Topics around Vapnik-Chevronenkis theory

Chapter 1

The Sauer-Shelah Lemma

1 Two equivalent frameworks

Let $\mathcal{U}$ and $\mathcal{V}$ be two sets and let $\varphi \subseteq \mathcal{U} \times \mathcal{V}$. In other words φ is a **bipartite graph**. We may call φ an (abstract) **incidence relation** and write $\varphi(x, y)$ for $\langle x, y \rangle \in \varphi$. Sets of the form

$$\boldsymbol{\varphi(\mathcal{U}, b)} \quad = \quad \{a \in \mathcal{U} : \varphi(a, b)\}.$$

are called **definable sets** or, when more that a relation is involved, **φ-definable sets**. The collection **collection definable subsets** is denoted by $\boldsymbol{\varphi(\mathcal{U}, b)_{b \in \mathcal{V}}}$.

Often we need to consider the **trace** of a definable set on some arbitrary $A \subseteq \mathcal{U}$.

$$\boldsymbol{\varphi(A, b)} \quad = \quad \{a \in A : \varphi(a, b)\}.$$

We call this a **definable subset of A**. And denote by $\boldsymbol{\varphi(A, b)_{b \in \mathcal{V}}}$ the collection of definable subsets on A.

If all subsets of A are definable, that is $\mathcal{P}A = \varphi(A, b)_{b \in \mathcal{V}}$, we say that A is **shattered** by φ. The following is called the **shatter function**

$$\boldsymbol{\pi_\varphi(n)} \quad = \quad \max \left\{ |\varphi(A, b)_{b \in \mathcal{V}}| \quad : \quad A \in \binom{\mathcal{U}}{n} \right\}.$$

So, $\pi_\varphi(n)$ gives the maximal number of definable subsets that a set of cardinality n may have. Trivially, $\pi_\varphi(n) \leq 2^n$ for all n. Moreover, if $\pi_\varphi(k) = 2^k$ for some k, then $\pi_\varphi(n) = 2^n$ for every $n \leq k$.

Given $A \subseteq \mathcal{U}$ define the equivalence relation $\equiv_{\varphi, A}$ defined by

$$\begin{aligned} \boldsymbol{b \equiv_{\varphi, A} b'} \quad &\Leftrightarrow \quad \varphi(a, b) \leftrightarrow \varphi(a, b') \quad \text{for ll } a \in A; \\ &\Leftrightarrow \quad \varphi(A, b) = \varphi(A, b'). \end{aligned}$$

We denote by $S_\varphi(A)$ the set of $\equiv_{\varphi, A}$-equivalence classes. We obviously have

$$\pi_\varphi(n) \quad = \quad \max \left\{ |S_\varphi(A)| \quad : \quad A \in \binom{\mathcal{U}}{n} \right\}.$$

The dual incidence relation φ^* is the relation on $\mathcal{V} \times \mathcal{U}$ which is sometimes denoted by φ^{-1}. Then dual scattering function is defined as follows (with the obvious meaning of the notation)

$$\boldsymbol{\pi^*_\varphi(n)} \quad = \quad \max \left\{ |\varphi(a, B)_{a \in \mathcal{U}}| \quad : \quad B \in \binom{\mathcal{V}}{n} \right\}.$$

1.1 Definition *The* ***Vapnik-Chervonenkis dimension*** *of φ, abbreviated by* ***VC-dimension****, is the maximal cardinality of a finite set $A \subseteq \mathcal{U}$ that is shattered by φ. Equivalently, it is the maximal k such that $\pi_\varphi(k) = 2^k$. If such a maximum does not exist, we say that φ has infinite VC-dimension.*

We will say ***dual VC-dimension*** *for the VC-dimension of φ^*.*

The **VC-density** *of φ is the infimum over all real number r such that $\pi_\varphi(n) \in O(n^r)$. It is infinite if no such r exist. The* **dual VC-density** *is defined accordingly.* □

Then if the VC-density is finite so is the VC-dimension. The converse is also true. In fact, in the next section we show that the VC-dimension bounds the VC-density.

1.2 Proposition *If φ has VC-dimension $< k$ then its dual VC-dimension is $< 2^{k+1}$.*

Proof Assume that the VC-dimension of φ^* is $\geq 2^k$. We prove that the VC-dimension of φ^* is $\geq k$. Let $B = \{b_I : I \subseteq k\}$ be a set of cardinality 2^k shattered by φ^*. That is, for every $\mathcal{J} \subseteq \mathcal{P}(k)$ there is $a_\mathcal{J}$ such that

$$\varphi(a_\mathcal{J}, b_I) \quad \Leftrightarrow \quad I \in \mathcal{J}$$

Let $a_i = a_{\{I \subseteq k : i \in I\}}$. Then from the equivalence above we obtain

$$\varphi(a_i, b_I) \quad \Leftrightarrow \quad i \in I$$

That is, φ shatters $\{a_i : i < k\}$. □

An alternative formalism uses **set systems** in place of incidence relations. A set system is a collection Φ of subsets of some set $\mathcal{U}$. We denote a set system by $(\mathcal{U}, \Phi)$. Given a set system, we immediately obtain an incidence relation with $\mathcal{V} = \mathcal{P}(\mathcal{U})$ by defining $\varphi(x, y)$ as $x \in y \in \Phi$.

Vice versa, to an incidence relation φ we associate $\Phi = \varphi(\mathcal{U}, b)_{b \in \mathcal{V}}$, the set system of the definable subsets of $\mathcal{U}$. The **VC-dimension of Φ** is defined to be that of the associated incidence relation. The shatter function $\boldsymbol{\pi_\Phi(n)}$ is defined in a similar way.

An incidence relation is **extensional** if $\varphi(\mathcal{U}, b') = \varphi(\mathcal{U}, b'')$ implies $b', b'' \in \mathcal{V}$. The correspondence between extensional incidence relations and set systems is one-to-one. In general, the correspondence is many-to-one. However, as we are mainly interested the set system associated, we my switch from one formalism to the other according to which one is more convenient.

1.3 Example a. Set systems of cardinality 1 are those with an incidence relation of the form $A \times \mathcal{V}$ for some $A \subseteq \mathcal{U}$. They shatter only the empty set, therefore they have VC-dimension 0. Their shatter function is identically 1.

b. Let Φ be a non trivial partition of $\mathcal{U}$. Then only singletons are shattered, so the VC-dimension is 1. The shatter function is $\pi_\Phi(n) = \min\{n, |\Phi|\}$.

c. If Φ is a non trivial chain of subsets of $\mathcal{U}$ the situation is identical to that described in b.

d. Let $\mathcal{U} = \mathbb{R}$ and let Φ be the collection of open intervals. Any set of 2 points is shattered but no set with 3 points can. So the VC-dimension is 2.

e. Let $\mathcal{U} = \mathbb{R}^2$ and let Φ be the collection of half planes. Any set of 3 non collinear points is shattered but no set with 4 points can (by Radon's Theorem). So the VC-dimension is 3.

f. Let $\Phi = U^{[\leq k]}$ be the collection of all subsets of $\mathcal{U}$ of cardinality $\leq k$. Then Φ has VC-dimension k and

$$\pi_\Phi(n) = \sum_{i=0}^{k} \binom{n}{i}.$$

g. Let $\mathcal{U} = \mathbb{R}^2$ and let Φ be the collection of polygons. Then Φ has VC-dimension ∞. □

2 The Sauer-Shelah Lemma

According to Gil Kalai in [3], Sauer-Shelah's Lemma can been described as an *eigen-theorem* because it is important in many different areas of mathematic (model theory, learning theory, probability theory, ergodic theory, Banach spaces, to name a few). No wonder it has been discovered and rediscovered may times.

It has been proved independently by Shelah [6], Sauer [5], and Vapnik-Chervonenkis [7] around 1970 (Shelah gives credit to Micha Perles). Saharon Shelah was working in model theory while Norbert Sauer, Vladimir Vapnik and Alexey Chervonenkis were in statistical learning theory.

We shall present three proofs of this lemma, one in this section and two in the next section. I am aware of a forth proof which uses linear algebra, see e.g. [2].

1.4 Proposition (Sauer-Shelah's Lemma) *If φ has VC-dimension k then for every $n \geq k$*

$$\pi_\varphi(n) \quad \leq \quad \sum_{i=0}^{k} \binom{n}{i}.$$

The set system presented in f of Example 1.3 shows that the bound is optimal.

Proof If $k = 0$, both sides of the inequality are 1. Now, assume the lemma is true for $k-1$. We prove by induction on n that for every A of cardinality n

$$\big|\varphi(A,b)_{b\in\mathcal{V}}\big| \quad \leq \quad \sum_{i=0}^{k} \binom{n}{i}.$$

If $n = k$ the r.h.s. of the inequality above is 2^n and the claim is trivial. So, assume the claim is true for $n-1$ and let A have cardinality n. Fix some $a \in A$ and let $A' = A \smallsetminus \{a\}$. We can assume that $\varphi(A,b) \triangle \varphi(A,b') = \{a\}$ for some b, b', otherwise $\big|\varphi(A,b)_{b\in\mathcal{V}}\big| = \big|\varphi(A',b)_{b\in\mathcal{V}}\big|$ and the claim follows immediately from the induction hypothesis.

Define a new incidence relation

$$\psi(x,y) \quad = \quad x \in A' \;\wedge\; \varphi(x,y) \;\wedge\; \exists y' \Big[\varphi(A,y) \triangle \varphi(A,y') = \{a\}\Big].$$

Note that if A'' is shattered by ψ then $A'' \cup \{a\}$ it is shattered by φ. Then the VC-dimension of ψ is at most $k-1$. We also have that

$$\big|\varphi(A,b)_{b\in\mathcal{V}}\big| \quad = \quad \big|\varphi(A',b)_{b\in\mathcal{V}}\big| \;+\; \big|\psi(A',b)_{b\in\mathcal{V}}\big|.$$

Hence by induction hypothesis

$$\begin{aligned}
\big|\varphi(A,b)_{b\in\mathcal{V}}\big| \quad &\leq \quad \sum_{i=0}^{k} \binom{n-1}{i} \;+\; \sum_{i=0}^{k-1} \binom{n-1}{i} \\
&= \quad \binom{n-1}{0} \;+\; \sum_{i=1}^{k} \left[\binom{n-1}{i} \;+\; \binom{n-1}{i-1}\right] \\
&= \quad \binom{n-1}{0} \;+\; \sum_{i=1}^{k} \binom{n}{i}
\end{aligned}$$

$$\leq \sum_{i=0}^{k} \binom{n}{i}$$

which completes the proof of the proposition. □

Next corollary states an important dichotomy. It says that the shatter function grows exponentially unless the VC-dimension is finite. In this case the growth is only polynomial. Therefore the VC-dimension is an upper bond to the VC-density.

1.5 Corollary *For every incidence relation φ one of the following obtains*

1. *the VC-dimension is infinite and $\pi_\varphi(n) = 2^n$ for every positive integer n;*
2. *the VC-dimension is k and $\pi_\varphi(n) \in O(n^k)$.*

Proof If VC-dimension is infinite claim 1 is obvious. So assume φ has VC-dimension is k. Then

$$\pi_\varphi(n) \leq \sum_{i=0}^{k} \binom{n}{i} \leq \sum_{i=0}^{k} \frac{n^i}{i!} \leq e\, n^k$$ □

3 Pajor variant and the method of shifting

An alternative proof of the Sauer-Shelah's Lemma derives it as corollary of a lemma by Alain Pajor [4]. The proof below is credited to Noga Alon by Kalai [3], to (many) others by [1].

1.6 Proposition (Pajor's Lemma) *Assume Φ is finite. Then Φ there are at least $|\Phi|$ sets shattered by Φ.*

We show how Sauer-Shelah's Lemma follows from Pajor's Lemma. Fix a set $A \subseteq \mathcal{U}$ of cardinality n such that $\pi_\varphi(n) = \big|\varphi(A,b)_{b\in\mathcal{V}}\big|$. Now, consider the set system $\Phi = \varphi(A,b)_{b\in\mathcal{V}}$. By Pajor's Lemma there are $|\Phi|$ subsets of $\mathcal{U}$ shattered by Φ. These subsets cannot have cardinality larger than the VC-dimension of φ, then $\Phi \subseteq A^{[\leq k]}$

$$\pi_\varphi(n) = |\Phi| \leq \big|A^{[\leq k]}\big| = \sum_{i=0}^{k} \binom{n}{i}.$$

Proof The proposition holds if $|\Phi| = 1$. Now, suppose it holds for set systems of cardinality $< |\Phi|$. Fix $a \in \mathcal{U}$ and let $\Phi_0 = \big\{B \in \Phi : a \notin B\big\}$ and $\Phi_1 = \Phi \smallsetminus \Phi_0$. We can choose a such that $\Phi_i \neq \Phi$. By the inductive hypothesis, both Φ_i shatter at least $\big|\Phi_i\big|$ sets. If no set is shattered by both Φ_i, the claim follows immediately. Otherwise, note that for each set shattered by both Φ_i there are two sets shattered by Φ, one containing a and one not containing a. The claim follows. □

We give a second proof of Pajor's Lemma by a method which is interesting in itself because of its many applications. I would not know whom to credit with this proof. The method has been introduced by Erdős, Ko and Rado to prove their eponymous theorem. They named it *compression*, but is now known as *shifting*.

When $A' \subseteq A$ and $A \smallsetminus A' = \{a\}$, we write $\boldsymbol{A' \subseteq_a A}$. We write

$$\boldsymbol{B_{\varphi,a}} = \big\{b \,:\, \varphi(a,b) \text{ and } \neg\exists b'\, \varphi(\mathcal{U},b') \subseteq_a \varphi(\mathcal{U},b)\big\}$$

We say that φ is **compressed** if $B_{\varphi,a} = \varnothing$ for every $a \in \mathcal{U}$.

1.7 Proposition *If φ is compressed then any co-finite subset of a definable set is definable. In particular φ shatters every finite subset of a definable set.*

Proof If for contradiction $\psi(\mathcal{U}, b) \smallsetminus \{a\}$ is not definable, then $b \in B_{\varphi,a}$. $\square$

1.8 Proposition *Let $\psi = \varphi \smallsetminus \{a\} \times B_{\varphi,a}$. Then*

$$\psi(\mathcal{U}, b) = \psi(\mathcal{U}, b') \quad \Leftrightarrow \quad \varphi(\mathcal{U}, b) = \varphi(\mathcal{U}, b') \qquad \textit{for all } b, b' \in \mathcal{V}.$$

Proof $\Rightarrow$ Assume $\varphi(\mathcal{U}, b) \neq \varphi(\mathcal{U}, b')$. We may also assume that $\varphi(\mathcal{U} \smallsetminus \{a\}, b) = \varphi(\mathcal{U} \smallsetminus \{a\}, b')$, otherwise $\psi(\mathcal{U}, b) \neq \psi(\mathcal{U}, b')$ is immediate. Then $\varphi(a, b) \not\Leftrightarrow \varphi(a, b')$, say $\varphi(a, b)$ and $\neg\varphi(a, b')$. Then $\varphi(\mathcal{U}, b') \subseteq_a \varphi(\mathcal{U}, b)$ and $b \notin B_{\varphi,a}$. So $\psi(a, b)$ and $\neg\psi(a, b')$.

$\Leftarrow$ Assume $\psi(\mathcal{U}, b) \neq \psi(\mathcal{U}, b')$. Again, we may also assume that $\psi(\mathcal{U} \smallsetminus \{a\}, b) = \psi(\mathcal{U} \smallsetminus \{a\}, b')$. Then $\psi(a, b) \not\Leftrightarrow \psi(a, b')$, say $\psi(a, b)$ and $\neg\psi(a, b')$. As $\varphi(a, b)$ is clear, we only have to prove that $\neg\varphi(a, b')$. Suppose for a contradiction that $\varphi(a, b')$. Then $b' \in B_{\varphi,a}$. This is a contradiction as $\psi(\mathcal{U}, b) \subseteq_a \psi(\mathcal{U}, b')$. $\square$

1.9 Proposition *Let $\psi = \varphi \smallsetminus \{a\} \times B_{\varphi,a}$. Then every set shattered by ψ is shattered by φ.*

Proof Assume A is shattered by ψ. We prove that for every b here is a b' such that $\psi(A, b) = \varphi(A, b')$. We may assume that $a \in A$. If $\psi(a, b)$ then we may chose $b = b'$. So, assume $\neg\psi(a, b)$. As ψ shatters A then there is c such that $\psi(A, c) = \psi(A, b) \cup \{a\}$. Then $\exists b' \, \varphi(\mathcal{U}, b') \subseteq_a \varphi(\mathcal{U}, c)$. Therefore $\psi(A, b) = \varphi(A, b')$. $\square$

1.10 Second proof of Pajor's Lemma Let φ be the incidence relation associated to Φ. We may assume that φ is finite. Let $\varphi_0 = \varphi$ and $\varphi_{i+1} = \varphi_i \smallsetminus \{a\} \times B_{\varphi_i,a}$ for some a. As φ is finite, we can assume that at some stage n we obtain a compressed relation $\psi = \varphi_n$. Let Ψ be the set of ψ-definable sets. By Proposition 1.9, Ψ shatters at least $|\Psi|$ sets. By Proposition 1.7, every set shattered by Ψ is shattered by Φ. By Proposition 1.8, $|\Psi| = |\Phi|$. $\square$

Define $E_{\varphi,a} \subseteq \mathcal{V}^2$ as the set of e pairs $\langle b', b \rangle$ such that $\varphi(\mathcal{U}, b') \subseteq_a \varphi(\mathcal{U}, b)$. Note incidentally that $E_{\varphi,a}$ is the graph of a partial injections of $\mathcal{V}$ into itself.

1.11 Proposition *Let $\psi = \varphi \smallsetminus \{a\} \times B_{\varphi,a}$. Then $E_\varphi \subseteq E_\psi$*

Proof Both the range and the domain of $E_{\varphi,a}$ are disjoint from $B_{\varphi,a}$. $\square$

We call $E_\varphi = \bigcup\limits_{a \in U} E_{\varphi,a}$ the unit distance diagraph of φ.

1.12 Proposition *Let φ be a finite incidence relation with* VC*-dimension k. Let Φ be the set system associated to it. Then $|E_\varphi| \leq k\,|\Phi|$.*

Proof Let ψ and Ψ be as defined in Proof 1.10. By Propositions 1.7 and 1.9, every has at most k ancestors. Therefore

$$\frac{|E_\varphi|}{|\Phi|} \quad \leq \quad \frac{|E_\psi|}{|\Psi|} \quad \leq \quad k$$ $\square$

4 Notes and references

[1] R. P. Anstee, Lajos Rónyai, and Attila Sali, *Shattering news*, Graphs Combin. **18** (2002), no. 1, 59–73.

[2] Timothy Gowers, *Dimension arguments in combinatorics*, Gowers's Weblog (2008).

[3] Gil Kalai, *Extremal Combinatorics III: Some Basic Theorems*, Combinatorics and more (2008).

[4] Alain Pajor, *Sous-espaces l_1^n des espaces de Banach*, Travaux en Cours [Works in Progress], vol. 16, 1985.

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

[6] Saharon Shelah, *A combinatorial problem; stability and order for models and theories in infinitary languages*, Pacific J. Math. **41** (1972), 247–261.

[7] V. N. Vapnik and A. Ya. Chervonenkis, *On the uniform convergence of relative frequencies of events to their probabilities*, Measures of complexity, Springer, Cham, 2015, pp. 11–30. Reprint of Theor. Probability Appl. **16** (1971), 264–280.

Chapter 2

Samples

1 Multisets

A sample is a sequence of elements of $\mathcal{U}$ where we disregard the order and only consider the number of times they appear. We formalize this with the notion of multiset.

A **multiset** is a function $A : \mathcal{U} \to \omega$. We interpret $A(a)$ as the multiplicity of the element a. The **support** of A is the set $\{a : A(a) \neq 0\}$. The **size of A** is defined as

$$|A| \quad = \quad \sum_{a \in \mathcal{U}} A(a).$$

The set of multisets of size n is denoted by $\mathcal{U}^{[n]}$.

If we identify plain sets with $\{0,1\}$-valued multisets, size generalizes cardinality. A tuple $\langle a_1, \ldots, a_n \rangle \in \mathcal{U}^n$ is an **enumeration** of A if $A(a) = \big|\{i : a_i = a\}\big|$ for all $a \in \mathcal{U}$. Clearly, any enumeration of A has length $n = |A|$.

We say that C is a submultiset of A if $C(a) \leq A(a)$ for all $a \in \mathcal{U}$. We write $C \subseteq A$. We write $A \cap C$ for the multiset $\min\{A(a), C(a)\}$. We write $A \smallsetminus C$ for the multiset $\max\{0, A(a) - C(a)\}$. Note that $|A| = |A \cap C| + |A \smallsetminus C|$.

When A is a multi-set we read $\varphi(A, b)$ as the intersection of A with $\varphi(\mathcal{U}, b)$.

Note that the definitions above (but for the notion of enumeration) generalize to **fractional multisets** i.e., function that take non negative real values. These will be used in the next chapter.

2 A uniform law of large numbers

Given a multiset A of finite size we define the **frequency** of $\varphi(\mathcal{U}, b)$

$$\operatorname{Fr}\varphi(A,b) \quad = \quad \frac{\big|\varphi(A,b)\big|}{|A|}$$

Recall the weak law of large number.

2.1 Proposition (Weak law of large numbers) *Let* Pr *be a probability measure on $\mathcal{U}$ that makes all definable sets measurable. Then for every $b \in \mathcal{V}$*

$$\Pr\Big(A \in \mathcal{U}^{[n]} \ :\ \Big|\Pr\varphi(\mathcal{U},b) - \operatorname{Fr}\varphi(A,b)\Big| \leq \varepsilon\Big) \quad \geq \quad 1 - \frac{1}{4n\varepsilon^2}$$

Proof Let $p = \mu\varphi(\mathcal{U}, b)$ then $\big|\varphi(A,b)\big|$ is a binomial random variable $B(p,n)$. Hence $(1/n)\big|\varphi(A,b)\big|$ has mean p and variance $p(1-p)/n \leq 1/4n$. Apply Chebyshev's inequality. □

An **ε-sample**, or **ε-approximation**, of a probability measure Pr is a multiset $A \subseteq \mathcal{U}$ such that for every $b \in \mathcal{V}$

$$\left|\Pr \varphi(\mathcal{U}, b) - \mathrm{Fr}\, \varphi(A, b)\right| \quad \leq \quad \varepsilon$$

We state the uniform law of large number for finite sample spaces. The generalization requires some technical nuisance, so it is moved to the end of the section.

2.2 Proposition (Uniform law of large numbers) *Let* Pr *be a probability measure on* $\mathcal{U}$*, which we assume to be finite set. Let* φ *have* VC*-density* d*. Then there is a multi-set* ε*-approximation of size*

$$n \quad \leq \quad C \frac{d}{\varepsilon^2} \ln \frac{1}{\varepsilon},$$

where C *is an absolute constant (in particular, it does not depend neither on* $\mathcal{U}$ *nor on* Pr*).*

3 Discrepancy

Given ε we are interested in the least n such that some ε-approximations of size n exist. The idea is to start with an approximation of large size and reduce size at the cost of slightly enlarging ε. We now introduce a powerful technique to achieve this.

It is convenient to include $\mathcal{U}$ among the definable sets. So, let $u \notin \mathcal{V}$ be a fresh element. We extend φ on $\mathcal{V} \cup \{u\}$ by defining $\boldsymbol{\varphi(\mathcal{U}, u) = \mathcal{U}}$.

A **coloring** of a multiset A is just a submultiset $C \subseteq A$. For $b \in \mathcal{V} \cup \{u\}$ we write

$$\delta_{A,C,b} \quad = \quad \frac{|\varphi(C, b)| - |\varphi(A \smallsetminus C, b)|}{|A|}.$$

The **(relative) discrepancy** of C is

$$\delta_{A,C} \quad = \quad \sup_{b \in \mathcal{V} \cup \{u\}} |\delta_{A,C,b}|$$

The **(relative) discrepancy** of A is

$$\delta_A \quad = \quad \inf_{C \subseteq A} \delta_{A,C}$$

It is immediate that if A is a set shattered by φ, then δ_A is large, i.e. close to $1/2$.

The next lemma is intuitive, if an ε-approximation has small discrepancy then we can halve its size at a small cost.

2.3 Lemma *Fix a probability measure* Pr*. Let* A *be an* ε*-approximation of size* n*. Let* $C \subseteq A$ *be a coloring of discrepancy* $\delta_{A,C}$*. Then either* C *or* $A \smallsetminus C$ *is an* $(\varepsilon + 2\delta_{A,C})$*-approximation of size* $\leq n/2$*.*

Proof Define also $n^+ = |C|$ and $n^- = |A \smallsetminus C|$. We may assume that $n^+ \leq n/2$, otherwise swap C and $A \smallsetminus C$. Then $\delta_{A,C,u} = (n^+ - n^-)/n < 0$.

1.
$$\begin{aligned} \frac{|\varphi(A,b)|}{n} &= \frac{|\varphi(C,b)| + |\varphi(A \smallsetminus C, b)|}{n} \\ &= \frac{2|\varphi(C,b)|}{n} - \delta_{A,C,b} \\ &\leq \frac{|\varphi(C,b)|}{n^+} + \delta_{A,C} \end{aligned}$$

We also have

2.
$$\frac{|\varphi(A,b)|}{n} = \frac{2|\varphi(C,b)|}{n} - \delta_{A,C,b}$$

$$
\begin{aligned}
&= \frac{|\varphi(C,b)|}{n^+}\big(1+\delta_{A,C,u}\big) \;-\; \delta_{A,C,b}\\
&\geq \frac{|\varphi(C,b)|}{n^+} \;-\; 2\delta_{A,C}
\end{aligned}
$$

Combining 1 and 2 we obtain

$$
\left|\frac{|\varphi(A,b)|}{n} \;-\; \frac{|\varphi(C,b)|}{n^+}\right| \;\leq\; 2\delta_{A,C}
$$

hence

$$
\left|\Pr\varphi(A,b) \;-\; \frac{|\varphi(C,b)|}{n^+}\right| \;\leq\; \varepsilon + 2\delta_{A,C}
$$

as claimed by the lemma. □

2.4 Corollary *Let A be an ε-approximation of size n and discrepancy δ_A. Then there is an $(\varepsilon + 2\delta_A)$-approximation of size $\leq n/2$.*

The lemma above tells that ε-approximations with small discrepancy are useful, but as yet we have no clue as to finding one. We are going to prove that when the number of definable subsets of A is relatively small, then the discrepancy of A is not too large. We use a probabilistic argument to prove this bound (when you don't have a clue how to do something, you might as well do it randomly).

First, we make a brief digression into probability theory. The following inequality is a classical tool in this context.

2.5 Lemma (Chernoff's bound, special case) *For $i = 1, \ldots, n$ let X_i be independent identically distributed random variables such that $\Pr(X_i = \pm 1) = 1/2$. Then for every $\delta > 0$*

$$
\Pr\left(\bar{X} \geq \delta\right) \;\leq\; \exp(-\frac{n}{2}\delta^2) \qquad\qquad \textit{where } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i
$$

Proof Let $t > 0$ be arbitrary. Then

$$
\sharp \qquad
\begin{aligned}
\Pr(\bar{X} \geq \delta) &= \Pr\left(e^{t\bar{X}} \geq e^{t\delta}\right)\\
&\leq e^{-t\delta}\,\mathrm{E}\big(e^{t\bar{X}}\big)
\end{aligned}
$$

In fact, the equality follows because the exponential is an increasing function and the inequality is Markov's inequality, which says that $\Pr(X \geq a) \leq a^{-1}\mathrm{E}(X)$ for every a and is immediate to verify. Now observe that

$$
\begin{aligned}
\mathrm{E}\big(e^{tX_i}\big) &= \frac{1}{2}e^{t} \;+\; \frac{1}{2}e^{-t}\\
&= \frac{1}{2}\sum_{i=0}^{\infty}\frac{t^i}{i!} \;+\; \frac{1}{2}\sum_{i=0}^{\infty}\frac{(-t)^i}{i!}\\
&= \sum_{i=0}^{\infty}\frac{t^{2i}}{(2i)!}\\
&\leq \sum_{i=0}^{\infty}\frac{(t^2/2)^i}{i!}\\
&= e^{t^2/2}
\end{aligned}
$$

From this, by independence we have

$$\mathrm{E}\big(e^{t\bar{X}}\big) \;=\; \prod_{i=1}^{n} e^{(t/n)X_i} \;=\; e^{t^2/2}$$

Subtituting in $\sharp$ gives $\Pr(\bar{X} \geq \delta) \leq e^{t^2/2 - t\delta}$. Finally Chernoff's inequality is obtained substituting δ for t. $\square$

2.6 Lemma *Let A be a multi-set of size $\leq n$. Assume the support of A has $\leq m$ definable subsets. Then $\delta_A \;\leq\; \sqrt{(2/n)\ln(2m)}$.*

Proof To prove that $\delta_A \leq \delta$ it suffices to show that there is a coloring $C \subseteq A$ such that $\delta_{A,C,b} \leq \delta$ for all b. It suffices to define a probability on C and show that

$$\Pr\big(\{C \subseteq A \;:\; \forall b \;\; \delta_{A,C,b} \leq \delta\}\big) \;>\; 0$$

or, by 3, that for every $b \in \mathcal{V}$

4. $$\Pr\big(\{C \subseteq A \;:\; \delta_{A,C,b} \geq \delta\}\big) \;\leq\; \frac{1}{2m}.$$

Let $\langle a_1, \dots, a_n\rangle$ be an enumeration of A. Imagine that the colorings are obtained by tossing n time a fair coin a toss to decide the color of each a_i independently. Denote by C_σ the coloring associated to the sequence of coin toss σ. Fix $b \in \mathcal{V}$ and let X_i be the random variables

$$X_i \;=\; \begin{cases} +1 & \text{if } a_i \in \varphi(C_\sigma, b) \\ -1 & \text{otherwise} \end{cases}$$

Then $\bar{X} = \delta_{A,C_\sigma,b}$ and we may apply Chernoff's bound to obtain

$$\begin{aligned} \Pr\big(\{C \subseteq A \;:\; \delta_{A,C,b} \geq \delta\}\big) \;&=\; \Pr(\bar{X} \geq \delta) \\ &\leq\; \exp\Big(-\frac{n\delta^2}{2}\Big), \end{aligned}$$

Then 4 is satisfied if $n\delta^2/2 \leq \ln(2m)$. This yields the required bound. $\square$

2.7 Theorem *Let φ have* VC*-density d. Let* $\Pr$ *be some probability measure that admits an η-approximation. Then there is a $(\varepsilon + \eta)$-approximation of size*

$\natural$ $$n \;\leq\; C\frac{k}{\varepsilon^2}\ln\frac{1}{\varepsilon}.$$

where C is an absolute constant. In particular, the bound above does not depend on the size of the η-approximation.

Note that the bound claimed by the theorem only depends on the dimension of φ and it is independent of Pr.

Proof For $i = 0, \dots, h$ we construct a decreasing chain A_i of η_i-approximations. By assumption we can require A_0 is an η_0-approximation for $\eta_0 = \eta$. We denote by n_i and δ_i the cardinality, respectively the discrepancy, of A_i. By lemma 2.3, we can require that $\eta_{i+1} \leq \eta_i + 2\delta_i$ and $n_i = 2^{-i} n_0$. We can assume $n_0 = 2^m$ for some m. Then

$$\eta_h \;=\; \eta \;+\; 2\sum_{i=1}^{h} \delta_i$$

The construction stops at the least h such that

$$\sum_{i=1}^{h} \delta_i \;\leq\; \sum_{i=1}^{h} \sqrt{(2/n_i)\ln(2n_i^d)}$$

$\varepsilon < \varepsilon_h + 2\delta_h$. So we have "only" have to prove that (independently of n_0) this h satisfies

$$\natural\natural \qquad 2^{-h} n_0 \;\leq\; 2^8 \frac{k}{\varepsilon^2} \ln \frac{1}{\varepsilon^2}.$$

We may rewrite the condition $\varepsilon < \varepsilon_h + 2\delta_h$ as

$$\varepsilon \;<\; 4 \sum_{i=1}^{h+1} \delta_i$$

To get rid of some annoying square roots that will appear soon, we substitute the latter inequality with

$$\varepsilon^2 \;<\; 2^5 \sum_{i=1}^{h+1} \delta_i^2$$

We do not know δ_i but Lemma 2.6, together with Sauer's Lemma **??**, gives an upper bound

$$\begin{aligned}
\delta_i^2 &\leq \frac{2}{n_i} \ln n_i^k \\
&= \frac{2^{i+1}}{n_0}\Big(k \ln 2^{-i} + k \ln n_0\Big) \\
&\leq \frac{2^{i+1}}{n_0} k \ln n_0
\end{aligned}$$

Then we obtain

$$\sum_{i=1}^{h+1} \delta_i^2 \;\leq\; 2^{h+2} k \, \frac{\ln n_0}{n_0}$$

The theorem is complete as any h that satisfies the inequality

$$\varepsilon^2 \;<\; 2^{h+7} k \, \frac{\ln n_0}{n_0}$$

verifies $\natural\natural$ (and in particular the h at which the constructions stops). The verification of is immediate, it suffices to substitute for ε^2 the r.h.s. of the inequality above and recall that we can assume that n_0 is sufficiently large. □

Then the uniform measure on $\mathfrak{U}$ admits a 0-approximation, namely $\mathfrak{U}$ itself. Then it admits ε-approximations of size at most $\natural$ for every $0 < \varepsilon < 1$.

(Even a 0-approximation, if Pr is rational valued.) □

2.8 Corollary *Let Φ be a finite set-system of* VC*-dimension $1 < k < \omega$. Then for every positive $\varepsilon < 1$, every probability measure* Pr *admits a multi-set ε-approximation of cardinality bounded by $\natural$ of Theorem 2.7.*

Proof Without loss of generality we can assume that Pr is rational valued. Then there is a uniform probability measure $\Pr'$ on some finite set $\mathfrak{U}'$ and a surjection $f : \mathfrak{U}' \to \mathfrak{U}$ such that $\Pr(f^{-1}\varphi) = \Pr(\varphi)$. By Theorem 2.7 $\Pr'$ admits an ε-approximation B' with cardinality by $\natural$, for every positive $\varepsilon < 1$. We know define a multi-set B such that $B(a) = |B' \cap f^{-1}|$ for every $a \in \mathfrak{U}$. As $|B'| = |B|$ and $|B' \cap f^{-1}\varphi| = |B \cap \varphi|$, this is the required multi-set ε-approximation. □

The following corollary is used in Theorem 3.1 below. Though it is sufficient for our application, stronger bound are known (essentially, we can replace ε for ε^2).

2.9 Corollary *Let Φ be a finite set-system of* VC*-dimension $1 < k < \omega$. Then for every positive $\varepsilon < 1$, every probability measure* Pr *admits a multi-set ε-net of cardinality bounded by $\natural$ of Theorem 2.7.*

2.10 Exercise Suppose that $\{a\}, \{b\} \in \Phi$ for some $a, b \in \mathcal{U}$. Show that, if Pr and ε are such that $0 < \varepsilon < \Pr(a)$ and $2\varepsilon < \Pr(b) - \Pr(a)$, then Pr admits no ε-approximation. $\square$

Chapter 3

A piercing problem

After Pierre Simon, after Jiri Matousek, after Noga Alon and Daniel Kleitman.

Let $q \leq p < \omega$ we say that Φ has the (p, q)-property if out of every p sets in Φ some q have non empty intersection. It has the dual (p, q)-property if for every p point in $\mathcal{U}$ some q are belong to the same set in Φ.

The following a is a particular case of a theorem of Matousek (based on a proof of Noga Alon and Daniel Kleitman). The proof by Pierre Simon is simpler. [I will add details, references, and much more.]

3.1 Theorem *Let $\mathcal{U}$ be finite and let Φ have* VC*-dimension $k < \omega$. There are some integers q and h that depends only on k such that, if every $B \subseteq \mathcal{U}$ of cardinality q is a subset of some $\varphi \in \Phi$, then $\mathcal{U}$ is covered by some $X \subseteq \Phi$ of cardinality h.*

This theorem is often stated in the dual form which explains why it is sometimes called a piercing (or Helly-type) theorem: there are some integers q and h that depends only on k such that, if every q sets in Φ have non-empty intersection, then there is a $B \subseteq \mathcal{U}$ of cardinality h that intersects every $\varphi \in \Phi$. See Exercise 3.3.

Recall that the dual system is $\mathcal{U}^* = \Phi$ and $\Phi^* = \{\Phi_i : i < m\}$, where $\Phi_i = \{\varphi \in \Phi : a_i \in \varphi\}$. By Proposition **??** the dual system has VC-dimension 2^k. Choose some $\varepsilon = 1/2$. From Theorem 2.7 we obtain an h such that every probability measure μ^* on Φ has an ε-approximation of cardinality h. As ε is fixed, this h only depends on k. Then there is a muti-set $X : \Phi \to \omega$ of cardinality h such that Moreover $|Y| \leq h$ as required. Suppose we can find some $x_j \in \mathbb{R}$ such that for every $i < m$ In fact, We can also rescale solutions so, setting $\mu^*(\varphi_j) \overset{\text{def}}{=} x_j$, we obtain a well-defined probability measure It suffices to show that clause 2 of Farkas' Lemma cannot obtain, that is, for every $\lambda_i \in \mathbb{R}_+$ there are some $x_j \in \mathbb{R}_+$ such that Note that for every j we have By assumption, there is a $\check{\jmath}$ such that $\{a : B(a) \neq 0\} \subseteq \varphi_{\check{\jmath}}$ hence $|B \cap \varphi_{\check{\jmath}}| = q$. Therefore $\mu(\varphi_{\check{\jmath}}) > 1 - \varepsilon$. Let $\check{x}_j$ be 1 for $j = \check{\jmath}$ and 0 otherwise. We claim that the tuple $\langle x_j : j < n \rangle$ is a solution of $\flat\flat$. In fact

Proof Let $\langle a_i : i < m \rangle$ and $\langle \varphi_j : j < n \rangle$ enumerate $\mathcal{U}$ and Φ without repetitions. Recall that the dual system is $\mathcal{U}^* = \Phi$ and $\Phi^* = \{\Phi_i : i < m\}$, where $\Phi_i = \{\varphi \in \Phi : a_i \in \varphi\}$. By Proposition **??** the dual system has VC-dimension 2^k. Choose some $\varepsilon = 1/2$. From Corollary 2.9 we obtain an h such that every probability measure μ^* on Φ admits an ε-net of cardinality h. As ε is fixed, this h only depends on k. Restating this explicitly there is a set $X \subseteq \mathcal{U}^*$ of cardinality h such that

$$\sharp \qquad |\mu^*(\Phi_i)| > \varepsilon \quad \Rightarrow \quad X \cap \Phi_i \neq \varnothing$$

So, if the probability measure on Φ is such that $\mu^*(\Phi_i) > \varepsilon$ for all $i < m$, then $X \subseteq \Phi$ cover $\mathcal{U}$ as required by the theorem. So we only need to find such a probability measure.

For $i < m$ and $j < n$ let $P_{i,j} = 1$ if $a_i \in \varphi_j$ and 0 otherwise. Suppose we can find some $x_j \in \mathbb{R}$ such that for every $i < m$

$\flat \qquad \sum_{j<n} (P_{i,j} - \varepsilon) x_j \quad > \quad 0$

Then we can assume that all x_j are positive. In fact,

$$\sum_{j<n} P_{i,j} \quad \geq \quad 1$$

(because every a_i belongs to some φ_j) so we can translate solutions of any positive quantity. We can also rescale solutions so, setting $\mu^*(\varphi_j) \stackrel{\text{def}}{=} x_j$, we obtain a well-defined probability measure

$$\mu^*(\Phi_i) \quad = \quad \sum_{j<n} P_{i,j}\, x_j \quad > \quad \varepsilon$$

We only have to prove that equation $\flat$ has a solution. It suffices to show that clause 2 of Farkas' Lemma cannot obtain, that is, for every $\lambda_i \in \mathbb{R}_+$ there are some $x_j \in \mathbb{R}_+$ such that

$\flat\flat \qquad \sum_{i<m} \lambda_i \sum_{j<n} (P_{i,j} - \varepsilon) x_j \quad > \quad 0.$

As we can assume the λ_i add to 1, setting $\mu(a_i) \stackrel{\text{def}}{=} \lambda_i$, we obtain a probability measure on $\mathcal{U}$. Note that for every j we have

$$\mu(\varphi_j) \quad = \quad \sum_{i<m} \lambda_i P_{i,j}.$$

Apply Corollary 2.9 once again to the set-system $\langle \mathcal{U}, \neg\Phi \rangle$. There is a q such that for every probability measure μ on $\mathcal{U}$ there is a set $B \subseteq \mathcal{U}$ of cardinality q such that

$\sharp\sharp \qquad \mu(\neg\varphi_j) > \varepsilon \quad \Rightarrow \quad B \not\subseteq \varphi_j$

As ε is fixed, q only depends on k. By assumption, there is a $\check{\jmath}$ such that $B \subseteq \varphi_{\check{\jmath}}$ and therefore $\mu(\varphi_{\check{\jmath}}) \geq 1 - \varepsilon$. Let $\check{x}_j$ be 1 for $j = \check{\jmath}$ and 0 otherwise. We claim that the tuple $\langle x_j : j < n \rangle$ is a solution of $\flat\flat$. In fact

$$\sum_{i<m} \lambda_i \sum_{j<n} (P_{i,j} - \varepsilon) \check{x}_j \quad = \quad \sum_{i<m} \lambda_i (P_{i,\check{\jmath}} - \varepsilon) \quad = \quad \mu(\varphi_{\check{\jmath}}) - \varepsilon \quad \geq \quad 1 - \varepsilon \quad > \quad 0.$$

This concludes the proof. □

The following is a classical result. The proof may be found in any introductory text of convex analysis or linear programming. There are many (sometimes non equivalent) ways to state it, we recommend Terence Tao's exposition which may be found in his mathematical blog (we have reversed strict and weak inequalities, but the proof is identical).

3.2 Proposition (Farkas' Lemma) *For $i < m$ let $P_i : \mathbb{R}^n \to \mathbb{R}$ be affine linear functions. Then the following are equivalent*

1. $\bigwedge_{i<m} P_i(x) > 0$ *for some* $x \in \mathbb{R}^n$;
2. *there are no* $\lambda_i \in \mathbb{R}_+$ *such that* $\sum_{i<m} \lambda_i P_i(x) \leq 0$ *for every* $x \in \mathbb{R}^n$. □

It will help induction Up to rescaling we can assume that $P_{i,j}(x, y)$ have the form

3.3 Exercise Let k, q and h be some integers as in Theorem 3.1 and let Φ be a set-system with dual VC-dimension k. Suppose that any q sets in Φ have non-empty intersection and prove that there is some set $B \subseteq \mathcal{U}$ of cardinality h that intersects every $\varphi \in \Phi$. □

1 Appendix: Farkas' lemma

Below, an affine version of Farkas' Lemma tailored to our purposes.

3.4 Proposition *Fix some $v_1, \ldots, v_n, u \in \mathbb{Q}^k$ and let $r_1, \ldots, r_n, s \in \mathbb{Q}$. Let*

$$X(r_1, \ldots, r_n) = \{x \in \mathbb{Q}^k \ : \ r_i \leq v_i \cdot x, \text{ for every } i = 1, \ldots, n\}.$$

Then the following are equivalent

1. *$s \leq u \cdot x$ for all $x \in X(r_1, \ldots, r_n)$;*
2. *there exist $q_1, \ldots, q_n \in \mathbb{Q}^+$ such that $\sum_{i=1}^{n} q_i v_i = u$.*

Proof Implication 2⇒1 is immediate. To prove 1⇒2 assume 1. We claim that

3. $0 \leq u \cdot x$ for all $x \in X(0, \ldots, 0)$.

Suppose not and let $x \in X(0, \ldots, 0)$ be such that $u \cdot x < 0$. Let $y \in X(r_1, \ldots, r_n)$. Then $y + ax \in X(r_1 \ldots, r_n)$ for every $a \in \mathbb{Q}^+$. From 1 we obtain $s \leq u \cdot (y + ax)$ which, for a is sufficiently large, is a contradiction.

From 3 it follows that if $x \cdot v_i = 0$ for $i = 1, \ldots, n$ then $0 = u \cdot x$. We can assume without loss of generality that $v_1, \ldots, v_n$ are linearly independent. Then by linear algebra $\sum q_i v_i = u$ for some $q_i \in \mathbb{Q}$. Now fix i and verify that q_i is non-negative. Let x_i such that $x_i \cdot v_j = 1$, if $i = j$, and 0 otherwise. Then $0 \leq x_i \cdot u = q_i$. □