Recall: Bayesian CRLB

prior: π (cts diff and bdd support)

Thm (van Trees) (under MSE) for any possible estimator δ (for $g(\theta)$)

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

"$\succcurlyeq$" for vector version

Proof. Key step in CRLB that requires unbiasedness.

$$\nabla g(\theta) = \int (\delta(X) - g(\theta)) \nabla \ell(\theta; x) e^{\ell(\theta; x)} d\mu(x).$$

"looks like" moving ∇.

Key obs:

(1) $\int \nabla_\theta(p_\theta(x) \cdot \pi(\theta)) d\theta = 0$

(2). $\int g(\theta) \cdot \nabla_\theta(p_\theta(x)\pi(\theta)) d\theta = -\int \nabla_\theta g(\theta) \cdot p_\theta(x)\pi(\theta) d\theta.$

Proof: "integration-by parts".

$$\oint_{\partial(\mathrm{supp}(\pi))} g(\theta) \cdot p_\theta(x) \cdot \pi(\theta)\, d\theta = 0$$

or $p_\theta(x)\pi(\theta)$

Back to Bayesian CRLB:

(by Eq(2))

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

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

Next, we continue on eq involving $\delta(X)$.

(by Eq(1))

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

Putting them together.

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

Applying Cauchy-Schwarz / non-negative quadratic form as in the proof of CRLB.

$$\iint \left(\nabla \log p_\theta(x) + \nabla \log \pi(\theta) \right)^{\otimes 2} p_\theta(x) \cdot \pi(\theta)\, d\theta\, d\mu(x)$$

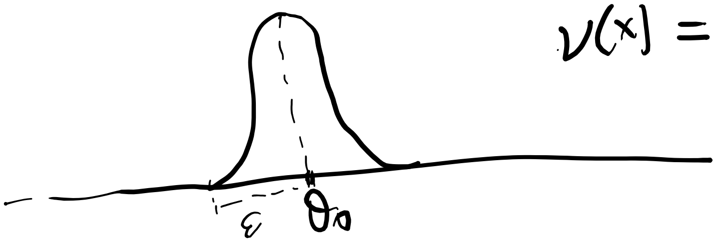
$$= \int I(\theta)\, \pi(\theta)\, d\theta + J(\pi).$$

e.g. (from last time) (ν for prior, π = 3.14159……).

$$\nu_0(x) = \cos^2\left(\frac{\pi x}{2}\right) \cdot \mathbb{1}_{\{|x| \le 1\}}$$ supported on $[-1,1]$

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

e.g. If we want ν to be supported on $[\theta_0-\varepsilon, \theta_0+\varepsilon]$

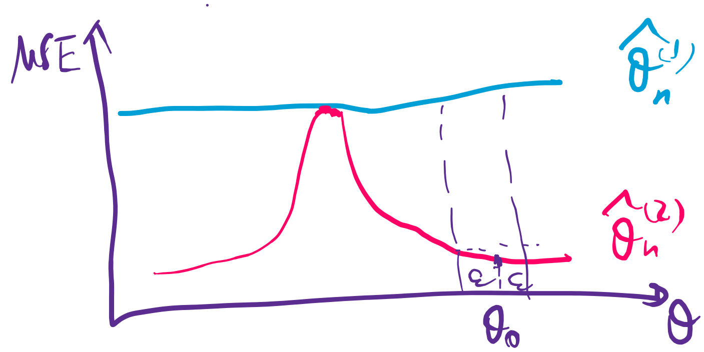


$$\nu(x) = \frac{1}{\varepsilon} \cdot \nu_0\left(\frac{x-\theta_0}{\varepsilon}\right)$$

$$J(\nu) = \frac{\pi^2}{\varepsilon^2}.$$

Why useful: local minimax.

$$\inf_{\hat{\theta}_n} \sup_{|\theta-\theta_0| \le \varepsilon} \mathbb{E}_\theta\left[|\hat{\theta}_n - \theta|^2\right] =: R_{local\text{-}minimax}(\theta_0, \varepsilon)$$



$$R_{local\text{-}minimax}(\theta_0, \varepsilon) \ge \inf_{\delta} r_\nu(\delta)$$

$$\ge \left(\int I_n(\theta)\nu(\theta)d\theta + J(\nu)\right)^{-1}$$

$$= \left(n\int I(\theta)\nu(\theta)d\theta + \frac{\pi^2}{\varepsilon^2}\right)^{-1}.$$

(taking $\varepsilon \gtrsim 1/\sqrt{n}$ first term dominates.)

Corollary: "LAM" thm, special case.

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

$$\ge \left(\int I(\theta)\nu(\theta)d\theta + \frac{\pi^2}{C^2}\right)^{-1}.$$

Rmk.

- Le Cam / Hajek. LAM theory. more general in two ways.
  - requires less regularity (DQM)
  - works for general loss functions (symmetric, bowl-shaped)

  But classical LAM holds only when $n\to+\infty$ while B-CRLB holds for finite n.
- Achived (up to high-order terms) by MLE.

Testing: basic framework.

$H_0: \theta\in\Theta_0\subseteq\Theta$ $\qquad \Theta_0\cap\Theta_1=\phi.$

$H_1: \theta\in\Theta_1\subseteq\Theta$

"critical function": $\phi(x)=\begin{cases}1 & \text{reject}\\ \pi\in(0,1) & \text{reject w.p. } \pi.\\ 0 & \text{do not reject}\end{cases}$

"significance level"

$$\alpha_\phi = \sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\phi(X)]$$

"Power" $\beta_\phi(\theta) = \mathbb{E}_\theta[\phi(X)]$ for $\theta \in \Theta_1$.

Goal: keep $\alpha_\phi \le \alpha$ for pre-specified α (e.g. 5%) while maximizing β_ϕ.

Uniformly Most Powerful (UMP):

We call ϕ^* is UMP (level-α) if for any other level-α test ϕ, we have $\beta_{\phi^*}(\theta) \ge \beta_\phi(\theta)$ $(\forall \theta \in \Theta_1)$.

(not always achievable)

Simple vs. simple testing

$\Theta_0 = \{\theta_0\}$ $\Theta_1 = \{\theta_1\}$

LRT: likelihood ratio: $L(x) = \dfrac{p_{\theta_1}(x)}{p_{\theta_0}(x)}$.

$$\varphi^*(x) = \begin{cases} 1 & L(x) > c \\ \pi & L(x) = c \\ 0 & L(x) < c. \end{cases}$$

for some (c, π).

To make it level-α test, solve (c, π) from the equation $\mathbb{E}_{\theta_0}[\varphi^*(X)] = \alpha$.

Lemma (Neymann - Pearson) Assuming p_0, p_1 has common support.

Given $\alpha \in (0,1)$, $\exists$ LRT φ^*_α w/ level $= \alpha$ and it is most powerful.

Proof. $(c=0, \pi=1) \Longleftrightarrow$ level $=1$

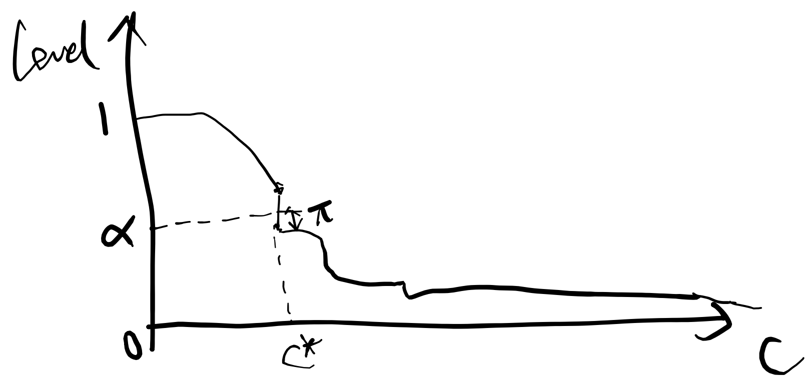
$(c \to +\infty, \pi=0)$

$$\text{level } \alpha_\phi = \int p_0(x) \mathbb{1}\{p_1(x) > c p_0(x)\} d\mu(x)$$

$$\leq \int p_0(x) \cdot \frac{p_1(x)}{c p_0(x)} d\mu(x)$$

$$= \frac{1}{c} \longrightarrow 0$$

Increase c, decrease $\pi \Rightarrow$ smaller α_ϕ.



This proves existence.

Optimality: Consider any othere level-α test ϕ.

$(c$ is threshold in LRT$)$

$$\mathbb{E}_1[\phi(X)] - c\mathbb{E}_0[\phi(X)]$$

$$= \underbrace{\int_{p_1 \geq c p_0} (p_1 - c p_0)(x)\, \phi(x)\, dx}_{\text{positive part}} + \underbrace{\int_{p_1 < c p_0} (p_1 - c p_0)(x)\, \phi(x)\, dx}_{\text{negative part}}.$$

$$\leq \int_{p_1 \geq c p_0} (p_1 - c p_0)(x)\,dx$$ (LRT ϕ^* maximizes the objective func)

$$= \mathbb{E}_1[\phi^*(X)] - c\mathbb{E}_0[\phi^*(X)].$$

$$\mathbb{E}_1[\phi(X)] \leq \mathbb{E}_1[\phi^*(X)] - c\underbrace{\mathbb{E}_0[\phi^*(X)]}_{=\alpha} + c\underbrace{\mathbb{E}_0[\phi(X)]}_{\leq\alpha}$$

$$\leq \mathbb{E}_1[\phi^*(X)]. \qquad \square$$

Key function : $\max_{\phi} \{ \mathbb{E}_1[\phi(X)] - c\mathbb{E}_0[\phi(X)] \}$.

Special case $c=1$.

$$\min_{\phi} \{ \mathbb{E}_0[\phi(X)] + \mathbb{E}_1[1-\phi(X)] \}$$

$$= \int \left(p_0(x)\cdot\phi(x) + p_1(x)\cdot(1-\phi(x)) \right) d\mu(x)$$

$$\geq \int \min(p_0(x), p_1(x))\, d\mu(x)$$

(achieved by LRT w/ $c=1$).

$$= 1 - d_{TV}(p_0, p_1)$$

where $d_{TV}(p_0, p_1) = \frac{1}{2}\int |p_1(x) - p_0(x)|\, d\mu(x)$

$$\left(= \sup_{\|f\|_\infty \leq 1} \left| \mathbb{E}_1[f(X)] - \mathbb{E}_0[f(X)] \right| \right).$$

To prove testing lower bounds:

Compute upper bound $d_{TV}(P_0, P_1)$.

(Larger distance, easier test).

e.g. $X_1, X_2, \cdots X_n \overset{iid}{\sim} \mathcal{N}(\theta, 1)$.

$$H_0: \theta = 0$$

$$H_1: \theta = \mu$$

$$d_{TV}\left(P_1^{\otimes n}, P_0^{\otimes n}\right) \le \sqrt{\frac{1}{2} D_{KL}\left(P_1^{\otimes n} \| P_0^{\otimes n}\right)}$$

(Pinsker's ineq)

$$D_{KL}(P \| Q) = \mathbb{E}_P\left[\log \frac{dP}{dQ}\right].$$

$$D_{KL}(P_1 \times P_2 \times \cdots P_n \| Q_1 \times \cdots \times Q_n) = \sum_{i=1}^{n} D_{KL}(P_i \| Q_i)$$

("tensorization property").

$$\le \sqrt{\frac{n}{2} D_{KL}(P_1 \| P_0)} = \sqrt{\frac{n\mu^2}{2}}.$$

when $\mu = \frac{1}{\sqrt{n}}$.

$$\mathbb{P}(\text{type I err}) + \mathbb{P}(\text{type-II err}) \ge 1 - \frac{1}{\sqrt{2}}$$

Rmk: choose $c=1$ to get d_{TV},

choose different c to get asymmetric notion of distances / indistinguishability.

Rmk: We can study estimation lower bound based on testing framework.

Le Cam's two-point method.

Idea: If we cannot distinguish P_0, P_1, $g(\theta)$ difference under P_0, P_1 then any estimator must pay this price.

Lemma (Le Cam). $\delta(X)$ estimator for $g(\theta)$. $\theta_0, \theta_1 \in \Theta$, then.

$$\inf_{\delta} \sup_{\theta \in \Theta} \mathbb{E}_\theta\left[|\delta(X) - g(\theta)|^2\right] \geq \frac{|g(\theta_1) - g(\theta_0)|^2}{8}\left(1 - d_{TV}(P_{\theta_1}, P_{\theta_0})\right)$$

(here $P_{\theta_1}, P_{\theta_0}$ denotes joint distribution of all observations).

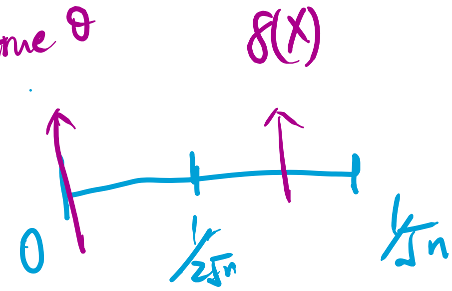
Proof. prior $\pi(\theta_0) = \frac{1}{2}$, $\pi(\theta_1) = \frac{1}{2}$.

$$\forall \delta, \quad \sup_{\theta \in \Theta} \mathbb{E}[\cdots] \geq r_\pi(\delta) = \frac{1}{2}\mathbb{E}_0\left[|\delta(X) - g(\theta_0)|^2\right] + \frac{1}{2}\mathbb{E}_1\left[|\delta(X) - g(\theta_1)|^2\right].$$

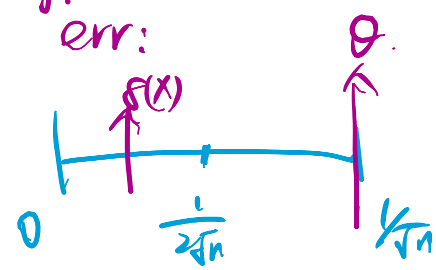
Construct a test

$$\phi(x) := \mathbb{1}\{|\delta(X) - g(\theta_1)| \leq |\delta(X) - g(\theta_0)|\}$$

Type I err:



Type 2 err:



Making an error $\iff |\delta(X) - g(\theta)| \geq \frac{1}{2} |g(\theta_1) - g(\theta_0)|$.

By testing lower bound,

$$\frac{1}{4}|g(\theta_1)-g(\theta_0)|^2 \frac{1}{2} \left(\text{type-I} + \text{type-II err}\right) \geq \frac{1}{2}\left(1 - d_{TV}(P_0 \| P_1)\right).$$

$$\le$$

$$\frac{1}{2} \mathbb{E}_{\theta_0} |\delta(X) - g(\theta_0)|^2 + \frac{1}{2} \mathbb{E}_{\theta_1} |\delta(X) - g(\theta_1)|^2.$$

$$\le$$

$$\sup_{\theta} \mathbb{E}_{\theta}\left[|\delta(X) - g(\theta)|^2\right].$$

Back to testing.

Idea: in some settings LRTs are all based on the same test stats for any $\theta_0 \in \Theta_0$, $\theta_1 \in \Theta_1$.

e.g. $X \sim N(\theta, 1)$

For $\theta_1 > \theta_0$

$$H_0 : \theta = \theta_0$$
vs.
$$H_1 : \theta = \theta_1$$

$$L(x) = \exp\left((\theta_1 - \theta_0) \cdot x - \frac{\theta_1^2 - \theta_0^2}{2} \right).$$

ϕ^* thresholds based on x.

What if we test

$$\Theta_0 : \theta \leq \theta^*$$
$$\Theta_1 : \theta > \theta^*..$$

Def. Monotone likelihood ratio (MLR) class.

$\theta \in \mathbb{R}$. if $\exists$ a summary statistics $T(X) \in \mathbb{R}$.

s.t. $\forall \theta_1 < \theta_2$ then $p_{\theta_2}(x) / p_{\theta_1}(x)$

is a non-decreasing function of T

and $\forall \theta_1 \neq \theta_2$, $p_{\theta_1} \neq p_{\theta_2}$

e.g. (exponential family) $p_\theta(x) = \exp(\eta(\theta) \cdot T(x) - B(\theta)) h(x)$.

$$p_{\theta_2}(x) / p_{\theta_1}(x) = \exp\left((\eta(\theta_2) - \eta(\theta_1)) \cdot T(x) - B(\theta_2) + B(\theta_1) \right).$$

If η is monotonic function of θ,

this is MLR.

Thm For the testing problem

$$H_0: \theta \le \theta_0 \quad \text{vs.} \quad H_1: \theta > \theta_0$$

Assuming MLR, then $\exists$ UMP test

$$\phi^*(x) = \begin{cases} 1 & T(X) > c \\ \pi & T(X) = c \\ 0 & T(X) < c \end{cases}$$

with $\mathbb{E}_{\theta_0}[\phi^*(X)] = \alpha$.

Proof ϕ^* is LRT for θ_0 vs. θ_1 $(\forall\, \theta_1 \in \Theta_1)$

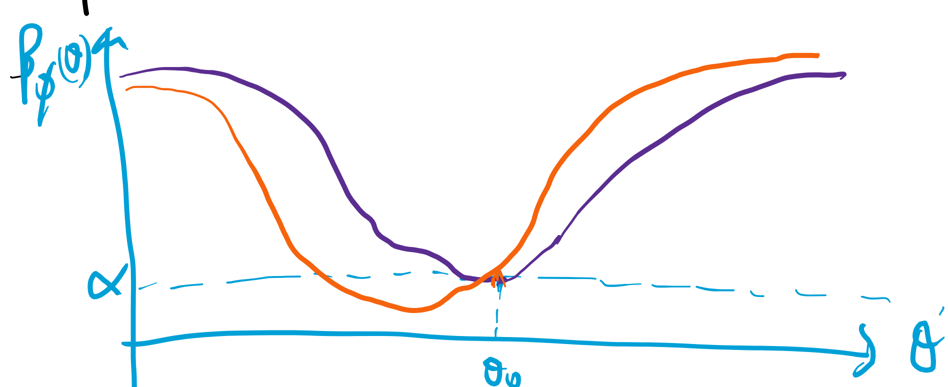
By NP lemma, ϕ^* is optimal for this simple vs. simple problem.

$\Theta \subseteq \mathbb{R}$.

Two-sided testing:

$$H_0: \theta = \theta_0 \quad \text{vs.} \quad H_1: \theta \ne \theta_0$$

More power on $\theta > \theta_0$ $\Leftrightarrow$ less power on $\theta < \theta_0$.



Def. "unbiased ~~test~~":

when $\mathbb{E}_{\theta}[\phi(X)] \geq \alpha$ $(\forall \theta \in \Theta_1)$.

We can study UMPU.