2 Estimation Theory
2.2: C-R Bound and Efficient Estimator

Estimation Theory

2.1

1. **Introduction to Estimation Theory**
    - Assessing Estimator Performance
    - Minimum Variance Unbiased Estimator
    - Function Estimation
2. **Cramer-Rao Bound and Efficient Estimator**
    - Cramer-Rao Bound
    - Examples
3. **Maximum Likelihood & Maximum a Posteriori Estimator**
    - Classical estimation: Maximum Likelihood Estimator
    - The Bayesian framework: Maximum a Posteriori Estimator

Cramer-Rao Bound and Efficient Estimator

1. **Introduction**
   - Minimum variance unbiased estimator
2. **Cramer-Rao Bound for parameters**
   - Likelihood function
   - Concept of efficient estimator
   - Examples
3. **Cramer-Rao Bound for vector parameters**
   - Fisher information matrix
   - Examples

MVU Estimator (I)

In the previous unit, we have been able to find the **Minimum Variance Unbiased Estimator** for the estimation of the mean value of a signal ($X[n]$), that can be modeled as a constant value (θ) embedded in zero-mean noise ($W[n]$).

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

$$\hat{\theta}_N = \underline{\mathrm{h}}^T \underline{\mathrm{x}} \quad \Rightarrow \quad \text{Unbiased if: } \underline{\mathrm{h}}^T \underline{1} = 1 \qquad \Rightarrow \qquad \sigma^2_{\hat{\theta}_N} = \underline{\mathrm{h}}^T \underline{\underline{\mathrm{R}}}_w \underline{\mathrm{h}}$$

To obtain the **Minimum Variance Unbiased (MVU)** estimator, we have solved the following problem of optimization with constraints :

$$\begin{array}{l} \min\limits_{\underline{\mathrm{h}}} \left(\underline{\mathrm{h}}^T \underline{\underline{\mathrm{R}}}_w \underline{\mathrm{h}}\right) \\ \text{subject to } \underline{\mathrm{h}}^T \underline{1} = 1 \end{array} \Rightarrow \quad \mathcal{L}(\underline{\mathrm{h}}, \lambda) = \left(\underline{\mathrm{h}}^T \underline{\underline{\mathrm{R}}}_w \underline{\mathrm{h}}\right) - \lambda\left(\underline{\mathrm{h}}^T \underline{1} - 1\right) \quad \Rightarrow \quad \underline{\mathrm{h}} = \frac{\underline{\underline{\mathrm{R}}}_w^{-1} \underline{1}}{\underline{1}^T \underline{\underline{\mathrm{R}}}_w^{-1} \underline{1}}$$

MVU Estimator (II)

- Prove that the resulting estimator is **unbiased** and compute the **minimum variance** that it achieves.

$$\hat{\theta}_N = \frac{\underline{1}^T \underline{\underline{R}}_w^{-1} \underline{x}}{\underline{1}^T \underline{\underline{R}}_w^{-1} \underline{1}}$$

Unbiased

$\underline{h}^T \underline{1} = 1$

$$\underline{h}^T \underline{1} = \left(\frac{\underline{\underline{R}}_w^{-1} \underline{1}}{\underline{1}^T \underline{\underline{R}}_w^{-1} \underline{1}} \right)^T \underline{1} = [\underline{\underline{R}}_w^{-1} = (\underline{\underline{R}}_w^{-1})^T] = \frac{\underline{1}^T \underline{\underline{R}}_w^{-1} \underline{1}}{\underline{1}^T \underline{\underline{R}}_w^{-1} \underline{1}} = 1$$

Variance

$\sigma^2_{\hat{\theta}_N} = \underline{h}^T \underline{\underline{R}}_w \underline{h}$ $\Rightarrow$

$$\sigma^2_{\hat{\theta}_N} = \left(\frac{\underline{\underline{R}}_w^{-1} \underline{1}}{\underline{1}^T \underline{\underline{R}}_w^{-1} \underline{1}} \right)^T \underline{\underline{R}}_w \frac{\underline{\underline{R}}_w^{-1} \underline{1}}{\underline{1}^T \underline{\underline{R}}_w^{-1} \underline{1}} = \frac{\underline{1}^T \underline{\underline{R}}_w^{-1}}{\underline{1}^T \underline{\underline{R}}_w^{-1} \underline{1}} \underline{\underline{R}}_w \frac{\underline{\underline{R}}_w^{-1} \underline{1}}{\underline{1}^T \underline{\underline{R}}_w^{-1} \underline{1}}$$

$$\sigma^2_{\hat{\theta}_N} = \frac{\underline{1}^T \underline{\underline{R}}_w^{-1}}{\underline{1}^T \underline{\underline{R}}_w^{-1} \underline{1}} \frac{\underline{1}}{\underline{1}^T \underline{\underline{R}}_w^{-1} \underline{1}}$$

$$\boxed{\sigma^2_{\hat{\theta}_N} = \frac{1}{\underline{1}^T \underline{\underline{R}}_w^{-1} \underline{1}}}$$

- Evaluate the results for the case of **stationary, white noise**.

$$\underline{\underline{R}}_w = \sigma_w^2 \underline{\underline{I}}$$

$$\underline{\underline{R}}_w^{-1} = \frac{1}{\sigma_w^2} \underline{\underline{I}}$$

$\Rightarrow$

$$\boxed{\hat{\theta}_N = \frac{1}{N} \underline{1}^T \underline{x}} \qquad \boxed{\sigma^2_{\hat{\theta}_N} = \frac{\sigma_w^2}{N}}$$

Cramer-Rao Bound and Efficient Estimator

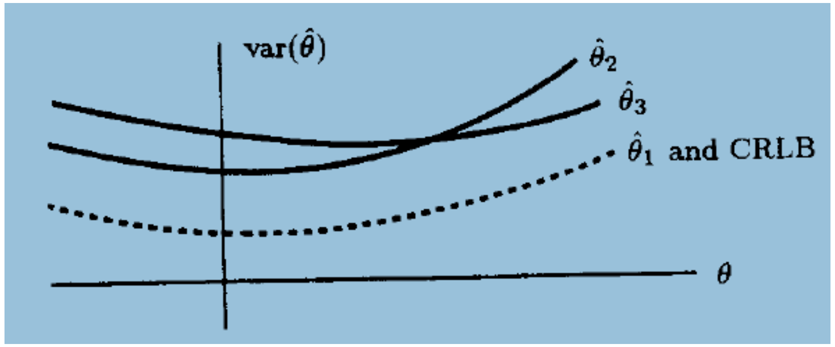
2.1

1. Introduction
    - Minimum variance unbiased estimator
2. **Cramer-Rao Bound for parameters**
    - Likelihood function
    - Concept of efficient estimator
    - Examples
3. **Cramer-Rao Bound for vector parameters**
    - Fisher information matrix
    - Examples

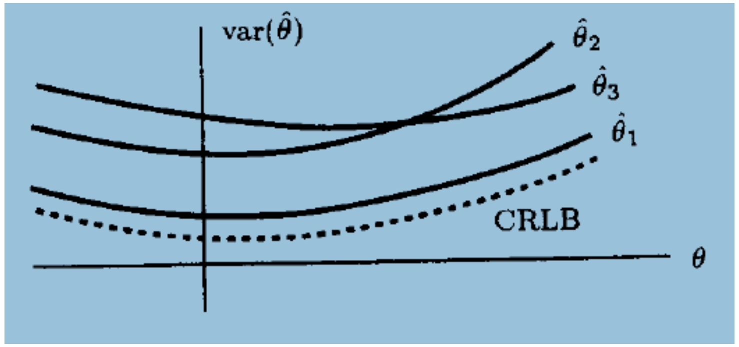
Usefulness of the C-R bound

However, there is no method that ensures that, if a MVU exists, we may be able to find it. Nevertheless, the **Cramer-Rao lower bound** (CRLB or CRB):

- Determines the minimum possible variance for any unbiased estimator:
  - This bound provides a benchmark for assessing any estimator performance
- Provides, in some cases, the expression for the MVU estimator
  - When a estimator attains the CRLB is said to be **efficient**
- Can be used to estimate the (non-linear) function of a parameter.



$\hat{\theta}_1$ is MVUE and attains the CRLB
Efficient estimator



$\hat{\theta}_1$ is MVUE and does not attain the CRLB
Not efficient estimator

Steven M. Kay, Fundamentals of Statistical Signal Processing, Prentice Hall Int. Ed.

Cramer-Rao Lower Bound

2.1

There exists a **lower bound** of the variance of the whole set of unbiased estimators of a parameter θ.

The bound is related to the **probability density function** of the data:

- When the pdf is viewed as a function of the unknown parameters (with $\underline{\mathbf{x}}$ fixed), it is termed the likelihood function:

$$f_{\underline{\mathbf{x}}}(x[1], x[2], \cdots, x[N]; \theta) = f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}; \theta)$$

Cramer-Rao Lower Bound

The **variance of any unbiased estimator** $\hat{\theta}$ must satisfy:

$$\operatorname{var}(\hat{\theta}) \geq \frac{1}{-E\left\{\frac{\partial^2 \ln f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}; \theta)}{\partial \theta^2}\right\}}$$

And the **equality is satisfied** when, for some function $k(\theta)$:

$$\frac{\partial \ln f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}; \theta)}{\partial \theta} = k(\theta)\left(\hat{\theta}_{opt}(\underline{\mathbf{x}}) - \theta\right)$$

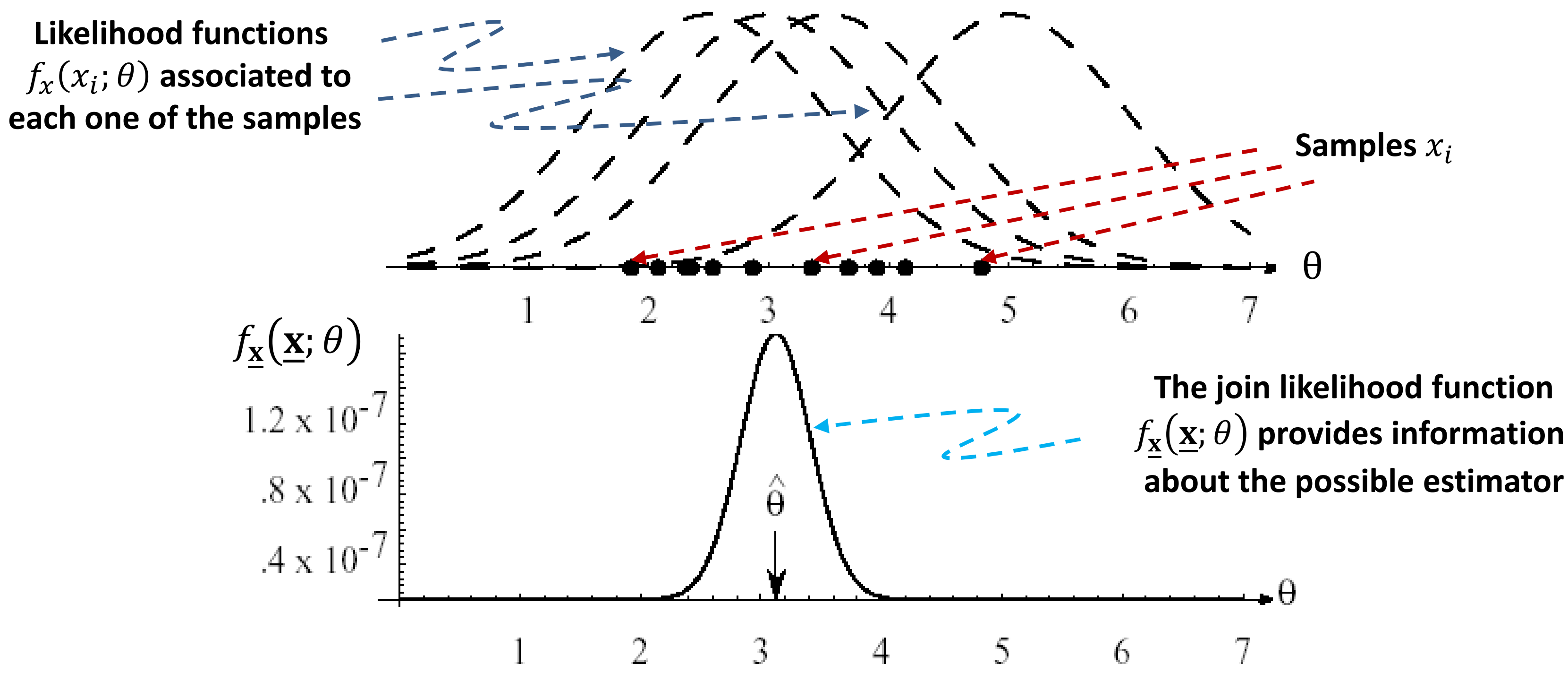
Likelihood function

2.1

Interpretation of the likelihood function:

Let us analyze the case of the likelihood function ($f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}};\theta)$) of a set of N Gaussian, independent samples:

$$f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}};\theta) = \prod_{i=1}^{N} f_x(x_i;\theta)$$



CRLB and the likelihood function

Interpretation of the CRLB:

The more informative the set of samples ($\underline{\mathbf{x}}$), the **sharper** the likelihood function ($f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}};\theta)$):

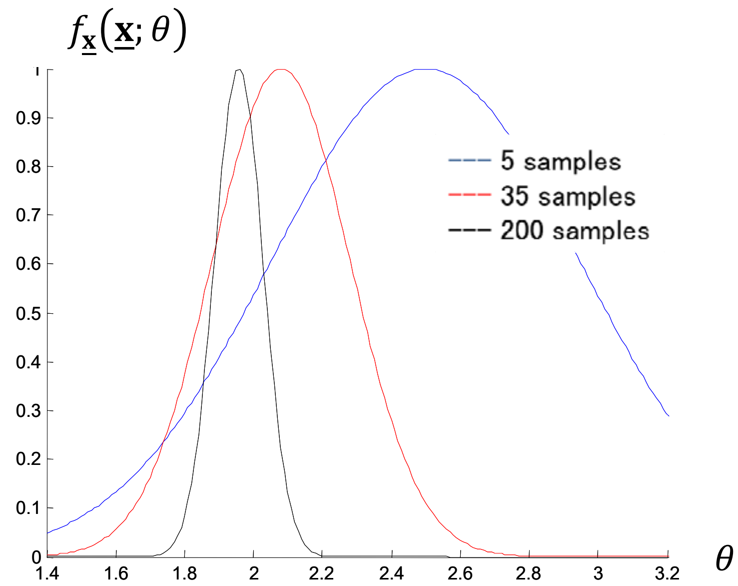
- A measure of sharpness its the **curvature**

The **curvature** is the negative of the second derivative of the log-likelihood function.

The larger the curvature, the smaller the Cramer-Rao bound on the variance.

The curvature depends on both:

- The number of samples (N) and
- The likelihood function ($f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}};\theta)$)



$$\mathrm{var}\left(\hat{\theta}\right) \geq \frac{1}{-E\left\{\dfrac{\partial^2 \ln f_{\underline{\mathbf{x}}}\left(\underline{\mathbf{x}};\theta\right)}{\partial\theta^2}\right\}}$$

CRLB and efficient estimators

2.1

- The **optimal (efficient) estimator** can be obtained through the condition of minimum variance:

$$\frac{\partial \ln f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}};\theta)}{\partial\theta} = k(\theta)\left(\hat{\theta}_{opt}(\underline{\mathbf{x}}) - \theta\right)$$

$$\hat{\theta}_{opt}(\underline{\mathbf{x}}) = \frac{1}{k(\theta)}\left(k(\theta)\theta + \frac{\partial \ln f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}};\theta)}{\partial\theta}\right)$$

◀ For the estimator to be feasible, the dependency with θ should cancel

- The **achieved minimum variance** is given by:

$$\mathrm{var}_{opt}(\hat{\theta}) = \frac{1}{k(\theta)}$$

- The denominator in the CRLB is referred to as the **Fisher information** $I(\theta)$:

$$I(\theta) = -E\left\{\frac{\partial^2 \ln f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}};\theta)}{\partial\theta^2}\right\} = E\left\{\left(\frac{\partial \ln f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}};\theta)}{\partial\theta}\right)^2\right\}$$

Cramer-Rao Bound and Efficient Estimator

2.1

1. **Introduction**
   - Minimum variance unbiased estimator
2. **Cramer-Rao Bound for parameters**
   - Likelihood function
   - Concept of efficient estimator
   - Examples
3. **Cramer-Rao Bound for vector parameters**
   - Fisher information matrix
   - Examples

Example (I)

2.1

- Given N samples of a process that can modeled as $\underline{\mathbf{x}} = \theta\underline{\mathbf{1}} + \underline{\mathbf{w}}$, compute and efficient estimator of its mean (θ).

  Note: $W[n]$ is a **Gaussian, stationary, white noise**.

Generic expression of a multivariate Gaussian ▶

$$f_{\underline{w}}(\underline{\mathbf{w}}) = \frac{1}{\sqrt{(2\pi)^N \left|\underline{\underline{\mathbf{C}}}_{\underline{w}}\right|}} \exp\left[-\frac{\left[\underline{\mathbf{w}} - \underline{\mathbf{m}}_{\underline{w}}\right]^T \underline{\underline{\mathbf{C}}}_{\underline{w}}^{-1} \left[\underline{\mathbf{w}} - \underline{\mathbf{m}}_{\underline{w}}\right]}{2}\right]$$

STATIONARY WHITE NOISE $\Rightarrow E\{w[n]\} = 0 \quad r_w[l] = \sigma_w^2 \delta[l]$

$$\underline{\underline{C}}_w = \left[E\{w[n]\} = 0\right] = \underline{\underline{R}}_w = \left[\text{WHITE, STAT.}\right] = \sigma_w^2 \underline{\underline{I}}$$

AS $\underline{x} = \theta \underline{1} + \underline{w} \Rightarrow \underline{w} = \underline{x} - \theta \underline{1} \leftarrow \begin{cases} \underline{w} : N(\underline{0}, \underline{\underline{R}}_w) \\ \underline{x} : N(\underline{\theta}, \underline{\underline{R}}_w) \end{cases}$

Example (I)

2.1

$$f_{\underline{w}}(\underline{x};\theta) = \frac{1}{\left[(2\pi)^N \sigma_w^{2N}\right]^{1/2}} \cdot \exp\left[-\frac{[\underline{x}-\underline{\theta}]^T \frac{1}{\sigma_w^2} \underline{\underline{I}} [\underline{x}-\underline{\theta}]}{2}\right] =$$

$$f(\underline{x};\theta) = \frac{1}{\left[2\pi\sigma_w^2\right]^{N/2}} \cdot \exp\left[-\frac{[\underline{x}-\underline{\theta}]^T[\underline{x}-\underline{\theta}]}{2\sigma_w^2}\right]$$

$$L(\underline{x};\theta) = \ln f(\underline{x},\theta) = -\frac{N}{2}\ln 2\pi\sigma_w^2 - \frac{[\underline{x}-\underline{\theta}]^T[\underline{x}-\underline{\theta}]}{2\sigma_w^2}$$

$$\frac{\partial L(\underline{x};\theta)}{\partial\theta} = \frac{\partial}{\partial\theta}\left[-\frac{1}{2\sigma_w^2}\left(\underline{x}^T\cdot\underline{x} - \underline{x}^T\cdot\theta\cdot\underline{1} - \theta\,\underline{1}^T\underline{x} + \theta^2\,\underline{1}^T\underline{1}\right]\right] =$$

$$= -\frac{1}{2\sigma_w^2}\left[-\underline{x}^T\cdot\underline{1} - \underline{1}^T\cdot\underline{x} + 2\theta N\right] = \frac{N}{\sigma_w^2}\left[\frac{1}{N}\underline{1}^T\cdot\underline{x} - \theta\right]$$

Example (I)

2.1

$$= -\frac{1}{2\sigma_w^2}\left[-\underline{x}^T\cdot\underline{1} - \underline{1}^T\cdot\underline{x} + 2\theta N\right] = \frac{N}{\sigma_w^2}\left[\frac{1}{N}\underline{1}^T\cdot\underline{x} - \theta\right]$$

$$\frac{\partial L(\underline{x};\theta)}{\partial\theta} = \frac{1}{\sigma^2_{OPT}(\hat{\theta})}\cdot\left[\hat{\theta}_{OPT}(\underline{x}) - \theta\right] \Leftarrow$$

MATCHING TERMS

$$\hat{\theta}_{OPT}(\underline{x}) = \frac{1}{N}\underline{1}^T\underline{x} \qquad \sigma^2_{OPT} = \frac{\sigma_w^2}{N}$$

$$-E\left\{\frac{\partial^2 \ln f(\underline{x};\theta)}{\partial\theta^2}\right\} = -E\left\{\frac{\partial}{\partial\theta}\left(\frac{N}{\sigma_w^2}\left[\frac{1}{N}\underline{1}^T\cdot\underline{x} - \theta\right]\right)\right\} =$$

$$= -E\left\{-\frac{N}{\sigma_w^2}\right\} = \frac{N}{\sigma_w^2} = \frac{1}{\sigma^2_{OPT}} \Rightarrow \text{EFFICIENT}$$

Example (II)

2.1

- Given N samples of a process that can modeled as $\underline{\mathbf{x}} = \theta \underline{\mathbf{1}} + \underline{\mathbf{w}}$, compute and efficient estimator of its mean (θ).

  Note: $W[n]$ is a **Gaussian, stationary, colored noise**.

Generic expression of a multivariate Gaussian ▶

$$f_{\underline{w}}(\underline{\mathbf{w}}) = \frac{1}{\sqrt{(2\pi)^N \left|\underline{\underline{\mathbf{C}}}_{\underline{w}}\right|}} \exp\left[-\frac{\left[\underline{\mathbf{w}} - \underline{\mathbf{m}}_{\underline{w}}\right]^T \underline{\underline{\mathbf{C}}}_{\underline{w}}^{-1} \left[\underline{\mathbf{w}} - \underline{\mathbf{m}}_{\underline{w}}\right]}{2}\right]$$

STATIONARY
COLORED NOISE $\Rightarrow E\{w[n]\} = \emptyset \quad r_w[l] \neq \sigma_w^2 \delta[l$

$$\underline{\underline{C}}_w = \left[E\{w[n]\} = \emptyset\right] = \underline{\underline{R}}_w = \cancel{\left[\text{WHITE, STAT.}\right] = \sigma_w^2 \underline{\underline{I}}}$$

AS $\underline{x} = \theta \underline{1} + \underline{w} \Rightarrow \underline{w} = \underline{x} - \theta \underline{1} \leftarrow \begin{cases} \underline{w} : N(\underline{\emptyset}, \underline{\underline{C}}_w) \\ \underline{x} : N(\underline{\theta}, \underline{\underline{C}}_w) \end{cases}$

Example (II)

2.1

$$f_{\underline{w}}(\underline{x};\theta) = \frac{1}{\left[(2\pi)^N |\underline{\underline{C}}_w|\right]^{1/2}} \cdot \exp\left[-\frac{1}{2}\left[\underline{x}-\theta\underline{1}\right]^T \underline{\underline{C}}_w^{-1}\left[\underline{x}-\theta\underline{1}\right]\right]$$

$$L(\underline{x};\theta) = \ln f(\underline{x};\theta) = -\frac{1}{2}\ln(2\pi)^N |\underline{\underline{C}}_w| - \frac{1}{2}\left[\underline{x}-\theta\underline{1}\right]^T \underline{\underline{C}}_w^{-1}\left[\underline{x}-\theta\underline{1}\right]$$

$$\frac{\partial L(\underline{x};\theta)}{\partial\theta} = -\frac{1}{2}\frac{\partial}{\partial\theta}\left[\underline{x}^T\underline{\underline{C}}_w^{-1}\underline{x} - \underline{x}^T\underline{\underline{C}}_w^{-1}\theta\underline{1} - \theta\underline{1}^T\underline{\underline{C}}_w^{-1}\underline{x} + \theta^2\underline{1}^T\underline{\underline{C}}_w^{-1}\underline{1}\right] =$$

$$\frac{\partial L(\underline{x};\theta)}{\partial\theta} = \frac{1}{2}\left[\underline{x}^T\underline{\underline{C}}_w^{-1}\underline{1} + \underline{1}^T\underline{\underline{C}}_w^{-1}\underline{x} - 2\theta\underline{1}^T\underline{\underline{C}}_w^{-1}\underline{1}\right] =$$

$$= \left[\underline{x}^T\underline{\underline{C}}_w^{-1}\underline{1} = [\text{esc}] = \left[\underline{x}^T\underline{\underline{C}}_w^{-1}\underline{1}\right]^T = \underline{1}^T\left[\underline{x}^T\underline{\underline{C}}_w^{-1}\right]^T = \underline{1}^T\underline{\underline{C}}_w^{-1}\underline{x}\right] =$$

$$\frac{\partial L(\underline{x};\theta)}{\partial\theta} = \underline{1}^T\underline{\underline{C}}_w^{-1}\underline{x} - \theta\,\underline{1}^T\underline{\underline{C}}_w^{-1}\underline{1} = \underline{1}^T\underline{\underline{C}}_w^{-1}\underline{1}\left[\frac{\underline{1}^T\underline{\underline{C}}_w^{-1}\underline{x}}{\underline{1}^T\underline{\underline{C}}_w\underline{1}} - \theta\right]$$

Example (II)

$$\frac{\partial L(\underline{x};\theta)}{\partial\theta} = \underline{1}^T \underline{C}_w^{-1} \underline{x} - \theta\, \underline{1}^T \underline{C}_w^{-1} \underline{1} = \underline{1}^T \underline{C}_w^{-1} \underline{1} \left[\frac{\underline{1}^T \underline{C}_w^{-1} \underline{x}}{\underline{1}^T \underline{C}_w^{-1} \underline{1}} - \theta \right]$$

$$\frac{\partial L(\underline{x};\theta)}{\partial\theta} = \frac{1}{\sigma^2_{OPT}(\hat{\theta})} \cdot \left[\hat{\theta}_{OPT}(\underline{x}) - \theta \right] \Leftarrow$$ MATCHING TERMS

$$\hat{\theta}_{OPT}(\underline{x}) = \frac{\underline{1}^T \underline{C}_w^{-1} \underline{x}}{\underline{1}^T \underline{C}_w^{-1} \underline{1}} \qquad \sigma^2_{OPT} = \frac{1}{\underline{1}^T \underline{C}_w^{-1} \underline{1}}$$

$$-E\left\{ \frac{\partial^2 \ln f(\underline{x};\theta)}{\partial\theta^2} \right\} = -E\left\{ \frac{\partial}{\partial\theta} \left(\underline{1}^T \underline{C}_w^{-1} \underline{1} \left[\frac{\underline{1}^T \underline{C}_w^{-1} \underline{x}}{\underline{1}^T \underline{C}_w^{-1} \underline{1}} - \theta \right] \right) \right\}$$

$$= -E\left\{ \underline{1}^T \underline{C}_w^{-1} \underline{1} \right\} = \frac{1}{\sigma^2_{OPT}} \Rightarrow$$ EFFICIENT

Cramer-Rao Bound and Efficient Estimator

2.1

1. **Introduction**
   - Minimum variance unbiased estimator
2. **Cramer-Rao Bound for parameters**
   - Likelihood function
   - Concept of efficient estimator
   - Examples
3. **Cramer-Rao Bound for vector parameters**
   - Fisher information matrix
   - Examples

CRLB for vector parameters

2.1

The extension to the case of a vector parameter $\underline{\boldsymbol{\theta}}$ is as follows:

$$f_{\underline{\mathbf{x}}}(x[1], x[2], \cdots, x[N];\ \theta_1, \theta_2, \cdots, \theta_P) = f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}; \underline{\boldsymbol{\theta}})$$

Cramer-Rao Lower Bound for vector parameters

The **variance of any unbiased estimator** $\hat{\theta}_i$ must satisfy:

$$\boxed{\operatorname{var}(\hat{\theta}_i) \geq \left[\underline{\underline{\mathbf{I}}}^{-1}(\underline{\boldsymbol{\theta}})\right]_{ii}}$$

where $\underline{\underline{\mathbf{I}}}(\underline{\boldsymbol{\theta}})$ is the $p \times p$ **Fisher information matrix**:

$$\boxed{\left[\underline{\underline{\mathbf{I}}}(\underline{\boldsymbol{\theta}})\right]_{ij} = -E\left\{\frac{\partial^2 \ln f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}; \underline{\boldsymbol{\theta}})}{\partial\theta_i \partial\theta_j}\right\}}$$

And the **equality is satisfied** when:

$$\boxed{\nabla_{\underline{\boldsymbol{\theta}}}\left(f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}; \underline{\boldsymbol{\theta}})\right) = \underline{\underline{\mathbf{I}}}^{-1}(\underline{\boldsymbol{\theta}})\left(\underline{\boldsymbol{\theta}}_{opt}(\underline{\mathbf{x}}) - \underline{\boldsymbol{\theta}}\right)}$$

Example (I)

2.1

- Given N samples of a process that can modeled as $\underline{\mathbf{x}} = A\underline{\mathbf{1}} + \underline{\mathbf{w}}$, compute and efficient estimator of its mean (A) and variance (σ^2) .
  Note: $W[n]$ is a **Gaussian, stationary, white noise**.

Generic expression of a multivariate Gaussian ▶

$$f_{\underline{W}}(\underline{\mathbf{w}}) = \frac{1}{\sqrt{(2\pi)^N\left|\underline{\underline{\mathbf{C}}}_{\underline{W}}\right|}} \exp\left[-\frac{\left[\underline{\mathbf{w}} - \underline{\mathbf{m}}_{\underline{W}}\right]^T \underline{\underline{\mathbf{C}}}_{\underline{W}}^{-1}\left[\underline{\mathbf{w}} - \underline{\mathbf{m}}_{\underline{W}}\right]}{2}\right]$$

STATIONARY WHITE NOISE $\Rightarrow E\{w[n]\} = \emptyset \quad r_w[\ell] = \sigma^2 \delta[\ell]$

$\underline{\underline{C}}_w = \left[E\{w[n]\} = \emptyset\right] = \underline{\underline{R}}_w = \left[\text{WHITE, STAT.}\right] = \sigma^2 \underline{\underline{I}}$

AS $\underline{x} = A\underline{1} + \underline{w} \Rightarrow \underline{w} = \underline{x} - A\underline{1} \leftarrow \begin{cases} \underline{w} : N(\underline{\emptyset}, \sigma^2\underline{\underline{I}}) \\ \underline{x} : N(\underline{A}, \sigma^2\underline{\underline{I}}) \end{cases}$

Example (I)

2.1

$$f_{\underline{w}}(\underline{x};\underline{\theta}) = \frac{1}{\left[(2\pi)^N \sigma^{2N}\right]^{1/2}} \cdot \exp\left[-\frac{[\underline{x}-\underline{A}]^T \frac{1}{\sigma^2}\underline{\underline{I}}\,[\underline{x}-\underline{A}]}{2}\right] =$$

$$f(\underline{x};\underline{\theta}) = \frac{1}{[2\pi\sigma^2]^{N/2}} \cdot \exp\left[-\frac{[\underline{x}-\underline{A}]^T[\underline{x}-\underline{A}]}{2\sigma^2}\right]$$

TWO COMMENTS:

* THE PARAMETERS ARE $\underline{\theta} = [A, \sigma^2]^T$ AND THE PDF IS NOT PARAMETERIZED WITH RESPECT TO $\underline{\theta}$ DIRECTLY
* WE ARE GOING TO DEVELOP THE STUDY USING SCALAR NOTATION. (VECTOR NOTATION IS POSSIBLE AS WELL)

Example (I)

2.1

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

$$L(\underline{x};\underline{\theta}) = \ln f(\underline{x},\underline{\theta}) = -\frac{N}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x[n]-A)^2$$

LET US FIRST COMPUTE ALL PARTIAL DERIVATIVES:

$$\frac{\partial L(\underline{x};\underline{\theta})}{\partial A} = \frac{\partial}{\partial A}\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x[n]-A)^2\right] =$$

$$= \frac{1}{2\sigma^2}\, 2\sum_{n=1}^{N}(x[n]-A) \qquad (1)$$

$$\frac{\partial L(\underline{x};\underline{\theta})}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2}\left[-\frac{N}{2}\ln\sigma^2\right] - \sum_{n=1}^{N}(x[n]-A)^2\frac{\partial}{\partial\sigma^2}\left[\frac{1}{2\sigma^2}\right] =$$

$$= -\frac{N}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{n=1}^{N}(x[n]-A)^2 \qquad (2)$$

Example (I)

2.1

$$\frac{\partial^2 L(\underline{x};\underline{\theta})}{\partial A^2} = \frac{\partial}{\partial A}\left[\frac{1}{\sigma^2}\sum_{n=1}^{N}(x[n]-A)\right] = -\frac{N}{\sigma^2} \qquad (3)$$

$$\frac{\partial^2 L(\underline{x};\underline{\theta})}{\partial \sigma^2 \partial A} = \frac{\partial}{\partial \sigma^2}\left[\frac{1}{\sigma^2}\sum_{n=1}^{N}(x[n]-A)\right] = -\frac{1}{\sigma^4}\sum_{n=1}^{N}(x[n]-A) \qquad (4)$$

$$\frac{\partial^2 L(\underline{x};\underline{\theta})}{(\partial \sigma^2)^2} = \frac{\partial}{\partial \sigma^2}\left[-\frac{N}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{n=1}^{N}(x[n]-A)^2\right] =$$

$$= \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{n=1}^{N}(x[n]-A)^2 \qquad (5)$$

TO COMPUTE $\underline{I}(\underline{\theta})$, WE HAVE TO TAKE THE EXPECTATIONS OF THE NEGATIVE OF THE VARIOUS SECOND DERIVATIVES:

Example (I)

2.1

$$(3) \Rightarrow E\left\{ \frac{\partial^2 L(\underline{x};\underline{\theta})}{\partial A^2} \right\} = E\left\{ \frac{N}{\sigma^2} \right\} = \frac{N}{\sigma^2}$$

$$(4) \Rightarrow E\left\{ \frac{\partial^2 L(\underline{x};\underline{\theta})}{\partial \sigma^2 \partial A} \right\} = E\left\{ \frac{1}{\sigma^4} \sum_{n=1}^{N} (x[n]-A) \right\} = \emptyset$$

$$(5) \Rightarrow E\left\{ \frac{\partial^2 L(\underline{x};\underline{\theta})}{(\partial \sigma^2)^2} \right\} = E\left\{ -\frac{N}{2\sigma^4} + \frac{1}{\sigma^6} \sum_{n=1}^{N} [x[n]-A]^2 \right\} =$$

$$= -\frac{N}{2\sigma^4} + \frac{1}{\sigma^6} \sum_{n=1}^{N} E\left\{ [x[n]-A]^2 \right\} = -\frac{N}{2\sigma^4} + \frac{1}{\sigma^6} \sum_{n=1}^{N} \sigma^2 =$$

$$= -\frac{N}{2\sigma^4} + \frac{1}{\sigma^6} \cdot N\sigma^2 = -\frac{N}{2\sigma^4} + \frac{N}{\sigma^4} = \frac{N}{2\sigma^4}.$$

$$\Rightarrow \underline{\underline{I}}(\underline{\theta}) = \begin{bmatrix} N/\sigma^2 & \emptyset \\ \emptyset & \frac{N}{2\sigma^4} \end{bmatrix} \Leftarrow \text{DIAGONAL MATRIX}$$

Example (I)

2.1

$$\Rightarrow \underline{\underline{I}}(\underline{\theta}) = \begin{bmatrix} N/\sigma^2 & \emptyset \\ \emptyset & \frac{N}{2\sigma^4} \end{bmatrix} \Leftarrow \text{DIAGONAL MATRIX}$$

$$\Rightarrow \underline{\underline{I}}^{-1}(\underline{\theta}) = \begin{bmatrix} \sigma^2/N & \emptyset \\ \emptyset & \frac{2\sigma^4}{N} \end{bmatrix}$$

AND WE CAN SUBSTITUTE IT IN $\nabla_{\underline{\theta}} L(\underline{x};\underline{\theta})$

$$\nabla_{\underline{\theta}} L(\underline{x};\underline{\theta}) = \left(\frac{\partial L(\underline{x};\underline{\theta})}{\partial A}, \frac{\partial L(\underline{x};\underline{\theta})}{\partial \sigma^2} \right)^T$$

$$(i) \Rightarrow \frac{\partial L(\underline{x};\underline{\theta})}{\partial A} = \frac{1}{\sigma^2}\sum_{n=1}^{N}(x[n]-A) = \frac{N}{\sigma^2}\left[\frac{1}{N}\sum_{n=1}^{N}x[n]-A\right]$$

$$\hat{A}_{OPT} = \frac{1}{N}\sum_{n=1}^{N}x[n] \quad \text{WITH} \quad \sigma(\hat{A}_{OPT}) = \frac{\sigma^2}{N} = \left[\underline{\underline{I}}^{-1}\right]_{11}$$

Example (I)

2.1

$$(2) \Rightarrow \frac{\partial L(\underline{x};\underline{\theta})}{\partial \sigma^2} = -\frac{N}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{n=1}^{N}(x[n]-A)^2 =$$

$$= \frac{N}{2\sigma^4}\left[\frac{1}{N}\sum_{n=1}^{N}(x[n]-A)^2 - \sigma^2\right]$$

$$\hat{\sigma}_{OPT} = \frac{1}{N}\sum_{n=1}(x[n]-A)^2 \Leftarrow \text{NOT FEASIBLE}$$