ST5215 Advanced Statistical Theory, Lecture 25

HUANG Dongming

National University of Singapore

12 Nov 2020

Overview

Last time

- Properties of LSE without normality
- Consistency of LSE

Today

- Asymptotic normality of LSE
- Bayes estimator and Minimaxity
- Empirical Bayes estimator and Inadmissiblity

Recap: BLUE and L_2 consistency

$$X = Z\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

- A *linear estimator* is a linear function of X, i.e., $\mathbf{c}^\top X$ for some fixed vector $\mathbf{c}$.
- Suppose $\ell \in \mathcal{R}(Z)$, the *best linear unbiased estimator* (BLUE) of $\ell^\top \boldsymbol{\beta}$ is $\ell^\top \hat{\boldsymbol{\beta}}$ if
  - $\mathrm{Var}(\epsilon) = \sigma^2 I$
  - $\mathrm{Var}(\epsilon) = \Sigma$ and $\mathbf{P}_Z \Sigma = \Sigma \mathbf{P}_Z$ (Robustness)
- If Z is of full rank and $\boldsymbol{\epsilon} \sim N_n(0, \Sigma)$, where Σ is an unknown positive definite matrix, then $\ell^\top \hat{\boldsymbol{\beta}}$ is a UMVUE of $\ell^\top \boldsymbol{\beta}$ for any $\ell \in \mathcal{R}^p$ if $\mathbf{P}_Z \Sigma = \Sigma \mathbf{P}_Z$
- If $\sup_n \|\Sigma\| < \infty$ and $\|(Z^\top Z)^-\| \to 0$, then $\ell^\top \hat{\boldsymbol{\beta}}$ is L_2-consistent for $\ell^\top \boldsymbol{\beta}$ for any $\ell \in \mathcal{R}(Z)$

Asymptotic Normality of LSE

- Consider model $X = Z\boldsymbol{\beta} + \epsilon$ under assumption A3
- Assume that ϵ_i's are independent with variances σ_i^2's
- For $\ell \in \mathcal{R}(Z)$, the asymptotic behavior of $\ell^\top(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is determined by $\ell^\top A_n Z^\top \epsilon$, where $A_n = (Z^\top Z)^-$

Theorem

Suppose that $\tau := \inf_n \sigma_n^2 > 0$ *and that*

$$\lim_{n\to\infty} \max_{1\le i\le n} Z_i^\top A_n Z_i = 0. \tag{2}$$

If (a) $\sup_i E|\epsilon_i|^{2+\delta} < \infty$ *or (b) if* ϵ_i*'s have the same distribution, then for any* $\ell \in \mathcal{R}(Z)$,

$$\ell^\top(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Big/ \sqrt{\mathrm{Var}(\ell^\top \hat{\boldsymbol{\beta}})} \to_d N(0,1). \tag{3}$$

Proof

- Since $\ell \in \mathcal{R}(Z)$,

$$\ell^\top (Z^\top Z)^- Z^\top Z\boldsymbol{\beta} - \ell^\top \boldsymbol{\beta} = 0$$

- Let $c_n = Z(Z^\top Z)^-\ell$, then

$$\ell^\top(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \ell^\top (Z^\top Z)^- Z^\top \epsilon = \sum_{j=1}^n c_{n,j}\epsilon_j$$

- Note that

$$\|c_n\|^2 = \left(Z(Z^\top Z)^-\ell\right)^\top Z(Z^\top Z)^-\ell = \ell^\top (Z^\top Z)^-\ell$$

  and

$$\max_{1\le j\le n} |c_{n,j}|^2 \le \max_{1\le j\le n} [\ell^\top (Z^\top Z)^- Z_j]^2 \le \ell^\top (Z^\top Z)^-\ell \max_{1\le j\le n} Z_j^\top (Z^\top Z)^- Z_j.$$

- These together with Condition (2) imply that

$$\lim_{n\to\infty}\left(\max_{1\le j\le n}|c_{n,j}|^2 \Big/ \|c_n\|^2\right) = 0.$$

Proof (Cont.)

The rest is to check the Lindeberg's condition holds for the triangular array $\{c_{n,j}\epsilon_j : 1 \le j \le n, n = 1, 2, \ldots\}$.

(a). Suppose $K = \sup_i E|\epsilon_i|^{2+\delta} < \infty$

- The row sum of variance is $s_n^2 = \sum_{j=1}^n c_{n,j}^2 \mathrm{Var}(\epsilon_j) \ge \|c_n\|^2 \tau$
- For any constant $a > 0$,

$$E\left(|c_{n,j}\epsilon_j|^2 I_{|c_{n,j}\epsilon_j| > as_n}\right) \le \frac{1}{a^\delta s_n^\delta} E|c_{n,j}\epsilon_j|^{2+\delta} \le \frac{c_{n,j}^{2+\delta}}{a^\delta s_n^\delta} K$$

- $\frac{1}{s_n^2} \sum_{j\le n} E\left(|c_{n,j}\epsilon_j|^2 I_{|c_{n,j}\epsilon_j| > as_n}\right) \le \frac{\sum_{j\le n} c_{n,j}^{2+\delta}}{s_n^{2+\delta}} \frac{K}{a^\delta} \le \frac{\sum_{j\le n} c_{n,j}^{2+\delta}}{\|c_n\|^{2+\delta}} \frac{K}{\tau a^\delta}$
- Note that $\sum_{j\le n} c_{n,j}^{2+\delta} \le \sum_{j\le n} c_{n,j}^2 \left(\max_{j\le n} c_{n,j}\right)^\delta$, we have

$$\frac{\sum_{j\le n} c_{n,j}^{2+\delta}}{\|c_n\|^{2+\delta}} \le \left(\frac{\max_{j\le n}|c_{n,j}|}{\|c_n\|}\right)^\delta = o(1)$$

(b). Suppose ϵ_i's have the same distribution. Then $\mathrm{Var}(\epsilon_j) = \tau$ for all j

- The row sum of variance is $s_n^2 = \sum_{j=1}^n c_{n,j} \mathrm{Var}(\epsilon_j) = \|c_n\|^2 \tau$
- For any constant $a > 0$,

$$\begin{aligned}
&\frac{1}{s_n^2} \sum_{j \le n} E\left(|c_{n,j}\epsilon_j|^2 I_{|c_{n,j}\epsilon_j| > a s_n} \right) \\
&= \frac{1}{\tau \|c_n\|^2} \sum_j c_{n,j}^2 E|\epsilon|^2 I_{|c_{n,j}\epsilon| > a s_n} \\
&\le \frac{\sum_j c_{n,j}^2}{\tau \|c_n\|^2} \max_{j \le n} E|\epsilon|^2 I_{|c_{n,j}\epsilon| > a s_n} \\
&= \tau^{-1} E|\epsilon|^2 I\{ \left(\max_{j \le n} c_{n,j}^2 / \|c_n\|^2 \right)^{1/2} |\epsilon| > a \},
\end{aligned}$$

which is $o(1)$ because $\max_{j \le n} c_{n,j}^2 / \|c_n\|^2 = o(1)$ and by the DCT

Remarks

- Theorem 3.12 of the textbook is a more general result that allows the noise to be block-wise independent. Its proof uses Corollary 1.3.
- Condition (2) is almost necessary for the consistency of the LSE (choose ℓ to be Z_i the ith row of Z)

Proposition (Exercise 3.6.80)

For any fixed $i \le n$, let $\hat{X}_i^{(n)} = Z_i^\top \hat{\beta}$ and $h_i^{(n)} = Z_i^\top (Z^\top Z)^- Z_i$. Suppose assumption A2 holds.

(a) *For any $\delta > 0$,*

$$P(|\hat{X}_i^{(n)} - E\hat{X}_i^{(n)}| \ge \delta) \ge \min\{P(\varepsilon_i \ge \delta/h_i^{(n)}), P(\varepsilon_i \le -\delta/h_i^{(n)})\}. \quad (4)$$

(b) $\hat{X}_i - E\hat{X}_i \xrightarrow{P} 0$ *if and only if* $h_i \to 0$.

If U and V are independent, then for $\epsilon > 0$,

$$P(|U + Y| \ge \epsilon) \ge P(U \ge \epsilon)P(Y \ge 0) + P(U \le -\epsilon)P(Y < 0)$$

Checking Condition (2)

Lemma (Lemma 3.3 of the textbook)

Each of the followings implies $\lim_{n\to\infty} \max_{1\le i\le n} \left(Z_i^\top (Z^\top Z)^- Z_i\right) = 0$

(a) $\lambda_+[(Z^\top Z)^-] \to 0$ *and* $Z_n^\top (Z^\top Z)^- Z_n \to 0$*, as* $n \to \infty$*.*

(b) *There is an increasing sequence* $\{a_n\}$ *such that* $a_n \to \infty$*,* $a_n/a_{n+1} \to 1$*, and* $Z^\top Z/a_n$ *converges to a positive definite matrix.*

- Denote $(Z^\top Z)^-$ by A_n and $h_i^{(n)} = Z_i^\top A_n Z_i$
- Let j_n be the integer such that $h_{j_n}^{(n)} = \max_{1\le j\le n} h_j^{(n)}$. Suppose $\{i_n\}$ is a subsequence of $\{j_n\}$ such that $h_{i_n}^{(n)}$ does not converge to 0
- If $\sup_n i_n \le c < \infty$, then

$$h_{i_n}^{(n)} = Z_{i_n}^\top A_n Z_{i_n} \le \max_{1\le i\le c} Z_i^\top A_n Z_i \le \lambda_+(A_n) \max_{1\le i\le c} \|Z_i\|^2 \to 0.$$

- If i_n is unbounded then it has a subsequence $\{k_n\}$ that goes to ∞, and

$$\lim_{n\to\infty} h_{k_n}^{(n)} = \lim_{n\to\infty} Z_{k_n}^\top A_n Z_{k_n} \le \lim_{n\to\infty} Z_{k_n}^\top A_{k_n} Z_{k_n} = 0,$$

because $A_{k_n} \succ A_n$. Either leads to a contradiction.

In the last slide, we have use the following
For any n,

$$Z^{(n),\top} Z^{(n)} = \sum_{j=1}^{n} Z_j Z_j^{\top} \succ \sum_{j=1}^{n-1} Z_j Z_j^{\top} = Z^{(n-1),\top} Z^{(n-1)}$$

so

$$A_{n-1} \succ A_n$$

Remark. Note that in this course, a vector is by default a column vector. The notation Z_j here is the transpose of the jth row of Z.

Example: Simple linear models

- In a simple linear model,

$$X_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 t_i + \epsilon_i, \quad i = 1, ..., n. \tag{5}$$

- In this case, $Z_i = (1, t_i)^\top, t_i \in \mathcal{R}, i = 1, \ldots, n.$ and

$$Z^\top Z = \begin{pmatrix} n & \sum_{i=1}^n t_i \\ \sum_{i=1}^n t_i & \sum_{i=1}^n t_i^2 \end{pmatrix},$$

- If $n^{-1} \sum_{i=1}^n t_i^2 \to c$ and $n^{-1} \sum_{i=1}^n t_i \to d$, then

$$n^{-1} Z^\top Z \to \begin{pmatrix} 1 & d \\ d & c \end{pmatrix}.$$

- The limit is positive definite iff $c > d^2$, in which case condition (b) in Lemma 3.3 is satisfied with $a_n = n$ and Theorem 3.12 applies.

Example: One-way ANOVA

- In the one-way ANOVA model (Example 3.13),

$$X_i = \mu_j + \epsilon_i, \qquad i = k_{j-1} + 1, ..., k_j, \; j = 1, ..., m, \tag{6}$$

where $k_0 = 0$, $k_j = \sum_{l=1}^{j} n_l$, $j = 1, ..., m$, and $(\mu_1, ..., \mu_m) = \beta$, and

$$Z = \begin{pmatrix} J_{n_1} & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & J_{n_m} \end{pmatrix},$$

where J_k is the k-vector of ones.

- Since $Z^\top Z = \mathrm{Diag}(n_1, \ldots, n_m)$, $(Z^\top Z)^{-1} = \mathrm{Diag}(n_1^{-1}, \ldots, n_m^{-1})$

$$\max_{1 \le i \le n} Z_i^\top (Z^\top Z)^{-} Z_i = \max_{1 \le j \le m} n_j^{-1}. \tag{7}$$

- Conditions related to Z in Theorem 3.12 are satisfied iff all $n_j \to \infty$.

Bayes Estimator

- Apart from MOM estimators, MLEs, and UMVUEs by Rao-Blackwellization, a popular estimator is the Bayes estimator
- Recall that a *Bayes rule* w.r.t. π is a rule that minimizes the Bayes risk $r_T(\pi) = \int_\Theta R_T(\theta)\, \mathrm{d}\pi(\theta)$ for any rule T, where π is a distribution on Θ.
- In an estimation problem, a Bayes rule is called a *Bayes estimator*.

Suppose $r_T(\theta) = E_\theta L(\theta, T)$.

- We can introduce two random elements $\tilde{\theta} \sim \pi$, and $X \mid \tilde{\theta} \sim P_{\tilde{\theta}}$, so that $r_T(\pi) = E\left[L(\tilde{\theta}, T(X))\right]$
- A Bayes estimator T_* can be obtained by minimizing the conditional risk

$$T_*(x) := \arg\min_a E\left[L(\tilde{\theta}, a) \mid X = x\right]$$

Example: Normal Priors for Normal Distributions

- Suppose $X \sim N(\theta, 1)$
- To find an estimator of θ, we postulate a prior distribution $\theta \sim \pi_\tau = N(0, \tau)$ and consider the squared error $L(\theta, a) = (\theta - a)^2$
- The posterior distribution is $\theta \mid X \sim N(\frac{\tau X}{1+\tau}, \frac{\tau}{1+\tau})$
- Note that $E\left[L(\theta, a)^2 \mid X\right]$ is minimized at $a = E(\theta \mid X)$, so a Bayes estimator is given by $\frac{\tau X}{1+\tau}$, whose Bayes risk is $\tau/(1+\tau)$.
- If τ is very large, the prior distribution is almost flat; it treats every possible value of θ equally and $\frac{\tau X}{1+\tau} \approx X$
- If τ is very close to 0, it indicates a strong preference towards small values of θ and the Bayes estimator shrinks the natural estimator X toward the prior mean 0

Bayes estimators with constant risk

Theorem (4.11 in the textbook)

If T is a Bayes estimator w.r.t. π and if

$$R_T(\theta) = \sup_{\theta'} R_T(\theta') \quad \pi\text{-a.e.}, \tag{8}$$

then T is minimax. If T is the unique Bayes estimator w.r.t. π then it is the unique minimax estimator.

- Let S be any other estimator of ϑ. Then

$$\sup_{\theta \in \Theta} R_S(\theta) \geq \int_{\Theta_\pi} R_S(\theta) d\pi \geq \int_{\Theta_\pi} R_T(\theta) d\pi = \sup_{\theta \in \Theta} R_T(\theta),$$

  where Θ_π is the complement of the null set in Eq (8).
- For the uniqueness, the second inequality becomes strict.

Example (Cont.)

Suppose $X \sim N(\theta, 1)$. We will show that X is minimax for estimating θ

- For a prior distribution $\theta \sim \pi_\tau = N(0, \tau)$, the Bayes estimator is $\delta_\tau = \frac{\tau X}{1+\tau}$ and its Bayes risk is $r_*(\pi_\tau) = \tau/(1+\tau)$.
- Note that $X = \lim_{\tau \to \infty} \frac{\tau X}{1+\tau}$ and

$$E(X-\theta)^2 = 1 = \lim_{\tau \to \infty} r_*(\pi_\tau).$$

- For any other estimator S,

$$\sup_{\theta \in \Theta} R_S(\theta) \geq \int R_S(\theta) d\pi_\tau \geq \int R_{\delta_\tau}(\theta) d\pi_\tau = r_*(\pi_\tau).$$

- Take $\tau \to \infty$, we have $\sup_{\theta \in \Theta} R_S(\theta) \geq R_X(\theta)$ for any θ

Remark. Generally, if T has a constant risk that equals to the limit of the optimal Bayes risks for a sequence of prior distributions $\{\pi_j\}_{j=1}^{\infty}$, then T is minimax (Theorem 4.12 in the textbook)

Example with unknown variance

If $X \sim N(\mu, \sigma^2)$ where $(\mu, \sigma^2) \in \mathcal{R} \times (0, c]$, then X is minimax for estimating μ

- If σ^2 is known, then X is minimax and has risk σ^2
- For any other estimator T, note that

$$\begin{aligned}
\sup_{(\mu,\sigma^2)\in\mathcal{R}\times(0,c]} R_T(\mu,\sigma^2) &\geq \sup_{(\mu,\sigma^2)\in\mathcal{R}\times\{c\}} R_T(\mu,\sigma^2) \\
&\geq \sup_{(\mu,\sigma^2)\in\mathcal{R}\times\{c\}} R_X(\mu,\sigma^2) \\
&= c \\
&\geq \sup_{(\mu,\sigma^2)\in\mathcal{R}\times(0,c]} R_X(\mu,\sigma^2)
\end{aligned}$$

Remark. Note that the minimaxity of X for any fixed σ^2 implies that

$$\sup_{(\mu,\sigma^2)\in\mathcal{R}\times(0,\infty)} R_T(\mu,\sigma^2) = \sup_{(\mu,\sigma^2)\in\mathcal{R}\times(0,\infty)} \sigma^2 = \infty, \qquad \forall T$$

so for $(\mu, \sigma^2) \in \mathcal{R} \times (0, \infty)$, minimaxity is meaningless

Empirical Bayes and Simultaneous Estimation

- The prior distribution used in the derivation of a Bayes estimator often depends on some parameters called *hyperparameters*
- If hyperparameters are unknown, we can estimate them using data. The resulting Bayes estimator is called *an empirical Bayes estimator*

Example:

- We have p different observations, each $X_i \sim N(\mu_i, 1)$
- Estimating μ_i's together is an example of *simultaneous estimation*
- Suppose a joint prior distribution is used: μ_i are i.i.d. from $N(0, \tau)$
- Here τ is the hyperparameter
- τ can be estimated by $\hat{\tau} = p^{-1} \sum_{i=1}^{p} X_i^2 - 1$ because integrating out μ_i's, X_i's is i.i.d. from $N(0, 1 + \tau)$
- The Bayes estimator of μ_i w.r.t. $\hat{\tau}$ is $\hat{\mu}_i^{EB} = (1 - \frac{1}{1+\hat{\tau}})X_i$ or

$$\hat{\mu}^{EB} = (1 - \frac{p}{\|X\|^2})X$$

Admissibility in Simultaneous Estimation

- Although $\hat{\mu}^{EB}$ is motivated from the Bayesian framework, it enjoys good frequentist's properties
- Using the technique in Exercise 3 of Tutorial 24, one can show that $R_{\hat{\mu}^{EB}}(\mu) = p - p(p-4)E\frac{1}{\|X\|^2}$, which is strictly smaller than $R_X(\mu)$ for all $\mu \in \mathcal{R}^p$ if $p > 4$
  - See Theorem 4.15 in the textbook for a general result.
- For $p \geq 3$, the *James-Stein estimator* and its improvement dominates X

$$\delta_{JS} = (1 - \frac{p-2}{\|X\|^2})X, \qquad \delta_{JS}^{+} = \min(0, 1 - \frac{p-2}{\|X\|^2})X.$$

- Above is surprising: although X is the MLE, the UMVUE, and also the minimax estimator of μ (extension of page 16), X is inadmissible
  - X is admissible if $p = 1$ and $p = 2$ (Stein, 1956).
  - δ_{JS} and δ_{JS}^{+} are also inadmissible, but can be substantially better than X if p is large and $\|\mu\|$ is close to 0

Tutorial

1. Let $X_1, \ldots, X_n$ be i.i.d. from $\mathrm{E}(0, \theta)$, where $\theta > 0$ is unknown. Let $\hat{p}_n = \# \{ i \leq n : X_i \geq 1 \} / n$ and $\bar{X}_n = (X_1 + \cdots + X_n) / n$. Determine the asymptotic relative efficiency of $-\log \hat{p}_n$ with respect to $1/\bar{X}_n$ for estimating $1/\theta$.
2. Let $X_1, \ldots, X_n$ be i.i.d. from $\mathrm{N}(\mu, 1)$, where $\mu > 0$ is unknown. Consider estimating μ by $\bar{X}_n = (X_1 + \cdots + X_n) / n$ and the sample median $m_n = X_{(\lfloor n/2 \rfloor)}$. Determine the asymptotic relative efficiency of $\bar{X}_n$ w.r.t. m_n.
3. Exercise 2.6.118 in JS
4. Exercise 4.6.112 in JS

Exercise 1

Let $X_1, \ldots, X_n$ be i.i.d. from $\mathrm{E}(0, \theta)$, where $\theta > 0$ is unknown. Let $\hat{p}_n = \#\{i \le n : X_i \ge 1\}/n$ and $\bar{X}_n = (X_1 + \cdots + X_n)/n$. Determine the asymptotic relative efficiency of $-\log \hat{p}_n$ with respect to $1/\bar{X}_n$ for estimating $1/\theta$.

Proof:

- By the CLT, $\sqrt{n}\left(\bar{X}_n - \theta\right) \xrightarrow{\mathcal{D}} N\left(0, \theta^2\right)$.
- Using δ-method with $f(x) = 1/x$, we have $\sqrt{n}\left(1/\bar{X}_n - 1/\theta\right) \xrightarrow{\mathcal{D}} N\left(0, 1/\theta^2\right)$.
- By the CLT for $I_{X_i \ge 1}$,
$$\sqrt{n}\left(\hat{p}_n - e^{-1/\theta}\right) \xrightarrow{\mathcal{D}} N\left[0, e^{-1/\theta}\left(1 - e^{-1/\theta}\right)\right],$$
because $I_{X_i \ge 1} \sim \mathrm{Bern}(e^{-1/\theta})$.
- Using δ-method with $f(x) = -\log x$, we have
$$\sqrt{n}\left(-\log \hat{p}_n - 1/\theta\right) \xrightarrow{\mathcal{D}} N\left(0, e^{1/\theta} - 1\right),$$
because $f(e^{-1/\theta}) = 1/\theta$ and $f'(e^{-1/\theta}) = e^{1/\theta}$
- So the asymptotic relative efficiency is $\theta^{-2}\left(e^{1/\theta} - 1\right)^{-1}$.

Exercise 2

Let $X_1, \ldots, X_n$ be i.i.d. from $\mathrm{N}(\mu, 1)$, where $\mu > 0$ is unknown. Consider estimating μ by $\bar{X}_n = (X_1 + \cdots + X_n)/n$ and the sample median $m_n = X_{(\lfloor n/2 \rfloor)}$. Determine the asymptotic relative efficiency of $\bar{X}_n$ w.r.t. m_n.

Proof:

- By the CLT, $\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{\mathcal{D}} N(0, 1)$.
- By the asymptotic normality of sample median, we have

$$\sqrt{n}\left(m_n - \mu\right) \xrightarrow{\mathcal{D}} N(0, \frac{2\pi}{4}\right),$$

  because the p.d.f. at $x = \mu$ is $1/\sqrt{2\pi}$.
- The asymptotic relative efficiency is $\frac{\pi}{2} > 1$

Exercise 2.6.118

Let $X_1, \ldots, X_n$ be i.i.d. from the $N\left(0, \sigma^2\right)$ distribution with an unknown $\sigma > 0$. Consider the estimation of $\vartheta = \sigma$. Find the asymptotic relative efficiency of $\sqrt{\pi/2} \sum_{i=1}^n |X_i| / n$ w.r.t. $\left(\sum_{i=1}^n X_i^2 / n\right)^{1/2}$

Proof:

- Since $E\left(\sqrt{\pi/2}\,|X_1|\right) = \sigma$ and $\operatorname{Var}\left(\sqrt{\pi/2}\,|X_1|\right) = \left(\frac{\pi}{2} - 1\right)\sigma^2$, by the central limit theorem, we obtain that

$$\sqrt{n}\,(T_{1n} - \sigma) \xrightarrow{\mathcal{D}} N\left(0, \left(\frac{\pi}{2} - 1\right)\sigma^2\right)$$

- Since $EX_1^2 = \sigma^2$ and $\operatorname{Var}(X_1) = 2\sigma^4$, by CLT,

$$\sqrt{n}\left(n^{-1}\sum_{i=1}^n X_i^2 - \sigma^2\right) \xrightarrow{\mathcal{D}} N\left(0, 2\sigma^4\right)$$

- By the δ-method with $g(t) = \sqrt{t}$ and $g'(t) = (2\sqrt{t})^{-1}$,

$$\sqrt{n}\,(T_{2n} - \sigma) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{2}\sigma^2\right)$$

Exercise 4.6.112

Let $X_1, \ldots, X_n$ be i.i.d. from the uniform distribution $U(0, \theta)$, where $\theta > 0$ is unknown. Let $\hat{\theta}$ be the MLE of θ and T be the UMVUE.
(a) Obtain the ratio $\mathrm{mse}_T(\theta)/\mathrm{mse}_{\hat{\theta}}(\theta)$ and show that the MLE is inadmissible when $n \geq 2$.
(b) Let $Z_{a,\theta}$ be a random variable having the exponential distribution $E(a, \theta)$. Prove $n(\theta - \hat{\theta}) \xrightarrow{\mathcal{D}} Z_{0,\theta}$ and $n(\theta - T) \xrightarrow{\mathcal{D}} Z_{-\theta,\theta}$. Obtain the asymptotic relative efficiency of $\hat{\theta}$ w.r.t. T.

Proof: Part (i): Let $X_{(n)}$ be the largest order statistic.

- $\hat{\theta} = X_{(n)}$ and $T(X) = \frac{n+1}{n} X_{(n)}$.
- The MSE of $\hat{\theta}$ is $E\left(X_{(n)} - \theta\right)^2 = \frac{2\theta^2}{(n+1)(n+2)}$
- The MSE of T is $E(T - \theta)^2 = \frac{\theta^2}{n(n+2)}$
- The ratio is $(n+1)/(2n) \leq 1$ for $n \geq 2$. Therefore, the MLE $\hat{\theta}$ is inadmissible.

(ii)

- Note that

$$
\begin{aligned}
P(n(\theta-\hat{\theta}) \leq x) &= P\left(X_{(n)} \geq \theta - \frac{x}{n}\right) \\
&= \theta^{-n} \int_{\theta - x/n}^{\theta} n t^{n-1} dt \\
&= 1 - \left(1 - \frac{x}{n\theta}\right)^n \\
&\to 1 - e^{-x/\theta}, \qquad \text{as } n \to \infty.
\end{aligned}
$$

  We conclude that $n(\theta-\hat{\theta}) \xrightarrow{\mathcal{D}} Z_{0,\theta}$.

- Note that

$$
n(\theta - T) = n(\theta - \hat{\theta}) - \hat{\theta}.
$$

- By Slutsky's theorem, we conclude that $n(\theta - T) \xrightarrow{\mathcal{D}} Z_{0,\theta} - \theta$, which has the same distribution as $Z_{-\theta,\theta}$.
- The asymptotic relative efficiency of $\hat{\theta}$ with respect to T is $E\left(Z^2_{-\theta,\theta}\right)/E\left(Z^2_{0,\theta}\right) = \theta^2/\left(\theta^2+\theta^2\right) = \frac{1}{2}$

Final Exam

What will not be asked

- Questions directly on probability theory such as the Q(3,4,6) in Homework 3; but probability will be used almost everywhere!
- The last part: Minimaxity and Bayes Estimator, empirical Bayes, Jame-Stein estimator

What you need to pay attention to

- Conditional expectation, Integration, SLLN, CLT, δ-method, USLLN
- Sufficiency, Completeness, UMVUE, Fisher information and C-R lower bound
- Consistency of Estimators
- Asymptotic Distributions of Estimators
- Linear model (assumptions and properties)