No. : 1

Date : .

DS215 : Assignment 2

Aneesh Panchal

06-18-01-10-12-24-1-25223

Sol^n 1. If an efficient estimator exists, the max. likelihood method will produce it.

Efficient MVUE estimator is given by,

$$\frac{\partial \ln p(x;\theta)}{\partial\theta} = I(\theta)(g(x)-\theta) = I(\theta)(\hat{\theta}-\theta)$$

Now for MLE we have,

$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ p(x;\theta)$ which is equivalent to,

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ \ln p(x;\theta)$$

Now for maxima we know the condition $\nabla_\theta f(\theta) = 0$

$$\Rightarrow \frac{\partial \ln p(x;\theta)}{\partial\theta} = 0 = I(\theta)(\hat{\theta}-\theta)$$

& as we know, minimum variance for MVUE estimator is $I^{-1}(\theta)$ which can't be ∞ & hence $I(\theta) > 0$

$\therefore I(\theta)(\hat{\theta}-\theta) = 0$

$\Rightarrow \hat{\theta} = \theta$

Hence, proved.

Sol^n 2. Given MAP estimator is,

$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ p(\theta|x)$ where, $\theta, \hat{\theta} \in \mathbb{R}^p$

Cost function, $C(\varepsilon) = \begin{cases} 1 & \text{when } \|\varepsilon\| > \delta \\ 0 & \text{when } \|\varepsilon\| < \delta \end{cases}$

where, $\varepsilon = \theta - \hat{\theta}$, $\|\varepsilon\|^2 = \Sigma_p \varepsilon_i^2$ and $\delta \to 0$

So, Bayes risk R is given by,

$$R = E(C(\varepsilon)) = \iint C(\theta-\hat{\theta})\, p(x;\theta)\, dx\, d\theta$$

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

Aneesh
25223

Punk

Assuming x to be fixed, we need to minimize for $\hat{\theta}$ & hence we get,

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

As here we have to choose $\theta : \|\epsilon\|_2 > \delta$ i.e. $\theta : \|\theta-\hat{\theta}\|_2 > \delta$

$$g(\hat{\theta}) = \int_{\{\theta : \|\theta-\hat{\theta}\|_2 > \delta\}} p(\theta|x)\, d\theta$$

But as we know $p(\theta|x)$ is pdf wrt. variable θ

$\therefore \int_{-\infty}^{\infty} p(\theta|x)\, d\theta = 1$

hence, we get, $g(\hat{\theta}) = 1 - \int_{\{\theta : \|\theta-\hat{\theta}\|_2 < \delta\}} p(\theta|x)\, d\theta = 1 - I$

$g(\hat{\theta})$ can be minimized by maximizing I

& for given condition $\delta \to 0$, I is maximized by choosing.

$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, p(\theta|x)$

Hence, proved

Sol^n 3. Given signal model,

$$s[n] = \begin{cases} A & \text{when } 0 \le n \le M-1 \\ -A & \text{when } M \le n \le N-1 \end{cases}$$

$x[n] = s[n] + w[n]$ for $n = 0, 1, \ldots, N-1$ are observed values

$w[n]$ are WGN with variance σ^2

So, error function in least square sense is given by,

$$J(\theta) = \sum_{n=0}^{N-1} (x[n] - s[n])^2$$

$$J(A) = \sum_{n=0}^{M-1} (x[n]-A)^2 + \sum_{n=M}^{N-1} (x[n]+A)^2 \quad \text{—— (i)}$$

Aneesh
25223

Least square estimate (LSE) is given by,

$\hat{A} = \underset{A}{\operatorname{argmin}}\, J(A)$

$$\frac{\partial J(A)}{\partial A} = -2\sum_{n=0}^{M-1} (x[n]-A) + 2\sum_{n=M}^{N-1} (x[n]+A) = 0 \quad \text{—— (ii)}$$

$$-2\sum_{n=0}^{M-1} x[n] + 2MA + 2\sum_{n=M}^{N-1} x[n] + 2(N-M)A = 0$$

hence, $\hat{A} = \frac{1}{N}\left(\sum_{n=0}^{M-1} x[n] - \sum_{n=M}^{N-1} x[n]\right)$ ______ (iii)

Minimum Least square error is given by,

$J_{min} = J(\hat{A})$

using (i) we get,

$$J_{min} = \sum_{n=0}^{M-1} (x[n]-\hat{A})(x[n]-\hat{A}) + \sum_{n=M}^{N-1} (x[n]+\hat{A})(x[n]+\hat{A})$$

Now using (ii) we get,

$$J_{min} = \sum_{n=0}^{M-1} x[n](x[n]-\hat{A}) + \sum_{n=M}^{N-1} x[n](x[n]+\hat{A})$$

$$= \sum_{n=0}^{N-1} (x[n])^2 - \hat{A}\left(\sum_{n=0}^{M-1} x[n] - \sum_{n=M}^{N-1} x[n]\right)$$

Now using (iii) we get,

$$J_{min} = \sum_{n=0}^{N-1} (x[n])^2 - N\hat{A}^2$$

As given $w[n]$ is WGN with variance σ^2, using (iii) we get,

$$E(\hat{A}) = \frac{1}{N}[MA - (N-M)A] = A$$

$$Var(\hat{A}) = \frac{1}{N^2}\left[Var\left(\sum_{n=0}^{M-1} x[n]\right) + Var\left(\sum_{n=0}^{N-1} x[n]\right)\right]$$

Since $w[n]$ are WGN & hence are iid, we get,

$$Var(\hat{A}) = \frac{1}{N^2}(M\sigma^2 + (N-M)\sigma^2) = \frac{\sigma^2}{N}$$

Aneesh
25223

Hence, $\hat{A} \sim \mathcal{N}(A, \sigma^2/N)$

because in (iii) we can see $\hat{A}$ is linear function of $x[n]$

Sol^n 4. Given data $x[n] = A + w[n]$ for $n = 0, 1, \dots, N-1$

unknown parameter A have prior pdf, $f(A) = \begin{cases} \lambda e^{-\lambda A} & \text{when } A \geq 0 \\ 0 & \text{when } A < 0 \end{cases}$

here $\lambda > 0$ & $w[n]$ is WGN with variance σ^2 independent of A.

Maximum A Posteriori (MAP) estimator is given by,

$\hat{\theta} = \underset{\theta}{\arg\max}\ p(\theta|x) = \underset{\theta}{\arg\max}\ p(x|\theta)\, p(\theta)$

which is equivalent to,

$\hat{A} = \underset{A}{\arg\max}\ p(x|A)\, p(A) = \underset{A}{\arg\max}\ \log_e [p(x|A)\, p(A)]$

$p(x|A)\, p(A) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{\frac{-1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n]-A)^2\right\} \lambda \exp\{-\lambda A\}$ when $A \geq 0$

$= 0$ when $A < 0$ (because $p(A) = 0$ when $A < 0$)

& hence we get,

$\ln [p(x|A)\, p(A)] = -\frac{N}{2} \ln (2\pi\sigma^2) - \frac{1}{2\sigma^2} \left[\sum_{n=0}^{N-1} (x[n]-A)^2\right] - \lambda A + \ln \lambda$

for maximizing this we use gradient,

$\frac{\partial \ln [p(x|A)\, p(A)]}{\partial A} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n]-A) - \lambda = 0$

& hence, $\hat{A} = \frac{1}{N} \left[\sum_{n=0}^{N-1} x[n] - \sigma^2 \lambda\right]$

Also we know $\hat{A} = 0$ when $\hat{A} < 0$ & hence,

$\hat{A} = \max \left\{0, \frac{1}{N}\left[\sum_{n=0}^{N-1} x[n] - \sigma^2\lambda\right]\right\}$

which is required MAP estimator of A.

Sol^n 5. Given model is $x[n] = A + Bn + w[n]$, $-M \leq n \leq M$

where $w[n]$ is WGN with variance σ^2

prior knowledge,

$\begin{bmatrix} A \\ B \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} A_0 \\ B_0 \end{bmatrix}, \begin{bmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{bmatrix}\right)$

Aneesh
25223

This can be seen as Bayesian linear model ($x[n] = H\theta + w[n]$)

where, $\theta = \begin{bmatrix} A \\ B \end{bmatrix}$ & $H = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ -M & -M+1 & \cdots & M \end{bmatrix}^T$

In Bayesian linear model, prior knowledge, $\theta \sim \mathcal{N}(\mu_\theta, C_\theta)$

w is WGN with distribution $\mathcal{N}(0, C_w)$

MMSE is given by $y \equiv \theta$ with conditions/equations,

$E(x) = H\mu_\theta$

$E(y) = \mu_\theta$

$C_{xx} = HC_\theta H^T + C_w$

$C_{xy} = C_\theta H^T$

& since x & θ are jointly Gaussian we have,

$$\hat{\theta} = E(\theta|x) = \mu_\theta + C_\theta H^T (HC_\theta H^T + C_w)^{-1}(x - H\mu_\theta) = \mu_\theta + (C_\theta^{-1} + H^T C_w^{-1} H)^{-1} H^T C_w^{-1}(x - H\mu_\theta)$$

for given question we have,

$$\mu_\theta = \begin{bmatrix} A_0 \\ B_0 \end{bmatrix}, \quad C_\theta = \begin{bmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{bmatrix}, \quad C_w = \sigma^2 I = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$$ & H is defined before.

$$H^T C_w^{-1} H = \frac{1}{\sigma^2} H^T H = \frac{1}{\sigma^2}\begin{bmatrix} n & 0 \\ 0 & \sum_{i=-M}^{M} i^2 \end{bmatrix}$$ where $n = 2M+1$

& let $\sum_{i=-M}^{M} i^2 = N$

$$H^T C_w^{-1} H = \frac{1}{\sigma^2}\begin{bmatrix} n & 0 \\ 0 & N \end{bmatrix}$$

$$(C_\theta^{-1} + H^T C_w^{-1} H)^{-1} = \begin{bmatrix} \left(\frac{1}{\sigma_A^2} + \frac{n}{\sigma^2}\right) & 0 \\ 0 & \left(\frac{1}{\sigma_B^2} + \frac{N}{\sigma^2}\right) \end{bmatrix}^{-1} = \begin{bmatrix} \left(\frac{1}{\sigma_A^2} + \frac{n}{\sigma^2}\right)^{-1} & 0 \\ 0 & \left(\frac{1}{\sigma_B^2} + \frac{N}{\sigma^2}\right)^{-1} \end{bmatrix}$$

$$H^T C_w^{-1}(x - H\mu_\theta) = \frac{1}{\sigma^2}\begin{bmatrix} \sum_{i=-M}^{M} x[i] - nA_0 \\ \sum_{i=-M}^{M} i\,x[i] - NB_0 \end{bmatrix}$$

Now, for $\hat{\theta} = \mu_\theta + (C_\theta^{-1} + H^T C_w^{-1} H)^{-1} H^T C_w^{-1}(x - H\mu_\theta)$ we get,

$$\hat{A} = A_0 + \left(\frac{1}{\sigma^2}\left(\sum_{i=-M}^{M} x[i] - nA_0\right)\right)\left(\frac{1}{\sigma_A^2} + \frac{n}{\sigma^2}\right)^{-1}$$

$$\hat{B} = B_0 + \left(\frac{1}{\sigma^2}\left(\sum_{i=-M}^{M} i\,x[i] - NB_0\right)\right)\left(\frac{1}{\sigma_B^2} + \frac{N}{\sigma^2}\right)^{-1}$$

which are required MMSE estimators of A & B.

Aneesh
25223

As we know,

$C_{\theta|x} = C_\theta - C_\theta H^T (HC_\theta H^T + C_w)^{-1} HC_\theta$

which can also be written as,

$C_{\theta|x} = (C_\theta^{-1} + H^T C_w^{-1} H)^{-1}$

or simply, $C_{\theta|x}^{-1} = C_\theta^{-1} + H^T C_w^{-1} H$

As we know Bayesian MSE is given by,

$BMSE(\hat{A}) = \int var(A|x)\, f(x)\, dx$

for the given question variables A & B are bivariate gaussian & independent hence we can see,

$BMSE(\hat{\theta}) = C_{\theta|x}$ [redacted] $(1-\rho^2) \int f(x)\, dx = C_{\theta|x}$ (because $\rho = 0$ & $\int f(x)\,dx = 1$)

So, minimum Bayesian MSE is given by

$BMSE(\hat{\theta}) = C_{\theta|x} = (C_\theta^{-1} + H^T C_w^{-1} H)^{-1}$

whose value is derived before hand & hence,

$BMSE(\hat{A}) = \left(\frac{1}{\sigma_A^2} + \frac{n}{\sigma^2}\right)^{-1}$ where $n = 2M+1$

Aneesh
25223

$BMSE(\hat{B}) = \left(\frac{1}{\sigma_B^2} + \frac{N}{\sigma^2}\right)^{-1}$ where $N = \sum_{i=-M}^{M} i^2$

Parameter A will benefit most from prior knowledge

As we know $\frac{n}{\sigma^2} < \frac{N}{\sigma^2}$ & hence $\left(\frac{n}{\sigma^2}\right)^{-1} > \left(\frac{N}{\sigma^2}\right)^{-1}$

$\therefore$ as we add terms σ_A^{-2} & σ_B^{-2} in respective BMSE we will see significant (larger) reduction in $BMSE(\hat{A})$ than compare to $BMSE(\hat{B})$.

Hence, from prior knowledge, parameter A will be benefitted most.