LAN

(1) The reason we want contiguity is to make the likelihood ratios $\frac{dQ_n}{dP_n}$ make sense.

(2) DQM + i.i.d $\longrightarrow$ LAN property $\longrightarrow$ $\log \prod_i \frac{P_{\theta+h/\sqrt{n}}}{P_\theta} \xrightarrow[d]{P_\theta} N(-\frac{1}{2}h' I_\theta h, h' I_\theta h)$

$X_n \xrightarrow{P_{\theta+h/\sqrt{n}}} L$ $\longleftarrow$ Le Cam's third lemma $\longleftarrow$

(3) $$l_{\theta+h}(x) = l_\theta(x) + h' \dot{l}_\theta(x) + \frac{1}{2} h' \ddot{l}_\theta(x) h + o(\|h\|^2)$$

(4) Investigating the asymptotic test with fixed alternative doesn't make much sense (as $n \to \infty$, the test will alway accept H_0). It makes sense when we consider shrinking alternative $\theta_0 + h/\sqrt{n}$

Cramér – Rao Lower bound

(1) Information inequality

$$\mathrm{Var}_\theta(\delta) \geq \frac{\left[\frac{\partial}{\partial\theta}\mathbb{E}_\theta \delta\right]^2}{I(\theta)} = \frac{[b'(\theta) + g'(\theta)]^2}{I(\theta)}$$

$\delta = \delta(X)$, $I(\theta)$ is the information in X w.r.t. θ.

If $X = X_1$, then $I(\theta) = I_1(\theta)$

If $X = (X_1, \cdots, X_n)$, then $I(\theta) = nI_1(\theta)$.

Equivariant

The reason we consider equivariant estimator is that, under invariant loss, the risk is a constant, which is easier to compare.

UMPU test

1. In some cases, the constraints are not enough, and hence UMP test doesn't exist. To get an optimal test, we need to add more constraints, restricting the class to be smaller. Unbiased, SOB are just the constraints added.

2. one-parameter exponential family

Promblem II: $H_0: \theta \le \theta_1$ or $\theta \ge \theta_2$ vs. $H_1: \theta_1 < \theta < \theta_2$

Problem III: $H_0: \theta_1 \le \theta \le \theta_2$ vs. $H_1: \theta < \theta_1$ or $\theta > \theta_2$

Problem II has UMP test; while Problem III has no UMP test, but UMPU test.

Problem II: maximize $\mathbb{E}_{\theta'} \varphi$ s.t. $\underbrace{\mathbb{E}_{\theta_1} \varphi \le \alpha, \ \mathbb{E}_{\theta_2} \varphi \le \alpha}$
$\theta_1 < \theta' < \theta_2$

$\Downarrow$

$K_1 > 0, \ K_2 > 0$

$\varphi(x) = 1$ when $P_{\theta'}(x) > K_1 P_{\theta_1}(x) + K_2 P_{\theta_2}(x)$

$(Q(\theta_1) < Q(\theta') < Q(\theta_2))$

$\mathbb{E}_{\theta_1} \varphi = \alpha, \ \mathbb{E}_{\theta_2} \varphi = \alpha$ has solution in this case.

So UMP test exists.

Problem II :

(1) Try to get UMP test :

maximize $\underset{\theta'<\theta_1 \text{ or } \theta'>\theta_2}{\mathbb{E}_{\theta'}\varphi}$ s.t. $\underbrace{\mathbb{E}_{\theta_1}\varphi \le \alpha \,,\ \mathbb{E}_{\theta_2}\varphi \le \alpha}$

$\Downarrow$

$K_1, K_2 \ge 0$

$\varphi(x) = 1$ when $P_{\theta'}(x) > K_1 P_{\theta_1}(x) + K_2 P_{\theta_2}(x)$

($Q(\theta') < Q(\theta_1) < Q(\theta_2)$ or

$Q(\theta') > Q(\theta_2) > Q(\theta_1)$)

$\mathbb{E}_{\theta_1}\varphi = \alpha$, $\mathbb{E}_{\theta_2}\varphi = \alpha$ has no solution.

So, no UMP test available,

(2) Try to get UMPU test : (or UMP SOB test)

maximize $\underset{\theta'<\theta_1 \text{ or } \theta'>\theta_2}{\mathbb{E}_{\theta'}\varphi}$ s.t. $\underbrace{\mathbb{E}_{\theta_1}\varphi = \alpha \,,\ \mathbb{E}_{\theta_2}\varphi = \alpha}$

$\Downarrow$

this is the difference,

K_1, K_2 can be positive or negative.

which makes $\mathbb{E}_{\theta_1}\varphi = \mathbb{E}_{\theta_2}\varphi = \alpha$ solvable

So, UMPU test exists in this case.

Another approach to see UMP exists for II, not III.

For $H_1: \theta > \theta_i$, the UMP test has form $\varphi_1(x) = \begin{cases} 1 & T > c \\ \gamma & T = c \\ 0 & T < c \end{cases}$

for $H_1: \theta < \theta_i$, the UMP test has form $\varphi_2(x) = \begin{cases} 1 & T < c \\ \gamma & T = c \\ 0 & T > c \end{cases}$

In problem II, $H_1: \theta_1 < \theta < \theta_2$

In problem III, $H_1: \theta > \theta_2$ or $\theta < \theta_1$,

for $\theta > \theta_2$, φ_1 is best, for $\theta < \theta_1$, φ_2 is best.

but φ_1 and φ_2 are different. so no UMP test.

Neyman structure

In the high dimensional case, we use conditioning to reduce to the univariate case. Since exponential family has good properties, we consider this family and can derive UMPU test.

Use $H_0: \theta \leq \theta_0$ vs. $\theta > \theta_0$ as an example.

Step 1: derive UMP conditional test.

Given $T = t$, $U | T \sim \exp(\theta u - A_t(\theta))$, and

$$\varphi_1 = \begin{cases} 1 & U > c(t) \\ \gamma(t) & U = c(t) \\ 0 & U < c(t) \end{cases} \quad \text{s.t. } \mathbb{E}_{\theta_0}[\varphi_1 \mid T = t] = h(t) = \alpha$$

We want the above to hold for $\forall t$, $\therefore h(T) \equiv \alpha$. and this is why we need Neyman structure, which is $\mathbb{E}_{\theta_0}[\varphi_1 \mid T] = \alpha$

Note that UMP conditional test has the following properties (Theorem 12.9):

(1) $\forall \theta > \theta_0$, $\mathbb{E}_{\theta,\eta}[\varphi_1 \mid T = t] \geq \mathbb{E}_{\theta,\eta}[\varphi \mid T = t]$

(2) the conditional power function is increasing.

(3) $\mathbb{E}_{\theta_0, \eta} \varphi_1 = \mathbb{E}_\eta \mathbb{E}_{\theta_0} [\varphi_1 | T] = \alpha \quad \therefore \varphi_1$ is SOB.

Step II : prove φ_1 is UMPU

use the relationship between unbiased and SOB and Neyman structure.

φ is SOB $\xRightarrow{T \text{ is complete ss.}}$ φ has Neyman structure

$\Longrightarrow$ φ is a valid conditional test

φ_1 is UMP conditional test + smoothing

$\Longrightarrow$ φ_1 is UMP SOB

power continuous + φ_1 is of level α

$\Longrightarrow$ φ_1 is UMPU

p - value

condition: nested

1. If $\sup_{\theta \in \Theta_0} P_\theta (X \in S_\alpha) \le \alpha$ for all $0 < \alpha < 1$, then

for all $\theta \in \Theta_0$, $\mathbb{P}_\theta (\hat{p} \le u) \le u$ for all $0 \le u \le 1$

There are several things to note:

① $\hat{p} = \inf \{ \alpha : X \in S_\alpha \}$ is a statistic.

Fix X, enlarge α, till S_α contains X.

② musted be nested rejection region.

$\{ \hat{p} \le u \} \Rightarrow \forall\, v > u,\ X \in S_v$

Let $v > u$, then $\mathbb{P}_\theta (\hat{p} \le u) \le \mathbb{P}_\theta (X \in S_v)$

$$\le \sup_{\theta \in \Theta_0} \mathbb{P}_\theta (X \in S_v) \le v$$

then let $v \to u$, we have $\mathbb{P}_\theta (\hat{p} \le u) \le u$

2. If for some $\theta \in \Theta_0$, $\mathbb{P}_\theta (X \in S_\alpha) = \alpha$, then for this θ,

$\mathbb{P}_\theta (\hat{p} \le u) = u$

$\{ X \in S_u \} \Rightarrow \hat{p} \le u \quad \Rightarrow \mathbb{P}_\theta (\hat{p} \le u) \ge \mathbb{P}_\theta (X \in S_u) = u$

$$\Rightarrow \mathbb{P} (\hat{p} \le u) = u$$

General Linear Model

$Y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n)$

1. Do some transformation to make the analysis easier.

   (1) Let $\xi = X\beta$, $\xi \in W = C(X)$ of rank r.

   then $Y = \xi + \varepsilon$

   (2) transform coordinate system : $Y = OZ$

   then $Z = O^T Y = O^T \xi + O^T \varepsilon$.

   Let $\eta = O^T \xi$, then $Z \sim N(\eta, \sigma^2 I_n)$

   (3) Decide the form of O.

   We want the distribution be as simple as possible and note that ξ is restricted to a rank-r space.

   Let $v_1, \cdots, v_r$ be the orthogonal basis of W.

   $v_{r+1}, \cdots, v_n$ be the remaining to form orthogonal basis of R^n.

   Then we have $\eta = \begin{pmatrix} v_1^T \\ \vdots \\ v_r^T \\ v_{r+1}^T \\ \vdots \\ v_n^T \end{pmatrix} \xi = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ (simpler)

2. Find complete and sufficient statistics.

   (Optimal estimator and tests are usually based on this)

   density is $(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{\sum_{i=1}^{n} z_i^2}{2\sigma^2} + \frac{\sum_{i=1}^{r} \eta_i z_i}{\sigma^2} - \frac{\sum_{i=1}^{r} \eta_i^2}{2\sigma^2} \right\}$

$\Rightarrow T = (z_1, \cdots, z_r, \sum_{i=1}^{n} z_i^2)$, or $(z_1, \cdots, z_r, \sum_{i=r+1}^{n} z_i^2)$

is complete and sufficient.

3. Now get UMVUE of $a'\hat{\xi}$

(1) $\mathbb{E} z_i = \eta_i \quad \therefore \hat{\eta}_i = z_i$ (is a function of T)

(2) $\xi = O\eta = (v_1, \cdots, v_n) \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \sum_{i=1}^{r} v_i \eta_i$

$$\therefore \hat{\xi} = \sum_{i=1}^{r} v_i z_i$$

$$= (v_1, \cdots, v_r) \begin{pmatrix} v_1^T \\ \vdots \\ v_r^T \end{pmatrix} Y = O_{[,1:r]} O_{[,1:r]}' Y = PY$$

$a'\hat{\xi} = a'PY$ is the UMVUE for $a'\xi$

4. Analyze the property of P

(1) $P' = P$, $P^2 = P \Rightarrow P$ is a projection matrix

(2) $PX = X$, $P\xi = \xi$ ($\because X, \xi \in \omega$)

(3) when X is of full rank, then $P = X(X'X)^{-1}X'$

(4) Implication of PY (projection)

① $PY = \underset{w \in \omega}{\arg\min} \|Y - w\|^2$

② $Y - PY \perp \omega$

(5) By projection, we seperate Y into two independent parts.

$\hat{\xi} = PY = \sum_{i=1}^{r} v_i z_i$, $e = Y - \hat{\xi} = Y - PY = \sum_{i=r+1}^{n} v_i z_i$

Also, note that $\|e\|^2 = \sum_{i=r+1}^{n} z_i^2$

PY has one-to-one correspondence to $(z_1, \cdots, z_r)$

$\therefore (PY, \|Y - PY\|^2)$ is complete and sufficient

5. Distribution of $\hat{\xi}$

(1) $\hat{\xi} = PY$, $Y \sim N(\xi, \sigma^2 I_n)$

$\therefore \hat{\xi} \sim N(P\xi, P\sigma^2 I_n P')$

$P\xi = \xi$, $PP' = P^2 = P$

$\therefore \hat{\xi} \sim N(\xi, \sigma^2 P)$

(2) $a'\hat{\xi} \sim N(a'\xi, a'\sigma^2 P a)$

$a'Pa = a'P'Pa = \|Pa\|^2$

$\therefore a'\hat{\xi} \sim N(a'\xi, \sigma^2 \|Pa\|^2)$

6. Go back to the original : $\hat{\beta}$.

$X\beta = \xi$, so $X\hat{\beta} = \hat{\xi}$

(1) $\hat{\xi} = \underset{w \in W}{\operatorname{argmin}} \|Y - w\|^2$ ($\because \hat{\xi} = PY$, definition)

$\therefore \hat{\beta} = \underset{\beta}{\operatorname{argmin}} \|Y - X\beta\|^2$ is least-square estimator.

(2) Get an equation of $\hat{\beta}$ based on X and Y.

$X\hat{\beta} = PY$,

$PX = X \Rightarrow X'P' = X' \Rightarrow X'P = X'$

$\therefore X'X\hat{\beta} = X'Y$

(3) particular case : when X is of full rank.

$X'X$ is invertible, $\hat{\beta} = (X'X)^{-1}X'Y$ is unique

$= (X'X)^{-1}X'\hat{\xi}$ is UMVUE

$\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$

7. Now focus on σ^2.

UMVUE is $S^2 = \frac{1}{n-r}\sum_{i=r+1}^{n} Z_i^2$

$= \frac{\|Y-\hat{\xi}\|^2}{n-r}$

$\frac{(n-r)S^2}{\sigma^2} \sim \chi^2_{n-r}$

8. Confidence interval for $a'\hat{\xi}$

$$\frac{a'\hat{\xi} - a'\xi}{\hat{\sigma}_{a'\hat{\xi}}} \sim t_{n-r}$$

9. Extend to general case

$\mathbb{E}Y = X\beta$, $Cov(Y) = \sigma^2 I_n$

Gauss-Markov Theorem : $a'\hat{\xi}$ is the best linear unbiased estimator (BLUE) for $a'\xi$

$a'\hat{\xi} = a'PY$, $\mathbb{E}\hat{\xi} = \xi$, $Cov(\hat{\xi}) = P\,Cov(Y)\,P'$

$= \sigma^2 P$

$\therefore \mathbb{E}\, a'\hat{\xi} = a'\xi$, $Var(a'\hat{\xi}) = \sigma^2 a'Pa = \sigma^2 \|Pa\|^2$

Consider $b'Y$, $\mathbb{E}\, b'Y = b'\xi = a'\xi$

$\Rightarrow (b-a)'\xi = 0 \Rightarrow b-a \perp W$

$Var(b'Y) = b'\sigma^2 I_n b = \sigma^2 \|b\|^2$

$$b = \underset{\in W}{Pa} + \underbrace{(I-P)a + b - a}_{\perp W}$$

$\therefore \|b\|^2 \geq \|Pa\|^2$ $\therefore Var(a'\hat{\xi}) \leq Var(b'Y)$

Geometry for parametric models

1.
The goal is to see whether not knowing the nuisance parameter will influence the estimation.

2. $\theta = (\nu, \eta)$, $q(\theta) = q(\nu, \eta) = \nu$, η is the nuisance parameter.

$\theta_0 = (\nu_0, \eta_0)$, $\dot{l} = \begin{pmatrix} \dot{l}_1 \\ \dot{l}_2 \end{pmatrix}$ is the score function for θ,

$\begin{pmatrix} \tilde{l}_1 \\ \tilde{l}_2 \end{pmatrix} = \tilde{l} = I^{-1}(\theta_0)\, \dot{l}$ is the efficient influence.

$I(\theta_0)$ is the information.

When η is not known: (use I to denote $I(\theta_0)$)

$$I = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}, \quad I^{-1} = \begin{pmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{pmatrix} = \begin{pmatrix} I_{11.2}^{-1} & -I_{11.2}^{-1} I_{12} I_{22}^{-1} \\ -I_{22.1}^{-1} I_{21} I_{11}^{-1} & I_{22.1}^{-1} \end{pmatrix},$$

$$I_{11.2} = I_{11} - I_{12} I_{22}^{-1} I_{21}$$

$$I_{22.1} = I_{22} - I_{21} I_{11}^{-1} I_{12}$$

$$\begin{pmatrix} \tilde{l}_1 \\ \tilde{l}_2 \end{pmatrix} = I^{-1} \begin{pmatrix} \dot{l}_1 \\ \dot{l}_2 \end{pmatrix} \Rightarrow \tilde{l}_1 = I^{11} \dot{l}_1 + I^{12} \dot{l}_2$$

$$= I_{11.2}^{-1} (\dot{l}_1 - I_{12} I_{22}^{-1} \dot{l}_2)$$

$$= I_{11.2}^{-1} \dot{l}_1^*$$

$\because E \dot{l}_1^* \dot{l}_1^{*T} = I_{11.2}$

$\therefore \dot{l}_1^* = \dot{l}_1 - I_{12} I_{22}^{-1} \dot{l}_2$ is the efficient score of ν,

$\tilde{l}_1$ is the efficient influence of ν,

$I_{11.2}$ is the information of ν

$I^{11} = I_{11.2}^{-1}$ is the information bound of ν.

$I_{11.2} = I_{11} - I_{12} I_{22}^{-1} I_{21} \le I_{11}$, smaller information.

when $I_{12} = 0$, adaptive estimator.

3. projection

(1) $I_{12} I_{22}^{-1} \dot{l}_2$ is the projection of $\dot{l}_1$ on $[\dot{l}_2]$,

so $\dot{l}_1^* = \dot{l}_1 - I_{12} I_{22}^{-1} \dot{l}_2$ is the projection of $\dot{l}_1$ on the orthocomplement of $[\dot{l}_2]$

(2) influence $\tilde{l}_1 : P$ $\quad I_{11}^{-1} \dot{l}_1 : P_1(\eta_0)$,

$I_{11}^{-1} \dot{l}_1$ is the projection of $\tilde{l}_1$ on $[\dot{l}_1]$

4. Other relationship

$\dot{l} = I \tilde{l} \Rightarrow \dot{l}_1 = I_{11} \tilde{l}_1 + I_{12} \tilde{l}_2$

$\Rightarrow \tilde{l}_1 = \underset{\substack{\uparrow \\ \text{projection of} \\ \tilde{l}_1 \text{ on } [\dot{l}_1]}}{I_{11}^{-1} \dot{l}_1} - \underset{\substack{\uparrow \\ \text{orthocomplement}}}{I_{11}^{-1} I_{12} \tilde{l}_2}$ $\quad$ (1) (influence)

take variance

$I_{11.2}^{-1} = I_{11}^{-1} + I_{11}^{-1} I_{12} I_{22.1}^{-1} I_{21} I_{11}^{-1}$ $\quad$ (2) (information bound)

5. For $P_1(\eta_0)$

efficient score : $\dot{l}_1$

efficient influence : $I_{11}^{-1}\dot{l}_1$

information : I_{11}

information bound : I_{11}^{-1}

Pass the derivative inside integral

Want to have $\frac{\partial}{\partial\theta}\int p_\theta(x)\,dx = \int \frac{\partial p_\theta(x)}{\partial\theta}\,dx$

$\Downarrow$

need to have

$$\sup_{h\in[-\varepsilon,\varepsilon]} \left| \frac{p_{\theta+h}(x) - p_\theta(x)}{h} \right| \leq K(x)$$

uniformly bounded .

Wald's theorem

① Note that a continuous function has maximum and minimum inside a compact set.

And an upper semicontinuous function on a compact set will surely achieves its maximum.

② Examples:

(1) $\{X_i\}_{i=1}^{n} \sim$ Cauchy (θ) , $\hat{\theta}_n \longrightarrow \theta_0$ a.s.

(2) Uniform $[0, \theta]$, $\hat{\theta}_n \longrightarrow \theta_0$ a.s.

we may use Wald's thm when we don't have close form solution

③ compactification

Lower bound

① If T_n is regular, then $Var(T_n) \geq I_\theta^{-1}$

② If T_n is arbitrary, then $Var(T_n) \geq I_\theta^{-1}$ a.e. θ

but there exists some point θ, $Var(T_n) < I_\theta^{-1}$.

In 2018 part 2 problem 4,

$\because$ α and trigamma(α) are continuous

$\therefore$ $\alpha \geq \frac{1}{\text{trigamma}(\alpha)}$ a.e. means always.

so $\alpha \geq \frac{1}{\text{trigamma}(\alpha)}$

Bayes and minimax

Consider θ and X, we know $P_\theta(X)$

(1) $\lambda(\theta)$ is known

then we can get the Bayes estimator and check the Bayes risk. If it's a constant, then is minimax

(2) $\lambda(\theta)$ is not known

(i) assume the prior is $\Lambda_m(\theta)$,

if $r_m \to r$ and $\sup_\theta(\theta, \delta) = r$,

then δ is minimax

(ii) assume the prior is $\Lambda_m(\theta)$,

use data to estimate the hyperparameters of $\Lambda_m(\theta)$. This is empirical Bayes.