DS 215

IDS ASSIGNMENT -2

Kaniki Bhikshapathi
24489

Question - 1

Requires to prove : "If an efficient estimator exists, the maximum likelihood method will produce it."

MLE is found by maximizing p or $\ln p$

i.e $\frac{\partial}{\partial \theta} \ln(p, \theta) = 0$

Assuming a scalar parameter, if efficient estimator exist then we have

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

$$I(\theta)(g(x) - \theta) = 0$$

$$[I(\theta)]_{(i,j)} = E\left[\frac{\partial^2 \ln p}{\partial x_i \partial x}\right]$$

& $I^{-1}(\theta)$ is the covarience of $g(x)$

multiply $I^{-1}(\theta)$ on both sides of ①

$$I^{-1}(\theta) I(\theta)(g(x) - \theta) = 0$$

$$(g(x) - \theta) = 0 \Rightarrow g(x) = \hat{\theta} = \theta$$

Hence $\hat{\theta}_{MLE}$ is $\hat{\theta}_{MVUE}$ (MVUE estimator) its exist.

Question-2

given $\hat{\theta} = \arg\max_{\theta} p(\theta/x)$

Prove

Bayes risk for cost function $c(\varepsilon) = \begin{cases} 1 & \text{when } \|\varepsilon\| > \delta \\ 0 & \text{when } \|\varepsilon\| \ \delta \end{cases}$

where $\theta, \hat{\theta} \in R^p$ $\varepsilon = \theta - \hat{\theta}$, $\|\varepsilon\|^2 = \sum_{i=1}^{p} \varepsilon_i^2$ & $\delta \to 0$

$$\text{Bayes risk} = \int_{x \in R^p} \int_{\theta \in R^p} c(\hat{\theta} - \theta)\, p(x, \theta)\, d\theta\, dx$$

$$= \int_{x \in R^p} \int_{\theta \in R^p} c(\hat{\theta} - \theta)\, p(\theta/x)\, p(x)\, d\theta\, dx \qquad \{\because p(x,\theta) = p(\theta/x)\, p(x)\}$$

$$= \int_{x \in R^p} \boxed{\int_{\theta \in R^p} c(\hat{\theta} - \theta)\, p(\theta/x)\, d\theta}\, dx$$

Let this be R

$$R = \int_{\theta \in R^p} c(\hat{\theta} - \theta)\, p(\theta/x)\, d\theta > 0 \text{ for all } x,$$

so rather than minimizing total Bayes risk, we minimize R.

$$\int_{\theta \in R^p} c(\hat{\theta} - \theta)\, p(\theta/x)\, d\theta = \int_{\theta = \|\theta - \hat{\theta}\| > \delta} c(\hat{\theta} - \theta)\, p(\theta/x)\, d\theta + \int_{\theta = \|\theta - \hat{\theta}\| \le \delta} c(\hat{\theta} - \theta)\, p(\theta/x)\, d\theta$$

$= 0$

(we know $c(\hat{\theta} - \theta) = 0$ for $\|\theta - \hat{\theta}\| < \delta$)

$$R = \int_{\|\theta - \hat{\theta}\| > \delta} p(\theta/x)\, d\theta = 1 - \int_{\|\theta - \hat{\theta}\| < \delta} p(\theta/x)\, d\theta$$

To minimize Bayes risk $\to$ minimize R

To minimize $R \to$ maximize $\int_{\|\theta - \hat{\theta}\| < \delta} p(\theta/x)\, d\theta$

achived by choosing $\boxed{\hat{\theta} = \arg\max_{\theta} (p(\theta/x))}$

3) $$x[N] = \begin{array}{l} m\left\{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ -1 \\ -1 \\ \vdots \\ -1 \end{bmatrix}\right. \\ N-m \end{array} A + W \quad ; \quad H = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ -1 \\ \vdots \\ -1 \end{bmatrix}$$

$$\hat{A}_{LSE} = (H^T H)^{-1} H^T x$$

$$= (N)^{-1} \left[\sum_{i=0}^{m-1} x_i - \sum_{i=m}^{N-1} x_i \right] \quad -①$$

$$LS_{error} = \sum_{i=0}^{m-1} (x_i - \hat{A})^2 + \sum_{m}^{N-1} (x_i + \hat{A})^2$$

$$= \sum_{i=0}^{m-1} (x_i^2 + \hat{A}^2 - 2\hat{A}x_i) + \sum_{i=m}^{N-1} (x_i^2 + \hat{A}^2 + 2\hat{A}x_i)$$

$$= \sum_{i=0}^{N-1} x_i^2 + N\hat{A}^2 - 2\hat{A}\left(\sum_{i=0}^{m-1} x_i - \sum_{i=m}^{N-1} x_i\right)$$

$$LS_{error} = \sum_{i=0}^{N-1} x_i^2 - N(\hat{A})^2$$

$$E[\hat{A}_{LSE}] = \frac{1}{N}\left(\sum_{i=0}^{m-1} E[x_i] - \sum_{i=m}^{N-1} E[x_i]\right) = \frac{1}{N}(NA)$$

$$= A$$

$$Var(\hat{A}_{LSE}) = \frac{1}{N^2} \sum_{i=0}^{N} var(x_i) = \frac{\sigma^2}{N}$$

$$\hat{A}_{LSE} \sim N(A, \sigma^2/N) \rightarrow \text{PDF of LSE}$$

Question- 4

MAP estimator of A → maximise posterior PDF

$$f(A/x) = \frac{f(x|A)\,f(A)}{f(x)}$$

$x[N] = A + w[N]$; $w[N] \sim N(0, \sigma^2)$

$x[N]$ are independent and normally distributed with mean A & varience σ^2

$$f(x/A) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x[N]-A)^2}{2\sigma^2}\right)$$

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

The prior distribution is $P(A) = \lambda e^{-\lambda A}$ $A>0$

and posterior distribution is

$$g(A) = P(A/x) \propto P(x/A)\,P(A)$$

$$= P(A/x) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[N]-A]^2\right)\lambda e^{-\lambda A}$$

$$\frac{\partial}{\partial A}\left(\ln p(A/x)\right) = 0$$

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

$$\frac{NA}{\sigma^2} = \sum_{n=0}^{N-1}\frac{x[n]}{\sigma^2} - \lambda \Rightarrow \hat{A}_{map} = \frac{\sigma^2}{N}\left(\frac{N}{\sigma^2}\bar{x} - \lambda\right)$$

$$\hat{A}_{map} = \bar{x} - \frac{\sigma^2}{N}\lambda$$

given PDF $f(A)$ defined only for $A \geq 0$

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

{as $A_{map} \geq 0$}

5) $x[n] = A + Bn + w[n] \qquad n \in [-u, u]$

It is Bayesian general linear model as the prior are jointly gaussian

$$x = \underbrace{\begin{bmatrix} 1 & -u \\ 1 & -u+1 \\ 1 & \vdots \\ \vdots & \vdots \\ 1 & u \end{bmatrix}_{(2u+1)\times 2}}_{H} \underbrace{\begin{bmatrix} A \\ B \end{bmatrix}_{2\times 1}}_{\theta} + \underbrace{w}_{N(0,\sigma^2 I_{2u+1})}$$

$$\hat{\theta} = \mu_\theta + \left(C_\theta^{-1} + H^T C_w^{-1} H\right)^{-1} H^T C_w^{-1} (x - H\mu_\theta)$$

$$\mu_\theta = \begin{bmatrix} A_0 \\ B_0 \end{bmatrix}, \quad C_\theta = \begin{bmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{bmatrix}, \quad C_w = \sigma^2 I_{2u+1}$$

$$\hat{\theta} = \begin{bmatrix} A_0 \\ B_0 \end{bmatrix} + \left(\begin{bmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{bmatrix} + \sigma^{-2} H^{-1} H \right)^{-1} \sigma^{-2} H^T (x - H\mu_0)$$

$$= \begin{bmatrix} A_0 \\ B_0 \end{bmatrix} + \frac{1}{\sigma^2} \left(\begin{bmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{bmatrix} + \sigma^{-2} \begin{bmatrix} 2u+1 & 0 \\ 0 & \frac{u(u+1)(2u+1)}{3} \end{bmatrix} \right)$$

$$= \begin{bmatrix} A_0 \\ B_0 \end{bmatrix} + \frac{1}{\sigma^2} \left(\begin{bmatrix} 2u+1+\sigma_A^{-2} & 0 \\ 0 & \frac{u(u+1)(2u+1)}{3\sigma^2} + \sigma_B^{-2} \end{bmatrix} \right)^{-1} H^T (x - H\mu_0)$$

$$\begin{bmatrix} \hat{A} \\ \hat{B} \end{bmatrix} = \begin{bmatrix} A_0 \\ B_0 \end{bmatrix} + \frac{1}{\sigma^2} \left(\begin{bmatrix} 2u+1+\sigma_A^{-2} & 0 \\ 0 & \frac{u(u+1)(2u+1)}{3\sigma^2} + \sigma_B^{-2} \end{bmatrix} \right)^{-1} \begin{bmatrix} \sum_{i=u}^{u} x_i - (2u+1)A_0 \\ \sum_{i=-u}^{u} i x_i - \frac{u(u+1)(2u+1)}{3} B_0 \end{bmatrix}$$

$$\hat{A} = A_0 + \frac{1}{\sigma^2}\left(\frac{\sum_{i=-u}^{u} x_i - (2u+1)A_0}{\frac{2u+1}{\sigma^2} + \frac{1}{\sigma_A^2}}\right)$$

$$\hat{B} = B_0 + \frac{1}{\sigma^2}\left(\frac{\sum_{i=-u}^{u} i x_i - \frac{u(u+1)(2u+1)B_0}{3}}{\frac{u(u+1)(2u+1)}{3\sigma^2} + \frac{1}{\sigma_B^2}}\right)$$

Now consider Bayesian MSE for parameter θ

$$BMSE = \iint_{x\,\theta} (\theta - \hat{\theta})^2 P(\theta/x)\, P(x)\, d\theta\, dx$$

$$= \iint_{x\,\theta} \left(\theta - E(\theta/x)\right)^2 P(\theta/x)\, P(x)\, d\theta\, dx$$

$$= \int_x var(\theta/x)\, P(x)\, dx = var(\theta/x) \int_x P(x)\, dx$$

$$BMSE = var(\theta/x)$$

By calculating the covarience matrix $\left[C_{A,B}^{-1} + H^T C_w^{-1} H\right]^{-1}$

$$= \begin{bmatrix} \frac{1}{\sigma_A^{-2} + \frac{2u+1}{\sigma^2}} & 0 \\ 0 & \frac{1}{\sigma_B^{-2} + \frac{u(u+1)(2u+1)}{3\sigma^2}} \end{bmatrix}$$

$$BMSE(\hat{A}) = var(A/x) = \frac{1}{\frac{1}{\sigma_A^2} + \frac{2u+1}{\sigma^2}}$$

$$BMSE(\hat{B}) = var(B/x) = \frac{1}{\frac{1}{\sigma_B^2} + \frac{u(2u+1)(u+1)}{3\sigma^2}}$$

we can see that $\hat{x}$ will benefit more from prior knowledge

as $\begin{cases} BMSE(\hat{B}) \rightarrow \text{denominator} \rightarrow \frac{u(u+1)(2u+1)}{3\sigma^2} \\ BMSE(\hat{A}) \rightarrow \text{denominator} \rightarrow \frac{2u+1}{\sigma^2} \end{cases}$ (effect of prior is lower compared with $\hat{A}$)