STA 3000F Lecture 15

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

. $\sqrt{n}$ -rate of convergence

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

$$H^* = \nabla^2 F(\theta^*), \quad \Sigma^* = \mathrm{cov}_{\theta^*}(\nabla f_{\theta^*}(X))$$

MLE. (Le Cam).

Quadratic Mean Differentiability. (QMD)

Def. $$\int \left(\sqrt{p_\theta} - \sqrt{p_{\theta^*}} - \frac{1}{2}(\theta-\theta^*)^T \dot{\ell}_{\theta^*}(x) \cdot \sqrt{p_{\theta^*}}\right)^2 d\mu(x)$$
$$= o(\|\theta - \theta^*\|_2^2).$$

If $p_\theta(x)$ is differentiable at θ^*, QMD

then $\dot{\ell}_{\theta^*} = \nabla \log p_{\theta^*}(x)$.

Intuition $H^* = \Sigma^*$, don't need 2nd order diff

$$I(\theta^*) = \mathrm{cov}(\dot{\ell}_{\theta^*}(X)).$$

Thm $\sqrt{n}(\hat{\theta}_n - \theta^*) \xrightarrow{d} N(0, I(\theta^*)^{-1})$.

QMD + (*). $(\mathbb{E}[\dot{\ell}_{\theta^*}(X)] = 0)$

e.g. $f_\theta(x) = -\mathbb{1}\{|x-\theta| \le 1\}$

$\hat{\theta}_n := \arg\min \frac{1}{n}\sum_{1}^{n} f_\theta(X_i)$ $\qquad \|\hat{\theta}_n - \theta^*\|_2 = O_p(n^{-1/3})$.

$\theta^* := \arg\min \mathbb{E}[f_\theta(X)]$.

$$\tilde{F}_n(h) = \underbrace{n^{2/3}(P_n - P)\left(f_{\theta^* + n^{-1/3}h} - f_{\theta^*}\right)}_{A_n(h)} + \underbrace{n^{2/3}\left(P(\theta^* + hn^{-1/3}) - P(\theta^*)\right)}_{C_n(h)}$$

$\hat{h}_n := \arg\min \tilde{F}_n(h)$

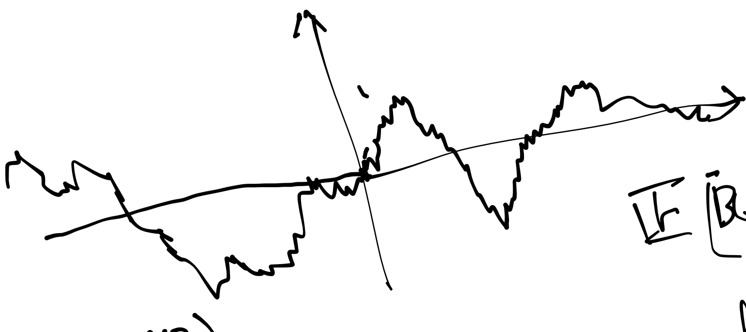
$$C_n(h) \to \frac{1}{2}h^2 \underbrace{\left(p'(\theta^*-1) - p'(\theta^*+1)\right)}_{\ge 0}$$

$\forall h_1, h_2 \in \mathbb{R}$

$$\mathbb{E}[A_n(h_1) \cdot A_n(h_2)] = n^{1/3} \cdot \operatorname{cov}\Big(\mathbb{1}\{|X-\theta^* - n^{-1/3}h_1| \le 1\} - \mathbb{1}\{|X-\theta^*| \le 1\},$$
$$\mathbb{1}\{|X-\theta^* - n^{-1/3}h_2| \le 1\} - \mathbb{1}\{|X-\theta^*| \le 1\}\Big)$$

$$\longrightarrow \left(p(\theta^*-1) + p(\theta^*+1)\right) \cdot \min(|h_1|, |h_2|) \cdot \mathbb{1}\{h_1 \cdot h_2 > 0\}$$

p: density of X



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

(Chernoff?)

$$n^{1/3}(\hat{\theta}_n - \theta^*) \xrightarrow{d} \arg\max_{h \in \mathbb{R}} \left\{ \sqrt{p(\theta^*-1) + p(\theta^*+1)}\, B(h) - \frac{1}{2}h^2\left(p'(\theta^*-1) - p'(\theta^*+1)\right) \right\}$$

$$\mathcal{G}_\eta := \{ f_{\theta^* + h_1 n^{-1/3}} - f_{\theta^* + h_2 n^{-1/3}} : h_1, h_2 \in K, |h_1 - h_2| \le \eta \}$$

$$\mathbb{E}\left[\sup_{g \in \mathcal{G}_\eta} \left| n^{2/3} (\mathbb{P}_n - \mathbb{P}) g \right| \right]$$

Roughly speaking.

$$\int_0 \sqrt{\log N_{[\,]}(\delta ; \mathcal{F}, L^2(\mathbb{P}))}\, d\delta < +\infty \implies \text{stochastic equicontinuity}$$

("Donsker").

Bayesian posterior. $X_1\ X_2 \cdots X_n \overset{iid}{\sim} P_{\theta^*}$

Prior π

$$\Pi_n(\theta | X_1^n) = \frac{\pi(\theta) \cdot \prod_{i=1}^n p_\theta(X_i)}{\int \text{------}\ d\theta'}.$$

- Consistency $\Pi_n(\theta : \|\theta - \theta^*\|_2 > \varepsilon | X_1^n) \xrightarrow{P} 0$ $(\forall \varepsilon > 0)$.
- Contraction rate $\Pi_n(\theta : \|\theta - \theta^*\|_2 \ge M_n \varepsilon_n | X_1^n) \xrightarrow{P} 0$ $(\forall M_n \to +\infty)$.
- Asymptotic posterior $d_{TV}\left(\Pi_n(\cdot | X_1^n), ?? \right) \xrightarrow{P} 0$.

$\pi = N(0,1)$

$X | \theta \sim N(\theta, 1)$.

$\Pi(\theta | X_1^n) = N(0, \frac{1}{n+1})$. $\varepsilon_n = \frac{1}{\sqrt{n}}$.

Posterior Consistency.

Thm (Schwartz). Suppose (i) $\forall \varepsilon > 0$,

$$\pi\Big(\theta : D_{KL}(P_{\theta^*} \| P_\theta) < \varepsilon\Big) > 0.$$

(ii) (Minimax testing radius) $\forall \delta > 0$, $\exists\, \phi_n$

s.t. $\mathbb{E}_{\theta^*}[\phi_n] \to 0$ and $\sup\limits_{\|\theta-\theta^*\|_2 \geq \delta} \mathbb{E}_\theta[1-\phi_n] \to 0$. $(*)$

Posterior consistency holds true.

Proof: Step I: Boost the error prob.

$(*) \Rightarrow n_0 > 0$, $\mathbb{P}_{\theta^*}\big(\phi_{n_0}(X) = 1\big) < \frac{1}{4}$,

$$\mathbb{P}_\theta\big(\phi_{n_0}(X) = 1\big) > 3/4 \qquad \forall\, \theta, \|\theta-\theta^*\| \geq \varepsilon.$$

$\forall n > 0$ $\quad l = \lceil n/n_0 \rceil$. divide into subgroups

$Y_1 \cdots Y_l \overset{iid}{\sim} \mathrm{Ber}(p)$

Under θ^*: $p < \frac{1}{4}$

Under $\theta \in \mathbb{B}(\theta^*, \varepsilon)^c$:

$p > 3/4$

$$\tilde{\phi}_n(X) = \mathbb{1}\Big\{\frac{1}{n}\sum_1^n Y_i > \frac{1}{2}\Big\}.$$

Claim. $\mathbb{E}_{\theta^*}\big[\tilde{\phi}_n(X)\big] \leq 2\exp\left(-\frac{n}{32 n_0}\right)$

$$\sup_{\|\theta-\theta^*\|\geq\varepsilon} \mathbb{E}_\theta\left[1-\phi_n(X)\right] \leq 2\exp\left(-\frac{n}{32 h_0}\right)$$

for universal const $c>0$. (Hoeffding bound).

Detour (Hoeffding bound). $X_1, \cdots, X_n \overset{iid}{\sim} P$, supported on $[0,1]$.

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_1^n x_i - \mathbb{E}[X]\right| > \varepsilon\right) \leq 2\exp\left(-\frac{n\varepsilon^2}{2}\right).$$

Proof: $\forall \lambda>0$, $\mathbb{E}\left[\exp\left(\lambda(X-\mathbb{E}(X))\right)\right] \leq e^{\frac{1}{2}\lambda^2}$

$$\mathbb{E}\exp\left(\lambda\sum_1^n(X_i-\mathbb{E}(X_i))\right) \leq e^{\frac{1}{2}\lambda^2 n}.$$

Markov ineq

$$\mathbb{P}\left(\lambda\sum_1^n(X_i-\mathbb{E}[X]) \geq n\lambda\varepsilon\right) \leq e^{-n\varepsilon\lambda}\cdot\mathbb{E}[\cdots\cdots]$$

$$\leq e^{-n\varepsilon\lambda}\, e^{\frac{1}{2}\lambda^2 n}$$

$(\lambda=\varepsilon)$ $$= \exp\left(-\frac{\varepsilon^2 n}{2}\right).$$

Step II (error decomposition).

$$U := \mathbb{B}(\theta^*, \delta) = \{\theta : \|\theta-\theta^*\|_2 \leq \delta\}.$$

$$\Pi(U^c \mid X_1^n) \leq \phi_n + (1-\phi_n)\frac{\int_{U^c}\prod_{i=1}^n p_\theta/p_{\theta^*}(X_i)\,d\pi(\theta)}{\int_\Theta \prod_{i=1}^n p_\theta/p_{\theta^*}(X_i)\,d\pi(\theta)}.$$

$$\mathbb{E}_{\theta^*}\left[(1-\phi_n(X))\int_{U^c}\prod_{i=1}^n p_\theta/p_{\theta^*}(X_i)\,d\pi(\theta)\right]$$

(Fubini) $$\int_{U^c} \mathbb{E}_{\theta^*}\left[(1-\phi_n(X)) \cdot \prod_{i=1}^{n} \frac{p_\theta}{p_{\theta^*}}(X_i)\right] d\pi(\theta)$$

$$= \int_{U^c} \mathbb{E}_{\theta}\left[(1-\phi_n(X))\right] d\pi(\theta)$$

$$\leq \sup_{\theta \in U^c} \mathbb{E}_{\theta}\left[1-\phi_n(X)\right]$$

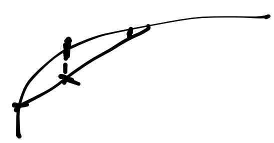
Step III : lower bound denominator.

$$\int_{\Theta} \prod_{i=1}^{n} \frac{p_\theta}{p_{\theta^*}}(X_i) \cdot d\pi(\theta). \qquad (\forall\, \Theta_0 \subseteq \Theta)$$

$$\geq \pi(\Theta_0) \cdot \int_{\Theta_0} \prod_{i=1}^{n} \frac{p_\theta}{p_{\theta^*}}(X_i)\, d\pi_0(\theta)$$

(where π_0 is π conditioned on Θ_0)

$$\log\left(\int_{\Theta_0} \prod_{i=1}^{n} \frac{p_\theta}{p_{\theta^*}}(X_i)\, d\pi_0(\theta)\right)$$



(Jensen)
$$\geq \int_{\Theta_0} \sum_{i=1}^{n} \left(\log p_\theta(X_i) - \log p_{\theta^*}(X_i)\right) d\pi_0(\theta).$$

(LLN) $$\frac{1}{n}\sum_{i=1}^{n} \int \cdots\cdots d\pi_0(\theta)$$

$$\xrightarrow{P} \mathbb{E}_{\theta^*}\left[\int_{\Theta_0} \log \frac{p_\theta}{p_{\theta^*}}(X)\, d\pi_0(\theta)\right]$$

$$= -\int_{\Theta_0} D_{KL}(p_{\theta^*} \| p_\theta)\, d\pi_0(\theta) \geq -\varepsilon.$$

$$\Theta_0 := \{ \theta \in \Theta : D_{KL}(p_{\theta^*} \| p_\theta) \leq \varepsilon \}$$

$$\mathbb{P}\left(\log \int_{\Theta} \prod_{i=1}^{n} \frac{p_\theta}{p_{\theta^*}}(X_i)\, d\pi(\theta) \leq \log \pi(\Theta_0) - 2n\varepsilon \right) \longrightarrow 0.$$

$\forall \Delta > 0$

$$\mathbb{P}\left((1-\phi_n) \frac{\int_{U^c} \prod_{i=1}^{n} p_\theta / p_{\theta^*}(X_i)\, d\pi(\theta)}{\int_{\Theta} \prod_{i=1}^{n} p_\theta / p_{\theta^*}(X_i)\, d\pi(\theta).} > \Delta \right).$$

$$\leq \mathbb{P}\left((1-\phi_n) \int_{U^c} \cdots\cdots > \Delta \cdot \pi(\Theta_0) \cdot e^{-2n\varepsilon} \right)$$

$$+ \underbrace{\mathbb{P}\left(\log \int_{\Theta} \cdots\cdots < \log \pi(\Theta_0) - 2n\varepsilon \right)}_{\longrightarrow 0}$$

$$\leq \frac{1}{\Delta \pi(\Theta_0)} \cdot e^{2n\varepsilon} \cdot \underbrace{\mathbb{E}\left[(1-\phi_n(X)) \int_{U^c} \cdots\cdots\, d\pi(\theta) \right]}_{\leq \exp\left(-\frac{n}{32 n_0}\right)}$$

Choose $\varepsilon < \frac{1}{128 n_0}$. $\qquad \longrightarrow 0.$

Composite test by covering/packing.

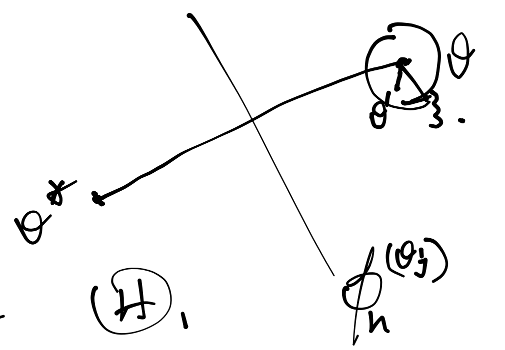
Two-point test: NP lemma.

Slightly stronger: $\forall \theta \in U^c, \exists \phi_n$.

$$\mathbb{E}_{\theta^*}\left[\phi_n(X)\right] \leq \delta_n$$

$\xi \ll \|\theta - \theta^*\|_2$

$$\sup_{\|\theta' - \theta\|_2 \leq \xi} \mathbb{E}_{\theta'}\left[\phi_n(X)\right] \leq \delta_n.$$



ξ- covering of Θ, $\{\theta_j\}_{j=1}^{N(\xi)}$

$\phi(X)$: accepts whenever $\phi_n^{(\theta_j)}$ all accept.

Failure prob $\leq N(\xi) \cdot \delta_n$.