The Weighted Approximation Method of Proving Limit Theorems for Functionals of the Uniform Empirical Process

Lectures Notes

David M. Mason
University of Delaware

The first part of these notes is an expansion and update of lectures that the author gave at Lunteren, Netherlands, to Dutch Ph.D. students in November of 1993. The second part is an elaboration of a number of talks that the author presented in seminars and conferences. A good source for any result in these notes for which a reference is not provided is the monograph: *Empirical Processes with Applications to Statistics* by Shorack and Wellner (1986). These notes have benefited by suggestions and corrections by Sándor Csörgő, Erich Haeusler and Claudia Kirch.

The Classical Empirical Process Technology Circa 1972

Uniform Empirical Distribution

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

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

G_n is a very good estimator of the *uniform cumulative distribution*

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

Glivenko–Cantelli Theorem

The *Glivenko–Cantelli Theorem* says that

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

Actually more is known.

Dvoretsky, Kiefer and Wolfowitz Inequality

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

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

Massart (1990) has shown that one can choose $K = 2$. Notice that when this inequality is combined with the Borel–Cantelli lemma, we get that

$$\sup_{0 \leq t \leq 1} \left| \widetilde{G}_n(t) - t \right| \to 0, \text{ a.s., as } n \to \infty, \tag{2}$$

for any sequence $\left\{\widetilde{G}_n\right\}$ of probabilistically equivalent versions of $\{G_n\}$, meaning that $\widetilde{G}_n =_d G_n$, for each $n \geq 1$.

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 $U_1, ..., U_n$.

Fact 2. (Poisson Approximation) There exists a standard rate one Poisson process $N(x)$, $x \geq 0$, and a sequence $\left\{\widetilde{G}_n\right\}$ of probabilistically equivalent versions of $\{G_n\}$ such that

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

Fact 3. For any sequence $a_n > 0$, $a_n \to 0$ and $na_n \to \infty$ as $n \to \infty$,

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

Fact 4. For any sequence $a_n > 0$, $a_n \to 0$ and $na_n / \log\log n \to \infty$ as $n \to \infty$,

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

For Fact 1 refer to Pyke and Shorack (1968). For the rest of these facts consult Wellner (1978) and the references therein. All of them are found in Shorack and Wellner (1986).

Uniform Empirical Process

The *uniform empirical process* based on $U_1, ..., U_n$, is defined to be

$$\alpha_n(t) = \sqrt{n}\{G_n(t) - t\}, \ t \in [0,1]. \tag{3}$$

It is readily checked that

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

and

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

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

$$(\alpha_n(t_1), \ldots, \alpha_n(t_m)) \to_d (Z_1, \ldots, Z_m), \text{ as } n \to \infty, \tag{4}$$

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

$$cov\,(Z_i,Z_j)=t_i\wedge t_j-t_it_j,\ 1\le i,j\le m.$$

Much more than (4) can be said.

Brownian Bridge

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

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

and

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

The Brownian bridge B has the following representation:

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

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 $0\le t\le 1$ 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).)

Donsker's famous and powerful functional central limit theorem implies that α_n converges in distribution to a Brownian bridge B. We shall soon see that much more can be said about how α_n converges to B. (Consult Billingsley (1968) for a proof of Donsker's theorem.)

The Skorokhod Representation Theorem

The *Skorokhod Representation Theorem* for the uniform empirical process α_n says that there exists a sequence $\{\widetilde{\alpha}_n\}$ of probabilistically equivalent versions of $\{\alpha_n\}$, meaning $\widetilde{\alpha}_n =_d \alpha_n$, for each $n\ge 1$, and a fixed Brownian bridge B such that

$$\sup_{0\le t\le 1}\left|\widetilde{\alpha}_n\,(t)-B\,(t)\right|\to 0,\ \text{a.s., as } n\to\infty. \tag{5}$$

Birnbaum–Marshall Inequality (1961)

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

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

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

$$P\left\{\sup_{0<s\le\delta}\left|\alpha_n(s)\right|/q\,(s)>r\right\}+P\left\{\sup_{0<s\le\delta}\left|\alpha_n(1-s)\right|/q\,(s)>r\right\}\le\frac{C}{r^2}\int_0^{\delta}\frac{du}{q^2\,(u)}$$

and

$$P\left\{\sup_{0<s\le\delta}\left|B(s)\right|/q\,(s)>r\right\}+P\left\{\sup_{0<s\le\delta}\left|B(1-s)\right|/q\,(s)>r\right\}\le\frac{C}{r^2}\int_0^{\delta}\frac{du}{q^2\,(u)}.$$

This says of course that for all $r > 0$

$$\lim_{\delta\searrow 0}\limsup_{n\to\infty}\left[P\left\{\sup_{0<s\le\delta}\left|\alpha_n(s)\right|/q\left(s\right)>r\right\}+P\left\{\sup_{0<s\le\delta}\left|\alpha_n(1-s)\right|/q\left(s\right)>r\right\}\right]=0 \qquad (6)$$

and

$$\lim_{\delta\searrow 0}\left[P\left\{\sup_{0<s\le\delta}\left|B(s)\right|/q\left(s\right)>r\right\}+P\left\{\sup_{0<s\le\delta}\left|B(1-s)\right|/q\left(s\right)>r\right\}\right]=0. \qquad (7)$$

Using this inequality one can show that for any probability space such that

$$\sup_{0\le t\le 1}\left|\widetilde{\alpha}_n\left(t\right)-B_n\left(t\right)\right|\to_p 0,$$

where $\{\widetilde{\alpha}_n\}$ is a sequence of probabilistically equivalent versions of $\{\alpha_n\}$, and $\{B_n\}$ is an appropriate sequence of Brownian bridges, one has for any positive function q on $(0,1)$, increasing on $(0,1/2]$ and decreasing on $[1/2,1)$ such that

$$\int_0^1\frac{du}{q^2\left(u\right)}<\infty$$

one has

$$\sup_{0<t<1}\left|\widetilde{\alpha}_n\left(t\right)-B_n\left(t\right)\right|/q\left(t\right)\to_p 0.$$

We shall return to this soon.

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 Skorokhod Representation Theorem;

in combination with the *probability integral transformation*, namely, that if X is a random variable with cumulative distribution function F, then

$$X=_d Q\left(U\right), \qquad (8)$$

where U is a Uniform $(0,1)$ random variable and Q is the inverse or quantile function of F defined to be

$$Q\left(s\right)=\inf\left\{x:F(x)\ge s\right\}\ \text{ for } 0<s<1.$$

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 $X, X_1,\ldots,X_n$ be i.i.d. with common cumulative distribution function F with corresponding quantile function Q and let $X_{1,n}\le\cdots\le X_{n,n}$ denote their order statistics. Consider the L–statistic

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

where $c_{1,n},\ldots,c_{n,n}$ are constants. By the probability integral transformation (8)

$$(X_1,\ldots,X_n) =_d (Q(U_1),\ldots,Q(U_n)),$$

where $U_1,\ldots,U_n$ are i.i.d. Uniform $(0,1)$ random variables. We then get that

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

where $U_{1,n} \le \cdots \le U_{n,n}$ are the order statistics of $U_1,\ldots,U_n$.

From now on for simplicity of presentation assume that

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

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

$$\sqrt{n}\int_0^1 J(\theta_n(t))(t - G_n(t))\,dQ(t) =_d \sqrt{n}(L_n - \mu). \tag{9}$$

We shall be using the tools:

1. The Glivenko–Cantelli Theorem;

2. Linearity in Probability;

3. Birnbaum–Marshall Inequality;

4. The Skorokhod Representation Theorem.

To obtain the asymptotic distribution of $\sqrt{n}(L_n - \mu)$ it is clear from (9) that it suffices to determine that of

$$\int_0^1 J(\theta_n(t))\,\alpha_n(t)\,dQ(t). \tag{10}$$

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

$$\sup_{0\le t\le 1} |\widetilde{\alpha}_n(t) - B(t)| \to 0,\ \text{a.s., as } n \to \infty.$$

So instead of (10), we shall investigate its probabilistically equivalent version

$$\int_0^1 J\left(\widetilde{\theta}_n\left(t\right)\right)\widetilde{\alpha}_n\left(t\right)dQ\left(t\right). \tag{11}$$

First by the Glivenko–Cantelli theorem (see (2) above)

$$\sup_{0\le t\le 1}\left|\widetilde{\theta}_n\left(t\right)-t\right|\to 0,\ \text{a.s., as } n\to\infty. \tag{12}$$

Therefore by continuity of J and B and (12) for each $0<\delta<1/2$,

$$\sup_{\delta\le t\le 1-\delta}\left|J\left(\widetilde{\theta}_n\left(t\right)\right)\widetilde{\alpha}_n\left(t\right)-J(t)B(t)\right|\to 0,\ \text{a.s., as } n\to\infty. \tag{13}$$

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

$$\int_0^1 J\left(\widetilde{\theta}_n\left(t\right)\right)\widetilde{\alpha}_n\left(t\right)dQ\left(t\right)\to\int_0^1 J\left(t\right)B\left(t\right)dQ\left(t\right), \tag{14}$$

from which it can be inferred that

$$\int_0^1 J\left(\widetilde{\theta}_n\left(t\right)\right)\widetilde{\alpha}_n\left(t\right)dQ\left(t\right)\to_d\int_0^1 J\left(t\right)B\left(t\right)dQ\left(t\right). \tag{15}$$

Since under suitable assumptions on J and Q the random variable $\int_0^1 J\left(t\right)B\left(t\right)dQ\left(t\right)$ 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 (15) that

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

We shall now show how to use the tools in 1, 2, 3 and 4 to establish (14). Choose any $0<\delta<1/2$ and decompose

$$\int_0^1 J\left(\widetilde{\theta}_n\left(t\right)\right)\widetilde{\alpha}_n\left(t\right)dQ\left(t\right)$$
$$=\int_\delta^{1-\delta} J\left(\widetilde{\theta}_n\left(t\right)\right)\widetilde{\alpha}_n\left(t\right)dQ\left(t\right)+\int_0^{\delta} J\left(\widetilde{\theta}_n\left(t\right)\right)\widetilde{\alpha}_n\left(t\right)dQ\left(t\right)+\int_{1-\delta}^{1} J\left(\widetilde{\theta}_n\left(t\right)\right)\widetilde{\alpha}_n\left(t\right)dQ\left(t\right)$$
$$=M_n\left(\delta\right)+L_n\left(\delta\right)+U_n\left(\delta\right).$$

Also write

$$\int_0^1 J\left(t\right)B\left(t\right)dQ\left(t\right)$$
$$=\int_\delta^{1-\delta} J\left(t\right)B\left(t\right)dQ\left(t\right)+\int_0^{\delta} J\left(t\right)B\left(t\right)dQ\left(t\right)+\int_{1-\delta}^{1} J\left(t\right)B\left(t\right)dQ\left(t\right)$$
$$=M\left(\delta\right)+L\left(\delta\right)+U\left(\delta\right).$$

Clearly by (13)

$$M_n\left(\delta\right)\to M\left(\delta\right),\ \text{a.s., as } n\to\infty. \tag{16}$$

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

$$|J(u)| \le K u^{-1/2+\nu_1} \text{ and } |J(1-u)| \le K u^{-1/2+\nu_2}, \text{ for } 0 < u \le 1/2. \tag{17}$$

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

$$\mathbb{B}_1 = \int_0^{1/2} t^{\mu_1} dQ(t) < \infty \text{ and } \mathbb{B}_2 = \int_{1/2}^1 (1-t)^{\mu_2} dQ(t) < \infty. \tag{18}$$

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 Fact 1 and the fact that $U_{1,n} \to_P 0$, for any $\varepsilon > 0$ we can choose a $\lambda > 1$ and n large enough so that with probability greater than or equal to $1 - \varepsilon$,

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

which implies by the first part of (17) that for all $U_{1,n} \le t \le \delta$,

$$|J(\theta_n(t))| \le K_\lambda t^{-1/2+\nu_1} \tag{19}$$

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

$$J(\theta_n(t))(t - G_n(t)) = J(\theta_n(t))\, t = \int_0^t J(u) du,$$

so that inequality (19) still holds. Therefore on this random set

$$|L_n(\delta)| \le K_\lambda \int_0^\delta |\alpha_n(t)|\, t^{-1/2+\nu_1} dQ(t)$$

$$\le K_\lambda \sup_{0<t\le\delta} |\alpha_n(t)|\, t^{-1/2+\nu_1+\mu_1} \mathbb{B}_1.$$

Notice that the function

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

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

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

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

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

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

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

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

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

Hence by "$\varepsilon-$squeezing" (see Theorem 4.2 of Billingsley (1968)),

$$\int_0^1 J\left(\widetilde{\theta}_n(t)\right)\widetilde{\alpha}_n(t)\,dQ(t) \to_P \int_0^1 J(t)\,B(t)\,dQ(t)\,.$$

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

$q-$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 Skorokhod representation

$$\sup_{0<t<1}\left|\widetilde{\alpha}_n(t)-B_n(t)\right|/q(t)\to_p 0,\ \text{ as } n\to\infty. \tag{20}$$

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 (20) to hold it suffices that

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

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

$$\sup_{\delta<t<1-\delta}\left|\widetilde{\alpha}_n(t)-B_n(t)\right|/q(t)\le \max\left(\frac{1}{q(\delta)},\frac{1}{q(1-\delta)}\right)\sup_{0<t<1}\left|\widetilde{\alpha}_n(t)-B_n(t)\right|,$$

$$\sup_{0<t\le\delta}\left|\widetilde{\alpha}_n(t)-B_n(t)\right|/q(t)\le \sup_{0<t\le\delta}\left|\widetilde{\alpha}_n(t)\right|/q(t)+\sup_{0<t\le\delta}\left|B_n(t)\right|/q(t),$$

with a similar bound for $\sup_{1-\delta<t\le 1}\left|\widetilde{\alpha}_n(t)-B_n(t)\right|/q(t)$. Using (5), (6) and (7), we see that (20) follows by "$\varepsilon-$squeezing".

Two 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. 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

$$\sup_{0<t<1}\left|\widetilde{\alpha}_n(t)-B_n(t)\right|\to_p 0,$$

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

$$\sup_{0<t<1}\left|\widetilde{\alpha}_n(t)-\widetilde{B}_n(t)\right|/q(t)\to_p 0$$

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

$$I(c,q) := \int_0^1 (s\,(1-s))^{-1} \exp\left(-\frac{cq^2\,(s)}{s(1-s)}\right) ds < \infty. \tag{21}$$

The crucial facts established by O'Reilly (1974) were

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

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

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, ..., and a sequence of Brownian bridges B_1, B_2, ..., such that for all $n \ge 1$ and $-\infty < x < \infty$,

$$P\left\{ \sup_{0\le t\le 1} |\alpha_n(t) - B_n(t)| \ge n^{-1/2}(a \log n + x) \right\} \le b \exp(-cx), \tag{22}$$

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

Notice that when inequality (22) is combined with the Borel–Cantelli lemma we get the rate of approximation

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

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 a, b and c in (22).

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

$$q\,(t)\,/\,(t \log\log(1/t))^{1/2} \to \infty \text{ and } q\,(1-t)\,/\,(t\log\log(1/t))^{1/2} \to \infty \text{ as } t \searrow 0. \tag{23}$$

Not all q for which (21) is finite for all $c > 0$ satisfy (23). (See M. Csörgő, S. Csörgő, Horváth and Mason [Cs-Cs-H-M] (1986).)

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 $\{B_n\}$.

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

$$\sup_{0 \le t \le 1} \frac{|\alpha_n(t) - B_n(t)|}{(t(1-t))^{1/2-\nu}} = O_p(n^{-\nu}). \tag{24}$$

Moreover, statement (24) remains true for $\nu = 0$ when B_n is replaced by $\overline{B}_n$, where

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

M. Csörgő, S. Csörgő, Horváth and Mason [Cs-Cs-H-M] (1986) first proved this result. Mason and van Zwet (1987) obtained the best possible version of it, allowing $0 \le \nu < \frac{1}{2}$. 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 Skorokhod embedding. More will be said about this later.

Examples of the Use of Weighted Approximations

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 $q(1-s)/s^{1/2} \to \infty$ as $s \searrow 0$. Any q function for which $I(c,q) < \infty$ for some $c > 0$ satisfies this condition. Assume that we are on the probability space of Theorem 1. It is easy to show that

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

when combined with

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

gives for any such q

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

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

$$\sup_{0 < s \le 1/n} |B_n(s)| / q(s) \to_P 0, \text{ as } n \to \infty,$$

and

$$\sup_{0 < s \le 1/n} |B_n(1-s)| / q(1-s) \to_P 0, \text{ as } n \to \infty.$$

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 $na_n \to \infty$. Csáki (1974) established by direct combinatorial methods the somewhat surprising result that

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

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

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

$$\left(\frac{a_n}{1-a_n}\right)^{1/2} \sup_{a_n \le s \le 1} \left| \frac{\alpha_n(s) - B_n(s)}{s} \right|$$

$$\le \left(\frac{a_n}{1-a_n}\right)^{1/2} a_n^{-1/2} n^{\nu} \sup_{a_n \le s \le 1} \left| \frac{\alpha_n(s) - B_n(s)}{s^{1/2-\nu}} \right| \frac{1}{(na_n)^{\nu}} = O_P(1) o(1) = o_P(1).$$

But

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

Thus

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

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 $\tau > 0$,

$$\left| \int_{\tau/n}^{1-\tau/n} v_n(t) d\mu_n(t) - \int_{\tau/n}^{1-\tau/n} B_n(t) d\mu_n(t) \right| \le$$

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

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

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

$$+\sum_{i=1}^{n}[Q(a)1\{X_i < Q(a)\} + Q(1-b)1\{X_i \geq Q(1-b)\}]$$

These sums can be written as

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

Set

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

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 $n \to \infty$,

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

that

$$Z_n(a_n, b_n) := \int_{a_n}^{1-b_n} \alpha_n(s)dQ(s)/\sigma(a_n,b_n) \underset{d}{\to} Z, \text{ as } n \to \infty, \tag{N}$$

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).

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

$$\leq \Delta_{n,\nu}(1)n^{-\nu}\int_{a_n}^{1/2}(s(1-s))^{1/2-\nu}dQ(s)/\sigma(a_n, 1/2)$$

$$+\Delta_{n,\nu}(1)n^{-\nu}\int_{1/2}^{1-b_n}(s(1-s))^{1/2-\nu}dQ(s)/\sigma(1/2,b_n),$$

where

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

Using the fact (e.g. Inequality 2.1 of Shorack (1997)) that for any $0<c<1-d<1$

$$\int_c^{1-d}(s(1-s))^{1/2-\nu}dQ(s)/\sigma(c,d)\le(3/\sqrt{\nu})(c\wedge d)^{-\nu}, \tag{25}$$

we see that this last bound is

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

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 $X_{1,n}\le\cdots\le X_{n,n}$. Consider integers k_n satisfying $1\le k_n\le n/2$, $n\ge 3$, $k_n\to\infty$ and $k_n/n\to 0$, and the intermediate trimmed sum

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

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

$$\frac{T_n\left(k_n\right)-n\int_{k_n/n}^{1-k_n/n}Q(u)du}{\sqrt{n}\sigma\left(k_n/n,k_n/n\right)}\to_d Z, \tag{Z}$$

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

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

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

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

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 $t>0$

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

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

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

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

$$\widehat{c}_n \to_P c, \text{ as } n \to \infty.$$

Let F be the cumulative distribution of X and Q be its inverse. We see by the probability integral transformation (8) that

$$\widehat{c}_n =_d \sum_{i=1}^{k_n} \frac{Q\left(U_{n+1-i,n}\right)}{k_n} - Q\left(U_{n-k_n,n}\right).$$

Set

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

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

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

where

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

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

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

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

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

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

which since

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

and

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

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*

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

See S. Csörgő, Horváth and Mason (1986), S. Csörgő and 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 and 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).

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).

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

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

and

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

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

Setting $d = n$ into these inequalities yields the original KMT inequality (22). 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 (24) is valid. The process β_n is defined below.

Mason and van Zwet Weighted Approximations

Mason and van Zwet (1987) pointed out that their inequality leads to the following useful weighted approximations. For any $0 \le \nu < 1/2$, $n \ge 1$, and $1 \le d < n$ let

$$\Delta_{n,\nu}^{(1)}(d) := \sup_{d/n \le t \le 1} \frac{n^{\nu}|\alpha_n(t) - B_n(t)|}{t^{1/2-\nu}}, \tag{26}$$

$$\Delta_{n,\nu}^{(2)}(d) := \sup_{0 \le t \le 1-d/n} \frac{n^{\nu}|\alpha_n(t) - B_n(t)|}{(1-t)^{1/2-\nu}}, \tag{27}$$

and

$$\Delta_{n,\nu}(d) := \sup_{d/n \le t \le 1-d/n} \frac{n^{\nu}|\alpha_n(t) - B_n(t)|}{(t(1-t))^{1/2-\nu}}. \tag{28}$$

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 $\Delta_{n,\nu}^{(2)}(1)$.

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

Mason (2001b) derived the following improved version of the Mason and van Zwet weighted approximations.

Theorem 3. (An Improved Mason and van Zwet Result). On the probability space of KMT (1975) for every $0 \le \nu < 1/2$ there exist positive constants A_ν and C_ν such that for all $n \ge 2$, $1 \le d < n$ and $0 \le x < \infty$,

$$\begin{aligned} &P\{\Delta_{n,\nu}(d) \ge x\} \\ &= P\left\{ \sup_{d/n \le t \le 1-d/n} \frac{n^{\nu}|\alpha_n(t) - B_n(t)|}{(t(1-t))^{1/2-\nu}} \ge x \right\} \\ &\le 2A_\nu \ \exp(d^{1/2-\nu}C_\nu) \ \exp\left(-\frac{d^{1/2-\nu}cx}{4}\right), \end{aligned}$$

with similar inequalities for $\Delta_{n,\nu}^{(1)}(d)$ and $\Delta_{n,\nu}^{(2)}(d)$.

A Moment Bound for the Weighted Approximation

Theorem 3 readily yields the following uniform moment bounds for (26), (27) and (28).

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

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

with the same statement holding with $\Delta_{n,\nu}(1)$ replaced by $\Delta_{n,\nu}^{(1)}(1)$ or $\Delta_{n,\nu}^{(2)}(1)$.

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 $0 \leq \nu < 1/2$

$$\sup_{n\geq 2}\sup_{r\in\mathcal{R}_n} d_n^{-1}\int_{1/n}^{1-1/n}(s(1-s))^{1/2-\nu}dr(s) =: M < \infty. \tag{29}$$

From Proposition 1 we obtain

Proposition 2. Let $\{R_n, n \geq 2\}$, denote a sequence of classes of nondecreasing left continuous functions on $[1/n, 1-1/n]$ satisfying (29) for some $0 \leq \nu < 1/2$. On the probability space of the KMT (1975) approximation (22) 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 $n^{\nu} I_n \leq \Delta_{n,\nu}(1)M$.

The Uniform Quantile Process

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

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

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

$$\beta_n(t) = \sqrt{n}\{t - U_n(t)\}, \quad \text{for} \ \ 0 \leq t \leq 1. \tag{30}$$

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.

More Exponential Inequalities

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

$$P\{K_{n,\nu} \ge x\} \le b_\nu \exp(-c_\nu x).$$

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

$$P\{\Gamma_{n,\nu} \ge x\} \le A_\nu \exp(-d_\nu x).$$

Example of the Use of Theorem 3: A Result of del Barrio, Giné and Matrán (1999)

We shall first need a definition.

The Domain of Attraction to a Normal Law

Let $X, X_1, X_2, \ldots,$ be a sequence of independent nondegenerate random variables with common distribution function F with left continuous inverse 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

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

where Z is a standard normal random variable. S. Csörgő, Hauesler and Mason (1988) show that one can always chose

$$c_n = nEX$$

and

$$b_n = \sqrt{n}\sigma\left(1/n, 1 - 1/n\right),$$

where

$$\sigma^2\left(1/n, 1 - 1/n\right) = \int_{1/n}^{1-1/n} \int_{1/n}^{1-1/n} \left(s \wedge t - st\right) dQ\left(s\right) dQ\left(t\right).$$

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

Let X, $X_1, X_2, \ldots,$ be a sequence of independent nondegenerate random variables with common distribution function F with left continuous inverse or quantile function Q. Introduce the empirical distribution function

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

Recall the Wasserstein distance between F_n and F,

$$\int_{-\infty}^{\infty} \left|F_n(x) - F(x)\right| dx.$$

Set

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

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. Its asymptotic distribution is highly dependent on the added conditions. Along the way they proved that whenever $F \in DN$, for all $0 < r < 2$,

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

We shall demonstrate how Theorem 3 leads to a quick proof of this result.

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

Observing that

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

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

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

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

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

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

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

A Weighted Approximation Approach to the del Barrio, Giné and Matrán Result

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

$$\sup_{n\geq 2} E\left[\sup_{1/n\leq t\leq 1-1/n} \frac{n^{\nu}|\alpha_n(t) - B_n(t)|}{(t(1-t))^{1/2-\nu}}\right]^r < \infty? \tag{31}$$

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

$$\sup_{n\geq 2} E\left|\frac{\int_{1/n}^{1-1/n} \{|\alpha_n(t)| - E\,|\alpha_n(t)|\}\, dQ(t)}{\sigma(1/n, 1-1/n)}\right|^r < \infty. \tag{32}$$

This was the motivation for Theorem 3, which implies (31). We shall first give a simple proof of (32) for the case $r = 2$ under no assumptions on F based upon the above moment result (31) being true, and then show by taking some pieces out of Barrio, Giné and Matrán that (32) holds for all $r > 0$, again under no assumptions on F. Their proof, based on Talagrand (1996), assumes $F \in DN$.

In our particular situation, we will see that our aim will be to transfer our study of the moment behavior of

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

to that of

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

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 (25))

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

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

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

where

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

Note that

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

Step 2.

This implies both

$$\sup_{n\geq 2} E\left|\frac{\int_{1/n}^{1-1/n} \{|\alpha_n(s)| - |B_n(s)|\}\, dQ(s)}{\sigma\,(1/n, 1-1/n)}\right|^{r} < \infty \tag{33}$$

for any $r > 0$, and (trivially)

$$\sup_{n\geq 2} \left|\frac{\int_{1/n}^{1-1/n} \{E|\alpha_n(s)| - E|B_n(s)|\}\, dQ(s)}{\sigma\,(1/n, 1-1/n)}\right| < \infty. \tag{34}$$

Step 3

To finish the proof when $r = 2$ it clearly suffices to show that

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

This will follow readily 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

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

In particular this implies that

$$0 \le Cov\left(|B(s)|\,,\ |B(t)|\right) \le Cov\left(B(s),\ B(t)\right)$$

and thus

$$E\left(\int_{1/n}^{1-1/n} \left(\{|B(t)| - E\,|B(t)|\}\,dQ\,(t)\right)^2\right.$$

$$= \int_{1/n}^{1-1/n}\int_{1/n}^{1-1/n} Cov\left(|B(s)|\,,\ |B(t)|\right) dQ\,(s)\,dQ\,(t)$$

$$\le \int_{1/n}^{1-1/n}\int_{1/n}^{1-1/n} Cov\left(B(s),\ B(t)\right) dQ\,(s)\,dQ\,(t) = \sigma^2\left(1/n, 1-1/n\right).$$

Implication

This obviously implies that

$$\sup_{n\ge 2} E\left(\frac{\int_{1/n}^{1-1/n}\{|B(t)| - E\,|B(t)|\}\,dQ(t)}{\sigma\left(1/n, 1-1/n\right)}\right)^2 \le 1.$$

Notice that absolutely no assumptions are required on the underlying distribution function (quantile function). But what about the general $r > 0$ case?

General Case

A stronger result is true. By recopying steps from the proof of Theorem 5.1 of Barrio, Giné and Matrán, (also see their Proposition 6.2), based on Borell's inequality, one gets the exponential inequality, for all $t > 0$

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

which, of course, implies that for all $r > 0$,

$$\sup_{n\ge 2} E\left|\frac{\int_{1/n}^{1-1/n}\{|B(t)| - E\,|B(t)|\}\,dQ(t)}{\sigma\left(1/n, 1-1/n\right)}\right|^r < \infty.$$

Notice again that absolutely no assumptions are required on the underlying F. This in combination with (33) and (34) finishes our proof of the Barrio, Giné and Matrán (1999) result based on weighted approximations.

One can say more

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

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

for suitable constants $A > 0$ and $C > 0$.

Notice once more that absolutely no assumptions are required on F. For additional investigations along this line consult Haeusler and Mason (2003), who study the asymptotic distribution of the moderately trimmed Wasserstein distance

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

where a_n is a sequence of positive constants satisfying $a_n \to 0$ and $na_n \to \infty$. As part of a general investigation of the trimmed p^{th} Mallows distance, Munk and Czado (1998) had previously looked at the trimmed Wasserstein distance when $0 < a_n = \alpha < 1/2$.

Some Further Progress on Weighted Approximations

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.

To establish this approximation in its full strength, i.e. (24) 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 (24) 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 Skorokhod embedding scheme as shown by Mason (1991) and M. Csörgő and Horváth (1986).

It naturally then comes to mind that the martingale version of the Skorokhod 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.

Exchangeable Processes

Shorack (1991) was the first to use the Skorokhod 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.

Einmahl and Mason (1992) generalized Shorack's results to exchangeable processes, i.e. to processes of the form

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

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

Assume that

(i) $\sum_{i=1}^{n} Y_n(i) = 0$,

(ii) $\frac{1}{n}\sum_{i=1}^{n} Y_n^2(i) \to_P \sigma^2$ for some $\sigma^2 > 0$, and

(iii) $\max_{1\le i\le n} Y_n^2(i)/n \to_P 0$.

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, Einmahl and Mason (1992) were able to obtained the following weighted approximation to ε_n :

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

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

and (iii) by

(v) $EY_n^4(1) \le K < \infty$ for some $K > 0$ and all $n \ge 1$.
Then on a suitable probability space there exist a sequence of probabilistically equivalent versions $\widetilde{\varepsilon}_n$ of ε_n and a sequence of Brownian bridges $B_1, B_2, \ldots$, such that for all $0 \le \nu < 1/4$ and $\tau > 0$

$$\sup_{\tau/n\le t\le 1-\tau/n} \frac{n^{\nu}|\widetilde{\varepsilon}_n(t) - \sigma B_n(t)|}{(t(1-t))^{1/2-\nu}} = O_P(1). \tag{35}$$

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

$$E\left|Y_n\right|^{\gamma}(1) \le K < \infty \text{ for some } \gamma > 2 \text{ and } K > 0 \text{ and all } n \ge 1, \tag{36}$$

with a corresponding restriction on ν in the conclusion (35). 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 (36) and (iv) by

$$\frac{1}{n}\sum_{i=1}^{n} Y_n^2(i) = \sigma^2 + O_P\left(n^{-2s}\right),$$

where $s = \min\left(\frac{\gamma-2}{2\gamma}, \frac{1}{4}\right)$, then (35) is valid for all $0 \le \nu < s$. 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).

Einmahl and Mason (1992) obtained the approximation (24) 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) have used the Einmahl and Mason (1992) weighted approximation to derive the limiting distribution of certain permutation tests for a change point.

Some Special Cases

1. Set

$$Y_n(i) = n\left\{G_n\left(\frac{i}{n}\right) - G_n\left(\frac{i-1}{n}\right)\right\} - 1,\ i = 1, \ldots, n.$$

This choice yields the weighted approximation (24) to the uniform empirical process given in Theorem 1.

2. Set

$$Y_n(i) = 1 - \frac{n\xi_i}{\xi_1 + \cdots + \xi_n},\ i = 1, \ldots, n,$$

where $\xi_1, \xi_2, \ldots$ 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 (30).

3. Let $c_n(1), \ldots, c_n(n)$, $n \geq 1$, be a triangular array of constants satisfying

$$\sum_{i=1}^{n} c_n(i) = 0, \ \sum_{i=1}^{n} c_n^2(i)/n = 1 \text{ and } \sum_{i=1}^{n} c_n^4(i)/n = O(1).$$

Consider the finite sampling processes

$$\Pi_n(t) = \sum_{i \leq tn} c_n(A_i), \ 0 \leq t \leq 1,$$

where $(A_1, \ldots, A_n)$ is a random permutation of $(1, \ldots, n)$ taken with probability $1/n!$. We get that for $0 \leq \nu < \frac{1}{4}$ and $d > 0$

$$\sup_{d/n \leq t \leq 1-d/n} \frac{|\widetilde{\Pi}_n(t) - \widetilde{B}_n(t)|}{(t(1-t))^{1/2-\nu}} = O_p(n^{-\nu}).$$

The same result holds for the so-called weighted empirical process of Koul (1970):

$$\alpha_{c,n}(t) = \sum_{i=1}^{n} c_n(i) 1\{U_i \leq t\}, \ 0 \leq t \leq 1.$$

Shorack (1991) first proved these results by means of the Skorokhod embedding for martingales. This was also the basic tool that 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.

Weighted Approximations to the Bootstrapped Empirical Process

Weighted Approximations to the Nonparametric Bootstrapped Empirical Process

From results in S. Csörgő and Mason (1989) one can derive the following weighted approximation:

On the same probability space there exist a sequence of i.i.d. F random variables $X_1, X_2, \ldots$, a triangular array

$$\{(M_{1,n}, \ldots, M_{n,n}) : n \geq 1\}$$

of Multinomial $\left(n; \frac{1}{n}, \ldots, \frac{1}{n}\right)$ random vectors and a sequence of Brownian bridges $B_1, B_2, \ldots$, where the $(M_{1,n}, \ldots, M_{n,n})$, $n \geq 1$, and $B_1, B_2, \ldots$, are independent of $X_1, X_2, \ldots$, 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

$$\alpha_{M,n}(x) = \sqrt{n}\left\{F_{M,n}(x) - F_n(x)\right\},$$

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

and

$$F_{M,n}(x) = n^{-1} \sum_{i=1}^{n} M_{i,n} 1\{X_i \le x\}, \quad -\infty < x < \infty.$$

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.

A Weighted Approximation to the General Weighted Bootstrapped Empirical Process

The results of 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

$$\{(W_{1,n}, \ldots, W_{n,n}) : n \ge 1\}$$

is a triangular array of exchangeable random variables satisfying

$$\sum_{i=1}^{n} W_{i,n} = 1, W_{i,n} \ge 0, E\,(nW_{1,n} - 1)^4 = O(1),$$

$$\frac{1}{n} \sum_{i=1}^{n} (nW_{i,n} - 1)^2 = \sigma^2 + O_P\left(n^{-1/2}\right), \text{ for some } \sigma^2 > 0,$$

and

$$\lim_{\varepsilon \searrow 0} \liminf_{n \to \infty} P\{nW_{1,n} > \varepsilon\} = 1.$$

Then on the same probability space there exist a sequence of i.i.d. F random variables $X_1, X_2, \ldots,$ a triangular array

$$\{(W_{1,n}. \ldots, W_{n,n}) : n \ge 1\}$$

as above and a sequence of Brownian bridges $B_1, B_2, \ldots,$ where the $(W_{1,n}, \ldots, W_{n,n})$, $n \ge 1$, and $B_1, B_2, \ldots,$ are independent of $X_1, X_2, \ldots,$ such that for all $0 \le \nu < 1/4$ and $\tau > 0$

$$\sup_{\tau/n \le F(x) \le 1-\tau/n} \frac{|\alpha_{W,n}(x) - \sigma B_n(F(x))|}{(F(x)(1 - F(x)))^{1/2-\nu}} = O_p(n^{-\nu}),$$

where

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

with

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

and

$$F_{W,n}(x) = \sum_{i=1}^{n} W_{i,n} 1\{X_i \le x\}, \quad -\infty < x < \infty.$$

A Weighted Approximation to a Sequence of Continuous Time Martingales

Some Technicalities

Fix any $0 < \bar{t} \le \infty$. For every integer $n \ge 1$, let

$$M_n = (M_n(t))_{0 \le t < \bar{t}}$$

be a sequence of mean zero martingales with respect to a filtration

$$\mathcal{F}_n = (\mathcal{F}_n(t))_{0 \le t < \bar{t}},$$

and satisfying $M_n(0) = 0$.

Assume $EM_n^2(t) < \infty$ for all $n \ge 1$ and $0 \le t < \bar{t}$. Also assume among other conditions, which are too technical to state here, that the predictable quadratic variation $< M_n >$ of M_n converges in a certain way (see Haeusler and Mason (1999) for details) to a function

$$D : [0, \bar{t}) \to [0, \infty)$$

which is continuous, non-decreasing and satisfies

$$D(0) = 0, \ \lim_{t \uparrow \bar{t}} D(t) = \infty.$$

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

$$(\tilde{M}_n)_{n \ge 1} \text{ of } (M_n)_{n \ge 1}, \text{ i.e. } \tilde{M}_n =_d M_n \text{ for each } n,$$

and a standard Wiener process W such that for all $0 < \nu < \beta$,

$$\sup_{t : \frac{1}{n-1} \le D(t) \le n-1} \frac{|\tilde{M}_n(t) - W(D(t))|}{D(t)^{1/2-\nu}(1 + D(t))^{2\nu}} = O_p(n^{-\nu}).$$

The constant $\beta > 0$ depends on a number of technical assumptions.

This result yields the 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 $n \ge 1$,

$$M_n(t) = \frac{\alpha_n(t)}{1-t} = \frac{\sqrt{n}(G_n(t) - t)}{1-t}, \ 0 \le t < 1.$$

Then

$$M_n = (M_n(t))_{0 \le t < 1}$$

is a sequence of mean zero martingales with respect to the filtration

$$\mathcal{F}_n = (\mathcal{F}_n(t))_{0 \le t < 1},$$

where for each $0 \le t < 1$,

$$\mathcal{F}_n(t) = \sigma(G_n(s), \ 0 \le s \le t).$$

In this case it turns out that

$$\langle M_n \rangle(t) = \frac{1}{n}\sum_{i=1}^{n} D_i(t),\ 0 \le t < 1,$$

where for each $i \ge 1$ and $0 \le t < 1$,

$$D_i(t) := \int_0^t \frac{1\{U_i \ge s\}}{(1-s)^3} ds$$

and

$$D(t) = \frac{t}{1-t},\ \text{for } 0 \le t < 1.$$

Applying Theorem 6 to this setup eventually yields the weighted approximation (24) to the uniform empirical process as stated in Theorem 1. In fact, Haeusler and Mason (1999) obtain the Cs-Cs-H-M (1986) approximation via a weighted approximation to the 'randomly' *weighted empirical process*

$$X_n(t) := \sum_{i=1}^{n} w_{i,n}(1\{U_i \le t\} - t),\ 0 \le t \le 1. \tag{37}$$

A special case of their general Theorem 6 yields the following weighted approximation for (37).

Theorem 7. Assume that the weights $w_{i,n}$, $1 \le i \le n$, satisfy the following two conditions:

$$\sum_{i=1}^{n} E w_{i,n}^4 = O(n^{-1}),$$

$$\sum_{i=1}^{n} w_{i,n}^2 - 1 = O_P(n^{-1/2}).$$

Then on a rich enough probability space there exists a sequence of probabilistically equivalent versions $(\widetilde{X}_n)_{n\ge1}$ of $(X_n)_{n\ge1}$ (i.e. $\widetilde{X}\ _n =_d X_n$ for every n) and a standard Brownian bridge B such that for all $0 \le \nu < 1/4$

$$\sup_{1/n \le t \le 1-1/n} \frac{|\widetilde{X}_n(t) - B(t)|}{(t(1-t))^{1/2-\nu}} = O_p(n^{-\nu}), \tag{38}$$

and moreover (38) remains true when the supremum is taken over the entire interval $(0,1)$ in the case $0 < \nu < 1/4$.

Shorack (1991) and 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 approximation result (24) stated in Theorem 1 by choosing $w_{i,n} = 1/\sqrt{n}$ for $i = 1, \ldots, n$.

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 sit a sequence of i.i.d. Uniform $(0,1)$ random variables $U_1, U_2, \ldots,$ it was shown in M. Csörgő, S. Csörgő, Horváth and Mason (1986) and Mason (1991) that one always has for any $0 \le \nu < 1/4$,

$$\sup_{1/n \le t \le 1-1/n} \frac{n^{\nu}\left|\alpha_n(t) - \beta_n(t)\right|}{(t(1-t))^{1/2-\nu}} = O_P(1). \tag{39}$$

However there are at least four methods to enlarge the space to include a sequence of Brownian bridges $B_1, B_2, \ldots$, such that for suitable $\nu_1 \geq 0$,

$$\sup_{1/n \leq t \leq 1-1/n} \frac{n^{\nu_1} \left|\alpha_n(t) - B_n(t)\right|}{(t(1-t))^{1/2-\nu_1}} = O_P(1) \tag{40}$$

and for suitable $\nu_2 \geq 0$,

$$\sup_{1/n \leq t \leq 1-1/n} \frac{n^{\nu_2} \left|\beta_n(t) - B_n(t)\right|}{(t(1-t))^{1/2-\nu_2}} = O_P(1)\,. \tag{41}$$

Method 1. M. Csörgő, S. Csörgő, Horváth and Mason (1986) used the KMT (1975, 1976) strong approximation to the partial sum process to contruct a probability space so that (41) is valid for all $0 \leq \nu_2 < 1/2$ and then inferred that (40) holds on this space for all $0 \leq \nu_1 < 1/4$ via (39). 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 (22) holds that (40) is valid for all $0 \leq \nu_1 < 1/2$ and then inferred that (41) holds on this space for all $0 \leq \nu_2 < 1/4$ via (39).

Method 3. M. Csörgő and Horváth (1986) and Mason (1991) used the Skorokhod embedding to the partial sum process to construct a probability space so that (41) is valid for all $0 \leq \nu_2 < 1/2$ and then inferred that (40) holds on this space for all $0 \leq \nu_1 < 1/4$ via (39).

Method 4. Einmahl and Mason (1992) and Haeusler and Mason (1999) constructed a probability space using the Skorokhod embedding for martingales so that (40) is valid for all $0 \leq \nu_1 < 1/4$ and then inferred that (41) holds on this space for all $0 \leq \nu_2 < 1/4$ via (39).

References

del Barrio, Eustasio; Cuesta-Albertos, Juan A.; Matrán, Carlos. Contributions of empirical and quantile processes to the asymptotic theory of goodness-of-fit tests. With discussion. Test 9 (2000), no. 1, 1–96.

del Barrio, Eustasio; Giné, Evarist; Matrán, Carlos. Central limit theorems for the Wasserstein distance between the empirical and the true distributions. Ann. Probab. 27 (1999), no. 2, 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), no. 4, 647–669 (1992)

Billingsley, Patrick. Convergence of probability measures. John Wiley & Sons, Inc., New York-London-Sydney 1968

Birnbaum, Z. W.; Marshall, Albert W. Some multivariate Chebyshev inequalities with extensions to continuous parameter processes. Ann. Math. Statist. 32 1961 687–703.

Bretagnolle, J.; Massart, P. Hungarian constructions from the nonasymptotic viewpoint. Ann. Probab. 17 (1989), no. 1, 239–256.

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, Miklós; Horváth, Lajos. Approximations of weighted empirical and quantile processes. Statist. Probab. Lett. 4 (1986), no. 6, 275–280.

Csörgő, Miklós; Csörgő, Sándor; Horváth, Lajos; Mason, David M. Weighted empirical and quantile processes. Ann. Probab. 14 (1986), no. 1, 31–85.

Csörgő, Miklós; Horváth, Lajos. 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, Sándor. A probabilistic approach to domains of partial attraction. Advances in Applied Mathematics 11 (1990), 282–327.

Csörgő, Sándor. Rényi confidence bands. Studia Sci. Math. Hungar. 34 (1998), 71–87.

Csörgő, Sándor. Weighted correlation tests for location-scale families. Hungarian applied mathematics and computer applications. Math. Comput. Modelling 38 (2003), no. 7-9, 753–762.

Csörgő, Sándor; Deheuvels, Paul; Mason, David Kernel estimates of the tail index of a distribution. Ann. Statist. 13 (1985), no. 3, 1050–1077.

Csörgő, Sándor; Haeusler, Erich; Mason, David M. A probabilistic approach to the asymptotic distribution of sums of independent, identically distributed random variables. Adv. in Appl. Math. 9 (1988a), no. 3, 259–333.

Csörgő, Sándor; Haeusler, Erich; Mason, David M. The asymptotic distribution of trimmed sums. Ann. Probab. 16 (1988b), no. 2, 672–699.

Csörgő, Sándor; Haeusler, Erich; Mason, David M. The asymptotic distribution of extreme sums. Ann. Probab. 19 (1991), no. 2, 783–811.

Csörgő, Sándor; Horváth, Lajos; Mason, David M. What portion of the sample makes a partial sum asymptotically stable or normal? Probab. Theory Relat. Fields 72 (1986), no. 1, 1–16.

Csörgő, Sándor; Mason, David M. Central limit theorems for sums of extreme values. Math. Proc. Cambridge Philos. Soc. 98 (1985), no. 3, 547–558.

Csörgő, Sándor; Mason, David M. The asymptotic distribution of sums of extreme values from a regularly varying distribution. Ann. Probab. 14 (1986), no. 3, 974–983.

Csörgő, Sándor; Mason, David M. Bootstrapping empirical functions. Ann. Statist. 17 (1989), no. 4, 1447–1471.

Csörgő, Sándor; Megyesi, Zoltán. 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, Sándor; Megyesi, Zoltán. Merging to semistable laws. Teoriya veroyatnosteĭ i ee primeneniya 47 (2002), 90–109. [Theory Probab. Appl. 47 (2002), 17–33.]

Csörgő, Sándor; Viharos, László. On the asymptotic normality of Hill's estimator. Math. Proc. Cambridge Philos. Soc. 118 (1995), 375–382.

Csörgő, Sándor; Viharos, László. Estimating the tail index. Asymptotic methods in probability and statistics (Ottawa, ON, 1997), 833–881, North-Holland, Amsterdam, 1998.

Csörgő, Sándor; Viharos, László. 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, Sándor; Viharos, László. Testing for small bias of tail index estimators. J. Comput. Appl. Math. 186 (2006), 232–252.

Deheuvels, Paul; Mason, David M. Bahadur-Kiefer-type processes. Ann. Probab. 18 (1990), no. 2, 669–697.

Deheuvels, Paul; Mason, David M.; Shorack, Galen R. Some results on the influence of extremes on the bootstrap. Ann. Inst. H. Poincaré Probab. Statist. 29 (1993), no. 1, 83–103.

Dudley, Richard 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.

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.

Einmahl, Uwe; Mason, David M. Approximations to permutation and exchangeable processes. J. Theoret. Probab. 5 (1992), no. 1, 101–126.

Groeneboom, P.; Lopuhaä, H. P.; de Wolf, P. P. Kernel-type estimators for the extreme value index. Ann. Statist. 31 (2003), no. 6, 1956–1995.

Haeusler, Erich; Mason, David M. Weighted approximations to continuous time martingales with applications. Scand. J. Statist. 26 (1999), no. 2, 281–295.

Haeusler, Erich; Mason, David 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, Erich; Mason, David M.; Turova, Tatyana S. A study of serial ranks via random graphs. Bernoulli 6 (2000), no. 3, 541–570.

Hájek, Jaroslav; Šidák, Zbyněk. Theory of rank tests. Academic Press, New York-London; Academia Publishing House of the Czechoslovak Academy of Sciences, Prague 1967

Kirch, Claudia. Permutationsprinzipien in der Changepoint Analyse. Diploma thesis Philipps–Universität Marburg, 2003.

Kirch, Claudia. Resampling Methods for the Change Analysis of Dependent Data. Dissertation. Universität zur Köln, 2006.

Kirch, Claudia; Steinebach, Josef. Permutation principles for the change analysis of stochastic processes under strong invariance. Journal of Computational and Applied Mathematics 186 (2006), no. 1, 64-88.

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), no. 1, 33–58.

Koul, Hira 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

Mason, David M. Laws of large numbers for sums of extreme values. Ann. Probab. 10 (1982), no. 3, 754–764.

Mason, David M. The asymptotic distribution of generalized Rényi statistics. Acta Sci. Math. (Szeged) 48 (1985), no. 1-4, 315–323.

Mason, David 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, David 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, David 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, David M.; Newton, Michael A. A rank statistics approach to the consistency of a general bootstrap. Ann. Statist. 20 (1992), no. 3, 1611–1624.

Mason, David M.; Shorack, Galen R. Necessary and sufficient conditions for asymptotic normality of trimmed L-statistics. J. Statist. Plann. Inference 25 (1990), no. 2, 111–139.

Mason, David M.; Shorack, Galen R. Necessary and sufficient conditions for asymptotic normality of L-statistics. Ann. Probab. 20 (1992), no. 4, 1779–1804.

Mason, David M.; van Zwet, Willem R. A refinement of the KMT inequality for the uniform empirical process. Ann. Probab. 15 (1987), no. 3, 871–884.

Massart, P. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. Ann. Probab. 18 (1990), no. 3, 1269–1283.

Megyesi, Zoltán. Coverage probabilities of Rényi confidence bands. Studia Sci. Math. Hungar. 34 (1998), 317–332.

Munk, Axel; Czado, Claudia Nonparametric validation of similar distributions and assessment of goodness of fit. J. R. Stat. Soc. Ser. B Stat. Methodol. 60 (1998), 223–241.

Nabeya, Seiji. Absolute moments in 2-dimensional normal distribution. Ann. Inst. Statist. Math., Tokyo 3, (1951). 2–6.

O'Reilly, Neville E. On the weak convergence of empirical processes in sup-norm metrics. Ann. Probability 2 (1974), 642–651.

Pyke, Ronald; Shorack, Galen R. Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems. Ann. Math. Statist. 39 (1968) 755–771.

Rio, Emmanuel. Local invariance principles and their application to density estimation. Probab. Theory Related Fields 98 (1994), no. 1, 21–45.

Shorack, Galen R. Functions of order statistics. Ann. Math. Statist. 43 (1972), 412–427.

Shorack, Galen R. Weak convergence of empirical and quantile processes in sup-norm metrics via KMT-constructions. Stochastic Process. Appl. 9 (1979), no. 1, 95–98.

Shorack, Galen R. Embedding the finite sampling process at a rate. Ann. Probab. 19 (1991), no. 2, 826–842.

Shorack, Galen R. Inequalities for quantile functions with a uniform Studentized CLT that includes trimming. J. Nonparametr. Statist. 8 (1997), no. 4, 307–335.

Shorack, Galen R. Probability for statisticians. Springer Texts in Statistics. Springer-Verlag, New York, 2000.

Shorack, Galen 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.

Talagrand, Michel. New concentration inequalities in product spaces. Invent. Math. 126 (1996), no. 3, 505–563.

Viharos, László. Asymptotic distributions of linear combinations of extreme values. Acta Sci. Math. (Szeged) 58 (1993), 211–231.

Viharos, László. 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, Jon A. Limit theorems for the ratio of the empirical distribution function to the true distribution function. Z. Wahrsch. Verw. Gebiete 45 (1978), no. 1, 73–88.