Empirical Processes with Applications in Statistics

2. Stochastic Convergence in Metric Spaces

Johan Segers

UCLouvain (Belgium)

Conférence Universitaire de Suisse Occidentale

Programme Doctoral en Statistique et Probabilités Appliquées

Les Diablerets, February 4–5, 2020

Overview

Introductory lecture: motivation why to consider weak convergence of non-measurable elements in $\ell^\infty(T)$ equipped with supremum distance

This lecture: weak convergence

1. first for general metric spaces $(\mathbb{D}, d)$
2. then specialized to $(\ell^\infty(T), \|\cdot\|_\infty)$

Theory goes back to Hoffmann-Jørgensen (1984, 1991) and others.
Main sources for this lecture:

- van der Vaart and Wellner (1996, Part 1)
- van der Vaart (1998, Chapter 18)

See historic notes in van der Vaart and Wellner (1996, pp. 75–78)

Stochastic Convergence in Metric Spaces

General metric spaces

Space of bounded functions

Outer expectation

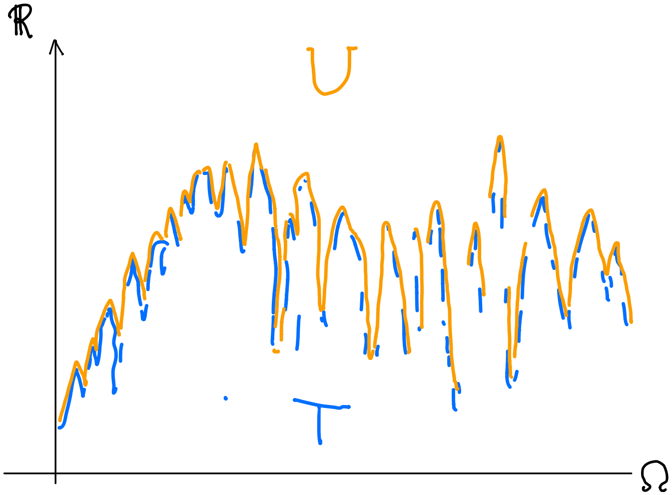
- Probability space $(\Omega, \mathcal{A}, \mathrm{P})$
- map $T : \Omega \to [-\infty, \infty]$, not necessarily measurable

Definition: Outer expectation.

$$\mathrm{E}^*[T] = \inf\Big\{\mathrm{E}[U] \;\Big|\; U : \Omega \to [-\infty, \infty] \text{ measurable}, \\ U \geqslant T \text{ and } \mathrm{E}[U] \text{ exists}\Big\}$$

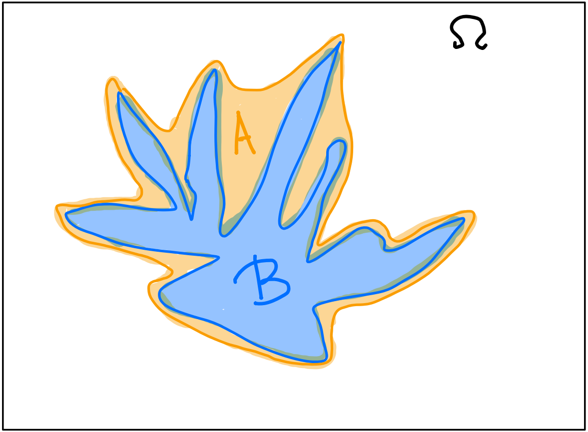
Measurable cover: Borel measurable $T^* : \Omega \to [-\infty, \infty]$ such that $T \leqslant T^* \leqslant U$ almost surely for any U as above

- $\mathrm{E}^*[T] = \mathrm{E}[T^*]$ (provided the latter expectation exists in $[-\infty, \infty]$)



ℝ
U
T
Ω

$B \subset A \subset \Omega$



Inner expectation:

$$\mathrm{E}_*[T] = -\,\mathrm{E}^*[-T]$$

Inner and outer probability:

$$\mathrm{P}^*(B) = \mathrm{E}^*[\mathbb{1}_B] = \inf\{\mathrm{P}(A) \mid B \subset A, \text{ measurable } A \subset \Omega\}$$
$$\mathrm{P}_*(B) = \mathrm{E}_*[\mathbb{1}_B] = 1 - \mathrm{P}^*(\Omega \setminus B)$$

Some care is required:

- $(S+T)^* \leqslant S^* + T^*$, but no equality in general
- Fubini's theorem no longer works
- "Law" of a random variable depends on underlying probability space
  - iid samples $X_1, X_2, \ldots, X_n, \ldots$ will be seen as coordinate projections on the product space $(\mathcal{X}^n, \mathcal{A}^n, P^n)$ or the sequence space $(\mathcal{X}^\infty, \mathcal{A}^\infty, P^\infty)$

Stochastic convergence

- Metric space $(\mathbb{D}, d)$, Borel σ-field
- Random elements $X_n, X : (\Omega, \mathcal{A}, \mathrm{P}) \to \mathbb{D}$
  X Borel measurable, but X_n not necessarily

Definition: Weak convergence. $X_n \rightsquigarrow X$ if

$$\lim_{n\to\infty} \mathrm{E}^*[f(X_n)] = \mathrm{E}[f(X)]$$

for every bounded, continuous $f : \mathbb{D} \to \mathbb{R}$

Other modes of convergence

Convergence:

- in *(outer) probability:* $X_n \overset{\mathrm{P}}{\to} X$ if $\mathrm{P}[d(X_n, X)^* > \varepsilon] \to 0$ for all $\varepsilon > 0$
- *(outer) almost surely:* $X_n \overset{\mathrm{as}*}{\to} X$ if $d(X_n, X)^* \to 0$ a.s.

The usual implications hold:

- $X_n \overset{\mathrm{as}*}{\to} X$ implies $X_n \overset{\mathrm{P}}{\to} X$
- $X_n \overset{\mathrm{P}}{\to} X$ implies $X_n \rightsquigarrow X$
- $X_n \overset{\mathrm{P}}{\to} c$ (constant) if and only if $X_n \rightsquigarrow c$

Characterizations of weak convergene

Portmanteau lemma. Are equivalent:

(i) $X_n \rightsquigarrow X$

(ii) $\liminf_{n\to\infty} \mathrm{P}_*(X_n \in G) \geqslant \mathrm{P}(X \in G)$ for every open $G \subset \mathbb{D}$

(iii) $\limsup_{n\to\infty} \mathrm{P}^*(X_n \in F) \leqslant \mathrm{P}(X \in F)$ for every closed $F \subset \mathbb{D}$

(iv) For Borel sets $B \subset \mathbb{D}$ such that $\mathrm{P}(X \in \partial B) = 0$,

$$\lim_{n\to\infty} \mathrm{P}_*(X_n \in B) = \lim_{n\to\infty} \mathrm{P}^*(X_n \in B) = \mathrm{P}(X \in B)$$

(v) ...

Characterizations mostly useful in proofs

Choice of state space not so important

Corollary

- Subset $\mathbb{D}_0 \subset \mathbb{D}$ equipped with the same metric d
- Maps $X_n, X : \Omega \to \mathbb{D}_0$

Then

$$X_n \rightsquigarrow X \text{ in } \mathbb{D} \iff X_n \rightsquigarrow X \text{ in } \mathbb{D}_0$$

Example:

- $\mathbb{D} = \ell^\infty([0, 1])$
- $\mathbb{D}_0 = \mathcal{C}([0, 1])$ or $\mathbb{D}_0 = \mathcal{D}([0, 1])$

Extracting new convergence relations

Continuous Mapping Theorem

- Metric spaces $\mathbb{D}, \mathbb{E}$
- $g : \mathbb{D} \to \mathbb{E}$ is continuous at every $x \in \mathbb{D}_0 \subset \mathbb{D}$
- $X_n \rightsquigarrow X$ in $\mathbb{D}$

If X takes values in $\mathbb{D}_0$, then $g(X_n) \rightsquigarrow g(X)$

- *Extended continuous mapping theorem*: mappings depend on n but $g_n(x_n) \to g(x)$ for sufficiently many sequences $x_n \to x$
- Will serve to prove the *functional delta method*

Extracting weakly convergent subsequences

Prohorov's Theorem. If the sequence of maps $X_n : \Omega \to \mathbb{D}$ is

- *asymptotically measurable*

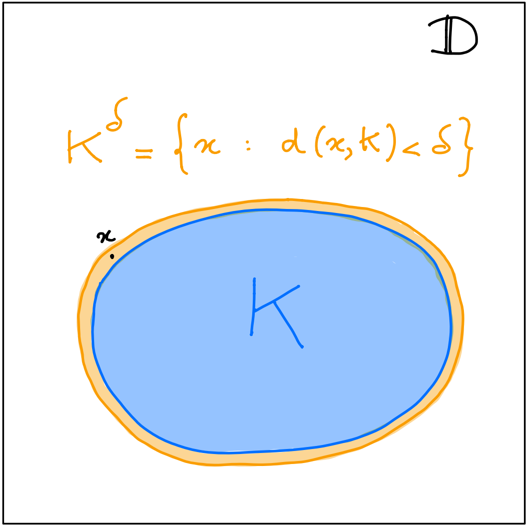
  $\lim_{n\to\infty} \mathrm{E}^*[f(X_n)] - \mathrm{E}_*[f(X_n)] = 0$ for every bounded, continuous $f : \mathbb{D} \to \mathbb{R}$

- *asymptotically tight*

  For every $\varepsilon > 0$ there exists a compact $K \subset \mathbb{D}$ such that, for every $\delta > 0$, we have $\liminf_{n\to\infty} \mathrm{P}_*[\exists y \in K : d(X_n, y) < \delta] \geqslant 1 - \varepsilon$

then it has a *subsequence* that converges weakly to a *tight* X.

For every $\varepsilon > 0$ there exists a compact $K \subset \mathbb{D}$ such that $\mathrm{P}(X \in K) \geqslant 1 - \varepsilon$.


$\mathbb{D}$
$K^\delta = \{x : d(x,K) < \delta\}$
x
K

Proof strategy

Strategy to prove weak convergence $X_n \rightsquigarrow X$:

1. Show that every subsequence of X_n has a further subsequence that convergence weakly
   - Via Prohorov's theorem
2. Show that all weak limits that can thus arise are the identical
   - For stochastic process: via finite-dimensional distributions

Little-oh calculus

Suppose $\mathbb{D}$ is actually a *Banach space*

Vector space over $\mathbb{R}$ equipped with a norm $\|\cdot\|$ such that the metric $d(x,y) = \|x - y\|$ is complete

Slutsky's lemma

If $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow c \in \mathbb{D}$ and if X is tight, then

$$X_n + Y_n \rightsquigarrow X + c$$

Common case: $c = 0 \in \mathbb{D}$. Then we write $Y_n = o_P(1)$ and thus

$$X_n \rightsquigarrow X \text{ tight} \implies X_n + o_P(1) \rightsquigarrow X$$

Stochastic Convergence in Metric Spaces

General metric spaces

Space of bounded functions

Space of bounded functions

Let $\mathbb{D}$ be the space of bounded real functions on some set T:

$$\ell^\infty(T) = \left\{ z : T \to \mathbb{R} \,\middle|\, \sup_{t \in T} |z(t)| < \infty \right\}$$

- Banach space with norm $\|z\|_\infty = \sup_{t\in T} |z(t)|$
- Supremum distance $d(z_1, z_2) = \|z_1 - z_2\|_\infty = \sup_{t\in T} |z_1(t) - z_2(t)|$
- Non-separable when T is uncountable: Borel σ-field is very large
- Natural space in which to study empirical processes such as

$$f \mapsto \mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^{n} f(X_i)$$

  indexed by $f \in \mathcal{F} = T$, for i.i.d. $X_1, X_2, \ldots$ taking values in some measurable space $(\mathcal{X}, \mathcal{A}, P)$ and $\mathcal{F}$ some subset of $L_1(P)$ or $L_2(P)$

Theorem: Weak convergence in $\ell^\infty(T)$.
Let $X_n : \Omega \to \ell^\infty(T)$. Then

$$X_n \text{ converges weakly to some tight limit}$$

if and only if the following two properties hold:

1. X_n is asymptotically tight
2. for every $(t_1, \ldots, t_k) \in T^k$, the random vectors $(X_n(t_1), \ldots, X_n(t_k))$ converge weakly in $\mathbb{R}^k$

Special case of proof strategy on slide 15

1. Asymptotic tightness: by controlling the oscillations of the trajectories (Arzèla–Ascoli)
2. Convergence of finite-dimensional distributions: via classical limit theorems

Theorem: Asymptotic tightness in $\ell^\infty(T)$.
Let $X_n : \Omega \to \ell^\infty(T)$ be such that $X_n(t) : \Omega \to \mathbb{R}$ is asymptotically tight for every $t \in T$. Are equivalent:

(i) X_n is asymptotically tight

(ii) For every $\varepsilon, \eta > 0$ there exists a finite partition $T = \bigcup_{i=1}^k T_i$ with

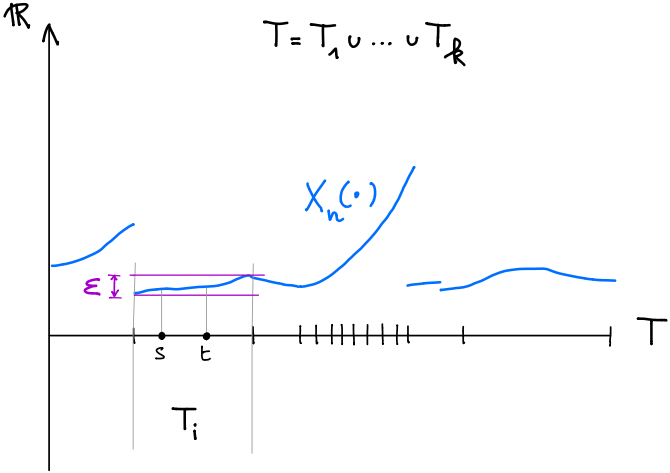
$$\limsup_{n\to\infty} \mathrm{P}^* \left[\max_{i=1,\dots,k} \sup_{s,t\in T_i} |X_n(s) - X_n(t)| > \varepsilon \right] < \eta$$

(iii) there exists a semimetric ρ on T such that (T, ρ) is totally bounded and such that for every $\varepsilon, \eta > 0$ there exists $\delta > 0$ with

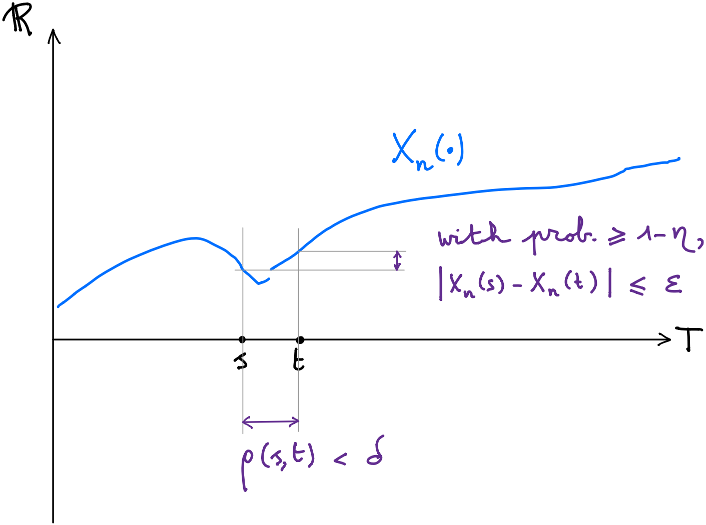
$$\limsup_{n\to\infty} \mathrm{P}^* \left[\sup_{\rho(s,t)<\delta} |X_n(s) - X_n(t)| > \varepsilon \right] < \eta$$

i.e., X_n is *asymptotically uniformly ρ-equicontinuous in probability*

(T, ρ) totally bounded: for every $\delta > 0$, there exists finitely many $t_1, \dots, t_\ell \in T$ such that for every $t \in T$, there is t_j with $\rho(t, t_j) < \delta$



ℝ
T = T1 ∪ … ∪ Tk
Xn(·)
ε
s
t
T
Ti


$\mathbb{R}$
$X_n(\cdot)$
with prob. $\geq 1-\eta$,
$|X_n(s) - X_n(t)| \leq \varepsilon$
T
s
t
$\rho(s,t) < \delta$

- If, moreover, $X_n \rightsquigarrow X$, then almost all trajectories $t \mapsto X(t)$ are uniformly ρ-continuous
- If, moreover, X is a Gaussian process, then the following semimetric always works:

$$\rho(s,t) = \left(\mathrm{E}\left[\{X(s) - X(t)\}^2\right]\right)^{1/2}, \qquad s,t \in T$$

- Techniques to control the probabilities in (ii) and (iii):
  - maximal inequalities
  - symmetrization
  - entropy: bracketing and covering numbers
  - ...

Summary

Classical theory of weak convergence in metric spaces as in Billingsley (1968) works well in *separable metric spaces*

- $\mathcal{C}([0,1])$ with supremum distance
- $\mathcal{D}([0,1])$ with Skorohod topology

The space $\ell^\infty(T)$ in empirical process theory is non-separable and requires handling non-measurable mappings: *Hoffmann-Jørgensen theory*

Classical results can be mostly recovered:

- Portmanteau lemma
- Continuous mapping theorem
- Prohorov's theorem
- Slutsky's lemma
- Tightness criteria for sequences of stochastic processes

References

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

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.

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.