## References


[Alo83] N. Alon. On the density of sets of vectors. *Discrete Mathematics*, 24, 177-184, 1983.

[BEHW89] A. Blumer, A. Ehrenfeucht, D. Haussler, and M.K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *JACM*, 36(4), 1989.

[Bon72] J.A. Bondy. Induced subsets. *Journal of Combinatorial Theory*, 12(B), 201-202, 1972.

[Dud84] R.M. Dudley. A course on empirical processes. *Lecture Notes in Mathematics*, 1097:2-142, 1984

[Fra83] P. Frankl. On the trace of finite sets. *Journal of Combinatorial Theory*, 34(A), 41-45, 1983.

[Hau89] 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*, IEEE, 1989.

[KM78] M.G. Karpovsky and V.D. Milman. Coordinate density of sets of vectors. *Discrete Mathematics*, 24, 177-184, 1978.

[Nat89] B.K. Natarajan. *Some Results on Learning.* Technical Report CMU-RI-TR-89-6, Carnegie Mellon, 1989.

[Pol84] D. Pollard. *Convergence of Stochastic Processes.* Springer-Verlag, 1984.

[Pol89] D. Pollard. *Empirical Processes : Theory and Applications.* Manuscript, December, 1989.

[Sau72] N. Sauer. On the density of families of sets. *J. Combinatorial Theory (A)*, 13:145-147, 1972.

[SAB89] J. Shawe-Taylor, M. Anthony and N.L. Biggs. Bounding sample size with the Vapnik-Chervonenkis dimension. Technical Report CSD-TR-618, University of London, Royal Holloway and Bedford New College, 1989.

[Ste78] J.M. Steele. Existence of submatrices with all possible columns. *Journal of Combinatorial Theory, Series A*, 24, 84-88, 1978.

[Tom81] P. Tomasta. "Dart calculus" of induced subsets. *Discrete Mathematics*, 34, 195-198, 1981.

[Val84] L. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134-1142, November, 1984

[Vap82] V.N. Vapnik. *Estimation of Dependences Based on Empirical Data.* Springer-Verlag, New York, 1982.

[Vap89] V.N. Vapnik. Inductive principles of the search for empirical dependences (methods based on weak convergence of probability measures). In *Proceedings of the 2nd Workshop on Computational Learning Theory*, Morgan Kaufmann, San Mateo, CA 1989.

[VC71] V.N. Vapnik and A.Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its applications*, XVI(2):264-280, 1971.

# 4. Conclusion

We have defined a dimension for a set of integer vectors, called the GC-dimension, and given tight bounds on the cardinality of a set of vectors taken from a particular domain which has a given GC-dimension. We have used this result to obtain similar tight bounds for generalizations of the VC-dimension which have been proposed by others, namely the pseudo dimension discussed by Pollard [Pol84] and the graph dimension introduced by Natarajan [Nat89]. 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 [Pol84] [Hau89]. These results can be extended to bound the sample size required for learning under the computational model of learnability discussed in [Hau89].

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 $m \in \mathbf{Z}^+, N \in \mathbf{N}$. Define $\gamma$ 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

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

Let $F \subseteq V^m$ be a clique in $G^m$. Let $I \subseteq \mathbf{N}_m$. 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

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

We define the GN-dimension of $F$ to be the cardinality of the largest subset of $\mathbf{N}_m$ shattered by $F$. Our conjecture is that if the GN-dimension of $F$ is no greater than $d$, then

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

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 1.5.

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 [Vap89].

We are also working on the problem of characterizing those functions $\phi$ (or $\gamma$) such that a bound on the $\phi$-dimension of a set gives the bounds of Theorem 1.2, or more generally, characterizing the functions $\phi$ 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 $\phi$-dimension.

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

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

the following is sufficient to guarantee Inequality 3.4:

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

Solving for $m$ yields

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

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

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

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

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

which is the desired bound. □

For comparison, we give the following theorem from [Hau89], which was obtained using a completely different technique, due to Pollard [Pol89, Theorem 4.7].

**Theorem 3.6:** *Let $F$ be a set[3] of random variables on $S$ such that there exists $M \in \mathbf{R}^+$ 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*

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

*Moreover, for*

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

*this probability is less than $\delta$.*

[3] Again, the same measurability assumptions as Theorem 3.1 are required.

**Theorem 3.5:** *Let $F$ be a set*[2] *of random variables on $S$ such that there exists $M \in \mathbf{R}^+$ 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*

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

*Moreover, for*

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

*this probability is less than $\delta$.*

Proof: First, from Corollary 3.3, we have that

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

Using the well known combinatorial identity that

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

and substituting

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

for each

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

we get

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

Applying Theorem 3.1 yields the first result.

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

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

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

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

Taking logs and rearranging terms yields the following equivalent expression:

$$\frac{\alpha m}{2k} \geq d\left(\ln m + \ln\frac{ke}{d}\right) + \ln 4/\delta. \tag{3.4}$$

[2]The same measurability assumptions as Theorem 3.1 are required.

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

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

since $y_j \in \mathbf{Z}$. 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 $j, 1 \leq j \leq k$, we have $f_{i_j} \geq 2\epsilon y_j$ 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$(G) \leq$ P-dim$(F) = d$.

Now, by Corollary 1.3,

$$|G| \leq \sum_{i=0}^{d} \binom{m}{i} \left( \left\lfloor \frac{M}{2\epsilon} \right\rfloor \right)^i.$$

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

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

which completes the proof. □

**Corollary 3.3:** *Let $M \in \mathbf{R}^+, m \in \mathbf{Z}^+$. Let $F \subseteq [0, M]^m$ be such that P-dim$(F) \leq d$. Let $\epsilon \in \mathbf{R}^+$. Then*

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

Proof: As discussed above

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

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 is similar to that used to in [SAB89] improve the bounds of [BEHW89]. The following lemma is useful in this derivation.

**Lemma 3.4 ([SAB89]):** *Let $x, y \in \mathbf{R}^+$. Then*

$$\ln x \leq xy - \ln ey.$$

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

$$f(x) = xy - \ln exy.$$

Then

$$f'(x) = y - 1/x.$$

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

$$f(1/y) = y(1/y) - \ln ey(1/y) = 0.$$

So $f(x) \geq 0$ for all $x \in \mathbf{R}^+$, 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.

Note that this definition is equivalent to that of the previous sections when $X = \mathbf{N}_m$ and we restrict the range of functions in $F$ to initial intervals of the nonnegative integers. If we assume that $X = \mathbf{N}_m$ 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 $\mathbf{R}$ has small P-dimension, then it can has a small $\epsilon$-cover in the $d_{L^\infty}$ metric.

**Lemma 3.2:** *Let $M \in \mathbf{R}^+, m \in \mathbf{Z}^+$. Let $F \subseteq [0, M]^m$ be such that P-dim$(F) \leq d$. Let $\epsilon \in \mathbf{R}^+$. Then*

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

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

Define $\beta : [0, M]^m \to \{0, ..., \left\lfloor \frac{M}{2\epsilon} \right\rfloor\}^m$ 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 $\epsilon$-cover for $F$ with respect to the $d_{L^\infty}$ 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

$$\begin{aligned} |f_i - h_i| &= \left| f_i - \left( 2\epsilon \left\lfloor \frac{f_i}{2\epsilon} \right\rfloor + \epsilon \right) \right| \\ &= 2\epsilon \left| \frac{f_i}{2\epsilon} - \left\lfloor \frac{f_i}{2\epsilon} \right\rfloor - \frac{1}{2} \right| \\ &\leq \epsilon. \end{aligned}$$

Since $i$ was chosen arbitrarily,

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

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

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

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

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 $g_{i_j} \geq y_j$ which is equivalent to

$$\left\lfloor \frac{f_{i_j}}{2\epsilon} \right\rfloor \geq y_j$$

which implies

$$\frac{f_{i_j}}{2\epsilon} \geq y_j$$

since $x \geq \lfloor x \rfloor$ for all $x \in \mathbf{R}$. Finally, the previous inequality implies

$$f_{i_j} \geq 2\epsilon y_j.$$

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

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

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

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

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

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

We are now ready for the following theorem. Similar results are given in [Dud84] [Pol84] [Vap82]. 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 $\epsilon$-covers for $F_{|\bar{\xi}}$.

**Theorem 3.1 ([Hau89]):** *Let $F$ be a set[1] of random variables on $S$ such that there exists $M \in \mathbf{R}^+$ 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*

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

*Then*

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

Let us generalize the definition of the $\phi$-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

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

Define

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

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

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

We say that $\bar{y}$ witnesses $F$'s $\phi$-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 $\phi$-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 $\phi$-dimension and $\phi$-shattering with $\phi$ defined by

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

[1] Further measurability assumptions are required. See [Hau89].

## 3. An application

In this section, we give an application of Corollary 1.3, 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 [Hau89]. These results can be extended to bound the sample size necessary for learning according to the computational model of learning discussed in [Hau89], an extension of that introduced in [Val84] 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 $\mathbf{R}^+$. Now, let $S$ be a set. Let $d : S \times S \to \mathbf{R}^+$. We say that $d$ is a metric on $S$ if for all $x, y, z \in S$,

$$x = y \quad \Leftrightarrow \quad d(x,y) = 0 \tag{3.1}$$

$$d(x,y) \quad = \quad d(y,x) \tag{3.2}$$

$$d(x,z) \quad \leq \quad d(x,y) + d(y,z). \tag{3.3}$$

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 $\epsilon \in \mathbf{R}^+$, a finite set $N$ is an $\epsilon$-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 $\epsilon$-cover for each $\epsilon \in \mathbf{R}^+$. In this case, we let $\mathcal{N}(\epsilon, T, d)$ denote the cardinality of the smallest $\epsilon$-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 [Hau89]. For each $\nu \in \mathbf{R}^+$, define $d_\nu : \mathbf{R}^+ \times \mathbf{R}^+ \to \mathbf{R}^+$ by

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

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

Let $(S, \mathcal{B}, D)$ be probability space with $D$ a probability measure on the set $S$, and $\mathcal{B}$ some appropriate $\sigma$-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

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

and $f \in F$, let

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

and

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

We can view $F_{|_{\bar{\xi}}}$ as a subspace of the metric space $(\mathbf{R}^m, d_{L^1})$, 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 $\mathbf{R}^m$,

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

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

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

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

**Claim 2.6:** *For all $u, v \in \mathbf{N}, 0 \leq u < v \leq N_m$, we have*

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

Proof (of Claim): Choose $u, v \in \mathbf{N}, 0 \leq u < v \leq N_m$. 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 2.3. 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

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

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

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

which completes the induction. □

Theorem 1.4 can now easily be established.

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

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

This completes the induction. □

Theorem 1.2 easily follows from the previous lemmas together with the discussion relating $GC_{\max}$ to $G_{\max}$ and $P_{\max}$.

Next, we turn to Theorem 1.4. The lower bound was established in Lemma 2.1. We obtain the upper bound with the following lemma, the proof of which is similar to that of Lemma 2.2.

**Lemma 2.5:** *Let $d, m \in \mathbf{Z}^+, N_1, ..., N_m \in \mathbf{N}$ be such that $d \leq m$. Let*

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

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

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

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

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

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

so since $F \subseteq X$, certainly

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

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

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

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

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

**Claim 2.3:**

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

Proof (of Claim 2.3): The restriction of $\pi$ to $F_-$ is 1-1 by construction of $F_-$. The set $\pi(F_-)$ has GC-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,

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

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

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

**Claim 2.4:** *For all* $n \in \mathbf{N}, 1 \leq n \leq N_m$,

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

Proof (of Claim 2.4): Choose $n \in \{1, ..., N_m\}$. We will show that the GC-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 GC-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 GC-shattered by $F$. Let $\bar{y}$ be the witness of $F_n$'s GC-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

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

satisfies $\bar{b}$. Since $\bar{b}$ was chosen arbitrarily, $I \cup \{m\}$ is GC-shattered by $F$. Since by assumption the GC-dimension of $F$ is no greater than $d$ and $m \notin I$, we have $|I| \leq d-1$. Since $I$ was chosen arbitrarily, the GC-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

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

**Lemma 2.2:** *Let* $d, m \in \mathbf{Z}^+, N_1, ..., N_m \in \mathbf{N}$ *be such that* $d \leq m$. *Let*

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

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

$$|F| \leq \sum_{i=0}^{d} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} N_k. \tag{2.1}$$

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

First we consider the case in which $d = 0$. Here, the bound 2.1 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 GC-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

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

so since $F \subseteq X$, certainly

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

establishing the result in this case.

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

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

Define

$$\alpha : \pi(F) \to \{0, ..., N_m\}$$

by

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

Define

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

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

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

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

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

Let us make the inductive assumption that the bound 2.1 holds for all sets $F$ of vectors of $m - 1$ elements. We claim that this implies the following.

## 2. Proofs of the results

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

**Lemma 2.1:** *Let $d, m \in \mathbf{Z}^+, N_1, ..., N_m \in \mathbf{N}$ be such that $d \leq m$. Then there exists*

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

*such that $F$ has G-, P-, GC-, and N-dimension $d$ and*

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

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

To prove that the G-, P-, GC- 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

$$\begin{aligned} \text{N-dim}(F) &\leq \text{GC-dim}(F) \\ \text{GC-dim}(F) &\leq \text{P-dim}(F) \\ \text{GC-dim}(F) &\leq \text{G-dim}(F). \end{aligned}$$

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

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

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-, GC- 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 $\mathbf{N}_m$ 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 GC-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 GC-dimension, and thereby that of sets of a given G- or P-dimension. Our argument is a generalization of that given by Sauer in [Sau72], and is similar to Natarajan's generalization of this argument in [Nat89].

which in turn implies that

$$\begin{aligned} P_{\max}(d,m,N_1,...,N_m) &\leq GC_{\max}(d,m,N_1,...,N_m) \\ G_{\max}(d,m,N_1,...,N_m) &\leq GC_{\max}(d,m,N_1,...,N_m) \\ GC_{\max}(d,m,N_1,...,N_m) &\leq N_{\max}(d,m,N_1,...,N_m) \end{aligned}$$

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

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

$$\Gamma_{mi} = \{S \subseteq \mathbf{N}_m : |S| = i\}.$$

**Theorem 1.2:** *For all $d, m \in \mathbf{Z}^+, N_1, ..., N_m \in \mathbf{N}$ such that $d \leq m$,*

$$\begin{aligned} G_{\max}(d,m,N_1,...,N_m) &= \\ P_{\max}(d,m,N_1,...,N_m) &= \\ GC_{\max}(d,m,N_1,...,N_m) &= \sum_{i=0}^{d} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} N_k. \end{aligned}$$

When there is an $N \in \mathbf{N}$ 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 [Hau89].

**Corollary 1.3:** *Let $d, m \in \mathbf{Z}^+, N \in \mathbf{N}$ be such that $d \leq m$. Let*

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

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

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

Proof: Follows from Theorem 1.2 by substituting $N$ for each $N_k$ and collecting terms. □

Using similar techniques, we can establish the following.

**Theorem 1.4:** *For all $d, m \in \mathbf{Z}^+, N_1, ..., N_m \in \mathbf{N}$ such that $d \leq m$,*

$$\begin{aligned} \sum_{i=0}^{d} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} N_k &\leq N_{\max}(d,m,N_1,...,N_m) \\ &\leq \sum_{i=0}^{d} \sum_{S \in \Gamma_{mi}} \prod_{k \in S} \binom{N_k + 1}{2}. \end{aligned}$$

This gives a result similar to that obtained by Natarajan [Nat89] in the special case above.

**Corollary 1.5:** *Let $d, m \in \mathbf{Z}^+, N \in \mathbf{N}$ be such that $d \leq m$. Let*

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

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

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

Note that both Corollary 1.3 and Corollary 1.5 give Sauer's result (Theorem 1.1) in the case $N = 1$.

with the corresponding definition for GC-shattering. The GC-dimension of $F$ is denoted by GC-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 $\mathbf{N}_m$ in the definition of VC-dimension as elements of $\{0,1\}^m$. Then $F \subseteq \{0,1\}^m$ shatters $I \subseteq \mathbf{N}_m$ exactly when $\{0,1\}^{|I|} \subseteq F_{|_I}$. It can now easily be verified that, for each function $\phi$ given above, $F$ is shattered exactly when $F$ is $\phi$-shattered.

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

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

$$\bigcup_{k \in \mathbf{N}} 2^{\mathbf{N}^k} \times \mathbf{N}^k \times \mathbf{N}^k$$

as above. We say that $I$ is $\gamma$-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

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

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

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

We say $F$ N-shatters $I$ if $F$ $\gamma$-shatters $I$ with $\gamma$ given by

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

The Natarajan-dimension (or N-dimension) of $F$ is defined to be its $\gamma$-dimension in this case. This definition appears in [Nat89]. The N-dimension also reduces to the VC-dimension when $N_i = 1$ for all $1 \leq i \leq m$.

Define

$$\begin{aligned}
P_{\max}(d, m, N_1, ..., N_m) &= \max\{|F| : F \subseteq \prod_{i=1}^{m} \{0, ..., N_i\}, \text{P-dim}(F) \leq d\} \\
G_{\max}(d, m, N_1, ..., N_m) &= \max\{|F| : F \subseteq \prod_{i=1}^{m} \{0, ..., N_i\}, \text{G-dim}(F) \leq d\} \\
GC_{\max}(d, m, N_1, ..., N_m) &= \max\{|F| : F \subseteq \prod_{i=1}^{m} \{0, ..., N_i\}, \text{GC-dim}(F) \leq d\} \\
N_{\max}(d, m, N_1, ..., N_m) &= \max\{|F| : F \subseteq \prod_{i=1}^{m} \{0, ..., N_i\}, \text{N-dim}(F) \leq d\}
\end{aligned}$$

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

$$\begin{aligned}
\text{N-dim}(F) &\leq \text{GC-dim}(F) \\
\text{GC-dim}(F) &\leq \text{P-dim}(F) \\
\text{GC-dim}(F) &\leq \text{G-dim}(F)
\end{aligned}$$

Define

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

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 [Alo83] [KM78] [Ste78]. Unfortunately, if $N_1 = ... = N_m = N$, the bounds obtained grow exponentially with $N$. For applications such as that given in section 3, a generalization of shattering which gives rise to bounds on $|F|$ that grow polynomially in $m$ and $N$ is desirable. We extend Sauer's result to some such generalizations which were given in [Nat89] [Pol84].

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

$$\bigcup_{k \in \mathbf{N}} \mathbf{N}^k \times \mathbf{N}^k$$

by defining

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

Extend $\phi$ further to $2^{\mathbf{N}^k} \times \mathbf{N}^k$ by defining

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

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

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

We say that $\bar{y}$ witnesses $F$'s $\phi$-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 $\phi$-dimension of $F$ is the cardinality of the largest subset of $\mathbf{N}_m$ shattered by $F$.

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

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

and define the Pollard-dimension (hereafter called the P-dimension) of $F$ to be its $\phi$-dimension in this case. We denote the P-dimension of $F$ by P-dim($F$). This definition is discussed in [Hau89][Pol84][Pol89]. It is called the pseudo dimension in [Pol89] and the combinatorial dimension in [Hau89].

Graph-shattering (G-shattering) and the Graph-dimension (G-dimension) are defined similarly with $\phi$ defined by

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

The G-dimension of $F$ is denoted by G-dim($F$). This definition is treated in [Nat89].

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

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

# 1. Introduction

Let $\mathbf{N}$ denote the positive integers, $\mathbf{Z}^+$ denote the nonnegative integers and $\mathbf{Z}$ denote the integers. Let $\mathbf{N}_0 = \emptyset$ and for each $m \in \mathbf{N}$, let $\mathbf{N}_m$ be the set $\{1, ..., m\}$.

We begin by stating Sauer's result [Sau72]. Vapnik and Chervonenkis independently proved a similar lemma in [VC71].

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

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

The Vapnik-Chervonenkis(VC) dimension of $F$ [VC71] is the cardinality of the largest subset of $\mathbf{N}_m$ shattered by $F$.

**Theorem 1.1 ([Sau72]):** *If the VC-dimension of $F$ is $d$, then*

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

*and this bound is tight; i.e., for all $d, m \in \mathbf{Z}^+, d \leq m$, there exists $F \subseteq 2^{\mathbf{N}_m}$ 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 1.1.

Following [Bon72], let $(m, k) \rightarrow (n, l)$ denote the statement: If $F \subseteq 2^{\mathbf{N}_m}$, $|F| = k$, then there exists $I \subseteq \mathbf{N}_m$ such that $|I| = n$

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

Sauer's result can now be stated as

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

Proofs of other statements of the form $(m, k) \rightarrow (n, l)$ are given in [Bon72] [Fra83] [Tom81].

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

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

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

Let $I = \{i_1, ..., i_k\} \subseteq \mathbf{N}_m$, with $i_1 < i_2 < \cdots < i_k$. Define

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

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

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

# Contents

**1. Introduction** **2**

**2. Proofs of the results** **6**

**3. An application** **11**

**4. Conclusion** **17**

**References** **18**

# A Generalization of Sauer's Lemma

David Haussler
Phil Long

UCSC-CRL-90-15
January 24, 1991

Board of Studies in Computer and Information Sciences
University of California at Santa Cruz
Santa Cruz, CA 95064

ABSTRACT

We generalize Sauer's lemma to multivalued functions. 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.