On the VC-dimension and Boolean functions with long runs*

Joel Ratsaby
Ben Gurion University of the Negev
P.O.B. 653, Be'er Sheva 84105
ISRAEL

April 30, 2006

Abstract

The Vapnik-Chervonenkis (VC) dimension and the Sauer-Shelah lemma have found applications in numerous areas including set theory, combinatorial geometry, graph theory and statistical learning theory. Estimation of the complexity of discrete structures associated with the search space of algorithms often amounts to estimating the cardinality of a simpler class which is effectively induced by some restrictive property of the search. In this paper we study the complexity of Boolean-function classes of finite VC-dimension which satisfy a local 'smoothness' property expressed as having long runs of repeated values. As in Sauer's lemma, a bound is obtained on the cardinality of such classes.

Keywords: Boolean functions, VC-dimension, Poisson approximation

1 Introduction

Let $[n] = \{1, \ldots, n\}$ and denote by $2^{[n]}$ the class of all 2^n functions $h : [n] \to \{0, 1\}$. Let $\mathcal{H}$ be a class of functions and for a set $A = \{x_1, \ldots, x_k\} \subseteq [n]$ denote by $h_{|A} = [h(x_1), \ldots, h(x_k)]$. A class $\mathcal{H}$ is said to *shatter* A if $\left|\{h_{|A} : h \in \mathcal{H}\}\right| = 2^k$. The Vapnik-Chervonenkis dimension of $\mathcal{H}$, denoted as $VC(\mathcal{H})$, is defined as the cardinality of the largest set shattered by $\mathcal{H}$. The following well known result obtained by [18, 19, 21] states an upper bound on the cardinality of classes $\mathcal{H}$ of VC-dimension d.

Lemma 1 (Sauer's Lemma) *For any* $1 \leq d < n$ *let*

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

Then

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

---

*A short version of the paper appeared at the Second International Workshop on Boolean Functions, Cryptography and Applications (BFCA'06)

More generally, the lemma holds for classes of finite VC-dimension on infinite domains. Aside of being an interesting combinatorial result in set theory (see Chapter 17 in [7]), Lemma 1 has been been extended in various directions notably [1, 2, 10, 13] and found applications in numerous fields such as combinatorial geometry [15], graph theory [4, 14], empirical processes [16] and statistical learning theory [8, 20]. In such areas, the complexity of analysis of algorithms on discrete structures, for instance, searching for best approximation of Boolean functions, typically reduces to the complexity of a simpler structure constrained by some 'smoothness' property which is induced by the search.

Consider Boolean functions $h : [n] \to \{0,1\}$. For $x \in [n]$, $y \in \{0,1\}$ define by $\omega_h(x,y)$ the largest $0 \leq a \leq n$ such that $h(z) = y$ for all $x-a \leq z \leq x+a$; if no such a exists then let $\omega_h(x,y) = -1$. We call this the *width* of h at x with respect to y. Denote by $\Xi = [n] \times \{0,1\}$. For a sample $\zeta = \{(x_i, y_i)\}_{i=1}^{\ell} \in \Xi^{\ell}$, define by $\omega_\zeta(h) = \min_{1 \leq i \leq \ell} \omega_h(x_i, y_i)$ the width of h with respect to ζ. For instance, Figure 1 displays a sample $\zeta = \{(x_1,y_1),(x_2,y_2)\}$ and two functions h_1, h_2 which have a width of 3 with respect to ζ.

In [17] we studied classes of Boolean functions that have a large width on a given fixed sample ζ. In this paper we study the complexity of classes of Boolean functions constrained by the width as follows:

$$\mathcal{H}_N(\ell) = \{h \in \mathcal{H} : \exists \zeta \in \Xi^{\ell}, \omega_\zeta(h) > N\}, \quad \ell \geq 1, N \geq 0 \tag{1}$$

where for brevity the dependence of $\mathcal{H}_N$ on $\mathcal{H}$ is left implicit. We obtain a bound (in the form of Lemma 1) for such classes.

The novelty of the paper is both in the results and in the bounding technique. Realizing that Boolean functions on $[n]$ can be represented both as finite binary sequences as well as finite sets in $[n]$ enables to use techniques from probability analysis and set-theory. The remainder of the paper is organized as follows: in the next section we state the main result. Section 3 contains the proof.

2 Main Result

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

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

or for brevity we write $\mathcal{D}$. It is easy to see that the class $\mathcal{D}$ is in one-to-one correspondence with $\mathcal{H}$. It will be convenient to view a function $h : [n] \to \{0,1\}$ as a binary sequence $x^{(n)}$ of n bits $X_1, \ldots, X_n$, where $X_i \in \{0,1\}$, $1 \leq i \leq n$. Denote by a *k-run* any subsequence in $x^{(n)}$ of k consecutive ones or consecutive zeros (the runs may be overlapping). For instance, suppose $k = 3$ then in the sequence $x^{(n)} = 01111100011$ there are four k-runs. Let $\zeta \in \Xi^{\ell}$ then for any $h \in \mathcal{H}$ with $\omega_\zeta(h) > N$, there exist ℓ runs of length $2(N+1)+1$ (of consecutive 0's or consecutive 1's) in the corresponding sequence $x^{(n)}$. This implies that the sequence corresponding to the difference function $\delta_h \in \mathcal{D}$ has at least ℓ runs of consecutive 1's of length $2(N+1)$. Letting

$$\mathcal{D}_N(\ell) \equiv \{\delta \in \mathcal{D} : \exists \, \ell \; 2(N+1)\text{-runs of 1's } \} \tag{3}$$

for $\ell \geq 1, N \geq 0$, then clearly

$$|\mathcal{H}_N(\ell)| \leq |\mathcal{D}_N(\ell)|, \tag{4}$$

where $\mathcal{H}_N(\ell)$ is defined in (1) and is based on the class $\mathcal{H}$ corresponding to $\mathcal{D}$. Our approach will be to bound from above the cardinality of the corresponding class $\mathcal{D}_N(\ell)$. We denote by

$$\mathrm{VC}_\Delta(\mathcal{H}) \equiv \mathrm{VC}(\mathcal{D}),$$

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 can be easily shown that $\mathrm{VC}(\mathcal{D}) \leq c\mathrm{VC}(\mathcal{H})$ for some small positive absolute constant c. Denote by $(n)_k \equiv n(n-1)\cdots(n-k+1)$ with $(n)_k = 0$ if $k > n$. Let $(a)_+ = a$ if $a \geq 0$ and $(a)_+ = 0$ otherwise. The following is the main result of the paper.

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

$$\max_{\mathcal{H} \subset 2^{[n]},\, VC_\Delta(\mathcal{H}) = d} |\mathcal{H}_N(\ell)| \leq \mathfrak{b}_d^{(\ell,N)}(n)$$

where $\mathcal{H}_N$ is dependent on $\mathcal{H}$ by its definition (1),

$$\mathfrak{b}_d^{(\ell,N)}(n) \equiv \sum_{i=0}^{d} \binom{n}{i} \eta(n, 2(N+1), \ell, n-i) \tag{5}$$

and

$$\begin{aligned}\eta(n,k,\ell,r) = & \left(\frac{(r-k+1)_+}{n-k}\right)^{\ell} e^{\lambda(\gamma-1)} \\ + \quad & (n-k+1)\frac{p^{k-1}}{q}\left(\frac{2p^{k-1}}{q}\left(\frac{p}{q}+k+1\right)+1\right) + \frac{(r)_{n/2}}{(n)_{n/2}},\end{aligned} \tag{6}$$

with $p = r/n$, $q = 1-p$, $\lambda = (n-r+1)(r)_k/(n)_k$ and $\gamma = 2(n-r)(n-k)(r-k+1)/((n/2+1)(r-k))$.

To understand this bound, first note that the form of $\mathfrak{b}_d^{(\ell,N)}(n)$ in (5) is similar to $\mathbb{S}(n,d)$ (of Lemma 1) with an additional factor of η. For any fixed value of n and ℓ the function $\mathfrak{b}_d^{(\ell,N)}(n)$ decreases at an exponential rate with respect to the width parameter value N. As an example, Figure 3 displays $\mathfrak{b}_d^{(\ell,N)}(n)$ versus $\mathbb{S}(n,d)$ for various values of N with $d = n^{0.6}$, $\ell = 0.3n$ (on a logarithmic scale). We now proceed with the proof of the theorem.

3 Proof of Theorem 1

For clarity, we split the proof into several subsections. We start by considering a class which is defined as

$$\hat{\mathcal{D}}_N(\ell) \equiv \{\delta : [n] \to \{0,1\} : \#\mathrm{ones}(\delta) \geq n-d, \exists \, \ell \; 2(N+1)\text{-runs of 1's } \} \tag{7}$$

where $1 \leq d, \ell \leq n$ and $N \geq 0$. We have the following result:

Lemma 2 *Let $1 \le d \le n$. Let $\mathcal{D}$ be any class of Boolean functions on $[n]$ with $VC(\mathcal{D}) = d$ and consider $\mathcal{D}_N(\ell) \subset \mathcal{D}$ as defined in (3). Then $|\mathcal{D}_N(\ell)| \le |\hat{\mathcal{D}}_N(\ell)|$.*

Proof: Complement each δ in $\mathcal{D}$ to obtain a new class $\overline{\mathcal{D}}$ where $\mathrm{VC}(\overline{\mathcal{D}}) = \mathrm{VC}(\mathcal{D}) = d$. There is a one-to-one correspondence between elements δ of $\mathcal{D}_N(\ell)$ and elements of the class $\overline{\mathcal{D}}_N(\ell) = \{\delta \in \overline{\mathcal{D}} : \exists\, \ell \;\; 2(N+1)\text{-runs of 0's}\}$ and clearly $\mathrm{VC}(\overline{\mathcal{D}}_N(\ell)) \le d$. So it suffices to show that $|\overline{\mathcal{D}}_N(\ell)| \le |\hat{\mathcal{D}}_N(\ell)|$. Let $\mathcal{F}$ be the set system corresponding to $\overline{\mathcal{D}}_N(\ell)$ which is defined as follows

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

Clearly, $|\mathcal{F}| = |\overline{\mathcal{D}}_N(\ell)|$. Fix a $\delta \in \overline{\mathcal{D}}_N(\ell)$ and consider the complement set $A_\delta^c \equiv [n] \setminus A_\delta$. Since δ, by definition, has at least ℓ $2(N+1)$-runs of 0's then A_δ has the following property P_N: there exist ℓ subsets $E_j \subseteq A_\delta^c$, of consecutive elements $i_j, i_j+1, \ldots, i_j+2N+1 \in [n]$ with $|E_j| = 2(N+1)$, $1 \le j \le \ell$. Hence for every element $A \in \mathcal{F}$, A satisfies P_N and this is denoted by $A \models P_N$. Define $G_{\mathcal{F}}(k) = \max\{|\{A \cap E : A \in \mathcal{F}\}| : E \subseteq [n], |E| = k\}$. The corresponding notion of VC-dimension for a class $\mathcal{F}$ of sets is the the so-called *trace number* [7, p.131] which is defined as $tr(\mathcal{F}) = \max\{m : G_{\mathcal{F}}(m) = 2^m\}$. Clearly, $tr(\mathcal{F}) = VC(\overline{\mathcal{D}}_N(\ell)) \le d$.

The proof proceeds as in the proof of Sauer's lemma [3, Theorem 3.6] which is based on the shifting method [see 7, Ch. 17, Theorem 1 & 4] [see also 10, 11, 12]. The idea is to transform $\mathcal{F}$ into $\mathcal{F}_0$ which is an *ideal* family of sets E, i.e., if $E \in \mathcal{F}_0$ then $S \in \mathcal{F}_0$ for every $S \subset E$, and such that $|\mathcal{F}| = |\mathcal{F}_0| \le |\hat{\mathcal{D}}_N(\ell)|$.

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

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

Consider now

$$\mathcal{F}_0 = T_1(T_2(\cdots T_n(\mathcal{F}) \cdots))$$

and denote the corresponding function class by $\overline{\mathcal{D}}_0$. Clearly, $|\overline{\mathcal{D}}_0| = |\mathcal{F}_0|$.

Now, $|\mathcal{F}_0| = |\mathcal{F}|$ 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 $x \in [n]$, $T_x(\mathcal{F}_0) = \mathcal{F}_0$ since for each $E \in \mathcal{F}_0$ 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 $\mathcal{F}_0$ is an ideal. Furthermore, since for all $A \in \mathcal{F}$, $A \models P_N$, then removing an element x from A which is equivalent to adding it to A^c, still leaves $A \setminus \{x\} \models P_N$. Hence for all $E \in \mathcal{F}_0$ we have $E \models P_N$.

Now, from Lemma 3 [7, p.133] we have $G_{\mathcal{F}_0}(k) \le G_{\mathcal{F}}(k)$, for all $1 \le k \le n$. Hence, since $tr(\mathcal{F}) \le d$ then $tr(\mathcal{F}_0) \le d$ and since $\mathcal{F}_0$ is an ideal then it follows that for all $E \in \mathcal{F}_0$, $|E| \le d$. Combined with the fact that for all $E \in \mathcal{F}_0$, $E \models P_N$ then it follows that the corresponding function class $\overline{\mathcal{D}}_0$ satisfies the following: for all $\delta \in \overline{\mathcal{D}}_0$, δ has at most d 1's and there exist ℓ $2(N+1)$-runs of 0's. It follows that the class $\mathcal{D}_0 = \{1 - \delta : \delta \in \overline{\mathcal{D}}_0\}$, whose cardinality equals that of $\overline{\mathcal{D}}_0$, has every $\delta \in \mathcal{D}_0$ with at least $n - d$ 1's and at least ℓ

$2(N+1)$-runs of 1's. From the above, $|\mathcal{D}_N(\ell)| = |\overline{\mathcal{D}}_N(\ell)| = |\mathcal{F}| = |\mathcal{F}_0| = |\overline{\mathcal{D}}_0| = |\mathcal{D}_0|$ and from (7) we have $|\mathcal{D}_0| \leq |\hat{\mathcal{D}}_N(\ell)|$. This proves the statement of the lemma. □

In order to prove Theorem 1 it suffices to show that $|\hat{\mathcal{D}}_N(\ell)| \leq \mathfrak{b}_d^{(\ell,N)}(n)$. We proceed to obtain a bound on $|\hat{\mathcal{D}}_N(\ell)|$.

3.1 Fixing the number of ones

For a sequence $x^{(n)}$ let $\#\text{runs}_k(x^{(n)})$ denote the number of k-runs of consecutive 1's in $x^{(n)}$. Fix n and d and consider the set of sequences

$$\hat{D}_{k,\ell} = \{x^{(n)} : \#\text{runs}_k(x^{(n)}) \geq \ell, \#\text{ones}(x^{(n)}) \geq n-d\}. \tag{8}$$

We proceed to derive an upper bound on $|\hat{D}_{k,\ell}|$. For any $1 \leq \alpha \leq n-k+1$, denote by

$$W_\alpha = \prod_{i=\alpha}^{\alpha+k-1} X_i.$$

Clearly, W_α equals 1 if and only if there is a k-run of $1's$ starting at X_α. Denote by

$$\hat{D}^{(r)} = \{x^{(n)} : \#\text{ones}(x^{(n)}) = r\} \tag{9}$$

and let $\mathbb{P}$ be a uniform probability law on $\hat{D}^{(r)}$ with

$$\mathbb{P}(x^{(n)}) = \frac{1}{\binom{n}{r}}, \qquad x^{(n)} \in \hat{D}^{(r)}. \tag{10}$$

It is clear that under this law the random variables W_α, $1 \leq \alpha \leq n-k+1$, are dependent. The expected value of W_α is

$$\begin{aligned} \mathbb{E}W_\alpha &= \mathbb{P}(W_\alpha = 1) \\ &= \mathbb{P}\left(X_\alpha = \cdots = X_{\alpha+k-1} = 1\right). \end{aligned} \tag{11}$$

The probability in (11) equals the number of sequences in $\hat{D}^{(r)}$ which have $X_\alpha = \cdots = X_{\alpha+k-1} = 1$, divided by $|\hat{D}^{(r)}|$. There are $\binom{n-k}{r-k}$ such sequences hence the probability is

$$\mathbb{P}\left(X_\alpha = \cdots = X_{\alpha+k-1} = 1\right) = \frac{\binom{n-k}{r-k}}{\binom{n}{r}}, \; k \leq r \tag{12}$$

and the probability is zero otherwise. We have

$$\binom{n-k}{r-k} \Big/ \binom{n}{r} = \frac{(r)_k}{(n)_k} \equiv \pi_k \tag{13}$$

where $(a)_k$ denotes $a(a-1)\cdots(a-(k-1))$.

The sum

$$\#\text{runs}_k(x^{(n)}) = \sum_{\alpha=1}^{n-k+1} W_\alpha$$

may be approximated by a Poisson random variable Z_λ with a mean of $(n-k+1)\pi_k$. The Chen-Stein method [5] may be used to upper bound the approximation error. Unfortunately, for our use, the bound does not decrease fast enough with respect to the run-length k.

3.2 Compound Poisson

A more accurate approximation of $\#\text{runs}_k(x^{(n)})$ is by a compound Poisson random variable [6, Section 3.1]. Let $\mathbb{N}$ denote the positive integers.

Definition 1 *Let M be a Poisson random variable with mean λ. Let X_i, $1 \le i \le M$, be mutually independent random variables defined on $\mathbb{N}$, independent of M and identically distributed according to a probability distribution μ. Then the sum $\sum_{i=1}^{M} X_i$ is distributed according to a* compound Poisson *distribution $CP(\lambda, \mu)$.*

The idea now is to represent $\#\text{runs}_k(x^{(n)})$ as a sum of a random number of clumps where a clump starting at α has a consecutive run of at least k $1's$ followed by a zero, for instance, 000111110101111 has a clump of length 6 starting at the 4^{th} bit.

In order to pick out the start of a clump at α we define

$$Y_\alpha = \begin{cases} (1 - X_{\alpha-1})W_\alpha, & \alpha = 2, \dots, n-k+1, \\ W_\alpha, & \alpha = 1, \end{cases}$$

i.e., Y_α indicates that a run of $1's$ of length at least k starts at α where there is no need to consider $\alpha > n - k + 1$ since such a clump cannot exist there.

Define R as

$$R = \sum_{\alpha=1}^{n-k+1} Y_\alpha.$$

Hence R counts the number of such clumps. Its expected value is

$$\begin{aligned} \mathbb{E}R &= \left((n-k)\binom{n-k-1}{r-k} + \binom{n-k}{r-k} \right) \Big/ \binom{n}{r} \\ &= \pi_k \left(n - r + 1\right). \end{aligned} \tag{14}$$

Since Y_α are (dependent) Bernoulli with small $P(Y_\alpha = 1) \le \pi_k$, with increasing n, if k and r increase at a rate such that $\mathbb{E}R \to \lambda$ then it easy to show using the Stein-Chen method [see for instance 5] that R may be approximated by a Poisson random variable with mean λ.

Next define

$$Y_{\alpha,l} = \begin{cases} (1 - X_{\alpha-1})X_\alpha \cdots X_{\alpha+k+l-2}(1 - X_{\alpha+k+l-1}), & 2 \le \alpha \le n-k+1 \\ X_\alpha \cdots X_{\alpha+k+l-2}(1 - X_{\alpha+k+l-1}), & \alpha = 1. \end{cases} \tag{15}$$

We may now express the number of k-runs as

$$\#\text{runs}_k(x^{(n)}) = \sum_{\alpha=1}^{n-k+1} \sum_{l \ge 1} l Y_{\alpha,l} \tag{16}$$

where the inner sum equals the size of a clump starting at α since every clump has only one unique indicator $Y_{\alpha,l}$ which equals 1 only when l is the size of the clump at α.

3.3 Truncating the sum

We continue now to estimate the cardinality of the set $\hat{D}_{k,\ell}$ defined in (8). Let

$$\hat{D}_{k,\ell}^{(r)} \equiv \{x^{(n)} : \#\text{runs}_k(x^{(n)}) \geq \ell, \#\text{ones}(x^{(n)}) = r\} \tag{17}$$

where $\hat{D}_{k,\ell}^{(r)} = \emptyset$ if $r < k + \ell - 1$. Then

$$|\hat{D}_{k,\ell}| = \sum_{r=n-d}^{n} |\hat{D}_{k,\ell}^{(r)}|.$$

Clearly, by (10), the cardinality of $\hat{D}_{k,\ell}^{(r)}$ can be expressed as

$$|\hat{D}_{k,\ell}^{(r)}| = \binom{n}{r} \mathbb{P}\left(\#\text{runs}_k(x^{(n)}) \geq \ell\right). \tag{18}$$

Let us simplify and limit the range of the clump size detected by the indicators $Y_{\alpha,l}$ to be $1 \leq l \leq n/2 - k - 1$. The sum of (16) thus becomes a restricted sum which we denote by

$$W^* = \sum_{\alpha=1}^{n-k+1} \sum_{l=1}^{n/2-k-1} l Y_{\alpha,l} \tag{19}$$

and, writing the dependence on $x^{(n)}$ explicitly, we have

$$W^*(x^{(n)}) = \#\text{runs}_k(x^{(n)}) - \sum_{\alpha=1}^{n-k+1} \sum_{l=n/2-k}^{n-k} l Y_{\alpha,l}.$$

For two random variables X, Y defined on a discrete space Ω, the total variation distance between the probability distribution of X and Y is defined as

$$\text{dist}(X, Y) = \sup_{A \in \Omega} |\mathbb{P}_X(A) - \mathbb{P}_Y(A)|$$

which for non-negative random variables X, Y with $\Omega = \{0, 1, \ldots\}$ amounts to $\text{dist}(X, Y) = \frac{1}{2} \sum_{j=0}^{\infty} |\mathbb{P}_X(j) - \mathbb{P}_Y(j)|$. Denote by $B = \{x^{(n)} \in \hat{D}^{(r)} : \nexists \text{ clump of size } > n/2 - k - 1\}$. Then, for $x^{(n)}$ randomly distributed according to (10), we have

$$\begin{aligned}
\text{dist}(W^*(x^{(n)}), \#\text{runs}_k(x^{(n)})) &= \sup_{A \subset \mathbb{N}} \left| \mathbb{P}(W^*(x^{(n)}) \in A) - \mathbb{P}(\#\text{runs}_k(x^{(n)}) \in A) \right| \\
= & \sup_{A \subset \mathbb{N}} \Big| \Big(\mathbb{P}(W^*(x^{(n)}) \in A | x^{(n)} \in B) - \mathbb{P}(\#\text{runs}_k(x^{(n)}) \in A | x^{(n)} \in B) \Big) \mathbb{P}(x^{(n)} \in B) \\
+ & \Big(\mathbb{P}(W^*(x^{(n)}) \in A | x^{(n)} \notin B) - \mathbb{P}(\#\text{runs}_k(x^{(n)}) \in A | x^{(n)} \notin B) \Big) \mathbb{P}(x^{(n)} \notin B) \Big| \\
\leq & \sup_{A \subset \mathbb{N}} \left| \mathbb{P}(W^*(x^{(n)}) \in A | x^{(n)} \in B) - \mathbb{P}(\#\text{runs}_k(x^{(n)}) \in A | x^{(n)} \in B) \right| \mathbb{P}(x^{(n)} \in B)
\end{aligned} \tag{20}$$

$$+ \sup_{A \subset \mathbb{N}} \left| \mathbb{P}(W^*(x^{(n)}) \in A | x^{(n)} \notin B) - \mathbb{P}(\#\text{runs}_k(x^{(n)}) \in A | x^{(n)} \notin B) \right| \mathbb{P}(x^{(n)} \notin B). \tag{21}$$

For $x^{(n)} \in B$, $W^*(x^{(n)}) = \#\text{runs}_k(x^{(n)})$ thus (20) equals zero. Now,

$$\mathbb{P}(x^{(n)} \notin B) = \mathbb{P}(x^{(n)} \notin B) \leq \frac{\binom{n-n/2}{r-n/2}}{\binom{n}{r}} = \pi_{n/2}$$

where π_k is defined in (13) and a clump of size greater than $n/2 - k - 1$ has at least $n/2$ ones. So (21) is bounded from above by $\pi_{n/2}$ and we have

$$\text{dist}(W^*(x^{(n)}), \#\text{runs}_k(x^{(n)})) \leq \pi_{n/2}. \tag{22}$$

We may therefore continue and bound (18) from above as

$$|\hat{D}^{(r)}_{k,\ell}| \leq \binom{n}{r} \left(\mathbb{P}\left(W^*(x^{(n)}) \geq \ell\right) + \pi_{n/2}\right). \tag{23}$$

3.4 Stein-Chen bound

The following result is based on Stein's method for Poisson process approximation [6, Section 2.1].

Lemma 3 *[6, CPA PP] Let Γ be an index set. Let $I_{\gamma,l}$ be an indicator of a clump of l events which occurs at $\gamma \in \Gamma$, $l \geq 1$. Let $B(\gamma, l) \subset \Gamma \times \mathbb{N}$ be a set containing $\{\gamma\} \times \mathbb{N}$ and let*

$$b_1 = \sum_{(\gamma,l)\in\Gamma\times\mathbb{N}} \sum_{(\beta,j)\in B(\gamma,l)} \mathbb{E}I_{\gamma,l}\mathbb{E}I_{\beta,j}$$

$$b_2 = \sum_{(\gamma,l)\in\Gamma\times\mathbb{N}} \sum_{\substack{(\beta,j)\in B(\gamma,l)\\(\beta,j)\neq(\gamma,l)}} \mathbb{E}\left(I_{\gamma,l}I_{\beta,j}\right) \tag{24}$$

and

$$b_3 = \sum_{(\gamma,l)\in\Gamma\times\mathbb{N}} \mathbb{E}\left|\mathbb{E}\left(I_{\gamma,l} - \mathbb{E}\left(I_{\gamma,l}\middle|\sigma\left(I_{\beta,j}; (\beta,j) \notin B(\gamma,l)\right)\right)\right)\right| \tag{25}$$

where $\sigma\left(I_{\beta,j}; (\beta,j) \notin B(\gamma,l)\right)$ denotes the σ-field of events generated by the random variables $I_{\beta,j}$ outside $B(\gamma,l)$. Let

$$W = \sum_{\gamma\in\Gamma}\sum_{l\geq 1} lI_{\gamma,l}$$

and let

$$M = \sum_{\gamma\in\Gamma}\sum_{l\geq 1} I_{\gamma,l}$$

be the total number of clumps. Let $\lambda \equiv \mathbb{E}M$ and define the probability distribution μ on $\mathbb{N}$ as

$$\mu(l) \equiv \lambda^{-1}\sum_{\gamma\in\Gamma}\mathbb{E}I_{\gamma,l},$$

$l \geq 1$. Then

$$dist(W, Z_{\lambda,\mu}) \leq b_1 + b_2 + b_3$$

where $Z_{\lambda,\mu}$ is a Compound Poisson random variable distributed as $CP(\lambda,\mu)$.

We now use this lemma by letting

$$\Gamma = \{1, \ldots, n-k+1\},$$

considering the variables $Y_{\alpha,l}$ as the indicators $I_{\gamma,l}$, the total number of clumps R as M and W^* as W. Thus from (14) we have

$$\lambda = \mathbb{E}R = \pi_k(n-r+1). \tag{26}$$

For $1 \le l \le r-k+1$ we have

$$\begin{aligned}
\sum_{\alpha} \mathbb{E}Y_{\alpha,l} &= \mathbb{E}Y_{1,l} + \sum_{\alpha=2}^{n-k+1} \mathbb{E}Y_{\alpha,l} \\
&= \frac{\binom{n-k-l}{r-k-(l-1)}}{\binom{n}{r}} + (n-k)\frac{\binom{n-k-l-1}{r-k-(l-1)}}{\binom{n}{r}} \\
&= \left(\frac{r_{k+l}}{n_{k+l}}\right)\left(\frac{n-r}{r-k-l+1}\right) + \frac{r_{k+l+1}}{n_{k+l+1}}\frac{(n-k)(n-r)(n-r-1)}{(r-k-l+1)(r-k-l)}
\end{aligned} \tag{27}$$

and hence according to the lemma

$$\begin{aligned}
\mu(l) &= \frac{1}{\pi_k(n-r+1)} \sum_{\alpha=1}^{n-k+1} \mathbb{E}Y_{\alpha,l} \\
&= \frac{(n)_k}{(r)_k(n-r+1)}\left(\frac{(r)_{k+l}}{(n)_{k+l}}\frac{n-r}{r-k-l+1} + \frac{(r)_{k+l+1}}{(n)_{k+l+1}}\frac{(n-k)(n-r)(n-r-1)}{(r-k-l+1)(r-k-l)}\right) \\
&= \left(\frac{(r-k)_{l-1}}{(n-k)_{l-1}}\right)\frac{n-r}{(n-r+1)(n-k-(l-1))}\left(1 + \frac{(n-r-1)(n-k)}{n-(k+l)}\right).
\end{aligned} \tag{28}$$

3.5 Approximation error

By its definition (19), the sum W^* may be approximated as a compound Poisson random variable. Applying Lemma 3 we obtain

$$\mathbb{P}(W^*(x^{(n)}) \ge \ell) \le \mathbb{P}(Z_{\lambda,\mu} \ge \ell) + \epsilon(n,k,r) \tag{29}$$

where $Z_{\lambda,\mu}$ is a compound Poisson random variable with λ and μ as in (26) and (28), respectively, and $\epsilon(n,k,r) = b_1 + b_2 + b_3$ as in Lemma 3. Let us now explicitly express $\epsilon(n,k,r)$. Let

$$L = \{1, 2, \ldots, n/2 - k - 1\} \tag{30}$$

and

$$B(\gamma, l) = \{(\beta, j) : j \in L, \gamma - k - j \le \beta \le \gamma + k + l\}.$$

We have

$$b_1 = \sum_{\gamma\in\Gamma, l\in L} \sum_{(\beta,j)\in B(\gamma,l)} \mathbb{E}Y_{\gamma,l}\mathbb{E}Y_{\beta,j}. \tag{31}$$

We have from (15),

$$\mathbb{E}Y_{\alpha,l} = \begin{cases} \frac{\binom{n-k-l}{r-k-(l-1)}}{\binom{n}{r}}, & \alpha = 1 \\ \\ \frac{\binom{n-k-l-1}{r-k-(l-1)}}{\binom{n}{r}}, & 2 \le \alpha \le n-k+1, \end{cases} \tag{32}$$

thus we may use as a bound on

$$\mathbb{E}Y_{\alpha,l} \le \mathbb{E}Y_{1,l}, \tag{33}$$

$1 \le \alpha \le n-k+1$.

Simplifying, we obtain

$$\begin{aligned} \mathbb{E}Y_{1,l} &= \frac{\binom{n-k-l}{r-k-(l-1)}}{\binom{n}{r}} \\ &= \frac{r_{k+l-1}}{n_{k-l-1}} \left(\frac{n-r}{n-k-(l-1)} \right) \\ &\le \frac{r_{k+l-1}}{n_{k-l-1}} \\ &= \pi_{k+l-1} \\ &\le p^{k+l-1} \end{aligned} \tag{34}$$

where we used $k + (l-1) \le r$, π_k is defined in (13),

$$p \equiv \frac{r}{n}$$

and we will denote by

$$q \equiv 1 - p.$$

We thus have

$$\begin{aligned} \sum_{(\beta,j)\in B(\gamma,l)} \mathbb{E}Y_{\beta,j} &= \sum_{j\in L} \sum_{\beta=\gamma-k-j}^{\gamma+k+l} \mathbb{E}Y_{\beta,j} \\ &\le p^{k-1} \sum_{j\in L} \sum_{\beta=\gamma-k-j}^{\gamma+k+l} p^j \\ &= p^{k-1} \sum_{j\in L} (2k+j+l+1)p^j \\ &\le p^{k-1} \left(\sum_{j\ge 0} jp^j + (2k+l+1) \sum_{j\ge 0} p^j \right) \\ &= \frac{p^{k-1}}{q} \left(\frac{p}{q} + 2k + l + 1 \right). \end{aligned}$$

Continuing from (31), we have

$$\begin{aligned} b_1 = \sum_{\gamma\in\Gamma, l\in L} \mathbb{E}Y_{\gamma,l} \sum_{(\beta,j)\in B(\gamma,l)} \mathbb{E}Y_{\beta,j} &\leq \sum_{\gamma\in\Gamma,l\in L} \mathbb{E}Y_{\gamma,l} \frac{p^{k-1}}{q}\left(\frac{p}{q}+2k+l+1\right) \\ &\leq (n-k+1)\frac{p^{2(k-1)}}{q}\sum_{l\geq 0}\left(\frac{p}{q}+2k+l+1\right)p^l \\ &\leq (n-k+1)\frac{p^{2(k-1)}}{q}\left(\left(\frac{p}{q}+2k+1\right)\sum_{l\geq 0}p^l+\sum_{l\geq 0}lp^l\right) \\ &= 2(n-k+1)\frac{p^{2(k-1)}}{q^2}\left(\frac{p}{q}+k+\frac{1}{2}\right). \end{aligned} \tag{35}$$

Next, we bound b_2. From (24) we have

$$b_2 = \sum_{(\gamma,l)\in\Gamma\times\mathbb{N}} \sum_{\substack{(\beta,j)\in B(\gamma,l)\\(\beta,j)\neq(\gamma,l)}} \mathbb{E}\left(Y_{\gamma,l}Y_{\beta,j}\right) \tag{36}$$

and considering the allowable range for (β, j) then only if β is at the leftmost point of $B(\gamma, l)$ then $\mathbb{E}\left(Y_{\gamma,l}Y_{\beta,j}\right) \neq 0$. Hence

$$\sum_{\substack{(\beta,j)\in B(\gamma,l)\\(\beta,j)\neq(\gamma,l)}} \mathbb{E}\left(Y_{\gamma,l}Y_{\beta,j}\right) = \sum_{j\in L}\mathbb{P}\left(Y_{\gamma,l}=1, Y_{\gamma-k-j,j}=1\right). \tag{37}$$

Denote by $s = 2k + j + l$. We have

$$\begin{aligned} \mathbb{P}\left(Y_{\gamma,l}=1, Y_{\gamma-k-j,j}=1\right) &\leq \frac{\binom{n-s}{r-(s-2)}}{\binom{n}{r}} \\ &= \frac{r_{s-2}}{n_{s-2}}\frac{(n-r)(n-(r+1))}{(n-(s-2))(n-(s-1))} \\ &\leq \frac{r_{s-2}}{n_{s-2}} \\ &= \pi_{s-2} \\ &\leq p^{s-2} \end{aligned}$$

since $s - 2 \leq r$. Hence the sum in (37) is bounded from above by

$$p^{2(k-1)+l}\sum_{j\geq 0}p^j = \frac{p^{2(k-1)+l}}{q}$$

and therefore from (36)

$$\begin{aligned} b_2 &\leq (n-k+1)\frac{p^{2(k-1)}}{q}\sum_{l\geq 0}p^l \\ &= (n-k+1)\frac{p^{2(k-1)}}{q^2}. \end{aligned} \tag{38}$$

Next, we bound b_3. Consider the inner expectation in (25) which may be written as

$$\mathbb{E}Y_{\gamma,l} - \mathbb{E}\left(Y_{\gamma,l}|\sigma_{\gamma,l}\right)$$

where $\sigma_{\gamma,l} = \sigma\left(Y_{\beta,j}; (\beta, j) \notin B(\gamma, l)\right)$ and we replaced $Y_{\gamma,l}$ instead of $I_{\gamma,l}$ in the lemma.

It is clear in this application of the lemma that there exists a dependence of $Y_{\gamma,l}$ on the set of random variables $\{Y_{\beta,j} : (\beta, j) \notin B(\gamma, l)\}$. For instance, let (β', j') be some point outside $B(\gamma, l)$ and consider the event A that $Y_{\beta',j'} = 1$. We have $\mathbb{P}(Y_{\gamma,l} = 1)$ which from (32) and (33) is no larger than $\binom{n-k-l}{r-k-(l-1)}/\binom{n}{r}$ while $\mathbb{P}(Y_{\gamma,l} = 1|A)$ is smaller since the event A uses up $(k + j - 1)$ $1's$ thereby leaving fewer sequences $x^{(n)}$ that have $Y_{\gamma,l}(x^{(n)}) = 1$. In the worst case, all available r $1's$ could be depleted by some event $A \in \sigma_{\gamma,l}$ in which case $\mathbb{P}(Y_{\gamma,l} = 1|A) = 0$. Thus we may bound b_3 as

$$\begin{aligned}
b_3 &= \sum_{(\gamma,l)\in\Gamma\times\mathbb{N}} \mathbb{E}\left|\mathbb{E}\left(Y_{\gamma,l} - \mathbb{E}(Y_{\gamma,l}|\sigma_{\gamma,l})\right)\right| \\
&\leq \sum_{(\gamma,l)\in\Gamma\times\mathbb{N}} \mathbb{E}\left|\mathbb{E}\left(Y_{\gamma,l} - 0\right)\right| \\
&= \sum_{(\gamma,l)\in\Gamma\times\mathbb{N}} \mathbb{E}Y_{\gamma,l} \\
&\leq (n-k+1)p^{k-1}\sum_{l\geq 0} p^l \\
&\leq (n-k+1)\frac{p^{k-1}}{q} \qquad (39)
\end{aligned}$$

where we used (34). Combining (35), (38) and (39) then (29) becomes

$$\begin{aligned}
&\mathbb{P}(W^*(x^{(n)}) \geq \ell) \leq \mathbb{P}(Z_{\lambda,\mu} \geq \ell) + \epsilon(n, k, r) \\
&\leq \mathbb{P}(Z_{\lambda,\mu} \geq \ell) + (n-k+1)\frac{p^{k-1}}{q}\left(\frac{p^{k-1}}{q}\left(2\left(\frac{p}{q} + k + \frac{1}{2}\right) + 1\right) + 1\right)
\end{aligned} \qquad (40)$$

where $p = (1 - q) = r/n$. Next, we upper bound the probability $\mathbb{P}(Z_{\lambda,\mu} \geq \ell)$.

3.6 Tail probability

We have the following bound on the tail probability of a compound Poisson random variable:

Lemma 4 *Let λ be as defined in (26), $m > 0$. Let M be a Poisson random variable with mean λ. Let Y_i, $1 \leq i \leq M$, be i.i.d. random variables taking positive integer values with a probability distribution μ (defined in (28)). Then the tail probability of their sum is*

$$\mathbb{P}\left(\sum_{i=1}^{M} Y_i \geq m\right) \leq \left(\frac{r-k+1}{n-k}\right)^m e^{\lambda(\gamma-1)}$$

where $\gamma = 2(n - r)(n - k)(r - k + 1)/((n/2 + 1)(r - k))$.

Proof: We have

$$\mathbb{P}\left(\sum_{i=1}^{M} Y_i \geq m\right) = \sum_{s=1}^{\infty} \mathbb{P}\left(\sum_{i=1}^{s} Y_i \geq m \middle| M = s\right) \mathbb{P}(M = s) \tag{41}$$

(note that if $M = 0$ then there is a zero probability that the sum is no less than m for $m > 0$). We now obtain an upper bound on the tail probability of

$$\mathbb{P}\left(\sum_{i=1}^{s} Y_i \geq m \middle| M = s\right), \quad s \geq 1$$

based on Cheronoff's method [9]. From Markov's inequality, for any $t > 0$, we have

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{s} Y_i \geq m \middle| M = s\right) &\leq e^{-mt}\mathbb{E}e^{t\sum_{i=1}^{s} Y_i} \\ &= e^{-mt}\prod_{i=1}^{s}\mathbb{E}e^{tY_i} \\ &= e^{-mt}\left(\mathbb{E}e^{tY_1}\right)^s \end{aligned}$$

where the last step follows from the Y_i, $1 \leq i \leq M$, being i.i.d.. We have

$$\mathbb{E}e^{tY_1} = \sum_{l \geq 1} \mu(l)e^{tl}. \tag{42}$$

Dividing both numerator and denominator of (28) by $(n - k)$ we obtain

$$\begin{aligned} &\left(\frac{(r-k)_{l-1}}{(n-k)_{l-1}}\right)\left(\frac{n-r-l/(n-k)}{(n-k-l)(1-(l-1)/(n-k))}\right)\left(\frac{n-r}{n-r+1}\right) \\ &\leq \left(\frac{(r-k)_{l-1}}{(n-k)_{l-1}}\right)\left(\frac{n-r-l/(n-k)}{(n-k-l)(1-(l-1)/(n-k))}\right). \end{aligned} \tag{43}$$

From (30) we have $l \leq n/2 - k - 1$ so the denominator of the second factor above is bounded from below by

$$(n - k - (n/2 - k - 1))\left(1 - \frac{n/2 - k - 1}{n - k}\right) \geq (n/2 + 1)(1 - 1/2)$$

hence we have

$$\mu(l) \leq \left(\frac{(r-k)_{l-1}}{(n-k)_{l-1}}\right)\frac{2(n-r)}{n/2+1}, \quad 1 \leq l \leq r - k + 1.$$

Let us denote the rational

$$\alpha \equiv \frac{r-k}{n-k}$$

and with

$$\alpha_l \equiv \frac{(r-k)_{l-1}}{(n-k)_{l-1}}$$

we therefore have

$$\mu(l) \leq \alpha_l \frac{2(n-r)}{n/2+1}.$$

We may therefore bound the expectation in (42) as

$$\begin{aligned}\mathbb{E}e^{tY_1} &= \sum_{1\leq l\leq r-k+1} \mu(l)e^{tl} \\ &\leq \frac{2(n-r)}{n/2+1} \sum_{1\leq l\leq r-k+1} \alpha_l e^{tl}\end{aligned}$$

and we have the simple bound

$$\alpha_l \leq \alpha^{l-1}$$

so therefore

$$\mathbb{E}e^{tl} \leq \frac{2(n-r)}{(n/2+1)\alpha} \sum_{1\leq l\leq r-k+1} \alpha^l e^{tl} \leq \frac{2(n-r)}{(n/2+1)\alpha} \frac{1}{(1-\alpha e^t)}$$

provided that $\alpha e^t < 1$. Choosing $t = \ln((n-k)/(r-k+1))$ we obtain

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{s} Y_i \geq m \middle| M = s\right) &\leq e^{-mt}\left(\frac{2(n-r)}{(n/2+1)\alpha}\right)^s \left(\frac{1}{(1-\alpha e^t)}\right)^s \\ &= \left(\frac{r-k+1}{n-k}\right)^m \left(\frac{2(n-r)(n-k)(r-k+1)}{(n/2+1)(r-k)}\right)^s \\ &= \left(\frac{r-k+1}{n-k}\right)^m \gamma^s.\end{aligned}$$

Now, since M is Poisson distributed with mean λ (denoted by $\mathbb{P}_\lambda(s)$) then the right side of (41) is bounded from above by

$$\begin{aligned}\left(\frac{r-k+1}{n-k}\right)^m \sum_{s=1}^{\infty} \mathbb{P}_\lambda(s)\gamma^s &\leq \left(\frac{r-k+1}{n-k}\right)^m e^{\lambda(\gamma-1)} \sum_{s=1}^{\infty} \frac{e^{-\lambda\gamma}(\lambda\gamma)^s}{s!} \\ &\leq \left(\frac{r-k+1}{n-k}\right)^m e^{\lambda(\gamma-1)}.\end{aligned}$$

□

By Lemma 4 it follows that the tail probability for $Z_{\lambda,\mu}$ in (29) satisfies

$$\mathbb{P}(Z_{\lambda,\mu} \geq \ell) \leq \left(\frac{r-k+1}{n-k}\right)^\ell e^{\lambda(\gamma-1)} \tag{44}$$

with γ and λ as defined in Lemma 4.

3.7 Combining

From (23), (40) and (44) it follows that as a bound on $|\hat{D}_{k,\ell}^{(r)}|$ (defined in (17)) we have

$$|\hat{D}_{k,\ell}^{(r)}| \leq \binom{n}{r} \eta(n,k,\ell,r) \tag{45}$$

where $\eta(n,k,\ell,r)$ is defined in (6). Hence the set $\hat{D}_{k,\ell}$ defined in (8) has cardinality

$$\begin{aligned} |\hat{D}_{k,\ell}| &\leq \sum_{r=n-d}^{n} \binom{n}{r} \eta(n,k,\ell,r) \\ &= \sum_{i=0}^{d} \binom{n}{i} \eta(n,k,\ell,n-i). \end{aligned} \tag{46}$$

The set $\hat{D}_{k,\ell}$ (defined in (8)) with $k = 2(N+1)$, is equivalent to the class $\hat{\mathcal{D}}_N(\ell)$ defined in (7). Thus

$$|\hat{\mathcal{D}}_N(\ell)| \leq \sum_{i=0}^{d} \binom{n}{i} \eta(n, 2(N+1), \ell, n-i) \equiv \mathfrak{b}_d^{(\ell,N)}(n). \tag{47}$$

Together with (4) and Lemma 2 it follows that for any $\mathcal{H}$ with $\mathrm{VC}_{\Delta}(\mathcal{H}) = d$, the corresponding class (see (1)) satisfies

$$|\mathcal{H}_N(\ell)| \leq \mathfrak{b}_d^{(\ell,N)}(n)$$

which completes the proof of Theorem 1.

4 Conclusion

The width of a Boolean function at x is defined as the degree to which it is smooth, i.e., constant around x. The paper extends the classical Sauer's lemma to classes of Boolean functions which are wide around a sample. An upper bound on the cardinality of any such class is obtained by counting binary sequences with long-runs using the Stein-Chen method of approximation. The result indicates that the cardinality decreases at an exponential rate with respect to the width parameter. The novelty of the paper is both in the results and in the bounding technique where Boolean functions on $[n]$ are represented both as finite binary sequences and as finite sets in $[n]$. This enables the use of techniques from probability analysis and set-theory.

Bibliography

[1] N. Alon. On the density of sets of vectors. *Discrete Math.*, 46:199–202, 1983.

[2] N. Alon, S. Ben-David, N. Cesa-Bianchi, and D. Haussler. Scale-sensitive dimensions, uniform convergence, and learnability. *J. ACM*, 44(4):616–631, 1997.

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

[4] M. Anthony, G. Brightwell, and C. Cooper. The Vapnik-Chervonenkis dimension of a random graph. pages 616–631, 1995.

[5] R. Arratia, L. Goldstein, and L. Gordon. Poisson approximation and the chen-stein method. *Statistical Science*, 5:403–434, 1990.

[6] A. D. Barbour and O. Chryssaphinou. Compound poisson approximation: A user's guide. *The Annals of Applied Probability*, 11(3):964–1002, 2001.

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

[8] S. Boucheron, O. Bousquet, and G. Lugosi. *Introduction to Statistical Learning Theory, In , O. Bousquet, U.v. Luxburg, and G. Rsch (Editors)*, pages 169–207. Springer, 2004.

[9] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.*, 23:493–507, 1952.

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

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

[12] 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.

[13] D. Haussler and P.M. Long. A generalization of Sauer's lemma. *Journal of Combinatorial Theory (A)*, 71(2):219–240, 1995.

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

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

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

[17] J. Ratsaby. A constrained version of Sauer's lemma. In *Algorithms, Trees Combinatorics and Probabilities*, volume III of *Mathematics and Computer Science*, pages 543–551. Birkhäuser Verlag, 2004.

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

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

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

[21] 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.

Appendix

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: $\omega_\zeta(h_1) = \omega_\zeta(h_2) = 3$

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

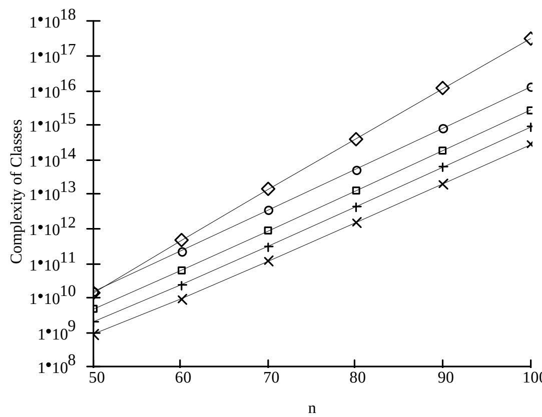

Figure 3: $\mathfrak{b}_d^{(\ell,N)}(n)$ for $N = 0.39n, 0.36n, 0.33n, 0.29n$, [x, $+, \square, \circ$ traces] v.s. $\mathbb{S}(n,d)$, [$\diamond$ trace]