Uniform Glivenko–Cantelli Classes

Constantine Caramanis *

May 10, 2000

1 Introduction

A class of sets, or functions, is said to be P–Glivenko–Cantelli if the empirical measure P_n converges in some sense to the true measure, P, as $n \to \infty$, uniformly over the class of sets or functions. Thus, the notions of Glivenko–Cantelli, and likewise uniform Glivenko–Cantelli are for the most part qualitative assessments of how "well–behaved" a collection of sets or functions is, in the sense of convergence of empirical measures. Since the ability to positively or negatively identify this property in a given family of sets or functions is quite useful, it is natural to seek a more specific, or rather, more tractably calculable criterion, or set of criteria, that is equivalent with some form of Glivenko–Cantelli.

This paper contains a discussion of the relation of measurability and capacity conditions, to various formulations of the Glivenko–Cantelli property. Specifically, we focus on weak and strong P–Glivenko–Cantelli classes, for some P given, and also on weak and strong universal and uniform Glivenko–Cantelli classes, as will be defined below.

Section 2 below will introduce the definitions we use, including those of measurability and capacity. Section three contains a discussion of Glivenko–Cantelli classes, some necessary and sufficient conditions to ensure Glivenko–Cantelli, and also the characterization of Glivenko–Cantelli given by Talagrand ([3]). Finally, section 4 contains similar results about so-called *universal* and *uniform* Glivenko–Cantelli classes. While a number of proofs are provided, some are omitted for length considerations, and instead references are given.

2 Definitions

Given some measurable space $(X, \mathcal{A})$, the set $\mathcal{F}$ of functions, or sets, is called a **universal** Glivenko–Cantelli class if and only if it is a Glivenko–Cantelli class for every probability measure $P \in \mathcal{P}(X, \mathcal{A})$. Moreover, if $\mathcal{F}$ is a universal Glivenko–Cantelli class, and furthermore the rate of convergence of the empirical measure is uniform over all $P \in \mathcal{P}(X) := \mathcal{P}(X, \mathcal{A})$:

$$\forall \varepsilon > 0,\ \exists N \text{ s.t. } \forall\, P \in \mathcal{P}(X),\ n > N \Rightarrow \Pr^*\{||P_n - P||_{\mathcal{F}} > \varepsilon\} < \varepsilon,$$

then we say that $\mathcal{F}$ is a **uniform** Glivenko–Cantelli class.

As mentioned above, we use notions of measurability and capacity in order to characterize the various versions of the Glivenko–Cantelli property. As examples below illustrate, there are still open questions as to the precise interplay between such conditions and the Glivenko–Cantelli property.

The idea of the "size" of a class of sets is captured by various definitions of capacity. One such measure of size is the so-called Vapnik–Červonenkis index of a set. This is defined as the supremum

*cmcaram@mit.edu

of cardinalities of sets shattered by our class of sets, where we say that a class $\mathcal{C}$ of sets shatters a finite set A if and only if for every subset $B \subset A$, we can write $B = A \cap C$, for some $C \in \mathcal{C}$. This paper discusses the connection between this definition and Glivenko–Cantelli classes. For a more in-depth discussion of Vapnik–Červonenkis classes and their relation to convergence of empirical measures, see, e.g. Dudley ([1]).

We are concerned with the weak or strong convergence of $||P_n - P||_{\mathcal{F}}$. However, the measurability of the cover function $||P_n - P||^*_{\mathcal{F}}$ is not always sufficient. We will see, for instance, that without stronger measurability conditions, a finite Vapnik–Červonenkis index is not sufficient to guarantee that a class $\mathcal{F}$ is Glivenko–Cantelli for all $P \in \mathcal{P}(X)$, while with some assumed measurability, a finite Vapnik–Červonenkis index is sufficient for uniform Glivenko–Cantelli. The measurability condition we use is given by the following

Definition 1 *We call a measurable space $(A, \mathcal{A})$ a* **Suslin space** *if and only if there is a Borel–measurable surjection from some Polish space Ω onto A.*

Combining this with the notion of admissibility gives us the condition we seek on the elements of $\mathcal{F}$.

Definition 2 *For $(\Omega, \mathcal{A})$ a measure space, $\mathcal{F}$ a set, and $X : \Omega \times \mathcal{F} \longrightarrow \mathbb{R}$, we call X* **image admissible Suslin via** $(Y, \mathcal{S}, T)$ *if and only if $(Y, \mathcal{S})$ is Suslin, $T : Y \to \Omega$ is a surjection, and the map*

$$(y, \omega) \mapsto X(T(y), \omega),$$

is measurable on $Y \times \Omega$. We call X **image admissible Suslin** *if it is image admissible Suslin via some $(Y, \mathcal{S}, T)$.*

We now begin to characterize the Glivenko–Cantelli property by means of the above definitions.

3 Glivenko–Cantelli

We say that $\mathcal{F}$ is a *weak* Glivenko–Cantelli class if the convergence of the empirical measures is in probability, that is, if

$$||P_n - P||^*_{\mathcal{F}} \longrightarrow 0,$$

in probability, as $n \to \infty$. On the other hand, $\mathcal{F}$ is called a *strong* Glivenko–Cantelli class if the convergence is almost sure for P.

Example 1 Weak Glivenko–Cantelli and strong Glivenko–Cantelli are not in general equivalent. Let $\mathcal{F}$ be the unit ball of the dual space H' of some infinite dimensional Hilbert space H with orthonormal basis $\{e_i\}$. Choose P on H that assigns probability c/n^2 to $k_n e_n$ and $-k_n e_n$, where $c = 3/\pi^2$ is the appropriate normalizing constant, and $k_n = n/(\log n + 1)$. Then some (non–trivial) work shows that $\mathcal{F}$ is a weak, but not strong, Glivenko–Cantelli class (see, e.g. [2]).

Under some boundedness conditions, however, these two notions do in fact coincide. Recall that a class $\mathcal{F} \subset \mathcal{L}^1(P)$ is called **order bounded** iff $F^*_{\mathcal{F}} \in \mathcal{L}^1(P)$, where

$$F_{\mathcal{F}}(x) := \sup\{|f(x)| \,:\, f \in \mathcal{F}\}.$$

We have the following, due to Talagrand ([3]):

Theorem 1 (Talagrand) *If $\mathcal{F} \subset \mathcal{L}^1(P)$, then the following are equivalent as $n \to \infty$:*

(i) $||P_n - P||^*_{\mathcal{F}} \to$ *almost surely;*

(ii) $||P_n - P||_{\mathcal{F}} \to$ *almost surely;*

(iii) $||P_n - P||^*_{\mathcal{F}} \to$ *in probability, and moreover $\mathcal{F}$ is order bounded up to additive constants, i.e. $\mathcal{F}_{0,P} := \{f - \int f\, dP \,:\, f \in \mathcal{F}\}$ is order bounded.*

Of course, it is not surprising that the added condition of boundedness is not nearly sufficient to ensure Glivenko–Cantelli. We have the following example.

Example 2 Consider the unit ball in $\mathcal{C}[0,1]$ under the supremum norm:

$$\mathcal{F} = \{f \in \mathcal{C}[0,1] \,:\, ||f||_\infty \leq 1\}.$$

This is neither a strong nor a weak Glivenko–Cantelli class (by the above theorem, since $\mathcal{F}$ is order bounded, these are equivalent) for $P = \lambda$, Lebesgue measure on the unit interval. The intuition is that the class of bounded continuous functions is quite large — more specifically, the class of sets given by the subgraphs of this class of functions is very large, in the Vapnik–Červonenkis sense.

This is made precise by the following characterization of Glivenko–Cantelli classes, due again to Talagrand ([3]), in terms of a particular notion of shattering. First we need some definitions. Suppose we have a probability space $(\Omega, \mathcal{A}, P)$. Then given any subset $B \subset \Omega$, real numbers $\alpha < \beta$, any $n \in \mathbb{N}$, and set $\mathcal{F}$, define

$$\begin{aligned} W(\mathcal{F}, B, \alpha, \beta, n) \quad := \quad & \{(x_1, \ldots, x_n, y_1, \ldots, y_n) \in B^{2n} \,:\, \exists\, f \in \mathcal{F}, \text{ such that} \\ & f(x_i) < \alpha < \beta < f(y_i), \text{ for all } i = 1, \ldots, n\}. \end{aligned}$$

Then we have the following two theorems.

Theorem 2 (Talagrand) *If $\mathcal{F}$ is order bounded for P, then it is Glivenko–Cantelli if and only if there do* **not** *exist real numbers $\alpha < \beta$, and a set $B \subset \Omega$ with $P(B) > 0$, such that*

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

for all positive integers n.

We also have the following:

Theorem 3 (Talagrand) *Suppose we have a class of sets $\mathcal{C}$ such that $\mathcal{F}_{\mathcal{C}}$, the set of corresponding indicator functions, is in turn such that for every $\alpha < \beta$ and integer n, $W(\mathcal{F}, \Omega, \alpha, \beta, n)$ is P^{2n}–measurable. Then $\mathcal{C}$ is P–Glivenko–Cantelli iff there exists no set B of positive measure such that P restricted to B is nonatomic, and furthermore for all n-tuples $x_1, \ldots, x_n$ in B, up to a set of P^n–measure zero in Ω^n, the set $\{x_1, \ldots, x_n\}$ is shattered by $\mathcal{C}$.*

From these two theorems it is immediate that the unit ball in $\mathcal{C}[0,1]$ fails to be a Glivenko–Cantelli class, precisely because while its elements are continuous, there are no restrictions on their derivatives, and thus the class can (easily) shatter any set of *distinct* n-tuples. Since the set of nondistinct n-tuples has measure zero for any nonatomic law P, the conclusion follows from the above theorem. In fact we see later that if we control the supremum norm of the first derivative almost everywhere, then in fact we obtain at least a universal Glivenko–Cantelli class.

Note that while many of the above quantities, such as the envelope function $F^*_{\mathcal{F}}$ for instance, are defined in terms of the probability measure P, the Vapnik–Červonenkis index is a purely combinatorial property. Thus it makes sense that this should be related to universal and uniform Glivenko–Cantelli classes. In the interest, then, of examining the relation between the Glivenko–Cantelli property and the Vapnik–Červonenkis index, we investigate universal and in particular, uniform Glivenko–Cantelli classes. To this we now turn.

4 Universal and Uniform Glivenko–Cantelli

First we look at universal Glivenko–Cantelli classes: classes of sets or functions that are Glivenko–Cantelli for all probability measures $P \in \mathcal{P}(\Omega)$. Then we consider the subset of these which have the added property of uniform convergence over the family of measures $\mathcal{P}(\Omega)$.

4.1 Universal Glivenko–Cantelli

In analogy to the case of a single, fixed distribution P, we have weak and strong universal Glivenko–Cantelli classes. We say that $\mathcal{F}$ is a weak universal Glivenko–Cantelli class if for every $n \in \mathbb{N}$ there exists some function,

$$a_n : \mathcal{F} \times \mathcal{P}(\Omega) \longrightarrow \mathbb{R},$$

such that we have

$$||P_n - a_n(\cdot, P)||_{\mathcal{F}}^* \longrightarrow 0,$$

in probability as $n \to \infty$. Unlike the case of a single P, the weak and strong universal Glivenko–Cantelli properties are in fact equivalent. This follows because by the following theorem, order boundedness of $\mathcal{F}$ is a *necessary* condition for $\mathcal{F}$ to be universal Glivenko–Cantelli (and recall that under this assumption, even for the case of single P, weak and strong Glivenko–Cantelli are equivalent). To see the equivalence of weak and strong universal Glivenko–Cantelli that results from the neccesity of order boundedness, suppose for the moment that the contrary holds. Suppose that $\mathcal{F}$ is a weak, but not a strong universal Glivenko–Cantelli class. Further, let us suppose that $M < \infty$ is the uniform bound on $\mathcal{F}$, that is, M is such that $\sup f - \inf f < M$ for all $f \in \mathcal{F}$. Then, by the failure of strong universal Glivenko–Cantelli, we can find some P, $\varepsilon > 0$, and sequence $n(k) \to \infty$, such that for some $g_k \in \mathcal{F}$,

$$\Big| \int g_k \, dP - a_{n(k)}(g_k, P) \Big| > \varepsilon.$$

By the assumed weak Glivenko–Cantelli property, for k sufficiently large,

$$\Pr\{\Big| \int g_k \, dP_{n(k)} - a_{n(k)}(g_k, P) \Big| > \varepsilon/2\} < \frac{1}{2}.$$

$$\text{Also,} \quad \Pr\{\Big| \int g_k \, d(P_{n(k)} - P) \Big| > \varepsilon/2\} \; < \; 4M^2/(\varepsilon^2 n(k)) < \frac{1}{2},$$

where the second inequality follows from Chebyshev's inequality. But then since each of these events occurs on sets with measure strictly less than one half, we can find a set of positive measure, and in particular a point, where neither event occurs. Applying the evaluation map at this point, and subtracting, results in a contradiction.

The following theorem shows that for $\mathcal{F}$ to be universal Glivenko–Cantelli, it must be uniformly bounded up to additive constants.

Theorem 4 *If the class of functions $\mathcal{F}$ is universal Glivenko–Cantelli, then the set*

$$\mathcal{F}_0 := \{f - \inf f \; : \; f \in \mathcal{F}\},$$

is uniformly bounded.

First we prove the following

Lemma 1 *Suppose we have a normed space $(X, || \cdot ||)$, and random elements $\{X_i\}$ that satisfy the following measurability and independence conditions:*

(i) *For each n, and for $s_i = 1$ or -1, $||s_1 X_1 + \cdots + s_n X_n||$, and $||X_n||$ are measurable,*

(ii) *The joint distribution of $||s_1 X_1 + \cdots + s_m X_m||$ for $m \leq n$, is independent of the $\{s_i\}$,*

(iii) *The random variables $||X_i||$ are i.i.d.*

Then, if $||S_n||/n \to 0$ in probability as $n \to \infty$, we also have $nP(||X_1|| > n) \to 0$, as $n \to \infty$.

PROOF. The proof is by contradiction. Under the given conditions, we can apply the Lévy inequality. Thus we have

$$\begin{aligned} P(||S_n|| \geq n/2) &\geq \frac{1}{2}P(\exists j \leq n : ||S_j|| \geq n/2) \\ &\geq \frac{1}{2}P(\exists j \leq n : ||X_j|| \geq n). \end{aligned}$$

Now, if $nP(||X_1|| > n) \nrightarrow 0$, then there exists some $\varepsilon > 0$, such that for any N, there exists some $n > N$ with $nP(||X_1|| > n) \geq \varepsilon$. But then there exists a subsequence $n_k \to \infty$ for which

$$P(||X_1|| > n_k) > \varepsilon/n_k, \quad \text{for every } k.$$

Then, since the X_i are i.i.d. by assumption,

$$P(\exists j \leq n : ||X_j|| \geq n) = 1 - (1 - \frac{\delta}{n})^n \gtrsim 1 - e^{-\varepsilon} > 0,$$

as $n \to \infty$, hence contradicting the assumption that

$$\frac{||S_n||}{n} \to 0, \quad \text{in probability.}$$

□

Now for the proof of the theorem:

PROOF. Suppose to the contrary, that $\mathcal{F}$ is a (weak) universal Glivenko–Cantelli class, but $\mathcal{F}_0$ (as given above) is not uniformly bounded. Then we can find $\mathcal{G} := \{f_k\} \subset \mathcal{F}$ and points x_k, y_k such that

$$|f_k(x_k) - f_k(y_k)| > 8^k, \text{ for } k = 1, 2, \ldots.$$

Now choose the atomic distribution P that assigns mass $2^{-(k+1)}$ to each of x_k and y_k, and let P_n, P_n' be independent empirical processes for P. Then,

$$(P_n - P_n') = \frac{1}{n}\sum_{i=1}^{n}(\delta_{X(i)} - \delta_{Y(i)}) =: \frac{1}{n}\sum_{i=1}^{n} V_i.$$

Since by assumption $\mathcal{F}$ is weak universal Glivenko–Cantelli, then surely so is $\mathcal{G}$, and thus

$$||P_n - P_n'||_{\mathcal{G}} = ||\frac{1}{n}\sum_{i=1}^{n} V_i||_{\mathcal{G}} \to 0, \text{ in probability.}$$

Applying lemma 1 above we obtain that

$$n\Pr(||V_1||_{\mathcal{G}}) > n) \to 0,$$

as $n \to \infty$. However, by our choice of P, for every k and $n = 8^k$, we also have

$$n\Pr(||V_1||_{\mathcal{G}}) > n) \geq n\Pr\{X(1) = x_k \text{ and } Y(1) = y_k\} \geq \frac{1}{4}2^k \longrightarrow \infty.$$

This contradiction concludes the proof. □

We now relate the universal Glivenko–Cantelli property to the combinatorial Vapnik–Červonenkis index.

Example 3 Consider some well-ordered uncountable set $(\Omega, \leq)$ (we can think of Ω as the least uncountable ordinal) such that the sets $I_x := \{y : y \leq x\}$ are each countable. Let our class of sets

$\mathcal{C}$ be precisely these "initial segments" I_x. Then consider any measure P that assigns measure zero to every countable set. Now, for every n, the empirical measure P_n will be of the form

$$P_n := \frac{1}{n}\sum_1^n \delta_{X_i}.$$

For any possible values of $X_1, \dots, X_n$, let x_0 be the maximum. Then

$$\sup(P_n(I_x)) \geq P_n(I_{x_0}) = 1.$$

Meanwhile, since the I_x are all countable, by our choice of P, $P(I_x) = 0$. Therefore

$$||P_n - P||_{\mathcal{C}} = 1 \;\; \forall\, n,$$

and thus $\mathcal{C}$ is not universal Glivenko–Cantelli. On the other hand, however, since $\mathcal{C}$ is linearly ordered by inclusion, we have $\mathcal{S}(\mathcal{C}) = 1$, and in particular $\mathcal{S}(\mathcal{C}) < \infty$.

This example, then, illustrates that a finite Vapnik–Červonenkis index is not sufficient to ensure that a class of sets is universal Glivenko–Cantelli. The following example demonstrates that it is also not necessary.

Example 4 Consider now any countable set X. We claim that the power-set σ–algebra $\mathcal{A} := 2^X$ is itself universal Glivenko–Cantelli. Recall that showing $\mathcal{A}$ is a weak or strong universal Glivenko–Cantelli class is equivalent. Choose any distribution P on X. Enumerate X as $\{r_1, r_2, \dots\}$, so that P assigns probability mass p_i to r_i. We want to show that $\forall\, \varepsilon > 0$,

$$\lim_{n\to\infty} \Pr(\sup_{A\in\mathcal{A}} |P_n(A) - P(A)| > \varepsilon) \to 0.$$

Given any $\varepsilon > 0$, choose M sufficiently large so that $\sum_{M+1}^{\infty} p_i \; < \varepsilon/4$. Then we have:

$$\begin{aligned}
\Pr(\;\; & \sup_{A\in\mathcal{A}} |P_n(A) - P(A)| > \varepsilon) \leq \Pr(\sup_{A\in\mathcal{A}} |P_n(A \cap \{r_1, \dots, r_M\}) \\
& \quad -P(A \cap \{r_1, \dots, r_M\})| > \varepsilon/2) + \Pr(\sup_{A\in\mathcal{A}} |P_n(A \cap \{r_{M+1}, \dots\}) \\
& \quad -P(\sup_{A\in\mathcal{A}} A \cap \{r_{M+1}, \dots\})| > \varepsilon/2) \\
& \leq \;\; \Pr(\sup_{A\in\mathcal{A}} |P_n(A \cap \{r_1, \dots, r_M\}) - P(A \cap \{r_1, \dots, r_M\})| > \varepsilon/2) \\
& \quad +\Pr(P_n(\{r_{M+1}, \dots\}) > \varepsilon/2).
\end{aligned}$$

Bounding the second term is a straightforward application of Hoeffding's inequality. As for the first term, call it (A), we again use Hoeffding's inequality to bound it:

$$(A) \leq \Pr(\sum_{j=1}^{M} |P_n(r_j) - P(r_j)| > \varepsilon/2) \leq Me^{-n\varepsilon^2/4}.$$

We have shown that for X any countable set, $\mathcal{A} := 2^X$ is universal Glivenko–Cantelli. But on the other hand, evidently, $\mathcal{S}(\mathcal{A}) = +\infty$.

Therefore, the two examples above illustrate that a finite Vapnik–Červonenkis index is neither necessary nor sufficient for the universal Glivenko–Cantelli property. However we will soon see that under measurability conditions such as those introduced in section 2, a finite Vapnik–Červonenkis index is in fact equivalent to the uniform Glivenko–Cantelli property, which of course implies the universal Glivenko–Cantelli property. We now turn our attention to the uniform Glivenko–Cantelli property.

4.2 Uniform Glivenko–Cantelli

In analogy to the definition of weak universal Glivenko–Cantelli classes, we say that some set $\mathcal{F}$ of (measurable) functions on $(\Omega, \mathcal{A})$, is *weak* uniform Glivenko–Cantelli if for every $n \in \mathbb{N}$ there exists some function,

$$a_n : \mathcal{F} \times \mathcal{P}(\Omega) \longrightarrow \mathbb{R},$$

such that for all $\varepsilon > 0$, we have

$$\sup_{P\in\mathcal{P}(\Omega,\mathcal{A})} \Pr^*\{||P_n - a_n(\cdot, P)||_{\mathcal{F}} > \varepsilon\} \longrightarrow 0,$$

as $n \to \infty$. We say that a set $\mathcal{F}$ is *strong* uniform Glivenko–Cantelli if for all $\varepsilon > 0$,

$$\sup_{P\in\mathcal{P}(\Omega,\mathcal{A})} \Pr^*\{\sup_{m\geq n} ||P_m - P||_{\mathcal{F}} > \varepsilon\} \longrightarrow 0.$$

Under measurability conditions, we will provide an entropy–related characterization of uniform Glivenko–Cantelli classes, and also demonstrate that, again, given these measurability conditions, weak and strong uniform Glivenko–Cantelli are equivalent.

If $\mathcal{F}$ is weak uniform Glivenko–Cantelli, then it must also be weak (and hence also strong) universal Glivenko–Cantelli, and hence by theorem 4 above, $\mathcal{F}$ must be uniformly bounded up to additive constants. As in the universal Glivenko–Cantelli case, we write

$$\mathcal{F}_0 := \{f - \inf f \,:\, f \in \mathcal{F}\}.$$

We have the following

Proposition 1 *The following are equivalent (and for all we assume that $\mathcal{F}$ consists of universally measurable functions):*

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

(ii) *$\mathcal{F}$ is a family of bounded functions, for which*

$$\sup_{P\in\mathcal{P}(\Omega,\mathcal{A})} Pr^*\{||P_m - P||_{\mathcal{F}} > \varepsilon\} \longrightarrow 0,$$

for all $\varepsilon > 0$.

(iii) *$\mathcal{F}$ is a family of bounded functions such that for all $r > 0$,*

$$\sup_{P\in\mathcal{P}(\Omega,\mathcal{A})} E^*||P_n - P||^r_{\mathcal{F}} = 0.$$

Note that as a consequence of this equivalence, and of theorem 4, uniform boundedness up to additive constants is implied by any of these conditions, $(i), (ii), (iii)$.

Proof. The sequence of implications $(iii) \Rightarrow (ii) \Rightarrow (i)$ is clear.
$(i) \Rightarrow (ii)$: By the triangle inequality we have

$$\begin{aligned}\sup_{P\in\mathcal{P}(\Omega,\mathcal{A})} \Pr^*\{||P_m - P||_{\mathcal{F}} > \varepsilon\} \quad &\leq \quad \sup_{P\in\mathcal{P}(\Omega,\mathcal{A})} \Pr^*\{||P_n - a_n(\cdot, P)||_{\mathcal{F}} > \varepsilon\} \\ &\quad\; + \Pr^*\{||P - a_n(\cdot, P)||_{\mathcal{F}} > \varepsilon\}.\end{aligned}$$

Since the first quantity above goes to zero by assumption, we need only prove that the second quantity goes to zero as $n \to \infty$. By the uniform boundedness implied by theorem 4, assume $\mathcal{F}_0$ is uniformly bounded by some $M < \infty$. By Chebyshev's inequality, for all $\varepsilon > 0$, $n \in \mathbb{N}$ and $f \in \mathcal{F}$,

$$\Pr\{|P_n f - Pf| \,>\, \varepsilon/2\} \leq \frac{4M^2}{n\varepsilon^2},$$

and thus for all $\delta > 0$ we can find a $N^{(1)}_{\varepsilon,\delta}$ such that if $n > N^{(1)}_{\varepsilon,\delta}$,

$$\Pr\{|P_n f - Pf| \,<\, \varepsilon/2\} > 1 - \frac{\delta}{2}.$$

From the definition of weak convergence, similarly, we can find some $N^{(2)}_{\varepsilon,\delta}$ such that if $n > N^{(2)}_{\varepsilon,\delta}$,

$$Pr\{|P_n f - a_n(f, P)| < \varepsilon/2\} > 1 - \frac{\delta}{2},$$

for all P, f. Together these imply the desired result.

$(ii) \Rightarrow (iii)$: Since $(ii) \Rightarrow (i)$, $\mathcal{F}_0$ is uniformly bounded, and we can replace $\mathcal{F}$ by $\mathcal{F}_0$ in both (ii) and (iii). Since $\Pr^*(f > \varepsilon) = \Pr(f^* > \varepsilon)$ for any $\mathbb{R}$–valued function f, and since our family is uniformly bounded up to additive constants, convergence in probability is equivalent to convergence in r^{th} moment, for any $r > 0$, and the result follows. $\square$

Recall that the strong and weak universal Glivenko–Cantelli properties are equivalent. For uniform Glivenko–Cantelli, the strong and weak formulations are equivalent if we make some additional measurability assumptions. Whether this equivalence holds without the added measurability is unknown. In addition, with the assumed measurability condition, we have a metric entropy type characterization of uniform Glivenko–Cantelli. For $x \in X^n$ we define the following functions on $\mathcal{F}_0$:

$$\begin{aligned} e_{x,p}(f,g) &:= \left[\frac{1}{n}\sum_{i=1}^{n}|f(x_i)-g(x_i)|^p\right]^{\frac{1}{p}\vee 1}, \quad 0<p<\infty; \\ e_{x,\infty}(f,g) &:= \max_{i\le n}|f(x_i)-g(x_i)|, \end{aligned}$$

and define

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

Then we have the following

Theorem 5 *If $\mathcal{F}$ is a set of bounded functions on $(X, \mathcal{A})$, and $\mathcal{F}_0$ is image admissible Suslin, then the following are equivalent:*

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

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

(iii) $\mathcal{F}_0$ is uniformly bounded, and for any $\varepsilon > 0$, and any (some) $0 < p \le \infty$,

$$\lim_{n\to\infty} H_{n,p}(\varepsilon, \mathcal{F}_0)/n = 0.$$

The proof is quite long, so we give the important ideas in each step.

Proof. Consider the condition (iii). By the assumed uniform boundedness of $\mathcal{F}_0$, for $0 < p < q < \infty$ we have

$$\begin{aligned} H_{n,p}(\varepsilon^{(p\wedge 1)/(q\wedge 1)}, \mathcal{F}_0) &\le H_{n,q}(\varepsilon, \mathcal{F}_0) \\ &\le H_{n,p}(\varepsilon^{(q\wedge 1)/(p\wedge 1)}/(2M)^{(q-p)/(p\vee 1)}, \mathcal{F}_0), \end{aligned}$$

and thus all the conditions in (iii) for $0 < p < \infty$ are equivalent. By Talagrand ([3]), (iii) for $p = 1$ implies (iii) for $p = \infty$. We can see this as follows. Take some $x = (x_1, \ldots, x_n) \in X^n$, and for $0 < \alpha < \varepsilon < 1/2$ choose some map $\pi : \mathcal{F}_0 \to \mathcal{F}_0$ such that

$$e_{x,1}(f, \pi f) \le \alpha\varepsilon/2 \quad \forall f \in \mathcal{F}_0,$$

and such that

$$\mathrm{card}\{\pi f \,:\, f \in \mathcal{F}_0\} = N(\alpha\varepsilon/2, \mathcal{F}_0, e_{x,1}).$$

By our restrictions on the map π, there are $N(\alpha\varepsilon/2, \mathcal{F}_0, e_{x,1})$ π–equivalence classes $\mathcal{E}_{[f]}$, where $g \in \mathcal{E}_{[f]}$ iff $\pi g = \pi f$. Now let $\mathcal{G} := \{f - \pi f \,:\, f \in \mathcal{F}_0\}$. Then for any equivalence class $\mathcal{E}_{[f]}$, we have

$$N(\varepsilon, \mathcal{E}_{[f]}, e_{x,\infty}) \le N(\varepsilon, \mathcal{G}, e_{x,\infty}),$$

and thus we have

$$N(\varepsilon, \mathcal{F}_0, e_{x,\infty}) \le N(\alpha\varepsilon/2, \mathcal{F}_0, e_{x,1}) \cdot N(\varepsilon, \mathcal{G}, e_{x,\infty}).$$

Now, $g \in \mathcal{G}$ iff $g = f - \pi f$ for some $f \in \mathcal{F}_0$, and therefore

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}|g(x_i)| &= \frac{1}{n}\sum_{i=1}^{n}|f(x_i) - \pi f(x_i)| \\ &= e_{x,1}(f, \pi f) \leq \alpha\varepsilon/2,\end{aligned}$$

and thus we have

$$\sum_{i=1}^{n}|g(x_i)| \leq \alpha\varepsilon n/2,$$

which implies that there are at most $m := [\alpha n]$ x_i's such that $|g(x_i)| > \varepsilon/2$. Now we approximate in the sup-norm the family $\mathcal{G}$, as follows. Let $\mathcal{H}$ be the family of functions with support on m of the n points $\{x_1, \ldots, x_n\}$, and such that they take values in $\{k\varepsilon/2 : |k| \leq 4M/\varepsilon\}$. Then $\forall\, g \in \mathcal{G}$, there is some $h \in \mathcal{H}$ that satisfies

$$\max_i |(g-h)(x_i)| < \varepsilon/2,$$

and therefore we have

$$\begin{aligned} & N(\varepsilon, \mathcal{G}, e_{x,\infty}) \leq \binom{n}{m}(1 + 8M/\varepsilon)^m \\ \Rightarrow\quad & N(\varepsilon, \mathcal{F}_0, e_{x,\infty}) \leq N(\alpha\varepsilon/2, \mathcal{F}_0, e_{x,1}) \cdot \binom{n}{m}(1 + 8M/\varepsilon)^m \\ \Rightarrow\quad & H_{n,\infty}(\varepsilon, \mathcal{F}_0) \leq \log N(\alpha\varepsilon/2, \mathcal{F}_0, e_{x,1}) + \log\binom{n}{m} + \log{(1 + 8M/\varepsilon)^m}.\end{aligned}$$

The first term on the RHS above is $H_{n,1}(\alpha\varepsilon/2, \mathcal{F}_0)$, and we know by assumption that $H_{n,1}(\varepsilon, \mathcal{F}_0)/n \to 0$ as $n \to \infty$. We use Sterling's approximation for the second term. Recall that by Sterling, we have

$$N! \sim e^{-N}N^N\sqrt{2\pi N}.$$

Therefore we have

$$\begin{aligned}\log\frac{n!}{m!(n-m)!} &\sim \log\frac{e^{-n}n^n\sqrt{2\pi n}}{e^{-n}m^m(n-m)^{(n-m)}2\pi\sqrt{mn}} \\ &= \log\frac{1}{\sqrt{2\pi m}} \cdot \left(\frac{n}{m}\right)^m \cdot \left(\frac{n}{n-m}\right)^{n-m} \\ &= \log\frac{1}{\sqrt{2\pi m}} + m\log|\alpha| + (n-m)\log|1-\alpha|.\end{aligned}$$

Thus we have

$$\begin{aligned} & \frac{1}{n}H_{n,\infty}(\varepsilon, \mathcal{F}_0) \leq \frac{1}{n}H_{n,1}(\alpha\varepsilon/2, \mathcal{F}_0) + \frac{1}{n}\log\frac{1}{\sqrt{2\pi m}} + \alpha\log|\alpha| \\ & \qquad + (1-\alpha)\log|1-\alpha| + \alpha\log(1 + 8M/\varepsilon), \\ \Rightarrow\quad & \limsup_{n\to\infty}\frac{1}{n}H_{n,\infty}(\varepsilon, \mathcal{F}_0) \leq \alpha\log|\alpha| + (1-\alpha)\log|1-\alpha| \\ & \qquad + \alpha\log(1 + 8M/\varepsilon),\end{aligned}$$

and letting $\alpha \to 0$ gives us the desired result. Finally, since (iii) for $p = \infty$ is easily seen to imply (iii) for all $0 < p < \infty$, all the conditions in (iii) are equivalent. It is also clear that (ii) implies (i). Then to conclude the theorem we will show that (iii) implies (ii), and (i) implies (iii). Since the condition in (iii) is equivalent for all $p \in (0, \infty]$, we are free in each of these steps to choose the p that is the most convenient for us.

First we show that (iii) implies (ii), for the case $p = 1$. By (iii), we have that when n is sufficiently large,

$$N(\varepsilon/8, \mathcal{F}_0, e_{x,1}) \leq \exp\{\varepsilon^2 n/(256M^2)\},$$

for all $x \in X^n$, where $M < \infty$ is the bound on $\mathcal{F}_0$, assumed to exist. Next, we look at the product space $(X, \mathcal{A}, P)^{\mathbb{N}} \otimes ([0,1], \mathcal{B}, \lambda)$, where X_i are coordinates on $(X, \mathcal{A}, P)^{\mathbb{N}}$, and $\{\varepsilon_i\}$ is a Rademacher sequence independent of the $\{X_i\}$. From here, by symmetrization, boundedness of $\mathcal{F}_0$, and also Hoeffding's inequality, we obtain the inequalities

$$\begin{aligned} \Pr\{||P_n - P||_{\mathcal{F}} > \varepsilon\} &\leq 4\Pr\left\{\Big\|\sum_{i=1}^{n} \varepsilon_i \delta_{X_i}/n\Big\|_{sF_0} > \varepsilon/4\right\} \\ &\leq 2(EN(\varepsilon/8, \mathcal{F}_0, e_{x_n(\omega),1}))\exp\{-\varepsilon^2 n/(128M^2)\}. \end{aligned}$$

This, together with the above inequality, concludes the proof that (iii) implies (ii).
To show (i) implies (iii), we choose $p = 2$. Again we can choose some finite M as the uniform bound on $\mathcal{F}_0$, this time because of the result of theorem 4. From here the result follows with quite a bit more work, using symmetrization arguments. See [2] for details. □

Indeed, measurability plays an important role here. As mentioned previously, whether the strong and weak uniform Glivenko–Cantelli properties are equivalent is unknown, if the image admissible Suslin condition is not imposed. Furthermore, in example 3 given above, the class $\mathcal{C}$ has Vapnik–Červonenkis index 1, and thus by Sauer's lemma, it satisfies

$$H_{n,p}(\varepsilon, \mathcal{C}) \leq \log(n+1), \quad 0 < p \leq \infty.$$

However, it is not universal Glivenko–Cantelli, and thus cannot possibly be uniform Glivenko–Cantelli.

On the other hand, regardless of measurability conditions, if condition (iii) in theorem 5 above holds for a class of sets $\mathcal{C}$, for $p = \infty$, then $\mathcal{C}$ is a Vapnik–Červonenkis class. And conversely, if a class $\mathcal{C}$ of sets is a Vapnik–Červonenkis class, then $H_{n,\infty}(\varepsilon, \mathcal{C}) \leq c \log n$, and in particular, condition (iii) in theorem 5 above holds for $p = \infty$. In other words, we have the following

Corollary 1 *If $\mathcal{C}$ is image admissible Suslin, then the following are equivalent:*

(i) *$\mathcal{C}$ is uniform Glivenko–Cantelli,*

(ii) *$\mathcal{C}$ is Vapnik–Červonenkis.*

Example 5 Recall example 2. There, theorem 3 shows that our set

$$\mathcal{F} = \{f \in \mathcal{C}[0,1] \,:\, ||f||_\infty \leq 1\},$$

is not a Glivenko–Cantelli class for Lebesgue measure on the unit interval, and hence neither universal nor uniform Glivenko–Cantelli. Moreover we saw that theorem 3 tells us that $\mathcal{F}$ fails to be Glivenko–Cantelli because in some sense, the class of its subgraphs is too big, as it can shatter any distinct n points on the line. We can sufficiently reduce this class by only considering continuous functions with some bound on the size of their first derivative. Requiring that a derivative exist a.e. and be uniformly bounded turns out to be enough. Consider the class

$$\mathcal{F}_M := \{f : S \to \mathbb{R} \,:\, ||f||_{BL} \leq M\}.$$

For S a separable metric space, the $\mathcal{F}_M$ are all universal Glivenko–Cantelli, and furthermore, they are each a uniform Glivenko–Cantelli class iff (S, d) is totally bounded.

That they form a universal Glivenko–Cantelli class is immediate, from the fact that the so-called dual-bounded-Lipschitz metric metrizes the convergence of laws, and a.s. convergence $P_n \Rightarrow P$ (Varadarajan's theorem).

Suppose that S is totally bounded. We show that $\mathcal{F}$ is also totally bounded for the sup-norm, and thus for any $\varepsilon > 0$, $H_{n,\infty}(\varepsilon, (\mathcal{F}_M)_0)$ is uniformly (in n) bounded, and thus as a corollary to theorem 5, $\mathcal{F}_M$ is uniform Glivenko–Cantelli. We must show then that $\mathcal{F}$ is totally bounded when S is. The idea is that once we cover S with N_ε ε-balls, a function $f \in \mathcal{F}_M$ can vary by at most εM in any ε-ball. Then partitioning $[-M, M]$ finely enough, say into $[M/\varepsilon]$ intervals of equal length, we can define $[M/\varepsilon]^{[N_\varepsilon]}$ continuous functions (not necessarily with the Lipschitz bound) such that in the sup-norm, any function in $\mathcal{F}$ is at most ε away from one of these functions. From here total boundedness follows.

Now suppose that S is not totally bounded. Then we can find an infinite set $A \subset S$ such that no two points of A are d-closer than ε apart. Then if $B \subset A$, we can define a function f that assumes the value ε on B and 0 on $A - B$, and $||f||_{BL} = 1 + \varepsilon$. Again by theorem 5, we conclude that $\mathcal{F}_{1+\varepsilon}$ is not a uniform Glivenko–Cantelli class, and hence by scaling, neither is $\mathcal{F}_M$.

5 Conclusion

The results summarized in this paper indicate that the law of large numbers type convergence of empirical measures to the true meaure, depends, essentially, on two notions of "niceness:" capacity and meaurability. Together, these two notions are sufficient, or in some cases come close to being sufficient, for characterizing the Glivenko–Cantelli classes. At the same time, however, we have given examples where a class of sets, or functions, may have measurability but be "large" in the sense of Vapnik–Červonenkis, or some other entropy sense, or vice versa, and it fails to be Glivenko–Cantelli. Conversely, we have seen that neither capacity nor measurability conditions are necessarily required for a class to be Glivenko–Cantelli.

References

[1] Dudley, R. M. "Uniform Central Limit Theorems," New York: Cambridge University Press, 1999.

[2] Dudley, R. M., Giné, E., and Zinn, J. (1991) "Uniform and universal Glivenko–Cantelli classes," *J. Theoret. Probab.* **4**, 485-510.

[3] Talagrand, M. (1987) "The Glivenko–Cantelli problem," *Ann. Probab.* **15**, 837-870.