# Section 2.4 and 2.5

# Weak Convergence and Empirical Processes

Lasse Vuursteen

# Today's program

We will discuss chapter's 2.4 and 2.5, in order. These chapters are respectively concerned with:

- What condition on $\mathscr{F}$ guarantees that it is $P$-Glivenko-Cantelli?
- What condition on $\mathscr{F}$ guarantees that it is $P$-Donsker?

# Glivenko-Cantelli Classes

A function class $\mathscr{F}$ is called *P-Glivenko-Cantelli* if

$$||\mathbb{P}_n - P||_{\mathscr{F}} \to 0 \text{ in outer probability}$$

where $||\cdot||_{\mathscr{F}}$ denotes the uniform norm.

# Conditions on $\mathscr{F}$

Showing $||\mathbb{P}_n - P||_{\mathscr{F}} \to 0$ is generally not straightforward.

- $\sup$ over $\mathscr{F}$ of infinite cardinality are generally nasty.
- Measurability issues.

A way to deal with these issues is through entropy conditions on $\mathscr{F}$: "$\sup$ smells like entropy"

Two flavours of entropy conditions:

- Bracketing entropy
- Metric entropy

# Bracketing entropy based Glivenko-Cantelli

## Theorem

*Let $\mathscr{F}$ be a class of measurable functions such that $N_{[]}(\varepsilon, \mathscr{F}, L_1(P)) < \infty$ for every $\varepsilon > 0$. Then, $\mathscr{F}$ is Glivenko-Cantelli.*

**Proof:** Fix $\varepsilon > 0$, choose a finite amount of brackets $[l_i, u_i]$ such that $P(u_i - l_i) < \varepsilon$.

Every $f \in \mathscr{F}$ lives in one of these brackets so

$$(\mathbb{P}_n - P)f \leq (\mathbb{P}_n - P)u_i + P(u_i - f) \leq (\mathbb{P}_n - P)u_i + \varepsilon.$$

This means

$$\sup_{f \in \mathscr{F}} (\mathbb{P}_n - P)f \leq \max_i (\mathbb{P}_n - P)u_i + \varepsilon$$

which is measurable. A similar argument can be made for the $\inf$. We conclude that $\limsup ||\mathbb{P}_n - P||^*_{\mathscr{F}} \leq \varepsilon$.

# (Random semi) metric entropy based Glivenko-Cantelli

## Theorem

*Let $\mathscr{F}$ be a $P$-measurable class with envelope $F$ such that $P^*F < \infty$. Let $\mathscr{F}_M$ be the class of functions $f1\{F \leq M\}$ when $f \in \mathscr{F}$.*

*Then, $||\mathbb{P}_n - P||^*_{\mathscr{F}} \to 0$ almost surely if and only if $n^{-1} \log N(\varepsilon, \mathscr{F}_M, L_1(\mathbb{P}_n)) \to 0$ in outer probability.*

*In the above case, the convergence is also in outer mean.*

# Proof: $\Longleftarrow$

Assume $n^{-1} \log N(\varepsilon, \mathscr{F}_M, L_1(\mathbb{P}_n)) \to 0$ in outer probability.

Recall the Symmetrization Lemma (2.3.1): For every nondecreasing,

convex $\Phi : \mathbb{R} \to \mathbb{R}$ and measurable $\mathscr{F}$,

$$E^* \Phi(||\mathbb{P}_n - P||_{\mathscr{F}}) \leq E^* \Phi\left(2||n^{-1} \sum_{i=1}^{n} \varepsilon_i f(X_i)||_{\mathscr{F}}\right).$$

# Proof: $\Longleftarrow$

This yields

$$E^*||\mathbb{P}_n - P||_{\mathscr{F}} \leq 2E_X E_\varepsilon||\frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i)||_{\mathscr{F}}.$$

which we can bound further using our envelope:

$$2E_X E_\varepsilon||\frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i)||_{\mathscr{F}_M} + 2P^* F\{F > M\}.$$

The first term can be dealt with using a maximal inequality.

# Maximal inequality (Lemma 2.2.2)

Let $\psi : \mathbb{R} \to \mathbb{R}$ convex, nondecreasing and nonzero except for in the point 0. Recall that *the Orlicz-norm* for $\psi$ is defined as

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

Suppose $\limsup_{x,y\to\infty} \psi(x)\psi(x)/\psi(cxy) < \infty$ for some $c \in \mathbb{R}$. Then,

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

for $K > 0$ depending only on $\psi$.

# Proof: $\Longleftarrow$

To apply this, note that we can choose a net $\mathscr{G}$ of cardinality $\log N(\varepsilon, \mathscr{F}_M, L_1(\mathbb{P}_n))$ such that

$$2E_X E_\varepsilon || \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f(X_i) ||_{\mathscr{F}_M} \leq \sqrt{1 + \log |\mathscr{G}|} \sup_{f \in \mathscr{G}} || \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f(X_i) ||_{\psi_2}$$

where we note that the $L_1$-norm on the right is bounded by the $\psi_2(x) := e^{x^2} - 1$ Orlicz-norm.

# Proof: $\Longleftarrow$

We have the Hoeffding increment bound for this choice of Orlicz norm from Chapter 2.2:

$$||X_s - X_t||_{\psi_2} \leq \sqrt{6} d(s,t).$$

We can apply this for the $L_2(\mathbb{P}_n)$ distance. By the envelope condition, $\mathbb{P}_n f^2 \leq M^2$ for all $f \in \mathscr{F}$. This gives us

$$\begin{aligned} E^* ||\mathbb{P}_n - P||_{\mathscr{F}} &\leq 2 E_X E_\varepsilon || \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) ||_{\mathscr{F}_M} + \varepsilon \\ &\leq E_X \sqrt{1 + \log N(\varepsilon, \mathscr{F}_M, L_1(\mathbb{P}_n))} \frac{1}{n} \sqrt{6} M + \varepsilon. \end{aligned}$$

Since $\varepsilon$ can be chosen arbitrarily small and $n^{-1} \log N(\varepsilon, \mathscr{F}_M, L_1(\mathbb{P}_n)) \to 0$ in outer probability.

# Proof: $\Longrightarrow$

For the other direction, assume $\|\mathbb{P}_n - P\|_{\mathscr{F}}^* \to 0$ almost surely.

Desymmetrization (Lemma 2.3.6) then implies

$$\frac{1}{2} E^* \left\| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i (f(X_i) - Pf) \right\|_{\mathscr{F}} \leq E^* \|\mathbb{P}_n - P\|_{\mathscr{F}}^* \to 0.$$

We can take $\varepsilon_i = Z_i \sim \mathscr{N}(0,1)$.

The link with entropy is provided by Sudakov's inequality.

# Sudakov's inequality

For a centered, separable Gaussian process $X$ indexed by $T$,

$$\sup_{\varepsilon>0}\varepsilon\sqrt{\log N(\varepsilon,T,\rho)}\leq\sqrt{2\pi\log 2}E\sup_{t\in T}X_t$$

where $\rho(s,t)=\sigma(X_s-X_t)$.

The symmetrized process $\{n^{-1}\sum_{i=1}^n Z_if(X_i):f\in\mathscr{F}\}$ is centered and Gaussian and $\rho$ here is the $L_2(\mathbb{P}_n)$ norm!

# Proof: $\Longrightarrow$

We obtain that $n^{-1/2}\sqrt{\log N(\varepsilon, \mathscr{F}, L_2(\mathbb{P}_n))} \to 0$ in outer expectation.

Note also that $N(\varepsilon, \mathscr{F}_M, L_2(\mathbb{P}_n)) \leq N(\varepsilon, \mathscr{F}_M, L_\infty(\mathbb{P}_n)) \leq (\frac{2M}{\varepsilon})^n$.

We finalize our proof:

$$\frac{1}{n}\log N(\varepsilon, \mathscr{F}_M, L_2(\mathbb{P}_n)) \leq \frac{1}{n}\sqrt{\log[(\frac{2M}{\varepsilon})^n]}\sqrt{\log N(\varepsilon, \mathscr{F}, L_2(\mathbb{P}_n))} \to 0$$

in outer expectation.

# Addendum to the theorem

Recall the last claim of the theorem:

**Theorem**

*Let $\mathscr{F}$ be a $P$-measurable class with envelope $F$ such that $P^* F < \infty$. Let $\mathscr{F}_M$ be the class of functions $f 1\{F \leq M\}$ when $f \in \mathscr{F}$.*

*Then, $||\mathbb{P}_n - P||^*_{\mathscr{F}} \to 0$ almost surely if and only if $n^{-1} \log N(\varepsilon, \mathscr{F}_M, L_1(\mathbb{P}_n)) \to 0$ in outer probability.*

***In the above case, the convergence is also in outer mean.***

This shown via a martingale argument.

# Hoffmann-Jorgensen inequality for moments (colloquial version)

Let $X_1, \ldots, X_n$ independent stochastic processes indexed by $T$. Then there exist $C > 0$ and $0 < v < 1$ such that

$$\mathbb{E}^* \max_{k \leq n} || \sum_{i=1}^{k} X_k ||_T \leq C \left( \mathbb{E}^* \max_{k \leq n} ||X_k||_T + F^{-1}(v) \right)$$

where $F^{-1}$ denotes the quantile function of $\max_{k \leq n} || \sum_{i=1}^{k} X_k ||_T$.

# Multiplier inequality (Lemma 2.9.1, colloquial version)

Let $Z_1, \ldots, Z_n$ iid standard normals and let $X_1, \ldots, X_n$ iid stochastic processes indexed by $T$, jointly independent.

$$\mathbb{E}^* \max_{k \leq n} \|\sum_{i=1}^{k} X_k\|_T \leq C \left( \mathbb{E}^* \max_{k \leq n} \|X_k\|_T + F^{-1}(v) \right)$$

where $F^{-1}$ denotes the quantile function of $\max_{k \leq n} \|\sum_{i=1}^{k} X_k\|_T$.

# Desymmetrization Lemma (2.3.1)

# Bracketing numbers of $\mathscr{F}$

A function class $\mathscr{F}$ is called *P-Donsker* if

$$\mathbb{G}_n := n^{-1/2}(\mathbb{P}_n - P) \rightsquigarrow \ell^\infty(\mathscr{F}).$$

# Weak convergence in $\ell^\infty(\mathscr{F})$

Recall: $X_\alpha$ converges weakly to a tight limit taking values in $\ell^\infty(\mathscr{F})$ if and only if $X_\alpha$ is asymptotically tight and its maringals converge weakly.
By the CLT, we already have marginal convergence, so asymptotic tightness of $\mathbb{G}_n$ is what we are after.

# Asymptotic tightness

Recall that a net $X_\alpha$ is asymptotically tight if for every $\varepsilon > 0$ there exists a compact set $K$ such that

$$\liminf P_*(X_\alpha \in K^\delta) \geq 1 - \varepsilon \quad \text{for every } \delta > 0.$$

This is hard to show directly for $\mathbb{G}_n$, but we can use a characterization: *asymptotic equicontinuity*.

# Asymptotic equicontinuity

A net $X_\alpha : \Omega \to \ell^\infty(T)$ is *asymptotically uniformly $\rho$-equicontinuous in probability* if for every $\varepsilon, \eta > 0$ there exists a $\delta > 0$ such that

$$\limsup_\alpha P^* \left( \sup_{\rho(s,t)<\delta} |X_\alpha(s) - X_\alpha(t)| > \varepsilon \right) < \eta .$$

**Theorem 1.5.7:** $X_\alpha$ is asymptotically tight if and only if $X_\alpha(t)$ is asympotically tight in $\mathbb{R}$ for every $t$ and there exists a semimetric $\rho$ on $T$ such that $(T, \rho)$ is totally bounded and $X_\alpha$ is asymptotically uniformly $\rho$-equicontinuous in probability.

The Donsker theorems come in two flavours (two types of conditions on $\mathscr{F}$):

- Based on the uniform entropy condition

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

- Based on bracketing entropy

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

These conditions are generally not comparable. Examples of function classes satisfying either or both are given in Chapter 2.7.

# Donsker Theorem's based on bracketing

**Theorem**

*Any class $\mathscr{F}$ of measurable functions satisfying*

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

*is $P$-Donsker.*

This result is a consequence of the following more general theorem.

# First, let's define a norm

Define the $L_{2,\infty}$-"norm" on $\mathscr{F}$ as

$$||f||_{P,2,\infty} = \sup_{x>0} \left( x^2 P(|f| > x) \right)^{1/2}.$$

This norm is weaker than the the $L_2(P)$ norm.

# Donsker Theorem's based on bracketing

## Theorem

*Any class $\mathscr{F}$ of measurable functions satisfying*

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathscr{F}, L_{2,\infty}(P))} d\varepsilon + \int_0^\infty \sqrt{\log N(\varepsilon, \mathscr{F}, L_2(P))} d\varepsilon < \infty$$

*is $P$-Donsker.*

This is more general than the previous result (why?).

# Sketch of the proof

The proof relies on applying the following result (Theorem 1.5.6) for proving asymptotic tightness:

A net $X_\alpha$ taking values in $\ell^\infty(T)$ is asymptotically tight if and only if $X_\alpha(t)$ is asymptotically tight in $\mathbb{R}$ for very $t \in T$ and if there exist a finite partition $T = \dot{\cup} T_i$ such that asymptotic equiconintuity holds uniformly over these partitions in the following sense:

$$\limsup P^* \left( \sup_i \sup_{s,t \in T_i} |X_\alpha(s) - X_\alpha(t)| > \varepsilon \right) < \eta.$$

The statement "$X_\alpha(t)$ is asymptotically tight in $\mathbb{R}$ for very $t \in T$" holds already in our situation..

# Sketch of the proof

The "joint entropy" condition

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathscr{F}, L_{2,\infty}(P))}d\varepsilon + \int_0^\infty \sqrt{\log N(\varepsilon, \mathscr{F}, L_2(P))}d\varepsilon < \infty$$

is used to construct the partition for which the second condition of 1.5.6 is satisfied. That is, a partition $\dot{\bigcup}\mathscr{F}_i = \mathscr{F}$ for which we can show $E^* \max_i ||\mathbb{G}_n||_{\mathscr{F}_i} \to 0$, which yields asymptotic equicontinuity through

$$P^*(\max_i ||\mathbb{G}_n||_{\mathscr{F}_i} > x) \leq \frac{1}{x} E^* \max_i ||\mathbb{G}_n||_{\mathscr{F}_i}.$$

# Sketch of the proof

In order to show $E^* \max_i ||\mathbb{G}_n||_{\mathscr{F}_i} \to 0$, we will aim to use the maximal inequality of Lemma 2.2.10:

Let $X_1, \ldots, X_n$ satisfy the tailbound

$$P(|X_i| > x) \leq 2e^{-\frac{1}{2}\frac{x^2}{b+ax}}$$

for all $x, i$ and fixed $a, b > 0$. Then,

$$||\max_i X_i||_{\psi_1} \leq K\left(a\log(1+m) + \sqrt{b}\sqrt{\log(1+m)}\right).$$

# Sketch of the proof

As a consequence of Bernstein's inequality, the condition of Lemma 2.2.10 is satisfied:

$$P(|\mathbb{G}_n f| > x) \leq 2e^{-\frac{1}{2}\frac{x^2}{Pf^2 + 2/3||f||_\infty x/\sqrt{n}}},$$

which holds for square integrable, uniformly bounded $f \in \mathscr{F}$.

# Sketch of the proof

So, if we partition $\mathscr{F} = \cup \mathscr{F}_i$ such that we can consider $||\mathbb{G}_n||_{\mathscr{F}_i}$ as (bounded by something) that runs over a finite amount of functions (say $m$ functions) we obtain

$$E^* ||\mathbb{G}_n||_{\mathscr{F}_i} \lesssim \max_{f_1,\dots,f_m} \frac{||f||_\infty}{\sqrt{n}} \log(1+m) + \max_{f_1,\dots,f_m} ||f||_{P,2} \sqrt{\log(1+m)}.$$

# Sketch of the proof

The partitions follow from the “joint” entropy condition and

$$\sqrt{\log N_{[]}(\varepsilon, \mathscr{F}, L_{2,\infty}(P))} \leq \sqrt{\log N(\varepsilon, \mathscr{F}, L_2(P))}$$

(note that the latter norm is stronger).

For a natural number $q \in \mathbb{N}$ we obtain $N_q^1$-$L_2(P)$ balls of radius $2^{-q}$ and similarly $N_q^2$-$L_{\infty,2}(P)$ balls of radius $2^{-q}$ such that

$$\underset{q \in \mathbb{N}}{2}^{-q} \sqrt{logN_q} < \infty$$

for $N_q = N_q^1 N_q^2$.

The intersection and disjointification of our balls form the partitions $\mathscr{F} = \cup \mathscr{F}_i^q$.

A (lengthy) chaining argument based on these partitions one can show that for all big enough $q_0$,

$$E^* \max_i ||\mathbb{G}_n||_{\mathscr{F}_i^q} \lesssim \sum_{\infty}^{q_0} [2^{-q} \sqrt{\log N_q}$$

which is finite by construction and hence goes to 0 as $q_0 \to \infty$, finishing the proof.

Thanks for listening!