Multiparameter and Gaussian Cramér-Rao Bounds

Aleksandar Dogandžić

2017-01-31

Contents

Reading: §3 in the textbook, (Hero 2015, §4.5.1 and 4.6), and (Sadler and Moore 2014).

Preliminaries

We introduce assumptions and notation used throughout this handout.

Assumptions

Assume that Assumptions 1 and 2 hold.

Assumption 1. The support set of $f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta})$:

$$\operatorname{supp}\{f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta})\} \triangleq \{\boldsymbol{x} \mid f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta}) > 0\} \tag{1}$$

does not depend on $\boldsymbol{\theta}$. For all $x \in \operatorname{supp}\{f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta})\}$ and $\boldsymbol{\theta}$ in the parameter space $\mathrm{sp}_{\boldsymbol{\Theta}}$, the score function

$$\frac{\partial \ln f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

exists and is finite.

Assumption 2. If $T(\boldsymbol{x})$ is a statistic that satisfies $\mathrm{E}_{\boldsymbol{X}|\boldsymbol{\Theta}}(|T(\boldsymbol{X})| \mid \boldsymbol{\theta}) < +\infty$ for all $\boldsymbol{\theta} \in \mathrm{sp}_{\boldsymbol{\Theta}}$, then integration over $\boldsymbol{x}$ and differentiation by $\boldsymbol{\theta}$ can be interchanged when applied to $\int T(\boldsymbol{x}) f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}$, i.e.,

$$\frac{\partial}{\partial \boldsymbol{\theta}}\Big[\int T(\boldsymbol{x}) f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}\Big] = \int T(\boldsymbol{x}) \frac{\partial}{\partial \boldsymbol{\theta}} f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x} \tag{2}$$

whenever the right-hand side is finite.

Notation

IF

$$\boldsymbol{a}(\boldsymbol{\theta}) = \begin{bmatrix} a_1(\boldsymbol{\theta}) \\ a_2(\boldsymbol{\theta}) \\ \vdots \\ a_m(\boldsymbol{\theta}) \end{bmatrix}, \qquad \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{bmatrix}$$

then

$$\frac{\partial \boldsymbol{a}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathsf{T}}} = \begin{bmatrix} \frac{\partial a_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial a_1(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial a_1(\boldsymbol{\theta})}{\partial \theta_d} \\ \frac{\partial a_2(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial a_2(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial a_2(\boldsymbol{\theta})}{\partial \theta_d} \\ \vdots & \vdots & \ddots & \ddots \\ \frac{\partial a_m(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial a_m(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial a_m(\boldsymbol{\theta})}{\partial \theta_d} \end{bmatrix}$$

and

$$\frac{\partial \boldsymbol{a}(\boldsymbol{\theta})^{\mathsf{T}}}{\partial \boldsymbol{\theta}} = \Big(\frac{\partial \boldsymbol{a}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathsf{T}}}\Big)^{\mathsf{T}}.$$

Multiparameter CRB

WE EXTEND the Cramér-Rao bound (CRB) to a $d \times 1$ parameter vector

$$\boldsymbol{\theta} = [\theta_1, \ldots, \theta_d]^{\mathsf{T}}.$$

Assume that the parameter space $\mathrm{sp}_{\boldsymbol{\Theta}}$ is an open subset of $\mathbb{R}^d$ and that the model distribution $f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta})$ satisfies Assumptions 1 and 2.[1]

[1] In Assumption 2, differentiation with respect to $\boldsymbol{\theta}$ is differentiation with respect to θ_i, $i = 1, 2, \ldots, d$.

Define the $d \times d$ Fisher information matrix (FIM) for the parameter vector $\boldsymbol{\theta}$ as

$$\mathcal{I}(\boldsymbol{\theta}) = (\mathcal{I}_{i,k}(\boldsymbol{\theta}))_{i,k=1}^{d}$$

where

$$\mathcal{I}_{i,k}(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{X}|\boldsymbol{\Theta}}\Big[\frac{\partial}{\partial\theta_i}\ln f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{X}\,|\,\boldsymbol{\theta})\frac{\partial}{\partial\theta_k}\ln f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{X}\,|\,\boldsymbol{\theta})\,\Big|\,\boldsymbol{\theta}\Big].$$

Score vector has mean zero and covariance matrix equal to FIM

A multivariate extension of Lemma 1 from handout `crb`:

$$\mathrm{E}_{\boldsymbol{X}|\boldsymbol{\Theta}}\Big[\frac{\partial}{\partial\theta_i}\ln f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{X}\,|\,\boldsymbol{\theta})\,\Big|\,\boldsymbol{\theta}\Big] = 0$$

$i = 1, \ldots, d$

and

$$(\mathcal{I}_{i,k})_{i,k=1}^{d} = \mathrm{cov}_{\boldsymbol{X}|\boldsymbol{\Theta}}\Big[\frac{\partial}{\partial\theta_i}\ln f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{X}\,|\,\boldsymbol{\theta}), \frac{\partial}{\partial\theta_k}\ln f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{X}\,|\,\boldsymbol{\theta})\,\Big|\,\boldsymbol{\theta}\Big].$$

Using the vector and matrix notation, we rewrite these results as

$$\mathrm{E}_{X|\boldsymbol{\Theta}}\Big[\frac{\partial}{\partial\boldsymbol{\theta}}\ln f_{X|\boldsymbol{\Theta}}(X\,|\,\boldsymbol{\theta})\,|\,\boldsymbol{\theta}\Big] = \mathbf{0}_{d\times 1}$$

$\mathbf{0}_{d\times 1}$ denotes the $d \times 1$ vector of zeros

and

$$\mathcal{I}(\boldsymbol{\theta}) = \mathrm{cov}_{\boldsymbol{X}|\boldsymbol{\Theta}}\Big[\frac{\partial}{\partial\boldsymbol{\theta}}\ln f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{X}\,|\,\boldsymbol{\theta})\,\Big|\,\boldsymbol{\theta}\Big].$$

I.I.D. Measurements

For independent, identically distributed (i.i.d.) measurements $(X[n])_{n=0}^{N-1}$, the FIM for $\boldsymbol{\theta}$ is

$$N\mathcal{I}_1(\boldsymbol{\theta})$$

where $\mathcal{I}_1(\boldsymbol{\theta})$ is the FIM for $\boldsymbol{\theta}$ and a single measurement $X[0]$, say (or $X[1]$, etc).

Fisher information as curvature of log likelihood

An alternative expression for the FIM.

Assumption 3. $f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x}\,|\,\boldsymbol{\theta})$ is twice differentiable and it is permitted to interchange integration with respect to $\boldsymbol{x}$ and differentiation with respect to $\boldsymbol{\theta}$.

If, in addition to Assumptions 1 and 2, $f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta})$ satisfies Assumption 3, then

$$(\mathcal{I}(\boldsymbol{\theta}))_{i,k=1}^{d} = -\mathrm{E}_{\boldsymbol{X}|\boldsymbol{\Theta}}\Big[\frac{\partial^2}{\partial\theta_i\,\partial\theta_k} \ln f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{X} \mid \boldsymbol{\theta}) \,\Big|\, \theta\Big]. \tag{3}$$

✻ EXAMPLE. Suppose

$$\{X \mid \boldsymbol{\theta}\} \sim \mathcal{N}(\mu, \sigma^2)$$

To simplify the notation, we use $\mathrm{E}_{\boldsymbol{X}}[\cdot]$ instead of $\mathrm{E}_{\boldsymbol{X}|\boldsymbol{\Theta}}[\cdot|\boldsymbol{\theta}]$.

and

$$\boldsymbol{\theta} = [\mu, \sigma^2]^\mathsf{T}.$$

Then

$$\begin{aligned}
\ln f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta}) &= -0.5\ln(2\pi) - 0.5\ln(\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2 \\
\mathcal{I}_{11}(\boldsymbol{\theta}) &= -\mathrm{E}_X\Big\{\frac{\partial^2}{\partial\mu^2}\ln[f_{X|\boldsymbol{\Theta}}(X \mid \boldsymbol{\theta})]\Big\} \\
&= \mathrm{E}_X(\sigma^{-2}) = (\sigma^2)^{-1} \\
\mathcal{I}_{12}(\boldsymbol{\theta}) &= -\mathrm{E}_X\Big\{\frac{\partial}{\partial\sigma^2}\frac{\partial}{\partial\mu}\ln[f_{X|\boldsymbol{\Theta}}(X \mid \boldsymbol{\theta})]\Big\} \\
&= -\sigma^{-4}\,\mathrm{E}_X(X-\mu) \\
&= 0 = \mathcal{I}_{21}(\boldsymbol{\theta}) \\
\mathcal{I}_{22}(\boldsymbol{\theta}) &= -\mathrm{E}_X\Big\{\frac{\partial^2}{\partial(\sigma^2)^2}\ln[f_{X|\boldsymbol{\Theta}}(X \mid \boldsymbol{\theta})]\Big\} = 0.5(\sigma^2)^{-2}.
\end{aligned}$$

Therefore

$$\mathcal{I}(\boldsymbol{\theta}) = \begin{bmatrix} (\sigma^2)^{-1} & 0 \\ 0 & 0.5(\sigma^2)^{-2} \end{bmatrix}. \tag{4}$$

✻ EXAMPLE. Multiple i.i.d. Gaussian observations:

$$\big\{(X[n])_{n=0}^{N-1} \mid \boldsymbol{\theta}\big\} \sim \mathcal{N}(\mu, \sigma^2)$$

with

$$\boldsymbol{\theta} = [\mu, \sigma^2]^\mathsf{T}.$$

Then, (4) implies

$$\mathcal{I}_1(\boldsymbol{\theta}) = \begin{bmatrix} (\sigma^2)^{-1} & 0 \\ 0 & 0.5(\sigma^2)^{-2} \end{bmatrix}$$

and, consequently,

$$\begin{aligned}
\mathcal{I}(\boldsymbol{\theta}) &= N\mathcal{I}_1(\boldsymbol{\theta}) \\
&= N\begin{bmatrix} (\sigma^2)^{-1} & 0 \\ 0 & 0.5(\sigma^2)^{-2} \end{bmatrix}.
\end{aligned} \tag{5}$$

Information Inequality

SUPPOSE that Assumptions 1 and 2 hold and that the FIM $\mathcal{I}(\boldsymbol{\theta})$ is positive definite and hence nonsingular. Then, for a d-dimensional statistic $\boldsymbol{T}(\boldsymbol{X}) = [T_1(\boldsymbol{X}), \ldots, T_d(\boldsymbol{X})]^\top$ and

$$\boldsymbol{\psi}(\boldsymbol{\theta}) \triangleq \mathrm{E}_{\boldsymbol{X}\,|\,\boldsymbol{\Theta}}[\boldsymbol{T}(\boldsymbol{X}) \,|\, \boldsymbol{\theta}] = \left[\psi_1(\boldsymbol{\theta}), \ldots, \psi_d(\boldsymbol{\theta})\right]^\top \tag{6a}$$

the following holds:

$A \geq B$ means $\boldsymbol{a}^\top(A - B)\boldsymbol{a} \geq 0$ for all $d \times 1$ vectors $\boldsymbol{a}$.

$$\mathrm{cov}_{\boldsymbol{X}\,|\,\boldsymbol{\Theta}}[\boldsymbol{T}(\boldsymbol{X}) \,|\, \boldsymbol{\theta}] \geq \frac{\partial \boldsymbol{\psi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \mathcal{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\psi}(\boldsymbol{\theta})^\top}{\partial \boldsymbol{\theta}} \tag{6b}$$

for all $\boldsymbol{\theta} \in \mathrm{sp}_{\boldsymbol{\Theta}}$.

$\mathcal{I}^{-1}(\boldsymbol{\theta}) \triangleq [\mathcal{I}(\boldsymbol{\theta})]^{-1}$

Unbiased Estimators

Cramér-Rao bound

ASSUME that Assumptions 1 and 2 hold and that $\boldsymbol{T}(\boldsymbol{X})$ is an unbiased estimator of $\boldsymbol{\theta}$, i.e.,

$$\mathrm{E}_{\boldsymbol{X}\,|\,\boldsymbol{\Theta}}[\boldsymbol{T}(\boldsymbol{X}) \,|\, \boldsymbol{\theta}] = \boldsymbol{\psi}(\boldsymbol{\theta}) = \boldsymbol{\theta}.$$

Then

$$\mathrm{cov}_{\boldsymbol{X}\,|\,\boldsymbol{\Theta}}[\boldsymbol{T}(\boldsymbol{X}) \,|\, \boldsymbol{\theta}] \geq \mathcal{I}^{-1}(\boldsymbol{\theta}) \triangleq \mathrm{CRB}(\boldsymbol{\theta})$$

where the equality holds if and only if, for some constant (nonrandom) matrix $C_{\boldsymbol{\theta}}$,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ln f_{\boldsymbol{X}\,|\,\boldsymbol{\Theta}}(\boldsymbol{x} \,|\, \boldsymbol{\theta}) = C_{\boldsymbol{\theta}}[\boldsymbol{T}(\boldsymbol{x}) - \boldsymbol{\theta}] \tag{7}$$

see also equation (3.7) in the textbook.

When the CRB is attainable, it is said to be a tight bound and (7) is called the *CRB tightness condition*. If Assumption 3 holds as well, it is easy to show that, when (7) holds, $C_{\boldsymbol{\theta}}$ is the FIM for $\boldsymbol{\theta}$:

$$C_{\boldsymbol{\theta}} = \mathcal{I}(\boldsymbol{\theta}).$$

Proof: Apply (3):

$$\mathcal{I}(\theta) = -\mathrm{E}_{\boldsymbol{X}\,|\,\boldsymbol{\Theta}}\left[\frac{\partial^2}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^\top} \ln f_{\boldsymbol{X}\,|\,\boldsymbol{\Theta}}(\boldsymbol{X} \,|\, \boldsymbol{\theta}) \,\Big|\, \boldsymbol{\theta}\right] = C_{\boldsymbol{\theta}}$$

by differentiating (7) once more. □

Elements of $\boldsymbol{T}(\boldsymbol{X})$

For an unbiased estimator $\boldsymbol{T}(\boldsymbol{X})$ of $\boldsymbol{\theta}$, consider

$$\psi(\boldsymbol{\theta}) = \theta_i$$

which corresponds to $T_i(\boldsymbol{X})$, where $T_i(\boldsymbol{X})$ is the ith element of $\boldsymbol{T}(\boldsymbol{X})$. Now,

$$\frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\top}} = [0, \ldots, 0, \underbrace{1}_{i\text{th place}}, 0, \ldots, 0]$$

and, consequently,

$$\begin{aligned} \operatorname{var}_{\boldsymbol{X}|\boldsymbol{\Theta}}[T_i(\boldsymbol{x}) \mid \boldsymbol{\theta}] &\geq [\mathcal{I}(\boldsymbol{\theta})^{-1}]_{i,i} \\ &= \mathrm{CRB}_{i,i}(\boldsymbol{\theta}). \end{aligned} \tag{8}$$

$\mathrm{CRB}_{i,i}(\boldsymbol{\theta})$ is the (i,i)th element of $\mathrm{CRB}(\boldsymbol{\theta})$, the CRB matrix for $\boldsymbol{\theta}$.

Decoupling

Recall that, for multiple i.i.d. Gaussian observations:

$$\left\{(X[n])_{n=0}^{N-1} \mid \boldsymbol{\theta}\right\} \sim \mathcal{N}(\mu, \sigma^2)$$

with

$$\boldsymbol{\theta} = [\mu, \sigma^2]^{\top}$$

we obtained the FIM

$$\mathcal{I}(\boldsymbol{\theta}) = N \begin{bmatrix} (\sigma^2)^{-1} & 0 \\ 0 & 0.5(\sigma^2)^{-2} \end{bmatrix} \tag{9}$$

which is diagonal. Therefore, CRB for μ remains the same whether or not σ^2 is known. Similarly, CRB for σ^2 is the same regardless of whether or not μ is known. In general, the more parameters[2], the larger (or equal) the CRB; the CRBs are equal in the case of decoupling. See problems 3.11 and 3.12 in the textbook and (Hero 2015, § 4.6, Case II: Handling nonrandom nuisance parameters).

[2] We have to compare nested models; otherwise, we would be comparing apples and oranges.

Multiparameter Canonical Exponential Family and Efficiency

Consider the canonical d-parameter exponential family:

$$f_{\boldsymbol{X}|\boldsymbol{\eta}}(\boldsymbol{x} \mid \boldsymbol{\eta}) = \exp\big[\boldsymbol{T}^{\top}(\boldsymbol{x})\boldsymbol{\eta} - A(\boldsymbol{\eta})\big]h(\boldsymbol{x}) \tag{10}$$

and assume that the parameter space $\mathrm{sp}_{\boldsymbol{\eta}}$ of $\boldsymbol{\eta}$ is an open subset of $\mathbb{R}^d$. Then

$$\frac{\partial \ln f_{\boldsymbol{X}|\boldsymbol{\eta}}(\boldsymbol{x} \mid \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \boldsymbol{T}(\boldsymbol{x}) - \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \tag{11}$$

and, by the multivariate extension of Lemma 1 from handout `crb`,

$$\begin{aligned} \mathrm{E}_{\boldsymbol{X}|\boldsymbol{\eta}}\Big[\frac{\partial}{\partial \boldsymbol{\eta}} \ln f_{X|\boldsymbol{\eta}}(\boldsymbol{X}|\boldsymbol{\eta}) \mid \boldsymbol{\eta}\Big] &= \boldsymbol{0}_{d\times 1} \\ &= \mathrm{E}_{\boldsymbol{X}|\boldsymbol{\eta}}[\boldsymbol{T}(\boldsymbol{X}) \mid \boldsymbol{\eta}] - \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \end{aligned}$$

and

$$\begin{aligned} \mathcal{I}(\boldsymbol{\eta}) &= \mathrm{cov}_{\boldsymbol{X}|\boldsymbol{\eta}}\Big[\frac{\partial}{\partial \boldsymbol{\eta}} \ln f_{\boldsymbol{X}|\boldsymbol{\eta}}(\boldsymbol{X} \mid \boldsymbol{\eta}) \,\Big|\, \boldsymbol{\eta}\Big] \\ &= \mathrm{cov}_{\boldsymbol{X}|\boldsymbol{\eta}}[\boldsymbol{T}(\boldsymbol{X}) \mid \boldsymbol{\eta}]. \end{aligned}$$

see (11)

We also know that, if Assumption 3 holds, the FIM for $\boldsymbol{\eta}$ is

$$\begin{aligned} \mathcal{I}(\boldsymbol{\eta}) &= -\mathrm{E}_{\boldsymbol{X}|\boldsymbol{\eta}}\Big[\frac{\partial^2}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} \ln f_{\boldsymbol{X}|\boldsymbol{\eta}}(\boldsymbol{X} \mid \boldsymbol{\eta}) \,\Big|\, \boldsymbol{\eta}\Big] \\ &= \underbrace{\frac{\partial^2 A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top}}_{d \times d \text{ matrix}} \end{aligned}$$

which implies

$$\mathrm{cov}_{\boldsymbol{X}|\boldsymbol{\eta}}[T(\boldsymbol{X}) \mid \boldsymbol{\eta}] = \frac{\partial^2 A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top}.$$

Hence, we have shown that

$$\mathrm{E}_{\boldsymbol{X}|\boldsymbol{\eta}}[\boldsymbol{T}(\boldsymbol{X}) \mid \boldsymbol{\eta}] = \frac{\partial A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}, \qquad \mathrm{cov}_{\boldsymbol{X}|\boldsymbol{\eta}}\big[T(\boldsymbol{X}) \mid \boldsymbol{\eta}\big] = \frac{\partial^2 A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} \tag{12}$$

whose scalar version we have seen in (3) of handout `expon_family`.

MVU estimators

For the canonical exponential family (10) with natural sufficient statistic vector $\boldsymbol{T}(\boldsymbol{x})$, each $T_i(\boldsymbol{x})$ is an minimum-variance unbiased (MVU) estimator of its expectation $\mathrm{E}_{X|\boldsymbol{\eta}}[T_i(\boldsymbol{x}) \mid \boldsymbol{\eta}]$.

Proof. Without loss of generality, we focus on $i = 1$. Note that

$$\begin{aligned} \mathrm{E}_{X|\boldsymbol{\eta}}[T_1(\boldsymbol{X}) \mid \boldsymbol{\eta}] &= \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_1} \triangleq \psi(\boldsymbol{\eta}) \\ \mathrm{var}_{X|\boldsymbol{\eta}}[T_1(\boldsymbol{X}) \mid \boldsymbol{\eta}] &= \frac{\partial^2 A(\boldsymbol{\eta})}{\partial \eta_1^2}. \end{aligned}$$

Therefore,

$$\frac{\partial \psi(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^\top} = \frac{\partial A(\boldsymbol{\eta})}{\partial \eta_1 \partial \boldsymbol{\eta}^\top}$$

is the first row of

$$\mathcal{I}(\boldsymbol{\eta}) = \frac{\partial^2 A(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top}$$

which implies

$$\frac{\partial \psi(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^\top}[\mathcal{I}(\boldsymbol{\eta})]^{-1} = \underbrace{[1, 0, \ldots, 0]\mathcal{I}(\boldsymbol{\eta})}_{\text{first row of } \mathcal{I}(\boldsymbol{\eta})} \mathcal{I}(\boldsymbol{\eta})^{-1} = [1, 0, \ldots, 0]$$

and, finally,

$$\begin{aligned} \frac{\partial \psi(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^\top}[\mathcal{I}(\boldsymbol{\eta})]^{-1} \frac{\partial \psi(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} &= \frac{\partial \psi(\boldsymbol{\eta})}{\partial \eta_1} \\ &= \frac{\partial^2 A(\boldsymbol{\eta})}{\partial \eta_1^2} \\ &= \operatorname{var}_{X \mid \boldsymbol{\eta}}[T_1(\boldsymbol{X}) | \mu] \end{aligned}$$

i.e., the scalar information inequality is satisfied *with equality* and $T_1(\boldsymbol{X})$ is MVU for $\mathrm{E}_{X \mid \boldsymbol{\eta}}[T_1(\boldsymbol{X}) \mid \boldsymbol{\eta}]$. □

✻ EXAMPLE. If $(X[n])_{n=0}^{N-1}$ given μ and σ^2 are i.i.d. $\mathcal{N}(\mu, \sigma^2)$, then

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

is the MVU estimator of μ and

$$\frac{1}{N} \sum_{n=0}^{N-1} X^2[n]$$

is the MVU estimator of $\mu^2 + \sigma^2$. This result follows by noting that

$$f_{\boldsymbol{X} \mid \boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta}) = (2\pi\sigma^2)^{-N/2} \exp\Big\{-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - \mu)^2\Big\}$$

$$= (2\pi\sigma^2)^{-N/2} \exp\Big(-\frac{N\mu^2}{2\sigma^2}\Big) \exp\Big[-\frac{1}{2\sigma^2}\Big(N \overbrace{\frac{1}{N} \sum_{n=0}^{N-1} x^2[n]}^{T_2(\boldsymbol{x})} - 2\mu N \overbrace{\bar{x}}^{T_1(\boldsymbol{x})}\Big)\Big]$$

belongs to the two-parameter exponential family of distributions and that $T_1(\boldsymbol{x})$ and $T_2(\boldsymbol{x})$ are natural sufficient statistics. But, it *does not* follow that

$$\frac{1}{N-1} \sum_{n=0}^{N-1} (X[n] - \bar{X})^2$$

is the MVU estimator of σ^2.

Gaussian CRB

SUPPOSE that $\boldsymbol{x}$ has an N-variate Gaussian distribution:

$$\boldsymbol{x} \sim \mathcal{N}\big(\boldsymbol{\mu}(\boldsymbol{\theta}), C(\boldsymbol{\theta})\big)$$

i.e.,

$$f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta}) = \frac{1}{\sqrt{\det(2\pi C)}} \exp\left[-0.5(\boldsymbol{x}-\boldsymbol{\mu})^{\mathsf{T}} C^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right].$$

Then, the (i,k)th element of the FIM for $\boldsymbol{\theta}$ is given by the Slepian-Bangs formula (Sadler and Moore 2014, eq. (8.16))

Here, we simplify the notation and omit the functional dependence of $\boldsymbol{\mu}$ and $\boldsymbol{C}$ on $\boldsymbol{\theta}$.

$$\mathcal{I}_{i,k} = \frac{\partial \boldsymbol{\mu}^{\mathsf{T}}}{\partial \theta_i} C^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_k} + \frac{1}{2} \operatorname{tr}\Big(C^{-1} \frac{\partial C}{\partial \theta_i} C^{-1} \frac{\partial C}{\partial \theta_k}\Big). \tag{13}$$

(13) is a convenient general formula for analysis.
Proof: See Appendix 3C in the textbook. □

Signal plus noise

CONSIDER the following signal-plus-noise model:

$$X[n] = s[n;\theta] + W[n]$$

where θ is the unknown parameter and $W[n]$ is additive white Gaussian noise (AWGN) with known variance σ^2. Then, we can write this model specification for N measurements in a vector form as follows:

$$\{\boldsymbol{x} \mid \theta\} = \boldsymbol{\mu}(\theta) + \boldsymbol{W} \sim \mathcal{N}(\boldsymbol{\mu}(\theta), \underbrace{\sigma^2 I_N}_{C}).$$

I_N is the $N \times N$ identity matrix

where $\boldsymbol{\mu}(\theta) = \big[s[0;\theta], s[1;\theta], \ldots, s[N-1;\theta]\big]^{\mathsf{T}}$ and C does not depend on θ (and, furthermore, is completely known).

$$\begin{aligned} \mathcal{I}(\theta) &= \frac{1}{\sigma^2} \frac{\partial \boldsymbol{\mu}^{\mathsf{T}}(\theta)}{\partial \theta} \frac{\partial \boldsymbol{\mu}(\theta)}{\partial \theta} \\ &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \Big(\frac{\partial s[n;\theta]}{\partial \theta}\Big)^2 \end{aligned}$$

which is the familiar expression that we derived earlier, see the sinusoidal frequency estimation example in handout `crb`.

What if we have a $d \times 1$ vector of parameters $\boldsymbol{\theta}$? In this case,

$$\begin{aligned} (\mathcal{I}_{i,k}(\boldsymbol{\theta}))_{i,k=1}^{d} &= \frac{1}{\sigma^2} \frac{\partial \boldsymbol{\mu}^{\mathsf{T}}(\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_k} \\ &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \frac{\partial s[n;\boldsymbol{\theta}]}{\partial \theta_i} \frac{\partial s[n;\boldsymbol{\theta}]}{\partial \theta_k} \end{aligned}$$

or, using the matrix notation,

$$\mathcal{I}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \frac{\partial \boldsymbol{\mu}^{\mathsf{T}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathsf{T}}}.$$

Noise parameters

CONSIDER AWGN $(X[n])_{n=0}^{N-1}$ with variance σ^2, i.e.,

$$\{\boldsymbol{x} \mid \sigma^2\} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_N).$$

* VARIANCE. If the variance σ^2 of the measurements is the unknown parameter, then

$$\begin{aligned}
\mathcal{I}_{\sigma^2}(\sigma^2) &= \frac{1}{2}\operatorname{tr}\Big(C^{-1}\frac{\partial C}{\partial \sigma^2}C^{-1}\frac{\partial C}{\partial \sigma^2}\Big) \\
&= \frac{1}{2(\sigma^2)^2}\operatorname{tr} I_N \\
&= \frac{N}{2(\sigma^2)^2}
\end{aligned}$$

and, therefore,

$$\mathrm{CRB}_{\sigma^2}(\sigma^2) = [\mathcal{I}_{\sigma^2}(\sigma^2)]^{-1} = \frac{2(\sigma^2)^2}{N}. \tag{14}$$

* STANDARD deviation. Consider computing the CRB for the standard deviation of the measurements $\sigma = \sqrt{\sigma^2}$:

$$\begin{aligned}
\mathrm{CRB}_{\sigma}(\sigma) &= [\mathcal{I}_{\sigma}(\sigma)]^{-1} \\
&= \Big[0.5\operatorname{tr}\Big(C^{-1}\frac{\partial C}{\partial \sigma}C^{-1}\frac{\partial C}{\partial \sigma}\Big)\Big]^{-1} \\
&= \frac{\sigma^2}{2N}
\end{aligned}$$

which can also be computed using change of variables: $\sigma = h(\sigma^2) = (\sigma^2)^{1/2}$, $h'(\sigma^2) = 0.5(\sigma^2)^{-1/2}$, and

$\sigma = \psi(\sigma^2) = (\sigma^2)^{1/2}$ is an invertible transform of the variance σ^2

$$\mathrm{CRB}_{\sigma}(\sigma) = |h'(\sigma^2)|^2\,\mathrm{CRB}_{\sigma^2}(\sigma^2) = 0.25\sigma^{-2}\frac{2(\sigma^2)^2}{N} = \frac{\sigma^2}{2N}.$$

HW: Show the above change-of-variables formula for CRB.

* MSE below CRB. Here, we consider the same measurement model as in the variance estimation example of `handout est_perf`. There, we studied the following family of estimators of σ^2:

$$\widehat{\sigma}^2 = c\frac{1}{N}\sum_{n=0}^{N-1} X^2[n]$$

and found that

$$c_{\mathrm{OPT}} = \frac{N}{N+2}$$

yields an estimator

$$\widehat{\sigma}^2_\star = c_{\mathrm{OPT}}\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] = \frac{1}{N+2}\sum_{n=0}^{N-1} x^2[n]$$

whose mean-square error (MSE) is the smallest within the family:

$$\mathrm{MSE}_{\mathrm{MIN}} = \frac{2(\sigma^2)^2}{N+2} < \frac{2(\sigma^2)^2}{N} = \mathrm{CRB}(\sigma^2)$$

see (14). Note that $\hat{\sigma}^2_\star$ is a *biased* estimator of σ^2 and that CRB is a lower bound on variance of *unbiased* estimators only.

Acronyms

AWGN additive white Gaussian noise. 9, 10

CRB Cramér-Rao bound. 2, 5, 6, 10, 11

FIM Fisher information matrix. 3, 5–7, 9

i.i.d. independent, identically distributed. 3, 4, 6, 8

MSE mean-square error. 11

MVU minimum-variance unbiased. 7, 8

References

Hero, Alfred O. (2015). *Statistical Methods for Signal Processing*. Lecture notes. Univ. Michigan, Ann Arbor, MI.

Sadler, Brian M. and Terrence J. Moore (2014). "Performance Analysis and Bounds". In: *Array and Statistical Signal Processing*. Ed. by Abdelhak M. Zoubir, Mats Viberg, Rama Chellappa, and Sergios Theodoridis. Vol. 3. Academic Press Library in Signal Processing. Elsevier. Chap. 8, pp. 297–322.