ST5215 Advanced Statistical Theory, Lecture 24

HUANG Dongming

National University of Singapore

10 Nov 2020

Overview

Last time

- Properties of LSE under Normality

Today

- Properties of LSE without normality
- Consistency of LSE

Recap: Assumptions and Estimability

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

A1: (Gaussian noise) $\epsilon \sim N_n(0, \sigma^2 I_n)$ with an unknown $\sigma^2 > 0$.

A2: (homoscedastic noise) $E(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = \sigma^2 I_n$ with an unknown $\sigma^2 > 0$.

A3: (general noise) $E(\epsilon) = 0$ and $\mathrm{Var}(\epsilon)$ is an unknown matrix.

Theorem

Assume model (1).

(i) *A necessary and sufficient condition for $\ell \in \mathcal{R}^p$ being $Q^\top c$ for some $c \in \mathcal{R}^r$ is $\ell \in \mathcal{R}(Z) = \mathcal{R}(Z^\top Z)$, where r is the rank of Z and Q is given in $Z = UQ$ for $Q \in \mathcal{R}^{r \times p}$.*

(ii) *Under assumption A3, if $\ell \in \mathcal{R}(Z)$, then the LSE $\ell^\top \hat{\boldsymbol{\beta}}$ is unique and unbiased for $\ell^\top \boldsymbol{\beta}$.*

(iii) *Under assumption A1, if $\ell \notin \mathcal{R}(Z)$, then $\ell^\top \boldsymbol{\beta}$ is not estimable.*

Recap: Properties Under Nomrality

Theorem (Theorem 3.7, 3.8 of the textbook)

Assume model $X = Z\boldsymbol{\beta} + \epsilon$ with assumption A1: ϵ is distributed as $N_n(0, \sigma^2 I_n)$ with an unknown $\sigma^2 > 0$.

- (i) *The LSE $\ell^\top \hat{\boldsymbol{\beta}}$ is the UMVUE of $\ell^\top \boldsymbol{\beta}$ for any estimable $\ell^\top \boldsymbol{\beta}$.*
- (ii) *The UMVUE of σ^2 is $\hat{\sigma}^2 = (n-r)^{-1}\|X - Z\hat{\boldsymbol{\beta}}\|^2$, where r is the rank of Z.*
- (iii) *For any estimable parameter $\ell^\top \boldsymbol{\beta}$, the UMVUE's $\ell^\top \hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are independent; the distribution of $\ell^\top \hat{\boldsymbol{\beta}}$ is $N(\ell^\top \boldsymbol{\beta}, \sigma^2 \ell^\top (Z^\top Z)^- \ell)$; and $(n-r)\hat{\sigma}^2/\sigma^2$ has the chi-square distribution χ^2_{n-r}.*

Summary of (i) and (ii)

Under A1,

- $T = (Z^\top X, \|X - Z\hat{\boldsymbol{\beta}}\|^2)$ is complete and sufficient for $\theta = (\boldsymbol{\beta}, \sigma^2)$
- $\ell^\top \hat{\boldsymbol{\beta}}$ is unbiased for $\ell^\top \boldsymbol{\beta}$ and, hence, $\ell^\top \hat{\boldsymbol{\beta}}$ is the UMVUE of $\ell^\top \boldsymbol{\beta}$
- $\hat{\sigma}^2$ is the UMVUE of σ^2 because $E\hat{\sigma}^2 = (n-r)^{-1} E\|X - Z\hat{\boldsymbol{\beta}}\|^2 = \sigma^2$

Generally,

- The fitted vector $Z\hat{\boldsymbol{\beta}} = Z\left(Z^\top Z\right)^{-} Z^\top X = \mathbf{P}_Z X$
- The residual vector $X - Z\hat{\boldsymbol{\beta}} = X - \mathbf{P}_Z X = \mathbf{P}_{Z\perp} X$
- They are orthogonal: $\langle Z\hat{\boldsymbol{\beta}}, X - Z\hat{\boldsymbol{\beta}}\rangle = 0$ because $\mathbf{P}_Z \mathbf{P}_{Z\perp} = 0$
- Under assumption A1, they are jointly normally distributed and are independent

Proof of (iii)

Based on the last remark, we only need to find the distributions of $\ell^\top \hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$

- Since $\ell^\top \boldsymbol{\beta}$ is estimable, $\ell \in \mathcal{R}(Z)$.
- Since $Z\hat{\boldsymbol{\beta}}$ is normally distributed, so is $\ell^\top \hat{\boldsymbol{\beta}}$.
- Its mean is $\ell^\top \boldsymbol{\beta}$ and variance is $\sigma^2 \ell^\top \left(Z^\top Z\right)^- \ell$, so

$$\ell^\top \hat{\boldsymbol{\beta}} \sim N(\ell^\top \boldsymbol{\beta}, \sigma^2 \ell^\top (Z^\top Z)^- \ell)$$

- $X - Z\hat{\boldsymbol{\beta}} = \mathbf{P}_{Z\perp} X = \mathbf{P}_{Z\perp} Z\boldsymbol{\beta} + \mathbf{P}_{Z\perp}\epsilon = \mathbf{P}_{Z\perp}\epsilon$
- Since $\mathbf{P}_{Z\perp}$ is the projection matrix onto the orthogonal complement of $\mathcal{R}(Z)$, one can find a matrix $W \in \mathcal{R}^{n\times(n-r)}$ such that $W^\top W = \boldsymbol{I}_{n-r}$ and $\mathbf{P}_{Z\perp} = WW^\top$.
- Therefore $W^\top \epsilon \sim N(0, \sigma^2 I_{n-r})$ and

$$\mathrm{SSR} = \|X - Z\hat{\boldsymbol{\beta}}\|^2 = (\mathbf{P}_{Z\perp}\epsilon)^\top \mathbf{P}_{Z\perp}\epsilon = \epsilon^\top WW^\top \epsilon = \|W^\top \epsilon\|^2,$$

which implies that $(n-r)\hat{\sigma}^2/\sigma^2$ has the chi-square distribution χ^2_{n-r}

Best Linear Unbiased Estimator

- A *linear estimator* for the linear model

$$X = Z\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2}$$

  is a linear function of X, i.e., $\mathbf{c}^\top X$ for some fixed vector $\mathbf{c}$.
- For example, $\ell^\top \hat{\boldsymbol{\beta}}$ is a linear estimator, since $\ell^\top \hat{\boldsymbol{\beta}} = \ell^\top (Z^\top Z)^- Z^\top X$ with $\mathbf{c} = Z(Z^\top Z)^- \ell$.
- The variance of $\mathbf{c}^\top X$ is given by

$$\mathbf{c}^\top \mathrm{Var}(X)\mathbf{c} = \mathbf{c}^\top \mathrm{Var}(\boldsymbol{\epsilon})\mathbf{c}$$

- The *best linear unbiased estimator* (BLUE) of $\ell^\top \boldsymbol{\beta}$ is the linear estimator that achieves the minimum variance in the class of linear unbiased estimators of $\ell^\top \boldsymbol{\beta}$

Properties Under Assumption A2

Under assumption A2: $E(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = \sigma^2 I_n$

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

$$\mathrm{Var}(\ell^\top \hat{\boldsymbol{\beta}}) = \ell^\top (Z^\top Z)^- Z^\top \mathrm{Var}(\boldsymbol{\epsilon}) Z (Z^\top Z)^- \ell = \sigma^2 \ell^\top (Z^\top Z)^- \ell.$$

- $\ell^\top \hat{\boldsymbol{\beta}}$ is the BLUE of $\ell^\top \boldsymbol{\beta}$

Theorem (Theorem 3.9 in JS)

Assume model $X = Z\boldsymbol{\beta} + \epsilon$ with assumption A2

(i) *A necessary and sufficient condition for the existence of a linear unbiased estimator of $\ell^\top \boldsymbol{\beta}$ is $\ell \in \mathcal{R}(Z)$.*

(ii) *(Gauss-Markov theorem). If $\ell \in \mathcal{R}(Z)$, then the LSE $\ell^\top \hat{\boldsymbol{\beta}}$ is the BLUE of $\ell^\top \boldsymbol{\beta}$*

Proof of Theorem 3.9

(i) Sufficiency: If $\ell \in \mathcal{R}(Z)$ then $\ell^\top \hat{\boldsymbol{\beta}}$ is unbiased (Theorem 3.6).
Necessity: Suppose $c^\top X$ is unbiased for $\ell^\top \boldsymbol{\beta}$. Then

$$\ell^\top \boldsymbol{\beta} = E(c^\top X) = c^\top EX = c^\top Z\boldsymbol{\beta}. \tag{3}$$

Since this holds for all $\boldsymbol{\beta}$, we have $\ell = Z^\top c$, i.e., $\ell \in \mathcal{R}(Z)$
(ii) Let $c^\top X$ be any linear unbiased estimator of $\ell^\top \boldsymbol{\beta}$.

- The proof of (i) implies that $Z^\top c = \ell$
- Under A2

$$\begin{aligned}
var(c^\top X) &= c^\top \mathrm{Var}(\epsilon) c \\
&= \sigma^2 c^\top c \\
&= \sigma^2 \left(c^\top \mathbf{P}_Z c + c^\top \mathbf{P}_{Z\perp} c \right) \\
&\geq \sigma^2 c^\top \mathbf{P}_Z c \\
&= \sigma^2 c^\top Z (Z^\top Z)^- Z^\top c \\
&= \sigma^2 \ell^\top (Z^\top Z)^- \ell = \mathrm{Var}(\ell^\top \hat{\boldsymbol{\beta}})
\end{aligned}$$

Another proof of (ii)

- Under A1, $\ell^\top \hat{\boldsymbol{\beta}}$ is the UMVUE of $\ell^\top \boldsymbol{\beta}$. In particular, it has the smallest variance among all linear unbiased estimators.
- As long as $\mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2 I$, the variances of any linear unbiased estimator remains the same.
- Hence $\ell^\top \hat{\boldsymbol{\beta}}$ is the BLUE of $\ell^\top \boldsymbol{\beta}$ under A2.

Remark. $\ell^\top \hat{\boldsymbol{\beta}}$ is the BLUE of $\ell^\top \boldsymbol{\beta}$ under either A1 or A2.

Robustness of BLUE

- A procedure having certain properties under an assumption is said to be *robust against violation of the assumption* if this procedure still has the same properties when the assumption is (slightly) violated.

Theorem (Theorem 3.10)

Assume model $X = Z\beta + \epsilon$ with assumption A3: $E(\epsilon) = 0$ and $\mathrm{Var}(\epsilon)$ is an unknown matrix. The following are equivalent.

(a) *$\ell^\top \hat{\beta}$ is the BLUE of $\ell^\top \beta$ for any $\ell \in \mathcal{R}(Z)$.*

(b) *$E(\ell^\top \hat{\beta}\eta^\top X) = 0$ for any $\ell \in \mathcal{R}(Z)$ and any η such that $E(\eta^\top X) = 0$.*

(c) *$Z^\top \mathrm{Var}(\epsilon) U = 0$, where U is a matrix such that $Z^\top U = 0$ and $\mathcal{R}(U^\top) + \mathcal{R}(Z^\top) = \mathcal{R}^n$.*

(d) *$\mathrm{Var}(\epsilon) = Z\Lambda_1 Z^\top + U\Lambda_2 U^\top$ for some Λ_1 and Λ_2, where U is a matrix such that $Z^\top U = 0$ and $\mathcal{R}(U^\top) + \mathcal{R}(Z^\top) = \mathcal{R}^n$.*

(e) *The matrix $Z(Z^\top Z)^- Z^\top \mathrm{Var}(\epsilon)$ is symmetric.*

Roadmap of proof: (a) $\Leftrightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (b).

(a) $\Leftrightarrow$ (b).

The proof is an analogue of Theorem 3.2(i).
If $\ell \in \mathcal{R}(Z)$, let $c = Z(Z^\top Z)^- \ell$. Then $\ell^\top \hat{\boldsymbol{\beta}} = c^\top X$.
Suppose (a) holds.

- Suppose there is some η such that $E(\eta^\top X) = 0$, and $E(\ell^\top \hat{\boldsymbol{\beta}} \eta^\top X) \neq 0$ (WLOG, assume > 0)
- Define $\tilde{c} = c - t\eta$. Then

$$\begin{aligned}\mathrm{Var}(\tilde{c}^\top X) &= \mathrm{Var}(\ell^\top \hat{\boldsymbol{\beta}} - t\eta^\top X) \\ &= \mathrm{Var}(\ell^\top \hat{\boldsymbol{\beta}}) + t^2 \mathrm{Var}(\eta^\top X) - 2t\mathrm{Cov}(\ell^\top \hat{\boldsymbol{\beta}}, \eta^\top X),\end{aligned}$$

  whose derivative w.r.t. t is $2t\mathrm{Var}(\eta^\top X) - 2E(\ell^\top \hat{\boldsymbol{\beta}} \eta^\top X) < 0$ for any t sufficiently close to 0.
- This indicates that it is possible to pick $t > 0$ such that $\mathrm{Var}(\tilde{c}^\top X) < \mathrm{Var}(\ell^\top \hat{\boldsymbol{\beta}})$, which contradicts with (a).

Suppose (b) holds.

- For any unbiased linear estimator $\tilde{c}^\top X$, let $\eta = c - \tilde{c}$. Then

$$\mathrm{Var}(\tilde{c}^\top X) = \mathrm{Var}(\ell^\top \hat{\boldsymbol{\beta}} - \eta^\top X) = \mathrm{Var}(\ell^\top \hat{\boldsymbol{\beta}}) + \mathrm{Var}(\eta^\top X) \geq \mathrm{Var}(\ell^\top \hat{\boldsymbol{\beta}})$$

(b) $\Rightarrow$ (c).

Suppose that (b) holds.

- For any $\eta \in \mathcal{R}(U^\top)$, $E(\eta^\top X) = \eta^\top Z\boldsymbol{\beta} = 0$.
- For any $\gamma \in \mathcal{R}^p$, let $\ell = (Z^\top Z)\gamma$. Then $\ell \in \mathcal{R}(Z)$.
- By (b),

$$\begin{aligned} 0 &= E(\ell^\top \hat{\boldsymbol{\beta}} \eta^\top X) \\ &= \mathrm{Cov}(\ell^\top \hat{\boldsymbol{\beta}}, \eta^\top X) \\ &= \mathrm{Cov}(\gamma^\top (Z^\top Z)(Z^\top Z)^- Z^\top X, \eta^\top X) \\ &= \gamma^\top (Z^\top Z)(Z^\top Z)^- Z^\top \mathrm{Cov}(X, X)\eta \\ &= \gamma^\top (Z^\top Z)(Z^\top Z)^- Z^\top \mathrm{Var}(\epsilon)\eta. \end{aligned}$$

- Since $(Z^\top Z)(Z^\top Z)^- Z^\top = Z^\top$ and since the last equality holds for all $\gamma \in \mathcal{R}^p$ and $\eta \in \mathcal{R}(U^\top)$, we have

$$0 = Z^\top \mathrm{Var}(\epsilon) U$$

(c) $\Rightarrow$ (d).

We need to use the following facts from the theory of linear algebra:
If $Z^\top U = 0$ and $\mathcal{R}(U^\top) + \mathcal{R}(Z^\top) = \mathcal{R}^n$, then there exists a nonsingular matrix C such that $\mathrm{Var}(\boldsymbol{\epsilon}) = CC^\top$ and $C = ZC_1 + UC_2$ for some matrices C_1 and C_2.

- Let $\Lambda_1 = C_1 C_1^\top$, $\Lambda_2 = C_2 C_2^\top$, and $\Lambda_3 = C_1 C_2^\top$.
- Then
$$\mathrm{Var}(\boldsymbol{\epsilon}) = Z\Lambda_1 Z^\top + U\Lambda_2 U^\top + Z\Lambda_3 U^\top + U\Lambda_3^\top Z^\top \tag{4}$$
and $Z^\top \mathrm{Var}(\boldsymbol{\epsilon}) U = Z^\top Z \Lambda_3 U^\top U$
- If (c) holds, $0 = Z^\top \mathrm{Var}(\boldsymbol{\epsilon}) U$ and thus
$$0 = Z(Z^\top Z)^- \left[Z^\top Z \Lambda_3 U^\top U \right] (U^\top U)^- U^\top = Z\Lambda_3 U^\top,$$
- Together with (4) , we have $\mathrm{Var}(\boldsymbol{\epsilon}) = Z\Lambda_1 Z^\top + U\Lambda_2 U^\top$

(d) $\Rightarrow$ (e).

If (d) holds, then $Z(Z^\top Z)^- Z^\top \mathrm{Var}(\boldsymbol{\epsilon}) = Z\Lambda_1 Z^\top$, which is symmetric.

(e) $\Rightarrow$ (b).

Suppose (e) holds.
For any $\ell \in \mathcal{R}(Z)$ and any η such that $E(\eta^\top X) = 0$,

- there exists some $\gamma \in \mathcal{R}^p$ such that $\ell = (Z^\top Z)\gamma$.
- $0 = E(\eta^\top X) = \eta^\top Z\boldsymbol{\beta}$ for all $\boldsymbol{\beta} \Rightarrow \eta^\top Z = 0$
- By the calculation we did in proof of "(b) $\Rightarrow$ (c)", we have

$$\begin{aligned} E(\ell^\top \hat{\boldsymbol{\beta}} \eta^\top X) &= \gamma^\top (Z^\top Z)(Z^\top Z)^- Z^\top \mathrm{Var}(\epsilon)\eta \\ &= \gamma^\top Z^\top \left[Z(Z^\top Z)^- Z^\top \mathrm{Var}(\epsilon) \right]^\top \eta \\ &= \gamma^\top Z^\top \mathrm{Var}(\epsilon) Z(Z^\top Z)^- Z^\top \eta \\ &= 0, \end{aligned}$$

where the second equation is due to (e) and the last is due to $\eta^\top Z = 0$

Robustness of UMVUE

The following result characterizes the robustness of UMVUE under the normal noise assumption against the violation of $\mathrm{Var}(\epsilon) = \sigma^2 \boldsymbol{I}_n$.

Corollary (Corollary 3.3 of the textbook)

Consider model $X = Z\boldsymbol{\beta} + \epsilon$ with a full rank Z, $\boldsymbol{\epsilon} \sim N_n(0, \Sigma)$, where Σ is an unknown positive definite matrix. Then $\ell^\top \hat{\boldsymbol{\beta}}$ is a UMVUE of $\ell^\top \boldsymbol{\beta}$ for any $\ell \in \mathcal{R}^p$ iff one of (b)-(e) in Theorem 3.10 holds.

"$\Rightarrow$" : because when $\ell^\top \hat{\boldsymbol{\beta}}$ is the UMVUE, it is the BLUE.

Proof of "$\Leftarrow$"

WLOG, we can assume $Z^\top Z = I_p$. Otherwise, re-parametrize $\tilde{\boldsymbol{\beta}} = DV^\top \boldsymbol{\beta}$ and let $\tilde{\ell} = D^{-1}V\ell$, where $Z = \tilde{Z}_{n\times p} D_{p\times p} V_{p\times p}^\top$ is the singular value decomposition of Z. Then the model becomes $X = \tilde{Z}\tilde{\boldsymbol{\beta}} + \epsilon$

Suppose (c) holds (since (a–e) are equivalent)

- Recall that $\mathrm{Var}(\ell^\top \hat{\boldsymbol{\beta}}) = \ell^\top (Z^\top Z)^{-1} Z^\top \Sigma Z (Z^\top Z)^{-1} \ell = \ell^\top Z^\top \Sigma Z \ell$
- Let $A \in \mathcal{R}^{n\times(n-p)}$ be an orthogonal matrix such that $A^\top Z = 0$ and $A^\top A = I_{n-p}$.
- Then $Z^\top \Sigma A = 0$ because of (c). One can show that $(Z^\top \Sigma Z)^{-1} = Z^\top \Sigma^{-1} Z$.
- The Fisher information is $I = Z^\top \Sigma^{-1} Z$, and the Cramé-Rao lower bound for $\ell^\top \boldsymbol{\beta}$ is

$$\ell^\top I^{-1} \ell = \ell^\top \left(Z^\top \Sigma^{-1} Z \right)^{-1} \ell = \ell^\top Z^\top \Sigma Z \ell,$$

  which is achieved by $\ell^\top \hat{\boldsymbol{\beta}}$

Asymptotic Properties of LSE

- Suppose $\ell \in \mathcal{R}(Z)$.
- Assume the linear model $X = Z\boldsymbol{\beta} + \epsilon$ under assumption A3, i.e., $E(\boldsymbol{\epsilon}) = 0$ and $\Sigma_n = \mathrm{Var}(\boldsymbol{\epsilon})$ is an unknown matrix.
- Consider the LSE $\ell^\top \hat{\boldsymbol{\beta}}$ for every n, where $\hat{\boldsymbol{\beta}} = (Z^\top Z)^- Z^\top X$
- Denote by $A_n = (Z^\top Z)^-$.
- We need some regularity conditions to ensure that, as n increase, the noise would not inflate (Σ_n is not too large) and the matrix of covariate is large (A_n is small)
- Denote by $\lambda_+[A]$ the largest eigenvalue of the matrix A

Theorem (Theorem 3.11 (Consistency) of the textbook)

Suppose that $\sup_n \lambda_+[\mathrm{Var}(\boldsymbol{\epsilon})] < \infty$ *and that* $\lim_{n\to\infty} \lambda_+[A_n] = 0$. *Then* $\ell^\top \hat{\boldsymbol{\beta}}$ *is consistent in MSE for any* $\ell \in \mathcal{R}(Z)$, *i.e.,* $\ell^\top \hat{\boldsymbol{\beta}} \to \ell^\top \boldsymbol{\beta}$ *in* L_2.

Proof

- From linear algebra, we have

$$v^\top A v \le \lambda_+(A) v^\top v \tag{5}$$

- The result follows from the fact that $\ell^\top \hat{\boldsymbol{\beta}}$ is unbiased and

$$\begin{aligned}
\mathrm{Var}(\ell^\top \hat{\boldsymbol{\beta}}) &= \ell^\top (Z^\top Z)^- Z^\top \mathrm{Var}(\epsilon) Z (Z^\top Z)^- \ell \\
&\le \lambda_+[\mathrm{Var}(\epsilon)] \ell^\top (Z^\top Z)^- Z^\top Z (Z^\top Z)^- \ell \\
&= \lambda_+[\mathrm{Var}(\epsilon)] \ell^\top (Z^\top Z)^- \ell \\
&\le \lambda_+[\mathrm{Var}(\epsilon)] \lambda_+((Z^\top Z)^-) \ell^\top \ell \\
&\to 0,
\end{aligned}$$

where the second and the fourth inequalities are due to Eq (5), and the last convergence is due to the conditions.

Tutorial

1. Let $(X_1, \ldots, X_n)$ be a random sample from the exponential distribution on (a, ∞) with scale parameter θ, where $a \in \mathcal{R}$ and $\theta > 0$ are unknown. Show that $T = (X_{(1)}, \sum_{i=1} X_i - nX_{(1)})$ is a complete statistic.
   Hint: Use the Rényi representation
2. Consider a linear model in matrix form $X_{n\times 1} = Z_{n\times p}\boldsymbol{\beta}_{p\times 1} + \epsilon_{n\times 1}$ with $p \leq n$ and with the assumption that $\epsilon \sim N(\mathbf{0}_n, \sigma^2 \boldsymbol{I}_n)$. Show that if each coordinate of $\boldsymbol{\beta}$ is estimable, then the rank of Z is p.
3. (James-Stein estimator) Suppose X is a p-random vector from $N(\theta, \boldsymbol{I}_p)$ with an unknown $\theta \in \mathcal{R}^p$. Consider the squared loss function for estimating θ:

$$L(\theta, a) = \|a - \theta\|^2 = \sum_{i=1}^{p} (a_i - \theta_i)^2 \,,$$

   where a_i and θ_i are the ith coordinates of the estimator and the estimand. Show that for any $p \geq 3$, the risk of the following estimator

$$\hat{\theta} = \left(1 - \frac{(p-2)}{\|X\|^2}\right) X$$

   is strictly smaller than X. Can you extend this result to the case where $X \sim N(\theta, D)$ with some known $p \times p$ positive definite matrix D?

Exercise 1

Let $(X_1, \ldots, X_n)$ be a random sample from the exponential distribution on (a, ∞) with scale parameter θ, where $a \in \mathcal{R}$ and $\theta > 0$ are unknown. Show that $T = (X_{(1)}, \sum_{i=1} X_i - nX_{(1)})$ is a boundedly complete statistic.
Hint: Use the Rényi representation

Proof:

- Last time, we have shown the joint Lebesgue p.d.f. of $x = (x_1, \ldots, x_n)$ is

$$\theta^{-n} \exp\left(-\theta^{-1} \sum_{i=1}^{n} (x_i - x_{(1)})\right) \exp\left(-n\theta^{-1}\left(x_{(1)} - a\right)\right) I_{(0, x_{(1)}]}(a)$$

  and T is sufficient for (a, θ)
- By Rényi representation, $T \stackrel{\mathcal{D}}{=} (a + Y_n/n, Y_1 + \cdots + Y_{n-1})$, where Y_i's are i.i.d. from $E(0, \theta)$. This shows that T_1 and T_2 are independent and the distribution of T_2 does not depend on a

- Suppose $h(T_1, T_2)$ is a bounded measurable function such that $Eh(T_1, T_2) = 0$ for all $\theta > 0$ and a.
- Let $g(t_1, \theta) = Eh(t_1, T_2)$. This only depend on θ but not on a. Furthermore, by the p.d.f. of T, this is a continuous function in θ for any fixed t_1
- Since T_1 and T_2 are independent, we have $E(h(T_1, T_2) \mid T_1) = g(T_1, \theta)$ by Proposition 1.10 (vii) in JS
- Therefore $0 = Eg(T_1, \theta)$ for all a and $\theta > 0$.
- For any fixed θ, the last equation is $0 = \int_a^\infty g(x, \theta) e^{-n\theta^{-1}(x-a)} \, \mathrm{d}x$, which implies $0 = \int_a^\infty g(x, \theta) e^{-n\theta^{-1}x} \, \mathrm{d}x$, for all a.
- Differentiate the last equation w.r.t. a, we have $g(x, \theta) e^{-n\theta^{-1}x} = 0$ a.e. So $g(x, \theta) = 0$ a.e.
- By Fubini's theorem, $0 = \int_{\mathcal{R}^+} \, \mathrm{d}\theta \int_{\mathcal{R}} |g(x, \theta)| \, \mathrm{d}x = \int_{\mathcal{R}} \, \mathrm{d}x \int_{\mathcal{R}^+} |g(x, \theta)| \, \mathrm{d}\theta$
- This together with the fact that $g(x, \theta)$ is continuous in θ shows that $g(x, \theta) = 0$ for all $\theta > 0$ except for a null set of x
- For a.e. x, $0 = \int h(x, y) y^{n-2} e^{-y/\theta} \, \mathrm{d}y$ for all θ. Fix x and apply the result of exponential family, we conclude that $h(x, y) = 0$ a.e.

Exercise 2

Consider a linear model in matrix form $X_{n\times 1} = Z_{n\times p}\boldsymbol{\beta}_{p\times 1} + \epsilon_{n\times 1}$ with $p \leq n$ and with the assumption that $\epsilon \sim N(\mathbf{0}_n, \sigma^2 \boldsymbol{I}_n)$. Show that if each coordinate of $\boldsymbol{\beta}$ is estimable, then the rank of Z is p.

Proof:

- Under the normality assumption, β_j being estimable implies that e_i the vector with elements 0 but 1 on the ith coordinate is in $\mathcal{R}(Z)$, i.e., $e_i = Z^\top \alpha_i$ for some $\alpha_i \in \mathcal{R}^n$.
- Since this holds for $i = 1, \ldots, p$, we have $[e_1, \ldots, e_p] = Z^\top [\alpha_1, \ldots, \alpha_p]$.
- Note that rank$(AB)\leq$ min(rank(A), rank(B)). So we have $p =$rank$(I_p)\leq$rank$(Z)\leq p$.

Exercise 3

(James-Stein estimator) Suppose X is a p-random vector from $N(\theta, \boldsymbol{I}_p)$ with an unknown $\theta \in \mathcal{R}^p$. Consider the squared loss function for estimating θ:

$$L(\theta, a) = \|a - \theta\|^2 = \sum_{i=1}^{p} (a_i - \theta_i)^2 ,$$

where a_i and θ_i are the ith coordinates of the estimator and the estimand. Show that for any $p \geq 3$, the risk of the following estimator

$$\hat{\theta} = \left(1 - \frac{(p-2)}{\|X\|^2}\right) X$$

is strictly smaller than X. Can you extend this result to the case where $X \sim N(\theta, D)$ with some known $p \times p$ positive definite matrix D?

Proof:

- Note that

$$\begin{aligned}
\mathbb{E}\|\theta - \hat{\theta}\|_2^2 &= \mathbb{E}\|\theta - X + X - \hat{\theta}\|_2^2 \\
&= p + \mathbb{E}\|X - \hat{\theta}\|_2^2 + 2\mathbb{E}(\theta - X)^\top (X - \hat{\theta}) \\
&= p + (p-2)^2 \mathbb{E}\frac{1}{\|X\|^2} - 2\mathbb{E}(X - \theta)^\top (X - \hat{\theta})
\end{aligned}$$

- The multivariate Stein's lemma:
  Suppose $X \sim N(\theta, \sigma^2 \boldsymbol{I}_p)$ and $f : \mathcal{R}^n \mapsto \mathcal{R}$ is differentiable satisfying $E|f(X)| < \infty$, we have

$$\frac{1}{\sigma^2}\mathbb{E}[(X_i - \theta_i) f(X)] = \mathbb{E}[\partial/\partial x_i f(X)]$$

- Let $f_i(x) = x_i/\|x\|^2$. Then $\partial/\partial x_i f_i(X) = 1/\|x\|^2 - 2x_i^2/\|x\|^4$.
- The lemma implies that

$$\begin{aligned}
\mathbb{E}(X-\theta)^\top(X-\hat{\theta}) &= (p-2)\mathbb{E}[\sum_{i\le p}(X_i - \theta_i) f_i(X)] \\
&= (p-2)\sum_{i\le p}\mathbb{E}[\partial/\partial x_i f_i(X)] \\
&= (p-2)\left(p\mathbb{E}\frac{1}{\|X\|^2} - 2\sum_{i\le p}\mathbb{E}\frac{X_i^2}{\|X\|^4}\right) \\
&= (p-2)^2\mathbb{E}\frac{1}{\|X\|^2}
\end{aligned}$$

- If $X \sim N(\theta, D)$ with some known $p \times p$ positive definite matrix D, the James-Stein estimator is defined as

$$\hat{\theta}_D = X - \frac{(p-2)}{\|D^{-1}X\|^2} D^{-1}X$$

- Let $D^{-1} = H^2$ for some p.s.d. matrix H (known as the square root).
- Let $Y = HX$. Then $Y \sim N(H\theta, HD^{-1}H = \boldsymbol{I}_p)$.
- Let $f_i(y) = y_i/\|Hy\|^2$, $f(y) = (f_1(y), \ldots, f_p(y))^\top$. Note that $\partial/\partial y_i f_i(Y) = 1/\|Hy\|^2 - 2(Hy)_i/\|Hy\|^4$.
- Stein's Lemma implies that

$$\begin{aligned}
\mathbb{E}(X-\theta)^\top (X - \hat{\theta}_D) &= \mathbb{E}(Y - H\theta)^\top f(Y) \\
&= (p-2) \sum_{i \le p} \mathbb{E}[\partial/\partial y_i f_i(Y)] \\
&= (p-2)^2 \mathbb{E} \frac{1}{\|HY\|^2}
\end{aligned}$$

- The rest is the same as before and $\mathbb{E}\|\theta - \hat{\theta}_D\|_2^2 = E\|\theta - X\|^2 - (p-2)^2 \mathbb{E}\frac{1}{\|D^{-1}X\|^2}$