COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 951

A FUNCTIONAL CENTRAL LIMIT THEOREM FOR STRONG MIXING STOCHASTIC PROCESSES

by

Donald W. K. Andrews and David Pollard

August 1990

A FUNCTIONAL CENTRAL LIMIT THEOREM FOR STRONG MIXING STOCHASTIC PROCESSES

by

Donald W. K. Andrews
Department of Economics
Yale University
Box 2125 Yale Station
New Haven, CT 06520

David Pollard
Department of Statistics
Yale University
Box 2179 Yale Station
New Haven, CT 06520

July 1990

ABSTRACT. This paper shows how the modern machinery for generating abstract empirical central limit theorems can be applied to arrays of dependent variables. It develops a bracketing approximation based on a moment inequality for sums of strong mixing arrays, in an effort to illustrate the sorts of difficulty that need to be overcome when adapting the empirical process theory for independent variables. Some suggestions for further development are offered. The paper is largely self-contained.

Keywords: strong mixing, functional central limit theorem, empirical process

§1 Introduction

Since the landmark paper of Dudley (1978), there have been many generalizations to abstract settings of Donsker's theorem for the empirical distribution function. For the most part, the generalized theory has treated empirical processes for independent summands. Exceptions have been Leventhal's (1988, 1989) work on regenerative processes and martingale difference arrays, and the general functional limit theorem of Andrews (1989a). The closely related theory for sums of random elements in Banach space (as in Dehling (1983), for example) and the work of Goldie and Greenwood (1986a, b) on set-indexed partial-sum processes do not translate easily into usable results for empirical processes.

The concentration on independent summands contrasts with the development of the theory in the one-dimensional case, where results for various types of dependence were discovered early. For example, Chapter 4 of Billingsley's (1968) influential book treated ϕ-mixing sequences.

In this paper we present an empirical central limit theorem for strong mixing triangular arrays of random processes. We chose to work with strong mixing, rather than with the more restrictive ϕ-mixing, because of the wider range of possible applications. We embarked upon the work leading to our theorem in response to the often posed question, How much of the theory for independent processes carries over to the dependent case? Some subtle difficulties made the task less straightforward than expected. As a guide to others who might want to extend empirical process theory to different types of dependent variables, we present in Section 5 an outline of general principles and a discussion of why certain plausible methods failed in our setting.

The formal statement of our limit theorem appears in Section 2. Our results apply to an empirical process ν_n indexed by a class of functions $\mathcal{F}$,

$$\nu_n f = \frac{1}{\sqrt{n}} \sum_{i \le n} \Big(f(\xi_{ni}) - \boldsymbol{P} f(\xi_{ni}) \Big),$$

where $\{\xi_{ni} : i \le n, n = 1, 2, \dots\}$ is a strong mixing triangular array. The empirical central limit theorem (Corollary 2.4) gives conditions under which ν_n converges in distribution, as a stochastic process indexed by $\mathcal{F}$, to an appropriate Gaussian process. The proof of such a theorem consists of the usual two steps: establish convergence of finite-dimensional distributions; then establish stochastic equicontinuity, a close relative of the familiar uniform tightness property. The literature already contains several results that can handle finite-dimensional convergence. Our Theorem 2.3 gives sufficient conditions for a convenient strengthening of stochastic equicontinuity. With an appropriate seminorm $\rho(\cdot)$ on $\mathcal{F}$

and an appropriate $\mathcal{L}^Q$ norm, it gives for each $\epsilon > 0$ a $\delta > 0$ such that

$$\limsup_{n\to\infty} \| \sup_{\rho(f-g)<\delta} |\nu_n f - \nu_n g| \|_Q < \epsilon.$$

The proof depends upon a moment inequality (Lemma 3.1) for sums of strong mixing sequences.

Stochastic equicontinuity is a most useful property even apart from its role in the functional central limit theorem. It implies that $|\nu_n f_n - \nu_n g_n| \to 0$ in probability for all sequences $\{f_n, g_n\}$, possibly random, from $\mathcal{F}$ such that $\rho(f_n, g_n) \to 0$ in probability. Andrews (1989b, c) has shown how this form of stochastic equicontinuity is the key to many semiparametric limit theorems; it was also the main hypothesis in the general central limit theorem for minimization estimators, established in Section VII.1 of Pollard (1984). We present a typical application in Section 4. By establishing conditions for stochastic equicontinuity under strong mixing assumptions, we automatically extend the range of application of all those results.

§2 Definitions and statement of results

Let $\{\xi_{ni} : i \leq n,\, n = 1, 2, \dots\}$ be a triangular array of random elements of a measurable space S. Define $\mathcal{A}_n(m)$ to be the σ-field generated by the variables ξ_{ni} for $i \leq m$ and $\mathcal{B}_n(m+d)$ to be the σ-field generated by the variables ξ_{ni} for $i \geq m+d$. We say that $\{\xi_{ni}\}$ is strong mixing if there is a sequence of numbers $\{\alpha(d)\}$ converging to zero for which

$$|\mathbb{P}AB - \mathbb{P}A\mathbb{P}B| \leq \alpha(d) \qquad \text{for all } A \in \mathcal{A}_n(m), \text{ all } B \in \mathcal{B}_n(m+d), \text{ all } m,\, d,\, n.$$

We define a uniform analogue of the $\mathcal{L}^2$ norm by $\rho(f) = \sup_{n,i} \|f(\xi_{ni})\|_2$. [In general, we write $\|Z\|_p$ for the $\mathcal{L}^p$ norm $(\mathbb{P}|Z|^p)^{1/p}$ of a random variable Z.]

Results for triangular arrays are more powerful than their analogues for a single strong mixing sequence. For example, local power calculations and asymptotic minimax theorems require triangular arrays.

Our main result is a maximal inequality for the empirical process indexed by a class of functions $\mathcal{F}$, with a bound involving a measure of complexity for $\mathcal{F}$ based on the concept of bracketing.

(2.1) **Definition:** The bracketing number $N(\delta) = N(\delta, \mathcal{F})$ equals the smallest value of N for which there exist functions $f_1, \dots, f_N$ in $\mathcal{F}$ and $b_1, \dots, b_N$ with $\rho(b_i) \leq \delta$ for each i such that: for each f in $\mathcal{F}$ there exists an i for which $|f - f_i| \leq b_i$.

Useful bounds on bracketing numbers can be obtained, for example, if $\mathcal{F}$ is a parametric family, $\mathcal{F} = \{f(\cdot,\theta) : \theta \in \Theta\}$, with Θ a bounded subset of some Euclidean space $\mathbb{R}^k$ and the functions subject to the condition: for some constants $C < \infty$ and $\lambda > 0$, and all r small enough,

$$\text{(2.2)} \qquad \sup_{n,i} \mathbb{P} \sup_{B(\theta,r)} |f(\xi_{ni},\theta') - f(\xi_{ni},\theta)|^2 \le C^2 r^{2\lambda} \qquad \text{for all } \theta$$

where $B(\theta,r)$ is the ball of radius r around θ. For example, such an inequality would follow from a Lipschitz condition,

$$|f(x,\theta) - f(x,\theta')| \le L(x)|\theta - \theta'|^\lambda,$$

if $\sup_{n,i} \|L(\xi_{ni})\|_2 = C < \infty$. If (2.2) holds, one takes the f_i in Definition 2.1 to correspond to the centers of the $O(r^k)$ many balls of radius $r = (\delta/C)^{1/\lambda}$ that are needed to cover the bounded set Θ. This gives bracketing numbers of order $O(\delta^{-k/\lambda})$, which is the sort of geometric bound needed for our theorem.

(2.3) **Theorem:** *Let $\{\xi_{ni}\}$ be a strong mixing triangular array whose mixing coefficients satisfy*

$$\sum_{d=1}^{\infty} d^{Q-2}\alpha(d)^{\gamma/(Q+\gamma)} < \infty$$

for some even integer $Q \ge 2$ and some $\gamma > 0$, and let $\mathcal{F}$ be a uniformly bounded class of real-valued functions whose bracketing numbers satisfy

$$\int_0^1 x^{-\gamma/(2+\gamma)} N(x)^{1/Q} dx < \infty$$

for the same Q and γ. Then for each $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\limsup_{n\to\infty} \| \sup_{\rho(f-g)<\delta} |\nu_n f - \nu_n g| \|_Q < \epsilon.$$

(2.4) **Corollary (Functional Central Limit Theorem):** *If the conditions of Theorem 2.3 are satisfied and if $(\nu_n f_1, \ldots, \nu_n f_k)$ has an asymptotic normal distribution for all choices of $f_1, \ldots, f_k$ from $\mathcal{F}$, then $\{\nu_n f : f \in \mathcal{F}\}$ converges in distribution to a Gaussian process indexed by $\mathcal{F}$ with ρ-continuous sample paths.*

The formal meaning of the Corollary and the general concept of convergence in distribution are explained in Sections 9 and 10 of Pollard (1990)—see Theorem 10.2 in particular.

The conditions of Theorem 2.3 require a balance between the rate of decrease in the mixing coefficients and the rate of growth in the bracketing numbers. For example, if

$N(x) = O(x^{-\beta})$ and $\alpha(d) = O(d^{-A})$ for some $\beta > 0$ and $A > 0$, then the requirements would be satisfied with Q equal to the smallest even integer greater than 2β and $\gamma = 2$, if $A > (Q-1)(1+Q/2)$. (These are not the best choices possible.) We have required Q to be an even integer merely to simplify the calculations in the Appendix. It is possible that the condition could be relaxed, to allow fractional Q, at the cost of a more delicate argument analogous to that of Yokoyama (1980). We have not included explicit conditions for the finite-dimensional convergence as part of the Corollary, because there many possiblities (Philipp 1969; McLeish 1977, Corollary 2.11; Herrndorf 1984, Corollary 1).

§3 Proof of Theorem 2.3

The proof depends upon a moment inequality applied to the increments $\nu_n f - \nu_n g$ of the empirical process. For independent summands the inequality is well known. For strong mixing arrays it extends results of Sen (1974) and Yokoyama (1980). It corresponds to Theorem 4 of Doukhan and Portal (1984) and Theorem 10 of Doukhan and Portal (1983). Because the last two papers offer only sketch proofs, and because typographical errors make the statement of their inequalities slightly confusing, we give a complete, self-contained proof of our inequality in the Appendix.

(3.1) Lemma: *Let $Z(1), Z(2), \ldots$ be a strong mixing sequence of random variables with mixing coefficients $\{\alpha(d)\}$. For some $\tau > 0$, $\gamma > 0$, and even integer $Q \geq 2$, suppose:*

(i) $|Z(i)| \leq 1$, $\mathbb{P}Z(i) = 0$, *and* $\mathbb{P}Z(i)^2 \leq \tau^{2+\gamma}$ *for every* i;

(ii) $\sum_{d=1}^{\infty} d^{Q-2}\alpha(d)^{\gamma/(Q+\gamma)} < \infty$.

Then

$$\mathbb{P}\Big|\sum_{i=1}^{n} Z(i)\Big|^Q \leq C\left((n\tau^2) + \cdots + (n\tau^2)^{Q/2}\right) \qquad \text{for all } n,$$

for some constant C that depends only on Q and the mixing coefficients. □

For empirical processes it will be most convenient to work with a new seminorm, $\tau(h) = \rho(h)^{2/(2+\gamma)}$, because then

$$\mathbb{P}|h(\xi_{ni}) - \mathbb{P}h(\xi_{ni})|^2 \leq \tau(h)^{2+\gamma} \qquad \text{for every } i \text{ and } n.$$

Without loss of generality we assume that $0 \leq f \leq 1/2$ for every f in $\mathcal{F}$. All the bracketing bounds b_i and all the differences $|f - f_i|$ that appear in Definition 2.1 can then be assumed less than $1/2$, and the moment bound from Lemma 3.1 will apply directly without the intrusion of extra scaling constants. Indeed, if $|h| \leq 1/2$ we may apply the lemma for each fixed n to the random variables $Z(i) = h(\xi_{ni}) - \mathbb{P}h(\xi_{ni})$ to get

$$\mathbb{P}|\nu_n h|^Q \leq n^{-Q/2} C \left((n\tau^2) + \cdots + (n\tau^2)^{Q/2} \right) \qquad \text{where } \tau = \tau(h).$$

When $\tau(h)$ is bigger than $1/\sqrt{n}$ the $(n\tau^2)^{Q/2}$ term dominates. Putting $C' = (CQ/2)^{1/Q}$ we deduce

$$\|\nu_n h\|_Q \leq C' \max\left(\frac{1}{\sqrt{n}}, \tau(h)\right). \tag{3.2}$$

We must be precise here with the form of the inequality, because usually we will assume only an upper bound for $\tau(h)$; the actual value might not satisfy the inequality $n\tau(h)^2 \geq 1$.

For $k = 1, 2, \ldots$ invoke the definition of bracketing numbers with $\delta = 2^{-k}$ to find approximating subclasses $\mathcal{F}_k$ and their corresponding bounding classes $\mathcal{B}_k$ with the property that to each f in $\mathcal{F}$ there is an f_k in $\mathcal{F}_k$ and a corresponding b_k in $\mathcal{B}_k$ for which $|f - f_k| \leq b_k$ and $\rho(b_k) \leq 2^{-k}$. If we define $\tau_k = 2^{-2k/(2+\gamma)}$ then $\tau(b_k) \leq \tau_k$. The class $\mathcal{F}_k$ need contain no more than $N_k = N(2^{-k}, \mathcal{F})$ functions.

In outline, the proof of the Theorem goes as follows. We first show that $\nu_n f$ is uniformly well approximated by $\nu_n f_{k(n)}$ if $k(n)$ diverges to infinity at a suitable rate. More precisely, we will choose $k(n)$ to ensure that

$$\|\sup_{\mathcal{F}} |\nu_n f - \nu_n f_{k(n)}|\,\|_Q < 2\epsilon \qquad \text{eventually.} \tag{3.3}$$

We then apply a chaining argument to show that, for some fixed m and n large enough, $\nu_n f_{k(n)}$ is uniformly well approximated by $\nu_n f_m$ for some $f_m \in \mathcal{F}_m$, in the sense that

$$\|\max_{f \in \mathcal{F}} |\nu_n f_{k(n)} - \nu_n f_m|\,\|_Q < 2\epsilon \qquad \text{eventually.} \tag{3.4}$$

Here we write max instead of sup, to emphasize that $f_{k(n)}$ and f_m range over only finitely many functions, even when $\mathcal{F}$ is infinite. The choice of f_m depends on n, but that does not disturb the subsequent steps in the proof. Together inequalities (3.3) and (3.4) imply

$$\|\sup_{\mathcal{F}} |\nu_n f - \nu_n f_m|\,\|_Q < 4\epsilon \qquad \text{eventually.} \tag{3.5}$$

Finally, using (3.5) and a subtle argument from Ledoux and Talagrand (1990), we reduce the comparisons between pairs f, g from $\mathcal{F}$ to comparisons between at most N_m^2 pairs.

Throughout the proof we rely on a simple maximal inequality due to Pisier (1983): for random variables $Z_1, \ldots, Z_N$,

$$\|\max_{i \leq N} |Z_i|\,\|_Q \leq N^{1/Q} \max_{i \leq N} \|Z_i\|_Q,$$

which is a consequence of the trivial bound

$$\mathbb{P}\max_{i\le N}|Z_i|^Q \le \sum_{i\le N}\mathbb{P}|Z_i|^Q.$$

When specialized to the empirical process evaluated at functions $h_1,\dots,h_N$ with $|h_i|\le 1/2$, Pisier's inequality together with the bound (3.2) gives

$$\|\max_{i\le N}|\nu_n h_i|\,\|_Q \le C'N^{1/Q}\max\left(\frac{1}{\sqrt{n}},\max_{i\le N}\tau(h_i)\right). \tag{3.6}$$

For all applications, the $1/\sqrt{n}$ term will be the smaller of the two terms in the max.

Proof of inequality (3.3). Let $k(n)$ be the smallest value of k for which $2\sqrt{n}2^{-k}\le\epsilon$. For each f in $\mathfrak{F}$,

$$\begin{aligned}|\nu_n f-\nu_n f_{k(n)}| &\le \frac{1}{\sqrt{n}}\sum_{i\le n}\Big(b_{k(n)}(\xi_{ni})+\mathbb{P}b_{k(n)}(\xi_{ni})\Big)\\ &\le \nu_n b_{k(n)}+\frac{2}{\sqrt{n}}\sum_{i\le n}\mathbb{P}b_{k(n)}(\xi_{ni}).\end{aligned}$$

The last sum is less than $2\sqrt{n}\rho(b_{k(n)})$, which is less than ϵ by the choice of $k(n)$. The defining inequality for $k(n)$ also ensures that

$$\tau_{k(n)}\ge\left(\epsilon/4\sqrt{n}\right)^{2/(2+\gamma)},$$

which is greater than $1/\sqrt{n}$ for n large enough, say $n\ge n(\epsilon)$. Invoking inequality (3.6), we deduce that

$$\begin{aligned}\|\sup_{\mathfrak{F}}|\nu_n f-\nu_n f_{k(n)}|\,\|_Q &\le \|\max_{b\in\mathfrak{B}_{k(n)}}|\nu_n b|\,\|_Q+\epsilon\\ &\le C'N_{k(n)}^{1/Q}\tau_{k(n)}+\epsilon \qquad \text{for } n\ge n(\epsilon)\\ &\le C'\int_0^{2^{-k(n)}}x^{-\gamma/(2+\gamma)}N(x,\mathfrak{F})^{1/Q}\,dx+\epsilon.\end{aligned}$$

The integral condition of the Theorem ensures that the last bound is eventually less than 2ϵ, as asserted by (3.3).

Proof of inequality (3.4). The integer m will soon be fixed at a value depending only on ϵ. Eventually $k(n)$ will be larger than m. To bridge the gap between m and $k(n)$ we argue recursively, relating the approximation via $\mathfrak{F}_k$ to the cruder approximation via $\mathfrak{F}_{k-1}$, for $k=m+1,\dots,k(n)$. A subtle difficulty now arises. If f and f' are functions in $\mathfrak{F}$ for which $f_k=f'_k$, there is no guarantee that $f_{k-1}=f'_{k-1}$. Potentially f_k-f_{k-1} could range over as many as $N_k N_{k-1}$ differences as f ranges over $\mathcal{F}$. To reduce the number of differences to N_k we recycle notation by redefining f_{k-1} inductively, for $k < k(n)$, to be the function from $\mathcal{F}_{k-1}$ that best approximates the function f_k in $\mathcal{F}_k$, in the sense of the τ distance. Certainly

$$\tau(f_k - f_{k-1}) \leq \tau_{k-1},$$

and $\tau_{k-1} \geq 1/\sqrt{n}$ if $k \leq k(n)$ and $n \geq n(\epsilon)$. Invoking inequality (3.6) again we get

$$\| \max_{f \in \mathcal{F}} |\nu_n f_k - \nu_n f_{k-1}| \|_Q \leq C' N_k^{1/Q} \tau_{k-1}.$$

As before, the max emphasizes that the differences range over only finitely many functions, N_k of them, as f ranges over $\mathcal{F}$. It follows that, for n large enough,

$$\begin{aligned} \| \max_{f \in \mathcal{F}} |\nu_n f_{k(n)} - \nu_n f_m| \|_Q &\leq \sum_{k=m+1}^{k(n)} \| \max_{f \in \mathcal{F}} |\nu_n f_k - \nu_n f_{k-1}| \|_Q \\ &\leq \sum_{k=m+1}^{\infty} C' N_k^{1/Q} \tau_{k-1} \\ &= \sum_{k=m+1}^{\infty} C' (2^{-k+1})^{2/(2+\gamma)} N(2^{-k}, \mathcal{F})^{1/Q}. \end{aligned}$$

For some constant C_γ, the sum is bounded by

$$C_\gamma \int_0^{\tau_m} x^{-\gamma/(2+\gamma)} N(x, \mathcal{F})^{1/Q} dx.$$

With m fixed so that the last bound is less than 2ϵ, we have (3.4). Notice that f_m depends on n, because it is the last function in a chain leading from $f_{k(n)}$.

Comparison of pairs. Define an equivalence relation on $\mathcal{F}$ by: $f \sim f'$ if $f_m = f'_m$. The relation serves to partition $\mathcal{F}$ into N_m equivalence classes $\mathcal{E}[1], \ldots, \mathcal{E}[N_m]$. (The partition actually depends on n, because of the way f_m depends on $f_{k(n)}$.) From (3.5) applied twice,

$$\| \sup_{f \sim f'} |\nu_n f - \nu_n f'| \|_Q \leq 8\epsilon \qquad \text{eventually.} \tag{3.7}$$

Define a distance between the classes by

$$d(\mathcal{E}[i], \mathcal{E}[j]) = \inf\{\rho(f - f') : f \in \mathcal{E}[i], f' \in \mathcal{E}[j]\}.$$

For a fixed $\delta > 0$ choose functions ϕ_{ij} in $\mathcal{E}[i]$ and ϕ_{ji} in $\mathcal{E}[j]$ such that

$$\rho(\phi_{ij} - \phi_{ji}) < d(\mathcal{E}[i], \mathcal{E}[j]) + \delta.$$

If $f \in \mathcal{E}[i]$ and $f' \in \mathcal{E}[j]$ and $\rho(f - f') < \delta$, then $\rho(\phi_{ij} - \phi_{ji}) < 2\delta$ and

$$|\nu_n f - \nu_n f'| \leq 2 \sup_{g \sim g'} |\nu_n g - \nu_n g'| + \max\{|\nu_n \phi_{ij} - \nu_n \phi_{ji}| : \rho(\phi_{ij} - \phi_{ji}) < 2\delta\}.$$

Notice that the last maximum runs over at most N_m^2 pairs. Taking norms of both sides, we deduce via (3.6) and (3.7) that

$$\| \sup_{\rho(f-f')<\delta} |\nu_n f - \nu_n f'| \|_Q < 16\epsilon + C' N_m^{2/Q} (2\delta)^{2/(2+\gamma)} \qquad \text{eventually.}$$

We have already fixed the value of m. We can therefore choose δ small enough to make the right-hand side less than 17ϵ. □

§4 An application of stochastic equicontinuity

We will sketch a typical example of how stochastic equicontinuity can be used to simplify asymptotic arguments.

Suppose $\mathcal{F} = \{f(\cdot, \theta) : \theta \in \Theta\}$ is a class of $\mathbb{R}^k$-valued functions indexed by a subset of $\mathbb{R}^k$. The ρ seminorm defines a new distance on Θ by

$$d(\theta, \theta') = \rho\left(f(\cdot, \theta) - f(\cdot, \theta')\right).$$

An m-estimator $\hat{\theta}_n$ might be chosen to make the random function

$$F_n(\theta) = \frac{1}{n} \sum_{i \leq n} f(\xi_{ni}, \theta)$$

close to zero, in the sense that

$$F_n(\hat{\theta}_n) = o_p(1/\sqrt{n}). \tag{4.1}$$

The true θ_0 might be identified as the root of the corresponding expected value,

$$M_n(\theta) = \mathbb{P} F_n(\theta) = \frac{1}{n} \sum_{i \leq n} \mathbb{P} f(\xi_{ni}, \theta),$$

in the sense that

$$M_n(\theta_0) = 0 \qquad \text{for all } n.$$

With a preliminary argument (often based on a uniform law of large numbers) one might be able to establish consistency, $\hat{\theta}_n \to \theta_0$ in probability. With mild continuity and domination conditions on the $f(\cdot, \theta)$ functions, this can usually be reinterpreted as

$$d(\hat{\theta}_n, \theta_0) \to 0 \qquad \text{in probability.} \tag{4.2}$$

With such preliminaries, asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ can then be deduced from the following three requirements on the processes.

(i) Uniform differentiability of the M_n functions at θ_0: for some nonsingular matrix D,

$$M_n(\theta) = D(\theta - \theta_0) + o(|\theta - \theta_0|),$$

with the $o(\cdot)$ term uniform in n. Notice that this requirement is weaker than pointwise differentiability of the $f(x, \cdot)$ functions, which can be useful in such applications as least absolute deviations estimation.

(ii) Asymptotic normality of $\nu_n f(\cdot, \theta_0)$.

(iii) Stochastic equicontinuity of ν_n at θ_0: for each $\epsilon > 0$ and $\eta > 0$ there exists a $\delta > 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\left\{ \sup_{d(\theta,\theta_0)<\delta} |\nu_n f(\cdot,\theta) - \nu_n f(\cdot,\theta_0)| > \eta \right\} < \epsilon.$$

When reduced to assertions about each of the components of the vector processes, requirement (iii) is weaker than the stochastic equicontinuity property delivered by Theorem 2.3. Together with (4.2) it implies that

$$\nu_n f(\cdot, \hat\theta_n) = \nu_n f(\cdot, \theta_0) + o_p(1). \tag{4.3}$$

In addition, from (i) we get

$$M_n(\hat\theta_n) = D(\hat\theta_n - \theta_0) + o(|\hat\theta_n - \theta_0|). \tag{4.4}$$

Substitution into (4.1) then gives

$$\begin{aligned} o_p(1/\sqrt{n}) &= F_n(\hat\theta_n) \\ &= M_n(\hat\theta_n) + \frac{1}{\sqrt{n}}\nu_n f(\cdot, \hat\theta_n) \\ &= D(\hat\theta_n - \theta_0) + o(|\hat\theta_n - \theta_0|) + \frac{1}{\sqrt{n}}\nu_n f(\cdot,\theta_0) + o_p(1/\sqrt{n}). \end{aligned} \tag{4.5}$$

First deduce from the nonsingularity of D that $|\hat\theta_n - \theta_0| = O_p(1/\sqrt{n})$:

$$\begin{aligned} \left|\frac{1}{\sqrt{n}}\nu_n f(\cdot,\theta_0) + o_p(1/\sqrt{n})\right| &= \left|D(\hat\theta_n - \theta_0) + o(|\hat\theta_n - \theta_0|)\right| \qquad \text{from (4.5)} \\ &\geq (\kappa - o_p(1))|\hat\theta_n - \theta_0| \end{aligned}$$

for some positive constant κ. The left-hand side is of order $O_p(1/\sqrt{n})$ by (ii). Next consolidate the error terms in (4.5) to get

$$\sqrt{n}(\hat\theta_n - \theta_0) = D^{-1}\nu_n f(\cdot,\theta_0) + o_p(1).$$

The random vector on the right-hand side has an asymptotic normal distribution.

§5 Some strategies

The main lesson that we learned from our efforts to develop empirical central limit theorems for dependent variables was: everything depends on the existence of good probabilistic bounds for the increments of the empirical process. Let us explain.

There have been two major lines of development in the literature on abstract empirical central limit theorems. One line, which starts from a symmetrization argument, has evolved from the method of Vapnik and Červonenkis (1971). It depends on the following two requirements.

(i) In the sense of either tail probability bounds or moment bounds, the quantity $\sup_{\mathcal{F}} |\nu_n f|$ is less variable than

$$\sup_{\mathcal{F}} \frac{1}{\sqrt{n}} \left| \sum_{i \leq n} \left(f(\xi_{ni}) - f(\xi'_{ni}) \right) \right|. \tag{5.1}$$

The new variables $\{\xi'_{ni}\}$ are typically an independent copy of the $\{\xi_{ni}\}$.

(ii) Conditional on certain information symmetric in both $\{\xi_{ni}\}$ and $\{\xi'_{ni}\}$, the process of sums appearing in (5.1) has a tractable distribution.

Requirement (ii) is analogous to the property that justifies the calculations with permutation distributions for experimental designs with randomization, as in Box and Andersen (1955). In practice it has required independence of $\{\xi_{ni}\}$ from $\{\xi'_{ni}\}$. For then the quantity in (5.1) has the same distribution as

$$\sup_{\mathcal{F}} \frac{1}{\sqrt{n}} \left| \sum_{i \leq n} \sigma_i \left(f(\xi_{ni}) - f(\xi'_{ni}) \right) \right|, \tag{5.2}$$

where the $\{\sigma_i\}$ are independent Rademacher variables ($+1$ and -1, each with probability $1/2$). The conditional distribution, given $\{\xi_{ni}\}$ and $\{\xi'_{ni}\}$, of this expression is amenable to various types of chaining argument (cf. Section 3), because good bounds exist for the increments of the underlying process: the Hoeffding inequalities give exponential bounds on tail probabilities and moments up to exponential order. It is largely a matter of taste whether one applies the chaining argument to moment quantities (as in Pollard 1990) or to tail probabilities (as in Pollard 1984). The moment bounds require slightly less machinery (one fewer sequence of constants to adjust correctly), at the slight cost of results not quite as refined as those for tail probabailities (Alexander 1984, Massart 1986). For other related applications, however, such as the U-processes of Nolan and Pollard (1987, 1988), only moment bounds seem to work.

Dependence between the $\{\xi_{ni}\}$ variables severely complicates the second requirement. Leventhal (1988) was able to extend the method to regenerative processes, in which excursions between renewal times are independent, by replacing the $\{\xi_{ni}\}$ by the whole excursion; he symmetrized over whole blocks of variables. Further progress with symmetrization applied to dependent variables seems unlikely, except in special cases that allow reduction to independence.

The second major line of development of abstract empirical central limit theorems has involved the use of bracketing arguments. These appear more promising for dependent variables because they work directly with the empirical process. Again one needs some sort of probabilistic bounds for the increments of the process in order to invoke a chaining argument. For the classical one-dimensional empirical central limit theorem in the independent case, a Tchebychev bound based on fourth moments of the binomial distribution suffices— for example, see pages 262-266 of Parthasarathy (1967).

For his abstract empirical central limit theorem (for independent summands) under a bracketing condition, Dudley (1978, 1981) applied the Bernstein exponential tail bound to the increments of the empirical process indexed by classes of sets and uniformly bounded classes of functions. The Bernstein inequality requires bounded summands. Ossiander (1987) combined the same bound with a delicate truncation argument to remove the boundedness assumption from the class of functions. Pollard (1991) has shown that a first moment version of Ossiander's method slightly simplifies the argument.

The main difficulty in Dudley's form of the bracketing method arises because the Bernstein bounds lose their power for increments with small variance—compare with the $1/\sqrt{n}$ that appears on the right-hand side of (3.6). A separate argument is needed to handle the contributons from such increments, as in our proof of (3.3).

Analogues of the Bernstein bound do exist for dependent summands, such as martingale difference arrays. Leventhal (1989) invoked such a bound, but was left with an unpalatable uniformity assumption involving the small increments. For ϕ-mixing sequences Collomb's (1984) Lemma 4.1 would support a chaining argument—Yukich (1986) has applied it with a simpler form of bracketing argument to derive uniform rates of convergence, but not for functional central limit theorems. For strong mixing arrays modified forms of Carbon's (1983) inequality (I) would seem to offer some promise. As it stands, that inequality is not suitable because it does not take advantage of small variances of the summands—the factor $\tilde{\alpha}(k)$ in the coefficient of α^2 does not decrease with D. By modifying the argument leading to the second moment bound at the top of Collomb's page 451, we were able to replace Carbon's $\tilde{\alpha}(k)$ by a factor involving the square root of an $\mathcal{L}_1$ norm. Unfortunately, we were not able to find an appropriate substitute for the last term in his exponential bound; we still do not have an appropriate exponential bound that could support a chaining argument for strong mixing arrays.

Without an adequate exponential bound for tail probabilities of increments, the Ossiander truncation method fails. That forces one to impose uniform boundedness on the class of functions, a requirement that limits the possible applications of the empirical central limit theorem. The basic chaining argument, however, can still be invoked even with much weaker control over the increments of the process, as in our Section 3, where moment bounds handled geometric rates of increase in bracketing numbers.

In summary, for strong mixing arrays we were only able to find moment bounds on the increments of the empirical process, which placed a constraint on the possible rate of increase in the bracketing numbers. With a better bound for the increments we might have a obtained a better theorem. That is the obvious place to start for anyone seeking to improve upon our results.

Appendix: Proof of Lemma 3.1

We will make repeated use of the following standard strong mixing inequality for random variables (Hall and Heyde 1980, Corollary A.2). For fixed n, m, and d, suppose X is $\mathcal{A}_n(m)$-measurable, Y is $\mathcal{B}_n(m+d)$-measurable. Let s, p, and q be positive numbers whose reciprocals sum to 1. Then

$$(A.1) \qquad |\mathbb{P}XY - \mathbb{P}X\mathbb{P}Y| \leq 8\alpha(d)^{1/s}\|X\|_p\|Y\|_q.$$

If X happens to be a product $X_1\cdots X_m$, Hölder's inequality bounds the factor $\|X\|_p$ by

$$\left(\prod_i \mathbb{P}|X_i|^{mp}\right)^{1/mp}.$$

If $|X_i| \leq 1$ and $\mathbb{P}X_i^2 \leq \tau^{2+\gamma} \leq 1$ for every i, and if $mp \geq 2$, the product is less than $\tau^{(2+\gamma)/p}$. If Y has a similar decomposition into a product of $(k-m)$ factors, $\|Y\|_q$ is similarly bounded by $\tau^{(2+\gamma)/q}$. Choosing $s = (Q+\gamma)/\gamma$ and $mp = (k-m)q = k/(1-1/s)$, then decreasing the resulting exponent of τ from $(2+\gamma)Q/(Q+\gamma)$ to 2, we arrive at our working inequality,

$$(A.2) \qquad |\mathbb{P}X_1\cdots X_mY_1\cdots Y_{k-m}| \leq |\mathbb{P}X_1\cdots X_m\,\mathbb{P}Y_1\cdots Y_{k-m}| + 8\alpha(d)^{\gamma/(Q+\gamma)}\tau^2.$$

Here the choice of p and q is not critical; we need only ensure that $mp \geq 2$ and $(k-m)q \geq 2$.

For positive integers k and n, with τ fixed, define a bounding function

$$B_n(k) = n\tau^2 + (n\tau^2)^2 + \cdots + (n\tau^2)^{\lfloor k/2 \rfloor},$$

where $\lfloor k/2 \rfloor$ stands for the integer part of $k/2$. We will establish the existence of constants C_k, for $k = 1, \ldots, Q$, such that

$$(A.3) \qquad \sum_{\mathbf{i}} |\mathbb{P} Z(i_1) \cdots Z(i_k)| \le C_k B_n(k) \qquad \text{for all } n,$$

where the sum runs over all choices of $\mathbf{i} = (i_1, \ldots, i_k)$ such that $1 \le i_1 \le \cdots \le i_k \le n$. With $k = Q$ the left-hand side of (A.3) is greater than $1/Q!$ times the Q^{th} moment quantity that we are seeking to bound.

Inequality (A.3) holds for $k = 1$, since the $Z(i)$ have zero expected values. We will argue inductively to establish it for a $k > 1$, assuming that it holds for all values less than k.

For a given $\mathbf{i} = (i_1, \ldots, i_k)$, let $G(\mathbf{i})$ denote the largest of the differences $i_{j+1} - i_j$, and let $m(\mathbf{i})$ be the smallest j for which the difference equals $G(\mathbf{i})$. To simplify the notation, write $\beta(d)$ for $\alpha(d)^{\gamma/(Q+\gamma)}$. Apply the inequality (A.2) to each term on the left-hand side of (A.3) to bound the sum by

$$(A.4) \qquad \sum_{m=1}^{k-1} \sum_{\mathbf{i}} \{m(\mathbf{i}) = m\} \Big(|\mathbb{P} Z(i_1) \cdots Z(i_m) \mathbb{P} Z(i_{m+1}) \cdots Z(i_k)| + 8\tau^2 \beta(G(\mathbf{i})) \Big).$$

Consider first the contribution from the product of expectations. If $m = 1$ or $m = k - 1$, one of the expectations is zero. For other values of m we invoke the inductive hypothesis. Fixing m and $i_1, \ldots, i_m$ and letting $i_{m+1}, \ldots, i_k$ range from 1 to n instead of just from i_m to n, we bound the contribution by

$$\sum_{m=2}^{k-2} \sum_{i_1, \ldots, i_m} |\mathbb{P} Z(i_1) \cdots Z(i_m)| C_{k-m} B_n(k-m),$$

which, by a second appeal to the inductive hypothesis, is less than

$$\sum_{m=2}^{k-2} C_m C_{k-m} B_n(m) B_n(k-m).$$

The product $B_n(m) B_n(k-m)$ is a polynomial in $n\tau^2$ of degree

$$\left\lfloor \frac{m}{2} \right\rfloor + \left\lfloor \frac{k-m}{2} \right\rfloor \le \left\lfloor \frac{k}{2} \right\rfloor.$$

Thus the product of expectations contributes to (A.4) at most a constant multiple of $B_n(k)$.

For the contribution to (A.4) from the mixing coefficients we further decompose the sum over $\mathbf{i}$ according to the location and size of the largest gap $G(\mathbf{i})$. The contribution equals

$$\sum_{m=1}^{k-1}\sum_{\ell=1}^{n}\sum_{g=1}^{n}\sum_{\mathbf{i}}\{m(\mathbf{i})=m, i_m=\ell, G(\mathbf{i})=g\}8\tau^2\beta(g).$$

Given $m(\mathbf{i}) = m$ and $i_m = \ell$ and $G(\mathbf{i}) = g$, the indices $i_1, \ldots, i_{m-1}$ are subject to the constraints

$$1 \le i_1 \le \cdots \le i_{m-1} \le i_m = \ell,$$
$$i_{j+1} - i_j \le g \qquad \text{for } j = 1, \ldots, m-1.$$

Summing first over i_1, for fixed $i_2, \ldots, i_m$, and then over i_2, and so on, we constrain each index to a range of $1+g$ or fewer integers. There are at most $(1+g)^{m-1}$ choices for the first $m-1$ indices. Similarly, because the two equalities $i_m = \ell$ and $G(\mathbf{i}) = g$ fix i_{m+1} at the value $\ell + g$, there are at most $(1+g)^{k-m-1}$ choices for $i_{m+1}, \ldots, i_k$. Thus the mixing coefficients contribute to (A.4) at most

$$8\tau^2\sum_{m=1}^{k-1}\sum_{\ell=1}^{n}\sum_{g=1}^{n}(1+g)^{m-1}(1+g)^{k-m-1}\beta(g) \le 8\tau^2 kn\sum_{g=1}^{\infty}(1+g)^{k-2}\beta(g).$$

Assumption (ii) ensures finiteness of the sum over g; the whole contribution to (A.4) is less than a constant multiple of $n\tau^2$, which can be absorbed into $B_n(k)$. □

References

ALEXANDER, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12** 1041-1067.

ANDREWS, D. W. K. (1989a). An empirical process central limit theorem for dependent non-identically distributed random variables. Cowles Foundation Discussion Paper #907, Yale University.

ANDREWS, D. W. K. (1989b). Asymptotics for semiparametric econometric models: I. Estimation and testing. Cowles Foundation Discussion Paper #908R, Yale University.

ANDREWS, D. W. K. (1989c). Asymptotics for semiparametric econometric models: II. Stochastic equicontinuity and nonparametric kernel estimation. Cowles Foundation Discussion Paper #909R, Yale University.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

BOX, G. E. P. and ANDERSEN, S. L. (1955). Permutation theory in the derivation of robust criteria and the study of departures from assumption. *J. Royal Statist. Soc.* **17** 1-34.

CARBON, M. (1983). Inégalité de Bernstein pour les processus fortement mélangeants, non nécessairement stationnaires. Applications. *C. R. Acad. Sc. Paris* **297** 303-306.

COLLOMB, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau. *Z. Wahrsch. verw. Gebiete* **66** 441-460.

DEHLING, H. (1983). Limit theorems for sums of weakly dependent Banach space valued random variables. *Z. Wahrsch. verw. Gebiete* **63** 393-432.

DOUKHAN, P. and PORTAL, F. (1983). Moments de variables aléatoires mélangeantes. *C. R. Acad. Sc. Paris* **297** 129-132.

DOUKHAN, P. and PORTAL, F. (1984). Vitesse de convergence dans le théorème central limite pour des variables aléatoires mélangeantes à valeurs dans un espace de Hilbert. *C. R. Acad. Sc. Paris* **298** 305-308.

DUDLEY, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899-929.

DUDLEY, R. M. (1981). Donsker classes of functions. In *Statistics and Related Topics* (M. Csörgő, D. A. Dawson, J. N. K. Rao and A. K. Md. E. Saleh, eds.) 341-352. North-Holland, Amsterdam.

GOLDIE, C. M. and GREENWOOD, P. E. (1986a). Characterisations of set-indexed Brownian motion and associated conditions for finite-dimensional convergence. *Ann. Probab.* **14** 802-816.

GOLDIE, C. M. and GREENWOOD, P. E. (1986b). Variance of set-indexed sums of mixing random variables and weak convergence of set-indexed processes. *Ann. Probab.* **14** 817-839.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York, NY.

HERRNDORF, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Probab.* **12** 141-153.

LEDOUX, M. and TALAGRAND, M. (1990). *Isoperimetry and Processes in Probability in Banach Spaces.* To appear.

LEVENTHAL, S. (1988). Uniform limit theorems for Harris recurrent markov chains. *Probab. Theory Related Fields* **80** 101-118.

LEVENTHAL, S. (1989). A uniform CLT for uniformly bounded families of martingale differences. *Journal of Theoretical Probability* **2** 271-287.

MASSART, P. (1986). Rates of convergence in the central limit theorem for empirical processes. *Ann. Inst. H. Poincaré Probab. Statist.* **22** 381-423.

MCLEISH, D. L. (1977). On the invariance principle for nonstationary mixingales. *Ann. Probab.* **5** 616-621.

NOLAN, D. and POLLARD, D. (1987). U-processes: rates of convergence. *Ann. Statist.* **15** 780-799.

NOLAN, D. and POLLARD, D. (1988). Functional limit theorems for U-processes. *Ann. Probab.* **16** 1291-1298.

OSSIANDER, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.* **15** 897-919.

PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces.* Academic, New York.

PHILIPP, W. (1969). The central limit problem for mixing sequences of random variables. *Z. Wahrsch. verw. Gebiete* **12** 155-171.

PISIER, G. (1983). Some applications of the metric entropy condition to harmonic analysis. *Banach Spaces, Harmonic Analysis, and Probability Theory. Lecture Notes in Math.* **995** 123-154. Springer, New York.

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications.* Institute of Mathematical Statistics, Hayward, CA. To appear in the Regional Conference Series in Probability and Statistics.

POLLARD, D. (1991). A maximal inequality for sums of independent processes under a bracketing condition. Preprint, Statistics Dept., Yale University.

SEN, P.K. (1974). Weak convergence of multidimensional empirical processes for stationary ϕ-mixing processes. *Ann. Probab.* **2** 147-154.

VAPNIK, V. N. and ČERVONENKIS, A. YA. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* **16** 264-280.

YOKOYAMA, R. (1980). Moment bounds for stationary mixing sequences. *Z. Wahrsch. verw. Gebiete* **52** 45-57.

YUKICH, J. E. (1986). Rates of convergence for classes of functions: the non-iid case. *J. Multivariate Analysis* **20** 175-189.