STA 3000 Lecture 8

Recap. $\hat{\theta}_n = \arg\min \{ P_n f_\theta \}$

$$\|\hat{\theta}_n - \theta^*\|^2 \leq Pf_{\hat{\theta}_n} - Pf_{\theta^*}$$

$$= (Pf_{\hat{\theta}_n} - P_n f_{\hat{\theta}_n}) + \underbrace{(P_n f_{\hat{\theta}_n} - P_n f_{\theta^*})}_{\leq 0} + (P_n f_{\theta^*} - Pf_{\theta^*})$$

$$\leq (P_n - P)(f_{\theta^*} - f_{\hat{\theta}_n}).$$

Fluctuation could cancel when $\|\hat{\theta}_n - \theta^*\|$ small.

Relate this to

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

Rate of convergence determined by fixed-pt $\delta_n^2 = \phi_n(\delta_n)$

- When $Pf_\theta$ satisfies a different local growth.

  e.g. $Pf_\theta - Pf_{\theta^*} \geq \|\theta - \theta^*\|_2^4$ in singular GMM.

$$\delta_n^4 = \phi_n(\delta_n)$$

- Failure prob is not tight.

Examples.

1. Regular parametric models.

Assume $\left| f_{\theta_1}(x) - f_{\theta_2}(x) \right| \leq M(x) \cdot \|\theta_1 - \theta_2\|_2$.

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

Assume. $P f_\theta - P f_{\theta^*} \geq \|\theta - \theta^*\|^2$.

~~Let $\Theta = \mathbb{B}(0, R)$ (for simplicity).~~

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

Envelope function.

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

$$\leq C \cdot \sqrt{\frac{\mathbb{E}\left[M(X)^2 u^2\right]}{n}}.$$

$$\int_0^1 \sqrt{\log \sup_Q N\left(u \cdot \delta \cdot \|M\|_{L^2(Q)} ; \mathcal{F}_u, \|\cdot\|_{L^2(Q)}\right)}\, d\delta.$$

where $\mathcal{F}_u := \left\{ f_\theta - f_{\theta^*} : \|\theta - \theta^*\|_2 \leq u,\ \theta \in \Theta \right\}$.

$$\left( \|\theta_1 - \theta_2\|_2 \leq \varepsilon \Rightarrow \|f_{\theta_1} - f_{\theta_2}\|_{L^2(Q)} \leq \varepsilon \cdot \|M\|_{L^2(Q)} \right).$$

$$N\left(u\delta \|M\|_{L^2(Q)} ; \mathcal{F}_u, \|\cdot\|_{L^2(Q)}\right)$$
$$\leq N\left(u\delta ; \{\theta \in \Theta : \|\theta - \theta^*\|_2 \leq u\}, \|\cdot\|_2\right)$$
$$\leq \left(\frac{c}{\delta}\right)^d .$$

So we conclude that

$$\int_0^1 \sqrt{\log \sup \cdots\cdots} \, d\delta \leq c'\sqrt{d}.$$

Conclusion: $(*)$ holds true w/

$$\phi_n(u) = c\sqrt{\frac{d}{n}} \cdot \|M\|_{L^2(P)} \cdot u$$

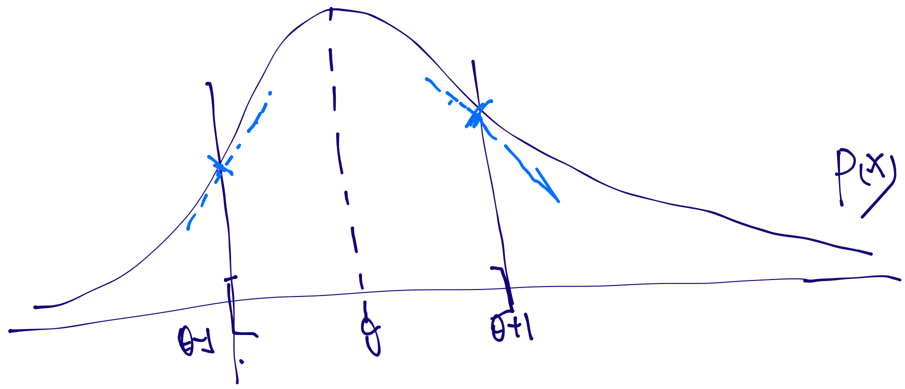
linearly growing in $u$ (slower than quadratic).

$$\delta_n = c\sqrt{\frac{d}{n}} \cdot \|M\|_{L^2(P)}.$$

$$\|\hat{\theta}_n - \theta^*\|_2 \lesssim \delta_n \quad \text{w.h.p.}$$

- In general, the bound might have bad dimension dependence from $\|M\|_{L^2(P)}$.

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



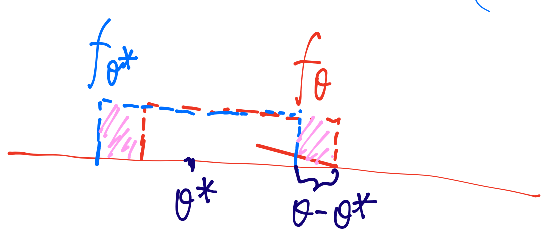

• $Pf_\theta = \mathbb{P}(x \in [\theta-1, \theta+1])$

Assuming $p(x)$ "nice ~~enough~~"

$$\frac{d^2}{d\theta^2} Pf_\theta \Big|_{\theta=\theta^*} = p'(\theta^*-1) - p'(\theta^*+1) > 0.$$

(Assume).

• Envelope function




$G_2(x)$

$$:= \mathbb{1}\{x \in [\theta^*+1-u, \theta^*+1+u] \text{ on } [\theta^*-1-u, \theta^*-1+u]\}$$

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

$$\le C\cdot\sqrt{\frac{\mathbb{E}[G(X)^2]}{n}} \cdot \int_0^1 \sqrt{\log \sup_Q N(\delta\cdot\|G\|_{L^2(Q)}; \mathcal{F}_u, \|\cdot\|_{L^2(Q)})}\, d\delta.$$

(i) - $\mathbb{E}[G(X)^2] = \mathbb{P}(X \in \theta^*+1 \pm u \text{ or } \theta^*-1 \pm u)$.

(Assume that $p(x) \le p_{max}$)

$$\le 4 p_{max} u.$$

(ii). $\mathcal{F}_u = \{ f_\theta - f_{\theta^*} : |\theta-\theta^*| \le u \}$

$VC(\mathcal{F}_u) \le 8$.

$$\log \sup_Q N(\delta\cdot\|G\|_{L^2(Q)}; \mathcal{F}_u, \|\cdot\|_{L^2(Q)}) \le c' \log(1/\delta).$$

$(*)$ is satisfied with

$$\phi_n(u) = C\cdot\sqrt{\frac{u p_{max}}{n}}. \quad = u^2(p'(\theta^*-1) - p'(\theta^*+1))$$

(growth slower than quadratic).

By thm, for $\delta_n = C\cdot\left(\frac{p_{max}}{n(p'(\theta^*-1)-p'(\theta^*+1))}\right)^{1/3}$. $\|\hat{\theta}_n - \theta^*\| \lesssim \delta_n$ w.h.p.

Asymptotic distributions.

- Classical method.

$$\hat{\theta}_n = \arg\min_{\theta \in \mathbb{R}^d} F_n(\theta).$$

$$0 = \nabla F_n(\hat{\theta}_n)$$

$$= \underbrace{\nabla F_n(\theta^*)}_{\text{sum of iid r.v.}} + \nabla^2 F_n(\theta^*)(\hat{\theta}_n - \theta^*) + \int_0^1 \nabla^3 F_n(\gamma\theta^* + (1-\theta)\hat{\theta}_n)[\hat{\theta}_n - \theta^*, \hat{\theta}_n - \theta^*] d\gamma.$$

$\nabla^2 F_n(\theta^*) \xrightarrow{p} \nabla^2 F(\theta^*)$.

$$\sqrt{n} \cdot \nabla F_n(\theta^*) \xrightarrow{d} \mathcal{N}(0, \Sigma^*)$$

Assume $|||\nabla^3_\theta f_\theta(x)|||_{\text{tsr}} \leq M(x)$ uniformly for $\theta$.

$$\leq \|\hat{\theta}_n - \theta^*\|_2^2 \cdot \frac{1}{n}\sum_{i=1}^n M(X_i)$$

Thm: $\Theta$ open set $\theta^* \in \Theta$

$\theta \mapsto f_\theta(x)$ 3rd cts differentiable

with $|||\nabla^3 f_\theta(x)|||_{\text{tsr}} \leq M(x)$ $\forall \theta \in \mathbb{B}(\theta^*, \varepsilon_0)$

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

Assume $\Sigma^* = \text{cov}(\nabla_\theta f_{\theta^*}(X))$ well defined

then $\sqrt{n}(\hat{\theta}_n - \theta^*) \xrightarrow{d} \mathcal{N}(0, \nabla^2 F(\theta^*)^{-1} \Sigma^* \nabla^2 F(\theta^*)^{-1})$.

eg. MLE $\quad f_\theta(x) = -\log p_\theta(x)$.

$$\Sigma^* = \operatorname{cov}\left(\nabla \log p_{\theta^*}(X)\right) = I(\theta^*).$$

$$H^* = \nabla^2 F(\theta^*) = I(\theta^*).$$

$$\sqrt{n}\left(\hat{\theta}_n - \theta^*\right) \xrightarrow{d} N\left(0, I(\theta^*)^{-1}\right).$$ matches (Bayesian) CR.

$$\boxed{\underbrace{\sqrt{n}\, \nabla^2 F(\theta^*)\cdot\left(\hat{\theta}_n - \theta^*\right)}_{\sqrt{n}\cdot g} \xrightarrow{d} N(0, \Sigma^*).}$$

$$\sqrt{n}\left(\hat{\theta}_n - \theta^*\right) = \nabla^2 F(\theta^*)^{-1} g$$

$$\operatorname{cov} = \nabla^2 F(\theta^*)^{-1} \cdot \mathbb{E}\left[g g^T\right] \nabla^2 F(\theta^*)^{-1}.$$

Drawbacks: (1) requires a lot of differentiable

eg. $X_1, X_2 \cdots X_n \overset{iid}{\sim} p(x)$

$$f_\theta(x) = |\theta - x|.$$

(2) Cannot tackle non-standard asymptotics.

---

Process convergence.

$X_n \xrightarrow{d} X \iff \forall$ bdd cts function $f$

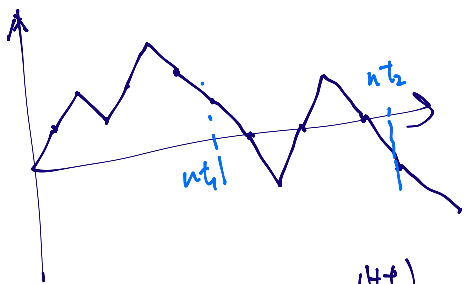
$$\mathbb{E}\left[f(X_n)\right] \to \mathbb{E}\left[f(X)\right].$$

What does it mean to say $f$ is cts at $x$?

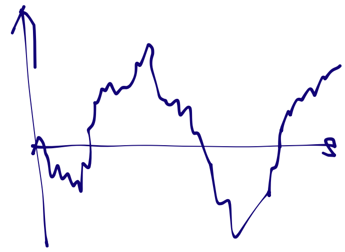
$\forall \varepsilon > 0$, $\exists \delta > 0$, s.t. whenever $\|x - y\|_\infty \leq \delta$

we have $|f(x) - f(y)| < \varepsilon$

e.g. Random walk and BM.




$\xrightarrow{d}$



$X_{t+1} = X_t + \varepsilon_{t+1}$ $(\forall t)$

$(\varepsilon_t \overset{iid}{\sim} \text{Rade})$.

$B(t) \sim GP$

$B(t) = 0 \quad \forall t \geq 0$

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

Claim: $\left( \frac{X_{nt}}{\sqrt{n}} : 0 \leq t \leq T \right) \xrightarrow{d} \left( B_t : 0 \leq t \leq T \right)$.

$(t_1 < t_2)$ $\mathbb{E}[X_{nt_1} \cdot X_{nt_2}] = \mathbb{E}[X_{nt_1}^2] = nt_1$

---

Thm. (Stochastic analogue of Arzela-Ascoli).

$(X_n(t) : t \in T)$ for $n = 1, 3 \cdots$

$(X(t) : t \in T)$

If (i). $\forall$ Finite collection $\{t_1, t_2, \dots, t_k\} \subseteq T$.

$(X_n(t_1), X_n(t_2), \dots X_n(t_k)) \xrightarrow{d} (X(t_1), \dots, X(t_k))$.

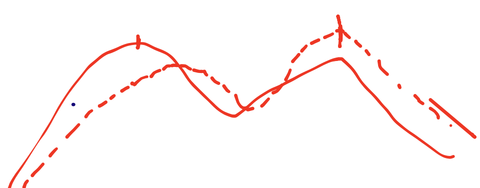
(ii). Stochastic Equicontinuity.

$$\limsup_{n\to+\infty} \mathbb{E}\Big[\underbrace{\sup_{\substack{\|s-t\|<\eta \\ s,t\in T}} |X_n(s)-X_n(t)|}_{\text{Asymptotic modulus of continuity}}\Big] \to 0 \quad (\text{as } \eta\to 0).$$

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

How to convert process convergence to desired result.

- $\forall$ bdd cts functional $g$ $\quad$ $\underline{g(X_n) \xrightarrow{d} g(X)}$
- argmax 's not continuous.



"~~argmax~~ convergence thm"

(i) $\forall$ compact subset $K$, $\quad (X_n(t) : t\in K) \xrightarrow{d} (X(t) : t\in K)$.

(ii) $(X(t) : t\in T)$ 's continuous

(iii) $\hat{t}_n \in \operatorname{argmax}\{X_n(t)\}$, $\quad t$ unique maximizer of $X(t)$

(iv) $t$ and $(\hat{t}_n)_{n\geq 1}$ are uniformly tight.

uniform control on tail

$\forall \varepsilon>0, \exists K$ st $\forall n$ $\quad \mathbb{P}(\|\hat{t}_n\| \geq K) \leq \varepsilon.$

Back to M-estimators.

Suppose $\left(F_n\left(\theta^* + h/\sqrt{n}\right) - F_n(\theta^*) : h \in K\right) \xrightarrow{d}$ Something

then $\hat{h}_n = \text{argmax}\{\cdots\}$

$= \sqrt{n}\left(\hat{\theta}_n - \theta^*\right) \xrightarrow{d}$ maximizer of something.

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

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

$B_n(h)$ deterministic (assuming $F$ $2^{nd}$ order diff at $\theta^*$).

$\longrightarrow \frac{1}{2} h^T \nabla^2 F(\theta^*) h$

Uniformly over any compact set $K$.

$A_n(h)$ : (i) Finite - dim

(ii) Stoch Equi-cts.

(i). $\text{cov}\left(A_n(h_1), A_n(h_2)\right)$

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

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

$$\xrightarrow{n\to+\infty} \mathbb{E}\left[ h_1^T \nabla f_{\theta^*}(X) \cdot \nabla f_{\theta^*}(X)^T h_2 \right].$$

$$= h_1^T \Sigma^* h_2.$$

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

$$\begin{bmatrix} A_n(h_1) \\ \vdots \\ A_n(h_K) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} h_1^T (\Sigma^*)^{1/2} Z \\ h_2^T (\Sigma^*)^{1/2} Z \\ \vdots \\ h_K^T (\Sigma^*)^{1/2} Z \end{bmatrix}$$

$$\left(A_n(h_1), \cancel{A_n(h_2)}\right) \xrightarrow{d} \begin{bmatrix} h_1^T (\Sigma^*)^{1/2} Z \\ h_2^T (\Sigma^*)^{1/2} Z \end{bmatrix}$$

where $Z \sim \mathcal{N}(0, I)$

Hope :— $\left(A_n(h) : h \in \Omega\right) \xrightarrow{d} \left(h^T (\Sigma^*)^{1/2} Z : h \in \Omega\right)$

for $Z \sim \mathcal{N}(0, I)$.

Suppose hope is true,

$$\sqrt{n}\left(\hat{\theta}_n - \theta^*\right) = \hat{h}_n \to \underset{h \in \mathbb{R}^d}{\mathrm{argmin}} \left\{ h^T (\Sigma^*)^{1/2} Z + \frac{1}{2} h^T \nabla^2 F(\theta^*) h \right\}$$

$$= \nabla^2 F(\theta^*)^{-1} \cdot (\Sigma^*)^{1/2} Z.$$

Thm :— $\hat{\theta}_n = \mathrm{argmin}\left(F_n(\theta)\right)$ $\quad \Sigma^* = \mathrm{cov}\left(\nabla f_{\theta^*}(X)\right)$ finite

(i) $f_\theta(x)$ differentiable at $\theta^*$ (a.s.).

(ii). $\left| f_{\theta_1}(x) - f_{\theta_2}(x) \right| \leq M(x) \|\theta_1 - \theta_2\|_2.$

(iii) $F$ twice cts diff at $\theta^*$ $\quad H^* = \nabla^2 F(\theta^*) \succ 0.$

(iv) $\hat{\theta}_n \xrightarrow{P} \theta^*$

Then $\sqrt{n}\left(\hat{\theta}_n - \theta^*\right) \xrightarrow{d} \mathcal{N}\left(0, (H^*)^{-1}\Sigma^*(H^*)^{-1}\right)$.

Proof. It remains to verify stoch Equicts.

$$\mathcal{G}_\eta := \left\{ f\left(\theta^* + h_1/\sqrt{n}\right) - f\left(\theta^* + h_2/\sqrt{n}\right) : \|h_1 - h_2\|_2 \leq \eta, \; h_1, h_2 \in K \right\}$$

$$n \cdot \mathbb{E}\left[\sup_{g \in \mathcal{G}_\eta} \left|(P_n - P) g\right|\right]$$

$$\leq C \cdot \sqrt{n} \cdot \|G\|_{L^2(P)} \int_0^1 \sqrt{\log N_{[\,]}\left(\delta \cdot \|G\|_{L^2(P)}; \mathcal{G}_\eta, L^2(P)\right)}\, d\delta.$$

Choice of envelope:

$$\left| f\left(\theta^* + \frac{h_1}{\sqrt{n}}\right) - f\left(\theta^* + \frac{h_2}{\sqrt{n}}\right)\right| \leq M(x) \cdot \frac{\eta}{\sqrt{n}}.$$

$$\leq C \cdot \eta \cdot \|M\|_{L^2(P)} \cdot \int_0^1 \sqrt{\log N_{[\,]}\left(\frac{\delta \cdot \eta}{\sqrt{n}} \|M\|_{L^2(P)}; \mathcal{G}_\eta, L^2(P)\right)}\, d\delta.$$

Construction of brackets

Let $\theta_1, \theta_2, \dots, \theta_N$ to be $\varepsilon$-covering of $K$.

Bracket $\quad l_i(x) = f_{\theta_i}(x) - \varepsilon \cdot M(x)$

$\quad u_i(x) = f_{\theta_i}(x) + \varepsilon \cdot M(x)$.

$\forall f_\theta$, $\theta \in \mathbb{B}(\theta^*; \varepsilon)$ $f_\theta \in [l_i, u_i]$.

Construction of brackets

Let $h_1, h_2, \cdots, h_K$ to be $\varepsilon$-covering of $K$.

Bracket $l_{ij}(x) = f_{\theta^* + h_i/\sqrt{n}}(x) - f_{\theta^* + h_j/\sqrt{n}}(x) + \frac{2\varepsilon}{\sqrt{n}} M(x)$

$u_{ij}(x) = f_{\theta^* + h_i/\sqrt{n}}(x) - f_{\theta^* + h_j/\sqrt{n}}(x) - \frac{2\varepsilon}{\sqrt{n}} M(x)$.

$$\log N_{[\,]}\left(\frac{\delta\eta}{\sqrt{n}} \cdot \|M\|_{L^2(P)}; \mathcal{G}_\eta, L^2(P)\right)$$

$$\leq \log N\left(\frac{\delta\eta}{4}; K, \|\cdot\|_2\right).$$

$$\leq C \cdot d \cdot \log\left(\frac{\operatorname{diam}(K)}{\delta\eta}\right).$$

Substitute back to integral.

$$n \cdot \mathbb{E}\left[\sup_{g \in \mathcal{G}_\eta} \left|(P_n - P) g\right|\right]$$

$$\leq C \cdot \eta \cdot \|M\|_{L^2(P)} \cdot \int_0^1 \sqrt{d\left(1 + \log\frac{1}{\delta\eta}\right)}\, d\delta.$$

$$\lesssim \|M\|_{L^2(P)} \cdot \eta \sqrt{d \cdot \log(1/\eta)} \longrightarrow 0.$$

QED.