Linear Estimation and Least Squares: What's the Connection?

Travis J. Arnold

University of Wisconsin-Madison
Department of Chemical and Biological Engineering
Systems Seminar

December 15, 2017

Introduction

- Linear regression and least squares techniques are commonly employed by engineers for parameter estimation because they are intuitive and often easy to implement.
- However, practitioners and researchers often miss some of the finer points of the theory underlying such methods.

Problem statement

- The context of this talk is the linear model:

$$y = X\theta + e$$
$$E[e] = 0 \quad E[ee'] = R$$

- **Goal:** Estimate the underlying parameter θ from the measurement y.
- Specifically, we would like to find a **minimum variance linear unbiased estimator (MVLUE)**, $\widehat{\theta}$, of θ.

What does MVLUE mean?

- **Linear:**

$$\begin{aligned} \widehat{\theta} &= My \\ E[\widehat{\theta}] &= MX\theta \\ \text{var}(\widehat{\theta}) &= MRM' \end{aligned}$$

- **Unbiased:**

$$\theta = E[\widehat{\theta}] = MX\theta \text{ for all } \theta \implies MX = I$$

- **Minimum variance:**

$$\text{var}(\widehat{\theta}^*) - \text{var}(\widehat{\theta}) \geq 0 \text{ for any linear unbiased estimator } \widehat{\theta}^*$$

Motivation: a problematic statement

- Consider this theorem from Rajamani and Rawlings (2007):

Theorem

For a linear model of the form $y = X\theta + e$ with $E[e] = 0$ and $E[ee'] = R$, the weighted least squares estimator for θ is formulated as

$$\min_{\theta} \; \|X\theta - y\|^2_{R^{-1}}$$

The weighted least squares estimator given by

$$\widehat{\theta} = (X'R^{-1}X)^{-1}X'R^{-1}y$$

then has the minimum variance among all linear unbiased estimators.

- This result isn't exactly wrong, but it is incomplete. To hold as stated it requires X to have full column rank and $R > 0$.

Overview for the rest of the talk

1. Show what the MVLUE is when X has full column rank and $R \geq 0$ (not necessarily strictly positive definite).
   - I will not provide a formal proof, but I will provide the intuition that underlies the result.
2. Demonstrate that the MVLUE is also a solution to a corresponding, well posed least squares problem.
3. Discuss what happens if X does hot have full column rank.

Intuition behind the MVLUE

- The minimum variance criterion is not immediately practicable—we cannot minimize a matrix.
- However, for any matrices A and B,

$$A - B \geq 0 \implies \operatorname{tr} A \geq \operatorname{tr} B$$

- This means that a MVLUE has the following property:

$$\operatorname{tr} \operatorname{var}(\widehat{\theta}^*) \geq \operatorname{tr} \operatorname{var}(\widehat{\theta}) \text{ for any linear unbiased estimator } \widehat{\theta}^*$$

- This motivates us to consider the following constrained optimization problem:

$$\min_M \ \frac{1}{2} \operatorname{tr} MRM'$$
$$\text{s.t. } MX = I$$

Solving the problem

- Form the Lagrangian and take derivatives:

$$L(M,\Lambda) = \tfrac{1}{2}\operatorname{tr} MRM' + \operatorname{tr} \Lambda'(MX - I)$$
$$\frac{dL}{dM} = MR + \Lambda X'$$
$$\frac{dL}{d\Lambda} = MX - I$$

- Setting the derivatives equal to zero gives the first order necessary conditions for a minimum:

$$\begin{bmatrix} R & X \\ X' & 0 \end{bmatrix} \begin{bmatrix} M^{o\prime} \\ \Lambda^{o\prime} \end{bmatrix} = \begin{bmatrix} 0 \\ I \end{bmatrix}$$

Solving the problem

- The block matrix on the left is called a bordered Gramian matrix. This equation has a solution if and only if X has full column rank, in which case the general solution is

$$M^o = (X'WX)^{-1}X'W + Q(I - W^+W)$$

  where $W = (R + XX')^+$ and Q is an arbitrary matrix of appropriate size.
    - The superscript + denotes Moore-Penrose pseudoinverse (Magnus and Neudecker, 1999, Section 2.5)
    - See Magnus and Neudecker (1999, Theorem 3.24) for proof of this result.
- This suggests that

$$\widehat{\theta} = (X'WX)^{-1}X'Wy$$

  is a good candidate for a MVLUE.

The MVLUE

- It turns out our intuition is correct. We give the following theorem without proof:

Theorem

The MVLUE of θ exists and is unique if and only if X is full column rank, in which case

$$\widehat{\theta} = (X'WX)^{-1}X'Wy$$
$$\text{cov}(\widehat{\theta}) = (X'WX)^{-1} - I$$

where $W = (R + XX')^{+}$.

- This is a simplified version of results found in Rao (1973).
- Rao's treatment is clunky and hard to read in my opinion. I wrote up a note for my own reference that gives a fairly expository proof—let me know if you'd like to see it!

X full column rank $\implies$ $X'WX$ is invertible

- Assume X is full column rank.
- W is symmetric, so we can write it's singular value decomposition as

$$W = (R + XX')^+ = USU' = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1' \\ U_2' \end{bmatrix} = U_1 \Sigma U_1'$$

  where U is orthogonal (so $U_1' U_1 = I$) and $\Sigma > 0$.
- Then an SVD of W^+ is

$$W^+ = R{+}XX' = US^+U' = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1' \\ U_2' \end{bmatrix} = U_1 \Sigma^{-1} U_1'$$

X full column rank $\implies X'WX$ is invertible

- U_1 is a basis for $\operatorname{col} W = \operatorname{col} W^+$, and

$$\operatorname{col} W = \operatorname{col} W^+ = \operatorname{col} \begin{bmatrix} R & XX' \end{bmatrix} \supseteq \operatorname{col} XX' = \operatorname{col} X$$

- Therefore we can write $X = U_1 Y$ for some Y. Furthermore, we know Y has full column rank (if it didn't, then X wouldn't have full column rank).
- Finally, this allows us to write

$$X'WX = Y'U_1'U_1\Sigma U_1'U_1 Y = Y'\Sigma Y > 0$$

The least squares problem

- Assume X has full column rank and consider the following least squares problem:

$$\min_{\theta} \ \tfrac{1}{2}\|X\theta - y\|_W^2$$

- Rewrite the objective function:

$$\begin{aligned} f(\theta) &:= \tfrac{1}{2}\|X\theta - y\|_W^2 \\ &= \tfrac{1}{2}(X\theta - y)'W(X\theta - y) \\ &= \tfrac{1}{2}\theta'X'WX\theta - \theta'X'Wy + \tfrac{1}{2}y'Wy \end{aligned}$$

- Sufficient conditions for an absolute minimum:

$$\begin{aligned} \frac{df}{d\theta} &= X'WX\theta - X'Wy = 0 \\ \frac{d^2f}{d\theta^2} &= X'WX > 0 \end{aligned}$$

The least squares problem

- These conditions are satisfied if and only if

$$\theta = \widehat{\theta} = (X'WX)^{-1}X'Wy$$

- This is a well posed optimization problem with a unique solution, and that solution is the MVLUE.
- Adding the XX' term in the weight is the secret sauce that makes it work out nicely.

What if X doesn't have full column rank?

- In this case, it's immediately obvious that the least squares problem does not have a unique solution.
  - For any θ_1, there exists $\theta_2 \neq \theta_1$ such that $X\theta_1 = X\theta_2$.
- It turns out there does not even exist a linear unbiased estimator of θ, let alone a MVLUE! This fact is less obvious however.

Conclusions

- For linear models, the MVLUE exists and is unique if and only if the model matrix X has full column rank.
- We do not require the covariance matrix R to be invertible for this to hold, but if we do not assume that R is invertible then we need to take some additional steps to solve the problem.
- The MVLUE is also the unique solution to a corresponding least squares problem.

References

J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley, New York, 1999.

M. R. Rajamani and J. B. Rawlings. Estimation of the disturbance structure from data using semidefinite programming and optimal weighting. Technical Report 2007–02, TWMCC, June 2007.

C. R. Rao. *Linear Statistical Inference and Its Applications*. John Wiley and Sons, New York, second edition, 1973.

Questions?