The Annals of Probability
1984, Vol. 12, No. 4, 929–989

SPECIAL INVITED PAPER

SOME LIMIT THEOREMS FOR EMPIRICAL PROCESSES[1]

BY EVARIST GINÉ AND JOEL ZINN

Texas A&M University

In this paper we provide a general framework for the study of the central limit theorem (CLT) for empirical processes indexed by uniformly bounded families of functions $\mathscr{F}$. From this we obtain essentially all known results for the CLT in this case; we improve Dudley's (1982) theorem on entropy with bracketing and Kolčinskii's (1981) CLT under random entropy conditions. One of our main results is that a combinatorial condition together with the existence of the limiting Gaussian process are necessary and sufficient for the CLT for a class of sets (modulo a measurability condition). The case of unbounded $\mathscr{F}$ is also considered; a general CLT as well as necessary and sufficient conditions for the law of large numbers are obtained in this case. The results for empiricals also yield some new CLT's in $C[0, 1]$ and $D[0, 1]$.

1. Introduction. Let $(S, \mathscr{S})$ be a measurable space and let P be a probability measure on $(S, \mathscr{S})$. Let $\{X_i\}_{i=1}^{\infty}$ be independent, identically distributed (i.i.d.) S-valued random variables with common law P. The empirical distributions P_n corresponding to the sequence $\{X_i\}_{i=1}^{\infty}$ are defined as the random

Received June 1983; revised May 1984.
[1] Research partially supported by NSF grants no. MCS-81-00728 and no. MCS-83-01367.
AMS 1980 *subject classifications*. Primary 60F17, 60B12; secondary 60F05, 62E20.
Key words and phrases. Central limit theorems, empirical processes, functional Donsker classes, Gaussian processes, metric entropy, laws of large numbers.

measures

$$P_n = (1/n) \sum_{i=1}^n \delta_{X_i},$$

where δ_x denotes point mass at $x \in S$. One of the most persistent problems in probability and statistics is that of recovering the law P from the random laws P_n. A subject of particular interest in this connection is that of the convergence of P_n to P uniformly over classes of sets or functions (a.s. convergence to zero of $P_n - P$ and weak convergence of $n^{1/2}(P_n - P)$ to a Gaussian process), the pioneering results in this area being the celebrated Glivenko–Cantelli and Kolmogorov–Smirnov theorems. A strong new impetus was given to this question by the works of Vapnik and Červonenkis (1971) on the weak law of large numbers and Dudley (1978) on the central limit theorem. (For a survey of recent work on empirical processes see Gaenssler and Stute, 1979.)

Modulo measurability considerations, Vapnik and Červonenkis characterize the families of sets $\mathscr{C}$ for which

$$\sup_{A\in\mathscr{C}} |P_n(A) - P(A)| \to 0 \quad \text{in pr.} \quad \text{(in fact a.s.)}$$

as those families of sets satisfying

$$(1/n)\ln \Delta^{\mathscr{C}}(X_1, \cdots, X_n) \to 0 \quad \text{in pr.},$$

where $\Delta^{\mathscr{C}}$ is the number of distinct sets of the form $C \cap \{X_1, \cdots, X_n\}$, $C \in \mathscr{C}$. Dudley proved weak convergence of

$$\text{“}\mathscr{L}\{n^{1/2}(P_n(C) - P(C)) : C \in \mathscr{C}\}\text{”}$$

to a Gaussian process indexed by the sets in $\mathscr{C}$ under several different conditions on $\mathscr{C}$, in particular, under some measurability hypotheses, in the case of polynomial growth (in n) of $\sup_{\{x_1,\cdots,x_n\}\subset S}\Delta^{\mathscr{C}}(x_1, \cdots, x_n)$ (Vapnik–Červonenkis classes). The methods of proof for the law of large numbers were mostly combinatorial, whereas the central limit theorems (CLT) were often proved using “metric entropy” techniques (see, e.g., Dudley, 1978, and references therein).

In this article we continue the study of the CLT for the empirical process

$$\left\{\int f\, d(P_n - P) : f \in \mathscr{F}\right\}$$

particularly in the case of a class $\mathscr{F}$ of measurable functions on S such that $\|f\|_\infty \le M < \infty$ for all $f \in \mathscr{F}$, although we also consider the case of unbounded $\mathscr{F}$ in connection with both the CLT and the LLN. We give general results which contain and unify most of the known CLT's for empirical processes, improve some of these and consider new situations. Our results for classes of sets $\mathscr{C}$ are rather complete. We obtain “random combinatorial” conditions which together with the existence of the limiting Gaussian process are necessary and sufficient for the class $\mathscr{C}$ to satisfy the CLT. (This last result may be considered as the analogue for the CLT of the Vapnik–Červonenkis law of large numbers for classes of sets.) Then we show that the random combinatorial quantities can be estimated with enough precision to obtain all the general CLT's (for sets) which are known to us. In addition to drawing on the work of Dudley (1978, 1982, etc.) and

Le Cam (1983b), we make extensive use of the theory of Gaussian and subgaussian processes.

Now we describe the contents of this paper. In Section 2 we collect basic prerequisites. The measurability hypotheses needed throughout are described. Randomization, one of our basic tools, is discussed: using ideas in Alexander (1982) and Pisier and Fernique (private communication), we show that $\sum_{i=1}^n(\delta_{X_i} - P)$ can be replaced by $\sum_{i=1}^n \varepsilon_i\delta_{X_i}$ and $\sum_{i=1}^n g_i\delta_{X_i}$ both in the CLT and in the law of large numbers, and to deal with both necessary and sufficient conditions for these theorems. Here $\{\varepsilon_i\}$ is a Rademacher sequence and $\{g_i\}$ is an orthogaussian sequence, and both are independent of the processes $\{f(X_i), f \in \mathscr{F}\}$. (These are standard practices for proving CLT's in Banach spaces; they originated in Kahane (1968, page 7: principle of reduction), and were first applied by Jain and Marcus (1975); see also Pisier (1975), Marcus (1978, 1981), Giné and Marcus (1980), Pollard (1981), Giné and Zinn (1983, Corollary 2.7) and Marcus and Pisier (1983).) We then explain exactly what we mean by CLT and state Dudley's equicontinuity condition, which is the starting point of our investigation. After the work of Dudley and Philipp (1983) by CLT we really mean an invariance principle in probability. Namely, that there exist i.i.d. centered Gaussian processes $Y_i(f), f \in \mathscr{F}$, with the covariance of $\delta_{X_i}(f)$, whose sample paths are bounded and uniformly continuous on $(\mathscr{F}, L_2(P)\text{-dist.})$ and such that the random quantities

$$n^{-1/2}\max_{k\le n}\sup_{f\in\mathscr{F}}\left|\sum_{i=1}^k\left(f(X_i) - \int f\,dP - Y_i(f)\right)\right|$$

converge to zero in Pr*. If this holds, $\mathscr{F}$ is called a functional P-Donsker class of functions (See Section 2(b)). (Under measurability hypotheses, this implies the weak convergence to $\mathscr{L}(Y_1)$ of the processes $n^{-1/2}\sum_{i=1}^n(f(X_i) - \int f\,dP)$, $f \in \mathscr{F}$, as random variables in $\ell^\infty(\mathscr{F})$.) For this convergence to take place it is necessary and sufficient that $\mathscr{F}$ both be totally bounded for $\rho_P^2(f, g) := \int (f - g)^2\,dP - (\int (f - g)\,dP)^2$ and that it satisfy the "asymptotic uniform equicontinuity" condition: for all $\varepsilon > 0$

$$\text{(1.1)}\quad \lim_{\delta\downarrow 0}\limsup_n \Pr{}^*\{\sup_{f,g\in\mathscr{F},\rho_P(f,g)\le\delta} n^{-1/2}\,|\,(P_n - P)(f - g)\,| > \varepsilon\} = 0$$

(Dudley and Philipp, 1983, Dudley, 1982). (Here and in what follows, if μ is a measure we write $\mu(f)$ for $\int f\,d\mu$). We show how (1.1) can be replaced by a "randomized" equicontinuity condition, with either $\{\varepsilon_i\}$ or $\{g_i\}$, Pr* or E^* (outer probability, upper integral: see Section 2). This result is useful later in the paper. In the next subsection we state all the basic results on Guassian processes that are used throughout (which are due to Dudley, Fernique, Marcus, Pisier, Shepp, Slepian and Sudakov). Finally, the last part of Section 2 is devoted to random distances

$$d_{n,p}(f, g) = (\textstyle\sum_{i=1}^n |\,f(X_i) - g(X_i)\,|^p/n)^{1/(p\vee 1)},$$

their associated random covering numbers $N_{n,p}(\varepsilon, \mathscr{F})$ and entropies $\ln N_{n,p}$, the related combinatorial quantities $\Delta^{\mathscr{C}}(X_1, \cdots, X_n)$ introduced by Vapnik and Červonenkis, and their law of large numbers.

Section 3 gives the basic framework for the CLT in the bounded case. We say that $\mathscr{F}$ is P-pregaussian (or G_PBUC) if the centered Gaussian process on $(\mathscr{F}, \rho_P)$ which has the covariance of the process $f \to f(X)$, admits a version with bounded uniformly continuous paths. Using Gaussian randomization and the theory of Gaussian processes we show that (under measurability assumptions) $\mathscr{F}$ is a functional P-Donsker class if and only if both $\mathscr{F}$ is P-pregaussian and satisfies a "reduced" randomized equicontinuity condition. This condition is

$$(1.2)\quad \lim_{\delta\downarrow 0}\lim\sup_n \Pr{}^*\{\sup_{f,g\in\mathscr{F}, P(f-g)^2<\varepsilon/n^{1/2}} |\textstyle\sum_{i=1}^n \varepsilon_i(f(X_i) - g(X_i))/n^{1/2}| > \varepsilon\} = 0$$

for all $\varepsilon > 0$ (where one can replace $\{\varepsilon_i\}$ by $\{g_i\}$ (and there are equivalent statements with E^* instead of $\Pr^*$). Here $P(f-g)^2 = \int (f-g)^2\, dP$. This is one of our main results (Theorem 3.2). Measurability may in some cases be a nuisance but it can be circumvented if instead of $\mathscr{F}$ P-pregaussian one assumes the L_2-entropy condition

$$(1.3)\qquad \int_0^\infty (\ln N(\varepsilon, \mathscr{F}, e_p))^{1/2}\, d\varepsilon < \infty$$

where $e_P^2(f, g) = P(f-g)^2$. This result (Theorem 3.1) is strong enough for most of the applications but, since condition (1.3) is not necessary for $\mathscr{F}$ to satisfy the CLT, Theorem 3.2 is qualitatively superior. We remark that the proof of the main result (Theorem 3.2) does not make use of the typical "chaining argument" associated with entropy (so commonly used in the CLT for empiricals), but rather a recent result of Fernique (1983) on comparison of Gaussian processes for sufficiency, and Sudakov's "minoration" for necessity—both applied conditionally. On the contrary, Theorem 3.1 uses the chaining argument, but we stop at the level $e_P^2(f, g) = \varepsilon/n^{1/2}$ because below this level the tails of the variables $n^{1/2}(P_n - P)(f - g)$ are not necessarily Gaussian-like. One should compare Theorem 3.1 with Le Cam's (1983a, Lemma 2) which is a similar result, and also to some of the statements at the end of Kolčinskiĭ (1981), however our proof is extracted from the proof of Theorem 5.1 of Dudley (1978).

In Section 4 we give several applications of Theorem 3.1 with no measurability hypotheses. In particular we obtain an improvement of Dudley's (e.g. 1982) theorem on "metric entropy with bracketing". We also obtain a sufficient condition for the CLT for classes of sets involving a condition on the speed of convergence to zero of the probability that about $\tau\varepsilon n^{1/2}$ X_i's fall simultaneously in one of the sets $A \,\Delta\, B$, $A, B \in \mathscr{C}$, $P(A \,\Delta\, B) \le \varepsilon/n^{1/2}$. Then we apply this last result to obtain quick proofs of the CLT under "metric entropy with inclusion" (with an improvement), for VC classes and for sequences of sets (Dudley, 1978) as well as in other situations.

In Section 5, under measurability, we approach the problem of handling the reduced equicontinuity condition (1.2) in more elaborate ways. Below the level $e_P^2(f, g) = \varepsilon/n^{1/2}$ the tails of $n^{1/2}(P_n - P)(f - g)$ are not Gaussian-like as remarked above, but if we randomize they are conditionally Gaussian-like, and we can either apply a simple form of the entropy argument (we apply a suitable modification of a "trick" of Le Cam (1983b) at a crucial step—see Lemma 5.2) or the ready-made upper estimates of Dudley (1973) and Marcus and Pisier (1978). In the first case the resulting condition is a tail condition for the random entropy

$$\ln N_{n,1}(\varepsilon/n^{1/2}, \mathscr{F}'_{\varepsilon,n}) \quad \text{where} \quad \mathscr{F}'_{\varepsilon,n} = \{f - g: f, g, \in \mathscr{F}, e_P^2(f, g) \leq \varepsilon/n^{1/2}\}$$

which in the case of families of sets turns out to be also *necessary* (Theorem 5.7) because of Sudakov's minoration. Theorem 5.7 and its variations is one of our main results.

In Section 6 we relate random entropies to the combinatorial quantities of Vapnik and Červonenkis and show how they can be computed in many cases. As examples, we obtain new results on metric entropy with inclusion, sequences of sets and the CLT for discrete distributions (the Borisov-Durst-Dudley theorem; see e.g. Dudley, 1982).

Section 7 contains applications of Section 5 to the CLT for uniformly bounded random variables (processes) taking values in $C[0, 1]$ and $D[0, 1]$. The results obtained do not seem to follow from known general theorems (such as the Jain–Marcus, 1975, CLT or the extensions by Hahn, 1978, and Pisier, 1980, of the CLT under the Kolmogorov conditions).

If the class $\mathscr{F}$ is not uniformly bounded not only does Bernstein's inequality fail to apply (at least directly) but neither does the other main ingredient of the proofs in Section 5 (Lemma 5.2). However, a direct application of any of the above randomizations to (1.1) makes it possible to use the standard Gaussian "majoration" conditionally. This gives a sufficient condition for the CLT in terms of the random entropy $N_{n,2}(\varepsilon, \mathscr{F})$. Sudakov's minoration gives a necessary condition which differs from the sufficient one by only "a log". The same techniques apply to the law of large numbers, in which case we obtain necessary and sufficient conditions (modulo measurability) in terms of $N_{n,p}(\varepsilon, \mathscr{F})$, $p \in (0, \infty]$.

For more concrete applications of the results of this article, we refer to Dudley (1978, 1982) and references there, Le Cam (1983a), Pollard (1979, 1980), Yatracos (1983), and also to the forthcoming books by Le Cam and by Pollard. Modulo measurability, all the CLT's for empirical processes that we are aware of have either been surveyed in this article, or follow from those surveyed here, with two exceptions: the CLT for the empirical characteristic function (Marcus, 1981) for which our results do not provide a significant simplification, and Dudley's (e.g. 1982) theorem on entropy with bracketing in the unbounded case.

2. Basic preliminaries.

In this section we collect several definitions and propositions used throughout.

(a) *Measurability and randomization.* Since suprema over uncountable families need not be measurable, some measurability assumptions are needed, and at times we will use outer measures and upper integrals, which are defined as follows: if $(A, \mathscr{A}, \mu)$ is a μ-complete probability space, then for every $E \subset A$,

$\mu^*(E) = \sup\{\mu(B): E \subset B, B \in \mathscr{A}\}$ and for every

$$f: A \to [0, \infty\}, \quad E^*f = \inf\left\{\int g\, d\mu: g \geq f,\, g\ \mathscr{F}\text{-measurable}\right\}.$$

We have chosen the following conditions, which are based on those in Alexander (1983). See Dudley (1978, 1982) for other approaches to the measurability problem.

First we give some notation used throughout. $(S, \mathscr{S})$ is a measurable space and P is a probability measure on $(S, \mathscr{S})$. $\mathscr{L}_p(P)$, $0 \leq p < \infty$, will denote the set of functions $f: S \to \mathbb{R}$ which are measurable and such that $\int |f|^p\, dP < \infty$ if $0 < p < \infty$. We let

$$Pf := \int f\, dP \quad \text{for all} \quad f \in \mathscr{L}_1(P),$$

$$\rho_P^2(f, g) = P(f-g)^2 - (P(f-g))^2,$$

$$e_P^2(f, g) = P(f-g)^2, \quad f, g \in \mathscr{L}_2(P).$$

For a class $\mathscr{F} \subset \mathscr{L}_2(P)$ and $\delta > 0$, we let

$$\mathscr{F}' = \{f - g: f, g \in \mathscr{F}\},$$

$$\mathscr{F}'_\delta = \{f - g: f, g \in \mathscr{F}, e_P^2(f, g) \leq \delta\}.$$

For later use we also define

$$\mathscr{F}'_{\varepsilon,n} := \mathscr{F}'_{\varepsilon/n^{1/2}}$$

In general we define, for any real function h on a set T,

$$\| h \|_T = \sup_{t \in T} | h(t) |.$$

Often our parameter set is a set of functions $\mathscr{G}$ and in this case, since it is more convenient to write $f(X)$ rather than $X(f)$ or even $\delta_X(f)$, we use notation such as the following: for $a_i, b \in \mathbb{R}$,

$$\| \textstyle\sum_i a_i f(X_i) - bPf \|_{\mathscr{G}} = \sup_{f \in \mathscr{G}} | \textstyle\sum_i a_i f(X_i) - bPf |$$

and

$$\| \textstyle\sum_i a_i f^2(X_i) \|_{\mathscr{G}} = \sup_{f \in \mathscr{G}} | \textstyle\sum_i a_i f^2(X_i) |.$$

We use the same notation for classes of sets $\mathscr{C}$, which we identify with their indicators.

In this paper we use the following general setup. $(S, \mathscr{S}, P)$ is a probability space and $\{X_i\}_{i=1}^\infty$ are the coordinate functions on $(S^{\mathbb{N}}, \mathscr{S}^{\mathbb{N}}, P^{\mathbb{N}})$. For a class $\mathscr{G}$ of real measurable functions on S consider the following statements.

(a) the quantities

$$\sup_{f \in \mathscr{G}} \{\textstyle\sum_{i=1}^n a_i f(X_i) - bPf\}, \quad a_i, b \in \mathbb{R}, \quad n \in \mathbb{N}$$

are $P^{\mathbb{N}}$-completion measurable;

(b) the quantities

$$\sup_{f\in\mathscr{S}}\{(\sum_{i=1}^{n} a_i f^2(X_i))^{1/2} - (\sum_{i=n+1}^{2n} a_i f^2(X_i))^{1/2}\}, \quad a_i \geq 0, \quad n \in \mathbb{N}$$

are $P^{\mathbb{N}}$-completion measurable.

The following definitions are close to those in Alexander (1983).

2.1. DEFINITION. $\mathscr{S}$ is *linearly supremum measurable for P* (LSM(P)) if (a) holds for $\mathscr{S}$ and moreover $\sup_{f\in\mathscr{S}} |f(s)| < \infty$ for all $s \in S$. $\mathscr{S}$ is *ℓ_2-supremum measurable for P* (SM(P)) if $\mathscr{S}$ is LSM(P) and satisfies (b).

2.2. DEFINITION. $\mathscr{S}$ is *ℓ_2-deviation measurable* (resp., *linearly deviation measurable*) *for P* if both $\mathscr{S}$ and $\mathscr{S}'_\delta$ for all $\delta > 0$ are SM(P) (resp., LSM(P)).

2.3. DEFINITION. $\mathscr{S}$ is *nearly* supremum measurable (NSM(P)) (resp. *nearly* linearly supremum measurable (NLSM(P))) for P if there exists a ℓ_2-supremum measurable (resp. linearly . . .) ${}_0\mathscr{S} \subset \mathscr{S}$ such that for all $n > 0$

$$\Pr{}^*\{\sup_{f\in\mathscr{S}} |\nu_n(f)| \neq \sup_{f\in{}_0\mathscr{S}} |\nu_n(f)|\} = 0.$$

where $\nu_n(f) = n^{-1/2} \sum_{i=1}^{n} (f(X_i) - Pf)$. $\mathscr{S}$ is *nearly* ℓ_2-deviation measurable (NDM(P)) (resp., *nearly* linearly deviation measurable (NLDM(P))) if there exists an ℓ_2-deviation measurable (resp., linearly deviation measurable) class ${}_0\mathscr{S} \subset \mathscr{S}$ such that for every $\delta > 0$ and $n \geq 1$

$$\Pr{}^*\{\sup_{f\in\mathscr{S}'_\delta} |\nu_n(f)| \neq \sup_{f\in{}_0\mathscr{S}'_\delta} |\nu_n(f)|\} = 0.$$

2.4. REMARKS. (1) Definition 2.3 is introduced in order to include the case considered by Pollard (1982), i.e. that the processes $f \to P_n f$, $n \in \mathbb{N}$, $f \in (\mathscr{F}, e_p)$, be stochastically separable. In this case, $\mathscr{F}$ is NDM and ${}_0\mathscr{F} \subset \mathscr{F}$ is countable.

(2) Let $\{\xi_i\}$ be real random variables defined on a probability space (Ω, Σ, Q). Then if $\mathscr{S}$ is ℓ_2-supremum measurable, the quantities

(a) $$\sup_{f\in\mathscr{S}}\{\sum_{i=1}^{n} \xi_i(f(X_i) - bPf)\}, \quad n \in \mathbb{N}$$

and

(b) $$\sup_{f\in\mathscr{S}}\{(\sum_{i=1}^{n} \xi_i^+ f^2(X_i))^{1/2} - (\sum_{i=1}^{n} \xi_i^- f^2(X_i))^{1/2}\}, \quad n \in \mathbb{N},$$

are each jointly measurable in the product probability space $(\Omega \otimes S^{\mathbb{N}}, \Sigma \otimes \mathscr{S}^{\mathbb{N}}, Q \otimes P^{\mathbb{N}})$ since they are continuous in the coefficients ξ_i, ξ_i^+, ξ_i^-. (Similarly if $\mathscr{S}$ is linearly supremum measurable then the quantities in (a) are measurable.) Usually $\{\xi_i\}$ will either be a Rademacher sequence or an orthogaussian sequence. A sequence $\{\varepsilon_i\}$ is Rademacher if the ε_i are i.i.d. and $Q\{\varepsilon_i = 1\} = Q\{\varepsilon_i = -1\} = \frac{1}{2}$, and a sequence $\{g_i\}$ is orthogaussian if the g_i are i.i.d. $N(0, 1)$. This is the context in which we always work even if we do not explicitly mention it. And in this context E_ξ (hence E_ε or E_g) denotes integration with respect to Q and E_X denotes integration with respect to $P^{\mathbb{N}}$. Also at times we use $P_\xi(P_\varepsilon, P_g)$ for Q and P_X for $P^{\mathbb{N}}$.

(3) Since $P^{\mathbb{N}}$ is invariant under permutations, if $\mathcal{G}$ is supremum measurable then

$$\mathcal{L}(\| \textstyle\sum_{i=1}^n \varepsilon_i(f(X_i) - f(X_{n+i})) \|_{\mathcal{G}}) = \mathcal{L}(\| \textstyle\sum_{i=1}^n (f(X_i) - f(X_{n+i})) \|_{\mathcal{G}}).$$

This remark is crucial in what follows.

The following lemma (Vapnik and Červonenkis, 1971, Lemma 2; Alexander, 1983, Lemma 3.2) bounds tail probabilities of $\| X \|$ by tail probabilities of $\| X - X' \|$ where X and X' are independent but not necessarily equidistributed.

2.5. LEMMA. *Let T be an index set and let $X(t)$, $X'(t)$, $t \in T$, be two indexed collections of random variables which are defined on $(\Omega \times \Omega', \Sigma \times \Sigma', \Pr = P \times P')$, X depending only on $\omega \in \Omega$ and X' on $\omega' \in \Omega'$. Then for all $s > 0$ and all $u > 0$ for which $\sup_{t\in T}\Pr\{|X'(t)| \geq u\} < 1$,*

$$\begin{aligned}(2.1)\qquad &\Pr^*\{\| X \|_T > s\} \\ &\quad \leq [1 - \sup_{t\in T}\Pr^*\{|X'(t)| \geq u\}]^{-1}\Pr^*\{\| X - X' \|_T > s - u\}.\end{aligned}$$

PROOF. We have

$$\begin{aligned}\Pr^*\{\| X - X' \|_T > s - u\} &\geq E_P^*(P')^*\{\| X - X' \|_T > s - u\} \\ &\geq P^*\{\| X \|_T > s\}\inf_{\{\omega:\| X(\omega) \| > s\}}(P')^*\{\| X(\omega) - X' \|_T > s - u\} \\ &\geq P^*\{\| X \|_T > s\}\inf_{t\in T}P'\{|X'(t)| < u\}\end{aligned}$$

as if $\| X(\omega) \| > s$ then $|X(\omega, t)| > s$ for some $t \in T$. □

Using Chebyshev's inequality, the previous lemma gives:

2.6. COROLLARY. *Under the hypotheses of Lemma 2.5, if $\theta > \sup_{t\in T}E(X'(t))^2$ then*

$$(2.2)\qquad P^*\{\| X \|_T > s\} \leq 2\Pr^*\{\| X - X' \|_T > s - (2\theta)^{1/2}\}, \quad s > 0.$$

In what follows the probability measure $Q \times P^{\mathbb{N}}$ of Remark 2.4 (2) will be denoted as Pr. In the proof of Lemma 2.7 below we use the following facts: $(Q \times P^{\mathbb{N}})^*(A) \geq E_Q^*(P^{\mathbb{N}})^*(A)$, $(Q \times P^{\mathbb{N}})^*(B \times C) = Q^*(B)(P^{\mathbb{N}})^*(C)$ and $P^{\mathbb{N}} = P^{\mathbb{N}} \circ \pi^{-1}$ for any permutation π of the coordinates.

2.7. LEMMA. *Let $\mathcal{F} \subset \mathcal{L}_0(P)$. Then*

(a) $$\begin{aligned}&\Pr^*\{\sup_{f\in\mathcal{F}} |\textstyle\sum_{i=1}^n \varepsilon_i f(X_i)| > t\} \\ &\quad \leq 2\max_{k\leq n}\Pr^*\{\sup_{f\in\mathcal{F}} |\textstyle\sum_{i=1}^k f(X_i)| > (t/2)\}, \quad t > 0, \quad n \geq 1;\end{aligned}$$

(b) *Assume*

$$\sup_{f\in\mathcal{F}} \int (f - Pf)^2 \, dP = \alpha^2 < \infty.$$

Then, for $t > 2^{1/2}\alpha n^{1/2}$ *and for all* $n \in \mathbb{N}$,

$$\Pr{}^*\{\sup_{f\in\mathscr{F}} | \textstyle\sum_{i=1}^n (f(X_i) - Pf) | > t\} \le 4 \Pr{}^*\{\sup_{f\in\mathscr{F}} | \sum_{i=1}^n \varepsilon_i f(X_i) | > (t - 2^{1/2}n^{1/2}\alpha)/2\}. \tag{2.3}$$

(In the right-hand side, $f(X_i)$ can also be replaced by $f(X_i) - Pf$.)

PROOF. Part (a) goes as follows:

$$\begin{aligned}\Pr\{ \| \textstyle\sum_{i=1}^n \varepsilon_i f(X_i) \|_{\mathscr{F}} > t\} &\le \textstyle\sum_{\{\tau_i=\pm1, i\le n\}} \Pr{}^*[\{\varepsilon_i = \tau_i,\ i \le n\} \times \{\| \sum_{\{\tau_i=1\}} f(X_i) - \sum_{\{\tau_j=-1\}} f(X_j) \|_{\mathscr{F}} > t\}] \\ &\le 2 \textstyle\sum_{\{\tau_i=\pm1, i\le n\}} \dfrac{1}{2^n} \max_{k\le n}(P^{\mathbb{N}})^*\left\{ \| \sum_{i=1}^k f(X_i) \|_{\mathscr{F}} > \dfrac{t}{2}\right\} \\ &= 2 \max_{k\le n}\Pr{}^*\left\{ \| \textstyle\sum_{i=1}^k f(X_i) \|_{\mathscr{F}} > \dfrac{t}{2}\right\}.\end{aligned}$$

For (b), we apply Corollary 2.6 to the process $f \to \sum_{i=1}^n f(X_i)$, $f \in \mathscr{F}$. Then, if $X_i' = X_{n+i}$, $i = 1, \cdots, n$, Corollary 2.6 gives (for any $\tau_i = \pm 1$)

$$\begin{aligned}\Pr\{ \| \textstyle\sum_{i=1}^n (f(X_i) - Pf) \|_{\mathscr{F}} > t\} &\le 2 \Pr{}^*\{ \| \textstyle\sum_{i=1}^n (f(X_i) - f(X_i')) \|_{\mathscr{F}} > t - 2^{1/2}n^{1/2}\alpha\} \\ &= 2(P^{\mathbb{N}})^*\{ \| \textstyle\sum_{i=1}^n \tau_i(f(X_i) - f(X_i')) \|_{\mathscr{F}} > t - 2^{1/2}n^{1/2}\alpha\} \\ &= 2E_Q^*(P^{\mathbb{N}})^*\{ \| \textstyle\sum_{i=1}^n \varepsilon_i(f(X_i) - f(X_i')) \|_{\mathscr{F}} > t - 2^{1/2}n^{1/2}\alpha\} \\ &\le 4 \Pr{}^*\{ \| \textstyle\sum_{i=1}^n \varepsilon_i f(X_i) \|_{\mathscr{F}} > (t - 2^{1/2}n^{1/2}\alpha)/2\}. \quad \square\end{aligned}$$

It is easy to show that part (b) admits a version for non-identically distributed variables X_i if $\mathscr{F}$ is linearly supremum measurable.

For symmetric variables it is often possible to replace probability statements by statements about expectations, which are easier to work with in some instances. The following lemma of Hoffmann-Jørgensen (1974) is the result we will use for this. We state it adapted to our needs.

2.8. LEMMA. *Let* $\mathscr{F} \subset \mathscr{L}_p(P)$, $p > 0$, *be a linearly supremum measurable class for* P. *Then for every* $n \in \mathbb{N}$,

$$E \sup_{f\in\mathscr{F}} | \textstyle\sum_{i=1}^n \varepsilon_i f(X_i) |^p \le 2 \cdot 3^p E \max_{i\le n}\sup_{f\in\mathscr{F}} | f(X_i) |^p + 8 \cdot 3^p t_0^p, \tag{2.4}$$

where

$$t_0 = \inf[t > 0: \Pr\{\sup_{f\in\mathscr{F}} | \textstyle\sum_{i=1}^n \varepsilon_i f(X_i) | > t\} \le 1/(8\cdot 3^p)].$$

In the next lemma we examine the relationship between $\sum_i \varepsilon_i f(X_i)$ and $\sum_i g_i f(X_i)$. It is a simple modification of an inequality communicated to us by G. Pisier and X. Fernique; the proof below may be new.

2.9. LEMMA. *Let $\mathscr{F} \subset \mathscr{L}_1(P)$ be a linearly supremum measurable class (for P). Then the following inequalities hold for every $0 \le n_0 < \infty$ and $n_0 < n \in \mathbb{N}$:*

$$
\begin{aligned}
M_1 E \sup_{f\in\mathscr{F}} \left| \frac{\sum_{k=1}^n \varepsilon_k f(X_k)}{n^{1/2}} \right| & \\
&\le E \sup_{f\in\mathscr{F}} \left| \frac{\sum_{k=1}^n g_k f(X_k)}{n^{1/2}} \right| \\
&\le n_0 (E \sup_{f\in\mathscr{F}} |f(X_1)|)(E \max_{k\le n} |g_k|/n^{1/2}) \\
&\quad + M_2 \max_{n_0<k\le n} E \sup_{f\in\mathscr{F}} \left| \frac{\sum_{n_0<\ell\le k} \varepsilon_\ell f(X_\ell)}{k^{1/2}} \right|,
\end{aligned}
\tag{2.5}
$$

where $M_1 = E|g_1| > 0$ and $M_2 = \int_0^\infty (P\{|g_1| > u\})^{1/2}\,du < \infty$.

PROOF. The left-hand side inequality is a simple application of Jensen's inequality: if $\{\varepsilon_k, g_k, X_k\}_{k=1}^\infty$ are all independent, then $\mathscr{L}(\{\varepsilon_k |g_k|\}_{k=1}^\infty) = \mathscr{L}(\{g_k\}_{k=1}^\infty)$ and

$$
\begin{aligned}
E \left\| \frac{\sum_{k=1}^n g_k f(X_k)}{n^{1/2}} \right\|_{\mathscr{F}} &= E \left\| \frac{\sum_{k=1}^n \varepsilon_k |g_k| f(X_k)}{n^{1/2}} \right\|_{\mathscr{F}} \\
&\ge E \left\| \frac{\sum_{k=1}^n \varepsilon_k f(X_k) E|g_1|}{n^{1/2}} \right\|_{\mathscr{F}}.
\end{aligned}
$$

More surprising is the right-hand side inequality, which we prove now. For $\{g_k\}_{k=1}^n$, let $g_1^* \ge g_2^* \ge \cdots \ge g_n^*$ be the ordered values of $\{|g_k|\}_{k=1}^n$. Then, as in Remark 2.4 (3), we have

$$
E \left\| \sum_{k=1}^n g_k f(X_k) \right\|_{\mathscr{F}} = E \left\| \sum_{k=1}^n |g_k| \varepsilon_k f(X_k) \right\|_{\mathscr{F}} = E \left\| \sum_{k=1}^n g_k^* \varepsilon_k f(X_k) \right\|_{\mathscr{F}}. \tag{2.6}
$$

Now let

$$
T_k(f) = \sum_{n_0<\ell\le k} \varepsilon_\ell f(X_\ell)/k^{1/2}
$$

(note $T_{n_0}(f) = 0$). Then, setting $g_{n+1}^* = 0$

$$
\sum_{n_0<k\le n} g_k^* \varepsilon_k f(X_k) = \sum_{n_0<k\le n} (g_k^* - g_{k+1}^*) k^{1/2} T_k(f);
$$

so,

$$
\begin{aligned}
E \left\| \sum_{n_0<k\le n} g_k^* \varepsilon_k f(X_k) \right\|_{\mathscr{F}} & \\
&\le \sum_{n_0<k\le n} k^{1/2} E(g_k^* - g_{k+1}^*) E \| T_k(f) \|_{\mathscr{F}} \\
&\le [\max_{n_0<k\le n} E \| T_k(f) \|_{\mathscr{F}}][E\{\textstyle\sum_{n_0<k<n} k^{1/2}(g_k^* - g_{k+1}^*)\}].
\end{aligned}
$$

But the last expected value is equal to

$$E\left\{\sum_{n_0<k\le n}\int_{g_{k+1}^*}^{g_k^*} k^{1/2}\,du\right\} = E\int_0^{g_{n_0+1}^*} (\#\{r \ge 1: g_r^* \ge u\})^{1/2}\,du$$

$$\le \int_0^\infty (nP\{|g_1| > u\})^{1/2}\,du = M_2 n^{1/2},$$

using Jensen's inequality. This, together with (2.6) gives

$$E\left\|\frac{\sum_{k=1}^n g_k f(X_k)}{n^{1/2}}\right\|_{\mathscr{F}}$$

$$\le E\left\|\frac{\sum_{k=1}^{n_0} g_k^*\varepsilon_k f(X_k)}{n^{1/2}}\right\|_{\mathscr{F}} + E\left\|\frac{\sum_{k=n_0+1}^n g_k^*\varepsilon_k f(X_k)}{n^{1/2}}\right\|_{\mathscr{F}}$$

$$\le n_0(E\|f(X_1)\|_{\mathscr{F}})(E\max_{k\le n}|g_k|/n^{1/2}) + M_2\max_{n_0<k\le n}E\|T_k(f)\|_{\mathscr{F}}. \quad \square$$

(b) *Functional P-Donsker classes. The asymptotic equicontinuity criterion.* For the reader's convenience, here we describe the setup and main result of Dudley (1982), Section 4.1. Then we obtain some useful variations of it.

Let P be a probability measure on $(S, \mathscr{S})$ and let $\{X_i\}_{i=1}^\infty$ be i.i.d. with law P. Define the random measures

(2.7) $$P_n = (1/n)\sum_{i=1}^n \delta_{X_i}, \quad \nu_n = n^{1/2}(P_n - P).$$

P_n is the *empirical measure* corresponding to $\{X_i\}$. With the notation

$$\mu(f) = \int f\,d\mu$$

for any measure μ on S, we have by the finite dimensional central limit theorem that for any finite set of functions $f_1, \cdots, f_r \in \mathscr{L}_2(P)$,

$$\mathscr{L}(\nu_n(f_1), \cdots, \nu_n(f_r)) \to_w \mathscr{L}(G_P(f_1), \cdots, G_P(f_r)),$$

where $G_P(f)$ is a centered Gaussian process on $\mathscr{L}_2(P)$ with covariance

$$EG_P(f)G_P(g) = P[(f - Pf)(g - Pg)], \quad f, g \in \mathscr{L}_2(P).$$

So, the "limit law" of $\{\sum_{i=1}^n (f(X_i) - Pf)/n^{1/2}: f \in \mathscr{F}\}$ should be the law of $\{G_P(f): f \in \mathscr{F}\}$ for some version of G_P (two processes are said to be versions of each other if they have the same finite dimensional distributions). The following is then a natural definition:

2.10. DEFINITION. A class of functions $\mathscr{F} \subset \mathscr{L}_2(P)$ is a G_PBUC class if the process $\{G_P(f): f \in \mathscr{F}\}$ has a version with all the sample functions bounded and uniformly continuous for ρ_P. We will also say (instead of G_PBUC) that $\mathscr{F}$ is *P-pregaussian.*

The following definition is convenient in view of the theorem that follows it. We let $(\Omega, \Sigma, \mathrm{Pr}) = (S^{\mathbb{N}}, \mathscr{S}^{\mathbb{N}}, P^{\mathbb{N}}) \times ([0, 1], \mathscr{B}, \lambda)$, where λ is Lebesgue measure and as in subsection 2(a), the X_i are the coordinate functionals of $S^{\mathbb{N}}$.

2.11. DEFINITION. A class of functions $\mathscr{F} \subset \mathscr{L}_2(P)$ is a *functional P-Donsker class* if it is G_PBUC and there exist $\{Y_j\}_{j=1}^{\infty}$, i.i.d. versions of G_P, defined on $(\Omega, \Sigma, \mathrm{Pr})$, which have all their sample functions bounded and ρ_P-uniformly continuous, such that

(2.8) $\mathrm{Pr}^*\{n^{-1/2}\max_{k\le n} \|\sum_{i=1}^{k} (f(X_i) - Pf - Y_i(f))\|_{\mathscr{F}} > \varepsilon\} \to 0$ as $n \to \infty$.

All the results in this article have the following theorem of Dudley and Philipp (1983) as their starting point. See Dudley (1982, Theorems 4.1.1 and 4.1.2) for the proof.

2.12. THEOREM. *$\mathscr{F} \subset \mathscr{L}_2(P)$ is a functional P-Donsker class if and only if both*
(a) *$(\mathscr{F}, \rho_P)$ is totally bounded;*
(b) *for every $\varepsilon > 0$*

(2.9) $$\lim_{\delta\downarrow 0}\limsup_n \mathrm{Pr}^*\{\sup_{f,g\in\mathscr{F};\rho_P(f,g)<\delta} |\nu_n(f-g)| > \varepsilon\} = 0.$$

Moreover, conditions (a) *and* (b) *with ρ_P replaced by e_P are sufficient* (*but not necessary*) *for $\mathscr{F}$ to be a functional P-Donsker class.*

In order to apply the available theory of Gaussian and subgaussian processes (see the next subsection) it is convenient to have a "randomized" L_1 version of Theorem 2.12. The following corollary is required in the proof.

2.13. COROLLARY. *Let $\mathscr{F} \subset \mathscr{L}_2(P)$ be a* NLDM(P) *class which satisfies Condition* (b) *in Theorem* 2.12. *Then*

(2.10) $$\lim_{\delta\downarrow 0}\limsup_{t\to\infty} t^2 \mathrm{Pr}\{\|f(X_1)\|_{{}_0\mathscr{F}'_\delta} > t\} = 0.$$

As a consequence,

(2.11) $$\lim_{\delta\downarrow 0}\limsup_n n^{-1/2} E \max_{k\le n} \|f(X_k)\|_{{}_0\mathscr{F}'_\delta} = 0.$$

PROOF. By (2.9), the definition of NLDM, and Lemma 2.7 (a) it follows that

$$\lim_{\delta\downarrow 0}\limsup_n \mathrm{Pr}\{\|\textstyle\sum_{k=1}^{n} \varepsilon_k f(X_k)\|_{{}_0\mathscr{F}'_\delta} > \varepsilon n^{1/2}\} = 0$$

for all $\varepsilon > 0$. Then, Lévy's inequalities (see e.g. Araujo and Giné, 1980, Theorem 3.2.6) imply that

$$\lim_{\delta\downarrow 0}\limsup_n \mathrm{Pr}\{\max_{k\le n} \|f(X_k)\|_{{}_0\mathscr{F}'_\delta} > \varepsilon n^{1/2}\} = 0$$

for all $\varepsilon > 0$. Since, for independent real random variables ξ_k,

$$P\{\max_{k\le n} |\xi_k| > t\} \ge 1 - \exp\{-\textstyle\sum_{k=1}^{n} P\{|\xi_k| > t\}\}$$
$$\ge \textstyle\sum_{k=1}^{n} P\{|\xi_k| > t\}/(1 + \sum_k P\{|\xi_k| > t\}),$$

it follows easily that

$$\lim_{\delta\downarrow 0}\limsup_n n \Pr\{\|f(X_1)\|_{{}_0\mathscr{F}'_\delta} > \varepsilon n^{1/2}\} = 0,$$

thus proving the limit (2.10). Given $\eta > 0$, we can find, by (2.10), $\delta_0 > 0$ and $t_0 < \infty$ such that

$$t^2\Pr\{\|f(X_1)\|_{{}_0\mathscr{F}'_\delta} > t\} < \eta^2, \quad t > t_0, \quad \delta < \delta_0$$

(note that this function is monotone in δ). Then, for $\delta < \delta_0$,

$$\begin{aligned}\limsup_n n^{-1/2}E\max_{k\le n}\|f(X_k)\|_{{}_0\mathscr{F}'_\delta} &\le \eta + \limsup_n n^{1/2}\int_{\eta n^{1/2}}^\infty \Pr\{\|f(X_1)\|_{{}_0\mathscr{F}'_\delta} > t\}\,dt\\ &\le \eta + \limsup_n n^{1/2}\int_{\eta n^{1/2}}^\infty \eta^2 t^{-2}\,dt = 2\eta.\end{aligned}$$

Since η is arbitrary, this proves the limit (2.11). □

The result that follows is a direct consequence of Theorem 2.12, Corollary 2.13 and the randomization inequalities of the previous subsection. In what follows we let

$$\mathscr{L}_2^0(P) = \{f \in \mathscr{L}_2(S, \mathscr{S}, P): Pf = 0\}.$$

Note that $\rho_P = e_P$ in $\mathscr{L}_2^0(P)$.

2.14. THEOREM. (1). *Let $\mathscr{F} \subset \mathscr{L}_2^0(P)$ be a* NLDM(P) *class of real functions on S. Then the following are equivalent:*

(a) *$\mathscr{F}$ is a functional P-Donsker class.*

(b) *$(\mathscr{F}, e_P)$ is totally bounded and*

$$\lim_{\delta\downarrow 0}\limsup_n \Pr\{\|\textstyle\sum_{i=1}^n \varepsilon_i f(X_i)\|_{{}_0\mathscr{F}'_\delta} > \varepsilon n^{1/2}\} = 0 \tag{2.12}$$

for all $\varepsilon > 0$.

(c) *$(\mathscr{F}, e_P)$ is totally bounded and*

$$\lim_{\delta\downarrow 0}\limsup_n E\left\|\frac{\sum_{i=1}^n \varepsilon_i f(X_i)}{n^{1/2}}\right\|_{{}_0\mathscr{F}'_\delta} = 0. \tag{2.13}$$

(d) *$(\mathscr{F}, e_P)$ is totally bounded and*

$$\lim_{\delta\downarrow 0}\limsup_n \Pr\{\|\textstyle\sum_{i=1}^n g_i f(X_i)\|_{{}_0\mathscr{F}'_\delta} > \varepsilon n^{1/2}\} = 0; \tag{2.14}$$

(e) *$(\mathscr{F}, e_P)$ is totally bounded and*

$$\lim_{\delta\downarrow 0}\limsup_n E\left\|\frac{\sum_{i=1}^n g_i f(X_i)}{n^{1/2}}\right\|_{{}_0\mathscr{F}'_\delta} = 0. \tag{2.15}$$

(2) *if the* NLDM(P) *class $\mathscr{F}$ is only in $\mathscr{L}_2(P)$, but is $\mathscr{L}_1$-bounded (that is, $\sup_{f\in\mathscr{F}}P|f| < \infty$), then the previous statements* (a) – (e) *are also all equivalent.*

(3) *If the* NLDM(P) *class* $\mathscr{F}$ *is only in* $\mathscr{L}_2(P)$, *then each of the conditions* (b), (c) *and* (d) *implies* (a).

PROOF. Condition (2.12) is equivalent to condition (2.9) by Lemma 2.7. Hence, by Theorem 2.12, (a) and (b) are equivalent. Condition (2.13) implies (2.12) by Chebyshev's inequality and therefore (c) implies (b). Suppose now that (b) holds. Then (2.9) holds, and so does (2.11) by Corollary 2.13. Using (2.12) and (2.11) in Hoffmann–Jørgensen's inequality (Lemma 2.8), we obtain (2.13). Hence (b) implies (c). That (e) implies (c) is a direct consequence of the left-hand side inequality in Lemma 2.9. We now show that (c) implies (e). First note that as in the proof of Corollary 2.13,

$$E \max_{k\le n} |g_k|/n^{1/2} \to 0 \quad \text{as} \quad n \to \infty.$$

(Alternatively, it is well known that E $\max_{k\le n}|g_k| \simeq (\ln n)^{1/2}$.) Hence there exists $k_n \to \infty$ such that

$$k_n E \max_{k\le n}|g_k|/n^{1/2} \to 0 \quad \text{as} \quad n \to \infty.$$

Now we just apply the right hand side of (2.5) in Lemma 2.9 with $n_0 = k_n$. (e) clearly implies (d). The implication "(d) implies (e)" follows again from Hoffmann–Jørgensen's inequality and Corollary 2.13.

(2) follows from (1) and the fact that a $\mathscr{L}_1$-bounded class is e_P-totally bounded iff it is ρ_P-totally bounded.

The sufficiency of the conditions (b), (c) and (d) in case the functions in $\mathscr{F}$ are neither centered nor uniformly bounded follows as in case (1). □

(c) *Gaussian and Subgaussian processes.* The processes which appear in conditions (2.12) and (2.14) of Theorem 2.14 above are respectively *subgaussian* and *Gaussian* for *fixed* $\{X_i(\omega)\}$. We will use to our advantage the fact that there exist very sharp "majorations" and "minorations" for these processes. We recall in this subsection the relevant facts on Gaussian and subgaussian processes.

Let Z_t, $t \in T$, be a stochastic process defined on some parameter set T. Z is a subgaussian process (more precisely a process with subgaussian increments) if there exists $\tau > 0$ such that for all $s, t \in T$ and for all real λ,

$$E \exp\{\lambda(Z(s) - Z(t))\} \le \exp\{\lambda^2 \tau E(Z(s) - Z(t))^2/2\}.$$

(This definition is taken from Jain and Marcus, 1978; see also Kahane, 1968.) We are interested in the following example of a subgaussian process: if $Z(t) = \sum_{i=1}^{\infty} \varepsilon_i h_i(t)$ and $\sum_{i=1}^{\infty} h_i^2(t) < \infty$ for all $t \in T$, then Z is subgaussian with $\tau = 1$. In fact, since

$$E \exp(\lambda\varepsilon_i) = \sum_{n=0}^{\infty} \frac{\lambda^{2n}}{(2n)!} \le \sum_{n=0}^{\infty} \left(\frac{\lambda^2}{2}\right)^n \Big/ n! = \exp\left(\frac{\lambda^2}{2}\right), \quad \lambda \in \mathbb{R},$$

it follows that

$$E \exp(\lambda(Z(s) - Z(t))) \le \exp\{\lambda^2 \sum_{i=1}^{\infty} (h_i(s) - h_i(t))^2/2\}.$$

This shows that, for fixed $\omega \in S^{\mathbb{N}}$, the random process

$$f \to \sum_{i=1}^{n} \varepsilon_i(\cdot) f(X_i(\omega)), \quad f \in \mathscr{F} \tag{2.16}$$

is subgaussian with $\tau = 1$. It is well known and easy to see from the above exponential inequality, that for any $\{a_i\} \subset \mathbb{R}$,

$$P\{|\sum_i a_i \varepsilon_i| > t\} \le 2 \exp\{(-t^2/2 \sum_i a_i^2)\}, \quad t > 0 \tag{2.17}$$

which we will refer to as the "standard subgaussian inequality."

A related inequality due to Bernstein (see, e.g., Bennett, 1962) is as follows: let $\{\xi_i\}_{i=1}^n$, $n \in \mathbb{N}$, be independent real valued random variables bounded by M and with mean zero; then for $t > 0$,

$$P\{\sum_{i=1}^n \xi_i \ge t\} \le \exp\{-t^2/(2 \sum_{i=1}^n E\xi_i^2 + 2Mt/3)\}. \tag{2.18}$$

Let (T, ρ) be a metric or pseudo-metric space. Then, the *covering number* $N(\varepsilon, T, \rho)$, $\varepsilon \in [0, |T|_\rho]$, where $|T|_\rho$ denotes the diameter of (T, ρ), is defined as

$$N(\varepsilon, T, \rho) = \min\{n: \exists t_1, \cdots, t_n \in T \text{ s.t. } \min_i \rho(t, t_i) \le \varepsilon \text{ for all } t \in T\}. \tag{2.19}$$

The *metric entropy* $H(\varepsilon, T, \rho)$ is defined as

$$H(\varepsilon, T, \rho) = \ln N(\varepsilon, T, \rho).$$

Given a subgaussian process $Z(t)$, $t \in T$, let

$$\sigma(s, t) = (E(Z(t) - Z(s))^2)^{1/2}, \tag{2.20}$$

and

$$\Phi(\delta) = 4\delta(\ln \ln 4 |T|_\sigma \delta^{-1})^{1/2}, \quad 0 < \delta \le |T|_\sigma. \tag{2.21}$$

With these definitions we have:

2.15. THEOREM. *Let $Z(t)$, $t \in T$, be a subgaussian process on T with $\tau = 1$ in the definition. Assume that (T, σ) satisfies the metric entropy condition*

$$\int_0^{|T|_\sigma} H^{1/2}(\varepsilon, T, \sigma)\, d\varepsilon < \infty. \tag{2.22}$$

Then Z has a version $\tilde{Z}$ with σ-uniformly continuous sample paths, and there exists $C > 0$ independent of Z such that

$$E \sup_{\sigma(s,t) \le \delta} |\tilde{Z}(t) - \tilde{Z}(s)| \le C \left[\int_0^\delta H^{1/2}(\varepsilon, T, \sigma)\, d\varepsilon + \Phi(\delta)\right]. \tag{2.23}$$

In particular, for any $t_0 \in T$

$$E \sup_{t \in T} |\tilde{Z}(t)| \le E|\tilde{Z}(t_0)| + C\left[\int_0^{|T|_\sigma} H^{1/2}(\varepsilon, T, \sigma)\, d\varepsilon + \Phi(|T|_\sigma)\right]. \tag{2.24}$$

This theorem is essentially due to Dudley (1967, 1973) for Gaussian processes, and it is stated and proved in the present form in Marcus and Pisier (1978, Theorem 3.1).

Note that if $Z(t) = \sum_{i=1}^n \varepsilon_i h_i(t)$, h_i continuous in (T, σ), then $Z = \tilde{Z}$ in Theorem 2.15.

To obtain necessary conditions for the central limit theorem we must consider the Gaussian processes (ω fixed)

$$f \to \sum_{i=1}^n g_i(\cdot) f(X_i(\omega)) \tag{2.25}$$

where $\{g_i\}$ is a $N(0, 1)$ i.i.d. sequence independent of $\{X_i\}$. We will require a lower bound for the supremum of a Gaussian process in terms of its metric entropy, which is due to Sudakov (1971) ("Sudakov's minoration": see e.g. Fernique (1974), 2.3.1 and 2.3.3), and an integrability result for Gaussian processes (see Fernique, 1974, 1.3.2, for a more general result).

2.16. THEOREM. *Let $\{Z(t), t \in T\}$ be a centered Gaussian process on a parameter set T and let σ be as in* (2.20). *Then there is a constant $C > 0$ such that*

$$E^* \sup_{t\in T} |Z(t)| \geq C \sup_{\lambda>0} \lambda H^{1/2}(\lambda, T, \sigma). \tag{2.26}$$

Furthermore

(a) *if $\sup_{t\in T} |Z(t)|$ is measurable and finite a.s., then $E \sup_{t\in T} |Z(t)| < \infty$, and*

(b) *if Z has almost all its trajectories bounded and uniformly continuous on (T, σ), then*

$$\lim_{\lambda\to 0} \lambda H^{1/2}(\lambda, T, \sigma) = 0. \tag{2.26$'$}$$

Note that (2.26)$'$ follows from (2.26) by considering the Karhunen–Loève expansion of Z and the inequality $N(\varepsilon, T, d_1 + d_2) \leq N(\varepsilon/4, T, d_1)N(\varepsilon/4, T, d_2)$ for metrics d_1 and d_2.

In some of the proofs below we will need to compare different Gaussian processes. Next we state two results in this direction. The first one is due to Fernique (1974, 2.12) after substantial contributions by Slepian and by Marcus and Shepp. The second one is a beautiful recent result of Fernique (1983, 1.6 and proof of 2.2).

2.17. THEOREM. *Let $\{Z_i(t), t \in T\}$, $i = 1, 2$, be centered Gaussian processes defined on a countable set T. Let $d_i(s, t) = [E(Z_i(s) - Z_i(t))^2]^{1/2}$, $s, t \in T$, $i = 1, 2$.*

(a) *Assume that the set T is countable and that for all $s, t \in T$,*

$$d_1(s, t) \leq d_2(s, t). \tag{2.27}$$

Then

$$E \sup_{t\in T} Z_1(t) \leq E \sup_{t\in T} Z_2(t). \tag{2.28}$$

(b) *Assume that T is a compact metric space and that Z_1 and Z_2 have continuous* *paths a.s. Then, if inequality* (2.27) *holds, we have for all* $\delta > 0$,

$$\begin{aligned} &E \sup_{d_2(s,t)\le\delta} |Z_1(s) - Z_1(t)| \\ &\qquad \le 4 \sup_{s\in T} E \sup_{d_2(s,t)\le\delta} |Z_2(s) - Z_2(t)| + 13\delta H^{1/2}(\delta/2, T, d_2). \end{aligned} \tag{2.29}$$

Another useful result is the following (see e.g. Dudley, 1973, Theorem 0.3):

2.18. LEMMA. *If* $\mathscr{F}$ *is* $\mathrm{G_P}$ BUC, *then so is its* ρ_P*-closed, convex symmetric hull, in particular* $\mathscr{F}' := \{f - g: f, g \in \mathscr{F}\}$ *is also* $\mathrm{G_P}$ BUC.

(d) *Random entropies and Vapnik-Červonenkis classes. The Vapnik-Červonenkis law of large numbers.* Given a sample $\{X_i\}_{i=1}^n$ from P we consider the random distances

$$\begin{cases} d_{n,p}(f, g) = [\sum_{i=1}^n |f(X_i) - g(X_i)|^p/n]^{1/(p\vee 1)}, & 0 < p < \infty. \\ d_{n,\infty}(f, g) = \max_{i\le n} |f(X_i) - g(X_i)| \end{cases} \tag{2.30}$$

for f, $g \in \mathscr{F}$ (and more generally, for f, g measurable). Then we define the covering numbers of $\mathscr{F}$ for these distances:

$$N_{n,p}(\varepsilon, \mathscr{F}) = N(\varepsilon, \mathscr{F}, d_{n,p}), \quad N_{n,\infty}(\varepsilon, \mathscr{F}) = N(\varepsilon, \mathscr{F}, d_{n,\infty}), \tag{2.31}$$

which are random, and the random entropies

$$H_{n,p}(\varepsilon, \mathscr{F}) = \ln N_{n,p}(\varepsilon, \mathscr{F}), \quad H_{n,\infty}(\varepsilon, \mathscr{F}) = \ln N_{n,\infty}(\varepsilon, \mathscr{F}).$$

Of particular interest to us in this article are $H_{n,1}$, $H_{n,2}$ and $H_{n,\infty}$. These random numbers may not be random variables. They have been first used in connection with empirical processes by Vapnik and Červonenkis (1971, 1981), and also by Kolčinskiĭ (1981).

Vapnik and Červonenkis (1981) proved the following result. In their version some measurability conditions seem to be missing; looking at their proof of necessity and using Theorem 8.3 below for sufficiency as well as Corollary 8.8, one obtains:

2.19. THEOREM. *Let* $\mathscr{F}$ *be* NLSM(P) *such that* $\|f\|_\infty \le 1$ *for all* $f \in \mathscr{F}$ *and such that* $N_{n,\infty}(\varepsilon, {}_0\mathscr{F})$ *and* $N_{n,1}(\varepsilon, {}_0\mathscr{F})$ *are measurable where* ${}_0\mathscr{F}$ *is as given in Definition* 2.3. *Then* (a) *and* (b) *below are equivalent.*

(a) $\|P_n - P\|_{\mathscr{F}} \to 0$ *a.s.*

(b) *for every* $\varepsilon > 0$

$$\lim_{n\to\infty} \frac{EH_{n,\infty}(\varepsilon, {}_0\mathscr{F})}{n} = 0. \tag{2.32}$$

If, moreover, the class $\mathscr{G} := \{|f - g|: f, g \in \mathscr{F}\}$ *is* LSM(P), *then* (a) *and* (b) *are also equivalent to*

(c) *there exists a finite function* $T(\varepsilon)$ *such that for all* $\varepsilon > 0$

$$\lim_{n\to\infty} \Pr\{N_{n,1}(\varepsilon, {}_0\mathscr{F}) > T(\varepsilon)\} = 0.$$

In the case of sets Vapnik and Červonenkis (1971) proved this theorem with convergence in probability. Steele (1978) showed that this convergence is a.s. (see also Kuelbs and Zinn, 1979, and Pollard, 1981). These proofs of a.s. convergence extend to the case of NSM(P) classes of functions.

In the set case condition (2.32) takes a special form. Let $\mathscr{C}$ be a family of measurable sets. As in Vapnik and Červonenkis (1971), define $\Delta^{\mathscr{C}}(X_1, \cdots, X_n)$ as the number of different sets in the collection $\{C \cap \{X_1, \cdots, X_n\}: C \in \mathscr{C}\}$, i.e.

$$(2.33) \qquad \Delta^{\mathscr{C}}(X_1, \cdots, X_n) = \#\{C \cap \{X_1, \cdots, X_n\}: C \in \mathscr{C}\}.$$

Identifying $\mathscr{C}$ with $\mathscr{F}_{\mathscr{C}} = \{I_C: C \in \mathscr{C}\}$, it is clear that for all $\varepsilon < 1$,

$$N_{n,\infty}(\varepsilon, \mathscr{C}) = \Delta^{\mathscr{C}}(X_1, \cdots, X_n),$$

so that condition (2.32) becomes

$$(2.32'') \qquad \lim_{n\to\infty} \frac{E^* \ln \Delta^{\mathscr{C}}_{0}(X_1, \cdots, X_n)}{n} = 0,$$

which is equivalent to

$$\frac{\ln \Delta^{\mathscr{C}}_{0}(X_1, \cdots, X_n)}{n} \to 0$$

in outer probability (see Lemma 2.20).

Now we show that this equivalence persists for $N_{n,p}$ as long as the functions in $\mathscr{F}$ are uniformly bounded. Although trivial, this is relevant for our results on the law of large numbers in Section 8.1.

2.20. LEMMA. *Let $\mathscr{F}$ be a uniformly bounded class of real functions on S and let $1 \ge \varepsilon > 0$, $0 < p \le \infty$, $0 < r < \infty$. Then,*

$$\lim_{n\to\infty} \Pr{}^* \left\{ \frac{\ln N_{n,p}(\varepsilon, \mathscr{F})}{n} > \delta \right\} = 0 \quad \textit{for all} \quad \delta > 0$$

if and only if

$$\lim_{n\to\infty} E^* \left(\frac{\ln N_{n,p}(\varepsilon, \mathscr{F})}{n} \right)^r = 0.$$

PROOF. We may assume that for $f \in \mathscr{F}$, $\sup_{s\in S} |f(s)| \le 1$. Let

$$R_n = \frac{\ln N_{n,p}(\varepsilon, \mathscr{F})}{n}.$$

We need only observe that since $N_{n,p}(\varepsilon, \mathscr{F}) \le N_{n,\infty}(\varepsilon^{(p^{-1}\vee 1)}, \mathscr{F}) \le (2/\varepsilon^{(p^{-1}\vee 1)})^n$,

$$E^* R_n^r \ge E^* \left(\frac{R_n^r}{1 + R_n^r} \right) \ge E^* \left(\frac{R_n^r}{1 + c^r} \right)$$

where $c = \ln(2/\varepsilon^{(p^{-1}\vee 1)})$. □

2.21. REMARK. In the case of $\Delta^{\mathscr{S}}(X_1, \cdots, X_n)$ one can modify an argument of Hoffmann-Jørgensen (1974), page 164, to obtain the equivalence of

$$\frac{\ln \Delta^{\mathscr{S}}(X_1, \cdots, X_n)}{a_n} \to 0 \quad \text{in Pr}^*$$

and

$$E^* \left(\frac{\ln \Delta^{\mathscr{S}}(X_1, \cdots, X_n)}{a_n} \right) \to 0$$

for any $r > 0$ and any sequence $a_n \to \infty$.

Finally we describe another remarkable result of Vapnik and Červonenkis (1971). Given a class $\mathscr{S}$ of subsets of S, define

$$m^{\mathscr{S}}(n) = \max\{\Delta^{\mathscr{S}}(s_1, \cdots, s_n), s_i \in S\}, \quad V(\mathscr{S}) = \inf\{n: m^{\mathscr{S}}(n) < 2^n\},$$
$$V(\mathscr{S}) = +\infty \quad \text{if} \quad m^{\mathscr{S}}(n) = 2^n \quad \text{for all} \quad n.$$

Dudley (1978) calls $\mathscr{S}$ a *Vapnik-Červonenkis class* of sets, (a VCC for short) if $V(\mathscr{S}) < \infty$. It is not difficult to prove by induction that if $\mathscr{A}(A_1, \cdots, A_k)$ denotes the algebra generated by the sets $A_1, \cdots, A_k$, then for any $k < +\infty$, the family of sets

$$\mathscr{A}_k(\mathscr{S}) = \cup \{\mathscr{A}(A_1, \cdots, A_k): A_1, \cdots, A_k \in \mathscr{S}\}$$

is a VCC if $\mathscr{S}$ is (Dudley 1978, Proposition 7.12). What is most remarkable and useful about VCC is the following combinatorial lemma of Vapnik and Červonenkis (1971):

2.22. THEOREM. *$m^{\mathscr{S}}(n) \le n^{V(\mathscr{S})} + 1$ for all $n \ge 1$, and $m^{\mathscr{S}}(n) \le n^{V(\mathscr{S})}$ for all $n \ge 2$.*

Dudley (1978, Lemma 7.13) uses this theorem to prove another beautiful result:

2.23. THEOREM. *Let $\mathscr{S}$ be a* VCC. *Then for any p.m. P on $(S, \mathscr{S})$, the metric entropy $N(\varepsilon, \mathscr{S}, \|\cdot\|_{L_2(P)}) := N(\varepsilon^2, \mathscr{S}, P)$ is bounded by*

$$N(\varepsilon^2, \mathscr{S}, P) \le A(v)(\varepsilon^{-1}\ln \varepsilon^{-1})^{2v}, \quad 0 < \varepsilon < ½$$

where $v = V(\mathscr{S})$, and $A(v) < \infty$.

Using the technique of the proof of Theorem 2.23, we have the following lemma which will be useful below.

2.24. LEMMA. *Let $F(x)$, $x > 0$, be a nonnegative differentiable function satisfying*

$$(2.34) \quad F(x) \uparrow \infty, \quad \frac{F(x)}{x} \downarrow 0 \quad (\textit{as } x \uparrow \infty) \quad \textit{and} \quad \int^{\infty} x^{-3/2} F(x)\, dx < \infty.$$

Let $\mathscr{F}$ be a class of measurable functions uniformly bounded by 1. *Suppose that there exist $\tau > 0$ and n_0 such that for all $n > n_0$,*

$$\Pr{}^*\left\{\frac{\ln N_{n,\infty}((\tau F(n)/n)^{1/2}, \mathscr{F})}{F(n)} > \tau\right\} < \frac{1}{2}. \tag{2.35}$$

Then, for all $n > n_0$,

$$\int_0^{\sqrt{24\tau F(n)/n}} \sqrt{\ln N(\varepsilon, \mathscr{F}, e_P)}\, d\varepsilon \le 2\sqrt{3\tau}\frac{F(n)}{n^{1/2}} + \sqrt{3\tau}\int_n^\infty \frac{F(x)}{x^{3/2}}\, dx \tag{2.36}$$

which converges to zero as $n \to \infty$.

PROOF. Let $f_1, \cdots, f_m \in \mathscr{F}$ be such that $E|f_i - f_j|^2 \ge \varepsilon^2$ for all $i \ne j$ and some $0 < \varepsilon < 1$. Let X_k be i.i.d. with law P. Then, for $0 < \delta < \varepsilon$

$$\begin{aligned}
&P\{\max_{k\le n}|f_i(X_k) - f_j(X_k)| \le \delta \text{ for some } i \ne j\}\\
&\qquad \le 1 \wedge \left(\frac{m(m-1)}{2}\cdot \max_{i\ne j}[P\{|f_i(X_1) - f_j(X_1)| \le \delta\}]^n\right)\\
&\qquad \le 1 \wedge \left(\frac{m^2}{2}\left(\frac{1-\varepsilon^2}{1-\delta^2}\right)^n\right) \le 1 \wedge (m^2\exp(-n(\varepsilon^2 - \delta^2))/2).
\end{aligned}$$

Therefore

$$\Pr{}^*\{N_{n,\infty}(\delta, F) \ge m\} \ge 1 - 1 \wedge ((1/2)\, m^2\exp(-n(\varepsilon^2 - \delta^2))). \tag{2.37}$$

Let now $M(\varepsilon^2) = \sup\{n\colon \exists f_1, \cdots, f_n \in \mathscr{F} \text{ such that } E|f_i - f_j|^2 \ge \varepsilon^2\ i \ne j\}$, and note that

$$N(2\varepsilon, \mathscr{F}, e_P) \le M(\varepsilon^2). \tag{2.38}$$

Let $\varepsilon_n^2 = \sup\{\varepsilon^2\colon M(\varepsilon^2) \ge e^{\tau F(n)}\}$. Then there exist $f_1, \cdots, f_{[e^{\tau F(n)}]}$ such that $E|f_i - f_j|^2 \ge (1/2)\varepsilon_n^2$. Take $m = \tau F(n)$ and $\delta = (\tau F(n)/2)^{1/2}$. Then (2.35) and (2.37) give that for $n > n_0$,

$$(1/2) > 1 - (1 \wedge (1/2)\exp\{3\tau F(n) - n\varepsilon_n^2/2\})$$

or

$$\exp\{3\tau F(n) - n\varepsilon_n^2/2\} > 1$$

or

$$\varepsilon_n^2 > 6\tau F(n)/n.$$

So, by (2.38) and the definition of ε_n,

$$N((24\tau F(n)/n)^{1/2}, \mathscr{F}, e_P) \le \exp(\tau F(n)).$$

Hence, for $n > n_0$, since $F(n)/n^{1/2} \to 0$, we have

$$\int_0^{\sqrt{24\tau F(n)/n}} (\ln N(\varepsilon, \mathscr{F}, e_P))^{1/2}\, d\varepsilon$$

$$\leq -4\sqrt{3}\,\tau \int_n^\infty (F(x))^{1/2} \left(\left(\frac{F(x)}{x}\right)^{1/2}\right)' dx$$

$$= 2\sqrt{3}\,\tau \frac{F(n)}{n^{1/2}} + \sqrt{3}\,\tau \int_n^\infty \frac{F(x)}{x^{3/2}}\, dx \to 0 \quad \text{as} \quad n \to \infty. \quad \square$$

3. The general CLT for uniformly bounded families of functions. Given $\mathscr{F} \subseteq \mathscr{L}^2(P)$ and a pseudo-metric ρ on $\mathscr{F}$ we say that $(\mathscr{F}, \rho)$ satisfies the *metric entropy condition* if

$$\text{(3.1)} \qquad \int_0 (\ln N(\varepsilon, \mathscr{F}, \rho))^{1/2}\, d\varepsilon < \infty.$$

It is known (Dudley, 1973) that the metric entropy condition for $(\mathscr{F}, \rho_P)$ implies sample continuity of $G_P|_{\mathscr{F}}$, but does not necessarily imply that $\mathscr{F}$ is a functional P-Donsker class.

The following theorem based on Dudley (1978, Theorem 5.1), shows that, however, for uniformly bounded classes $\mathscr{F}$, condition (3.1) with $\rho = \rho_P$ (or equivalently, with $\rho = e_P$) does control the size of "part" of (2.9), namely, of

$$\sup_{f,g \in \mathscr{F}, \varepsilon/n^{1/2} < e_P^2(f,g) < \delta} |\nu_n(f - g)|$$

for all $\varepsilon > 0$.

3.1. THEOREM. *Let $\mathscr{F}$ be a family of measurable functions on $(S, \mathscr{S})$ such that $\|f\|_\infty \leq 1$ for all $f \in \mathscr{F}$. Let ρ be an e_p-uniformly continuous pseudo-metric on $\mathscr{F}$ such that $e_P(f, g) \leq \rho(f, g)$, $f, g \in \mathscr{F}$, and $\rho(f, g) = 0$ if $f = g$ P a.s. Assume that $(\mathscr{F}, \rho)$ satisfies the metric entropy condition* (3.1). *Then $\mathscr{F}$ is a functional P-Donsker class if and only if there exist τ, $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$,*

$$\text{(3.2)} \qquad \lim_{n\to\infty} \Pr{}^*\{\sup_{f,g \in \mathscr{F}, \rho^2(f,g) \leq \varepsilon/n^{1/2}} |\nu_n(f - g)| > \tau\varepsilon\} = 0.$$

PROOF. The metric entropy condition on $(\mathscr{F}, \rho)$ implies (together with the assumption $e_P \leq \rho$) that both $(\mathscr{F}, \rho)$ and $(\mathscr{F}, e_P)$ are totally bounded. So, by Theorem 2.12, in order to prove sufficiency, we must show that (2.9) with e_P (instead of ρ_P) holds. We will actually prove (2.9) with ρ replacing ρ_P. Now, let $\tilde{\mathscr{F}}$ be the set of equivalence classes of $\mathscr{F}$ mod P (that is, $f \sim g$ iff $f = g$ P- a.s.), and let $\bar{\mathscr{F}}$ be the completion of $\tilde{\mathscr{F}}$ with respect to ρ. Then $(\bar{\mathscr{F}}, \rho)$ is a compact metric space and $(\bar{\mathscr{F}}, e_P)$ is Hausdorff. Hence the identity map $(\bar{\mathscr{F}}, \rho) \to (\bar{\mathscr{F}}, e_P)$ is a homeomorphism, in particular the inverse map is continuous. So, (2.9) for ρ implies (2.9) for e_P.

The metric entropy condition also implies that given $\varepsilon > 0$ we may take μ

large enough so that

$$\sum_{i\geq\mu} (2^{-i}\ln N(2^{-i/2}, \mathscr{F}, \rho))^{1/2} < \varepsilon/32 \tag{3.3}$$

and

$$\sum_{i\geq 0} \exp\left\{-\frac{2^{i+\mu}\varepsilon^2}{2^{11}(i+1)^4}\right\} < \frac{\varepsilon}{16}. \tag{3.4}$$

Let $r \geq \mu$ be large enough so that $\delta_0 := 2^{-r} \leq \varepsilon^2$. Let $\delta_k = \delta_0/2^k = 2^{-k-r}$ and $b_k^2 = 2^{-k}\ln N(\delta_k^{1/2}, \mathscr{F}, \rho)$. Choose $n_0 > \varepsilon^2/256\delta_0^2$, and for each $n \geq n_0$ let $k = k(n)$ be such that

$$\frac{1}{2} < \frac{8\delta_k n^{1/2}}{\varepsilon} \leq 1. \tag{3.5}$$

For $i = 0, 1, \cdots$, let $N_i = N(\delta_i^{1/2}, \mathscr{F}, \rho)$ and let $\mathscr{F}_i = \{h_1^i, \cdots, h_{N_i}^i\} \subset \mathscr{F}$ such that for each $f \in \mathscr{F}$, there exists $1 \leq j(i, f) \leq N_i$ with

$$\rho(f, h_{j(i,f)}^i) \leq \delta_i^{1/2}.$$

Now assume $f, g \in \mathscr{F}$ with $\rho(f, g) \leq \delta_0^{1/2}$. We have:

$$\begin{aligned}|\nu_n(f-g)| \leq{}& |\nu_n(f - h_{j(k,f)}^k)| + |\nu_n(g - h_{j(k,g)}^k)|\\ &+ \textstyle\sum_{0\leq i<k} |\nu_n(h_{j(i,f)}^i - h_{j(i+1,f)}^{i+1})|\\ &+ \textstyle\sum_{0\leq i<k} |\nu_n(h_{j(i,g)}^i - h_{j(i+1,g)}^{i+1})|\\ &+ |\nu_n(h_{j(0,f)}^0 - h_{j(0,g)}^0)|.\end{aligned}$$

Since, e.g., $\rho(h_{j(i,f)}^i, h_{j(i+1,f)}^{i+1}) \leq (3\delta_i)^{1/2}$, $i \geq 0$, and $\rho(h_{j(0,f)}^0, h_{j(0,g)}^0) \leq 3\delta_0^{1/2}$, we have

$$\begin{aligned}\sup_{f,g\in\mathscr{F},\rho(f,)\leq\delta_0^{1/2}} |\nu_n(f-g)| \leq{}& 2\sup_{f,g\in\mathscr{F},\rho(f,g)\leq\delta_k^{1/2}} |\nu_n(f-g)|\\ &+ 3\textstyle\sum_{0\leq i<k} \sup_{f,g\in\mathscr{F}_i\cup\mathscr{F}_{i+1},\rho(f,g)\leq 3\delta_i^{1/2}} |\nu_n(f-g)|.\end{aligned}$$

Hence, for τ as in (3.2),

$$\begin{aligned}&\Pr{}^*\{\sup_{f,g\in\mathscr{F},\rho(f,g)\leq\delta_0^{1/2}} |\nu_n(f-g)| \geq (2\tau+3)\varepsilon\}\\ &\quad\leq \Pr{}^*\{\sup_{f,g\in\mathscr{F},\rho^2(f,g)\leq\varepsilon/n^{1/2}} |\nu_n(f-g)| \geq \tau\varepsilon\}\\ &\quad\quad+ \Pr{}^*\{\textstyle\sum_{0\leq i<k} \sup_{f,g\in\mathscr{F}_i\cup\mathscr{F}_{i+1},\rho(f,g)\leq 3\delta_i^{1/2}} |\nu_n(f-g)| > \varepsilon\}\\ &\quad:= (\mathrm{I}) + (\mathrm{II}).\end{aligned}$$

By (3.2), (I) $\leq \varepsilon/2$ for n sufficiently large. To estimate II, let

$$d_i = \max\{(i+1)^{-2}\varepsilon/32,\ 2^{1-r/2}b_{i+1}\}, \quad i = 0, 1, \cdots.$$

Then, by (3.3)

$$\textstyle\sum_{i=0}^{\infty} d_i < \varepsilon/8.$$

Hence,

$$(\mathrm{II}) \leq \textstyle\sum_{0\leq i<k} 4N_{i+1}^2 \sup_{f,g\in\mathscr{F},\rho(f,g)\leq 3\delta_i^{1/2}} \Pr{}^*\{|\nu_n(f-g)| > 8d_i\}.$$

Since $\|f - g\|_\infty \le 2$, $\rho_P(f, g) \le \rho(f, g) \le 3\delta_i^{1/2}$, and for $0 \le i < k$, $n^{-1/2}d_i \le \varepsilon n^{-1/2}/8 < 2\delta_k \le \delta_i$ (see (3.5)), Bernstein's inequality (2.18) gives

$$(\mathrm{II}) \le 8 \sum_{0\le i<k} N_{i+1}^2 \exp\{-64d_i^2/(64/3 + 18)\delta_i\}.$$

Therefore, using (3.4) and the definition of d_i, we obtain

$$(\mathrm{II}) \le 8 \sum_{i=0}^\infty \exp\left\{4\cdot 2^i b_{i+1}^2 - \frac{64\cdot 3}{118}\frac{d_i^2}{\delta_i}\right\}$$

$$\le 8 \sum_{i=0}^\infty \exp\left\{-\frac{74}{118}\frac{d_i^2}{\delta_i}\right\} \le 8 \sum_{i=0}^\infty \exp\left\{-\frac{2^{i+r}\varepsilon^2}{2^{11}(i+1)^4}\right\} < \frac{\varepsilon}{2}.$$

Hence sufficiency is proved.

Necessity of (3.2) follows directly from Theorem 2.12 since $\rho_P \le e_P$. □

While the entropy condition (3.1) is sufficient for the existence of the limiting sample continuous Gaussian process in Theorem 3.1, it is not necessary (Dudley, 1967, Proposition 6.10). In the next theorem, essentially at the expense of some measurability, we replace the entropy condition by the necessary condition that $\mathscr{F}$ be pregaussian.

3.2. THEOREM. *Let $\mathscr{F}$ be a uniformly bounded* NLDM(P) *class of functions on S. Then the conditions*

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

(ii) *there exist τ, $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$,*

$$\lim \sup_n \Pr{}^*\left\{\sup_{f,g\in{}_0\mathscr{F};P(f-g)^2\le\varepsilon/n^{1/2}} \left|\frac{\sum_{i=1}^n \varepsilon_i(f(X_i) - g(X_i))}{n^{1/2}}\right| > \tau\varepsilon\right\} = 0 \tag{3.6}$$

are necessary and (together) sufficient for $\mathscr{F}$ to be a functional P-Donsker class.

PROOF. Sufficiency of (i) and (ii). We assume without loss of generality that $\|f\|_\infty \le 1$ for all $f \in \mathscr{F}$. For simplicity of notation we write $\mathscr{F}$ for ${}_0\mathscr{F}$. By Theorem 2.16 and condition (i), $(\mathscr{F}, \rho_P)$ is totally bounded, and, by uniform boundedness, $(\mathscr{F}, e_P)$ is also totally bounded. Hence for every $\varepsilon > 0$, and $n \in \mathbb{N}$, the number $m(\varepsilon, n) = \max\{m$: there exist $h_1, \cdots, h_m \in \mathscr{F}$ such that $P(h_i - h_j)^2 > \varepsilon/4n^{1/2}$ for $i \ne j\}$ is finite. Let $\mathscr{A} = \mathscr{A}(\varepsilon, n)$ be a collection of $m(\varepsilon, n)$ functions $h_i \in \mathscr{F}$ satisfying $P(h_i - h_j)^2 > \varepsilon/4n^{1/2}$, $i \ne j$. Note in particular that

$$\sup_{f\in\mathscr{F}} \min_{h_i\in\mathscr{A}} P(f - h_i)^2 \le \varepsilon/n^{1/2}. \tag{3.7}$$

Then by (3.7), for every $\delta > 0$ and for n sufficiently large (depending on δ),

$$\Pr\left\{\left\|\frac{\sum_{i=1}^n \varepsilon_i f(X_i)}{n^{1/2}}\right\|_{\mathscr{F}'_\delta} > 3\tau\varepsilon\right\}$$

$$\le 2\Pr{}^*\left\{\left\|\frac{\sum_{i=1}^n \varepsilon_i f(X_i)}{n^{1/2}}\right\|_{\mathscr{F}'_{\varepsilon,n}} > \tau\varepsilon\right\} + \Pr\left\{\left\|\frac{\sum_{i=1}^n \varepsilon_i f(X_i)}{n^{1/2}}\right\|_{\mathscr{A}'_{4\delta}} > \tau\varepsilon\right\}.$$

So, by Theorem 2.14 and (3.6), it is enough to show

$$\lim_{\delta\to 0}\limsup_n \Pr\left\{ \left\| \frac{\sum_{i=1}^n \varepsilon_i f(X_i)}{n^{1/2}} \right\|_{\mathscr{A}'_{4\delta}} > \tau\varepsilon \right\} = 0. \tag{3.8}$$

Let $\mathscr{H}(\varepsilon, n) = \mathscr{A}'\setminus\{0\}$ and let

$$A(\varepsilon, n) = \left\{ \sup_{f\in\mathscr{H}(\varepsilon,n)} \frac{\sum_{i=1}^n f^2(X_i)}{nPf^2} < 4 \right\}.$$

($\mathscr{H}$ and A for brevity). We decompose the probability in (3.8) as

$$\begin{aligned} &\Pr\left\{ \left\| \frac{\sum_{i=1}^n \varepsilon_i h(X_i)}{n^{1/2}} \right\|_{\mathscr{A}'_{4\delta}} > \tau\varepsilon \right\} \\ &\quad \le \Pr\{A^c\} + (\tau\varepsilon)^{-1} E_X E_\varepsilon \left(\left\| \frac{\sum_{i=1}^n \varepsilon_i h(X_i)}{n^{1/2}} \right\|_{\mathscr{A}'_{4\delta}} 1_A \right) := \text{(I)} + \text{(II)} \end{aligned} \tag{3.9}$$

(see Remark 2.4(2) for the notation E_X and E_ε). Now we compute the limit of (I). By condition (i) there exists a centered Gaussian process $Y(f)$, $f\in\mathscr{F}$, with the covariance of G_P (as in Definition 2.10) and with all its sample paths bounded and uniformly continuous with respect to ρ_P. If g is a $N(0,1)$ variable independent of the process Y, then the process

$$W(f) = Y(f) + g(Pf), \quad f\in\mathscr{F}, \tag{3.10}$$

has uniformly continuous paths with respect to e_P and

$$E(W(f) - W(g))^2 = P(f-g)^2 \quad (= (e_P(f,g))^2).$$

So, by Theorem 2.16 (Sudakov),

$$\lim_{\lambda\to 0}\lambda(\ln N(\lambda, \mathscr{F}, e_P))^{1/2} = 0. \tag{3.11}$$

Since $\#\mathscr{A}(\varepsilon, n) \le N(\varepsilon^{1/2}/2n^{1/4}, \mathscr{F}, e_P)$, (3.11) implies the existence of $c_n = c_n(\varepsilon) \to 0$ (as $n\to\infty$) such that

$$\#\mathscr{A}(\varepsilon, n) \le \exp(c_n n^{1/2}/\varepsilon) \tag{3.12}$$

for all ε and n. Then, using (3.12), $\|f\|_\infty \le 1$ and Bernstein's inequality (2.18) we obtain

$$\begin{aligned} &\lim_{\delta\downarrow 0}\limsup_n \Pr\{A^c\} \\ &\quad \le \lim_{\delta\downarrow 0}\limsup_n (\#\mathscr{A})^2 \sup_{f\in\mathscr{H}} \Pr\{\textstyle\sum_{i=1}^n f^2(X_i) > 4nP(f^2)\} \\ &\quad \le \lim_{\delta\downarrow 0}\limsup_n \exp\{2c_n n^{1/2}/\varepsilon - 9\varepsilon n^{1/2}/16\} = 0. \end{aligned} \tag{3.13}$$

Hence, by (3.8) and (3.9), we need only show that $\lim_\delta \limsup_n \text{(II)} = 0$.

By the first inequality in Lemma 2.9 applied conditionally on $\{X_i\}$ (i.e. considering $\{X_i\}$ fixed) we can replace $\{\varepsilon_i\}$ in (II) by $\{g_i\}$. So, for each $n\in\mathbb{N}$ and $\omega \in A$ let us consider the Gaussian process

$$Z_{\omega,n}(h) = \textstyle\sum_{i=1}^n g_i h(X_i(\omega))/n^{1/2}, \quad h\in\mathscr{A}.$$

Its increments obviously satisfy, for all $h, h' \in \mathscr{H}$,

$$E_g(Z_{\omega,n}(h) - Z_{\omega,n}(h'))^2 \le E(2W(h) - 2W(h'))^2.$$

Hence, by Theorem 2.17 (b)

$$E_g \left\| \frac{\sum_{i=1}^n g_i f(X_i(\omega))}{n^{1/2}} \right\|_{\mathscr{H}'_{4\delta}} \le 8 \sup_{f\in\mathscr{F}} E \sup_{g\in\mathscr{F}, P(f-g)^2\le 4\delta} |W(f) - W(g)|$$

$$(3.14) \qquad + 26(\delta \ln N(\delta^{1/2}/2, \mathscr{F}, e_P))^{1/2}.$$

By the well known integrability properties of Gaussian vectors (Fernique, 1974, 1.3.2), and since $\sup_{f\in\mathscr{F}'_\delta} |W(f)| \to 0$ a.s., the first term on the right side of inequality (3.14) tends to zero as $\delta \to 0$. The second term also tends to zero by (3.11). Since moreover these two terms are independent of $\omega \in A$ and $n \in \mathbb{N}$, $\lim_{\delta\to 0}\lim\sup_n(\text{II}) = 0$; this proves sufficiency.

For necessity we note that (i) follows by definition and (ii) follows from Theorem 2.14 (2). □

3.3. REMARKS. (1). $\{\varepsilon_i\}$ can be replaced by $\{g_i\}$ in (3.6).
(2) As in Theorem 3.1, other distances $\rho \ge e_P$ can be used in Theorem 3.2. For instance, the following can be proved: Let $\rho \ge e_P$ be the distance induced on $\mathscr{F}$ by a sample continuous Gaussian process on $(\mathscr{F}, e_P)$ and assume that ρ is e_P-uniformly continuous and that $\rho(f, g) = 0$ if $f = g$ P—a.s. If there exist $\tau, \varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$,

$$\lim_{\delta\downarrow 0}\lim\sup_n \Pr\left\{\sup_{f,g\in {}_0\mathscr{F};\rho^2(f,g)<\varepsilon/n^{1/2}} \left| \frac{\sum_{i=1}^n \varepsilon_i(f(X_i) - g(X_i))}{n^{1/2}} \right| > \tau\varepsilon\right\} = 0,$$

then $\mathscr{F}$ is a functional P-Donsker class.

In order to compare the differences between Theorems 3.1 and 3.2, we describe a randomized version of Theorem 3.1. We skip the proof because it is completely analogous to the proof of Theorem 2.14 (except that Corollary 2.13 is not needed).

3.4. PROPOSITION. *Let $\mathscr{F}$ and ρ satisfy the hypotheses of Theorem* 3.1 (*including* (3.1)). *Assume further that $\mathscr{F}$ is* NLDM(P). *Then the following conditions are equivalent* (*where $\{\xi_i\}$ is either a Rademacher or an orthogaussian sequence*):

(a) *$\mathscr{F}$ is a functional P-Donsker class.*

(b) *There exist $\tau, \varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$*

$$(3.15) \qquad \lim_n \Pr\left\{\sup_{f,g\in {}_0\mathscr{F};\rho^2(f,g)\le\varepsilon/n^{1/2}} \left| \frac{\sum_{i=1}^n \xi_i(f(X_i) - g(X_i))}{n^{1/2}} \right| > \tau\varepsilon\right\} = 0.$$

(c)

$$(3.16) \qquad \lim_n E \sup_{f,g\in {}_0\mathscr{F};\rho^2(f,g)\le\varepsilon/n^{1/2}} \left| \frac{\sum_{i=1}^n \xi_i(f(X_i) - g(X_i))}{n^{1/2}} \right| = 0.$$

4. CLT's without measurability assumptions. In this section we apply Theorem 3.1, which does not require measurability, to derive a strengthening of a result of Dudley on metric entropy "with bracketing" (see e.g. Dudley, 1982, Theorem 6.2.1) and to prove a general CLT for classes of sets with applications (old and new). (But our main results for classes of sets, which do require measurability, are given in Section 5; the rest of the paper is independent of this section except for Theorem 4.3, so it may be skipped by those readers who are not concerned about measurability constraints.)

Given a family of measurable functions $\mathscr{G}$, and $\varepsilon > 0$, define

$$M_0(\mathscr{G}, \varepsilon, P) \quad (=M_0(\mathscr{G}, \varepsilon))$$

(4.1) $$= \inf\{r\colon \text{there exist measurable functions } f_1, \cdots, f_r \text{ such that for each } g \in \mathscr{G}, f_i \geq |g| \text{ and } \textstyle\int (f_i - |g|)\, dP < 2^{-1}\varepsilon \text{ for some } 1 \leq i \leq r\}.$$

For any family of functions $\mathscr{F}$, pseudo-distance ρ, $\varepsilon > 0$, and $n \in \mathbb{N}$, set

(4.2) $$\mathscr{F}'(\varepsilon, n, \rho) = \{f - g\colon f, g \in \mathscr{F}, \rho(f, g) \leq \varepsilon/n^{1/2}\}.$$

Finally, let $\| f \|_1 = P|f|$. With this notation we have:

4.1. THEOREM. *Let $\mathscr{F}$ be a uniformly bounded family of measurable functions on S. Assume:*

(a) $\int_0 \ln(N(\varepsilon, \mathscr{F}, \|\cdot\|_1^{1/2}))^{1/2}\, d\varepsilon < \infty$

(b) $n^{-1/2}\ln M_0(\mathscr{F}'(\varepsilon, n, \|\cdot\|_1), \varepsilon/n^{1/2}, P) \to 0$ *as* $n \to \infty$ *for all* $\varepsilon > 0$.

Then $\mathscr{F}$ is a functional P-Donsker class.

PROOF. We may assume without loss of generality that $\sup_{f\in\mathscr{F}} \| f \|_\infty \leq 1/2$. Obviously the pseudo-distance $\rho(f, g) = \| f - g \|_1^{1/2}$ verifies the hypotheses of Theorem 3.1. Hence, by (a), the theorem will be proved if we show that for some $\tau > 0$ and all $\varepsilon > 0$,

(4.3) $$\lim_{n\to\infty}\Pr{}^*\left\{\sup_{f\in\mathscr{F}'(\varepsilon,n,\|\cdot\|_1)} \left| \frac{\sum_{i=1}^n (f(X_i) - Pf)}{n^{1/2}} \right| > \tau\varepsilon\right\} = 0.$$

Since $P|f| \leq \varepsilon/n^{1/2}$ for $f \in \mathscr{F}'(\varepsilon, n, \|\cdot\|_1)$, it will be enough to show

(4.4) $$\lim_{n\to\infty}\Pr{}^*\left\{\sup_{f\in\mathscr{F}'(\varepsilon, n, \|\cdot\|_1)} \frac{\sum_{i=1}^n |f(X_i)|}{n^{1/2}} > \sigma\varepsilon\right\} = 0$$

for some $\sigma > 0$ and all $\varepsilon > 0$. We may assume $\sigma > 3/2$. Let $m = M_0(\mathscr{F}'(\varepsilon, n, \|\cdot\|_1), \varepsilon/n^{1/2}, P)$ and $f_1, \cdots, f_m$ be the functions given by the definition of M_0 (in (4.1)), which we may assume to be bounded by 1. Then

(4.5) $$\sup_{f\in\mathscr{F}'(\varepsilon, n, \|\cdot\|_1)} \frac{\sum_{i=1}^n |f(X_i)|}{n^{1/2}} \leq \max_{j\leq m} \frac{\sum_{i=1}^n |f_j(X_i)|}{n^{1/2}}.$$

Taking into account that $P(|f_j| - P|f_j|)^2 \leq P|f_j| \leq 3\varepsilon/2n^{1/2}$ we have that for each j,

$$\Pr\{\sum_{i=1}^{n} |f_j(X_i)| > \sigma\varepsilon n^{1/2}\}$$

$$(4.6) \qquad \le \Pr\{\sum_{i=1}^{n} (|f_j(X_i)| - E|f_j(X_i)|) \ge (\sigma - 3/2)\varepsilon n^{1/2}\}$$

$$\le \exp\{-(\sigma - 3/2)^2(2 + 2\sigma/3)^{-1}\varepsilon n^{1/2}\}$$

by Bernstein's inequality (2.18). Hence, by (4.5) and (4.6), the probability in (4.4) is bounded by

$$(\exp\{-(\sigma - 3/2)^2(2 + 2\sigma/3)^{-1}\varepsilon n^{1/2}\})M_0(\mathscr{F}(\varepsilon, n, \|\cdot\|_1), \varepsilon/n^{1/2}) \to 0. \quad \square$$

4.2. REMARK. (1) Dudley (1982) defines metric entropy with bracketing for a class $\mathscr{F} \subset \mathscr{L}_1(S, \mathscr{S}, P)$ as $\log N_{[\,]}^{(1)}(\varepsilon, \mathscr{F}, P)$, where $N_{[\,]}^{(1)}(\varepsilon, \mathscr{F}, P)$ is the smallest r such that for some $f_1, \cdots, f_r \in \mathscr{L}_1(P)$ and for each $f \in \mathscr{F}$ there exists some i, $j \le r$ with $f_i \le f \le f_j$ and $\|f_j - f_i\|_1 \le \varepsilon$. Then it is obvious that

$$M_0(\mathscr{F}'(\varepsilon, n, \|\cdot\|_1), \varepsilon/n^{1/2}) \le 2[N_{[\,]}^{(1)}(\varepsilon/4n^{1/2}, \mathscr{F})]^4.$$

(Use as the approximating functions for $\mathscr{F}'(\varepsilon, n, \|\cdot\|_1)$ the functions $\max(f_j - f_k, f_\ell - f_m)$, where the f_i's come from the definition of $N_{[\,]}^{(1)}$.) Hence, hypothesis (b) in Theorem 4.1 may be replaced by

$$(\mathrm{b}') \qquad \lim_{\varepsilon\to 0} \varepsilon \ln N_{[\,]}^{(1)}(\varepsilon, \mathscr{F}, P) = 0.$$

Hypotheses (a) and (b′) are weaker than the condition

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

and therefore, Theorem 4.1 implies Theorem 6.2.1 in Dudley (1982).

(2) Here we translate Theorem 4.1 to sets. It includes Theorem 5.1 on metric entropy "with inclusion" in Dudley (1978) (see page 917 and 918 in the same paper to appreciate its wide applicability). For any class of measurable subsets of S, $\mathscr{D}$, define

$$M_0(\mathscr{D}, \varepsilon, P) = M_0(\mathscr{D}, \varepsilon)$$

$$(4.7) \qquad = \inf\{r: r \ge 1, \exists A_1, \cdots, A_r \in \mathscr{S} \text{ such that for all } A \in \mathscr{D} \text{ and for some } j, A \subseteq A_j \text{ and } P(A_j \setminus A) \le \varepsilon/2\}.$$

Given a class of measurable sets, here and in what follows, we write

$$(4.8) \qquad \mathscr{C}' = \{A \,\Delta\, B: A, B \in \mathscr{C}\} \quad \text{and} \quad \mathscr{C}'_{\varepsilon,n} = \{C \in \mathscr{C}': PC \le \varepsilon/n^{1/2}\}$$

(for $\varepsilon > 0$ and $n \in \mathbb{N}$), where Δ denotes symmetric difference; and we denote by $N(\lambda, \mathscr{C}, P)$ the λ-covering number of $\mathscr{C}$ for the distance $d(A, B) = P(A \,\Delta\, B)$. With this notation, Theorem 4.1 for classes of sets $\mathscr{C}$ becomes: if

$$(1) \qquad \int_0^1 (\ln N(\varepsilon^2, \mathscr{C}, P))^{1/2}\, d\varepsilon < \infty$$

and

(2) $$n^{-1/2}\ln M_0(\mathscr{C}'_{\varepsilon,n}, \varepsilon/n^{1/2}, P) \to 0 \quad \text{as} \quad n \to \infty, \quad \text{for all} \quad \varepsilon > 0,$$

then $\mathscr{C}$ is a functional P-Donsker class of sets.

Next we consider classes of sets (i.e., classes of indicator functions). The difference with the case of functions is that for sets we can dispense with the centering, even in the nonrandomized version of Theorem 3.1: if $e_P^2(I_A, I_B) = P(A \,\Delta\, B) \le \varepsilon/n^{1/2}$, then $n^{1/2}\,|\,PA - PB\,| \le \varepsilon$. So, Theorem 3.1 becomes

4.3. THEOREM. *Let $\mathscr{C} \subset \mathscr{S}$ satisfy the entropy condition*

$$\int_0^1 (\ln N(\varepsilon^2, \mathscr{C}, P))^{1/2}\, d\varepsilon < \infty. \tag{4.9}$$

Then $\mathscr{C}$ is a functional P-Donsker class of sets if and only if there exist τ, $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$,

$$\lim_{n\to\infty}\Pr{}^*\left\{\sup_{A,B\in\mathscr{C};P(A\Delta B)\le\varepsilon/n^{1/2}}\left|\frac{\sum_{i=1}^n(\delta_{X_i}(A) - \delta_{X_i}(B))}{n^{1/2}}\right| > \tau\varepsilon\right\} = 0. \tag{4.10}$$

Sometimes, $\delta_{X_i}(A \,\Delta\, B) = |\,\delta_{X_i}(A) - \delta_{X_i}(B)\,|$ is easier to deal with than $\delta_{X_i}(A) - \delta_{X_i}(B)$. So, we state the following

4.4. COROLLARY. *Let $\mathscr{C} \subset \mathscr{S}$ satisfy* (4.9). *Then $\mathscr{C}$ is a functional P-Donsker class of sets if there exist τ, $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$,*

$$\lim_{n\to\infty}\Pr{}^*\left\{\sup_{A,B\in\mathscr{C};P(A\Delta B)\le\varepsilon/n^{1/2}}\frac{\sum_{i=1}^n \delta_{X_i}(A \,\Delta\, B)}{n^{1/2}} \ge \tau\varepsilon\right\} = 0. \tag{4.11}$$

4.5. REMARK. (a) If $\mathscr{C}$ is countable, condition (4.11) is also necessary for the CLT. To see this note that for X_i fixed, the (conditionally) Gaussian process $\sum_{i=1}^n g_i(\delta_{X_i}(A) - \delta_{X_i}(B))$ dominates the process $\sum_{i=1}^n g_i\delta_{X_i}(A \,\Delta\, B)$ in the sense of Theorem 2.17; then the result follows by application of Theorem 2.17 (as in (3.14)) and Proposition 3.4.

(b) Note that the conclusions of both Theorem 4.3 and Corollary 4.4 still hold if we replace condition (4.9) by $\mathscr{C}$ being NLDM(P) and P-pregaussian.

The main result in what follows is Theorem 4.8 where condition (4.11) is replaced by a rate of convergence to zero of the probability that $[\tau\varepsilon n^{1/2}]$ random variables X_i (i.i.d. with law P) fall simultaneously in some set $A \,\Delta\, B$, $A, B \in \mathscr{C}$, with $P(A \,\Delta\, B) \le \varepsilon/n^{1/2}$. From this we rederive several results from Dudley (1978) and consider also a new situation.

We let $r = r(\tau, \varepsilon, n)$ denote the smallest integer larger than or equal to $\tau\varepsilon n^{1/2}$ ($\tau, \varepsilon > 0$, $n \in \mathbb{N}$).

4.6. LEMMA. *Let $\{X_i\}$ be i.i.d. with law P, and $\mathscr{S}$ a family of measurable subsets of S. Then*

$$\Pr{}^*\left\{\sup_{A\in\mathscr{S}'_{\varepsilon,n}} \frac{\sum_{k=1}^n \delta_{X_k}(A)}{n^{1/2}} \geq 3\tau\varepsilon\right\}$$

$$\leq \binom{n}{r}\,[2\Pr{}^*\{\sup_{A\in\mathscr{S},PA\leq 3\varepsilon/n^{1/2}}\delta_{X_1}(A)\cdots\delta_{X_r}(A)=1\}$$

$$+\Pr{}^*\{\sup_{A,B\in\mathscr{S},PA,PB>2\varepsilon/n^{1/2};P(A\Delta B)\leq\varepsilon/n^{1/2}}\delta_{X_1}(A\,\Delta\,B)\cdots\delta_{X_r}(A\,\Delta\,B)=1\}].$$

PROOF.

$$\Pr{}^*\left\{\sup_{A,B\in\mathscr{S};P(A\Delta B)\leq\varepsilon/n^{1/2}} \frac{\sum_{k=1}^n \delta_{X_k}(A\,\Delta\,B)}{n^{1/2}} \geq 3\tau\varepsilon\right\}$$

$$\leq \Pr{}^*\left\{\sup_{A,B\in\mathscr{S};PA,PB\leq 3\varepsilon/n^{1/2}} \frac{\sum_{k=1}^n \delta_{X_k}(A)}{n^{1/2}} + \frac{\sum_{k=1}^n \delta_{X_k}(B)}{n^{1/2}} \geq 2\tau\varepsilon\right\}$$

$$+ \Pr{}^*\left\{\sup_{A,B\in\mathscr{S};P(A\Delta B)\leq\varepsilon/n^{1/2};PA,PB>2\varepsilon n^{1/2}} \frac{\sum_{k=1}^n \delta_{X_k}(A\,\Delta\,B)}{n^{1/2}} \geq \tau\varepsilon\right\}$$

$$\leq 2\Pr{}^*\left\{\sup_{A\in\mathscr{S};PA\leq 3\varepsilon/n^{1/2}} \frac{\sum_{k=1}^n \delta_{X_k}(A)}{n^{1/2}} \geq \tau\varepsilon\right\}$$

$$+ \Pr{}^*\left\{\sup_{A,B\in\mathscr{S};P(A\Delta B)\leq\varepsilon/n^{1/2};PA,PB>2\varepsilon/n^{1/2}} \frac{\sum_{k=1}^n \delta_{X_k}(A\,\Delta\,B)}{n^{1/2}} \geq \tau\varepsilon\right\}.$$

Now,

$$\Pr{}^*\left\{\sup_{A\in\mathscr{S};PA\leq 3\varepsilon/n^{1/2}} \frac{\sum_{k=1}^n \delta_{X_k}(A)}{n^{1/2}} \geq \tau\varepsilon\right\}$$

$$= \Pr{}^*[\cup_{j_1<\cdots<j_{r(n,\tau,\varepsilon)}}\{\sup_{A\in\mathscr{S};PA\leq 3\varepsilon/n^{1/2}}\delta_{X_{j_1}}(A)\cdots\delta_{X_{j_r}}(A)=1\}]$$

$$\leq \binom{n}{r}\Pr{}^*\{\sup_{A\in\mathscr{S};PA\leq 3\varepsilon/n^{1/2}}\delta_{X_1}(A)\cdots\delta_{X_r}(A)=1\},$$

and similarly

$$\Pr{}^*\left\{\sup_{A,B\in\mathscr{S};P(A\Delta B)\leq\varepsilon/n^{1/2};PA,PB>2\varepsilon/n^{1/2}} \frac{\sum_{k=1}^n \delta_{X_k}(A\,\Delta\,B)}{n^{1/2}} \geq \tau\varepsilon\right\}$$

$$\leq \binom{n}{r}\Pr{}^*\{\sup_{A,B\in\mathscr{S};P(A\Delta B)\leq\varepsilon/n^{1/2};PA,PB>2\varepsilon/n^{1/2}}\delta_{X_1}(A\,\Delta\,B)\cdots\delta_{X_r}(A\,\Delta\,B)=1\}.$$

The lemma is proved. □

4.7. REMARK. Exponentials are easier to use than factorials. Let us note to this effect that, since $\ln r! > \int_1^r \ln x\,dx \geq r\ln r - r = \ln(r/e)^r$, we have

$$\binom{n}{r} \leq \left(\frac{en}{r}\right)^r,$$

and that $(en/r)^r \uparrow$ for $0 < r \leq n$. It is worth mentioning that the argument in the last part of Lemma 4.6 and this inequality give:

$$P\{\mathrm{Bin}(n, p) \geq k\} \leq \binom{n}{k}p^k \leq \left(\frac{enp}{k}\right)^k.$$

This inequality is comparable to the usual exponential inequalities.

Combining this remark, Lemma 4.6 and Corollary 4.4, we obtain the following theorem, which is the main result of this section.

4.8. THEOREM. *If*

(1) $$\int_0^1 (\ln N(\varepsilon^2, \mathscr{C}, P))^{1/2}\,d\varepsilon < \infty,$$

(2) $$\lim_{n\to\infty}\left(\frac{en}{r(\tau, \varepsilon, n)}\right)^{r(\tau,\varepsilon,n)} \Pr{}^*\{\sup\nolimits_{A\in\mathscr{C};PA\leq\varepsilon/n^{1/2}}\delta_{X_1}(A)\cdots\delta_{X_{r(\varepsilon,\tau,n)}}(A) = 1\} = 0,$$

(3) $$\lim_{n\to\infty}\left(\frac{en}{r(\tau, \varepsilon, n)}\right)^{r(\varepsilon,\tau,n)}$$
$$\cdot \Pr{}^*\{\sup\nolimits_{A,B\in\mathscr{C};P(A\Delta B)\leq\varepsilon/n^{1/2};PA,PB>2\varepsilon/n^{1/2}}\delta_{X_1}(A\,\Delta\,B)\cdots\delta_{X_{r(\varepsilon,\tau,n)}}(A\,\Delta\,B) = 1\}$$
$$= 0$$

for all $0 < \varepsilon \leq \varepsilon_0$, *for some* ε_0, $\tau > 0$, *then* $\mathscr{C}$ *is a functional P-Donsker class of sets.*

4.9. COROLLARY. *If*

(1) $$\int_0^1 (\ln N(\varepsilon^2, \mathscr{C}, P))^{1/2}\,d\varepsilon < \infty$$

and

(2′) $$\lim_{n\to\infty}\left(\frac{en}{r(\tau, \varepsilon, n)}\right)^{r(\tau,\varepsilon,n)} \Pr{}^*\{\sup\nolimits_{A\in\mathscr{C}'_{\varepsilon,n}}\delta_{X_1}(A)\cdots\delta_{X_{r(\tau,\varepsilon,n)}}(A) = 1\} = 0$$

for some $\tau > 0$, $\varepsilon_0 > 0$ *and all* $\varepsilon_0 > \varepsilon > 0$, *then* $\mathscr{C}$ *is a functional P-Donsker class of sets.*

PROOF. Obvious (assume $\phi \in \mathscr{C}$). □

4.10. REMARK. Condition (2′) in Corollary 4.9 is not far from being necessary for the CLT. In fact, it is not difficult to check, using Kuelbs' or de Acosta's exponential bounds for Banach space valued r.v.'s (see e.g. de Acosta, 1981, Theorem 2.2) that

$$\Pr{}^*\{\sup_{A\in\mathscr{S};PA\le\varepsilon/n^{1/2}}\delta_{X_1}(A)\cdots\delta_{X_r}(A)=1\}$$

$$\le \Pr{}^*\left\{\sup_{A\in\mathscr{S}'}\frac{\sum_{i=1}^r(\delta_{X_i}(A)-P(A))}{r}\ge 1-\frac{\varepsilon}{n^{1/2}}\right\}\le c\exp(-dn^{1/2})$$

for some $c, d \in (0, \infty)$.

A first consequence of this result is another (easier) proof of the CLT under metric entropy with inclusion (Remark 4.2 (2)), which follows directly from the simple observation that

(4.12) $\Pr^*\{\sup_{A\in\mathscr{S}'_{\varepsilon,n}}\delta_{X_1}(A)\cdots\delta_{X_r}(A)=1\}\le(3\varepsilon/2n^{1/2})^r M_0(\mathscr{S}'_{\varepsilon,n}, \varepsilon/n^{1/2}, P).$

We skip the details.

As another consequence of Theorem 4.8 we prove a CLT result for sequences of sets under summability conditions for their probabilities (Dudley, 1978, Theorem 2.1).

4.11. COROLLARY. *Let $\mathscr{C}=\{A_m\}\subset\mathscr{S}$, and let $p_m = PA_m$. If for some $s > 0$ $\sum_{m=1}^{\infty} p_m^s < \infty$, then $\mathscr{C}$ is a functional P-Donsker class of sets.*

PROOF. We may assume $p_m\downarrow$, $p_m < 1/2$, hence that $p_m < m^{-1/s}$ (we may discard a finite number of sets A_m from our collection if necessary). Then $N(m^{-1/s}, \mathscr{C}, P)\le m$, and therefore the metric entropy condition (1) of Theorem 4.8 holds for $\mathscr{C}$ and P. Next we verify condition (2) in Theorem 4.8. With $r = r(\tau, \varepsilon, n)$ and $\tau > e$, we have

$$\left(\frac{en}{r}\right)^r\Pr\{\sup_{A_m:PA_m\le\varepsilon/n^{1/2}}\delta_{X_1}(A_m)\cdots\delta_{X_r}(A_m)=1\}$$

$$\le\left(\frac{en}{r}\right)^r\sum_{\{m:PA_m\le\varepsilon/n^{1/2}\}}(PA_m)^r\le\left(\frac{en}{r}\right)^r\left[\sum_{m\le(n^{1/2}/\varepsilon)^s}\left(\frac{\varepsilon}{n^{1/2}}\right)^r+\sum_{m>(n^{1/2}/\varepsilon)^s}m^{-r/s}\right]$$

$$\le\left(\frac{e}{\tau}\right)^{\tau\varepsilon n^{1/2}}\left(\frac{n^{1/2}}{\varepsilon}\right)^s+\left(\frac{e}{\tau}\right)^{\tau\varepsilon n^{1/2}}\left(\frac{n^{1/2}}{\varepsilon}\right)^s(\tau\varepsilon n^{1/2}s^{-1}-1)^{-1}\to 0\quad\text{as}\quad n\to\infty.$$

As for (3) in Theorem 4.8, we observe that there are at most $(n^{1/2}/2\varepsilon)^s$ sets A_m with $PA_m > 2\varepsilon/n^{1/2}$, and therefore,

$$\left(\frac{en}{r}\right)^r\Pr\{\sup_{PA_m,PA_\ell>2\varepsilon/n^{1/2}}\delta_{X_1}(A_m\,\Delta\,A_\ell)\cdots\delta_{X_r}(A_m\,\Delta\,A_\ell)=1\}$$

$$\le\left(\frac{en}{r}\right)^r\left(\frac{\varepsilon}{n^{1/2}}\right)^r\left(\frac{n^{1/2}}{2\varepsilon}\right)^{2s}\le\left(\frac{e}{\tau}\right)^{\tau\varepsilon n^{1/2}}\left(\frac{n^{1/2}}{2\varepsilon}\right)^{2s}\to 0\quad\text{as}\quad n\to\infty.$$

Now the result follows from Theorem 4.8. □

Next we derive from Theorem 4.3 the CLT for VC classes (in the following sections we give other proofs of this result; here we are only interested in showing that it follows from Theorem 4.3). This is one of the main results in Dudley (1978). The measurability hypothesis required in this proof is different from that in Dudley (1978) and weaker than the one required in the proof given in Section 6.

4.12. COROLLARY. *Let $\mathscr{C}$ be a* NLDM(P) VC *class. Then $\mathscr{C}$ is a functional P-Donsker class.*

4.13. REMARK. To prove this corollary from Corollary 4.9 we need to use an exponential bound. But this need not be the strongest; in fact, the Vapnik-Červonenkis bound (1971, Theorem 2) suffices.

PROOF. By Lemma 7.13 in Dudley (1978) (see Theorem 2.23) the metric entropy condition (1) in Corollary 4.9 is satisfied. So it is enough to prove (2′). By NLDM(P) it is enough to prove (2′) for ${}_0\mathscr{C}$. For ease of notation we use also $\mathscr{C}$ for ${}_0\mathscr{C}$.

Given $m > 0$, let G be the permutation group of $\{1, \cdots, m\}$ and let μ be the uniform probability on $(G, 2^G)$. Denote elements of G by $\pi = (\pi_1, \cdots, \pi_m)$. Define new variables Y_i: $(S^{\mathbb{N}} \times G, \mathscr{S}^{\mathbb{N}} \times 2^G, \Pr \times \mu) \to \mathbb{R}$ (where $\Pr = P^{\mathbb{N}}$) by

$$Y_i(\omega, \omega') = X_{\pi_i(\omega')}(\omega).$$

Since $\mathscr{C}$ is VC, so is $\mathscr{C}'$. Let $v = V(\mathscr{C}')$. For $m > 1$, let $\Delta_m(\omega) = \Delta_{\epsilon,n}^{\mathscr{C}'}(X_1(\omega), \cdots, X_m(\omega))$. Then, by Theorem 2.22 $\Delta_m(\omega) \le m^v$, and if we call $A_1(\omega), \cdots, A_{\Delta_m(\omega)}(\omega)$ the $\Delta_m(\omega)$ different sets in $\mathscr{C}'_{\epsilon,n} \cap \{X_1(\omega), \cdots, X_m(\omega)\}$, we have, for $m \ge r$,

$$\begin{aligned}
&\mu\{\omega': \sup_{A\in\mathscr{C}'_{\epsilon,n}} \delta_{Y_1(\omega,\omega')}(A) \cdots \delta_{Y_r(\omega,\omega')}(A) = 1\}\\
&\quad\le \textstyle\sum_{i=1}^{\Delta_m(\omega)} \mu\{\omega': \delta_{X_{\pi_1(\omega')}(\omega)}(A_i(\omega)) \cdots \delta_{X_{\pi_r(\omega')}}(A_i(\omega)) = 1\}\\
&\quad= \textstyle\sum_{i=1}^{\Delta_m(\omega)} \binom{mP_m(A_i(\omega))}{r} \Big/ \binom{m}{r} \quad \text{where we let } \binom{\ell}{r} = 0 \text{ if } \ell < r \qquad (4.13)\\
&\quad\le m^v \sup_{A\in\mathscr{C}'_{\epsilon,n}} \left(\frac{1}{m} \textstyle\sum_{j=1}^{m} \delta_{X_j(\omega)}(A)\right)^r.
\end{aligned}$$

Since G is finite, the measurability assumption implies that the variable $\sup_{A\in\mathscr{C}'_{\epsilon,n}} \delta_{Y_1}(A) \cdots \delta_{Y_r}(A)$ is jointly measurable in (ω, ω'). Then, since obviously

$$\Pr\{\sup_{A\in\mathscr{C}'_{\epsilon,n}} \delta_{Y_1}(A) \cdots \delta_{Y_r}(A) = 1\} = \Pr\{\sup_{A\in\mathscr{C}'_{\epsilon,n}} \delta_{X_1}(A) \cdots \delta_{X_r}(A) = 1\}$$

for all $\omega' \in G$, the bound (4.13) gives

$$\begin{aligned}
&\Pr\{\sup_{A\in\mathscr{C}'_{\epsilon,n}} \delta_{X_1}(A) \cdots \delta_{X_r}(A) = 1\} \qquad (4.14)\\
&\quad= E_\omega \mu\{\sup_{A\in\mathscr{C}'_{\epsilon,n}} \delta_{Y_1}(A) \cdots \delta_{Y_r}(A) = 1\} \le m^v E \| (1/m) \textstyle\sum_{i=1}^{m} \delta_{X_i} \|^r_{\mathscr{C}'_{\epsilon,n}}.
\end{aligned}$$

Now by Theorem 2 in Vapnik–Červonenkis (1971) we get

$$\Pr\left\{ \left\| \frac{1}{m} \sum_{j=1}^{m} \delta_{X_i}(A) \right\|_{\mathscr{S}'_{\varepsilon,n}} > \frac{2\varepsilon}{n^{1/2}} \right\}$$

$$\le \Pr\left\{ \left\| \frac{1}{m} \sum_{j=1}^{m} (\delta_{X_j} - P) \right\|_{\mathscr{S}'} > \frac{\varepsilon}{n^{1/2}} \right\} \le 4(2m)^v \exp\left\{ -\frac{\varepsilon^2 m}{8n} \right\}$$

for $m \ge 2n/\varepsilon^2$. We take $m = [2^{r/v}]$ and $\tau = 8e$. Then by integrating over the event

$$\left\{ \sup_{A \in \mathscr{S}'_{\varepsilon,n}} \frac{1}{m} \sum_{j=1}^{m} \delta_{X_j}(A) > \frac{2\varepsilon}{n^{1/2}} \right\}$$

and over its complement in the last term of (4.14) we obtain

$$\Pr\{\sup_{A \in \mathscr{S}'_{\varepsilon,n}} \delta_{X_1}(A) \cdots \delta_{X_r}(A) = 1\} \le m^v \left[\left(\frac{2\varepsilon}{n^{1/2}} \right)^r + 2^{v+2} m^v \exp - \left\{ \frac{\varepsilon^2 m}{8n} \right\} \right].$$

Therefore,

$$\left(\frac{en}{r} \right)^r \Pr\{\sup_{A \in \mathscr{S}'_{\varepsilon,n}} \delta_{X_1}(A) \cdots \delta_{X_r}(A) = 1\} \le 2\left(\frac{1}{2} \right)^r$$

for n large enough. This proves 4.12 by Corollary 4.9. □

Let us finally apply Theorem 4.8 in a situation which does not fall in any of the last three cases.

4.14. COROLLARY. *Let $\{\mathscr{S}_m\}_{m=1}^{\infty}$ be a countable number of independent collections of measurable sets and let $\mathscr{S} = \bigcup_{m=1}^{\infty} \mathscr{S}_m$. Assume*:

(a) $$\sum_m [\sup_{A \in \mathscr{S}_m} PA]^s < \infty \quad \textit{for some} \quad s > 0$$

and either

(b) $$\varepsilon \ln[\sup_m M_0(\mathscr{S}'_m, \varepsilon, P)] \to \infty \quad \textit{as} \quad \varepsilon \to 0 \quad \textit{and}$$

$$\int_0 (\ln N(\varepsilon^2, \mathscr{S}, P))^{1/2}\, d\varepsilon < \infty$$

or

(b′) $\mathscr{S}_m$, $m = 1, \cdots$, *are* NLDM(P) VC *classes with uniformly bounded* VC *constants.*

Then $\mathscr{S}$ is a functional P-Donsker class of sets.

PROOF. As in Corollary 4.11 we may assume without loss of generality that $p_m = \sup_{A \in \mathscr{S}_m} PA$ decreases, $p_m < 1/2$ and $p_m < m^{-1/s}$. Then, under either (b) or (b′), $\mathscr{S}$ satisfies the metric entropy condition. Let us now assume (b). To prove (2) in Theorem 4.8 we assume $\tau > 3e/2$ and n large enough so that

$M_0(\mathscr{S}_m, \varepsilon/n^{1/2}) \le (3e/2\tau)^{(1/2)\tau\varepsilon n^{1/2}}$. We then have

$$\left(\frac{en}{r}\right)^r \Pr\{\sup_{A\in\mathscr{S};PA\le\varepsilon/n^{1/2}} \delta_{X_1}(A)\cdots\delta_{X_r}(A) = 1\}$$

$$\le \left(\frac{en}{r}\right)^r \sum_m \Pr\{\sup_{A\in\mathscr{S};PA\le\varepsilon/n^{1/2}} \delta_{X_1}(A)\cdots\delta_{X_r}(A) = 1\}$$

$$\le \left(\frac{en}{r}\right)^r \sum_m M_0\left(\mathscr{S}_m, \frac{\varepsilon}{n^{1/2}}\right)\left(\frac{3\varepsilon}{n^{1/2}} \wedge \frac{1}{m^{1/s}}\right)^r$$

$$\le \left(\frac{en}{r}\right)^r \left(\frac{2n^{1/2}}{3\varepsilon}\right)^s \left(\frac{3e}{2\tau}\right)^{-(1/2)\tau\varepsilon n^{1/2}} \left(\frac{3\varepsilon}{2n^{1/2}}\right)^r$$

$$+ \left(\frac{en}{r}\right)^r \left(\frac{3e}{2\tau}\right)^{-(1/2)\tau\varepsilon n^{1/2}} \frac{1}{\tau\varepsilon n^{1/2}s^{-1} - 1} \left(\frac{3\varepsilon}{2n^{1/2}}\right)^{r-s}$$

$$= \left(\frac{3e}{2\tau}\right)^{(1/2)\tau\varepsilon n^{1/2}} \left(\frac{2n^{1/2}}{3\varepsilon}\right)^s$$

$$+ \left(\frac{3e}{2\tau}\right)^{(1/2)\tau\varepsilon n^{1/2}} \left(\frac{2n^{1/2}}{3\varepsilon}\right)^s \frac{1}{\tau\varepsilon n^{1/2}s^{-1} - 1} \to 0 \quad \text{as} \quad n \to \infty.$$

To prove (3) in Theorem 4.8 let us notice that $PA, PB > 2\varepsilon/n^{1/2}$, $PA, PB < \frac{1}{2}$, $P(A \,\Delta\, B) \le \varepsilon/n^{1/2}$ is impossible if A and B are independent. Therefore,

$$\left(\frac{en}{r}\right)^r P\{\sup_{A,B\in\mathscr{S};PA,PB>2\varepsilon/n^{1/2};P(A\Delta B)\le\varepsilon/n^{1/2}} \delta_{X_1}(A \,\Delta\, B)\cdots\delta_{X_r}(A \,\Delta\, B) = 1\}$$

$$\le \left(\frac{en}{r}\right)^r \sum_{m\le(n^{1/2}/2\varepsilon)^s} P\{\sup_{A,B\in\mathscr{S}_m';P(A\Delta B)\le\varepsilon/n^{1/2}} \delta_{X_1}(A \,\Delta\, B)\cdots\delta_{X_r}(A \,\Delta\, B) = 1\}$$

and if we assume n large enough so that

$$M_0\left(\mathscr{S}'_m, \frac{\varepsilon}{n^{1/2}}\right) \le \left(\frac{3e}{2\tau}\right)^{-(1/2)\tau\varepsilon n^{1/2}},$$

then, as in (4.12) this is bounded by

$$\left(\frac{en}{r}\right)^r \left(\frac{n^{1/2}}{2\varepsilon}\right)^s \left(\frac{3e}{2\tau}\right)^{-(1/2)\tau\varepsilon n^{1/2}} \left(\frac{3\varepsilon}{2n^{1/2}}\right)^r \le \left(\frac{3e}{2\tau}\right)^{(1/2)\tau\varepsilon n^{1/2}} \left(\frac{n^{1/2}}{2\varepsilon}\right)^s \to 0$$

as $n \to \infty$, assuming $2\tau > 3e$.

This shows that if $\mathscr{S}$ satisfies (a) and (b) then it is a P-Donsker class. A similar proof, using computations similar to those of Corollary 4.10 (instead of (4.12)), gives the CLT under (a) and (b′). We skip it. □

4.15. EXAMPLE. Take $(X, \Sigma, P) = ([0, 1]^{\mathbb{N}}, B^{\mathbb{N}}, \lambda^{\mathbb{N}})$, and

$$\mathscr{L}_m = \{[0, 1] \times \cdots \times [0, 1] \times A \times [0, 1] \times \cdots: \quad A = [a, b],\ a, b \in [0, 1],\ b - a \leq (1/m)\}.$$

Then $\mathscr{L} = \cup \mathscr{L}_m$ is a P-Donsker class by Corollary 4.14.

4.16. REMARK. Corollary 4.15 cannot be improved to general $\{\mathscr{L}_m\}$ since for example if $\mathscr{L}_m = \{C \in \Sigma: \#C \leq m\}$, then $\mathscr{L}$ = all finite subsets, which is not Donsker. Also, condition (b′) is not stated in full generality: the VC constants of $\mathscr{L}_m$ can be allowed to increase moderately, but we do not pursue this issue, since this is just an illustration of our main results.

5. The CLT under random entropy conditions. In this section we replace the equicontinuity conditions (3.15) from Proposition 3.4 and (3.6) from Theorem 3.2 by conditions expressed in terms of the random entropies $N_{n,p}$ defined in (2.31), obtaining sufficient conditions for the CLT in the case of families of functions, and necessary and sufficient conditions in the case of families of sets.

(a) *Families of functions.* Although we present results in the i.i.d. case and for a single class $\mathscr{F}$, since our estimates are made for each fixed n and since i.i.d. is not required for randomization, most of the results in this section admit generalizations to non-i.i.d. variables and to varying classes of functions $\mathscr{F}_n$, as considered by Le Cam (1983a).

The starting points are either Theorem 3.1 (or Proposition 3.4) or, preferably, Theorem 3.2.

5.1. THEOREM. *Let $\mathscr{F}$ be a* NDM(P) *uniformly bounded class of real measurable functions on S such that:*

(i) *$\mathscr{F}$ is P-pregaussian,*

(ii) *there exist γ, σ, $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$,*

$$\lim_n \Pr{}^*\left\{\frac{\ln N_{n,1}(\gamma\varepsilon/n^{1/2}, {}_0\mathscr{F}'_{\varepsilon,n})}{n^{1/2}} > \sigma\varepsilon\right\} = 0. \tag{5.1}$$

Then $\mathscr{F}$ is a functional P-Donsker class.

The proof of Theorem 5.1 is based upon Theorem 3.2 and the lemma that follows, which elaborates on a technique of Le Cam (1983b). This lemma, the "square root trick", seems to be of independent interest (see Pollard's forthcoming book for applications in other situations).

If $\mathscr{G}^{1/2}$ is a class of functions, we let

$$\mathscr{G} = \{h^2: h \in \mathscr{G}^{1/2}\}.$$

5.2. LEMMA. *Let $\mathscr{G}^{1/2}$ be a class of functions such that $\mathscr{G}$ is* NDM(P) *and* *uniformly bounded, say, by r. Set*

$$M_n = n^{1/2}\sup_{g\in\mathscr{G}} Pg, \quad n \in \mathbb{N}. \tag{5.2}$$

Let $t > 0$ *and* $\rho > 0$ *be such that* $\lambda := 2^{-1/2}t^{1/2} - 2^{1/2}M_n^{1/2} - 2\rho > 0$. *Then for all* n, $m > 0$,

$$\begin{aligned}&\Pr{}^*\{\sup_{g\in\mathscr{G}} \textstyle\sum_{i=1}^n g(X_i) > tn^{1/2}\}\\ &\qquad \le 4 \Pr{}^*\{N_{n,2}(\rho/n^{1/4}, \mathscr{G}^{1/2}) > m\} + 8m \exp(-\lambda^2 n^{1/2}/2r).\end{aligned} \tag{5.3}$$

PROOF. As usual, we write $\mathscr{G}$ for ${}_0\mathscr{G}$. Let

$$N_+(g) = \textstyle\sum_{\{i\le n:\varepsilon_i=1\}} g(X_i), \quad N_-(g) = \sum_{\{i\le n:\varepsilon_i=-1\}} g(X_i), \quad g \in \mathscr{G}.$$

Then N_+ and N_- are equidistributed, they are conditionally independent given $\{\varepsilon_i\}_{i=1}^n$, $N_+(g) - N_-(g) = \sum_{i=1}^n \varepsilon_i g(X_i)$ and $N_+(g) + N_-(g) = \sum_{i=1}^n g(X_i)$. We also have

$$E(N_-^{1/2}(g))^2 = EN_-(g) \le n^{1/2}M_n \tag{5.4}$$

and

$$\begin{aligned}&|(N_+^{1/2}(g) - N_-^{1/2}(g)) - (N_+^{1/2}(h) - N_-^{1/2}(h))|\\ &\qquad \le 2[\textstyle\sum_{i=1}^n (g^{1/2}(X_i) - h^{1/2}(X_i))^2]^{1/2} = 2n^{1/2}d_{n,2}(g^{1/2}, h^{1/2}).\end{aligned} \tag{5.5}$$

Then, using these properties of N_+ and N_- ((5.4) and (5.5)) together with Lemma 2.7 and Fubini's theorem, we obtain, with $\alpha = 2^{-1/2}t^{1/2} - 2^{1/2}M_n^{1/2}$,

$$\begin{aligned}&\Pr\{\|\textstyle\sum_{i=1}^n g(X_i)\|_{\mathscr{G}} > tn^{1/2}\}\\ &\quad \le 2 \Pr\{\|N_+^{1/2}(g)\|_{\mathscr{G}} \ge 2^{-1/2}t^{1/2}n^{1/4}\}\\ &\quad \le 4 E_\varepsilon P_X\{\|N_+^{1/2}(g) - N_-^{1/2}(g)\|_{\mathscr{G}} > \alpha n^{1/4}\}\\ &\quad = 4 E_X P_\varepsilon\{\|N_+^{1/2}(g) - N_-^{1/2}(g)\|_{\mathscr{G}} > \alpha n^{1/4}\}\\ &\quad \le 4 \Pr{}^*\{N_{n,2}(\rho/n^{1/4}, \mathscr{G}^{1/2}) > m\}\\ &\qquad + 4E_X^* P_\varepsilon\{\|N_+^{1/2}(g) - N_-^{1/2}(g)\|_{\mathscr{G}} > \alpha n^{1/4}, N_{n,2}(\rho/n^{1/4}, \mathscr{G}^{1/2}) \le m\}.\end{aligned} \tag{5.6}$$

Let $\mathscr{D}$ be a $d_{n,2}(\rho/n^{1/4})$-dense net of $\mathscr{G}^{1/2}$ of cardinality $N_{n,2}(\rho/n^{1/4}, \mathscr{G}^{1/2})$. Then the P_ε probability of the last event is bounded from above by

$$\begin{aligned}&m \sup_{g\in\mathscr{D}} P_\varepsilon\{|N_+^{1/2}(g) - N_-^{1/2}(g)| > \lambda n^{1/4}\}\\ &\qquad = m \sup_{g\in\mathscr{D}} P_\varepsilon\left\{\frac{|N_+(g) - N_-(g)|}{N_+^{1/2}(g) + N_-^{1/2}(g)} > \lambda n^{1/4}\right\}\\ &\qquad \le m \sup_{g\in\mathscr{D}} P_\varepsilon\left\{\frac{|\sum_{i=1}^n \varepsilon_i g(X_i)|}{(\sum_{i=1}^n g(X_i))^{1/2}} > \lambda n^{1/4}\right\}\\ &\qquad \le m \sup_{g\in\mathscr{D}} P_\varepsilon\left\{\frac{|\sum_{i=1}^n \varepsilon_i g(X_i)|}{(\sum_{i=1}^n g^2(X_i))^{1/2}} > \frac{\lambda n^{1/4}}{r^{1/2}}\right\} \le m \exp\left(\frac{-\lambda^2 n^{1/2}}{2r}\right).\end{aligned}$$

This and (5.6) give (5.3). □

PROOF OF THEOREM 5.1. We will assume without loss of generality that $\|f\|_\infty \le 1/2$ for all $f \in \mathscr{F}$, and as usual we write $\mathscr{F}$ for ${}_0\mathscr{F}$. Let us fix $\varepsilon \in (0, \varepsilon_0)$. By Theorem 3.2, it is enough to show

$$\lim_{n\to\infty}\Pr\left\{\left\|\frac{\sum_{i=1}^n \varepsilon_i f(X_i)}{n^{1/2}}\right\|_{\mathscr{F}'_{\varepsilon,n}} > \tau\varepsilon\right\} = 0 \tag{5.7}$$

for some $\tau > 0$. Let $\mathscr{D}$ denote the set of centers of a minimal covering of $\mathscr{F}'_{\varepsilon,n}$ by $d_{n,1}$-balls of radius at most $\gamma\varepsilon/n^{1/2}$. Then

$$\#\mathscr{D} = N_{n,1}(\gamma\varepsilon/n^{1/2}, \mathscr{F}'_{\varepsilon,n}). \tag{5.8}$$

Let $X'_i = X_{n+i}$, $i = 1, \cdots, n$. Let

$$u = (\tau - \gamma)\varepsilon n^{1/2}. \tag{5.9}$$

Then

$$\begin{aligned}
&\Pr\{\|\textstyle\sum_{i=1}^n \varepsilon_i f(X_i)\|_{\mathscr{F}'_{\varepsilon,n}} > \tau\varepsilon n^{1/2}\}\\
&= E_X P_\varepsilon\{\|\textstyle\sum_{i=1}^n \varepsilon_i f(X_i)\|_{\mathscr{F}'_{\varepsilon,n}} > \tau\varepsilon n^{1/2}\}\\
&\le \Pr{}^*\left\{N_{n,1}\left(\frac{\gamma\varepsilon}{n^{1/2}}, \mathscr{F}'_{\varepsilon,n}\right) > m\right\} + \Pr\left\{\|\textstyle\sum_{i=1}^n f^2(X_i)\|_{\mathscr{F}'_{\varepsilon,n}} > \frac{u}{2}\right\}\\
&\quad + E_X^* P_\varepsilon\Big\{\|\textstyle\sum_{i=1}^n \varepsilon_i f(X_i)\|_{\mathscr{F}'_{\varepsilon,n}} > \tau\varepsilon n^{1/2},\\
&\qquad N_{n,1}\left(\frac{\gamma\varepsilon}{n^{1/2}}, \mathscr{F}'_{\varepsilon,n}\right) \le m,\ \|\textstyle\sum_{i=1}^n f^2(X_i)\|_{\mathscr{F}'_{\varepsilon,n}} \le \frac{u}{2}\Big\}.
\end{aligned} \tag{5.10}$$

The P_ε-probability of this last event is bounded from above by (see (5.8))

$$m \sup_{f\in\mathscr{D}} P_\varepsilon\{|\textstyle\sum_{i=1}^n \varepsilon_i f(X_i)| > u, \sum_{i=1}^n f^2(X_i) \le u/2\} \le 2me^{-u}$$

where in the last inequality we use the standard subgaussian estimate (2.17). Summarizing from (5.10) we get:

$$\begin{aligned}
&\Pr\{\|\textstyle\sum_{i=1}^n \varepsilon_i f(X_i)\|_{\mathscr{F}'_{\varepsilon,n}} > \tau\varepsilon n^{1/2}\}\\
&\le \Pr{}^*\left\{N_{n,1}\left(\frac{\gamma\varepsilon}{n^{1/2}}, \mathscr{F}'_{\varepsilon,n}\right) > m\right\} + \Pr\left\{\|\textstyle\sum_{i=1}^n f^2(X_i)\|_{\mathscr{F}'_{\varepsilon,n}} > \frac{u}{2}\right\} + 2me^{-u}.
\end{aligned} \tag{5.11}$$

We now use Lemma 5.2 to estimate $\Pr\{\|\sum_{i=1}^n f^2(X_i)\|_{\mathscr{F}'_{\varepsilon,n}} > u/2\}$. Take $\mathscr{F}'_{\varepsilon,n}$ as $\mathscr{G}^{1/2}$ in Lemma 5.2 and observe that for $f, g \in \mathscr{F}'_{\varepsilon,n}$, $d_{n,2}(f, g) \le 2^{1/2} d_{n,1}^{1/2}(f, g)$ so that if $d_{n,1}(f, g) \le \gamma\varepsilon/n^{1/2}$ then $d_{n,2}(f, g) \le 2^{1/2}\gamma^{1/2}\varepsilon^{1/2}/n^{1/4}$. So the quantities appearing in Lemma 5.2 become $r = 1$, $M_n = \varepsilon$, $\rho = 2^{1/2}\gamma^{1/2}\varepsilon^{1/2}$ and $2\lambda = [(\tau - \gamma)^{1/2} - 2^{3/2}(1 + 2\gamma^{1/2})]\varepsilon^{1/2}$, and we obtain

$$\begin{aligned}
&\Pr\left\{\|\textstyle\sum_{i=1}^n \varepsilon_i f(X_i)\|_{\mathscr{F}'_{\varepsilon,n}} > \frac{u}{2}\right\}\\
&\le 4\Pr{}^*\left\{N_{n,1}\left(\frac{\gamma\varepsilon}{n^{1/2}}, \mathscr{F}'_{\varepsilon,n}\right) > m\right\} + 8\, m \exp\left(\frac{-\lambda^2 n^{1/2}}{2}\right)
\end{aligned} \tag{5.12}$$

which applied in (5.11) gives

$$\Pr\{\|\textstyle\sum_{i=1}^n \varepsilon_i f(X_i)\|_{\mathscr{F}'_{\varepsilon,n}} > \tau\varepsilon n^{1/2}\}$$

$$(5.13)\qquad \le 5\ \Pr^*\left\{N_{n,1}\left(\frac{\gamma\varepsilon}{n^{1/2}},\ \mathscr{F}'_{\varepsilon,n}\right) > m\right\} + 2me^{-u} + 8m\exp\left(\frac{-\lambda^2 n^{1/2}}{32}\right).$$

Choose now $m = \exp(\sigma\varepsilon n^{1/2})$ and τ large enough (depending only on σ and γ) so that $\sigma\varepsilon - un^{-1/2}$ and $\sigma\varepsilon - \lambda^2/2$ are both negative. Then the limit in (5.13) is zero by (5.1), and this proves (5.7). □

5.3. REMARK. Since the ε-covering number of $\mathscr{F}'$ is bounded by the square of the $\varepsilon/2$-covering number of $\mathscr{F}$, each of the following conditions implies (5.1):

$$(5.14)\qquad \lim_{\delta\downarrow 0}\limsup_n \Pr^*\left\{\frac{\ln N_{n,1}(\gamma\varepsilon/n^{1/2},\ {}_0\mathscr{F}'_\delta)}{n^{1/2}} > \sigma\varepsilon\right\} = 0;$$

$$(5.15)\qquad \limsup_n \Pr^*\left\{\frac{\ln N_{n,1}(\gamma\varepsilon/n^{1/2},\ {}_0\mathscr{F})}{n^{1/2}} > \sigma\varepsilon\right\} = 0.$$

Also, ${}_0\mathscr{F}$ can be replaced by $\mathscr{F}$ in these expressions.

From the point of view of the CLT it seems more natural to impose conditions on $d_{n,2}$ than on $d_{n,1}$. The estimate (5.12) can clearly be given in terms of $N_{n,2}$, and we can use Theorem 2.15 in the first part of the proof. This yields the following result:

5.4. THEOREM. *Let $\mathscr{F}$ be a uniformly bounded* NDM(P) *class of functions. Assume*:

(i) *$\mathscr{F}$ is P-pregaussian,*
(ii) *for all $\varepsilon > 0$,*

$$(5.16)\qquad \lim_n E^*\left[1 \wedge \int_0^{n^{-1/4}} (\ln N_{n,2}(\lambda,\ {}_0\mathscr{F}'_{\varepsilon,n}))^{1/2}\, d\lambda\right] = 0.$$

Then $\mathscr{F}$ is a functional P-Donsker class.

PROOF. As in 5.1 we assume $\|f\|_\infty \le \frac{1}{2}$ for all $f \in \mathscr{F}$, and write $\mathscr{F}$ for ${}_0\mathscr{F}$. So,

$$\Pr\{\|\textstyle\sum_{i=1}^n \varepsilon_i f(X_i)\|_{\mathscr{F}'_{\varepsilon,n}} > \tau\varepsilon n^{1/2}\}$$

$$\le \Pr\{\|\textstyle\sum_{i=1}^n f^2(X_i)\|_{\mathscr{F}'_{\varepsilon,n}} > 4^{-1}n^{1/2}\}$$

$$(5.17)\qquad + E_X P_\varepsilon\{\|\textstyle\sum_{i=1}^n \varepsilon_i f(X_i)\|_{\mathscr{F}'_{\varepsilon,n}} > \tau\varepsilon n^{1/2},\ \|\sum_{i=1}^n f^2(X_i)\|_{\mathscr{F}'_{\varepsilon,n}} \le 4^{-1}n^{1/2}\}$$

$$= (\mathrm{I}) + (\mathrm{II}).$$

To estimate (I) we apply Lemma 5.2 as in the previous theorem, to get (for $\varepsilon < 1/64$):

(5.18) $\quad (\mathrm{I}) \le 4 \Pr^*\{N_{n,2}(2^{-4}n^{-1/4}, \mathscr{F}'_{\varepsilon,n}) > m\} + 8m \exp(-2^{-6}n^{1/2}).$

Then, by taking $m = \exp(2^{-7}n^{1/2})$ we see by (5.16) that $\limsup_n(\mathrm{I}) = 0$.

To estimate (II) we use Theorem 2.15 (2.24) with $t_0 = 0$, $T = \mathscr{F}'_{\varepsilon,n}$ and $Z = \sum_{i=1}^n \varepsilon_i \delta_{X_i}/n^{1/2}$. So,

$$(\mathrm{II}) \le E^*\left[1 \wedge \frac{c}{\tau\varepsilon}\left(\int_0^{n^{-1/4}} (\ln N_{n,2}(\lambda, \mathscr{F}'_{\varepsilon,n}))^{1/2}\, d\lambda + \Phi(n^{-1/4})\right)\right] \to 0$$

as $n \to \infty$. This and (5.18) give that the probability in the left side of (5.17) tends to zero as $n \to \infty$. Now the result follows from Theorem 3.2. □

5.5. REMARK. As in Remark 5.4, it is obvious that we can replace ${}_0\mathscr{F}'_{\varepsilon,n}$ by $\mathscr{F}$ (at the expense of some weakening of the resulting statement).

We may ask whether certain random entropy conditions already imply that $\mathscr{F}$ is pregaussian or even that $\mathscr{F}$ satisfies the L_2-entropy condition ((3.1) with $\rho = e_P$). Lemma 2.24 shows that a condition stronger than (5.1) but of the same type does indeed imply (3.1). Combining Theorem 5.1 with Lemma 2.24 (taking into account that $N_{n,1}(\delta, \mathscr{F}'_{\varepsilon,n}) \le N^2_{n,1}(\delta, \mathscr{F}) \le N^2_{n,2}(\delta, \mathscr{F})$ and that if F satisfies conditions (2.34) then $n^{1/2}/F(n) \to 0$) we obtain the following result.

5.6. THEOREM. *Let $\mathscr{F} \subset \mathscr{L}_2(P)$ be a uniformly bounded* NDM(P) *class. Assume that for some function F verifying the conditions* (2.34), *the sequence*

(5.19) $$\left\{\frac{\ln N_{n,\infty}(\delta/n^{1/2}, \mathscr{F})}{F(n)}\right\}_{n=1}^{\infty}$$

is $\Pr^*$*-stochastically bounded for every $\delta > 0$. Then $\mathscr{F}$ is a funtional P-Donsker class.*

We do not know if $F(n)$ can be replaced by $n^{1/2}$ in Theorem 5.6, but as we show in Section 6, this theorem is sharp (and so are 5.1 and 5.4). Now we pass to classes of sets; for examples of application of the previous results see Sections 6 and 7.

(b) *Families of sets.* Let us recall from (4.8) that the indicator functions of sets in $\mathscr{C}'_{\varepsilon,n}$ have the form

$$\{|I_A - I_B|: A, B \in \mathscr{C}, P(|I_A - I_B|^2) \le \varepsilon/n^{1/2}\},$$

which are the absolute values of the functions in $\mathscr{F}'_{\varepsilon,n}$ if $\mathscr{F} = \{I_C: C \in \mathscr{C}\}$. Hence random entropies of $\mathscr{C}'_{\varepsilon,n}$ are smaller than those of $\mathscr{F}'_{\varepsilon,n}$. On the other hand, by a proof which is similar but simpler than that of Theorem 5.1, using

Theorem 3.2 (via Remark 4.5 (b)) we obtain:

5.7. THEOREM. *Let* $\mathscr{C}$ *be an* NDM(P) *class of measurable subsets of S. Then* $\mathscr{C}$ *is a functional P-Donsker class if and only if both:*

(i) $\mathscr{C}$ *is P-pregaussian, and*

(ii) *for some* γ, σ, $\varepsilon_0 > 0$ *and all* $0 < \varepsilon < \varepsilon_0$,

$$(5.1)' \qquad \limsup_n \Pr{}^*\left\{\frac{\ln N_{n,1}(\gamma\varepsilon/n^{1/2}, {}_0\mathscr{C}'_{\varepsilon,n})}{n^{1/2}} > \sigma\varepsilon\right\} = 0.$$

PROOF. As usual, we write $\mathscr{C}$ for ${}_0\mathscr{C}$. In contrast to the proof of Theorem 5.1, we apply Lemma 5.2 directly to prove (4.11) in Corollary 4.4 (see Remark 4.5 (b)). This proves that (i) and (ii) imply that $\mathscr{C}$ is functional P-Donsker. Now assume that $\mathscr{F} = \{I_C: C \in \mathscr{C}\}$ is a functional P-Donsker class. Since (i) is necessary by definition, we only have to show that (ii) holds. By Theorem 2.14 (2) it follows that if $\mathscr{F}$ is a functional P-Donsker class, then

$$\lim_n E\left\|\frac{\sum_{i=1}^n g_i f(X_i)}{n^{1/2}}\right\|_{\mathscr{F}'_{\varepsilon,n}} = 0.$$

Then, Theorem 2.16 applied conditionally to the process $\sum_{i=1}^n g_i f(X_i)/n^{1/2}$, $f \in \mathscr{F}'_{\varepsilon,n}$, gives:

$$(5.20) \qquad \lim_n E \sup_{\lambda>0}\lambda(\ln N_{n,2}(\lambda, \mathscr{F}'_{\varepsilon,n}))^{1/2} = 0.$$

But any $f \in \mathscr{F}'_{\varepsilon,n}$ takes on only the values 1, 0, -1, therefore for $f, g \in \mathscr{F}'_{\varepsilon,n}$,

$$(5.21) \qquad \tfrac{1}{2}(d_{n,2}(f, g))^2 \le d_{n,1}(f, g) \le (d_{n,2}(f, g))^2.$$

This, together with (5.20) implies that

$$(5.22) \qquad \limsup_n E \sup_{\lambda>0}(\lambda \ln N_{n,1}(\lambda, \mathscr{F}'_{\varepsilon,n}))^{1/2} = 0.$$

This implies condition (5.1) as $N_{n,1}(\lambda, \mathscr{F}'_{\varepsilon,n}) \ge N_{n,1}(\lambda, \mathscr{C}'_{\varepsilon,n})$. □

As we see from the previous proof, there is a variety of necessary and sufficient conditions for the CLT involving random entropies; the weakest (hence the best sufficient condition) being (i) + (5.1)′, and the strongest (or the best necessary), (i) + (5.20). Let us also record this last one (which by (5.21) can be given in terms of $N_{n,p}$ for any $0 < p < \infty$ but which we give in terms of $N_{n,2}$ for the purpose of comparison with Theorem 5.4).

5.7′. THEOREM. Let $\mathscr{F}$ be an NDM(P) *class of indicator functions. Then* $\mathscr{F}$ *is a funtional P-Donsker class if and only if both,*

(i) $\mathscr{F}$ *is P-pregaussian and*

(ii) $\lim_n E \sup_{\lambda>0}(\lambda^2 \ln N_{n,2}(\lambda, {}_0\mathscr{F}'_{\varepsilon,n}))^{1/2} = 0.$

In view of Theorem 5.7, it is natural to ask whether the $N_{n,1}$ entropy condition

(5.1) is also necessary for uniformly bounded classes of functions $\mathscr{F}$ to be functional P-Donsker. We now give an example to the contrary, even satisfying the usual L_2-entropy condition.

5.8. EXAMPLE. Let $S = c_0$, the space of sequences of real numbers converging to zero. Let $\mathscr{F} = \{f_k\}_{k=1}^{\infty}$ be the family of coordinate functionals of c_0, enlarged with the zero functional, i.e. for any $\{\lambda_r\} \in c_0$

$$f_k(\{\lambda_r\}) = \lambda_k, \quad f_\infty(\{\lambda_r\}) = 0, \quad k \in \mathbb{N}.$$

Define a c_0-valued random variable as

$$Y(\omega) = \{a_k \varepsilon_k(\omega)\}_{k=1}^{\infty}$$

where $\{\varepsilon_k\}$ is a Rademacher sequence, and

$$a_k = (\ln k)^{-1}, \quad k > 1; \quad a_1 = 1.$$

Let $P = \mathscr{L}(Y)$ on $(c_0, \mathscr{B}(c_0))$. We show that $\mathscr{F}$ is a funtional P-Donsker class by applying Theorem 5.4 but that nevertheless $\mathscr{F}$ does not satisfy condition (5.1). Let $\{X_r\}_{r=1}^{\infty}$ be independent copies of Y which we may assume to be the coordinate functions on $(c_0^{\mathbb{N}}, (\mathscr{B}(c_0))^{\mathbb{N}}, P^{\mathbb{N}})$ to comply with the measurability framework of Section 2, but since $\mathscr{F}$ is a countable class this is not needed. Then the random variables $\{f_k(X_r)\}_{k,r}$ are independent with law $\mathscr{L}(a_k\varepsilon)$, ε a Rademacher variable. For convenience we write

$$f_k(X_r) = a_k \varepsilon_k^r.$$

Note then that

$$E(f_k(X) - f_\ell(X))^2 = a_k^2 + a_\ell^2 \le 2(a_k \vee a_\ell)^2$$

and that

$$d_{n,2}^2(f_k, f_\ell) = \textstyle\sum_{r=1}^n |f_r(X_k) - f_\ell(X_k)|^2/n$$
$$= \textstyle\sum_{r=1}^n |a_k\varepsilon_r^k - a_\ell\varepsilon_r^\ell|^2/n \le 2(a_k \vee a_\ell)^2.$$

For $\varepsilon < 1$, $a_k < \varepsilon$ if $k > k_\varepsilon := e^{1/\varepsilon}$. So, the $L_2(P)$- and $d_{n,2}$-balls of radius ε and centers $f_1, \cdots, f_{k_\varepsilon}$ cover $\mathscr{F}$. Hence

$$\ln N(\varepsilon, \mathscr{F}, P) \le \varepsilon^{-1}, \quad \ln N_{n,2}(\varepsilon, \mathscr{F}) \le \varepsilon^{-1}$$

showing that the L_2-entropy hypothesis (3.1) (with $\rho = e_P$) and hypothesis (5.16) of Theorem 5.4 are satisfied. So, $\mathscr{F}$ is P-Donsker. We now estimate $d_{n,1}$. For $\varepsilon < n^{1/2}$, it follows from the previous computations that

$$\mathscr{F}'_{\varepsilon,n} \supseteq \{f_i - f_j : i, j \ge \exp[n^{1/4}/\varepsilon^{1/2}]\}.$$

Since $0 \in \mathscr{F}$, this class contains the family of functions

$$\mathscr{F}(\varepsilon, n) = \{f_i : i \ge \exp[n^{1/4}/\varepsilon^{1/2}]\}.$$

So, in order to show that the random entropy condition (5.1) does not hold, it is enough to prove that for any γ, $\sigma > 0$ there exists ε_0 such that for all $\varepsilon \in (0, \varepsilon_0]$,

$$\lim_{n\to\infty}\Pr\{\inf[d_{n,1}(f_i, f_j): i \neq j, \exp(n^{1/4}/\varepsilon^{1/2}) \le i, j \le 2\exp(\sigma n^{1/2})] > \gamma\varepsilon/n^{1/2}\} = 1.$$

We take $\varepsilon_0 < (4\gamma\sigma)^{-1}$. Since for $i, j \le \exp(\sigma n^{1/2})$,

$$d_{n,1}(f_i, f_j) \ge \sum_{k=1}^n I_{[\varepsilon_i^k \neq \varepsilon_j^k]}/2\sigma n^{3/2},$$

we need only show that

$$\lim_{n\to\infty}\Pr\{\inf_{i\neq j, i,j\le 2\exp(\sigma n^{1/2})} \sum_{k=1}^n I_{[\varepsilon_i^k\neq\varepsilon_j^k]} > 2\gamma\sigma\varepsilon/n^{1/2}\} = 1,$$

or that

$$\lim_{n\to\infty}\Pr\{\sum_{k=1}^n I_{[\varepsilon_i^k\neq\varepsilon_j^k]} \le 2\gamma\sigma\varepsilon/n^{1/2} \text{ for some } i \neq j \text{ with } i, j \le 2\exp(\sigma n^{1/2})\} = 0.$$

This probability is clearly bounded from above by

$$\begin{aligned} 4\exp(2\sigma n^{1/2})&\Pr\{\sum_{k=1}^n I_{[\varepsilon_i^k=\varepsilon_j^k]} > (1 - 2\gamma\sigma\varepsilon)n\} \\ &= 4\exp(2\sigma n^{1/2})\Pr\{\sum_{k=1}^n (I_{[\varepsilon_i^k=\varepsilon_j^k]} - 1/2) > (1/2 - 2\gamma\sigma\varepsilon)n\} \\ &\le 4\exp(2\sigma n^{1/2} - 6n/45) \to 0, \end{aligned}$$

by Bernstein's inequality (2.18). □

We do not know whether the conditions in Theorem 5.7′ are sufficient in the case of functions.

6. Estimation of random entropies for families of sets with applications. In the case of sets random entropies can be related to the combinatorial quantities $\Delta^{\mathscr{S}}(X_1, \cdots, X_n)$ introduced by Vapnik and Červonenkis (1971) for the Glivenko-Cantelli theorem (see (2.23) for the definition). In fact for any family of sets $\mathscr{S}$, $n \in \mathbb{N}$, $0 < p \le \infty$, $\varepsilon > 0$,

$$\text{(6.1)} \qquad \begin{cases} N_{n,p}(\varepsilon, \mathscr{S}) \le \Delta^{\mathscr{S}}(X_1, \cdots, X_n), & \varepsilon \ge 0, \\ N_{n,p}(\varepsilon, \mathscr{S}) = \Delta^{\mathscr{S}}(X_1, \cdots, X_n), & 0 \le \varepsilon < n^{-1/(p\vee 1)} \end{cases}$$

(with $1 = n^{-1/\infty}$) as if two sets A and B determine the same subset of the sample $X_1, \cdots, X_n$, then $d_{n,p}(I_A, I_B) = 0$ and otherwise, $d_{n,p}(I_A, I_B) \ge n^{-1/(p\vee 1)}$.

6.1. PROPOSITION. *Let $\mathscr{S}$ be a* NDM(P) *class of sets such that*

(i) *$\mathscr{S}$ is P-pregaussian, and*

(ii) *there exist $\sigma > 0$ such that for all $\varepsilon > 0$*

$$\text{(6.2)} \qquad \lim_{n\to\infty}\Pr^*\left\{\frac{\ln \Delta^{\mathscr{S}'_{\varepsilon,n}}(X_1, \cdots, X_n)}{n^{1/2}} > \sigma\varepsilon\right\} = 0.$$

Then $\mathscr{S}$ is a functional P-Donsker class.

6.2. COROLLARY. *Let $\mathscr{S}$ be a* NDM(P) *class of sets such that*
(i) *$\mathscr{S}$ is P-pregaussian and*
(ii)

$$\lim_{n\to\infty} \frac{\ln \Delta^{\mathscr{S}}(X_1, \cdots, X_n)}{n^{1/2}} = 0 \quad \textit{in probability.} \tag{6.3}$$

Then $\mathscr{S}$ is a functional P-Donsker class.

Proposition 6.1 is related to Theorem 1 in Le Cam (1983b).

6.3. PROPOSITION. *Let $\mathscr{S}$ be* NDM(P) *and let F be a function verifying the conditions* (2.34). *Assume that the sequence*

$$\left\{\frac{\ln \Delta^{o\mathscr{S}}(X_1, \cdots, X_n)}{F(n)}\right\}_{n=1}^{\infty} \tag{6.4}$$

is Pr*-*stochastically bounded. Then $\mathscr{S}$ is a funtional P-Donsker class.*

Typical functions $F(n)$ are $F(n) = n^{1/2}/(\ln n)(\ln_2 n) \cdots (\ln_{k-1} n)(\ln_k n)^{1+\delta}$ for some $k \geq 1$ and $\delta > 0$, where $\ln_r n = \ln(\cdots (\ln n) \cdots)$.

6.4. REMARK. We may ask how close the conditions on $\Delta^{\mathscr{S}}$ in the previous two theorems are from being necessary for the CLT. In this respect we note the following: if the "bounded" CLT holds for $\mathscr{S}$ (i.e. if the sequence of laws $\{\mathscr{L}(\| \sum_{i=1}^n (\delta_{X_i} - P)/n^{1/2} \|_{\mathscr{S}})\}_{n=1}^\infty$ is tight) then the sequence

$$\left\{\frac{\ln \Delta^{\mathscr{S}}(X_1, \cdots, X_n)}{n^{1/2}\ln n}\right\}_{n=1}^{\infty} \tag{6.5}$$

is stochastically bounded (we assume $\mathscr{S}$ countable). To prove this, note that if this sequence is not stochastically bounded, then there exist $\tau > 0$, $A_k \uparrow \infty$, $n_k \uparrow \infty$ such that

$$\Pr\{\Delta^{\mathscr{S}}(X_1, \cdots, X_{n_k}) > \exp(A_k n_k^{1/2} \ln n_k)\} > \tau; \tag{6.6}$$

then by Sauer's theorem (Sauer, 1972), with probability larger than τ, for each k, $\mathscr{S}$ completely shatters a subset of $\{X_1, \cdots, X_{n_k}\}$ of cardinality $[A_k n_k^{1/2}/3]$ (i.e., there exists $J_k(\omega) \subset \{1, \cdots, n_k\}$ with $\#J_k(\omega) = [A_k n_k^{1/2}/3]$ such that $\Delta^{\mathscr{S}}\{x_i: i \in J_k\} = 2^{[A_k n_k^{1/2}/3]}$) with probability at least τ. Since $\{\sum_{i=1}^n \varepsilon_i/n^{1/2}\}$ is stochastically bounded and the bounded CLT holds, by Lemma 2.7,

$$\{\| \textstyle\sum_{n=1}^n \varepsilon_i \delta_{X_i}/n^{1/2} \|_{\mathscr{S}}\}_{n=1}^\infty$$

is stochastically bounded, but this is a contradiction with the following (which follows from the previous discussion):

$$\lim_{t\to\infty}\lim_{k\to\infty} 2\ \Pr\{\|\textstyle\sum_{i=1}^{n_k}\varepsilon_i\delta_{X_i}/n_k^{1/2}\|_{\mathscr{C}} > 4t\}$$

$$\geq \lim_{t\to\infty}\lim_{k\to\infty} E_X P_\varepsilon\{\|\textstyle\sum_{i\in J_k}\varepsilon_i\delta_{X_i}/n_k^{1/2}\|_{\mathscr{C}} > 4t\}$$

$$\geq \lim_{t\to\infty}\lim_{k\to\infty} E_X P_\varepsilon\{\max[\#J_k\cap\{i: \varepsilon_i = 1\}, \#J_k\cap\{i: \varepsilon_i = -1\}] > 4tn_k^{1/2}\}$$

$$\geq \tau > 0.$$

This shows that the conditions on $\Delta^{\mathscr{C}}$ in the previous propositions cannot be much weakened, partially answering a question of Le Cam (1983a). It would be interesting if one could obtain more precise conditions on the CLT in terms of K instead of $\ln\Delta$, where K is the maximum of the cardinalities of the subsets of $\{X_1, \cdots, X_n\}$ which are completely shattered by $\mathscr{C}$.

Next we show that the previous propositions are useful and sharp. Certainly, modulo measurability, all of them imply the CLT for VC classes (by Theorems 2.22 and 2.23). Proposition 6.1 implies also a CLT under a metric entropy with inclusion condition (with only a pregaussian assumption). Even the weaker Proposition 6.3 implies the CLT for sequences of sets (Corollary 4.11) with interesting rates on Δ as well as the theorem of Durst and Dudley (1981) (see also Dudley, 1982, Section 6.3) for $\mathscr{C} = 2^S$ and P discrete. For M_0 as defined in (4.7) we have

6.5. Corollary. *Let $\mathscr{C}$ be a* NDM(P) *class of sets such that*

(i) *$\mathscr{C}$ is P-pregaussian, and*

(ii) *$n^{1/2}\ln M_0(\mathscr{C}'_{\varepsilon,n}, \varepsilon/n^{1/2}, P) \to 0$ as $n\to\infty$ for all $\varepsilon > 0$.*

Then $\mathscr{C}$ is a functional P-Donsker class.

Proof. Let $M_0 := M_0(\mathscr{C}'_{\varepsilon,n}, \varepsilon/n^{1/2}, P)$ and let $A_1, \cdots, A_{M_0}$ be sets in $\mathscr{S}$ such that for each $A \in \mathscr{C}'_{\varepsilon,n}$ there is $j_A \leq M_0$ satisfying $A_{j_A} \supset A$ and $PA_{j_A}\backslash A < \varepsilon/2n^{1/2}$. For any set B, let $X\cap B$ denote the intersection $\{X_1, \cdots, X_n\}\cap B$, and $\#X\cap B$ its cardinality. Then for σ to be chosen below and for all $\varepsilon > 0$,

$$\Pr^*\{\Delta^{\mathscr{C}_{\varepsilon,n}}(X_1, \cdots, X_n) \geq \exp(\sigma\varepsilon n^{1/2})\}$$

$$\leq \Pr\{\textstyle\sum_{j=1}^{M_0} 2^{\#X\cap A_j} > \exp(\sigma\varepsilon n^{1/2})\}$$

$$\leq M_0\max_{j\leq M_0}\Pr\{2^{\#X\cap A_j} \geq \exp(\sigma\varepsilon n^{1/2})/M_0\} = (\mathrm{I}).$$

Let m denote the smallest integer larger than or equal to $(\sigma\varepsilon n^{1/2} - \ln M_0)\log_2 e$. Then

$$(\mathrm{I}) = M_0\max_{j\leq M_0}\Pr\{\text{there are at least } m\ X_i\text{'s in } A_j\}$$

$$\leq M_0\left(\frac{en}{m}\right)^m\left(\frac{3\varepsilon}{2n^{1/2}}\right)^m.$$

By the hypothesis on M_0, there exists n_0 such that

$$M_0 \leq \exp(\sigma\varepsilon n^{1/2}/2), \quad n > n_0,$$

and therefore, for $\sigma > 3e^2/(\log_2 e)$,

$$(\mathrm{I}) \le \exp(\sigma\varepsilon n^{1/2}/2)\left(\frac{3e}{(\log_2 e)\sigma}\right)^{(\sigma\varepsilon\log_2 e/2)n^{1/2}} \to 0 \quad \text{as} \quad n \to \infty.$$

Therefore, condition (ii) above implies condition (ii) in Proposition 6.1. □

As in Remark 4.2 (1), we can compare this result with Theorem 5.1 in Dudley (1978). Corollary 6.5 shows that for a NDM(P) P-pregaussian class $\mathscr{C}$, the condition

$$\lim_{\varepsilon\to 0}\varepsilon \ln N_I(\varepsilon, \mathscr{C}, P) = 0$$

implies that $\mathscr{C}$ is a functional P-Donsker class. (See Dudley, loc. cit., for the definition of N_I.)

In view of Proposition 6.3, the following corollary implies the CLT for sequences of sets (Corollary 4.11).

6.6. **Proposition.** *Let $\mathscr{C} = \{A_m\}_{m=1}^{\infty}$ and $p_m = PA_m$. Then, for all $s > 0$ there exists $0 < c_s < \infty$ such that*

$$\sum_{m=1}^{\infty} p_m^s < \infty \Rightarrow \Pr\left\{\frac{\ln \Delta^{\mathscr{C}}(X_1, \cdots, X_n)}{\ln n} > c_s\right\} \to 0 \quad \textit{as} \quad n \to \infty.$$

Proof. Let $t_s = [2s] + 1$, $m_n = t_s\binom{n}{t_s}$ and assume $2n > t_s$. We first observe that if for every A_m, $m > m_n$, fewer than t_s X_i's (out of $X_1, \cdots, X_n$) are in A_m, then

$$\Delta^{\mathscr{C}}(X_1, \cdots, X_n) \le m_n + \sum_{k=0}^{t_s-1}\binom{n}{k} \le 2t_s\binom{n}{t_s}.$$

Therefore, since $p_m \le m^{-1/s}$ for $m > m_n$ and n large enough, we have

$$\Pr\left\{\Delta^{\mathscr{C}}(X_1, \cdots, X_n) > 2t_s\binom{n}{t_s}\right\}$$
$$\le \Pr\{\#[\{X_1, \cdots, X_n\} \cap A_m] \ge t_s \text{ for some } m > m_n\}$$
$$\le \binom{n}{t_s}\Pr\{\sup_{m>m_n}\delta_{X_1}(A_n) \cdots \delta_{X_{t_s}}(A_m) = 1\} \le \binom{n}{t_s}\sum_{m>m_n} p_m^{t_s} \to 0$$
$$\text{as} \quad n \to \infty. \quad \square$$

This proposition should be compared to Theorem 9.3.2 in Dudley (1982) which yields that no infinite sequence of nontrivial independent sets is a Vapnik–Červonenkis collection.

Durst and Dudley (1981) and Borisov (1981) give two characterizations of those discrete probability measures for which $\mathscr{C} = 2^{\mathbb{N}}$ is a functional P-Donsker class, in particular showing that the metric entropy with inclusion result is sharp. Next we give yet another characterization of these P which shows that Theorem 5.5 (the weakest theorem in Section 5), is sharp.

6.7. COROLLARY. *Let $S = \mathbb{N}$, $\mathscr{C} = 2^{\mathbb{N}}$, and let P be a p.m. on S. Then $\mathscr{C}$ is a functional P-Donsker class if and only if there exists a function F satisfying the conditions* (2.34) *such that the sequence*

$$\left\{\frac{\ln \Delta^{\mathscr{C}}(X_1, \cdots, X_n)}{F(n)}\right\}_{n=1}^{\infty} \tag{6.4}$$

is stochastically bounded.

PROOF. Clearly $\mathscr{C}$ is NDM(P). Let $P\{i\} = p_i$. By Proposition 6.3, condition (6.4) is sufficient for $\mathscr{C}$ to be a functional P-Donsker class. If $\mathscr{C}$ is P-pregaussian, then $\sum_{i=1}^{\infty} p_i^{1/2} < \infty$ (Durst and Dudley, 1981). So, to prove the corollary it is enough to show that

$$\sum_i p_i^{1/2} < \infty \Rightarrow \left\{\frac{\ln \Delta^{\mathscr{C}}(X_1, \cdots, X_n)}{F(n)}\right\}_{n=1}^{\infty} \tag{6.7}$$

is stochastically bounded for some F satisfying (2.34).

Let $\sum p_i^{1/2} < \infty$, $p_i \downarrow$. By a theorem of Boas (1960), if $\mathscr{L}(X) = P$, then

$$\int^{\infty} \left(\frac{\Pr\{X > u\}}{u}\right)^{1/2} du < \infty.$$

Let Z be a nonnegative real random variable with a differentiable distribution function such that

$$\int^{\infty} \left(\frac{P\{Z > u\}}{u}\right)^{1/2} du < \infty \quad \text{and} \quad P\{Z > u\} \geq P\{X > u\} \quad \text{for all} \quad u > 0.$$

Set $G(u) = P\{Z > u\}/u$. We claim that the function

$$F(u) = G^{-1}\left(\frac{1}{4eu}\right)$$

satisfies the required conditions. F satisfies the conditions (2.34):

$$F(u)\uparrow \quad (\text{as } G\downarrow), \quad \frac{F(u)}{u} = 4e \Pr\left\{Z > G^{-1}\left(\frac{1}{4eu}\right)\right\}\downarrow, \quad \text{and}$$

$$\int^{\infty} \frac{F(u)}{u^{3/2}}\, du = 2\sqrt{e} \int_0 G^{-1}(y^2)\, dy < \infty$$

by Corollary 1 in Boas and Marcus (1973). The sequence (6.4) is bounded in probability:

$$\Pr\left\{\frac{\ln \Delta^{\mathscr{S}}(X_1, \cdots, X_n)}{F(n)} > 2 \ln 2\right\}$$

$$= \Pr\{\Delta^{\mathscr{S}}(X_1, \cdots, X_n) > 2^{2F(n)}\} \geq \Pr\{\#\{X_1, \cdots, X_n\} \geq [2F(n)] + 1\}$$

$$\leq \Pr\{\text{at least } [F(n)]\ x_i\text{'s are larger than } F(n)\}$$

$$\leq \left(\frac{en \Pr\{X > F(n)\}}{[F(n)]}\right)^{[F(n)]} \leq \left(\frac{2en \Pr\{Z > F(n)\}}{F(n)}\right)^{[F(n)]}$$

$$= (1/2)^{[F(n)]} \to 0 \quad \text{as} \quad n \to \infty. \quad \square$$

(Note that by taking $F(u) = G^{-1}(1/4eu) \vee u^{1/3}$, one can even ensure that the sequence (6.4) is a.s. bounded if $\sum p_i^{1/2} < \infty$.)

The results 6.5–6.7 show that it is possible in many instances to estimate efficiently $\Delta^{\mathscr{S}}(X_1, \cdots, X_n)$, and that Propositions 6.1–6.3 are sharp. It is difficult to compare these results with Theorem 4.8: both types of results cover the best known cases and we do not know of any example for which one type of theorem applies and not the other (aside from measurability). Theorem 4.8 has the advantage of not requiring measurability whereas the results in this section have the advantage of generalizing to classes of functions, and of course providing essentially necessary and sufficient conditions in the set case. Note also that in the applications we have not used the full strength of Theorem 5.1 as we considered $\Delta = N_{n,1}(1/2n)$ instead of the smaller $N_{n,1}(n^{-1/2})$.

7. Applications to the CLT for uniformly bounded processes. Let $\{X(t): t \in T\}$, T compact metric, be a stochastic process with trajectories in $C(T)$ (or if $T = [0, 1]^d$, in $D[0, 1]^d$: see e.g. Billingsley (1968) for definitions), and let P_X be the law of X. Let $\mathscr{F} = \{\delta_t: t \in T\}$. If $\mathscr{F}$ is a functional P_X-Donsker class and if $\{X_i\}_{i=1}^{\infty}$ are i.i.d. with law P_X, then it is obvious that

$$\mathscr{L}\left(\textstyle\sum_{i=1}^{n} (X_i - EX_i)/n^{1/2}\right) \to \mathscr{L}(Z) \tag{7.1}$$

where Z is a Gaussian process with trajectories in $C(T)$ (or $D[0, 1]^d$) as given in Definition 2.11, and conversely. If this limit holds we write $X \in$ CLT.

Theorems 5.1 and 5.4 and some observations of Kolčinskiĭ are relevant to the CLT in $C(T)$. Below we also obtain results of this type as well as some results in $D[0, 1]$.

In the theorems below the hypotheses easily imply that the class is NDM(P).

7.1. THEOREM. *Let X be a centered, uniformly bounded, pregaussian process on T. Suppose that there exists a nonnegative real random variable L and a continuous pseudo-distance ρ on T such that for some $r \in (0, 1]$,*

$$EL^r < \infty, \quad \lim_{\varepsilon \to 0} \varepsilon \ln N(\varepsilon, T, \rho^r) = 0 \tag{7.2}$$

and

$$(7.3)\qquad |X(t,\omega) - X(s,\omega)| \le L(\omega)\rho(s,t), \quad s, t \in T, \quad \omega \in \Omega.$$

Then $X \in CLT$ in $C(T)$.

PROOF. By Theorem 5.1, it is enough to check condition (5.15) for $\mathscr{F} = \{\delta_t: t \in T\}$ and $\{X_i\}_{i=1}^{\infty}$ i.i.d. with law P. We assume $\|X\|_\infty \le 1/2$. Now,

$$d_{n,1}(\delta_s, \delta_t) = \sum_{i=1}^n |X_i(t) - X_i(s)|/n \le \left(\sum_{i=1}^n L_i^r/n\right)\rho^r,$$

where $\{L_i\}$ are independent copies of L. So, it follows that

$$\Pr\left\{\frac{\ln N_{n,1}(\gamma\varepsilon/n^{1/2}, \mathscr{F})}{n^{1/2}} > \varepsilon\right\}$$

$$\le \Pr\left\{\frac{\ln N(\gamma\varepsilon/n^{1/2}, T, (\sum_{i=1}^n L_i^r/n)\rho^r)}{n^{1/2}} > \varepsilon\right\}$$

$$\le \Pr\left\{\frac{\sum_{i=1}^n L_i^r}{n} > 2EL^r\right\} + \Pr\left\{\frac{\ln N(\gamma\varepsilon/2(EL^r)n^{1/2}, T, \rho^r)}{n^{1/2}} > \varepsilon\right\}$$

$$\to 0 \quad \text{as} \quad n \to \infty$$

(the first term by the law of large numbers, and the second by hypothesis). This proves (5.15) and the result follows from Theorem 5.1. □

7.2. REMARK. Now we compare this result with the Jain–Marcus theorem (1975). For this purpose assume, as usual, $\|X\|_\infty \le 1/2$ and

$$|X(t,\omega) = X(s,\omega)| \le L(\omega)\rho(t,s).$$

Then, for $0 < r < 2$

$$|X(t,\omega) - X(s,\omega)| \le L^{r/2}(\omega)\rho^{r/2}(t,s).$$

Now assume $EL^r < \infty$ and that $\rho^{r/2}$ satisfies the metric entropy condition

$$(7.4)\qquad \int_0 (\ln N(\varepsilon, T, \rho^{r/2}))^{1/2}\, d\varepsilon > \infty.$$

Then $E(L_{r/2})^2 < \infty$, so by the Jain–Marcus CLT (1975) $X \in$ CLT in $C(T)$.

However, notice that (for $0 < r < 1$) (7.4) is a stronger condition than (7.2).

If $1 < r \le 2$, an application of Theorem 5.4 gives the CLT for uniformly bounded processes satisfying condition (7.3) for $L \in L_r(\Omega)$ and ρ verifying (7.4). However this is no improvement on the Jain–Marcus CLT.

Next we refine a result of Kolčinskiĭ for processes whose sample paths have bounded p-variation.

A function $x\colon [0, 1] \to \mathbb{R}$ is of bounded p-variation if

$$V_p^p(x) := \sup \textstyle\sum_{j=0}^{r-1} |x(t_{j+1}) - x(t_j)|^p < \infty$$

where the supremum is taken over all finite partitions $0 = t_0 < t_1 < \cdots < t_r = 1$.

By definition there exist $N = N_{n,p}(\varepsilon, \mathscr{F})$ ordered points in $[0, 1]$, $t_1 < \cdots < t_N$, such that $d_{n,p}(t_i, t_j) \geq \varepsilon/2$ for $i \neq j$. Then for $p \geq 1$ and $N_{n,1}(\varepsilon) \geq 2$ we have

$$\tag{7.5} \begin{aligned} \frac{1}{n} \textstyle\sum_{i=1}^{n} V_p^p(X_i) &\geq \frac{1}{n} \textstyle\sum_{j=1}^{N-1} \sum_{i=1}^{n} |X_i(t_{j+1}) - X_i(t_j)|^p \\ &\geq \frac{\varepsilon^p(N-1)}{2^p} \geq \frac{\varepsilon^p N_{n,1}(\varepsilon)}{2^{p+1}}. \end{aligned}$$

This simple observation (due to Kolčinskiĭ, 1981) together with Theorem 5.1 gives the following:

7.3. THEOREM. *Let $\{X(t)\colon t \in [0, 1]\}$ be a centered, uniformly bounded process whose trajectories are right continuous and have left limits at every $t \in [0, 1]$. Assume that X is a pregaussian process and that for some $p \geq 1$*

$$\tag{7.6} \lim_{u\to\infty} u^2 \Pr\{\ln_+ V_p(X) > u\} = 0.$$

Then $X \in$ CLT in $(D[0, 1], \|\cdot\|_\infty)$, and if X is sample continuous then $X \in$ CLT in $C[0, 1]$.

PROOF. By Theorem 5.1 it is enough to verify condition (5.15). For n large enough we have (using (7.5) and (7.6)):

$$\begin{aligned} &\Pr{}^*\{\ln N_{n,1}(\gamma\varepsilon/n^{1/2}, \mathscr{F}) > \varepsilon n^{1/2}\} \\ &\quad\leq \Pr\left\{\frac{2^{p+1} n^{p/2-1}}{\gamma^p \varepsilon^p} \textstyle\sum_{i=1}^{n} V_p^p(X_i) > \exp(\varepsilon n^{1/2})\right\} \\ &\quad\leq n \Pr\left\{V_p^p(X) > \frac{\gamma^p \varepsilon^p}{2^{p+1} n^{p/2}} \exp(\varepsilon n^{1/2})\right\} \leq n \Pr\left\{\ln_+ V_p(X) > \frac{\varepsilon n^{1/2}}{2p}\right\} \\ &\quad\to 0 \quad \text{as} \quad n \to \infty. \quad \square \end{aligned}$$

We note, in connection with the previous theorem, that the L_2-entropy condition together with only finiteness of V_p are not sufficient for the CLT: see the last example in Strassen and Dudley (1969).

Theorem 7.3 has a surprising corollary on the CLT under incremental moment conditions:

7.4. THEOREM. *Let $\{X(t)\colon t \in [0, 1]\}$ be a centered stochastically continuous uniformly bounded process whose trajectories are right continuous and have left limits at every $t \in [0, 1]$. Assume that for some $c > 0$, some nondecreasing function*

$F \in D[0, 1]$ and for all $s, t \in [0, 1]$,

$$E\,|\,X(t) - X(s)\,| \le C\,|\,F(t) - F(s)\,|. \tag{7.7}$$

Then $X \in$ CLT in $(D[0, 1], \|\cdot\|_\infty)$, and if X is sample continuous, then $X \in$ CLT in $C[0, 1]$.

PROOF. It is obvious that boundedness and (7.7) implies the L_2-metric entropy condition and hence X is pregaussian. Now if $\pi_n = \{t_i^n\}$, $n = 1, \cdots$, is a sequence of nested finite partitions increasing to a dense set, then for any function x in $D(0, 1]$,

$$V(x) = \uparrow\lim_{n\to\infty} \textstyle\sum_i |\,x(t_{i+1}^n) - x(t_i^n)\,|.$$

Therefore, by (7.7) and monotone convergence,

$$EV(X) = \lim_{n\to\infty} \textstyle\sum_i E\,|\,X(t_{i+1}^n) - X(t_i^n)\,| \le C(F(1) - F(0)) := C'.$$

Hence, we can apply Chebyshev's inequality to get

$$\lim_{u\to\infty} u^2 \Pr\{\ln_+ V(X) > u\} \le \lim_{u\to\infty} C' u^2 e^{-u} = 0.$$

This proves (7.6) and the result follows from Theorem 7.3. □

This result should be compared to Theorem 12.3 and 15.6 in Billingsley (1968). See also Hahn (1978) and Pisier (1980).

7.5. EXAMPLES. Let θ be a uniform random variable on the circle $\mathbb{T}$ (note that Theorem 7.4 also holds for processes on the circle $\mathbb{T}$). Let $\tau = \tau_k \in (0, 1)$ with probability p_k, be independent of θ. Define

$$X(t) = \varepsilon I_{[\theta-\tau,\theta+\tau]}(t)$$

where $\varepsilon = \pm 1$ with probability ½ and is independent of θ and τ. Then the variation $V_1(X)$ equals 2 and since

$$E\,|\,X(t) - X(s)\,| \le \textstyle\sum_k p_k \min(2\,|\,t - s\,|, 4\tau_k) \le 2\,|\,t - s\,|,$$

the L_2-entropy condition holds. Hence Theorem 7.3 applies (alternatively Theorem 7.4 applies with $F(t) = t$). Next we give an example showing that Theorem 7.3 and 7.4 are not true for general unbounded processes. Let now $\{C_k\}$ be an unbounded sequence and choose $\{p_k\}$ such that for $1 < r < 2$, $\sum p_k C_k^r = \infty$, but $\sum p_k C_k < \infty$. Then, for $Y = CX$, where $C = C_k$ when $\tau = \tau_k$, $V_1(Y) = 2C$ and therefore $EV_1(Y) < \infty$. In fact

$$E\,|\,Y(t) - Y(s)\,| \le 2(\textstyle\sum_k p_k C_k)\,|\,t - s\,|.$$

However $E\,\|\,Y\,\|_\infty^r = \infty$ and therefore $Y \notin$ CLT (see e.g. Jain, 1976, Theorem 5.7). With some extra care Y can even be chosen to be pregaussian.

The Jain–Marcus CLT implies (see Remark 7.2) that if a uniformly bounded process $X(t)$ satisfies that for some $\alpha, \beta > 0$ and all $s, t \in [0, 1]$

$$|X(t) - X(s)| \le L(\omega)|t - s|^{\alpha}, \quad EL^{\beta} < \infty$$

then $X \in$ CLT in $C[0, 1]$. Next we show, as another application of Theorem 5.1, that the integrability requirement on L can be considerably weakened for $\alpha \le 1$.

7.6. THEOREM. *Let $\{X(t): t \in [0, 1]\}$ be a centered stochastically continuous uniformly bounded stochastic process. Assume there exist $r \le 1$, a nondecreasing function $F \in D[0, 1]$ and a real random variable $L > 0$ such that*

$$(7.8) \qquad u^2 \Pr\{\ln_+ L > u\} \to 0 \quad as \quad u \to \infty,$$

and for which

$$(7.9) \qquad |X(t, \omega) - X(s, \omega)| \le L(\omega)(F(t) - F(s))^r \quad \omega \in \Omega, \quad 0 \le s \le t \le 1.$$

Then $X \in$ CLT in $(D[0, 1], \|\cdot\|_\infty)$ (and $X \in$ CLT in $C[0, 1]$ if X is sample continuous).

PROOF. We first show that under (7.8) the L_2-metric entropy condition for σ_X holds. By (7.8) there exists $e < u_0 < \infty$ such that

$$\Pr\{L > u\} \le (\ln u)^{-2} \quad \text{for} \quad u \ge u_0.$$

Hence there exists a constant $b < \infty$ such that

$$E|X(t) - X(s)|^2$$
$$\le 2\int_0^2 u \Pr\{L|F(t) - F(s)|^r > u\}\, du \le b\left[\ln\left(\frac{b}{|F(t) - F(s)|}\right)\right]^{-2}.$$

This then yields $(\ln N(\varepsilon, \sigma_X))^{1/2} \le K\varepsilon^{-1/2}$ for some $K < \infty$, which implies that X satisfies the L_2-entropy condition.

Now we obtain from (7.5) that

$$\varepsilon^p N_{n,1}(\varepsilon)/2^{p+1} \le \left(\textstyle\sum_{i=1}^n L_i^p/n\right)(F(1) - F(0))$$

where $p = r^{-1} \ge 1$, L_i are i.i.d. copies of L and we assume $N_{n,1}(\varepsilon) \ge 2$. If we let $C = F(1) - F(0)$, then we have that for n large enough

$$\Pr{}^*\{\ln N_{n,1}(\gamma\varepsilon/n^{1/2}, \mathscr{F}) > \varepsilon n^{1/2}\}$$
$$\le \Pr\left\{\frac{2^{p+1}Cn^{p/2}}{\gamma^p\varepsilon^p}\frac{\sum_{i=1}^n L_i}{n} > \exp(\varepsilon n^{1/2})\right\}$$
$$\le n\Pr\left\{L > \frac{\gamma^p\varepsilon^p}{2^{p+1}Cn^{p/2}}\exp(\varepsilon n^{1/2})\right\} \le n\Pr\left\{\ln_+ L > \frac{\varepsilon n^{1/2}}{2}\right\} \to 0$$

as $n \to \infty$ by (7.8), and Theorem 5.1 applies. □

7.7. REMARK. Theorem 7.6 can also be obtained from Theorem 5.3 by using essentially the same computations.

8. Remarks on the unbounded case. In this section we are concerned with the CLT and the law of large numbers in the unbounded case.

(a) *The law of large numbers.* In this subsection we treat the strong law of large numbers. Our results extend those in Vapnik and Červonenkis (1981) for uniformly bounded $\mathscr{F}$.

8.1. DEFINITION. A family of measurable functions $\mathscr{F} \subset \mathscr{L}_1(P)$ satisfies the Glivenko–Cantelli theorem for P (and we write $\mathscr{F} \in \mathrm{GC(P)}$) if

$$\Pr^*\left\{ \left\| \frac{\sum_{i=1}^n (f(X_i) - Pf)}{n} \right\|_{\mathscr{F}} \not\to 0 \right\} = 0. \tag{8.1}$$

8.2. REMARKS. Assume $\mathscr{F}$ is NLSM(P). Then: (1) If $E\|f(X_1)\| < \infty$, a.s. convergence to zero of the random variables (8.1) is equivalent to their convergence in probability (Steele, 1978, Kuelbs and Zinn, 1979, Pollard, 1981). The same remark applies to

$$\left\{ \left\| \frac{\sum_{i=1}^n \varepsilon_i f(X_i)}{n} \right\|_{\mathscr{F}} \right\}_{n=1}^{\infty}. \tag{8.2}$$

(2) If the sequence in (8.1) converges a.s., then the Borel–Cantelli lemma implies that

$$E\|f(X_1) - Pf\|_{\mathscr{F}} < \infty. \tag{8.3}$$

Hence, for $\mathscr{L}_1(P)$-bounded families $\mathscr{F}$, the condition

$$E\|f(X_1)\|_{\mathscr{F}} < \infty \tag{8.4}$$

is necessary for $\mathscr{F} \in \mathrm{GC(P)}$.

(3) If (8.4) holds, then one shows by the usual proof of the law of large numbers in $\mathbb{R}$ (i.e., using truncation, recentering and Chebyshev's inequality) that

$$\sup_{f\in\mathscr{F}}\Pr\left\{ \left| \frac{\sum_{i=1}^n (f(X_i) - Pf)}{n} \right| > \varepsilon \right\} \to 0$$

for all $\varepsilon > 0$. Hence Lemma 2.7 applies and we conclude that a NLSM(P) $\mathscr{L}_1(P)$-bounded class $\mathscr{F}$ satisfies the Glivenko–Cantelli theorem for P if and only if both (8.4) holds and

$$\left\| \frac{\sum_{i=1}^n \varepsilon_i f(X_i)}{n} \right\|_{0\mathscr{F}} \to 0 \quad \text{in probability.} \tag{8.5}$$

The following theorem and its corollaries generalize the Vapnik–Červonenkis (1981) law of large numbers (see Theorem 2.19) to unbounded classes. It also generalizes Steele's (1978) strengthening of the VC law of large numbers for sets.

Here and in what follows we let

$$F(x) = \sup_{f\in {}_0\mathscr{F}} |f(x)|, \quad x \in S.$$

It is appropriate for the reader at this point to recall Lemma 2.20.

8.3. THEOREM. *Let* $\mathscr{F}$ *be a* NLSM(P) $\mathscr{L}_1(P)$*-bounded class of functions. Then* $\mathscr{F} \in$ GC(P) *if and only if both:*

(i) $F \in \mathscr{L}_1(P)$, *and*

(ii) $$\lim_{n\to\infty} \frac{E^*\ln N_{n,2}(\varepsilon, {}_0\mathscr{F}_M)}{n} = 0 \quad \textit{for all} \quad M < \infty \quad \textit{and} \quad \varepsilon > 0,$$

where ${}_0\mathscr{F}_M = \{fI_{[F\le M]}: f \in {}_0\mathscr{F}\}$.

PROOF. As usual we use $\mathscr{F}$ to denote ${}_0\mathscr{F}$.

Sufficiency of (i) *and* (ii): Let $f_M = fI_{[F\le M]}$ and $f^M = f - f_M$, $f \in \mathscr{F}$. Then as in Pollard (1982) page 243, we have that for all $\varepsilon > 0$,

$$\lim_{M\to\infty}\lim_{n\to\infty}\Pr\left\{\sup_{f\in\mathscr{F}}\left|\frac{\sum_{i=1}^n (f^M(X_i) - Pf^M)}{n}\right| > \varepsilon\right\}$$

(8.6)

$$\le \lim_{M\to\infty}\lim_{n\to\infty}\Pr\left\{\frac{\sum_{i=1}^n (F^M(X_i) - PF^M)}{n} + 2\ PF^M > \varepsilon\right\} = 0.$$

Hence, it is enough to prove that for all $0 < M < \infty$,

$$\Pr - \lim_{n\to\infty}\sup_{f\in\mathscr{F}} \frac{|\sum_{i=1}^n \varepsilon_i f_M(X_i)|}{n} = 0 \tag{8.7}$$

(see Remark 8.2 (3), (8.5)). For this we apply Theorem 2.15 (2.24), as follows:

$$\Pr\left\{\left\|\frac{\sum_{i=1}^n \varepsilon_i f(X_i)}{n}\right\|_{\mathscr{F}_M} > \varepsilon\right\}$$

$$\le E_X\left[1 \wedge \left(\frac{1}{\varepsilon n^{1/2}} E_\varepsilon \left\|\frac{\sum_{i=1}^n \varepsilon_i f(X_i)}{n^{1/2}}\right\|_{\mathscr{F}_M}\right)\right]$$

$$\le CE^* \int_0^{2M} \left(\frac{\ln N_{n,2}(\varepsilon, \mathscr{F}_M)}{n}\right)^{1/2} d\varepsilon + 8\varepsilon^{-1}CM(\ln \ln 4)^{1/2}n^{-1/2} \tag{8.8}$$

$$\le C \int_0^{2M} \left(\frac{E^*\ln N_{n,2}(\varepsilon, \mathscr{F}_M)}{n}\right)^{1/2} d\varepsilon + 8\varepsilon^{-1}CM(\ln \ln 4)^{1/2}n^{-1/2}$$

$$\to 0 \quad \text{as} \quad n \to \infty,$$

where in the last inequality we use the fact that $\int_0^{2M}$ is a Riemann integral, and convergence to zero follows by dominated convergence and hypothesis (ii) (just note that $N_{n,2}(\varepsilon, \mathscr{F}_M) \le N_{n,\infty}(\varepsilon, \mathscr{F}_M, P) \le (M/\varepsilon)^n$, so that $((\ln N_{n,2}(\varepsilon, \mathscr{F}_M))/n)^{1/2} \le (\ln(M/\varepsilon))^{1/2}$, which is integrable on $[0, M]$). So, (8.7) is proved, and by (8.6), the proof of sufficiency is completed.

Necessity of (i) *and* (ii). Necessity of (i) follows from Remark 8.2 (1). By inequality (2.5) we have

$$(8.9) \qquad E\left\| \frac{\sum_{i=1}^n g_i f(X_i)}{n} \right\|_{\mathscr{F}} \le M_2 \max_{n_0<k\le n} E \left\| \frac{\sum_{i=1}^k \varepsilon_i f(X_i)}{k} \right\|_{\mathscr{F}} + o(1).$$

Since $EF < \infty$, we have $E \max_{k\le n} \| f(X_k)/n \|_{\mathscr{F}} \to 0$ as $n \to \infty$. So, by Hoffmann-Jørgensen's inequality (Lemma 2.8), if $\mathscr{F} \in \mathrm{GC(P)}$ and is NLSM(P), then

$$E\left\| \frac{\sum_{i=1}^n \varepsilon_i f(X_i)}{n} \right\|_{\mathscr{F}} \to 0 \quad \text{as} \quad n \to \infty.$$

Therefore, by (8.9)

$$E\left\| \frac{\sum_{i=1}^n g_i f(X_i)}{n} \right\|_{\mathscr{F}} \to 0 \quad \text{as} \quad n \to \infty.$$

Then the Sudakov minoration (Theorem 2.16) implies

$$(8.10) \qquad n^{-1/2} E^* \sup_{\lambda>0} \lambda (\ln N_{n,2}(\lambda, \mathscr{F}))^{1/2} \to 0 \quad \text{as} \quad n \to \infty.$$

Since $N_{n,2}(\lambda, \mathscr{F}) \ge N_{n,2}(\lambda, \mathscr{F}_M)$ for all $M < \infty$ it follows that

$$(8.11) \qquad E^* \frac{\ln N(\varepsilon, \mathscr{F}_M)}{n} \le \left(\ln\left(\frac{2M}{\varepsilon}\right)\right)^{1/2} E^*\left(\frac{\ln N(\varepsilon, \mathscr{F}_M)}{n}\right)^{1/2} \to 0$$

for all $\varepsilon > 0$. □

The previous proof shows also the following:

8.4. COROLLARY. *Let* $\mathscr{F}$ *be a* NLSM(P) $\mathscr{L}_1(P)$*-bounded class of functions. Then* $\mathscr{F} \in \mathrm{GC(P)}$ *if and only if both*:

(i) $F \in \mathscr{L}_1(P)$ *and*

(ii) $\lim_{n\to\infty} E^*\left(\frac{\ln N_{n,2}(\varepsilon, {}_0\mathscr{F})}{n}\right)^{1/2} = 0$ *for all* $\varepsilon > 0$.

Since $N_{n,p}(\varepsilon, \mathscr{F})$ increases with p, and also, on ${}_0\mathscr{F}_M$, $d_{n,2} \le (2M)^{(2-p)/2} d_{n,p}^{(p\vee 1)/2}$ if $0 < p < 2$ and $d_{n,2} \ge (2M)^{(2-p)/2} d_{n,p}^{p/2}$ if $2 < p < \infty$, it follows that in Theorem 8.3, $N_{n,2}$ can be replaced by $N_{n,p}$ for any $p \in (0, \infty)$ for necessity and $p \in (0, \infty]$ for sufficiency of conditions (i) and (ii). Combining this observation with the necessity part of Theorem 2.19 gives:

8.5. COROLLARY. *Let* $\mathscr{F}$ *be a* NLSM(P) $\mathscr{L}_1(P)$*-bounded class of functions. Then* $\mathscr{F} \in \mathrm{GC(P)}$ *if and only if both*:

(i) $F \in \mathscr{L}_1(P)$ *and*

$$\text{(ii)} \qquad \lim_{n\to\infty} \frac{E^*\ln N_{n,p}(\varepsilon, {}_0\mathscr{F}_M)}{n} = 0 \quad \textit{for all} \quad \varepsilon > 0, \quad M < \infty,$$

and some (*all*) $p \in (0, \infty]$.

Another consequence:

8.6. COROLLARY. *Let* $\mathscr{F}$ *be a* LSM(P) *uniformly bounded class of functions. Then if* $\mathscr{F}$ *satisfies the Glivenko–Cantelli theorem for P, so does*

$$H\mathscr{F} = \{Hf: f \in \mathscr{F}\} \quad \textit{for all} \quad H \in \mathscr{L}_1(P).$$

PROOF. Let $M = \sup\{|f(s)|: s \in S, f \in \mathscr{F}\}$, and $k \equiv M$. By considering $M^{-1}[\mathscr{F} \cup \{k\}]$ we may assume $\sup_{f\in\mathscr{F}} |f(s)| = 1$ for all $s \in S$. Then, for every $0 < M < \infty$ and $f, g \in \mathscr{F}$ we have $d_{n,2}(HfI_{[|H|\le M]}, HgI_{[|H|\le M]}) \le Md_{n,2}(f, g)$, so that $N_{n,2}(\varepsilon, (H\mathscr{F})_M) \le N_{n,2}(\varepsilon/M, \mathscr{F})$. Hence Theorem 8.3 applies. □

The following, which is used below, is easy to see from a close look at the proof of Theorem 8.3 (see (8.8)):

8.7. COROLLARY. *Let* $\mathscr{F}$ *be a* NLSM(P) $\mathscr{L}_1(P)$ *bounded class. Then* $\mathscr{F} \in$ GC(P) *if and only if both* $F \in \mathscr{L}_1(P)$ *and*

$$\text{(8.12)} \qquad \lim_{n\to\infty} E^*\left(1 \wedge n^{-1/2} \int_0^{2M} (\ln N_{n,2}(\varepsilon, {}_0\mathscr{F}_M))^{1/2}\, d\varepsilon\right) = 0$$

for all $M < \infty$.

Finally we extend Theorem 2.17 ((a) $\Leftrightarrow$ (c)) (only necessity is interesting).

8.8. COROLLARY. *Let* $\mathscr{F}$ *be* $\mathscr{L}_1$-*(P)-bounded and assume that it is* NLSM(P) *and that* $\mathscr{G} := \{|f - g|: f, g \in \mathscr{F}\}$ *is* LSM(P). *Then* $\mathscr{F} \in$ GC(P) *if and only if both* $F \in \mathscr{L}_1(P)$ *and there exists a finite function* $T(\varepsilon)$, $\varepsilon > 0$, *such that*

$$\text{(8.13)} \qquad \lim_{n\to\infty} \Pr{}^*\{N_{n,1}(\varepsilon, {}_0\mathscr{F}) > T(\varepsilon)\} = 0, \quad \varepsilon > 0.$$

PROOF. As in Theorem 8.3, in order to prove sufficiency it is enough to show that (8.7) holds for all M whenever (8.13) does. Now, the usual arguments (see, e.g., the proof of Theorem 5.1) give

$$\Pr\left\{ \left\| \frac{\sum_{i=1}^n \varepsilon_i f(X_i)}{n} \right\|_{{}_0\mathscr{F}_M} > 2\varepsilon \right\}$$

$$\le \Pr{}^*\{N_{n,1}(\varepsilon, {}_0\mathscr{F}_M) > T(\varepsilon)\} + T(\varepsilon) E_X \sup_{f\in{}_0\mathscr{F}_M} P_\varepsilon\left[\left| \frac{\sum_{i=1}^n \varepsilon_i f(X_i)}{n} \right| > \varepsilon \right\}$$

$$\le \Pr{}^*\{N_{n,1}(\varepsilon, {}_0\mathscr{F}) > T(\varepsilon)\} + T(\varepsilon) M^2 \varepsilon^{-2} n^{-1} \to 0 \quad \text{as} \quad n \to \infty.$$

Necessity follows approximately as in the proof of Corollary 4 in Vapnik and

Červonenkis (1981). Assume $\mathscr{F}$ satisfies the GC theorem. Then, by Theorem 8.4, so does $\mathscr{G}$, that is

$$\Pr\{\sup_{f,g\in{}_0\mathscr{F}} | d_{n,1}(f, g) - \| f - g \|_1 | > \varepsilon\} \to 0$$

for all $\varepsilon > 0$. Therefore for any $\varepsilon > 0$ there exist $n(\varepsilon)$ and measurable sets $\Omega_{n,\varepsilon} \subset S^{\mathbb{N}}$, $n \geq n(\varepsilon)$, such that $\Pr(\Omega_{n,\varepsilon}) \to 1$ as $n \to \infty$ and

$$\sup_{f,g\in{}_0\mathscr{F}} | d_{n,1}(\omega)(f, g) - \| f - g \|_1 | \leq \varepsilon \quad \text{for all} \quad \omega \in \Omega_{n,\varepsilon}.$$

On the other hand, since ${}_0\mathscr{F}$ is $d_{n,2}(\omega)$-totally bounded for almost every ω (also by Theorem 8.6) and $d_{n,1} \leq d_{n,2}$, it follows that there exists a measurable set $\Omega \subset S^{\mathbb{N}}$ with $\Pr(\Omega) = 1$ such that ${}_0\mathscr{F}$ is $d_{n,1}(\omega)$-totally bounded for all $\omega \in \Omega$. These two facts imply that

$$N(2\varepsilon, {}_0\mathscr{F}, \| \cdot \|_1) \leq N_{n,1}(\omega)(\varepsilon, {}_0\mathscr{F}) < \infty$$

for all $\omega \in \Omega_{n,\varepsilon} \cap \Omega$. This proves the corollary with $T(\varepsilon) = N(2\varepsilon, {}_0\mathscr{F}, \| \cdot \|_1)$. □

Note that the sufficiency in the previous corollary does not require any measurability assumptions on $\mathscr{G}$.

8.9. Remark. It is interesting that under appropriate measurability,

$$\Pr^*\text{-}\lim_{n\to\infty} \frac{\ln N_{n,1}(\varepsilon, {}_0\mathscr{F}_M)}{n} = 0 \quad \text{for all} \quad \varepsilon > 0 \quad \text{and} \quad M > 0$$

implies the existence of a finite $T(\varepsilon)$ for which (8.13) holds. Compare with Theorem 2.22.

(b) *The central limit theorem.* Here is a general CLT for unbounded families of functions. This theorem combines some of the ideas in Giné and Marcus (1981) (which in turn originated in Marcus, 1978, and Jain and Marcus, 1975) and Pollard (1982). In what follows $\mathscr{F}$ is NLDM(P) and we let $F = \sup_{f\in{}_0\mathscr{F}} | f |$.

8.10. Theorem. *Let* $\mathscr{F} \subset \mathscr{L}_2(P)$ *be* NLDM(P) *and such that* $F \in \mathscr{L}_2(P)$. *Let* $\mathscr{G} := \{(f-g)^2 : f, g \in {}_0\mathscr{F}\}$ *be* NLSM(P). *Assume*

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} E^*\left(1 \wedge \int_0^\delta (\ln N_{n,2}(\varepsilon, {}_0\mathscr{F}))^{1/2}\, d\varepsilon\right) = 0. \tag{8.14}$$

Then $\mathscr{F}$ *is a functional P-Donsker class.*

Proof. We let $\mathscr{F} = {}_0\mathscr{F}$ in this proof as usual. We first show that

$$\mathscr{G} = \{(f-g)^2 : f, g \in \mathscr{F}\} \in \mathrm{GC(P)}. \tag{8.15}$$

Note that $G = \sup_{g\in{}_0\mathscr{G}} | g | \leq 4F^2 \in \mathscr{L}_1(P)$ so that the integrability hypothesis in

Corollary 8.7 holds. Since for $f_i, g_i \in \mathscr{F}$, $i = 1, 2$,

$$|(f_1 - g_1)^2 - (f_2 - g_2)^2| I_{[G\le M]} \le |f_1 - f_2 + g_2 - g_1| |f_1 - g_1 + f_2 - g_2| I_{[G\le M]}$$
$$\le 2M^{1/2}(|f_1 - f_2| + |g_1 - g_2|),$$

it follows that

$$N_{n,2}(\varepsilon, {}_0\mathscr{G}_M) \le N_{n,2}^2(\varepsilon/4M^{1/2}, \mathscr{F}).$$

To verify (8.12) for $\mathscr{G}$ note that for n large enough

$$n^{-1/2}\int_\delta^{2M} (\ln N_{n,2}(\varepsilon, \mathscr{G}_M))^{1/2}\, d\varepsilon \le 2Mn^{-1/2}\left(2 \ln N_{n,2}\left(\frac{\delta}{4M^{1/2}}, \mathscr{F}\right)\right)^{1/2}$$
$$\le \int_0^{\delta/4M^{1/2}} (\ln N_{n,2}(\varepsilon, \mathscr{F}))^{1/2}\, d\varepsilon.$$

Therefore (8.14) implies (8.12) for $\mathscr{G}$ and, by Corollary 8.7, that $\mathscr{G}$ is GC(P).

Next we show that $\mathscr{F}$ is totally bounded for e_P. Given $\varepsilon > 0$, the set

$$\left\{\omega \in \Omega: \sup_{g\in\mathscr{G}} \left|\frac{\sum_{i=1}^n g(X_i(\omega))}{n} - Pg\right| < \varepsilon\right\} \cap \{\omega \in \Omega: N_{n,2}(\varepsilon^{1/2}, \mathscr{F})(\omega) < \infty\}$$

has strictly positive inner probability for n large enough by (8.14) and (8.15). So, for ω in this set we can find $N_{n,2}(\varepsilon^{1/2}, \mathscr{F})(\omega) < \infty$ functions $g_i \in \mathscr{F}$ such that $\min_i P(f - g_i) < 2\varepsilon$, i.e. $\mathscr{F}$ is totally bounded for e_P.

Hence by Theorem 2.14 it is enough to prove the symmetrized probable equicontinuity condition (2.12). Now, for $\varepsilon, \sigma \ge 0$

$$\Pr\left\{\sup_{f,g\in\mathscr{F}; e_P(f,g)\le\delta} \left|\frac{\sum_{i=1}^n \varepsilon_i(f-g)(X_i)}{n^{1/2}}\right| > \varepsilon\right\}$$
$$\le \Pr\left\{\sup_{h\in\mathscr{G}}\left|\frac{\sum_{i=1}^n h(X_i)}{n} - Ph\right| > \delta^2\right\}$$
$$+ E_X P_\varepsilon\left\{\sup_{f,g\in\mathscr{F}; e_P(f,g)<\delta}\left|\frac{\sum_{i=1}^n \varepsilon_i(f-g)(X_i)}{n^{1/2}}\right| > \varepsilon,\right. \tag{8.16}$$
$$\left.\sup_{f,g\in\mathscr{F}} |e_P^2(f,g) - d_{n,2}^2(f,g)| \le \delta^2\right\}$$
$$:= (\mathrm{I}) + (\mathrm{II}).$$

But,

$$(\mathrm{I}) \to 0 \quad \text{as} \quad n \to \infty \tag{8.17}$$

for all $\delta > 0$ by (8.15). And we apply Theorem 2.15 (2.23) to bound (II): for

$0 < \varepsilon < C$ (C is as in (2.23)) we have

$$(\mathrm{II}) < E_X\left[1 \wedge \left(\frac{1}{\varepsilon} E_\varepsilon \sup_{f,g\in\mathscr{F};e_P(f,g)<\delta}\left|\frac{\sum_{i=1}^n \varepsilon_i(f-g)(X_i)}{n^{1/2}}\right|\right)\right]$$

$$\le \frac{1}{\varepsilon} E_X^*\left[\varepsilon \wedge C\left\{\int_0^{2^{1/2}\delta} (\ln N_{n,2}(u, \mathscr{F}))^{1/2}\,du + \Phi(2^{1/2}\delta)\right\}\right] \tag{8.18}$$

$$\le \frac{C}{\varepsilon} E_X^*\left[1 \wedge \int_0^{2^{1/2}\delta} (\ln N_{n,2}(u, \mathscr{F}))^{1/2}\,du\right] + \frac{C}{\varepsilon}\Phi(2^{1/2}\delta).$$

Then (8.14) and (8.18) give

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}(\mathrm{II}) = 0. \tag{8.19}$$

(8.16)–(8.19) prove condition (2.12) in Theorem 2.14. □

8.11. REMARKS. (a) Using Theorem 2.14 ((a) ⇒ (e)) and Sudakov's minoration (Theorem 2.16), we get that a necessary condition for $\mathscr{F} \subset \mathscr{L}_2^0(P)$ and NLDM(P) to be a funtional P-Donsker class is that

$$\lim_{\delta\downarrow 0}\limsup_n E^*\sup_{\lambda>0}\lambda(\ln N_{n,2}(\lambda, {}_0\mathscr{F}_\delta', P))^{1/2} = 0.$$

(b) The idea of using the law of large numbers to replace e_P by $d_{n,1}$ in (8.16) is due to Pollard (1982).

Theorem 8.10 should be compared with Theorem 5.8; it is not difficult to check that Theorem 8.10 is weaker than Theorem 5.8 in the uniformly bounded case.

Next we prove Pollard's theorem (Pollard, 1982); this theorem is used by Pollard to prove that if $\mathscr{C}$ is a Vapnik–Červonenkis class and $F \in \mathscr{L}_2(P)$, then $\mathscr{F} = \{FI_C\colon C \in \mathscr{C}\}$ is a functional P-Donsker class of functions. Let $\mathscr{F} \subset \mathscr{L}_2(P)$. For any finite set $T \subset S$, define, following Pollard (1982),

$$\delta_T^{(2)}(f, g) = \left(\frac{\sum_{x\in T}(f-g)^2(x)}{\sum_{x\in T}F^2(x)}\right)^{1/2}, \tag{8.20}$$

and for all $\varepsilon > 0$

$$N^{(2)}(\varepsilon, \mathscr{F}) = \sup_{T\subset S, T\,\mathrm{finite}} N(\varepsilon, \mathscr{F}, \delta_T^{(2)}). \tag{8.21}$$

Then, Pollard's CLT is as follows:

8.12. COROLLARY. *Let $\mathscr{F}$ (and $\mathscr{G}$) satisfy the measurability hypothesis of* 8.10 *and let $\mathscr{F} \in \mathscr{L}_2(P)$. Assume that*

$$\int_0^\infty (\ln N^{(2)}(\varepsilon, \mathscr{F}))^{1/2}\,d\varepsilon < \infty. \tag{8.22}$$

Then $\mathscr{F}$ is a functional P-Donsker class.

PROOF. We show that condition (8.22) implies (8.14) in Theorem 8.10 and this will conclude the proof of the theorem. Since

$$d_{n,2}^2(f, g) \le \|F\|_{n,2}^2(\delta_{T_n}^{(2)}(f, g))^2, \quad f, g \in \mathscr{F}$$

where $\|F\|_{\mathfrak{p},2}^2 = \sum_{i=1}^n F^2(X_i)/n$ and $T_n = \{X_1, \cdots, X_n\}$, it follows that

$$N_{n,2}(\varepsilon, \mathscr{F}) \le N^{(2)}\left(\frac{\varepsilon}{\|F\|_{n,2}}, \mathscr{F}\right)$$

so that

$$\int_0^\delta (\ln N_{n,2}(\varepsilon, \mathscr{F}))^{1/2}\, d\varepsilon$$

$$\le \|F\|_{n,2} \int_0^{\delta/\|F\|_{n,2}} (\ln N^{(2)}(u, \mathscr{F}))^{1/2}\, du$$

$$\le \|F\|_{n,2} \int_0^\delta (\ln N^{(2)}(u, F))^{1/2}\, du + \delta(\ln N^{(2)}(\delta, F))^{1/2}.$$

Since $E\|F\|_{n,2} \to (PF^2)^{1/2} < \infty$, condition (8.22) gives immediately that $\mathscr{F}$ satisfies condition (8.14) in Theorem 8.10. □

Acknowledgement. We thank: R. M. Dudley for sending us his 1982 manuscript, for directing our attention to the recent papers of Vapnik and Červonenkis (1981) and Kolčinskiĭ (1981) and for several discussions; L. Le Cam for sending us his unpublished work at a crucial moment of our research (we had a weaker version of Theorem 5.1—the present Theorem 5.8—which was improved by use of Lemma 5.2, based on Le Cam's work); M. Marcus for several stimulating conversations, particularly on Section 7; and D. Pollard for a careful reading of the manuscript (of course, any mistakes are ours alone).

REFERENCES

DE ACOSTA, A. (1981). Strong exponential integrability of sums of independent B-valued random vectors. *Probab. Math. Statist.* **1** 133–150.

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

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

BENNETT, G. (1962). Probability inequalities for sums of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

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

BOAS, R. P. (1960). Inequalities for monotonic series. *J. Math. Anal. Appl.* **1** 121–126.

BOAS, R. P. and MARCUS, M. B. (1973). Inequalities involving a function and its inverse. *SIAM J. Math. Anal.* **4** 585–591.

BORISOV, I. S. (1981). Some limit theorems for empirical distributions. (In Russian.) Abstracts of reports, *Third Vilnius Conf. Probability Th. Math. Statist.* **I** 71–72.

DUDLEY, R. M. (1967). The size 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. (1980). Some recent results on empirical processes. *Lecture Notes in Math.* **860** (Probability in Banach Spaces III), 107–123.
DUDLEY, R. M. (1981). Vapnik–Červonenkis Donsker classes of functions. In *Aspects Statistiques et Aspects Physiques des Processus Gaussiens.* Coll. Int. du CNRS, no. 307. Ed. du CNRS. Paris, 251–270.
DUDLEY, R. M. (1984). A course on empirical processes. *Lecture Notes in Math.* Springer, New York. To appear.
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 trajectoires des functions aleatoires gaussiennes. *Lecture Notes in Math.* **480** 1–96. Springer, Berlin.
FERNIQUE, X. (1983). Sur la convergence étroite des mésures gaussiennes. Publ. de l'IRMA, 2241/P-130, 27–34. Strasbourg.
GAENSSLER, P. and STUTE, W. (1979). Empirical processes: a survey of results for independent identically distributed random variables. *Ann. Probab.* **7** 193–242.
GINÉ, E. and MARCUS, M. B. (1981). On the central limit theorem in $C(K)$. In *Aspects Statistiques et Aspects Physiques des Processus Gaussiens.* Coll. Int. du CNRS, no. 307. Paris, 361–383.
GINÉ, E. and ZINN, J. (1983). Central limit theorems and weak laws of large numbers in certain Banach spaces. *Z. Wahrsch. verw. Gebiete* **62** 323–354.
HAHN, M. G. (1978). Central limit theorems in $D[0, 1]$. *Z. Wahrsch. verw. Gebiete* **44** 89–101.
HOFFMANN-JØRGENSEN, J. (1974). Sums of independent Banach space valued random variables. *Studia Math.* **52** 159–186.
JAIN, N. C. (1976). Central limit theorem in Banach space. *Lecture Notes in Math.* **526** 113–130. Springer, Berlin.
JAIN, N. C. and MARCUS, M. B. (1975). Central limit theorems for $C(S)$-valued random variables. *J. Funct. Anal.* **19** 216–231.
JAIN, N. and MARCUS, M. B. (1978). Continuity of subgaussian processes. *Adv. Probab.* **4** 81–196. Dekker, New York.
KAHANE, J.-P. (1968). *Some Random Series of Functions.* Heath, Lexington, Massachusetts.
KOLČINSKIĬ, V. I. (1981). On the central limit theorem for empirical measures. *Theory Probab. Math. Statist.* **24** 71–82.
KUELBS, J. and ZINN, J. (1979). Some stability results for vector valued random variables. *Ann. Probab.* **7** 75–84.
KUELBS, J. and ZINN, J. (1983). Some results on LIL behavior. *Ann. Probab.* **11** 506–557.
LE CAM, L. (1983a). A remark on empirical measures. In *A Festschrift for Erich L. Lehman in Honor of His Sixty-Fifth Birthday* 305–327. Wadsworth, Belmont, California.
LE CAM, L. (1983b). Manuscript of a forthcoming book.
MARCUS, M. B. (1978). Continuity and the central limit theorem for random trigonometric series. *Z. Wahrsch. verw. Gebiete* **42** 35–36.
MARCUS, M. B. (1981). Weak convergence of the empirical characteristic function. *Ann. Probab.* **9** 194–201.
MARCUS, M. B. and PISIER, G. (1978). *Random Fourier Series with Applications to Harmonic Analysis.* Princeton U. Press, Princeton, New Jersey.
MARCUS, M. B. and PISIER, G. (1982). Characterizations of almost surely continuous p-stable random Fourier series and strongly stationary processes. *Acta Math.* To appear.
MARCUS, M. B. and SHEPP, L. A. (1972). Sample behavior of Gaussian processes. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **II** 423–439.
PISIER, G. (1975). Le théorème limite central et la loi du logarithme iteré dans les espaces de Banach. Séminaire Maurey-Schwartz 1975–76, exposés III et IV. Ecole Polytechnique, Paris.
PISIER, G. (1980). Conditions d'entropie assurant la continuité de certains processus et applications a l'Analyse Harmonique. Séminaire d'Analyse Fonctionelle 1979–80, Ecole Polytechnique, Paris. Exp XIII–XIV.

POLLARD, D. (1979). General chi-square goodness of fit tests with data dependent cells. *Z. Wahrsch. verw. Gebiete* **50** 317–331.

POLLARD, D. (1980). The minimum distance method of testing. *Metrika* **27** 43–70.

POLLARD, D. (1981). Limit theorems for empirical processes. *Z. Wahrsch. verw. Gebiete* **57** 181–185.

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

SAUER, N. (1972). On the density of families of sets. *J. Combin. Theory Ser. A* **13** 145–147.

STEELE, J. M. (1978). Empirical discrepancies and subadditive processes. *Ann. Probab.* **6** 118–127.

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

VAPNIK, V. N. and ČERVONENKIS, A. JA. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* **16** 164–280.

VAPNIK, V. N. and ČERVONENKIS, A. JA. (1981). Necessary and sufficient conditions for the uniform convergence of means to their expectations. *Theory Probab. Appl.* **26** 532–553.

YATRACOS, Y. (1983). Rates of convergence of minimum distance estimators and Kolmogorov's entropy. U. C. Berkeley, Ph.D. dissertation.

TEXAS A&M UNIVERSITY
DEPARTMENT OF MATHEMATICS
COLLEGE STATION, TEXAS 77843