Video 34: Proof of GM theorem

Scribe: Arghya Sarkar
bs2021

August 2022

Introduction

This lecture contains a proof of GM theorem stated earlier.

Theorem 1 (Gauss-Markov theorem)**.** *Assume the setup for linear regression. Assume X^TX is non-singular. Then for all $\vec{\mathbf{c}}$, $\vec{\mathbf{c}}'\hat{\vec{\beta}}$ is the 'unique' best linear unbiased estimator (BLUE) of $\vec{\mathbf{c}}'\vec{\beta}$.*
Here by 'unique' we mean if $\vec{l}'\vec{y}$ is also a BLUE then $\mathbb{P}_{\vec{\beta}}(\vec{l}'\vec{y} = \vec{\mathbf{c}}'\hat{\vec{\beta}}) = 1$ for all $\vec{\beta}$

Proof. We already showed, $\mathbb{E}\hat{\vec{\beta}} = \vec{\beta}$, this implies, $\mathbb{E}[\vec{\mathbf{c}}'\hat{\vec{\beta}}] = \vec{\mathbf{c}}'\vec{\beta}$. So $\vec{\mathbf{c}}'\hat{\vec{\beta}}$ is an unbiased estimator of $\vec{c}'\vec{\beta}$. Suppose $\vec{a}'\vec{y}$ is another unbiased estimator of $\vec{c}'\vec{\beta}$. So

$$\mathbb{E}[\vec{a}'\vec{y}] = \vec{c}'\vec{\beta} \implies \vec{a}'X\vec{\beta} = \vec{c}'\vec{\beta} \qquad \forall\vec{\beta}$$

Which means

$$\vec{a}'X = \vec{c}' \tag{1}$$

We will need this property later.
Now note we can write $\vec{a}'\vec{y}$ as,

$$\vec{a}'\vec{y} = \vec{\mathbf{c}}'\hat{\vec{\beta}} + (\vec{a}'\vec{y} - \vec{\mathbf{c}}'\hat{\vec{\beta}})$$

This implies, $\mathrm{Var}[\vec{a}'\vec{y}] = \mathrm{Var}(\vec{\mathbf{c}}'\hat{\vec{\beta}}) + \mathrm{Var}(\vec{a}'\vec{y} - \vec{\mathbf{c}}'\hat{\vec{\beta}}) + \mathrm{Cov}(\vec{\mathbf{c}}'\hat{\vec{\beta}}, \vec{a}'\vec{y} - \vec{\mathbf{c}}'\hat{\vec{\beta}})$. We will show that $\mathrm{Cov}(\vec{\mathbf{c}}'\hat{\vec{\beta}}, \vec{a}'\vec{y} - \vec{\mathbf{c}}'\hat{\vec{\beta}}) = 0$. If we can do that then we will get $\mathrm{Var}[\vec{a}'\vec{y}] = \mathrm{Var}(\vec{\mathbf{c}}'\hat{\vec{\beta}}) + \mathrm{Var}(\vec{a}'\vec{y} - \vec{\mathbf{c}}'\hat{\vec{\beta}}) \implies \mathrm{Var}[\vec{a}'\vec{y}] \geq \mathrm{Var}(\vec{\mathbf{c}}'\hat{\vec{\beta}})$. Which means $\vec{\mathbf{c}}'\hat{\vec{\beta}}$ has lowest varience among all unbiased linear estimators of $\vec{\mathbf{c}}'\vec{\beta}$, i.e. $\vec{\mathbf{c}}'\hat{\vec{\beta}}$ is a BLUE of $\vec{c}'\vec{\beta}$.
So it is enough to show,

$$\mathrm{Cov}(\vec{\mathbf{c}}'\hat{\vec{\beta}}, \vec{a}'\vec{y} - \vec{\mathbf{c}}'\hat{\vec{\beta}}) = 0 \tag{2}$$

.
Since X^TX is nonsingular. $\hat{\vec{\beta}} = (X^TX)^{-1}X^T\vec{y}$. So

$$\begin{aligned}
\mathrm{Cov}(\vec{\mathbf{c}}'\hat{\vec{\beta}}, \vec{a}'\vec{y} - \vec{\mathbf{c}}'\hat{\vec{\beta}}) &= \mathrm{Cov}(\vec{\mathbf{c}}'(X^TX)^{-1}X^T\vec{y}, (\vec{a}' - \vec{\mathbf{c}}'(X^TX)^{-1}X^T)\vec{y}) \\
&= (\vec{a}' - \vec{\mathbf{c}}'(X^TX)^{-1}X^T)\mathrm{Var}(\vec{y})X(X^TX)^{-1}\vec{\mathbf{c}} \\
&= (\vec{a}' - \vec{\mathbf{c}}'(X^TX)^{-1}X^T)\sigma^2 IX(X^TX)^{-1}\vec{\mathbf{c}} \\
&= \sigma^2((\vec{a}' - \vec{\mathbf{c}}'(X^TX)^{-1}X^T)X(X^TX)^{-1}\vec{\mathbf{c}}) \\
&= \sigma^2(\vec{a}'X(X^TX)^{-1}\vec{\mathbf{c}} - \vec{\mathbf{c}}'(X^TX)^{-1}\vec{\mathbf{c}}) \\
&= \sigma^2(\vec{a}'X - \vec{\mathbf{c}}')(X^TX)^{-1}\vec{\mathbf{c}} = 0
\end{aligned}$$

In the last line we used (1). Now since we proved (2) we conclude $\hat{\vec{\beta}}$ is a BLUE of $\vec{\beta}$.
For 'uniqueness' part, note if $\vec{l}'\vec{y}$ is another BLUE of $\vec{\mathbf{c}}'\vec{\beta} \quad \forall\vec{\beta}$. Then we know, $\mathrm{Var}[\vec{l}'\vec{y}] = \mathrm{Var}(\vec{\mathbf{c}}'\hat{\vec{\beta}}) + \mathrm{Var}(\vec{l}'\vec{y} - \vec{\mathbf{c}}'\hat{\vec{\beta}})$. But $\vec{l}'\vec{y}$ and $\vec{\mathbf{c}}'\hat{\vec{\beta}}$ are both BLUE of $\vec{\mathbf{c}}'\vec{\beta}$ so $\mathrm{Var}[\vec{l}'\vec{y}] = \mathrm{Var}(\vec{\mathbf{c}}'\hat{\vec{\beta}})$. So

$$\mathrm{Var}(\vec{l}'\vec{y} - \vec{\mathbf{c}}'\hat{\vec{\beta}}) = 0 \qquad \forall\vec{\beta}$$

Which implies

$$(\vec{l}'\vec{y} - \vec{\mathbf{c}}'\hat{\vec{\beta}}) = \text{Constant} \qquad \text{with probability 1}$$

But note both $\vec{l}'\vec{y}$ and $\vec{\mathbf{c}}'\hat{\vec{\beta}}$ are linear function of y so the constant need to be zero.
So $\vec{l}'\vec{y} = \vec{\mathbf{c}}'\hat{\vec{\beta}}$ with probability 1.
This completes the proof. □