# Scombinatorics

# Notation

Let $\mathcal{U}$ and $\mathcal{V}$ be two (large) 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 $A$.

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

$$\binom{A}{k} \;=\; A^{(k)} \;=\; \left\{A'\subseteq A : |A'|=k\right\}$$

# 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

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

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. $|B'| \;\leq\; \left| \bigcup_{b \in B'} \varphi(A;b) \right|$ *for every $B' \subseteq B$.*

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

$$|B'| \;=\; \left| \bigcup_{b \in B'} \varphi(A';b) \right| \;\subseteq\; \left| \bigcup_{b \in B'} \varphi(A;b) \right|.$$

(2$\Rightarrow$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 $\varphi(A \smallsetminus \{a\}; B \smallsetminus \{b\})$ 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

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

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

3

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

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

By the definition of $A',B'$

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

The two sets above are disjunct, hence

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

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. $\qquad\square$

# 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

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

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 $a \in A \smallsetminus A'$.* $\qquad\square$

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 $\varphi(A'\,;B\smallsetminus B')^*$ 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 $A', B'$. $\qquad\square$

# 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 $a < a'$ for every $a, a' \in A'$. A chain is a set $A' \subseteq \mathcal{U}$ such that $a < a' \ \lor \ a' < a$ 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 cardinal-

ity of a partition of $A$ into chains.

Let $A_1, \dots, A_k$ 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 $A_1, \dots, A_k$ 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 $\varphi(A\,;f[A])$. We claim that $A \smallsetminus (A_1 \cup f^{-1}[B_1])$ 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 $a_0, b_0, a_1, b_1, \dots, a_m, b_m, a_{m+1}$ 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 $\leq |A| \smallsetminus |A''| \leq |A'|$. $\qquad\square$

# 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

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

The following classical theorem it is not possible to improve 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 $\Pi$ 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 $k$

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

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

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

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

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

Now, the inequality is evident. $\qquad\square$

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

$$\mathbb{P}_k(\{A'\}) \;=\; \begin{cases} 0 & \text{if } |A'| \neq k \\[2mm] \binom{n}{k}^{-1} & \text{if } |A'| = k \end{cases}$$

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

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

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 $m \cdot n \cdot |H|/|G| \leq kn$. $\qquad\square$

**2.4  Erdős-Ko-Rado Theorem**  *Let $A \subseteq \mathcal{U}$ 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 $m = \left| \varphi(A\,;b)_{b\in V} \right|$. Consider the graph

$$G = \binom{A}{k},$$

$$E(G) = \left\{ \{A', A''\} \;:\; A' \cap A'' \neq \varnothing \right\}.$$

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

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

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

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

therefore

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

# Chapter 3

# Stable and NIP relations

## 1 Stable formulas

The ladder-dimension 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 $n$ of a chain of the form

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

for some set $A\subseteq\mathcal{U}$ and some $b_0,\ldots,b_{n-1}\in\mathcal{V}$. If a maximal length exists we say that $\varphi(x\,;z)$ is stable otherwise we say that $\varphi(x\,;z)$ is unstable.

## 2 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

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

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

The Vapnik-Chervonenkis dimension of $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$, 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.

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 $\varphi(\mathcal{U}\,;b)_{b\in\mathcal{V}}$.

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

well Levine.

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

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

**3.4 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. □

**3.5 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). □

**3.6 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

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

□

We call the VC-dimension of $\varphi(x; z)^*$ the dual VC-dimension of $\varphi(x; z)$.

**3.7 Proposition** *If $\varphi(x; z)$ has VC-dimension $< k$ then its dual VC-dimension is $< 2^k$.*

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

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

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

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

That is, $\varphi(x; z)$ shatters $A = \{a_i : i \in k\}$. □

# 3 The Sauer-Shelah lemma

According to Gil Kalai in [5], 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 [8], Sauer [7], and Vapnik-Chervo-nenkis [9] 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.

**3.8 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 3.6 shows that the bound is optimal.

An alternative proof of the Sauer-Shelah Lemma derives it as corollary of a lemma by Alain Pajor [6].

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

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

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

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

Notice that

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

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

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

By induction hypothesis,

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

and

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

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

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. $\qquad\square$

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

$$\pi_\varphi(n) = \max_{|A|=n} \left| \varphi(A,b)_{b\in\mathcal{V}} \right|$$

$$\pi_\varphi(n) \leq \max_{|A|=n} \left| \{ C \subseteq A \,:\, C \text{ shattered by } \varphi(x\,;z) \} \right| \quad \text{by Pajor's Lemma}$$

$$\leq \sum_{i=0}^{k} \binom{n}{i} \qquad \text{because } \varphi(x\,;z) \text{ has VC-dimension } k \qquad \square$$

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 $\pi_\varphi(n) \in O(n^r)$. 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.

$$\sum_{i=0}^{k} \binom{n}{i} = \sum_{i=0}^{k} \frac{n!}{i!\,(n-i)!}$$

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

$$\leq \sum_{i=0}^{k} \frac{n^i \, k!}{i!(k-i)!}$$

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

$$= (n+1)^k \qquad \text{by the binomial theorem.}$$

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

$$\sum_{i=0}^{k} \binom{n}{i} \leq \left(\frac{n}{k}\right)^k \sum_{i=0}^{k} \left(\frac{k}{n}\right)^i \binom{n}{i} \qquad \text{because } \frac{k}{n} < 1$$

$$\leq \left(\frac{n}{k}\right)^k \sum_{i=0}^{n} \left(\frac{k}{n}\right)^i \binom{n}{i}$$

$$= \left(\frac{n}{k}\right)^k \left(1 + \frac{k}{n}\right) \qquad \text{by the binomial theorem}$$

$$\leq \left(\frac{n \, e}{k}\right)^k \qquad \text{where } e \text{ is the base of the natural logarithm.}$$

# Chapter 4

# 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 doesn't 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 finite $\mathcal{U}$, but they are easily seen to hold in general.

**4.1 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*

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

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

**4.2 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. □

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.

**4.3 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 $A = \{a \in \mathcal{U} : X(a) \geq \varepsilon\}$.

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

**4.4 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 $\mathbb{P}\big(X \geq \varepsilon\big) = \mathbb{P}\big(X^k \geq \varepsilon^k\big)$. □

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

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

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

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

**4.6 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 $t > 0$

$$\mathbb{P}\big(X \geq \varepsilon\big) \;=\; \mathbb{P}\big(e^{tX} \geq e^{t\varepsilon}\big) \qquad\qquad\qquad \text{because } e^{tx} \text{ is increasing}$$

15

$$\leq \ e^{-t\varepsilon}\, \mathbb{E}\big[e^{tX}\big],$$

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

**4.7 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*

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

**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,

$$\mathbb{E}\Big[e^{tX}\Big] \ \leq \ \frac{b\, e^{ta} - a\, e^{tb}}{d}$$

$$\log \mathbb{E}\Big[e^{tX}\Big] \ \leq \ \log \frac{b\, e^{ta} - a\, e^{tb}}{d}$$

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 $\leq 1/4$) we obtain

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

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

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

*Then for every $\varepsilon > 0$*

$$\mathbb{P}\Big(M \geq \varepsilon\Big) \ \leq \ \exp\Big(-\frac{2\varepsilon^2}{nd^2}\Big),$$

$$\mathbb{P}\Big(M \leq -\varepsilon\Big) \ \leq \ \exp\Big(-\frac{2\varepsilon^2}{nd^2}\Big).$$

*Clearly, the two inequalities above imply the following*

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

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

$$
\begin{aligned}
\mathbb{P}\Big(M \geq \varepsilon\Big) \ &\leq \ e^{-t\varepsilon}\, \mathbb{E}\Big[e^{tM}\Big] && \text{by Chernoff's method (4.6)}\\
&= \ e^{-t\varepsilon}\, \prod_{i=1}^{n} \mathbb{E}\Big[e^{t\,(X_i-\mu_i)}\Big] && \text{by independence.}\\
&\leq \ e^{-t\varepsilon}\, \prod_{i=1}^{n} \exp\Big(\frac{t^2 d^2}{8}\Big) && \text{by Hoeffding's Lemma (4.7).}
\end{aligned}
$$

16

$$= \exp\left(\frac{n\,t^2 d^2}{8} - t\varepsilon\right)$$

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

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.

**4.9 Lemma** *Let $\sigma$ be a random sign variable. Then for every $t > 0$*

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

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

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

**4.10 Second proof of Hoeffding's Lemma** Recall that Hoeffding's Lemma claims that, if $a \leq X \leq b$, then

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

where $\mathbb{E}[X] = \mu$ and $d = b - a$.

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

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

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

$$
\begin{aligned}
&= \mathbb{E}\left[e^{t\sigma(X-X')}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[e^{t\sigma(X-X')} \mid X, X'\right]\right] \\
&\leq \mathbb{E}\left[e^{t^2(X-X')^2/2}\right] && \text{by Lemma 4.9} \\
&\leq e^{t^2 d^2/2} && \text{because } |X - X'| \leq d.
\end{aligned}
$$

This yields the bound above (only with 2 in place of 8). $\square$

# 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 $s$

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

It is convenient to rewrite it using indicator functions

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

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

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

and that, for $n$ large enough, $\mathrm{some\_bound}(\varepsilon, n) < 1$. Then a sample of size $\leq n$ that approximate the probability within $\varepsilon$ 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 $S = S_1, \ldots, S_n$ be of independent random elements of $\mathcal{U}$. Then $\mathbb{I}_{S_i \in A}$ as a Bernoulli random variable with probability of success $\mathbb{P}(A)$ and

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

is (up to the factor $1/n$) a binomial random variable.

**4.11  Weak Law of Large Numbers**  *For every event $A \subseteq \mathcal{U}$ and every $n > 0$*

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

**Proof**  Let $S_1, \ldots, S_n$ be a random sample from $\mathcal{U}$. Up to the factor $1/n$, the distribution of $\mathrm{Fr}(S, A)$ is binomial with parameters $n$ and $\mathbb{P}(A)$. Therefore it 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. $\square$

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

**4.12  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*

$$\left| \operatorname{Fr}(s, A) - \mathbb{P}(A) \right| < \varepsilon.$$

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

$$\frac{1}{n\varepsilon^2} < 1 \qquad\qquad \square$$

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

**4.13** **Weak Law of Large Numbers (with exponential bound)** *For every event $A \subseteq \mathcal{U}$ and every $n > 0$*

$$2e^{-2n\varepsilon^2} \geq \mathbb{P}\Big(s \in \mathcal{U}^n \ : \ \left| \operatorname{Fr}(s, A) - \mathbb{P}(A) \right| \geq \varepsilon \Big).$$

**Proof** Let $S_1, \ldots, S_n$ be a random sample from $\mathcal{U}$. 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

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

and this follows immediately from Hoeffding inequality. $\qquad \square$

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 4.12.

**4.14** **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*

$$\left| \operatorname{Fr}(s, A) - \mathbb{P}(A) \right| < \varepsilon. \qquad\qquad \square$$

# 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

$$\mathbb{P}(b) = \mathbb{P}\Big(\varphi(\mathcal{U}; b)\Big)$$

$$\operatorname{Fr}(s, b) = \operatorname{Fr}\Big(s, \varphi(\mathcal{U}; b)\Big)$$

$$\mathbb{I}_{s,b} = \mathbb{I}_{\varphi(s;b)}$$

An ε-approximation is a sample $s$ such that

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

We are interested in estimating the minimal size of an $\varepsilon$-approximation.

The main theorem of this section is this famous result of Vapnik-Chervonenkis [9].

**4.15 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 $b \in \mathcal{V}$*

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

N.B. Some technical hypothesis of measurability are necessary when $\varphi(\mathcal{U};b)_{b\in\mathcal{V}}$ is uncountable. This 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

$$\mathbb{P}\left( \left| \mathrm{Fr}(S,b) - \mathbb{P}(b) \right| \geq \varepsilon \right) \;\leq\; \mathbb{P}\left( \left| \mathrm{Fr}(S,b) - \mathrm{Fr}(S',b) \right| \geq \frac{\varepsilon}{2} \right) \;+\; (*)$$

where, by the Weak Law of Large Numbers 4.13,

$$(*) \;=\; \mathbb{P}\left( \left| \mathrm{Fr}(S',b) - \mathbb{P}(b) \right| \geq \frac{\varepsilon}{2} \right)$$

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

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

$$\mathbb{P}\left( \left| \mathrm{Fr}(S,b) - \mathrm{Fr}(S',b) \right| \geq \frac{\varepsilon}{2} \right) \;=\; \mathbb{P}\left( \sum_{i=1}^{n} \left| \mathbb{I}_{S_i,b} - \mathbb{I}_{S'_i,b} \right| \geq \frac{n\varepsilon}{2} \right)$$

Then, again by the triangular inequality

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

Putting together the inequalities above we obtain

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

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

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

Finally, note that the r.h.s. of (2) only depends on $\varphi\big(\mathrm{range}(s),b\big)$. 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 $b_1,\dots,b_m \in \mathcal{V}$ such that

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

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

$$\mathbb{P}\left( \sup_{i\leq m} X_i \geq \varepsilon \right) \;=\; \mathbb{P}\left( \bigcup_{i=0}^{m} X_i \geq \varepsilon \right)$$

$$\leq \sum_{i=1}^{m} \mathbb{P}(X_i \geq \varepsilon)$$

Hence, continuing from (2) we obtain

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

$$\leq 2\,\pi_\varphi(n) \exp\left(-\frac{n\varepsilon^2}{32}\right),$$

where the last inequality is obtained from Hoeffding's Inequality 4.8. In fact, Hoeffding's Inequality, applied to $X_i = \sigma_i \, \mathbb{I}_{s_i, b}$ with $n\varepsilon/4$ for $\varepsilon$, yields

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

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

$$\leq 4\pi_\varphi(n) \exp\left(-\frac{n\varepsilon^2}{32}\right) + 2 \exp\left(-\frac{n\varepsilon^2}{2}\right)$$

$$\leq 6\pi_\varphi(n) \exp\left(-\frac{n\varepsilon^2}{32}\right),$$

which finally proves the theorem. $\qquad\square$

**4.16 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*

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

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

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

By the Vapnik-Chervonenkis inequality 4.15, it suffices that

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

By the Sauer-Shelah Lemma 3.8, $\pi_\varphi(n)$ grows polynomially. Hence the inequality holds for $n$ large enough. $\qquad\square$

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

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

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

*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 3.8 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 $c$. □

# 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 [3].

**4.18 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 $b \in \mathcal{V}$*

$$\mathbb{E} \left| \operatorname{Fr}(S, b) - \mathbb{P}(b) \right| \ \leq \ 2 \sqrt{\frac{\log \ (2\, \pi_\varphi(n))}{n}}.$$ □

The same caveat on measurability apply as for Inequality 4.15.

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

$$\mathbb{E} \left| \operatorname{Fr}(S, b) - \mathbb{P}(b) \right| \ < \ c \sqrt{\frac{k}{n}},$$

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.

**4.19 Lemma** *Let $X_1, \dots, X_m$ be real valued random variables. Let $c$ be such that*

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

*Then*

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

*If in addition*

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

*then*

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

**Proof** By Jensen's inequality,

$$\exp \left( t \cdot \mathbb{E}\left[ \max_{i \leq m} X_i \right] \right) \ \leq \ \mathbb{E}\left[ \exp\left( \max_{i \leq m} t X_i \right) \right]$$

$$= \ \mathbb{E}\left[ \max_{i \leq m} e^{tX_i} \right]$$

22

$$= \mathbb{E}\left[\sum_{i \leq m} e^{tX_i}\right]$$

$$= \sum_{i \leq m} \mathbb{E}\left[e^{tX_i}\right]$$

$$= m\, e^{c^2 t^2 / 2}$$

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.) $\qquad\square$

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

(1) $$\mathbb{E}\left[\sup_{b \in \mathcal{V}} \left| \mathrm{Fr}(S, b) - \mathbb{P}(b)\right|\right] \leq \mathbb{E}\left[\sup_{b \in \mathcal{V}} \left| \mathrm{Fr}(S, b) - \mathrm{Fr}(S', b)\right|\right]$$

In fact,

$$\mathrm{Fr}(S, b) - \mathbb{P}(b) = \mathrm{Fr}(S, b) - \mathbb{E}\left[\mathrm{Fr}(S', b)\right]$$

$$= \mathbb{E}\left[\mathrm{Fr}(S, b) - \mathrm{Fr}(S', b) \mid S\right].$$

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

(2) $$\sup_{b \in \mathcal{V}} \mathbb{E}\left[\cdots\right] \leq \mathbb{E}\left[\sup_{b \in \mathcal{V}} (\cdots)\right].$$

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

$$\left| \mathrm{Fr}(S, b) - \mathrm{Fr}(S', b)\right| = \frac{1}{n}\left| \sum_{i=1}^{n} \left(\mathbb{I}_{S_i, b} - \mathbb{I}_{S'_i, b}\right)\right|$$

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 $\sigma_i \left(\mathbb{I}_{S_i, b} - \mathbb{I}_{S'_i, b}\right)$ hence

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

Inserting this into (1) we obtain

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

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

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

and, by what remarked in (2)

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

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 4.19 (the second inequality, with $m = \pi_\varphi(2n)$ and $i$ ranging over the definable subsets of $A$), for an appropriate constant $c$,

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

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

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

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

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

The Vapnik-Chervonenkis inequality is proved if we can show the assumption of Lemma 4.19 holds with $c = \sqrt{n}$.

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

As $\sigma_i \left(\mathbb{I}_{s_i,b} - \mathbb{I}_{s'_i,b}\right)$ takes values in $\{-1, 1\}$ with mean 0, by Lemma 4.9

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

and the same holds for $-\sigma_i \left(\mathbb{I}_{s_i,b} - \mathbb{I}_{s'_i,b}\right)$. $\qquad\square$

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).

**4.20 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,*

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

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

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

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

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

hence, by the Vapnik-Chervonenkis inequality 4.15

$$\leq \; \sqrt{2 \frac{\log \, \left(2 \, \pi_\varphi(n)\right)}{n}}.$$

It is clear that $\varphi(\mathcal{U}; b)_{b \in \mathcal{V}}$ has VC-dimension 1. Hence, by the Sauer-Shelah Lemma 3.8 and the inequalities proved thereafter, $\pi_\varphi(n) \leq n + 1$. The theorem follows. $\qquad\square$

# Chapter 5

## 1 Transversals and packings

A subset $A \subseteq \mathcal{U}$ is a transversal if $\varphi(A, b) \neq \varnothing$ for every $b \in \mathcal{V}$. Equivalently, if

$$\mathcal{V} = \bigcup_{a \in A} \varphi(a, \mathcal{V}).$$

The transversal number of $\varphi(x\, ; z)$, denoted by $\tau(\varphi)$, is the smallest cardinality of a transversal of $\varphi(x\, ; z)$.

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

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

A fractional multi-set $A$ over $\mathcal{U}$ is a fractional transversal if $|\varphi(A, b)| \geq 1$ for every $b \in \mathcal{V}$. The fractional transversal number of $\varphi$, denoted by $\tau^*(\varphi)$, is the infimum of the size of the fractional transversals of $\varphi$.

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

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

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

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

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

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

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

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

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

## 2  Nets

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

$\varepsilon$-Nets, like all threshold structures are mathematically awkward. The advantage is that $\varepsilon$-nets have a smaller size.

**5.4 Proposition**  *For any $k, \varepsilon > 0$, every measure on a set system $\Phi$ of VC-dimension $\leq k$, has an $\varepsilon$-net of cardinality $\leq c\,\dfrac{k}{\varepsilon}\,\ln\dfrac{1}{\varepsilon}$ where $c$ is an absolute constant.*

**Proof**  We only prove a weaker bound by observing that every $\varepsilon$-approximation is in particular an $\varepsilon$-net. Hence from Proposition **??** we obtain an $\varepsilon$-net of cardinality $\leq c\,(k/\varepsilon^2)\,\ln(k/\varepsilon)$. $\qquad\square$

**5.5 Proposition**  *Let $\Phi$ be a finite set system with VC-dimension $\leq k$. Then*
$$\tau(\Phi) \;\leq\; c\,k\,\tau^*(\Phi)\,\ln\tau^*(\Phi)$$
*where $c$ is an absolute constant.*

**Proof**  Let $r = \tau^*(\Phi)$ ad let $A$ be an optimal fractional transversal. As $\Phi$ is finite we can assume that the support of $A$ is finite. After normalizing, $A$ defines a probability measure on $\mathcal{U}$ concentrated on $\operatorname{supp} A$. Namely, $\mathbb{P}(\{x\}) = (1/r)A(x)$ for all $x \in \mathcal{U}$. By the definition of fractional transversal every set in $\Phi$ has measure at last $1/r$. Therefore every $(1/r)$-net is a transversal. By Proposition 5.4 we obtain a transversal of cardinality $\leq c\,k\,r\ln r$. $\qquad\square$

## 3  Helly-type properies

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

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

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

The following property is more robust.

**5.7** **Definition** *Let $\varphi$ be an incidence relation. We say that $\varphi$ has fractional Helly number $k$ if for every $\alpha > 0$ there is a $\beta > 0$ such that:*

*if $b_i \in \mathcal{V}$, for $i = 1, \ldots, n$, are such that*

$$\bigcap_{i \in I} \varphi(\mathcal{U}, b_i) \ \neq \ \varnothing \qquad \qquad \text{for at least } \ \alpha \binom{n}{k} \ \text{ sets } \ I \in \binom{[n]}{k}$$

*then*

$$\bigcap_{i \in J} \varphi(\mathcal{U}, b_i) \ \neq \ \varnothing \qquad \qquad \text{for some } J \subseteq [n] \text{ of cardinality} \geq \beta \, n.$$

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

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

**5.8** **Theorem (Matoušek [?M2004])** *Let $\varphi$ be an incidence relation with $\pi^*(\varphi) \in o(n^k)$. Then $\varphi$ has fractional Helly number $\leq k$.*

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

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

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

Fix some $b_i \in \mathcal{V}$, for $i = 1, \ldots, n$ and assume

$$\bigcap_{i \in I} \varphi(\mathcal{U}, b_i) \ \neq \ \varnothing \ \text{ for at least } \ \alpha \binom{n}{k} \ \text{ sets } \ I \in \binom{[n]}{k}$$

We need to show that $\bigcap_{i \in J} \varphi(\mathcal{U}, b_i) \neq \varnothing$ for some set $J \subseteq [n]$ of cardinality $\geq \beta \, n$. So, assume no such $J$ exists and reason for a contradiction. Note that we can assume that $n > 2m$ because for $n \leq 2m$ we can take take $|J| = 1$.

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

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

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

Assume the claim and proceed with the proof. We can think that the random pair in

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

We now prove the claim. There is another equivalent way to pick a random pair in $P$. First we choose at random $I \subseteq [n]$ of cardinality $k$, then we choose at random $m - k$ elements from $[n] \smallsetminus I$. By assumption, the probability that $I$ is such that $\bigcap_{i \in I} \varphi(\mathcal{U}, b_i) \neq \varnothing$ is at least $\alpha$. So, assume that $I$ is such, and fix any $a \in \bigcap_{i \in I} \varphi(\mathcal{U}, b_i)$. By the assumption that we would like to contradict, there are $< \beta n$ many $i \in [n]$ such that $\varphi(a, b_i)$. Then the probability that all $i \in J \smallsetminus I$ are such that $\neg \varphi(a, b_i)$ is at least

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

$$\geq \prod_{i=0}^{m-k} \frac{n - \beta n - m}{n - m} \geq \left( \frac{n - \beta n - m}{n - m} \right)^m$$

$$= \left( 1 - \frac{\beta n}{n - m} \right)^m$$

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

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

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

# 4 The (p,q)-theorem

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

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

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

By Proposition 5.5, $\tau(\varphi)$ is bounded by a function of $\tau^*(\varphi)$, so it is enough to bound $\tau^*(\varphi)$. By Proposition 5.3, $\tau^*(\varphi) = \nu^*(\varphi)$, so it is enough to bound $\nu^*(\varphi)$.

Let $A$ be an optimal fractional packing. By Theorem 5.3 we may assume that $A$ is rational valued. Therefore $A = (1/m) C$ where $m$ is a positive integer and $C$ is a (integral valued) multi-subset of $\mathcal{V}$.

As explained in Section **??**, after replacing each $b \in \mathcal{V}$ with $m$ copies $b_1, \ldots, b_m$ and denote the resultig set by $\mathcal{V}_m$. Replace each arc $\langle a, b \rangle$ in $\varphi$ with the arcs $\langle a, b_1 \rangle, \ldots, \langle a, b_m \rangle$. Denote by $\varphi_m$ the new incidence relation. We can assume that $C$ is a plain subset of $\mathcal{V}_m$ with the property that $|\varphi(a, C)| \leq m$ for every $a \in \mathcal{U}$.

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

A $p'$-collection of $\varphi_m$ definable sets either contains $q$ copies of the same $\varphi$-definable set or it contains $p$ distinct $\varphi$-definable sets. In both case it has a subcollection of $q$ sets with non-empty intersection. This proves Claim 1.

*Claim 2.* There is an $\alpha = \alpha(p, q) > 0$ such that for every $n$ and every $b_1, \ldots, b_n \in \mathcal{V}'$

$$\bigcap_{i \in I} \varphi(\mathcal{U}, b_i) \; \neq \; \varnothing \qquad\qquad \text{for at least } \; \alpha \binom{n}{q} \; \text{ sets } \; I \in \binom{[n]}{q}.$$

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

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

Therefore, claim 2 holds with $\alpha = \dbinom{p'}{q}^{-1}$.

*Claim 3.* There is a $\beta = \beta(p, q) > 0$ such that for every $n$ and every $b_1, \ldots, b_n \in \mathcal{V}'$ there is an $a \in \mathcal{U}$ such that $\varphi(a, b_i)$ holds for at least $\beta n$ many $b_i$.

As $\varphi'$ has the same VC-codensity as $\varphi$, Caim 3 follows immediately from Claim 2 by Helly fractional theorem.

Now, apply Claim 3 with the elements of $C$ as $b_1, \ldots, b_n$. Then $\nu^*(\varphi) = n/m$. So, from $\beta n \leq |\varphi(a, C)| \leq m$, we obtain $\nu^*(\varphi) \leq 1/\beta$. So $N = 1/\beta$ is the bound claimed by the theorem. $\qquad\square$

# References

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

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

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

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

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

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

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

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

[9] 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.