ECE 564: Detection and Estimation

Point Estimation:
Cramér-Rao Lower Bound

http://courseinfo.cit.cornell.edu/courses/ECE564/

Lang Tong
School of Electrical and Computer Engineering
Cornell University, Ithaca, NY 14853
ltong@ece.cornell.edu

Outline

Topics

- Fisher information matrix and CRB.
- CRB for functions of parameters.
- CRB for Gaussian models.
- Chapman-Robbins, Bhattachayya bounds.
- CRB for random parameters.
- CRB for complex models.

References:

1. H.V. Poor, An Introduction to Signal Detection and Estimation, 2nd Ed., Springer-Verlag, 1994, Chapter IV-C.
2. S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice Hall, 1993.
3. P.J. Bickel and K.A. Doksum, Mathematical Statistics: Basic Ideas and Selected Topics, Prentice Hall, Englewood Cliffs, NJ, 1977.
4. E.L. Lehmann, Theory of Point Estimation, Chapman & Hall, New York, 1991.

Motivations

To Find UMVU:

1. Find the complete sufficient $\mathbf{T} = \mathbf{t}(\mathbf{Y})$.
2. Two ways:
   (a) Find an unbiased estimator $\hat{\mathbf{g}}(\mathbf{T})$.
   (b) Find any unbiased estimator $\hat{\mathbf{g}}(\mathbf{Y})$ and $\hat{\mathbf{g}}_*(\mathbf{T}) = \mathbb{E}(\hat{\mathbf{g}}(\mathbf{Y})|\mathbf{T})$

Difficulties:

1. Complete sufficient statistics may be difficult to find.
2. $\hat{\mathbf{g}}_*(\mathbf{T}) = \mathbb{E}(\hat{\mathbf{g}}(\mathbf{Y})|\mathbf{T})$ may be hard to compute.
3. It is difficult to know, without finding UMVU, whether certain performance can be achieved.

An alternative strategy:

- Find a tight lower bound on MSE among all unbiased estimators.
- Check if the lower bound can be achieved.

Schur Complement

Block Diagonalization

Consider a block matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

where $\mathbf{A}_{ii}$ are square and nonsingular. Matrix $\mathbf{A}$ can be diagonalized by

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Delta}_{11} \end{bmatrix}.$$

where the Schur Complement of $\mathbf{A}_{11}$ is defined as

$$\mathbf{\Delta}_{11} \triangleq \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$$

Decorrelation

If $\mathbf{A} \geq \mathbf{0}$ is the covariance matrix† of a zero mean random vector $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$. The vector $\mathbf{x}$ can be decorrelated via transform

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{X}_1 \end{bmatrix}$$

with covariance $\text{Cov}(\mathbf{Y}) = \text{diag}\{\mathbf{A}_{11}, \mathbf{\Delta}_{11}\}$, and

$$\mathbf{\Delta}_{11} \triangleq \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \geq 0$$

with equality iff

$$\mathbf{X}_2 = \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{X}_1 \quad \text{a.s.}$$

† By $\mathbf{A} \geq 0$ we mean that matrix $\mathbf{A}$ is positive semidefinite, *i.e.*, , for any column vector $\mathbf{v}$, $\mathbf{v}'\mathbf{A}\mathbf{v} \geq 0$, which implies that all diagonal blocks of $\mathbf{A}$ are also positive semidefinite.

Score Function and Fisher Information

Definition

Consider the real vector model $p(\mathbf{y};\boldsymbol{\theta}), \boldsymbol{\theta} \in \mathcal{R}^K$. The score function is defined by

$$\mathbf{s}(\mathbf{y};\boldsymbol{\theta}) \triangleq \begin{bmatrix} \frac{\partial}{\partial\theta_1}\ln p(\mathbf{y};\boldsymbol{\theta}) \\ \vdots \\ \frac{\partial}{\partial\theta_K}\ln p(\mathbf{y};\boldsymbol{\theta}) \end{bmatrix}$$

Under regularity conditions, $\mathbb{E}_{\boldsymbol{\theta}}(\mathbf{s}(\mathbf{Y};\boldsymbol{\theta})) = \mathbf{0}$.

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}}(\frac{\partial}{\partial\theta_i}\ln p(\mathbf{Y};\boldsymbol{\theta})) &= \int p(\mathbf{y};\boldsymbol{\theta})\frac{\partial}{\partial\theta_i}\ln p(\mathbf{y};\boldsymbol{\theta})d\mathbf{y} \\ &= \int \frac{\partial}{\partial\theta_i}p(\mathbf{y};\boldsymbol{\theta})d\mathbf{y} = \frac{\partial}{\partial\theta_i}\int p(\mathbf{y};\boldsymbol{\theta})d\mathbf{y} \end{aligned}$$

Fisher Information Matrix

The covariance matrix of $\mathbf{s}(\mathbf{Y};\boldsymbol{\theta})$ is the Fisher Information Matrix

$$\mathbf{I}(\boldsymbol{\theta}) \triangleq \mathbb{E}(\mathbf{s}(\mathbf{Y};\boldsymbol{\theta})\mathbf{s}'(\mathbf{Y};\boldsymbol{\theta})) \geq \mathbf{0}$$

The (i,j)th entry of $\mathbf{I}(\boldsymbol{\theta})$ can also be written as

$$\mathbf{I}_{ij}(\boldsymbol{\theta}) = \mathbb{E}(\frac{\partial}{\partial\theta_i}\ln p(\mathbf{y};\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}\ln p(\mathbf{y};\boldsymbol{\theta})) = -\mathbb{E}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln p(\mathbf{y};\boldsymbol{\theta})$$

where the second equality is based on

$$\mathbb{E}(\frac{1}{p(\mathbf{y};\boldsymbol{\theta})}\frac{\partial^2}{\partial\theta_i\partial\theta_j}p(\mathbf{y};\boldsymbol{\theta})) = \mathbf{0}$$

The Cramér-Rao Lower Bound

Theorem (The scalar case.)

Given $\mathbf{Y} \sim p(\mathbf{y};\theta)$, let $\hat{\theta}$ be a scaler unbiased estimator of θ. Then, under regularity conditions[‡],

$$\text{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

where $I(\theta)$ is the Fisher Information. The equality holds if and only if the scoring function satisfies

$$s(y;\theta) \triangleq \frac{\partial}{\partial\theta} \ln p(y;\theta) = I(\theta)(\hat{\theta}(y) - \theta)$$

Proof:

- For any unbiased estimator $\hat{\theta}$, Consider vector $\mathbf{z} \triangleq \begin{bmatrix} \mathbf{s}(\mathbf{y};\theta) \\ \hat{\theta}(y) - \theta \end{bmatrix}$. We have $\mathbb{E}(\mathbf{z}) = \mathbf{0}$.
- Compute the covariance $\text{Cov}(\mathbf{z}) = \begin{bmatrix} I(\theta) & 1 \\ 1 & \text{Var}(\hat{\theta}) \end{bmatrix}$. The Schur complement of $I(\theta)$ implies

  $$\text{Var}(\hat{\theta}) - I^{-1}(\theta) \geq 0$$

  with equality holds if and only if

  $$\hat{\theta}(y) - \theta = I^{-1}(\theta)s(\mathbf{y};\theta) \quad \text{almost surely}$$

Generalization

For biased estimator, $\mathbb{E}(\hat{\theta}) = \Phi(\theta)$, then

$$\text{Var}(\hat{\theta}) \geq \frac{[\Phi'(\theta)]^2}{I(\theta)}$$

with equality iff

$$s(x;\theta) = I(\theta)(\hat{\theta}(y) - \Phi(\theta))$$

[‡] The regularity conditions involve (i) The support of $p(\mathbf{x};\theta)$ does not depend on θ. (ii) All derivatives exist. (iii) Switch between $\mathbb{E}\{\cdot\}$ and $\frac{\partial}{\partial\theta}$.

An Alternative Proof

- Unbiasedness:
$$\mathbb{E}(\hat{\theta}) = \theta \rightarrow \int \hat{\theta}\frac{\partial}{\partial\theta}p = 1.$$
- Variation of the likelihood function $p(\mathbf{y};\theta)$ at the true parameter:
$$\frac{\partial p}{\partial\theta} = p\frac{\partial}{\partial\theta}\ln p, \quad \mathbb{E}(\frac{\partial}{\partial\theta}\ln p) = 0$$
- Substitution:
$$\mathbb{E}\{(\hat{\theta}-\theta)\frac{\partial}{\partial\theta}\ln p\} = 1.$$
- Schwarz Inequality: $|\mathbb{E}(XY)|^2 \leq \mathbb{E}(X^2)E(Y^2)$ with equality iff $Y = cX$.
$$\text{Var}(\theta)\mathbb{E}(\frac{\partial}{\partial\theta}\ln p)^2 \geq 1$$
with equality only when
$$c(\theta)(\hat{\theta}-\theta) = \frac{\partial}{\partial\theta}\ln p$$
- Note:
$$\frac{\partial}{\partial\theta}\int p\frac{\partial}{\partial\theta}\ln p = \mathbb{E}(\frac{\partial}{\partial\theta}\ln p)^2 + \mathbb{E}(\frac{\partial^2}{\partial\theta^2}\ln p) = 0$$
- Finally, to find $c(\theta)$, because $\hat{\theta}$ is unbiased,
$$c(\theta) = -\mathbb{E}(\frac{\partial^2}{\partial\theta^2}\ln p) = I(\theta)$$

Efficiency

Definition An unbiased estimator is efficient if it achieves CRB.

Theorem

If there exists an efficient estimator $\hat{\theta}$, then the distribution of the observation must be belong to the exponential family. The efficient estimator can be found by the maximum likelihood (ML) estimator:

$$\hat{\theta}_{\text{ML}} = \arg\max_{\theta} \ln p(\mathbf{y};\theta)$$

Proof: If the CRB is achieved by an unbiased estimator $\hat{\theta}(\mathbf{y})$,

$$\frac{\partial}{\partial\theta} \ln p(\mathbf{y};\theta) = I(\theta)(\hat{\theta}(\mathbf{y}) - \theta) \text{ a.s.}$$

which implies

$$p(\mathbf{y};\theta) = h(\mathbf{y}) exp\{\hat{\theta} \int_{-\infty}^{\theta} I(u)du - \int_{-\infty}^{\theta} I(u)udu\}$$

and $\hat{\theta}$ is a complete sufficient statistic. To show that $\hat{\theta}$ is the maximum likelihood estimator, we note that

$$\frac{\partial}{\partial\theta} \ln p(\mathbf{y};\theta)|_{\theta=\hat{\theta}_{\text{ML}}} = I(\theta)(\hat{\theta} - \hat{\theta}_{\text{ML}}) = 0.$$

Remark An efficient estimator is UMVU but a UMVU estimator may not be efficient (when CRB is not achievable).

Example: Estimating Signal Amplitude

Example: Sinusoid in Noise:

$$x_n = \alpha cos(\omega_0 n + \phi) + w_n, \quad n = 0, \cdots, N-1,$$

where $w_n \sim \mathcal{N}(0, \sigma^2)$ and i.i.d.. All variables except α are known. In vector form:

$$\mathbf{x} = \mathbf{h}\alpha + \mathbf{w},$$

where

$$\begin{aligned} \mathbf{x} &= [x_0, \cdots, x_{N-1}]^t, \mathbf{w} = [w_0, \cdots, w_{N-1}]^t, \\ \mathbf{h} &= [cos(\phi), \cdots, cos(\omega_0(N-1) + \phi)]^t; \end{aligned} \tag{1}$$

1. Log-likelihood function. Denote $\mathbf{x} = [x_0, \cdots, x_{N-1}]'$.

$$\ln p(\mathbf{x}; \alpha) = -\frac{||\mathbf{x} - \mathbf{h}\alpha||^2}{2\sigma^2} + const.$$

2. The score function:

$$s(\mathbf{x}; \alpha) = \frac{||\mathbf{h}||^2}{\sigma^2}(\frac{\mathbf{x}^t\mathbf{h}}{||\mathbf{h}||^2} - \alpha)$$

3. Fisher Information:

$$I(\alpha) = \frac{||\mathbf{h}||^2}{\sigma^2}$$

4. CRLB:

$$\text{Var}(\hat{\alpha}) \geq \frac{\sigma^2}{||\mathbf{h}||^2}$$

   with equality with the least squares estimator

$$\hat{\alpha}_{LS} = \arg\min_{\alpha} ||\mathbf{x} - \alpha\mathbf{h}||^2 = \frac{\mathbf{x}^t\mathbf{h}}{||\mathbf{h}||^2},$$

   The least squares estimator is unbiased and is UMVU.

5. Asymptotic Performance: As $N \to \infty$, $Var(\alpha_{LS}) \to 0$. Consistent.

The estimator $\hat{\alpha}_{LS}$ is (i) UMVU, (ii) efficient, (iii) Gaussian, (iii) and consistent.

Example: Estimating Signal Phase

Example: Sinusoid in Noise:

$$x_n = \alpha cos(\omega_0 n + \phi) + w_n, \quad n = 0, \cdots, N-1,$$

where $w_n \sim \mathcal{N}(0, \sigma^2)$ and i.i.d.. All variables except ϕ are known. In vector form:

$$\mathbf{x} = \mathbf{h}\alpha + \mathbf{w},$$

where

$$\begin{aligned} \mathbf{x} &= [x_0, \cdots, x_{N-1}]^t, \mathbf{w} = [w_0, \cdots, w_{N-1}]^t, \\ \mathbf{h} &= [cos(\phi), \cdots, cos(\omega_0(N-1) + \phi)]^t; \end{aligned} \tag{2}$$

1. Log-likelihood function. Denote $\mathbf{x} = [x_0, \cdots, x_{N-1}]'$.

$$\ln p(\mathbf{x}; \phi) = -\frac{||\mathbf{x} - \mathbf{h}\alpha||^2}{2\sigma^2} + const.$$

2. The score function:

$$s(\mathbf{x}; \phi) = -\frac{\alpha}{\sigma^2}(\sum_i x_i sin(i\omega_0 + \phi) - \frac{\alpha}{2}\sum_i sin(2i\omega_0 + 2\phi))$$

3. Fisher Information:

$$\begin{aligned} I(\phi) &= \frac{\alpha^2}{\sigma^2}(\sum_i cos^2(i\omega_0 + \phi) - \sum_i cos(2i\omega_0 + 2\phi)) \\ &= \frac{N\alpha^2}{2\sigma^2} - \frac{\alpha^2}{2\sigma^2}\sum_i cos(2i\omega_0 + 2\phi) \approx \frac{N\alpha^2}{2\sigma^2} \end{aligned}$$

4. CRLB:

$$Var(\hat{\phi}) \geq \frac{2\sigma^2}{N\alpha^2}$$

but unachievable. (Not the one-parameter exp. family.!)

Example: UMVU and CRB

Example: Let X has the Poisson Distribution with parameter θ:

$$\Pr\{X=k\} = \frac{e^{-\theta}\theta^k}{k!}, \quad k = 0, 1 \cdots \tag{3}$$

To estimate $e^{-\theta}$, consider the estimator

$$T(X) = \begin{cases} 1 & \text{if } X = 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

UMVU For an estimator $g(x)$ to be unbiased, we have

$$\sum g(k)\frac{e^{-\theta}\theta^k}{k!} = e^{-\theta}, \quad \forall\theta$$

which implies that $g(X) = T(X)$, *i.e.*, there is only one unbiased estimator. Hence T is UMVU.

CRB

$$\begin{aligned} \text{CRB} &= \theta e^{-2\theta} \quad (5) \\ \text{Var}(T) &= e^{-2\theta}(e^{\theta} - 1) \geq \theta e^{-2\theta}. \quad (6) \end{aligned}$$

Remark:

The UMVU estimator may not achieve CRLB.

CRB: The Vector Case

Theorem Let $\hat{\theta}$ be an unbiased estimator of $\boldsymbol{\theta}$. Then

$$E\{(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^t\} \geq \mathbf{I}^{-1}(\boldsymbol{\theta})$$

where $I(\boldsymbol{\theta})$ is the Fisher Information Matrix

$$\begin{aligned} [I(\theta)]_{ij} &= \mathbb{E}\{\frac{\partial \ln p(\mathbf{Y};\boldsymbol{\theta})}{\partial \theta_i}\frac{\partial \ln p(\mathbf{Y};\boldsymbol{\theta})}{\partial \theta_j}\} = -\mathbb{E}\{\frac{\partial^2 \ln p(\mathbf{Y};\theta)}{\partial \theta_i \partial \theta_j}\} \\ \mathbf{I}(\boldsymbol{\theta}) &= \mathbb{E}\{[\nabla_\theta \ln p(\mathbf{Y};\boldsymbol{\theta})][\nabla_\theta \ln p(\mathbf{Y};\boldsymbol{\theta})]^T\} = -\mathbb{E}\{\nabla^2 \ln p(\mathbf{Y};\boldsymbol{\theta})\} \end{aligned}$$

The equality holds iff

$$\nabla_\theta \ln p(\mathbf{Y};\boldsymbol{\theta}) = \mathbf{I}(\theta)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$$

Remarks:

- $A \geq B$ iff $A - B$ is positive semidefinite $\Rightarrow A_{ii} \geq B_{ii}$.
- Definition: For a real-valued function $f(\boldsymbol{\theta})$

$$\underbrace{\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial f}{\partial \theta_1} \\ \vdots \\ \frac{\partial f}{\partial \theta_n} \end{pmatrix}}_{\text{Gradient}} \qquad \underbrace{\nabla^2_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) \triangleq \nabla_\theta(\nabla_{\boldsymbol{\theta}} f(\theta)^T) = \begin{pmatrix} \frac{\partial^2 f}{\partial \theta_1^2} & \cdots & \frac{\partial^2 f}{\partial \theta_1 \partial \theta_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial \theta_n \partial \theta_1} & \cdots & \frac{\partial^2 f}{\partial \theta_n^2} \end{pmatrix}}_{\text{Hessian}}$$

- Useful formulae:

$$\begin{aligned} \nabla_\theta \boldsymbol{\theta}^t &= \mathbf{I}, \quad \nabla_{\boldsymbol{\theta}} \mathbf{a}^t \boldsymbol{\theta} = \mathbf{a} \\ \nabla_\theta \mathbf{a}^t(\boldsymbol{\theta})\mathbf{b}(\boldsymbol{\theta}) &= (\nabla_\theta \mathbf{a}^t(\boldsymbol{\theta}))\mathbf{b}(\boldsymbol{\theta}) + (\nabla_\theta \mathbf{b}^t(\theta))\mathbf{a}(\theta) \\ \nabla_\theta \boldsymbol{\theta}^t \mathbf{Q} \boldsymbol{\theta} &= 2\mathbf{Q}\boldsymbol{\theta} \quad (\mathbf{Q}^T = \mathbf{Q}) \\ \nabla_\theta \mathbf{a}^t(\boldsymbol{\theta})\mathbf{Q}\mathbf{a}(\boldsymbol{\theta}) &= 2(\nabla_\theta \mathbf{a}^t(\boldsymbol{\theta}))\mathbf{Q}\mathbf{a}(\boldsymbol{\theta}) \quad (\mathbf{Q}^T = \mathbf{Q}) \end{aligned}$$

Proof

We call the following two properties

- The unbiasedness

$$\mathbb{E}(\hat{\boldsymbol{\theta}}^T) = \boldsymbol{\theta}^T \rightarrow \mathbb{E}(\nabla_\theta \ln p(\mathbf{y};\theta)\hat{\boldsymbol{\theta}}^T) = \mathbf{I}$$

- Block diagonalization:

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_{11} \end{bmatrix}.$$

  where

$$\boldsymbol{\Delta}_{11} \stackrel{\Delta}{=} \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$$

  is the Schur compliment of $\mathbf{A}_{11}$. If $\mathbf{A} \geq \mathbf{0}$, then $\boldsymbol{\Delta}_{11} \geq \mathbf{0}$.

- The "score" function is given by

$$\mathbf{s}(\mathbf{y};\theta) \stackrel{\Delta}{=} \nabla_\theta \ln p(\mathbf{y};\boldsymbol{\theta})$$

- Consider $\mathbf{x} = \begin{pmatrix} s(\mathbf{y};\boldsymbol{\theta}) \\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \end{pmatrix}$. We then have

$$\mathbb{E}(\mathbf{x}\mathbf{x}') = \begin{pmatrix} \mathbf{I}(\boldsymbol{\theta}) & \mathbf{I} \\ \mathbf{I} & \mathsf{Cov}(\hat{\boldsymbol{\theta}}) \end{pmatrix} \geq \mathbf{0}$$

  The Schur compliment of $\mathbf{I}(\boldsymbol{\theta})$ satisfies

$$\mathsf{Cov}(\hat{\boldsymbol{\theta}}) - \mathbf{I}^{-1}(\boldsymbol{\theta}) \geq \mathbf{0}$$

- The CRLB is achieved when $\mathbf{x}$ degenerates,

$$s(\mathbf{y};\boldsymbol{\theta}) = \mathbf{C}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{I}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$

The Linear Gaussian Model

- The linear Gaussian model is given by

$$\mathbf{X} = \mathbf{H}\boldsymbol{\theta} + \mathbf{W},$$

where $\mathbf{H} \in \mathcal{R}^{n\times m}$ is known, $\boldsymbol{\theta} \in \mathcal{R}^m$ deterministic unknown parameter, and $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ random Gaussian.

- The log-likelihood function is given by

$$\ln p(\mathbf{x};\boldsymbol{\theta}) = -\frac{||\mathbf{x} - \mathbf{H}\boldsymbol{\theta}||^2}{2\sigma^2} + const.$$

- The score function is

$$\begin{aligned} s(\mathbf{x};\theta) &= \nabla_\theta \ln p(\mathbf{x};\boldsymbol{\theta}) \\ &= -\frac{1}{2\sigma^2}(2\mathbf{H}^t\mathbf{H}\boldsymbol{\theta} - 2\mathbf{H}^t\mathbf{x}) \\ &= \frac{(\mathbf{H}^t\mathbf{H})}{\sigma^2}((\mathbf{H}^t\mathbf{H})^{-1}\mathbf{H}^t\mathbf{x} - \boldsymbol{\theta}) \end{aligned}$$

- The Fisher Information Matrix is

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{(\mathbf{H}^t\mathbf{H})}{\sigma^2}$$

It is unusual that the FIM is independent of $\boldsymbol{\theta}$!

- The estimator

$$\hat{\boldsymbol{\theta}}(\mathbf{x}) = (\mathbf{H}^t\mathbf{H})^{-1}\mathbf{H}^t\mathbf{x}$$

is UMVU, efficient, Gaussian, ML, LS estimator.

Extension–Function of parameters

Theorem

Let $\hat{\mathbf{g}}$ be an unbiased estimator of $\mathbf{g}(\boldsymbol{\theta})$, and $\mathbf{G}(\boldsymbol{\theta})$ be the Jacobian

$$\mathbf{G}(\boldsymbol{\theta}) \stackrel{\Delta}{=} [\nabla_\theta \mathbf{g}^t(\boldsymbol{\theta})]^t = \begin{pmatrix} \frac{\partial g_1}{\partial \theta_1} & \cdots & \frac{\partial g_1}{\partial \theta_n} \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial \theta_1} & \cdots & \frac{\partial g_m}{\partial \theta_n} \end{pmatrix}$$

Then

$$\mathbb{E}\{(\hat{\mathbf{g}} - \mathbf{g}(\theta))(\hat{\mathbf{g}} - \mathbf{g}(\theta))^t\} \geq \mathbf{G}(\theta)\mathbf{I}^{-1}(\theta)\mathbf{G}^t(\theta),$$

where the equality holds iff

$$\hat{\mathbf{g}} - \mathbf{g}(\boldsymbol{\theta}) = \mathbf{G}(\boldsymbol{\theta})\mathbf{I}^{-1}(\boldsymbol{\theta})\frac{\partial}{\partial \boldsymbol{\theta}} \ln p(\mathbf{y}; \boldsymbol{\theta})$$

Proof:

- Let $\mathbf{y} = \begin{pmatrix} \hat{\mathbf{g}} - \mathbf{g} \\ \nabla_\theta \ln p(\mathbf{x}; \theta) \end{pmatrix}$.

$$\mathbb{E}(\mathbf{y}\mathbf{y}^t) = \begin{pmatrix} \mathsf{Cov}(\hat{\mathbf{g}}) & \mathbf{G}(\theta) \\ \mathbf{G}^t(\theta) & \mathbf{I}(\theta) \end{pmatrix}$$

- Apply a useful identity:

$$\begin{pmatrix} \mathbf{I} & -\mathbf{D}\mathbf{B}^{-1} \\ 0 & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A} & \mathbf{D} \\ \mathbf{C} & \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{I} & 0 \\ -\mathbf{B}^{-1}\mathbf{C} & \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{A} - \mathbf{D}\mathbf{B}^{-1}\mathbf{C} & 0 \\ 0 & \mathbf{B} \end{pmatrix}$$

  Note $\mathbf{A} - \mathbf{D}\mathbf{B}^{-1}\mathbf{C}$ is the Schur Complement.

CRLB for the Gaussian Case

Theorem

Let $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$. Then

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = [\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i}]^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_j}] + \frac{1}{2} tr\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_j}\}$$

where

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i} = \begin{pmatrix} \frac{\partial \mu_1(\theta)}{\partial \theta_i} \\ \vdots \\ \frac{\partial \mu_n(\theta)}{\partial \theta_i} \end{pmatrix} \qquad \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_k} = [\frac{\partial \boldsymbol{\Sigma}_{ij}(\boldsymbol{\theta})}{\partial \theta_k}]$$

Proof: See Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall.

Example: The Linear Model

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{H}\boldsymbol{\theta} = \sum \mathbf{h}_i \boldsymbol{\theta}_i, \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \sigma^2 \mathbf{I}.$$

From the general formula for the Gaussian case,

$$\mathbf{I}(\boldsymbol{\theta}) = [\mathbf{h}_i^t (\frac{1}{\boldsymbol{\sigma}^2}) \mathbf{h}_j]_{ij} = \frac{1}{\sigma^2} \mathbf{H}^T \mathbf{H}$$

Other Bounds

- The Chapman-Robbins Inequality. (Use finite differences)

$$\mathbf{x} \triangleq \begin{pmatrix} \mathbf{s}(\mathbf{y};\boldsymbol{\theta}) \\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \end{pmatrix}, \mathbf{s}(\mathbf{y};\boldsymbol{\theta}) = [\frac{p(\mathbf{y};\boldsymbol{\theta}+\Delta\mathbf{e}_i)}{p(\mathbf{y};\boldsymbol{\theta})} - 1]$$

We then have $\mathbb{E}(\mathbf{s}(\mathbf{Y};\boldsymbol{\theta})) = \mathbf{0}$, using again the Schur Complement:

$$\text{Cov}(\boldsymbol{\theta}) \geq \mathbf{J}_{\Delta}^{-1}(\boldsymbol{\theta})$$

- The Bhattachayya System of Inequality: (Use high-order derivatives)

$$\mathbf{y} \triangleq \begin{pmatrix} \mathbf{s}(\mathbf{y};\boldsymbol{\theta}) \\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \end{pmatrix}, \mathbf{s}(\mathbf{y};\boldsymbol{\theta}) = \frac{1}{p(\mathbf{y};\boldsymbol{\theta})} \begin{pmatrix} \frac{\partial^k p}{\partial \theta_1^k} \\ \vdots \\ \frac{\partial^k p}{\partial \theta_n^k} \end{pmatrix}$$

We then have $\mathbb{E}(\mathbf{s}(\mathbf{Y};\boldsymbol{\theta})) = 0$, using again the Schur Complement:

$$\text{Cov}(\boldsymbol{\theta}) \geq \mathbf{J}_{k}^{-1}(\boldsymbol{\theta})$$

Note that for $k = 1$, we have the CRLB.

Extension to Random Parameters

Theorem:

Consider $(\mathbf{Y}, \boldsymbol{\Theta}) \sim p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) = p(\mathbf{y}, \boldsymbol{\theta})$ and $\Theta_i \in (-\infty, \infty)$, and $p(\mathbf{y}, \boldsymbol{\theta}) > 0$. Let $\hat{\boldsymbol{\Theta}}$ be a Bayesian estimator of $\boldsymbol{\Theta}$. Subject to regularity conditions

1. $p(\mathbf{y}, \boldsymbol{\theta})$ is absolutely continuous with respect to $\boldsymbol{\theta}$;
2. $\lim_{\theta_i \to \pm\infty} \theta_i p(\mathbf{y}, \theta_i) = 0 \quad \forall i$, or

2' the conditional bias satisfies

$$\lim_{\theta_i \to \pm\infty} \mathbb{E}(\hat{\Theta}_i - \theta_i | \boldsymbol{\Theta} = \boldsymbol{\theta}) p(\theta_i) = 0, \quad \forall i$$

we have

$$\mathcal{M}(\hat{\boldsymbol{\Theta}}) \stackrel{\Delta}{=} \mathbb{E}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})^T \geq \mathbf{J}^{-1}$$

where, assuming all expectations are finite and inverse exist,

$$\mathbf{J} = \mathbb{E}\{[\nabla_\theta \ln p(\mathbf{Y}, \boldsymbol{\Theta})][\nabla_\theta \ln p(\mathbf{Y}, \boldsymbol{\Theta})]^T\}$$

The equality holds iff $p(\theta|\mathbf{y})$ is Gaussian.

Remarks:

- The bound has the same form except that the Fisher Information Matrix here is averaged over both $\mathbf{Y}$ and $\boldsymbol{\Theta}$.
- Note that the unbiased assumption is not imposed. Indeed, MMSE estimator in general is biased.

Proof

The proof uses the same technique as that used in the proof of the vector CRLB for deterministic parameters. We will only show the case when (2) holds. See Van Trees for the proof of (2'). Defining

$$\mathbf{s}(\mathbf{y},\boldsymbol{\theta}) = \nabla_\theta \ln p(\mathbf{y};\boldsymbol{\theta}), \quad \mathbf{z} \triangleq \begin{pmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \\ \mathbf{s}(\mathbf{y},\boldsymbol{\theta}) \end{pmatrix}$$

we only need to show that

$$\mathbb{E}(\mathbf{z}\mathbf{z}^T) = \begin{pmatrix} \mathcal{M}(\hat{\boldsymbol{\theta}}) & -\mathbf{I} \\ -\mathbf{I} & \mathbf{J} \end{pmatrix}.$$

Under (2),

$$\begin{aligned}
\mathbb{E}(\hat{\theta}_i(\mathbf{y})\frac{\partial \ln p(\mathbf{y},\boldsymbol{\theta})}{\partial \theta_j}) &= \int \hat{\theta}_i(\mathbf{y})\frac{\partial p(\mathbf{y},\boldsymbol{\theta})}{\partial \theta_j} d\boldsymbol{\theta} d\mathbf{y} = \int \frac{\partial p(\mathbf{y},\theta_j)}{\partial \theta_j} d\theta_i d\mathbf{y} = 0 \\
\mathbb{E}(\theta_i \frac{\partial \ln p(\mathbf{y},\boldsymbol{\theta})}{\partial \theta_j}) &= \int \theta_i \frac{\partial p(\mathbf{y},\boldsymbol{\theta})}{\partial \theta_j} d\boldsymbol{\theta} d\mathbf{y} = \int \theta_i \frac{\partial p(\mathbf{y},\theta_j)}{\partial \theta_j}) d\theta_i d\theta_j d\mathbf{y} \\
&= \begin{cases} -1 & \text{i=j} \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

where we used the integration by parts under (1) in the last equality. The equality holds iff

$$\nabla_\theta \ln p(\mathbf{y},\boldsymbol{\theta}) = \mathbf{K}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \rightarrow \nabla_\theta^2 \ln p(\mathbf{y},\boldsymbol{\theta}) = -\mathbf{K}$$

Equivalently,

$$\nabla_\theta^2 \ln p(\boldsymbol{\theta}|\mathbf{y}) = -\mathbf{K}$$

which implies that $p(\boldsymbol{\theta}|\mathbf{y})$ is Gaussian.

Extension to Proper Complex Data Model

Complex Data Representation:

$$\mathbf{x} = \mathbf{x}_R + j\mathbf{x}_I \Leftrightarrow \mathbf{y} = \begin{pmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{pmatrix}$$

Proper Complex Random Vector:

$\mathbf{x}$ is proper or symmetrical if

$$cov(\mathbf{x}\mathbf{x}^t) = \mathbf{0} \;\Rightarrow\; \begin{cases} cov(\mathbf{x}_R, \mathbf{x}_R^t) = cov(\mathbf{x}_I, \mathbf{x}_I^t) \\ cov(\mathbf{x}_R, \mathbf{x}_I^t) = -cov(\mathbf{x}_I, \mathbf{x}_R^t) \end{cases}$$

$$\Rightarrow\; cov(\mathbf{y}\mathbf{y}^t) = \begin{pmatrix} cov(\mathbf{x}_R, \mathbf{x}_R) & cov(\mathbf{x}_R, \mathbf{x}_I) \\ -cov(\mathbf{x}_R, \mathbf{x}_I) & cov(\mathbf{x}_R, \mathbf{x}_R) \end{pmatrix}$$

Note:

- If $\mathbf{x}$ is proper, then all second-order statistics of $\mathbf{x}$ is contained in $cov(\mathbf{x}, \mathbf{x}^H)$.

$$\begin{aligned} cov(\mathbf{x}, \mathbf{x}^H) &= cov(\mathbf{x}_R, \mathbf{x}_R^t) + cov(\mathbf{x}_I, \mathbf{x}_I^t) \\ &\quad -j(cov(\mathbf{x}_R, \mathbf{x}_I^t) - cov(\mathbf{x}_I, \mathbf{x}_R^t)) \\ &= 2cov(\mathbf{x}_R, \mathbf{x}_R^t) + 2jcov(\mathbf{x}_I, \mathbf{x}_R^t) \end{aligned}$$

- If $\mathbf{x}$ is proper, then $\mathbf{A}\mathbf{x} + \mathbf{b}$ is also symmetrical (invariant under affine transforms).
- The properness of the random vector enables the use of complex arithmetics at lower dimension by changing transpose to Hermitian.

Complex Gaussian Random Vectors

Random vector $\mathbf{x}$ is complex Gaussian if (i) $\mathbf{x}$ is proper and (ii) $\begin{pmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{pmatrix}$ is Gaussian.

- Distribution: $\mathbf{x} \sim \mathcal{N}_c(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ implies

$$\begin{aligned} \mathbb{E}(\mathbf{x}) &= \boldsymbol{\mu}, cov(\mathbf{x}, \mathbf{x}^H) = \boldsymbol{\Sigma}, \\ p(\mathbf{x}) &= \frac{1}{\pi^n |\boldsymbol{\Sigma}|} exp\{-(\mathbf{x} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\}. \end{aligned}$$

- $\mathbf{x}$ and $\begin{pmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{pmatrix}$ have the same PDF.
- Fourth-order moments:

$$\mathbb{E}(x_1^* x_2 x_3^* x_4) = E(x_1^* x_2) E(x_3^* x_4) + E(x_1^* x_4) E(x_3^* x_2)$$

  Note the difference from the real case.

- $\mathbf{x}$ inherits all the properties of a real Gaussian random vector:
  - Any sub-vector of $\mathbf{x}$ is complex Gaussian.
  - Affine transform: $\mathbf{A}\mathbf{x} + \mathbf{b} \sim \mathcal{N}_c(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^H)$.
  - If $\mathbf{x} = \begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix} \sim \mathcal{N}(\begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_z \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yz} \\ \boldsymbol{\Sigma}_{zy} & \boldsymbol{\Sigma}_{zz} \end{pmatrix})$, then $p(\mathbf{y}|\mathbf{z})$ is complex Gaussian with

$$\begin{aligned} \mathbb{E}(\mathbf{y}|\mathbf{z}) &= \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yz} \boldsymbol{\Sigma}_{zz}^{-1} (\mathbf{z} - \boldsymbol{\mu}_z) \\ cov(\mathbf{y}, \mathbf{y}^H | \mathbf{z}) &= \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yz} \boldsymbol{\Sigma}_{zz}^{-1} \boldsymbol{\Sigma}_{zy} \end{aligned}$$

The CRLB for the Complex Gaussian Case

Let $\mathbf{y} \sim \mathcal{N}_c(\boldsymbol{\mu}(\theta), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$, and $\boldsymbol{\theta}$ is real. Then

$$
\begin{aligned}
[\mathbf{I}(\boldsymbol{\theta})]_{ij} &= 2Re[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i}]^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_j}] \\
&\quad + tr\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_j}\}
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i} &= \begin{pmatrix} \frac{\partial \mu_1(\theta)}{\partial \theta_i} \\ \vdots \\ \frac{\partial \mu_n(\theta)}{\partial \theta_i} \end{pmatrix} \\
\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_k} &= [\frac{\partial}{\partial \theta_k}(Re\{\Sigma_{ij}(\boldsymbol{\theta})\} + jIm\{\Sigma_{ij}(\boldsymbol{\theta})\})]
\end{aligned}
$$

CRLB for Complex Parameters

Theorem:
Let $\boldsymbol{\theta} \in \mathcal{C}^K$ and proper $\mathbf{y} \in \mathcal{C}^n$ with $\mathbf{y} \sim p(\mathbf{y}; \boldsymbol{\theta})$ be proper. If $\hat{\boldsymbol{\theta}}$ is unbiased, then

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \geq \mathbf{I}_c^{-1}(\boldsymbol{\theta})$$

where $\mathbf{I}_c^{-1}(\boldsymbol{\theta})$ is the complex Fisher Information Matrix defined by

$$\mathbf{I}_c(\boldsymbol{\theta}) \stackrel{\Delta}{=} E\{[\nabla_{\theta^*} \ln p(\mathbf{y}; \boldsymbol{\theta})][\nabla_{\theta^*} \ln p(\mathbf{y}; \boldsymbol{\theta})]^H\}.$$

The proof of this theorem is identical to the real vector case.

Theorem
Let $\boldsymbol{\theta} \in \mathcal{C}^K$ and proper $\mathbf{y} \in \mathcal{C}^n$ with $\mathbf{y} \sim p(\mathbf{y}; \boldsymbol{\theta})$ be proper. If the scoring function

$$\mathbf{s}(\mathbf{y}; \boldsymbol{\theta}) \stackrel{\Delta}{=} \nabla_{\theta^*} \ln p(\mathbf{y}; \boldsymbol{\theta})$$

is proper, then the complex CRLB matches the real CRLB defined by the model

$$\tilde{\mathbf{y}} = \begin{pmatrix} \Re(\mathbf{y}) \\ \Im(\mathbf{y}) \end{pmatrix}, \tilde{\boldsymbol{\theta}} = \begin{pmatrix} \Re(\boldsymbol{\theta}) \\ \Im(\boldsymbol{\theta}) \end{pmatrix}, \quad \tilde{\mathbf{y}} \sim \tilde{p}(\tilde{\mathbf{y}}; \tilde{\boldsymbol{\theta}})$$

Proof of the Equivalence

For a general vector $\mathbf{a}$, we use the notation $\mathbf{a}_R = \Re(\mathbf{a}), \mathbf{a}_I = \Im(\mathbf{a})$.

- From the definition of complex gradient

$$\mathbf{s}(\mathbf{y};\boldsymbol{\theta}) \triangleq \frac{1}{2}[\nabla_{\theta_R} \ln p(\mathbf{x};\boldsymbol{\theta}) - j\nabla_{\theta_I} \ln p(\mathbf{x};\boldsymbol{\theta})],$$

  we have, for proper $\mathbf{s}(\mathbf{y};\boldsymbol{\theta})$,

$$\begin{aligned} \mathbb{E}(\mathbf{s}_R(\mathbf{y};\boldsymbol{\theta}\mathbf{s}_R^T(\mathbf{y};\boldsymbol{\theta})) &= \mathbb{E}(\mathbf{s}_I(\mathbf{y};\boldsymbol{\theta}\mathbf{s}_I^T(\mathbf{y};\boldsymbol{\theta})) \\ \mathbb{E}(\mathbf{s}_R(\mathbf{y};\boldsymbol{\theta}\mathbf{s}_I^T(\mathbf{y};\boldsymbol{\theta})) &= -\mathbb{E}(\mathbf{s}_I(\mathbf{y};\boldsymbol{\theta}\mathbf{s}_R^T(\mathbf{y};\boldsymbol{\theta})) \end{aligned}$$

  Hence

$$\begin{aligned} \mathbf{I}_c(\boldsymbol{\theta}) &= \mathbb{E}\{\mathbf{s}(\mathbf{y};\boldsymbol{\theta})\mathbf{s}^H(\mathbf{y};\boldsymbol{\theta})\} \\ &= 2\mathbb{E}\{\mathbf{s}_R(\mathbf{y};\boldsymbol{\theta})\mathbf{s}_R^T(\mathbf{y};\boldsymbol{\theta})\} + 2j\mathbb{E}\{\mathbf{s}_I(\mathbf{y};\boldsymbol{\theta})\mathbf{s}_R^T(\mathbf{y};\boldsymbol{\theta})\} \\ &= \frac{1}{2}\mathbf{I}_1(\tilde{\boldsymbol{\theta}}) + \frac{j}{2}\mathbf{I}_2(\tilde{\boldsymbol{\theta}}) \end{aligned}$$

  where

$$\begin{aligned} \mathbf{I}_1(\tilde{\boldsymbol{\theta}}) &\triangleq \mathbb{E}\{\nabla_{\theta_R} \ln p(\mathbf{y};\boldsymbol{\theta})\nabla_{\theta_R}^T \ln p(\mathbf{y};\boldsymbol{\theta})\} \\ \mathbf{I}_2(\tilde{\boldsymbol{\theta}}) &\triangleq \mathbb{E}\{\nabla_{\theta_I} \ln p(\mathbf{y};\boldsymbol{\theta})\nabla_{\theta_R}^T \ln p(\mathbf{y};\boldsymbol{\theta})\} \end{aligned}$$

- On the other hand, the (real) Fisher information matrix has the form

$$\mathbf{I}(\tilde{\boldsymbol{\theta}}) = \begin{pmatrix} \mathbf{I}_1(\tilde{\boldsymbol{\theta}}) & -\mathbf{I}_2(\tilde{\boldsymbol{\theta}}) \\ \mathbf{I}_2(\tilde{\boldsymbol{\theta}}) & \mathbf{I}_1(\tilde{\boldsymbol{\theta}}) \end{pmatrix}$$

  The real CRLB has the following form

$$\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}}) = \begin{pmatrix} \mathbf{E} & -\mathbf{F} \\ \mathbf{F} & \mathbf{E} \end{pmatrix}$$

- To relate the two CRLBs,

$$\mathbf{I}_c(\boldsymbol{\theta})(\mathbf{E} + j\mathbf{F}) = \frac{1}{2}\mathbf{I}$$

  Hence

$$\mathbf{I}_c^{-1}(\boldsymbol{\theta}) = 2(\mathbf{E} + j\mathbf{F}) \rightarrow [\mathbf{I}_c^{-1}(\boldsymbol{\theta})]_{ii} = 2[\mathbf{I}(\tilde{\boldsymbol{\theta}})]_{ii}$$

Summary and Preview

Summary

- The CRLB:

$$\text{Cov}(\hat{\boldsymbol{\theta}}) \geq \mathbf{I}^{-1}(\boldsymbol{\theta})$$

  where $\mathbf{I}(\theta)$ is the Fisher Information Matrix– more information about the model, the better the estimate.

- The estimator achieving CRLB is an *efficient* UMVU. In general, UMVU may not be efficient.

- The use of Schur complement to prove variance bound.

- If efficient UMVU exists, it is maximum likelihood:

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

Preview

- The Maximum Likelihood Estimators