AMS 578 (Spring, 2004) Kenny Ye

HANDOUT: PROOF OF GAUSS-MARKOV THEOREM

THEOREM (GAUSS-MARKOV). Suppose $\mathbf{y} = \mathbf{X}\beta + \epsilon$, where $E(\epsilon) = \mathbf{0}$ and $Var(\epsilon) = \sigma^2\mathbf{I}$. Then the least square estimate $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ is the Best Linear Unbiased Estimate (BLUE) of β.

PROOF: Consider an arbitrary estimable linear combination of coefficients $\mathbf{c}^T\beta$. Let $\mathbf{a}^T\mathbf{y}$ be an unbiased estimate of $\mathbf{c}^T\beta$. That is,

$$E(\mathbf{a}^T\mathbf{y}) = \mathbf{a}^T\mathbf{X}\beta = \mathbf{c}^T\beta,$$

for all β. This implies that

$$\mathbf{a}^T\mathbf{X} = \mathbf{c}^T.$$

Since $\mathbf{X}^T\mathbf{X}$ is full rank, $\mathbf{c}$ is linear combination of column vectors of $\mathbf{X}^T\mathbf{X}$. Hence,

$$\mathbf{c} = \mathbf{X}^T\mathbf{X}\lambda$$

$$\mathbf{c}^T\hat{\beta} = \lambda^T\mathbf{X}^T\mathbf{X}\hat{\beta} = \lambda^T\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \lambda^T\mathbf{X}^T\mathbf{y}$$

Now, compute the variance of $\mathbf{a}^T\mathbf{y}$.

$$\begin{aligned} Var(\mathbf{a}^T\mathbf{y}) =& Var(\mathbf{a}^T\mathbf{y} - \mathbf{c}^T\hat{\beta} + \mathbf{c}^T\hat{\beta}) \\ =& Var(\mathbf{a}^T\mathbf{y} - \mathbf{c}^T\hat{\beta}) + Var(\mathbf{c}^T\hat{\beta}) + 2Cov(\mathbf{a}^T\mathbf{y} - \mathbf{c}^T\hat{\beta}, \mathbf{c}^T\hat{\beta}). \end{aligned}$$

But

$$\begin{aligned} Cov(\mathbf{a}^T\mathbf{y} - \mathbf{c}^T\hat{\beta}, \mathbf{c}^T\hat{\beta}) =& Cov(\mathbf{a}^T\mathbf{y} - \lambda^T\mathbf{X}^T\mathbf{y}, \lambda^T\mathbf{X}^T\mathbf{y}) \\ =& (\mathbf{a}^T - \lambda^T\mathbf{X}^T)Var(\mathbf{y})(\lambda^T\mathbf{X}^T)^T \\ =& (\mathbf{a}^T - \lambda^T\mathbf{X}^T)\sigma^2\mathbf{I}\mathbf{X}\lambda \\ =& (\mathbf{a}^T\mathbf{X} - \lambda^T\mathbf{X}^T\mathbf{X})\sigma^2\lambda \\ =& (\mathbf{c}^T - \mathbf{c}^T)\sigma^2\lambda = 0 \end{aligned}$$

Therefore,

$$Var(\mathbf{a}^T\mathbf{y}) = Var(\mathbf{a}^T\mathbf{y} - \mathbf{c}^T\hat{\beta}) + Var(\mathbf{c}^T\hat{\beta}).$$

Since variance cannot be negative, we have that

$$Var(\mathbf{a}^T\mathbf{y}) \geq Var(\mathbf{c}^T\hat{\beta}).$$

In other words, $\mathbf{c}^T\hat{\beta}$ has the smallest variance among all linear unbiased estimators of $\mathbf{c}^T\beta$. In addition, $Var(\mathbf{a}^T\mathbf{y}) \geq Var(\mathbf{c}^T\hat{\beta})$ if and only if $Var(\mathbf{a}^T\mathbf{y} - \mathbf{c}^T\hat{\beta}) = 0$, which implies $\mathbf{a}^T\mathbf{y} = \mathbf{c}^T\hat{\beta}$. Therefore, the least square estimate of β is also unique BLUE.