STA3000F Lecture 11

$\Psi_n(\theta) = \frac{1}{n}\sum_1^n \psi(\theta; X_i)$. $\Psi(\theta) := \mathbb{E}[\psi(\theta; X)]$ $\Psi(\theta^*) = 0$

$\Sigma^* = \mathbb{E}[\psi(\theta^*; X)\psi(\theta^*; X)^T]$. $\Psi_n(\hat{\theta}_n) = 0$.

Thm. Under suitable conditions

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

$|||\nabla^2\psi(\theta; x)|||_{tsr} \leq L(x)$ in a neighborhood of $\theta^*$ $\mathbb{B}(\theta^*, \varepsilon_0)$.

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

$$0 = \Psi_n(\hat{\theta}_n) = \Psi_n(\theta^*) + \nabla\Psi_n(\theta^*)\cdot\boxed{(\hat{\theta}_n - \theta^*)} + \underbrace{\frac{1}{2}\int_0^1 \nabla^2\Psi_n(\gamma\theta^* + (1-\gamma)\hat{\theta}_n)[\hat{\theta}_n - \theta^*, \hat{\theta}_n - \theta^*]\,d\gamma}_{R_n}$$

- $\sqrt{n}\cdot\Psi_n(\theta^*) \xrightarrow{d} \mathcal{N}(0, \Sigma^*)$.
- $\|\nabla\Psi_n(\theta^*) - \nabla\Psi(\theta^*)\|_{op} = o_p(1)$.
- $\|R_n\|_2 \leq \frac{1}{2}\cdot\sup_{\gamma\in[0,1]} |||\nabla^2\Psi_n(\gamma\theta^* + (1-\gamma)\hat{\theta}_n)|||_{tsr}\cdot\|\hat{\theta}_n - \theta^*\|_2^2$.

$\leq \frac{1}{2}\left(\sum_1^n L(X_i)\right)\cdot\|\hat{\theta}_n - \theta^*\|_2^2$ when $\|\hat{\theta}_n - \theta^*\|_2 \leq \varepsilon_0$.

$\xrightarrow{P} \mathbb{E}[L(X)]$. w.p. $\to 1$.

$$\sqrt{n}(\hat{\theta}_n - \theta^*) = \sqrt{n}\nabla\Psi_n(\theta^*)^{-1}\Psi_n(\theta^*) + \sqrt{n}\cdot\nabla\Psi_n(\theta^*)^{-1}R_n$$

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

---

Asymptotic normality from another perspective.

$\hat{\theta}_n = \arg\max F_n(\theta)$, $\theta^* = \arg\max F(\theta)$.

$$F_n\left(\theta^* + \frac{h}{\sqrt{n}}\right) - F_n(\theta^*)$$

$$\approx \underbrace{\frac{h^T}{\sqrt{n}} \nabla F_n(\theta^*)} + \frac{1}{2n} h^T \nabla^2 F(\theta^*) h$$

where $Z \sim N(0, I_d)$. $\Sigma^* := \mathrm{cov}(\nabla f(\theta^*; X))$.

$$\sqrt{n}\, \nabla F_n(\theta^*) \approx (\Sigma^*)^{1/2} Z$$

$$n\left(F_n\left(\theta^* + \frac{h}{\sqrt{n}}\right) - F_n(\theta^*)\right) \xrightarrow{d} N\left(\frac{1}{2} h^T \nabla^2 F(\theta^*) h,\ h^T \Sigma^* h\right)$$

(Lindeberg-Feller CLT).

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \xrightarrow{d} \operatorname{argmin}\left\{\underline{\frac{1}{2} h^T \nabla^2 F(\theta^*) h + h^T (\Sigma^*)^{1/2} Z}\right\}.(?).$$

$$\left\{ n\left(F_n\left(\theta^* + \frac{h}{\sqrt{n}}\right) - F_n(\theta^*)\right) : h \in \Omega \right\} \xrightarrow{d} ?$$ $L^\infty$-norm.

---

Motivation. non-asymptotic rate of convergence.

$$F(\hat{\theta}_n) - F(\theta^*) = \underbrace{\boxed{F(\hat{\theta}_n) - F_n(\hat{\theta}_n)}}_{\le \sup_{\theta \in ?} |F(\theta) - F_n(\theta)|} + \underbrace{F_n(\hat{\theta}_n) - F_n(\theta^*)}_{\le 0} + \underbrace{F_n(\theta^*) - F(\theta^*)}_{\text{LLN / concentration ineq.}}$$

local neighbor of $\theta^*$. $\|\hat{\theta}_n - \theta^*\|_2 \le r_n$

---

An introduction to empirical process.

Notations. $P_n f := \frac{1}{n} \sum_1^n f(X_i)$

$P f := \mathbb{E}[f(X)]$.

$$\boxed{\sup_{f \in \mathcal{F}} |P_n f - P f|.}$$

→ M-estimation

→ Unif CLT.

Key tools: discretization, symmetrization & chaining.

Define $\mathcal{R}_n(\mathcal{F}) := \mathbb{E}\left[\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_1^n \varepsilon_i f(X_i)\right|\right]$ (Rademacher complexity).

where $\varepsilon_i \overset{iid}{\sim} \mathrm{Unif}(\{-1,+1\})$ (Rademacher r.v.).

Thm. $\mathbb{E}\left[\sup_{f\in\mathcal{F}} |P_n f - Pf|\right] \leq 2\mathcal{R}_n(\mathcal{F})$.

Proof: $X_1' \cdots X_n' \overset{iid}{\sim} P$ (indep of $X_i$'s).

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} |P_n f - Pf|\right] = \mathbb{E} \sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_1^n f(X_i) - \frac{1}{n}\sum_1^n \mathbb{E}[f(X_i')]\right|.$$

$$\overset{(\text{Jensen})}{\leq} \mathbb{E} \sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_1^n (f(X_i) - f(X_i'))\right|.$$

$$\{f(X_i)-f(X_i')\}_{i=1}^n \overset{d}{=} \{\varepsilon_i(f(X_i)-f(X_i'))\}_{i=1}^n$$

$$= \mathbb{E}\left[\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_1^n \varepsilon_i (f(X_i) - f(X_i'))\right|\right]$$

$$\overset{(\text{Jensen})}{\leq} 2\,\mathbb{E}\left[\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_1^n \varepsilon_i f(X_i)\right|\right] = \mathcal{R}_n(\mathcal{F}).$$

---

Symmetrization allows condition on $(X_i)_{i=1}^n$.

For $A \subseteq \mathbb{R}^n$, need to bound

$$R_n(A) := \mathbb{E}\left[\sup_{a\in A} \frac{1}{n}\sum_1^n \varepsilon_i a_i\right]$$

$A = \{(f(x_1), f(x_2) \cdots f(x_n)) \mid f \in \mathcal{F}\}$

How to bound $R_n(A)$? Finite $A$.

- A weak bound. $R_n(A) \le \sum_{a\in A} \mathbb{E}\left[\left|\frac{1}{n}\sum_1^n \varepsilon_i a_i\right|\right] \le \frac{|A|}{\sqrt{n}} \sqrt{\sum_1^n a_i^2/n}$

- Another trick. $R_n(A)^2 \le \mathbb{E}\left[\sup_{a\in A}\left(\frac{1}{n}\sum_1^n \varepsilon_i a_i\right)^2\right]$

$$\le \sum_{a\in A} \mathbb{E}\left[\left(\frac{1}{n}\sum_1^n \varepsilon_i a_i\right)^2\right] \le \frac{|A|}{n}\cdot\frac{\sum a_i^2}{n}.$$

$$R_n(A) \lesssim \frac{\sqrt{|A|}}{\sqrt{n}}.$$

- Exponentiate it. $(\forall \lambda > 0)$.

$$R_n(A) = \mathbb{E}\left[\frac{1}{\lambda}\log\left(\sup_{a\in A} \exp\left(\frac{\lambda}{n}\sum_{i=1}^n \varepsilon_i a_i\right)\right)\right].$$

$$\text{(Jensen)} \le \frac{1}{\lambda}\log \mathbb{E}\left[\exp\left(\max_{a\in A}\frac{\lambda}{n}\sum_{i=1}^n \varepsilon_i a_i\right)\right]$$

$$\text{(union bound)} \le \frac{1}{\lambda}\log\left(\sum_{a\in A}\mathbb{E}\left[\exp\left(\frac{\lambda}{n}\sum_{i=1}^n \varepsilon_i a_i\right)\right]\right).$$

---

$$\mathbb{E}\left[\exp\left(\frac{\lambda}{n}\sum_1^n \varepsilon_i a_i\right)\right] = \prod_{i=1}^n\left(\frac{1}{2}e^{\frac{\lambda}{n}a_i} + \frac{1}{2}e^{-\frac{\lambda}{n}a_i}\right)$$

$\left(\frac{1}{2}(e^x + e^{-x}) \le e^{\frac{1}{2}x^2}\right)$

$$\le \exp\left(\frac{1}{2}\sum_{i=1}^n \frac{\lambda^2}{n^2}a_i^2\right)$$

$\left(\|a\|_n^2 := \frac{1}{n}\sum_{i=1}^n a_i^2\right)$

$$= \exp\left(\frac{\lambda^2}{2n}\|a\|_n^2\right)$$

---

$$R_n(A) \le \frac{1}{\lambda}\cdot\log\left(\sum_{a\in A}\exp\left(\frac{\lambda^2}{2n}\|a\|_n^2\right)\right)$$

$$\le \frac{1}{\lambda}\cdot\log\left(|A|\cdot\max_a \exp\left(\frac{\lambda^2}{2n}\|a\|_n^2\right)\right)$$

$$\leq \frac{\log|A|}{\lambda} + \frac{\lambda}{2n} \cdot \max_a \left(\|a\|_n^2\right)$$

$$\left(\lambda = \sqrt{2n \log|A| \cdot \max_a \|a\|_n^2}\right)$$

$$R_n(A) \leq \sqrt{\frac{2\log|A|}{n}} \cdot \max_a \|a\|_n$$

---

Going from finite to cts. "chaining"

$$\left(\{a^{(j)}\}_{j=1}^N\right).$$

Naïve approach. $N := N(\delta)$ min-$\delta$ covering # of $A$ under $\|\cdot\|_n$.

$\forall a \in A$, $\pi(a) :=$ closest point to "a" in the covering.

$$R_n(A) \leq \mathbb{E}\left[\max_{j\in[N]} \frac{1}{n}\varepsilon^T a^{(j)}\right] + \mathbb{E}\left[\sup_{a\in A} \frac{1}{n}\varepsilon^T(a-\pi(a))\right].$$

$$\leq \sqrt{\frac{2\log N(\delta)}{n}} \cdot \max_a \|a\|_n + \delta$$

$$\begin{aligned} &|\varepsilon^T(a-\pi(a))| \\ &\leq \|\varepsilon\|_2 \cdot \|a-\pi(a)\|_2 \\ &\leq \sqrt{n}\cdot\sqrt{n}\cdot\delta \end{aligned}$$

**Thm (chaining).** $(0 \in A)$.

$$R_n(A) \leq \frac{C}{\sqrt{n}} \int_0^{+\infty} \sqrt{\log N(\delta)}\, d\delta$$

for some universal constant $C > 0$. $(C \simeq 12)$

---

Proof: Given $A$, and $m > 0$. $D := \max_{a\in A} \|a\|_n$.

Lee $A_m$ be $D/2^m$-min-covering of $A$

$|A_m| = N(D/2^m)$. $\quad A_0 = \{0\}$.

$\forall a \in A$, $m \in \mathbb{N}_+$, $\quad \pi_m(a) :=$ best approx of "$a$" within the set $A_m$.

$$\frac{1}{n}\varepsilon^T a = \sum_{m=0}^{+\infty} \frac{1}{n}\varepsilon^T(\pi_{m+1}(a) - \pi_m(a)) \quad \left(\pi_m(a) \xrightarrow{m\to+\infty} a\right)$$

---

$$\mathbb{E}\left[\max_{a\in A} \frac{1}{n}\varepsilon^T(\pi_m(a) - \pi_{m+1}(a))\right]$$

$$\leq \underbrace{\text{Some dimeter}} \cdot \sqrt{\frac{2\log(|A_m|\cdot|A_{m+1}|)}{n}}$$

$a \in A$

$(\pi_m(a), \pi_{m+1}(a)) \in \underbrace{A_m \times A_{m+1}}_{\text{Finite.}}$

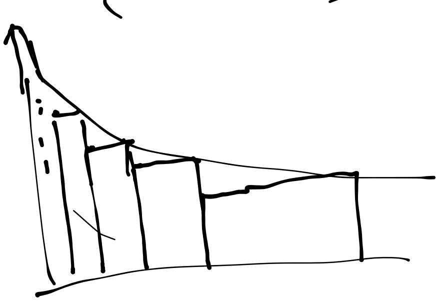
$$\text{Some dimeter} := \max_{a\in A} \|\pi_m(a) - \pi_{m+1}(a)\|_n.$$

$$\leq \max_{a\in A} \|\pi_m(a) - a\|_n + \max_{a\in A} \|\pi_{m+1}(a) - a\|_n$$

$$\leq \frac{3}{2^m}\cdot D.$$

$$\mathbb{E}\left[\max_{a\in A} \frac{1}{n}\varepsilon^T a\right] \leq \sum_{m=0}^{+\infty} \mathbb{E}\left[\max_a \frac{1}{n}\varepsilon^T(\pi_{m+1}(a) - \pi_m(a))\right]$$

$$\leq 6\cdot\sum_{m=0}^{+\infty} \frac{D}{2^m}\cdot\sqrt{\frac{\log N(D/2^{m+1})}{n}}$$



$$\leq 12\int_0^{+\infty}\sqrt{\frac{\log N(\delta)}{n}}\, d\delta.$$

Thm (main) $\quad |f(x)| \leq F(x) \quad (\forall f \in \mathcal{F})$.

$$\mathbb{E} \sup_{f \in \mathcal{F}} |P_n f - Pf| \leq C \cdot \sqrt{\frac{\mathbb{E}[F(X)^2]}{n}} \cdot \int_0^1 \sqrt{\log \sup_Q N(\delta \cdot \|F\|_{L^2(Q)}; \mathcal{F}, L^2(Q))}\, d\delta$$

where $N(t; \mathcal{F}, L^2(Q)) :=$ minimal covering # of $\mathcal{F}$ under $L^2(Q)$.

---

Proof: Conditioned on $(X_i)_{i=1}^n$

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} \left|\frac{1}{n}\sum_1^n \varepsilon_i f(X_i)\right| \,\middle|\, (X_i)_{i=1}^n\right]$$

$\left(P_n F^2 = \frac{1}{n}\sum_1^n F(X_i)^2\right)$

$$\leq \frac{12}{\sqrt{n}} \int_0^{\sqrt{P_n F^2}} \sqrt{\log N(\delta; \mathcal{F}, L^2(P_n))}\, d\delta.$$

$$\leq 12 \sqrt{\frac{P_n F^2}{n}} \cdot \int_0^1 \sqrt{\log N(\delta \cdot \|F\|_n, \mathcal{F}, L^2(P_n))}\, d\delta.$$

$$\leq \int_0^1 \sup_Q \sqrt{\log N(\delta \|F\|_{L^2(Q)}; \mathcal{F}, L^2(Q))}\, d\delta$$

$$\mathbb{E}\left[\sqrt{\frac{P_n F^2}{n}}\right] \underset{(C\text{-}S.)}{\leq} \sqrt{\frac{\mathbb{E}[F^2(X)]}{n}}.$$