Bayesian philosophy

Raj Thilak Rajan

Overview

Estimation of a deterministic parameter

- Example *constant in noise* data model: $x[n] = A + w[n]$
- Finding an estimator $\hat{A}$
  - Mean Square Error (MSE)
  - Minimum variance Unbiased Estimator (MVUE)
  - Cramér-Rao lower bound (CRLB)
  - Maximum Likelihood Estimator (MLE)
  - Best Linear Unbiased Estimator (BLUE)
  - Least squares (LS)

Example 1: Classical estimation (1)

- Consider the estimation of A

$$x[n] = A + w[n], \quad n = 0, \cdots, N-1, \quad w[n] \sim \mathcal{N}(0, \sigma^2).$$

- PDF:

$$p(\mathbf{x}; A) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]$$

- Score:

$$\begin{aligned}\frac{\partial \ln p(\mathbf{x}; A)}{\partial A} &= \frac{\partial}{\partial A}\left[-\ln[(2\pi\sigma^2)^{N/2}] - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right] \\ &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A) = \underbrace{\frac{N}{\sigma^2}}_{I(A)}\left(\underbrace{\frac{1}{N}\sum_{n=0}^{N-1}x[n]}_{\hat{A}} - A\right)\end{aligned}$$

Example 1 (2)

- CRLB:

$$var(\hat{A}) \geq \frac{1}{I(A)} = \frac{1}{-\mathbb{E}\left[\frac{\partial^2 \ln p(\mathbf{x};A))}{\partial A^2}\right]} = \frac{\sigma^2}{N}$$

- MVU:

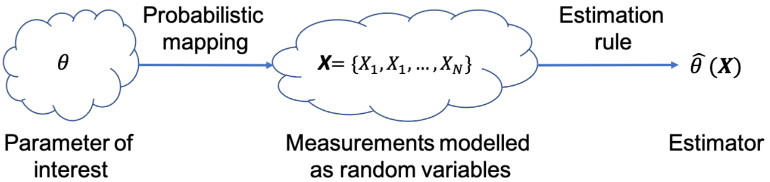
$$\frac{1}{N}\sum_{n=0}^{N-1} x[n]$$

- MLE:

$$\begin{aligned}\frac{\partial \ln p(\mathbf{x};A)}{\partial A} &= \frac{\partial}{\partial A}\left[-\ln[(2\pi\sigma^2)^{N/2}] - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right] \\ &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A) = 0 \quad \Rightarrow \hat{A} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]\end{aligned}$$

- LS, WLS and BLUE also offer identical solutions, under certain conditions

Estimation Philosophy



- Let $X = \{X_1, X_2, \ldots, X_N\}$ be a set of random samples drawn from probability distributions $f_{X_n}(x_n; \boldsymbol{\theta})\ \forall\ 1 \leq n \leq N$, where $\boldsymbol{\theta}$ is the parameter of interest
- We aim to
  - (a) recover the unknown $\boldsymbol{\theta}$ from the measurements X, and
  - (b) provide a performance measure of the estimated $\boldsymbol{\theta}$
- Bayesian philosophy : θ is a random variable and a *prior* $p_\Theta(\theta)$ is known, or the statistics of θ is known.

Bayesian MSE

- θ is viewed as a random variable and we must estimate its particular realization. This allows us to use prior knowledge about θ, i.e., its prior pdf $p(\theta)$. Again, we would like to minimize the MSE

$$Bmse(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2]$$

  but this time both $\mathbf{x}$ and θ are random and the statistics of $\hat{\theta}$ depend on the statistics of both $\mathbf{x}$ and θ.

- Note the difference between these two MSEs:

$$\mathrm{mse}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] = \int (\hat{\theta} - \theta)^2 p(\mathbf{x}; \theta) d\mathbf{x}$$

$$Bmse(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] = \int \int (\hat{\theta} - \theta)^2 p(\mathbf{x}, \theta) d\mathbf{x} d\theta$$

- Note that *mse* depends on θ, but *Bmse* does not, only on its statistics.

Bayes Theorem

Given two random variables X, Y,

- Product rule:
$$p(x,y) = p(x|y)p(y) = p(y|x)p(x)$$
- Bayes theorem
$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(y|x)p(x)}{p(y)}$$
  where
  - $p(x,y)$ is the joint PDF
  - $p(x|y)$ is the posterior PDF
  - $p(y)$ is the marginal PDF of y
  - $p(x)$ is the marginal PDF of x
- If $y \triangleq \theta$ is the unknown parameter of interest, then $p(\theta)$ is the *prior* of θ

Minimum mean square estimation (MMSE)

- We know from Bayes' theorem $p(\mathbf{x},\theta) = p(\theta|\mathbf{x})p(\mathbf{x})$, and hence

$$Bmse(\hat{\theta}) = \int\int(\hat{\theta}-\theta)^2 p(\mathbf{x},\theta)d\mathbf{x}d\theta = \int\left[\int(\hat{\theta}-\theta)^2 p(\theta|\mathbf{x})d\theta\right]p(\mathbf{x})d\mathbf{x},$$

and since $p(\mathbf{x}) \geq 0$ for all $\mathbf{x}$, we minimize the inner integral for each $\mathbf{x}$ i.e.,

$$\text{Solve:} \qquad \underset{\hat{\theta}}{min}\int(\hat{\theta}-\theta)^2 p(\theta|\mathbf{x})d\theta$$

- Solution: Setting the derivative with respect to $\hat{\theta}$ to zero we obtain:

$$\begin{aligned}\frac{\partial}{\partial\hat{\theta}}\int(\hat{\theta}-\theta)^2 p(\theta|\mathbf{x})d\theta &= 2\int(\hat{\theta}-\theta)p(\theta|\mathbf{x})d\theta \\ &= 2\hat{\theta} - 2\int\theta p(\theta|\mathbf{x})d\theta = 0\end{aligned}$$

or

$$\hat{\theta} = \int\theta p(\theta|\mathbf{x})d\theta = \mathbb{E}(\theta|\mathbf{x})$$

Example 2

Consider the probability distributions

$$p_X(x) = 2x, \qquad p_{Y|X}(y|x) = 2xy - x + 1,$$

which exist for $0 < x < 1$, $0 < y < 1$, and are 0 otherwise.Find the MMSE estimate for X given observation $Y = y$.

Solution: $\hat{X} = \dfrac{6y + 1}{2(4y + 1)}$

Example 2 : Gaussian prior (1)

- Consider the estimation of A

$$x[n] = A + w[n], \quad n = 0, \cdots, N-1, \quad w[n] \sim \mathcal{N}(0, \sigma^2) \quad A \sim \mathcal{N}(\mu_A, \sigma_A^2)$$

- Conditional and prior PDF:

$$\begin{aligned} p(\mathbf{x}|A) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right] \\ p(A) &= \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left[-\frac{1}{2\sigma_A^2}(A-\mu_A)^2\right] \end{aligned}$$

- Since both $p(\mathbf{x}|A)$ and $p(A)$ are Gaussian, and assuming $A \perp w[n] \; \forall \; n = 0, 1, \ldots, N-1$, the posterior PDF $p(A|\mathbf{x})$ is also Gaussian:

$$p(A|\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma_{A|x}^2}} \exp\left[-\frac{1}{2\sigma_{A|x}^2}(A-\mu_{A|x})^2\right]$$

with $\sigma_{A|x}^2 = \frac{1}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}}$ and $\mu_{A|x} = (\frac{N}{\sigma^2}\bar{x} + \frac{\mu_A}{\sigma_A^2})\sigma_{A|x}^2$

Example 2 : Gaussian prior (2)

MMSE estimator:

$$\hat{A} = \mathbb{E}(A|\mathbf{x}) = \mu_{A|x} = \frac{\frac{N}{\sigma^2}\bar{x} + \frac{\mu_A}{\sigma_A^2}}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}} = \frac{\sigma_A^2\bar{x} + \frac{\sigma^2}{N}\mu_A}{\frac{\sigma^2}{N} + \sigma_A^2} = \alpha\bar{x} + (1-\alpha)\mu_A \quad (1)$$

where $\alpha = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}}$ and $0 \leq \alpha \leq 1$.

Remarks:

- α: the interplay between the prior knowledge (μ_A) and the data ($\bar{x}$).
- For small N or large σ^2: $\alpha \to 0$, $\sigma_A^2 << \sigma^2/N$ and $\hat{A} = \mu_A$.
- For larger N or small σ^2: $\alpha \approx 1$ and $\hat{A} = \bar{x}$.
- For larger N, the narrower the posterior PDF (and less uncertainty), since

$$\sigma_{A|x}^2 = var[A|\mathbf{x}] = \mathbb{E}\left[(A - E(A|x))^2|A\right] = \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}}$$

Example 2 : Gaussian prior (3)

- MMSE estimate:

$$\begin{aligned}\hat{A} &= \mathbb{E}(A|\mathbf{x}) = \mu_{A|x} \\ p(A|\mathbf{x}) &= \frac{1}{\sqrt{2\pi\sigma^2_{A|x}}}\exp[-\frac{1}{2\sigma^2_{A|x}}(A-\mu_{A|x})^2]\end{aligned}$$

where

$$\mu_{A|x} = \left(\frac{N}{\sigma^2}\bar{x} + \frac{\mu_A}{\sigma^2_A}\right)\sigma^2_{A|x}, \quad \sigma^2_{A|x} = \frac{1}{\frac{N}{\sigma^2}+\frac{1}{\sigma^2_A}}$$

- Remarks:
  - If $N \to \infty$, then $\hat{A} \to \bar{x}$.
  - No prior knowledge i.e., $\sigma^2_A \to \infty$, then $\hat{A} \to \bar{x}$ (i.e., classical estimator)

Bayesian MSE versus Classical MSE

$$
\begin{aligned}
\mathrm{Bmse}(\hat{A}) = E[(A-\hat{A})^2] &= \int\int (A - E[A|\mathbf{x}])^2 p(\mathbf{x}, A) d\mathbf{x} dA \\
&= \int\int (A - E[A|\mathbf{x}])^2 p(A|\mathbf{x}) dA p(\mathbf{x}) d\mathbf{x} \\
&= \int var[A|\mathbf{x}] p(\mathbf{x}) d\mathbf{x} \\
&= \int \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}} p(\mathbf{x}) d\mathbf{x} = \frac{\sigma^2}{N}\left(\frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}}\right) = \frac{\alpha\sigma^2}{N}
\end{aligned}
$$

Hence,

$$
Bmse(\hat{A}) = \frac{\alpha\sigma^2}{N} < \underbrace{\frac{\sigma^2}{N}}_{\text{CRB for classical estimators}} \qquad \because\ 0 \le \alpha \le 1
$$

Using prior knowledge we can improve the estimation accuracy.

Bivariate Gaussian process

If x and y are jointly Gaussian, with joint mean and covariance matrix

$$\mathbb{E}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \mathbb{E}(x) \\ \mathbb{E}(y) \end{bmatrix}, \mathbf{C} = \begin{bmatrix} var(x) & cov(x,y) \\ cov(y,x) & var(y) \end{bmatrix}$$

such that

$$p(x,y) = \frac{1}{(2\pi)\sqrt{\det(\mathbf{C})}} \exp \mathbf{Q}$$

where

$$\mathbf{Q} = -\frac{1}{2}\left[\begin{bmatrix} x - \mathbb{E}(x) \\ y - \mathbb{E}(y) \end{bmatrix}^T \mathbf{C}^{-1} \begin{bmatrix} x - \mathbb{E}(x) \\ y - \mathbb{E}(y) \end{bmatrix}\right]$$

then the conditional PDF $p(y|x)$ is also Gaussian with mean and variance

$$\begin{aligned} \mathbb{E}(y|x) &= \mathbb{E}(y) + \frac{cov(y,x)}{var(x)}(x - \mathbb{E}(x)) \\ var(y|x) &= var(y) - \frac{cov(x,y)^2}{var(x)} = var(y)\left(1 - \frac{cov(x,y)^2}{var(x)var(y)}\right) \\ &= var(y)\left(1 - \rho^2\right) \end{aligned}$$

Multivariate Gaussian process

If $\mathbf{x}$ and $\mathbf{y}$ are jointly Gaussian, where $\mathbf{x}$ is $k \times 1$ and $\mathbf{y}$ is $l \times 1$, with joint mean and covariance matrix

$$\mathbb{E}\left(\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}\right) = \begin{bmatrix}\mathbb{E}(\mathbf{x})\\ \mathbb{E}(\mathbf{y})\end{bmatrix}, \mathbf{C} = \begin{bmatrix}\mathbf{C}_{xx} & \mathbf{C}_{xy}\\ \mathbf{C}_{yx} & \mathbf{C}_{yy}\end{bmatrix}$$

such that

$$p(\mathbf{x}, \mathbf{y}) = \frac{1}{\sqrt{(2\pi)^{k+l}\det(\mathbf{C})}} \exp \mathbf{Q}$$

where

$$\mathbf{Q} = -\frac{1}{2}\left[\begin{bmatrix}\mathbf{x} - \mathbb{E}(\mathbf{x})\\ \mathbf{y} - \mathbb{E}(\mathbf{y})\end{bmatrix}^T \mathbf{C}^{-1} \begin{bmatrix}\mathbf{x} - \mathbb{E}(\mathbf{x})\\ \mathbf{y} - \mathbb{E}(\mathbf{y})\end{bmatrix}\right]$$

then the conditional PDF $p(\mathbf{y}|\mathbf{x})$ is also Gaussian with mean and covariance matrix

$$\begin{aligned} \mathbb{E}(\mathbf{y}|\mathbf{x}) &= \mathbb{E}(\mathbf{y}) + \mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}(\mathbf{x} - \mathbb{E}(\mathbf{x})) \\ \mathbf{C}y|x &= \mathbf{C}yy - \mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy} \end{aligned}$$

Example 5: Vector process (1)

Let us assume now that the prior distribution of A is Gaussian: $A \sim \mathcal{N}(0, \sigma_A^2)$, and $w[n]$ white Gaussian noise, i.e., for $n = 0, ..., N-1$ $w[n] \sim \mathcal{N}(0, \sigma^2)$,

$$\mathbf{x} = \mathbf{1}A + \mathbf{w}.$$

then, $\mathbf{x}$ and A are jointly Gaussian ($k = N$ and $l = 1$), with zero mean and covariance matrix

$$\mathbf{C}_{\mathbf{x},A} = E\left[\begin{bmatrix}\mathbf{x}\\ A\end{bmatrix}\begin{bmatrix}\mathbf{x}^T, A\end{bmatrix}\right] = \begin{bmatrix}\sigma_A^2\mathbf{1}\mathbf{1}^T + \sigma^2\mathbf{I} & \sigma_A^2\mathbf{1}\\ \sigma_A^2\mathbf{1}^T & \sigma_A^2\end{bmatrix}$$

Example 5: Vector process (2)

- Recollect:

$$\begin{aligned}\mathbb{E}(\mathbf{y}|\mathbf{x}) &= \mathbb{E}(\mathbf{y}) + \mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}(\mathbf{x} - \mathbb{E}(\mathbf{x})) \\ \mathbf{C}_{y|x} &= \mathbf{C}_{yy} - \mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}\end{aligned}$$

- Substituting with

$$\mathbb{E} = \begin{bmatrix}\mathbf{x} \\ A\end{bmatrix} = \mathbf{0}, \quad \mathbf{C}_{\mathbf{x},A} = \mathbb{E}\left[\begin{bmatrix}\mathbf{x} \\ A\end{bmatrix}\begin{bmatrix}\mathbf{x}^T, A\end{bmatrix}\right] = \begin{bmatrix}\sigma_A^2\mathbf{1}\mathbf{1}^T + \sigma^2\mathbf{I} & \sigma_A^2\mathbf{1} \\ \sigma_A^2\mathbf{1}^T & \sigma_A^2\end{bmatrix},$$

- Hence, we have

$$\begin{aligned}\mathbb{E}(A|\mathbf{x}) &= \sigma_A^2\mathbf{1}^T(\sigma_A^2\mathbf{1}\mathbf{1}^T + \sigma^2\mathbf{I})^{-1}\mathbf{x} \\ \mathbf{C}_{A|x} &= \sigma_A^2 - \sigma_A^4\mathbf{1}^T(\sigma_A^2\mathbf{1}\mathbf{1}^T + \sigma^2\mathbf{I})^{-1}\mathbf{1}\end{aligned}$$

Example 5 (3)

- Using the matrix inversion lemma (MIL)

$$(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}$$

- Conditional mean

$$\begin{aligned}\mathbb{E}(A|\mathbf{x}) &= \sigma_A^2\mathbf{1}^T(\sigma_A^2\mathbf{1}\mathbf{1}^T+\sigma^2\mathbf{I})^{-1}\mathbf{x} \\ &= (\sigma_A^{-2}+\sigma^{-2}\mathbf{1}^T\mathbf{1})^{-1}\sigma^{-2}\mathbf{1}^T\mathbf{x}\end{aligned}$$

- Conditional covariance

$$\mathbf{C}_{A|x} = \sigma_A^2[1-(\sigma_A^{-2}+\sigma^{-2}\mathbf{1}^T\mathbf{1})^{-1}\sigma^{-2}\mathbf{1}^T\mathbf{1}] = \frac{1}{\frac{1}{\sigma_A^2}+\frac{N}{\sigma^2}}$$

General Linear Gaussian model

- Consider the generalized linear Gaussian model:

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}, \quad \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$$

  where $\boldsymbol{\theta}$ is a random vector with distribution $\mathcal{N}(\boldsymbol{\mu}_\theta, \mathbf{C}_\theta)$.
- Here, $p(\boldsymbol{\theta}|\mathbf{x})$ is also Gaussian with mean and covariance matrix

$$\begin{aligned}\mathbb{E}(\boldsymbol{\theta}|\mathbf{x}) &= \boldsymbol{\mu}_\theta + \mathbf{C}_\theta \mathbf{H}^T (\mathbf{H}\mathbf{C}_\theta \mathbf{H}^T + \mathbf{C})^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_\theta) \\ \mathbf{C}_{\theta|x} &= \mathbf{C}_\theta - \mathbf{C}_\theta \mathbf{H}^T (\mathbf{H}\mathbf{C}_\theta \mathbf{H}^T + \mathbf{C})^{-1} \mathbf{H}\mathbf{C}_\theta\end{aligned}$$

- Alternative formulation using Matrix inversion lemma:

$$\begin{aligned}\mathbb{E}(\boldsymbol{\theta}|\mathbf{x}) &= \boldsymbol{\mu}_\theta + (\mathbf{C}_\theta^{-1} + \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_\theta) \\ \mathbf{C}_{\theta|x} &= (\mathbf{C}_\theta^{-1} + \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1}\end{aligned}$$

MMSE estimator: Properties

- MMSE estimator is linear in data for jointly Gaussian distributions:

$$\hat{\boldsymbol{\theta}} = \mathbb{E}(\boldsymbol{\theta}|\mathbf{x}) = \mathbb{E}(\boldsymbol{\theta}) + \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}(\mathbf{x} - \mathbb{E}(\mathbf{x}))$$

- MMSE estimator has an additivity property for independent data sets
- MMSE estimator commutes over affine transformations: Consider the estimation of $\alpha = \mathbf{A}\boldsymbol{\theta} + \mathbf{b}$, where $\mathbf{A}$ and $\mathbf{b}$ are deterministic and known, then the MMSE estimator is

$$\hat{\alpha} = \mathbb{E}(\mathbf{A}\boldsymbol{\theta} + \mathbf{b}|\mathbf{x}) = \mathbf{A}\mathbb{E}(\boldsymbol{\theta}|\mathbf{x}) + \mathbf{b} = \mathbf{A}\hat{\boldsymbol{\theta}} + \mathbf{b} = \mathbb{E}(\alpha|\mathbf{x})$$

MMSE estimator: Vector process (1)

- Consider the estimation of random vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_N]^T$ from $\mathbf{x}$, and let $p(\mathbf{x}|\boldsymbol{\theta})$ and $p(\boldsymbol{\theta})$ be the conditional and prior PDFs
- θ_i can be estimated by viewing the other parameters as nuisance parameters , i.e., for the i^{th} element

$$\hat{\theta}_i = \int \theta_i p(\theta_i|\mathbf{x}) d\theta_i = \int \theta_i \left[\int \cdots \int p(\boldsymbol{\theta}|\mathbf{x}) d\theta_1 \ldots d\theta_{i-1} d\theta_{i+1} \ldots d\theta_p \right] d\theta_i$$
$$= \int \theta_i p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \quad \forall i = 1, 2, \ldots, p$$

- In vector form, we have the MMSE estimator as

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} \int \theta_1 p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \\ \int \theta_2 p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \\ \vdots \\ \int \theta_p p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \end{pmatrix} = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} = E[\boldsymbol{\theta}|\mathbf{x}]$$

MMSE estimator: Vector process (2)

- Minimum Bmse for $\theta_i \; \forall i = 1, 2, \ldots, p$

$$\begin{aligned}\text{Bmse}(\hat{\theta}_i) &= \mathbb{E}[(\theta_i - \hat{\theta}_i)^2] = \int (\theta_i - \hat{\theta}_i)^2 p(\mathbf{x}, \theta_i) d\theta_i d\mathbf{x} \\ &= \int \left[\int (\theta_i - \hat{\theta}_i)^2 p(\theta_i|\mathbf{x}) d\theta_i \right] p(\mathbf{x}) d\mathbf{x} = \int var(\theta_i|\mathbf{x}) p(\mathbf{x}) d\mathbf{x}\end{aligned}$$

- Substituting $p(\theta_1|\mathbf{x}) = \int \ldots \int p(\boldsymbol{\theta}|\mathbf{x}) d\theta_1 ... d\theta_{i-1} d\theta_{i+1} ... d\theta_p$

$$\begin{aligned}\text{Bmse}(\hat{\theta}_i) &= \int \left[\int (\theta_i - \mathbb{E}(\theta_i|\mathbf{x}))^2 p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \right] p(\mathbf{x}) d\mathbf{x} \\ &= \int \left[\mathbf{C}_{\theta|x} \right]_{ii} p(\mathbf{x}) d\mathbf{x}\end{aligned}$$

where

$$\mathbf{C}_{\theta|x} = \mathbb{E}_{\theta|x} \left[(\boldsymbol{\theta} - E(\theta|\mathbf{x})(\boldsymbol{\theta} - E(\theta|\mathbf{x})^T) \right]$$

Summary

Key points:

- Bayesian philosophy : Unknown parameter is *random* and the statistics are known *apriori*
- Minimum Mean Square Error (MMSE) estimator is the mean of the posterior PDF, and is the optimal estimator which minimizes the Bayesian mean square error (Bmse)
- Conditional independence : If X, Y, Z are conditionally independent, then

$$p(x, y|z) = p(x|z)p(y|z)$$

- When the measurements and unknown parameter are jointly Gaussian, then posterior and marginal PDFs are also Gaussian
- MMSE estimator is linear in data for jointly Gaussian Distributions, has an additivity property, and commutes over affine transformations

Next session:

- Bayes Risk, MAP and LMMSE estimators

Assignments

Solve:

- Problems 10.4, 10.5 and 10.8

Reading:

- Appendix 10A: Derivation of Conditional Gaussian PDF
- Kay-I, Section 10.3: Prior knowledge and estimation