STA 3000F — Lecture 3

From last lecture.

Thm (factorization): $\{P_\theta : \theta \in \Theta\}$, assuming density $\exists$,

$T(X)$ is sufficient $\iff$ $p_\theta(x) = g_\theta(T(x)) \cdot h(x)$ $\forall \theta \in \Theta$

"$\Rightarrow$". $g_\theta(t) = P_\theta(T(X)=t)$ $\quad h(x) = P_{\theta_0}(X=x \mid T=T(x))$.

(choice of θ_0 arbitrary).

$$P_\theta(T(X)=t; X=x) = \begin{cases} 0 & \underline{T(x)\neq t} \\ p_\theta(x) & T(x)=t. \end{cases}$$

$$g(T(x))\, h(x) = \mathbb{P}_\theta(T(X)=T(x)) \cdot \mathbb{P}_\theta(X=x \mid T=T(x))$$

$$= \mathbb{P}_\theta(X=x)$$

Thm. Full-rank exp family $p_\theta(x) = \exp(\eta(\theta)^T T(x) - B(\theta)) \cdot h(x)$,

T is complete.

Proof — ν := density of $T(X)$. under P_{θ_0}

$\mathbb{B}(\eta(\theta_0), r_0) \subseteq \eta(\Theta)$

$$\int_{\mathbb{R}^k} f(t) \cdot \exp(\eta^T t)\, \nu(t)\, dt = 0$$

For any $\eta = \eta(\theta) - \eta(\theta_0)$.

$$H(\eta) = \int_{\mathbb{R}^k} f(t) \exp(\eta^T t)\, \nu(t)\, dt \quad \text{for} \quad \eta \in \mathbb{C}^k \quad \|Re(\eta)\|_2 \le r_0$$

$\forall j$ Fix $\{\eta_i : i \neq j\}$, $H(\eta_j, \cdot)$ analytic.

Conclusion. $H(\eta) = 0 \quad \forall \eta \quad \text{s.t.} \ \|Re(\eta)\|_2 \le r_0.$

$$\widehat{f \cdot \nu} = 0 \implies f \cdot \nu = 0 \Rightarrow f = 0.$$

Thm. (Lehmann-Scheffé).

T is sufficient and complete $(P_\theta : \theta \in \Theta)$.

Suppose $\exists \delta$ unbiased for $g(\theta)$. $\mathrm{var}_\theta(\delta(X)) < +\infty$ $(\forall \theta \in \Theta)$.

Then $\delta^*(T) = \mathbb{E}[\delta(X) | T]$ is unique UMVU estimator.

Proof: δ^* unbiased

δ' another unbiased estimator.

$$\tilde{\delta}(T) = \mathbb{E}[\delta'(X) | T]. \qquad \mathrm{var}_\theta(\tilde{\delta}) \le \mathrm{var}_\theta(\delta') \quad (\forall \theta).$$

$$\mathbb{E}_\theta[\tilde{\delta}(T) - \delta^*(T)] = 0 \quad (\forall \theta \in \Theta).$$

$$\tilde{\delta} = \delta^*.$$

Find UMVU.

1. Find sufficient & complete $T(X)$
2. Find an unbiased $\delta(X)$
3. $\mathbb{E}[\delta(X) \mid T]$.

---

eg. $X_1, \dots, X_n \overset{iid}{\sim} N(\theta, 1)$. $g(\theta) = \theta$.

$T(x) = \frac{1}{n} \sum_i^n X_i$ $\delta(x) = X_1$

$\mathbb{E}[\delta(X) \mid T(X)] = T(X)$.

eg. $X_1, \dots, X_n \overset{iid}{\sim} Ber(\theta)$ $g(\theta) = \theta^k$.

$T = \sum_1^n X_i$. $\delta(X) = X_1 \cdot X_2 \cdots X_k$.

$\mathbb{E}[\delta(X) \mid T(X)] = \frac{T}{n} \cdot \frac{T-1}{n-1} \cdots \frac{T-k+1}{n-k+1}$.

$X \mid T(X) \sim Unif(\{x \in \{0,1\}^n : \sum x_i = T\})$

eg. $X_1, \dots X_n \overset{iid}{\sim} N(0, \sigma^2)$ $g(\sigma^2) = \sigma^2$.

$T = \sum_{i=1}^n X_i^2$. $\frac{T(X)}{n}$ is UMVU.

$\delta_c(T) = c \cdot T$

$R(\sigma^2, \delta_c) = \sigma^4 (2c^2 n + (nc-1)^2)$.

minimized at $c = \frac{1}{n+2}$.

eg. Truncated Poisson. $p_\theta(x) = \dfrac{\theta^x e^{-\theta}}{x!(1-e^{-\theta})}$

$(x = 1, 2, \cdots)$.

$g(\theta) = e^{-\theta}$

$X \sim p_\theta$.

$$e^{-\theta} = \mathbb{E}[\delta(X)] = \sum_{x=1}^{+\infty} \frac{\theta^x e^{-\theta}}{x!(1-e^{-\theta})} \cdot \delta(x).$$

$$\sum_{x=1}^{+\infty} \frac{\theta^x}{x!} \cdot \delta(x) = 1 - e^{-\theta} \quad (\forall \theta)$$

$$\delta(x) = (-1)^{x+1}.$$

Cramér-Rao lower bound.

Thm. $\delta(X)$ is unbiased estimator for $g(\theta) \in \mathbb{R}$.

then we have

$$(\forall \theta). \quad \mathrm{var}_\theta(\delta(X)) \geq \nabla g(\theta)^T I(\theta)^{-1} \nabla g(\theta).$$

where $I(\theta) = \mathbb{E}_\theta[\nabla \log p_\theta(x) \cdot \nabla \log p_\theta(x)^T]$.

"Fisher information"

(Assuming $\{x : p_\theta(x) > 0\}$ same for all θ

$\log p_\theta(x)$ differentiable, $\|\nabla \log p_\theta\|^2$ integrable).

Preliminaries. $\ell(\theta;X) := \log p_\theta(X)$

- $$\mathbb{E}_\theta\left[\nabla_\theta \ell(\theta;X)\right] = \int_{\mathcal{X}} \underbrace{\nabla_\theta \ell(\theta;X)\cdot e^{\ell(\theta;X)}}_{=\nabla_\theta\left(e^{\ell(\theta;X)}\right)} d\mu(x)$$

$$= \nabla_\theta \cdot \left(\int_{\mathcal{X}} e^{\ell(\theta;x)}\, d\mu(x)\right).$$

$$= 0.$$

- (Assuming C^2). $\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta_k}\ell(\theta;X)\right] = 0.$

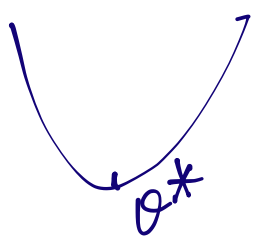
$$0 = \frac{\partial}{\partial\theta_j}\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta_k}\ell(\theta;X)\right]$$

$$= \int \frac{\partial}{\partial\theta_j}\left(\frac{\partial}{\partial\theta_k}\ell(\theta;X)\cdot e^{\ell(\theta;X)}\right) d\mu(x).$$

$$= \int \left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}\ell(\theta;X) + \frac{\partial\ell}{\partial\theta_j}\cdot\frac{\partial\ell}{\partial\theta_k}\right)\cdot e^{\ell(\theta;X)}\, d\mu(x).$$

$$\mathbb{E}_\theta\left[\nabla^2\ell(\theta;X)\right] + \mathbb{E}_\theta\left[\nabla\ell(\theta;X)\cdot\nabla\ell(\theta;X)^T\right] = 0$$

$$F(\theta) = -\mathbb{E}_\theta\left[\ell(\theta;X)\right]$$



$$F(\theta) \approx F(\theta^*) + \cancel{\langle\nabla F(\theta^*), \theta-\theta^*\rangle} + \frac{1}{2}(\theta-\theta^*)\nabla^2 F(\theta^*)(\theta-\theta^*)$$

Proof (1-D). $\theta \in \mathbb{R}$

$$g'(\theta) \underset{\text{(unbiased)}}{=} \frac{d}{d\theta}\left(\int_{\mathcal{X}} \delta(x)\, e^{\ell(\theta;x)}\, d\mu(x)\right)$$

$$= \int_{\mathcal{X}} \delta(x) \cdot \ell'(\theta;x) \cdot e^{\ell(\theta;x)}\, d\mu(x).$$

$$\int_{\mathcal{X}} \ell'(\theta;x) \cdot e^{\ell(\theta;x)}\, d\mu(x) = 0.$$

$$g'(\theta) = \int_{\mathcal{X}} \left(\delta(x) - g(\theta)\right) \cdot \ell'(\theta;x) \cdot e^{\ell(\theta;x)}\, d\mu(x).$$

$$= \mathbb{E}_\theta\left[\left(\delta(X) - g(\theta)\right) \cdot \ell'(\theta;X)\right].$$

(Cauchy-Schwarz)
$$\leq \sqrt{\operatorname{var}_\theta(\delta(X))} \cdot \sqrt{\mathbb{E}\left[\ell'(\theta;X)^2\right]}.$$

$$\operatorname{var}_\theta(\delta(X)) \geq \frac{g'(\theta)^2}{\mathbb{E}\left[\ell'(\theta;X)^2\right]}.$$

$\to$ Fisher info.

$\theta \in \mathbb{R}^d$.

$u \in \mathbb{R}^d$, $u \mapsto \mathbb{E}\left[\left(u^T \nabla_\theta \ell(\theta;X) - (\delta(X) - g(\theta))\right)^2\right].$

$$0 \leq u^T\, \mathbb{E}\left[\nabla_\theta \ell \cdot \nabla_\theta \ell^T\right] u - 2 \cdot \mathbb{E}\left[\left(\delta(X) - g(\theta)\right) \cdot \nabla_\theta \ell\right]^T u$$
$$+ \mathbb{E}\left[\left(\delta(X) - g(\theta)\right)^2\right].$$

$\forall u \in \mathbb{R}^d$, $u^T I(\theta) u - 2\nabla g(\theta)^T u + \mathrm{var}_\theta(\delta(X)) \geq 0$

$$\mathrm{var}_\theta(\delta(X)) \geq \nabla g(\theta)^T I(\theta)^{-1} \nabla g(\theta).$$

Corollary: $g: \mathbb{R}^d \to \mathbb{R}^k$,

$$\underbrace{\mathbb{E}\left[(\delta(X) - g(\theta))^{\otimes 2}\right]}_{k \times k} \succeq \underbrace{(\nabla g(\theta))^T}_{k \times d} \underbrace{I(\theta)^{-1}}_{d \times d} \underbrace{\nabla g(\theta)}_{d \times k}.$$

$(a^{\otimes 2} := aa^T.)$.

Proof: For any $v \in \mathbb{R}^k$

apply CRLB $\quad g_v(\theta) := g(\theta)^T v$.

Corollary. $g(\theta) = \theta \in \mathbb{R}^d$,

For any unbiased δ,

$$\mathbb{E}_\theta\left[\|\delta(X) - \theta\|_2^2\right] \geq \mathrm{tr}(I(\theta)^{-1}).$$

CRLB in the iid case. $\quad X_1, \dots X_n \overset{iid}{\sim} P_\theta$.

$$\ell_n(\theta; (X_i)_{i=1}^n) = \sum_{i=1}^n \ell(\theta; X_i).$$

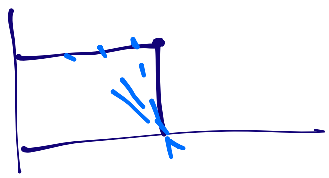
$$I_n(\theta) = n \cdot I(\theta).$$

$$\mathbb{E}\left[(\delta(X) - g(\theta))^2\right] \geq n^{-1} \cdot (\nabla g(\theta))^T I(\theta)^{-1} \nabla g(\theta).$$

"Achieved asymptotically by MLE".

Without regularity cond.

e.g. $X_1, X_2, \cdots X_n \overset{iid}{\sim} \mathrm{Unif}([0, \theta])$.



$$\delta = \max_i X_i \cdot \frac{n+1}{n}$$

$$\mathbb{E}[\delta(X)] = \theta.$$

$$\mathrm{Var}_\theta(\delta(X)) = \frac{\theta^2}{n(n+2)} \quad \text{(faster rate)}.$$

(In general, singularity/discontinuity make rate faster).

Bayesian CRLB. (van Trees' inequality).

Recall decision-theoretic framework.

$$L(\theta, a) = (g(\theta) - a)^2.$$

π over Θ, $\quad r_\pi(\delta) = \int_\Theta R(\theta; \delta)\, \pi(\theta)\, d\theta.$

Thm.

$$r_\pi(\delta) \geq \left(\int_\Theta \nabla g(\theta)\, \pi(\theta)\, d\theta\right)^T \cdot \left(\int_\Theta I(\theta)\, \pi(\theta)\, d\theta + J(\pi)\right)^{-1} \cdot \left(\int_\Theta \nabla g(\theta)\, \pi(\theta)\, d\theta\right)^T$$

where $J(\pi) := \int_{\Theta} \left(\nabla_\theta \log \pi(\theta) \, \nabla_\theta \log \pi(\theta)^T \right) \pi(\theta) \, d\theta.$

"Info theorists' Fisher info".

Proof idea

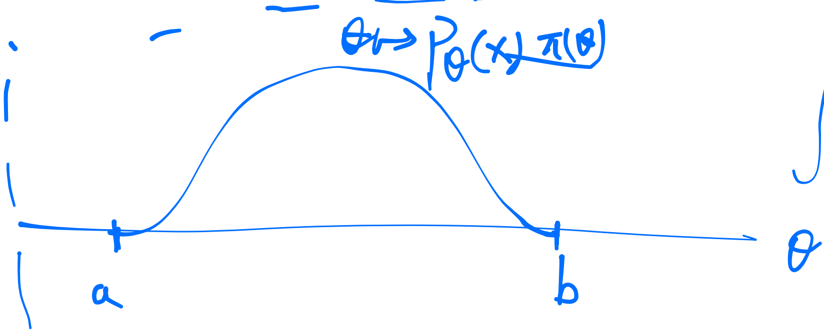
$$\nabla g(\theta) = \int \left(\delta(x) - g(\theta) \right) \cdot \nabla \ell(\theta; x) \cdot e^{\ell(\theta; x)} \, d\mu(x)$$

Breaks down for biased δ.

Key observations: (If $\pi \in C_c^1(\mathbb{R}^k)$).

$$\int_{\mathbb{R}^k} \nabla_\theta \left(p_\theta(x) \pi(\theta) \right) d\theta = 0.$$

"Integration by parts"



$$\int \cdots = p_b(x)\pi(b) - p_a(x)\pi(a) = 0.$$

$$0 = \int_{\mathbb{R}^k} \left(\nabla \log p_\theta(x) + \nabla \log \pi(\theta) \right) \cdot p_\theta(x) \pi(\theta) \, d\theta.$$

- $$\int_{\mathbb{R}^k} g(\theta) \cdot \nabla_\theta \left(p_\theta(x) \pi(\theta) \right) d\theta$$

$$= - \int_{\mathbb{R}^k} \nabla_\theta g(\theta) \cdot p_\theta(x) \pi(\theta) \, d\theta.$$

(Also by integration by-parts).

$$\int_{\mathbb{R}^k} \int_{\mathbb{X}} \nabla_\theta g(\theta) \cdot p_\theta(x) \cdot \pi(\theta) \, d\mu(x) \, d(\theta)$$

$$= - \int_{\mathbb{X}} \int_{\mathbb{R}^k} g(\theta) \cdot \left(\nabla_\theta \log p_\theta(x) + \nabla \log \pi(\theta) \right) \cdot p_\theta(x) \pi(\theta) \, d\theta \, d\mu(x).$$

$$\text{LHS} = \int_{\mathbb{R}^k} \nabla g(\theta) \, \pi(\theta) \, d\theta.$$

$$0 = \int_{\mathbb{X}} \delta(x) \int_{\mathbb{R}^k} \left(\nabla \log p_\theta(x) + \nabla \log \pi(\theta) \right) \cdot p_\theta(x) \pi(\theta) \, d\theta \, d\mu(x).$$

$$\text{RHS} = \int_{\mathbb{X}} \int_{\mathbb{R}^k} \left(\delta(x) - g(\theta) \right) \cdot \left(\nabla \log p_\theta(x) + \nabla \log \pi(\theta) \right) p_\theta(x) \pi(\theta) \, d\theta \, d\mu(x).$$

$$H(u) = \int_{\mathbb{X}} \int_{\mathbb{R}^k} \left[\left(g(\theta) - \delta(x) \right) - u^T \left(\nabla \log p_\theta(x) + \nabla \log \pi(\theta) \right) \right]^2 p_\theta(x) \pi(\theta) \, d\theta \, d\mu(x).$$

$$H(u) = u^T\left(\int_{\mathbb{R}^k} I(\theta)\,\pi(\theta)\,d\theta + J(\pi)\right)u$$

$$-2u^T\int_{\mathbb{R}^k} \nabla g(\theta)\,\pi(\theta)\,d\theta$$

$$+\, r_\pi(\delta). \quad \geq 0.$$ □

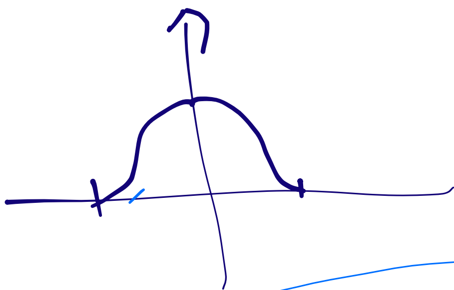
e.g. $\theta_0 \in \Theta \subseteq \mathbb{R}$ $\quad [\theta_0 - a, \theta_0 + a] \subseteq \Theta.$

$\forall n,$ $\quad X_1, X_2, \dots, X_n \overset{iid}{\sim} P_\theta$

$$\inf_{\hat\theta_n} \sup_{|\theta-\theta_0| \leq \frac{c}{\sqrt{n}}} \mathbb{E}_\theta\left[|\hat\theta_n - \theta|^2\right] \geq \frac{1}{n \cdot \sup_{|\theta-\theta_0|\leq \frac{c}{\sqrt{n}}} I(\theta) + \frac{\pi^2}{c^2}}$$

Proof — $R_{minimax}(\hat\theta_n) \geq r_\nu(\hat\theta_n).$

$$\nu_0(x) = \cos^2\left(\frac{\pi x}{2}\right)\cdot \mathbb{1}_{|x|\leq 1}$$

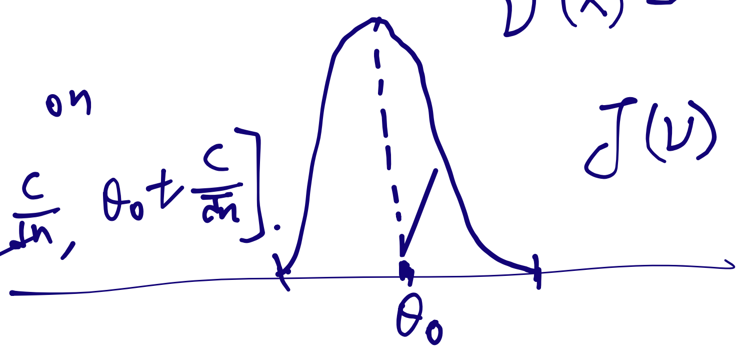


$$J(\nu_0) = \pi^2$$

$$\nu(x) = \frac{\sqrt{n}}{c}\cdot \nu_0\left(\frac{\sqrt{n}(x-\theta_0)}{c}\right).$$

$$J(\nu) = \pi^2\cdot\frac{n}{c^2}$$

supported on $\left[\theta_0 - \frac{c}{\sqrt{n}}, \theta_0 + \frac{c}{\sqrt{n}}\right].$



- $\int \nabla g(\theta)\, \pi(\theta)\, d\theta = 1$

- $\int I_n(\theta)\, \pi(\theta)\, d\theta \leq n \cdot \sup_{|\theta-\theta_0| \leq c/\sqrt{n}} I(\theta)$

- $J(\nu) = \frac{n\pi^2}{c^2}$.

If $I(\cdot)$ is cts in θ.

$$\liminf_{c \to +\infty} \liminf_{n \to +\infty} \inf_{\hat{\theta}_n} \sup_{|\theta-\theta_0| \leq c/\sqrt{n}} \left\{ n \cdot \mathbb{E}_\theta \left[|\hat{\theta}_n - \theta|^2 \right] \right\} \geq \frac{1}{I(\theta_0)}.$$

$$\geq \frac{1}{I(\theta_0) + \frac{\pi^2}{c^2}}$$

"Local asymptotic ~~minimax~~".