Empirical Processes

MAA110 - EPFL

Nikitas Georgakis*, Myrto Limnios*

29/04/2025

Notations. Without additional notice, we consider the same notations as in the lecture notes. We use the sup-norm for any function $h : \mathcal{X} \to \mathbb{R}$, defined by $\|h\|_\infty = \sup_{t\in\mathcal{X}} |h(t)|$.

Exercise 1. *Let $X_1, \ldots, X_n$ be an i.i.d. random sample of X valued in $\mathbb{R}$, of canonical empirical c.d.f. $F_n(t) = (1/n)\sum_{i=1}^n 1\{X_i \le t\}$, being an estimator of the true c.d.f. $F(t) = \mathbb{P}(X \le t)$, for all $t \in \mathbb{R}$. Prove that*

$$\mathbb{P}\left(\|F_n - F\|_\infty \ge \frac{C}{\sqrt{n}} + u\right) \le 2e^{-nu^2/8} \ , \ \forall u > 0 \ , \tag{1}$$

using Dudley's entropy integral.

Exercise 2 (Gaussian complexity for parametric classes of functions)**.** *Consider the parametric class of functions*

$$\mathcal{H} := \{h_\theta : x \in [0,1] \mapsto 1 - e^{-\theta x} \in \mathbb{R}, \quad \theta \in [0,1]\} \ ,$$

endowed with the uniform distance $\|f - g\|_\infty = \sup_{x\in[0,1]} |f(x) - g(x)|$. Recall that (exercise 1, week 5) $(\mathcal{H}, \|\cdot\|_\infty)$ is a metric space and that the covering number w.r.t. the uniform norm by $N(\varepsilon, \mathcal{H}, \|\cdot\|_\infty)$ is upperbounded for any $\varepsilon > 0$

$$N(\varepsilon) \le 2 + \frac{1}{2\varepsilon} \ .$$

1. *Prove that for all $\varepsilon > 0$*

   $$1 + \lfloor \frac{1 - 1/e}{2\varepsilon} \rfloor \le N(\varepsilon) \ .$$

   Hint: Prove a lowed bound on the packing number.

2. *Let an i.i.d. sequence of standard Gaussian r.v.s $\eta_1, \ldots, \eta_n$, independent of the X's, and define the Gaussian complexity by*

   $$\mathcal{G}_n = \mathbb{E}\left[\sup_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^n \eta_i X_i\right] \ .$$

   Prove that there exists a universal constant $C > 0$, such that

   $$\mathcal{G}_n \le C\sqrt{\frac{\log(n)}{n}} \ , \tag{2}$$

3. *Use Dudley's entropy integral to prove that the rate $O_{\mathbb{P}}(1/\sqrt{n})$ can be obtained. Precisely, show that there exists a universal constant $C' > 0$, such that*

   $$\mathcal{G}_n \le \frac{C'}{\sqrt{n}} \int_0^2 \sqrt{\log(1 + 1/u)}\mathrm{d}u \ . \tag{3}$$

***{first.last}@epfl.ch**, Office MA1493, CM1618

Exercise 3 (Lower bound for Gaussian maxima)**.** *Let* $\{X_1, \dots, X_M\}$ *be a set of centered Gaussian i.i.d. r.v.s of variance* $\sigma^2 > 0$.

1. *Prove that*

$$\mathbb{E}\left[\max_{i \le M} X_i\right] \ge u(1 - (\mathbb{P}(X_1 \le u))^n) - \mathbb{E}|X_1| \ ,$$

   for all $u > 0$. *Hint: Write* $\max_{i \le M} X_i = \max\{\max_{i \le M} X_i, 0\} + \min\{\max_{i \le M} X_i, 0\}$

2. *Conclude that there exists a small constant* $c > 0$ *such that*

$$\mathbb{E}\left[\max_{i \le M} X_i\right] \ge c\sigma\sqrt{\log M} \ .$$