

Estimation Theory and Inference

Hanchen Wang

Ph.D Candidate in
Engineering, University of Cambridge

January 23, 2020

Proof: OLS is the Unique BLUE for the GLM

- here we are going to show that OLS(Ordinary Least Square) is the unique BLUE(Best¹ Linear Unbiased Estimator) for the GLM(General Linear Model):
 - from the lecture notes, we have known that the variance of OLS estimator is the lower bound of the GLS parameters estimation under a few assumptions of the observation noise²:

$$\text{var}(\hat{\theta}_i) \geq \text{var}(\theta_i^{OLS})$$

where $\hat{\theta}_i$ and θ_i^{OLS} is the i-th estimated GLM parameter by arbitrary estimator and OLS estimator respectively, to prove the unique BLUE of OLS estimator, we first show that OLS estimator is BLUE of GLM

$$\text{var}(\hat{\theta}_i) = \text{var}(\theta_i^{OLS}) + \sigma_e^2 \cdot \text{diag}(\Delta\Delta^T)$$

here $\Delta\Delta^T$ is the product of the full-rank $n \times n$ matrix and its transpose, thus it is positive semi-definite, thus its diagonal elements are ensured to be non-negative, thus OLS is BLUE of GLM

¹in terms of lowest variance unbiased estimation, some biased estimators such as *James-Stein estimator* and *ridge regression* exist with lower variance than OLS, yet they are not BLUE

²more formally, aka *Gauss-Markov assumptions*

Example: Power Spectrum of Sine Wave

- to show the uniqueness, we only need to show that '=' holds only when $\hat{\theta}_i = \theta_i^{OLS} \forall i$, suppose we have another BLUE $\hat{\theta}_i$, thus:

$$\begin{aligned} \text{var}(\hat{\theta}_i) &\geq \text{var}(\theta_i^{OLS}) \text{ and } \text{var}(\hat{\theta}_i) \leq \text{var}(\theta_i^{OLS}) \Rightarrow \text{var}(\hat{\theta}_i) = \text{var}(\theta_i^{OLS}) \\ \Rightarrow \text{diag}(\Delta\Delta^T) &= \mathbf{0} \end{aligned}$$

recall that $\Delta\Delta^T$ are semi-definite and now its trace and determinant are all zero, thus all the eigenvalues are zero. Recall the theorem that 'leading principal submatrix of a positive semidefinite matrix should also be positive semidefinite', and we can show that its 1×1 are zero matrix, then induce that $2 \times 2, 3 \times 3 \dots n \times n$ (itself) are all zero matrices, thus:

$$\Delta = \mathbf{0} \Rightarrow \mathbf{D} = \mathbf{C} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T$$

as desired

Some Examples on the Design Matrix G

- recall the expression of GLM:

$$\mathbf{x} = \mathbf{G}\theta + \mathbf{e}$$

where \mathbf{e} is supposed to have zero mean and constant variance σ_e^2

\mathbf{G} is called design matrix, which depends on the data \mathbf{x} we are trying to generate.

-

TBD

MLE in non-Gaussian cases

- basically to summarize some MLE cases when the noise is not Gaussian:
 - TBD

Differences between OLS, MLE, MAP Estimator

- basically to summarize the differences between OLS, MLE and MAP
 - TBD