Empirical processes : uniform law of large numbers

ULLN

① Weak convergence and empirical processes

② Empirical processes with application in M-estimation

③ [Sen]

$X_1 \cdots X_n \overset{iid}{\sim} F_X$ c.d.f.

Estimate c.d.f.

Empirical c.d.f. $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1\{X_i \le x\}$

$1\{X_i \le x\} \sim \text{Bernoulli}(F(x))$

R.k. consistent estimator

$\lim_{n \to \infty} F_n(x) = F_X(x)$ for any given x

pointwise

Plug-in estimation $\rightarrow \theta = T(F)$ T: is functional

$\theta = \int r(x)\, dF(x)$; quantile : inverse CDF

$F^{-1}(\alpha) = \inf\{t : F(t) \ge \alpha\}$

$\hat{\theta} = \int r(x)\, dF_n(x)$

$= T(F_n)$

$\hat{\theta} - \theta$ as $n \to \infty$

Question.

$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \to 0$ a.s. $\to$ ULLN

$\sqrt{n}\, |F_n(x) - F(x)| \xrightarrow{d}$ stochastic process UCLT

$\hookrightarrow$ Brownian bridge

$\hookrightarrow$ Glivenko - Cantelli theorem.

$\text{Cov}(F_n(x), F_n(y)) = \min\{F_n(x), F_n(y)\} - F_n(x) F_n(y)$

Today: $\mathbb{E}\ \sup_x |F_n(x) - F(x)| \to 0$

$\mathbb{P}$: population distribution
$\mathbb{P}_n$: empirical measure $\quad \mathbb{P}_n = \frac{1}{n}\sum_i \delta_{x_i}$

Def $f \in \mathcal{F}$

$$\mathbb{P} f = \int f \, dP$$

$$\mathbb{P}_n f = \int f \, d\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} f(x_i) \to \mathbb{P} f$$

Study $\sup_{f\in\mathcal{F}} |\mathbb{P}_n f - \mathbb{P} f| \to 0$ as $n \to \infty$ $\quad (*)$

$$\mathcal{F} := \{ 1\{t \le x\} : x \in \mathbb{R} \}$$

Question: When does $(*)$ hold?

Answer depends $\mathcal{F}$: $\begin{cases}\text{complexity of } \mathcal{F} \\ \text{size}\end{cases}$

Empirical processes $\mathbb{P}_n f$ is a stochastic process indexed by $\mathcal{F}$.

Def: If for some $\mathcal{F}$, $(*)$ holds, then $\mathcal{F}$ is called Glivenko–Cantelli class.

$$\mathcal{F} := \{ \ell(x;\theta) = (f(x;\theta) - y_i)^2 \}$$
$$\theta \in \Theta$$

$$\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} := \sup_{f\in\mathcal{F}} |\mathbb{P}_n f - \mathbb{P} f|$$

$\|P_n - P\|_{\mathcal{F}}$

Symmetrization.

Rade comp of $\{(f(x_1), \dots, f(x_n)) : f \in \mathcal{F}\}$

Lemma:

$$\mathbb{E}_{X_i} \|P_n - P\|_{\mathcal{F}} \le 2\, \mathbb{E}_{\varepsilon_i, X_i} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f(X_i) \right|$$

ε_i: Rademacher random variable

$$P(\varepsilon_i = 1) = P(\varepsilon_i = -1) = \frac{1}{2}$$

Def: Rademacher complexity of $S \subseteq \mathbb{R}^n$.

$$\mathrm{Rad}(S) := \mathbb{E} \sup_{s \in S} \frac{1}{n} \left| \sum_{i=1}^{n} \varepsilon_i s_i \right|$$

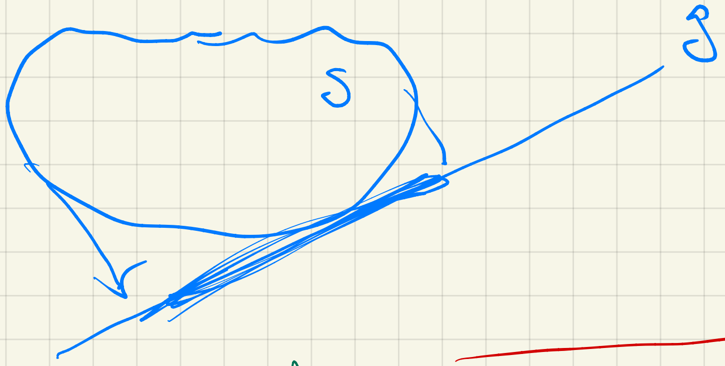
$\mathrm{conv}(S)$

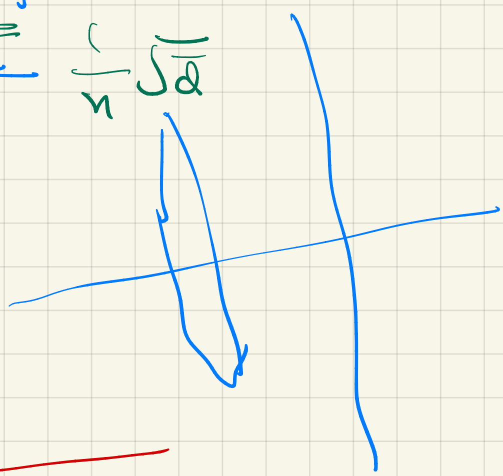
$\vec{s} = (s_1, \dots, s_n) \in S$

Gaussian width $\sup_{s \in S} \langle g, s \rangle$

$$w(S) = \mathbb{E} \sup_{s \in S} \frac{1}{n} \left| \sum_{i=1}^{n} g_i s_i \right| \qquad g_i \sim N(0, I)$$

$\frac{1}{n}\sqrt{d}$





$S \subseteq \mathbb{R}^d$ $\qquad S \subseteq \mathbb{R}^{\tilde{d}}$ $\qquad S = [s_1 \cdots s_{\tilde{d}}, 0 \cdots 0]$

Pf: Jensen's inequality

$$\mathbb{E}_X \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} := \mathbb{E}_X \frac{1}{n} \sup_f \left| \sum_{i=1}^n (f(X_i) - \mathbb{E} f(X)) \right|$$

$$= \mathbb{E}_X \frac{1}{n} \sup_f \underline{\left| \sum_{i=1}^n (f(X_i) - \mathbb{E}_{X'} f(X_i')) \right|}$$

X_i' is an independent copy of X_i

Jensen

$$\le \mathbb{E}_X \frac{1}{n} \underline{\underline{\sup_f}}\ \underline{\mathbb{E}_{X'}} \left| \sum_{i=1}^n (f(X_i) - f(X_i')) \right|$$

$$\le \mathbb{E}_{X,X'} \frac{1}{n} \sup_f \left| \sum_{i=1}^n (\underline{f(X_i) - f(X_i')}) \right|$$

$$f(X_i) - f(X_i') \overset{d}{=} \varepsilon_i (f(X_i) - f(X_i'))$$

$\varepsilon_i \sim$ Rade r.v.

$$= \mathbb{E}_{X,X'} \mathbb{E}_\varepsilon \frac{1}{n} \sup_f \left| \sum_{i=1}^n \varepsilon_i (\underline{f(X_i) - f(X_i')}) \right|$$

$$\le 2 \mathbb{E}_{X,\varepsilon} \frac{1}{n} \sup_f \left| \sum_{i=1}^n \varepsilon_i f(X_i) \right|$$

Ex. $\mathbb{E} \sup_x |F_n(x) - F_X(x)| = O\left(\sqrt{\frac{\log n}{n}}\right)$

Pf. $\mathbb{E} \sup_x |F_n(x) - F_X(x)| \le 2 \mathbb{E}_{X,\varepsilon} \sup_{f \in \mathcal{F}} \frac{1}{n} \left| \sum_{i=1}^n \varepsilon_i f(X_i) \right|$

$f_x(t) = 1_{\{t \le x\}}$

$$= 2 \mathbb{E}_{X,\varepsilon} \sup_x \frac{1}{n} \left| \sum_{i=1}^n \varepsilon_i 1_{\{X_i \le x\}} \right|$$

$$= 2\,\mathbb{E}_{X,\varepsilon} \sup_x \frac{1}{n} \left| \sum_{i=1}^n \varepsilon_i \mathbb{1}\{X_i \le x\} \right| .$$

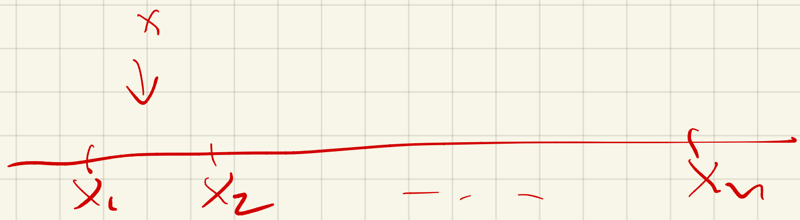
$$S_x = \{ (\mathbb{1}_{\{X_1 \le x\}} \cdots \mathbb{1}_{\{X_n \le x\}}) \mid x \in \mathbb{R} \}$$

$$= 2\,\mathbb{E}_X \left\{ \left(\mathbb{E}_\varepsilon \sup \frac{1}{n} \left| \sum_{i=1}^n \varepsilon_i \mathbb{1}\{X_i \le x\} \right| \;\middle|\; X_1 \cdots X_n \right\} \right.$$

w.l.o.g $X_1 \le X_2 \le \cdots \le X_n$

$$S_x = \{ (1,0,\cdots,0) \quad (0,\cdots,0) \quad (1,1,0\cdots 0) \cdots (1,\cdots,1) \}$$

n+1 possibility



$|S_x| \le n+1$

$\Downarrow$

finite set

If $|S| < \infty$, what is Rad(S)?

Massart's lemma

$$\mathbb{E} \sup_{\vec{s}_i \in S} |\langle \vec{\varepsilon}, \vec{s} \rangle| \le R\sqrt{2 \log 2|S|}$$

R: radius of S $\quad R := \max_s \|s\|$

$$\le \mathbb{E}_X \left(\frac{1}{n} \cdot \underset{\sqrt{n}}{\underset{\downarrow}{R}} \cdot \sqrt{2 \log 2(n+1)} \right)$$

$$= \sqrt{\frac{2 \log 2(n+1)}{n}}$$

$$\mathbb{E} \sup_{s \in \underline{\underline{S}}} \left| \sum_i \varepsilon_i s_i \right| \le \mathbb{E} \underbrace{\sup_{s \in S \cup (-S)} \sum_{i=1}^{n} \varepsilon_i s_i}_{= Z}$$

For any $\lambda > 0$

$$\underline{\underline{e^{\lambda \mathbb{E} Z}}} \le \underline{\mathbb{E} e^{\lambda Z}}$$

$$\le \sum_{s \in S \cup (-S)} \underline{\mathbb{E} \exp\left(\lambda \sum_{i=1}^{n} \varepsilon_i s_i\right)}$$

$$= \sum_{s \in S \cup (-S)} \prod_{i=1}^{n} \boxed{\mathbb{E}\left[\underline{\exp(\lambda \varepsilon_i s_i)}\right]} \le \exp\left(\frac{\lambda^2 s_i^2}{2}\right)$$

$$\le 2|S| \exp\left(\frac{\lambda^2}{2} \|s\|^2\right)$$

$$\underline{\mathbb{E} Z \le \frac{1}{\lambda}\left(\log 2|S| + \frac{\lambda^2}{2} R^2\right)}$$

$\lambda = \dots$

$$\mathbb{E}_{X, \varepsilon} \frac{1}{n} \sup_{f \in \underline{\mathcal{F}}} \underline{\left| \sum_{i=1}^{n} \varepsilon_i \underline{\underline{f(x_i)}} \right|}$$

$[\underline{f(x_1)} \cdots \underline{f(x_n)}]$? $\mathcal{F}$

Net - argument

Def: η-net: We say $\mathcal{G}$ is an η-net of $\mathcal{F}$ under d, if $\forall f$ in $\mathcal{F}$, $\exists g \in \mathcal{G}$. such that $d(f, g) \le \eta$

$$\bigcup_{g \in \mathcal{G}} B_d(g, \eta) \supseteq \mathcal{F}$$

Def: Covering number of $\mathcal{F}$ under d: the smallest cardinality of η-net

$$\mathbb{E}_{X,\varepsilon}\ \frac{1}{n}\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right|$$

Thm. Let $F(x) := \sup_{f\in\mathcal{F}} f(x)$, ~~is~~ be the envelop of $\mathcal{F}$. Suppose

① $\mathbb{P}F < \infty$ ✓

② $\frac{1}{n}\log \boxed{N(\mathcal{F}, L_1(\mathbb{P}_n), \varepsilon)} \to 0$ as $n\to\infty$. $\left(\frac{1}{\varepsilon}\right)^d$

then $\mathcal{F}$ is Glivenko–Cantelli $\Rightarrow \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \to 0$ a.s.

$L_1(\mathbb{P}_n) = \|f\|_{L_1(\mathbb{P}_n)} = \frac{1}{n}\sum_i |f(X_i)|$

$N(\mathcal{F}, L_1(\mathbb{P}_n), \varepsilon)$ is the covering number.

$\log N(\cdots)$ metric entropy

Assume $\mathcal{F}$ is uniformly bounded. $\forall f\in\mathcal{F}$, $|f|\le M$.

①
$$\mathbb{E}_{X_i}\left(\mathbb{E}_\varepsilon \frac{1}{n}\sup_f\left|\sum_{i=1}^n \varepsilon_i f(X_i)\,\mathbb{1}\{f(X_i)\le M\}\right| \,\middle|\, X_1,\dots,X_n\right) = \mathrm{I}$$

$$+\mathbb{E}\frac{1}{n}\sup_f\left|\sum_{i=1}^n \varepsilon_i f(X_i)\,\mathbb{1}\{f(X_i)\ge M\}\right| = \mathrm{II}$$

$$\mathrm{II} \le \mathbb{E}\sup_f \sum_{i=1}^n |f(X_i)|\cdot \mathbb{1}\{f(X_i)\ge M\}$$

$$\le \mathbb{E}\,|F|\cdot \mathbb{1}_{\{F\ge M\}}$$

② Net-argument. Let $\mathcal{G}$ be an η-net

For any $f\in\mathcal{F}$ $\exists\, g\in\mathcal{G}$, such

$$\|f-g\|_{L_1(\mathbb{P}_n)} = \frac{1}{n}\sum |f(X_i) - g(X_i)| \le \eta.$$

$$\mathbb{E}_\varepsilon \frac{1}{n} \sup_{f\in\mathcal{F}} \left| \sum \varepsilon_i \left(f(x_i) - g(x_i)\right)\right|$$

$$+ \mathbb{E}_\varepsilon \frac{1}{n} \sup_{g\in\mathcal{G}} \left| \sum \varepsilon_i\, g(x_i) \right|$$

$\rightarrow$ Massart

$$\leq \eta + \mathbb{E}_\varepsilon \frac{1}{n} \sup_{g\in\mathcal{G}} \left| \sum \varepsilon_i\, g(x_i) \right|$$

$$|\mathcal{G}| = N(\mathcal{F}, L_1(P_n), \eta) < \infty$$

$$\leq \eta + \frac{1}{n} \sqrt{\sup_{g\in\mathcal{G}} \underbrace{\sum_{i=1}^n g^2(x_i)}_{\leq n M^2}} \cdot \sqrt{2 \log 2|\mathcal{G}|}$$

$$\leq \eta + M \cdot \sqrt{\frac{2\log 2|\mathcal{G}|}{n}} \quad \xrightarrow{n\to\infty} \leq \eta$$

$$\lim_{n\to\infty} \mathbb{E}_\varepsilon \left(\|P_n - P\|_{\mathcal{F}} \mid x_1 \cdots x_n \right) = 0$$

$$|P_n f| \leq M ,$$

dominating convergence

$$\Rightarrow \quad \mathbb{E}\, \|P_n - P\|_{\mathcal{F}} \to 0 \quad \text{as } n\to\infty$$

almost sure convergence