2 Estimation Theory

2.1: Introduction to Estimation Theory

Estimation Theory

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

Introduction to Estimation Theory

1. **Introduction**
    - The relevance of model selection
2. **Assessing Estimator Performance**
    - Bias, Variance and MSE
    - Vector notation
3. **Minimum Variance Unbiased Estimator**
    - Derivative of a scalar function with respect to a vector
    - MVU: Lagrange optimization
4. **Function Estimation**
    - Autocorrelation
    - Power Spectral Density

Usefulness of Parameter Estimation

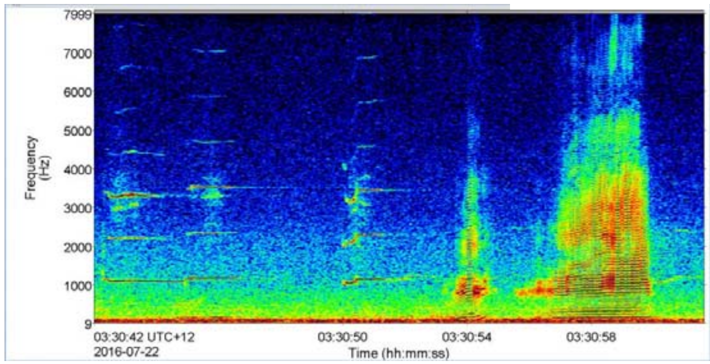
2.1

Given an N- point data set $\{x[1], x[2], \cdots, x[N]\}$ which **depends on an unknown parameter** θ (or set of parameters $\underline{\boldsymbol{\theta}}$), we wish to determine θ based on the data, through the definition of an estimator:

$$\boxed{\hat{\theta} = g(x[1], x[2], \cdots, x[N]) = g(\underline{\mathbf{x}})}$$

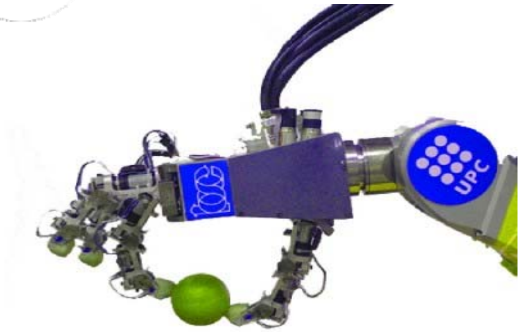
where $g(.)$ is some function.

Spectrogram and sound clip of a humpback whale



http://www.lab.upc.edu

Multifingered hand for dexterity and versatility



https://www.iri.upc.edu

Model selection (I)

2.1

The dependence of the available data ($\underline{\mathbf{x}}$) with respect to the parameters ($\underline{\boldsymbol{\theta}}$) is captured by the **model** that is proposed.

As data is random in nature, we represent it by its **probability density function** (pdf):

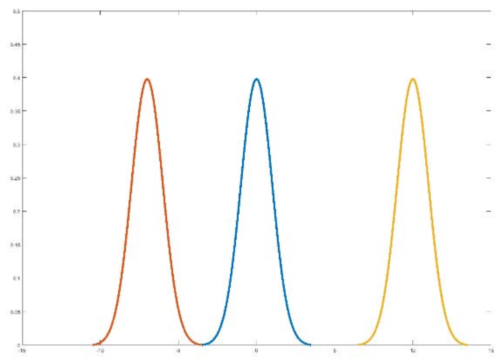
$$f_{\underline{\mathbf{x}}}\left(x[1], x[2], \cdots, x[N]; \underline{\boldsymbol{\theta}}\right) = f_{\underline{\mathbf{x}}}\left(\underline{\mathbf{x}}; \underline{\boldsymbol{\theta}}\right)$$

The **pdf is parameterized** by the unknown (vector of) parameter(s) $\underline{\boldsymbol{\theta}}$.

Case 1: We are given a pdf.

For instance, $N = 1$ ($x[1] = x$) and θ is the mean, the pdf could be:

$$f_x(x; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(x-\theta)^2}{2\sigma^2}\right]$$



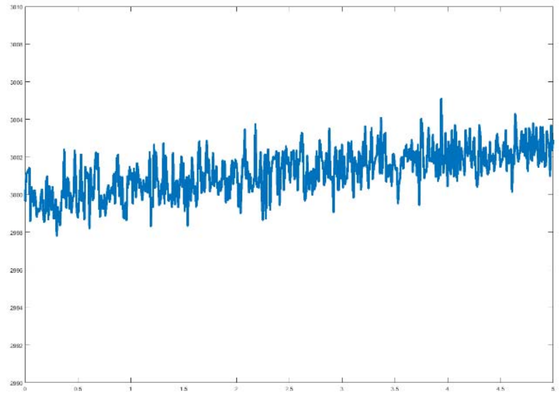
Example of three Gaussian pdf's parameterized with respect to θ

Model selection (II)

2.1

Case 2: Usually, we are given data and we have to **chose a model**:

1) Models should be **consistent** with the **problem** and **previous knowledge**
2) Models should be **mathematically tractable**



(1)

$$x[n] = A + Bn + w[n]$$

(2)

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

Case 3: **Bayessian approach**: We can assume that the parameters to be estimated are random variables (instead of deterministic but unknown):

- The knowledge about its pdf can be included

$$f(\underline{\mathbf{x}}, \underline{\boldsymbol{\theta}}) = f(\underline{\mathbf{x}} / \underline{\boldsymbol{\theta}})\; f(\underline{\boldsymbol{\theta}})$$

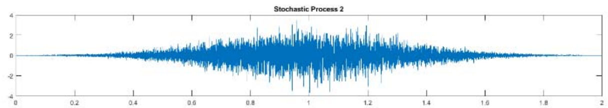
Several estimators are possible (I)

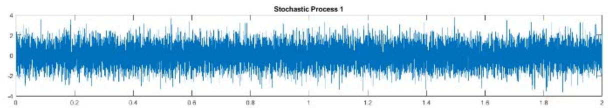
2.1

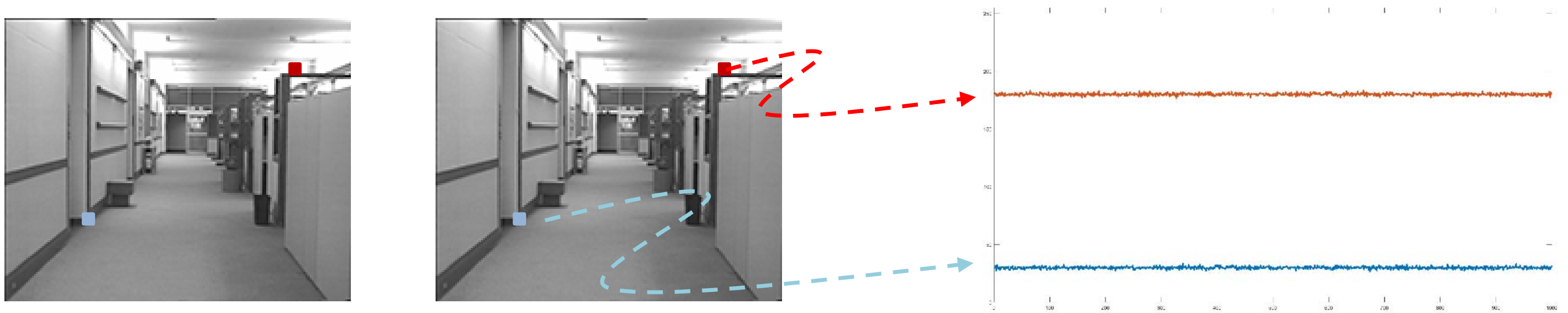
In several situations, we want to **estimate the mean value of a random process** that can be modeled as a constant value (θ) embedded in stationary white noise ($W[n]$):

- **White noise**: Each sample has a probability distribution with zero mean and finite variance, and samples are statistically independent and $r_w[n,l] = \sigma_w^2[n]\delta[l]$
- **Stationary white noise**: All variance samples have the same value and the autocorrelation function is $r_w[l] = \sigma_w^2\delta[l]$

$$\boxed{X[n] = \theta + W[n]}$$







The value of two pixels through time (n)

Several estimators are possible (II)

How can we **estimate the mean value of a random process** given a set of observations (N) of a single realization?

We can propose different estimators:

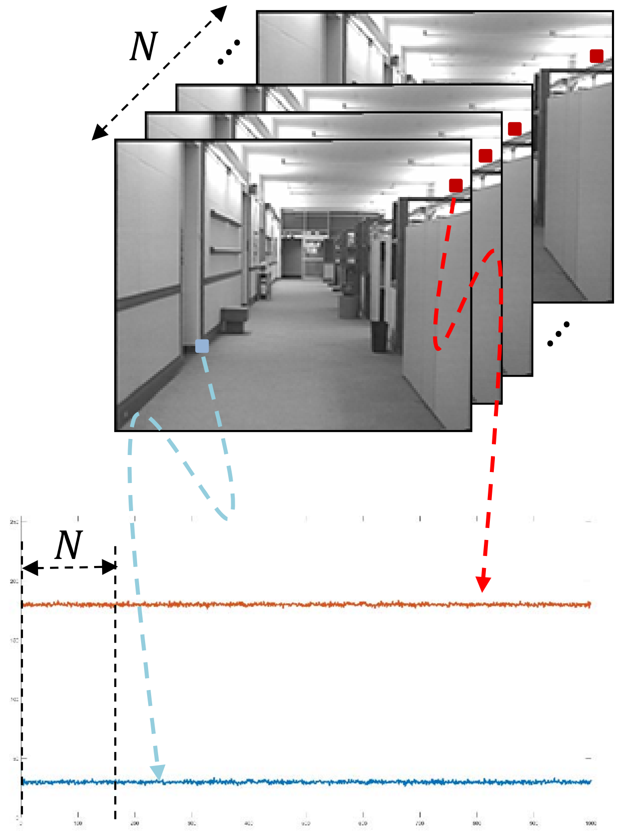
$$\hat{m}_X^{(1)} = \frac{1}{N}\sum_{n=1}^{N} x[n]$$

$$\hat{m}_X^{(2)} = \mathrm{median}(x[1], x[2], \cdots, x[N])$$

$$\hat{m}_X^{(3)} = \frac{\max(x[1], \cdots, x[N]) + \min(x[1], \cdots, x[N])}{2}$$

We need to assess **the performance of the estimators** to decide which one should be used.

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



Estimation of the mean value based on N samples

Introduction to Estimation Theory

2.1

1. **Introduction**
   - The relevance of model selection
2. **Assessing Estimator Performance**
   - Bias, Variance and MSE
   - Vector notation
3. **Minimum Variance Unbiased Estimator**
   - Derivative of a scalar function with respect to a vector
   - MVU: Lagrange optimization
4. **Function Estimation**
   - Autocorrelation
   - Power Spectral Density

Assessing the estimator

How can we **estimate the mean value of a random process** given a set of samples (N) of a single realization?

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

Let us assume that we select the average of the available samples (**sample mean**) as estimate of the mean value of the process. For this selection to be correct, we have to assume that:

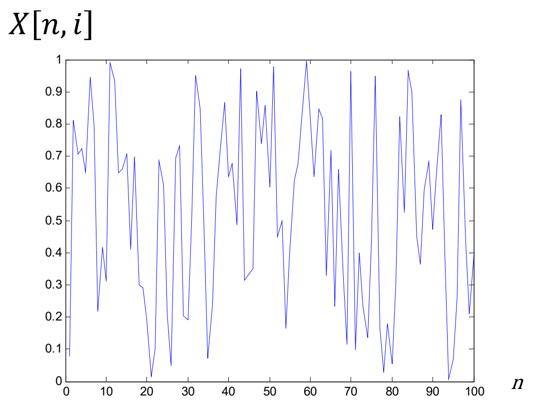
$$\hat{m}_X^{(1)} = \frac{1}{N}\sum_{n=1}^{N} x[n]$$

- **Stationarity**: The parameter to be estimated does not change through time

$$m_x[n] = m_x$$

- **Ergodicity**: Any realization of the process ($X[n, i]$) assumes the statistical properties of the whole process

$$m_X = E\{X[n]\} = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} X[n,i]$$



Any estimator is a random variable

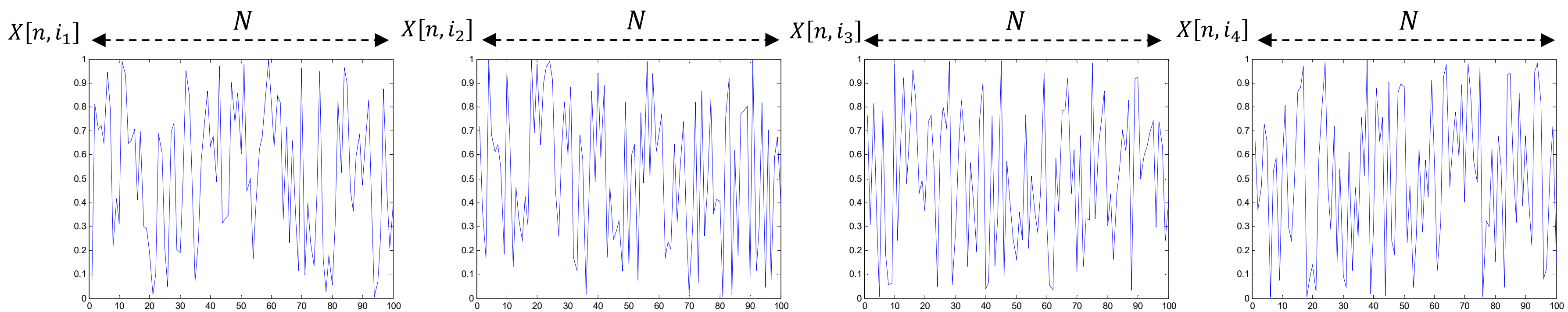
Estimators operate on the **samples of a given realization**. The estimated value depends on:

- The **available realization** ($X[n,i]$)
- The **selected window** (n, N)

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

$$\widehat{m}_x^{(1)} = \frac{1}{N}\sum_{n=1}^{N} X[n,i]$$

Thus, **any estimator is a random variable!**



$\widehat{m}_x^{(1)}(0,i_1) = 0.5205$ $\widehat{m}_x^{(1)}(0,i_2) = 0.5217$ $\widehat{m}_x^{(1)}(0,i_3) = 0.5045$ $\widehat{m}_x^{(1)}(0,i_4) = 0.5100$

$\widehat{m}_x^{(3)}(0,i_1) = 0.5010$ $\widehat{m}_x^{(3)}(0,i_2) = 0.5007$ $\widehat{m}_x^{(3)}(0,i_3) = 0.5008$ $\widehat{m}_x^{(3)}(0,i_4) = 0.4990$

Four realizations of a process with samples uniformly distributed between 0 and 1

PDF of an estimator

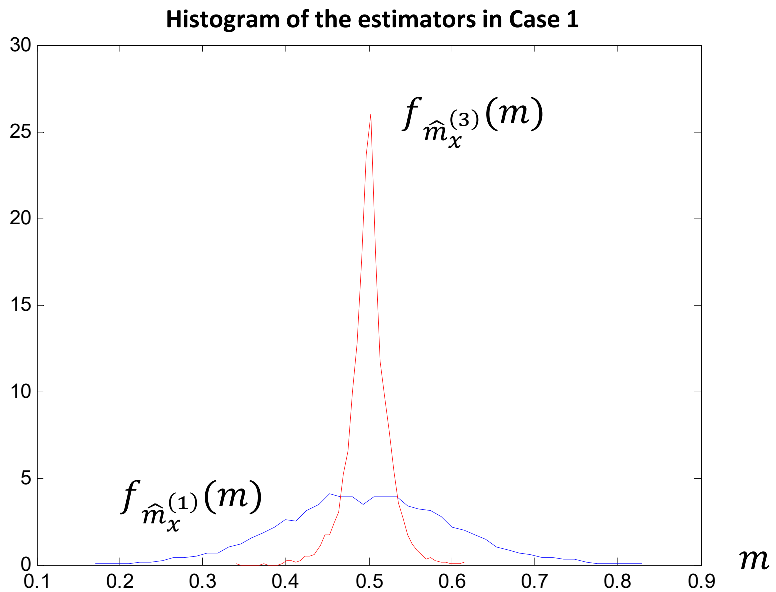
2.1

Examples: Estimators as **random variables** (pdf):

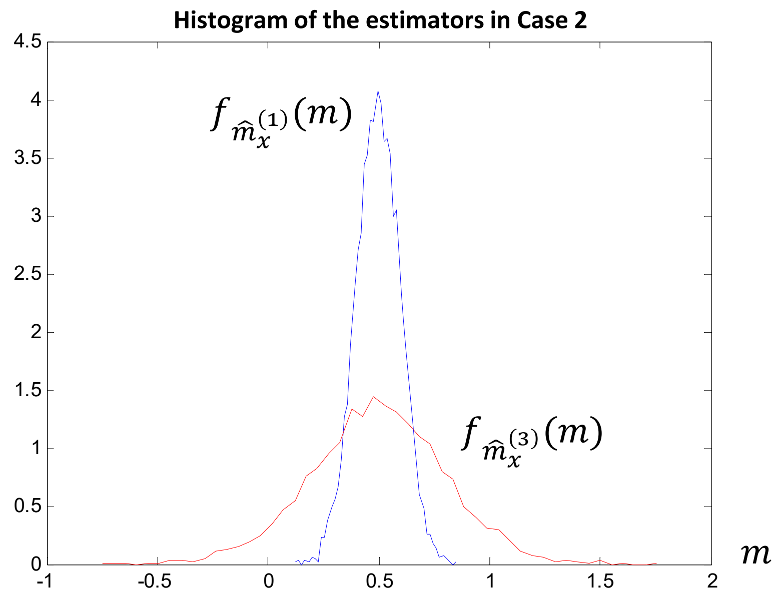
- **Case 1:** 5000 realizations of 100 samples from a white **uniform** process $(m_X = 0.5, \sigma_X^2 = 1)$
- **Case 2:** 5000 realizations of 100 samples from a white **Gaussian** process $(m_X = 0.5, \sigma_X^2 = 1)$

$$\hat{m}_x^{(1)} = \frac{1}{N}\sum_{n=1}^{N} x[n] = \frac{1}{N}\underline{\mathbf{1}}^T\underline{\mathbf{x}}$$

$$\hat{m}_x^{(3)} = \frac{\max(\underline{\mathbf{x}}) + \min(\underline{\mathbf{x}})}{2}$$



Estimators behave differently with the same data



Estimators behave differently when the pdf varies

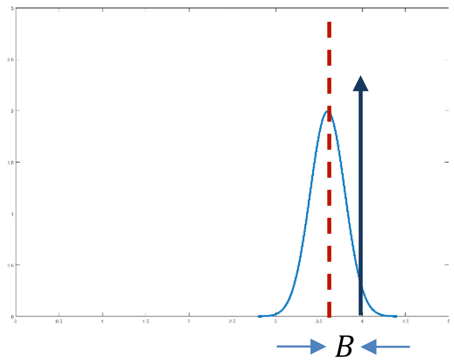
Features of an estimator (I)

2.1

The **bias of an estimator** is the difference between the expected value of the estimator and the true value of the parameter being estimated:

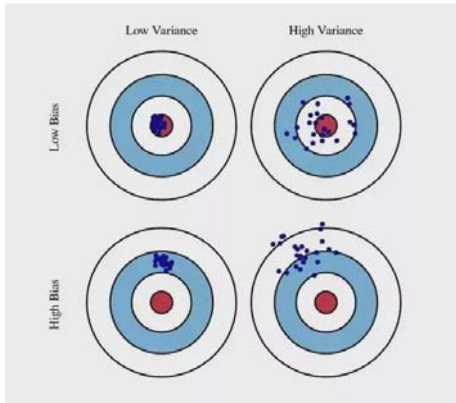
- Estimations delivered by a biased estimator are **consistently different** from the parameter to be estimated
- An estimator without biased is named **unbiased**

$$B(\hat{\theta}) = |\theta - E\{\hat{\theta}\}|$$



- ❑ Given the signal model ($X[n] = \theta + W[n]$), calculate the bias of the estimator.
  Note: $W[n]$ is a stationary, white noise.

$$\hat{\theta}_N = \frac{1}{N}\sum_{n=1}^{N} x[n]$$



Graphical comparison of bias and variance

Bias of the estimator

2.1

$$B(\hat{\theta}_N) = \left| \theta - E\{\hat{\theta}_N\} \right| = \qquad \left[\hat{\theta}_N = \frac{1}{N}\sum_{n=1}^{N} x[n] \right]$$

$$= \left| \theta - E\left\{ \frac{1}{N}\sum_{n=1}^{N} x[n] \right\} \right| = \qquad \left[x[n] = \theta + w[n] \right]$$

$$= \left| \theta - E\left\{ \frac{1}{N}\sum_{n=1}^{N} (\theta + w[n]) \right\} \right| = \qquad \left[\text{OPERAD. LSN.} \right]$$

$$= \left| \theta - \frac{1}{N}\sum_{n=1}^{N} E\{\theta + w[n]\} \right| = \qquad \left[\theta \ \text{DETERM.} \right]$$

$$= \left| \theta - \frac{1}{N}\sum_{n=1}^{N} (\theta + E\{w[n]\}) \right| = \qquad \left[E\{w[n]\} = \emptyset \right]$$

$$= \left| \theta - \frac{1}{N}\sum_{n=1}^{N} (\theta + \emptyset) \right| = \left| \theta - \frac{1}{N} N\theta \right| = \emptyset$$

Features of an estimator (II)

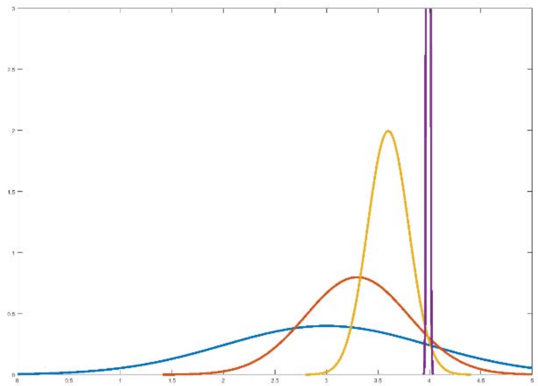
2.1

The unbiased constrain is desirable and, among all unbiased estimators, that of minimum **variance** is preferred (**Minimum Variance Unbiased: MVU**)

$$\sigma_{\hat{\theta}}^2 = E\left\{\left(\hat{\theta} - E\{\hat{\theta}\}\right)^2\right\}$$

An estimator is **consistent** if, as the number of samples (N) increases, the resulting sequence of estimates converges to θ

$$\lim_{N\to\infty} E\{\hat{\theta}\} \longrightarrow \theta \qquad \lim_{N\to\infty} \sigma_{\hat{\theta}}^2 \longrightarrow 0$$



A consistent estimator of $\theta = 4$

- Given the signal model ($X[n] = \theta + W[n]$), calculate the variance of the estimator $\hat{\theta}_N$. Is it consistent?
  Note: $W[n]$ is a stationary, white noise.

$$\hat{\theta}_N = \frac{1}{N}\sum_{n=1}^{N} x[n]$$

Variance of the estimator (I)

2.1

$$\sigma^2_{\hat{\theta}_N} = E\left\{\left(\hat{\theta}_N - E\{\hat{\theta}_N\}\right)^2\right\} = \qquad \left[E\{\hat{\theta}_N\} = \theta\right]$$

$$= E\left\{\left(\hat{\theta}_N - \theta\right)^2\right\} = \qquad \left[\hat{\theta}_N = \frac{1}{N}\sum_{n=1}^{N} x[n]\right]$$

$$= E\left\{\left(\frac{1}{N}\sum_{n=1}^{N} x[n] - \theta\right)^2\right\} = \qquad \left[x[n] = \theta + w[n]\right]$$

$$= E\left\{\left(\frac{1}{N}\sum_{n=1}^{N}\left(\theta + w[n]\right) - \theta\right)^2\right\} = \qquad \left[\text{SIMPLIF. EL } (\cdot)^2\right]$$

$$= E\left\{\left(\frac{1}{N}\cdot N\theta + \frac{1}{N}\sum_{n=1}^{N} w[n] - \theta\right)^2\right\} =$$

$$= E\left\{\frac{1}{N^2}\left(\sum_{n=1}^{N} w[n]\right)^2\right\} = \qquad \left[\Sigma \text{ INDEX DIF}\right]$$

$$= E\left\{\frac{1}{N^2}\sum_{n=1}^{N} w[n] \sum_{m=1}^{N} w[m]\right\} =$$

Variance of the estimator (II)

2.1

$$\sigma^2_{\hat{\theta}_N} = E\left\{\left(\hat{\theta}_N - E\{\hat{\theta}_N\}\right)^2\right\} = \qquad \left[E\{\hat{\theta}_N\} = \theta\right]$$

$$= E\left\{\frac{1}{N^2}\sum_{n=1}^{N} w[n] \sum_{m=1}^{N} w[m]\right\} =$$

$$= \frac{1}{N^2} E\left\{\sum_{n=1}^{N}\sum_{m=1}^{N} w[n]\,w[m]\right\} = \qquad [\text{OPERAD. LINEAL}]$$

$$= \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N} E\{w[n]\,w[m]\} = \qquad \left[r_w[\ell] = \sigma_w^2\,\delta[\ell]\right]$$

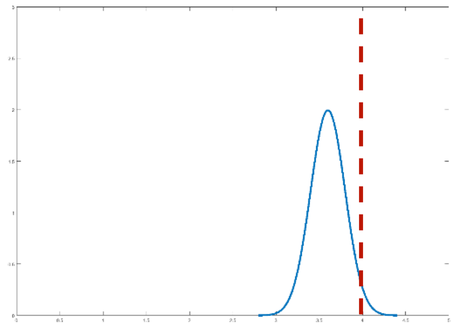
$$= \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N} \delta[n-m]\,\sigma_w^2 = \qquad \left[\sum_{n=1}^{N}\sum_{m=1}^{N}\delta[n-m] = N\right]$$

$$= \frac{1}{N^2}\cdot N\,\sigma_w^2 \quad \Rightarrow \quad \sigma^2_{\hat{\theta}_N} = \frac{\sigma_w^2}{N} \quad \Leftarrow \text{CONSISTENT}$$

Features of an estimator (III)

2.1

If the estimator is biased, the dispersion of the estimations with respect to the actual value to be estimated (θ) is not the variance but the **Mean Square Error** of the estimator ($MSE(\hat{\theta})$)



- The MSE can be a measure of assessment for a given estimator, but to define an estimator **optimizing the MSE usually leads to unrealizable estimators**

$$\sigma_{\hat{\theta}}^2 = E\left\{\left(\hat{\theta} - E\{\hat{\theta}\}\right)^2\right\} = [\text{Biased}] \neq E\left\{\left(\hat{\theta} - \theta\right)^2\right\} = MSE\left(\hat{\theta}\right)$$

❑ Demonstrate that, for a given estimator ($\hat{\theta}$),

$$MSE\left(\hat{\theta}\right) = \sigma_{\hat{\theta}}^2 + B^2\left(\hat{\theta}\right)$$

MSE of the estimator

2.1

$$MSE(\hat{\theta}) = E\{(\hat{\theta}-\theta)^2\} = E\{(\hat{\theta} - E\{\hat{\theta}\} - \theta + E\{\hat{\theta}\})^2\} =$$

$$= E\{((\hat{\theta}-E\{\hat{\theta}\}) - (\theta - E\{\hat{\theta}\}))^2\} =$$

$$= E\{(\hat{\theta} - E\{\hat{\theta}\})^2 - 2(\hat{\theta} - E\{\hat{\theta}\})(\theta - E\{\hat{\theta}\}) + (\theta - E\{\hat{\theta}\})^2\} =$$

$$= E\{(\hat{\theta} - E\{\hat{\theta}\})^2\} - 2E\{(\hat{\theta} - E\{\hat{\theta}\})(\theta - E\{\hat{\theta}\})\} + E\{(\theta - E\{\hat{\theta}\})^2\} =$$

$$= \sigma_{\hat{\theta}}^2 - 2E\{(\hat{\theta} - E\{\hat{\theta}\})(\theta - E\{\hat{\theta}\})\} + B^2(\hat{\theta}) =$$

$$= \sigma_{\hat{\theta}}^2 + B^2(\hat{\theta}) - 2E\{\hat{\theta} - E\{\hat{\theta}\}\}(\theta - E\{\hat{\theta}\}) =$$

$$= \sigma_{\hat{\theta}}^2 + B^2(\hat{\theta}) - 2(E\{\hat{\theta}\} - E\{E\{\hat{\theta}\}\})(\theta - E\{\hat{\theta}\}) =$$

$$= [E\{E\{\hat{\theta}\}\} = E\{\hat{\theta}\}] = \sigma_{\hat{\theta}}^2 + B^2(\hat{\theta}) \leftarrow MSE(\hat{\theta}) = \sigma_{\hat{\theta}}^2 + B^2(\hat{\theta})$$

Exercise 2.1

2.1

2.1. Suppose that we have N samples of a random process $X[n] = A + W[n]$, where A is an unknown constant and $W[n]$ is a stationary, white noise (σ_W^2). Consider the modified estimator $\check{A}$ for some constant a. Find a which results in the minimum MSE and discuss the result.

$$\boxed{\check{A} = a\frac{1}{N}\sum_{n=1}^{N} x[n]}$$

$$\check{A} = a\frac{1}{N}\sum_{n=1}^{N} x[n] = a\hat{\theta}_N \quad \Rightarrow \quad \begin{cases} m_{\check{A}} = a m_{\hat{\theta}_N} = aA \\ \sigma_{\check{A}}^2 = a^2\sigma_{\hat{\theta}}^2 = a^2\dfrac{\sigma_W^2}{N} \end{cases}$$

$$MSE(\hat{\theta}) = \sigma_{\hat{\theta}}^2 + B^2(\hat{\theta}) \quad \Rightarrow \quad MSE(\check{A}) = a^2\frac{\sigma_W^2}{N} + (a-1)^2A^2$$

$$\frac{dMSE(\check{A})}{da} = 0 \quad \Rightarrow \quad \boxed{a_{\text{opt}} = \frac{A^2}{A^2 + \sigma_W^2/N}}$$

“The MSE can be a measure of assessment for a given estimator, but to define an estimator optimizing the MSE usually leads to unrealizable estimators”

Exercise 2.2

2.2. Suppose that we have N uncorrelated samples of a stationary random process denoted by $X[n]$. From these samples, we estimate the mean m_X and the variance σ_X^2 of the process using the following expressions. Determine:

a) The reasoning behind the definition of the variance estimator

b) The bias of the variance estimator σ_X^2.

c) An unbiased estimator for the variance of the process σ_X^2.

$$\widehat{m}_X = \frac{1}{N}\sum_{n=1}^{N} x[n]$$

$$\hat{\sigma}_X^2 = \frac{1}{N}\sum_{n=1}^{N} (x[n] - \widehat{m}_X)^2$$

Introduction to Estimation Theory

2.1

1. **Introduction**
    - The relevance of model selection
2. **Assessing Estimator Performance**
    - Bias, Variance and MSE
    - Vector notation
3. **Minimum Variance Unbiased Estimator**
    - Derivative of a scalar function with respect to a vector
    - MVU: Lagrange optimization
4. **Function Estimation**
    - Autocorrelation
    - Power Spectral Density

The sample mean estimator as a filter

2.1

The previous **sample mean** estimator can be interpreted as a **filter** and, this way, we can generalize the study of its properties:

$$\hat{\theta}_N = \frac{1}{N}\sum_{n=1}^{N} x[n] = \frac{1}{N}\underline{1}^T\underline{x} \quad \Rightarrow \quad \boxed{\hat{\theta}_N = \underline{h}^T\underline{x}}$$

◀ Estimator **linear** in the (N) data

- Given the signal model ($X[n] = \theta + W[n]$), analyze the **bias** of the estimator $\hat{\theta}_N$.

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

$$B(\hat{\theta}_N) = 0 \quad \Rightarrow \quad \boxed{\underline{h}^T\underline{1} = 1}$$

One possible solution is the previous result: $\underline{h}^T = \frac{1}{N}\underline{1}^T$

Bias of the estimator

2.1

$$B(\hat{\theta}_N) = |\theta - E\{\hat{\theta}_N\}| \Rightarrow E\{\hat{\theta}_N\}$$

$$E\{\hat{\theta}_N\} = E\{\underline{h}^T \underline{x}\} = \underline{h}^T E\{\underline{x}\} = [\underline{x} = \underline{\theta} + \underline{w}] =$$

$$= \underline{h}^T E\{\underline{\theta} + \underline{w}\} = [\theta \text{ DETERM.}] = \underline{h}^T [\underline{\theta} + E\{\underline{w}\}] =$$

$$= [E\{\underline{w}\} = \underline{0}] = \underline{h}^T \cdot \underline{\theta} = \theta\, \underline{h}^T \cdot \underline{1}$$

$$B(\hat{\theta}_N) = |\theta - \theta\, \underline{h}^T \cdot \underline{1}| = |\theta [1 - \underline{h}^T \underline{1}]|$$

$$B(\hat{\theta}_N) = \emptyset \Rightarrow \underline{h}^T \cdot \underline{1} = 1 \Rightarrow \text{FOR INSTANCE } \underline{h}^T = \frac{1}{N}\underline{1}$$

The sample mean estimator as a filter

2.1

The previous **sample mean** estimator can be interpreted as a **filter** and, this way, we can generalize the study of its properties:

$$\hat{\theta}_N = \frac{1}{N}\sum_{n=1}^{N} x[n] = \frac{1}{N}\underline{1}^T\underline{x} \quad \Rightarrow \quad \boxed{\hat{\theta}_N = \underline{h}^T\underline{x}}$$

◀ Estimator **linear** in the (N) data

- Given the signal model ($X[n] = \theta + W[n]$), analyze the **variance** of the estimator $\hat{\theta}_N$. Note: $W[n]$ is a stationary, white noise

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

◀ It is the **variance** (power) **of the filtered noise**

- Evaluate this result for a **stationary**, white noise and $\underline{h} = \frac{1}{N}\underline{1}$

Variance of the estimator

2.1

$$\hat{\sigma}^2_{\hat{\theta}_N} = E\{(\hat{\theta}_N - E\{\hat{\theta}_N\})^2\} =$$

$$= [\text{UNBIASED CASE}] = [\underline{h}^T \cdot \underline{1} = 1] = [E\{\hat{\theta}_N\} = \theta] =$$

$$E\{(\hat{\theta}_N - \theta)^2\} = E\{(\underline{h}^T \cdot \underline{x} - \theta)^2\} = [\underline{x} = \underline{\theta} + \underline{w}] =$$

$$= E\{(\underline{h}^T(\underline{\theta} + \underline{w}) - \theta)^2\} = E\{(\underline{h}^T \cdot \underline{\theta} + \underline{h}^T \underline{w} - \underline{\theta})^2\} =$$

$$= [\underline{h}^T \cdot \underline{\theta} = \theta \cdot \underline{h}^T \cdot \underline{1} = \theta] = E\{(\theta + \underline{h}^T \cdot \underline{w} - \theta)^2\} =$$

$$= E\{(\underline{h}^T \cdot \underline{w})^2\} = E\{\underline{h}^T \cdot \underline{w} \cdot \underline{h}^T \cdot \underline{w}\} = E\{\underline{h}^T \cdot \underline{w}\, \underline{w}^T \cdot \underline{h}\} =$$

$$= [\underline{h} \text{ DETERM.}] = \underline{h}^T E\{\underline{w} \cdot \underline{w}^T\} \cdot \underline{h} = \underline{h}^T \underline{\underline{R}}_w \cdot \underline{h}$$

Introduction to Estimation Theory

2.1

1. **Introduction**
   - The relevance of model selection
2. **Assessing Estimator Performance**
   - Bias, Variance and MSE
   - Vector notation
3. **Minimum Variance Unbiased Estimator**
   - Derivative of a scalar function with respect to a vector
   - MVU: Lagrange optimization
4. **Function Estimation**
   - Autocorrelation
   - Power Spectral Density