<u>MA329 Statistical linear models</u>

Assignment 4 (Due date: Nov 30, 11pm. For late submission, each day costs 10 percent)

1. (10 marks) Let

$$\begin{aligned} Y_1 &= \theta + \epsilon_1 \\ Y_2 &= 2\theta - \phi + \epsilon_2 \\ Y_3 &= \theta + 2\phi + \epsilon_3 \end{aligned}$$

   where $E[\epsilon_i] = 0$ $(i = 1; 2; 3)$. Find the least squares estimates of θ and ϕ.

2. (15 marks) In order to estimate two parameters θ and ϕ, it is possible to make observations of three types:

   (a) the first type have expectation θ,

   (b) the second type have expectation $\theta + \phi$, and

   (c) the third type have expectation $\theta - 2\phi$.

   All observations are subject to independent normal errors with zero means and common variance σ^2. If m observations of type (a), m observations of type (b) and n observations of type (c) are made, find the least squares estimates $\hat{\theta}$ and $\hat{\phi}$. Prove that these estimates are uncorrelated if $m = 2n$.

3. (15 marks) Consider the linear regression model

$$y = \boldsymbol{X}_{\boldsymbol{n\times p}}\boldsymbol{\beta}_{\boldsymbol{p\times 1}} + \boldsymbol{\epsilon}$$

   where $\boldsymbol{\epsilon} \sim N(\mathbf{0}; \sigma^2\boldsymbol{I})$. Let $\hat{\boldsymbol{\beta}}$ be the least squares estimate of $\boldsymbol{\beta}$. Define $\tilde{\boldsymbol{\beta}} = c\hat{\boldsymbol{\beta}}$ where $c \leq 1$. The mean squared error (MSE) of $\tilde{\boldsymbol{\beta}}$ is

$$MSE(\tilde{\boldsymbol{\beta}}) = \mathrm{E}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

   (a) Prove that $MSE(\tilde{\boldsymbol{\beta}}) = c^2\sigma^2\mathrm{tr}(\boldsymbol{X}'\boldsymbol{X})^{-1} + (c-1)^2\boldsymbol{\beta}'\boldsymbol{\beta}$.

   (b) Let c^* be the value of c such that $MSE(\tilde{\boldsymbol{\beta}})$ is a minimum. Find c^*.

   (c) Let $p = 5$, $\sigma^2 = 1$, $\boldsymbol{\beta}' = (1; 2; 3; 4; 5)$ and the eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$ be 1, 2, 3, 4, 5. Evaluate c^*.

4. (20 marks) For the following model

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2 v_i) \text{ iid}, \; i = 1, \ldots, n, \tag{1}$$

   where v_i is given but σ^2 and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)'$ are unknown.

(a) For $\boldsymbol{\beta}$, find its generalized LSE $\hat{\boldsymbol{\beta}}$ and its distribution;

(b) Find the distribution of $\tilde{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, where $\boldsymbol{y}$ and $\boldsymbol{X}$ are vector and matrix forms of equation (1).

(c) Prove that the variance of $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ is not larger than the variance of $\boldsymbol{\lambda}'\tilde{\boldsymbol{\beta}}$ for any $(k+1)-$dimensional vector $\boldsymbol{\lambda}$. You need to write down the detailed procedure.