STAT 510 Homework 2

Yifan Zhu

January 25, 2017

1. $\boldsymbol{z} \in \mathcal{C}(\boldsymbol{X}) \Rightarrow \boldsymbol{z} = \boldsymbol{X}\boldsymbol{b}$ for some $\boldsymbol{b}$. Hence

$$
\begin{aligned}
&(\boldsymbol{y} - \boldsymbol{P_X}\boldsymbol{y})^T(\boldsymbol{P_X}\boldsymbol{y} - \boldsymbol{z}) \\
=&(\boldsymbol{y} - \boldsymbol{P_X}\boldsymbol{y})^T(\boldsymbol{P_X}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) \\
=&\boldsymbol{y}^T(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{P_X}\boldsymbol{y} - \boldsymbol{y}^T(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{X}\boldsymbol{b} \\
=&\boldsymbol{y}^T(\boldsymbol{P_X} - \boldsymbol{P_X})\boldsymbol{y} - \boldsymbol{y}^T(\boldsymbol{X} - \boldsymbol{X})\boldsymbol{b} \\
=&\boldsymbol{0}
\end{aligned}
$$

We also have $\boldsymbol{z} \neq \boldsymbol{P_X}\boldsymbol{y} \Rightarrow \boldsymbol{P_X}\boldsymbol{y} - \boldsymbol{z} \neq 0$. Thus we have

$$\|\boldsymbol{y} - \boldsymbol{z}\|^2 = \|\boldsymbol{y} - \boldsymbol{P_X}\boldsymbol{y} + \boldsymbol{P_X}\boldsymbol{y} - \boldsymbol{z}\|^2 > \|\boldsymbol{y} - \boldsymbol{P_X}\boldsymbol{y}\|^2$$

2. For projection matrix $\boldsymbol{P_X}$ we have $\boldsymbol{P_X}\boldsymbol{X} = \boldsymbol{X}$. Let $\boldsymbol{X} = \begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_p\end{bmatrix} = [x_{ij}]_{n\times p}$ and $\boldsymbol{P_X} = \begin{bmatrix}\boldsymbol{\epsilon}_1 & \boldsymbol{\epsilon}_2 & \cdots & \boldsymbol{\epsilon}_n\end{bmatrix}$. Hence we have

$$
\begin{bmatrix}\boldsymbol{\epsilon}_1 & \boldsymbol{\epsilon}_2 & \cdots & \boldsymbol{\epsilon}_n\end{bmatrix}
\begin{bmatrix}
x_{11} & x_{12} & \cdots & x_{1n} \\
x_{21} & x_{22} & \cdots & x_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
x_{n1} & x_{n2} & \cdots & x_{nn}
\end{bmatrix}
= \begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_p\end{bmatrix}
$$

Thus

$$\boldsymbol{x}_j = \sum_{i=1}^{n} x_{ij}\boldsymbol{\epsilon}_i \Rightarrow \mathcal{C}(\boldsymbol{X}) \subset \mathcal{C}(\boldsymbol{P_X})$$

We also have

$$\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T = \boldsymbol{P_X}$$

Let $(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T = [a_{ij}]_{p\times n}$, thus

$$
\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_p\end{bmatrix}
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{p1} & a_{p2} & \cdots & a_{pn}
\end{bmatrix}
= \begin{bmatrix}\boldsymbol{\epsilon}_1 & \boldsymbol{\epsilon}_2 & \cdots & \boldsymbol{\epsilon}_n\end{bmatrix}
$$

Thus

$$\boldsymbol{\epsilon}_j = \sum_{i=1}^{p} a_{ij}\boldsymbol{x}_i \Rightarrow \mathcal{C}(\boldsymbol{P_X}) \subset \mathcal{C}(\boldsymbol{X})$$

Hence we have $\mathcal{C}(\boldsymbol{P_X}) = \mathcal{C}(\boldsymbol{X})$.

3.

$$\begin{aligned}&\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{y}\\=&\boldsymbol{X}^T\boldsymbol{P_X}\boldsymbol{y}\\=&\boldsymbol{X}^T\boldsymbol{P_X}^T\boldsymbol{y}\\=&(\boldsymbol{P_X}\boldsymbol{X})^T\boldsymbol{y}\\=&\boldsymbol{X}^T\boldsymbol{y}\end{aligned}$$

Hence $(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{y}$ is a solution of $\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{b}=\boldsymbol{X}^T\boldsymbol{y}$.

4. (a) $\boldsymbol{C}\hat{\boldsymbol{\beta}}=C(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon})=\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{\epsilon}=\boldsymbol{A}\boldsymbol{P_X}\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{A}\boldsymbol{P_X}\boldsymbol{\epsilon}=\boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{A}\boldsymbol{P_X}\boldsymbol{\epsilon}=\boldsymbol{C}\boldsymbol{\beta}+\boldsymbol{A}\boldsymbol{P_X}\boldsymbol{\epsilon}$.
$\boldsymbol{\epsilon}\sim N(\boldsymbol{0},\sigma^2\boldsymbol{I})$, thus $\boldsymbol{C}\hat{\boldsymbol{\beta}}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, where

$$\begin{aligned}&\boldsymbol{\mu}=\boldsymbol{C}\boldsymbol{\beta}\\&\boldsymbol{\Sigma}=\boldsymbol{A}\boldsymbol{P_X}\sigma^2\boldsymbol{I}\boldsymbol{P_X}^T\boldsymbol{A}^T=\sigma^2\boldsymbol{A}\boldsymbol{P_X}\boldsymbol{A}^T=\sigma^2\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{A}^T=\sigma^2\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T\end{aligned}$$

(b) Let $\boldsymbol{G}=(\boldsymbol{X}^T\boldsymbol{X})^-$ be one of the generalized inverse of $\boldsymbol{X}^T\boldsymbol{X}$ and $\boldsymbol{G}^T$ be its transpose. Thus

$$Var(\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{y})=\boldsymbol{C}\boldsymbol{G}\boldsymbol{X}^T\sigma^2\boldsymbol{I}\boldsymbol{X}\boldsymbol{G}^T\boldsymbol{C}^T=\sigma^2\boldsymbol{C}\boldsymbol{G}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{G}^T\boldsymbol{C}^T$$

(c) We know that $\boldsymbol{c}^T\hat{\boldsymbol{\beta}}\sim N(\boldsymbol{c}^T\boldsymbol{\beta},\sigma^2\boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{c})$, thus $\boldsymbol{c}^T\hat{\boldsymbol{\beta}}-d\sim N(\boldsymbol{c}^T\boldsymbol{\beta}-d,\sigma^2\boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{c})\Rightarrow \frac{1}{\sigma\sqrt{\boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{c}}}(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}-d)\sim(\frac{\boldsymbol{c}^T\boldsymbol{\beta}-d}{\sqrt{\sigma^2\boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{c}}},1)$.
Also we know $\hat{\sigma^2}=\frac{\boldsymbol{y}^T(\boldsymbol{I}-\boldsymbol{P}_X)\boldsymbol{y}}{n-r}\Rightarrow\frac{n-r}{\sigma^2}\hat{\sigma^2}\sim\chi^2_{n-r}$.
And from $\boldsymbol{c}^T\hat{\boldsymbol{\beta}}=\boldsymbol{A}\boldsymbol{P_X}\boldsymbol{y}$ and $\hat{\sigma^2}=\frac{\boldsymbol{y}^T(\boldsymbol{I}-\boldsymbol{P}_X)\boldsymbol{y}}{n-r}$ we know $\boldsymbol{A}\boldsymbol{P_X}\sigma^2\boldsymbol{I}(\boldsymbol{I}-\boldsymbol{P_X})/(n-r)=\boldsymbol{0}\Rightarrow(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}-d)\perp\hat{\sigma^2}$.
Hence,

$$\frac{\boldsymbol{c}^T\hat{\boldsymbol{\beta}}-d}{\sqrt{\hat{\sigma^2}\boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{c}}}=\frac{\frac{1}{\sigma\sqrt{\boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{c}}}(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}-d)}{\sqrt{\frac{n-r}{\sigma^2}\hat{\sigma^2}/(n-r)}}\sim t_{n-r}(\delta)$$

where $\delta=\frac{\boldsymbol{c}^T\boldsymbol{\beta}-d}{\sqrt{\sigma^2\boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{c}}}$.

5. (a)

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

(b) $E(\boldsymbol{y})=\begin{bmatrix}\beta_1-\beta_2\\\beta_3-\beta_4\\\beta_2-\beta_5\\\beta_1-\beta_5\end{bmatrix}$ and $\beta_1-\beta_2=\begin{bmatrix}1&0&0&0\end{bmatrix}E(\boldsymbol{y})$, thus $\beta_1-\beta_2$ is estimable.

(c) It is not estimable. Suppose $\beta_1-\beta_3$ is estimable, then there exists $\lambda_1,\cdots,\lambda_4$ such that $\lambda_1(\beta_1-\beta_2)+\lambda_2(\beta_3-\beta_4)+\lambda_3(\beta_2-\beta_5)+\lambda_4(\beta_1-\beta_5)=\beta_1-\beta_3$, which means $(\lambda_1+\lambda_4)\beta_1+(-\lambda_1+\lambda_3)\beta_2+\lambda_2\beta_3+(-\lambda_2)\beta_4+(-\lambda_3-\lambda_4)\beta_5=\beta_1-\beta_3$. The coefficient of β_3 should be -1 and coefficient of β_4 should be 0 $\Rightarrow\lambda_2=-1$ and $-\lambda_2=0$. It is impossible, thus $\beta_1-\beta_3$ is not estimable.

(d)

$$(\boldsymbol{X}^T\boldsymbol{X})^- = ginv(\boldsymbol{X}^T\boldsymbol{X}) = \begin{bmatrix} 2/9 & -1/9 & 0 & 0 & -1/9 \\ -1/9 & 2/9 & 0 & 0 & -1/9 \\ 0 & 0 & 1/4 & -1/4 & 0 \\ 0 & 0 & -1/4 & 1/4 & 0 \\ -1/9 & -1/9 & 0 & 0 & 2/9 \end{bmatrix}$$

(e) $\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}^T\boldsymbol{y} \Rightarrow$

$$\begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 2 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ -1 & -1 & 0 & 0 & 2 \end{bmatrix} \boldsymbol{b} = \begin{bmatrix} y_{12} + y_{15} \\ -y_{12} + y_{25} \\ y_{34} \\ -y_{34} \\ -y_{25} - y_{15} \end{bmatrix}$$

(f)

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{y} = \begin{bmatrix} \frac{y_{12}+y_{15}}{3} \\ \frac{-y_{12}+y_{25}}{3} \\ \frac{y_{34}}{2} \\ \frac{-y_{34}}{2} \\ \frac{-y_{25}-y_{1}5}{3} \end{bmatrix}$$

(g) $OLS(\beta_1 - \beta_5) = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 \end{bmatrix} \hat{\boldsymbol{\beta}} = \frac{y_{12}+2y_{15}+y_{25}}{3}$.

(h) $E(y_{15}) = \beta_1 - \beta_5$, hence y_{15} is an unbiased estimator of $\beta_1 - \beta_5$. By the uniqueness of OLS estimator we know y_{15} is not an OLS estimator of $\beta_1 - \beta_5$.