Empirical Processes (MATH-522)

Lecture 3: Measure theoretic aspects of stochastic processes

Myrto Limnios

MATH, Ecole Polytéchnique Fédérale de Lausanne

March 4, 2025

What we saw last week

- We studied how to derive probabilistic bounds quantifying the deviation rate of averages around their mean, known as *concentration inequalities*.
- We focused on bounds that have exponential decay for **fixed** sample size, under various assumptions on the moments of the r.v.s

What we will focus on today

- We will extend theoretical tools to extend those results **uniformly** over a class of functionals/sets.
- We will focus on weak convergence for stochastic processes, that will result in a new set of characterizations requiring the definition of *outer-measure*.
- We will go from separable finite-dimensional metric spaces to non-separable metric spaces.

$$\sqrt{n}\left(F_n(t) - F(t)\right) \rightsquigarrow W$$

$$\sup_{t \in \mathbb{R}} \left| \sqrt{n} \left(\quad " \quad \right) \right| \rightsquigarrow B \quad \text{on } [0,1]$$

w. convergence.

Today's outline

Borel measurability

Let $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space, and $(\mathbb{D}, \mathcal{D}, d)$ a metric spaces endowed with a metric d and the σ-field/algebra $\mathcal{D}$.

- A map $h : \Omega \to \mathbb{D}$ is *$\mathcal{A}/\mathcal{D}$-measurable* if the preimage $h^{-1}(U) = \{x \in \Omega,\ h(x) \in U\}$ is measurable in $\mathcal{A}$ for all sets $U \in \mathcal{D}$.
- The *Borel σ-field* $\mathcal{B}(\mathbb{D})$ of $\mathbb{D}$ is the smallest σ-field containing all the open sets of $\mathbb{D}$.
- A function is *Borel measurable* relative to two metric spaces if it is measurable w.r.t. their Borel σ-field.
- A Borel-measurable map $X : \Omega \to \mathbb{D}$ defined on the p.s. $(\Omega, \mathcal{A}, \mathbb{P})$ is referred to as a *random element/map* valued in $\mathbb{D}$.

Remark

For Euclidean spaces, Borel measurability is the usual measurability.

We lastly recall an important result.

Lemma

A continuous map between two metric spaces is Borel measurable.

Tightness

Tightness characterizes when a measure **concentrates** on a compact set almost surely.

Definition

Let $(\mathbb{D}, d)$ be a metric space. A Borel probability measure P is *tight* if

$$\forall \varepsilon > 0, \quad \exists K \subset \mathbb{D} \text{ compact}, \quad P(K) \geq 1 - \varepsilon \ .$$

A Borel map X of distribution P is tight if P is tight.

Remark (Key fact)

This property ensures a kind of 'smoothness' for the r.v. X. Weak convergence will be extended for tight limiting r.v.s.
We can say that tightness is equivalent to being a σ-compact set (countable union of compacts) that has P-measure equal to one.

The following results show the importance of tightness.

Theorem

Let $\mathbb{D}$ be a separable and complete metric space. Then, every probability measure on $(\mathbb{D}, \mathcal{B}(\mathbb{D}))$ is tight.

Lemma

Let X and Y be two tight Borel-measurable processes in $\mathbb{D} = \ell^\infty(T)$, then $\boxed{X = Y}$ iff. their finite-dimensional (fidi) marginal distributions are equal, i.e.

$$\forall t_1, \ldots, t_k \in T, \qquad (X(t_1), \ldots, X(t_k)) = (Y(t_1), \ldots, Y(t_k)) \ ,$$

for all integer $k > 1$.

Reminder (Bounded functions)

Let T be an arbitrary set. We denote by $\ell^\infty(T)$ the class of all bounded real-valued functions $x : T \to \mathbb{R}$. We will endow the space by the *uniform norm* on T:

$$\|x\|_T = \sup_{t \in T} |x(t)| ,$$

where we define pointwise the sum $(x_1 + x_2)(t) = x_1(t) + x_2(t)$ and product with a scalar $(\alpha x)(t) = \alpha x(t)$, for all $t \in T$.

The space $\ell^\infty(T)$ contains all functions of finite sup-norm, i.e., such that $\|x\|_T < \infty$.

Property: It is separable iff. the set T is countable.

Separability

A weaker requirement to tightness is separability.

Definition

Let $(\mathbb{D}, d)$ be a metric space. We say that $X : \Omega \to \mathbb{D}$ (or its p.m P) are *separable*, if there exists a measurable separable set (i.e. it has a countable dense subset) with probability one, i.e., if $\exists K \subset \mathbb{D}$ such that $P(K) = 1$.

Definition

A σ-field is *separable* if it is generated by a countable collection of subsets.

Remark

If a topological space $\mathbb{D}$ is separable, then its Borel σ-field is separable as well.
If X is tight or separable, then it is Borel measurable. Notice that tightness and separability are independent on the metric.

Example

A Euclidean space $\mathbb{R}^d$ is separable as it is generated by a dense countable subset composed of vectors with rational coordinates.

Separability

A weaker requirement to tightness is separability.

Definition

Let $(\mathbb{D}, d)$ be a metric space. We say that $X : \Omega \to \mathbb{D}$ (or its p.m P) are *separable*, if there exists a measurable separable set (i.e. it has a countable dense subset) with probability one, i.e., if $\exists K \subset \mathbb{D}$ such that $P(K) = 1$.

Definition

A σ-field is *separable* if it is generated by a countable collection of subsets.

Remark

If a topological space $\mathbb{D}$ is separable, then its Borel σ-field is separable as well.
If X is tight or separable, then it is Borel measurable. Notice that tightness and separability are independent on the metric.

Example

A Euclidean space $\mathbb{R}^d$ is separable as it is generated by a dense countable subset composed of vectors with rational coordinates.

Separability

A weaker requirement to tightness is separability.

Definition

Let $(\mathbb{D}, d)$ be a metric space. We say that $X : \Omega \to \mathbb{D}$ (or its p.m P) are *separable*, if there exists a measurable separable set (i.e. it has a countable dense subset) with probability one, i.e., if $\exists K \subset \mathbb{D}$ such that $P(K) = 1$.

Definition

A σ-field is *separable* if it is generated by a countable collection of subsets.

Remark

If a topological space $\mathbb{D}$ is separable, then its Borel σ-field is separable as well.
If X is tight or separable, then it is Borel measurable. Notice that tightness and separability are independent on the metric.

Example

A Euclidean space $\mathbb{R}^d$ is separable as it is generated by a dense countable subset composed of vectors with rational coordinates.

Separable stochastic processes

We now define another type of separability than that related to Borel measurability of the stochastic process defined as a random map.

Definition

Let $\{Z(t),\ t \in T\}$ be a real-valued stochastic process, indexed by a separable set T.

We say that the process Z is *separable* if there exists a countable set $T' \subset T$, such that a.s.

$$\sup_{t \in T} \inf_{s \in T'} |X(t) - X(s)| = 0 \ .$$

Example

Brownian process, sub-Gaussian processes, and in particular Rademacher processes are separable in that sense.

Consequences

- Suppose T to be endowed with a semimetric ρ.
  For any point t, and for a sequence $t_m \in T'$ such that $\rho(t, t_m) \to 0$, then $|X(t) - X(t_m)| \to 0$ a.s.
- Many processes in applications will be separable in that sense, while not being Borel measurable and thus not separable w.r.t. $\mathbb{D}$.
- However, the limiting process will often be in $\mathbb{D} = \ell^\infty(T)$, where T is usually a class of functions defined on the sample space.
- Separability allows us to extract a countable sub-family of elements $\mathbb{D}_0 \subset \mathbb{D}$ of elements converging in $\overline{\mathbb{D}_0}$.

sup $|X|$

Remark

Tightness and separability are fundamental properties for random maps. Prohorov's Theorem is the key result in probability theory (not studied in this class).

Today's outline

1. First definitions: tightness and separability
2. Complete separable metric spaces
   - Random vectors
   - Finite-dimensional metric spaces
3. Non-separable metric spaces

Weak convergence of random vectors

Let $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space. Suppose the r.v.s to be valued in $\mathbb{D} \subset \mathbb{R}^d$, endowed with the Euclidean norm.

Definition

Suppose the random sequence $X_1, X_2, \ldots$ to have a distribution function F_n and p.d. P_n. X_n converges in *distribution/weakly/in law* to a r.v. X, of d.f. F and drawn from P if, for all points $x \in \mathbb{D}$ for which F is continuous,

$$F_n(x) \underset{n\to\infty}{\longrightarrow} F(x) \ .$$

We denote it by $X_n \rightsquigarrow X$ (or $P_n \rightsquigarrow P$).

Remark

- Weak convergence is **inherent to the underlying distributions** of the random maps.
- The goal is to study the properties of the limit of distribution functions when n tends to infinity (notice as well that we should consider a sequence of probability spaces but we ignore this technicality).

Weak convergence of random vectors

Let $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space. Suppose the r.v.s to be valued in $\mathbb{D} \subset \mathbb{R}^d$, endowed with the Euclidean norm.

Definition

Suppose the random sequence $X_1, X_2, \ldots$ to have a distribution function F_n and p.d. P_n. X_n converges in *distribution/weakly/in law* to a r.v. X, of d.f. F and drawn from P if, for all points $x \in \mathbb{D}$ for which F is continuous,

$$F_n(x) \underset{n \to \infty}{\longrightarrow} F(x) \, .$$

We denote it by $X_n \rightsquigarrow X$ (or $P_n \rightsquigarrow P$).

Remark

- Weak convergence is **inherent to the underlying distributions** of the random maps.
- The goal is to study the properties of the limit of distribution functions when n tends to infinity (notice as well that we should consider a sequence of probability spaces but we ignore this technicality).

Portmanteau Theorem

Weak convergence provides a series of equivalent properties referred to as *Portmanteau Theorem* that we state below.

Theorem (Portmanteau Theorem I)

Consider a random sequence of vectors $X_1, \ldots, X_n$, of p.d. P_n, and $X \sim P$. The following assertions are equivalent.

1. $X_n \rightsquigarrow X$ *or* $P_n \rightsquigarrow P$
2. $P_n h \underset{n\to\infty}{\longrightarrow} Ph$, *for all* $h \in \mathcal{C}_b(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *i.e. continuous and bounded function*
3. $\liminf_{n\infty} P_n(U) \geq P(U)$, *for all open sets* $U \subset \mathbb{R}^d$
4. $\limsup_{n\infty} P_n(F) \leq P(F)$, *for all closed sets* $F \subset \mathbb{R}^d$
5. $P_n(A) \underset{n\to\infty}{\longrightarrow} P(A)$, *for all* P*-continuity sets* A, *i.e., such that* $P(\partial A) = 0$ *where* ∂A *denotes the boundary of* A.

Continuous Mapping Theorem

Now that we have established the main characterizations, we are able to state the *Continuous Mapping Theorem* (CMT) that is fundamental to any statistical problem.

Theorem (CMT I)

Let $C \subseteq \mathbb{R}^d$ be a set such that $\mathbb{P}(X \in C) = 1$.
If $X_n \rightsquigarrow X$, then $\Phi(X_n) \rightsquigarrow \Phi(X)$ for any function $\Phi : \mathbb{R}^d \to \mathbb{R}^q$ that is continuous on C.

Proof.

Let $F \subset \mathbb{R}^d$ be a fixed closed set.

- Notice that Φ, $\{\Phi(X_n) \in F\} = \{X_n \in \Phi^{-1}(F)\}$.
- Consider $x \in \overline{\Phi^{-1}(F)}$, then, by definition of the closure, it contains all limit points of the elements in the preimge $\Phi^{-1}(F)$. Thus, there exists a sequence x_n of elements in $\Phi^{-1}(F)$, such that $x_n \to x$ and $\Phi(x_n) \in F$. If $x \in C$ then, because F is closed, $\Phi(x_n) \in F$. Otherwise $x \in C^c$. Thus

$$\Phi^{-1}(F) \subset \overline{\Phi^{-1}(F)} \subset \Phi^{-1}(F) \cup C^c \ .$$

- By the Portmanteau Theorem,

$$\limsup \mathbb{P}(\Phi(X_n) \in F) \leq \limsup \mathbb{P}(X_n \in \overline{\Phi^{-1}(F)}) \leq \mathbb{P}(X \in \overline{\Phi^{-1}(F)})$$

  Recall that $\mathbb{P}(X \in C) = 1$, hence $\mathbb{P}(X \in C^c) = 0$.
- By the last inclusion and inferring that Φ is continuous on C yields $\mathbb{P}(X \in \overline{\Phi^{-1}(F)}) \leq \mathbb{P}(X \in \Phi^{-1}(F)) = \mathbb{P}(\Phi(X) \in F)$.
- Hence assertion 4 of the Portmanteau theorem is proved, hence, because it is true for any arbitrary closed subset F, we proved by equivalence that (1) is fulfilled: $\Phi(X_n) \rightsquigarrow \Phi(X)$.

$\square$

Finite-dimensional metric spaces: Weak convergence

$\stackrel{?}{=} \mathcal{B}(\mathbb{D})$

- Consider the r.v. X as a random map $X : (\Omega, \mathcal{A}) \to (\mathbb{D}, \mathcal{D})$.
- Define the set $\mathcal{C}_b(\mathbb{D}, \mathcal{D})$ to be composed of all bounded and continuous functions $h : \mathbb{D} \to \mathbb{R}$, that are $\mathcal{D}/\mathcal{B}(\mathbb{R})$-measurable.

Definition

We say that the random sequence $\{X_n\}_{n \geq 1}$, defined on the p.s. $(\Omega, \mathcal{A}, \mathbb{P})$, converges weakly to a r.v. X, denoted $X_n \rightsquigarrow X$, if

$$\boxed{\mathbb{E}h(X_n) \underset{n\to\infty}{\longrightarrow} \mathbb{E}h(X)},$$

for all $h \in \mathcal{C}_b(\mathbb{D}, \mathcal{D})$. And we say that a sequence of p.m.s $\{P_n\}_{n\geq 1}$ converges weakly to P, denoted $P_n \rightsquigarrow P$, if

$$\boxed{P_n h \underset{n\to\infty}{\longrightarrow} Ph ,}$$

for all $h \in \mathcal{C}_b(\mathbb{D}, \mathcal{D})$.

We will try to extend the Portmanteau Theorem and the CMT in the following paragraphs, depending on whether the maps are measurable w.r.t. the Borel $\mathcal{B}(\mathbb{D})$, or not.

Borel σ-field.

We suppose for now that the space $\mathbb{D}$ is endowed with the metric d, and that $\boxed{\mathcal{D} = \mathcal{B}(\mathbb{D})}$.

Theorem (Protmanteau Theorem II)

Consider a random sequence $X_1, \dots, X_n$, of p.d. P_n, and $X \sim P$ of the metric space $(\mathbb{D}, \mathcal{D})$. The following assertions are equivalent.

1. $X_n \rightsquigarrow X$ *or* $P_n \rightsquigarrow P$
2. $\liminf_{n\infty} P_n(U) \geq P(U)$, *for all open sets* $U \subset \mathbb{R}^d$
3. $\limsup_{n\infty} P_n(F) \leq P(F)$, *for all closed sets* $F \subset \mathbb{R}^d$
4. $P_n(A) \underset{n\to\infty}{\longrightarrow} P(A)$, *for all P-continuity sets A, i.e., such that* $P(\partial A) = 0$.

The Portmanteau Theorem allows us to state the CMT II.

CMT II

Theorem (CMT II)

Consider two metric spaces $(\mathbb{D}, \mathcal{B}(\mathbb{D}))$ *and* $(\mathbb{E}, \mathcal{B}(\mathbb{E}))$*, and let the sequence of r.v.s* $X_1, \dots, X_n$ *valued in* $(\mathbb{D}, \mathcal{D})$*.*

If $X_n \rightsquigarrow X$*, then* $\Phi(X_n) \rightsquigarrow \Phi(X)$ *in* $(\mathbb{E}, \mathcal{B}(\mathbb{E}))$*, for any measurable continuous function* $\Phi : \mathbb{D} \to \mathbb{E}$*.*

Proof.

Let $h \in \mathcal{C}_b(\mathbb{D}, \mathcal{D})$, then, by continuity of $\Phi : \mathbb{D} \to \mathbb{E}$, the function $h \circ \Phi$ is bounded and continuous. Suppose that $X_n \rightsquigarrow X$, then by definition of weak convergence,

$$\mathbb{E}h \circ \Phi(X_n) \underset{n \to \infty}{\longrightarrow} \mathbb{E}h \circ \Phi(X)$$

in $(\mathbb{E}, \mathcal{B}(\mathbb{E}))$. As it holds true for any $h \in \mathcal{C}_b(\mathbb{D}, \mathcal{B}(\mathbb{D}))$, then the theorem is proved. □

In fact, we can prove the following convenient formulation of the CMT, where, as in CMT I (12), we only require the function Φ **to be continuous on the image set of the limiting r.v.**

CMT II

Theorem (CMT II)

Consider two metric spaces $(\mathbb{D}, \mathcal{B}(\mathbb{D}))$ *and* $(\mathbb{E}, \mathcal{B}(\mathbb{E}))$*, and let the sequence of r.v.s* $X_1, \ldots, X_n$ *valued in* $(\mathbb{D}, \mathcal{D})$*.*

If $X_n \rightsquigarrow X$*, then* $\Phi(X_n) \rightsquigarrow \Phi(X)$ *in* $(\mathbb{E}, \mathcal{B}(\mathbb{E}))$*, for any measurable continuous function* $\Phi : \mathbb{D} \to \mathbb{E}$*.*

Proof.

$\in \mathcal{C}_b$

Let $h \in \mathcal{C}_b(\mathbb{D}, \mathcal{D})$, then, by continuity of $\Phi : \mathbb{D} \to \mathbb{E}$, the function $\boxed{h \circ \Phi}$ is bounded and continuous. Suppose that $X_n \rightsquigarrow X$, then by definition of weak convergence,

$$\boxed{\mathbb{E}h \circ \Phi(X_n) \underset{n\to\infty}{\longrightarrow} \mathbb{E}h \circ \Phi(X)}$$

in $(\mathbb{E}, \mathcal{B}(\mathbb{E}))$. As it holds true for any $h \in \mathcal{C}_b(\mathbb{D}, \mathcal{B}(\mathbb{D}))$, then the theorem is proved. □

In fact, we can prove the following convenient formulation of the CMT, where, as in CMT I (12), we only require the function Φ **to be continuous on the image set of the limiting r.v.**

CMT IIbis

Theorem (CMT IIbis)

Consider two metric spaces $(\mathbb{D}, \mathcal{B}(\mathbb{D}))$ *and* $(\mathbb{E}, \mathcal{B}(\mathbb{E}))$*, and let the sequence of r.v.s* $X_1, \ldots, X_n$ *valued in* $(T, \mathcal{T})$*. Let* $C \subseteq \mathbb{R}^d$ *be a set such that* $\mathbb{P}(X \in C) = 1$*.*

If $X_n \rightsquigarrow X$*, then* $\Phi(X_n) \rightsquigarrow \Phi(X)$ *in* $(\mathbb{E}, \mathcal{B}(\mathbb{E}))$*, for any measurable function* $\Phi : \mathbb{D} \to \mathbb{E}$ *that is continuous on* C*.*

Proof.

Board □

Sub-field of $\mathcal{B}(\mathbb{D})$.

We suppose now the space T to be endowed with the metric d, and that $\mathcal{D}$ is a sub-σ-field of $\mathcal{B}(\mathbb{D})$. We will see if we can prove a similar CMT IIbis in that context, for any measurable map $\Phi : (\mathbb{D}, \mathcal{D}) \to (\mathbb{E}, \mathcal{E})$ continuous on the image set of points of X.

We show that CMT IIbis holds true if we suppose in addition that C is separable and $\mathcal{D}$-measurable.

Theorem (CMT III)

Consider two metric spaces $(\mathbb{D}, \mathcal{B}(\mathbb{D}))$ and $(\mathbb{E}, \mathcal{B}(\mathbb{E}))$, and let the sequence of r.v.s $X_1, \dots, X_n$ valued in $(\mathbb{D}, \mathcal{D})$. Let $C \subseteq \mathbb{R}^d$ be a separable and $\mathcal{D}$-measurable set such that $\mathbb{P}(X \in C) = 1$. If $X_n \rightsquigarrow X$, then $\Phi(X_n) \rightsquigarrow \Phi(X)$ in $(\mathbb{E}, \mathcal{B}(\mathbb{E}))$, for any measurable function $\Phi : \mathbb{D} \to \mathbb{E}$ that is continuous on C.

Exercise.

Hint: Let $h \in \mathcal{C}_b(\mathbb{D}, \mathcal{D})$, then the function $h \circ \Phi$ is continuous, bounded on C. To prove that $\Phi(X_n) \rightsquigarrow \Phi(X)$, we construct a countable sub-family of $\boxed{\mathcal{G} = \{g \in \mathcal{C}_b(\mathbb{D}, \mathcal{D}),\ g \leq h \circ \Phi,\ \text{uniformly continuous}\}}$ such that there exists an increasing sequence $\{g_k\}_{k \geq 1}$, for any completely regular point $x \in C$, $\sup_k g_k(x) = \sup_{g \in \mathcal{G}} g(x) = (h \circ \Phi)(x)$. Inferring Portmanteau II.2 and Theorem of monotone convergence concludes the proof. □

Corollary

If $\mathbb{E}h(X_n) \underset{n\to\infty}{\longrightarrow} \mathbb{E}h(X)$ *for any function* h *that is bounded, uniformly continuous and* $\mathcal{D}$*-measurable, and if* X *concentrates on a separable set of completely regular points, then* $\boxed{X_n \rightsquigarrow X}$*.*

We highlight the importance of building a countable subset of a family of the type of $\mathcal{G}$ that will be fundamental in the next chapters.

Today's outline

1. First definitions: tightness and separability

2. Complete separable metric spaces

3. Non-separable metric spaces
   - Important example of non-measurability
   - Outer measure: definition and properties
   - Bounded stochastic processes

Space of cadlag functions on a compact: Skorohod space

$\bar{R} = [-\infty, \infty]$

- Suppose that $T = [a, b]$ possibly the extended real line. The space of functions $h : [a, b] \to \mathbb{R}$, being right-continuous with left limits that exist (càdlàg) is defined by $D(T, \mathbb{R})$ (or $D(T)$).
- The space $D([a, b])$ is NOT separable

What does it mean and why do we care about it?

Empirical measures

$([0,1])$

- Let the i.i.d. sequence of Uniform r.v.s $X_1, \ldots, X_n$ defined on the probability space $\Omega =$ $([0,1], \mathcal{B}, \lambda)$, with $\mathcal{B}$ the Borel σ-field and λ the Lebesgue measure both on $[0,1]$.
- Then, the empirical c.d.f. is defined by

$$\boxed{F_n(t) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{0 \le X_i \le t\}}} \ , \quad , t \in [0,1] .$$

"LLN"

GC

and we can define the standard empirical process by

$$Z_n(t) = \sqrt{n}(F_n(t) - F(t)) = \sqrt{n}(F_n(t) - t) \ .$$

"CLT"

Donsker

- Then, both F_n, Z_n are maps defined on $[0,1]$ and valued in the space of $\ell^\infty([0,1])$ (in fact $\boxed{D[0,1]}$, but are not continuous). If we endow this space with the sup-norm on $[0,1]$, then those maps are no longer Borel measurable, insofar as $F_n^{-1}(\mathcal{D}) \not\subset \mathcal{B}^n$, where $\mathcal{D}$ is the Borel σ-field of $\mathbb{D}$.

$\shortparallel$

$D[0,1]$

Proof.

"F_n, Z_n."

- Let $K \subset [0,1]$ be not a Borel set.
- Conisder $X(\omega, t) = 1\{\omega \le t \le 1\}$, for any event ω. Then define the union of uniform open balls of events in K by $G = \cup_{k \in K}\{y, \; \|y(k) - 1_{[k,1]}\|_{[0,1]} < 1/2\}$.
- Because G is an uncountable union of open sets, it is open.
- If X was $\mathcal{B}/\mathcal{B}(D([0,1]))$-measurable, then the set $\{\omega \in [0,1], \; X(\omega) \in G\}$ would belong to $\mathcal{B}$. But

  [0,1]
- The map X valued at the event $\omega \in [0,1]$ is in G iff. $\boxed{\omega = k.}$
- Conclude that $\{\omega \in [0,1], \; X(\omega) \in G\} = \boxed{K.}$
- This is true for any subset of $[0,1]$, hence X is Borel measurable iff. any $K \subset [0,1]$ is a Borel set. Hence, we can found a non Borel set such that X is not Borel measurable.

□

Remark

- Notice that the limit process of Z_n is a Brownian Bridge, that is Borel measurable.
- This example illustrates that even for the most classical problem, the empirical processes are not necessarily Borel measurable random maps w.r.t. the sup-norm, i.e., $\boxed{Z_n^{-1}(U)}$ might be *too* large to be measurable. The interpretation being that the Borel σ-field generated by the Uniform distribution contains too many sets.
- In addition, the space $(D[0,1], \mathcal{B}(D[0,1]))$ is not separable.

We can prove that by considering the *ball σ-field*, endowed with the sup-norm, we can extend a characterization of weak convergence by considering fidi projections of distributions, but we will not go through this path (cf. lecture notes for more details).

Outer-measures for non-separable metric spaces

In fact, it has been highlighted (J. Hoffmann-Jørgensen) that Borel measurability of each X_n is not necessary for weak convergence, as soon as the limiting variable **IS** Borel-measurable, and requiring the convergence in expectation in terms of *outer expectations*.

Outer measure

Let $(\Omega, \mathcal{A}, P)$ be an arbitrary probability space and let $X : \Omega \to \overline{\mathbb{R}}$ be an arbitrary random map.

Definition

The *outer integral* w.r.t. P is defined by

$$E^* X = \inf\{EU : \quad U \geq X,\ U : \Omega \to \overline{\mathbb{R}} \ \text{ measurable and } \ EU \ \text{ exists}\}\ ,$$

where EU exists if both its positive and negative parts are finite.

In particular, if $B \subset \Omega$, then the *outer probability* is given by

$$P^*(B) = \inf\{P(A) : B \subset A,\ A \in \mathcal{A}\}\ .$$

Minimal measurable majorant

The infimum in both outer integral and probability are always achieved, and in particular if there exists a measurable *envelope function*.

Lemma

For any map $X : \Omega \to \overline{\mathbb{R}}$*, there exists a measurable function* $X^* : \Omega \to \overline{\mathbb{R}}$ *such that*

- $X^* \geq X$
- $X^* \leq U$ *a.s., for every measurable map* U *such that* $U \geq X$ *a.s.*

If EX^* *exists (in particular if* $E^*X < \infty$*), and if both statements are fulfilled, then* $E^*X = EX^*$.

In that case, X^* *is called the minimal measurable majorant/ measurable cover or envelope function of* X.

Remark

- We can see X^* as the smallest measurable function above X.

Maximal measurable minorant

Similarly, we can define a maximal measurable minorant as follows.

Lemma

For any map $X : \Omega \to \overline{\mathbb{R}}$, *there exists a measurable function* $X_* : \Omega \to \overline{\mathbb{R}}$ *such that*

- $X_* \leq X$
- $X_* \geq L$ *a.s., for every measurable map* L *such that* $L \leq X$ *a.s.*

If EX_* *exists (in particular if* $E_* X < \infty$*), and if both statements are fulfilled, then* $\boxed{E_* X = EX_*}$

In that case, X^* *is called the maximal measurable minorant of* X.

Definition

The *inner probability* is of an arbitrary subset $B \subset \Omega$ is given by

$$\boxed{P_*(B) = 1 - P^*(\Omega - B)}.$$

If X measurable

$X^* = X_* = X$

Remark

- We can see X_* as the biggest measurable function smaller than X.
- Notice that $\boxed{E_* X = -E^*[-X]}$.

Weak convergence

Now we can extend weak convergence using outer integrals.

Definition

Let a sequence of random maps $X_n : \Omega \to \mathbb{D}$ be defined on the probability space $(\Omega, \mathcal{A}, P)$. We say X_n *converges weakly* to a Borel measurable $X : \Omega \to \mathbb{D}$, denoted by $X_n \rightsquigarrow X$, if

$$\boxed{E^* h(X_n) \underset{n\to\infty}{\longrightarrow} Eh(X)}, = \int h(x) dP(x)$$

because X is measurable.

for every $h \in C_b(\mathbb{D}, \mathbb{R})$, where the limit can be written in terms of the law of X.

Notice that because the limit r.v. is Borel, then it has a distribution.

Convergence in probability and almost surely

Definition

A sequence of random maps $X_n : \Omega \to \mathbb{D}$ *converges in probability* to X if, for all $\varepsilon > 0$,

$$P^*(d(X_n, X) > \varepsilon) \underset{n\to\infty}{\longrightarrow} 0 \ .$$

We denote it by $X_n \xrightarrow{\mathbb{P}^*} X$.

Definition

A sequence of random maps $X_n : \Omega \to \mathbb{D}$ *converges almost surely* to X if, there exists a sequence of measurable random variables δ_n, such that

$$\boxed{d(X_n, X) \leq \delta_n,} \quad \text{and} \quad \delta_n \xrightarrow{a.s.} 0 \ .$$

We denote it by $\boxed{X_n \xrightarrow{a.s.*} X.}$

Theorem (Portmanteau Theorem III)

Let a sequence of random maps $X_n : \Omega \to \mathbb{D}$, and $X : \Omega \to \mathbb{D}$. Then the following assertions are equivalent.

1. $E^*h(X_n) \underset{n\to\infty}{\longrightarrow} Eh(X)$ *for any real-valued bounded continuous function* h.
2. $E^*h(X_n) \underset{n\to\infty}{\longrightarrow} Eh(X)$ *for any real-valued bounded Lipschitz function* h.
3. $\liminf_{n\infty} P_n^*(U) \geq P(U)$, *for all open sets* $U \subset \mathbb{D}$.
4. $\limsup_{n\infty} P_n^*(F) \leq P(F)$, *for all closed sets* $F \subset \mathbb{D}$.
5. $P_n^*(A) \underset{n\to\infty}{\longrightarrow} P(A)$, *for all Borel* P*-continuity sets* A $(P(\partial A) = 0)$.

Theorem (CMT IV)

Let $\Phi : \mathbb{D} \to \mathbb{E}$ *be a continuous mapping for all points in* $\mathbb{D}_0 \subset \mathbb{D}$*. Suppose the process* $X_n \rightsquigarrow X$*, with* X *being valued in* $\mathbb{D}_0$*, then* $\Phi(X_n) \rightsquigarrow \Phi(X)$*.*

Example

Suppose the process Z_n to be indexed by a Donsker class of measurable functions $\mathcal{H}$. Then, because the sup-norm can be viewed as a UC mapping on $\ell^\infty(\mathcal{H})$: $|\|x\|_\mathcal{H} - \|y\|_\mathcal{H}| \leq \|x - y\|_\mathcal{H}$. Then, we can build confidence intervals for the sup-norm of the scaled empirical process $\|\sqrt{n}(P_n - P)\|_\mathcal{H}$, where we apply Theorem 27 with $\Phi = \|\cdot\|_\mathcal{H}$.

Bounded stochastic processes

- A stochastic process $X = \{X_t,\ t \in T\}$ is a collection of r.v.s $X_t : \Omega \to \mathbb{R}$, indexed by a set T and defined on a p.s.
- For fixed $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is called a *sample path*. **It is useful to think of a stochastic process as a random function, of realizations being the sample paths, instead of a collection of r.v.s.**
- If every sample path is **bounded**, then we can view X as a random map $X : \Omega \to \ell^\infty(T)$.

  Because T is usually not finite, the space $\ell^\infty(T)$ is not separable, but we can extend the theory of weak convergence if the limit laws are **tight Borel p.m. on** $\ell^\infty(T)$.

Key result

Weak convergence of a sequence of sample bounded processes $\Longleftrightarrow$ weak convergence of the fidi distributions + asymptotic equicontinuity !!!

Bounded stochastic processes

- A stochastic process $X = \{X_t,\ t \in T\}$ is a collection of r.v.s $X_t : \Omega \to \mathbb{R}$, indexed by a set T and defined on a p.s.
- For fixed $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is called a *sample path*. **It is useful to think of a stochastic process as a random function, of realizations being the sample paths, instead of a collection of r.v.s.**
- If every sample path is **bounded**, then we can view X as a random map $X : \Omega \to \ell^\infty(T)$.

  Because T is usually not finite, the space $\ell^\infty(T)$ is not separable, but we can extend the theory of weak convergence if the limit laws are **tight Borel p.m. on** $\ell^\infty(T)$.

Key result

Weak convergence of a sequence of sample bounded processes $\Longleftrightarrow$ weak convergence of the fidi distributions + asymptotic equicontinuity !!!

Bounded stochastic processes

- A stochastic process $X = \{X_t,\ t \in T\}$ is a collection of r.v.s $X_t : \Omega \to \mathbb{R}$, indexed by a set T and defined on a p.s.
- For fixed $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is called a *sample path*. **It is useful to think of a stochastic process as a random function, of realizations being the sample paths, instead of a collection of r.v.s.**
- If every sample path is **bounded**, then we can view X as a random map $X : \Omega \to \ell^\infty(T)$.

  Because T is usually not finite, the space $\ell^\infty(T)$ is not separable, but we can extend the theory of weak convergence if the limit laws are **tight Borel p.m. on** $\ell^\infty(T)$.

Key result

Weak convergence of a sequence of sample bounded processes $\Longleftrightarrow$ weak convergence of the fidi distributions + asymptotic equicontinuity !!!

Bounded stochastic processes

- A stochastic process $X = \{X_t,\ t \in T\}$ is a collection of r.v.s $X_t : \Omega \to \mathbb{R}$, indexed by a set T and defined on a p.s.
- For fixed $\omega \in \Omega$, the map $\boxed{t \mapsto X_t(\omega)}$ is called a *sample path*. **It is useful to think of a stochastic process as a random function, of realizations being the sample paths, instead of a collection of r.v.s.**
- If every sample path is **bounded**, then we can view X as a random map $\boxed{X : \Omega \to \ell^\infty(T)}$

  $X : \Omega \to D$

  Because T is usually not finite, the space $\ell^\infty(T)$ is not separable, but we can extend the theory of weak convergence if the limit laws are **tight Borel p.m. on** $\ell^\infty(T)$.

Key result

Weak convergence of a sequence of sample bounded processes $\Longleftrightarrow$ weak convergence of the fidi distributions + asymptotic equicontinuity !!!

Theorem

We say that $X_n : \Omega_n \to \boxed{\ell^\infty(T)}$ *converges weakly to a* $\boxed{\textit{tight}}$ *process* X *iff. the following two assertions hold true:*

$\forall k \geq 1 \quad \forall t_1, \dots, t_k \in T$

(i) *Convergence of all finite-dimensional distributions:* $\boxed{X_n \overset{fidi}{\rightsquigarrow} X}$ $\quad (X(t_1), \dots, X(t_k)) \rightsquigarrow (X(t_1), \dots, X(t_k))$

(ii) *Asymptotic equicontinuity: there exists a semimetric* ρ *that makes* T *totally bounded, and*

$$\forall \varepsilon > 0, \quad \lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}^* \left\{ \sup_{\rho(s,t) < \delta,\ s,t \in T} |X_n(t) - X_n(s)| > \varepsilon \right\} = 0 \ .$$

$\rho(s,t) < \delta$

T

$t_1 \quad s$

$X(t) \quad X(s)$

$|\cdot|$ increment

Remark (How should we prove and interpret the second condition?)

For (ii) to hold true, we can upperbound it by using Markov Inequality, for all $\varepsilon > 0$,

$$\mathbb{P}^* \left\{ \sup_{\rho(s,t) < \delta,\ s,t \in T} |X_n(t) - X_n(s)| > \varepsilon \right\} \leq (1/\varepsilon) \mathbb{E}^* \left[\sup_{\rho(s,t) < \varepsilon,\ s,t \in T} |X_n(t) - X_n(s)| \right]$$

We hope that controlling the tail fluctuations of the increments $X_n(t) - X_n(s)$, would result in a *nice* behavior over the whole sample path.

Theorem

We say that $X_n : \Omega_n \to \ell^\infty(T)$ converges weakly to a tight process X iff. the following two assertions hold true:

(i) *Convergence of all finite-dimensional distributions:* $X_n \overset{fidi}{\rightsquigarrow} X$

(ii) *Asymptotic equicontinuity: there exists a semimetric ρ that makes T totally bounded, and*

$$\forall \varepsilon > 0, \quad \lim_{\delta \to 0} \limsup_{n \to \infty} \underbrace{\mathbb{P}^* \left\{ \sup_{\rho(s,t)<\delta,\ s,t \in T} |X_n(t) - X_n(s)| > \varepsilon \right\}} = 0 \ .$$

Remark (How should we prove and interpret the second condition?)

For (ii) to hold true, we can upperbound it by using Markov Inequality, for all $\varepsilon > 0$,

$$\mathbb{P}^* \left\{ \overbrace{\sup_{\rho(s,t)<\delta,\ s,t \in T} |X_n(t) - X_n(s)|}^{=: Z} > \varepsilon \right\} \leq (1/\varepsilon) \mathbb{E}^* \left[\underbrace{\sup_{\rho(s,t)<\varepsilon,\ s,t \in T} |X_n(t) - X_n(s)|}_{=: Z} \right]$$

We hope that controlling the tail fluctuations of the increments $X_n(t) - X_n(s)$, would result in a *nice* behavior over the whole sample path.

To sum up

- We rigorously defined weak convergence for separable and non-separable spaces.
- Key ingredients for the proof are related to considering a subclass of functions with measurable envelope function.
- The outer-measure ($\mathbb{E}^*, \mathbb{P}^*$) yields a novel definition of weak convergence for the classes of stochastic processes that we will study until the end of the semester.

References

- Aad W. Vaart, Jon A. Wellner, *Weak Convergence and Empirical Processes. With Applications to Statistics.* Springer Series in Statistics (SSS), Springer New York, 1996.
- Aad W. Vaart, *Asymptotic Statistics.* Cambridge University Press; 1998.
- Michael R. Kosorok, *Introduction to Empirical Processes and Semiparametric Inference.* Springer Series in Statistics (SSS), Springer New York, 2008.

Next lecture's program

$\{Z(t),\ t \in T\}$
size$(Z_t, t \in T) \Leftrightarrow$ structure / complexity of T.

It will be on the 18th of March!!

- We will study how to characterize the *size* of an index class of stochastic processes through their *complexity*.
- We will see that it is a key property to bound this size in order to control the uniform deviations of the process.