

ECE531 Screencast 4.3: Maximum Likelihood Estimation vs. Minimum Variance Unbiased Estimation

D. Richard Brown III

Worcester Polytechnic Institute

Introduction

We've seen examples where the ML estimator and the MVU estimator are the same. In fact, we know that when an efficient estimator exists, we have $\hat{\theta}_{\text{ML}}(y) = \hat{\theta}_{\text{MVU}}(y)$.

What happens when an efficient estimator doesn't exist?

To get some insight into this question, let's consider the following example: Suppose we have random observations given by

$$Y_k \stackrel{\text{i.i.d.}}{\sim} \theta e^{-\theta y_k}$$

for $k = 0, \dots, n - 1$ and $y_k \geq 0$. The unknown parameter $\theta > 0$.

Does an Efficient Estimator Exist?

Since everything is i.i.d., we can write the joint pdf

$$p_Y(y; \theta) = \prod_k p_{Y_k}(y_k; \theta) = \theta^n \exp \left\{ -\theta \sum_k y_k \right\}$$

and we can compute

$$\frac{\partial}{\partial \theta} \ln p_Y(y; \theta) = \frac{\partial}{\partial \theta} \left(n \ln \theta - \theta \sum_k y_k \right) = \frac{n}{\theta} - n\bar{y}$$

where $\bar{y} = \frac{1}{n} \sum_k y_k$. We can compute the Fisher information

$$I(\theta) = -\mathbb{E} \left[\frac{\partial^2}{\partial \theta^2} \ln p_Y(Y; \theta) \right] = \frac{n}{\theta^2}$$

Since we can't write $\frac{\partial}{\partial \theta} \ln p_Y(y; \theta)$ in the form $I(\theta)(g(y) - \theta)$, the CRLB is not attainable and an efficient estimator does not exist in this example.

MVU Estimator

Even though an efficient estimator doesn't exist, it is possible in this problem to find an MVU estimator via the RBLS theorem. Since $p_Y(y; \theta)$ is a one-parameter exponential family, it isn't too hard to show

$$\hat{\theta}_{\text{MVU}}(y) = \left(\frac{1}{n-1} \sum_{k=0}^{n-1} y_k \right)^{-1} = \frac{n-1}{n\bar{y}}.$$

Assuming $n \geq 3$, the variance of the MVU estimator is then

$$\text{var} \left\{ \hat{\theta}_{\text{MVU}}(Y) \right\} = \frac{\theta^2}{n-2} > \frac{\theta^2}{n} = I^{-1}(\theta).$$

Maximum Likelihood Estimator

Let's set up the likelihood equation $\frac{\partial}{\partial \theta} \ln p_Y(y; \theta) = 0 \dots$

We can use our previous result to write

$$\frac{\partial}{\partial \theta} \ln p_Y(y; \theta) = \frac{n}{\theta} - n\bar{y} = 0$$

which has the unique solution

$$\hat{\theta}_{\text{ML}}(y) = \frac{1}{\bar{y}}.$$

You can easily confirm that this solution is a maximum of $p_Y(y; \theta)$ since $\frac{\partial^2}{\partial \theta^2} \ln p_Y(y; \theta) = \frac{-n}{\theta^2} < 0$.

Note $\hat{\theta}_{\text{ML}}(y) \neq \hat{\theta}_{\text{MVU}}(y)$. Which is better, $\hat{\theta}_{\text{ML}}(y)$ or $\hat{\theta}_{\text{MVU}}(y)$?

Performance of the Maximum Likelihood Estimator

Assuming $n \geq 2$, the mean of the ML estimator can be computed as

$$\mathbb{E} \left\{ \hat{\theta}_{\text{ML}}(Y) \right\} = \frac{n}{n-1} \theta$$

Note that $\hat{\theta}_{\text{ML}}(y)$ is **biased**. Unlike MVU estimators, MLEs can be biased.

Assuming $n \geq 3$, the variance of the ML estimator can be computed as

$$\text{var} \left\{ \hat{\theta}_{\text{ML}}(Y) \right\} = \frac{n^2 \theta^2}{(n-1)^2 (n-2)} > \frac{\theta^2}{n} = I^{-1}(\theta)$$

where $I(\theta)$ is the Fisher information.

Caution: The CRLB is for unbiased estimators. So this inequality is a bit bogus. It is actually possible for a biased estimator to have variance smaller than $I^{-1}(\theta)$.

MVU vs. MLE

A fair comparison requires returning to the **squared error** cost function.

$$\begin{aligned}
 \mathbb{E} \left\{ (\hat{\theta}_{\text{ML}}(Y) - \theta)^2 \right\} &= \text{var} \left\{ \hat{\theta}_{\text{ML}}(Y) \right\} + \left(\frac{\theta}{n-1} \right)^2 \\
 &= \frac{n+2}{n-1} \cdot \frac{\theta^2}{n-2} \\
 &> \frac{\theta^2}{n-2} \\
 &= \text{var} \left\{ \hat{\theta}_{\text{MVU}}(Y) \right\} \\
 &> I(\theta).
 \end{aligned}$$

In this problem, the MVU estimator is preferable to the ML estimator. Asymptotically, as $n \rightarrow \infty$ however, their squared error performance is equivalent.