Empirical Processes with Applications in Statistics

1.Introduction

Johan Segers

UCLouvain (Belgium)

Conférence Universitaire de Suisse Occidentale

Programme Doctoral en Statistique et Probabilités Appliquées

Les Diablerets, February 4–5, 2020

A classical example

- $X_1, X_2, \ldots$ i.i.d. random variables on $\mathbb{R}$
- Cumulative distribution functions: for $x \in \mathbb{R}$,

$$F(x) = \mathrm{P}(X_i \leqslant x)$$

$$\mathbb{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \leqslant x\}$$

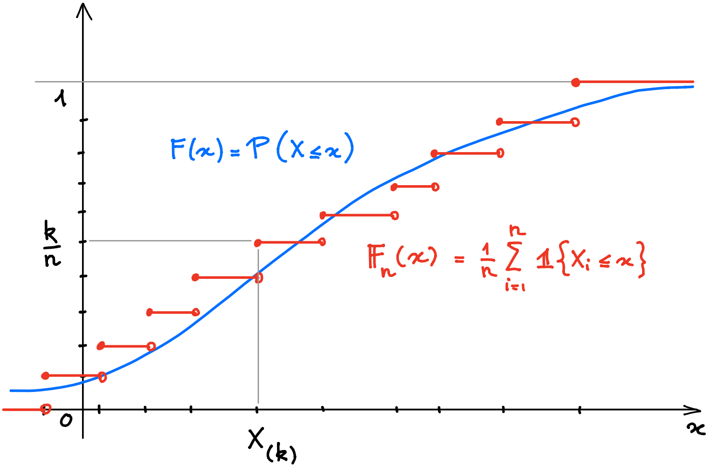
- **Law of large numbers**: for every $x \in \mathbb{R}$,

$$\mathbb{F}_n(x) \to F(x) \quad \text{a.s.}, \quad n \to \infty$$

- **Multivariate central limit theorem**: for every $(x_1, \ldots, x_\ell) \in \mathbb{R}^\ell$,

$$\left(\sqrt{n} \left\{ \mathbb{F}_n(x_j) - F(x_j) \right\} \right)_{j=1}^{\ell} \xrightarrow{d} \mathcal{N}_\ell(0, \Sigma), \qquad n \to \infty$$

with covariances $\sigma_{jk} = F(x_j \wedge x_k) - F(x_j)F(x_k)$



$F(x) = \mathbb{P}(X \leq x)$
$\mathbb{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \leq x\}$
1
$\frac{k}{n}$
0
$X_{(k)}$
x

Goodness-of-fit testing

Test of a simple hypothesis on the distribution: for a given F_0,

$$H_0 : F = F_0 \qquad \text{vs} \qquad H_1 : F \neq F_0$$

Test statistic:

$$T_n = \sup_{x \in \mathbb{R}} \sqrt{n} \left| \mathbb{F}_n(x) - F_0(x) \right|$$

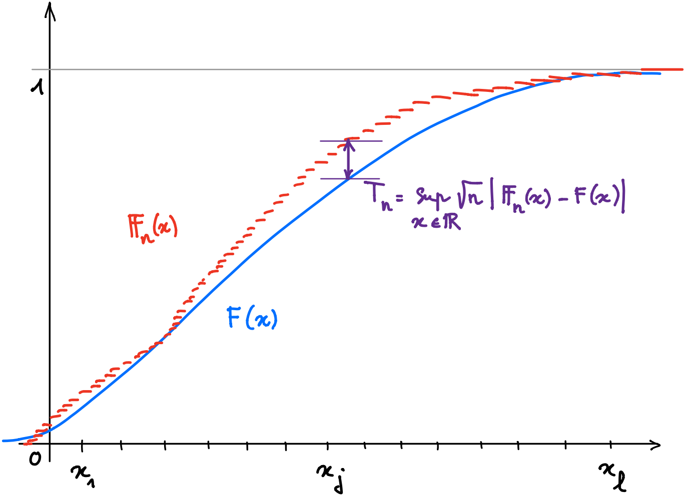
Distribution under H_0? Limit distribution as $n \to \infty$?

For a vector $(x_1, \ldots, x_\ell) \in \mathbb{R}$, we could consider

$$T_n(x_1, \ldots, x_\ell) := \max_{j=1\ldots,\ell} \sqrt{n} \left| \mathbb{F}_n(x_j) - F_0(x_j) \right| \xrightarrow{d} \max_{j=1\ldots,\ell} |B(x_j)|$$

where $(B(x_1), \ldots, B(x_\ell)) \sim \mathcal{N}(0, \Sigma)$ as on the previous slide.

How to go from "$\max_{x_1,\ldots,x_\ell}$" to "$\sup_{x \in \mathbb{R}}$?



1
0
$\mathbb{F}_n(x)$
$F(x)$
$T_n = \sup_{x \in \mathbb{R}} \sqrt{n} \, | \mathbb{F}_n(x) - F(x) |$
x_1
x_j
x_ℓ

Kolmogorov (1933), Doob (1949), Donsker (1952)
If F is continuous, the law of T_n under H_0 does not depend on F_0, and

$$T_n \xrightarrow{d} \sup_{t \in [0,1]} |B(t)|, \qquad n \to \infty$$

where $(B(t))_{t \in [0,1]}$ is a *Brownian bridge*

Brownian bridge:

- zero-mean Gaussian process
  (= collection of jointly normal random variables)
- (almost surely) continuous trajectories $t \mapsto B(t)$
- covariance function

$$\forall s, t \in [0, 1], \qquad \mathrm{E}[B(s)B(t)] = s \wedge t - st$$

Pointwise convergence: not sufficient for suprema

- Kolmogorov's result does *not* follow from the multivariate CLT: pointwise convergence does *not* imply convergence of suprema
- Sufficient (but not necessary) is *uniform convergence*

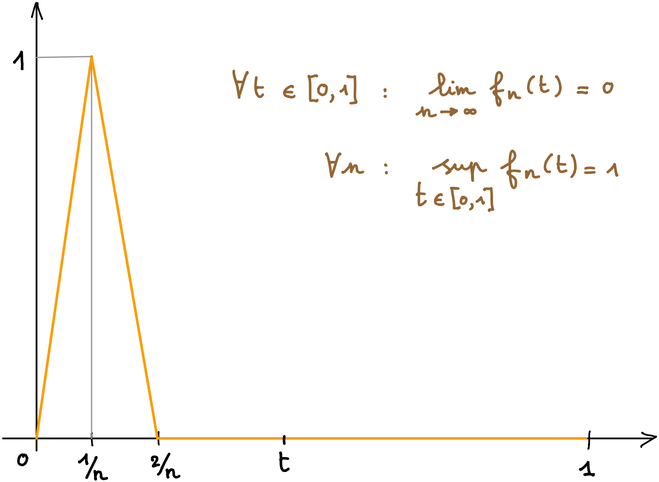
Example: construct functions $f_n : [0,1] \to [0,1]$ such that, at the same time

$$\lim_{n\to\infty} f_n(t) = 0 \qquad \forall t \in [0,1]$$

$$\sup_{t\in[0,1]} f_n(t) = 1 \qquad \forall n$$

even though, for all $(t_1, \ldots, t_\ell) \in [0,1]^\ell$, we must have

$$\max_{j=1,\ldots,\ell} f_n(t_j) \to 0, \qquad n \to \infty$$



$$\forall t \in [0,1] \ : \ \lim_{n \to \infty} f_n(t) = 0$$

$$\forall n \ : \ \sup_{t \in [0,1]} f_n(t) = 1$$

Uniform convergence of stochastic processes?

For simplicity, assume F is uniform on $[0,1]$

Kolmogorov's result would follow from a "uniform" version of

$$\left(\sqrt{n}\,\{\mathbb{F}_n(t) - F(t)\}\right)_{t\in[0,1]} \xrightarrow{d} (B(t))_{t\in[0,1]}, \qquad n \to \infty$$

Meaning? Certainly *not*[1]

$$\sup_{t\in[0,1]} \left|\sqrt{n}\,\{\mathbb{F}_n(t) - F(t)\} - B(t)\right| \xrightarrow{??} 0, \qquad n \to \infty$$

[1] Although such statements can be given a formal meaning via Skorohod constructions

Random elements in a function space

View $t \mapsto \sqrt{n}\{\mathbb{F}_n(t) - F(t)\}$ as a *random function*:
map from probability space Ω carrying the X_i into some *function space*

- Space should contain *almost all trajectories* of $\sqrt{n}(\mathbb{F}_n - F)$
- Function space to be equipped with a *metric* or norm, so we can consider the corresponding Borel σ-field
- Metric should be strong enough so that convergence of functions has useful consequences
  - E.g.: $z_n \to z$ should imply $\sup_t |z_n(t)| \to \sup_t |z(t)|$

In that function space, show weak convergence $\sqrt{n}(\mathbb{F}_n - F) \xrightarrow{d} B$

Which function space?

$\mathcal{C}([0,1])$ Continuous functions, supremum norm

$$\|z\|_\infty = \sup_{0 \leqslant t \leqslant 1} |z(t)|$$

- $t \mapsto \mathbb{F}_n(t)$ is not continuous, but we could make it so
- Enforcing continuity could get awkward on other domains

$\mathcal{D}([0,1])$ càdlàg functions with (some) Skorohod topology. Drawbacks:

- Addition of functions is not continuous in Skorohod metric
- Difficult to generalize to other domains than intervals

Billingsley (1968)

$L_p([0,1])$ Convergence in p-th mean does not even imply pointwise convergence, let alone convergence of suprema

Space of bounded functions

$\ell^\infty([0,1])$ Space of all bounded functions $z : [0,1] \to \mathbb{R}$ equipped with the supremum norm

$$\|z\|_\infty := \sup_{0 \leqslant t \leqslant 1} |z(t)|$$

- Strong notion of convergence, easy to apply
- No worries about regularity of trajectories
- Easy to extend from $[0,1]$ to general domains

But...

[Error on slide 12: non-measurable mapping]

Non-measurability of the empirical process

The map

$$\Psi : [0, 1] \to \ell^\infty([0, 1]) : x \mapsto \mathbb{1}_{[0,x]}$$

is *not* Borel measurable.

Proof.

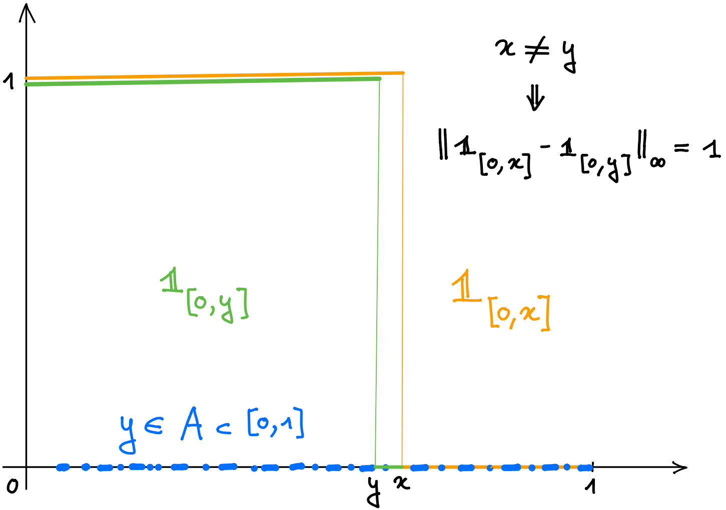
Let $A \subset [0, 1]$ be *not* Borel measurable. The set

$$G = \bigcup_{y \in A} \left\{ z \in \ell^\infty([0, 1]) \;\middle|\; \|z - \mathbb{1}_{[0,y]}\|_\infty \leqslant 1/2 \right\}$$

is a union of open balls in $\ell^\infty([0, 1])$, therefore open, therefore Borel measurable. But

$$\mathbb{1}_{[0,x]} \in G \iff x \in A$$

so $\Psi^{-1}(G) = A$, which is not Borel measurable. □


x ≠ y
⇓
‖𝟙[0,x] - 𝟙[0,y]‖∞ = 1
1
𝟙[0,y]
𝟙[0,x]
y ∈ A ⊂ [0,1]
0
y
x
1

Possible fix $\#1$: the ball σ-field

Lack of measurability is not to be taken lightly (ask S. Banach and A. Tarski)

Possible solution: on $\ell^\infty([0,1])$, consider a smaller σ-field, the one generated by open balls (Dudley, 1966; Pollard, 1984)

- The set G on slide 13 is *not* in the ball σ-field
  The union there is uncountable

Drawback: Natural link with topology on $\ell^\infty([0,1])$ is lost

- Continuous mappings on $\ell^\infty([0,1])$ are not automatically ball-measurable
- Continuous mapping theorem and functional delta method become potentially more difficult to apply

Possible fix #2: allow for non-measurable mappings

Hoffmann-Jørgensen (1984, 1991):

- Use *inner/outer expectation* and *inner/outer probability* for non-measurable mappings and sets
- Insist on measurability only in the limit

Theory works especially well when limit Borel probability measure L on $\ell^\infty([0,1])$ is *tight*:

- For all $\varepsilon > 0$, there exists compact $K \subset \ell^\infty([0,1])$ such that $L(K) \geqslant 1 - \varepsilon$
- Then L concentrates on stochastic processes with sample paths that are uniformly continuous with respect to some nice semimetric on $[0,1]$

Extends to $\ell^\infty(T)$, where T is any set:
$\implies$ T is a family of functions: empirical processes indexed by *functions*

Empirical process indexed by functions

- $X_1, X_2, \dots$ i.i.d. random variables on some probability space Ω and into some measurable space $(\mathcal{X}, \mathcal{A})$ with common distribution P
- Family $\mathcal{F} \subset L_2(P)$ of P-square integrable functions $f : \mathcal{X} \to \mathbb{R}$

Definition: Empirical process indexed by $\mathcal{F}$. For $f \in \mathcal{F}$,

$$\mathbb{G}_n f = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathrm{E}[f(X_1)]\right)$$

Examples:

- The empirical process $\sqrt{n}(\mathbb{F}_n - F)$ above: all $f = \mathbb{1}_{[0,x]}$ for $x \in \mathbb{R}$
- Multivariate weighted empirical distribution functions
- Functions f_θ indexed by a parameter θ, e.g., parametric models
- ...

Convergence of finite-dimensional distributions

Multivariate central limit theorem. For every $(f_1, \ldots, f_\ell) \in \mathcal{F}^\ell$,

$$(\mathbb{G}_n f_1, \ldots, \mathbb{G}_n f_\ell) \xrightarrow{d} (\mathbb{G} f_1, \ldots, \mathbb{G} f_\ell), \qquad n \to \infty$$

The limit vector is centered multivariate normal with covariances

$$\mathrm{E}[\mathbb{G} f_j \, \mathbb{G} f_k] = \operatorname{cov}(f_j(X_1), f_k(X_1))$$

- *Cannot* conclude that

$$\sup_{f \in \mathcal{F}} |\mathbb{G}_n f| \xrightarrow{d} \sup_{f \in \mathcal{F}} |\mathbb{G} f|, \qquad n \to \infty$$

- CLT *uniform* in $f \in \mathcal{F}$?

Empirical process as a random bounded function

Suppose $\mathcal{F}$ is such that almost every trajectory $f \mapsto \mathbb{G}_n f$ is bounded:

- $\sup_{f \in \mathcal{F}} \left| f(x) \right| < \infty$ for P-almost every $x \in \mathcal{X}$
- $\sup_{f \in \mathcal{F}} \left| \mathrm{E}[f(X_1)] \right| < \infty$

View $\mathbb{G}_n$ as a possibly nonmeasurable map into $\ell^\infty(\mathcal{F})$:

$$\mathbb{G}_n : \Omega \to \ell^\infty(\mathcal{F})$$

Weak convergence in $\ell^\infty(\mathcal{F})$?

$$\mathbb{G}_n \overset{?}{\rightsquigarrow} \mathbb{G}, \qquad n \to \infty$$

Limit process $\mathbb{G}$ should be a *P-Brownian bridge*

- centered Gaussian process indexed by $\mathcal{F}$
- covariances as on slide 18
- "nice" trajectories $f \mapsto \mathbb{G} f$

Approach will work for some, but not all, families $\mathcal{F}$

For weak convergence of $\mathbb{G}_n$ in $\ell^\infty(\mathcal{F})$, the choice of $\mathcal{F}$ is crucial:

- For $\mathcal{X} = \mathbb{R}$ and any P, weak convergence always holds for

$$\mathcal{F} = \left\{ \mathbb{1}_{(-\infty,x]} \,\middle|\, x \in \mathbb{R} \right\}$$

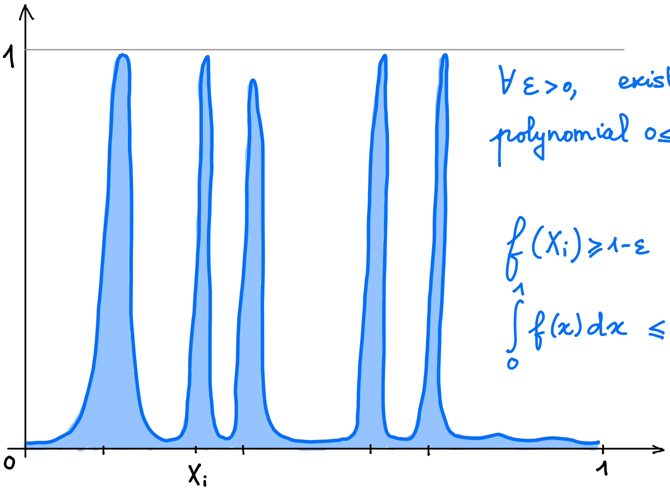
- Counterexample: for $\mathcal{X} = [0,1]$ with P the uniform distribution and

$$\mathcal{F} = \{\text{all polynomial } f : [0,1] \to [0,1]\}$$

  we have, by the Stone–Weierstrass theorem,

$$\sup_{f \in \mathcal{F}} |\mathbb{G}_n f| = \sup_{f \in \mathcal{F}} \sqrt{n} \left| \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \int_0^1 f(x)\, dx \right| = \sqrt{n}$$

$\implies$ $\mathcal{F}$-based Kolmogorov–Smirnov test breaks down



$\forall \varepsilon > 0$, exists polynomial $0 \leq f \leq 1$:

$$f(X_i) \geq 1 - \varepsilon \quad \forall i$$

$$\int_0^1 f(x)\,dx \leq \varepsilon$$

Course outline

1. Introduction
2. Stochastic Convergence in Metric Spaces
3. Glivenko–Cantelli and Donsker Theorems
4. Tools to Work with Empirical Processes

Main sources for these lectures:

- van der Vaart and Wellner (1996)
- van der Vaart (1998)
- Kosorok (2008)

References

Billingsley, P. (1968). *Convergence of Probability Measures*. New York: John Wiley.

Donsker, M. D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems. *The Annals of Mathematical Statistics 23*, 277–381.

Doob, J. L. (1949). Heuristic approach to the Kolmogorov–Smirnov theorems. *The Annals of Mathematical Statistics 20*, 292–403.

Dudley, R. (1966). Weak convergence of measures on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois Journal of Mathematics 10*, 109–126.

Hoffmann-Jørgensen, J. (1984). Stochastic Processes on Polish Spaces. Unpublished.

Hoffmann-Jørgensen, J. (1991). *Stochastic Processes on Polish Spaces*, Volume 39 of *Various Publication Series*. Aarhus, Denmark: Aarhus Universitet.

Kolmogorov, A. (1933). Sulla determinazione empirica di une legge di distributzione. *Giornale dell'Instituto Italiano degli Attuari 4*, 83–91.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer Sciences+Business Media.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer-Verlag.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*. New York: Springer Sciences+Business Media.