The Annals of Probability
1994, Vol. 22, No. 1, 94–116

RATES OF CONVERGENCE FOR EMPIRICAL PROCESSES OF STATIONARY MIXING SEQUENCES[1]

BY BIN YU

University of Wisconsin—Madison

Classical empirical process theory for Vapnik-Červonenkis classes deals mainly with sequences of independent variables. This paper extends the theory to stationary sequences of dependent variables. It establishes rates of convergence for β-mixing and ϕ-mixing empirical processes indexed by classes of functions. The method of proof depends on a coupling of the dependent sequence with sequences of independent blocks, to which the classical theory can be applied. A uniform $O(n^{-s/(1+s)})$ rate of convergence over V-C classes is established for sequences whose mixing coefficients decay slightly faster than $O(n^{-s})$.

1. Introduction. There has been a great deal of research work on empirical processes indexed by classes of functions. It was Vapnik and Čveronenkis (1971) who showed the uniform convergence of the empirical processes indexed by Vapnik-Červonenkis (V-C) classes in the iid case. Many papers followed; for example, Dudley (1978), Dudley and Philipp (1983), Giné and Zinn (1984), Le Cam (1984) and Pollard (1982). Most of this work, however, concentrates on the independent case.

In this paper, we extend some of the previous results in the iid case to the dependent case, that is, we obtain uniform convergence and rates of convergence for empirical processes of β-mixing (completely regular) or ϕ-mixing sequences for which the CLT does not hold. Conditions on the mixing rate of the sequence and the metric entropy of the index class are imposed. The entropy condition is in terms of the L^1 random semimetric. The main technique used is the construction of an independent block (IB) sequence which enables us to employ the symmetrization technique and an exponential inequality available in the iid case. Note that the blocking technique can be traced back to Bernstein (1927).

In the case of V-C index classes, Nobel and Dembo (1993) obtained the uniform convergence result under weaker β-mixing conditions than in this paper. Philipp (1984), Yukich (1986) and Massart (1988) considered the dependent case for general index classes, but under bracketing conditions. In particular, Yukich (1986) derived rates of convergence results for ϕ-mixing sequences, while Philipp (1984) and Massart (1988) obtained invariance principle results on top of the rates of convergence for ϕ- and α-mixing sequences,

Received September 1991; revised December 1992.
[1]Research supported in part by NSF Grants MC–84–03239, DMS–87–0146, DMS–90–01710 and ARO Grant DAAL03–91–G–007.
AMS 1991 *subject classifications*. Primary 60F05, 60F17; secondary 60G10.
Key words and phrases. Empirical process, β–mixing, rate of convergence, random metric entropy, V–C class, blocking.

and we note that α-mixing conditions are weaker than the corresponding β-mixing ones. More recently Andrews and Pollard (1993) gave a nice exposition of CLTs for dependent sequences. Their work is closely related to Massart (1988) and Philipp (1984). Other related works are Levental (1988, 1989) on Harris recurrent Markov chains and martingale difference sequences. Levental (1988) took advantage of the independent regenerative block structure of Harris recurrent Markov chains and used iid techniques on those blocks, similar in spirit to our approach here. Note that Nobel and Dembo (1993) and the work presented here might be viewed as an extension of Vapnik–Červeronenkis–type theory to weakly dependent (mixing) sequences. In this context we note the remark of Massart (1988), who stated that "we don't know whether the above weakly dependent framework [as in Massart (1988)] could support a general 'Vapnik–Červeronenkis type theory' or not."

The rest of the paper is organized as follows. Section 2 gives preliminaries on mixing sequences and metric entropies. Section 3 includes the main results, that is, rates of convergence of empirical processes for bounded index classes (Theorem 3.1) and uniform convergence (Theorem 3.4) for general index classes. Section 4 contains the proofs, starting with a construction of an independent block sequence which is the cornerstone of the proofs.

The measurability issues will be dealt with in the appendix in a way similar to Pollard [(1984), Appendix C] although the construction of the independent block sequence does complicate the issue. We assume the index class to be permissible to ensure that the necessary measurability requirements are satisfied. The definition of a permissible class can be found in the appendix.

The original version of the paper when submitted also contained a CLT [cf. Yu (1990a, b)] which has since been improved by Arcones and Yu (1994) using the same blocking technique. Interested readers are referred to that paper for details.

2. Preliminaries.

This section contains the preliminary material on mixing sequences and metric entropies. The size of the index class of the empirical process can be regulated through metric entropy conditions related to the empirical L^1 norm. The algebraic decay condition of the covering number of an index class is also introduced.

Let $\underline{X} = (X_i)_{i \geq 1}$ be a strictly stationary real-valued sequence with distribution P, which implies X_i, $i \geq 1$, all have the same distribution P. For the sequence $\underline{X}$, let

$$\sigma_l = \sigma(X_1, X_2, \dots, X_l)$$

and

$$\sigma'_{l+k} = \sigma(X_{l+k}, X_{l+k+1}, \dots,).$$

Many kinds of mixing conditions exist in the literature. The weakest among those most commonly used is called strong mixing or α-mixing.

2.1 DEFINITION. For any sequence $\underline{X}$, the α-mixing coefficient α_k is defined as follows:

$$\alpha_k(\underline{X}) = \tfrac{1}{2}\sup\{E|P(B|\sigma_l) - P(B)|: B \in \sigma'_{l+k}, l \ge 1\}.$$

Other mixings are the following.

2.2 DEFINITION. For any sequence $\underline{X}$, the β-mixing (or completely regular) coefficient β_k is defined as follows:

$$\beta_k(\underline{X}) = \tfrac{1}{2}\sup\left\{\sum_{i=1}^{I}\sum_{j=1}^{J}|P(A_i \cap B_j) - P(A_i)P(B_j)|: \{A_i\} \text{ any finite partition in } \sigma_l, \{B_j\} \text{ any finite partition in } \sigma'_{l+k}, l \ge 1\right\}.$$

Note that this definition [cf. Bradley (1983)] is more convenient when dealing with the measureability issues than the following equivalent one:

$$\beta_k(\underline{X}) = \sup_{l\ge 1} E\sup\{|P(B|\sigma_l) - P(B)|: B \in \sigma'_{l+k}\}.$$

2.3 DEFINITION. For any sequence $\underline{X}$, the ϕ-mixing coefficient ϕ_k is defined as follows:

$$\phi_k(\underline{X}) = \sup\{|P(B|A) - P(B)|: A \in \sigma_l, B \in \sigma_{l+k}, l \ge 1\}.$$

Moreover, we assume throughout the paper that there are positive constants r_c (c standing for α-, β- and ϕ-mixings) such that

$$c_k = O(k^{-r_c}).$$

The three mixing coefficients are ordered as follows [see Philipp (1986)]:

$$\alpha_k \le \beta_k \le \phi_k.$$

Note that the stationary sequence $\{f(X_i): i = 1, 2, \dots\}$ for a measurable function f has its α-mixing or β-mixing or ϕ-mixing rate bounded by the corresponding rate of the original sequence, since for any measurable f the σ-field of $f(X)$ is contained in the σ-field of X, that is,

$$\alpha_k \ge \alpha_k(f), \qquad \beta_k \ge \beta_k(f) \quad \text{and} \quad \phi_k \ge \phi_k(f).$$

Therefore, if the sequence $\{X_i\}$ satisfies an α- or β- or ϕ-mixing condition, then so does the sequence $\{f(X_i)\}$.

For examples of mixing sequences, see Athreya and Pantula (1986), Ibragimov and Rozanov [(1978), IV.4], Mokkadem (1988), Pham and Tran (1985) and Withers (1981). In particular, a Markov chain is ϕ-mixing with $\theta_\phi < 1$ under some regularity conditions [cf. Section 5.5 in Doob (1953)].

We take Pollard's (1984) linear functional notation and use P instead of E to denote expectations. Hence, $Pf = \int f\,dP = Ef(X_1) = Ef(X_n)$ for all $n \geq 1$.

It is known that if the mixing rate of sequence tends to zero fast, the variance of the sum of function values over n successive observations is $O(n)$. This is necessary for the CLT to hold, but we have a general bound on the variance even when the CLT does not hold.

2.4 Lemma. *Suppose that f is a bounded measurable function $|f| < M$, $\underline{X}$ a strictly stationary β-mixing sequence and $Pf = 0$. Then the following hold*:

(i)

$$P\left(\sum_1^n f(X_i)\right)^2 \leq n\left(1 + 20\sum_1^n \alpha_k\right)M^2. \tag{2.1}$$

(ii) *For* $0 < r_\beta \leq 1$ *and* $\delta_\beta = I_{\{r_\beta = 1\}}$ (*i.e.*, $\delta_\beta = 1$ *if* $r_\beta = 1$, $\delta_\beta = 0$ *if* $0 < r_\beta < 1$), *there exists a constant* $C = C(M, r_\beta)$ *such that*

$$P\left(\sum_1^n f(X_i)\right)^2 \leq C_n^{2-r_\beta}(\log n)^{\delta_\beta}. \tag{2.2}$$

Proof. Since f is bounded by M, (2.1) follows directly from Lemma 3.2 of Dehling (1983). Relation (2.2) follows from (2.1) since the β-mixing coefficient β_k bounds the α-mixing coefficient α_k from above and $\Sigma_k^n k^{-r_\beta} = O(n^{1-r_\beta}(\log n)^{\delta_\beta})$. □

We may define the β-mixing and ϕ-mixing coefficients for any probability measure Q on a product measure space $(\Omega_1 \times \Omega_2, \Sigma_1 \times \Sigma_2)$ as follows.

2.5 Definition. Suppose that Q_1 and Q_2 are the marginal probability measures of Q on (Ω_1, Σ_1) and (Ω_2, Σ_2), respectively. Then we define

$$\beta(\Sigma_1, \Sigma_2, Q) = P\sup\{|Q(B|\Sigma_1) - Q_2(B)|: B \in \Sigma_2\},$$
$$\phi(\Sigma_1, \Sigma_2, Q) = \sup\{|Q(B|A) - Q_2(B)|: A \in \Sigma_1, B \in \Sigma_2\}.$$

The following lemma is straightforward from our definition of the β-mixing coefficient if we approximate a measurable function by a sequence of simple functions, or it can be found in Volkonskii and Rozanov (1959) and Eberlein (1984).

2.6 Lemma. *Suppose that $h(x, y)$ is a measurable function with bound M_h, and P is the product measure $Q_1 \times Q_2$. Then we have*

$$|Qh - Ph| \leq M_h\beta(\Sigma_1, \Sigma_2, Q) \leq M_h\phi(\Sigma_1, \Sigma_2, Q).$$

By induction and Lemma 2.6, we have the following.

2.7 COROLLARY. *Let $m \geq 1$, and suppose that h is a bounded measurable function on a product probability space $(\prod_{i=1}^m \Omega_i, \prod_{i=1}^m \Sigma_i)$. Let Q be a probability measure on the product space with marginal measures Q_i on (Ω_i, Σ_i), and let Q^{i+1} be the marginal measure of Q on $(\prod_{j=1}^{i+1} \Omega_j, \prod_{j=1}^{i+1} \Sigma_j)$, $i = 1, \dots, m-1$. Write*

$$\phi(Q) = \sup_{1 \leq i \leq m-1} \phi\left(\prod_{j=1}^{i} \Sigma_j, \Sigma_{i+1}, Q^{i+1} \right),$$

and define $\beta(Q)$ similarly and let $P = \prod_{i=1}^m Q_i$. Then

$$|Qh - Ph| \leq (m-1) M_h \beta(Q) \leq (m-1) M_h \phi(Q).$$

REMARK. Corollary 2.7 is the key to connecting the mixing sequence and the independent block sequence (see Section 4 for details).

Before we introduce the definitions of metric entropy, we need some notation.

For any mixing sequence $X_1, X_2, \dots$ and Borel-measurable function f, denote by P_n the empirical measure of the first n observations:

$$P_n f = \frac{1}{n} \sum_1^n f(X_j).$$

For any family $\mathbf{F}$ of measurable functions, we call F an envelope function of $\mathbf{F}$ if $|f| \leq F$ for all f in $\mathbf{F}$.

Since we are interested in the uniform performance of the empirical measure, intuition suggests that we might have to regulate the size of $\mathbf{F}$. One measure of the size of $\mathbf{F}$ is the covering number or metric entropy. Along the lines of Pollard (1984), we define the covering number as follows.

2.8 DEFINITION (Covering number). The covering number $N(\varepsilon, d, \mathbf{F})$ related to a semimetric d on $\mathbf{F}$ is defined as

$$N(\varepsilon, d, \mathbf{F}) = \min_m \left\{ \text{there are } g_1, \dots, g_m \text{ in } L^1(P), \text{ such that } \min_{1 \leq j \leq m} d(f, g_j) \leq \varepsilon \quad \text{for any } f \text{ in } \mathbf{F} \right\}.$$

The quantity $\log N(\varepsilon, d, \mathbf{F})$ is called the *metric entropy* at ε. See Kolmogorov and Tihomirov (1959).

If we take the following random L^1-semimetric $\rho_{1,n}$ as d:

$$\rho_{1,n}(f, g) = P_n(|f - g|),$$

then the related covering number is random. Denote by ρ_1 the L^1-semimetric

$$\rho_1(f, g) = P|f - g|.$$

The following algebraic decay conditions on covering numbers will be imposed in the later theorems. They are known to be satisfied by V-C classes

[Dudley (1978)]:

(2.3) $N(\varepsilon, \rho_1, \mathbf{F}) = O(\varepsilon^{-w})$ for some $w > 0$, as $\varepsilon \to 0$;

(2.4) $N(\varepsilon, \rho_{1,n}, \mathbf{F}) = O_P(\varepsilon^{-w})$ for some $w > 0$, as $\varepsilon \to 0$ and $n \to \infty$.

3. Limit theorems: The main results. This section contains the statements of the main results, that is, the rates of convergence and uniform convergence for the empirical process of a stationary mixing sequence indexed by a class of functions. The main results are Theorem 3.1 (the rates of convergence) and Theorem 3.3 (uniform convergence). We concentrate on the case where $0 < r_\beta \le 1$ because when $r_\beta > 1$ a CLT holds [cf. Arcones and Yu (1994)]. Since ϕ-mixing is a stronger condition, all the results in this section hold under corresponding ϕ-mixing conditions. The proofs are left to Section 4.

For the uniform convergence theorem, we first restrict ourselves to bounded index classes. We then find proper conditions to ensure that the law of large numbers holds for the envelope function, which is in turn the condition necessary to generalize the uniform convergence result to classes having nonconstant envelope functions. Moreover, we use $\mathbf{P}$ to denote a probability measure on a (sometimes unspecified) measurable space whose precise definition can be found in the Appendix.

3.1 THEOREM (Rates of convergence). *Let $\underline{X}$ be a stationary β-mixing sequence and let $\mathbf{F}_M$ be a permissible index class with a constant envelope function M. Assume $0 < r_\beta \le 1$. For any given $0 < s < r_\beta$ and $h_n \to \infty$ as $n \to \infty$, let $a_n = [n^{1/(1+s)}]$ and $\mu_n = [n/(2a_n)] = [n^{s/(1+s)}/2]$. Then if, for all $c > 0$,*

$$\log N\big(cn^{-s/(1+s)}h_n, \rho_{1,n}, \mathbf{F}_M\big) = o_{\mathbf{P}}(h_n), \tag{3.1}$$

we have

$$\mathbf{P}\Big\{\sup_{f\in\mathbf{F}_M} |P_n f - Pf| > \varepsilon n^{-s/(1+s)}h_n\Big\} \to 0 \quad \text{as } n \to \infty,$$

that is,

$$\sup_{f\in\mathbf{F}_M} |P_n f - Pf| = o_{\mathbf{P}}\big(n^{-s/(1+s)}h_n\big) \quad \text{as } n \to \infty.$$

3.2 COROLLARY. *Suppose that $\mathbf{F}_M$ is a bounded permissible index class satisfying* (2.4), *that is,*

$$N(\varepsilon, \rho_{1,n}, \mathbf{F}_M) = O_{\mathbf{P}}(\varepsilon^{-w}) \quad \text{for some } w > 0, \text{ as } \varepsilon \to 0 \quad \text{and } n \to \infty.$$

If $0 < r_\beta \le 1$, then for any given $0 < s < r_\beta$,

$$\mathbf{P}\Big\{\sup_{f\in\mathbf{F}_M} |P_n f - Pf| > \varepsilon n^{-s/(1+s)}\Big\} \to 0 \quad \text{as } n \to \infty,$$

that is,

$$\sup_{f\in\mathbf{F}_M} |P_n f - Pf| = o_{\mathbf{P}}(n^{-s/(1+s)}) \quad \text{as } n \to \infty.$$

3.3 THEOREM (Uniform convergence for bounded families). *Suppose that* $\underline{X}$ *is a stationary* β*-mixing sequence. Assume that* $\mathbf{F}_M$ *is a bounded permissible index class,* $0 < r_\beta \le 1$, *and for some* $0 < \alpha_0 < r_\beta/(1 + r_\beta)$,

$$\log N(\varepsilon, \rho_{1,n}, \mathbf{F}_M) = o_{\mathbf{P}}(n^{\alpha_0}) \quad \text{for } \varepsilon > 0 \quad \text{as } n \to \infty. \tag{3.2}$$

Then for any given $\varepsilon > 0$, *we have*

$$\mathbf{P}\left\{ \sup_{f\in\mathbf{F}_M} |P_n f - Pf| > \varepsilon \right\} \to 0 \quad \text{as } n \to \infty,$$

that is,

$$\sup_{f\in\mathbf{F}_M} |P_n f - Pf| = o_{\mathbf{P}}(1) \quad \text{as } n \to \infty.$$

REMARKS. (i) Arcones and Yu (1994) contains results on CLT's. In particular, they show that, when $r_\beta > 1$, a CLT holds for the empirical process if the index class is a V-C class.

(ii) As a straightforward consequence of Corollary 3.2 and Theorem 3.3, for a bounded permissible class $\mathbf{F}_M$ satisfying (2.4), especially a V-C class, one has

$$\mathbf{P}\left\{ \sup_{f\in\mathbf{F}_M} |P_n f - Pf| > \varepsilon n^{-s/(1+s)} \right\} \to 0 \quad \text{as } n \to \infty$$

for $0 < s < r_\beta \le 1$ and

$$\mathbf{P}\left\{ \sup_{f\in\mathbf{F}_M} |P_n f - Pf| > \varepsilon \right\} \to 0 \quad \text{as } n \to \infty,$$

provided that $r_\beta > 0$.

3.4 THEOREM (Uniform convergence for general families). *Assume that* $\mathbf{F}$ *is a permissible index class of functions with an* $L^1(P)$ *envelope function* F. *If* $0 < r_\beta \le 1$ *and for some* $0 < \alpha_0 < r_\beta/(1 + r_\beta)$, *and for* $\varepsilon > 0$,

$$\log N(\varepsilon, \rho_{1,n}, \mathbf{F}) = o_{\mathbf{P}}(n^{\alpha_0}), \tag{3.3}$$

then for any given $\varepsilon > 0$, *we have*

$$\mathbf{P}\left\{ \sup_{f\in\mathbf{F}} |P_n f - Pf| > \varepsilon \right\} \to 0 \quad \text{as } n \to \infty,$$

that is,

$$\sup_{f\in\mathbf{F}} |P_n f - Pf| = o_{\mathbf{P}}(1) \quad \text{as } n \to \infty.$$

4. Proofs. This section contains the proofs for Theorems 3.1–3.4. First, we explain how a blocking technique enables us to get the exponential inequality, which is the key step in proving our results. Then the rigorous proofs are given in the form of a few lemmas after constructing an independent block (IB) sequence. Some useful inequalities relating the original sequence to the IB sequence are also presented (Lemma 4.2).

After observing that all the results we want are either in terms of distribution, or can be interpreted in terms of probabilities, we construct an IB sequence from the original stationary mixing sequence such that the IB sequence is very close in distribution to the mixing sequence. We then transfer the problem to the IB sequence to which the standard techniques of the independent case can be applied. Symmetrization and Hoeffding's inequality are used for the IB sequence in the proof of Theorem 3.1.

We divide the n-sequence $X_n = (X_1, X_2, \dots, X_n)$ into blocks of length a_n, one after the other. We eliminate every other block and work with the remaining odd-numbered blocks. Depending on the mixing and metric entropy conditions to be assumed, we choose a_n large enough so that the odd-numbered a_n-blocks are "almost" independent, but at the same time choose a_n small enough so that the odd-numbered a_n-blocks together behave similarly to the original mixing sequence. Then we construct an independent sequence of blocks where each block has the same distribution as one of the a_n-blocks of the original sequence.

More precisely, for any integer pair (a_n, μ_n) with $\mu_n = [n/2a_n]$, we divide the strictly stationary n-sequence $\underline{X}_n = (X_1, X_2, \dots, X_n)$ into $2\mu_n$ blocks of length a_n and the remainder block of length $n - 2\mu_n a_n$. Denote the indices in the blocks alternately by H's and T's, and denote the indices in the remainder block by R_e. These indices depend on n, but for simplicity we suppress n. That is,

$$H_1 = \{i: 1 \le i \le a_n\},$$
$$T_1 = \{i: a_n + 1 \le i \le 2a_n\}.$$

Generally, for $1 \le j \le \mu_n$,

$$H_j = \{i: 2(j-1)a_n + 1 \le i \le (2j-1)a_n\},$$
$$T_j = \{i: (2j-1)a_n + 1 \le i \le (2j)a_n\}.$$

Denote the random variables that correspond to the H_j and T_j indices as

$$X(H_j) = \{X_i, i \in H_j\}, \qquad X(T_j) = \{X_i, i \in T_j\}.$$

Further, let the whole sequence of H-blocks be denoted by $X_{a_n} = \{X(H_j): j = 1, 2, \dots, \mu_n\}$.

Now, we take a sequence of identically distributed independent blocks $\{\Xi(H_j): j = 1, \dots, \mu_n\}$, where $\Xi(H_j) = \{\xi_i: i \in H_j\}$, such that the sequence is independent of $\underline{X}_n$ and each block has the same distribution as a block from the original sequence:

$$\mathscr{L}(\Xi(H_j)) = \mathscr{L}(X(H_j)) = \mathscr{L}(X(H_1)).$$

We call this constructed sequence the independent block a_n-sequence (IB sequence). Denote the IB sequence as Ξ_{a_n}. (Note that a proper measurable space can be found to host both sequences and on this space measurability issues can be addressed; see the Appendix.) Because of the mixing condition, we can relate X_{a_n} and Ξ_{a_n} in the following way.

4.1 LEMMA. *Let the distributions of* X_{a_n} *and* Ξ_{a_n} *be* Q *and* $\tilde{Q}$, *respectively. For any measurable function* h *on* $R^{\mu_n a_n}$ *with bound* M,

$$\left|Qh(X_{a_n}) - \tilde{Q}h(\Xi_{a_n})\right| \le M(\mu_n - 1)\beta_{a_n}. \tag{4.1}$$

PROOF. This is a direct application of Corollary 2.7. In the corollary, take $Q =$ the probability distribution of the a_n-sequence with $\Omega_i = R^{a_n}$, $\Sigma_i =$ product Borel σ-field on R^{a_n} and $m = \mu_n$. Then P in the corollary equals the probability distribution of the IB a_n-sequence, that is, $\tilde{Q}$. Notice that $\beta(Q) \le \beta_{a_n}$. □

REMARKS. (i) This is the key lemma that is used throughout the subsequent proofs. Different functions h are used in the application of this lemma; in particular, h is often taken to be an indicator function.

(ii) The β-mixing (or ϕ-mixing) condition is required for the uniform convergence result because in our approach this lemma is crucial in connecting the original sequence with the IB sequence. We are unable to obtain this lemma under α-mixing conditions and doubt that it is true.

Recall that an index class with an envelope function F is denoted by $\mathbf{F}$. For simplicity, we assume $Pf = 0$ for all f in $\mathbf{F}$. Then the empirical measure on $\{f(X_i)\colon i = 1, 2, \ldots, n\}$ is

$$P_n f = \frac{1}{n}\sum_i^n f(X_j).$$

For the original sequence $\underline{X}$, we write

$$Y_{j,f}(X_{a_n}) = \sum_{i \in H_j} f(X_i) \quad \text{and} \quad Y_{1,j,f}(X_{1,a_n}) = \sum_{i \in T_j} F(X_i).$$

For the constructed IB sequence Ξ, define

$$Z_{j,f}(\Xi_{a_n}) = \sum_{i \in H_j} f(\xi_i) \quad \text{and} \quad \tilde{P}_{\mu_n} f = \frac{1}{n}\sum_1^{\mu_n} Z_{j,f}.$$

Associated with this empirical measure (note that it is not a probability measure if $a_n > 0$) are two random semimetrics,

$$\rho_{1,\mu_n}(f,g) = \frac{1}{n}\sum_1^{\mu_n} |Y_{j,f-g}|$$

and

$$\tilde{\rho}_{1,\mu_n}(f,g) = \frac{1}{n}\sum_1^{\mu_n} |Z_{j,f-g}|.$$

We shall compare these two semimetrics with the L^1 empirical metric

$$\rho_{1,n}(f,g) = P_n|f-g|.$$

The L^1 random covering numbers corresponding to ρ_{1,μ_n}, $\tilde{\rho}_{1,\mu_n}$ and $\rho_{1,n}$ are denoted by

$$N(\varepsilon, \rho_{1,\mu_n}, \mathbf{F}), \qquad N(\varepsilon, \tilde{\rho}_{1,\mu_n}, \mathbf{F}) \quad \text{and} \quad N(\varepsilon, \rho_{1,n}, \mathbf{F}),$$

respectively.

For simplicity, from now on assume that (μ_n, a_n) is always an integer pair satisfying $n/2 - a_n \le \mu_n a_n \le n/2$, and $\mu_n \to \infty$, $a_n \to \infty$. Thus, $a_n = o(n)$, and $\mu_n a_n = O(n)$. Moreover, the IB sequence is assumed to be defined in terms of a pair of integers (μ_n, a_n) implicitly.

The following lemma allows us to replace P_n by $\tilde{P}_{\mu_n}$ with only an error of order $\mu_n \beta_{a_n}$.

4.2 LEMMA. *Suppose that $\mathbf{F}_M$ is a permissible bounded class, and $b_n = O(1)$, as $n \to \infty$. If $\mu_n b_n \to \infty$, then*

$$\mathbf{P}\Big(\sup_{f\in\mathbf{F}} |P_n f| \ge \varepsilon b_n\Big) \le 2\mathbf{P}\Big(\sup_{f\in\mathbf{F}} |\tilde{P}_{1,\mu_n} f| \ge \frac{\varepsilon}{4} b_n\Big) + 2\mu_n\beta_{a_n}. \tag{4.2}$$

PROOF. Note that the sum of f over the remainder block R_e is uniformly bounded by $M(2a_n)n^{-1} = O(\mu_n^{-1})$, which tends to zero faster than b_n since $\mu_n b_n \to \infty$, and X_{a_n} has the same distribution as $X_{1,a_n} = \{X_i\colon i \in T_j \text{ for } 1 \le j \le \mu_n\}$. Therefore, for n sufficiently large, we have

$$\begin{aligned}
&\mathbf{P}\Big(\sup_{f\in\mathbf{F}} |P_n f| \ge \varepsilon\Big) \\
&\quad \le \mathbf{P}\left(\sup_{f\in\mathbf{F}} \left|\frac{1}{n}\sum_{j=1}^{\mu_n} Y_{j,f}(X_{a_n}) + \frac{1}{n}\sum_{j=1}^{\mu_n} Y_{j,f}(X_{1,a_n})\right| > \frac{\varepsilon}{2}\right) \\
&\quad \le \mathbf{P}\left(\sup_{f\in\mathbf{F}} \left|\frac{1}{n}\sum_{j=1}^{\mu_n} Y_{j,f}(X_{a_n})\right| > \frac{\varepsilon}{4}\right) \\
&\qquad + \mathbf{P}\left(\sup_{f\in\mathbf{F}} \left|\frac{1}{n}\sum_{j=1}^{\mu_n} Y_{1,j,f}(X_{1,a_n})\right| > \frac{\varepsilon}{4}\right) \\
&\quad = 2\mathbf{P}\left(\sup_{f\in\mathbf{F}} \left|\frac{1}{n}\sum_{j=1}^{\mu_n} Y_{j,f}(X_{a_n})\right| > \frac{\varepsilon}{4}\right).
\end{aligned} \tag{4.3}$$

Taking for h the indicator function of the event

$$\left\{\sup_{f\in\mathbf{F}} \left|\frac{1}{n}\sum_{j=1}^{\mu_n} Y_{j,f}(X_{a_n})\right| > \frac{\varepsilon}{4}\right\},$$

Lemma 4.1 gives the following bound on the left-hand side of (4.3):

$$2\mathbf{P}\left(\sup_{f\in\mathbf{F}} \left|\frac{1}{n}\sum_{j=1}^{\mu_n} Z_{j,f}(X_{a_n})\right| > \frac{\varepsilon}{4}\right) + 2\mu_n\beta_{a_n}. \qquad \square$$

Since the IB sequence consists of iid blocks, we can use the standard symmetrization technique in the independent case to get the uniform convergence (rates of convergence) for the pseudoempirical measure $\tilde{P}_{1,\mu_n}$ of the IB sequence. The entropy conditions needed will be in terms of the IB sequence. Therefore, we need to relate the entropy conditions on the original sequence to those on the IB sequence. Because the entropy conditions are random, that is, they can be stated in terms of probabilities, we can easily transfer the entropy condition about the original sequence to the IB sequence by Lemma 4.1.

4.3 LEMMA. *Assume that* $\mathbf{F}$ *is a permissible index class and* $F = M$.

(i) *For* $b_n = O(1)$, *if* $\mu_n \beta_{a_n} = o(1)$ *and* $\log N(\varepsilon, \rho_{1,n}, \mathbf{F}) = o_{\mathbf{P}}(b_n)$, *then*

$$\log N\big(\varepsilon, \tilde{\rho}_{1,\mu_n}, \mathbf{F}\big) = o_{\mathbf{P}}(b_n). \tag{4.4}$$

(ii) *For any* $0 < s < r_\beta$ *and* $h_n \to \infty$ *as* $n \to \infty$, *let* $b_n = n^{-s/(1+s)} h_n$, $a_n = [n^{1/(1+s)}]$ *and* $\mu_n = [n^{s/(1+s)}/2]$. *Then for* n *large there exists a* $\delta(\varepsilon) > 0$ *for which we have*

$$\begin{aligned} \mathbf{P}\Big(\sup_{f\in\mathbf{F}} |\tilde{\mathbf{P}}_{\mu_n} f| \geq \varepsilon b_n\Big) &\leq 10 \exp(-O(h_n)) \\ &\quad + 16\mathbf{P}\Big(h_n \leq O\Big(\log N\Big(\delta n^{-s/(1+s)} h_n, \tilde{\rho}_{1,\mu_n}, \mathbf{F}\Big)\Big)\Big). \end{aligned} \tag{4.5}$$

(iii) *Under the assumptions in* (ii) *and the further assumption that, for all* $c > 0$,

$$\log N\Big(c n^{-s/(1+s)} h_n, \tilde{\rho}_{1,\mu_n}, \mathbf{F}\Big) = o_{\mathbf{P}}(h_n), \tag{4.6}$$

then, for any $\varepsilon > 0$ *and* n *large,*

$$\mathbf{P}\Big(\sup_{f\in\mathbf{F}} |\tilde{P}_{\mu_n} f| \geq \varepsilon b_n\Big) < \varepsilon.$$

PROOF. (i) By the triangle inequality, we have

$$\rho_{1,\mu_n}(f, g) \leq \frac{1}{n} \sum_1^{\mu_n} Y_{j,|f-g|} \leq \frac{1}{n} \sum_{i=1}^{n} |f(X_i) - g(X_i)| \leq \rho_{1,n}(f, g).$$

Thus,

$$N\big(\varepsilon, \rho_{1,\mu_n}, \mathbf{F}\big) \leq N(\varepsilon, p_{1,n}, \mathbf{F}).$$

This together with the assumption in (i) implies

$$\log N\big(\varepsilon, \rho_{1,\mu_n}, \mathbf{F}\big) = o_{\mathbf{P}}(b_n).$$

Then taking h in Lemma 4.1 to be the indicator function of the event

$$\left\{\frac{\log N\big(\varepsilon, \rho_{1,\mu_n}, \mathbf{F}\big)}{b_n} > \varepsilon\right\},$$

we obtain the conclusion of (i).

(ii) This can be obtained by the standard symmetrization technique for the iid case, for the Z's are iid for any fixed f in $\mathbf{F}_M$, and by (2.2) in Lemma 2.4,

$$P\left(Z_{j,f}^2\right) \le C a_n^{2-r_\beta}(\log a_n)^{\delta_\beta},$$

which implies, if $a_n^{r_\beta}\mu_n b_n^2 \to \infty$ as $n \to \infty$,

$$\mathbf{P}\left(|\tilde{P}_{\mu_n} f| \ge \varepsilon b_n\right) \le \frac{4}{\varepsilon^2 b_n^2 n^2}\mu_n C a_n^{2-r_\beta}(\log a_n)^{\delta_\beta} = O\left(\frac{1}{a_n^{r_\beta}\mu_n b_n^2}\right) < 1 - b = \frac{1}{2}.$$

Since $Z_{j,f}$'s are iid blocks we may use arguments from the iid case.

Let $\{\pi_n f\colon f \in \mathbf{F}\}$ be the $(\varepsilon b_n/2)$-net of $\mathbf{F}$ in terms of $\tilde{\rho}_{1,\mu_n}$. Then, for any $f \in \mathbf{F}$,

$$\tilde{P}_{\mu_n}|f - \pi_n f| \le \varepsilon b_n/2.$$

Denote

$$A_n = \left\{\Xi_{a_n}\colon \log N\left(\frac{\varepsilon b_n}{8}, \tilde{\rho}_{1,\mu_n}, \mathbf{F}\right) \le \frac{\varepsilon^2 n b_n^2}{64\nu_n},\right.$$

$$\left.\log N\left(\frac{\nu_n}{16 M a_n}, \tilde{\rho}_{1,\mu_n}, \mathbf{F}\right) \le \frac{\mu_n \nu_n}{16 M^2 a_n}\right\},$$

$$B_n = \left\{\Xi_{a_n}\colon \sup_{f\in\mathbf{F}}\left|\sum_{j=1}^{\mu_n} Z_{j,f}^2\right| \le \nu_n n\right\}$$

for some ν_n to be chosen later. Let

$$\tilde{P}_{\mu_n,\sigma} f = \frac{1}{n}\sum_{j=1}^{\mu_n} \sigma_j Z_{j,f}$$

for an iid sequence $\sigma_1, \ldots, \sigma_{\mu_n}$ such that $P(\sigma_i = \pm 1) = 1/2$. We have

$$\begin{aligned}
&\mathbf{P}\left(\sup_{f\in\mathbf{F}}|\tilde{P}_{\mu_n} f| \ge \varepsilon b_n\right)\\
&\quad\le 2\mathbf{P}\left(\sup_{f\in\mathbf{F}}|\tilde{P}_{\mu_n,\sigma} f| \ge \varepsilon b_n\right)\\
&\quad\le 2\mathbf{P}\left(\sup_{f\in\mathbf{F}}\left|\tilde{P}_{\mu_n,\sigma}(f - \pi_n f)\right| + \sup_{f\in\mathbf{F}}\left|\tilde{P}_{\mu_n,\sigma}(\pi_n f)\right| \ge \varepsilon b_n\right)\\
&\quad\le 2\mathbf{P}\left(\sup_{f\in\mathbf{F}}\tilde{P}_{\mu_n}|f - \pi_n f| + \sup_{f\in\mathbf{F}}\left|\tilde{P}_{\mu_n,\sigma}(\pi_n f)\right| \ge \varepsilon b_n\right)\\
&\quad\le 2\mathbf{P}\left(\sup_{f\in\mathbf{F}}\tilde{P}_{\mu_n,\sigma}|\pi_n f| \ge \frac{\varepsilon b_n}{2}\right)\\
&\quad\le 2\mathbf{P}\left(\left\{\sup_{f\in\mathbf{F}}|\tilde{P}_{\mu_n,\sigma}\pi_n f| \ge \frac{\varepsilon b_n}{2}\right\} \cap A_n \cap B_n\right) + 2\mathbf{P}(A_n^c) + 2\mathbf{P}(B_n^c).
\end{aligned}$$

Note that conditioning on $\Xi_{a_n} \in B_n$, by Hoeffding's inequality,

$$P\left(|\tilde{P}_{\mu_n,\sigma}\pi_n f| \geq \frac{\varepsilon b_n}{2}\middle|\Xi_{a_n}\right) \leq \exp\left(-\frac{1}{8}\frac{n^2\varepsilon^2 b_n^2}{4\sum_{j=1}^{\mu_n} Z_{j,\pi_n f}^2}\right)$$

$$\leq \exp\left(-\frac{1}{32}\frac{n^2\varepsilon^2 b_n^2}{\nu_n n}\right)$$

$$\leq \exp\left(-\frac{\varepsilon^2 n b_n^2}{32\nu_n}\right).$$

Thus,

$$\mathbf{P}\left(\sup_{f\in\mathbf{F}}\left|\tilde{P}_{\mu_n}(\pi_n f)\right| \geq \frac{\varepsilon b_n}{2} \cap A_n \cap B_n\right)$$

$$= \int_{\{\Xi_{a_n}\in A_n\cap B_n\}} P\left(\sup_{f\in\mathbf{F}}|\tilde{P}_{\mu_n}(\pi_n f)| \geq \frac{\varepsilon b_n}{2}\middle|\Xi_{a_n}\right) P(d\Xi_{a_n})$$

$$= \int_{\{\Xi_{a_n}\in A_n\cap B_n\}} N\left(\frac{\varepsilon b_n}{2}, \tilde{\rho}_{1,\mu_n}, \mathbf{F}\right) \max_f P\left(|\tilde{P}_{\mu_n}(\pi_n f)| \geq \frac{\varepsilon b_n}{2}\middle|\Xi_{a_n}\right) P(d\Xi_{a_n})$$

$$\leq \int_{\{\Xi_{a_n}\in A_n\cap B_n\}} N\left(\frac{\varepsilon b_n}{2}, \tilde{\rho}_{1,\mu_n}, \mathbf{F}\right) \exp\left(-\frac{\varepsilon^2 n b_n^2}{32\nu_n}\right) P(d\Xi_{a_n})$$

$$\leq \int_{\{\Xi_{a_n}\in A_n\cap B_n\}} \exp\left(\frac{\varepsilon^2 n b_n^2}{64\nu_n}\right) \exp\left(-\frac{\varepsilon^2 n b_n^2}{32\nu_n}\right) P(d\Xi_{a_n})$$

$$\leq \int_{\{\Xi_{a_n}\in A_n\cap B_n\}} \exp\left(-\frac{\varepsilon^2 n b_n^2}{64\nu_n}\right) P(d\Xi_{a_n})$$

$$\leq \exp\left(-\frac{\varepsilon^2 n b_n^2}{64\nu_n}\right).$$

Let $Q_{\mu_n}(f^2) = (1/\mu_n)\sum_{j=1}^{\mu_n}[Z_{j,f}/(a_n M)]^2$. Since the summation in Q_{μ_n} is over iid blocks and $|Z_{j,f}/(a_n M)| \leq 1$, we can follow the same arguments as in the proof of Lemma 33 in Pollard [(1984), page 31] and obtain, for $\delta^2 \geq PZ_{j,f}^2/(a_n M)^2$,

$$(4.7)\quad \mathbf{P}\left(\sup_{\mathbf{F}}\left|Q_{\mu_n}(f^2)\right| > 8\delta^2\right) \leq 4P\max\left[N\left(\delta, \rho_{Q_{\mu_n}}, \mathbf{F}\right)\exp\left(-\mu_n\delta^2\right), 1\right],$$

where

$$\begin{aligned}\rho^2_{Q_{\mu_n}}(f,g) &= Q_{\mu_n}\big((f-g)^2\big) = \frac{1}{\mu_n}\sum_{j=1}^{\mu_n}\frac{Z^2_{j,f-g}}{a_n^2M^2}\\ &= \frac{1}{\mu_n a_n^2 M^2}\sum_{j=1}^{\mu_n}(Z_{j,f}-Z_{j,g})^2\\ &\le \frac{2}{\mu_n a_n M}\sum_{j=1}^{\mu_n}|Z_{j,f}-Z_{j,g}|\\ &= \frac{4}{M}\frac{1}{n}\sum_{j=1}^{\mu_n}Z_{j,|f-g|}\\ &= \frac{4}{M}\tilde{\rho}_{1,\mu_n}(f,g).\end{aligned}$$

Take ν_n such that $8\delta^2 := \nu_n n/(\mu_n a_n^2 M^2) \ge C a_n^{-r_\beta}(\log a_n)^{\delta_\beta} \ge P(Z_{j,f}/(a_n M))^2$, that is, $\delta^2 = \nu_n/(4M^2 a_n)$ and $\nu_n \ge C a_n^{1-r_\beta}(\log a_n)^{\delta_\beta}$. One has

$$\begin{aligned}\mathbf{P}(B_n^c) &= \mathbf{P}\left(\sup_{f\in\mathbf{F}}\left|\sum_{j=1}^{\mu_n}Z^2_{j,f}\right| \ge \nu_n n\right)\\ &= \mathbf{P}\left(\sup_{f\in\mathbf{F}}\left|\frac{1}{\mu_n}\sum_{j=1}^{\mu_n}\frac{Z^2_{j,f}}{a_n^2M^2}\right| \ge \frac{\nu_n n}{a_n^2\mu_n M^2}\right)\\ &= \mathbf{P}\left(\sup_{f\in\mathbf{F}}\left|\frac{1}{\mu_n}\sum_{j=1}^{\mu_n}\frac{Z^2_{j,f}}{a_n^2M^2}\right| \ge 8\delta^2\right)\\ &\le 4P\max\left[N\left(\frac{\nu_n}{16Ma_n},\tilde{\rho}_{1,\mu_n},\mathbf{F}\right)\exp\left(-\frac{\mu_n\nu_n}{8M^2a_n}\right),1\right] \quad [\text{by } (4.7)]\\ &\le 4\int_{\{\Xi_{a_n}\in A_n\}}N\left(\frac{\nu_n}{16Ma_n},\tilde{\rho}_{1,\mu_n},\mathbf{F}\right)\exp\left(-\frac{\mu_n\nu_n}{8M^2a_n}\right)P(d\Xi_{a_n}) + \mathbf{P}(A_n^c)\\ &\le 4\int_{\{\Xi_{a_n}\in A_n\}}\exp\left(\frac{\mu_n\nu_n}{16M^2a_n}\right)\exp\left(-\frac{\mu_n\nu_n}{8M^2a_n}\right)\mathbf{P}(d\Xi_{a_n}) + \mathbf{P}(A_n^c)\\ &\le 4\exp\left(-\frac{\mu_n\nu_n}{16M^2a_n}\right) + 2\mathbf{P}(A_n^c).\end{aligned}$$

To summarize, we have

$$\mathbf{P}\left(\sup_{f\in\mathbf{F}}|\tilde{P}_{\mu_n}f| > \varepsilon b_n\right) \le 2\exp\left(-\frac{\varepsilon^2 n b^2}{64\nu_n}\right) + 8\exp\left(-\frac{\mu_n\nu_n}{16M^2a_n}\right) + 8\mathbf{P}(A_n^c),$$

provided that $\mu_n b_n^2 a_n^{r_\beta} \to \infty$, and $a_n^{1-r_\beta}(\log a_n)^{\delta_\beta} \ll \nu_n$ as $n \to \infty$.

For any $0 < s < r_\beta$, let us take $a_n = [n^{1/(1+s)}]$, $\mu_n = [n^{s/(1+s)}/2]$, $\nu_n = n^{(1-s)/(1+s)}h_n$, and $b_n = \sqrt{(\nu_n/n)h_n} = n^{-s/(1+s)}h_n$ for some $h_n \geq 1$. Then, as $n \to \infty$,

$$\mu_n \beta_{a_n} = n^{s/(1+s) - r_\beta/(1+s)} \to 0, \quad \mu_n b_n^2 a_n^{r_\beta} = O(n^{(r_b - s)/(1+s)}) \to \infty,$$

$$a_n^{1-r_\beta}(\log a_n)^{\delta_\beta} = O\big(n^{(1-r_b)/(1+s)}(\log n)^{\delta_\beta}\big) \ll n^{(1-s)/(1+s)} \leq \nu_n$$

and

$$\text{(4.8)} \quad \mathbf{P}\Big(\sup_{f \in \mathbf{F}} |\tilde{P}_{\mu_n} f| \geq \varepsilon b_n\Big) < 2\exp\left(-\frac{\varepsilon^2 h^n}{64}\right) + 8\exp\left(-\frac{h_n}{16M^2}\right) + 8\mathbf{P}(A_n^c)$$

which goes to zero if $h_n \to \infty$ and $\mathbf{P}(A_n^c) \to 0$.

Note that

$$A_n = \left\{\Xi_{a_n} : \max\left(\frac{64}{\varepsilon^2}\log N\left(\frac{\varepsilon n^{-s/(1+s)}h_n}{8}, \tilde{\rho}_{1,\mu_n}, \mathbf{F}\right),\right.\right.$$
$$\left.\left. 16M^2 \log N\left(\frac{n^{-s/(1+s)}h_n}{16M^2}\right), \tilde{\rho}_{1,\mu_n}, \mathbf{F}\right) \leq h_n\right\},$$

so

$$\mathbf{P}(A_n^c) \leq \mathbf{P}\left(\frac{64}{\varepsilon^2}\log N\big(\varepsilon n^{-s/(1+s)}, \tilde{\rho}_{1,\mu_n}, \mathbf{F}\big) \geq h_n\right)$$
$$+ \mathbf{P}\left(16M^2 \log N\left(\frac{n^{-s/(1+s)}}{16M^2}, \tilde{\rho}_{1,\mu_n}, \mathbf{F}\right) \geq h_n\right),$$

which, together with (4.8), implies (4.5).

(iii) This is straightforward from (ii) and the assumption (4.6). □

REMARK. The measurability issue arising from the symmetrization is taken care of in the Appendix under "Measurable Cross Sections."

PROOF OF THEOREM 3.1 (Rates of convergence). For the choices of μ_n, a_n, and b_n in Lemma 4.3 and Theorem 3.1, $\mu_n b_n = h_n \to \infty$ and $\mu_n \beta_{a_n} = o(1)$ as $n \to \infty$. Hence if we combine Lemmas 4.2, 4.3(i) and 4.3(iii), we obtain Theorem 3.1. □

PROOF OF COROLLARY 3.2. Note that, for any $s > 0$ and $h_n \geq 1$, from assumption (2.4)

$$\log N_1\big(cn^{-s/(1+s)}h_n, \rho_{1,n}, \mathbf{F}\big) \leq \log N_1\big(cn^{-s/(1+s)}, \rho_{1,n}, \mathbf{F}\big) = O_{\mathbf{P}}(\log n).$$

We can take $h_n = (\log n)^2$ in Theorem 3.1 and note that for any $0 < s < r_\beta$, we can find s' such that $s < s' < r_\beta$ and $n^{-s'/(1+s')}h_n \ll n^{-s/(1+s)}$. □

PROOF OF THEOREM 3.3 (Uniform convergence for bounded families). For any $0 < \alpha_0 < r_\beta/(1 + r_\beta)$, there is an s such that $\alpha_0 = s/(1 + s)$ and $0 < s < r_\beta$. Take $h_n' = n^{s/(1+s)}$ in Theorem 3.1. We have

$$P\Big(\sup_{f \in \mathbf{F}} |P_n f - Pf| > \varepsilon n^{-s/(1+s)} h_n\Big) = P\Big(\sup_{f \in \mathbf{F}} |P_n f - Pf| > \varepsilon\Big) \to 0$$

provided that

$$\log N_1\big(cn^{-s/(1+s)} h_n, \rho_{1,n} \mathbf{F}\big) = \log N_1(c, \rho_{1,n}, \mathbf{F}) = o_{\mathbf{P}}(h_n) = o_{\mathbf{P}}(n^{s/(1+s)}),$$

which is our hypothesis (3.3) □

Before we prove the uniform convergence theorem for an index class with nonconstant envelope function, we first recall a result from Doob [(1953), page 465] on the law of large numbers for strictly stationary sequences.

4.4 DEFINITION. A strictly stationary sequence $\underline{Y} = (Y_1, \dots)$ is called *metrically transitive* if all invariant sets of its shift transformation T (i.e., $TY_i = Y_{i+1}$) have probability 0 or 1.

4.5 PROPOSITION. *Suppose that $\underline{X}$ is strictly stationary with the stationary distribution P. If $\underline{X}$ is metrically transitive and $F \in L^1(P)$, we have*

$$\frac{1}{n} \sum_{i=1}^{n} F(X_i) \to PF \quad \textit{in probability}.$$

PROOF OF THEOREM 3.4 (Uniform convergence for general families)

(a) Assume $F \equiv M$. This case is covered by Theorem 3.3.

(b) For any given M, take $\mathbf{F}_M = \{f_M = fI_{\{F<M\}}\colon f \in G\}$. Then,

$$|P_n f - Pf| \le |P_n f_M - Pf_M| + P_n(FI_{\{F>M\}}) + P(FI_{\{F>M\}}).$$

This implies

$$\sup_{\mathbf{F}} |P_n f - Pf| \le \sup_{\mathbf{F}_M} |P_n f_M - Pf_M| + P_n(FI_{\{F>M\}}) + P(FI_{\{F>M\}}).$$

For any fixed $\varepsilon > 0$, take M such that $P(FI_{\{f>M\}}) < \varepsilon$. Since $\{X_i\}$ is mixing, so is $\{(FI_{\{F>M\}})(X_i)\}$; hence it is metrically transitive by Corollary 17.1.1 of Ibragimov and Linnik (1971). Thus Proposition 4.5 gives the law of large numbers for $FI_{\{f>M\}}$ with this M. Therefore, both $P(FI_{\{F>M\}})$ and $P_n(FI_{\{F>M\}})$ can be bounded by $3P(FI_{\{F>M\}})$ in probability when n is sufficiently large. Note that for this fixed M the supremum over $\mathbf{F}_M$ tends to 0 in probability by part (a) and the fact that the covering number of $\mathbf{F}_M$ is bounded by that of $\mathbf{F}$. □

APPENDIX

Measurability. In this Appendix we deal with the measurability issues raised in earlier sections and for completeness we also deal with the measurability issues related to the CLT [cf. Arcones and Yu (1994)]. We follow the order of Pollard [(1984), Appendix C], and refer to it when necessary to avoid repetition. Moreover, let DM denote Dellacherie and Meyer (1978), and DM II.32 means Dellacherie and Meyer [(1978), Chapter II, Section 32], for example.

Assume that $\underline{X} = (X_1, X_2, \dots, X_n, \dots)$ is a measurable map from $(\Omega, \mathbf{E}, P)$ to $(R^\infty, \mathbf{B}^\infty)$.

It is noted that the original space $(\Omega, \mathbf{E})$ might not be rich enough to support the independent block sequences Ξ_n. We can, however, always regard $\underline{X}$ as a measurable map from $(R^\infty, \mathbf{B}^\infty)$ to itself, and on $(R^\infty, \mathbf{B}^\infty)$ the independent block sequence exists. Note that the probabilities of events of interest regarding the original sequence take the same values regardless of which space $[(\Omega, \mathbf{E})$ or $(R^\infty, \mathbf{B}^\infty)]$ we view as the domain of the measurable map $\underline{X}$. Therefore, we can work with $(R^\infty, \mathbf{B}^\infty)$ to construct the block sequence Ξ_n, connect it with the original sequence $\underline{X}$ and bound the probabilities regarding $\underline{X}$ by those regarding Ξ_n [depending on (μ_n, a_n)]. For clarity, let $(\tilde{\Omega}, \tilde{\mathbf{E}}) \equiv (R^\infty, \mathbf{B}^\infty)$. Then $\underline{X}$ may be regarded as a measurable map from $(\tilde{\Omega}, \tilde{\mathbf{E}})$ to $(R^\infty, \mathbf{B}^\infty)$. For simplicity, we will use the same $\underline{X}$ to denote either the map from the original $(\Omega, \mathbf{E})$ or from the new $(\tilde{\Omega}, \tilde{\mathbf{E}})$.

Pollard's measurability treatment involves the probability measure of the iid sequence under consideration so he requires the probability space $(\Omega, \mathbf{E}, P)$ to be complete. In our case, however, we have more than one measure: the measure of the original sequence plus the measures of the independent block sequences. So we prefer to take a different approach in which the measurability problem is handled independently of any particular probability measure.

For any measure space $(\Omega, \mathbf{E})$, denote by $\mathbf{A}(\mathbf{E})$ the class of $\mathbf{E}$-analytic sets of Ω, and by $(\Omega, \mathbf{E}^u)$ the *universal completion* of $(\Omega, \mathbf{E})$ (DM II.32). Then the following hold:

(a) $\mathbf{A}(\mathbf{E}) \subset \mathbf{E}^u$ (straightforward from DM III.33 and the definition of universal completion);

(b) every probability measure λ on $\mathbf{E}$ can be uniquely extended to a probability measure λ^u on $\mathbf{E}^u$, and the mapping $\lambda \to \lambda^u$ is a bijection of the set of probability measures on $\tilde{\mathbf{E}}$ onto the set of probability measures on $\mathbf{E}^u$ (DM II.32);

(c) if $(\Omega', \mathbf{E}')$ is a probability space and X is a measurable mapping of $(\Omega, \mathbf{E})$ into $(\Omega', \mathbf{E}')$, then X is a measurable mapping of $(\Omega, \mathbf{E}^u)$ into $(\Omega', \mathbf{E}'^u)$ (DM II.32).

Observe that the probability $\mathbf{P}$ used in previous sections can now be taken as the unique extension, from $\mathbf{E}$ to $\mathbf{E}^u$ (or $\tilde{\mathbf{E}}$ to $\tilde{\mathbf{E}}^u$), of the probability of the original sequence $\underline{X}$; or the unique extension, from $\mathbf{E}$ to $\mathbf{E}^u$, of the constructed independent block sequence Ξ_n. Thus when $\mathbf{F}$ is suitably restricted as below, the convergence rate theorem and the CLT hold on $(\Omega, \mathbf{E}^u)$. As also will be seen below, the suprema of interest are $\mathbf{E}^u$ (or $\tilde{\mathbf{E}}^u$)-measurable. Hence by (b) and (c) the probabilities based on these suprema concerning the original sequence $\underline{X}$ have the same value (as numbers) no matter which space, $(\Omega, \mathbf{E})$ or $(\tilde{\Omega}, \tilde{\mathbf{E}})$, we put $\underline{X}$ on.

Now we are ready to show that arguments similar to Pollard's will give $\tilde{\mathbf{E}}^u$-measurability of suprema from the original sequence or the independent block sequence; in addition, cross sections needed for the symmetrization of the independent block sequence will be shown to be $\tilde{\mathbf{E}}^u$-measurable. We will also show that the empirical process E_n is $\tilde{\mathbf{E}}^u/\mathbf{B}^P$-measurable. Moreover, since $\underline{X}$ is a measurable map from $(\Omega, \mathbf{E}^u)$ to $(\tilde{\Omega}, \tilde{E}^u)$, by (c), the $\tilde{\mathbf{E}}^u$-measurability of functions of the original sequence guarantees their $\mathbf{E}^u$-measurability.

Permissible Class. Suppose that the class $\mathbf{F}$ is indexed by a parameter t that ranges over some set T. That is, $\mathbf{F} = \{f(\cdot, t): t \in T\}$. We may take $T = \mathbf{F}$ as a subset of $L^1(P)$. Assume that T is a separable metric space with the Borel σ-field $\mathbf{B}(T)$.

A.1 DEFINITION. The class $\mathbf{F}$ is permissible if it can be indexed by a T in such a way that the following hold:

(i) the function $f(\cdot,\cdot)$ is $\mathbf{B} \times \mathbf{B}(T)$-measurable as a function from $R^1 \times T$ into the real line;

(ii) T is isomorphic to an analytic subset of a compact metrizable space $\overline{T}$ [or, equivalently, T is a Souslin measurable space (DM III.16)].

Note that (i) is needed for the measurability of suprema, but (ii) is needed for symmetrization.

Measurable Suprema. Assume that $\mathbf{F}$ is permissible and that $Pf_t < \infty$ for each t. For any $s \in R^n$, let

$$g(s,t) = n^{-1} \sum_{i=1}^{n} [f(s_i, t) - Pf_t].$$

Since $Pf_t < \infty$ and $f(\cdot,\cdot)$ is $\mathbf{B} \times \mathbf{B}(T)$-measurable, then $Pf_t \equiv n^{-1}\Sigma_{i=1}^n Pf_t$ is $\mathbf{B}(T)$-measurable by Fubini's theorem. Moreover, by (ii) of the definition of a permissible class, $n^{-1}\Sigma_{i=1}^n f(s_i, t)$ is $\mathbf{B}^n \times \mathbf{B}(T)$-measurable. Thus $g(s,t)$ is a $\mathbf{B}^n \times \mathbf{B}(T)$-measurable real function on $R^n \times T$. Write $G(s) = \sup_t g(s,t)$. For any $a \in R^1$, let $H_a = \{s,t): g(s,t) > a\}$. Since g is $\mathbf{B}^n \times \mathbf{B}(T)$ measurable,

$$H_a \in \mathbf{B}^n \times \mathbf{B}(T) \subset \mathbf{A}(\mathbf{B}^n \times \mathbf{B}(T)).$$

Thus $\{G(s) > a\}$ as the projection of H_a onto R^n is also analytic, that is, $\{G(s) > a\} \in \mathbf{A}(\mathbf{B}^n) \subset \mathbf{A}(\mathbf{B}^\infty) \subset (\mathbf{B}^\infty)^u \equiv \tilde{\mathbf{E}}^u$. Moreover, $\underline{X}_n$ is measurable from

$(\tilde{\Omega}, \tilde{\mathbf{E}}^u)$ to $(R^n, (\mathbf{B}^n)^u)$ by (c) and hence

$$\{\omega: G(\underline{X}_n(\omega)) > a\} = \underline{X}_n^{-1}(\{G(s) > a\}) \in \tilde{\mathbf{E}}^u,$$

that is, $\omega \to G(\underline{X}_n(\omega))$ is $\tilde{\mathbf{E}}^u$-measurable, hence $\mathbf{E}^u$-measurable, and similarly for $G(\Xi_n(\omega))$ and the suprema from the even blocks, the odd blocks and the remainder block. Note that the suprema over $[\delta]$ in the equicontinuity lemma are also $\tilde{\mathbf{E}}^u$- (or $\mathbf{E}^u$-) measurable and this is needed in Arcones and Yu (1994). Here we will only demonstrate that the supremum related to the empirical process $E_n = \sqrt{n}(P_n - P)$ is $\tilde{\mathbf{E}}^u$-measurable. Let

$$g(s,t,t') = n^{-1/2} \sum_{i=1}^{n} [f(s_i,t) - Pf_t - f(s_i,t') + Pf_{t'}].$$

This is $\mathbf{B}^n \times \mathbf{B}(T) \times \mathbf{B}(T)$-measurable. Thus $V_\delta \equiv \{(t,t'): \rho_1(f_t, f_{t'}) < \delta\}$ is $\mathbf{B}(T) \times \mathbf{B}(T)$-measurable by Fubini's theorem. If we define

$$H_\alpha = \{(s,t,t'): g(s,t,t') > \alpha\},$$
$$G(s) = \sup_{t,t': \rho_1(f_t, f_{t'}) < \delta} g(s,t,t'),$$

then $\{G > \alpha\}$ is the projection onto R^n of $H_\alpha \cap V_\delta$ which is $\mathbf{B}^n \times \mathbf{B}(T) \times \mathbf{B}(T)$-measurable, therefore analytic. Thus $\{G > \alpha\}$ is $\mathbf{A}(\mathbf{b}^n)$-analytic, so it is $\tilde{\mathbf{E}}^u$-measurable, and hence $\mathbf{E}^u$-measurable by (c).

It should be clear that extending the probability measure of $\underline{X}$ from $\mathbf{E}$ (or $\tilde{\mathbf{E}}$) to $\mathbf{E}^u$ (or $\tilde{\mathbf{E}}^u$) does not change the α-, β-, or ϕ-mixing coefficients, because for any $A \in \tilde{\mathbf{E}}$ there is an $a' \in \tilde{\mathbf{E}}^u$ such that $P(A) = P(A')$, and all three mixing coefficients are defined in terms of probabilities.

Measurable Cross Sections. In the symmetrization of the empirical process from the independent block sequence, we needed the following assertion:

> If $\mathbf{F}$ is a permissible class, then for a stochastic process $Z(\omega, t)$ on $(\tilde{\Omega}, \tilde{\mathbf{E}})$, there exists a $\tilde{\mathbf{E}}^u/\mathbf{B}(T)$-measurable ($T$-valued) τ such that $Z(\omega, \tau(\omega)) > \varepsilon$ whenever $\sup_t |z(\omega, t)| > \varepsilon$.

When Definition 1(ii) is satisfied and since both T and $\tilde{\Omega} \equiv R^\infty$ are metrizable, by DM III.19, every Borel [or $\mathbf{B}(T) \times \mathbf{B}^\infty$-analytic] subset of $T \times R^\infty$ is Souslin (because T and R^1 are both Souslin and a countable product of Souslin spaces in Souslin). So from a theorem in DM (Appendix III.81) and Remark (a) we have the following:

(d) if H is a $\mathbf{B}(T) \times \mathbf{B}^\infty$-analytic subset of $T \times \tilde{\Omega} \equiv T \times R^\infty$, there exists an $\tilde{\mathbf{E}}^u/\mathbf{B}(T)$-measurable mapping τ of $\tilde{\Omega}$ into T such that, for all ω in the projection of H onto $\tilde{\Omega}$, $(\tau(\omega), \omega)$ belongs to H (τ is a complete cross section of H).

Then our assertion can be proved similarly to Pollard (1984), except that we use our (d) above instead the (d) stated there.

For any given $\varepsilon > 0$, let

$$A = \left\{\omega \colon \sup_t |Z(\omega, t)| < \varepsilon\right\},$$

$$B = \{(\omega, t) \colon |Z(\omega, t)| > \varepsilon\}.$$

The set A belongs to $\tilde{\mathbf{E}}^u$, and the set B belongs to $\tilde{\mathbf{E}} \times \mathbf{B}(T) \equiv \mathbf{B}^\infty \times \mathbf{B}(T)$. Hence by (d), we can choose an $\tilde{\mathbf{E}}^u$-measurable cross section τ_0. Set τ equal to τ_0 on A, and outside A set it equal to t_0 for a fixed element t_0 of T. Since A is $\tilde{\mathbf{E}}^u$-measurable, τ is $\tilde{\mathbf{E}}^u$-measurable. Moreover,

$$A = \left\{\omega \colon \sup_t |Z(\omega, t)| > \varepsilon\right\}$$

and on A $(\omega, \tau(\omega))$ belongs to B; that is,

$$Z(\omega, \tau(\omega)) > \varepsilon$$

as required. So the assertion is proved.

Now symmetrization can be carried out by taking two independent copies Z and Z' on a product space $\tilde{\Omega} \times \tilde{\Omega}'$ with the product measure $\mathbf{P} \times \mathbf{P}'$ on $\tilde{\mathbf{E}}^u \times (\tilde{\mathbf{E}}')^u$,

$$Z(\omega, \omega', t) = Z(\omega, t),$$

$$Z'(\omega, \omega', t) = Z'(\omega', t).$$

The τ constructed above depends only on ω. For all ω,

$$\mathbf{P}'\{\omega' \colon |Z'(\omega', \tau(\omega))| \le \alpha\} \ge \beta.$$

The rest of the proof goes through as in Pollard [(1984), page 14, Symmetrization Lemma 8], with Fubini's theorem formalizing the conditioning argument.

Covering Numbers. We may interpret a condition like

$$\log N(\delta, P_n, \mathbf{F}) = o_p(b_n) \quad \text{as } n \to \infty$$

to mean that $\mathbf{P}\{Z_n > b_n \varepsilon\} \to 0$ and $n \to \infty$ for some $\tilde{\mathbf{E}}$- (or $\mathbf{E}^u$-) measurable random variable Z_n greater than $\log N(\delta, P_n, \mathbf{F})$. Alternatively, we may use packing numbers instead of covering numbers as in Pollard [(1984), Appendix C]: Define $M(\delta, P_n, \mathbf{F})$ as the smallest m for which there exist functions $f_1, \dots, f_m$ in $\mathbf{F}$ with $P_n|f_j - f_k| > \delta$ for $j \ne k$. This is an $\tilde{\mathbf{E}}$- (therefore $\mathbf{E}^u$-) measurable function of ω since the set

$$\{\omega \colon M(\delta, P_n(\omega, \cdot), \mathbf{F}) \ge m\}$$

equals the projection on $\tilde{\Omega}$ of the $\tilde{\mathbf{E}} \times \mathbf{B}(t)$-measurable set

$$\left\{(\omega, t) \colon \min_{j \ne k} n^{-1} \sum_{i=1}^{n} |f(X_i(\omega), t_j) - f(X_i(\omega), t_k)| > \delta\right\}.$$

The packing numbers are closely related to covering numbers:

$$M(2\delta, P_n, \mathbf{F}) \le N(\delta, P_n, \mathbf{F}) \le M(\delta, P_n, \mathbf{F}).$$

Similarly, we can use the corresponding packing numbers for P_{μ_n} and $\tilde{P}_{\mu_n}$. They are $\tilde{\mathbf{E}}$-measurable, hence $\tilde{\mathbf{E}}^u$-measurable (or $\mathbf{E}^u$-measurable).

The Function Space $\mathscr{X}$. This session deals with the measurability issue related to the weak convergence of the empirical process E_n to its limit process. Let us first introduce the space $C(\mathbf{F}, P) \subseteq \mathscr{X}$ of real bounded functionals on $\mathbf{F}$, equipped with the uniform norm $\|x\| = \sup_{\mathbf{F}}|x(f)|$ [cf. Pollard (1984)].

DEFINITION. $C(\mathbf{F}, P)$ is the set of all functionals $x(\cdot)$ of $\mathbf{F}$ which are uniformly continuous with respect to the $L^1(P)$ seminorm ρ_1 on $\mathbf{F}$. That is, for each functional $x(\cdot)$ and $\varepsilon > 0$, there should exist a $\delta > 0$ such that $|x(f) - x(g)| < \varepsilon$ whenever $\rho_1(f - g) < \delta$. Define $\mathbf{B}^P$ as the smallest σ-field on $\mathscr{X}$ which contains all the closed balls in terms of $\|\cdot\|$ with centers in $C(\mathbf{F}, P)$ and makes all the finite-dimensional projections measurable.

Denote by E_P the limiting P-bridge process indexed by $\mathbf{F}$, which is a tight, Gaussian random element of $\mathscr{X}$ whose sample paths all belong to $C(\mathbf{F}, P)$ and has a covariance structure as follows:

$$\mathrm{Cov}\big(E_p(f), E_p(g)\big) = Pf(X_1)g(X_1) + 2\sum_1^{\infty} P[f(X_1)g(X_{k+1})]$$

for all $f, g \in \mathbf{F}$.

Assume that $\mathbf{F}$ is permissible and separable under the ρ_1 seminorm. Using essentially the same arguments as in Pollard (1984), we can prove that E_n is $\mathbf{E}^u/\mathbf{B}^P$-measurable. This is needed to make sense of the weak convergence of E_n to E_p as random elements of $(\mathscr{X}, \mathbf{B}^P)$. To be self-contained again, we repeat Pollard's arguments:

Recall that $C(\mathbf{F}, P)$ is the set of functions in $\mathscr{X}$ that are uniformly continuous with respect to the ρ_1 seminorm on $\mathbf{F}$ and that the σ-field $\mathbf{B}^P$ is the smallest for which the following hold:

1. all the closed balls (for $\|x\| \equiv \sup_{\mathbf{F}}|x(f)|$) belong to $\mathbf{B}^P$;
2. all the finite-dimensional projections are $\mathbf{B}^P$-measurable.

Because each $E_n(\cdot, f)$ is a real random variable, the finite-dimensional projections create no difficulty for $\mathbf{E}^u/\mathbf{B}^P$-measurability. So we only need to show that $\{\omega: \|E_n(\omega, \cdot) - x(\cdot)\| \le r\}$ belongs to $\mathbf{E}^u$. Introduce T as in Definition 1. Since $\mathbf{F}$ is separable under ρ_1, Problem 1 of Pollard [(1984), page 200] establishes the $\mathbf{B}(T)$-measurability of $x(f_t)$. So for $s \in R^n$ the function

$$g(s,t) = \left| n^{-1/2} \sum_{i=1}^{n} [f(s_i, t) - x(f_t)] \right|$$

is $\mathbf{B}^n \times \mathbf{B}(T)$-measurable. The argument in "Measurable Suprema" establishes $\mathbf{E}^u$-measurability of $\sup_t g(X(\omega), t)$, which equals $\|E_n(\omega, \cdot) - x(\cdot)\|$.

Acknowledgments. This research is based on part of the author's Ph.D. dissertation submitted to the University of California at Berkeley. The author is deeply indebted to Professor Lucien Le Cam for his guidance, invaluable discussions and encouragement during the course of the work. The author also wants to express her thanks to Professor Deborah Nolan for many helpful discussions and suggestions for improvement, and to Professor Terry Speed for commenting on the drafts. Special thanks are due to Professor Walter Philipp for pointing out the differences between this work and the related work of his and Massart's, and to Dr. Miguel Arcones and two referees for helpful comments.

REFERENCES

ANDREWS, D. and POLLARD, D. (1993). An introduction to functional central limit theorems for dependent stochastic processes. *Internat. Statist. Rev.* **61**. To appear.

ARCONES, M. A. and YU, B. (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *J. Theoret. Probab.* To appear.

ATHREYA, K. B. and PANTULA, S. G. (1986). Mixing properties of Harris chains and autoregressive processes. *J. Appl. Probab.* **23** 880–892.

BERNSTEIN, S. N. (1927). Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dependantes. *Math. Ann.* **97** 1–59.

BRADLEY, R. C. (1983). Absolute regularity and functions of Markov chains. *Stochastic Process. Appl.* **14** 67–77.

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

DELLACHERIE, C. and MEYER, P. A. (1978). *Probabilities and Potential. Part A.* North-Holland, Amsterdam. [Translation of the 1975 French second edition of the first four chapters from Meyer (1966).]

DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

DUDLEY, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899–929. [Correction: (1979) **7** 909–911.]

DUDLEY, R. M. and PHILIPP, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. Verw. Gebiete* **62** 509–552.

EBERLEIN, E. (1984). Weak convergence of partial sums of absolutely regular sequences. *Statist. Probab. Lett.* **2** 291–293

GINÉ, E. and ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* **12** 929–989.

IBRAGIMOV, I. A. and LINNIK, YU. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters–Noordhoff, Groningen.

IBRAGIMOV, I. A. and ROZANOV, Y. A. (1978). *Gaussian Random Processes. Appl. Math.* **9**. Springer, New York.

KOLMOGOROV, A. N. and TIHOMIROV, V. M. (1959). ε–entropy and ε–capacity of sets in functional spaces. *Uspekhi Mat. Nauk.* **14** 3–86. (*Amer. Math. Soc. Transl. Ser. 2* **17** 277–364.)

LE CAM, L. M. (1984). A remark on empirical measures. In *A Festschrift for Erich L. Lehmann* (P. Bickel, K. Doksum and J. Hodges, eds.) 305–327. Wadsworth, Belmont, CA.

LEVENTAL, S. (1988). Uniform limit theorems for Harris recurrent Markov Chains. *Probab. Theory Related Fields* **80** 101–118.

LEVENTAL, S. (1989). A uniform CLT for uniformly bounded families of martingale differences *J. Theoret. Probab.* **2** 271–287.

MASSART, P. (1988). Invariance principles for empirical processes: The weakly dependent case. Ph.D. dissertation, Univ. Paris.

MOKKADEM, A. (1988). Mixing properties of ARMA processes. *Stochastic Process. Appl.* **29** 309–315.

NOBEL, A. and DEMBO, A. (1993). A note on uniform laws of averages for dependent processes. *Statist. Probab. Lett.* **17** 169–172.

PHAM, T. D. and TRAN, L. T. (1985). Some mixing properties of time series models. *Stochastic Process. Appl.* **19** 297–303.

PHILIPP, W. (1984). Invariance principles for sums of mixing random elements and the multivariate empirical process. In *Limit Theorems in Probability and Statistics. Colloq. Math. Soc. János Bolyai* **36**. North-Holland, Amsterdam.

PHILIPP, W. (1986). Invariance principles for independent and weakly dependent random variables. In *Dependence in Probability and Statistics: A Survey of Recent Results* (E. Eberlein and M. S. Taqqu, eds.) 225–268. Birkhäuser, Boston.

POLLARD, D. (1982). A central limit theorem for empirical processes. *J. Austral. Math. Soc. Ser. A* **33** 235–248.

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

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.

VOLKONSKII, V. A. and ROZANOV, YU. A. (1959). Some limit theorems for random functions I. *Theory Probab. Appl.* **4** 178–197.

WITHERS, C. S. (1981). Conditions for linear processes to be strong-mixing. *Z. Wahrsch. Verw. Gebiete* **57** 477–480

YU, B. (1990a). Some results on empirical processes and stochastic complexity. Ph.D. dissertation, Univ. California, Berkeley.

YU, B. (1990b). Rates of convergence and central limit theorems for empirical processes of stationary mixing sequences. Technical Report 260, Dept. Statistics, Univ. California, Berkeley.

YUKICH, J. E. (1986). Rates of convergence for classes of functions: The non-i.i.d. case. *J. Multivariate Anal.* **20** 175–189.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706