# Mathematical foundations of Infinite-Dim Statistical models
# 3.7 Empirical Processes (3.7.1 $\sim$ 3.7.2)

Presented by Jiin Seo

January 18, 2019

# Outline

3.7 Limit Theorems for Empirical Processes

3.7.1 Some Measurability

3.7.2 Uniform Laws of Large Numbers (Glivenko-Cantelli Theorems)

# Outline


3.7 Limit Theorems for Empirical Processes

3.7.1 Some Measurability

3.7.2 Uniform Laws of Large Numbers (Glivenko-Cantelli Theorems)

# 3.7 Limit Theorems for Empirical Processes

- Asymptotic properties of empirical processes.
- Givenko-Cantelli LLN and Kolmogorov-Doob-Donsker-Dudley CLT
- $F$ : cumulative distn. ftn of a probability measure $P$
  $F_n$ : cumulative distn. ftn corresponding to an indep. sample from $P$

$$||F_n - F||_\infty \to 0 \ a.s.,$$

$$\sqrt{n}(F_n(t) - F(t)) \xrightarrow{\text{law in } l_\infty} G_P, \text{(Centred Gaussian process)} \ , t \in \mathbb{R}$$

- By the continuous mapping theorem,

$$\sqrt{n}||F_N - F||_\infty \xrightarrow{dist^n} ||G_P||_\infty$$

- The same is true for any other continuous functional on $l_\infty(\mathbb{R})$.

# 3.7 Limit Theorems for Empirical Processes

- LLN
- CLT for the empirical process indexed by a class $\mathcal{F}$ of functions by carefully defining convergence in law of processes with bounded paths. (i.e. random elements defined on the space $l_\infty(\mathcal{F})$ of all bdd ftns $H : \mathcal{F} \mapsto \mathbb{R}$)
- $l_\infty(\mathcal{F})$ is a non-separable metric space, and in order to recover the uniform tightness property associated to convergence in law, the definition asks for the limiting process to have a tight Borel probability law in this space.
- Skorokhod representation
- CLT for empirical processes(permanence properties and extension by convexity)

# Outline


3.7 Limit Theorems for Empirical Processes

3.7.1 Some Measurability

3.7.2 Uniform Laws of Large Numbers (Glivenko-Cantelli Theorems)

## 3.7.1 Some Measurability

- The definition of convergence in law in the nonseparable space $l_\infty(\mathcal{F})$ does require the notion of outer expectation as soon as $\mathcal{F}$ is infinite

Definition [ Outer probability ]
Let $(\Omega, \Sigma, P)$ be a prob. space, and let $A \subset \Omega$ be a not necessarily measurable set.
The outer probability $P^*(A)$ of $A \subseteq \Omega$ is defined as

$$P^*(A) = inf\{P(C) : A \subseteq C, C \in \Sigma\}. \qquad (3.241)$$

$P^*(A) = P(A)$ if $A$ is measurable.

# 3.7.1 Some Measurability

Definition [ Outer expectation ]
With the notation $Eg := \int gdP$ for $g$ measurable, if $f : \Omega \mapsto [-\infty, \infty]$ is not measurable, we may also define its outer expectation or integral as

$$\int^* fdP = E^*f = inf\{Eg : g \geq f, g\ measurable, [-\infty, \infty]-valued\}, \qquad (3.242)$$

except that $E^*f$ is undefined if there exist a measurable function $g \geq f$ s.t. $Eg^+ = Eg^- = \infty$ and no measurable function $g \geq f$ s.t. $Eg = -\infty$.

$Eg$ exists if at most one of $Eg^+$ and $Eg^-$ is infinite, and then
$Eg = Eg^+ - Eg^-$.

## 3.7.1 Some Measurability

Set

$$\mathcal{C}_A = \{C : A \subseteq C, C \in \Sigma\}, \; \mathcal{G}_f = \{g \geq f : g \; measurable, [-\infty, \infty]-valued\}$$

and note that $\Omega \in \mathcal{C}_A$ and $\infty \in \mathcal{G}_f$, so outer probabilities always exist and outer expectations exist or are undefined.

## 3.7.1 Some Measurability

The infimum in (3.241) and (3.242) are, respectively, attained at a P a.s. unique set in $\mathcal{C}_A$ and a $P$ a.s. unique function in $\mathcal{G}_f$ :

**Proposition 3.7.1**

(a) For every set $A \subset \Omega$, the infimum in the definition (3.241) of $P^*(A)$ is attained at a measurable set $A^* \in \mathcal{C}_A$ which is $P$ a.s. uniquely determined. In particular, $P^*(A) = P(A^*)$.

(b) For every function $f : \Omega \mapsto \bar{R}$, there exists a $P$ a.s. unique function $f^* \in \mathcal{G}_f$ s.t. $f^* \leq g$ $P$ a.s. for every $g \in \mathcal{G}_f$ . Then, if either of $E^*f$ or $Ef^*$ is defined, both are equal, as is the case, for example, if f is bounded above or below.

(c) For any set $A \in \Omega$, $(I_A)^* = I_{A^*}$ a.s. and hence $P^*(A) = E^*(I_A)$.

## 3.7.1 Some Measurability

- $A^*$ : P-measurable cover of the set $A$
- $f^*$ : P-measurable cover of the function $f$
- It will also be convenient to call a function $F$ a P-measurable envelope of $f$ if $F \geq f^*$ P a.s. and likewise for sets.
  Note that if $P$ and $Q$ are mutually absolutely continuous, the $P-$ and $Q-$measurable covers of $f$ coincide and likewise for measurable envelopes.

## 3.7.1 Some Measurability

**Proposition 3.7.2**

(a) For any two functions $f, g : \Omega \mapsto (-\infty, \infty]$, we have

$$(f+g)^* \leq f^* + g^* \ a.s. \text{ and } (f-g)^* \geq f^* - g^*,$$

where the second inequality requires that both sides be defined.

(b) For $f : \Omega \mapsto \mathbb{R}$, $t \in \mathbb{R}$ and $\epsilon > 0$,

$$P^*\{f > t\} = P\{f^* > t\} \text{ and } P^*\{f \geq t\} \leq P\{f^* \geq t\} \leq P^*\{f \geq t \grave{} \epsilon\}.$$

(c) If $B$ is a vector space with a pseudo-norm $||\cdot||$, , then for any functions $f, g : \Omega \mapsto B$,

$$||f+g||^* \leq ||f||^* + ||g||^* a.s. \text{ and } ||cf||^* = |c|||f||^* a.s.$$

## 3.7.1 Some Measurability

The following one-sided Fubini-Tonelli theorem is an important tool in the calculus of nonmeasurable functions and it will be used often:

Proposition 3.7.3 [ One-sided Fubini-Tonelli theorem]
Let $(X \times Y, \mathcal{A} \bigotimes \mathcal{B}, P \times Q)$ be a product probability space. Let $f : X \times Y \mapsto [0, \infty)$ , and let $f^*$ be its measurable cover w.r.t. $P \times Q$. Let $E_P^*$ and $E_Q^*$ denote, respectively, the outer expectations w.r.t. $P$ and $Q$. Then

$$E_P^* E_Q^* f \leq E(f^*), \qquad E_Q^* E_P^* f \leq E(f^*).$$

If, moreover, $\mathcal{Q}$ is discrete and $\mathcal{B}$ is the collection of all the subsets of $Y$, then

$$E_P^* E_{\mathcal{Q}} f \leq E(f^*) = E_{\mathcal{Q}} E_P^* f.$$

## 3.7.1 Some Measurability

We introduce a concept that will be useful when extending to the nonmeasurable setting Skorokhod's theorem about a.s. convergent representations of sequences of random variables that converge in distribution.

Let $\phi : (\tilde{\mathbf{X}}, \tilde{\mathcal{A}}) \mapsto (\mathbf{X}, \mathcal{A})$ be measurable, let $\tilde{P}$ be a probability measure on $\tilde{\mathcal{A}}$ and let $\tilde{P} \circ \phi^{-1}$ be the probability law of $\phi$. Then, if $f : X \mapsto \mathbb{R}$ is arbitrary, we have $f^* \circ \phi \geq f \circ \phi$, where $f^*$ is the $\tilde{P} \circ \phi^{-1}$-measurable cover of $f$ and hence $f^* \circ \phi$ is $\tilde{P}$-measurable and therefore $f^* \circ \phi \geq (f \circ \phi)^*$ $\tilde{P}$ a.s.

## 3.7.1 Some Measurability

Definition 3.7.5 [ $\tilde{P}$-perfect ] A measurable map $\phi : \tilde{X} \mapsto X$ is $\tilde{P}$-perfect if $f^* \circ \phi = (f \circ \phi)^*$ $\tilde{P}$ a.s. for every bounded function $f : X \mapsto \mathbb{R}$ , where $(f \circ \phi)^*$ is the $\tilde{P}$-measurable cover of $f \circ \phi$ and $f^*$ is the $P \circ \phi^{-1}$-measurable cover of $f$.

Then, if $\phi$ is perfect and $f$ is bounded,

$$\begin{aligned} E^*_{\tilde{P}}(f \circ \phi) &= \int (f \circ \phi)^* d\tilde{P} = \int f^* \circ \phi d\tilde{P} = \int f * d(\tilde{P} \circ \phi^{-1})`1) \\ &= \int^* fd(\tilde{P} \circ \phi^{-1}) = E^*_{\tilde{P} \circ \phi^{-1}} f, \qquad (3.243) \end{aligned}$$

or, for indicators, $\tilde{P}^*\{\phi \in A\} = (\tilde{P} \circ \phi^{-1})^*(A)$ for any $A \subset X$ . It is this property that will make perfectness useful.

# Outline

3.7 Limit Theorems for Empirical Processes

3.7.1 Some Measurability

3.7.2 Uniform Laws of Large Numbers (Glivenko-Cantelli Theorems)

# 3.7.2 Uniform Laws of Large Numbers (G-C Theorems)

- Given $X_i, i \in \mathbb{N}$, on $(\Omega, \Sigma, Pr) := (S, \mathcal{S}, P)^{\mathbb{N}}$, the product of countably many copies of $(S, \mathcal{S}, P)$ and a collection of real-valued measurable functions $\mathcal{F}$ on $S$.
- We are now interested in obtaining conditions on $\mathcal{F}$ and $P$ ensuring that LLN holds uniformly in $f \in \mathcal{F}$, that is, so that

$$\lim_{n\to\infty} ||P_n - P||^*_{\mathcal{F}} = 0 \ a.s.,$$

  where $P_n = \sum_{i=1}^{n} \delta_{X_i}/n \quad (1 \leq i \leq n, n \in \mathbb{N})$.
- Let $F$ be the P-measurable cover of the function $x \mapsto sup_{f\in\mathcal{F}}|f(x)|$. We call this function the measurable cover of $\mathcal{F}$.

## 3.7.2 Uniform Laws of Large Numbers (G-C Theorems)

Proposition 3.7.8 If $PF < \infty$ then the sequence $\{||P_n - P||^*_{\mathcal{F}}\}_{n=1}^{\infty}$ converges a.s. and in $L^1$ to a finite limit.

The limit in proposition3.7.8 may not be zero: if, for example, $P$ gives mass zero to all finite sets of $\mathbb{R}$ and $\mathcal{F}$ is the collection of indicators of all finite sets in $\mathbb{R}$, then $||P_n - P||_{\mathcal{F}} = ||(1/n)\sum_{i=n}^{n} \delta_{X_i}(\{X_1, \cdots, X_n\})||_{\mathcal{F}} = 1$

# 3.7.2 Uniform Laws of Large Numbers (G-C Theorems)

Corollary 3.7.9 If $PF < \infty$ and $\{||P_n - P||^*_{\mathcal{F}}\}$ converges in probability to zero, then it converges a.s. to zero.

In other words, under integrability of the measurable cover of the class $\mathcal{F}$, the weak law of large numbers uniform in $f \in \mathcal{F}$ implies the uniform strong law.

## 3.7.2 Uniform Laws of Large Numbers (G-C Theorems)

The following definition is given by analogy with the classical Glivenko-Cantelli theorem for the empirical distribution function in $\mathbb{R}$, which is just the law of large numbers for the empirical process over $\mathcal{F} = \{I_{(-\infty,x]} : x \in \mathbb{R}\}$.

Definition 3.7.10 [ P-Glivenko-Cantelli class ] A class of functions $\mathcal{F}$ is a P-Glivenko-Cantelli class if $||P_n - P||_{\mathcal{F}}^* \to 0$ a.s., where $P_n$ is the empirical process based on the coordinate projections $X_i, i = 1, \cdots, n, n \in \mathbb{N}$, of the product probability space $(S, \mathcal{S}, P)^{\mathbb{N}}$.

# 3.7.2 Uniform Laws of Large Numbers (G-C Theorems)

Definition 3.7.11 A class of functions $\mathcal{F}$ is P-measurable, or P-empirically measurable, or just measurable, if for each $\{a_i, b\} \subset \mathbb{R}$ and $n \in \mathbb{N}$ , the quantity $||\sum_{i=1}^n a_i f(X_i) + bP||_{\mathcal{F}}$ is measurable for the completion of $P^n$.

For example, if $\mathcal{F}$ is countable, then it is P-measurable for every $P$.
If $\mathcal{F}_0 \subset \mathcal{F}$ is P-measurable and for each$\{a_i, b\} \subset \mathbb{R}$ and $n \in \mathbb{N}$,

$$Pr^*\{||\sum_{i=1}^n a_i f(X_i) + bPf||_{\mathcal{F}} \neq ||\sum_{i=1}^n a_i f(X_i) + bPf||_{\mathcal{F}_0}\} = 0,$$

then $\mathcal{F}$ is P-measurable;

## 3.7.2 Uniform Laws of Large Numbers (G-C Theorems)

In the measurable case, the Glivenko-Cantelli property for $\mathcal{F}$ can be characterised by a condition on the metric entropies of $\mathcal{F}$ w.r.t the $L_p(P_n)$ pseudo-metrics, for any $0 < p \leq \infty$. These metric entropies are random, so the result does not constitute a complete solution to the problem, but it does simplify it.

- Definition of the empirical $L^p$ pseudo-distances :
  $e_{n,p}(f,g) = ||f - g||_{L^p(P_n)}$
  - $p = \infty$, $e_{n,\infty}(f,g) = max_{1 \leq i \leq n} |f(X_i) - g(X_i)|$
  - $0 < p < \infty$, $e_{n,p}(f,g) = [\sum_{i=1}^{n} |f(X_i) - g(X_i)|^p]^{1/(p \wedge 1)}$ .
- Covering numbers and packing numbers of $(T, d)$ : $N(T, d, \epsilon)$ , $D(T, d, \epsilon)$

Given a class of functions $\mathcal{F}$ and a positive number $M$, we set

$$\mathcal{F}_M = \{ f I_{F \leq M} : f \in \mathcal{F} \},$$

where $F$ is the P-measurable cover of $\mathcal{F}$ (determined only P a.s.).

## 3.7.2 Uniform Laws of Large Numbers (G-C Theorems)

Theorem 3.7.14 Let $\mathcal{F}$ be class of functions with an everywhere finite measurable cover $F$ and such that $\mathcal{F}_M$ is P-measurable for all $M \leq \infty$. Assume also that $\mathcal{F}$ is $L^1(P)$ -bounded, that is, $sup_{f \in \mathcal{F}} P|f| < \infty$ . Then the following are equivalent:

(a) $\mathcal{F}$ is a P-Glivenko-Cantelli class of functions.

(b) $PF < \infty$ and $||P_n - P||_{\mathcal{F}} \xrightarrow{\text{prob.}} 0$

(c) $PF < \infty$, and for all $M < \infty, \epsilon > 0$ and $p \in (0, \infty]$, $(logN^*(\mathcal{F}_M, e_{n,p}, \epsilon))/n \xrightarrow{\text{prob.}} 0$ (in $L^r$ for any $0 < r < \infty$).

(d) $PF < \infty$, and for all $M < \infty$ and $\epsilon > 0$ and for some $p \in (0, \infty]$, $(logN^*(\mathcal{F}_M, e_{n,p}, \epsilon))/n \xrightarrow{\text{prob.}} 0$ (in $L^r$ for any $0 < r < \infty$).

(e) $PF < \infty$, and for all $M < \infty$ and $\epsilon > 0$ ,

$$E(1 \wedge (1/\sqrt{n}) \int_0^{2M} \sqrt{logN^*(\mathcal{F}_M, e_{n,2}, \tau)} d\tau) \to 0$$

# 3.7.2 Uniform Laws of Large Numbers (G-C Theorems)

Corollary 3.7.15 Let $\mathcal{F}$ be an $L^1(P)$-bounded, P-measurable class of functions, and let $F$ be its P-measurable cover. Then $\mathcal{F}$ is P-Glivenko-Cantelli if and only if

(a) $PF < \infty$ , and

(b) $(1/n) log N^*(\mathcal{F}, e_{n,2}, \epsilon) \xrightarrow{\text{prob.}} 0$ (or in $L^{1/2}$ ) .

For classes of sets $\mathcal{C}$, recall the definition of $\Delta^{\mathcal{C}}(A)$ for finite sets $A$ in Section 3.6.1, $\Delta^{\mathcal{C}}(A) = Card\{A \cap C : C \in \mathcal{C}\}$ , and note that for $A(\omega) = \{X_1(\omega), \cdots, X_n(\omega)\}$, if $C \cap \{X_1, \cdots, X_n\} = D \cap \{X_1, \cdots, X_n\}$, then $e_{n,p}(C, D) = 0$ for all $0 < p \le \infty$ and that $e_{n,p}(C, D) \ge n^{-1/(p \vee 1)}$ otherwise. Hence, $N(\mathcal{C}, e_{n,p}, \epsilon) \le \Delta^{\mathcal{C}}(X_1, \cdots, X_n)$ for all $\epsilon > 0$ , with equality for $0 < \epsilon \le n^{-1/(p \vee 1)}$.

# 3.7.2 Uniform Laws of Large Numbers (G-C Theorems)

This observation and Theorem 3.7.14 for $p = \infty$ give the following result for classes of sets:

Corollary 3.7.16 Let $\mathcal{C}$ be an $L^1(P)$-bounded, P-measurable class of sets. Then $||P_n - P||_{\mathcal{C}}^* \to 0$ $a.s.$ if and only if

$$\lim_{n\to\infty} \frac{1}{n} log(\Delta^{\mathcal{C}}(X_1, \cdots, X_n))^* = 0 \text{ in prob.( or in } L^r \text{ for any } r < \infty)$$

# 3.7.2 Uniform Laws of Large Numbers (G-C Theorems)

Combining Corollary3.7.15 with Theorem3.6.9 about the empirical metric entropy properties of VC type classes of functions, we obtain the following uniform law of large numbers.

Corollary 3.7.17 Let $P$ be any probability measure on $(S, \mathcal{S})$, and let $\mathcal{F}$ be a P-measurable class of functions whose measurable cover $F$ is P-integrable. Assume that
(a) $\mathcal{F}$ is VC subgraph or, more generally, of VC type, or
(b) $\mathcal{F}$ is VC hull.

Then $\mathcal{F}$ is P-Glivenko-Cantelli.

## 3.7.2 Uniform Laws of Large Numbers (G-C Theorems)

Remark 3.7.18 Since, as mentioned in the preceding proof, $||P_n - P||_{\mathcal{F}} = ||P_n - P||_{\bar{co}\mathcal{F}}$ , it follows that the Glivenko-Cantelli property is preserved by taking pointwise closures of convex hulls; that is, $\mathcal{F}$ is P-Glivenko-Cantelli if and only if $\bar{co}\mathcal{F}$ is.

# 3.7.2 Uniform Laws of Large Numbers (G-C Theorems)

For larger classes, one may use the random entropies in Corollary 3.7.15 and Theorem 3.7.14;
The following criterion for the Glivenko-Cantelli property based on $L^1(P)$ bracketing is more user friendly when it applies:

Theorem 3.7.20 If $\mathcal{F} \subset L^1(S, \mathcal{S}, P)$ and $N_{[]}(\mathcal{F}, L^1(P), \epsilon) < \infty$ for all $\epsilon > 0$ , then $||P_n - P||^*_{\mathcal{F}} \to 0 a.s.$

## 3.7.2 Uniform Laws of Large Numbers (G-C Theorems)

It also implies LLN in separable Banach spaces. For a random variable $X$ in a Banach space $B$, the expectation $EX$ is defined in the Bochner sense.

Corollary 3.7.21 (Mourier law of large numbers) Let $B$ be a separable Banach space, and let $X, X_i$ be i.i.d. B-valued random vectors such that $E||X|| < \infty$ . Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to EX \ a.s.$$