Weak Convergence and Empirical Processes

Chapter 2.14: Maximal Inequalities and Tail Bounds

Bart Eggen

Leiden University

June 15, 2020

What's on the menu today?

Tail bounds for the empirical process

$X_1, \ldots, X_n \sim P, \quad \mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$

Main goal

$$P^* \left(\|\mathbb{G}_n\|_{\mathcal{F}} > t \right) \leq \ \textbf{?}$$

Remember

$$\|\mathbb{G}_n\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\mathbb{G}_n f|$$

Tail bounds for the empirical process

$X_1, \ldots, X_n \sim P, \quad \mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$

Main goal

$$P^* \left(\|\mathbb{G}_n\|_{\mathcal{F}} > t \right) \leq \ \textbf{?}$$

Methods

- Finite uniform-entropy integral (2.5.1)

$$\sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}\mathcal{F}, L_2(Q)\right)}\, d\varepsilon$$

- Finite bracketing integral (2.5.2)

$$\int_0^\delta \sqrt{1 + \log N_{[\,]}\left(\varepsilon \|F\|, \mathcal{F}, \|\cdot\|\right)}\, d\varepsilon$$

Tail bounds for the empirical process

$X_1, \ldots, X_n \sim P, \quad \mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$

Main goal

$$P^* \left(\|\mathbb{G}_n\|_{\mathcal{F}} > t \right) \leq \ \textbf{?}$$

Methods

- Uniform bounded classes (2.6)

$$\sup_Q N(\varepsilon, \mathcal{F}, L_2(Q)) \leq \left(\frac{K}{\varepsilon} \right)^V, \quad \text{for every } 0 < \varepsilon < K$$

- Uniform bounded classes (2.7)

$$N_{[\,]}(\varepsilon, \mathcal{F}, L_2(P)) \leq \left(\frac{K}{\varepsilon} \right)^V, \quad \text{for every } 0 < \varepsilon < K$$

Tail Bounds

$$\|X\|_\psi = \inf\left\{C > 0 : E\psi\left(\frac{|X|}{C}\right) \leq 1\right\}$$

Markov's inequality

Let X be a random variable, ψ a nondecreasing, convex function with $\psi(0) = 0$, then by Markov's inequality

$$P(|X| > t) \leq P(\psi(|X|/\|X\|_\psi) \geq \psi(t/\|X\|_\psi)) \leq \frac{1}{\psi(t/\|X\|_\psi)}$$

Tail Bounds

$$\|X\|_\psi = \inf\left\{C > 0 : E\psi\left(\frac{|X|}{C}\right) \leq 1\right\}$$

Markov's inequality

Let X be a random variable, ψ a nondecreasing, convex function with $\psi(0) = 0$, then by Markov's inequality

$$P\left(|X| > t\right) \leq P\left(\psi\left(|X|/\|X\|_\psi\right) \geq \psi\left(t/\|X\|_\psi\right)\right) \leq \frac{1}{\psi\left(t/\|X\|_\psi\right)}$$

L_p-norm

For the L_p-norm, we have $\psi(x) = x^p$ and therefore

$$P\left(|X| > t\right) \leq \frac{\|X\|_p^p}{t^p}$$

Tail Bounds

$$\|X\|_\psi = \inf\left\{C > 0 : E\psi\left(\frac{|X|}{C}\right) \le 1\right\}$$

Markov's inequality

Let X be a random variable, ψ a nondecreasing, convex function with $\psi(0) = 0$, then by Markov's inequality

$$P(|X| > t) \le P\left(\psi\left(|X|/\|X\|_\psi\right) \ge \psi\left(t/\|X\|_\psi\right)\right) \le \frac{1}{\psi\left(t/\|X\|_\psi\right)}$$

$\|\cdot\|_{\psi_p}$-norm

With $\psi_p(x) = e^{x^p} - 1$ we have for t sufficiently large

$$P(|X| > t) \le \frac{1}{e^{t^p/\|X\|_{\psi_p}^p} - 1} \le 2e^{-t^p/\|X\|_{\psi_p}}$$

Uniform Entropy Integrals

Definition

For $\mathcal{F}$ a class of measurable functions with measurable envelope function F, the *uniform entropy integral* is defined by

$$J(\delta, \mathcal{F} \mid F, L_2) := \sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\varepsilon,$$

where the supremum is taken over all discrete probability measures Q with $\|F\|_{Q,2} > 0$.

Uniform Entropy Integrals

Theorem 2.14.1

Let $\mathcal{F}$ be a P-measurable class of measurable functions with measurable envelope function F. Then, for $p \geq 1$,

$$\left\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \right\|_{P,p} \lesssim \left\| J\left(\theta_n, \mathcal{F}|F, L_2\right) \|F\|_n \right\|_{P,p} \lesssim J\left(1, \mathcal{F}|F, L_2\right) \|F\|_{P,2\vee p}.$$

Here $\theta_n = \left\| \|f\|_n \right\|_{\mathcal{F}}^* / \|F\|_n$, where $\|\cdot\|_n$ is the $L_2(\mathbb{P}_n)$-seminorm and the inequalities are valid up to constants depending only on the p involved in the statement.

$J(\delta, \mathcal{F} \mid F, L_2) := \sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\varepsilon$

$\left\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \right\|_{P,p} \lesssim \left\| J\left(\theta_n, \mathcal{F} | F, L_2\right) \|F\|_n \right\|_{P,p}$

Proof idea

- Remember chapter 2.2 with maximal inequalities and Orlicz-norms
- Use symmetrization
- Work out the details

$J(\delta, \mathcal{F} \mid F, L_2) := \sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\varepsilon$

$$\left\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \right\|_{P,p} \lesssim \left\| J\left(\theta_n, \mathcal{F} | F, L_2\right) \|F\|_n \right\|_{P,p}$$

Proof idea

- Remember chapter 2.2 with maximal inequalities and Orlicz-norms
- Use symmetrization
- Work out the details

Chapter 2.2

$$\|X\|_p \leq p! \|X\|_{\psi_1} \leq p! (\log 2)^{-1/2} \|X\|_{\psi_2}$$

$J(\delta, \mathcal{F} \mid F, L_2) := \sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\varepsilon$

$\left\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \right\|_{P,p} \lesssim \left\| J\left(\theta_n, \mathcal{F} | F, L_2\right) \|F\|_n \right\|_{P,p}$

Orlicz-norm

Let ψ be a non-decreasing, convex function with $\psi(0) = 0$, then

$$\left\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \right\|_\psi = \inf \left\{ C > 0 : E\psi\left(\frac{\|\mathbb{G}_n\|_{\mathcal{F}}^*}{C} \right) \leq 1 \right\}.$$

We will use $\psi(x) = \psi_2(x) = e^{x^2} - 1$.

$J(\delta, \mathcal{F} \mid F, L_2) := \sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\varepsilon$

$\left\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \right\|_{P,p} \lesssim \left\| J\left(\theta_n, \mathcal{F} | F, L_2\right) \|F\|_n \right\|_{P,p}$

Orlicz-norm

Let ψ be a non-decreasing, convex function with $\psi(0) = 0$, then

$$\left\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \right\|_\psi = \inf \left\{ C > 0 : E\psi\left(\frac{\|\mathbb{G}_n\|_{\mathcal{F}}^*}{C}\right) \leq 1 \right\}.$$

We will use $\psi(x) = \psi_2(x) = e^{x^2} - 1$.

To show

$$\left\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \right\|_{\psi_2} \leq 2 \left\| \|\mathbb{G}_n^o\|_{\mathcal{F}}^* \right\|_{\psi_2}$$

$J(\delta, \mathcal{F} \mid F, L_2) := \sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\varepsilon$

$\left\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \right\|_{P,p} \lesssim \left\| J\left(\theta_n, \mathcal{F} | F, L_2\right) \|F\|_n \right\|_{P,p}$

Orlicz-norm

Let ψ be a non-decreasing, convex function with $\psi(0) = 0$, then

$$\left\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \right\|_\psi = \inf \left\{ C > 0 : E\psi\left(\frac{\|\mathbb{G}_n\|_{\mathcal{F}}^*}{C} \right) \leq 1 \right\}.$$

We will use $\psi(x) = \psi_2(x) = e^{x^2} - 1$.

To show

$$\left\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \right\|_{\psi_2} \leq 2 \left\| \|\mathbb{G}_n^o\|_{\mathcal{F}}^* \right\|_{\psi_2}$$

Remember

$$\mathbb{G}_n^o = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i f(X_i)$$

Assume $\|\mathbb{G}_n^o\|_{\mathcal{F}}^* = \|\mathbb{G}_n^o\|_{\mathcal{F}}$. See p 109-110 in old book.

To show

$$\Big\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \Big\|_{\psi_2} \leq 2 \Big\| \|\mathbb{G}_n^o\|_{\mathcal{F}}^* \Big\|_{\psi_2}$$

$J(\delta, \mathcal{F} \mid F, L_2) := \sup_Q \int_0^\delta \sqrt{1 + \log N(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q))} d\varepsilon$

$\left\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \right\|_{P,p} \lesssim \left\| J(\theta_n, \mathcal{F}|F, L_2) \|F\|_n \right\|_{P,p}$

Theorem 2.2.4

Let ψ be a convex, nondecreasing, nonzero function with $\psi(0) = 0$ and $\limsup_{x,y\to\infty} \psi(x)\psi(y)/\psi(cxy) < \infty$, for some constant c. Let $\{X_t : t \in T\}$ be a seperable stochastic process with

$$\|X_s - X_t\|_\psi \leq Cd(s, t), \quad \text{for every } s, t,$$

for some semimetric d on T and a constant C. Then, for any $\eta, \delta > 0$,

$$\left\| \sup_{d(s,t)\leq\delta} |X_s - X_t| \right\|_\psi \leq K \left[\int_0^\eta \psi^{-1}(D(\varepsilon, d))\, d\varepsilon + \delta\psi^{-1}\left(D^2(\eta, d)\right) \right],$$

for a constant K depending on ψ and C only.

Remember

$$N(\varepsilon, d) \leq D(\varepsilon, d) \leq N(\varepsilon/2, d)$$

To show

Given $X_1, \ldots, X_n$, $\mathbb{G}_n^o$ is sub-Gaussian w.r.t. the $L_2(\mathbb{P}_n)$ norm.
Consequence

$$\|\mathbb{G}_n^o f - \mathbb{G}_n^o g\|_{\psi_2} \leq \sqrt{6} d_{L^2(\mathbb{P}_n)}(f, g), \quad \text{for every } f, g \in \mathcal{F}.$$

$J(\delta, \mathcal{F} \mid F, L_2) := \sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\varepsilon$

$\left\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \right\|_{P,p} \lesssim \left\| J\left(\theta_n, \mathcal{F} | F, L_2\right) \|F\|_n \right\|_{P,p}$

Consequence

$$\left\| \|\mathbb{G}_n^0\|_{\mathcal{F}} \right\|_{\psi_2|X} \leq K \int_0^{\operatorname{diam} \mathcal{F}} \psi_2^{-1}\left(D\left(\varepsilon, \mathcal{F}, L_2\left(\mathbb{P}_n\right)\right)\right) d\varepsilon$$

$J(\delta, \mathcal{F} \mid F, L_2) := \sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\varepsilon$

$\left\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \right\|_{P,p} \lesssim \left\| J\left(\theta_n, \mathcal{F} | F, L_2\right) \|F\|_n \right\|_{P,p}$

Consequence

$$\left\| \|\mathbb{G}_n^0\|_{\mathcal{F}} \right\|_{\psi_2|X} \leq K \int_0^{\operatorname{diam} \mathcal{F}} \psi_2^{-1}\left(D\left(\varepsilon, \mathcal{F}, L_2\left(\mathbb{P}_n\right)\right)\right) d\varepsilon$$

$J(\delta, \mathcal{F} \mid F, L_2) := \sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\varepsilon$

$\left\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \right\|_{P,p} \lesssim \left\| J\left(\theta_n, \mathcal{F} | F, L_2\right) \|F\|_n \right\|_{P,p}$

Inequality

$$\left\| \|\mathbb{G}_n^0\|_{\mathcal{F}} \right\|_{\psi_2 | X} \lesssim J\left(\theta_n, \mathcal{F} | F, L_2\right) \|F\|_n$$

Result

$$\left\| \|\mathbb{G}_n\|^*_{\mathcal{F}} \right\|_{P,p} \lesssim J\left(1, \mathcal{F}|F, L_2\right) \|F\|_{P,2\vee p}$$

Additional bounds

$J(\delta, \mathcal{F} \mid F, L_2) := \sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\varepsilon$

Theorem 2.14.2

Let $\mathcal{F}$ be a P-measurable class of measurable functions with envelope function $F \leq 1$ and such that $\mathcal{F}^2$ is P-measurable. If $Pf^2 < \delta^2 PF^2$, for everyt f and some $\delta \in (0, 1)$, then

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J\left(\delta, \mathcal{F}, L_2\right) \left(1 + \frac{J\left(\delta, \mathcal{F}, L_2\right)}{\delta^2 \sqrt{n} \|F\|_{P,2}}\right) \|F\|_{P,2}$$

Additional bounds

$J(\delta, \mathcal{F} \mid F, L_2) := \sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\varepsilon$

Theorem 2.14.2

Let $\mathcal{F}$ be a P-measurable class of measurable functions with envelope function $F \leq 1$ and such that $\mathcal{F}^2$ is P-measurable. If $Pf^2 < \delta^2 PF^2$, for everyt f and some $\delta \in (0,1)$, then

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J(\delta, \mathcal{F}, L_2) \left(1 + \frac{J(\delta, \mathcal{F}, L_2)}{\delta^2 \sqrt{n} \|F\|_{P,2}}\right) \|F\|_{P,2}$$

Question?

Not accurate when δ is 'very small' relative to n and the entropy integral.

Additional bounds

$J(\delta, \mathcal{F} \mid F, L_2) := \sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\varepsilon$

Theorem 2.14.7

Let $\mathcal{F}$ be a P-measurable class of measurable functions with envelope function F such that $PF^{(4p-2)/(p-1)} < \infty$ for some $p > 1$ and such that $\mathcal{F}^2$ and $\mathcal{F}^4$ are P-measurable. If $Pf^2 < \delta^2 PF^2$ for every f and some $\delta \in (0, 1)$, then

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J\left(\delta, \mathcal{F}, L_2\right) \left(1 + \frac{J\left(\delta^{1/p}, \mathcal{F}, L_2\right)}{\delta^2 \sqrt{n}} \frac{\|F\|_{P, \frac{4p-2}{p-1}}^{2-1/p}}{\|F\|_{P,2}^{2-1/p}}\right)^{\frac{p}{2p-1}} \|F\|_{P,2}$$

Additional bounds

$J(\delta, \mathcal{F} \mid F, L_2) := \sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\varepsilon$

Theorem 2.14.8

Let $\mathcal{F}$ be a P-measurable class of measurable functions with envelope function F such that $Pe^{F^{p+\rho}} < \infty$ for some $p, \rho > 0$ and such that $\mathcal{F}^2$ and $\mathcal{F}^4$ are P-measurable. If $Pf^2 < \delta^2 PF^2$ for every f and some $\delta \in (0, 1)$, then for a constant c depending on p, PF^2, PF^4 and $Pe^{F^{p+\rho}}$,

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \leq cJ\left(\delta, \mathcal{F}, L_2\right) \left(1 + \frac{J\left(\delta (\log(1/\delta))^{1/p}, \mathcal{F}, L_2\right)}{\delta^2 \sqrt{n}}\right)$$

Bracketing integrals

Definition

The *bracketing integral* of a class of functions $\mathcal{F}$ with measurable envelope function F relative to a given norm is given by

$$J_{[\,]}(\delta, \mathcal{F}|F, \|\cdot\|) = \int_0^\delta \sqrt{1 + \log N_{[\,]}(\varepsilon\|F\|, \mathcal{F}, \|\cdot\|)}d\varepsilon$$

Bracketing integrals

$J_{[\,]}(\delta, \mathcal{F}|F, \|\cdot\|) = \int_0^\delta \sqrt{1 + \log N_{[\,]}(\varepsilon\|F\|, \mathcal{F}, \|\cdot\|)}d\varepsilon$

Theorem 2.14.15

Let $\mathcal{F}$ be a class of measurable functions with measurable envelope function F. For given $\eta > 0$, set

$$a(\eta) = \eta\|F\|_{P,2}/\sqrt{1 + N_{[\,]}\left(\eta\|F\|_{P,2}, \mathcal{F}, L_2(P)\right)}.$$

Then for every $\eta > 0$,

$$\Big\|\|\mathbb{G}_n\|_{\mathcal{F}}^*\Big\|_{P,1} \lesssim J_{[\,]}(\eta,\mathcal{F}|F, L_2(P))\|F\|_{P,2} + \sqrt{n}PF\left\{F > \sqrt{n}a(\eta)\right\} + \Big\|\|f\|_{P,2}\Big\|_{\mathcal{F}}\sqrt{1 + \log N_{[\,]}\left(\eta\|F\|_{P,2}, \mathcal{F}, L_2(P)\right)}.$$

Consequently, if $\|f\|_{P,2} < \delta\|F\|_{P,2}$ for every $f \in \mathcal{F}$, then

$$\Big\|\|\mathbb{G}_n\|_{\mathcal{F}}^*\Big\|_{P,1} \lesssim J_{[\,]}\left(\delta, \mathcal{F}|F, L_2(P)\right)\|F\|_{P,2} + \sqrt{n}PF\left\{F > \sqrt{n}a(\delta)\right\}.$$

Bracketing integrals

$J_{[\,]}(\delta, \mathcal{F}|F, \|\cdot\|) = \int_0^\delta \sqrt{1 + \log N_{[\,]}(\varepsilon\|F\|, \mathcal{F}, \|\cdot\|)}\,d\varepsilon$

Theorem 2.14.15

For any class $\mathcal{F}$,

$$\big\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \big\|_{P,1} \lesssim J_{[\,]}(1, \mathcal{F}|F, L_2(P))\, \|F\|_{P,2}$$

Additional bounds

$J_{[\,]}(\delta, \mathcal{F}|F, \|\cdot\|) = \int_0^\delta \sqrt{1 + \log N_{[\,]}(\varepsilon\|F\|, \mathcal{F}, \|\cdot\|)}d\varepsilon$

Theorem 2.14.16

Let $\mathcal{F}$ be a class of measurable functions such that $Pf^2 < \delta^2 PF^2$ and $\|f\|_\infty \leq 1$ for every $f \in \mathcal{F}$. Then

$$E_P^*\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J_{[\,]}(\delta, \mathcal{F}|F, L_2(P)) \left(1 + \frac{J_{[\,]}(\delta, \mathcal{F}|F, L_2(P))}{\delta^2\sqrt{n}\|F\|_{P,2}}\right) \|F\|_{P,2}$$

Additional bounds

$J_{[\,]}(\delta, \mathcal{F}|F, \|\cdot\|) = \int_0^\delta \sqrt{1 + \log N_{[\,]}(\varepsilon\|F\|, \mathcal{F}, \|\cdot\|)}\,d\varepsilon$

Bernstein "norm"

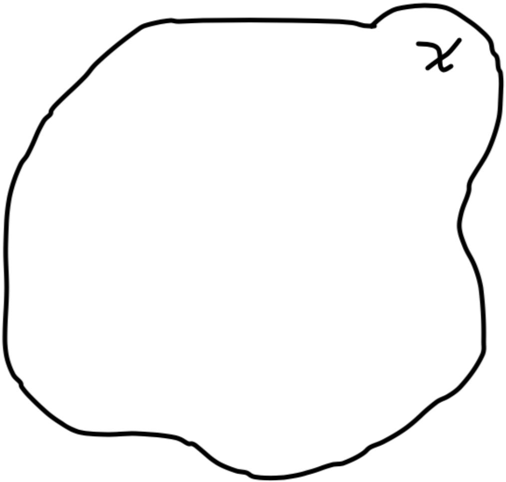
$$\|f\|_{P,B} := \left(2P\left(e^{|f|} - 1 - |f|\right)\right)^{1/2}$$

Theorem 2.14.17

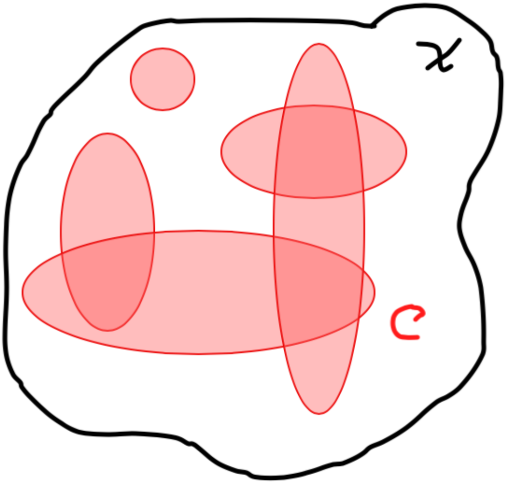
Let $\mathcal{F}$ be a class of measurable functions such that $\|f\|_{P,B} \leq \delta\|F\|_{P,B}$ for every $f \in \mathcal{F}$. Then

$$E_P^*\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J_{[\,]}\left(\delta, \mathcal{F}|F, \|\cdot\|_{P,B}\right)\left(1 + \frac{J_{[\,]}\left(\delta, \mathcal{F}|F, \|\cdot\|_{P,B}\right)}{\delta^2\sqrt{n}\|F\|_{P,B}}\right)\|F\|_{P,B}$$

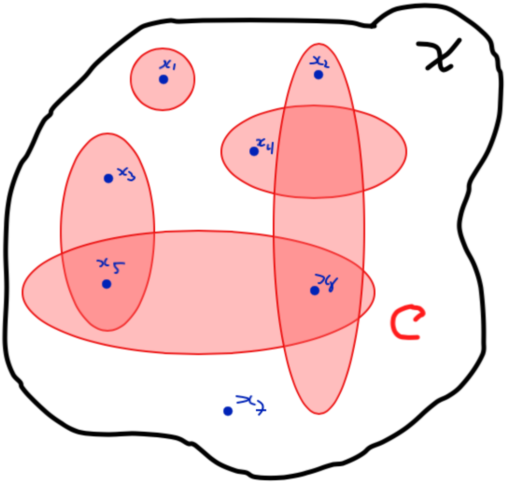
VC-Classes of Sets



VC-Classes of Sets

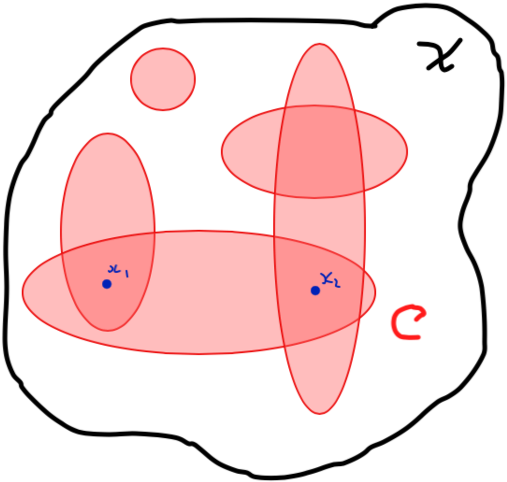


VC-Classes of Sets



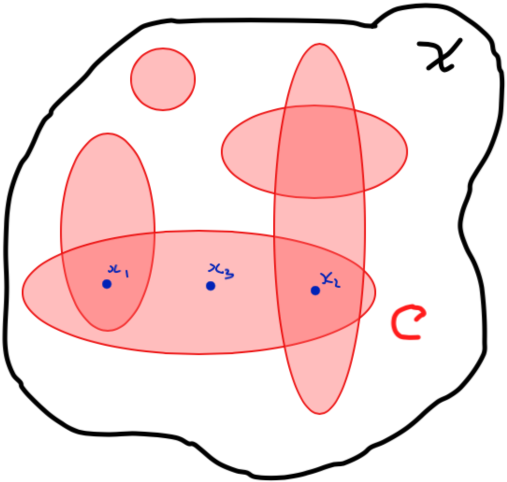
$\mathcal{C}$ **picks out** a subset $A \subseteq \{x_1, \ldots, x_n\}$ if there exists a $C \in \mathcal{C}$ with $C \cap \{x_1, \ldots, X_n\} = A$

VC-Classes of Sets



$\mathcal{C}$ **shatters** $\{x_1, \ldots, x_n\}$ if each of its 2^n subsets can be picked out

VC-Classes of Sets



The **VC-index** $V(\mathcal{C})$ of the class $\mathcal{C}$ is the smallest n for which no set of size n is shattered by $\mathcal{C}$

VC-Classes of Sets

VC-index

The **VC-index** of $\mathcal{C}$ is defined as

$$V(\mathcal{C}) = \inf\{n : \max_{x_1,\ldots,x_n} \Delta_n(\mathcal{C}, x_1, \ldots, x_n) < 2^n\},$$

where

$$\Delta_n(\mathcal{C}, x_1, \ldots, x_n) = \#\left\{C \cap \{x_1, \ldots, x_n\} : C \in \mathcal{C}\right\}.$$

A collection of measurable sets $\mathcal{C}$ is called a **VC-class** if its index is finite.

VC-Classes of Sets

VC-index

The **VC-index** of $\mathcal{C}$ is defined as

$$V(\mathcal{C}) = \inf\{n : \max_{x_1,\ldots,x_n} \Delta_n(\mathcal{C}, x_1, \ldots, x_n) < 2^n\},$$

where

$$\Delta_n(\mathcal{C}, x_1, \ldots, x_n) = \#\left\{C \cap \{x_1, \ldots, x_n\} : C \in \mathcal{C}\right\}.$$

A collection of measurable sets $\mathcal{C}$ is called a **VC-class** if its index is finite.

Convention

The infimum over the empty set is taken to be infinity

Subgraphs

Subgraph

The *subgraph* of a function $f : \mathcal{X} \to \mathbb{R}$ is the subset of $\mathcal{X} \times \mathbb{R}$ given by

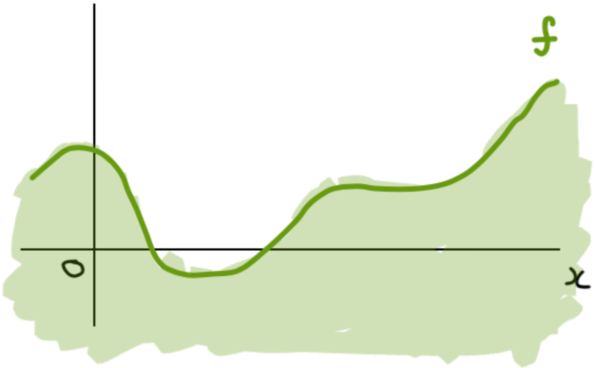
$$\{(x, t) : t < f(x)\}$$

Subgraphs

Subgraph

The *subgraph* of a function $f : \mathcal{X} \to \mathbb{R}$ is the subset of $\mathcal{X} \times \mathbb{R}$ given by

$$\{(x, t) : t < f(x)\}$$



Subgraphs

Subgraph

The *subgraph* of a function $f : \mathcal{X} \to \mathbb{R}$ is the subset of $\mathcal{X} \times \mathbb{R}$ given by

$$\{(x, t) : t < f(x)\}$$

VC-Class

A collection $\mathcal{F}$ of measurable functions is called a *VC-class*, if the collection of all subgraphs of the functions in $\mathcal{F}$ forms a VC-class of sets in $\mathcal{X} \times \mathbb{R}$.

Combine results

Theorem 2.14.1

Let $\mathcal{F}$ be a P-measurable class of measurable functions with measurable envelope function F. Then, for $p \geq 1$,

$$\left\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \right\|_{P,p} \lesssim \sup_Q \int_0^1 \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)}\, d\varepsilon \cdot \|F\|_{P,2\vee p}.$$

Theorem 2.6.7

For a VC-class of functions with measurable envelope function F and $r \geq 1$, one has for any probability measure Q with $\|F\|_{Q,r} > 0$,

$$N\left(\epsilon \|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})} \left(\frac{1}{\epsilon}\right)^{rV(\mathcal{F})},$$

for a universal constant K and $0 < \epsilon < 1$.

$$N\left(\epsilon\|F\|_{Q,2},\mathcal{F},L_2(Q)\right)\leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\tfrac{1}{\epsilon}\right)^{2V(\mathcal{F})}$$

$$\left\|\|\mathbb{G}_n\|_{\mathcal{F}}^*\right\|_{P,p}\lesssim\sup_Q\int_0^1\sqrt{1+\log N\left(\varepsilon\|F\|_{Q,2},\mathcal{F},L_2(Q)\right)}d\varepsilon\cdot\|F\|_{P,2\vee p}$$

Explicit and small constants

Lemma 2.14.20

For any P-measurable VC-class $\mathcal{C}$ of sets and $2n \geq V$,

$$\mathbb{E}^* \|\mathbb{G}_n\|_{\mathcal{C}} \leq 2\sqrt{2V(\mathcal{C})\left(1 + \frac{\log n}{V(\mathcal{C})}\right)}$$

Refine ψ_p

General Orlicz norm bounds

Theorem 2.14.21

Let $\mathcal{F}$ be a class of measurable functions with measurable envelope function F. Then

$$\big\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \big\|_{P,p} \lesssim \big\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \big\|_{P,1} + n^{-1/2+1/p} \|F\|_{P,p} \qquad (p \geq 2)$$

$$\big\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \big\|_{P,\psi_p} \lesssim \big\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \big\|_{P,1} + n^{-1/2}(1 + \log n)^{1/p} \|F\|_{P,\psi_p} \qquad (0 < p \leq 1)$$

$$\big\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \big\|_{P,\psi_p} \lesssim \big\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \big\|_{P,1} + n^{-1/2+1/q} \|F\|_{P,\psi_p} \qquad (1 < p \leq 2)$$

Consequence

Tail bounds

$$\Big\| \|\mathbb{G}_n\|_{\mathcal{F}}^* \Big\|_{P,\psi_p} \lesssim \|F\|_{P,\psi_p}, \quad (0 < p \leq 2)$$

Uniformly Bounded Classes

Assumption

$$0 \le f \le 1, \quad \text{for every } f \in \mathcal{F}$$

Uniformly Bounded Classes

Assumption

$$0 \leq f \leq 1, \quad \text{for every } f \in \mathcal{F}$$

Entropy integral

$$\sup_Q \int_0^1 \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\varepsilon$$

Uniformly Bounded Classes

Assumption

$$0 \leq f \leq 1, \quad \text{for every } f \in \mathcal{F}$$

Entropy integral

$$\sup_Q \int_0^1 \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\varepsilon$$

Cases

- $\sup_Q N(\varepsilon, \mathcal{F}, L_2(Q)) \leq \left(\frac{K}{\varepsilon}\right)^V, \quad$ for every $0 < \varepsilon < K$,
- $N_{[\,]}(\varepsilon, \mathcal{F}, L_2(P)) \leq \left(\frac{K}{\varepsilon}\right)^V, \quad$ for every $0 < \varepsilon < K$.

$\sup_Q N(\varepsilon, \mathcal{F}, L_2(Q)) \leq \left(\frac{K}{\varepsilon}\right)^V$

$N_{[\,]}(\varepsilon, \mathcal{F}, L_2(P)) \leq \left(\frac{K}{\varepsilon}\right)^V$

Theorem 2.14.25

Let $\mathcal{F}$ be a class of measurable functions $f : \mathcal{X} \to [0, 1]$ that satisfies one of the above inequalities. Then, for every $t > 0$,

$$P^*\left(\|\mathbb{G}_n\|_{\mathcal{F}} > t\right) \leq \left(\frac{Dt}{\sqrt{V}}\right)^V e^{-2t^2},$$

for a constant d that depends on K only.

Additional case

Case

For some constants $0 < W < 2$ and K,

$$\sup_Q \log N(\varepsilon, \mathcal{F}, L_2(Q)) \leq K \left(\frac{1}{\varepsilon}\right)^W, \quad \text{for every } \varepsilon > 0.$$

Additional case

Case

For some constants $0 < W < 2$ and K,

$$\sup_Q \log N\left(\varepsilon, \mathcal{F}, L_2(Q)\right) \le K \left(\frac{1}{\varepsilon}\right)^W, \quad \text{for every } \varepsilon > 0.$$

Theorem 2.14.26

Let $\mathcal{F}$ be a class of measurable functions $f : \mathcal{X} \to [0, 1]$ that satisfies the above inequality. Then, for every $\delta > 0$ and $t > 0$,

$$P^*\left(\|\mathbb{G}_n\|_{\mathcal{F}} > t\right) \le C e^{D t^{U+\delta}} e^{-2t^2},$$

where $U = W(6 - W)/(2 + W)$ and the constants C and D depend on K, W, and δ only.

Sharper bounds using sets

Cases

Suppose that $\mathcal{C}$ is a class of sets such that, for given consztntz K and V, either of the following holds

$$\sup_Q N(\varepsilon, \mathcal{C}, L_1(Q)) \leq \left(\frac{K}{\varepsilon}\right)^V, \quad \text{for every } 0 < \varepsilon < K,$$

$$N_{[\,]}(\varepsilon, \mathcal{C}, L_1(P)) \leq \left(\frac{K}{\varepsilon}\right)^V, \quad \text{for every } 0 < \varepsilon < K.$$

Sharper bounds using sets

Cases

Suppose that $\mathcal{C}$ is a class of sets such that, for given constzntz K and V, either of the following holds

$$\sup_Q N\left(\varepsilon, \mathcal{C}, L_1(Q)\right) \leq \left(\frac{K}{\varepsilon}\right)^V, \quad \text{for every } 0 < \varepsilon < K,$$

$$N_{[\,]}\left(\varepsilon, \mathcal{C}, L_1(P)\right) \leq \left(\frac{K}{\varepsilon}\right)^V, \quad \text{for every } 0 < \varepsilon < K.$$

Theorem 2.14.29

Let $\mathcal{C}$ be a class of sets that satisfies one of the inequalities above. Then

$$P^*\left(\|\mathbb{G}_n\|_{\mathcal{C}} > t\right) \leq \frac{D}{t}\left(\frac{DKt^2}{V}\right)^V e^{-2t^2},$$

for every $t > 0$ and a constant D that depends on K only.

$\sup_Q N(\varepsilon, \mathcal{C}, L_1(Q)) \leq \left(\frac{K}{\varepsilon}\right)^V$

$N_{[\,]}(\varepsilon, \mathcal{C}, L_1(P)) \leq \left(\frac{K}{\varepsilon}\right)^V$

Definition

$$\mathcal{C}_\delta := \{C \in \mathcal{C} : |P(C) - 1/2| \leq \delta\}$$

$\sup_Q N(\varepsilon, \mathcal{C}, L_1(Q)) \leq \left(\frac{K}{\varepsilon}\right)^V$

$N_{[\,]}(\varepsilon, \mathcal{C}, L_1(P)) \leq \left(\frac{K}{\varepsilon}\right)^V$

Definition

$$\mathcal{C}_\delta := \{C \in \mathcal{C} : |P(C) - 1/2| \leq \delta\}$$

Theorem 2.14.30

Let $\mathcal{C}$ be a class of sets that satisfies one of the above inequalities, and suppose moreover that

$$N(\varepsilon, \mathcal{C}_\delta, L_1(P)) \leq K' \delta^W \varepsilon^{-V'}, \quad \text{for every } \delta \geq \varepsilon > 0,$$

for some constant K'. Then

$$P^*\left(\|\mathbb{G}_n\|_{\mathcal{C}} > t\right) \leq D t^{2V' - 2W} e^{-2t^2},$$

for every $t > K\sqrt{W}$ and a constant D that depends on K, K', W, V, and V' only.