Bernoulli **18**(4), 2012, 1310–1319
DOI: 10.3150/11-BEJ379

Uniform approximation of Vapnik–Chervonenkis classes

TERRENCE M. ADAMS[1] and ANDREW B. NOBEL[2]

[1]*Department of Defense, 9800 Savage Rd. Suite 6513, Ft. Meade, MD 20755, USA*
[2]*Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3260, USA. E-mail: nobel@email.unc.edu*

For any family of measurable sets in a probability space, we show that either (i) the family has infinite Vapnik–Chervonenkis (VC) dimension or (ii) for every $\varepsilon > 0$ there is a finite partition π such the essential π-boundary of each set has measure at most ε. Immediate corollaries include the fact that a separable family with finite VC dimension has finite bracketing numbers, and satisfies uniform laws of large numbers for every ergodic process. From these corollaries, we derive analogous results for VC major and VC graph families of functions.

Keywords: bracketing numbers; finite approximation; uniform law of large numbers; Vapnik–Chervonenkis class; VC graph class; VC major class

1. Introduction

Let $(\mathcal{X}, \mathcal{S}, \mu)$ be a probability space and let $\mathcal{C} \subseteq \mathcal{S}$ be a given family of measurable sets. The Vapnik–Chervonenkis dimension of $\mathcal{C}$ is a measure of its combinatorial complexity, specifically, the ability of $\mathcal{C}$ to separate finite sets of points. Given a finite set $D \subseteq \mathcal{X}$, let $\{C \cap D\colon C \in \mathcal{C}\}$ be the collection of subsets of D selected by the members of $\mathcal{C}$. The family $\mathcal{C}$ is said to shatter D if its elements can select every subset of D, or equivalently, if $|\{C \cap D\colon C \in \mathcal{C}\}| = 2^{|D|}$. Here and in what follows, $|A|$ denotes the cardinality of a given set A. The Vapnik–Chervonenkis (VC) dimension of $\mathcal{C}$, denoted $\dim(\mathcal{C})$, is the largest integer k such that $\mathcal{C}$ is able to shatter *some* set of cardinality k (see Vapnik and Chervonenkis [14]). If $\mathcal{C}$ can shatter arbitrarily large finite sets, then $\dim(\mathcal{C}) = +\infty$. A family of sets $\mathcal{C}$ is said to be a VC class if $\dim(\mathcal{C})$ is finite.

Let π be a finite, measurable partition of $\mathcal{X}$. For every set $C \in \mathcal{C}$, the π-boundary of C, denoted $\partial(C\colon \pi)$, is the union of all the cells in π that intersect both C and its complement with positive probability. Formally,

$$\partial(C\colon \pi) = \bigcup \{A \in \pi\colon \mu(A \cap C) \cdot \mu(A \cap C^c) > 0\}.$$

Note that $\partial(C\colon \pi)$ depends on μ, though this dependence is suppressed in our notation. We will call a family $\mathcal{C}$ *finitely approximable* if for every $\varepsilon > 0$ there exists a finite, measurable partition π of $\mathcal{X}$ such that $\mu(\partial(C\colon \pi)) \leq \varepsilon$ for every $C \in \mathcal{C}$. Our principal result is the following.

Theorem 1.1. *Let $(\mathcal{X}, \mathcal{S})$ be a measurable space. If $\mathcal{C} \subseteq \mathcal{S}$ has finite VC dimension, then $\mathcal{C}$ is finitely approximable for every probability measure μ on $(\mathcal{X}, \mathcal{S})$.*

Note that Theorem 1.1 does not impose any cardinality or regularity constraints on the family of sets $\mathcal{C}$ beyond the purely combinatorial requirement that $\mathcal{C}$ have finite VC dimension. It should be noted that there exist families $\mathcal{C}$ that are finitely approximable for all μ, but have infinite VC dimension. This is the case, for example, if $\mathcal{X}$ is a countable set, $\mathcal{S}$ contains all the subsets of $\mathcal{X}$, and $\mathcal{C}$ is the family of all finite subsets of $\mathcal{X}$. In general, the finite approximability of $\mathcal{C}$ will depend on the measure μ.

Gaenssler and Stute [7] studied π-boundaries in work concerning uniform convergence of measures. In conjunction with Theorem 1.1, their results show that, if for some separable VC-class $\mathcal{C}$ and some sequence $\{\mu_n\}$ of finite measures, $\mu_n(A) \to \mu(A)$ for every $A \in \sigma(\mathcal{C})$, then this convergence is uniform over $\mathcal{C}$. One may establish the same conclusion using Corollary 2.1 below.

Families of sets with finite VC-dimension figure prominently in machine learning, empirical process theory, and combinatorial geometry, and have been widely studied in these and other fields, see Pollard [10], van der Vaart and Wellner [12], Devroye, Györfi and Lugosi [5], Dudley [6], Vapnik [13], and Matousek [9]. The majority of this work concerns the combinatorial properties of VC-classes, and related exponential probability inequalities for uniform laws of large numbers under independent sampling (see Section 3 below). The uniform approximation guaranteed by Theorem 1.1 provides new insights into the structure of VC-classes. Some immediate corollaries of Theorem 1.1 are explored in Sections 2 and 3 below, including new results on the bracketing properties of VC major and VC graph classes of functions. For separable classes of functions with finite fat-shattering (scale sensitive) dimension (see Kearns and Schapire [8]), approximation properties analogous to those of Theorems 1.1 may be established by extending the arguments in Section 4. The Glivenko Cantelli properties of such classes are explored in Adams and Nobel [1].

The dual VC dimension. The proof of Theorem 1.1 makes use of an equivalent version of the VC dimension that we now describe. Recall that the join of k sets $A_1, \ldots, A_k \subseteq \mathcal{X}$, denoted $J = \bigvee_{i=1}^{k} A_i$, is the finite partition of $\mathcal{X}$ consisting of all nonempty intersections $\tilde{A}_1 \cap \cdots \cap \tilde{A}_k$, where $\tilde{A}_i \in \{A_i, A_i^c\}$ for $i = 1, \ldots, k$. Equivalently, J consists of the non-empty atoms of the field generated by $A_1, \ldots, A_k$. The collection $A_1, \ldots, A_k \subseteq \mathcal{X}$ is said to be Boolean independent if J has (maximal) cardinality 2^k. The dual VC dimension, denoted $\dim^*(\mathcal{C})$, is the largest k such that $\mathcal{C}$ contains k Boolean independent sets. If $\mathcal{C}$ contains Boolean independent families of every finite size, then $\dim^*(\mathcal{C}) = +\infty$. The dual VC-dimension was introduced by Assouad [4], and is so named because $\dim^*(\mathcal{C})$ is the VC-dimension of the dual family $\{D_x\colon x \in \mathcal{X}\} \subseteq 2^{\mathcal{C}}$, where $D_x = \{C \in \mathcal{C}\colon x \in C\}$. We will make use of the following, elementary result, whose proof can be found in Assouad [4]; see also Matousek [9] and Adams and Nobel [3].

Lemma 1.1. *Let $\mathcal{C}$ be any collection of subsets of $\mathcal{X}$. The VC-dimension* $\dim(\mathcal{C})$ *is finite if and only if the dual VC-dimension* $\dim^*(\mathcal{C})$ *is finite.*

In proving Theorem 1.1, we begin with the assumption that $\mathcal{C}$ is not finitely approximable for some μ, and then deduce from this that $\dim^*(\mathcal{C}) = +\infty$. Specifically, we show that for every $L \geq 1$ the family $\mathcal{C}$ contains a sub-family of L Boolean independent sets. The construction proceeds in stages. At each stage, a splitting set is produced by means of a weak limit, and is then incorporated in the construction of the splitting sets at subsequent stages. The resulting sequence of splitting sets is used to identify Boolean independent collections of arbitrary finite size.

Building in part on the results in Adams and Nobel [2,3], von Handel [15] has recently obtained general approximation results that extend those of Theorem 1.1 and Corollary 2.1 to families of real-valued functions having finite fat-shattering dimension. His principal result provides an elegant characterization of universal Glivenko–Cantelli classes.

1.1. Overview

The next two sections are devoted to corollaries of Theorem 1.1 for families of sets and functions with bounded combinatorial complexity. In Section 2, we establish that separable VC classes of sets have finite bracketing numbers, and deduce similar results for VC major and VC graph families of functions. In Section 3, we show that separable VC classes satisfy uniform laws of large numbers for every ergodic process. The proof of Theorem 1.1 is presented in Section 4.

2. Bracketing of VC classes of sets and functions

Let $\mathcal{F}$ be a family of measurable functions $f : \mathcal{X} \to \mathbb{R}$. We recall some basic definitions from the theory of empirical processes. A measurable function $F : \mathcal{X} \to [0, \infty)$ is said to be an envelope for $\mathcal{F}$ if $|f(x)| \leq F(x)$ for each $x \in \mathcal{X}$ and $f \in \mathcal{F}$. We call $\mathcal{F}$ separable if there is a countable sub-family $\mathcal{F}_0 \subseteq \mathcal{F}$ such that each function $f \in \mathcal{F}$ is a pointwise limit of a sequence of functions in $\mathcal{F}_0$. For each pair of measurable functions $g, h : \mathcal{X} \to \mathbb{R}$ with $g \leq h$, the bracket $[g, h]$ denotes the set of all measurable functions f such that $g \leq f \leq h$ pointwise on $\mathcal{X}$. In particular, $[g, h]$ is said to be an ε-bracket if $\int (h - g)\, d\mu \leq \varepsilon$. For $\varepsilon > 0$, the bracketing number $N_{[\,]}(\varepsilon, \mathcal{F}, \mu)$ of $\mathcal{F}$ is the least number of ε-brackets needed to cover $\mathcal{F}$. In general, the functions defining the minimal brackets need not be elements of $\mathcal{F}$.

2.1. VC classes of sets

Let a measure μ and family $\mathcal{C} \subseteq \mathcal{S}$ be fixed. The notions of separability and bracketing may be applied to $\mathcal{C}$ if we regard its elements as indicator functions. In this case, we may assume, without loss of generality, that the lower and upper limits of each bracket are themselves indicator functions.

Corollary 2.1. *If $\mathcal{C}$ is a separable VC-class, then $N_{[\,]}(\varepsilon, \mathcal{C}, \mu)$ is finite for every $\varepsilon > 0$.*

Proof. If $\mathcal{C}_0$ is a countable, pointwise approximating subfamily of $\mathcal{C}$, then it is straightforward to show that $N_{[\,]}(\varepsilon, \mathcal{C}, \mu) = N_{[\,]}(\varepsilon, \mathcal{C}_0, \mu)$. Thus, we assume in what follows that $\mathcal{C}$ is countable. Fix $\varepsilon > 0$. Let $\pi = \{A_1, \dots, A_m\}$ be a finite measurable partition of $\mathcal{X}$ such that $\mu(\partial(C : \pi)) < \varepsilon$ for every $C \in \mathcal{C}$, and assume without loss of generality that each cell of π has positive μ-measure. For each $C \in \mathcal{C}$, remove all points in C from A_j if $\mu(A_j \cap C) = 0$, and remove all points in C^c from A_j if $\mu(A_j \cap C^c) = 0$. Denote the resulting set by B_j. Clearly, $B_j \subseteq A_j$ and $\mu(A_j \setminus B_j) = 0$

as $\mathcal{C}$ is countable. The definition of B_j ensures that for each $C \in \mathcal{C}$ exactly one of the following relations holds: $B_j \subseteq C$, $B_j \subseteq C^c$, or $\mu(B_j \cap C) \cdot \mu(B_j \cap C^c) > 0$. Let $B_0 = \mathcal{X} \setminus \bigcup_{j=1}^m B_j$, and define the partition $\pi' = \{B_0, B_1, \ldots, B_m\}$. Given $C \in \mathcal{C}$ let $C_l = \bigcup\{B \in \pi' \colon B \subseteq C\}$ and $C_u = \bigcup\{B \in \pi' \colon B \cap C \neq \emptyset\}$. A straightforward argument shows that $C_l \subseteq C \subseteq C_u$, and that $\mu(C_u \setminus C_l) = \mu(\partial(C\colon \pi')) = \mu(\partial(C\colon \pi)) < \varepsilon$. It follows that $\Theta = \{[C_l, C_u]\colon C \in \mathcal{C}\}$ is a collection of ε-brackets covering $\mathcal{C}$. The cardinality of Θ is at most $2^{2|\pi'|}$. □

2.2. VC major families

Let $\mathcal{F}$ be a family of measurable functions $f: \mathcal{X} \to \mathbb{R}$ with envelope F. For $f \in \mathcal{F}$ and $\alpha \in \mathbb{R}$ let $L_f(\alpha) = \{x\colon f(x) \le \alpha\}$ be the α-level set of f. Define

$$\mathcal{C}_\alpha = \{L_f(\alpha)\colon f \in \mathcal{F}\}$$

to be the family of α-level sets associated with functions in $\mathcal{F}$.

Proposition 2.1. *Suppose that for every $\alpha \in \mathbb{R}$ the family $\mathcal{C}_\alpha$ is separable and* $\dim(\mathcal{C}_\alpha)$ *is finite. If μ is any probability measure on $(\mathcal{X}, \mathcal{S})$ such that $\int F \,\mathrm{d}\mu$ is finite, then $N_{[\,]}(\varepsilon, \mathcal{F}, \mu)$ is finite for every $\varepsilon > 0$.*

Proof. Suppose first that $\mathcal{F}$ is bounded, with constant envelope $M < \infty$. Fix $\varepsilon > 0$ and let K be an integer such that $2M/K \le \varepsilon$. For each $f \in \mathcal{F}$ define the approximation

$$\tilde{f}(x) = M - \frac{2M}{K} \sum_{j=1}^{K} I\bigl(x \in L_f(\alpha_j)\bigr) \qquad \text{with } \alpha_j = M - \frac{2Mj}{K}.$$

The choice of K ensures that $\tilde{f}(x) - \varepsilon \le f(x) \le \tilde{f}(x)$ for each $x \in \mathcal{X}$. The dimension of $\mathcal{C}_{\alpha_j}$ is finite by assumption, and it then follows from Corollary 2.1 that there is a finite collection Θ_j of $\varepsilon/2M$-brackets that covers the level sets $\{L_f(\alpha_j)\colon f \in \mathcal{F}\}$. For each $f \in \mathcal{F}$, let $[g_f^j, h_f^j]$ be a bracket in Θ_j containing $L_f(\alpha_j)$. With this identification, define upper and lower approximations of f as follows:

$$\tilde{f}_l = M - \frac{2M}{K} \sum_{j=1}^{K} h_f^j(x) - \varepsilon \quad \text{and} \quad \tilde{f}_u = M - \frac{2M}{K} \sum_{j=1}^{K} g_f^j(x).$$

An easy argument shows that $\tilde{f}_l \le f \le \tilde{f}_u$, and the family of brackets $\Theta = \{[\tilde{f}_l, \tilde{f}_u]\colon f \in \mathcal{F}\}$ is finite, as $|\Theta| \le \prod_{j=1}^K |\Theta_j|$. Moreover,

$$\tilde{f}_u - \tilde{f}_l \le \frac{2M}{K} \sum_{j=1}^{K} \bigl(h_f^j(x) - g_f^j(x)\bigr) + \varepsilon,$$

and therefore $\int (\tilde{f}_u - \tilde{f}_l)\,\mathrm{d}\mu \le 2\varepsilon$. Thus, Θ is a finite family of 2ε-brackets covering $\mathcal{F}$.

Suppose now that $\mathcal{F}$ has an envelope F such that $\int F\,\mathrm{d}\mu < \infty$. Given $\varepsilon > 0$ let $M < \infty$ be such that $\int_{F>M} F\,\mathrm{d}\mu < \varepsilon$. For each $f \in \mathcal{F}$ define the truncation $f_M(x) = (f(x) \vee -M) \wedge M$, and let $\mathcal{F}_M = \{f_M\colon\ f \in \mathcal{F}\}$. By the preceding argument, there is a finite family Θ of ε-brackets covering $\mathcal{F}_M$. Let $[g,h]$ be an element of Θ; without loss of generality, we may assume that $|g|, |h| \le M$. Define

$$g' = g \wedge \bigl(-FI(F > M)\bigr) \quad \text{and} \quad h' = h \vee \bigl(FI(F > M)\bigr)$$

and note that $g' \le g \le h \le h'$. Moreover, $f_M \in [g,h]$ implies $f \in [g', h']$, so the finite family of brackets $\{[g', h']\colon\ [g,h] \in \Theta\}$ covers $\mathcal{F}$. It is easy to see that

$$h' - g' = (h-g)I(F \le M) + 2FI(F > M),$$

and therefore $\int (h' - g')\,\mathrm{d}\mu \le \int (h-g)\,\mathrm{d}\mu + 2\int_{F>M} F\,\mathrm{d}\mu \le 3\varepsilon$. □

2.3. VC graph families

Let $\mathcal{F}$ be a family of measurable functions $f\colon\ \mathcal{X} \to \mathbb{R}$ with envelope $F(x)$. The graph of $f \in \mathcal{F}$ is defined by

$$G_f = \{(x,s)\colon\ x \in \mathcal{X} \text{ and } 0 \le s \le f(x) \text{ or } f(x) \le s \le 0\} \subseteq \mathcal{X} \times \mathbb{R}.$$

Let $\mathcal{G}(\mathcal{F}) = \{G_f\colon\ f \in \mathcal{F}\}$ be the family of graphs of functions in $\mathcal{F}$.

Proposition 2.2. *Suppose that $\mathcal{G}(\mathcal{F})$ is separable and that* $\dim(\mathcal{G}(\mathcal{F}))$ *is finite. If μ is any probability measure on $(\mathcal{X}, \mathcal{S})$ such that $\int F\,\mathrm{d}\mu$ is finite, then $N_{[\,]}(\varepsilon, \mathcal{F}, \mu)$ is finite for each $\varepsilon > 0$.*

Proof. Suppose first that $\mathcal{F}$ is bounded, with constant envelope $M < \infty$. The finiteness of the bracketing numbers is unaffected if we replace each function $f \in \mathcal{F}$ by $(f + M)/2M$, and we therefore assume that every $f \in \mathcal{F}$ takes values in [0, 1]. With this restriction,

$$G_f = \{(x,s)\colon\ x \in \mathcal{X} \text{ and } 0 \le s \le f(x) \le 1\} \subseteq \mathcal{X} \times [0,1].$$

Let $\lambda(\cdot)$ denote Lebesgue measure on the Borel subsets $\mathcal{B}$ of [0, 1], and define the product measure $\nu = \mu \otimes \lambda$ on $(\mathcal{X} \times [0,1], \mathcal{S} \otimes \mathcal{B})$.

Fix $\varepsilon > 0$. As $\mathcal{G}(\mathcal{F})$ is separable and has finite VC dimension, Corollary 2.1 ensures that $\mathcal{G}(\mathcal{F})$ is covered by a finite collection Θ of ε-brackets. Without loss of generality, we may represent the brackets in Θ in the form $[A, B]$, where $A, B \in \mathcal{S} \otimes \mathcal{B}$ and $A \subseteq B$. Let $[A, B]$ be a fixed bracket in Θ, and define associated functions $g, h : \mathcal{X} \to [0,1]$ as follows:

$$g(x) = \text{ess-sup}\bigl(\{s\colon\ (x,s) \in A\}\bigr) \quad \text{and} \quad h(x) = \text{ess-inf}\bigl(\{s\colon\ (x,s) \in B^c\}\bigr),$$

where for $U \subseteq [0,1]$ the essential supremum $\text{ess-sup}(U) = \inf\{\alpha\colon\ \lambda(U \cap [0,\alpha]) = \lambda(U)\}$, and $\text{ess-inf}(U)$ is defined analogously.

A straightforward argument shows that, for $v \in [0,1]$, $g^{-1}(v,1]$ is equal to the set of x such that $\int_v^1 I_A(x,s)\,\mathrm{d}\lambda(s) > 0$. It follows that $g(\cdot)$ is $\mathcal{S}$-measurable, and a similar argument shows that $h(\cdot)$ is $\mathcal{S}$-measurable as well. It is clear that $g \le h$, and that $\nu(A \setminus G_g) = \nu(B^c \setminus G_h^c) = 0$. Moreover, for every function $f : \mathcal{X} \to [0,1]$ it is easy to see that $G_f \in [A,B]$ implies $G_f \in [G_g, G_h]$, which implies in turn that $g \le f \le h$.

It follows from the arguments above that the finite family Θ_0 of brackets $[g,h]$ derived from the elements of Θ covers $\mathcal{F}$. In order to assess the size of these brackets, note that

$$(G_h \setminus G_g)_x = \{s\colon (x,s) \in G_h \setminus G_g\} = \{s\colon g(x) < s \le h(x)\},$$

and therefore by Fubini's theorem

$$\int \bigl(h(x) - g(x)\bigr)\,\mathrm{d}\mu(x) = \int \lambda\bigl((G_h \setminus G_g)_x\bigr)\,\mathrm{d}\mu(x) = \nu(G_h \setminus G_g) \le \nu(B \setminus A) \le \varepsilon.$$

Thus, every element $[g,h]$ of Θ_0 is an ε-bracket under μ.

The argument for an unbounded family $\mathcal{F}$ with an integrable envelope F is similar to that for VC Major families. Given $\varepsilon > 0$, let $M < \infty$ be such that $\int_{F>M} F\,\mathrm{d}\mu < \varepsilon$. For each $f \in \mathcal{F}$ define the truncation $f_M(x) = (f(x) \vee -M) \wedge M$, and let $\mathcal{F}_M = \{f_M\colon f \in \mathcal{F}\}$. As $G_{f_M} = G_f \cap (\mathcal{X} \times [-M,M])$, it is easy to see that the dimension of $\mathcal{G}(\mathcal{F}_M)$ is no greater than that of $\mathcal{G}(\mathcal{F})$, and is therefore finite. The preceding argument shows that there is a finite collection of ε-brackets covering $\mathcal{F}_M$, and these can be extended to 3ε-brackets covering $\mathcal{F}$ following the proof of Proposition 2.1. □

3. Uniform laws of large numbers

Let $\mathbf{X} = X_1, X_2, \ldots$ be a stationary ergodic process taking values in $(\mathcal{X}, \mathcal{S})$. The ergodic theorem ensures that for every measurable set C the sample averages $n^{-1}\sum_{i=1}^n I_C(X_i)$ converge almost surely to $P(X \in C)$. A family $\mathcal{C} \subseteq \mathcal{S}$ satisfies a uniform law of large numbers with respect to $\mathbf{X}$ if the discrepancy

$$\Delta_n(\mathcal{C}\colon \mathbf{X}) = \sup_{C \in \mathcal{C}} \left| \frac{1}{n} \sum_{i=1}^n I_C(X_i) - P(X \in C) \right|$$

is measurable and tends to zero almost surely as n tends to infinity, so that the relative frequencies of sets in $\mathcal{C}$ converge uniformly to their limiting probabilities.

For i.i.d. processes $\mathbf{X}$, Vapnik and Chervonenkis [14] gave necessary and sufficient conditions under which $\Delta_n(\mathcal{C}\colon \mathbf{X}) \to 0$, under the assumption that the discrepancies are measurable for each n. For VC-classes they established exponential inequalities of the form $\mathbb{P}(\Delta_n(\mathcal{C}\colon \mathbf{X}) > t) \le a \cdot n^{\dim(\mathcal{C})} \cdot \exp\{-bt^2\}$, where a, b are positive constants independent of $\mathbf{X}$ and $\mathcal{C}$. Thus, under appropriate measurability conditions, a VC class $\mathcal{C}$ satisfies a uniform law of large numbers for any i.i.d. process. Talagrand [11] studied necessary and sufficient conditions for uniform laws of large numbers for families of real valued functions. Restricted to families of sets, his results strengthen those of Vapnik and Chervonenkis [14]. Under appropriate measurability conditions, they imply that if the distribution of X_1 is non-atomic, then $\Delta_n(\mathcal{C}\colon \mathbf{X}) \not\to 0$ if and only if there is a set $A \in \mathcal{S}$ with $P(A) > 0$ such that with probability one $\mathcal{C}$ shatters every finite subset of $\{X_i\colon X_i \in A\}$.

Using the bracketing properties of VC classes established in the previous section, one may establish uniform laws of large numbers for separable VC classes in the general ergodic case. The following theorem appears in Adams and Nobel [3] (under an additional Polish assumption), where there is also a discussion of related work on uniform laws of large numbers under dependent sampling.

Theorem 3.1. *If $\mathcal{C}$ is a separable VC-class of sets and $\mathbf{X}$ is a stationary ergodic process, then $\Delta_n(\mathcal{C}\colon \mathbf{X}) \to 0$ almost surely as n tends to infinity.*

Proof. The stated convergence follows easily from Corollary 2.1 and standard arguments for the Blum DeHardt law of large numbers (see van der Vaart and Wellner [12] or Dudley [6]). □

One may also establish uniform laws of large numbers for separable VC major and VC graph classes of functions in the general ergodic case using the bracketing results in Propositions 2.1 and 2.2, respectively. In Adams and Nobel [3], these results are derived directly from Theorem 3.1. Related work for separable families of functions having finite fat-shattering (scale sensitive) dimension can be found in Adams and Nobel [1] and von Handel [15].

4. Proof of the main theorem

In the case where $\mathcal{X}$ is a complete separable metric space and $\mathcal{S}$ is the Borel subsets of $\mathcal{X}$, one may establish Theorem 1.1 using arguments similar to those used in Adams and Nobel [3] to establish uniform laws of large numbers for VC classes under ergodic sampling. The details can be found in Adams and Nobel [2], which contains an earlier version of the results presented here. Below we provide a simpler argument that does not require the Polish assumption. The new argument employs several simplifications and improvements that were suggested by an anonymous referee of Adams and Nobel [3], in particular, the use of Hilbert space weak limits in the definition of splitting sets.

4.1. Proof of Theorem 1.1

It follows from standard results on the L_p-covering numbers of VC classes (e.g., Theorem 2.6.4 of van der Vaart and Wellner [12]) that there exists a countable sub-family $\mathcal{C}_0$ of $\mathcal{C}$ such that $\inf_{C' \in \mathcal{C}_0} \mu(C' \triangle C) = 0$ for each $C \in \mathcal{C}$. An elementary argument then shows that

$$\sup_{C \in \mathcal{C}} \mu\bigl(\partial(C\colon \pi)\bigr) = \sup_{C \in \mathcal{C}_0} \mu\bigl(\partial(C\colon \pi)\bigr)$$

for every finite partition π, and we may therefore assume that $\mathcal{C}$ is countable. Let $\mathcal{C} = \{C_1, C_2, \ldots\}$ and let $\mathcal{S}_0 = \sigma(\mathcal{C}) \subseteq \mathcal{S}$ be the sigma field generated by $\mathcal{C}$. Suppose that the uniform approximation property fails to hold for $\mathcal{C}$. Then there exists a number $\eta > 0$ such that

$$\sup_{C\in\mathcal{C}} \lambda\big(\partial(C\colon \pi)\big) > \eta \qquad \text{for every finite measurable partition } \pi. \tag{4.1}$$

Using this relationship, we will construct a sequence of "splitting sets" $S_1, S_2, \ldots \subseteq \mathcal{X}$ in a stage-wise fashion. At the kth stage the splitting set S_k is obtained from a sequential procedure that makes use of the splitting sets $S_1, \ldots, S_{k-1}$ produced at previous stages. The splitting sets are used to identify arbitrarily large finite collections of sets in $\mathcal{C}$ having full join. The existence of these collections implies that $\mathcal{C}$ has infinite VC dimension by Lemma 1.1.

First stage. Define the refining sequence of joins $J_1(n) = C_1 \vee \cdots \vee C_n$ for $n \geq 1$. It follows from (4.1) that for each n there is a set $C_1(n) \in \mathcal{C}$ whose boundary $G_1(n) = \partial(C_1(n)\colon J_1(n))$ has measure greater than η. Note that the sets $\{G_1(n)\colon n \geq 1\}$ are measurable $\mathcal{S}_0$. By standard results in functional analysis, there exists a subsequence $\{n_m\}$ of the positive integers and an $\mathcal{S}_0$-measurable function h_1 such that $\int g I_{G_1(n_m)}\,\mathrm{d}\mu \to \int g h_1\,\mathrm{d}\mu$ as m tends to infinity for every function $g \in L_2(\mathcal{X}, \mathcal{S}_0, \mu)$. The function h_1 is the weak limit of the indicator functions $\{I_{G_1(n_m)}\}$. It follows that $0 \leq h_1 \leq 1$ almost surely, and that $\int h_1\,\mathrm{d}\mu \geq \eta$. Define the splitting set $S_1 = \{h_1 > 0\}$ and note that $\mu(S_1) \geq \eta$.

Subsequent stages. Suppose now that we have constructed splitting sets S_j at stages $j = 1, \ldots, k-1$, and wish to construct a splitting set S_k at stage k. Begin by defining the refining sequence of joins $J_k(n) = S_1 \vee \cdots \vee S_{k-1} \vee C_1 \vee \cdots \vee C_n$, for $n \geq 1$. Note that $J_k(n)$ includes the splitting sets produced at previous stages. It follows from (4.1) that for each n there is a set $C_k(n) \in \mathcal{C}$ whose boundary $G_k(n) = \partial(C_k(n)\colon J_k(n))$ has measure greater than η. Let $\{n_m\}$ be the subsequence used to define S_{k-1}. Taking a further subsequence if necessary, we may assume that the indicator functions $\{I_{G_k(n_m)}\}$ have a weak limit $h_k \in L_2(\mathcal{X}, \mathcal{S}_0, \mu)$ such that $0 \leq h_k \leq 1$ almost surely, and $\int h_k\,\mathrm{d}\mu \geq \eta$. Define the splitting set $S_k = \{h_k > 0\}$ and note that $\mu(S_k) \geq \eta$.

Construction of full joins. Fix an integer $L \geq 1$. As the measure of each splitting set S_k is at least η, there exist positive integers $k_1 < k_2 < \cdots < k_{L+1}$ such that $\mu(\bigcap_{j=1}^{L+1} S_{k_j}) > 0$. Suppose for simplicity, and without loss of generality, that $k_j = j$. For $l = 1, \ldots, L+1$ define

$$Q_l = \bigcap_{j=1}^{l} S_j.$$

In what follows, we will make repeated use of the elementary fact that $\int_B (h_1 \cdots h_l)\,\mathrm{d}\mu > 0$ if and only if $\mu(B \cap Q_l) > 0$. We claim that there exist sets $D_1, \ldots, D_L \in \mathcal{C}$ such that for each $l = 1, \ldots, L$,

$$\int_B (h_1 \cdots h_l)\,\mathrm{d}\mu > 0 \qquad \text{for every } B \in D_l \vee \cdots \vee D_L. \tag{4.2}$$

The inequalities (4.2) are established by reverse induction, beginning with the case $l = L$. To this end, note that

$$0 < \int (h_1 \cdots h_{L+1})\,\mathrm{d}\mu = \lim_{m\to\infty} \int (h_1 \cdots h_L) I_{G_{L+1}(n_m)}\,\mathrm{d}\mu,$$

and therefore $\mu(Q_L \cap G_{L+1}(n_m)) > 0$ for all m sufficiently large. Fix such an m and let $D = C_{L+1}(n_m)$. It follows from the definition of $G_{L+1}(n_m)$ that for some cell $A \in J_{L+1}(n_m)$,

$$\mu(Q_L \cap A) > 0 \quad \text{and} \quad \mu(A \cap D) \cdot \mu(A \cap D^c) > 0. \tag{4.3}$$

The presence of the sets $S_1, \ldots, S_L$ in the definition of the joins $J_{L+1}(n)$ ensures that Q_L is a finite union of cells of $J_{L+1}(n_m)$. The first relation in (4.3) then implies that A is necessarily a subset of Q_L, and it follows from the second relation that $\mu(Q_L \cap D) \cdot \mu(Q_L \cap D^c) > 0$. Letting $D_L = D$ the last inequality implies (4.2) in the case $l = L$.

Suppose now that for some $1 < l < L$ we have identified sets $D_l, D_{l+1}, \ldots, D_L$ such that (4.2) holds. Then for each cell B in the join $D_l \vee \cdots \vee D_L$,

$$0 < \int_B (h_1 \cdots h_l) \,\mathrm{d}\mu = \lim_{m\to\infty} \int_B (h_1 \cdots h_{l-1}) I_{G_l(n_m)} \,\mathrm{d}\mu.$$

Thus, there exists an integer m_0 such that $\mu(B \cap Q_{l-1} \cap G_l(n_m)) > 0$ for every $B \in D_l \vee \cdots \vee D_L$ and every $m \geq m_0$. As the join $J_l(n_m)$ includes the first n_m elements of $\mathcal{C}$, there is an $m \geq m_0$ sufficiently large that $J_l(n_m)$ includes $D_l, \ldots, D_L$. Let $D = C_l(n_m)$ and let B be any cell of $D_l \vee \cdots \vee D_L$. The definition of $G_l(n_m)$ and the choice of m together imply that for some cell $A \in J_l(n_m)$,

$$\mu(B \cap Q_{l-1} \cap A) > 0 \quad \text{and} \quad \mu(A \cap D) \cdot \mu(A \cap D^c) > 0. \tag{4.4}$$

Both Q_{l-1} and B are equal to a union of cells of the partition $J_l(n_m)$, so the first relation in (4.4) implies that $A \subseteq B \cap Q_{l-1}$. It then follows from the second relation that $\mu(B \cap Q_{l-1} \cap D)$ and $\mu(B \cap Q_{l-1} \cap D^c)$ are positive. As these inequalities hold for each $B \in D_l \vee \cdots \vee D_L$, we have $\int_{B'} (h_1 \cdots h_{l-1}) \,\mathrm{d}\mu > 0$ for every $B' \in D \vee D_l \vee \cdots \vee D_L$. Letting $D_{l-1} = D$ completes the induction, and the proof of (4.2).

It follows from (4.2) that the sets $D_1, \ldots, D_L$ have full join, and as $L \geq 1$ was arbitrary, Lemma 1.1 implies that $\mathcal{C}$ has infinite VC dimension. This completes the proof of the theorem.

Remark. An inspection of the proof shows that the approximating partitions π in the theorem can be taken to be measurable $\sigma(\mathcal{C})$. A simple counterexample shows that π may not be chosen from the smaller family $\bigcup_{n=1}^\infty \sigma(C_1 \vee C_2 \vee \cdots \vee C_n)$. Let $\mathcal{X} = [0, 1]$ and let μ be Lebesgue measure. Let $a_1, a_2, \ldots$ be a sequence of positive real numbers such that $s = \sum_{n=1}^\infty a_n < 1$. Define $s_0 = 0$ and $s_n = \sum_{i=1}^n a_i$ for $n \geq 1$, and let $C_n = [s_{n-1}, s_n)$. Clearly, the VC-dimension of the class $\{C_1, C_2, \ldots\}$ equals 1, since the sets are disjoint. Define $J_n = C_1 \vee C_2 \vee \cdots \vee C_n$. Then the set $A_n = [s_n, 1]$ is a single element in J_n with measure $1 - s_n > 1 - s > 0$. Moreover, both $A_n \cap C_{n+1}$ and $A_n \cap C'_{n+1}$ have positive measure. Thus, for $n \geq 1$, $A_n \subseteq \partial(C_{n+1}\colon G_n)$ and $\mu(\partial(C_{n+1}\colon G_n)) > 1 - s$.

Acknowledgements

The authors are indebted to an anonymous referee of the earlier paper Adams and Nobel [3], who suggested the general form of Theorem 1.1, and whose detailed comments led to a simpler and more general proof. The authors would also like to acknowledge helpful discussions with Ramon van Handel, who provided feedback on an earlier version of this work, Adams and Nobel [2]. The work presented in this paper was supported in part by NSF Grant DMS-09-07177.

References

[1] Adams, T.M. and Nobel, A.B. (2010). The gap dimension and uniform laws of large numbers for ergodic processes. Preprint. Available at arXiv:1007.2964v1.

[2] Adams, T.M. and Nobel, A.B. (2010). Uniform approximation and bracketing properties of VC classes. Preprint. Available at arXiv:1007.4037v1.

[3] Adams, T.M. and Nobel, A.B. (2010). Uniform convergence of Vapnik–Chervonenkis classes under ergodic sampling. *Ann. Probab.* **38** 1345–1367. MR2663629

[4] Assouad, P. (1983). Densité et dimension. *Ann. Inst. Fourier* (*Grenoble*) **33** 233–282. MR0723955

[5] Devroye, L., Györfi, L. and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. *Applications of Mathematics* (*New York*) **31**. New York: Springer. MR1383093

[6] Dudley, R.M. (1999). *Uniform Central Limit Theorems*. *Cambridge Studies in Advanced Mathematics* **63**. Cambridge: Cambridge Univ. Press. MR1720712

[7] Gaenssler, P. and Stute, W. (1976). On uniform convergence of measures with applications to uniform convergence of empirical distributions. In *Empirical Distributions and Processes* (*Selected Papers*, *Meeting on Math*. *Stochastics*, *Oberwolfach*, 1976). *Lecture Notes in Math*. **566** 45–56. Berlin: Springer. MR0433534

[8] Kearns, M.J. and Schapire, R.E. (1994). Efficient distribution-free learning of probabilistic concepts. *J. Comput. System Sci*. **48** 464–497. MR1279411

[9] Matoušek, J. (2002). *Lectures on Discrete Geometry*. *Graduate Texts in Mathematics* **212**. New York: Springer. MR1899299

[10] Pollard, D. (1984). *Convergence of Stochastic Processes*. *Springer Series in Statistics*. New York: Springer. MR0762984

[11] Talagrand, M. (1987). The Glivenko–Cantelli problem. *Ann. Probab.* **15** 837–870. MR0893902

[12] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. *Springer Series in Statistics*. New York: Springer. MR1385671

[13] Vapnik, V.N. (2000). *The Nature of Statistical Learning Theory*, 2nd ed. *Statistics for Engineering and Information Science*. New York: Springer. MR1719582

[14] Vapnik, V.N. and Chervonenkis, A.Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl*. **16** 264–280.

[15] von Handel, R. (2011). The universal Glivenko–Cantelli property. Preprint. Available at ArXiv: 1009.4434v3.

Received November 2010 and revised May 2011