Empirical Processes

MAA110 - EPFL

Nikitas Georgakis*, Myrto Limnios*

18/02/2025

Binary classification. 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), and define the regression function (posterior distribution) by

$$\eta : x \in \mathcal{X} \mapsto P(Y = 1 | X = x) \; .$$

Exercise 1. Derive the explicit formula for η as function of θ and p, when we consider first that $\mathcal{X} \subset [0, 1]$, and P such that:

1. the conditional distribution of X given $Y = 0$ is $P_0 = \mathcal{U}([0, \theta])$, with $\theta \in (0, 1)$
2. the conditional distribution of X given $Y = 1$ is $P_1 = \mathcal{U}([0, 1])$
3. $p = \mathbb{P}(Y = 1) \in (0, 1)$.

Exercise 2 (Bayes classifier)**.** Suppose now that $\mathcal{X} \subset \mathbb{R}_+$, and denote by P_X the marginal of X, as well as the regression function given by:

$$\eta : x \in \mathcal{X} \mapsto \frac{x}{x + \theta} \; ,$$

for all $\theta > 0$. Consider a collection of classifiers $\mathcal{H} = \{h : \mathcal{X} \longrightarrow \{0, 1\}, \quad h \text{ measurable}\}$. The associate binary loss is defined as

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

i.e., equals to 1 if the predictor h mislabels the input x. The goal is to find the optimal classifier known as the *Bayes classifier* h^*, that minimizes the associated risk

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

1. Recall the minimization problem for which η is the unique solution.
2. Derive the explicit formula for the risk as function of η.
3. Prove that the Bayes risk equals to

   $$\mathcal{R}(h^*) = \int_{\mathcal{X}} \min(\eta(x), 1 - \eta(x)) \mathrm{d}P_X(x)$$

   Deduce the explicit formula for h^* as function of η.

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

4. Derive the Bayes risk for $P_X = \mathcal{U}([0, \alpha\theta])$, with $\alpha > 1$

Exercise 3. Consider the r.v. $X = (T, U, V)$, of independent coordinates distributed from a standard exponential distribution. Let $\theta > 0$ be fixed and define the response variable by $Y = 1\{T + U + V \leq \theta\}$.

1. Derive the Bayes function when V is not observed, i.e., $h^*(T, U)$, and the associated Bayes risk.
2. Consider now that T is not observed. Continue the above computations and compare the respective risk functions for $\theta = 9$.
3. Propose a classifier when none of the coordinates of X are observed. Derive its risk.

Uniform convergence. This exercise shows that if we are able to approximate (discretize) the function class $\mathcal{H}$ into an (arbitrarily) finite class $\mathcal{H}_\varepsilon$, depending on ε, then we can prove a ULLN.

Exercise 4 (Glivenko-Cantelli's Theorem with bracketing)**.** Let $(\Omega, \mathcal{A}, P)$ be a probability space. Consider a class $\mathcal{H}$ of measurable functions $h : \mathcal{X} \to \mathbb{R}$, such that $P|h| < \infty$, and X a r.v. of probability distribution P. Let $X_1, \ldots, X_n$ an i.i.d. sample drawn from P, with $n \in \mathbb{N}^*$. Define the standard empirical process by

$$(h, \omega) \in \mathcal{H} \times \Omega \mapsto \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (h(X_i) - \mathbb{E}h(X)) = \sqrt{n}(P_n(\omega) - P)(h) \in \mathbb{R} .$$

Let $\varepsilon > 0$ be fixed. Suppose that there exists a class of functions $\mathcal{H}_\varepsilon$ such that we can bound any function $h \in \mathcal{H}$ by elements of $\mathcal{H}_\varepsilon$. Specifically, for all $h \in \mathcal{H}$ we can find $h_l, h_u \in \mathcal{H}_\varepsilon$, such that $h_l \leq h \leq h_u$, and $P(h_l - h_u) \leq \varepsilon$.

Prove that

$$\|P_n - P\|_{\mathcal{H}} := \sup_{h \in \mathcal{H}} |P_n h - Ph| \xrightarrow[n \to \infty]{a.s.} 0 .$$

Exercise 5. Prove that any continuous map between two metric spaces is Borel-measurable. Hint: Use the definition of measurable maps.