Applicable Analysis and Discrete Mathematics (April, 2007)

DENSITY OF SMOOTH BOOLEAN FUNCTIONS

Joel Ratsaby

The Sauer-Shelah lemma has been instrumental in the analysis of algorithms in many areas including learning theory, combinatorial geometry, graph theory. Algorithms over discrete structures, for instance, sets of Boolean functions, often involve a search over a constrained subset which satisfies some properties. In this paper we study the complexity of classes of functions h of finite VC-dimension which satisfy a local 'smoothness' property expressed as having long repeated values around elements of a given sample. A tight upper bound is obtained on the density of such classes. It is shown to possess a sharp threshold with respect to the smoothness parameter.

1 INTRODUCTION

Let $[n] = \{1, \ldots, n\}$ and denote by $2^{[n]}$ the class of all 2^n functions $h : [n] \rightarrow \{0, 1\}$. Let $\mathcal{H}$ be a class of functions and for a set $A = \{x_1, \ldots, x_k\} \subseteq [n]$ denote by $h_{|A} = [h(x_1), \ldots, h(x_k)]$. The *trace* of $\mathcal{H}$ on A is defined as $\mathrm{tr}_A(\mathcal{H}) = \{h_{|A} : h \in \mathcal{H}\}$. Define the *density function* $\rho_{\mathcal{H}}(k)$ of $\mathcal{H}$ as

$$\rho_{\mathcal{H}}(k) = \max_{A \subseteq [n] : |A| = k} \frac{|\mathrm{tr}_A(\mathcal{H})|}{2^k}.$$

The Vapnik-Chervonenkis dimension of $\mathcal{H}$, denoted as $\mathrm{VC}(\mathcal{H})$, is defined as the largest k such that $\rho_{\mathcal{H}}(k) = 1$. The following well known result obtained by [15, 12, 13] states that if $\mathrm{VC}(\mathcal{H}) < n$, then $\rho_{\mathcal{H}}(n)$ decreases at a rate of $O\left(\frac{n^{VC(\mathcal{H})}}{2^n}\right)$.

Lemma 1 *For any $1 \leq d < n$ let*

$$\mathbb{S}(n, d) = \sum_{k=0}^{d} \binom{n}{k}.$$

2006 Mathematics Subject Classification: 68Q25, 03C13, 51E20, 06E30, 11B05

Keywords: Binary functions, Binary sequences, Runs, VC-dimension

Then

$$\max_{\mathcal{H}\subset 2^{[n]}:\mathrm{VC}(\mathcal{H})=d} \rho_{\mathcal{H}}(n) = \frac{\mathbb{S}(n,d)}{2^n}.$$

Aside of being an interesting combinatorial result (see Chapter 17 in [4]), Lemma 1 has been instrumental in analysis of algorithms in statistical learning theory [14], combinatorial geometry [10], graph theory [9, 3] and in the theory of empirical processes [11]. In many problems which involve the analysis of discrete classes of structures, for instance, sets of Boolean functions, a search for some optimal element (target) in this set is employed based on an algorithm which uses available partial information, for instance in the form of a sample. This information effectively induces a smaller class of possible functions. The estimation of the density of such a class is important for analyzing the accuracy and the convergence properties of the algorithm. In this paper we study the density of finite VC-dimension classes of Boolean functions which are locally-smooth, i.e., have a repeated value over subsets of consecutive elements of $[n]$. In practice, this type of property is easy to measure and is a typical form of prior knowledge about the unknown target function.

Formally, such classes may be introduced by defining the following measure: for $h:[n]\to\{0,1\}$, $x\in[n]$ and $y\in\{0,1\}$ let the *width* $\omega_h(x,y)$ of h at x with respect to y be the largest $0\le a\le n$ such that $h(z)=y$ for all $x-a\le z\le x+a$; if no such a exists then let $\omega_h(x,y)=-1$. Denote by $\Xi=[n]\times\{0,1\}$. For a sample $\zeta_\ell=\{(x_i,y_i)\}_{i=1}^{\ell}\in\Xi^\ell$, define by $\omega_{\zeta_\ell}(h)=\min_{1\le i\le\ell}\omega_h(x_i,y_i)$ the width of h with respect to ζ. For instance, Figure 1 displays a sample $\zeta_2=\{(x_1,y_1),(x_2,y_2)\}$ and

y					y_1=1									y_2=0							
h_1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0
h_2	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	0
[n]	1	2	.	.	x_1	.	.	.	.	.	.	.	.	x_2	.	.	.	.	.	.	n

Figure 1: $\omega_{\zeta_2}(h_1)=\omega_{\zeta_2}(h_2)=3$

two functions h_1, h_2 which have a width of 3 with respect to ζ_2. The classes of Boolean functions on $[n]$ which we study have a constraint on the width, i.e.,

$$\mathcal{H}_N(\zeta_\ell)=\{h\in\mathcal{H}:\omega_{\zeta_\ell}(h)>N\},\ N\ge 0 \tag{1}$$

where $\zeta_\ell=\{(x_i,y_i)\}_{i=1}^{\ell}\in\Xi^\ell$ is a given sample. In this paper we obtain tight bounds (in the form of Lemma 1) on the density of such a class. As it turns out, the bounds have sharp thresholds with respect to the width parameter value. In subsequent sections we investigate this in detail.

For a function $h:[n]\to\{0,1\}$ let the *difference* function be defined as

$$\delta_h(x)=\begin{cases}1 & \text{if } h(x-1)=h(x)\\ 0 & \text{otherwise}\end{cases}$$

where we assume that any h satisfies $h(0) = 0$ (see Figure 2). Define

h	1	1	1	1	1	0	0	0	0	1	1	1	0	0	0	0	1	0	1	0	0	0
δ_h	0	1	1	1	1	0	1	1	1	0	1	1	0	1	1	1	0	0	0	0	1	1
[n]	1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	n

Figure 2: h and the corresponding δ_h

$$\mathcal{D}_{\mathcal{H}} \equiv \{\delta_h : h \in \mathcal{H}\}, \tag{2}$$

or for brevity we write $\mathcal{D}$. It is easy to see that the class $\mathcal{D}$ is in one-to-one correspondence with $\mathcal{H}$. For $N \geq 0$ and any sample ζ, if $\omega_h(x, y) \leq N$ for $(x, y) \in \zeta$ then the corresponding δ_h has $\omega_{\delta_h}(x, 1) \leq N$. So in order to obtain estimates on the cardinality of classes $\mathcal{H}_N(\zeta)$, it suffices to estimate the cardinality of the corresponding difference classes $\mathcal{D}_N(\zeta_+)$, defined based on $\zeta_+ = \{(x_i, 1) : (x_i, y_i) \in \zeta, 1 \leq i \leq \ell\}$, which turns out to be simpler. We denote by $\mathrm{VC}_\Delta(\mathcal{H})$ the VC-dimension of the difference class $\mathcal{D} = \{\delta_h : h \in \mathcal{H}\}$ and use it to characterize the complexity of $\mathcal{H}$ (it is straightforward to show that $\mathrm{VC}_\Delta(\mathcal{H}) \leq c\mathrm{VC}(\mathcal{H})$ for a small positive absolute constant c). Henceforth we use d as a parameter value of $VC_\Delta(\mathcal{H})$.

The remaining parts of the paper are organized as follows: in Section 2 we state the main results, Section 3 contains the lemmas used for proving the first two results and the sketches of the proof of the remaining results.

2 MAIN RESULTS

The first result concerns classes of functions constrained by an upper bound on the width. For any class $\mathcal{H}$ of binary functions on $[n]$ define

$$\mathcal{H}_N = \{h \in \mathcal{H} : \omega_h(x, h(x)) \leq N, x \in [n]\},\ N \geq 0 \tag{3}$$

where, as for $\mathcal{H}_N(\zeta_\ell)$ in (1), the dependence of $\mathcal{H}_N$ on $\mathcal{H}$ is left implicit.

Theorem 1 *Let $1 \leq d \leq n$ and $N \geq 0$. Then*

$$\max_{\mathcal{H} \subset 2^{[n]} : \mathrm{VC}_\Delta(\mathcal{H}) = d} \rho_{\mathcal{H}_N}(n) = \frac{\beta_d^{(N)}(n)}{2^n} \tag{4}$$

where $\beta_d^{(N)}(n)$ is defined in Lemma 3.

The proof follows from Lemma 4 in Section 3.1.

If we define the threshold of $\beta_d^{(N)}(n)$ as the point N^* at which $\beta_d^{(N)}(n)$ reaches half of its maximal value then $\beta_d^{(N)}(n)$ has a sharp transition at N^*; an example is displayed in Figure 3. The next result states an estimate for N^*.

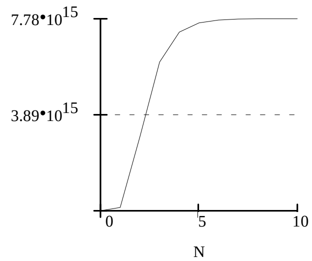

Figure 3: $\beta_{20}^{(N)}(60)$

Theorem 2 *Let $0 < \alpha < 1/2$ and $d = d_n = \alpha n$ then for large n, N^* is approximated by $c \ln d$ for some c dependent on α.*

The proof follows from Lemma 6 in Section 3.2. The next two results concern classes of functions with a lower-bound on the width as defined in (1).

Theorem 3 *Let $1 \leq d, \ell \leq n$ and $N \geq 0$. Then*

$$\max_{\mathcal{H} \subset 2^{[n]}, \zeta \in \Xi^{\ell} : \mathrm{VC}_{\Delta}(\mathcal{H}) = d} \rho_{\mathcal{H}_N(\zeta)}(n) = \frac{\mathbb{S}(n - \ell - 2N - 1, d)}{2^n} \tag{5}$$

which is bounded from above by $(1 + e^{-(\ell+2N+1)/n} \mathbb{S}(n, d)) 2^{-n}$.

The proof is in Section 3.3.

Next, consider an extremal case where the width of h is larger than N *only* on elements of ζ, for all $h \in \mathcal{H}_N(\zeta)$. In this case the class is defined as

$$\mathcal{H}_N^*(\zeta) = \{h \in \mathcal{H} : \omega_h(x, h(x)) > N \text{ iff } (x, h(x)) \in \zeta\}, \; N \geq 0.$$

This type of class arises in certain applications where given a sample ζ an algorithm obtains a solution, i.e., a binary function, which maximizes the width on ζ.

Theorem 4 *Let $1 \leq d, \ell \leq n$ and $N \geq 0$. Then*

$$\max_{\mathcal{H} \subset 2^{[n]}, \zeta \in \Xi^{\ell} : \mathrm{VC}_{\Delta}(\mathcal{H}) = d} \rho_{\mathcal{H}_N^*(\zeta)}(n) = \frac{\beta_d^{(N)}(n - \ell - 2N - 1)}{2^n} \tag{6}$$

where

$$\beta_d^{(N)}(n - \ell - 2N - 1) \leq 3 e^{-e^{-(2N+1)}} \left(1 + e^{-(\ell+2N+1)/n} \mathbb{S}(n, d)\right). \tag{7}$$

Its maximum value with respect to N is approximated by

$$N' = (\ln(n) - 1)/2. \tag{8}$$

The sketch of the proof is in Section 3.4.

Comparing (6) against (5) then N' is a critical point where, roughly, only when $N \leq N'$ the bound on the extremal class $\mathcal{H}_N^*(\zeta)$ is smaller than the bound on $\mathcal{H}_N(\zeta)$ while for $N > N'$ they are approximately equal. An example of their ratio is shown in Figure 4.

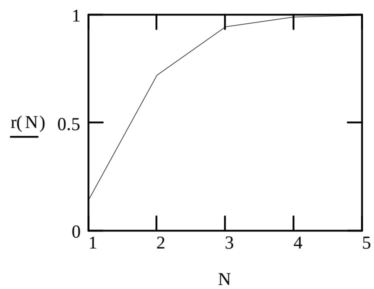

Figure 4: $r(N) = \beta_d^{(N)}(n' - 2N)/\mathbb{S}(n' - 2N, d)$ for $n' = n - \ell - 1$, $n = 300$, $d = 20$, $\ell = 20$, $N' \approx 3$

3 TECHNICAL WORK

We start with several lemmas used in proving the first Theorem.

3.1 LEMMAS FOR THEOREM 1

Let $\binom{n}{k}$ denote the following function

$$\binom{n}{k} = \begin{cases} n!/(k!(n-k)!) & \text{if } 0 \leq k \leq n \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathbb{I}(\mathbb{E})$ denote the indicator function which equals 1 if the expression E is true and 0 otherwise.

Lemma 2 *For any integer n, $\nu \geq 0$, $m \leq n$, define the following:*

$$w_{m,\nu}(n) = \begin{cases} 0 & \textit{if } n < 0 \\ \mathbb{I}(n = 0) & \textit{if } m = 0 \textit{ or } \nu = 0 \\ \displaystyle\sum_{i=0,\nu+1,2(\nu+1),\ldots}^{n} (-1)^{i/(\nu+1)} \binom{m}{i/(\nu+1)} \binom{n-i+m-1}{n-i} & \textit{if } m \geq 1. \end{cases}$$

Then for a nonnegative integer n, the number of standard (one-dimensional) ordered partitions of n into m parts each no larger than ν is equal to $w_{m,\nu}(n)$.

Proof: The generating function (g.f.) for $w_{m,\nu}(n)$ is

$$W(x) = \sum_{n \geq 0} w_{m,\nu}(n) x^n = \left(\frac{1 - x^{\nu+1}}{1 - x} \right)^m .$$

When $m = 0$ or $\nu = 0$ the only non-zero coefficient is of x^0 and it equals 1 so $w_{m,\nu}(n) = \mathbb{I}(n = 0)$. Let $T(x) = (1 - x^{\nu+1})^m$ and $S(x) = \left(\frac{1}{1-x}\right)^m$. Then

$$T(x) = \sum_{i=0}^{m} (-1)^i \binom{m}{i} x^{i(\nu+1)}$$

which generates the sequence $t_\nu(n) = \binom{m}{n/(\nu+1)} (-1)^{n/(\nu+1)} \mathbb{I}(n \bmod (\nu + 1) = 0)$. Similarly, for $m \geq 1$, it is easy to show $S(x)$ generates $s(n) = \binom{n+m-1}{n}$. The product $W(x) = T(x)S(x)$ generates their convolution $t_\nu(n) \star s(n)$, namely,

$$w_{m,\nu}(n) = \sum_{i=0,\nu+1,2(\nu+1),\ldots,}^{n} (-1)^{i/(\nu+1)} \binom{m}{i/(\nu+1)} \binom{n - i + m - 1}{n - i}. \quad \blacksquare$$

Remark 1 *While our interest is in $[n] = \{1, \ldots, n\}$, we allow $w_{m,\nu}(n)$ to be defined on $n \leq 0$ for use by Lemma 3.*

Remark 2 *This expression may alternatively be expressed as*

$$w_{m,\nu}(n) = \sum_{k=0}^{m} (-1)^k \binom{m}{k} \binom{n + m - 1 - k(\nu + 1)}{m - 1},$$

over $m \geq 1$.

We need two additional lemmas for proving (4) of Theorem 1.

Lemma 3 *Let the integer $N \geq 0$ and consider the class F of all binary-valued functions f on $[n]$, or equivalently, sequences $f = f(1), \ldots, f(n)$, satisfying: (a) f has no more than r 1's (b) every run of consecutive 1's in f is no longer than $2N + 1$, except for a run that starts at $f(1)$ which may be of length $2(N + 1)$. Then*

$$|F| = \beta_r^{(N)}(n)$$

where

$$\beta_r^{(N)}(n) \equiv \sum_{k=0}^{r} \sum_{m=1}^{n} c(k, n - k; m, N)$$

with

$$c(k, n - k; m, N) = \binom{n - k}{m - 1} \cdot$$
$$\cdot \Big(w_{m,2N}(k - m + 1) + w_{m-1,2N}(k - m - 2N) + w_{m-1,2N}(k - m - 2N - 1) \Big).$$

Remark 3 *Note that when $r \leq 2N+1$, $\beta_r^{(N)}(n) = \mathbb{S}(n,d)$.*

Proof: Consider the integer pair $[k, n-k]$, where $n \geq 1$ and $0 \leq k \leq n$. A two-dimensional ordered m-partition of $[k, n-k]$ is an ordered partition into m two-dimensional parts, $[a_j, b_j]$ where $0 \leq a_j, b_j \leq n$ but not both are zero and where $\sum_{j=1}^{m}[a_j, b_j] = [k, n-k]$. For instance, $[2,1] = [0,1] + [2,0] = [1,1] + [1,0] = [2,0] + [0,1]$ are three partitions of $[2,1]$ into two parts (for more examples see [1]).

Suppose we add the constraint that only a_1 or b_m may be zero while all remaining

$$a_j, b_k \geq 1, \quad 2 \leq j \leq m,\ 1 \leq k \leq m-1. \tag{9}$$

Denote any partition that satisfies this as *valid*. For instance, let $k = 2$, $m = 3$ then the m-partitions of $[k, n-k]$ are: $\{[0,1][1,1][1,n-4]\}$,$\{[0,1][1,2][1,n-5]\}$, $\ldots$,$\{[0,1][1,n-3][1,0]\}$, $\{[0,2][1,1][1,n-5]\}$, $\{[0,2][1,2][1,n-6]\}$, $\ldots$, $\{[0,2][1,n-4][1,0]\}$, $\ldots$, $\{[0,n-3][1,1][1,0]\}$. For $[k, n-k]$, let $\mathcal{P}_{n,k}$ be the collection of all valid partitions of $[k, n-k]$.

Let F_k denote all binary functions on $[n]$ which take the value 1 over exactly k elements of $[n]$. Define the mapping $\Pi : F_k \to \mathcal{P}_{n,k}$ where for any $f \in F_k$ the partition $\Pi(f)$ is defined by the following procedure: Start from the first element of $[n]$, i.e., 1. If f takes the value 1 on it then let a_1 be the length of the constant 1-segment, i.e., the set of all elements starting from 1 on which f takes the constant value 1. Otherwise if f takes the value 0 let $a_1 = 0$. Then let b_1 be the length of the subsequent 0-segment on which f takes the value 0. Let $[a_1, b_1]$ be the first part of $\Pi(f)$. Next, repeat the following: if there is at least one more element of $[n]$ which has not been included in the preceding segment, then let a_j be the length of the next 1-segment and b_j the length of the subsequent 0-segment. Let $[a_j, b_j]$, $j = 1, \ldots, m$, be the resulting sequence of parts where m is the total number of parts. Only the last part may have a zero valued b_m since the function may take the value 1 on the last element n of $[n]$ while all other parts, $[a_j, b_j]$, $2 \leq j \leq m-1$, must have $a_j, b_j \geq 1$. The result is a valid partition of $[k, n-k]$ into m parts.

Clearly, every $f \in F_k$ has a unique partition. Therefore Π is a bijection. Moreover, we may divide $\mathcal{P}_{n,k}$ into mutually exclusive subsets V_m consisting of all valid partitions of $[k, n-k]$ having exactly m parts, where $1 \leq m \leq n$. Thus

$$|F_k| = \sum_{m=1}^{n} |V_m|.$$

Consider the following constraint on components of parts:

$$a_i \leq \begin{cases} 2N+1 & \text{if } 2 \leq i \leq m \\ 2(N+1) & \text{if } i = 1. \end{cases} \tag{10}$$

Denote by $V_{m,N} \subset \mathcal{P}_{n,k}$ the collection of valid partitions of $[k, n-k]$ into m parts each of which satisfies this constraint.

Let $F_{k,N} = F \cap F_k$ consist of all functions satisfying the run-constraint in the statement of the lemma and having exactly k ones. If f has no run of consecutive 1's starting at $f(i)$ of length larger than $2N+1$ then there does not exist a segment a_i of length larger than $2N+1$, $i \geq 2$ (and similarly with a run of size $2(N+1)$ starting at $f(1)$). Hence the parts of $\Pi(f)$ satisfy (10) and for any $f \in F_{k,N}$, its unique valid partition $\Pi(f)$ must be in $V_{m,N}$. We therefore have

$$|F_{k,N}| = \sum_{m=1}^{n} |V_{m,N}|. \tag{11}$$

By definition of F it follows that

$$|F| = \sum_{k=0}^{r} |F_{k,N}|. \tag{12}$$

Let us denote by

$$c(k, n-k; m, N) \equiv |V_{m,N}| \tag{13}$$

the number of valid partitions of $[k, n-k]$ into exactly m parts whose components satisfy (10). In order to determine $|F|$ it therefore suffices to determine $c(k, n-k; m, N)$.

We next construct the generating function

$$G(t_1, t_2) = \sum_{\alpha_1 \geq 0} \sum_{\alpha_2 \geq 0} c(\alpha_1, \alpha_2; m, N) t_1^{\alpha_1} t_2^{\alpha_2}. \tag{14}$$

For $m \geq 1$,

$$\begin{aligned} G(t_1, t_2) = {} & (t_1^0 + t_1^1 + \cdots + t_1^{2N+2})(t_2^1 + t_2^2 + \cdots)^{\mathbb{I}(m \geq 2)} \\ & \cdot \left((t_1^1 + \cdots + t_1^{2N+1})(t_2^1 + t_2^2 + \cdots)\right)^{(m-2)_+} \\ & \cdot (t_1^1 + \cdots + t_1^{2N+1})^{\mathbb{I}(m \geq 2)}(t_2^0 + t_2^1 + \cdots) \end{aligned} \tag{15}$$

where the values of the exponents of all terms in the first and second factors represent the possible values for a_1 and b_1, respectively. The values of the exponents in the middle $m-2$ factors are for the values of a_j, b_j, $2 \leq j \leq m-1$ and those in the factor before last and last are for a_m and b_m, respectively. Equating this to (14) implies the coefficient of $t_1^{\alpha_1} t_2^{\alpha_2}$ equals $c(\alpha_1, \alpha_2; m, N)$ which we seek.

The right side of (15) equals

$$t_1^{m-1} t_2^{m-1} \left(\frac{1}{1-t_2}\right)^m \left(\left(\frac{1-t_1^{2N+1}}{1-t_1}\right)^m + t_1^{2N+1}(1+t_1)\left(\frac{1-t_1^{2N+1}}{1-t_1}\right)^{m-1}\right). \tag{16}$$

Let $W(x) = \left(\frac{1-x^{2N+1}}{1-x}\right)^{m-1}$ generate $w_{m-1,2N}(n)$ which is defined in Lemma 2 and denote by $s(n) = \binom{n+m-1}{n}$. So (16) becomes

$$\sum_{\alpha_1,\alpha_2\geq 0} s(\alpha_2)t_2^{\alpha_2+m-1}\left(w_{m,2N}(\alpha_1)t_1^{\alpha_1+m-1} + w_{m-1,2N}(\alpha_1)t_1^{\alpha_1+m+2N}(1+t_1)\right). \quad (17)$$

Equating the coefficients of $t_1^{\alpha_1'}t_2^{\alpha_2'}$ in (14) and (17) yields

$$c(\alpha_1',\alpha_2';m,N) = s(\alpha_2'-m+1)\cdot$$
$$\cdot\Big(w_{m,2N}(\alpha_1'-m+1) + w_{m-1,2N}(\alpha_1'-m-2N) + w_{m-1,2N}(\alpha_1'-m-2N-1)\Big).$$

Replacing $s(\alpha_2'-m+1)$ by $\binom{\alpha_2'}{m-1}$, substituting k for α_1', $n-k$ for α_2' and combining (11), (12) and (13) yields the result. ■

The next lemma extends the result of Lemma 3 to the class $\mathcal{H}_N$ defined in (3).

Lemma 4 *Let $1 \leq d \leq n$ and $N \geq 0$. For any class $\mathcal{H}$ with $VC_\Delta(\mathcal{H}) = d$, the cardinality of the corresponding class $\mathcal{H}_N$ defined in (3) is no larger than $\beta_d^{(N)}(n)$. This bound is tight.*

Proof: Denote by $\mathcal{D}_N = \{\delta_h : h \in \mathcal{H}_N\}$. Clearly, $|\mathcal{D}_N| = |\mathcal{H}_N|$. Consider any $h \in \mathcal{H}_N$. Since for all $x \in [n]$, $\omega_h(x, h(x)) \leq N$ then the corresponding δ_h in $\mathcal{D}_N$ satisfies the following: every run of consecutive 1's is of length no larger than $2N+1$, except for a run which starts at $x = 1$ whose length may be as large as $2(N+1)$. Let $\mathcal{F}_N$ be the set system corresponding to the class $\mathcal{D}_N$ which is defined as follows:

$$\mathcal{F}_N = \{A_\delta : \delta \in \mathcal{D}_N\},\ A_\delta = \{x \in [n] : \delta(x) = 1\}.$$

Clearly, $|\mathcal{F}_N| = |\mathcal{D}_N|$. Note that the above constraint on δ translates to A_δ possessing the property P_N defined as having every subset $E \subseteq A_\delta$ which consists of consecutive elements $E = \{i, i+1, \ldots, j-1, j\}$ be of cardinality $|E| \leq 2N+1$, except for such an E that contains the element $\{1\}$ which may have cardinality as large as $2(N+1)$. Hence for every element $A \in \mathcal{F}_N$, A satisfies P_N. This is denoted by $A \models P_N$. Let $G_{\mathcal{F}}(k) \equiv \max\{|\{A \cap E : A \in \mathcal{F}_N\}| : E \subseteq [n], |E| = k\}$. The corresponding notion of VC-dimension for a class $\mathcal{F}_N$ of sets is the the so-called *trace number* ([4], p.131) and is defined as $tr(\mathcal{F}_N) = \max\{m : G_{\mathcal{F}_N}(m) = 2^m\}$. Clearly, $tr(\mathcal{F}_N) = VC(\mathcal{D}_N) \leq \mathrm{VC}(\mathcal{D}) \equiv \mathrm{VC}_\Delta(\mathcal{H}) = d$ (where $\mathcal{D}$ is defined in (2)).

The proof proceeds as in the proof of Lemma 1 (for instance [2], Theorem 3.6) which is based on the shifting method (see [4] Ch. 17, Theorem 1 & 4 and also [8, 6, 5]). The idea is to transform $\mathcal{F}_N$ into an *ideal* family $\mathcal{F}_N'$ of sets E, i.e., if $E \in \mathcal{F}_N'$ then $S \in \mathcal{F}_N'$ for every $S \subset E$, and such that $|\mathcal{F}_N| = |\mathcal{F}_N'| \leq \beta_d^{(N)}(n)$.

Start by defining the operator T_x on $\mathcal{F}_N$ which removes an element $x \in [n]$ from every set $A \in \mathcal{F}_N$ provided that this does not duplicate any existing set. It is defined as follows:

$$T_x(\mathcal{F}_N) = \{A \setminus \{x\} : A \in \mathcal{F}_N\} \cup \{A \in \mathcal{F}_N : A \setminus \{x\} \in \mathcal{F}_N\}.$$

Consider now

$$\mathcal{F}'_N = T_1(T_2(\cdots T_n(\mathcal{F}_N)\cdots))$$

and denote the corresponding function class by $\mathcal{D}'_N$. Clearly, $|\mathcal{D}'_N| = |\mathcal{F}'_N|$.

We have $|\mathcal{F}'_N| = |\mathcal{F}_N|$ since the only time that the operator T_x changes an element A into a different set $A^* = T_x(A)$ is when A^* does not already exist in the class so no additional element in the new class can be created. It is also clear that for all $x \in [n]$, $T_x(\mathcal{F}'_N) = \mathcal{F}'_N$ since for each $E \in \mathcal{F}'_N$ there exists a G that differs from it on exactly one element hence it is not possible to remove any element $x \in [n]$ from all sets without creating a duplicate. Applying this repeatedly implies that $\mathcal{F}'_N$ is an ideal. Furthermore, since for all $A \in \mathcal{F}_N$, $A \models P_N$ then removing an element x from A still leaves $A \setminus \{x\} \models P_N$. Hence for all $E \in \mathcal{F}'_N$ we have $E \models P_N$.

From Lemma 3 ([4], p.133) we have $G_{\mathcal{F}'_N}(k) \le G_{\mathcal{F}_N}(k)$, for all $1 \le k \le n$. Since $tr(\mathcal{F}_N) \le d$ then $tr(\mathcal{F}'_N) \le d$. Together with $\mathcal{F}'_N$ being an ideal it follows that for all $E \in \mathcal{F}'_N$, $|E| \le d$. For all $E \in \mathcal{F}'_N$, $E \models P_N$ hence the corresponding class $\mathcal{D}'_N$ satisfies the following: for all $\delta \in \mathcal{D}'_N$, δ has at most d 1's and every run of consecutive 1's is of length no larger than $2N+1$ except possibly for a run which starts at $x = 1$ which may be as large as $2(N+1)$. By Lemma 3 above, we therefore have $|\mathcal{D}'_N| \le \beta_d^{(N)}(n)$. We conclude that $|\mathcal{H}_N| = |\mathcal{D}_N| = |\mathcal{F}_N| = |\mathcal{F}'_N| = |\mathcal{D}'_N|$ and hence $|\mathcal{H}_N| \le \beta_d^{(N)}(n)$. This bound is tight since consider $\mathcal{H}^*$ whose corresponding class $\mathcal{D}^*$ has *all* functions on $[n]$ with at most d 1's. Clearly, $\mathrm{VC}_\Delta(\mathcal{H}^*) = \mathrm{VC}(\mathcal{D}^*) = d$. The cardinality of $\mathcal{H}^*_N$ equals that of $\mathcal{D}^*_N$ which consists of all $\delta \in \mathcal{D}^*$ that satisfy the above condition on runs of 1's. Clearly, $|\mathcal{D}^*_N| = \beta_d^{(N)}(n)$. ■

Remark 4 *As indicated in Remark 3, when N is greater than $(d-1)/2$ the bound $\beta_d^{(N)}(n)$ is as in Lemma 1 and hence the effect of N is void. It turns out that this starts to happen at a much smaller value of N (see Remark 5).*

In the following section we study the function $\beta_d^{(N)}(n)$ with respect to N.

3.2 LEMMAS FOR THEOREM 2

We start with a lemma that estimates $c(k, n-k; m, N)$ (defined in (9)) which is the number of two-dimensional valid ordered m-partitions of $[k, n-k]$ satisfying (10) where a valid partition is defined according to (9).

Lemma 5 *For $n \ge k \ge m-1 \ge 1$ we have*

$$c(k, n-k; m, N) = b_1\,(1 + b_2\alpha)\,(1-\alpha)^m \binom{k}{m-1}\binom{n-k}{m-1}$$

for some absolute positive constants $b_1 \le 1$, $b_2 \le 2$ and $\alpha = \alpha(N, m, k) \equiv e^{-(2N+1)(m-1)/k}$.

Proof sketch: By definition, from (9) the quantity $c(k, n-k; m, N)$ involves a sum of three terms, $w_{m,2N}(k-m+1)$, $w_{m-1,2N}(k-m-2N-1)$ and $w_{m-1,2N}(k-m-2N)$. Using Remark 2 the first equals

$$w_{m,2N}(k-m+1) = \sum_{l=0}^{m}(-1)^l \binom{m}{l}\binom{k-l(2N+1)}{m-1}. \tag{18}$$

By Lemma 2 we have $w_{m-1,2N}(k-m-2N) \leq w_{m,2N}(k-m-2N)$ and $w_{m-1,2N}(k-m-2N-1) \leq w_{m,2N}(k-m-2N-1)$. We have

$$w_{m,2N}(k-m-2N) = \sum_{l=0}^{m}(-1)^l \binom{m}{l}\binom{k-l(2N+1)-(2N+1)}{m-1}$$

and similarly for $w_{m,2N}(k-m-2N-1)$. Hence

$$c(k, n-k; m, N) \quad = \quad \binom{n-k}{m-1}\sum_{l=0}^{m}(-1)^l \binom{m}{l}\binom{k-l(2N+1)}{m-1}(1+\epsilon(m,k,N,l))$$

where

$$0 < \epsilon(m,k,N,l) \leq \frac{\binom{k-l(2N+1)-(2N+1)}{m-1}}{\binom{k-l(2N+1)}{m-1}} + \frac{\binom{k-l(2N+1)-2(N+1)}{m-1}}{\binom{k-l(2N+1)}{m-1}}$$

which for all $0 \leq l \leq m$ is bounded from above by

$$\frac{\binom{k-(2N+1)}{m-1}}{\binom{k}{m-1}} + \frac{\binom{k-2(N+1)}{m-1}}{\binom{k}{m-1}}. \tag{19}$$

Using a standard combinatoric identity it is easy to show that both terms of (19) are bounded from above by $\alpha = \alpha(m,k,N) = \exp(-(2N+1)(m-1)/k)$. The same argument applied on $\binom{k-l(2N+1)}{m-1}$ completes the stated result. ■

Lemma 6 *Let N^* be the value at which the function $\beta_d^{(N)}(n)$ reaches half of its maximum value. Assume $1 \leq d < n/2$ and denote by $t = 1 + d(n-d)/n$ then N^* is approximated by*

$$\frac{n}{2(n-d)} \ln\left(\frac{2b_2 t}{b_2 - t + \sqrt{(b_2+t)^2 - 2tb_2/b_1}}\right)$$

for some absolute positive constants $b_1 \leq 1$, $b_2 \leq 2$.

Remark 5 *It follows that for $0 < \alpha < 1/2$, $d = d_n = \alpha n$ then for large n, N^* is approximated by $c \ln d$ for some $c > 0$ dependent on α.*

Proof sketch: We seek the solution N^* of the equation

$$\sum_{k=0}^{d}\sum_{m=1}^{n} c(k, n-k; m, N) = \frac{1}{2}\sum_{k=0}^{d}\binom{n}{k}$$

which, using Lemma 5 and a common identity (see [7](5.23)), can be approximated by the solution of

$$\sum_{k=0}^{d}\sum_{m=1}^{n}\binom{k}{m-1}\binom{n-k}{m-1}\left(f(m)-\frac{1}{2}\right)=0$$

where

$$f(m) = b_1\left(1+b_2 e^{-(2N+1)(m-1)/k}\right)\left(1-e^{-(2N+1)(m-1)/k}\right)^m,$$

$0 < b_1, b_2 \leq 2$. The first sum is approximated as

$$\sum_{m=1}^{n}\binom{k}{m-1}\binom{n-k}{m-1}f(m) \approx f(m^*)\sum_{m=1}^{n}\binom{k}{m-1}\binom{n-k}{m-1} = f(m^*)\binom{n}{k}$$

where $\binom{k}{m-1}\binom{n-k}{m-1}$ peaks at $m = m^* \equiv 1 + k(n-k)/n$. Hence the solution may be approximated by solving

$$\sum_{k=0}^{d}\binom{n}{k}\left(b_1\left(1+b_2 e^{-(2N+1)(m^*-1)/k}\right)\left(1-e^{-(2N+1)(m^*-1)/k}\right)^{m^*}-\frac{1}{2}\right)=0$$

for N. For $1 \leq d < n/2$, the dominant term is $k = d$. Simple calculus then yields the result. ∎

3.3 SKETCH OF PROOF OF THEOREM 3

Fix any $(x, y) \in \zeta$. The condition $\omega_h(x, y) > N$ implies that h must have a constant value of y over all elements z, $x - N - 1 \leq z \leq x + N + 1$. For this x, the uniquely corresponding δ_h has a constant value of 1 over the interval $I_N(x) \equiv \{z : x - N \leq z \leq x + N + 1\}$. By definition of $\mathcal{H}_N(\zeta)$ this holds for any $(x, y) \in \zeta$. Denote by $\mathcal{D}_N(\zeta_+) = \{\delta_h : h \in \mathcal{H}_N(\zeta)\}$ where $\zeta_+ = \{x_i : (x_i, y_i) \in \zeta, 1 \leq i \leq \ell\}$. Clearly, $|\mathcal{D}_N(\zeta_+)| = |\mathcal{H}_N(\zeta)|$. Hence we seek an upper bound on $|\mathcal{D}_N(\zeta_+)|$ for any ζ_+ and $\mathcal{H}$ with $\mathrm{VC}_\Delta(\mathcal{H}) = d$.

Let $R(\zeta_+) = \bigcup_{x\in\zeta_+} I_N(x)$. Since for every $\delta \in \mathcal{D}_N(\zeta_+)$, $\delta(z) = 1$ for all $z \in R(\zeta_+)$ then the cardinality of the restriction $\mathcal{D}_N(\zeta_+)_{|R(\zeta_+)}$ of the class $\mathcal{D}_N(\zeta_+)$ on the set $R(\zeta_+)$ equals one. Denote by $R^c(\zeta_+) \equiv [n] \setminus R(\zeta_+)$ then we have

$$|\mathcal{D}_N(\zeta_+)| = |\mathcal{D}_N(\zeta_+)_{|R^c(\zeta_+)}|.$$

Since $\mathrm{VC}(\mathcal{D}_N(\zeta_+)) \leq \mathrm{VC}_\Delta(\mathcal{H}) = d$ then by Lemma 1 it follows that

$$|\mathcal{D}_N(\zeta_+)_{|R^c(\zeta_+)}| \leq \mathbb{S}(|R^c(\zeta_+)|, d). \tag{20}$$

We also have

$$\max\{|R^c(S)| : S \subset [n], |S| = \ell\} = n - \ell - 2N - 1 \tag{21}$$

which is achieved for instance by a set $S' = \{N+2, \ldots, N+\ell+1\}$ with $R(S') = \{2, \ldots, 2(N+1)+\ell\}$. Hence for any ζ_+ as above we have

$$|\mathcal{D}_N(\zeta_+)| \leq \mathbb{S}(n - 2N - \ell - 1, d). \tag{22}$$

Since the bound of Lemma 1 is tight then there exists a class $\mathcal{D}_N(\zeta_+)$ (with a corresponding class $\mathcal{H}_N(\zeta)$) of this size. The first claim of Theorem 3 follows. The right side of (22) may be bounded as in the statement of the theorem using a similar argument as in the proof of Lemma 5. ■

3.4 SKETCH OF PROOF OF THEOREM 4

The proof follows that of Theorem 3 up to (20) with $\mathcal{H}_N^*(\zeta)$ instead of $\mathcal{H}_N(\zeta)$. By Theorem 1 we have

$$|\mathcal{D}_N^*(\zeta_+)_{|R^c(\zeta_+)}| \leq \beta_d^{(N)}(|R^c(\zeta_+)|)$$

and from (21) the statement of (6) follows. By the tightness of the bound in Theorem 1 there exists a class $\mathcal{D}_N^*(\zeta_+)$ and hence $\mathcal{H}_N^*(\zeta)$ of this size. We now sketch the proof of the approximation statement of the theorem. Using Lemma 5 we have

$$\beta_d^{(N)}(n - \ell - 2N - 1) \leq 3 \sum_{k=0}^{d} \sum_{m=1}^{n'} \binom{k}{m-1} \binom{n'-k}{m-1} \left(1 - e^{-\frac{(2N+1)(m-1)}{k}}\right)^m \tag{23}$$

where $n' = n - \ell - 2N - 1$. Denote by

$$\mathbb{P}(m) = \frac{\binom{k}{m-1}\binom{n'-k}{m-1}}{\sum_{m=1}^{n'} \binom{k}{m-1}\binom{n'-k}{m-1}}$$

and consider bounding from above the quantity

$$\mathbb{E}\left(1 - e^{-\frac{(m-1)(2N+1)}{k}}\right)^m$$

where expectation is taken with respect to $\mathbb{P}$. Using Jensen's inequality, this leads to the following bound on the right side of (23),

$$3 \sum_{k=0}^{d} \binom{n'}{k} \left(1 - e^{-\frac{(\mu-1)(2N+1)}{k}}\right) \tag{24}$$

where μ is the mean of a random variable with probability distribution $\mathbb{P}$. Solving for the generating function of the sequence $f(n) \equiv \sum_{m\geq 1} m\binom{k}{m-1}\binom{n-k}{m-1}$ we obtain that $f(n) = k\binom{n-1}{k} + \binom{n}{k}$ which then yields

$$\mu = \frac{k(n-k)}{n} + 1.$$

Replacing n by n' above, substituting this for μ in (24) and using the inequality $1 - a \leq e^{-a}$ which holds for all $a \in \mathbb{R}$ gives (7). Using this estimate of $\beta_d^{(N)}(n - \ell - 2N - 1))$ we solve for the N' at which it is maximized. Simple calculus yields (8). ■

4 CONCLUSIONS

Letting the width of a binary function at x denote the degree to which it is smooth, i.e., constant around x, the paper extends the classical Sauer's lemma to Vapnik-Chervonenkis classes of binary functions which are smooth at elements of a sample. Using a novel approach based on a bijection between a class of such functions and integer partitions, the cardinality of such a class is computed. Tight upper bounds with a dependence on the width parameter N are obtained and shown to exhibit a sharp threshold with respect to N.

REFERENCES

[1] G. E. Andrews. *The Theory of Partitions*. Cambridge University Press, 1998.

[2] M. Anthony and P. L. Bartlett. *Neural Network Learning:Theoretical Foundations*. Cambridge University Press, 1999.

[3] M. Anthony, G. Brightwell, and C. Cooper. The Vapnik-Chervonenkis dimension of a random graph. *Discrete Mathematics*, 138(1-3):43–56, 1995.

[4] B. Bollobás. *Combinatorics: Set Systems, Hypergraphs, Families of vectors, and combinatorial probability*. Cambridge University Press, 1986.

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

[6] P. Frankl. The shifting technique in extremal set theory. In C. Whitehead, editor, *Surveys in Combinatorics*, pages 81–110. Cambridge University Press, 1987.

[7] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics*. Addison-Wesley, 1994.

[8] D. Haussler. Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension. *Journal of Combinatorial Theory, Series A*, 69:217–232, 1995.

[9] D. Haussler and E. Welzl. Epsilon-nets and simplex range queries. *Discrete Computational Geometry*, 2:127–151, 1987.

[10] J. Pach and Pankaj K. Agarwal. *Combinatorial Geometry*. Wiley-Interscience Series, 1995.

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

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

[13] S. Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific Journal of Mathematics*, 41:247–261, 1972.

[14] V. Vapnik. *Statistical Learning Theory*. Wiley, 1998.

[15] V. N. Vapnik and A. Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Apl.*, 16:264–280, 1971.

P.O.B. 11438
Tel Aviv 61116
ISRAEL
E–mail: ratsaby@gmail.com