Chapter 2.3 - Symmetrization and Measurability

Amine Hadji

University of Leiden

November 18, 2019

Overview

- Chapter 2.2 (Recap and Applications)
- Chapter 2.3 Symmetrization

Orlicz norm

Definition

Let ψ be a nonzero, nondecreasing, convex function with $\psi(0) = 0$ and X a random variable. Then the *Orlicz norm* $\|X\|_\psi$ is defined as

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

Orlicz norm

Definition

Let ψ be a nonzero, nondecreasing, convex function with $\psi(0) = 0$ and X a random variable. Then the *Orlicz norm* $\|X\|_\psi$ is defined as

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

Examples

- Let $p \geq 1$, let $\psi(x) = x^p$. Then for any random variable X $\|X\|_\psi = \|X\|_p$.
- Let $p \geq 1$, let $\psi_p(x) = e^{x^p} - 1$

Covering numbers

Definition

Let (T, d) an arbitrary semi-metric space. Then, the covering number $N(\epsilon, d)$ is the minimal number of balls of radius ϵ needed to cover T. The packing number $D(\epsilon, d)$ is the maximum number of ϵ-separated points in T (i.e. collection of points such that the distance between each pair of points is strictly larger than ϵ).

Through the inequalities

$$N(\epsilon, d) \leq D(\epsilon, d) \leq N(\epsilon/2, d),$$

covering number and packing number are generally used interchangeably.

Maximal Inequalities

Theorem

Let ψ a nonzero, nondecreasing, convex function with $\psi(0) = 0$ and $\limsup_{x,y\to+\infty} \psi(x)\psi(y)/\psi(cxy) < +\infty$ for some constant c. Let $\{X_t : t \in T\}$ be a separable stochastic process on with

$$\|X_s - X_t\|_\psi \leq Cd(s,t), \quad \textit{for every } s,t,$$

for some semimetric d and a constant C. Then, for any $\eta, \delta > 0$,

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

for a constant K depending on ψ and C only.

Maximal Inequalities

Corollary

The constant K can be chosen such that

$$\| \sup_{d(s,t)\leq\delta} |X_s - X_t| \|_\psi \leq K \int_0^{diam T} \psi^{-1}(D(\epsilon, d))d\epsilon.$$

Maximal Inequalities

Corollary

The constant K can be chosen such that

$$\| \sup_{d(s,t)\leq\delta} |X_s - X_t| \|_\psi \leq K \int_0^{diam T} \psi^{-1}(D(\epsilon, d)) d\epsilon.$$

General Bounds:

$$\| \sup_t |X_t| \|_\psi \leq \|X_{t_0}\|_\psi + K \int_0^{diam T} \psi^{-1}(D(\epsilon, d)) d\epsilon$$

Adaptation - Frequentist Bayes

Definition

Given a collection of possible models, we want to find a single procedure which: (1) works well for all models and (2) is specifically targetted to the correct model.
Here, "correct" means "which contains the true distribution of the data".

Adaptation - Frequentist Bayes

Definition

Given a collection of possible models, we want to find a single procedure which: (1) works well for all models and (2) is specifically targetted to the correct model.
Here, "correct" means "which contains the true distribution of the data".

Data	$X_1, ..., X_n$ i.i.d p_0
Models	$\mathcal{P}_{n,\alpha}$ for $\alpha \in A$
Prior	$\Pi_{n,\alpha}$ on $\mathcal{P}_{n,\alpha}$

where p_0 is contained in (or close to) $\mathcal{P}_{n,\beta}$ for some $\beta \in A$

Adaptation to smoothness

Let us consider the Gaussian white noise model

$$Y(t) = \int_0^t f_0(s)ds + \frac{1}{\sqrt{n}} W_t, t \in [0, 1].$$

Assume the true function f_0 belongs to a hyper-rectangle regularity class, i.e.

$$f_0 \in \Theta^\beta(M) = \{f = \sum_i f_i \psi_i \in L^2[0, 1] : \sup_i f_i^2 i^{2\beta+1} \leq M\}$$

Adaptation to smoothness

Let us consider the Gaussian white noise model

$$Y(t) = \int_0^t f_0(s)ds + \frac{1}{\sqrt{n}} W_t, t \in [0, 1].$$

Assume the true function f_0 belongs to a hyper-rectangle regularity class, i.e.

$$f_0 \in \Theta^\beta(M) = \{f = \sum_i f_i \psi_i \in L^2[0, 1] : \sup_i f_i^2 i^{2\beta+1} \leq M\}$$

We put a prior $f|a \sim \bigotimes_i \mathcal{N}(0, a^{-1} e^{-i/a})$, such that the posterior is

Adaptation to smoothness

Let us consider the Gaussian white noise model

$$Y(t) = \int_0^t f_0(s)ds + \frac{1}{\sqrt{n}} W_t, t \in [0, 1].$$

Assume the true function f_0 belongs to a hyper-rectangle regularity class, i.e.

$$f_0 \in \Theta^\beta(M) = \{f = \sum_i f_i \psi_i \in L^2[0, 1] : \sup_i f_i^2 i^{2\beta+1} \leq M\}$$

We put a prior $f|a \sim \bigotimes_i \mathcal{N}(0, a^{-1} e^{-i/a})$, such that the posterior is

$$f|a, Y \sim \bigotimes_i \mathcal{N}\left(\frac{nY_i}{ae^{i/a} + n}, \frac{1}{ae^{i/a} + n}\right)$$

Adaptation to smoothness

Appendix E. Proof of Theorem A.1

First note that the derivative of the marginal likelihood function $\ell_n(a)$ is

$$\text{(E.1)} \qquad \mathbb{M}_n(a) = \frac{1}{2}\Big(\sum_{i=1}^{\infty} \frac{n^2 Y_i^2 e^{i/a}(i-a)}{a(ae^{i/a}+n)^2} - \sum_{i=1}^{\infty} \frac{n(i-a)}{a^2(ae^{i/a}+n)}\Big),$$

with expected value

$$\text{(E.2)} \qquad E_0[\mathbb{M}_n(a)] = \frac{1}{2}\Big(\sum_{i=1}^{\infty} \frac{n^2(i-a)e^{i/a} f_{0,i}^2}{a(ae^{i/a}+n)^2} - \sum_{i=1}^{\infty} \frac{n^2(i-a)}{a^2(ae^{i/a}+n)^2}\Big).$$

In the following subsections we show with the help of the score function $\mathbb{M}_n(a)$ that the marginal likelihood function $\ell_n(a)$ with probability tending to one has its global maximum outside of the set $[1, \underline{a}_n) \cup (\overline{a}_n, A_n]$.

Adaptation to smoothness

E.2. $\mathbb{M}_n(a)$ **on** $[\overline{a}_n, A_n]$. By assuming $\overline{a}_n > K_0$, we have $h_n(a, f_0) \leq b$ for $a \in [\overline{a}_n, A_n]$. Next we prove that for sufficiently large choice of $K_0 > 0$

$$\text{(E.8)} \qquad \limsup_n \sup_{f_0 \in \ell_2(M)} \sup_{a \in [\overline{a}_n, A_n]} E_0\left[\frac{\mathbb{M}_n(a)}{\log^2(n/a)}\right] < -2^{-5},$$

$$\boxed{\text{(E.9)}} \qquad \limsup_n \sup_{f_0 \in \ell_2(M)} E_0\left[\sup_{a \in [\overline{a}_n, A_n]} \frac{|\mathbb{M}_n(a) - E_0[\mathbb{M}_n(a)]|}{\log^2(n/a)}\right] \leq 2^{-6}.$$

Adaptation to smoothness

E.2. $\mathbb{M}_n(a)$ **on** $[\overline{a}_n, A_n]$. By assuming $\overline{a}_n > K_0$, we have $h_n(a, f_0) \leq b$ for $a \in [\overline{a}_n, A_n]$. Next we prove that for sufficiently large choice of $K_0 > 0$

$$\text{(E.8)} \qquad \limsup_n \sup_{f_0 \in \ell_2(M)} \sup_{a \in [\overline{a}_n, A_n]} E_0\left[\frac{\mathbb{M}_n(a)}{\log^2(n/a)}\right] < -2^{-5},$$

$$\text{(E.9)} \qquad \limsup_n \sup_{f_0 \in \ell_2(M)} E_0\left[\sup_{a \in [\overline{a}_n, A_n]} \frac{|\mathbb{M}_n(a) - E_0[\mathbb{M}_n(a)]|}{\log^2(n/a)}\right] \leq 2^{-6}.$$

- $\psi(x) = x^2$ induces an Orlicz norm
- $\{\frac{|\mathbb{M}_n(a) - E_0\mathbb{M}_n(a)|}{\log^2(n/a)} : a \in [\overline{a}_n, A_n]\}$ separable for all n
- $\|\frac{|\mathbb{M}_n(a_1) - E_0\mathbb{M}_n(a_1)|}{\log^2(n/a_1)} - \frac{|\mathbb{M}_n(a_2) - E_0\mathbb{M}_n(a_2)|}{\log^2(n/a_2)}\|_\psi =$ $V_0\left(\frac{\mathbb{M}_n(a_1)}{\log^2(n/a_1)} - \frac{\mathbb{M}_n(a_2)}{\log^2(n/a_2)}\right)$

Adaptation to smoothness

Proof of assertion (E.9): In view of Corollary 2.2.5 in [34] (applied with $\psi(x) = x^2$) it is sufficient to show that there exist universal constants $K_1, K_2 > 0$ such that for any $a \in [\overline{a}_n, A_n]$

$$V_0\big(\mathbb{M}_n(a)/\log^2(n/a)\big) \leq K_1/\log(n/a), \tag{E.11}$$

$$\int_0^{diam_n} \sqrt{N(\varepsilon, [\overline{a}_n, A_n], d_n)}d\varepsilon \leq K_2/K_0^{1/4}, \tag{E.12}$$

where d_n is the semimetric defined by $d_n^2(a_1, a_2) := V_0\Big(\frac{\mathbb{M}_n(a_1)}{\log^2(n/a_1)} - \frac{\mathbb{M}_n(a_2)}{\log^2(n/a_2)}\Big)$, $diam_n$ is the diameter of $[\overline{a}_n, A_n]$ relative to d_n and $N(\varepsilon, S, d_n)$ is the minimal number of d_n-balls of radius ε needed to cover the set S, since by sufficiently large choice of K_0 ($K_0 \geq (2^6 K_2)^4$ is sufficiently large) assertion (E.9) holds.

Adaptation to smoothness - Examples

Knapik B.T., Szabó B.T., van der Vaart A. W. and van Zanten J. H.
Bayes procedures for adaptive inference in inverse problems for the white noise model
Probability Theory and Related Fields, 164 (2016), pp. 771813.

Szabó B.T., van der Vaart A. W. and van Zanten J. H.
Frequentist coverage of adaptive nonparametric Bayesian credible sets
Annals of Statistics, 43 (2015), pp. 13911428.

Hadji A. and Szabó B.T.
Can we trust Bayesian uncertainty quantification from Gaussian process priors with squared exponential covariance kernel?
arXiv preprint arXiv:1904.01383.

Sub-Gaussian processes

Definition

A stochastic process $\{X_t : t \in T\}$ is called *sub-Gaussian* with respect to the semi-metric d on T if

$$P(|X_s - X_t| > x) \leq 2e^{-\frac{1}{2}x^2/d^2(s,t)}, \text{ for every } s, t \in T, x > 0.$$

Sub-Gaussian processes

Definition

A stochastic process $\{X_t : t \in T\}$ is called *sub-Gaussian* with respect to the semi-metric d on T if

$$P(|X_s - X_t| > x) \leq 2e^{-\frac{1}{2}x^2/d^2(s,t)}, \text{ for every } s, t \in T, x > 0.$$

Examples

- Any Gaussian process is sub-Gaussian for the standard deviation metric $d(s,t) = \sigma(X_s - X_t)$
- The *Rademacher process*, defined as $X_a = \sum_{i=1}^n a_i \varepsilon_i$ for $a \in \mathbb{R}^n$ and $\varepsilon_1, \ldots, \varepsilon_n$ Rademacher variables.
  The Rademacher process is a sub-Gaussian process for the Euclidiean distance (Hoeffding's inequality).

Symmetrization

We already now the empirical process for $(X_1, ..., X_n)$ i.i.d random variables

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

However, proving empirical limit theorems for a class of function $\mathcal{F}$ using this form is difficult (impossible).

Symmetrization

We already now the empirical process for $(X_1, ..., X_n)$ i.i.d random variables

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

However, proving empirical limit theorems for a class of function $\mathcal{F}$ using this form is difficult (impossible).
Let $\varepsilon_1, ..., \varepsilon_n$ i.i.d Rademacher random variables independent of $(X_1, ..., X_n)$. We'll consider the symmetrized process

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

Symmetrization

Lemma

For every nondecreasing, convex $\Phi : \mathbb{R} \to \mathbb{R}$ and class of measurable functions $\mathcal{F}$,

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

In most if not all of the chapter, $\Phi(x) = x$.

Symmetrization

Lemma

For every nondecreasing, convex $\Phi : \mathbb{R} \to \mathbb{R}$ and class of measurable functions $\mathcal{F}$,

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

In most if not all of the chapter, $\Phi(x) = x$.

Idea of the proof:

We know that $\|\mathbb{P}_n - P\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} \frac{1}{n}\Big|\sum_i (f(X_i) - Ef(X))\Big|$ where X follows the same distribution as the X_i's. But somehow, we want to get rid of $Ef(X)$, hence we create Y_i's independent copies of X_i's.

Proof of the Lemma

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

$$\|\mathbb{P}_n - P\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} \frac{1}{n}\Big|\sum_i (f(X_i) - Ef(Y_i))\Big|$$

Proof of the Lemma

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

$$\begin{aligned}\|\mathbb{P}_n - P\|_{\mathcal{F}} &= \sup_{f\in\mathcal{F}} \frac{1}{n}\Big|\sum_i (f(X_i) - Ef(Y_i))\Big| \\ &\leq E_Y^* \sup_{f\in\mathcal{F}} \frac{1}{n}\Big|\sum_i (f(X_i) - f(Y_i))\Big|\end{aligned}$$

Proof of the Lemma

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

$$\begin{aligned}
\|\mathbb{P}_n - P\|_{\mathcal{F}} &= \sup_{f\in\mathcal{F}} \frac{1}{n}\Big|\sum_i (f(X_i) - Ef(Y_i))\Big| \\
&\leq E_Y^* \sup_{f\in\mathcal{F}} \frac{1}{n}\Big|\sum_i (f(X_i) - f(Y_i))\Big| \\
&= E_Y \Big\|\frac{1}{n}\Big|\sum_i (f(X_i) - f(Y_i))\Big|\Big\|_{\mathcal{F}}^{*Y}
\end{aligned}$$

Proof of the Lemma

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

Using Jensen's unequality, we get

$$\Phi(\|\mathbb{P}_n - P\|_{\mathcal{F}}) = E_Y\Phi\left(\left\|\frac{1}{n}|\sum_i (f(X_i) - f(Y_i))|\right\|_{\mathcal{F}}^{*Y}\right)$$

Proof of the Lemma

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

$$\Phi(\|\mathbb{P}_n - P\|_{\mathcal{F}}) = E_Y\Phi\left(\left\|\frac{1}{n}\Big|\sum_i (f(X_i) - f(Y_i))\Big|\right\|_{\mathcal{F}}^{*Y}\right)$$

$$E_X^*\Phi(\|\mathbb{P}_n - P\|_{\mathcal{F}}) \leq E_X^* E_Y^*\Phi\left(\left\|\frac{1}{n}\Big|\sum_i (f(X_i) - f(Y_i))\Big|\right\|_{\mathcal{F}}\right)$$

Proof of the Lemma

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

$$\Phi(\|\mathbb{P}_n - P\|_{\mathcal{F}}) = E_Y\Phi\left(\left\|\frac{1}{n}\Big|\sum_i (f(X_i) - f(Y_i))\Big|\right\|_{\mathcal{F}}^{*Y}\right)$$

$$E_X^*\Phi(\|\mathbb{P}_n - P\|_{\mathcal{F}}) \leq E_X^*E_Y^*\Phi\left(\left\|\frac{1}{n}\Big|\sum_i (f(X_i) - f(Y_i))\Big|\right\|_{\mathcal{F}}\right)$$

For any $(e_1, ..., e_n) \in \{-1, 1\}^n$, the expression

$$E_X^*E_Y^*\Phi\left(\left\|\frac{1}{n}\Big|\sum_i e_i(f(X_i) - f(Y_i))\Big|\right\|_{\mathcal{F}}\right)$$

is identical.

Proof of the Lemma

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

$$\begin{aligned} E^*\Phi(\|\mathbb{P}_n - P\|_{\mathcal{F}}) &\leq E_\epsilon E_X^* E_Y^* \Phi\left(\left\|\frac{1}{n}\Big|\sum_i \varepsilon_i (f(X_i) - f(Y_i))\Big|\right\|_{\mathcal{F}}\right) \\ &\leq E_\epsilon E_{X,Y}^* \Phi\left(\left\|\frac{1}{n}\Big|\sum_i \varepsilon_i (f(X_i) - f(Y_i))\Big|\right\|_{\mathcal{F}}\right) \end{aligned}$$

Proof of the Lemma

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

$$E^*\Phi(\|\mathbb{P}_n - P\|_{\mathcal{F}}) \leq E_\epsilon E_X^* E_Y^* \Phi\left(\Big\|\frac{1}{n}\Big|\sum_i \varepsilon_i (f(X_i) - f(Y_i))\Big|\Big\|_{\mathcal{F}}\right)$$

$$\leq E_\epsilon E_{X,Y}^* \Phi\left(\Big\|\frac{1}{n}\Big|\sum_i \varepsilon_i (f(X_i) - f(Y_i))\Big|\Big\|_{\mathcal{F}}\right)$$

Using the triangle inequality, we can separate the X's and the Y's and use Φ's convexity to get the bound

$$\frac{1}{2}E_\epsilon E_{X,Y}^*\Phi\left(2\Big\|\frac{1}{n}\Big|\sum_i \varepsilon_i f(X_i)\Big|\Big\|_{\mathcal{F}}\right) + \frac{1}{2}E_\epsilon E_{X,Y}^*\Phi\left(2\Big\|\frac{1}{n}\Big|\sum_i \varepsilon_i f(Y_i)\Big|\Big\|_{\mathcal{F}}\right).$$

Measurable class

The lemma is valid for any class $\mathcal{F}$. The next step would be to apply a maximal inequality to the right side. Unfortunately, Fubini's theorem is not valid for outer expectations.

Measurable class

The lemma is valid for any class $\mathcal{F}$. The next step would be to apply a maximal inequality to the right side. Unfortunately, Fubini's theorem is not valid for outer expectations.

Definition

A class $\mathcal{F}$ of measurable functions $f : \mathcal{X} \to \mathbb{R}$ on a probability space $(\mathcal{X}, \mathcal{A}, P)$ is called a *P-measurable class* if

$$(X_1, ..., X_2) \mapsto \| \sum_i e_i f(X_i) \|$$

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

More about symmetrization

In order to generalize the notation, instead of the empirical distribution, we consider sums $\sum_i Z_i$ of independent stochastic processes $\{Z_i(f) : f \in \mathcal{F}\}$.

Lemma

For every nondecreasing, convex $\Phi : \mathbb{R} \to \mathbb{R}$ and class of measurable functions $\mathcal{F}$, let $Z_1, ..., Z_n$ be independent stochastic process with mean zero. Then

$$E^*\Phi\left(\frac{1}{2}\|\Sigma_i\varepsilon_i Z_i|_{\mathcal{F}}\right) \leq E^*\Phi\left(\|\Sigma_i Z_i|_{\mathcal{F}}\right) \leq E^*\Phi\left(2\|\Sigma_i\varepsilon_i(Z_i - \mu_i)|_{\mathcal{F}}\right),$$

for arbitrary $\mu_i : \mathcal{F} \to \mathbb{R}$.

Final results

One important result of this chapter is Lemma 2.3.11 which gives sufficient conditions for convergence of the empirical process $(\mathcal{P}_n - P)f$

Final results

One important result of this chapter is Lemma 2.3.11 which gives sufficient conditions for convergence of the empirical process $(\mathcal{P}_n - P)f$

Lemma

Let $Z_1, Z_2, \ldots$ be i.i.d stochastic processes, linear in f. Set $\rho_Z(f,g) = \sigma(Z_1(f) - Z_1(g))$ and $\mathcal{F}_\delta = \{f - g : \rho_Z(f,g) < \delta\}$. Then the following statements are equivalent:

Final results

One important result of this chapter is Lemma 2.3.11 which gives sufficient conditions for convergence of the empirical process $(\mathcal{P}_n - P)f$

Lemma

Let $Z_1, Z_2, \ldots$ be i.i.d stochastic processes, linear in f. Set $\rho_Z(f,g) = \sigma(Z_1(f) - Z_1(g))$ and $\mathcal{F}_\delta = \{f - g : \rho_Z(f,g) < \delta\}$. Then the following statements are equivalent:

- *(i) $n^{-1/2}\sum_i Z_i$ converges weakly to a tight limit in $\ell^\infty(\mathcal{F})$;*
- *(ii) $(\mathcal{F}, \rho_Z)$ is totally bounded and $\|n^{-1/2}\sum_i Z_i\|_{\mathcal{F}_{\delta_n}} \xrightarrow{P^*} 0$ for every $\delta_n \downarrow 0$;*

Final results

One important result of this chapter is Lemma 2.3.11 which gives sufficient conditions for convergence of the empirical process $(\mathcal{P}_n - P)f$

Lemma

Let $Z_1, Z_2, \ldots$ be i.i.d stochastic processes, linear in f. Set $\rho_Z(f,g) = \sigma(Z_1(f) - Z_1(g))$ and $\mathcal{F}_\delta = \{f - g : \rho_Z(f,g) < \delta\}$. Then the following statements are equivalent:

- *(i) $n^{-1/2}\sum_i Z_i$ converges weakly to a tight limit in $\ell^\infty(\mathcal{F})$;*
- *(ii) $(\mathcal{F}, \rho_Z)$ is totally bounded and $\|n^{-1/2}\sum_i Z_i\|_{\mathcal{F}_{\delta_n}} \xrightarrow{P^*} 0$ for every $\delta_n \downarrow 0$;*
- *(iii) $(\mathcal{F}, \rho_Z)$ is totally bounded and $E^*\|n^{-1/2}\sum_i Z_i\|_{\mathcal{F}_{\delta_n}} \to 0$ for every $\delta_n \downarrow 0$.*

Final results

Corollary

Let $\mathcal{F}$ be a class of measurable functions. Then the following statements are equivalent:

Final results

Corollary

Let $\mathcal{F}$ be a class of measurable functions. Then the following statements are equivalent:

- *(i) $\mathcal{F}$ is P-Donsker;*
- *(ii) $(\mathcal{F}, \rho_P)$ is totally bounded and $\sqrt{n}\|\mathbb{P}_n - P\|_{\mathcal{F}_{\delta_n}} \xrightarrow{P^*} 0$ for every $\delta_n \downarrow 0$;*

Final results

Corollary

Let $\mathcal{F}$ be a class of measurable functions. Then the following statements are equivalent:

- *(i) $\mathcal{F}$ is P-Donsker;*
- *(ii) $(\mathcal{F}, \rho_P)$ is totally bounded and $\sqrt{n}\|\mathbb{P}_n - P\|_{\mathcal{F}_{\delta_n}} \xrightarrow{P^*} 0$ for every $\delta_n \downarrow 0$;*
- *(iii) $(\mathcal{F}, \rho_Z)$ is totally bounded and $E^*\sqrt{n}\|\mathbb{P}_n - P\|_{\mathcal{F}_{\delta_n}} \to 0$ for every $\delta_n \downarrow 0$.*

Final results

Corollary

Let $\mathcal{F}$ be a class of measurable functions. Then the following statements are equivalent:

- *(i) $\mathcal{F}$ is P-Donsker;*
- *(ii) $(\mathcal{F}, \rho_P)$ is totally bounded and $\sqrt{n}\|\mathbb{P}_n - P\|_{\mathcal{F}_{\delta_n}} \xrightarrow{P^*} 0$ for every $\delta_n \downarrow 0$;*
- *(iii) $(\mathcal{F}, \rho_Z)$ is totally bounded and $E^*\sqrt{n}\|\mathbb{P}_n - P\|_{\mathcal{F}_{\delta_n}} \to 0$ for every $\delta_n \downarrow 0$.*

Moreover, every P-Donsker class $\mathcal{F}$ satisfies $P(\|f - Pf\|^*_{\mathcal{F}} > x) = o(x^{-2})$ as $x \to +\infty$. Therefore, if $\|Pf\|_\infty < +\infty$, then $\mathcal{F}$ possesses an envelope function F with $P(F > x) = o(x^{-2})$.