The Annals of Probability
1987, Vol. 15, No. 4, 1306–1326

UNIVERSAL DONSKER CLASSES AND METRIC ENTROPY[1]

BY R. M. DUDLEY

Massachusetts Institute of Technology

Let $(X, \mathscr{A})$ be a measurable space and $\mathscr{F}$ a class of measurable functions on X. $\mathscr{F}$ is called a universal Donsker class if for every probability measure P on $\mathscr{A}$, the centered and normalized empirical measures $n^{1/2}(P_n - P)$ converge in law, with respect to uniform convergence over $\mathscr{F}$, to the limiting "Brownian bridge" process G_P. Then up to additive constants, $\mathscr{F}$ must be uniformly bounded. Several nonequivalent conditions are shown to imply the universal Donsker property. Some are connected with the Vapnik–Červonenkis combinatorial condition on classes of sets, others with metric entropy. The implications between the various conditions are considered. Bounds are given for the metric entropy of convex hulls in Hilbert space.

0. Introduction. When $\mathscr{F}$ is a universal Donsker class, then for independent, identically distributed (i.i.d.) observations $X_1, \dots, X_n$ with an unknown law P, for any f_i in $\mathscr{F}$, $i = 1, \dots, m$, $n^{-1/2}\{f_i(X_1) + \cdots + f_i(X_n)\}_{1 \le i \le m}$ is asymptotically normal with mean vector $n^{1/2}\{\int f_i(x)\, dP(x)\}_{1 \le i \le m}$ and covariance matrix $\int f_i f_j\, dP - \int f_i\, dP \int f_j\, dP$, uniformly for $f_i \in \mathscr{F}$. Then, for certain statistics formed from the $f_i(X_k)$, even where f_i may be chosen depending on the X_k, there will be asymptotic distributions as $n \to \infty$. For example, for χ^2 statistics, where f_i are indicators of disjoint intervals, depending suitably on $X_1, \dots, X_n$, whose union is the real line, X^2 quadratic forms have limiting distributions [Roy (1956) and Watson (1958)] which may, however, not be χ^2 distributions and may depend on P [Chernoff and Lehmann (1954)]. Universal Donsker classes of sets are, up to mild measurability conditions, just classes satisfying the Vapnik–Červonenkis combinatorial conditions defined later in this section [Durst and Dudley (1981) and Dudley (1984) Chapter 11]. The use of such classes allows a variety of extensions of the Roy–Watson results to general (multidimensional) sample spaces [Pollard (1979) and Moore and Stubblebine (1981)]. Vapnik and Červonenkis (1974) indicated applications of their families of sets to classification (pattern recognition) problems. More recently, the classes have been applied to tree-structured classification [Breiman, Friedman, Olshen and Stone (1984), Chapter 12].

The use of functions more general than indicators gives additional potentially useful freedom in constructing statistics. For example, there may be advantages to procedures based on spaces of smooth functions, which contain no nontrivial indicators. Le Cam, Mahan and Singh (1983) give a rather general extension of the Chernoff–Lehmann approach to "quadratic forms, or related objects."

Or, if $\mathscr{F}$ is an (infinite-dimensional) ellipsoid, the square of the supremum of an empirical measure over $\mathscr{F}$ is a sum of squared integrals, approximable by

Received August 1985; revised August 1986.

[1]This research was partially supported by National Science Foundation Grant DMS-8506638.

AMS 1980 *subject classifications*. Primary 60F17, 60F05; secondary 60G17, 60G20.

Key words and phrases. Central limit theorems, Vapnik–Červonenkis classes.

finite sums, thus easier to compute than suprema over most families of functions. Ellipsoids seem difficult to obtain as symmetric convex hulls of classes of indicators.

The present paper will give no specific statistical procedures, but rather a general approach to sufficient conditions for the universal Donsker property.

First, some terminology and previous results will be recalled. Most appeared in Dudley (1984). Let $(X, \mathscr{A}, P)$ be a probability space. Let $P(f) := \int f\,dP$ for each integrable function f. Let G_P be a Gaussian process indexed by $\mathscr{L}^2(X, \mathscr{A}, P)$ with mean 0 and covariance $EG_P(f)G_P(g) = P(fg) - P(f)P(g)$ for all f and g in $\mathscr{L}^2$. Let $\rho_P(f, g) := (E((G_P(f) - G_P(g))^2))^{1/2}$. Such a process will be called *coherent* if each sample function $G_P(\cdot)(\omega)$ is bounded and uniformly continuous on $\mathscr{F}$ with respect to ρ_P. A class $\mathscr{F} \subset \mathscr{L}^2$ will be called *pregaussian* or *P-pregaussian* (formerly called G_PBUC) iff a coherent G_P process exists on $\mathscr{F}$. Let $(X^\infty, \mathscr{A}^\infty, P^\infty)$ be the countable product of copies of $(X, \mathscr{A}, P)$, with coordinates $x(1), x(2), \dots$. Let $P_n := n^{-1}(\delta_{x(1)} + \cdots + \delta_{x(n)})$, the nth empirical measure for P. Let $\nu_n := n^{1/2}(P_n - P)$. Let $(\Omega, \mathscr{S}, \Pr)$ be the product of $(X^\infty, \mathscr{A}^\infty, P^\infty)$ and the unit interval with Lebesgue measure. For any real-valued function G on $\mathscr{F}$ let $\|G\|_{\mathscr{F}} := \sup\{|G(f)|: f \in \mathscr{F}\}$. Call $\mathscr{F}$ a *functional Donsker class* for P iff on Ω, there exist independent coherent G_P processes $Y_1, Y_2, \dots,$ such that for each $\varepsilon > 0$,

$$\lim_{n \to \infty} \Pr^*\left\{n^{-1/2} \max_{k \le n} \|k(P_k - P) - Y_1 - \cdots - Y_k\|_{\mathscr{F}} > \varepsilon\right\} = 0.$$

An alternate definition, due to Hoffmann-Jørgensen (1984), is that $\mathscr{F}$ be P-pregaussian and for every bounded $\|\cdot\|_{\mathscr{F}}$-continuous real function h on the set of all bounded functions on $\mathscr{F}$,

$$\int^* h(\nu_n)\,d\Pr \to \int h(G_P)\,d\Pr;$$

this definition is equivalent to the previous one according to Hoffmann-Jørgensen (1984), Talagrand (1987) and/or Dudley (1985), Theorem 5.2. Thus, if, in addition, $h(\nu_n)$ is Pr-measurable, it converges in law to $h(G_P)$. We have uniform convergence of $Eh(\nu_n)$ to $Eh(G_p)$ for h in any uniformly bounded, $\|\cdot\|_{\mathscr{F}}$-equicontinuous family of functions h for which $h(\nu_n)$ are measurable, as follows from an extended Wichura theorem [Dudley (1985), Theorem 4.1].

A collection $\mathscr{C}$ of sets is said to *shatter* a set F iff every subset of F is of the form $B \cap F$ for some $B \in \mathscr{C}$. The supremum of cardinalities of finite sets shattered by $\mathscr{C}$ will be called $S(\mathscr{C})$. A collection $\mathscr{C}$ is called a *Vapnik–Červonenkis* (VC) *class* iff $S(\mathscr{C}) < \infty$.

A measurable space $(S, \mathscr{B})$ will be called *standard* iff there exists a metric d on S for which (S, d) is a complete separable metric space and $\mathscr{B}$ is the Borel σ-algebra. A measurable space $(Y, \mathscr{U})$ will be called *Suslin* iff $\mathscr{U}$ is countably generated and separates points of Y and there is a standard space $(S, \mathscr{B})$ and a measurable function from S onto Y. Given a measurable space $(X, \mathscr{A})$, a collection $\mathscr{F}$ of measurable functions on X will be called *image admissible Suslin via* $(Y, \mathscr{U}, T)$ iff $(X, \mathscr{A})$ and $(Y, \mathscr{U})$ are Suslin and T is a function from Y

onto $\mathscr{F}$ such that the function $\langle x, y\rangle \mapsto T(y)(x)$ is jointly measurable. Here, equivalently, $(Y, \mathscr{U})$ can be taken to be standard.

For a measurable space $(X, \mathscr{A})$, a collection $\mathscr{F}$ of measurable real-valued functions on X will be called a *universal Donsker class* iff it is a functional Donsker class for every probability measure (law) P on $\mathscr{A}$. If $\mathscr{F}$ is a collection of indicators of sets, $\mathscr{F} = \{1_B\colon B \in \mathscr{C}\}$, and a universal Donsker class, or even pregaussian for all P, then $\mathscr{C}$ must be a Vapnik–Červonenkis class [Durst and Dudley (1981) and Dudley (1984), Theorem 11.4.1]. Conversely, an image admissible Suslin VC class is a universal Donsker class [Dudley (1978), Section 7, and (1984), Theorems 11.1.2 and 11.3.1]. The measurability hypotheses cannot be completely removed [Durst and Dudley (1981) and Dudley (1984), Theorem 11.4.2]. The question remains open as far as I know: What (measurability) conditions on a family of sets, together with the VC property, are equivalent to the universal Donsker property?

For classes of functions a quantitative condition characterizing the universal Donsker property up to measurability is not known (to me). Several nonequivalent extensions of the VC property to families of functions will be shown to be sufficient (under suitable measurability), but not necessary.

The remaining sections of the paper are: 1. Statements of conditions. 2. The easier implications. 3. Small non-VC hull classes and dual density. 4. Stability properties. 5. Metric entropy of convex hulls in Hilbert space. 6. Sharpness of the conditions. 7. Notes on weighted processes.

1. Statements of conditions.

Here is a first result. For any real function f let $\operatorname{diam}(f) := \sup f - \inf f$.

(1.1) PROPOSITION. *For any universal Donsker class $\mathscr{F}$,*

$$\sup_{f\in\mathscr{F}} \operatorname{diam}(f) < \infty.$$

PROOF. If not, take $x_n := x(n) \in X$, $y_n := y(n) \in X$ and $f_n \in \mathscr{F}$ with $f_n(x_n) - f_n(y_n) > 2^n$ for all n. Let

$$P := \sum_{n=1}^{\infty} \left(\delta_{x(n)} + \delta_{y(n)}\right)/2^{n+1},$$

a law on all subsets of X and thus on $\mathscr{A}$. Then for P,

$$E(f_n - Ef_n)^2 \geq 2^{-n-1} \inf_{c\in\mathbb{R}} \left\{(f_n(x_n) - c)^2 + (f_n(y_n) - c)^2\right\}$$
$$= 2^{-n-2}(f_n(x_n) - f_n(y_n))^2 > 2^{n-2}, \qquad n = 1,2,\dots.$$

Thus, for the ρ_P (standard deviation) metric, $\mathscr{F}$ is unbounded, so not a Donsker class for P [Dudley (1984), Theorem 4.1.1] and not a universal Donsker class. □

Say a function h on $\mathscr{F}$ *ignores additive constants* if $h(f) = h(f + c)$ whenever $f \in \mathscr{F}$, c is a constant and $f + c \in \mathscr{F}$. Let Y be a coherent G_P

process. Then if $f \in \mathscr{F}$ and $f + c \in \mathscr{F}$ for a constant c, $\rho_P(f, f + c) = 0$ so $Y(f) \equiv Y(f + c)$. Then Y is consistently extended to the set $\mathscr{F} + \mathbb{R}$ of all functions $f + c$, $f \in \mathscr{F}$, $c \in \mathbb{R}$, setting $Y(f + c) \equiv Y(f)$. This extension is coherent, so $\mathscr{F} + \mathbb{R}$ is pregaussian.

Now let $\mathscr{F}$ be a functional Donsker class for P. Note that each $k(P_k - P)$ is defined and ignores additive constants on $\mathscr{F} + \mathbb{R}$. Extending each Y_j in the definition to $\mathscr{F} + \mathbb{R}$ as above, we have $\|\alpha\|_{\mathscr{F}+\mathbb{R}} \equiv \|\alpha\|_{\mathscr{F}}$ for each $\alpha = k(P_k - P) - Y_1 - \cdots - Y_k$. Thus, $\mathscr{F} + \mathbb{R}$ is a functional Donsker class for P. Hence, if $\mathscr{F}$ is a universal Donsker class, so is $\mathscr{F} + \mathbb{R}$.

Any subset of a functional Donsker class for P is also, thus any subset of a universal Donsker class is a universal Donsker class. For an arbitrary real function $c(\cdot)$ on $\mathscr{F}$, $\mathscr{F}$ is a universal Donsker class iff $\{f - c(f): f \in \mathscr{F}\}$ is. Let $b\mathscr{F} := \{bf: f \in \mathscr{F}\}$ for a constant $b \neq 0$. Let $U_j(bf) := bY_j(f)$, $f \in \mathscr{F}$, for Y_j as in the definition of a functional Donsker class. Then each U_j is a coherent G_P process on $b\mathscr{F}$, so that replacing $\mathscr{F}$ by $b\mathscr{F}$ and Y_j by U_j we see that $b\mathscr{F}$ is a functional Donsker class iff $\mathscr{F}$ is, and thus a universal Donsker class iff $\mathscr{F}$ is. So in finding sufficient conditions for a class $\mathscr{F}$ to be a universal Donsker class, it will be enough to consider uniformly bounded classes of functions [letting $c(f) := \inf f$], where we can also assume $0 \leq f \leq 1$ for all $f \in \mathscr{F}$.

Now several conditions to be shown sufficient for the universal Donsker property will be defined.

DEFINITIONS. Let $\mathscr{F}$ be a class of functions on X with $f(x) \geq 0$ for all $f \in \mathscr{F}$ and $x \in X$. Let $\mathscr{C}$ be a class of subsets of X.

For each $f \in \mathscr{F}$ and $t \in \mathbb{R}$, the set $\{x \in X: f(x) > t\}$ will be called a *major set* of f and of $\mathscr{F}$. I call $\mathscr{F}$ a *major class* for $\mathscr{C}$ iff all the major sets of $\mathscr{F}$ are in $\mathscr{C}$. If $\mathscr{C}$ is a VC class, then I call $\mathscr{F}$ a *VC major class* (for $\mathscr{C}$).

The *subgraph* of the function $f \geq 0$ will be defined as

$$\mathrm{sub}(f) := \{\langle x, t\rangle \in X \times \mathbb{R}: 0 \leq t \leq f(x)\}.$$

For a class $\mathscr{D}$ of subsets of $X \times \mathbb{R}$, say $\mathscr{F}$ is a *subgraph class* for $\mathscr{D}$ iff $\mathrm{sub}(f) \in \mathscr{D}$ for all $f \in \mathscr{F}$. If $\mathscr{D}$ is a VC class, then I call $\mathscr{F}$ a *VC subgraph class* (for $\mathscr{D}$). [VC subgraph classes have previously been called "VC graph" classes [Alexander (1984, 1987)] or "Polynomial classes" [Pollard (1984), pages 17 and 34, and (1985), Section 6].]

Given a class $\mathscr{F}$ of functions and $0 < M < \infty$, let $H(\mathscr{F}, M)$ be M times the symmetric convex hull of $\mathscr{F}$, that is, the class of all functions g such that for some $k < \infty$, $f_j \in \mathscr{F}$ and real t_j, $j = 1, \ldots, k$, $\sum_{j=1}^k |t_j| \leq M$ and $g(x) = \sum_{j=1}^k t_j f_j(x)$ for all x. Then let $\overline{H}_s(\mathscr{F}, M)$ [sequential closure of $H(\mathscr{F}, M)$] be the smallest class $\mathscr{G}$ of functions including $H(\mathscr{F}, M)$ such that for any $g_n \in \mathscr{G}$ with $g_n(x) \to g(x)$ as $n \to \infty$ for all x, we have $g \in \mathscr{G}$. Say that $\mathscr{F}$ is a *hull class* for $\mathscr{C}$ (a class of sets identified with its class of indicator functions) iff $\mathscr{F} \subset \overline{H}_s(\mathscr{C}, M)$ for some $M < \infty$. Then $\mathscr{F}$ will be called a *VC hull class* if $S(\mathscr{C}) < \infty$. I call $\mathscr{F}$ a *VC subgraph hull class* if $\mathscr{F} \subset \overline{H}_s(\mathscr{G}, M)$ for some $M < \infty$ and VC subgraph class $\mathscr{G}$. [The "secondary VC graph difference classes," for which Alexander

(1985b), Theorem 2.2, proves extended Kiefer inequalities, are VC subgraph hull classes.]

For any metric space (S, d) and $\varepsilon > 0$ let $D(\varepsilon, S) := D(\varepsilon, S, d) := \sup\{m\colon$ for some $x_1, \dots, x_m \in S$, $d(x_i, x_j) > \varepsilon$ for $1 \le i < j \le m\}$. Given a law Q on $\mathscr{A}$, $1 \le p < \infty$ and $\varepsilon > 0$ let $D^{(p)}(\varepsilon, \mathscr{F}, Q) := D(\varepsilon, \mathscr{F}, e_{Q,p})$, where $e_{Q,p}(f, g) := (\int |f - g|^p \, dQ)^{1/p}$. Let $D^{(p)}(\varepsilon, \mathscr{F}) := \sup_Q D^{(p)}(\varepsilon, \mathscr{F}, Q)$, where the supremum is over all laws Q concentrated in finite sets. [These definitions are due to Kolčinskiĭ (1981) and Pollard (1982).] Say that $\mathscr{F}$ satisfies *Pollard's entropy condition* iff

$$(1.2) \qquad \int_0^1 \left(\log D^{(2)}(\varepsilon, \mathscr{F})\right)^{1/2} d\varepsilon < \infty.$$

Let $D^{(\infty)}(\varepsilon, \mathscr{F}) := D(\varepsilon, \mathscr{F}, d_\infty)$, where $d_\infty(f, g) := \sup_x |(f - g)(x)|$.

Then clearly $D^{(2)}(\varepsilon, \mathscr{F}) \le D^{(\infty)}(\varepsilon, \mathscr{F})$.

Note that for any uniformly bounded class $\mathscr{F}$ of measurable functions, $M < \infty$ and any law Q on $\mathscr{A}$, $H(\mathscr{F}, M)$ is dense in $\overline{H}_s(\mathscr{F}, M)$ for $e_{Q,p}$, $1 \le p < \infty$, as will be used for $p = 2$.

Some other conditions to be considered are as follows:

$$(1.3) \qquad \text{For some } r < \infty, \quad D^{(2)}(\varepsilon, \mathscr{F}) = O(\varepsilon^{-r}) \text{ as } \varepsilon \downarrow 0.$$

(1.4) $\mathscr{F}$ is a sequence $\{f_j\}_{j \ge 1}$ of measurable functions on X with $\operatorname{diam}(f_j) = o((\log j)^{-1/2})$ as $j \to \infty$.

$$(1.4)^{\mathrm{co}} \qquad \mathscr{F} \subset \overline{H}_s(\{f_j\}, M), \quad \text{where } M < \infty \text{ and } \{f_j\} \text{ satisfies (1.4).}$$

Condition (1.2) with sup instead of L^2 norm becomes

$$(1.5) \qquad \int_0^1 \left(\log D^{(\infty)}(x, \mathscr{F})\right)^{1/2} dx < \infty.$$

Most of the rest of the paper is devoted to showing what implications hold between the various conditions just defined, all of which imply the universal Donsker property. Figure 1 illustrates the implications, where h) requires some measurability conditions. Some implications which are obvious (given others) are not indicated. Most of the cases are handled by Theorem 2.1 below; "a)," "b)," etc. in Figure 1 indicate parts of Theorem 2.1. The condition

$$(1.6) \qquad \log D^{(2)}(\varepsilon, \mathscr{F}) = O(\varepsilon^{-2}), \quad \text{as } \varepsilon \downarrow 0$$

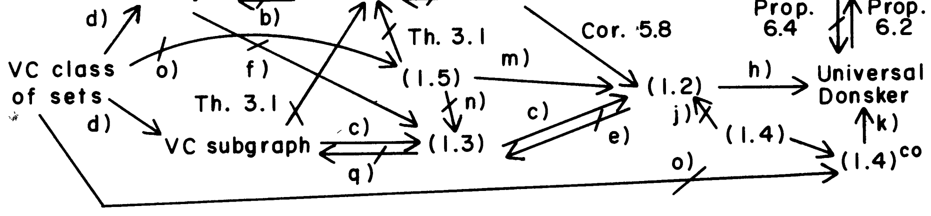

FIG. 1.

is shown in Section 6 to be a sharp (Proposition 6.3) necessary condition (Proposition 6.2) for the universal Donsker property but not sufficient (Proposition 6.4). These facts and Theorem 2.1h) show that $D^{(2)}$ comes close to characterizing the universal Donsker property, but does not. [On (1.4) see Theorem 2.1o).] Some of the stronger conditions are often easier to check and apply. Also, they may imply stronger inequalities or speeds of convergence, which may vary notably even between VC classes of sets, according to Beck (1985). A statistician has some freedom of choice in the class of functions.

2. The easier implications. This section will prove most of the implications and nonimplications in Figure 1 relatively easily.

(2.1) THEOREM. *Let $\mathscr{F}$ be a class of measurable functions f on X with $0 \le f(x) \le 1$ for all x, and $\mathscr{C}$ a class of measurable subsets of X.*

a) *If all the major sets of $\mathscr{F}$ are in $\mathscr{C}$, then $\mathscr{F}$ is also a hull class for $\mathscr{C}$. Thus, every VC major class $\mathscr{F}$ is also a VC hull class.*

b) *There exist VC hull classes $\mathscr{F}$ which are not VC major classes [i.e., $\{\{x: f(x) > t\}: f \in \mathscr{F}, t \in \mathbb{R}\}$ is not a VC class].*

c) *(Pollard) For every VC subgraph class $\mathscr{F}$,* (1.3) *and* (1.2) *hold.*

d) *For any VC class $\mathscr{C}$ of sets, its set of indicator functions is both a VC subgraph and a VC major class.*

e) *There exist classes $\mathscr{F}$ of functions satisfying* (1.2) *and not* (1.3).

f) *There exist VC major (thus VC hull) classes which do not satisfy* (1.3), *thus are not VC subgraph classes.*

g) *There are VC subgraph classes [thus satisfying* (1.3) *and* (1.2)] *which are not VC major.*

h) *(Pollard) If $\mathscr{F}$ satisfies* (1.2) *and is image admissible Suslin, it is a universal Donsker class.*

i) *If $\mathscr{F}$ is a functional Donsker class for P, then so is $\overline{H}_s(\mathscr{F}, M)$ for any $M < \infty$.*

j) *There exist sequences $\{f_j\}$ satisfying* (1.4) *but not Pollard's entropy condition* (1.2).

k) *(Paulauskas and Heinkel) If $\mathscr{F}$ satisfies* (1.4) *or* $(1.4)^{co}$, *then $\mathscr{F}$ is a universal Donsker class.*

m) *Condition* (1.5) *implies* (1.2).

n) (1.5) *does not imply* (1.3), *thus does not imply that $\mathscr{F}$ is a VC subgraph class.*

o) *An infinite collection of indicators of sets is never included in $\overline{H}_s(\mathscr{F}, M)$ for any $M < \infty$ and class $\mathscr{F}$ satisfying* (1.5) *or sequence $\mathscr{F} = \{f_j\}$ satisfying* (1.4).

p) (1.5) *does not imply that $\mathscr{F}$ is a VC major class.*

q) (1.3) *does not imply that $\mathscr{F}$ is a VC subgraph class.*

REMARK. As noted, if $\mathscr{F}$ is the class of indicators of sets in $\mathscr{C}$, then up to measurability conditions, the universal Donsker property of $\mathscr{F}$ is equivalent to the Vapnik–Červonenkis property of $\mathscr{C}$, and hence to the intermediate conditions in Figure 1, namely the VC major, VC hull and VC subgraph (hull) properties of $\mathscr{F}$ as well as (1.2) and (1.3) [that (1.2) implies (1.3) for classes of sets may be surprising]. Some of these equivalences can be seen directly (without measurability assumptions). Conditions (1.4) and (1.5) do not join in the equivalence.

PROOF. For the examples in parts e), j) and q), $A(n)$ will always be independent sets with $P(A(n)) = 1/2$, $n = 1, 2, \dots$, for some law P, specifically Lebesgue measure on $[0, 1]$.

a) Let $\mathscr{F}$ be a major class for $\mathscr{C}$. Given $f \in \mathscr{F}$ let

$$f_n := \frac{1}{n} \sum_{j=1}^{n} 1_{\{f > j/n\}} = \sum_{j=0}^{n-1} \frac{j}{n} 1_{\{j/n < f \le (j+1)/n\}}.$$

Then $f_n \to f$ as $n \to \infty$ (even uniformly), so $\mathscr{F}$ is a hull class for $\mathscr{C}$ [this argument was used previously in Dudley (1981), Theorem 1.9].

b) Let $X = \mathbb{R}^2$ with usual Borel σ-algebra. Let $\mathscr{C}$ be the collection of all open lower left quadrants

$$\{\langle x, y\rangle : x < a,\ y < b\}, \qquad a \in \mathbb{R},\ b \in \mathbb{R}.$$

Then $\mathscr{C}$ is a VC class [$S(\mathscr{C}) = 2$, see Wenocur and Dudley (1981), Proposition 2.3, and Dudley (1984), Corollary 9.2.15]. Let

$$\mathscr{F} := \left\{ \sum_{n=1}^{\infty} 1_{C(n)}/2^n : C(n) \in \mathscr{C} \right\}.$$

Then $\mathscr{F}$ is clearly a hull class for $\mathscr{C}$ (this did not depend on the particular choice of $\mathscr{C}$). The sets $\{f > 0\}$, $f \in \mathscr{F}$, are exactly the countable unions of sets in $\mathscr{C}$. These are all the open "lower layers" in the plane [e.g., Dudley (1984), Section 7.2] and do not form a VC class. For example, their intersections with the line $x + y = 1$ shatter any finite subset of the line.

c) See Pollard (1984), page 27, Lemma 25, with $F \equiv 1$, and page 34, Lemma 36.

d) $\mathscr{C}$ is a VC class in X iff $\{B \times [0, 1] : B \in \mathscr{C}\}$ is a VC class in $X \times \mathbb{R}$: "if" is straightforward, and "only if" is a special case of Assouad (1983), Proposition 2.5b, and Dudley (1984), Theorem 9.2.6, for $\times$. The class of major sets for $\{1_A : A \in \mathscr{C}\}$ is $\mathscr{C} \cup \{\varnothing, X\}$, a VC class iff $\mathscr{C}$ is.

e) and n) Let $Lx := \max(1, \log x)$ and $f_n := 1_{A(n)}/Ln$. Then the sequence $\mathscr{F} := \{f_n\}_{n \ge 1}$ satisfies (1.4), (1.5) and (1.2). For each subset F of $\{1, \dots, n\}$ choose a point

$$x(F) \in \bigcap_{j \in F} A(j) \cap \bigcap_{j \notin F,\ j \le n} X \setminus A(j).$$

Set $Q := 2^{-n} \sum_F \delta_{x(F)}$. Then $A(j)$ for $j = 1, \dots, n$ are independent for Q with probability 1/2. We have for $i < j \le n$,

$$\int (f_i - f_j)^2 \, dQ = \int f_i^2 + f_j^2 \, dQ - \frac{1}{2LiLj}$$
$$= \frac{1}{2}\left(\frac{1}{(Li)^2} + \frac{1}{(Lj)^2} - \frac{1}{LiLj}\right) \ge \frac{1}{2(Lj)^2} \ge \frac{1}{2(Ln)^2}.$$

So $D^{(2)}(\varepsilon, \mathscr{F}, Q) \ge \exp(1/2^{1/2}\varepsilon) - 1$ for $0 < \varepsilon < 1$ and (1.3) fails, proving n) and e).

f) Let $\mathscr{F}$ be the set of all nondecreasing right-continuous functions f on $\mathbb{R}$ with $0 \le f \le 1$. Then $\mathscr{F}$ is a VC major class, since half-lines $[x, \infty[$ or $]x, \infty[$, $-\infty \le x \le \infty$, form a VC class $\mathscr{C}$ with $S(\mathscr{C}) = 1$. To show that (1.3) fails for $\mathscr{F}$, first note that for any measurable f and g and law Q, the L^2 distance $e_Q(f, g) := e_{Q,2}(f, g) \ge \int |f - g| \, dQ$, so $D^{(2)}(\varepsilon, \mathscr{F}) \ge D^{(1)}(\varepsilon, \mathscr{F})$, $\varepsilon > 0$.

The next steps use some facts about lower layers [Dudley (1984), Section 7.2]. For Lebesgue measure P on $[0, 1]$,

$$D^{(1)}(\varepsilon, \mathscr{F}, P) = D(\varepsilon, \mathscr{L}\mathscr{L}_{2,1}, d_\lambda), \qquad 0 < \varepsilon < 1,$$

where $\mathscr{L}\mathscr{L}_{2,1}$ is the set of all lower layers in the unit square I^2 in $\mathbb{R}^2$ and d_λ is the Lebesgue measure of the symmetric difference of sets in I^2. Thus, for some $c > 0$, $D^{(1)}(\varepsilon, \mathscr{F}, P) \ge e^{c/\varepsilon}$ as $\varepsilon \downarrow 0$. For each ε with $0 < \varepsilon < 1$, there is a law Q with finite support and $D^{(1)}(\varepsilon, \mathscr{F}) \ge D^{(1)}(\varepsilon, \mathscr{F}, Q) \ge e^{c/\varepsilon} - 1$, so (1.3) fails, proving f).

g) and p) These will follow from Theorem 3.1, but here are short proofs. Let $f_n := n^{-1} + n^{-2} 1_{B(n)}$, $n = 1, 2, \dots$, where $B(n)$ is any sequence of measurable sets. Then $f_n \downarrow 0$, so the subgraphs of f_n also decrease. Being linearly ordered by inclusion, they form a VC class $\mathscr{C}$ with $S(\mathscr{C}) = 1$ [Wenocur and Dudley (1981), Corollary 2.2, and Dudley (1984), Theorem 9.2.4). So $\mathscr{F} := \{f_n\}_{n \ge 1}$ is a VC subgraph class. We have $\{f_n > 1/n\} = B(n)$, and the $B(n)$ need not form a VC class, so $\mathscr{F}$ need not be a VC major class, proving g). Now $\mathscr{F}$ has $D^{(\infty)}(\varepsilon, \mathscr{F}) \le 2 + 1/\varepsilon$ for $0 < \varepsilon < 1$, so $\mathscr{F}$ satisfies (1.5), proving p).

h) This essentially follows from a theorem of Pollard (1982), proved with the specific "image admissible Suslin" measurability condition in Dudley (1984), Section 11.3, with $F \equiv 1$.

i) For any subset $\mathscr{F}$ of a real vector space V and real function Y on $\mathscr{F}$, call Y *prelinear* iff whenever $\sum_{i=1}^m a_i f_i = 0$ for real a_i and f_i in $\mathscr{F}$, then $\sum_{i=1}^m a_i Y(f_i) = 0$.

(2.2) LEMMA. *For any coherent G_P process Y on some $\mathscr{F}$ there is such a process Z on the ρ_P-closed convex symmetric hull $\mathscr{H}$ of $\mathscr{F}$ with all sample functions of Z prelinear on $\mathscr{H}$ and a.s. $Z \equiv Y$ on $\mathscr{F}$. Thus, almost all sample functions of Y are prelinear on $\mathscr{F}$.*

PROOF. See Dudley (1985), Theorem 5.1, in whose proof the three-series theorem should and need only be applied when $f \in \mathscr{F}$. □

Let $\mathscr{F}$ be a uniformly bounded functional P-Donsker class and $\mathscr{H} := \{h$: for some $f_n \in \mathscr{F}$, $f_n(x) \to h(x)$ for all $x\}$. Each $k(P_k - P)$ is continuous for bounded pointwise convergence, and each coherent Y_j in the definition of functional P-Donsker class extends naturally to $\mathscr{H}$. For each $\alpha := k(P_k - P) - Y_1 - \cdots - Y_k$, we have $\|\alpha\|_{\mathscr{H}} \equiv \|\alpha\|_{\mathscr{F}}$. Thus, $\mathscr{H}$ is a functional P-Donsker class. Hence, so is $\mathscr{G} := \overline{H}_s(\mathscr{F}, M)$, we have $\|\alpha\|_{\mathscr{G}} = M\|\alpha\|_{\mathscr{F}}$, and i) follows.

j) An example will be based on one in the proof of Dudley (1967), Proposition 6.10. Let $a_n := \alpha_n(Ln)^{-1/2}$ where $\alpha_n \downarrow 0$ slowly; specifically, $\alpha_n := (LLn)^{-1/2}$. Let $f_n := a_n 1_{A(n)}$. Then for each n, taking Q as in the proof of e), for $1 \le i < j \le n$ we have

$$\int (f_i - f_j)^2 \, dQ = \tfrac{1}{2}\Big(\alpha_i^2 (Li)^{-1} + \alpha_j^2 (Lj)^{-1} - \alpha_i \alpha_j (LiLj)^{-1/2}\Big)$$

$$\ge \tfrac{1}{2}\alpha_j^2 (Lj)^{-1} > \varepsilon^2$$

for $j \le n$ if $\alpha_n^2 > 2\varepsilon^2 Ln$, or, equivalently, $LnLLn < 1/(2\varepsilon^2)$. Let $n(\varepsilon)$ be the largest n for which this holds. Then as $\varepsilon \downarrow 0$, $LLn(\varepsilon) \sim 2L(1/\varepsilon)$, so $Ln(\varepsilon) \sim (4\varepsilon^2 L(1/\varepsilon))^{-1}$, and $\int_0^1 (Ln(\varepsilon))^{1/2} \, d\varepsilon = +\infty$. So $\{f_j\}_{j \ge 1}$ satisfies (1.4) but not (1.2), proving j).

k) By part i), we may assume (1.4), then apply a central limit theorem in the Banach space c_0 stated by Paulauskas (1980), Corollary. Heinkel (1983), Theorem 1, gave a relatively long proof. Here is a direct proof.

By an inequality of Hoeffding (1963), (2.3) of Theorem 2, for any measurable f with $0 \le f \le 1$, $n = 1, 2, \dots$, and any law P,

$$\Pr\{\nu_n(f) \ge y\} \le \exp(-2y^2), \text{ for each } y \ge 0.$$

Considering $1 - f$, we have

$$\Pr\{|\nu_n(f)| \ge y\} \le 2\exp(-2y^2).$$

Let $a_j := \operatorname{diam}(f_j) = \alpha_j (Lj)^{-1/2}$, so that $\alpha_j \to 0$ as $j \to \infty$. We may assume that $0 \le f_j \le a_j > 0$. Then for any $\varepsilon > 0$,

$$\Pr\Big\{|\nu_n(f_j)| > \varepsilon\Big\} = \Pr\Big\{|\nu_n(f_j/a_j)| > \varepsilon/a_j\Big\} \le 2\exp\big(-2\varepsilon^2/a_j^2\big)$$

$$= 2\exp\big(-(2\varepsilon^2 Lj)/\alpha_j^2\big)$$

$$= 2j^{-2\varepsilon^2/\alpha(j)^2}, \text{ where } \alpha(j) := \alpha_j,\ j \ge 3.$$

Since $\varepsilon^2/\alpha_j^2 \ge 1$ for j large,

$$\sum_j \Pr\Big\{|\nu_n(f_j)| > \varepsilon\Big\} \text{ converges, uniformly in } n.$$

If $f_i = f_j$ a.s. for P, then $\nu_n(f_i) = \nu_n(f_j)$ a.s. for P^∞. Thus, $\{f_j\}_{j \ge 1}$ form a functional Donsker class [by the proof of Theorem 5.2.1 of Dudley (1984)], proving k).

m) follows easily from the definitions.

Recall that n) was proved with e) and p) with g).

For o), recall that the convex hull of a totally bounded set is totally bounded [by Mazur's theorem, e.g., Dunford and Schwartz (1958), page 416].

q) Take $\mathscr{F} := \{1_{A(n)}/n\}_{n\geq 1}$. Then for $0 < \varepsilon < 1$, $D^{(2)}\varepsilon, \mathscr{F}) \leq D^{(\infty)}(\varepsilon, \mathscr{F}) \leq 1 + 1/\varepsilon$, so (1.3) holds. For $m = 1, 2, \dots$, the sets $A(1), \dots, A(2^m)$ shatter some set B with m elements [Assouad (1983), 2.12 and 2.13b, and Dudley (1984), Theorem 9.3.2, which lacks the Assouad reference]. Then the class $\mathscr{D}$ of all subgraphs of functions in $\mathscr{F}$ shatters $\{\langle x, 2^{-m-1}\rangle: x \in B\}$, so $S(\mathscr{D}) = +\infty$, proving q) and Theorem 2.1. □

3. Small non-VC hull classes and dual density. This section will show that there are sequences converging to 0 in supremum norm like any power of n which are not VC hull classes. The proof applies the notion of dual density [Assouad (1983)].

(3.1) THEOREM. *There exist sequences* $\mathscr{F} := \{b_n 1_{A(n)}\}_{n\geq 1}$, *specifically with* $b_n = 1/n^v$ *for any positive integer* v, *which satisfy* (1.3), (1.4) *and* (1.5), *and there exist VC subgraph classes* $\mathscr{G}$, *such that* $\mathscr{F}$ *and* $\mathscr{G}$ *are not VC hull classes.*

PROOF. Two lemmas will be proved.

(3.2) LEMMA. *Let* $A_1, \dots, A_n$ *be jointly independent events in* X *with* $P(A_j) = 1/2$, $j = 1, \dots, n$. *Let* $B_1, \dots, B_n$ *be any events and set* $D := \bigcup_{1\leq j\leq n} A_j \triangle B_j$. *Then the algebra* $\mathscr{B}$ *generated by* $B_1, \dots, B_n$ *has at least* $2^n(1 - P(D))$ *atoms.*

PROOF. For any $F \subset \{1, \dots, n\}$, let

$$A_F := \bigcap_{j\in F} A_j \cap \bigcap_{j\notin F,\, j\leq n} X\setminus A_j,$$

and likewise define B_F. The atoms of $\mathscr{B}$ are those B_F which are nonempty. For each F, $P(A_F) = 1/2^n$. If a point of A_F is not in D, it is also in B_F, which is then nonempty. Since D can include at most $2^n P(D)$ of the 2^n events A_F, the lemma follows. □

(3.3) LEMMA. *Suppose* $A_j := A(j)$ *are independent events with* $P(A_j) = 1/2$ *for all* j, *and* $\mathscr{C}$ *is a class of events such that for some* $K < \infty$ *and* $u < \infty$, *for each* j *there is an event* D_j *such that* $P(A_j \triangle D_j) < \eta_j$, *where* D_j *is in an algebra generated by at most* Kj^u *elements of* $\mathscr{C}$ *and* $\sum_j \eta_j < 1$. *Then* $\mathscr{C}$ *is not a VC class.*

PROOF. By Lemma 3.2, for each $m = 1, 2, \dots$, the algebra $\mathscr{D}$ generated by $D_1, \dots, D_m$ has at least $2^m\alpha$ atoms, where $\alpha := 1 - \sum_j \eta_j > 0$. On the other hand, $\mathscr{D}$ is generated by at most

$$\sum_{1\leq j\leq m} Kj^u \leq K(m+1)^{u+1}/(u+1)$$

sets in $\mathscr{C}$. Now, for a VC class $\mathscr{C}$, according to Assouad (1983), 1.3d and 2.13e, there is some $t < \infty$ and a $C < \infty$ such that the number of atoms of the algebra generated by k elements of $\mathscr{C}$ is at most Ck^t. Then $2^m\alpha$ is bounded above by a polynomial in m [of degree $(u + 1)t$], a contradiction. □

Now to prove Theorem 3.1, let $A(m)$ and P be as in the proof of Theorem 2.1. Suppose the class $\mathscr{F} := \{1_{A(m)}/m^v\}_{m\geq 1}$ were a VC hull class for some $\mathscr{C}$. Clearly, $D^{(\infty)}(\varepsilon, \mathscr{F}) \leq m$ for the least m such that $1/m^v \leq \varepsilon$, so $m < \varepsilon^{-1/v} + 1$. Thus, (1.4) and (1.5) both hold for $\mathscr{F}$ as stated.

Suppose $P(A) = 1/2$ and that $a + b1_A \in \overline{H}_s(\mathscr{C}, M)$ for some $a \geq 0$, $b > 0$ and $0 < M < \infty$. Replacing a by a/M and b by b/M, we can assume that $M = 1$. If $S(\mathscr{C}) = r < \infty$, then for any $w > r$ there is a constant $C < \infty$ such that $D(\varepsilon, \mathscr{C}, d_P) < C\varepsilon^{-w}$ for $0 < \varepsilon < 1$ [Assouad (1983), Proposition 4.3, and Dudley (1984), Theorem 9.3.1], where $d_P(B, C) := P(B \triangle C)$. Take $w = r + 1$. Given $0 < \beta < 1$, take $C(j) \in \mathscr{C}$ and t_j with $\Sigma_j|t_j| \leq 1$ such that $P(|a + b1_A - \Sigma_j t_j 1_{C(j)}|) < \beta$. Choose $D_i := D(i) \in \mathscr{C}$, $i = 1, \ldots, m$, such that for each j, $P(C_j \triangle D_i) \leq \beta$ for some $i := i(j)$, with $m \leq C\beta^{-r-1}$. Then $P(|a + b1_A - f|) < 2\beta$, where $f := \Sigma_j t_j 1_{D(i(j))}$. Let $B := \{f > a + b/2\}$. Then B is in the algebra generated by $D_1, \ldots, D_m$, and $P(B \triangle A) < 4\beta/b$. We apply this to $A = A(k)$ for each k, with $a = a(k) = 0$, $b = b(k) = 1/k^v$ and $\beta = \beta_k := 1/(8k^{2+v})$. We obtain sets $B := B(k)$ and $m := m(k) \leq Nk^u$ for some $N < \infty$ and $u := (r + 1)(2 + v) < \infty$. To apply Lemma 3.3, let $\eta_k := 1/(2k^2)$, giving a contradiction if $\mathscr{C}$ is a VC class. We obtain a VC subgraph class $\mathscr{G}$ as in the proof of Theorem 2.1g), letting $a := a(k) := 1/k$ and $b := b(k) = 1/k^2$. So Theorem 3.1 is proved. □

NOTES. The counterexamples in the proofs of Theorem 2.1b), e), f), g), j), n), o), p), q) and Theorem 3.1, all are, or can be taken to be, image admissible Suslin.

It is not settled here whether classes satisfying (1.5) are necessarily VC subgraph hull classes.

4. Stability properties. Some of the properties treated above are preserved by some operations. If $\mathscr{F}$ is a major class for $\mathscr{C}$, then so is $\{cf\colon f \in \mathscr{F}, c > 0\}$, clearly. The property of being a subgraph class for some $\mathscr{D}$ is not necessarily preserved by taking constant multiples, but it is for many of the classes $\mathscr{D}$ in applications.

(4.1) PROPOSITION [Assouad (1983), Proposition 2.15]. *Let $\mathscr{C}$ be a VC class and let $\mathscr{E}$ be its closure for pointwise convergence of indicator functions. Then $\mathscr{E}$ is also a VC class, with $S(\mathscr{E}) = S(\mathscr{C})$.*

PROOF (for completeness). Let F be a finite set, $A \in \mathscr{E}$, and let $A(\alpha)$ be a net of sets in $\mathscr{C}$ with $1_{A(\alpha)}(x) \to 1_A(x)$ for all x. Then for some α, $A(\alpha) \cap F = A \cap F$, so $\mathscr{C}$ and $\mathscr{E}$ give the same intersections with finite sets and shattered finite sets. □

Sometimes, to preserve norms for empirical measures and/or for measurability, it may be useful to consider subsets of the closure, such as the sequential closure. Given $\mathscr{C}$, its *monotone derived class* will be the class $\mathscr{D}$ of all sets $D \subset X$ such that for some $C_n \in \mathscr{C}$, either $C_n \uparrow D$ or $C_n \downarrow D$. If $\mathscr{F}$ is a VC major class for $\mathscr{C}$, then the sets $\{x \in X\colon f(x) \geq t\}$, $f \in \mathscr{F}$, $t \geq 0$, are in $\mathscr{D}$, so they can also be assumed to be in $\mathscr{C}$ without increasing $S(\mathscr{C})$.

(4.2) PROPOSITION. *If $\mathscr{F}$ is a major class for $\mathscr{C}$, $\mathscr{H}$ is the set of all nondecreasing functions from $\mathbb{R}$ into $[0,1]$, and*

$$\mathscr{G} := \{h \circ f \colon f \in \mathscr{F},\ h \in \mathscr{H}\},$$

then $\mathscr{G}$ is a major class for the monotone derived class $\mathscr{D}$ of $\mathscr{C}$. Thus, if $\mathscr{F}$ is a VC major class, so is $\mathscr{G}$.

PROOF. Given $h \in \mathscr{H}$ and $t \geq 0$, let $s := \inf\{y\colon h(y) > t\}$. Then $h^{-1}(]t,\infty[) =]s,\infty[$ or possibly $[s,\infty[$ if h has a jump just to the left of s, $h(s^-) < h(s)$. So for $f \in \mathscr{F}$,

$$(h \circ f)^{-1}(]t,\infty[) = f^{-1}(]s,\infty[) \in \mathscr{C} \subset \mathscr{D}$$

or

$$f^{-1}([s,\infty[) \in \mathscr{D}. \qquad \square$$

For any VC class $\mathscr{C}$, if we set

$$\mathscr{E} := \mathscr{C} \cup \{X \setminus A\colon A \in \mathscr{C}\}, \tag{4.3}$$

then $\mathscr{E}$ is also a VC class [e.g., by Vapnik and Červonenkis (1971), Theorem 1]. Thus, in Proposition 4.2 if $\mathscr{H}$ is replaced by the set of all monotone functions from $\mathbb{R}$ into $[0,1]$, $\mathscr{F}$ is a major class for the monotone derived class of $\mathscr{E}$ and thus still a VC major class. Here is a further extension:

(4.4) PROPOSITION. *Given $0 < M < \infty$ let V_M be the set of all functions from $\mathbb{R}$ into itself with total variation $\leq M$. Let $\mathscr{F}$ be a major class for $\mathscr{C}$, a functional Donsker class for P. Then $\mathscr{J} := \{h \circ f\colon h \in V_M,\ f \in \mathscr{F}\}$ is a functional Donsker class for P.*

PROOF. By subtracting constants which do not matter, as noted after Proposition 1.1, we may consider just functions h in V_M with $h(x) \to 0$ as $x \to -\infty$. For each such h, we have $h = u - v$ for some nondecreasing functions u and v, each with total variation $\leq M$, also going to 0 at $-\infty$. Let $\mathscr{G}$ be the class of all functions $u \circ f$, where u is nondecreasing, $f \in \mathscr{F}$ and $0 \leq u \leq 1$. Then by Proposition 4.2, $\mathscr{G}$ is a major class for $\mathscr{D}$, the monotone derived class of $\mathscr{C}$. By Theorem 2.1a) and its proof, $\mathscr{G}$ is a hull class for $\mathscr{D}$, with $\mathscr{G} \subset \overline{H}_s(\mathscr{D},1)$. It follows that $\mathscr{J} \subset \overline{H}_s(\mathscr{D},2M) \subset \overline{H}_s(\mathscr{C},2M)$. Thus, Theorem 2.1i) gives the result. $\square$

If $\mathscr{F}$ is a finite-dimensional real vector space of measurable functions on X, then either it contains the constants or the vector space W spanned by $\mathscr{F}$ and constants has dimension one larger. The class of sets

$$\{\{x\colon f(x) > t\}\colon f \in \mathscr{F},\ t \in \mathbb{R}\} = \{\{x\colon g(x) > 0\}\colon g \in W\}$$

is a VC class $\mathscr{C}$ with index $S(\mathscr{C})$ equal to the dimension of W [Dudley (1978), Theorem 7.2].

Let $\mathscr{F}$ be a VC major class which is image admissible Suslin via $(Y, \mathscr{U}, T)$. For example, $\mathscr{F}$ may be any finite-dimensional real vector space of measurable functions on X. Let $\mathscr{C} := \{\{x: f(x) > t\}: f \in \mathscr{F}, t \in \mathbb{R}\}$. Then $\mathscr{C}$ is a VC class and is image admissible Suslin via $(Y \times \mathbb{R}, \mathscr{U} \times \mathscr{B}, T')$, where $\mathscr{B}$ is the Borel σ-algebra and $T'(y,t)(x) := 1_{]t,\infty[}(T(y)(x))$. Then for V_M in Proposition 4.4, the given class of compositions is a universal Donsker class. On some possible applications and related results see Pollard (1985), Section 6.

Next, classes of products of functions will be considered.

(4.5) PROPOSITION. *Let $\mathscr{F}$ and $\mathscr{G}$ be classes of real functions on X and $\mathscr{C}$ and $\mathscr{D}$ classes of subsets of X. Let*

$$\mathscr{C} \underline{\cap} \mathscr{D} := \{A \cap B: A \in \mathscr{C}, B \in \mathscr{D}\}.$$

If for some K and M, $\mathscr{F} \subset \overline{H}_s(\mathscr{C}, K)$ and $\mathscr{G} \subset \overline{H}_s(\mathscr{D}, M)$, then $\{fg: f \in \mathscr{F}, g \in \mathscr{G}\} \subset \overline{H}_s(\mathscr{C} \underline{\cap} \mathscr{D}, KM)$.

PROOF. Suppose $\Sigma_i |s_i| \le K$, $\Sigma_j |t_j| \le M$, $A(i) \in \mathscr{C}$ and $B(j) \in \mathscr{D}$. Then $(\Sigma_i s_i 1_{A(i)} \Sigma_j t_j 1_{B(j)}) \equiv \Sigma_{i,j} s_i t_j 1_{A(i) \cap B(j)}$ and $\Sigma_{i,j} |s_i t_j| \le KM$. Taking limits of sequences then gives the result. □

Recall that if $\mathscr{C}$ and $\mathscr{D}$ are VC classes, then so is $\mathscr{C} \underline{\cap} \mathscr{D}$ [Assouad (1983), Proposition 2.5b, and Dudley (1984), Theorem 9.2.6]. Thus, if $\mathscr{F}$ and $\mathscr{G}$ are VC hull classes, so is their set of products fg.

(4.6) PROPOSITION. *There exist VC major classes $\mathscr{F}$ and $\mathscr{G}$ such that $\{\langle x, y \rangle \mapsto f(x)g(y): f \in \mathscr{F}, g \in \mathscr{G}\}$ is not a VC major class. Both $\mathscr{F}$ and $\mathscr{G}$ can be taken as the set of increasing functions on $\mathbb{R}$.*

PROOF. Let g be a fixed positive, continuous, strictly increasing function on $\mathbb{R}$, say $g(y) := e^y$. Let h be any continuous, strictly decreasing function. Let $f(x) := 1/g(h(x))$. Then f is positive, continuous and strictly increasing. We have $f(x)g(y) = 1$ on the graph $y = h(x)$. Thus, $f(x)g(y) < 1$ for $y < h(x)$ and $f(x)g(y) > 1$ iff $y > h(x)$. But the class of all sets where $y > h(x)$ for h continuous and strictly decreasing is not a VC class; it shatters all finite subsets of the line $x + y = 0$, for example. □

Thus, the VC hull property, with more stability, and which follows from the VC major property, may be more useful. Inequalities for suprema over VC subgraph hull classes, as good as those for VC classes $\mathscr{C}$ of sets [Alexander (1984)] extend immediately to VC subgraph hull classes, as near the end of the proof of Theorem 2.1i). In this sense VC subgraph hull classes are "just as good" as VC classes of sets despite the failure of several converse implications in Figure 1.

5. Metric entropy of convex hulls in Hilbert space. Let H be a real Hilbert space. For any subset B of H let $\mathrm{co}(B)$ be its convex hull.

(5.1) THEOREM. *Suppose $\|x\| \le 1$ for all $x \in B$ and for some $K < \infty$ and $\lambda < \infty$, $D(\varepsilon, B) \le K\varepsilon^{-\lambda}$ for $0 < \varepsilon \le 1$. Let $s := 2\lambda/(2 + \lambda)$. Then for any $t > s$*

there are constants C_1 and C_2, depending only on K, λ and t, such that

$$D(\varepsilon, \mathrm{co}(B)) \le C_1 \exp(C_2 \varepsilon^{-t}), \qquad 0 < \varepsilon \le 1.$$

PROOF. We may assume $K \ge 1$. Choose any $x_1 \in B$. Given $B(n) := \{x_1, \dots, x_{n-1}\}$, $n \ge 2$, let $d(x, B(n)) := \inf_{y \in B(n)} \|x - y\|$ and $\delta_n := \sup_{x \in B} d(x, B(n))$. If $\delta_n > 0$ choose $x_n \in B$ with $d(x_n, B(n)) > \delta_n/2$. (The estimates to be proved in general will also hold in case B is finite, so that $\delta_n = 0$ for some n.) Then for all n, $K(2/\delta_n)^\lambda \ge D(\delta_n/2, B) \ge n$, so $\delta_n \le Mn^{-1/\lambda}$ for all n where $M := 2K^{1/\lambda}$.

Let $0 < \varepsilon \le 1$. Let $N := N(\varepsilon)$ be the next integer larger than $(4M/\varepsilon)^\lambda$. Then $\delta_N < \varepsilon/4$. Let $G := B(N)$. For each $x \in B$ there is an $i = i(x) \le N$ with $\|x - x_i\| \le \delta_N$. For any convex combination $z = \sum_{x \in B} z_x x$, where $z_x \ge 0$ and $\sum_{x \in B} z_x = 1$, let $z_N := \sum_{x \in B} z_x x_{i(x)}$. Then $\|z - z_N\| \le \delta_N < \varepsilon/4$, so

$$D(\varepsilon, \mathrm{co}(B)) \le D(\varepsilon/2, \mathrm{co}(G)). \tag{5.2}$$

To bound the latter, let $m := m(\varepsilon) := [\varepsilon^{-s}]$ (integer part). Then for each i with $m < i \le N$, there is a $j := j(i) \le m$ such that

$$\|x_i - x_j\| \le \delta_{m+1} \le M\varepsilon^{s/\lambda}. \tag{5.3}$$

Let $\Lambda_m := \{\{\lambda_j\}_{1 \le j \le m}: \lambda_j \ge 0, \sum_{1 \le j \le m} \lambda_j = 1\}$. On $\mathbb{R}^m$ we have the l_p metrics

$$\rho_p(\{x_j\}, \{y_j\}) := \Bigl(\sum_{1 \le j \le m} |x_j - y_j|^p\Bigr)^{1/p}.$$

Let $\gamma := \varepsilon/6$ and $\delta := \gamma/(2m^{1/2})$. The δ-neighborhood of Λ_m for ρ_2 is included in a ball of radius $1 + \delta < 13/12$. We have $D(2\delta, \Lambda_m, \rho_2)$ centers of disjoint balls of radius δ included in the neighborhood. Comparing volumes of balls gives

$$\begin{aligned} D(\gamma, \Lambda_m, \rho_1) &\le D(2\delta, \Lambda_m, \rho_2) \le 13^m m^{m/2} \varepsilon^{-m} \\ &\le \exp(m\{L(1/\varepsilon) + (Lm)/2 + \log(13)\}) \\ &\le \exp(C_3 \varepsilon^{-s} L(1/\varepsilon)) \le \exp(C_4 \varepsilon^{-t}), \qquad 0 < \varepsilon \le 1, \end{aligned}$$

for some constants C_3, C_4.

For each $j = 1, \dots, m$, let $A(j)$ consist of x_j and the set of all x_i, $i = m+1, \dots, N$, such that $j(i) = j$. Take a maximal set $S = S(\varepsilon) \subset \Lambda_m$ with $\rho_1(x, y) > \gamma$ for any $x \ne y$ in S. For a given $\lambda = (\lambda_1, \dots, \lambda_m) \in S$, let

$$F_\lambda := \Bigl\{x \in \mathrm{co}(G): x = \sum_{y \in G} \mu_y y, \text{ where } \mu_y \ge 0 \text{ and } \sum_{y \in A(j)} \mu_y = \lambda_j, \text{ for all } j = 1, \dots, m\Bigr\}.$$

For any $x \in \mathrm{co}(G)$, and μ_y as in the definition of F_λ, there is a $\lambda \in S$ and $z \in F_\lambda$ with $\|z - x\| \le \gamma$. Thus,

$$\begin{aligned} D(\varepsilon, \mathrm{co}(B)) &\le D(\varepsilon/2, \mathrm{co}(G)) \le D\Bigl(\gamma, \bigcup_{\lambda \in S} F_\lambda\Bigr) \\ &\le \mathrm{card}(S) \max_\lambda D(\gamma, F_\lambda) \le \exp(C_4 \varepsilon^{-t}) \max_\lambda D(\gamma, F_\lambda). \end{aligned} \tag{5.4}$$

To estimate the latter factor, let $\lambda \in S$. We may assume $\lambda_j > 0$ for all j. For any $x \in F_\lambda$, let $x^{(j)} := \sum_{y \in A(j)} \mu_y y$. Let Y_j be a random variable with values in $A(j)$ and $P(Y_j = y) = \mu_y / \lambda_j$ for each $y \in A(j)$. Then $EY_j = x^{(j)}/\lambda_j := z_j$. Take $Y_1, \dots, Y_m$ to be independent and let $Y := \sum_{1 \le j \le m} \lambda_j Y_j$. Then $EY = x$ and

$$E\|Y - x\|^2 = E\left\| \sum_{1 \le j \le m} \lambda_j (Y_j - z_j) \right\|^2 = \sum_{1 \le j \le m} \lambda_j^2 E\|Y_j - z_j\|^2,$$

since $Y_j - z_j$ are independent and have mean 0.

Now the diameter of $A(j)$ is at most $2M\varepsilon^{s/\lambda}$ by (5.3), and z_j is a convex combination of elements of $A(j)$. Thus,

$$E\|Y_j - z_j\|^2 \le 4M^2 \varepsilon^{2s/\lambda}$$

and for any set $F \subset \{1, \dots, m\}$,

$$E\left\| \sum_{j \in F} \lambda_j (Y_j - z_j) \right\|^2 \le \sum_{j \in F} \lambda_j^2 4M^2 \varepsilon^{2s/\lambda} \le 4M^2 \left(\max_{j \in F} \lambda_j \right) \varepsilon^{2s/\lambda}$$

and

$$E\left\| \sum_{j \in F} \lambda_j (Y_j - z_j) \right\| \le 2M \left(\max_{j \in F} \lambda_j \right)^{1/2} \varepsilon^{s/\lambda}.$$

The following argument is based on an idea of Maurey, see the proofs of Pisier (1981), Lemma 2, and Carl (1982), Lemma 1. Let $Y_{j1}, Y_{j2}, \dots, Y_{jk}$ be independent with the distribution of Y_j, with Y_{ji} also independent for different j. Then

$$E\left\| \sum_{j \in F} \lambda_j \left(\left(\frac{1}{k} \sum_{i=1}^{k} Y_{ji} \right) - z_j \right) \right\| \le 2M \left(\max_{j \in F} \lambda_j \right)^{1/2} \varepsilon^{s/\lambda} / k^{1/2}.$$

Thus, there exist $y_{ji} \in A(j)$, $i = 1, \dots, k$, such that

$$\left\| \sum_{j \in F} \lambda_j \left(\left(\frac{1}{k} \sum_{i=1}^{k} y_{ji} \right) - z_j \right) \right\| \le 2M \left(\max_{j \in F} \lambda_j \right)^{1/2} \varepsilon^{s/\lambda} / k^{1/2}. \tag{5.5}$$

Take $v > 0$ such that $s + v < t$. Let $F(0) := \{j \le m: \lambda_j \ge \varepsilon^v\}$. Let $k(0)$ be the smallest integer k such that

$$k \ge 6400 M^2 \varepsilon^{-2 + 2s/\lambda} = 6400 M^2 \varepsilon^{-s}.$$

Then the expressions in (5.5) are at most $\varepsilon/40$.

Let r be the smallest positive integer such that $\varepsilon^v/4^r \le (\varepsilon^{1-s/\lambda}/(80M))^2$. For $u = 1, 2, \dots, r$, let

$$F(u) := \left\{ j \le m: \varepsilon^v/4^u \le \lambda_j < \varepsilon^v/4^{u-1} \right\},$$

and let $k = k(u)$ be the smallest integer k such that $2^{2-u} M \varepsilon^{s/\lambda}/k^{1/2} < \varepsilon/(40r)$, i.e., $k > 100 M^2 4^{4-u} r^2 \varepsilon^{-2+2s/\lambda}$. Thus, for some constant C_5, $k(u) \le 1 + C_5 4^{-u} \varepsilon^{-s} (L(1/\varepsilon))^2$. The y_{ji} for $F = F(u)$ will be called $y_{ji}^{(u)}$ (they depend on x).

Let $F(r+1) := \{j \le m\colon \lambda_j < \varepsilon^v/4^r\}$. Let $k(r+1) = 1$. For $F = F(r+1)$ and $k = k(r+1)$, (5.5) is bounded above by $\varepsilon/40$. Let $y_{j1}^{(r+1)} := x_j$, a single choice for each j.

Combining terms for $u = 0, 1, \dots, r+1$, we see that each $x \in F_\lambda$ can be approximated within $3\varepsilon/40 < \varepsilon/12$ by a convex combination determined uniquely by the $k(u)$-tuples $(y_{j1}^{(u)}, \dots, y_{jk(u)}^{(u)})$, $j \in F(u)$, $u = 0, 1, \dots, r+1$. Each $A(j)$ has at most N elements, so that for given $u \le r$ and $j \le m$, there are at most $N^{k(u)}$ ways of choosing the $y_{ji}^{(u)}$. Now $\operatorname{card}(F(u)) \le 4^u/\varepsilon^v$, so the number of ways to choose the $y_{ji}^{(u)}$ for given $u \le r$ is at most $\exp\{(\log N)(4^u\varepsilon^{-v} + C_5\varepsilon^{-s-v}L(1/\varepsilon)^2)\}$. Thus, the total number of ways to choose all the $y_{ji}^{(u)}$ gives $D(\varepsilon/6, F_\lambda) \le \exp(C_6\{\varepsilon^{-s-v}L(1/\varepsilon)^4 + L(1/\varepsilon)4^r/\varepsilon^v\})$ for some C_6. By definition of r, $4^r/\varepsilon^v \le C_7\varepsilon^{2s/\lambda-2} = C_7\varepsilon^{-s}$ for some C_7. Thus, $D(\varepsilon/6, F_\lambda) \le \exp(C_8\varepsilon^{-t})$ for some C_8. Combining with (5.4) completes the proof of Theorem 5.1. □

(5.6) Remark. The exponent $s = 2\lambda/(2+\lambda)$ in Theorem 5.1 is sharp, as shown by Dudley (1967), Proposition 6.12, where $B = \{\pm n^{-1/\lambda}e_n\}_{n\ge1}$ and $\{e_n\}$ is an orthonormal basis. More generally, Carl (1982) treats Banach spaces of type 2 in place of Hilbert space.

(5.7) Corollary. *If $\mathscr{G}$ is a uniformly bounded class of measurable functions such that for some $K < \infty$ and $0 < \lambda < \infty$, $D^{(2)}(\varepsilon, \mathscr{G}) \le K\varepsilon^{-\lambda}$ for $0 < \varepsilon \le 1$, then for any $t > s = 2\lambda/(2+\lambda)$, and for the constants C_1 and C_2 of Theorem* 5.1,

$$D^{(2)}\big(\varepsilon, \overline{H}_s(\mathscr{G},1)\big) \le C_1\exp\big(C_2\varepsilon^{-t}\big), \quad \textit{for } 0 < \varepsilon \le 1.$$

(5.8) Corollary. *If $\mathscr{F}$ is a VC subgraph hull class, then $\mathscr{F}$ satisfies Pollard's entropy condition* (1.2).

Proof. Let $\mathscr{F} \subset \overline{H}_s(\mathscr{G}, M)$ for $M < \infty$, where $\mathscr{G}$ is a VC subgraph class. Theorem 2.1c) gives condition (1.3) for $\mathscr{G}$, so Corollary 5.7 applies, with $s < t < 2$, and (1.2) for $\mathscr{F}$ follows. □

(5.9) Example. Let $\mathscr{C}$ be the collection of all intervals $]a, b]$, $0 \le a \le b \le 1$. Let G be the space of functions f satisfying $|f(x)| \le 1/2$ and $|f(x) - f(y)| \le |x - y|$ for all $x, y \in]0,1]$, $f(x) = 0$ elsewhere. Functions in G have total variation at most 2 (at most 1 on the open interval $]0,1[$ and $1/2$ at each end). As in the proof of Proposition 4.4, each such $f = g - h$, where g and h are each nondecreasing and 0 for $x \le 0$ and $df = dg - dh$ is the Jordan decomposition of the signed measure df. Thus, g and h have equal total variations ≤ 1 and $G \subset \overline{H}_s(\mathscr{C}, 2)$. Next, $D^{(2)}(\varepsilon, G) \ge \exp(c/\varepsilon)$ as $\varepsilon \downarrow 0$ for some constant $c > 0$, by considering Lebesgue measure P on $[0,1]$ (or finite approximations to it), noting that the $L^2(P)$ norm is larger than the $L^1(P)$ norm, and applying Dudley (1984), Theorem 7.1.10. Here since $S(\mathscr{C}) = 2$, the exponent λ can be any number larger than 2 [Assouad (1983), Proposition 4.3], so for $s = 2\lambda/(2+\lambda)$, $\lambda \downarrow 2$, t can be any number larger than 1. So the exponent s of Theorem 5.1 is also sharp for this example.

6. Sharpness of the conditions. How sharp are the various sufficient conditions for the universal Donsker property in Figure 1? First note that all but (1.4), $(1.4)^{\mathrm{co}}$ and (1.5) hold for every universal Donsker class of indicators of sets, since such classes of sets are VC [Durst and Dudley (1981) and exposition in Dudley (1984), Theorem 11.4.1]. So we have sharpness when restricted to classes of indicators. Thus, we have:

(6.1) REMARK. Let $\mathscr{C}$ be a class of measurable sets. If either every uniformly bounded major class for $\mathscr{C}$, or the convex hull of the indicators of sets in $\mathscr{C}$, is a universal Donsker class, then $\mathscr{C}$ is a VC class.

In this sense "VC major (with enough measurability) implies universal Donsker" is sharp in that "VC" cannot be replaced by any weaker condition on families of sets. On the other hand, "VC major" is one of the strongest conditions in Figure 1. It turns out that $D^{(2)}$ entropy conditions, much like and because of the L^2 entropy conditions for Gaussian processes, come close to characterizing, but do not characterize, the universal Donsker property. Note that by Pollard's theorem (2.1h) above), for any $\delta > 0$,

$$\log D^{(2)}(\varepsilon, \mathscr{F}) = O\big(1/\varepsilon^{2-\delta}\big), \quad \text{as } \varepsilon \downarrow 0$$

is (with measurability) sufficient for $\mathscr{F}$ to be a universal Donsker class.

(6.2) PROPOSITION. *$\log D^{(2)}(\varepsilon, \mathscr{F}) = O(\varepsilon^{-2})$ as $\varepsilon \downarrow 0$ is necessary for $\mathscr{F}$ to be a universal Donsker class.*

PROOF. If not, there exists a universal Donsker class $\mathscr{F}$ and $\varepsilon_k \downarrow 0$ such that $\log D^{(2)}(\varepsilon_k, \mathscr{F}) > k^3/\varepsilon_k^2$ for all $k = 1, 2, \dots$. Take probability laws P_k such that

$$\log D^{(2)}(\varepsilon_k, \mathscr{F}, P_k) > k^3/\varepsilon_k^2, \qquad k = 2, 3, \dots,$$

and let P be a law with $P \geq \sum_{k \geq 2} P_k/k^2$. Then for any measurable f and g,

$$\left(\int (f-g)^2\, dP\right)^{1/2} \geq \left(\int (f-g)^2\, dP_k\right)^{1/2} \Big/ k.$$

Let $\delta_k := \varepsilon_k/k$. Then

$$\log D^{(2)}(\delta_k, \mathscr{F}, P) \geq \log D^{(2)}(\varepsilon_k, \mathscr{F}, P_k) > k^3/\varepsilon_k^2 = k/\delta_k^2.$$

Thus, $\mathscr{F}$ is not pregaussian for P [Dudley (1973), Theorem 1.1 (c)], so not a universal Donsker class. □

An example will show that Proposition 6.2 is sharp. Let $A(j)$ be disjoint, nonempty measurable sets, $j = 1, 2, \dots$. For $x = \{x_j\}_{j \geq 1}$ let $\|x\|_2 := (\sum_j x_j^2)^{1/2}$. Let

$$\mathscr{E} := \mathscr{E}\big(\{1_{A(j)}\}_{j \geq 1}\big) := \Big\{\sum_j x_j 1_{A(j)} \colon \|x\|_2 \leq 1\Big\},$$

the ellipsoid with center 0 and semiaxes $1_{A(j)}$.

(6.3) PROPOSITION. *$\mathscr{E}$ is a universal Donsker class and*

$$\liminf_{\delta \downarrow 0} \delta^2 \log D(\delta, \mathscr{E}) > 0.$$

PROOF. Let P be any probability law with $p_j := P(A(j))$, $j = 1, 2, \ldots$. We may assume that $p_j > 0$ for all j, since the union of all those $A(j)$ with $p_j = 0$ is a fixed set of probability 0 on which G_P and ν_n can be taken to be identically 0 a.s.

Let $\varepsilon > 0$. For any k and n, let

$$\|\nu_n\|_{2,k} := \left(\sum_{j \ge k} \nu_n(A(j))^2 \right)^{1/2}.$$

Then for all n,

$$E\|\nu_n\|_{2,k}^2 \le \sum_{j \ge k} p_j \to 0 \quad \text{as } k \to \infty.$$

Take $k = k(\varepsilon)$ large enough so that

$$\Pr\{\|\nu_n\|_{2,k} > \varepsilon/3\} < \varepsilon/2.$$

If $\|\nu_n\|_{2,k} \le \varepsilon/3$, $\|x\|_2 \le 1$ and $\|y\|_2 \le 1$, then

$$\left| \nu_n\left(\sum_{j \ge k} (x_j - y_j) 1_{A(j)} \right) \right| \le 2\varepsilon/3.$$

Also, $\Pr\{\|\nu_n\|_{2,1} > 2/\varepsilon\} < \varepsilon/2$.

Let $\delta := \delta(\varepsilon) := (\min_{j<k} p_j)\varepsilon^2/6$. For each $x = \{x_j\}_{j \ge 1}$ let $f_x := \sum_j x_j 1_{A(j)}$. Then if $f_x, f_y \in \mathscr{E}$ and $(\int (f_x - f_y)^2 \, dP)^{1/2} < \delta$, $(\sum_{j<k} (x_j - y_j)^2)^{1/2} < \varepsilon^2/6$, so

$$\left| \sum_{j<k} \nu_n\big((x_j - y_j) 1_{A(j)}\big) \right| \le \|\nu_n\|_{2,1} \varepsilon^2/6 < \varepsilon/3,$$

if $\|\nu_n\|_{2,1} \le 2/\varepsilon$. Then except on one event with probability ε, we have $|\nu_n(f_x - f_y)| < \varepsilon$ for all such x and y.

Clearly, $\mathscr{E}$ is totally bounded in $L^2(P)$. Thus, $\mathscr{E}$ is a P-Donsker class [Dudley (1984), Theorem 4.1.10] and since P was arbitrary, a universal Donsker class.

Now let $0 < \delta < 1$. To bound $D^{(2)}(\delta, \mathscr{E})$ from below, let $m = [1/(4\delta^2)]$ (integer part). Let $P = P^{(m)}$ be a law with $P(A(j)) = 1/m$ for $j = 1, \ldots, m$. Then in $L^2(P)$, $\mathscr{E}$ is an m-dimensional ball of radius $m^{-1/2}$. The balls of radius 2δ with centers at a maximal set of points δ apart in $\mathscr{E}$ cover the ball of radius $m^{-1/2} + \delta$. Thus,

$$D^{(2)}(\delta, \mathscr{E}, P) \ge (m^{-1/2} + \delta)^m/(2\delta)^m = \left(\frac{1}{2} \left(\frac{1}{\delta m^{1/2}} + 1 \right) \right)^m$$

$$\ge \left(\tfrac{3}{2}\right)^m \ge \left(\tfrac{2}{3}\right)\exp\left(\log\left(\tfrac{3}{2}\right)/4\delta^2\right),$$

and the result follows. □

(6.4) PROPOSITION. *There is a uniformly bounded class $\mathscr{F}$ of measurable functions, which is not a universal Donsker class, such that*

$$\log D^{(2)}(\varepsilon, \mathscr{F}) \le 3/\big(\varepsilon^2 L(1/\varepsilon)\big), \quad \textit{as } \varepsilon \downarrow 0.$$

PROOF. Let $B(j)$, $j = 1, 2, \ldots$, be disjoint, nonempty measurable sets, $a_j := 1/(jLj)^{1/2}$ and $\mathscr{F} := \{\sum_{j \ge 1} x_j 1_{B(j)}: \ x_j = \pm a_j \text{ for all } j\}$. Take c such that

$\sum_j p_j = 1$ where $p_j = c(a_j/LLj)^2$. Then $\sum_j a_j p_j^{1/2} = +\infty$. Take a probability measure P with $P(B(j)) = p_j$ for all j. Then

$$E \sup_{f \in \mathscr{F}} |G_P(f)| = E \sum_j a_j \left| G_P(1_{B(j)}) \right|$$

$$= \sum_j a_j (2/\pi)^{1/2} \left(p_j(1 - p_j) \right)^{1/2} \geq \alpha \sum_j a_j p_j^{1/2} = +\infty,$$

where $\alpha = (1 - p_1)^{1/2}/2 > 0$. Thus, by the Landau–Shepp–Fernique theorem [Fernique (1970)], $\mathscr{F}$ is not pregaussian for P, hence not a universal Donsker class.

For any probability law Q, $r < \infty$, and $x_j = \pm a_j$,

$$\left(\int \left(\sum_{j \geq r} x_j 1_{B(j)} \right)^2 dQ \right)^{1/2} = \left(\sum_{j \geq r} a_j^2 Q(B(j)) \right)^{1/2} \leq a_r.$$

Given $\varepsilon > 0$, take the smallest $r = r(\varepsilon)$ for which $a_r < \varepsilon/2$. Then $D^{(2)}(\varepsilon, \mathscr{F}) \leq 2^r$ so $\log D^{(2)}(\varepsilon, \mathscr{F}) \leq r \log 2$. As $\varepsilon \downarrow 0$, we have $a_r \sim \varepsilon/2$, $\log(1/\varepsilon) \sim \log(2/\varepsilon) \sim -\log a_r \sim \log(r)/2$ and $\varepsilon/2 \sim 1/(r(\varepsilon)(2L(1/\varepsilon)))^{1/2}$, so $r(\varepsilon) \sim 2/(\varepsilon^2 L(1/\varepsilon))$. □

Propositions 6.3 and 6.4 together show that there is no characterization of the universal Donsker property in terms of $D^{(2)}$. Pollard's integral condition (1.2) is not sharp in the example of Proposition 6.4. [*Added in proof*: M. Talagrand has shown that (1.2) is sharp.] On this point see also Giné and Zinn (1984), Section 5.

7. Notes on weighted processes. In Dudley (1985), for Theorem 6.3 I gave credit to "O'Reilly, Chibisov et al." The "et al." should include Csörgő, Csörgő and Horváth (1986) and Csörgő, Csörgő, Horváth and Mason (1986), who proved parts e), f) and g) of that theorem, also removed the continuity assumption from the weight function, and gave an example like Example 7.2 of my paper. So I claim no priority (rather than "somewhat... perhaps") for these matters. I am very grateful to Miklos and Sandor Csörgő for sending me their and co-authors' reports (in 1983) and for pointing out their specific results. See also Alexander (1985a, b, 1987).

Acknowledgment. I thank the referee for useful comments.

REFERENCES

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

ALEXANDER, K. S. (1985a). The central limit theorem for weighted empirical processes indexed by sets. *J. Multivariate Anal.* To appear.

ALEXANDER, K. S. (1985b). Rates of growth for weighted empirical processes. In *Proc. of the Conference Berkeley in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) **2** 475–493. Wadsworth, Monterey, Calif.

ALEXANDER, K. S. (1987). The central limit theorem for empirical processes on Vapnik–Červonenkis classes. *Ann. Probab.* **15** 178–203.

ASSOUAD, P. (1983). Densité et dimension. *Ann. Inst. Fourier (Grenoble)* **33** (3) 233–282.

BECK, J. (1985). Lower bounds on the approximation of the multivariate empirical process. *Z. Wahrsch. verw. Gebiete* **70** 289–306.

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees.* Wadsworth, Belmont, Calif.

CARL, B. (1982). On a characterization of operators from l_q into a Banach space of type p with some applications to eigenvalue problems. *J. Funct. Anal.* **48** 394–407.

CHERNOFF, H. and LEHMANN, E. L. (1954). The use of maximum likelihood estimates in χ^2 tests for goodness of fit. *Ann. Math. Statist.* **25** 579–586.

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

CSÖRGŐ, M., CSÖRGŐ, S., HORVÁTH, L. and MASON, D. (1986). Weighted empirical and quantile processes. *Ann. Probab.* **14** 31–85.

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

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

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

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

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

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

DUNFORD, N. and SCHWARTZ, J. T. (1958). *Linear Operators, Part I: General Theory.* Wiley, New York.

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

FERNIQUE, X. (1970). Intégrabilité des vecteurs gaussiens. *C. R. Acad. Sci. Paris Sér. A* **270** 1698–1699.

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

HEINKEL, B. (1983). Majorizing measures and limit theorems for c_0-valued random variables. In *Probability in Banach Spaces IV* (A. Beck and K. Jacobs, eds.). *Lecture Notes in Math.* **990** 136–149. Springer, New York.

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

HOFFMANN-JØRGENSEN, J. (1984). Personal communication.

KOLČINSKIĬ, V. I. (1981). On the central limit theorem for empirical measures. *Theory Probab. Math. Statist.* **24** 71–82.

LE CAM, L., MAHAN, C. and SINGH, A. (1983). An extension of a theorem of H. Chernoff and E. L. Lehmann. In *Recent Advances in Statistics: Papers in Honor of Herman Chernoff* (M. H. Rizvi, J. S. Rustagi and D. Siegmund, eds.) 303–337. Academic, New York.

MOORE, D. S. and STUBBLEBINE, J. B. (1981). Chi-square tests for multivariate normality with application to common stock prices. *Comm. Statist. A—Theory Methods* **10** 713–738.

PAULAUSKAS, V. (1980). On the central limit theorem in the Banach space c_0. *Dokl. Akad. Nauk SSSR* **254** 286–288.

PISIER, G. (1981). Remarques sur un résultat non publié de B. Maurey. *Séminaire d'Analyse Fonctionelle 1980–1981* V.1–V.12. Ecole Polytechnique, Centre de Mathématiques, Palaiseau.

POLLARD, D. (1979). General chi-square goodness-of-fit tests with data-dependent cells. *Z. Wahrsch. verw. Gebiete* **50** 317–331.

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

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

POLLARD, D. (1985). New ways to prove central limit theorems. Preprint.

ROY, A. R. (1956). On χ^2 statistics with variable intervals. Technical Report 1, Dept. of Statistics, Stanford Univ.

TALAGRAND, M. (1987). The Glivenko–Cantelli problem. *Ann. Probab.* **15** 837–870.

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

VAPNIK, V. N. and ČERVONENKIS, A. YA. (1974). *Teoriya Raspoznavaniya Obrazov: Statisticheskie Problemy Obucheniya* (*Theory of Pattern Recognition: Statistical Problems of Learning*). (In Russian.) Nauka, Moscow. German ed.: *Theorie der Zeichenerkennung*, by W. N. Wapnik and A. J. Tscherwonenkis, translated by K. G. Stöckel and B. Schneider (S. Unger and K. Fritzsch, eds.). Akademie-Verlag, Berlin, 1979 (*Elektronisches Rechnen und Regeln*, Sonderband **28**).

WATSON, G. S. (1958). On chi-square goodness-of-fit tests for continuous distributions. *J. Roy. Statist. Soc. Ser. B* **20** 44–61.

WENOCUR, R. S. and DUDLEY, R. M. (1981). Some special Vapnik–Červonenkis classes. *Discrete Math.* **33** 313–318.

DEPARTMENT OF MATHEMATICS
ROOM 2-245
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139