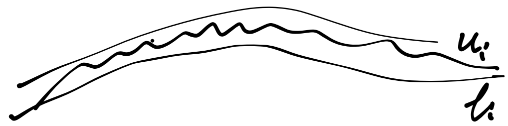
STA 3000F Lecture 13.

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} |\mathbb{P}_n f - Pf|\right] \le \frac{C}{\sqrt{n}}\int_0^1 \sqrt{\log \boxed{\sup_Q} N(\cdots)}\, d\delta$$

Bracketing # $N_{[\,]}(\varepsilon; \mathcal{F}, L^2(P))$

$\{[l_i, u_i]\}_{i=1}^{N}$



ε- bracket covering of $\mathcal{F}$

if $\forall f \in \mathcal{F}$, $\exists i$, $f \in [l_i, u_i]$

and $\|u_i - l_i\|_{L^2(P)} \le \varepsilon$.

Thm. $$\mathbb{E}\left[\sup_{f\in\mathcal{F}} |(\mathbb{P}_n - P)f|\right] \le \frac{C}{\sqrt{n}}\|F\|_{L^2(P)}\int_0^1 \sqrt{\log N_{[\,]}(\delta\|F\|_{L^2(P)}; \mathcal{F}, \|\cdot\|_{L^2(P)})}\, d\delta$$

$|f(x)| \le F(x)$ $(\forall f \in \mathcal{F})$.

---

e.g. K compact subset of $\mathbb{R}^d$ $\quad \mathcal{F} = \{f_\theta : \theta \in \Theta\}$

$|f_{\theta_1}(x) - f_{\theta_2}(x)| \le M(x)\|\theta_1 - \theta_2\|_2$.

Given $\theta_1, \theta_2, \ldots, \theta_N$ be a δ-covering of K.

$l_i(x) = f_{\theta_i}(x) - \delta \cdot M(x)$ $\qquad \|u_i - l_i\|_{L^2(P)} = 2\delta\|M\|_{L^2(P)}$

$u_i(x) = f_{\theta_i}(x) + \delta \cdot M(x)$

Back to M-estimators

$$R(\hat\theta_n) - R(\theta^*) = \underbrace{R(\hat\theta_n) - R_n(\hat\theta_n)}_{\text{Hard.}} + \underbrace{R_n(\hat\theta_n) - R_n(\theta^*)}_{\leq 0} + \underbrace{R_n(\theta^*) - R(\theta^*)}_{\text{Easy}}$$

$$\leq \sup_{\theta \in ?} | R(\theta) - R_n(\theta) |.$$

Thm (Localization). $F(\theta) - F(\theta^*) \geq \|\theta - \theta^*\|_2^2$ (Actually needed in local neighbor of θ^*)

$(\theta \in \mathbb{B}(\theta^*, r_0))$

$$\mathbb{E}\left[\sup_{\substack{\theta \in \Theta \\ \|\theta - \theta^*\|_2 \leq u}} \left| (P_n - P)(f_\theta - f^*) \right| \right] \lesssim \phi_n(u)$$

satisfying $\phi_n(cx) \leq c^\alpha \phi_n(x)$ $(\forall c > 0, x > 0)$

$(\alpha < 2)$

Then for δ_n s.t. $\phi_n(\delta_n) \leq \delta_n^2$, we have

$\forall \varepsilon > 0$, $\exists c_\varepsilon > 0$ s.t. $\|\hat\theta_n - \theta^*\|_2 < c_\varepsilon \cdot \delta_n$ w.p. $1 - \varepsilon$.

Proof

$$\mathbb{P}(\|\hat\theta_n - \theta^*\|_2 \geq 2^M \delta_n)$$

$$= \sum_{j \geq M+1} \mathbb{P}\left(2^{j-1}\delta_n \leq \|\hat\theta_n - \theta^*\|_2 < 2^j \delta_n \right)$$

$$\|\hat\theta_n - \theta^*\|_2^2 \leq F(\hat\theta_n) - F(\theta^*) \leq \left| (P_n - P)(f_{\hat\theta_n} - f_{\theta^*}) \right|.$$

$$j\text{-th term} \leq \mathbb{P}\left(\|\hat{\theta}_n - \theta^*\|_2 \leq 2^j \delta_n,\ |(\mathbb{P}_n - \mathbb{P})(f_{\theta^*} - f_{\hat{\theta}_n})| \geq 2^{2j-2} \delta_n^2 \right)$$

$$\leq \mathbb{P}\left(\sup_{\substack{\theta \in \Theta \\ \|\theta - \theta^*\| \leq 2^j \delta_n}} |(\mathbb{P}_n - \mathbb{P})(f^* - f_\theta)| \geq 2^{2j-2} \delta_n^2 \right).$$

$$\leq \frac{1}{2^{2j-2} \delta_n^2} \mathbb{E}\left[\sup_{\substack{\theta \in \Theta \\ \|\theta - \theta^*\|_2 \leq 2^j \delta}} |\cdots\cdots| \right]$$

$$\leq \frac{1}{2^{2j-2} \delta_n^2} \cdot \phi_n(2^j \delta_n)$$

$$\leq \frac{4\, \phi_n(\delta_n)}{\delta_n} \cdot 2^{(\alpha-2)j}$$

$$\mathbb{P}\left(\|\hat{\theta}_n - \theta^*\|_2 \geq 2^M \delta_n \right) \leq 4 \cdot \sum_{j=M+1}^{+\infty} 2^{(\alpha-2)j}$$

$$= \frac{4 \cdot 2^{(\alpha-2)M}}{1 - 2^{\alpha-2}}$$

---

e.g. K compact subset of $\mathbb{R}^d$ $\quad \mathcal{F} = \{ f_\theta : \theta \in \Theta \}$

$$|f_{\theta_1}(x) - f_{\theta_2}(x)| \leq M(x) \|\theta_1 - \theta_2\|_2. \qquad \mathbb{E}[M^2(X)] < +\infty.$$

Conclusion. $$\mathbb{E}\left[\sup_{\substack{\theta \in \Theta \\ \|\theta - \theta^*\|_2 \leq u}} |(\mathbb{P}_n - \mathbb{P})(f_\theta - f_{\theta^*})| \right]$$

$$\leq C \sqrt{\frac{\mathbb{E}[F(X)^2]}{n}} \int_0^1 \sqrt{\log\left[\left(1 + \frac{2}{\delta}\right)^d\right]}\, d\delta \leq C \cdot u \cdot \sqrt{\frac{d}{n}}.$$

$$F(x) = M(x) \cdot u. \qquad |f_\theta(x) - f_{\theta^*}(x)| \leq F(x)$$

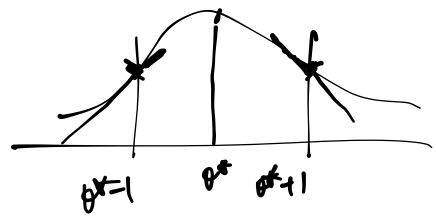
$$\phi_n(\delta_n) = c\cdot\sqrt{\frac{d}{n}}\cdot\delta_n = \delta_n^2$$

$$\delta_n = c\cdot\sqrt{\frac{d}{n}}.$$

$$\|\theta^* - \hat{\theta}_n\|_2 \le c\sqrt{\frac{d}{n}} \quad \text{w.h.p.}$$

---

e.g. $f_\theta(x) = -\mathbb{1}\{x \in [\theta-1, \theta+1]\}$



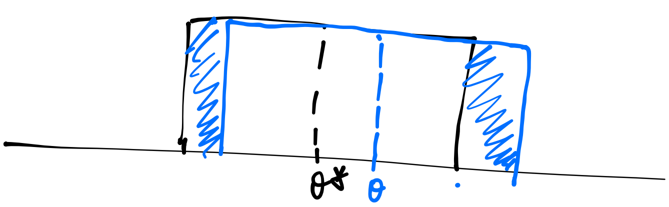
$$F(\theta) = -\mathbb{P}(|X-\theta| \le 1).$$

Assume. $F''(\theta^*) = p'(\theta^*-1) - p'(\theta^*+1) > 0$

$$\mathbb{E}\left[\sup_{|\theta-\theta^*|\le u} \left|(P_n-P)(f_\theta - f_{\theta^*})\right|\right] \le c\sqrt{\frac{p_{max}\cdot u}{n}},$$

$$|f_\theta(x) - f_{\theta^*}(x)| \le \underbrace{\mathbb{1}\{|x-\theta^*+1|\le u\} + \mathbb{1}\{|x-\theta^*-1|\le u\}}_{F(x).}$$

$|\theta^* - \theta| \le u.$



$p(x) \le p_{max} < +\infty$

$$\mathbb{E}[F(X)^2] = \mathbb{P}(|X-\theta^*+1|\le u) + \mathbb{P}(|X-\theta^*-1|\le u).$$
$$\le 4p_{max}\cdot u.$$

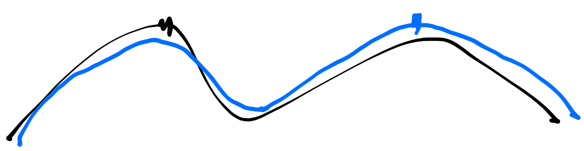
$$\delta_n^2 = \phi_n(\delta_n) = C\cdot\sqrt{\frac{p_{max}\,\delta_n}{n}}.$$

$$\delta_n = C\cdot\left(\frac{p_{max}}{n}\right)^{1/3}.$$

$$n^{1/3}\left(\hat{\theta}_n - \theta^*\right) \xrightarrow{d} \text{Something.}$$

Def. $X_n \xrightarrow{d} X \iff \forall$ bdd, cts f $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$.

$$\boxed{\begin{array}{l}\text{A functional } h \text{ is cts in } L^\infty(K). \text{ if } \forall \varepsilon>0 \\ f\in L^\infty(K),\ \exists\, \delta>0, \text{ s.t.} \\ \|g-f\|_\infty \le \delta \implies |h(g)-h(f)| \le \varepsilon.\end{array}}$$



Argmax cts mapping thm.

$\{F_n(t): t\in T\}$, $\{F(t): t\in T\}$, if

(i) $\forall$ compact K, $\{F_n(t): t\in K\} \xrightarrow{d} \{F(t): t\in K\}$.

(ii) F is cts a.s.

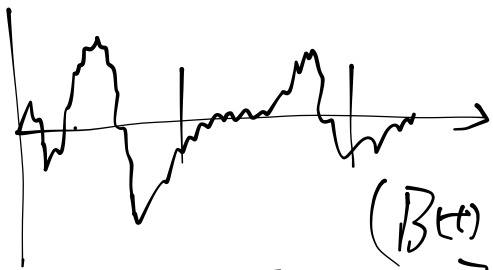
(iii) $\hat{t}_n$ maximizes F_n over T, t uniquely maximizes F over T.

(iv). $t, \{\hat{t}_n\}_{n\ge 1}$ is uniformly tight ($O_p(1)$).

($\hat{t}_n = \sqrt{n}(\hat{\theta}_n - \theta^*)$, $\|\hat{\theta}_n - \theta^*\|_2 \le \frac{1}{\sqrt{n}}$ w.h.p.)

---

e.g. RW and BM. within time $[t, t+\Delta t]$



add indp Gaussian noise $\sim \mathcal{N}(0, \Delta t)$.

$(B(t) : t \ge 0)$ GP. $\mathbb{E}[B(t)] = 0$

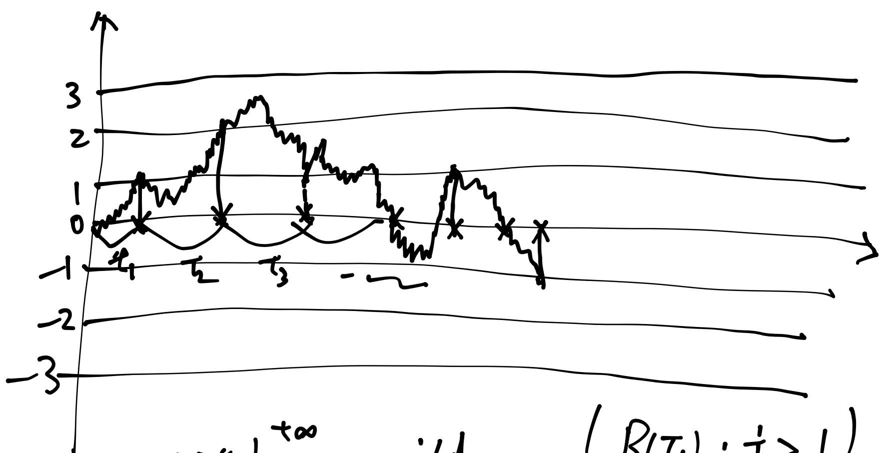
$$\mathbb{E}[B(s) \cdot B(t)] = \min\{s, t\}$$

SRW: $X_{t+1} = X_t + \varepsilon_{t+1}$ $\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{Unif}(\{\pm 1\})$.

Fact. (Donsker). $\forall T > 0$.

$$\left\{ \frac{X_{nt}}{\sqrt{n}} : 0 \le t \le T \right\} \xrightarrow{d} \{ B(t) : t \in [0, T] \}$$

Proof.



$\{\tau_i\}_{i=1}^{+\infty}$ iid. $(B(\tau_i) : i \ge 1)$ is RW.

$$\left\{ \frac{X(nt)}{\sqrt{n}} : t \in [0, T] \right\} \xrightarrow{P} \{ B(t) : t \in [0, T] \}.$$

Apply to [illegible]. Goal:

$$\left\{ n \left(F_n\left(\theta^* + \frac{h}{\sqrt{n}}\right) - F_n(\theta^*) \right) : h \in K \right\} \xrightarrow{d} \text{something.}$$

- Finite-dim convergence.
- "Path is regular"

Lindeberg-Feller CLT

$\forall n,\quad Y_{n1}, Y_{n2}, \dots, Y_{nk_n}$ indep r.v.

$$\sum_{i=1}^{k_n} \operatorname{cov}(Y_{ni}) \to \Sigma$$

$\forall \varepsilon > 0,\quad \sum_{i=1}^{k_n} \mathbb{E}\left[\|Y_{ni}\|_2^2 \, \mathbb{1}\{\|Y_{ni}\|_2 > \varepsilon\} \right] \to 0.$

$$\Downarrow$$

$$\sum_1^n Y_{ni} \xrightarrow{d} N(0, \Sigma)$$

$$\boxed{Y_{ni} = \frac{f_{\theta + \frac{h}{\sqrt{n}}}(X_i)}{\sqrt{n}}}$$

$$\begin{array}{llll}
Y_{11} & & & \\
Y_{21} & Y_{22} & & \\
\vdots & & & \\
Y_{n1} & - & - & Y_{nn}.
\end{array}$$

if dominating function F_n.

Easy to verify

$\mathbb{E}\left[F_n^2 \cdot \mathbb{1}\{F_n > \varepsilon\sqrt{n}\} \right] \to 0 \quad (n \to \infty)$

$\forall \varepsilon > 0.$

Stochastic Equicontinuity.

$$\boxed{\lim_{\eta\to 0^+}\ \limsup_{n\to+\infty}\ \mathbb{E}\left[\sup_{\substack{\|s-t\|\le\eta\\ s,t\in T}} |X_n(s)-X_n(t)|\right]} \to 0.$$

Thm (Arzela–Ascoli)

Finite dim convergence + Stochastic Equicts

$\Rightarrow$ limiting process $\{X(t): t\in T\}$ exists

$\{X_n(t): t\in T\} \xrightarrow{d} \{X(t): t\in T\}$.

K compact subset of $\mathbb{R}^d$ $\quad \mathcal{F} = \{f_\theta : \theta\in\Theta\}$

$|f_{\theta_1}(x) - f_{\theta_2}(x)| \le M(x)\|\theta_1-\theta_2\|_2.$

$\mathbb{E}[M(X)^2] < +\infty.$

$$\hat{h}_n = \sqrt{n}(\hat{\theta}_n - \theta^*)$$

$$\tilde{F}_n(h) := n\cdot\left(F_n\left(\theta^* + \frac{h}{\sqrt{n}}\right) - F_n(\theta^*)\right).$$

$$= \underbrace{n\cdot(P_n - P)\left(f_{\theta^*+h/\sqrt{n}} - f_{\theta^*}\right)}_{A_n(h)} + \underbrace{n\cdot\left(F\left(\theta^*+\frac{h}{\sqrt{n}}\right) - F(\theta^*)\right)}_{B_n(h)}$$

$B_n(h) \to \frac{1}{2} h^T \nabla F(\theta^*) h$ uniformly on any compact set.

$$\mathrm{cov}\left(A_n(h_1),\ A_n(h_2)\right)$$

$$= n \cdot \mathbb{E}\left[\left(f_{\theta^* + \frac{h_1}{\sqrt{n}}}(x) - f_{\theta^*}(x)\right) \cdot \left(f_{\theta^* + \frac{h_2}{\sqrt{n}}}(x) - f_{\theta^*}(x)\right)\right]$$

$$- n \cdot \left(F\left(\theta^* + \frac{h_1}{\sqrt{n}}\right) - F(\theta^*)\right) \cdot \left(F\left(\theta^* + \frac{h_2}{\sqrt{n}}\right) - F(\theta^*)\right)$$

$$\xrightarrow{(\mathrm{DCT})} \mathbb{E}\left[h_1^T \nabla f(\theta^*; x)\ \nabla f(\theta^*; x)^T h_2\right].$$

$$\begin{bmatrix} A_n(h_1) \\ \vdots \\ A_n(h_k) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} (\Sigma^*)^{1/2} z \cdot h_1 \\ \vdots \\ (\Sigma^*)^{1/2} z \cdot h_k \end{bmatrix}$$

$z \sim N(0, I_d)$

$\Sigma^* = \mathrm{cov}\left(\nabla f_{\theta^*}(x)\right).$