



Comunicaciones del CIMAT

CLASSICAL EMPIRICAL PROCESS THEORY AND
WEIGHTED APPROXIMATIONS

David M. Mason

Comunicación del CIMAT No I-15-03/09-11-2015
(PE/CIMAT)



CIMAT

Classical Empirical Process Theory and Weighted Approximations

David M Mason

UNIVERSITY OF DELAWARE, NEWARK, DE 19716, USA
E-mail address: davidm@udel.edu

CIMAT

ABSTRACT. I will survey classical empirical process theory and its application to the study of the large sample properties of nonparametric statistics. I will discuss such important and useful results as the Skorohod representation, the KMT Brownian bridge approximation to the empirical and quantile processes and *weighted approximations* to these processes and their martingale generalizations. In the process I will demonstrate their use in proving central limit theorems for L-statistics and trimmed sums, among other examples, as well as in the derivation of the asymptotic distribution of various goodness-of-fit tests. This material forms what may be called the *Seattle-Hungarian School of Empirical Processes*. Much of the material will be taken from the text: *Empirical Processes with Applications to Statistics*, by Galen Shorack and Jon Wellner, and my research papers.

Contents

Preface	iii
Chapter 1. Introduction	1
Chapter 2. Basic Notions, Definitions and Facts	7
Chapter 3. Empirical Process Technology Circa 1972	21
Chapter 4. Intermediate Steps Towards Weighted Approximations	27
Chapter 5. The First Weighted Approximation	31
Chapter 6. The Mason and van Zwet Refinement of KMT	39
Chapter 7. Use of Theorem 3 to Study the Wasserstein Distance	43
Chapter 8. The Quantile Process	57
Chapter 9. A Non-Hungarian Construction	61
Chapter 10. Some Further Progress on Weighted Approximations	71
10.1. Exchangeable Processes	71
10.2. Nonparametric Bootstrapped Empirical Process	73
10.3. Weighted Bootstrapped Empirical Process	74
10.4. Approximation to Continuous Time Martingales	75
10.5. Some Final Remarks About Probability Spaces	80
Appendix A. Basic Large Sample Theory Facts	83
Appendix B. Properties of Order Statistics	91
Bibliography	97

Preface

These notes began as a series of lectures that the author delivered at Lunteren, Netherlands, to Dutch Ph.D. students in November of 1993. They were then reworked and elaborated with material taken from talks that the author had presented in seminars and conferences, and formed the basis of a week long course that he gave to graduate students at Masaryk University, Brno, Czech Republic in October of 2006. Later on he further augmented these notes and presented them as a mini-course on classical empirical process theory at the Centro de Investigación en Matemáticas (CIMAT), Guanajuato, Mexico, in February 2011 and in December 2014. This is an edited version of his CIMAT lectures. They are largely about the remarkable properties of the *uniform empirical distribution function* and its application to the study of nonparametric statistics and estimators. They have benefited by suggestions and corrections by the late Sándor Csörgő, Uwe Einmahl, Erich Haeusler, Péter Kevei, Claudia Kirch and Galen Shorack. A good source for any result in them for which a reference is not provided is the monograph: *Empirical Processes with Applications to Statistics* by Shorack and Wellner (1986).

CHAPTER 1

Introduction

I was introduced to the Seattle School of empirical processes in a sequence of courses on measure theoretic probability that I took in the Winter and Spring Quarters of 1973 taught by Galen Shorack at the University of Washington. By chance, Jon Wellner was a fellow student. Some 13 years later the Shorack and Wellner [SW] (1986) famous tome on empirical processes and their applications made its appearance. Both Wellner and I wrote our Ph.D. dissertations under Shorack's direction and subsequently devoted much of our research careers to the study of empirical processes.

Shorack's treatment of empirical process theory revolved around the uniform empirical distribution function, which had already shown itself by 1973 to be very useful in the study of nonparametric statistics. We shall begin with the definition of this function and indicate some of its uses in nonparametric statistics.

Uniform (0, 1) Random Variable

The Uniform (0, 1) random variable U has cumulative distribution function [cdf]

$$F_U(t) = \begin{cases} 1, & t \geq 1, \\ t, & 0 \leq t < 1, \\ 0, & t < 0. \end{cases}$$

Uniform Empirical Distribution

Let U, U_1, U_2, \dots , be independent Uniform (0, 1) random variables. For each integer $n \geq 1$ the *empirical distribution function* based on U_1, \dots, U_n , is defined to be

$$(1.1) \quad G_n(t) = \frac{1}{n} \sum_{i=1}^n 1\{U_i \leq t\}, \quad -\infty < t < \infty.$$

The uniform empirical distribution function G_n is a very good estimator of F_U .

Uniform Order Statistics

For future reference, for each integer $n \geq 1$, let $U_{1,n} \leq \dots \leq U_{n,n}$ denote the order statistics of U_1, \dots, U_n .

We shall next turn to some motivation for the study of the uniform empirical distribution.

Cumulative Distribution Function

Let X be a random variable with *cumulative distribution function* [cdf] F , i.e. $F(x) = P(X \leq x)$, for all x . The cdf F is right-continuous since for all x , by a property of probability measures, that says that if $A_{m+1} \subset A_m$, $m \geq 1$, then $\lim_{m \rightarrow \infty} P\{A_m\} = P\{A\}$, where $A = \bigcap_{m=1}^{\infty} A_m$, we have

$$P\{X \leq x\} = P\{\bigcap_{\varepsilon \searrow 0} \{X \leq x + \varepsilon\}\} = \lim_{\varepsilon \searrow 0} P\{X \leq x + \varepsilon\}.$$

Empirical Distribution Function

For each integer $n \geq 1$, let X, X_1, \dots, X_n be an i.i.d. random sample from a cdf F and consider the empirical cumulative distribution function F_n defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1\{X_i \leq x\}, \quad -\infty < x < \infty.$$

Order Statistics

Let $X_{1,n} \leq \dots \leq X_{n,n}$ denote the order statistics of X_1, \dots, X_n .

Kolmogorov-Smirnov Statistic

Consider the statistic

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|.$$

The statistic D_n is called the *Kolmogorov-Smirnov Goodness of Fit Test*. The following two theorems describe two of its important properties.

Theorem K1 *As long as F is continuous, the distribution of D_n does not depend on F , i.e. D_n is distribution free.*

Proof. It is not difficult to show that

$$D_n = \max_{1 \leq i \leq n} \max\left\{\left|\frac{i}{n} - F(X_{i,n})\right|, \left|\frac{i-1}{n} - F(X_{i,n})\right|\right\}.$$

Now since $F(X) =_d U$, we see that $(F(X_{1,n}), \dots, F(X_{n,n}))$ has the same joint distribution as $(U_{1,n}, \dots, U_{n,n})$, and thus D_n has the same distribution as

$$\max_{1 \leq i \leq n} \max \left\{ \left| \frac{i}{n} - U_{i,n} \right|, \left| \frac{i-1}{n} - U_{i,n} \right| \right\} = \sup_{0 \leq t \leq 1} |G_n(t) - t|.$$

□

Theorem K2 For all $x > 0$,

$$P \{ \sqrt{n} D_n \leq x \} \rightarrow H(x), \text{ as } n \rightarrow \infty,$$

where

$$H(x) = 1 - 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2 x^2}.$$

The Kolmogorov-Smirnov Goodness of Fit Test has the following two main uses:

A) Let F_0 be a continuous cdf. Suppose we want to test the *null hypothesis*

$$H_0 : F = F_0, \text{ where } F_0 \text{ is a specified distribution function}$$

versus the *alternative hypothesis*

$$H_1 : F \neq F_0 \text{ at level } \alpha.$$

We reject H_0 if $\sqrt{n} D_n > \varepsilon_{n,\alpha}$, where $P \{ \sqrt{n} D_n > \varepsilon_{n,\alpha} \} = \alpha$.

(B) Confidence Bands For F

Now when F is continuous

$$\begin{aligned} P \{ \sqrt{n} D_n \leq \varepsilon_{n,\alpha} \} &= 1 - \alpha \\ &= P \left\{ F_n(x) - \frac{\varepsilon_{n,\alpha}}{\sqrt{n}} \leq F(x) \leq F_n(x) + \frac{\varepsilon_{n,\alpha}}{\sqrt{n}}, \text{ for all } x \right\} \\ &= P \left\{ \max[0, F_n(x) - \frac{\varepsilon_{n,\alpha}}{\sqrt{n}}] \leq F(x) \leq \min[F_n(x) + \frac{\varepsilon_{n,\alpha}}{\sqrt{n}}, 1], \text{ for all } x \right\}. \end{aligned}$$

Thus a $(1 - \alpha)100\%$ confidence band for F is any observation of the random $(1 - \alpha)100\%$ band for F

$$\left[\max[0, F_n(x) - \frac{\varepsilon_{n,\alpha}}{\sqrt{n}}], \min[F_n(x) + \frac{\varepsilon_{n,\alpha}}{\sqrt{n}}, 1] \right], \text{ for all } x.$$

Notice that when F is continuous the critical value $\varepsilon_{n,\alpha}$ is determined on the assumption that $F = F_U$ and $F_n = G_n$.

The uniform empirical distribution function also arises naturally in the study of rank statistics. For example, consider:

The Mann-Whitney Wilcoxon test

Model: X, X_1, \dots, X_m i.i.d. F , continuous, and Y, Y_1, \dots, Y_n i.i.d. G , continuous.

The cdf G is said to be *stochastically larger* than F if for all x , $P\{Y > x\} \geq P\{X > x\}$ and $P\{Y > x\} > P\{X > x\}$ for at least one x .

A special case of this is when $Y =_d X + \Delta$, with $\Delta > 0$, where F is the cdf of X and G is the cdf of Y .

The cdf G is said to be *stochastically smaller* than F if for all x , $P\{Y \leq x\} \geq P\{X \leq x\}$ and $P\{Y \leq x\} > P\{X \leq x\}$ for at least one x .

A special case of this is when $Y =_d X + \Delta$, with $\Delta < 0$, where F is the cdf of X and G is the cdf of Y .

One may be interested in testing one of the following three hypothesis testing situations:

- (I) $H_0 : F = G$ versus $H_1 : G$ is *stochastically larger* than F ;
- (II) $H_0 : F = G$ versus $H_1 : G$ is *stochastically smaller* than F ;
- (III) $H_0 : F = G$ versus $H_1 : G \neq F$.

A widely used statistic for testing these three hypotheses is the *Wilcoxon statistic*

$$W = \sum_{j=1}^n R_j,$$

where R_1, \dots, R_n denote the ranks of Y_1, \dots, Y_n among the combined sample $X_1, \dots, X_m, Y_1, \dots, Y_n$. Note that for each $1 \leq j \leq n$

$$R_j = \sum_{i=1}^m 1\{X_i \leq Y_j\} + \sum_{k=1}^n 1\{Y_k \leq Y_j\}.$$

Clearly in hypothesis testing situation (I) we reject H_0 if W is too large; in situation (II), we reject H_0 if W is too small; and in situation (III), we reject H_0 if W is too large or small.

Let F_m^X be the empirical distribution function based on X_1, \dots, X_m and F_n^Y be the empirical distribution function based on Y_1, \dots, Y_n . Further let

$$H_{m+n} = (mF_m^X + nF_n^Y) / (m + n).$$

We see that

$$W = (m + n) n \int_{\mathbb{R}} H_{m+n}(x) dF_n^Y(x),$$

which, when $F = G$, has the same distribution as when $F = G = F_U$, that is, W is *distribution free*. To see this note that when F is continuous and $F = G$

$$W =_d (m+n)n \int_0^1 G_{m+n}(x) dG_n^Y(x),$$

$$G_{m+n}(x) = \frac{1}{m+n} \sum_{i=1}^{m+n} 1\{U_i \leq x\} \quad \text{and} \quad G_n^Y(x) = \frac{1}{n} \sum_{i=1}^n 1\{U_{i+m} \leq x\},$$

where U_1, \dots, U_{m+n} are i.i.d. Uniform $(0, 1)$.

It turns out that the Wilcoxon statistic is a special case of a Chernoff-Savage (1958) statistic. Pyke and Shorack (1968) published a path breaking paper that showed how the asymptotic properties of the uniform empirical distribution can be used to prove central limit theorems for such statistics. The main message that I received from Shorack's lectures in 1973 was that the study of the asymptotic distribution of nonparametric statistics very often reduces to the study of the asymptotic distribution of functionals of the *uniform empirical distribution function*. In the first part of these notes we discuss the properties of G_n that are important to prove such results.

My research direction took a radical turn after I attended a series of lectures presented by Miklós Csörgő based on his recent monograph with Pál Révész, *Strong approximations in probability and statistics*, at an N.S.F. Regional Conference on Quantile Processes held in July, 1981 at Texas A&M University. A year later I met Miklós' younger brother Sándor at the *Conference on Limit Theorems in Probability and Statistics* in June, 1982 at Veszprém, Hungary. This was both my introduction and induction into the Hungarian school of empirical processes. The next year I spent 6 months visiting Sándor at the University of Szeged, where, working with Miklós and Lajos Horváth, Sándor's Ph.D. student at the time, we created a theory of weighted approximations to the uniform empirical and quantile processes, which in the following years I developed in collaboration with Sándor and Erich Haeusler into a general quantile-transform-empirical-process approach to limit theorems. Much of my research over roughly the next 20 years, the bulk of it done with talented coauthors, was devoted to extensions and applications of this theory to large sample problems in probability theory and statistics. The second part of these notes describe the main features of this work. Also included are two appendices. The first is a purview of elementary large sample theory to fix notation and definition, and the second derives some basic facts about

uniform order statistics that are used in the proofs. In total, this material forms a survey of what may be called the *Seattle-Hungarian School of Empirical Processes*.



CIMAT

CHAPTER 2

Basic Notions, Definitions and Facts

We shall begin with a description and derivation of some of the salient properties on the uniform empirical distribution function.

Glivenko–Cantelli Theorem

The *Glivenko–Cantelli Theorem* says that

$$(2.1) \quad \sup_{0 \leq t \leq 1} |G_n(t) - t| \rightarrow 0, \text{ a.s., as } n \rightarrow \infty.$$

Actually more is known.

Dvoretzky, Kiefer and Wolfowitz Inequality

The *Dvoretzky, Kiefer and Wolfowitz (1956) Inequality* says even more, namely that for some constant $K > 0$, all $n \geq 1$ and any $r > 0$

$$(2.2) \quad P \left\{ \sup_{0 \leq t \leq 1} |G_n(t) - t| > r \right\} \leq K \exp(-2r^2n).$$

Massart (1990) has shown that one can choose $K = 2$. (After Massart announced this remarkable result in an invited talk presented at the 18th European Meeting of Statisticians in August 1988 in Berlin, DDR, Sándor Csörgő and I stood up and applauded.) Notice that when this inequality is combined with the Borel–Cantelli lemma, we get that

$$(2.3) \quad \sup_{0 \leq t \leq 1} |\tilde{G}_n(t) - t| \rightarrow 0, \text{ a.s., as } n \rightarrow \infty,$$

for any sequence $\{\tilde{G}_n\}_{n \geq 1}$ of equivalent versions of $\{G_n\}_{n \geq 1}$, meaning that $\tilde{G}_n =_d G_n$, for each $n \geq 1$.

The Glivenko-Cantelli Theorem and the DKW Inequality for F_n

It turns out that if X has cdf F then

$$\{1\{X \leq x\}, x \in \mathbb{R}\} =_d \{1\{U \leq F(x)\}, x \in \mathbb{R}\}.$$

From this it follows that as a process in $n \geq 1$,

$$(2.4) \quad \{F_n(x), x \in \mathbb{R}\}_{n \geq 1} =_d \{G_n(F(x)), x \in \mathbb{R}\}_{n \geq 1}.$$

Thus from (2.1) and (2.2) we get

$$(2.5) \quad \sup_{-\infty < x < \infty} |F_n(x) - F(x)| \rightarrow 0, \text{ a.s., as } n \rightarrow \infty$$

and for all $r > 0$,

$$(2.6) \quad P \left\{ \sup_{-\infty < x < \infty} |F_n(x) - F(x)| \geq r \right\} \leq K \exp(-2r^2n).$$

The Glivenko-Cantelli theorem can be proved directly using the strong law of large numbers, which implies for all x

$$F_n(x) \rightarrow F(x), \text{ a.s., as } n \rightarrow \infty,$$

combined with an elementary argument using a grid and right-continuity of F .

A Useful Probabilistically Equivalent Version of G_n

Here is a very useful probabilistically equivalent version of G_n . Let $\omega_1, \omega_2, \dots$, be a sequence of i.i.d. exponential random variables with mean 1. Set for $j \geq 1$, $S_j = \omega_1 + \dots + \omega_j$. One has for each integer $n \geq 1$

$$(2.7) \quad \left(\frac{S_1}{S_{n+1}}, \frac{S_2}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}} \right) =_d (U_{1,n}, U_{2,n}, \dots, U_{n,n}),$$

where $U_{1,n} \leq U_{2,n} \leq \dots \leq U_{n,n}$ are the order statistics of n i.i.d. Uniform $(0, 1)$ random variables. For a proof of this fact see Theorem 4B in Appendix B. We see then that for each $n \geq 1$,

$$(2.8) \quad \tilde{G}_n(t) = \frac{1}{n} \sum_{i=1}^n 1 \left\{ \frac{S_i}{S_{n+1}} \leq t \right\}, \quad -\infty < t < \infty,$$

has the same distribution as

$$G_n(t) = \frac{1}{n} \sum_{i=1}^n 1 \{U_{i,n} \leq t\}, \quad -\infty < t < \infty,$$

which is the uniform empirical distribution function (1.1). Thus for each $n \geq 1$, \tilde{G}_n is a probabilistically equivalent version of G_n . However, as a sequence in $n \geq 1$,

$$\left\{ \tilde{G}_n \right\}_{n \geq 1} \neq_d \left\{ G_n \right\}_{n \geq 1}.$$

The reason for this is that the representation (2.7) does not hold as a sequence in $n \geq 1$.

A Useful Martingale

Here we point out that there is continuous time martingale lurking here. Introduce the filtration

$$\begin{aligned}\mathcal{F}_n(s) &= \sigma\{1\{U_i \leq r\}, i = 1, \dots, n, 0 \leq r \leq s\}, \text{ for } 0 < s < 1, \\ \mathcal{F}_n(0) &= \{\emptyset, \Omega\}.\end{aligned}$$

Martingale Fact

$$M_n(t) = \frac{G_n(t) - t}{1 - t}, 0 \leq t < 1,$$

is a martingale, that is, for all $0 \leq s \leq t < 1$,

$$(2.9) \quad E(M_n(t) | \mathcal{F}_n(s)) = M_n(s).$$

Proof. Choose $0 \leq s \leq t < 1$. We see that

$$\frac{\sum_{i=1}^n 1\{U_i > t\}}{1 - t} = \frac{\sum_{i=1}^n 1\{U_i > s\} - \sum_{i=1}^n 1\{s < U_i \leq t\}}{1 - t}.$$

Now $\sum_{i=1}^n 1\{s < U_i \leq t\} | \sum_{i=1}^n 1\{U_i > s\} = m$ is binomial with parameters m and $p = \frac{t-s}{1-s}$. Hence

$$E\left(\sum_{i=1}^n 1\{s < U_i \leq t\} | \sum_{i=1}^n 1\{U_i > s\} = m\right) = \frac{m(t-s)}{1-s},$$

which gives

$$\begin{aligned}E\left(\frac{\sum_{i=1}^n 1\{U_i > t\}}{1 - t} | \sum_{i=1}^n 1\{U_i > s\} = m\right) &= \frac{m}{1 - t} - \frac{m(t-s)}{(1-s)(1-t)} \\ &= \frac{m(1-t)}{(1-s)(1-t)} = \frac{m}{1-s} = \frac{\sum_{i=1}^n 1\{U_i > s\}}{1-s}.\end{aligned}$$

Noting that

$$M_n(t) = \frac{G_n(t) - t}{1 - t} = 1 - \frac{\sum_{i=1}^n 1\{U_i > t\}}{n(1 - t)},$$

this implies that for $0 \leq s \leq t < 1$,

$$E(M_n(t) | \mathcal{F}_n(s)) = 1 - \frac{\sum_{i=1}^n 1\{U_i > s\}}{n(1 - s)} = M_n(s),$$

that is, (2.9) holds. \square

An Easier to Establish Version of the DKW Inequality

Armed with the martingale fact we shall now prove a less precise, but easier to establish version of the DKW inequality. (I first saw the idea of this proof in Jon Wellner's 1975 Ph.D. thesis. Also see Corollary

2 of Wellner (1977).) First of all notice that since $(U_1, \dots, U_n) =_d (1 - U_1, \dots, 1 - U_n)$,

$$(2.10) \quad \sup_{0 \leq t \leq 1/2} |G_n(t) - t| =_d \sup_{1/2 \leq t \leq 1} |G_n(t) - t|.$$

Next, since the right-continuous process

$$M_t := \frac{G_n(t) - t}{1 - t}, \quad 0 \leq t \leq 1/2$$

is a martingale, for any $u > 0$ the processes $Y_t^{(1)} = \exp(uM_t)$ and $Y_t^{(2)} = \exp(-uM_t)$ are submartingales. Thus by Doob's inequality for right-continuous submartingales (see the remark below), for any $z \geq 0$ and $u > 0$

$$P \left\{ \sup_{0 \leq t \leq 1/2} uM_t > uz \right\} = P \left\{ \sup_{0 \leq t \leq 1/2} Y_t^{(1)} > e^{uz} \right\} \leq EY_{1/2}^{(1)} \exp(-uz)$$

and

$$P \left\{ \sup_{0 \leq t \leq 1/2} -uM_t > uz \right\} = P \left\{ \sup_{0 \leq t \leq 1/2} Y_t^{(2)} > e^{uz} \right\} \leq EY_{1/2}^{(2)} \exp(-uz).$$

Now

$$EY_{1/2}^{(1)} = EY_{1/2}^{(2)} = \left(\left(\exp\left(\frac{u}{n}\right) + \exp\left(-\frac{u}{n}\right) \right) / 2 \right)^n,$$

which by using the elementary inequality

$$(\exp(u) + \exp(-u)) / 2 \leq \exp(u^2/2),$$

gives

$$\left(\left(\exp\left(\frac{u}{n}\right) + \exp\left(-\frac{u}{n}\right) \right) / 2 \right)^n \leq \exp\left(\frac{u^2}{2n}\right).$$

Thus by setting $u = zn$, we get

$$\begin{aligned} EY_{1/2}^{(1)} \exp(-uz) &= EY_{1/2}^{(2)} \exp(-uz) \\ &\leq \exp\left(\frac{u^2}{2n}\right) \exp(-uz) = \exp(-nz^2/2). \end{aligned}$$

Putting everything together we have

$$\begin{aligned} &P \left\{ \sup_{0 \leq t \leq 1/2} |G_n(t) - t| > z \right\} \\ &\leq P \left\{ \sup_{0 \leq t \leq 1/2} \frac{G_n(t) - t}{1 - t} > z \right\} + P \left\{ \sup_{0 \leq t \leq 1/2} \frac{t - G_n(t)}{1 - t} > z \right\} \\ &\leq 2 \exp(-nz^2/2). \end{aligned}$$

By (2.10) this implies that

$$P \left\{ \sup_{0 \leq t \leq 1} |G_n(t) - t| > z \right\} \leq 4 \exp(-nz^2/2).$$

Remark Doob's inequality for right-continuous submartingales says that for a non-negative submartingale Y_s , $a \leq s \leq b$, for all $\lambda > 0$,

$$P \left\{ \sup_{a \leq s \leq b} Y_s > \lambda \right\} \leq \lambda^{-1} EY_b.$$

(See, for instance, Inequality 5 on page 874 of SW(1986).)

Linearity of G_n

A very useful property of G_n is its linearity in various senses.

Fact 1. (Linearity in probability) For all $\varepsilon > 0$ there exists a $\lambda > 1$ such that for all $n \geq 1$,

$$P \left\{ \frac{1}{\lambda} < \frac{G_n(t)}{t} < \lambda \text{ for all } U_{1,n} \leq t \leq 1 \right\} \geq 1 - \varepsilon,$$

where $U_{1,n}$ denotes the minimum of U_1, \dots, U_n .

Fact 2. (Poisson approximation) There exists a standard rate one Poisson process $N(x)$, $x \geq 0$, and a sequence $\{\tilde{G}_n\}_{n \geq 1}$ of probabilistically equivalent versions of $\{G_n\}$ such that

$$\sup_{0 \leq x \leq n} \left| \frac{n\tilde{G}_n(x/n)}{x} - \frac{N(x)}{x} \right| \rightarrow_P 0, \text{ as } n \rightarrow \infty.$$

Fact 3 (Linearity close to zero, in probability) For any sequence $a_n > 0$, such that $a_n \rightarrow 0$ and $na_n \rightarrow \infty$, as $n \rightarrow \infty$,

$$\sup_{a_n \leq t \leq 1} \left| \frac{G_n(t)}{t} - 1 \right| \rightarrow_P 0, \text{ as } n \rightarrow \infty.$$

Fact 4. (Linearity close to zero, almost surely) For any sequence $a_n > 0$, such that $a_n \rightarrow 0$ and $na_n/\log \log n \rightarrow \infty$ as $n \rightarrow \infty$,

$$\sup_{a_n \leq t \leq 1} \left| \frac{G_n(t)}{t} - 1 \right| \rightarrow 0, \text{ a.s., as } n \rightarrow \infty.$$

A proof of Fact 1 is given in Pyke and Shorack (1968). We shall provide our own proof here which will give you an example of the kind of arguments used to prove *linearity* results. For the rest of these facts consult Wellner (1978) and the references therein. All of them are found in SW(1986).

The following two results establish Fact 1.

Daniels (1945) For all $n \geq 1$ and $\lambda \geq 1$

$$(2.11) \quad P \left\{ \sup_{0 \leq s < 1} \frac{G_n(s)}{s} \leq \lambda \right\} = 1 - \lambda^{-1}.$$

Proof. We shall prove this by induction. Note that when $n = 1$

$$P \left\{ \sup_{0 \leq s < 1} \frac{G_1(s)}{s} \leq \lambda \right\} = P \left\{ \frac{1}{U} \leq \lambda \right\} = 1 - \lambda^{-1}.$$

Now consider

$$P \left\{ \sup_{0 \leq s < 1} \frac{G_{n+1}(s)}{s} \leq \lambda \right\}.$$

Let $U_{1,n+1} \leq \dots \leq U_{n+1,n+1}$ denote the order statistics of U_1, \dots, U_{n+1} .

For $0 \leq s < U_{n+1,n+1}$

$$G_{n+1}(s) = \frac{1}{n+1} \sum_{i=1}^n 1 \{U_{i,n+1} \leq s\}$$

and for $U_{n+1,n+1} \leq s < 1$, $G_{n+1}(s) = 1$. Thus

$$\begin{aligned} & \sup_{0 \leq s < 1} \frac{G_{n+1}(s)}{s} \\ &= \max \left\{ \sup_{0 \leq s < U_{n+1,n+1}} \frac{1}{s(n+1)} \sum_{i=1}^n 1 \{U_{i,n+1} \leq s\}, \frac{1}{U_{n+1,n+1}} \right\}. \end{aligned}$$

Next notice that

$$\begin{aligned} & \sup_{0 \leq s < U_{n+1,n+1}} \frac{1}{s(n+1)} \sum_{i=1}^n 1 \{U_{i,n+1} \leq s\} \\ &= \frac{n}{(n+1)U_{n+1,n+1}} \sup_{0 \leq t < 1} \frac{1}{tn} \sum_{i=1}^n 1 \left\{ \frac{U_{i,n+1}}{U_{n+1,n+1}} \leq t \right\}, \end{aligned}$$

and since $\left(\frac{U_{1,n+1}}{U_{n+1,n+1}}, \dots, \frac{U_{n,n+1}}{U_{n+1,n+1}} \right) =_d (U_{1,n}, \dots, U_{n,n})$,

$$\sup_{0 \leq t < 1} \frac{1}{tn} \sum_{i=1}^n 1 \left\{ \frac{U_{i,n+1}}{U_{n+1,n+1}} \leq t \right\} =_d \sup_{0 \leq t < 1} \frac{G_n(t)}{t}.$$

An application of Theorem 4B in Appendix B shows that $\frac{U_{i,n+1}}{U_{n+1,n+1}}$, $i = 1, \dots, n$ are independent of $U_{n+1,n+1}$. Moreover, a special case of Theorem 3B gives that $U_{n+1,n+1}$ has density

$$g_{n+1}(u) = (n+1)u^n 1 \{u \in [0, 1]\}.$$

By induction $P \left\{ \sup_{0 \leq s < 1} \frac{G_n(s)}{s} \leq \lambda \right\} = 1 - \lambda^{-1}$. Therefore

$$\begin{aligned} & P \left\{ \max \left\{ \sup_{0 \leq s < U_{n+1, n+1}} \frac{1}{s(n+1)} \sum_{i=1}^n 1 \{U_{i, n+1} \leq s\}, \frac{1}{U_{n+1, n+1}} \right\} \leq \lambda \right\} \\ &= P \left\{ \max \left\{ \frac{n}{(n+1)U_{n+1, n+1}} \sup_{0 \leq t < 1} \frac{G_n(t)}{t}, \frac{1}{U_{n+1, n+1}} \right\} \leq \lambda \right\} \\ &= P \left\{ \sup_{0 \leq t < 1} \frac{G_n(t)}{t} \leq \frac{(n+1)\lambda U_{n+1, n+1}}{n} \text{ and } U_{n+1, n+1} \geq \frac{1}{\lambda} \right\} \end{aligned}$$

Now this last probability is equal to

$$\begin{aligned} & \int_{\lambda^{-1}}^1 \left(1 - \frac{n}{u(n+1)\lambda} \right) g_{n+1}(u) du \\ &= \int_{\lambda^{-1}}^1 \left(1 - \frac{n}{u(n+1)\lambda} \right) (n+1) u^n du \\ &= (n+1) \int_{\lambda^{-1}}^1 u^n du - \frac{n}{\lambda} \int_{\lambda^{-1}}^1 u^{n-1} du \\ &= 1 - \left(\frac{1}{\lambda} \right)^{n+1} - \frac{1}{\lambda} \left(1 - \left(\frac{1}{\lambda} \right)^n \right) = 1 - \lambda^{-1}. \end{aligned}$$

□

Here is how the second half of Fact 1 is proved. We see that for any $U_{i, n} \leq t < U_{i+1, n}$, $i = 1, \dots, n$, where $U_{n+1, n} = 1$,

$$\frac{G_n(t)}{t} \geq \frac{i}{nU_{i+1, n}}.$$

Recall the distributional representation given in (2.7) for each $n \geq 1$ for the uniform order statistics $U_{1, n}, U_{2, n}, \dots, U_{n, n}$. Thus

$$\inf_{U_{1, n} \leq t \leq 1} \frac{G_n(t)}{t} \geq \min_{1 \leq i \leq n} \frac{i}{nU_{i+1, n}} =_d \min_{1 \leq i \leq n} \frac{S_{n+1}^i}{nS_{i+1}} =: V_n.$$

Our proof will be based on the following elementary fact.

Elementary Fact Let $\{Y_j\}_{j \geq 1}$ be a sequence of strictly positive random variables on a probability space (Ω, A, P) such that for some $c > 0$, $Y_j \rightarrow c$, a.s., as $j \rightarrow \infty$. Then $P\{0 < V < \infty\} = 1$, where $V = \inf_{j \geq 1} Y_j$.

Proof By assumption for almost every $\omega \in \Omega$ there exists a $j(\omega) > 1$ such that $Y_j \geq c/2$ for all $j > j(\omega)$. Thus for almost every ω , $Y_j > 0$ for all $1 \leq j \leq j(\omega)$ and $Y_j \geq c/2$ for all $j > j(\omega)$. Hence $P\{0 < V < \infty\} = 1$. □

Apply the elementary fact with $Y_j = j/S_{j+1}$, $j \geq 1$, and noting that $Y_j \rightarrow 1$, a.s, as $j \rightarrow \infty$, we see that $P\{0 < V < \infty\} = 1$, where $V = \inf_{j \geq 1} Y_j$. Next observe that

$$V_n = \frac{S_{n+1}}{n} \min_{1 \leq i \leq n} \frac{i}{S_{i+1}} \rightarrow_d V.$$

Thus for all $\rho > 0$ such that ρ is a continuity point of the cdf of V ,

$$P\{V_n > \rho\} \rightarrow P\{V > \rho\}.$$

Since $P\{V > 0\} = 1$, for all $0 < \varepsilon < 1$ there exists a $\rho > 0$ satisfying $P\{V > \rho\} > 1 - \varepsilon/4$ and an $n(\varepsilon) > 1$ such that $P\{V_n > \rho\} > 1 - \varepsilon/2$ for all $n > n(\varepsilon)$. Now since $P\{V_n > 0\} = 1$ for all $n \geq 1$, there exists a $0 < \rho' < \rho$ such that $P\{V_n > \rho'\} > 1 - \varepsilon/2$ for each $1 \leq n \leq n(\varepsilon)$. Therefore $P\{V_n > \rho'\} > 1 - \varepsilon/2$ for all $n \geq 1$. Hence there exists $\lambda > 1$ such that

(2.12)

$$P\left\{\frac{1}{\lambda} < \frac{G_n(t)}{t} \text{ for all } U_{1,n} \leq t \leq 1\right\} \geq P\{V_n > \lambda^{-1}\} \geq 1 - \varepsilon/2.$$

We see that for any $\lambda > 1 \vee (2/\varepsilon)$ satisfying (2.12) and by (2.11)

$$\begin{aligned} & P\left\{\frac{1}{\lambda} < \frac{G_n(t)}{t} < \lambda \text{ for all } U_{1,n} \leq t \leq 1\right\} \\ & \geq P\left\{\frac{1}{\lambda} < \frac{G_n(t)}{t} \text{ for all } U_{1,n} \leq t \leq 1\right\} - P\left\{\sup_{0 \leq s < 1} \frac{G_n(s)}{s} \geq \lambda\right\} \\ & \geq 1 - \varepsilon/2 - \lambda^{-1} > 1 - \varepsilon. \end{aligned}$$

Thus completes the prove of Fact 1.

Uniform Empirical Process

The *uniform empirical process* based on U_1, \dots, U_n , $n \geq 1$, is defined to be

$$(2.13) \quad \alpha_n(t) = \sqrt{n}\{G_n(t) - t\}, \quad t \in [0, 1].$$

It is readily checked that

$$\alpha_n(0) = \alpha_n(1) = 0, \quad E\alpha_n(t) = 0 \text{ for all } t \in [0, 1]$$

and

$$\text{Cov}(\alpha_n(s), \alpha_n(t)) = s \wedge t - st, \quad s, t \in [0, 1],$$

where $s \wedge t = \min(s, t)$. The multivariate central limit theorem implies that for any choice of $t_1, \dots, t_m \in [0, 1]$, $m \geq 1$,

$$(2.14) \quad (\alpha_n(t_1), \dots, \alpha_n(t_m)) \rightarrow_d (Z_1, \dots, Z_m), \text{ as } n \rightarrow \infty,$$

where (Z_1, \dots, Z_m) is multivariate normal with mean vector zero and

$$\text{cov}(Z_i, Z_j) = t_i \wedge t_j - t_i t_j, \quad 1 \leq i, j \leq m.$$

Much more than (2.14) can be said.

Brownian Bridge

A *Brownian Bridge* is a continuous Gaussian process on $[0, 1]$ such that

$$B(0) = B(1) = 0, \quad EB(t) = 0 \text{ for all } t \in [0, 1]$$

and

$$\text{Cov}(B(s), B(t)) = s \wedge t - st, \quad s, t \in [0, 1].$$

The Brownian bridge B has the following representation:

$$B(t) = W(t) - tW(1), \quad t \in [0, 1],$$

where W is a standard Wiener process, i.e. W is a continuous Gaussian process on $[0, 1]$ with $W(0) = 0$, $EW(t) = 0$ for $0 \leq t \leq 1$ and $E(W(t)W(s)) = s \wedge t$, $s, t \in [0, 1]$. (For more about the Brownian bridge see pages 182–184 of Hájek and Šidák (1967).)

Doob (1949) was the first to notice that for all $x > 0$,

$$P \left\{ \sup_{0 \leq s \leq 1} |B(s)| \leq x \right\} = H(x),$$

where H is as in Theorem K2. Noting that

$$P \left\{ \sqrt{n}D_n \leq x \right\} = P \left\{ \sup_{0 \leq s \leq 1} |\alpha_n(s)| \leq x \right\},$$

we see then by Theorem K2 that

$$P \left\{ \sup_{0 \leq s \leq 1} |\alpha_n(s)| \leq x \right\} \rightarrow P \left\{ \sup_{0 \leq s \leq 1} |B(s)| \leq x \right\}, \text{ as } n \rightarrow \infty.$$

Donsker's famous and powerful functional central limit theorem implies that $\mathbb{T}(\alpha_n)$ converges in distribution to $\mathbb{T}(B)$, where B is a Brownian bridge, for a suitable class of functionals \mathbb{T} , which includes $\mathbb{T}(\alpha_n) = \sup_{0 \leq s \leq 1} |\alpha_n(s)|$. (Consult Billingsley (1968) for a proof of Donsker's theorem.) We shall soon see that much more can be said about how α_n converges to B .

The Skorohod Representation Theorem

The *Skorohod Representation Theorem* for the uniform empirical process α_n says that there exists a sequence $\{\tilde{\alpha}_n\}_{n \geq 1}$ of probabilistically

equivalent versions of $\{\alpha_n\}$, meaning $\tilde{\alpha}_n =_d \alpha_n$, for each $n \geq 1$, and a fixed Brownian bridge B such that

$$(2.15) \quad \sup_{0 \leq t \leq 1} |\tilde{\alpha}_n(t) - B(t)| \rightarrow 0, \text{ a.s., as } n \rightarrow \infty.$$

(See Theorem 1 of SW (1986) or pages 757-758 of Pyke and Shorack (1968).) Notice that this implies that $\mathbb{T}(\alpha_n) \rightarrow_d \mathbb{T}(B)$ for any functional on the space of bounded functions on $[0, 1]$ that is continuous in the supremum norm.

The General Empirical Process

The general empirical process based on F_n is

$$\alpha_{n,F}(x) = \sqrt{n} \{F_n(x) - F(x)\}, \quad -\infty < x < \infty.$$

Notice that for any x, y

$$\text{cov}(\alpha_{n,F}(x), \alpha_{n,F}(y)) = F(x \wedge y) - F(x)F(y)$$

and we get from (2.4), the Skorohod representation

$$(2.16) \quad \sup_{-\infty < x < \infty} |\tilde{\alpha}_{n,F}(x) - B(F(x))| \rightarrow 0, \text{ a.s., as } n \rightarrow \infty,$$

where for each $n \geq 1$, $\tilde{\alpha}_{n,F}(x) := \tilde{\alpha}_n(F(x))$ and $\{\tilde{\alpha}_{n,F}(x)\}_{x \in (-\infty, \infty)} =_d \{\alpha_{n,F}(x)\}_{x \in (-\infty, \infty)}$.

Birnbaum–Marshall Inequality (1961)

The first step towards a *weighted approximation* to α_n was based upon the following application of the Birnbaum–Marshall inequality, which says that for any positive function q on $(0, 1)$, increasing on $(0, 1/2]$ and decreasing on $[1/2, 1)$ such that

$$(2.17) \quad \int_0^1 \frac{du}{q^2(u)} < \infty,$$

there exists a constant C such that for all $r > 0$ and $0 < \delta < 1/2$,

$$(2.18) \quad \begin{aligned} &P \left\{ \sup_{0 < s \leq \delta} |\alpha_n(s)|/q(s) > r \right\} + P \left\{ \sup_{0 < s \leq \delta} |\alpha_n(1-s)|/q(1-s) > r \right\} \\ &\leq \frac{C}{r^2} \int_0^\delta \left(\frac{1}{q^2(u)} + \frac{1}{q^2(1-u)} \right) du \end{aligned}$$

and

$$P \left\{ \sup_{0 < s \leq \delta} |B(s)|/q(s) > r \right\} + P \left\{ \sup_{0 < s \leq \delta} |B(1-s)|/q(1-s) > r \right\}$$

$$(2.19) \quad \leq \frac{C}{r^2} \int_0^\delta \left(\frac{1}{q^2(u)} + \frac{1}{q^2(1-u)} \right) du.$$

(These inequalities actually follow from the Hájek-Rényi inequality. See below.) This says of course that for all $r > 0$

$$(2.20) \quad \lim_{\delta \searrow 0} \limsup_{n \rightarrow \infty} \left[P \left\{ \sup_{0 < s \leq \delta} \frac{|\alpha_n(s)|}{q(s)} > r \right\} + P \left\{ \sup_{0 < s \leq \delta} \frac{|\alpha_n(1-s)|}{q(1-s)} > r \right\} \right] = 0$$

and

$$(2.21) \quad \lim_{\delta \searrow 0} \left[P \left\{ \frac{\sup_{0 < s \leq \delta} |B(s)|}{q(s)} > r \right\} + P \left\{ \frac{\sup_{0 < s \leq \delta} |B(1-s)|}{q(1-s)} > r \right\} \right] = 0.$$

Let q be any positive function on $(0, 1)$, increasing on $(0, 1/2]$ and decreasing on $[1/2, 1)$ such that (2.17) holds. Using this inequality one can show that for any probability space such that

$$\sup_{0 \leq t \leq 1} |\tilde{\alpha}_n(t) - B_n(t)| \rightarrow_p 0,$$

where $\{\tilde{\alpha}_n\}$ is a sequence of probabilistically equivalent versions of $\{\alpha_n\}$, and $\{B_n\}$ is an appropriate sequence of Brownian bridges, we have the following *weighted convergence* in probability to zero

$$(2.22) \quad \sup_{0 < t < 1} |\tilde{\alpha}_n(t) - B_n(t)| / q(t) \rightarrow_p 0.$$

We shall return to this soon.

A Digression about the Hájek-Rényi Inequality

Let ξ_1, \dots, ξ_m be martingale difference sequence, where each ξ_k has mean 0 and finite variance σ_k^2 . (This means that S_j , $j = 0, \dots, m$, forms a martingale, where $S_0 = 0$ and for $j = 1, \dots, m$, $S_j = \sum_{i=1}^j \xi_i$.) Then for any non-increasing sequence of positive constants c_1, \dots, c_m , $m > 1$, we have for any $\varepsilon > 0$ and positive integers $n < m$

$$P \left\{ \max_{n \leq k \leq m} c_k \left| \sum_{j=1}^k \xi_j \right| \geq \varepsilon \right\} \leq \frac{1}{\varepsilon^2} \left(c_n^2 \sum_{k=1}^n \sigma_k^2 + \sum_{k=n+1}^m c_k^2 \sigma_k^2 \right).$$

This is Theorem 8 (iii) on page 243 of Chow and Teicher (1978). The original version of the Hájek-Rényi inequality was stated for sums of independent random variables.

To see how the Birnbaum and Marshall inequality follows from the Hájek-Rényi inequality, let q be a positive and increasing function on $(0, 1/2]$ and choose any $N \geq 4$ and $0 < \delta < 1/2$. Set $c_k = 1/q(k/N)$ for

$k = 1, \dots, \lceil N\delta \rceil \vee 2$ and $\xi_j = \frac{\alpha_n(j/N)}{1-j/N} - \frac{\alpha_n((j-1)/N)}{1-(j-1)/N}$ for $j = 1, \dots, \lceil N\delta \rceil \vee 2$. We see that for $k = 1, \dots, \lceil N\delta \rceil \vee 2$,

$$\sum_{j=1}^k \xi_j = \frac{\alpha_n(k/N)}{1-k/N} = \frac{\sqrt{n}M_n(k/N)}{1-k/N},$$

which by the above forms a martingale difference sequence. Therefore by the Hájek-Rényi inequality with $n = 1$ and $m = \lceil N\delta \rceil$,

$$\begin{aligned} & P \left\{ \max_{1 \leq k \leq \lceil N\delta \rceil \vee 2} \left| \frac{\alpha_n(k/N)}{(1-k/N)q(k/N)} \right| \geq \varepsilon \right\} \\ & \leq \frac{1}{\varepsilon^2} \left(q^{-2}(1/N)\sigma_1^2 + \sum_{k=2}^{\lceil N\delta \rceil \vee 2} q^{-2}(k/N)\sigma_k^2 \right). \end{aligned}$$

Now for some $D > 0$ independent of $N \geq 4$, $0 < \delta < 1/2$ and $k = 1, \dots, \lceil N\delta \rceil \vee 2$,

$$\sigma_k^2 = \text{Var} \left(\frac{\alpha_n(k/N)}{1-k/N} - \frac{\alpha_n((k-1)/N)}{1-(k-1)/N} \right) \leq \frac{D}{N}.$$

Thus

$$\begin{aligned} & q^{-2}(1/N)\sigma_1^2 + \sum_{k=2}^{\lceil N\delta \rceil \vee 2} q^{-2}(k/N)\sigma_k^2 \\ & \leq \frac{D}{\varepsilon^2 N} \sum_{k=1}^{\lceil N\delta \rceil \vee 2} q^{-2}(k/N) \leq \frac{D}{\varepsilon^2} \int_0^{(\lceil N\delta \rceil \vee 2)/N} \frac{du}{q^2(u)}. \end{aligned}$$

Now by letting $N \rightarrow \infty$, and noting that, with probability 1, for each $n \geq 1$, U_1, \dots, U_n miss the discontinuity points of q , we can conclude that for some constant $C' > 0$ independent $0 < \delta < 1/2$ and $\varepsilon > 0$,

$$P \left\{ \sup_{0 < s \leq \delta} \left| \frac{\alpha_n(s)}{q(s)} \right| \geq \varepsilon \right\} \leq P \left\{ \sup_{0 < s \leq \delta} \left| \frac{\alpha_n(s)}{(1-s)q(s)} \right| \geq \varepsilon \right\} \leq \frac{C'}{\varepsilon^2} \int_0^\delta \frac{du}{q^2(u)}.$$

Quantile Function

Let X be a random variable with cdf F , i.e. $F(x) = P(X \leq x)$. The *inverse* or *quantile function* of F , written Q , is defined for $s \in (0, 1)$ to be

$$(2.23) \quad Q(s) = \inf\{x : F(x) \geq s\}.$$

($\inf\{x : F(x) \geq s\}$ means infimum $\{x : F(x) \geq s\}$, that is, $Q(s)$ is that value for which $F(Q(s) + \varepsilon) \geq s$ for $\varepsilon > 0$ and $F(Q(s) - \varepsilon) < s$ for $\varepsilon > 0$.) The quantile function Q has a number of useful properties.

- (i) Q is nondecreasing;
- (ii) for any $s \in (0, 1)$, $Q(s) \leq x$ if and only if $F(x) \geq s$;
- (ii') for any $s \in (0, 1)$, $Q(s) > x$ if and only if $F(x) < s$;
- (iii) Q left-continuous on $(0, 1)$, since F is right-continuous.

Note that (i), (ii) and (ii') are obvious. To see why (iii) is true, notice that by (i), $Q(s-) \leq Q(s)$. Suppose that $Q(s-) < Q(s)$. Clearly for any $Q(s-) < x < Q(s)$, we have by (ii') that $F(x) < s$. Moreover by right-continuity of F , for all $\varepsilon > 0$ small enough, $F(x + \varepsilon) < s$, which is impossible since we know by (ii) that $s \leq F(x)$. Thus $Q(s-) = Q(s)$, which says that Q is left continuous.

For more details see Appendix 1 of Reiss (1989).

Probability Integral Transformation

If X is a random variable with cdf F , then

$$(2.24) \quad X =_d Q(U).$$

(This is called the *probability integral transformation*.)

Proof. By (ii), $Q(U) \leq x$ if and only if $F(x) \geq U$. Thus for each x ,

$$P(Q(U) \leq x) = P(F(x) \geq U) = F(x).$$

□

Note that the *probability integral transformation* implies that if X_1, \dots, X_n are independent random variables with common cdf F and U_1, \dots, U_n are Uniform $(0, 1)$ random variables,

$$(2.25) \quad (X_1, \dots, X_n) =_d (Q(U_1), \dots, Q(U_n)).$$

Comunicaciones del CIMAT



CIMAT

CHAPTER 3

Empirical Process Technology Circa 1972

The Classical Empirical Process Technology Circa 1972 consisted of the following basic ingredients:

1. The Glivenko–Cantelli theorem;
2. Linearity in probability;
3. Birnbaum–Marshall inequality;
4. The Skorohod Representation theorem;
5. The probability integral transformation.

An Example of the Use of the 1972 Technology: The Asymptotic Normality of L-statistics

The basic ideas in this section originate from Shorack (1972). Let X, X_1, \dots, X_n be i.i.d. with common cdf F with corresponding quantile function Q and let $X_{1,n} \leq \dots \leq X_{n,n}$ denote their order statistics. Consider the L-statistic

$$L_n = \sum_{i=1}^n c_{i,n} X_{i,n},$$

where $c_{1,n}, \dots, c_{n,n}$ are constants. By (2.25) we get that

$$L_n =_d \sum_{i=1}^n c_{i,n} Q(U_{i,n}),$$

where $U_{1,n} \leq \dots \leq U_{n,n}$ are the order statistics of U_1, \dots, U_n .

From now on for simplicity of presentation assume that

$$c_{i,n} = \int_{(i-1)/n}^{i/n} J(u) du, \quad i = 1, \dots, n,$$

with J being a continuous integrable function on $(0, 1)$. Write

$$\mu = \int_0^1 Q(u) J(u) du,$$

where we assume $\int_0^1 |Q(u)J(u)| du < \infty$. (One can weaken the continuity assumption on J to the requirement that J and Q do not share discontinuity points.)

It was observed by Shorack (1972) that

$$\int_0^1 \int_{G_n(t)}^t J(u) du dQ(t) = \sum_{i=1}^n c_{i,n} Q(U_{i,n}) - \mu =_d L_n - \mu.$$

Now by applying the mean value theorem for each $t \in (0, 1)$ we can find a $\theta_n(t)$ between $G_n(t)$ and t so that

$$J(\theta_n(t))(t - G_n(t)) = \int_{G_n(t)}^t J(u) du.$$

So we get that

$$(3.1) \quad \sqrt{n} \int_0^1 J(\theta_n(t))(t - G_n(t)) dQ(t) =_d \sqrt{n}(L_n - \mu).$$

We shall be using the tools 1, 2, 3 and 4.

To obtain the asymptotic distribution of $\sqrt{n}(L_n - \mu)$, it is clear from (3.1) that it suffices to determine that of

$$(3.2) \quad \int_0^1 J(\theta_n(t)) \alpha_n(t) dQ(t).$$

We shall now switch to the probability space of the Skorohod representation. We shall work with a sequence $\{\tilde{\alpha}_n\}_{n \geq 1}$ of probabilistically equivalent versions of $\{\alpha_n\}$ and a fixed Brownian bridge B such that

$$\sup_{0 \leq t \leq 1} |\tilde{\alpha}_n(t) - B(t)| \rightarrow 0, \text{ a.s., as } n \rightarrow \infty.$$

So instead of (3.2), we shall investigate its probabilistically equivalent version

$$(3.3) \quad \int_0^1 J(\tilde{\theta}_n(t)) \tilde{\alpha}_n(t) dQ(t).$$

First by the Glivenko–Cantelli theorem (see (2.3) above)

$$(3.4) \quad \sup_{0 \leq t \leq 1} |\tilde{\theta}_n(t) - t| \rightarrow 0, \text{ a.s., as } n \rightarrow \infty.$$

Therefore by continuity of J and B and (3.4) for each $0 < \delta < 1/2$,

$$(3.5) \quad \sup_{\delta \leq t \leq 1-\delta} |J(\tilde{\theta}_n(t)) \tilde{\alpha}_n(t) - J(t)B(t)| \rightarrow 0, \text{ a.s., as } n \rightarrow \infty.$$

Hence it is natural then to assume that somehow in some stochastic sense

$$(3.6) \quad \int_0^1 J(\tilde{\theta}_n(t)) \tilde{\alpha}_n(t) dQ(t) \rightarrow \int_0^1 J(t) B(t) dQ(t),$$

from which it can be inferred that

$$(3.7) \quad \int_0^1 J(\tilde{\theta}_n(t)) \tilde{\alpha}_n(t) dQ(t) \rightarrow_d \int_0^1 J(t) B(t) dQ(t).$$

Since under suitable assumptions on J and Q the random variable

$$\int_0^1 J(t) B(t) dQ(t)$$

is a normal random variable with mean 0 and variance

$$\sigma^2(J) = \int_0^1 \int_0^1 (s \wedge t - st) J(s) J(t) dQ(s) dQ(t) < \infty,$$

we could conclude from (3.7) that

$$\sqrt{n}(L_n - \mu) \rightarrow_d N(0, \sigma^2(J)).$$

We shall now show how to use the tools in 1, 2, 3 and 4 to establish (3.6). Choose any $0 < \delta < 1/2$ and decompose

$$\begin{aligned} & \int_0^1 J(\tilde{\theta}_n(t)) \tilde{\alpha}_n(t) dQ(t) \\ &= \int_\delta^{1-\delta} J(\tilde{\theta}_n(t)) \tilde{\alpha}_n(t) dQ(t) + \int_0^\delta J(\tilde{\theta}_n(t)) \tilde{\alpha}_n(t) dQ(t) \\ & \quad + \int_{1-\delta}^1 J(\tilde{\theta}_n(t)) \tilde{\alpha}_n(t) dQ(t) = M_n(\delta) + L_n(\delta) + U_n(\delta). \end{aligned}$$

Also write

$$\begin{aligned} \int_0^1 J(t) B(t) dQ(t) &= \int_\delta^{1-\delta} J(t) B(t) dQ(t) + \int_0^\delta J(t) B(t) dQ(t) \\ & \quad + \int_{1-\delta}^1 J(t) B(t) dQ(t) = M(\delta) + L(\delta) + U(\delta). \end{aligned}$$

Clearly by (3.5)

$$(3.8) \quad M_n(\delta) \rightarrow M(\delta), \text{ a.s., as } n \rightarrow \infty.$$

Now impose the following assumptions: for some $\nu_1 > 0$ and $\nu_2 > 0$ with $-1/2 + \nu_1 < 0$ and $-1/2 + \nu_2 < 0$

$$(3.9) \quad |J(u)| \leq K u^{-1/2+\nu_1} \text{ and } |J(1-u)| \leq K u^{-1/2+\nu_2}, \text{ for } 0 < u \leq 1/2.$$

Further assume that for some $\nu_1 > \mu_1 > 0$ and $\nu_2 > \mu_2 > 0$ with $-1/2 < -1/2 + \nu_1 - \mu_1 < 0$ and $-1/2 < -1/2 + \nu_2 - \mu_2 < 0$

$$(3.10) \quad \mathbb{B}_1 = \int_0^{1/2} t^{\mu_1} dQ(t) < \infty \text{ and } \mathbb{B}_2 = \int_{1/2}^1 (1-t)^{\mu_2} dQ(t) < \infty.$$

The conditions on J and Q imply that $\sigma^2(J) < \infty$. (From now on to ease notation we shall drop the \sim 's.) Using the linearity in probability (keeping in mind that $\theta_n(t)$ is between $G_n(t)$ and t), by Fact 1 and the fact that $U_{1,n} \rightarrow_P 0$, for any $\varepsilon > 0$ we can choose a $\lambda > 1$ and n large enough so that with probability greater than or equal to $1 - \varepsilon$,

$$\frac{1}{\lambda} < \frac{G_n(t)}{t} < \lambda \text{ for all } U_{1,n} \leq t \leq 1 \text{ and } U_{1,n} < \delta,$$

which implies by the first part of (3.9) that for all $U_{1,n} \leq t \leq \delta$,

$$(3.11) \quad |J(\theta_n(t))| \leq K_\lambda t^{-1/2+\nu_1}$$

for some $K_\lambda > 0$. Now for $0 < t < U_{1,n} \leq 1/2$, by definition,

$$J(\theta_n(t)) = t^{-1} \int_0^t J(u) du,$$

so that inequality (3.11) still holds by increasing K_λ , if necessary. Therefore on this random set

$$\begin{aligned} |L_n(\delta)| &\leq K_\lambda \int_0^\delta |\alpha_n(t)| t^{-1/2+\nu_1} dQ(t) \\ &\leq K_\lambda \sup_{0 < t \leq \delta} |\alpha_n(t)| t^{-1/2+\nu_1-\mu_1} \mathbb{B}_1. \end{aligned}$$

Notice that the function

$$q(t) = t^{1/2-\nu_1+\mu_1}, \quad 0 < t \leq 1/2,$$

satisfies the conditions of the Birnbaum–Marshall inequality, so that for all $\varepsilon > 0$,

$$\lim_{\delta \searrow 0} \limsup_{n \rightarrow \infty} P \left\{ \sup_{0 < t \leq \delta} |\alpha_n(t)| t^{-1/2+\nu_1-\mu_1} > \varepsilon \right\} = 0.$$

Thus we get for all $\varepsilon > 0$,

$$\lim_{\delta \searrow 0} \limsup_{n \rightarrow \infty} P \left\{ \sup_{0 < t \leq \delta} |L_n(\delta)| > \varepsilon \right\} = 0.$$

In the same way one can show that for all $\varepsilon > 0$,

$$\lim_{\delta \searrow 0} \limsup_{n \rightarrow \infty} P \left\{ \sup_{0 < t \leq \delta} |U_n(\delta)| > \varepsilon \right\} = 0.$$

Moreover, similarly, one can prove that for all $\varepsilon > 0$,

$$\lim_{\delta \searrow 0} \left[P \left\{ \sup_{0 < t \leq \delta} |L(\delta)| > \varepsilon \right\} + P \left\{ \sup_{0 < t \leq \delta} |U(\delta)| > \varepsilon \right\} \right] = 0.$$

Hence by “ ε -squeezing” (see Theorem 4.2 of Billingsley (1968)),

$$\int_0^1 J(\tilde{\theta}_n(t)) \tilde{\alpha}_n(t) dQ(t) \rightarrow_P \int_0^1 J(t) B(t) dQ(t).$$

We have just proved a simplified version of Theorem 1 of Shorack (1972). Also see Section 16.4 of Shorack (2000). For further advances in central limit theorems for L-statistics refer to Mason and Shorack (1990, 1992). In particular, see Theorem 3.2 of Mason and Shorack (1990). There the proofs are based on the weighted approximation stated in Theorem 1 below.

ε -squeezing

Suppose we want to show that the sequence of random variables $(Y_n)_{n \geq 1}$ converges in probability to the random variable Y , i.e., we want to show that for all $\varepsilon > 0$,

$$(3.12) \quad P \{|Y_n - Y| > \varepsilon\} \rightarrow 0,$$

that is, $Y_n \rightarrow_P Y$. Sometimes one can do this by “ ε -squeezing”. Suppose that for each $0 < \delta < 1$, there exist sequences of random variables $\Delta_n(\delta)$ and $M_n(\delta)$ such that

$$|Y_n - Y| \leq \Delta_n(\delta) + M_n(\delta).$$

Assume that for all $0 < \delta < 1$ and $\varepsilon > 0$

$$P \{|M_n(\delta)| > \varepsilon\} \rightarrow 0,$$

that is, $M_n(\delta) \rightarrow_P 0$. Further assume that for all $\varepsilon > 0$

$$\limsup_{\delta \searrow 0} \limsup_{n \rightarrow \infty} P \{|\Delta_n(\delta)| > \varepsilon\} \rightarrow 0.$$

Then we get that for all $\varepsilon > 0$

$$P \{|Y_n - Y| > 2\varepsilon\} \leq P \{|M_n(\delta)| > \varepsilon\} + P \{|\Delta_n(\delta)| > \varepsilon\}.$$

Therefore

$$\begin{aligned} \limsup_{n \rightarrow \infty} P \{|Y_n - Y| > 2\varepsilon\} &= \limsup_{\delta \searrow 0} \limsup_{n \rightarrow \infty} P \{|Y_n - Y| > 2\varepsilon\} \\ &\leq \limsup_{\delta \searrow 0} \limsup_{n \rightarrow \infty} P \{|M_n(\delta)| > \varepsilon\} + \limsup_{\delta \searrow 0} \limsup_{n \rightarrow \infty} P \{|\Delta_n(\delta)| > \varepsilon\} = 0. \end{aligned}$$

Since $\varepsilon > 0$ can be made arbitrarily small, this implies (3.12).

q -metric convergence

Pyke and Shorack (1968) were interested in characterizing those positive functions q on $(0, 1)$ increasing on $(0, 1/2]$ and decreasing on $[1/2, 1)$ such that for the Skorohod representation

$$(3.13) \quad \sup_{0 < t < 1} |\tilde{\alpha}_n(t) - B(t)|/q(t) \rightarrow_p 0, \text{ as } n \rightarrow \infty.$$

They called this q -metric convergence of the uniform empirical process to a Brownian bridge. An application of the Birnbaum–Marshall inequality shows that for (3.13) to hold it suffices that

$$\int_0^1 \frac{dt}{q^2(t)} < \infty.$$

Here is the argument. We have for any $0 < \delta < 1/2$,

$$\begin{aligned} & \sup_{\delta < t < 1-\delta} |\tilde{\alpha}_n(t) - B(t)|/q(t) \\ & \leq \max\left(\frac{1}{q(\delta)}, \frac{1}{q(1-\delta)}\right) \sup_{0 < t < 1} |\tilde{\alpha}_n(t) - B(t)|, \\ & \sup_{0 < t \leq \delta} |\tilde{\alpha}_n(t) - B(t)|/q(t) \leq \sup_{0 < t \leq \delta} |\tilde{\alpha}_n(t)|/q(t) + \sup_{0 < t \leq \delta} |B(t)|/q(t), \end{aligned}$$

with a similar bound for $\sup_{1-\delta < t \leq 1} |\tilde{\alpha}_n(t) - B(t)|/q(t)$. Using (2.15), (2.20) and (2.21), we see that (3.13) follows by “ ε -squeezing”.

CHAPTER 4

Intermediate Steps Towards Weighted Approximations

O'Reilly's Theorem (1974)

O'Reilly's theorem was in a sense an intermediate step towards the development of the weighted approximation methodology, since the search for an easy and transparent proof of it led to the creation of the first weighted approximation of the uniform empirical process by a sequence of Brownian bridges. Here is a statement of O'Reilly's theorem.

O'Reilly's Theorem *Let q be a positive function on $(0, 1)$, increasing on $(0, 1/2]$ and decreasing on $[1/2, 1)$. For any probability space such that*

$$\sup_{0 < t < 1} |\tilde{\alpha}_n(t) - B_n(t)| \rightarrow_p 0,$$

where $\{\tilde{\alpha}_n\}$ is a sequence of probabilistically equivalent versions of $\{\alpha_n\}$, and $\{B_n\}$ is an appropriate sequence of Brownian bridges, one also has

$$\sup_{0 < t < 1} |\tilde{\alpha}_n(t) - B_n(t)| / q(t) \rightarrow_p 0$$

if and only if for all $c > 0$,

$$(4.1) \quad I(c, q) := \int_0^1 (s(1-s))^{-1} \exp\left(-\frac{cq^2(s)}{s(1-s)}\right) ds < \infty.$$

The crucial fact established by O'Reilly (1974) was that

$$\lim_{\delta \searrow 0} \left(\sup_{0 < s \leq \delta} |B(s)| / q(s) + \sup_{0 < s \leq \delta} |B(1-s)| / q(1-s) \right) = 0, \text{ a.s.}$$

if and only if for all $c > 0$, $I(c, q) < \infty$.

Chung's Law of the Iterated Logarithm for α_n

Let X_1, X_2, \dots , be a sequence of i.i.d. random variables with mean μ and variance σ^2 . The law of the iterated logarithm [LIL] says that with

probability 1,

$$\limsup_{n \rightarrow \infty} \pm \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{2n \log \log n}} = \sigma.$$

Applying this to the uniform empirical process α_n at a fixed $0 \leq t \leq 1$, gives with probability 1,

$$\limsup_{n \rightarrow \infty} \pm \frac{\alpha_n(t)}{\sqrt{2 \log \log n}} = \sqrt{t(1-t)}.$$

The Chung (1949) Law of the Iterated Logarithm for α_n says that with probability 1,

$$\limsup_{n \rightarrow \infty} \frac{\sup_{0 \leq t \leq 1} \pm \alpha_n(t)}{\sqrt{2 \log \log n}} = \sup_{0 \leq t \leq 1} \sqrt{t(1-t)} = \frac{1}{2}.$$

For the general empirical process we get that with probability 1,

$$\limsup_{n \rightarrow \infty} \frac{\sup_{-\infty < x < \infty} \pm \alpha_{n,F}(x)}{\sqrt{2 \log \log n}} = \sup_{-\infty < x < \infty} \sqrt{F(x)(1-F(x))}.$$

Note that if F is not continuous it may happen that

$$\sup_{-\infty < x < \infty} \sqrt{F(x)(1-F(x))} < \frac{1}{2}.$$

One of the implications of the Chung LIL is that it is impossible to define a sequence of independent Uniform $(0, 1)$ random variables U_1, U_2, \dots , and a fixed Brownian bridge B on the same probability space so that with probability 1

$$(4.2) \quad \lim_{n \rightarrow \infty} \sup_{0 \leq t \leq 1} |\alpha_n(t) - B(t)| = 0.$$

To see this, clearly we have

$$\limsup_{n \rightarrow \infty} \frac{\sup_{0 \leq t \leq 1} |B(t)|}{\sqrt{\log \log n}} = 0,$$

but, with probability 1,

$$\limsup_{n \rightarrow \infty} \frac{\sup_{0 \leq t \leq 1} |\alpha_n(t)|}{\sqrt{2 \log \log n}} = \frac{1}{2},$$

which contradicts (4.2).

Our next result shows that (4.2) can hold when B is replaced by an appropriate sequence of Brownian bridges $\{B_n\}_{n \geq 1}$.

The KMT (1975) Approximation

Komlós, Major and Tusnády [KMT] (1975) published the following remarkable Brownian bridge approximation to the uniform empirical process.

Theorem [KMT] *There exists a probability space (Ω, A, P) with independent Uniform $(0, 1)$ random variables U_1, U_2, \dots , and a sequence of Brownian bridges B_1, B_2, \dots , such that for all $n \geq 1$ and $-\infty < x < \infty$,*

$$(4.3) \quad P \left\{ \sup_{0 \leq t \leq 1} |\alpha_n(t) - B_n(t)| \geq n^{-1/2}(a \log n + x) \right\} \leq b \exp(-cx),$$

where a, b and c are suitable positive constants independent of n and x .

Notice that when inequality (4.3) is combined with the Borel–Cantelli lemma we get the rate of approximation

$$\sup_{0 \leq t \leq 1} |\alpha_n(t) - B_n(t)| = O \left(\frac{\log n}{\sqrt{n}} \right), \text{ a.s.}$$

For some time people did not know what to do with the KMT (1975) approximation to the uniform empirical process. This was complicated by the fact that KMT (1975) only provided a sketch of its proof. Complete proofs are now available. Consult Mason and van Zwet (1987) with additional notes in Mason (2001a), Péter Major’s website, Bretagnolle and Massart (1989), Major (1999) and Dudley (2000). Bretagnolle and Massart (1989) determined values for the constants in (4.3), namely $a = 12$, $b = 2$ and $c = 1/6$. In Section 1.4 of the second edition of his book *Uniform Central Limit Theorems*, Dudley (2014) includes a full proof of the Bretagnolle and Massart (1989) result. A detailed discussion of the quantile transform methodology upon which these proofs are based is given in Mason and Zhou (2012).

As aside we should mention that KMT (1975) also proved the following *Kiefer process approximation* to α_n .

Theorem [KMT(KP)] *There exists a probability space (Ω, A, P) with independent Uniform $(0, 1)$ random variables U_1, U_2, \dots , and a sequence of independent Brownian bridges B_1, B_2, \dots , such that for all $n \geq 1$ and $-\infty < x < \infty$,*

$$(4.4) \quad P \left\{ \sup_{0 \leq t \leq 1} |\alpha_n(t) - n^{-1/2} \sum_{i=1}^n B_i(t)| \geq n^{-1/2} \log n (a_1 \log n + x) \right\} \leq b_1 \exp(-c_1 x),$$

where a_1, b_1 and c_1 are suitable positive constants independent of n and x .

We should point out that the probability space of KMT is not the same at the probability space of KMT(KP). We see that on the probability space of KMT(KP)

$$(4.5) \quad \sup_{0 \leq t \leq 1} |\alpha_n(t) - n^{-1/2} \sum_{i=1}^n B_i(t)| = O\left(\frac{(\log n)^2}{\sqrt{n}}\right), \text{ a.s.}$$

Shorack (1979) was able to use KMT (1975) to give a simple proof of O'Reilly's theorem under the additional assumption that $q(t)/t^{1/2} \nearrow \infty$ and $q(1-t)/t^{1/2} \nearrow \infty$ as $t \searrow 0$. In this case, it is readily verified that $I(q, c) < \infty$ for all $c > 0$ is equivalent to, as $t \searrow 0$,

$$(4.6) \quad q(t)/(t \log \log(1/t))^{1/2} \rightarrow \infty \text{ and } q(1-t)/(t \log \log(1/t))^{1/2} \rightarrow \infty.$$

However, not all q for which (4.1) is finite for all $c > 0$ satisfy (4.6). (See M. Csörgő (1983).)

CHAPTER 5

The First Weighted Approximation

A much stronger result than the O'Reilly theorem is the following weighted approximation in probability of special versions of the α_n 's by a sequence of Brownian bridges $\{B_n\}_{n \geq 1}$.

Theorem 1. *On a rich enough probability space there exists a sequence of independent Uniform $(0, 1)$ random variables U_1, U_2, \dots , and a sequence of Brownian bridges B_1, B_2, \dots , such that for the uniform empirical processes α_n based on the U_i 's and all $0 < \nu < \frac{1}{4}$*

$$(5.1) \quad \sup_{0 \leq t \leq 1} \frac{|\alpha_n(t) - B_n(t)|}{(t(1-t))^{1/2-\nu}} = O_p(n^{-\nu}).$$

Moreover, statement (5.1) remains true for $\nu = 0$ when B_n is replaced by \bar{B}_n , where for $n \geq 2$

$$(5.2) \quad \bar{B}_n(t) = B_n(t) 1_{\{t \in [1/n, 1 - 1/n]\}}.$$

M. Csörgő, S. Csörgő, Horváth and Mason [Cs-Cs-H-M] (1986) proved this result in 1983 during a hot and sweaty Szeged summer. Mason and van Zwet (1987) obtained the best possible version of it during a deeply overcast March 1985 in Leiden, allowing $0 \leq \nu < \frac{1}{2}$. (Extreme weather must be good for mathematics.) Both of these results were based upon the strong approximation methods and results of KMT (1975). Later it was discovered that a very useful version of this result could be derived using the Skorohod embedding. More will be said about this later.

Miklós Csörgő, Sándor Csörgő and I introduced our results at a March 1984 Oberwolfach Conference on Order Statistics, Quantile Processes, and Extreme Value theory in a three part series of lectures entitled, *A new approximation for the uniform empirical and quantile processes with applications*. (Lajos Horváth was unable to attend on account of visa problems.) Our lectures created a bit of stir. At the meeting, Sándor Csörgő, Paul Deheuvels and I wrote one of the first papers to apply these approximations. In the years that followed I was able to attract a number of talented collaborators to work with me on extensions and applications of Cs-Cs-H-M (1986). They include Jan Beirlant, Paul Deheuvels, John Einmahl, Uwe Einmahl, Erich Haeusler, Galen

Shorack, Tatyana Turova and Willem van Zwet. Much of my work with them will be described in the notes that follow.

The Goal of Weighted Approximations

The goal of the weighted approximation technique is to transfer the asymptotic distributional analysis of a sequence of functionals of the uniform empirical process α_n to that of a sequence of functionals of Brownian bridges B_n .

Example 1: O'Reilly's Theorem Revisited

Only assume that $q(s)/s^{1/2}$ and $q(1-s)/s^{1/2} \rightarrow \infty$ as $s \searrow 0$. (Any q function for which $I(c, q) < \infty$ for some $c > 0$ satisfies this condition, however, for every function satisfying this condition it is not always true that $I(c, q) < \infty$ for some $c > 0$.) Assume that we are on the probability space of Theorem 1. It is easy to show that

$$\sup_{0 \leq t \leq 1} |\alpha_n(t) - B_n(t)| = o_p(1),$$

when combined with

$$\sup_{0 \leq t \leq 1} \frac{|\alpha_n(t) - \bar{B}_n(t)|}{(t(1-t))^{1/2}} = O_p(1)$$

gives for any such q

$$\sup_{0 < t < 1} \frac{|\alpha_n(t) - \bar{B}_n(t)|}{q(s)} = o_p(1).$$

So clearly the underlying rationale behind O'Reilly's conditions was to characterize when

$$\sup_{0 < s \leq 1/n} |B_n(s)|/q(s) \rightarrow_P 0, \text{ as } n \rightarrow \infty,$$

and

$$\sup_{0 < s \leq 1/n} |B_n(1-s)|/q(1-s) \rightarrow_P 0, \text{ as } n \rightarrow \infty.$$

Example 2: Asymptotic Distribution of Rényi-Type Statistics

Let a_n be any sequence of positive constants such that $0 < a_n < \beta < 1$, for some $0 < \beta < 1$, and $na_n \rightarrow \infty$. Csáki (1974) established by direct combinatorial methods the somewhat surprising result that

$$\left(\frac{a_n}{1-a_n} \right)^{1/2} \sup_{a_n \leq s \leq 1} \frac{\alpha_n(s)}{s} \rightarrow_d \sup_{0 \leq s \leq 1} W(s),$$

where W is a standard Wiener process on $[0, 1]$.

Proof. Choose $0 < \nu < 1/4$. Now on the probability space of Theorem 1,

$$\begin{aligned} & \left(\frac{a_n}{1-a_n} \right)^{1/2} \sup_{a_n \leq s \leq 1} \left| \frac{\alpha_n(s) - B_n(s)}{s} \right| \\ & \leq \left(\frac{a_n}{1-a_n} \right)^{1/2} \frac{n^\nu}{\sqrt{a_n}} \sup_{a_n \leq s \leq 1} \left| \frac{\alpha_n(s) - B_n(s)}{s^{1/2-\nu}} \right| \frac{1}{(na_n)^\nu} = o_P(1). \end{aligned}$$

But

$$\begin{aligned} & \left\{ \left(\frac{a_n}{1-a_n} \right)^{1/2} \frac{B(s)}{s}, a_n \leq s \leq 1 \right\} \\ & =_d \left\{ W \left(\left(\frac{a_n}{1-a_n} \right) \frac{1-s}{s} \right), a_n \leq s \leq 1 \right\}. \end{aligned}$$

Thus

$$\sup_{a_n \leq s \leq 1} \left(\frac{a_n}{1-a_n} \right)^{1/2} \frac{B(s)}{s} =_d \sup_{0 \leq t \leq 1} W(t),$$

which completes the proof. For a generalized version of this result refer to Mason (1985), and for applicable versions of the Rényi confidence bands, also obtained by similar ideas, see S. Csörgő (1998) and Megyesi (1998).

A Typical Application of Weighted Approximations

Often one is interested in establishing the asymptotic normality of an integral function of a process v_n , say,

$$I_n = \int_0^1 v_n(t) d\mu_n(t),$$

where μ_n is some measure on $(0, 1)$. Whenever there exists a weighted approximation of v_n by a Brownian bridge B_n , one can typically show that for any $\tau > 0$,

$$\begin{aligned} & \left| \int_{\tau/n}^{1-\tau/n} v_n(t) d\mu_n(t) - \int_{\tau/n}^{1-\tau/n} B_n(t) d\mu_n(t) \right| \leq \\ & \sup_{\frac{\tau}{n} \leq t \leq 1-\frac{\tau}{n}} \frac{n^\nu |v_n(t) - B_n(t)|}{(t(1-t))^{1/2-\nu}} \left[\frac{\int_{\tau/n}^{1-\tau/n} (t(1-t))^{1/2-\nu} d\mu_n(t)}{n^\nu} \right] = o_p(1). \end{aligned}$$

This is the crucial step to approximate I_n directly by the normal random variable

$$\int_0^1 B_n(t) d\mu_n(t)$$

and establish the asymptotic normality of I_n —should it, in fact, be asymptotically normal.

Example 3: A Central Limit Theorem for Winsorized-type Sums

Let X, X_1, X_2, \dots , be a sequence of i.i.d. nondegenerate random variables with cdf F with left-continuous inverse function Q . Choose $0 < a < 1 - b < 1$ and $n \geq 1$, and consider the Winsorized-type sum

$$W_n(a, b) := \sum_{i=1}^n [X_i 1\{Q(a) < X_i \leq Q(1-b)\}] \\ + \sum_{i=1}^n [Q(a) 1\{X_i \leq Q(a)\} + Q(1-b) 1\{X_i > Q(1-b)\}].$$

Integrating by parts these sums can be written as

$$(5.3) \quad n^{-1/2} \{W_n(a, b) - EW_n(a, b)\} = - \int_a^{1-b} \alpha_n(s) dQ(s).$$

Set

$$(5.4) \quad \sigma^2(a, b) = \int_a^{1-b} \int_a^{1-b} (s \wedge t - st) dQ(s) dQ(t) = \text{Var } W_1(a, b).$$

We show below that if a_n and b_n are sequences of positive constants such that $0 < a_n < 1 - b_n < 1$ for $n \geq 1$, and as $n \rightarrow \infty$,

$$a_n \rightarrow 0, \quad na_n \rightarrow \infty, \quad b_n \rightarrow 0 \quad \text{and} \quad nb_n \rightarrow \infty,$$

then

$$(N) \quad Z_n(a_n, b_n) := \int_{a_n}^{1-b_n} \alpha_n(s) dQ(s) / \sigma(a_n, b_n) \rightarrow_d Z, \quad \text{as } n \rightarrow \infty,$$

where Z is a standard normal random variable. This was a crucial step in the S. Csörgő, Haeusler and Mason (1988a) probabilistic approach to the asymptotic distribution of sums of independent, identically distributed random variables; see also S. Csörgő (1990) and S. Csörgő and Megyesi (2002). The approach used here was first introduced by Shorack (1972) to derive central limit theorems for sums of functions of order statistics.

Proof of (N).

Denote the standard normal random variable

$$Z_n := \int_{a_n}^{1-b_n} B_n(s) dQ(s) / \sigma(a_n, b_n).$$

Notice that on the probability space of Theorem 1,

$$|Z_n(a_n, b_n) - Z_n| \leq \int_{a_n}^{1/2} |\alpha_n(s) - B_n(s)| dQ(s) / \sigma(a_n, 1/2)$$

$$+ \int_{1/2}^{1-b_n} |\alpha_n(s) - B_n(s)| dQ(s) / \sigma(1/2, b_n),$$

which for any $0 < \nu < 1/4$ is

$$\begin{aligned} &\leq \Delta_{n,\nu}(1) n^{-\nu} \int_{a_n}^{1/2} (s(1-s))^{1/2-\nu} dQ(s) / \sigma(a_n, 1/2) \\ &+ \Delta_{n,\nu}(1) n^{-\nu} \int_{1/2}^{1-b_n} (s(1-s))^{1/2-\nu} dQ(s) / \sigma(1/2, b_n), \end{aligned}$$

where

$$\Delta_{n,\nu}(1) := \sup_{1/n \leq t \leq 1-1/n} \frac{n^\nu |\alpha_n(t) - B_n(t)|}{(t(1-t))^{1/2-\nu}}.$$

Using the fact (e.g. Inequality 2.1 of Shorack (1997)) that for any $0 < c < 1-d < 1$ and $0 < \nu < 1/2$,

$$(5.5) \quad \int_c^{1-d} (s(1-s))^{1/2-\nu} dQ(s) / \sigma(c, d) \leq (3/\sqrt{\nu})(c \wedge d)^{-\nu},$$

we see that this last bound is

$$\leq (3/\sqrt{\nu})(na_n)^{-\nu} O_p(1) + (3/\sqrt{\nu})(nb_n)^{-\nu} O_p(1) = o_p(1).$$

Use of this result to prove asymptotic normality of intermediate trimmed sums

Let X_1, \dots, X_n be i.i.d. F with order statistics $X_{1,n} \leq \dots \leq X_{n,n}$. Consider integers k_n satisfying $1 \leq k_n \leq n/2$, $n \geq 3$, $k_n \rightarrow \infty$ and $k_n/n \rightarrow 0$, and the intermediate trimmed sum

$$T_n(k_n) = \sum_{i=k_n+1}^{n-k_n} X_{i,n}.$$

Under certain necessary and sufficient conditions (see S. Csörgő and Haeusler and Mason (1988b))

$$(Z) \quad \frac{T_n(k_n) - n \int_{k_n/n}^{1-k_n/n} Q(u) du}{\sqrt{n} \sigma(k_n/n, k_n/n)} \rightarrow_d Z,$$

where Z is standard normal. The reason for the normality is that the necessary and sufficient conditions for (Z) to hold give

$$\frac{T_n(k_n) - n \int_{k_n/n}^{1-k_n/n} Q(u) du}{\sqrt{n} \sigma(k_n/n, k_n/n)} + \frac{\int_{k_n/n}^{1-k_n/n} \alpha_n(u) dQ(u)}{\sigma(k_n/n, k_n/n)} = o_P(1)$$

and, as we have just shown, it is always true that

$$\frac{\int_{k_n/n}^{1-k_n/n} \alpha_n(u) dQ(u)}{\sigma(k_n/n, k_n/n)} \rightarrow_d Z.$$

Example 4: Central Limit Theorem for the Hill Estimator (S. Csörgő and Mason (1985))

Let Y, Y_1, \dots, Y_n be i.i.d. G with a regularly varying upper tail with index $1/c$, $c > 0$, that is for all $t > 0$

$$\frac{1 - G(xt)}{1 - G(x)} \rightarrow t^{-1/c}, \text{ as } x \rightarrow \infty.$$

Now set $X = \log(\max(Y, 1))$, $X_i = \log(\max(Y_i, 1))$, $i = 1, \dots, n$. Further let $X_{1,n} \leq \dots \leq X_{n,n}$ denote the order statistics of X_1, \dots, X_n . The *Hill estimator* of c is

$$\hat{c}_n = \sum_{i=1}^{k_n} \frac{X_{n+1-i,n}}{k_n} - X_{n-k_n,n},$$

where k_n is a sequence of positive integers satisfying $1 \leq k_n < n$, $k_n \rightarrow \infty$ and $k_n/n \rightarrow 0$. Mason (1983) showed that for any such sequence

$$\hat{c}_n \rightarrow_P c, \text{ as } n \rightarrow \infty.$$

Let F be the cdf of X and Q be its inverse. We see by the probability integral transformation (2.24) that

$$\hat{c}_n \stackrel{d}{=} \sum_{i=1}^{k_n} \frac{Q(U_{n+1-i,n})}{k_n} - Q(U_{n-k_n,n}).$$

Set

$$c_n = \frac{n}{k_n} \int_{1-k_n/n}^1 (1-s) dQ(s).$$

One can verify that $c_n \rightarrow c$ as $n \rightarrow \infty$. Under additional assumptions (see S. Csörgő and Mason (1985)) it can be shown that on the probability space of Theorem 1,

$$\sqrt{k_n}(\hat{c}_n - c_n) = Z_n + o_P(1),$$

where

$$Z_n := -\sqrt{\frac{n}{k_n}} \int_{1-k_n/n}^1 B_n(s) dQ(s) + c \sqrt{\frac{n}{k_n}} B_n\left(1 - \frac{k_n}{n}\right).$$

The random variable Z_n is normal with mean 0 and a variance, which converges to c^2 as $n \rightarrow \infty$. The essential step in the proof is the replacement

$$\left| \sqrt{\frac{n}{k_n}} \int_{1-k_n/n}^1 \alpha_n(s) dQ(s) - \sqrt{\frac{n}{k_n}} \int_{1-k_n/n}^1 B_n(s) dQ(s) \right|,$$

which for any $0 < \nu < 1/4$ is

$$\begin{aligned} &\leq \sqrt{\frac{n}{k_n}} \int_{1-k_n/n}^1 \frac{|\alpha_n(s) - B_n(s)|}{(1-s)^{1/2-\nu}} (1-s)^{1/2-\nu} dQ(s) \\ &\leq \frac{n^\nu}{k_n^\nu} \sup_{1-\frac{k_n}{n} \leq s \leq 1} \frac{|\alpha_n(s) - B_n(s)|}{(1-s)^{1/2-\nu}} \int_{1-k_n/n}^1 (1-s)^{1/2-\nu} dQ(s) \left(\frac{n}{k_n}\right)^{1/2-\nu}, \end{aligned}$$

which since

$$n^\nu \sup_{1-\frac{k_n}{n} \leq s \leq 1} \frac{|\alpha_n(s) - B_n(s)|}{(1-s)^{1/2-\nu}} = O_P(1)$$

and

$$\int_{1-k_n/n}^1 (1-s)^{1/2-\nu} dQ(s) \left(\frac{n}{k_n}\right)^{1/2-\nu} \rightarrow \frac{c}{1/2-\nu}, \text{ as } n \rightarrow \infty,$$

is equal to $o_P(1)$.

For generalizations of this estimator refer to S. Csörgő, Deheuvels and Mason (1985) and Groeneboom, Lopuhaä and de Wolf (2003). In both of these papers the weighted approximation in Theorem 1 is the crucial tool used in the derivation of the asymptotic distribution of the estimators. For related applications of the method we refer to S. Csörgő and Viharos (1995, 1998, 2002, 2006).

Further Applications of this Type

The Cs-Cs-H-M (1986) weighted approximation has been applied very successfully in the study of

1. *Central Limit Theorems for Trimmed Sums (as already pointed out)*

$$\sum_{i=k_n+1}^{n-k_n} X_{i,n}.$$

See S. Csörgő, Horváth and Mason (1986), S. Csörgő, Haeusler and Mason (1988b) and S. Csörgő and Megyesi (2001).

2. *Central Limit Theorems for Sums of Extreme Values*

$$\sum_{i=1}^{k_n} X_{i,n}.$$

See S. Csörgő and Mason (1986), S. Csörgő, Haeusler and Mason (1991) and Viharos (1993, 1995).

3. *Central Limit Theorems for L-statistics*

$$\sum_{i=1}^n c_{i,n} X_{i,n}.$$

See Mason and Shorack (1990, 1992).

4. *Bootstrap* See S. Csörgő and Mason (1989) and Deheuvels, Mason and Shorack (1993).

5. *Bahadur–Kiefer Processes* See Deheuvels and Mason (1990) and Beirlant, Deheuvels, J. Einmahl and Mason (1991). The Bahadur–Kiefer process is defined below.

6. *Goodness of fit tests* See del Barrio, Cuesta–Albertos and Matrán (2000) and S. Csörgő (2003).

For further applications refer to the proceedings volume edited by Hahn, Mason and Weiner (1991), the monograph by M. Csörgő and Horváth (1993) and the graduate probability text by Shorack (2000).

CHAPTER 6

The Mason and van Zwet Refinement of KMT

Mason and van Zwet (1987) obtained the following refinement of the KMT (1975) Brownian bridge approximation to the uniform empirical process.

Theorem 2. *There exists a probability space (Ω, A, P) with independent Uniform $(0, 1)$ random variables U_1, U_2, \dots , and a sequence of Brownian bridges B_1, B_2, \dots , such that for all $n \geq 1$, $1 \leq d \leq n$, and $-\infty < x < \infty$,*

$$(6.1) \quad P \left\{ \sup_{0 \leq t \leq d/n} |\alpha_n(t) - B_n(t)| \geq n^{-1/2}(a \log d + x) \right\} \leq b \exp(-cx)$$

and

$$(6.2) \quad P \left\{ \sup_{1-d/n \leq t \leq 1} |\alpha_n(t) - B_n(t)| \geq n^{-1/2}(a \log d + x) \right\} \leq b \exp(-cx),$$

where a, b and c are suitable positive constants independent of n, d and x .

Remark The probability space of Theorem 3 is in fact the KMT (1975) space of Theorem KMT. Setting $d = n$ into these inequalities yields the original KMT inequality (4.3). Rio (1994) has computed values for the constants in these inequalities. Cs-Cs-H-M (1986) had earlier established that the analogs to these inequalities held with α_n replaced by β_n (the uniform quantile process) on the probability space that they constructed so that (5.1) is valid. The process β_n is defined below and the statements of the Cs-Cs-H-M inequalities are given in Chapter 8.

Mason and van Zwet Weighted Approximations

Mason and van Zwet (1987) pointed out that by arguing just as in Cs-Cs-H-M (1983) their inequality leads to the following useful weighted approximations. For any $0 \leq \nu < 1/2$, $n \geq 1$, and $1 \leq d \leq n$ let

$$(6.3) \quad \Delta_{n,\nu}^{(1)}(d) := \sup_{d/n \leq t \leq 1} \frac{n^\nu |\alpha_n(t) - B_n(t)|}{t^{1/2-\nu}},$$

$$(6.4) \quad \Delta_{n,\nu}^{(2)}(d) := \sup_{0 \leq t \leq 1-d/n} \frac{n^\nu |\alpha_n(t) - B_n(t)|}{(1-t)^{1/2-\nu}},$$

and for $1 \leq d \leq n/2$,

$$(6.5) \quad \Delta_{n,\nu}(d) := \sup_{d/n \leq t \leq 1-d/n} \frac{n^\nu |\alpha_n(t) - B_n(t)|}{(t(1-t))^{1/2-\nu}}.$$

On the probability space of Theorem 2, one has

$$\Delta_{n,\nu}(1) = O_p(1),$$

with the same holding with $\Delta_{n,\nu}(1)$ replaced by $\Delta_{n,\nu}^{(1)}(1)$ and $\Delta_{n,\nu}^{(2)}(1)$. (Note that Theorem 1 would give these results in the restricted range $0 \leq \nu < 1/4$.) Our next theorem improves these results.

An Exponential Inequality for the Weighted Approximation to the Uniform Empirical Process

Motivated by a intriguing question brought to him by Evarist Giné, Mason (2001b) derived the following exponential inequality for the Mason and van Zwet weighted approximations.

Theorem 3. (An Improved Mason and van Zwet Result). *On the probability space of Theorem KMT, for every $0 \leq \nu < 1/2$ there exist positive constants A_ν and C_ν such that for all $n \geq 2$, $1 \leq d \leq n/2$ and $0 \leq x < \infty$,*

$$(6.6) \quad P \{ \Delta_{n,\nu}^{(1)}(d) \geq x \} \leq A_\nu \exp(d^{1/2-\nu} C_\nu) \exp\left(-\frac{d^{1/2-\nu} cx}{2}\right),$$

$$(6.7) \quad P \{ \Delta_{n,\nu}^{(2)}(d) \geq x \} \leq A_\nu \exp(d^{1/2-\nu} C_\nu) \exp\left(-\frac{d^{1/2-\nu} cx}{2}\right),$$

and

$$(6.8) \quad P \{ \Delta_{n,\nu}(d) \geq x \} \leq 2A_\nu \exp(d^{1/2-\nu} C_\nu) \exp\left(-\frac{d^{1/2-\nu} cx}{4}\right).$$

Proof. Recall by the previous remark that the probability space of Theorem KMT is that of Theorem 2. First consider (6.6). For any $1 \leq i < i+1 \leq n$ write

$$\delta_{i,n} = P \left\{ \sup_{i/n \leq t \leq (i+1)/n} \frac{n^\nu |\alpha_n(t) - B_n(t)|}{t^{1/2-\nu}} \geq x \right\}.$$

Set $x = 2a_\nu + z$, where a_ν satisfies

$$a_\nu i^{1/2-\nu} > a \log(i+1) \text{ for all } i \geq 1$$

and the constant a is as in (6.1). We get then that

$$\begin{aligned} \delta_{i,n} &\leq P \left\{ \sup_{0 \leq t \leq (i+1)/n} |\alpha_n(t) - B_n(t)| \geq n^{-1/2} i^{1/2-\nu} x \right\} \\ &\leq P \left\{ \sup_{0 \leq t \leq (i+1)/n} |\alpha_n(t) - B_n(t)| \right. \\ &\quad \left. \geq n^{-1/2} (a \log(i+1) + i^{1/2-\nu} a_\nu + i^{1/2-\nu} z) \right\}, \end{aligned}$$

which by (6.1) is

$$\leq b \exp(-i^{1/2-\nu} a_\nu c) \exp(-i^{1/2-\nu} cz).$$

We see then that for any $1 \leq d < n$

$$\begin{aligned} P \left\{ \Delta_{n,\nu}^{(1)}(d) \geq x \right\} &\leq \sum_{i=[d]}^{n-1} \delta_{i,n} \leq b \sum_{i=[d]}^{\infty} \left\{ \exp(-i^{1/2-\nu} a_\nu c) \exp(-i^{1/2-\nu} cz) \right\} \\ &\leq A_\nu \exp(-d^{1/2-\nu} cz/2) = A_\nu \exp(d^{1/2-\nu} C_\nu) \exp(-d^{1/2-\nu} cx/2), \end{aligned}$$

where

$$A_\nu = b \sum_{i=1}^{\infty} \exp(-i^{1/2-\nu} a_\nu c) \text{ and } C_\nu = a_\nu c.$$

This proves inequality (6.6). Inequality (6.7) follows in the same way and inequality (6.8) is an immediate consequence of (6.6) and (6.7). \square

A Moment Bound for the Weighted Approximation

Theorem 3 readily yields the following uniform moment bounds for (6.3), (6.4) and (6.5).

Proposition 1. *On the KMT (1975) approximation probability space for all $0 \leq \nu < 1/2$ there exists a $\gamma > 0$ such that*

$$\sup_{n \geq 2} E \exp(\gamma \Delta_{n,\nu}(1)) < \infty,$$

with the same statement holding with $\Delta_{n,\nu}(1)$ replaced by $\Delta_{n,\nu}^{(1)}(1)$ or $\Delta_{n,\nu}^{(2)}(1)$. In particular, we have for all $r > 0$

$$\sup_{n \geq 2} E \Delta_{n,\nu}^r(1) < \infty.$$

A Functional Version

Now for each integer $n \geq 2$ let R_n denote a class of nondecreasing left-continuous functions r on $[1/n, 1-1/n]$. Assume there exists a sequence of positive constants d_n such that for some $0 \leq \nu < 1/2$

$$(6.9) \quad \sup_{n \geq 2} \sup_{r \in R_n} d_n^{-1} \int_{1/n}^{1-1/n} (s(1-s))^{1/2-\nu} dr(s) =: M < \infty.$$

From Proposition 1 we obtain

Proposition 2. *Let $\{R_n, n \geq 2\}$ denote a sequence of classes of non-decreasing left-continuous functions on $[1/n, 1 - 1/n]$ satisfying (6.9) for some $0 \leq \nu < 1/2$. On the probability space of the KMT (1975) approximation (4.3) there exists a $\gamma > 0$ such that*

$$\sup_{n \geq 2} E \exp(\gamma n^\nu I_n) < \infty,$$

where

$$I_n := \sup_{r \in \mathcal{R}_n} d_n^{-1} \int_{1/n}^{1-1/n} |\alpha_n(s) - B_n(s)| dr(s).$$

Proposition 2 follows trivially from Proposition 1 by observing that $n^\nu I_n \leq \Delta_{n,\nu}(1)M$.

CHAPTER 7

Use of Theorem 3 to Study the Wasserstein Distance

We shall first need some background and a definition.

The Domain of Attraction to a Normal Law

Let X, X_1, X_2, \dots , be a sequence of independent nondegenerate random variables with cdf F and left-continuous inverse or quantile function Q . We say that F is in the domain of attraction of a normal law, written $F \in DN$, if there exist norming and centering constants b_n and c_n such that

$$\frac{\sum_{i=1}^n X_i - c_n}{b_n} \rightarrow_d Z,$$

where Z is a standard normal random variable. S. Csörgő, Haeusler and Mason (1988a) show that when $F \in DN$ one can always choose for $n \geq 2$, $c_n = nEX$ and $b_n = \sqrt{n}\sigma(1/n)$, where for any $0 < u < 1/2$

$$(7.1) \quad \sigma^2(u) := \sigma^2(u, u) = \int_u^{1-u} \int_u^{1-u} (s \wedge t - st) dQ(s) dQ(t).$$

For future reference we shall write for any $0 < u < 1/2$

$$(7.2) \quad \tau^2(u) = \left(\int_u^{1-u} \sqrt{s(1-s)} dQ(s) \right)^2,$$

and note that

$$(7.3) \quad \tau^2(u) \geq \sigma^2(u).$$

It turns out, recall (5.4), that

$$\sigma^2(u) = \text{Var}W_1(u) = \text{Var} \left(\int_u^{1-u} (1\{U \leq t\} - t) dQ(t) \right),$$

where U is Uniform $(0, 1)$. Furthermore $\sigma^2(0) < \infty$ if and only if $\text{Var}X$ is finite.

Now with $F_n(x) = \frac{1}{n} \sum_{i=1}^n 1 \{X_i \leq x\}$, $-\infty < x < \infty$, we have by (2.25)

$$\frac{\sum_{i=1}^n X_i - nEX}{b_n} = \frac{\int_{-\infty}^{\infty} \sqrt{n} \{F(x) - F_n(x)\} dx}{\sigma(1/n)} =_d \frac{-\int_0^1 \alpha_n(s) dQ(s)}{\sigma(1/n)}.$$

In fact one can use the weighted approximation technology as we did in our discussion of Winsorized sums to show that whenever $F \in DN$

$$\frac{\int_0^1 \alpha_n(s) dQ(s)}{\sigma(1/n)} = \frac{\int_{1/n}^{1-1/n} B_n(s) dQ(s)}{\sigma(1/n)} + o_p(1) =_d Z + o_p(1).$$

Crucial to the proof is the fact established in Corollary 1 of S. Csörgő, Haeusler and Mason (1988a) that $F \in DN$ if and only if

$$(7.4) \quad \lim_{u \searrow 0} u (Q^2(\lambda u) + Q^2(1 - \lambda u)) / \sigma^2(u) = 0, \text{ for all } \lambda > 0$$

if and only if σ is slowly varying at zero, i.e.

$$(7.5) \quad \lim_{u \searrow 0} \sigma^2(\lambda u) / \sigma^2(u) = 1, \text{ for all } \lambda > 0.$$

The Wasserstein distance

Recall that the Wasserstein distance between any two cdfs F and G with finite means is

$$d_1(G, F) = \int_{-\infty}^{\infty} |G(x) - F(x)| dx.$$

In particular, the (empirical) Wasserstein distance between F_n , based on X_1, \dots, X_n i.i.d. F , and F is

$$d_1(F_n, F) = \int_{-\infty}^{\infty} |F_n(x) - F(x)| dx =_d \int_0^1 |G_n(t) - t| dQ(t),$$

where Q is defined as in (2.23). Note that $d_1(F_n, F)$ is finite as long as $E|X_1| < \infty$. $d_1(F_n, F)$ also has the representation

$$d_1(F_n, F) = \int_0^1 |Q_n(t) - Q(t)| dt,$$

with Q_n as in (8.5), (See for instance, Exercise 3 on page 64 of SW(1986).) Observe that by (2.25)

$$\sqrt{n}d_1(F_n, F) =_d \int_0^1 |\alpha_n(t)| dQ(t).$$

We shall show how to use Theorem 3 and Proposition 1 to obtain rates of convergence of $Ed_1(F_n, F)$ to zero. This will lead to refinements and

complements of Theorem 6.7 of Bobkov and Ledoux (2014), which in our notation says that for a universal constant $c > 0$,

$$(7.6) \quad \begin{aligned} & 2c \int_{\{t:t(1-t) \leq \frac{1}{4n}\}} t(1-t)dQ(t) + \frac{2c}{\sqrt{n}} \int_{\{t:t(1-t) > \frac{1}{4n}\}} \sqrt{t(1-t)}dQ(t) \\ & \leq Ed_1(F_n, F) \\ & \leq \int_{\{t:t(1-t) \leq \frac{1}{4n}\}} t(1-t)dQ(t) + \frac{1}{\sqrt{n}} \int_{\{t:t(1-t) > \frac{1}{4n}\}} \sqrt{t(1-t)}dQ(t), \end{aligned}$$

where c may be chosen to be $\frac{1}{2}5^{-4}$. They base their proof on their Lemma 3.8, a version of which is stated in (7.23) below. Our first result along this line is the following proposition.

Proposition 3 *For any quantile function Q and $p > 1$,*

$$(7.7) \quad \int_{\frac{1}{n}}^{1-\frac{1}{n}} E|\alpha_n(t)|dQ(t) = \int_{\frac{1}{n}}^{1-\frac{1}{n}} E|B(t)|dQ(t)(1 + O(1))$$

$$(7.8) \quad = \sqrt{\frac{2}{\pi}} \int_{\frac{1}{n}}^{1-\frac{1}{n}} \sqrt{t(1-t)}dQ(t)(1 + O(1)),$$

where B is a Brownian bridge on $[0, 1]$ and the big Oh term in (7.8) is bounded in absolute value by $c_p(r(1/n))^{1-1/p}$ for some constant c_p depending on p and

$$(7.9) \quad r(1/n) = \frac{|Q(\frac{1}{n})| + |Q(1 - \frac{1}{n})|}{n^{1/2} \int_{\frac{1}{n}}^{1-\frac{1}{n}} \sqrt{t(1-t)}dQ(t)}.$$

Furthermore, if

$$(7.10) \quad r(1/n) \rightarrow 0, \text{ as } n \rightarrow \infty,$$

then

$$(7.11) \quad \int_{\frac{1}{n}}^{1-\frac{1}{n}} E|\alpha_n(t)|dQ(t) = \int_{\frac{1}{n}}^{1-\frac{1}{n}} E|B(t)|dQ(t)(1 + o(1)).$$

Proof. Note that for any finite measure μ on $[\frac{1}{n}, 1 - \frac{1}{n}]$ and functions f and g in $L_1([\frac{1}{n}, 1 - \frac{1}{n}], \mu)$,

$$\begin{aligned} & \left| \int_{\frac{1}{n}}^{1-\frac{1}{n}} E|f(t)|d\mu(t) - \int_{\frac{1}{n}}^{1-\frac{1}{n}} E|g(t)|d\mu(t) \right| \\ & \leq \int_{\frac{1}{n}}^{1-\frac{1}{n}} E|f(t) - g(t)|d\mu(t). \end{aligned}$$

Applying this fact we get with obvious choices of f , g and μ

$$\begin{aligned} & \left| \int_{\frac{1}{n}}^{1-\frac{1}{n}} E |\alpha_n(t)| dQ(t) - \int_{\frac{1}{n}}^{1-\frac{1}{n}} E |B(t)| dQ(t) \right| \\ & \leq \int_{\frac{1}{n}}^{1-\frac{1}{n}} E |\alpha_n(t) - B_n(t)| dQ(t). \end{aligned}$$

This last bound is, in turn, with $v = 1/2 - 1/(2p)$ ($1/2 - v = 1/(2p)$)

$$\begin{aligned} & \leq E \Delta_{n,\nu}(1) \int_{\frac{1}{n}}^{1-\frac{1}{n}} (t(1-t))^{1/2-v} dQ(t) n^{-v} \\ & = E \Delta_{n,\nu}(1) \int_{\frac{1}{n}}^{1-\frac{1}{n}} (t(1-t))^{1/(2p)} dQ(t) n^{-1/2+1/(2p)}, \end{aligned}$$

which by an application of Proposition 1 is for some positive constant C_p ,

$$(7.12) \quad \leq C_p \int_{\frac{1}{n}}^{1-\frac{1}{n}} (t(1-t))^{1/(2p)} dQ(t) n^{-1/2+1/(2p)}$$

and by Hölder's inequality is

$$\begin{aligned} & \leq C_p \left(\int_{\frac{1}{n}}^{1-\frac{1}{n}} \sqrt{t(1-t)} dQ(t) \right)^{1/p} \\ & \quad \times \left(\left| Q\left(\frac{1}{n}\right) \right| + \left| Q\left(1 - \frac{1}{n}\right) \right| \right)^{1-1/p} n^{-1/2+1/(2p)} \\ & = C_p \int_{\frac{1}{n}}^{1-\frac{1}{n}} \sqrt{t(1-t)} dQ(t) \left(\frac{\left| Q\left(\frac{1}{n}\right) \right| + \left| Q\left(1 - \frac{1}{n}\right) \right|}{n^{1/2} \int_{\frac{1}{n}}^{1-\frac{1}{n}} \sqrt{t(1-t)} dQ(t)} \right)^{1-1/p}. \end{aligned}$$

Noting that for each $t \in (0, 1)$,

$$(7.13) \quad E |B(t)| = E |Z| \sqrt{t(1-t)} = \sqrt{\frac{2}{\pi}} \sqrt{t(1-t)},$$

we see that for $c_p = C_p \sqrt{\frac{\pi}{2}}$ the last bound

$$= \int_{\frac{1}{n}}^{1-\frac{1}{n}} E |B(t)| dQ(t) c_p (r(1/n))^{1-1/p}.$$

Notice that by (7.3)

$$(7.14) \quad r^2(1/n) \leq 2n^{-1} \left(Q^2\left(\frac{1}{n}\right) + Q^2\left(1 - \frac{1}{n}\right) \right) / \sigma^2(1/n),$$

where $\sigma^2(1/n)$ is defined in (7.1). It is shown in the proof of Lemma 2.1 of Csörgő, Haeusler and Mason (1988b) that

$$(7.15) \quad \limsup_{n \rightarrow \infty} n^{-1} \left(Q^2 \left(\frac{1}{n} \right) + Q^2 \left(1 - \frac{1}{n} \right) \right) / \sigma^2(1/n) \leq 1.$$

Therefore under absolutely no conditions on Q we have (7.7). Whereas if (7.10) holds we have (7.11). \square

Corollary 1 *If $F \in DN$, then*

$$(7.16) \quad \int_0^1 E |\alpha_n(t)| dQ(t) = \int_{\frac{1}{n}}^{1-\frac{1}{n}} E |B(t)| dQ(t) (1 + o(1)).$$

Proof. If $F \in DN$, by (7.14) and (7.4), (7.10) holds. Thus

$$(7.17) \quad \int_{\frac{1}{n}}^{1-\frac{1}{n}} E |\alpha_n(t)| dQ(t) = \int_{\frac{1}{n}}^{1-\frac{1}{n}} E |B(t)| dQ(t) (1 + o(1)).$$

To finish the proof it suffices to prove that

$$(7.18) \quad \left(\int_0^{\frac{1}{n}} E |\alpha_n(t)| dQ(t) + \int_{1-\frac{1}{n}}^1 E |\alpha_n(t)| dQ(t) \right) / \tau(1/n) \rightarrow 0.$$

Since $\tau(1/n) \geq \sigma(\frac{1}{n})$, to show this it is enough to verify that

$$(7.19) \quad \left(\int_0^{\frac{1}{n}} E |\alpha_n(t)| dQ(t) + \int_{1-\frac{1}{n}}^1 E |\alpha_n(t)| dQ(t) \right) / \sigma(1/n) \rightarrow 0.$$

Notice that since $E |\alpha_n(t)| \leq 2\sqrt{nt}$, we have

$$\begin{aligned} \int_0^{\frac{1}{n}} E |\alpha_n(t)| dQ(t) / \sigma(1/n) &\leq 2\sqrt{n} \int_0^{\frac{1}{n}} t dQ(t) / \sigma(1/n) \\ &\leq \left(\frac{2}{\sqrt{n}} \left| Q \left(\frac{1}{n} \right) \right| + 2\sqrt{n} \int_0^{\frac{1}{n}} |Q(t)| dt \right) / \sigma(1/n) \\ &\leq \left(\frac{2}{\sqrt{n}} \left| Q \left(\frac{1}{n} \right) \right| + 2\sqrt{n} \int_0^{\frac{1}{n}} t^{-1/2} \sigma(t) dt \sup_{0 < t \leq 1/n} \frac{\sqrt{t} |Q(t)|}{\sigma(t)} \right) / \sigma(1/n) \\ &= \frac{2}{\sqrt{n}} \frac{|Q(\frac{1}{n})|}{\sigma(1/n)} + o(1) = o(1), \end{aligned}$$

where in the last step we use the facts that $F \in DN$ is equivalent to σ being slowly varying at zero and that $F \in DN$ implies $\sup_{0 < t \leq 1/n} \frac{\sqrt{t} |Q(t)|}{\sigma(t)} = o(1)$. (We pointed out these two facts in (7.4) and (7.5) above.) This proves the first part of (7.19). The second part of (7.19) is proved in the same way. \square

Remark Notice that in the special case when F is symmetric about zero and $F(x) = \frac{1}{2}(1+x)^{-2}$ for $x \geq 0$, we have $F \in DN$ and

$$\int_0^1 E|\alpha_n(t)| dQ(t) \sim \frac{\log n}{\sqrt{\pi}}, \text{ as } n \rightarrow \infty.$$

Remark Clearly (7.16) holds whenever $r(1/n) \rightarrow 0$, as $n \rightarrow \infty$. It also implies, as $n \rightarrow \infty$,

$$(7.20) \quad \left(\int_0^{1/n} E|\alpha_n(t)| dQ(t) + \int_{1-1/n}^1 E|\alpha_n(t)| dQ(t) \right) / \tau(1/n) \rightarrow 0.$$

The proof of Corollary 1 shows that whenever $F \in DN$, both $r(1/n) \rightarrow 0$, as $n \rightarrow \infty$, and (7.20) hold.

Assuming that $E|X| < \infty$, write for $0 < u \leq 1/2$ and $n \geq 2$,

$$(7.21) \quad \bar{\beta}_n(u) = \sqrt{n} \int_0^u t dQ(t) + \sqrt{n} \int_{1-u}^1 (1-t) dQ(t) =: \beta_{n,(-)}(u) + \beta_{n,(+)}(u).$$

Observation Whenever $E|X| < \infty$, (7.20) is satisfied if and only if

$$(7.22) \quad \bar{\beta}_n(1/n) / \tau(1/n) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Proof. Lemma 3.8 of Bobkov and Ledoux (2014) says that for an absolute constant $c > 0$ for all $0 \leq t \leq 1$,

$$(7.23) \quad c \min \left\{ 2\sqrt{nt}(1-t), \sqrt{t(1-t)} \right\} \leq E|\alpha_n(t)| \leq \min \left\{ \sqrt{n}2t(1-t), \sqrt{t(1-t)} \right\},$$

where c may chosen to be $\frac{1}{2}5^{-4}$. This implies that for all $0 < t \leq 1/n$ with $1/n \leq 1/2$

$$(7.24) \quad \frac{c\sqrt{nt}}{2} = c \min \left\{ \frac{\sqrt{nt}}{2}, \frac{\sqrt{t}}{2} \right\} \leq c \min \left\{ \sqrt{nt}(1-t), \sqrt{t(1-t)} \right\} \leq E|\alpha_n(t)| \leq 2\sqrt{nt}(1-t).$$

Using this inequality, we get for $n \geq 2$,

$$\frac{c\sqrt{n}}{2} \int_0^{1/n} t dQ(t) \leq \int_0^{1/n} E|\alpha_n(t)| dQ(t) \leq 2\sqrt{n} \int_0^{1/n} t dQ(t).$$

Obviously this implies that $\beta_{n,(-)}(1/n) / \tau(1/n) \rightarrow 0$, as $n \rightarrow \infty$, if and only if

$$\left(\int_0^{1/n} E|\alpha_n(t)| dQ(t) \right) / \tau(1/n) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

In the same way using the version of inequality (7.24) with t replaced by $1 - t$, we get $\beta_{n,(+)}(1/n)/\tau(1/n) \rightarrow 0$, as $n \rightarrow \infty$, if and only if

$$\left(\int_{1-1/n}^1 E |\alpha_n(t)| dQ(t) \right) / \tau(1/n) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

□

Remark Whenever

$$(7.25) \quad 0 < \int_{-\infty}^{\infty} \sqrt{F(x)(1-F(x))} dx = \int_0^1 \sqrt{s(1-s)} dQ(s) < \infty,$$

we have

$$\begin{aligned} \text{Var} X &= \sigma^2(0) = \int_0^1 \int_0^1 (s \wedge t - st) dQ(s) dQ(t) \\ &\leq \left(\int_0^1 \sqrt{s(1-s)} dQ(s) \right)^2 < \infty, \end{aligned}$$

which implies $0 < \text{Var} X < \infty$, and thus $F \in DN$. Hence we infer from (7.20)

$$\int_0^1 E |\alpha_n(t)| dQ(t) - \int_{\frac{1}{n}}^{1-\frac{1}{n}} E |\alpha_n(t)| dQ(t) \rightarrow 0$$

and from (7.25) that

$$\int_0^1 E |B(t)| dQ(t) - \int_{\frac{1}{n}}^{1-\frac{1}{n}} E |B(t)| dQ(t) \rightarrow 0,$$

and thus since $r^2(1/n) \rightarrow 0$ we can conclude by (7.16) that

$$\int_0^1 E |\alpha_n(t)| dQ(t) \rightarrow \int_0^1 E |B(t)| dQ(t) < \infty.$$

Proposition 4 For any quantile function Q , any $p > 1$ and any sequences of positive numbers $0 < c_n < 1 - d_n < 1$, $n \geq 1$,

$$(7.26) \quad \left| \int_{c_n}^{1-d_n} E |\alpha_n(t)| dQ(t) - \int_{c_n}^{1-d_n} E |B(t)| dQ(t) \right| \leq \sqrt{\frac{\pi}{2}} C_p (3/\sqrt{\nu}) (n(c_n \wedge d_n))^{-\nu} \int_{c_n}^{1-d_n} E |B(t)| dQ(t),$$

where $\nu = 1/2 - 1/(2p)$. In particular, if $n(c_n \wedge d_n) \rightarrow \infty$, as $n \rightarrow \infty$,

$$(7.27) \quad \int_{c_n}^{1-d_n} E |\alpha_n(t)| dQ(t) = \int_{c_n}^{1-d_n} E |B(t)| dQ(t) (1 + o(1)).$$

Proof. Notice that for $0 < c < 1 - d < 1$

$$\begin{aligned}\sigma^2(c, d) &= \int_c^{1-d} \int_c^{1-d} (s \wedge t - st) dQ(s) dQ(t) \\ &\leq \left(\int_c^{1-d} \sqrt{s(1-s)} dQ(s) \right)^2,\end{aligned}$$

and we get by the Shorack (1997) fact (5.5) that for any $0 < c < 1 - d < 1$ and $0 < v < 1/2$,

$$\int_c^{1-d} (s(1-s))^{1/2-v} dQ(s) / \int_c^{1-d} \sqrt{s(1-s)} dQ(s) \leq (3/\sqrt{v})(c \wedge d)^{-v}.$$

We see then that as in the proof of Proposition 3 that for any $p > 1$ with $v = 1/2 - 1/(2p)$,

$$\begin{aligned}& \left| \int_{c_n}^{1-d_n} E |\alpha_n(t)| dQ(t) - \int_{c_n}^{1-d_n} E |B(t)| dQ(t) \right| \\ & \leq C_p \int_{c_n}^{1-d_n} (t(1-t))^{1/(2p)} dQ(t) n^{-1/2+1/(2p)} \\ & = C_p \frac{\int_{c_n}^{1-d_n} (t(1-t))^{1/(2p)} dQ(t)}{\int_{c_n}^{1-d_n} \sqrt{s(1-s)} dQ(s)} n^{-1/2+1/(2p)} \int_{c_n}^{1-d_n} \sqrt{s(1-s)} dQ(s) \\ & \leq \sqrt{\frac{\pi}{2}} C_p (3/\sqrt{v}) (n(c_n \wedge d_n))^{-v} \sqrt{\frac{2}{\pi}} \int_{c_n}^{1-d_n} \sqrt{s(1-s)} dQ(s) \\ & = \sqrt{\frac{\pi}{2}} C_p (3/\sqrt{v}) (n(c_n \wedge d_n))^{-v} \int_{c_n}^{1-d_n} E |B(t)| dQ(t).\end{aligned}$$

□

We immediately get the following corollary.

Corollary 2 *If $E|X| < \infty$, then for all $0 < \varepsilon < 1$ there exists a $k > 0$ such that for $k/n \leq 1/2$*

$$\begin{aligned}\int_{(0,1)-[k/n,1-k/n]} E |G_n(t) - t| dQ(t) + \frac{1}{\sqrt{n}} \int_{k/n}^{1-k/n} E |B(t)| dQ(t) (1 - \varepsilon) \\ \leq Ed_1(F_n, F)\end{aligned}$$

(7.28)

$$\leq \int_{(0,1)-[k/n,1-k/n]} E |G_n(t) - t| dQ(t) + \frac{1 + \varepsilon}{\sqrt{n}} \int_{k/n}^{1-k/n} E |B(t)| dQ(t).$$

Note that whenever $E|X| < \infty$, by applying inequality (7.23), we get for $n \geq k/2$,

$$(7.29) \quad \int_{(0,1)-[k/n,1-k/n]} E|G_n(t) - t| dQ(t) \leq 2 \int_{(0,1)-[k/n,1-k/n]} t(1-t) dQ(t),$$

where the right side of (7.29) is finite. Furthermore we can say that for any $0 < \varepsilon < 1$ there exists a $k > 0$ such that for large enough $n \geq 1$

$$\begin{aligned} \frac{1}{\sqrt{n}} \int_{k/n}^{1-k/n} E|B(t)| dQ(t) (1-\varepsilon) &\leq Ed_1(F_n, F) \\ &\leq \varepsilon + \frac{1}{\sqrt{n}} \int_{k/n}^{1-k/n} E|B(t)| dQ(t) (1+\varepsilon). \end{aligned}$$

A Result of del Barrio, Giné and Matrán (1999)

Set

$$W_n = n \int_{-\infty}^{\infty} |F_n(x) - F(x)| dx.$$

Del Barrio, Giné and Matrán (1999) using the weighted approximation of Theorem 1, derived the asymptotic distribution of W_n whenever $F \in DN$ and satisfies some additional conditions. For instance, if (7.25) is satisfied then

$$(7.30) \quad \sqrt{n} \int_{-\infty}^{\infty} |F_n(x) - F(x)| dx =_d \int_0^1 |\alpha_n(s)| dQ(s) \rightarrow_d \int_0^1 |B(s)| dQ(s).$$

Condition (7.25) is a bit stronger than $0 < Var X = \sigma^2 < \infty$ and it is necessary for the limit integral to exist. Notice that if we remove the absolute values signs in (7.30) we get the usual central limit theorem, namely, in the $0 < \sigma^2 < \infty$ case

$$\sqrt{n} \int_{-\infty}^{\infty} \{F_n(x) - F(x)\} dx \rightarrow_d \sigma Z =_d \int_0^1 B(s) Q(s).$$

Along the way, in their study, del Barrio, Giné and Matrán (1999) proved that whenever $F \in DN$, for all $0 < r < 2$,

$$(7.31) \quad \sup_{n \geq 1} E \left| \frac{W_n - EW_n}{b_n} \right|^r < \infty.$$

We shall demonstrate how Theorem 3 leads to a quick proof of this result. This was the original motivation for the Giné question alluded to in the previous chapter.

An Equivalent Version of the del Barrio, Giné and Matrán Result (7.31)

Observing that by the probability integral transformation,

$$W_n =_d n \int_0^1 |G_n(t) - t| dQ(t),$$

we see that their result is equivalent to, for all $0 < r < 2$,

$$(7.32) \quad \sup_{n \geq 2} E \left| \frac{\int_0^1 \{|\alpha_n(t)| - E|\alpha_n(t)|\} dQ(t)}{\sigma(1/n)} \right|^r < \infty.$$

In a separate technical lemma they showed that whenever $F \in DN$, for all $0 < r < 2$,

$$(7.33) \quad \sup_{n \geq 2} E \left| \frac{\int_{[1/n, 1-1/n]^c} \{|\alpha_n(t)| - E|\alpha_n(t)|\} dQ(t)}{\sigma(1/n)} \right|^r < \infty$$

and they used Talagrand's (1996) exponential inequality to prove that for all $r > 0$,

$$(7.34) \quad \sup_{n \geq 2} E \left| \frac{\int_{1/n}^{1-1/n} \{|\alpha_n(t)| - E|\alpha_n(t)|\} dQ(t)}{\sigma(1/n)} \right|^r < \infty.$$

Clearly (7.33) and (7.34) imply (7.31).

A Weighted Approximation Approach to (7.34)

Giné asked the question whether it is true that on the space of Theorem 2 for all $r > 0$,

$$(7.35) \quad \sup_{n \geq 2} E \left[\sup_{1/n \leq t \leq 1-1/n} \frac{n^\nu |\alpha_n(t) - B_n(t)|}{(t(1-t))^{1/2-\nu}} \right]^r < \infty?$$

In which case, a weighted approximation approach could be used to show that for all $r > 0$, (7.34) holds.

This was the motivation for the author to establish Theorem 2, which we have shown in Proposition 1 implies (7.35). We shall use Proposition 2 and some pieces from del Barrio, Giné and Matrán (1999) to prove that (7.34) holds for all $r > 0$, under no assumptions on F . Their proof of (7.34), based on Talagrand (1996), assumes $F \in DN$.

Our aim will be to transfer our study of the moment behavior of

$$\frac{\int_{1/n}^{1-1/n} \{|\alpha_n(t)| - E|\alpha_n(t)|\} dQ(t)}{\sigma(1/n)}$$

to that of

$$\frac{\int_{1/n}^{1-1/n} \{|B_n(t)| - E|B_n(t)|\} dQ(t)}{\sigma(1/n)}.$$

What follows is somewhat technical, however, it demonstrates nicely the power of Theorem 3.

Step 1.

For any quantile function Q , one has for any $0 < \nu < 1/2$ (see the Shorack (1997) fact (5.5))

$$\sup_{n \geq 2} \frac{\int_{1/n}^{1-1/n} (s(1-s))^{1/2-\nu} dQ(s)}{n^\nu \sigma(1/n)} \leq \frac{3}{\sqrt{\nu}}.$$

Thus from Proposition 2, (with $M = \frac{3}{\sqrt{\nu}}$ and $d_n = n^\nu \sigma(1/n)$), we get for any $0 < \nu < 1/2$, on the probability space of the KMT (1975) approximation there exists a $\gamma > 0$ such that

$$(7.36) \quad \sup_{n \geq 2} E \exp(\gamma n^\nu I_n) < \infty,$$

where

$$I_n := \frac{\int_{1/n}^{1-1/n} |\alpha_n(s) - B_n(s)| dQ(s)}{n^\nu \sigma(1/n)}.$$

Step 2.

Noting that

$$n^\nu I_n = \frac{\int_{1/n}^{1-1/n} |\alpha_n(s) - B_n(s)| dQ(s)}{\sigma(1/n)},$$

we see that (7.36) implies that for any $r > 0$

$$(7.37) \quad \sup_{n \geq 2} E \left| \frac{\int_{1/n}^{1-1/n} \{|\alpha_n(s)| - |B_n(s)|\} dQ(s)}{\sigma(1/n)} \right|^r < \infty$$

Step 3

To finish the proof it clearly suffices to show that for all $r > 0$

$$(7.38) \quad \sup_{n \geq 2} E \left| \frac{\int_{1/n}^{1-1/n} \{|B_n(t)| - E|B_n(t)|\} dQ(t)}{\sigma(1/n)} \right|^r < \infty.$$

By recopying steps from the proof of Theorem 5.1 of del Barrio, Giné and Matrán, (also see their Proposition 6.2), based on the Borell (1975) inequality, one gets the exponential inequality, for all $t > 0$

$$P \left\{ \frac{\left| \int_{1/n}^{1-1/n} \{|B(t)| - E|B(t)|\} dQ(t) \right|}{\sigma(1/n)} > t \right\} \leq 2 \exp\left(-\frac{2t^2}{\pi^2}\right),$$

which clearly implies (7.38).

Notice again that absolutely no assumptions are required on the underlying F . This in combination with (7.37) finishes our proof based on weighted approximations of the del Barrio, Giné and Matrán (1999) result (7.34). In the end they decided to stick to their original proof of (7.34) based on Talagrand (1996), assuming $F \in DN$.

Remark In passing we record that a simple proof of (7.38) when $r = 2$ follows from a covariance formula of Nabeya (1951).

Nabeya's (1951) Covariance Formula

Let Z_1 and Z_2 be two standard normal random variables with correlation ρ . Then the covariance

$$0 \leq Cov(|Z_1|, |Z_2|) = \frac{2}{\pi} \left[\rho \arcsin \rho + \sqrt{1 - \rho^2} - 1 \right] \leq |\rho|.$$

In particular this implies that

$$0 \leq Cov(|B(s)|, |B(t)|) \leq Cov(B(s), B(t))$$

and thus

$$\begin{aligned} & E \left(\int_{1/n}^{1-1/n} (|B(t)| - E|B(t)|) dQ(t) \right)^2 \\ &= \int_{1/n}^{1-1/n} \int_{1/n}^{1-1/n} Cov(|B(s)|, |B(t)|) dQ(s) dQ(t) \\ &\leq \int_{1/n}^{1-1/n} \int_{1/n}^{1-1/n} Cov(B(s), B(t)) dQ(s) dQ(t) = \sigma^2(1/n). \end{aligned}$$

This obviously implies that

$$(7.39) \quad \sup_{n \geq 2} E \left(\frac{\int_{1/n}^{1-1/n} \{|B(t)| - E|B(t)|\} dQ(t)}{\sigma(1/n)} \right)^2 \leq 1.$$

Notice that absolutely no assumptions is required on the underlying distribution function (quantile function) for (7.39) to hold .

Concluding observations

Piecing all of our inequalities together we can conclude that for suitable constants $A > 0$ and $C > 0$, for all $n \geq 2$ and $t > 0$,

$$(7.40) \quad P \left\{ \frac{\left| \int_{1/n}^{1-1/n} \{ |\alpha_n(t)| - E |\alpha_n(t)| \} dQ(t) \right|}{\sigma(1/n)} > t \right\} \leq A \exp(-Ct).$$

Notice once more that absolutely no assumptions are required on F for (7.40) to hold.

For additional investigations along this line consult Haeusler and Mason (2003), who study the asymptotic distribution of the appropriately centered and normed moderately trimmed Wasserstein distance

$$\int_{Q(a_n/n)}^{Q(1-a_n/n)} |F_n(x) - F(x)| dx =_d \int_{a_n/n}^{1-a_n/n} |G_n(t) - t| dQ(t),$$

where a_n is a sequence of positive constants satisfying $a_n \rightarrow 0$ and $na_n \rightarrow \infty$. See Haeusler and Mason (2003) for motivation. As part of a general investigation of the trimmed p^{th} Mallows distance, Munk and Czado (1998) had previously looked at a somewhat different version of the trimmed Wasserstein distance when $0 < a_n = \alpha < 1/2$. Check their paper for details.

Comunicaciones del CIMAT



CIMAT

CHAPTER 8

The Quantile Process

M. Csörgő and Révész (1981) and M. Csörgő (1983) have shown the notion of a quantile process to be very useful in the study of the asymptotic properties of statistical estimators and tests. In this chapter we shall review some of the highlights of what is known about Gaussian approximations (unweighted and weighted) to this process. We begin with the definition of the uniform quantile process.

The Uniform Quantile Process

For each $n \geq 1$, let $U_{1,n} \leq \dots \leq U_{n,n}$ denote the order statistics of U_1, \dots, U_n . Define the empirical quantile function on $[0, 1]$

$$U_n(t) = U_{k,n}, \quad (k-1)/n < t \leq k/n, \quad \text{for } k = 1, \dots, n,$$

and $U_n(0) = U_{1,n}$, and the *uniform quantile process*

$$(8.1) \quad \beta_n(t) = \sqrt{n}\{t - U_n(t)\}, \quad \text{for } 0 \leq t \leq 1.$$

One easily checks that

$$U_n(t) = G_n^{-1}(t) = \inf \{x : G_n(x) \geq t\}, \quad 0 < t < 1,$$

and

$$\sup_{0 \leq t \leq 1} |\beta_n(t)| = \sup_{0 \leq t \leq 1} |\alpha_n(t)|.$$

The M. Csörgő and Révész (1979) Brownian Bridge Approximation to β_n

Theorem [MCsR] *There exists a probability space (Ω, A, P) with independent Uniform $(0, 1)$ random variables U_1, U_2, \dots , and a sequence of Brownian bridges B_1, B_2, \dots , such that for all $n \geq 1$ and $-\infty < x < \infty$,*

$$(8.2) \quad P \left\{ \sup_{0 \leq t \leq 1} |\beta_n(t) - B_n(t)| \geq n^{-1/2}(a \log n + x) \right\} \leq b \exp(-cx),$$

where a, b and c are suitable positive constants independent of n and x .

Notice that on the probability space of Theorem [MCsR], (8.2) implies that with probability 1,

$$(8.3) \quad \sup_{0 \leq t \leq 1} |\beta_n(t) - B_n(t)| = O\left(\frac{\log n}{\sqrt{n}}\right).$$

The Brownian bridge that appears in (8.2) is not the same one in the KMT approximation to α_n . The approximation (8.2) is obtained by noting that

$$\begin{aligned} \beta_n(k/n) &= \sqrt{n}\{k/n - U_{k,n}\} \quad \text{for } k = 1, \dots, n, \\ &= \sqrt{n}\{k/n - S_k/S_{n+1}\} \quad \text{for } k = 1, \dots, n, \end{aligned}$$

where $S_k = \sum_{i=1}^k \omega_i$, for $k = 1, \dots, n+1$, and $\omega_1, \dots, \omega_{n+1}$ are i.i.d. exponential random variables with mean 1 and applying the following special case of the Komlós, Major, and Tusnády [KMT] (1975) Brownian motion approximation to the partial sum process given in the following result:

Theorem [KMT] *Let ω be an exponential random variable with variance 1. Then on the same probability space there exist i.i.d. ω random variables $\omega_1, \omega_2, \dots$, and i.i.d. standard normal random variables Z_1, Z_2, \dots , such that for positive constants C, D and λ for all $x \in \mathbb{R}$ and $n \geq 1$,*

$$(8.4) \quad P\left\{\max_{1 \leq k \leq n} \left| \sum_{i=1}^k \omega_i - k - \sum_{i=1}^k Z_i \right| > D \log n + x\right\} \leq C \exp(-\lambda x).$$

The General Quantile Process

For each $n \geq 1$, let $X_{1,n} \leq \dots \leq X_{n,n}$ denote the order statistics of X_1, \dots, X_n i.i.d. with common cdf F and with quantile function Q . Define the empirical quantile function on $[0, 1]$

$$(8.5) \quad Q_n(t) = X_{k,n}, \quad (k-1)/n < t \leq k/n, \quad \text{for } k = 1, \dots, n,$$

and $Q_n(0) = X_{1,n}$. It is readily checked that for $t \in (0, 1)$

$$Q_n(t) = \inf \{x : F_n(x) \geq t\},$$

so that $Q_n(t)$ is the quantile or inverse function of the empirical distribution function F_n . We define the general quantile process

$$(8.6) \quad q_n(t) = \sqrt{n}\{Q(t) - Q_n(t)\}, \quad \text{for } 0 < t < 1.$$

Notice that we avoid $t = 0$ and $t = 1$ since it could happen that $Q(0) = -\infty$ or $Q(1) = \infty$. Assume that F has a density function f , which is strictly positive on the set where $0 < F(x) < 1$, plus some other regularity conditions. M. Csörgő and Révész (1979, 1981) showed that there exists a probability space (Ω, A, P) with X_1, X_2, \dots , i.i.d.

with common cdf F and a sequence of Brownian bridges B_1, B_2, \dots , such that, with probability 1,

$$(8.7) \quad \sup_{0 < t < 1} |f(Q(t))q_n(t) - B_n(t)| = O\left(\frac{\log n}{\sqrt{n}}\right).$$

The proof is based on (8.3), using the mean value theorem, noting that $Q'(t) = 1/f(Q(t))$, and applying the fact that

$$\{q_n(t), 0 < t < 1\}_{n \geq 1} =_d \{\sqrt{n}\{Q(t) - Q(U_n(t))\}, 0 < t < 1\}_{n \geq 1}.$$

For more details see M. Csörgő and Révész (1979, 1981).

The Bahadur-Kiefer process The process

$$\alpha_n(t) - \beta_n(t), \quad 0 \leq t \leq 1,$$

is called the *Bahadur-Kiefer process*. Surprisingly one can show that, with probability 1, for each $0 \leq t \leq 1$,

$$\limsup_{n \rightarrow \infty} \frac{n^{1/4} \pm (\alpha_n(t) - \beta_n(t))}{(\log \log n)^{3/4}} = 2^{5/4} 3^{-3/4} (t(1-t))^{1/4}.$$

Refer to Kiefer (1967) and Arcones and Mason (1997). Moreover, one has, with probability 1,

$$\limsup_{n \rightarrow \infty} \frac{n^{1/4} \sup_{0 \leq t \leq 1} |\alpha_n(t) - \beta_n(t)|}{\sqrt{\log n} \sqrt[4]{\log \log n}} = 2^{-1/4}$$

and

$$\lim_{n \rightarrow \infty} \frac{n^{1/4} \sup_{0 \leq t \leq 1} |\alpha_n(t) - \beta_n(t)|}{\sqrt{\log n} \sqrt{\sup_{0 \leq t \leq 1} |\alpha_n(t)|}} = 1.$$

See Kiefer (1970) and for more details, Deheuvels and Mason (1990). This second statement implies that

$$\lim_{n \rightarrow \infty} \frac{n^{1/4} \sup_{0 \leq t \leq 1} |\alpha_n(t) - \beta_n(t)|}{\sqrt{\log n}} \rightarrow_d \sqrt{\sup_{0 \leq t \leq 1} |B(t)|}.$$

For any $n \geq 2$ and $0 \leq \nu < 1/4$ set

$$K_{n,\nu} = \sup_{1/n \leq t \leq 1-1/n} \frac{n^\nu |\alpha_n(t) - \beta_n(t)|}{(t(1-t))^{1/2-\nu}}$$

and

$$\Gamma_{n,\nu} = \sup_{1/n \leq t \leq 1-1/n} \frac{n^\nu |\beta_n(t) - B_n(t)|}{(t(1-t))^{1/2-\nu}}.$$

Cs-Cs-H-M (1986) (see also Mason (1991)) proved that for any $0 \leq \nu < 1/4$, $K_{n,\nu} = O_p(1)$. This implies that on the probability space of the Mason and van Zwet theorem one has $\Gamma_{n,\nu} = O_p(1)$. On the

Cs-Cs-H-M (1986) space $\Gamma_{n,\nu} = O_p(1)$ for any $0 \leq \nu < 1/2$, whereas $\Delta_{n,\nu}(1) = O_p(1)$ for any $0 \leq \nu < 1/4$. So the probability spaces of Theorems 1 and 2 are, in a sense, the duals of each other. Cs-Cs-H-M (1986) used ideas from M. Csörgő and Révész (1979) to prove the following weighted approximation for the uniform quantile process:

Theorem Cs-Cs-H-M (1986). *On a rich enough probability space there exists a sequence of independent Uniform $(0, 1)$ random variables U_1, U_2, \dots , and a sequence of Brownian bridges B_1, B_2, \dots , such that for the uniform empirical processes β_n based on the U_i 's and all $0 \leq \nu < \frac{1}{2}$*

$$(8.8) \quad \sup_{1/(n+1) \leq t \leq 1-n/(n+1)} \frac{|\beta_n(t) - B_n(t)|}{(t(1-t))^{1/2-\nu}} = O_p(n^{-\nu}).$$

They inferred the weighted approximation (5.1) for α_n for $0 \leq \nu < \frac{1}{4}$ from the fact that for any such ν

$$\sup_{1/(n+1) \leq t \leq 1-n/(n+1)} \frac{n^\nu |\beta_n(t) - \alpha_n(t)|}{(t(1-t))^{1/2-\nu}} = O_p(1).$$

On the probability space of Theorem 3 the ranges ν are reversed.

More Exponential Inequalities

Mason (2001b) derived the following exponential inequalities for $K_{n,\nu}$ and $\Gamma_{n,\nu}$.

Theorem 4. *For every $0 \leq \nu < 1/4$ there exist positive constants b_ν and c_ν such that for all $n \geq 2$ and $0 \leq x < \infty$,*

$$(8.9) \quad P \{K_{n,\nu} \geq x\} \leq b_\nu \exp(-c_\nu x)$$

and there exist positive constants A_ν and d_ν such that for all $n \geq 2$ and $0 \leq x < \infty$,

$$(8.10) \quad P \{\Gamma_{n,\nu} \geq x\} \leq A_\nu \exp(-d_\nu x).$$

More about weighted approximations to the quantile process and their applications can be found in M. Csörgő, M. and Horváth (1993).

CHAPTER 9

A Non-Hungarian Construction

The original proof of Theorem 1 given by Cs-Cs-H-M (1986) was based on the KMT (1975, 1976) Wiener process strong approximation to the partial sum process. Mason and van Zwet (1987) derived their version through their refinement of the KMT (1975) Brownian bridge approximation to the uniform empirical process stated in Theorem 2 above. For this reason SW (1986) refer to these weighted approximations as Hungarian constructions.

To establish this approximation in its full strength, i.e. (5.1) holds for all $0 \leq \nu < \frac{1}{2}$, the use of the KMT construction seems to be unavoidable. For the overwhelming majority of situations, it suffices for (5.1) to hold for e.g. $0 < \nu < \frac{1}{4}$. But for this range of ν 's such a construction can be obtained by a much less involved tool, namely, the *Skorohod embedding* scheme as shown by Mason (1991) and M. Csörgő and Horváth (1986). It is based on the fact that one can use the Skorohod (1965) embedding, cf. Breiman (1967), to show that there exist on the same probability space a standard Wiener process W on $[0, \infty)$ and a sequence of i.i.d. exponential random variables $\omega_1, \omega_2, \dots$ with mean 1 such that for all $2 < p < 4$, with probability 1,

$$(9.1) \quad m^{-1/p} |S_m - m - W(m)| \rightarrow 0, \text{ as } m \rightarrow \infty,$$

where for any $m \geq 1$, $S_m = \omega_1 + \dots + \omega_m$. More specifically, applying Breiman (1967) to this situation we get that there exist a sequence of i.i.d. positive random variables τ_1, τ_2, \dots , such that $E\tau_1 = Var(\omega_1)$ and a standard Wiener process W on $[0, \infty)$ such that as a sequence of random variables,

$$\{S_m - m\}_{m \geq 1} =_d \left\{ W \left(\sum_{i=1}^m \tau_i \right) \right\}_{m \geq 1}$$

and (9.1) holds. The following distributional representation of the uniform order statistics (previously stated in (2.7)) for each $n \geq 1$,

$$(U_{1,n}, \dots, U_{n,n}) =_d \left(\frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}} \right),$$

also comes into play.

Theorem S. *On a rich enough probability space there exists a sequence of independent Uniform $(0, 1)$ random variables U_1, U_2, \dots , and a sequence of Brownian bridges B_1, B_2, \dots , such that for the uniform empirical processes α_n based on the U_i 's and all $0 < \nu < \frac{1}{4}$*

$$(9.2) \quad \sup_{0 \leq s \leq 1} \frac{|\alpha_n(s) - B_n(s)|}{(s(1-s))^{1/2-\nu}} = O_p(n^{-\nu}).$$

Moreover, statement (9.2) remains true for $\nu = 0$ when B_n is replaced by \bar{B}_n , where \bar{B}_n is defined as in (5.2). Furthermore,

$$(9.3) \quad \sup_{1/(n+1) \leq s \leq n/(n+1)} \frac{|\beta_n(s) - B_n(s)|}{(s(1-s))^{1/2-\nu}} = O_p(n^{-\nu}).$$

Proof. As in Cs-Cs-H-M (1986) let $\{W^{(i)}(s) : 0 \leq s < \infty\}$, $i = 1, 2$, be two independent standard Wiener processes sitting on the same probability space. On this probability space construct using the Skorohod embedding (9.1) two independent sequences of independent exponential random variables with expectation one, $Y_1^{(i)}, Y_2^{(i)}, \dots$, as a function of $W^{(i)}$, $i = 1, 2$. These sequences have the property that for each $i = 1, 2$, and $2 < p < 4$, cf. Breiman (1967),

$$(9.4) \quad m^{-1/p} |S_m^{(i)} - m - W^{(i)}(m)| \rightarrow 0, \text{ a.s. as } m \rightarrow \infty,$$

where for $i = 1, 2$, and $m \geq 1$

$$S_m^{(i)} = \sum_{j=1}^m Y_j^{(i)}.$$

For each integer $n \geq 2$, let

$$Y_j(n) = \begin{cases} Y_j^{(1)} & \text{for } j = 1, \dots, \lfloor n/2 \rfloor \\ Y_{n+2-j}^{(2)} & \text{for } j = \lfloor n/2 \rfloor + 1, \dots, n+1, \end{cases}$$

where $\lfloor x \rfloor$ is the integer part of x , and set

$$S_m(n) = \sum_{j=1}^m Y_j(n) \text{ for } m = 1, \dots, n+1.$$

For notational convenience, we write from now on Y_j and S_m for $Y_j(n)$ and $S_m(n)$. Also for each integer $n \geq 2$, let $W_n(s) = W^{(1)}(s)$ for $0 \leq s \leq \lfloor n/2 \rfloor$, and $W_n(s) = W^{(1)}(\lfloor n/2 \rfloor) + W^{(2)}(n+1 - \lfloor n/2 \rfloor) - W^{(2)}(n+1 - s)$ for $\lfloor n/2 \rfloor < s \leq n+1$. It is easily checked that for each integer $n \geq 2$, $\{W_n(s) : 0 \leq s \leq n+1\}$ is a standard Wiener process on $[0, n+1]$.

We will first show that for each $0 \leq \nu < \frac{1}{4}$,

$$(9.5) \quad C_{n,1} = O_P(1),$$

where $C_{n,1} =$

$$\max_{1 \leq i \leq n/2} \left| n \left\{ \frac{S_i}{S_{n+1}} - \frac{i}{n} \right\} - \left\{ W_n(i) - \frac{i}{n} W_n(i) \right\} \right| i^{-1/2+\nu},$$

and

$$(9.6) \quad C_{n,2} = O_P(1),$$

where $C_{n,2} =$

$$\max_{n/2+1 \leq i \leq n} \left| n \left\{ \frac{S_i}{S_{n+1}} - \frac{i}{n} \right\} - \left\{ W_n(i) - \frac{i}{n} W_n(i) \right\} \right| (n+1-i)^{-1/2+\nu}.$$

First consider (9.5). Notice that by the law of large numbers

$$\begin{aligned} & \max_{1 \leq i \leq n/2} n \left| \frac{S_i}{S_{n+1}} - \frac{S_i}{S_n} \right| i^{-1/2+\nu} \\ &= \max_{1 \leq i \leq n/2} n i^{-1/2+\nu} \frac{S_i Y_{n+1}}{S_n S_{n+1}} = O_P(n^{-1/2+\nu}) = o_P(1). \end{aligned}$$

Also by the law of large numbers, the law of the iterated logarithm and the central limit theorem

$$\begin{aligned} & \max_{1 \leq i \leq n/2} \left| \frac{n}{S_n} \left\{ S_i - i - \frac{i}{n} (S_n - n) \right\} - \left\{ S_i - i - \frac{i}{n} (S_n - n) \right\} \right| i^{-1/2+\nu} \\ & \leq \left\{ \max_{1 \leq i \leq n/2} |S_i - i| i^{-1/2+\nu} + n^{-1/2+\nu} |S_n - n| \right\} \frac{|S_n - n|}{S_n} \\ & = \left\{ O_P((\log \log n)^{1/2} n^\nu) + O_P(n^\nu) \right\} O_P(n^{-1/2}) = o_P(1). \end{aligned}$$

Thus in light of the above two $o_P(1)$ statements and a little algebra, to prove (9.5) it is enough to show that

$$(9.7) \quad D_n = O_P(1),$$

where $D_n =$

$$\max_{1 \leq i \leq n/2} \left| S_i - i - \frac{i}{n} (S_n - n) - \left\{ W_n(i) - \frac{i}{n} W_n(n) \right\} \right| i^{-1/2+\nu}.$$

Observe that

$$\begin{aligned}
 D_n &\leq \max_{1 \leq i \leq n/2} \left| S_i^{(1)} - i - W_n^{(1)}(i) \right| i^{-1/2+\nu} \\
 &\quad + n^{-1/2+\nu} \left| S_{\lfloor \frac{n}{2} \rfloor}^{(1)} - \lfloor \frac{n}{2} \rfloor - W_n^{(1)}\left(\lfloor \frac{n}{2} \rfloor\right) \right| \\
 &\quad + n^{-1/2+\nu} \left| S_{n+1-\lfloor \frac{n}{2} \rfloor}^{(2)} - \left(n+1 - \lfloor \frac{n}{2} \rfloor\right) - W_n^{(2)}\left(n+1 - \lfloor \frac{n}{2} \rfloor\right) \right| \\
 &\quad + n^{-1/2+\nu} \left| Y_1^{(2)} - 1 \right| + n^{-1/2+\nu} |W_n(n+1) - W_n(n)|.
 \end{aligned}$$

Since $1/2 - \nu > 0$, we see by (9.4) that the first term on the right side of this inequality is $O_P(1)$ and the next two terms are $o_P(1)$. The last two terms are obviously $o_P(1)$ random variables. Hence we have established (9.7) and therefore (9.5). Assertion (9.6) is proved in almost the same way using the symmetry of the construction given above.

Next set for each integer $n \geq 2$

$$\tilde{U}_{i,n} = S_i/S_{n+1} \text{ for } i = 1, \dots, n,$$

and

$$\tilde{B}_n(s) = n^{-1/2} (sW_n(n) - W_n(sn)) \text{ for } 0 \leq s \leq 1.$$

We see that for each integer $n \geq 2$

$$\left(\tilde{U}_{1,n}, \dots, \tilde{U}_{n,n} \right) =_d (U_{1,n}, \dots, U_{n,n}),$$

and \tilde{B}_n is a Brownian bridge. Let $\tilde{\beta}_n$ denote the uniform quantile process based on $\tilde{U}_{1,n}, \dots, \tilde{U}_{n,n}$. We claim that

$$(9.8) \quad E_{1,n} := \sup_{1/(n+1) \leq s \leq \lfloor n/2 \rfloor / 2} \frac{n^\nu |\tilde{\beta}_n(s) - \tilde{B}_n(s)|}{s^{1/2-\nu}} = O_p(1)$$

and

$$(9.9) \quad E_{2,n} := \sup_{\lfloor n/2 \rfloor / 2 < s \leq n/(n+1)} \frac{n^\nu |\tilde{\beta}_n(s) - \tilde{B}_n(s)|}{(1-s)^{1/2-\nu}} = O_p(1).$$

Notice that

$$\begin{aligned}
 E_{1,n} &\leq 2 \max_{1 \leq i \leq n/2} \frac{n^\nu |\tilde{\beta}_n(i/n) - \tilde{B}_n(i/n)|}{(i/n)^{1/2-\nu}} \\
 &\quad + \max_{1 \leq i \leq [n/2]-1} \sup_{i/n \leq s \leq (i+1)/n} \frac{n^\nu |\tilde{B}_n(\frac{i+1}{n}) - \tilde{B}_n(s)|}{(i/n)^{1/2-\nu}} \\
 (9.10) \quad &\quad + \sup_{1/(n+1) \leq s \leq 1/n} \frac{n^\nu |\tilde{B}_n(\frac{1}{n}) - \tilde{B}_n(s)|}{(1/(n+1))^{1/2-\nu}} + 1.
 \end{aligned}$$

The first term on the right side of (9.10), we recognize to be $2C_{n,1}$, which has just been proved to be $O_p(1)$. To show that the next two terms on the right side of (9.10) are $O_p(1)$, we require the following probability inequality: for any $0 < a < 1$, $h > 0$ and $0 < u < \infty$

$$(9.11) \quad P \left\{ \sup_{s \in [a-h, a+h] \cap [0,1]} |B(a) - B(s)| \geq u\sqrt{h} \right\} \leq Au^{-1} \exp(-u^2/8),$$

where B denotes a Brownian bridge and A is a suitably chosen universal constant, cf. (1.11) of Cs-Cs-H-M (1986).

Using (9.11) it is routine to verify that for all $\varepsilon > 0$ there exists a $0 < M < \infty$ such that for all integers $n \geq 2$

$$\sum_{i=1}^{[n/2]-1} P \left\{ \sup_{i/n \leq s \leq (i+1)/n} |\tilde{B}_n(\frac{i+1}{n}) - \tilde{B}_n(s)| \geq \frac{M}{\sqrt{n}} i^{1/2-\nu} \right\} < \varepsilon,$$

which proves that the second term on the right side of (9.10) is $O_p(1)$. The fact that the third term on the right of (9.10) is $o_p(1)$ follows easily from (9.11). Thus we have established (9.8). Assertion (9.9) is proved similarly using (9.6) and (9.11).

Combining (9.8) and (9.9), we get that

$$(9.12) \quad \sup_{1/(n+1) \leq s \leq n/(n+1)} \frac{n^\nu |\tilde{\beta}_n(s) - \tilde{B}_n(s)|}{(s(1-s))^{1/2-\nu}} = O_p(1).$$

Up to this stage in the proof, we have constructed a probability space on which sit a sequence $\tilde{\beta}_n$ of versions of β_n , i.e. for each integer $n \geq 2$

$$\left\{ \tilde{\beta}_n(s) : 0 \leq s \leq 1 \right\} =_d \left\{ \beta_n(s) : 0 \leq s \leq 1 \right\},$$

and a sequence of Brownian bridges \tilde{B}_n such that (9.12) holds for all $0 \leq \nu < 1/4$. In order to construct a probability space with a sequence of Uniform $(0, 1)$ random variables U_1, U_2, \dots , and a sequence of Brownian bridges B_1, B_2, \dots , such that (9.3) holds, we now follow the procedure

given in Lemma 3.1.1 of M. Csörgő (1983). This finishes the proof of the (9.3) part of Theorem S, (9.2).

We now turn to the proof of the (9.2) part of Theorem S. First note that an easy application of the Birnbaum-Marshall inequality (see (2.18) and (2.19)) that for all $0 < \nu < 1/4$,

$$\sup_{0 < s \leq 1/n} |\alpha_n(s)| / s^{1/2-\nu} + \sup_{0 < s \leq 1/n} |\alpha_n(1-s)| / (1-s)^{1/2-\nu} = O_P(1)$$

and

$$\sup_{0 < s \leq 1/n} |B_n(s)| / s^{1/2-\nu} + \sup_{0 < s \leq 1/n} |B_n(1-s)| / (1-s)^{1/2-\nu} = O_P(1).$$

This combined with the following proposition completes the proof of (9.2) as well as Theorem S. \square

Proposition S For any $0 < d < n/2$, $n \geq 2$ and $0 \leq \nu < \frac{1}{4}$

$$(9.13) \quad \sup_{d/n \leq s \leq 1-d/n} \frac{n^\nu |\alpha_n(s) - \beta_n(s)|}{(s(1-s))^{1/2-\nu}} = O_p(1).$$

Proof. Since trivially

$$(9.14) \quad \sup_{0 \leq s \leq 1} \frac{n^\nu |\alpha_n(U_n(s)) - \beta_n(s)|}{(s(1-s))^{1/2-\nu}} = O(n^{-1/2}), \text{ a.s.},$$

to prove (9.14) it suffices to show that both

$$(9.15) \quad \sup_{d/n \leq s \leq 1} \frac{n^\nu |\alpha_n(U_n(s)) - \alpha_n(s)|}{s^{1/2-\nu}} := M_{n,1}(d) = O_p(1)$$

and

$$(9.16) \quad \sup_{0 \leq s \leq 1-d/n} \frac{n^\nu |\alpha_n(U_n(s)) - \alpha_n(s)|}{(1-s)^{1/2-\nu}} := M_{n,2}(d) = O_p(1).$$

Also on account of $M_{n,1}(d) =_d M_{n,2}(d)$ it is enough to prove (9.15). Assertion (9.15) will be an easy consequence of the following lemmas.

Lemma i For any $0 < d < 1$ and $0 \leq 2\delta < \frac{1}{2}$

$$(9.17) \quad \sup_{d/n \leq s \leq 1} \frac{n^{2\delta} |U_n(s) - s|}{s^{1/2-2\delta}} = O_p(1).$$

Proof. In Mason (1983), (also see Remark 4.4 of Marcus and Zinn (1984), it is proved that for any $0 \leq 2\delta < \frac{1}{2}$

$$\sup_{0 \leq s \leq 1} \frac{n^{2\delta} |G_n(s) - s|}{s^{1/2-2\delta}} = O_p(1).$$

In addition, from Lemma 2 of Wellner (1978) one has for any $0 < d < 1$

$$\sup_{d/n \leq s \leq 1} U_n(s)/s = O_P(1).$$

These two facts when combined with (9.14) yield (9.17). \square

For any $0 < a \leq \frac{1}{2}$, $0 \leq b < c \leq 1$ and integer $n \geq 1$ set $\omega_n(a, b, c) =$

$$\sup \{ |\alpha_n(s+h) - \alpha_n(s)| : 0 \leq s+h \leq 1, 0 \leq |h| \leq a, b \leq s \leq c \}.$$

Lemma ii For universal positive constants A and B for all $0 < a \leq \frac{1}{2}$, $0 \leq b < c \leq 1$, $n \geq 1$ and $\lambda > 0$

$$(9.18) \quad \begin{aligned} & P \{ \omega_n(a, b, c) > \lambda \sqrt{a} \} \\ & \leq \left\{ \left(\frac{c-b}{a} \right) \vee 1 \right\} A \exp \left(-B \lambda^2 \psi \left(\frac{\lambda}{\sqrt{na}} \right) \right), \end{aligned}$$

where for all $x \geq 0$

$$\psi(x) = 2x^{-2} \{ (x+1) \log(x+1) - x \}.$$

Proof. The proof is essentially contained in that of Inequality 1 in Mason, Shorack and Wellner (1983). Also see Inequality 1 of J. Einmahl and Mason (1988) (the ba^{-1} there should be replaced by $(ba^{-1}) \vee 1$). \square

For future reference we record the fact that for $x \geq 0$

$$(9.19) \quad \psi(x) \searrow \text{ as } x \nearrow.$$

Choose $0 \leq \nu < \delta < \frac{1}{4}$ and set $\rho = \delta - \nu$. Also for any $\gamma > 0$, $m \geq 1$ and $1 \leq i \leq n2^m$, let

$$\Delta_n(i) = \omega_n \left(\frac{\gamma i^{1-2\delta}}{n}, \frac{i}{n2^m}, \frac{i+1}{n2^m} \right).$$

Lemma iii For any $\varepsilon > 0$, $m \geq 1$ and $0 \leq \nu < \delta < \frac{1}{4}$, there exists a $\gamma > 0$ such that for all n large enough

$$(9.20) \quad P \left(\max_{1 \leq i < n2^m} n^\nu \Delta_n(i) / (i/n)^{1/2-\nu} > \gamma \right) < \varepsilon.$$

Proof. For all large enough n inequality (9.18) is applicable to give uniformly in $1 \leq i < n2^m$, after a little algebra

$$\begin{aligned} & P \{ \Delta_n(i) > \gamma n^{-1/2} i^{1/2-\nu} \} \\ & \leq \left\{ \left(\frac{2^m}{\gamma} \right) \vee 1 \right\} A \exp \left(-B i^{2\rho} \gamma \psi \left(\frac{i^{2\delta-\nu-1/2}}{n} \right) \right), \end{aligned}$$

which by $2\delta - \nu - 1/2 < 0$ and (9.19) is

$$\leq \left\{ \left(\frac{2^m}{\gamma} \right) \vee 1 \right\} A \exp(-B i^{2\rho} \gamma \psi(1)) := P_i(\gamma).$$

Since $\rho > 0$, for every $\varepsilon > 0$ we can choose a $\gamma > 0$ such that $\sum_{i=1}^{\infty} P_i(\gamma) < \varepsilon$. \square

Returning to the proof of (9.15), by Lemma i for any $m \geq 1$ and $\varepsilon > 0$ there exists a $\gamma > 0$ such that for all $n \geq 1$

$$P \left\{ \sup_{1/(n2^m) \leq s \leq 1} \frac{n^{2\delta} |U_n(s) - s|}{s^{1-2\delta}} \leq 2^{m(1-2\delta)} \frac{\gamma}{2} \right\} > 1 - \varepsilon.$$

Hence with probability greater than $1 - \varepsilon$

$$M_{n,1}(2^{-m}, \nu) \leq \max_{1 \leq i < n2^m} \widetilde{M}_n(i, \delta, \nu),$$

where $\widetilde{M}_n(i, \delta, \nu) =$

$$\sup \left\{ n^\nu \frac{|\alpha_n(s+h) - \alpha_n(s)|}{s^{1/2-\nu}} : 0 \leq s+h \leq 1, 0 \leq |h| \leq \frac{\gamma(i+1)^{1-2\delta}}{2n}, \right. \\ \left. \frac{i}{n2^m} \leq s \leq \frac{i+1}{n2^m} \right\},$$

which since $(\frac{i+1}{i})^{1-2\delta} / 2 < 1$, is obviously

$$\leq 2^{m(1/2-\nu)} \max_{1 \leq i < n2^m} n^\nu \Delta_n(i) / (i/n)^{1/2-\nu}.$$

A straightforward argument based on Lemma iii now verifies that $M_{n,1}(2^{-m}, \nu) = O_P(1)$. \square

Remark In Mason (1999) the proof of Proposition S is refined to give the exponential inequality for $K_{n,\nu}$ stated in Theorem 4 above.

Remark The construction just given is nearly the same as that in Cs-Cs-H-M (1986) with an important difference. Instead of using the Skorohod embedding, they used the KMT (1975,1976) approximation. This says that one can construct two independent standard Wiener processes $\{W^{(i)}(s) : 0 \leq s < \infty\}$, $i = 1, 2$, and two independent sequences of independent exponential random variables with expectation one, $Y_1^{(i)}, Y_2^{(i)}, \dots$, $i = 1, 2$, sitting on the same probability space such that for universal constants C, K and λ and all $x \geq 0$

$$P \left\{ \max_{1 \leq m \leq n} |S_m^{(i)} - W^{(i)}(m)| \geq n^{-1/2}(C \log x + x) \right\} \leq K \exp(-\lambda x),$$

where as above for $i = 1, 2$, and $m \geq 1$, $S_m^{(i)} = \sum_{j=1}^m Y_j^{(i)}$. They obtained the following.

Theorem Cs-Cs-H-M Inequalities (1986) *There exists a probability space (Ω, A, P) with independent Uniform $(0, 1)$ random variables $U_1,$*

U_2, \dots , and a sequence of Brownian bridges B_1, B_2, \dots , such that for all $n \geq 1$, $1 \leq d \leq n$, and $0 \leq x < \infty$,

$$P \left\{ \sup_{0 \leq t \leq d/n} |\beta_n(t) - B_n(t)| \geq n^{-1/2}(a \log d + x) \right\} \leq b \exp(-cx)$$

and

$$P \left\{ \sup_{1-d/n \leq t \leq 1} |\beta_n(t) - B_n(t)| \geq n^{-1/2}(a \log d + x) \right\} \leq b \exp(-cx),$$

where a, b and c are suitable positive constants independent of n, d and x .

The original version of these inequalities restricted $0 \leq x \leq \sqrt{d}$. In M. Csörgő and Horváth (1993) this restriction was lifted. From these inequalities they proved that the weighted approximation (8.8) holds on the probability space of Theorem Cs-Cs-H-H Inequalities (1986).

Comunicaciones del CIMAT



CIMAT

Some Further Progress on Weighted Approximations

Given the construction given in the previous chapter it naturally comes to mind that the martingale version of the Skorohod embedding might also be used to prove weighted approximation results for more general processes than α_n as long as they possess a certain martingale structure.

10.1. Exchangeable Processes

Shorack (1991) was the first to use the Skorohod embedding for martingales in this way. He used it to establish a weighted approximation to the finite sampling process and a weighted uniform empirical process.

Motivated by discussions with Shorack during a visit to the University of Washington in the summer of 1990, Uwe Einmahl and I, when not hiking in the North Cascades, found time to generalize Shorack's results to exchangeable processes, i.e. to processes of the form

$$(10.1) \quad \varepsilon_n(t) = n^{-1/2} \sum_{i \leq nt} Y_n(i), \quad 0 \leq t \leq 1,$$

where for every $n \geq 1$ the random variables $Y_n(1), \dots, Y_n(n)$ are exchangeable.

Assume that

- (i) $\sum_{i=1}^n Y_n(i) = 0$,
- (ii) $\frac{1}{n} \sum_{i=1}^n Y_n^2(i) \rightarrow_P \sigma^2$ for some $\sigma^2 > 0$, and
- (iii) $\max_{1 \leq i \leq n} Y_n^2(i)/n \rightarrow_P 0$.

Then by Theorem 24.3 of Billingsley (1968) one concludes that ε_n converges weakly to σB , where B is a Brownian bridge. Under additional regularity conditions, U. Einmahl and Mason (1992) were able to obtain the following weighted approximation to ε_n .

Theorem 5. *Assume (i) and replace (ii) by*

$$(iv) \quad \frac{1}{n} \sum_{i=1}^n Y_n^2(i) = \sigma^2 + O_P(n^{-1/2}),$$

and (iii) by

(v) $EY_n^4(1) \leq K < \infty$ for some $K > 0$ and all $n \geq 1$.

Then on a suitable probability space there exist a sequence of probabilistically equivalent versions $\tilde{\varepsilon}_n$ of ε_n and a sequence of Brownian bridges B_1, B_2, \dots , such that for all $0 \leq \nu < 1/4$ and $\tau > 0$

$$(10.2) \quad \sup_{\tau/n \leq t \leq 1-\tau/n} \frac{n^\nu |\tilde{\varepsilon}_n(t) - \sigma B_n(t)|}{(t(1-t))^{1/2-\nu}} = O_P(1).$$

U. Einmahl and Mason (1992) point out that condition (v) can be weakened to

$$(10.3) \quad E|Y_n|^\gamma(1) \leq K < \infty \text{ for some } \gamma > 2 \text{ and } K > 0 \text{ and all } n \geq 1,$$

with a corresponding restriction on ν in the conclusion (10.2). Kirch (2003) has carried out the needed analysis to verify this. (Also see Theorem D.1 in the Appendix of Kirch (2006).) Her calculations show that when (v) is replaced by (10.3) and (iv) by

$$\frac{1}{n} \sum_{i=1}^n Y_n^2(i) = \sigma^2 + O_P(n^{-2s}),$$

where $s = \min\left(\frac{\gamma-2}{2\gamma}, \frac{1}{4}\right)$, then (10.2) is valid for all $0 \leq \nu < s$. This result could also be derived with some difficulty from the general weighted approximation to continuous time martingales given in Theorem 1 of Haeusler and Mason (1999).

U. Einmahl and Mason (1992) obtained the approximation (5.1) stated in Theorem 1 and those in Shorack (1991) as special cases of their approximation, as well as weighted approximations for a number of other interesting examples. Recently Kirch and Steinebach (2006) and Kirch (2006, 2007, 2008) have used the U. Einmahl and Mason (1992) weighted approximation to derive the limiting distribution of certain permutation tests for a change point.

Some Special Cases

1. Set

$$Y_n(i) = n \left\{ G_n\left(\frac{i}{n}\right) - G_n\left(\frac{i-1}{n}\right) \right\} - 1, \quad i = 1, \dots, n.$$

This choice yields a version of the weighted approximation (5.1) to the uniform empirical process given in Theorem 1.

2. Set

$$Y_n(i) = 1 - \frac{n\xi_i}{\xi_1 + \dots + \xi_n}, \quad i = 1, \dots, n,$$

where ξ_1, ξ_2, \dots are i.i.d. exponential random variables with mean 1. This choice yields a weighted approximation to the uniform quantile process β_n as defined in (8.1).

3. Let $c_n(1), \dots, c_n(n)$, $n \geq 1$, be a triangular array of constants satisfying

$$\sum_{i=1}^n c_n(i) = 0, \quad \sum_{i=1}^n c_n^2(i)/n = 1 \quad \text{and} \quad \sum_{i=1}^n c_n^4(i)/n = O(1).$$

Consider the finite sampling processes

$$\Pi_n(t) = \sum_{i \leq tn} c_n(A_i), \quad 0 \leq t \leq 1,$$

where (A_1, \dots, A_n) is a random permutation of $(1, \dots, n)$ taken with probability $1/n!$. We get that for $0 \leq \nu < \frac{1}{4}$ and $d > 0$

$$\sup_{d/n \leq t \leq 1-d/n} \frac{|\tilde{\Pi}_n(t) - \tilde{B}_n(t)|}{(t(1-t))^{1/2-\nu}} = O_p(n^{-\nu}).$$

The same result holds for the so-called weighted empirical process of Koul (1970):

$$\alpha_{c,n}(t) = \sum_{i=1}^n c_n(i) 1\{U_i \leq t\}, \quad 0 \leq t \leq 1.$$

Shorack (1991) first proved these results by means of the Skorohod embedding for martingales. This was also the basic tool that U. Einmahl and Mason (1992) used to obtain their general approximation to exchangeable processes. Replacing the $c_n(i)$, by random exchangeable weights $W_{i,n} - 1/n$, one readily derives weighted approximations to the weighted bootstrap empirical process of Mason and Newton (1992). See the discussion of the weighted approximation to the general weighted bootstrapped empirical process below.

10.2. Nonparametric Bootstrapped Empirical Process

From results in S. Csörgő and Mason (1989) one can derive the following weighted approximation to the nonparametric bootstrapped empirical process:

On the same probability space there exist a sequence of i.i.d. F random variables X_1, X_2, \dots , a triangular array

$$\{(M_{1,n}, \dots, M_{n,n}) : n \geq 1\}$$

of Multinomial $(n; \frac{1}{n}, \dots, \frac{1}{n})$ random vectors and a sequence of Brownian bridges $\{B_n\}_{n \geq 1}$, where the $(M_{1,n}, \dots, M_{n,n})$, $n \geq 1$, and B_1, B_2, \dots , are independent of X_1, X_2, \dots , such that for all $0 \leq \nu < 1/4$ and $\tau > 0$

$$\sup_{\tau/n \leq F(x) \leq 1-\tau/n} \frac{|\alpha_{M,n}(x) - B_n(F(x))|}{(F(x)(1-F(x)))^{1/2-\nu}} = O_p(n^{-\nu}),$$

where

$$\alpha_{M,n}(x) = \sqrt{n} \{F_{M,n}(x) - F_n(x)\},$$

$$F_n(x) = n^{-1} \sum_{i=1}^n 1\{X_i \leq x\}, \quad -\infty < x < \infty,$$

and

$$F_{M,n}(x) = n^{-1} \sum_{i=1}^n M_{i,n} 1\{X_i \leq x\}, \quad -\infty < x < \infty.$$

Such a weighted approximation to the bootstrapped empirical process has proved useful in establishing the weak consistency of nonparametric bootstrapped functions of the empirical process. See S. Csörgő and Mason (1989) for details and many examples.

10.3. Weighted Bootstrapped Empirical Process

The results of U. Einmahl and Mason (1992) described above yield the following weighted approximation to the general weighted bootstrapped empirical process introduced by Mason and Newton (1992). It includes as a special case the S. Csörgő and Mason (1989) result just cited:

Assume that $\{(W_{1,n}, \dots, W_{n,n}) : n \geq 1\}$ is a triangular array of exchangeable random variables satisfying

$$\sum_{i=1}^n W_{i,n} = 1, W_{i,n} \geq 0, E(nW_{1,n} - 1)^4 = O(1),$$

$$\frac{1}{n} \sum_{i=1}^n (nW_{i,n} - 1)^2 = \sigma^2 + O_p(n^{-1/2}), \text{ for some } \sigma^2 > 0,$$

and

$$\lim_{\varepsilon \searrow 0} \liminf_{n \rightarrow \infty} P\{nW_{1,n} > \varepsilon\} = 1.$$

Then on the same probability space there exist a sequence of i.i.d. F random variables X_1, X_2, \dots , a triangular array

$$\{(W_{1,n}, \dots, W_{n,n}) : n \geq 1\}$$

as above and a sequence of Brownian bridges $\{B_n\}_{n \geq 1}$, where the

$$(W_{1,n}, \dots, W_{n,n}), n \geq 1,$$

and B_1, B_2, \dots , are independent of X_1, X_2, \dots , such that for all $0 \leq \nu < 1/4$ and $\tau > 0$

$$\sup_{\tau/n \leq F(x) \leq 1-\tau/n} \frac{|\alpha_{W,n}(x) - \sigma B_n(F(x))|}{(F(x)(1-F(x)))^{1/2-\nu}} = O_p(n^{-\nu}),$$

where

$$\alpha_{W,n}(x) = \sqrt{n} \{F_{W,n}(x) - F_n(x)\}, \quad -\infty < x < \infty,$$

with

$$F_n(x) = n^{-1} \sum_{i=1}^n 1\{X_i \leq x\}, \quad -\infty < x < \infty,$$

and

$$F_{W,n}(x) = \sum_{i=1}^n W_{i,n} 1\{X_i \leq x\}, \quad -\infty < x < \infty.$$

10.4. Approximation to Continuous Time Martingales

We shall describe some research of Haeusler and Mason (1999), which in many ways generalizes the results in the previous subsections. Erich Haeusler and I accomplished this work during a very busy period crammed with trips to Civil War battlefields and hikes in the Pennsylvania mountains, while he was visiting me at the University of Delaware in the winter of 1996. Our paper would have appeared earlier if the editor had not misplaced it on his desk for a year.

Some Technicalities

Fix any $0 < \bar{t} \leq \infty$. Let

$$\{M_n\}_{n \geq 1} = \{(M_n(t))_{0 \leq t < \bar{t}}\}_{n \geq 1}$$

be a sequence of mean zero martingales with respect to filtrations $\mathcal{F}_n = (\mathcal{F}_n(t))_{0 \leq t < \bar{t}}$, and satisfying $M_n(0) = 0$.

Assume $EM_n^2(t) < \infty$ for all $n \geq 1$ and $0 \leq t < \bar{t}$. Also assume among other conditions, which are too technical to state here, that the predictable quadratic variation $\langle M_n \rangle$ of M_n converges in a certain way (see Haeusler and Mason (1999) for details) to a function

$$D : [0, \bar{t}] \rightarrow [0, \infty)$$

which is continuous, non-decreasing and satisfies

$$D(0) = 0, \quad \lim_{t \uparrow \bar{t}} D(t) = \infty.$$

Under the above assumptions, Haeusler and Mason (1999) obtained the following:

Theorem 6. *On a rich enough probability space there exists a sequence of versions*

$$(\tilde{M}_n)_{n \geq 1} \text{ of } (M_n)_{n \geq 1} \text{ i.e. } \tilde{M}_n =_d M_n \text{ for each } n,$$

and a standard Wiener process W such that for all $0 < \nu < \beta$,

$$\sup_{t: \frac{1}{n-1} \leq D(t) \leq n-1} \frac{|\tilde{M}_n(t) - W(D(t))|}{D(t)^{1/2-\nu}(1+D(t))^{2\nu}} = O_p(n^{-\nu}).$$

The constant $\beta > 0$ depends on a number of technical assumptions.

This result yields the U. Einmahl and Mason (1992) theorem as a special case. In a related paper, Haeusler, Mason and Turova (2000) used these ideas to construct a weighted approximation to a serial rank process.

The Empirical Process seen as a Martingale

Set for $n \geq 1$,

$$M_n(t) = \frac{\alpha_n(t)}{1-t} = \frac{\sqrt{n}(G_n(t) - t)}{1-t}, \quad 0 \leq t < 1.$$

As we have already pointed out,

$$M_n = (M_n(t))_{0 \leq t < 1}$$

is a sequence of mean zero martingales with respect to the filtrations

$$\mathcal{F}_n = (\mathcal{F}_n(t))_{0 \leq t < 1},$$

where for each $0 \leq t < 1$,

$$\mathcal{F}_n(t) = \sigma(G_n(s), 0 \leq s \leq t).$$

In this case it turns out that

$$\langle M_n \rangle(t) = \frac{1}{n} \sum_{i=1}^n D_i(t), \quad 0 \leq t < 1,$$

where for each $i \geq 1$ and $0 \leq t < 1$,

$$D_i(t) := \int_0^t \frac{1\{U_i \geq s\}}{(1-s)^3} ds \text{ and } D(t) = \frac{t}{1-t}, \text{ for } 0 \leq t < 1.$$

Applying Theorem 6 to this setup eventually yields a version of the weighted approximation (5.1) to the uniform empirical process as stated

in Theorem 1. In fact, Haeusler and Mason (1999) obtain this approximation via a special case of a weighted approximation to the ‘randomly’ *weighted empirical process*

$$(10.4) \quad X_n(t) := \sum_{i=1}^n w_{i,n}(1\{U_i \leq t\} - t), \quad 0 \leq t \leq 1,$$

where the weights $w_{1,n}, 1 \leq i \leq n$, are independent of U_1, U_2, \dots . A special case of their general Theorem 6 yields the following weighted approximation for (10.4).

Theorem 7. *Assume that the weights $w_{i,n}$, $1 \leq i \leq n$, satisfy the following two conditions:*

$$\sum_{i=1}^n Ew_{i,n}^4 = O(n^{-1}) \quad \text{and} \quad \sum_{i=1}^n w_{i,n}^2 - 1 = O_P(n^{-1/2}).$$

Then on a rich enough probability space there exists a sequence of probabilistically equivalent versions $(\tilde{X}_n)_{n \geq 1}$ of $(X_n)_{n \geq 1}$ (i.e. $\tilde{X}_n \stackrel{d}{=} X_n$ for every n) and a standard Brownian bridge B such that for all $0 \leq \nu < 1/4$

$$(10.5) \quad \sup_{1/n \leq t \leq 1-1/n} \frac{|\tilde{X}_n(t) - B(t)|}{(t(1-t))^{1/2-\nu}} = O_P(n^{-\nu}),$$

and moreover (10.5) remains true when the supremum is taken over the entire interval $(0, 1)$ in the case $0 < \nu < 1/4$.

Shorack (1991) and U. Einmahl and Mason (1992) established special cases of this result under the additional but unnecessary assumption that $\sum_{i=1}^n w_{i,n} = 0$. Clearly Theorem 7 also gives as a special case a version of the approximation result (5.1) stated in Theorem 1 by choosing $w_{i,n} = 1/\sqrt{n}$ for $i = 1, \dots, n$.

10.4.1. Application to Eicker-Jaeschke and Rényi-type statistics. Weighted approximations are especially useful to establish asymptotic distributional results when usual weak convergence techniques fail or are extremely cumbersome to apply. For example, consider the following statistics. Let X_1, X_2, \dots , be a sequence of independent random variables with a common continuous cdf F . For each $n \geq 1$ define as before the empirical distribution function based upon X_1, \dots, X_n to be

$$F_n(x) := \frac{1}{n} \sum_{i=1}^n 1\{X_i \leq x\}, \quad -\infty < x < \infty.$$

Introduce the Eicker-Jaeschke statistic

$$E_n = \sup_{-\infty < x < \infty} \frac{\sqrt{n}|F_n(x) - F(x)|}{\sqrt{F(x)(1 - F(x))}},$$

and the Rényi statistic

$$R_n(r) = \sup_{\{x: r \leq F(x) < \infty\}} \frac{\sqrt{n}|F_n(x) - F(x)|}{F(x)},$$

where $0 < r < 1$. These two statistics, along with their versions formed by replacing in the denominator F by F_n , have been shown by Révész (1982), Mason and Schuenemeyer (1983) and Calitz (1987) to be more sensitive to various types of heavy and light tail alternatives than the usual Kolmogorov-Smirnov statistic.

Eicker (1979) and Jaeschke (1979) proved, among other results, that for all $-\infty < x < \infty$

$$(10.6) \quad P\{a_n E_n - b_n \leq x\} \rightarrow E^2(x), \text{ as } n \rightarrow \infty,$$

where for $n \geq 3$

$$a_n = \sqrt{2 \log \log n}, \quad b_n = 2 \log \log n + 2^{-1} \log \log \log n - 2^{-1} \log \pi$$

and $E(x) = \exp(-\exp(-x))$. Eicker proved (10.6) using methods from classical probability, whereas Jaeschke based his proof upon the powerful KMT (1975) Brownian bridge approximation to the empirical process. Later, Cs-Cs-H-M (1986) provided a proof of (10.6) based upon their weighted approximation to the uniform empirical process given in Theorem 1, which is a special case of Theorem 7.

Cs-Cs-H-M (1986) also used Theorem 1 to prove the following result for the Rényi statistic: Let r_n be any sequence of positive constants such that for some $0 < \beta < 1$, we have $0 < r_n \leq \beta$ for all large enough n , and $nr_n \rightarrow \infty$. Then for all $x \geq 0$, as $n \rightarrow \infty$,

$$(10.7) \quad P \left\{ \sqrt{\frac{r_n}{1 - r_n}} R_n(r_n) \leq x \right\} \rightarrow P \left\{ \sup_{0 \leq t \leq 1} |W(t)| \leq x \right\}.$$

The special case when $r_n = \beta$ was first proved by Rényi (1953) and the case when $r_n \rightarrow 0$ and $nr_n \rightarrow \infty$ by Csáki (1974).

The Cs-Cs-H-M (1986) result (10.7) can be obtained, though much less efficiently, using weak convergence results for the local empirical process established by U. Einmahl and Mason (1997). So far no approach to the Eicker-Jaeschke limiting distribution theorem (10.6) based entirely upon weak convergence results for empirical processes seems to be known.

Consider now the following generalized versions of the statistics E_n and $R_n(r)$, respectively, given by

$$E_{n,w} = \sup_{-\infty < x < \infty} \frac{|\sum_{i=1}^n w_{i,n}(1\{X_i \leq x\} - F(x))|}{\sqrt{F(x)(1 - F(x))}}$$

and

$$R_{n,w}(r) = \sup_{\{x: r \leq F(x) < \infty\}} \frac{|\sum_{i=1}^n w_{i,n}(1\{X_i \leq x\} - F(x))|}{F(x)}.$$

Assuming that the weights $w_{i,n}$, $1 \leq i \leq n$, satisfy the conditions of Theorem 7, and noting that $(F(X_1), \dots, F(X_n))$ is equal in distribution to (U_1, \dots, U_n) , when F is continuous, we can apply the weighted approximation by Theorem 7, in exactly the same way as Theorem 1 was used in the proofs of Theorems 4.4.1 and 4.5.1 of Cs-Cs-H-M (1986) to show that both (10.6) and (10.7) remain true when E_n and R_n are replaced by $E_{n,w}$ and $R_{n,w}(r_n)$, respectively. For the proof for the Rényi statistic based on Theorem 1 see Example 2 in Chapter 5.

Closely related to the statistics $E_{n,w}$ and $R_{n,w}(r)$ are the Kolmogorov-Smirnov and Rényi-type statistics proposed by Hájek and Šidák (1967) for regression-type alternatives which are functions of the random sums

$$\sum_{i=1}^k c_{D_i,n}, \quad 1 \leq k \leq n,$$

where $(c_{1,n}, \dots, c_{n,n})$ is a vector of *regression constants* and (D_1, \dots, D_n) is an anti-rank vector uniformly distributed over the permutations of $(1, \dots, n)$. Consider the following Eicker-Jaesche version of the Hájek and Šidák (1967) Kolmogorov-Smirnov statistic

$$\widehat{E}_{n,C} = \max_{1 \leq k \leq n} \frac{|\sum_{i=1}^k c_{D_i,n} - k\bar{c}_n|}{\sqrt{\sum_{i=1}^n (c_{i,n} - \bar{c}_n)^2} \sqrt{\frac{k}{n} \left(\frac{n+1-k}{n}\right)}},$$

where $\bar{c}_n = \sum_{i=1}^n c_{i,n}/n$, and, further, the Hájek and Šidák (1967) Rényi-type statistic defined with $0 < r < 1$ to be

$$\widehat{R}_{n,C}(r) = \max_{rn \leq k \leq n} \frac{n|\sum_{i=1}^k c_{D_i,n} - k\bar{c}_n|}{k\sqrt{\sum_{i=1}^n (c_{i,n} - \bar{c}_n)^2}}.$$

By applying the weighted approximation result of Shorack (1991) for the finite sampling process or that of U. Einmahl and Mason (1992) for the exchangeable process (10.1) to the *anti-rank process*

$$A_n(t) := \frac{\sum_{i \leq tn} c_{D_i,n} - [tn]\bar{c}_n}{\sqrt{\sum_{i=1}^n (c_{i,n} - \bar{c}_n)^2}}, \quad 0 \leq t \leq 1,$$

(both of which follow from our Theorem 6), we obtain under the assumptions on the weights

$$w_{i,n} := \frac{c_{i,n} - \bar{c}_n}{\sqrt{\sum_{i=1}^n (c_{i,n} - \bar{c}_n)^2}}, \quad 1 \leq i \leq n,$$

given in Theorem 7 that both (10.6) and (10.7) hold, as before, with E_n and R_n replaced by $\widehat{E}_{n,C}$ and $\widehat{R}_{n,C}(r_n)$, respectively. (Hájek and Šidák (1967) obtained (10.7) for $\widehat{R}_{n,C}(r_n)$ in the case when $r_n = r$, using weak convergence. Their method fails in the case when $r_n \rightarrow 0$ and $nr_n \rightarrow \infty$.)

It is difficult to conceive how one would prove all of the limiting distributional results presented in this subsection for the general Eicker-Jaeschke and the Rényi-type statistics introduced here without these weighted approximations. This should convince the reader of the power of the weighted approximation methodology. (This subsection was taken almost verbatim from Haeusler and Mason (1999).)

10.5. Some Final Remarks About Probability Spaces

With respect to weighted approximations to the uniform empirical and quantile processes, there are at least four probability spaces on which they hold for suitable values of ν . First of all on any probability space on which sits a sequence of i.i.d. Uniform $(0, 1)$ random variables U_1, U_2, \dots , it was shown in M. Csörgő, S. Csörgő, Horváth and Mason (1986) and Mason (1991) that one always has for any $0 \leq \nu < 1/4$,

$$(10.8) \quad \sup_{1/(n+1) \leq s \leq n/(n+1)} \frac{n^\nu |\alpha_n(t) - \beta_n(t)|}{(t(1-t))^{1/2-\nu}} = O_P(1).$$

However there are at least four methods to enlarge the space to include a sequence of Brownian bridges $\{B_n\}_{n \geq 1}$ such that for suitable $\nu_1 \geq 0$,

$$(10.9) \quad \sup_{0 \leq t \leq 1} \frac{n^{\nu_1} |\alpha_n(t) - \bar{B}_n(t)|}{(t(1-t))^{1/2-\nu_1}} = O_P(1),$$

and for suitable $\nu_2 \geq 0$, where \bar{B}_n is defined as in (5.2),

$$(10.10) \quad \sup_{1/(n+1) \leq s \leq n/(n+1)} \frac{n^{\nu_2} |\beta_n(t) - B_n(t)|}{(t(1-t))^{1/2-\nu_2}} = O_P(1).$$

Method 1. M. Csörgő, S. Csörgő, Horváth and Mason (1986) used the KMT (1975, 1976) strong approximation to the partial sum process to construct a probability space so that (10.10) is valid for all $0 \leq \nu_2 < 1/2$ and then inferred that (10.9) holds on this space for all $0 \leq \nu_1 < 1/4$ via (10.8). In the process they proved that on their probability space

the analogs to the inequalities in Theorem 2 held with α_n replaced by β_n .

Method 2. Mason and van Zwet (1987) showed that on the probability space on which the KMT (1975) Brownian bridge approximation to uniform empirical process (4.3) holds that (10.9) is valid for all $0 \leq \nu_1 < 1/2$ and then inferred that (10.10) holds on this space for all $0 \leq \nu_2 < 1/4$ via (10.8).

Method 3. M. Csörgő and Horváth (1986) and Mason (1986, 1991) used the Skorohod embedding to the partial sum process to construct a probability space so that (10.10) is valid for all $0 \leq \nu_2 < 1/4$ and then inferred that (10.9) holds on this space for all $0 \leq \nu_1 < 1/4$ via (10.8).

Method 4. U. Einmahl and Mason (1992) and Haeusler and Mason (1999) constructed a probability space using the Skorohod embedding for martingales so that (10.9) is valid for all $0 \leq \nu_1 < 1/4$ and then inferred that (10.10) holds on this space for all $0 \leq \nu_2 < 1/4$ via (10.8).

Comunicaciones del CIMAT



CIMAT

APPENDIX A

Basic Large Sample Theory Facts

A good reference for the material in this appendix, along with much more, is Chapter 1 of Serfling (1980).

A σ -field of events Let Ω be a non-empty sample space of outcomes. A set \mathcal{A} of subsets of Ω is called a σ -field if

- (a) $\Omega \in \mathcal{A}$.
- (b) If $A \in \mathcal{A}$ then $A^C \in \mathcal{A}$.
- (c) If $A_1, A_2, \dots, \in \mathcal{A}$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$.

The pair (Ω, \mathcal{A}) is called a *measurable space*.

Probability Given a sample space Ω with a σ -field of events \mathcal{A} , a probability P is a function defined on \mathcal{A} , which satisfies the following properties:

- (a) $P(A) \geq 0$ for all $A \in \mathcal{A}$.
- (b) $P(\Omega) = 1$.
- (c) If $A_1, A_2, \dots, \in \mathcal{A}$ are pairwise disjoint then

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

The triple (Ω, \mathcal{A}, P) is called a *probability space*.

A real valued random variable A real valued random variable X on a probability space (Ω, \mathcal{A}, P) is a measurable function from Ω to \mathbb{R} , meaning that for each $\omega \in \Omega$, $X(\omega) \in \mathbb{R}$ and for each $x \in \mathbb{R}$, $\{\omega : X(\omega) \leq x\} \in \mathcal{A}$.

Convergence in probability A sequence of random variables X_n converges in probability to a random variable X if for all $\varepsilon > 0$

$$P(|X_n - X| > \varepsilon) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

This is usually written as

$$X_n \rightarrow_P X.$$

Note that X is allowed to be a constant.

Convergence with probability 1 (w.p. 1) or almost surely (a.s.) A sequence of random variables X_n converges w.p.1 or a.s. to a random variable X if

$$P\{X_n \rightarrow X, \text{ as } n \rightarrow \infty\} = 1.$$

This is usually written as

$$X_n \rightarrow X, \text{ w.p.1, as } n \rightarrow \infty, \text{ or } X_n \rightarrow X, \text{ a.s., as } n \rightarrow \infty.$$

Note that convergence in probability does not imply convergence almost surely, but the converse is obviously true.

Convergence in distribution A sequence of random variables X_n converges in distribution to a random variable X with cdf F if for every continuity point x of F , as $n \rightarrow \infty$,

$$P(X_n \leq x) \rightarrow P(X \leq x) =: F(x).$$

This is often written

$$X_n \rightarrow_d X.$$

As further notation we write

$$X =_d Y$$

to mean that X and Y have the same cdf.

Empirical distribution function The *empirical distribution function* based on a random sample X_1, \dots, X_n is the cumulative function F_n defined for each x to be

$$(A.1) \quad F_n(x) = \frac{\#\{X_i \leq x\}}{n} = n^{-1} \sum_{i=1}^n 1\{X_i \leq x\}.$$

To construct F_n as in (A.1), it is convenient to know the order statistics based on X_1, \dots, X_n .

Order statistics Let X_1, \dots, X_n be independent random variables with cdf F . Their order statistics are defined to be the ordered values $X_{1,n} \leq \dots \leq X_{n,n}$. $X_{k,n}$ is called the *k*th order statistic based upon X_1, \dots, X_n .

The weak law of large numbers

For each $n \geq 1$, let X_1, \dots, X_n be independent random variables with common cdf F . If $E|X_1| < \infty$, then

$$(A.2) \quad \int_{-\infty}^{\infty} x dF_n(x) =: \bar{X} \rightarrow_P \mu = \int_{-\infty}^{\infty} x dF(x).$$

The strong law of large numbers

If X_1, X_2, \dots , is an infinite sequence of i.i.d. F random variables sitting on the same probability space then, with probability 1, $\bar{X} \rightarrow \mu$.

The Central limit theorem

Let X_1, \dots, X_n be independent random variables with common cdf F . If $0 < \sigma^2 = \text{Var}X_1 < \infty$, then

$$(A.3) \quad \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{d} Z,$$

where Z is a standard normal random variable.

Fact (A) $X_n \rightarrow_P X$ implies $X_n \rightarrow_d X$.

Proof. Choose any continuity point x of F , the cdf of X , and any $\varepsilon > 0$.

$$\begin{aligned} P(X \leq x - \varepsilon, |X_n - X| \leq \varepsilon) &\leq P(X_n \leq x) \\ &\leq P(X \leq x + \varepsilon, |X_n - X| \leq \varepsilon) + P(|X_n - X| > \varepsilon). \end{aligned}$$

Hence

$$\begin{aligned} P(X \leq x - \varepsilon) - P(|X_n - X| > \varepsilon) &\leq P(X_n \leq x) \\ &\leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon). \end{aligned}$$

Since $P(|X_n - X| > \varepsilon) \rightarrow 0$ as $n \rightarrow \infty$, we have for all n sufficiently large

$$F(x - \varepsilon) - \varepsilon \leq P(X_n \leq x) \leq F(x + \varepsilon) + \varepsilon.$$

Since F is continuous at x and ε can be made arbitrarily small

$$P(X_n \leq x) \rightarrow F(x).$$

□

Fact (B) (Baby Skorohod representation theorem) It is not in general true that $X_n \rightarrow_d X$ implies $X_n \rightarrow_P X$. However, the following is true: if $X_n \rightarrow_d X$ then there exists a sequence of random variables $\{Y_n\}_{n \geq 1}$ such that $Y_n \stackrel{d}{=} X_n$ for all $n \geq 1$ and

$$Y_n \rightarrow_P X.$$

We shall only prove this theorem in the case when $X = c$, where c is a constant. In this case we can choose $Y_n = X_n$, since $X_n \rightarrow_P c$.

Proof. $X_n \rightarrow_d c$ means $P(X_n \leq c + \varepsilon) \rightarrow 1$ for all $\varepsilon > 0$ and $P(X_n \leq c - \varepsilon) \rightarrow 0$ for all $\varepsilon > 0$. This implies that both $P(X_n - c < -\varepsilon) \rightarrow 0$ and $P(X_n - c > \varepsilon) \rightarrow 0$ as $n \rightarrow \infty$, which is equivalent to $P(|X_n - c| > \varepsilon) \rightarrow 0$ as $n \rightarrow \infty$. (What we have actually shown here is that $X_n \rightarrow_d c$ if and only if $X_n \rightarrow_P c$. The only if part comes from Fact (A).) □

For a complete proof of Fact (B) refer to Theorem 1.6.3 of Serfling (1980). There it is shown that one can choose $X = Q(U)$ and $Y_n = Q_n(U)$, $n \geq 1$, where Q is the quantile function of X and Q_n is the quantile function of X_n , $n \geq 1$, and, in fact, $Y_n \rightarrow X$, a.s.

More generally, Skorohod (1956) proved that if $\{P_n\}_{n \geq 1}$ is a sequence of probability measures on a complete separable metric space S converging weakly to a probability measure P on S , then there exist S -valued random variables X_n each with distribution P_n such that X_n converges almost surely to X , where X has distribution P . (For more about the Skorohod representation theorem consult Billingsley (1999).)

Fact (C) If g is continuous on the image of X and $X_n \rightarrow_P X$, then $g(X_n) \rightarrow_P g(X)$.

Proof. We shall only prove this in the case when $X = c$, where c is a constant. Choose any $\varepsilon > 0$ and $\delta > 0$ such that $|g(x) - g(c)| \leq \varepsilon$ whenever $|x - c| \leq \delta$. We see that

$$P(|g(X_n) - g(c)| > \varepsilon) \leq P(|X_n - c| > \delta) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Thus $g(X_n) \rightarrow_P g(c)$ if $X_n \rightarrow_P X$. \square

Combining Facts (B) and (C), we get

Fact (D)(Mann-Wald theorem) If g is continuous then $X_n \rightarrow_d X$ implies $g(X_n) \rightarrow_d g(X)$.

Fact (E) If $X_n \rightarrow_P X$ and $Y_n \rightarrow_P Y$ then $X_n + Y_n \rightarrow_P X + Y$.

Proof. By the triangle inequality

$$|X_n + Y_n - X - Y| \leq |X_n - X| + |Y_n - Y|.$$

Thus, as $n \rightarrow \infty$,

$$\begin{aligned} & P(|X_n + Y_n - (X + Y)| > \varepsilon) \\ & \leq P\left(|X_n - X| > \frac{\varepsilon}{2}\right) + P\left(|Y_n - Y| > \frac{\varepsilon}{2}\right) \rightarrow 0. \end{aligned}$$

\square

Big O_p and little o_p notation Let Y_n be a sequence of random variables and γ_n be a sequence of positive constants. The notation

$$Y_n = O_p(\gamma_n),$$

means that the sequence Y_n/γ_n is bounded in probability, that is, for all $\varepsilon > 0$ there is an M such that for all $n \geq 1$,

$$P(|Y_n| > M\gamma_n) < \varepsilon.$$

Also we shall use the notation

$$Y_n = o_p(\gamma_n),$$

to denote that $Y_n/\gamma_n \rightarrow_P 0$. In particular, $Y_n = O_p(1)$ means that the sequence of random variables is bounded in probability and $Y_n = o_p(1)$ means that $Y_n \rightarrow_P 0$.

Fact (F) If $X_n \rightarrow_P 0$ and $Y_n = O_p(1)$, then $X_n Y_n \rightarrow_P 0$.

Proof. Notice that

$$P(|X_n Y_n| > \varepsilon) \leq P(|X_n| > \varepsilon^2) + P(|Y_n| > \varepsilon^{-1})$$

and both of the terms on the right hand side of this inequality can be made as small as desired by choosing $\varepsilon > 0$ small enough and by letting $n \rightarrow \infty$. \square

Fact (G) If $A_n \rightarrow_P 0$ and $X_n \rightarrow_P X$ then $A_n X_n \rightarrow_P 0$.

This is a special case of (F).

One can also show that

Fact (G') If $A_n \rightarrow_P 0$ and $X_n \rightarrow_d X$ then $A_n X_n \rightarrow_P 0$.

Fact (H) If $X_n \rightarrow_P X$ and $Y_n \rightarrow_P Y$ then $X_n Y_n \rightarrow_P XY$.

Proof. Observe that

$$XY - X_n Y_n = (X - X_n)Y - (X_n - X)(Y_n - Y) + X(Y - Y_n).$$

Now by (F)

$$(X_n - X)Y \rightarrow_P 0, (X_n - X)(Y_n - Y) \rightarrow_P 0 \text{ and } X(Y - Y_n) \rightarrow_P 0.$$

Hence by (E)

$$XY - X_n Y_n \rightarrow_P 0.$$

\square

Fact (I) If $X_n \rightarrow_d X$ and $X_n - Y_n \rightarrow_P 0$ then $Y_n \rightarrow_d X$.

Proof. Choose any continuity point x of F and $\varepsilon > 0$ such that $x + \varepsilon$ is a continuity point of F .

$$P(Y_n \leq x) \leq P(X_n \leq x + \varepsilon) + P(|X_n - Y_n| > \varepsilon).$$

Hence

$$\limsup_{n \rightarrow \infty} P(Y_n \leq x) \leq F(x + \varepsilon).$$

Similarly choose $\delta > 0$ such that $x - \delta$ is a continuity point of F . We get in a similar manner

$$\liminf_{n \rightarrow \infty} P(Y_n \leq x) \geq F(x - \delta).$$

Since both $\delta > 0$ and $\varepsilon > 0$ can be made arbitrarily small

$$\lim_{n \rightarrow \infty} P(Y_n \leq x) = F(x).$$

□

Fact (J) (Slutsky's theorem) If $X_n \rightarrow_d X$ and for constants A and B ,

$$A_n \rightarrow_P A \text{ and } B_n \rightarrow_P B,$$

then

$$A_n X_n + B_n \rightarrow_d AX + B.$$

Proof. Notice that by Fact (D),

$$AX_n + B \rightarrow_d AX + B.$$

Next write

$$A_n X_n + B_n - AX_n - B = (A_n - A)X_n + B_n - B.$$

Applying (G), we see that $(A_n - A)X_n \rightarrow_P 0$. This combined with $B_n - B \rightarrow_P 0$, implies by (E) that

$$A_n X_n + B_n - (AX_n + B) \rightarrow_P 0,$$

Hence by (I)

$$A_n X_n + B_n \rightarrow_d AX + B.$$

□

Fact (K) Let a_n be a sequence of constants converging to infinity such that

$$a_n(X_n - b) \rightarrow_d W.$$

If g is any function which is differentiable at b with derivative $g'(b)$, then

$$a_n(g(X_n) - g(b)) \rightarrow_d g'(b)W.$$

Proof. Observe that since $a_n(X_n - b) \rightarrow_d W$ and $a_n \rightarrow \infty$, we get that $X_n \rightarrow_P b$. Let h be a function defined as follows:

$$h(x) = \begin{cases} \frac{g(x) - g(b)}{x - b}, & x \neq b \\ g'(b), & \text{when } x = b. \end{cases}$$

Since g is differentiable at b , h is continuous at b . Hence

$$h(X_n) \rightarrow_P g'(b)$$

by (C). Now

$$a_n(g(X_n) - g(b)) = h(X_n)a_n(X_n - b),$$

which by Slutsky's Theorem converges in distribution to $g'(b)W$. □

A shorthand way to write a number of the above facts is the following:

Let

$$X_n = O_P(1), Y_n = O_P(1), W_n = o_P(1), Z_n = o_P(1),$$

Then

$$X_n + Y_n = O_P(1), X_n Y_n = O_P(1), W_n + Z_n = o_P(1),$$

$$W_n Z_n = o_P(1) \text{ and } X_n W_n = o_P(1).$$

Fact (L) (Borel–Cantelli lemma) Let $\{A_n\}_{n \geq 1}$ be a sequence of events on a probability space (Ω, \mathcal{A}, P) . The event

$$\{A_n, \text{ i.o.}\} = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m,$$

where “i.o.” means “infinitely often”. Notice that this is the set of $\omega \in \Omega$ such that $\omega \in A_n$ for infinitely many A_n in the sequence $\{A_n\}_{n \geq 1}$.

The Borel–Cantelli lemma says that if

$$\sum_{n=1}^{\infty} P(A_n) < \infty,$$

then $P\{A_n, \text{ i.o.}\} = 0$, and if $\{A_n\}_{n \geq 1}$ are independent, then

$$\sum_{n=1}^{\infty} P(A_n) = \infty,$$

implies $P\{A_n, \text{ i.o.}\} = 1$.

Moment inequalities

A good source for moment inequalities is the set of notes: *Properties of moments of random variables*, by Jean-Marie Dufour.

http://www2.cirano.qc.ca/~dufourj/Web_Site/ResE/Dufour_1999.C_TS_Moments.pdf

Comunicaciones del CIMAT



CIMAT

APPENDIX B

Properties of Order Statistics

Let X_1, \dots, X_n be independent random variables with cdf F . Their order statistics are defined to be the ordered values $X_{1,n} \leq \dots \leq X_{n,n}$. Note the special cases

$$X_{1,n} = \min_{1 \leq i \leq n} X_i \text{ and } X_{n,n} = \max_{1 \leq i \leq n} X_i.$$

To derive the cdf of an order statistic $X_{i,n}$, $1 \leq i \leq n$, we will need the following equation that relates the Beta distribution to the Binomial distribution.

Relation between the Beta and the Binomial distribution

Theorem 1B For all $1 \leq k \leq n$ and $0 \leq p \leq 1$,

$$(B.1) \quad \int_0^p n \binom{n-1}{k-1} u^{k-1} (1-u)^{n-k} du = \sum_{m=k}^n \binom{n}{m} p^m (1-p)^{n-m},$$

or in other words,

$$(B.2) \quad P \{ \text{Beta}(k, n-k+1) \leq p \} = P \{ \text{Bin}(n, p) \geq k \}.$$

Proof. Set

$$P_m = \int_0^p n \binom{n-1}{m-1} u^{m-1} (1-u)^{n-m} du.$$

Notice that

$$P_n = \int_0^p n \binom{n-1}{n-1} u^{n-1} (1-u)^{n-n} du = p^n = \binom{n}{n} p^n.$$

Next by integration by parts for any $m < n$

$$\begin{aligned} P_m &= n \binom{n-1}{m-1} \frac{p^m}{m} (1-p)^{n-m} \\ &\quad + \int_0^p n \binom{n-1}{m-1} \left(\frac{n-m}{m} \right) u^m (1-u)^{n-m-1} du \\ &= n \binom{n-1}{m-1} \frac{p^m}{m} (1-p)^{n-m} + \int_0^p n \binom{n-1}{m} u^m (1-u)^{n-m-1} du \end{aligned}$$

$$= \binom{n}{m} p^m (1-p)^{n-m} + P_{m+1}.$$

Thus

$$\sum_{m=k}^{n-1} (P_m - P_{m+1}) = P_k - P_n = \sum_{m=k}^{n-1} \binom{n}{m} p^m (1-p)^{n-m}.$$

This completes the proof. \square

Distribution of order statistics

Theorem 2B. Let X_1, \dots, X_n be independent random variables with common cdf F , then

(i) for all x
(B.3)

$$P\{X_{k,n} \leq x\} = \int_{-\infty}^x n \binom{n-1}{k-1} F(w)^{k-1} (1-F(w))^{n-k} dF(w);$$

(ii) for all $x < y$ and $1 \leq k < l \leq n$,

$$\begin{aligned} & P\{X_{k,n} \leq x, X_{l,n} \leq y\} \\ &= \int_{-\infty}^x \int_{-\infty}^y n(n-1) I\{w \leq z\} \binom{n-2}{k-1, l-k-1, n-l} \\ & \times F(w)^{k-1} (F(z) - F(w))^{l-k-1} (1-F(z))^{n-l} dF(w) dF(z) \end{aligned}$$

and (iii) for all $x \geq y$

$$(B.5) \quad P\{X_{k,n} \leq x, X_{l,n} \leq y\} = P\{X_{l,n} \leq y\}.$$

Proof of (i). Let

$$S_n(x) = \sum_{i=1}^n I\{X_i \leq x\}.$$

Clearly

$$\begin{aligned} P\{X_{k,n} \leq x\} &= P\{S_n(x) \geq k\} \\ &= \sum_{m=k}^n \binom{n}{m} F(x)^m (1-F(x))^{n-m}, \end{aligned}$$

which by (B.1) equals

$$\int_0^{F(x)} n \binom{n-1}{k-1} u^{k-1} (1-u)^{n-k} du.$$

This last integral by the change of variables $u = F(w)$ is equal to

$$\int_{-\infty}^x n \binom{n-1}{k-1} F(w)^{k-1} (1-F(w))^{n-k} dF(w).$$

Proof of (ii). Choose any $x < y$. We see that

$$\begin{aligned} P\{X_{k,n} \leq x, X_{l,n} \leq y\} &= P\{S_n(x) \geq k, S_n(y) \geq l\} \\ &= \sum_{k+j+i \geq l} P\{S_n(y) - S_n(x) = i, S_n(x) = k+j\} \\ &= \sum_{k+j+i \geq l} \binom{n}{i, k+j, n-k-i-j} F(x)^{k+j} (F(y) - F(x))^i \\ &\times (1-F(y))^{n-k-i-j} \end{aligned} \tag{B.6}$$

$$\begin{aligned} &= \int_0^{F(x)} \int_0^{F(y)} n(n-1) I\{u \leq v\} \binom{n-2}{k-1, l-k-1, n-l} \\ &\times u^{k-1} (v-u)^{l-k-1} (1-v)^{n-l} dudv, \end{aligned} \tag{B.7}$$

which by the change of variables $v = F(z)$ and $u = F(w)$, equals (B.1).

Part (iii) is obvious. \square

Theorem 3B. Let X_1, \dots, X_n be independent random variables with common probability density function [pdf f], then

(i) $X_{k,n}$ has pdf

$$f_{X_{(k)}}(x) = n \binom{n-1}{k-1} F(x)^{k-1} (1-F(x))^{n-k} f(x); \tag{B.8}$$

(ii) for $1 \leq k < l \leq n$, the pair $(X_{k,n}, X_{l,n})$ has joint pdf

$$\begin{aligned} f_{X_{k,n}, X_{l,n}}(x, y) &= n(n-1) I\{x \leq y\} \times \\ &\binom{n-2}{k-1, l-k-1, n-l} F(x)^{k-1} (F(y) - F(x))^{l-k-1} \\ &\times (1-F(y))^{n-l} f(x) f(y); \end{aligned} \tag{B.9}$$

(iii) $(X_{1,n}, \dots, X_{n,n})$ has joint pdf

$$f_{(X_{1,n}, \dots, X_{n,n})}(x_{(1)}, \dots, x_{(n)}) = \begin{cases} n! f(x_{(1)}) \dots f(x_{(n)}), & x_{(1)} \leq \dots \leq x_{(n)} \\ 0, & \text{elsewhere.} \end{cases}$$

Proof. We only have to prove part (iii). Choose any permutation i_1, \dots, i_n of $1, \dots, n$. We see that

$$P\{X_{1,n} \leq x_1, \dots, X_{n,n} \leq x_n \text{ and } X_{1,n} = X_{i_1}, \dots, X_{n,n} = X_{i_n}\}$$

$$\begin{aligned}
 &= \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} I \{w_{i_1} \leq \dots \leq w_{i_n}\} f(w_{i_1}) \dots f(w_{i_n}) dw_{i_1} \dots dw_{i_n} \\
 &= \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} I \{w_1 \leq \dots \leq w_n\} f(w_1) \dots f(w_n) dw_1 \dots dw_n.
 \end{aligned}$$

Now clearly

$$\begin{aligned}
 &P \{X_{1,n} \leq x_1, \dots, X_{n,n} \leq x_n\} \\
 &= \sum_{\text{permutations } i_1, \dots, i_n} P \{X_{1,n} \leq x_1, \dots, X_{n,n} \leq x_n \\
 &\hspace{15em} \text{and } X_{1,n} = X_{i_1}, \dots, X_{n,n} = X_{i_n}\} \\
 &= \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} n! I \{w_1 \leq \dots \leq w_n\} f(w_1) \dots f(w_n) dw_1 \dots dw_n.
 \end{aligned}$$

□

Application

Let $(U_{1,n}, \dots, U_{n,n})$ denote the order statistics of n independent uniform random variables U_1, \dots, U_n . Here we shall establish a useful representation for the joint distribution of the uniform order statistics. Let $\omega_1, \dots, \omega_{n+1}$, be independent exponential random variables with mean 1. For each $1 \leq k \leq n + 1$, let

$$S_k = \sum_{i=1}^k \omega_i.$$

Note that since both $U_{k,n}$ and S_k/S_{n+1} are $Beta(k, n - k + 1)$ random variables, we have

$$U_{k,n} =_d \frac{S_k}{S_{n+1}}.$$

More generally we have the following.

Theorem 4B For all $n \geq 1$

$$(U_{1,n}, \dots, U_{n,n}) =_d \left(\frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}} \right) =: V$$

and S_{n+1} is independent of V .

Proof. Consider the transformation

$$\begin{aligned}
 &(y_1, \dots, y_{n+1}) \xrightarrow{g} \\
 &\left(\frac{y_1}{y_1 + \dots + y_{n+1}}, \dots, \frac{y_1 + \dots + y_n}{y_1 + \dots + y_{n+1}}, y_1 + \dots + y_{n+1} \right) \\
 &= (v_1, \dots, v_n, v_{n+1}).
 \end{aligned}$$

This transformation has inverse

$$\begin{aligned} & g^{-1}(v_1, \dots, v_n, v_{n+1}) \\ &= (v_{n+1}v_1, v_{n+1}v_2 - v_{n+1}v_1, \dots, v_{n+1}v_n - v_{n+1}v_{n-1}, v_{n+1} - v_{n+1}v_n) \\ &= (h_1, \dots, h_{n+1}). \end{aligned}$$

Note that the joint *pdf* of $(\omega_1, \dots, \omega_{n+1})$ is

$$f_{E_1, \dots, E_{n+1}}(y_1, \dots, y_{n+1}) = e^{-y_1} \dots e^{-y_{n+1}} I \{y_1 > 0, \dots, y_{n+1} > 0\}.$$

We get then that the joint *pdf* of

$$\left(\frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}}, S_{n+1} \right) = (V_1, \dots, V_n, V_{n+1})$$

is

$$\begin{aligned} & f_{V_1, \dots, V_{n+1}}(v_1, \dots, v_{n+1}) \\ &= f_{\omega_1, \dots, \omega_{n+1}}(g^{-1}(v_1, \dots, v_n, v_{n+1})) |J(v_1, \dots, v_n, v_{n+1})|, \end{aligned}$$

where

$$J(v_1, \dots, v_n, v_{n+1}) = \det \begin{vmatrix} \frac{dh_1}{dv_1} & \dots & \frac{dh_1}{dv_{n+1}} \\ \frac{dh_{n+1}}{dv_1} & \dots & \frac{dh_{n+1}}{dv_{n+1}} \end{vmatrix}.$$

One finds that

$$\begin{aligned} f_{V_1, \dots, V_n}(v_1, \dots, v_n) &= \int_0^\infty f_{V_1, \dots, V_{n+1}}(v_1, \dots, v_{n+1}) dv_{n+1} \\ &= n! I \{0 \leq v_1 \leq \dots \leq v_n \leq 1\} = f_{U_{1,n}, \dots, U_{n,n}}(v_1, \dots, v_n). \end{aligned}$$

and

$$f_{V_1, \dots, V_{n+1}}(v_1, \dots, v_{n+1}) = f_{V_1, \dots, V_n}(v_1, \dots, v_n) f_{V_{n+1}}(v_{n+1}).$$

□

Comunicaciones del CIMAT



CIMAT

Bibliography

- Arcones, M. A.; Mason, D. M. A general approach to Bahadur-Kiefer representations for M -estimators. *Math. Methods Statist.* 6 (1997), 267–292.
- del Barrio, E.; Cuesta-Albertos, J. A.; Matrán, C. Contributions of empirical and quantile processes to the asymptotic theory of goodness-of-fit tests. With discussion. *Test* 9 (2000), 1–96.
- del Barrio, E.; Giné, E.; Matrán, C. Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Ann. Probab.* 27 (1999), 1009–1071.
- Beirlant, J.; Deheuvels, P.; Einmahl, J. H. J.; Mason, D. M. Bahadur-Kiefer theorems for uniform spacings processes. *Teor. Veroyatnost. i Primenen.* 36 (1991), no. 4, 724–743; translation in *Theory Probab. Appl.* 36 (1991), 647–669 (1992)
- Billingsley, P. *Convergence of probability measures.* John Wiley & Sons, Inc., New York-London-Sydney 1968
- Billingsley, P. *Convergence of Probability Measures.* New York: John Wiley & Sons, Inc 1999
- Birnbaum, Z. W.; Marshall, A. W. Some multivariate Chebyshev inequalities with extensions to continuous parameter processes. *Ann. Math. Statist.* 32 (1961), 687–703.
- Bobkov, S.; Ledoux, M. One-dimensional empirical measures, order statistics and Kantorovich transport distances. Preprint (2014).
- Borell, C. The Brunn–Minkowski inequality in Gauss space. *Invent. Math.* 30 (1975), 207–216.
- Breiman, L. On the tail behavior of sums of independent random variables. *Z. Wahrsch. verw. Gebiete* 9 (1967), 20–25.
- Bretagnolle, J.; Massart, P. Hungarian constructions from the nonasymptotic viewpoint. *Ann. Probab.* 17 (1989), 239–256.
- Calitz, F. An alternative to the Kolmogorov-Smirnov test for goodness of fit. *Comm. Statist. Theory Methods* 16 (1987), 3519–3534.

- Chernoff, H.; Savage, I. R. Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist.* 29 (1958), 972–994.
- Chow, Y.S.; Teicher, H. *Probability theory. Independence, interchangeability, martingales.* Springer-Verlag, New York-Heidelberg, 1978.
- Chung, K.-L. An estimate concerning the Kolmogoroff limit distribution. *Trans. Amer. Math. Soc.* 67 (1949), 36–50.
- Csáki, E. Investigations concerning the empirical distribution. *MTA III. Osztály Közleményi* 23, 239–327 (English translation in *Selected Transl. Math. Statist. Probab.* 15 (1981), 229–317)
- Csörgő, M. *Quantile processes with statistical applications.* CBMS-NSF Regional Conference Series in Applied Mathematics, 42. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1983.
- Csörgő, M.; Horváth, L. Approximations of weighted empirical and quantile processes. *Statist. Probab. Lett.* 4 (1986), 275–280.
- Csörgő, M.; Csörgő, S.; Horváth, L.; Mason, D. M. Weighted empirical and quantile processes. *Ann. Probab.* 14 (1986), 31–85.
- Csörgő, M.; Horváth, L. *Weighted approximations in probability and statistics.* With a foreword by David Kendall. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics.* John Wiley & Sons, Ltd., Chichester, 1993
- Csörgő, M.; Révész, P. Strong approximations of the quantile process. *Ann. Statist.* 6 (1978), 882–894.
- Csörgő, M.; Révész, P. *Strong approximations in probability and statistics.* *Probability and Mathematical Statistics.* Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1981.
- Csörgő, S. A probabilistic approach to domains of partial attraction. *Advances in Applied Mathematics* 11 (1990), 282–327.
- Csörgő, S. Rényi confidence bands. *Studia Sci. Math. Hungar.* 34 (1998), 71–87.
- Csörgő, S. Weighted correlation tests for location-scale families. *Hungarian applied mathematics and computer applications.* *Math. Comput. Modelling* 38 (2003), 753–762.
- Csörgő, S.; Deheuvels, P.; Mason, D. M. Kernel estimates of the tail index of a distribution. *Ann. Statist.* 13 (1985), 1050–1077.

Csörgő, S.; Haeusler, E.; Mason, D. M. A probabilistic approach to the asymptotic distribution of sums of independent, identically distributed random variables. *Adv. in Appl. Math.* 9 (1988a), 259–333.

Csörgő, S.; Haeusler, E.; Mason, D. M. The asymptotic distribution of trimmed sums. *Ann. Probab.* 16 (1988b), 672–699.

Csörgő, S.; Haeusler, E.; Mason, D. M. The quantile transform empirical process approach to limit theorems for sums of order statistics. Sums, trimmed sums and extremes, 215–267, *Progr. Probab.*, 23, Birkhäuser Boston, Boston, MA, 1991.

Csörgő, S.; Haeusler, E.; Mason, D. M. The asymptotic distribution of extreme sums. *Ann. Probab.* 19 (1991), 783–811.

Csörgő, S.; Horváth, L.; Mason, D. M. What portion of the sample makes a partial sum asymptotically stable or normal? *Probab. Theory Relat. Fields* 72 (1986), 1–16.

Csörgő, S.; Mason, D. M. Central limit theorems for sums of extreme values. *Math. Proc. Cambridge Philos. Soc.* 98 (1985), 547–558.

Csörgő, S.; Mason, D. M. The asymptotic distribution of sums of extreme values from a regularly varying distribution. *Ann. Probab.* 14 (1986), 974–983.

Csörgő, S.; Mason, D. M. Bootstrapping empirical functions. *Ann. Statist.* 17 (1989), 1447–1471.

Csörgő, S.; Megyesi, Z. Trimmed sums from the domain of geometric partial attraction of semistable laws. *State of the art in probability and statistics (Leiden, 1999)*, 173–194, *IMS Lecture Notes Monogr. Ser.*, 36, Inst. Math. Statist., Beachwood, OH, 2001.

Csörgő, S.; Megyesi, Z. Merging to semistable laws. *Teoriya veroyatnostei i ee primeneniya* 47 (2002), 90–109. [Theory Probab. Appl. 47 (2002), 17–33.]

Csörgő, S.; Viharos, L. On the asymptotic normality of Hill's estimator. *Math. Proc. Cambridge Philos. Soc.* 118 (1995), 375–382.

Csörgő, S.; Viharos, L. Estimating the tail index. *Asymptotic methods in probability and statistics (Ottawa, ON, 1997)*, 833–881, North-Holland, Amsterdam, 1998.

Csörgő, S.; Viharos, L. Joint asymptotic normality of tail-index estimators. *Limit theorems in probability and statistics, Vol. I (Balatonlelle, 1999)*, 457–478, János Bolyai Math. Soc., Budapest, 2002.

Csörgő, S.; Viharos, L. Testing for small bias of tail index estimators. *J. Comput. Appl. Math.* 186 (2006), 232–252.

Daniels, H. E. The statistical theory of the strength of bundles of thread. *Proc. Roy. Soc. London Ser. A* 183, (1945), 405–435.

Deheuvels, P.; Mason, D. M. Bahadur-Kiefer-type processes. *Ann. Probab.* 18 (1990), 669–697.

Deheuvels, P.; Mason, D. M.; Shorack, G. R. Some results on the influence of extremes on the bootstrap. *Ann. Inst. H. Poincaré Probab. Statist.* 29 (1993), 83–103.

Doob, J. L. Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statistics* 20 (1949), 393–403.

Dudley, R. M. (2000). An exposition of Bretagnolle and Massart's proof of the KMT (1975) theorem for the uniform empirical process. In: Notes on empirical processes, lectures notes for a course given at Aarhus University, Denmark, August 2000.

Dudley, R. M. *Uniform Central Limit Theorems*. Cambridge University Press, 2nd ed., 2014.

Dvoretzky, A.; Kiefer, J.; Wolfowitz, J. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* 27 (1956), 642–669.

Eicker, F. The asymptotic distribution of the suprema of the standardized empirical process. *Ann. Statist.* 7 (1979), 116–138.

Einmahl, J. H. J.; Mason, D. M. Laws of the iterated logarithm in the tails for weighted uniform empirical processes. *Ann. Probab.* 16 (1988), 126–141.

Einmahl, U.; Mason, D. M. Approximations to permutation and exchangeable processes. *J. Theoret. Probab.* 5 (1992), 101–126.

Einmahl, U.; Mason, D. M. Gaussian approximation of local empirical processes indexed by functions. *Probab. Th. Rel. Fields.* 107 (1997), 283–311.

Groeneboom, P.; Lopuhaä, H. P.; de Wolf, P. P. Kernel-type estimators for the extreme value index. *Ann. Statist.* 31 (2003), 1956–1995.

Haeusler, E.; Mason, D. M. Weighted approximations to continuous time martingales with applications. *Scand. J. Statist.* 26 (1999), 281–295.

- Haeusler, E.; Mason, D. M. Asymptotic distributions of trimmed Wasserstein distances between the true and the empirical distribution function. *Stochastic inequalities and applications*, 279–298, *Progr. Probab.*, 56, Birkhäuser, Basel, 2003.
- Haeusler, E.; Mason, D. M.; Turova, T. S. A study of serial ranks via random graphs. *Bernoulli* 6 (2000), 541–570.
- Hahn, M. G.; Mason, D. M., Weiner, D. C. (eds) *Sums, trimmed sums and extremes*, *Progr. Probab.*, 23, Birkhäuser Boston, Boston, MA, 1991
- Hájek, J.; Šidák, Z.. *Theory of rank tests*. Academic Press, New York-London; Academia Publishing House of the Czechoslovak Academy of Sciences, Prague 1967
- Jaeschke, D. (1979). The asymptotic distribution of the supremum of the standardized empirical distribution on subintervals. *Ann. Statist.* 7 (1979), 108-115.
- Janssen, A.; Mayer, C-D. Conditional Studentized survival tests for randomly censored models. *Scand. J. Statist.* 28 (2001), 283–293.
- Janssen, A.; Pauls, T. How do bootstrap and permutation tests work? *Ann. Statist.* 31 (2003), 768–806.
- Kiefer, J. On Bahadur's representation of sample quantiles. *Ann. Math. Statist.* 38 (1967), 1323–1342.
- Kiefer, J. Old and new methods for studying order statistics and sample quantiles. *1970 Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969)* pp. 349–357 Cambridge Univ. Press, London
- Kirch, C. *Permutationsprinzipien in der Change-point Analyse*. Diploma thesis Philipps-Universität Marburg, 2003.
- Kirch, C. *Resampling Methods for the Change Analysis of Dependent Data*. Dissertation. Universität zur Köln, 2006.
- Kirch, C.; Steinebach, Josef. Permutation principles for the change analysis of stochastic processes under strong invariance. *Journal of Computational and Applied Mathematics* 186 (2006), 64-88.
- Kirch, C. Block permutation principles for the change analysis of dependent data. *J. Statist. Plann. Inference* 137 (2007), 2453–2474.
- Kirch, C. Bootstrapping sequential change-point tests. *Sequential Anal.* 27 (2008), 330–349.

Komlós, J.; Major, P.; Tusnády, G. An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 32 (1975), 111–131.

Komlós, J.; Major, P.; Tusnády, G. An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 34 (1976), 33–58.

Koul, H. L. Some convergence theorems for ranks and weighted empirical cumulatives. *Ann. Math. Statist.* 41 (1970), 1768–1773.

Major, P. (1999). The approximation of the normalized empirical distribution function by a Brownian bridge,

<http://www.renyi.hu/~major/probability/empir.html>

Mallows, C. L. A note on asymptotic joint normality. *Ann. Math. Statist.* 43 (1972), 508–515.

Marcus, M. B.; Zinn, J. The bounded law of the iterated logarithm for the weighted empirical distribution process in the non-i.i.d. case. *Ann. Probab.* 12 (1984), 335–360.

Mason, D. M. Laws of large numbers for sums of extreme values. *Ann. Probab.* 10 (1982), 754–764.

Mason, D. M. The asymptotic distribution of weighted empirical distribution functions. *Stochastic Process. Appl.* 15 (1983), 99–109.

Mason, D. M. The asymptotic distribution of generalized Rényi statistics. *Acta Sci. Math. (Szeged)* 48 (1985), 315–323.

Mason, D. M. A note on weighted approximations to the uniform empirical and quantile processes. University of Munich technical report, January 1986.

Mason, D. M. A note on weighted approximations to the uniform empirical and quantile processes. Sums, trimmed sums and extremes, 269–283, *Progr. Probab.*, 23, Birkhäuser Boston, Boston, MA, 1991.

Mason, D. M. Notes on the KMT Brownian bridge approximation to the uniform empirical process. *Asymptotic methods in probability and statistics with applications (St. Petersburg, 1998)*, 351–369, *Stat. Ind. Technol.*, Birkhäuser Boston, Boston, MA, 2001a.

Mason, D. M. An exponential inequality for a weighted approximation to the uniform empirical process with applications. *State of the art in probability and statistics (Leiden, 1999)*, 477–498, *IMS Lecture Notes Monogr. Ser.*, 36, Inst. Math. Statist., Beachwood, OH, 2001b.

- Mason, D. M.; Newton, M. A. A rank statistics approach to the consistency of a general bootstrap. *Ann. Statist.* 20 (1992), 1611–1624.
- Mason, D. M.; Shorack, G. R. Necessary and sufficient conditions for asymptotic normality of trimmed L-statistics. *J. Statist. Plann. Inference* 25 (1990), no. 2, 111–139.
- Mason, D. M.; Shorack, G. R. Necessary and sufficient conditions for asymptotic normality of L-statistics. *Ann. Probab.* 20 (1992), no. 4, 1779–1804.
- Mason, D. M.; Schuenemeyer, J.H. A modified Kolmogorov-Smirnov test sensitive to tail alternatives. *Ann. Statist.* 11 (1983), 933–946.
- Mason, D. M.; Shorack, G. R.; Wellner, Jon A. Strong limit theorems for oscillation moduli of the uniform empirical process. *Z. Wahrsch. Verw. Gebiete* 65 (1983), no. 1, 83–97.
- Mason, D. M.; van Zwet, W. R. A refinement of the KMT inequality for the uniform empirical process. *Ann. Probab.* 15 (1987), no. 3, 871–884.
- Mason, D. M.; Zhou, H. H. Quantile coupling inequalities and their applications. *Probab. Surv.* 9 (2012), 439–479.
- Massart, P. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.* 18 (1990), no. 3, 1269–1283.
- Megyesi, Z. Coverage probabilities of Rényi confidence bands. *Studia Sci. Math. Hungar.* 34 (1998), 317–332.
- Munk, A.; Czado, C. Nonparametric validation of similar distributions and assessment of goodness of fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 60 (1998), 223–241.
- Nabeya, S. Absolute moments in 2-dimensional normal distribution. *Ann. Inst. Statist. Math., Tokyo* 3 (1951), 2–6.
- O'Reilly, N. E. On the weak convergence of empirical processes in sup-norm metrics. *Ann. Probability* 2 (1974), 642–651.
- Pyke, R.; Shorack, G. R. Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems. *Ann. Math. Statist.* 39 (1968) 755–771.
- Reiss, R.-D. Approximate distributions of order statistics. With applications to nonparametric statistics. Springer Series in Statistics. Springer-Verlag, New York, 1989
- Révész, P. A joint study of the Kolmogorov-Smirnov and Eicker-Jaeschke statistics. *Statist. and Decisions.* 1 (1982), 57–65.

Rio, E. Local invariance principles and their application to density estimation. *Probab. Theory Related Fields* 98 (1994), no. 1, 21–45.

Serfling, R. J. *Approximation theorems of mathematical statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1980

Shorack, G. R. Functions of order statistics. *Ann. Math. Statist.* 43 (1972), 412–427.

Shorack, G. R. Weak convergence of empirical and quantile processes in sup-norm metrics via KMT-constructions. *Stochastic Process. Appl.* 9 (1979), 95–98.

Shorack, G. R. Embedding the finite sampling process at a rate. *Ann. Probab.* 19 (1991), 826–842.

Shorack, G. R. Inequalities for quantile functions with a uniform Studentized CLT that includes trimming. *J. Nonparametr. Statist.* 8 (1997), 307–335.

Shorack, G. R. *Probability for statisticians*. Springer Texts in Statistics. Springer-Verlag, New York, 2000.

Shorack, G. R.; Wellner, Jon A. *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986.

Skorohod, A.V. Limit theorems for stochastic processes. (Russian) *Teor. Veroyatnost. i Primenen.* 1 (1956), 289–319..

Skorohod, A.V. *Studies in the Theory of Random Processes*. Reading, Mass. Addison–Wesley, 1965.

Talagrand, M. New concentration inequalities in product spaces. *Invent. Math.* 126 (1996), 505–563.

Viharos, L. Asymptotic distributions of linear combinations of extreme values. *Acta Sci. Math. (Szeged)* 58 (1993), 211–231.

Viharos, L. Limit theorems for linear combinations of extreme values with applications to inference about the tail of a distribution. *Acta Sci. Math. (Szeged)* 60 (1995), 761–777.

Wellner, J. A. A martingale inequality for the empirical process. *Ann. Probability* 5 (1977), 303–308.

Wellner, J. A. Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. Verw. Gebiete* 45 (1978), 73–88.



CIMAT