Scombinatorics

Notation and preliminaries **3**

1 Three minimax theorems **4**
- 1 Hall's Marriage Theorem 4
- 2 Kőnig's Minimax Theorem 5
- 3 Dilworth's Theorem 6

2 Set systems **8**
- 1 Sperner's Theorem 8
- 2 The Erdős-Ko-Rado Theorem 9

3 Stability **11**
- 1 The order property 11
- 2 Approximable sets 15

4 Vapnik-Chervonenkis theory **18**
- 1 The Vapnik-Chervonenkis dimension 18
- 2 The Sauer-Shelah lemma 20

5 Law(s) of large numbers **22**
- 1 Inequalities 22
- 2 Two Weak Laws of Large Numbers 26
- 3 A Uniform Law of Large Numbers 28
- 4 A Uniform Law of Large Numbers, again 30

6 Small transversals **34**
- 1 Transversals and packings 34
- 2 Helly-type properties 36
- 3 The (p,q)-theorem 38

7 Zarankiewicz problem(s) **40**
- 1 The Kővári-Sós-Turán Theorem 40
- 2 The NIP case 41

8 Szemerédi regularity lemma **42**
- 1 The regularity lemma 42
- 2 Stable regularity lemma 42
- 3 Distal regularity lemma 42

References 43

Notation and preliminaries

Let $\mathcal{U}$ and $\mathcal{V}$ be two infinite sets. Let $\varphi(x\,;z)$ be a relation symbol, or a formula, whatever. We denote by $\varphi(\mathcal{U}\,;\mathcal{V})$ the set $\{\langle a,b\rangle\in\mathcal{U}\times\mathcal{V}:\varphi(a\,;b)\}$ which we call: the relation defined by $\varphi(x\,;z)$. Sets of the form $\varphi(\mathcal{U}\,;b)=\{a\in\mathcal{U}:\varphi(a\,;b)\}$, for some $b\in\mathcal{V}$, are called definable sets.

In the first chapters we always restrict the study to the trace of $\varphi(\mathcal{U}\,;\mathcal{V})$ on some finite set $A\times B$, where $A\subseteq\mathcal{U}$ and $B\subseteq\mathcal{V}$. We wite $\varphi(A\,;B)$ for $\varphi(\mathcal{U}\,;\mathcal{V})\cap A\times B$. Similarly, we write $\varphi(A\,;b)$ for the trace of $\varphi(\mathcal{U}\,;b)$ on A, that is, the set $\varphi(\mathcal{U}\,;b)\cap A$. We call it a definable subset of A.

We denote by $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$ and $\varphi(A\,;b)_{b\in\mathcal{V}}$ the collection of definable sets, respectively definable subsets of A.

A similar notations is used with the roles of $\mathcal{U}$ and $\mathcal{V}$ interchanged. I.e., we think of $\mathcal{U}$ as the set of parameters defining subsets of $\mathcal{V}$. We write $\varphi(x\,;z)^{\mathrm{op}}$ to signal that the formula $\varphi(x\,;z)$ is considered in this dual setting.

For the model theoretic minded, it would be more appropriate to call the definable sets *global types*, respectively *types over* A. But this would make the terminology (if possible) more obscure.

For $k\leq|A|$ we use following notation interchangeably

$$\binom{A}{k} \quad=\quad A^{(k)} \quad=\quad \Big\{A'\subseteq A\,:\,|A'|=k\Big\}$$

For n a non negative integer we write

$$(n] \quad=\quad \{1,\dots,n\},$$

$$n \quad=\quad [n) \quad=\quad \{0,\dots,n-1\},$$

and

$$[n] \quad=\quad \{0,\dots,n\}$$

Note that the latter conflicts with the notation used in combinatorics.

Chapter 1

Three minimax theorems

Though apparently unrelated, the three theorems in this chapter can be derived one each other. We prove them in an arbitrary order.

As evident from the statement, the last two theorems are minimax theorems. The first theorem less so, hence the title is only approximately correct.

1 Hall's Marriage Theorem

Let $\varphi(x\,;z)$ be given. Let $A \subseteq \mathcal{U}$ and $B \subseteq \mathcal{V}$ be finite sets.

We say that $A' \subseteq A$ is a set of distinct representatives for $\varphi(A\,;B)$ if

$$|\varphi(A';b)| \quad = \quad |\varphi(a\,;B)| \quad = \quad 1 \qquad \text{for every } a,b \in A',B,$$

or, in other words, if $\varphi(A';B)$ is the graph of a bijection.

1.1 Hall's Marriage Theorem For every finite $B \subseteq \mathcal{V}$, the following are equivalent

1. $\varphi(A\,;B)$ has a set of distinct representatives;
2. $|B'| \ \leq \ \Big|\bigcup_{b\in B'} \varphi(A\,;b)\Big|$ for every $B' \subseteq B$.

Proof **(1⇒2)** The following holds for any set of distinct representatives A' and every set $B' \subseteq B$

$$|B'| \quad = \quad \Big|\bigcup_{b\in B'} \varphi(A';b)\Big| \quad \subseteq \quad \Big|\bigcup_{b\in B'} \varphi(A\,;b)\Big|.$$

(2⇒1) Reason by induction on the cardinality of B. If $|B| = 1$, the claim is clear. Now assume $|B| > 1$ and consider two cases.

a. This is the case when the inequality in 2 is strict for all nonempty $B' \subset B$. Pick any pair $a,b \in A,B$ such that $\varphi(a\,;b)$. Then $\varphi(A \setminus \{a\}\,;B \setminus \{b\})$ still satisfy 2. By induction hypothesis, it has a set of distinct representatives A'. Then $A' \cup \{a\}$ is a set of distinct representatives for $\varphi(A\,;B)$.

b. Suppose instead that for some nonempty $B' \subset B$ the inequality in 2 holds with equality. Define

$$A' \quad = \quad \bigcup_{b\in B'} \varphi(A\,;b)$$

It is clear that 2 holds for $\varphi(A';B')$. Below we prove that 2 also holds for $\varphi(A \setminus A'\,;\, B \setminus B')$. Once this claim is proved, we apply the induction hypothesis to obtain sets of distinct representatives for these two relations and note that their union is a set of distinct representatives for $\varphi(A\,;B)$.

To prove the claim assume that there is a set $B'' \subseteq B \smallsetminus B'$ that contradicts 2, then

$$\Big|\bigcup_{b\in B''} \varphi(A\smallsetminus A'\,;\,b)\Big| \quad < \quad |B''|.$$

By the definition of A', B'

$$\begin{aligned}\bigcup_{b\in B'\cup B''} \varphi(A\,;b) \quad &= \quad \bigcup_{b\in B'} \varphi(A\,;b) \;\cup\; \bigcup_{b\in B''} \varphi(A\,;b)\\ &= \qquad A' \qquad \cup\; \bigcup_{b\in B''} \varphi(A\,;b)\\ &= \qquad A' \qquad \cup\; \bigcup_{b\in B''} \varphi(A\smallsetminus A'\,;b)\end{aligned}$$

The two sets above are disjunct, hence

$$\Big|\bigcup_{b\in B'\cup B''} \varphi(A\,;b)\Big| \quad > \quad |A'|+|B''|$$

As $|A'| = |B'|$, by the choice of A', B', we obtain that $B' \cup B''$ contradicts the inequality in 2. This prove the claim and with it the theorem. $\square$

2 Kőnig's Minimax Theorem

Let $\varphi(x\,;z)$ be given. Let $A \subseteq \mathcal{U}$ and $B \subseteq \mathcal{V}$ be finite sets.

A matching of $\varphi(A\,;B)$ is a pair of sets $A' \subseteq A$ and $B' \subseteq B$ such that $\varphi(A'\,;B')$ is the graph of a bijection between A' and B' in other words

$$|\varphi(A';b)| \quad = \quad |\varphi(a\,;B')| \quad = \quad 1 \quad \text{for every } a,b \in A',B'.$$

Yet in other words, A' is a set of distinctive representatives for $\varphi(A\,;B')$

We call $|A'| = |B'|$ the cardinality of the matching. The matching number of $\varphi(A\,;B)$ is the maximal cardinality of a matching.

Note that is A' is a set of distinct representatives for $\varphi(A\,;B)$, then there is a $B' \subseteq B$ such that A', B'. Hence the matching number is less or equal than the cardinality of any set of distinct representatives (if it exists).

A (vertex) cover of $\varphi(A\,;B)$ is a pair of sets $A' \subseteq A$ and $B' \subseteq B$ such that $\varphi(A\,;B)$ is contained in $\big(A' \times (B \smallsetminus B')\big) \cup \big((A \smallsetminus A') \times B'\big)$. We will mainly use this property as characterized by the easy fact below.

1.2 Fact The following are equivalent

1. A', B' is a cover;
2. $\varphi(A\,;b) \subseteq A'$ for every $b \in B \smallsetminus B'$;
3. $\varphi(a\,;B) \subseteq B'$ for every $a \in A \smallsetminus A'$. $\square$

We call $|A'| + |B'|$ the cardinality of the cover. The cover number of $\varphi(A\,;B)$ is the minimal cardinality of a cover.

1.3 Kőnig's Minimax Theorem For any given $\varphi(A\,;B)$, matching number = cover number. That is, the maximal cardinality of a matching equals the minimal cardinality of a cover.

Proof ($\leq$) We prove that $|A''| \leq |A'| + |B'|$ for every cover A', B' and every matching A'', B''.

As $\varphi(A\,;b) \subseteq A'$ for every $b \in B \smallsetminus B'$, in particular we have that $\varphi(A'';b) \subseteq A'$ for every $b \in B'' \smallsetminus B'$. By the definition of matching, all these sets $\varphi(A'';b)$ are distinct singletons. Hence $|B''| - |B'| \leq |B \smallsetminus B'| \leq |A'|$ is clear.

($\geq$) Let A', B' be a cover of minimal cardinality. We prove that there is a matching of cardinality at least $|A'| + |B'|$.

We break $\varphi(A\,;B)$ into two relations, find a matching of each of these and join them together to obtain a matching of cardinality $\geq |A'| + |B'|$. Precisely, first we show that $\varphi(A \smallsetminus A'\,;B')$ has a set of distinct representatives $A_1 \subseteq A \smallsetminus A'$. Hence A_1, B' is a matching. Second, we apply the same argument shows that $\varphi(A'\,;B \smallsetminus B')^{\mathrm{op}}$ has a set of distinct representatives $B_2 \subseteq B \smallsetminus B'$. Hence A', B_1 is a matching. Then $(A_1 \cup A')$, $(B' \cup B_1)$ is a matching of $\varphi(A\,;B)$. The cardinality of this matching is $|A_1| + |A'| = |B_1| + |A'| = |B'| + |A'|$.

We use Hall's Marriage Theorem to prove the first claim above. The second is proved by the symmetric argument (using 3 of the fact above in place of 2).

We need to check that $\varphi(A \smallsetminus A'\,;B')$ satisfies 2 of Theorem 1.1. Suppose not. Then there is a set $B'' \subseteq B'$ such that $|A''| < |B''|$, where

$$A'' \quad = \quad \bigcup_{b \in B''} \varphi(A \smallsetminus A', b)$$

Then $(A' \cup A''), (B' \smallsetminus B'')$ would be a cover of cardinality $< |A'| + |B'|$. This contradicts the minimality of A', B'. $\square$

3 Dilworth's Theorem

Dilworth's Theorem is minimax theorem essentially equivalent to Kőnig's Theorem. To highlight the connection we choose to prove it using Kőnig's Theorem. Alternatively we could have proved Dilworth's Theorem directly and derived Kőnig's and Hall's Theorem from it.

Let $<$ be a strict partial order on $\mathcal{U}$. An antichain is a set $A' \subseteq \mathcal{U}$ such that $a \not< a'$ for every $a, a' \in A'$. A chain is a set $A' \subseteq \mathcal{U}$ such that $a < a' \ \vee \ a' < a$ for every distinct $a, a' \in A'$.

1.4 Dilworth's Theorem Let $A \subseteq \mathcal{U}$ be finite. The maximal cardinality of an antichain $A' \subseteq A$ equals the minimal cardinality of a partition of A into chains.

Proof ($\leq$) We prove that the cardinality of an antichain cannot exceed the cardinality of a partition of A into chains.

Let $A_1, \ldots, A_k$ be a partition of A into chains and let A' be an antichain. A chain can contain at most one element of A', hence $|A'| \leq k$.

($\geq$) Let $A' \subseteq A$ be an antichain of maximal cardinality. We prove that there is a partition $A_1, \ldots, A_k$ into chains for some $k \leq |A'|$.

Let $\mathcal{V}$ be a disjoint copy of $\mathcal{U}$. Let $f : \mathcal{U} \to \mathcal{V}$ the bijection that maps each element of $\mathcal{U}$ to its copy in $\mathcal{V}$. For $a, b \in \mathcal{U}$ such that $a < b$ let $\varphi(a\,; fb)$. Let A_1, B_1 be a cover of $\varphi\big(A\,; f[A]\big)$. We claim that $A \smallsetminus \big(A_1 \cup f^{-1}[B_1]\big)$ is an antichain. In fact, if $a < b$ then either $a \in A_1$ or $fb \in B_1$, by the definition of cover. This proves the claim.

As A' has maximal cardinality, $|A| - |A'| \leq |A_1 \cup f^{-1}[B_1]| \leq |A_1 \cup B_1|$. If we choose a over A_1, B_1 of minimal cardinality, by Kőnig's Theorem there is a matching A'', B'' of cardinality $|A''| \geq |A_1 \cup B_1|$. Hence $|A''| \geq |A \smallsetminus A'|$.

We construct a chain-partition of A as follow. Pick an element of $a_0 \in A''$ and construct the longest possible chain $a_0, b_0, a_1, b_1, \ldots, a_m, b_m, a_{m+1}$ where $a_i \in A''$ for all $i \leq m$, and $b_i \in B''$ is the (unique) element such that $\varphi(a_i\,; b_i)$ and $a_{i+1} \in A$ is the copy of $b_i \in B''$. The construction halts at the first $a_{m+1} \notin A''$. Then we start a new chain from some fresh element of A'' until the chains $a_0 < a_1 < \cdots < a_m < a_{m+1}$ constructed in this way cover the whole of A''. Note that these chains are pairwise disjoint. Finally, put each element of A not covered by these chains in a chain on its own.

Notice that the elements of A'' belongs to a chain of length at least 2. Therefore the number k of chains necessary to cover A is $\leq |A| \smallsetminus |A''| \leq |A'|$. $\square$

Chapter 2

Set systems

1 Sperner's Theorem

We say that $\varphi(A\,;b)_{b\in\mathcal{V}}$ is an antichain if there is no pair of distinct elements $b, b' \in \mathcal{V}$ such that $\varphi(A\,;b) \subset \varphi(A\,;b')$. Antichains are also called Sperner systems.

If all sets in $\varphi(A\,;b)_{b\in\mathcal{V}}$ are distinct and of equal cardinality, then we clearly have an antichain. If $|A| = n$, the cardinality of a collection of subsets of A, all of cardinality k, is maximal when $k = \lfloor n/2 \rfloor$ or $k = \lceil n/2 \rceil$. In this case

$$\begin{aligned} \left|\varphi(A\,;b)_{b\in\mathcal{V}}\right| &= \binom{n}{\lfloor n/2 \rfloor} \\ &= \binom{n}{\lceil n/2 \rceil}. \end{aligned}$$

By the following classical theorem, this bound holds for all antichain. This is one of the first results of external combinatorics (though the term has been coined a few years later).

2.1 Sperner's Theorem Let $A \subseteq \mathcal{U}$ have cardinality n, finite. If $\varphi(A\,;b)_{b\in\mathcal{V}}$ is an antichain then

$$\left|\varphi(A\,;b)_{b\in\mathcal{V}}\right| \leq \binom{n}{\lfloor n/2 \rfloor}.$$

Proof Clearly, $\varphi(A\,;b)_{b\in\mathcal{V}}$ is the disjoint union of the sets $\binom{A}{k} \cap \varphi(A\,;b)_{b\in\mathcal{V}}$ for k ranging over $\{0,\dots,n\}$. Then

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

As for every $k \leq n$

$$\binom{n}{k} \leq \binom{n}{\lfloor n/2 \rfloor},$$

the theorem follows immediately from the LYM inequality that we prove below. □

The acronym LYM stands for Lubell-Yamamoto-Meshalkin.

2.2 Lemma (LYM inequality) Let $A \subseteq \mathcal{U}$ have cardinality n, finite. If $\varphi(A\,;b)_{b\in\mathcal{V}}$ is an antichain then

$$\sum_{k=0}^{n} \left|\binom{A}{k} \cap \varphi(A\,;b)_{b\in\mathcal{V}}\right| \cdot \binom{n}{k}^{-1} \leq 1.$$

Proof Let Π be uniform random variable that ranges over the set of permutations of $A = \{a_1, \dots, a_n\}$. For any $\varphi(A\,;b)$ of cardinality k

$$\mathbb{P}\Big(\Pi\{a_1,\dots,a_k\} = \varphi(A\,;b)\Big) \quad = \quad \binom{n}{k}^{-1}.$$

The events above are disjoint for distinct sets $\varphi(A\,;b)$, hence

$$\mathbb{P}\Big(\Pi\{a_1,\dots,a_k\} \in \varphi(A\,;b)_{b\in\mathcal{V}}\Big) \quad = \quad \left|\binom{A}{k} \cap \varphi(A\,;b)_{b\in\mathcal{V}}\right| \cdot \binom{n}{k}^{-1}.$$

As $\varphi(A\,;b)_{b\in\mathcal{V}}$ is an antichain, for distinct k the events above are disjoint, hence

$$\mathbb{P}\Big(\bigcup_{k=0}^{n} \Pi\{a_1,\dots,a_k\} \in \varphi(A\,;b)_{b\in\mathcal{V}}\Big) \quad = \quad \sum_{k=0}^{n}\left|\binom{A}{k} \cap \varphi(A\,;b)_{b\in\mathcal{V}}\right| \cdot \binom{n}{k}^{-1}.$$

Now, the inequality is evident. □

Let $\mathbb{P}_k$ be the probability measure on the subsets of A that is concentrated and uniform on $A^{(k)}$. Namely, for $A' \subseteq A$

$$\mathbb{P}_k\big(\{A'\}\big) \quad = \quad \begin{cases} 0 & \text{if } \ |A'| \neq k \\ \binom{n}{k}^{-1} & \text{if } \ |A'| = k \end{cases}$$

Then the the LYM inequality asserts that if $\varphi(A\,;b)_{b\in\mathcal{V}}$ is an antichain then

$$\sum_{k=0}^{n} \mathbb{P}_k\big(\varphi(A\,;b)_{b\in\mathcal{V}}\big) \quad \leq \quad 1.$$

This inequality is strict when $\varphi(A\,;b)_{b\in\mathcal{V}} = A^{(k)}$ for some k. In the next section we show that these are the only cases.

2 The Erdős-Ko-Rado Theorem

2.3 Lemma (Peter J. Cameron) Let G be a 1-transitive finite graph. If G contains a clique of cardinality m, then every subgraph $H \subseteq G$ contains a clique of cardinality

$$\geq \ m\frac{|H|}{|G|}.$$

Proof Let C be a clique in G of cardinality m. Let k the cardinality of the largest clique in H. Let $n = |\mathrm{Aut}(G)|$. By 1-transitivity, the sets $\{f \in \mathrm{Aut}(G) : fa = b\}$, for any fixed $a \in G$ and b ranging over G, have all the same cardinality. Hence, for any given pair $\langle a, b\rangle$, they have cardinality $n/|G|$.

Count the pairs $\langle a, f\rangle \in C \times \mathrm{Aut}(G)$ such that $fa \in H$. For every $a \in C$ there are $n \cdot |H|$ automorphisms. So the number of pairs is $m \cdot n \cdot |H|/|G|$

On the other hand for each $f \in \mathrm{Aut}(G)$ there are at most k choices of $a \in C$. So $m \cdot n \cdot |H|/|G| \leq k\,n$. □

2.4 Erdős-Ko-Rado Theorem Let $A \subseteq \mathfrak{U}$ be a finite set of cardinality n. Let $k \leq n/2$. Let $\varphi(A\,;b)_{b\in\mathcal{V}}$ be an intersecting family of sets of cardinality k. Then

$$\left|\varphi(A\,;b)_{b\in\mathcal{V}}\right| \ \leq \ \binom{n-1}{k-1}.$$

Proof Let $m = \left|\varphi(A\,;b)_{b\in\mathcal{V}}\right|$. Consider the graph

$$\begin{aligned} G &= \binom{A}{k}, \\ E(G) &= \Big\{\{A', A''\} \;:\; A' \cap A'' \neq \varnothing\Big\}. \end{aligned}$$

Enumerate the elements of A, say $A = \{a_0, \ldots, a_{n-1}\}$. Consider the following subgraph of G

$$H = \Big\{\{a_i, \ldots, a_{i+k-1}\} \;:\; 0 \leq i < n\Big\},$$

where the indices are intended modulo n. As $k \leq n$, the largest clique in H has cardinality k. As $\varphi(A'\,;b)_{b\in\mathcal{V}}$ is a clique of G, by the lemma above,

$$k \geq m\frac{|H|}{|G|} = m \cdot n \cdot \binom{n}{k}^{-1}$$

therefore

$$m \leq \binom{n-1}{k-1}$$

□

Chapter 3

Stability

1 The order property

The chain index of $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$, or of $\varphi(x\,;z)$ when $\mathcal{U}$ and $\mathcal{V}$ are clear, is the maximal length (that is, $n+1$) of a chain of the form

ch $\qquad \varphi(A\,;b_0)\ \subset\ \ldots\ \subset\ \varphi(A\,;b_n)$

for some set $A \subseteq \mathcal{U}$ and some $b_0,\ldots,b_n \in \mathcal{V}$. Note that we allow $\varphi(A\,;b_0)$ to be empty. This choice produces a small asymmetry below in the definition of ladder; see also Fact 3.3.

3.1 Example If $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$ consists of just one set, the chain index is 1. If it contains two distinct sets, the chain index is at least 2 and it is exactly 2 if there are no more two sets, or if all sets are disjoint. $\square$

If a maximal length does not exist, we say that $\varphi(x\,;z)$ is unstable, or that it has the order-property. Otherwise we say that it is stable.

In place of requiring the existence of the chain in ch, we could equivalently ask for a pair of tuples $a_1,\ldots,a_n \in \mathcal{U}$ and $b_0,\ldots,b_n \in \mathcal{V}$ such that

ld $\qquad \varphi(a_h\,;b_k)\ \Leftrightarrow\ h \leq k.$

We call this pair of tuples a ladder of length $n+1$. We may also say ladder index instead of chain index. Setting $A = \{a_1,\ldots,a_n\}$ we easily obtain a chain from a ladder, the converse is left as an easy exercise for the reader.

3.2 Exercise Let $\varphi(x\,;z)$ have chain index $n+1$ or more. Let $A \subseteq \mathcal{U}$ be a minimal set such that a chain as in ch obtains for some $b_0,\ldots,b_n \in \mathcal{V}$. Prove that there is a ladder $a_1,\ldots,a_n$ and $b_0,\ldots,b_n$ such that $A = \{a_1,\ldots,a_n\}$ and conclude that $\varphi(x\,;z)$ has ladder index $n+1$. $\square$

The following facts are obvious but worth noting.

3.3 Fact *Let $\varphi(x\,;z)$ have ladder index $n+1$. Then $\varphi(x\,;z)^{\rm op}$ has ladder index $\geq n$.* $\square$

3.4 Fact *Let $\varphi(x\,;z)$ have ladder index n. Then $\neg\varphi(x\,;z)$ has ladder index n.* $\square$

Proof If for all $h \in (n]$ and $k \in [n]$

$$\neg\varphi(a_h\,;b_k)\ \Leftrightarrow\ h \leq k$$

then $a'_h = a_{n+1-h}$ and $b'_k = b_{n-k}$ satisfy

$$\begin{aligned} \varphi(a'_h\,;b'_k) &\Leftrightarrow \varphi(a_{n+1-h}\,;b_{n-h}) \\ &\Leftrightarrow n-k > n+1-h \\ &\Leftrightarrow h \le k \end{aligned}$$

□

The following definition is connected with those above, though in a less evident manner. We write ${}^n2$ for the set of binary sequences of length n or, more precisely, the set of functions $s : [n) \to [2)$. We write s_h for the value of s at h, and $s{\restriction}h$ for the restriction of s to $[h)$. We define ${}^{<n}2 = \big\{r \,:\, r \in {}^h2,\ h \in [n)\big\}$.

A branching tree of hight n for the formula $\varphi(x\,;z)$ is a function

$$\begin{aligned} \bar{a} \;:\; {}^{<n}2 &\;\to\; \mathcal{U} \\ r &\;\mapsto\; a_r, \end{aligned}$$

which we may also present by writing $\bar{a} = \langle a_r : r \in {}^{<n}2\rangle$, such that

$$\varnothing \;\neq\; \bigcap_{h=0}^{n-1} \neg^{s_h}\,\varphi(a_{s\restriction h}\,;\mathcal{V}) \qquad\qquad \text{for all } s \in {}^n2. \tag{2r}$$

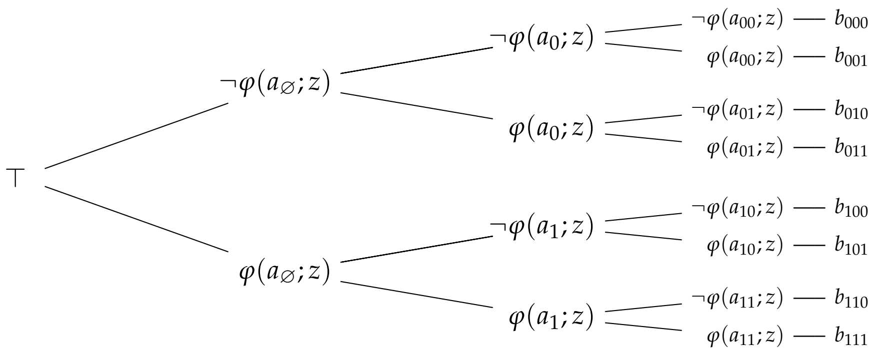
where $\neg^i$, for i a non negative integer, denotes a negation symbol repeated i times. In other words 2r requires the existence of some $\langle b_s : s \in {}^n2\rangle$ such that

$$\varphi(a_{s\restriction h}\,;b_s) \;\Leftrightarrow\; s_h = 1 \qquad\qquad \text{for all pairs } s \in {}^n2 \text{ and } h \in [n)$$

or, with slightly different notation,

$$\varphi(a_r\,;b_s) \;\Leftrightarrow\; r^\frown 1 \subseteq s \qquad\qquad \text{for all pairs } r \subset s \in {}^n2.$$

It helps to represent a branching tree as follows. For definiteness, fix $n = 3$. Consider a full binary tree of height $n+1$ and assign to each internal node (different from root and leaves) a formula as depicted below. Then 2r requires that all formulas in each branch $s \in 2^n$ are satified by some $b_s \in \mathcal{V}$.



The Shelah local 2-rank of $\varphi(x\,;z)$ is the maximal height of a branching tree for $\varphi(x\,;z)$. If such a maximal integer does not exist, we say that the 2-rank is infinite.

3.5 Example If $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$ consists of just one set, the only branching tree for $\varphi(x\,;z)$ is the empty tree. Therefore the 2-rank is 1. If there are at least two distinct definable sets, then we can always find $a_\varnothing, b_0, b_1$ such that $\varphi(a_\varnothing, b_0) \wedge \varphi(a_\varnothing, b_1)$ hence the 2-rank is at least 2. □

A branching tree $\bar{a}' = \langle a'_r : r \in {}^{m}2\rangle$ is a subtree of $\bar{a}$ if there is an $\subseteq$-preserving map $f : {}^{m}2 \to {}^{n}2$ such that $a'_r = a_{fr}$.

For the theorem below we need the following Ramsey-like lemma.

3.6 Lemma Let $\bar{a} = \langle a_r : r \in {}^{<n}2\rangle$ be a branching tree for $\varphi(x\,;z)$. Let $k \in [n]$ be given. Then, for every red-blue coloring of range$(\bar{a})$, there is a monochromatic branching subtree, say $\bar{a}' = \langle a'_r : r \in {}^{n'}2\rangle$, and either all nodes of $\bar{a}'$ are red and $n' = k$, or they are all blue and $n' = n - k$.

Proof Induction on n. First, note that the lemma is trivial when $k = 0$ or $k = n$. In particular the lemma holds for $n = 1$.

Assume the lemma true for n and prove it for $n+1$. Let $\bar{a} = \langle a_r : r \in {}^{n+1}2\rangle$ be given. Fix some $k \in [n+1]$. We want a branching tree $\bar{a}'$ that is either red of height k or blue of height $n+1-k$.

If we discard the trivial cases, we can assume $k \in (n]$.

For $i = 0,1$ we define $\bar{a}_i = \langle a_{i^\frown r} : r \in {}^{<n}2\rangle$.

First suppose that $a_\varnothing$ is blue. If, for either $i = 0$ or $i = 1$, there is a red branching tree $\bar{a}'_i$ of height k, we are done. Otherwise, for both $i = 0,1$ there is a blue branching tree $\bar{a}'_i$ of hight $n - k$. Then we graft these two trees on $a_\varnothing$, which is also blue, and obtain the required blue tree of height $n - k + 1$.

Now suppose that $a_\varnothing$ is red. If, for either $i = 0$ or $i = 1$, there is a blue branching tree $\bar{a}'_i \subseteq \bar{a}_i$ of height $n - (k-1)$, we are done. Otherwise, we graft on $a_\varnothing$ two red trees of hight $k - 1$ to obtain a red tree of height k. □

We are now ready to characterize stability via the 2-rank. The proof is based on Hodges [7]. It is a direct proof which yields an explicit bound on the 2-rank given the ladder index (it yields also the converse, but this is easy). This bound may be far from optimal.

Shelah was the first to prove the equivalence 1⇔2 below. His proof is model theoretic and does not gives explicit bounds. However it introduces some deep insight on sable formulas that we will present in the next section.

3.7 Theorem The following are equivalent

1. $\varphi(x\,;z)$ is stable;
2. $\varphi(x\,;z)$ has finite 2-rank.

Precisely, if $n_{\rm ld}$ and $n_{\rm 2r}$ are the ladder index and the 2-rank, respectively, then

$$\begin{aligned} n_{\rm ld} &< 2^{n_{\rm 2r}+1}; \\ n_{\rm 2r} &< 2^{n_{\rm ld}+1} - 2. \end{aligned}$$

Proof (2⇒1) We prove the contrapositive. We show that if there is a ladder of length $m = 2^n$, say $a_1, \dots, a_{m-1}$ and $b_0, \dots, b_{m-1}$, then there is a branching tree $\bar{a}'$ of height n. This also proves the first inequality above. In fact, if $n_{\rm ld} \geq 2^{n_{\rm 2r}+1}$, there would exist a branching tree of height $n_{2\mathrm{r}}+1$ which is a contradiction.

The branching tree $\bar{a}' = \langle a'_r : r \in {}^{<n}2 \rangle$ is defined as follows

$a'_r = a_h$ where h is obtained reading $r^\frown 1^\frown 0^{n-|r|-1}$ as an n-digit binary number.

To verify 2r we define for $s \in {}^n 2$

$b'_s = b_k$ where k is obtained reading s as an n-digit binary number.

Then it is easy to verify that for all pairs $r \subset s \in {}^n 2$

$$\begin{aligned}
\varphi(a'_r\,; b'_s) &\Leftrightarrow \varphi(a_h\,; b_k) && \text{where } h \text{ and } k \text{ are like above}\\
&\Leftrightarrow h \leq k\\
&\Leftrightarrow r^\frown 1^\frown 0^{n-|r|-1} \;\leq\; s && \text{as } n\text{-digit binary numbers}\\
&\Leftrightarrow r^\frown 1 \;\subseteq\; s
\end{aligned}$$

(1 ⇒ 2) We prove the contrapositive. We claim that if there is a branching tree $\bar{a}$ of height 2^n-2 then there is a ladder $a_1,\ldots,a_{n-1}$ and $b_0,\ldots,b_{n-1}$, of length n, such that $a_1,\ldots,a_{n-1} \in \mathrm{range}(\bar{a})$ and $b_0,\ldots,b_{n-1} \in \{b_s : s \in {}^{2^n-2}2\}$. This yields also the second inequality of the theorem. In fact, if $n_{2\mathrm{r}} \geq 2^{n_{\mathrm{ld}}+1}-2$, there would exist a ladder of length $n_{\mathrm{ld}}+1$ which is a contradiction.

As $n_{\mathrm{ld}} \geq 1$, we start the induction from $n=2$. In this case the claim is witnessed by ladder $a_\varnothing$ and b_0, b_1. Now we assume the claim is true for n and prove it for $n+1$.

Let $\langle a_r : r \in {}^{<2m+2}2 \rangle$, where $m = 2^n-2$, be a branching tree of height $2^{n+1}-2$. To each $b \in \{b_s : s \in {}^{2m+2}2\}$ we associate a red-blue coloring of $\mathrm{range}(\bar{a})$ as follows. A node $a \in \mathrm{range}(\bar{a})$ is colored

red if $\varphi(a\,;b)$ holds;

blue otherwise, that is, $\neg\varphi(a\,;b)$.

We consider two cases that are exhaustive by Lemma 3.6. Note that we are applying the lemma only to the subtree $\langle a_{1^\frown r} : r \in {}^{2m+1}2 \rangle$.

Case 1: for some b there is a red subtree of $\langle a_{1^\frown r} : r \in {}^{2m+1}2 \rangle$ of height $m+1$. Let $\bar{a}'$ be this red tree and consider its subtree $\langle a'_{0^\frown r} : r \in {}^{<m}2 \rangle$. By induction hypothesis, there are $A \subseteq \{a'_{0^\frown r} : r \in {}^{<m}2\}$ and $b_0,\ldots,b_{n-1}$ such that

(1) $\quad \varphi(A\,;b_0) \;\subset\; \ldots \;\subset\; \varphi(A\,;b_{n-1})$

Let $A' = A \cup \{a'_\varnothing\}$ then

(2) $\quad \varphi(A'\,;b_0) \;\subset\; \ldots \;\subset\; \varphi(A'\,;b_{n-1})$

In fact, as $b_0,\ldots,b_{n-1} \in \neg\varphi(a'_\varnothing\,;\mathcal{V})$, this is the same chain as (1). Therefore, if we extend the chain on the right with $\varphi(A'\,;b) = A'$, we obtain the required chain of length $n+1$.

Case 2: for every b there is a blue subtree of $\langle a_{1^\frown r} : r \in {}^{2m+1}2 \rangle$ of height m. Pick any $b \in \{b_s : s \in {}^{2m+2}2\}$ such that $\neg\varphi(a_\varnothing, b)$ and let $\bar{a}'$ the corresponding blue subtree. Apply the induction hypothesis to obtain $A \subseteq \mathrm{range}(\bar{a}')$ and $b_0,\ldots,b_{n-1}$ such that (1). We claim that (2) above holds with $A' = A \cup \{a_\varnothing\}$. In fact, $b_0,\ldots,b_{n-1} \in \varphi(a_\varnothing\,;\mathcal{V})$ so (2) is the chain in (1) with all sets augmented by $a_\varnothing$. We can extend the chain on the left with $\varphi(A';b) = \varnothing$ and obtain the required chain of length $n+1$. □

2 Approximable sets

The notion of approximable set is a combinatorial counterpart of the model theoretical notion of externally definable set.

3.8 Definition We say that $\mathcal{A} \subseteq \mathcal{U}$ is approximable if for every finite set $A \subseteq \mathcal{U}$ there is a $b \in \mathcal{V}$ such that $\varphi(A;b) = \mathcal{A} \cap A$. If we also have that $\varphi(\mathcal{U};b) \subseteq \mathcal{A}$, then we say that $\mathcal{A}$ is approximable from below. □

The following is immediate.

3.9 Fact The following are equivalent for every $\mathcal{A} \subseteq \mathcal{U}$

1. $\mathcal{A}$ is approximable from below;
2. for every finite set $A \subseteq \mathcal{A}$ there is a $b \in \mathcal{V}$ such that $A \subseteq \varphi(\mathcal{U};b) \subseteq \mathcal{A}$. □

Towards the main theorem of this section we prove three separated lemmas.

3.10 Lemma Let $\varphi(x;z)$ be a stable formula. Every set $\mathcal{A} \subseteq \mathcal{U}$ approximable by $\varphi(x;z)$ is approximable from below by the formula

$$\psi(x;z_0,\ldots,z_n) \quad = \quad \bigwedge_{i=0}^{n} \varphi(x;z_i)$$

where $n+1$ is the ladder index of $\varphi(x;z)$.

Proof Let $A \subseteq \mathcal{A}$ be finite. We prove that $A \subseteq \psi(\mathcal{U};b_0,\ldots,b_n) \subseteq \mathcal{A}$ for some $b_0,\ldots,b_n$. To obtain the b_k, we construct a ladder inductively. Pick any b_0 be such that $A \subseteq \varphi(\mathcal{U};b_0)$, which we can do by approximability. Now, suppose that $a_1,\ldots,a_k$ and $b_0,\ldots,b_k$ have been defined. If

$$A \quad \subseteq \quad \bigcap_{i=0}^{k} \varphi(\mathcal{U};b_i) \quad \subseteq \quad \mathcal{A}$$

we set $b_{k+1} = \cdots = b_n = b_k$ and stop as we already have the required parameters. Otherwise pick any element

$$a_{k+1} \quad \in \quad \bigcap_{i=0}^{k} \varphi(\mathcal{U};b_i) \smallsetminus \mathcal{A}$$

and some b_{k+1} such that

$$A \quad \subseteq \quad \varphi(\mathcal{U};b_{k+1}) \quad \subseteq \quad \mathcal{U} \smallsetminus \{a_1,\ldots,a_{k+1}\}.$$

Such a parameter b_{k+1} exists because $\mathcal{A}$ is approximable (apply Definition 3.8 with $A \cup \{a_1,\ldots,a_{k+1}\}$ for A). Note that at stage n we have constructed a ladder of length n for $\neg\varphi(x;z)$. In fact for all $h \in (n]$ and $k \in [n]$ we have

$$\neg\varphi(a_h;b_k) \quad \Leftrightarrow \quad h \leq k.$$

As $\varphi(x\,;z)$ has ladder index $n+1$, by Fact 3.4 it is not possible to further prolong this chain, hence

$$A \;\subseteq\; \bigcap_{i=0}^{n} \varphi(\mathcal{U}\,;b_i) \;\subseteq\; \mathcal{A}$$

as required. $\square$

We prove that the formula $\psi(x\,;z_0,\ldots,z_n)$ in the lemma above is itself stable, though with a larger ladder index.

3.11 Lemma Let $\psi_i(x\,;z)$, where $i=1,\ldots,m$, be formulas with ladder index n_i. Let

$$\varphi(x\,;z) \;=\; \bigwedge_{i=1}^{m} \varphi_i(x\,;z)$$

Then $\varphi(x\,;z)$ has ladder index $< R(n_1+1,\ldots,n_m+1)$, the Ramsey number for m-colorings.

Proof Suppose for a contradiction that there is a ladder $a_1,\ldots,a_n$ and $b_0,\ldots,b_n$, where $n = R(n_1+1,\ldots,n_m+1)-1$. Let C_i contains the pairs $\{h,k\}$ such that $\neg\varphi_i(a_h\,;b_k)$ and $0\le k<h\le n$. Then from ld we obtain

$$\bigcup_{i=1}^{m} C_i \;=\; \binom{[n]}{2}$$

By the definition of n, for some $i\in[m]$, there is a set H of cardinality n_i+1 such that $H^{(2)}\subseteq C_i$. Assume i=1 for definiteness.

Write $a'_1,\ldots,a'_{n_1}$ and $b'_0,\ldots,b'_{n_1}$ for the tuples obtained by restricting $a_1,\ldots,a_n$ and $b_0,\ldots,b_n$ to the indexes in H. These tuples witness that $\varphi_1(x\,;z)$ has ladder index at least n_1+1, which contradicts the assumption of the lemma. $\square$

To apply the proposition to the formula $\psi(x\,;z_0,\ldots,z_n)$ in the lemma above, let $z=z_0,\ldots,z_n$ and let $\varphi_i(x\,;z)=\varphi(x\,;z_i)$. That is, $\varphi_i(x\,;z)$ defines a relation between $\mathcal{U}$ and $\mathcal{V}^{n+1}$ that depends only on the i-th coordinate. The ladder index of every $\varphi_i(x\,;z)$ is the same as $\varphi(x\,;z_0)$ as a relation between $\mathcal{U}$ and $\mathcal{V}$.

3.12 Lemma If $\mathcal{A}$ is approximated from below by a stable formula $\varphi(x\,;z)$ then for some $b_0,\ldots,b_n\in\mathcal{V}$

$$\mathcal{A} \;=\; \bigcup_{i=0}^{n} \varphi(\mathcal{U}\,;b_i)$$

where $n+1$ is the ladder index of $\varphi(x\,;z)$.

Proof Let $A\subseteq\mathcal{U}$ be finite. To obtain the b_k, we construct a ladder by stages. Pick any b_0 be such that $\varphi(\mathcal{U}\,;b_0)\subseteq\mathcal{A}$. Now, suppose that $a_1,\ldots,a_k$ and $b_0,\ldots,b_k$ have been defined. If

$$\mathcal{A} \;\subseteq\; \bigcup_{i=1}^{k} \varphi(\mathcal{U}\,;b_i)$$

we set $b_{k+1}=\cdots=b_n=b_k$ and stop. As the construction guarantees the converse inclusion, we have the required parameters. Otherwise pick any element

$$a_{k+1} \;\in\; \mathcal{A} \smallsetminus \bigcup_{i=1}^{k} \varphi(\mathcal{U}\,;b_i)$$

and some b_{k+1} such that

$$\{a_1,\dots,a_{k+1}\} \;\subseteq\; \varphi(\mathcal{U}\,;b_{k+1}) \;\subseteq\; \mathcal{A}.$$

Such a parameter b_{k+1} exists because $\mathcal{A}$ is approximable from below. Note that at stage n we have constructed a ladder of length n for $\varphi(x\,;z)$. In fact for all $h \in (n]$ and $k \in [n]$ we have

$$\neg\varphi(a_h\,;b_k) \;\Leftrightarrow\; h \leq k.$$

As $\varphi(x\,;z)$ has ladder index $n+1$, it is not possible to prolonge this chain, hence

$$\mathcal{A} \;=\; \bigcup_{i=0}^{n} \varphi(\mathcal{U}\,;b_i)$$

as required. □

The main theorem of this section is an immediate corollary is the three lemmas above.

3.13 Theorem Let $\varphi(x\,;z)$ be stable. Then there are m,n such every set $\mathcal{A} \subseteq \mathcal{U}$ approximable by $\varphi(x\,;z)$ is definable by $\psi(x\,;\bar{b})$, where

$$\psi(x\,;\bar{z}) \;=\; \bigvee_{j=0}^{m} \bigwedge_{i=0}^{n} \varphi(z\,;z_{i,j})$$

and $\bar{b} \in \mathcal{V}^{n\times m}$ is some $n\times m$-tuple of parameters. The numbers m,n only depend on the ladder index of $\varphi(x\,;z)$. □

We conclude this section with the sketch of an argument that shows that stable formulas have finite 2-rank, that is, 1⇒2 in Theorem 3.7.

First of all note that Theorem 3.13 remains valid, with the same formula $\psi(x\,;\bar{z})$, when $\mathcal{U}$ is replaced by a subset $\mathcal{U}' \subseteq \mathcal{U}$. In fact, the ladder index of $\varphi(x\,;z)$ does not decrease when moving to a subset.

Assume that the 2-rank of $\psi(x\,;\bar{z})$ is infinite. We claim that, with the appropriate $\mathcal{U}$ and $\mathcal{V}$, there is a branching tree of height ω, say $\{a_s : s \in 2^{<\omega}\}$.

Model theoretically, the claim is immediate. A low-tech combinatorial proof is also not difficult but it requires some handwork which we entrust to the reader.

For $s \in 2^{\omega}$, let $\mathcal{U}_s = \{a_{s\restriction n} : n < \omega\} \subseteq \mathcal{U}$. Let $\mathcal{A}_s = \{a_{s\restriction n} : n < \omega,\ s_n = 1\} \subseteq \mathcal{U}_s$. Then $\mathcal{A}_s$ is approximable with respect to the collection $\varphi(\mathcal{U}_s\,;b)_{b\in\mathcal{V}}$. From the theorem above, we obtain $\mathcal{A}_s = \psi(\mathcal{U}_s\,;\bar{b}_s)$ for some $\bar{b}_s \in \mathcal{V}^{n\times m}$, where the formula $\psi(x\,;\bar{z})$ does non depend on s.

For $s \neq s'$, let n be least such that $s_n \neq s'_n$. Then $a_{s\restriction n} = a_{s'\restriction n}$ belongs to $\mathcal{U}_s \cap \mathcal{U}_{s'}$ and to $\mathcal{A}_s \triangle \mathcal{A}_{s'}$. Therefore $\bar{b}_s \neq \bar{b}_{s'}$, hence there is a continuum of tuples $\bar{b}_s$.

On the other hand, without loss of generality we can assume that $\mathcal{V}$ is a countable set. A contradiction.

Chapter 4

Vapnik-Chervonenkis theory

1 The Vapnik-Chervonenkis dimension

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

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

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

The Vapnik-Chervonenkis dimension of $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$, or of $\varphi(x\,;z)$, abbreviated by VC-dimension, is the maximal cardinality of a finite set $A \subseteq \mathcal{U}$ that is shattered by $\varphi(x\,;z)$. Equivalently, it is the maximal k such that $\pi_\varphi(k) = 2^k$. If such a maximum does not exist, we say that the VC-dimension is infinite or that $\varphi(x\,;z)$ has IP (the independence property). Otherwise, we say that $\varphi(x\,;z)$ has NIP (not the independence property). We may also say: *is* IP, or *is* NIP.

As $\mathcal{U}$ and $\mathcal{V}$ are usually clear from the context, we may say VC-dimension of $\varphi(x\,;z)$ for the VC-dimension of $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$.

4.1 Example If $\varphi(x\,;z)$ is either $\top$ or $\bot$, then it shatters only the empty set, therefore it has VC-dimension 0. □

4.2 Example If $\varphi(x\,;z)$ has ladder dimension n then it has VC-dimension at most n. Hence stable formulas are NIP. □

4.3 Example If $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$ is a non trivial chain of sets, then its VC-dimension is 1. □

4.4 Example Let $\mathcal{U} = \mathbb{R}$ and $\mathcal{V} = \mathbb{R}^2$. Let $\varphi(x\,;z_1,z_2)$ be the formula $z_1 < x < z_2$. Then its VC-dimension 2. □

4.5 Example Let $\mathcal{U} = \mathcal{V} = \mathbb{R}^2$. Let $\varphi(x_1,x_2\,;z_1,z_2)$ be the formula $y < z_1 \cdot x + z_2$. Then its VC-dimension 3 (by Radon's Theorem). □

4.6 Example If $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$ is the set of all subsets of $\mathcal{U}$ of cardinality $\leq k$. Then its VC-dimension is k and

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

Incidentally, we note that this is also the shatter function of the collection of all subsets of $\mathcal{U}$ of cardinality exactly k. In fact we always assume $\mathcal{U}$ is infinite (or at least very large, in this case, size $\geq 2k$ suffices). $\square$

The VC-dimension of $\varphi(x\,;z)^{\mathrm{op}}$ is called the dual VC-dimension or VC-codimension of $\varphi(x\,;z)$.

4.7 Proposition If $\varphi(x\,;z)$ has VC-dimension k, then its VC-codimension is $\leq 2^{k+1}-1$.

Proof Suppose that the VC-dimension of $\varphi(x\,;z)^{\mathrm{op}}$ is $\geq 2^{k+1}$. We prove that the VC-dimension of $\varphi(x\,;z)$ is at least $k+1$. Let $B = \big\{b_I \ : \ I \subseteq [k+1]\big\}$ be a set of cardinality 2^{k+1} shattered by $\varphi(x\,;z)^{\mathrm{op}}$. That is, for every $\mathcal{J} \subseteq \mathcal{P}[k+1]$ there is $a_{\mathcal{J}}$ such that

$$\varphi(a_{\mathcal{J}}, b_I) \ \Leftrightarrow\ I \in \mathcal{J} \qquad \text{for all } I \subseteq [k+1]$$

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

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

That is, $\varphi(x\,;z)$ shatters $A = \big\{a_i \ : \ i \in [k+1]\big\}$. $\square$

We prove that the bound in the proposition above is optimal.

4.8 Proposition For every k, there is a formula $\varphi(x\,;z)$ that has VC-dimension k and VC-codimension $2^{k+1}-1$.

Proof Let k be given. We claim that there is a formula $\varphi(x\,;z)$ with VC-dimension $2^{k+1}-1$ and VC-codimension k. As the dual of the dual is the primal, this claim is equivalent to the proposition. Let $\mathcal{U}$ an infinite set, let $\mathcal{V} = \mathcal{P}(\mathcal{U})$. Fix an equivalence relation on $\mathcal{U}$ with $2^{k+1}-1$ many classes. We say that $b \in \mathcal{V}$ is compatible if it is union of equivalence classes. Define

$$\varphi(x\,;z) \ = \ x \in z \ \wedge\ z \text{ is compatible}$$

(Recall that in the preliminaries to this notes, we have agreed that $\mathcal{U}$ and $\mathcal{V}$ are some fixed infinite sets. Without this constraint we could have taken as $\mathcal{U}$ a set of cardinality $2^{k+1}-1$; defined $\varphi(x\,;z) = x \in z$; and forgot the equivalence relation altogether.)

Clearly, $\varphi(x\,;z)$ has VC-dimension $2^{k+1}-1$. Note that the dimension of $\varphi(x\,;z)^{\mathrm{op}}$ is at least k. Otherwise $\varphi(x\,;z)$ would have dimension $\leq 2^k-1$ by Proposition 4.7.

So, it suffices to prove that the dimension of $\varphi(x\,;z)^{\mathrm{op}}$ is exactly k. Assume for a contradiction that some $B = \big\{b_i \ : \ i \in [k+1]\big\} \subseteq \mathcal{V}$ is shattered by $\varphi(x\,;z)^{\mathrm{op}}$. Then there is some set $A = \big\{a_I : I \subseteq [k+1]\big\} \subseteq \mathcal{U}$ such that

$$\varphi(a_I, b_i) \ \Leftrightarrow\ i \in I$$

The $b_i \in B$ are necessarily compatible sets, therefore any two distinct $a_I \in A$ are non equivalent. But this is impossible by cardinality reasons. $\square$

4.9 Exercise Let $\varphi(x\,;z)$ have VC-dimension k. Assume that there is no $A \subseteq \mathcal{U}$ of cardinality $\leq k$ such that $\varphi(a\,;\mathcal{V})_{a\in A}$ covers $\mathcal{V}$. Prove that $\varphi(x\,;z)$ has VC-codimension $\leq 2^{k+1}-2$. Prove that the bound is optimal. Hint: let $\mathcal{U}$ be any infinite set $\mathcal{V} = \mathcal{U}^{(\leq m)}$, where $m = 2^{k+1}-2$. Let $\varphi(x\,;z) = x \in z$. Prove that $\varphi(x\,;z)$ has VC-codimension k. □

2 The Sauer-Shelah lemma

According to Gil Kalai in [8], 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 [13], Sauer [12], and Vapnik-Chervonenkis [14] 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.

4.10 Sauer-Shelah Lemma If $\varphi(x\,;z)$ has VC-dimension k then for every $n \geq k$

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

□

The set system presented in Example 4.6 shows that the bound is optimal.

An alternative proof of the Sauer-Shelah Lemma derives it as corollary of a lemma by Alain Pajor [11].

4.11 Pajor's Lemma Let $A \subseteq \mathcal{U}$ be finite.

$$\big|\varphi(A,b)_{b\in\mathcal{V}}\big| \;\leq\; \Big|\{C \subseteq A \;:\; C \text{ is shattered by } \varphi(x\,;z)\}\Big|.$$

Proof If A is empty then $|\varphi(A,b)_{b\in\mathcal{V}}| = 1$ and $\varnothing$ is the only subset of A that φ shatters, so the inequality holds trivially. Otherwise, pick an $a \in A$ and assume the lemma holds for $A' = A \smallsetminus \{a\}$. Define

$$\psi(x\,;y) \;=\; \varphi(x\,;y)\;\wedge\neg\varphi(a\,;y)\;\wedge\;\exists y'\Big[\varphi(a\,;y')\;\wedge\;\varphi(A'\,;y') = \varphi(A'\,;y)\Big].$$

Notice that

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

as the two sets in the r.h.s. are disjoint

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

By induction hypothesis,

$$\big|\varphi(A',b)_{b\in\mathcal{V}}\big| \;\leq\; \Big|\{C \subseteq A' \;:\; C \text{ is shattered by } \varphi(x\,;z)\}\Big| \tag{1}$$

and

$$\big|\psi(A',b)_{b\in\mathcal{V}}\big| \;\leq\; \Big|\{C \subseteq A' \;:\; C \text{ is shattered by } \psi(x\,;z)\}\Big|$$

$$= \left|\{C \subseteq A' \,:\, C \cup \{a\} \text{ is shattered by } \varphi(x\,;z)\}\right|. \tag{2}$$

In fact, $C \subseteq A'$ is shattered by $\psi(x\,;y)$ if an only if $C \cup \{a\}$ it is shattered by $\varphi(x\,;y)$. Clearly,

$$(1) + (2) \;=\; \left|\{C \subseteq A \,:\, C \text{ is shattered by } \varphi(x\,;z)\}\right|,$$

so the lemma follows. $\square$

Proof of the Sauer-Shelah Lemma Assume $\varphi(x\,;z)$ has VC-dimension k and let $n \geq k$. Then

$$\begin{aligned}
\pi_\varphi(n) &= \max_{|A|=n} \left|\varphi(A,b)_{b\in\mathcal{V}}\right| \\
\pi_\varphi(n) &\leq \max_{|A|=n} \left|\{C \subseteq A \,:\, C \text{ shattered by } \varphi(x\,;z)\}\right| && \text{by Pajor's Lemma} \\
&\leq \sum_{i=0}^{k} \binom{n}{i} && \text{because } \varphi(x\,;z) \text{ has VC-dimension } k
\end{aligned}$$

$\square$

We write $f(n) = O(g(n))$ if there is a constant C such that $|f(n)| \leq Cg(n)$ holds for all (sufficiently large) n.

The VC-density of $\varphi(x\,;z)$ is the infimum over all real number r such that $\pi_\varphi(n) = O(n^r)$. It is infinite if no such r exist. The dual VC-density is defined accordingly.

By the Sauer-Shelah lemma the VC-density is at most as large as the VC-dimension. It could be smaller, however it is usually rather difficult to compute.

We conclude this section with a couple of inequalities that is useful to have at hand.

$$\begin{aligned}
\sum_{i=0}^{k} \binom{n}{i} &= \sum_{i=0}^{k} \frac{n!}{i!\,(n-i)!} \\
&\leq \sum_{i=0}^{k} \frac{n^i}{i!} \\
&\leq \sum_{i=0}^{k} \frac{n^i\,k!}{i!(k-i)!} \\
&= \sum_{i=0}^{k} n^i \binom{k}{i} \\
&= (n+1)^k && \text{by the binomial theorem.}
\end{aligned}$$

There is a second bound, which is better when $k \geq 3$ and holds for $n > k$

$$\begin{aligned}
\sum_{i=0}^{k} \binom{n}{i} &\leq \left(\frac{n}{k}\right)^k \sum_{i=0}^{k} \left(\frac{k}{n}\right)^i \binom{n}{i} && \text{because } \frac{k}{n} < 1 \\
&\leq \left(\frac{n}{k}\right)^k \sum_{i=0}^{n} \left(\frac{k}{n}\right)^i \binom{n}{i} \\
&= \left(\frac{n}{k}\right)^k \left(1 + \frac{k}{n}\right)^n && \text{by the binomial theorem} \\
&\leq \left(\frac{n\,e}{k}\right)^k && \text{where } e \text{ is the base of the natural logarithm.}
\end{aligned}$$

Chapter 5

Law(s) of large numbers

Quoting from some unpublished notes by Carlos C. Rodríguez

> What is a Law of Large Numbers? I am glad you asked! The Laws of Large Numbers, or LLNs for short, come in three basic flavors: Weak, Strong and Uniform. They all state that the observed frequencies of events tend to approach the actual probabilities as the number of observations increases. Saying it in another way, the LLNs show that under certain conditions, we can asymptotically learn the probabilities of events from their observed frequencies. To add some drama we could say that if God is not cheating and S/he doesnt change the initial standard probabilistic model too much then, in principle, we (or other machines, or even the universe as a whole) could eventually find out the Truth, the whole Truth, and nothing but the Truth.
>
> Bull! The Devil, is in the details.
>
> I suspect that for reasons not too different in spirit to the ones above, famous minds of the past took the slippery slope of defining probabilities as the limits of relative frequencies. They became known as "frequentist". They wrote books and indoctrinated generations of confused students.

1 Inequalities

Throughout this and the next section we work with a given probability space $\mathcal{U}, \mathbb{P}$. For simplicity, the following two propositions are proved for discrete $\mathbb{P}$, but they are easily seen to hold in general.

5.1 Definition A function $f : \mathbb{R} \to \mathbb{R}$ is convex if for every tuples of real numbers p_i and x_i such that

$$\sum_{i=1}^{n} p_i \;=\; 1$$

we have

$$f\Big(\sum_{i=1}^{n} p_i x_i\Big) \;\leq\; \sum_{i=1}^{n} p_i f(x_i). \qquad \square$$

Note that, though the definition is usually given with $n = 2$, the general property above follows easily.

5.2 Jensen's Inequality Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function. Then

$$f\Big(\mathbb{E}[X]\Big) \;\leq\; \mathbb{E}\Big[f(X)\Big]$$

Proof For simplicity, assume that the sample space $\mathcal{U}$ is finite. Then the claim is obvious from the definition. $\square$

The following is arguably the most basic inequality in probability theory. Although it is almost trivial, it will be required several times in this chapter.

5.3 Markov's Inequality Let X be a nonnegative random variable with finite mean. Then for every $\varepsilon > 0$

$$\mathbb{P}\Big(X \geq \varepsilon\Big) \;\leq\; \frac{\mathbb{E}[X]}{\varepsilon}$$

Proof For simplicity, assume that the sample space $\mathcal{U}$ is finite. (The theorem holds in general, but we only need the finite case.) Define $A = \{a \in \mathcal{U} \,:\, X(a) \geq \varepsilon\}$.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{a\in\mathcal{U}} \mathbb{P}(a)\, X(a) \\
&= \sum_{a\in A} \mathbb{P}(a)\, X(a) + \sum_{a\notin A} \mathbb{P}(a)\, X(a) \\
&\geq \sum_{a\in A} \mathbb{P}(a)\, X(a) \\
&\geq \varepsilon \sum_{a\in A} \mathbb{P}(a) \\
&= \varepsilon\, \mathbb{P}(X \geq \varepsilon)
\end{aligned}$$

$\square$

5.4 Corollary Let X be a nonnegative random variable. If $\mathbb{E}\big[X^k\big]$ exists, then for every $\varepsilon > 0$

$$\mathbb{P}\Big(X \geq \varepsilon\Big) \;\leq\; \frac{\mathbb{E}[X^k]}{\varepsilon^k}$$

Proof By Markov's inequality, since $\mathbb{P}\big(X \geq \varepsilon\big) = \mathbb{P}\big(X^k \geq \varepsilon^k\big)$. $\square$

Chebyshev's inequality (a.k.a. Chebysheff, Chebyshov, Tschebyscheff, Tschebycheff) is a special case of the corollary above.

5.5 Chebyshev's Inequality Let X be a random variable with finite mean and variance. Then for every $\varepsilon > 0$

$$\mathbb{P}\Big(\big|X - \mathbb{E}[X]\big| \geq \varepsilon\Big) \;\leq\; \frac{\mathrm{Var}[X]}{\varepsilon^2}$$

$\square$

To obtain exponential bounds, we frequently apply the following trick.

5.6 Chernoff's method Let X be a random variable with finite mean. Then for every $t > 0$

$$\mathbb{P}\big(X \geq \varepsilon\big) \;\leq\; e^{-t\varepsilon}\, \mathbb{E}\big[e^{tX}\big]$$

Proof For every $t > 0$

$$\begin{aligned} \mathbb{P}(X \geq \varepsilon) &= \mathbb{P}(e^{tX} \geq e^{t\varepsilon}) && \text{because } e^{tx} \text{ is increasing} \\ &\leq e^{-t\varepsilon}\,\mathbb{E}[e^{tX}], \end{aligned}$$

by Markov's inequality, which me may apply since e^{tX} is always positive. □

5.7 Hoeffding's lemma Let X be a bounded random variable, say $a \leq X \leq b$. Let $\mathbb{E}[X] = \mu$ and $d = b - a$. Then

$$\mathbb{E}\left[e^{t(X-\mu)}\right] \leq \exp\left(\frac{t^2d^2}{8}\right).$$

Proof For clarity, assume $\mu = 0$. The general result follows easily from this special case by centralization. Recall that, by convexity, for every $x \in [a, b]$

$$e^{tx} \leq \frac{x-a}{d}\,e^{tb} + \frac{b-x}{d}\,e^{ta}$$

Then

$$e^{tX} \leq \frac{X-a}{d}\,e^{tb} + \frac{b-X}{d}\,e^{ta}$$

By the linearity of expectation,

$$\begin{aligned} \mathbb{E}\left[e^{tX}\right] &\leq \frac{b\,e^{ta} - a\,e^{tb}}{d} \\ \log\mathbb{E}\left[e^{tX}\right] &\leq \log\frac{b\,e^{ta} - a\,e^{tb}}{d} \end{aligned}$$

taking the Taylor series expansion of the r.h.s. at $t = 0$ (the first and second derivatives vanish at 0; the second derivative is always $\leq d^2/4$) we obtain

$$\log\mathbb{E}\left[e^{tX}\right] \leq \frac{t^2d^2}{8}.$$

□

5.8 Hoeffding's Inequality Let $X_1, \ldots, X_n$ be independent random variables with bounded range, say $a \leq X_i \leq b$. Define $d = b - a$.

$$M = \sum_{i=1}^{n}\left(X_i - \mathbb{E}[X_i]\right)$$

Then for every $\varepsilon > 0$

$$\begin{aligned} \mathbb{P}\left(M \geq \varepsilon\right) &\leq \exp\left(-\frac{2\varepsilon^2}{nd^2}\right), \\ \mathbb{P}\left(M \leq -\varepsilon\right) &\leq \exp\left(-\frac{2\varepsilon^2}{nd^2}\right). \end{aligned}$$

Clearly, the two inequalities above imply the following

$$\mathbb{P}\left(|M| \geq \varepsilon\right) \leq 2\exp\left(-\frac{2\varepsilon^2}{nd^2}\right).$$

Proof Define $\mathbb{E}[X_i] = \mu_i$. Let $t > 0$ be arbitrary.

$$\begin{aligned} \mathbb{P}\left(M \geq \varepsilon\right) &\leq e^{-t\varepsilon}\,\mathbb{E}\left[e^{tM}\right] && \text{by Chernoff's method (5.6)} \\ &= e^{-t\varepsilon}\prod_{i=1}^{n}\mathbb{E}\left[e^{t\,(X_i-\mu_i)}\right] && \text{by independence.} \end{aligned}$$

$$\begin{aligned} &\le e^{-t\varepsilon}\prod_{i=1}^{n}\exp\left(\frac{t^2d^2}{8}\right) && \text{by Hoeffding's Lemma (5.7).}\\ &= \exp\left(\frac{n\,t^2d^2}{8}-t\varepsilon\right) \end{aligned}$$

Now substitute $4\varepsilon/nd^2$ for t. □

We prove Hoeffding's lemma with a slightly weaker bound (2 for 8). The purpose is to present two clever tricks *ghost sample* and *symmetrization* which in the following section is applied in a more complex setting.

First we need the following lemma. A random sign variable (a.k.a. Rademacher random variable) is a random variable $\sigma \in \{-1,1\}$ with mean 0.

5.9 Lemma Let σ be a random sign variable. Then for every t

$$\mathbb{E}\left[e^{t\sigma}\right] \le e^{t^2/2}$$

Proof Replace e^x with its Taylor expansion around $x = 0$

$$\begin{aligned} \mathbb{E}\left[e^{t\sigma}\right] &= \sum_{i=0}^{\infty}\frac{t^i\mathbb{E}\left[\sigma^i\right]}{i!}\\ &= \sum_{i=0}^{\infty}\frac{t^{2i}}{(2i)!} && \text{since } \mathbb{E}\left[\sigma^i\right] = \begin{cases}1 & i \text{ even}\\ 0 & i \text{ odd}\end{cases}\\ &= \sum_{i=0}^{\infty}\frac{(t^2/2)^i}{i!}\\ &= e^{t^2/2}. \end{aligned}$$

□

5.10 Second proof of Hoeffding's Lemma Recall that Hoeffding's Lemma claims that, if $a \le X \le b$, then

$$\mathbb{E}\left[e^{t(X-\mu)}\right] \le \exp\left(\frac{t^2d^2}{8}\right),$$

where $\mathbb{E}[X] = \mu$ and $d = b - a$. Here we prove a marginally weaker bound (2 in place of 8).

Let X' be an independent copy of X (a.k.a. ghost sample). In particular $\mu = \mathbb{E}(X')$. Then

$$\begin{aligned} \mathbb{E}\left[e^{t(X-\mu)}\right] &= \mathbb{E}\left[e^{t(X-\mathbb{E}[X'])}\right]\\ &\le \mathbb{E}\left[\mathbb{E}\left[e^{t(X-X')}\mid X\right]\right] && \text{by Jensen's inequality}\\ &\le \mathbb{E}\left[e^{t(X-X')}\right] \end{aligned}$$

Let σ be a random sign variable independent of X, X'. Then $\sigma(X-X')$ has the same distribution of $X - X'$.

$$\begin{aligned} &= \mathbb{E}\left[e^{t\sigma(X-X')}\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[e^{t\sigma(X-X')}\mid X, X'\right]\right] \end{aligned}$$

$$\leq \ \mathbb{E}\left[e^{t^2(X-X')^2/2}\right] \qquad \text{by Lemma 5.9}$$

$$\leq \ e^{t^2d^2/2} \qquad \text{because } |X-X'| \leq d.$$

This yields the bound above with 2 in place of 8. □

2 Two Weak Laws of Large Numbers

A sample s is a sequence $s_1, \ldots, s_n$ of elements of $\mathcal{U}$. Its length $|s| = n$ is also called size or dimension. We write range(s) for the set $\{s_1, \ldots, s_n\}$. Note that this set may have cardinality $< n$.

To a sample s of size n we associate a finite probability measure on the subsets of $\mathcal{U}$ namely, for any event $A \subseteq \mathcal{U}$, we define the empirical frequence of A given s

$$\mathrm{Fr}\,(s, A) \quad = \quad \frac{1}{n} \cdot \big|\{i \ : \ s_i \in A\}\big|.$$

It is convenient to rewrite it using indicator functions

$$\mathrm{Fr}(s, A) \quad = \quad \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{s_i \in A}.$$

We are interested in the *existence* of samples that approximate the probability within ε. Suppose that, for a given event A, we can prove that

$$(1) \qquad \mathbb{P}\Big(s \in \mathcal{U}^n : \big|\,\mathrm{Fr}(s, A) - \mathbb{P}(A)\big| \geq \varepsilon\Big) \quad \leq \quad \text{some\_bound}(\varepsilon, n)$$

and that, for n large enough, some_bound$(\varepsilon, n) < 1$. Then a sample of size $\leq n$ that approximate the probability within ε is guaranteed to exit.

Random variables are convenient formalism to discuss these probabilities. We say random element of $\mathcal{U}$ for a random variables S such that $\mathbb{P}(S \in A) = \mathbb{P}(A)$ for every $A \subseteq \mathcal{U}$. A random sample from $\mathcal{U}$ is a tuple $S = S_1, \ldots, S_n$ of independent random elements of $\mathcal{U}$. Then $\mathbb{I}_{S_i \in A} = \mathbb{I}_A \circ S_i$ is a Bernoulli random variable with probability of success $\mathbb{P}(A)$ and

$$\mathrm{Fr}(S, A) \quad = \quad \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{S_i \in A}$$

is (up to the factor $1/n$) a binomial random variable. This random variable is used below to find the bound in (1), in fact

$$\mathbb{P}\Big(\big|\,\mathrm{Fr}(S, A) - \mathbb{P}(A)\big| \geq \varepsilon\Big)$$

equals the probability in (1) but is easier to estimate.

5.11 Weak Law of Large Numbers For every event $A \subseteq \mathcal{U}$ and every tuple $S = S_1, \ldots, S_n$ of independent random elements of $\mathcal{U}$

$$\mathbb{P}\Big(\big|\,\mathrm{Fr}(S, A) - \mathbb{P}(A)\big| \geq \varepsilon\Big) \quad \leq \quad \frac{1}{n\varepsilon^2}.$$

Proof The random variable $\mathrm{Fr}(S, A)$ has expected value $\mathbb{P}(A)$ and variance $\leq 1/n$. By Chebyshev's inequality we obtain

$$\mathbb{P}\Big(\big|\,\mathrm{Fr}(S,A) - \mathbb{P}(A)\big| \geq \varepsilon\Big) \;\leq\; \frac{1}{n\varepsilon^2}$$

which proves the theorem. ☐

Sometime we are interested in the minimal size of a sample that approximates the probability up to a given ε.

5.12 Corollary Assume $\mathcal{U}$ is finite (of arbitrary cardinality, tough). For every $A \subseteq \mathcal{U}$ and every $\varepsilon > 0$ there is a sample s of size

$$|s| \;=\; \left\lfloor \frac{1}{\varepsilon^2} + 1 \right\rfloor$$

such that

$$\Big|\,\mathrm{Fr}(s,A) - \mathbb{P}(A)\Big| \;<\; \varepsilon.$$

Proof By the Weak Law of Large Numbers above, a sample of size n exists if

$$\frac{1}{n\varepsilon^2} \;<\; 1$$

☐

In the following section we need a better bound for the Weak Law of Large Numbers. This is obtained with a similar proof.

5.13 Weak Law of Large Numbers (with exponential bound) For every event $A \subseteq \mathcal{U}$ and every tuple $S = S_1, \ldots, S_n$ of independent random elements of $\mathcal{U}$

$$\mathbb{P}\Big(\big|\,\mathrm{Fr}(S,A) - \mathbb{P}(A)\big| \geq \varepsilon\Big) \;\leq\; 2\,e^{-2n\varepsilon^2}.$$

Proof Define

$$M \;=\; \sum_{i=1}^{n} \Big(\mathbb{I}_{S_i \in A} - \mathbb{E}[\mathbb{I}_{S_i \in A}]\Big)$$

As $\mathbb{E}[\mathbb{I}_{S_i \in A}] = \mathbb{P}(A)$, the inequality we have to prove can be rewritten as

$$\mathbb{P}\Big(|M| \geq n\varepsilon\Big) \;\leq\; 2\,e^{-2n\varepsilon^2}$$

and this follows immediately from Hoeffding inequality. ☐

Using the exponential bounds above, we can improve (by a constant factor) the size of the minimal sample size that approximates the probability obtained in Corollary 5.12.

5.14 Corollary For every $A \subseteq \mathcal{U}$ and every $\varepsilon > 0$ there is a sample s of size n where

$$n \;=\; \left\lfloor \frac{\log 2}{2\varepsilon^2} + 1 \right\rfloor$$

such that

$$\Big|\,\mathrm{Fr}(s,A) - \mathbb{P}(A)\Big| \;<\; \varepsilon.$$

☐

3 A Uniform Law of Large Numbers

Throughout this section we work with a fixed family of definable subsets $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$ that are events of the sample space $\mathcal{U},\mathbb{P}$. It is convenient to introduce some abbreviations

$$
\begin{aligned}
\mathbb{P}(b) &= \mathbb{P}\Big(\varphi(\mathcal{U}\,;b)\Big)\\
\operatorname{Fr}(s,b) &= \operatorname{Fr}\Big(s,\varphi(\mathcal{U}\,;b)\Big)\\
\mathbb{I}_{s,b} &= \mathbb{I}_{\varphi(s;b)}
\end{aligned}
$$

An ε-approximation is a sample s such that

$$
\Big|\operatorname{Fr}(s,b)-\mathbb{P}(b)\Big| \;<\; \varepsilon \qquad\qquad \text{for every } b\in\mathcal{V}.
$$

We are interested in estimating the minimal size of an ε-approximation.

The main theorem of this section is this famous result of Vapnik-Chervonenkis [14].

5.15 Vapnik-Chervonenkis Inequality Let $\pi_\varphi(n)$ be the shatter function of $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$. Let $S=S_1,\dots,S_n$ be a random sample from $\mathcal{U}$. Then, for every $b\in\mathcal{V}$

$$
\mathbb{P}\Big(\big|\operatorname{Fr}(S,b)-\mathbb{P}(b)\big|\geq\varepsilon\Big) \;\leq\; 6\,\pi_\varphi(n)\exp\Big(-\frac{n\varepsilon^2}{32}\Big).
$$

N.B. Some technical measure-theoretical assumptions are necessary when $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$ is uncountable. These have been omitted in the statement above, and will be discussed below.

Proof Let $S'=S'_1,\dots,S'_n$ be an independent copy of S. Then, by the triangular inequality

$$
\mathbb{P}\Big(\big|\operatorname{Fr}(S,b)-\mathbb{P}(b)\big|\geq\varepsilon\Big) \;\leq\; \mathbb{P}\Big(\big|\operatorname{Fr}(S,b)-\operatorname{Fr}(S',b)\big|\geq\frac{\varepsilon}{2}\Big) \;+\; (*)
$$

where,

$$
(*) \;=\; \mathbb{P}\Big(\big|\operatorname{Fr}(S',b)-\mathbb{P}(b)\big|\geq\frac{\varepsilon}{2}\Big)
$$

which, by the Weak Law of Large Numbers 5.13,

$$
\leq\; 2\,e^{-n\varepsilon^2/2}
$$

Let $\sigma=\sigma_1,\dots,\sigma_n$ be a tuple of independent sign random variables. Then

$$
\begin{aligned}
\mathbb{P}\Big(\big|\operatorname{Fr}(S,b)-\operatorname{Fr}(S',b)\big|\geq\frac{\varepsilon}{2}\Big) &= \mathbb{P}\Big(\Big|\sum_{i=1}^{n}\mathbb{I}_{S_i,b}-\mathbb{I}_{S'_i,b}\Big|\geq\frac{n\varepsilon}{2}\Big)\\
&= \mathbb{P}\Big(\Big|\sum_{i=1}^{n}\sigma_i\big(\mathbb{I}_{S_i,b}-\mathbb{I}_{S'_i,b}\big)\Big|\geq\frac{n\varepsilon}{2}\Big)
\end{aligned}
$$

Then, again by the triangular inequality

$$
\leq\; 2\,\mathbb{P}\Big(\Big|\sum_{i=1}^{n}\sigma_i\,\mathbb{I}_{S_i,b}\Big|\geq\frac{n\varepsilon}{4}\Big)
$$

Putting together the inequalities above we obtain

$$\mathbb{P}\Big(\sup_{b\in\mathcal{V}}\big|\operatorname{Fr}(S,b)-\mathbb{P}(b)\big|\geq\varepsilon\Big) \;\leq\; 2\,\mathbb{P}\Big(\sup_{b\in\mathcal{V}}\Big|\sum_{i=1}^{n}\sigma_i\,\mathbb{I}_{S_i,b}\Big|\geq\frac{n\varepsilon}{4}\Big) \;+\; 2\,e^{-n\varepsilon^2/2}$$

Let $s = s_1,\ldots,s_n$ be a possible realization of S.

(1) $$\mathbb{P}\Big(\sup_{b\in\mathcal{V}}\Big|\sum_{i=1}^{n}\sigma_i\,\mathbb{I}_{S_i,b}\Big|\geq\frac{n\varepsilon}{4}\Big) \;=\; \mathbb{P}\Big(\sup_{b\in\mathcal{V}}\Big|\sum_{i=1}^{n}\sigma_i\,\mathbb{I}_{S_i,b}\Big|\geq\frac{n\varepsilon}{4}\;\Big|\;S=s\Big)$$

Finally, note that the r.h.s. of (1) only depends on $\varphi(\{s_1,\ldots,s_n\}\,;b)$. Hence the $\sup_{b\in\mathcal{V}}(\cdot)$ on the r.h.s. is actually a maximum among $m=\pi_\varphi(n)$ events. Say, we can choose $b_1,\ldots,b_m\in\mathcal{V}$ such that

(2) $$=\;\mathbb{P}\Big(\max_{j\leq m}\Big|\sum_{i=1}^{n}\sigma_i\,\mathbb{I}_{s_i,b_j}\Big|\geq\frac{n\varepsilon}{4}\Big)$$

Note that, in general, for any real random variables $X_1,\ldots,X_m$ we have

$$\begin{aligned}
\mathbb{P}\Big(\max_{i\leq m}X_i\geq\varepsilon\Big) \;&=\; \mathbb{P}\Big(\bigcup_{i=0}^{m}X_i\geq\varepsilon\Big)\\
&\leq\; \sum_{i=1}^{m}\mathbb{P}\big(X_i\geq\varepsilon\big)
\end{aligned}$$

Hence, continuing from (2) we obtain

$$\begin{aligned}
&\leq\; \sum_{j=1}^{m}\mathbb{P}\Big(\Big|\sum_{i=1}^{n}\sigma_i\,\mathbb{I}_{s_i,b_j}\Big|\geq\frac{n\varepsilon}{4}\Big).\\
&\leq\; 2\,\pi_\varphi(n)\,\exp\Big(-\frac{n\varepsilon^2}{32}\Big),
\end{aligned}$$

where the last inequality is obtained from Hoeffding's Inequality 5.8. In fact, Hoeffding's Inequality, applied to $X_i=\sigma_i\,\mathbb{I}_{s_i,b}$ with $n\varepsilon/4$ for ε, yields

$$\mathbb{P}\Big(\Big|\sum_{i=1}^{n}\sigma_i\,\mathbb{I}_{s_i,b}\Big|\geq\frac{n\varepsilon}{4}\Big) \;\leq\; 2\,\exp\Big(-\frac{n\varepsilon^2}{32}\Big).$$

Unsurprisingly, the bound does not depends on b. Finally, proceeding from (1) we obtain

$$\begin{aligned}
&\leq\; 4\pi_\varphi(n)\exp\Big(-\frac{n\varepsilon^2}{32}\Big) \;+\; 2\,\exp\Big(-\frac{n\varepsilon^2}{2}\Big)\\
&\leq\; 6\pi_\varphi(n)\exp\Big(-\frac{n\varepsilon^2}{32}\Big),
\end{aligned}$$

which finally proves the theorem. □

5.16 Corollary Let $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$ have finite VC-dimension. For every every $\varepsilon>0$ there is a finite sample s such that

$$\Big|\operatorname{Fr}(s,b)-\mathbb{P}(b)\Big| \;<\; \varepsilon \qquad \text{for every } b\in\mathcal{V}.$$

Proof It suffices to require that $n=|s|=|S|$ is large enough to guarantee

$$\mathbb{P}\Big(\big|\operatorname{Fr}(S,b)-\mathbb{P}(b)\big|\geq\varepsilon\Big) \;<\; 1 \qquad \text{for every } b\in\mathcal{V}.$$

By the Vapnik-Chervonenkis inequality 5.15, it suffices that

$$(3) \qquad 6\,\pi_\varphi(n)\,\exp\Big(-\frac{n\varepsilon^2}{32}\Big) \;<\; 1$$

By the Sauer-Shelah Lemma 4.10, $\pi_\varphi(n)$ grows polynomially. Hence the inequality holds for n large enough. □

The corollary above is sufficient for our intended applications. For completeness, the following proposition gives an explicit bound.

5.17 Proposition There is a sample s as in the corollary above of size

$$n \;\le\; c\,\frac{k}{\varepsilon^2}\log\frac{k}{\varepsilon}$$

where c is an absolute constant and k is the VC-dimension of $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$.

Proofsketch By (3) in the proof above and the inequality proved after the Sauer-Shelah Lemma 4.10 it suffices that $n = |s|$ satisfies

$$\log 6 + k\,\log(n+1) \;<\; \frac{n\varepsilon^2}{32},$$

which is the case if n satisfies the following inequality

$$c'\,\frac{k}{\varepsilon^2} \;<\; \frac{n}{\log n},$$

for some absolute constant c'. Finally, the latter inequality is satisfied if

$$c''\,\frac{k}{\varepsilon^2}\log\frac{k}{\varepsilon^2} \;<\; n$$

for some absolute constant c'', see the exercise below. □

5.18 Exercise Prove that for all $x, y > 1$

$$2x\log x < y \;\Rightarrow\; x < \frac{y}{\log y}$$ □

4 A Uniform Law of Large Numbers, again

We prove a second version of the Vapnik-Chervonenkis Inequality. Which, I conjecture, is due to Devroye and Lugosi [5].

5.19 Vapnik-Chervonenkis Inequality (2) Let $\pi_\varphi(n)$ be the shatter function of $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$. Let $S = S_1,\dots,S_n$ be a random sample from $\mathcal{U}$. Then, for every $b \in \mathcal{V}$

$$\mathbb{E}\,\Big|\,\mathrm{Fr}(S,b) - \mathbb{P}(b)\Big| \;\le\; 2\sqrt{\frac{\log\,\big(2\,\pi_\varphi(n)\big)}{n}}.$$ □

The same caveat on measurability apply as for Inequality 5.15.

We note that the bound is not optimal, using a clever techniques called *chaining*, Dudley could prove that

$$\mathbb{E}\,\Big|\,\mathrm{Fr}(S,b) - \mathbb{P}(b)\Big| \;<\; c\sqrt{\frac{k}{n}},$$

where k is the VC-dimension and c is absolute constant.

Before embarking in the proof of the theorem above, we prove the following (easy, although mysterious) lemma, which also has independent interest.

5.20 Lemma Let $X_1, \ldots, X_m$ be real valued random variables. Let c be such that

$$\mathbb{E}\big[e^{tX_i}\big] \;\leq\; e^{c^2t^2/2} \qquad \text{for every } i \leq m \text{ and every } t > 0.$$

Then

$$\mathbb{E}\big[\max_{i\leq m} X_i\big] \;\leq\; c\sqrt{2\log m}.$$

If in addition

$$\mathbb{E}\big[e^{-tX_i}\big] \;\leq\; e^{c^2t^2/2} \qquad \text{for every } i \leq m \text{ and every } t > 0,$$

then

$$\mathbb{E}\big[\max_{i\leq m} |X_i|\big] \;\leq\; c\sqrt{2\log(2m)}.$$

Proof By Jensen's inequality,

$$\begin{aligned}
\exp\Big(t\cdot\mathbb{E}\big[\max_{i\leq m} X_i\,\big]\Big) &\leq \mathbb{E}\Big[\exp\big(\max_{i\leq m} tX_i\,\big)\Big] \\
&= \mathbb{E}\Big[\max_{i\leq m} e^{tX_i}\Big] \\
&\leq \mathbb{E}\Big[\sum_{i\leq m} e^{tX_i}\Big] \\
&= \sum_{i\leq m} \mathbb{E}\big[e^{tX_i}\big] \\
&\leq m\, e^{c^2t^2/2}
\end{aligned}$$

Taking the logarithm of both sides and replacing t with $\dfrac{\sqrt{2\log m}}{c}$, we obtain the first inequality of the lemma.

To prove the second inequality, apply the first one to $X_1, \ldots, X_m, -X_1, \ldots, -X_m$. (N.B. note that independence is not assumed.) □

Proof of the Vapnik-Chervonenkis inequality Let $S' = S'_1, \ldots, S'_n$ be an independent copy of S. We claim that

$$\text{(1)} \qquad \mathbb{E}\Big[\sup_{b\in\mathcal{V}} \big|\operatorname{Fr}(S,b) - \mathbb{P}(b)\big|\Big] \;\leq\; \mathbb{E}\Big[\sup_{b\in\mathcal{V}} \big|\operatorname{Fr}(S,b) - \operatorname{Fr}(S',b)\big|\Big]$$

In fact,

$$\begin{aligned}
\operatorname{Fr}(S,b) - \mathbb{P}(b) &= \operatorname{Fr}(S,b) - \mathbb{E}\big[\operatorname{Fr}(S',b)\big] \\
&= \mathbb{E}\Big[\operatorname{Fr}(S,b) - \operatorname{Fr}(S',b) \;\big|\; S\Big].
\end{aligned}$$

Now, apply Jensen's inequality to the absolute value function, then use that

$$\text{(2)} \qquad \sup_{b\in\mathcal{V}} \mathbb{E}\big[\,\cdots\big] \;\leq\; \mathbb{E}\big[\sup_{b\in\mathcal{V}}(\cdots)\big].$$

Write $\mathbb{I}_b$ for the indicator function of $\varphi(\mathcal{U}\,;b)$. Then

$$\big|\operatorname{Fr}(S,b) - \operatorname{Fr}(S',b)\big| \;=\; \frac{1}{n}\Big|\sum_{i=1}^{n}\Big(\mathbb{I}_{S_i,b} - \mathbb{I}_{S'_i,b}\Big)\Big|$$

Let $\sigma = \sigma_1, \ldots, \sigma_n$ be a tuple of independent sign random variable. The random variable $\mathbb{I}_{S_i,b} - \mathbb{I}_{S'_i,b}$ has the same distribution of $\sigma_i\big(\mathbb{I}_{S_i,b} - \mathbb{I}_{S'_i,b}\big)$ hence

$$= \frac{1}{n}\mathbb{E}\bigg|\sum_{i=1}^{n}\sigma_i\Big(\mathbb{I}_{S_i,b} - \mathbb{I}_{S'_i,b}\Big)\ \Big|\ S, S'\bigg|$$

Inserting this into (1) we obtain

$$\mathbb{E}\Big[\sup_{b\in\mathcal{V}}\big|\operatorname{Fr}(S,b) - \mathbb{P}(b)\big|\Big] \leq \frac{1}{n}\,\mathbb{E}\bigg[\sup_{b\in\mathcal{V}}\mathbb{E}\bigg|\sum_{i=1}^{n}\sigma_i\Big(\mathbb{I}_{S_i,b} - \mathbb{I}_{S'_i,b}\Big)\ \Big|\ S, S'\bigg|\bigg]$$

Let s, s' be a generic realization of S, S'

$$\leq \frac{1}{n}\sup_{s,s'}\sup_{b\in\mathcal{V}}\mathbb{E}\bigg|\sum_{i=1}^{n}\sigma_i\big(\mathbb{I}_{s_i,b} - \mathbb{I}_{s'_i,b}\big)\bigg|$$

and, by what remarked in (2)

$$\leq \frac{1}{n}\sup_{s,s'}\mathbb{E}\bigg[\sup_{b\in\mathcal{V}}\bigg|\sum_{i=1}^{n}\sigma_i\big(\mathbb{I}_{s_i,b} - \mathbb{I}_{s'_i,b}\big)\bigg|\bigg]$$

Observe that once s, s' is fixed, $\sup_{b\in\mathcal{V}}$ is actually a maximum among $\pi_\varphi(2n)$ sets, in fact, $\pi_\varphi(2n)$ is the number of definable subsets of $A = \{s_1, \ldots, s_n, s'_1, \ldots, s'_n\}$. Then, by Lemma 5.20 (the second inequality, with $m = \pi_\varphi(2n)$ and i ranging over the definable subsets of A), for an appropriate constant c,

$$\leq \frac{1}{n}\sup_{s,s'} c\sqrt{2\log\big(2\,\pi_\varphi(2n)\big)}.$$

As the r.h.s. does not depend on s, s',

$$\leq \frac{c}{n}\sqrt{2\log\big(2\,\pi_\varphi(2n)\big)}$$

Finally, as $\pi_\varphi(2n) \leq \pi_\varphi(n)^2$,

$$\leq \frac{2c}{n}\sqrt{\log\big(2\,\pi_\varphi(n)\big)}$$

The Vapnik-Chervonenkis inequality is proved if we can show the assumption of Lemma 5.20 holds with $c = \sqrt{n}$.

$$\mathbb{E}\bigg[\exp\bigg(t\sum_{i=1}^{n}\sigma_i\big(\mathbb{I}_{s_i,b} - \mathbb{I}_{s'_i,b}\big)\bigg)\bigg] = \prod_{i=1}^{n}\mathbb{E}\Big[\exp\Big(t\,\sigma_i\big(\mathbb{I}_{s_i,b} - \mathbb{I}_{s'_i,b}\big)\Big)\Big]$$

As $\sigma_i\big(\mathbb{I}_{s_i,b} - \mathbb{I}_{s'_i,b}\big)$ takes values in $\{-1, 1\}$ with mean 0, by Lemma 5.9

$$\leq e^{nt^2/2}$$

and the same holds for $-\sigma_i\big(\mathbb{I}_{s_i,b} - \mathbb{I}_{s'_i,b}\big)$. □

As an application we prove the Glivenko-Cantelli Theorem, an important theorem of mathematical statistics. The theorem says that the empirical cumulative distribution function converges uniformly to the true one. We prove an informative variant which gives the rate of convergence (though, this is not optimal).

5.21 Glivenko-Cantelli Theorem Let $X = X_1, \ldots, X_n$ be i.i.d. random variables. Let $F(z) = \mathbb{P}(X_i \leq z)$ be their common cumulative distribution function. Let $F_e(x)$ be the empirical cumulative distribution function, that is,

$$F_e(X,z) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{X_i \le z}.$$

Then, for every $z \in \mathbb{R}$,

$$\mathbb{E}\left|F_e(X,z) - F(z)\right| \le 2\sqrt{\frac{\log(2n+2)}{n}}.$$ □

Proof Take $\mathcal{U} = \mathcal{V} = \mathbb{R}$ and $\varphi(x\,;z) = x \le z$. Assign to $\varphi(\mathcal{U}\,;b) = (-\infty, b]$ probability $\mathbb{P}(X_i \le b)$. Then, if $S = S_1, \ldots, S_n$ is a random sample from $\mathcal{U}$

$$\mathbb{E}\left|F_e(X,z) - F(z)\right| = \mathbb{E}\left|\operatorname{Fr}(S,b) - \mathbb{P}(b)\right|$$

hence, by the Vapnik-Chervonenkis inequality 5.15

$$\le \sqrt{2\frac{\log\ (2\,\pi_\varphi(n))}{n}}.$$

It is clear that $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$ has VC-dimension 1. By the Sauer-Shelah Lemma 4.10 and the inequalities proved thereafter, $\pi_\varphi(n) \le n+1$. The theorem follows. □

Chapter 6

Small transversals

Unlike the rest of these notes, this chapter is not self contained, as it relies on the duality of linear programming. The reader can use the result as a black box. Otherwise, we recommend [10, Chapter 6], a lively and conceptual introduction to linear programming (a rarity for an otherwise rather dry subject).

1 Transversals and packings

Let $\varphi(x\,;z)$ be given. Let $A \subseteq \mathcal{U}$ and $B \subseteq \mathcal{V}$ be finite sets.

A subset $A' \subseteq A$ is a transversal if $\varphi(A', b) \neq \varnothing$ for every $b \in B$. Equivalently, if the sets $\varphi(a, B)_{a \in A'}$ cover B, i.e.

$$B \;=\; \bigcup_{a \in A'} \varphi(a, B).$$

The transversal number is the smallest cardinality of a transversal A'. It denoted by τ.

A subset $B' \subseteq B$ is a packing if $\varphi(A, b) \cap \varphi(A, b') = \varnothing$ for every distinct $b, b' \in B'$. Equivalently if $|\varphi(a, B')| \leq 1$ for every $a \in A$. The packing number is the largest cardinality of a packing B'. It is denoted by ν.

We may write $\tau_{\varphi(A,B)}$ and $\nu_{\varphi(A,B)}$ when ambiguity is of concern.

If A' is a transversal and $B' \subseteq B$, then the sets $\varphi(a, B')_{a \in A'}$ cover B'. Now, suppose B' is a packing, then these sets contain at most one element, hence $|B'| \leq |A'|$. Therefore, we always have $\nu \leq \tau$. Very little can be said in general about the reverse direction.

6.1 Example Let $\mathcal{U} = \mathbb{R}^2$ and $\mathcal{V}$ is the set of lines in $\mathbb{R}^2$. Let $\varphi(x\,;z)$ be the incidence (that is, membership) relation. Let $A \subseteq \mathcal{U}$ and let $B \subseteq \mathcal{V}$ be a set of n lines in generic position (any to lines intersect and every point is contained in at most two lines). Then $\tau = \lceil n/2 \rceil$, as each point belongs to at most two lines, while $\nu = 1$, as any two lines intersect. □

A (fractional) multiset over $\mathcal{U}$ is a real-valued function $A' : \mathcal{U} \to \mathbb{R}$. The support support of a multiset A' is the set where it takes nonzero values. In this chapter will only consider nonnegative multisets with finite support. These can be interpreted as measures concentrated on a finite set.

If $A'' : \mathcal{U} \to \mathbb{R}$ is another multiset, we write $A' \cdot A''$ for the pointwise product of the two. We write $A' \leq A''$ if $A'(a) \leq A''(a)$ for every $a \in \mathcal{U}$.

We define the size of A' to be

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

We write $\varphi(A', b)$ for the multiset $A' \cdot \mathbb{I}_{\varphi(\mathcal{U}, b)}$.

Multisets over $\mathcal{V}$ are defined analogously.

A fractional transversal is a multiset $A' \leq \mathbb{I}_A$ such that $|\varphi(A', b)| \geq 1$ for every $b \in B$. The fractional transversal number of $\varphi(x\,;z)$, denoted by τ^*, is the infimum of the size of the fractional transversals of $\varphi(x\,;z)$.

A fractional multiset $B' \leq \mathbb{I}_B$ over $\mathcal{V}$ is a fractional packing if $|\varphi(a, B')| \leq 1$ for every $a \in A$. The fractional packing number of $\varphi(x\,;z)$, denoted by ν^*, is the supremum of the size of the fractional packings of $\varphi(x\,;z)$.

6.2 Example The sets $\mathcal{U}, \mathcal{V}$ and the relation $\varphi(A\,;B)$ are as in the example 6.1. Let B' be a multiset that assigns $1/2$ to every line in B. Then $|\varphi(a, B')| \leq 1$ holds because each point is contained in at most two lines. Then $\nu^* \geq |B'| = n/2$. It is easy to see that $\tau^* \geq n/2$. We claim that $\tau \leq n/2$. If n is even, use the same transversal A' as in Example 6.1 is even. 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. □

6.3 Exercise Let $\mathcal{U} = \mathbb{R}$ and $\mathcal{V}$ is a set of finitely many closed intervals. Let $\varphi(x\,;z)$ be the membership relation. Then $\nu = \tau$. Hint: use induction on ν. □

6.4 Theorem For all $\varphi(x\,;z)$ and all finite sets $A \subseteq \mathcal{U}$ and $B \subseteq \mathcal{V}$, we have $\nu^* = \tau^*$ and this value is rational.

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

$$\begin{aligned} \tau^* &= \inf \{ 1_m^{\mathrm{T}} x \ : \ F^{\mathrm{T}} x \geq 1_n, \ 0 \leq x \}; \\ \nu^* &= \sup \{ 1_n^{\mathrm{T}} y \ : \ F y \leq 1_m, \ 0 \leq y \}. \end{aligned}$$

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

As τ^* is the minimum of the linear function $x \mapsto 1_m^{\mathrm{T}} x$ over a polyhedron, such minimum is attained at vertex. The inequalities describing the polyhedron have rational coefficients, so also the vertices have rational coordinates (elaborate on this). □

When $\varphi(x\,;z)$ has finite VC-dimension, the transversal number τ is bounded by a function of τ^*.

6.5 Proposition Let $\varphi(x\,;z)$ have VC-dimension k. Then for all finite sets $A \subseteq \mathcal{U}$ and $B \subseteq \mathcal{V}$

$$\tau \leq c\,k\,(\tau^*)^2 \ln(k\,\tau^*)$$

where c is an absolute constant.

Proof Let A' be an optimal fractional transversal. After normalizing, A' defines a probability measure $\mathbb{P}$ on $\mathcal{U}$. Namely, $\mathbb{P}(\{a\}) = A'(a)/\tau^*$ for $a \in \mathcal{U}$. By the definition of fractional transversal, every set $\varphi(A\,;b)_{b\in B}$ has measure at last $1/\tau^*$. By Proposition 5.17, for every $\varepsilon > 0$ there is a sample s of size

$$n \;\leq\; c\,\frac{k}{\varepsilon^2}\log\frac{k}{\varepsilon}$$

If we set $\varepsilon = 1/\tau^*$, then $\mathrm{range}(s)$ is a transversal and we obtain the required bound. □

The bound in the proposition above can be improved. One can replace $(\tau^*)^2$ with τ^* at the cost of a more difficult proof.

2 Helly-type properties

We now investigate methods of bounding $\tau^* = \nu^*$. As motivation we cite a classical theorem of Helly.

6.6 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.

We will deal with the following property, which is more robust. It says that if there is *plenty* of *small* collections of sets with nonempty intersection, then there is a *large* collection with nonempty intersection.

6.7 Definition We say that $\varphi(x\,;z)$ has fractional Helly number k if for all $\alpha > 0$ there is a $\beta > 0$ such that for every finite $A \subseteq \mathcal{U}$ and $B \subseteq \mathcal{V}$ the following holds (write n for $|B|$):

(1) $$\bigcap_{b\in B'} \varphi(A,b) \;\neq\; \varnothing \qquad \text{for at least } \alpha\binom{n}{k} \text{ sets } B' \in \binom{B}{k}$$

then

(2) $$\bigcap_{b\in B''} \varphi(A,b) \;\neq\; \varnothing \qquad \text{for some } B'' \subseteq B \text{ of cardinality} \geq \beta n.$$

We say that φ has the fractional Helly property if it has fractional Helly number k for some finite k. The fractional Helly number of $\varphi(x\,;z)$ is the smallest number k satisfying the property above. □

For further reference, we note that (2) in the definition above can be rewritten as

(2′) $$\big|\varphi(a,B)\big| \;\geq\; \beta n \qquad \text{for some } a \in A.$$

The following theorem proves that NIP formulas have the fractionally Helly property (in a strong sense).

6.8 Theorem (Matoušek) Let $\varphi(x\,;z)$ have VC-codimension $< k$. Then $\varphi(x\,;z)$ has fractional Helly number k. Moreover, β in the definition above only depends on α and k.

Proof Let α be arbitrary and set $\beta = 1/2m$ where m is such that

$$\sum_{i=0}^{k-1}\binom{m}{i} \;<\; \frac{\alpha}{4}\binom{m}{k}.$$

Note that, by the Sauer-Shelah Lemma 4.10, the r.h.s. is strictly larger than $\pi^*_\varphi(m)$ for all $\varphi(x\,;z)$ with VC-codimension $< k$.

Assume for a contradiction that some finite $A \subseteq \mathcal{U}$ and $B \subseteq \mathcal{V}$ contradicts the definition above. That is,

$$(1) \qquad \bigcap_{b\in B'} \varphi(A,b) \;\neq\; \varnothing \qquad\qquad \text{for at least } \alpha\binom{n}{k} \text{ sets } B' \in \binom{B}{k}$$

and

$$(2) \qquad \big|\varphi(a\,;B)\big| \;<\; \beta\, n \qquad\qquad \text{for all } a \in A.$$

Note that we can assume that $n > 2m$ otherwise $\beta\, n < 1$ and (2) never occur. We will find a set $B'' \subseteq B$ of cardinality m with more than $\pi^*(m)$ distinct $\varphi(x\,;z)^{\text{op}}$-definable subsets, a contradiction.

Let P be the set of pairs $B' \subseteq B'' \subseteq B$ such that $|B'| = k$ and $|B''| = m$. We say that a pair $B' \subseteq B''$ in P is *good* if there is $a \in A$ such that $B' = \varphi(a\,;B'')$. That is, B' is a $\varphi(x\,;z)^{\text{op}}$-definable subset of B''.

Claim 1. Assume the uniform probability on P. Then the probability that a random pair is good is $\geq \alpha/4$.

Assume Claim 1 for now and continue with the proof. We can think that the random pair in P is chosen by first picking $B'' \in \binom{B}{m}$ with the uniform distribution and than $B' \in \binom{B''}{k}$ again with the uniform distribution. (To put 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 $B'' \in \binom{B}{m}$ the probability of finding a good subset B' is $\geq \alpha/4$. Therefore, B'' has $\geq \frac{\alpha}{4}\binom{m}{k} > \pi^*(m)$ 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 $B' \subseteq B$ of cardinality k then obtain B'' by adding $m-k$ random elements from $B \smallsetminus B'$. By (1), the probability that B' is such that $\bigcap_{b\in B'} \varphi(A,b) \neq \varnothing$ is at least α. So, assume that B' is such, and fix any a is this intersection. By (2), there are $|\varphi(a,B)| < \beta n$. Then the probability that all $b \in B'' \smallsetminus B'$ are such that $\neg\varphi(a,b)$ is at least

$$\binom{n-\beta n}{m-k}\Big/\binom{n-k}{m-k} \;=\; \prod_{i=0}^{m-k-1}\frac{n-\beta n-i}{n-k-i} \qquad\qquad \text{we write } \beta n \text{ for } \lfloor \beta n\rfloor$$

$$\geq\; \prod_{i=0}^{m-k-1}\frac{n-\beta n-m}{n-m}$$

$$
\begin{aligned}
&= \left(\frac{n-\beta n-m}{n-m}\right)^m \\
&= \left(1-\frac{\beta n}{n-m}\right)^m \\
&\geq (1-2\beta)^m && \text{because } n > 2m \\
&\geq \left(1-\frac{1}{m}\right)^m && \text{because } \beta \leq 1/2m.
\end{aligned}
$$

As we can assume $m \geq 2$, the probability that a random pair in P is good is at least $\alpha/4$. This proves Claim 1 and with it the theorem. □

3 The (p,q)-theorem

For integers $p \geq q$ we say that $\varphi(x\,;z)$ has the (p,q)-property if for every finite $A \subseteq \mathcal{U}$ and every $B \subseteq \mathcal{V}$ of cardinality p there is some $B' \subseteq B$ of cardinality q such that

$$\bigcap_{b\in B'} \varphi(A\,;b) \neq \varnothing$$

For ease of speaking we call $\varphi(A\,;b)_{b\in B}$ a p-collection of definable subsets of A. Note that, strictly speaking, these is not collection of sets but collection of parameters defining sets (though not intrinsically relevant, it is convenient not to assume extensionality).

Hence, we can rephrase the (p,q)-property in plain English: out of any p-collection of definable sets there are at least q sets with nonempty intersection.

Helly's theorem says that any finite collection of convex sets in $\mathbb{R}^d$ satisfying the $(d+1,d+1)$-property has non-empty intersection, i.e. admits a transversal of size 1. A generalization of this was conjectured by Hadwiger and Debrunner, and many years later proved by Alon and Kleitman [1], [2]. Subsequently, after proving Theorem 6.8, Matoušek [9] noted that the method in [2] applies also to collections of sets with finite VC-dimension. More precisely, Matoušek proved the existence of a bound to the cardinality of the transversal number τ of NIP formulas with the (p,q)-property, and that this bound depends only on p,q and the VC-codimension of the formula. (Formally, the results [2] and [9] do not imply each other.)

6.9 Theorem (Alon, Kleitman + Matoušek) Let $p \geq q > k$ be natural numbers. There is a number $N = N(k,p,q)$ such that $\tau_{\varphi(A\,;B)} \leq N$ for all formulas $\varphi(x\,;z)$ with the (p,q)-property and VC-codimension $< k$, and every $A \subseteq \mathcal{U}$ and $B \subseteq \mathcal{V}$.

Proof As we are not trying to optimize N, we may prove the theorem for $q = k+1$. By Proposition 6.5, the transversal number is bounded by a function of τ^*, so it is enough to bound τ^*. By Theorem 6.4, we can equivalently bound the fractional packing number ν^* because it coincides with τ^*.

Let $B' \subseteq B$ be an optimal fractional packing. That is, $\nu^* = |B'|$ where $B' \leq \mathbb{I}_B$ is such that $|\varphi(a, B')| \leq 1$ for every $a \in A$. As we rather work with regular sets than with fractional multisets, we apply a trick that allows to replace B' with a regular set C'.

By Theorem 6.4 we may assume that B' is rational valued. Therefore $B' = (1/m)\, C$ where m is a positive integer and C is a integral valued multiset over $\mathcal{V}$. Replace $\mathcal{V}$ with $\mathcal{V} \times [m]$. Define $\varphi_{\times}(a\,; b, i)$ to be the relation that holds if and only if $\varphi(a\,; b)$ holds. Then we can replace the multiset C with a regular set $C' \subseteq \mathcal{V} \times [m]$ such that $|\varphi_{\times}(a\,; C')| \leq m$ for every $a \in A$.

If write n for $|C'|$, then $\nu^* = n/m$.

Claim 1. $\varphi_{\times}(x\,; z, y)$ satisfies the (qp, q)-property.

Let $D \subseteq \mathcal{V} \times [m]$ have cardinality qp. If the qp-collection $\varphi_{\times}(A\,; b, i)_{b,i \in D}$ contains q copies of the same definable set $\varphi(A\,; b)$, then we immediately have the required q-collection with nonempty intersection. So, suppose not. Then $\varphi_{\times}(A\,; b, i)_{b,i \in D}$ contains p distinct sets $\varphi(A, b_1), \ldots, \varphi(A, b_p)$. Then the q-collection with nonempty intersection is obtained from the (p, q)-property of $\varphi(x\,; z)$.

Claim 2. There is an $\alpha = \alpha(p, q) > 0$ such that

$$\bigcap_{b,i \in D} \varphi_{\times}(A\,; b, i) \;\neq\; \varnothing \qquad\qquad \text{for at least } \alpha \binom{n}{q} \text{ sets } D \in \binom{C'}{q}.$$

By Claim 1, every qp-collection of $\varphi_{\times}$-definable sets contains at least one q-collection with non-empty intersection. Every q-collection is contained in $\binom{n-q}{qp-q}$ many qp-collections. Therefore the number q-collections with non-empty intersection is at least

$$\binom{n}{qp} \Big/ \binom{n-q}{qp-q} \;=\; \binom{n}{q} \Big/ \binom{qp}{q}.$$

Therefore, the claim holds with $1/\alpha = \binom{qp}{q}$.

Now we can resume the proof of the theorem (recall that our goal is to bound ν^* by a function of p, q, and k). Let $\beta = \beta(\alpha, k)$ be as in Theorem 6.8. As $\varphi_{\times}(x\,; z, y)$ has the same VC-codimension as $\varphi(x\,; z)$, by Claim 2 there is an $a \in A$ such that $\varphi(a, C')$ has cardinality at least $\beta\, n$. So, from $\beta\, n \leq |\varphi(a, C')| \leq m$ we obtain $\nu^* \leq 1/\beta$. $\square$

Chapter 7

Zarankiewicz problem(s)

Let us start with presenting Zarankiewicz problem. Let $A \subseteq \mathcal{U}$ and $B \subseteq \mathcal{V}$ have cardinality m respectively n. Given two integers s, t (both at least 2 to avoid trivialities), what is the maximal cardinality of a graph $\varphi(A\,;B)$ that does not contain $A' \times B'$ for any $A' \subseteq A$ and $B' \subseteq B$ of cardinality s respectively t? Denote this maximum by $z(m,n;s,t)$. Define also $z(n;t) = z(m,n;s,t)$. In 1951 Zarankiewicz posed the problem of determining $z(n;3)$ for $n = 4,5,6$ and the general problem has also become known as the problem of Zarankiewicz.

1 The Kővári-Sós-Turán Theorem

The theorem in this section is a classical result of Kővári, Sós, and Turán. They actually proved it for $z(n;t)$ but the proof easily generalizes.

In the proof we need the following generalization of the binomial coefficient. For any positive integer t and any real x we define

$$\binom{x}{t} = \frac{x(x-1)\cdots(x-t+1)}{t!}.$$

It is easy to verify that for any fixed t, this is a convex and strictly increasing function of x.

7.1 Kővári-Sós-Turán Theorem For all $2 \leq s \leq m$ and $2 \leq t \leq n$

$$z(m,n;s,t) < (s-1)^{1/t}(n-t+1)m^{1-1/t} + (t-1)m$$

or also

$$< c\left(n\, m^{1-1/t} + m\right)$$

for some $c = c(s,t)$.

Proof Let $A \subseteq \mathcal{U}$ and $B \subseteq \mathcal{V}$ have cardinality m respectively n. Let P be the set of pairs $\langle a, B' \rangle$ such that $B' \subseteq \varphi(a\,;B)$ and B' has cardinality t. Writing d_a for the cardinality of $\varphi(a\,;B)$, we have

$$|P| = \sum_{a \in A} \binom{d_a}{t}$$

Write z for $|\varphi(A\,;B)|$, and note that

$$z = \sum_{a \in A} d_a.$$

As $\binom{x}{t}$ is a convex function of x,

$$\binom{z/m}{t} \leq \frac{1}{m}|P|. \tag{1}$$

Now, suppose that $z = z(m, n; s, t)$. Then for any fixed $B' \subseteq B$ there are at most $s-1$ pairs $\langle a, B' \rangle \in P$. Hence

$$(2) \qquad |P| \leq (s-1)\binom{n}{t}.$$

Together, (1) and (2) yield

$$m\binom{z/m}{t} \leq (s-1)\binom{n}{t}.$$

As $\binom{x}{t}$ is strictly increasing function of x, and $z/m < n$,

$$m\left(\frac{z}{m} - t + 1\right)^t < (s-1)\left(n - t + 1\right)^t$$

which easily yields the theorem. □

2 The NIP case

Chapter 8

Szemerédi regularity lemma

1 The regularity lemma

2 Stable regularity lemma

3 Distal regularity lemma

References

[1] N. Alon and D. J. Kleitman, *A purely combinatorial proof of the Hadwiger Debrunner (p, q) conjecture*, Electron. J. Combin. **4** (1997).

[2] Noga Alon and Daniel J. Kleitman, *Piercing convex sets and the Hadwiger-Debrunner (p, q)-problem*, Adv. Math. **96** (1992).

[3] R. P. Anstee, Lajos Rónyai, and Attila Sali, *Shattering news*, Graphs Combin. **18** (2002).

[4] Luc Devroye, László Györfi, and Gábor Lugosi, *A probabilistic theory of pattern recognition*, Springer-Verlag, 1996.

[5] Luc Devroye and Gábor Lugosi, *Combinatorial methods in density estimation*, Springer Series in Statistics, Springer-Verlag, 2001.

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

[7] Wilfrid Hodges, *Encoding orders and trees in binary relations*, Mathematika **28** (1981), 67–71.

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

[9] Jiří Matoušek, *Bounded VC-dimension implies a fractional Helly theorem*, Discrete Comput. Geom. **31** (2004).

[10] Jĭri Matousek and Bernd Gartner, *Understanding and Using Linear Programming*, Springer-Verlag, 2007.

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

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

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

[14] 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.