Stat 849: Application of the Gauss-Markov Theorem

Sündüz Keleş

Department of Statistics
Department of Biostatistics and Medical Informatics
University of Wisconsin, Madison

September 14, 2008

Problem

A rocket is launched from ground level (altitude $y = 0$) at time $t = 0$. Its upward acceleration is a. The altitude of the rocket is measured by two separate instruments at times $t_1, t_2, \cdots, t_n$. At time t_i, the altitudes $y_{i,1}$ and $y_{i,2}$ measured by the respective instruments satisfy the model

$$y_{i,j} = \frac{1}{2}at_i^2 + e_{i,j}, \quad 1 \leq i \leq n \quad j = 1, 2.$$

The measurement errors $\{e_{i,j} : 1 \leq i \leq n, j = 1, 2\}$ are taken to be independent, identically distributed. Both the acceleration a and the variance σ^2 are unknown.

Candidate estimators for a

- Least squares estimator:

$$\hat{a}_{LSE} = \frac{\sum_{i=1}^{n} t_i^2 (y_{i,1} + y_{i,2})}{\sum_{i=1}^{n} t_i^4}.$$

- Another estimator:

$$\tilde{a} = \sum_{i=1}^{n} (y_{i,1} + y_{i,2}) / \sum_{i=1}^{n} t_i^2.$$

You are allowed to choose between the two estimators. Which one would you use and why?

Properties of the estimators

- Linearity?

Properties of the estimators

- Linearity?

Properties of the estimators

- Linearity? Both estimators are linear.
- Bias:

Properties of the estimators

- Linearity? Both estimators are linear.
- Bias:

Properties of the estimators

- Linearity? Both estimators are linear.
- Bias:
  $\hat{a}_{LSE}$ is unbiased by the LSE theory.

Properties of the estimators

- Linearity? Both estimators are linear.
- Bias:
  $\hat{a}_{LSE}$ is unbiased by the LSE theory.

$$E[\tilde{a}] = \frac{\sum_{i=1}^{n} E(y_{i,1} + y_{i,2})}{\sum_{i=1}^{n} t_i^2} = \frac{\sum_{i=1}^{n}(at_i^2/2 + at_i^2/2)}{\sum_{i=1}^{n} t_i^2}$$

$$= a\frac{\sum_{i=1}^{n} t_i^2}{\sum_{i=1}^{n} t_i^2} = a.$$

Properties of the estimators

- Linearity? Both estimators are linear.
- Bias:
  $\hat{a}_{LSE}$ is unbiased by the LSE theory.

$$E[\tilde{a}] = \frac{\sum_{i=1}^{n} E(y_{i,1} + y_{i,2})}{\sum_{i=1}^{n} t_i^2} = \frac{\sum_{i=1}^{n}(at_i^2/2 + at_i^2/2)}{\sum_{i=1}^{n} t_i^2}$$

$$= a\frac{\sum_{i=1}^{n} t_i^2}{\sum_{i=1}^{n} t_i^2} = a.$$

Properties of the estimators

- Linearity? Both estimators are linear.
- Bias:
  $\hat{a}_{LSE}$ is unbiased by the LSE theory.

$$E[\tilde{a}] = \frac{\sum_{i=1}^{n} E(y_{i,1} + y_{i,2})}{\sum_{i=1}^{n} t_i^2} = \frac{\sum_{i=1}^{n}(at_i^2/2 + at_i^2/2)}{\sum_{i=1}^{n} t_i^2} = a\frac{\sum_{i=1}^{n} t_i^2}{\sum_{i=1}^{n} t_i^2} = a.$$

- Both are linear and biased. By the Gauss-Markov Theorem, the LSE $\hat{a}_{LSE}$ is better (has smaller mean squared error, e.g., variance).

Variances of the estimators

$$
\begin{aligned}
Var(\hat{a}_{LSE}) &= \frac{2\sigma^2}{\sum_{i=1}^{n} t_i^4}, \\
Var(\tilde{a}) &= \frac{2n\sigma^2}{\left(\sum_{i=1}^{n} t_i^2\right)^2}.
\end{aligned}
$$

Variances of the estimators

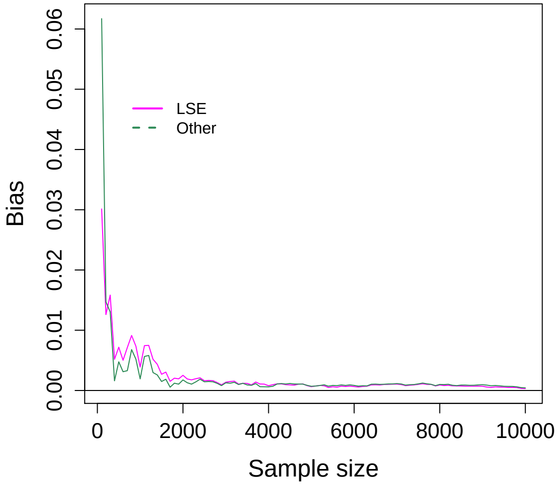
$$
\begin{aligned}
Var(\hat{a}_{LSE}) &= \frac{2\sigma^2}{\sum_{i=1}^{n} t_i^4}, \\
Var(\tilde{a}) &= \frac{2n\sigma^2}{\left(\sum_{i=1}^{n} t_i^2\right)^2}.
\end{aligned}
$$

Can estimate these variances by replacing σ^2 by its least squares estimator s^2.

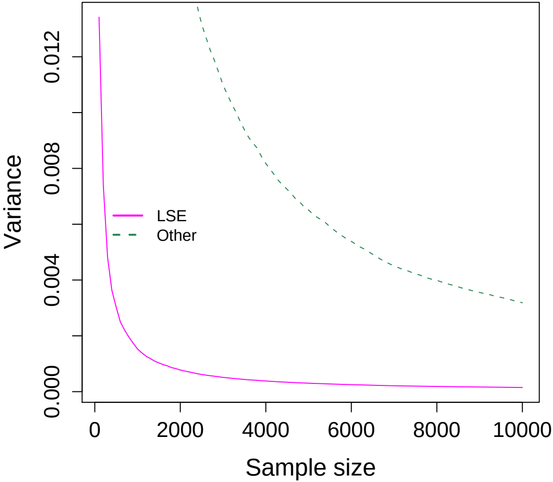
A simulation study to see this in action

We will generate 1000 data points. Then, we will look at the variance and bias of the two estimators as a function of the sample size ($n = \{10, 20, \cdots, 1000\}$).

Results: Bias



Results: Variance



Notes on the R code

- The code to generate these plots are available on the course website.
- The code works with the estimator of $2a$ rather than a.
- $Var(2\tilde{a}) = 4Var(\tilde{a})$.
- Note that for each t_i^2 we have a $y_{i,1}$ and $y_{i,2}$, $i = 1, \cdots, n$. So the effective sample size of the data is $2n$.
- You are encouraged to experiment with the value of a, σ^2, and sample size.

Exercise

- Calculate variances of the both estimators.
- Analytically show that $\hat{a}_{LSE}$ has a smaller variance (*Hint:* Use Cauchy-Schwarz inequality).