ECE531 Screencast 4.4: Vector Parameter Maximum Likelihood Estimation

D. Richard Brown III

Worcester Polytechnic Institute

Vector Parameter Maximum Likelihood Estimator

In the case of a vector parameter, the same principle applies with the only difference being that our parameter space Λ is multidimensional. Hence, the goal is to find

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

and the likelihood equation (satisfied when $\theta = \hat{\theta}_{\mathsf{ML}}(y)$) is

$$\nabla_\theta \ln p_Y(y\,;\,\theta) = 0.$$

For a p-parameter estimation problem, there will be p simultaneous equations to solve. The solution may not be unique. Once you get a solution, you should compute the Hessian to check that you did indeed find a maximum.

Example: Estimating a The Mean and Variance of WGN

Suppose we have random observations given by $Y_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ for $k = 0, \dots, n-1$. The unknown vector parameter $\theta = [\mu, \sigma^2]$ where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. The joint density is given as

$$p_Y(y\,;\,\theta) \quad = \quad \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ \frac{-\sum_{k=0}^{n-1}(y_k - \mu)^2}{2\sigma^2} \right\}$$

Rather than computing the gradient $\nabla_\theta \ln p_Y(y\,;\,\theta)$, setting the result equal to zero, and then trying to solve the simultaneous equations, we will use a shortcut here. Recognize that the value of μ that maximizes $p_Y(y\,;\,\theta)$ does not depend on σ^2...

Example: Estimating a The Mean and Variance of WGN

We already know $\hat{\mu}_{\mathsf{ML}}(y)$ for estimating μ:

$$\hat{\mu}_{\mathsf{ML}}(y) \quad = \quad \bar{y} \qquad \text{(unbiased, same as MVU estimator).}$$

We can use this result and now we just need to solve

$$\hat{\sigma}^2_{\mathsf{ML}}(y) \quad = \quad \arg\max_{a>0} \left(\ln \left\{ \frac{1}{(2\pi a)^{n/2}} \exp \left\{ \frac{-\sum_{k=0}^{n-1}(y_k - \bar{y})^2}{2a} \right\} \right\} \right)$$

Skipping the details (standard calculus, see example 7.12 in your textbook), it can be shown that

$$\hat{\sigma}^2_{\mathsf{ML}}(y) = \frac{1}{n} \sum_{k=0}^{n-1} (y_k - \bar{y})^2$$

Example: Estimating a The Mean and Variance of WGN

Is $\hat{\sigma}^2_{\mathsf{ML}}$ biased in this case?

$$\begin{aligned}
\mathrm{E}\left\{\hat{\sigma}^2_{\mathsf{ML}}(Y)\right\} &= \frac{1}{n}\sum_{k=0}^{n-1}\mathrm{E}[(Y_k-\mu)^2 - 2(Y_k-\mu)(\bar{Y}-\mu) + (\bar{Y}-\mu)^2] \\
&= \frac{1}{n}\sum_{k=0}^{n-1}\sigma^2 - \frac{2\sigma^2}{n} + \frac{\sigma^2}{n} \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}$$

Our estimator $\hat{\mu}_{\mathsf{ML}}(y)$ is unbiased, but we see that $\hat{\sigma}^2_{\mathsf{ML}}$ is biased.

The steps in the previous analysis can be followed to show that $\hat{\sigma}^2_{\mathsf{ML}}$ is unbiased if μ is a known parameter. The unknown mean, even though we have an unbiased efficient estimator of it, causes $\hat{\sigma}^2_{\mathsf{ML}}(y)$ to be biased here.

Example: Estimating a The Mean and Variance of WGN

It can also be shown that

$$\mathrm{var}\left\{\hat{\sigma}^2_{\mathsf{ML}}(Y)\right\} = \frac{2(n-1)\sigma^4}{n^2} \quad < \quad \frac{2\sigma^4}{n-1} = \mathrm{var}\left\{\hat{\sigma}^2_{\mathsf{MVU}}(Y)\right\}$$

Which is better, $\hat{\sigma}^2_{\mathsf{ML}}(y)$ or $\hat{\sigma}^2_{\mathsf{MVU}}(y)$? To answer this, let's compute the mean squared error (MSE) of the ML estimator for σ^2:

$$\begin{aligned}
\mathrm{E}\left\{(\hat{\sigma}^2_{\mathsf{ML}}(Y) - \sigma^2)^2\right\} &= \mathrm{var}\left\{\hat{\sigma}^2_{\mathsf{ML}}(Y)\right\} + \left(\mathrm{E}\left\{\hat{\sigma}^2_{\mathsf{ML}}(Y)\right\} - \sigma^2\right)^2 \\
&= \frac{2(n-1)\sigma^4}{n^2} + \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 \\
&= \frac{(2n-1)\sigma^4}{n^2} \\
&< \frac{2\sigma^4}{n-1} = \mathrm{E}\left\{(\hat{\sigma}^2_{\mathsf{MVU}}(Y) - \sigma^2)^2\right\}
\end{aligned}$$

Hence the ML estimator has uniformly lower mean squared error (MSE) performance than the MVU estimator. The increase in MSE due to the bias is more than offset by the decreased variance of the ML estimator.

More Properties of ML Estimators

From our examples, we can say that

- Maximum likelihood estimators may be biased.
- A biased ML estimator may, in some cases, outperform a MVU estimator in terms of overall mean squared error.
- It seems that ML estimators are often asymptotically unbiased:

$$\lim_{n\to\infty} \mathrm{E}\left\{\hat{\theta}_{\mathsf{ML}}(Y)\right\} = \theta$$

  Is this always true?

- It seems that ML estimators are often asymptotically efficient:

$$\lim_{n\to\infty} \mathrm{var}\left\{\hat{\theta}_{\mathsf{ML}}(Y)\right\} = I^{-1}(\theta)$$

  Is this always true?