Empirical processes and chaining

Young researchers' days - June 2026

Maël Duverger

Introduction

- Let $\mathcal{F}$ be a set of functions from $\mathcal{X}$ to $\mathbb{R}$.
- To simplify the presentation, assume that $\sup_{f \in \mathcal{F}} \|f\|_\infty < +\infty$.
- Let $X_1, ..., X_n$ be i.i.d. random variables on $\mathcal{X}$ with probability distribution $\mathbb{P}$.

Let

- $\mathbb{P}f = \mathbb{E}_{\mathbb{P}}[f(X)] = \int f(x) d\mathbb{P}(x)$
- $\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^n f(X_i) = \int f(x) d\mathbb{P}_n(x)$ $\quad \rightarrow$ ($\mathbb{P}_n$=empirical measure)
- $G_n(f) = \sqrt{n}(\mathbb{P}_n f - \mathbb{P}f) = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^n f(X_i) - \mathbb{E}_{\mathbb{P}}[f(X_i)] \right)$
  $\hookrightarrow$ G_n=empirical process

For any $k \in \mathbb{N}$ and $f_1, ..., f_K \in \mathcal{F}$, by the CLT we have

$$\left(G_n(f_1), ..., G_n(f_K)\right) \xrightarrow[n \to +\infty]{(d)} \mathcal{N}_K(0, V) \qquad \text{(marginal convergence)}$$

where V is the $K \times K$ matrix whose entry (i, j) is given by $V_{i,j} = Cov(f_i(X_1), f_j(X_1)) = \mathbb{P}f_i f_j - \mathbb{P}f_i \mathbb{P}f_j$.

Goal

- The empirical process G_n is a process indexed by $\mathcal{F}$: $G_n = (G_n(f), f \in \mathcal{F})$.
- $\mathcal{L}_\infty(\mathcal{F}) = \{h : \mathcal{F} \to \mathbb{R},\ h \text{ bounded}\}$ equipped with the norm $\|h\|_\mathcal{F} = \sup_{f\in\mathcal{F}} |h(f)|$.

$\hookrightarrow$ We view G_n as a random variable with values in $\mathcal{L}_\infty(\mathcal{F})$.

- $\mathcal{L}_\infty(\mathcal{F})$ can be non separable, they are thus measurability issues, we will not speak about it.
- We want to prove that

$$G_n \overset{\mathcal{L}_\infty(\mathcal{F})}{\rightsquigarrow} G\ , \quad n \to +\infty$$

which means that for all bounded continuous function $g : \mathcal{L}_\infty(\mathcal{F}) \to \mathbb{R}$,

$$\mathbb{E}\big[g(G_n)\big] \xrightarrow{n\to+\infty} \mathbb{E}\big[g(G)\big]$$

Gaussian limit

Recall the marginal convergence :

$$\big(G_n(f_1), ..., G_n(f_K)\big) \xrightarrow{(d)} \mathcal{N}_K(0, V).$$

Thus, if

$$G_n \overset{\mathcal{L}_\infty(\mathcal{F})}{\rightsquigarrow} G$$

then the limit G is necessarily (by Kolmogorov extension theorem) a Gaussian random variable in $\mathcal{L}_\infty(\mathcal{F})$.
$\hookrightarrow$ So $G_n \overset{\mathcal{L}_\infty(\mathcal{F})}{\rightsquigarrow} G$ is a kind of CLT in a Banach space.

Simple application : empirical cumulative distribution function

- Recall that the cumulative distribution function $F : \mathbb{R} \to [0; 1]$ is given by

$$F(t) = \mathbb{P}(]-\infty, t]) = \mathbb{P}f_t$$

with $f_t(x) = \mathbb{1}_{]-\infty;t]}(x)$. Note that f_t is in $\mathcal{F}$=Skorohod space (equipped with sup-norm).

- A "natural" estimator is the empirical cumulative distribution function $F_n : \mathbb{R} \to [0; 1]$ defined by

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{]-\infty,t]}(X_i) = \mathbb{P}_n f_t$$

- In this case, $G_n \overset{\mathcal{L}_\infty(\mathcal{F})}{\rightsquigarrow} G$ means that $\sqrt{n}(F_n - F) \xrightarrow{\mathcal{L}_\infty(\mathcal{F})} B$ where B is a Brownian bridge. $\hookrightarrow$ uniform convergence of F_n to F at rate $1/\sqrt{n}$ and limiting shape.

- Numerous applications in statistics : M-estimation, non parametric problems...

Prokhorov argument

• We say that G_n is asymptotically tight, if (roughly) for every $\epsilon \in]0,1[$, there exists a compact K of $\mathcal{L}_\infty(\mathcal{F})$ such that

$$\liminf_{n \to +\infty} \mathbb{P}\big(G_n \in K\big) \geq 1 - \epsilon$$

Prokhorov theorem : asymptotic tightness + "marginal convergence" implies that

$$G_n \overset{\mathcal{L}_\infty(\mathcal{F})}{\rightsquigarrow} G$$

$\hookrightarrow$ How large $\mathcal{F}$ can be ? How to measure its size ?

$\rightarrow$ a class of functions $\mathcal{F}$ such that $G_n \overset{\mathcal{L}_\infty(\mathcal{F})}{\rightsquigarrow} G$ is called a $\mathbb{P}$-Donsker class.

Remark on tightness

- On a separable space, a probability measure is always tight, but here $\mathcal{L}_\infty(\mathcal{F})$ is not separable as soon as $\mathcal{F}$ is not finite.

- For applications, we need the limit G to be tight which is a priori not always true.

- However, if G_n is asymptotically tight then $G_n \to G$ and one can show that G is also tight $\to$ we are fine

Sufficient condition for tightness

Let's fix a metric ρ on $\mathcal{F}$.

Definition

We say that G_n is *asymptotically ρ-equicontinuous* if for every $\epsilon > 0$ and $\eta > 0$, there exists $\delta > 0$ such that

$$\limsup_{n \to +\infty} \mathbb{P}\left(\sup_{\rho(f,f') \leq \delta} |G_n(f) - G_n(f')| > \epsilon \right) \leq \eta$$

Property

If for some metric ρ, $(\mathcal{F}, \rho)$ is totally bounded and G_n is asymptotically ρ-equicontinuous, then G_n is asymptotically tight.

$\hookrightarrow$ think about Arzela-Ascoli theorem

Verifying the asymptotic equicontinuity condition

- To verify the asymptotic equicontinuity condition, we have to control the supremum of a collection of random variables : **how to do this efficiently (=without a stringent condition on $\mathcal{F}$) ?**

- In other words (little change of paradigm), given a collection of real-valued centered random va ($Y_f, f \in \mathcal{F}_0$) and $u > 0$, how can we obtain a sharp upper bound on

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}_0} |Y_f - Y_{f_0}| > u \right)$$

for some arbitrary $f_0 \in \mathcal{F}_0$.

Chaining

• Chaining technique : pioneered by Kolmogorov, then developed by Talagrand who proved its optimality, see the book "Talagrand - The Generic Chaining - Springer".

• Assume that the process $(Y_f)_{f\in\mathcal{F}_0}$ is separable :

$$\mathbb{P}\Big(\sup_{f\in\mathcal{F}_0}|Y_f - Y_{f_0}| > u\Big) = \sup\Big\{\mathbb{P}\Big(\sup_{f\in\mathcal{T}}|Y_f - Y_{f_0}| > u\Big), \mathcal{T}\subset\mathcal{F}_0 \quad \textit{finite}\Big\}$$

so that without loss of generality, we can consider that $\mathcal{F}_0$ is finite.

• We also assume that for some metric d

$$\mathbb{P}\Big(|Y_f - Y_{f'}| > u\Big) \leq \exp\Big(-\frac{u^2}{2d^2(f,f')}\Big)$$

$\hookrightarrow$ make a comment here.

Chaining : main idea

- We first have

$$\mathbb{P}\Big(\sup_{f\in\mathcal{F}_0}|Y_f - Y_{f_0}| > u\Big) \leq \sum_{f\in\mathcal{F}_0}\mathbb{P}\Big(|Y_f - Y_{f_0}| > u\Big) \qquad \text{(union bound)}$$

$\hookrightarrow$ it is not bad if the variables $(Y_f - Y_{f_0})_f$ are independent but it is a disaster if the variables $(Y_f - Y_{f_0})_f$ are very correlated (think about the identical case).

$\hookrightarrow$ Let $\mathcal{F}_0^1$ be a subset of $\mathcal{F}_0$ and for $f \in \mathcal{F}_0$, denote by $Y_{\pi_1(f)}$ the closest element of $\mathcal{F}_0^1$ to f. We have

$$Y_f - Y_{f_0} = \underbrace{Y_f - Y_{\pi_1(f)}}_{\textit{lower magnitude}} + \underbrace{Y_{\pi_1(f)} - Y_{f_0}}_{\textit{few variables+rather different}} \qquad (\textit{make a draw})$$

Iterated process

We consider now a sequence $(\mathcal{F}_0^n)_n$ of subsets of $\mathcal{F}_0$ such that $Card(\mathcal{F}_0^n) \leq 2^{2^n}$ and we have

$$Y_f - Y_{f_0} = \sum_{n \geq 1} Y_{\pi_n(f)} - Y_{\pi_{n-1}(f)}$$

with by convention $\pi_0(f) = f_0$. Now, on the event

$$\Omega_u = \left\{ \forall n \geq 1, \forall f, |Y_{\pi_n(f)} - Y_{\pi_{n-1}(f)}| \leq u 2^{n/2} d(\pi_n(f), \pi_{n-1}(f)) \right\}$$

we have that

$$\sup_{f \in \mathcal{F}_0} |Y_f - Y_{f_0}| \leq u \sum_{n \geq 1} 2^{n/2} d(\pi_n(f), \pi_{n-1}(f)) \leq 2u \sum_{n \geq 0} 2^{n/2} d(f, \mathcal{F}_0^n)$$

Also, by a union bound and the sub-gaussian assumption, for $u \geq 2$:

$$\mathbb{P}(\Omega_u^c) \leq \sum_{n \geq 1} Card(\mathcal{F}_0^n \times \mathcal{F}_0^{n-1}) \exp\left(-u^2 2^n\right) \leq \sum_{n \geq 1} \left(2^{2^n}\right)^2 \exp\left(-u^2 2^n\right) \lesssim \exp(-u^2/2)$$

We finally have

$$\mathbb{P}\Big(\sup_{f\in\mathcal{F}_0} |Y_f - Y_{f_0}| > u \sup_{f\in\mathcal{F}_0} \sum_{n\geq 0} 2^{n/2} d(f,\mathcal{F}_0^n)\Big) \lesssim \exp\big(-u^2/2\big)$$

$\hookrightarrow \sup_{f\in\mathcal{F}_0} \sum_{n\geq 0} 2^{n/2} d(f,\mathcal{F}_0^n)$ is optimal !

$\hookrightarrow$ now, we upper bound $\sup_{f\in\mathcal{F}_0} \sum_{n\geq 0} 2^{n/2} d(f,\mathcal{F}_0^n)$ to obtain a more explicit final upper bound :

$$\sup_{f\in\mathcal{F}_0} \sum_{n\geq 0} 2^{n/2} d(f,\mathcal{F}_0^n) \leq \sum_{n\geq 0} 2^{n/2} \sup_{f\in\mathcal{F}_0} d(f,\mathcal{F}_0^n)$$

Covering number

Definition

For some $\epsilon > 0$ and the metric d, the *covering number* of $\mathcal{F}_0$, denoted $\mathcal{N}(d, \epsilon, \mathcal{F}_0)$ is the minimal number of d-balls of radius ϵ needed to cover $\mathcal{F}_0$.

$\hookrightarrow$ examples of covering number for some functional sets if you want !

• Choose $\epsilon = \epsilon_n$ such that $\mathcal{N}(d, \epsilon_n, \mathcal{F}_0) \leq 2^{2^n}$, let $\mathcal{F}_0^n$ be the centering points of this covering (*make a draw*), then

$$\sum_{n \geq 0} 2^{n/2} \sup_{f \in \mathcal{F}_0} d(f, \mathcal{F}_0^n) \lesssim \sum_{n \geq 0} 2^{n/2} \epsilon_n \leq \ldots \lesssim \int_0^{+\infty} \sqrt{\log(\mathcal{N}(d, \epsilon, \mathcal{F}_0))} d\epsilon$$

$\hookrightarrow$ this last upper bound is called Dudley entropy bound : it is not optimal but it is easier to compute.

Application to empirical process

We come back to our initial problem on empirical process indexed by $\mathcal{F}$:

Theorem

If some mild and technical conditions are satisfied and more importantly if

$$\int_0^{+\infty} \sqrt{\log(\mathcal{N}(L_2(\mathbb{P}), \epsilon, \mathcal{F}))} d\epsilon < +\infty,$$

then $\mathcal{F}$ is a Donsker class :

$$G_n \xrightarrow[n \to +\infty]{\mathcal{L}_\infty(\mathcal{F})} G$$

$\hookrightarrow$this is a slightly abusive entropy condition

Reference

References :

- Weak Convergence and Empirical processes, van der Vaart and Wellner, Springer
- The Generic Chaining, Talagrand, Springer