Notes on Maximum A Posteriori (MAP) Estimation

Paul F. Roysdon, Ph.D.

I. Estimator Background

When the PDF is viewed as a function of the unknown parameter (with $\boldsymbol{x}$ fixed) it is termed the "likelihood function". Given a generic model

$$\boldsymbol{x}[0] = \mathbf{A} + \boldsymbol{\eta}[0]$$

where

$$\boldsymbol{\eta}[0] \sim \mathcal{N}(0, \sigma^2)$$

and PDF

$$\mathrm{p}_i(\boldsymbol{x}[0]; \mathbf{A}) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{\left[-\frac{1}{2\sigma_i^2}(\boldsymbol{x}[0]-\mathbf{A})^2\right]}.$$

The log-likelihood simplifies the equation

$$\ln \mathrm{p}(\boldsymbol{x}[0]; \mathbf{A}) = -\ln\sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(\boldsymbol{x}[0] - \mathbf{A})^2.$$

The first-derivative is

$$-\frac{\partial \ln \mathrm{p}(\boldsymbol{x}[0]; \mathbf{A})}{\partial \mathbf{A}} = \frac{1}{\sigma^2}(\boldsymbol{x}[0] - \mathbf{A}).$$

The second-derivative is

$$-\frac{\partial^2 \ln \mathrm{p}(\boldsymbol{x}[0]; \mathbf{A})}{\partial \mathbf{A}^2} = \frac{1}{\sigma^2}.$$

The efficiency of an estimator is measured by the curvature of the estimator. The curvature can be computed by the variance of the estimator

$$\mathrm{var}\left\langle \hat{\mathbf{A}} \right\rangle = \frac{1}{-\frac{\partial^2 \ln \mathrm{p}(\boldsymbol{x}[0];\mathbf{A})}{\partial \mathbf{A}^2}},$$

or more conveniently as the expected value

$$-\mathrm{E}\left\langle \frac{\partial^2 \ln \mathrm{p}(\boldsymbol{x}[0]; \mathbf{A})}{\partial \mathbf{A}^2} \right\rangle.$$

II. General Model

Assume a linear model

$$\boldsymbol{x} = \mathbf{H}\boldsymbol{\theta} + \boldsymbol{\eta}$$

where $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$, $\mathbf{H} \in \mathbb{R}^{n\times p}$ with $n > p$ and $\mathrm{rank}(\mathbf{H}) = p$, $\boldsymbol{\theta} \in \mathbb{R}^{p\times 1}$, and $\boldsymbol{\eta} \in \mathbb{R}^{n\times 1}$ and $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$.

The PDF of $\boldsymbol{x}$ is

$$\mathrm{p}(\boldsymbol{x}; \boldsymbol{\theta}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\mathbf{C})}} e^{\left[-\frac{1}{2}(\boldsymbol{x}-\mathbf{H}\boldsymbol{\theta})^\intercal \mathbf{C}^{-1}(\boldsymbol{x}-\mathbf{H}\boldsymbol{\theta})\right]}$$

III. Cramer-Rao Lower Bound

The Cramer-Rao Lower Bound (CRLB) is the lower bound for all unbiased estimators, thus it is the Minimum Variance Unbiased Estimator (MVUE). Given the model in Section II, the CRLB is defined as

$$\frac{\partial \ln \mathrm{p}(\boldsymbol{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = (\mathbf{H}^\intercal \mathbf{C}^{-1}\mathbf{H})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$

where

$$\hat{\boldsymbol{\theta}}_{\text{CRLB}} = (\mathbf{H}^\intercal \mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^\intercal \mathbf{C}^{-1}\boldsymbol{x}$$

Note if $\hat{\boldsymbol{\theta}}$ is the MVUE, it is an "efficient estimator", and the minimum variance is the diagonal entries of the covariance matrix

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathbf{I}^{-1}(\boldsymbol{\theta}) = (\mathbf{H}^\intercal \mathbf{C}^{-1}\mathbf{H})^{-1}$$

IV. Maximum Likelihood Estimator

Given the model in Section II, the Maximum Likelihood Estimator (MLE) maximizes the likelihood of the parameter $\boldsymbol{\theta}$ such that

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\text{MLE}} &= \underset{\boldsymbol{\theta}}{\mathit{arg\,max}}\ \ \mathrm{p}(\boldsymbol{x}; \boldsymbol{\theta}) \\ &= (\mathbf{H}^\intercal \mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^\intercal \mathbf{C}^{-1}\boldsymbol{x}\end{aligned}$$

If $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ is "efficient", then $\hat{\boldsymbol{\theta}}_{\text{MLE}} \mapsto$ CRLB and thus is the MVUE.

Note, in general the MLE is not optimal. However, as $n \to \infty$ then the MLE converges to the MVUE asymptotically.

V. Maximum A Posteriori Estimator

Given the model in Section II, the Maximum A Posteriori (MAP) Estimator maximizes the posterior PDF, such that

$$\hat{\boldsymbol{\theta}}_{\text{MAP}} = \underset{\boldsymbol{\theta}}{\mathit{arg\,max}}\ \ \mathrm{p}(\boldsymbol{\theta}|\boldsymbol{x}),$$

where

$$\mathrm{p}(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{\mathrm{p}(\boldsymbol{x}|\boldsymbol{\theta})\mathrm{p}(\boldsymbol{\theta})}{\mathrm{p}(\boldsymbol{x})},$$

also known as "Bayes Risk". Therefore

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\text{MAP}} &= \underset{\boldsymbol{\theta}}{\mathit{arg\,max}}\ \ \mathrm{p}(\boldsymbol{x}|\boldsymbol{\theta})\mathrm{p}(\boldsymbol{\theta}) \\ &= \underset{\boldsymbol{\theta}}{\mathit{arg\,max}} \left[\ln \mathrm{p}(\boldsymbol{x}|\boldsymbol{\theta}) + \ln \mathrm{p}(\boldsymbol{\theta})\right].\end{aligned}$$

Note, under normality constraints the MAP is equivalent to the Minimum Mean Square Error (MMSE).