A Generalization of Sauer's Lemma

David Haussler*
Computer Science Department
UC Santa Cruz
Santa Cruz, CA 95064

Philip M. Long*
Computer Science Department
UC Santa Cruz
Santa Cruz, CA 95064

*The authors gratefully acknowledge the support of ONR grant N00014-86-K-0454.

Abstract

We generalize Sauer's lemma to multivalued functions, proving tight bounds on the cardinality of subsets of $\prod_{i=1}^{m}\{0,...,N_m\}$ which avoid certain patterns. In addition, we give an application of this result, bounding the uniform rate of convergence of empirical estimates of the expectations of a set of random variables to their true expectations.

1 Introduction

Let $\boldsymbol{N}$ denote the positive integers, Z^+ denote the nonnegative integers and Z denote the integers. Let $[0] = \emptyset$ and for each $m \in \boldsymbol{N}$, let $[m]$ be the set $\{1, ..., m\}$.

We begin by stating Sauer's result [17]. Vapnik and Chervonenkis independently proved a similar lemma in [24].

Let $m \in Z^+$. Let F be a family of subsets of $[m]$. If I is a subset of $[m]$, we say that I is shattered by F if and only if

$$\{f \cap I : f \in F\} = 2^I. \tag{1}$$

The Vapnik-Chervonenkis(VC) dimension of F [24] is the cardinality of the largest subset of $[m]$ shattered by F.

Theorem 1 ([17]) *If the VC-dimension of F is d, then*

$$|F| \leq \sum_{i=0}^{d} \binom{m}{i} \tag{2}$$

and this bound is tight; i.e., for all $d, m \in Z^+, d \leq m$, there exists $F \subseteq 2^{[m]}$ of VC-dimension d that meets this upper bound.

In this paper, we look at some generalizations of the above definition of dimension and of Theorem 1.

Following [8], let $(m, k) \rightarrow (n, l)$ denote the statement: If $F \subseteq 2^{[m]}$, $|F| = k$, then there exists $I \subseteq [m]$ such that $|I| = n$ and

$$|\{f \cap I : f \in F\}| \geq l. \tag{3}$$

Sauer's result can now be stated as

$$\left(m, 1 + \sum_{i=0}^{d-1} \binom{m}{i}\right) \rightarrow (d, 2^d). \tag{4}$$

Proofs of other statements of the form $(m, k) \rightarrow (n, l)$ and related results are given in [3] [5] [6] [8] [11] [10] [20].

Let $m \in Z^+$. Let $N_i \in \boldsymbol{N}, 1 \leq i \leq m$. Let

$$F \subseteq X = \prod_{i=1}^{m} \{0, ..., N_i\}. \tag{5}$$

Note that when $N_i = 1$ for all $i, 1 \leq i \leq m$, F is essentially a family of subsets of $[m]$, as in Sauer's lemma. For $\bar{f} \in X$, denote by f_i the ith coordinate of $\bar{f}$, and similarly for all cartesian products used in the paper.

Let $I = \{i_1, ..., i_k\} \subseteq [m]$, with $i_1 < i_2 < \cdots < i_k$. Define

$$X_{|_I} = \prod_{j=1}^{k} \{0, ..., N_{i_j}\}. \tag{6}$$

For each $\bar{f} \in X$, let

$$\bar{f}_{|_I} = (f_{i_1}, ..., f_{i_k}). \tag{7}$$

Define

$$F_{|_I} = \{\bar{f}_{|_I} : \bar{f} \in F\}. \tag{8}$$

Suppose that we extend Sauer's definition of shattering to say that F shatters I if and only if $F_{|_I} = X_{|_I}$. Generalizations of Sauer's result using this extension of the definition of shattering are given in [2] [13] [19] [4]. Unfortunately, the bounds obtained grow exponentially with m. For applications such as that given in

section 3, a generalization of shattering which gives rise to bounds on $|F|$ that grow polynomially in m is desirable. We extend Sauer's result to some such generalizations which were given in [14] [15].[1]

Choose $\phi : \boldsymbol{N} \times \boldsymbol{N} \to \{0, 1, \star\}$. Extend ϕ to

$$\bigcup_{k \in \boldsymbol{N}} \boldsymbol{N}^k \times \boldsymbol{N}^k \tag{9}$$

by defining

$$\phi(\bar{x}, \bar{y}) = (\phi(x_1, y_1), ..., \phi(x_k, y_k)). \tag{10}$$

Extend ϕ further to $2^{\boldsymbol{N}^k} \times \boldsymbol{N}^k$ by defining

$$\phi(S, \bar{y}) = \{\phi(\bar{s}, \bar{y}) : s \in S\}. \tag{11}$$

We say that I is ϕ-shattered by F if there exists $\bar{y} \in X_{|_I}$ such that

$$\{0, 1\}^{|I|} \subseteq \phi(F_{|_I}, \bar{y}). \tag{12}$$

We say that $\bar{y}$ witnesses F's ϕ-shattering of I and that $\bar{f} \in F$ satisfies $\bar{b} \in \{0, 1\}^{|I|}$ if and only if $\phi(\bar{f}_{|_I}, \bar{y}) = \bar{b}$. The ϕ-dimension of F is the cardinality of the largest subset of $[m]$ shattered by F.

We say $F \subseteq X$ Pollard-shatters (P-shatters) I if F ϕ-shatters I with ϕ defined by

$$\phi(i, j) = \begin{cases} 1 & \text{if } i \geq j \\ 0 & \text{if } i < j \end{cases} \tag{13}$$

and define the Pollard-dimension (hereafter called the P-dimension) of F to be its ϕ-dimension in this case. We denote the P-dimension of F by P-dim(F). This definition is discussed in [12][15][16]. It is called the pseudo dimension in [16] and the combinatorial dimension in [12].

Graph-shattering (G-shattering) and the Graph-dimension (G-dimension) are defined similarly with ϕ defined by

$$\phi(i, j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} . \tag{14}$$

The G-dimension of F is denoted by G-dim(F). This definition is treated in [14].

For the purpose of bounding the cardinality of sets of a given dimension using either of the previous two definitions, we define the GP-dimension of F to be its ϕ-dimension when ϕ is given by

$$\phi(i, j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i < j \\ \star & \text{if } i > j \end{cases} \tag{15}$$

with the corresponding definition for GP-shattering. The GP-dimension of F is denoted by GP-dim(F).

Note that if $N_i = 1$ for all $i, 1 \leq i \leq m$, all of these dimensions reduce to the VC-dimension. This can be seen by viewing the subsets of $[m]$ in the definition of VC-dimension as elements of $\{0, 1\}^m$. Then $F \subseteq \{0, 1\}^m$ shatters $I \subseteq [m]$ exactly when $\{0, 1\}^{|I|} \subseteq F_{|_I}$. It can now easily be verified that, for each function ϕ given above, F is shattered exactly when F is ϕ-shattered.

We now look at a fourth generalization of the VC-dimension.

Choose $\gamma : \boldsymbol{N}^3 \to \{0, 1, \star\}$. Extend γ to

$$\bigcup_{k \in \boldsymbol{N}} 2^{\boldsymbol{N}^k} \times \boldsymbol{N}^k \times \boldsymbol{N}^k \tag{16}$$

as above. We say that I is γ-shattered by F if there exist $\bar{y}, \bar{z} \in X_{|_I}$ such that for all $i, 1 \leq i \leq k$, $y_i < z_i$ and

$$\{0, 1\}^{|I|} = \gamma(F_{|_I}, \bar{y}, \bar{z}). \tag{17}$$

[1] Bounds polynomial in m and N for yet another generalization of the VC-dimension are implicit in [1], research pursued independently from ours. It appears, however, that the bounds of this paper are more useful when applying the results of section 3 (see e.g. [12]).

Here we say that $\bar{y}$ and $\bar{z}$ witness F's γ-shattering of I and say $\bar{f} \in F$ satisfies $\bar{b} \in \{0,1\}^{|I|}$ if and only if

$$\gamma(\bar{f}_{|_I}, \bar{y}, \bar{z}) = \bar{b}. \tag{18}$$

We say F N-shatters I if F γ-shatters I with γ given by

$$\gamma(i,j,l) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i = l \\ \star & \text{otherwise} \end{cases}. \tag{19}$$

The Natarajan-dimension (or N-dimension) of F is defined to be its γ-dimension in this case. This definition appears in [14]. The N-dimension also reduces to the VC-dimension when $N_i = 1$ for all $1 \leq i \leq m$.

Define

$$P_{\max}(d, m, N_1, ..., N_m) = \max\{|F| : F \subseteq \prod_{i=1}^{m} \{0, ..., N_i\}, \text{P-dim}(F) \leq d\} \tag{20}$$

$$G_{\max}(d, m, N_1, ..., N_m) = \max\{|F| : F \subseteq \prod_{i=1}^{m} \{0, ..., N_i\}, \text{G-dim}(F) \leq d\} \tag{21}$$

$$GP_{\max}(d, m, N_1, ..., N_m) = \max\{|F| : F \subseteq \prod_{i=1}^{m} \{0, ..., N_i\}, \text{GP-dim}(F) \leq d\} \tag{22}$$

$$N_{\max}(d, m, N_1, ..., N_m) = \max\{|F| : F \subseteq \prod_{i=1}^{m} \{0, ..., N_i\}, \text{N-dim}(F) \leq d\} \tag{23}$$

It is easily verified that if a set F N-shatters a set, it also GP-shatters it, and if F GP-shatters a set, it also G-shatters it and P-shatters it. This implies that

$$\text{N-dim}(F) \leq \text{GP-dim}(F) \tag{24}$$

$$\text{GP-dim}(F) \leq \text{P-dim}(F) \tag{25}$$

$$\text{GP-dim}(F) \leq \text{G-dim}(F) \tag{26}$$

which in turn implies that

$$P_{\max}(d, m, N_1, ..., N_m) \leq GP_{\max}(d, m, N_1, ..., N_m) \tag{27}$$

$$G_{\max}(d, m, N_1, ..., N_m) \leq GP_{\max}(d, m, N_1, ..., N_m) \tag{28}$$

$$GP_{\max}(d, m, N_1, ..., N_m) \leq N_{\max}(d, m, N_1, ..., N_m) \tag{29}$$

for all relevant $d, m \in Z^+, N_1, ..., N_m \in \boldsymbol{N}$.

Our main result is stated below, and will be proved in the following section. In the following, for each $i, m \in Z^+$, let $\Gamma_{mi} \subseteq 2^{[m]}$ be defined by

$$\Gamma_{mi} = \{S \subseteq [m] : |S| = i\}. \tag{30}$$

Theorem 2 *For all $d, m \in Z^+, N_1, ..., N_m \in \boldsymbol{N}$ such that $d \leq m$,*

$$G_{\max}(d, m, N_1, ..., N_m) = \tag{31}$$

$$P_{\max}(d, m, N_1, ..., N_m) = \tag{32}$$

$$GP_{\max}(d, m, N_1, ..., N_m) = \sum_{i=0}^{d} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} N_k. \tag{33}$$

When there is an $N \in \boldsymbol{N}$ such that $N_i = N$ for all $i, 1 \leq i \leq m$, we obtain the following corollary, which is useful for obtaining learning results such as those in [12].

Corollary 3 *Let $d, m \in Z^+, N \in \boldsymbol{N}$ be such that $d \leq m$. Let*

$$F \subseteq \{0, ..., N\}^m \tag{34}$$

such that F has G-, P- or GP-dimension no greater than d. Then

$$|F| \leq \sum_{i=0}^{d} \binom{m}{i} N^i. \tag{35}$$

Proof: Follows from Theorem 2 by substituting N for each N_k and collecting terms. □

Using similar techniques, we can establish the following.

Theorem 4 *For all $d, m \in Z^+, N_1, ..., N_m \in \boldsymbol{N}$ such that $d \leq m$,*

$$\sum_{i=0}^{d} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} N_k \leq N_{\max}(d, m, N_1, ..., N_m) \tag{36}$$

$$\leq \sum_{i=0}^{d} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} \binom{N_k + 1}{2}. \tag{37}$$

This gives a result similar to that obtained by Natarajan [14] in the special case above.

Corollary 5 *Let $d, m \in Z^+, N \in \boldsymbol{N}$ be such that $d \leq m$. Let*

$$F \subseteq \{0, ..., N\}^m \tag{38}$$

such that F has N-dimension no greater than d. Then

$$|F| \leq \sum_{i=0}^{d} \binom{m}{i} \binom{N+1}{2}^i \tag{39}$$

Note that both Corollary 3 and Corollary 5 give Sauer's result (Theorem 1) in the case $N = 1$.

2 Proofs of the results

We begin by exhibiting large sets of a given G-, P-, GP-, and N-dimension.

Lemma 6 *Let $d, m \in Z^+, N_1, ..., N_m \in \boldsymbol{N}$ be such that $d \leq m$. Then there exists*

$$F \subseteq X = \prod_{i=1}^{m} \{0, ..., N_i\} \tag{40}$$

such that F has G-, P-, GP-, and N-dimension d and

$$|F| = \sum_{i=0}^{d} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} N_k. \tag{41}$$

Proof: Define F to be all the elements of X with at most d nonzero entries. We claim F has G-, P-, GP- and N-dimension d, and $|F|$ is as given above.

To prove that the G-, P-, GP- and N-dimensions of F are all no greater than d, it is sufficient to prove that G-dim$(F) \leq d$ and P-dim$(F) \leq d$, since as discussed above

$$\text{N-dim}(F) \leq \text{GP-dim}(F) \tag{42}$$

$$\text{GP-dim}(F) \leq \text{P-dim}(F) \tag{43}$$

$$\text{GP-dim}(F) \leq \text{G-dim}(F). \tag{44}$$

First, we show that G-dim$(F) \leq d$. Assume G-dim$(F) > d$ for contradiction. Let $\bar{y}$ witness F's G-shattering of I with $|I| = k > d$. Form $\bar{b} \in \{0,1\}^k$ by

$$b_i = \begin{cases} 0 & \text{if } y_i = 0 \\ 1 & \text{otherwise} \end{cases} . \tag{45}$$

Let $\bar{f} \in F$ satisfy $\bar{b}$. Let $\bar{g} = \bar{f}_{|I}$. By definition of G-shattering, we have $g_i \neq y_i$ if $y_i = 0$ and $g_i = y_i$ if $y_i \neq 0$, so $g_i \neq 0$ for all i, which implies $f_j \neq 0$ for all $j \in I$ which contradicts the definition of F, since $|I| > d$.

Next, we need to show that P-dim$(F) \leq d$. Again, assume P-dim$(F) > d$ for contradiction. Let $\bar{y}$ witness F's P-shattering of $I = \{i_1, ..., i_k\}$ with $|I| = k > d$. Let $\bar{f} \in F$ satisfy $(0, 0, ..., 0)$. Since $y_j > f_{i_j}$ for all $j, 1 \leq j \leq k$, we have $y_j > 0$ for all $j, 1 \leq j \leq k$. Let $\bar{g} \in F$ satisfy $(1, 1, ..., 1)$. Since $g_{i_j} \geq y_j$ for all $j, 1 \leq j \leq k$, we have $g_{i_j} > 0$ for all $j, 1 \leq j \leq k$, which again contradicts the definition of F.

We can see that the G-, P-, GP- and N-dimensions of F are all no less than d, since for each of the definitions of shattering, any subset I of d elements of $[m]$ is shattered, since it is trivially N-shattered (taking $\bar{y} = (0, 0, ..., 0)$, $\bar{z} = (1, 1, ..., 1)$), and as discussed previously, the N-shattering of I implies its G-, P- and GP-shattering.

We can see that F's cardinality is as given in the lemma by breaking the elements of F up into subsets consisting of the elements with exactly i non-zero elements, $0 \leq i \leq d$, and for each i further breaking these up according to which i elements are nonzero. □

For our next lemma, we give an upper bound on the cardinality of sets of a given GP-dimension, and thereby that of sets of a given G- or P-dimension. Our argument is a generalization of that given by Sauer in [17], and is similar to Natarajan's generalization of this argument in [14].

Lemma 7 *Let $d, m \in Z^+, N_1, ..., N_m \in \boldsymbol{N}$ be such that $d \leq m$. Let*

$$F \subseteq X = \prod_{i=1}^{m} \{0, ..., N_i\} \tag{46}$$

be such that GP-dim$(F) \leq d$. Then

$$|F| \leq \sum_{i=0}^{d} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} N_k. \tag{47}$$

Proof: Our proof is by double induction on m and d.

First we consider the case in which $d = 0$. Here, the bound (47) reduces to $|F| \leq 1$. If $|F| > 1$, then F must have two distinct elements f and g. Let i be an index on whose entry f and g differ. Then $\{i\}$ is shattered by F, so the GP-dimension of F is at least 1, which contradicts the assumption that $d = 0$, so $|F| \leq 1$ and the lemma holds.

Next, suppose that $d = m$. By partitioning the elements of the domain as discussed above, we can see that

$$|X| \leq \sum_{i=0}^{m} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} N_k. \tag{48}$$

so since $F \subseteq X$, certainly

$$|F| \leq \sum_{i=0}^{m} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} N_k, \tag{49}$$

establishing the result in this case.

Now, choose $d, m \in Z^+$ such that $0 < d < m$. Define $\pi : X \rightarrow \prod_{i=1}^{m-1} \{0, ..., N_i\}$ by

$$\pi(\bar{f}) = (f_1, ..., f_{m-1}). \tag{50}$$

Define

$$\alpha : \pi(F) \rightarrow \{0, ..., N_m\} \tag{51}$$

by

$$\alpha(w_1, ..., w_{m-1}) = \min\{v : (w_1, ..., w_{m-1}, v) \in F\} \tag{52}$$

Define

$$F_- = \{(f_1, ..., f_{m-1}, \alpha(f_1, ..., f_{m-1})) : \bar{f} \in F\} \tag{53}$$

and for each $n \in \boldsymbol{N}, 1 \leq n \leq N_m$, define

$$F_n = \{\bar{f} \in F - F_- : f_m = n\}. \tag{54}$$

Since the above sets are disjoint and their union is all of F, we have

$$|F| = |F_-| + \sum_{n=1}^{N_m} |F_n|. \tag{55}$$

Let us make the inductive assumption that the bound (47) holds for all sets F of vectors of $m-1$ elements. We claim that this implies the following.

Claim 8

$$|F_-| \leq \sum_{i=0}^{d} \sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S} N_k. \tag{56}$$

Proof (of Claim 8): The restriction of π to F_- is 1-1 by construction of F_-. The set $\pi(F_-)$ has GP-dimension no greater than d since any set of indices shattered by $\pi(F_-)$ is also shattered by F_-, and therefore by F. By the induction hypothesis,

$$|\pi(F_-)| \leq \sum_{i=0}^{d} \sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S} N_k, \tag{57}$$

so since π's restriction to F_- is 1-1, the claim is verified. □

Next, under the same induction hypothesis, we make the following claim.

Claim 9 *For all* $n \in \boldsymbol{N}, 1 \leq n \leq N_m$,

$$|F_n| \leq \sum_{i=0}^{d-1} \sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S} N_k. \tag{58}$$

Proof (of Claim 9): Choose $n \in \{1, ..., N_m\}$. We will show that the GP-dimension of F_n is at most $d-1$. The claim then follows by an argument similar to that of the previous claim. Let I be a set of indices GP-shattered by F_n and let $|I| = l$. Note that $m \notin I$, since $f_m = n$ for all $\bar{f} \in F_n$.

Now we show that $I \cup \{m\}$ is GP-shattered by F. Let $\bar{y}$ be the witness of F_n's GP-shattering of I. Consider $(y_1, ..., y_l, n)$. Choose $\bar{b} \in \{0,1\}^{l+1}$. Let $\bar{f} \in F_n$ satisfy $(b_1, ..., b_l)$ (with respect to I).

If $b_{l+1} = 1$, then $\bar{f}$ satisfies $\bar{b}$, and if $b_{l+1} = 0$, then

$$(f_1, ..., f_{m-1}, \alpha(f_1, ..., f_{m-1})) \tag{59}$$

satisfies $\bar{b}$. Since $\bar{b}$ was chosen arbitrarily, $I \cup \{m\}$ is GP-shattered by F. Since by assumption the GP-dimension of F is no greater than d and $m \notin I$, we have $|I| \leq d-1$. Since I was chosen arbitrarily, the GP-dimension of F_n is no greater than $d-1$, which is sufficient to prove this claim, as discussed above. □

From the previous two claims, we have that

$$|F| \leq \left[\sum_{i=0}^{d} \sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S} N_k\right] + N_m \sum_{i=0}^{d-1} \sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S} N_k \tag{60}$$

$$= \left[\sum_{i=0}^{d} \sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S} N_k\right] + \sum_{i=0}^{d-1} \sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S \cup \{m\}} N_k \tag{61}$$

$$= \left[1 + \sum_{i=1}^{d} \sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S} N_k\right] + \sum_{i=1}^{d} \sum_{S \in \Gamma_{(m-1)(i-1)}} \prod_{k \in S \cup \{m\}} N_k \tag{62}$$

$$= 1 + \sum_{i=1}^{d} \left\{ \left[\sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S} N_k \right] + \left[\sum_{S \in \Gamma_{(m-1)(i-1)}} \prod_{k \in S \cup \{m\}} N_k \right] \right\} \tag{63}$$

$$= 1 + \sum_{i=1}^{d} \left\{ \left[\sum_{S \in \Gamma_{mi}, m \notin S} \prod_{k \in S} N_k \right] + \left[\sum_{S \in \Gamma_{mi}, m \in S} \prod_{k \in S} N_k \right] \right\} \tag{64}$$

$$= \sum_{i=0}^{d} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} N_k. \tag{65}$$

This completes the induction. □

Theorem 2 easily follows from the previous lemmas together with the discussion relating $GP_{\max}$ to $G_{\max}$ and $P_{\max}$.

Next, we turn to Theorem 4. The lower bound was established in Lemma 6. We obtain the upper bound with the following lemma, the proof of which is similar to that of Lemma 7.

Lemma 10 *Let $d, m \in Z^+$, $N_1, ..., N_m \in \boldsymbol{N}$ be such that $d \leq m$. Let*

$$F \subseteq X = \prod_{i=1}^{m} \{0, ..., N_i\} \tag{66}$$

be such that N-dim$(F) \leq d$. Then

$$|F| \leq \sum_{i=0}^{d} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} \binom{N_k + 1}{2}. \tag{67}$$

Proof: As before, our proof is by double induction on m and d.

Using the same argument as the previous lemma, we can establish this lemma for the case $d = 0$.

Next, suppose that $d = m$. By partitioning the elements of the domain as discussed above, we can see that

$$|X| \leq \sum_{i=0}^{m} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} N_k \tag{68}$$

$$\leq \sum_{i=0}^{m} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} \binom{N_k + 1}{2} \tag{69}$$

so since $F \subseteq X$, certainly

$$|F| \leq \sum_{i=0}^{m} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} \binom{N_k + 1}{2}. \tag{70}$$

Now, choose $d, m \in Z^+$ such that $0 < d < m$. Define α and F_- as in the previous lemma and for each pair of distinct elements $u, v \in \boldsymbol{N}, 0 \leq u < v \leq N_m$, define

$$F_{uv} = \{\bar{f} \in F - F_- : f_m = v, \alpha(f_1, ..., f_{m-1}) = u\}. \tag{71}$$

Since each of the above sets are disjoint and their union is all of F, we have

$$|F| = |F_-| + \sum_{u=0}^{N_m - 1} \sum_{v=u+1}^{N_m} |F_{uv}|. \tag{72}$$

Using the same argument as in the previous lemma, under the inductive hypothesis that the lemma holds for all sets F of vectors of $m-1$ elements, we have

$$|F_-| \leq \sum_{i=0}^{d} \sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S} \binom{N_k+1}{2}. \tag{73}$$

Now, we wish to establish the following claim under the same inductive hypothesis.

Claim 11 *For all $u, v \in \boldsymbol{N}, 0 \leq u < v \leq N_m$, we have*

$$|F_{uv}| \leq \sum_{i=0}^{d-1} \sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S} \binom{N_k+1}{2}. \tag{74}$$

Proof (of Claim): Choose $u, v \in \boldsymbol{N}, 0 \leq u < v \leq N_m$. We will show that the N-dimension of F_{uv} is at most $d-1$. The claim then follows by an argument similar to that of Claim 8. Let I be a set of indices shattered by F_{uv} with $|I| = l$. Note that $m \notin I$, since $f_m = v$ for all $\bar{f} \in F_{uv}$.

Now we show that $I \cup \{m\}$ is shattered by F. Let $\bar{y}$ and $\bar{z}$ be the witnesses of F_{uv}'s N-shattering of I. Consider $(y_1, ..., y_l, u)$ and $(z_1, ..., z_l, v)$. We claim that these vectors witness F's N-shattering of $I \cup \{m\}$. Choose $\bar{b} \in \{0,1\}^{l+1}$. Let $\bar{f} \in F_{uv}$ satisfy $(b_1, ..., b_l)$ (with respect to I).

If $b_{l+1} = 1$, then $\bar{f}$ satisfies $\bar{b}$, and if $b_{l+1} = 0$, then

$$(f_1, ..., f_{m-1}, \alpha(f_1, ..., f_{m-1})) = (f_1, ..., f_{m-1}, u) \tag{75}$$

satisfies $\bar{b}$. Since $\bar{b}$ was chosen arbitrarily, $I \cup \{m\}$ is N-shattered by F. Since by assumption the N-dimension of F is no greater than d and $m \notin I$, we have $|I| \leq d-1$. Since I was chosen arbitrarily, the N-dimension of F_{uv} is no greater than $d-1$, which completes our proof of this claim, by the discussion above. □

From the previous two claims, we have that

$$\begin{aligned}
|F| &\leq \left[\sum_{i=0}^{d} \sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S} \binom{N_k+1}{2}\right] + \binom{N_m+1}{2} \sum_{i=0}^{d-1} \sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S} \binom{N_k+1}{2} && (76)\\
&= \left[\sum_{i=0}^{d} \sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S} \binom{N_k+1}{2}\right] + \sum_{i=0}^{d-1} \sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S \cup \{m\}} \binom{N_k+1}{2} && (77)\\
&= \left[1 + \sum_{i=1}^{d} \sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S} \binom{N_k+1}{2}\right] + \sum_{i=1}^{d} \sum_{S \in \Gamma_{(m-1)(i-1)}} \prod_{k \in S \cup \{m\}} \binom{N_k+1}{2} && (78)\\
&= 1 + \sum_{i=1}^{d} \left\{ \left[\sum_{S \in \Gamma_{(m-1)i}} \prod_{k \in S} \binom{N_k+1}{2}\right] + \left[\sum_{S \in \Gamma_{(m-1)(i-1)}} \prod_{k \in S \cup \{m\}} \binom{N_k+1}{2}\right] \right\} && (79)\\
&= 1 + \sum_{i=1}^{d} \left\{ \left[\sum_{S \in \Gamma_{mi}, m \notin S} \prod_{k \in S} \binom{N_k+1}{2}\right] + \left[\sum_{S \in \Gamma_{mi}, m \in S} \prod_{k \in S} \binom{N_k+1}{2}\right] \right\} && (80)\\
&= \sum_{i=0}^{d} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} \binom{N_k+1}{2} && (81)
\end{aligned}$$

which completes the induction. □

Theorem 4 can now easily be established.

3 An application

In this section, we give an application of Corollary 3, bounding the sample size necessary to obtain uniformly good empirical estimates for the expectations of all random variables of a given class F in terms of a

generalization of the definition of P-dimension given above to classes of real valued functions, in this case, random variables. We will measure the deviation of the estimates from the true expectations using a metric introduced in [12]. These results can be extended to bound the sample size necessary for learning according to the computational model of learning discussed in [12], an extension of that introduced in [21] which incorporates additional methods from previous work in Pattern Recognition.

We begin with some definitions. First, we will denote the set of positive real numbers by R^+. Now, let S be a set. Let $d : S \times S \to R^+$. We say that d is a metric on S if for all $x, y, z \in S$,

$$x = y \quad \Leftrightarrow \quad d(x, y) = 0 \tag{82}$$

$$d(x, y) \quad = \quad d(y, x) \tag{83}$$

$$d(x, z) \quad \leq \quad d(x, y) + d(y, z). \tag{84}$$

In this case, we say (S, d) is a metric space. Let $T \subseteq S$. We say T is bounded if $\sup\{d(x, y) : x, y \in T\}$ is finite. For any $\epsilon \in R^+$, a finite set N is an ϵ-cover for T if and only if for all $x \in T$, there exists $y \in N$ with $d(x, y) \leq \epsilon$. We say T is totally bounded if T has a finite ϵ-cover for each $\epsilon \in R^+$. In this case, we let $\mathcal{N}(\epsilon, T, d)$ denote the cardinality of the smallest ϵ-cover of T (w.r.t. S and d).

Now, we define the metric relative to which we prove uniform convergence results in this section. This metric was introduced and its utility as a measure of accuracy for an approximation of a function was discussed in [12]. For each $\nu \in R^+$, define $d_\nu : R^+ \times R^+ \to R^+$ by

$$d_\nu(r, s) = \frac{|r - s|}{\nu + r + s}. \tag{85}$$

It is straightforward but tedious to verify that for all $\nu \in R^+$, d_ν is a metric on R^+.

Let $(S, \mathcal{B}, D)$ be probability space with D a probability measure on the set S, and $\mathcal{B}$ some appropriate σ-algebra on S. Let F be a set of (measurable) random variables on S. For $m \geq 1$, denote by S^m the m-fold product space with the usual product probability measure. For any

$$\bar{\xi} = (\xi_1, ..., \xi_m) \in S^m \tag{86}$$

and $f \in F$, let

$$\hat{E}_{\bar{\xi}}(f) = \frac{1}{m} \sum_{i=1}^{m} f(\xi_i). \tag{87}$$

and

$$F_{|_{\bar{\xi}}} = \{(f(\xi_1), ..., f(\xi_m)) : f \in F\}. \tag{88}$$

We can view $F_{|_{\bar{\xi}}}$ as a subspace of the metric space (R^m, d_{L^1}), where d_{L^1} is the usual L^1 metric, i.e., for any $\bar{x} = (x_1, ..., x_m)$ and $\bar{y} = (y_1, ..., y_m)$ in R^m,

$$d_{L^1}(\bar{x}, \bar{y}) = \frac{1}{m} \sum_{i=1}^{m} |x_i - y_i|. \tag{89}$$

Also, we denote by $\mathcal{N}(\epsilon, F_{|_{\bar{\xi}}}, d_{L^1})$ the size of the smallest ϵ-cover of $F_{|_{\bar{\xi}}}$ in the d_{L^1} metric by elements of R^m.

Similary, we can view $F_{|_{\bar{\xi}}}$ as a subspace of (R^m, d_{L^∞}), where d_{L^∞} is defined as follows. For $\bar{x} = (x_1, ..., x_m)$ and $\bar{y} = (y_1, ..., y_m)$ in R^m,

$$d_{L^\infty}(\bar{x}, \bar{y}) = \max\{|x_i - y_i| : 1 \leq i \leq m\}. \tag{90}$$

Denote by $\mathcal{N}(\epsilon, F_{|_{\bar{\xi}}}, d_{L^\infty})$ the size of the smallest ϵ-cover of $F_{|_{\bar{\xi}}}$ in the d_{L^∞} metric by elements of R^m. Since clearly for all $\bar{x}, \bar{y} \in R^m$, $d_{L^1}(\bar{x}, \bar{y}) \leq d_{L^\infty}(\bar{x}, \bar{y})$, any ϵ-cover in the d_{L^∞} metric also serves as a ϵ-cover in the d_{L^1} metric, which implies

$$\mathcal{N}(\epsilon, F_{|_{\bar{\xi}}}, d_{L^1}) \leq \mathcal{N}(\epsilon, F_{|_{\bar{\xi}}}, d_{L^\infty}). \tag{91}$$

We are now ready for the following theorem. Similar results are given in [9] [15] [22]. In general, these theorems bound deviation of estimates $\hat{E}_{\bar{\xi}}(f)$ from true means $E(f)$ for functions f in F in terms of sizes of ϵ-covers for $F_{|_{\bar{\xi}}}$.

Theorem 12 ([12]) *Let F be a set[2] of random variables on S such that there exists $M \in R^+$ with $0 \leq f(\xi) \leq M$ for all $f \in F$ and $\xi \in S$. Assume $\nu > 0$, $0 < \alpha < 1$ and $m \geq 1$. Suppose that $\bar{\xi} \in S^m$ is generated by m independent random draws according to the fixed measure D on S. Let*

$$p(\alpha, \nu, m) = \boldsymbol{Pr}\left\{\exists f \in F : d_\nu(\hat{E}_{\bar{\xi}}(f), E(f)) > \alpha\right\}. \tag{92}$$

Then

$$p(\alpha, \nu, m) \leq 2E\left(\min(2\mathcal{N}(\alpha\nu/8, F_{|_{\bar{\xi}}}, d_{L^1})e^{-\alpha^2\nu m/8M}, 1)\right). \tag{93}$$

Let us generalize the definition of the ϕ-dimension given above for sets of integer vectors to sets of real valued functions. F be a set of real valued functions defined on some linearly ordered domain X. Let $I = \{x_1, ..., x_k\} \subseteq X$, with $x_1 < x_2 < \cdots < x_k$. For $f \in F$, let

$$f_{|_I} = (f(x_1), ..., f(x_k)). \tag{94}$$

Define

$$F_{|_I} = \{f_{|_I} : f \in F\}. \tag{95}$$

Choose $\phi : R \times R \to \{0, 1, \star\}$. Extend ϕ to $2^{R^k} \times R^k$ as in Section 1. We say that I is ϕ-shattered by F if there exists $\bar{y} \in R^k$ such that

$$\{0,1\}^k \subseteq \phi(F_{|_I}, \bar{y}). \tag{96}$$

We say that $\bar{y}$ witnesses F's ϕ-shattering of I and that $f \in F$ satisfies $\bar{b} \in \{0,1\}^k$ if and only if $\phi(f_{|_I}, \bar{y}) = \bar{b}$. The ϕ-dimension of F is the cardinality of the largest subset of X shattered by F.

As in Section 1, we define the P-dimension and P-shattering to be the ϕ-dimension and ϕ-shattering with ϕ defined by

$$\phi(i,j) = \begin{cases} 1 & \text{if } i \geq j \\ 0 & \text{if } i < j \end{cases}. \tag{97}$$

Note that this definition is equivalent to that of the previous sections when $X = [m]$ and we restrict the range of functions in F to initial intervals of the nonnegative integers. If we assume that $X = [m]$ and the range of all functions in F is $[0, M]$ for some positive real M, we obtain from the above definition a definition of the P-dimension of a subset of $[0, M]^m$ analogous to that of the previous section for integer vectors. This will prove useful.

Now we wish to show that if a subset of a product of closed intervals of R has small P-dimension, then it has a small ϵ-cover in the d_{L^∞} metric.

Lemma 13 *Let $M \in R^+, m \in Z^+$. Let $F \subseteq [0, M]^m$ be such that P-dim$(F) \leq d$. Let $\epsilon \in R^+$. Then*

$$\mathcal{N}(\epsilon, F, d_{L^\infty}) \leq \sum_{i=0}^{d} \binom{m}{i} \left\lfloor \frac{M}{2\epsilon} \right\rfloor^i \tag{98}$$

Proof: If $S \subseteq R^m$, $\bar{y} \in R^m$ and $c \in R$, denote by $cS + \bar{y}$ the set $\{c\bar{s} + \bar{y} : \bar{s} \in S\}$.

Define $\beta : [0, M]^m \to \{0, ..., \lfloor \frac{M}{2\epsilon} \rfloor\}^m$ by $\beta(\bar{f}) = \bar{g}$, where $g_i = \left\lfloor \frac{f_i}{2\epsilon} \right\rfloor$ for all $i, 1 \leq i \leq m$. Let $G = \beta(F)$. Let $H = 2\epsilon G + (\epsilon, \epsilon, ..., \epsilon)$.

First, we claim that H is an ϵ-cover for F with respect to the d_{L^∞} metric. Choose $\bar{f} \in F$. Let $\bar{h} = 2\epsilon\beta(\bar{f}) + (\epsilon, \epsilon, ..., \epsilon)$. Choose $i, 1 \leq i \leq m$. Then we have

$$|f_i - h_i| = \left| f_i - \left(2\epsilon \left\lfloor \frac{f_i}{2\epsilon} \right\rfloor + \epsilon \right) \right| \tag{99}$$

$$= 2\epsilon \left| \frac{f_i}{2\epsilon} - \left\lfloor \frac{f_i}{2\epsilon} \right\rfloor - \frac{1}{2} \right| \tag{100}$$

$$\leq \epsilon. \tag{101}$$

[2] Further measurability assumptions are required. See [12].

Since i was chosen arbitrarily,

$$d_{L^\infty}(f,h) = \max\{|f_i - h_i| : 1 \le i \le m\} \le \epsilon. \tag{102}$$

Since $\bar{f} \in F$ was chosen arbitrarily, H is an ϵ-cover for F.

Next, we wish to show that P-dim$(G) \le d$. Let $I = \{i_1, ..., i_k\}$ be a set shattered by G and let

$$\bar{y} \in \{0, ..., \left\lfloor \frac{M}{2\epsilon} \right\rfloor\}^k \tag{103}$$

witness this shattering. We claim that $2\epsilon\bar{y}$ witnesses F's shattering of I. Choose $\bar{b} \in \{0,1\}^k$. Let $\bar{g} \in G$ satisfy $\bar{b}$. Choose $\bar{f} \in F$, such that $\beta(\bar{f}) = \bar{g}$.

If $b_j = 1$, we have $g_{i_j} \ge y_j$ which is equivalent to

$$\left\lfloor \frac{f_{i_j}}{2\epsilon} \right\rfloor \ge y_j \tag{104}$$

which implies

$$\frac{f_{i_j}}{2\epsilon} \ge y_j \tag{105}$$

since $x \ge \lfloor x \rfloor$ for all $x \in R$. Finally, the previous inequality implies

$$f_{i_j} \ge 2\epsilon y_j. \tag{106}$$

So if $b_j = 1$, $f_{i_j} \ge 2\epsilon y_j$.

Suppose $b_j = 0$ and $f_{i_j} \ge 2\epsilon y_j$. This implies $f_{i_j}/2\epsilon \ge y_j$, which in turn implies

$$g_{i_j} = \left\lfloor \frac{f_{i_j}}{2\epsilon} \right\rfloor \ge y_j, \tag{107}$$

since $y_j \in Z$. But this is a contradiction, since $g_{i_j} < y_j$, which holds because $b_j = 0$ and g satisfies $\bar{b}$. So if $b_j = 0$, we have $f_{i_j} < 2\epsilon y_j$.

In the preceding two paragraphs we have established that for all $j, 1 \le j \le k$, we have $f_{i_j} \ge 2\epsilon y_j$ if and only if $b_j = 1$, and thereby that $\bar{f}$ satisfies $\bar{b}$. Since $\bar{b}$ was chosen arbitrarily, I is shattered by F. Since I was chosen arbitrarily, P-dim$(G) \le$ P-dim$(F) = d$.

Now, by Corollary 3,

$$|G| \le \sum_{i=0}^{d} \binom{m}{i} \left\lfloor \frac{M}{2\epsilon} \right\rfloor^i. \tag{108}$$

Since H is an ϵ-cover of F and $|G| = |H|$, we have

$$\mathcal{N}(\epsilon, F, d_{L^\infty}) \le \sum_{i=0}^{d} \binom{m}{i} \left\lfloor \frac{M}{2\epsilon} \right\rfloor^i, \tag{109}$$

which completes the proof. □

Corollary 14 *Let $M \in R^+, m \in Z^+$. Let $F \subseteq [0,M]^m$ be such that P-dim$(F) \le d$. Let $\epsilon \in R^+$. Then*

$$\mathcal{N}(\epsilon, F, d_{L^1}) \le \sum_{i=0}^{d} \binom{m}{i} \left(\left\lfloor \frac{M}{2\epsilon} \right\rfloor \right)^i \tag{110}$$

Proof: As discussed above

$$\mathcal{N}(\epsilon, F, d_{L^1}) \le \mathcal{N}(\epsilon, F, d_{L^\infty}). \tag{111}$$

The corollary then follows from the previous lemma. □

The technique by which we obtain bounds on the sample size necessary for the uniform convergence of estimates to true means for a sequence of random variables has elements which are similar to that used to in [18] improve the bounds of [7]. The following lemma is useful in this derivation.

Lemma 15 ([18]) *Let $x, y \in \boldsymbol{R}+$. Then*

$$\ln x \leq xy - \ln ey. \tag{112}$$

Proof: Fix $y \in \boldsymbol{R}+$. Consider $f : \boldsymbol{R}+ \to R$ defined by

$$f(x) = xy - \ln exy. \tag{113}$$

Then

$$f'(x) = y - 1/x. \tag{114}$$

Clearly, $f'(x)$ is positive when $x > 1/y$ and negative when $x < 1/y$ and f is continuous and differentiable over its domain, so f assumes its minimum at $1/y$ and

$$f(1/y) = y(1/y) - \ln ey(1/y) = 0. \tag{115}$$

So $f(x) \geq 0$ for all $x \in \boldsymbol{R}+$, which yields the desired result. □

Finally, we are ready to bound the sample size necessary to ensure that with high probability an empirical estimate of the expected value of a random variable chosen from a set of a small P-dimension is accurate.

Theorem 16 *Let F be a set[3] of random variables on S such that there exists $M \in R^+$ with $0 \leq f(\xi) \leq M$ for all $f \in F$ and $\xi \in S$. Assume $\nu > 0$, $0 < \alpha < 1$ and $m \geq 1$. Suppose that $\bar{\xi}$ is generated by m independent random draws according to the fixed measure D on S. Suppose also that P-dim$(F) \leq d$. Then*

$$\boldsymbol{Pr}\left\{\exists f \in F : d_\nu(\hat{E}_{\bar{\xi}}(f), E(f)) > \alpha\right\} \leq 4\left(\frac{4M}{\alpha\nu}\right)^d \left(\frac{em}{d}\right)^d e^{-\alpha^2\nu m/8M}. \tag{116}$$

Moreover, for

$$m \geq \frac{9M}{\alpha^2\nu}\left(2d\ln\frac{17M}{(\alpha\sqrt{\alpha})\nu} + \ln\frac{4}{\delta}\right), \tag{117}$$

this probability is less than δ.

Proof: First, from Corollary 14, we have that

$$\mathcal{N}(\alpha\nu/8, F_{|_{\bar{\xi}}}, d_{L^1}) \leq \sum_{i=0}^{d}\binom{m}{i}\left\lfloor\frac{4M}{\alpha\nu}\right\rfloor^i. \tag{118}$$

Using the well known combinatorial identity that

$$\sum_{i=0}^{d}\binom{m}{i} \leq (em/d)^d \tag{119}$$

and substituting

$$\left(\frac{4M}{\alpha\nu}\right)^d \tag{120}$$

for each

$$\left\lfloor\frac{4M}{\alpha\nu}\right\rfloor^i, \tag{121}$$

we get

$$\mathcal{N}(\alpha\nu/8, F_{|_{\bar{\xi}}}, d_{L^1}) \leq \left(\frac{4M}{\alpha\nu}\right)^d \left(\frac{em}{d}\right)^d. \tag{122}$$

Applying Theorem 12 yields the first result.

[3] The same measurability assumptions as Theorem 12 are required.

Now, we wish to determine a lower bound on m which guarantees that

$$4\left(\frac{4M}{\alpha\nu}\right)^d\left(\frac{em}{d}\right)^d e^{-\alpha^2\nu m/8M} \leq \delta. \tag{123}$$

Set $k = \frac{4M}{\alpha\nu}$. Then the above expression simplifies to

$$4k^d\left(\frac{em}{d}\right)^d e^{-\frac{\alpha m}{2k}} \leq \delta. \tag{124}$$

Taking logs and rearranging terms yields the following equivalent expression:

$$\frac{\alpha m}{2k} \geq d\left(\ln m + \ln\frac{ke}{d}\right) + \ln 4/\delta. \tag{125}$$

Since by the preceding lemma for any $\lambda \in R, 0 < \lambda < 1$,

$$\ln m \leq \left(\frac{\lambda\alpha}{kd}\right)m + \left(\ln\frac{kd}{\lambda\alpha e}\right), \tag{126}$$

the following is sufficient to guarantee Inequality (125):

$$\begin{aligned}\frac{\alpha m}{2k} &\geq d\left(\frac{\lambda\alpha}{kd}m + \ln\frac{kd}{\lambda\alpha e} + \ln\frac{ke}{d}\right) + \ln 4/\delta &(127)\\ &= \frac{\lambda\alpha}{k}m + d\ln\frac{k^2}{\lambda\alpha} + \ln 4/\delta. &(128)\end{aligned}$$

Solving for m yields

$$m \geq \frac{2k}{\alpha(1-2\lambda)}\left(2d\ln\frac{k}{\sqrt{\lambda\alpha}} + \ln\frac{4}{\delta}\right) \tag{129}$$

and resubstituting $k = \frac{4M}{\alpha\nu}$ gives

$$m \geq \frac{8M}{\alpha^2\nu(1-2\lambda)}\left(2d\ln\frac{4M}{\alpha\nu\sqrt{\lambda\alpha}} + \ln\frac{4}{\delta}\right). \tag{130}$$

We choose $\lambda = 1/18$ for readability, yielding

$$m \geq \frac{9M}{\alpha^2\nu}\left(2d\ln\frac{17M}{(\alpha\sqrt{\alpha})\nu} + \ln\frac{4}{\delta}\right) \tag{131}$$

which is the desired bound. □

For comparison, we give the following theorem from [12], which was obtained using a completely different technique, due to Pollard [16, Theorem 4.7].

Theorem 17 *Let F be a set[4] of random variables on S such that there exists $M \in R^+$ with $0 \leq f(\xi) \leq M$ for all $f \in F$ and $\xi \in S$. Assume $0 < \nu \leq 4M/d$, $0 < \alpha < 1$ and $m \geq 1$. Suppose that $\bar{\xi}$ is generated by m independent random draws according to the fixed measure D on S. Suppose also that P-dim$(F) \leq d$. Then*

$$\boldsymbol{Pr}\left\{\exists f \in F : d_\nu(\hat{E}_{\bar{\xi}}(f), E(f)) > \alpha\right\} \leq 8\left(\frac{16eM}{\alpha\nu}\ln\frac{16eM}{\alpha\nu}\right)^d e^{-\alpha^2\nu m/8M}. \tag{132}$$

Moreover, for

$$m \geq \frac{8M}{\alpha^2\nu}\left(2d\ln\frac{8eM}{\alpha\nu} + \ln\frac{8}{\delta}\right), \tag{133}$$

this probability is less than δ.

[4] Again, the same measurability assumptions as Theorem 12 are required.

4 Conclusion

We have given tight bounds on the cardinality of a subset of $\prod_i \{0, ..., N_i\}$ of a certain dimension for two generalizations of the VC-dimension: namely the pseudo dimension discussed by Pollard [15] and the graph dimension introduced by Natarajan [14]. We also have used a similar technique obtain tighter bounds for another generalization of the VC-dimension introduced by Natarajan, which we have called the Natarajan dimension. The problem of obtaining tight bounds for the Natarajan dimension remains open.

In addition, we have applied this result to bound the rate of convergence of empirical estimates of the expectations of a sequence of random variables to their true expectations, obtaining bounds similar to those already derived in [15] [12]. These results can be extended to bound the sample size required for learning under the computational model of learnability discussed in [12].

The primary motivation for this research, however, was to attempt to find some simple property of a class of functions that would characterize the uniform rate of convergence of estimates to true means. While the finiteness of any of the dimensions discussed in this paper is sufficient for rapid convergence, none of them are necessary. We hope that the insight gained by studying these generalizations of the VC-dimension will aid us in this pursuit.

Towards this end, we are currently investigating the following conjecture. Let $m \in Z^+$, $N \in \boldsymbol{N}$. Define γ as in the definition of N-dimension given previously. Let $G = (V, E)$ be a graph with $V = \{0, ..., N\}$. Form $G^m = (V^m, E^m)$ as follows. Let $V^m = \{0, ..., N\}^m$, as the notation suggests. Let

$$E^m = \{\langle \bar{f}, \bar{g} \rangle : \bar{f}, \bar{g} \in V^m, \exists i, 1 \leq i \leq m, \langle f_i, g_i \rangle \in E\}. \tag{134}$$

Let $F \subseteq V^m$ be a clique in G^m. Let $I \subseteq [m]$. We say I is GN-shattered by F if there exist $\bar{f}, \bar{g} \in F$ such that for all $i \in I$, $\langle f_i, g_i \rangle \in G$, and

$$\{0, 1\}^{|I|} \subseteq \gamma(F_{|_I}, \bar{f}, \bar{g}). \tag{135}$$

We define the GN-dimension of F to be the cardinality of the largest subset of $[m]$ shattered by F. Our conjecture is that if the GN-dimension of F is no greater than d, then

$$|F| \leq \sum_{i=0}^{d} \binom{m}{i} \binom{N+1}{2}^i. \tag{136}$$

Note that if G is the complete graph, any $F \subseteq V^m$ induces a clique, and the GN-dimension of F reduces to its N-dimension, so the above bound follows from Corollary 5.

If such a result could be obtained, it would lead to a new characterization of conditions that ensure rapid uniform convergence similar to the conditions outlined by Vapnik [23].

We are also working on the problem of characterizing those functions ϕ (or γ) such that a bound on the ϕ-dimension of a set gives the bounds of Theorem 2, or more generally, characterizing the functions ϕ such that there exist bounds polynomial in m and N on the cardinality of subsets of $\prod_{i=1}^{m} \{0, ..., N\}$ of a given ϕ-dimension.

References

[1] R.E.L. Aldred and R.P. Anstee. On the density of sets of divisors. Unpublished manuscript, 1991.

[2] N. Alon. On the density of sets of vectors. *Discrete Mathematics*, 24:177–184, 1983.

[3] R.P. Anstee. General forbidden configuration theorems. *Journal of Combinatorial Theory (A)*, 40:108–124, 1985.

[4] R.P. Anstee. A forbidden configuration theorem of Alon. *Journal of Combinatorial Theory (A)*, 47:16–27, 1988.

[5] R.P. Anstee. On a conjecture concerning forbidden submatrices. Unpublished manuscript, 1991.

[6] R.P. Anstee and Z. Furedi. Forbidden submatrices. *Discrete Math*, 62:225–243, 1986.

[7] A. Blumer, A. Ehrenfeucht, D. Haussler, and M.K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *JACM*, 36(4):929–965, 1989.

[8] J.A. Bondy. Induced subsets. *Journal of Combinatorial Theory (B)*, 12:201–202, 1972.

[9] R.M. Dudley. A course on empirical processes. *Lecture notes in mathematics*, 1097:2–142, 1984.

[10] P. Frankl. On the trace of finite sets. *Journal of Combinatorial Theory (A)*, 34:41–45, 1983.

[11] P. Frankl, Z. Furedi, and J. Pach. Bounding one-way differences. *European Journal of Combinatorics*, 3:341–347, 1987.

[12] D. Haussler. Generalizing the PAC model: sample size bounds from metric dimension-based uniform convergence results. *Proceedings of the 30th Annual Symposium on the Foundations of Computer Science*, 1989.

[13] M.G. Karpovsky and V.D. Milman. Coordinate density of sets of vectors. *Discrete Mathematics*, 24:177–184, 1978.

[14] B.K. Natarajan. On learning sets and functions. *Machine Learning*, 4:67–97, 1989.

[15] D. Pollard. *Convergence of Stochastic Processes*. Springer Verlag, 1984.

[16] D. Pollard. *Empirical Processes : Theory and Applications*, volume 2 of *NSF-CBMS Regional Conference Series in Probability and Statistics*. Institute of Math. Stat. and Am. Stat. Assoc., 1990.

[17] N. Sauer. On the density of families of sets. *J. Combinatorial Theory (A)*, 13:145–147, 1972.

[18] J. Shawe-Taylor, M. Anthony, and R.L. Biggs. Bounding sample size with the vapnik-chervonenkis dimension. Technical Report CSD-TR-618, University of London, Royal Halloway and New Bedford College, 1989.

[19] J.M. Steele. Existence of submatrices with all possible columns. *Journal of Combinatorial Theory, Series A*, 24:84–88, 1978.

[20] P. Tomasta. Dart calculus of induced subsets. *Discrete Mathematics*, 34:195–198, 1981.

[21] L.G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[22] V.N. Vapnik. *Estimation of Dependencies based on Empirical Data*. Springer Verlag, 1982.

[23] V.N. Vapnik. Inductive principles of the search for empirical dependences (methods based on weak convergence of probability measures). *The 1989 Workshop on Computational Learning Theory*, 1989.

[24] V.N. Vapnik and A.Y. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16(2):264–280, 1971.