LONDON TAUGHT COURSE CENTRE

LTCC Basic Course — Statistical Modelling and Estimation

Exercise Sheet 2: Solutions — **February/March 2012**

---

1. Let $\mathbf{X}$ be the design matrix of a linear model. Show that if $\mathbf{G}$ and $\mathbf{H}$ are g-inverses of $\mathbf{X}'\mathbf{X}$, then

   (a) $\mathbf{XGX}'\mathbf{X} = \mathbf{XHX}'\mathbf{X} = \mathbf{X}$,

   We assume that $\mathbf{X}$ has n rows and p columns. Let $\mathbf{u}$ be a vector in $\mathbb{R}^n$. It is known from Linear Algebra that $\mathbf{u}$ can be written as $\mathbf{u} = \mathbf{v} + \mathbf{w}$, where $\mathbf{v} \in C(\mathbf{X})$ and $\mathbf{w} \in C(\mathbf{X})^\perp$ are uniquely determined. Here, $C(\mathbf{X})$ is the column space of $\mathbf{X}$ and $C(\mathbf{X})^\perp$ the orthogonal complement of $C(\mathbf{X})$. Since $\mathbf{v} \in C(\mathbf{X})$, there exists a vector $\mathbf{b} \in \mathbb{R}^p$ such that $\mathbf{v} = \mathbf{Xb}$. It then follows that

   $$\mathbf{u}'\mathbf{XGX}'\mathbf{X} = (\mathbf{v}+\mathbf{w})'\mathbf{XGX}'\mathbf{X} = \mathbf{v}'\mathbf{XGX}'\mathbf{X} + \mathbf{w}'\mathbf{XGX}'\mathbf{X} = \mathbf{v}'\mathbf{XGX}'\mathbf{X},$$

   as $\mathbf{w}$ is orthogonal to every column of $\mathbf{X}$. Since $\mathbf{v} = \mathbf{Xb}$ and since $\mathbf{G}$ is a g-inverse of $\mathbf{X}'\mathbf{X}$ we further have

   $$\mathbf{v}'\mathbf{XGX}'\mathbf{X} = \mathbf{b}'\mathbf{X}'\mathbf{XGX}'\mathbf{X} = \mathbf{b}'\mathbf{X}'\mathbf{X} = \mathbf{v}'\mathbf{X} = \mathbf{u}'\mathbf{X}$$

   and so $\mathbf{u}'(\mathbf{XGX}'\mathbf{X} - \mathbf{X}) = \mathbf{0}$. Since this is true for every vector $\mathbf{u} \in \mathbb{R}^n$ it finally follows that $\mathbf{XGX}'\mathbf{X} - \mathbf{X} = \mathbf{0}$ and so $\mathbf{XGX}'\mathbf{X} = \mathbf{X}$. Note that this result does not depend on the choice of the g-inverse. In other words, if $\mathbf{H}$ is another g-inverse of $\mathbf{X}'\mathbf{X}$ then exactly the same argument shows that $\mathbf{XHX}'\mathbf{X} = \mathbf{X}$.

   (b) $\mathbf{XGX}' = \mathbf{XHX}'$.

   As in part (a) let $\mathbf{u}$ be a vector in $\mathbb{R}^n$ and write $\mathbf{u} = \mathbf{v} + \mathbf{w}$, where $\mathbf{v} \in C(\mathbf{X})$ and $\mathbf{w} \in C(\mathbf{X})^\perp$. As before $\mathbf{v} = \mathbf{Xb}$ for some $\mathbf{b} \in \mathbb{R}^p$. The result in part (a) then implies that

   $$\mathbf{XGX}'\mathbf{u} = \mathbf{XGX}'\mathbf{v} = \mathbf{XGX}'\mathbf{Xb} = \mathbf{XHX}'\mathbf{Xb} = \mathbf{XHX}'\mathbf{v} = \mathbf{XHX}'\mathbf{u}.$$

   Since this is true for every $\mathbf{u} \in \mathbb{R}^n$ it follows that $\mathbf{XGX}' = \mathbf{XHX}'$.

2. Let $\mathbf{X}$ be the design matrix of a linear model and $\mathbf{M} = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'$. Show that regardless of the choice of $(\mathbf{X}'\mathbf{X})^-$

   (a) $\mathbf{M}^2 = \mathbf{M}$,

   This follows from the result in Question 1(a) since

   $$\mathbf{M}^2 = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}' = \mathbf{M}.$$

   (b) $\mathbf{M}$ is symmetric.

   The construction on p. 10 of the lecture notes shows that the Moore-Penrose inverse $\mathbf{G}^+$ of any real symmetric matrix $\mathbf{G}$ is symmetric. So in particular $(\mathbf{X}'\mathbf{X})^+$ is symmetric which implies that

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

The result in Question 1(b) then allows us to verify that $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ is also symmetric for any other g-inverse $(\mathbf{X}'\mathbf{X})^{-}$, since it implies that

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

Thus regardless of the choice of the g-inverse of $\mathbf{X}'\mathbf{X}$ we have that $\mathbf{M}' = \mathbf{M}$, that is $\mathbf{M}$ is symmetric.

3. Consider a linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $V(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$. For $S(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ show that

$$\frac{\partial S}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}.$$

By using the notation in Section 1.2 of the lecture notes we can write

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left(Y_i - \sum_{s=1}^{p} f_s(\mathbf{x}_i)\beta_s\right)^2.$$

For every $t = 1, \ldots, p$ differentiating this with respect to β_t gives

$$\begin{aligned}\frac{\partial S(\boldsymbol{\beta})}{\partial \beta_t} &= -2\sum_{i=1}^{n} f_t(\mathbf{x}_i)\left(Y_i - \sum_{s=1}^{p} f_s(\mathbf{x}_i)\beta_s\right)\\ &= -2\sum_{i=1}^{n} f_t(\mathbf{x}_i)Y_i + 2\sum_{i=1}^{n}\left(f_t(\mathbf{x}_i)\sum_{s=1}^{p} f_s(\mathbf{x}_i)\beta_s\right)\end{aligned}$$

and so

$$\begin{aligned}\frac{\partial S}{\partial \boldsymbol{\beta}} &= \begin{bmatrix}\frac{\partial S(\boldsymbol{\beta})}{\partial \beta_1}\\ \vdots \\ \frac{\partial S(\boldsymbol{\beta})}{\partial \beta_p}\end{bmatrix} = -2\begin{bmatrix}\sum_{i=1}^n f_1(\mathbf{x}_i)Y_i\\ \vdots\\ \sum_{i=1}^n f_p(\mathbf{x}_i)Y_i\end{bmatrix} + 2\begin{bmatrix}\sum_{i=1}^n \left(f_1(\mathbf{x}_i)\sum_{s=1}^p f_s(\mathbf{x}_i)\beta_s\right)\\ \vdots\\ \sum_{i=1}^n \left(f_p(\mathbf{x}_i)\sum_{s=1}^p f_s(\mathbf{x}_i)\beta_s\right)\end{bmatrix}\\ &= -2\begin{bmatrix} f_1(\mathbf{x}_1) & \cdots & f_1(\mathbf{x}_n)\\ \vdots & & \vdots\\ f_p(\mathbf{x}_1) & \cdots & f_p(\mathbf{x}_n)\end{bmatrix}\begin{bmatrix}Y_1\\ \vdots\\ Y_n\end{bmatrix} + 2\begin{bmatrix} f_1(\mathbf{x}_1) & \cdots & f_1(\mathbf{x}_n)\\ \vdots & & \vdots\\ f_p(\mathbf{x}_1) & \cdots & f_p(\mathbf{x}_n)\end{bmatrix}\begin{bmatrix}\sum_{s=1}^p f_s(\mathbf{x}_1)\beta_s\\ \vdots\\ \sum_{s=1}^p f_s(\mathbf{x}_n)\beta_s\end{bmatrix}\\ &= -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\begin{bmatrix} f_1(\mathbf{x}_1) & \cdots & f_p(\mathbf{x}_1)\\ \vdots & & \vdots\\ f_1(\mathbf{x}_n) & \cdots & f_p(\mathbf{x}_n)\end{bmatrix}\begin{bmatrix}\beta_1\\ \vdots\\ \beta_p\end{bmatrix}\\ &= -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}.\end{aligned}$$

4. Consider the linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $V(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$.

   (a) Prove the equivalence of parts (a) and (c) of Theorem 2.3, that is show that $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable if and only if $\boldsymbol{\lambda}$ is in the column space $C(\mathbf{X}'\mathbf{X})$ of the matrix $\mathbf{X}'\mathbf{X}$.

   We assume that the design matrix $\mathbf{X}$ has n rows and p columns. Suppose that $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable, so that $\boldsymbol{\lambda}' = \mathbf{a}'\mathbf{X}$ for some vector $\mathbf{a} \in \mathbb{R}^n$ by Theorem 2.1. Let

$\mathbf{M} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ be the orthogonal projection matrix onto $C(\mathbf{X})$, that is the projection matrix onto the column space of $\mathbf{X}$. It follows that $\mathbf{X} = \mathbf{M}\mathbf{X}$ and so

$$\boldsymbol{\lambda} = \mathbf{X}'\mathbf{a} = \mathbf{X}'\mathbf{M}'\mathbf{a} = \mathbf{X}'\mathbf{M}\mathbf{a} = \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{a} = \mathbf{X}'\mathbf{X}\mathbf{b},$$

where $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{a}$ is a vector in $\mathbb{R}^p$. Hence $\boldsymbol{\lambda}$ is an element of the column space $C(\mathbf{X}'\mathbf{X})$. Note that we have used that $\mathbf{M}$ is symmetric.

For the converse assume that $\boldsymbol{\lambda} \in C(\mathbf{X}'\mathbf{X})$, so that $\boldsymbol{\lambda} = \mathbf{X}'\mathbf{X}\mathbf{b}$ for some vector $\mathbf{b} \in \mathbb{R}^p$. It then follows that $\boldsymbol{\lambda} = \mathbf{X}'\mathbf{a}$ or equivalently $\boldsymbol{\lambda}' = \mathbf{a}'\mathbf{X}$, where $\mathbf{a} = \mathbf{X}\mathbf{b}$ is a vector in $\mathbb{R}^n$. Hence $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable by Theorem 2.1.

(b) Use the equivalence you proved in part (a) to prove Theorem 2.4, that is show that $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable if and only if

$$\boldsymbol{\lambda}'\{\mathbf{I} - (\mathbf{X}'\mathbf{X})^{-}(\mathbf{X}'\mathbf{X})\} = \mathbf{0}.$$

Obviously, if $\boldsymbol{\lambda}'\{\mathbf{I} - (\mathbf{X}'\mathbf{X})^{-}(\mathbf{X}'\mathbf{X})\} = \mathbf{0}$, then $\boldsymbol{\lambda}' = \boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-}(\mathbf{X}'\mathbf{X})$ and transposing the latter equation shows that $\boldsymbol{\lambda} \in C(\mathbf{X}'\mathbf{X})$. Hence it follows from part (a) that $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable. Conversely, if $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable, then $\boldsymbol{\lambda} \in C(\mathbf{X}'\mathbf{X})$ by part (a) and so there exists a vector $\mathbf{b} \in \mathbb{R}^p$ such that $\boldsymbol{\lambda} = (\mathbf{X}'\mathbf{X})\mathbf{b}$. Thus

$$\boldsymbol{\lambda}'\{\mathbf{I} - (\mathbf{X}'\mathbf{X})^{-}(\mathbf{X}'\mathbf{X})\} = \mathbf{b}'\{\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}(\mathbf{X}'\mathbf{X})\} = \mathbf{b}'\{\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{X}\} = \mathbf{0}.$$

(c) Show that all linear functions $\boldsymbol{\lambda}'\boldsymbol{\beta}$ are estimable, if and only if $\text{rank}(\mathbf{X}) = p$.

This result can be proved in several ways. The probably quickest proof uses Theorem 2.4. First note that the $n \times p$ design $\mathbf{X}$ has rank p if and only if the $p \times p$ matrix $\mathbf{X}'\mathbf{X}$ has rank p which in turn is equivalent to the fact that $\mathbf{X}'\mathbf{X}$ has an inverse $(\mathbf{X}'\mathbf{X})^{-1}$. Further, according to Theorem 2.4 all linear functions $\boldsymbol{\lambda}'\boldsymbol{\beta}$ are estimable if and only if $\boldsymbol{\lambda}'\{\mathbf{I} - (\mathbf{X}'\mathbf{X})^{-}(\mathbf{X}'\mathbf{X})\} = \mathbf{0}$ for every $\boldsymbol{\lambda} \in \mathbb{R}^p$.

If $\text{rank}(\mathbf{X}) = p$, then $(\mathbf{X}'\mathbf{X})^{-} = (\mathbf{X}'\mathbf{X})^{-1}$ for every g-inverse of $\mathbf{X}'\mathbf{X}$. It then follows that $\mathbf{I} - (\mathbf{X}'\mathbf{X})^{-}(\mathbf{X}'\mathbf{X}) = \mathbf{I} - (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X}) = \mathbf{0}$ which implies that every function $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable. For the converse, note that $\boldsymbol{\lambda}'\{\mathbf{I} - (\mathbf{X}'\mathbf{X})^{-}(\mathbf{X}'\mathbf{X})\} = \mathbf{0}$ for every $\boldsymbol{\lambda} \in \mathbb{R}^p$ implies that $\mathbf{I} - (\mathbf{X}'\mathbf{X})^{-}(\mathbf{X}'\mathbf{X}) = \mathbf{0}$ and so $\mathbf{X}'\mathbf{X}$ has an inverse from which it follows that $\text{rank}(\mathbf{X}) = p$.

5. Consider the multiple regression model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

for $i = 1, \ldots, 4$ with design matrix

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

and parameter vector $\boldsymbol{\beta} = [\beta_0\ \beta_1\ \beta_2]'$. The random errors $\epsilon_1, \ldots, \epsilon_4$ are assumed to be uncorrelated with mean zero and constant variance σ^2.

(a) Find rank($\mathbf{X}$).

The first two columns of $\mathbf{X}$ are linearly independent, while the third is equal to the second column, so rank($\mathbf{X}$) = 2.

(b) Find a generalized inverse of $\mathbf{X}'\mathbf{X}$.

We can use $\mathbf{X}^* = [1 \;\; -1 \quad 1]$ to augment $\mathbf{X}$, for which

$$\mathbf{X}^{*\prime}\mathbf{X}^* = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}.$$

Since

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 4 \\ 0 & 4 & 4 \end{bmatrix},$$

we have

$$\mathbf{X}'\mathbf{X} + \mathbf{X}^{*\prime}\mathbf{X}^* = \begin{bmatrix} 5 & -1 & 1 \\ -1 & 5 & 3 \\ 1 & 3 & 5 \end{bmatrix},$$

so that

$$(\mathbf{X}'\mathbf{X})^- = (\mathbf{X}'\mathbf{X} + \mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} = \frac{1}{8}\begin{bmatrix} 2 & 1 & -1 \\ 1 & 3 & -2 \\ -1 & -2 & 3 \end{bmatrix}.$$

Note that there are other generalized inverses.

(c) Hence find a least squares estimator of $\boldsymbol{\beta}$.

Using the previous results a least squares estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{Y} \quad = \frac{1}{4}\begin{bmatrix} Y_1 + Y_2 + Y_3 + Y_4 \\ Y_3 + Y_4 \\ -Y_1 - Y_2 \end{bmatrix} = \begin{bmatrix} \bar{Y} \\ \frac{1}{4}(Y_3 + Y_4) \\ -\frac{1}{4}(Y_1 + Y_2) \end{bmatrix}.$$

(d) Check whether or not β_1 is estimable.

$\beta_1 = \boldsymbol{\lambda}'\boldsymbol{\beta}$, where $\boldsymbol{\lambda}' = [0\ 1\ 0]$. To check whether this is estimable we use Theorem 2.4. Now

$$(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X} = \frac{1}{2}\begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix}$$

and

$$\boldsymbol{\lambda}'\{\mathbf{I} - (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}\} = \frac{1}{2}[0\ 1\ 0]\begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix} = \frac{1}{2}[-1 \quad 1 \; -1],$$

so β_1 is not estimable.

(e) Check whether or not $\beta_1 + \beta_2$ is estimable.

$\beta_1 + \beta_2 = \boldsymbol{\lambda}'\boldsymbol{\beta}$, where $\boldsymbol{\lambda}' = [0\ 1\ 1]$. Hence $\boldsymbol{\lambda}'\{\mathbf{I} - (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}\} = [0\ 0\ 0]$ and so $\beta_1 + \beta_2$ is estimable by Theorem 2.4.