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 will 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 will 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 these 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

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

Chapter 1

Three minimax theorems

Though apparently unrelated, the three theorems in this chapter could be derived one from the other in any order. The order we choose is arbitrary.

The last two theorems are, evidently, minimax theorems. The first one is not, so 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 $B' \subseteq B$

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

(2⇒1) Reason by induction on the cardinality of B. If B is empty, the claim is clear. Now assume $|B| > 0$ and consider two cases.

a. This is the case when the inequality in 2 is strict for all $B' \subseteq B$. 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 $B' \subseteq B$ 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 $\varphi(A \smallsetminus A'\,;\,B \smallsetminus B')$. Once this claim is proved, we apply the induction hypothesis to obtain sets of distinct representatives for these two relations and note that 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. □

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.

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 $(A' \times B) \cup (A \times B')$. 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 $a \in A \smallsetminus A'$.* □

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 $b \in B'' \smallsetminus B'$. Therefore all elements of A'' are in A' but for at most $|B'|$ elements. Hence $|A''| \leq |A'| + |B'|$ 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_1 is a matching for some $B_1 \subseteq B'$. 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_2, B_2 is a matching for some $A_2 \subseteq A'$. Then $(A_1 \cup A_2)$, $(B_1 \cup B_2)$ is a matching of $\varphi(A\,;B)$. The cardinality of this maching is $|A_1| + |A_2| = |B_1| + |A_2| \leq |B'| + |A'|$.

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'. □

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 $\mathfrak{U}$. An antichain is a set $A \subseteq \mathfrak{U}$ such that $a \nless a'$ for every $a, a' \in A$. A chain is a set A such that $a < a' \;\vee\; a' < a$ for every distinct $a, a' \in A$.

1.4 Dilworth's Theorem *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 antichains. A chain can contain at most one element of A', hence $|A'| \leq k$.

($\geq$) Let $A \smallsetminus A' \subseteq A$ be an antichain (for uniformity with the notation in Kőnig's Theorem, here we denote by A' the complement of the chain). We prove that there is a partition $A_1, \ldots, A_k$ into chains for some $k \leq |A \smallsetminus A'|$.

Let $\mathcal{V}$ be a disjoint copy of $\mathcal{U}$. For $a, b \in \mathcal{U}, \mathcal{V}$ let $\varphi(a\,; b)$ hold if $a <$ (the copy in $\mathcal{U}$ of) b. Let $B' \subseteq B \subseteq \mathcal{V}$ be the copy of $A' \subseteq A \subseteq \mathcal{U}$. Then $\varphi(A\,; b) \subseteq A'$ for every $b \in B \smallsetminus B'$. By Kőnig's Theorem there is a matching A'', B'' of cardinality $|A''| = |B''| \geq |A' \cup B'| = |A'|$.

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 $\leq |A| \smallsetminus |A''| \leq |A| \smallsetminus |A'|$. $\square$

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 elements $b, b' \in \mathcal{V}$ such that $\varphi(A\,;b) \subset \varphi(A\,;b')$. Antichains are also called Sperner systems.

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

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

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 $\{0,\ldots,n\}$. Then

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

As for every $k \leq n$

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

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 $A \subseteq \mathfrak{U}$ have cardinality n, finite. If $\varphi(A\,;b)_{b\in\mathcal{V}}$ is an antichain then*

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

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

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

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

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

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

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

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 $A' \subseteq A$

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

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 $A \subseteq \mathcal{U}$ 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*

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

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

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

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

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

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

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.

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

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

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

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

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

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. $\square$

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. $\square$

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 [4], 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 [7], Sauer [6], and Vapnik-Chervonenkis [8] 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 [5].

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

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

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 $A' = A \setminus \{a\}$. Define

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

Notice that

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

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

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

By induction hypothesis,

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

and

$$\begin{aligned}\big|\psi(A',b)_{b\in\mathcal{V}}\big| \quad &\leq \quad \Big|\{C \subseteq A' \ : \ C \text{ is shattered by } \psi(x\,;z)\}\Big|.\\ &= \quad \Big|\{C \subseteq A' \ : \ C \cup \{a\} \text{ is shattered by } \varphi(x\,;z)\}\Big| \qquad (2)\end{aligned}$$

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,

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

so the lemma follows. □

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

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

□

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

Chapter 4

Law(s) of large numbers

Quoting some 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 "frequentists". They wrote books and indoctrinated generations of confused students.

1 Inequalities

Throughout this and the next section we work with a given sample space $\Omega, \mathbb{P}$. For simplicity, the following two propositions are proved for finite Ω, but they are easily seen to hold in general. The finiteness will be an essential hypothesis In Section 3.

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

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

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

4.1 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 Ω 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.

4.2 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 Ω is finite. (The theorem holds in general, but we only need the finite case.) Define $A = \{a \in \Omega \,:\, X(a) \geq \varepsilon\}$.

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

$\square$

4.3 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.

4.4 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.

4.5 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 $t > 0$

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

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

4.6 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,

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

taking the Taylor series expansion of the r.h.s. at $t = 0$ we obtain (the first and second derivatives vanish at 0; the second derivative is always $\leq 1/4$)

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

□

4.7 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 $\varepsilon > 0$

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

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.

$$\begin{aligned} \mathbb{P}\Big(M \geq \varepsilon\Big) \;&\leq\; e^{-t\varepsilon}\, \mathbb{E}\Big[e^{tM}\Big] && \text{by Chernoff's method (4.5)} \\ &=\; e^{-t\varepsilon} \prod_{i=1}^{n} \mathbb{E}\Big[e^{t\,(X_i - \mu_i)}\Big] && \text{by independence.} \\ &\leq\; e^{-t\varepsilon} \prod_{i=1}^{n} \exp\Big(\frac{t^2 d^2}{8}\Big) && \text{by Hoeffding's Lemma (4.6).} \\ &=\; \exp\Big(\frac{n\, t^2 d^2}{8} - t\varepsilon\Big) \end{aligned}$$

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 a clever trick called *symmetrization* which in the following section is applied in a more complex setting.

First we need the following lemma.

4.8 Lemma *Let σ be a random sign variable (a.k.a. Rademacher random variable). That is, $\sigma \in \{-1, 1\}$ with uniform distribution. 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 $x = 0$

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

□

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

$$\begin{aligned}
&= \mathbb{E}\left[e^{t\sigma(X-X')}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[e^{t\sigma(X-X')} \mid X, X'\right]\right] \\
&\leq \mathbb{E}\left[e^{t^2(X-X')^2/2}\right] && \text{by Lemma 4.8} \\
&\leq e^{t^2 d^2/2} && \text{because } |X - X'| \leq d.
\end{aligned}$$

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

2 Two Weak Laws of Large Numbers

A sample s is a sequence $s_0, \dots, s_{n-1}$ of elements of Ω. Its length $|s| = n$ is also called size or dimension. We write range(s) for the set $\{s_0, \dots, s_{n-1}\}$. 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 Ω, namely for any $A \subseteq \Omega$ we define

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

Let $S = S_1, \dots, S_n$ be independent random elements of Ω, that is, random variables such that $\mathbb{P}(S_i \in A) = \mathbb{P}(A)$ for every $A \subseteq \Omega$. Write $\mathbb{I}_A$ for the indicator function of A. Then $\mathbb{I}_A \circ S_i$ as a Bernoulli random variable with probability of success $\mathbb{P}(A)$. Moreover

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

4.10 Weak Law of Large Numbers *For every event $A \subseteq \Omega$ and every $n > 0$*

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

Proof Let $S_1, \dots, S_n$ be independent random elements of Ω. 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

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

which proves the theorem. □

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

4.11 Corollary *For every $A \subseteq \Omega$ and every $\varepsilon > 0$ there is a sample s of size*

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

such that

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

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

$$1 \quad > \quad \frac{1}{n\varepsilon^2}$$ □

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

4.12 Weak Law of Large Numbers (with exponential bound) *For every event $A \subseteq \Omega$ and every $n > 0$*

$$2e^{-2n\varepsilon^2} \quad \geq \quad \mathbb{P}\Big(s \in \Omega^n \;:\; \big|\mathrm{Fr}(s, A) - \mathbb{P}(A)\big| \geq \varepsilon\Big).$$

Proof Let $S_1, \dots, S_n$ be independent random elements of Ω. Define

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

As $\mathbb{E}[\mathbb{I}_{A}\circ S_i] = \mathbb{P}(A)$, 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. □

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 4.11.

4.13 Corollary *For every $A \subseteq \Omega$ and every $\varepsilon > 0$ there is a sample s of size*

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

such that

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

3 The Uniform Law of Large Numbers

In this section we work with a fixed formula $\varphi(x\,;z)$ and with the family of definable subsets $\varphi(\Omega\,;b)_{b\in\mathcal{V}}$ of a given finite sample space $\Omega, \mathbb{P}$. It is convenient to introduce e some abbreviations

$$\begin{aligned}\mathbb{P}(b) &= \mathbb{P}\Big(\varphi(\Omega\,;b)\Big)\\ \mathrm{Fr}(s,b) &= \mathrm{Fr}\,\Big(s,\varphi(\Omega\,;b)\Big)\end{aligned}$$

An ε-approximation is a sample s such that

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

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 [8]. Following Devroye and Lugosi [2], we prove a slightly better bound.

4.14 Vapnik-Chervonenkis inequality *Let $\pi_\varphi(n)$ be the shatter function of $\varphi(\Omega\,;b)_{b\in\mathcal{V}}$. Let $S = S_1,\ldots,S_n$ be independent random elements of Ω. Then, for every $b \in \mathcal{V}$*

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

Before embarking of the proof of the theorem above, we prove the following easy but mysterious lemma which is of independent interest.

4.15 Lemma *Let $X_1,\ldots,X_m$ real valued random variable (the sample space $\Omega, \mathbb{P}$ is arbitrary). Let c be such that*

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

Then

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

Proof By Jensen's inequality,

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

The lemma is obtained taking the logarithm and substituting $t=\dfrac{\sqrt{2\log m}}{c}$. □

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

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

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

$$\sup_{b\in\mathcal{V}}\mathbb{E}\big[\cdots\big] \;\le\; \mathbb{E}\big[\sup_{b\in\mathcal{V}}(\cdots)\big].$$

Write $\mathbb{I}_b$ for the indicator function of $\varphi(\Omega)$. Then

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

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

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

Inserting this into (1) we obtain

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

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

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

Observe that once s,s' is fixed, $\sup_{b\in\mathcal{V}}$ is actually a maximum among $m=\pi_\varphi(2n)$ sets, that is,the definable subsets of $\{s_1,\dots,s_n,s'_1,\dots,s'_n\}$. Then, by Lemma 4.15, for the appropiate c we obtain

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

Finally, as the l.h.s. does not depend on s, s'

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

We are only left with proving that the assumption of Lemma 4.15 holds with $c = \sqrt{n}$.

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

As $\sigma_i\left(\mathbb{I}_b\, s_i - \mathbb{I}_b\, s'_i\right)$ takes values in $\{-1, 1\}$ with mean 0, by Lemma 4.8

$$\leq e^{nt^2/2}. \qquad \square$$

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

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

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

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

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

[8] 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.