线性回归第五周作业

宋歌 2015080086 数52

3/27/2018

1. 考虑简单线性模型$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$，假定$\varepsilon_1, \dots, \varepsilon_n$ $iid \sim U(-\theta, \theta)$，$\theta > 0$是未知参数。试研究$\beta_0, \beta_1$的LSE和MLE的渐近性质，同时考虑$\theta$ 的估计性质。

   1) $\theta = 1,\ x_i \equiv 0$

   In this case, $Y_i = \beta_0 + \varepsilon_i \ iid\ \sim U(\beta_0 - 1, \beta_0 + 1)$

   By the definition of LSE, we have

$$\hat{\beta}_{0,LSE} = \arg\min_{\beta_0} \sum_{i=1}^{n} (Y_i - \beta_0)^2 = \arg\min_{\beta_0} ||\mathbf{Y} - \beta_0||^2 = \bar{\mathbf{Y}}$$

   Thus,

$$\hat{\beta}_{0,LSE} = \bar{\mathbf{Y}} \xrightarrow{P} E(Y_i) = \beta_0$$

   By the definition of MLE, we have

$$\begin{aligned}\hat{\beta}_{0,MLE} &= \arg\max_{\beta_0} L(\beta_0|y_1, \dots, y_n) = \arg\max_{\beta_0} \prod_{i=1}^{n} f(y_i|\beta_0) \\ &= \arg\max_{\beta_0} [(\frac{1}{2})^n \mathbf{I}_{(\beta_0 - 1, \infty)}(y_{(1)}) \mathbf{I}_{(-\infty, \beta_0 + 1)}(y_{(n)})] \\ &= \lambda(y_{(n)} - 1) + (1 - \lambda)(y_{(1)} + 1),\ \forall \lambda \in (0, 1)\end{aligned}$$

   For any $0 < \varepsilon < 2$

$$\begin{aligned}P(|y_{(n)} - (\beta_0 + 1)| > \varepsilon) &= P(\beta_0 + 1 - y_{(n)} > \varepsilon) = P(y_{(n)} < \beta_0 + 1 - \varepsilon) \\ &= \prod_{i=1}^{n} P(y_i < \beta_0 + 1 - \varepsilon) = (\frac{2 - \varepsilon}{2})^n \to 0\end{aligned}$$

that is to say, $y_{(n)} \xrightarrow{P} \beta_0 + 1$, in the same way, we obtain $y_{(1)} \xrightarrow{P} \beta_0 - 1$

Thus, by Slutsky Theorem,

$$\hat{\beta}_{0,MLE} \xrightarrow{P} \lambda\beta_0 + (1-\lambda)\beta_0 = \beta_0$$

2) $x_i \equiv 0$

In this case, $Y_i = \beta_0 + \varepsilon_i \ iid \ \sim U(\beta_0 - \theta, \beta_0 + \theta)$

By the conclusion in 1), we still have

$$\hat{\beta}_{0,LSE} = \bar{\mathbf{Y}} \xrightarrow{P} E(Y_i) = \beta_0$$

By the definition of MLE, we have

$$\begin{aligned}
(\hat{\beta}_{0,MLE}, \hat{\theta}_{MLE}) &= \arg\max_{\beta_0,\theta} L(\beta_0, \theta | y_1, \ldots, y_n) = \arg\max_{\beta_0,\theta} \prod_{i=1}^{n} f(y_i|\beta_0, \theta) \\
&= \arg\max_{\beta_0,\theta} [(\frac{1}{2\theta})^n \mathbf{I}_{(\beta_0-\theta,\infty)}(y_{(1)}) \mathbf{I}_{(-\infty,\beta_0+\theta)}(y_{(n)})] \\
&= \arg \max_{\substack{\theta>0,\ \beta_0 \in \\ (y_{(n)}-\theta, y_{(1)}+\theta)}} (\frac{1}{2\theta})^n = (\frac{y_{(n)} + y_{(1)}}{2}, \frac{y_{(n)} - y_{(1)}}{2})
\end{aligned}$$

We have proved that $y_{(n)} \xrightarrow{P} \beta_0 + \theta,\ y_{(1)} \xrightarrow{P} \beta_0 - \theta$. Thus, by Slutsky Theorem,

$$\begin{aligned}
\hat{\beta}_{0,MLE} &= \frac{y_{(n)} + y_{(1)}}{2} \xrightarrow{P} \frac{\beta_0 + \theta + \beta_0 - \theta}{2} = \beta_0 \\
\hat{\theta}_{MLE} &= \frac{y_{(n)} - y_{(1)}}{2} \xrightarrow{P} \frac{\beta_0 + \theta - \beta_0 + \theta}{2} = \theta
\end{aligned}$$

3) (The answer seems to be wrong...)

Consider the general Simple Linear Model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, where $\varepsilon_1, \ldots, \varepsilon_n \ iid \sim U(-\theta, \theta),\ \theta > 0$ is an unknown parameter.

We already know that, when $\sum_{i=1}^{n}(x_i - \bar{\mathbf{X}})^2 \to \infty$

$$\begin{aligned}
\hat{\beta}_{1,LSE} &= \beta_1 + \sum_{i=1}^{n} b_i \varepsilon_i \xrightarrow{P} \beta_1 \\
\hat{\beta}_{0,LSE} &= \bar{\mathbf{Y}} - \hat{\beta}_{1,LSE}\bar{\mathbf{X}} \xrightarrow{P} \beta_0
\end{aligned}$$

Now consider

$$
\begin{aligned}
(\hat{\beta}_{0,MLE}, \hat{\beta}_{1,MLE}, \hat{\theta}_{MLE}) &= \arg\max_{\beta_0,\beta_1,\theta} L(\beta_0, \beta_1, \theta | y_1, \ldots, y_n) \\
&= \arg\max_{\beta_0,\beta_1,\theta} \prod_{i=1}^{n} f(y_i|\beta_0, \beta_1, \theta) \\
&= \arg\max_{\beta_0,\beta_1,\theta} [(\frac{1}{2\theta})^n \mathbf{I}_{(\beta_0+\beta_1 X_i - \theta < y_i < \beta_0 + \beta_1 X_i + \theta)}]
\end{aligned}
$$

When $\beta_1 > 0$, the maximum of the likelihood function can be written as

$$
\begin{aligned}
\max_{\beta_0,\beta_1,\theta} L &= \max_{\beta_0,\beta_1>0,\theta>0} [(\frac{1}{2\theta})^n \mathbf{I}_{(\beta_0+\beta_1 X_{(n)}-\theta,\infty)}(y_{(1)}) \mathbf{I}_{(-\infty,\beta_0+\beta_1 X_{(1)}+\theta)}(y_{(n)})] \\
&= \max_{\substack{\theta>0,\ \beta_1>0 \\ \beta_0 \in (y_{(n)}-\theta-\beta_1 X_{(1)}, y_{(1)}+\theta-\beta_1 X_{(n)})}} (\frac{1}{2\theta})^n \\
&= \max_{\beta_1>0} L\left(\frac{y_{(n)} + y_{(1)} - \beta_1(X_{(n)} + X_{(1)})}{2}, \beta_1, \frac{y_{(n)} - y_{(1)} + \beta_1(X_{(n)} - X_{(1)})}{2}\right)
\end{aligned}
$$

When $\beta_1 < 0$, the maximum of the likelihood function can be written as

$$
\begin{aligned}
\max_{\beta_0,\beta_1,\theta} L &= \max_{\beta_0,\beta_1<0,\theta>0} [(\frac{1}{2\theta})^n \mathbf{I}_{(\beta_0+\beta_1 X_{(1)}-\theta,\infty)}(y_{(1)}) \mathbf{I}_{(-\infty,\beta_0+\beta_1 X_{(n)}+\theta)}(y_{(n)})] \\
&= \max_{\substack{\theta>0,\ \beta_1<0 \\ \beta_0 \in (y_{(n)}-\theta-\beta_1 X_{(n)}, y_{(1)}+\theta-\beta_1 X_{(1)})}} (\frac{1}{2\theta})^n \\
&= \max_{\beta_1<0} L\left(\frac{y_{(n)} + y_{(1)} - \beta_1(X_{(n)} + X_{(1)})}{2}, \beta_1, \frac{y_{(n)} - y_{(1)} - \beta_1(X_{(n)} - X_{(1)})}{2}\right)
\end{aligned}
$$

Thus, for any $\beta_1 \in \mathbb{R}$, we have

$$
(\hat{\beta}_{0,MLE}, \hat{\beta}_{1,MLE}, \hat{\theta}_{MLE}) = \left(\frac{y_{(n)} + y_{(1)} - \beta_1(X_{(n)} + X_{(1)})}{2}, \beta_1, \frac{y_{(n)} - y_{(1)} + |\beta_1|(X_{(n)} - X_{(1)})}{2}\right)
$$

where

$$
\begin{aligned}
\hat{\beta}_{0,MLE} &= \frac{y_{(n)} + y_{(1)} - \beta_1(X_{(n)} + X_{(1)})}{2} \xrightarrow{P} \beta_0 \\
\hat{\theta}_{MLE} &= \frac{y_{(n)} - y_{(1)} + |\beta_1|(X_{(n)} - X_{(1)})}{2} \xrightarrow{P} \frac{\beta_1 + |\beta_1|}{2}(X_{(n)} - X_{(1)}) + \theta
\end{aligned}
$$

2. Let $\mathbf{A} = \mathbf{A}^T$ be symmetric. Prove the following results:

   a) Let $\mathbf{Y} \sim N_n(\mathbf{0}, \mathbf{I})$. Then $\mathbf{Y}^T \mathbf{A} \mathbf{Y} \sim \chi^2_r$ iff $\mathbf{A}$ is idempotent of rank $r = tr(\mathbf{A})$ since then $\mathbf{A}$ is a projection matrix.

By Spectral Decomposition Theorem, there exists an orthogonal matrix $\mathbf{\Gamma}$ such that

$$\mathbf{\Gamma}^T\mathbf{A}\mathbf{\Gamma} = diag(\lambda_1, \ldots, \lambda_n)$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of matrix $\mathbf{A}$.

i) If $\mathbf{A}$ is idempotent of rank $r = tr(\mathbf{A})$, we know that the eigenvalues of $\mathbf{A}$ are either 0 or 1, then

$$\mathbf{\Gamma}^T\mathbf{A}\mathbf{\Gamma} = \begin{bmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{bmatrix}$$

Let $\mathbf{Y} = \mathbf{\Gamma}\mathbf{X}$, then $\mathbf{X} = \mathbf{\Gamma}^T\mathbf{Y} \sim N_n(\mathbf{0}, \mathbf{\Gamma}^T\mathbf{I_n}\mathbf{\Gamma}) = N_n(\mathbf{0}, \mathbf{I}_n)$

Now we have

$$\mathbf{Y}^T\mathbf{A}\mathbf{Y} = \mathbf{X}^T\mathbf{\Gamma}^T\mathbf{A}\mathbf{\Gamma}\mathbf{X} = \sum_{i=1}^{r} X_i^2$$

Since $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{I}_n)$, we know that $X_i\ iid \sim N(0,1)$, thus,

$$\mathbf{Y}^T\mathbf{A}\mathbf{Y} = \sum_{i=1}^{r} X_i^2 \sim \chi_r^2$$

ii) If $\mathbf{Y}^T\mathbf{A}\mathbf{Y} \sim \chi_r^2$, let $\mathbf{Y} = \mathbf{\Gamma}\mathbf{X}$, then

$$\mathbf{Y}^T\mathbf{A}\mathbf{Y} = \mathbf{X}^T\mathbf{\Gamma}^T\mathbf{A}\mathbf{\Gamma}\mathbf{X} = \mathbf{X}^T diag(\lambda_1, \ldots, \lambda_n)\mathbf{X} = \sum_{i=1}^{n} \lambda_i X_i^2 \sim \chi_r^2$$

By the fact that $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{I}_n)$ and $X_i\ iid \sim N(0,1)$, we have

$$\begin{aligned} f_{\mathbf{Y}^T\mathbf{A}\mathbf{Y}}(t) &= E[\exp(it\sum_{i=1}^{n} \lambda_i X_i^2)] = E[\prod_{k=1}^{n} \exp(it\lambda_k X_k^2)] \\ &= \prod_{k=1}^{n} E[\exp(it\lambda_k X_k^2)] = \prod_{k=1}^{n} f_{\lambda_k X_k^2}(t) \\ &= \prod_{k=1}^{n} (1-2it\lambda_k)^{-1/2} = (1-2it)^{-r/2} \end{aligned}$$

that is to say, $\lambda_1 = \cdots = \lambda_r = 1, \lambda_{r+1} = \cdots = \lambda_n = 0$

thus,

$$\mathbf{\Gamma}^T\mathbf{A}\mathbf{\Gamma} = \begin{bmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{A}^2 = \mathbf{\Gamma}^T \begin{bmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{bmatrix} \mathbf{\Gamma}\mathbf{\Gamma}^T \begin{bmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{bmatrix} \mathbf{\Gamma} = \mathbf{A}$$

$tr(\mathbf{A}) = \sum_{k=1}^{n} \lambda_k = r$, so $\mathbf{A}$ is idempotent of rank $r = tr(\mathbf{A})$

b) Let $\mathbf{Y} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$. Then

$$\frac{\mathbf{Y^T A Y}}{\sigma^2} \sim \chi^2_r \; or \; \mathbf{Y^T A Y} \sim \sigma^2 \chi^2_r$$

iff $\mathbf{A}$ is idempotent of rank r.

Let $\mathbf{Z} = \mathbf{Y}/\sigma \sim N_n(\mathbf{0}, \mathbf{I})$. Hence by a)

$$\frac{\mathbf{Y^T A Y}}{\sigma^2} = \mathbf{Z^T A Z} \sim \chi^2_r$$

iff $\mathbf{A}$ is idempotent of rank r.

c) if $\mathbf{Y} \sim N_n(\mathbf{0}, \mathbf{\Sigma})$ where $\mathbf{\Sigma} > 0$, then $\mathbf{Y}^T\mathbf{A}\mathbf{Y} \sim \chi^2_r$ iff $\mathbf{A\Sigma}$ is idempotent of rank $r = rank(\mathbf{A})$.

Since $\mathbf{\Sigma}$ is symmetric and $\mathbf{\Sigma} > 0$, there exists an invertible matrix $\mathbf{P}$ such that $\mathbf{P}^T\mathbf{\Sigma}\mathbf{P} = \mathbf{I}_n$, then $\mathbf{P}^T\mathbf{Y} \sim N_n(\mathbf{0}, \mathbf{I}_n)$

$$\mathbf{Y}^T\mathbf{A}\mathbf{Y} = \mathbf{Y}^T\mathbf{P}\mathbf{P}^{-1}\mathbf{A}\mathbf{P}^{-T}\mathbf{P}^T\mathbf{Y} = (\mathbf{P}^T\mathbf{Y})^T(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}^{-T})(\mathbf{P}^T\mathbf{Y})$$

by a), we know that $\mathbf{Y}^T\mathbf{A}\mathbf{Y} \sim \chi^2_r$ iff $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}^{-T}$ is idempotent of rank $tr(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}^{-T})$ that is

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P}^{-T}\mathbf{P}^{-1}\mathbf{A}\mathbf{P}^{-T} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}^{-T}$$

Since $\mathbf{P}$ is an invertible matrix, we also have

$$\mathbf{A}\mathbf{P}^{-T}\mathbf{P}^{-1}\mathbf{A} = \mathbf{A}$$

Thus,

$$(\mathbf{A\Sigma})^2 = \mathbf{A\Sigma A\Sigma} = \mathbf{A}\mathbf{P}^{-T}\mathbf{P}^{-1}\mathbf{A\Sigma} = \mathbf{A\Sigma}$$

So $\mathbf{A\Sigma}$ is idempotent, and

$$rank(\mathbf{A\Sigma}) = tr(\mathbf{A\Sigma}) = tr(\mathbf{A}\mathbf{P}^{-T}\mathbf{P}^{-1}) = tr(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}^{-T}) = rank(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}^{-T}) = rank(\mathbf{A})$$

3. 教材中习题：11.6 - 11.12

11.6 Write the following quantities as $\mathbf{b^T Y}$ or $\mathbf{Y^T A Y}$ or $\mathbf{AY}$.

Let $\mathbf{H} = \mathbf{X}(\mathbf{X^T X})^{-1}\mathbf{X^T}$, then

a) $\bar{Y} = [1/n, \ldots, 1/n][Y_1, \ldots, Y_n]^T$

b) $\sum_i (Y_i - \hat{Y}_i)^2 = (\mathbf{Y} - \mathbf{HY})^T(\mathbf{Y} - \mathbf{HY}) = \mathbf{Y^T}(\mathbf{I} - \mathbf{H})^{\mathbf{2}}\mathbf{Y}$

c) $\sum_i (\hat{Y}_i)^2 = (\mathbf{HY})^{\mathbf{T}}(\mathbf{HY}) = \mathbf{Y^T H^2 Y}$

d) $\hat{\boldsymbol{\beta}} = (\mathbf{X^T X})^{-1}\mathbf{X^T Y}$

e) $\hat{\mathbf{Y}} = \mathbf{HY}$

11.7 Show that $\mathbf{I} - \mathbf{H} = \mathbf{I} - \mathbf{X}(\mathbf{X^T X})^{-1}\mathbf{X^T}$ is idempotent.

$$\begin{aligned}(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H}) &= [\mathbf{I} - \mathbf{X}(\mathbf{X^T X})^{-1}\mathbf{X^T}][\mathbf{I} - \mathbf{X}(\mathbf{X^T X})^{-1}\mathbf{X^T}] \\ &= \mathbf{I} - \mathbf{2X}(\mathbf{X^T X})^{-1}\mathbf{X^T} + \mathbf{X}(\mathbf{X^T X})^{-1}\mathbf{X^T X}(\mathbf{X^T X})^{-1}\mathbf{X^T} \\ &= \mathbf{I} - \mathbf{X}(\mathbf{X^T X})^{-1}\mathbf{X^T} = \mathbf{I} - \mathbf{H}\end{aligned}$$

11.8 Let $Y \sim N(\mu, \sigma^2)$ so that $E(Y) = \mu$ and $Var(Y) = \sigma^2 = E(Y^2) - [E(Y)]^2$. If $k \geqslant 2$ is an integer, then

$$E(Y^k) = (k-1)\sigma^2 E(Y^{k-2}) + \mu E(Y^{k-1})$$

Let $Z = (Y - \mu)/\sigma \sim N(0, 1)$. Hence $\mu_k = E(Y - \mu)^k = \sigma^k E(Z^k)$. Use this fact and the above recursion relationship $E(Z^k) = (k-1)E(Z^{k-2})$ to find μ_3 and μ_4.

$$\begin{aligned}\mu_3 &= \sigma^3 E(Z^3) = 2\sigma^3 E(Z) = 0 \\ \mu_4 &= \sigma^4 E(Z^4) = 3\sigma^4 E(Z^2) = 3\sigma^4\end{aligned}$$

11.9 Let $\mathbf{A}$ and $\mathbf{B}$ be matrices with the same number of rows. If $\mathbf{C}$ is another matrix such that $\mathbf{A} = \mathbf{BC}$, is it true that $rank(\mathbf{A}) = rank(\mathbf{B})$? Prove of give a counterexample.

Counterexample:

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & 0 \end{bmatrix}$$

where $rank(A) = 1$ and $rank(B) = n$

11.10 Let $\mathbf{x}$ be an $n \times 1$ vector and let $\mathbf{B}$ be an $n \times n$ matrix. Show that $\mathbf{x^T B x} = \mathbf{x^T B^T x}$.

$$\begin{aligned}
\mathbf{x^T B x} &= (x_1, \ldots, x_n) \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= x_1 \sum_{i=1}^{n} x_i b_{i1} + x_2 \sum_{i=1}^{n} x_i b_{i2} + \cdots + x_n \sum_{i=1}^{n} x_i b_{in} \\
&= \sum_{i=1}^{n} b_{ii} x_i^2 + \sum_{i,j} b_{ij} x_i x_j
\end{aligned}$$

$$\begin{aligned}
\mathbf{x^T B^T x} &= (x_1, \ldots, x_n) \begin{bmatrix} b_{11} & \cdots & b_{n1} \\ \vdots & & \vdots \\ b_{1n} & \cdots & b_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= x_1 \sum_{i=1}^{n} x_i b_{1i} + x_2 \sum_{i=1}^{n} x_i b_{2i} + \cdots + x_n \sum_{i=1}^{n} x_i b_{ni} \\
&= \sum_{i=1}^{n} b_{ii} x_i^2 + \sum_{i,j} b_{ji} x_i x_j
\end{aligned}$$

Thus, $\mathbf{x^T B x} = \mathbf{x^T B^T x}$.

11.11 Consider the model $Y_i = \beta_1 + \beta_2 X_{i,2} + \cdots + \beta_p X_{i,p} + e_i = \mathbf{x_i^T} \boldsymbol{\beta} + e_i$. The least squares estimator $\hat{\boldsymbol{\beta}}$ minimizes

$$Q_{OLS}(\boldsymbol{\eta}) = \sum_{i=1}^{n} (Y_i - \mathbf{x_i^T} \boldsymbol{\eta})^2$$

and the weighted least squares estimator minimizes

$$Q_{WLS}(\boldsymbol{\eta}) = \sum_{i=1}^{n} w_i (Y_i - \mathbf{x_i^T} \boldsymbol{\eta})^2$$

where the w_i, Y_i and $\mathbf{x_i}$ are known quantities. Show that

$$\sum_{i=1}^{n} w_i (Y_i - \mathbf{x_i}^T \boldsymbol{\eta})^2 = \sum_{i=1}^{n} (\tilde{Y}_i - \tilde{\mathbf{x}}_i^T \boldsymbol{\eta})^2$$

by identifying $\tilde{Y}_i$, and $\tilde{\mathbf{x}}_i$. (Hence the WLS estimator is obtained from the least squares regression of $\tilde{Y}_i$ on $\tilde{\mathbf{x}}_i$ without an intercept.)

Let $\tilde{Y}_i = \sqrt{w_i}Y_i,\ \tilde{\mathbf{x}}_i = \sqrt{w_i}\mathbf{x}_i$, then we have

$$\sum_{i=1}^{n} w_i(Y_i - \mathbf{x_i}^T\boldsymbol{\eta})^2 = \sum_{i=1}^{n}(\tilde{Y}_i - \tilde{\mathbf{x}}_i^T\boldsymbol{\eta})^2$$

11.12 Suppose that $\mathbf{X}$ is an $n \times p$ matrix but the rank of $\mathbf{X} < p < n$. Then the normal equations $\mathbf{X^TX}\boldsymbol{\beta} = \mathbf{X^TY}$ have infinitely many solutions. Let $\hat{\boldsymbol{\beta}}$ be a solution to the normal equations. So $\mathbf{X^TX}\hat{\boldsymbol{\beta}} = \mathbf{X^TY}$. Let $\mathbf{G} = (\mathbf{X^TX})^-$ be a generalized inverse of $(\mathbf{X^TX})$. Assume that $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ and $Cov(\mathbf{Y}) = \sigma^2\mathbf{I}$. It can be shown that all solutions to the normal equations have the form $\mathbf{b_z}$ given below.

a) Show that $\mathbf{b_z} = \mathbf{GX^TY} + (\mathbf{GX^TX} - \mathbf{I})\mathbf{z}$ is a solution to the normal equations where the $p \times 1$ vector $\mathbf{z}$ is arbitrary.

By the definition of generalized inverse matrix, we have

$$\begin{aligned}(\mathbf{X}^T\mathbf{X})\mathbf{G}(\mathbf{X}^T\mathbf{X}) &= \mathbf{X}^T\mathbf{X}\\ \mathbf{G}(\mathbf{X}^T\mathbf{X})\mathbf{G} &= \mathbf{X}^T\mathbf{X}\end{aligned}$$

Then we obtain

$$\begin{aligned}\mathbf{X}^T\mathbf{X}\mathbf{b_z} &= \mathbf{X}^T\mathbf{X}\Big[\mathbf{G}\mathbf{X}^T\mathbf{Y} + (\mathbf{G}\mathbf{X}^T\mathbf{X} - \mathbf{I})\mathbf{z}\Big]\\ &= \mathbf{X}^T\mathbf{X}\mathbf{G}\mathbf{X}^T\mathbf{Y} + \mathbf{X}^T\mathbf{X}\mathbf{G}\mathbf{X}^T\mathbf{X} - \mathbf{X}^T\mathbf{X}\mathbf{z}\\ &= \mathbf{X}^T\mathbf{X}\mathbf{G}\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}^T\mathbf{X}\mathbf{z} - \mathbf{X}^T\mathbf{X}\mathbf{z}\\ &= \mathbf{X}^T\mathbf{X}\mathbf{G}\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{Y}\end{aligned}$$

Thus, for any $p \times 1$ vector $\mathbf{z}$, $\mathbf{b_z}$ is a solution to the normal equations.

b) Show that $E(\mathbf{b_z}) \neq \boldsymbol{\beta}$

We know that $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, so we obtain

$$\begin{aligned}E(\mathbf{b_z}) &= E(\mathbf{G}\mathbf{X}^T\mathbf{Y}) + E(\mathbf{G}\mathbf{X}^T\mathbf{X} - \mathbf{I})\mathbf{z}\\ &= E(\mathbf{G}\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}}) + (\mathbf{G}\mathbf{X}^T\mathbf{X} - \mathbf{I})\mathbf{z}\\ &= \mathbf{G}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} + (\mathbf{G}\mathbf{X}^T\mathbf{X} - \mathbf{I})\mathbf{z}\end{aligned}$$

If $E(\mathbf{b_z}) = \boldsymbol{\beta}$, then $\mathbf{G}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} + (\mathbf{G}\mathbf{X}^T\mathbf{X} - \mathbf{I})\mathbf{z} = \boldsymbol{\beta}$, that is to say

$$(\mathbf{G}\mathbf{X}^T\mathbf{X} - \mathbf{I})(\boldsymbol{\beta} + \mathbf{z}) = 0, \ \forall \mathbf{z}$$

Thus $\mathbf{G}\mathbf{X}^T\mathbf{X} = \mathbf{I}$, $\mathbf{X}^T\mathbf{X}$ is invertible and $r(\mathbf{X}^T\mathbf{X}) = p$, which contradicts with

$$r(\mathbf{X}^T\mathbf{X}) \leqslant \min\{r(\mathbf{X}), r(\mathbf{X}^T)\} < p$$

c) Show that $Cov(\mathbf{b_z}) = \sigma^2\mathbf{G}\mathbf{X}^T\mathbf{X}\mathbf{G}^T$.

By the above results, we have

$$\begin{aligned} \mathbf{b_z} &= \mathbf{G}\mathbf{X}^T\mathbf{Y} + (\mathbf{G}\mathbf{X}^T\mathbf{X} - \mathbf{I})\mathbf{z} \\ E(\mathbf{b_z}) &= \mathbf{G}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} + (\mathbf{G}\mathbf{X}^T\mathbf{X} - \mathbf{I})\mathbf{z} \end{aligned}$$

So we obtain

$$\begin{aligned} Cov(\mathbf{b_z}) &= E\Big[(\mathbf{b_z} - E\mathbf{b_z})(\mathbf{b_z} - E\mathbf{b_z})^T\Big] \\ &= E\Big[\mathbf{G}\mathbf{X}^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{X}\mathbf{G}^T\Big] \\ &= \mathbf{G}\mathbf{X}^T E\Big[(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T\Big]\mathbf{X}\mathbf{G}^T \\ &= \mathbf{G}\mathbf{X}^T Cov(\mathbf{Y})\mathbf{X}\mathbf{G}^T = \sigma^2\mathbf{G}\mathbf{X}^T\mathbf{X}\mathbf{G}^T \end{aligned}$$

d) Although $\mathbf{G}$ is not unique, the projection matrix $\mathbf{P} = \mathbf{X}\mathbf{G}\mathbf{X}^\mathbf{T}$ onto $C(\mathbf{X})$ is unique. Use this fact to show that $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b_z}$ does not depend on $\mathbf{G}$ or $\mathbf{z}$.

Since $\mathbf{P}\mathbf{X} = \mathbf{X}$, we obtain

$$\begin{aligned} \hat{\mathbf{Y}} = \mathbf{X}\mathbf{b_z} &= \mathbf{X}\Big[\mathbf{G}\mathbf{X}^T\mathbf{Y} + (\mathbf{G}\mathbf{X}^T\mathbf{X} - \mathbf{I})\mathbf{z}\Big] \\ &= \mathbf{X}\mathbf{G}\mathbf{X}^T\mathbf{Y} + \mathbf{X}\mathbf{G}\mathbf{X}^T\mathbf{X}\mathbf{z} - \mathbf{X}\mathbf{z} \\ &= \mathbf{P}\mathbf{Y} + \mathbf{P}\mathbf{X}\mathbf{z} - \mathbf{X}\mathbf{z} = \mathbf{P}\mathbf{Y} \end{aligned}$$

Thus, $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b_z} = \mathbf{P}\mathbf{Y}$ does not depend on $\mathbf{G}$ or $\mathbf{z}$.

e) There are two ways to show that $\mathbf{a}^T\boldsymbol{\beta}$ is an estimable function. Either show that there exists a vector $\mathbf{c}$ such that $E(\mathbf{c}^\mathbf{T}\mathbf{Y}) = \mathbf{a}^T\boldsymbol{\beta}$, or show that $\mathbf{a} \in C(\mathbf{X}^\mathbf{T})$.

Suppose that $\mathbf{a} = \mathbf{X^T w}$ for some fixed vector $\mathbf{w}$. Show that $E(\mathbf{a^T b_z}) = \mathbf{a}^T\boldsymbol{\beta}$

$$
\begin{aligned}
E(\mathbf{a^T b_z}) &= E\Bigg[\mathbf{w}^T\mathbf{X}\Bigg(\mathbf{G}\mathbf{X}^T\mathbf{Y} + (\mathbf{G}\mathbf{X}^T\mathbf{X} - \mathbf{I})\mathbf{z}\Bigg)\Bigg] \\
&= E(\mathbf{w}^T\mathbf{X}\mathbf{G}\mathbf{X}^T\mathbf{Y}) + E(\mathbf{w}^T\mathbf{X}\mathbf{G}\mathbf{X}^T\mathbf{X}\mathbf{z}) - E(\mathbf{w}^T\mathbf{X}\mathbf{z}) \\
&= \mathbf{w}^T\mathbf{X}\mathbf{G}\mathbf{X}^T E(\mathbf{Y}) + \mathbf{w}^T\mathbf{X}\mathbf{G}\mathbf{X}^T\mathbf{X}\mathbf{z} - \mathbf{w}^T\mathbf{X}\mathbf{z} \\
&= \mathbf{w}^T\mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{w}^T\mathbf{P}\mathbf{X}\mathbf{z} - \mathbf{w}^T\mathbf{X}\mathbf{z} \\
&= \mathbf{w}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{a}^T\boldsymbol{\beta}
\end{aligned}
$$

f) Suppose that $\mathbf{a} = \mathbf{X^T w}$ for some fixed vector $\mathbf{w}$. Show that $Var(\mathbf{a^T b_z}) = \sigma^2\mathbf{w}^T\mathbf{P}\mathbf{w}$.

By the above results, we have

$$
\begin{aligned}
&\mathbf{P} = \mathbf{X}\mathbf{G}\mathbf{X}^T,\ \mathbf{P}\mathbf{X} = \mathbf{X} \\
&\mathbf{P}\mathbf{P}^T = \mathbf{X}\mathbf{G}\mathbf{X}^T\mathbf{X}\mathbf{G}^T\mathbf{X} = \mathbf{P}\mathbf{X}\mathbf{G}^T\mathbf{X} = \mathbf{X}\mathbf{G}\mathbf{X}^T = \mathbf{P} \\
&\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b_z} = \mathbf{P}\mathbf{Y}
\end{aligned}
$$

Then we obtain

$$
\begin{aligned}
Var(\mathbf{a^T b_z}) &= Var(\mathbf{w}^T\mathbf{X}\mathbf{b_z}) = Var(\mathbf{w}^T\hat{\mathbf{Y}}) = Var(\mathbf{w}^T\mathbf{P}\mathbf{Y}) \\
&= E(\mathbf{w}^T\mathbf{P}\mathbf{Y})^T(\mathbf{w}^T\mathbf{P}\mathbf{Y}) - \Bigg[E(\mathbf{w}^T\mathbf{P}\mathbf{Y})\Bigg]^T\Bigg[E(\mathbf{w}^T\mathbf{P}\mathbf{Y})\Bigg] \\
&= E(\mathbf{Y}^T\mathbf{P}^T\mathbf{w}\mathbf{w}^T\mathbf{P}\mathbf{Y}) - \Bigg[\mathbf{w}^T\mathbf{P}E(\mathbf{Y})\Bigg]^T\Bigg[\mathbf{w}^T\mathbf{P}E(\mathbf{Y})\Bigg] \\
&= tr\Bigg(\mathbf{P}^T\mathbf{w}\mathbf{w}^T\mathbf{P}Cov(\mathbf{Y})\Bigg) + E^T(\mathbf{Y})\mathbf{P}^T\mathbf{w}\mathbf{w}^T\mathbf{P}E(\mathbf{Y}) - E^T(\mathbf{Y})\mathbf{P}^T\mathbf{w}\mathbf{w}^T\mathbf{P}E(\mathbf{Y}) \\
&= tr\Bigg(\mathbf{P}^T\mathbf{w}\mathbf{w}^T\mathbf{P}\sigma^2\mathbf{I}\Bigg) = \sigma^2 tr(\mathbf{P}^T\mathbf{w}\mathbf{w}^T\mathbf{P}) = \sigma^2 tr(\mathbf{w}^T\mathbf{P}\mathbf{P}^T\mathbf{w}) \\
&= \sigma^2 tr(\mathbf{w}^T\mathbf{P}\mathbf{w}) = \sigma^2\mathbf{w}^T\mathbf{P}\mathbf{w}
\end{aligned}
$$