Estimation Notes 3 (Problems 3.10 - 3.20) pp.65 - 66

Problem 3.10

By using (3.23) prove that the Fisher information matrix is **positive semidefinite** for all θ. In practice, we **assume** it to be positive definite and hence invertible, although this is not always the case. Consider the data model in Problem 3.3 with the modification that r is unknown. Find the Fisher information matrix for $\boldsymbol{\theta} = [A, r]^{\mathrm{T}}$. Are there any values of $\boldsymbol{\theta}$ for which $\mathbf{I}(\boldsymbol{\theta})$ is not positive definite?

Answer:

1) **positive semidefinite** implies that (we only talk about real matrix this time)

$$\mathbf{M} \text{ is positive definite} \iff \mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{x} \geq 0 \quad \text{for all} \quad \mathbf{x} \in \mathbb{R}^n$$

where $\mathbf{M}$ is a $n \times n$ symmetric matrix, $\mathbf{x}^{\mathrm{T}}$ is the transpose of $\mathbf{x}$

2) so we have

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = E\left\{\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i}\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_j}\right\}$$

$$\Rightarrow \mathbf{I}(\boldsymbol{\theta}) = E\left\{\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}(\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^{\mathrm{T}}\right\}$$

$$\Rightarrow \mathbf{a}^{\mathrm{T}}\mathbf{I}(\boldsymbol{\theta})\mathbf{a} = E\left\{\mathbf{a}^{\mathrm{T}}\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}(\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^{\mathrm{T}}\mathbf{a}\right\} = E\left\{(\mathbf{a}^{\mathrm{T}}\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^2\right\} \geq 0$$

where for all $\mathbf{a}$. Therefore, $\mathbf{I}(\boldsymbol{\theta})$ is positive semidefinite for all $\boldsymbol{\theta}$.

3) $\boldsymbol{\theta} = [Ar]^{\mathrm{T}}$ and using (3.21)

$$[\boldsymbol{I}(\boldsymbol{\theta})]_{ij} = -E\left\{\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right\} \quad i = 1,\ldots,p; j = 1,\ldots,p \tag{3.21}$$

since

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = [A, Ar, \ldots, Ar^{N-1}]^{\mathrm{T}}$$

$$\Rightarrow \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial A} = [1, r, r^2, \ldots, r^{N-1}]^{\mathrm{T}}, \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial r} = A[0, 1, 2r, 3r^2, \ldots, (N-1)r^{N-2}]^{\mathrm{T}}$$

then we can obtain that

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = \frac{1}{\sigma^2}\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i}\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_j}$$

we know that $\theta_1 = A$ and $\theta_2 = r$, so

$$[\mathbf{I}(\boldsymbol{\theta})]_{11} = \frac{1}{\sigma^2}(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_1})^{\mathrm{T}}\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_1} = \frac{1}{\sigma^2}(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial A})^{\mathrm{T}}\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial A}$$

$$= \frac{1}{\sigma^2}[1, r, r^2, \ldots, r^{N-1}] \cdot [1, r, r^2, \ldots, r^{N-1}]^{\mathrm{T}} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1} r^n$$

$$[\mathbf{I}(\boldsymbol{\theta})]_{12} = \frac{1}{\sigma^2}[1, r, r^2, \ldots, r^{N-1}] \cdot A[0, 1, 2r, 3r^2, \ldots, (N-1)r^{N-2}]^{\mathrm{T}} = \frac{A}{\sigma^2}\sum_{n=0}^{N-1} nr^{2n-1}$$

$$[\mathbf{I}(\boldsymbol{\theta})]_{21} = \frac{1}{\sigma^2}A[0, 1, 2r, 3r^2, \ldots, (N-1)r^{N-2}] \cdot [1, r, r^2, \ldots, r^{N-1}]^{\mathrm{T}} = \frac{A}{\sigma^2}\sum_{n=0}^{N-1} nr^{2n-1}$$

$$[\mathbf{I}(\boldsymbol{\theta})]_{22} = \frac{1}{\sigma^2}A[0, 1, 2r, 3r^2, \ldots, (N-1)r^{N-2}] \cdot A[0, 1, 2r, 3r^2, \ldots, (N-1)r^{N-2}]^{\mathrm{T}} = \frac{A^2}{\sigma^2}\sum_{n=0}^{N-1} n^2r^{2n-2}$$

finally we have

$$\boldsymbol{I}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{\sigma^2}\sum_{n=0}^{N-1} r^n & \frac{A}{\sigma^2}\sum_{n=0}^{N-1} nr^{2n-1} \\ \frac{A}{\sigma^2}\sum_{n=0}^{N-1} nr^{2n-1} & \frac{A^2}{\sigma^2}\sum_{n=0}^{N-1} n^2r^{2n-2} \end{bmatrix}$$

If $A = 0$, $\boldsymbol{I}(\boldsymbol{\theta})$ is not positive definitive. Clearly, in this case there is no information in the data about r.

Problem 3.11

For a 2×2 Fisher information matrix

$$\boldsymbol{I}(\boldsymbol{\theta}) = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

which is positive definite, show that

$$[\boldsymbol{I}^{-1}(\boldsymbol{\theta})]_{11} = \frac{c}{ac-b^2} \geq \frac{1}{a} = \frac{1}{[\boldsymbol{I}(\boldsymbol{\theta})]_{11}}$$

What does this say about estimating a parameter when a second parameter is either known or unknown? When does quality hold and why?

Answer:

1) since $\mathbf{I}(\boldsymbol{\theta})$ is positive definite, we can obtain that $a > 0$, $c > 0$ and $\det(\mathbf{I}(\boldsymbol{\theta})) = ac - b^2 > 0$. Then

$$\mathbf{I}^{-1}(\boldsymbol{\theta}) = \frac{1}{ac-b^2} \begin{bmatrix} c & -b \\ -b & a \end{bmatrix}$$

$$\Rightarrow [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{11} = \frac{c}{ac-b^2}$$

$$\Rightarrow \frac{c}{ac-b^2} - \frac{1}{a} = \frac{ac-(ac-b^2)}{a(ac-b^2)} = \frac{b^2}{a(ac-b^2)} \geq 0$$

$$\Rightarrow [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{11} = \frac{c}{ac-b^2} \geq \frac{1}{a} = \frac{1}{[\mathbf{I}(\boldsymbol{\theta})]_{11}}$$

2) Thus, the CRLB is almost always increased when we estimate additional parameters. Equality holds if and only if $b = 0$ or the Fisher information matrix is decoupled, i.e., it is diagonal. In this case the additional parameter does not affect the CRLB.

Problem 3.12

Prove that

$$[\mathbf{I}^{-1}(\boldsymbol{\theta})]_{ii} \geq \frac{1}{[\mathbf{I}(\boldsymbol{\theta})]_{ii}}$$

This generalizes the result of Problem 3.11. Additionally, it provides another lower bound on the variance, although it is usually not attainable. Under what conditions will the new bound be achieved? Hint: Apply the Cauchy-Schwarz inequality to $\mathbf{e}_i^{\mathrm{T}}\sqrt{\mathbf{I}(\boldsymbol{\theta})}\sqrt{\mathbf{I}^{-1}(\boldsymbol{\theta})}\mathbf{e}_i$, where $\mathbf{e}_i$ is the vectors of all zeros except for a 1 as the ith element. $\sqrt{\mathbf{A}}$ is the square root of a positive definite matrix $\mathbf{A}$. More precisely, $\sqrt{\mathbf{A}}$ is defined to be the matrix with the same eigenvectors as $\mathbf{A}$ but whose eigenvalues are the square roots of those of $\mathbf{A}$.

Answer:

the Cauchy-Schwarz inequality is

$$|\mathbf{uv}|^2 \leq |\mathbf{u}|^2 \cdot |\mathbf{v}|^2$$

let $\mathbf{u} = \mathbf{e}_i^{\mathrm{T}}\sqrt{\mathbf{I}(\boldsymbol{\theta})}$ and $\mathbf{v} = \sqrt{\mathbf{I}^{-1}(\boldsymbol{\theta})}\mathbf{e}_i$, then we have

$$1^2 = (\mathbf{e}_i^{\mathrm{T}}\sqrt{\mathbf{I}(\boldsymbol{\theta})}\sqrt{\mathbf{I}^{-1}(\boldsymbol{\theta})}\mathbf{e}_i)^2 \leq \mathbf{e}_i^{\mathrm{T}}\mathbf{I}(\boldsymbol{\theta})\mathbf{e}_i\mathbf{e}_i^{\mathrm{T}}\mathbf{I}^{-1}(\boldsymbol{\theta})\mathbf{e}_i = [\mathbf{I}(\boldsymbol{\theta})]_{ii}[\mathbf{I}^{-1}(\boldsymbol{\theta})]_{ii}$$

$$\Rightarrow [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{ii} \geq \frac{1}{[\mathbf{I}(\boldsymbol{\theta})]_{ii}}$$

New bound achieved when an efficient estimation and $\mathbf{I}(\boldsymbol{\theta})$ is **diagonal**.

Problem 3.13

Consider a generalization of the line fitting problem as described in Example 3.7, termed **polynomial** or **curve fitting**. The data model is

$$x[n] = \sum_{k=0}^{p-1} A_k n^k + w[n]$$

for $n = 0, 1, \ldots, N-1$. As before, $w[n]$ is WGN with variance σ^2. It is desired to estimate $\{A_0, A_1, \ldots, A_{p-1}\}$. Find the Fisher information matrix for this problem.

Answer:

We know that

$$[I(\boldsymbol{\theta})]_{ij} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\frac{\partial s[n;\boldsymbol{\theta}]}{\partial\theta_i}\frac{\partial s[n;\boldsymbol{\theta}]}{\partial\theta_j} \tag{3.33}$$

$$\Rightarrow [I(\boldsymbol{\theta})]_{ij} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(\frac{\partial}{\partial\theta_i}\sum_{k=0}^{p-1}A_k n^k)(\frac{\partial}{\partial\theta_j}\sum_{k=0}^{p-1}A_k n^k)$$

Problem 3.14

For the data model in Example 3.11 consider the estimator $\hat{\sigma_A^2} = (\hat{A})^2$, where $\hat{A}$ is the sample mean. Assume we observe a given data set in which the realization of the random variable A is the value A_0. Show that $\hat{A} \to A_0$ as $N \to \infty$ by verifying that

$$E(\hat{A}|A = A_0) = A_0$$
$$var(\hat{A}|A = A_0) = \frac{\sigma^2}{N}$$

Hence, $\hat{\sigma_A^2} \to \hat{A_0^2}$ as $N \to \infty$ for the given realization $A = A_0$. Next find the variance of $\hat{\sigma_A^2}$ as $N \to \infty$ by determining $var(A^2)$, where $A \sim \mathcal{N}(0, \sigma_A^2)$, and compare it to the CRLB. Explain why σ_A^2 cannot be estimated without error even for $N \to \infty$

Answer:

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

we condition the mean and variance on A, then we can regard A as the observed value A_0. Hence, from Example 3.11, we have

$$E\{\hat{A}|A = A_0\} = A_0$$
$$var\{\hat{A}|A = A_0\} = \frac{\sigma^2}{N}$$

and since $\sigma^2/N \to 0$ as $N \to \infty$, $\hat{A} \to A_0$

Now consider A as a random variable as $N \to \infty$

$$var\{\hat{\sigma_A^2}\} = var\{\hat{A}^2\} \to var\{A^2\}$$

But

$$var\{A^2\} = E\{A^4\} - E^2\{A^2\} = 3\sigma_A^4 - \sigma_A^4 = 2\sigma_A^4$$

So that

$$var\{\hat{\sigma_A^2}\} \to 2\sigma_A^4$$

which is approach the CRLB as $N \to \infty$. σ_A^2 cannot be estimated without error even as $N \to \infty$, although we can modify the noise effect by averaging ($\hat{A} \to A_0$), we cannot reduce the random ( ) of A. This is because we have only one ( ) of A. Since $\hat{\sigma_A^2}$ is the ( ) of $\hat{A}$, it Will ( )( ) the ( ).

Problem 3.15

Consider a generalization of the line fitting problem as

Answer:

We know that