Exercise4

Chutian Chen cc4515; Congcheng Yan cy2550; Mingrui Liu ml4404

4/12/2020

(1)

$Var(\delta_i) = \mathbb{E}\left[\left(\hat{Y}_i - \mathbb{E}\left[Y_i\right]\right)^2\right] - \mathbb{E}\left[\hat{Y}_i - \mathbb{E}\left[Y_i\right]\right]^2$

So $\mathbb{E}\left[\mathrm{RSS}(\hat{\mathbf{Y}}) - \sum_{i=1}^{n} \operatorname{var}\left(\hat{\varepsilon}_i\right) + \sum_{i=1}^{n} \operatorname{var}\left(\delta_i\right)\right]$

$= \sum_{i=1}^{n}(\mathrm{E}[\left(Y_i - \hat{Y}_i\right)]^2 + \mathbb{E}\left[\left(\hat{Y}_i - \mathbb{E}\left[Y_i\right]\right)^2\right] - \mathbb{E}\left[\hat{Y}_i - \mathbb{E}\left[Y_i\right]\right]^2)$

$= \mathbb{E}\left[\sum_{i=1}^{n}\left(\hat{Y}_i - \mathbb{E}\left[Y_i\right]\right)^2\right]$

$(\mathrm{E}[\left(Y_i - \hat{Y}_i\right)] = 0,\ \mathbb{E}\left[\hat{Y}_i - \mathbb{E}\left[Y_i\right]\right] = 0)$

So $\Gamma = \frac{1}{\sigma^2}\mathbb{E}\left[\sum_{i=1}^{n}\left(\hat{Y}_i - \mathbb{E}\left[Y_i\right]\right)^2\right]$

(2)

Let $\hat{\delta} = \hat{\mathbf{Y}} - \mathbb{E}\left[\mathbf{Y}\right]$, $\hat{\epsilon} = \mathbf{Y} - \hat{\mathbf{Y}}$.

Then $\hat{\epsilon} = (\mathbf{I} - \mathbf{S})\mathbf{Y}$, $\hat{\delta} = \mathbf{S}\mathbf{Y} - \mathbb{E}\left[\mathbf{Y}\right]$.

So $\sigma^2\Gamma = \mathbb{E}\left[\mathrm{RSS}(\hat{\mathbf{Y}}) - \sum_{i=1}^{n} \operatorname{var}\left(\hat{\varepsilon}_i\right) + \sum_{i=1}^{n} \operatorname{var}\left(\delta_i\right)\right]$

$= \mathbb{E}\left[\mathrm{RSS}(\hat{\mathbf{Y}}) - tr(\mathrm{Var}(\hat{\epsilon})) + tr(\mathrm{Var}(\hat{\delta}))\right]$

$tr(\mathrm{Var}(\hat{\epsilon})) = tr((\mathbf{I} - \mathbf{S})^T \mathrm{Var}(\mathbf{Y})(\mathbf{I} - \mathbf{S})) = \sigma^2 tr(\mathbf{S}^T\mathbf{S} - \mathbf{S}^T - \mathbf{S} + \mathbf{I})$

$tr(\mathrm{Var}(\hat{\delta})) = tr(\mathrm{Var}(\mathbf{S}\mathbf{Y})) = \sigma^2 tr(\mathbf{S}^T\mathbf{S})$

So $\sigma^2\Gamma = \mathbb{E}\left[\mathrm{RSS}(\hat{\mathbf{Y}}) - \sum_{i=1}^{n} \operatorname{var}\left(\hat{\varepsilon}_i\right) + \sum_{i=1}^{n} \operatorname{var}\left(\delta_i\right)\right]$

$= \mathbb{E}\left[\mathrm{RSS}(\hat{\mathbf{Y}}) + \sigma^2 tr(2\mathbf{S} - \mathbf{I})\right] = \mathrm{RSS}(\hat{\mathbf{Y}}) + 2\sigma^2 \operatorname{tr}(\mathbf{S}) - \sigma^2 n = \sigma^2 C$

So C is an unbiased estimator of Γ

(3)

Because $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{Y}$, $\mathbf{S} = \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top$

$tr(\mathbf{S}) = tr(\mathbf{X}^\top\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}) = tr(\mathbf{I}) = p$

So $C_p = \frac{1}{\sigma^2}\mathrm{RSS}(\hat{\mathbf{Y}}) + 2p - n$

Note that $\mathrm{AIC}(\hat{\boldsymbol{\beta}}) = n\log SS(\hat{\boldsymbol{\beta}}) + 2p$ and here SS is the same as RSS.

So AIC and C_p are both based on RSS and p.

(4)

$\mathrm{P}(\mathrm{AIC}\left(\hat{\boldsymbol{\beta}}_{q+1}\right) < \mathrm{AIC}\left(\hat{\boldsymbol{\beta}}_{q}\right)) = \mathrm{P}(nlog(\frac{SS(\hat{\beta}_{q+1})}{SS(\hat{\beta}_q)}) < -2)$

Because $SS(\hat{\beta}_q) \sim \sigma^2\chi^2_{n-q}$, $SS(\hat{\beta}_{q+1}) \sim \sigma^2\chi^2_{n-q-1}$ and they're independent.

$SS(\hat{\beta}_q) - SS(\hat{\beta}_{q+1}) \sim \sigma^2\chi^2_1$

$\frac{SS(\hat{\beta}_{q+1})}{SS(\hat{\beta}_q)} \sim 1 - F_{1,n-q}/(n-q) = 1 - t^2_{n-q}/(n-q)$

Then $P = P(\frac{SS(\hat{\beta}_{q+1})}{SS(\hat{\beta}_q)} < e^{-2/n}) \approx P(\frac{SS(\hat{\beta}_{q+1})}{SS(\hat{\beta}_q)} < 1)$
$= P(1 - t^2_{n-q}/(n-q) < 1) = P(t^2_{n-q} > 0) = 1$

Then proved.

(5)

Same reason,

$\mathrm{P}(\mathrm{BIC}\left(\hat{\boldsymbol{\beta}}_{q+1}\right) < \mathrm{BIC}\left(\hat{\boldsymbol{\beta}}_{q}\right)) = \mathrm{P}(nlog(\frac{SS(\hat{\beta}_{q+1})}{SS(\hat{\beta}_q)}) < -logn)$
$= P(nlog(1 - t^2_{n-q}/(n-q)) < -logn) \approx P(-nt^2_{n-q}/(n-q) < -logn)$
$= P(t^2_{n-q} > logn) = 0$

Proved.