Midterm exam, Statistical Inference II: (2016 Spring): [+40points]

Name: Shih Jia-Han

- Proofs must be understandable to the instructor.
- Avoid typos and undefined notations in your proofs.

Q1 [+5] Prove or disprove (with a counter example) the following statements.

1) [+3] The unique Bayes estimator is minimax.

Answer:

Assume that

$$X \sim \text{Bin}(n, p), \quad p \sim \text{Beta}(a, b) \Rightarrow p \mid x \sim \text{Beta}(x+a, n-x+b).$$

Therefore, the Bayes estimator of p can be written as

$$\delta_\Lambda = E(p \mid x) = \frac{x-a}{n+a+b},$$

which is an unique Bayes estimator for arbitrary a, b. Then consider the risk function of δ_Λ

$$\begin{aligned}R(p, \delta_\Lambda) &= E\left\{\left(\frac{x+a}{n+a+b} - p\right)^2\right\} = \frac{1}{(n+a+b)^2} E[\{(a+x)-(a+b+n)p\}^2] \\ &= \frac{1}{(n+a+b)^2} E([(x-np)+\{a(1-p)-bp\}]^2) \\ &= \frac{1}{(n+a+b)^2}[np(1-p)+\{a(1-p)-bp\}^2].\end{aligned}$$

Solve for

$$\begin{aligned}\frac{\partial R(p, \delta_\Lambda)}{\partial p} = 0 &\Rightarrow n - 2np - 2(a+b)\{a-(a+b)p\} = 0 \\ &\Rightarrow \{2(a+b)^2 - 2n\}p + n - 2a(a+b) = 0.\end{aligned}$$

That is

$$\begin{cases}(a+b)^2 = n \\ 2a(a+b) = n\end{cases} \Rightarrow \begin{cases}a = \sqrt{n}/2 \\ b = \sqrt{n}/2\end{cases}$$

Thus, let

$$\delta_M = \frac{x - \sqrt{n}/2}{n + \sqrt{n}}.$$

Therefore, for all arbitrary a, b, δ_Λ is an unique Bayes estimator but

$$\sup_p R(p, \delta_\Lambda) > \sup_p R(p, \delta_M).$$

Hence unique Bayes estimator is not minimax. □

2) [+2] The unique Bayes estimator is admissible.

Answer:

Let δ_Λ be an unique Bayes estimator of θ. Assume δ_Λ is not admissible, that is there exist δ' such that $R(\theta, \delta') \le R(\theta, \delta_\Lambda)$. Since δ_Λ is a Bayes estimator, we have

$$R(\theta, \delta') \le R(\theta, \delta_\Lambda) \le R(\theta, \delta), \text{ for all } \delta.$$

Therefore, δ' is also a Bayes estimator which contradict to the uniqueness. Hence the unique Bayes estimator is admissible. □

Q2 [+3]. Let $X_1,..,X_n \overset{iid}{\sim}$ with $E[X_i]=\xi$, and $Var[X_i]=\sigma^2$. Derive the asymptotic distribution of $n\{h(\bar{X})-h(\xi)\}$ when $h'(\xi)=0$ and $h''(\xi)\neq 0$. (with proof)

Answer:

By a Taylor expansion, we have

$$h(\bar{X})=h(\xi)+(\bar{X}-\xi)h'(\xi)+\frac{1}{2}(\bar{X}-\xi)^2h''(\xi)+O_p(|\bar{X}-\xi|^3)$$

$$\Rightarrow n\{h(\bar{X})-h(\xi)\}=\frac{1}{2}h''(\xi)n(\bar{X}-\xi)^2+nO_p(|\bar{X}-\xi|^3).$$

By the Central Limit Theorem,

$$\sqrt{n}(\bar{X}-\xi)\xrightarrow{d}N(0,\sigma^2)\Rightarrow\frac{\sqrt{n}}{\sigma}(\bar{X}-\xi)\xrightarrow{d}N(0,1)$$

$$\Rightarrow\frac{n}{\sigma^2}(\bar{X}-\xi)^2\xrightarrow{d}\chi^2_{\text{df}=1}\Rightarrow n(\bar{X}-\xi)^2\xrightarrow{d}\sigma^2\chi^2_{\text{df}=1},\ \text{as}\ \ n\to\infty.$$

By the Weak Law of Large Number,

$$\bar{X}-\xi\xrightarrow{p}0,\ \text{as}\ \ n\to\infty.$$

By the Slusky's Theorem,

$$n(\bar{X}-\xi)^3=n(\bar{X}-\xi)^2(\bar{X}-\xi)=O_p(1)(\bar{X}-\xi)\xrightarrow{p}0,\ \text{as}\ \ n\to\infty.$$

Combining the results above, again, by the Slusky's Theorem, we have

$$n\{h(\bar{X})-h(\xi)\}\xrightarrow{d}\frac{h''(\xi)}{2}\sigma^2\chi^2_{\text{df}=1},\ \text{as}\ \ n\to\infty.\ \square$$

Q3 [+12]. Consider an exponential family

$$\mathbf{X}=(X_1,\ldots,X_p)\sim p_{\boldsymbol{\eta}}(\mathbf{x})=\exp\left[\sum_{i=1}^{s}\eta_i T_i(\mathbf{x})-A(\boldsymbol{\eta})\right]h(\mathbf{x}),$$

where $p_{\boldsymbol{\eta}}$ is the density with respect to some measure, and let $\boldsymbol{\eta}\sim\pi(\boldsymbol{\eta})$, where π is the prior density with respect to the Lebesgue measure. Let $m(\mathbf{x})$ be the marginal density of $\mathbf{X}$.

(1) [+10]Express $E\left[\sum_{i=1}^{s}\eta_i\frac{\partial T(\mathbf{X})}{\partial X_i}\middle|\mathbf{X}\right]$ by $m(\mathbf{x})$, $h(\mathbf{x})$, and their derivatives.

<u>**Answer:**</u>

Since

$$\frac{\partial p_{\boldsymbol{\eta}}(\mathbf{x})}{\partial x_j}=\left(\sum_{i=1}^{s}\eta_i\frac{\partial T_i(\mathbf{x})}{\partial x_j}+\frac{\partial\log h(\mathbf{x})}{\partial x_j}\right)p_{\boldsymbol{\eta}}(\mathbf{x})$$

$$\Rightarrow\left(\sum_{i=1}^{s}\eta_i\frac{\partial T_i(\mathbf{x})}{\partial x_j}\right)p_{\boldsymbol{\eta}}(\mathbf{x})=\frac{\partial p_{\boldsymbol{\eta}}(\mathbf{x})}{\partial x_j}-\frac{\partial\log h(\mathbf{x})}{\partial x_j}p_{\boldsymbol{\eta}}(\mathbf{x}).$$

Therefore,

$$E\left[\sum_{i=1}^{s}\eta_i\frac{\partial T(\mathbf{X})}{\partial X_i}\middle|\mathbf{X}\right]=\int\left(\sum_{i=1}^{s}\eta_i\frac{\partial T_i(\mathbf{x})}{\partial x_j}\right)\frac{p_{\boldsymbol{\eta}}(\mathbf{x})\pi(\eta)}{m(\mathbf{x})}d\eta$$

$$=\frac{1}{m(\mathbf{x})}\int\left(\sum_{i=1}^{s}\eta_i\frac{\partial T_i(\mathbf{x})}{\partial x_j}\right)p_{\boldsymbol{\eta}}(\mathbf{x})\pi(\eta)d\eta$$

$$=\frac{1}{m(\mathbf{x})}\int\frac{\partial p_{\boldsymbol{\eta}}(\mathbf{x})}{\partial x_j}\pi(\eta)d\eta-\frac{1}{m(\mathbf{x})}\int\frac{\partial\log h(\mathbf{x})}{\partial x_j}p_{\boldsymbol{\eta}}(\mathbf{x})\pi(\eta)d\eta$$

$$=\frac{1}{m(\mathbf{x})}\frac{\partial}{\partial x_j}\int p_{\boldsymbol{\eta}}(\mathbf{x})\pi(\eta)d\eta-\frac{1}{m(\mathbf{x})}\frac{\partial\log h(\mathbf{x})}{\partial x_j}\int p_{\boldsymbol{\eta}}(\mathbf{x})\pi(\eta)d\eta$$

$$=\frac{1}{m(\mathbf{x})}\frac{\partial m(\mathbf{x})}{\partial x_j}-\frac{\partial\log h(\mathbf{x})}{\partial x_j}$$

$$=\frac{\partial\log m(\mathbf{x})}{\partial x_j}-\frac{\partial\log h(\mathbf{x})}{\partial x_j}.$$

(2) [+2] Let $T_i(\mathbf{x}) = x_i \quad i = 1, \ldots, p$. Express the Bayes estimator of η_i by $m(\mathbf{x})$, $h(\mathbf{x})$, and their derivatives.

Answer:

Since $T_i(\mathbf{x}) = x_i$, $i = 1, \ldots, p$, by the formula above, we have

$$E\left[\sum_{i=1}^{s} \eta_i \frac{\partial T(\mathbf{X})}{\partial X_i} \middle| \mathbf{X}\right] = \frac{\partial \log m(\mathbf{x})}{\partial x_j} - \frac{\partial \log h(\mathbf{x})}{\partial x_j}$$

$$\Rightarrow E[\eta_j \mid \mathbf{x}] = \frac{\partial \log m(\mathbf{x})}{\partial x_j} - \frac{\partial \log h(\mathbf{x})}{\partial x_j}.$$

Hence the Bayes estimator of η_i is

$$\frac{\partial \log m(\mathbf{x})}{\partial x_i} - \frac{\partial \log h(\mathbf{x})}{\partial x_i},$$

for $i = 1, \ldots, p$.

Q4 [+10]. Let $X \sim Bin(n,p)$ and $g(p) = p$ be the estimand.

1) [+3]Show that $\delta(X) = X/n$ is a minimax estimator under <u>some loss function</u> (also calculate the risk under the loss function).

<u>Answer:</u>

Let the loss function

$$L(\delta, p) = \frac{(p-\delta)^2}{p(1-p)}.$$

Suppose the prior distribution of p follows $U(0,1)$. The Bayes estimator is

$$\delta = \frac{\int_0^1 \frac{1}{p(1-p)} p \cdot p^x (1-p)^{n-x} dp}{\int_0^1 \frac{1}{p(1-p)} p^x (1-p)^{n-x} dp} = \frac{\int_0^1 p^{x+1-1}(1-p)^{n-x-1} dp}{\int_0^1 p^{x-1}(1-p)^{n-x-1} dp}$$

$$= \frac{\frac{\Gamma(x+1)\Gamma(n-x)}{\Gamma(n+1)}}{\frac{\Gamma(x)\Gamma(n-x)}{\Gamma(n)}} = \frac{x}{n}.$$

And the risk function is

$$R\left(p, \frac{x}{n}\right) = \frac{1}{p(1-p)} E\left(p - \frac{x}{n}\right)^2 = \frac{1}{p(1-p)n^2} E(x-np)^2 = \frac{np(1-p)}{p(1-p)n^2} = \frac{1}{n}.$$

The risk function does not depend on p. Therefore, $\delta(X) = X/n$ is a Bayes estimator with constant risk, that is, $\delta(X) = X/n$ is a minimax estimator.

2) [+3]Find a minimax estimator under the square loss $L(a, p) = | p - a |^2$.

Answer:

Let

$$p \sim \text{Beta}(a, b) \Rightarrow p \mid x \sim \text{Beta}(x + a, n - x + b).$$

Therefore, the Bayes estimator of p is

$$\delta_\Lambda = E(p \mid x) = \frac{x - a}{n + a + b}.$$

Then consider the risk function of δ_Λ

$$R(p, \delta_\Lambda) = E\left\{\left(\frac{x + a}{n + a + b} - p\right)^2\right\} = \frac{1}{(n + a + b)^2} E[\{(a + x) - (a + b + n)p\}^2]$$

$$= \frac{1}{(n + a + b)^2} E([(x - np) + \{a(1 - p) - bp\}]^2)$$

$$= \frac{1}{(n + a + b)^2} [np(1 - p) + \{a(1 - p) - bp\}^2].$$

Solve for

$$\frac{\partial R(p, \delta_\Lambda)}{\partial p} = 0 \Rightarrow n - 2np - 2(a + b)\{a - (a + b)p\} = 0$$

$$\Rightarrow \{2(a + b)^2 - 2n\}p + n - 2a(a + b) = 0.$$

That is

$$\begin{cases} (a + b)^2 = n \\ 2a(a + b) = n \end{cases} \Rightarrow \begin{cases} a = \sqrt{n}/2 \\ b = \sqrt{n}/2 \end{cases}$$

Thus, let

$$\delta_M = \frac{x - \sqrt{n}/2}{n + \sqrt{n}}.$$

Since δ_M is a Bayes estimator with constant risk then it is a minimax estimator.

3) [+4]Is the following estimator minimax under the square loss $L(a,p) = |p-a|^2$?

$$\delta(X) = \begin{cases} X/n & \text{with probability } n/(n+1) \\ 1/2 & \text{with probability } 1/(n+1) \end{cases}$$

[Hint: calculate the risk]

Answer:

$$R(p,\delta) = E(\delta - p)^2 = E\left(\frac{x}{n} - p\right)^2 \frac{n}{n+1} + E\left(\frac{1}{2} - p\right)^2 \frac{1}{n+1}$$

$$= \frac{E(x-np)^2}{n(n+1)} + \frac{p^2 - p + \frac{1}{4}}{n+1} = \frac{p-p^2}{n+1} + \frac{p^2 - p + \frac{1}{4}}{n+1} = \frac{1}{4(n+1)}.$$

Consider the minimax estimator δ_M in the previous problem, that is

$$\delta_M = \frac{x - \sqrt{n}/2}{n + \sqrt{n}}.$$

The risk function of δ_M is

$$R(p,\delta_M) = \frac{1}{(n+\sqrt{n})^2}\left[np(1-p) + \left\{ \frac{\sqrt{n}}{2}(1-p) - \frac{\sqrt{n}}{2}p \right\}^2 \right]$$

$$= \frac{1}{4(n+\sqrt{n})^2}\{ 4np(1-p) + n(1-2p)^2 \}$$

$$= \frac{1}{4(n+\sqrt{n})^2}\{ 4np - 4np^2 + 4np^2 + n - 4np \}$$

$$= \frac{n}{4(n+\sqrt{n})^2} = \frac{1}{4(1+\sqrt{n})^2}.$$

Since

$$R(p,\delta) = \frac{1}{4n+4} > \frac{1}{4n+4+8\sqrt{n}} = R(p,\delta_M).$$

Hence δ is not a minimax estimator.

Q5 [+10]. Let $X_1,..,X_n \overset{iid}{\sim} f(x|\theta)$, $\theta \in \Omega \subset R$, where the usual regularity conditions are assumed. Let $\tilde{\theta}_n$ be an initial estimator of θ.

1. [+3] Define and explain what is the one-step estimator

Answer:

Assume that $\ell'(\hat{\theta}_n) = 0$, that is, $\hat{\theta}_n$ is the MLE of θ. Then the one-step estimator is

$$\hat{\theta}_n \approx \tilde{\theta}_n - \frac{\ell'(\tilde{\theta}_n)}{\ell''(\tilde{\theta}_n)}.$$

This can be derived by a Taylor expansion (this can also be explained by slope),

$$\ell'(\hat{\theta}_n) \approx \ell'(\tilde{\theta}_n) + (\hat{\theta}_n - \tilde{\theta}_n)\ell''(\tilde{\theta}_n) \overset{\text{set}}{=} 0 \Rightarrow \hat{\theta}_n = \tilde{\theta}_n - \frac{\ell'(\tilde{\theta}_n)}{\ell''(\tilde{\theta}_n)}.$$

If $\tilde{\theta}_n$ is close to $\hat{\theta}_n$ then approximate $\hat{\theta}_n$ by the one-step estimator.

2. [+7]Prove the asymptotic efficiency of the one-step estimator. In the proof, please explain what conditions are used.

Answer:

Suppose initial estimator $\tilde{\theta}_n$ is $\sqrt{n}$-consistent, that is,

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) = O_p(1),$$

where θ_0 is the true parameter. Then the one-step estimator is

$$\delta_n = \tilde{\theta}_n - \frac{\ell'(\tilde{\theta}_n)}{\ell''(\tilde{\theta}_n)}.$$

By a Taylor expansion, we have

$$\ell'(\tilde{\theta}_n) = \ell'(\theta_0) + (\tilde{\theta}_n - \theta_0)\ell''(\theta_0) + \frac{1}{2}(\tilde{\theta}_n - \theta_0)^2 \ell'''(\theta_n^*),$$

where θ_n^* is between $\tilde{\theta}_n$ and θ_0.

By the definition of one-step estimator, we have

$$\sqrt{n}(\delta_n - \theta_0) = \sqrt{n}(\tilde{\theta}_n - \theta_0) - \frac{\frac{1}{\sqrt{n}}\ell'(\tilde{\theta}_n)}{\frac{1}{n}\ell''(\tilde{\theta}_n)}$$

$$= \sqrt{n}(\tilde{\theta}_n - \theta_0) - \frac{\frac{1}{\sqrt{n}}\left\{\ell'(\theta_0) + (\tilde{\theta}_n - \theta_0)\ell''(\theta_0) + \frac{1}{2}(\tilde{\theta}_n - \theta_0)^2\ell'''(\theta_n^*)\right\}}{\frac{1}{n}\ell''(\tilde{\theta}_n)}$$

$$= \frac{\frac{1}{\sqrt{n}}\ell'(\theta_0)}{-\frac{1}{n}\ell''(\tilde{\theta}_n)} + \sqrt{n}(\tilde{\theta}_n - \theta_0)\left\{1 - \frac{\ell''(\theta_0)}{\ell''(\tilde{\theta}_n)} - \frac{1}{2}\frac{(\tilde{\theta}_n - \theta_0)\ell'''(\theta_n^*)}{\ell''(\tilde{\theta}_n)}\right\}.$$

Since we have,

$$E\left(\frac{\partial}{\partial\theta}\log f(x|\theta_0)\right) = 0,$$

and

$$\text{var}\left(\frac{\partial}{\partial\theta}\log f(x|\theta_0)\right) = E\left(\frac{\partial}{\partial\theta}\log f(x|\theta_0)\right)^2 = I(\theta_0).$$

By the Central Limit Theorem,

$$\frac{1}{\sqrt{n}}\ell'(\theta_0) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\log f(x|\theta_0) - 0\right) \xrightarrow{d} N(0, I(\theta_0)), \text{ as } n \to \infty.$$

By the Weak Law of Large Number,

$$-\frac{1}{n}\ell''(\tilde{\theta}_n) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log f(x|\theta_0) \xrightarrow{p} -E\left(\frac{\partial^2}{\partial\theta^2}\log f(x|\theta_0)\right) = I(\theta_0),$$

as $n \to \infty$. By the Slusky's Theorem,

$$\frac{\frac{1}{\sqrt{n}}\ell'(\theta_0)}{-\frac{1}{n}\ell''(\tilde{\theta}_n)} \xrightarrow{d} N(0, I^{-1}(\theta_0)), \text{ as } n \to \infty.$$

By assumption ($\tilde{\theta}_n$ is $\sqrt{n}$-consistent),

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) = O_p(1), \text{ and } \tilde{\theta}_n - \theta_0 \xrightarrow{p} 0, \text{ as } n \to \infty.$$

Since

$$\frac{1}{n}\ell''(\tilde{\theta}_n)=\frac{1}{n}\ell''(\theta_0)+\frac{1}{n}(\tilde{\theta}_n-\theta_0)\ell'''(\theta_n^{**}),$$

where θ_n^{**} is between $\tilde{\theta}_n$ and θ_0. Therefore, we have

$$\frac{\ell''(\theta_0)}{\ell''(\tilde{\theta}_n)}\xrightarrow{p}1,\text{ as }\ n\to\infty.$$

By the regularity condition provided that $\ell'''(\theta_n^*)$ is bounded, we have

$$\frac{\ell'''(\theta_n^*)}{\ell''(\tilde{\theta}_n)}\xrightarrow{p}\text{constant},\text{ as }\ n\to\infty.$$

Then combining the results above, by the Slusky's Theorem, we have

$$\sqrt{n}(\delta_n-\theta_0)\xrightarrow{d}N(0,I^{-1}(\theta_0))+O_p(1)\times\left(1-1-\frac{1}{2}\cdot 0\cdot 0\right)=N(0,I^{-1}(\theta_0)),$$

as $n\to\infty$. Hence the one-step estimator is efficient. □