ECE531 Screencast 4.2: Maximum Likelihood Estimation

D. Richard Brown III

Worcester Polytechnic Institute

The Maximum Likelihood Criterion

Recall our goal is to find the parameter θ that maximizes the likelihood of a given observation $Y = y$. We wish to solve

$$\hat{\theta}_{\mathsf{ML}}(y) = \arg\max_{\theta \in \Lambda} p_Y(y\,;\,\theta)$$

Note that, since $\ln$ is strictly monotonically increasing,

$$\hat{\theta}_{\mathsf{ML}}(y) = \arg\max_{\theta \in \Lambda} p_Y(y\,;\,\theta) = \arg\max_{\theta \in \Lambda} \ln p_Y(y\,;\,\theta)$$

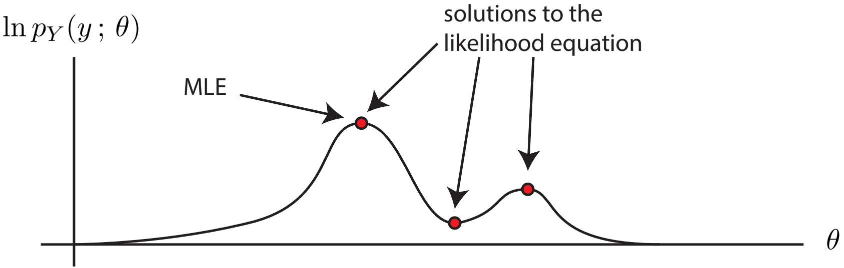
Assuming that $\ln p_Y(y\,;\,\theta)$ is differentiable, we can find the maximum likelihood estimator by looking looking at members of the set

$$\hat{\theta}_{\mathsf{ML}}(y) \in \mathcal{M} = \left\{\theta \,:\, \frac{\partial}{\partial\theta} \ln p_Y(y\,;\,\theta) = 0\right\}$$

where the partial derivative is taken to mean the gradient ∇_θ for multiparameter estimation problems. $\frac{\partial}{\partial\theta} \ln p_Y(y\,;\,\theta) = 0$ is called the "likelihood equation".

The Likelihood Equation and the MLE

For a particular value of y, one might have



In this example, the MLE is unique, but the set of $\theta(y)$ that satisfy the likelihood equation has three members, i.e. $|\mathcal{M}| = 3$.

Maximum Likelihood and Efficiency

Recall that we can attain the CRLB **if and only if** we can write

$$\frac{\partial}{\partial\theta}\ln p_Y(y\,;\,\theta) = I(\theta)\left[g(y)-\theta\right]$$

for all $y \in \mathcal{Y}$ and all $\theta \in \Lambda$. The MVU estimator is $g(y)$ and is efficient.

Now look at the likelihood equation again assuming this condition is true. We have

$$\left.\frac{\partial}{\partial\theta}\ln p_Y(y\,;\,\theta)\right|_{\theta\in\mathcal{M}} = 0 \quad \Rightarrow \quad I(\theta)\left[g(y)-\theta\right]_{\theta\in\mathcal{M}} = 0$$

which, as long as $I(\theta) > 0$, has solution(s) $g(y) = \theta \in \mathcal{M}$.

Hence, if an estimator $g(y)$ is efficient, it must be a solution to the likelihood equation. Not every solution to the likelihood equation is an efficient estimator, however.

An Example of MLE Non-Uniqueness

Suppose you get one observation $Y \sim \mathcal{U}(\theta - 0.5, \theta + 0.5)$ and wish to estimate θ. We can write the parameterized pdf of the observation as

$$p_Y(y\,;\,\theta) = I_{y \in (\theta - 0.5, \theta + 0.5)}$$

where I is the indicator function equal to one if its argument is true and equal to zero otherwise. Since we are thinking of y as fixed and θ as the free parameter, we can rewrite this as

$$p_Y(y\,;\,\theta) = I_{\theta \in (y - 0.5, y + 0.5)}$$

Note that any choice of $\theta \in (y - 0.5, y + 0.5)$ makes the observation y equally likely. Hence, there are an infinite number of possibilities for $\hat{\theta}_{\text{ML}}(y)$.

Some Initial Properties of Maximum Likelihood Estimators

Maximum likelihood estimator definition:

$$\hat{\theta}_{\mathsf{ML}}(y) = \arg\max_{\theta \in \Lambda} p_Y(y\,;\,\theta) = \arg\max_{\theta \in \Lambda} \ln p_Y(y\,;\,\theta)$$

Likelihood equation (satisfied when $\theta = \hat{\theta}_{\mathsf{ML}}(y)$):

$$\frac{\partial}{\partial \theta} \ln p_Y(y\,;\,\theta) = 0$$

Properties:

- If an estimator is efficient, then it is MVU and it must be a solution to the likelihood equation. In this case we have $\hat{\theta}_{\mathsf{ML}}(y) = \hat{\theta}_{\mathsf{MVU}}(y)$.
- There may be multiple solutions to the likelihood equation.
- The MLE may not be unique.