The Annals of Probability
1992, Vol. 20, No. 4, 1983–2030

UNIFORM DONSKER CLASSES OF FUNCTIONS

BY ANNE SHEEHY[1] AND JON A. WELLNER[2]

Eidgenössiche Technische Hochschule and University of Washington

A class $\mathscr{F}$ of measurable functions on a probability space $(A, \mathbb{A}, P)$ is called a P-Donsker class and we also write $\mathscr{F} \in \mathrm{CLT}(P)$, if the empirical processes $\mathbb{X}_n^P \equiv \sqrt{n}(\mathbb{P}_n - P)$ converge weakly to a P-Brownian bridge G_P having bounded uniformly continuous sample paths almost surely. If this convergence holds for every probability measure P on $(A, \mathbb{A})$, then $\mathscr{F}$ is called a universal Donsker class and we write $\mathscr{F} \in \mathrm{CLT}(\mathbf{M})$, where $\mathbf{M} \equiv$ {all probability measures on $(A, \mathbb{A})$}. If the convergence holds uniformly in all P, then $\mathscr{F}$ is called a uniform Donsker class and we write $\mathscr{F} \in \mathrm{CLT}_u(\mathbf{M})$. For many applications the latter concept is too restrictive and it is useful to focus instead on a fixed subcollection $\mathscr{P}$ of the collection $\mathbf{M}$ of all probability measures on $(A, \mathbb{A})$. If the empirical processes converge weakly to G_P uniformly for all $P \in \mathscr{P}$, then we say that $\mathscr{F}$ is a $\mathscr{P}$-uniform Donsker class and write $\mathscr{F} \in \mathrm{CLT}_u(\mathscr{P})$. We give general sufficient conditions for the $\mathscr{P}$-uniform Donsker property and establish basic equivalences in the uniform (in $P \in \mathscr{P}$) central limit theorem for $\mathbf{X}_n$, including a detailed study of the equivalences to the "functional" or "process in n" formulations of the CLT. We give applications of our uniform convergence results to sequences of measures $\{P_n\}$ and to bootstrap resampling methods.

0. Introduction. Limit theory for empirical processes has grown and developed enormously in the past ten years. Most of the recently developed central limit theorems for general empirical processes are for a fixed underlying probability measure P on the given sample space $(A, \mathbb{A})$. Dudley (1987) has investigated classes of functions $\mathscr{F}$ for which the central limit theorem holds for all probability measures P on $(A, \mathbb{A})$, and calls such classes *universal Donsker classes*. Giné and Zinn (1991) have studied classes $\mathscr{F}$ for which the central limit theorem holds uniformly in all P on $(A, \mathbb{A})$ and call such classes *uniform Donsker classes*. For many applications in statistics, uniformity of the convergence in P is of interest and importance, but requiring convergence uniformly in all P is too restrictive. Instead, it is useful to focus on some fixed subcollection $\mathscr{P}$ of all measures and ask that the convergence be uniform over P in this subcollection. We call such a class of functions $\mathscr{F}$ a *$\mathscr{P}$-uniform Donsker class*.

Received March 1990; revised June 1991.

[1]Research supported by NSF Grant DMS-87-08083 and NSF Switzerland Grant 21/29013.90. On leave from Stanford University.

[2]Research supported by NSF Grants DMS-84-00893 and DMS-87-23011, and the John Simon Guggenheim Foundation.

AMS 1980 *subject classifications*. Primary 60F05, 62G05, 62G30; secondary 60B12, 60F17.

Key words and phrases. Bootstrap, central limit theorem, empirical process, functional central limit theorem, Gaussian processes, parametric bootstrap, nonparametric bootstrap, regular estimators, sequential empirical process, uniformity, uniform integrability, weak approximation.

Our goal in this paper is to study various equivalent conditions for the $\mathscr{P}$-uniform Donsker property of a class of functions $\mathscr{F}$ and to indicate briefly how the properties of such classes can be exploited to study the behavior of bootstrap methods.

Here is how the paper is organized: We begin in Section 1 with a review of equivalences in the central limit theorem (CLT) for a fixed P. Most of the results in this section are known, but we make one addition to the list of equivalent formulations. Section 2 presents corresponding equivalences in the $\mathscr{P}$-uniform CLT and useful sufficient conditions. Section 3 gives applications to sequences of probability measures $\{P_n\}$, with regularity properties of the empirical measure as corollaries.

Section 4 contains applications to bootstrap methods in statistics. We show that for certain $\mathscr{P}$-uniform Donsker classes $\mathscr{F}$, the bootstrap works for the empirical process if the bootstrap sample size m and the original sample size n both go to infinity (in an arbitrary way). Proofs for the main results in Section 2 are presented in Sections 5 and 6.

1. *P*-Donsker classes $\mathscr{F}$. This section introduces notation and terminology and reviews what is known about equivalences in the definition of the central limit theorem (CLT) for the empirical process for a fixed probability measure P.

Let P be a fixed probability measure on the measurable space $(A, \mathbb{A})$. Let G_P be a *P-Brownian bridge process* and let W_P be a *P-Brownian motion process*; that is, G_P and W_P are Gaussian processes indexed by $L^2(A, \mathbb{A}, P) \equiv L^2(P)$ with mean zero and covariances

$$\mathrm{Cov}(G_P(f), G_P(g)) = P(fg) - P(f)P(g) \quad \text{for all } f, g \in L^2(P),$$
$$\mathrm{Cov}(W_P(f), W_P(g)) = P(fg) \quad \text{for all } f, g \in L^2(P),$$

where $P(f) \equiv \int f\,dP$. We let

$$\rho_P^2(f, g) \equiv E(G_P(f) - G_P(g))^2 = \mathrm{Var}_P(f(X) - g(X))$$

and

$$e_P^2(f, g) \equiv E(W_P(f) - W_P(g))^2 = E_P(f(X) - g(X))^2,$$

where X has distribution P. G_P is *coherent* if each sample function $G_P(\cdot)(\omega)$ is bounded and uniformly continuous on $\mathscr{F}$ with respect to ρ_P; and a collection $\mathscr{F} \subset L^2(P)$ is *P-pregaussian* ($\mathscr{F} \in \mathrm{PG}(P)$) (earlier this was called G_PBUC) if and only if a coherent G_P process exists. Let

$$(\Omega, \Sigma, \mathrm{Pr}_P) \equiv (A^\infty, \mathbb{A}^\infty, P^\infty) \times ([0,1], \mathbf{B}, \text{Lebesgue}),$$

and let $X_1, X_2, \ldots$ be the coordinate functions on $(A^\infty, \mathbb{A}^\infty, P^\infty)$ so that $X_1, X_2, \ldots$ are iid P. The *empirical measure* of the first n X's is

$$\mathbb{P}_n \equiv n^{-1} \sum_{i=1}^{n} \delta_{X_i},$$

and the *empirical process* $\mathbb{X}_n \equiv \mathbb{X}_n^P$ is

$$\mathbb{X}_n^P = \sqrt{n}\,(\mathbb{P}_n - P).$$

Of course for any finite collection $\{f_1, \ldots, f_m\} \subset L^2(P)$,

$$(\mathbb{X}_n(f_1), \ldots, \mathbb{X}_n(f_m)) \to_d (G_P(f_1), \ldots, G_P(f_m)) \\ \sim N_m\big(0, \big(P(f_i f_j) - P(f_i)P(f_j)\big)\big),$$

but our main concern in this section is the central limit theorem for larger collections $\mathscr{F} \subset L^2(P)$ for a fixed P. We assume that $\sup_{f \in \mathscr{F}} |f(x) - P(f)| < \infty$ for all $x \in A$; then we view $\mathbb{X}_n$ as an element in $l^\infty(\mathscr{F})$, the space of all bounded functions from $\mathscr{F}$ to R. We let $C(\mathscr{F}, \rho_P)$ denote the set of all functions x in $l^\infty(\mathscr{F})$ that are uniformly continuous with respect to ρ_P.

Here are several definitions related to the central limit theorem for $\mathbb{X}_n$.

DEFINITION 1.1. $\mathscr{F} \in$ CLT(P) (or $\mathscr{F}$ is P-Donsker) if $\mathscr{F}$ is P-pregaussian and $\mathbb{X}_n \Rightarrow \mathbb{X} \sim G_P$ in $l^\infty(\mathscr{F})$.

Here the weak convergence $\Rightarrow$ is in the sense of Hoffman-Jørgensen (1984); see, for example, Anderson and Dobrić (1987), Anderson (1985), Dudley (1985) and van der Vaart and Wellner (1990):

$$E^* h(\mathbb{X}_n) \to Eh(\mathbb{X}) \quad \text{for all } h \in C_b(l^\infty(\mathscr{F})),$$

where E^* denotes the upper (or outer) integral computed under $\Pr_P$ and $C_b(l^\infty(\mathscr{F}))$ is the set of all bounded $\|\ \|_{\mathscr{F}}$-continuous real-valued functions on $l^\infty(\mathscr{F})$. We suppress the dependence of $\mathbb{X}_n \equiv \mathbb{X}_n^P$ on P since it is fixed.

We now let

$$\mathscr{F}' \equiv \{f - g: f, g \in \mathscr{F}\}, \qquad (\mathscr{F}')^2 \equiv \{(f-g)^2: f, g \in \mathscr{F}\},$$

and for a pseudometric d on $L^2(P)$ and $\delta > 0$,

$$\mathscr{F}'(\delta, d) \equiv \{(f, g) \in \mathscr{F} \times \mathscr{F}: d(f, g) \le \delta\}.$$

Then for any real-valued function ψ on $\mathscr{F}$,

$$\|\psi\|_{\mathscr{F}} \equiv \sup_{f \in \mathscr{F}} |\psi(f)|, \qquad \|\psi\|_{\mathscr{F}'(\delta, d)} \equiv \sup_{(f,g) \in \mathscr{F}'(\delta, d)} |\psi(f) - \psi(g)|.$$

DEFINITION 1.2. We say that $\{\mathbb{X}_n\}_{n \ge 1}$ is *asymptotically equicontinuous with respect to* d on $\mathscr{F}$, or $\mathscr{F} \in$ AEC(P, d), if and only if for every $\varepsilon > 0$,

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \Pr_P^*\{\|\mathbb{X}_n\|_{\mathscr{F}'(\delta, d)} > \varepsilon\} = 0.$$

DEFINITION 1.3. We say that $\mathscr{F}$ *admits a weak gaussian approximation of the empirical process* $\mathbb{X}_n$ or $\mathscr{F} \in$ WGA(P), if and only if there exists a sequence of coherent G_P processes $\mathbb{X}^{(1)}, \mathbb{X}^{(2)}, \ldots$ such that for every $\varepsilon > 0$,

$$\lim_{n \to \infty} \Pr_P^*\{\|\mathbb{X}_n - \mathbb{X}^{(n)}\|_{\mathscr{F}} > \varepsilon\} = 0.$$

DEFINITION 1.4. We say that *$\mathscr{F}$ admits a weak functional gaussian approximation of the empirical process(es)* and write $\mathscr{F} \in \mathrm{WFGA}(P)$, if and only if there exists a sequence of independent coherent G_P processes $\mathbb{Y}_1, \mathbb{Y}_2, \dots$ such that for every $\varepsilon > 0$,

$$\lim_{n \to \infty} \Pr_P^* \left\{ n^{-1/2} \max_{m \le n} \left\| m(\mathbb{P}_m - P) - \sum_{i=1}^{m} \mathbb{Y}_i \right\|_{\mathscr{F}} > \varepsilon \right\} = 0.$$

If the convergence to zero in probability is replaced by convergence in L_p for some $p > 0$, then we say that *$\mathscr{F}$ admits an $L_p(\Omega, \Sigma, \Pr_P)$-functional gaussian approximation* and write $\mathscr{F} \in L_p\mathrm{FGA}(P)$; that is,

$$\lim_{n \to \infty} E_P^* \left| n^{-1/2} \max_{m \le n} \left\| m(\mathbb{P}_m - P) - \sum_{i=1}^{m} \mathbb{Y}_i \right\|_{\mathscr{F}} \right|^p = 0.$$

Another gaussian process with which we will be concerned is the *P-Kiefer process* $\mathbb{Z}_P$; it is a gaussian process indexed by $[0, 1] \times L^2(P) \equiv I \times L^2(P)$ with mean zero and covariances

$$\mathrm{Cov}(\mathbb{Z}_P(s, f), \mathbb{Z}_P(t, g)) = (s \wedge t)(P(fg) - P(f)P(g)) \qquad \text{for all } f, g \in L^2(P).$$

A class $\mathscr{F} \subset L^2(P)$ will be called a *functional pregaussian class*, and we write $\mathscr{F} \in \mathrm{FPG}(P)$ if there is a version of $\mathbb{Z}_P$ with bounded and uniformly $\tilde{\rho}_P$-continuous sample functions on $\tilde{\mathscr{F}} \equiv I \times \mathscr{F}$; here $\tilde{\rho}_P((s, f), (t, g)) \equiv |s - t| \vee \rho_P(f, g)$.

For our last definition, we let $I \equiv [0, 1]$ and define the *sequential empirical process* $\mathbb{Z}_n$: $I \times \mathscr{F} \to R$ by

$$\begin{aligned} \mathbb{Z}_n(s, f) &\equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{[ns]} (f(X_i) - P(f)) \\ &= \frac{[ns]}{\sqrt{n}} (\mathbb{P}_{[ns]} - P)(f), \qquad (s, f) \in I \times \mathscr{F}. \end{aligned}$$

Thus $\mathbb{Z}_n$ can be regarded as an element of $l^\infty(I \times \mathscr{F}) \equiv l^\infty(\tilde{\mathscr{F}})$.

DEFINITION 1.5. We say that *$\mathscr{F}$ satisfies the functional central limit theorem* and write $\mathscr{F} \in \mathrm{FCLT}(P)$ if and only if $\mathscr{F} \in \mathrm{FPG}(P)$ and $\mathbb{Z}_n \Rightarrow \mathbb{Z}$ in $l^\infty(\tilde{\mathscr{F}})$.

The following theorem is due to Dudley (1984), (1985), Dudley and Philipp (1983) and Hoffman-Jørgensen (1984); also see Andersen and Dobrić (1987) and Giné and Zinn (1986), Theorem 1.3. Although (E) and (F) are clearly very closely related, we have not seen it stated before with the $\mathbb{Z}_n$ part (statement F) included.

THEOREM 1.1. *The following are equivalent*:
(A) $\mathscr{F} \in CLT(P)$.
(B) $\mathscr{F} \in AEC(P, \rho_P)$ *and* $\mathscr{F}$ *is* P*-pregaussian*.
(B)′ $\mathscr{F} \in AEC(P, \rho_P)$ *and* $\mathscr{F}$ *is* ρ_P*-totally bounded*.
(C) $\mathscr{F} \in WGA(P)$.
(D) $\mathscr{F} \in WFGA(P)$.
(E) $\mathscr{F} \in L_p FGA(P)$ *for all* $0 < p < 2$.
(F) $\mathscr{F} \in FCLT(P)$.

PROOF. All the conditions except (B)′ imply that $\mathscr{F}$ is P-pregaussian by definition. Dudley [(1984), Theorem 4.1.1] proves (B)′ equivalent to (D). The equivalence of (A), (B) and (D) was shown by Dudley [(1985), Theorem 5.2, page 158]. Since (D) implies (C) trivially (take $\mathbb{X}^{(n)} = n^{-1/2}\sum_{i=1}^{n}\mathbb{Y}_i$) and (C) implies (B) [by Dudley's (1984), Theorem 4.1.1 proof], (C) can be included in the list. (E) implies (D) trivially, while (D) implies (E) was proved by Dudley and Philipp [(1983), Theorem 1.3, pages 525–526]. Note that (F) implies (A) trivially.

It remains only to show that (E) implies (F). We will prove this by noting that (F) is equivalent to $\mathscr{F} \in \mathrm{FPG}(P)$ and

$$\text{(a)} \qquad \sup_{h \in BL_1} |E^* h(\mathbb{Z}_n) - Eh(\mathbb{Z})| \to 0 \quad \text{as } n \to \infty,$$

where

$$BL_1 \equiv BL_1(\tilde{\mathscr{F}}) \equiv \{h : l^\infty(\tilde{\mathscr{F}}) \to R;\, \|h\|_\infty \vee \|h\|_L \le 1\},$$

$$\|h\|_\infty \equiv \sup\{|h(x)| : x \in l^\infty(\tilde{\mathscr{F}})\},$$

$$\|h\|_L \equiv \sup_{x \ne y} \frac{|h(x) - h(y)|}{\|x - y\|_{\tilde{\mathscr{F}}}}.$$

See Dudley [(1990), Theorem B] or van der Vaart and Wellner [(1990), Corollary 1.5].

We first show that $\mathscr{F} \in \mathrm{FPG}(P)$. Since $\mathscr{F} \in \mathrm{PG}(P)$, $(\mathscr{F}, \rho_P)$ is totally bounded [see, e.g., Dudley (1984), page 28] and hence $M^2 \equiv \sup_{f \in \mathscr{F}} \rho_P^2(f, 0) = \sup_{f \in \mathscr{F}}(P(f^2) - (Pf)^2) < \infty$. Furthermore, there exist both a sample continuous Brownian motion process B indexed by $I = [0, 1]$, and a ρ_P-uniformly continuous G_P which we may take to be independent. Define $Z_1 = \mathbb{Z}_P$ on $I \times \mathscr{F} \equiv \tilde{\mathscr{F}}$ and $Z_2(s, f) = G_P(f) + MB(s)$ for $(s, f) \in I \times \mathscr{F} \equiv \tilde{\mathscr{F}}$. Then, Z_2 has bounded and $\tilde{\rho}_P$-uniformly continuous sample functions on $\tilde{\mathscr{F}}$ and, for $(s, f), (t, g) \in \tilde{\mathscr{F}}$,

$$\begin{aligned} &E(Z_1(s, f) - Z_1(t, g))^2 \\ &\quad = (s \wedge t)\mathrm{Var}_P(f - g) + (t - s)^+ \mathrm{Var}_P(g) + (s - t)^+ \mathrm{Var}_P(f) \\ &\quad \le \mathrm{Var}_P(f - g) + |t - s| M^2 \\ &\quad = E(Z_2(s, f) - Z_2(t, g))^2. \end{aligned}$$

It then follows from a gaussian comparison theorem due to Marcus and Shepp [Jain and Marcus (1978), page 145, and, for a convenient statement, Giné and Zinn (1986), Theorem 4.4, pages 73 and 74, where (4.7) should read (4.10)] that $Z_1 \equiv \mathbb{Z}_P$ can be chosen to be bounded and $\tilde{\rho}_P$-uniformly continuous on $\tilde{\mathscr{F}} = I \times \mathscr{F}$, or $\mathscr{F} \in \mathrm{FPG}(P)$.

Now we prove (a). Let

$$\mathbb{Z}^{(n)}(s, f) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{[ns]} \mathbb{Y}_i(f) \quad \text{for } (s, f) \in I \times \mathscr{F} \equiv \tilde{\mathscr{F}}.$$

Then

$$\text{(b)} \qquad \|\mathbb{Z}_n - \mathbb{Z}^{(n)}\|_{\tilde{\mathscr{F}}} = n^{-1/2} \max_{k \le n} \left\| k(\mathbb{P}_k - P) - \sum_{i=1}^{k} \mathbb{Y}_i \right\|_{\mathscr{F}}$$

so that

$$\text{(c)} \qquad \begin{aligned} &\sup_{h \in BL_1} \left| E_P^* h(\mathbb{Z}_n) - Eh(\mathbb{Z}^{(n)}) \right| \\ &\le E_P^* \|\mathbb{Z}_n - \mathbb{Z}^{(n)}\|_{\tilde{\mathscr{F}}} \\ &= E_P^* \left(n^{-1/2} \max_{k \le n} \left\| k(\mathbb{P}_k - P) - \sum_{i=1}^{k} \mathbb{Y}_i \right\|_{\mathscr{F}} \right) \qquad [\text{by (b)}] \\ &\to 0 \qquad [\text{by } E \text{ with } p = 1]. \end{aligned}$$

But $\mathbb{Z}^{(n)} \Rightarrow \mathbb{Z}$ in $l^\infty(\tilde{\mathscr{F}})$ follows easily from Fernique [(1985), Corollary 2.2] and this in turn implies

$$\text{(d)} \qquad \lim_{n \to \infty} \sup_{h \in BL_1} \left| Eh(\mathbb{Z}^{(n)}) - Eh(\mathbb{Z}) \right| \to 0.$$

Combining (c) and (d) yields (a) and completes the proof. □

The metric e_P is sometimes useful. For it the following is straightforward:

PROPOSITION 1.1. *$(\mathscr{F}, e_P)$ is totally bounded if and only if both $(\mathscr{F}, \rho_P)$ is totally bounded and $\|P\|_{\mathscr{F}} \equiv \sup\{|P(f)|: f \in \mathscr{F}\} < \infty$.*

2. $\mathscr{P}$-uniform Donsker classes $\mathscr{F}$. Now let $\mathscr{P}$ be a collection of probability measures on the measurable space $(A, \mathbb{A})$. For each $P \in \mathscr{P}$, G_P (or $\mathbb{X}^P$) will denote a P-Brownian bridge process, W_P will denote a P-Brownian motion process and $\mathbb{Z}_P$ will denote a P-Kiefer process.

Our goal in this section is to establish a result like Theorem 1.1, but with the convergence in the central limit theorem uniform over all $P \in \mathscr{P}$. Therefore we first need uniform in $P \in \mathscr{P}$ generalizations of the definitions in Section 1. We begin by generalizing the notion of a P-pregaussian class $\mathscr{F} \subset L^2(P)$.

DEFINITION 2.0. A class $\mathscr{F} \subset L^2(A, \mathbb{A}, P)$ for all $P \in \mathscr{P}$ will be called a $\mathscr{P}$-uniform pregaussian class and we write $\mathscr{F} \in \mathrm{PG}_u(\mathscr{P})$ if and only if the P-Brownian bridge processes $G_P(f)(\omega)$, $f \in \mathscr{F}$, $\omega \in \Omega$, $P \in \mathscr{P}$ can be chosen so that

$$\lim_{\lambda \to \infty} \sup_{P \in \mathscr{P}} \Pr_P\{\|G_P\|_{\mathscr{F}} > \lambda\} = 0; \tag{2.1}$$

and, for every $\varepsilon > 0$,

$$\lim_{\delta \downarrow 0} \sup_{P \in \mathscr{P}} \Pr_P\{\|G_P\|_{\mathscr{F}'(\delta, \rho_P)} > \varepsilon\} = 0. \tag{2.2}$$

If (2.1) and (2.2) hold but with ρ_P replaced by e_P in (2.2), we write $\mathscr{F} \in \mathrm{PG}_u(\mathscr{P}, e_P)$. A class $\mathscr{F} \subset L^2(A, \mathbb{A}, P)$ for all $P \in \mathscr{P}$ will be called a *$\mathscr{P}$-uniform functional pregaussian class* and we write $\mathscr{F} \in \mathrm{FPG}_u(\mathscr{P})$ if the P-Kiefer processes $\mathbb{Z}_P(s, f)(\omega)$, $(s, f) \in I \times \mathscr{F}$, $\omega \in \Omega$, $P \in \mathscr{P}$ can be chosen so that

$$\lim_{\lambda \to \infty} \sup_{P \in \mathscr{P}} \Pr_P\{\|\mathbb{Z}_P\|_{\tilde{\mathscr{F}}} > \lambda\} = 0; \tag{2.3}$$

and, for every $\varepsilon > 0$,

$$\lim_{\delta \downarrow 0} \sup_{P \in \mathscr{P}} \Pr_P\{\|\mathbb{Z}_P\|_{\tilde{\mathscr{F}}'(\delta, \tilde{\rho}_P)} > \varepsilon\} = 0. \tag{2.4}$$

Here $\tilde{\mathscr{F}} \equiv I \times \mathscr{F}$ and

$$\tilde{\mathscr{F}}'(\delta, \tilde{\rho}_P) \equiv \Big\{((s, f), (t, g)) \in \tilde{\mathscr{F}} \times \tilde{\mathscr{F}}: \tilde{\rho}_P((s, f), (t, g)) \equiv |t - s| \vee \rho_P(f, g) \le \delta\Big\}.$$

As noted by Giné and Zinn (1991), by Borell's (1975) inequality [or, more conveniently, its version for expectations given by Pisier (1986)], (2.1) and (2.2) are equivalent to

$$\sup_{P \in \mathscr{P}} E\|G_P\|_{\mathscr{F}}^r < \infty \quad \text{and} \quad \lim_{\delta \downarrow 0} \sup_{P \in \mathscr{P}} E\|G_P\|_{\mathscr{F}'(\delta, \rho_P)}^r = 0 \tag{2.5}$$

for any $r > 0$.

Note that if $\mathscr{P}$ contains just a single measure P, then $\mathscr{F} \in \mathrm{PG}_u(\{P\})$ just means that $\mathscr{F}$ is a P-pregaussian class in the terminology of Dudley (1987) and Giné and Zinn (1986). On the other hand, if $\mathscr{P} = \mathbf{M} \equiv$ {all probability measures on A}, then $\mathscr{F} \in \mathrm{PG}_u(\mathbf{M})$ will be abbreviated to just $\mathscr{F} \in \mathrm{PG}_u$; this is written as $\mathscr{F} \in \mathrm{UPG}$ in Giné and Zinn (1991). If $\mathbf{M}_f \equiv \{P \in \mathbf{M}$: P has finite support$\}$, then $\mathscr{F} \in \mathrm{PG}_u(\mathbf{M}_f)$ is written as $\mathscr{F} \in \mathrm{UPG}_f$ in Giné and Zinn (1991).

Now we give the uniform in $P \in \mathscr{P}$ versions of Definitions 1.1–1.5. Since the dependence on P is now important, we will now emphasize it by writing, for example, $\mathbb{X}_n^P$ and $\mathbb{Z}_n^P$ rather than $\mathbb{X}_n$ and $\mathbb{Z}_n$ as in Section 1.

DEFINITION 2.1. We say that $\mathscr{F} \in \mathrm{CLT}(P)$ *uniformly in* $P \in \mathscr{P}$ and write $\mathscr{F} \in \mathrm{CLT}_u(\mathscr{P})$ (or, $\mathscr{F}$ is $\mathscr{P}$-uniform Donsker), if and only if both:

(i) $\mathscr{F} \in \mathrm{PG}_u(\mathscr{P})$; and

(ii) $\lim_{n \to \infty} \sup_{P \in \mathscr{P}} d_{BL^*}(\mathbb{X}_n^P, G_P) \equiv \lim_{n \to \infty} \sup_{P \in \mathscr{P}} \sup_{h \in BL_1} |E^* h(\mathbb{X}_n^P) - Eh(G_P)| = 0$.

DEFINITION 2.2. We say that $\{\mathbb{X}_n^P\}_{n\geq 1}$ is *asymptotically equicontinuous with respect to* d_P *on* $\mathscr{F}$ *uniformly in* $P \in \mathscr{P}$ and write $\mathscr{F} \in \mathrm{AEC}_u(\mathscr{P}, d_P)$ if and only if for every $\varepsilon > 0$,

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \sup_{P\in\mathscr{P}} \Pr_P^*\left\{\|\mathbb{X}_n^P\|_{\mathscr{F}'(\delta, d_P)} > \varepsilon\right\} = 0. \tag{2.6}$$

Here we have in mind usually $d_P = \rho_P$ or sometimes $d_P = e_P$.

DEFINITION 2.3. We say that $\mathscr{F}$ *admits a weak gaussian approximation of the empirical process* $\mathbb{X}_n^P$ *uniformly in* $P \in \mathscr{P}$, or $\mathscr{F} \in \mathrm{WGA}_u(\mathscr{P})$, if and only if $\mathscr{F} \in \mathrm{PG}_u(\mathscr{P})$ and for each P, there exists a sequence of coherent G_P processes $\mathbb{X}^{(1)}, \mathbb{X}^{(2)}, \dots$ such that for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \sup_{P\in\mathscr{P}} \Pr_P^*\left\{\left\|\mathbb{X}_n^P - \mathbb{X}^{(n)}\right\|_{\mathscr{F}} > \varepsilon\right\} = 0. \tag{2.7}$$

DEFINITION 2.4. $\mathscr{F} \in \mathrm{WFGA}_u(\mathscr{P})$, or $\mathscr{F} \in \mathrm{WFGA}(P)$ *uniformly in* $P \in \mathscr{P}$, if and only if $\mathscr{F} \in \mathrm{PG}_u(\mathscr{P})$ and for each $P \in \mathscr{P}$, there is a sequence of independent coherent G_P-processes $\mathbb{Y}_1, \mathbb{Y}_2, \dots$ such that, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \sup_{P\in\mathscr{P}} \Pr_P^*\left\{n^{-1/2} \max_{m\leq n}\left\| m(\mathbb{P}_m - P) - \sum_{i=1}^m \mathbb{Y}_i\right\|_{\mathscr{F}} > \varepsilon\right\} = 0. \tag{2.8}$$

If instead

$$\lim_{n\to\infty} \sup_{P\in\mathscr{P}} E_P^*\left| n^{-1/2} \max_{m\leq n}\left\| m(\mathbb{P}_m - P) - \sum_{i=1}^m \mathbb{Y}_i\right\|_{\mathscr{F}}\right|^p = 0 \tag{2.9}$$

for some $p > 0$, then we say that $\mathscr{F} \in L_p\mathrm{FGA}_u(\mathscr{P})$.

DEFINITION 2.5. $\mathscr{F} \in \mathrm{FCLT}_u(\mathscr{P})$, or $\mathscr{F} \in \mathrm{FCLT}(P)$ *uniformly in* $P \in \mathscr{P}$, if and only if both (i) $\mathscr{F} \in \mathrm{FPG}_u(\mathscr{P})$; and (ii) $\lim_{n\to\infty} \sup_{P\in\mathscr{P}} d_{BL^*}(\mathbb{Z}_n^P, \mathbb{Z}_P) \equiv \lim_{n\to\infty} \sup_{P\in\mathscr{P}} \sup_{h\in BL_1(\bar{\mathscr{F}})} |E^*h(\mathbb{Z}_n^P) - Eh(\mathbb{Z}_P)| = 0$.

Let F be a measurable envelope function for $\mathscr{F}$; that is, F is measurable and $|f(x)| \leq F(x)$ for all $f \in \mathscr{F}$ and $x \in A$. Thus $F(x) \geq (\sup_{f\in\mathscr{F}} f(x))^*$ for all $x \in A$, where h^* is the least measurable function dominating a given function h; see, for example, Dudley (1985) or van der Vaart and Wellner (1990).

The following theorem is a uniform in P analogue of Theorem 1.1:

THEOREM 2.1. *Suppose that* $\mathscr{F}$ *has envelope function* F *which is* $\mathscr{P}$*-uniformly square integrable*:

$$\lim_{\lambda\to\infty} \sup_{P\in\mathscr{P}} P\left(F^2 1_{[F\geq\lambda]}\right) = 0.$$

Then the following are equivalent:

(A) $\mathscr{F} \in CLT_u(\mathscr{P})$.

(B) $\mathscr{F} \in AEC_u(\mathscr{P}, \rho_P)$ *and* $\mathscr{F} \in PG_u(\mathscr{P})$.

(B)′ $\mathscr{F} \in AEC_u(\mathscr{P}, \rho_P)$ *and* $(\mathscr{F}, \rho_P)$ *is totally bounded uniformly in* P.

(C) $\mathscr{F} \in WGA_u(\mathscr{P})$.

(D) $\mathscr{F} \in WFGA_u(\mathscr{P})$.

(E) $\mathscr{F} \in L_pFGA_u(\mathscr{P})$ *for all* $0 < p < 2$.

(F) $\mathscr{F} \in FCLT_u(\mathscr{P})$.

PROOF. Our proof will proceed as follows:

$$\begin{tikzcd}
(A) \arrow[rr] \arrow[dr] & & (B) \arrow[rr] & & (D) \arrow[rr, leftrightarrow] \arrow[dl] & & (E) \arrow[rr] & & (F) \arrow[llllllll, bend left, dash] \\
& (B)' \arrow[ur] & & (C) \arrow[ul] & & & & &
\end{tikzcd}$$

The implications (E) implies (D), (D) implies (C), and (F) implies (A) are all trivial. The proof that (E) implies (F) is similar to the proof of the same part of Theorem 1.1, using the Lipschitz property of the bounded Lipschitz functions. [Here we need to show that $\mathscr{F} \in \mathrm{PG}_u(\mathscr{P})$ implies $\mathscr{F} \in \mathrm{FPG}_u(\mathscr{P})$ also; we defer this proof to Section 5.]

To see that (C) implies (B), let $N(\varepsilon)$ be so large that

$$\text{(a)} \qquad \sup_{P \in \mathscr{P}} \Pr_P^*\left\{\left\|\mathbb{X}_n^P - \mathbb{X}^{(n)}\right\|_{\mathscr{F}} > \varepsilon/3\right\} < \varepsilon/2, \qquad n \geq N(\varepsilon).$$

Since $\mathscr{F} \in \mathrm{PG}_u(\mathscr{P})$ and $\mathbb{X}^{(n)} \sim G_P$ for each n and $P \in \mathscr{P}$, there is a $\delta = \delta(\varepsilon)$ such that

$$\text{(b)} \qquad \sup_{P \in \mathscr{P}} \Pr_P\left\{\left\|\mathbb{X}^{(n)}\right\|_{\mathscr{F}'(\delta, \rho_P)} > \varepsilon/3\right\} < \varepsilon/2.$$

But

$$\text{(c)} \qquad \|\mathbb{X}_n^P\|_{\mathscr{F}'(\delta, \rho_P)} \leq 2\left\|\mathbb{X}_n^P - \mathbb{X}^{(n)}\right\|_{\mathscr{F}} + \left\|\mathbb{X}^{(n)}\right\|_{\mathscr{F}'(\delta, \rho_P)},$$

so combining (a), (b) and (c) shows that (B) holds.

Now we show that (A) implies (B). For fixed P, this is proved via the portmanteau theorem; see, for example, Giné and Zinn [(1986), pages 60 and 61, and Andersen and Dobrić (1987), Remark 2.13, page 168, or van der Vaart and Wellner (1990), Lemma 1.3, for the validity of the portmanteau theorem in the Hoffman-Jørgensen weak convergence theory]. Here we argue directly, taking care to obtain the needed uniformity in P.

Suppose that A holds and let $\varepsilon > 0$ and $\delta > 0$. Then the set

$$F_{\delta, \varepsilon} \equiv \left\{x \in l^\infty(\mathscr{F}) \colon \|x\|_{\mathscr{F}'(\delta, \rho_P)} \geq \varepsilon\right\}$$

is closed and

$$G \equiv \left\{x \in l^\infty(\mathscr{F}) \colon \|x - F_{\delta, \varepsilon}\|_{\mathscr{F}} < \varepsilon/4\right\}$$

is open. Further, $x \in G$ implies $x \in F_{\delta, \varepsilon/4}$ or $G \subset F_{\delta, \varepsilon/4}$. Now let $\phi(u) \equiv \{(1-u) \wedge 1\} \vee 0$ for $u \in R$ and $h(x) = \phi(4\|x - F_{\delta,\varepsilon}\|_{\mathscr{F}}/\varepsilon)$, so that

$$1_{F_{\delta,\varepsilon}} \le h \le 1_G \le 1_{F_{\delta,\varepsilon/4}}. \tag{d}$$

Then we have

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{P\in\mathscr{P}} \Pr\nolimits_P^*\{\mathbb{X}_n^P \in F_{\delta,\varepsilon}\} \\
&\quad \le \limsup_{n\to\infty} \sup_{P\in\mathscr{P}} E_P^* h(\mathbb{X}_n^P) \qquad [\text{by (d)}] \\
&\quad = \limsup_{n\to\infty} \sup_{P\in\mathscr{P}} \left(E_P^* h(\mathbb{X}_n^P) - Eh(G_P) + Eh(G_P)\right) \\
&\quad \le 0 + \sup_{P\in\mathscr{P}} \Pr\nolimits_P\{G_P \in F_{\delta,\varepsilon/4}\} \qquad [\text{by (A) and (d)}] \\
&\quad = \sup_{P\in\mathscr{P}} \Pr\nolimits_P\{\|G_P\|_{\mathscr{F}'(\delta,\rho_P)} \ge \varepsilon/4\} \to 0 \qquad [\text{as } \delta \downarrow 0],
\end{aligned} \tag{e}$$

and hence (B) holds.

As we show in (g) below, $\mathscr{F} \in \mathrm{PG}_u(\mathscr{P})$ implies that $(\mathscr{F}, \rho_P)$ is totally bounded uniformly in $P \in \mathscr{P}$, so (A) implies (B)′. To show that B′ implies B, we need to show that (B)′ implies $\mathscr{F} \in \mathrm{PG}_u(\mathscr{P})$. By the uniform in $P \in \mathscr{P}$ multivariate CLT, Theorem 5.1, for any finite subset $\mathbf{G} \subset \mathscr{F}$, we can pass from $\mathbb{X}_n$ in the condition for $\mathscr{F} \in \mathrm{AEC}_u(\mathscr{P}, \rho)$ to a similar statement for G_P; that is, for $\delta < \delta(\varepsilon)$,

$$\sup_{P\in\mathscr{P}} \Pr\nolimits_P\{\|G_P\|_{\mathbf{G}'(\delta,\rho_P)} > \varepsilon/2\} < 2\varepsilon.$$

Since $\delta(\varepsilon)$ does not depend on $\mathbf{G}$ or $P \in \mathscr{P}$, we can let $\mathbf{G}$ increase to a countable dense set $\mathbf{H}_P \subset \mathscr{F}$ for $L^2(P)$ and thereby obtain

$$\sup_{P\in\mathscr{P}} \Pr\nolimits_P\{\|G_P\|_{\mathbf{H}'_P(\delta,\rho_P)} > 2\varepsilon\} \le 2\varepsilon$$

for $\delta < \delta(\varepsilon)$. Thus G_P defined for $f \in \mathscr{F}$ by

$$\lim\{G_P(h) \colon h \in \mathbf{H}_P,\ \rho_P(f, h) \to 0\}$$

exists a.s. and defines a (family of) version(s) of G_P satisfying (2.2).

To prove that (2.1) holds, let $\mathscr{F}(\delta)$ be a δ-net for $(\mathscr{F}, \rho_P)$ with $k \equiv \#(\mathscr{F}(\delta))$ independent of $P \in \mathscr{P}$ by the uniform in $P \in \mathscr{P}$ total boundedness of $(\mathscr{F}, \rho_P)$. Also, note that the $\mathscr{P}$-uniform square integrability of F implies that there is a constant M satisfying $\sup_{P\in\mathscr{P}} P(F^2) \le M^2 < \infty$. Let Π_δ denote a map from $\mathscr{F}$ to a nearest point in $\mathscr{F}(\delta)$. Then, since

$$\|G_P\|_{\mathscr{F}} \le \Big\|G_P - G_P \circ \prod_\delta\Big\|_{\mathscr{F}} + \|G_P\|_{\mathscr{F}(\delta)} \le \|G_P\|_{\mathscr{F}'(\delta,\rho_P)} + \|G_P\|_{\mathscr{F}(\delta)},$$

it follows, for $0 < \varepsilon/2 < \lambda$, that

$$\begin{aligned}\sup_{P\in\mathscr{P}} &\Pr_P(\|G_P\|_{\mathscr{F}} > \lambda)\\ &\le \sup_{P\in\mathscr{P}} \Pr_P\left\{\|G_P\|_{\mathscr{F}'(\delta,\rho_P)} > \frac{\varepsilon}{2}\right\} + \sup_{P\in\mathscr{P}} \Pr_P\left\{\|G_P\|_{\mathscr{F}(\delta)} > \lambda - \frac{\varepsilon}{2}\right\}\\ &\le \frac{\varepsilon}{2} + \frac{kM^2}{(\lambda - (\varepsilon/2))^2} \qquad \left[\text{by (2.2) for } \delta \le \delta\left(\frac{\varepsilon}{4}\right)\right]\\ &\le \varepsilon\end{aligned}$$

for $\lambda \ge \varepsilon/2 + (2kM^2/\varepsilon)^{1/2}$, so (2.1) holds and $\mathscr{F} \in \mathrm{PG}_u(\mathscr{P})$.

It remains only to prove that (B) implies (D) and that (D) implies (E).

The proof that (B) implies (D) (given that $\mathscr{F} \in \mathrm{PG}_u(\mathscr{P})$ and the envelope F is $\mathscr{P}$-uniformly square integrable) is lengthy, but follows the argument used by Dudley (1984) to prove the corresponding part of Theorem 1.1, after preparation in the form of a uniform extension of the results of Philipp (1980) to handle the finite-dimensional laws. The only major modification of Dudley's (1984) proof which we use here is the following: We argue that $\mathscr{F} \in \mathrm{PG}_u(\mathscr{P})$ implies that $(\mathscr{F}, \rho_P)$ is totally bounded uniformly in $P \in \mathscr{P}$. By the Sudakov (1971) minorization theorem [see, e.g., Giné and Zinn (1986), page 73] and (2.5),

$$\text{(f)} \qquad \infty > \sup_{P\in\mathscr{P}} E\|G_P\|_{\mathscr{F}} \ge C \sup_{P\in\mathscr{P}} \varepsilon\big(\log N(\varepsilon, \mathscr{F}, \rho_P)\big)^{1/2} \quad \text{for all } \varepsilon > 0$$

for some constant C, and hence there is a constant $K < \infty$ such that

$$\text{(g)} \qquad \sup_{P\in\mathscr{P}} \log N(\varepsilon, \mathscr{F}, \rho_P) \le \frac{K}{\varepsilon^2}$$

for all $\varepsilon > 0$, and this gives the total boundedness of $(\mathscr{F}, \rho_P)$ uniformly in $P \in \mathscr{P}$.

The proofs of (B) implies (D) and of (D) implies (E) are given in Sections 5 and 6. □

To provide sufficient conditions for the $\mathscr{P}$-uniform central limit theorem, our methods of proof will need $\mathscr{F}$ to satisfy enough measurability so that, for each $P \in \mathscr{P}$, $\|\mathbb{X}_n\|_{\mathscr{F}'(\delta,\rho_P)}$ is completion measurable and Fubini's theorem can be applied to $\|\Sigma_{i=1}^n Y_i \delta_{X_i}\|_{\mathscr{F}'(\delta,\rho_P)}$, where the Y_i are iid real-valued symmetric rv's independent of the X_i's and defined on ([0, 1], **B**, Lebesgue). In the terminology of Giné and Zinn (1986), we require $\mathscr{F} \in \mathrm{NLDM}(P)$ for each $P \in \mathscr{P}$. When this holds, we say $\mathscr{F}$ is *$\mathscr{P}$-measurable* or $\mathscr{F} \in \mathrm{NLDM}(\mathscr{P})$. If $\mathscr{F}$ is **M**-measurable, we abbreviate this to simply saying that $\mathscr{F}$ is *measurable*. It is well-known that $\mathscr{F}$ is measurable if it is countable, or if the empirical processes $\mathbb{X}_n^P$ are stochastically separable or if $\mathscr{F}$ is image admissible Suslin [see, e.g., Dudley (1984), Section 10.3].

If we insist on uniformity of convergence over all $P \in \mathbf{M}$, then of course $\mathscr{F} \in \mathrm{CLT}(P)$ for all $P \in \mathbf{M}$ or $\mathscr{F}$ is *universal Donsker* in the terminology of Dudley (1987). As shown by Dudley (1987), this implies that $\sup_{f \in \mathscr{F}} \operatorname{diam}(f) < \infty$ [where $\operatorname{diam}(f) \equiv \sup_{x \in A} f(x) - \inf_{x \in A} f(x)$]. For classes of this type, Giné and Zinn (1991) used Gaussian comparison methods to prove the following beautiful characterization of classes $\mathscr{F}$ satisfying the $\mathrm{CLT}_u = \mathrm{CLT}_u(\mathbf{M})$. Recall that $\mathbf{M} \equiv$ {all probability measures on A} and $\mathbf{M}_f \equiv \{P \in \mathbf{M}$: P has finite support$\}$.

THEOREM 2.2 [Giné and Zinn (1991)]. *Suppose that $\mathscr{F}$ is measurable. Then $\mathscr{F} \in PG_u(\mathbf{M}_f)$ implies $\mathscr{F} \in CLT_u$.*

Note that CLT_u implies $\mathscr{F} \in \mathrm{PG}_u(\mathbf{M}) \subset \mathrm{PG}_u(\mathbf{M}_f)$, so the converse is trivially true.

Combining the Giné and Zinn Theorem 2.2 with Theorem 2.1 shows that if $\mathscr{F} \in \mathrm{PG}_u(\mathbf{M}_f)$ and is measurable and, with $c \equiv 1/\sup_{f \in F}(\operatorname{diam}(f))$, $c_f \equiv \inf_x f(x)$, $\mathscr{F}_c \equiv \{c(f - c_f)\colon f \in \mathscr{F}\}$, then all of (A)–(E) of Theorem 2.1 hold for $\mathscr{F}_c$ with $\mathscr{P} = \mathbf{M}$ and hence for $\mathscr{F}$ with $\mathscr{P} = \mathbf{M}$. In particular, (A)–(E) of Theorem 2.1 hold with $\mathscr{P} = \mathbf{M}$ if $\mathscr{F} = \{1_C\colon C \in \mathbf{C}\}$, where $\mathbf{C}$ is a measurable Vapnik–Chervonenkis class of subsets of A.

When the function class $\mathscr{F}$ has $\sup_{f \in \mathscr{F}}(\operatorname{diam}(f))$ unbounded, then the uniform Donsker property cannot hold for all measures $\mathbf{M}$, but it may still hold for a very substantial subcollection $\mathscr{P} \subset \mathbf{M}$. To state our first result in this direction, for $r > 0$, let $N_F^{(r)}(\varepsilon, \mathscr{F})$, $\varepsilon > 0$, denote Pollard's (1982) (rth power) combinatorial entropy of $\mathscr{F}$ relative to an envelope function F of $\mathscr{F}$; also see Dudley [(1984), Chapter 11]. The following definition is due to Pollard (1982).

DEFINITION 2.6. We say that $(\mathscr{F}, F)$ is sparse (or $\mathscr{F}$ is F-sparse) if and only if

$$\int_0^1 \left(\log N_F^{(2)}(\varepsilon, \mathscr{F})\right)^{1/2} d\varepsilon < \infty. \tag{2.10}$$

By extending Pollard's (1982) argument, Sheehy and Wellner (1988) proved the following theorem for classes of functions satisfying (2.10).

THEOREM 2.3. *Suppose that $\mathscr{F}$ is $\mathscr{P}$-measurable and satisfies*: (i) *F is $\mathscr{P}$-uniformly square integrable.* (ii) *$(\mathscr{F}, F)$ is sparse. Then all of* (A)–(F) *in Theorem* 2.2 *hold.*

Theorem 2.3 will be proved in the second half of Section 5.

Note that for uniformly bounded classes $\mathscr{F}$ with $F = 1$, all the classes $\mathscr{F}$ in Dudley's (1987) Theorem 2.1 and Figure 1 except (1.4), $(1.4)^{co}$ and (1.6) (and, of course, universal Donsker itself) satisfy (ii) in Theorem 2.3, so Theorem 2.3 shows that all of those classes are, in fact, $\mathbf{M}$-uniform Donsker as well as universal Donsker. Giné and Zinn (1991) show that Dudley's classes (1.4) ($\mathscr{F} = \{f_j\}_{j \ge 2}$ with $\operatorname{diam}(f_j) = o((\log j)^{-1/2})$) satisfy $\mathscr{F} \in \mathrm{PG}_u(\mathbf{M}_f)$ and hence $\mathscr{F} \in \mathrm{CLT}_u$ by Theorem 2.2.

Also note that Theorem 2.3 allows classes $\mathscr{F}$ for which $\sup_{f \in \mathscr{F}}(\operatorname{diam}(f))$ is infinite; for example, $\mathscr{F} = \{F1_C: C \in \mathbf{C}\}$ satisfies both (i) and (ii) if the function F satisfies (i) and $\mathbf{C}$ is any measurable VC class of subsets of A, even though $\sup_{C \in \mathbf{C}} \operatorname{diam}(F1_C)$ may be infinite if F is unbounded.

Massart [(1986), Theorem 5.10, page 411], has established rates of convergence for (C) in Theorem 1.1 under growth rate hypotheses on $N_F^{(2)}(\varepsilon, \mathscr{F})$ and additional moment hypotheses on the envelope function F. In view of Theorem 2.3 it is quite plausible, although not yet proved, that the same rates apply to (C) in Theorem 2.1.

3. $\{P_n\}$-Donsker classes $\mathscr{F}$. Now we consider sequences of measures on $(A, \mathbb{A})$. For each $n = 1, 2, \dots,$ we suppose that $X_{n1}, \dots, X_{nm}$ are row independent, iid P_n, (A-valued) random variables and $m = m(n)$ is any sequence of integers $\uparrow \infty$ as $n \to \infty$. We assume that the resulting triangular array is defined on a common probability space

$$
\begin{aligned}
&(\Omega, \Sigma, \Pr) \\
(3.1) \qquad &\equiv \left(A^{m(1)}, \mathbb{A}^{m(1)}, P_1^{m(1)}\right) \times \cdots \times \left(A^{m(n)}, \mathbb{A}^{m(n)}, P_n^{m(n)}\right) \\
&\times \cdots \times ([0,1], \mathbf{B}, \lambda),
\end{aligned}
$$

where λ denotes Lebesgue measure. We define the empirical measure $\mathbb{P}_m$ of the m random variables in the nth row of the array by

$$
\mathbb{P}_m \equiv \frac{1}{m} \sum_{i=1}^{m} \delta_{X_{ni}} \tag{3.2}
$$

and the empirical process by

$$
\mathbb{X}_m^{P_n} \equiv \mathbb{X}_m \equiv \sqrt{m}\,(\mathbb{P}_m - P_n). \tag{3.3}
$$

Suppose that a collection of real-valued functions $\mathscr{F}$ on A is $\mathscr{P}$-uniform Donsker for a collection $\mathscr{P}$ of probability measures on A as defined in Section 2; $\mathscr{F} \in \mathrm{CLT}_u(\mathscr{P})$. Then if $\{P_n\}$ is any sequence in $\mathscr{P}$, we deduce from Theorem 2.1 that, for example: For every $\varepsilon > 0$,

$$
\lim_{n \to \infty} \Pr\left\{\left\|\mathbb{X}_m^{P_n} - G_{P_n}^{(m)}\right\|_{\mathscr{F}} > \varepsilon\right\} = 0, \tag{3.4}
$$

where $G_{P_n}^{(m)} = G_{P_n}^{(m(n))}$ is a P_n-Brownian bridge process for each $n \ge 1$, even if the sequence $\{P_n\}$ does not converge. If $\{P_n\}$ converges in one of several senses to a measure P_0, then we can replace $G_{P_n}^{(m)}$ in (3.4) by a sequence $G_{P_0}^{(m)}$ of P_0-Brownian bridges and deduce that $\mathbb{X}_m^{P_n} \Rightarrow G_{P_0}$. Our goal in this section is to make this more precise.

DEFINITION 3.1. $\mathscr{F} \in \mathrm{CLT}(\{P_n\}_{n \ge 0})$ (or $\mathscr{F}$ is *$\{P_n\}_{n \ge 0}$-Donsker*) if and only if $\mathscr{F}$ is P_0-pregaussian and $\mathbb{X}_m^{P_n} \Rightarrow \mathbb{X}_0 \sim G_{P_0}$ in $l^\infty(\mathscr{F})$ for some fixed probability measure P_0.

For a given class $\mathscr{F}$, set $\mathbf{G} \equiv \mathscr{F} \cup \mathscr{F}^2 \cup (\mathscr{F}')^2$. Note that for $P, Q \in \mathscr{P}$ and $f, g \in \mathscr{F}$,

$$\left| e_P^2(f, g) - e_Q^2(f, g) \right| \le \|P - Q\|_{\mathbf{G}}. \tag{3.5}$$

Here is a basic theorem of the type we need.

THEOREM 3.1. *Suppose that*:

(i) F *is* $\{P_n\}$ – *uniformly square integrable*:

$$\lim_{\lambda \to \infty} \limsup_{n \to \infty} P_n\left(F^2 1_{[F \ge \lambda]}\right) = 0.$$

(ii) $\|P_n - P_0\|_{\mathbf{G}} \to 0$ *as* $n \to \infty$.

(iii) $(\mathscr{F}, \rho_{P_0})$ *is totally bounded and* $\mathscr{F} \in AEC(\{P_n\}, \rho)$: *That is, for every* $\varepsilon > 0$,

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \Pr{}^*_{P_n}\left\{\|\mathbb{X}_m^{P_n}\|_{\mathscr{F}'(\delta, \rho_n)} > \varepsilon\right\} = 0,$$

where $\mathscr{F}_n'(\delta, \rho_n) \equiv \mathscr{F}'(\delta, \rho_{P_n})$.

Then $\mathscr{F} \in CLT(\{P_n\}_{n \ge 0})$.

Now we can use the sufficient conditions given in Section 2 to verify hypothesis (iii) for various classes $\mathscr{F}$. The following corollary is a consequence of Theorems 2.3 and 3.1:

COROLLARY 3.1. *Suppose that* $\mathscr{F} \in NLDM(\{P_n\})$ *and that*:

(i) F *is* $\{P_n\}$-*uniformly square integrable*.
(ii) $(\mathscr{F}, F)$ *is sparse*; *that is*, (2.10) *holds*.
(iii) $\|P_n - P_0\|_{\mathbf{G}} \to 0$ *as* $n \to \infty$.

Then $\mathscr{F} \in CLT(\{P_n\}_{n \ge 0})$.

While the preceding corollary handles unbounded classes of functions under a Pollard type entropy condition, the following corollary assumes the uniform Donsker property and therefore yields somewhat more for classes $\mathscr{F}$ with $\sup_{f \in \mathscr{F}}(\operatorname{diam}(f))$ finite.

COROLLARY 3.2. *Suppose that* $\mathscr{F}$ *is measurable and that*:

(i) $\mathscr{F} \in CLT_u$.
(ii) $\|P_n - P_0\|_{\mathbf{G}} \to 0$ *as* $n \to \infty$.

Then $\mathscr{F} \in CLT(\{P_n\}_{n \ge 0})$.

PROOF. We first need a lemma concerning normal distributions in R^d. Let π and d_{BL^*} denote the Prohorov and bounded Lipschitz metrics for probability measures; see, for example, Dudley [(1989), Section 11.3].

LEMMA 3.1. *Suppose that* $\Sigma = (\sigma_{ij})$ *and* $\bar{\Sigma} = (\bar{\sigma}_{ij})$ *are two* $d \times d$ *covariance matrices on* R^d *and let* $N(0, \Sigma)$, $N(0, \bar{\Sigma})$ *be the corresponding mean zero gaussian laws. Then*

$$\pi\big(N(0,\Sigma), N(0,\bar{\Sigma})\big) \le M_d \|\Sigma - \bar{\Sigma}\|_1^{1/4} \le \tilde{M}_d \|\Sigma - \bar{\Sigma}\|_\infty^{1/4}$$

and

$$d_{BL^*}\big(N(0,\Sigma), N(0,\bar{\Sigma})\big) \le C_d \|\Sigma - \bar{\Sigma}\|_1^{1/4} \le \tilde{C}_d \|\Sigma - \bar{\Sigma}\|_\infty^{1/4},$$

where M_d, $\tilde{M}_d$, C_d, $\tilde{C}_d$ *are constants depending only on d and*

$$\|\Sigma - \bar{\Sigma}\|_1 \equiv \max_{1 \le j \le d} \sum_{i=1}^{d} |\sigma_{ij} - \bar{\sigma}_{ij}|, \qquad \|\Sigma - \bar{\Sigma}\|_\infty \equiv \max_{1 \le i, j \le d} |\sigma_{ij} - \bar{\sigma}_{ij}|.$$

PROOF. This follows from Theorem 7 of Dehling (1983) (by using $1 + |\log x|^{1/2} \le \text{constant } x^{-1/12}$, $x \le 1$) upon noting that $d_{BL^*}(P,Q) \le 2\pi(P,Q)$, where π denotes the Prohorov distance; see, for example, Dudley (1989), Corollary 11.6.5, page 322. □

PROOF OF THEOREM 3.1. First note that (ii) and (iii) imply that $\mathscr{F} \in \mathrm{AEC}(\{P_n\}, \rho_{P_0})$; for every $\varepsilon > 0$,

$$\text{(a)} \qquad \lim_{\delta \downarrow 0} \limsup_{n \to \infty} \Pr{}^*_{P_n}\Big\{\|\mathbb{X}_m^{P_n}\|_{\mathscr{F}'(\delta, \rho_{P_0})} > \varepsilon\Big\} = 0.$$

Now let $\mathscr{F}_d = \{f_1, \dots, f_d\}$ be any finite subset of $\mathscr{F}$ and let F_n denote the laws of $(\mathbb{X}_m^{P_n}(f_1), \dots, \mathbb{X}_m^{P_n}(f_d))$ on R^d. Note that if $\Sigma_n \equiv (\mathrm{Cov}_{P_n}(f_i, f_j))$, $\Sigma_0 \equiv (\mathrm{Cov}_{P_0}(f_i, f_j))$, then by (i) and (ii),

$$\|\Sigma_n - \Sigma_0\|_\infty \le \text{constant} \|P_n - P_0\|_{\mathbf{G}} \to 0.$$

Hence it follows from Theorem 5.1 and Lemma 3.1 that $\pi(F_n, N_d(0, \Sigma_0)) \to 0$ as $n \to \infty$, where $N(0, \Sigma_0)$ is the law of $(G_{P_0}(f_1), \dots, G_{P_0}(f_d))$; that is, the finite-dimensional laws of $\mathbb{X}_m^{P_n}$ converge to those of G_{P_0}:

$$\text{(b)} \qquad \mathbb{X}_m^{P_n} \to_{\text{f.d.}} \mathbb{X}_0 \sim G_{P_0}.$$

But (a) and (b) together with total boundedness of $(\mathscr{F}, \rho_{P_0})$ yield $\mathbb{X}_m^{P_n} \Rightarrow \mathbb{X}_0 \sim G_{P_0}$ by a standard argument; see, for example, Giné and Zinn (1986), Theorem 1.3, (ii) implies (i), or Pollard (1989), Theorem 10.2. Furthermore, (a) and (b) imply that $\mathscr{F}$ is P_0 pregaussian by the same argument as given for (B)′ implies $\mathscr{F} \in \mathrm{PG}_u(\mathscr{P})$ in the proof of Theorem 2.1. □

PROOF OF COROLLARY 3.1. It suffices to check the hypotheses of Theorem 3.1: (i) and (ii) hold by assumption. By (i) and (ii) of Corollary 3.1, (i) and (ii) of the hypotheses of Corollary 2.2 hold for $\mathscr{P} \equiv \{P_n\}$ and hence Corollary 2.2 yields (iii) of Theorem 3.1. Theorem 3.1 then yields the conclusion. □

PROOF OF COROLLARY 3.2. Again, we check the hypotheses of Theorem 3.1: (i) holds trivially, since we can without loss of generality suppose that $F = 1$.

(ii) holds by assumption. But assumption (i) of Corollary 3.2 implies that $\mathscr{F} \in \mathrm{CLT}_u$ and hence $\mathscr{F} \in \mathrm{AEC}_u(\mathbf{M}, \rho)$ by Theorem 2.1. Thus (iii) of Theorem 3.1 holds and the conclusion follows from Theorem 3.1. □

4. Applications to bootstrap and estimation. Now we examine some of the consequences of the $\mathscr{P}$-uniform Donsker theorems in Sections 2 and 3 for various versions of the bootstrapped empirical process.

To begin, suppose that $X_1, \dots, X_n, \dots$ are independent and identically distributed with distribution P on $(A, \mathbb{A})$ and let

$$\mathbb{P}_n \equiv n^{-1} \sum_{i=1}^{n} \delta_{X_i} \tag{4.1}$$

be the empirical measure of the first n X's. Efron's nonparametric bootstrap (ENPB) proceeds by sampling from $\mathbb{P}_n^{\omega} \equiv \mathbb{P}_n(\cdot, \omega)$: Suppose that $X_{n1}^{\#}, \dots, X_{nm}^{\#}$ are independent and identically distributed with distribution $\mathbb{P}_n^{\omega}$ on $(A, \mathbb{A})$ and let

$$\mathbb{P}_m^{\#} \equiv m^{-1} \sum_{i=1}^{m} \delta_{X_{ni}^{\#}}, \qquad \mathbb{X}_{m,n}^{\#}(\omega) \equiv \sqrt{m}\left(\mathbb{P}_m^{\#} - \mathbb{P}_n^{\omega}\right). \tag{4.2}$$

Thus $X_{n1}^{\#}, \dots, X_{nm}^{\#}$ is the "bootstrap sample", $\mathbb{P}_m^{\#}$ is the "bootstrap empirical measure" and $\mathbb{X}_{m,n}^{\#}(\omega)$ is the "bootstrap empirical process."

Giné and Zinn (1990) use Poissonization and symmetrization techniques together with recent results of Ledoux and Talagrand [(1986), (1988)] to prove the following theorem.

THEOREM 4.1 [Giné and Zinn (1990)]. *Suppose that $\mathscr{F} \in M(P)$. Then*:
(A) *The following are equivalent*:
(i) $P(F^2) < \infty$ *and* $\mathscr{F} \in CLT(P)$;
(ii) P^{∞} *a.s.*, $\mathbb{X}_{n,n}^{\#}(\omega) \Rightarrow \mathbb{X}_0^{\#} \sim G_P$ *in* $l^{\infty}(\mathscr{F})$.
(B) *The following are equivalent*:
(i) $\mathscr{F} \in CLT(P)$;
(ii) $d_{BL^*}(\mathbb{X}_{n,n}^{\#}(\omega), G_P) \to 0$ *in* Pr*.

In particular, if either $\mathscr{F}$ has $P(F^2) < \infty$ and $(\mathscr{F}, F)$ is sparse so that Pollard's (1982) CLT holds [see, e.g., Dudley (1984), Theorem 11.3.1] or if $\mathscr{F}$ satisfies

$$\int_0^1 \left(\log N_{[\,]}^{(2)}(\varepsilon, \mathscr{F}, P)\right)^{1/2} d\varepsilon < \infty,$$

where $N_{[\,]}^{(2)}(\varepsilon, \mathscr{F}, P)$ is the L^2-entropy of $\mathscr{F}$ with bracketing [see, e.g., Dudley (1984), Section 6.1] so that Ossiander's (1987) CLT holds, then A.(i) holds, and hence P^{∞} a.s., $\mathbb{X}_{n,n}^{\#}(\omega) \Rightarrow \mathbb{X}_0^{\#} \sim G_P$ in $l^{\infty}(\mathscr{F})$ by Theorem 4.1.A.

Giné and Zinn [(1990), page 852] say the main feature of their theorem, "aside from its generality, is that no assumptions are made on local uniformity (about P) of the CLT ... ." Our goal here is to briefly explore several related results. We want to know if we can decouple the bootstrap sample size m and the original sample size n: Does the bootstrap empirical process still converge if just $m \wedge n \to \infty$? (Note that the Giné and Zinn, Theorem 4.1, sets $m = n$.) Here we do make uniformity assumptions on $\mathscr{F}$. We then prove bootstrap central limit theorems when $m \wedge n \to \infty$, which, as far as we know, have not been proved in the setting of Theorem 4.1.

The methods developed in Sections 2 and 3 also have applications to the stability of bootstrap methods (i.e., when the underlying P depends on n with P_n converging to P); and to parametric or model-based bootstrap resampling methods as suggested by Efron [(1982), Sections 5.2 and 5.3]. We intend to elaborate on these elsewhere.

Here is a limit theorem for the bootstrap empirical process which allows $m \to \infty$ and $n \to \infty$ in any way.

THEOREM 4.2 (A bootstrap limit theorem for $\mathscr{F} \in \mathrm{CLT}_u$ with $m \neq n$). *Suppose that $\mathscr{F}$ is measurable and that $\mathscr{F} \in CLT_u$. Then, for almost all sample sequences $X_1(\omega), \dots, X_n(\omega), \dots$, $\mathbb{X}^{\#}_{m,n}(\omega) \Rightarrow \mathbb{X}^{\#}_0 \sim G_P$ in $l^{\infty}(\mathscr{F})$ as $m \wedge n \to \infty$.*

PROOF. Suppose that $\mathscr{F} \in \mathrm{CLT}_u$. We first claim that it suffices to prove the theorem for classes $\mathscr{F}$ that are uniformly bounded by 1; that is, such that $F(x) \leq 1$ for all $x \in A$. To see this we argue as in Claim 1 of the proof of Theorem 2.3 of Giné and Zinn (1991): $\mathscr{F} \in \mathrm{CLT}_u$ if and only if $\tilde{\mathscr{F}} \in \mathrm{CLT}_u$, where $\tilde{\mathscr{F}} \equiv \{\tilde{f} \equiv c(f - c_f)\colon f \in \mathscr{F}\}$ for some $c \neq 0$ and arbitrary finite constants c_f. Furthermore, $\mathbb{X}^{\#}_{m,n}(\omega) \Rightarrow \mathbb{X}^{\#}_0$ in $l^{\infty}(\mathscr{F})$ a.s. P^{∞} if and only if $\mathbb{X}^{\#}_{m,n}(\omega) \Rightarrow \mathbb{X}^{\#}_0$ in $l^{\infty}(\tilde{\mathscr{F}})$ a.s. P^{∞}. Since $\mathscr{F} \in \mathrm{CLT}_u$ implies that $\mathscr{F}$ is universal Donsker, it follows from Dudley (1987) that $\sup_{f \in \mathscr{F}} \mathrm{diam}(f) < \infty$. Hence taking $c_f \equiv \inf(f)$ and $c \equiv \{\sup_{f \in \mathscr{F}} \mathrm{diam}(f)\}^{-1}$ yields a class of functions $\tilde{\mathscr{F}}$ with $0 \leq \tilde{f} \leq 1$.

It remains only to prove the theorem for a class $\mathscr{F} \in \mathrm{CLT}_u$ with $0 \leq f \leq 1$ for all $f \in \mathscr{F}$. We do this by verifying the hypotheses of Theorem 3.1: Since $F \leq 1$, it follows that

$$\mathbb{P}^{\omega}_n\left(F^2 1_{[F \geq \lambda]}\right) \leq \mathbb{P}^{\omega}_n\left(1_{[1 \geq \lambda]}\right) = 1_{[1 \geq \lambda]},$$

so $\mathbb{P}^{\omega}_n$-uniform square integrability of F holds trivially for each fixed ω. By Theorem 2.1, $\mathscr{F} \in \mathrm{AEC}_u(\mathbf{M}, \rho_P)$. In particular, $\mathscr{F} \in \mathrm{AEC}_u(\{\mathbb{P}^{\omega}_n\}, \rho_P)$ for all ω. Furthermore, as in Giné and Zinn (1991),

$$\|\mathbb{P}_n - P\|^{*}_{\mathbf{G}} \to_{\text{a.s.}} 0 \quad \text{as } n \to \infty.$$

Thus for P^{∞}-a.e. ω, the hypotheses of Theorem 3.1 are satisfied by $\mathscr{F}$ for the sequence $\{\mathbb{P}^{\omega}_n\}$ of probability measures on $(A, \mathbb{A})$. Hence the conclusion follows from Theorem 3.1 with $P_0 \equiv P$. □

Here is a variant of the previous theorem for classes $\mathscr{F}$ which may not have $\sup_{f \in \mathscr{F}}(\mathrm{diam}\ f) < \infty$.

THEOREM 4.3 (A bootstrap limit theorem for sparse classes $\mathscr{F}$ with $m \neq n$). *Suppose that $\mathscr{F}$ is measurable and that*: (i) $P(F^2) < \infty$. (ii) $(\mathscr{F}, F)$ *is sparse. Then, for almost all sample sequences $X_1(\omega), \dots, X_n(\omega), \dots$, $\mathbb{X}^{\#}_{m,n}(\omega) \Rightarrow \mathbb{X}^{\#}_0 \sim G_P$ in $l^\infty(\mathscr{F})$ as $m \wedge n \to \infty$.*

PROOF. First note that F is a.s. $\{\mathbb{P}_n\}$-uniformly square integrable: By the strong law of large numbers,

$$\text{(a)} \qquad \lim_{n\to\infty} \mathbb{P}_n F^2 1_{[F \geq \lambda]} = P\big(F^2 1_{[F\geq\lambda]}\big) \quad \text{a.s.},$$

for any countable collection of λ's and the right side $\to 0$ as $\lambda \to \infty$ by (i). Furthermore, by Pollard [(1982), Theorem 12, page 243], (i) and (ii) imply that

$$\text{(b)} \qquad \|\mathbb{P}_n - P\|^*_{\mathbb{G}} \to_{\text{a.s.}} 0 \quad \text{as } n \to \infty.$$

Thus for P^∞-a.e. ω, the hypotheses of Corollary 3.1 are satisfied by $\mathscr{F}$ for the sequence $\{\mathbb{P}^\omega_n\}$ of probability measures on $(A, \mathbb{A})$. Hence the conclusion follows from Corollary 3.1. □

5. Proofs for Section 2: Main steps. What remains is to complete the proof of Theorem 2.1 and to prove Theorem 2.3. We first show that $\mathscr{F} \in \mathrm{PG}_u(\mathscr{P})$ implies $\mathscr{F} \in \mathrm{FPG}_u(\mathscr{P})$ if $\sup_{P\in\mathscr{P}} P(F^2) \equiv M^2 < \infty$.

PROPOSITION 5.1. *If $\mathscr{F} \in PG_u(\mathscr{P})$ and has envelope function F satisfying $\sup_{P\in\mathscr{P}} P(F^2) \equiv M^2 < \infty$, then $\mathscr{F} \in FPG_u(\mathscr{P})$.*

Our proof of the proposition will depend on the following lemma.

LEMMA 5.1. *If $\mathscr{F} \in PG_u(\mathscr{P})$, then*

$$\lim_{\delta\downarrow 0}\left\{\sup_{P\in\mathscr{P}} \delta\sqrt{\log N(\delta, \mathscr{F}, \rho_P)}\right\} = 0.$$

PROOF. Let $\delta, \varepsilon > 0$. Let $\mathscr{F}(\varepsilon) = \{f_1, \dots, f_m\}$ be an ε-net for $(\mathscr{F}, \rho_P)$ with $m = N(\varepsilon, \mathscr{F}, \rho_P)$. If

$$\text{(a)} \qquad B_k \equiv \{f \in \mathscr{F}: \rho_P(f, f_k) \leq \varepsilon\} \quad \text{for } f_k \in \mathscr{F}(\varepsilon),$$

then $\mathscr{F} \subset \bigcup_{k=1}^m B_k$. Furthermore,

$$\text{(b)} \qquad \begin{aligned} N(\delta, \mathscr{F}, \rho_P) &\leq N\left(\delta, \bigcup_{k=1}^m B_k, \rho_P\right) \\ &\leq m \max_k N(\delta, B_k, \rho_P) = N(\varepsilon, \mathscr{F}, \rho_P) \max_k N(\delta, B_k, \rho_P). \end{aligned}$$

Let

$$\text{(c)} \qquad B'_k \equiv \{f - f_k: f \in \mathscr{F}, \rho_P(f, f_k) < \varepsilon\}$$

and

$$\text{(d)} \qquad \mathscr{F}''(\varepsilon, \rho_P) = \{f - g: f, g \in \mathscr{F}, \rho_P(f, g) < \epsilon\}.$$

Then for each k, we have

(e) $$B'_k \subset \mathscr{F}''(\varepsilon, \rho_P)$$

and

(f) $$N(\delta, B'_k, \rho_P) = N(\delta, B_k, \rho_P).$$

Hence, using Sudakov's inequality [(f) in Section 2 above] repeatedly,

(g)
$$\begin{aligned}
\delta\sqrt{\log N(\delta, \mathscr{F}, \rho_P)} &\le \delta\sqrt{\log N(\varepsilon, \mathscr{F}, \rho_P)} + \delta \max_{1\le k\le m} \sqrt{\log N(\delta, B'_k, \rho_P)} \\
&\qquad\qquad [\text{by (b) and (f)}] \\
&\le \delta\sqrt{\log N(\varepsilon, \mathscr{F}, \rho_P)} + C \max_{1\le k\le m} E\|G_P\|_{B'_k} \\
&\qquad [\text{by Sudakov's inequality applied to the second term}] \\
&\le \frac{C\delta}{\varepsilon} E\|G_P\|_{\mathscr{F}} + CE\|G_P\|_{\mathscr{F}''(\varepsilon,\rho_P)} \\
&\qquad [\text{by Sudakov's inequality applied to the first term and (e)}],
\end{aligned}$$

where C is an absolute constant. Taking the supremum over $P \in \mathscr{P}$ and then letting $\delta \downarrow 0$ yields, by (2.5),

(h) $$\lim_{\delta\downarrow 0}\left\{\sup_{P\in\mathscr{P}} \delta\sqrt{\log N(\delta, \mathscr{F}, \rho_P)}\right\} \le C \sup_{P\in\mathscr{P}} E\|G_P\|_{\mathscr{F}''(\varepsilon,\rho_P)}$$

for every $\varepsilon > 0$; letting $\varepsilon \downarrow 0$ yields the conclusion, again using (2.5). □

Proof of Proposition 5.1. Let $Z_1(s, f) \equiv \mathbb{Z}_P(s, f)$ and $Z_2(s, f) \equiv G_P(f) + MB(s)$, where G_P is bounded and ρ_P-uniformly continuous uniformly in $P \in \mathscr{P}$ and B is standard Brownian motion. Then, as in the proof of Theorem 1.1, for $(s, f), (t, g) \in \tilde{\mathscr{F}}$,

$$d_1((s, f), (t, g)) \le d_2((s, f), (t, g)),$$

where

$$\begin{aligned}
d_1^2((s, f), (t, g)) &\equiv E(Z_1(s, f) - Z_1(t, g))^2 \\
&= (s \wedge t)\mathrm{Var}_P(f - g) + (t - s)^+\mathrm{Var}_P(g) \\
&\quad + (s - t)^+\mathrm{Var}_P(f), \\
d_2^2((s, f), (t, g)) &\equiv E(Z_2(s, f) - Z_2(t, g))^2 = \mathrm{Var}_P(f - g) + |t - s|M^2
\end{aligned}$$

and

$$\sup_{P\in\mathscr{P}} \mathrm{Var}_P(f) \le \sup_{P\in\mathscr{P}} P(F^2) \equiv M^2 < \infty.$$

Thus it follows from the Gaussian comparison theorems [e.g., Giné and Zinn (1986), Theorem 4.4, page 74] that

$$E\|\mathbb{Z}_P\|_{\tilde{\mathscr{F}}} = E\|Z_1\|_{\tilde{\mathscr{F}}} \le 2E\|Z_2\|_{\tilde{\mathscr{F}}} \le 2\{E\|G_P\|_{\mathscr{F}} + ME\|B\|_I\},$$

and hence, by (2.5),

$$\text{(a)} \qquad \sup_{P \in \mathscr{P}} E\|\mathbb{Z}_P\|_{\tilde{\mathscr{F}}} < \infty.$$

Furthermore, by an inequality of Fernique (1985) [see, e.g., Giné and Zinn (1986), Theorem 4.4.(c), page 74, where (4.7) should read (4.10)],

$$\text{(b)} \qquad \begin{aligned} E\|\mathbb{Z}_P\|_{\tilde{\mathscr{F}}'(\delta, d_2)} &= E\|Z_1\|_{\tilde{\mathscr{F}}'(\delta, d_2)} \\ &\leq 4E\|Z_2\|_{\tilde{\mathscr{F}}'(\delta, d_2)} + 13\delta\left(\log N\left(\delta/2, \tilde{\mathscr{F}}, d_2\right)\right)^{1/2}. \end{aligned}$$

Since $d_2^2 \geq \rho_P^2 \vee M^2|\cdot|$, the first term in (b) is bounded by

$$\text{(c)} \qquad E\|G_P\|_{\mathscr{F}'(\delta, \rho_P)} + ME\|B\|_{I(\delta^2/M^2)}.$$

Furthermore,

$$N\left(\delta, \tilde{\mathscr{F}}, d_2\right) \leq N\left(\frac{\delta}{2}, \mathscr{F}, \rho_P\right) N\left(\frac{3\delta^2}{4M^2}, I, |\cdot|\right)$$

[since (a) $\delta/2$-net for $(\mathscr{F}, \rho_P)$ and a $3\delta^2/(4M^2)$-net for I yield a δ-net for $(\tilde{\mathscr{F}}, d_2)$], so the second term on the right side of (b) is bounded by

$$\text{(d)} \qquad \begin{aligned} &13\delta\left\{\sqrt{\log N(\delta/4, \mathscr{F}, \rho_P)} + \sqrt{\log N(3\delta^2/(16M^2), I, |\cdot|)}\right\} \\ &\quad \leq 13\delta\sqrt{\log N(\delta/4, \mathscr{F}, \rho_P)} + 13\delta\sqrt{\log(16M^2/(3\delta^2))}. \end{aligned}$$

Combining (c) and (d) with (b) yields

$$\text{(e)} \qquad \begin{aligned} &\sup_{P \in \mathscr{P}} E\|\mathbb{Z}_P\|_{\tilde{\mathscr{F}}'(\delta, d_2)} \\ &\quad \leq 4 \sup_{P \in \mathscr{P}} E\|G_P\|_{\mathscr{F}'(\delta, \rho_P)} + 13 \sup_{P \in \mathscr{P}} \left\{\delta\sqrt{\log N(\delta/4, \mathscr{F}, \rho_P)}\right\} \\ &\qquad + 4ME\|B\|_{I(\delta^2/M^2)} + 13\delta\sqrt{\log(16M^2/(3\delta^2))} \end{aligned}$$

so that

$$\text{(f)} \qquad \lim_{\delta \downarrow 0} \sup_{P \in \mathscr{P}} E\|\mathbb{Z}_P\|_{\tilde{\mathscr{F}}'(\delta, d_2)} = 0$$

by (2.5) and Lemma 5.1. Finally, using

$$\rho_P^2(f, g) \leq P(f - g)^2 \leq 4P(F^2) \leq 4M^2,$$

it follows that

$$\text{(g)} \qquad \begin{aligned} d_2^2((s, f), (t, g)) &= \rho_P^2(f, g) + M^2|t - s| \\ &\leq \{2M + M^2\}\tilde{\rho}_P((s, f), (t, g)). \end{aligned}$$

We conclude from (f) and (g) that

$$\text{(h)} \qquad \lim_{\delta \downarrow 0} \sup_{P \in \mathscr{P}} E\|\mathbb{Z}_P\|_{\tilde{\mathscr{F}}'(\delta, \tilde{\rho}_P)} = 0.$$

The conclusion follows from (a) and (h). □

To complete the proof of Theorem 2.1, it remains to prove that (B) implies (D) and that (D) implies (E). The first of these is the big job. As we noted before, the proof of (B) implies (D) is long, but follows closely the "if" part of Dudley's (1984) proof of his Theorem 4.1.1, pages 27–31. The main tools needed are: (i). Uniform in $P \in \mathscr{P}$ bounds on the Prohorov distance between the finite-dimensional laws of the process $\mathbb{X}_n$ and the corresponding P-Brownian bridge $\mathbb{X} \cong G_P$. These are obtained from Theorem 5.1 below which is derived from the bounds of Yurinskii (1977). (ii) A uniform in $P \in \mathscr{P}$ weak approximation for sums of mean-zero uniformly square integrable random vectors based on the bounds of (i). This result adds uniformity in P to the weak approximation result of Philipp (1980) [see Dudley (1984), Theorem 1.1.3, page 7] and extends (without interpolation to continuous time) the uniform in $P \in \mathscr{P}$ weak approximation result of Lai (1978) from R^1 to R^d.

THEOREM 5.1. *For $P \in \mathscr{P}$, let $X_1, \dots, X_n$ be iid P in R^d with $E_P X_i = 0$ and $\Sigma_P = E_P(X_i X_i^T)$. Suppose that $|X|$ is uniformly square integrable over $P \in \mathscr{P}$ (where $|\cdot|$ denotes the usual Euclidean norm):*

$$\sup_{P \in \mathscr{P}} E_P|X|^2 1_{[|X| \geq \lambda]} \to 0 \quad \text{as } \lambda \to \infty. \tag{5.1}$$

Let F_n denote the law of $n^{-1/2} S_n \equiv n^{-1/2} \sum_{i=1}^n X_i$ on R^d and let $F_{N(0, \Sigma_P)}$ denote the $N_d(0, \Sigma_P)$ law on R^d. Then for every $\varepsilon > 0$, the Prohorov distance $\pi(F_n, F_{N(0,\Sigma_P)})$ is bounded by

$$\begin{aligned} \pi\big(F_n, F_{N(0,\Sigma_P)}\big) \leq{}& 2\Big\{\varepsilon^{-2} E_P|X|^2 1_{[|X| \geq \varepsilon\sqrt{n}]} \vee \varepsilon\Big\} \\ &+ 2^{2/3}\Big(E_P|X|^2 1_{[|X| \geq \varepsilon\sqrt{n}]}\Big)^{1/3} \\ &+ C(dM)^{1/4}\varepsilon^{1/4}\Big(1 + \big|\log(2^3 \varepsilon M/d)\big|^{1/2}\Big), \end{aligned} \tag{5.2}$$

where C is an absolute constant and $\sup_{P \in \mathscr{P}} E_P|X|^2 \leq M$ [which follows for some constant M from (5.1)]. Hence

$$\sup_{P \in \mathscr{P}} \pi\big(F_n, F_{N(0,\Sigma_P)}\big) \to 0 \quad \text{as } n \to \infty. \tag{5.3}$$

Theorem 5.1 is the basic result needed to prove the following uniform in P weak approximation for sums of independent random variables in R^d.

THEOREM 5.2. *Let $\mathscr{P}$ be a collection of laws on R^d with mean zero and covariance matrices $\Sigma_P = E_P(XX^T)$ and suppose that (5.1) holds. Then there exists a family of probability spaces $\{(\tilde\Omega, \tilde\Sigma, \Pr_P)\colon P \in \mathscr{P}\}$ with random variables $\{X_i\}_{i \geq 1}$ and $\{Z_i\}_{i \geq 1}$ defined thereon (for each $P \in \mathscr{P}$) such that: X_i are iid P, Z_i are iid $N_d(0, \Sigma_P)$ and, for every $\varepsilon > 0$,*

$$\sup_{P \in \mathscr{P}} \Pr_P\left(n^{-1/2} \max_{m \leq n}\left|\sum_{i=1}^m X_i - \sum_{i=1}^m Z_i\right| \geq \varepsilon\right) \to 0 \quad \text{as } n \to \infty. \tag{5.4}$$

Complete proofs of Theorems 5.1, 5.2 and the proof of (B) implies (D) in Theorem 2.1 are given in Section 6.

PROOF OF (D) IMPLIES (E) IN THEOREM 2.1. We first show that (D) implies

$$\text{(a)} \qquad \limsup_{n\to\infty} \sup_{P\in\mathscr{P}} E_P \|\mathbb{X}_n\|_{\mathscr{F}}^{*p} < \infty \quad \text{for each } 0 < p < 2.$$

The following proof of (a) follows Pisier [(1975), Proposition 2.1]; or see Hoffman-Jørgensen [(1976), Lemma 4.9, page 94]. Write $\mathbb{X}_n = n^{-1/2}\sum_{i=1}^n \mathbb{V}_i$ (so $\mathbb{V}_i = \delta_{X_i} - P$). Let $X_1', X_2', \dots$ be independent copies of $X_1, X_2, \dots,$ let $\mathbb{X}_n' = n^{-1/2}\sum_{i=1}^n \mathbb{V}_i'$ denote the corresponding empirical process and let $\mathbb{X}_n^s \equiv \mathbb{X}_n - \mathbb{X}_n' = n^{-1/2}\sum_{i=1}^n (\mathbb{V}_i - \mathbb{V}_i') \equiv n^{-1/2}\sum_{i=1}^n \mathbb{V}_i^s$ denote the symmetrized process. Let G_P, G_P' denote independent P-Brownian bridge processes and $G_P^s \equiv G_P - G_P'$. Then it follows from (D) and $G_P^s =_d \sqrt{2}\, G_P$ that

$$\text{(b)} \qquad \begin{aligned} \limsup_{n\to\infty} \sup_{P\in\mathscr{P}} \Pr\nolimits_P^*\left(\|\mathbb{X}_n^s\|_{\mathscr{F}} \ge \lambda\right) &\le \sup_{P\in\mathscr{P}} \Pr\nolimits_P\left(\|G_P^s\|_{\mathscr{F}} \ge \frac{\lambda}{2}\right) \\ &\le \sup_{P\in\mathscr{P}} \Pr\nolimits_P\left(\|G_P\|_{\mathscr{F}} \ge \frac{\lambda}{2\sqrt{2}}\right) \end{aligned}$$

for each $\lambda > 0$. Hence, for every $0 < \varepsilon \le 1/4$, there is a $\lambda = \lambda(\varepsilon) > 0$ such that

$$\text{(c)} \qquad \sup_{n\ge 1} \sup_{P\in\mathscr{P}} \Pr\nolimits_P^*\left(\|\mathbb{X}_n^s\|_{\mathscr{F}} > \lambda\right) \le \varepsilon.$$

Then, by Lévy's inequality [see, e.g., Araujo and Giné (1980), Theorem 2.6, page 10; for the first part of this for nonmeasurable maps, see Dudley (1984), Lemma 3.2.11; the analogue of the second inequality in Araujo and Giné's Theorem 2.6 for nonmeasurable maps which we use here can be proved similarly],

$$\text{(d)} \qquad \sup_{n\ge 1} \sup_{P\in\mathscr{P}} \Pr\nolimits_P\left(\max_{1\le i\le n} \|\mathbb{V}_i^s\|_{\mathscr{F}}^* > \lambda\sqrt{n}\right) \le 2\varepsilon,$$

and (d) implies that

$$\text{(e)} \qquad \sup_{P\in\mathscr{P}} \Pr\nolimits_P\left(\|\mathbb{V}_1^s\|_{\mathscr{F}}^* > \lambda\sqrt{n}\right) \le 1 - (1-2\varepsilon)^{1/n} \le \frac{\log 2}{n}$$

for all $n \ge 1$ since $\varepsilon \le 1/4$. But (e) yields, by monotonicity,

$$\text{(f)} \qquad \sup_{t>\lambda_0} \left\{ t^2 \sup_{P\in\mathscr{P}} \Pr\nolimits_P\left(\|\mathbb{V}_1^s\|_{\mathscr{F}}^* > t\right) \right\} \le 2\lambda_0^2 \log 2,$$

where $\lambda_0 \equiv \lambda(1/4)$. Now the argument proceeds by scaling on sample size: For a fixed integer $k \ge 1$, define

$$\text{(g)} \qquad \begin{aligned} Z_n^k &\equiv \left\{ \sqrt{nk}\,\mathbb{X}_{nk}^s - \sqrt{(n-1)k}\,\mathbb{X}_{(n-1)k}^s \right\} / \sqrt{k} \\ &= (1/\sqrt{k}) \sum_{j=(n-1)k+1}^{nk} \mathbb{V}_j^s. \end{aligned}$$

Then $Z_1^k, Z_2^k, \ldots$ are iid and can play the role of $\{\mathbb{Y}_i^s: i = 1, 2, \ldots\}$ in (c) and (f) with the same $\lambda = \lambda(\varepsilon)$ since we have

$$\text{(h)} \qquad \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_j^k = \mathbb{X}_{nk}^s,$$

and hence (D) implies that (b) holds with $\mathbb{X}_n^s$ replaced by $\mathbb{X}_{nk}^s$. The conclusion is that (f) holds, with $\mathbb{Y}_1^s$ replaced by Z_1^k for every $k = 1, 2, \ldots$. But since $Z_1^k = \mathbb{X}_k^s$, we conclude that

$$\text{(i)} \qquad \sup_{t > \lambda_0} \left\{ t^2 \sup_{P \in \mathscr{P}} \Pr_P(\|\mathbb{X}_k^s\|_{\mathscr{F}}^* > t) \right\} \le 2\lambda_0^2 \log 2$$

for every $k = 1, 2, \ldots$. But

$$\text{(j)} \qquad \Pr_P(\|\mathbb{X}_n\|_{\mathscr{F}}^* > t + a) \le \frac{\Pr_P(\|\mathbb{X}_n^s\|_{\mathscr{F}}^* > t)}{\Pr_P(\|\mathbb{X}_n\|_{\mathscr{F}}^* \le a)}$$

for $t, a > 0$ and, by D,

$$\begin{aligned} \liminf_{n\to\infty} \inf_{P \in \mathscr{P}} \Pr_P(\|\mathbb{X}_n\|_{\mathscr{F}}^* \le a) &= 1 - \limsup_{n\to\infty} \sup_{P\in\mathscr{P}} \Pr_P(\|\mathbb{X}_n\|_{\mathscr{F}}^* > a) \\ \text{(k)} \qquad &\ge 1 - \sup_{P\in\mathscr{P}} \Pr_P(\|G_P\|_{\mathscr{F}} > a/2) \\ &\ge 1/2 \end{aligned}$$

for any $a \ge a_0$ sufficiently large. It follows easily from (i)–(k) that

$$\sup_{t>\lambda_0} \left\{ (t+a)^2 \sup_{P\in\mathscr{P}} \Pr_P(\|\mathbb{X}_n\|_{\mathscr{F}}^* > t + a) \right\} \le 4(\lambda_0 + a)^2 \log 2, \qquad n \ge \text{some } N_0,$$

for $a \ge a_0$, and hence, with $b \equiv \lambda_0 + a_0$, that

$$\text{(l)} \qquad \sup_{s>b} \left\{ s^2 \sup_{P\in\mathscr{P}} \Pr_P(\|\mathbb{X}_n\|_{\mathscr{F}}^* > s) \right\} \le 4b^2 \log 2, \qquad n \ge \text{some } N_0.$$

But (l) implies that

$$\begin{aligned} E_P \|\mathbb{X}_n\|_{\mathscr{F}}^{*p} &= \int_0^\infty p t^{p-1} \Pr_P(\|\mathbb{X}_n\|_{\mathscr{F}}^* > t)\, dt \\ \text{(m)} \qquad &\le p \int_0^b t^{p-1}\, dt + 4b^2 \log 2 p \int_b^\infty t^{p-3}\, dt \\ &= b^p \left(1 + 4 \log 2 \frac{p}{2-p} \right) < \infty \end{aligned}$$

uniformly in $P \in \mathscr{P}$ and $n \ge N_0$; that is, (a) holds.

Next, Ottaviani's inequality yields uniform integrability of

$$\left(n^{-1/2} \max_{k \le n} \|k(\mathbb{P}_k - P)\|_{\mathscr{F}}^* \right)^p$$

uniformly in $P \in \mathscr{P}$ and hence uniform integrability of

$$\left(n^{-1/2} \max_{k \le n} \left\| \sum_{i=1}^{k} \mathbb{Y}_i \right\|_{\mathscr{F}}\right)^p$$

uniformly in $P \in \mathscr{P}$ just as in Dudley and Philipp [(1983), page 526] and Philipp [(1980), page 80], but with the additional uniformity in $P \in \mathscr{P}$. Hence $(n^{-1/2} \max_{k \le n} \|k(\mathbb{P}_k - P) - \sum_{i=1}^{k} \mathbb{Y}_i\|_{\mathscr{F}}^{*})^p$ is uniformly integrable uniformly in $P \in \mathscr{P}$ and (E) follows via the following elementary proposition.

PROPOSITION 5.2. *Suppose that Y_n are nonnegative rv's defined on $\{(\Omega, \Sigma, \Pr_P)$: $P \in \mathscr{P}\}$ and that*:

$$\text{(i)} \qquad \lim_{n \to \infty} \sup_{P \in \mathscr{P}} \Pr_P(Y_n \ge \varepsilon) = 0 \quad \textit{for every } \varepsilon > 0.$$

$$\text{(ii)} \qquad \lim_{\lambda \to \infty} \limsup_{n \to \infty} \sup_{P \in \mathscr{P}} E_P Y_n 1_{[Y_n \ge \lambda]} = 0.$$

Then $\lim_{n \to \infty} \sup_{P \in \mathscr{P}} E_P Y_n = 0$; *that is*, $Y_n \to_1 0$ *uniformly in* $P \in \mathscr{P}$.

PROOF. Now for any $\lambda > 0$,

$$\begin{aligned} E_P Y_n &\le E_P Y_n 1_{[Y_n \le \lambda]} + E_P Y_n 1_{[Y_n \ge \lambda]} \\ \text{(a)} \qquad &\le \int_{(0, \lambda]} \Pr_P(Y_n \ge t)\, dt + E_P Y_n 1_{[Y_n \ge \lambda]}. \end{aligned}$$

Hence

$$\text{(b)} \qquad \sup_{P \in \mathscr{P}} E_P Y_n \le \int_{(0, \lambda]} \sup_{P \in \mathscr{P}} \Pr_P(Y_n \ge t)\, dt + \sup_{P \in \mathscr{P}} E_P Y_n 1_{[Y_n \ge \lambda]}.$$

This yields, by dominated convergence and (i),

$$\text{(c)} \qquad \lim_{n \to \infty} \sup_{P \in \mathscr{P}} E_P Y_n \le \limsup_{n \to \infty} \sup_{P \in \mathscr{P}} E_P Y_n 1_{[Y_n \ge \lambda]}$$

for any $\lambda > 0$; letting $\lambda \to \infty$ yields the conclusion by (ii). □

To prove Theorem 2.3, we first need to establish the following $\mathscr{P}$-uniform Glivenko–Cantelli theorem. The measurability condition we will need is that $\mathscr{F}$ be nearly linearly supremum measurable for all $P \in \mathscr{P}$ [see, e.g., Giné and Zinn (1984), Definition 2.3, page 935]; when this holds we write $\mathscr{F} \in$ NLSM($\mathscr{P}$). Recall that for $r > 0$, $N_F^{(r)}(\varepsilon, \mathscr{F})$, $\varepsilon > 0$, denotes Pollard's (1982) (rth power) combinatorial entropy of $\mathscr{F}$ relative to an envelope function F of $\mathscr{F}$.

THEOREM 5.3. *Suppose $\mathscr{F} \in NLSM(\mathscr{P})$ and that*:

(i) $N_K^{(1)}(\delta, \mathscr{F}_K) < \infty$ *for every* $\delta > 0$ *and* $K > 0$, *where*

$$\mathscr{F}_K \equiv \{f 1_{[F \le K]} : f \in \mathscr{F}\}.$$

(ii) *F is $\mathscr{P}$-uniformly integrable*; *that is*, $\sup_{P \in \mathscr{P}} E_P F 1_{[F \ge \lambda]} \to 0$ *as* $\lambda \to \infty$.

Then, for every $\varepsilon > 0$,

$$\sup_{P \in \mathscr{P}} \Pr_P^* \Big\{ \max_{m \geq n} \|\mathbb{P}_m - P\|_{\mathscr{F}} \geq \varepsilon \Big\} \to 0 \quad \text{as } n \to \infty. \tag{5.5}$$

Recall that $\mathbf{G} \equiv \mathscr{F} \cup \mathscr{F}^2 \cup (\mathscr{F}')^2$.

COROLLARY 5.1. *Suppose that* $\mathbf{G} \in NLSM(\mathscr{P})$ *and that*:

(i) $N_F^{(2)}(\delta, \mathscr{F}) < \infty$ *for every* $\delta > 0$.
(ii) F *is* $\mathscr{P}$*-uniformly square integrable.*

Then, for every $\varepsilon > 0$,

$$\sup_{P \in \mathscr{P}} \Pr_P^* \Big\{ \max_{m \geq n} \|\mathbb{P}_m - P\|_{\mathbf{G}} \geq \varepsilon \Big\} \to 0 \quad \text{as } n \to \infty. \tag{5.6}$$

Our proofs of Theorem 5.3 and Corollary 5.1 depend on the following lemmas.

LEMMA 5.2. *If* $N_F^{(r)}(\delta, \mathscr{F}) < \infty$ *and* $\mathscr{F}_K \equiv \{f 1_{[F \leq K]}\colon f \in \mathscr{F}\}$, *then* $N_K^{(r)}(\delta, \mathscr{F}_K) \leq N_F^{(r)}(\delta, \mathscr{F}) < \infty$.

PROOF. Since $N_F^{(r)}(\delta, \mathscr{F}) < \infty$, given a set S we can find $\{f_1, \dots, f_m\} \subset \mathscr{F}$ so that for each $f \in \mathscr{F}$, there is an i with

$$\begin{aligned}
\sum_{x \in S} \left| f(x) 1_{[F(x) \leq K]} - f_i(x) 1_{[F(x) \leq K]} \right|^r &= \sum_{x \in S \cap [F(x) \leq K]} \left| f(x) - f_i(x) \right|^r \\
&\leq \delta^r \sum_{x \in S \cap [F(x) \leq K]} F(x)^r \\
&\leq \delta^r \sum_{x \in S \cap [F(x) \leq K]} K^r \\
&\leq \delta^r \sum_{x \in S} K^r,
\end{aligned}$$

where

$$m \leq N_K^{(r)}(\delta, S, \mathscr{F}_K) \leq N_F^{(r)}(\delta, \mathscr{F}) < \infty. \qquad \square$$

LEMMA 5.3. *Suppose that* $\mathscr{F}$ *has envelope function* K. *If* $N_K^{(2)}(\delta, \mathscr{F}) < \infty$ *for all* $\delta > 0$ *and* $\tilde{\mathscr{F}} \equiv \{fg\colon f, g \in \mathscr{F}\}$, *then* $N_{K^2}^{(1)}(\delta, \tilde{\mathscr{F}}) < \infty$ *for all* $\delta > 0$.

PROOF. We show, in fact, that

$$N_{K^2}^{(1)}(2\delta, \tilde{\mathscr{F}}) \leq \binom{N_K^{(2)}(\delta, \mathscr{F}) + 1}{2} = \binom{N_K^{(2)}(\delta, \mathscr{F})}{2} + N_K^{(2)}(\delta, \mathscr{F}).$$

Let $\{f_1, \dots, f_m\} \equiv \mathscr{F}_\delta$ be chosen so that given any $f \in \mathscr{F}$, there exists $f' \in \mathscr{F}_\delta$ so that

$$\sum_{x \in S} [f(x) - f'(x)]^2 \leq \delta^2 \sum_{x \in S} K^2, \tag{a}$$

where $m \le N_K^{(2)}(\delta, \mathscr{F})$. Then for any $f, g \in \mathscr{F}$ with f_i and g_i chosen so that (a) is true (if $f = g$ we can choose $f_i = g_i$), we have, with $n \equiv \#(S)$,

$$\begin{aligned}
\frac{1}{n}\sum_{x\in S} |fg - f_i g_i| &\le \frac{1}{n}\sum_{x\in S}|f|\,|g-g_i| + \frac{1}{n}\sum_{x\in S}|g_i|\,|f-f_i| \\
&\le K\left\{\frac{1}{n}\sum_{x\in S}|g-g_i| + \frac{1}{n}\sum_{x\in S}|f-f_i|\right\} \\
&\le K\left\{\left(\frac{1}{n}\sum_{x\in S}|g-g_i|^2\right)^{1/2} + \left(\frac{1}{n}\sum_{x\in S}|f-f_i|^2\right)^{1/2}\right\} \\
&\le K\{2\delta K\} = 2\delta K^2. \qquad \square
\end{aligned}$$

LEMMA 5.4. *Suppose that $\mathscr{F} \in NLSM(\mathscr{P})$ and has envelope function K, a constant, and that $N_K^{(1)}(\delta, \mathscr{F}) < \infty$ for all $\delta > 0$. Then, for every $\varepsilon > 0$,*

$$\sup_{P\in\mathscr{P}} \Pr_P^*\left\{\max_{m\ge n}\|\mathbb{P}_m - P\|_{\mathscr{F}} > \varepsilon\right\} \to 0 \quad \text{as } n \to \infty. \tag{5.7}$$

PROOF. Let $\varepsilon > 0$. We will prove (5.7) by showing that we can choose $n(\varepsilon)$ so large that

$$\sup_{P\in\mathscr{P}} \Pr_P^*\left\{\max_{m\ge n}\|\mathbb{P}_m - P\|_{\mathscr{F}} > \varepsilon\right\} < \varepsilon \quad \text{for } n \ge n(\varepsilon). \tag{a}$$

In fact, a choice of $n(\varepsilon)$ that works is

$$n(\varepsilon) \ge \max\left\{\frac{8K^2}{\varepsilon^2}, \frac{256K^2}{\varepsilon^2}\log N_K^{(1)}\left(\frac{\varepsilon}{8K}, \mathscr{F}\right), n(\varepsilon, K)\right\}, \tag{b}$$

where $n(\varepsilon, K)$ is so large that

$$8\sum_{m=n(\varepsilon,K)}^{\infty} \exp\left(-\frac{m\varepsilon^2}{256K^2}\right) < \varepsilon. \tag{c}$$

First, note that since $\mathscr{F} \in$ NLSM$(\mathscr{P})$, for each $P \in \mathscr{P}$, there exists ${}_0\mathscr{F} = {}_0\mathscr{F}_P \subset \mathscr{F}$ with ${}_0\mathscr{F} \in$ LSM(P) satisfying

$$\Pr_P^*\left\{\max_{m\ge n}\|\mathbb{P}_m - P\|_{\mathscr{F}} > \varepsilon\right\} = \Pr_P\left\{\max_{m\ge n}\|\mathbb{P}_m - P\|_{{}_0\mathscr{F}} > \varepsilon\right\}$$

and hence

$$\text{(d)}\quad \sup_{P\in\mathscr{P}} \Pr_P^*\Big\{\max_{m\geq n}\|\mathbb{P}_m - P\|_{\mathscr{F}} > \varepsilon\Big\} = \sup_{P\in\mathscr{P}} \Pr_P\Big\{\max_{m\geq n}\|\mathbb{P}_m - P\|_{{}_0\mathscr{F}} > \varepsilon\Big\}.$$

For simplicity of notation we will not distinguish between ${}_0\mathscr{F}$ and $\mathscr{F}$ in the rest of this proof and we will write Pr instead of Pr* except in the final line. The proof uses the symmetrized empirical measure

$$\mathbb{P}_n^0(A) \equiv \frac{1}{n}\sum_{i=1}^{n} \sigma_i 1_A(X_i),$$

where $\sigma_1, \sigma_2, \dots$ are iid Rademacher rv's [so that $P(\sigma_i = \pm 1) = 1/2$]. By symmetrization [e.g., Pollard (1984), equation (11), page 15, or Giné and Zinn (1986), Lemma 2.3, page 63] it follows that

$$\text{(e)}\quad \begin{aligned} &\sup_{P\in\mathscr{P}} \Pr_P\{\|\mathbb{P}_n - P\|_{\mathscr{F}} > \varepsilon\} \\ &\leq 4 \sup_{P\in\mathscr{P}} \Pr_P\Big\{\|\mathbb{P}_n^0\|_{\mathscr{F}} > \frac{\varepsilon}{4}\Big\} = 4 \sup_{P\in\mathscr{P}} E\Big[\Pr_P\Big\{\|\mathbb{P}_n^0\|_{\mathscr{F}} > \frac{\varepsilon}{4}\Big|\underline{X}_n\Big\}\Big] \end{aligned}$$

for $n \geq 8K^2/\varepsilon^2$. The inequality in (e) depends on

$$\begin{aligned} &\sup_{P\in\mathscr{P}}\sup_{f\in\mathscr{F}} \Pr_P\Big\{|(\mathbb{P}_n - P)(f)| > \frac{\varepsilon}{2}\Big\} \\ &\leq \sup_{P\in\mathscr{P}}\sup_{f\in\mathscr{F}} \frac{4E_P f^2}{n\varepsilon^2} \\ &\leq \frac{4K^2}{n\varepsilon^2} \leq \frac{1}{2} \quad \text{for } n \geq \frac{8K^2}{\varepsilon^2} \end{aligned}$$

to obtain the factor of 4 on the right hand side. Given $\underline{X}_n$, choose functions $g_1, \dots, g_M$, $M = N_K^{(1)}(\varepsilon/8K, \mathscr{F})$ so that

$$\min_j \mathbb{P}_n|f - g_j| \leq \frac{\varepsilon}{8} \quad \text{for each } f \in \mathscr{F}.$$

Write f^* for a g_j at which the minimum is achieved. For any function g,

$$|\mathbb{P}_n^0 g| = \left|n^{-1}\sum_{i=1}^{n}\sigma_i g(X_i)\right| \leq n^{-1}\sum_{i=1}^{n}|g(X_i)| = \mathbb{P}_n|g|.$$

Choose $g = f - f^*$ for each f in turn to obtain, by Hoeffding's inequality at the next to last step [Hoeffding (1963), Theorem 2; see, e.g., Shorack and

Wellner (1986), Inequality A.4.6],

$$
\begin{aligned}
&\mathrm{Prob}\left\{\sup_{\mathscr{F}}|\mathbb{P}_n^0 f| > \frac{\varepsilon}{4}\,\Big|\,\underline{X}_n\right\}\\
&\quad\le \mathrm{Prob}\left\{\sup_{\mathscr{F}}\left[|\mathbb{P}_n^0 f^*| + \mathbb{P}_n|f - f^*|\right] > \frac{\varepsilon}{4}\,\Big|\,\underline{X}_n\right\}\\
&\quad\le \mathrm{Prob}\left\{\max_j|\mathbb{P}_n^0 g_j| > \frac{\varepsilon}{8}\,\Big|\,\underline{X}_N\right\} \qquad \left[\text{because } \mathbb{P}_n|f - f^*| \le \frac{\varepsilon}{8}\right]\\
\text{(f)}\qquad &\quad\le N_K^{(1)}(\varepsilon/8K, \mathscr{F})\max_j \mathrm{Prob}\left\{|\mathbb{P}_n^0 g_j| > \frac{\varepsilon}{8}\,\Big|\,\underline{X}_n\right\}.\\
&\quad\le N_K^{(1)}(\varepsilon/8K, \mathscr{F})\max_j \mathrm{Prob}\left\{\left|\sum_{i=1}^n \sigma_i g_j(X_i)\right| > n\frac{\varepsilon}{8}\,\Big|\,\underline{X}_N\right\}\\
&\quad\le 2\exp\left[H_K^{(1)}\left(\frac{\varepsilon}{8K}, \mathscr{F}\right) - 2(n\varepsilon/8)^2 \Big/ \sum_{i=1}^n \left(2g_j(X_i)\right)^2\right]\\
&\qquad\qquad\qquad\qquad\qquad\qquad [\text{by Hoeffding's inequality}]\\
&\quad\le 2\exp\left[-n\varepsilon^2/256K^2\right] \qquad \left[\text{using } |g_j| \le K \text{ and (b)}\right].
\end{aligned}
$$

Combining (c), (d), (e) and (f) yields, for $n \ge n(\varepsilon)$,

$$
\text{(g)}\qquad
\begin{aligned}
\sup_{P\in\mathscr{P}} \mathrm{Pr}_P^*\left\{\max_{m\ge n}\|\mathbb{P}_m - P\|_{\mathscr{F}} > \varepsilon\right\} &\le 4\sum_{m=n}^{\infty}\sup_{P\in\mathscr{P}}\mathrm{Pr}_P^*\left\{\|\mathbb{P}_m^0\|_{\mathscr{F}} > \frac{\varepsilon}{4}\right\}\\
&\le 8\sum_{m=n}^{\infty}\exp\left(-\frac{n\varepsilon^2}{256K^2}\right) < \varepsilon,
\end{aligned}
$$

and hence (a) holds. □

With these three lemmas as preparation we can prove Theorem 5.3.

PROOF OF THEOREM 5.3. In view of (ii) we can choose K so large that

$$
\text{(a)}\qquad \sup_{P\in\mathscr{P}} E_P F 1_{[F>K]} < \frac{\varepsilon}{4}.
$$

Then, since

$$
\begin{aligned}
\|\mathbb{P}_m - P\|_{\mathscr{F}} &\le \|\mathbb{P}_m - P\|_{\mathscr{F}_K} + \sup_{f\in\mathscr{F}}\left|(\mathbb{P}_m - P)f1_{[F>K]}\right|\\
&\le \|\mathbb{P}_m - P\|_{\mathscr{F}_K} + \mathbb{P}_m F1_{[F>K]} + PF1_{[F>K]}\\
\text{(b)}\qquad &\le \|\mathbb{P}_m - P\|_{\mathscr{F}_K} + \left|(\mathbb{P}_m - P)F1_{[F>K]}\right| + 2PF1_{[F>K]}\\
&\le \|\mathbb{P}_m - P\|_{\mathscr{F}_K} + \left|(\mathbb{P}_m - P)F1_{[F>K]}\right| + \frac{\varepsilon}{2} \qquad [\text{by (a)}],
\end{aligned}
$$

it follows that

$$
\begin{aligned}
&\sup_{P\in\mathscr{P}} \Pr_P^*\Big\{\max_{m\ge n}\|\mathbb{P}_m - P\|_{\mathscr{F}} > \varepsilon\Big\}\\
\text{(c)}\qquad &\le \sup_{P\in\mathscr{P}} \Pr_P^*\Big\{\max_{m\ge n}\|\mathbb{P}_m - P\|_{\mathscr{F}_K} > \frac{\varepsilon}{4}\Big\}\\
&\quad + \sup_{P\in\mathscr{P}} \Pr_P\Big\{\max_{m\ge n}\big|(\mathbb{P}_m - P)F1_{[F>K]}\big| > \frac{\varepsilon}{4}\Big\}\\
&\le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \qquad [\text{for } n \ge \text{some } N(\varepsilon)],
\end{aligned}
$$

by Lemma 5.4 [since $N_K^{(1)}(\delta, \mathscr{F}_K) < \infty$ by Lemma 5.2] and Chung's (1951) uniform strong law of large numbers, respectively. □

Proof of Corollary 5.1. Now $N_F^{(2)}(\delta, \mathscr{F}) < \infty$ for every $\delta > 0$ implies that $N_K^{(2)}(\delta, \mathscr{F}_K) < \infty$ for every $\delta > 0$ by Lemma 5.2. By Lemma 5.3, this implies that $N_{K^2}^{(1)}(\delta, \tilde{\mathscr{F}}_K) < \infty$ for all $\delta > 0$, where

$$\tilde{\mathscr{F}}_K \equiv \{fg1_{[F^2\le K^2]}: f, g \in \mathscr{F}\}.$$

Hence the conditions of Theorem 5.3 are satisfied with

$$\tilde{\mathscr{F}} \equiv \{fg: f, g \in \mathscr{F}\}$$

replacing $\mathscr{F}$. Thus (5.6) holds with $\|\mathbb{P}_n - P\|_{\mathbf{G}}$ replaced by $\|\mathbb{P}_n - P\|_{\tilde{\mathscr{F}}}$. Since $\|\mathbb{P}_n - P\|_{\mathbf{G}} \le 4\|\mathbb{P}_n - P\|_{\tilde{\mathscr{F}}}$, (5.6) also holds. □

Proof of Theorem 2.3. Our proof of Theorem 2.3 is a modification of Pollard's (1982) proof of his Theorem 7. For the record, we give our proof in detail.

First note that (i) implies that there is a constant $M < \infty$ so that

$$\text{(a)}\qquad \sup_{P\in\mathscr{P}} P(F^2) = \sup_{P\in\mathscr{P}} E_P F^2 \le M^2 < \infty.$$

This constant enters repeatedly in the remainder of the proof.

Now let $\delta_j \equiv 2^{-j}$ for $j \ge 1$ and set $H_j \equiv \log N_F^{(2)}(2^{-j}, \mathscr{F})$ so that $\sum_{j=1}^\infty \delta_j H_j^{1/2} < \infty$ by condition (ii). Select a sequence of positive numbers $\{\eta_j\}$ for which

$$\text{(b)}\qquad \sum_{j=1}^\infty \eta_j < \infty,$$

$$\text{(c)}\qquad \eta_j \ge \left(144M^2\delta_j^2 H_j\right)^{1/2} \quad \left(\text{so that } H_j \le \frac{\eta_j^2}{144M^2\delta_j^2}\right),$$

$$\text{(d)}\qquad \sum_{j=1}^\infty \exp\left(-\frac{\eta_j^2}{72\delta_j^2 M^2}\right) < \infty.$$

This is possible because of the growth condition (ii) on $\log N_F^{(2)}(\cdot, \mathscr{F})$. For example, $\eta_j \equiv \max\{j\delta_j, 12M\delta_j H_j^{1/2}\}$ works.

We now give our choices of $\delta = \delta(\varepsilon) > 0$ and $n = n(\varepsilon, \delta)$ (not dependent on $P \in \mathscr{P}$) which yield $\mathscr{F} \in \mathrm{AEC}_u(\mathscr{P}, e_p)$; that is, (2.6) holds with $d_p = e_p$. Choose an integer $r = r(\varepsilon)$ so large that, with $\eta \equiv \varepsilon/8$,

$$\text{(e)} \qquad 2\sum_{j=r+1}^{\infty} \exp\left(-\frac{\eta_j^2}{72M^2\delta_j^2}\right) < \frac{\varepsilon}{16} \qquad [\text{by (d)}],$$

$$\text{(f)} \qquad \sum_{j=r+1}^{\infty} \eta_j < \frac{\varepsilon}{16} \qquad [\text{by (b)}],$$

$$\text{(g)} \qquad 2\exp\left(-\frac{\eta^2}{72\delta_r^2M^2}\right) < \frac{\varepsilon}{16} \qquad [\text{since } \delta_r \to 0],$$

$$\text{(h)} \qquad \eta^2 \geq 144M^2H_r\delta_r^2 \qquad [\text{by condition (i)}],$$

all hold. Now choose $\delta \equiv \delta(\varepsilon) > 0$ so that

$$\text{(i)} \qquad \delta \leq \min\left\{\sqrt{\frac{2}{3}}\,\delta_{r(\varepsilon)}M, \frac{\varepsilon}{\sqrt{2}}\right\}.$$

With this $\delta = \delta(\varepsilon)$, we choose $n \equiv n(\varepsilon, \delta)$ so large that

$$\text{(j)} \qquad \sup_{P\in\mathscr{P}} \Pr_P^*\left\{\|\mathbb{P}_n - P\|_{\mathbf{G}} > \frac{\delta^2}{2}\right\} \leq \frac{\varepsilon}{16}$$

and

$$\text{(k)} \qquad \sup_{P\in\mathscr{P}} \Pr_P\{\mathbb{P}_nF^2 > PF^2 + M^2\} < \frac{\varepsilon}{16}.$$

Such a choice is possible in view of the $\mathscr{P}$-uniform Glivenko–Cantelli Theorem 5.3.

To prove $\mathscr{F} \in \mathrm{AEC}_u(\mathscr{P}, \rho_P)$, we will show first that this choice of δ and $n(\varepsilon,\delta)$ imply that

$$\text{(l)} \qquad \sup_{P\in\mathscr{P}} \Pr_P^*\{\|\mathbb{X}_n\|_{\mathscr{F}'(\delta, e_P)} > \varepsilon\} \leq \varepsilon \quad \text{for all } n \geq n(\varepsilon,\delta);$$

that is, that $\mathscr{F} \in \mathrm{AEC}_u(\mathscr{P}, e_P)$. Write $\underline{X}_n \equiv \{X_1, \dots, X_n\}$.

Since $\mathscr{F} \in \mathrm{NLDM}(\mathscr{P})$, for each $P \in \mathscr{P}$ there exists ${}_0\mathscr{F} = {}_0\mathscr{F}_P \subset \mathscr{F}$ with ${}_0\mathscr{F} \in \mathrm{LDM}(P)$ satisfying

$$\Pr_P^*\{\|\mathbb{X}_n\|_{\mathscr{F}'(\delta, e_P)} > \varepsilon\} = \Pr_P\{\|\mathbb{X}_n\|_{{}_0\mathscr{F}'(\delta, e_P)} > \varepsilon\}$$

and hence

$$\sup_{P\in\mathscr{P}} \Pr_P^*\{\|\mathbb{X}_n\|_{\mathscr{F}'(\delta, e_P)} > \varepsilon\} = \sup_{P\in\mathscr{P}} \Pr_P\{\|\mathbb{X}_n\|_{{}_0\mathscr{F}'(\delta, e_P)} > \varepsilon\}.$$

For simplicity of notation we will not distinguish between ${}_0\mathscr{F}$ and $\mathscr{F}$ in the rest of this proof and we will write Pr instead of Pr* in the remainder of the proof.

By symmetrization [e.g., Pollard (1984), Lemma II.8, page 14; or Giné and Zinn (1986), Lemma 2.3, page 63]

$$\text{(m)} \qquad \sup_{P\in\mathscr{P}} \Pr_P\{\|\mathbb{X}_n\|_{\mathscr{F}'(\delta,e_P)} > \varepsilon\} \le 4\sup_{P\in\mathscr{P}} \Pr_P\left\{\|\mathbb{P}_n^0\|_{\mathscr{F}'(\delta,e_P)} > \frac{\varepsilon}{4\sqrt{n}}\right\}$$

for all $n \ge 1$, where $\mathbb{P}_n^0$ is the symmetrized empirical measure $\mathbb{P}_n^0 \equiv n^{-1}\sum_{i=1}^n \varepsilon_i \delta_{X_i}$ and $\{\varepsilon_i\}$ are iid Rademacher rv's. The validity of (m) depends on

$$\begin{aligned}
&\sup_{P\in\mathscr{P}} \sup_{(f,g)\in\mathscr{F}'(\delta,e_P)} \Pr_P\left\{\left|\sqrt{n}(\mathbb{P}_n - P)(f-g)\right| > \varepsilon\right\} \\
&\qquad \le \sup_{P\in\mathscr{P}} \sup_{(f,g)\in\mathscr{F}'(\delta,e_P)} \frac{\operatorname{Var}_P(f-g)}{\varepsilon^2} \le \frac{\delta^2}{\varepsilon^2} \le \frac{1}{2}
\end{aligned}$$

by the choice (i) of δ.

We use (j) to replace $\mathscr{F}'(\delta, e_P)$ on the right side in (m) by $\mathscr{F}'((3/2)^{1/2}\delta, e_{\mathbb{P}_n})$. Let $D_n \equiv \|\mathbb{P}_n - P\|_{\mathbf{G}}$ and let B_n^c denote the event $[\mathbb{P}_n F^2 > PF^2 + M^2]$ in (k). It follows from our choice of n and (j) and (k) that the right side of (m) is bounded by

$$\text{(n)} \qquad \begin{aligned}
&4\sup_{P\in\mathscr{P}} \Pr_P\left\{\|\mathbb{P}_n^0\|_{\mathscr{F}'(\delta,e_P)} > \frac{\varepsilon}{4\sqrt{n}}, D_n > \frac{\delta^2}{2}\right\} \\
&\qquad + 4\sup_{P\in\mathscr{P}} \Pr_P\left\{\|\mathbb{P}_n^0\|_{\mathscr{F}'(\delta,e_P)} > \frac{\varepsilon}{4\sqrt{n}}, D_n \le \frac{\delta^2}{2}\right\} \\
&\le \frac{\varepsilon}{4} + 4\sup_{P\in\mathscr{P}} \Pr_P\left\{\|\mathbb{P}_n^0\|_{\mathscr{F}'((3/2)^{1/2}\delta, e_{\mathbb{P}_n})} > \frac{\varepsilon}{4\sqrt{n}}\right\} \\
&\le \frac{\varepsilon}{2} + 4\sup_{P\in\mathscr{P}} E_P\left\{\operatorname{Prob}\left\{\|\mathbb{P}_n^0\|_{\mathscr{F}'((3/2)^{1/2}\delta, e_{\mathbb{P}_n})} > \frac{\varepsilon}{4\sqrt{n}} \,\middle|\, \underline{X}_n\right\} 1_{B_n}\right\}.
\end{aligned}$$

In view of (n), the desired inequality (1) will hold if we show that

$$\text{(o)} \qquad \operatorname{Prob}\left\{\|\mathbb{P}_n^0\|_{\mathscr{F}'((3/2)^{1/2}\delta, e_{\mathbb{P}_n})} > \frac{\varepsilon}{4\sqrt{n}} \,\middle|\, \underline{X}_n\right\} < \frac{\varepsilon}{8} \quad \text{on } B_n.$$

Choose finite subclasses $\mathscr{F}(1), \mathscr{F}(2), \dots$ of $\mathscr{F}$ such that

$$\text{(p)} \qquad \min_{\phi\in\mathscr{F}(i)} e_{\mathbb{P}_n}^2(f,\phi) \le \delta_i^2(\mathbb{P}_n F^2) \quad \text{for each fixed } f \in \mathscr{F}.$$

By Definition 1.1, $\mathscr{F}(i)$ need contain at most $\exp(H_i)$ functions [recall that $\delta_i \equiv 2^{-i}$ and $H_i \equiv \log N_F^{(2)}(2^{-i}, \mathscr{F})$]. For a given $f \in \mathscr{F}$, denote by f_i a function ϕ in $\mathscr{F}(i)$ for which the left-hand side of (p) achieves its minimum. Note that $e_{\mathbb{P}_n}(f, f_i) \to 0$ as $i \to \infty$. Thus, for any fixed r,

$$\text{(q)} \qquad f - f_r = \sum_{j=r+1}^{\infty} (f_j - f_{j-1})$$

pointwise on $\underline{X}_n$.

The proof of (o) breaks into two parts. The first is to show that for our choice of $r \geq r(\varepsilon)$, we have

$$\text{(r)} \qquad \operatorname{Prob}\left\{ \sup_{\mathscr{F}} \left| \mathbb{P}_n^0(f - f_r) \right| > \frac{\varepsilon}{16\sqrt{n}} \middle| \underline{X}_n \right\} < \frac{\varepsilon}{16}$$

on

$$B_n \equiv \left[\mathbb{P}_n F^2 \leq P(F^2) + M^2 \right].$$

The second part is to show that for our choice of r, we have

$$\text{(s)} \quad \operatorname{Prob}\left\{ \sup\left\{ \left| \mathbb{P}_n^0(f_r - g_r) \right| : f, g \in \mathscr{F}'\left((3/2)^{1/2}\delta, e_{\mathbb{P}_n} \right) \right\} > \frac{\varepsilon}{8\sqrt{n}} \middle| \underline{X}_n \right\} < \frac{\varepsilon}{16}$$

on B_n. Since

$$\text{(t)} \quad \begin{aligned} &\sup\left\{ \left| \mathbb{P}_n^0(f - g) \right| : f, g \in \mathscr{F}'\left((3/2)^{1/2}\delta, e_{\mathbb{P}_n} \right) \right\} \\ &\quad \leq 2 \sup_{\mathscr{F}} \left| \mathbb{P}_n^0(f - f_r) \right| + \sup\left\{ \left| \mathbb{P}_n^0(f_r - g_r) \right| : f, g \in \mathscr{F}'\left((3/2)^{1/2}\delta, e_{\mathbb{P}_n} \right) \right\}, \end{aligned}$$

the inequality (o) follows from (r)–(t).

To prove (r), use (f) and (q) to bound the left side of (r) by

$$\text{(u)} \quad \begin{aligned} &\operatorname{Prob}\left\{ \sup_{\mathscr{F}} \left| \mathbb{P}_n^0(f - f_r) \right| > \frac{1}{\sqrt{n}} \sum_{j=r+1}^{\infty} \eta_j \middle| \underline{X}_n \right\} \\ &\quad \leq \sum_{j=r+1}^{\infty} \operatorname{Prob}\left\{ \sup_{\mathscr{F}} \left| \mathbb{P}_n^0(f_j - f_{j-1}) \right| > \frac{\eta_j}{\sqrt{n}} \middle| \underline{X}_n \right\} \\ &\quad \leq \sum_{j=r+1}^{\infty} |\mathscr{F}_j| \, |\mathscr{F}_{j-1}| \sup_{\mathscr{F}} \operatorname{Prob}\left\{ \left| \mathbb{P}_n^0(f_j - f_{j-1}) \right| > \frac{\eta_j}{\sqrt{n}} \middle| \underline{X}_n \right\}, \end{aligned}$$

where $|\mathscr{F}_j| = \exp(H_j)$. Consider one of these last conditional probabilities, noting that

$$\mathbb{P}_n^0(f_j - f_{j-1}) = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i (f_j - f_{j-1})(X_i) \equiv \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i h_i.$$

Thus by Theorem 2 of Hoeffding (1963) [see, e.g., Inequality A.4.6 of Shorack and Wellner (1986)],

$$\text{(v)} \qquad \operatorname{Prob}\left\{ \left| n^{-1/2} \sum_{i=1}^{n} \varepsilon_i h_i \right| > \eta_j \middle| \underline{X}_n \right\} \leq 2 \exp\left(-\frac{2n\eta_j^2}{4\Sigma h_i^2} \right),$$

where

$$\sum_{i=1}^{n} h_i^2 = n e_{\mathbb{P}_n}^2(f_j, f_{j-1})$$
$$\le n\big(e_{\mathbb{P}_n}(f, f_j) + e_{\mathbb{P}_n}(f, f_{j-1})\big)^2$$
$$\le n(\mathbb{P}_n F^2)(\delta_j + \delta_{j-1})^2$$
$$\le 9n\delta_j^2[PF^2 + M^2] \qquad [\text{on } B_n]$$
$$\le 18n\delta_j^2 M^2 \qquad [\text{by (a)}].$$

Therefore on B_n, the sum in (u) is less than

$$\sum_{j=r+1}^{\infty} \exp(2H_j) 2\exp\left(-\frac{\eta_j^2}{36\delta_j^2 M^2}\right)$$
$$\le 2\sum_{j=r+1}^{\infty} \exp\left(-\frac{\eta_j^2}{72\delta_j^2 M^2}\right) \qquad [\text{by (c)}]$$
$$\le \frac{\varepsilon}{16} \qquad [\text{by (e)}];$$

hence (r) holds.

To prove (s), note that on the event $B_n \cap [e_{\mathbb{P}_n}(f, g) < \sqrt{3/2}\,\delta]$, we have

$$\text{(w)} \qquad \begin{aligned} e_{\mathbb{P}_n}(f_r, g_r) &\le e_{\mathbb{P}_n}(f_r, f) + e_{\mathbb{P}_n}(f, g) + e_{\mathbb{P}_n}(g, g_r) \\ &\le \left(\tfrac{3}{2}\right)^{1/2}\delta + 2\delta_r(\mathbb{P}_n F^2)^{1/2} \\ &\le \left(\tfrac{3}{2}\right)^{1/2}\delta + 2\delta_r[P(F^2) + M^2]^{1/2} \qquad [\text{on } B_n] \\ &< (1 + 2\sqrt{2})\delta_r M < 3\sqrt{2}\,\delta_r M \end{aligned}$$

by (a) and our choice of δ in (i). Recall that $\eta \equiv \varepsilon/8$ as in (e)–(h). Use of Hoeffding's inequality again allows us to bound the left side of (s) on B_n by

$$|\mathscr{F}(r)|^2 \sup_{\mathscr{F}'((3/2)^{1/2}\delta,\, e_{\mathbb{P}_n})} 2\exp\left(-\frac{\eta^2}{2e_{\mathbb{P}_n}^2(f_r, g_r)}\right)$$
$$\le 2\exp\left(2H_r - \frac{\eta^2}{36\delta_r^2 M^2}\right) \qquad [\text{by (w)}]$$
$$\le 2\exp\left(-\frac{\eta^2}{72\delta_r^2 M^2}\right) \qquad [\text{by (h)}]$$
$$\le \frac{\varepsilon}{16} \qquad [\text{by (g)}].$$

Hence (s) holds and this completes the proof that $\mathscr{F} \in \mathrm{AEC}_u(\mathscr{P}, e_P)$.

We now show that $\mathscr{F} \in \mathrm{AEC}_u(\mathscr{P}, \rho_P)$. Define

$$\mathscr{F}_M \equiv \{ f - c \colon f \in \mathscr{F}, |c| < M \}.$$

Then, under the hypotheses of the theorem, (i) and (ii) are true with $\mathscr{F}$ replaced by $\mathscr{F}_M$ and F replaced by $F + M$. To see this, note that $\mathscr{F}_M^1 \equiv \{c \colon |c| \leq M\}$ satisfies

$$\int_0^1 \left(\log N_M^{(2)}(x, \mathscr{F}_M^1)\right)^{1/2} dx \leq \int_0^1 (1 + \log(1/x))^{1/2} dx < \infty,$$

and then apply Theorem 10 of Pollard (1982). Thus by the preceding argument, $\mathscr{F}_M \in \mathrm{AEC}_u(\mathscr{P}, e_P)$.

Now define

$$\mathscr{F}_P \equiv \{ f - Pf \colon f \in \mathscr{F} \},$$

$$\mathscr{F}_P'(\delta, e_P) \equiv \{ (f, g) \in \mathscr{F}_P \times \mathscr{F}_P \colon e_P(f, g) < \delta \}$$

and

$$\mathscr{F}_M'(\delta, e_P) \equiv \{ (f, g) \in \mathscr{F}_M \times \mathscr{F}_M \colon e_P(f, g) < \delta \}.$$

Since $|P(f)| \leq P(F) \leq (P(F^2))^{1/2} \leq M$ for all $f \in \mathscr{F}$ and $P \in \mathscr{P}$,

$$\mathscr{F}_P \subset \mathscr{F}_M \quad \text{and} \quad \mathscr{F}_P'(\delta, e_P) \subset \mathscr{F}_M'(\delta, e_P).$$

Also,

$$\|\mathbb{X}_n\|_{\mathscr{F}'(\delta, \rho_P)} = \|\mathbb{X}_n\|_{\mathscr{F}_P'(\delta, e_P)} \leq \|\mathbb{X}_n\|_{\mathscr{F}_M'(\delta, e_P)}. \tag{x}$$

Thus for every $\varepsilon > 0$,

$$\begin{aligned} &\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \sup_{P \in \mathscr{P}} \Pr_P^*\{\|\mathbb{X}_n\|_{\mathscr{F}'(\delta, \rho_P)} > \varepsilon\} \\ &\quad \leq \lim_{\delta \downarrow 0} \limsup_{n \to \infty} \sup_{P \in \mathscr{P}} \Pr_P^*\{\|\mathbb{X}_n\|_{\mathscr{F}_M'(\delta, e_P)} > \varepsilon\} = 0, \end{aligned}$$

since $\mathscr{F}_M \in \mathrm{AEC}_u(\mathscr{P}, e_P)$; that is, $\mathscr{F} \in \mathrm{AEC}_u(\mathscr{P}, \rho_P)$. □

6. Proofs for Sections 2 and 5: Completion. In Section 5 we outlined the proof of (B) implies (D) in Theorem 2.1 and discussed its dependence on Theorems 5.1 and 5.2. Now we give proofs of Theorems 5.1 and 5.2 and then use Theorems 5.1 and 5.2 to complete the proof of (B) implies (D) in Theorem 2.1.

PROOF OF THEOREM 5.1. Let $0 < \varepsilon < 1$. Define the truncated variables $X_{in} \equiv X_i 1_{[|X_i| \leq \varepsilon\sqrt{n}]}$ and the corresponding centered variables $X_{in}^c \equiv X_{in} - E(X_{in})$, $i = 1, \dots, n$. If $F_n^{\#}$ is the law of $n^{-1/2}\Sigma_1^n X_{in}$ and F_n^* is the law of $n^{-1/2}S_n^* \equiv n^{-1/2}\Sigma_1^n X_{in}^c$, it is easy to show (using Strassen's theorem) that

$$\pi\left(F_n, F_n^{\#}\right) \leq \varepsilon^{-2} E_P |X|^2 1_{[|X| \geq \varepsilon\sqrt{n}]} \vee \varepsilon \tag{a0}$$

and that

$$\pi\left(F_n^{\#}, F_n^*\right) \leq \varepsilon^{-2} E_P |X|^2 1_{[|X| \geq \varepsilon\sqrt{n}]} \vee \varepsilon. \tag{b0}$$

Furthermore,

$$E_P|X_{in}^c|^3 \le 2^3 E_P|X_{in}|^3 \le 2^3\varepsilon\sqrt{n}\, E_P|X|^2 \le 2^3\varepsilon\sqrt{n}\, M,$$

since (5.1) implies that

$$\sup_{P\in\mathscr{P}} E_P|X|^2 \le M$$

for some absolute constant M. Thus by Yurinskii's (1977), Theorem 1 as corrected in Dehling [(1983), Theorem B, page 395],

$$\text{(c0)}\quad \begin{aligned}\pi\left(F_n^*, F_{N(0,\Sigma_n)}\right) &\le Cd^{1/4}\left(2^3\varepsilon\sqrt{n}\, M\right)^{1/4} n^{-1/8}\left(1+\left|\log\left(\frac{2^3\varepsilon\sqrt{n}\, M}{\sqrt{n}\ d}\right)\right|^{1/2}\right)\\ &= Cd^{1/4}\left(2^3\varepsilon M\right)^{1/4}\left(1+\left|\log\left(\frac{2^3\varepsilon M}{d}\right)\right|^{1/2}\right),\end{aligned}$$

where $\Sigma_n \equiv E(X_{in}^c X_{in}^{cT})$. Finally,

$$E_P|X_i - X_{in}^c|^2 \le 2^2 E_P|X|^2 1_{[|X|\ge\varepsilon\sqrt{n}]}$$

so that

$$\text{(d0)}\qquad \pi\left(F_{N(0,\Sigma_n)}, F_{N(0,\Sigma_P)}\right) \le 2^{2/3}\left(E_P|X|^2 1_{[|X|\ge\varepsilon\sqrt{n}]}\right)^{1/3}$$

by Dehling [(1983), Lemma 2.1, page 402].

Combining (a0)–(d0) yields (5.2), and (5.3) follows from (5.2). Alternatively, (5.3) can be proved using a proof by contradiction from (5.1) and the classical Lindeberg–Feller central limit theorem. □

PROOF OF THEOREM 5.2. Our proof follows Philipp (1980) with corrections as indicated in Philipp (1986). Our present situation is much simpler than Philipp's (excepting the issue of uniformity in P) since $G = N_d(0,\Sigma_P)$ and $a(n) = n^{1/2}$ throughout. The main changes are in the inequality given in (b) below and a somewhat more careful bookkeeping of constants involved.

Let $Z \sim N_d(0,\Sigma_P)$; it follows from (5.1) that

$$\text{(a)}\qquad \sup_{P\in\mathscr{P}} E_P|Z|^2 = \sup_{P\in\mathscr{P}} E_P|X|^2 \le M < \infty$$

for some M. A simple modification of Philipp's argument is to replace his (3.1) by the following moment inequality [which will suffice for the present proof and is simpler than the exponential bound suggested in Philipp (1986)]: for

$r > 0$, $\lambda > 0$,

$$\sup_{P\in\mathscr{P}} \Pr_P(|Z| \geq \lambda) \leq \sup_{P\in\mathscr{P}} \frac{E_P|Z|^r}{\lambda^r}$$

$$\text{(b)} \qquad \leq C_{r,2} \sup_{P\in\mathscr{P}} \frac{\{E_P|Z|^2\}^{r/2}}{\lambda^r}$$

$$\leq C_{r,2} \frac{M^{r/2}}{\lambda^r},$$

where $C_{r,2}$ is an absolute constant; here the first inequality is Markov's, the second follows from Pisier [(1986), Corollary 2.5, page 179] and the third inequality follows from (a).

Let $0 < \varepsilon < 10^{-2} \wedge (2^5 M)^{-1/2}$ be given and set

$$\text{(c)} \qquad t_k \equiv \begin{cases} \left[(1+\varepsilon^4)^k\right], & k \geq 1, \\ 0, & k = 0, \end{cases}$$

$$\text{(d)} \qquad n_k \equiv t_{k+1} - t_k,$$

$$\text{(e)} \qquad s \equiv 4\left[\frac{-\log\varepsilon}{\log(1+\varepsilon^4)}\right] \leq 4\varepsilon^{-5}.$$

Note that since $n_k/t_k \to \varepsilon^4$ as $k \to \infty$,

$$\text{(f)} \qquad \max_{m\leq n_k}\left(\frac{m}{t_k}\right)^{1/2} \leq 2\varepsilon^2 \quad \text{for } k \geq K_0(\varepsilon).$$

LEMMA 6.1. *Given any $\varepsilon > 0$ as above, there exists $K_1(\varepsilon) \geq K_0(\varepsilon)$ so that for $k \geq K_1(\varepsilon)$,*

$$\sup_{P\in\mathscr{P}} \Pr_P\left(\left|\sum_{j=t_k+1}^{t_{k+1}} X_j\right| > \varepsilon\sqrt{t_k}\right) \leq C_1\varepsilon^6$$

and

$$\sup_{P\in\mathscr{P}} \Pr_P\left(\max_{m\leq n_k}\left|\sum_{j=t_k+1}^{t_k+m} X_j\right| > \varepsilon\sqrt{t_k}\right) \leq C_2\varepsilon^6,$$

where $C_1 \equiv 1 + 2^{12}C_{6,2}M^3$ and $C_2 \equiv 2(1 + 2^{18}C_{6,2}M^3)$.

PROOF. We prove the second inequality first. Let $S_n = \Sigma_{j=1}^n X_j$. By stationarity and Ottaviani's inequality,

$$\text{(a1)} \qquad \Pr_P\left(\max_{m\leq n_k}|S_m| > \varepsilon\sqrt{t_k}\right) \leq (1-c_P)^{-1}\Pr_P\left(\left|S_{n_k}\right| > \frac{1}{2}\varepsilon\sqrt{t_k}\right),$$

where

$$
\begin{aligned}
c_P &\equiv \max_{m \le n_k} \Pr_P\left(|S_{n_k} - S_m| > \frac{1}{2}\varepsilon\sqrt{t_k}\right)\\
&= \max_{m \le n_k} \Pr_P\left(|S_m| > \frac{1}{2}\varepsilon\sqrt{t_k}\right)\\
\text{(b1)}\qquad &\le \max_{m \le n_k} \frac{4E_P|S_m|^2}{\varepsilon^2 t_k}\\
&\le \frac{4Mn_k}{\varepsilon^2 t_k}\\
&\le 16M\varepsilon^2 \le \frac{1}{2}d \qquad [\text{for } k \ge K_0(\varepsilon) \text{ by (f)}],
\end{aligned}
$$

uniformly in $P \in \mathscr{P}$ since $\varepsilon^2 \le (2^5M)^{-1}$.

Let F_n be the distribution of $n^{-1/2}S_n$. By Theorem 5.1,

$$
\text{(c1)}\qquad \pi_n \equiv \sup_{P \in \mathscr{P}} \pi\left(F_n, F_{N(0,\Sigma_P)}\right) \to 0 \quad \text{as } n \to \infty,
$$

so we can choose $N_0(\varepsilon)$ so large that

$$
\text{(d1)}\qquad \pi_n \le \varepsilon^6 \quad \text{for } n \ge N_0(\varepsilon).
$$

Then, for $n_k \ge N_0(\varepsilon)$,

$$
\begin{aligned}
\Pr_P\left(\frac{1}{\sqrt{n_k}}|S_{n_k}| \ge \frac{1}{4}\varepsilon^{-1}\right) &\le \Pr_P\left(|Z| \ge \frac{1}{8}\varepsilon^{-1}\right) + \varepsilon^6\\
\text{(e1)}\qquad &\qquad [\text{by the definition of Prohorov distance and (c1), (d1)}]\\
&\le (2^{18}C_{6,2}M^3 + 1)\varepsilon^6\\
&\qquad [\text{uniformly in } P \in \mathscr{P} \text{ by (b) with } r = 6]\\
&\equiv \tilde{C}\varepsilon^6.
\end{aligned}
$$

Hence we have, for $n_k \ge N_0(\varepsilon)$ and $k \ge K_0(\varepsilon)$,

$$
\begin{aligned}
\Pr_P\left(|S_{n_k}| \ge \tfrac{1}{2}\varepsilon\sqrt{t_k}\right) &\le \Pr_P\left(n_k^{-1/2}|S_{n_k}| \ge \tfrac{1}{2}\varepsilon(t_k/n_k)^{1/2}\right)\\
\text{(f1)}\qquad &\le \Pr_P\left(n_k^{-1/2}|S_{n_k}| \ge \tfrac{1}{4}\varepsilon^{-1}\right) \qquad [\text{by (f)}]\\
&\le \tilde{C}\varepsilon^6 \qquad [\text{uniformly in } P \in \mathscr{P} \text{ by (e1)}].
\end{aligned}
$$

Combining (a1), (b1) and (f1) yields the second inequality with $K_1(\varepsilon) \ge K_0(\varepsilon)$ chosen so large that $n_k \ge N_0(\varepsilon)$ for $k \ge K_1(\varepsilon)$.

The proof of the first inequality is similar but easier, using just (c1)–(f1), but starting with ε instead of $\varepsilon/2$ in the first line of (f1). □

Now we return to the proof of Theorem 5.2. Define, for $k \geq 0$, $H_k \equiv (t_k, t_{k+1}]$ and

(g) $$V_k \equiv n_k^{-1/2} \sum_{j \in H_k} X_j.$$

Let $\tilde{F}_k$ be the distribution of V_k, $k = 0, 1, 2, \dots$. By stationarity and Theorem 5.1,

$$\tilde{\pi}_k \equiv \sup_{P \in \mathscr{P}} \pi\left(\tilde{F}_k, F_{N(0, \Sigma_P)}\right) \to 0 \quad \text{as } k \to \infty.$$

Hence by Strassen's (1965) theorem, we can redefine the sequence $\{V_k;\ k \geq 0\}$ on a new probability space on which there exists a sequence $\{Y_k;\ k \geq 0\}$ of independent $N_d(0, \Sigma_P)$ random variables such that

(h) $$\sup_{P \in \mathscr{P}} \Pr_P\left(|V_k - Y_k| > 2\tilde{\pi}_k\right) \leq 2\tilde{\pi}_k.$$

Let $K_2(\varepsilon)$ be so large that $\tilde{\pi}_k \leq \varepsilon^6$ for $k \geq K_2(\varepsilon)$.

Let $\{Z_k;\ k \geq 0\}$ be a sequence of independent $N_d(0, \Sigma_P)$ rv's and note that

$$\mathbf{L}\left(n_k^{-1/2} \sum_{j \in H_k} Z_j\right) = \mathbf{L}(Y_k), \qquad k \geq 0.$$

Let $\tilde{Y}_k \equiv n_k^{-1/2} \Sigma_{j \in H_k} Z_j$. For $k = 0, 1, 2, \dots,$ let R_k^d be a copy of R^d. Then by Dudley [(1984), Lemma 1.2.2] with $X = Y = Z = \prod_{k=0}^{\infty} R_k^d$, $P = \mathbf{L}(\{V_k, Y_k)\}_{k \geq 0})$ and $Q = \mathbf{L}(\{(\tilde{Y}_k, Z_k)\}_{k \geq 0})$, we can take $Y_k = \tilde{Y}_k$, $k \geq 0$. Similarly, we can take V_k as in (g) and by Dudley [(1984), Lemma 1.2.3], we can take the new probability space to be $(\Omega, \Sigma, \Pr_P) = ((R^d)^\infty, (\mathbf{B}^d)^\infty, P^\infty) \times ([0,1], \mathbf{B}, \lambda)$.

Let

$$S_m \equiv \sum_{j \leq m} X_j, \qquad T_m \equiv \sum_{j \leq m} Z_j,$$

and for $k \geq 1$, set

$$S(t_k) \equiv S_{t_k}, \qquad T(t_k) \equiv T_{t_k}.$$

We need one more lemma. For a given integer J, let $J' \equiv J - s$.

LEMMA 6.2. *For* $J \geq J_0 \equiv J_0(\varepsilon)$,

$$\sup_{P \in \mathscr{P}} \Pr_P\left\{\max_{J' < k \leq J} |S(t_k) - T(t_k)| \geq \varepsilon\sqrt{t_J}\right\} \leq \left(2^{-16}C_1 + 8\right)\varepsilon$$

for the constant C_1 *of Lemma* 6.1.

PROOF. First, note that for $J \geq$ some $\tilde{J}_0(\varepsilon)$,

(a2) $$\frac{\Sigma_{J' < k \leq J} n_k^{1/2}}{t_J^{1/2}} \leq \frac{\Sigma_{J' < k \leq J}\left\{(1+\varepsilon^4)^{k+1} - \left((1+\varepsilon^4)^k - 1\right)\right\}^{1/2}}{(1+\varepsilon^4)^{J/2}\left(1 - (1+\varepsilon^4)^{-J}\right)^{1/2}}$$

$$\leq 2\varepsilon^2 s \leq 8\varepsilon^{-3} \qquad [\text{by (e)}],$$

(b2) $$\sqrt{t_{J-1}} > 2^{-1}\sqrt{t_J}$$

and

(c2) $$t_{J'} \le n_J + 1 \le 2n_J \qquad [\text{by (c)–(e)}].$$

Now the probability in question is bounded by

(d2)
$$\begin{aligned}
&\Pr_P\Big\{\max_{J'<k\le J}\big|S(t_k) - S(t_{J'}) - (T(t_k) - T(t_{J'}))\big| \ge \tfrac{1}{2}\varepsilon\sqrt{t_J}\Big\}\\
&\quad + \Pr_P\Big\{\big|S(t_{J'})\big| \ge \tfrac{1}{4}\varepsilon\sqrt{t_J}\Big\} + \Pr_P\Big\{\big|T(t_{J'})\big| \ge \tfrac{1}{4}\varepsilon\sqrt{t_J}\Big\}\\
&= \mathrm{I}_P + \mathrm{II}_P + \mathrm{III}_P.
\end{aligned}$$

Thus in view of (c2), if $J \ge \max\{\tilde{J}_0(\varepsilon), K_1(\varepsilon/8)\}$, by the first inequality of Lemma 6.1 it follows that

(e2) $$\sup_{P\in\mathscr{P}} \mathrm{II}_P \le 2^{-17}C_1\varepsilon^6.$$

Since the same argument applies to sums of the Z's,

(f2) $$\sup_{P\in\mathscr{P}} \mathrm{III}_P \le 2^{-17}C_1\varepsilon^6.$$

Now by definition of V_k and Y_k,

$$S(t_k) - S(t_{J'}) = \sum_{J'<j\le k} \sqrt{n_j}\,V_j$$

and

$$T(t_k) - T(t_{J'}) = \sum_{J'<j\le k} \sqrt{n_j}\,Y_j.$$

Hence, for $J \ge \max\{K_2(\varepsilon) + 4\varepsilon^{-5}, \tilde{J}_0(\varepsilon)\}$, so that $J' = J - s \ge J - 4\varepsilon^{-5} \ge K_2(\varepsilon)$,

(g2)
$$\begin{aligned}
\mathrm{I}_P &= \Pr_P\left\{\max_{J'<k\le J}\left|\sum_{J'<j\le k}\sqrt{n_j}\,(V_j - Y_j)\right| \ge \frac{1}{2}\varepsilon\sqrt{t_J}\right\}\\
&\le \Pr_P\left\{\sum_{J'<j\le J}\sqrt{n_j}\,|V_j - Y_j| \ge \frac{1}{2}\varepsilon\sqrt{t_J}\right\}\\
&\le \Pr_P\left\{\sum_{J'<j\le J}\frac{\sqrt{n_j}}{\sqrt{t_J}}|V_j - Y_j| \ge \frac{1}{2}\varepsilon\frac{\sum_{J'<j\le J}(n_j/t_J)^{1/2}}{8\varepsilon^{-3}}\right\} \qquad [\text{by (a2)}]\\
&\le \sum_{J'<j\le J}\Pr_P\left\{\frac{\sqrt{n_j}}{\sqrt{t_J}}|V_j - Y_j| \ge \frac{1}{16}\varepsilon^4(n_j/t_J)^{1/2}\right\}\\
&\le s\max_{J'<j\le J}\Pr_P\left\{|V_j - Y_j| \ge \frac{1}{16}\varepsilon^4 > 2\varepsilon^6\right\}\\
&\qquad\qquad [\text{using } \varepsilon < 10^{-2} < 2^{-5/2} \text{ as assumed before (c)}]\\
&\le s2\varepsilon^6 \qquad [\text{by (h) and the definition of } K_2(\varepsilon) \text{ just below (h)}]\\
&\le 8\varepsilon \qquad [\text{by (e)}]
\end{aligned}$$

uniformly in $P \in \mathscr{P}$. Combining (e2)–(g2) with (d2) shows that the lemma holds, with $J_0(\varepsilon) \equiv \max\{K_1(\varepsilon/8), K_2(\varepsilon) + 4\varepsilon^{-5}, \tilde{J}_0(\varepsilon)\}$. □

Finally, let n be so large that J defined by $t_{J-1} < n \le t_J$ satisfies

$$J \ge J_1(\varepsilon) \equiv \max\left\{K_1(\varepsilon/8) + 4\varepsilon^{-5}, K_2(\varepsilon) + 4\varepsilon^{-5}, \tilde{J}_0(\varepsilon)\right\}$$

[where $K_1(\varepsilon)$, $K_2(\varepsilon)$, $\tilde{J}_0(\varepsilon)$ are as defined in Lemma 6.1, the first line after (h) and Lemma 6.2, respectively]. Let $J' \equiv J - s$. Then

$$\begin{aligned}
&\Pr_P\left\{\max_{m \le n}|S(m) - T(m)| \ge 10\varepsilon\sqrt{n}\right\}\\
&\quad\le \Pr_P\left\{\max_{m \le t_{J'}}|S(m)| \ge 2\varepsilon\sqrt{n}\right\}\\
&\qquad + \Pr_P\left\{\max_{m \le t_{J'}}|T(m)| \ge 2\varepsilon\sqrt{n}\right\}\\
&\qquad + \Pr_P\left\{\max_{J' < k \le J}|S(t_k) - T(t_k)| \ge 2\varepsilon\sqrt{n}\right\}\\
&\qquad + \Pr_P\left\{\max_{J' < k \le J}\max_{j \le n_k}|S(t_k + j) - S(t_k)| \ge 2\varepsilon\sqrt{n}\right\}\\
&\qquad + \Pr_P\left\{\max_{J' < k \le J}\max_{j \le n_k}|T(t_k + j) - T(t_k)| \ge 2\varepsilon\sqrt{n}\right\}\\
&\quad\equiv \mathrm{I}_P + \mathrm{I}_P + \mathrm{III}_P + \mathrm{IV}_P + \mathrm{V}_P.
\end{aligned}$$

Now $t_{J'} \le n_J + 1$ by (c2) and $\sqrt{n} \ge \sqrt{t_{J-1}} > 2^{-1}\sqrt{t_J}$ by (b2), so Lemma 6.1 yields, since $J > K_1(\varepsilon/8) \ge K_1(\varepsilon/2)$ without loss of generality,

$$\text{(j)} \qquad \sup_P \mathrm{I}_P \le \sup_{P \in \mathscr{P}} \Pr_P\left\{\max_{m \le n_J + 1}|S(m)| \ge \varepsilon\sqrt{t_J}\right\} \le 2^{-5}C_2\varepsilon^6 < C_2\varepsilon.$$

Similarly,

$$\text{(k)} \qquad \sup_P \mathrm{II}_P \le C_2\varepsilon.$$

Now again using $\sqrt{n} > 2^{-1}\sqrt{t_J}$ and (e) in the last step,

$$\text{(l)} \qquad \begin{aligned}
\sup_{P \in \mathscr{P}} \mathrm{IV}_P &\le \sum_{J' < k \le J} \Pr_P\left\{\max_{j \le n_k}|S(t_k + j) - S(t_k)| \ge \varepsilon\sqrt{t_J}\right\}\\
&\le sC_2\varepsilon^6 \le 4C_2\varepsilon \qquad [\text{by Lemma 6.1, since } J' \ge K_1(\varepsilon)]
\end{aligned}$$

and in the same way,

$$\text{(m)} \qquad \sup_{P \in \mathscr{P}} V_P \le 4C_2\varepsilon.$$

The remaining term is

$$\begin{aligned}
\sup_{P\in\mathscr{P}} \mathrm{III}_P &= \sup_{P\in\mathscr{P}} \mathrm{Pr}_P\Big\{\max_{J'<k\le J}|S(t_k) - T(t_k)| \ge 2\varepsilon\sqrt{n}\Big\}\\
\text{(n)}\qquad &\le \sup_{P\in\mathscr{P}} \mathrm{Pr}_P\Big\{\max_{J'<k\le J}|S(t_k) - T(t_k)| \ge \varepsilon\sqrt{t_J}\Big\}\\
&\le (2^{-16}C_1 + 8)\varepsilon \qquad [\text{by Lemma 6.2, since } J \ge J_1(\varepsilon) > J_0(\varepsilon)].
\end{aligned}$$

Combining (j)–(n) with (i) yields

$$\begin{aligned}
&\sup_{P\in\mathscr{P}} \mathrm{Pr}_P\Big\{\max_{m\le n}|S(m) - T(m)| \ge 10\varepsilon\sqrt{n}\Big\}\\
\text{(o)}\qquad &\le 2C_2\varepsilon + 8C_2\varepsilon + (2^{-16}C_1 + 8)\varepsilon\\
&= (10C_2 + 2^{-16}C_1 + 8)\varepsilon \equiv C_3\varepsilon.
\end{aligned}$$

If we start with $\tilde{\varepsilon} > 0$, then

$$\begin{aligned}
&\sup_{P\in\mathscr{P}} \mathrm{Pr}_P\Big\{\max_{m\le n}|S(m) - T(m)| \ge \tilde{\varepsilon}\sqrt{n}\Big\}\\
\text{(p)}\qquad &\le \sup_{P\in\mathscr{P}} \mathrm{Pr}_P\left\{\max_{m\le n}|S(m) - T(m)| \ge \frac{10\tilde{\varepsilon}}{C_3}\sqrt{n}\right\}\\
&\qquad\qquad [\text{since } 1 > 10/C_3]\\
&\le \tilde{\varepsilon} \qquad [\text{by (o) with } \varepsilon = \tilde{\varepsilon}/C_3],
\end{aligned}$$

and this almost gives (5.4).

The remaining problem is that the constructed sequence still depends on ε. We therefore universalize as in Philipp (1986) and Major (1976) [in fact, our argument below essentially follows Dudley (1984), pages 29–31; a similar argument is used in the proof of Theorem 2.1] to get rid of the dependence on ε as follows: For given $k \ge 1$, choose sequences $\{X_{kj};\ j \ge 1\}$ and $\{Y_{kj};\ j \ge 1\}$ of independent random variables with partial sums S_{km} and T_{km}, respectively, such that for all $n \ge n_k \equiv n(k)$,

$$\text{(q)}\qquad \sup_{P\in\mathscr{P}} \mathrm{Pr}_P\Big\{n^{-1/2}\max_{m\le n}|S_{km} - T_{km}| \ge 2^{-k}\Big\} \le 2^{-k}.$$

We may suppose without loss, that the X sequences are independent of one another for different k and that the Y sequences are also independent of one another for different k. We may also take $1 \equiv n_0 < n_1 < \cdots$.

For each $j = 1, 2, \dots,$ if $n_k \le j < n_{k+1}$, write $k = k(j)$ and set

$$\text{(r)}\qquad X_j \equiv X_{k(j)j}, \qquad Z_j \equiv Z_{k(j)j} \quad \text{if } n_k \le j < n_{k+1}.$$

Then $\{X_j;\ j \ge 1\}$ and $\{Z_j;\ j \ge 1\}$ are sequences of independent random variables with common distributions P and $F_{N(0,\Sigma_P)}$ on R^d, respectively. By another application of Dudley's (1984) Lemmas 1.2.2 and 1.2.3, we can arrange everything to be defined on the probability space $(\Omega, \Sigma, \mathrm{Pr}_P)$.

Now we show that (5.4) holds for this (universal) choice of the X and Z sequences. Let $\varepsilon > 0$ be given and let $k \equiv k(\varepsilon)$ satisfy $2^{3-k} \le \varepsilon$. Let $M_k \ge n_k$

be so large that

$$\text{(s)}\qquad \sup_{P\in\mathscr{P}} \Pr_P\left\{\max_{m\le n(k)}(|S_m|+|T_m|) > 2^{1-k}\sqrt{n}\right\} \le 2^{1-k} \quad \text{for } n \ge M_k.$$

Fix $n \ge M_k$ and choose r so that $n_r \le n < n_{r+1}$. Then

$$\text{(t)}\qquad \begin{aligned} \max_{m\le n}|S_m - T_m| &\le \max_{m\le n_k}|S_m - T_m| + \sum_{i=k}^{r-1} \max_{n_i \le m < n_{i+1}} \left|\sum_{j=n_i}^{m}(X_j - Z_j)\right| \\ &\quad + \max_{n_r\le m\le n}\left|\sum_{j=n_r}^{m}(X_j - Z_j)\right| \\ &\equiv L_k + \sum_{i=k}^{r-1} A_i + A_r. \end{aligned}$$

Now $L_k \le \max_{m\le n_k}(|S_m| + |T_m|)$, so

$$\text{(u)}\qquad \sup_{P\in\mathscr{P}} \Pr_P\left\{L_k \ge 2^{1-k}\sqrt{n}\right\} \le 2^{1-k}$$

by (s), since $n \ge M_k$. Furthermore,

$$\text{(v)}\qquad \begin{aligned} &\sup_{P\in\mathscr{P}} \Pr_P\left\{A_i \ge 2^{1-i}\sqrt{n}\right\} \\ &\quad = \sup_{P\in\mathscr{P}} \Pr_P\left\{\max_{n_i\le m<n_{i+1}}\left|\sum_{j=n_i}^{m}(X_j - Z_j)\right| \ge 2^{1-i}\sqrt{n}\right\} \\ &\quad = \sup_{P\in\mathscr{P}} \Pr_P\left\{\max_{n_i\le m<n_{i+1}}\left|\sum_{j=n_i}^{m}(X_{ij} - Z_{ij})\right| \ge 2^{1-i}\sqrt{n}\right\} \\ &\quad \le 2^{1-i} \qquad [\text{by (q) since } n \ge n_r \ge n_{i+1} \text{ for } i < r] \end{aligned}$$

[and by writing $\sum_{j=n_i}^{m}(X_{ij} - Z_{ij}) = (S_{im} - T_{im}) - (S_{i,n_i} - 1 - T_{i,n_i} - 1]$, and

$$\text{(w)}\qquad \begin{aligned} &\sup_{P\in\mathscr{P}} \Pr_P\left\{A_r \ge 2^{1-r}\sqrt{n}\right\} \\ &\quad = \sup_{P\in\mathscr{P}} \Pr_P\left\{\max_{n_r\le m\le n}\left|\sum_{j=n_r}^{m}(X_j - Z_j)\right| \ge 2^{1-r}\sqrt{n}\right\} \\ &\quad = \sup_{P\in\mathscr{P}} \Pr_P\left\{\max_{n_r\le m\le n}\left|\sum_{j=n_r}^{m}(X_{rj} - Z_{rj})\right| \ge 2^{1-r}\sqrt{n}\right\} \\ &\quad \le 2^{1-r} \qquad [\text{by (q) since } n \ge n_r]. \end{aligned}$$

Thus, by (t)–(w), for $n \ge M_k = M_{k(\varepsilon)}$,

$$\begin{aligned}
\sup_{P \in \mathscr{P}} \Pr_P\Big\{\max_{m \le n} |S_m - T_m| > \varepsilon\sqrt{n}\Big\}
&\le \sup_{P \in \mathscr{P}} \Pr_P\big\{L_k \ge 2^{1-k}\sqrt{n}\big\} \\
&\quad + \sum_{i=k}^{r-1} \sup_{P \in \mathscr{P}} \Pr_P\big\{A_i \ge 2^{1-i}\sqrt{n}\big\} + \sup_{P \in \mathscr{P}} \Pr_P\big\{A_r \ge 2^{1-r}\sqrt{n}\big\} \\
&\le 2^{1-k} + \sum_{i=k}^{r} 2^{1-i} \\
&< 2^{1-k}(1+2) < 2^{3-k} \le \varepsilon. \qquad \square
\end{aligned}$$

Now we have the tools to complete the proof of Theorem 2.1.

PROOF OF (B) IMPLIES (D) IN THEOREM 2.1. The following proof is essentially the same as Dudley's (1984) proof of his Theorem 4.1.1 with care to keep everything uniform in $P \in \mathscr{P}$ and with only minor corrections/changes; there is a $\sqrt{n}$ missing in Dudley's (4.1.9).

Suppose that (B) of Theorem 2.1 holds. For $k = 1, 2, \ldots,$ take $\varepsilon = 2^{-k}$ in (2.6) of Definition 2.2 with $d = \rho_P$; we thereby obtain $\delta = \delta_k > 0$ and $N \equiv N_k$, so that (2.6) holds with $d = \rho_P$, $\varepsilon = 2^{-k}$ and $\delta = \delta_k$ for $n \ge N_k$. For $k = 1, 2, \ldots,$ let $\mathscr{F}_{k,P}$ be a finite subset of $\mathscr{F}$ such that

$$\sup_{f \in \mathscr{F}} \inf\{\rho_P(f, g): g \in \mathscr{F}_{k,P}\} < \delta_k.$$

Let $T_{k,P}$ denote the finite-dimensional space of all real functions on $\mathscr{F}_{k,P}$ also with supremum norm $\|\cdot\|_k = \|\cdot\|_{k,P}$. Let $\mathscr{F}_{k,P} = \{g_{1,P}, g_{2,P}, \ldots, g_{m(k,P),P}\}$. By the Sudakov minorization (f) at the end of the proof of Theorem 2.1 in Section 2, we know that $m(k,P) \le m(k) < \infty$ for all $P \in \mathscr{P}$. For each $f \in \mathscr{F}$ and $P \in \mathscr{P}$, let $f_{k,P} = g_{j,P}$ for the least j such that $\rho_P(f, g_{j,P}) < \delta_k$. For any $\phi \in l^\infty(\mathscr{F})$, let $\phi_{k,P}(f) = \phi(f_{k,P})$, $f \in \mathscr{F}$. Then $\phi_{k,P} \in l^\infty(\mathscr{F})$. Write $\mathbb{E}_{j,P} = \delta_{X_j} - P \in l^\infty(\mathscr{F})$, $j \ge 1$. Let $\Lambda_{k,P}(\phi) = \phi_{k,P}$ and $\mathbb{E}_{kj,P} = \mathbb{E}_{j,P} - \Lambda_{k,P}\mathbb{E}_{j,P}$.

The union of the finite-dimensional ranges of the $\Lambda_{k,P}$, $k = 1, 2, \ldots,$ is included in a complete separable subspace T_P of $l^\infty(\mathscr{F})$ with $C(\mathscr{F}, \rho_P) \subset T_P$ for each $P \in \mathscr{P}$. Note that $\|\phi_{k,P}\| \equiv \|\phi_{k,P}\|_{\mathscr{F}} \le \|\phi\|_{\mathscr{F}} \equiv \|\phi\|$ for all k, P and all $\phi \in l^\infty(\mathscr{F})$. Then by (2.6) with the choices made above we have

$$\text{(a3)} \qquad \sup_{P \in \mathscr{P}} \Pr_P\left\{ n^{-1/2} \left\| \sum_{j=1}^{n} \mathbb{E}_{kj,P} \right\| > 2^{-k} \right\} \le 2^{-k}, \qquad n \ge N_k.$$

By the same argument as in Dudley [(1984), page 29] (which involves the Ottaviani–Skorohod inequality), it then follows that

$$\text{(b3)} \qquad \sup_{P \in \mathscr{P}} \Pr_P\left\{ n^{-1/2} \max_{m \le n} \left\| \sum_{j=1}^{m} \mathbb{E}_{kj,P} \right\| > 2^{2-k} \right\} \le 2^{2-k}, \qquad n \ge 2N_k.$$

Let P_k be the law on $T_{k,P}$ of $f \to f(X_1) - Pf$, $f \in \mathscr{F}_{k,P}$. Then by Theorem 5.2, there exist random variables $V_{kj,P}$ iid P_k, $W_{kj,P}$ iid with a Gaussian law $Q_{k,P}$, all defined on some common probability space and some $n_k \equiv n_k \geq 2N_k$, $k \geq 1$, with

$$\text{(c3)} \qquad \sup_{P \in \mathscr{P}} \Pr_P \left\{ n^{-1/2} \max_{m \leq n} \left\| \sum_{j=1}^{m} V_{kj,P} - W_{kj,P} \right\|_k > 2^{-k} \right\} \leq 2^{-k}$$

for all $n \geq n_k$. Also, $Q_{k,P}$ must be the law of the restriction of $\mathbb{X} = G_P$ to $\mathscr{F}_{k,P}$. We assume that the sequences $\{(V_{kj,P}, W_{kj,P})\}_{j \geq 1}$ are independent of each other for different k. We also take $1 \equiv n_0 < n_1 < \cdots$.

For each $j = 1, 2, \ldots$, if $n_k \leq j < n_{k+1}$, write $k = k(j)$ and set $V_{j,P} = V_{k(j)j,P}$, $W_{j,P} \equiv W_{k(j)j,P}$ and $\mathscr{F}_{(j),P} \equiv \mathscr{F}_{k(j)j,P}$. Then $\{V_{j,P}\}_{j \geq 1}$ and $\{W_{j,P}\}_{j \geq 1}$ are sequences of independent random variables, each with values in a countable product of Polish spaces which itself is Polish. Let $\{\mathbb{Z}_{j,P}\}_{j \geq 1}$ be iid copies of $\mathbb{X} = G_P$ on $U_P \equiv C(\mathscr{F}, \rho_P)$. Then W_j has the law of the restriction $\mathbb{Z}_{j,P}|_{\mathscr{F}_{k,P}}$, $n_k \leq j < n_{k+1}$. Let $T_{kj,P}$ be a copy of $T_{k,P}$ and $U_{(j),P}$ a copy of U_P for each j (these are spaces—not variables!) and set $T_{(j),P} \equiv T_{k(j)j,P}$. Then by Dudley [(1984), Lemma 1.2.2], with

$$X = Y = \prod_{j=1}^{\infty} T_{(j),P}, \qquad Z = \prod_{j=1}^{\infty} U_{(j),P},$$

$$P = \mathbf{L}\Big(\big\{(V_{j,P}, W_{j,P})\big\}_{j \geq 1}\Big)$$

and

$$Q = \mathbf{L}\Big(\Big\{\big(\mathbb{Z}_{j,P}|_{\mathscr{F}_{(j),P}}, \mathbb{Z}_{j,P}\big)\Big\}_{j \geq 1}\Big),$$

we can take $W_{j,P} = \mathbb{Z}_{j,P}|_{\mathscr{F}_{(j),P}}$.

Then, for $\omega = (\{x_j\}_{j \geq 1}, t) \in A^\infty \times [0,1] \equiv \Omega$, let $S_1 \equiv X$ as above and

$$V_P(\omega) \equiv \left\{ f \to f(x_j) - \int f \, dP \colon f \in \mathscr{F}_{(j),P} \right\}_{j \geq 1} \in S_1.$$

Let $T_1 \equiv Y \times Z$ and

$$Q = \mathbf{L}\Big(\big\{(V_{j,P}, W_{j,P}, \mathbb{Z}_{j,P})\big\}_{j \geq 1}\Big) \quad \text{on } S_1 \times T_1.$$

By Dudley [(1984), Lemma 1.2.3], the $V_{j,P}$ and $\mathbb{Z}_{j,P}$ can be taken as defined on Ω with $(V_P(\omega))_j = V_{j,P}$ for all j.

Now let $\varepsilon > 0$ and choose k so large that $2^{6-k} < \varepsilon$.

Let $M_k \geq n_k$ be large enough so that for all $n \geq M_k$,

$$\text{(d3)} \qquad \sup_{P \in \mathscr{P}} \Pr_P^* \left\{ n^{-1/2} \sum_{j=1}^{n_k} \|\Lambda_{k,P} \mathbb{E}_{j,P}\| + \|\mathbb{Z}_{j,P}\| > 2^{-k} \right\} < 2^{-k}.$$

Fix $n \geq M_k$ and choose r so that $n_r \leq n < n_{r+1}$. Then

$$\begin{aligned} D_n &\equiv \max_{m \leq n} \left\| \sum_{j=1}^{m} \mathbb{E}_{j,P} - \mathbb{Z}_{j,P} \right\| \\ &\leq \max_{m \leq n_k} \left\| \sum_{j=1}^{m} \mathbb{E}_{j,P} - \mathbb{Z}_{j,P} \right\| \\ &\quad + \sum_{i=k}^{r-1} \max_{n_i \leq m < n_{i+1}} \left\| \sum_{j=n_i}^{m} \mathbb{E}_{j,P} - \mathbb{Z}_{j,P} \right\| \qquad \text{(e3)} \\ &\quad + \max_{n_r \leq m \leq n} \left\| \sum_{j=n_r}^{m} \mathbb{E}_{j,P} - \mathbb{Z}_{j,P} \right\| \\ &\equiv L_k + \sum_{i=k}^{r-1} A_i + A_r. \end{aligned}$$

Then

$$L_k \leq \max_{m \leq n_k} \left\{ \left\| \sum_{j \leq m} \mathbb{E}_{kj,P} \right\|^* + \sum_{j \leq m} \|\Lambda_{k,P}\mathbb{E}_{j,P}\| + \|\mathbb{Z}_{j,P}\| \right\},$$

so that, by (b3) and (d3),

$$\text{(f3)} \qquad \sup_{P \in \mathscr{P}} \Pr_P\left\{ L_k^* \geq 2^{3-k}\sqrt{n} \right\} \leq 2^{3-k}.$$

[The $\sqrt{n}$ is missing in Dudley's (4.1.9); it should be inserted there, too!] Since restriction to $\mathscr{F}_{i,P}$ is a linear isometry from $\Lambda_{i,P}(l^\infty(\mathscr{F}))$ to $(T_{i,P}, \|\cdot\|_i)$, for $k \leq i \leq r$ we have

$$\begin{aligned} A_i &= \max_{n_i \leq m < n_{i+1}} \left\| \sum_{j=n_i}^{m} \mathbb{E}_{j,P} - \mathbb{Z}_{j,P} \right\| \\ &\leq \max_{n_i \leq m < n_{i+1}} \left\{ \left\| \sum_{j=n_i}^{m} \mathbb{E}_{ij,P} \right\|^* + \left\| \sum_{j=n_i}^{m} V_{j,P} - W_{j,P} \right\|_i \right. \qquad \text{(g3)} \\ &\qquad\qquad \left. + \left\| \sum_{j=n_i}^{m} \Lambda_i \mathbb{Z}_{j,P} - \mathbb{Z}_{j,\mathbb{P}} \right\| \right\} \\ &\equiv \mathrm{I} + \mathrm{II} + \mathrm{III}. \end{aligned}$$

But since $n \geq n_r \geq 2N_i$, it follows from (b3) and stationarity that

$$\text{(h3)} \quad \sup_{P \in \mathscr{P}} \Pr_P\left\{ \max_{n_i \leq m \leq n_{i+1}} \left\| \sum_{j=n_i}^{m} \mathbb{E}_{ij,P} \right\|^* > 2^{2-i}\sqrt{n} \right\} \leq 2^{2-i}, \qquad k \leq i < r.$$

Similarly, by (c3),

$$\text{(i3)} \quad \sup_{P} \Pr_P\left\{ \max_{n_i \leq m \leq n_{i+1}} \left\| \sum_{n_i}^{m} V_{j,P} - W_{j,P} \right\|_i > 2^{-i}\sqrt{n} \right\} < 2^{-i}, \qquad k \leq i < r.$$

Finally,

$$\mathbf{L}_P\left((n_{i+1}-n_i)^{-1/2}\sum_{j=n_i}^{n_{i+1}-1}\mathbb{Z}_{j,P}\right)=\mathbf{L}(\mathbb{Z}_{1,P})=\mathbf{L}(G_P)$$

with ρ_P-uniformly continuous sample functions uniformly in $P\in\mathscr{P}$. Since $n\ge n_{i+1}-n_i$ for $i<r$, it follows from Lévy's inequality, (2.2) and the choices made above, that

$$\text{(j3)}\qquad \sup_{P\in\mathscr{P}}\Pr_P\left\{\max_{n_i\le m<n_{i+1}}\left\|\sum_{j=n_i}^{m}\mathbb{Z}_{j,P}-\Lambda_i\mathbb{Z}_{j,P}\right\|>2^{2-i}\sqrt{n}\right\}$$

$$\le 2\sup_{P\in\mathscr{P}}\Pr_P\left\{n^{-1/2}\left\|\sum_{j=n_i}^{n_{i+1}-1}\mathbb{Z}_{j,P}-\Lambda_i\mathbb{Z}_{j,P}\right\|>2^{2-i}\right\}$$

$$\le 2\sup_{P\in\mathscr{P}}\Pr_P\{\|\mathbb{Z}_{1,P}-\Lambda_i\mathbb{Z}_{1,P}\|>2^{2-i}\}$$

$$\le 2^{3-i}\qquad[\text{by }(2.2)].$$

Collecting terms, (g3)–(j3) imply that

$$\text{(k3)}\qquad \sup_{P\in\mathscr{P}}\Pr_P\{A_i>2^{4-i}\sqrt{n}\}\le 2^{4-i},\qquad k\le i<r.$$

Similarly, for $i=r$, replacement of n_{i+1} by n throughout yields the same bound. Thus, by (e3), (f3), (k3) and $2^{6-k}<\varepsilon$,

$$\sup_{P\in\mathscr{P}}\Pr_P\left\{\max_{m\le n}\left\|\sum_{j=1}^{m}\mathbb{E}_{j,P}-\mathbb{Z}_{j,P}\right\|^*\ge\sqrt{n}\,\varepsilon\right\}$$

$$\le 2^{3-k}+\sum_{i=k}^{r}2^{4-i}$$

$$\le 2^{3-k}(1+4)<2^{6-k}<\varepsilon.\qquad\square$$

Acknowledgments. We owe thanks to Richard M. Dudley, Peter Gaenssler, Richard Gill and Evarist Giné for conversations, interest and support at various times during the writing of this paper. In particular, Evarist Giné provided the present proof of Lemma 5.1. We also owe thanks to a referee for many constructive comments and suggestions.

REFERENCES

ANDERSEN, N. T. (1985). The central limit theorem for nonseparable valued functions. *Z. Wahrsch. Verw. Gebiete* **70** 445–455.

ANDERSEN, N. T. and DOBRIĆ, V. (1987). The central limit theorem for stochastic processes. *Ann. Probab.* **15** 164–177.

ARAUJO, A. and GINÉ, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables*. Wiley, New York.

BORELL, C. (1975). The Brunn–Minkowski inequality in Gauss space. *Invent. Math.* **30** 205–216.

CHUNG, K. L. (1951). The strong law of large numbers. *Proc. Second Berkeley Symp. Math. Statist. Probab.* 341–352. Univ. California Press, Berkeley.

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

DUDLEY, R. M. (1984). A course on empirical processes. *Ecole d'Eté de Probabilités de St. Flour. Lecture Notes in Math.* **1097** 2–142. Springer, New York.

DUDLEY, R. M. (1985). An extended Wichura theorem, definitions of Donsker class, and weighted empirical distributions. *Probability in Banach Spaces V. Lecture Notes in Math.* **1153** 141–178. Springer, New York.

DUDLEY, R. M. (1987). Universal Donsker classes and metric entropy. *Ann. Probab.* **15** 1306–1326.

DUDLEY, R. M. (1989). *Real Analysis and Probability*. Wadsworth and Brooks/Cole, Pacific Grove.

DUDLEY, R. M. (1990). Nonlinear functionals of empirical measures and the bootstrap. In *Probability in Banach Spaces. Progress in Probability* **21** 63–82. Birkhäuser, Boston.

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.

EFRON, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. SIAM, Philadelphia.

FERNIQUE, X. (1985). Sur la convergence étroite des mesures gaussiennes. *Z. Wahrsch. Verw. Gebiete* **68** 331–336.

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

GINÉ, E. and ZINN, J. (1986). Lectures on the central limit theorem for empirical processes. *Lecture Notes in Math.* **1221** 50–113. Springer, Berlin.

GINÉ, E. and ZINN, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.* **18** 851–869.

GINÉ, E. and ZINN, J. (1991). Gaussian characterization of uniform Donsker classes of functions. *Ann. Probab.* **19** 758–782.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

HOFFMANN-JØRGENSEN, J. (1977). Probability in Banach spaces. *Ecole d'Eté de Probabilités de Saint-Flour VI. Lecture Notes in Math.* **598** 1–186. Springer, New York.

HOFFMANN-JØRGENSEN, J. (1984). Stochastic processes on Polish spaces. Unpublished manuscript.

JAIN, N. and MARCUS, M. B. (1978). Continuity of subgaussian processes. In *Probability in Banach Spaces. Advances in Probability* **4** 81–196. Birkhäuser, Boston.

LAI, T. L. (1978). Pitman efficiencies of sequential tests and uniform limit theorems in nonparametric statistics. *Ann. Statist.* **6** 1027–1047.

LE CAM, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Statist.* **41** 802–828.

LEDOUX, M. and TALAGRAND, M. (1986). Conditions d'intégrabilité pour les multiplicateurs dans le TLC Banachique. *Ann. Probab.* **14** 916–921.

LEDOUX, M. and TALAGRAND, M. (1988). Un critere sur les petites boules dan le théorème limite central. *Probab. Theory Related Fields* **77** 29–47.

MAJOR, P. (1976). Approximation of partial sums of i.i.d. r.v.'s when the summands have only two moments. *Z. Wahrsch. Verw. Gebiete* **35** 221–229.

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

OSSIANDER, M. (1987). A central limit theorem under metric entropy with $L-2$ bracketing. *Ann. Probab.* **15** 897–919.

PHILIPP, W. (1980). Weak and L^p-invariance principles for sums of B-valued random variables. *Ann. Probab.* **8** 68–82.

PHILIPP, W. (1986). Correction to "Weak and L^p-invariance principles for sums of B-valued random variables." *Ann. Probab.* **14** 1095–1101.

PISIER, G. (1975). Le théorème de la limite centrale et la loi du logarithme itérée dans les espaces de Banach. *Séminaire Maurey–Schwartz 1975–1976*. Exposé IV, Ecole Polytechnique, Paris.

PISIER, G. (1986). Probabilistic methods in the geometry of Banach spaces. *Lecture Notes in Math.* **1206** 167–241. Springer, Berlin.

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.

POLLARD, D. (1989). *Empirical Processes: Theory and Applications*. SIAM/IMS, Philadelphia.

SHEEHY, A. and WELLNER, J. A. (1988). Uniformity in P of some limit theorems for empirical measures and processes. Technical Report 134, Dept. of Statist., Univ. Washington, Seattle.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.

SUDAKOV, V. N. (1971). Gaussian random processes and measures of solid angles in Hilbert spaces. *Dokl. Akad. Nauk SSSR* **197** 43–45.

VAN DER VAART, A. W. (1988). Statistical estimation in large parameter spaces. *CWI Tract* **44**. Centrum voor Wiskunde en Informatica, Amsterdam.

VAN DER VAART, A. W. and WELLNER, J. A. (1990). Prohorov and continuous mappings theorems in the Hoffmann-Jøgensen weak convergence theory with applications to convolution and asymptotic minimax theorems. Technical Report 157, Dept. Statist., Univ. Washington, Seattle.

YURINSKII, V. V. (1977). On the error of the Gaussian approximation for convolutions. *Theory Probab. Appl.* **22** 236–247.

SEMINAR FÜR STATISTIK
SONEGGSTR. 33.
ETH
8092 ZÜRICH
SWITZERLAND

DEPARTMENT OF STATISTICS, GN-22
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195