Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

"As far as the laws of mathematics refer to reality, they are not certain,

Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

"As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality. "
(A. Einstein)

Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

"As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality. "
(A. Einstein)

"It is scientific only to say what is more likely and what is less likely"
(R. Feynman)

Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

"As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality. "
(A. Einstein)

"It is scientific only to say what is more likely and what is less likely"
(R. Feynman)

"Probability is good sense reduced to a calculus"
(S. Laplace)

Introducing the logic of uncertainty

- No collection of formulae.

Introducing the logic of uncertainty

- No collection of formulae.
- No collection of tests "with Russian names".

Introducing the logic of uncertainty

- No collection of formulae.
- No collection of tests "with Russian names".
- Try to build up a consistent theory
  that can be used for a broad range of applications.
  - Avoid unneeded '*principles*'

Introducing the logic of uncertainty

- No collection of formulae.
- No collection of tests "with Russian names".
- Try to build up a consistent theory
  that can be used for a broad range of applications.
  - Avoid unneeded '*principles*'...
    whose results will *possibly* be reobtained
    as *approximations under well stated conditions*.

Please be patient

Please be patient

"...today I'll learn to read,

Please be patient

"...today I'll learn to read,
tomorrow to write,

Please be patient

"...today I'll learn to read,
tomorrow to write,
and the day after tomorrow
I'll do arithmetic."

Please be patient

"...today I'll learn to read,
tomorrow to write,
and the day after tomorrow
I'll do arithmetic."

Please be patient

"...today I'll learn to read,
tomorrow to write,
and the day after tomorrow
I'll do arithmetic."

["Then, clever as I am,
I can earn a lot of money."]

Please be patient

“. . . today I’ll learn to read,
tomorrow to write,
and the day after tomorrow
I’ll do arithmetic.”

["Then, clever as I am,
I can earn a lot of money."]

- No rush to get formulae

Please be patient

"...today I'll learn to read,
tomorrow to write,
and the day after tomorrow
I'll do arithmetic."

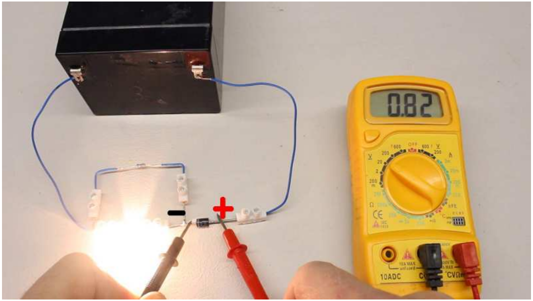
["Then, clever as I am,
I can earn a lot of money."]

- No rush to get formulae

→ If you understand the basic reasoning
you can derive many formulae by yourself'!

What is measurement?

What is measurement?

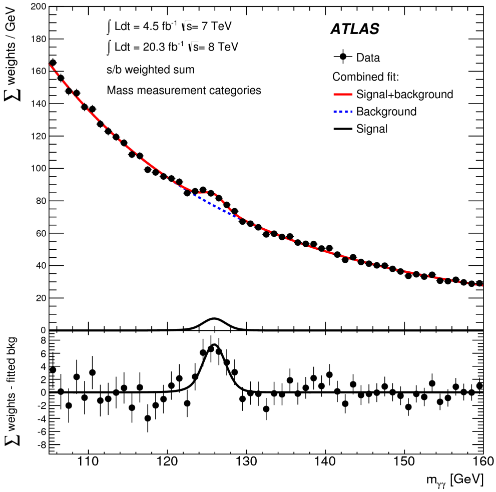


What is measurement?

What is measurement?

What is measurement?

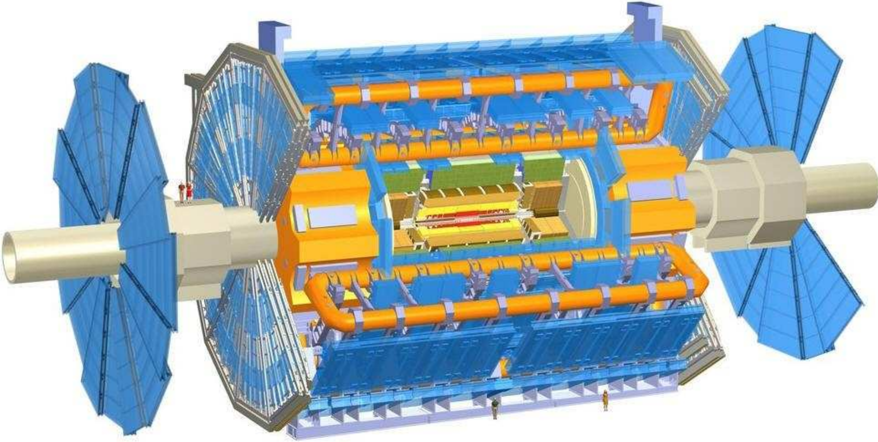
Higgs $\to \gamma\gamma$ (2012)



Two-photon *invariant mass*

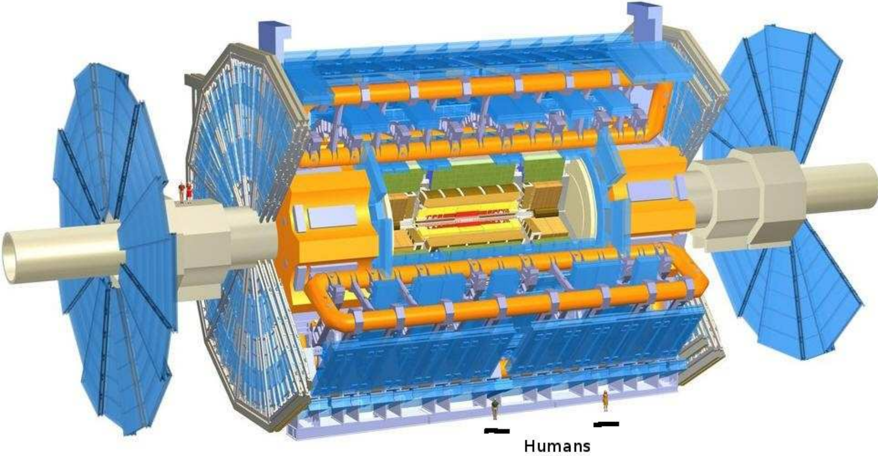
What is measurement?

ATLAS Experiment at LHC (CERN, Geneva)



What is measurement?

ATLAS Experiment at LHC [length: 46 m; ⌀ 25 m]

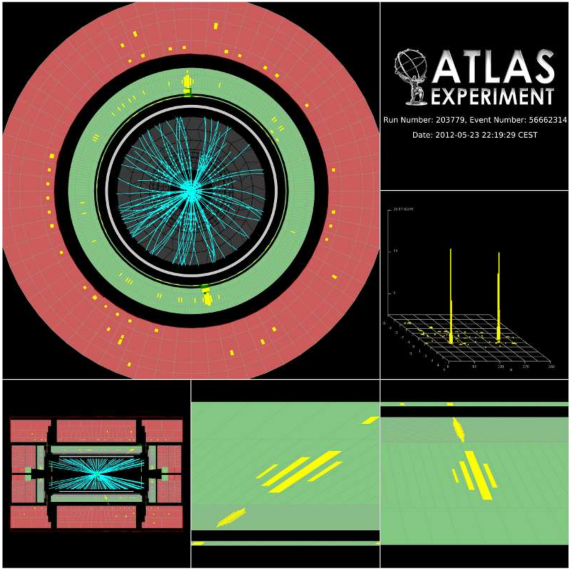


≈ 3000 km cables

≈ 7000 tonnes

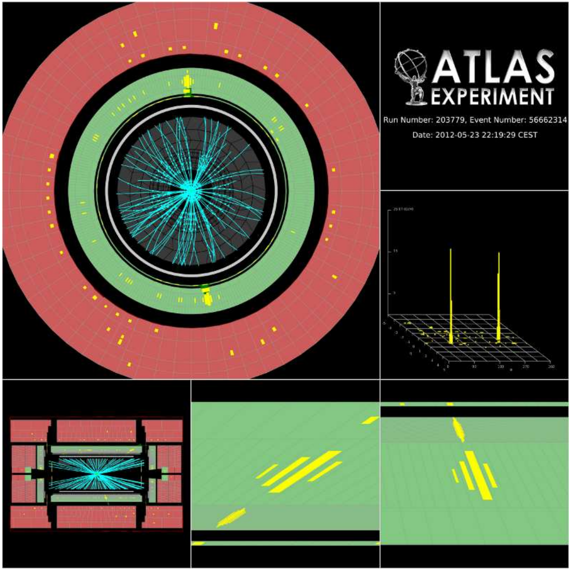
≈ 100 millions electronic channels

What is measurement?



Two flashes of ‘light’ (2 γ’s) in a ‘noisy’ environment.

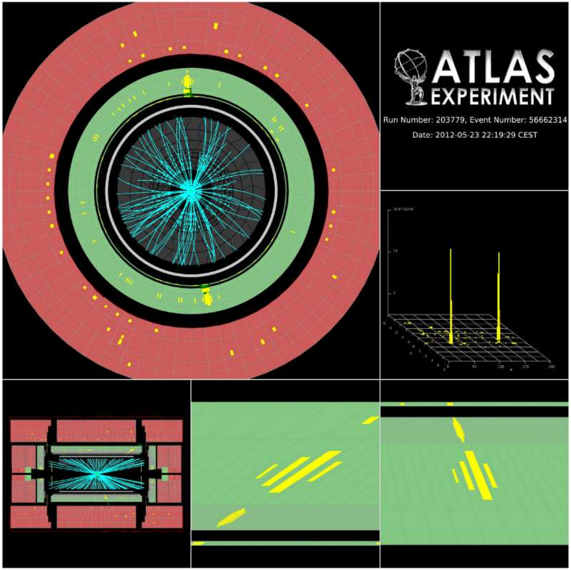
What is measurement?



Two flashes of 'light' (2 γ's) in a 'noisy' environment.

Higgs $\rightarrow \gamma\gamma$?

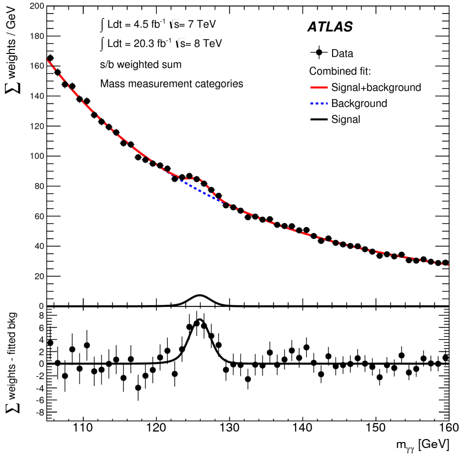
What is measurement?



Two flashes of ‘light’ (2 γ’s) in a ‘noisy’ environment.
Higgs $\rightarrow \gamma\gamma$? Probably not. . .

What is measurement?

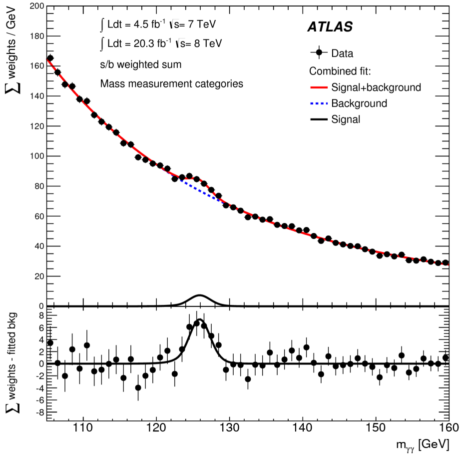
Higgs $\to \gamma\gamma$



$\Rightarrow$ { Mass value
Production rate

What is measurement?

Higgs $\to \gamma\gamma$

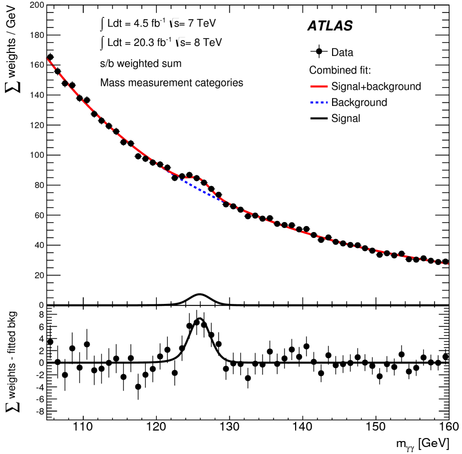


$\Rightarrow \begin{cases} \text{Mass value} \\ \text{Production rate} \end{cases}$

(with uncertainties)

What is measurement?

Higgs $\to \gamma\gamma$



$\Rightarrow$ { Mass value, Production rate (with uncertainties)

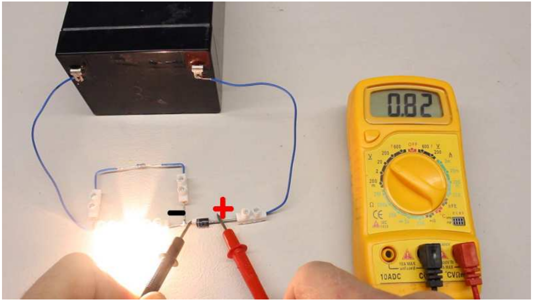
Quite indirect measurements of something we do not "see"!

Can we “see” physics quantities?

But, can we <u>see</u> our mass?

Can we "see" physics quantities?

... or a voltage?



Can we "see" physics quantities?

. . . or our blood pressure?

Can we “see” physics quantities?

Certainly not!

Can we “see” physics quantities?

Certainly not!

. . . although for some quantities we can have

a ‘vivid impression’ (in the David Hume’s sense)

Measuring a mass on a scale

Equilibrium:

$$\begin{aligned} {\color{red}m}g - k{\color{blue}\Delta x} &= 0 \\ {\color{blue}\Delta x} &\rightarrow {\color{blue}\theta} \rightarrow \text{scale reading} \end{aligned}$$

(with 'g' *gravitational acceleration;* 'k' spring constant.)

Measuring a mass on a scale

Equilibrium:

$$mg - k\Delta x = 0$$
$$\Delta x \rightarrow \theta \rightarrow \text{scale reading}$$

(with 'g' *gravitational acceleration;* 'k' spring constant.)

From the reading to the value of the mass:

$$\text{scale reading} \xrightarrow[\text{given } g,\ k,\ \text{"etc."}\ldots]{} m$$

Measuring a mass on a balance

$$\text{scale reading} \xrightarrow[\text{given } g,\ k,\ \text{“etc.”}\ldots]{} m$$

Dependence on 'g': $$g \stackrel{?}{=} \frac{GM_{\oplus}}{R_{\oplus}^2}$$

Measuring a mass on a balance

$$\text{scale reading} \xrightarrow[\text{given } g,\ k,\ \text{“etc.”}\ldots]{} m$$

Dependence on 'g': $$g \stackrel{?}{=} \frac{GM_{\oplus}}{R_{\oplus}^2}$$

- Position is usually <u>not</u> at "$R_{\oplus}$" from the Earth center;
- Earth not spherical...
- ...not even ellipsoidal...
- ...and not even homogeneous.
- Moreover we have to consider centrifugal effects
- ...and even the effect from the Moon

Measuring a mass on a balance

$$\text{scale reading} \xrightarrow[\text{given } g,\ k,\ \text{“etc.”}\ldots]{} m$$

Dependence on 'g': $$g \stackrel{?}{=} \frac{GM_{\oplus}}{R_{\oplus}^2}$$

- Position is usually <u>not</u> at "$R_{\oplus}$" from the Earth center;
- Earth not spherical. . .
- . . . not even ellipsoidal. . .
- . . . and not even homogeneous.
- Moreover we have to consider centrifugal effects
- . . . and even the effect from the Moon

Certainly not to watch our weight 

Measuring a mass on a balance

$$\text{scale reading} \xrightarrow[\text{given } g,\ k,\ \text{“etc.”}\ldots]{} m$$

Dependence on 'g': $$g \stackrel{?}{=} \frac{GM_{\oplus}}{R_{\oplus}^2}$$

- Position is usually <u>not</u> at "$R_{\oplus}$" from the Earth center;
- Earth not spherical...
- ...not even ellipsoidal...
- ...and not even homogeneous.
- Moreover we have to consider centrifugal effects
- ...and even the effect from the Moon

Certainly not to watch our weight

But think about it!

Measuring a mass on a balance

scale reading $\xrightarrow[\text{given } g,\ k,\ \text{“etc.”}\dots]{}$ m

Dependence on 'k':

- temperature
- non linearity
- ...

Measuring a mass on a balance

$$\text{scale reading} \xrightarrow[\text{given } g,\ k,\ \text{“etc.”}\ldots]{} m$$

Dependence on 'k':

- temperature
- non linearity
- ...

Δx $\to \theta \to$ scale reading:

- left to your imagination...

Measuring a mass on a balance

$$\text{scale reading} \xrightarrow[\textit{given } g,\ k,\ \text{“etc.”}\ldots]{} m$$

Dependence on 'k':

- temperature
- non linearity
- ...

Δx $\rightarrow \theta \rightarrow$ scale reading:

- left to your imagination...

+ randomic effects:

- stopping position of damped oscillation;
- variability of all quantities of influence (in the ISO-GUM sense);
- reading of analog scale.

Measuring a mass on a balance

$$\text{scale reading} \xrightarrow[\text{given } g,\ k,\ \text{“etc.”}\ldots]{} m$$

Dependence on 'k':

- temperature
- non linearity
- ...

$\mathbf{\Delta x} \rightarrow \theta \rightarrow$ scale reading:

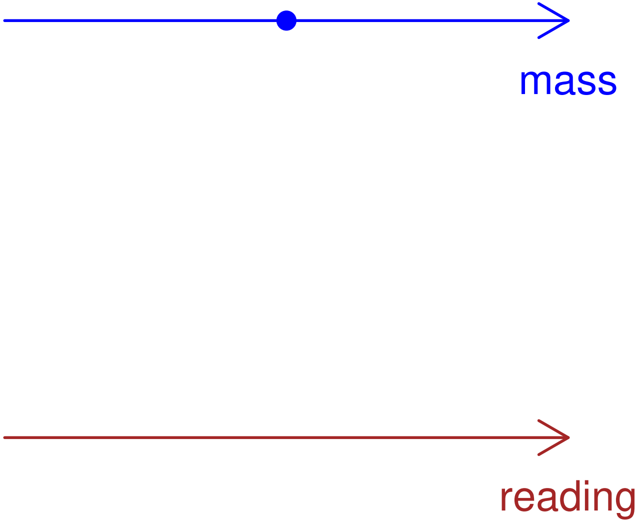
- left to your imagination...

$+$ **randomic effects:**

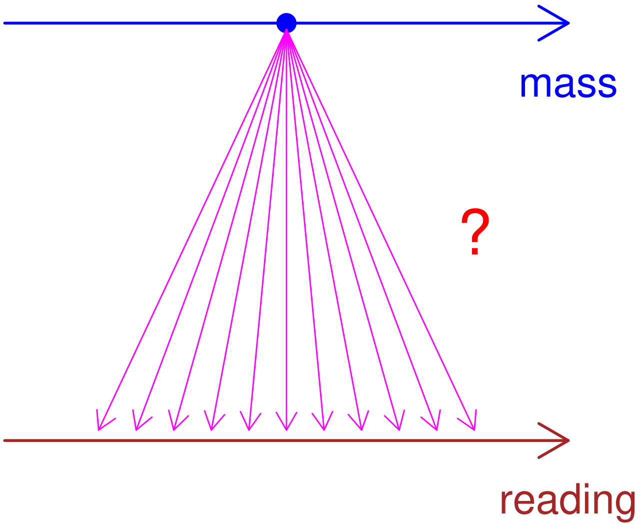
- stopping position of damped oscillation;
- variability of all quantities of influence (in the ISO-GUM sense);
- reading of analog scale.

$\Rightarrow \; m$??

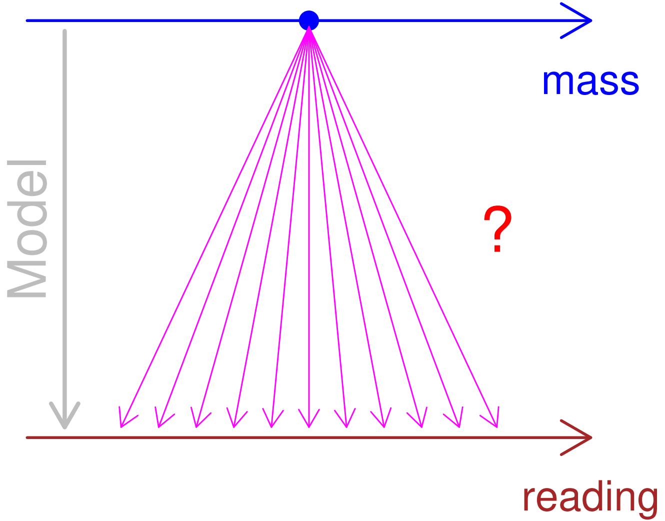
Mass ⟶ Reading



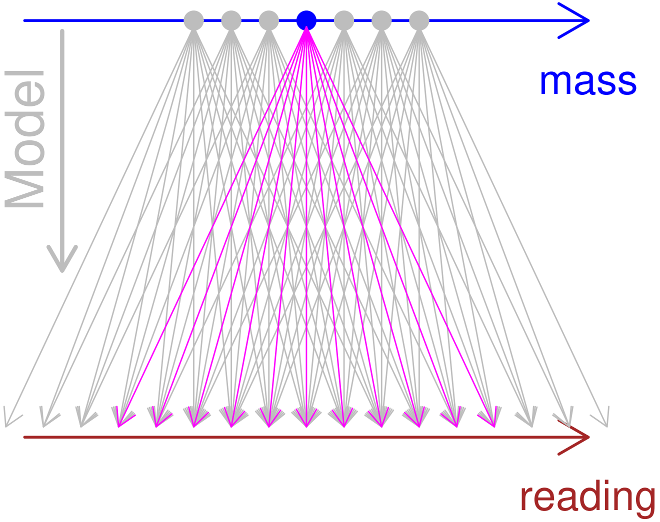
Mass $\longrightarrow$ Reading



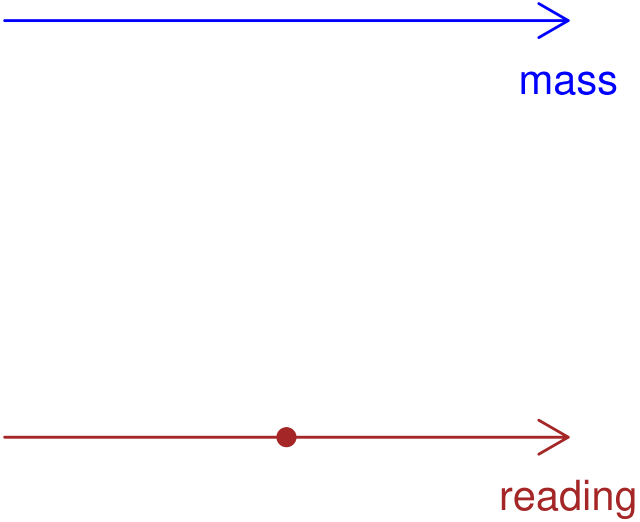
Mass ⟶ Reading



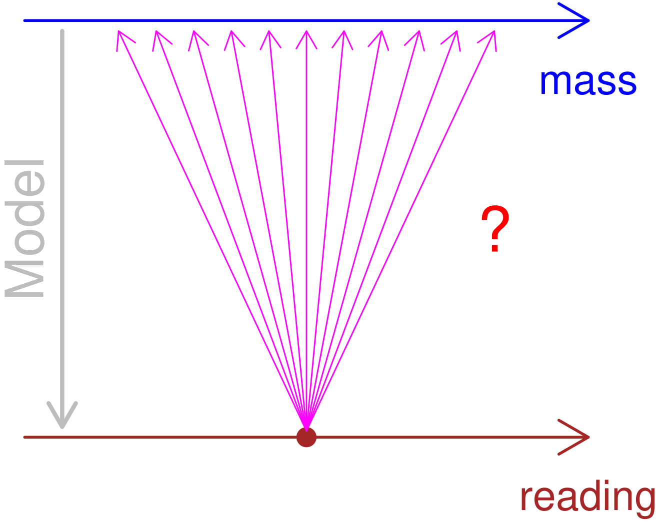
Mass ⟶ reading



Reading $\longrightarrow$ 'true' mass



Reading ⟶ 'true' mass



Sources of uncertainties (from ISO GUM)

1. *incomplete definition of the measurand*

Sources of uncertainties (from ISO GUM)

1 *incomplete definition of the measurand*

→ g

→where?

→inertial effects subtracted?

Sources of uncertainties (from ISO GUM)

1. *incomplete definition of the measurand*
   - → g
     - →where?
     - →inertial effects subtracted?
2. *imperfect realization of the definition of the measurand*

Sources of uncertainties (from ISO GUM)

1. *incomplete definition of the measurand*
   → g
   →where?
   →inertial effects subtracted?
2. *imperfect realization of the definition of the measurand*
   → scattering on neutron
   →how to realize a neutron target?

Sources of uncertainties (from ISO GUM)

1. *incomplete definition of the measurand*
   - → *g*
     →where?
     →inertial effects subtracted?
2. *imperfect realization of the definition of the measurand*
   - → scattering on neutron
     →how to realize a neutron target?
3. *non-representative sampling — the sample measured may not represent the measurand;*

Sources of uncertainties (from ISO GUM)

1. *incomplete definition of the measurand*
   - → g
     - →where?
     - →inertial effects subtracted?
2. *imperfect realization of the definition of the measurand*
   - → scattering on neutron
     - →how to realize a neutron target?
3. *non-representative sampling — the sample measured may not represent the measurand;*
4. *inadequate knowledge of the effects of environmental conditions on the measurement, or imperfect measurement of environmental conditions;*

Sources of uncertainties (from ISO GUM)

1. *incomplete definition of the measurand*
   - → g
     - →where?
     - →inertial effects subtracted?
2. *imperfect realization of the definition of the measurand*
   - → scattering on neutron
     - →how to realize a neutron target?
3. *non-representative sampling — the sample measured may not represent the measurand;*
4. *inadequate knowledge of the effects of environmental conditions on the measurement, or imperfect measurement of environmental conditions;*
5. *personal bias in reading analogue instruments;*

Sources of uncertainties (from ISO GUM)

6 *finite instrument resolution or discrimination threshold;*

Sources of uncertainties (from ISO GUM)

6. *finite instrument resolution or discrimination threshold;*
7. *inexact values of measurement standards and reference materials;*

Sources of uncertainties (from ISO GUM)

6. *finite instrument resolution or discrimination threshold;*
7. *inexact values of measurement standards and reference materials;*
8. *inexact values of constants and other parameters obtained from external sources and used in the data-reduction algorithm;*

Sources of uncertainties (from ISO GUM)

6. *finite instrument resolution or discrimination threshold;*
7. *inexact values of measurement standards and reference materials;*
8. *inexact values of constants and other parameters obtained from external sources and used in the data-reduction algorithm;*
9. *approximations and assumptions incorporated in the measurement method and procedure;*

Sources of uncertainties (from ISO GUM)

6. *finite instrument resolution or discrimination threshold;*
7. *inexact values of measurement standards and reference materials;*
8. *inexact values of constants and other parameters obtained from external sources and used in the data-reduction algorithm;*
9. *approximations and assumptions incorporated in the measurement method and procedure;*
10. *variations in repeated observations of the measurand under apparently identical conditions.*

Sources of uncertainties (from ISO GUM)

6. *finite instrument resolution or discrimination threshold;*
7. *inexact values of measurement standards and reference materials;*
8. *inexact values of constants and other parameters obtained from external sources and used in the data-reduction algorithm;*
9. *approximations and assumptions incorporated in the measurement method and procedure;*
10. *variations in repeated observations of the measurand under apparently identical conditions.*

Note

- Sources not necessarily independent
- In particular, sources 1-9 may contribute to 10 (e.g. not-monitored electric fluctuations)

Half scale spacing? (mezza divisione, “mezza tacca”. . .)

Half scale spacing? (mezza divisione, “mezza tacca”. . .)

- “Line scales mainly have a scale numbering with regular spacing and are mostly intended for a continuous indication of measured values”. (DIN 1319, part 2, 6.1.1)

Half scale spacing? (mezza divisione, “mezza tacca”. . .)

- “Line scales mainly have a scale numbering with regular spacing and are mostly intended for a continuous indication of measured values”. (DIN 1319, part 2, 6.1.1)
- “Unduly small scale spacing (less than approx. 0.7 mm) should be avoided, since such scales are tiring to read and in particular the estimating of tenths is impossible so that the observation is rendered less certain.” (DIN 1319, part 2, 6.3)

Half scale spacing? (mezza divisione, “mezza tacca”. . .)

- “Line scales mainly have a scale numbering with regular spacing and are mostly intended for a continuous indication of measured values”. (DIN 1319, part 2, 6.1.1)
- “Unduly small scale spacing (less than approx. 0.7 mm) should be avoided, since such scales are tiring to read and in particular the estimating of tenths is impossible so that the observation is rendered less certain.” (DIN 1319, part 2, 6.3)
- “In some areas of metrology the term “resolution” is used. This is understood to mean the small change in the value of the measurand which is necessary to produce a perceptible (often specified) small change in the response (in the case of measuring instruments with scale indication, for example, 1/5 of the scale interval)”. (DIN 1319, part 2, 9)

Half scale spacing? (mezza divisione, “mezza tacca”...)

- “In un formato per osservatore umano l’incertezza di lettura dipende dalle caratteristiche costruttive della scala e dell’indice, dalle modalità d’osservazione, dal rumore eventuale e dall’abilità dell’osservatore. Per esempio se si ammette che un osservatore di normale abilità, leggendo lo strumento nella posizione appropriata, possa stimare 1/5 di divisione, si indicherà come incertezza di lettura ± 0.1 divisioni.” (UNI 4546, 5.5)

Half scale spacing? (mezza divisione, “mezza tacca”...)

- “In un formato per osservatore umano l’incertezza di lettura dipende dalle caratteristiche costruttive della scala e dell’indice, dalle modalità d’osservazione, dal rumore eventuale e dall’abilità dell’osservatore. Per esempio se si ammette che un osservatore di normale abilità, leggendo lo strumento nella posizione appropriata, possa stimare 1/5 di divisione, si indicherà come incertezza di lettura ± 0.1 divisioni.” (UNI 4546, 5.5)
- **Remark 1:** Essere praticamente sicuri che il valore sia entro il 1/5 di divisione, vuol dire che, se ci si sforza al interpolare al meglio, ci si aspetta una deviazione standard dell’errore di lettura di circa $0.2/\sqrt{12}$ divisioni, compatibile al valore di $\approx$ **0.07 che si osserva sperimentalmente**.

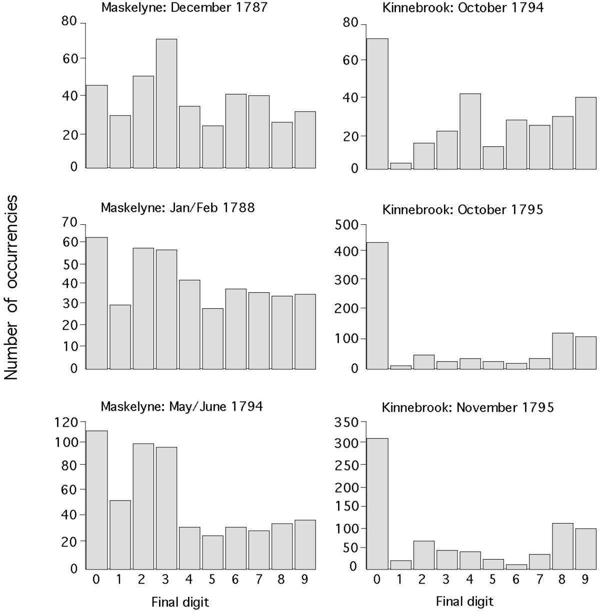
Half scale spacing? (mezza divisione, "mezza tacca"...)

- "In un formato per osservatore umano l'incertezza di lettura dipende dalle caratteristiche costruttive della scala e dell'indice, dalle modalità d'osservazione, dal rumore eventuale e dall'abilità dell'osservatore. Per esempio se si ammette che un osservatore di normale abilità, leggendo lo strumento nella posizione appropriata, possa stimare 1/5 di divisione, si indicherà come incertezza di lettura ± 0.1 divisioni." (UNI 4546, 5.5)
- **Remark 1:** Essere praticamente sicuri che il valore sia entro il 1/5 di divisione, vuol dire che, se ci si sforza al interpolare al meglio, ci si aspetta una deviazione standard dell'errore di lettura di circa $0.2/\sqrt{12}$ divisioni, compatibile al valore di $\approx$ **0.07 che si osserva sperimentalmente**.
- **Remark 2:** No absolute rule: it depends of the persons and on the working conditions!

Half scale spacing? (mezza divisione, “mezza tacca”. . .)

- “In un formato per osservatore umano l’incertezza di lettura dipende dalle caratteristiche costruttive della scala e dell’indice, dalle modalità d’osservazione, dal rumore eventuale e dall’abilità dell’osservatore. Per esempio se si ammette che un osservatore di normale abilità, leggendo lo strumento nella posizione appropriata, possa stimare 1/5 di divisione, si indicherà come incertezza di lettura ± 0.1 divisioni.” (UNI 4546, 5.5)
- **Remark 1:** Essere praticamente sicuri che il valore sia entro il 1/5 di divisione, vuol dire che, se ci si sforza al interpolare al meglio, ci si aspetta una deviazione standard dell’errore di lettura di circa $0.2/\sqrt{12}$ divisioni, compatibile al valore di $\approx$ **0.07 che si osserva sperimentalmente**.
- **Remark 2:** No absolute rule: it depends of the persons and on the working conditions!
- Just try!
  No dogmatism!

An historical case (Nevil Maskelyne)



Half of the scale distance?

- It is not the reading error

Half of the scale distance?

- It is not the reading error
- It is not even (always) the systematic error

Half of the scale distance?

- It is not the reading error
- It is not even (always) the systematic error

More on "Errori e incertezze di misura – rassegna critica e proposte per l'insegnamento"
⇒ http://www.roma1.infn.it/~dagos/teaching.html

An app to check you ability

$\Rightarrow$ ErroriLettura.apk on the course web site

ISO dictionary

ISO: International Organization for Standardization

GUM: Guides to the expression of Uncertainty in Measurement

ISO dictionary

ISO: International Organization for Standardization

GUM: Guides to the expression of Uncertainty in Measurement

Measurand: *"particular quantity subject to measurement."*

ISO dictionary

ISO: International Organization for Standardization

GUM: Guides to the expression of Uncertainty in Measurement

Measurand: *"particular quantity subject to measurement."*

True value: *"a value compatible with the definition of a given particular quantity."*

ISO dictionary

ISO: International Organization for Standardization

GUM: Guides to the expression of Uncertainty in Measurement

Measurand: *"particular quantity subject to measurement."*

True value: *"a value compatible with the definition of a given particular quantity."*

Result of a measurement: *"value attributed to a measurand, obtained by measurement."*

ISO dictionary

ISO: International Organization for Standardization

GUM: Guides to the expression of Uncertainty in Measurement

Measurand: *"particular quantity subject to measurement."*

True value: *"a value compatible with the definition of a given particular quantity."*

Result of a measurement: *"value attributed to a measurand, obtained by measurement."*

Uncertainty: *"a parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand."*

ISO dictionary

ISO: International Organization for Standardization

GUM: Guides to the expression of Uncertainty in Measurement

Measurand: *"particular quantity subject to measurement."*

True value: *"a value compatible with the definition of a given particular quantity."*

Result of a measurement: *"value attributed to a measurand, obtained by measurement."*

Uncertainty: *"a parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand."*

Error: *"the result of a measurement minus a true value of the measurand."*

ISO dictionary

ISO: International Organization for Standardization

GUM: Guides to the expression of Uncertainty in Measurement

Measurand: *"particular quantity subject to measurement."*

True value: *"a value compatible with the definition of a given particular quantity."*

Result of a measurement: *"value attributed to a measurand, obtained by measurement."*

Uncertainty: *"a parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand."*

Error: *"the result of a measurement minus a true value of the measurand."*

Error and *uncertainty* are not synonyms!

Type A and Type B uncertainties

Type A evaluation (of uncertainty): *"method of evaluation of uncertainty by the statistical analysis of series of observations."*

Type A and Type B uncertainties

Type A evaluation (of uncertainty): *"method of evaluation of uncertainty by the statistical analysis of series of observations."*

Type B evaluation (of uncertainty): *"method of evaluation of uncertainty by means other than the statistical analysis of series of observations."*

Type A and Type B uncertainties

Type A evaluation (of uncertainty): *"method of evaluation of uncertainty by the statistical analysis of series of observations."*

Type B evaluation (of uncertainty): *"method of evaluation of uncertainty by means other than the statistical analysis of series of observations."*

$\Rightarrow$ *"... the standard uncertainty $u(x_i)$ is evaluated by scientific judgement based on all of the available information on the possible variability of X_i.*

Type A and Type B uncertainties

Type A evaluation (of uncertainty): *"method of evaluation of uncertainty by the statistical analysis of series of observations."*

Type B evaluation (of uncertainty): *"method of evaluation of uncertainty by means other than the statistical analysis of series of observations."*

⇒ *"... the standard uncertainty $u(x_i)$ is evaluated by scientific judgement based on all of the available information on the possible variability of X_i. The pool of information may include*

- *previous measurement data;*

Type A and Type B uncertainties

Type A evaluation (of uncertainty): *"method of evaluation of uncertainty by the statistical analysis of series of observations."*

Type B evaluation (of uncertainty): *"method of evaluation of uncertainty by means other than the statistical analysis of series of observations."*

$\Rightarrow$ *"...the standard uncertainty $u(x_i)$ is evaluated by scientific judgement based on all of the available information on the possible variability of X_i.*
The pool of information may include

- *previous measurement data;*
- *experience with or general knowledge of the behaviour and properties of relevant materials and instruments;*

Type A and Type B uncertainties

Type A evaluation (of uncertainty): *"method of evaluation of uncertainty by the statistical analysis of series of observations."*

Type B evaluation (of uncertainty): *"method of evaluation of uncertainty by means other than the statistical analysis of series of observations."*

$\Rightarrow$ *"... the standard uncertainty $u(x_i)$ is evaluated by scientific judgement based on all of the available information on the possible variability of X_i.*
The pool of information may include

- *previous measurement data;*
- *experience with or general knowledge of the behaviour and properties of relevant materials and instruments;*
- *manufacturer's specifications;*

Type A and Type B uncertainties

Type A evaluation (of uncertainty): *"method of evaluation of uncertainty by the statistical analysis of series of observations."*

Type B evaluation (of uncertainty): *"method of evaluation of uncertainty by means other than the statistical analysis of series of observations."*

$\Rightarrow$ *"... the standard uncertainty $u(x_i)$ is evaluated by scientific judgement based on all of the available information on the possible variability of X_i.*
The pool of information may include

- *previous measurement data;*
- *experience with or general knowledge of the behaviour and properties of relevant materials and instruments;*
- *manufacturer's specifications;*
- *data provided in calibration and other certificates;*

Type A and Type B uncertainties

Type A evaluation (of uncertainty): *"method of evaluation of uncertainty by the statistical analysis of series of observations."*

Type B evaluation (of uncertainty): *"method of evaluation of uncertainty by means other than the statistical analysis of series of observations."*

$\Rightarrow$ *"... the standard uncertainty $u(x_i)$ is evaluated by scientific judgement based on all of the available information on the possible variability of X_i.*
The pool of information may include

- *previous measurement data;*
- *experience with or general knowledge of the behaviour and properties of relevant materials and instruments;*
- *manufacturer's specifications;*
- *data provided in calibration and other certificates;*
- *uncertainties assigned to reference data taken from handbooks."*

Usual handling of measurement uncertainties

Uncertainties due to statistical errors are currently treated using the frequentistic concept of 'confidence interval'

Usual handling of measurement uncertainties

Uncertainties due to statistical errors are currently treated using the frequentistic concept of 'confidence interval', although

- there are well-know cases — of great relevance in frontier physics — in which the approach is not applicable (e.g. small number of observed events, or measurement close to the edge of the physical region);

Usual handling of measurement uncertainties

Uncertainties due to statistical errors are currently treated using the frequentistic concept of 'confidence interval', although

- there are well-know cases — of great relevance in frontier physics — in which the approach is not applicable (e.g. small number of observed events, or measurement close to the edge of the physical region);
- the procedure is rather unnatural, and in fact the interpretation of the results is unconsciously (intuitively) probabilistic (see later).

Usual handling of measurement uncertainties

Uncertainties due to statistical errors are currently treated using the frequentistic concept of 'confidence interval', although

- there are well-know cases — of great relevance in frontier physics — in which the approach is not applicable (e.g. small number of observed events, or measurement close to the edge of the physical region);
- the procedure is rather unnatural, and in fact the interpretation of the results is unconsciously (intuitively) probabilistic (see later).

  $\rightarrow$ Intuitive reasoning $\Longleftrightarrow$ statistics education

Usual handling of measurement uncertainties

Uncertainties due to statistical errors are currently treated using the frequentistic concept of 'confidence interval', although

- there are well-know cases — of great relevance in frontier physics — in which the approach is not applicable (e.g. small number of observed events, or measurement close to the edge of the physical region);
- the procedure is rather unnatural, and in fact the interpretation of the results is unconsciously (intuitively) probabilistic (see later).
  $\rightarrow$ Intuitive reasoning $\Longleftrightarrow$ statistics education

These cases have not to be seen as "the exception that confirms the rule" [in physics exceptions falsify laws!]

Usual handling of measurement uncertainties

Uncertainties due to statistical errors are currently treated using the frequentistic concept of 'confidence interval', although

- there are well-know cases — of great relevance in frontier physics — in which the approach is not applicable (e.g. small number of observed events, or measurement close to the edge of the physical region);
- the procedure is rather unnatural, and in fact the interpretation of the results is unconsciously (intuitively) probabilistic (see later).

  $\rightarrow$ Intuitive reasoning $\Longleftrightarrow$ statistics education

These cases have not to be seen as "the exception that confirms the rule" [in physics exceptions falsify laws!], but as symptoms of something flawed in the reasoning, that could seriously effects also results that are not as self-evidently paradoxical as in these cases!

Usual handling of measurement uncertainties

There is no satisfactory theory or model to treat uncertainties due to systematic errors:

Usual handling of measurement uncertainties

There is no satisfactory theory or model to treat uncertainties due to systematic errors:

- *"my supervisor says ..."*

Usual handling of measurement uncertainties

There is no satisfactory theory or model to treat uncertainties due to systematic errors:

- *"my supervisor says ..."*
- *"add them linearly"*;

Usual handling of measurement uncertainties

There is no satisfactory theory or model to treat uncertainties due to systematic errors:

- *"my supervisor says . . ."*
- *"add them linearly"*;
- *"add them linearly if . . . , else add them quadratically"*;

Usual handling of measurement uncertainties

There is no satisfactory theory or model to treat uncertainties due to systematic errors:

- *"my supervisor says . . ."*
- *"add them linearly"*;
- *"add them linearly if . . . , else add them quadratically"*;
- *"don't add them at all"*.

Usual handling of measurement uncertainties

There is no satisfactory theory or model to treat uncertainties due to systematic errors:

- *"my supervisor says ..."*
- *"add them linearly"*;
- *"add them linearly if ..., else add them quadratically"*;
- *"don't add them at all"*.

The modern *fashion*: add them quadratically if they are considered to be independent, or build a covariance matrix of statistical and systematic contributions in the general case.

Usual handling of measurement uncertainties

There is no satisfactory theory or model to treat uncertainties due to systematic errors:

- *"my supervisor says . . ."*
- *"add them linearly"*;
- *"add them linearly if . . . , else add them quadratically"*;
- *"don't add them at all"*.

The modern *fashion*: add them quadratically if they are considered to be independent, or build a covariance matrix of statistical and systematic contributions in the general case.
In my opinion, simply the reluctance to combine linearly 10, 20 or more contributions to a global uncertainty, as the (out of fashion) 'theory' of maximum bounds would require.
→ Right in most cases!
→ Good sense of physicists ⟺ cultural background

A simple case

n independent measurements of the same quantity μ (with n large enough and no systematic effects, to avoid, for the moment, extra complications).

A simple case

n independent measurements of the same quantity μ (with n large enough and no systematic effects, to avoid, for the moment, extra complications).

Evaluate $\overline{x}$ and σ from the data

report result: $\rightarrow \mu = \overline{x} \pm \sigma/\sqrt{n}$

A simple case

n independent measurements of the same quantity μ (with n large enough and no systematic effects, to avoid, for the moment, extra complications).

Evaluate $\overline{x}$ and σ from the data

report result: $\rightarrow \mu = \overline{x} \pm \sigma/\sqrt{n}$

- what does it mean?

1 For the large majority of physicists
$P(\overline{x} - \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + \frac{\sigma}{\sqrt{n}}) = 68\%$

A simple case

n independent measurements of the same quantity μ (with n large enough and no systematic effects, to avoid, for the moment, extra complications).

Evaluate $\overline{x}$ and σ from the data

report result: $\rightarrow \mu = \overline{x} \pm \sigma/\sqrt{n}$

- what does it mean?

1 For the large majority of physicists $P(\overline{x} - \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + \frac{\sigma}{\sqrt{n}}) = 68\%$

2 And many explain (also to students!) that *"this means that, if I repeat the experiment a great number of times, then I will find that in roughly 68% of the cases the observed average will be in the interval* $\left[\overline{x} - \sigma/\sqrt{n},\ \overline{x} + \sigma/\sqrt{n}\right]$*."*

A simple case

n independent measurements of the same quantity μ (with n large enough and no systematic effects, to avoid, for the moment, extra complications).

Evaluate $\overline{x}$ and σ from the data

report result: $\rightarrow \mu = \overline{x} \pm \sigma/\sqrt{n}$

- what does it mean?

1. For the large majority of physicists $P(\overline{x} - \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + \frac{\sigma}{\sqrt{n}}) = 68\%$
2. And many explain (also to students!) that *"this means that, if I repeat the experiment a great number of times, then I will find that in roughly 68% of the cases the observed average will be in the interval* $\left[\overline{x} - \sigma/\sqrt{n},\ \overline{x} + \sigma/\sqrt{n}\right]$*."*
3. Statistics experts tell that the interval $\left[\overline{x} - \sigma/\sqrt{n},\ \overline{x} + \sigma/\sqrt{n}\right]$ covers the true μ in 68% of cases

A simple case

n independent measurements of the same quantity μ (with n large enough and no systematic effects, to avoid, for the moment, extra complications).

Evaluate $\overline{x}$ and σ from the data

report result: $\rightarrow \mu = \overline{x} \pm \sigma/\sqrt{n}$

- what does it mean?

1 For the large majority of physicists $P(\overline{x} - \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + \frac{\sigma}{\sqrt{n}}) = 68\%$

2 And many explain (also to students!) that *"this means that, if I repeat the experiment a great number of times, then I will find that in roughly 68% of the cases the observed average will be in the interval* $\left[\overline{x} - \sigma/\sqrt{n},\ \overline{x} + \sigma/\sqrt{n}\right]$*."*

3 Statistics experts tell that the interval $\left[\overline{x} - \sigma/\sqrt{n},\ \overline{x} + \sigma/\sqrt{n}\right]$ covers the true μ in 68% of cases

Objections?

Meaning of $\mu = \overline{x} \pm \sigma/\sqrt{n}$

1 $P(\overline{x} - \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + \frac{\sigma}{\sqrt{n}}) = 68\%$
OK to me, and perhaps no objections by many of you

Meaning of $\mu = \overline{x} \pm \sigma/\sqrt{n}$

1 $P(\overline{x} - \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + \frac{\sigma}{\sqrt{n}}) = 68\%$

OK to me, and perhaps no objections by many of you

- But it depends on what we mean by probability

Meaning of $\mu = \overline{x} \pm \sigma/\sqrt{n}$

1. $P(\overline{x} - \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{x} + \frac{\sigma}{\sqrt{n}}) = 68\%$

   OK to me, and perhaps no objections by many of you

   - But it depends on what we mean by probability
   - If probability is the "limit of the frequency", this statement is meaningless, because the 'frequency based' probability theory only speak about

$$P(\mu - \frac{\sigma}{\sqrt{n}} \le \overline{X} \le \mu + \frac{\sigma}{\sqrt{n}}) = 68\%\,,$$

(that is a probabilistic statement about $\overline{X}$: probabilistic statements about μ are not allowed by the theory).

Meaning of $\mu = \overline{x} \pm \sigma/\sqrt{n}$

2 *“if I repeat the experiment a great number of times, then I will find that in roughly 68% of the cases the observed average will be in the interval* $\left[\overline{x} - \sigma/\sqrt{n},\ \overline{x} + \sigma/\sqrt{n}\right]$*.”*

Meaning of $\mu = \overline{x} \pm \sigma/\sqrt{n}$

2 *"if I repeat the experiment a great number of times, then I will find that in roughly 68% of the cases the observed average will be in the interval* $\left[\overline{x} - \sigma/\sqrt{n},\ \overline{x} + \sigma/\sqrt{n}\right]$*."*

- Nothing wrong in principle (in my opinion)

Meaning of $\mu = \overline{x} \pm \sigma/\sqrt{n}$

2 *"if I repeat the experiment a great number of times, then I will find that in roughly 68% of the cases the observed average will be in the interval* $\left[\overline{x} - \sigma/\sqrt{n},\ \overline{x} + \sigma/\sqrt{n}\right]$*."*

- Nothing wrong in principle (in my opinion)
- but a $\sqrt{2}$ mistake in the width of the interval

Meaning of $\mu = \overline{x} \pm \sigma/\sqrt{n}$

2 *"if I repeat the experiment a great number of times, then I will find that in roughly 68% of the cases the observed average will be in the interval* $\left[\overline{x} - \sigma/\sqrt{n},\ \overline{x} + \sigma/\sqrt{n}\right]$*."*

- Nothing wrong in principle (in my opinion)
- but a $\sqrt{2}$ mistake in the width of the interval

$\rightarrow$ $P(\overline{x} - \sigma/\sqrt{n}\ \leq\ \overline{x}_f\ \leq\ \overline{x} + \sigma/\sqrt{n}) = 52\%$,
where $\overline{x}_f$ stands for future averages;

Meaning of $\mu = \overline{x} \pm \sigma/\sqrt{n}$

2 *"if I repeat the experiment a great number of times, then I will find that in roughly 68% of the cases the observed average will be in the interval* $\left[\overline{x} - \sigma/\sqrt{n},\ \overline{x} + \sigma/\sqrt{n}\right]$*."*

- Nothing wrong in principle (in my opinion)
- but a $\sqrt{2}$ mistake in the width of the interval

$\rightarrow$ $P(\overline{x} - \sigma/\sqrt{n} \ \leq\ \overline{x}_f \ \leq\ \overline{x} + \sigma/\sqrt{n}) = 52\%$,
where $\overline{x}_f$ stands for future averages;
or $P(\overline{x} - \sqrt{2}\,\sigma/\sqrt{n} \ \leq\ \overline{x}_f \ \leq\ \overline{x} + \sqrt{2}\,\sigma/\sqrt{n}) = 68\%$,
as we shall see later ($\rightarrow$ 'predictive distributions').

‘Confidence intervals’ and ‘frequentistic coverage’

3 Frequentistic coverage $\rightarrow$ “several problems”

‘Confidence intervals’ and ‘frequentistic coverage’

3 Frequentistic coverage $\rightarrow$ “several problems”

- ‘Trivial’ interpretation problem: $\rightarrow$ taken by most users as if it were a probability interval

‘Confidence intervals’ and ‘frequentistic coverage’

3 Frequentistic coverage $\rightarrow$ “several problems”

- ‘Trivial’ interpretation problem: $\rightarrow$ taken by most users as if it were a probability interval (not just semantic!)

‘Confidence intervals’ and ‘frequentistic coverage’

3 Frequentistic coverage $\rightarrow$ “several problems”

- ‘Trivial’ interpretation problem: $\rightarrow$ taken by most users as if it were a probability interval (not just semantic!)
- It fails in frontier cases

‘Confidence intervals’ and ‘frequentistic coverage’

3 Frequentistic coverage $\rightarrow$ “several problems”

- ‘Trivial’ interpretation problem: $\rightarrow$ taken by most users as if it were a probability interval (not just semantic!)
- It fails in frontier cases
  - 'technically' [see e.g. G. Zech, *Frequentistic and Bayesian confidence limits*, EPJdirect C12 (2002) 1]

‘Confidence intervals’ and ‘frequentistic coverage’

3 Frequentistic coverage $\rightarrow$ “several problems”

- ‘Trivial’ interpretation problem: $\rightarrow$ taken by most users as if it were a probability interval (not just semantic!)
- It fails in frontier cases
  - 'technically' [see e.g. G. Zech, *Frequentistic and Bayesian confidence limits*, EPJdirect C12 (2002) 1]
  - ‘in terms of performance’ $\rightarrow$ ‘very strange’ that no quantities show in ‘other side’ of a 95% C.L. bound !
- Not suited to express our confidence! Simply because it was not invented for that purpose!

‘Confidence intervals’ and ‘frequentistic coverage’

The **pretended** peculiar characteristic of frequentistic coverage is not to express confidence, but, when it works, to ‘ensure’ that, when applied a great number of times, in a defined percentage of the report the coverage statement is true.

‘Confidence intervals’ and ‘frequentistic coverage’

The **pretended** peculiar characteristic of frequentistic coverage is not to express confidence, but, when it works, to ‘ensure’ that, when applied a great number of times, in a defined percentage of the report the coverage statement is true.

> *“Carry out your experiment, calculate the confidence interval, and state that c belong to this interval.*

‘Confidence intervals’ and ‘frequentistic coverage’

The **pretended** peculiar characteristic of frequentistic coverage is not to express confidence, but, when it works, to ‘ensure’ that, when applied a great number of times, in a defined percentage of the report the coverage statement is true.

> *“Carry out your experiment, calculate the confidence interval, and state that c belong to this interval. If you are asked whether you ‘believe’ that c belongs to the confidence interval you must refuse to answer.*

‘Confidence intervals’ and ‘frequentistic coverage’

The **pretended** peculiar characteristic of frequentistic coverage is not to express confidence, but, when it works, to ‘ensure’ that, when applied a great number of times, in a defined percentage of the report the coverage statement is true.

> *“Carry out your experiment, calculate the confidence interval, and state that c belong to this interval. If you are asked whether you ‘believe’ that c belongs to the confidence interval you must refuse to answer. In the long run your assertions, if independent of each other, will be right in approximately a proportion α of cases.”*
> *(Neyman)*

‘Confidence intervals’ and ‘frequentistic coverage’

The **pretended** peculiar characteristic of frequentistic coverage is not to express confidence, but, when it works, to ‘ensure’ that, when applied a great number of times, in a defined percentage of the report the coverage statement is true.

> *“Carry out your experiment, calculate the confidence interval, and state that c belong to this interval. If you are asked whether you ‘believe’ that c belongs to the confidence interval you must refuse to answer. In the long run your assertions, if independent of each other, will be right in approximately a proportion α of cases.”*
> *(Neyman)*

> *“that technological and commercial apparatus which is known as an acceptance procedure”*
> *(Fisher, referring to Neyman’s statistical confidence method)*

‘Confidence intervals’ and ‘frequentistic coverage’

- Are we sure that our aim is to be right e.g. 68% of the times?

'Confidence intervals' and 'frequentistic coverage'

- Are we sure that our aim is to be right e.g. 68% of the times?
- For that we don't need to make an experiment!

  The ultimate 68.3% C.L. confidence interval calculator:

‘Confidence intervals’ and ‘frequentistic coverage’

- Are we sure that our aim is to be right e.g. 68% of the times?
- For that we don’t need to make an experiment!

  The ultimate 68.3% C.L. confidence interval calculator: a random number generator that gives
  - $[-10^{+9999}, +10^{+9999}]$ with 68.3% probability
  - $[1.00000001 \times 10^{-300}, 1.00000002 \times 10^{-300}]$ with 31.7% probability.

‘Confidence intervals’ and ‘frequentistic coverage’

- Are we sure that our aim is to be right e.g. 68% of the times?
- For that we don’t need to make an experiment!

  The ultimate 68.3% C.L. confidence interval calculator:
  a random number generator that gives

  - $[-10^{+9999}, +10^{+9999}]$ with 68.3% probability
  - $[1.00000001 \times 10^{-300}, 1.00000002 \times 10^{-300}]$ with 31.7% probability.

  If you do not like it, it might be you do not really care about ‘coverage’.

‘Confidence intervals’ and ‘frequentistic coverage’

- Are we sure that our aim is to be right e.g. 68% of the times?
- For that we don’t need to make an experiment!

  The ultimate 68.3% C.L. confidence interval calculator:
  a random number generator that gives
  - $[-10^{+9999}, +10^{+9999}]$ with 68.3% probability
  - $[1.00000001 \times 10^{-300}, 1.00000002 \times 10^{-300}]$ with 31.7% probability.

  If you do not like it, it might be you do not really care about ‘coverage’. You, as a physicist who care about your physical quantity, think in terms of 'confidence'

'Confidence intervals' and 'frequentistic coverage'

- Are we sure that our aim is to be right e.g. 68% of the times?
- For that we don't need to make an experiment!

  The ultimate 68.3% C.L. confidence interval calculator:
  a random number generator that gives
  - $[-10^{+9999}, +10^{+9999}]$ with 68.3% probability
  - $[1.00000001 \times 10^{-300}, 1.00000002 \times 10^{-300}]$ with 31.7% probability.

  If you do not like it, it might be you do not really care about 'coverage'. You, as a physicist who care about your physical quantity, think in terms of 'confidence':

  $\Rightarrow$ How much you are confident that the value of your quantity of interest is in a given interval.

'Confidence intervals' and 'frequentistic coverage'

- Are we sure that our aim is to be right e.g. 68% of the times?
- For that we don't need to make an experiment!

  The ultimate 68.3% C.L. confidence interval calculator:
  a random number generator that gives
  - $[-10^{+9999}, +10^{+9999}]$ with 68.3% probability
  - $[1.00000001 \times 10^{-300}, 1.00000002 \times 10^{-300}]$ with 31.7% probability.

  If you do not like it, it might be you do not really care about 'coverage'. You, as a physicist who care about your physical quantity, think in terms of 'confidence':

  $\Rightarrow$ How much you are confident that the value of your quantity of interest is in a given interval. We do not play a lottery!

Arbitrary probability inversions

How do we turn, just 'intuitively'

$$P(\mu - \frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + \frac{\sigma}{\sqrt{n}}) = 68\%$$

into

$$P(\overline{x} - \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + \frac{\sigma}{\sqrt{n}}) = 68\%\,?$$

Arbitrary probability inversions

How do we turn, just 'intuitively'

$$P(\mu - \frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + \frac{\sigma}{\sqrt{n}}) = 68\%$$

into

$$P(\overline{x} - \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + \frac{\sigma}{\sqrt{n}}) = 68\%\,?$$

We can paraphrase as

"the dog and the hunter"

The dog and the hunter

We know that a dog has a 50% probability of being 100 m from the hunter

⇒ if we observe the dog, what can we say about the hunter?

The dog and the hunter

We know that a dog has a 50% probability of being 100 m from the hunter

⇒ if we observe the dog, what can we say about the hunter?

The terms of the analogy are clear:

$$\begin{aligned} \text{hunter} &\leftrightarrow \text{true value} \\ \text{dog} &\leftrightarrow \text{observable}\,. \end{aligned}$$

The dog and the hunter

We know that a dog has a 50% probability of being 100 m from the hunter

⇒ if we observe the dog, what can we say about the hunter?

The terms of the analogy are clear:

$$\begin{aligned} \text{hunter} &\quad\leftrightarrow\quad \text{true value} \\ \text{dog} &\quad\leftrightarrow\quad \text{observable}\,. \end{aligned}$$

Intuitive and reasonable answer:

> *"The hunter is, with 50% probability, within 100 m of the position of the dog."*

The dog and the hunter

- dog has a 50% probability of being 100 m from the hunter

The dog and the hunter

- dog has a 50% probability of being 100 m from the hunter
- hunter has a 50% probability of being 100 m from the dog

The dog and the hunter

- dog has a 50% probability of being 100 m from the hunter
- hunter has a 50% probability of being 100 m from the dog

Easy to understand that this conclusion is based on some tacit assumptions:

The dog and the hunter

- dog has a 50% probability of being 100 m from the hunter
- hunter has a 50% probability of being 100 m from the dog

Easy to understand that this conclusion is based on some tacit assumptions:

- the hunter can be anywhere around the dog

The dog and the hunter

- dog has a 50% probability of being 100 m from the hunter
- hunter has a 50% probability of being 100 m from the dog

Easy to understand that this conclusion is based on some tacit assumptions:

- the hunter can be anywhere around the dog
- the dog has no preferred direction of arrival at the point where we observe him.

The dog and the hunter

- dog has a 50% probability of being 100 m from the hunter
- hunter has a 50% probability of being 100 m from the dog

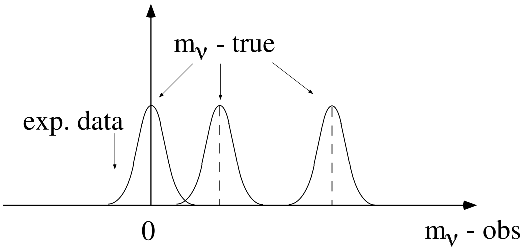
Easy to understand that this conclusion is based on some tacit assumptions:

- the hunter can be anywhere around the dog
- the dog has no preferred direction of arrival at the point where we observe him.

$\rightarrow$ not always valid!

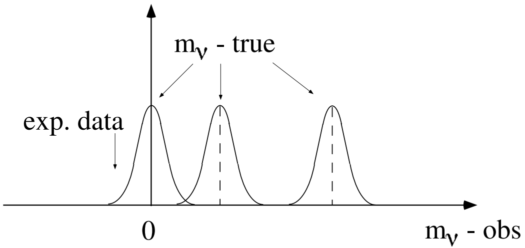
Measurement at the edge of a physical region

Electron-neutrino experiment, mass resolution $\sigma = 2\,\mathrm{eV}$, independent of m_ν.



Measurement at the edge of a physical region

Electron-neutrino experiment, mass resolution $\sigma = 2\,\mathrm{eV}$, independent of m_ν.

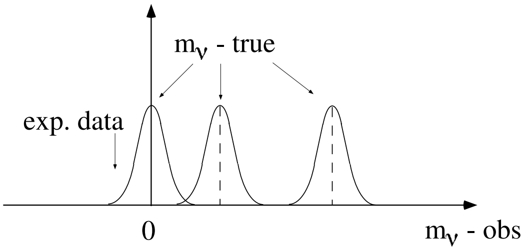


Observation: $-4\,\mathrm{eV}$.

What can we tell about m_ν?

Measurement at the edge of a physical region

Electron-neutrino experiment, mass resolution $\sigma = 2\,\text{eV}$, independent of m_ν.



Observation: $-4\,\text{eV}$.

What can we tell about m_ν?

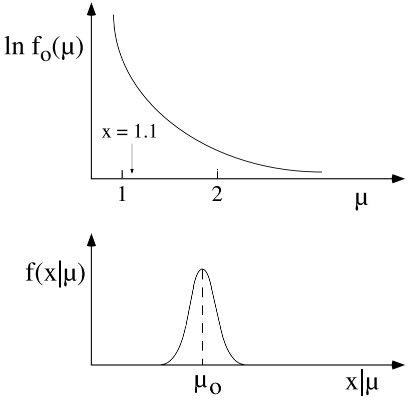
$m_\nu = -4 \pm 2$ eV ?

$P(-6 \leq m_\nu/\text{eV} \leq -2) = 68\%$?

$P(m_\nu \leq 0 \text{ eV}) = 98\%$?

Non-flat distribution of a physical quantity

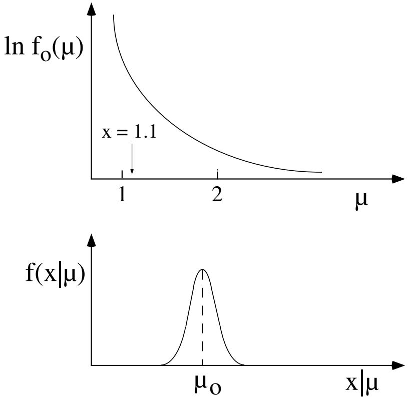
Imagine a cosmic ray particle or a bremsstrahlung γ.



Observed $x = 1.1$.

Non-flat distribution of a physical quantity

Imagine a cosmic ray particle or a bremsstrahlung γ.

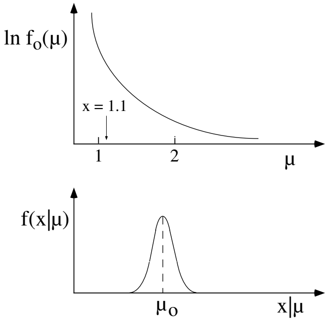


Observed $x = 1.1$.

What can we say about the true value μ that **has caused** this observation?

Non-flat distribution of a physical quantity

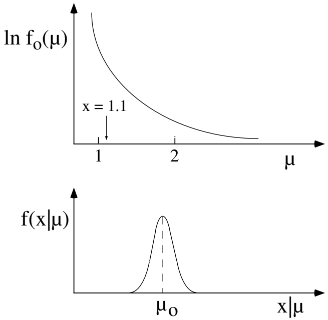
Imagine a cosmic ray particle or a bremsstrahlung γ.



Also in this case the formal definition of the confidence interval does not work.

Non-flat distribution of a physical quantity

Imagine a cosmic ray particle or a bremsstrahlung γ.

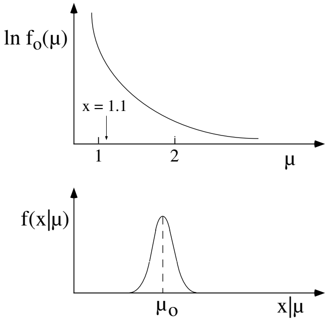


Also in this case the formal definition of the confidence interval does not work.

Intuitively, we feel that there is more chance that μ is on the left of 1.1 than on the right one.

Non-flat distribution of a physical quantity

Imagine a cosmic ray particle or a bremsstrahlung γ.



Also in this case the formal definition of the confidence interval does not work.

Intuitively, we feel that there is more chance that μ is on the left of 1.1 than on the right one.

In the jargon of the experimentalists, *"there are more migrations from left to right than from right to left"*.

Asymmetric detector response

These two examples deviate from the dog-hunter picture only because of an asymmetric possible position of the 'hunter', i.e our expectation about μ is not uniform.

Asymmetric detector response

These two examples deviate from the dog-hunter picture only because of an asymmetric possible position of the 'hunter', i.e our expectation about μ is not uniform.

But there are also interesting cases in which the response of the apparatus $f(x \mid \mu)$ is not symmetric around μ, e.g. the reconstructed momentum in a magnetic spectrometer.

Asymmetric detector response

These two examples deviate from the dog-hunter picture only because of an asymmetric possible position of the 'hunter', i.e our expectation about μ is not uniform.

But there are also interesting cases in which the response of the apparatus $f(x \,|\, \mu)$ is not symmetric around μ, e.g. the reconstructed momentum in a magnetic spectrometer.

Summing up:

the intuitive inversion of probability

$$P(\ldots \leq \overline{X} \leq \ldots) \Longrightarrow P(\ldots \leq \mu \leq \ldots)\,,$$

besides being theoretically unjustifiable in the frequestist approach to probability, yields results which are numerically correct only in the case of symmetric problems.

Summary about standard methods

Situation is not satisfactory in the critical situations that often occur in HEP, both in

- hypotheses tests
- confidence intervals

Summary about standard methods

Situation is not satisfactory in the critical situations that often occur in HEP, both in

- hypotheses tests
- confidence intervals

Moreover there are issues not easy to treat in that frame
[and I smile at the heroic effort to get some result :-)]

- systematic errors
- background

Implicit assumptions

We have seen clearly what are the hidden assumptions in the ‘naive probability inversion’ (that corresponds more or less to the prescriptions to build confidence intervals).

> We shall see that, similarly, there are hidden assumptions behind the naive probabilistic inversions.

Implicit assumptions

We have seen clearly what are the hidden assumptions in the 'naive probability inversion' (that corresponds more or less to the prescriptions to build confidence intervals).

We shall see that, similarly, there are hidden assumptions behind the naive probabilistic inversions.

Curiously enough, these methods are advertised as objective because they do not need as input our scientific expectations of where the value of the quantity might lie, or of which physical hypothesis seems more reasonable!

Implicit assumptions

We have seen clearly what are the hidden assumptions in the 'naive probability inversion' (that corresponds more or less to the prescriptions to build confidence intervals).

We shall see that, similarly, there are hidden assumptions behind the naive probabilistic inversions.

Curiously enough, these methods are advertised as objective because they do not need as input our scientific expectations of where the value of the quantity might lie, or of which physical hypothesis seems more reasonable!

But if we are convinced (by logic, or by the fact that neglecting that knowledge paradoxical results can be achieved) that prior expectation is relevant in inferences, we cannot accept methods which systematically neglect it and that, for that reason, solve problems different from those we are interested in!

The End