Chapter 2.1: Introduction

Geerten Koers

Reading group *Weak Convergence and Empirical Processes*

2019-10-07

Outline

Empirical measures

Definition

The *empirical measure* is defined by

$$\mathbb{P}_n(C) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i \in C)$$

for $C \subset \mathcal{X}$.

Empirical measures

Definition

The *empirical measure* is defined by

$$\mathbb{P}_n(C) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i \in C)$$

for $C \subset \mathcal{X}$.

This induces the map

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

for any $f : \mathcal{X} \to \mathbb{R}$.

Empirical process

Definition

The *empirical process* $\mathbb{G}_n$ is defined by

$$f \mapsto \mathbb{G}_n f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f(X_i) - Pf)$$

with $Pf = \int f dP$.

Limiting theorems

For a fixed $f : \mathcal{X} \to \mathbb{R}$ **we have**

$$\begin{aligned} \mathbb{P}_n f &\to Pf, \\ \mathbb{G}_n &\rightsquigarrow N(0, P(f - Pf)^2) \end{aligned}$$

if $P|f| < \infty$ **and** $Pf^2 < \infty$**.**

Limiting theorems

For a fixed $f : \mathcal{X} \to \mathbb{R}$ we have

$$\begin{aligned} \mathbb{P}_n f &\to Pf, \\ \mathbb{G}_n &\rightsquigarrow N(0, P(f - Pf)^2) \end{aligned}$$

if $P|f| < \infty$ and $Pf^2 < \infty$.

For what $\mathcal{F} \subset \mathbb{R}^{\mathcal{X}}$ can we make these limits uniform?

Uniform convergence

Definition

For a measure Q on $\mathcal{X}$, define

$$\|Q\|_{\mathcal{F}} := \sup\{|Qf| : f \in \mathcal{F}\}.$$

Uniform convergence

Definition

For a measure Q on $\mathcal{X}$, define

$$\|Q\|_{\mathcal{F}} := \sup\{|Qf| : f \in \mathcal{F}\}.$$

Definition

If $\|\mathbb{P}_n - P\|_{\mathcal{F}} \to 0$ in outer probability, then $\mathcal{F}$ is called *Glivenko-Cantelli*.

Uniform convergence

Definition

For a measure Q on $\mathcal{X}$, define

$$\|Q\|_{\mathcal{F}} := \sup\{|Qf| : f \in \mathcal{F}\}.$$

Definition

If $\|\mathbb{P}_n - P\|_{\mathcal{F}} \to 0$ in outer probability, then $\mathcal{F}$ is called *Glivenko-Cantelli*.

$$\sup\left\{\frac{1}{n}|\sum_{i=1}^{n} f(X_i) - Pf(X_1)| : f \in \mathcal{F}\right\} \to 0.$$

Limiting theorems

Assume

$$\sup_{f \in \mathcal{F}} |f(x) - Pf| < \infty.$$

Now $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ **is a map into** $\ell^\infty(\mathcal{F})$**.**

Assume

$$\sup_{f \in \mathcal{F}} |f(x) - Pf| < \infty.$$

Now $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ **is a map into** $\ell^\infty(\mathcal{F})$**.**

Definition

If

$$\mathbb{G}_n \rightsquigarrow \mathbb{G}$$

for a tight Borel measurable element $\mathbb{G} \in \ell^\infty(\mathcal{F})$, then $\mathcal{F}$ is called a *P-Donsker class*.

Limiting theorems

Assume

$$\sup_{f \in \mathcal{F}} |f(x) - Pf| < \infty.$$

Now $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ **is a map into** $\ell^\infty(\mathcal{F})$**.**

Definition

If

$$\mathbb{G}_n \rightsquigarrow \mathbb{G}$$

for a tight Borel measurable element $\mathbb{G} \in \ell^\infty(\mathcal{F})$, then $\mathcal{F}$ is called a *P-Donsker class*.

$\mathbb{G}$ **is a centered Gaussian process with covariance**

$$\mathbb{E}\mathbb{G}f_1\mathbb{G}f_2 = Pf_1f_2 - Pf_1Pf_2.$$

Technicality

We define the X_i on the product space $(\mathcal{X}^\infty, \mathcal{B}^\infty, P^\infty)$. This is possible since the projection on $\mathcal{X}^n$ is perfect.

Entropy

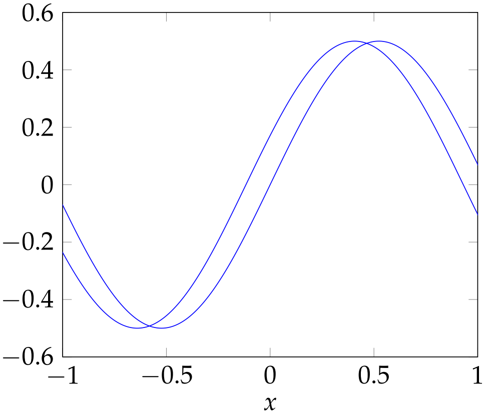
Definition (2.1.5)

The *covering number* $N(\epsilon, \mathcal{F}, \|\cdot\|)$ is the mimimum number of balls $B_\epsilon(f) := \{g \in \mathcal{F} : \|g - f\| < \epsilon\}$ such that $\mathcal{F}$ is covered. The *entropy without bracketing* is $\log N(\epsilon, \mathcal{F}, \|\cdot\|)$.

Entropy

Definition (2.1.5)

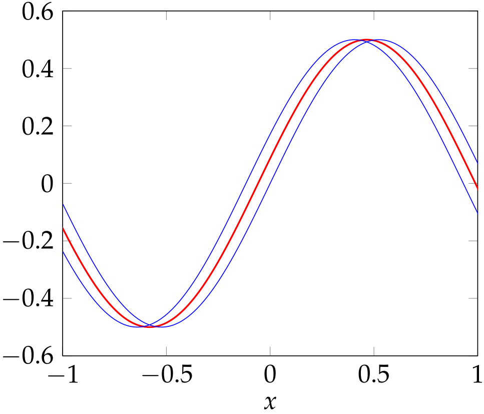
The *covering number* $N(\epsilon, \mathcal{F}, \|\cdot\|)$ is the mimimum number of balls $B_\epsilon(f) := \{g \in \mathcal{F} : \|g - f\| < \epsilon\}$ such that $\mathcal{F}$ is covered. The *entropy without bracketing* is $\log N(\epsilon, \mathcal{F}, \|\cdot\|)$.



Entropy

Definition (2.1.5)

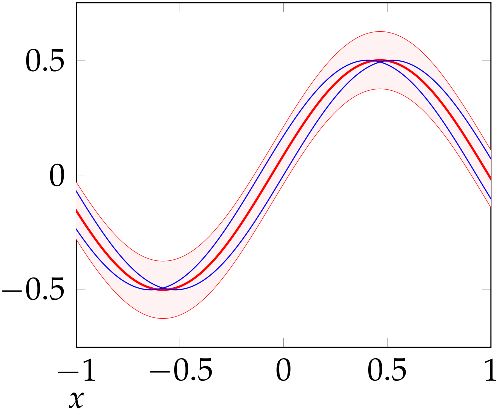
The *covering number* $N(\epsilon, \mathcal{F}, \|\cdot\|)$ is the mimimum number of balls $B_\epsilon(f) := \{g \in \mathcal{F} : \|g - f\| < \epsilon\}$ such that $\mathcal{F}$ is covered. The *entropy without bracketing* is $\log N(\epsilon, \mathcal{F}, \|\cdot\|)$.



Entropy

Definition (2.1.5)

The *covering number* $N(\epsilon, \mathcal{F}, \|\cdot\|)$ is the mimimum number of balls $B_\epsilon(f) := \{g \in \mathcal{F} : \|g - f\| < \epsilon\}$ such that $\mathcal{F}$ is covered.
The *entropy without bracketing* is $\log N(\epsilon, \mathcal{F}, \|\cdot\|)$.



Entropy

Definition (2.1.6)

The *bracketing number* $N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$ is the mimimum number of brackets $[l, u] := \{g \in \mathcal{F} : l \leq g \leq u\}$ with $\|u - l\| < \epsilon$ such that $\mathcal{F}$ is covered.
The *entropy with bracketing* is $\log N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$.

Entropy

Definition (2.1.6)

The *bracketing number* $N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$ is the mimimum number of brackets $[l, u] := \{g \in \mathcal{F} : l \leq g \leq u\}$ with $\|u - l\| < \epsilon$ such that $\mathcal{F}$ is covered.
The *entropy with bracketing* is $\log N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$.

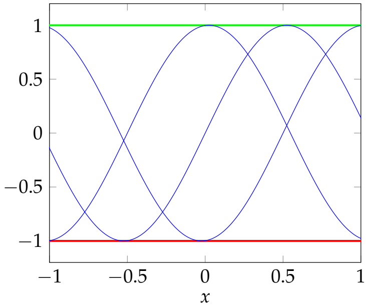


Figure: **Lower and upper bounds on sine functions.**

Relation between coverings

For norms such that $|f| \leq |g|$ implies $\|f\| \leq \|g\|$, we have

$$N(\epsilon, \mathcal{F}, \|\cdot\|) \leq N_{[]}(2\epsilon, \mathcal{F}, \|\cdot\|).$$

Relation between coverings

For norms such that $|f| \leq |g|$ implies $\|f\| \leq \|g\|$, we have

$$N(\epsilon, \mathcal{F}, \|\cdot\|) \leq N_{[]}(2\epsilon, \mathcal{F}, \|\cdot\|).$$

- **There is no general reverse inequality.**

Relation between coverings

For norms such that $|f| \leq |g|$ implies $\|f\| \leq \|g\|$, we have

$$N(\epsilon, \mathcal{F}, \|\cdot\|) \leq N_{[]}(2\epsilon, \mathcal{F}, \|\cdot\|).$$

- **There is no general reverse inequality.**
- **For the uniform norm, this is an equality.**

Entropy and envelopes

Definition

An *envelope function* for a class $\mathcal{F}$ is a function $F : \mathcal{X} \to \mathbb{R}$ with $|f(x)| \leq F(x)$ for all $x \in \mathcal{X}$, $f \in \mathcal{F}$.

Entropy and envelopes

Definition

An *envelope function* for a class $\mathcal{F}$ is a function $F : \mathcal{X} \to \mathbb{R}$ with $|f(x)| \leq F(x)$ for all $x \in \mathcal{X}$, $f \in \mathcal{F}$.

Definition

The *uniform entropy numbers* are defined as

$$\sup_Q \log N(\epsilon \|F\|_{Q,r}, \mathcal{F}, L_r(Q)),$$

with the supremum over all probability measures Q on $(\mathcal{X}, \mathcal{A})$ with $0 < QF^r < \infty$.

Theorem

$\mathcal{F}$ is a P-Glivenko-Cantelli class if

$$N_{[]}(\epsilon, \mathcal{F}, L_1(P)) < \infty, \qquad \textit{for all } \epsilon > 0.$$

This is a corollary of Theorem 2.4.3**.**

Theorem (2.4.3)

$\mathcal{F}$ is a P-Glivenko-Cantelli class if

$$\sup_Q N(\epsilon \|F\|_{Q,1}, \mathcal{F}, L_1(Q)) < \infty, \qquad \textit{for all } \epsilon > 0$$

with $P^* F < \infty$ and

$$(X_1, \ldots, X_n) \mapsto \left\| \sum_{i=1}^n e_i f(X_i) \right\|_{\mathcal{F}}$$

is measurable in the completion of $(\mathcal{X}^n, \mathcal{A}^n, P^n)$ for every vector $(e_1, \ldots, e_n) \in \mathbb{R}^n$.

Symmetrization

The measurability of

$$(X_1, \dots, X_n) \mapsto \left\| \sum_{i=1}^{n} e_i f(X_i) \right\|_{\mathcal{F}}$$

is needed for the bound

$$\mathbb{E}^* \Phi \left(\|\mathbb{P}_n - P\|_{\mathcal{F}} \right) \leq \mathbb{E}^* \Phi \left(2 \|\mathbb{P}_n^o\|_{\mathcal{F}} \right),$$

with $\Phi : \mathbb{R} \to \mathbb{R}$ nondecreasing, convex,

Symmetrization

The measurability of

$$(X_1, \dots, X_n) \mapsto \left\| \sum_{i=1}^{n} e_i f(X_i) \right\|_{\mathcal{F}}$$

is needed for the bound

$$\mathbb{E}^* \Phi \left(\| \mathbb{P}_n - P \|_{\mathcal{F}} \right) \leq \mathbb{E}^* \Phi \left(2 \| \mathbb{P}_n^o \|_{\mathcal{F}} \right),$$

with $\Phi : \mathbb{R} \to \mathbb{R}$ nondecreasing, convex, and

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

where $\mathbb{P}(\epsilon_1 = 1) = \mathbb{P}(\epsilon_1 = -1) = \frac{1}{2}$.

2.5: Donsker Theorems

Theorem

$\mathcal{F}$ is a P-Donsker class if

$$\int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(P))}\, d\epsilon < \infty.$$

2.5: Donsker Theorems

Theorem

$\mathcal{F}$ is a P-Donsker class if

$$\int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(P))}\, d\epsilon < \infty.$$

Theorem (2.5.2)

$\mathcal{F}$ is a P-Donsker class if

$$\int_0^\infty \sup_Q \sqrt{\log N(\epsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q))}\, d\epsilon < \infty$$

for all P such that $P^* F^2 < \infty$. The supremum is taken over all finitely discrete probability measures Q on $(\mathcal{X}, \mathcal{A})$ with $\int F^2 dQ > 0$.

2.6: VC-Classes

Definition

$\mathcal{C} \subset \mathcal{P}(\mathcal{X})$ *picks out* a subset A of $\{x_1, \ldots, x_n\} \subseteq \mathcal{X}$ if there is a $C \in \mathcal{C}$ such that $A = \{x_1, \ldots, x_n\} \cap C$.

Definition

$\mathcal{C} \subset \mathcal{P}(\mathcal{X})$ *picks out* a subset A of $\{x_1, \ldots, x_n\} \subseteq \mathcal{X}$ if there is a $C \in \mathcal{C}$ such that $A = \{x_1, \ldots, x_n\} \cap C$.

Definition

$\mathcal{C}$ *shatters* $\{x_1, \ldots, x_n\}$ if $\mathcal{C}$ picks out every $A \subseteq \{x_1, \ldots, x_n\}$.

2.6: VC-Classes

Definition

$\mathcal{C} \subset \mathcal{P}(\mathcal{X})$ *picks out* a subset A of $\{x_1, \ldots, x_n\} \subseteq \mathcal{X}$ if there is a $C \in \mathcal{C}$ such that $A = \{x_1, \ldots, x_n\} \cap C$.

Definition

$\mathcal{C}$ *shatters* $\{x_1, \ldots, x_n\}$ if $\mathcal{C}$ picks out every $A \subseteq \{x_1, \ldots, x_n\}$.

Definition

The *VC-index* $V(\mathcal{C})$ is the smallest n such that for all $B \subseteq \mathcal{X}$ with $|B| \geq n$, $\mathcal{C}$ does not shatter B.

2.6: VC-Classes

Definition

$\mathcal{C} \subset \mathcal{P}(\mathcal{X})$ *picks out* a subset A of $\{x_1, \ldots, x_n\} \subseteq \mathcal{X}$ if there is a $C \in \mathcal{C}$ such that $A = \{x_1, \ldots, x_n\} \cap C$.

Definition

$\mathcal{C}$ *shatters* $\{x_1, \ldots, x_n\}$ if $\mathcal{C}$ picks out every $A \subseteq \{x_1, \ldots, x_n\}$.

Definition

The *VC-index* $V(\mathcal{C})$ is the smallest n such that for all $B \subseteq \mathcal{X}$ with $|B| \geq n$, $\mathcal{C}$ does not shatter B.

Definition

A collection of measurable sets $\mathcal{C}$ is a *VC-class* if $V(C) < \infty$.

2.6: VC-Classes and Covering Numbers

A VC-class picks out $O(n^{V(\mathcal{C})-1})$ subsets of $\{x_1, \ldots, x_n\}$ if $n \geq V(\mathcal{C})$.

2.6: VC-Classes and Covering Numbers

A VC-class picks out $O(n^{V(\mathcal{C})-1})$ subsets of $\{x_1, \ldots, x_n\}$ if $n \geq V(\mathcal{C})$.

"Some thought shows that the number of subsets picked out (...) is closely related to the covering numbers (...)".

A VC-class picks out $O(n^{V(\mathcal{C})-1})$ subsets of $\{x_1, \ldots, x_n\}$ if $n \geq V(\mathcal{C})$.

"Some thought shows that the number of subsets picked out (...) is closely related to the covering numbers (...)".

Theorem (2.6.4)

There exists a universal K such that for all VC-classes $\mathcal{C}$ and measures Q,

$$N(\epsilon, \mathcal{C}, L_r(Q)) \leq KV(\mathcal{C})(4e)^{V(\mathcal{C})} \left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{C})-1)}$$

for all $r \geq 1$ and $0 < \epsilon < 1$.

2.6: VC-Classes and Covering Numbers

A VC-class picks out $O(n^{V(\mathcal{C})-1})$ **subsets of** $\{x_1, \ldots, x_n\}$ **if** $n \geq V(\mathcal{C})$.

"Some thought shows that the number of subsets picked out (...) is closely related to the covering numbers (...)".

Theorem (2.6.4)

There exists a universal K such that for all VC-classes $\mathcal{C}$ *and measures Q,*

$$N(\epsilon, \mathcal{C}, L_r(Q)) \leq KV(\mathcal{C})(4e)^{V(\mathcal{C})} \left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{C})-1)}$$

for all $r \geq 1$ *and* $0 < \epsilon < 1$.

This thought requires 2.5 **pages.**

Definition

A general $\mathcal{F}$ is called a *VC-subgraph class* if the collection of all subgraphs

$$\{(x,t) \in \mathcal{X} \times \mathbb{R} : t \leq f(x)\}$$

of $f \in \mathcal{F}$ is a VC-class in $\mathcal{X} \times \mathbb{R}$.

2.6: VC-classes of Functions

Definition

A general $\mathcal{F}$ is called a *VC-subgraph class* if the collection of all subgraphs

$$\{(x,t) \in \mathcal{X} \times \mathbb{R} : t \leq f(x)\}$$

of $f \in \mathcal{F}$ is a VC-class in $\mathcal{X} \times \mathbb{R}$.

If $\mathcal{F}$ is a VC-subgraph class, it is a Glivenko-Cantelli and Donsker class.

Classes of smooth functions have nice entropy properties.

$C_1^\alpha(\mathcal{X})$ **is (roughly) the set of functions on** $\mathcal{X} \subseteq \mathbb{R}^d$ **with uniformly bounded partial derivatives up to order** α**.**

2.7: Bracketing Numbers

Classes of smooth functions have nice entropy properties.

$C_1^\alpha(\mathcal{X})$ is (roughly) the set of functions on $\mathcal{X} \subseteq \mathbb{R}^d$ with uniformly bounded partial derivatives up to order α.

Theorem (2.7.1 and 2.7.2)

$$\log N(\epsilon, C_1^\alpha(\mathcal{X}), \|\cdot\|_\infty) \leq K\epsilon^{-d/\alpha},$$

and

$$\log N_{[]}(\epsilon, C_1^\alpha(\mathcal{X}), L_r(Q)) \leq K\epsilon^{-d/\alpha}.$$

2.7: Bracketing Numbers

Classes of smooth functions have nice entropy properties.

$C_1^\alpha(\mathcal{X})$ is (roughly) the set of functions on $\mathcal{X} \subseteq \mathbb{R}^d$ with uniformly bounded partial derivatives up to order α.

Theorem (2.7.1 and 2.7.2)

$$\log N(\epsilon, C_1^\alpha(\mathcal{X}), \|\cdot\|_\infty) \le K\epsilon^{-d/\alpha},$$

and

$$\log N_{[]}(\epsilon, C_1^\alpha(\mathcal{X}), L_r(Q)) \le K\epsilon^{-d/\alpha}.$$

Theorem (2.7.5)

The class $\mathcal{F}$ of monotone functions $f : \mathbb{R} \to [0,1]$ satisfies

$$\log N_{[]}(\epsilon, \mathcal{F}, L_r(Q)) \le K\epsilon^{-1}.$$

2.8: Uniformity in the Underlying Distribution

Some results can be stated uniformly in the distribution of the X_i's.

2.8: Uniformity in the Underlying Distribution

Some results can be stated uniformly in the distribution of the X_i's.

Theorem (2.8.1)

Let $\mathcal{F}$ be a P-measurable class of functions on a measurable space for every probability measure P in a class $\mathcal{P}$. Suppose that, for some measurable envelope function F,

$$\lim_{M\to\infty} \sup_{P\in\mathcal{P}} PF\{F > M\} = 0,$$

$$\sup_{Q\in\mathcal{Q}_n} \log N(\epsilon\|F\|_{Q,1}, \mathcal{F}, L_1(Q)) = o(n), \qquad \text{for every } \epsilon > 0,$$

where the supremum is taken over the set $\mathcal{Q}_n$ of all discrete probability measures with atoms of size integer multiples of $1/n$. Then $\mathcal{F}$ is Glivenko-Cantelli uniformly in $P \in \mathcal{P}$.

2.8: Uniformity in the Underlying Distribution

Theorem (2.8.2)

Let $\mathcal{F}$ be a class of measurable functions with envelope function F that is square integrable uniformly in $P \in \mathcal{P}$. Then the following statements are equivalent:

1. *$\mathcal{F}$ is Donsker and pre-Gaussian, both uniformly in $P \in \mathcal{P}$;*
2. *the sequence $\mathbb{G}_{n,P}$ is asymptotically ρ_P-equicontinuous uniformly in $P \in \mathcal{P}$ and $\sup_{P \in \mathcal{P}} N(\epsilon, \mathcal{F}, \rho_P) < \infty$ for every $\epsilon > 0$.*

2.9: Multiplier Central Limit Theorem

When are the following equivalent:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\delta_{X_i} - P) \rightsquigarrow \mathbb{G},$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i(\delta_{X_i} - P) \rightsquigarrow \mathbb{G},$$

with convergence in $\ell^\infty(\mathcal{F})$.

When are the following equivalent:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\delta_{X_i}-P)\rightsquigarrow \mathbb{G},$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i(\delta_{X_i}-P)\rightsquigarrow \mathbb{G},$$

with convergence in $\ell^\infty(\mathcal{F})$.

Theorem

If ξ_i are i.i.d. with zero mean and variance 1, and

$$\|\xi_1\|_{2,1}=\int_0^\infty\sqrt{P(|\xi_1|>t)}dt<\infty,$$

then it is equivalent to the $\mathcal{F}$ being Donsker.

If $\mathcal{F}$ and $\mathcal{G}$ are Donsker, then so are

1. **The convex hull of $\mathcal{F}$;**

If $\mathcal{F}$ and $\mathcal{G}$ are Donsker, then so are

1. **The convex hull of $\mathcal{F}$;**
2. $\{\phi(f_1, \ldots, f_k), f_i \in \mathcal{F}\}$ **if ϕ is a Lipschitz function on $\mathbb{R}^k$;**

2.10: Donsker Permanence

If $\mathcal{F}$ and $\mathcal{G}$ are Donsker, then so are

1. **The convex hull of $\mathcal{F}$;**
2. $\{\phi(f_1, \dots, f_k), f_i \in \mathcal{F}\}$ **if ϕ is a Lipschitz function on $\mathbb{R}^k$;**
3. $\mathcal{F} + \mathcal{G}$**,** $\mathcal{F} \wedge \mathcal{G}$**, and** $\mathcal{F} \cup \mathcal{G}$**;**

2.10: Donsker Permanence

If $\mathcal{F}$ and $\mathcal{G}$ are Donsker, then so are

1. **The convex hull of $\mathcal{F}$;**
2. $\{\phi(f_1, \ldots, f_k), f_i \in \mathcal{F}\}$ **if ϕ is a Lipschitz function on $\mathbb{R}^k$;**
3. $\mathcal{F} + \mathcal{G}$**,** $\mathcal{F} \wedge \mathcal{G}$**, and** $\mathcal{F} \cup \mathcal{G}$**;**
4. $\mathcal{F}\mathcal{G}$ **and** $\mathcal{F}/\mathcal{G}$ **if they are uniformly bounded from above, or uniformly bounded away from zero.**

2.11: Central limit theorem for Processes

We can extend the Donsker theorem to sums of independent, but not identically distributed processes $\sum_{i=1}^{n} Z_{n,i}$.

For $s \in [0,1]$, consider

$$\mathbb{Z}_n(s,f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} (f(X_i) - Pf).$$

For $s \in [0,1]$, consider

$$\mathbb{Z}_n(s,f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} (f(X_i) - Pf).$$

It converges in $\ell^\infty([0,1] \times \mathcal{F})$ to a tight limit if and only if $\mathcal{F}$ is Donsker.

For $s \in [0,1]$**, consider**

$$\mathbb{Z}_n(s,f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} (f(X_i) - Pf).$$

It converges in $\ell^\infty([0,1] \times \mathcal{F})$ **to a tight limit if and only if** $\mathcal{F}$ **is Donsker.**

It converges to the centered process $\mathbb{Z}$ **with**

$$\textbf{cov}(\mathbb{Z}(s,f), \mathbb{Z}(t,g)) = (s \wedge t)(Pfg - PfPg).$$

2.13: Other Donsker Classes

Theorem (2.13.1)

Any sequence $\{f_i\}$ of square-integrable, measurable functions with $\sum_{i=1}^{\infty} P(f_i - Pf_i)^2 < \infty$ is P-Donsker.

Theorem (2.13.1)

Any sequence $\{f_i\}$ of square-integrable, measurable functions with $\sum_{i=1}^{\infty} P(f_i - Pf_i)^2 < \infty$ is P-Donsker.

Theorem (2.13.2)

Let $\{f_i\}$ be a sequence of measurable functions such that $Pf_if_j = 0$ for every $i \neq j$ and $\sum_{i=1}^{\infty} Pf_i^2 < \infty$. Then the class of all pointwise converging series $\sum_{i=1}^{\infty} c_i f_i$, such that $\sum_{i=1}^{\infty} c_i^2 \leq 1$, is P-Donsker.

Asymptotic Equicontinuity

By Kolmogorov's Extension Theorem, there always exists a zero-mean Gaussian process $\{\mathbb{G}f : f \in \mathcal{F}\}$ with covariance function

$$\mathbb{E}\mathbb{G}f_1\mathbb{G}f_2 = Pf_1f_2 - Pf_1Pf_2.$$

Asymptotic Equicontinuity

By Kolmogorov's Extension Theorem, there always exists a zero-mean Gaussian process $\{\mathbb{G}f : f \in \mathcal{F}\}$ with covariance function

$$\mathbb{E}\mathbb{G}f_1\mathbb{G}f_2 = Pf_1f_2 - Pf_1Pf_2.$$

$\mathcal{F}$ is pre-Gaussian if $\mathbb{G}$ is a tight, Borel measurable map into $\ell^\infty(\mathcal{F})$.

Pre-Gaussianity is necessary for being Donsker, but not sufficient.

Asymptotic Equicontinuity

Pre-Gaussianity is necessary for being Donsker, but not sufficient. $\mathcal{F}$ is Donsker if and only if

$$\lim_{\delta \searrow 0} \limsup_{n \to \infty} \mathbb{P}^* \left(\sup_{\rho_P(f-g)<\delta} |\mathbb{G}_n(f-g)| > \epsilon \right) = 0,$$

and $\mathcal{F}$ is totally bounded in $\mathcal{L}_2(P)$.

Asymptotic Equicontinuity

Pre-Gaussianity is necessary for being Donsker, but not sufficient. $\mathcal{F}$ is Donsker if and only if

$$\lim_{\delta \searrow 0} \limsup_{n \to \infty} \mathbb{P}^* \left(\sup_{\rho_P(f-g)<\delta} |\mathbb{G}_n(f-g)| > \epsilon \right) = 0,$$

and $\mathcal{F}$ is totally bounded in $\mathcal{L}_2(P)$.

Here

$$\rho_P(f) = (P(f - Pf)^2)^{1/2}.$$

Exercises

2.1.1: **If $\mathcal{F}$ is totally bounded in $L_2(P)$, then it is totally bounded for the seminorm ρ_P. If $\mathcal{F}$ is totally bounded for ρ_P and $\|P\|_{\mathcal{F}} = \sup\{|Pf| : f \in \mathcal{F}\}$ is finite, then it is totally bounded in $L_2(P)$.**

Exercises

2.1.2**: Suppose** $\sup\{|Pf| : f \in \mathcal{F}\}$ **is finite. Then a class** $\mathcal{F}$ **of measurable functions is** P**-Donsker if and only if** $\mathcal{F}$ **is totally bounded in** $L_2(P)$ **and the empirical process is asymptotically equicontinuous in probability for the** $L_2(P)$**-semimetric.**

Exercises

2.1.5: **If** $a_n : [0,1] \to [0,\infty)$ **is a sequence of nondecreasing functions, then there exist** $\delta_n \searrow 0$ **such that** $\limsup a_n(\delta_n) = \lim_{\delta \to 0} \limsup_{n\to\infty} a_n(\delta)$**.**

Exercises

2.1.5: **If** $a_n : [0,1] \to [0,\infty)$ **is a sequence of nondecreasing functions, then there exist** $\delta_n \searrow 0$ **such that** $\limsup a_n(\delta_n) = \lim_{\delta\to 0} \limsup_{n\to\infty} a_n(\delta)$**.**

Deduce that $\lim_{\delta\to 0} \limsup_{n\to\infty} a_n(\delta) = 0$ **if and only if** $a_n(\delta_n) \to 0$ **for every** $\delta_n \to 0$**.**

Chapter 2.1: Introduction

Geerten Koers

Reading group *Weak Convergence and Empirical Processes*

2019-10-07

Chapter 2.2: Maximal Inequalities

Geerten Koers

Reading group ***Weak Convergence and Empirical Processes***

2019-10-07

Orlicz norm

Definition

The *Orlicz norm* is defined as

$$\|X\|_\psi = \inf \left\{ C > 0 : \mathbb{E}\psi\left(\frac{|X|}{C}\right) \leq 1 \right\},$$

with ψ non-decreasing, convex and $\psi(0) = 0$.

Orlicz norm

Definition

The *Orlicz norm* is defined as

$$\|X\|_\psi = \inf\left\{C > 0 : \mathbb{E}\psi\left(\frac{|X|}{C}\right) \leq 1\right\},$$

with ψ non-decreasing, convex and $\psi(0) = 0$.

1. **If $\psi(x) = x^p$, then $\|X\|_\psi = \|X\|_p = (E|X|^p)^{1/p}$, $p \geq 1$.**

Orlicz norm

Definition

The *Orlicz norm* is defined as

$$\|X\|_\psi = \inf\left\{C > 0 : \mathbb{E}\psi\left(\frac{|X|}{C}\right) \leq 1\right\},$$

with ψ non-decreasing, convex and $\psi(0) = 0$.

1. **If** $\psi(x) = x^p$**, then** $\|X\|_\psi = \|X\|_p = (E|X|^p)^{1/p}$**,** $p \geq 1$**.**
2. $\psi_p(x) = e^{x^p}$.

Orlicz norm

$$\|X\|_\psi = \inf\left\{C > 0 : \mathbb{E}\psi\left(\frac{|X|}{C}\right) \leq 1\right\}$$

is indeed a norm:

1. $\|X\|_\psi \geq 0$**;**

Orlicz norm

$$\|X\|_\psi = \inf\left\{C > 0 : \mathbb{E}\psi\left(\frac{|X|}{C}\right) \le 1\right\}$$

is indeed a norm:

1. $\|X\|_\psi \ge 0$;
2. $\|\lambda X\|_\psi = |\lambda| \, \|X\|_\psi$;

Orlicz norm

$$\|X\|_\psi = \inf\left\{C > 0 : \mathbb{E}\psi\left(\frac{|X|}{C}\right) \leq 1\right\}$$

is indeed a norm:

1. $\|X\|_\psi \geq 0$;
2. $\|\lambda X\|_\psi = |\lambda|\ \|X\|_\psi$;
3. $\|X\|_\psi = 0$ **if and only if** $X = 0$ **a.s.;**

Orlicz norm

$$\|X\|_\psi = \inf\left\{C > 0 : \mathbb{E}\psi\left(\frac{|X|}{C}\right) \leq 1\right\}$$

is indeed a norm:

1. $\|X\|_\psi \geq 0$;
2. $\|\lambda X\|_\psi = |\lambda| \ \|X\|_\psi$;
3. $\|X\|_\psi = 0$ **if and only if** $X = 0$ **a.s.;**
4. $\|X + Y\|_\psi \leq \|X\|_\psi + \|Y\|_\psi$.

Orlicz norm

Exponential norms give better bounds than L_p-norms:

- $\|X\|_{\psi_p} \leq \|X\|_{\psi_q}(\log 2)^{p/q}, \qquad p \leq q,$

Orlicz norm

Exponential norms give better bounds than L_p-norms:

- $\|X\|_{\psi_p} \leq \|X\|_{\psi_q} (\log 2)^{p/q}, \qquad p \leq q,$
- $\|X\|_p \leq p! \|X\|_{\psi_1}, \qquad p \geq 1.$

Bound on norm

Lemma (2.2.1)

If for all $x \geq 0$, $\mathbb{P}(|X| > x) \leq Ke^{-Cx^p}$ *for some constants* K *and* C*, then*

$$\|X\|_{\psi_p} \leq ((1+K)/C)^{1/p}.$$

Bound on norm

Lemma (2.2.1)

If for all $x \geq 0$, $\mathbb{P}(|X| > x) \leq Ke^{-Cx^p}$ for some constants K and C, then

$$\|X\|_{\psi_p} \leq ((1+K)/C)^{1/p}.$$

This follows from

$$\mathbb{E}(e^{D|X|^p} - 1) = \mathbb{E}\int_0^{|X|^p} De^{Ds}ds = \int_0^{\infty} \mathbb{P}(|X| > s^{1/p})De^{Ds}ds.$$

Orlicz norm

Theorem (2.2.2)

Assume ψ is a convex, nondecreasing, nonzero function, with $\psi(0) = 0$ and such that $\limsup_{x,y\to\infty} \psi(x)\psi(y)/\psi(cxy) < \infty$ for some $c \in \mathbb{R}$. Then we have

$$\left\| \max_{1\leq i\leq m} X_i \right\|_\psi \leq K\psi^{-1}(m) \max_{1\leq i\leq m} \|X_i\|_\psi,$$

for K depending only on ψ.

Chaining

Theorem (2.2.4)

Assume that ψ satisfies the previous conditions. Then for any $\eta, \delta > 0$, we have

$$\left\| \sup_{d(s,t) \leq \delta} |X_s - X_t| \right\|_\psi \leq K \left[\int_0^\eta \psi^{-1}(N(\epsilon, d)) d\epsilon + \delta \psi^{-1}(N^2(\eta, d)) \right],$$

with $d(s,t) = \|X_s - X_t\|_\psi$.

Chaining

Definition

If the ψ_2-norms of $X_s - X_t$ are finite, then $\{X_t : t \in T\}$ is *sub-Gaussian*.

Chaining

Definition

If the ψ_2-norms of $X_s - X_t$ are finite, then $\{X_t : t \in T\}$ is *sub-Gaussian*.

This happens if and only if

$$\mathbb{P}(|X_s - X_t| > x) \leq 2e^{-\frac{1}{2}x^2/d^2(s,t)}, \qquad x > 0.$$

Chaining

Definition

If the ψ_2-norms of $X_s - X_t$ are finite, then $\{X_t : t \in T\}$ is *sub-Gaussian*.

This happens if and only if

$$\mathbb{P}(|X_s - X_t| > x) \leq 2e^{-\frac{1}{2}x^2/d^2(s,t)}, \qquad x > 0.$$

Now the equality in the theorem can be simplified to

$$\mathbb{E} \sup_{d(s,t)\leq\delta} |X_s - X_t| \leq K \int_0^\delta \sqrt{\log N(\epsilon, T, d)}\, d\epsilon.$$