Parameter Estimation - 1

Prof. C. Emre Koksal

ECE 7001: Stochastic Processes, Detection, and Estimation

The Ohio State University

Outline

Parameter Estimation

- Detection: Choose among a discrete set of statistical situations.
- Estimation: Choose among a continuum set of statistical situations.

	estimation	binary detection
true value:	$\theta \in \Lambda \subset \mathbb{R}$	$i \in \{0,1\}$
estimate / decision:	$\hat{\theta} \in \Lambda'$ (not necessarily Λ)	$j \in \{0,1\}$
observation:	$\vec{y} \in \Gamma$	$\vec{y} \in \Gamma$
estimator / detector	$\hat{\theta}(\vec{y})$: $\Gamma \rightarrow \Lambda'$	$\delta(\vec{y}) : \Gamma \rightarrow \{0,1\}$

Outline:

1. Bayesian estimation: Θ is random with known prior distribution. Define cost, minimize average cost.
2. Non-random parameter estimation: no statistical information on Θ. Find minimum variance unbiased estimator.
3. Maximum likelihood estimator.

Bayesian Parameter Estimation

- Setting: Similar to Bayesian detection
- Model: $\Theta \in \Lambda \subset \mathbb{R}$ with $p(\Theta)$ (like priors Π_i)
- Cost: $C(\hat{\theta}, \theta)$ of estimating θ by $\hat{\theta}$ (like $L_{i,j}$)
- Conditional Risk: $R_\theta(\hat{\theta}) = E_{\vec{y}|\theta}[C(\hat{\theta}(\vec{y}), \Theta) \big| \Theta = \theta]$ (like $R_i(\Gamma_j)$)
- Bayes risk: $r(\hat{\theta}) = E_\Theta[R_\Theta(\hat{\theta})] = E[C(\hat{\theta}(\vec{y}), \Theta)]$ (like $r(\delta)$)

Bayes estimator: $\hat{\theta}(\cdot) = \arg\min \ r(\hat{\theta})$

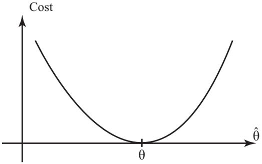
$$\begin{aligned}
\min_{\hat{\theta}(\cdot)} r(\hat{\theta}) &= \min_{\hat{\theta}(\cdot)} E[C(\hat{\theta}(\vec{y}), \Theta)] \\
&= \min_{\hat{\theta}(\cdot)} E_{\vec{y}} \left[E[C(\hat{\theta}(\vec{y}), \Theta) \big| \vec{Y}] \right] \\
&= \min_{\hat{\theta}(\cdot)} \int_\Gamma E[C(\hat{\theta}(\vec{y}), \Theta) \big| \vec{Y} = \vec{y}] \ p(\vec{y}) \ d\vec{y} \\
&\geq \int_\Gamma \min_{\hat{\theta}(\cdot)} E[C(\hat{\theta}(\vec{y}), \Theta) \big| \vec{Y} = \vec{y}] \ p(\vec{y}) \ d\vec{y}
\end{aligned}$$

with equality if and only if

$$\hat{\theta}(\vec{y}) \triangleq \underbrace{\hat{\theta}_B(\vec{y}) = \arg\min_{\hat{\theta}(\cdot)} E[C(\hat{\theta}(\vec{y}), \Theta) \big| \vec{Y} = \vec{y}]}_{\text{different Bayes estimators} \longleftrightarrow \text{different cost functions}}$$

Minimum Mean Squared Error (MMSE) Estimation (Bayes Least Squares (BLS))

$$C(\hat{\theta}(\vec{y}), \theta) = (\hat{\theta}(\vec{y}) - \theta)^2$$



Posterior cost:

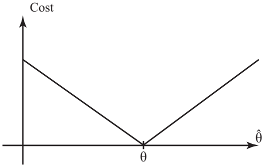
$$\begin{aligned} E[(\hat{\theta}(\vec{y}) - \Theta)^2 | \vec{Y} = \vec{y}] &= E[\hat{\theta}^2(\vec{y}) | \vec{Y} = \vec{y}] + E[\Theta^2 | \vec{Y} = \vec{y}] - 2E[\hat{\theta}(\vec{y})\Theta | \vec{Y} = \vec{y}] \\ &= \hat{\theta}^2(\vec{y}) + E[\Theta^2 | \vec{Y} = \vec{y}] - 2\hat{\theta}(\vec{y}) E[\Theta | \vec{Y} = \vec{y}] \end{aligned}$$

Convex and differentiable fn. of $\hat{\theta}(\vec{y})$

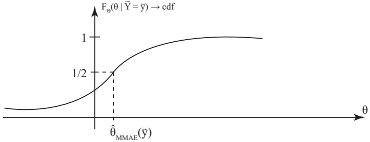
$$\left.\begin{aligned} \frac{\partial}{\partial \hat{\theta}(\vec{y})}(\cdot) &= 2\hat{\theta}(\vec{y}) - 2E[\Theta | \vec{Y} = \vec{y}] \\ \frac{\partial^2}{\partial \hat{\theta}^2(\vec{y})}(\cdot) &= 2 > 0 \ \Rightarrow \ \text{convex} \end{aligned}\right\} \hat{\theta}_{MMSE}(\vec{y}) = E[\Theta | \vec{Y} = \vec{y}]$$

Minimum Mean Absolute Error (MMAE) Estimation

$$C(\hat{\theta}(\vec{y}), \theta) = |\hat{\theta}(\vec{y}) - \theta|$$

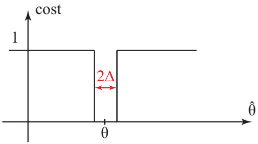


$$\hat{\theta}_{MMAE}(\vec{y}) \left\{ \theta \middle| P(\Theta < \theta | \vec{Y} = \vec{y}) = P(\Theta > \theta | \vec{Y} = \vec{y}) = \frac{1}{2} \right\}$$



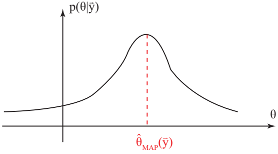
Maximum A Posteriori (MAP) Estimation

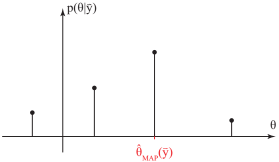
$$C(\hat{\theta}(\vec{y}), \theta) = \begin{cases} 0, & \text{if } |\theta - \hat{\theta}(\vec{y})| \le \Delta \\ 1, & \text{otherwise} \end{cases}, \quad \text{for some } 0 < \Delta \ll 1$$



$$\begin{aligned} E[C(\hat{\theta}(\vec{y}), \Theta) \big| \vec{Y} = \vec{y}] &= P(|\Theta - \hat{\theta}(\vec{y})| > \Delta \big| \vec{Y} = \vec{y}) \cdot 1 + P(|\cdot| \le \Delta \big| \vec{Y} = \vec{y}) \cdot 0 \\ &= 1 - P(|\Theta - \hat{\theta}(\vec{y})| \le \Delta \big| \vec{Y} = \vec{y}) \\ &= 1 - \int_{\hat{\theta}(\vec{y})-\Delta}^{\hat{\theta}(\vec{y})+\Delta} p(\theta \big| \vec{y})\, d\theta \quad \overset{\Delta \ll 1}{\rightarrow} \quad 1 - 2\Delta p(\hat{\theta}(\vec{y}) \big| \vec{y}) \end{aligned}$$

$$\Rightarrow \hat{\theta}_{MAP}(\vec{y}) = \arg\min_{\theta} p(\theta \big| \vec{y}) \Rightarrow$$ conditional mode of the distribution





- Summary:

$$\left.\begin{array}{ll} \text{MMSE} & - \text{mean} \\ \text{MMAE} & - \text{median} \\ \text{MAP} & - \text{mode} \end{array}\right\} \quad \text{of } p(\theta \big| \vec{y})$$

- Note: Given $p(\vec{y}|\theta)$, $p(\theta)$, $p(\theta\big|\vec{y}) = \frac{p(\vec{y}\big|\theta)p(\theta)}{p(\vec{y})}$
- minor note: If $p(\theta\big|\vec{y})$ is symmetric and unimodal $\Rightarrow$ MMSE = MMAE = MAP

Bayesian Estimation Example

Estimating the parameter of an exponential pdf:

Problem setup: $y \in \Lambda = [0, \infty)$, $\theta \in \Gamma = [0, \infty)$

$$p_\theta(y) = \begin{cases} \theta e^{-\theta y}, & y \geq 0 \\ 0, & y < 0 \end{cases}$$ exponential pdf with unknown parameter θ

$$w(\theta) = \begin{cases} \alpha e^{-\alpha \theta}, & \theta \geq 0 \\ 0, & \theta < 0 \end{cases}$$ known $\alpha > 0$

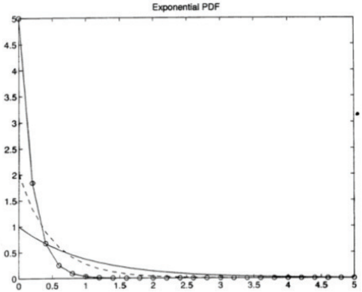
Bayes estimators based on conditional pdf $w(\theta\big|y)$:

$$w(\theta\big|y) = \frac{p_\theta(y)w(\theta)}{p(y)} = \frac{p_\theta(y)w(\theta)}{\int_\Gamma p_\theta(y)w(\theta)d\theta}$$
$$= \frac{\theta e^{-\theta y}\alpha e^{-\alpha\theta}}{\int_0^\infty \theta e^{-\theta y}\alpha e^{-\alpha\theta}d\theta}$$
$$= (\alpha + y)^2\theta e^{-\theta(\alpha+y)} \quad \theta, y > 0$$

Since

$$p(y) = \alpha\int_0^\infty \theta e^{-(y+\alpha)\theta}d\theta = \frac{\alpha}{(\alpha+y)^2} \Rightarrow E[Y] = \infty$$

Consider the exponential pdf: $p_\alpha(x) = \alpha e^{-\alpha x}$



$$E[X] = 1/\alpha$$

$$var(X) = 1/\alpha^2$$

$$E[X^2] = (1/\alpha)^2 + 1/\alpha^2 = 2/\alpha^2$$

MMSE Estimate

The conditional mean estimator:

$$\begin{aligned}
\hat{\theta}_{MMSE}(y) &= E[\Theta|y] = \int_0^\infty \theta\omega(\theta|y)d\theta \\
&= \int_0^\infty (\alpha+y)^2\theta^2 e^{-\theta(\alpha+y)}d\theta \\
&= (\alpha+y)\underbrace{\int_0^\infty \theta^2(\alpha+y)e^{-\theta(\alpha+y)}d\theta}_{=\frac{2}{(\alpha+y)^2} \text{ via } E[X^2] \text{ property}} \\
&= \frac{2}{\alpha+y} = \frac{1}{\frac{\alpha+y}{2}} = \frac{1}{\text{avg of } y \text{ and } \alpha}
\end{aligned}$$

What is the Bayes cost?

$$\begin{aligned}
r(\hat{\theta}_{MMSE}) &= \text{MMSE} = E[(\hat{\theta}_{MMSE}(Y)-\Theta)^2] \\
&= E_Y[E_{\Theta|Y}[(\hat{\theta}_{MMSE}(Y)-\Theta)^2|Y]] \\
&= E_Y[E_{\Theta|Y}[E[\Theta|Y]-\Theta)^2|Y]] \\
&= E_Y[var(\Theta|Y)]
\end{aligned}$$

Using $var(\Theta|Y) = E[\Theta^2|Y] - E^2[\Theta|Y]$, where $E[\Theta|Y] = \frac{2}{\alpha+Y}$ from earlier

$$\begin{aligned}
E[\Theta^2|Y] &= (\alpha+Y)\int_0^\infty \theta^3(\alpha+Y)e^{-\theta(\alpha+Y)}d\theta \\
&= (\alpha+Y)\left[-\theta^3 e^{-\theta(\alpha+Y)}\Big|_0^\infty + \int_0^\infty 3\theta^2 e^{-\theta(\alpha+Y)}d\theta\right] \\
&= 3\int_0^\infty \theta^2(\alpha+Y)e^{-\theta(\alpha+Y)}d\theta \\
&= 3\frac{2}{(\alpha+Y)^2} \text{ via } E[X^2] \text{ property}
\end{aligned}$$

we find

$$\begin{aligned}
var(\Theta|Y) &= \frac{6}{(\alpha+Y)^2} - \left(\frac{2}{\alpha+Y}\right)^2 = \frac{2}{(\alpha+Y)^2} \\
\text{MMSE} &= E\left[\frac{2}{(\alpha+Y)^2}\right] \\
&= \int_0^\infty \frac{2}{(\alpha+y)^2}\frac{\alpha}{(\alpha+y)^2}dy = \frac{2}{3\alpha^2}
\end{aligned}$$

MMAE Estimate

The conditional median estimator:

$$\hat{\theta}_{MMSE}(y) = \{\hat{\theta} : Pr[\Theta < \hat{\theta}|y] = Pr[\Theta > \hat{\theta}|y]\}$$

Since $\omega(\theta|y)$ continuous,

$$\exists\hat{\theta} \text{ s.t. } \frac{1}{2} = Pr[\Theta < \hat{\theta}|y] = Pr[\Theta > \hat{\theta}|y]$$

Thus can find MMAE estimator via

$$\begin{aligned}
\frac{1}{2} &= \int_{\hat{\theta}_{MMAE}(y)}^{\infty} \omega(\theta|y)d\theta = \int_{\hat{\theta}_{MMAE}(y)}^{\infty} (\alpha + y)^2 \theta e^{-\theta(\alpha+Y)} d\theta \\
&= [1 + (\alpha + y)\hat{\theta}_{MMAE}(y)]e^{-(\alpha+Y)\hat{\theta}_{MMAE}(y)} \\
\Rightarrow \hat{\theta}_{MMAE}(y) &= \frac{T_\theta}{\alpha + y} \\
\text{where } T_0 \text{ solves } \frac{1}{2} &= (1 + T_0)e^{-T_0} \\
T_0 &\approx 1.68
\end{aligned}$$

Compare $\hat{\theta}_{MMSE}(y)$ to $\hat{\theta}_{MMAE}(y)$ ...

MAP Estimate

The conditional mode estimator:

$$\begin{aligned}\hat{\theta}_{MMSE}(y) &= \arg\max_{\theta}\left\{\omega(\theta \middle| y)\right\} \\ &= \arg\max_{\theta}\left\{\frac{p_\theta(y)\omega(\theta)}{p(y)}\right\} \\ &= \arg\max_{\theta}\left\{p_\theta(y)\omega(\theta)\right\} \\ &= \arg\max_{\theta}\left\{\log p_\theta(y) + \log \omega(\theta)\right\}\end{aligned}$$

and for this example

$$\hat{\theta}_{MAP}(y) = \arg\max_{\theta}\left\{\log\theta - \theta y + \log\alpha - \alpha\theta\right\}$$

Since function is concave, have global maximum.

$$\frac{\partial}{\partial\theta}\{\log\theta - \theta y + \log\alpha - \alpha\theta\} = \frac{1}{\theta} - (\alpha + y) = 0$$
$$\Rightarrow \hat{\theta}_{MAP}(y) = \frac{1}{\alpha + y}$$

- All three estimators have the same form of constant $\alpha + y$

Orthogonality Theorem

- Definition: Estimation error is $\hat{\theta}(\vec{y}) - \Theta$
- Theorem: An estimator $\hat{\theta}(\vec{y})$ is MMSE if and only if the associated estimation error $\hat{\theta}(\vec{y}) - \Theta$ is orthogonal to any function $g(\vec{y})$ of the observation:

$$E[(\hat{\theta}(\vec{y}) - \Theta)g(\vec{y})] = 0$$

- Proof: If $\rightarrow$ $\hat{\theta} \equiv \hat{\theta}_{MMSE}$. For any arbitrary $g(\cdot)$,

$$\begin{aligned} E[\Theta g(\vec{y})] &= E[E[\Theta g(\vec{y}) \big| \vec{Y} = \vec{y}]] \\ &= E[g(\vec{y}) E[\Theta \big| \vec{Y} = \vec{y}]] \\ &= E[g(\vec{y}) \hat{\theta}_{MMSE}(\vec{y})] \end{aligned}$$

$$
\begin{aligned}
\text{Only if } \rightarrow \; 0 &= E[\Theta \; g(\vec{y})] - E[\hat{\theta}(\vec{y}) \; g(\vec{y})] \\
&= E[E[\Theta \; g(\vec{y}) \big| \vec{Y} = \vec{y}]] - E[\hat{\theta}(\vec{y}) \; g(\vec{y})] \\
&= E[g(\vec{y}) \; E[\Theta \big| \vec{Y} = \vec{y}]] - E[\hat{\theta}(\vec{y}) \; g(\vec{y})] \\
&= E[g(\vec{y}) \; (E[\Theta \big| \vec{Y} = \vec{y}] - \hat{\theta}(\vec{y}))]
\end{aligned}
$$

for all $g(\vec{y})$. Thus, the equality holds for $g(\vec{y}) = E[\Theta \big| \vec{Y} = \vec{y}] - \hat{\theta}(\vec{y})$. Then,

$$
\begin{aligned}
E[g^2(\vec{y})] = 0 \; \Rightarrow \; g(\vec{y}) = 0 \quad \text{w.p. } 1 \\
\Rightarrow \; \hat{\theta}(\vec{y}) = E[\Theta \big| \vec{Y} = \vec{y}] = \hat{\theta}_{MMSE}(\vec{y}) \quad \text{w.p. } 1
\end{aligned}
$$

MMSE Estimation of a Gaussian Signal From a Jointly Gaussian Observation

Let $\Theta \in \mathbb{R}$ and $\vec{Y} \in \mathbb{R}^n$ be zero-mean jointly Gaussian: for any $\vec{C} \in \mathbb{R}^{n+1}$, $\vec{C}^T[\vec{Y}\Theta]^T$ is a Gaussian random variable.

- Theorem: $\Theta = \vec{\alpha}^T\vec{Y} + N$, N and $\vec{Y}$ are independent, and $\vec{\alpha}^T = K_{\Theta\vec{Y}}\Sigma_{\vec{Y}}^{-1}$ ($K_{\Theta\vec{Y}} = E[\Theta\vec{Y}^T]$, $\Sigma_{\vec{Y}} = E[\vec{Y}\vec{Y}^T]$). N is Gaussian with zero mean and $\sigma_N^2 = \sigma_\Theta^2 - K_{\Theta\vec{Y}}\Sigma_{\vec{Y}}^{-1}K_{\Theta\vec{Y}}^T$.
- Implication: Suppose we want to estimate Θ based on $\vec{Y}$.

$$\begin{aligned}
\hat{\theta}_{MMSE}(\vec{y}) &= E[\Theta|\vec{Y} = \vec{y}] = E[\vec{\alpha}^T\vec{Y} + N|\vec{Y} = \vec{y}] \\
&= E[\vec{\alpha}^T\vec{Y}] + E[N|\vec{Y} = \vec{y}] \\
&= \vec{\alpha}^T\vec{y}
\end{aligned}$$

- estimation error: $\hat{\theta}(\vec{y}) - \Theta = -N \quad (\sigma_N^2 \leq \sigma_\Theta^2)$
- By orthogonality Thm., $\hat{\theta}_{MMSE}(\vec{y}) \perp N$

For non-zero mean: $\hat{\theta}_{MMSE}(\vec{y}) = E[\Theta] + K_{\Theta\vec{Y}}\Sigma_{\vec{Y}}^{-1}(\vec{y} - E[\vec{Y}])$

Linear Least Squared Error (LLSE) Estimation

A linear estimator has the form: $\hat{\theta}(\vec{y}) = \vec{\alpha}^T \vec{y} + b$

- For jointly Gaussian $\vec{Y}$ and Θ, MMSE $\equiv$ LLSE
- For everything else,

$$\vec{\alpha}_{LLSE}, b_{LLSE} = \underset{\vec{\alpha}\in\mathbb{R}, b\in\mathbb{R}}{\arg\min}\, E[(\Theta - (\vec{\alpha}^T \vec{Y} + b))^2]$$

$b = 0$, $E[(\Theta - (\vec{\alpha}^T \vec{Y}))^2] = \sigma_\Theta^2 - 2K_{\Theta\vec{Y}}\vec{\alpha} + \vec{\alpha}^T \Sigma_{\vec{Y}} \vec{\alpha}$ is to be minimized.

$$\frac{\partial}{\partial \vec{\alpha}}(\sigma_\Theta^2 - 2K_{\Theta\vec{Y}}\vec{\alpha} + \vec{\alpha}^T \Sigma_{\vec{Y}} \vec{\alpha}) = 2\vec{\alpha}^T \Sigma_{\vec{Y}} - 2K_{\Theta\vec{Y}} = 0 \;\Rightarrow\; \vec{\alpha}^T = K_{\Theta\vec{Y}} \Sigma_{\vec{Y}}^{-1}$$

$$\hat{\theta}_{LLSE}(\vec{y}) = E[\Theta] = K_{\Theta\vec{Y}}\Sigma_{\vec{Y}}^{-1}(\vec{y} - E[\vec{Y}])$$

associated estimation error variance: $\sigma_{\Theta}^2 - K_{\Theta\vec{Y}}\Sigma_{\vec{Y}}^{-1}K_{\Theta\vec{Y}}^T$

$\Rightarrow$ LLSE estimator minimizes MSE in jointly Gaussian scenario, but it is suboptimal for everything else.

Important Examples

1. Signal + noise
   - $Y = X + Z$, X and Z are independent Gaussian, $X \sim \mathcal{N}(0, \sigma_X^2)$, $Z \sim \mathcal{N}(0, \sigma_Z^2)$
   - Y and X are jointly Gaussian, since for all $\alpha, b \in \mathbb{R}$, $\alpha X + bY = (\alpha + b)X + bN$ is Gaussian.
   - Using our main theorem, given $Y = y$, $X \sim \mathcal{N}(\alpha y, \sigma_{X|Y}^2)$, $\alpha = K_{XY} K_Y^{-1}$ and $\sigma_{X|Y}^2 = K_X - K_{XY} K_Y^{-1} K_{XY}^T$, $K_{XY} = E[XY] = \sigma_X^2$ and $K_Y = E[(X+Z)^2] = \sigma_X^2 + \sigma_Z^2$

$$\Rightarrow \hat{X}(y) = \alpha y = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_Z^2} y \quad \text{and} \quad \hat{Z}(y) = \frac{\sigma_Z^2}{\sigma_X^2 + \sigma_Z^2} y$$

   - $\hat{X}(y) + \hat{Z}(y) = y$
   - $\xi = E[(X - \hat{X}(y))^2 \big| Y = y] = \sigma_X^2 - \frac{\sigma_X^4}{\sigma_X^2 + \sigma_Y^2} \Rightarrow \frac{1}{\sigma_\xi^2} = \frac{1}{\sigma_X^2} + \frac{1}{\sigma_Z^2}$

2. Alternated signal + noise

$Y = hX + Z$, where $X \sim \mathcal{N}(0, \sigma_X^2)$, $Z \sim \mathcal{N}(0, \sigma_Z^2)$ are independent.

Now, $K_{XY} = h\sigma_X^2$ and $K_Y = h^2\sigma_X^2 + \sigma_Z^2$

$$\Rightarrow \hat{X}(y) = \frac{h^2\sigma_X^2}{h^2\sigma_X^2 + \sigma_Z^2} y \quad \text{and} \quad \frac{1}{\sigma_\xi^2} = \frac{1}{\sigma_X^2} + \frac{h^2}{\sigma_Z^2}$$

$$\frac{\sigma_\xi^2}{\sigma_X^2} = \frac{1}{\frac{h^2\sigma_X^2}{\sigma_Z^2} + 1} = \frac{1}{1 + SNR}$$

Recall: for $\vec{\Theta} = \Theta_1, ..., \Theta_n$, $\vec{Y} = Y_1, ..., Y_m$ zero mean jointly Gaussian r.v.

$\vec{\Theta} = G\vec{Y} + \vec{V}$ where $\vec{V}$ independent from $\vec{Y}$, zero mean Gaussian

$G = K_{\Theta Y} K_Y^{-1}$, $K_V = K_\Theta - K_{\Theta Y} K_Y^{-1} K_{\Theta Y}^T$

Then,

$$\hat{\theta}_{MMSE}(\vec{y}) = E[\vec{\Theta}|\vec{Y} = \vec{y}] = E[G\vec{Y} + \vec{V}|\vec{Y} = \vec{y}] = G\vec{y} \quad (1)$$

Let $\vec{\xi} = \vec{\Theta} - \hat{\Theta}_{MMSE}(\vec{y})$. Then $\vec{\xi}$ is zero mean Gaussian

$K_{\vec{\xi}} = K_{\vec{V}} = K_\Theta - K_{\Theta Y} K_Y^{-1} K_{\Theta Y}^T$ (2)

Attenuated scalar signal

When $\vec{\Theta} = \Theta$, $\vec{Y} = Y$ such that $Y = h\Theta + Z$, where Θ, Y, Z are scalar zero mean Gaussian, $(\Theta | Y = y) \sim \mathcal{N}(K_{\Theta Y} K_Y^{-1} y, K_\Theta - K_{\Theta Y} K_Y^{-1} K_{\Theta Y}^T)$ where $K_{\Theta Y} = h^2 \sigma_X^2$, $K_Y = \sigma_Y^2 = h^2 \sigma_X^2 + \sigma_Z^2$

Therefore (1) and (2) becomes

$$\hat{\theta}_{MMSE}(y) = \frac{h\sigma_\Theta^2}{h^2\sigma_\Theta^2 + \sigma_Z^2} y$$

$$\sigma_\xi^2 = \sigma_\Theta^2 - \frac{h^2\sigma_\Theta^4}{h^2\sigma_\Theta^2 + \sigma_Z^2} = \frac{\sigma_\Theta^2 \sigma_Z^2}{h^2\sigma_\Theta^2 + \sigma_Z^2}$$

$$\frac{1}{\sigma_\xi^2} = \frac{1}{\sigma_\Theta^2} + \frac{h^2}{\sigma_Z^2}$$

If $E[\Theta] = \bar{\Theta} \neq 0$, then

$$\hat{\theta}_{MMSE}(y) = \bar{\Theta} + \frac{h\sigma_\Theta^2 (y - h\bar{\Theta})}{h^2\sigma_\Theta^2 + \sigma_Z^2} \quad (3)$$

Scalar Iterative Estimation

Multiple noisy observations of $\Theta \sim \mathcal{N}(\bar{\Theta}, \sigma_\Theta^2)$;

$$Y_1 = h_1\Theta + Z_1$$
$$\vdots$$
$$Y_n = h_n\Theta + Z_n$$

$Z_i \sim \mathcal{N}(0, \sigma_{Z_i}^2)$ are independent of Θ.

- One method is to estimate Θ jointly based on $\vec{Y} = [Y_1, ..., Y_n]^T$ using (1).
- Here, we develop an iterative approach:

  Let $Y_m^n = [Y_m, Y_{m+1}, ..., Y_n]^T$

  1. We obtain $\hat{\theta}(y_1)$ solely based on y_1
  2. Using $\hat{\theta}(y_1)$ and y_2, obtain $\hat{\theta}(y_1^2)$
  3. Using $\hat{\theta}(y_1)$ and y_3, obtain $\hat{\theta}(y_1^3)$

     $\vdots$

  n. Obtain $\hat{\theta}(y_1^n) \equiv \hat{\theta}(\vec{y})$

Step (1): Obtain $\hat{\theta(y_1)}$. We solved this in (3),

$$\hat{\theta}(y_1) = \bar{\Theta} + \frac{h_1 \sigma_\Theta^2 [y_1 - h_1 \bar{\Theta}]}{h^2 \sigma_\Theta^2 + \sigma_{Z_1}^2}, \qquad \frac{1}{\sigma_\xi^2} = \frac{1}{\sigma_\Theta^2} + \frac{h^2}{\sigma_Z^2}$$

Step 2: Note that $(\Theta \big| Y_1 = y_1) \sim \mathcal{N}(\hat{\theta}(y_1), \xi_1^2)$, $\Theta \big| Y_1 = y_1$ is independent of Z_2. Then, we estimate $(\Theta \big| Y_1 = y_1)$ based on $Y_2 = y_2$.

$\Rightarrow$ In the equation (3), where we use $\theta(\hat{y_1})$ instead of $\bar{\Theta}$, $\sigma_{\xi_1}^2$ instead of σ_Θ^2 and h_2 instead of h.

$$\hat{\theta}(y_1^2) = \hat{\theta}(y_1) + \frac{h_2 \sigma_{\xi_1}^2 [y_2 - h_2 \hat{\theta}(y_1)]}{h_2^2 \sigma_{\xi_1}^2 + \sigma_{Z_2}^2}$$

$$\frac{1}{\sigma_{\xi_2}^2} = \frac{1}{\sigma_{\xi_1}^2} + \frac{h_2^2}{\sigma_Z^2}$$

Step3-n: Note that, $(\Theta | Y_1 = y_1, Y_2 = y_2) \sim \mathcal{N}(\hat{\theta}(y_1^2), \xi_2^2)$

- We first conditioned on $Y_1 = y_1$, then on $Y_2 = y_2$

- This is identical to conditioning on $(Y_1 = y_1, Y_2 = y_2)$, since

$$f_{\Theta|Y_1,Y_2}(\theta|y_1,y_2) = \frac{f_{Y_2|\Theta Y_1}(y_2|\theta y_1)}{f_{Y_2|Y_1}(y_2|y_1)} f_{\Theta|Y_1}(\theta|y_1)$$

After n steps, obtain

$$\hat{\theta}(\vec{y}_1^n) = \hat{\theta}(\vec{y}) = \hat{\theta}(y_1^{n-1}) + \frac{h_n \sigma^2_{\xi_n - 1}[y_n - h_n \hat{\theta}(y_1^{n-1})]}{h_n \sigma^2_{\xi_n - 1} + \sigma^2_{Z_n}}$$

$$\frac{1}{\sigma^2_{\xi_n}} = \frac{1}{\sigma^2_{\xi_{n-1}}} + \frac{h_2^2}{\sigma^2_{Z_n}}$$

Also, we could have jointly solved to obtain

$$\hat{\theta}(y_1^n) = \sigma^2_{\xi_n} \sum_{j=1}^{n} \frac{h_j y_j}{\sigma^2_{Z_j}}, \quad \frac{1}{\sigma^2_{\xi_n}} = \frac{1}{\sigma^2_{\Theta}} + \sum_{i=1}^{n} \frac{h_i^2}{\sigma^2_{Z_i}}$$

Scalar Kalman Filter

Let $\Theta_1, \Theta_2, \ldots,$ be a Gauss-Markov process such that $\Theta_n \sim \mathcal{N}(\bar{\Theta}_n, \sigma^2_{\Theta_n})$ for any n and $\Theta_{n+1} = \alpha_n \Theta_n + W_n, W_n \sim \mathcal{N}(0, \sigma^2_{W_n})$ are independent of Θ_n.

Noisy observations are:

$$Y_n = h_n \Theta_n + Z_n, \quad Z_n \sim \mathcal{N}(0, \sigma^2_{Z_n})$$

We want MMSE estimate of Θ_n and Θ_{n+1} based on Y_1^n. We will use the iterative method we develop.

Step (1): Based on $Y_1 = y_1$, estimate: $\hat{\theta}_1(y_1)$ and $\hat{\theta}_2(y_1)$.
Let $\xi_n = \Theta_n - \hat{\Theta}_n(y_1^n), \zeta_N = \Theta_n - \hat{\Theta}_n(y_1^{n-1})$
Then, we know that

$$\hat{\theta}_1(y_1) = \bar{\Theta}_1 + \frac{h_1 \sigma^2_{\Theta_1}[y_1 - h_1 \bar{\Theta}_1]}{h_1^2 \sigma^2_{\Theta_1} + \sigma^2_{Z_1}}$$

$$\frac{1}{\sigma^2_{\xi_1}} = \frac{1}{\sigma^2_{\Theta_1}} + \frac{h_1^2}{\sigma^2_{Z_1}}$$

Note that, $(\Theta_1 \big| Y_1 = y_1) \sim \mathcal{N}(\hat{\theta}_1(y_1), \sigma^2_{\xi_1})$.
Therefore,

$$(\Theta_2 \big| Y_1 = y_1) = (\alpha\Theta_1 + W_1 \big| Y_1 = y_1) \sim \mathcal{N}(\alpha_1 \hat{\theta}_1(y_1), \alpha_1^2 \sigma^2_{\xi_1} + \sigma^2_{W_1})$$

since W_1 is independent of Y_1, Θ_1

$$\Rightarrow \hat{\theta}_2(y_1) = \alpha_1 \hat{\theta}_1(y_1), \quad \sigma^2_{\zeta_2} = \alpha_1^2 \sigma^2_{\xi_1} + \sigma^2_{W_1}$$

Step (2): Now estimate $\hat{\theta}_2(y_1^2)$, $\hat{\theta}_3(y_1^2)$ based on $\hat{\theta}_1(y_1), \hat{\theta}_2(y_1)$.

$$\hat{\theta}_2(y_1^2) = \hat{\theta}_2(y_1) + \frac{h_2\sigma^2_{\zeta_2}[y_2 - h_n\hat{\theta}_2(y_1)]}{h_2^2\sigma^2_{\zeta_2} + \sigma^2_{Z_n}}$$

$$\frac{1}{\sigma^2_{\xi_2}} = \frac{1}{\sigma^2_{\zeta_2}} + \frac{h_2^2}{\sigma^2_{Z_2}}, \quad \text{and similarly get} \quad \hat{\theta}_3(y_1^2)$$

After n steps, we obtain,

$$\hat{\theta}_n(y_1^{n-1}) = \alpha_{n-1}\hat{\theta}_{n-1}(y_1^{n-1})$$

$$\sigma^2_{\zeta_n} = \alpha^2_{n-1}\sigma^2_{\xi_{n-1}} + \sigma^2_{W_{n-1}}$$

$$\hat{\theta}_n(y_1^n) = \hat{\theta}_n(y_1^{n-1}) + \frac{h_n\sigma^2_{\zeta_n}[y_n - \hat{\theta}_n(y_1^{n-1})]}{h_n^2\sigma^2_{\zeta_n} + \sigma^2_{Z_n}}$$

$$\frac{1}{\sigma^2_{\xi_n}} = \frac{1}{\sigma^2_{\zeta_n}} + \frac{h_n^2}{\sigma^2_{Z_n}}$$

- The variance terms $\sigma^2_{\xi_n}, \sigma^2_{\zeta_n}$ could be precomputed.
- The estimates are linear in the observations and $\bar{\Theta}_1$.

For $\sigma^2_{\xi_n}$, we can get the following recursion:

$$\frac{1}{\sigma^2_{\xi_n}} = \frac{1}{\alpha^2_{n-1}\sigma^2_{\xi_{n-1}} + \sigma^2_{W_{n-1}}} + \frac{h_n^2}{\sigma^2_{Z_n}}$$

Special Case: Let $h_n = h$ are constant, $0 < \alpha < 1$. $\alpha_n = \alpha$, $\sigma^2_{Z_n} = \sigma^2_Z$, $\sigma^2_{W_n} = \sigma^2_W$.

- Then $E[\Theta_n] = E[\alpha\Theta_{n-1} + W_{n-1}] = E[\alpha^2\Theta_{n-2}] = \alpha^n E[\bar{\Theta_1}]$ goes to zero with increasing n.
- Similarly, $\sigma^2_{\Theta_n}$ varies monotonically from $\sigma^2_{\Theta_1}$, for $n = 1$ to $\frac{\sigma_{W^2}}{1-\alpha}$, for $n \to \infty$
- $\forall\, \sigma^2_{\xi_n}$ is also monotonic in n, approaches a finite limiting value of λ given by the positive root of

$$\alpha^2 h^2 \sigma_Z^{-2} \lambda^2 + [h^2 \sigma_Z^{-2} \sigma_{W^2} + 1 - \alpha^2]\lambda - \sigma^2_W = 0$$