Empirical Processes

MAA110 - EPFL

Nikitas Georgakis*, Myrto Limnios*

25/02/2025

Exercise 1. Let ψ be a real-valued convex, and continuously differentiable functions defined on $[0, b)$, $b \in (0, \infty]$, such that $\psi(0) = \psi'(0) = 0$.

1. Prove that the function ψ^* defined below is nonnegative, convex, and nondecreasing,

$$\psi^* : u \geq 0 \mapsto \inf_{t \in (0,b)} (tu - \psi(t)) \ .$$

2. Prove that the generalized inverse of ψ^* is defined by

$$\psi^{*-1}(u) = \inf_{t \in (0,b)} \frac{u + \psi(t)}{t} \ .$$

Exercise 2. 1. Prove Bennett's inequality in Theorem 2.1.7.
Hint: Define $g(u) = e^u - u - 1$. Prove that $u \mapsto u^2 g(u)$ is a nondecreasing function.

2. Prove Bernstein's inequality under the same assumptions.

Exercise 3. Let X be a nonnegative r.r.v., such that for all $u > 0$, there exists $C \geq 2$, $c > 0$ satisfying

$$\mathbb{P}(X \geq u) \leq Ce^{-u^2/c^2} \ .$$

Prove that

$$\mathbb{E}X \leq cK\sqrt{\log C} \ ,$$

where $K > 0$ is a universal constant.

Exercise 4 (Binary classification)**.** Consider the binary classification problem from Week 1. Consider an input r.v. valued in a measurable space $\mathcal{X} \subset \mathbb{R}^d$, with $d \in \mathbb{N}^*$, and an output r.v. Y valued in $\{0, 1\}$. Denote by P the joint distribution of (X, Y). Consider a class of a collection of classifiers $\mathcal{H} = \{h : \mathcal{X} \longrightarrow \{0, 1\}, \quad h \text{ measurable}\}$. The associate binary loss is defined as

$$\ell_h : (x, y) \in \mathcal{X} \times \{0, 1\} \mapsto 1\{h(x) \neq y\} \ ,$$

and the associated risk

$$\mathcal{R}(h) = \mathbb{E}[\ell_h(X, Y)] = \mathbb{P}(h(X) \neq Y) \ .$$

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

We suppose that the class $\mathcal{H}$ is finite, and totally separates the labels in $\{0, 1\}$, i.e., $\min_{h\in\mathcal{H}} \mathcal{R}(h) = 0$.

Consider an i.i.d. sample $\{(X_i, Y_i)\}_{i\leq n}$ drawn from P, and denote by $\widehat{h}$ the empirical minizer of the empirical risk function defined as follows:

$$\widehat{h} \in \arg\min_{h\in\mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} 1\{h(X_i) \neq Y_i\} =: \arg\min_{h\in\mathcal{H}} \mathcal{R}_n(h) \ .$$

1. Prove that $\min_{h\in\mathcal{H}} \mathcal{R}_n(h) = 0$ a.s.
2. Prove that $\mathbb{P}(R(\widehat{h}) > \varepsilon) \leq |\mathcal{H}|e^{-n\varepsilon}$, for all $\varepsilon > 0$.
3. Conclude that
$$\mathbb{E}[R(\widehat{h})] \leq \frac{1 + \log |\mathcal{H}|}{n} \ .$$