LONDON TAUGHT COURSE CENTRE

LTCC Basic Course **Statistical Modelling and Estimation**

Exercise Sheet 1: Solutions **February/March 2012**

---

1. Assume $Y_i \sim N(\mu, \sigma^2)$ for $i = 1, \ldots, n$, with all random variables independent. Show how this can be written as a linear model, i.e. specify the matrix $\mathbf{X}$ and the vector $\boldsymbol{\beta}$.

   The linear model in matrix notation for this situation is
   $$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$
   where the design matrix $\mathbf{X} = [1\ 1\ \cdots\ 1]'$ has only a single column and $\boldsymbol{\beta} = [\mu]$. It is assumed that $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $V(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$.

2. Consider the model $Y_{ij} \sim N(\mu_i, \sigma^2)$, for data from two groups $i = 1, 2$, $j = 1, \ldots, n$ with $\mu_i = \mu + \alpha_i$, and all random variables independent.

   (a) Write this as a linear model.

   For simplicity of notation we assume that the n observations from the first group come first and the n observations from the second group next. The corresponding linear model is then
   $$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$
   where the $2n \times 3$ design matrix is given by
   $$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{bmatrix}$$
   and the parameter vector is $\boldsymbol{\beta} = [\mu\ \alpha_1\ \alpha_2]'$. The first n rows of $\mathbf{X}$ contain a 1 in the second and a 0 in the third column, whereas for the remaining n rows the second column contains a 0 and the third column the value 1. Note that here we are considering a normal linear model in which in addition to assuming $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $V(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$ it is also assumed that the random errors have a normal distribution.

   (b) Draw a sketch to show how a histogram of such data would look (for large n).

   The histogram would be bimodal. The point is that even in a normal linear model we should not expect a plot of the raw data to look as if they come from a single normal distribution.

3. Verify that the matrix $(\mathbf{X}'\mathbf{X})^-$ in Example 1.7, where the simple linear regression model was considered, is a generalized inverse of $\mathbf{X}'\mathbf{X}$.

   In Example 1.7 we have
   $$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & nx \\ nx & nx^2 \end{bmatrix}$$

and

$$(\mathbf{X}'\mathbf{X})^- = \frac{1}{(n+1)(nx^2+x^{*2})-(nx+x^*)^2}\begin{bmatrix} nx^2+x^{*2} & -(nx+x^*) \\ -(nx+x^*) & n+1 \end{bmatrix},$$

where $x^* \neq x$. We need to show that

$$\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}.$$

This is however true since

$$\begin{aligned}
&\begin{bmatrix} n & nx \\ nx & nx^2 \end{bmatrix}\begin{bmatrix} nx^2+x^{*2} & -(nx+x^*) \\ -(nx+x^*) & n+1 \end{bmatrix}\begin{bmatrix} n & nx \\ nx & nx^2 \end{bmatrix} \\
= &\begin{bmatrix} nx^{*2}-nxx^* & -nx^*+nx \\ nxx^{*2}-nx^2x^* & -nxx^*+nx^2 \end{bmatrix}\begin{bmatrix} n & nx \\ nx & nx^2 \end{bmatrix} \\
= &\begin{bmatrix} n^2x^{*2}-2n^2xx^*+n^2x^2 & n^2xx^{*2}-2n^2x^2x^*+n^2x^3 \\ n^2xx^{*2}-2n^2x^2x^*+n^2x^3 & n^2x^2x^{*2}-2n^2x^3x^*+n^2x^4 \end{bmatrix} \\
= &\ n(x^*-x)^2\begin{bmatrix} n & nx \\ nx & nx^2 \end{bmatrix}
\end{aligned}$$

and

$$(n+1)(nx^2+x^{*2})-(nx+x^*)^2 = n(x^*-x)^2.$$

4. Let $\mathbf{G}$ be a square matrix. Show that if $\mathbf{G}$ possesses an inverse $\mathbf{G}^{-1}$, then $\mathbf{G}^{-1}$ is the unique g-inverse of $\mathbf{G}$.

   Suppose that $\mathbf{G}$ has an inverse $\mathbf{G}^{-1}$ and that $\mathbf{G}^-$ is some g-inverse of $\mathbf{G}$. Let $\mathbf{I}$ be the identity matrix with the same number of rows and columns as $\mathbf{G}$. It follows that

   $$\mathbf{G}\mathbf{G}^- = \mathbf{G}\mathbf{G}^-\mathbf{I} = \mathbf{G}\mathbf{G}^-\mathbf{G}\mathbf{G}^{-1} = \mathbf{G}\mathbf{G}^{-1}$$

   since $\mathbf{G}\mathbf{G}^-\mathbf{G} = \mathbf{G}$ by definition of a g-inverse. Multiplying the equation $\mathbf{G}\mathbf{G}^- = \mathbf{G}\mathbf{G}^{-1}$ from the left with $\mathbf{G}^{-1}$ then shows that $\mathbf{G}^- = \mathbf{G}^{-1}$. Since this is true for every g-inverse, it follows that every g-inverse of $\mathbf{G}$ is equal to $\mathbf{G}^{-1}$. Thus $\mathbf{G}^{-1}$ is the unique g-inverse of $\mathbf{G}$.

5. Let $\mathbf{G}$ be a symmetric real $m \times m$ matrix. Denote the eigenvalues of $\mathbf{G}$ by $\lambda_1, \ldots, \lambda_m$ and let $\mathbf{P}$ be an $m \times m$ matrix whose columns are corresponding orthonormal eigenvectors. For $i = 1, \ldots, m$ set $\gamma_i = 1/\lambda_i$ if $\lambda_i \neq 0$ and $\gamma_i = 0$ otherwise.

   Show that $\mathbf{G}^+ = \mathbf{P}\text{diag}(\gamma_1, \ldots, \gamma_m)\mathbf{P}'$ is the Moore-Penrose inverse of $\mathbf{G}$, that is show that the given $\mathbf{G}^+$ has the four properties in Definition 1.2.

   The columns of $\mathbf{P}$ are orthonormal and so $\mathbf{P}'\mathbf{P} = \mathbf{I}$, where $\mathbf{I}$ is the $m \times m$ identity matrix. Thus $\mathbf{P}$ is an orthogonal matrix with inverse $\mathbf{P}^{-1} = \mathbf{P}'$. The fact that the columns of $\mathbf{P}$ are the eigenvectors corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_m$ of $\mathbf{G}$ implies that $\mathbf{P}'\mathbf{G}\mathbf{P} = \text{diag}(\lambda_1, \ldots, \lambda_m)$ and so $\mathbf{G} = \mathbf{P}\text{diag}(\lambda_1, \ldots, \lambda_m)\mathbf{P}'$.

   We need to show that the matrix $\mathbf{G}^+$ defined above has the four defining properties (a)-(d) of the Moore-Penrose inverse in Definition 1.2. Property (a) follows from

   $$\begin{aligned}
   \mathbf{G}\mathbf{G}^+\mathbf{G} &= \mathbf{P}\text{diag}(\lambda_1, \ldots, \lambda_m)\mathbf{P}'\mathbf{P}\text{diag}(\gamma_1, \ldots, \gamma_m)\mathbf{P}'\mathbf{P}\text{diag}(\lambda_1, \ldots, \lambda_m)\mathbf{P}' \\
   &= \mathbf{P}\text{diag}(\lambda_1, \ldots, \lambda_m)\text{diag}(\gamma_1, \ldots, \gamma_m)\text{diag}(\lambda_1, \ldots, \lambda_m)\mathbf{P}' \\
   &= \mathbf{P}\text{diag}(\lambda_1, \ldots, \lambda_m)\mathbf{P}' = \mathbf{G},
   \end{aligned}$$

where it has been used that $\lambda_i\gamma_i\lambda_i = \lambda_i$ for every $i = 1, \dots, m$. Similarly, the calculation

$$\begin{aligned}
\mathbf{G}^+\mathbf{G}\mathbf{G}^+ &= \mathbf{P}\text{diag}(\gamma_1, \dots, \gamma_m)\mathbf{P}'\mathbf{P}\text{diag}(\lambda_1, \dots, \lambda_m)\mathbf{P}'\mathbf{P}\text{diag}(\gamma_1, \dots, \gamma_m)\mathbf{P}' \\
&= \mathbf{P}\text{diag}(\gamma_1, \dots, \gamma_m)\text{diag}(\lambda_1, \dots, \lambda_m)\text{diag}(\gamma_1, \dots, \gamma_m)\mathbf{P}' \\
&= \mathbf{P}\text{diag}(\gamma_1, \dots, \gamma_m)\mathbf{P}' = \mathbf{G}^+
\end{aligned}$$

shows that $\mathbf{G}^+$ has property (b). Finally, the symmetry properties (c) and (d) are also fulfilled since

$$\begin{aligned}
(\mathbf{G}\mathbf{G}^+)' &= (\mathbf{P}\text{diag}(\lambda_1, \dots, \lambda_m)\mathbf{P}'\mathbf{P}\text{diag}(\gamma_1, \dots, \gamma_m)\mathbf{P}')' \\
&= \mathbf{P}\text{diag}(\gamma_1, \dots, \gamma_m)\mathbf{P}'\mathbf{P}\text{diag}(\lambda_1, \dots, \lambda_m)\mathbf{P}' \\
&= \mathbf{P}\text{diag}(\lambda_1, \dots, \lambda_m)\mathbf{P}'\mathbf{P}\text{diag}(\gamma_1, \dots, \gamma_m)\mathbf{P}' = \mathbf{G}\mathbf{G}^+
\end{aligned}$$

and

$$\begin{aligned}
(\mathbf{G}^+\mathbf{G})' &= (\mathbf{P}\text{diag}(\gamma_1, \dots, \gamma_m)\mathbf{P}'\mathbf{P}\text{diag}(\lambda_1, \dots, \lambda_m)\mathbf{P}')' \\
&= \mathbf{P}\text{diag}(\lambda_1, \dots, \lambda_m)\mathbf{P}'\mathbf{P}\text{diag}(\gamma_1, \dots, \gamma_m)\mathbf{P}' \\
&= \mathbf{P}\text{diag}(\gamma_1, \dots, \gamma_m)\mathbf{P}'\mathbf{P}\text{diag}(\lambda_1, \dots, \lambda_m)\mathbf{P}' = \mathbf{G}^+\mathbf{G}.
\end{aligned}$$