Biostat 250A HW7

Elvis Cui
Han Cui
Dept. of Biostat
UCLA

Biostat 250A Homework 7 due Wednesday 12/02/20

Read sections 3.7-3.8 from text.

1. Find the maximum likelihood estimates for the px1 parameters in the standard linear model, where X is a nxp matrix with full column rank and the model is given by

$$Y = X\beta + e$$

where $Y \sim N_n(0, \sigma^2 I_n)$. Compute the information matrix and deduce the asymptotic covariance matrix of the maximum likelihood estimates.

2. Refer to Question 1 and use the Spectral Decomposition Theorem, or otherwise, to find the maximum likelihood estimates for the model parameters in Question 1 if the covariance matrix of Y is now a positive definite matrix V and not $\sigma^2 I_n$.

3. Suppose $Y = X\beta + z\delta + e$ where X is nxp of rank p and z is a nx1 vector. Assume rank $(X \mid z) = p+1$ and the covariance of the error vector e is $\sigma^2 I$.
   a. What is the least squares estimate for δ and what is its variance?
   b. Provide an unbiased estimate for σ^2.
   c. Suppose n = 5 and p = 2, and the first predicted value of Y is

$$\hat{Y}_1 = -.3Y_1 + .2Y_2 + .4Y_3 - 0.1Y_4 + .2Y_5$$

   Is it possible that this model has an intercept term? Justify.
   d. Show that all off-diagonal entries in an orthogonal projection matrix are between -0.5 and 0.5.

4-5. Miscellaneous Ex 3 #1, #11.

6. Suppose under the standard linear model E(y) = Xb we have X is a 10 by 3 matrix with orthonormal columns, $X^T y = (1,2,3)^T$ and $y^T y = 20$. Find the least squares estimate $\hat{b}$ of b $= (b_1, b_2, b_3)$ subject to $b_1 + b_2 + b_3 = 2$. Find the covariance matrix for the estimator $\hat{b}$, $\text{cov}(\hat{b})$ and determine whether $\text{cov}(\hat{b})-\text{cov}(b^*)$ is positive definite, where b* is the unconstrained least squares estimator of b.

7-9. Ex 3g #2, #3, #5

Problem 1:

1. Find the maximum likelihood estimates for the px1 parameters in the standard linear model, where X is a nxp matrix with full column rank and the model is given by

$$Y = X\beta + e$$

where $Y \sim N_n(0, \sigma^2 I_n)$. Compute the information matrix and deduce the asymptotic covariance matrix of the maximum likelihood estimates.

Sol. $e \sim \mathcal{N}_n(0, \sigma^2 I_n) \Rightarrow$

$$L(\beta,\sigma^2) = f(y|\beta,\sigma^2) = \frac{1}{(2\pi)^{\frac{p}{2}}\sigma^n} \exp\left(-\frac{1}{2\sigma^2}\|Y-X\beta\|_2^2\right)$$

$$\Rightarrow \underbrace{\log L(\beta,\sigma^2)}_{\ell} = -\frac{p}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\|Y-X\beta\|_2^2$$

$$\Rightarrow \frac{\partial \ell}{\partial \beta} = +\frac{1}{\sigma^2}X^T(Y-X\beta) = 0$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{1}{2}\frac{n}{\sigma^2} + \frac{1}{2\sigma^4}\|Y-X\beta\|_2^2 = 0$$

$$\Rightarrow \begin{cases} \hat{\beta}_{MLE} = (X^TX)^{-1}X^TY & \text{since } X \text{ has full col. rank} \\ \hat{\sigma}^2_{MLE} = \frac{1}{n}\|Y - X\hat{\beta}_{MLE}\|_2^2 \end{cases}$$

This is the maxima since MVN is in the exponential family and its cumulant function is concave.

Next, $$\frac{\partial^2 \ell}{\partial\beta\partial\beta^T} = -\frac{1}{\sigma^2}X^TX$$

$$\frac{\partial^2 \ell}{\partial\beta\partial\sigma^2} = \frac{\partial^2 \ell}{\partial\sigma^2\partial\beta} = \frac{1}{\sigma^4}X^T(X\beta - Y)$$

$$\frac{\partial^2 \ell}{\partial\sigma^2\partial\sigma^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\|Y-X\beta\|_2^2$$

Thus, the info matrix is

$$I_n(\beta,\sigma^2) = E\left(-\begin{pmatrix} \frac{\partial^2 \ell}{\partial\beta\partial\beta^T} & \frac{\partial^2 \ell}{\partial\beta\partial\sigma^2} \\ \frac{\partial^2 \ell}{\partial\sigma^2\partial\beta^T} & \frac{\partial^2 \ell}{(\partial\sigma^2)^2} \end{pmatrix}\right)$$

$$= E\begin{bmatrix} \frac{1}{\sigma^2}X^TX & \frac{1}{\sigma^4}(Y-X\beta)^TX \\ \frac{1}{\sigma^4}X^T(Y-X\beta) & -\frac{\partial^2 \ell}{\partial\sigma^2\partial\sigma^2} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sigma^2}X^TX & 0^T \\ 0 & \frac{n}{\sigma^4} \end{bmatrix}$$

MLE in this case is AN. Thus,

$$\sqrt{n}\left[\begin{pmatrix} \hat{\beta}_{MLE} \\ \hat{\sigma}^2_{MLE} \end{pmatrix} - \begin{pmatrix} \beta \\ \sigma^2 \end{pmatrix}\right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, I(\beta,\sigma^2)^{-1})$$

where $I(\beta,\sigma^2) =$

$$\begin{bmatrix} \lim_{n\to\infty}\frac{1}{\sigma^2}X_n^TX_n & 0^T \\ 0 & \frac{1}{\sigma^4} \end{bmatrix}$$

provided the limit $\frac{1}{\sigma^2}\lim_n X_n^TX_n$ exists.

where $X_n = [1 x_1 \cdots x_p]$ is a nxp matrix.

& $I(\beta,\sigma^2)^{-1} = \begin{bmatrix} \sigma^2 & 0^T \\ 0 & \sigma^4 \end{bmatrix}$ if X_n is orthogonal.

Problem 2:

2. Refer to Question 1 and use the Spectral Decomposition Theorem, or otherwise, to find the maximum likelihood estimates for the model parameters in Question 1 if the covariance matrix of Y is now a positive definite matrix V and not $\sigma^2 I_n$.

Sol. Let the SD of V be:

$$V = TDT^T,\ TT^T = I.\ D \text{ diagonal.}$$

Thus, Let $V^{\frac{1}{2}} = TD^{\frac{1}{2}}T^T$ & $V = V^{\frac{1}{2}}V^{\frac{1}{2}}$

Now Let $Z = V^{-\frac{1}{2}}Y$, then

$$Z \sim \mathcal{N}(V^{-\frac{1}{2}}X\beta, I_n).$$

Thus, by the invariance principle of MLE, we have:

$$\hat{\beta}_{MLE} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T Z$$

where $\tilde{X} = V^{-\frac{1}{2}}X$.

Or, equivalently,

$$\hat{\beta}_{MLE} = (X^TV^{-1}X)^{-1}X^TV^{-1}y.$$

Another Version of Problem 2:

Find MLE of μ & Σ if $Y_1, \cdots, Y_n \sim \mathcal{N}_p(\mu, \Sigma)$.

Sol.

$$L(Y|\mu,\Sigma) = \prod_{i=1}^{n} |\Sigma|^{-\frac{1}{2}}(2\pi)^{-\frac{p}{2}} e^{-\frac{1}{2}(Y_i-\mu)^T\Sigma^{-1}(Y_i-\mu)}$$

$$\ell(y|\mu,\Sigma) = \ln L(Y|\mu,\Sigma)$$

$$= C - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(Y_i-\mu)^T\Sigma^{-1}(Y_i-\mu)$$

$$= C - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(Y_i-\bar{Y})^T\Sigma^{-1}(Y_i-\bar{Y}) - \frac{n}{2}(\bar{Y}-\mu)^T\Sigma^{-1}(\bar{Y}-\mu) \quad (\Delta)$$

where $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

But $(\Delta) \le C - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(Y_i-\bar{Y})^T\Sigma^{-1}(Y_i-\bar{Y})$

with "=" iff $\bar{Y} = \mu$.

Thus, $\hat{\mu}_{MLE} = \bar{Y}$.

Define $A = \frac{1}{n}\left(\sum_{i=1}^{n}(Y_i-\bar{Y})(Y_i-\bar{Y})^T\right)$

Then $\ell(y|\hat{\mu},\Sigma) =$

$$C - \frac{n}{2}\ln|\Sigma| - \frac{n}{2}Tr(A\Sigma^{-1})$$

$$= C - \frac{n}{2}\ln|\Sigma| - \frac{n}{2}Tr(A^{\frac{1}{2}}\Sigma^{-1}A^{\frac{1}{2}})$$

Since A is positive definite a.s.

$$= \frac{n}{2}\ln|A^{\frac{1}{2}}\Sigma^{-1}A^{\frac{1}{2}}| - \frac{n}{2}Tr(A^{\frac{1}{2}}\Sigma^{-1}A^{\frac{1}{2}}) + C$$

The problem becomes $\max\left[\ln|TDT^T| - Tr\, TDT^T\right]$ where $A^{\frac{1}{2}}\Sigma^{-1}A^{\frac{1}{2}} = TDT^T$ is the spectral decomp.

But this is equivalent to $\max(\ln|D| - Tr\, D)$ or $\max\left(\sum_{i=1}^{p}\ln\lambda_i - \sum_{i=1}^{p}\lambda_i\right)$. $(*)$

λ_i's are eig-vals of $A^{\frac{1}{2}}\Sigma^{-1}A^{\frac{1}{2}}$.

$$\frac{\partial(*)}{\partial\lambda_i} = \frac{1}{\lambda_i} - 1 = 0,\ \frac{\partial^2(*)}{\partial\lambda_i^2} = -\frac{1}{\lambda_i^2} < 0 \Rightarrow \lambda_i = 1,\ i=1,\cdots,p. \Rightarrow \left(\widehat{A^{\frac{1}{2}}\Sigma^{-1}A^{\frac{1}{2}}}\right)_{MLE} = TT^T = I$$

Or, $\hat{\Sigma}_{MLE} = A = \frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})(Y_i-\bar{Y})^T$. $\square$.

Problem 3:

3. Suppose $Y = X\beta + z\delta + e$ where X is nxp of rank p and z is a nx1 vector.
   Assume rank $(X \mid z) = p+1$ and the covariance of the error vector e is $\sigma^2 I$.
   a. What is the least squares estimate for δ and what is its variance?
   b. Provide an unbiased estimate for σ^2.
   c. Suppose n = 5 and p = 2, and the first predicted value of Y is

   $$\hat{Y}_1 = -.3Y_1 + .2Y_2 + .4Y_3 - 0.1Y_4 + .2Y_5$$

   Is it possible that this model has an intercept term? Justify.
   d. Show that all off-diagonal entries in an orthogonal projection matrix are between -0.5 and 0.5.

Sol. **(a):** define

$P_X = X(X^TX)^{-1}X^T$ the projection operator.

Then $Y = X\beta + P_X z\delta + (I-P_X)z\delta + e$

$$= X\alpha + Rz\delta + e.$$

where $\alpha = \beta + (X^TX)^{-1}X^Tz\delta$

$R = I - P_X$

Then let $W = (X, Rz)$, $\gamma = \begin{pmatrix}\alpha\\ \delta\end{pmatrix}$

Note $W^TW = \begin{pmatrix}X^T\\ (Rz)^T\end{pmatrix}(X \;\; Rz)$

$$= \begin{pmatrix}X^TX & 0\\ 0 & z^TRz\end{pmatrix}$$

Since $R^T = R = R^2$ & $XR = 0$.

$$\Rightarrow (W^TW)^{-1} = \begin{pmatrix}(X^TX)^{-1} & 0\\ 0 & (z^TRz)^{-1}\end{pmatrix}$$

Since z^TRz is invertible.

- $\hat{\gamma}_{LSE} = (W^TW)^{-1}W^Ty$

$$= \begin{pmatrix}(X^TX)^{-1} & \\ & (z^TRz)^{-1}\end{pmatrix}\begin{pmatrix}X^T\\ z^TR\end{pmatrix}y$$

$$= \begin{pmatrix}(X^TX)^{-1}X^Ty\\ (z^TRz)^{-1}z^TRy\end{pmatrix}$$

Therefore, $\hat{\delta}_{LSE} = (z^TRz)^{-1}z^TRy$

& $Var(\hat{\delta}_{LSE}) = (z^TRz)^{-1}z^TR(\sigma^2 I)Rz(z^TRz)^{-1}$

$$= \sigma^2(z^TRz)^{-1}$$

Since $R = R^T = R^2$.

(b): $\hat{\sigma}^2 = \frac{1}{n-p-1}\|Y - W\hat{\gamma}_{LSE}\|_2^2$

Pf: $W\hat{\gamma}_{LSE} = (X \;\; Rz)\begin{pmatrix}(X^TX)^{-1}X^T\\ (z^TRz)^{-1}z^TR\end{pmatrix}y$

$$= Ay$$

where $A = P_X + Rz(z^TRz)^{-1}z^TR$

Since $C(P_X) \perp C(R)$

& $C(Rz(z^TRz)^{-1}z^TR) \subseteq C(R)$

we have

$$A^2 = P_X^2 + \left(Rz(z^TRz)^{-1}z^TR\right)^2$$

$$= P_X + Rz(z^TRz)^{-1}z^TR = A.$$

Thus, $\|Y - W\hat{\gamma}_{LSE}\|_2^2 = Y^T(I-A)Y$

& $(I-A)^2 = I-A$. Now by fundamental thm,

$$\frac{1}{\sigma^2}Y^T(I-A)Y \sim \chi_r^2(m),$$

$m = (EY)^T(I-A)(EY)$

$$= \gamma^TW^T(I-A)W\gamma = 0$$

$r = Tr(I-A) = Tr(I) - Tr(A)$

$= n - Tr(P_X) - Tr(z^TRz(z^TRz)^{-1})$

$= n-p-1.$ → Next Page

Thus, $\frac{1}{\sigma^2} Y^T (I-A) Y \sim \chi^2_{n-p-1}(0)$

$\Rightarrow \frac{1}{n-p-1} Y^T (I-A) Y \sim \frac{\sigma^2}{n-p-1} \chi^2_{n-p-1}(0)$

But $\mathbb{E}(\chi^2_{n-p-1}(0)) = n-p-1$

$\Rightarrow \mathbb{E}(\hat{\sigma}^2) = \sigma^2$ $\square$.

(c): Recall $\hat{Y} = X\hat{\beta} = P_X Y$, $P_X = X(X^TX)^{-1}X^T$

If the model has an intercept, then we have $1 \in C(X)$, thus:

$$1 = P_X 1.$$

So if $y = 1$, we should have $\hat{Y} = 1$.

But if $Y_1 = \cdots = Y_5 = 1$.

$$-0.3Y_1 + 0.2Y_2 + 0.4Y_3 - 0.1Y_4 + 0.2Y_5 = 0.4$$

So the model does not have an intercept.

(d): Since $P_\Omega^2 = P_\Omega$, we have

$$\Sigma_j P_{ij} P_{ji} = P_{ii} \text{ where}$$

- P_{ij} is the $(ij)^{th}$ element of the projection matrix P_Ω.

So $P_{ii} = \Sigma_j P_{ij}^2 = P_{ii}^2 + \sum_{j \neq i} P_{ij}^2 \geq P_{ii}^2$

Thus, $0 \leq \sum_{j \neq i}^{n} P_{ij}^2 = P_{ii} - P_{ii}^2 \leq \frac{1}{4}$

But $P_{ij}^2 \leq \sum_{j \neq i}^{n} P_{ij}^2 \leq \frac{1}{4} \Rightarrow$

$$P_{ij} \in [-\tfrac{1}{2}, \tfrac{1}{2}]. \quad \square.$$

Problem 4 & 5:

MISCELLANEOUS EXERCISES 3

1. Let $Y_i = a_i\beta_1 + b_i\beta_2 + \varepsilon_i$ $(i = 1, 2, \ldots, n)$, where the a_i, b_i are known and the ε_i are independently and identically distributed as $N(0, \sigma^2)$. Find a necessary and sufficient condition for the least squares estimates of β_1 and β_2 to be independent.

11. Let

$$\begin{aligned} Y_1 &= \theta_1 + \theta_2 + \varepsilon_1, \\ Y_2 &= \theta_1 - 2\theta_2 + \varepsilon_2, \end{aligned}$$

and

$$Y_3 = 2\theta_1 - \theta_2 + \varepsilon_3,$$

where $E[\varepsilon_i] = 0$ $(i = 1, 2, 3)$. Find the least squares estimates of θ_1 and θ_2. If the equations above are augmented to

$$\begin{aligned} Y_1 &= \theta_1 + \theta_2 + \theta_3 + \varepsilon_1, \\ Y_2 &= \theta_1 - 2\theta_2 + \theta_3 + \varepsilon_2, \\ Y_3 &= 2\theta_1 - \theta_2 + \theta_3 + \varepsilon_3, \end{aligned}$$

find the least squares estimate of θ_3.

1:

Sol. Rewrite $Y = a\beta_1 + b\beta_2 + \varepsilon$

$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$, $a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$, $b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$, $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$

Let $X = (a\ b)$, $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$, then

$\hat{\beta}_{LSE} = (X^TX)^{-1}X^Ty$, since $\varepsilon \sim N(0, \sigma^2 I_n)$

We have $\hat{\beta}_{1LSE} \perp\!\!\!\perp \hat{\beta}_{2LSE}$ iff

$(X^TX)^{-1}X^T$ is orthogonal. That is,

$$(X^TX)^{-1}X^TX(X^TX)^{-1} = (X^TX)^{-1} = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix}$$

But $X^TX = \begin{bmatrix} a^Ta & a^Tb \\ b^Ta & b^Tb \end{bmatrix}$, thus,

$$(X^TX)^{-1} = \frac{1}{\|a\|_2^2\|b\|_2^2 - (a^Tb)^2} \begin{bmatrix} \|b\|_2^2 & -a^Tb \\ -b^Ta & \|a\|_2^2 \end{bmatrix}$$

Thus,

$\hat{\beta}_{1LSE} \perp\!\!\!\perp \hat{\beta}_{2LSE}$ iff

$a \perp b$ or $a^Tb = 0$

11: ① Let $X = \begin{bmatrix} 1 & 1 \\ 1 & -2 \\ 2 & -1 \end{bmatrix}$, then

$$\hat{\theta}_{LSE} = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = (X^TX)^{-1}X^T \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$

$$= \frac{1}{3} \begin{bmatrix} Y_1 + Y_3 \\ Y_1 - Y_2 \end{bmatrix}$$

② Let $R = I - P_X = I - X(X^TX)^{-1}X^T$

$$= \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{-1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{-1}{3} & \frac{1}{3} \end{bmatrix}$$

Then by previous problem,

$$\hat{\theta}_3 = (Z^TRZ)^{-1}Z^TRY$$

where $Z = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

$$\Rightarrow \hat{\theta}_3 = Y_1 + Y_2 - Y_3$$

Problem 6:

6. Suppose under the standard linear model E(y) = Xb we have X is a 10 by 3 matrix with orthonormal columns, $X^T y = (1,2,3)^T$ and $y^T y = 20$. Find the least squares estimate $\hat{b}$ of b = (b_1, b_2, b_3) subject to $b_1 + b_2 + b_3 = 2$. Find the covariance matrix for the estimator $\hat{b}$, cov($\hat{b}$) and determine whether cov($\hat{b}$)-cov(b*) is positive definite, where b* is the unconstrained least squares estimator of b.

- $\min_b \|y - Xb\|_2^2$ is the problem.
  s.t. $1^T b = 2$

Also note: $X^T X = I$, $X^T y = \begin{pmatrix}1\\2\\3\end{pmatrix}$, $y^T y = 20$.

First, we have b^*: unconstrained LSE:

$$b^* = (X^T X)^{-1} X^T y = \begin{pmatrix}1\\2\\3\end{pmatrix}.$$

Next, using the formula in class, we have

$$\hat{b} = b^* - (X^T X)^{-1} 1 (1^T (X^T X)^{-1} 1)^{-1} (1^T b^* - 2)$$

here $A = 1^T$ & $C = 2$.

$$= b^* - \frac{4}{3} 1 = \begin{pmatrix}1\\2\\3\end{pmatrix} - \begin{pmatrix}\frac{4}{3}\\ 4/3 \\ 4/3\end{pmatrix}$$

$$= \begin{pmatrix}\frac{1}{3}\\ 2/3 \\ 5/3\end{pmatrix}$$

In general,

$$\hat{b} = b^* - \frac{1}{3} 1 (1^T b^* - 2)$$

$$= (I - \frac{1}{3} 1 1^T) X^T y + \frac{2}{3} 1$$

Thus, $\mathrm{Cov}(\hat{b}) = (I - \frac{1}{3} 1 1^T) X^T (6^2 I) X (I - \frac{1}{3} 1 1^T)$

$$= 6^2 (I - \frac{1}{3} 1 1^T)^2$$

$$= 6^2 (I - \frac{1}{3} 1 1^T)$$

Also, $\mathrm{Cov}(b^*) = 6^2 (X^T X)^{-1} = 6^2 I$

Thus,

$\mathrm{Cov}(\hat{b}) - \mathrm{Cov}(b^*)$

$= -\frac{6^2}{3} 1 1^T < 0$

is negative definite.

Problem 7-9:

2. By considering the identity $\mathbf{Y} - \hat{\mathbf{Y}}_H = \mathbf{Y} - \hat{\mathbf{Y}} + \hat{\mathbf{Y}} - \hat{\mathbf{Y}}_H$, prove that
$$\|\mathbf{Y} - \hat{\mathbf{Y}}_H\|^2 = \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2 + \|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_H\|^2.$$

3. Prove that
$$\mathrm{Var}[\hat{\beta}_H] = \sigma^2 \left\{ (\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}' \left[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}'\right]^{-1} \mathbf{A}(\mathbf{X}'\mathbf{X})^{-1} \right\}.$$
Hence deduce that
$$\mathrm{var}[\hat{\beta}_{Hj}] \leq \mathrm{var}[\hat{\beta}_j],$$
where $\hat{\beta}_{Hj}$ and $\hat{\beta}_j$ are the jth elements of $\hat{\beta}_H$ and $\hat{\beta}$, respectively.

4. Show that
$$\|\mathbf{Y} - \hat{\mathbf{Y}}_H\|^2 - \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2 = \sigma^2 \hat{\boldsymbol{\lambda}}_H' \left(\mathrm{Var}[\hat{\boldsymbol{\lambda}}_H]\right)^{-1} \hat{\boldsymbol{\lambda}}_H.$$

5. If $\mathbf{X}$ is $n \times p$ of rank p and $\mathbf{B}$ is $p \times q$ of rank q, show that $\mathrm{rank}(\mathbf{XB}) = q$.

Sol. **2.** Let $\Omega = C(X)$,
$$W = \mathcal{N}(A(X^TX)^{-1}X^T) \cap \Omega$$
Then $\hat{Y}_H = P_W Y$, $\hat{Y} = P_\Omega Y$

where P_W & P_Ω are the orthogonal projections onto spaces W & Ω.

Thus, $\|Y - \hat{Y}_H\|_2^2$
$$= \|Y - P_\Omega Y + P_\Omega Y - P_W Y\|_2^2$$
$$= \|Y - P_\Omega Y\|_2^2 + \|P_\Omega Y - P_W Y\|_2^2 + 2y^T(I-P_\Omega)^T(P_\Omega - P_W)y \quad (*)$$

But
$$(I-P_\Omega)^T(P_\Omega - P_W)$$
$$= (I-P_\Omega)(P_\Omega - P_W)$$
$$= P_\Omega - P_W - P_\Omega^2 + P_\Omega P_W$$
$$= P_\Omega - P_W - P_\Omega + P_W$$
$= 0$ since $W \subseteq \Omega$, $P_\Omega P_W = P_W$ & $P_\Omega^2 = P_\Omega = P_\Omega^T$.

Thus,
$$(*) = \|Y - P_\Omega Y\|_2^2 + \|P_\Omega Y - P_W Y\|_2^2$$
$$= \|Y - \hat{Y}\|^2 + \|\hat{Y} - \hat{Y}_H\|^2.$$

3. Since
$$\hat{\beta}_H = \hat{\beta} - (X^TX)^{-1}A^T(A(X^TX)^{-1}A^T)^{-1}(A\hat{\beta} - c)$$
we have
$\mathrm{Var}(\hat{\beta}_H) = \mathrm{Var}(M\hat{\beta})$, where
$$M = I - (X^TX)^{-1}A^T(A(X^TX)^{-1}A^T)^{-1}A$$
$$\Rightarrow \mathrm{Var}(\hat{\beta}_H) = \sigma^2 M(X^TX)^{-1}M^T$$
$$= \sigma^2(X^TX)^{-1} - \sigma^2(X^TX)^{-1}A^T(A(X^TX)^{-1}A^T)^{-1}A(X^TX)^{-1}$$
$$\Rightarrow \mathrm{Var}(\hat{\beta}_H) - \mathrm{Var}(\hat{\beta}_{LSE}) =$$
$$-\sigma^2(X^TX)^{-1}A^T(A(X^TX)^{-1}A^T)^{-1}A(X^TX)^{-1}$$
which is negative definite, thus, the diagonal elements are all negative $\Rightarrow$
$$\mathrm{Var}(\hat{\beta}_{Hj}) \leq \mathrm{Var}(\hat{\beta}_j)$$

5. XB is a $n \times q$ matrix.

Thus, $r(XB) = q$ if $(XBy = 0$ iff $y = 0)$.

But $XBy = 0$ implies $By = 0$ since X has full col. rank

But $By = 0$ implies $y = 0$ since B has full col. rank $\square$.