Chase Joyner

802 Homework 1

January 20, 2017

Consider the Gauss-Markov model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \;\; \boldsymbol{\epsilon} \sim N\big(\mathbf{0}, \sigma^2\mathbf{I}\big),$$

where $\mathbf{X} \in \mathbb{R}^{n\times(k+1)}$ is of full-column rank (i.e., full rank with $k+1 < n$) and both $\boldsymbol{\beta}$ and σ^2 are unknown. Let SSE denote the residual sum of squares from fitting this model.

1. Derive the maximum likelihood estimator of $\boldsymbol{\beta}$ and σ^2.

   Solution: The likelihood function is $N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, i.e.

$$L\big(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{y}\big) \propto \big(\sigma^2\big)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\}.$$

   Then, the log likelihood function is

$$\begin{aligned}\ell\big(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{y}\big) &\propto -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \\ &\propto -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\Big[\mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\Big]\end{aligned}$$

   Differentiating the log likelihood function above, we have

$$\begin{aligned}\frac{d\ell}{d\boldsymbol{\beta}} &= \frac{1}{\sigma^2}\mathbf{X}'\mathbf{y} - \frac{1}{\sigma^2}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ \frac{d\ell}{d\sigma^2} &= -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\big(\sigma^2\big)^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}).\end{aligned}$$

   Setting these partial derivatives equal to zero and solving, the MLEs are

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \big(\mathbf{X}'\mathbf{X}\big)^{-1}\mathbf{X}'\mathbf{y} \\ \widehat{\sigma^2} &= \frac{1}{n}\Big(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}\Big)'\Big(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}\Big) = \frac{SSE}{n}.\end{aligned}$$

2. Check whether or not the MLE of σ^2 is unbiased. If not, modify the MLE to get an unbiased estimator of σ^2.

Solution: We will show the MLE of σ^2 is biased. First, notice that

$$SSE = \left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right)' \left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right) = \mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = \mathbf{y}'\left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{y}.$$

Therefore, by a known result, we have that

$$\begin{aligned} E[SSE] &= \mathrm{tr}\left[\left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\sigma^2\mathbf{I}\right] + \boldsymbol{\beta}'\mathbf{X}'\left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{X}\boldsymbol{\beta} \\ &= \sigma^2\mathrm{tr}(\mathbf{I}) - \sigma^2\mathrm{tr}\left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right) + 0 \\ &= \sigma^2 \cdot n - \sigma^2 \cdot \mathrm{rank}(\mathbf{X}) \qquad (1) \\ &= \sigma^2 \cdot n - \sigma^2 \cdot (k+1) \\ &= \sigma^2(n-k-1), \end{aligned}$$

where (1) follows since $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is a projection matrix onto $C(\mathbf{X})$. Thus, we have

$$E\left[\widehat{\sigma^2}\right] = E\left[\frac{SSE}{n}\right] = \frac{1}{n}E[SSE] = \frac{n-k-1}{n}\sigma^2$$

and hence the MLE is biased. Notice that if we change the MLE to

$$\widehat{\sigma^2} = \frac{1}{n-k-1}SSE$$

then it will be an unbiased estimator of σ^2.

3. Derive the distribution of SSE/σ^2.

Solution: Notice that $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$. Also, let

$$\mathbf{A} = \left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)$$

and so from the first argument in part 2., we have

$$\frac{SSE}{\sigma^2} = \frac{1}{\sigma^2}\mathbf{y}'\mathbf{A}\mathbf{y}.$$

Next, we show $\mathbf{A}$ is idempotent. Notice that

$$\begin{aligned} \mathbf{A}\mathbf{A} &= \left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right) \\ &= \mathbf{I} - 2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{A}. \end{aligned}$$

Thus, $\mathbf{A}$ is idempontent and has rank $n-k-1$. Lastly, the noncentrality parameter is

$$\frac{1}{2\sigma^2}\boldsymbol{\beta}'\mathbf{X}'\mathbf{A}\mathbf{X}\boldsymbol{\beta} = 0$$

and so the distribution of SSE/σ^2 is

$$\frac{SSE}{\sigma^2} \sim \chi^2_{n-k-1}.$$

4. Partition the design matrix $\mathbf{X}$ as $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, where $\mathbf{X}_2 \in \mathbb{R}^{n\times h}, h \geq 1$. Keeping conformity, partition $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)$, as well. Let SSE_1 denote the residual of squares obtained by fitting $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1^\star + \boldsymbol{\epsilon}$.

   (a) Show that $SSE_1 \geq SSE$.

   Solution: First notice that we write

$$SSE = \mathbf{y}'(\mathbf{I} - \mathbf{P}_{C(\mathbf{X})})\mathbf{y}$$
$$SSE_1 = \mathbf{y}'(\mathbf{I} - \mathbf{P}_{C(\mathbf{X}_1)})\mathbf{y}.$$

   Recall in class that we showed

$$\mathbf{P}_{C(\mathbf{X})} = \mathbf{P}_{C(\mathbf{X}_1)} + \mathbf{P}_{C(\mathbf{X})\cap C^\perp(\mathbf{X}_1)}$$

   and from here it follows that

$$\mathbf{P}_{C(\mathbf{X})} - \mathbf{P}_{C(\mathbf{X}_1)} = \mathbf{P}_{C(\mathbf{X})\cap C^\perp(\mathbf{X}_1)}.$$

   Using the property that all projection matrices are positive semidefinite, we have

$$SSE_1 - SSE = \mathbf{y}'(\mathbf{P}_{C(\mathbf{X})} - \mathbf{P}_{C(\mathbf{X}_1)})\mathbf{y} = \mathbf{y}'\mathbf{P}_{C(\mathbf{X})\cap C^\perp(\mathbf{X}_1)}\mathbf{y} \geq 0.$$

   Therefore, we see that

$$SSE_1 \geq SSE.$$

   (b) Show that $(SSE_1 - SSE)/\sigma^2$ has a χ^2 distribution. Find the degrees of freedom and the noncentrality parameter.

   Solution: By part (a), we see that

$$\frac{SSE_1 - SSE}{\sigma^2} = \frac{1}{\sigma^2}\mathbf{y}'\mathbf{P}_{C(\mathbf{X})\cap C^\perp(\mathbf{X}_1)}\mathbf{y} \sim \chi^2_d(p)$$

   since $\mathbf{P}_{C(\mathbf{X})\cap C^\perp(\mathbf{X}_1)}$ is a projection matrix and is therefore idempotent. The degrees of freedom d is given by

$$d = \text{rank}(\mathbf{P}_{C(\mathbf{X})\cap C^\perp(\mathbf{X}_1)}) = \text{tr}\big(\mathbf{P}_{C(\mathbf{X})}\big) - \text{tr}\big(\mathbf{P}_{C(\mathbf{X}_1)}\big)$$
$$= \text{rank}(\mathbf{X}) - \text{rank}(\mathbf{X}_1) = h.$$

   The noncentrality parameter p is

$$p = \frac{1}{2\sigma^2}\boldsymbol{\beta}'\mathbf{X}'\mathbf{P}_{C(\mathbf{X})\cap C^\perp(\mathbf{X}_1)}\mathbf{X}\boldsymbol{\beta}$$
$$= \frac{1}{2\sigma^2}\boldsymbol{\beta}_2'\mathbf{X}_2'\mathbf{X}_2\boldsymbol{\beta}_2.$$

(c) Show that SSE and $SSE_1 - SSE$ are independent.

Solution: Again notice that we can write

$$SSE = \mathbf{y}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} = \mathbf{y}'(\mathbf{I} - \mathbf{P}_{C(\mathbf{X})})\mathbf{y}$$
$$SSE_1 - SSE = \mathbf{y}'(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')\mathbf{y} = \mathbf{y}'\mathbf{P}_{C(\mathbf{X})\cap C^{\perp}(\mathbf{X}_1)}\mathbf{y}.$$

Then, it is a known result that since $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, these are independent if and only if

$$(\mathbf{I} - \mathbf{P}_{C(\mathbf{X})}) \cdot \sigma^2\mathbf{I} \cdot \mathbf{P}_{C(\mathbf{X})\cap C^{\perp}(\mathbf{X}_1)} = \mathbf{0}$$
$$\iff \quad (\mathbf{I} - \mathbf{P}_{C(\mathbf{X})})\mathbf{P}_{C(\mathbf{X})\cap C^{\perp}(\mathbf{X}_1)} = \mathbf{0}.$$

This is clear since $C(\mathbf{X}) \cap C^{\perp}(\mathbf{X}_1) \subseteq C(\mathbf{X})$ and $(\mathbf{I} - \mathbf{P}_{C(\mathbf{X})}) = \mathbf{P}_{C^{\perp}(\mathbf{X})}$.

(d) Derive the likelihood ratio test statistic to test the hypothesis $H_0\colon \boldsymbol{\beta}_2 = \mathbf{0}$ vs $H_1\colon \boldsymbol{\beta}_2 \neq \mathbf{0}$.

Solution: First notice that the MLE for σ^2 is a function of $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$. Therefore, under the null hypothesis,

$$\widehat{\sigma^2}_0 = SSE_1/n.$$

Then, the likelihood ratio test statistic for this hypothesis is given by

$$\Lambda(\mathbf{y}) = \frac{L\left(\widehat{\boldsymbol{\beta}_1}, \mathbf{0}, \widehat{\sigma^2}_0 \mid \mathbf{y}\right)}{L\left(\widehat{\boldsymbol{\beta}_1}, \widehat{\boldsymbol{\beta}_2}, \widehat{\sigma^2} \mid \mathbf{y}\right)} = \frac{\left(\widehat{\sigma^2}_0\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\widehat{\sigma^2}_0}(\mathbf{y} - \mathbf{X}_1\widehat{\boldsymbol{\beta}}_1)'(\mathbf{y} - \mathbf{X}_1\widehat{\boldsymbol{\beta}}_1)\right\}}{\left(\widehat{\sigma^2}\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\widehat{\sigma^2}}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})\right\}}$$
$$= \left(\frac{SSE_1}{SSE}\right)^{-\frac{n}{2}} \frac{\exp\left\{-\frac{n}{2}\right\}}{\exp\left\{-\frac{n}{2}\right\}} = \left(\frac{SSE_1}{SSE}\right)^{-\frac{n}{2}}.$$

(e) Explicitly derive the level α rejection region of this likelihood ratio test.

Solution: Notice that we can rewrite the likelihood ratio test as

$$\Lambda(\mathbf{y}) = \left(\frac{SSE_1}{SSE}\right)^{-\frac{n}{2}} = \left(\frac{SSE_1 - SSE}{SSE} + 1\right)^{-\frac{n}{2}}$$
$$= \left(\frac{(SSE_1 - SSE)/h}{SSE/(n-k-1)} \cdot \frac{h}{n-k-1} + 1\right)^{-\frac{n}{2}}$$
$$= \left(\frac{h}{n-k-1} \cdot F_{h,n-k-1} + 1\right)^{-\frac{n}{2}}$$

where we get the central $F_{h,n-k-1}$ statistic using the previous parts and the fact that under the null, the centrality parameter p in part (b) is 0. Then, we notice that the likelihood ratio test statistic is monotone decreasing in the F statistic. As a consequence, high values of $F_{h,n-k-1}$ favors the full model, and reduces the value of $\Lambda(\mathbf{y})$, which also favors the full model. Therefore, they are equivalent tests and so the level α rejection region is given by $F_{1-\alpha,h,n-k-1}$ since everything else is just constant.