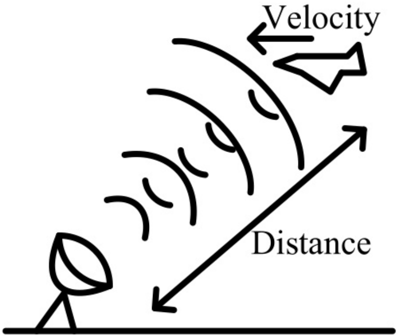
ELEG 5633 Detection and Estimation
Estimation Theory I

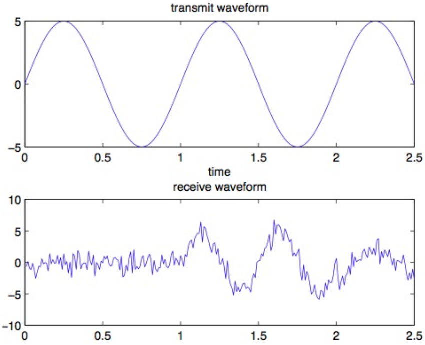
Jingxian Wu

Department of Electrical Engineering
University of Arkansas

Outline

- Bayesian Estimation
  - Basic Ingredients of Estimation Theory
  - Minimum Absolute Error (MAE) Estimator
  - Maximum A Posteriori (MAP) Estimator
  - Minimum Mean Square Error (MMSE) Estimator
  - Linear Minimum MSE (LMMSE) Estimators
  - Orthogonality Principle
- Classical Estimation
  - Maximum Likelihood Estimation
  - Minimum Variance Unbiased (MVUB) Estimators





The received waveform is time-dilated and shifted version of original waveform $g(t)$ plus noise

$$x(t) = g(\alpha t - \tau) + w(t)$$

The parameter is $\theta = [\alpha, \tau]^T$. α is related to velocity, Doppler shift. τ is related to distance, $\tau = \frac{2d}{c}$. Here c is the speed of light.

$$x \to \hat{\tau} \to \hat{d} = \frac{c\hat{\tau}}{2}$$

Observe a moving object with noise. The image is blurry and noisy. We aim to restore the image by deblurring and denoising.

$$\mathbf{x} = \mathbf{h} * \boldsymbol{\theta} + \mathbf{w}$$

where $\boldsymbol{\theta}$ is the ideal image, $\mathbf{h}$ captures the motion blur effect.

Basic Ingredients of Estimation Theory

We observe X, and the goal is to determine the parameter θ that produced X.

- Observation model

$$X \sim p(x|\theta) \quad \theta \in \Theta, x \in \mathcal{X}$$

- Estimator

$$\hat{\theta} : \mathcal{X} \to \Theta$$

  a mapping from $\mathcal{X}$ to Θ.

- Estimate: Given a particular observation of X, say x, $\hat{\theta}(x)$ is called the estimate of θ given observation x.
- Loss/error function

$$\ell : \Theta \times \Theta \to \mathbb{R}^+$$

  $\ell(\theta, \hat{\theta})$ measures proximity of $\hat{\theta}$ to θ

Bayesian Inference

- Basic quantities for Bayesian inference
  - Prior distribution $p(\theta)$
  - likelihood $p(x|\theta)$
  - Loss/error function $\ell(\theta, \hat{\theta})$

- Posterior distribution $p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)d\theta}$

Bayesian Inference

- How data is generated:

$$p(\theta) \to \theta \to p(x|\theta) \to x$$

  which involves the prior and likelihood.

- What we are interested in:

$$x \to p(\theta|x) \to \hat{\theta}$$

  which boils down to computing the posterior distribution.

Loss Functions

- Absolute Error (ℓ_1 loss)

$$\ell(\theta, \hat{\theta}) = \|\theta - \hat{\theta}\|_1 = \sum_{i=1}^{n} |\theta_i - \hat{\theta}_i|$$

- Squared Error (ℓ_2 loss)

$$\ell(\theta, \hat{\theta}) = \|\theta - \hat{\theta}\|_2^2 = \sum_{i=1}^{n} (\theta_i - \hat{\theta}_i)^2$$

- Hit-or-miss loss

$$\ell(\theta, \hat{\theta}) = \begin{cases} 1 & |\theta - \hat{\theta}| > \epsilon \\ 0 & |\theta - \hat{\theta}| < \epsilon \end{cases}$$

Bayes Risk

- Bayes Risk (average/expected loss)

$$R = \mathbb{E}_{x,\theta}[\ell(\theta, \hat{\theta})] = \int_\Theta \int_\mathcal{X} \ell(\theta, \hat{\theta}(x)) p(x, \theta) dx d\theta$$
$$= \int_\mathcal{X} \left[\int_\Theta \ell(\theta, \hat{\theta}(x)) p(\theta | x) d\theta \right] p(x) dx$$

- Estimation:

$$\hat{\theta} = \underset{\theta \in \Theta}{\text{argmin}} \int_\Theta \ell(\theta, \hat{\theta}(x)) p(\theta | x) d\theta$$

Minimum Absolute Error (MAE) Estimator

Consider Absolute Error (ℓ_1 loss) for a scalar case, $\ell(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$.

$$\hat{\theta}_{\text{MAE}}(x) = \arg\min \int_{\Theta} |\theta - \hat{\theta}(x)| p(\theta|x) d\theta := g(\hat{\theta})$$

Since

$$g(\hat{\theta}) = \int_{-\infty}^{\hat{\theta}} (\hat{\theta}(x) - \theta) p(\theta|x) d\theta + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta}(x)) p(\theta|x) d\theta$$

Differentiate with respect to $\hat{\theta}$, we have

$$\frac{dg(\hat{\theta})}{d\hat{\theta}} = \int_{-\infty}^{\hat{\theta}} p(\theta|x) d\theta - \int_{\hat{\theta}}^{\infty} p(\theta|x) d\theta = 0$$

$\hat{\theta}_{\text{MAE}}$ is the median of the posterior PDF of θ!

Minimum Uniform Cost (MUC) Estimator

Consider the hit-or-miss loss

$$\ell(\theta, \hat{\theta}) = \begin{cases} 1 & |\theta - \hat{\theta}| > \epsilon \\ 0 & |\theta - \hat{\theta}| < \epsilon \end{cases}$$

$$\begin{aligned} \hat{\theta}_{\mathsf{MUC}}(x) &= \arg\min \left[1 - \int_{|\theta - \hat{\theta}(x)| < \epsilon} p(\theta|x) d\theta \right] \\ &= \arg\max \int_{|\theta - \hat{\theta}(x)| < \epsilon} p(\theta|x) d\theta \\ &= \arg\max \int_{\hat{\theta}(x) - \epsilon}^{\hat{\theta}(x) + \epsilon} p(\theta|x) d\theta \end{aligned}$$

When $\epsilon \to 0$, MUC becomes the Maximum A Posteriori (MAP) Estimator.

$$\hat{\theta}_{\mathsf{MAP}}(x) = \arg\max p(\theta|x) = \lim_{\epsilon \to 0} \hat{\theta}_{\mathsf{MUC}}(x)$$

$\hat{\theta}_{\mathsf{MAP}}$ is the mode of the posterior PDF of θ!

Minimum Mean Square Error (MMSE) Estimator

Consider the Squared Error (ℓ_2 loss) for a scalar case, , $\ell(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$

$$\hat{\theta}_{\mathsf{MMSE}}(x) = \arg\min \int_{-\infty}^{\infty} (\theta - \hat{\theta}(x))^2 p(\theta|x) d\theta = \mathbb{E}[(\theta - \hat{\theta})^2 | X = x]$$

Taking derivative with respect to $\hat{\theta}$,

$$-2 \int_{-\infty}^{\infty} \theta p(\theta|x) d\theta + 2\hat{\theta}(x) \int_{-\infty}^{\infty} p(\theta|x) d\theta = 0$$

Thus,

$$\hat{\theta}_{\mathsf{MMSE}}(x) = \int_{-\infty}^{\infty} \theta p(\theta|x) d\theta = \mathbb{E}[\theta | X = x]$$

$\hat{\theta}_{\mathsf{MMSE}}$ is the mean of the posterior PDF of θ!

Computation of MAP Estimator

- Computation of MAP Estimator $\hat{\boldsymbol{\theta}}_{MAP}(x)$
  Since
  $$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$
  Then,
  $$\begin{aligned}\hat{\theta}_{\mathsf{MAP}}(x) &= \underset{\theta}{\mathsf{argmax}}\, p(x|\theta)p(\theta) \\ &= \underset{\theta}{\mathsf{argmax}}\, [\log p(x|\theta) + \log p(\theta)]\end{aligned}$$
- Vector extension $\boldsymbol{\theta} \in \mathbb{R}^n$
  $$\hat{\boldsymbol{\theta}}_{\mathsf{MAP}}(x) = \underset{\boldsymbol{\theta}\in\mathbb{R}^n}{\mathsf{argmax}}\, p(\boldsymbol{\theta}|x)$$

Example

Consider N observations $x_i, i = 0, \ldots, N-1$, which are conditionally i.i.d with

$$p(x_i|\theta) = \begin{cases} \theta \exp(-\theta x_i) & x_i > 0 \\ 0 & x_i < 0 \end{cases}$$

$$p(\theta) = \begin{cases} \lambda \exp(-\lambda \theta) & \theta > 0 \\ 0 & \theta < 0 \end{cases}$$

Find the MAP estimator of θ.

Example

Consider N conditionally i.i.d observations generated according to $X_i = A + W_i$, where $W_i \sim \mathcal{N}(0, \sigma^2)$, A is a random parameter uniformly distributed on $[-A_0, A_0]$. A and W_i are independent. What is the MAP estimator of A?

Solution: Example 11.3 on Page 352, [M. Kay Volumn 1].

$$\log p(x|\theta) + \log p(\theta) =$$

$$\log\left(\frac{1}{\sqrt{2\pi\sigma^2}^N}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - A)^2 - \log(2A_0), -A_0 \le A \le A_0$$

Example

We now modify our prior knowledge for the previous example by assuming $A \sim \mathcal{N}(0, \sigma_A^2)$. What are the MAP estimators?