Maximum Likelihood Example

Estimation of Signal Amplitude

Setup:

$$Y_k = N_k + \mu s_k \quad \text{for} \quad k = 1, \ldots, n$$
$$\underline{N} \sim \mathcal{N}(\underline{0}, \sigma^2 I)$$
$$s_k \quad \text{known}$$

Objective: Estimate $\mu \in \mathbb{R}$ and $\sigma^2 \in (0, \infty)$.

Recall the MVUE's:

$$\hat{\mu}_{\mathsf{MVUE}}(\underline{y}) = \frac{1}{n} \frac{\sum_{k=1}^n s_k y_k}{\bar{s^2}} \quad \text{where} \quad \bar{s^2} = \frac{1}{n} \sum_{k=1}^n s_k^2$$

$$\hat{\sigma}^2_{\mathsf{MVUE}}(\underline{y}) = \frac{1}{n-1} \sum_{k=1}^n \left(y_k - \hat{\mu}_{\mathsf{MVUE}}(\underline{y}) s_k\right)^2$$

$$\begin{bmatrix} T_1(\underline{y}) \\ T_2(\underline{y}) \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^n s_k y_k \\ \sum_{k=1}^n y_k^2 \end{bmatrix} \quad \text{complete sufficient statistics}$$

Form likelihood equation:

$$\log p_\theta(\underline{y}) = -\frac{1}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{k=1}^n (y_k - \mu s_k)^2$$

Two equations, two unknowns:

$$\frac{\partial}{\partial \mu} \log p_\theta(\underline{y}) = \frac{1}{\sigma^2} \sum_{k=1}^n (y_k - \mu s_k)\, s_k \Bigg|_{\substack{\mu = \hat{\mu}_{\mathsf{ML}} \\ \sigma^2 = \hat{\sigma}^2_{\mathsf{ML}}}} = 0$$

$$\frac{\partial}{\partial \sigma^2} \log p_\theta(\underline{y}) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{k=1}^n (y_k - \mu s_k)^2 \Bigg|_{\substack{\mu = \hat{\mu}_{\mathsf{ML}} \\ \sigma^2 = \hat{\sigma}^2_{\mathsf{ML}}}} = 0$$

Solving the first equation for $\hat{\mu}_{\mathsf{ML}}$ yields (likelihood global max)

$$\hat{\mu}_{\mathsf{ML}}(\underline{y}) = \frac{1}{n} \frac{\sum_{k=1}^n s_k y_k}{\bar{s^2}} = \hat{\mu}_{\mathsf{MVUE}}(\underline{y})$$

Then using $\hat{\mu}_{\mathsf{ML}}$ determine $\hat{\sigma}^2_{\mathsf{ML}}$, the likelihood global max is

$$\hat{\sigma}^2_{\mathsf{ML}}(\underline{y}) = \frac{1}{n} \sum_{k=1}^n \left(y_k - \hat{\mu}_{\mathsf{ML}}(\underline{y}) s_k\right)^2 = \frac{n-1}{n}\, \hat{\sigma}^2_{\mathsf{MVUE}}(\underline{y})$$

So can see that $\hat{\sigma}^2_{\mathsf{ML}}(\underline{y})$ is *biased*.

Performance: μ unknown, σ^2 known

Determine the estimator variance and Fisher's information...

$$\begin{aligned}
\mathbf{Var}_\theta\left(\hat{\mu}_{\text{ML}}(\underline{Y})\right) &= \frac{\sigma^2}{n\bar{s}^2} \quad \text{(MVUE result previously derived)} \\
I_\theta &= \mathbf{E}_\theta\left\{-\frac{\partial^2}{\partial\mu^2}\log p_\theta(\underline{Y})\right\} \\
&= \mathbf{E}_\theta\left\{-\frac{\partial^2}{\partial\mu^2}\left[-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(Y_k-\mu s_k)^2\right]\right\} \\
&= \frac{1}{\sigma^2}\sum_{k=1}^{n}s_k^2 = \frac{n\bar{s}^2}{\sigma^2} \\
\text{CRLB} &= \frac{1}{I_\theta} = \mathbf{Var}_\theta\left(\hat{\mu}_{\text{ML}}(\underline{Y})\right)
\end{aligned}$$

Makes sense, since with $\theta = \mu$ we can write,

$$\frac{\partial}{\partial\theta}\log p_\theta(\underline{y}) = k(\theta)\left[\hat{\theta}_{\text{ML}}(\underline{y}) - \theta\right] \quad \text{with} \quad k(\theta) = I_\theta$$

meaning CRLB achievable.

Performance: μ known, σ^2 unknown

Determine the estimator variance and Fisher's information...

$$\begin{aligned}
\hat{\sigma}^2_{\text{ML}}(\underline{y}) &= \frac{1}{n}\sum_{k=1}^{n}(y_k-\mu s_k)^2 \\
\mathbf{E}_\theta\left\{\hat{\sigma}^2_{\text{ML}}(\underline{Y})\right\} &= \sigma^2 \ \text{unbiased} \\
\mathbf{Var}_\theta\left\{\hat{\sigma}^2_{\text{ML}}(\underline{Y})\right\} &= \frac{2\sigma^4}{n} \\
I_\theta &= \mathbf{E}_\theta\left\{-\frac{\partial^2}{(\partial\sigma^2)^2}\log p_\theta(\underline{Y})\right\} \\
&= \mathbf{E}_\theta\left\{-\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{k=1}^{n}(Y_k-\mu s_k)^2\right\} \\
&= -\frac{n}{2\sigma^4} + \frac{n\sigma^2}{\sigma^6} = \frac{n}{2\sigma^4} \\
\Rightarrow \text{CRLB} &= \frac{1}{I_\theta} = \mathbf{Var}_\theta\left(\hat{\sigma}^2_{\text{ML}}(\underline{Y})\right)
\end{aligned}$$

Again makes sense, since with $\theta = \sigma^2$ we can write,

$$\frac{\partial}{\partial\theta}\log p_\theta(\underline{y}) = k(\theta)\left[\hat{\theta}_{\text{ML}}(\underline{y}) - \theta\right] \quad \text{with} \quad k(\theta) = I_\theta$$

meaning CRLB achievable.

Performance: μ and σ^2 both unknown

Determine the estimator variances and Fisher's information...

$$\begin{aligned}
\mathbf{Var}_\theta\left\{\hat{\sigma}^2_{\text{ML}}(\underline{Y})\right\} &= \mathbf{Var}_\theta\left(\frac{n-1}{n}\hat{\sigma}^2_{\text{MVUE}}(\underline{Y})\right) \\
&= \frac{(n-1)^2}{n^2}\mathbf{Var}_\theta\left\{\hat{\sigma}^2_{\text{MVUE}}(\underline{Y})\right\} \\
&= \frac{(n-1)^2}{n^2}\frac{2\sigma^4}{n-1} = \frac{2(n-1)\sigma^4}{n^2} \\
&< \frac{2\sigma^4}{n-1} = \mathbf{Var}_\theta\left\{\hat{\sigma}^2_{\text{MVUE}}(\underline{Y})\right\}
\end{aligned}$$

How can ML have lower variance than MVUE???
Forcing unbiasedness increased the variance!

Consider mean-squared error:

$$\begin{aligned}
\mathbf{MSE}_{\text{ML}} &= \mathbf{Var}_\theta\left\{\hat{\sigma}^2_{\text{ML}}(\underline{Y})\right\} + \left(\mathbf{E}_\theta\{\hat{\sigma}^2_{\text{ML}}\} - \sigma^2\right)^2 \\
&= \frac{\sigma^4(2n-1)}{n^2} \\
\mathbf{MSE}_{\text{MVUE}} &= \mathbf{Var}_\theta\left\{\hat{\sigma}^2_{\text{MVUE}}(\underline{Y})\right\} = \frac{2\sigma^4}{n-1} \\
\rightarrow \frac{\mathbf{MSE}_{\text{MVUE}}}{\mathbf{MSE}_{\text{ML}}} &= \left(\frac{n}{n-1}\right)\left(\frac{2n}{2n-1}\right) > 1
\end{aligned}$$

Thus ML outperforms MVUE in both variance and MSE!

Summary

- $\hat{\mu}_{\text{ML}} = \hat{\mu}_{\text{MVUE}}$
- for μ known, $\hat{\sigma}^2_{\text{ML}} = \hat{\sigma}^2_{\text{MVUE}}$
- with μ unknown, $\hat{\sigma}^2_{\text{ML}}$ outperformed $\hat{\sigma}^2_{\text{MVUE}}$ in variance and MSE.
- $\mathbf{Var}_\theta\left\{\hat{\sigma}^2_{\text{MVUE}}\right\}_{\mu\text{ known}} < \mathbf{Var}_\theta\left\{\hat{\sigma}^2_{\text{MVUE}}\right\}_{\mu\text{ unknown}}$