ECE 275A: Parameter Estimation I
Bayesian Estimation – Part II

Florian Meyer

Electrical and Computer Engineering Department
University of California San Diego

UC San Diego
JACOBS SCHOOL OF ENGINEERING

Optimality of $E(\theta|y)$

$$
\begin{aligned}
&\mathrm{BMSE}(\hat{\boldsymbol{\theta}}) \\
&= E_{\mathbf{y},\boldsymbol{\theta}}\left(\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^{\mathsf{T}}\right) \\
&= E_{\mathbf{y},\boldsymbol{\theta}}\left((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathsf{T}}\right) \\
&= E_{\mathbf{y},\boldsymbol{\theta}}\left((\underbrace{\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{\mathrm{MMSE}}}_{\Delta\hat{\boldsymbol{\theta}}} + \underbrace{\hat{\boldsymbol{\theta}}_{\mathrm{MMSE}} - \boldsymbol{\theta}}_{\tilde{\boldsymbol{\theta}}_{\mathrm{MMSE}}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathsf{T}}\right) \\
&= E_{\mathbf{y},\boldsymbol{\theta}}\left((\Delta\hat{\boldsymbol{\theta}} + \tilde{\boldsymbol{\theta}}_{\mathrm{MMSE}})(\Delta\hat{\boldsymbol{\theta}} + \tilde{\boldsymbol{\theta}}_{\mathrm{MMSE}})^{\mathsf{T}}\right) \\
&= E_{\mathbf{y},\boldsymbol{\theta}}\left(\Delta\hat{\boldsymbol{\theta}}\Delta\hat{\boldsymbol{\theta}}^{\mathsf{T}}\right) + 2E_{\mathbf{y},\boldsymbol{\theta}}\left(\Delta\hat{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}_{\mathrm{MMSE}}^{\mathsf{T}}\right) + E_{\mathbf{y},\boldsymbol{\theta}}\left(\tilde{\boldsymbol{\theta}}_{\mathrm{MMSE}}\tilde{\boldsymbol{\theta}}_{\mathrm{MMSE}}^{\mathsf{T}}\right) \\
&= E_{\mathbf{y},\boldsymbol{\theta}}\left(\Delta\hat{\boldsymbol{\theta}}\Delta\hat{\boldsymbol{\theta}}^{\mathsf{T}}\right) + 2E_{\mathbf{y}}\left(\Delta\hat{\boldsymbol{\theta}}\,\underbrace{E_{\boldsymbol{\theta}|\mathbf{y}}\left(\tilde{\boldsymbol{\theta}}_{\mathrm{MMSE}}^{\mathsf{T}}|\boldsymbol{y}\right)}_{=0}\right) + E_{\mathbf{y},\boldsymbol{\theta}}\left(\tilde{\boldsymbol{\theta}}_{\mathrm{MMSE}}\tilde{\boldsymbol{\theta}}_{\mathrm{MMSE}}^{\mathsf{T}}\right) \\
&= \underbrace{E_{\mathbf{y},\boldsymbol{\theta}}\left(\Delta\hat{\boldsymbol{\theta}}\Delta\hat{\boldsymbol{\theta}}^{\mathsf{T}}\right)}_{\succcurlyeq 0} + \underbrace{E_{\mathbf{y},\boldsymbol{\theta}}\left(\tilde{\boldsymbol{\theta}}_{\mathrm{MMSE}}\tilde{\boldsymbol{\theta}}_{\mathrm{MMSE}}^{\mathsf{T}}\right)}_{\mathrm{BMSE}(\hat{\boldsymbol{\theta}}_{\mathrm{MMSE}})} \succcurlyeq \mathrm{BMSE}(\hat{\boldsymbol{\theta}}_{\mathrm{MMSE}})
\end{aligned}
$$

- The equality holds iff $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{\mathrm{MMSE}}$

Example: Binary Symbols in Noise

- Suppose we have a scalar observation $y = x + n$ where $n \sim \mathcal{N}(0, 1)$ and the prior distribution of the unknown parameter $x \in \{-1, 1\}$ is

$$p(x) = \frac{1}{2}\delta(x+1) + \frac{1}{2}\delta(x-1)$$

where $\delta(\cdot)$ is the Kronecker delta and $x \perp\!\!\!\perp n$

- The marginal distribution of y is given by

$$p(y) = \frac{1}{2}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y+1)^2}{2}\right) + \frac{1}{2}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-1)^2}{2}\right)$$

Example: Binary Symbols in Noise

- The posterior distribution $p(x|y)$ can be calculated as

$$
\begin{aligned}
p(x|y) &= \frac{p(y|x)p(x)}{p(y)} \\
&= \frac{\frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(y-x)^2}{2}\right)}{\frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(y+1)^2}{2}\right)+\frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(y-1)^2}{2}\right)}\delta(x+1) \\
&+ \frac{\frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(y-x)^2}{2}\right)}{\frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(y+1)^2}{2}\right)+\frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(y-1)^2}{2}\right)}\delta(x-1) \\
&= \frac{\exp(xy)}{\exp(-y)+\exp(y)}\delta(x+1)+\frac{\exp(xy)}{\exp(-y)+\exp(y)}\delta(x-1)
\end{aligned}
$$

- Thus, we obtain

$$\hat{x}_{\mathrm{MMSE}} = E(x|y) = \frac{\exp(y)-\exp(-y)}{\exp(y)+\exp(-y)} = \tanh(y)$$

- A MAP estimate can be obtained as

$$\hat{x}_{\mathrm{MAP}} = \begin{cases} 1 & \text{if } y \geqslant 0 \\ -1 & \text{if } y < 0 \end{cases}$$

Linear Minimum Mean Square Error Estimation

- Now we restrict the estimator to be a linear function with respect to the observation, i.e.,

$$\hat{\boldsymbol{\theta}} = \boldsymbol{C}\mathbf{y} + \boldsymbol{d}$$

where $\boldsymbol{C} \in \mathbb{R}^{p \times m}$ and $\boldsymbol{d} \in \mathbb{R}^{p}$

- We aim to find the linear estimator $\hat{\boldsymbol{\theta}}_{\mathrm{LMMSE}} = \boldsymbol{A}\mathbf{y} + \boldsymbol{b}$ that minimizes the Bayesian mean square error

- By using the orthogonality principle, we have

$$\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\mathrm{LMMSE}} \perp \mathrm{LF}(\mathbf{y}) \subset \mathscr{L}(\Omega, \Theta)$$

where

$$\mathrm{LF}(\mathbf{y}) = \left\{ \boldsymbol{C}\mathbf{y} + \boldsymbol{d} \,\middle|\, \forall\, \boldsymbol{C} \in \mathbb{R}^{p \times m}, \boldsymbol{d} \in \mathbb{R}^{p} \right\}$$

Linear Minimum Mean Square Error Estimation

$$\begin{aligned}\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\text{LMMSE}} \perp \text{LF}(\mathbf{y}) &\Leftrightarrow \langle \hat{\boldsymbol{\theta}}, \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\text{LMMSE}} \rangle = 0,\ \forall \hat{\boldsymbol{\theta}} \in \text{LF}(\mathbf{y}) \\ &\Rightarrow \langle \boldsymbol{C}\mathbf{y} + \boldsymbol{d}, \boldsymbol{\theta} - \boldsymbol{A}\mathbf{y} - \boldsymbol{b} \rangle = 0,\ \forall\, \boldsymbol{C}, \boldsymbol{d} \\ &\Rightarrow E_{\boldsymbol{\theta},\mathbf{y}}\big((\boldsymbol{C}\mathbf{y} + \boldsymbol{d})^{\mathsf{T}}(\boldsymbol{\theta} - \boldsymbol{A}\mathbf{y} - \boldsymbol{b})\big) = 0,\ \forall\, \boldsymbol{C}, \boldsymbol{d}\end{aligned}$$

- Let $\boldsymbol{C} = \mathbf{0}$, $\boldsymbol{m}_{\theta} = E(\boldsymbol{\theta})$, and $\boldsymbol{m}_{\mathbf{y}} = E(\mathbf{y})$

$$\begin{aligned}&E_{\boldsymbol{\theta},\mathbf{y}}\big(\boldsymbol{d}^{\mathsf{T}}(\boldsymbol{\theta} - \boldsymbol{A}\mathbf{y} - \boldsymbol{b})\big) = 0 \\ \Rightarrow &\boldsymbol{d}^{\mathsf{T}}(\boldsymbol{m}_{\theta} - \boldsymbol{A}\boldsymbol{m}_{\mathbf{y}} - \boldsymbol{b}) = 0\end{aligned}$$

- Now we set $\boldsymbol{d} = \boldsymbol{m}_{\theta} - \boldsymbol{A}\boldsymbol{m}_{\mathbf{y}} - \boldsymbol{b}$

$$\begin{aligned}\Rightarrow &\|\boldsymbol{m}_{\theta} - \boldsymbol{A}\boldsymbol{m}_{\mathbf{y}} - \boldsymbol{b}\|^2 = 0 \\ \Rightarrow &\boldsymbol{b} = \boldsymbol{m}_{\theta} - \boldsymbol{A}\boldsymbol{m}_{\mathbf{y}}\end{aligned}$$

- Hence

$$\hat{\boldsymbol{\theta}}_{\text{LMMSE}} = \boldsymbol{m}_{\theta} + \boldsymbol{A}(\mathbf{y} - \boldsymbol{m}_{\mathbf{y}}) \tag{1}$$

Linear Minimum Mean Square Error Estimation

- Let $\boldsymbol{d} = \mathbf{0}$

$$
\begin{aligned}
& E_{\boldsymbol{\theta},\mathbf{y}}\left(\mathbf{y}^{\mathsf{T}}\boldsymbol{C}^{\mathsf{T}}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\mathrm{LMMSE}})\right) = 0 \\
\Rightarrow & E_{\boldsymbol{\theta},\mathbf{y}}\left(\mathbf{y}^{\mathsf{T}}\boldsymbol{C}^{\mathsf{T}}(\underbrace{\boldsymbol{\theta} - \boldsymbol{m}_{\theta}}_{\Delta\boldsymbol{\theta}} - \boldsymbol{A}(\underbrace{\mathbf{y} - \boldsymbol{m}_{\mathbf{y}}}_{\Delta\mathbf{y}}))\right) = 0 \\
\Rightarrow & E_{\boldsymbol{\theta},\mathbf{y}}\left(\mathbf{y}^{\mathsf{T}}\boldsymbol{C}^{\mathsf{T}}(\Delta\boldsymbol{\theta} - \boldsymbol{A}\Delta\mathbf{y})\right) = 0 \\
\Rightarrow & E_{\boldsymbol{\theta},\mathbf{y}}\left(\mathbf{y}^{\mathsf{T}}\boldsymbol{C}^{\mathsf{T}}(\Delta\boldsymbol{\theta} - \boldsymbol{A}\Delta\mathbf{y})\right) - \boldsymbol{m}_{\mathbf{y}}^{\mathsf{T}}\boldsymbol{C}^{\mathsf{T}}\underbrace{E_{\boldsymbol{\theta},\mathbf{y}}\left(\Delta\boldsymbol{\theta} - \boldsymbol{A}\Delta\mathbf{y}\right)}_{=\mathbf{0}} = 0 \\
\Rightarrow & E_{\boldsymbol{\theta},\mathbf{y}}\left(\Delta\mathbf{y}^{\mathsf{T}}\boldsymbol{C}^{\mathsf{T}}(\Delta\boldsymbol{\theta} - \boldsymbol{A}\Delta\mathbf{y})\right) = 0 \\
\Rightarrow & \mathrm{tr}\left(\boldsymbol{C}^{\mathsf{T}}E_{\boldsymbol{\theta},\mathbf{y}}\left((\Delta\boldsymbol{\theta} - \boldsymbol{A}\Delta\mathbf{y})\Delta\mathbf{y}^{\mathsf{T}}\right)\right) = 0 \\
\Rightarrow & \mathrm{tr}\left(\boldsymbol{C}^{\mathsf{T}}(\boldsymbol{C}_{\theta y} - \boldsymbol{A}\boldsymbol{C}_{yy})\right) = 0 \quad \forall \boldsymbol{C}
\end{aligned}
$$

- For $\boldsymbol{C} = \boldsymbol{C}_{\theta y} - \boldsymbol{A}\boldsymbol{C}_{yy}$, we have $\|\boldsymbol{C}_{\theta y} - \boldsymbol{A}\boldsymbol{C}_{yy}\|_F^2 = 0$, i.e.,

$$\boldsymbol{A} = \boldsymbol{C}_{\theta y}\boldsymbol{C}_{yy}^{-1} \tag{2}$$

Linear Minimum Mean Square Error Estimation

- Combining equation (1) and (2), we have

$$\hat{\boldsymbol{\theta}}_{\text{LMMSE}} = \boldsymbol{m}_{\theta} + \boldsymbol{C}_{\theta \mathbf{y}} \boldsymbol{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - \boldsymbol{m}_{\mathbf{y}})$$

- The estimation error can be expressed as

$$\tilde{\boldsymbol{\theta}}_{\text{LMMSE}} = \hat{\boldsymbol{\theta}}_{\text{LMMSE}} - \boldsymbol{\theta} = -\Delta\boldsymbol{\theta} + \boldsymbol{A}\Delta\mathbf{y}$$

- The covariance of the estimator is obtained as

$$\begin{aligned}
\text{Cov}(\hat{\boldsymbol{\theta}}_{\text{LMMSE}}) &= E_{\boldsymbol{\theta},\mathbf{y}}(\tilde{\boldsymbol{\theta}}_{\text{LMMSE}} \tilde{\boldsymbol{\theta}}_{\text{LMMSE}}^{\text{T}}) \\
&= E_{\boldsymbol{\theta},\mathbf{y}}\big((\boldsymbol{A}\Delta\mathbf{y} - \Delta\boldsymbol{\theta})(\boldsymbol{A}\Delta\mathbf{y} - \Delta\boldsymbol{\theta})^{\text{T}}\big) \\
&= \boldsymbol{C}_{\theta\theta} - \boldsymbol{C}_{\theta \mathbf{y}} \boldsymbol{C}_{\mathbf{yy}}^{-1} \boldsymbol{C}_{\mathbf{y}\theta}
\end{aligned}$$

- Note that if $\boldsymbol{\theta}$ and $\mathbf{y}$ are jointly Gaussian, we get $\hat{\boldsymbol{\theta}}_{\text{LMMSE}} = \hat{\boldsymbol{\theta}}_{\text{MMSE}}$