ST5215 Advanced Statistical Theory, Lecture 23

HUANG Dongming

National University of Singapore

5 Nov 2020

Overview

Last time

- Asymptotic Efficiency
- Linear Models

Today

- Properties of Least Squares Estimators

Recap: Asymptotic Efficiency

- Consider the estimators s.t.

$$[V_n(\theta)]^{-1/2}(\hat{\theta}_n - \theta) \xrightarrow{D} N_k(0, I_k),$$

  where $V_n(\theta)$ is called *the asymptotic covariance matrix*
- $V_n(\theta)$ is usually of the form $n^{-\delta}V(\theta)$ for some $\delta > 0$
- We can compare different estimators by their asymptotic covariance matrices

Under some regularity conditions, we have

- The information inequality

$$V_n(\theta) \succeq [I_n(\theta)]^{-1}$$

  holds for θ except for a null set (Lebesgue measure $= 0$)
- An estimator with $V_n(\theta) = [I_n(\theta)]^{-1}$ is called *asymptotically efficient*
- If a MLE is always the unique RLE, then it is asymptotically efficient
- The one-step MLE is asymptotically efficient if the initial estimator is $\sqrt{n}$-consistent

Linear Models

A matrix form of a linear model is

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

where

- $X = (X_1, ..., X_n)^\top$: the vector of responses
- $\epsilon = (\epsilon_1, ..., \epsilon_n)^\top$: the vector of noise
- Z be the $n \times p$ matrix whose ith row is the vector $Z_i^\top$, $i = 1, ..., n$: the design matrix, or the matrix of covariates
- $\boldsymbol{\beta}$ is a p-vector of unknown parameters (main parameters of interest), $p < n$; in this course, that the range of $\boldsymbol{\beta}$ is $\mathcal{R}^p$.

A *least squares estimator (LSE)* of $\boldsymbol{\beta}$ is defined to be any $\hat{\boldsymbol{\beta}}$ such that

$$\|X - Z\hat{\boldsymbol{\beta}}\|^2 = \min_{\mathbf{b}} \|X - Z\mathbf{b}\|^2,$$

which is given by

$$\hat{\boldsymbol{\beta}} = (Z^\top Z)^{-} Z^\top X,$$

Example 3.13: One-way ANOVA Models

- Suppose that $n = \sum_{j=1}^{m} n_j$ with m positive integers $n_1, ..., n_m$ and that Consider the model:

$$X_{ij} = \mu_i + \epsilon_{ik}, \qquad j = 1, \ldots, n_i, i = 1, \ldots, m,$$

where ϵ_{ij} are i.i.d random errors with mean 0 and variance σ^2.

- This model is called a one-way ANOVA model.
- Let $\mathbf{X}_i = (X_{i1} \ldots, X_{in_i})^\top$ and $\mathbf{X} = (\mathbf{X}_1^\top, \ldots, \mathbf{X}_m^\top)^\top$. Let J_k be the k-vector of ones and

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

Let $\boldsymbol{\beta} = (\mu_1, ..., \mu_m)^\top$ and $\boldsymbol{\epsilon} = (\epsilon_{11}, \ldots, \epsilon_{1n_1}, \ldots, \epsilon_{m1}, \ldots, \epsilon_{mn_m})^\top$. Then the one-way ANOVA model can be expressed as

$$\mathbf{X} = Z\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Example (Cont.)

- Since $Z^\top Z = \mathrm{Diag}(n_1, \ldots, n_m)$, $(Z^\top Z)^{-1} = \mathrm{Diag}(n_1^{-1}, \ldots, n_m^{-1})$.
- Hence the unique LSE of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (Z^\top Z)^{-1} Z^\top \mathbf{X} = (\bar{X}_{1\cdot}, \ldots, \bar{X}_{m\cdot})^\top,$$

  where $\bar{X}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$.

Sometimes the model is expressed as

$$X_{ij} = \mu + \alpha_i + \epsilon_{ij}, j = 1, ..., n_i, i = 1, ..., m,$$

with constraint $\sum \alpha_i = 0$.

- Without this constrain, the model is not identifiable (Homework 1.10)
- Let $\boldsymbol{\beta} = (\mu, \alpha_1, \ldots, \alpha_m)^\top$. The LSE of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = \left(\bar{X}, \bar{X}_{1\cdot} - \bar{X}, ..., \bar{X}_{m\cdot} - \bar{X}\right),$$

  where $\bar{X}$ is total sample mean.

Model Assumptions

To study properties of LSE's of $\boldsymbol{\beta}$, we need some assumptions on the distribution of X or ϵ (conditional on Z if Z is random).

A1: (Gaussian noise) ϵ is distributed as $N_n(0, \sigma^2 I_n)$ 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.

- If the matrix Z is not of full rank, then the model is not identifiable.
- Suppose that the rank of Z is $r \le p$. Then there is an $n \times r$ submatrix U of Z such that $Z = UQ$ and U is of rank r, where Q is a fixed $r \times p$ matrix. The model is identifiable if we consider the reparameterization $\tilde{\boldsymbol{\beta}} = Q\boldsymbol{\beta}$.

Estimating Linear Combinations of Coefficients

- In many applications, we are interested in estimating $\vartheta = \ell^\top \boldsymbol{\beta}$ for some $\ell \in \mathcal{R}^p$.
- But $\ell^\top \boldsymbol{\beta}$ is meaningless unless $\ell = Q^\top c$ for some $c \in \mathcal{R}^r$ so that

$$\ell^\top \boldsymbol{\beta} = c^\top Q \boldsymbol{\beta} = c^\top \tilde{\boldsymbol{\beta}}.$$

- Denoted by $\mathcal{R}(A)$ the smallest linear subspace containing all rows of A

Theorem (Theorem 3.6 of the textbook)

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 Q is given in $Z = UQ$.*

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

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

Proof of (i)

This is a result in linear algebra.

- Note that $a \in \mathcal{R}(A)$ iff $a = A^\top b$ for some vector b.
- If $\ell = Q^\top c$, then
$$\ell = Q^\top c = Q^\top U^\top U(U^\top U)^{-1}c = Z^\top[U(U^\top U)^{-1}c].$$
Hence $\ell \in \mathcal{R}(Z)$.
- If $\ell \in \mathcal{R}(Z)$, then $\ell = Z^\top \zeta$ for some ζ and
$$\ell = (UQ)^\top \zeta = Q^\top[U^\top \zeta].$$

Remark. If $Z = UQ$ such that $U^\top U$ and $QQ^\top$ are invertible, then
$$\mathcal{R}(Z) = \mathcal{R}(Q).$$

Since $Z^\top Z = Q^\top U^\top UQ = \tilde{U}Q$, where $\tilde{U}^\top \tilde{U} = U^\top UQQ^\top U^\top U$ is invertible, we have

$$\mathcal{R}(Z) = \mathcal{R}(Z^\top Z)$$

Proof of (ii)

Suppose $\ell \in \mathcal{R}(Z) = \mathcal{R}(Z^\top Z)$: $\ell = Z^\top Z\zeta$ for some ζ.

- (Uniqueness) If $\bar{\boldsymbol{\beta}}$ is any other LSE of $\boldsymbol{\beta}$, then $Z^\top Z\bar{\boldsymbol{\beta}} = Z^\top X$, which implies

$$\ell^\top \hat{\boldsymbol{\beta}} - \ell^\top \bar{\boldsymbol{\beta}} = \zeta^\top (Z^\top Z)(\hat{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}) = \zeta^\top (Z^\top X - Z^\top X) = 0.$$

- (Unbiasedness) Since $\hat{\boldsymbol{\beta}} = (Z^\top Z)^- Z^\top X$, we have

$$\begin{aligned} E(\ell^\top \hat{\boldsymbol{\beta}}) &= E[\ell^\top (Z^\top Z)^- Z^\top X] = \zeta^\top Z^\top Z (Z^\top Z)^- Z^\top Z \boldsymbol{\beta} \\ &= \zeta^\top Z^\top Z \boldsymbol{\beta} = \ell^\top \boldsymbol{\beta}, \end{aligned}$$

  where the second equation is due to the linearity of expectation and assumption A3.

Proof of (iii)

Proof by Contraposition:
Suppose there is an estimator $h(X, Z)$ unbiased for $\ell^\top \boldsymbol{\beta}$.

- Under A1,

$$\ell^\top \boldsymbol{\beta} = \int_{\mathcal{R}^n} h(x, Z)(2\pi)^{-n/2}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\|x - Z\boldsymbol{\beta}\|^2\right\} dx.$$

- Differentiate w.r.t. $\boldsymbol{\beta}$, which can be exchanged with the integral sign for natural exponential families (Theorem 2.1)

$$\ell = Z^\top \int_{\mathcal{R}^n} h(x, Z)(2\pi)^{-n/2}\sigma^{-n-2}(x - Z\boldsymbol{\beta})\exp\left\{-\frac{1}{2\sigma^2}\|x - Z\boldsymbol{\beta}\|^2\right\} dx,$$

  which implies $\ell \in \mathcal{R}(Z)$.

Properties Under Assumption A1 (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}.*

Proof of (i)

- If $\hat{\boldsymbol{\beta}}$ is an LSE of $\boldsymbol{\beta}$, then as a function of x, $\hat{\boldsymbol{\beta}}(x) = (Z^\top Z)^- Z^\top x$, and $Z^\top Z\hat{\boldsymbol{\beta}} = Z^\top x$,
$$(x - Z\hat{\boldsymbol{\beta}})^\top Z(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (x^\top Z - x^\top Z)(\hat{\boldsymbol{\beta}}(x) - \boldsymbol{\beta}) = 0, \forall \boldsymbol{\beta} \in \mathcal{R}^p$$
- Hence,
$$\begin{aligned}\|x - Z\boldsymbol{\beta}\|^2 &= \|x - Z\hat{\boldsymbol{\beta}}(x) + Z\hat{\boldsymbol{\beta}}(x) - Z\boldsymbol{\beta}\|^2 \\ &= \|x - Z\hat{\boldsymbol{\beta}}(x)\|^2 + \|Z\hat{\boldsymbol{\beta}}(x) - Z\boldsymbol{\beta}\|^2 \\ &= \|x - Z\hat{\boldsymbol{\beta}}(x)\|^2 - 2\boldsymbol{\beta}^\top Z^\top x + \|Z\boldsymbol{\beta}\|^2 + \|Z\hat{\boldsymbol{\beta}}(x)\|^2.\end{aligned}$$
- Under assumption A1, the joint Lebesgue p.d.f. of X can be written as:
$$f_{\boldsymbol{\beta}}(x) = (2\pi\sigma^2)^{-n/2}\exp\left\{\frac{\boldsymbol{\beta}^\top Z^\top x}{\sigma^2} - \frac{\|x - Z\hat{\boldsymbol{\beta}}(x)\|^2 + \|Z\hat{\boldsymbol{\beta}}(x)\|^2}{2\sigma^2} - \frac{\|Z\boldsymbol{\beta}\|^2}{2\sigma^2}\right\}$$
- $\hat{\boldsymbol{\beta}}(x)$ is a function of $Z^\top x$, the statistic $T = (Z^\top X, \|X - Z\hat{\boldsymbol{\beta}}\|^2)$ is complete and sufficient for $\theta = (\boldsymbol{\beta}, \sigma^2)$ (by properties of exponential families, Proposition 2.1)
- If $\ell^\top \boldsymbol{\beta}$ is estimable, then $\ell^\top \hat{\boldsymbol{\beta}}$ is unbiased for $\ell^\top \boldsymbol{\beta}$ (Theorem 3.6) and, hence, $\ell^\top \hat{\boldsymbol{\beta}}$ is the UMVUE of $\ell^\top \boldsymbol{\beta}$ since it is a function of T

Remarks

- In general,

$$\mathrm{Var}\left(\ell^\top \hat{\boldsymbol{\beta}}\right) = \ell^\top \left(Z^\top Z\right)^- Z^\top \mathrm{Var}(\varepsilon) Z \left(Z^\top Z\right)^- \ell$$

- If $\ell \in \mathcal{R}(Z)$ and $\mathrm{Var}(\varepsilon) = \sigma^2 I_n$ (assumption A2), then the use of the generalized inverse matrix in (3.34) leads to $\mathrm{Var}\left(\ell^\top \hat{\boldsymbol{\beta}}\right) = \sigma^2 \ell^\top \left(Z^\top Z\right)^- \ell$

Proof of (ii)

- Since $\hat{\sigma}^2$ is a function of the complete sufficient statistic, it is the UMVUE of σ^2 if we can show

$$E\hat{\sigma}^2 = (n-r)^{-1}E\|X - Z\hat{\boldsymbol{\beta}}\|^2 = \sigma^2.$$

- Since $\|X - Z\boldsymbol{\beta}\|^2 = \|X - Z\hat{\boldsymbol{\beta}}\|^2 + \|Z\hat{\boldsymbol{\beta}} - Z\boldsymbol{\beta}\|^2$ and $E(Z\hat{\boldsymbol{\beta}}) = Z\boldsymbol{\beta}$,

$$\begin{aligned}
E\|X - Z\hat{\boldsymbol{\beta}}\|^2 &= E(X - Z\boldsymbol{\beta})^\top(X - Z\boldsymbol{\beta}) - E(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top Z^\top Z(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\
&= \mathrm{tr}\left(\mathrm{Var}(X) - \mathrm{Var}(Z\hat{\boldsymbol{\beta}})\right) \\
&= \sigma^2[n - \mathrm{tr}\left(Z(Z^\top Z)^- Z^\top Z(Z^\top Z)^- Z^\top\right)] \\
&= \sigma^2[n - \mathrm{tr}\left((Z^\top Z)^- Z^\top Z\right)].
\end{aligned}$$

- It remains to show $\mathrm{tr}\left((Z^\top Z)^- Z^\top Z\right) = r$. This can be showed using the singular value decomposition in Lecture 22 (Page 16) and is left for exercise.

Remarks

- The vector $X - Z\hat{\boldsymbol{\beta}}$ is called the residual vector and $\|X - Z\hat{\boldsymbol{\beta}}\|^2$ is called the sum of squared residuals and is denoted by SSR.
- The estimator $\hat{\sigma}^2$ is then equal to $SSR/(n-r)$
- Note that both $X - Z\hat{\boldsymbol{\beta}} = \left[I_n - Z\left(Z^\top Z\right)^- Z^\top\right] = \mathbf{P}_{Z\perp}X$ and $Z\hat{\boldsymbol{\beta}} = Z\left(Z^\top Z\right)^- Z^\top X = \mathbf{P}_Z X$ are linear in X, they are jointly normally distributed under assumption A1.
- Furthermore, we can check that

$$\mathbf{P}_{Z\perp}\mathbf{P}_Z = (I_n - \mathbf{P}_Z)\mathbf{P}_Z = \mathbf{P}_Z - \mathbf{P}_Z^2 = \mathbf{P}_Z - \mathbf{P}_Z = 0$$

  so $X - Z\hat{\boldsymbol{\beta}}$ and $Z\hat{\boldsymbol{\beta}}$ are independent
- It follows that for any estimable $\ell^\top\boldsymbol{\beta}$, $\hat{\sigma}^2$ and $\ell^\top\hat{\boldsymbol{\beta}}$ 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 $\ell^\top \boldsymbol{\beta}$.
- 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}

Properties Under Assumption A2

- 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 $\mathbf{c}$.
- $\ell^\top \hat{\boldsymbol{\beta}}$ 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 $\mathbf{c}^\top \mathrm{Var}(X)\mathbf{c} = \mathbf{c}^\top \mathrm{Var}(\boldsymbol{\epsilon})\mathbf{c}$.

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 *best linear unbiased estimator* (BLUE) of $\ell^\top \boldsymbol{\beta}$ in the sense that it has the minimum variance in the class of linear unbiased estimators of $\ell^\top \boldsymbol{\beta}$

Theorem 3.9 in JS

Theorem

Assume model $X = Z\boldsymbol{\beta} + \epsilon$ *with assumption A2:* $E(\epsilon) = 0$ *and* $\mathrm{Var}(\epsilon) = \sigma^2 I_n$ *with an unknown* $\sigma^2 > 0$.

(i) *A necessary and sufficient condition for the existence of a linear unbiased estimator of* $\ell^\top \boldsymbol{\beta}$ *(i.e., an unbiased estimator that is linear in* X*) 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) If $\ell \in \mathcal{R}(Z)$ then $\ell^\top \hat{\boldsymbol{\beta}}$ is unbiased (Theorem 3.6)

- Suppose $c^\top X$ be unbiased for $\ell^\top \boldsymbol{\beta}$. Then

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

- So $\ell = Z^\top c$, 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

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

Another proof of (ii)

- Under A1, $\ell^\top \hat{\boldsymbol{\beta}}$ is the UMVUE. In particular, it has the smallest variance among all linear unbiased estimators.
- However, 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 under A2.

Tutorial

1. Let $(Y_1, \ldots, Y_n)$ be a random sample such that Y_i is distributed as $N(\theta, \theta)$ with an unknown $\theta > 0$.

   Show that one of the solutions of the likelihood equation is the unique MLE of θ. Obtain the asymptotic distribution of the MLE of θ.

2. 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. Obtain the asymptotic relative efficiency of the MLE of a with respect to the UMVUE of a.

3. Consider a linear model in matrix form $X_{n\times 1} = Z_{n\times p}\boldsymbol{\beta}_{p\times 1} + \epsilon_{n\times 1}$. Under the assumption that $\epsilon \sim N(\mathbf{0}_n, \sigma^2 \boldsymbol{I}_n)$ where σ is known, compute the Fisher information $I(\boldsymbol{\beta})$. When is $I(\boldsymbol{\beta})$ positive definite?

Exercise 1

Let $(Y_1, \ldots, Y_n)$ be a random sample such that Y_i is distributed as $N(\theta, \theta)$ with an unknown $\theta > 0$.
Show that one of the solutions of the likelihood equation is the unique MLE of θ. Obtain the asymptotic distribution of the MLE of θ.

Proof:

- Let $T = n^{-1} \sum_{i=1}^n Y_i^2$ and $\ell(\theta)$ be the likelihood function.
- Then

$$\log \ell(\theta) = -\frac{n \log \theta}{2} - \frac{1}{2\theta} \sum_{i=1}^{n} (Y_i - \theta)^2 ,$$

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = \frac{n}{2} \left(\frac{T}{\theta^2} - \frac{1}{\theta} - 1 \right),$$

and

$$\frac{\partial^2 \log \ell(\theta)}{\partial \theta^2} = \frac{n}{2\theta^2} - \frac{nT}{\theta^3}$$

Ex 1 (Cont.)

- The likelihood equation $\frac{\partial \log \ell(\theta)}{\partial \theta} = 0$ has two solutions

$$\frac{\pm\sqrt{1+4T}-1}{2}$$

- The only positive solution is $\hat{\theta} = (\sqrt{1+4T}-1)/2$
- For $\theta > \hat{\theta}$, $\frac{\partial \log \ell(\theta)}{\partial \theta} < 0$; for $\theta \in (0, \hat{\theta})$, $\frac{\partial \log \ell(\theta)}{\partial \theta} > 0$.
- So $\hat{\theta}$ is the unique MLE of θ
- Since $EY_1^2 = \theta + \theta^2$, the Fisher information is

$$I_n(\theta) = -E\left(\frac{\partial^2 \log \ell(\theta)}{\partial \theta^2}\right) = E\left(\frac{n}{2\theta^2} - \frac{nT}{\theta^3}\right) = \frac{(2\theta+1)n}{2\theta^2}$$

- Thus, $\sqrt{n}(\hat{\theta}-\theta) \to_d N\left(0, 2\theta^2/(2\theta+1)\right)$ because the regularity conditions are satisfied by the natural exponential family with $\eta = -1/(2\theta)$

Exercise 2

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. Obtain the asymptotic relative efficiency of the MLE of a with respect to the UMVUE of a.

Proof:

- Let $X_{(1)}$ be the smallest order statistic and $T = \sum_i X_i$
- The joint Lebesgue p.d.f. of $x = (x_1, \ldots, x_n)$ is

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

- So $\left(X_{(1)}, T - nX_{(1)}\right)$ is complete and sufficient for (a, θ) (the completeness is left for next time)
- By memoryless property and relation between $E(0,1)$ and $\Gamma(n,1)$, $\left(T - nX_{(1)}\right)/\theta$ has the gamma distribution $\Gamma(n-1)$ Then $E\left(T - nX_{(1)}\right) = (n-1)\theta$.
- By the famous Rènyi representation, $X_{(1)} \stackrel{\mathcal{D}}{=} a + n^{-1} E(0, \theta)$ and $EX_{(1)} = a + \theta/n$. The UMVUE of a is $X_{(1)} - \left(T - nX_{(1)}\right)/[n(n-1)]$
- The MLE of a is $\hat{a} = X_{(1)}$

- Note that $n(\hat{a} - a) = n\left(X_{(1)} - a\right) \overset{\mathcal{D}}{=} Z \sim E(0, \theta)$. So amse of $\hat{a}$ is $2\theta^2/n^2$
- For the UMVUE $\tilde{a}$

$$n(\tilde{a} - a) = n\left(X_{(1)} - a\right) - \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - X_{(1)}\right) \xrightarrow{\mathcal{D}} Z - \theta,$$

  since $\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - X_{(1)}\right) \to_p \theta$.
- The amse of $\tilde{a}$ is $E(Z - \theta)^2/n^2 = \theta^2/n^2$
- The asymptotic relative efficiency of $\hat{a}$ with respect to $\tilde{a}$ is $1/2$

Exercise 3

Consider a linear model in matrix form $X_{n\times 1} = Z_{n\times p}\boldsymbol{\beta}_{p\times 1} + \epsilon_{n\times 1}$.
Under the assumption that $\epsilon \sim N(\mathbf{0}_n, \sigma^2 \boldsymbol{I}_n)$ where σ is known, compute the Fisher information $I(\beta)$. When is $I(\beta)$ positive definite?

Solution:

The Fisher information matrix is

$$\frac{1}{\sigma^2} Z^\top Z.$$

It is positive definite only if Z is of full rank.