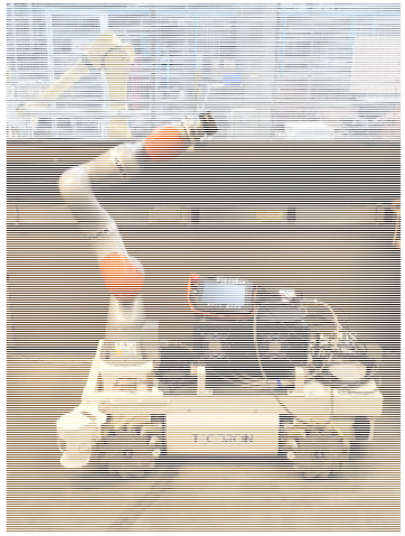
ESTIMATION THEORY AND PARAMETRIC IDENTIFICATION

LINEAR ESTIMATION THEORY



Mechbal N. / Guskov M. / Rebillat M.

*Laboratoire **PIMM**, UMR-CNRS*
Arts et Métiers ParisTech (ENSAM), Paris

OUTLINE TOPICS

There are 3 main Parts in this course:

Part 1 Estimation Theory – N. Mechbal – 6 lectures

- Review of probability theory
- Linear Estimation Theory – Deterministic and Bayesian methods
- Nonlinear estimation – EKF, UKF and PF

Part 2 Parametric identification theory – N. Mechbal & M. Rebillat

- Mathematical foundations of system identification – **NM/1 Lecture**
- Parameter estimation - **NM /2 Lectures**
- Sensors and Signal processing for identification – discrete time processing, frequency analysis, denoising - **MR/3 Lectures**

Part 3 Model identification of flexible manipulators – M. Guskov

- Modeling of flexible manipulators - **MG/1Lectures**
- Modeling for vibration analysis - **MG/1Lectures**
- Application: Robotic machining - **MG/1Lectures**

PART I: OUTLINE TOPICS

- **There are 3 Topics in Part 1:**
  - Chapter 1 Preliminaries: Probability Theory
  - Chapter 2 Linear estimation theory
  - Chapter 3 Non Linear Estimation

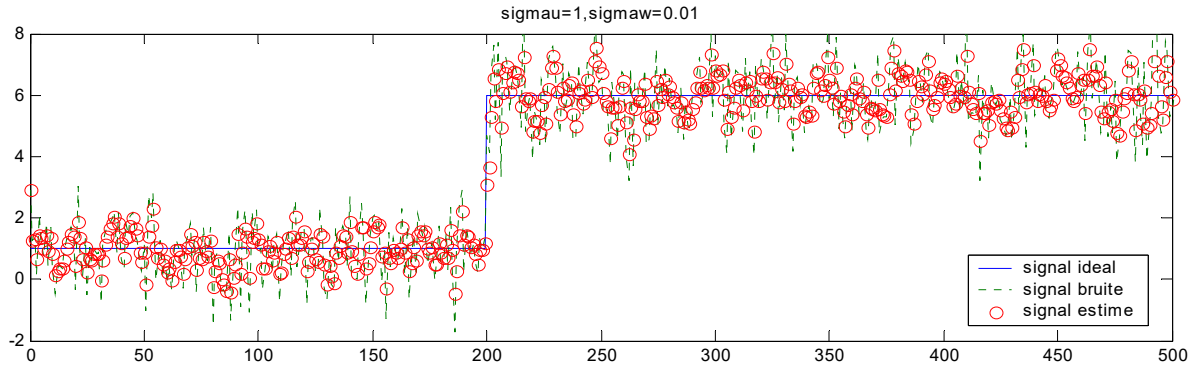
SYLLABUS

- **Lecture 1:**
  - Course organization
  - Motivations
  - Introductive Examples
  - Review of Probability
- **Lecture 2:**
  - **Estimation theory**
  - **Exercises**
- **Lecture 3:**
  - Linear estimators: Least Square, Maximum Likelihood
  - Exercise
- **Lecture 4:**
  - Linear estimator: Kalman Filter
  - Exercise & Matlab class
- **Lecture 5:**
  - Nonlinear estimation: EKF and UKF
  - Exercises & Matlab class
- **Lecture 6:**
  - Nonlinear estimation: PF and RBPF
  - Exercise & Matlab class

... no one believes an hypothesis except its originator but everyone believes an experiment except the experimenter.

W. I. B. Beveridge, 1950.

LINEAR ESTIMATION THEORY



OUTLINE

I. Introduction

II. Estimators properties

III. Estimation techniques

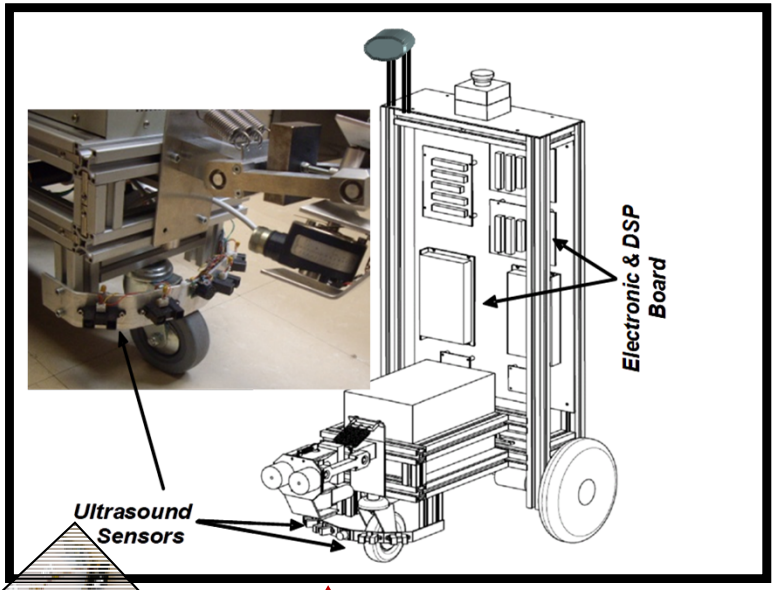
IV. Kalman filter

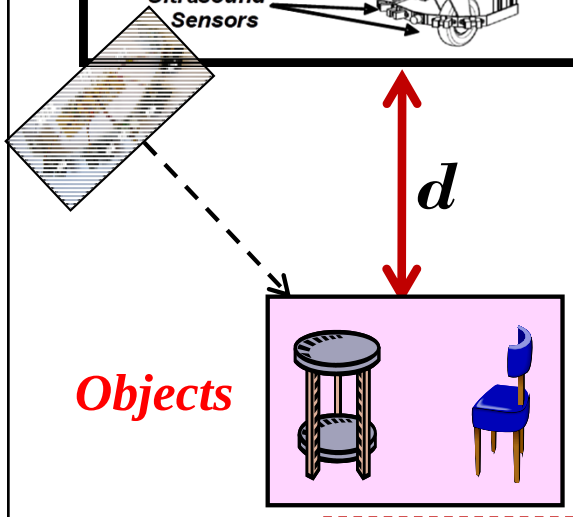
Brief History :

- The first attempts to systematically approach the estimation problem, as it is known today, were taken by **Gauss and Legendre** in studying astronomical problems during the late **18th and the early 19th century**. More specifically, they tried to estimate the positions of planets and comets using telescopic measurements. Gauss made use of the method of least-squares for the first time in **1795** at the age of 18. However, it was not until 1809 that he published his results in his book "*Theoria Motus Corporum Celestium*" (Gauss, 1809). A few years earlier, in 1805 Legendre had independently invented and published the method in his book "*Nouvelles méthodes pour la détermination des orbites des comètes*" This gave rise to a big dispute between Gauss and Legendre, concerning who was the inventor of the least-squares method (Sorenson, 1970). A thorough discussion of the early contributions to estimation theory is provided by Seal (1967) and Sorenson (1970).
- The next major development in the study of the estimation problem came in the **1940s,** with the filtering work of **Wiener** (1949) and Kolmogorov.
- … and **The breakthrough came with the Kalman filter (Kalman, 1960)** …

(T. Schon, 2006)

Example 1:





Objective: *To determine the distance,* ***d,*** *between the robot and the object.*

Ultrasound sensors*: generate high frequency sound waves and evaluate the echo which is received back by the sensor.*

In theory, sensor measure the time interval, t, between transmitting the signal and receiving it echo, in order to determine the distance, d, to an object, i.e.:

$$t = \frac{2\,d}{C}$$

C: wave velocity

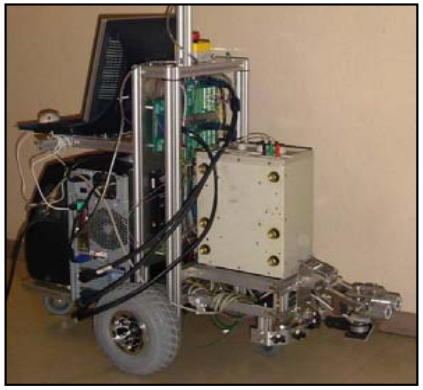
In practice, the wave undergoes during its propagation a number of distortions that make impossible the exact determination of t.

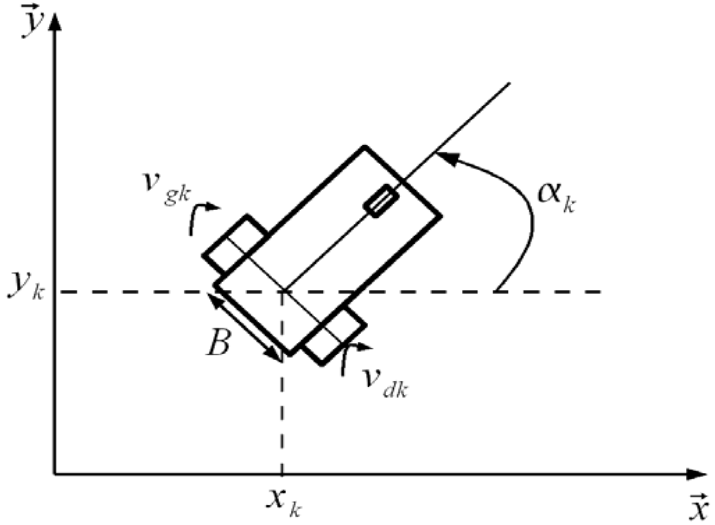
$$t = \frac{2\,d}{C} + b$$

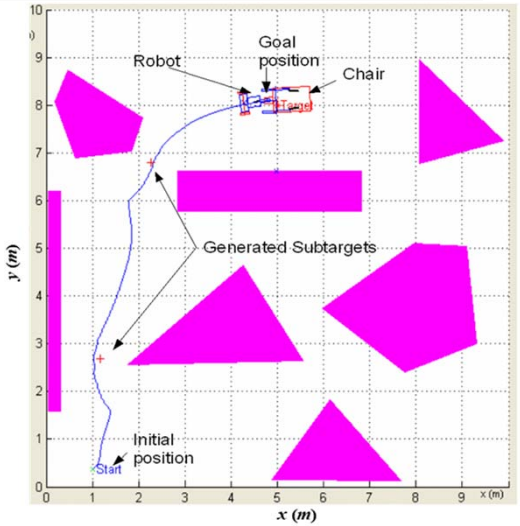
b: Measurement noise

Using measurements of t, we can only find an approximation of d, i.e., an ***estimation*** $\hat{d}$ *d.*

Example 2:







Objective: *at every moment, k, determine the position and orientation of the robot relative to its initial state, i.e., determine:* $\underline{X}_k = [x_k \quad y_k \quad \alpha_k]^T$

Sensors:

$$Odometry \Rightarrow \underline{u}_k = [\underline{v}_{dk} \quad \underline{v}_{gk}]^T + \underline{\omega}_{uk}^T$$

$$GPS \Rightarrow \underline{Y}_k = [x_k \quad y_k]^T + \underline{\omega}_{yk}^T$$

Measurement noise

Model: *State space representation*

$$\begin{cases} \underline{X}_{k+1} = f(\underline{X}_k; \underline{u}_k + \underline{\omega}_{uk}) + \underline{\varsigma}_k \\ \underline{Y}_k = C\underline{X}_k + \underline{\omega}_{yk} \end{cases}$$

Model noise

Measurement noise

Because of the noise, we could not determine the exact value of the state, X, one can only find an estimated $\hat{X}_k$

Example 3:

- Consider the classic example of measuring the exact value of a resistance R. For that, we perform N independent experiments where one measures the current $i(k)$ and voltage $u(k)$ at its terminals. These measures are all subject to errors and are grouped into the following vector:

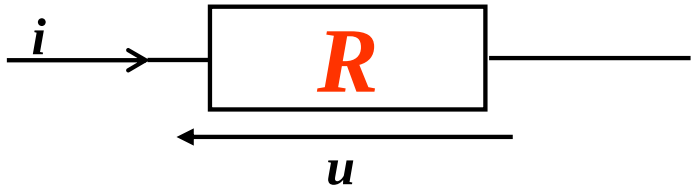
$$\underline{z} = [i(1) \quad u(1) \quad i(2) \quad u(2) \quad \cdots \quad i(N) \quad u(N)]^T \in \Re^{2N}$$

The parameter to be estimated is $\theta = R$.

The goal is then to build an estimator: $\hat{\theta} = \hat{R} = g(\underline{z})$

that can best estimate the value the resistance R.

We can propose several estimators, such



$$\hat{\theta}_1 = \frac{\sum_{k=1}^{N} u(k)i(k)}{\sum_{k=1}^{N} i^2(k)} \qquad \hat{\theta}_2 = \frac{\sum_{k=1}^{N} u(k)}{\sum_{k=1}^{N} i(k)}$$

We will see later that these two estimators, which seems reasonable, have different properties.

Problem statement

- Consider a physical phenomenon which involves the characterization of one or more parameters. The observation of the phenomenon gives us some knowledge of the parameters although this observation is generally subject to errors.

- The presence of these errors has the consequence, that it is impossible to determine the exact value of the various parameters that characterize the phenomenon.

- The estimation problem consists in extracting the maximum information from observations to derive the **optimal value** of parameters within the sense of a certain criterion.

- ***We call Estimator the results of this operation.***

Problem statement

- We seek to estimate: $\underline{\theta} = [\theta_1\ \theta_2\ \dots\ \theta_p]^T \in \mathbb{R}^p$
- We use:
  - measurment: $\underline{z}^N = [z_1 \quad z_2 \quad \cdots \quad z_N]^T \in \mathbb{R}^N$
  - model linking $\underline{\theta}$ and $\underline{z}^N$: $\underline{z}^N = h(\underline{\theta}, \underline{b})$

Noise

Example 1 :

$$\underline{\theta} = d$$

$$z^N = [t_1\ t_2\ \dots\ t_N]^T$$

$$b^N = [b_1\ b_2\ \dots\ b_N]^T \qquad t_i = \frac{2d}{C} + b_i$$

$$\underline{z}^N = \begin{bmatrix} 2/C \\ \vdots \\ 2/C \end{bmatrix} d + \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} = h(\underline{\theta}, \underline{b})$$

Example 2 :

$$\underline{\theta} = X_k = [x_k \quad y_k \quad \alpha_k]^T$$

$$z^N = [y_1^T\ u_1^T\ \dots\ y_N^T\ u_N^T]^T$$

- if $N < k$: prediction
- if $N = k$: filtering
- if $N > k$: smoothing

Problem statement

In general, the observed signal consists of the expected signal and noise or disturbance. The measurement model is as follows:

$$\underline{y} = h(\underline{\theta}, \underline{b})$$

- $\underline{y} \in \Re^N$ is the vector composed of N observations or measurements
- $\underline{\theta} \in \Re^p$ is the vector containing the parameters to be estimated.
- b is the noise vector. This is a random signal with statistical characteristics.

Definition:

Consider the measurement model above, ***an estimator*** $\underline{\hat{\theta}}$ *of* $\underline{\theta}$ *is a certain function* $g(\underline{y})$ *that for each realization* $\underline{z}$ *of* $\underline{y}$ *matches a particular value* $\underline{\hat{\theta}}$

$$\underline{\hat{\theta}} = g(\underline{z})$$

The variable $\underline{\tilde{\theta}} = \underline{\hat{\theta}} - \underline{\theta}$ *is called the* ***estimation error***.

- **There are two cases to consider when designing an estimator:**
  - Classical approach: we suppose here that the unknown parameters to be estimated are deterministic. Therefore, based on the measurement model $\underline{y}$ has a PDF that is parameterized by $\underline{\theta}$, *i.e.* : $p\left(\underline{y}\left(\underline{\theta}\right)\right)$
    - **Example 1:** estimate the deterministic distance with ultrasound sensor
  - Bayesian approach: we use *a priori* information that is available on $\underline{\theta}$. In this case, $\underline{\theta}$ is the realization of a random variable with probability density function.
    - **Example** 2 - *localization of a robot*: you have a map of the environment and the odometry sensor through which we have a priori knowledge of d. For example, d is Gaussian with mean d_0 and variance σ^2

Models:

To solve any estimation problem, we **need a model**. Here, in this part of the course, we will look deeper into *three specific* models:

- *Linear Model*: In this model, the relationship between $\underline{y}$ and θ is given by

$$\underline{y} = a\theta + \underline{\omega}$$

where a is the (known) model vector and $\underline{\omega}$ is the noise vector, which is assumed to have mean and covariance.

- *The linear Gaussian model*: This model is a special case of the linear model, where the noise vector $\underline{\omega}$ is assumed to be Gaussian (or normal) distributed.

$$\underline{\omega} \sim N(0,\Gamma) \;\Rightarrow\; p_{\omega}(\omega) = (2\pi)^{\frac{-N}{2}} \det(\Gamma)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\underline{\omega}^{T}\Gamma^{-1}\underline{\omega}\right]$$

- *Linear discrete state space model*: we use here a state space model with process noise and measurement noises, *i.e.*,

$$\begin{cases} \underline{X}_{k+1} = F\underline{X}_k + G\underline{u}_k + \underline{\varsigma}_k \\ \underline{Y}_k = C\underline{X}_k + \underline{\omega}_k \end{cases}$$

Estimation techniques

- There are different techniques used to generate estimators and they depend on the **nature of $\underline{\theta}$** :

- We can view the unknown parameter $\underline{\theta}$ as a **deterministic** variable vector
  - Minimum Variance Unbiased Estimator (**MVU**)
  - Linear Minimum Variance Unbiased Estimator (**LMVU**)
  - Maximum Likelihood Estimator (**ML**)
  - Least Squares Estimator (**LS**)

- The Bayesian philosophy: $\underline{\theta}$ is viewed as a **random** variable vector
  - Bayesian Mean Square Estimator (**BMS**)
  - Linear Bayesian Mean Square Estimator (**LBMS**)
  - Maximum A Posteriori estimator (**MAP**)
  - Kalman Filter (**KF**)
  - Particular Filter (**PF**)

ESTIMATORS PROPERTIES

Objective

- *If we build an estimator, it is important to know whether it provides a "**good**" estimate. It may also be useful to know if there is a "**best**" estimator and if we have several estimators to know how **to compare** them.* We present in this section some **properties** by which we can judge of the quality of an estimator

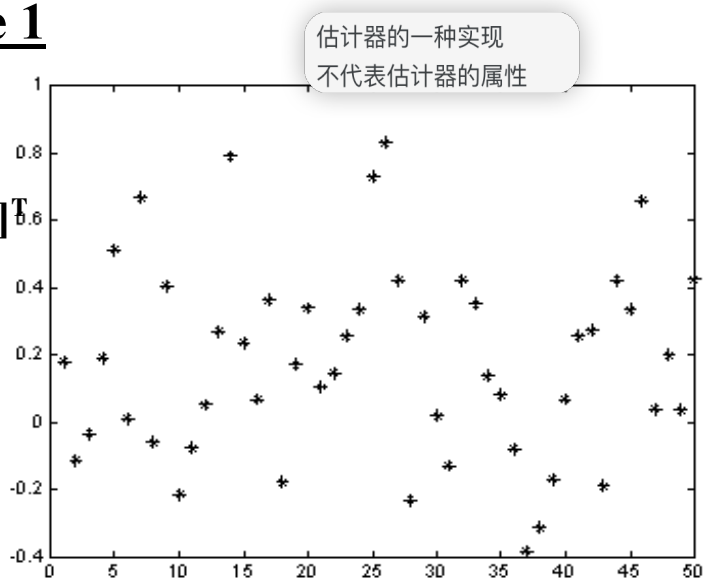
Nature of the properties

- Recall that an estimator is a random variable. And for different realizations of the measurement vector, we will have different achievements. One realization of the estimator is not representative of the properties of an estimator. We must therefore consider all possible outcomes or estimated.

Example 1

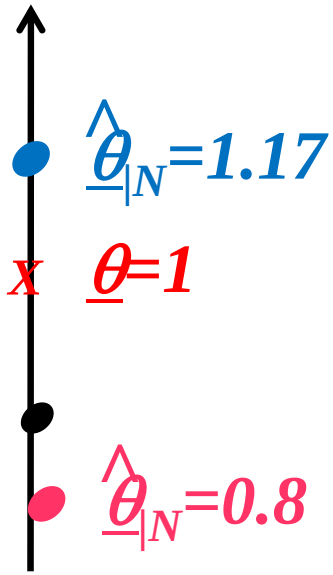
θ=1 N=50

$\underline{z}^N=[t_1\ t_2 \dots t_N]^T$



Estimator **:**

$$\hat{\underline{\theta}}_{|N} = \frac{C}{2N}\sum_{i=1}^{N} t_i$$



Bias of an estimator:

- **Definition**: *We call Bias of an estimator the quantity that characterizes the difference between the **mean** of estimates and the **mean** of possible values of the parameter to be estimated. It is defined by:*

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

- An estimator is called unbiased if

$$B(\underline{\theta}) = 0 \quad \Rightarrow \quad E\{\underline{\hat{\theta}}\} = E\{\underline{\theta}\}$$

- If the parametric is supposed deterministic then we have

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

Example:

- Take the Example, dealing with the estimation of the precise value of a resistance R from the measurement of the current $i(k)$ and voltage $u(k)$.
- Suppose that the same current i_0 is passing through the resistance which produce the voltage, $u_0 = R\, i_0$. Suppose also that these measures are corrupted by noise, *i.e.*,

$$i(k) = i_o + b_i(k) \qquad u(k) = u_o + b_u(k)$$

- We assume that

$$E\{b_u(k)\} = 0 \quad E\{b_i(k)\} = 0 \quad E\{b_i(j)b_u(k)\} = 0 \quad E\{b_i^2(k)\} = \sigma_i^2 \quad E\{b_u^2(k)\} = \sigma_u^2$$

- Let us examine if the first proposed estimator is biased or not:

$$\hat{\theta}_1 = \hat{R}_1 = \frac{\sum_{k=1}^{N} u(k)i(k)}{\sum_{k=1}^{N} i^2(k)} = \frac{\frac{1}{N}\sum_{k=1}^{N}\left(i_o + b_u(k)\right)\left(u_o + b_i(k)\right)}{\frac{1}{N}\sum_{k=1}^{N}\left(i_o + b_u(k)\right)^2} = \frac{R+\alpha}{1+\beta}$$

$$\text{with} \quad \alpha = \frac{1}{N}\sum_{k=1}^{N}\left(R\frac{b_i(k)}{i_o} + \frac{b_u(k)}{i_o} + \frac{b_u(k)b_i(k)}{i_o^2}\right) \quad \text{and} \quad \beta = \frac{1}{N}\sum_{k=1}^{N}\left(2\frac{b_i(k)}{i_o} + \frac{b_i^2(k)}{i_o^2}\right)$$

Example:

- We obtain by approximating

$$\hat{R}_1 = \frac{R+\alpha}{1+\beta} \approx (R+\alpha)(1-\beta+\beta^2+\ldots) = R+\alpha-\beta R-\alpha\beta+\beta^2 R+\alpha\beta^2+\ldots$$

- The expectation of the estimator is then given by

$$E\{\hat{R}_1\} = R + E\{\alpha\} - RE\{\beta\} - E\{\alpha\beta\} + RE\{\beta^2\} + E\{\alpha\beta^2\} + \ldots$$

- By neglecting third order terms and using statistical assumption on noises, we have:

$$E\{\alpha\} = 0, \qquad E\{\beta\} = \frac{\sigma_i^2}{i^2}, \qquad E\{\alpha\beta\} = R\frac{2}{N}\frac{\sigma_i^2}{i^2}, \qquad E\{\beta^2\} = \frac{4}{N}\frac{\sigma_i^2}{i^2}$$

- Therefore, we get

$$E\{\hat{R}_1\} \approx R\left(1 - \frac{\sigma_i^2}{i^2} + \frac{2}{N}\frac{\sigma_i^2}{i^2}\right)$$

- The bias is then

$$B(\hat{R}_1) = E\{\hat{R}_1\} - R \approx R\left(-\frac{\sigma_i^2}{i^2} + \frac{2}{N}\frac{\sigma_i^2}{i^2}\right) \neq 0$$

Biased estimator !!

Evaluation of the Bias: Monte-Carlo simulations

- Computational complexity is such that it is not always possible to proceed to an analytical calculation of the bias. We then use simulations called "*Monte-Carlo simulations*" to evaluate the properties of the estimator (e.g., using Matlab).

- **Principe**: M-C simulations are a class of computational algorithms that rely on repeated random sampling to compute their results (**law of large numbers**). M-C methods are often used in simulating physical and mathematical systems:

"The characteristics of a random sample are closer to the statistical characteristics of the population as the sample size increases "

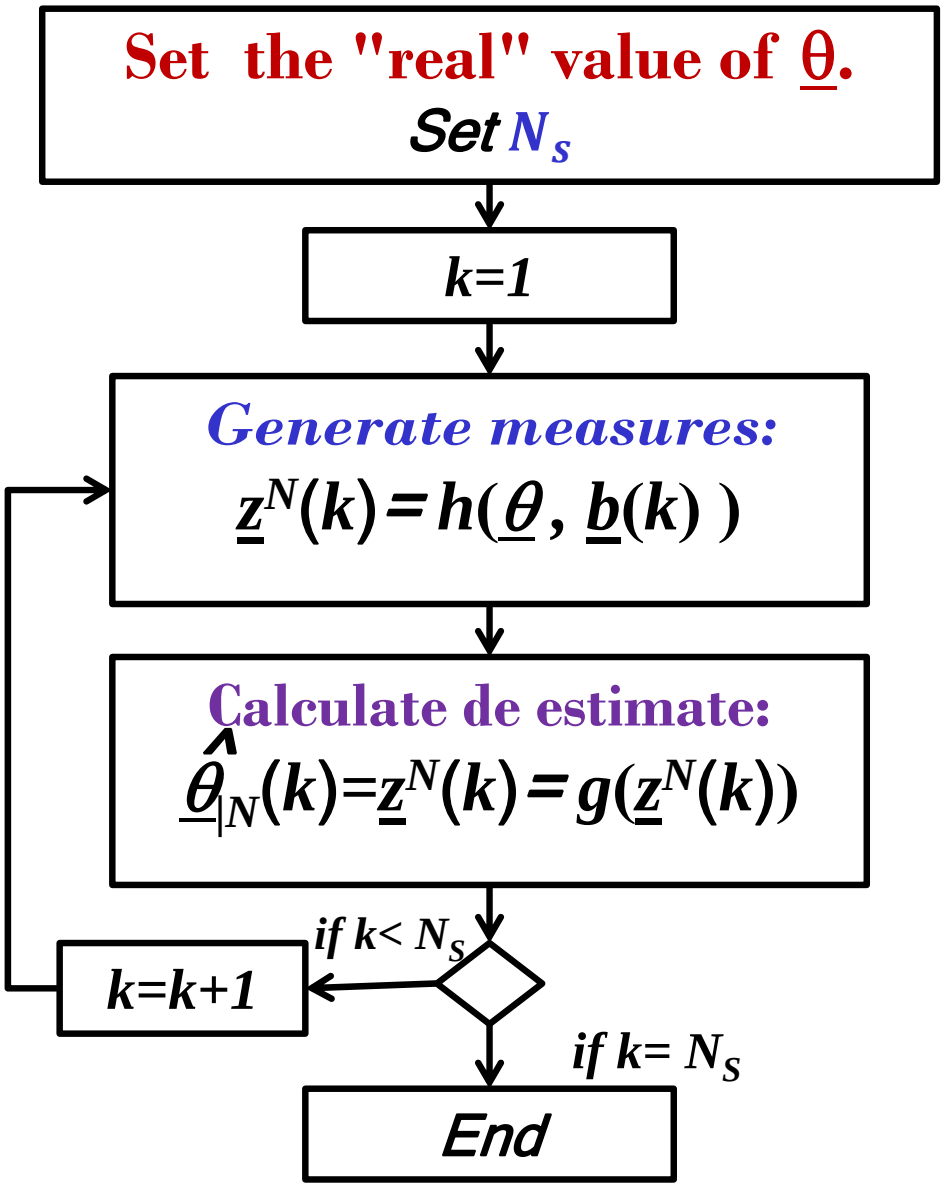
"Les caractéristiques d'un échantillon aléatoire se rapprochent d'autant plus des caractéristiques statistiques de la population que la taille de l'échantillon augmente"

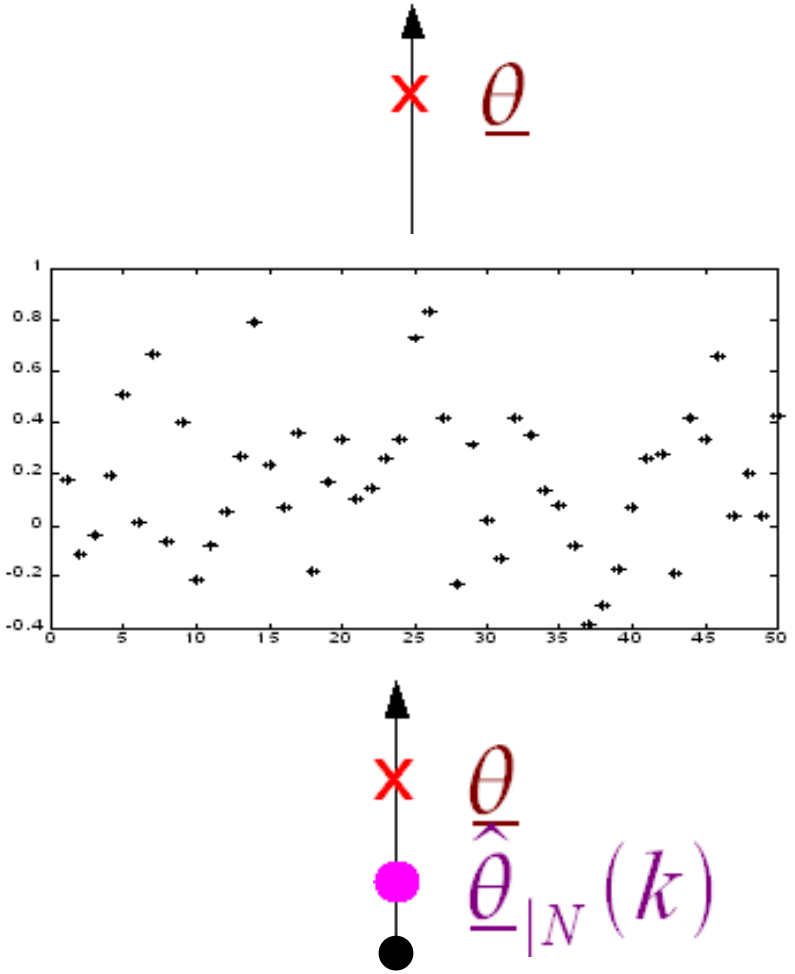
- The Monte-Carlo method was coined in **the 1940s** by *J. von Neumann*, *S. Ulam* and *N. Metropolis*, while they were working on nuclear weapon projects in the Los Alamos National Laboratory. *The name refers to gambling games practiced at Monte-Carlo casino.*

- To estimate the bias, we shall perform N_s **simulations**

Evaluation of the Bias: Monte-Carlo simulations

(case of deterministic $\underline{\theta}$)





Evaluation of the Bias: Monte-Carlo simulations

- Approximation of the expectation: we recall that the expectation of a random vector $\underline{x}$ is given by:

$$E\{\underline{x}\} = \int_{-\infty}^{+\infty} \underline{x}\, f(\underline{x}) d\underline{x}$$

- If now we consider N_s realization of the a vector $\underline{x}$, the **empirical mean** is:

$$\bar{x} = \frac{1}{Ns} \sum_{k=1}^{Ns} \underline{x}(k)$$

- We can show that for large N_s, the expectation can be approximate by this empirical mean (**law of large numbers**) :

$$\lim_{N_s \to \infty} \bar{x} = E\{\underline{x}\} \quad \Rightarrow \quad \text{Large } N_s \Rightarrow \bar{x} \approx E\{\underline{x}\}$$

Evaluation of the Bias: Monte-Carlo simulations

- **Procedure**: The evaluation of the bias by Monte Carlo simulations is based on this property (empirical expectation). Indeed, if one has a sufficiently high number N_s of achievements of the estimator, we can then use the following approximation

$$\lim_{N_s \to \infty} \bar{\hat{\theta}} = E\{\underline{\hat{\theta}}\} \qquad with \qquad \bar{\hat{\theta}} = \frac{1}{N_s} \sum_{k=1}^{N_s} \underline{\hat{\theta}}(k)$$

- We perform N_s realization

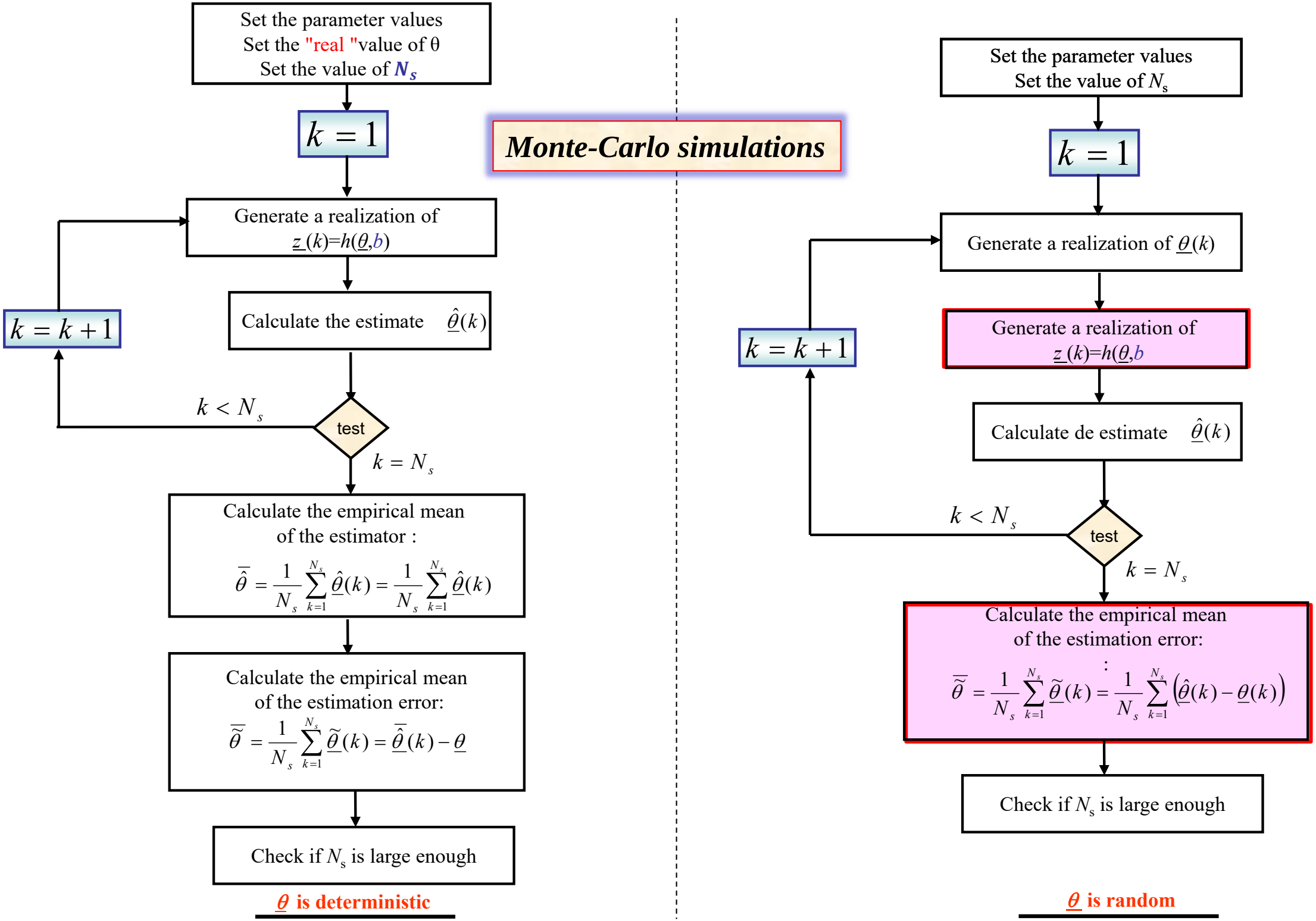
$$\underline{\hat{\theta}} = \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_{N_s}$$

- by simulating N_s realization of measurement vector

$$\underline{z} = z_1, z_2, \ldots, z_{N_s} \quad i.e., \quad \underline{z}(1), \underline{z}(2), \ldots, \underline{z}(N_s)$$

- If $\underline{\theta}$ is a **random vector** parameter then its expectation is also approximated, *i.e.*

$$\lim_{N_s \to \infty} \bar{\underline{\theta}} = E\{\underline{\theta}\} \qquad with \qquad \bar{\underline{\theta}} = \frac{1}{N_s} \sum_{k=1}^{N_s} \underline{\theta}(k)$$


Monte-Carlo simulations
Set the parameter values
Set the "real "value of θ
Set the value of N_s
$k = 1$
Generate a realization of
$\underline{z}(k)=h(\underline{\theta},b)$
Calculate the estimate $\hat{\underline{\theta}}(k)$
test
$k < N_s$
$k = k+1$
$k = N_s$
Calculate the empirical mean
of the estimator :
$\overline{\hat{\theta}} = \frac{1}{N_s}\sum_{k=1}^{N_s}\hat{\underline{\theta}}(k) = \frac{1}{N_s}\sum_{k=1}^{N_s}\hat{\underline{\theta}}(k)$
Calculate the empirical mean
of the estimation error:
$\overline{\tilde{\theta}} = \frac{1}{N_s}\sum_{k=1}^{N_s}\tilde{\underline{\theta}}(k) = \overline{\hat{\underline{\theta}}}(k) - \underline{\theta}$
Check if N_s is large enough
$\underline{\theta}$ is deterministic
Set the parameter values
Set the value of N_s
$k = 1$
Generate a realization of $\underline{\theta}(k)$
Generate a realization of
$\underline{z}(k)=h(\underline{\theta},b)$
Calculate de estimate $\hat{\underline{\theta}}(k)$
test
$k < N_s$
$k = k+1$
$k = N_s$
Calculate the empirical mean
of the estimation error:
$\overline{\tilde{\theta}} = \frac{1}{N_s}\sum_{k=1}^{N_s}\tilde{\underline{\theta}}(k) = \frac{1}{N_s}\sum_{k=1}^{N_s}\left(\hat{\underline{\theta}}(k) - \underline{\theta}(k)\right)$
Check if N_s is large enough
$\underline{\theta}$ is random

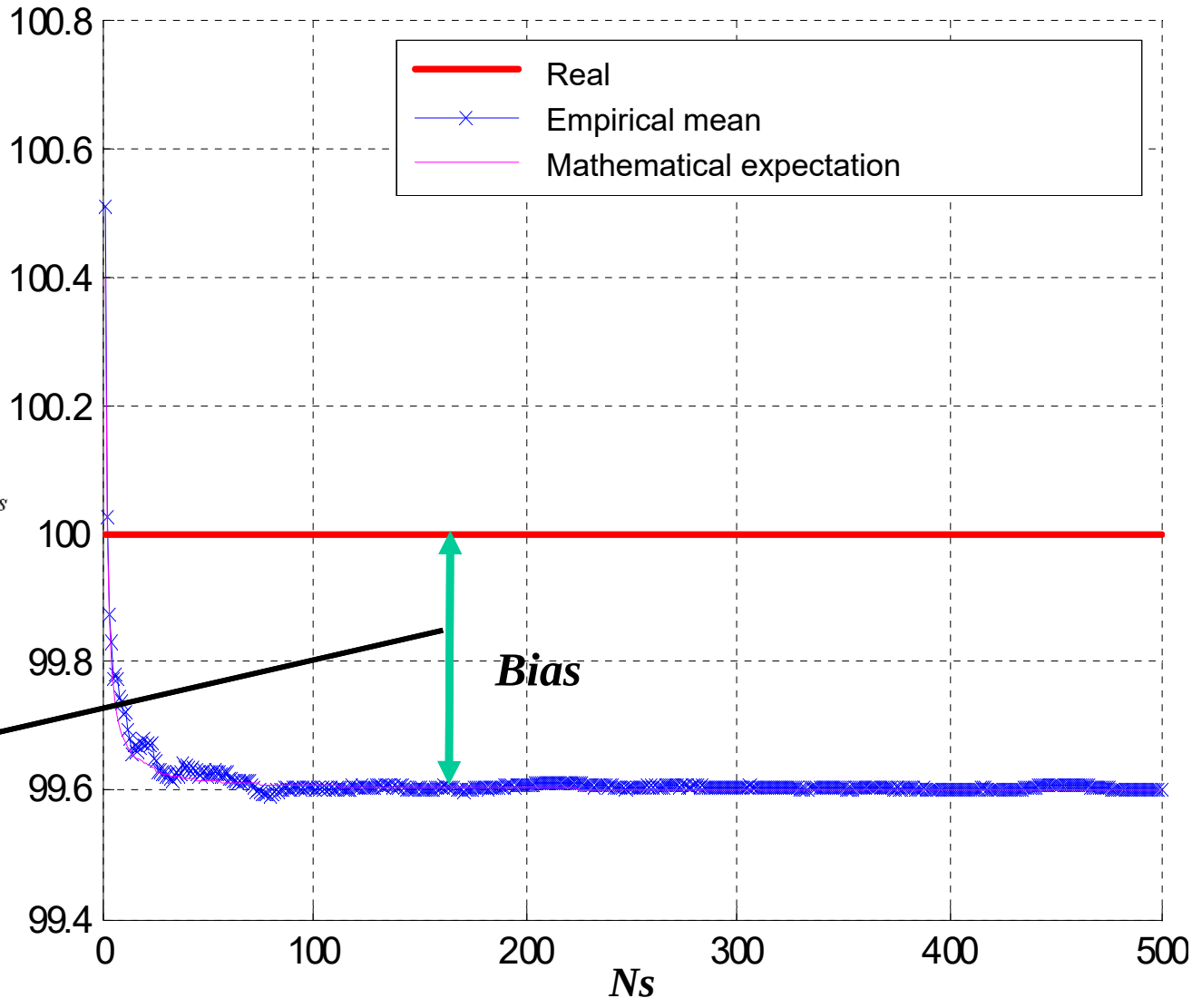
- ***Example :*** *Estimation of the Resistance value*
  - Let us verify by MC simulations *if the* first estimator of *Example 3* *is biased*
  - Simulation parameters

$$R = 100\,\Omega \quad i = 5\,A \quad u = 500\,V$$

$$\sigma_i^2 = 0.1 \quad \sigma_u^2 = 0.02$$

$$N = 10 \qquad Ns = 500$$

$$\overline{\hat{R}_{|N}}(k) = \frac{1}{k}\sum_{i=1}^{k} \hat{R}_{|N}(i) \quad \text{pour} \quad k = 1, \cdots, N_s$$



Biased estimator !!

Exercise :

- Recall the problem of measuring the obstacle avoidance by ultrasound sensor.
- N measurements are performed

$$\underline{z} = [t_1 \quad \dots \quad t_N]^T \quad with \qquad t_i = \frac{2d}{C} + b_i$$

- The noise b_i is Gaussian, white noise with zero mean

- Two estimators are proposed to estimate the robot-obstacle distance:

$$\hat{d}_1 = \frac{C}{2} t_1 \qquad \hat{d}_2 = \frac{C}{2N} \sum_{i=1}^{N} t_i$$

- Evaluate the bias of these two estimators:
  - By analytical calculus
  - By MC simulations

Exercise : Evaluation by simulations

- Using the example of the ultrasound presented in Exercise 2. The Monte Carlo simulation was performed with the following parameters:

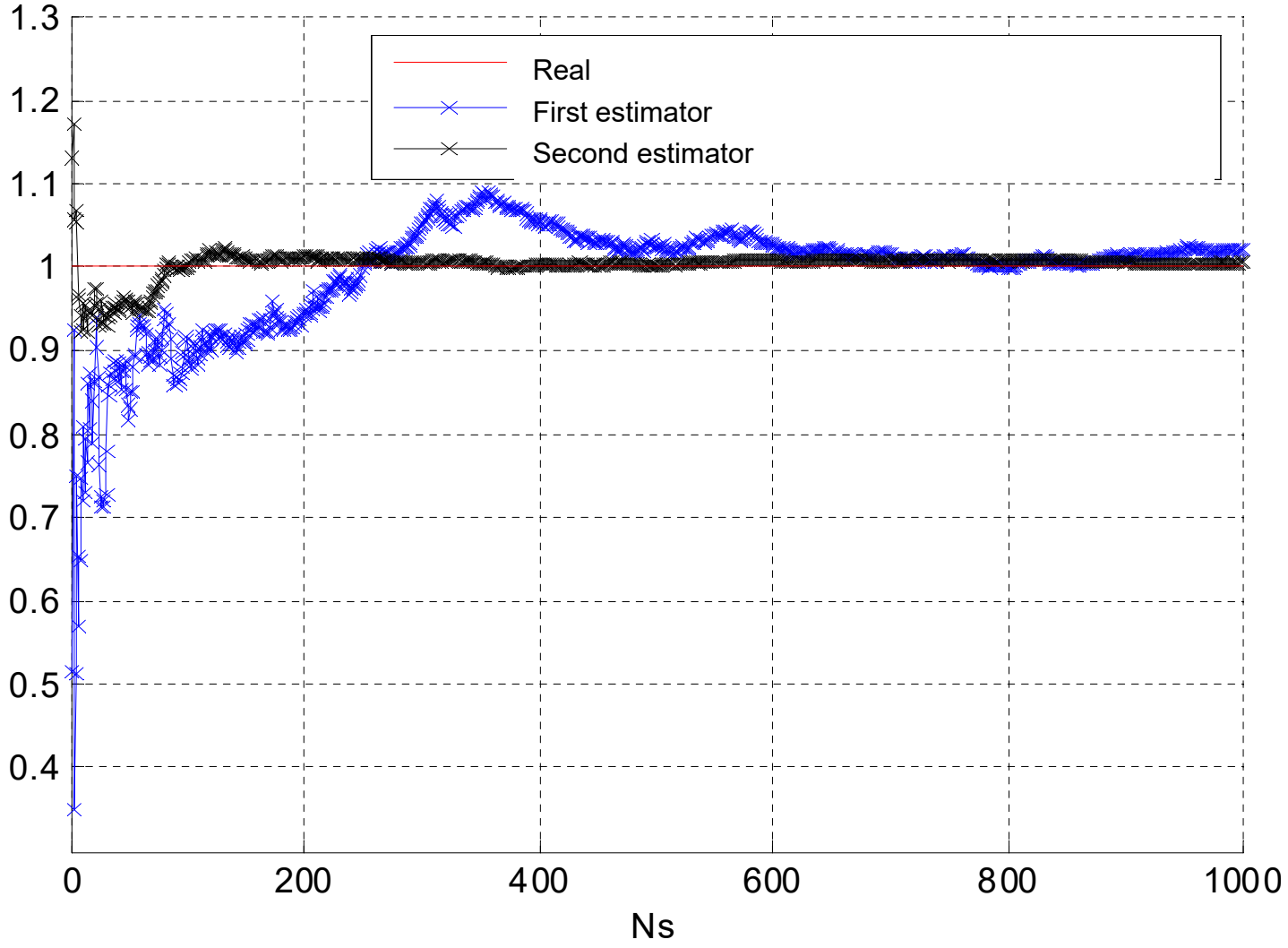
$$\text{Target value} : d = 1m \quad , \quad Ns = 1000 \quad , \quad C = 10 \quad , \quad N = 20 \quad and \quad \sigma_{b_i}^2 = 0.05$$

- We have generated the vector of measurements $\underline{z}$ using the "*randn.m*" function of Matlab

$$\underline{z} = [t_1 \quad \dots \quad t_N]^T \quad with \quad t_i = \frac{2d}{C} + b_i$$

- Following the procedure describe before in the deterministic case, we have the subsequent results

Exercise : Evaluation by simulations



✓ ***Empirical mean for the first estimator :***

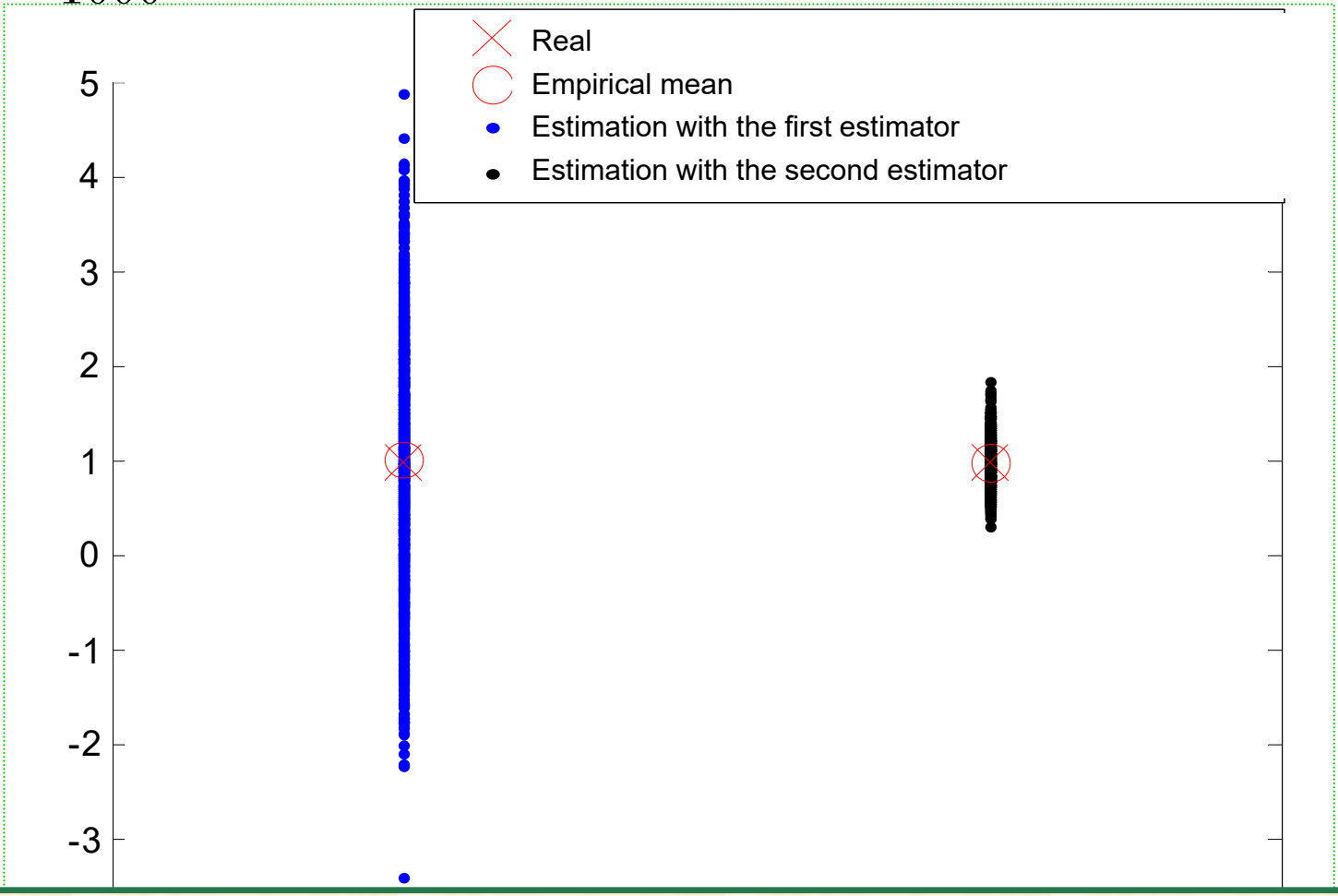
$$\hat{d}_1 \Rightarrow \overline{\hat{d}_1}(N) = 1.0209$$

✓ ***Empirical mean for the second estimator :***

$$\hat{d}_2 \Rightarrow \overline{\hat{d}_2}(N) = 1.0043$$

Exercise : Evaluation by simulations

- Monte Carlo simulation results: Fluctuation of estimates around the mean value for Ns = 1000



To compare performances of estimators we must also compare their variances

Covariance of an estimator

- Definition: The covariance matrix of an estimator is defined by the following symmetric, positive semi-definite matrix:

$$Cov(\underline{\hat{\theta}}) = E\left\{(\underline{\hat{\theta}} - m_{\hat{\theta}})(\underline{\hat{\theta}} - m_{\hat{\theta}})^T\right\} \qquad avec \quad m_{\hat{\theta}} = E\left\{\underline{\hat{\theta}}\right\}$$

- The diagonal elements of the covariance matrix reflect the fluctuations of the various components of the estimator around the value

Mean square error (MSE)

Remark: *if* $m_{\hat{\theta}} = 0$ *then* $\boldsymbol{cov(\hat{\theta})=MSE(\hat{\theta})}$

- Definition: The MSE of an estimator is the matrix defined by:

$$E\left\{\underline{\tilde{\theta}}\,\underline{\tilde{\theta}}^T\right\} \quad with \quad \underline{\tilde{\theta}} = \underline{\hat{\theta}} - \underline{\theta} \; the \; Estimation \; error$$

Efficiency

- **Definition** : An unbiased estimator $\underline{\hat{\theta}}$ of $\underline{\theta}$ is said **efficient** if its variance is lower than the variance of any other unbiased estimator $\underline{\hat{\theta}'}$ of $\underline{\theta}$, i.e.,

$$Cov(\underline{\hat{\theta}}) \leq Cov(\underline{\hat{\theta}'})$$

- We talk about efficient estimator or **minimum variance** estimator

Covariance evaluation: Monte-Carlo simulations

- As for the bias evaluation we shall use empirical calculus based on Monte-Carlo simulations to evaluate the covariance of an estimator.

- Recall that the covariance of a random vector is given by

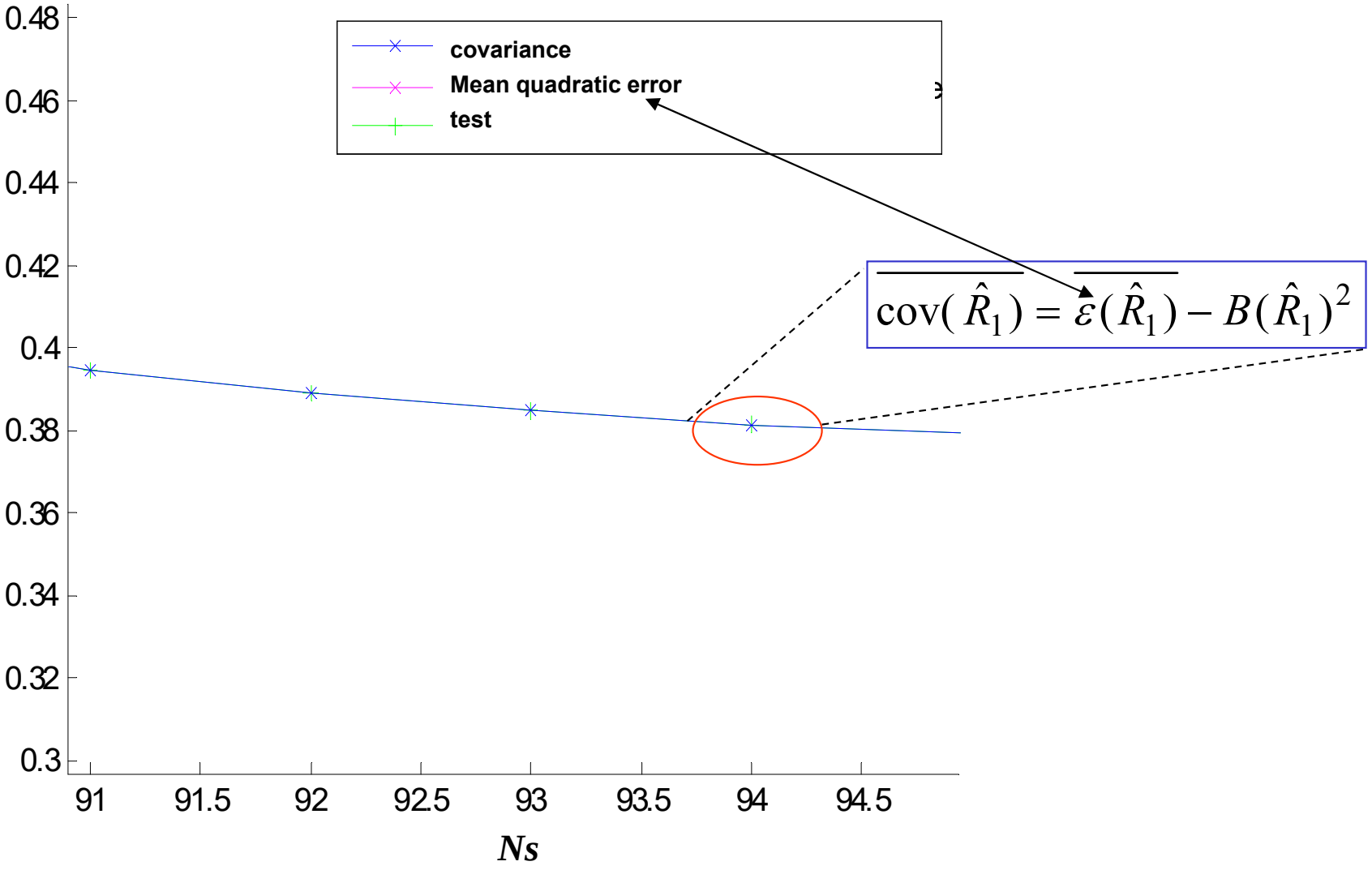
$$Cov\{\underline{x}\} = E\left\{(\underline{x} - m_x)(\underline{x} - m_x)^T\right\} \text{ with } m_x = E\{\underline{x}\}$$

- By performing $N_s$realization (**law of large numbers**), the empirical covariance is given by:

$$\overline{Cov\{\underline{x}\}} = \frac{1}{Ns}\sum_{k=1}^{Ns}(\underline{x}(k) - \bar{x})(\underline{x}(k) - \bar{x})^T \quad with \quad \bar{x} = \frac{1}{Ns}\sum_{k=1}^{Ns}\underline{x}(k)$$

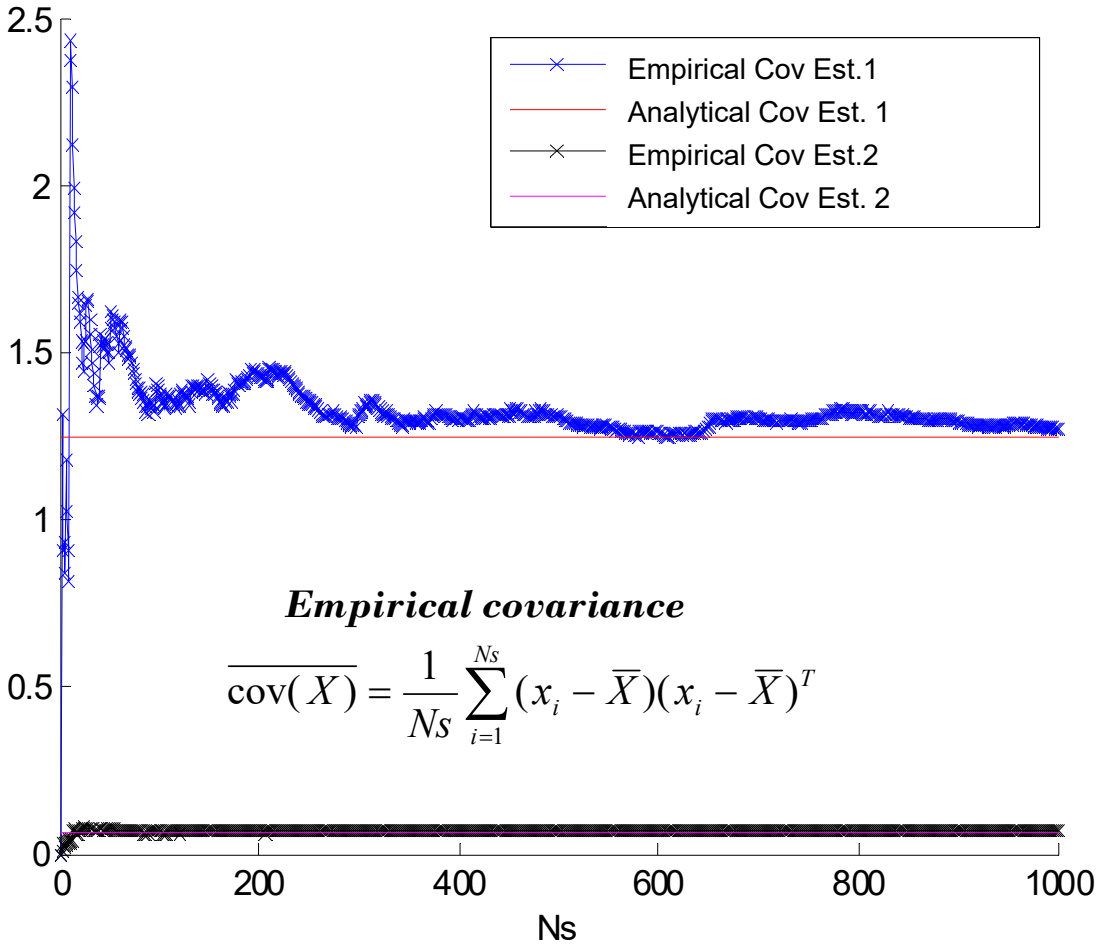
$$\lim_{Ns \to \infty} \overline{Cov\{\underline{x}\}} = Cov\{\underline{x}\}$$

Example : Estimation of the resistance R – Covariance and quadratic error



Exercise: Ultrasound sensor

- Calculate the covariance of the two estimators by performing analytical calculus and simulations.
- Monte-Carlo simulations



Empirical covariance

$$\overline{\mathrm{cov}(X)} = \frac{1}{Ns}\sum_{i=1}^{Ns}(x_i - \bar{X})(x_i - \bar{X})^T$$

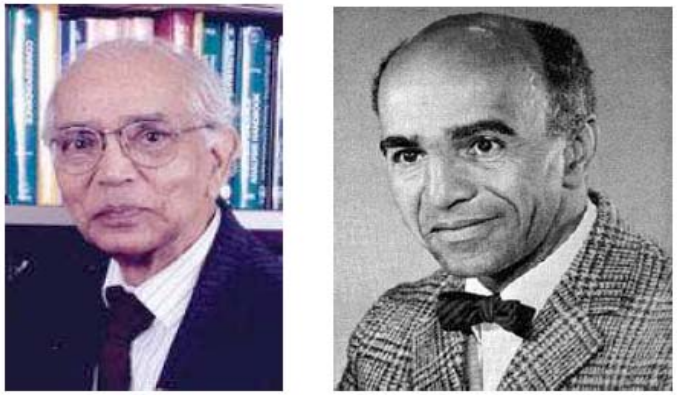
- ✓ ***Analytical covariance:***
  - ***Estimator 1 : 1.25***
  - ***Estimator 2 : 0.0625***
- ✓ ***Empirical covariance Ns = 1000***
  - ***Estimator 1 : 1.2719***
  - ***Estimator 2 :0.06469***

We prefer estimator 2

But is it the "best"?

Fisher Matrix and Cramer-Rao lower bound

- In a problem of determining an estimator, it is interesting to know the optimal performances that can be claimed. Regarding the bias, the lower bound to be achieved is **zero**. Similarly, we look for a *lower bound for the covariance* to hand out as a reference.
- It is the Cramer-Rao Lower Bound (CRLB). It is a *lower bound* on the variance of *any unbiased estimator*. Hence, *it is a powerful tool to compare estimators to a theoretical lower bound*.
- In fact, one of the most important results of the estimation theory is that for any unbiased estimator, the variance of the estimate is always greater than an expression depending on the probability density of measures.



Rao and Blackwell.

Fisher Matrix and Cramer-Rao lower bound (CRLB)

- Consider the case where the vector parameter $\underline{\theta} \in \mathbb{R}^p$ is deterministic and that the vector $\underline{z}(k)$ has $f_{\underline{z},\underline{\theta}}(\underline{z}/\underline{\theta})$ as PDF. We also assume that this density is differentiable with respect to $\underline{\theta}$:

$$\boldsymbol{V} = \frac{\partial}{\partial\theta}\boldsymbol{ln}\left(\boldsymbol{f}_{\underline{z},\underline{\theta}}(\underline{\boldsymbol{z}}/\underline{\boldsymbol{\theta}})\right) = \frac{\frac{\partial}{\partial\theta} f_{\underline{z},\underline{\theta}}(\underline{z}/\underline{\theta})}{f_{\underline{z},\underline{\theta}}(\underline{z}/\underline{\theta})}$$

$ln\left(f_{\underline{z},\underline{\theta}}(\underline{z}/\underline{\theta})\right)$ is *the* ***log-likelihood***

- Define the ***score*** ***by***

$$[I(\underline{\theta})]_{i,j} = \boldsymbol{E}\{(\boldsymbol{V} - \boldsymbol{E}(\boldsymbol{V}))^2\} = E\left\{\left(\frac{\partial \boldsymbol{ln}(f_{\underline{z},\underline{\theta}}(z/\theta))}{\partial\theta}\right)^2\right\} = -E\left\{\frac{\partial^2 \boldsymbol{ln}(f_{\underline{z},\underline{\theta}}(z/\theta))}{\partial\theta_i\partial\theta_j}\right\}$$

The score is a **random variable**. Its mean is $\boldsymbol{E(V)=0}$

- Then define the **Fisher matrix**, $\boldsymbol{I(\underline{\theta})}$, as the covariance of this score variable

Theorem: *For any unbiased estimator* $\hat{\underline{\theta}}$ *of* $\underline{\theta}$ *the following inequality holds*:

$$Cov(\hat{\underline{\theta}}) \geq I^{-1}(\underline{\theta}) \quad \Rightarrow \quad Cov(\hat{\underline{\theta}}) - I^{-1}(\underline{\theta}) \text{ is positive semi - definite}$$

$\boldsymbol{\theta}$ follows, in first approximation, a Gaussian law of expectation $\hat{\boldsymbol{\theta}}$ and variance $\boldsymbol{I^{-1}(\theta)}$
The bound is the best estimation precision attainable in the absence of a priori knowledge..

- This result is known as the **inequality of Cramer-Rao** and $I^{-1}(\underline{\theta})$ is the CRLB

Fisher Matrix and Cramer-Rao lower bound (CRLB)

Remarks:

- Note that the inequality of Fischer implies that the diagonal elements of the matrix $Cov\,(\hat{\underline{\theta}}) - I^{-1}(\underline{\theta})$ are nonnegative. Therefore,

$$\text{var}(\hat{\theta}) = E\left\{(\hat{\theta}_i - \theta_i)^2\right\} \geq \left[I^{-1}(\theta)\right]_{i,j}$$

- If an **unbiased** estimator reaches the Cramer-Rao lower bound, we can state that this estimator is **efficient**.
- For a given estimation problem, it is possible that the Cramer-Rao is not achievable. *That, does not mean that there is no efficient estimator.* We then use "**sufficient statistic**" to determine the estimator.
- In our applications, the vector of measurement $\underline{z}(k)$ consists often of N independent observations, that are supposed to be ***independent and identically distributed* (*i.i.d***). Hence, we have:
- which implies $$f_{\underline{z},\underline{\theta}}(\underline{z}/\underline{\theta}) = \prod_{k=1}^{N} f_{z_k,\underline{\theta}}(\boldsymbol{z(k)}/\underline{\boldsymbol{\theta}})$$

$$I(\theta) = N\,\boldsymbol{M(\theta)} \;\; with \;\; [\boldsymbol{M(\theta)}]_{i,j} = -E\left\{\frac{\partial^2 \boldsymbol{ln}\left(f_{z_1,\underline{\theta}}(z_1/\underline{\theta})\right)}{\partial\theta_i\partial\theta_j}\right\} \; i = 1,\ldots,p; j = 1,\ldots,p$$

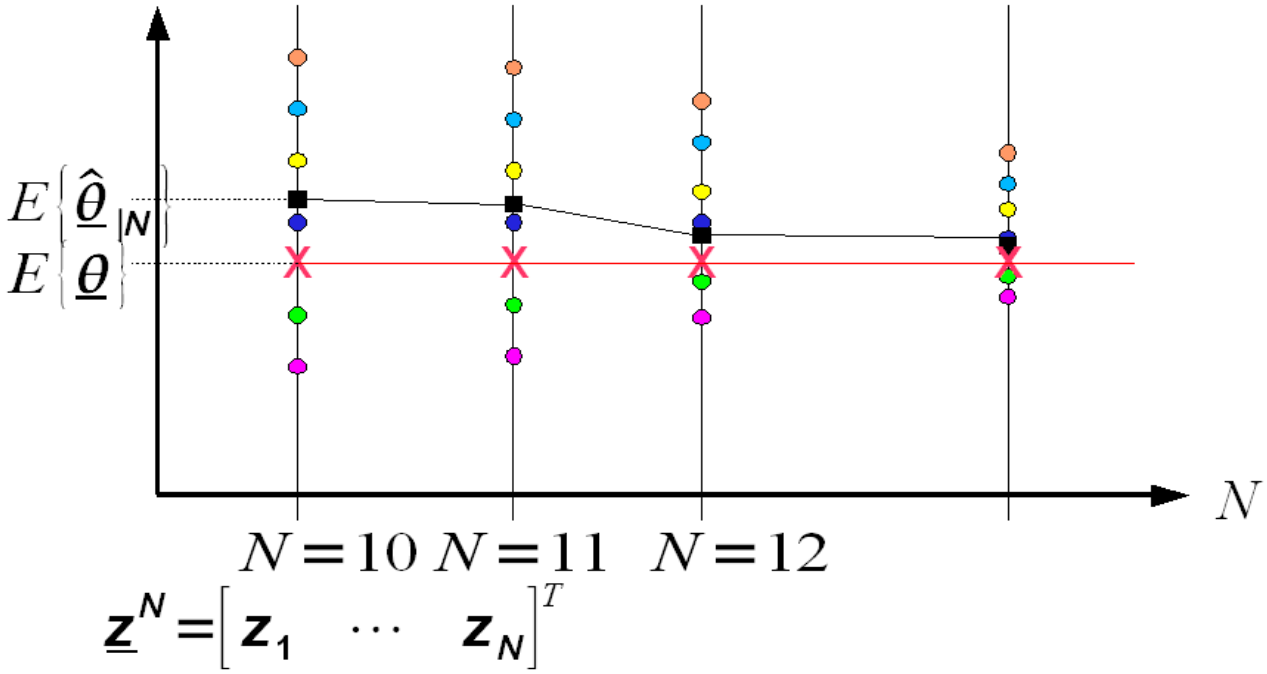
- The Cramer-Rao inequality may be, then, rewritten as

$$\boxed{Cov(\hat{\underline{\theta}}) \geq I^{-1}(\underline{\theta}) = \frac{1}{N}\boldsymbol{M^{-1}}(\underline{\theta})}$$

⇒ ***The CRLB is inversely proportional to the number N of observations.***

Asymptotic properties and convergence

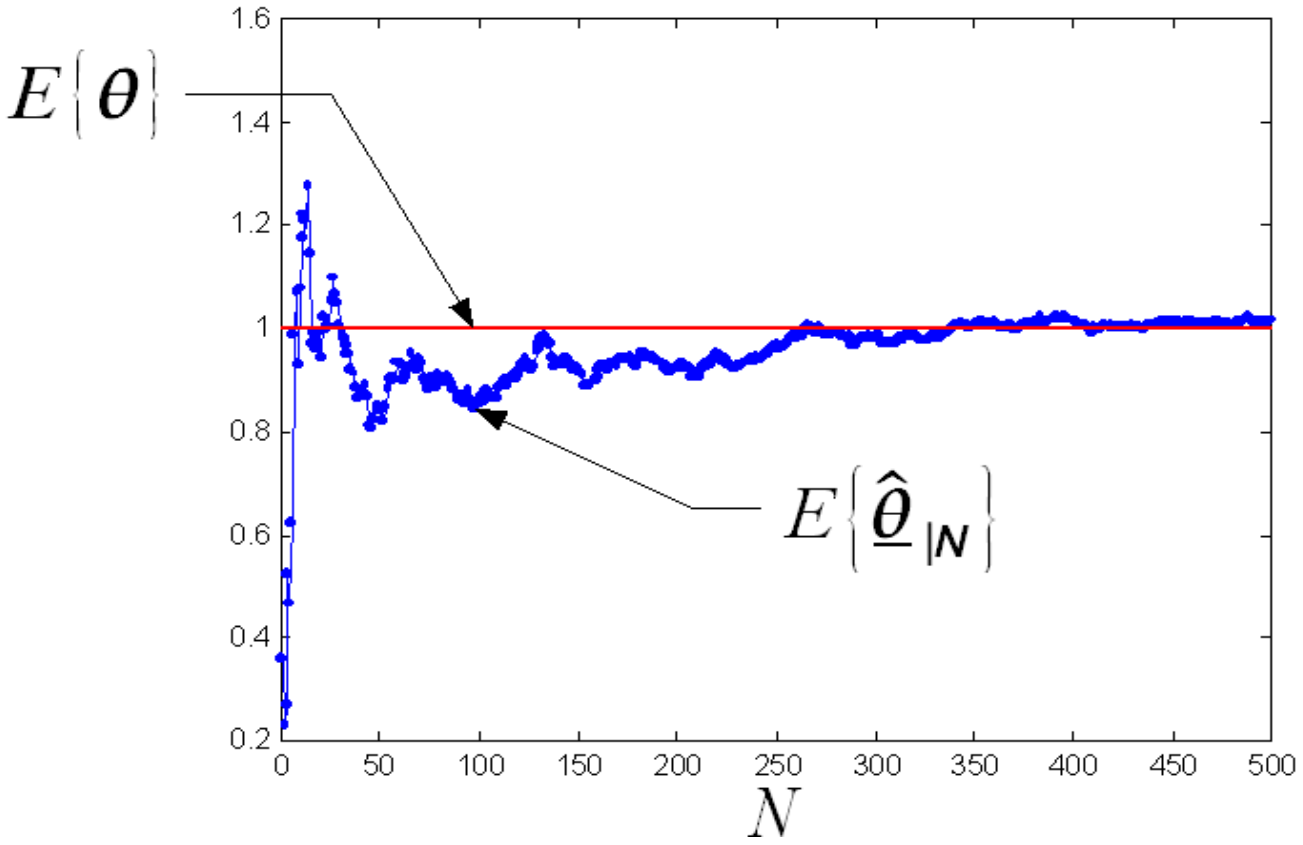
- Objective: inspect the properties of the estimator when the dimension of the observation vector $\underline{z}(k)$ **tend to infinity** (k or $N \to \infty$).
- We'll discuss two questions in particular:
  1) *Does our estimator work properly asymptotically? Or, does it provide us with the correct answer if we give it enough data?*
  2) *How asymptotically efficient is our estimator?*



Asymptotic properties and convergence

- An estimator is asymptotically unbiased when

$$\lim_{k\to\infty} E\left\{\underline{\hat{\theta}}(k)\right\} = E\left\{\underline{\theta}\right\}$$



Asymptotic properties and convergence

- There are several types of stochastic convergence. We give below the two most common definitions. We consider here a sequence of random vectors:

$$\underline{\hat{\theta}}_1, \underline{\hat{\theta}}_2, \ldots, \underline{\hat{\theta}}_k$$

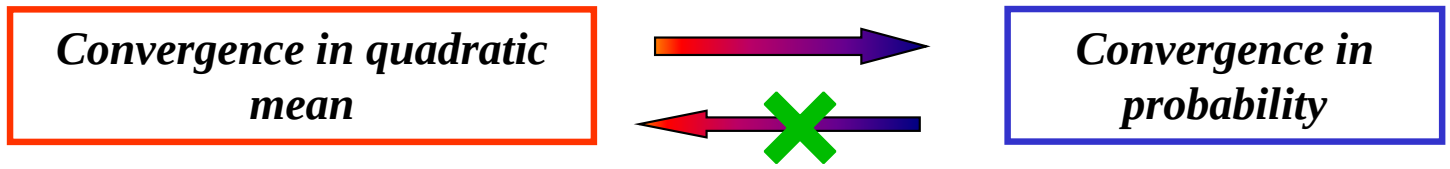
- **Consistency**: We say an estimator is *consistent or convergent in probability* if it provides us with the correct answer asymptotically. Hence, in terms of probability we have

$$\forall \varepsilon > 0 \quad : \lim_{k \to \infty} \Pr \text{ob} \left[\left\| \underline{\hat{\theta}}_k - \underline{\theta} \right\| > \varepsilon \right] = 0$$

- **Convergence in quadratic mean:** An estimator converge in quadratic mean if ,

$$\lim_{k \to \infty} E \left\{ \left\| \underline{\hat{\theta}}_k - \underline{\theta} \right\|^2 \right\} = 0$$

- Convergence in quadratic mean, indicates that the fluctuations around the parameter tend to zero when k tends to infinity.



Sufficiency

- Let's now, consider the statisticians view of the estimation problem:

"Statistical inference corresponds to an inverse problem:

"given data, we want to answer questions about the "true" underlying state, e.g., the true parameter indexing the distribution that gave rise to our observed data. Estimation is about choosing between a continuum of possible parameters …"

- The concept "sufficiency," or "data reduction" is then *to look how to decide which aspects of the data matter for inference, and which aspects can be safely ignored.*
- For example:
  - if we have ***i.i.d.*** data, it doesn't really matter what **order** the data appeared in. Hence, we can throw out the order of the data and do inference given the unordered data just as well.
  - we don't need to remember everything about Gaussian data, just the sample mean and sample variance. This is quite a savings: *we've compressed N data points into just two* !!

Sufficiency

- **Statistic**: we call "***statistic***" any function $T(\boldsymbol{z})$ of the data $\boldsymbol{z}$ (and only of the data $\boldsymbol{z}$). For example

  is a statistic. $$T(z) = \sum_{k=0}^{N-1} \boldsymbol{z}(\boldsymbol{k})$$

- **Sufficient statistic** (*statistique exhaustive*): if the knowledge of any other statistic does not provide any additional information on $\underline{\theta}$ and therefore cannot be estimated better. Somehow, $T(\boldsymbol{z})$ *summarizes all available information on the parameter from the components of* $\boldsymbol{z}$. This also means, that $p(\boldsymbol{z}/\theta)$ depends only on $\underline{\theta}$.

- Indeed, even if that was the case, it would mean that some values of $\boldsymbol{z}$ provide further information on $\underline{\theta}$ which contradicts the definition of sufficiency.

- **Neyman Fisher factorization theorem** : *$T(\boldsymbol{z})$ is a sufficient statistic to estimate $\underline{\theta}$ if and only if the probability density $p(\boldsymbol{z}/\theta)$ can be written in the following form for some function $g(.)$ and $h(.)$:*

  $$p(\mathbf{z}/\underline{\theta}) = g\big(T(\mathbf{z}), \underline{\theta}\big)h(\mathbf{z})$$

- It can easily be seen that the maximum likelihood estimate of $\underline{\theta}$ interact with $\boldsymbol{z}$ only through $T(\boldsymbol{z})$.

Exercise:

We want to estimate the value of a constant ***C*** immersed in a white Gaussian noise (*DC level in noise*) with zero mean and variance σ^2. The observation equation is:

$$y(k) = C + b(k)$$

We also suppose that

$$E\{b(i)b(j)\} = 0 \qquad \forall i \neq j$$

We propose two estimators

1. $\hat{C}_1(k) = y(0)$
2. $\hat{C}_2(k) = \frac{1}{N}\sum_{k=0}^{N-1} y(k)$

1. **Compare these two estimators**
2. **Are they efficient ?**
3. **Do we need to construct another estimator ?**

Exercise

- Consider the following measurement model:

$$y(k) = c_1 + c_2 k + n(k)$$

where c_1 and c_2 are two constants to be identified and $n(k)$ a Gaussian white noise with zero mean and σ^2 as variance.

1. Calculate the Fisher matrix associated with estimation of the two constants
2. Is there an efficient estimator? if so, determine it.
3. Which of these parameters, c_1 and c_2, could be estimate with the highest accuracy ?
4. Which constant parameter c_1 or c_2 changes, $x(k)$ is more sensitive ?

SYLLABUS

- **Lecture 1:**
  - Course organization
  - Motivations
  - Introductive Examples
  - Review of Probability
- **Lecture 2:**
  - Estimation theory
  - Exercises
- **Lecture 3:**
  - Linear estimators
  - Exercise
- **Lecture 4:**
  - Linear estimator: Kalman Filter
  - Exercises & Matlab class
- **Lecture 5:**
  - Nonlinear estimation: EKF and UKF
  - Exercises & Matlab class
- **Lecture 6:**
  - Nonlinear estimation: PF and RBPF
  - Exercises & Matlab Class