HISTORICAL SURVEY
OF EMPIRICAL PROCESS
AND QUANTILE FUNCTION METHODS

BY GALEN R. SHORACK

University of Washington

Let $X_1, \ldots, X_n$ denote independent rv’s with $dfF$. Let $X_{1,n} \leq \cdots \leq X_{n,n}$ denote the order statistics, and let

\begin{equation}
F_n(x) \equiv \frac{1}{n} \sum_{i=1}^{n} 1_{[X_i \leq x]} \quad \text{for } -\infty < x < \infty
\end{equation}

denote the empirical $df$; here $1_A$ denotes the indicator function of the set $A$. Then

\begin{equation}
e_n(x) \equiv \sqrt{n}[F_n(x) - F(x)] \quad \text{for } -\infty < x < \infty
\end{equation}

is called the empirical process.

Kolmogorov (1933) determined the asymptotic $df$ of the statistic

$\sqrt{n}\|F_n - F\|$ in case $F$ is continuous, here $\| \|$ denotes the supremum norm. This distribution is well known, but we will not record its formula here; we call it simply Kolmogorov’s distribution.

Let $U$ denote a Uniform $(0,1)$ rv. Let $\overset{d}{=} \text{ mean “is distributed as”}$. Then

\begin{equation}
F(X) \overset{d}{=} U \text{ whenever } X \text{ has a continuous } dfF
\end{equation}

\footnote{Supported in part by National Science Foundation Grand DMS-8801083}
This is the *probability integral transformation*. We also have

\[(1) \quad X \equiv F^{-1}(U) \text{ has } dfF, \text{ for any } dfF.\]

This is the *inverse transformation*. These transformations allow one to go back and forth from the general case of the previous paragraph to the special case of Uniform (0, 1) rv's in the next paragraph.

Let \(U_1, \ldots, U_n\) denote iid Uniform (0, 1) rv's with order statistics \(0 \leq U_{1:n} \leq \cdots \leq U_{n:n} \leq 1\) and empirical \(dfG_n\). Then

\[(5) \quad \alpha_n(t) \equiv \sqrt{n}[G_n(t) - t] \quad \text{for } 0 \leq t \leq 1\]

is called the *uniform empirical process*. Also, using the left continuous inverse of a \(df\),

\[(6) \quad \beta_n(t) \equiv \sqrt{n}[G_n^{-1}(t) - t] \quad \text{for } 0 \leq t \leq 1\]

is called the *uniform quantile process* since \(G_n^{-1}(t) = u_{i:n}\) for \((i - 1)/n < t \leq i/n\) and \(1 \leq i \leq n\). Notice that \(G_n(t)\) has the Binomial \((n, t)\) distribution, so the ordinary CLT implies that \(\alpha_n(t) \to_d N(0, t(1 - t))\) as \(n \to \infty\) for each \(0 \leq t \leq 1\). Let \(B(t)\) denote a *Brownian bridge*; that is, \(B\) is a normal process with continuous paths, mean value function 0 and covariance function \(s \wedge t - st\) for \(0 \leq s, t \leq 1\). The ordinary multivariate CLT implies that all *finite dimensional distributions* of \(\alpha_n\) *converge* in distribution to the corresponding distributions of \(B\); we denote this by \(\alpha_n \to_{l.d.} B\) as \(n \to \infty\). This conclusion is not strong enough to imply that \(T(\alpha_n)\) converges in distribution to \(T(B)\)
for continuous (we will leave the metric unspecified) functionals $T$. Next we record the elementary relationship

$$
\beta_n(t) = \sqrt{n}[G_n^{-1}(t) - t] = -\sqrt{n}[G_n(G_n^{-1}(t)) - G_n^{-1}(t)] + \sqrt{n}[G_n(G_n^{-1}(t)) - t] = -\alpha_n(G_n^{-1}(t)) + \sqrt{n}[G_n(G_n^{-1}(t)) - t].
$$

(7)

The SLLN shows that $F_n(x) \rightarrow_{a.s.} x$, and this was strengthened in Glivenko (1933) and Cantelli (1933) to $\|F_n - F\| \rightarrow_{a.s.} 0$. Thus

$$
\|G_n - I\| \rightarrow_{a.s.} 0 \text{ as } n \rightarrow \infty
$$

for the identity function $I$. Since $\|G_n^{-1} - I\| = \|G_n - I\|$ and

$$
\sqrt{n}\|G_n(G_n^{-1}) - I\| \leq n^{-1/2},
$$

it is clear from identity (7) that $\beta_n \rightarrow_{f.d.} -B$.

Because of the probability integral transformation, Kolmogorov’s statistic $\sqrt{n}\|F_n - F\|$ has the same distribution as does $\|\alpha_n\|$ when $F$ is continuous. That is,

$$
\sqrt{n}\|F_n - F\| \overset{d}{=} \|\alpha_n\| \text{ for any continuous } dfF.
$$

Doob’s (1949) paper is a clear landmark in the historical development of these methods. He observed that the result $\alpha_n \rightarrow_{f.d.} B$ suggests that $\|\alpha_n\| \rightarrow_d \|B\|$. Then he applied a reflection principle to derive the distribution of $\|B\|$, and when he obtained Kolmogorov’s distribution he felt sure the technique was of value. Donsker (1952) verified that Doob’s approach could be rigorized. He also sketched a proof that $T(\alpha_n) \rightarrow_d T(B)$ for all functionals $T$ that are a.s. $\|\|$-continuous. Today, we would denote this by saying that $\alpha_n$ converges weakly to $B$ in the norm topology, and write $\alpha_n \Rightarrow B$ as
$n \to \infty$; we would also have to be very careful in our definitions so as to avoid measurability difficulties. We say more on this below.

We next single out the work of Hájek in the early sixties, which was presented with great impact in Hájek and Šidák (1967). An $R$-statistic is of the form

$$\frac{1}{n} \sum_{i=1}^{n} c_{ni} h(R_{in}/(n + 1))$$

for known constants $c_{ni}$, a known function $h$ and for $R_{in}$ denoting the rank of $X_i$ among $X_1, \ldots, X_n$. Hájek's projection technique for proving limit theorems was beautifully applied to $R$-statistics; in this approach a statistic is replaced by the nearest, in the sense of mean square error, statistic belonging to a class that is fundamentally easier to work with. Hájek extended his results from the null hypothesis to local alternatives by applying LeCam's concept of contiguity, in the process of which he gave contiguity great popular appeal.

Led by Tukey (1949), the Statistical Research Group at Princeton had considered trimmed means and Winsorized means in the beginnings of their studies on robustness. Various limiting results had been proved for such statistics, but the first we will single out is Chernoff, Gastwirth and Johns (1967). They, in fact, considered the more general class of $L$-statistics defined by

$$T_n \equiv \frac{1}{n} \sum_{i=1}^{n} c_{ni} h(X_{i,n})$$

for a specified function $h$ and for specified constants $c_{ni}$ typically defined in
terms of some function $J$ by either

$$c_{ni} = J(i/(n+1)) \quad \text{or} \quad c_{ni}/n = \int_{(i-1)/n}^{i/n} J(t) dt.$$  

The key to their approach is that if $V_{i,n} \leq \cdots \leq V_{n,n}$ are Exponential (1) order statistics, then we may assume that

$$(11) \quad V_{i,n} = \sum_{j=1}^{i} \eta_j/(n-j+1), 1 \leq i \leq n, \text{ with } \eta_j \text{'s iid Exponential (1).}$$

This is known as Rényi's (1973) representation for $V_{i,n}$. Note that

$$(12) \quad \nu_{i,n} \equiv EV_{i,n} = \sum_{j=1}^{i} (n-j+1)^{-1}, 1 \leq i \leq n.$$ 

Since $h(X_{i,n}) = h(F^{-1}(U_{i,n})) = h(F^{-1}(1 - \exp(-V_{i,n}))) \equiv \hat{H}(V_{i,n})$, a Taylor series expansion of $T_n$ yields

$$T_n = \frac{1}{n} \sum_{i=1}^{n} c_{ni} \hat{H}(\nu_{i,n}) + \frac{1}{n} \sum_{i=1}^{n} c_{ni}(V_{i,n} - \nu_{i,n})\hat{H}'(\nu_{i,n}) + R_n$$

$$\equiv \mu_n + \frac{1}{n} \sum_{i=1}^{n} c_{ni} \sum_{j=1}^{i} \frac{\eta_j - 1}{n-j+1} \hat{H}'(\nu_{i,n}) + R_n$$

$$= \mu_n + \frac{1}{n} \sum_{j=1}^{n} \left[ \sum_{i=j}^{n} \frac{c_{ni} \hat{H}'(\nu_{in})}{n-j+1} \right] (\eta_j - 1) + R_n$$

$$(13) \quad \equiv \mu_n + \frac{1}{n} \sum_{j=1}^{n} \gamma_{nj}(\eta_j - 1) + R_n$$

for an appropriate remainder $R_n$. Thus $\sqrt{n}(T_n - \mu_n)$ is nothing but a weighted sum of the iid rv's $(\eta_j - 1)$. Adding smoothness assumptions to allow simpler approximations to the $\gamma_{nj}$ and $R_n$, they established a good result on the asymptotic normality of $T_n$.  

5
Bickel (1967) attacked the $L$-statistic problem from a different point of view. His tools were the weak convergence results developed by Prohorov (1956), which we will hear more of later. Letting $M_n(t) \equiv \sum_{i \leq nt} c_{ni}/n$, Bickel noted that

$$T_n \equiv \frac{1}{n} \sum_{i=1}^{n} c_{ni} X_{i,n} = \int_0^1 F^{-1}(nt/(n+1)) dM_n(t) + \int_0^1 q_n(t) dM_n(t)$$

(14)

$$\equiv \mu_n + \int_0^1 q_n(t) dM_n(t)$$

where

$$Q_n(t) \equiv \sqrt{n}[X_{i,n} - F^{-1}(i/(n+1))]$$

$$= \sqrt{n}[F^{-1}(U_{i,n}) - F^{-1}(i/(n+1))]$$

for $1 \leq i \leq n$

with appropriate interpolation between these points. It is natural to call $q_n$ the quantile process. Now $e_n$ and $q_n$ are the general processes, which in the special case of Uniform $(0,1)$ rv’s reduce to $\alpha_n$ and (up to trivial perturbation) $\beta_n$. A Taylor series expansion shows that

$$q_n \approx (F^{-1})' \beta_n = \beta_n/f(F^{-1})$$

(16)

so one would suspect that $q_n$ converges in some sense to $-B/f(F^{-1})$. The difficulty with such convergence arises in the extreme tails, where even $df$’s $f$ having “nice” densities $f$ may well have $1/f(F^{-1}(t))$ growing to infinity at extremely fast rates as $t$ approaches 0 or 1. Thus Bickel only showed that $q_n \Rightarrow -B/f(F^{-1})$ on $[a,b]$ for $0 < a < b < 1$, and his limit theorem took the form

$$\sqrt{n}(T_n - \mu_n) \to_d - \int_a^b [B/f(F^{-1})] dM$$

(17)
provided $M_n \to_d M$ on $[a, b]$ with $c_{ni}$ equal to 0 if $i < na$ or $i > nb$ and provided $f$ was smooth enough.

Consider the limiting rv in Bickel's result. Now any rv of the type $\int_0^1 gBdM$ is defined to be the limit of approximating sums of the type $\sum g(t_i)B(t_i)\Delta M_i$. This approximating rv is normal with mean 0 and variance

$$\sum \sum g(t_i)g(t_j)\text{Cov}[B(t_i), B(t_j)]\Delta M_i\Delta M_j \to \int_0^1 \int_0^1 g(s)g(t)(s \wedge t - st)dM(s)dM(t),$$

where the convergence claims will require hypotheses. Thus

$$\int_0^1 gBdM \overset{d}{=} N(0, \int_0^1 \int_0^1 g(s)g(t)(s \wedge t - st)dM(s)dM(t))$$

under regularity on $g$ and $M$. Since this variance can also be written as

$$\int_0^1 g^2dM - \left(\int_0^1 gdM\right)^2$$

using integration by parts, as in Chernoff and Savage (1958), we could also write

$$\int_0^1 gBdM \overset{d}{=} N(0, \sigma_K^2) \text{ for } K \text{ defined by } dK = gdM.$$

Stigler (1969) applied the projection technique of Hájek to $L$-statistics. Thus $X_{i,n}$ was replaced by its projection $\hat{X}_{i,n}$ onto the space generated by $L_2$ functions of $X_1, \ldots, X_n$. This $\hat{X}_{i,n}$ was nothing but a weighted sum of functions of the individual $X_i$'s. Hence the projection $\hat{T}_n$ of $T_n$ was just a weighted sum of independent rv's. Under regularity, this weighted sum $\hat{T}_n$
and the difference $T_n - \hat{T}_n$ were approximated simply, so that a good theorem on asymptotic normality of $T_n$ followed.

Another front had actually begun earlier that involved the investigation of the partial sums $X_k \equiv (X_1 - \mu) + \cdots + (X_k - \mu), 1 \leq k \leq n$. Erdős and Kac (1946) had established that for iid $X_i$'s with mean 0 and variance 1 the rv's

$$\text{(20)} \quad \max_{1 \leq k \leq n} S_k / \sqrt{n}, \max_{1 \leq k \leq n} |S_k| / \sqrt{n}, \sum_{k=1}^{n} |S_k| / n^{3/2} \text{ and } \sum_{k=1}^{n} S_k^2 / n^2$$

have limiting distributions independent of the underlying $df F$. Their method of proof was to show that the limiting $df$ was independent of $F$, and then establish its form in the simplest possible special case; this method was called the invariance principle. Donsker (1951) defined a partial sum process $S_n$ by letting

$$\text{(21)} \quad S_n(t) \equiv S_{[nt]} / \sqrt{n} \text{ for } 0 \leq t \leq 1.$$ 

He noted from the elementary CLT that $S_n \rightarrow_{d} W$, where $W$ is a normal process with continuous paths, mean value function 0 and covariance function $s \wedge t$ for $0 \leq s, t \leq 1$. This process is known as the Weiner process or Brownian motion. His motivation was the fact that the four rv's of Erdős and Kac can be written as $\|S_n^+\|, \|S_n\|, \int_0^1 |S_n(t)| dt$ and $\int_0^1 S_n^2(t) dt$. Doob's reasoning leads one to conclude that the limiting distribution of these rv's should agree with those of $\|W^+\|, \|W\|, \int_0^1 |W(t)| dt$ and $\int_0^1 W^2(t) dt$. Donsker went further by showing that $S_n \Rightarrow W$ in the $\| \|\cdot\|$-topology; in this first paper of Donsker, possible measurability difficulties mentioned earlier are much less severe as
the process $S_n$ is only discontinuous at the fixed points of the form $k/n$. Donsker (1952) then established the result for $\alpha_n$ mentioned earlier.

Prohorov (1956) presented a unified theory of weak convergence, and applied it to partial sum and empirical processes. The natural space on which to consider these random processes was $D$, defined to be the collection of all right continuous functions on $[0,1]$ possessing left hand limits at each point. The natural metric to use is the supremum norm $\| \|$. Prohorov’s theory of weak convergence was developed for processes defined on complete separable metric spaces. Herein lay the rub that would lead to measurability difficulties; the $\| \|$-topology on $D$ is a complete, but nonseparable, metric space. One way to get around this is to use appropriate linear interpolation in the definition of the processes $S_n, \alpha_n$, etc. so that they are processes taking values in $C$, the set of all continuous functions on $[0,1]$. The $\| \|$-topology on $C$ is a complete separable metric space. Getting all this straight took some time and led to problems in some very useful papers.

We need to discuss Prohorov’s general theory of weak convergence on a complete separable metric space $(M, \delta)$ in a bit of detail before we can go on. We will suppose $M$ is a collection of functions defined on some subinterval of the real line. Let $\{X_n : n \geq 0\}$ denote measurable transformations from the underlying probability space $(\Omega, \mathcal{A}, \text{Prob})$ to the metric space $(M, \delta)$ that induces probability measures $P_n, n \geq 0$, on $(M, \mathcal{B})$, where $\mathcal{B}$ denotes the Borel $\sigma$-field generated by the $\delta$-open subsets of $M$. Then $X_n$ is said to converge
weakly to $x_0$ as $n \to \infty$ provided $h(x_n) \to_d h(x_0)$ for all $\mathcal{B}$-measurable functionals $h$ that are a.s. $\delta$-continuous; this is actually an equivalent form of the standard definition. We write

$$x_n \Rightarrow x_0 \text{ on } (M, \mathcal{B}, \delta)$$

to denote this weak convergence. The major theorem about weak convergence is that $x_n \Rightarrow x_0$ on $(M, \mathcal{B}, \delta)$ provided that $x_n \to_{L_d} x_0$ and provided there are some $\delta$-compact subsets $K_\varepsilon$ of functions of $M$ for which $P_n(K_\varepsilon) \geq 1-\varepsilon$ for all $n \geq 1$ holds for each $\varepsilon > 0$. Processes $x_n$ satisfying the latter condition are called tight. To show the processes $x_n$ are tight one has to characterize the $\delta$-compact sets in $M$ and then show the existence of the set $K_\varepsilon$ above. Prohorov did all this, and applied it to standard processes such as $\xi_n$ and $\alpha_n$ on $D$. Chentsov (1956) gave another, particularly useful, condition for tightness. So far we have not mentioned Prohorov's metric on $D$. Most useful for us, Prohorov showed that one can work with a metric on $D$ that was devised by Skorokhod (1956). Such work was polished and presented in the later work of Billingsley (1968) that did much to popularize this theory. In particular, Billingsley developed very useful moment inequalities, in the spirit of Chentsov, that implied tightness and were easy to verify for $\xi_n$ and $\alpha_n$. He also gave a very careful discussion of various metrics on $D$ that generate the topology used by Skorokhod.

We embark on a brief discussion of properties of the so called Skorokhod topology on $D$ and of a certain metric $d$ that generates it; see Billingsley.
Let \( \{X(t) : 0 \leq t \leq 1 \} \) be a collection of rv’s on \((\Omega, \mathcal{A}, \text{Prob})\) taking values in \(D\), so that \(X\) is a measurable transformation relative to the \(\sigma\)-field \(\mathcal{D}\) generated by the finite dimensional subsets of \(D\). This transformation thus induces a measure \(P_X(F) \equiv \text{Prob}(X^{-1}(F))\) for all \(F \in \mathcal{D}\). Now consider a \(\|\cdot\|\)-continuous functional \(h\) from \(D\) to the real line. This need not induce a measure on \((R, \mathcal{B})\), with \(\mathcal{B}\) the Borel subsets of the real line, since the \(\sigma\)-field \(\mathcal{D}_{\|\cdot\|}\) generated by the \(\|\cdot\|\)-open subsets of \(D\) happens to be strictly larger than \(\mathcal{D}\). Thus a measurability problem arises. The Skorokhod topology avoids this problem because the \(\sigma\)-field \(\mathcal{D}_d\) generated by the \(d\)-open subsets of \(D\) is equal to \(\mathcal{D}\). Moreover, the \(d\)-topology on \(D\) is complete and separable so that Prohorov’s theory can be applied. Finally, if \(d(x_n, x) \to 0\) with \(x \in \mathcal{C}\), then \(\|x_n - x\| \to 0\). Skorokhod (1956) made a truly important advance: if \(X_n \Rightarrow X_0\) on \((M, \mathcal{B}, \delta)\), as described above, where \((M, \delta)\) is a complete separable metric space, then there exist processes \(X_n^*\) on some new probability space \((\Omega^*, \mathcal{A}^*, \text{Prob}^*)\) that induce the same distributions \(P_n\) on \((M, \mathcal{B})\) as do the \(X_n\) and that satisfy the even stronger convergence statement \(\delta(X_n^*, X_0^*) \to_{a.s.} 0\) as \(n \to \infty\). We refer to this as the Skorokhod construction of equivalent processes.

We remark on the simplest special case of the Skorokhod construction. Suppose \(F_n \rightarrow_d F_0\). Then with \(X_n^* \equiv F_n^{-1}(U)\) for a single Uniform \((0,1)\) rv \(U\), it is easy to show that \(X_n^* \rightarrow_{a.s.} X_0^*\). The hypothesis says that most vertical distances between \(F_n\) and \(F\) go to 0, while the conclusion says that most
horizontal distances between $F_n$ and $F$ go to 0.

That $S_n \Rightarrow W$ and $\alpha_n \Rightarrow B$ on $(D, \mathcal{D}, d)$ follows from Prohorov. Skorokhod's work can be applied to claim that

$$\|S_n - W\| \rightarrow_{a.s.} 0$$

and

$$\|\alpha_n - B\| \rightarrow_{a.s.} 0$$

for Skorokhod constructions of these processes (where we now drop the subscript *). Of course, the identity (7) then gives

$$\|\beta_n - (-B)\| \rightarrow_{a.s.} 0$$

for the Skorokhod construction.

Suppose $F$ is continuous. Kolmogorov's statistic is distributionally just $\|\alpha_n\|$. Instead, one could consider the Cramér (1928) or von Mises (1931) $\|\|\|$ norm defined by

$$\|\alpha_n\|^2 \equiv \int_0^1 \alpha_n^2(t)dt$$

or the Anderson and Darling (1952) type of generalization

$$\|\alpha_n\|_{\psi}^2 \equiv \int_0^1 \alpha_n^2(t)\psi(t)dt.$$

At the heart of Chibisov's (1964, 1965) investigation of such tests was his consideration of the convergence of $\alpha_n$ to $B$ in the metric $d_{\psi}(x, y) \equiv d(x\psi, y\psi)$ on $D_{\psi} \equiv \{x : x\psi \in D\}$. Chibisov's (1965) result shows that for a positive $\psi$,
symmetric about \( t = 1/2 \) and \( \gamma \) on \([0, 1/2]\) we have:

\[
\alpha_n \Rightarrow B \text{ on } (D_\psi, \mathcal{B}, d_\psi)
\]

if and only if

\[
\int_0^{1/2} t^{-1/2} \exp(-c\psi^2(t)/t)dt < \infty \text{ for all } \epsilon > 0 \text{ where } \psi \equiv 1/\psi. 
\]

We call this Chibisov’s condition. Chibisov’s (1961) paper had claimed that \( \alpha_n \Rightarrow B \text{ on } (D_\psi, \mathcal{B}, \rho_\psi) \) if and only if his condition held, where \( \rho_\psi(x, y) \equiv ||(x - y)\psi||; \) but he soon corrected this to the result above because of the previously mentioned measurability difficulties. O’Reilly (1974) obtained this same result for \( \beta_n \). The class of all functions \( \psi \) satisfying the requirements above will be denoted by \( Q \). We mention that any \( \psi_n(t) \equiv [t(t - t)]^{1/2} - \nu \) with \( 0 < \nu \leq 1/2 \) satisfies Chibisov’s condition, but \( \psi_0(t) \equiv [t(1 - t)]^{1/2} \) does not. Note that \( \alpha_n(t)/\psi_0(t) \) has mean 0 and variance 1 for all \( t \).

Suppose \( F \) has mean 0 and variance 1. Skorokhod (1965) opened another area of research when he showed how to define a random stopping time \( \tau \) so that Brownian motion \( W \) stopped at time \( \tau \) has \( dfF \); in fact, \( X \equiv W(\tau) \) has \( dfF \) and \( E\tau = 1 \). Since \( W_1(t) \equiv W(t + \tau) - W(\tau) \) is a Brownian motion independent of \( W \), one can keep going in this manner to produce a sequence of iid rv’s \( X_1, \ldots, X_n \) with partial sums \( S_k \equiv X_1 + \cdots + X_k \) for which \( S_k = W(\tau_1 + \cdots + \tau_k) \) and \( \tau_1, \tau_2, \ldots \) are iid rv’s with mean 1. Thus the partial sum process \( S_n = W(\sum_{k \leq [n]} \tau_k)/\sqrt{n} \) of these specially constructed rv’s has been embedded in the Brownian motion \( W \). A number of authors
devised various means of defining the stopping times $\tau_k$. We will refer to all such schemes under the generic title of Skorokhod embedding. Note that 

$$W_n(t) = W(nt)/\sqrt{n}$$

is again Brownian motion with $W_n(k/n) = W(k)/\sqrt{n}$ while $S_n(k/n) = W\left(\sum_{i=1}^k \tau_i\right)/\sqrt{n}$ with $E\left(\sum_{i=1}^k \tau_i\right) = k$. There is thus reason to hope that $\|S_n - W_n\|$ converges to 0. Strassen's (1967) scheme satisfies

$$\|S_n - W_n\| = O((\log n)^{1/2}(\log_2 n)^{1/4}/n^{1/4}) \text{ a.s. provided } EX^4 < \infty.$$ 

This began a considerable amount of research directed towards obtaining best rates of convergence. Also, Breiman (1968) presented an embedding of a triangular array of row independent rv's $X_{ni}$ with $dfF$ having mean 0 and variance 1 for which the partial sum process $S_n$ of the rv's in the $n$-th row satisfies

$$\|S_n - W\| \to_p 0;$$

note that the Brownian motion $W$ appearing here does not have a subscript. Most important though, Breiman took advantage of an old observation. If $\eta_1, \ldots, \eta_{n+1}$ are independent Exponential (1) rv's, then

$$\frac{\eta_1 + \cdots + \eta_i}{\eta_1 + \cdots + \eta_{n+1}}, 1 \leq i \leq n,$$

are jointly distributed

as are the Uniform (0,1) order statistics $U_{i,n}$, $1 \leq i \leq n$.

Letting $S_{n+1}$ denotes the partial sum process of row independent rv's $\eta_{ni} - 1$ and letting $\beta_n$ denote the uniform quantile process of uniform order statistics defined by $U_{i,n} \equiv (\eta_1 + \cdots + \eta_{ni})/(\eta_1 + \cdots + \eta_{n,n+1})$, the approximation

$$\beta_n \approx S_{n+1} - IS_{n+1}(1)$$
is so close that it is trivial to show that

\[(32) \| \beta_n - B \| \to_d 0 \]

where

\[(33) - B \equiv W - IW(1) \text{ is a Brownian bridge.} \]

Because of the companion identity to (7), the empirical process \( \alpha_n \) of the same \( U_{i,n} \)'s satisfies

\[(34) \| \alpha_n - B \| \to_d 0. \]

From this point on, Strassen's scheme, and all improvements made in it, would be immediately applied to obtain rates in the convergence of \( \alpha_n \) and \( \beta_n \) to sequence of Brownian bridges.

Meanwhile Pyke and Shorack (1968) used Skorokhod's (1956) construction to give a new proof of a Chernoff and Savage (1958) limit theorem for \( R \)-statistics under fixed alternatives in the two sample problem. Of fundamental importance in this paper was that the value of replacing \( \alpha_n \Rightarrow B \) on \( (D, D, d) \) by the fact that \( \| \alpha_n - B \| \to_{a.s.} 0 \) for a Skorokhod construction of these processes in order to obtain technically simpler proofs of \( \to_d \) for functionals of empirical processes was clearly recognized and illustrated. They also introduced \( \| /q\| \)-metrics into these a.s. convergent constructions, in that their proofs actually required \( \|(\alpha_n - B)/q\| \to_p 0 \) for \( q \in Q \) satisfying \( \int_0^1 q^{-2}(t)dt < \infty. \) They also overcame the measurability difficulties through the simple device of restricting attention to \( \| \|\)-continuous functionals of \( D \)-measurable processes on \( D. \) Having recognized the power of this technique,
Pyke and Shorack set out to popularize it. Among these were Pyke (1969) and Shorack (1969), both of which considered the \( L \)-statistics \( T_n \) of (10), but which started from different identities. In the process both considered convergence of the quantile process \( Q_n \) in appropriate \( \| q \|_{-}\)-metric, and on all of \([0,1]\). Shorack (1972) summed the \( L \)-statistics by parts so that the required regularity conditions fell mainly on the \( J \) function controlled by the statistician and less heavily on the \( df F \) controlled by nature. This is a good result on the asymptotic normality of the \( L \)-statistic \( T_n \). A rough sketch of the proof follows, simplified to the special case

\[
c_{n}/n = \int_{(t-1)/n}^{t/n} J(t) dt
\]

for a streamlined presentation. For

\[
\mu \equiv \int_{0}^{1} J(t) g(t) dt, \quad g = h \circ F^{-1}
\]

and \( \Gamma \) defined by \( d\Gamma = J ds \) we have

\[
\sqrt{n}(T_n - \mu) = \sqrt{n} \left[ \int_{0}^{1} g(G_n^{-1}(t)) J(t) dt - \mu \right] = \sqrt{n} \int_{0}^{1} gd(\Gamma(G_n) - \Gamma)
\]

(35) \quad \sqrt{n}(T_n - \mu) = \sqrt{n} \left[ \int_{0}^{1} g(G_n^{-1}(t)) J(t) dt - \mu \right] = \sqrt{n} \int_{0}^{1} gd(\Gamma(G_n) - \Gamma)

(36) \quad \mu = - \int_{0}^{1} \sqrt{n} [\Gamma(G_n) - \Gamma] dg

(37) \quad \mu = - \int_{0}^{1} \alpha_n J dg + o_n(1)

(38) \quad \mu \stackrel{d}{\rightarrow} - \int_{0}^{1} BJ dg \stackrel{d}{\equiv} N(0, \sigma^2_K) \text{ for } dK \equiv JdG, \text{ see (19),}

under appropriate assumptions that roughly require \( K(U) \) to have a finite \( 2 + \nu \) moment for some \( \nu > 0 \). The details of such a proof rely on arguments
like
\[
\left| \int_0^1 \alpha_n J dg - \int_0^1 BJ dg \right| \leq \| (\alpha_n - B)/q \| \int_0^1 q |J| dg |
\]
\[
= o_p(1)0(1) = o_p(1)
\]
in which the Skorokhod construction is used to justify the $o_p(1)$ term and \( \int_0^1 q |J| dg \) is assumed finite.

At this point we have three good theorems on $L$-statistics based on the representation (11), on projection and on empirical process results. Many other authors proved good asymptotic normality results for $L$-statistics, even though we have not chosen to discuss them here.

The landmark papers dealing with constructions of $s_n$, $W_n$ and $\alpha_n$, $B_n$ (with $B_n$, $n \geq 1$, denoting a sequence of Brownian bridges) that converge at best possible rates were Komlós, Major and Tusnády (1975,1976). A corollary to this KMT construction was
\[
\| \alpha_n - B_n \| = O(\sqrt{\log n} / \sqrt{n}) \text{ a.s.}
\]
Various important papers preceded this and many variations on their result would follow. We single out two of these.

If one uses the construction of Breiman (1968) mentioned earlier to construct a version of the empirical processes $\alpha_n$, then the rv's $\{U_{i,n} : 1 \leq i \leq n\}$ so obtained are not a subset of the rv's $\{U_{i,n+1} : 1 \leq i \leq n + 1\}$ as they should be if only a single sequence of iid Uniform (0,1) rv's was involved. Keifer (1972) showed how a different construction that avoided this problem could be made, and he analyzed its rate of convergence. He did
this by embedding the sequence of empirical processes $\alpha_n$ of a single sequence $U_1, U_2, \ldots$ into what we now call a Kiefer process. This Kiefer process
\[ \{K(s, t) : 0 \leq s \leq 1, t \geq 0\} \]
is normal with continuous paths, 0 means and covariance function $(s_1 \wedge s_2 - s_1 s_2)(t_1 \wedge t_2)$ for $0 \leq s_1, s_2 \leq 1$ and $t_1, t_2 \geq 0$. It is true that for each fixed $t \geq 0$ the process $K(\cdot, t)/\sqrt{t}$ is a Brownian bridge.

Note that $\sqrt{\ln s}/n \alpha(t) \to_d K(s, t)$ for all $0 \leq s \leq 1$, $t \geq 0$.

M. Csörgö, S. Csörgö, Horváth and Mason (1986a) showed that one can construct a triangular array of row independent Uniform $(0, 1)$ rv's $\{U_{ni} : 1 \leq i \leq n, n \geq 1\}$ and a sequence of Brownian bridges $B_n$ such that the uniform empirical process $\alpha_n$ of the rv's in the $n$-th row satisfies

\[ \left\| (\alpha_n - B_n)/(I(1 - I))^{(1/2)-\nu} \right\|_{1/n}^{1-1/n} = O_p(n^{-\nu}) \quad \text{as } n \to \infty \]

and
\[ \left\| (\beta_n - B_n)/(I(1 - I))^{(1/2)-\nu} \right\|_{1/n}^{1-1/n} = O_p(n^{-\nu}) \]

for each fixed $0 < \nu < 1/4$. The order statistics $U_{i,n}$ of these $U_{ni}$ are actually of the form

\[ U_{i,n} = (\eta_1^0 + \cdots + \eta_i^0)/(\eta_1^0 + \cdots + \eta_{n+1}^0) \equiv \frac{S_i^0}{S_{n+1}^0} \]

for iid Exponential $(1)$ rv's $\eta_i^0$. Thus

\[ nG_n(t/n) = \sum_{i=1}^n 1_{U_{i,n} \leq t/n} = \sum_{i=1}^n 1_{\tau_i^0 \leq S_{n+1}^0/n} = N^0(t S_{n+1}^0/n) \]

where $N^0$ is a Poisson process with arrival times $S_i^0$. Moreover, this construction uses an analogous method to construct the rv's $1 - U_{a,n}, 1 - U_{a-1,n}, \ldots$.
starting at the right hand end. This leads to an independent Poisson process $\mathcal{N}^1$ associated with the right hand tail. These authors also point out how the above construction can be modified to obtain a single sequence of iid Uniform (0,1) rv's $U_1, \ldots, U_n, \ldots$ for which (41) and (42) hold. However, (44) is no longer valid on this modified probability space. We refer to these as the CsCsHM constructions. These results were shown strong enough to yield the Chibisov and O'Reilly results cited above, as well as a host of others, which the KMT construction alone can not. M. Csörgö, S. Csörgö, Horváth and Mason (1986b) is also full of applications.

That the CsCsHM constructions represent a major advance over previous constructions is also shown in our consideration of the $L$-statistic $T_n$ of (35). Using the Skorokhod construction in the argument (38) with the function $q_\theta$, for $0 \leq \theta < 1/4$, causes us to “lose” the factor $[\theta (1 - t)]^\theta$, and we end up needing a bit more than a second moment in the proof. However, when we do this with the CsCsHM construction, we still have the factor $O_\theta(n^{-\gamma})$ to use. This allows one to obtain best possible results, as we presently relate.

Interest in the asymptotic behaviour of trimmed means underwent a resurgence in the eighties, as evidenced by this volume. We will focus on S. Csörgö, Haeusler and Mason (1988a,1988b), and then consider Mason and Shorack (1988) on $L$-statistics again.

We now let $c_{ni}/n = \int_{(i-1)/n}^{i/n} J(t)dt$, $g = h(F^{-1})d\gamma = JdT$ and consider
the trimmed $L$-statistic

\begin{equation}
T_n \equiv T_n(k_n, m_n) \equiv \frac{1}{n} \sum_{i=k_n+1}^{m_n} c_{ni} g(U_{i,n}).
\end{equation}

for some $0 \leq k_n \leq m_n \leq n$, with centering and scaling constants

\begin{equation}
\mu_n \equiv \mu(k_n/n, m_n/n) \equiv \int_{k_n/n}^{m_n/n} g(t)J(t)dt
\end{equation}

and

\begin{equation}
\sigma_n \equiv \sigma[k_n/n, m_n/n)
\end{equation}

where

\begin{equation}
\sigma^2[a, b] \equiv \text{Var}\left[ \int_{[a, b]} BdK \right] \text{ with } dK = Jdg.
\end{equation}

The centered and scaled statistic of interest is

\begin{equation}
S_n \equiv S_n(k_n, m_n) \equiv \sqrt{n}(T_n(k_n, m_n) - \mu_n(k_n/n, m_n/n)) / \sigma_n.
\end{equation}

We have used $J$ and $h$ in this paragraph so that we don’t have to repeat all this notation again. However, we stress that what we now present was born in S. Csörgő, Hecussler and Mason (1988a,1988b) with $J \equiv 1$, $c_{ni} \equiv 1$, $g = F^{-1}$ and $K = F^{-1}$ so that

\begin{align}
S_n & \equiv S_n(k_n, m_n) \\
& = \sqrt{n} \left[ \frac{1}{n} \sum_{i=k_n+1}^{m_n} X_{i,n} - \int_{k_n/n}^{m_n/n} F^{-1}(t)dt \right] / \text{StDev} \left[ \int_{[k_n/n, m_n/n]} BdF^{-1} \right]
\end{align}

is the centered and scaled trimmed mean.
We carry on, still using \( J, c_n, g \) and \( K \) in the notation, but with the understanding for now that these equal 1, 1, \( F^{-1} \) and \( F^{-1} \). Integration by parts gives us for \( 1 \leq l \leq r < n \) that

\[
S_n(l, r) = \int_{l/n}^{r/n} \sqrt{n} \int_{l}^{G_n(t)} J(s) ds dg(t) / \sigma_n \\
- \int_{\xi_n}^{1/n} \sqrt{n} \int_{l/n}^{G_n(t)} J(s) ds dg(t) / \sigma_n + \int_{r/n}^{G_n(t)} \int_{l/n}^{G_n(t)} J(s) ds dg(t) / \sigma_n
\]

(51) \( \equiv \theta_n[l, r] + \gamma_n(l) - \gamma_n(r) \).

Suppose now that

(52) \( k_n \wedge (n - m_n) \to \infty, k_n/n \to 0 \) and \( m_n/n \to 1 \).

An easy Taylor series expansion shows that

\[
\theta_n[k_n, m_n] \approx \int_{k_n/n}^{m_n/n} \alpha_n J dg / \sigma_n \\
\approx \int_{k_n/n}^{m_n/n} B_n dK / \sigma_n \quad \text{using (41)}
\]

(53) \( \equiv N(0, 1) \).

Likewise, letting \( x = \sqrt{n(t - k_n/n)/\sqrt{k_n/n}} \),

\[
\gamma_n(k_n) \approx \int_{U_{n, n}}^{k_n/n} \sqrt{n}[G_n(t) - t] dK / \sigma_n \\
\approx - \int_{\sqrt{n} \xi_n}^{\sqrt{n} \xi_n} \left[ \frac{\alpha_n \left( \frac{k_n}{n} \left( 1 + \frac{x}{\sqrt{k_n/n}} \right) \right)}{\sqrt{k_n/n}} + x \right] d\Psi_n(x)
\]

where

(54) \( \Psi_n(x) \equiv \sqrt{k_n/n} \left[ K \left( \frac{k_n}{n} \right) + x \sqrt{k_n/n} \right] - K \left( \frac{k_n}{n} \right) / \sigma_n \)

21
for $|x| \leq \sqrt{k_n}/2$. Thus

$$\alpha_n(k_n) \approx - \int_{-B_n(k_n/n)/\sqrt{k_n/n}}^0 \left[ \frac{B_n(k_n/n)}{\sqrt{k_n/n}} + x \right] d\Psi_0(x)$$

$$= - \int_{-Z_0}^0 (Z_0 + x) d\Psi_0(x)$$

(55)

$$\to_d \int_{-Z_1}^0 [Z_1 + x] d\Psi(x) \text{ if } \Psi_{0n} \to \Psi_0,$$

where $\to_d$ denote convergence in distribution of generalized df's.

Let $\Psi_{1n}$ be analogous to $\Psi_{0n}$, but for the right tail. Then

$$\gamma_n(m_n) \approx \ldots \approx - \int_{-Z_1}^0 [Z_1 + x] d\Psi_{1n}(x)$$

(56)

$$\to_d \int_{-Z_1}^0 [Z_1 + x] d\Psi_1(x) \text{ if } \Psi_{1n} \to \Psi_1.$$  

Since $[Z_0 + x]$ is of one sign on $[0, -Z_0)$, (55) suggests that $\gamma_n(k_n)$ blows up if $\lim|\Psi_{0n}(x)| = +\infty$ for any $x$; likewise $\gamma_n(m_n)$ blows up if $\lim|\Psi_{1n}(x)| = +\infty$ for any $x$. If both $\lim$'s are finite for all $x$, then from any subsequence $n'$ we can find a further subsequence $n''$ for which $\Psi_{in} \to \Psi_i$ (some $\Psi_i$) for $i = 0, 1$.

We then expect convergence in distribution to the rv that is the sum of (53), (55) and (56). This latter sum seems clearly to be normal if and only if $\Psi_1 = \Psi_2 = 0$. This suggests that we have

(57)  

$$S_n(k_n/n, m_n/n) \to_d N(0, 1) \text{ under (52)}$$

if and only if

(58)  

$$\Psi_{in}(x) \to 0 \text{ as } n \to \infty \text{ for all } x \text{ and } i = 0, 1.$$
Moreover, we seem to have

\[(59) \quad S_n(k_n/n, m_n/n) \text{ is stochastically compact under (52)}\]

if and only if

\[(60) \quad \lim_{n \to \infty} |\Psi_{i n}(x)| < \infty \text{ for all } x \text{ and } i = 0, 1,\]

with all possible subsequential limits expressible as a sum of (53), (55) and (56). Indeed, these suggested results for the trimmed mean $S_n$ of (50) were rigorously established by S. Csörgö, Haeussler and Mason (1988a).

The CsCsHIM constructions were also strong enough for S. Csörgö, Haeussler and Mason (1988, 1989) to establish necessary and sufficient conditions for asymptotic normality, and stochastic compactness, with a represenation of all possible subsequential limits in case either

\[(61) \quad \sqrt{n}(k_n/n - a) \to 0 \text{ and } \sqrt{n}(m_n/n - b) \to 0 \text{ with } 0 < a < b < 1\]

or

\[(62) \quad k_n = k \text{ and } n - m_n = m \text{ for all } n.\]

In case (61), these are the results of Stigler (1973). In case (62), the asymptotic normality result included and generalized the classical results on the domain of attraction of the stable distributions. In fact a multitude of results on domains of partially attraction, and other generalizations, were established by these methods. S. Csörgö, Horváth and Mason (1986) considered the case of convergence to stable laws in great detail.
A key result in much of this work was the discovery that

\begin{equation}
\limsup_{a \to 0, b \to 1} \frac{a[F^{-1}(a)]^2 + b[F^{-1}(b)]^2}{\sigma^2[a, b]} \leq 1 \text{ for all } F^{-1}.
\end{equation}

Consider now the extension of (57)-(58) and (59)-(60) from trimmed means to trimmed \( L \)-statistics under (52). The heuristics carry through from earlier as soon as we introduce the device of replacing \( G_n(t) \) by \( G_n^*(t) \), defined to be \( k/n, G_n(t), 1 - m/n \) according as \( t \leq U_{k,n}, U_{k,n} \leq t \leq U_{n-m,n}, U_{n-m,n} \leq t \); this alternative definition prevents certain terms from assuming an infinite value. The technicalities also require some regularity, the most notable condition being that \( J(s) \) and \( J(1-s) \) are regularly varying at 0. See Mason and Shorack (1988) for this, and the following. Under (61), there is always a distributional limit provided \( J \) is Lipschitz, and this limit is normal if and only if \( \Delta K(a) = \Delta K(b) = 0 \). Under (62), the \( L \)-statistic \( T_n \) is asymptotically normal with \( \sqrt{n} \) norming if and only if \( Y \equiv K(U) \) has finite variance and it is asymptotically normal with arbitrary norming if and only if \( Y \) is in the domain of attraction of a normal law. Stochastic compactness is also characterized. This solved a long outstanding problem in \( L \)-statistics for a large class of smooth \( J \) functions.

Quite a number of other related topics are treated in Shorack and Wellner (1986).

REFERENCES


15. **Csörgő, S., Haeussler, E. and Mason, D.** (1988b). A probabilistic approach to the asymptotic distribution of sums of indepen-
dent, identically distributed random variables. *Adv. in Appl. Math.*, 9
259-333.

of the sample makes a partial sum asymptotically stable or normal?
*Prob. Theory Relat. Fields* 72 1-16.


19. DOOB, J. (1949). Heuristic approach to the Kolmogorov-Smirnov theo-

20. ERDÖS, P. AND KAC, M. (1946). On certain limit theorems of the


Press, New York.

23. KIEFER, J. (1972). Skorokhod embedding of multivariate rv’s, and the

27


DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195