Detection & Estimation Lecture 5

Bayesian II

Xiliang Luo

Prior Knowledge

- Knowing that the unknown parameter must lie in a known interval, we may assign a uniform PDF and model the true value as a realization of a random variable
- Bayesian MSE

$$Bmse(\hat{A}) = E_{A,\boldsymbol{x}}\left[\left(A - \hat{A}\right)^2\right]$$

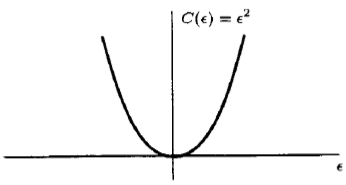
Minimum Bmse Estimator

$$Bmse(\hat{A}) = \int \left[\int (A - \hat{A})^2 p(A|x) dA \right] p(x) dx$$

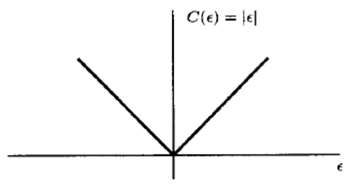
minimize

$$\hat{A} = E[A|x] = \int A p(A|x) dA$$

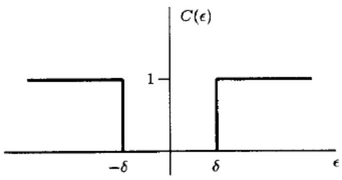
Risk Functions



(a) Quadratic error



(b) Absolute error



(c) Hit-or-miss error

$$R = E[C(\epsilon)]$$

$$\epsilon = \theta - \hat{\theta}$$

Bayes Risk

$$R = E[C(\epsilon)] = \iint C(\theta - \hat{\theta})p(x,\theta)dxd\theta$$

$$= \int dx\, p(x) \int d\theta\, C(\theta - \hat{\theta})p(\theta|x)$$

Absolute Cost Function

$$g(\hat{\theta}) = \int |\theta - \hat{\theta}| p(\theta|x)d\theta$$

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

Median of the posterior PDF

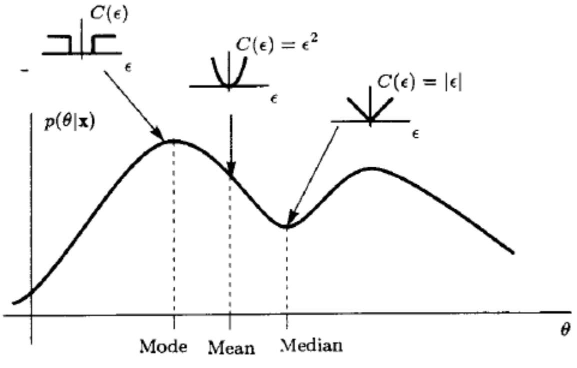
Hit-or-Miss Cost Function

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

$$= 1 - \int_{\hat{\theta}-\delta}^{\hat{\theta}+\delta} p(\theta|x)d\theta$$

Mode (location of the maximum) of the posterior PDF

Mean, Median, Mode



MMSE Estimator

- If $\boldsymbol{\theta}$ is a vector parameter: px1, we can estimate the first parameter by treating the other parameters as nuisance parameters.

$$p(\theta_1|x) = \int \dots \int p(\boldsymbol{\theta}|x)\, d\theta_2 \dots d\theta_p$$

$$\widehat{\theta_1} = E(\theta_1|x) = \int \theta_1 p(\theta_1|x) d\theta_1 = \int \theta_1 p(\boldsymbol{\theta}|x) d\boldsymbol{\theta}$$

$$\hat{\boldsymbol{\theta}} = \boldsymbol{E}(\boldsymbol{\theta}|x)$$

MAP Estimator

$$\hat{\theta} = \arg\max_{\theta} p(\theta|x) = \arg\max_{\theta} p(x|\theta)p(\theta)$$
$$= \arg\max_{\theta} [\ln p(x|\theta) + \ln p(\theta)]$$

Observe the difference/similarity with respect to the MLE

Example: Exponential PDF

$$p(x[n]|\theta) = \begin{cases} \theta \exp(-\theta x[n]) & x[n] > 0 \\ 0 & x[n] < 0 \end{cases}$$

$$p(\mathbf{x}|\theta) = \prod_{n=0}^{N-1} p(x[n]|\theta)$$

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

MAP gives: $\hat{\theta} = \frac{1}{\bar{x} + \frac{\lambda}{N}}$

Linear Bayesian Estimators

- Previous optimal estimators are too computationally intensive to implement
  - Mean
  - Median
  - Mode
- For jointly Gaussian signals, all the above estimators coincide and are linear
- When we are not able to make the Gaussian assumption
  - MMSE
  - linear estimator

LMMSE

- data set:

$$x[0], x[1], \dots, x[N-1]$$

- random parameter to estimate:

$$\theta$$

- Linear estimator:

$$\hat{\theta} = a_0 x[0] + a_1 x[1] + \cdots + a_{N-1} x[N-1] + a_N$$

LMMSE

- Bmse

$$Bmse(\hat{\theta}) = E\left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n] - a_N\right)^2\right]$$

$$a_N = E(\theta) - \sum_{n=0}^{N-1} a_n E[x[n]]$$

$$Bmse(\hat{\theta}) = E\left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n] - E(\theta) + \sum_{n=0}^{N-1} a_n E[x[n]]\right)^2\right]$$

$$Bmse(\hat{\theta}) = E\left[\left(\tilde{\theta} - \sum_{n=0}^{N-1} a_n \widetilde{x[n]}\right)^2\right] \Rightarrow \boldsymbol{a} = C_{xx}^{-1} C_{x\theta}$$

LMMSE

$\hat{\theta} = \boldsymbol{a}^T \boldsymbol{x} + a_N$

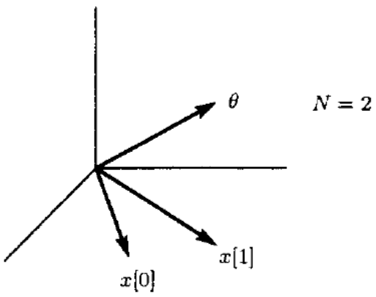
$a_N = E(\theta) - \boldsymbol{a}^T E(\boldsymbol{x})$

$\boldsymbol{a} = C_{xx}^{-1} C_{x\theta}$

$\hat{\theta} = E(\theta) + C_{\theta x} C_{xx}^{-1} (\boldsymbol{x} - E(\boldsymbol{x}))$

$Bmse(\hat{\theta}) = C_{\theta\theta} - C_{\theta x} C_{xx}^{-1} C_{x\theta}$

Geometric View



$\theta, x[0], \dots, x[N-1]$ are all vectors in the vector space of all zero-mean r.v.

The inner product of x and y is defined as:

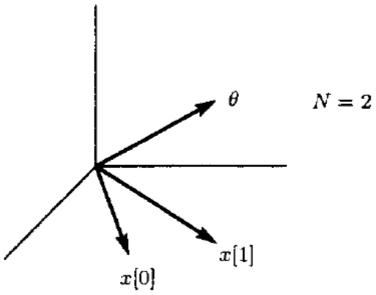
$$(x, y) = E(x \cdot y)$$

Then the length of a vector is:

$$||x||:= \sqrt{E(x^2)}$$

Two vectors are orthogonal when they are uncorrelated

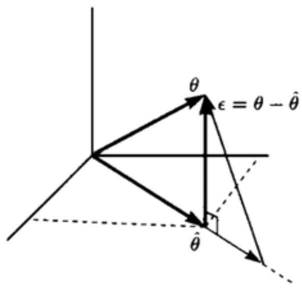
Geometric View



LMMSE for the parameter is to find a linear combination of the data vectors to minimize the length of the error vector!

Orthogonality principle:

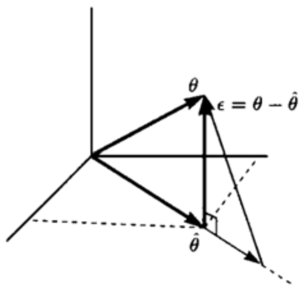
$\epsilon \perp x[0], x[1], \dots, x\{N-1]$



Orthogonality Principle

$E[(\theta-\hat{\theta})x[n]] = 0, n = 0,1,\dots,N-1$

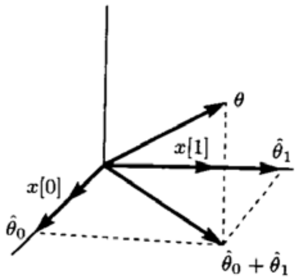
$$\sum_{m=0}^{N-1} a_m E(x[m]x[n]) = E(\theta x[n]), n = 0,\dots,N-1$$



$$\begin{bmatrix} E(x^2[0]) & E(x[0]x[1]) & \dots & E(x[0]x[N-1]) \\ E(x[1]x[0]) & E(x^2[1]) & \dots & E(x[1]x[N-1]) \\ \vdots & \vdots & \ddots & \vdots \\ E(x[N-1]x[0]) & E(x[N-1]x[1]) & \dots & E(x^2[N-1]) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-1} \end{bmatrix} = \begin{bmatrix} E(\theta x[0]) \\ E(\theta x[1]) \\ \vdots \\ E(\theta x[N-1]) \end{bmatrix}$$

$$C_{xx}\boldsymbol{a} = C_{x\theta}$$

Estimation with Independent Data



Independent
←→
orthogonal in the vector space

$\theta - \widehat{\theta_1} - \widehat{\theta_2} \perp$ {data set 1} & {data set 2}

Sequential LMMSE Estimation

- DC Level in WGN

$x[n] = A + w[n]$

$A \sim N(0, \sigma_A^2)$

$w[n] \sim N(0, \sigma^2)$

$$\hat{A}[N-1] = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}} \bar{x} \qquad Bmse\big(\hat{A}[N-1]\big) = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}} \frac{\sigma^2}{N}$$

Woodbury's Formula:

$$(A + uu^H)^{-1} = A^{-1} - A^{-1}u(1 + u^H A^{-1} u)^{-1} u^H A^{-1}$$

Sequential LMMSE Estimation

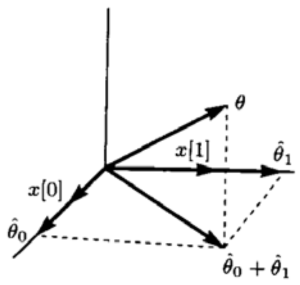
- Question:
  - how to update this LMMSE estimator as a new data sample is available, i.e. $x[N]$?

- S-LMMSE:
  - Estimator Update: $\hat{A}[N] = \hat{A}[N-1] + K[N](x[N] - \hat{A}[N-1])$

$$K[N] = \frac{\mathrm{Bmse}(\hat{A}[N-1])}{\mathrm{Bmse}(\hat{A}[N-1]) + \sigma^2}$$

  - MSE Update: $\mathrm{Bmse}(\hat{A}[N]) = (1 - K[N])\mathrm{Bmse}(\hat{A}[N-1]).$

Sequential LMMSE Estimation



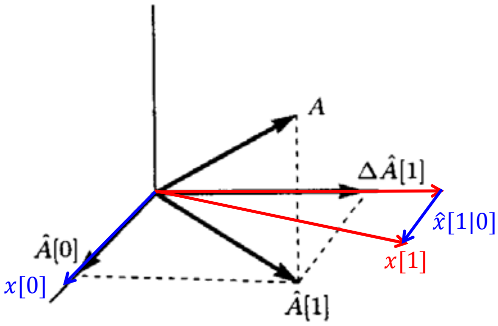
Independent

↔

orthogonal in the vector space

$\theta - \widehat{\theta_1} - \widehat{\theta_2} \perp$ {data set 1} & {data set 2}

Sequential LMMSE Estimation



$\hat{x}[1|0] = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2} x[0]$

$\tilde{x}[1] = x[1] - \hat{x}[1|0]$: innovation

$\hat{x}[1|0] = \hat{A}[0]$

$$\Delta\hat{A}[1] = \frac{E(A\tilde{x}[1])}{E(\tilde{x}[1]\tilde{x}[1])}\tilde{x}[1] = K[1]\tilde{x}[1]$$

HW

- Chapter 11
  - 11.7, 11.11, 11.12
- Chapter 12
  - 12.2, 12. 14, 12.19, 12.20