Intelligent distributed systems

Daniele Fontanelli

Department of Industrial Engineering
University of Trento
E-mail address: *daniele.fontanelli@unitn.it*

2022/2023

UNIVERSITÀ DI TRENTO | Dipartimento di Ingegneria Industriale

Outline

1. Data Analysis
2. Measurement processes
   - The sensor
   - Random effects
3. Estimation Algorithms
4. Take home message

Outline

1. Data Analysis
2. Measurement processes
   - The sensor
   - Random effects
3. Estimation Algorithms
4. Take home message

Data Analysis

We already saw that there is an undisputed need of *analysing the measured data* to extract meaningful information from a production site, a plant, an urban area, etc.
Perhaps the most relevant application in this case is a *SCADA* system, which is critically related to the collection of measurements, to their analysis and to the representation of those data to a human operator with a HMI.
Usually *data collection analysis and transformation*, *data storage* and *data aggregation* are responsible for data management.

Data management

For any possible application, the main issues related to the *data management* are:

- *Data protection*: This is mainly related to *privacy*, that is how to control owned information. Keywords are: *certification* and *data anonymising*.
- *Data processing*: This is mainly related to the execution of data analysis. The keywords here are *big data* and *data ownership*.
- *Data analysis*, which is the *focus for this course*.

Data analysis

There are three main goals for *data analysis*:

- *Extract meaningful information from data*: this is usually related to *model estimation*, in order to illustrate the behaviour of a given complex system and to predict its future evolution.
- *Resources optimisation*: mainly bandwidth, communication QoS and components lifetime.
- *Estimate actual values from measurements*: this is the classic problem of *estimation*, *main focus for this course*.

Data analysis

Depending on the type of information to be extracted from the data, different approaches are available:

- The identification of possible problem in production systems is carried out using *process mining*: *what are the possible point of failures inside the production process?*
- If aggregated information is needed, *statistics* is adopted: *what is the mean number of users of a service (e.g. electric power, communication network, etc.) in the day-time?*
- To derive relations in the datasets, *machine learning* is usually adopted.

Data analysis

Machine Learning examples

- The relation, if any, between *dependent variables* from *independent variables* given a data set can be retrieved using *regression*: *what is the relationship between weather condition and power consumption?*
- The automatic definition of *categories* (aka *clusters*) in a data set is the goal of *clustering* algorithms: *what is the typical behavioural pattern of the users of a certain service?*
- Identifying to which *category* a certain variable belongs to is instead a *classification* problem (a subset of *pattern recognition*): *is this power grid state leading to a possible incoming black-out?*

In this course we will consider only the *estimation problems*, which are of major concerns for *distributed autonomous systems*.

Outline

1 Data Analysis

2 Measurement processes
- The sensor
- Random effects

3 Estimation Algorithms

4 Take home message

Measurement process

Any state estimation or localisation problem relies on the availability of *measurements*.

Definition (*International Vocabulary of Metrology* (VIM))

The *measurement* is the process of experimentally obtaining one or more *quantity values* that can reasonably be attributed to a *quantity*.

Definition (VIM)

The quantity intended to be measured is called the *measurand*.

The result of the *measurement process* is the *measurement result*.

Measurement process

Definition (VIM)

The *quantity* is the property of a phenomenon, body, or substance, where the property has a magnitude that can be expressed as a number and a *reference*.

The *reference* can be a *measurement unit* (e.g. the metre), a *measurement procedure* or a *reference material* (e.g. in chemical reactions).
The main issue is that the measurement result *cannot be* expressed by just a number.
Moreover, the measurement result can only provide a *finite amount of information* about the measurand.
So what is a *measurement result*?

Measurement process

Definition (VIM)

The *measurement result* is the set of quantity values being attributed to a measurand together with *any other available relevant information*.

The *main* attribute is called *uncertainty*. Therefore:

Definition (VIM)

The *measurement result* is generally expressed as a *single measured quantity value* and a *measurement uncertainty*.

All the previous definitions pertain to the field of science called *metrology*.

Measurement process

"*The* measurement process *maps a single manifestation* q *of a property (for instance weight, length,...) of an object (that is a phenomenon, a body, or a substance) belonging to the empirical world onto a symbol* x *that belongs to the symbolical world.*" (A. Ferrero, D. Petri, P. Carbone and M. Catelani, "MODERN MEASUREMENTS Fundamentals and Applications", IEEE Press, 2015)
Formally, let us define the set of possible manifestations of q as $\mathcal{Q}$ and let $\mathcal{Z}$ be the class of all the symbols z, than a *measurement process* is

$$\eta : \mathcal{Q} \to \mathcal{Z},$$

while

$$z = \eta(q).$$

is the *measurement result*.

Measurement process

Uncertainty

According to the *Guide to the Expression of Uncertainty in Measurement* (GUM), the word *uncertainty* means *doubt* about the information provided by the measurement result.
Notice that *uncertainty* expresses also the *quantitative measure* of the doubt.
There are different origins of the *uncertainty*:

- *Definitional uncertainty*: it is related to the modelling errors (e.g. assuming that a surface is actually planar);
- *Interaction uncertainty*: physical interaction of the measurement system with the measurand (e.g. a thermometer measuring the temperature of a fluid);
- *Instrumental uncertainty*: the imperfect behaviour of the measurement system (e.g. the use of reference quantities for comparison, which are actually uncertain).

Measurement process

Uncertainty

The *uncertainty* has different effects on the *measurement results*, which can be grouped in two broad sets: *systematic* and *random* effects.
Systematic effects are constant in time and very dangerous.
A standard way to detect them is to compare the measurement results obtained by using *two independent measuring systems*, where one of the two is known to be with *negligible* systematic effects.
In general, the *systematic effects* depend on some *influence quantities* (e.g. the temperature).

Measurement process

Systematic effects

In the GUM it is recognised the *high relevance* of such uncertainties: "*It is assumed that the result of a measurement has been corrected* for all *recognised significant systematic effects and that every effort has been made to identify such effects.*"
Usually the systematic effects are characterised by the *manufacturer* or by the *calibration certificates*.
Let us clarify how the systematic effect can be mitigated using *regression techniques*.

Outline

1. Data Analysis
2. Measurement processes
   - The sensor
   - Random effects
3. Estimation Algorithms
4. Take home message

Measurement process

The sensor

The term *sensor* defines the set of all the possible (usually) electro–mechanical systems that convert physical phenomena to measurable signals, typically voltages or currents.
The sensor is one component of a *measuring instrument*.
Sensors are also called *transducers* and implements the mapping $\eta : \mathcal{Q} \to \mathcal{Z}$ implementing the *measurement process*.
Consider a simple pressure measuring device: the output voltage proportionally increases as the pressure increases. A digital device could measure the voltage, and convert it to a pressure. Each of this processes introduces *uncertainties*.
Most of these sensors are based on subtle electrical properties of materials and devices. As a result the signals often require *signal conditioners* (e.g., amplifiers).

The sensor

The basic physical phenomena typically measured with sensors include:

(1) Angular or linear position;
(2) Acceleration;
(3) Temperature;
(4) Pressure or flow rates;
(5) Stress, strain or force;
(6) Light intensity;
(7) Sound.

The sensor

Critical characteristics

- Real sensors convert an input phenomena to an output of a different type. This transformation relies upon a manufactured device, with limitations and imperfection. Therefore real sensors are *uncertain*;
- Real sensors returns a *partial description of the environment*, i.e., only finite and well precise phenomenon can be directly observed;
- Real sensors are *difficult to be modelled*.

As a consequence, data coming from real sensors should be managed carefully, taking into account their *uncertainties* and adopting model–based reconstruction algorithms.
Let us take a closer look...

The sensor

Characteristics

Sensors are mainly characterised by three parameters, directly derived from *metrology*:

- *Accuracy*: closeness of agreement between a measured quantity value and a true quantity value of a *measurand*. A measurement is said to be more accurate when it offers a smaller *measurement error*. Hence, a measurement is said to be *more accurate* when it offers a smaller measurement error (see VIM, 2.13, Note 1).

The sensor

Characteristics

- *Precision*: closeness of agreement between *measurement results* obtained by replicate measurements on the same or similar objects under specified conditions. Measurement precision is usually expressed numerically by measures of *uncertainty*, such as *standard deviation*, *variance*, or coefficient of variation under the specified conditions of measurement. Measurement precision is used to define *measurement repeatability*, *intermediate measurement precision*, and *measurement reproducibility*.

Measurement Processes

Pre

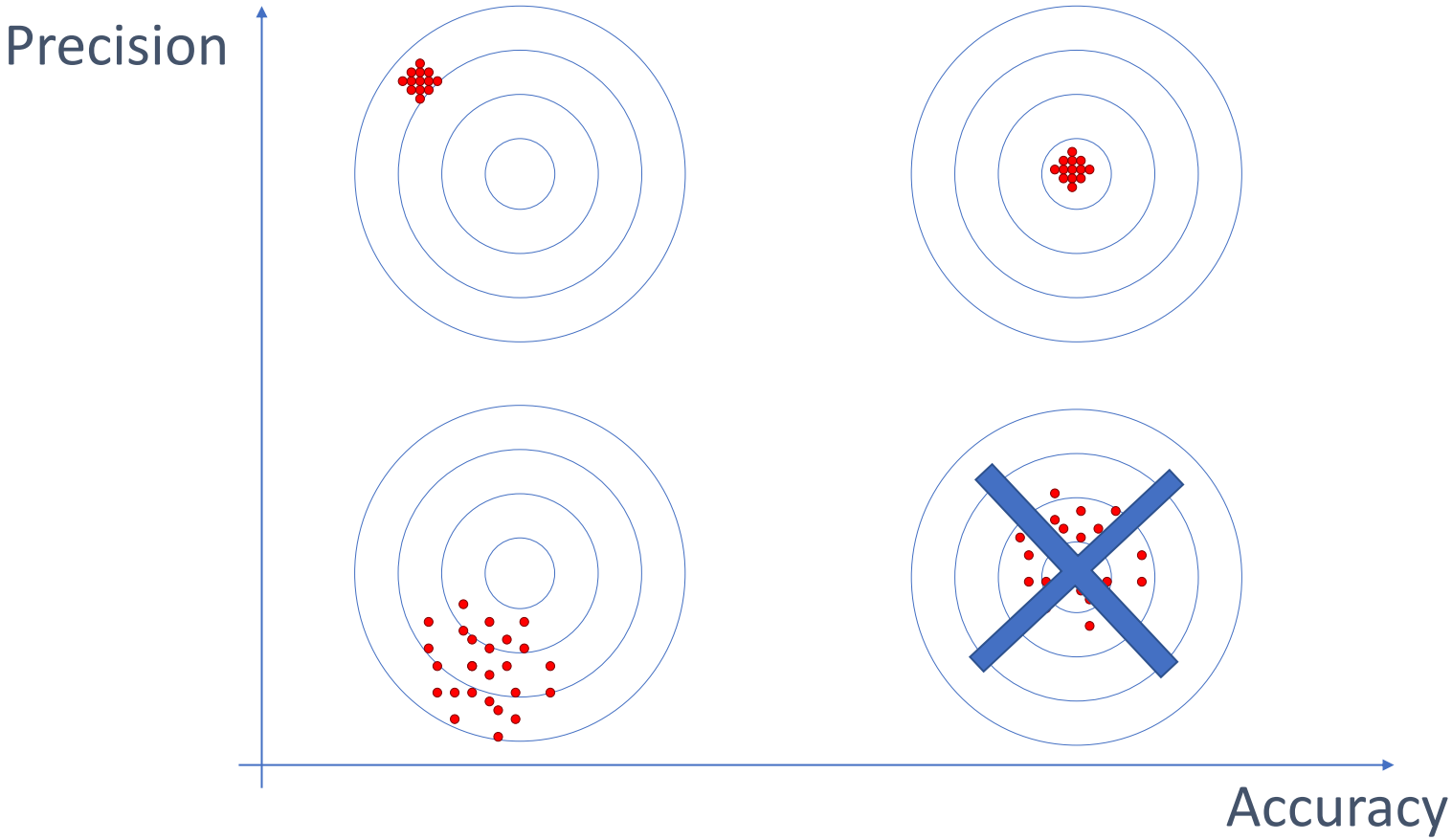


Figure: Precision and accuracy of measurements.

The sensor

Characteristics

- *Repeatability*: measurement precision under a set of *repeatability conditions of measurement*, e.g., the condition of measurement, out of a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time. Roughly speaking, this concept deals with *time invariance of the measurement results*.

The sensor

Characteristics

Other important properties are:

- *Resolution*: the smallest increment that the sensor can detect (due to *quantisation*).
- *Linearity*: a linear sensor has a linear relationship between the input phenomenon and the output signal. This way, it is easier to derive the measured quantity from the sensor readings.
- *Range*: the range of the measurable inputs and the range of the associated outputs.
- *Dynamic Response*: the frequency range for nominal operation of the sensor (e.g., cameras or microphones). Typically, there is an upper operation frequency, while a lower frequency is occasional.
- *Robustness*: tolerance with respect to deviations from nominal operational conditions, e.g., changes in temperature, humidity, dirt/oil, pressure, mechanical vibrations.

The sensor

Characteristics

- *Calibration*: many sensors need some calibration to determine or set the relationship between the input phenomena and the output. Calibration is usually performed when they are manufactured or installed. Nevertheless, mainly due to environmental characteristics, it may be needed frequently (e.g. due to *systematic effects*).
- *Speed of operation*: the rate at which the sensor returns the measurements in continuous mode and/or the delay until the measurement is returned when it is requested intermittently.
- *Error rate*: the time of inter-arrival of erroneous measurements, e.g., outliers or missed measurements.
- *Computational cost*: the amount of data processing needed to achieve the desired information. For example, a switch may not require any computation, while a vision processing algorithm can be very time consuming.

The sensor

Characteristics

- *Cost*: more precision costs more. In the sensor cost, should be also considered the cost of the conditioning equipment. For example, some sensors are very inexpensive, but the signal conditioning equipment costs are significant.
- *Power, weight and size*: the physical encumbrance and the power requested for the operation (some sensors can be switched on and off).

The sensor

Classic sensor response

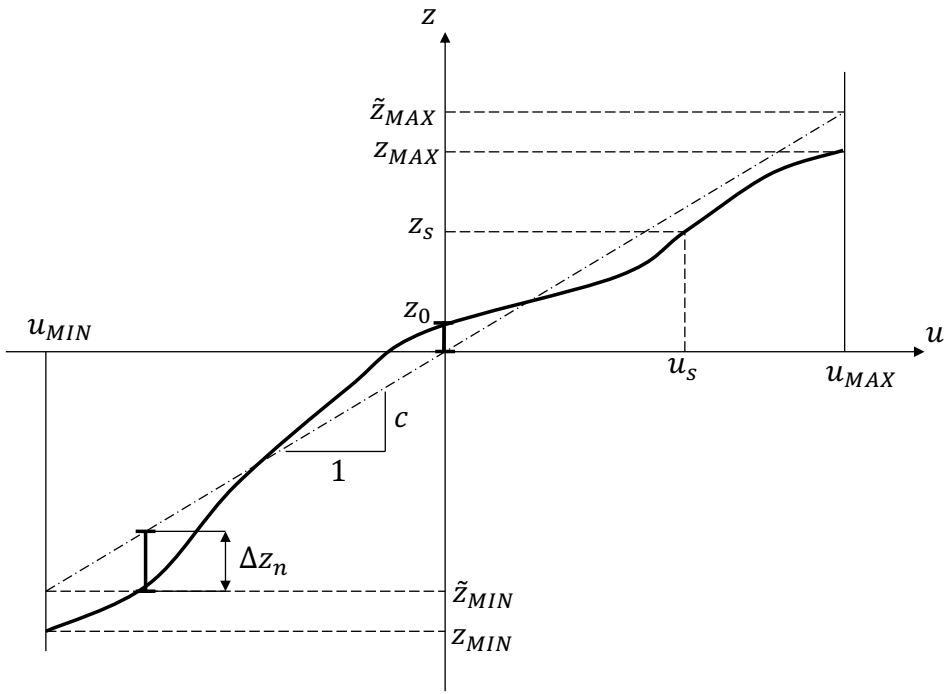


Figure: Classic quasi-static sensor response.

The *sensor response* relates the quantity to be measured u, i.e., input phenomenon, with the measured value z, i.e., the sensor output.

The sensor

Sensor calibration

The sensor

Calibration

In order to *calibrate* the sensor, the curve of the sensor response has to estimated.
Suppose the sensor response can be described as
$z = a_0 + a_1 u + a_2 u^2 + \cdots + a_q u^q$.
Suppose that a sufficient number of measurements $n \geq q + 1$ is collected and stored in a matrix form, i.e.,

$$z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} 1 & u_1 & u_1^2 & \dots & u_1^q \\ 1 & u_2 & u_2^2 & \dots & u_2^q \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & u_n & u_n^2 & \dots & u_n^q \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_q \end{bmatrix} = Ua,$$

where $z \in \mathbb{R}^n$, $U \in \mathbb{R}^{n \times q+1}$ and $a \in \mathbb{R}^{q+1}$.

The sensor

Calibration

To find the best set of the *calibration parameters* $a = [a_0, a_1, \ldots, a_q]^T$ it is possible to minimise the *mean squared error*

$$J(\widehat{a}) = \sum_{i=1}^{n} [z_i - (\hat{a}_0 + \hat{a}_1 u_i + \hat{a}_2 u_i^2 + \cdots + \hat{a}_q u_i^q)]^2,$$

or equivalently in matrix form

$$J(\widehat{a}) = (z - U\widehat{a})^T (z - U\widehat{a}) = z^T z - \widehat{a}^T U^T z - z^T U\widehat{a} + \widehat{a}^T U^T U\widehat{a}.$$

The optimal choice of the parameters is then given by

$$\widehat{a} = \arg \min_{\widehat{a} \in \mathbb{R}^{q+1}} J(\widehat{a}).$$

This is an *unconstrained minimisation problem*, i.e. a *regression*.

The sensor

Calibration

It is then sufficient to impose that the *gradient* of the cost index is equal to zero for the optimal value, i.e.,

$$\frac{dJ(\widehat{a})}{d\widehat{a}} = 0.$$

To compute the derivative we first notice to properties of the matrix product derivatives (obtained by simple algebra):

- Given $x, y \in \mathbb{R}^n$, $\frac{dx^T y}{dx} = \frac{dy^T x}{dx} = y$;
- Given $x \in \mathbb{R}^n$ and $B \in \mathbb{R}^{n \times n}$, $\frac{dx^T Bx}{dx} = Bx + B^T x$. If B is also *symmetric* (i.e., $B = B^T$), it follows that $\frac{dx^T Bx}{dx} = 2Bx$.

The sensor

Calibration

Therefore

$$\frac{dJ(\widehat{a})}{d\widehat{a}} = -2U^T z + 2U^T U\widehat{a}.$$

Since the *Hessian* is given by

$$\frac{d^2 J(\widehat{a})}{d\widehat{a}^2} = 2U^T U,$$

which is positive definite, the minimum is given by imposing

$$\frac{dJ(\widehat{a})}{d\widehat{a}} = 0 \Rightarrow \widehat{a} = \left(U^T U\right)^{-1} U^T z = U^\dagger z,$$

where $U^\dagger$ is the *pseudo-inverse* of U.

The sensor

The pseudo-inverse

Notice that $U^\dagger$ exists whenever $U^T U$ is *full rank*, i.e., a number of *independent* measures $n \geq q$.
Moreover, $U^\dagger = U^{-1}$ if $n = q$.

The sensor

Solution

The solution thus found is *optimal* with respect to the cost index $J(\widehat{a})$ chosen, which is the sum of the *squared residuals* between the measurements and the model, i.e., the sum of the squares of the errors made in the results of every single equation.
Hence, we have solved the *Least Squares* (LS) problem, which is a standard approach to the approximate solution of overdetermined systems.
In practice, calibration is used to reduce at most the *measurement error*, which affects the *accuracy*.

Measurement Processes

Pre

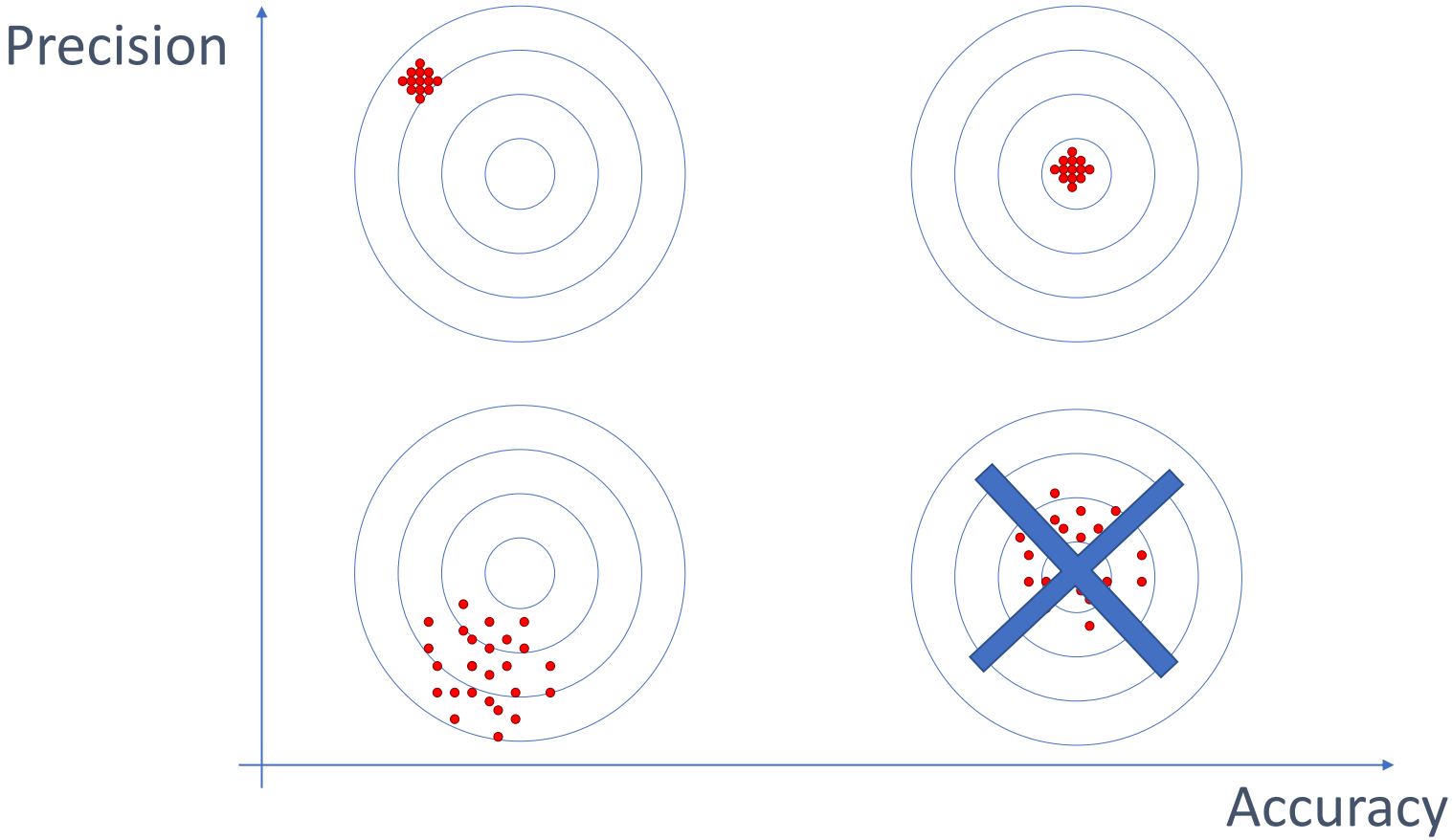


Figure: Precision and accuracy of measurements.

The sensor

Accuracy

Remark

Accuracy *should be part of the* characterisation of a measurement instrument *but not part of the characterisation of measurement results, since evaluating accuracy requires a* reference value *to be known, which is not the case for measurement results.*

Definition (GUM)

Measurement uncertainty is a parameter, associated with the result of a measurement, that characterises the *dispersion of the values that could reasonably be attributed to the measurand*.

Definition (ISO 5725)

The *trueness* is the closeness of agreement between the *average value* obtained from a large series of test results and an accepted reference value.

Measurement Processes

Pre

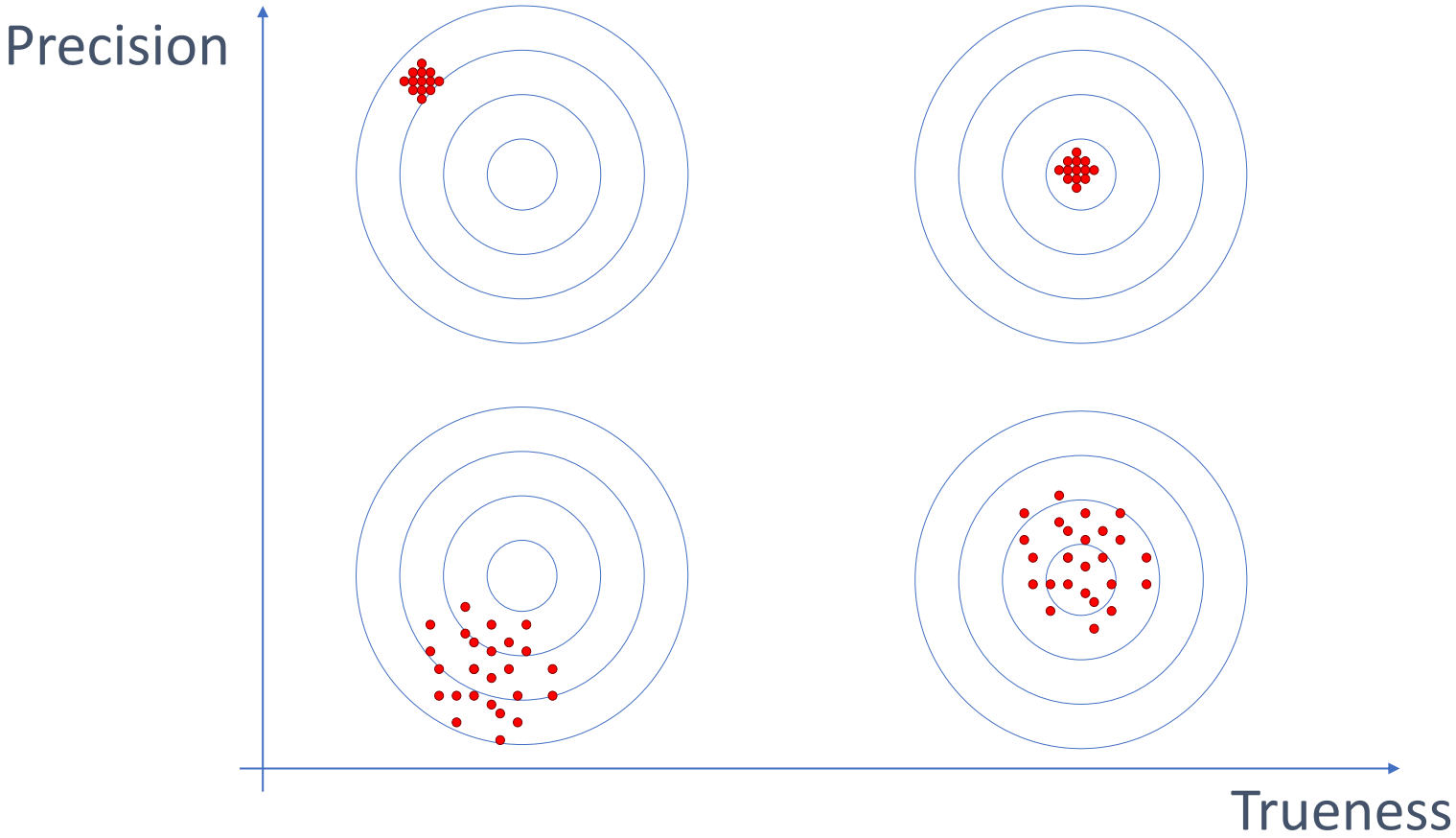


Figure: Precision and trueness of measurements.

Measurement Processes

Precision and accuracy

With the previous figure in mind, we can draw some remarks:

- We *cannot* express a measurement result *with a single value*, because the same value will be barely obtained by a new measurement!
- The single measurement value *cannot* be used in a comparison.
- The *measurement result* is represented by the whole set of values!
- It is clear that the *measurement uncertainty*, i.e. the *random effects*, can be characterised by the *dispersion* of the data around the *measured quantity*.

Outline

1. Data Analysis
2. Measurement processes
   - The sensor
   - Random effects
3. Estimation Algorithms
4. Take home message

Random effects

As reported in the VIM, it is now clear why a *measurement result* cannot be expressed by a single number.
According to the GUM, to apply a *correct measurement* we have to ensure that:

- The method to express the uncertainty should be *universal*, i.e. applicable to any situation and measurand;
- The expression of the uncertainty should be *internally consistent*, i.e. directly derivable from the components that contribute to it, and *transferable*, i.e. the uncertainty of a measurement should be derived by the components uncertainty.

Moreover the GUM states that the uncertainty should be provided by an *interval* with an associated *level of confidence* or *coverage probability*.

Random effects

The GUM suggests to *model the measured values as realisations of a random variable*.
Provided that the *systematic effects* are negligible, the GUM assumes that *the best estimate of the measurand is the mean value of the different realisations*.
Hence, the distribution of the measured variables is represented by a *probability density function* (pdf) and the related *standard deviation* provides a quantitative estimate of its dispersion.

Measurement Processes

Precision, accuracy and trueness

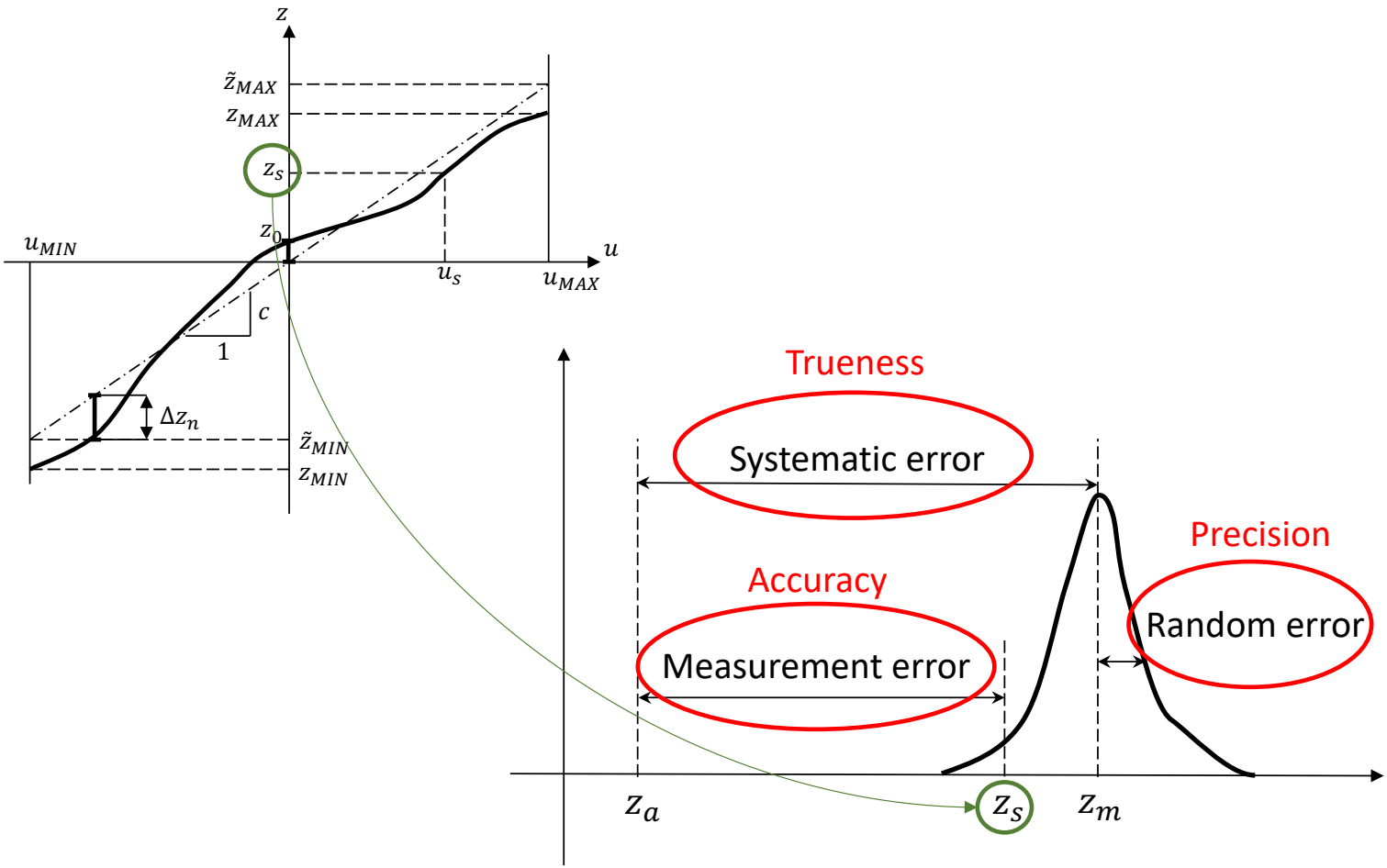


Figure: Precision, trueness and accuracy of measurements.

Random effects

Definition (GUM)

The *standard uncertainty* is the uncertainty of the result of a measurement expressed as a *standard deviation*.

Usually, the *standard uncertainty* of the measured value z is denoted with $u(z)$, even though we will use interchangeably the classic notation σ_z.
The *coverage probability* (or *level of confidence*) is expressed with respect to an interval about its mean value μ_z, i.e. $[\mu_z - u(z), \mu_z + u(z)]$.
For example, if the pdf is Gaussian, the *coverage probability* is given by the well known 68.27%.

Random effects

Since the GUM requires that the *coverage interval* comprises a *large* fraction of the possible values of the measurand:

Definition (GUM)

The *expanded uncertainty* is the quantity defining an interval about the result of a measurement that *may be expected to encompass a large fraction of the distribution of values* that could reasonably be attributed to the measurand.

The *large fraction* can be viewed as the *coverage probability*, hence the *expanded uncertainty* $U(z)$ exists if and only if the pdf is *known*.
The *expanded uncertainty* is given with a *coverage factor*. For the previous example, $[\mu_z - U(z), \mu_z + U(z)]$ with 99.73% if the *coverage factor* is 3.

Random effects

In the GUM, there are two ways in which the uncertainty can be evaluated: *Type A* and *Type B* evaluation.

Definition (GUM)

Type A evaluation method of uncertainty uses the *statistical analysis* of a series of repeated *observations*.

Let us make an example.

Random effects

Type A evaluation

When a measure is carried out, the result is represented by a *numeric array* or a *function*.
To measure the *uncertainty* one may use the *measurement error*

$$w_i = z_i - z_a,$$

being z_i the i–th value that is measured and z_a the actual value.
For example, $z_a = 25^\circ$ Celsius is the actual temperature, while the thermometer measures $z_i = 24.8^\circ$ $(w_i = -0.2^\circ)$.
It has to be noted that this approach cannot be actually pursued since the *actual value* of a measure is *ideal* and is not accessible, i.e. the actual value of the *measurand* is *unavailable*.

Random effects

Type A evaluation

By computing n times the measure, we obtain a series of *random values*, generated by the time-varying nature of any measurement process, i.e. the *random effects*.
These are due by the limited *resolution*, *precision* and *accuracy* of the measurements.
For example, $z_a = 25°$ Celsius is the actual temperature, while the thermometer measures a sequence of
$Z = [24.8°, 23.9°, 25.1°, 24.9°, 25.3°]$.
To the sequence of *measurement results* it is possible to associate the *statistical description*.

Random effects

Type A evaluation: Classic sensor response

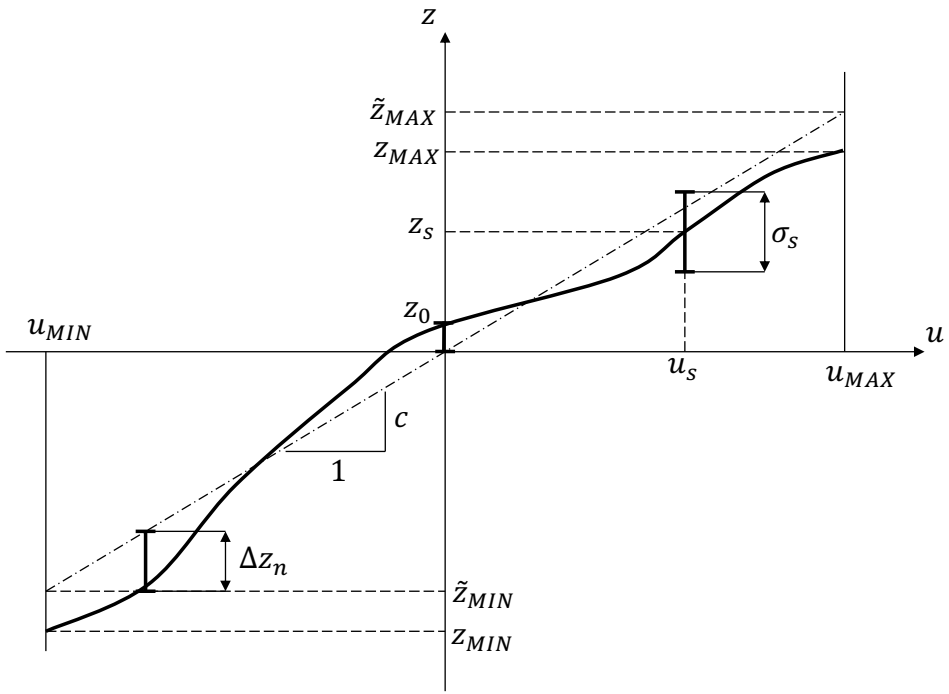


Figure: Classic quasi-static sensor response with *uncertainties*.

The *sensor response* is now affected by uncertainties, clearly visible for the quantity u_s.

Random effects

Type A evaluation

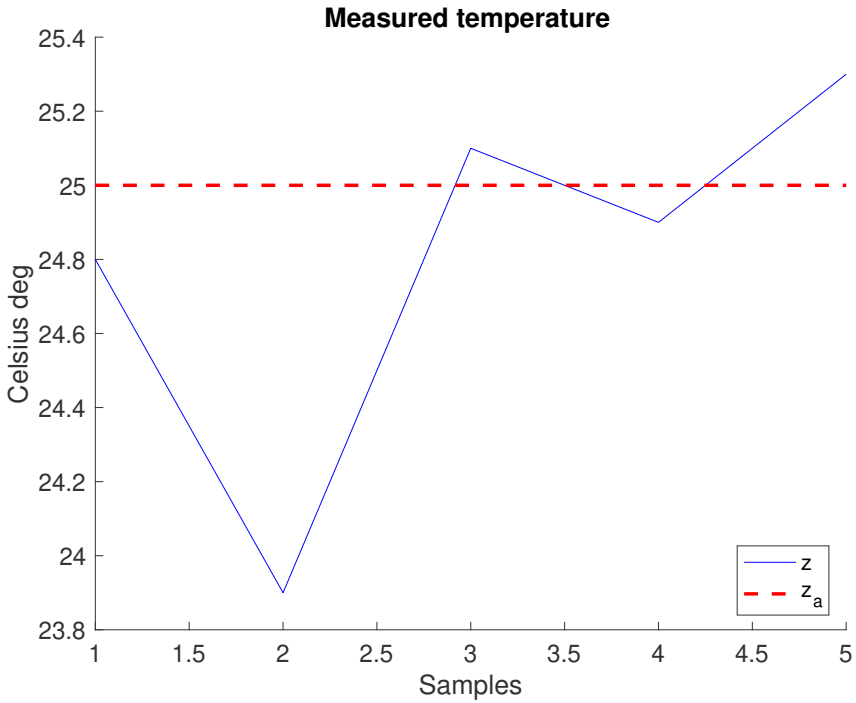
Suppose we have a sequence of n measures belonging to a certain set $\mathcal{Z} \subset \mathbb{R}$.
The n sampled values will be ordered along one axis, e.g., the time axis, and then *collected* according to a constant interval Δ_z.
For any $z \in \mathcal{Z}$, the function $N_\Delta(z)$ returns the number of samples belonging to the interval Δ_z to which z belongs.
As a consequence it is possible to define the *histogram* as the following function:

$$P_\Delta(z) = \frac{1}{\Delta_z} \frac{N_\Delta(z)}{n}.$$

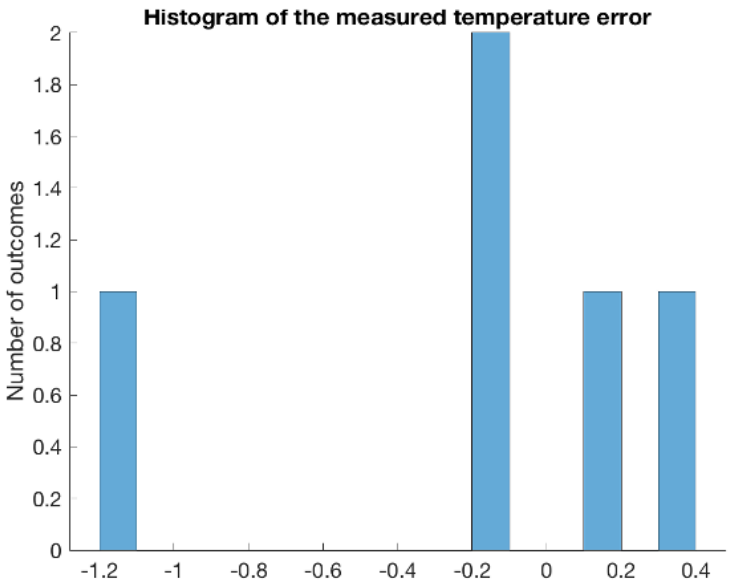
Notice how $\Delta_z P_\Delta(z)$ is *numerically equivalent* to the probability of having z in a certain interval Δ_z.

Random effects

Statistical description: Histograms



(a)

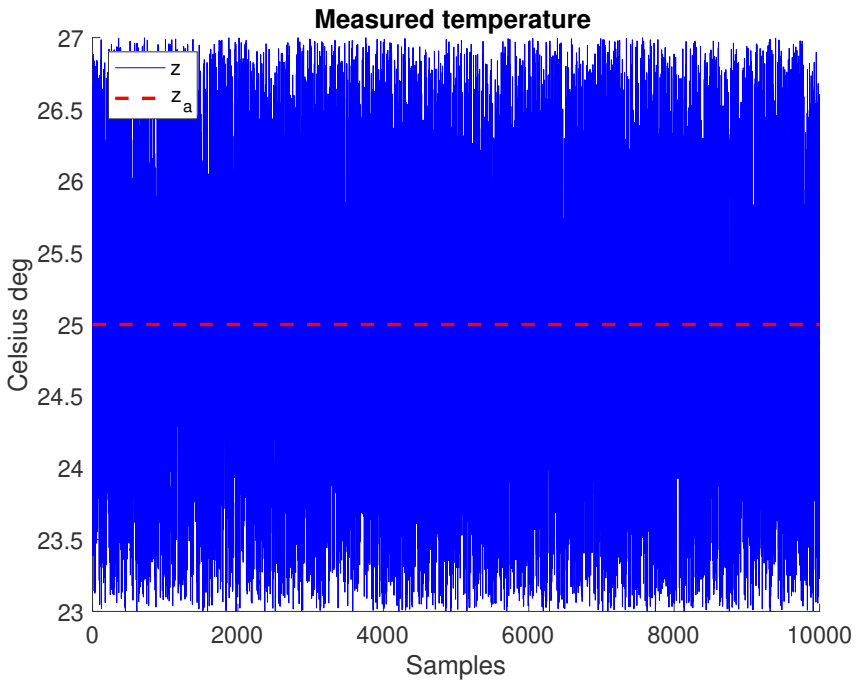


(b)

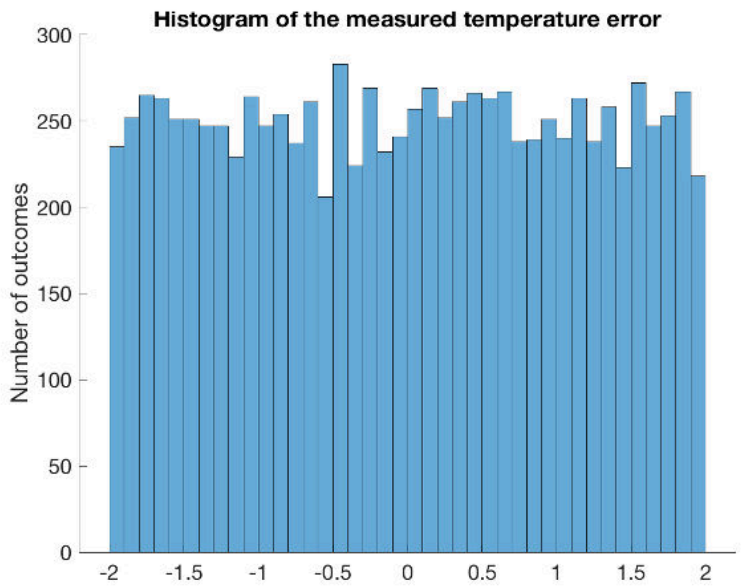
Figure: Examples of the temperatures: $z_a = 25°$ Celsius is the actual temperature, while the thermometer measures a sequence of $Z = [24.8°, 23.9°, 25.1°, 24.9°, 25.3°]$. (a) Time evolution, (b) Histogram.

Random effects

Statistical description: Histograms



(a)

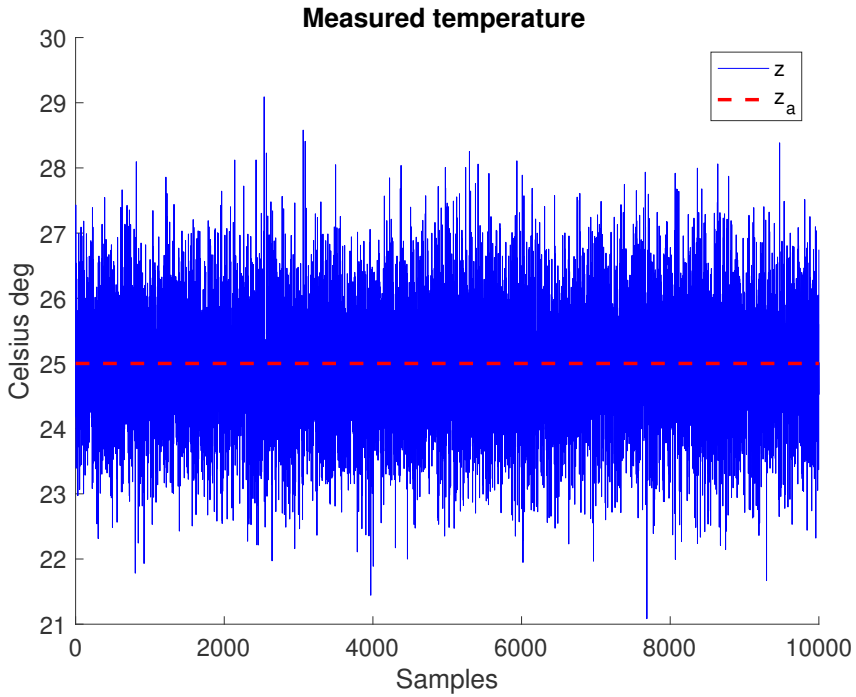


(b)

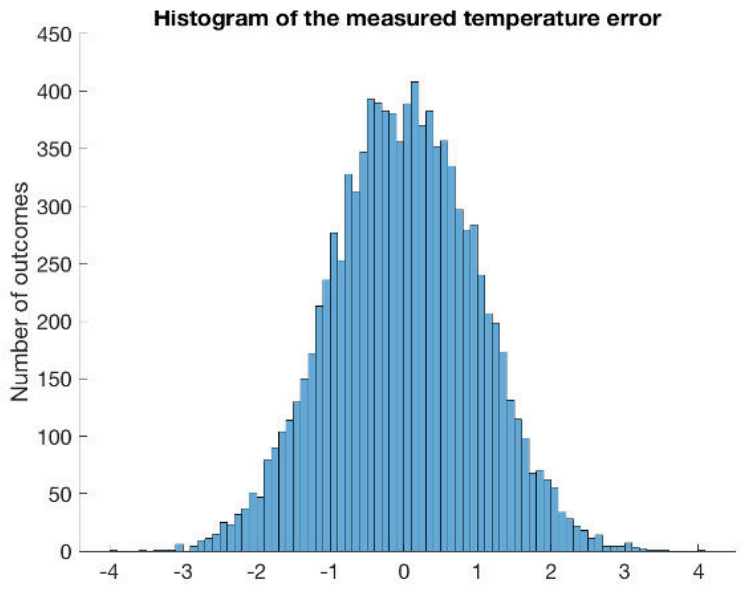
Figure: Examples of the temperatures: $z_a = 25°$ Celsius is the actual temperature, while the thermometer measures a sequence of 10^4 samples $z \in [23°, 27°]$. *Confidence level* 100%. (a) Time evolution, (b) Histogram.

Random effects

Statistical description: Histograms



(a)



(b)

Figure: Examples of the temperatures: $z_a = 25°$ Celsius is the actual temperature, while the thermometer measures a sequence of 10^4 samples $z \in [20°, 30°]$. *Confidence level* 99.9%. (a) Time evolution, (b) Histogram.

Random effects

Type A evaluation

In the previous three cases, we can identify two different situations:

- In the first example, there are *too few* observations to clearly identify the underlying statistical behaviour, hence the *Type A evaluation* cannot be carried out (more on this in the next slides);
- In the last two cases, the mean value approaches the actual value, i.e. no *systematic effects* are present. Hence, a *Type A evaluation* is feasible;
- Instead, if the *systematic effects* are still there, the analysis of the *Type A evaluation* cannot be carried out.

Random effects

Type A evaluation

Let us clarify one issue. As stated previously and according to the GUM, the *the best estimate of the measurand is the mean value* provided that the *systematic effects* are negligible.
Hence, recalling the definition of the *measurement error*

$$w_i = z_i - z_a,$$

we can state that the *systematic effects* o can be given by

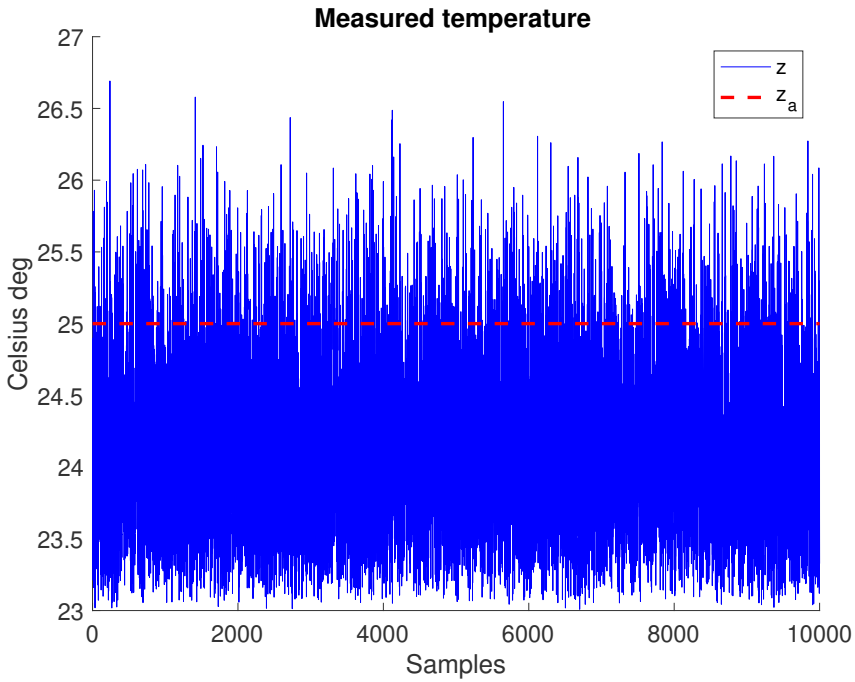
$$o = \frac{1}{n} \sum_{i=1}^{n} w_i.$$

Indeed, defining $\bar{z}_i = z_i - o$ as the *measurement with compensated and negligible systematic effects*, we have

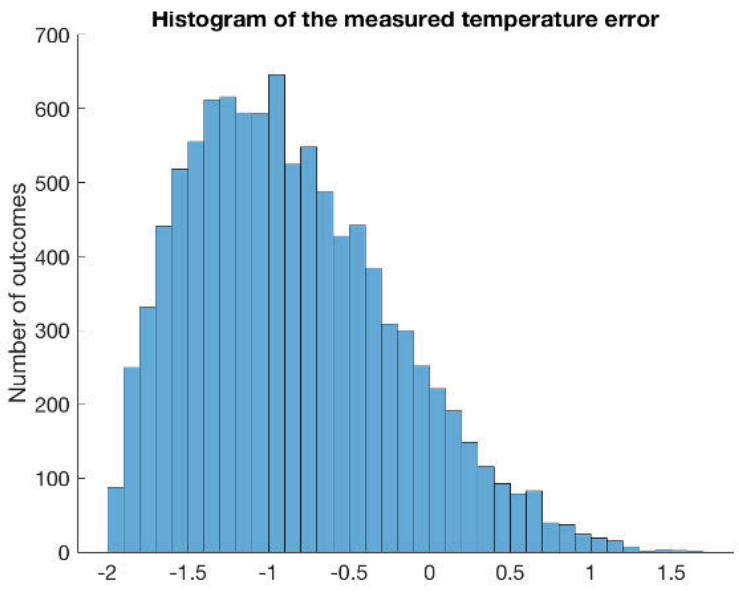
$$\frac{1}{n} \sum_{i=1}^{n} \bar{z}_i = z_a.$$

Random effects

Statistical description: Histograms



(a)



(b)

Figure: Examples of the temperatures: $z_a = 25°$ Celsius is the actual temperature, while the thermometer measures a sequence of 10^4 samples $z \in [23°, 27°]$. *Confidence level* 100%. (a) Time evolution, (b) Histogram. This sensor *has systematic effects*.

Random effects

Type A evaluation

To further analyse the choice of the *arithmetic mean*, i.e. the *mean value*, in the *Type A evaluation*, let us make an example.
Consider the *mean squared error* (MSE)

$$m^2 = \frac{\sum_{i=1}^{n}(z_i - z_a)^2}{n},$$

from which we can also compute the *root mean squared error* (RMSE), i.e. the square root of the MSE, also known as *mean error*.
The question now is: *what is a good estimate of z_a, i.e., $\hat{z}_a$ or simply $\hat{z}$, obtained using the measurements and how to compute it?*

Random effects

Type A evaluation

Notice that:

- Whenever an estimate $\hat{z}$ is available, an estimate of the quantity of interest $\hat{x}$ is available due to *sensor calibration*. Therefore, the more $\hat{z} \to z_a$, the more $\hat{x} \to x_a$;
- How to find $\hat{z}$ using z_i is the first example of an *estimator design*!

Random effects

Type A evaluation

An *estimate* $\hat{z}$ can be the value that *minimises* the the MSE $m_{\hat{z}}^2$ of the estimated value.
In this case, this is called the *Minimum MSE* (MMSE) estimator, i.e. the one that maximises the *precision* of the estimates.
In analytical terms:

$$\hat{z} = \arg\min_{\hat{z}} \sum_{i=1}^{n} (z_i - \hat{z})^2.$$

The solution to this problem is given exactly by the *arithmetic mean*, i.e.,

$$\hat{z} = \frac{\sum_{i=1}^{n} z_i}{n},$$

computed nullifying the first derivative with respect to the estimates $\hat{z}$.

Random effects

Type A evaluation

Definition

We will denote with

$$\hat{z}[z_1, \ldots, z_n] = \frac{\sum_{i=1}^{n} z_i}{n},$$

the *estimator*, i.e. the function or set of operations generating the estimates.

In this case the *estimator* is the *arithmetic mean*.

Definition

We will denote with $\hat{z}$ the *estimates* of z_a, i.e. the result of the *estimator* applied to the *available measurements*.

Random effects

Type A evaluation

Remark

This solution has been obtained finding a solution to a maximum likelihood (ML) criterium, since the values around $\hat{z}$ have the maximum probability to be measured.

Remark

This estimate is centred since if the set of the infinite measures is split into m subsets upon which the mean is computed, the mean of all the mean would be exactly the value z_a of interest.

We have proved that the *arithmetic mean* is the optimal estimator with respect to the MSE when the *systematic effects* can be *neglected* (recall the GUM definition).
We are now in a position to establish what is the *precision* of this estimator, i.e., the MSE attained by the *arithmetic mean*.

Random effects

Type A evaluation

The objective is to compute the MSE of $\hat{z}$, which we will denote with $m^2_{\hat{z}}$. First, let us notice that $\hat{z}$ depends on a set of *measurement values* z_i that are affected by *uncertainties*, i.e.

$$\hat{z} = \arg\min_{\hat{z}} \sum_{i=1}^{n} (z_i - \hat{z})^2.$$

The previous observation, and what follows, is valid if we consider $\hat{z}$ either as an *estimate*, i.e. the result of an *estimator*, or an *indirect measurement*, i.e. a measurement as a function or combination of other measurements. In general, $\hat{z} = f(z_1, z_2, \ldots, z_n)$.

Random effects

Type A evaluation

If $f(z_1, z_2, \ldots, z_n)$ is a function of the observed quantities $z_1, z_2, \ldots, z_n$, whose *actual values* are $z_{a_1}, z_{a_2}, \ldots, z_{a_n}$, it is possible to consider its Taylor expansion around those points, i.e.,

$$\begin{aligned} f(z_1, z_2, \ldots, z_n) = & f_a + \frac{\partial f}{\partial z_1}(z_1 - z_{a_1}) + \frac{\partial f}{\partial z_2}(z_2 - z_{a_2}) + \cdots + \\ & + \frac{1}{2}\frac{\partial^2 f}{\partial z_1^2}(z_1 - z_{a_1})^2 + \frac{1}{2}\frac{\partial^2 f}{\partial z_2^2}(z_2 - z_{a_2})^2 + \ldots \end{aligned}$$

where $f_a = f(z_{a_1}, z_{a_2}, \ldots, z_{a_n})$.
Assuming that the errors $z_i - z_{a_i}$ are small enough, we can limit the Taylor expansion up *to the first order*.

Random effects

Type A evaluation

Therefore

$$f(z_1, z_2, \ldots, z_n) - f_a = \sum_{i=1}^{n} \frac{\partial f}{\partial z_i}(z_i - z_{a_i})$$

and applying the squares

$$(f(z_1, z_2, \ldots, z_n) - f_a)^2 = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial z_i}\right)^2 (z_i - z_{a_i})^2 +$$

$$+ 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{\partial f}{\partial z_i} \frac{\partial f}{\partial z_j}(z_i - z_{a_i})(z_j - z_{a_j})$$

Assuming that the measurement processes generating z_i and z_j, $\forall i \neq j$, are *independent*, we can neglect all the mixed terms.

Random effects

Type A evaluation

We can now compute the MSE of $f(\cdot)$ using the MSE of each measurement, i.e.

$$m_f^2 = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial z_i} \right)^2 m_{z_i}^2$$

where $m_{z_i}^2$ is the MSE of the estimate related to the i–th measure.
In the particular case of the mean function

$$f(z_1, z_2, \ldots, z_n) = \hat{z} = \frac{\sum_{i=1}^{n} z_i}{n} \Rightarrow \frac{\partial f}{\partial z_1} = \frac{\partial f}{\partial z_2} = \cdots = \frac{1}{n}.$$

Moreover, we have the same MSE for all the measures, i.e., $m_{z_1}^2 = m_{z_2}^2 = \cdots = m^2$.

Random effects

Type A evaluation

Therefore,

$$m_f^2 = m_{\hat{z}}^2 = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial z_1} \right)^2 m_{z_i}^2 = \sum_{i=1}^{n} \frac{m^2}{n^2} = \frac{m^2}{n},$$

where we have to remember that m^2 is the MSE of the *measurement results*.

Remark

$\hat{z}$, that is the estimate of z_a, is a random variable.

Remark

The more measurement values we have, the less is the MSE!

Random effects

Type A evaluation

Recall that the MSE of all the measures m^2 is given by

$$m^2 = \frac{\sum_{i=1}^{n}(z_i - z_a)^2}{n},$$

which *involves the knowledge of* z_a that is *not available*.
This is the reason why the concept of *measurement error*, i.e. $w_i = z_i - z_a$, cannot be used *when the observations are considered*, because it is *philosophically incorrect*: the *actual value* of a quantity is an abstract representation of reality!

Random effects

Type A evaluation

To solve this problem, we first notice that $m_{\hat{z}}^2 = \frac{m^2}{n}$ can be computed by definition as $m_{\hat{z}}^2 = \frac{1}{n}\sum_{i=1}^{n}(\hat{z} - z_a)^2 = (\hat{z} - z_a)^2$, hence

$$nm^2 = \sum_{i=1}^{n}(z_i - z_a)^2 = \sum_{i=1}^{n}(z_i - \hat{z} + \hat{z} - z_a)^2 = \sum_{i=1}^{n}(z_i - \hat{z} + m_{\hat{z}})^2$$
$$= \sum_{i=1}^{n}(z_i - \hat{z})^2 + m_{\hat{z}}^2 + 2m_{\hat{z}}(z_i - \hat{z}).$$

Since $\sum_{i=1}^{n}(z_i - \hat{z}) = 0$ by definition, one has

$$nm^2 = \sum_{i=1}^{n}(z_i - \hat{z})^2 + \sum_{i=1}^{n} m_{\hat{z}}^2 = \sum_{i=1}^{n}(z_i - \hat{z})^2 + nm_{\hat{z}}^2 = \sum_{i=1}^{n}(z_i - \hat{z})^2 + m^2.$$

Random effects

Type A evaluation

Finally

$$m^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \hat{z})^2.$$

Remark

Notice that the MSE of the measurement values can be computed without the knowledge of the actual value.

Remark

This equation explains why, with Type A evaluation, the arithmetic mean is considered as the central value to compare all the measurements.

Random effects

Type A evaluation

Let us go back to the *histogram* definition:

$$P_\Delta(z) = \frac{1}{\Delta_z} \frac{N_\Delta(z)}{n}.$$

and recall that $\Delta_z P_\Delta(z)$ is *numerically equivalent* to the probability of having z in a certain interval Δ_z.

Random effects

Type A evaluation

Defining the infinite set $\mathcal{D} = \cup_i \Delta_{z_i} = \mathbb{R}$, we get that $\Delta_{z_i} P_\Delta(z)$ is always nonnegative and $\sum_i \Delta_{z_i} P_\Delta(z) = 1$, i.e., the *overall area covered by the histogram* sums up to 1.
In practice, if $n \to +\infty$, it is possible to let $\Delta_z \to 0$ and hence the *discrete function* $P_\Delta(z)$ turns to be a *continuos function* named *probability density function* (pdf) $p(z)$.
The *distribution function*, homologous of $\sum_i \Delta_{z_i} P_\Delta(z)$, is instead the integral of the pdf and, hence, expresses the probability, i.e.,

$$P(\bar{z}) = \mathsf{Pr}\left[z < \bar{z}\right] = \int_{-\infty}^{\bar{z}} p(z)dz,$$

or, in general,

$$P(\bar{z}'') - P(\bar{z}') = \mathsf{Pr}\left[\bar{z}' < z < \bar{z}''\right] = \int_{\bar{z}'}^{\bar{z}''} p(z)dz.$$

Random effects

Type A evaluation

Assuming that *the actual value exists*, we can consider $w = z - z_a$.
Under the following assumptions:

- The *measurement error* is *symmetric*, i.e., $p(w)$ is symmetric;
- *Small errors* are more frequent than *large errors*;
- The *greater* is the number of small errors, the *smaller* is the number of large errors;

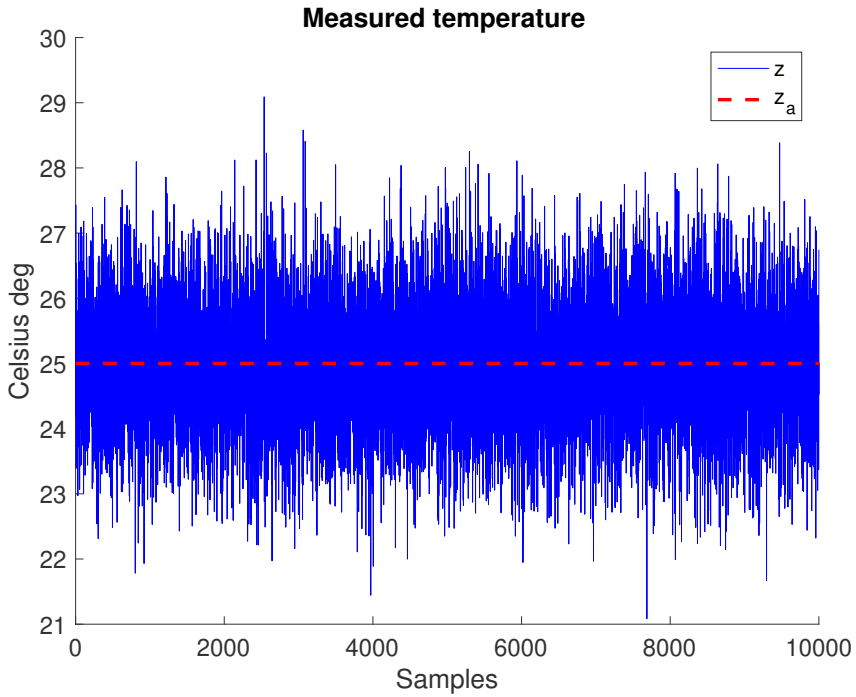
The pdf turns to be a *Gaussian* (or *Normal*) pdf, i.e.,

$$p(w) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{w^2}{2\sigma^2}}.$$

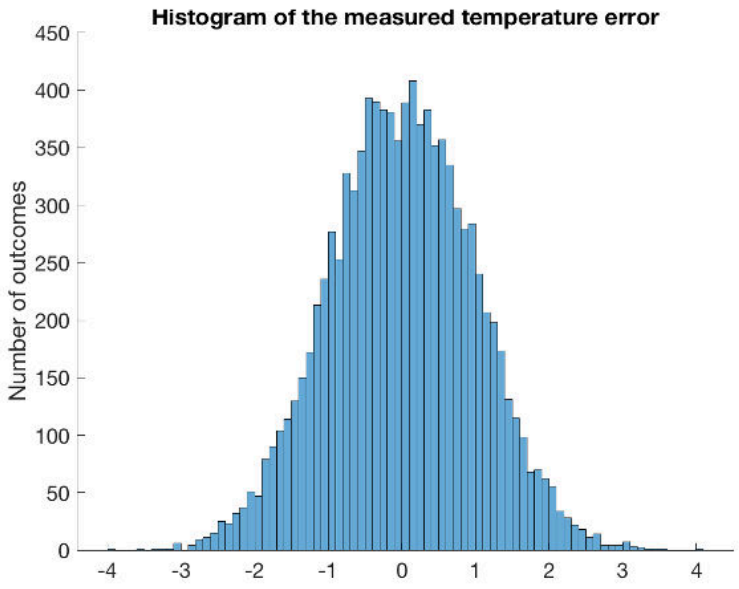
σ is the *standard deviation* and it is a function of the *precision* of the system, i.e. *standard uncertainty*.

Random effects

Statistical description: Histograms



(a)



(b)

Figure: Examples of the temperatures: $z_a = 25°$ Celsius is the actual temperature, while the thermometer measures a sequence of 10^4 samples $z \in [20°, 30°]$. (a) Time evolution, (b) Histogram.

Random effects

Type A evaluation

Again, it is evident that the *smaller* is the standard deviation, the *more precise* will be the measurement value.
In particular, the square of the standard deviation, i.e., the *variance*, is given by

$$\sigma^2 = \lim_{n \to +\infty} \frac{\sum_{i=1}^{n} (z_i - z_a)^2}{n}.$$

Notice that σ^2 can be known *up to a certain accuracy* since n is finite.

Random effects

Type A evaluation

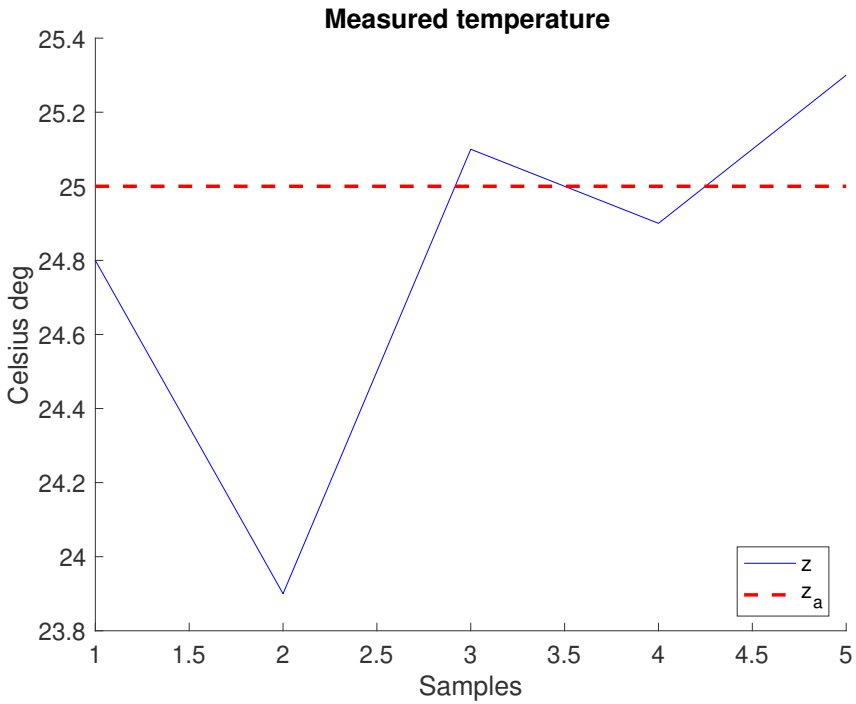
The MSE

$$m^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \hat{z})^2$$

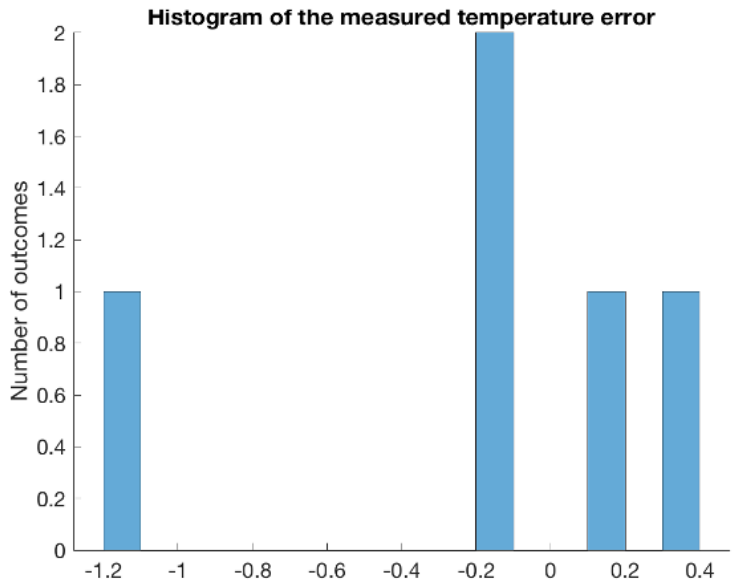
is a *statistically consistent* estimate of the *variance* σ^2 when the number of measurements n is finite and the *mean value* is null. In other words, for $n \to +\infty$, the *probability that m^2 differs from σ^2 tends towards zero*. Indeed, for $n \to +\infty$, $\hat{z} \to z_a$ (recall that $m_{\hat{z}}^2 = \frac{m^2}{n}$).

Random effects

Type A evaluation



(a)



(b)

Figure: The fact that $n \to +\infty$ exemplifies why for this example of the temperature we cannot carry out the *Type A Evaluation*: there is an *insufficient* number of samples.

Random effects

Type B evaluation

Sometimes it happens that the *systematic errors* cannot be recognised or compensated.
Moreover, it may happens that *it is not possible* to get access to observations.
In those cases, we can resort to *Type B evaluations*:

Definition (GUM)

Type B evaluation method of uncertainty by means *other than* the *statistical analysis* of a series of repeated *observations*.

Random effects

Type B evaluation

According to the GUM, the possible sources of information are:

- Previous measurement data;
- Experience or previous knowledge;
- Manufacturer's specifications;
- Data provided in calibration and other certificates;
- Uncertainties assigned to reference data taken from handbooks.

Random effects

Type B evaluation

Moreover, assumptions on a) the pdf of the *random effects* and b) the *coverage probability* are needed.
With those information, the *standard uncertainty* can be computed.
For example, assume that a thermometer measure $T_m = 24.845°$ Celsius and that the manufacturer reports a range of possible values within 0.04% of the reading. What is the *standard uncertainty*?
We first have that $\tau = 24.845 \cdot 4 \cdot 10^{-4} = 9.9 \cdot 10^{-3}$. Then, $T_m \pm \tau$ encompasses all the possible values with *coverage probability* 100%. Then, if no assumption can be made on the pdf, the maximum entropy assumption is taken (i.e., *uniform pdf*), finally having $u(T_m) = \frac{\tau}{\sqrt{3}} = 5.7 \cdot 10^{-3}$.

Random effects

Type A and Type B evaluation

When an estimate of both types of uncertainty is available, say $u_a(z)$ and $u_b(z)$, the combined uncertainty is given by:

$$u(z) = \sqrt{u_a(z)^2 + u_b(z)^2}.$$

Random effects

Uncertainty propagation law

In the previous example of the MSE, we noticed that when one *measurement value* is a nonlinear combination of other measurements $z = f(x_1, \ldots, x_n)$ (i.e. *indirect measurement*), $u(z)$ can be determined using the Taylor expansion (e.g. the velocity inferred from position measurements).
However, the *coverage probability* is difficult to be computed since the pdf of z can be *hardly computed*.
This is a crucial problem also for *estimators*, as we will see in the next classes.
In those cases, *Monte Carlo simulations* should be used, unless the *Central Limit theorem* can be applied.

Random effects

Modelling

Let us talk about the *modelling* of the measurement process.
By computing n times the measure, we obtain a series of *random values*, generated by the time-varying nature of any measurement process, i.e. the *random effects*.
This is due by the limited *resolution*, *precision* and *accuracy* of the measurements, which are affected by a *random measurement noise*, e.g. ε.
For example, $z_a = 25°$ Celsius is the actual temperature, while the thermometer measures a sequence of $z = [24.8°, 23.9°, 25.1°, 24.9°, 25.3°]$.
In other words

$$z = z_a + \varepsilon = 25° + [-0.2°, -1.1°, 0.1°, -0.1°, 0.3°],$$

where ε is the sequence of *measurement noises*.
To the sequence of *measurement noises* ε it is possible to associate a *statistical description*.

Random effects

Overall view

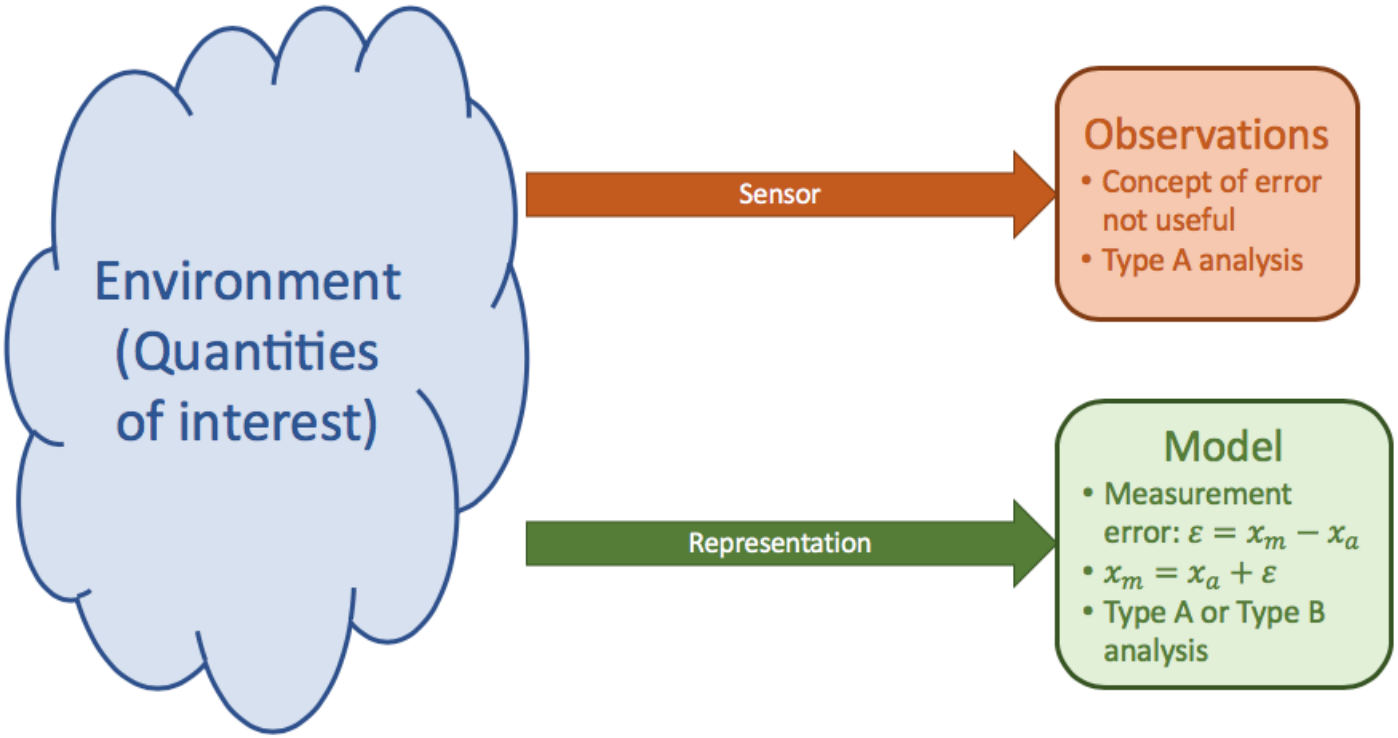


Figure: Two possible views of the environment using *observations* or using a suitable *model*.

Outline

1. Data Analysis
2. Measurement processes
   - The sensor
   - Random effects
3. Estimation Algorithms
4. Take home message

Random effects

Type A evaluation

In practice, *accuracy* is related to the *closeness of the mean value of repeated measurements* to the actual value, while *precision* is related to the *dispersion of repeated measurements* with respect to the actual value. Roughly speaking:

- *Accuracy* can be improved by calibration;
- *Precision* asks for *estimation algorithms* on a *calibrated* sensor;
- *Trueness* defines the quality of the measurements *on average*.

Fundaments of estimation

Let us assume that we want to *estimate* (that is to *increase our knowledge*) about a certain quantity α.
A *measurement process* can be formally defined as

$$z = g(\alpha, w, t),$$

where $g(\cdot)$ can represent either a *direct* measurement process, e.g. the thermometer measuring the temperature, or an *indirect* measurement, e.g. the velocity inferred from position measurements.
The data $z \in \mathbb{R}^m$ are the *measurements*.
The vector $w \in \mathbb{R}^m$ is the *measurement noise*.
t expresses the *time variability* of the measurement process.

Fundaments of estimation

Given the *measurement process* $g(\cdot)$ we usually define a *model* for it as well as for the quantities α.
The model of the quantities α is usually represented by a vector $x \in \mathbb{R}^n$.
An example is the representation by means of position and velocity.
Notice that the model is crucial, since there might be models that *better fits the specific system at hand*, e.g. change of coordinates or Euler angles.

Remark

We need always to think carefully to the quantities to be measured and to the their representation.

Fundaments of estimation

For the *measurement model*, we usually adopt a *static map*

$$z(k) = h(x(k)),$$

where the index k represents the k-th measurement (e.g., the sample taken at time $t = k\Delta_t$, $k \in \mathbb{N}$, with *sampling time* Δ_t).
For example, assume that you are able to measure the *distance* (i.e. the quantity α) of an object in the plane with coordinates (x, y) from a wireless anchor of position (x_a, y_a) using the *Time of Flight*.
In such a case the forward map is given by

$$z(k) = \sqrt{(x - x_a)^2 + (y - y_a)^2} = h(x, y, x_a, y_a).$$

Fundaments of estimation

How does the *static map model* $h(x)$ relates to the *measurement process* $g(\alpha)$?
Whenever we have a *model* we have a *modelling error*!
The problem of determining the *modelling error* is a complex problem per se.
In the ideal case (no measurement error), the *modelling error* can be represented as

$$e_h = h(x(\alpha)) - g(\alpha).$$

Usually, the measurement error is *larger* than the modelling error, and this is what we will assume in the following.
The model should be simple enough and in the *easiest case* is *linear*, i.e.

$$z(k) = Hx(k).$$

If this is not the case, usually a *Taylor approximation* is employed.

Fundaments of estimation

Usually in the linear model $H \in \mathbb{R}^{m \times n}$. The objective of the *estimator* is to retrieve $x \in \mathbb{R}^n$ from $z \in \mathbb{R}^m$.

- If $n < m$ the problem is *over-determined* and the problem is usually solvable in the linear and nonlinear cases;
- If $n = m$ the problem is more involved since it may happens that it does not have solutions for the linear and the nonlinear cases;
- If $n > m$ the problem is not solvable unless *prior knowledge* is available.

Fundaments of estimation

The problem of finding an *estimation algorithm* to determine the value of a certain (scalar or vector) unknown parameter x depends on the *nature* of x.
If the parameter x is *time-invariant*, then *static estimation algorithms* (such as the WLS) are adopted, otherwise *dynamic estimation algorithm* (such as the Kalman Filter).

Fundaments of estimation

In more strict terms, the problem of estimation of x given the set of measures

$$z(j) = h(j, x, w(j)), \quad j = 1, \ldots, k,$$

where $w(j)$ is a sequence of measurement noises (or *disturbances*) amounts to find the *estimates*

$$\hat{x}(k) = \hat{x}[k, Z^k],$$

being Z^k the set of all measures up to time k, i.e.,

$$Z^k = \{z(j)\}_{j=1,\ldots,k}.$$

Recall that the $\hat{x}(k)$ are the *estimates* of x, the function $\hat{x}[k, Z^k]$ is instead the *estimator*.

Fundaments of estimation

Suppose

$$z(j) = x + w(j), \quad j = 1, \ldots, k,$$

where $w(j)$ is the *measurement noise*.
Some possible estimators are:

- $\hat{x}(k) = \frac{1}{k} \sum_{j=1}^{k} z(j)$;
- $\hat{x}(k) = \frac{\max(Z^k) + \min(Z^k)}{2}$;
- $\hat{x}(k) = z(1)$.

Each of the previous estimators works, but is there a systematic way to *design an estimator*?
Using just one value it is impossible to judge: by chance we can derive the incorrect answer! A statistical analysis is needed (recall the definition of *measurement result*).

Fundaments of estimation

The *measurement error* w is the error affecting the measures at time k

$$z(k) = x(k) + w(k),$$

where $w(k)$ is a *random variable*.
The *estimation error* at time k is instead given by

$$\tilde{x}(k) = x(k) - \hat{x}(k),$$

where $\hat{x}(k)$ is a function of the measurements $z(k)$, $\forall k$.

Fundaments of estimation

To design an *estimator* we resort to the following description:

- The role of an estimator is to retrieve a *correct estimate* $\hat{x}(k)$, possibly the *best estimate*, i.e. the one with the smallest *estimation error* $\tilde{x}$;
- The information available are: a) the *measurements* $z(k)$, which are *noisy*; b) an idea about the *system output model*, i.e. the function $h(\cdot)$; c) maybe, some *additional knowledge*, e.g. the quantity is always positive;
- *All* the available knowledge *must be exploited*, so the estimator needs to incorporate it properly;
- We can treat or not the quantity to estimate as a *random variable*.

Outline

1. Data Analysis
2. Measurement processes
   - The sensor
   - Random effects
3. Estimation Algorithms
4. Take home message

Basis of Estimation Theory

Any *measurement result* is a *collection of data* that is affected by *uncertainties*.
There are *systematic* (aka *measurement biases*) and *random* (aka *measurement fluctuations*) effects acting on measurement.
According to the GUM, there is the possibility to carry out *Type A* (i.e. using observations) or *Type B* (i.e. any other means different from observations) evaluations.
To retrieve the actual value of a measured quantity, an *estimator* should be designed.
To tackle the problems related to estimation, we need some tools to properly manage *random variables*.