Lecture 3: Empirical Processes

Tianpei Xie

Feb. 1st., 2023

Contents

1 Uniform Law of Large Numbers **2**
- 1.1 Functional of Cumulative Distribution Function . . . 2
- 1.2 Glivenko-Cantelli Theorem . . . 2
- 1.3 Glivenko-Cantelli Class . . . 2

2 Empirical Processes **2**
- 2.1 Definitions . . . 2
- 2.2 Tail bounds for Empirical Processes . . . 5
- 2.3 Maximal Inequalities . . . 5
- 2.4 Symmetrization . . . 5
- 2.5 Uniform Convergence via Rademacher Complexity . . . 5

3 Expected Value of Suprema of Empirical Process **5**
- 3.1 Metric Entropy . . . 5
- 3.2 Chaining and Dudley's Entropy Integral . . . 5
- 3.3 Contraction Inequality . . . 5
- 3.4 Vapnik-Chervonenkis Class . . . 5
- 3.5 Comparison Theorems . . . 5

1 Uniform Law of Large Numbers

1.1 Functional of Cumulative Distribution Function

1.2 Glivenko-Cantelli Theorem

1.3 Glivenko-Cantelli Class

2 Empirical Processes

2.1 Definitions

- **Definition** (***Empirical Measure***) [Wellner et al., 2013, Giné and Nickl, 2021]
  Let $(\mathcal{X}, \mathscr{F}, \mathcal{P})$ be a *probability space*, and let $X_i, i \in \mathbb{N}$, be the *coordinate functions* of *the* ***infinite product probability space*** $(\Omega, \mathscr{B}, \mathbb{P}) := (\mathcal{X}^{\infty}, \mathscr{F}^{\infty}, \mathcal{P}^{\infty})$, $X_i : \mathcal{X}^{\infty} \to \mathcal{X}$, which are ***independent*** *identically distributed* $\mathcal{X}$-valued random variables with law $\mathcal{P}$.

  The empirical measure corresponding to the '*observations*' $X_1, \ldots, X_n$, for any $n \in \mathbb{N}$, is defined as ***the random discrete probability measure***

  $$\mathcal{P}_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \tag{1}$$

  where δ_x is *Dirac measure* at x, that is, unit mass at the point x. In other words, for each event A, $\mathcal{P}_n(A)$ is *the* ***proportion*** *of* ***observations*** X_i*,* $i = 1, \ldots, n$*, that fall in* A; that is,

  $$\mathcal{P}(A) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}(A) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \in A\}, \quad A \in \mathscr{F}.$$

- **Remark** (***Probability Measure with Operator Notation***) [Wellner et al., 2013, Giné and Nickl, 2021]
  For any measure μ and μ-integrable function f, we will use the following ***operator notation*** for the integral of f with respect to μ:

  $$\mu f \equiv \mu(f) = \int_{\Omega} f d\mu.$$

  This is valid since there exists an isomorphism between *the space of probability measure* and *the space of bounded linear functional* on $\mathcal{C}_0(\Omega)$ by Riesz-Markov representation theorem (assuming Ω is *locally compact*). By this notion the expectation $\mathcal{P}f = \mathbb{E}_{\mathcal{P}}[f]$.

- **Definition** (***Empirical Process***) [Wellner et al., 2013, Giné and Nickl, 2021]
  Let $\mathcal{F}$ be a *collection of* $\mathcal{P}$*-integrable functions* $f : \mathcal{X} \to \mathbb{R}$, usually infinite. For any such class of functions $\mathcal{F}$, ***the empirical measure*** defines a ***stochastic process***

  $$f \to \mathcal{P}_n f, \quad f \in \mathcal{F} \tag{2}$$

  which we may call ***the empirical process indexed by*** $\mathcal{F}$, although we prefer to reserve the notation '*empirical process*' for *the* ***centred*** *and* ***normalised*** *process*

  $$f \to \nu_n(f) := \sqrt{n}\left(\mathcal{P}_n f - \mathcal{P} f\right), \quad f \in \mathcal{F}. \tag{3}$$

- **Remark** An explicit notion of *(centered and normalized) empirical process* is

$$\sqrt{n}\,(\mathcal{P}_n f - \mathcal{P} f) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(f(X_i) - \mathbb{E}_{\mathcal{P}}\left[f(X) \right] \right), \quad f \in \mathcal{F}.$$

  where $X_1, \ldots, X_n \sim \mathcal{P}$ are i.i.d random variables. Note that it is a stochastic process since *the function f is changing* in $\mathcal{F}$, i.e. the process $(\mathcal{P}_n - \mathcal{P})\, f$ is indexed by function $f \in \mathcal{F}$ not finite dimensional variable.

- **Remark** (***Random Measure***)
  Normally we assume that data are sampled from some distribution $\mathcal{P}$ and the data itself is random. However, the empirical measure

$$\mathcal{P}_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

  itself is considered as a ***random*** probability measure. That is, *the sampling mechanism itself contains randomness* and it is not sampling from one distribution but ***a system of distributions depending on the choice of dataset*** $X_1, \ldots, X_n$, which in turn were sampled from some *prior* $\mathcal{P}$. Due to this randomness, $\mathcal{P}_n f = \mathbb{E}_{\mathcal{P}_n}[f]$ is not a fixed expectation number but a random variable. In fact, this is the empirical mean (i.e. sample mean)

$$\mathcal{P}_n f = \mathbb{E}_{\mathcal{P}_n}[f] = \frac{1}{n} \sum_{i=1}^{n} f(X_i).$$

  The critical difference between mean of empirical measure vs. sample mean is that we now assume that f is ***not fixed***.

- **Remark** (***Object of Empirical Process Theory***)
  The ***object*** *of empirical process theory* is to study *the* ***properties*** *of the* ***approximation*** of $\mathcal{P}f$ by $\mathcal{P}_n f$, ***uniformly in*** $\mathcal{F}$, concretely, to obtain both ***probability estimates*** *for the random quantities*

$$\|\mathcal{P}_n - \mathcal{P}\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |\mathcal{P}_n f - \mathcal{P} f|$$

  and ***probabilistic limit theorems*** for the processes $\{(\mathcal{P}_n - \mathcal{P})(f) : f \in \mathcal{F}\}$.

  Note that the quantity $\|\mathcal{P}_n - \mathcal{P}\|_{\mathcal{F}}$ is a ***random variable*** since $\mathcal{P}_n$ is a ***random measure***.

- **Remark** (***Measurability Problem***)
  There may be a ***measurability problem*** for

$$\|\mathcal{P}_n - \mathcal{P}\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |\mathcal{P}_n f - \mathcal{P} f|$$

  since *the* ***uncountable*** *suprema* of measurable functions *may not be measurable.*

  However, there are many situations where this is actually a ***countable supremum***. For instance, for probability distribution on $\mathbb{R}$

$$\|\mathcal{P}_n - \mathcal{P}\|_{\infty} := \sup_{t \in \mathbb{R}} |(\mathcal{P}_n - \mathcal{P})\,(-\infty, t)| = \sup_{t \in \mathbb{Q}} |F_n(t) - F(t)| = \sup_{t \in \mathbb{Q}} |(\mathcal{P}_n - \mathcal{P})\,(-\infty, t)|$$

where $F(t) = \mathcal{P}(-\infty, t)$ is the cumulative distribution function. If $\mathcal{F}$ is *countable* or if there exists $\mathcal{F}_0$ *countable* such that

$$\|\mathcal{P}_n - \mathcal{P}\|_{\mathcal{F}} = \|\mathcal{P}_n - \mathcal{P}\|_{\mathcal{F}_0}, \quad \text{a.s.}$$

then the measurability problem disappears.

For the next few sections we will simply assume that the class $\mathcal{F}$ is *countable.*

- **Remark** (***Bounded Assumption***)
  If we assume that

$$\sup_{f \in \mathcal{F}} |f(x) - \mathcal{P}f| < \infty, \quad \forall x \in \mathcal{X}, \tag{4}$$

  then the maps from $\mathcal{F}$ to $\mathbb{R}$,

$$f \to f(x) - \mathcal{P}f, \quad x \in \mathcal{X},$$

  are ***bounded functionals*** over $\mathcal{F}$, and therefore, so is $f \to (\mathcal{P}_n - \mathcal{P})(f)$. That is,

$$\mathcal{P}_n - \mathcal{P} \in \ell_\infty(\mathcal{F}),$$

  where $\ell_\infty(\mathcal{F})$ is ***the space of bounded real functionals*** *on* $\mathcal{F}$, a *Banach space* if we equip it with the supremum norm $\|\cdot\|_{\mathcal{F}}$.

  A large literature is available on *probability in* ***separable Banach spaces***, but unfortunately, $\ell_\infty(\mathcal{F})$ is ***only separable*** when the class $\mathcal{F}$ is ***finite***, and ***measurability problems*** arise because *the probability law* of the process $\{(\mathcal{P}_n - \mathcal{P})(f) : f \in \mathcal{F}\}$ ***does not extend*** *to the Borel* σ*-algebra of* $\ell_\infty(\mathcal{F})$ even in simple situations.

- **Remark** This chapter addresses ***three main questions*** about the empirical process:

  1. The first question has to do with ***concentration*** of $\|\mathcal{P}_n - \mathcal{P}\|_{\mathcal{F}}$ about *its* ***mean*** when $\mathcal{F}$ is ***uniformly bounded***. Recall that $\|\mathcal{P}_n - \mathcal{P}\|_{\mathcal{F}}$ is a random variable itself, due to randomness of the empirical measure. We mainly use the *non-asymptotic analysis* to obtain *the exponential bound for concentration.*

  2. The second question is do ***good estimates*** for ***mean*** $\mathbb{E}\left[\|\mathcal{P}_n - \mathcal{P}\|_{\mathcal{F}}\right]$ exist? We will examine two main techniques that give answers to this question, both related to ***metric entropy*** and ***chaining***. One of them, called ***bracketing***, uses *chaining* in combination with *truncation* and *Bernstein's inequality.* The other one applies to ***Vapnik-Cervonenkis (VC) classes of functions***.

  3. Finally, the last question about the empirical process refers to ***limit theorems***, mainly ***the uniform law of large numbers*** and the ***central limit theorem***, in fact, the analogues of ***the classical Glivenko-Cantelli*** and *Donsker theorems* for the empirical distribution function.

     Formulation of *the central limit theorem* will require some more *measurability* because we will be considering ***convergence in law*** of random elements in ***not necessarily separable Banach spaces***.

2.2 Tail bounds for Empirical Processes

2.3 Maximal Inequalities

2.4 Symmetrization

2.5 Uniform Convergence via Rademacher Complexity

3 Expected Value of Suprema of Empirical Process

3.1 Metric Entropy

3.2 Chaining and Dudley's Entropy Integral

3.3 Contraction Inequality

3.4 Vapnik-Chervonenkis Class

3.5 Comparison Theorems

References

Evarist Giné and Richard Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge university press, 2021.

Jon Wellner et al. *Weak convergence and empirical processes: with applications to statistics*. Springer Science & Business Media, 2013.