The Annals of Probability
1987, Vol. 15, No. 1, 178–203

THE CENTRAL LIMIT THEOREM FOR EMPIRICAL PROCESSES ON VAPNIK–ČERVONENKIS CLASSES[1]

BY KENNETH S. ALEXANDER[2]

University of Washington

Sufficient conditions, which under weak additional assumptions are also necessary, are found for the central limit theorem for empirical processes indexed by classes of functions which have the Vapnik–Červonenkis property. This improves an earlier theorem of Pollard (1982a), and leads to necessary and sufficient conditions for the CLT for weighted empirical processes indexed by Vapnik–Červonenkis classes of sets.

1. Introduction. Let $X, X_1, X_2, \ldots$ be i.i.d. random variables defined on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ and taking values in a measure space $(T, \mathscr{B})$, and let $P := \mathscr{L}(X)$. Define the nth empirical measure

$$P_n := n^{-1} \sum_{i \le n} \delta_{X_i}$$

and the nth normalized empirical process

$$\nu_n := n^{1/2}(P_n - P).$$

Given a function f and signed measure Q on $(T, \mathscr{B})$, we write $Q(f)$ for $\int f\,dQ$; in this way, given a class $\mathscr{F}$ on functions on T, ν_n may be viewed as a stochastic process indexed by $\mathscr{F}$:

$$\nu_n(f) = n^{-1/2} \sum_{i \le n} (f(X_i) - Ef(X_i)).$$

More precisely, when the *envelope* F of $\mathscr{F}$, defined by

$$F(t) := \sup_{f \in \mathscr{F}} |f(t)|,$$

is finite P-a.e., ν_n may be viewed as an element of the space $l^\infty(\mathscr{F})$ of all bounded functions on $\mathscr{F}$, endowed with the sup norm $\|\cdot\|_{\mathscr{F}}$. ν_n may also be viewed as indexed by some class $\mathscr{C}$ of subsets of T, in which case we identify these sets with their indicator functions. Taking $\mathscr{C}$ to be $\{[0, t]: t \in [0,1]\}$ gives the classical case of empirical distribution functions.

Let e_P be the $L^2(P)$ pseudometric on $\mathscr{F}$, and ρ_P the centered $L^2(P)$ pseudometric:

$$\rho_P(f, g) := e_P(f - P(f), g - P(g)).$$

We say (following Dudley (1978)) that $\mathscr{F}$ is a *P-Donsker class* if ν_n converges

Received January 1985; revised August 1985.

[1]Research supported by an NSF Postdoctoral Fellowship, Grant No. MCS 83-11686.

[2]Author now at University of Southern California.

AMS 1980 *subject classifications*. 60F05, 60F17, 60B12.

Key words and phrases. Empirical process, central limit theorem, Vapnik–Červonenkis class, metric entropy, weighted empirical process.

weakly in $l^\infty(\mathscr{F})$ to a Gaussian process, call it G_P, indexed by $\mathscr{F}$ which has bounded, ρ_P-uniformly continuous sample paths. G_P necessarily has mean 0 and the same covariance as ν_n:

$$\operatorname{cov}(G_P(f), G_P(g)) = P(fg) - P(f)P(g).$$

When such a process G_P exists, we say $\mathscr{F}$ is *P-pregaussian* (Dudley (1978) called such $\mathscr{F}$ "$G_P BUC$").

Considerable efforts have gone into finding conditions on $\mathscr{F}$ and P under which $\mathscr{F}$ is P-Donsker. Some of the best successes involve classes $\mathscr{F}$ which have the Vapnik–Červonenkis (or VC) property (defined below—it is a combinatorial condition on the regions determined in $T \times \mathbb{R}$ by the graphs of functions in $\mathscr{F}$). This is a natural property in our context because Dudley (1978) and Durst and Dudley (1981) showed that, under some measurability assumptions, a class of sets is P-Donsker for all P if and only if it is a VC class. Pollard (1982a, 1984) showed that a class of functions with the VC property is P-Donsker provided the envelope satisfies $P(F^2) < \infty$.

The latter condition is not necessary, however: For a nonnegative increasing function q on [0, 1] we can consider

$$\mathscr{F} = \{1_{[0,t]}/q(t) \colon t \in [0,1]\} \tag{1.1}$$

and P uniform; ν_n then becomes a weighted empirical d.f., and by a theorem of Cibisov (1964) and O'Reilly (1974), $\mathscr{F}$ is P-Donsker if and only if

$$\int_0^1 \exp(-\varepsilon q^2(t)/t)\, dt/t < \infty \quad \text{for all } \varepsilon > 0, \tag{1.2}$$

which is weaker than the requirement that $F = 1/q \in L^2(P)$. Therefore we would like to improve Pollard's theorem if possible.

Indeed, we will prove (under some measurability hypotheses) that $\mathscr{F}$ is P-Donsker provided

$$\mathscr{F} \text{ is } P\text{-pregaussian} \tag{1.3}$$

and

$$u^2 P[F > u] \to 0 \quad \text{as } u \to \infty. \tag{1.4}$$

Of course (1.3) is necessary, and we will see that (1.4) is necessary if $|P(f)|$ is bounded over $\mathscr{F}$. It is clear that $\mathscr{F}$ is P-Donsker if and only if

$$\mathscr{F}_P := \{f - P(f) \colon f \in \mathscr{F}\}$$

is P-Donsker; therefore it is always necessary for $\mathscr{F}$ to be P-Donsker that the envelope

$$F_P(t) := \sup_{f \in \mathscr{F}} |f(t) - P(f)|$$

of $\mathscr{F}_P$ satisfy

$$u^2 P[F_P > u] \to 0 \quad \text{as } u \to \infty. \tag{1.5}$$

Thus (1.3) and (1.5) are necessary and sufficient for $\mathscr{F}$ to be P-Donsker whenever either $\mathscr{F}_P$ has the VC property or $\sup_{f \in \mathscr{F}} |P(f)| < \infty$.

Given a VC class $\mathscr{C}$ of sets and a nonnegative function q on $\mathscr{C}$, we can, motivated by (1.1), define the weighted empirical process $\nu_n(C)/q(C)$, $C \in \mathscr{C}$, which is equivalent to taking

$$\mathscr{F} := \{1_C/q(C) \colon C \in \mathscr{C}\}.$$

If $q(C)$ is small whenever $P(C)$ is small, this is a way of expanding the normalized deviation ν_n of P_n from P on small sets C, which is analogous to weighting the tails when considering the Kolmogorov–Smirnov distance between d.f.'s. We will use our main theorem to obtain necessary and sufficient conditions for ν_n/q to converge weakly to G_P/q, where G_P is the weak limit of ν_n on $\mathscr{C}$. (Throughout this paper, ratios which evaluate as $0/0$—such as $G_P(C)/q(C)$ or $\nu(C)/q(C)$ when $P(C) = 0$—should be interpreted as being 0. This is justified by Lemma 3.2 below.)

In two special cases we will obtain more readily checkable versions of these conditions. These are, first, when $q(C)$ depends only on $P(C)$, and second, in the multidimensional d.f. case ($\mathscr{C} = \{[0, t]\colon t \in [0,1]^d\}$, P uniform).

Thanks to a result of Dudley (1984, 1985) and Dudley and Philipp (1983), the same conditions implying $\mathscr{F}$ is P-Donsker also imply an invariance principle in probability for ν_n on $\mathscr{F}$. Therefore the CLT we obtain is a functional one.

2. The main results. Given a class $\mathscr{C}$ of subsets of a set T, and a finite subset E of T, we say $\mathscr{C}$ *shatters* E if every $D \subset E$ is of the form $E \cap C$ for some $C \in \mathscr{C}$. $\mathscr{C}$ is called a *Vapnik–Červonenkis* (or VC) *class* if for some $n \geq 1$, no n-element subset of T is shattered by $\mathscr{C}$. The least such n is called the *index* of $\mathscr{C}$ and denoted $V(\mathscr{C})$. Examples in $\mathbb{R}^d$ include the classes of all rectangles, all ellipsoids, and all polyhedra of at most k sides (any fixed k). If $\mathscr{G}$ is a finite-dimensional vector space of functions on T, then

$$\operatorname{pos}(\mathscr{G}) := \{\{t\colon g(t) > 0\} \colon g \in \mathscr{G}\}$$

is a VC class. Any subset of a VC class is a VC class, and $\{C \vartriangle D\colon C, D \in \mathscr{C}\}$, $\{C \setminus D\colon C, D \in \mathscr{C}\}$, and $\{C \cap D\colon C, D \in \mathscr{C}\}$ are VC classes if $\mathscr{C}$ is one. If $\mathscr{C}$ and $\mathscr{D}$ are VC classes in T and U, respectively, then $\{C \times D\colon C \in \mathscr{C}, D \in \mathscr{D}\}$ is a VC class in $T \times U$. These and more facts about VC classes may be found in Dudley (1978, 1984).

A class $\mathscr{F}$ of real functions on T is called a *VC graph class* if the class

$$\mathscr{R} := \{\{(t, x)\colon 0 \leq x \leq f(t) \text{ or } f(t) \leq x \leq 0\} \colon f \in \mathscr{F}\}$$

of regions in $T \times \mathbb{R}$ which lie between $T \times \{0\}$ and the graph of some $f \in \mathscr{F}$ is a VC class of sets. The *index* $V(\mathscr{F})$ of $\mathscr{F}$ is defined to be $V(\mathscr{R})$. To avoid trivialities, we always tacitly assume $V(\mathscr{F}) \geq 2$. Any finite-dimensional vector space of functions (e.g., polynomials of bounded degree on $\mathbb{R}^d$) is a VC graph class (Dudley, 1984). Further, it follows from the above remarks that if $\mathscr{C}$ is a VC class of sets and q a real function on $\mathscr{C}$, then the class $\{1_C/q(C)\colon C \in \mathscr{C}\}$ corresponding to a weighted empirical process is a VC graph class. If $\mathscr{F}$ is a VC graph class and ψ a monotone function on $\mathbb{R}$, then $\{\psi \circ f\colon f \in \mathscr{F}\}$ is a VC graph class.

Because we are working in the possibly nonseparable Banach space $l^\infty(\mathscr{F})$, and because we take suprema over possibly uncountable classes $\mathscr{F}$, we must concern ourselves with measurability. First, we always assume the r.v.'s X_i are *canonically formed*, that is, the space $(\Omega, \mathscr{A}, \mathbb{P})$ on which the X_i are defined is the product of $(T^\infty, \mathscr{B}^\infty, P^\infty)$ and a copy $([0, 1], \Sigma, \lambda)$ of the unit interval with Lebesgue measure, where $(T^\infty, \mathscr{B}^\infty, P^\infty)$ is a countable product of copies of $(T, \mathscr{B}, P)$, and X_i is the ith coordinate function.

Second, we must specify just what we mean by weak convergence. Let Z_n and Z be $l^\infty(\mathscr{F})$-valued functions on $(\Omega, \mathscr{A}, \mathbb{P})$, with Z taking values a.s. in a separable subspace L of $l^\infty(\mathscr{F})$. We say Z_n *converges weakly* to Z ($Z_n \Rightarrow Z$) if

$$(2.1) \quad \int^* \varphi(Z_n)\, d\mathbb{P} \to \int \varphi(Z)\, d\mathbb{P}$$

for all bounded continuous real functions φ on $l^\infty(\mathscr{F})$,

where $\int^*$ denotes the upper integral, $\int^* \varphi\, d\mathbb{P} = \inf\{\int \psi\, d\mathbb{P}\colon \psi \geq \varphi,\ \psi \text{ measurable}\}$. (Note that the second integral in (2.1) is always defined since L is separable.) This kind of weak convergence was studied by Dudley (1966) and Hoffmann-Jørgensen (1986). For our applications, L will be the space of all bounded, uniformly ρ_P-continuous functions on $\mathscr{F}$.

For any $[-\infty, \infty]$-valued function φ on T, there exists a $[-\infty, \infty]$-valued $\mathscr{B}$-measurable function, denoted φ^*, on T with $\varphi \leq \varphi^*$ and with $\varphi^* \leq \psi$ a.s. for any measurable $\psi \geq \varphi$. The same holds for functions φ on $\Omega = T^\infty$. This φ^* satisfies $P[\varphi^* > u] = P^*[\varphi > u]$ for all u. See Chapter 3 of Dudley (1984).

Finally, we need some measurability assumptions to carry out particular calculations. First define

$$\mathscr{F}'(\alpha) := \{f - g\colon f, g \in \mathscr{F}, e_P(f, g) < \alpha\},$$

$$\mathscr{F}''(\alpha) := \{f - g\colon f, g \in \mathscr{F}'(\infty), e_P(f, g) < \alpha\}.$$

We say $\mathscr{F}$ is *supremum measurable* (for P) if $\sup_{\mathscr{F}} |Q(f, (X_i)_{i \geq 1})|$ is $\mathbb{P}$-completion measurable for each function Q which is a linear or quadratic function of finitely many of the $f(X_i)$. We say $\mathscr{F}$ is *deviation measurable* if the classes $\mathscr{F}$ and $\mathscr{F}'(\alpha)$ are supremum measurable for all $\alpha > 0$, and *admissibly measurable* if

$$\left\{f 1_{[a < F^* \leq b]}\colon f \in \mathscr{F}''(\infty), P\left(f^2 1_{[a < F^* \leq b]}\right) < \alpha\right\}$$

is deviation measurable for all $\alpha > 0$ and $0 \leq a < b < \infty$.

We say $\mathscr{F}$ is a *functional P-Donsker class* if $\mathscr{F}$ is P-pregaussian and there exists an i.i.d. sequence $\{Y_j\}$ of copies of the Gaussian process G_P, defined on $(\Omega, \mathscr{A}, \mathbb{P})$, such that

$$(2.2) \quad n^{-1/2} \max_{k \leq n} \sup_{f \in \mathscr{F}} \left| \sum_{j \leq k} \left(f(X_j) - P(f) - Y_j(f) \right) \right| \to 0 \quad \text{in probability.}$$

This apparently stronger property is actually equivalent to $\mathscr{F}$ being P-Donsker, as the following result of Dudley ((1984), Theorems 4.1.1. and 4.1.10; (1985), Theorem 5.2) and Dudley and Philipp (1983) shows.

THEOREM 2.1. *The following are equivalent:*

(i) $\mathscr{F}$ *is a P-Donsker class;*
(ii) $\mathscr{F}$ *is a functional P-Donsker class;*
(iii) *both*

$$\mathscr{F} \text{ is } \rho_P\text{-totally bounded, and} \tag{2.3}$$

$$\lim_{\alpha\to 0}\limsup_{n\to\infty}\mathbb{P}^*\left[\sup\{|\nu_n(f-g)|: f,g\in\mathscr{F}, \rho_P(f,g)<\alpha\}>\eta\right]=0 \quad \text{for all } \eta>0. \tag{2.4}$$

It is sufficient for each of (i)–(iii) *that* (2.3) *and* (2.4) *hold with* ρ_P *replaced by* e_P.

Here now is our main result.

THEOREM 2.2. *Let* $\mathscr{F}$ *be an admissibly measurable VC graph class of functions on* $(T, \mathscr{B}, P)$ *with envelope F. If*

$$\mathscr{F} \text{ is P-pregaussian} \tag{2.5}$$

and

$$u^2P[F^* > u] \to 0 \quad \text{as } u\to\infty, \tag{2.6}$$

then $\mathscr{F}$ *is a (functional) P-Donsker class. Conversely if* $\mathscr{F}$ *is a (functional) P-Donsker class then* (2.5) *holds and*

$$u^2P[F_P^* > u] \to 0 \quad \text{as } u\to\infty. \tag{2.7}$$

COROLLARY 2.3. *Let* $\mathscr{F}$ *be as in Theorem 2.2. If either* $\sup_{\mathscr{F}}|P(f)| < \infty$ *or* $\mathscr{F}_P$ *is an admissibly measurable VC graph class, then* $\mathscr{F}$ *is a (functional) P-Donsker class if and only if* $\mathscr{F}$ *is P-pregaussian and* $u^2P[F_P^* > u] \to 0$.

As discussed in the introduction, Corollary 2.3 is immediate from Theorem 2.2.

If $\mathscr{F}$ is contained in a finite-dimensional vector space of functions on T, then so is $\mathscr{F}_P$, and (assuming enough measurability) Corollary 2.3 applies.

REMARK 2.4. The condition in Corollary 2.3 that $\mathscr{F}_P$ be a VC graph class is not automatic. For example, let $\mathscr{G}$ be the class of all piecewise linear functions g on $[0,1]$ with finitely many pieces, with $-\frac{1}{2} < g < \frac{1}{2}$, with all vertices of the graph of g having rational coordinates, and with $P(g) = 0$, where P is the uniform law. Then $\mathscr{G}$ is countable, say $\mathscr{G} = \{g_1, g_2, \dots\}$, but $\mathscr{G}$ is easily seen not to be a VC graph class. Let $\mathscr{F} := \{n + g_n: n \geq 1\}$. Then the graph region class of $\mathscr{F}$ is a nested sequence of sets, so $\mathscr{F}$ is a VC graph class. But $\mathscr{F}_P = \mathscr{G}$ is not one.

For some classes $\mathscr{F}$ of functions, it may be difficult to verify (or conceivably even false) that $\mathscr{F}$ is a VC graph class, but easy to show $\mathscr{F}$ is of a form such as

$$\mathscr{F} = \{g - h: g\in\mathscr{G}, h\in\mathscr{H}\} \tag{2.8}$$

with $\mathscr{G}$, $\mathscr{H}$ VC graph classes. For example, suppose $\mathscr{F} = \{f(\cdot - t): t\in\mathbb{R}\}$ is the class of all translates of a fixed function f of bounded variation on $\mathbb{R}$. Since the class of all translates of a fixed monotone function on $\mathbb{R}$ forms a VC graph class (Pollard (1982b)), this $\mathscr{F}$ has form (2.8). For examples like these, the following results are useful. Each is an extension of a slightly weaker result of Dudley (1981).

PROPOSITION 2.5. *Any subset of a P-Donsker class is P-Donsker.*

PROPOSITION 2.6. *Let $\mathscr{F}_1, \dots, \mathscr{F}_m$ be P-Donsker classes of functions on $(T, \mathscr{B})$, and $M > 0$. Then $\{\Sigma_{i \le m} a_i f_i\colon f_i \in \mathscr{F}_i, \Sigma_{i \le m} |a_i| \le M\}$ is P-Donsker.*

COROLLARY 2.7. *$\mathscr{F} \cup \mathscr{G}$ is P-Donsker if and only if $\mathscr{F}$ and $\mathscr{G}$ both are P-Donsker.*

3. Weighted empirical processes. We turn next to weighted empirical processes, our main application of Theorem 2.2. Suppose we have a VC class $\mathscr{C}$ of sets and nonnegative function q on $\mathscr{C}$. We may study the weighted empirical process ν_n/q on $\mathscr{C}$ by considering the VC graph class

$$\mathscr{F} := \{1_C/q(C)\colon C \in \mathscr{C}\}. \tag{3.1}$$

It is most natural, however, to view ν_n and G_P as indexed by $\mathscr{C}$, not $\mathscr{F}$. Define

$$f_C := 1_C/q(C), \qquad C \in \mathscr{C},$$

and define pseudometrics on $\mathscr{C}$ by

$$r_P(C, D) := \rho_P(f_C, f_D),$$

$$d_P(C, D) := P(C \triangle D) = e_P(1_C, 1_D)^2.$$

For $\varphi \in l^\infty(\mathscr{C})$, set

$$\|\varphi\|_{\mathscr{C}, q} := \|\varphi/q\|_{\mathscr{C}}$$

and define

$$l_q^\infty(\mathscr{C}) := \{\varphi \in \ell^\infty(\mathscr{C})\colon \|\varphi\|_{\mathscr{C}, q} < \infty\},$$

endowed with norm $\|\cdot\|_{\mathscr{C}, q}$. The following is true from the definitions.

LEMMA 3.1. *The following are equivalent:*

(i) *$\mathscr{F}$ (of (3.1)) is P-Donsker;*

(ii) *$\nu_n/q \Rightarrow G_P/q$ in $l^\infty(\mathscr{C})$, and G_P/q has uniformly r_P-continuous sample paths on $\mathscr{C}$;*

(iii) *$\nu_n \Rightarrow G_P$ in $l_q^\infty(\mathscr{C})$, and G_P/q has uniformly r_P-continuous sample paths on $\mathscr{C}$.*

Of these formulations, (iii) is essentially the one used by O'Reilly (1974) for empirical d.f.'s.

To completely clarify the relation between $\nu_n/q \Rightarrow G_P/q$ in $l^\infty(\mathscr{C})$ and $\mathscr{F}$ being P-Donsker, we might ask whether the r_P-continuity of sample paths mentioned in Lemma 3.1 (which is part of the definition of $\mathscr{F}$ being P-Donsker) is really a natural requirement to impose. Could there be, for example, a natural notion of weak convergence in $l^\infty(\mathscr{C})$ for which we would have $\nu_n/q \Rightarrow G_P/q$, but with the sample paths of G_P/q not r_P-continuous, perhaps even with $\mathscr{L}(G_P/q)$ not concentrated on any separable subspace? (This latter possibility is not ruled out even in the distribution function case, since O'Reilly (1974) implicitly assumed $\mathscr{L}(G_P)$ concentrated on a separable subspace of $l_q^\infty(\mathscr{C})$ when he took a.s.-convergent versions of weakly convergent sequences.) The answer to this question is that r_P-continuity *is* natural, as our next two results show. First define

$$\mathscr{C}(a,b) := \{C \in \mathscr{C} : a \le P(C) \le b\}.$$

LEMMA 3.2. *Suppose $\mathscr{C}$ is P-pregaussian, suppose ν_n/q and G_P/q are $l^\infty(\mathscr{C})$-valued a.s., and suppose $\int \varphi(\nu_n/q)\,d\mathbb{P} \to \int \varphi(G_P/q)\,d\mathbb{P}$ whenever φ is a bounded continuous function on $l^\infty(\mathscr{C})$ and all these integrals are defined. Then $G_P(C)/q(C) \to 0$ a.s. as $P(C) \to 0$ or 1.*

PROPOSITION 3.3. *Let $\mathscr{C}$ be P-pregaussian. The following are equivalent:*

(i) *$\mathscr{F}$ (of (3.1)) is P-pregaussian;*

(ii) *$G_P(C)/q(C) \to 0$ a.s. as $P(C) \to 0$ or 1, and q is bounded away from 0 on $\mathscr{C}(\varepsilon, 1-\varepsilon)$ for all $\varepsilon > 0$;*

(iii) *$G_P/q \in l^\infty(\mathscr{C})$ a.s. and has uniformly r_P-continuous sample paths.*

If q is uniformly d_P-continuous on $\{C \in \mathscr{C}\colon 0 < P(C) < 1\}$ then these are equivalent to

(iv) *$G_P/q \in l^\infty(\mathscr{C})$ a.s. and has uniformly d_P-continuous sample paths.*

Under any reasonable notion of weak convergence in $l^\infty(\mathscr{C})$, the hypotheses of Lemma 3.2 will be satisfied. Therefore by Lemma 3.2, Proposition 3.3 ((ii) $\Rightarrow$ (iii)), and Lemma 3.1 (ii) $\Rightarrow$ (i)), we have $\nu_n/q \Rightarrow G_P/q$ in $l^\infty(\mathscr{C})$ if and only if $\mathscr{F}$ is P-Donsker.

Since $\nu_n(-1_{C^c}) = \nu_n(C)$ and similarly for G_P, ν_n on $\mathscr{F}$ is equivalent to ν_n on

$$\tilde{\mathscr{F}} := \{1_C/q(C)\colon C \in \mathscr{C}, P(C) \le \tfrac{1}{2}\} \cup \{-1_{C^c}/q(C)\colon C \in \mathscr{C}, P(C) > \tfrac{1}{2}\}.$$

Further, $\{C^c\colon C \in \mathscr{C}\}$ is a VC class. By Corollary 2.7, then, it is sufficient to consider $\mathscr{C}$ with $P(C) \le \frac{1}{2}$ for all $C \in \mathscr{C}$. Let us call such a class $\mathscr{C}$ *half-bounded*.

In our discussions of weighted empiricals, F and F_P will always refer to the envelopes of $\mathscr{F}$ and $\mathscr{F}_P$, where $\mathscr{F}$ is from (3.1). Thus

$$F(t) = 1/\inf\{q(C)\colon t \in C\}.$$

If $\mathscr{C}$ is half-bounded, then

$$P(f_C) \le 2P(C)(1 - P(C))/q(C) \le 2\int_C F_P^*\,dP \le 2P(F_P^*),$$

so if $u^2 P[F_P^* > u] \to 0$, then

$$\sup_{\mathscr{F}} |P(f)| \le 2P(F_P^*) < \infty.$$

It follows that (2.6) and (2.7) are equivalent here. Further, defining for any function φ on $\mathscr{C}$,

$$a_\varphi(\varepsilon) := P^*\left[\bigcup\{C \in \mathscr{C}: \varphi(C) \le \varepsilon\}\right],$$

(2.6) is equivalent to

$$a_q(\varepsilon) = o(\varepsilon^2) \quad \text{as } \varepsilon \to 0.$$

Define

$$\mathscr{C}'(a, b) := \{C^c: C \in \mathscr{C}, a \le P(C) \le b\},$$
$$q'(C) := q(C^c).$$

We have proved the following:

THEOREM 3.4. *Let $\mathscr{C}$ be a half-bounded VC class of subsets of $(\dot{T}, \mathscr{B}, P)$ and q a nonnegative function on $\mathscr{C}$. Suppose $\mathscr{F}$ (of (3.1)) is admissibly measurable. Then $\nu_n/q \Rightarrow G_P/q$ in $l^\infty(\mathscr{C})$ if and only if*

(3.2) $$a_q(\varepsilon) = o(\varepsilon^2) \quad \textit{as } \varepsilon \to 0$$

and

(3.3) $$G_P(C)/q(C) \to 0 \quad \textit{a.s. as } P(C) \to 0.$$

For $\mathscr{C}$ not half-bounded it is necessary and sufficient that (3.2) *and* (3.3) *hold both for $\mathscr{C}(0, \frac{1}{2})$ and when $\mathscr{C}$ and q are replaced by $\mathscr{C}'(\frac{1}{2}, 1)$ and q'.*

This theorem clarifies one point about the CLT for weighted empiricals: The only function of regularity conditions on q is to allow the replacement of (3.3), or possibly (3.2), with more readily checkable conditions, e.g., integral tests for (3.3). They do not prevent any pathologies intrinsically related to weak convergence. For an extensive discussion of regularity conditions in the distribution function case (and for the Brownian bridge) see Csörgő, Csörgő, Horváth and Mason (1986), Section 2.

Given P and the class $\mathscr{C}$, define

$$\psi(\varepsilon) := \left(2\varepsilon\left\{\log\left(\varepsilon^{-1} a_P(\varepsilon)\right) + \log\log \varepsilon^{-1}\right\}\right)^{1/2}.$$

Theorem 1.2 from Alexander (1986) tells us that

$$\limsup_{P(C) \to 0} |G_P(C)|/\psi(P(C)) < \infty \quad \text{a.s.},$$

so a sufficient condition for (3.3) is that

(3.4) $$q(C)/\psi(P(C)) \to \infty \quad \text{as } P(C) \to 0.$$

For this to be practical we have to be able to calculate $a_P(\cdot)$; examples will be given below.

Suppose $q(C)$ depends only on $P(C)$, i.e., $q = \varphi \circ P$ for some φ on [0, 1]. Then (3.2) becomes

$$\varphi^2(\varepsilon)/a_P(\varepsilon) \to \infty \quad \text{as } \varepsilon \to 0. \tag{3.5}$$

Suppose φ is nondecreasing and define

$$\bar{\varphi}(\varepsilon) := \inf\{\varphi(P(C)) \colon P(C) \geq \varepsilon, C \in \mathscr{C}\};$$

thus $\bar{\varphi}$ is the largest nondecreasing function for which $\bar{\varphi} \circ P = \varphi \circ P$ on $\mathscr{C}$. Then under (3.5), there is an integral test for (3.3), as is reflected in the following result.

THEOREM 3.5. *Let $\mathscr{C}$ be a half-bounded VC class of subsets of $(T, \mathscr{B}, P)$ and φ a nonnegative nondecreasing function on $[0, \frac{1}{2}]$. Suppose $\mathscr{F}$ (of (3.1)) is admissibly measurable. Then $\nu_n/\varphi \circ P \Rightarrow G_P/\varphi \circ P$ in $l^\infty(\mathscr{C})$ if and only if*

$$\varphi^2(\varepsilon)/a_P(\varepsilon) \to \infty \quad \textit{as } \varepsilon \to 0 \tag{3.6}$$

and

$$\int_0^{1/2} e^{-\varepsilon\bar{\varphi}^2(s)/s} \frac{ds}{s} < \infty \quad \textit{for all } \varepsilon > 0. \tag{3.7}$$

Similar considerations to those in Theorem 3.4 apply when $\mathscr{C}$ is not half-bounded.

For (3.7) it suffices that

$$\varphi(\varepsilon)/(\varepsilon \log\log \varepsilon^{-1})^{1/2} \to \infty \quad \text{as } \varepsilon \to 0.$$

By (3.6), the CLT can hold for nontrivial φ only if $a_P(\varepsilon) \to 0$ as $\varepsilon \to 0$. Unfortunately, many classes of interest do not satisfy this—the subintervals of [0, 1] under the uniform law, for example. Fortunately, a different sort of CLT may still hold—one involving truncation. Specifically, for an appropriate sequence $\tau_n \to 0$ we can define

$$\psi_n(C) := \frac{\nu_n(C)}{\varphi(P(C))} 1_{[P(C) \geq \tau_n]}$$

and look for conditions under which $\psi_n \Rightarrow G_P/q$. This problem is examined in a separate paper (Alexander, 1985).

The function $a_P(\cdot)$ was calculated in Alexander (1984b) for the following examples.

EXAMPLE 3.6. The weighted empirical d.f. Here $T = [0, 1]$, P is uniform, and $\mathscr{C}$ is $\{[0, s] \colon s \in [0, 1]\}$. (3.7) is exactly the Cibisov–O'Reilly condition (1.2). Clearly $a_P(\varepsilon) = \varepsilon$, so (3.6) says $\varphi^2(\varepsilon)/\varepsilon \to \infty$, which is easily shown to be a consequence of (3.7) for monotone φ. Thus we obtain O'Reilly's (1974) result that (3.7) is necessary and sufficient for $\nu_n/\varphi \Rightarrow G_P/\varphi$.

EXAMPLE 3.7. The multidimensional weighted uniform empirical d.f. Here $T = [0,1]^d$ $(d > 1)$, P is uniform, $\mathscr{C}$ is $\{[0,s]\colon\ s \in [0,1]^d\}$, and $a_P(\varepsilon) \sim c_d\varepsilon(\log \varepsilon^{-1})^{d-1}$ for some constant c_d. Therefore (3.6) implies (3.7), so it is necessary and sufficient for weak convergence that $\varphi^2(\varepsilon)/\varepsilon(\log \varepsilon^{-1})^{d-1} \to \infty$ as $\varepsilon \to 0$.

EXAMPLE 3.8. Let $T = \mathbb{R}^d$ $(d > 1)$, let P be a nondegenerate normal law, and let $\mathscr{C}$ consist of all closed half spaces. Then $a_P(\varepsilon) \sim c_d\varepsilon(\log \varepsilon^{-1})^{(d-1)/2}$ as $\varepsilon \to 0$ for some constant c_d, so again (3.6) implies (3.7). Therefore $\nu_n/\varphi \circ P \Rightarrow G_P/\varphi \circ P$ if and only if $\varphi^2(\varepsilon)/\varepsilon(\log \varepsilon^{-1})^{(d-1)/2} \to \infty$ as $\varepsilon \to 0$.

We can actually do better than Example 3.7 for weighted multidimensional uniform empirical d.f.'s. Here the limit process G_P is the tied-down version W_0 of the Brownian sheet W on $[0,1]^d$, so (3.3) clearly holds if and only if it holds for W. Write $q(s)$ for $q([0,s])$ and $|s|$ for $\prod_{i \le d} s_i$. Then Theorem 1 of Kalinauskaite (1979) provides an integral test for (3.3), which leads to the following.

COROLLARY 3.9. *Let q be a continuous function on $[0,1]^d$, nondecreasing in each argument. Let P be the uniform law on $T = [0,1]^d$. Then the weighted empirical d.f. ν_n/q on $\mathscr{C} = \{[0,s]\colon\ s \in [0,1]^d\}$ converges weakly to W_0/q in $l^\infty([0,1]^d)$ if and only if*

$$P[q < \varepsilon] = o(\varepsilon^2) \quad \text{as } \varepsilon \to 0$$

and

$$(3.8) \qquad \int_{[0,1]^d} e^{-\varepsilon q^2(s)/|s|} \frac{ds}{|s|} < \infty \quad \text{for all } \varepsilon > 0.$$

Here the weak convergence can also be viewed as occurring in the space $D([0,1]^d)$ of Wichura (1969), since both ν_n/q and G_P/q are in this subspace of $l^\infty([0,1]^d)$ a.s.

REMARK 3.10. Kalinauskaite proved the integral test (3.8) for (3.3) under the added assumptions that q be invariant under permutation of coordinates, with $q(s) \to 0$ as $|s| \to 0$. However, these are unnecessary restrictions. For q may be replaced by $\tilde{q}(s) := \min_{\sigma \in S_d} q(s_\sigma)$, where S_d is the dth permutation group and $s_\sigma := (s_{\sigma(1)}, \dots, s_{\sigma(d)})$; this does not affect (3.8) or (3.3). Further, q may be replaced by $\bar{q}(s) := q(s) \wedge q_0(s)$, where q_0 is any function satisfying (3.3) and (3.8) with $q_0(s) \to 0$ as $|s| \to 0$; this does not affect (3.8) or (3.3) either.

4. Proof of the main theorem.

Let us first sketch the proof of the sufficiency part of Theorem 2.2. Let $a_k \to \infty$ geometrically fast, and write $f \in \mathscr{F}'(\alpha)$ as $\sum_k f 1_{[a_k \le F < a_{k+1}]}$. The class $\{f 1_{[a_k \le F < a_{k+1}]}\colon\ f \in \mathscr{F}'(\alpha)\}$ is (almost) a VC graph class, with envelope at most $F_k := a_{k+1} 1_{[F \ge a_k]}$. By (2.6), $P(F_k^2)$ is bounded in k. This enables us to bound the number of functions $f_k \in \mathscr{F}'(\alpha)$ needed to approximate each $f 1_{[a_k \le F < a_{k+1}]}$ by $f_k 1_{[a_k \le F < a_{k+1}]}$ with an error (in $L^2(P)$

distance) of at most δ_k, for a suitable sequence (δ_k) of small constants. The sum over k can be truncated, and we obtain using bounds from Alexander (1984a):

$$\nu_n(f) \approx \sum_{k \le k_n} \nu_n\left(f_k 1_{[a_k \le F < a_{k+1}]}\right). \tag{4.1}$$

A method first applied to empirical processes by Giné and Zinn (1984) is now used: by attaching i.i.d. $N(0,1)$ coefficients g_i to each $f(X_i)$, the sup over $\mathscr{F}'(\alpha)$ of the second process in (4.1) can be compared to the sup of $G_n(f) := \sum_{k \le k_n} W_P^{(k)}(f_k 1_{[a_k \le F < a_{k+1}]})$, where $(W_P^{(k)}, k \ge 1)$ are i.i.d. copies of the non-tied-down form W_P of G_P. G_n, in turn, is close to a copy of W_P, which has a small sup over $\mathscr{F}'(\alpha)$ since $\mathscr{F}$ is P-pregaussian. Therefore we obtain (2.4) for e_P. (2.3) follows from (2.5).

The above method is similar in many ways to one applied in Alexander and Pyke (1986) to partial-sum processes. Variants of the idea of approximating to within a different δ_k on each stratum $[a_k \le F < a_{k+1}]$ were used by Bass and Pyke (1985) and Bass (1985).

Let us list some of the lemmas we will need. Let (ε_i) be a Rademacher sequence, i.e., an i.i.d. sequence with $P[\varepsilon_i = 1] = P[\varepsilon_i = -1] = \frac{1}{2}$, independent of the sequence (X_i), and define the symmetrized empirical process ν_n^0 by

$$\nu_n^0(f) := n^{-1/2} \sum_{i \le n} \varepsilon_i f(X_i).$$

The following symmetrization inequality is proved by an argument which dates back at least to Pollard (1982a)—see Giné and Zinn (1984), Lemma 2.7.

LEMMA 4.1. *For all* η, $\alpha > 0$,

$$\mathbb{P}\left[\|\nu_n\|_{\mathscr{F}'(\alpha)} > \eta\right] \le 2\mathbb{P}^*\left[\|\nu_n^0\|_{\mathscr{F}'(\alpha)} > \left(\eta - (2\alpha)^{1/2}\right)/2\right].$$

Lemma 4.3 enables us to use ν_n^0 in place of ν_n in (2.4). The next two lemmas will help us compare the tails of $\|\nu_n^0\|_{\mathscr{F}'(\alpha)}$ to those of $\|G_P\|_{\mathscr{F}'(\alpha)}$. The first appeared in Alexander and Pyke (1986) and is based on Lemma 2.9 of Giné and Zinn (1984). The second is essentially Theorem 2.17 of the Giné and Zinn paper, which is based on a result of Fernique (1974).

LEMMA 4.2. *Let* $\{\varphi_j\colon j \in J\}$ *be a finite set of real functions on a space* S, *and let* $\{Y_j\colon j \in J\}$ *and* $\{Z_j\colon j \in J\}$ *be r.v.'s with* $\{Y_j\}$ *independent of* $\{Z_j\}$ *and* $EZ_j = 1$. *Then*

$$E\left\| \sum_{j \in J} Y_j \varphi_j \right\|_S \le E\left\| \sum_{j \in J} Y_j Z_j \varphi_j \right\|_S$$

provided both these suprema are measurable.

LEMMA 4.3. *Let* Y_1 *and* Y_2 *be centered Gaussian processes indexed by a countable set* S. *Suppose for all* $s, t \in S$,

$$E\left(Y_1(t) - Y_1(s)\right)^2 \le E\left(Y_2(t) - Y_2(s)\right)^2,$$

and suppose $0 \in \text{range}(Y_1)$ *a.s. Then*

$$E\|Y_1\|_S \le 2E\|Y_2\|_S.$$

Given a set S with a pseudometric d on it, define for $\varepsilon > 0$:

$$D(\varepsilon, S, d) := \max\{k \ge 1\colon \text{there exist } s_1, \dots, s_k \in S \text{ such that } d(s_i, s_j) \ge \varepsilon \text{ for all } i \ne j\},$$

$$N(\varepsilon, S, d) := \min\Big\{k \ge 1\colon \text{there exist } s_1, \dots, s_k \in S \text{ such that } \min_{i \le k} d(s, s_i) < \varepsilon \text{ for all } s \in S\Big\}.$$

Then

$$N(\varepsilon, S, d) \le D(\varepsilon, S, d) \le N(\varepsilon/2, S, d). \tag{4.2}$$

The function $\log N(\cdot, S, d)$ is called the *metric entropy* of S, and the set $\{s_1, \dots, s_k\}$ in the definition of N is called an *ε-net*. The following result is essentially due to Pollard (1984) and is based on a lemma of Dudley (1978).

Lemma 4.4. *Let $\mathscr{F}$ be a VC graph class of functions on $(T, \mathscr{B})$ with envelope F, and $v := 2(V(\mathscr{F}) - 1)$. Then there exists a constant $K_0 = K_0(V(\mathscr{F}))$ such that*

$$D\big(\varepsilon P(F^2)^{1/2}, \mathscr{F}, e_P\big) \le K_0 \varepsilon^{-v}$$

for all $0 < \varepsilon \le 1$ and all laws P on $(T, \mathscr{B})$.

Given $\mathscr{F}$ P-pregaussian, let ξ be a normal $N(0, 1)$ r.v. independent of G_P and define W_P on $\mathscr{F}$ by

$$W_P(f) := G_P(f) + P(f)\xi.$$

Then

$$\big(E(W_P(f) - W_P(g))^2\big)^{1/2} = e_P(f, g)$$

and W_P is linear a.s., so W_P extends to all $\mathscr{F}'(\alpha)$.

Lemma 4.5. *There exists a universal constant K_1 such that*

$$E\|W_P\|_{\mathscr{F}''(\alpha)} \le K_1 \int_0^\alpha \big(\log D(\varepsilon, \mathscr{F}, e_P)\big)^{1/2} d\varepsilon$$

for all P-pregaussian $\mathscr{F}$ and all $\alpha > 0$.

Proof. This follows directly from the proof of Theorem 2.1 of Dudley (1973), (4.2), and the observation that $N(\varepsilon, \mathscr{F}''(\alpha), e_P) \le N(\varepsilon/8, \mathscr{F}, e_P)^2$. See also Theorem 2.15 of Giné and Zinn (1984). □

Let

$$Lx := \log(\max(x, e)).$$

The following result is essentially an immediate consequence of Theorem 2.8 of Alexander (1984a) and Bernstein's inequality (see Hoeffding (1963)), though minor modifications of the former are needed to obtain a bound for $\|\nu_n^0\|_{\mathscr{F}''(\alpha)}$ instead of for $\|\nu_n\|_{\mathscr{F}}$.

PROPOSITION 4.6. *Let $\mathscr{F}$ be a VC graph class of functions on $(T, \mathscr{B}, P)$, uniformly bounded in magnitude by $a > 0$. Let $\alpha > 0$ and suppose $\mathscr{F}''(\alpha)$ is deviation measurable. There exists a constant $K_2 = K_2(V(\mathscr{F}))$ such that if*

$$aM/n^{1/2}\alpha^2 \geq 3 \quad \text{and} \quad Mn^{1/2}/a \geq K_2 L(na^2\alpha^{-2}), \tag{4.3}$$

then

$$\mathbb{P}^*\left[\|\nu_n^0\|_{\mathscr{F}''(\alpha)} > M\right] \leq 16\exp\left(-Mn^{1/2}/16a\right).$$

PROOF OF THEOREM 2.1. We begin with sufficiency of (2.5) and (2.6). By Theorem 2.1, it suffices to prove (2.3) and (2.4), with e_P replacing ρ_P. But (2.5) implies $(\mathscr{F}, \rho_P)$ is totally bounded (Dudley (1967)) and (2.6) implies $\mathscr{F}$ is bounded in $L^1(P)$, so $(\mathscr{F}, e_P)$ is totally bounded. It remains to prove (2.4) for e_P.

By Lemma 4.1, it suffices to prove (2.4) for ν_n^0. Our first step is truncation. By (2.6), there exist $\tau_n \to 0$ such that

$$\lim_n nP\left(F^* > \tau_n n^{1/2}\right) = 0. \tag{4.4}$$

Fix η, $\delta > 0$ and define

$$\gamma_n := \inf\left\{\gamma > 0: nP\left(F^* > \gamma_n n^{1/2}\right) < \eta\delta\tau_n^{-1}\right\}.$$

Then $\gamma_n \leq \tau_n$ for large n, so $\gamma_n \to 0$. Consider now the processes

$$\nu_n'(f) := \nu_n^0\left(f 1_{[F^* > \tau_n n^{1/2}]}\right),$$

$$\nu_n''(f) := \nu_n^0\left(f 1_{[\gamma_n n^{1/2} < F^* \leq \tau_n n^{1/2}]}\right).$$

By (4.4),

$$\begin{aligned}\mathbb{P}^*\left[\|\nu_n'\|_{\mathscr{F}'(\alpha)} > \eta\right] &\leq \mathbb{P}\left[F^*(X_i) > \tau_n n^{1/2} \text{ for some } i \leq n\right] \\ &= o(1) \quad \text{as } n \to \infty.\end{aligned} \tag{4.5}$$

Further,

$$\|\nu_n''\|_{\mathscr{F}'(\alpha)} \leq n^{-1/2} \sum_{i \leq n} 2\tau_n n^{1/2} 1_{[F^* > \gamma_n n^{1/2}]},$$

and the expectation of the latter is at most $2\eta\delta$. Therefore

$$\mathbb{P}^*\left[\|\nu_n''\|_{\mathscr{F}'(\alpha)} > \eta\right] \leq 2\delta. \tag{4.6}$$

Thus we may work with the truncated process

$$\nu_n^t(f) := \nu_n^0(f) - \nu_n'(f) - \nu_n''(f) = \nu_n^0\left(f 1_{[F^* \leq \gamma_n n^{1/2}]}\right).$$

(We point out that the reason for making our second truncation is (4.12) below.)

Next we stratify the process ν_n^t. Let (a_k) be a geometrically increasing sequence, say $a_k = 4^k$, and define

$$k_n := \max\{k \geq 0 \colon a_k < \gamma_n n^{1/2}\},$$

$$I_k := \begin{cases} (a_k, a_{k+1}], & \text{if } 0 \leq k < k_n, \\ (a_{k_n}, \gamma_n n^{1/2}], & \text{if } k = k_n, \end{cases}$$

$$p_k := P[F^* > a_k].$$

Since F is only used as an upper bound, we may assume $F > 1$. Then

$$\nu_n^t(f) = \sum_{k \leq k_n} \nu_n^0\big(f 1_{[F^* \in I_k]}\big).$$

Now we approximate f on each stratum. Let K_0 and v be the constants of Lemma 4.4 and define

$$H(\varepsilon) := 1 + L\big(K_0(4\varepsilon^{-1})^v\big);$$

we may assume $K_0 \geq 1$. Let θ_1 and θ_2 be constants satisfying

(4.7) $$\theta_1 \leq 3/256, \qquad 4\theta_1\theta_2 > 3 \quad \text{and} \quad \theta_2\theta_1^{-1} \geq \max(2K_2, 64),$$

where K_2 is the constant of Proposition 4.6. Fix n and α and define δ_k and η_k for $k \leq k_n$ by

$$a_k = \theta_1\big(n\delta_k^2/H(\delta_k)\big)^{1/2}, \qquad \eta_k := \theta_2\delta_k H(\delta_k)^{1/2}.$$

Then since $\gamma_n \to 0$ we have

$$\delta_{k_n} = \max_{k \leq k_n} \delta_k \to 0 \quad \text{as } n \to \infty,$$

while by an easy calculation,

(4.8) $$\sum_{k \leq k_n} \eta_k \leq 5\theta_2 \int_0^{\delta_{k_n}} H(\varepsilon)^{1/2} d\varepsilon = o(1), \quad \text{as } n \to \infty.$$

Let

$$\mathscr{F}_k := \big\{f 1_{[F^* \in I_k]} \colon f \in \mathscr{F}\big\},$$

$$\mathscr{G}_k^\alpha := \big\{f 1_{[F^* \in I_k]} \colon f \in \mathscr{F}'(\alpha)\big\}.$$

Then $\mathscr{F}_k$ is a VC graph class of index $V(\mathscr{F}_k) \leq V(\mathscr{F})$, with envelope at most

$$F_k := a_{k+1} 1_{[F^* > a_k]}.$$

By (2.6), $P(F_k^2) \to 0$ as $k \to \infty$, so, by rescaling if necessary, we may assume $P(F_k^2) \leq 1$ for all k. Then by (4.2) and Lemma 4.4,

$$D(\delta_k, \mathscr{G}_k^\alpha, e_P) \leq D(\delta_k, \mathscr{G}_k^\infty, e_P) \leq N(\delta_k/2, \mathscr{G}_k^\infty, e_P)$$
$$\leq N(\delta_k/4, \mathscr{F}_k, e_P)^2 \leq D(\delta_k/4, \mathscr{F}_k, e_P)^2 \leq \exp\big(H(\delta_k)\big).$$

Thus there is a finite subset $\mathscr{H}_k^\alpha$ of $\mathscr{F}'(\alpha)$ of cardinality at most $\exp(H(\delta_k))$

such that

$$e_P\left(f1_{[F^*\in I_k]}, g1_{[F^*\in I_k]}\right) \geq \delta_k \quad \text{if } f, g \in \mathscr{H}_k^\alpha \text{ and } f \neq g \tag{4.9}$$

and such that for each $f \in \mathscr{F}'(\alpha)$, there is an element of $\mathscr{H}_k^\alpha$, call it $Q_k f$, such that

$$e_P\left(f1_{[F^*\in I_k]}, (Q_k f)1_{[F^*\in I_k]}\right) < \delta_k. \tag{4.10}$$

Thus we may write $\nu_n^t = \nu_n^{(1)} + \nu_n^{(2)}$, where

$$\nu_n^{(1)}(f) := \sum_{k\leq k_n} \nu^0\left((Q_k f)1_{[F^*\in I_k]}\right),$$

$$\nu_n^{(2)}(f) := \sum_{k\leq k_n} \nu_n^0\left((f - Q_k f)1_{[F^*\in I_k]}\right).$$

To handle the error term $\nu_n^{(2)}$, observe that by (4.10),

$$\left\{(f - Q_k f)1_{[F^*\in I_k]}: f \in \mathscr{F}'(\alpha)\right\} \subset \mathscr{F}_k''(\delta_k).$$

Therefore by (4.8), for large n,

$$\mathbb{P}^*\left[\|\nu_n^{(2)}\|_{\mathscr{F}'(\alpha)} > \eta\right] \leq \mathbb{P}^*\left[\|\nu_n^0\|_{\mathscr{F}_k''(\delta_k)} > \eta_k \text{ for some } k \leq k_n\right], \tag{4.11}$$

so we would like to use Proposition 4.6. Fix $k \leq k_n$ and an integer $l \geq 1$. Define a new probability measure $P^{(k)}$ on $(T, \mathscr{B})$ by

$$P^{(k)}(\cdot) = P(\cdot|F^* > a_k)$$

and let $\mathbb{P}_k := (P^{(k)})^\infty \times \lambda$ be the corresponding product measure on $(T^\infty, \mathscr{B}^\infty) \times ([0,1], \Sigma)$. On the event $[nP_n(F^* > a_k) = l]$, the distribution of ν_n^0 on $\mathscr{F}_k''(\delta_k)$ under $\mathbb{P}$ is the same as the distribution of $(l/n)^{1/2}\nu_l^0$ under $\mathbb{P}_k$.

Before we take advantage of this latter fact, we must consider what values of l will arise. From the definitions we get

$$\min_{k\leq k_n} nP(F^* > a_k) = nP\left(F^* > a_{k_n}\right) \to \infty. \tag{4.12}$$

Since the events $\{[F^* > a_k],\ k \geq 1\}$ are nested, the process $\{nP_n(F^* > a_k),\ k \geq 1\}$ can be imbedded in a uniform empirical d.f. Therefore (4.12) and a theorem of Chang (1964) tell us that the event

$$R_n := \left[n^{-1} \leq P_n(F^* > a_k) \leq 2P(F^* > a_k) \text{ for all } k \leq k_n\right]$$

satisfies

$$\mathbb{P}(R_n^c) \to 0 \quad \text{as } n \to \infty. \tag{4.13}$$

Therefore we need only consider $1 \leq l \leq 2np_k$.

Define

$$\mathscr{F}_k''(\beta, P^{(k)}) := \{f - g: f, g \in \mathscr{F}_k'(\infty), e_{P^{(k)}}(f,g) < \beta\}.$$

For functions f, g supported on $[F^* > a_k]$, we have $e_{P^{(k)}}(f,g) = p_k^{-1/2}e_P(f,g)$. It follows that $\mathscr{F}_k''(\delta_k) = \mathscr{F}_k''(p_k^{-1/2}\delta_k, P^{(k)})$. It therefore follows from the above discussion (using the fact the X_i are canonically formed to avoid measurability difficulties) that

$$\begin{aligned}
&\mathbb{P}^*\Big[\|\nu_n^0\|_{\mathscr{F}_k''(\delta_k)} > \eta_k \text{ for some } k \le k_n\Big] \\
&\quad \le \mathbb{P}(R_n^c) + \sum_{k \le k_n} \mathbb{P}^*\Big[\|\nu_n^0\|_{\mathscr{F}_k''(\delta_k)} > \eta_k;\, R_n\Big] \\
(4.14)\qquad &\quad = o(1) + \sum_{k \le k_n} \sum_{l \le 2np_k} \mathbb{P}_k^*\Big[\|\nu_l^0\|_{\mathscr{F}_k''(p_k^{-1/2}\delta_k,\, P^{(k)})} \le (n/l)^{1/2}\eta_k\Big] \\
&\qquad\qquad \times \mathbb{P}\big[nP_n(F^* > a_k) = l\big].
\end{aligned}$$

Hence, provided we can verify (4.3) for the appropriate constants, we can use Proposition 4.6, (4.7) and (4.8) to obtain

$$\begin{aligned}
\mathrm{RHS}(4.14) &\le o(1) + \sum_{k \le k_n} 16\exp\big(-n^{1/2}\eta_k/16a_{k+1}\big) \\
&= o(1) + \sum_{k \le k_n} 16\exp\big(-(64\theta_1)^{-1}\theta_2 H(\delta_k)\big) \\
(4.15)\qquad &\le o(1) + \sum_{k \le k_n} \delta_k \\
&\le o(1) + \sum_{k \le k_n} \eta_k = o(1) \quad \text{as } n \to \infty.
\end{aligned}$$

Combining this with (4.14) and (4.11) gives

$$(4.16)\qquad \mathbb{P}^*\Big[\|\nu_n^{(2)}\|_{\mathscr{F}'(\alpha)} > \eta\Big] = o(1) \quad \text{as } n \to \infty.$$

Let us now verify (4.3) as required, with $a_{k+1}, (n/l)^{1/2}\eta_k$, l, and $p^{-1/2}\delta_k$ in the roles of the a, M, n, and α, respectively, in (4.3). Since $l \le 2np_k$, the first inequality follows from

$$a_{k+1}n^{1/2}\eta_k p_k/l\delta_k^2 \ge 4a_k\eta_k/n^{1/2}\delta_k^2 = 4\theta_1\theta_2 > 3.$$

To prove the second, recall we have assumed $a_{k+1}^2 p_k = P(F_k^2) \le 1$ for all k. Hence for large n,

$$\begin{aligned}
L\big(la_{k+1}^2 p_k/\delta_k^2\big) &\le L(np_k) + L\big(\delta_k^{-2}\big) \\
&\le L\big(na_k^{-2}\big) + L\big(\delta_k^{-2}\big) \le 6L\big(\delta_k^{-1}\big) \le 8H(\delta_k),
\end{aligned}$$

while by (4.7),

$$(n/l)^{1/2}\eta_k l^{1/2}/a_{k+1} = 4\theta_2\theta_1^{-1}H(\delta_k) \ge 8K_2 H(\delta_k).$$

Thus (4.3) is established.

It remains to handle $\nu_n^{(1)}$, which we rewrite as

$$\nu_n^{(1)}(f) = \sum_{i \le n} \varepsilon_i U_{ni}(f),$$

where

$$U_{ni}(f) := \sum_{k \le k_n} n^{-1/2}(Q_k f)(X_i) 1_{[F^*(X_i) \in I_k]}.$$

Observe that since each $\mathscr{H}_k^\alpha$ is finite, measurability is no longer a problem. Let (ξ_i) be an i.i.d. sequence of normal $N(0,1)$ r.v.'s, independent of (X_i) and (ε_i), and let $\mu := E|\xi_1|$. Let E_ξ denote expectation with respect to the ξ_i's, i.e., with the (ε_i) and (X_i) held fixed, and define E_ε similarly. Since $\mathscr{L}(\varepsilon_i\xi_i) = \mathscr{L}(\xi_i)$, it follows from Lemma 4.2 that

$$E_\varepsilon\|\nu_n^{(1)}\|_{\mathscr{F}'(\alpha)} \le \mu^{-1}E_\xi\|\nu_n^\xi\|_{\mathscr{F}'(\alpha)}, \tag{4.17}$$

where

$$\nu_n^\xi(f) := \sum_{i\le n}\xi_i U_{ni}(f).$$

Conditionally given (X_i), ν_n^ξ is a Gaussian process, to which we would like to apply Lemma 4.3. Note that since $0 \in \mathscr{F}'(\alpha)$, we have $0 \in \operatorname{range}(\nu_n^\xi)$ a.s. Since $\mathscr{F}_k$ is a uniformly bounded VC graph class, it is P-pregaussian by Lemma 4.4 above and Theorem 1.2 of Dudley (1973). Therefore so is $\mathscr{G}_k^\alpha$, and we may take a copy W_k of W_P on each $\mathscr{G}_k^\alpha$, with the W_k's independent. Define

$$G_n(f) := \sum_{k\le k_n} 2W_k\Big((Q_k f)1_{[F^*\in I_k]}\Big),$$

$$G_n'(f) := \sum_{k\le k_n} 2W_k\Big(f1_{[F^*\in I_k]}\Big)$$

for $f \in \mathscr{F}'(\alpha)$. For $f, g \in \mathscr{F}'(\alpha)$,

$$E_\xi\big(\nu_n^\xi(f) - \nu_n^\xi(g)\big)^2 = \sum_{i\le n}(U_{ni}(f) - U_{ni}(g))^2$$

$$= \sum_{k\le k_n} P_n\Big((Q_k f - Q_k g)^2 1_{[F^*\in I_k]}\Big),$$

while

$$E(G_n(f) - G_n(g))^2 = \sum_{k\le k_n} 4P\Big((Q_k f - Q_k g)^2 1_{[F^*\in I_k]}\Big).$$

Thus on the event

$$L_n := \Big[P_n\Big((Q_k f - Q_k g)^2 1_{[F^*\in I_k]}\Big) \le 4P\Big((Q_k f - Q_k g)^2 1_{[F^*\in I_k]}\Big)$$
$$\text{for all } k \le k_n \text{ and } f, g \in \mathscr{F}'(\alpha)\Big],$$

we obtain

$$E_\xi\big(\nu_n^\xi(f) - \nu_n^\xi(g)\big)^2 \le E(G_n(f) - G_n(g))^2$$

for all $f, g \in \mathscr{F}'(\alpha)$. Therefore by Lemma 4.3, on L_n,

$$E_\xi\|\nu_n^\xi\|_{\mathscr{F}'(\alpha)} \le 2E\|G_n\|_{\mathscr{F}'(\alpha)}. \tag{4.18}$$

Meanwhile, by Lemma 4.5, Lemma 4.4 and (4.8),

$$(4.19)\qquad \begin{aligned} E\|G_n - G_n'\|_{\mathscr{F}'(\alpha)} &\le \sum_{k \le k_n} E\|W_k\|_{\mathscr{F}_k''(\delta_k)} \\ &= O\Big(\sum_{k\le k_n} \delta_k H(\delta_k)^{1/2}\Big) \\ &= o(1) \quad \text{as } n \to \infty. \end{aligned}$$

Since

$$\begin{aligned} E(G_n'(f) - G_n'(g))^2 &= P\big((f-g)^2 1_{[F^* \le \gamma_n n^{1/2}]}\big) \\ &\le P\big((f-g)^2\big) = E(W_P(f) - W_P(g))^2, \end{aligned}$$

Lemma 4.3 gives

$$(4.20)\qquad E\|G_n'\|_{\mathscr{F}'(\alpha)} \le 2E\|W_P\|_{\mathscr{F}'(\alpha)}.$$

By standard integrability properties of Gaussian r.v.'s, $E\|G_P\|_{\mathscr{F}'(\alpha)}$ is finite for all α; since $\mathscr{F}$ is L^1-bounded, so is $E\|W_P\|_{\mathscr{F}'(\alpha)}$. Thus by dominated convergence and the uniform e_P-continuity of W_P on $\mathscr{F}$,

$$E\|W_P\|_{\mathscr{F}'(\alpha)} \to 0 \quad \text{as } \alpha \to 0.$$

Combining this with (4.17), (4.18), (4.19) and (4.20) shows

$$\begin{aligned} \mathbb{P}\big[\|\nu_n^{(1)}\|_{\mathscr{F}'(\alpha)} > \eta\big] &\le \mathbb{P}(L_n^c) + \sup_{L_n} \eta^{-1} E_\varepsilon \|\nu_n^{(1)}\|_{\mathscr{F}'(\alpha)} \\ &= \mathbb{P}(L_n^c) + o(1) \end{aligned}$$

with the $o(1)$ being when we let $n \to \infty$ and then $\alpha \to 0$.

We will be finished when we show $\mathbb{P}(L_n^c) \to 0$ (regardless of α). Fix $k \le k_n$ and $f, g \in \mathscr{F}'(\alpha)$, and let $h_k := (Q_k f - Q_k g)^2 1_{[F^* \in I_k]}$. Then $\|h_k\|_\infty \le 16a_{k+1}^2$ and $\operatorname{var}(h_k(X)) \le P(h_k^2) \le 16a_{k+1}^2 P(h_k)$. Further, by (4.9), if $h_k \ne 0$ then $P(h_k) \ge \delta_k^2$. Therefore by Bernstein's inequality,

$$\begin{aligned} \mathbb{P}[P_n(h_k) > 4P(h_k)] &\le \mathbb{P}\big[\nu_n(h_k) > 3n^{1/2}P(h_k)\big] \\ &\le \exp\big(-9nP(h_k)/64a_{k+1}^2\big) \\ &\le \exp\big(-9H(\delta_k)/256\theta_1\big). \end{aligned}$$

Therefore, using (4.7),

$$\begin{aligned} \mathbb{P}(L_n^c) &\le \sum_{k\le k_n} |\mathscr{H}_k^\alpha|^2 \exp\big(-9H(\delta_k)/256\theta_1\big) \\ &\le \sum_{k \le k_n} \exp(-H(\delta_k)) \\ &= o(1) \end{aligned}$$

as in (4.15), and the sufficiency part of the theorem is proved.

Since (2.5) is necessary by definition, we turn to (2.7). But (2.7) is equivalent to

$$(4.21)\qquad n\mathbb{P}^*\big[\|n^{-1/2}(\delta_X - P)\|_{\mathscr{F}} > \delta\big] \to 0 \quad \text{for all } \delta > 0,$$

which is necessary by Theorem 2.2 and Remark 2.3 of Alexander (1984c). (Note that necessity of the analog of (4.21) in separable Banach spaces is well known.) □

Proposition 2.5 is immediate from Theorem 2.1, so we move on to Proposition 2.6.

PROOF OF PROPOSITION 2.6. It is sufficient to show that, given $M > 0$ and P-Donsker classes $\mathscr{F}$ and $\mathscr{G}$, $\mathscr{J} := \{f + g\colon f \in \mathscr{F}, g \in \mathscr{G}\}$ and $\mathscr{K} := \{af\colon f \in \mathscr{F}, |a| \le M\}$ are P-Donsker. Further, we may assume $P(f) = 0$ for all $f \in \mathscr{F} \cup \mathscr{G}$.

Clearly $\mathscr{J}$ and $\mathscr{K}$ are totally bounded, since $\mathscr{F}$ and $\mathscr{G}$ are by Theorem 2.1. Therefore we must prove (2.4) for $\mathscr{J}$ and $\mathscr{K}$ assuming (2.4) for $\mathscr{F}$ and $\mathscr{G}$.

Let $\mathscr{F}_\delta$ and $\mathscr{G}_\delta$ be δ-nets in $\mathscr{F}$ and $\mathscr{G}$, with $|\mathscr{F}_\delta| = N(\delta, \mathscr{F}, \rho_P)$ and $|\mathscr{G}_\delta| = N(\delta, \mathscr{G}, \rho_P)$. By Theorem 2.16(b) of Giné and Zinn (1984), which is based on a result of Sudakov (1971), since $\mathscr{F}$ and $\mathscr{G}$ are P-pregaussian we have

$$\delta^2 \log N(\delta, \mathscr{F}, \rho_P) \to 0 \quad \text{as } \delta \to 0, \tag{4.22}$$

and similarly for $\mathscr{G}$.

Suppose $e, f \in \mathscr{F}$, $g, h \in \mathscr{G}$, and $(e + g) - (f + h) \in \mathscr{J}'(\delta)$. Let $e_\delta \in \mathscr{F}_\delta$ with $\rho_P(e, e_\delta) < \delta$, and similarly for f_δ, g_δ and h_δ. Then

$$\begin{aligned} |\nu_n((e+g) - (f+h))| \le\ & 2\|\nu_n\|_{\mathscr{F}'(\delta)} + 2\|\nu_n\|_{\mathscr{G}'(\delta)} \\ & + |\nu_n((e_\delta - g_\delta) - (f_\delta + h_\delta))|. \end{aligned}$$

Since

$$\rho_P(e_\delta + g_\delta, f_\delta + h_\delta) \le 5\delta,$$

it follows that

$$\begin{aligned} \limsup_n \mathbb{P}^*\left[\|\nu_n\|_{\mathscr{J}'(\delta)} > 6\eta\right] \le\ & \limsup_n \mathbb{P}^*\left[\|\nu_n\|_{\mathscr{F}'(\delta)} > \eta\right] \\ & + \limsup_n \mathbb{P}^*\left[\|\nu_n\|_{\mathscr{G}'(\delta)} > \eta\right] \\ & + 2|\mathscr{F}_\delta|^2 |\mathscr{G}_\delta|^2 \exp(-\eta^2/50\delta^2), \end{aligned}$$

and (2.4) for $\mathscr{J}$ follows from (2.4) for $\mathscr{F}$ and $\mathscr{G}$ and from (4.22).

The proof for $\mathscr{K}$ is similar, but now if $f, g \in \mathscr{F}$, $|a| \le M$, $|B| \le M$, and $af - bg \in \mathscr{K}'(\delta)$, then we approximate a and b by the nearest integer multiples of δ. There are at most $O(M\delta^{-1})$ such multiples, and the error in ν_n created by this approximation is at most $2\delta\|\nu_n\|_{\mathscr{F}}$. Since $\|\nu_n\|_{\mathscr{F}}$ is bounded in probability, the proof can proceed similarly to the proof for $\mathscr{J}$. □

5. Proofs for weighted empiricals. We will need the following.

LEMMA 5.1. *Let $\mathscr{C}$ be a P-pregaussian class of sets and q a real function on $\mathscr{C}$. If $G_P/q \in l^\infty(\mathscr{C})$ a.s., then*

$$q^2(C)/P(C) \to \infty \quad \text{as } P(C) \to 0. \tag{5.1}$$

PROOF. Suppose (5.1) is false. Then take a sequence (C_i) in $\mathscr{C}$ with

$$P(C_{i+1}) \le P(C_i)/2 \quad \text{and} \quad q^2(C_i)/P(C_i) \le M < \infty$$

for all i. If $i > j$ then

$$P\big(1_{C_i}/q(C_i) - 1_{C_j}/q(C_j)\big)^2 \ge P(C_j/C_i)/q^2(C_j) \ge 1/2M.$$

Therefore $\mathscr{F} := \{1_C/q(C)\colon\ C \in \mathscr{C}\}$ is not e_P-totally bounded. But by Proposition 3.4 of Dudley (1967), this contradicts $W_P/q \in l^\infty(\mathscr{C})$ a.s., hence also $G_P/q \in l^\infty(\mathscr{C})$ a.s. □

PROOF OF LEMMA 3.2. Since $\nu_n/q \in l^\infty(\mathscr{C})$ a.s., we have

$$\|\nu_1/q\|_{\mathscr{C}} = \|(\delta_X - P)/q\|_{\mathscr{C}} < \infty \quad \text{a.s.}$$

It follows that a.s., $\delta_X(C) = 0$ for all C with both $P(C)$ and $q(C)$ sufficiently small. Therefore by Lemma 5.1,

$$\limsup_{\substack{P(C)\to 0\\ q(C)\to 0}} |\nu_n(C)/q(C)| = \limsup_{\substack{P(C)\to 0\\ q(C)\to 0}} n^{1/2}P(C)/q(C) = 0 \quad \text{a.s.}$$

Define the bounded continuous function φ on $l^\infty(\mathscr{C})$ by

$$\varphi(\psi) := \limsup_{\substack{P(C)\to 0\\ q(C)\to 0}} |\psi(C)|.$$

By our hypothesis,

$$\int \varphi(G_P/q)\, d\mathbb{P} = \lim_n \int \varphi(\nu_n/q)\, d\mathbb{P} = 0,$$

so $\varphi(G_P/q) = 0$ a.s. It follows from the fact that G_P is uniformly continuous, and convergent in probability to 0 as $P(C) \to 0$, that for each $\varepsilon > 0$,

$$\limsup_{\substack{P(C)\to 0\\ q(C)\ge \varepsilon}} |G_P(C)/q(C)| = 0 \quad \text{a.s.}$$

The lemma for $P(C) \to 0$ now follows. The proof as $P(C) \to 1$ is similar. □

PROPOSITION 5.2. *Let $\mathscr{C}$ be a P-pregaussian class of sets and q a real function on $\mathscr{C}$. If $G_P/q \in l^\infty(\mathscr{C})$ a.s. then*

$$\limsup_{P(C)\to 0} G_P(C)/q(C) = \theta \quad a.s.$$

for some finite constant θ.

PROOF. It suffices (by Lemma 5.1) to prove the same results for W_P. By Theorem III.3.5 of Jain and Marcus (1978), we have the representation (at least in law) of W_P as a uniformly convergent series:

$$W_P(C) = \sum_{j\ge 1} \varphi_j(C)\xi_j,$$

where (ξ_j) is an i.i.d. normal $N(0,1)$ sequence and $\varphi_j(C) = \langle 1_C, g_j\rangle_{L^2(P)}$ for some $g_j \in L^2(P)$. Therefore by Lemma 5.1,

$$|\varphi_j(C)/q(C)| \le \|g_j\|_{L^2(P)} P(C)^{1/2}/|q(C)|$$
$$\to 0 \quad \text{as } P(C) \to 0.$$

It follows that $\limsup_{P(C)\to 0} W_P(C)/q(C)$ is a tail r.v. of the sequence (ξ_j), hence is constant a.s. by the Kolmogorov 0–1 law. □

Proof of Proposition 3.3. First observe that the obvious analogs of Lemma 5.1 and Proposition 5.2 as $P(C) \to 1$ are also valid.

(i) ⇔ (iii) by definition. Suppose (iii) holds. Then $r_P(C, \phi) \to 0$ as $P(C) \to 0$ or 1 by Lemma 5.1, so G_P/q approaches a limit a.s. as $P(C) \to 0$ or 1. By Proposition 5.2 and symmetry of G_P, this limit must a.s. be 0. Further, boundedness a.s. of G_P/q implies q bounded away from 0 on $\mathscr{C}(\varepsilon, 1-\varepsilon)$ for all $\varepsilon > 0$. Thus (iii) ⇒ (ii). Similarly, (iv) ⇒ (ii).

Suppose (ii) holds. Then clearly $G_P/q \in l^\infty(\mathscr{C})$ a.s. Uniform r_P-continuity on $\mathscr{C}(\varepsilon, 1-\varepsilon)$ for all $\varepsilon > 0$ is also clear. If $r_P(C_n, C) \to 0$ and $P(C) = 0$ or 1, then $P(C_n) \to 0$ or 1 or $q(C_n) \to \infty$, so $G_P(C_n)/q(C_n) \to 0 = G_P(C)/q(C)$. Thus G_P/q has r_P-continuous sample paths, and (iii) holds.

Suppose q is uniformly d_P-continuous on $\{C \in \mathscr{C}\colon 0 < P(C) < 1\}$. If (iii) holds then using (ii) and Lemma 5.1 we see that r_P is uniformly d_P-continuous. Thus (iii) ⇒ (iv). □

It remains to prove Theorem 3.5. In establishing necessity of (3.7), the following will be critical.

Proposition 5.3. *Let $\mathscr{C}$ and φ be as in Theorem 3.5. If $\bar{\varphi} > 0$ on $(0, \frac{1}{2}]$ and*

(5.2) $$G_P(C)/\varphi(P(C)) \to 0 \quad a.s.\ as\ P(C) \to 0,$$

then

(5.3) $$\int_0^{1/2} e^{-\varepsilon\bar{\varphi}^2(s)/s}\, ds/s < \infty \quad \text{for all } \varepsilon > 0.$$

Proof. We may assume $\varphi(0) = 0$ and that $\mathscr{C}$ contains sets of arbitrarily small positive probability.

Suppose the integral in (5.3) is infinite for some $\varepsilon > 0$. We will show that there is a sequence (C_m) in $\mathscr{C}$ and a constant $\theta > 0$ such that

(5.4) $$P(C_{m+1}) \le P(C_m)/3$$

and

(5.5) $$\sum_{m \ge 1} \exp\left(-\theta\varphi^2(P(C_m))/P(C_m)\right) = \infty.$$

Observe that by Lemma 5.1,

(5.6) $$\bar{\varphi}^2(s)/s \to \infty \quad \text{as } s \to 0.$$

Let

$$E := \{P(C): C \in \mathscr{C}\}.$$

Then we can write

$$(0,1/2) \setminus \overline{E} = \bigcup_{0 \le n < N} (r_n, v_n),$$

a union of disjoint intervals, for some $0 \le N \le \infty$. Let $I := \{n \ge 0: r_n \le v_n/3\}$, $M := \text{card}(I)$, and $S := \bigcup_{n \in I}(r_n, v_n)$. We can write $S = \bigcup_{0 \le m < M}(a_m, b_m)$, again a union of disjoint intervals, with $b_0 > b_1 > \cdots$ and $M \le \infty$. Let

$$\sigma := \inf(S)$$

with $\inf(\phi)$ interpreted as $\frac{1}{2}$. Then either $M = \infty$ or $\sigma > 0$.

CASE 1. $M = \infty$. Then there exists $\varepsilon > 0$ such that

$$\begin{aligned} \infty &= \int_0^{b_0} \exp(-\varepsilon\overline{\varphi}^2(s)/s)\, ds/s \\ &\le \sum_{m \ge 0} \int_{a_m}^{b_m} e^{-\varepsilon\overline{\varphi}^2(b_m)/s}\, ds/s + \sum_{m \ge 0} \int_{b_{m+1}}^{a_m} e^{-\varepsilon\overline{\varphi}^2(s)/s}\, ds/s. \end{aligned} \tag{5.7}$$

Thus we get two subcases.

CASE 1a. The first sum on the right side of (5.7) is infinite. Then by (5.7), if m_0 is large,

$$\begin{aligned} \infty &= \sum_{m \ge m_0} \int_0^{b_m} \exp(-\varepsilon\overline{\varphi}ay^2(b_m)/s)\, ds/s \\ &\le \sum_{m \ge m_0} (b_m/\varepsilon\overline{\varphi}^2(b_m))\exp(-\varepsilon\overline{\varphi}^2(b_m)/b_m) \\ &\le \sum_{m \ge m_0} \exp(-\varepsilon\overline{\varphi}^2(b_m)/b_m). \end{aligned} \tag{5.8}$$

Now for each $m \ge 1$ there is a set $C_m \in \mathscr{C}$ with $b_m \le P(C_m) \le b_{m-1}/3$ and $\varphi^2(P(C_m))/P(C_m) \le 2\overline{\varphi}^2(b_m)/b_m$. (5.4) and (5.5), with $\theta = \varepsilon/2$, now follow from (5.8).

CASE 1b. The second sum on the right of (5.7) is infinite. For $m \ge 0$ define

$$V_m := \begin{cases} [b_{m+1}, a_m) \cap \{3^{2l}b_{m+1}: l \ge 0\}, & \text{if } b_{m+1} < a_m, \\ \{b_{m+1}\}, & \text{if } b_{m+1} = a_m, \end{cases}$$

and let $\{s_k: k \ge 0\}$ be the set $\bigcup_{m \ge 0} V_m$ in decreasing order. Then

$$\begin{aligned} \infty &= \sum_{m \ge m_0} \sum_{s_k \in [b_{m+1}, a_m]} \int_{s_k}^{9s_k} \exp(-\varepsilon\overline{\varphi}^2(s)/s)\, ds/s \\ &\le \sum_{k \ge 0} 8\exp(-\varepsilon\overline{\varphi}^2(s_k)/9s_k). \end{aligned} \tag{5.9}$$

Now $\inf([s_k, \infty) \cap S) < 3s_k$ for all $k \geq 1$, so we can find $C_k \in \mathscr{C}$ with $s_k \leq P(C_k) < 3s_k$, $\varphi(P(C_k)) \leq 2\bar{\varphi}(s_k)$, and $P(C_k)$ in the same interval $[b_{m+1}, a_m]$ as s_k. Then (5.4) holds, and $\varphi^2(P(C_k))/P(C_k) \leq 4\bar{\varphi}^2(s_k)/s_k$, so (5.5), with $\theta = \varepsilon/36$, follows from (5.9).

CASE 2. $\sigma > 0$. Let $s_m := \sigma/9^m$. For some $\varepsilon > 0$ we have

$$\begin{aligned}\infty &= \sum_{m\geq 0} \int_{s_{m+1}}^{s_m} \exp\left(-\varepsilon\varphi^2(s)/s\right) ds/s \\ &\leq \sum_{m\geq 0} (s_m - s_{m+1}) s_{m+1}^{-1} \exp\left(-\varepsilon\varphi^2(s_m)/9s_m\right) \\ &= 8 \sum_{m\geq 0} \exp\left(-\varepsilon\varphi^2(s_m)/9s_m\right).\end{aligned}$$

As in Case 1b we have $\inf([s_m, \infty) \cap S) < 3s_m$, so (5.4) and (5.5) follow similarly.

Thus we can satisfy (5.4) and (5.5) in all cases. Define

$$D_m := C_m \setminus \bigcup_{j>m} C_j, \qquad m \geq 1,$$

so the D_m are disjoint, and by (5.4),

$$P(D_m) \geq P(C_m)/4.$$

Using (5.6) we have for large m,

$$\begin{aligned}\mathbb{P}\left[\frac{|W_P(D_m)|}{\varphi(P(C_m))} > \theta^{1/2}/2\right] &\geq \mathbb{P}\left[\frac{|W_P(D_m)|}{P(D_m)^{1/2}} > \frac{\theta^{1/2}\varphi(P(C_m))}{P(C_m)^{1/2}}\right] \\ &\geq \exp\left(-\frac{\theta\varphi^2(P(C_m))}{P(C_m)}\right).\end{aligned}$$

Since the D_m are independent, it follows from this and (5.5) that

$$\mathbb{P}\left[\frac{|W_P(D_m)|}{\varphi(P(C_m))} > \theta^{1/2}/2 \text{ i.o.}\right] = 1.$$

We can therefore define τ_j, $j \geq 1$, to be the jth index m such that $|W_P(D_m)|/\varphi(P(C_m)) > \theta^{1/2}/2$. Then $[\tau_j = k]$ is an event in the σ-algebra generated by $W_P(D_1), \dots, W_P(D_k)$ for all $k \geq 1$, so is independent of $W_P(C_k \setminus D_k)$, since W_P is independent on disjoint sets. Fix $\lambda > 1$. Now $W_P(C_m) = W_P(D_m) + W_P(C_m \setminus D_m)$ a.s. for all m, so using (5.6), for large k we have

$$\begin{aligned}\mathbb{P}\left[\frac{|W_P(C_{\tau_j})|}{\varphi(P(C_{\tau_j}))} > \frac{\theta^{1/2}}{4} \middle| \tau_j = k\right] &\leq \mathbb{P}\left[\frac{|W_P(C_k \setminus D_k)|}{\varphi(P(C_k))} \leq \frac{\theta^{1/2}}{4} \middle| \tau_j = k\right] \\ &= \mathbb{P}\left[|W_P(C_k \setminus D_k)| \leq \theta^{1/2}\varphi(P(C_k))/4\right] \\ &\geq \mathbb{P}\left[|W_P(C_k \setminus D_k)| \leq \lambda P(C_k \setminus D_k)^{1/2}\right] \\ &\geq 1 - \lambda^{-2}.\end{aligned}$$

Since $\tau_j \geq j$ this shows

$$\mathbb{P}\big[|W_P(C_m)|/\varphi\big(P(C_m)\big) > \theta^{1/2}/4 \text{ i.o.}\big] = 1.$$

By (5.6) the same holds for G_P, which contradicts (5.2). □

PROOF OF THEOREM 3.5. Let h be a nonnegative function on $(0, \frac{1}{2}]$ and define

$$\eta(s) := \big(2s\{V(\mathscr{C})L\big(s^{-1}a_P(s)\big) + h(s)\}\big)^{1/2}.$$

The proof of Theorem 1.2 of Alexander (1986) shows that if $sh(s) \sim m(s)$ as $s \to 0$ for some positive nondecreasing function m and

$$\sum_{j \geq 0} \exp\big(-h(\mu^j s)\big) \to 0 \quad \text{as } s \to 0 \text{ for all } 0 < \mu < 1,$$

then

$$\limsup_{P(C) \to 0} |W_P(C)|/\eta\big(P(C)\big) \leq 1 \quad \text{a.s.}$$

(Note that this gives (3.4) above when $h(s)$ is LLs^{-1}.)

Suppose (3.6) and (3.7) hold. Fix $\delta > 0$ and define $h(s)$ by

$$\delta\overline{\varphi}(s) = \big(2s\{V(\mathscr{C})L\big(s^{-1}a_P(s)\big) + h(s)\}\big)^{1/2}.$$

From (3.6) it follows that $(sL(s^{-1}a_P(s)))^{1/2} = o(\overline{\varphi}(s))$ as $s \to 0$, so $sh(s) \sim \frac{1}{2}\delta^2\overline{\varphi}^2(s)$. Hence for small s,

$$h(s) \geq \delta^2\overline{\varphi}^2(s)/4s.$$

Therefore, since δ is arbitrary, we will have

$$\limsup_{P(C) \to 0} |W_P(C)|/\varphi\big(P(C)\big) = 0 \quad \text{a.s.,} \tag{5.10}$$

provided we can show

$$\sum_{j \geq 0} \exp\big(-\delta\overline{\varphi}^2(\mu^j s)/\mu^j s\big) \to 0 \quad \text{as } s \to 0 \text{ for all } \delta > 0 \text{ and } 0 < \mu < 1. \tag{5.11}$$

But

$$\begin{aligned}\sum_{j \geq 0} \exp\big(-\delta\overline{\varphi}^2(\mu^j s)/\mu^j s\big) &\leq \sum_{j \geq 0} (1-\mu)^{-1} \int_{\mu^{j+1}s}^{\mu^j s} \exp\big(-\delta\mu\overline{\varphi}^2(r)/r\big)\, dr/r \\ &= (1-\mu)^{-1} \int_0^s \exp\big(-\delta\mu\overline{\varphi}^2(r)/r\big)\, dr/r,\end{aligned}$$

and so (5.11) follows from (3.7).

By (3.6), (5.10) also holds for G_P. Since (3.2), with $q = \varphi \circ P$, is equivalent to (3.6), weak convergence now follows from Theorem 3.4.

Conversely if the weak convergence holds then so do (3.2) and (3.3) for $q = \varphi \circ P$, so (3.7) follows from Proposition 5.3 and (3.6) from (3.2). □

REFERENCES

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

ALEXANDER, K. S. (1984b). Rates of growth and sample moduli for weighted empirical processes indexed by sets. To appear in *Probab. Theory Rel. Fields.*

ALEXANDER, K. S. (1984c). Central limit theorems for stochastic processes under random entropy conditions. To appear in *Probab. Theory Rel. Fields.*

ALEXANDER, K. S. (1985). The central limit theorem for weighted empirical processes indexed by sets. Unpublished manuscript.

ALEXANDER, K. S. (1986). Sample moduli for set-indexed Gaussian processes. *Ann. Probab.* **14** 598–611.

ALEXANDER, K. S. and PYKE, R. (1986). A uniform central limit theorem for set-indexed partial-sum processes with finite variance. *Ann. Probab.* **14** 582–597.

BASS, R. F. (1985). Law of the iterated logarithm for set-indexed partial sum processes with finite variance. *Z. Wahrsch. verw. Gebiete* **70** 591–608.

BASS, R. F. and PYKE, R. (1985). The space $D(A)$ and weak convergence for set-indexed processes. *Ann. Probab.* **13** 860–884.

CHANG, L.-C. (1964). On the ratio of the empirical distribution function to the theoretical distribution function. *Selected Transl. Math. Statist. Probab.* **4** 17–38.

CIBISOV, D. M. (1964). Some theorems on the limiting behavior of empirical distribution functions. *Selected Transl. Math. Statist. Probab.* **6** 147–156.

CSÖRGŐ, M., CSÖRGŐ, S., HORVÁTH, L. and MASON, D. M. (1986). *An Asymptotic Theory for Empirical Reliability and Concentration Processes. Lecture Notes in Statist.* **33**. Springer, New York.

DUDLEY, R. M. (1966). Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10** 109–126.

DUDLEY, R. M. (1967). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Funct. Anal.* **1** 290–330.

DUDLEY, R. M. (1973). Sample functions of the Gaussian process. *Ann. Probab.* **1** 66–103.

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

DUDLEY, R. M. (1981). Donsker classes of functions. *Statistics and Related Topics* (Proc. Symp. Ottawa, 1980) 341–352. North-Holland, New York.

DUDLEY, R. M. (1984). A course on empirical processes. *Ecole d'Été de Probabilités de Saint-Flour, XII—1982. Lecture Notes in Math.* **1097** 1–142. Springer, New York.

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

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.

DURST, M. and DUDLEY, R. M. (1981). Empirical processes, Vapnik–Červonenkis classes and Poisson processes. *Probab. Math. Statist.* **1** 109–115.

FERNIQUE, X. (1974). Regularité des trajectories des functions aléatoires gaussiennes. *Ecole d'Été de Probabilités de Saint-Flour IV—1974. Lecture Notes in Math.* **480** 1–96. Springer, Berlin.

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

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

HOFFMANN-JØRGENSEN, J. (1986). Stochastic processes on Polish spaces. Unpublished manuscript.

JAIN, N. C. and MARCUS, M. B. (1978). Continuity of subgaussian processes. In *Advances in Probability* (J. Kuelbs, ed.) **4** 1–68. Dekker, New York.

KALINAUSKAITE, N. (1979). Local sample behavior of random fields with independent increments. *Lithuanian Math. J.* **19** 60–67. (Original in Russian in *Litovsk. Mat. Sb.* **19** 87–96.)

O'REILLY, N. (1974). On the convergence of empirical processes in sup-norm metrics. *Ann. Probab.* **2** 642–651.

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

POLLARD, D. (1982b). Rates of strong uniform convergence. Unpublished manuscript.

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

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

WICHURA, M. J. (1969). Inequalities with applications to the weak convergence of random processes with multi-dimensional time parameters. *Ann. Math. Statist.* **40** 681–687.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
LOS ANGELES, CALIFORNIA 90089-1113