Topics around Vapnik-Chevronenkis theory

This is a subset of the content of the first chapters of Artem Chernicov's Topics in combinatorics which is also the main source of these notes.

These notes follow closely my expositions in the class (some proofs are different from Chernikov's). However, my notes are less complete and in general less reliable. So, students are encouraged to study from both.

Chapter 1

The Sauer-Shelah Lemma

1 Two equivalent frameworks

Let $\mathcal{U}$ and $\mathcal{V}$ be two sets and let $\varphi \subseteq \mathcal{U} \times \mathcal{V}$ be a binary relation. 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 than one relation is involved, **φ-definable sets**. The collection of all definable sets 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.

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

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

The $\equiv_{\varphi, A}$-equivalence class are called **types over A**. We denote by $S_\varphi(A)$ the set types over A. As there is a one-to-one correspondence between definable subsets and types, $|S_\varphi(A)| = |\varphi(A, b)_{b \in \mathcal{V}}|$.

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**

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

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 $n \leq k$.

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)

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

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. Fix any $\mathcal{V}$ containing Φ and define $\varphi(x, y)$ 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 $\mathcal{U}$. 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.

The **VC-dimension of Φ** is defined to be that of the associated incidence relation. We repeat this explicitly in order to set the notation. Define

$$\boldsymbol{\Phi{\restriction}A} \quad = \quad \{A \cap \mathcal{B} \ : \ \mathcal{B} \in \Phi\}.$$

Then A is shattered by Φ if $\mathcal{P}A = \Phi{\restriction}A$. The shatter function $\boldsymbol{\pi_\Phi(n)}$ is

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

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 infinite 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

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

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 $A' = A \setminus \{a\}$. We can assume that $\varphi(A,b) \triangle \varphi(A,b') = \{a\}$ for some b, b', otherwise $\left|\varphi(A,b)_{b\in\mathcal{V}}\right| = \left|\varphi(A',b)_{b\in\mathcal{V}}\right|$ 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

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

Hence by induction hypothesis

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

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].

1.6 Proposition (Pajor's Lemma) *Let $A \subseteq \mathcal{U}$ be finite. Then φ shatters at least $|S_\varphi(A)|$ subsets of A.*

We show how Sauer-Shelah's Lemma follows from Pajor's Lemma. Fix a set $A \subseteq \mathcal{U}$ of cardinality n such that $\pi_\varphi(n) = \left|S_\varphi(A)\right|$. By Pajor's Lemma there are $|S_\varphi(A)|$ subsets of A shattered by φ. These subsets cannot have cardinality larger than the VC-dimension of φ, then

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

Proof If A is empty then $|S_\varphi(A)| = 1$ and $\varnothing$ is the only subset of A that φ shatters. Fix $a \in A$ and assume the lemma holds for $A' = A \smallsetminus \{a\}$. Let ψ be the relation defined in the proof of Prosition 1.4. Recall that

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

and that if $A'' \subseteq A'$ is shattered by ψ then $A'' \cup \{a\}$ it is shattered by φ. By induction hypothesis φ shatters least $\left|\varphi(A',b)_{b\in\mathcal{V}}\right|$ subsets of A' and at least $\left|\psi(A',b)_{b\in\mathcal{V}}\right|$

containing a. The lemma follows.

We present a different wording of essentially the same proof. Consider the set system $\Phi = \varphi(A,b)_{b\in\mathcal{V}}$ and reason by induction of $|\Phi|$. The proposition holds trivially if $|\Phi| = 1$. Now, suppose it holds for set systems of cardinality $< |\Phi|$. Fix $a \in A$ and let $\Phi_0 = \big\{B \in \Phi : a \notin B\big\}$ and $\Phi_1 = \Phi \smallsetminus \Phi_0$. As $|\Phi| > 1$, we can choose a such that $\Phi_i \neq \Phi$. By the inductive hypothesis, both Φ_i shatter at least $\big|\Phi_i\big|$ subsets of A. 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 different proof of Pajor's Lemma by a method which is interesting in itself because of its many applications. The method has been introduced by Erdős, Ko and Rado to prove their eponymous theorem. They named it *compression*, but is also known as *shifting*.

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

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

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 $b \in B_{\varphi,a}$. □

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 $\psi(\mathcal{U},b) \subseteq_a \psi(\mathcal{U},b')$. □

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 $\psi(A,b) = \varphi(A,b')$. □

1.10 Second proof of Pajor's Lemma Let φ be a finite extensional incidence relation associated to Φ. 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 1.7, Ψ shatters at least $|\Psi|$ sets. By Proposition 1.9, every set shattered by Ψ is shattered by Φ. By Proposition 1.8, $|\Psi| = |\Phi|$. □

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 $B_{\varphi,a}$. □

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 extensional incidence relation with* VC*-dimension k. Then* $|E_\varphi| \leq k\,|\mathcal{V}|$.

Proof Let ψ be as defined in Proof 1.10. By Propositions 1.7 and 1.9, every $b \in \mathcal{V}$ has at most k ancestors. Therefore

$$\frac{|E_\varphi|}{|\mathcal{V}|} \quad\leq\quad \frac{|E_\psi|}{|\mathcal{V}|} \quad\leq\quad k$$ □

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 and approximations of measures

1 Samples and subsamples

A sample is a sequence of elements of $\mathcal{U}$ where we disregard the order and only consider the number of times they appear. Formally, a **sample** is a **multi-subset of $\mathcal{U}$**, that is is a function $A : \mathcal{U} \to \mathbb{N}$.

We interpret $A(x)$ as the **multiplicity** of the element $x \in \mathcal{U}$. The **support** of A is the set **supp(A)**$= \{x : A(x) \neq 0\}$. The **size of A** is defined as

$$|A| \;=\; \sum_{x \in \mathcal{U}} A(x).$$

If we identify sets with $\{0,1\}$-valued samples, the size generalizes cardinality.

We say that C is a **subsample** of A, and write $C \subseteq A$, if $C(x) \leq A(x)$ for all $x \in \mathcal{U}$. We also define the **intersection** and the **difference** of two samples. The element $x \in \mathcal{U}$ has multiplicity $A(x) \wedge C(x)$ in $A \cap C$ and multiplicity $A(x) \dot{-} C(x)$ in $A \smallsetminus C$. Note that $|A| = |A \cap C| + |A \smallsetminus C|$.

The definitions above and those in next sections easily generalize to **fractional multi-sets** i.e., function $A : \mathcal{U} \to \mathbb{R}$ with non negative values. These will have some applications in the next chapters.

As an alternative to multi-sets, we could use the following approach. Replace $\mathcal{U}$ with $\mathcal{U}' = \mathcal{U} \times \omega$ and Φ with $\Phi' = \{\mathcal{B} \times \omega \;:\; \mathcal{B} \in \Phi\}$. Then $\mathcal{U}'$ contains infinitely many copies of each element of $\mathcal{U}$. These copies are indistinguishable by sets in Φ'. Then to every finite subset of $A \subseteq \mathcal{U}'$ we can associate a multi-subset of $\mathcal{U}$. (The correspondence is not one-to-one, but this is not relevant for the intended applications.)

If $\mathcal{B} \subseteq \mathcal{U}$ we define the **frequency** of $\mathcal{B}$ over A

$$\mathbf{Fr}\left(\mathcal{B}/A\right) \;=\; \frac{|\mathcal{B} \cap A|}{|A|}$$

Note that $\mathrm{Fr}(\cdot/A)$ is a probability measure on $\mathcal{U}$.

If Pr is a probability measure on $\mathcal{U}$. An **ε-approximation** of Pr is a sample C such that

$$\Big|\mathrm{Pr}(\mathcal{B}) - \mathrm{Fr}(\mathcal{B}/C)\Big| \;\leq\; \varepsilon$$

It is interesting to note that not all probability measures have an approximation. For instance let $\mathcal{U} = \aleph_1$ and let Φ be the σ-algebra of the countable and co-countable sets. Let $\mathrm{Pr}(\mathcal{B}) = 1$ if $\mathcal{B}$ contains an end-segment, $\mathrm{Pr}(\mathcal{B}) = 0$ otherwise. This defines a σ-additive measure which has no ε-approximations for $\varepsilon < 1$.

The **ε-approximation of a sample A** is a subsample $C \subseteq A$ such that for every definable set $\mathcal{B}$

$$\left| \mathrm{Fr}(\mathcal{B}/A) - \mathrm{Fr}(\mathcal{B}/C) \right| \leq \varepsilon$$

In other words, it is an ε-approximation of the probability measure $\mathrm{Fr}(\cdot/A)$.

2 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.

In general $\mathcal{U}$ may not be among the definable sets. As it is often convenient to include it, we write $\mathbf{\Phi'}$ for $\Phi \cup \{\mathcal{U}\}$.

Let $C \subseteq A$ and $\mathcal{B} \in \Phi'$. We call the quantity

$$\mathbf{\Delta_{A,C,\mathcal{B}}} = |C \cap \mathcal{B}| - |(A \smallsetminus C) \cap \mathcal{B}|$$

the **discrepancy** of C in $\mathcal{B}$. The **discrepancy** of C is

$$\mathbf{\Delta_{A,C}} = \max_{\mathcal{B} \in \Phi'} \left| \Delta_{A,C,\mathcal{B}} \right|.$$

Finally, the **discrepancy** of A is

$$\mathbf{\Delta_A} = \min_{C \subseteq A} \Delta_{A,C}.$$

We will use the same notation but with a lowercase δ for the **relative discrepancy**. This is obtained obtained dividing the discrepancy by the size of A.

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

2.1 Lemma *Let A be a sample of size n. Let $C \subseteq A$ have relative discrepancy $\delta_{A,C}$. Then either C or $A \smallsetminus C$ is an ε-approximation of size $\leq n/2$ for $\varepsilon = 2\delta_{A,C}$.*

Proof Define $n^+ = |C|$ and $n^- = |A \smallsetminus C|$. We may assume that $n^+ \leq n/2$, otherwise swap C and $A \smallsetminus C$. Then $\delta_{A,C,\mathcal{U}} = (n^+ - n^-)/n < 0$. Now, let $\mathcal{B} \in \Phi'$ be arbitrary

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

We also have

2.
$$\begin{aligned}
(*) &= \frac{|C \cap \mathcal{B})|}{n^+} \big(1 + \delta_{A,C,\mathcal{U}}\big) - \delta_{A,C,\mathcal{B}} \\
&\geq \frac{|C \cap \mathcal{B}|}{n^+} - 2\delta_{A,C}
\end{aligned}$$

Combining 1 and 2 we obtain

$$\begin{aligned}
\left| \mathrm{Fr}(\mathcal{B}/A) - \mathrm{Fr}(\mathcal{B}/C) \right| &\leq \left| \frac{|\mathcal{B} \cap A|}{n} - \frac{|\mathcal{B} \cap C)|}{n^+} \right| \\
&\leq 2\delta_{A,C}
\end{aligned}$$

as claimed by the lemma. □

3 Random colorings

We say that a tuple $a = \langle a_1, \ldots, a_n \rangle \in \mathcal{U}^n$ is an **enumeration** of the sample A if for all $x \in \mathcal{U}$

$$A(x) \quad = \quad \big|\{i \,:\, a_i = x\}\big|.$$

Clearly, all enumerations of A have length $n = |A|$. We write $\mathbf{rng}(a)$ for the sample enumerated by a.

Fix an enumeration $a = \langle a_1, \ldots, a_n \rangle$ of A. A tuple $c = \langle c_1, \ldots, c_n \rangle \in \{-1, +1\}^n$ is called a **coloring**. To each coloring c we associate a subsample of A, which we denote by $\mathbf{smpl}(a, c)$. It assigns to $x \in \mathcal{U}$ multiplicity $\big|\{i : a_i = x,\ c_i = +1\}\big|$.

Note that the quantity

$$\Delta_{a,c,\mathcal{B}} \quad = \quad \sum_{a_i \in \mathcal{B}} c_i.$$

coincide with the discrepancy of $\mathrm{smpl}(a, c)$ over $\mathcal{B}$.

There is probability measure on the subsamples of A which is computationally convenient. We assume that $C = \mathrm{smpl}(a, c)$ where $c = \langle c_1, \ldots, c_n \rangle \in \{\pm 1\}^n$ is obtained tossing n times a fair coin.

That is, we assume an uniform distribution on the colorings and this induces a multivariate binomial distribution on the subsamples. Note that this distribution s independent of the enumeration of A.

4 Useful inequalities

2.2 Proposition (Markov's inequality) *Let X be a random variable with finite mean*

$$\Pr\Big(X \geq \varepsilon\Big) \quad \leq \quad \frac{\mathrm{E}(X)}{\varepsilon}$$

Proof For simplicity, we will assume that the sample space Ω is finite. Define $A = \{a \in \Omega \,:\, X(a) > \varepsilon\}$.

$$\begin{aligned}
\mathrm{E}(X) \quad &= \quad \sum_{a \in \Omega} \Pr(a)\, X(a) \\
&= \quad \sum_{a \in A} \Pr(a)\, X(a) + \sum_{a \notin A} \Pr(a)\, X(a) \\
&\geq \quad \sum_{a \in A} \Pr(a)\, X(a) + \varepsilon \sum_{a \notin A} \Pr(a) \\
&\geq \quad \varepsilon \sum_{a \notin A} X(a) \leq \varepsilon \Pr(a) \\
&= \quad \varepsilon \Pr(X \leq \varepsilon)
\end{aligned}$$

□

2.3 Proposition (Chebychev's inequality) *Let X be a random variable with finite mean and finite variance.*

$$\Pr\Big(\big|X - \mathrm{E}(X)\big| \geq \varepsilon\Big) \quad \leq \quad \frac{\mathrm{Var}(X)}{\varepsilon^2}$$

Proof For simplicity, we will assume that the sample space Ω is finite. Define $A = \{a \in \Omega \,:\, |E(X) - X(a)| \geq \varepsilon\}$.

$$
\begin{aligned}
\mathrm{Var}(X) &= \sum_{a\in\Omega} \Pr(a)\,\big(X(a)-\mathrm{E}(X)\big)^2 \\
&= \sum_{a\in A} \Pr(a)\,\big(X(a)-\mathrm{E}(X)\big)^2 + \sum_{a\notin A} \Pr(a)\,\big(X(a)-\mathrm{E}(X)\big)^2 \\
&\geq \sum_{a\in A} \Pr(a)\,\big(X(a)-\mathrm{E}(X)\big)^2 + \varepsilon^2 \sum_{a\notin A} \Pr(a) \\
&\geq \varepsilon^2 \sum_{a\notin A} \Pr(a) \\
&= \varepsilon^2 \Pr\Big(\big|X-\mathrm{E}(X)\big| \leq \varepsilon\Big) \qquad \Box
\end{aligned}
$$

2.4 Proposition (Weak Law of Large Numbers) *Let* Pr *be a probability measure on* $\mathcal{U}$*. Then, for every event* $\mathcal{B}\subseteq\mathcal{U}$ *and every* $n,\varepsilon>0$

$$
\Pr\Big(c\in\mathcal{U}^n \;:\; \Big|\Pr(\mathcal{B})-\mathrm{Fr}(\mathcal{B}/c)\Big| \geq \varepsilon\Big) \quad\leq\quad \frac{1}{4n\varepsilon^2}
$$

Proof We can rephrase the inequality above as

$$
\Pr\Big(\big|p-\bar{X}\big|\geq\varepsilon\Big) \quad\leq\quad \frac{1}{4n\varepsilon^2} \qquad\qquad \text{where} \quad \bar{X}=\frac{1}{n}\sum_{i=1}^{n} X_i
$$

and X_i are Bernoulli variables with success probability p. In fact, let $p=\mu\mathcal{B}$ and let X_i output 1 if $c_i\in\mathcal{B}$ and 0 otherwise. Hence $\mathrm{E}[\bar{X}]=p$ and $\mathrm{Var}[\bar{X}]=p(1-p)/n$. Apply Chebyshev's inequality. $\Box$

We want to prove a better bound for the Weak Law of Large Numbers. So, we need to prove Hoeffding's inequality. For clarity we first prove a special case

2.5 Lemma (Chernoff's bound) *For* $i=1,\ldots,n$ *let* X_i *be independent identically distributed random variables such that* $\Pr(X_i=\pm1)=1/2$*. Then for every* $\varepsilon>0$

$$
\Pr\big(M\geq\varepsilon\big) \quad\leq\quad \exp\Big(-\frac{\varepsilon^2}{2n}\Big) \qquad\qquad \textit{where } M=\sum_{i=1}^{n} X_i
$$

Proof Let $t>0$ be arbitrary (we will set it at the end of the proof). Then

$$
\sharp \qquad
\begin{aligned}
\Pr(M\geq\varepsilon) &= \Pr\big(e^{tM}\geq e^{t\varepsilon}\big) \\
&\leq e^{-t\varepsilon}\,\mathrm{E}\big(e^{tM}\big)
\end{aligned}
$$

In fact, the equality follows because the exponential is an increasing function and the inequality is Markov's inequality. Now observe that

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

From this, by independence we have

$$\mathrm{E}\big(e^{tM}\big) \quad = \quad \prod_{i=1}^{n} e^{t\,X_i} \quad = \quad e^{n\,t^2/2}$$

Substituting in $\sharp$ gives $\Pr(M \geq \varepsilon) \leq e^{n\,t^2/2 - t\varepsilon}$. Finally Chernoff's inequality is obtained substituting ε/n for t. $\square$

2.6 Proposition (Hoeffding's inequality) *Let Y_i, for $i \in [n]$, be independent random variables with bounded range, say $a_i \leq Y_i \leq b_i$. Define*

$$M \quad = \quad \sum_{i=1}^{n} \big(Y_i - \mathrm{E}[Y_i]\big) \qquad \textit{and} \qquad c \quad = \quad \sum_{i=1}^{n} (a_i - b_i)^2$$

and assume $d > 0$ to avoid trivialities. Then for every $\varepsilon > 0$

$$\Pr\big(M \geq \varepsilon\big) \quad \leq \quad \exp\Big(-\frac{2\varepsilon^2}{d}\Big).$$

Proof Let $X_i = Y_i - \mathrm{E}[Y_i]$. The proof is identical to the one above, only the bound $\flat$ needs a new proof. Let $Z_i = (d_i + X_i)/2d_i$. Note that $0 \leq Z_i \leq 1$. By the convexity of the exponential function.

$$\begin{aligned} e^{tX_i} \quad &= \quad e^{td_i Z_i - td_i(1-Z_i)} \\ &\leq \quad Z_i\, e^{td_i} + (1 - Z_i)\, e^{-td_i} \end{aligned}$$

As $\mathrm{E}[Z_i] = 1/2$ we obtain

$$\begin{aligned} \mathrm{E}\big(e^{tX_i}\big) \quad &\leq \quad \frac{1}{2}e^{td_i} + \frac{1}{2}e^{-td_i} \\ &= \quad e^{t^2 d_i^2} \end{aligned}$$

Then, proceeding as in Proposition 2.5 we obtain $\mathrm{E}[M] = e^{t^2 d}$. By the Markov inequality $\Pr\big(M \geq \varepsilon\big) \leq e^{t^2 d - t\varepsilon}$. Substituting ε/d for t we obtain the required inequality. $\square$

2.7 Corollary (Weak Law of Large Numbers with exponential bound) *Let $\Pr$ be a probability measure on $\mathcal{U}$. Then, for every event $\mathcal{B} \subseteq \mathcal{U}$ and every $n, \varepsilon > 0$*

$$\Pr\Big(c \in \mathcal{U}^n \ :\ \Big|\Pr(\mathcal{B}) - \mathrm{Fr}(\mathcal{B}/c)\Big| \geq \varepsilon\Big) \quad \leq \quad 2\exp\Big(-\frac{n\varepsilon^2}{2}\Big)$$

Proof Reason as in the proof of Proposition 2.4 and apply Hoeffding's inequality with $M = n\bar{X}$ and $d = n$. $\square$

For future reference we include here the following simple fact.

2.8 Proposition *For every y and every $x > 1$*

$$\frac{x}{\ln x} \leq y \quad \Rightarrow \quad x \leq 2\,y\,\ln y$$

Proof It suffices to prove that for every $x > 1$

$$\frac{x}{\ln x} \leq y \quad \Rightarrow \quad \ln x \leq 2\,\ln y$$

As the latter inequality is equivalent to $\sqrt{x} \leq y$, it suffices to verify that $\sqrt{x} \leq \dfrac{x}{\ln x}$ for all $x > 1$. $\square$

5 A uniform law of large numbers

We want to estimate the minimal size of an ε-approximation of A. We want a bound that depends solely on ε, not on the size of A. It is also important to note that the definition of approximation includes a requirement of uniformity: the same subsample works for all definable sets.

If we allow $C \subseteq A$ to depend on $\mathcal{B}$, the existence of a subsample that satisfies the inequality

$$\Big|\operatorname{Fr}(\mathcal{B}/A) - \operatorname{Fr}(\mathcal{B}/C)\Big| \quad \leq \quad \varepsilon$$

follows easily from the law of large numbers and its size only depends on ε. This is essentially the weak law of large numbers for the probability measure $\operatorname{Fr}(\cdot/A)$.

2.9 Proposition *For every sample A, every $\mathcal{B} \subseteq \mathcal{U}$ and every $\varepsilon > 0$ there is a C of size $\leq 2\exp\big(-n\,\varepsilon^2/2\big)$ such that $\big|\operatorname{Fr}(\mathcal{B}/A) - \operatorname{Fr}(\mathcal{B}/C)\big| \leq \varepsilon$.*

More explicitly, let C ranges over the samples of size n with the product probability measure. Then

$$\Pr\{\big|\operatorname{Fr}(\mathcal{B}/A) - \operatorname{Fr}(\mathcal{B}/C)\big| \geq \varepsilon\} \quad \leq \quad 2\exp\Big(-\frac{n\varepsilon^2}{2}\Big)$$

Note that to avoid technicalities we do not insist in requiring $C \subseteq A$. The proof only works for $\operatorname{supp} C \subseteq \operatorname{supp} A$.

Proof Clearly, the first claim follows from the second. Interpret $\operatorname{Fr}(\cdot/A)$ as a probability measure on $\operatorname{supp} A$ an apply the weak law of large numbers with exponential bounds. □

The key result in VC-theory is the following theorem of Vapnik and Chervonenkis which is a uniform version of the weak law of large numbers.

2.10 Theorem (Uniform law of large numbers) *Assume Φ has VC-density d. Let A be any finite sample and let C range over the samples of size n with the product probability measure. Then*

$$\Pr\Big(\sup_{\mathcal{B}\in\Phi}\{\big|\operatorname{Fr}(\mathcal{B}/A) - \operatorname{Fr}(\mathcal{B}/C)\big|\} > \varepsilon\Big) \quad \leq \quad 8\,d\exp\Big(-\frac{n\varepsilon^2}{32}\Big)$$ □

We will not prove this theorem. The main result of this chapter is the following which a corollary which we proof independently in the next section. The corollary is all what we need in the sequel.

2.11 Corollary *Assume Φ has VC-density d. Then every sample A has an ε-approximation of size*

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

where c is an absolute constant.

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).

2.12 Lemma *Let A be a sample of size $\leq n$. Assume the support of A has $\leq m$ definable subsets. Then $\Delta_A \leq \sqrt{2n\ln(m+2)}$.*

Proof Let $\varepsilon = \sqrt{2n\ln(m+2)}$. Then we need to prove that $\Delta_A \leq \varepsilon$ which is equivalent to

$$\exists C \subseteq A\ \forall \mathcal{B} \in \Phi'\ \Delta_{A,C,\mathcal{B}} \quad \leq \quad \varepsilon$$

Let $\Pr(\cdot)$ denote the probability measure on the subsamples of A discussed in Section 3 (though, clearly, any other probability measure would do for the implication)

$$\Uparrow$$

$$\Pr\left(\forall \mathcal{B} \in \Phi' \ \ \Delta_{A,C,\mathcal{B}} \leq \varepsilon\right) \quad > \quad 0.$$

As Φ' has at most $m+1$ elements

$$\Uparrow$$

$$\forall \mathcal{B} \in \Phi' \ \Pr\left(\Delta_{A,C,\mathcal{B}} \geq \varepsilon\right) \quad \leq \quad \frac{1}{m+2}.$$

Fix an enumeration a of A. Suppose $c = \langle c_1, \ldots, c_n \rangle \in \{\pm 1\}^n$ is obtained tossing n times a fair coin. Consider the Bernoulli random variables X_i that on input c output c_i. For a given $\mathcal{B}$ define $M_{\mathcal{B}} = \sum\limits_{a_i \in \mathcal{B}} X_i$. Then $\Delta_{A,C,\mathcal{B}} = \Delta_{a,c,\mathcal{B}} = M_{\mathcal{B}}(c)$. Therefore

$$\Pr\left(\Delta_{A,C,\mathcal{B}} \geq \varepsilon\right) \quad \leq \quad \frac{1}{m+2}.$$

$$\Updownarrow$$

$$\Pr\left(M_{\mathcal{B}} \geq \varepsilon\right) \quad \leq \quad \frac{1}{m+2}.$$

By the Chernoff's bound

$$\Uparrow$$

$$\exp\left(-\frac{\varepsilon^2}{2|\mathcal{B}|}\right) \quad \leq \quad \frac{1}{m+2}.$$

$$\Updownarrow$$

$$2|\mathcal{B}|\ln(m+2) \quad \leq \quad \varepsilon^2$$

As $|\mathcal{B}| \leq n$, the latter is clearly true. □

2.13 Proof of Corollary 2.11 Set $A_0 = A$ and $\varepsilon_0 = 0$. We construct a decreasing chain A_i of ε_i-approximations. We denote by n_i and δ_i the cardinality, respectively the discrepancy, of A_i. By lemma 2.1, we can require that $\varepsilon_{i+1} = \varepsilon_i + 2\delta_i$ and $n_{i+1} \leq n_i/2$. Then

$$\varepsilon_h \quad = \quad 2\sum_{i=1}^{h} \delta_i$$

Let h be the largest such that $\varepsilon_h \leq \varepsilon$. The theorem claims that $2^{-h}n_0 \leq c\,(d/\varepsilon^2)\log(1/\varepsilon)$, independently of n_0.

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

and, as $n_i = 2^{-i}n_0$, the sum above is proportional to its largest term,

$$\leq \quad c\sqrt{(d/n_h)\ln(n_h)}$$

Hence a sufficient condition for $\varepsilon_h < \varepsilon$ is

$$c\sqrt{(d/n_h)\ln(n_h)} \quad \leq \quad \varepsilon$$

which is equivalent to

$$\frac{c^2 d}{\varepsilon^2} \quad \leq \quad \frac{n_h}{\ln n_h}$$

Let h be the maximal (or, n_h the minimal) that satisfies this inequality. Hence

$$\frac{c^2 d}{\varepsilon^2} \quad > \quad \frac{n_{h+1}}{\ln n_{h+1}}$$

By Proposition 2.8

$$n_{h+1} \quad \leq \quad 2\frac{c^2 d}{\varepsilon^2}\ln\frac{c^2 d}{\varepsilon^2}$$

hence

$$n_h \quad \leq \quad 4\frac{c^2 d}{\varepsilon^2}\ln\frac{c^2 d}{\varepsilon^2}$$

which proves the theorem.

Chapter 3

Packings and transversals

1 Linear programming / convex optimization

The following is a classical result. The proof may be found in any introductory text of linear programming. There are many (sometimes non equivalent) ways to state it. The following somewhat unusual phrasing is taken from Terence Tao's blog [3]. It stresses the analogies between Farkas' lemma and Hilbert's Nullstellensatz.

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

1. $\displaystyle\bigwedge_{i=1}^{m} P_i(x) \geq 0$ *has a solution* $x \in \mathbb{R}^n$;
2. *there are no* $0 \leq y \in \mathbb{R}^m$ *such that* $\displaystyle\sum_{i=1}^{m} y_i P_i = -1$. □

We translate the lemma in a more standard form. See [1] for a discussion of different formulations and for a proof of the following important theorem.

3.2 Proposition (Farkas' Lemma, second formulation) *Let A be a $m \times n$ real matrix. Let $b \in \mathbb{R}^m$ be a column vector. Then the following are equivalent*

1. $Ax \geq b$ *has a solution* $x \in \mathbb{R}^n$;
2. *all* $0 \leq y \in \mathbb{R}^m$ *such that* $A^{\mathrm{T}}y = 0$ *also satisfy* $b^{\mathrm{T}}y \geq 0$. □

3.3 Proposition (Duality for LP) *Let A be a $m \times n$ real matrix. Let $c \in \mathbb{R}^m$ and $b \in \mathbb{R}^n$ be column vectors. Then the following maximum and minimum exist and coincide.*

1. $\min\{\, c^{\mathrm{T}} x \;:\; A\,x \geq b,\ 0 \leq x\}$;
2. $\max\{\, b^{\mathrm{T}} y \;:\; A^{\mathrm{T}} y \leq c,\ 0 \leq y\}$. □

2 Transversals and packings

A subset $A \subseteq \mathcal{U}$ is a **tansversal** if $\varphi(A,b) \neq \varnothing$ for every $b \in \mathcal{V}$. Equivalently, if $\varphi(a,\mathcal{V})_{a\in A}$ covers $\mathcal{V}$. The **transversal number** of φ, denoted by $\tau(\varphi)$, is the smallest cardinality of a transversal of φ.

A subset $B \subseteq \mathcal{V}$ is a **packing** if $\varphi(\mathcal{U},b) \cap \varphi(\mathcal{U},b') = \varnothing$ for every $b,b' \in B$. Equivalently if $|\varphi(a,B)| \leq 1$ for every $a \in \mathcal{U}$. The **packing number** of φ, denoted by $\nu(\varphi)$, is the largest cardinality of a packing $B \subseteq \mathcal{V}$.

As any transversal is at least as large as any packing, we always have $\nu(\varphi) \leq \tau(\varphi)$. Very little can be said in the reverse direction in general.

A fractional multi-set A over $\mathcal{U}$ is a **fractional tansversal** if $|\varphi(A,b)| \geq 1$ for every

$b \in \mathcal{V}$. The **fractional transversal number** of φ, denoted by $\tau^*(\varphi)$, is the infimum of the size of the fractional transversals of φ.

A fractional multi-set B over $\mathcal{V}$ is a **fractional packing** if $|\varphi(a, B)| \leq 1$ for every $a \in \mathcal{U}$. The **fractional packing number** of φ, denoted by $\nu^*(\varphi)$, is the supremum of the size of the fractional packings of φ.

3.4 Example Let $\mathcal{U} = \mathbb{R}^2$ and $\mathcal{V}$ is a set of n lines in general position. Let φ be the incidence relation. Then $\nu(\varphi) = 1$, as any two lines intersect. And $\tau(\varphi) = \lceil n/2 \rceil$, as each point belongs to at most two lines.

Let B assign $1/2$ to every line. Then $|\varphi(a, B)| \leq 1$ holds because each point is contained in at most two lines. Then $\nu^*(\varphi) \geq n/2$. It is easy to see that $\tau^*(\varphi) = n/2$. If n is even, it is clear. If n is odd, take 3 any lines, and assign $1/2$ to the three intersection points. Proceed as in the even case with the other lines. □

3.5 Exercise Let $\mathcal{U} = \mathbb{R}$ and $\mathcal{V}$ is a set of finitely many closed intervals. Let φ be the incidence relation. Then $\nu(\varphi) = \tau(\varphi)$. Hint: use induction on $\nu(\varphi)$.

3.6 Theorem *For every finite incidence relation $\nu^*(\varphi) = \tau^*(\varphi)$ and this value is rational.*

Proof Let $\mathcal{U} = \{a_1, \ldots, a_m\}$ and $\mathcal{V} = \{b_1, \ldots, b_n\}$. Let F be the $\{0,1\}$-valued incidence matrix $\varphi(a_i, b_j)$. A multi-set over $\mathcal{U}$ is a naturally associated to a vector $0 \leq x \in \mathbb{R}^m$. A multi-set over $\mathcal{U}$ is associated to a vector $0 \leq y \in \mathbb{R}^n$. Then it is easy to verify that

$$\begin{aligned} \tau^*(\varphi) &= \min \left\{ 1_n^{\mathrm{T}}\, x \ : \ F\, x \geq 1_m,\ 0 \leq x \right\}. \\ \nu^*(\varphi) &= \max \left\{ 1_n^{\mathrm{T}}\, y \ : \ F^{\mathrm{T}} y \leq 1_n,\ 0 \leq y \right\}; \end{aligned}$$

Therefore, by he duality of linear programming $\nu^*(\varphi) = \tau^*(\varphi)$.

As $\tau^*(\varphi)$ is the minumum of the linear function $x \mapsto 1_n^{\mathrm{T}}\, x$ over a polyhedron, such minumum is attained at vertex. The inequalities describing the polyhedron have rational coefficients, so also the vertices have rational coordinates. □

3 Nets

Fix a set system Φ and let Pr be a probability measure on $\mathcal{U}$ such that all sets in Φ are measurable. Fix also $\varepsilon > 0$. An **ε-net** is a set $A \subseteq \mathcal{U}$ that intersects all sets in Φ of measure at least ε. In other words, an ε-net is a transversal for the set system $\Phi_\varepsilon = \{\mathcal{B} \in \Phi : \Pr(\mathcal{B}) \geq \varepsilon\}$.

3.7 Proposition *For any $d, \varepsilon > 0$, every measure on a set system Φ of VC-density $\leq d$, has an ε-net of cardinality $\leq c\,\dfrac{d}{\varepsilon}\,\ln\dfrac{1}{\varepsilon}$.*

Proof We only prove a weaker bound by observing that every ε-approximation is in particular an ε-net. Hence from Proposition 2.11 we obtain an ε-net of cardinality $\leq c\,(d/\varepsilon^2)\,\ln(1/\varepsilon)$. □

3.8 Proposition *Let Φ be a finite set system with VC-dimension $\leq k$. Then*

$$\tau(\Phi) \quad \leq \quad c\,\tau^*(\Phi)\,\ln\tau^*(\Phi)$$

where c is an absolute constant.

Proof Let $r = \tau^*(\Phi)$ ad let A be an optimal fractional transversal. As Φ is finite we can assume that the support of A is finite. After normalizing, A defines a probability measure on $\mathcal{U}$ concentrated on $\operatorname{supp} A$. Namely, $\Pr(\{x\}) = (1/r)A(x)$ for all $x \in \mathcal{U}$. By the definition of fractional transversal every set in Φ has measure at last $1/r$. Therefore every $(1/r)$-net is a transversal. By Proposition 3.7 we obtain a transversal of cardinality $\leq cr \ln r$. □

4 Helly-type properies

We now investigate methods of bounding $\tau^*(\varphi) = \nu^*(\varphi)$. Recall a classical theorem of Helly.

3.9 **Proposition (Helly Theorem)** *Let Φ be a finite family of convex sets in $\mathbb{R}^d$. Assume that any $d+1$ sets from Φ have non empty intersection. Then the whole family Φ has non empty intersection.*

Note that Helly's theorem does not hold for families of finite VC-dimension. A counter example of VC-dimension 2 can be constructed with a family containing sets that are unions of two finite intervals of the real line.

The following property is more robust.

3.10 **Definition** *Let Φ be an infinite family of sets. We say that Φ* ***has fractional Helly number k*** *if for every $\alpha > 0$ there is a $\beta > 0$ such that:*

if $S_i \in \Phi$, for $i = 1, \dots, n$ are such that

$$\bigcap_{i \in I} S_i \neq \varnothing \qquad \text{for at least } \alpha\binom{n}{k} \text{ sets } I \in \binom{[n]}{k}$$

then

$$\bigcap_{i \in J} S_i \neq \varnothing \qquad \text{for some } J \subseteq [n] \text{ of cardinality} \geq \beta n.$$

We say that Φ has the ***fractional Helly property*** *if it has some finite Helly number. The* ***fractional Helly number*** *of Φ is the smallest number k satisfying the property above.* □

For future reference we remark that the β above may be used to obtain a bound to $\nu(\varphi)$. In fact, if at least βn sets in Φ intersect, then $\nu(\Phi) \leq n - \beta n + 1$.

3.11 **Theorem (Matoušek [2])** *Let Φ be a set system with $\pi^*(\varphi) \in o(n^k)$. Then Φ has fractional Helly number $\leq k$.*

Recall that $\pi^*(\varphi) \in o(n^k)$ means that $\lim_{n\to\infty} \pi^*(\varphi)/n^k = 0$. Which occurs in particular when the VC-density is $< k$ or when the VC-dimension is $\leq k$.

Proof Let Φ and k be as in the assumptions of the theorem. Let α be arbitrary and set $\beta = 1/2m$ where m is such that

$$\pi^*(m) \quad < \quad \frac{\alpha}{4}\binom{m}{k}.$$

Fix some $S_i \in \Phi$, for $i = 1, \dots, n$ and assume

$$\bigcap_{i\in I} S_i \quad \neq \quad \varnothing \quad \text{for at least } \alpha\binom{n}{k} \text{ sets } I \in \binom{[n]}{k}$$

We will abbreviate the intersection above with S_I. We need to show that there is a set $J \subseteq [n]$ of cardinality $\geq \beta n$ such that $S_J \neq 0$. So, assume not and reason for a contradiction. Note that we can assume $n > 2m$ because for $n \leq 2m$ we can take take $|J| = 1$.

Identify $J \subseteq [n]$ with subsets of $\mathcal{V}$. We will find a set $J \subseteq [n]$ of cardinality m with many φ^*-definable subsets, more than $\pi^*(m)$, a contradiction.

Let P be the set of pairs $I \subseteq J \subseteq [n]$ such that $|I| = k$ and $|J| = m$. We say that a pair $I \subseteq J$ in P is *good* if there is $a \in \mathcal{U}$ such that $I = \{i \in J \ : \ a \in S_i\}$. That is, I is a φ^*-definable subset of J.

Claim. Assume on P the uniform probability. Then the probability that a random pair is good is $\geq \alpha/4$.

Assume the claim and proceed with the proof. We can think that the random pair in P is chosen by first picking $J \in \binom{[n]}{m}$ with the uniform distribution and than $I \in \binom{J}{k}$ again with the uniform distribution. (To say it more pedantically, we are applying the theorem of total probability.) If the the probability that a pair is good is $\geq \alpha/4$, then for at least one $J \in \binom{[n]}{m}$ the probability of finding a good subset I is $\geq \alpha/4$. So, that J has $\geq \frac{\alpha}{4}\binom{m}{k}$ good subsets. A contradiction which proves the theorem (given the claim).

We now prove the claim. There is another equivalent way to pick a random pair in P. First we choose at random $I \in \binom{[n]}{k}$ of cardinality k, then we choose at random $m-k$ elements from $[n] \smallsetminus I$. By assumption, the probability that $S_I \neq \varnothing$ is at least α. If $S_I \neq \varnothing$, we fix $a \in S_I$. By the assumption that we would like to contradict, a is contained in $< \beta n$ sets S_i. Then the probability that no set $S_i \in J \smallsetminus I$ contains a is at least

$$\begin{aligned}
\frac{\binom{n-\beta n}{m-k}}{\binom{n-k}{m-k}} \quad &\geq \quad \prod_{i=0}^{m-k-1} \frac{n-\beta n - i}{n-k-i} \quad \geq \quad \prod_{i=0}^{m-k} \frac{n-\beta n - i}{n-k-i} \\
&\geq \quad \prod_{i=0}^{m-k} \frac{n-\beta n - m}{n-m} \quad \geq \quad \left(\frac{n-\beta n-m}{n-m}\right)^m \\
&= \quad \left(1 - \frac{\beta n}{n-m}\right)^m
\end{aligned}$$

recall that $n > 2m$ and $\beta \leq 1/2m$, then

$$\geq \quad (1-2\beta)^m \quad = \quad \left(1 - \frac{1}{m}\right)^m \quad \geq \quad \frac{1}{4}$$

Therefore the probability that a random pair in P is good is at least $\alpha/4$. This proves the claim and with it the theorem. $\square$

5 The (p,q)-theorem

For integers $p \geq q$ we say that φ has the **(p,q)-property** if out of any p definable sets there are q sets with non empty intersection.

3.12 Theorem (Alon, Kleitman + Matoušek) *Let $p \geq q \geq d+1$ be natural numbers. Then there is a number $N = N(d,p,q)$ such that if φ has VC-codensity $\leq d$ satisfies the (p,q)-property then $\tau(\varphi) \leq N$.*

Proof As we are not trying to optimize N, we may prove the theorem for $q = d+1$.

By Proposition 3.8, $\tau(\varphi)$ is bounded by a function of $\tau^*(\varphi)$, so it is enough to bound $\tau^*(\varphi)$. By Proposition 3.6, $\tau^*(\varphi) = \nu^*(\varphi)$, so it is enough to bound $\nu^*(\varphi)$.

Let A be an optimal fractional packing. By Theorem 3.6 we may assume that A is rational valued. Therefore $A = (1/m)\,C$ where m is a positive integer and C is a (integral valued) multi-subset of $\mathcal{V}$.

As explained in Section 1, after replacing each $b \in \mathcal{V}$ with m copies $b_1, \ldots, b_m$, and each arc $\langle a, b\rangle$ in φ with the arcs $\langle a, b_1\rangle, \ldots, \langle a, b_m\rangle$, we can assume that C is a plain set. Call φ' the new incidence relation.

Claim 1. φ' satisfies the (p', q)-property with $p' = q(p-1)+1$.

A p'-collection of φ' definable sets either contains q copies of the same φ-definable set or it contains p distinct φ-definable sets. In both case it has a subcollection of q sets with non-empty intersection.

Claim 2. For some $\alpha = \alpha(p,q) > 0$ there are $\alpha\dbinom{n}{q}$ many collections in $\dbinom{\Phi}{q}$ with non-empty intersection.

Every p'-collection of definable sets contains at least one q-collection with non-empty intersection. Every q-collection is contained in $\binom{n-q}{p'-q}$ many p'-collections. Therefore the number q-collections of definable sets with non-empty intersection is

$$\frac{\dbinom{n}{p'}}{\dbinom{n-q}{p'-q}} = \frac{\dbinom{n}{q}}{\dbinom{p'}{q}}.$$

Therefore, the claim is proved with $\alpha = \dbinom{p'}{q}^{-1}$.

Claim 2. We claim that for some $\beta = \beta(p,q) > 0$ there are a least βn definable sets with non-empty intersection.

This follows from Claim 1 by Helly fractional theorem.

□

6 Notes and references

[1] Bernd Gärtner and Jiří Matoušek, *Approximation algorithms and semidefinite programming*, Springer, Heidelberg, 2012.

[2] Jiří Matoušek, *Bounded VC-dimension implies a fractional Helly theorem*, Discrete Comput. Geom. **31** (2004), no. 2, 251–255.

[3] Terence Tao, *The Hahn-Banach theorem, Menger's theorem, and Helly's theorem*, What's new (2008).