THE UNIVERSAL GLIVENKO-CANTELLI PROPERTY

BY RAMON VAN HANDEL*

Princeton University

Let $\mathcal{F}$ be a separable uniformly bounded family of measurable functions on a standard measurable space $(X,\mathcal{X})$, and let $N_{[]}(\mathcal{F},\varepsilon,\mu)$ be the smallest number of ε-brackets in $L^1(\mu)$ needed to cover $\mathcal{F}$. The following are equivalent:

1. $\mathcal{F}$ is a universal Glivenko-Cantelli class.
2. $N_{[]}(\mathcal{F},\varepsilon,\mu)<\infty$ for every $\varepsilon>0$ and every probability measure μ.
3. $\mathcal{F}$ is totally bounded in $L^1(\mu)$ for every probability measure μ.
4. $\mathcal{F}$ does not contain a Boolean σ-independent sequence.

In particular, universal Glivenko-Cantelli classes are uniformity classes for general sequences of almost surely convergent random measures.

1. Main results. Let $(X,\mathcal{X})$ be a measurable space, and let $\mathcal{F}$ be a family of measurable functions on $(X,\mathcal{X})$. Given a probability measure μ on $(X,\mathcal{X})$, the family $\mathcal{F}$ is said to be a *μ-Glivenko-Cantelli class* if

$$\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{k=1}^{n} f(X_k)-\mu(f)\right| \xrightarrow{n\to\infty} 0 \quad \text{a.s.},$$

where $(X_k)_{k\ge1}$ is the i.i.d. sequence of X-valued random variables with distribution μ, defined on its canonical product probability space.[1] The class $\mathcal{F}$ is said to be a *universal Glivenko-Cantelli class* if it is μ-Glivenko-Cantelli for every probability measure μ on $(X,\mathcal{X})$. The goal of this paper is to obtain, under mild regularity assumptions, a precise characterization of universal Glivenko-Cantelli classes. Somewhat surprisingly, we find that universal Glivenko-Cantelli classes are in fact uniformity classes for convergence of (random) probability measures in a very general setting, so that their applicability extends substantially beyond the setting of laws of large numbers for i.i.d. sequences that is inherent in their definition.

*This work was partially supported by NSF award DMS-1005575.

AMS 2000 subject classifications: 60F15, 60B10, 41A46

Keywords and phrases: universal Glivenko-Cantelli classes, uniformity classes, uniform convergence of random measures, entropy with bracketing, Boolean independence

[1] Recall that a sequence of possibly non-measurable real-valued functions $(Z_n)_{n\ge1}$ on a probability space is said to converge a.s. to a function Z if there exists a sequence of nonnegative measurable functions Δ_n converging to zero a.s. such that $|Z_n-Z|\le\Delta_n$ pointwise for all $n\ge1$.

The following probability-free independence properties for families of functions will play a fundamental role in this paper. These notions date back to Marczewski [14] (for sets) and Rosenthal [17] (for functions, see also [8]).

DEFINITION 1.1. A family $\mathcal{F}$ of functions on a set X is said to be *Boolean independent at levels* (α, β) if for every finite subfamily $\{f_1, \ldots, f_n\} \subseteq \mathcal{F}$

$$\bigcap_{j\in F}\{f_j < \alpha\} \cap \bigcap_{j\notin F}\{f_j > \beta\} \neq \varnothing \quad \text{for every } F \subseteq \{1, \ldots, n\}.$$

A sequence $(f_i)_{i\in\mathbb{N}}$ is said to be *Boolean σ-independent at levels* (α, β) if

$$\bigcap_{j\in F}\{f_j < \alpha\} \cap \bigcap_{j\notin F}\{f_j > \beta\} \neq \varnothing \quad \text{for every } F \subseteq \mathbb{N}.$$

A family (sequence) of functions is said to be Boolean (σ-)independent if it is Boolean (σ-)independent at levels (α, β) for some $\alpha < \beta$.

We also recall the well-known notions of bracketing and covering numbers.

DEFINITION 1.2. Let $\mathcal{F}$ be a class of functions on a measurable space $(X, \mathcal{X})$. Given $\varepsilon > 0$ and a probability measure μ on $(X, \mathcal{X})$, a pair of measurable functions f^+, f^- such that $f^- \leq f^+$ pointwise and $\mu(f^+ - f^-) \leq \varepsilon$ defines an *ε-bracket in $L^1(\mu)$* $[f^+, f^-] := \{f : f^- \leq f \leq f^+ \text{ pointwise}\}$. Denote by $N_{[]}(\mathcal{F}, \varepsilon, \mu)$ the cardinality of the smallest collection of ε-brackets in $L^1(\mu)$ covering $\mathcal{F}$, and by $N(\mathcal{F}, \varepsilon, \mu)$ the cardinality of the smallest covering of $\mathcal{F}$ by ε-balls in $L^1(\mu)$.

A class of functions $\mathcal{F}$ on a set X will be said to be *separable*[2] if it contains a countable dense subset for the topology of pointwise convergence in $\mathbb{R}^X$. Recall that a measurable space $(X, \mathcal{X})$ is said to be *standard* if it is Borel-isomorphic to a Polish space. We can now formulate our main result.

THEOREM 1.3. *Let $\mathcal{F}$ be a separable uniformly bounded family of measurable functions on a standard measurable space $(X, \mathcal{X})$. The following are equivalent:*

1. *$\mathcal{F}$ is a universal Glivenko-Cantelli class.*
2. *$N_{[]}(\mathcal{F}, \varepsilon, \mu) < \infty$ for every $\varepsilon > 0$ and every probability measure μ.*
3. *$N(\mathcal{F}, \varepsilon, \mu) < \infty$ for every $\varepsilon > 0$ and every probability measure μ.*
4. *$\mathcal{F}$ contains no Boolean σ-independent sequence.*

[2] Note that the notion of separability used here is a slightly weaker assumption than a common notion of poinwise measurability employed, for example, in [21], Example 2.3.4.

A notable aspect of this result is that the four equivalent conditions of Theorem 1.3 are quite different in nature: roughly speaking, the first condition is probabilistic, the second and third are geometric and the fourth is combinatorial.

The implication $1 \Rightarrow 2$ in Theorem 1.3 is the most important result of this paper. A consequence of this implication is that universal Glivenko-Cantelli classes can be characterized as uniformity classes in a much more general setting, as in the following Corollary. The first condition is due to Topsøe [20], while the remaining implications are straightforward up to measurability issues.

COROLLARY 1.4. *Under the assumptions of Theorem 1.3, the following are equivalent to the equivalent conditions 1–4 of Theorem 1.3:*

5. *For any probability measure μ on $(X, \mathcal{X})$ and net of probability measures $(\mu_\alpha)_{\alpha \in I}$ such that $\mu_\alpha \to \mu$ setwise, we have $\sup_{f \in \mathcal{F}} |\mu_\alpha(f) - \mu(f)| \to 0$.*
6. *For any probability measure μ on $(X, \mathcal{X})$ and sequence of random probability measures (kernels) $(\mu_n)_{n \in \mathbb{N}}$ such that $\mu_n(A) \to \mu(A)$ a.s. for every $A \in \mathcal{X}$, we have $\sup_{f \in \mathcal{F}} |\mu_n(f) - \mu(f)| \to 0$ a.s.*
7. *For any countably generated reverse filtration $(\mathcal{G}_{-n})_{n \in \mathbb{N}}$ and X-valued random variable Z, $\sup_{f \in \mathcal{F}} |\mathbf{P}_{\mathcal{G}_{-n}}(f(Z)) - \mathbf{P}_{\mathcal{G}_{-\infty}}(f(Z))| \to 0$ a.s.*
8. *For any strictly stationary sequence $(Z_n)_{n \in \mathbb{Z}}$ of X-valued random variables, $\sup_{f \in \mathcal{F}} |\frac{1}{n} \sum_{k=1}^{n} f(Z_k) - \mathbf{P}_{\mathcal{I}}(f(Z_0))| \to 0$ a.s. ($\mathcal{I}$ is the invariant σ-field).*

Here $\mathbf{P}_{\mathcal{G}}$ denotes any version of the regular conditional probability $\mathbf{P}[\,\cdot\,|\mathcal{G}]$.

The separability assumption in Theorem 1.3 is crucial: without it, easy counterexamples show that a characterization along the lines of this paper is impossible. For example, consider the class $\mathcal{F}$ consisting of all indicator functions of finite subsets of X. It is clear that this class is not μ-Glivenko-Cantelli for any nonatomic measure μ, yet it is easily seen that condition 3 of Theorem 1.3 holds. On the other hand, [3], section 1.2 gives a simple example of a universal Glivenko-Cantelli class (in fact, a Vapnik-Chervonenkis class that satisfies standard measurability assumptions) for which condition 8 of Corollary 1.4, and therefore condition 2 of Theorem 1.3, are violated. It therefore appears that separability plays a fundamental role here and cannot be replaced by a weaker measurability assumption.

On the other hand, the assumption that $\mathcal{F}$ is uniformly bounded is not a restriction: indeed, it is easily seen that any universal Glivenko-Cantelli class is uniformly bounded up to additive constants (see, for example, [11], Proposition 4). The importance of the assumption that the measurable space $(X, \mathcal{X})$ is standard is less obvious. Its main role in the proof is to ensure that the implication $4 \Rightarrow 2$ holds in the absence of a Boolean σ-independent sequence rather than a Boolean independent sequence. For this purpose it would suffice to restrict attention to quasi-compact measures as defined in [18], but this does not yield a substantial weakening of the assumptions. For the case where $(X, \mathcal{X})$ is a general measurable space we will prove the following quantitative result, which is of independent interest.

DEFINITION 1.5. Let $\gamma > 0$. A family $\mathcal{F}$ of functions on a set X is said to *γ-shatter* a subset $X_0 \subseteq X$ if there exist levels $\alpha < \beta$ with $\beta - \alpha \geq \gamma$ such that, for every finite subset $\{x_1, \ldots, x_n\} \subseteq X_0$, the following holds:

$$\forall\, F \subseteq \{1, \ldots, n\}, \quad \exists\, f \in \mathcal{F} \text{ so that } f(x_j) < \alpha \text{ for } j \in F, \quad f(x_j) > \beta \text{ for } j \notin F.$$

The *γ-dimension* of $\mathcal{F}$ is the maximal cardinality of γ-shattered finite subsets of X.

THEOREM 1.6. *Let $\mathcal{F}$ be a separable uniformly bounded family of measurable functions on a measurable space $(X, \mathcal{X})$, and let $\gamma > 0$. Consider:*

a. $\mathcal{F}$ has finite γ-dimension.
b. No sequence in $\mathcal{F}$ is Boolean independent at levels (α, β) with $\beta - \alpha \geq \gamma$.
c. $N_{[]}(\mathcal{F}, \varepsilon, \mu) < \infty$ for every $\varepsilon > \gamma$ and every probability measure μ.

Then the implications $a \Rightarrow b \Rightarrow c$ hold.

The notion of γ-dimension appears in [4] (called $V_{\gamma/2}$-dimension there). The implication $a \Rightarrow c$ of Theorem 1.6 contains the recent results of Adams and Nobel [1, 2, 3]. Let us note that condition b is strictly weaker than condition a: for example, the class $\mathcal{F} = \{\mathbf{1}_C : C \text{ is a finite subset of } \mathbb{N}\}$ has infinite γ-dimension for $\gamma < 1$, but does not contain a Boolean independent sequence. Condition b is dual (in the sense of Assouad [6]) to the nonexistence of a γ-shattered sequence in X. A connection between the latter and the universal Glivenko-Cantelli property for families of indicators is considered by Dudley, Giné and Zinn [11].

An interesting question arising from Theorem 1.6 is as follows. It is known, see [4] and [15], that if $\mathcal{F}$ has finite γ-dimension for all $\gamma > 0$ then $\sup_\mu N(\mathcal{F}, \gamma, \mu) < \infty$ for all $\gamma > 0$, that is, the covering numbers of the class $\mathcal{F}$ are bounded uniformly with respect to the underlying probability measure. If in addition $\mathcal{F}$ is a family of indicators, then we even have the polynomial bound $\sup_\mu N(\mathcal{F}, \varepsilon, \mu) \lesssim \varepsilon^{-d}$ for some constant $d > 0$. In view of Theorem 1.6, one might expect that one can similarly obtain quantitative bounds on the bracketing numbers. Unfortunately, this is not the case: $N_{[]}(\mathcal{F}, \varepsilon, \mu)$ can blow up arbitrarily quickly as $\varepsilon \downarrow 0$.

PROPOSITION 1.7. *There exists a countable class $\mathcal{C}$ of subsets of $\mathbb{N}$, whose Vapnik-Chervonenkis dimension is two (that is, the γ-dimension of $\{\mathbf{1}_C : C \in \mathcal{C}\}$ is two for all $0 < \gamma < 1$) such that the following holds: for any function $n(\varepsilon) \uparrow \infty$ as $\varepsilon \downarrow 0$, there is a probability measure μ on $\mathbb{N}$ such that $N_{[]}(\mathcal{C}, \varepsilon, \mu) \geq n(\varepsilon)$ for all $0 < \varepsilon < 1/3$. In particular, $\sup_\mu N_{[]}(\mathcal{C}, \varepsilon, \mu) = \infty$ for all $0 < \varepsilon < 1/3$.*

Probabilistically, this result has the following consequence. In Theorem 1.3, we established that the universal Glivenko-Cantelli property is characterized (under mild regularity assumptions) in terms of the bracketing numbers. In contrast, Proposition 1.7 shows that neither the uniform Glivenko-Cantelli property nor the universal Donsker property can be characterized in terms of bracketing numbers (both these properties are characterized by finiteness of the Vapnik-Chervonenkis dimension for classes of sets, see [9], p. 225 and p. 215, respectively). Indeed, the former would require a uniform bound on the bracketing numbers, while the latter would require finiteness of the bracketing integral for every probability measure (see [16]), both of which are ruled out by Proposition 1.7.

The original motivation of the author was an attempt to characterize uniformity classes for certain reverse martingales that appear in the theory of nonlinear filtering. In a remarkable recent paper, Adams and Nobel [3] showed that Vapnik-Chervonenkis classes of sets are uniformity classes for the convergence of empirical measures of stationary ergodic sequences; their proof could be extended to more general random measures. A simplified version of the argument, which makes the natural connection with bracketing, appeared subsequently in [2]. While attempting to understand the results of [3], the author realized that the techniques used in the proof are very closely related to a set of techniques developed in Banach space theory by Bourgain, Fremlin and Talagrand [8, 19] in order to study pointwise compact sets of measurable functions. The proof of Theorem 1.3 is based on this elegant theory, which does not appear to be widely known in the probability literature (however, the proofs in this paper are intended to be essentially self-contained). A key innovation is the construction in section 2 of the "weakly dense" set which allows to fully exploit the techniques of [8, 19].

The remainder of this paper is organized as follows. We first prove Theorem 1.6 in section 2. The proofs of Theorem 1.3, Corollary 1.4, and Proposition 1.7 are subsequently given in sections 3, 4, and 5, respectively.

2. Proof of Theorem 1.6. In this section, we fix a measurable space $(X, \mathcal{X})$ and a separable uniformly bounded family of measurable functions $\mathcal{F}$. Let $\mathcal{F}_0 \subseteq \mathcal{F}$ be a countable family that is dense in $\mathcal{F}$ in the pointwise convergence topology.

DEFINITION 2.1. Denote by $\Pi(X, \mathcal{X})$ the collection of all finite measurable partitions of X. For $\pi, \pi' \in \Pi(X, \mathcal{X})$, we write $\pi \preceq \pi'$ if π is finer than π'. For any pair of sets $A, B \in \mathcal{X}$, finite partition $\pi \in \Pi(X, \mathcal{X})$, and probability measure μ on $(X, \mathcal{X})$, define the μ-essential π-boundary of (A, B) as

$$\partial_\pi^\mu(A, B) = \bigcup \{P \in \pi : \mu(P \cap A) > 0 \text{ and } \mu(P \cap B) > 0\}.$$

We begin by proving an approximation result.

LEMMA 2.2. *Let μ be a probability measure on $(X,\mathcal{X})$ and let $\gamma>0$. If*

$$\inf_{\pi\in\Pi(X,\mathcal{X})}\sup_{f\in\mathcal{F}_0}\mu\big(\partial^\mu_\pi(\{f<\alpha\},\{f>\beta\})\big)=0\quad\text{for all}\quad\beta-\alpha\ge\gamma,$$

then $N_{[]}(\mathcal{F},\varepsilon,\mu)<\infty$ for every $\varepsilon>\gamma$.

PROOF. There is clearly no loss of generality in assuming that every $f\in\mathcal{F}$ takes values in $[0,1]$. Fix $k\in\mathbb{N}$, and let $\delta:=\gamma/k$. Choose $\pi\in\Pi(X,\mathcal{X})$ such that

$$\sup_{f\in\mathcal{F}_0}\mu\left(\Xi(f)\right)<\delta,\qquad \Xi(f):=\bigcup_{1\le j\le\lfloor\delta^{-1}\rfloor}\partial^\mu_\pi(\{f<j\delta\},\{f>j\delta+\gamma\}).$$

For each $f\in\mathcal{F}_0$, define the functions f^+ and f^- as follows:

$$f^+=\delta\left\lceil\delta^{-1}\right\rceil\mathbf{1}_{\Xi(f)}+\sum_{P\in\pi:P\not\subseteq\Xi(f)}\delta\left\lceil\delta^{-1}\operatorname{ess\,sup}_P f\right\rceil\mathbf{1}_P,$$

$$f^-=\sum_{P\in\pi:P\not\subseteq\Xi(f)}\delta\left\lfloor\delta^{-1}\operatorname{ess\,inf}_P f\right\rfloor\mathbf{1}_P.$$

Here $\operatorname{ess\,sup}_P f$ ($\operatorname{ess\,inf}_P f$) denotes the essential supremum (infimum) of f on the set P with respect to μ. By construction, $f^-\le f\le f^+$ outside a μ-null set and $\mu(f^+-f^-)<\gamma+2\delta$. Moreover, as f^+,f^- are constant on each $P\in\pi$ and take values in the finite set $\{j\delta:0\le j\le\lceil\delta^{-1}\rceil\}$, there is only a finite number of such functions. As $\mathcal{F}_0$ is countable, we can eliminate the null set to obtain a finite number of $(\gamma+2\delta)$-brackets in $L^1(\mu)$ covering $\mathcal{F}_0$. But $\mathcal{F}_0$ is pointwise dense in $\mathcal{F}$, so $N_{[]}(\mathcal{F},\gamma+2\delta,\mu)<\infty$, and we may choose $\delta=\gamma/k$ arbitrarily small. □

To proceed, we need the notion of a "weakly dense" set, which is the measure-theoretic counterpart of the corresponding topological notion defined in [8].

DEFINITION 2.3. Given a measurable set $A\in\mathcal{X}$ and a probability measure μ on $(X,\mathcal{X})$, the family of functions $\mathcal{F}$ is said to be *μ-weakly dense over A at levels* (α,β) if $\mu(A)>0$ and for any finite collection of measurable sets $B_1,\ldots,B_p\in\mathcal{X}$ such that $\mu(A\cap B_i)>0$ for all $1\le i\le p$, there exists $f\in\mathcal{F}$ such that $\mu(A\cap B_i\cap\{f<\alpha\})>0$ and $\mu(A\cap B_i\cap\{f>\beta\})>0$ for all $1\le i\le p$.

The key idea of this section, which lies at the heart of the results in this paper, is that we can construct such a set if the bracketing numbers fail to be finite.

PROPOSITION 2.4. *Suppose there exists a probability measure μ on $(X,\mathcal{X})$ such that $N_{[]}(\mathcal{F},\varepsilon,\mu)=\infty$ for some $\varepsilon>\gamma$. Then there exist $\alpha<\beta$ with $\beta-\alpha\ge\gamma$ and a measurable set $A\in\mathcal{X}$ such that $\mathcal{F}_0$ is μ-weakly dense over A at levels (α,β).*

PROOF. By Lemma 2.2, there exist $\alpha < \beta$ with $\beta - \alpha \geq \gamma$ such that

$$\inf_{\pi \in \Pi(X,\mathcal{X})} \sup_{f \in \mathcal{F}_0} \mu\big(\partial_\pi^\mu(\{f < \alpha\}, \{f > \beta\})\big) > 0.$$

Choose for every $\pi \in \Pi(X,\mathcal{X})$ a function $f_\pi \in \mathcal{F}_0$ such that

$$\mu\big(\partial_\pi^\mu(\{f_\pi < \alpha\}, \{f_\pi > \beta\})\big) \geq \frac{1}{2} \sup_{f \in \mathcal{F}_0} \mu\big(\partial_\pi^\mu(\{f < \alpha\}, \{f > \beta\})\big).$$

Define $A_\pi := \partial_\pi^\mu(\{f_\pi < \alpha\}, \{f_\pi > \beta\})$. Then $(\mathbf{1}_{A_\pi})_{\pi \in \Pi(X,\mathcal{X})}$ is a net of random variables in the unit ball of $L^2(\mu)$. By weak compactness, we may extract a subnet $(\mathbf{1}_{A_{\pi(\tau)}})_{\tau \in T}$ that converges weakly in $L^2(\mu)$ to a random variable H. We claim that $\mathcal{F}_0$ is μ-weakly dense over $A := \{H > 0\}$ at levels (α, β).

To prove the claim, let us first note that as $\inf_\pi \mu(A_\pi) > 0$, clearly $\mu(A) > 0$. Now fix $B_1, \ldots, B_p \in \mathcal{X}$ such that $\mu(A \cap B_i) > 0$ for all i. This trivially implies that $\mu(H\mathbf{1}_{A \cap B_i}) > 0$ for all i, so we can choose $\tau_0 \in T$ such that

$$\mu(A_{\pi(\tau)} \cap A \cap B_i) > 0 \quad \forall\, 1 \leq i \leq p,\ \tau \preceq \tau_0.$$

Let π_0 be the partition generated by $A, B_1, \ldots, B_p$, and choose $\tau^* \in T$ such that $\tau^* \preceq \tau_0$ and $\pi^* := \pi(\tau^*) \preceq \pi_0$. As $A \cap B_i$ is a union of atoms of π^* by construction, $\mu(A_{\pi^*} \cap A \cap B_i) > 0$ must imply that $A \cap B_i$ contains an atom $P \in \pi^*$ such that $\mu(P \cap \{f_{\pi^*} < \alpha\}) > 0$ and $\mu(P \cap \{f_{\pi^*} > \beta\}) > 0$. Therefore

$$\mu(A \cap B_i \cap \{f_{\pi^*} < \alpha\}) > 0 \quad \text{and} \quad \mu(A \cap B_i \cap \{f_{\pi^*} > \beta\}) > 0 \quad \forall\, i.$$

Thus $\mathcal{F}_0$ is μ-weakly dense over A at levels (α, β) as claimed. □

We can now complete the proof of Theorem 1.6.

PROOF OF THEOREM 1.6.

$a \Rightarrow b$: The proof of Theorem 4.6.2 in [9] (see also [6]) shows that if $\mathcal{F}$ contains a finite subset of cardinality 2^n that is Boolean independent at levels (α, β) with $\beta - \alpha \geq \gamma$, then $\mathcal{F}$ γ-shatters a finite subset of X of cardinality n. Therefore, if condition b fails, there must clearly exist γ-shattered finite subsets of X of arbitrarily large cardinality, in contradiction with condition a.

$b \Rightarrow c$: Suppose that condition c fails. By Lemma 2.2 and Proposition 2.4, there exist a probability measure μ, levels $\alpha < \beta$ with $\beta - \alpha \geq \gamma$, and a measurable set $A \in \mathcal{X}$ such that $\mathcal{F}_0$ is μ-weakly dense over A at levels (α, β). We can now iteratively apply Definition 2.3 to construct a Boolean independent sequence. Indeed, applying first the definition with $p = 1$ and $B_1 = X$, we choose $f_1 \in \mathcal{F}_0$ such that $\mu(A \cap \{f_1 < \alpha\}) > 0$ and $\mu(A \cap \{f_1 > \beta\}) > 0$. Then applying the definition with $p = 2$ and $B_1 = \{f_1 < \alpha\}$, $B_2 = \{f_1 > \beta\}$, we choose $f_2 \in \mathcal{F}_0$ such that $\mu(A \cap \{f_1 < \alpha\} \cap \{f_2 < \alpha\}) > 0$, $\mu(A \cap \{f_1 < \alpha\} \cap \{f_2 > \beta\}) > 0$, $\mu(A \cap \{f_1 > \beta\} \cap \{f_2 < \alpha\}) > 0$, and $\mu(A \cap \{f_1 > \beta\} \cap \{f_2 > \beta\}) > 0$. Repeating this procedure yields the desired sequence $(f_i)_{i\in\mathbb{N}}$. □

3. Proof of Theorem 1.3. Throughout this section, we fix a standard measurable space $(X, \mathcal{X})$ and a separable uniformly bounded family of measurable functions $\mathcal{F}$. We will prove Theorem 1.3 by proving the implications $1 \Rightarrow 4 \Rightarrow 2 \Rightarrow 1$ and $2 \Rightarrow 3 \Rightarrow 4$. Below, we consider each of these implications in turn.

3.1. $1 \Rightarrow 4$. Suppose there exists a sequence $(f_i)_{i\in\mathbb{N}} \subseteq \mathcal{F}$ that is Boolean σ-independent at levels (α, β) for some $\alpha < \beta$. Clearly we must have

$$\kappa_- < \alpha < \beta < \kappa_+, \qquad \kappa_- := \inf_{f\in\mathcal{F}} \inf_{x\in X} f(x), \quad \kappa_+ := \sup_{f\in\mathcal{F}} \sup_{x\in X} f(x).$$

Define the measurable set

$$X_0 = \bigcap_{n\in\mathbb{N}} (\{f_n < \alpha\} \cup \{f_n > \beta\}).$$

As $(f_i)_{i\in\mathbb{N}}$ is Boolean σ-independent, this set is nonempty (in fact uncountable). Define $C_n := \{f_n < \alpha\} \cap X_0$, and note that $X_0 \backslash C_n = \{f_n > \beta\} \cap X_0$ by construction. Therefore, the Boolean σ-independence property can be expressed as

$$\bigcap_{j\in F} C_j \cap \bigcap_{j\notin F} X_0 \backslash C_j \neq \varnothing \quad \text{for every } F \subseteq \mathbb{N}.$$

Define on X_0 the σ-field $\mathcal{X}_0 := \sigma\{C_n : n \in \mathbb{N}\}$. By a result of Marczewski [14], p. 25, there exists a probability measure μ_0 on $(X_0, \mathcal{X}_0)$ such that $(C_n)_{n\in\mathbb{N}}$ is an i.i.d. sequence of sets with $\mu_0(C_n) := p$, where we choose $(\kappa_+ - \beta)/(\kappa_+ - \alpha) < p < 1$. As $(X, \mathcal{X})$ is a standard measurable space, the extension theorem of [13] implies that there exists a probability measure μ on $(X, \mathcal{X})$ that is supported on X_0 and such that the restriction of μ to $(X_0, \mathcal{X}_0)$ coincides with μ_0.

We now claim that $\mathcal{F}$ is not μ-Glivenko-Cantelli, which yields the desired contradiction. To this end, note that we can trivially estimate for any $f \in \mathcal{F}$

$$\beta\,\mathbf{1}_{f>\beta} + \kappa_-\,\mathbf{1}_{f\le\beta} \le f \le \alpha\,\mathbf{1}_{f<\alpha} + \kappa_+\,\mathbf{1}_{f\ge\alpha}.$$

We therefore have

$$\begin{aligned}\sup_{f\in\mathcal{F}} \left| \frac{1}{n} \sum_{k=1}^n f(X_k) - \mu(f) \right| &\ge \sup_{j\in\mathbb{N}} \frac{1}{n} \sum_{k=1}^n \{f_j(X_k) - \mu(f_j)\} \\ &\ge (\kappa_- - \beta) \inf_{j\in\mathbb{N}} \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{f_j\le\beta}(X_k) + \varepsilon,\end{aligned}$$

where we have used that $\mu(f_n < \alpha) = p = (\kappa_+ - \beta + \varepsilon)/(\kappa_+ - \alpha)$ for some $\varepsilon > 0$. But if $(X_k)_{k\geq 1}$ are i.i.d. with distribution μ then, by construction, the family of random variables $\{\mathbf{1}_{f_j \leq \beta}(X_k) : j, k \in \mathbb{N}\}$ is i.i.d. with $\mathbf{P}[\mathbf{1}_{f_j\leq\beta}(X_k) = 0] > 0$. Using the Borel-Cantelli lemma, it is easily established that

$$\inf_{j\in\mathbb{N}} \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{f_j\leq\beta}(X_k) = 0 \quad \text{a.s.} \quad \text{for all } n \in \mathbb{N}.$$

Thus $\mathcal{F}$ is not a μ-Glivenko-Cantelli class. This completes the proof.

3.2. $4 \Rightarrow 2$. Suppose there exists a probability measure μ and $\varepsilon > 0$ such that $N_{[]}(\mathcal{F}, \varepsilon, \mu) = \infty$. By Lemma 2.2 and Proposition 2.4, there exist levels $\alpha < \beta$ and a measurable set $A \in \mathcal{X}$ such that $\mathcal{F}$ is μ-weakly dense over A at levels (α, β). We will presently construct a Boolean σ-independent sequence, which yields the desired contradiction. The idea is to repeat the proof of Theorem 1.6, but now exploiting the fact that $(X, \mathcal{X})$ is a standard measurable space to ensure that the infinite intersections in the definition of Boolean σ-independence are nonempty.

As $(X, \mathcal{X})$ is standard, we may assume without loss of generality that X is Polish and that $\mathcal{X}$ is the Borel σ-field. Thus μ is inner regular. We now apply Definition 2.3 as follows. First, setting $p = 1$ and $B_1 = X$, choose $f_1 \in \mathcal{F}$ such that

$$\mu(A \cap \{f_1 < \alpha\}) > 0, \qquad \mu(A \cap \{f_1 > \beta\}) > 0.$$

As μ is inner regular, we may choose compact sets $F_1 \subseteq \{f_1 < \alpha\}$ and $G_1 \subseteq \{f_1 > \beta\}$ such that $\mu(A \cap F_1) > 0$ and $\mu(A \cap F_2) > 0$. Applying the definition with $p = 2$, $B_1 = F_1$, and $B_2 = G_1$, we can choose $f_2 \in \mathcal{F}$ such that

$$\begin{aligned} &\mu(A \cap F_1 \cap \{f_2 < \alpha\}) > 0, \qquad &&\mu(A \cap F_1 \cap \{f_2 > \beta\}) > 0, \\ &\mu(A \cap G_1 \cap \{f_2 < \alpha\}) > 0, \qquad &&\mu(A \cap G_1 \cap \{f_2 > \beta\}) > 0. \end{aligned}$$

Using again inner regularity, we can now choose compact sets $F_2 \subseteq \{f_2 < \alpha\}$ and $G_2 \subseteq \{f_2 > \beta\}$ such that $\mu(A \cap F_1 \cap F_2) > 0$, $\mu(A \cap F_1 \cap G_2) > 0$, $\mu(A \cap G_1 \cap F_2) > 0$, and $\mu(A \cap G_1 \cap G_2) > 0$. Iterating the above steps, we construct a sequence of functions $(f_i)_{i\in\mathbb{N}} \subseteq \mathcal{F}$ and compact sets $(F_i)_{i\in\mathbb{N}}$, $(G_i)_{i\in\mathbb{N}}$ such that $F_i \subseteq \{f_i < \alpha\}$, $G_i \subseteq \{f_i > \beta\}$ for every $i \in \mathbb{N}$, and for any $n \in \mathbb{N}$

$$\mu\left(\bigcap_{j\in Q} F_j \cap \bigcap_{j\in\{1,\ldots,n\}\setminus Q} G_j\right) > 0 \quad \text{for every } Q \subseteq \{1, \ldots, n\}.$$

Now suppose that the sequence $(f_i)_{i\in\mathbb{N}}$ is not Boolean σ-independent. Then

$$\bigcap_{j\in R}\{f_j < \alpha\} \cap \bigcap_{j\notin R}\{f_j > \beta\} = \varnothing$$

for some $R \subseteq \mathbb{N}$. Thus we certainly have

$$\bigcap_{j\in R} F_j \cap \bigcap_{j\notin R} G_j = \varnothing.$$

Choose arbitrary $\ell \in R$ (if R is the empty set, replace F_ℓ by G_1 throughout the following argument). Then clearly $\{X\backslash F_j : j \in R\} \cup \{X\backslash G_j : j \notin R\}$ is an open cover of F_ℓ. Therefore, there exists finite subsets $Q_1 \subseteq R$, $Q_2 \subseteq \mathbb{N}\backslash R$ such that $\{X\backslash F_j : j \in Q_1\} \cup \{X\backslash G_j : j \in Q_2\}$ covers F_ℓ. But then

$$F_\ell \cap \bigcap_{j\in Q_1} F_j \cap \bigcap_{j\in Q_2} G_j = \varnothing,$$

a contradiction. Thus $(f_i)_{i\in\mathbb{N}}$ is Boolean σ-independent at levels (α,β).

3.3. $2 \Rightarrow 1$. This is the usual Blum-DeHardt argument, included here for completeness. Fix a probability measure μ and $\varepsilon > 0$, and suppose that $N_{[]}(\mathcal{F},\varepsilon,\mu) < \infty$. Choose ε-brackets $[f_1,g_1],\ldots,[f_N,g_N]$ in $L^1(\mu)$ covering $\mathcal{F}$. Then

$$\begin{aligned}\sup_{f\in\mathcal{F}} |\mu_n(f) - \mu(f)| &= \sup_{f\in\mathcal{F}}\{\mu_n(f) - \mu(f)\} \vee \sup_{f\in\mathcal{F}}\{\mu(f) - \mu_n(f)\} \\ &\le \max_{i=1,\ldots,N}\{\mu_n(g_i) - \mu(f_i)\} \vee \max_{i=1,\ldots,N}\{\mu(g_i) - \mu_n(f_i)\},\end{aligned}$$

where we define the empirical measure $\mu_n := \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$ for an i.i.d. sequence $(X_k)_{k\in\mathbb{N}}$ with distribution μ. The right hand side in the above expression is measurable and converges a.s. to a constant not exceeding ε by the law of large numbers. As $\varepsilon > 0$ and μ were arbitrary, $\mathcal{F}$ is universal Glivenko-Cantelli.

3.4. $2 \Rightarrow 3 \Rightarrow 4$. As $N(\mathcal{F},\varepsilon,\mu) \le N_{[]}(\mathcal{F},2\varepsilon,\mu)$, the implication $2 \Rightarrow 3$ is trivial. It therefore remains to prove the implication $3 \Rightarrow 4$.

To this end, suppose that there exists a sequence $(f_i)_{i\in\mathbb{N}} \subseteq \mathcal{F}$ that is Boolean σ-independent at levels (α,β) for some $\alpha < \beta$. Construct the probability measure μ as in the proof of the implication $1 \Rightarrow 4$. We claim that $N(\mathcal{F},\varepsilon,\mu) = \infty$ for $\varepsilon > 0$ sufficiently small, which yields the desired contradiction.

To prove the claim, it suffices to note that for any $i \ne j$

$$\begin{aligned}\mu(|f_i - f_j|) &\ge \mu(|f_i - f_j|\mathbf{1}_{f_j<\alpha}\mathbf{1}_{f_i>\beta}) \\ &\ge (\beta-\alpha)\,\mu(\{f_j < \alpha\} \cap \{f_i > \beta\}) = (\beta-\alpha)p(1-p) > 0\end{aligned}$$

by the construction of μ. Therefore $\mathcal{F}$ contains an infinite set of $(\beta-\alpha)p(1-p)$-separated points in $L^1(\mu)$, so $N(\mathcal{F},(\beta-\alpha)p(1-p)/2,\mu) = \infty$.

4. Proof of Corollary 1.4. Throughout this section, we fix a standard measurable space $(X, \mathcal{X})$ and a separable uniformly bounded family of measurable functions $\mathcal{F}$. We will prove Corollary 1.4 by proving the implications $2 \Leftrightarrow 5$ and $2 \Rightarrow \{6, 7, 8\} \Rightarrow 1$. Below, we consider each of these implications in turn.

4.1. $2 \Leftrightarrow 5$. The implication $2 \Rightarrow 5$ follows by the Blum-DeHardt argument as in section 3.3. For the implication $5 \Rightarrow 2$, we employ a result due to Topsøe [20] that can be stated as follows. For any probability measure μ, function $f \in \mathcal{F}$, and finite partition $\pi \in \Pi(X, \mathcal{X})$, define the *$\mu$-average π-oscillation of f* as

$$\varpi_\pi^\mu f = \sum_{P \in \pi} \left\{ \sup_{x \in P} f(x) - \inf_{x \in P} f(x) \right\} \mu(P).$$

By [20], Theorem 1, condition 5 holds if and only if

$$\inf_{\pi \in \Pi(X, \mathcal{X})} \sup_{f \in \mathcal{F}} \varpi_\pi^\mu f = 0 \quad \text{for every probability measure } \mu.$$

We claim that the latter property implies condition 2.

To prove the claim, we may clearly assume that every $f \in \mathcal{F}$ takes values in $[0, 1]$. Fix a probability measure μ and $k \in \mathbb{N}$. Choose $\pi \in \Pi(X, \mathcal{X})$ such that $\sup_{f \in \mathcal{F}} \varpi_\pi^\mu f \leq k^{-1}$. For each $f \in \mathcal{F}$, define the functions f^+ and f^- as follows:

$$f^+ = \sum_{P \in \pi} k^{-1} \lceil k \sup\nolimits_{x \in P} f(x) \rceil \, \mathbf{1}_P,$$
$$f^- = \sum_{P \in \pi} k^{-1} \lfloor k \inf\nolimits_{x \in P} f(x) \rfloor \, \mathbf{1}_P.$$

By construction, $f^- \leq f \leq f^+$ pointwise and $\mu(f^+ - f^-) \leq 3k^{-1}$ for every $f \in \mathcal{F}$. Moreover, as f^+, f^- are constant on each $P \in \pi$ and take values in the finite set $\{jk^{-1} : 0 \leq j \leq k\}$, there is only a finite number of such functions. Thus $N_{[]}(\mathcal{F}, 3k^{-1}, \mu) < \infty$. As μ and k are arbitrary, the claim is established.

4.2. $2 \Rightarrow \{6, 7, 8\}$. The implication $2 \Rightarrow 6$ follows immediately from the Blum-DeHardt argument as in section 3.3. The complication for the implications $2 \Rightarrow \{7, 8\}$ is that the limiting measure is a random measure (unlike $2 \Rightarrow 6$ where the limiting measure is assumed to be nonrandom). Intuitively one can simply condition on $\mathcal{G}_{-\infty}$ or $\mathcal{I}$, respectively, so that the problem reduces to the implication $2 \Rightarrow 6$ under the conditional distribution. The main work in the proof consists of resolving the measurability issues that arise in this approach.

Let $\mathcal{F}_0 \subseteq \mathcal{F}$ be a countable family that is dense in $\mathcal{F}$ in the topology of pointwise convergence. We first show that $\mathcal{F}_0$ is also $L^1(\mu)$-dense in $\mathcal{F}$ for any μ: this is not obvious, as the dominated convergence theorem does not hold for nets.

LEMMA 4.1. *If $N_{[]}(\mathcal{F},\varepsilon,\mu)<\infty$ for all $\varepsilon>0$, then $\mathcal{F}_0$ is $L^1(\mu)$-dense in $\mathcal{F}$.*

PROOF. Fix $\varepsilon>0$, and choose ε-brackets $[f_1,g_1],\ldots,[f_N,g_N]$ in $L^1(\mu)$ covering $\mathcal{F}$. As topological closure and finite unions commute, for every $f\in\mathcal{F}$ there exists $1\le i\le N$ such that f is in the pointwise closure of $[f_i,g_i]\cap\mathcal{F}_0$. But then clearly $f\in[f_i,g_i]$, and choosing any $g\in[f_i,g_i]\cap\mathcal{F}_0$ we have $\mu(|f-g|)\le\mu(g_i-f_i)\le\varepsilon$. As $\varepsilon>0$ is arbitrary, the proof is complete. □

We can now reduce the suprema in conditions 7 and 8 to countable suprema.

COROLLARY 4.2. *Suppose that $N_{[]}(\mathcal{F},\varepsilon,\mu)<\infty$ for every $\varepsilon>0$ and probability measure μ. Then for any pair of probability measures μ,ν we have*

$$\sup_{f\in\mathcal{F}}|\mu(f)-\nu(f)|=\sup_{f\in\mathcal{F}_0}|\mu(f)-\nu(f)|.$$

In particular, this holds when μ and ν are random measures.

PROOF. Fix (nonrandom) probability measures μ,ν, and define $\rho=\{\mu+\nu\}/2$. Then $\mathcal{F}_0$ is $L^1(\rho)$-dense in $\mathcal{F}$ by Lemma 4.1. In particular, for every $f\in\mathcal{F}$ and $\varepsilon>0$, we can choose $g\in\mathcal{F}_0$ such that $\mu(|f-g|)+\nu(|f-g|)\le\varepsilon$. Now let $(f_n)_{n\in\mathbb{N}}\subseteq\mathcal{F}$ be a sequence such that $\sup_{f\in\mathcal{F}}|\mu(f)-\nu(f)|=\lim_{n\to\infty}|\mu(f_n)-\nu(f_n)|$. For each f_n, choose $g_n\in\mathcal{F}_0$ such that $\mu(|f_n-g_n|)+\nu(|f_n-g_n|)\le n^{-1}$. Then

$$\sup_{f\in\mathcal{F}}|\mu(f)-\nu(f)|=\lim_{n\to\infty}|\mu(g_n)-\nu(g_n)|\le\sup_{f\in\mathcal{F}_0}|\mu(f)-\nu(f)|,$$

which clearly yields the result (as $\mathcal{F}_0\subseteq\mathcal{F}$). In the case of random probability measures, we simply apply the nonrandom result pointwise. □

To prove $2\Rightarrow 8$ we use the ergodic decomposition. Consider a strictly stationary sequence $(Z_n)_{n\in\mathbb{Z}}$ of X-valued random variables on an underlying probability space $(\Omega,\mathcal{G},\mathbf{P})$. As $(X,\mathcal{X})$ (hence also $(X^{\mathbb{Z}},\mathcal{X}^{\otimes\mathbb{Z}})$) is standard, Theorem 7.6 in [12] states that there exists a version $\mathbf{P}_{\mathcal{I}}$ of the regular conditional probability $\mathbf{P}[(Z_n)_{n\in\mathbb{Z}}\in\cdot\,|\mathcal{I}]$ such that the law $\mathbf{P}_{\mathcal{I}}^{\omega}$ is stationary and ergodic for every $\omega\in\Omega$. Applying the Blum-DeHardt argument as in section 3.3 for fixed $\omega\in\Omega$ gives

$$\mathbf{P}_{\mathcal{I}}^{\omega}\left[\sup_{f\in\mathcal{F}_0}\left|\frac{1}{n}\sum_{k=1}^{n}f(Z_k)-\mathbf{P}_{\mathcal{I}}^{\omega}(f(Z_0))\right|\xrightarrow{n\to\infty}0\right]=1\quad\text{for all }\omega\in\Omega.$$

Integrating this expression with respect to $\mathbf{P}(d\omega)$ gives

$$\sup_{f\in\mathcal{F}_0}\left|\frac{1}{n}\sum_{k=1}^{n}f(Z_k)-\mathbf{P}_{\mathcal{I}}(f(Z_0))\right|\xrightarrow{n\to\infty}0\quad\text{a.s.}$$

using the disintegration of measures. Applying Corollary 4.2 and using the fact that regular conditional probabilities are a.s. unique (so the result holds for an arbitrary choice of the regular conditional probability) yields the desired result.

To prove the implication $2 \Rightarrow 7$, we aim to repeat the proof of $2 \Rightarrow 8$ with a suitable tail decomposition, given presently, replacing the ergodic decomposition.

PROPOSITION 4.3. *Let $(Z_{-k})_{k\geq 0}$ be random variables on a probability space $(\Omega, \mathcal{G}, \mathbf{P})$ taking values in the standard space $(X, \mathcal{X})$, and let $\mathcal{F}_0$ be a countable family of bounded measurable functions on $(X, \mathcal{X})$. Let $\mathcal{G}_{-n} := \sigma\{Z_{-k} : k \geq n\}$ and $\mathcal{G}_{-\infty} := \bigcap_n \mathcal{G}_{-n}$. Then there exists for every $0 \leq n \leq \infty$ a version $\mathbf{P}_{\mathcal{G}_{-n}}$ of the regular conditional probability $\mathbf{P}[(Z_{-k})_{k\geq 0} \in \cdot\,|\mathcal{G}_{-n}]$ such that:*

1. *For $\mathbf{P}$-a.e. $\omega \in \Omega$*

$$\mathbf{P}^{\omega}_{\mathcal{G}_{-\infty}}\left[\mathbf{P}^{\omega}_{\mathcal{G}_{-\infty}}[f(Z_0)|\mathcal{G}_{-n}] = \mathbf{P}_{\mathcal{G}_{-n}}(f(Z_0)) \quad \forall\, f \in \mathcal{F}_0,\ n \in \mathbb{N}\right] = 1.$$

2. *$\mathcal{G}_{-\infty}$ is $\mathbf{P}^{\omega}_{\mathcal{G}_{-\infty}}$-a.s. trivial for $\mathbf{P}$-a.e. $\omega \in \Omega$.*

PROOF. Define on $(\Omega \times \Omega, \mathcal{G} \otimes \mathcal{G})$ the random variables $(Z^1_{-n}, Z^2_{-n})_{n\geq 0}$ as $Z^1_{-n}(\omega, \omega') = Z_{-n}(\omega)$, $Z^2_{-n}(\omega, \omega') = Z_{-n}(\omega')$, and define the probability measure $\mathbf{Q}$ such that $\mathbf{Q}(A \times B) = \mathbf{P}(A \cap B)$ for all $A, B \in \mathcal{G}$. That is, $\mathbf{Q}$ is supported on the diagonal $\{(\omega, \omega') : \omega = \omega'\}$ and its marginals coincide with $\mathbf{P}$.

Choose for every $n \leq \infty$ a version μ_n of the regular conditional probability $\mathbf{P}[(Z_{-k})_{k\geq 0} \in \cdot\,|\mathcal{G}_{-n}]$ (it exists by [10], Theorem 10.2.2 as $(X^{\mathbb{Z}_+}, \mathcal{X}^{\otimes \mathbb{Z}_+})$ is standard). It is easily seen that $(\omega, \omega') \mapsto \mu^{\omega}_{\infty}$ is a version of the regular conditional probability $\mathbf{Q}[(Z^2_{-k})_{k\geq 0} \in \cdot\,|\mathcal{G}_{-\infty} \otimes \{\varnothing, \Omega\}]$ and that $(\omega, \omega') \mapsto \mu^{\omega'}_n$ is a version of the regular conditional probability $\mathbf{Q}[(Z^2_{-k})_{k\geq 0} \in \cdot\,|\mathcal{G}_{-\infty} \otimes \mathcal{G}_{-n}]$ for every $n < \infty$. As $\mathcal{G}_{-n}$ is countably generated, the result in [22], pp. 95–96 states that

$$\mu^{\omega}_{\infty}[f(Z_0)|\mathcal{G}_{-n}] = \mu_n(f(Z_0)) \qquad \mathbf{P}\text{-a.e. } \omega \in \Omega$$

for all $f \in \mathcal{F}_0$ and $n \in \mathbb{N}$. As $\mathcal{F}_0, \mathbb{N}$ are countable we have proved the first part of the result. The second part of the result is proved in [7], Theorem 15. □

We now prove $2 \Rightarrow 7$. Let $(\mathcal{G}_{-n})_{n\in\mathbb{N}}$ be a reverse filtration such that $\mathcal{G}_{-n}$ is countably generated for each $n \in \mathbb{N}$, and consider a random variable Z taking values in the standard space $(X, \mathcal{X})$. Choose for every $n \in \mathbb{N}$ a countable generating class $(H_{n,j})_{j\in\mathbb{N}} \subseteq \mathcal{G}_{-n}$, and define the $\{0,1\}^{\mathbb{N}}$-valued random variable $Y_{-n} = (\mathbf{1}_{H_{n,j}})_{j\in\mathbb{N}}$. Then, by construction, $\mathcal{G}_{-n} = \sigma\{Y_{-k} : k \geq n\}$. As $\{0,1\}^{\mathbb{N}}$ is Polish, it is Borel-isomorphic with $(X, \mathcal{X})$, so we can construct for every $n \in \mathbb{N}$ an $(X, \mathcal{X})$-valued random variable Z_{-n} such that $\mathcal{G}_{-n} = \sigma\{Z_{-k} : k \geq n\}$. Finally, define $Z_0 = Z$. This puts us in the setting of Proposition 4.3.

Define the tail σ-field $\mathcal{G}_{-\infty} = \bigcap_n \mathcal{G}_{-n}$ (of course $\mathcal{G}_{-\infty}$ is not countably generated, but this is not needed). Using Proposition 4.3, the martingale convergence theorem, and applying the Blum-DeHardt argument as in section 3.3 gives

$$\mathbf{P}^{\omega}_{\mathcal{G}_{-\infty}}\left[\sup_{f\in\mathcal{F}_0}\left|\mathbf{P}_{\mathcal{G}_{-n}}(f(Z)) - \mathbf{P}^{\omega}_{\mathcal{G}_{-\infty}}(f(Z))\right| \xrightarrow{n\to\infty} 0\right] = 1 \quad \text{for } \mathbf{P}\text{-a.e. } \omega\in\Omega.$$

Integrating this expression with respect to $\mathbf{P}(d\omega)$, using the disintegration of measures, and applying Corollary 4.2 yields the desired result.

4.3. $\{6,7,8\} \Rightarrow 1$. These implications are immediate, as each of the conditions $\{6,7,8\}$ contains condition 1 as a special case. Indeed, for the implication $6 \Rightarrow 1$, it suffices to choose μ_n to be the empirical measure of an i.i.d. sequence with distribution μ. Similarly, the implication $8 \Rightarrow 1$ follows from the fact that an i.i.d. sequence is stationary and ergodic. Finally, the implication $7 \Rightarrow 1$ follows from Theorem 6.1.6 in [9] and the Kolmogorov zero-one law.

5. Proof of Proposition 1.7.

5.1. *Construction.* The construction of the class $\mathcal{C}$ in Proposition 1.7 is based on a combinatorial construction due to Alon, Haussler, and Welzl [5], Theorem A(2). We begin by recalling the essential results in that paper.

DEFINITION 5.1. Let $q\in\mathbb{N}$ be a prime power and define $m = q^2+q+1$. The *finite projective plane* $\mathrm{PG}(2,q)$ is a family of m subsets of $\{1,\ldots,m\}$ such that every set $C\in\mathrm{PG}(2,q)$ has cardinality $q+1$, every point $x\in\{1,\ldots,m\}$ belongs to exactly $q+1$ elements of $\mathrm{PG}(2,q)$, and for every pair of points $x,x'\in\{1,\ldots,m\}$, $x\neq x'$ there is a unique set $C\in\mathrm{PG}(2,q)$ with $x,x'\in C$.

The finite projective plane $\mathrm{PG}(2,q)$ is known to exist whenever q is a prime power (see [5] and the references therein). For our purposes, the key result about finite projective planes is the following (see [5], p. 336 for the proof).

PROPOSITION 5.2. *Let $q\in\mathbb{N}$ be a prime power and define $m=q^2+q+1$. Then for any partition π of $\{1,\ldots,m\}$ such that*

$$\max_{C\in\mathrm{PG}(2,q)} \frac{|\partial_\pi C|}{m} \le \varepsilon,$$

we have $|\pi|^2 > m^{1/2}(1-\varepsilon)$. Here $|\cdot|$ denotes the cardinality of a set, and we have defined the π-boundary $\partial_\pi C := \bigcup\{P\in\pi : P\cap C\neq\varnothing$ and $P\not\subseteq C\}$.

We now proceed to construct the class $\mathcal{C}$ in Proposition 1.7. Let $q_j \uparrow \infty$ be an increasing sequence of prime powers, and define $m_j = q_j^2 + q_j + 1$. We now partition $\mathbb{N}$ into consecutive blocks of length m_j, as follows:

$$\mathbb{N} = \bigcup_{j=1}^{\infty} N_j, \qquad N_j = \left\{ \sum_{i=1}^{j-1} m_j + 1, \ldots, \sum_{i=1}^{j} m_j \right\}.$$

Define $\mathcal{C}$ as the disjoint union of copies of $\mathrm{PG}(2, q_j)$ on the blocks N_j:

$$\mathcal{C} = \bigcup_{j=1}^{\infty} \mathcal{C}_j, \qquad \mathcal{C}_j = \{B \subseteq N_j : B \cap N_j = C,\ C \in \mathrm{PG}(2, q_j)\}.$$

We claim that the countable class $\mathcal{C}$ of subsets of $\mathbb{N}$ has γ-dimension two.

LEMMA 5.3. *$\mathcal{C}$ has Vapnik-Chervonenkis dimension two.*

PROOF. Choose any three distinct points $n_1, n_2, n_3 \in \mathbb{N}$. If two of these points are in distinct intervals N_j, then no set in $\mathcal{C}$ contains both points. On the other hand, suppose that all three points are in the same interval N_j. Then by the definition of the finite projective plane, either there is no set in $\mathcal{C}$ that contains all three points, or there is no set that contains two of the points but not the third (as each pair of points must lie in a unique set in $\mathcal{C}$). Thus we have shown that no family of three points $\{n_1, n_2, n_3\}$ is γ-shattered for $0 < \gamma < 1$. On the other hand, it is clear from the definition of the finite projective plane that any pair of points $\{n_1, n_2\}$ belonging to the same interval N_j is γ-shattered for $0 < \gamma < 1$. ☐

We now turn to the proof of Proposition 1.7.

5.2. *Proof of Proposition 1.7.* We will use the following crude lemma to obtain lower bounds on the bracketing numbers.

LEMMA 5.4. *Let μ be a probability measure on $\mathbb{N}$. Then*

$$\inf_{|\pi| \le 3^N} \sup_{C \in \mathcal{C}} \mu(\partial_\pi C) > \varepsilon \qquad \textit{implies} \qquad N_{[]}(\mathcal{C}, \varepsilon, \mu) > N,$$

where the infimum ranges over all partitions of $\mathbb{N}$ with $|\pi| \le 3^N$.

PROOF. Suppose that $N_{[]}(\mathcal{C}, \varepsilon, \mu) \le N$. Then we may choose $k \le N$ pairs $\{C_i^+, C_i^-\}_{1 \le i \le k}$ of subsets of $\mathbb{N}$ such that $\mu(C_i^+ \backslash C_i^-) \le \varepsilon$ for all $1 \le i \le k$, and for every $C \in \mathcal{C}$, there exists $1 \le i \le k$ such that $C_i^- \subseteq X \subseteq C_i^+$. Let π be the partition generated by $\{C_i^+, C_i^- : 1 \le i \le k\}$. Then $|\pi| \le 3^N$, as π is the common refinement of at most N partitions $\{C_i^-, C_i^+ \backslash C_i^-, \mathbb{N}\backslash C_i^+\}$ of size three.

Now choose any $C \in \mathcal{C}$, and choose $1 \le i \le k$ such that $C_i^- \subseteq C \subseteq C_i^+$. As C_i^- and $\mathbb{N}\backslash C_i^+$ are unions of atoms of π by construction, and as $C_i^- \subseteq C$ and $(\mathbb{N}\backslash C_i^+) \cap C = \varnothing$, we evidently have $\partial_\pi C \subseteq C_i^+ \backslash C_i^-$. Thus $\mu(\partial_\pi C) \le \varepsilon$. As this holds for any $C \in \mathcal{C}$, we complete the proof by contradiction. □

Denote by μ_j the uniform distribution on N_j. Let $(p_j)_{j\in\mathbb{N}}$ be a sequence of nonnegative numbers $p_j \ge 0$ so that $\sum_j p_j = 1$, and define the probability measure

$$\mu = \sum_{j=1}^{\infty} p_j \mu_j.$$

We first obtain a lower bound on $N_{[]}(\mathcal{C}, \varepsilon, \mu)$. Subsequently, we will be able to choose the sequence $(p_j)_{j\in\mathbb{N}}$ such that this bound grows arbitrarily quickly.

To obtain a lower bound, let us suppose that $N_{[]}(\mathcal{C}, \varepsilon, \mu) \le N$. Then applying Lemma 5.4, there exists a partition π of $\mathbb{N}$ with $|\pi| \le 3^N$ such that

$$\sup_{j\in\mathbb{N}} p_j \min_{|\pi'|\le 3^N} \max_{C\in \mathrm{PG}(2,q_j)} \frac{|\partial_{\pi'} C|}{m_j} \le \sup_{j\in\mathbb{N}} p_j \sup_{C\in\mathcal{C}_j} \mu_j(\partial_\pi C) \le \sup_{C\in\mathcal{C}} \mu(\partial_\pi C) \le \varepsilon.$$

By Proposition 5.2,

$$\min_{|\pi'|\le 3^N} \max_{C\in \mathrm{PG}(2,q_j)} \frac{|\partial_{\pi'} C|}{m_j} \le \frac{\varepsilon}{p_j} \quad \text{implies} \quad m_j^{1/4} \sqrt{1 - \frac{\varepsilon}{p_j} \wedge 1} < 3^N.$$

Therefore, $N_{[]}(\mathcal{C}, \varepsilon, \mu) \le N$ implies that

$$N > \frac{1}{4} \log_3 m_j + \frac{1}{2} \log_3 \left(1 - \frac{\varepsilon}{p_j} \wedge 1\right)$$

for every $j \in \mathbb{N}$. Conversely, we have shown that

$$N_{[]}(\mathcal{C}, \varepsilon, \mu) \ge \sup_{j\in\mathbb{N}} \left\lfloor \frac{1}{4} \log_3 m_j + \frac{1}{2} \log_3 \left(1 - \frac{\varepsilon}{p_j} \wedge 1\right)\right\rfloor.$$

This bound holds for any choice of $(p_j)_{j\in\mathbb{N}}$.

Fix $n(\varepsilon) \uparrow \infty$ as $\varepsilon \downarrow 0$. We now choose $(p_j)_{j\in\mathbb{N}}$ such that $N_{[]}(\mathcal{C}, \varepsilon, \mu) \ge n(\varepsilon)$. First, as $m_j \uparrow \infty$, we can choose a subsequence $j(k) \uparrow \infty$ such that

$$m_{j(\lfloor \log_2(2/3\varepsilon)\rfloor)} \ge 3^{4n(\varepsilon)+6} \qquad \text{for all } 0 < \varepsilon < 1/3.$$

Now define $(p_j)_{j\in\mathbb{N}}$ as follows:

$$p_{j(k)} = 2^{-k} \quad \text{for } k \in \mathbb{N}, \qquad p_j = 0 \quad \text{for } j \notin \{j(k) : k \in \mathbb{N}\}.$$

Then we clearly have, setting $J(\varepsilon) = j(\lfloor \log_2(2/3\varepsilon) \rfloor)$,

$$N_{[]}(\mathcal{C}, \varepsilon, \mu) \geq \left\lfloor \frac{1}{4} \log_3 m_{J(\varepsilon)} + \frac{1}{2} \log_3 \left(1 - \frac{\varepsilon}{p_{J(\varepsilon)}} \wedge 1 \right) \right\rfloor \geq \lfloor n(\varepsilon) + 1 \rfloor \geq n(\varepsilon)$$

for all $0 < \varepsilon < 1/3$. This completes the proof.

Acknowledgment. The author would like to thank Terry Adams and Andrew Nobel for making available an early version of [2] and for interesting discussions.

REFERENCES

[1] ADAMS, T. M. AND NOBEL, A. B. (2010a). The gap dimension and uniform laws of large numbers for ergodic processes. Preprint arXiv:1007.2964.

[2] ADAMS, T. M. AND NOBEL, A. B. (2010b). Uniform approximation and bracketing properties of VC classes. Preprint arXiv:1007.4037.

[3] ADAMS, T. M. AND NOBEL, A. B. (2010c). Uniform convergence of Vapnik-Chervonenkis classes under ergodic sampling. *Ann. Probab.* **38**, 4, 1345–1367.

[4] ALON, N., BEN-DAVID, S., CESA-BIANCHI, N., AND HAUSSLER, D. (1997). Scale-sensitive dimensions, uniform convergence, and learnability. *J. ACM* **44**, 4, 615–631. MR1481318 (99b:68154)

[5] ALON, N., HAUSSLER, D., AND WELZL, E. (1987). Partitioning and geometric embedding of range spaces of finite Vapnik-Chervonenkis dimension. In *SCG '87: Proceedings of the third annual symposium on Computational geometry*. ACM, New York, 331–340.

[6] ASSOUAD, P. (1983). Densité et dimension. *Ann. Inst. Fourier (Grenoble)* **33**, 3, 233–282. MR723955 (86j:05022)

[7] BERTI, P. AND RIGO, P. (2007). 0-1 laws for regular conditional distributions. *Ann. Probab.* **35**, 2, 649–662. MR2308591 (2008e:60087)

[8] BOURGAIN, J., FREMLIN, D. H., AND TALAGRAND, M. (1978). Pointwise compact sets of Baire-measurable functions. *Amer. J. Math.* **100**, 4, 845–886. MR509077 (80b:54017)

[9] DUDLEY, R. M. (1999). *Uniform central limit theorems*. Cambridge Studies in Advanced Mathematics, Vol. **63**. Cambridge University Press, Cambridge. MR1720712 (2000k:60050)

[10] DUDLEY, R. M. (2002). *Real analysis and probability*. Cambridge Studies in Advanced Mathematics, Vol. **74**. Cambridge University Press, Cambridge. MR1932358 (2003h:60001)

[11] DUDLEY, R. M., GINÉ, E., AND ZINN, J. (1991). Uniform and universal Glivenko-Cantelli classes. *J. Theoret. Probab.* **4**, 3, 485–510. MR1115159 (92i:60009)

[12] GRAY, R. M. (2009). *Probability, random processes, and ergodic properties*, second ed. Springer, New York.

[13] LANDERS, D. AND ROGGE, L. (1974). On the extension problem for measures. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 30*, 167–169. MR0357722 (50 #10190)

[14] MARCZEWSKI, E. (1948). Ensembles indépendants et leurs applications à la théorie de la mesure. *Fund. Math. 35*, 13–28. MR0027313 (10,287e)

[15] MENDELSON, S. AND VERSHYNIN, R. (2003). Entropy and the combinatorial dimension. *Invent. Math.* **152**, 1, 37–55. MR1965359 (2004d:60047)

[16] OSSIANDER, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.* **15**, 3, 897–919. MR893905 (88k:60067)

[17] ROSENTHAL, H. P. (1974). A characterization of Banach spaces containing l^1. *Proc. Nat. Acad. Sci. U.S.A. 71*, 2411–2413. MR0358307 (50 #10773)

[18] RYLL-NARDZEWSKI, C. (1953). On quasi-compact measures. *Fund. Math. 40*, 125–130. MR0059997 (15,610d)

[19] TALAGRAND, M. (1984). Pettis integral and measure theory. *Mem. Amer. Math. Soc.* **51**, 307, ix+224. MR756174 (86j:46042)

[20] TOPSØE, F. (1977). Uniformity in convergence of measures. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **39**, 1, 1–30. MR0443025 (56 #1398)

[21] VAN DER VAART, A. W. AND WELLNER, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. MR1385671 (97g:60035)

[22] VON WEIZSÄCKER, H. (1983). Exchanging the order of taking suprema and countable intersections of σ-algebras. *Ann. Inst. H. Poincaré Sect. B (N.S.)* **19**, 1, 91–100. MR699981 (85c:28001)

SHERRERD HALL, ROOM 227,
PRINCETON UNIVERSITY,
PRINCETON, NJ 08544, USA.
E-MAIL: rvan@princeton.edu