# Mathematical Statistics2 Tutoring3

Seung Bong Jung

Seoul National University

October 7, 2021

## 2018 Midterm Problem3

Let $X_1, \ldots, X_n$ be i.i.d. with p.d.f. $f_{\theta_1,\theta_2}$ $(0 < \theta_1, \theta_2)$, where

$$f_{\theta_1,\theta_2}(x) = \begin{cases} (\theta_1 + \theta_2)^{-1} e^{-x/\theta_1} & \text{if } x \geq 0 \\ (\theta_1 + \theta_2)^{-1} e^{x/\theta_2} & \text{if } x < 0 \end{cases}.$$

(a) Find a MLE $\hat{\theta}$ of $\theta = (\theta_1, \theta_2)^t$.
(b) Find a 2-dimensional vector $\mu$ and $2 \times 2$ matrix $\Sigma$ such that $\sqrt{n}(\hat{\theta} - \mu) \xrightarrow{\text{d}} N(0, \Sigma)$.

# One-step Approximation of MLE

- **Motivation)** If initial choice $\hat{\theta}^{[0]}$ lies in a $n^{-1/2}-$ neighborhood of the true parameter $\theta_0$ in probability, then one-step update in Newton-Raphson iteration is enough.

$$\hat{\theta}^{[1]} = \hat{\theta}^{[0]} - \ddot{l}(\hat{\theta}^{[0]})^{-1}\dot{l}(\hat{\theta}^{[0]}).$$

**Theorem 0.1**

*Under (R0)-(R7) and assuming* $\hat{\theta}^{[0]} = \theta_0 + O_p(n^{-1/2})$ *under* $P_{\theta_0}$*,*

$$\sqrt{n}(\hat{\theta}^{[1]} - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1})$$

*in* $P_{\theta_0}-$*probability.*

# Least Square Estimator

Consider the following model for

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon,\ \epsilon \sim N(0, \sigma^2)$$

for $p \geq 1$.

- $\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2 = \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\arg\min} ||\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}||^2$ where $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^t, \mathbf{X} = (\mathbf{1}, \mathbf{x_1}, \ldots, \mathbf{x_p})$ and $\mathbf{x}_j = (x_{1j}, \ldots, x_{nj})^t$.
- Denote the column space of $\mathbf{X}$ by $\mathcal{C}_{\mathbf{X}}$(the linear space in $\mathbb{R}^n$ spanned by the columns of the matrix $\mathbf{X}$.
- Then $\mathbf{X}\hat{\boldsymbol{\beta}} = \Pi(\mathbf{Y}|\mathcal{C}_{\mathbf{X}})$
- Least squared estimator: $\hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}$

# Confidence Region for $\boldsymbol{\beta}$

- Recall that $\hat{\sigma}^2 = 1/(n-p-1)||\mathbf{Y} - \mathbf{X}\hat{\beta}||^2 = 1/n\text{SSE}$.
- Distribution of $\hat{\boldsymbol{\beta}}$: $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^t\mathbf{X})^{-1})$ so that

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^t\mathbf{X}^t\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\sigma^2 \sim \chi^2(p+1).$$

- It can be shown that $\text{SSE}/\sigma^2 \sim \chi^2(n-p-1)$.
- Using the independence of $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^t(\mathbf{X}^t\mathbf{X})^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and SSE, one can see that $100(1-\alpha)\%$ confidence region for $\boldsymbol{\beta}$ is given by

$$\hat{\mathcal{I}} = \{\boldsymbol{\beta} : (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^t(\mathbf{X}^t\mathbf{X})^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq (p+1)\cdot\hat{\sigma}^2 \cdot F_\alpha(p+1, n-p-1)\}.$$

# Least Square Estimator of Subvector $\beta_2$

Write $\boldsymbol{\beta}^t = (\boldsymbol{\beta}_1^t, \boldsymbol{\beta}_2^t)$ and $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, so that $\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2$.

- Orthogonal decomposition of the column space:

$$\mathcal{C}_{\mathbf{X}} = \mathcal{C}_{\mathbf{X}_1, \mathbf{X}_2 - \Pi_1 \mathbf{X}_2} = \mathcal{C}_{\mathbf{X}_1} \oplus \mathcal{C}_{\mathbf{X}_2 - \Pi_1 \mathbf{X}_2}$$

  where $\Pi_1 = \mathbf{X}_1(\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t$.

- Writing $\mathbf{X}_{2,\perp} = \mathbf{X}_2 - \Pi_1\mathbf{X}_2$, we get

$$\begin{aligned} \mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}_1[\hat{\boldsymbol{\beta}}_1 + (\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{X}_2\hat{\boldsymbol{\beta}}_2] + \mathbf{X}_{2,\perp}\hat{\boldsymbol{\beta}}_2 \\ \Pi(\mathbf{Y}|\mathcal{C}_{\mathbf{X}}) &= \Pi_1\mathbf{Y} + \mathbf{X}_{2,\perp}(\mathbf{X}_{2,\perp}^t\mathbf{X}_{2,\perp})^{-1}\mathbf{X}_{2,\perp}^t\mathbf{Y} \end{aligned}$$

- This gives

$$\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t(\mathbf{Y} - \mathbf{X}_2\hat{\boldsymbol{\beta}}_2), \hat{\boldsymbol{\beta}}_2 = (\mathbf{X}_{2,\perp}^t\mathbf{X}_{2,\perp})^{-1}\mathbf{X}_{2,\perp}^t\mathbf{Y}.$$

# Confidence Region for $\boldsymbol{\beta}_2$

- Distribution of $\hat{\boldsymbol{\beta}}_2$: $(\mathbf{X}_{2,\perp}^t \mathbf{X}_{2,\perp})^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \sim N_q(\mathbf{0}_q, \sigma^2 \mathbf{I}_q)$, so that $(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)^t (\mathbf{X}_{2,\perp}^t \mathbf{X}_{2,\perp})(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \sim \sigma^2 \chi^2(q)$
- Distribution of SSE: SSE $\sim \sigma^2 \chi^2(n-p-1)$
- SSE and $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)$ are independent.
- Similarly, $100(1-\alpha)\%$ confidence region for $\boldsymbol{\beta}$ is given by

$$\hat{\mathcal{I}} = \{\boldsymbol{\beta}_2 : (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)^t \mathbf{X}_{2,\perp}^t \mathbf{X}_{2,\perp} (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \leq q \cdot \hat{\sigma}^2 \cdot F_\alpha(q, n-p-1)\}.$$

# Idea of Maximum Likelihood Test

Suppose that we observe a random sample $X_1, \ldots, X_n$ from a distribution $P_\theta$ with pdf $f(\cdot;\theta)$ for $\theta \in \Theta$.

- The problem: Testing

$$H_0 : \theta \in \Theta_0 \quad \text{against} \quad H_1 : \theta \in \Theta_1$$

  at a significance level $0 < \alpha < 1$, where $\Theta_0$ is a subset of $\Theta$ and $\Theta_1 = \Theta \setminus \Theta_0$.

- Fisher's idea: Compare the maximum likelihoods in $\Theta_0$ and $\Theta_1$. For a given observation $x \equiv (x_1, \ldots, x_n)$ of $X \equiv (X_1, \ldots, X_n)$, reject $H_0$ if

$$\max_{\theta \in \Theta_1} L(\theta; x) / \max_{\theta \in \Theta_0} L(\theta; x) \geq c$$

  for a critical value $c$ that is determined by the level $\alpha$ as follows.

# Idea of Maximum Likelihood Test

- Determination of critical value: Choose $c$ such that (the size of the test)$= \alpha$, i.e.,

$$\sup_{\theta \in \Theta_0} P_\theta\left(\frac{\max_{\theta \in \Theta_1} L(\theta; x)}{\max_{\theta \in \Theta_0} L(\theta; x)} > c\right) = \alpha.$$

- The maximization of the likelihood in a restricted set, such as $\Theta_1$, is more involved than the maximization of $\Theta$.
- Note that, for $c > 1$,

$$\begin{aligned} R_0 &\equiv \frac{\max_{\theta \in \Theta_1} L(\theta; X)}{\max_{\theta \in \Theta_0} L(\theta; X)} \geq c \\ &\Leftrightarrow R \equiv \frac{\max_{\theta \in \Theta} L(\theta; X)}{\max_{\theta \in \Theta_0} L(\theta; X)} = \max\{1, \frac{\max_{\theta \in \Theta_1} L(\theta; X)}{\max_{\theta \in \Theta_0} L(\theta; X)}\} \geq c. \end{aligned}$$

If there exists $c > 0$ such that $\sup_{\theta \in \Theta_0} P_\theta(R \geq c) = \alpha$, then $c > 1$ since $R \geq 1$ always and $\alpha < 1$. This means that the LRT may be based on $R$, rather than $R_0$.

# Likelihood Ratio Test

Let $\hat{\theta}^{\Theta}$ and $\hat{\theta}^{\Theta_0}$ denote the MLEs in $\Theta$ and $\Theta_0$, respectively. The likelihood ratio test (LRT), for $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$, rejects $H_0$ when

$$\frac{\max_{\theta \in \Theta} L(\theta; x)}{\max_{\theta \in \Theta_0} L(\theta; x)} \geq c$$

or equivalently when

$$2(l(\hat{\theta}^{\Theta}; x) - l(\hat{\theta}^{\Theta_0}; x)) \geq c',$$

where $c$ and $c'$ are determined by the given level $\alpha$.

# Examples of LRT: Normal Mean with Unknown Variance

Suppose that we observe a random sample $X_1, \ldots, X_n$ from $N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$, and want to test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ at a level $0 < \alpha < 1$.

- Rejection region: Let $s^2 = (n-1)^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$.

  Reject $H_0$ if $|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}| \geq t_{\alpha/2}(n-1)$.

- Wilk's phenomenon: Since $T_n \equiv (\bar{X} - \mu)/(S/\sqrt{n}) \stackrel{d}{=} t(n-1)$ under $H_0$ and $t(n-1) \xrightarrow{d} Z \stackrel{d}{=} N(0,1)$, we get

$$2(l(\hat{\theta}^{\Theta}) - l(\hat{\theta}^{\Theta_0})) = n \log(1 + T_n^2/(n-1)) = T_n^2 + o_p(1)$$
$$\xrightarrow{d} Z^2 \stackrel{d}{=} \chi^2(1)$$

under $H_0$. Note that (d.f.)=dim$(\Theta)$ − dim$(\Theta_0) = 2 - 1 = 1$.

# Exercises

## 2019 Final Problem5

Consider the following model

$$Y_{ij} = \mu + \alpha_i + \beta x_{ij} + \epsilon_{ij}$$

where $\epsilon_{ij} \overset{\text{i.id.}}{\sim} N(0, \sigma^2)$, $(i = 1, \ldots, I, j = 1, \ldots, n_i)$. Here $\mu, \alpha_i, \beta, \sigma^2$ are unknown parameters ($\mu, \alpha_i, \beta \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}_+$) and $x_{ij}$ are known covariates. We further assume that

$$\sum_{i=1}^{I} n_i \alpha_i = 0 \text{ and } \sum_{i=1}^{I} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 \neq 0 \text{ where } \bar{x}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}.$$

(a) Find the MLE of $(\mu, \alpha_i, \beta, \sigma^2)$ under the full model.
(b) For the following hypothesis $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$, find the LRT with significance level $0 < \alpha < 1$.

# Exercises

## 2019 Final Problem4

Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be i.i.d. samples where

$$(X_i, Y_i) \sim N\left(\left(\begin{array}{c} \mu_1 \\ \mu_2 \end{array}\right), \left(\begin{array}{cc} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{array}\right)\right).$$

Here $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)^t \in \mathbb{R}^2 \times \mathbb{R}^2 \times (-1, 1)$. Assume $n \geq 5$ and denote the sample correlation by $\hat{\rho}$.

(a) Find the distribution of $\hat{\rho}/\sqrt{1 - \hat{\rho}^2}$.

(b) Find the LRT with significance level $0 < \alpha < 1$ for testing hypothesis $H_0 : \rho \leq 0$ versus $H_1 : \rho > 0$.

# Exercises

## 2018 Final Problem1

Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be i.i.d. samples where

$$(X_i, Y_i) \sim N\left(\left(\begin{array}{c}\theta_1\\ \theta_2\end{array}\right), \left(\begin{array}{cc}1 & \rho\\ \rho & 1\end{array}\right)\right) \text{ for known } \rho \in (-1, 1).$$

We wish to test

$$H_0 : \theta_1 = \theta_2 = 0 \text{ versus } H_1 : \text{not } H_0.$$

(a) Suppose that the parameter space is given by $\Theta = \{(\theta_1, \theta_2) : \theta_2 \leq c\theta_1, \theta_1 \geq 0\}$ for $c \in \mathbb{R}$ and $\rho = 0$. Derive LRT of significance level $0 < \alpha < 1$.
(b)Suppose that the parameter space is given by $\Theta = \{(\theta_1, \theta_2) : \theta_1, \theta_2 \geq 0\}$. Derive LRT of significance level $0 < \alpha < 1$.