3. Fisher Information Matrix and Cramér–Rao Bound

3.1 Scalar Estimation

A scalar parameter θ is estimated by an estimator $\hat{\theta}(x^{(1)}, \dots, x^{(N)})$ based on N statistics $x^{(1)}, \dots, x^{(N)}$ with the joint PDF $f_{x^{(1)},\dots,x^{(N)}}(x^{(1)}, \dots, x^{(N)}; \theta)$. Note that $\hat{\theta}(x^{(1)}, \dots, x^{(N)})$ is a random variable as it depends on the statistics $x^{(1)}, \dots, x^{(N)}$. The score function is defined as

$$g(x^{(1)}, \dots, x^{(N)}; \theta) = \frac{\partial \ell(x^{(1)}, \dots, x^{(N)}; \theta)}{\partial \theta} = \frac{\partial \log\left(L(x^{(1)}, \dots, x^{(N)}; \theta)\right)}{\partial \theta},$$

where $L(x^{(1)}, \dots, x^{(N)}; \theta)$ is the likelihood function. Assuming that $L(x^{(1)}, \dots, x^{(N)}; \theta)$ satisfies the regularity conditions (here for $L(x; \theta)$)

1) $L(x; \theta) > 0 \; \forall x \in \mathbb{X}, \; \forall \theta \in \Theta$,

2) $L(x; \theta)$ differentiable with respect to θ on Θ,

3) $\int_{\mathbb{X}} \frac{\partial}{\partial \theta} L(x; \theta) \, \mathrm{d}x = \frac{\partial}{\partial \theta} \int_{\mathbb{X}} L(x; \theta) \, \mathrm{d}x$,

the variance of the score function

$$\mathrm{Var}\left[g(x^{(1)}, \dots, x^{(N)}; \theta)\right] = I_{\mathrm{F}}^{(N)}(\theta)$$

is the Fisher information $I_{\mathrm{F}}^{(N)}(\theta)$ and serves as a metric for the information on θ contained in the N statistics $x^{(1)}, \dots, x^{(N)}$. If the estimator is unbiased, the inverse of the Fisher information gives the Cramér–Rao bound (CRB) on the variance of the estimator $\hat{\theta}(x^{(1)}, \dots, x^{(N)})$.

a) Assuming that $L(x^{(1)}, \dots, x^{(N)}; \theta)$ satisfies the regularity conditions, show that N i.i.d. statistics $x^{(1)}, \dots, x^{(N)}$ carry N times as much information about θ as one statistic $x^{(1)}$. This means that for N i.i.d. statistics

$$I_{\mathrm{F}}^{(N)}(\theta) = N I_{\mathrm{F}}^{(1)}(\theta),$$

where $I_{\mathrm{F}}^{(i)}(\theta)$ is the Fisher information of i statistics.

b) Consider N i.i.d. statistics $x^{(1)}, \dots, x^{(N)}$, where $x^{(i)}$ is

1) uniformly distributed on the interval $[0, \theta]$ with unknown $\theta \in (0, \infty)$, i.e., $x^{(i)} \sim \mathcal{U}(0, \theta)$,

2) Gaussian with unknown mean $\theta \in (-\infty, \infty)$ and known variance σ^2, i.e., $x^{(i)} \sim \mathcal{N}(\theta, \sigma^2)$,

3) binomially distributed with unknown success probability $\theta \in (0, 1)$ and known number of trials K, i.e., $x^{(i)} \sim \mathcal{B}(K, \theta)$.

Determine, if applicable, the CRB and the Fisher information for an *unbiased* estimation of the parameter θ using N i.i.d. statistics $x^{(1)}, \dots, x^{(N)}$ for each distribution. Justify your answer.

3.2 Multivariate Estimation I

A parameter $\boldsymbol{\theta} \in \mathbb{R}^M$ is estimated using an estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}) \in \mathbb{R}^M$ based on N statistics $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}$ with the joint PDF $f_{\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(N)}}(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}; \boldsymbol{\theta})$. The score function is defined as

$$\boldsymbol{g}(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}; \boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial \log\left(L(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}; \boldsymbol{\theta})\right)}{\partial \boldsymbol{\theta}},$$

where $L(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}; \boldsymbol{\theta})$ is the likelihood function. Assuming that $L(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}; \boldsymbol{\theta})$ satisfies the regularity conditions (cf. problem 3.1), the covariance matrix of the score function

$$\mathrm{Cov}\left[\boldsymbol{g}(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}; \boldsymbol{\theta})\right] = \boldsymbol{I}_{\mathrm{F}}^{(N)}(\boldsymbol{\theta})$$

is known as the Fisher information matrix. If $\hat{\boldsymbol{\theta}}(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)})$ is unbiased, the CRB is given as

$$\mathrm{Cov}\left[\hat{\boldsymbol{\theta}}(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)})\right] \geq \left(\boldsymbol{I}_{\mathrm{F}}^{(N)}(\boldsymbol{\theta})\right)^{-1}.$$

Consequently, the CRB for the estimate of the m-th entry of $\boldsymbol{\theta}$ is given as

$$\mathrm{Var}\left[\hat{\theta}_m(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)})\right] \geq \left[\left(\boldsymbol{I}_{\mathrm{F}}^{(N)}(\boldsymbol{\theta})\right)^{-1}\right]_{m,m}.$$

In the following, we investigate the CRB for the estimation of the channel gain $\boldsymbol{h}$ of the transmission system discussed in problem 2.1, i.e., of

$$\boldsymbol{x}^{(i)} = \boldsymbol{h} s^{(i)} + \boldsymbol{n}^{(i)}.$$

As the noise vectors $\boldsymbol{n}^{(1)}, \ldots, \boldsymbol{n}^{(N)}$ are assumed to be i.i.d. with $\boldsymbol{n}^{(i)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{C} = \sigma^2 \boldsymbol{I}_M)$, the statistics $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}$ are independent and Gaussian. In addition, the regularity conditions on $L(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}; \boldsymbol{\theta})$ are satisfied. The ML estimator $\hat{\boldsymbol{h}}_{\mathrm{ML}}$ for $\boldsymbol{h}$ derived in problem 2.1b) is used.

a) Is the ML estimator $\hat{\boldsymbol{h}}_{\mathrm{ML}}$ unbiased?

b) Why does

$$\boldsymbol{I}_{\mathrm{F}}^{(N)}(\boldsymbol{h}) = N \boldsymbol{I}_{\mathrm{F}}^{(1)}(\boldsymbol{h})$$

not hold in this case?

c) Determine the Fisher information matrix

$$\boldsymbol{I}_{\mathrm{F}}^{(N)}(\boldsymbol{h}) = \mathrm{Cov}\left[\boldsymbol{g}(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}; \boldsymbol{h})\right].$$

d) Determine a lower bound on the variance of the m-th entry in $\hat{\boldsymbol{h}}_{\mathrm{ML}}$, i.e., the CRB for the estimate of the m-th channel gain.

3.3 Advanced: Multivariate Estimation II

A parameter $\boldsymbol{\theta} \in \mathbb{R}^M$ is estimated by an unbiased estimator $\hat{\boldsymbol{\theta}}(\mathsf{x}) \in \mathbb{R}^M$ based on one statistic x with the PDF $f_{\mathsf{x}}(\boldsymbol{x};\boldsymbol{\theta})$. The covariance matrix of the estimator $\hat{\boldsymbol{\theta}}(\mathsf{x})$ is given as

$$\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} = \mathrm{E}\left[\left(\hat{\boldsymbol{\theta}}(\mathsf{x}) - \mathrm{E}\left[\hat{\boldsymbol{\theta}}(\mathsf{x})\right]\right)\left(\hat{\boldsymbol{\theta}}(\mathsf{x}) - \mathrm{E}\left[\hat{\boldsymbol{\theta}}(\mathsf{x})\right]\right)^{\mathrm{T}}\right].$$

In the following, assume that $L(\mathsf{x};\boldsymbol{\theta})$ satisfies the regularity conditions (cf. problem 3.1). Then, the score function

$$\boldsymbol{g}(\mathsf{x};\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}}\ell(\mathsf{x};\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}}\log\left(L(\mathsf{x};\boldsymbol{\theta})\right)$$

can be used in order to obtain the CRB on $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}}$.

a) Let $\boldsymbol{t}(\mathsf{x},\boldsymbol{\theta})$ be a function of x and $\boldsymbol{\theta}$. Show that

$$\mathrm{E}\left[\boldsymbol{g}(\mathsf{x};\boldsymbol{\theta})\boldsymbol{t}^{\mathrm{T}}(\mathsf{x},\boldsymbol{\theta})\right] = \frac{\partial\,\mathrm{E}\left[\boldsymbol{t}^{\mathrm{T}}(\mathsf{x},\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}} - \mathrm{E}\left[\frac{\partial\boldsymbol{t}^{\mathrm{T}}(\mathsf{x},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right].$$

Hint: Start with $\frac{\partial\mathrm{E}\left[\boldsymbol{t}^{\mathrm{T}}(\mathsf{x},\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}$ and use the definition of the score function $\boldsymbol{g}(\mathsf{x};\boldsymbol{\theta})$.

b) Let $\boldsymbol{t}(\mathsf{x},\boldsymbol{\theta}) = \hat{\boldsymbol{\theta}}(\mathsf{x})$ be an unbiased estimator for $\boldsymbol{\theta}$. Use the result from sub-problem a) to determine $\mathrm{E}\left[\boldsymbol{g}(\mathsf{x};\boldsymbol{\theta})\hat{\boldsymbol{\theta}}^{\mathrm{T}}(\mathsf{x})\right]$. Note that $\hat{\boldsymbol{\theta}}(\mathsf{x})$ is not a function of $\boldsymbol{\theta}$.

c) Show that $\mathrm{E}\left[\boldsymbol{g}(\mathsf{x};\boldsymbol{\theta})\right] = \boldsymbol{0}$.

d) The covariance matrix of the score function is called Fisher information matrix $\boldsymbol{I}_{\mathrm{F}}(\boldsymbol{\theta})$. Let $\boldsymbol{t}(\mathsf{x},\boldsymbol{\theta}) = \boldsymbol{g}(\mathsf{x};\boldsymbol{\theta})$ and determine $\boldsymbol{I}_{\mathrm{F}}(\boldsymbol{\theta})$ using the results from a) and c).

e) Now consider the random variables $\mathsf{u} = \boldsymbol{a}^{\mathrm{T}}\hat{\boldsymbol{\theta}}(\mathsf{x})$, $\boldsymbol{a} \in \mathbb{R}^M$ and $\mathsf{v} = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{g}(\mathsf{x};\boldsymbol{\theta})$, $\boldsymbol{b} \in \mathbb{R}^M$. Starting with

$$\rho_{\mathsf{u},\mathsf{v}} = \frac{\mathrm{Cov}\left[\mathsf{u},\mathsf{v}\right]}{\sqrt{\mathrm{Var}\left[\mathsf{u}\right]\mathrm{Var}\left[\mathsf{v}\right]}} \leq 1,$$

show that

$$\mathrm{Var}\left[\hat{\theta}_m(\mathsf{x})\right] \geq \left[(\boldsymbol{I}_{\mathrm{F}}(\boldsymbol{\theta}))^{-1}\right]_{m,m}.$$

f) The efficiency of an estimator is defined as the ratio of the least possible variance and the actual variance. It can be shown that, under certain conditions (cf. chapter 4.6) and assuming i.i.d. statistics $\mathsf{x}^{(1)},\ldots,\mathsf{x}^{(N)}$, every ML estimator is asymptotically unbiased, i.e.,

$$\lim_{N\to\infty}\mathrm{E}\left[\hat{\boldsymbol{\theta}}_{\mathrm{ML}}(\mathsf{x}^{(1)},\ldots,\mathsf{x}^{(N)})\right] = \boldsymbol{\theta},$$

and asymptotically efficient, i.e.,

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\mathrm{ML}}(\mathsf{x}^{(1)},\ldots,\mathsf{x}^{(N)}) - \boldsymbol{\theta}\right) \overset{N\to\infty}{\sim} \mathcal{N}\left(\boldsymbol{0},\left(I_{\mathrm{F}}^{(1)}(\boldsymbol{\theta})\right)^{-1}\right).$$

Why does not every practical implementation of an estimator use the ML principle?