Journal of Theoretical Probability, Vol. 4, No. 3, 1991

Uniform and Universal Glivenko–Cantelli Classes

R. M. Dudley,[1] E. Giné,[2] and J. Zinn[3]

Received January 28, 1990; revised January 28, 1991

A class $\tilde{F}$ of measurable functions on a probability space is called a Glivenko–Cantelli class if the empirical measures P_n converge to the true P uniformly over $\tilde{F}$ almost surely. $\tilde{F}$ is a universal Glivenko–Cantelli class if it is a Glivenko–Cantelli class for all laws P on a measurable space, and a uniform Glivenko–Cantelli class if the convergence is also uniform in P. We give general sufficient conditions for the Glivenko–Cantelli and universal Glivenko–Cantelli properties and examples to show that some stronger conditions are not necessary. The uniform Glivenko–Cantelli property is characterized, under measurability assumptions, by an entropy condition.

KEY WORDS: Laws of large numbers; Vapnik–Červonenkis classes.

1. INTRODUCTION

Let $(X, \tilde{A}, P)$ be a probability space. Let $X_1, X_2, \ldots,$ be independent, identically distributed variables with values in X and distribution P, specifically coordinates on a countable product of copies of $(X, \tilde{A}, P)$. Let $X(i) := X_i$ and let P_n be the usual empirical measures, $P_n := (\delta_{X(1)} + \cdots + \delta_{X(n)})/n$. Let $\tilde{F}$ be a class of measurable functions on X. For any real-valued function G on $\tilde{F}$, such as a signed measure for which all functions in $\tilde{F}$ are integrable, let $\|G\|_{\tilde{F}} := \sup\{|G(f)|: f \in \tilde{F}\}$. If all functions in $\tilde{F}$ are integrable for P, then $\tilde{F}$ will be called a (strong) *Glivenko–Cantelli class* for P iff $\|P_n - P\|_{\tilde{F}} \to 0$ almost surely as $n \to \infty$. For such a class Talagrand[28] has shown that although the $\|P_n - P\|_{\tilde{F}}$ are not necessarily measurable they

[1] Massachusetts Institute of Technology, Department of Mathematics, Cambridge, Massachusetts 02139.
[2] University of Connecticut, Department of Mathematics, Storrs, Connecticut 06269-3009.
[3] Texas A & M University, Department of Mathematics, College Station, Texas 77843.

0894-9840/91/0700-0485$06.50/0

are majorized by measurable functions converging to zero; on this and other definitions of and conditions for Glivenko–Cantelli properties for one P see Section 2 below, especially Theorem 1.

Let $(X, \tilde{A})$ be a measurable space and $\tilde{P} = \tilde{P}(X, \tilde{A})$ the set of all laws (probability measures) on it. A subset of X is called *universally measurable* if it is measurable for the completion of every law in $\tilde{P}$. The collection of all universally measurable sets forms a σ-algebra $\tilde{U} \supset \tilde{A}$. A class of real-valued universally measurable functions on X will be called a *universal* Glivenko–Cantelli class if it is a Glivenko–Cantelli class for every probability measure P defined on $\tilde{A}$. $\tilde{F}$ will be called a *uniform* Glivenko–Cantelli class if for every $\varepsilon > 0$ there is an n_0 such that for any probability measure P on $\tilde{A}$ and any $n \geqslant n_0$, $\Pr^*\{\|P_n - P\|_{\tilde{F}} > \varepsilon\} < \varepsilon$.

For a class $\tilde{C}$ of sets, the seminorm $\|\cdot\|_{\tilde{C}}$ is naturally defined as $\|\cdot\|_{\tilde{F}}$ where $\tilde{F}$ is the set of indicator functions of sets in $\tilde{C}$. So a class of measurable sets will be called a Glivenko–Cantelli class if its class of indicator functions is such a class, and likewise for uniform and universal Glivenko–Cantelli classes.

The notion of Glivenko–Cantelli class (for one P) and its vital relation with the notion of shattering are due to Vapnik and Červonenkis,[29],[30] with notable later work by Steele[26] and Talagrand.[28] Let $\tilde{C}$ be a collection of subsets of X. Then $\tilde{C}$ *shatters* a finite set F iff every subset of F is of the form $A \cap F$ for some $A \in \tilde{C}$. Let $S(\tilde{C})$ be the supremum of cardinalities of shattered sets. If $S(\tilde{C}) < \infty$, then $\tilde{C}$ is called a *Vapnik–Červonenkis class*. An infinite set Y will be said to be *finitely shattered* by $\tilde{C}$ iff every finite subset of Y is shattered by $\tilde{C}$.

We will see (Proposition 4) that a universal (and thus a uniform) Glivenko–Cantelli class of functions must be uniformly bounded up to additive constants. For a preview of our further results, first consider classes of sets. A class of sets satisfying suitable measurability conditions is a uniform Glivenko–Cantelli class if and only if it is a Vapnik–Červonenkis class (Proposition 11 below). On the other hand the collection of all subsets of a countable set is a universal but not uniform Glivenko–Cantelli class. In fact, if the sample space is a countable union of measurable sets A_m, and on each A_m, $\tilde{C}$ induces a Vapnik–Červonenkis class $\tilde{C}_m$ (with suitable measurability), then $\tilde{C}$ is a universal Glivenko–Cantelli class, although $S(\tilde{C}_m)$ can grow with m (Corollary 4). The existence of such A_m is not necessary for the universal Glivenko–Cantelli property (Proposition 6). A more general sufficient condition will be given based on the following.

Let $(X, \tilde{A}, P)$ be a probability space. For any real-valued function g on X let g^* be a measurable cover function for g, i.e., a measurable function $g^* \geqslant g$ everywhere such that for any measurable function $h \geqslant g$ everywhere, we have $h \geqslant g^*$ a.s. (Ref. 5, Section 3.1). Here g^* has values in $\mathbb{R} \cup \{+\infty\}$ and may be infinite with positive probability even if g is finite everywhere.

Definitions. Let $\tilde{D}_m := \tilde{D}(m)$, $m = 1, 2,...,$ be a sequence of classes of measurable subsets of X. Then we say $\{\tilde{D}_m\}_{m \geqslant 1}$ is *P-asymptotically null* if $\|P\|_{\tilde{D}(m)} \to 0$ as $m \to \infty$ and $\|P_n\|^*_{\tilde{D}(m)} \to 0$ in probability as $m \to \infty$ and $n \to \infty$. If $(X, \tilde{A})$ is a measurable space and the sets in $\tilde{D}_m$ are universally measurable for each m, $\{\tilde{D}_m\}_{m \geqslant 1}$ will be called *universally* asymptotically null if it is P-asymptotically null for every nonatomic law P on $(X, \tilde{A})$.

Let $\tilde{F}$ be a uniformly bounded class of universally measurable functions such that for some universally asymptotically null sequence of classes $\{\tilde{D}_m\}$ and some universal Glivenko–Cantelli classes $\tilde{F}_m$, for each $f \in \tilde{F}$ and m there is a $D \in \tilde{D}_m$ such that the restriction of f to the complement of D is in $\tilde{F}_m$. Then $\tilde{F}$ is a universal Glivenko–Cantelli class of functions (Theorem 5). We do not know of any universal Glivenko–Cantelli classes of sets other than those obtained in this way where the $\tilde{F}_m$ are suitably measurable Vapnik–Červonenkis classes.

In our main result on uniform Glivenko–Cantelli classes (Theorem 6), such classes, if they satisfy a mild measurability condition, are characterized by an entropy condition.

2. GLIVENKO–CANTELLI CLASSES

Let $(X, \tilde{A}, P)$ be a probability space. A collection $\tilde{F}$ of functions integrable for P is called *order-bounded* iff for the envelope function $F_{\tilde{F}}(x) := \sup\{|f(x)|: f \in \tilde{F}\}$, $F^*_{\tilde{F}}$ is integrable for P. In other words, there exists $u \in \mathscr{L}^1(P)$ such that for all $f \in \tilde{F}$, $|f(x)| \leqslant u(x)$ for all x. For any $\tilde{F} \subset \mathscr{L}^1(P)$ let $F_{0,P} := \{f - \int f\,dP: f \in \tilde{F}\}$. Then clearly $\tilde{F}$ is a Glivenko–Cantelli class for P if and only if $\tilde{F}_{0,P}$ is. Here are some characterizations of Glivenko–Cantelli classes:

Theorem 1. (Talagrand.) For any class $\tilde{F}$ of functions integrable for P, the following are equivalent as $n \to \infty$:

(a) $\|P_n - P\|^*_{\tilde{F}} \to 0$ in probability and $\tilde{F}_{0,P}$ is order bounded

(b) $\|P_n - P\|^*_{\tilde{F}} \to 0$ almost surely.

(c) $\|P_n - P\|_{\tilde{F}} \to 0$ almost surely

Proof. This is a corollary of Talagrand (Ref. 28, Theorem 22, p. 860). Specifically, each of (a), (b), or (c) for $\tilde{F}$ is equivalent to the same statement for $\tilde{F}_{0,P}$; we restrict attention to $\tilde{F}_{0,P}$. Now (b) implies (c), which is equivalent to Talagrand's (I) (for $\tilde{F}_{0,P}$). Also (a) is intermediate between

Talagrand's equivalent (VI) and (VII), so it is also equivalent. Talagrand's (V) implies (b). □

$\tilde{F}$ is a Glivenko–Cantelli class for P, by definition, if (c) holds, and so if and only if any of the three equivalent conditions in Theorem 1 holds. Call $\tilde{F}$ a *weak* Glivenko–Cantelli class for P if $\|P_n - P\|_{\tilde{F}}^* \to 0$ in probability as $n \to \infty$. An example of a weak Glivenko–Cantelli class which is not a Glivenko–Cantelli class will be given in Proposition 3.

In treating (strong) Glivenko–Cantelli classes we can (up to additive constants) restrict ourselves to order-bounded classes. The main result of Talagrand (Ref. 28, Theorem 2, p. 838) characterizes non-Glivenko–Cantelli classes, as follows. For any measurable set $A \subset X$, real $\alpha < \beta$, and positive integer n let

$$W(\tilde{F}, A, \alpha, \beta, n) := \{(x_1,..., x_n, y_1,..., y_n) \in A^{2n}\colon \text{for some } f \in \tilde{F},$$

$$f(x_i) < \alpha < \beta < f(y_i) \text{ for all } i = 1,..., n\}$$

Theorem 2. (Talagrand.) An order bounded class $\tilde{F}$ for P fails to be a Glivenko–Cantelli class for P if and only if there exist some real $\alpha < \beta$ and a measurable set A with $P(A) > 0$ such that for all n,

$$P^{2n*}(W(\tilde{F}, A, \alpha, \beta, n)) = P(A)^{2n}$$

A class $\tilde{F}$ of measurable functions is said to satisfy condition (M) if for every $\alpha < \beta$ and positive integer n, $W(\tilde{F}, X, \alpha, \beta, n)$ is P^{2n} measurable, and hence so is $W(\tilde{F}, A, \alpha, \beta, n)$ for any measurable A. For classes of sets, Talagrand (Ref. 28, Theorem 5, p. 839) obtains the following:

Theorem 3. (Talagrand.) A class $\tilde{C}$ of measurable sets satisfying (M) fails to be a Glivenko–Cantelli class for P if and only if there is a measurable set A with $P(A) > 0$ on which P is nonatomic such that for all n and P^n-almost all choices of $x_1,..., x_n$ in A, $\tilde{C}$ shatters $\{x_1,..., x_n\}$.

Theorems 2 and 3 say that when the Glivenko–Cantelli property fails, the failure can be in a sense localized on some set, in terms of shattering. It can be asked, though, when the Glivenko–Cantelli property does hold for a class $\tilde{C}$ of sets, can the good property also be localized: Are there sets with probability close to zero on whose complements $\tilde{C}$ is a Vapnik–Červonenkis class? This turns out not to be necessary for any fixed sets, but a sufficient condition will be given where the sets of probability near zero, in an asymptotically null class, can depend on a set in $\tilde{C}$ being approximated.

If $\{\tilde{D}_m\}_{m \geq 1}$ is P-asymptotically null, and $\tilde{E}_m$ is the class of all measurable subsets of sets in $\tilde{D}_m$, then clearly $\tilde{E}_m$ is also P-asymptotically null.

Example 1. For X equal to the unit square I^2 with Lebesgue measure P, let $\tilde{D}_m$ be the class of all measurable sets A such that for some b, c with $0 < c - b < 1/m$, $b < y < c$ for all (x, y) in A. Then $\{\tilde{D}_m\}_{m \geq 1}$ is P-asymptotically null but not universally asymptotically null (consider one-dimensional Lebesgue measures on lines $y = \text{constant}$).

Example 2. Let $X = \bigcup_{i \geq 0} X_i$ where X_i are disjoint (universally) measurable sets. Let $\tilde{D}_m$ consist of the one set $Y_m := X_0 \cup \bigcup_{i \geq m} X_i$, or of Y_m and all its (universally) measurable subsets. Then $\{\tilde{D}_m\}_{m \geq 1}$ is P-asymptotically null if $P(X_0) = 0$ and universally asymptotically null if X_0 is universally null, meaning that $P(X_0) = 0$ for all nonatomic laws P on $(X, \tilde{A})$; about such sets see Section 3 below.

Example 3. Let (X, d) be a separable metric space with Borel σ-algebra. Let $\delta_m \downarrow 0$. Let $\tilde{D}_m$ be the class of all universally measurable sets A with diameters $\operatorname{diam} A := \sup\{d(x, y)\colon x \in A,\ y \in A\} \leq \delta_m$. Then $\tilde{D}_m$ is universally asymptotically null.

Next, here is a way to extend the Glivenko–Cantelli property from simpler to more complex classes of functions.

Theorem 4. Let $(X, \tilde{A}, P)$ be a probability space and $\{\tilde{D}_m\}_{m \geq 1}$ a P-asymptotically null sequence of classes of measurable sets. For $m = 1, 2, \ldots$, let $\tilde{F}_m$ be a Glivenko–Cantelli class of functions for P. Let $\tilde{F}$ be a class of measurable functions, order-bounded in $\mathscr{L}^1(P)$. Suppose that for any $f \in \tilde{F}$ and $m = 1, 2, \ldots$, there is a set $A \in \tilde{D}_m$ with $f(1 - 1_A) \in \tilde{F}_m$. Then $\tilde{F}$ is a Glivenko–Cantelli class for P.

Proof. Let $|f| \leq u \in \mathscr{L}^1(P)$ for all $f \in \tilde{F}$. For any $\varepsilon > 0$, there is a $\delta > 0$ such that if $P(A) < \delta$, then $|\int_A f\, dP| \leq \int_A u\, dP < \varepsilon/3$ for all $f \in \tilde{F}$. Also, let $x_1, x_2, \ldots$, be i.i.d. (P) and $X_i := u(x_i)$. By the strong law of large numbers, we have

$$\frac{1}{n} \sum_{i=1}^{n} X_i 1\{X_i \geq M\} \to 0 \text{ a.s.} \qquad \text{as} \quad n \to \infty \text{ and } M \to \infty$$

Thus since $\|P\|_{\tilde{D}(m)} \to 0$ and $\|P_n\|^*_{\tilde{D}(m)} \to 0$ in probability as $m \to \infty$ and $n \to \infty$, we have $(\sup\{P_n(u 1_A)\colon A \in \tilde{D}(m)\})^* \to 0$ in probability, considering cases $X_i < M$ and $X_i \geqslant M$ for large M. Take m_0 and n_0 large enough so that for $m \geqslant m_0$ and $n \geqslant n_0$,

$$\|P\|_{\tilde{D}(m)} < \delta$$

$$\Pr\{(\sup\{P_n(u1_A): A \in \tilde{D}(m)\})^* > \varepsilon/3\} < \varepsilon/2$$

and

$$\Pr\{\|P_n - P\|^*_{\tilde{F}(m)} > \varepsilon/3\} < \varepsilon/2$$

For any $f \in \tilde{F}$, take $A \in \tilde{D}(m)$ with $f(1 - 1_A) \in \tilde{F}_m$. Then

$$\begin{aligned} |(P_n - P)(f)| &\leqslant |(P_n - P)(f(1 - 1_A))| + |(P_n - P)(f1_A)| \\ &\leqslant \|P_n - P\|_{\tilde{F}(m)} + P_n(u1_A) + P(u1_A) \\ &\leqslant \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon \end{aligned}$$

except on a set with (outer) probability at most $\varepsilon/2 + \varepsilon/2 = \varepsilon$, so the conclusion follows (from Theorem 1). □

A Vapnik–Červonenkis class satisfying some measurability conditions is a Glivenko–Cantelli class.[29] A Vapnik–Červonenkis class satisfying a measurability condition enough to imply universal Glivenko–Cantelli will be called *sufficiently measurable.* The image admissible Suslin condition implies sufficiently measurable (Ref. 5, Sections 10.3 and 11.1). Measurability conditions cannot be completely removed (Ref. 5, example 10.3, p. 101).

Corollary 1. Let the hypotheses of Theorem 4 hold for some $\{\tilde{D}_m\}$ where $\tilde{F}$ and $\tilde{F}_m$ are classes of indicators of sets in classes $\tilde{C}$ and $\tilde{C}_m$, respectively, and $\tilde{C}_m$ are sufficiently measurable Vapnik–Červonenkis classes. Then $\tilde{C}$ is a Glivenko–Cantelli class for P.

We do not know whether the sufficient condition for the Glivenko–Cantelli property of a class of sets in Corollary 1 is necessary.

Applying Example 2 before Theorem 4 gives the following:

Corollary 2. Supose that X can be decomposed as a union of disjoint measurable sets X_i, $i = 0, 1, 2,...$, and $\tilde{C}$ is a collection of measurable subsets of X such that for each $i > 0$, the collection $\tilde{C}_i$ of all sets $C \cap X_i$ for $C \in \tilde{C}$ is a sufficiently measurable Vapnik–Červonenkis class. Let P be any law such that $P(X_0) = 0$. Then $\tilde{C}$ is a Glivenko–Cantelli class for P.

Glivenko–Cantelli classes of sets for a given P, at least under some measurability conditions, have been variously characterized (Vapnik and Červonenkis[29],[30]; Steele,[26] Talagrand[28]). Although it does not seem clear from these characterizations (such as Theorem 3 above) whether the sufficient condition of Corollary 2 is necessary, in fact it is not:

Proposition 1. There exist a probability measure P and a Glivenko–Cantelli class $\tilde{C}$ for P such that $\tilde{C}$ does not satisfy the condition in Corollary 2, and $\tilde{C}$ is not a universal Glivenko–Cantelli class.

Proof. Let P be Lebesgue measure on the unit square $I^2 := \{(x, y): 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1\}$. Let $\tilde{C}$ be the collection of all measurable subsets of horizontal intervals, on each of which y has a fixed value. Then for each $A \in \tilde{C}$, $P(A) = 0$. For each n, $\|P_n\|_{\tilde{C}} = 1/n$ a.s., so $\tilde{C}$ is a Glivenko–Cantelli class. It is not a universal Glivenko–Cantelli class: Consider a nonatomic law (having no atoms) concentrated on one horizontal interval. Next, if P satisfies the condition of Corollary 2, there is a set X_i of positive measure on which $\tilde{C}$ is a Vapnik–Červonenkis class. By the Tonelli–Fubini theorem, the intersection of X_i with some horizontal interval has positive linear measure and is thus infinite, a contradiction. □

An example of nonexistence of the decomposition where $\tilde{C}$ is a universal Glivenko–Cantelli class will be given in Proposition 6 below.

Kolmogorov[15] proved the folowing. Let $X_1, X_2, \ldots,$ be i.i.d. real random variables and $S_n := X_1 + \cdots + X_n$. Then the following are equivalent:

(i) For some constants a_n, $|S_n/n - a_n| \to 0$ in probability as $n \to \infty$.

(ii) $nP(|X_1| > n) \to 0$ as $n \to \infty$.

For random variables in a separable Banach space, with absolute value replaced by norm, (i) always implies (ii), whereas (ii) implies (i) if and only if the Banach space is B-convex (B-convexity is defined in Section 4). See Araujo and Giné,[2] Mandrekar and Zinn,[17] and Marcus and Woyczynski.[18] For completeness we give a direct proof of (i) ⇒ (ii) for symmetric elements of a normed space:

Proposition 2. Let $X_1, X_2, \ldots,$ be random elements of a normed space $(X, \|\cdot\|)$ such that for each n and $s_j = 1$ or -1 for each $j = 1, \ldots, n$, $\|s_1 X_1 + \cdots + s_n X_n\|$ and $\|X_n\|$ are measurable, the joint distribution of $\|s_1 X_1\|, \|s_1 X_1 + s_2 X_2\|, \ldots, \|s_1 X_1 + \cdots + s_n X_n\|$ does not depend on $s_1, \ldots, s_n$, and $\|X_1\|, \|X_2\|, \ldots$ are i.i.d. Let $S_n := X_1 + \cdots + X_n$. Then (a) implies (b):

(a) $\|S_n\|/n \to 0$ in probability as $n \to \infty$.

(b) $nP(\|X_1\| > n) \to 0$.

Proof. Assume (a). Then

$$P(\max_{1 \leqslant j \leqslant n} \|X_j\| > n) \leqslant P(\max_{j \leqslant n} \|S_j\|/n > \tfrac{1}{2})$$

which by the P. Lévy inequality (which holds under the given conditions) is $\leqslant 2P(\|S_n\|/n > \frac{1}{2}) \to 0$ as $n \to \infty$. Then for $p_n := P(\|X_1\| > n)$, $(1-p_n)^n \to 1$ implies $np_n \to 0$, which is (b). □

Proposition 3. There is a weak Glivenko–Cantelli class which is not a Glivenko–Cantelli class.

Proof. Let H be a separable Hilbert space with an orthonormal basis $\{e_i\}_{1 \leqslant i < \infty}$. Let $\tilde{F} = B_1$, the unit ball of the (dual) Hilbert space. So on H, $\|\cdot\|_{\tilde{F}}$ is just the Hilbert norm $\|\cdot\|$. Let P on H put probability $p_i/2$ at $k_i e_i$ and also at $-k_i e_i$ for each i, where $p_i = c/i^2$ $(c = 6/\pi^2)$ and $k_i = i/(1+\log i)$, $i = 1, 2,...$. Then for each $f \in \tilde{F}$, $\int |f|\, dP = \sum_i p_i k_i |f(e_i)| < \infty$ by the Cauchy inequality, and $\int f\, dP = 0$ by symmetry. We have

$$nP(\|x\| > n) \leqslant n \sum \{p_i : k_i > n\} \leqslant n \sum \{p_i : i > n(1+\log n)\}$$

$$\leqslant c/\log n \to 0 \qquad \text{as} \quad n \to \infty$$

so (since Hilbert space is B-convex) by the references given before Proposition 2, (a) in Proposition 2 follows, so $\tilde{F}$ is a weak Glivenko–Cantelli class.

Now $\int \|x\|\, dP(x) = +\infty$, which says in this case that $\tilde{F} = \tilde{F}_{0,P}$ is not order-bounded, so $\tilde{F}$ is not a Glivenko–Cantelli class by Theorem 1 (or by the Mourier[19] law of large numbers). □

3. UNIVERSAL GLIVENKO–CANTELLI CLASSES

A collection $\tilde{F}$ of universally measurable functions on X will be called a *weak universal* Glivenko–Cantelli class if there exist some real-valued $a_n(f, P)$, defined for each $f \in \tilde{F}$, $n = 1, 2,...$, and law P on $(X, \tilde{A})$, such that $\|P_n - a_n(\cdot, P)\|^*_{\tilde{F}} \to 0$ in probability as $n \to \infty$. This notion turns out to be equivalent to that of universal Glivenko–Cantelli class defined early in Section 1 (although for a single P equivalence does not hold—the weak law of large numbers is not equivalent to the strong law, as is well known–see Propositions 2 and 3 above).

Proposition 4. For any measurable space $(X, \tilde{A})$ and class $\tilde{F}$ of universally measurable real-valued functions on X, $\tilde{F}$ is a weak universal

Glivenko–Cantelli class if and only if it is a universal Glivenko–Cantelli class. Also, each function f in such a class $\tilde{F}$ must be bounded, and the collection of functions $\{f - \inf f : f \in \tilde{F}\}$ is uniformly bounded ($\tilde{F}$ is uniformly bounded up to additive constants).

Proof. Suppose $\tilde{F}$ is a weak universal Glivenko–Cantelli class. If $\tilde{F}$ is not uniformly bounded up to additive constants, take $f_k \in \tilde{F}$ and points x_k and y_k such that $|f_k(x_k) - f_k(y_k)| > 8^k$ for $k = 1, 2, \ldots$. Let $\tilde{G}$ be the sequence $\{f_k\}_{k \geq 1}$. Since $\tilde{G}$ is countable, the following $\|\cdot\|_{\tilde{G}}$ seminorms of random elements are measurable (even though the P_n may not necessarily belong to a normed space separable for $\|\cdot\|_{\tilde{G}}$ or be Borel measurable for it). Let P be the sum of point masses $1/2^{k+1}$ at x_k and at y_k for each k. Let P'_n be an independent copy of P_n. Then $\|P_n - P'_n\|_{\tilde{G}} \to 0$ in probability since $\tilde{F}$ is a weak universal Glivenko–Cantelli class. Now, $P_n - P'_n$ is $1/n$ times a sum of n i.i.d. random elements $V_i = \delta_{X(i)} - \delta_{Y(i)}$ which are symmetric [V_i can be interchanged with $-V_i$ for any set of i's without changing the distribution(s) of the $\|\cdot\|_{\tilde{G}}$ seminorms of any partial sums]. Under these conditions, it follows from Proposition 2 that $n \Pr(\|V_1\|_{\tilde{G}} > n) \to 0$ as $n \to \infty$. But now by definition of P in this case, for each k, and $n = 8^k$,

$$n \Pr(\|V_1\|_{\tilde{G}} > n) \geq n \Pr\{X(1) = x_k \text{ and } Y(1) = y_k\} \geq 8^k/4^{k+1} \to \infty$$

a contradiction. So $\tilde{F}$ is uniformly bounded up to additive constants: for some $M < \infty$, $\sup f - \inf f \leq M$ for all $f \in \tilde{F}$.

To finish the proof, we need to show that for each law P, as $n \to \infty$, $\sup\{|\int f\, dP - a_n(f, P)| : f \in \tilde{F}\} \to 0$. If not, then for some P, some $\varepsilon > 0$, sequence $n(k) \to \infty$ and $g_k \in \tilde{F}$, $|\int g_k\, dP - a_{n(k)}(g_k, P)| > \varepsilon$. For k large enough, $\Pr\{|\int g_k\, dP_{n(k)} - a_{n(k)}(g_k, P)| > \varepsilon/2\} < \frac{1}{2}$ and, by Chebyshev's inequality, $\Pr\{|\int g_k\, d(P_{n(k)} - P)| > \varepsilon/2\} < 4M^2/(\varepsilon^2 n(k)) < \frac{1}{2}$. Evaluating at a point where neither of the latter two events occur and subtracting gives a contradiction, proving Proposition 4. □

A notable example of a universal Glivenko–Cantelli class is the collection of all subsets of a countable set. So, for the Glivenko–Cantelli property the atomic parts of laws pose no problem. We then clearly have the following:

Proposition 5. A class $\tilde{C}$ is a universal Glivenko–Cantelli class if and only if it is a Glivenko–Cantelli class for all nonatomic P in $\tilde{P}$.

Here is a way to extend the universal Glivenko–Cantelli property to fairly general classes of functions from simpler classes, just as in Theorem 4 for one P:

Theorem 5. For $m = 1, 2, \ldots$, let $\tilde{F}_m$ be a universal Glivenko–Cantelli class of functions. Let $\{\tilde{D}_m\}_{m \geqslant 1}$ be a universally asymptotically null sequence of classes of universally measurable sets. Let $\tilde{F}$ be a uniformly bounded class of measurable functions such that for any $f \in \tilde{F}$ and each m there is a set $A \in \tilde{D}_m$ such that $f(1 - 1_A) \in \tilde{F}_m$. Then $\tilde{F}$ is a universal Glivenko–Cantelli class.

Proof. This follows from Theorem 4 and its proof. Actually, a direct proof would be slightly easier since $\tilde{F}$ is assumed uniformly bounded. □

Likewise, Corollary 1 for sets has its counterpart in the universal case, where "sufficiently measurable" is as in Corollary 1:

Corollary 3. Assume that $\tilde{C}$ is a class of universally measurable sets; $\{\tilde{D}_m\}_{m \geqslant 1}$ is universally asymptotically null; for each m, $\tilde{C}_m$ is a sufficiently measurable Vapnik–Červonenkis class; and that for each $C \in \tilde{C}$ and each m there is an $A \in \tilde{D}_m$ with $C \backslash A \in \tilde{C}_m$. Then $\tilde{C}$ is a universal Glivenko–Cantelli class of sets.

We do not know whether the condition in Corollary 3 is necessary for the universal Glivenko–Cantelli property for a class of sets.

A set $C \subset X$ is called *universally null* if it has outer measure zero for all nonatomic laws on $\tilde{A}$ (and so is universally measurable, with measure zero for the completion of any nonatomic law). If all singletons are in $\tilde{A}$, then all countable sets trivially are universally null. Uncountable universally null sets, otherwise quite unusual, may turn up in theorems on universal Glivenko–Cantelli classes. For example, by Proposition 5 and Proposition 8 below, the collection $\tilde{C}$ of all subsets of a set D is a universal Glivenko–Cantelli class if and only if D is universally null. Then, for any nonatomic P, $\|P_n - P\|_{\tilde{C}} \equiv 0$ a.s. A fixed universally null set could be included in all sets in universally asymptotically null classes (Theorem 5, Corollary 3).

In a Polish (complete separable metric) space, an uncountable Borel set B cannot be universally null, since by the Borel isomorphism theorem (e.g., Ref. 7, Section 13.1) there exists a nonatomic law P supported on B. But in any uncountable Polish space, there exist uncountable universally null sets. In fact, there are uncountable sets, called Lusin sets, having countable intersection with any nowhere dense compact set, as shown by Lusin,(16) assuming the continuum hypothesis. It is known that any Lusin set (in the current sense) is universally null. For completeness here is a proof, kindly conveyed to us by W. Adamski via P. Gaenssler. For any nonatomic law P on a Polish space and Borel set B, $P(B) = \sup\{P(K)$: $K \subset B$, K compact and nowhere dense$\}$. Suppose there is a Lusin set C and a nonatomic law μ such that the outer measure $\mu^*(C) > 0$. Let B be a Borel set with $C \subset B$ and $\mu^*(C) = \mu(B)$. Also, there are closed, nowhere dense sets $F_1 \subset F_2 \subset \cdots \subset B$ with $\mu(B) = \sup_n \mu(F_n)$. Then for any compact $K \subset B$,

$$\mu(K) = \mu\left(K \cap \bigcup_n F_n\right) = \mu^*\left(K \cap \bigcup_n F_n \cap C\right) + \mu_*\left(\left(K \cap \bigcup_n F_n\right)\backslash C\right) = 0$$

since (e.g., Ref. 7, Theorem 3.1.11) $\mu^*(K \cap \bigcup_n F_n \cap C) = \lim_n \mu^*(K \cap F_n \cap C) = 0$, because each $K \cap F_n \cap C$ is countable and μ is nonatomic, and since

$$\mu_*\left(\left(K \cap \bigcup_n F_n\right)\backslash C\right) \leqslant \mu_*(K\backslash C) \leqslant \mu_*(B\backslash C) = 0$$

So, $0 < \mu^*(C) = \mu(B) = \sup\{\mu(K)\colon K \text{ compact, } K \subset B\} = 0$, a contradiction.

Sierpiński and Szipilrajn-Marczewski[25] on the basis of results of Hausdorff,[13] without the continuum hypothesis, proved that there exist uncountable universally null (not necessarily Lusin) sets in $[0, 1]$. A recent reference, giving others, is Shortt.[24]

Corollary 2 has the following straightforward counterpart:

Corollary 4. Suppose X is the union of disjoint universally measurable sets $A_0, A_1, A_2, \ldots$, such that A_0 is universally null. Let $\tilde{C}$ be a collection of sets such that for each $i > 0$, $\tilde{C}_i := \{C \cap A_i\colon C \in \tilde{C}\}$ is a sufficiently measurable Vapnik–Červonenkis class. Then $\tilde{C}$ is a universal Glivenko–Cantelli class.

Since the example in Proposition 1 is not universal Glivenko–Cantelli, a different example is needed of a universal Glivenko–Cantelli class for which the conditions of Corollary 4 fail.

Proposition 6. There exists a universal Glivenko–Cantelli class of finite sets in $I := [0, 1]$ not satisfying the conditions in Corollary 4.

Proof. Let $\tilde{D}_m$ be the class of all intervals of length at most $1/m$ in I, $m = 1, 2, \ldots$. Then $\{\tilde{D}_m\}_{m \geqslant 1}$ is universally asymptotically null. Let $\tilde{C}$ be the collection of all finite subsets of I such that each set C in $\tilde{C}$ of cardinality m is included in some interval A in $\tilde{D}_m$. Then Corollary 3 applies as follows: Let $\tilde{C}_m$ be the class of all sets with at most $m-1$ members, so $S(\tilde{C}_m) = m-1$. For any $C \in \tilde{C}$, if C has fewer than m elements, take A to be arbitrary, say $A = [0, 1/m]$, and $C\backslash A \in \tilde{C}_m$. If C has m or more members, take an interval $A \supset C$ with $A \in \tilde{D}_m$. Then $C\backslash A$ is empty and so in $\tilde{C}_m$. Since each $\tilde{C}_m$ is a direct image of I^{m-1}, it has suitable measurability properties (Ref. 5, pp. 101, 108), so it is a universal Glivenko–Cantelli class and, by Corollary 3, so is $\tilde{C}$.

If a decomposition as in Corollary 4 existed, then some A_i must have positive Lebesgue measure, and for $\tilde{E}_i := \{A_i \cap C\colon C \in \tilde{C}\}$, then $S(\tilde{E}_i) = m < \infty$. But A_i has infinitely many points in some interval of length less than $1/(m+1)$; any $m+1$ of these points form a shattered set, contradicting Corollary 4. The proof of Proposition 6 is complete. □

Theorem 5 and Corollary 3 provided sufficient conditions for the universal Glivenko–Cantelli property. Here is a necessary condition:

Proposition 7. If $\tilde{C}$ is a universal Glivenko–Cantelli class of sets, then any set finitely shattered by $\tilde{C}$ must be universally null.

Proof. Suppose A is finitely shattered by $\tilde{C}$ and not universally null. Then for some nonatomic probability measure P on $(X, \tilde{A})$, $P^*(A) > 0$. There is a measurable cover B of A, so that $A \subset B$, B is measurable, and $P(B) = P^*(A)$. Replacing P by Q defined by $Q(C) = P(B \cap C)/P(B)$ for all measurable sets C, we can assume that $P^*(A) = 1$.

We symmetrize: If Q_n are also empirical measures for P, independent of P_n, then $\|P_n - Q_n\|^*_{\tilde{C}} \to 0$ a.s. as $n \to \infty$. Since P is nonatomic, the $2n$ points supporting $P_n + Q_n$ are almost surely all distinct. With outer probability 1, all these points belong to A (e.g., Ref. 5, Lemma 3.1.4), so $\|P_n - Q_n\|^*_{\tilde{C}} = 1$ a.s., a contradiction. □

For example, in a Polish space, if $\tilde{C}$ is a universal Glivenko–Cantelli class there cannot exist any uncountable Borel set K finitely shattered by $\tilde{C}$, since as noted above such a set is not universally null. Condition (M) in Section 2, Theorem 3, above was defined (implicitly) in terms of P^{2n} measurability for one probability measure P. A collection $\tilde{F}$ will be said to satisfy condition (MU) if the sets in the definition of condition (M) are all P^{2n} measurable for every law P on $(X, \tilde{A})$. We then have another sufficient condition for the universal Glivenko–Cantelli property:

Proposition 8. Let $\tilde{C}$ be a class of measurable sets satisfying (MU) for which there are no infinite, finitely shattered sets; or, all such sets are included in one universally null set T. Then $\tilde{C}$ is a universal Glivenko–Cantelli class.

Proof. Removing T, we can assume that there are no infinite, finitely shattered sets. Suppose $\tilde{C}$ is not universal Glivenko–Cantelli, so it is not Glivenko–Cantelli for some P. Apply a theorem of Talagrand (Section 2, Theorem 3, above) to get a set A as described there. Let $Q(C) := P(C \cap A)/P(A)$ for any measurable set C. Then Q has no atoms and for all n and Q^n-almost all $x_1,..., x_n$, $\tilde{C}$ shatters $\{x_1,..., x_n\}$. Then there exist sequences $\{x_i\}_{i=1}^{\infty}$ of distinct points whose ranges are finitely shattered by $\tilde{C}$, in fact almost all such sequences for Q^{∞}, a contradiction. □

In the following Proposition 9, part (ii) shows that the necessary condition for the universal Glivenko–Cantelli property in Proposition 7 is not sufficient, and part (i) shows that the sufficient condition in Proposition 8 is not necessary, even assuming (MU):

Proposition 9. Assuming the continuum hypothesis, $I := [0, 1]$ can be decomposed as a union $I = \bigcup_{\alpha} Y_{\alpha}$ of disjoint, uncountable, Lusin universally null sets. Let $\tilde{C}$ be the union over α of the collections of all finite subsets of Y_{α}. Then Y_{α} can be chosen so that either:

(i) $\tilde{C}$ is a universal Glivenko–Cantelli class of subsets of I, satisfying (MU), with $\|P\|_{\tilde{C}} = 0$ and $\|P_n\|_{\tilde{C}} = 1/n$ a.s. for any nonatomic P, and where the condition of Corollary 4 holds with $I = A_0 \cup A_1$ where A_0 is universally null and $S(\{C \cap A_1 \colon C \in \tilde{C}\}) = 1$; or

(ii) $\tilde{C}$ is not a Glivenko–Cantelli class for any nonatomic measure P on I ($\tilde{C}$ is a "universal non-Glivenko–Cantelli class").

Proof. For (i), by the continuum hypothesis there exists a Lusin, universally null set A_0 of cardinality c. Let $A_1 := I \setminus A_0$. Let $A_1 = \{y_{\alpha}\}_{\alpha < c}$ be an enumeration of A_1 (without repetitions), indexed by the countable ordinals. Also, we can write A_0 as a union of c disjoint uncountable sets T_{α}, $\alpha < c$. Let $Y_{\alpha} := T_{\alpha} \cup \{y_{\alpha}\}$. Then the Y_{α} are disjoint, Lusin universally null sets, as claimed. If $x_1, x_2, \ldots$, are independent with any nonatomic distribution P on I, then there is probability zero that any two x_i are equal, or that any of them belongs to A_0. So the set on which two or more of them are in Y_{α} for the same α has P^n probability zero for any n.

For the measurability property (MU) of $\tilde{C}$, consider the set $W := \{x \in I^{2n} \colon$ for some $C \in \tilde{C}$, $x_i \in C$ if and only if $n < i \leqslant 2n\}$. To show W is P^{2n} measurable for any law P on the Borel sets of I, let $P = \lambda Q + (1 - \lambda)\mu$ where λ is a number in $[0, 1]$, Q is purely atomic, $Q(A_0) = 1$, and $\mu(A_1) = 1$. Let $B \subset A_0$ be a countable set with $Q(B) = 1$. Let $V_1, V_2, \ldots$, be the nonempty sets of the form $Y_{\alpha} \cap B$. Let $W_0 := \{x \in I^{2n} \colon$ for some j and finite set $F \subset V_j$, for all $i = 1, \ldots, 2n$, $x_i \in F$ iff $i > n\}$, $W_1 := \{x \in I^{2n} \colon x_{n+1} = \cdots = x_{2n} = y$ for some $y \notin B$ and $x_i \neq y$, $i = 1, \ldots, n\}$. Then W_0, W_1, and $Z := W_0 \cup W_1$ are Borel sets in I^{2n}, $Z \subset W$, and $W \setminus Z \subset \{x \colon x_i \in A_0 \setminus B$ for some $i > n\}$, so $P^{2n}(W \setminus Z) = 0$ and W is P^{2n} measurable. So (MU) holds. The other conclusions in (i) follow straightforwardly.

Now for (ii), the collection of all compact, nowhere dense sets in I has cardinality c, so by the continuum hypothesis we can write such sets as $\{L_{\beta}\}_{1 \leqslant \beta < c}$ so that β runs through the set of nonzero countable ordinals.

Each Y_α will be defined recursively as $\{y_{\alpha\beta}\}_{\beta<c}$, where α will run through the set of all countable ordinals, and β through that set and a finite sequence of negative integers $-1,\dots,-j_\alpha$ depending on α. We also use the continuum hypothesis to well-order I as $I=\{x_\gamma\}_{\gamma<c}$. Let $y_{00}=x_0$. Given $y_{0\beta}$ for $\beta<\delta$ choose $y_{0\delta}$ such that $y_{0\delta}\neq y_{0\beta}$ for $\beta<\delta$ and $y_{0\delta}\notin\bigcup_{\beta<\delta}L_\beta$, as is possible by the category theorem. Given Y_ζ for $\zeta<\alpha$, Y_α will be defined recursively.

Consider the set of pairs (j, K) where j is an integer, $j\geqslant 2$, and K is a compact subset of I^j such that $P^j(K)>0$ for some nonatomic P (depending on K). The set of all such pairs has cardinal c and so can be indexed as $\{(j_\alpha, K_\alpha)\}_{1\leqslant\alpha<c}$. So j_α is defined. By assumption on K_α, there exists a point t of K_α none of whose coordinates t_i, $i=1,\dots,j_\alpha$, belong to any of the universally null sets Y_ζ for $\zeta<\alpha$. Let $y_{\alpha\beta}=t_{-\beta}$ for $\beta=-1,\dots,-j_\alpha$. Let $y_{\alpha 0}=x_\gamma$ for the least γ such that x_γ is neither in Y_ζ for any $\zeta<\alpha$ (countable many universally null sets) nor equal to any $y_{\alpha\beta}$, $\beta<0$. Then, for any countable ordinal $\delta>0$, given $y_{\alpha\beta}$ for $\beta<\delta$, choose $y_{\alpha\beta}=y$ such that:

(a) $y\neq y_{\alpha\beta}$ for any $\beta<\delta$

(b) $y\notin L_\beta$ for any $\beta<\delta$

(c) $y\notin\bigcup_{\zeta<\alpha}Y_\zeta$.

Such a y exists because $\bigcup_{\beta<\delta}L_\beta\cup\{y_{\alpha\beta}\}$, a countable union of nowhere dense compact sets, has complement an uncountable Borel set which cannot (as noted above) be included in a universally null set, specifically in $\bigcup_{\zeta<\alpha}Y_\zeta$. So all the Y_α are defined. Since $y_{\alpha\delta}$ can belong to L_β only for $\delta\leqslant\beta$, $Y_\alpha\cap L_\beta$ is countable for all α and β, so Y_α is a Lusin set and hence universally null. By (a) each Y_α is uncountable, and by (c) the Y_α for different α are disjoint.

Let P be a nonatomic law on I. To see that $\tilde{C}$ is not Glivenko–Cantelli for P, consider the P^n outer probability, call it p^*, of the set of all $x=(x_1,\dots,x_n)$ such that for some α, all x_i are in Y_α. If $p^*<1$, then there is some compact set $K\subset I^n$ with $P^n(K)>0$, disjoint from Y_α^n for all α. But then $K=K_\alpha$ for some α, and by choice of $y_{\alpha\beta}$ for $\beta<0$, K_α intersects Y_α^n, a contradiction. So $p^*=1$ and, by definition of $\tilde{C}$, with outer probability 1 it shatters $\{x_1,\dots,x_n\}$ and by a theorem of Talagrand (Theorem 2 in Section 2 above) (taking n even, and $\alpha=\frac{1}{3}$, $\beta=\frac{2}{3}$), $\tilde{C}$ is not a Glivenko–Cantelli class for P, finishing the proof. □

4. UNIFORM GLIVENKO–CANTELLI CLASSES

In analogy with universal Glivenko–Cantelli classes, a collection $\tilde{F}$ of universally measurable functions on $(X, \tilde{A})$ will be called a *weak uniform*

Glivenko–Cantelli class if there exist some real-valued $a_n(f, P)$, $f \in \tilde{F}$, $n = 1, 2, \ldots$, $P \in \tilde{P}(X, \tilde{A})$, such that for all $\varepsilon > 0$,

$$\lim_{n \to \infty} \sup_{P \in \tilde{P}(X, \tilde{A})} \Pr^*\{\|P_n - a_n(\cdot, P)\|_{\tilde{F}} > \varepsilon\} = 0$$

Since weak uniform Glivenko–Cantelli classes are weak universal, it follows from Proposition 4 that each function f in such a class is bounded and that these classes are uniformly bounded up to additive constants. Given a collection $\tilde{F}$ of bounded functions, we will write $F_0 = \tilde{F}$ if $\tilde{F}$ is bounded and $\tilde{F}_0 = \{f - \inf f : f \in \tilde{F}\}$ otherwise, and $F = \sup\{|f| : f \in \tilde{F}_0\}$.

Proposition 10. The following are equivalent:

(a) $\tilde{F}$ is a weak uniform Glivenko–Cantelli class.

(b) $\tilde{F}$ is a class of bounded universally measurable functions and for all $\varepsilon > 0$,

$$\lim_{n \to \infty} \sup_{P \in \tilde{P}(X, \tilde{A})} \Pr^*\{\|P_n - P\|_{\tilde{F}} > \varepsilon\} = 0$$

(c) $\tilde{F}$ is a class of bounded universally measurable functions and for all $r > 0$ (or, for some $r > 0$)

$$\lim_{n \to \infty} \sup_{P \in \tilde{P}(X, \tilde{A})} E^* \|P_n - P\|_{\tilde{F}}^r = 0$$

If any of (a), (b), or (c) hold, $\tilde{F}_0$ is uniformly bounded.

Proof. Obviously (c) $\Rightarrow$ (b) $\Rightarrow$ (a). In order to prove (a) $\Rightarrow$ (b) it suffices to show

$$\lim_{n \to \infty} \sup_P \sup_{f \in \tilde{F}} |Pf - a_n(f, P)| = 0$$

The class $\tilde{F}$ is uniformly bounded up to additive constants by Proposition 4, therefore, $\tilde{F}_0$ is uniformly bounded. Then, with $M = \|F\|_\infty < \infty$, we have $\Pr\{|P_n f - Pf| > \varepsilon\} \leqslant 4M^2/(n\varepsilon^2)$ for all $n = 1, 2, \ldots$ and $f \in \tilde{F}$, by Chebyshev's inequality. By definition, (a) implies

$$\lim_{n \to \infty} \sup_P \sup_{f \in \tilde{F}} \Pr\{|P_n f - a_n(f, P)| > \varepsilon\} = 0$$

Thus for all $\varepsilon > 0$ there is $n_\varepsilon < \infty$ such that for all $n > n_\varepsilon$, $f \in \tilde{F}$, and $P \in \tilde{P}(X, \tilde{A})$ there exists ω for which $|P_n(\omega)(f) - Pf| \leqslant \varepsilon$ and $|P_n(\omega)(f) - a_n(f, P)| \leqslant \varepsilon$, and the conclusion follows. (b) $\Rightarrow$ (c) follows easily since, in both, $\tilde{F}$ can be replaced by $\tilde{F}_0$, which is uniformly bounded since (b) implies (a); for any real function g, $\Pr^*(g>\varepsilon)=\Pr(g^*>\varepsilon)$ [Ref. 5, Lemma 3.1.6(a)], and for a uniformly bounded sequence of measurable functions, convergence in probability is equivalent to convergence in rth moment for any $r>0$, which holds in this case uniformly in P. □

We call $\tilde{F}$ a *strong uniform Glivenko–Cantelli class* if for all $\varepsilon>0$,

$$\lim_{n\to\infty} \sup_{P\in\tilde{P}(X,\tilde{A})} \Pr^*\{\sup_{m\geqslant n} \|P_m-P\|_{\tilde{F}}>\varepsilon\}=0$$

We will show that the weak and strong uniform Glivenko–Cantelli properties are equivalent, under measurability hypotheses, to a condition on the size of $\tilde{F}_0$ for certain metrics (which are nonrandom). These are different from, but related to, some metrics used by Vapnik and Červonenkis,[30] Kolčinskii,[14] and Pollard[23]; see also Giné and Zinn.[10] The measurability that the proof will require is that $\tilde{F}_0$ be nearly linearly supremum measurable for all $P\in\tilde{P}(X,\tilde{A})$ (e.g., Giné and Zinn[10]), a condition which is satisfied if $\tilde{F}$ is image admissible Suslin (Ref. 5, p. 101). For $x=(x_1,\dots,x_n)\in X^n$, $n=1,2,\dots$, and $p\in(0,\infty)$ we define on $\tilde{F}_0$ the pseudodistances

$$e_{x,p}(f,g)=\left[n^{-1}\sum_{i=1}^{n}|f(x_i)-g(x_i)|^p\right]^{p^{-1}\wedge 1}$$

$$e_{x,\infty}(f,g)=\max_{i\leqslant n}|f(x_i)-g(x_i)|,\qquad f,\ g\in\tilde{F}_0$$

Let $N(\varepsilon,\tilde{F}_0,e_{x,p})$ denote the ε-covering number of $(\tilde{F}_0,e_{x,p})$, $\varepsilon>0$. Then, we define, for $n=1,2,\dots$, $\varepsilon>0$ and $p\in(0,\infty]$, the quantities

$$H_{n,p}(\varepsilon,\tilde{F}_0)=\sup_{x\in X^n}\log N(\varepsilon,\tilde{F}_0,e_{x,p})$$

Theorem 6. Let $\tilde{F}$ be a family of bounded functions on $(X,\tilde{A})$ such that $\tilde{F}_0$ is image admissible Suslin. Then the following are equivalent:

(a) $\tilde{F}$ is a weak uniform Glivenko–Cantelli class.

(b) $\tilde{F}$ is a strong uniform Glivenko–Cantelli class.

For $0<p\leqslant\infty$,

(c_p) $\quad\lim_{n\to\infty} H_{n,p}(\varepsilon,\tilde{F}_0)/n=0\qquad$ for all $\varepsilon>0$

and $\tilde{F}_0$ is uniformly bounded.

Proof. We show first that (c_1) implies (b). Let $\{\varepsilon_i\}$ be a Rademacher sequence independent of $\{X_i\}$. (We assume $(\Omega, \Sigma, \Pr) = (X, \tilde{A}, P)^{\mathbb{N}} \otimes ([0, 1], \tilde{B}, \lambda)$, the X_i being the coordinates on $(X, \tilde{A}, P)^{\mathbb{N}}$ and $\varepsilon_i(\{X_j\}, u)$ depending only on $u \in [0, 1]$.) We denote by E_P integration with respect to $P^{\mathbb{N}}$, and by $\Pr_\varepsilon$ probability conditioned on $\{X_i\}$, i.e., λ-probability. By symmetrization (e.g., Ref. 10, p. 936f) and boundedness of $\tilde{F}_0$, we have that for all $\varepsilon > 0$ and for all n large enough (depending on ε)

$$\Pr\{\|P_n - P\|_{\tilde{F}} > \varepsilon\} \leqslant 4 \Pr\left\{\left\|\sum_{i=1}^{n} \varepsilon_i \delta_{X_i}/n\right\|_{\tilde{F}_0} > \varepsilon/4\right\}$$

For $n = 1, 2, \ldots$, let $x_n(\omega) = (X_1(\omega), \ldots, X_n(\omega))$. By definition of $N(\varepsilon, \tilde{F}_0, e_{x,1})$, for each ω there is a map $\pi_n = \pi_n^\omega\colon \tilde{F}_0 \mapsto \tilde{F}_0$ such that $\operatorname{card}\{\pi_n f\colon f \in \tilde{F}_0\} = N(\varepsilon/8, \tilde{F}_0, e_{x_n(\omega),1})$ and

$$e_{x_n(\omega),1}(f, \pi_n f) \leqslant \varepsilon/8, f \in \tilde{F}_0$$

Let $M = \|F\|_\infty < \infty$. Then, by Hoeffding's inequality,

$$\Pr\left\{\left\|\sum_{i=1}^{n} \varepsilon_i \delta_{X_i}/n\right\|_{\tilde{F}_0} > \varepsilon/4\right\} \leqslant E_P \Pr_\varepsilon\left\{\left\|\sum_{i=1}^{n} \varepsilon_i \pi_n f(X_i)/n\right\|_{\tilde{F}_0} > \varepsilon/8\right\}$$
$$\leqslant 2(EN(\varepsilon/8, \tilde{F}_0, e_{x_n(\omega),1})) \exp\{-\varepsilon^2 n/(128 M^2)\}$$

where measurability follows from the image admissible Suslin condition. By (c_1), for all n large enough $N(\varepsilon/8, \tilde{F}_0, e_{x,1}) \leqslant \exp\{\varepsilon^2 n/(256 M^2)\}$ for all $x \in X^n$. Hence, for all $\varepsilon > 0$ there is n_ε such that if $n > n_\varepsilon$ then for some $C = C(M, \varepsilon) = 8(1 - \exp\{-\varepsilon^2/(256 M^2)\})^{-1} < \infty$,

$$\sup_{P \in \tilde{P}(X, \tilde{A})} \sum_{k \geqslant n} \Pr\{\|P_k - P\|_{\tilde{F}} > \varepsilon\} \leqslant C \exp\{-\varepsilon^2 n/(256 M^2)\}$$

and (b) is proved.

Next we show that (a) implies (c_2). Here $M = \|F\|_\infty < \infty$ by Proposition 4. Symmetrizing,

$$2E\|P_n - P\|_{\tilde{F}} \geqslant E\left\|\sum_{j=1}^{n} \varepsilon_j(\delta_{X_j} - P)/n\right\|_{\tilde{F}_0} \geqslant E\left\|\sum_{j=1}^{n} \varepsilon_j \delta_{X_j}/n\right\|_{\tilde{F}_0} - ME\left|\sum_{j=1}^{n} \varepsilon_j/n\right|$$

So, if $\tilde{F}$ is a weak uniform Glivenko–Cantelli class and $\tilde{F}_0$ is image admissible Suslin, Proposition 10 gives that

$$\lim_{n \to \infty} \sup_{P \in \tilde{P}(X, \tilde{A})} E\left\|\sum_{j=1}^{n} \varepsilon_j \delta_{X_j}/n\right\|_{\tilde{F}_0} = 0$$

Then, by Lemma 2.9 in Giné and Zinn,[10] we also have

$$\lim_{n\to\infty} \sup_{P\in \tilde{P}(X,\tilde{A})} E\left\| \sum_{j=1}^{n} g_j \delta_{X_j}/n \right\|_{\tilde{F}_0} = 0$$

where $\{g_i\}$ is an orthogaussian sequence independent of $\{X_i\}$ (we take $\{g_i\}$ defined on $([0,1], \tilde{B}, \lambda)$, like $\{\varepsilon_i\}$). Let now $x = (x_1,..., x_n) \in X^n$, $P_x = n^{-1} \sum_{1\leqslant i \leqslant n} \delta_{x(i)}$ where $x(i) := x_i$. For $j = 1,..., n$, let $m(j)(\cdot)$ be i.i.d. random variables uniformly distributed over $\{1,..., n\}$, independent of $\{g_i\}$ (and not necessarily defined on the original probability space Ω). Then $x_{m(j)}$ are i.i.d. (P_x) random variables independent of $\{g_i\}$. Specializing the last displayed equation to P_x gives

$$\lim_{n\to\infty} \sup_{x\in X^n} E\left\| \sum_{j=1}^{n} g_j \delta_{x_{m(j)}}/n \right\|_{\tilde{F}_0} = 0$$

Now we claim that

$$E\left\| \sum_{i=1}^{n} g_i \delta_{x_i}/n \right\|_{\tilde{F}_0} \leqslant (1-e^{-1})^{-1} E\left\| \sum_{j=1}^{n} g_j \delta_{x_{m(j)}}/n \right\|_{\tilde{F}_0}$$

To prove this inequality we will proceed as in the proof of an inequality in Giné and Zinn,[12] Proposition 2.2, which uses symmetrization arguments due to Pisier,[21] Proposition 5.1. Let $A_{ij} = \{m(j) = i\}$, $i, j = 1,..., n$. Then for each $j \leqslant n$ the sets $A_{1j},..., A_{nj}$ form a disjoint partition of a probability space, the sets in different partitions are independent, and $\Pr(A_{ij}) = 1/n$, i, $j = 1,..., n$. Let g_{ij} for $i, j = 1,..., n$ be i.i.d. $N(0, 1)$ random variables. By disjointness and independence, the two sets of random vectors

$$\{(g_j 1_{A_{1j}},..., g_j 1_{A_{nj}})\}_{j=1}^{n} \qquad \text{and} \qquad \{(g_{1j} 1_{A_{1j}},..., g_{nj} 1_{A_{nj}})\}_{j=1}^{n}$$

have the same distribution. Therefore,

$$E\left\| \sum_{j=1}^{n} g_j \delta_{x_{m(j)}}/n \right\|_{\tilde{F}_0} = E\left\| \sum_{j=1} g_j \left(\sum_{i=1}^{n} 1_{A_{ij}} \delta_{x_i} \right) \Big/ n \right\|_{\tilde{F}_0} = E\left\| \sum_{i,j=1}^{n} g_{ij} 1_{A_{ij}} \delta_{x_i}/n \right\|_{\tilde{F}_0}$$

Conditionally on the events $\{A_{ij}\}$, the random variables $\sum_{j=1}^{n} g_{ij} 1_{A_{ij}}$, $i = 1,..., n$, are independent $N(0, \sum_{j=1}^{n} 1_{A_{ij}})$. Hence, using the Tonelli–Fubini theorem, Jensen's inequality, and the fact that $E(\sum_{j=1}^{n} 1_{A_{ij}})^{1/2} \geqslant 1 - e^{-1}$, we obtain $E\|\sum_{i,j=1}^{n} g_{ij} 1_{A_{ij}} \delta_{x_i}/n\|_{\tilde{F}_0} = E\|\sum_{i=1}^{n} (\sum_{j=1}^{n} 1_{A_{ij}})^{1/2} g_i \delta_{x_i}/n\|_{\tilde{F}_0} \geqslant (1-e^{-1}) E\|\sum_{i=1}^{n} g_i \delta_{x_i}/n\|_{\tilde{F}_0}$, and the claim is proved. Consequently,

$$\lim_{n\to\infty} \sup_{x\in X^n} E\left\| \sum_{i=1}^{n} g_i \delta_{x_i}/n \right\|_{\tilde{F}_0} = 0$$

By Sudakov's[27] inequality (e.g., Ref. 8, p. 26), there is a universal constant C such that

$$E\left\|\sum_{i=1}^{n} g_i \delta_{x_i}/n\right\|_{\tilde{F}_0} \geqslant C \sup_{\varepsilon>0} \varepsilon(\log N(\varepsilon, \tilde{F}_0, e_{x,2}))^{1/2}/n^{1/2}$$

(c_2) follows from this inequality and the previous limit.

By uniform boundedness of $\tilde{F}_0$, for $0<p<q<\infty$ we have

$$H_{n,p}(\varepsilon^{(p\wedge 1)/(q\wedge 1)}, \tilde{F}_0) \leqslant H_{n,q}(\varepsilon, \tilde{F}_0)$$
$$\leqslant H_{n,p}(\varepsilon^{(q\wedge 1)/(p\wedge 1)}/(2M)^{(q-p)/(p\vee 1)}, \tilde{F}_0)$$

and therefore conditions (c_p), $0<p<\infty$, are all equivalent.

Obviously (c_∞) implies (c_p) for all p. So, to finish the proof of the theorem it only remains to be shown that for example, (c_1) implies (c_∞). For this we follow an argument of Talagrand (Ref. 28, p. 982). Let $x=(x_1,..., x_n)\in X^n$ and $0<\alpha<\varepsilon<\frac{1}{2}$. Let $\pi: \tilde{F}_0 \mapsto \tilde{F}_0$ satisfy $e_{x,1}(f, \pi f) \leqslant \alpha\varepsilon/2$ for all $f\in\tilde{F}_0$ and $\text{card}\{\pi f: f\in\tilde{F}_0\} = N(\alpha\varepsilon/2, \tilde{F}_0, e_{x,1})$, and let $\tilde{G}=\{f-\pi f: f\in\tilde{F}_0\}$. If f and g are in the same π-equivalence class of $\tilde{F}_0$, then $\pi f=\pi g$ and therefore the $e_{x,\infty}$ ε-covering number of each of these $N(\alpha\varepsilon/2, \tilde{F}_0, e_{x,1})$ equivalence classes is not larger than $N(\varepsilon, \tilde{G}, e_{x,\infty})$. So,

$$N(\varepsilon, \tilde{F}_0, e_{x,\infty}) \leqslant N(\alpha\varepsilon/2, \tilde{F}_0, e_{x,1})\, N(\varepsilon, \tilde{G}, e_{x,\infty})$$

If $g\in\tilde{G}$, then $\sum_{i=1}^n |g(x_i)| \leqslant \alpha\varepsilon n/2$. Hence there are at most $m=[\alpha n]$ i's for which $|g(x_i)|>\varepsilon/2$. Let $\tilde{H}$ be the set of functions on $\{x_1,..., x_n\}$ which are zero at $n-m$ of these points and take values of the form $k\varepsilon/2$, $k\in\mathbb{Z}$, $|k|\leqslant 4M/\varepsilon$, at the remaining m points. Then, $\min\{\max_i |(g-h)(x_i)|: h\in\tilde{H}\} \leqslant \varepsilon/2$, and we obtain

$$N(\varepsilon, \tilde{G}, e_{x,\infty}) \leqslant \binom{n}{m}\left(1+\frac{8M}{\varepsilon}\right)^m$$

(note that $\tilde{H}$ is not included in $\tilde{G}$, but we can replace each $h\in\tilde{H}$ such that the $\varepsilon/2$-ball centered at h, $B(h, \varepsilon/2)$, intersects $\tilde{G}$ so that $B(h, \varepsilon/2)\subset B(g, \varepsilon)$). From this, (c_1) and Stirling's formula yield

$$\limsup_{n\to\infty} H_{n,\infty}(\varepsilon, \tilde{F}_0)/n \leqslant \alpha\,|\log\alpha| + (1-\alpha)\,|\log(1-\alpha)| + \alpha\log(1+8M/\varepsilon)$$

Letting $\alpha\to 0$ we get (c_∞). □

Remark. Let $y(1), y(2),...,$ be distinct points and $f_k := k1_{\{y(k)\}}$. Let $\tilde{F}=\{f_k\}_{k\geqslant 1}$. Then $H_{n,p}(\varepsilon, \tilde{F}_0) \equiv \log(n+1)$, so (c_p) holds for all $p>0$ except that $\tilde{F}_0$ is not uniformly bounded.

We do not know if the weak and the strong uniform Glivenko–Cantelli properties are equivalent without any measurability assumptions on $\tilde{F}$. But there are Vapnik–Červonenkis classes $\tilde{C}$ (with bad measurability properties) of index $S(\tilde{C})=1$, and so by Sauer's lemma (e.g., Ref. 5, Theorem 9.1.2) satisfying

$$H_{n,p}(\varepsilon, \tilde{C}) \leqslant \log(n+1), \quad 0<p\leqslant\infty$$

which are not universal Glivenko–Cantelli. One example (Ref. 5, 10.3, pp. 103–104) is the class $\tilde{C}$ of all initial segments of the least uncountable ordinal $\mathscr{H}_1$. Then for every probability measure P on $\mathscr{H}_1$ which is zero on countable sets, $\|P_n-P\|_{\tilde{C}}\equiv 1$. Assuming the continuum hypothesis, this happens simultaneously for every nonatomic Borel probability measure on an uncountable Polish space S, using a well-ordering of S. Also Dobrić (Ref. 4, Theorem 3.1), shows, given any particular nonatomic law P on a Polish space (without the continuum hypothesis), that there is a class $\tilde{C}$ of disjoint countable sets, so that again $S(\tilde{C})=1$, such that $P^{\infty *}(\|P_n-P\|_{\tilde{C}}=1 \text{ for all } n)=1$.

Given a class $\tilde{C}$ of subsets of X, for $x_1,..., x_n \in X$ let $\Delta^{\tilde{C}}(x_1,..., x_n)$ be the cardinality of the collection of sets of the form $\{x_1,...,x_n\} \cap C$, $C\in\tilde{C}$. Then, if $x=(x_1,..., x_n)$, $\Delta^{\tilde{C}}(x_1,..., x_n)=N(\varepsilon, \tilde{C}, e_{x,\infty})$ for all $\varepsilon\in(0,1)$. Hence, if condition (c_∞) holds for $\tilde{C}$, then for some n, $\sup\{\Delta^{\tilde{C}}(x): x\in X^n\}<2^n$ and $\tilde{C}$ is a Vapnik–Červonenkis class. Conversely, if $\tilde{C}$ is a Vapnik–Červonenkis class, then Vapnik and Červonenkis[29] proved that $H_{n,\infty}(\varepsilon, \tilde{C})\leqslant c\log n$ for some $c<\infty$ and all $n\in\mathbb{N}$, so a condition much stronger than (c_∞) holds for $\tilde{C}$. Thus, we have the following:

Proposition 11. Let $\tilde{C}$ be an image admissible Suslin class of sets. Then $\tilde{C}$ is uniform Glivenko–Cantelli if and only if $\tilde{C}$ is Vapnik–Červonenkis.

This proposition is not new: Vapnik and Červonenkis[29] prove, at least implicitly, that if $\tilde{C}$ is Vapnik–Červonenkis (and satisfies measurability conditions which follow from image admissible Suslin) then $\tilde{C}$ is uniform Glivenko–Cantelli at rate $1/n^\gamma$ for any $\gamma<\frac{1}{2}$; see also Pisier.[22] Moreover, in the other direction, Assouad and Dudley[1] prove the stronger result that if $\tilde{Q}$ is the class of all laws with finite support and $\sup_{P\in\tilde{Q}} E^P\|P_n-P\|_{\tilde{C}}<\frac{1}{2}$ for some n (and so a fortiori if $\tilde{C}$ is uniform Glivenko–Cantelli) then $\tilde{C}$ is Vapnik–Červonenkis.

The universal Glivenko–Cantelli class of all finite subsets (or all subsets) of a countably infinite set is not Vapnik–Červonenkis and hence not uniform Glivenko–Cantelli by Proposition 11. To define other examples of such classes of functions, for any metric space (S, d) and real-valued function f on S, let

$$\|f\|_{BL} := \sup_{x} |f(x)| + \sup_{x \neq y} |f(x) - f(y)|/d(x, y)$$

Then we have the following:

Proposition 12. For any separable metric space (S, d) and $0 < M < \infty$, $\tilde{F}_M := \{f\colon \|f\|_{BL} \leqslant M\}$ is a universal Glivenko–Cantelli class. It is a uniform Glivenko–Cantelli class if and only if (S, d) is totally bounded.

Proof. The universal Glivenko–Cantelli property holds since almost surely $P_n \to P$ for convergence of laws (Ref. 31), which is metrized by $\beta(P, Q) := \sup\{|\int f d(P-Q)|\colon \|f\|_{BL} \leqslant 1\}$ (e.g., Ref. 7, Theorems 11.4.1 and 11.3.3).

If (S, d) is not totally bounded, then for some $\delta > 0$ and infinite $A \subset S$, $d(x, y) > \delta$ for all $x \neq y$ in A. For any subset $B \subset A$, there is some f_B with $f = \delta$ on B, $f = 0$ on $A \backslash B$, and $\|f\|_{BL} \leqslant 1 + \delta$ (Ref. 7, Proposition 11.2.3). It follows from Theorem 6 that $\tilde{F}_{1+\delta}$ and so (by constant multiples) any $\tilde{F}_M$ with $M > 0$ is not a uniform Glivenko–Cantelli class.

If (S, d) is totally bounded, then for $M < \infty$, $\tilde{F}_M$ is totally bounded for the supremum norm. So for any $\varepsilon > 0$, $H_{n,\infty}(\varepsilon, (\tilde{F}_M)_0)$ is bounded uniformly in n, and $\tilde{F}_M$ is uniform Glivenko–Cantelli by Theorem 6. □

The non i.i.d. version of the Glivenko–Cantelli theorem also holds uniformly in uniform Glivenko–Cantelli classes. This is the subject of the next theorem. Let Q_i be probability measures on $(X, \tilde{A})$, let X_i be the coordinates of $\Pi_i(X, \tilde{A}, Q_i)$ and let P_n be the empirical measure of $X_1, \ldots, X_n$, $n = 1, 2, \ldots$. Then, we have the following:

Theorem 7. Let $\tilde{F}$ be a class of bounded functions such that $\tilde{F}_0$ is image admissible Suslin. Then the following are equivalent:

(i) $\tilde{F}$ is a (weak or strong) uniform Glivenko–Cantelli class.

(ii) $\lim_{n \to \infty} \sup\{E\|P_n - n^{-1} \sum_{i=1}^{n} Q_i\|_{\tilde{F}}\colon (Q_i) \in \tilde{P}(X, \tilde{A})^{\mathbb{N}}\} = 0$

(iii) $\lim_{n \to \infty} \sup\{\Pr\{\sup_{m \geqslant n} \|P_m - m^{-1} \sum_{i=1}^{m} Q_i\|_{\tilde{F}} > \varepsilon\}\colon (Q_i) \in \tilde{P}(X, \tilde{A})^{\mathbb{N}}\} = 0$ for all $\varepsilon > 0$.

Proof. The proof of $(c_1) \Rightarrow$ (b) in Theorem 6 does not make use of the equidistribution of the variables X_i and it applies verbatim to show that $(c_1) \Rightarrow$ (iii). Therefore, by Theorem 6, (i) $\Rightarrow$ (iii). (iii) implies (b) in Proposition 10, which implies $\tilde{F}_0$ is uniformly bounded, so (iii) $\Rightarrow$ (ii) by the proof of (b) ⇒ (c) in Proposition 10, which also extends, with its cited references, to the non i.i.d. case. And obviously (ii) ⇒ (i) by taking $Q_i = P$, $i = 1, 2,...$. □

Finally, here are some permanence properties of uniform Glivenko–Cantelli classes. It follows from the definition that if $\tilde{F}$ and $\tilde{G}$ are uniform Glivenko–Cantelli then so are $a\tilde{F} + b\tilde{G}$, a, $b \in \mathbb{R}$, and $\tilde{F} \cup \tilde{G}$. Condition (c_∞) in Theorem 6 can be used to prove the following: Let $p \in (0, \infty)$ and let $\tilde{F}$ be a bounded image admissible Suslin uniform Glivenko–Cantelli class. Then $\tilde{F}^p = \{|f|^p \operatorname{sgn}(f): f \in \tilde{F}\}$ is a uniform Glivenko–Cantelli class [this follows by comparison of $|f^p(s) - g^p(s)|$ with $c\,|f(s) - g(s)|^{p \wedge 1}$ for some $c < \infty$. The same comment applies to $|\tilde{F}|^p = \{|f|^p: f \in \tilde{F}\}$.

Let now B be a separable Banach space, $P \in \tilde{P}(B)$ and X_i i.i.d. (P). Mourier[19] showed that if $\int \|x\| \, dP(x) < \infty$ then $\|\sum_{i=1}^n X_i/n - \int x \, dP(x)\| \to 0$ a.s. In particular, letting $X = B_1$, the unit ball of B,‘ and $\tilde{F} = B'_1$, the unit ball of the dual B' of B, $\tilde{F}$ is universal Glivenko–Cantelli. The example of all the finite subsets of X countable (which, for the law of large numbers, is equivalent to the power set example) is of this type: Just take $X = \{e_k\}_{k=1}^\infty \subset l_1$, the canonical basis of l_1, and $\tilde{F}$ equal to the set of extreme points in the unit ball of c_0 with nonnegative coordinates. We see next that in the context of separable Banach spaces this is in some sense the only example of universal but not uniform Glivenko–Cantelli classes.

A Banach space B will be called *uniform for the law of large numbers* if

$$\lim_{n \to \infty} \sup_{P \in \tilde{P}(B_1)} E \left\| \sum_{i=1}^n (X_i^P - EX_1^P)/n \right\| = 0$$

where $\{X_i^P\}$ is an i.i.d. (P) sequence of random variables. The relevant geometric concept for this property is *B-convexity* (Ref. 3; see, e.g., Woyczyński[32]). B is called B-convex if there exist $k \in \mathbb{N}$ and $\varepsilon \in (0, 1)$ such that

$$\inf_{\varepsilon_i = \pm 1} \left\| \sum_{i=1}^k \varepsilon_i x_i \right\| < k(1 - \varepsilon)$$

for any $x_i \in B$ with $\|x_i\| = 1$, $i = 1,..., k$. l_1 is said to be finitely representable in a Banach space B if for every $n = 1, 2,...,$ there exist a linear map I_n: $l_1^n \mapsto B$ such that $\|x\|/2 \leqslant \|I_n x\| \leqslant 2\|x\|$, $x \in l_1^n$ (where l_1^n is $\mathbb{R}^n$ with the l_1-norm $\|x\| = \sum_{i=1}^n |x_i|$). An important property of B-convex spaces is the following (Ref. 9; see, e.g., Woyczyński[32]): B is B-convex if and only if l_1 is not finitely representable in B. A second important property is that B is B-convex if and only if B is of Rademacher type p for some $p \in (1, 2]$

(Ref. 20; see, e.g., Woyczyński[32]). As mentioned in Section 2, B-convex spaces are also characterized by the weak law of large numbers. Here is another characterization of B-convexity based on the law of large numbers.

Theorem 8. A separable Banach space B is uniform for the law of large numbers if and only if it is B-convex.

Proof. The "if" part of this theorem is known (see Woyczyński[32]); we give a proof of both parts, based on the above properties, for completeness. If B is B-convex, then it is of type p for some $p \in (1, 2]$, and therefore we have for some fixed constant c,

$$E\left\|\sum_{i=1}^{n} (X_i^P - EX_1^P)/n\right\|^p \leqslant cE\|X_1^P\|^p/n^{p-1} \to 0 \text{ uniformly in } P \in \tilde{P}(B_1)$$

Conversely, if B is not B-convex them l_1 is finitely representable in B. Let, for $N = 1, 2,..., e_{iN} = I_N e^i$, $i = 1,..., N$, where $\{e_i\}_{i=1}^N$ is the canonical basis of l_1^N and I_N is an operator as in the definition of finite representability. Let $Q_N = N^{-1} \sum_{i=1}^N \delta_{e_{iN}}$ and let $\{X_i^N\}_{i=1}^N$ be i.i.d. (Q_N) random variables. Then, the left-hand side part of Proposition 2.2 in Giné and Zinn[12] gives

$$E\left\|\sum_{i=1}^{N} \varepsilon_i X_i^N/N\right\| \geqslant 2^{-1}(1-e^{-1})E\left\|\sum_{i=1}^{N} \varepsilon_i e_{iN}/N\right\|$$
$$\geqslant 4^{-1}(1-e^{-1})E\left\|\sum_{i=1}^{N} \varepsilon_i e_i/N\right\| = 4^{-1}(1-e^{-1})$$

where $\{\varepsilon_i\}$ is a Rademacher sequence independent of $\{X_i\}$. Since

$$2E\left\|\sum_{i=1}^{N} (X_i^N - EX_1^N)/N\right\| \geqslant E\left\|\sum_{i=1}^{N} \varepsilon_i X_i^N/N\right\| - 2E\left|\sum_{i=1}^{N} \varepsilon_i/N\right|$$

it follows that

$$\limsup_{N\to\infty} \sup_{P\in\tilde{P}(B_1)} E\left\|\sum_{i=1}^{N} (X_i^P - EX_1^P)/N\right\| \geqslant 8^{-1}(1-e^{-1}) > 0$$

that is, B is not uniform for the law of large numbers. □

Theorems 6 and 7 give other different necessary and sufficient conditions for B to be uniform for the law of large numbers. Another equivalent condition is: Whenever a family of laws $\tilde{H} \subset \tilde{P}(B)$ satisfies the uniform integrability condition $\lim_{M\to\infty} \sup_{P\in\tilde{H}} \int \|x\| 1_{\{\|x\|>M\}} dP(x) = 0$, then $\tilde{H}$ also satisfies $\lim_{n\to\infty} \sup_{P\in\tilde{H}} E\|\sum_{i=1}^n (X_i^P - EX_1^P)/n\| = 0$. This condition on B clearly implies that B is uniform for the law of large numbers. On the other hand, if B is uniform for the law of large numbers, this property follows by an easy truncation argument.

Example 4. Dudley[6] and Giné and Zinn[11] prove that if $\{A(i)\}$ is a partition of X then the class $\tilde{F}=\{\sum_{i=1}^{\infty} x_i 1_{A(i)}: \sum_i x_i^2 \leqslant 1\}$ is universal Donsker but not uniform Donsker. Let B' be the dual of a B-convex sequence space, e.g., $B'=l_p$, $1<p<\infty$, and let B'_1 be its unit ball. Then Theorem 7 immediately gives that $\tilde{F}=\{\sum_{i=1}^{\infty} x_i 1_{A(i)}: (x_i)\in B'_1\}$ is a uniform Glivenko–Cantelli class: To each P on X assign $\bar{P}$ on B_1 by the equation $\bar{P}\{e_i\}=PA(i)$, to obtain that, by duality, $\tilde{F}$ is uniform Glivenko–Cantelli if and only if the law of large numbers holds on B uniformly over all the probability measures supported by the canonical basis. Since this last property does not hold for $B=l_1$ (see the proof of Theorem 8), the class of functions $\tilde{F}=\{\sum_{i=1}^{\infty} x_i 1_{A(i)}: |x_i|\leqslant 1\}$ is not uniform Glivenko–Cantelli. On the other hand, $\tilde{F}=\{\sum_{i=1}^{\infty} x_i 1_{A(i)}: |x_i|\leqslant a_i\}$ with $a_i\to 0$ is uniform Glivenko–Cantelli: Assuming without loss of generality $1\geqslant a_i\downarrow 0$, and letting, for each $\varepsilon>0$, $n(\varepsilon)=\max\{i: 2a_i\geqslant\varepsilon\}$ we have for $n=1,2,...$, $H_{n,\infty}(\varepsilon,\tilde{F})\leqslant \log\prod_{i=1}^{n(\varepsilon)}([2a_i/\varepsilon]+1)\leqslant\sum_{i=1}^{n(\varepsilon)} 2a_i/\varepsilon\leqslant 2n(\varepsilon)/\varepsilon$; therefore $H_{n,\infty}(\varepsilon,\tilde{F})/n\to 0$ for all $\varepsilon>0$ and Theorem 6 gives the result.

Example 5. Let $\tilde{F}=\{f_m\}_{m=1}^{\infty}$. It is proved in Giné and Zinn[11] that $\tilde{F}$ is uniform Donsker if $\|f_m\|_\infty=o(1/(\log m)^{1/2}$ (and there are classes $\tilde{F}$ with $\|f_m\|_\infty=O(1/(\log m)^{1/2}$ which are not even pregaussian for some P, e.g., $f_m=\varepsilon_m/(\log m)^{1/2}$, $m\geqslant 2$, where ε_m are Rademacher for P). If $\|f_m\|_\infty\to 0$ (and $\|f_m\|_\infty<\infty$ for all m) then $\tilde{F}$ is uniform Glivenko–Cantelli by Theorem 6, since either (a) or (c_∞) there is easy to check. On the other hand there is a class $\tilde{F}=\{f_m\}$ with $\|f_m\|_\infty\equiv 1$ which is not Glivenko–Cantelli for some P (and so not even universal Glivenko–Cantelli): a sequence of indicators of sets A_m independent for P with $P(A_m)=\frac{1}{2}$ for all m.

ACKNOWLEDGMENTS

R. M. Dudley was partially supported by NSF grant DMS-8803305; E. Giné was partially supported by NSF grant DMS-8619411; J. Zinn was partially supported by NSF grant DMS-8601250 and Texas Advanced Research Program Grant 3825.

Some of the questions treated in this paper arose in conversations with Miguel Martin and Joseph Yukich. We are very grateful to W. Adamski for a proof on Lusin sets (Section 3) and to Peter Gaenssler for numerous helpful comments.

REFERENCES

1. Assouad, P., and Dudley, R. M. (1989). Minimax nonparametric estimation over classes of sets. Preprint.
2. Araujo, A., and Giné, E. (1978). On tails and domains of attraction of stable measures in Banach spaces. *Trans. Amer. Math. Soc.* **248**, 105–119.
3. Beck, A. (1962). A convexity condition in Banach spaces and the strong law of large numbers. *Proc. Amer. Math. Soc.* **13**, 329–334.
4. Dobrić, V. (1987). The law of large numbers, examples and counterexamples. *Math. Scand.* **60**, 273–291.
5. Dudley, R. M. (1984). *A Course on Empirical Processes. Ecole d'été de probabilités de St.-Flour, 1982.* Lecture Notes in Mathematics, Vol. 1097, pp. 1–142, Springer, Berlin.
6. Dudley, R. M. (1987). Universal Donsker classes and metric entropy. *Ann. Prob.* **15**, 1306–1326.
7. Dudley, R. M. (1989). *Real Analysis and Probabililty.* Wadsworth & Brooks/Cole, Pacific Grove, CA.
8. Fernique, X. (1974). *Régularité des Trajectoires des Fonctions Aléatoires Gaussiennes.* Lecture Notes in Mathematics, Vol. 480, pp. 1–96, Springer, Berlin.
9. Giesy, D. P., and James, R. C. (1973). Uniformly non-l_1 and B-convex Banach spaces. *Studia Math.* **48**, 61–69.
10. Giné, E., and Zinn, J. (1984). Some limit theorems for empirical processes. *Ann. Prob.* **12**, 929–989.
11. Giné, E., and Zinn, J. (1989). Gaussian characterization of uniform Donsker classes of functions. *Ann. Prob.* **19** (to appear).
12. Giné, E., and Zinn, J. (1990). Bootstrapping general empirical measures. *Ann. Prob.* **18**, 851–869.
13. Hausdorff, F. (1936). Summen von $\mathscr{H}_1$ Mengen. *Fund. Math.* **26**, 241–255.
14. Kolčinskii, V. I. (1981). On the central limit theorem for empirical measures. *Theor. Prob. Math. Statist.* **24**, 71–82.
15. Kolmogorov, A. N. (1929). Bemerkungen zu meiner Arbeit "Über die Summen zufälliger Grössen." *Math. Ann.* **102**, 484–488.
16. Lusin, N. (1914). Sur un problème de M. Baire. *Comptes Rendus Acad. Sci. Paris* **158**, 1258–1261.
17. Mandrekar, V., and Zinn, J. (1980). Central limit problem for symmetric case: convergence to non-Gaussian laws. *Studia Math.* **67**, 279–296.
18. Marcus, M. B., and Woyczynski, W. A. (1979). Stable measures and central limit theorems in spaces of stable type. *Trans. Amer. Math. Soc.* **251**, 71–102.
19. Mourier, E. (1953). Eléments aléatoires dans un espace de Banach. *Ann. Inst. H. Poincaré* **13**, 161–244.
20. Pisier, G. (1973). Sur les espaces de Banach qui ne contiennent pas uniformément de l_n^1. *Comptes Rendus Acad. Sci. Paris Sér. A* **277**, 991–994.
21. Pisier, G. (1975). Le théorème limite central et la loi du logarithme itéré dans les espaces de Banach. *Séminaire Maurey–Schwartz 1973–1974, Exposés III et IV.* Ecole Polytechnique, Paris.
22. Pisier, G. (1984). Remarques sur les classes de Vapnik–Červonenkis. *Ann. Inst. H. Poincaré Prob. Statist.* **20**, 287–298.
23. Pollard, D. (1982). A central limit theorem for empirical processes. *J. Austral. Math. Soc. Sec. A* **33**, 235–248.
24. Shortt, Rae M. (1984). Universally measurable spaces: An invariance theorem and diverse characterizations. *Fund. Math.* **121**, 169–176.

25. Sierpiński, W., and Szpilrajn-Marczewski, E. (1936). Remarque sur le problème de la mesure. *Fund. Math.* **26**, 256–261.
26. Steele, J. Michael (1978). Empirical discrepancies and subadditive processes. *Ann. Prob.* **6**, 118–127.
27. Sudakov, V. N. (1971). Gaussian random processes and measures of solid angles in Hilbert space. *Dokl. Akad. Nauk SSSR* **197**, 43–45 (*Soviet Math. Dokl.* **12**, 412–415).
28. Talagrand, M. (1987). The Glivenko–Cantelli problem. *Ann. Prob.* **15**, 837–870.
29. Vapnik, V. N., and A. Ya. Červonenkis (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theor. Prob. Appl.* **16**, 264–280.
30. Vapnik, V. N., and A. Ya. Červonenkis (1981). Necessary and sufficient conditions for the uniform convergence of means to their expectations. *Theor. Prob. Appl.* **26**, 532–553.
31. Varadarajan, V . S. (1958). On the convergence of sample probability distributions. *Sankhyā* **19**, 23–26.
32. Woyczyński, W. (1978). Geometry and martingales in Banach spaces—Part II: Independent increments. In J. Kuelbs (Ed.), *Probability on Banach Spaces*, pp. 267–521, Marcel Dekker, New York.