Empirical Process & VC (DasGupta, PSML, Ch. 16)

$\{X_i\}_{i=1}^{n}$ – iid $\sim F$ (cdf)

Empirical CDF:

$$F_n(t) = \frac{\#\{i : X_i \le t\}}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \le t\}}$$

Glivenko-Cantelli Thm:

$$F_n(t) \xrightarrow{n\to\infty} F(t) \quad \text{uniformly in } t$$

$$\|F_n - F\|_{L^\infty(\mathbb{R})} = \sup_{t\in\mathbb{R}} |F_n(t) - F(t)| \xrightarrow[a.s.]{n\to\infty} 0$$

random

CLT (Scaling)

$$\beta_n(t) = \sqrt{n}\left(F_n(t) - F(t)\right) \xrightarrow[n\to\infty]{\mathcal{D}} N\left(0, F(t)(1-F(t))\right) \quad (t\text{ - fixed})$$

Kolmogorov-Smirnov Statistics

$$D_n = \sqrt{n}\,\|F_n - F\|_\infty$$

Well-known Example: $F \sim U([0,1])$

$$\beta_n(\cdot) \xrightarrow[t\in[0,1]]{\mathcal{D}} B_u(\cdot)$$ Brownian Bridge

$$D_n \xrightarrow{\mathcal{D}} \sup_{t\in[0,1]} B_u(t)$$

(convergence as a process)

(Thm 16.1, Example 16.1, 16.2)

K-S   Cramer-von Mises

More general Empirical Process, indexed by sets $C \in \mathcal{C}$, or functions $f \in \mathcal{F}$

$$\overset{\shortparallel}{\mathbb{1}_C}, \quad C \subseteq S$$

$\{X_i\}_{i=1}^n \sim P,$

$$\mathbb{E}_n \mathbb{1}_C = P_n(C) = \frac{1}{n}\#\{i : X_i \in C\}$$

$$= \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{X_i \in C\}$$

$$\mathbb{E}_n f = \mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^{n} f(X_n)$$

(Classical empirical process, $C_t = (-\infty, t]$)

<u>C, f - fixed</u>

LLN, $\mathbb{P}_n(C) \xrightarrow{n\to\infty} P(C)$, $\mathbb{E}_n f \longrightarrow Ef = \int f\,dP$

CLT, $\sqrt{n}\,(\mathbb{P}_n(C) - P(C)) \xrightarrow{\mathcal{D}} N(0,\ P(C)(1-P(C)))$

$\sqrt{n}\,(E_n f - Ef) \Longrightarrow N(0,\ E(f^2) - E^2(f))$

<u>Q</u>:

$$\left.\begin{array}{l} \sup\limits_{C\in\mathcal{C}} |\mathbb{P}_n(C) - P(C)| \\ \sup\limits_{f\in\mathcal{F}} |E_n f - Ef| \end{array}\right\} \xrightarrow[a.s]{n\to\infty} 0\ ?$$

$$\left.\begin{array}{l} \sqrt{n}\,(\mathbb{P}_n(\cdot) - P(\cdot)) \\ \sqrt{n}\,(E_n(\cdot) - E(\cdot)) \end{array}\right\} \xRightarrow[n\to\infty]{\mathcal{D}}$$ Gaussian Process indexed by C, f?

Vapnik–Chervonenkis Theory

(A measurement of the complexity of $\mathcal{C}$.)

$$\Delta^{\mathcal{C}}(x_1, x_2, \cdots, x_n) = \underbrace{\mathrm{Card}\left(\{x_1, \cdots x_n\} \cap C : C \in \mathcal{C}\right)}$$

number of subsets of $\{x_1, \cdots, x_n\}$ picked up or "seen" by $\mathcal{C}$

$$S(n, \mathcal{C}) = \max_{x_1, x_2, \cdots, x_n \in S} \Delta^{\mathcal{C}}(x_1, \cdots, x_n)$$

$(\leq 2^n$, clearly$)$

Shattering coefficient

Def 1.6.4 $\quad VC(\mathcal{C}) = \min\{n : S(n, \mathcal{C}) < 2^n\} - 1$

$$= \max\{n : S(n, \mathcal{C}) = 2^n\}$$

$\mathcal{C}$ is called VC if $VC(\mathcal{C}) < \infty$

Thm (Sauer) Either,

$$S(n, \mathcal{G}) = 2^n \quad \forall n \quad \text{or} \qquad \text{exp. in } n$$

$$S(n, \mathcal{G}) \leq \sum_{i=0}^{VC(\mathcal{G})} \binom{n}{i} \quad \forall n \qquad \text{poly in } n$$

$$\frac{n(n-1)(n-2)\cdots(n-i+1)}{i!}$$

$$= \sum_{i=0}^{VC \wedge n} \binom{n}{i} \leq 2^n$$

$$\lesssim O\left(n^{VC(\mathcal{G})}\right)$$

<u>Example</u> $\mathcal{G} = \{ C_t = (-\infty, t] : t \in \mathbb{R} \}$

$n = 2$ $\quad \{x_1, x_2\} \quad x_1 < x_2$

$$S(2, \mathcal{G}) = 1 < 2^2 \Rightarrow VC(\mathcal{G}) = 1$$

Thm 16.10 $\{X_i\}_{i=1}^{n} \sim iid\ P$, $X_i \in \mathbb{R}^d$

$$P\left(\sup_{C \in \mathcal{C}} |P_n(C) - P(C)| > \varepsilon\right)$$

$$\le 8\, \underline{E}\left[\Delta^{\mathcal{C}}(X_1, \cdots X_n)\right] e^{-n\varepsilon^2/32}$$

$$\le 8\, S(n, \mathcal{C})\, e^{-n\varepsilon^2/32}$$

$$\Longrightarrow \|P_n - P\|_\infty \xrightarrow{n \to \infty} 0 \text{ in Prob.}$$

Thm 16.11

(a) $\|P_n - P\|_\infty \xrightarrow{a.s} 0 \iff \xrightarrow{\text{in Prob}} 0$

(b) $\|P_n - P\|_\infty \xrightarrow{a.s.} 0 \iff$

$$\frac{\log \Delta^{\mathcal{C}}(X_1, \cdots, X_n)}{n} \xrightarrow{\text{in Prob}} 0$$

$$VC(\mathcal{C}) < \infty \Longrightarrow \|P_n - P\|_\infty \xrightarrow{a.s.} 0$$

CLT for Empirical Process

f - fixed $\in \mathcal{F}$

$$\sqrt{n}\left(\mathbb{E}_n f - Ef\right) \to N\left(0, E(f^2) - E^2 f\right)$$

Gaussian Process

$$f \in \mathcal{F} \longrightarrow B_P(f) \sim N(0, \cdots)$$

$$\left(\text{Compare: } t \in [0,1] \longrightarrow B_U(t) \sim N(0, t(1-t))\right)$$

Brownian Bridge

$$B_n(t) = \sqrt{n}\left[F_n(t) - F(t)\right] \xrightarrow[n\to\infty]{\mathcal{D}} B_U(t)$$

vs. $$B_n(\cdot) = \sqrt{n}\left[F_n(\cdot) - F(\cdot)\right] \xrightarrow{\mathcal{D}} B_U(\cdot)$$

We need tightness condition on $\{B_n(\cdot)\}$

$\{P_n\}_{n\geq 1}$, prob. meas on X is tight if

$\forall \varepsilon > 0$, $\exists K^{\text{compact}} \subseteq X$ s.t. $P_n(K) \geq 1 - \varepsilon$ $\forall n$

Characteization of Compact Set on $C[0,1]$

Arzela-Ascoli Thm

$\{f_n\}_{n\geq 1} \subseteq C[0,1]$ is compact iff

(1) $\sup_n \|f_n\|_{L^\infty} < \infty$

(2) $\forall \varepsilon, \exists \delta$ s.t. $\forall n, \ \forall |x-y| \leq \delta,$

$$|f_n(x) - f_n(y)| \leq \varepsilon$$

$\{P_n\}$ — prob. meas on $C([0,1])$ iff.

(a) $\forall \eta > 0, \exists a > 0 \ni$

$$\overline{\lim_n} P_n\{f : |f(0)| \geq a\} < \eta$$

(b) $\forall \varepsilon, \ \lim_{\delta \to 0} P_n\{f : \omega_f(\delta) \geq \varepsilon\} = 0$

modulus of cont.

$$\omega_f(\delta) = \sup_{|x-y|\leq\delta} |f(x) - f(y)|$$

Giné-Nickl (Bk) Thm 3.7.23

P_n – prob. measures on $\ell^\infty(T)$ arbitrary index set

$$\ell^\infty(T) = \{X : T \to \mathbb{R},\ \sup_{t\in T}|X(t)| < \infty\}$$

$\{P_n\}$ on $\ell^\infty(T)$ is tight iff

$\exists$ a pseudometric on T s.t.

(1) T is totally bounded;

(2) $\forall \varepsilon > 0$,

$$\lim_{\delta\to 0}\ \overline{\lim_{n\to\infty}}\ P_n\{X : \sup_{d(s,t)\le\delta} |X(t) - X(s)| \ge \varepsilon\} = 0$$

G.-N. (Bk) Thm 3.7.31

$\mathcal{F} \subseteq L^2(X, \mathcal{B}, P)$, satisfies:

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

$$\left(\sup_{f \in \mathcal{F}} |f(x)|, \ \sup_{f \in \mathcal{F}} |Pf| < \infty, \ \forall x \in X \right)$$

$\mathcal{F}$ is Donsker iff

(1) $(\mathcal{F}, d_P)$ is totally bounded

(2) $\forall \varepsilon > 0$,

$$\lim_{\delta \to 0} \overline{\lim_{n}} \, \mathbb{P} \left\{ \sup_{\substack{f, g \in \mathcal{F} \\ d_P(f,g) \le \delta}} \left| \sqrt{n} (P_n - P)(f - g) \right| \geq \varepsilon \right\} = 0$$

$$d_P(f-g) = \|f-g\|_{L^2(\mathbb{P})}$$

Envelope of $\mathcal{F}$

① $F(x) = \sup\limits_{f \in \mathcal{F}} |f(x)|$

② $F \in L^2(P)$, $E(F^2) < \infty$

Thm 16.12 If

(1) $F \in L^2(P)$

(2) Subgraph of $\mathcal{F}$ is VC

Then

$$\sqrt{n}\left(E_n(\cdot) - E(\cdot)\right) \xrightarrow[n\to\infty]{\mathcal{D}} B_P(\cdot)$$

$\mathcal{F}$ is Donsker.

$C_f^+ = \{(x,y) : 0 \le y \le f(x)\}$

$C_f^- = \{(x,y) : f(x) \le y \le 0\}$

$\mathcal{C} = \{C_f^+, C_f^-, f \in \mathcal{F}\}$

Thm 16.13 $F \in L^2(P)$, $\mathcal{F}$ is Donsker if

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

or

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

Sup over any Prob. meas.

$N(\delta, \mathcal{F}, \|\ \|)$ $(g_i \in \tilde{\mathcal{F}},\ \mathcal{F} \subseteq \tilde{\mathcal{F}})$

$= \min\{n : \exists\, g_1, g_2, \ldots, g_n \text{ s.t. } \mathcal{F} \subseteq \bigcup_i B_\delta(g_i)\}$

$N_{[\,]}(\delta, \mathcal{F}, \|\ \|)$

$= \min\{n : \exists\, [l_i, u_i],\ i = 1, \ldots, n \text{ s.t. }$

$\|l_i - u_i\| \leq \delta,\ \mathcal{F} \subseteq \bigcup_{i=1}^n [l_i, u_i]\}$

$[l, u] = \{f : l \leq f \leq u\}$

Connection between N & VC

$\forall r \geq 1, \quad 0 < \varepsilon < 1$

$$N(\varepsilon \|F\|_{r,Q}, \mathcal{F}, L_r(Q))$$
$$\leq C(VC(\mathcal{G}_{\mathcal{F}})) \left(\frac{1}{\varepsilon}\right)^{rVC(\mathcal{G}_{\mathcal{F}})}$$

N is related to the "smoothness" of $\mathcal{F}$

(van der Vaart - Wellner, Bk) p. 134

(1) $\sup_Q \log N(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L^2(Q)) \lesssim \left(\frac{1}{\varepsilon}\right)^{2-\delta}$

(2) $\sup_Q N(\varepsilon \|F\|_{L^2(Q)}, \mathcal{F}, L^2(Q)) \lesssim \left(\frac{1}{\varepsilon}\right)^{V}$

(3) Thm 2.7.1

$$\log \underset{(N_{[\,]})}{N}(\varepsilon, C_1^\alpha(X), \|\cdot\|_\infty) \lesssim \left(\frac{1}{\varepsilon}\right)^{d/\alpha}$$

$(X \subseteq \mathbb{R}^d)$

(4) van de Geer (Bk) Thm 2.4

$$\log N(\varepsilon, \mathcal{F}, L^2) \lesssim \left(\frac{1}{\varepsilon}\right)^{\frac{1}{m}}$$

$$\rotatebox{90}{$\in$}$$

$$\int_0^1 \left(g^{(m)}(x)\right)^2 dx$$

Pakes-Pollard (Sim. & Asymp. of Opt. Est.)
Econometrica, 1989

$$G(\theta) = \int h(x, \theta)\, dP(x)$$ ← true fct to be min.

$$G_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} h(x_i, \theta)$$ ← sample approx. of G

θ^* — true min $\qquad\qquad$ $\hat{\theta}_n$ — approx. min.

(Assume $\{G_n\}$ - Glivenko-Cantelli, & Donsker)

"Thm 3.1" If

(a) $\|G_n(\hat{\theta}_n)\| \lesssim o_p(1) + \inf_{\theta \in \Theta} \|G_n(\theta)\|$

(b) $|G_n(\theta_*) - G(\theta_*)| = o_p(1)$

(c) $\sup_{\|\theta - \theta_*\| > \delta} \|G_n(\theta)\|^{-1} = O_p(1)$

Then $\hat{\theta}_n \longrightarrow \theta_*$ in Prob.

"Thm 3.3"

(a) $\|\hat{G}_n(\hat{\theta}_n)\| \leq o_p\left(\frac{1}{\sqrt{n}}\right) + \inf_{\theta} \|\hat{G}_n(\theta)\|$

(b) G is diff. at θ_*,

$DG(\theta_*)$, full rank

(c) $\sqrt{n}\left(\hat{G}_n(\theta_*) - G(\theta_*)\right) \longrightarrow N(0, V)$

(d) $\sqrt{n}\,(\hat{\theta}_n - \theta_*) \xrightarrow{\mathcal{D}}$

$$N\left(0, (\Gamma^T\Gamma)^{-1}\Gamma^T V \Gamma (\Gamma^T\Gamma)^{-1}\right)$$

$$\left((\Gamma)^{-1} V (\Gamma^T)^{-1}\right) ?$$