LONDON TAUGHT COURSE CENTRE

LTCC Basic Course — Statistical Modelling and Estimation

Exercise Sheet 4: Solutions — **February/March 2012**

---

1. Consider the normal linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where the components $\epsilon_1, \ldots, \epsilon_n$ of $\boldsymbol{\epsilon}$ are independent random variables with $\epsilon_i \sim N(0, \sigma^2)$ for $i = 1, \ldots, n$. Derive the maximum likelihood estimator of σ^2.

   From the likelihood

$$L(\boldsymbol{\beta}, \sigma^2|\mathbf{Y}) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right\}.$$

   given in lectures, the log-likelihood is

$$\log L(\boldsymbol{\beta}, \sigma^2|\mathbf{Y}) = \log[(2\pi)^{-n/2}] - \frac{n}{2}\log[\sigma^2] - \frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).$$

   When doing maximum likelihood estimation we simultaneously maximize $\log L(\boldsymbol{\beta}, \sigma^2|\mathbf{Y})$ as a function of $\boldsymbol{\beta}$ and σ^2. Thus we need to calculate $\partial \log L(\boldsymbol{\beta}, \sigma^2|\mathbf{Y})/\partial\boldsymbol{\beta}$ and also $\partial \log L(\boldsymbol{\beta}, \sigma^2|\mathbf{Y})/\partial\sigma^2$ and to equate both to zero. As was explained in the lecture, the equation $\partial \log L(\boldsymbol{\beta}, \sigma^2|\mathbf{Y})/\partial\boldsymbol{\beta} = \mathbf{0}$ is equivalent to the normal equations and so a least squares estimate $\hat{\boldsymbol{\beta}}$ is also a maximum likelihood estimate of $\boldsymbol{\beta}$.

   The partial derivative of the log-likelihood with respect to σ^2 is

$$\frac{\partial \log L(\boldsymbol{\beta}, \sigma^2|\mathbf{Y})}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).$$

   Equating this to zero and multiplying through with $2\sigma^4$, we get

$$-n\sigma^2 + (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = 0$$

   or

$$\sigma^2 = \frac{(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})}{n}.$$

   Since we are simultaneously maximizing the log-likelihood with respect to $\boldsymbol{\beta}$ and σ^2 it follows that the maximum likelihood estimator of σ^2 is then given by

$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n},$$

   where $\hat{\boldsymbol{\beta}}$ is a least squares estimate of $\boldsymbol{\beta}$.

2. Let $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $V(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$, be a linear model and suppose that $\sigma^2 > 0$. Prove that the BLUE of any estimable function $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is unique and hence equal to the least squares estimate $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$.

   Since $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable there exists a vector $\mathbf{a} \in \mathbb{R}^n$ such that $\boldsymbol{\lambda}' = \mathbf{a}'\mathbf{X}$. Let $\mathbf{b}'\mathbf{Y}$ be any linear unbiased estimator of $\boldsymbol{\lambda}'\boldsymbol{\beta}$. In the proof of the Gauss-Markov Theorem 2.6 it was shown that

$$Var(\mathbf{b}'\mathbf{Y}) = Var(\mathbf{b}'\mathbf{Y} - \mathbf{a}'\mathbf{M}\mathbf{Y}) + Var(\mathbf{a}'\mathbf{M}\mathbf{Y}),$$

where $\mathbf{M}$ is the orthogonal projection matrix onto $C(\mathbf{X})$. Moreover, it was noted that $Var(\mathbf{a}'\mathbf{M}\mathbf{Y})$ is the variance of the least squares estimator $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\lambda}'\boldsymbol{\beta}$.

If $\mathbf{b}'\mathbf{Y}$ is a BLUE, then its variance has to be equal to the variance of $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$, since by the Gauss-Markov Theorem the least squares estimator is a best linear unbiased estimator. It follows that $Var(\mathbf{b}'\mathbf{Y} - \mathbf{a}'\mathbf{M}\mathbf{Y}) = 0$ and further that

$$0 = Var((\mathbf{b}' - \mathbf{a}'\mathbf{M})\mathbf{Y}) = (\mathbf{b}' - \mathbf{a}'\mathbf{M})V(\mathbf{Y})(\mathbf{b}' - \mathbf{a}'\mathbf{M})' = \sigma^2(\mathbf{b}' - \mathbf{a}'\mathbf{M})(\mathbf{b}' - \mathbf{a}'\mathbf{M})'.$$

Hence $\mathbf{b}' - \mathbf{a}'\mathbf{M}$ is a zero row vector since $\sigma^2 > 0$ and so $\mathbf{b}' = \mathbf{a}'\mathbf{M}$. Thus

$$\mathbf{b}'\mathbf{Y} = \mathbf{a}'\mathbf{M}\mathbf{Y} = \mathbf{a}'\mathbf{X}\hat{\boldsymbol{\beta}} = \boldsymbol{\lambda}'\hat{\boldsymbol{\beta}},$$

by Theorem 2.2. In other words the BLUE $\mathbf{b}'\mathbf{Y}$ is equal to the least squares estimator of $\boldsymbol{\lambda}'\boldsymbol{\beta}$.

Since $\mathbf{b}'\mathbf{Y}$ could have been any BLUE it follows that every best linear unbiased estimator of $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is equal to the least squares estimator, which shows that the BLUE is unique and equal to the least squares estimator.

3. A random variable T has a t_m distribution, that is a t distribution with m degrees of freedom, if it can be written as

$$T = \frac{X}{\sqrt{Y/m}}$$

where X and Y are independent random variables which respectively have a standard normal and a chi-square distribution with m degrees of freedom, that is $X \sim N(0,1)$ and $Y \sim \chi^2_m$. By using this relationship and Theorems 3.1-3.3 from the lectures prove Theorem 3.4 in the lecture notes.

Theorem 3.1 shows that $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\lambda}'\boldsymbol{\beta}, \sigma^2\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^-\boldsymbol{\lambda})$ and so

$$X = \frac{\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} - \boldsymbol{\lambda}'\boldsymbol{\beta}}{\sqrt{\sigma^2\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^-\boldsymbol{\lambda}}}$$

has a standard normal distribution. Further it follows from Theorem 3.2 that

$$Y = \frac{n-r}{\sigma^2}\text{MSE} \sim \chi^2_{n-r},$$

where $\text{MSE} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})/(n-r)$ is the mean square error and $r = \text{rank}(\mathbf{X})$. Theorem 3.3 implies that $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ and $(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ are independent. Now X is a function of $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ and Y is a function of $(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$, hence X and Y are also independent. Thus $X \sim N(0,1)$ and $Y \sim \chi^2_{n-r}$ and since the two random variables are independent it follows that

$$\frac{X}{\sqrt{Y/(n-r)}} \sim t_{n-r}.$$

The result then follows by noting that

$$\frac{X}{\sqrt{Y/(n-r)}} = \frac{\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} - \boldsymbol{\lambda}'\boldsymbol{\beta}}{\sqrt{\sigma^2\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^-\boldsymbol{\lambda}}}\sqrt{\frac{\sigma^2}{\text{MSE}}} = \frac{\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} - \boldsymbol{\lambda}'\boldsymbol{\beta}}{\sqrt{\text{MSE}\,\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^-\boldsymbol{\lambda}}}.$$

4. Consider the full second-order polynomial regression model (see Example 1.6)

$$Y_i = \beta_0 + \sum_{r=1}^{3} \beta_r x_{ir} + \sum_{r=1}^{3} \beta_{rr} x_{ir}^2 + \sum_{r=1}^{2} \sum_{s=r+1}^{3} \beta_{rs} x_{ir} x_{is} + \epsilon_i,$$

with independent random errors $\epsilon_i \sim N(0, \sigma^2)$ for the half-replicate fractional factorial design for three factors, each at two levels, plus two centre points shown below:

X_1	X_2	X_3
-1	-1	-1
-1	1	1
1	-1	1
1	1	-1
0	0	0
0	0	0

(a) Show that the function $\beta_1 - \beta_{23}$ of the model parameters is estimable.

The design matrix for the given design with $n = 6$ observations in the second-order polynomial regression model with three factors is

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and has rank $r = 5$. The corresponding parameter vector is

$$\boldsymbol{\beta} = [\beta_0\ \beta_1\ \beta_2\ \beta_3\ \beta_{11}\ \beta_{22}\ \beta_{33}\ \beta_{12}\ \beta_{13}\ \beta_{23}]'.$$

The linear function $\beta_1 - \beta_{23}$ can be written as $\boldsymbol{\lambda}'\boldsymbol{\beta}$ where $\boldsymbol{\lambda}' = [0\,1\,0\,0\,0\,0\,0\,0\,0\,-1]$. By using Theorem 2.4 and, for example, the generalized inverse

$$(\mathbf{X}'\mathbf{X})^- = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & 0 & 0 & 0 \\ 0 & \frac{1}{16} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{16} \\ 0 & 0 & \frac{1}{16} & 0 & 0 & 0 & 0 & 0 & -\frac{1}{16} & 0 \\ 0 & 0 & 0 & \frac{1}{16} & 0 & 0 & 0 & -\frac{1}{16} & 0 & 0 \\ -\frac{1}{6} & 0 & 0 & 0 & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & 0 & 0 & 0 \\ -\frac{1}{6} & 0 & 0 & 0 & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & 0 & 0 & 0 \\ -\frac{1}{6} & 0 & 0 & 0 & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{16} & 0 & 0 & 0 & \frac{1}{16} & 0 & 0 \\ 0 & 0 & -\frac{1}{16} & 0 & 0 & 0 & 0 & 0 & \frac{1}{16} & 0 \\ 0 & -\frac{1}{16} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{16} \end{bmatrix}$$

of $\mathbf{X}'\mathbf{X}$ it can then be shown that $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable.

(b) Assuming that the least squares estimate of $\beta_1 - \beta_{23}$ is equal to $\widehat{\beta_1 - \beta_{23}} = 10.03$ and and that the mean square error is MSE = 5.34, find a 95% confidence interval for $\beta_1 - \beta_{23}$. How would the interval be interpreted?

To find the 95% confidence interval for $\beta_1 - \beta_{23} = \boldsymbol{\lambda}'\boldsymbol{\beta}$ we note that according to Theorem 3.4 (see also Question 3)

$$\frac{\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} - \boldsymbol{\lambda}'\boldsymbol{\beta}}{\sqrt{\text{MSE}\,\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{\lambda}}} \sim t_{n-r}.$$

Here we have $n = 6$, $r = 5$. It follows that a 95% confidence interval for $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is given by

$$\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} \pm t_{1,0.975}\sqrt{\text{MSE}\,\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{\lambda}},$$

where $t_{1,0.975} = 12.71$ is the upper 2.5% point of the t distribution with one degree of freedom, which can be found in Tables such as the New Cambridge Statistical Tables.

The residual mean square is given as MSE = 5.34. By using the g-inverse on the previous page it follows that $\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{\lambda} = \frac{1}{4}$. Note, however, than any other g-inverse would give the same result, since $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable. By further using the given least squares estimate $\widehat{\beta_1 - \beta_{23}} = 10.03$ the 95% confidence interval for $\beta_1 - \beta_{23}$ is obtained as

$$10.03 \pm 12.71\sqrt{5.34 \times 0.25},$$

or $(-4.655, 24.715)$.

This can be interpreted in the usual way for the difference between the effect of factor X_1 and the interaction of X_2 and X_3, that is if we were to calculate the 95% confidence interval for a large number of experiments which use the same design, then 95% of these intervals would contain the true difference $\beta_1 - \beta_{23}$ of the parameters. Since the interval contains the value 0 we can also conclude that a two-sided test at the 5% level of significance would not reject the null hypothesis $H_0 : \beta_1 = \beta_{23}$.

5. Suppose that $\mathbf{M}$ and $\mathbf{M}_0$ are orthogonal projection matrices onto $C(\mathbf{M})$ and $C(\mathbf{M}_0)$, where $C(\mathbf{M}_0) \subset C(\mathbf{M})$.

(a) Prove that $C(\mathbf{M} - \mathbf{M}_0)$ is equal to the orthogonal complement of $C(\mathbf{M}_0)$ with respect to $C(\mathbf{M})$, that is $C(\mathbf{M} - \mathbf{M}_0) = C(\mathbf{M}) \cap C(\mathbf{M}_0)^{\perp}$.

The matrices $\mathbf{M}$ and $\mathbf{M}_0$ are symmetric and idempotent since they are orthogonal projection matrices. Moreover, $\mathbf{M}\mathbf{M_0} = \mathbf{M}_0$ since $C(\mathbf{M}_0) \subset C(\mathbf{M})$.

Suppose that $\mathbf{v} \in C(\mathbf{M} - \mathbf{M}_0)$ so that $\mathbf{v} = (\mathbf{M} - \mathbf{M}_0)\mathbf{u}$ for some vector $\mathbf{u}$. It follows that $\mathbf{v} = \mathbf{M}\mathbf{u} - \mathbf{M}_0\mathbf{u}$. Since $\mathbf{M}\mathbf{u} \in C(\mathbf{M})$, $\mathbf{M}_0\mathbf{u} \in C(\mathbf{M}_0)$ and $C(\mathbf{M}_0) \subset C(\mathbf{M})$ it follows that $\mathbf{v} \in C(\mathbf{M})$. Further

$$\begin{aligned}
\mathbf{v}'\mathbf{M}_0 &= \mathbf{u}'(\mathbf{M} - \mathbf{M}_0)'\mathbf{M}_0 = \mathbf{u}'(\mathbf{M}' - \mathbf{M}_0')\mathbf{M}_0 = \mathbf{u}'(\mathbf{M} - \mathbf{M}_0)\mathbf{M}_0 \\
&= \mathbf{u}'\mathbf{M}\mathbf{M}_0 - \mathbf{u}'\mathbf{M}_0^2 = \mathbf{u}'\mathbf{M}_0 - \mathbf{u}'\mathbf{M}_0 = \mathbf{0}.
\end{aligned}$$

Hence $\mathbf{v}$ is orthogonal to every column of $\mathbf{M}_0$ which implies that $\mathbf{v} \in C(\mathbf{M}_0)^{\perp}$. Thus $\mathbf{v} \in C(\mathbf{M}) \cap C(\mathbf{M}_0)^{\perp}$.

Now assume that $\mathbf{v} \in C(\mathbf{M}) \cap C(\mathbf{M}_0)^{\perp}$. It follows that

$$(\mathbf{M} - \mathbf{M}_0)\mathbf{v} = \mathbf{M}\mathbf{v} - \mathbf{M}_0\mathbf{v} = \mathbf{v} - \mathbf{M}_0'\mathbf{v} = \mathbf{v}$$

and so $\mathbf{v} \in C(\mathbf{M} - \mathbf{M}_0)$.

(b) Show that $\mathbf{M} - \mathbf{M}_0$ is symmetric and idempotent.

The symmetry of $\mathbf{M} - \mathbf{M}_0$ follows immediately from the symmetry of $\mathbf{M}$ and $\mathbf{M}_0$. Further

$$\begin{aligned}(\mathbf{M} - \mathbf{M}_0)^2 &= \mathbf{M}^2 - \mathbf{M}\mathbf{M}_0 - \mathbf{M}_0\mathbf{M} + \mathbf{M}_0^2 = \mathbf{M} - \mathbf{M}_0 - (\mathbf{M}\mathbf{M}_0)' + \mathbf{M}_0 \\ &= \mathbf{M} - \mathbf{M}_0 - \mathbf{M}_0 + \mathbf{M}_0 = \mathbf{M} - \mathbf{M}_0\end{aligned}$$

and so $\mathbf{M} - \mathbf{M}_0$ is also idempotent.