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."

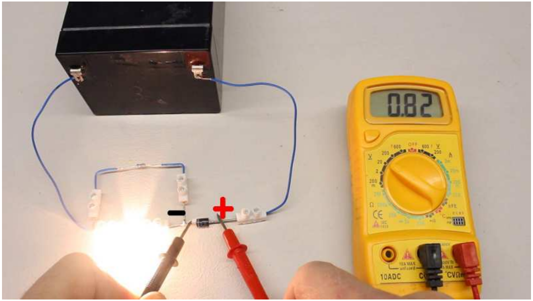
["Then, clever as I am,
I can earn a lot of money."]

- No rush to get formulae

$\rightarrow$ If you understand the basic reasoning
you can derive many formulae by yourself'!

What is measurement?

What is measurement?

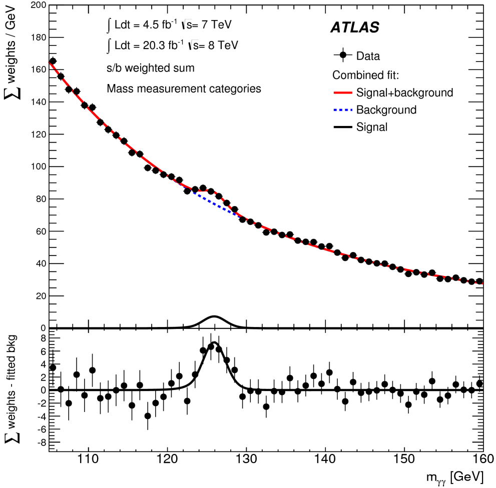


What is measurement?

What is measurement?

What is measurement?

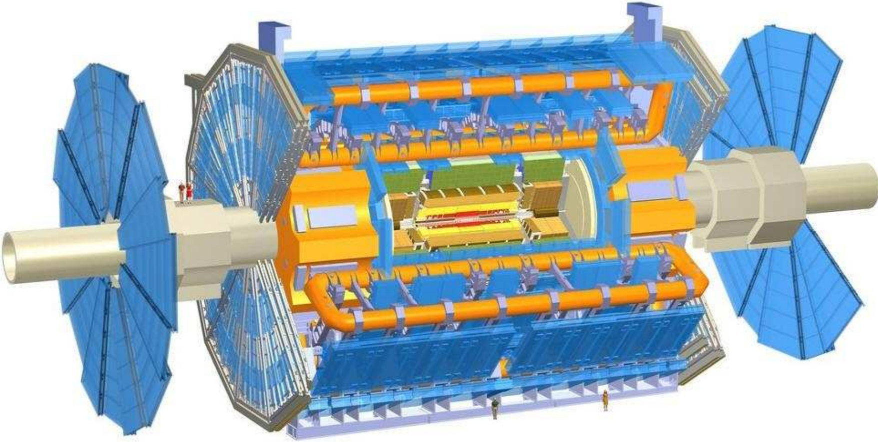
Higgs $\rightarrow \gamma\gamma$ (2012)



Two-photon *invariant mass*

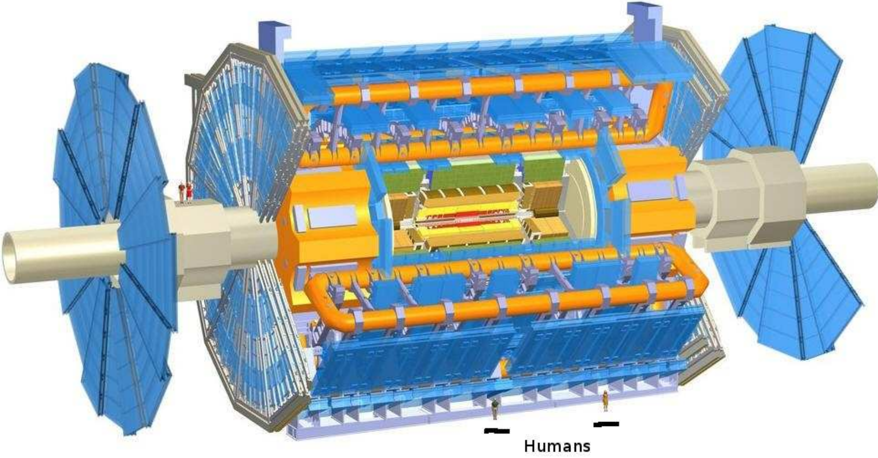
What is measurement?

ATLAS Experiment at LHC (CERN, Geneva)



What is measurement?

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

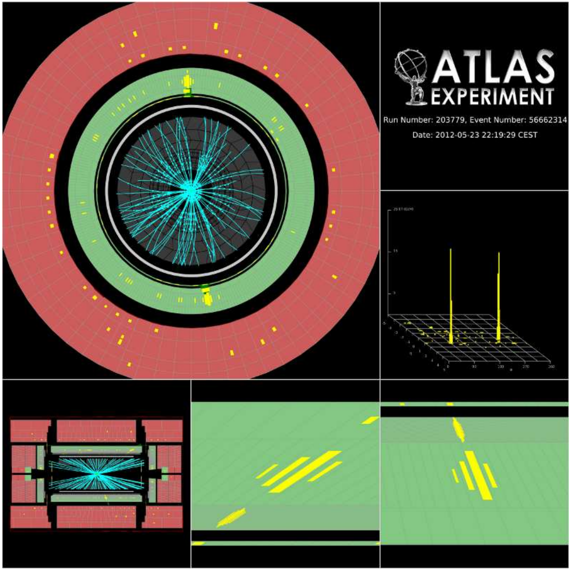


≈ 3000 km cables

≈ 7000 tonnes

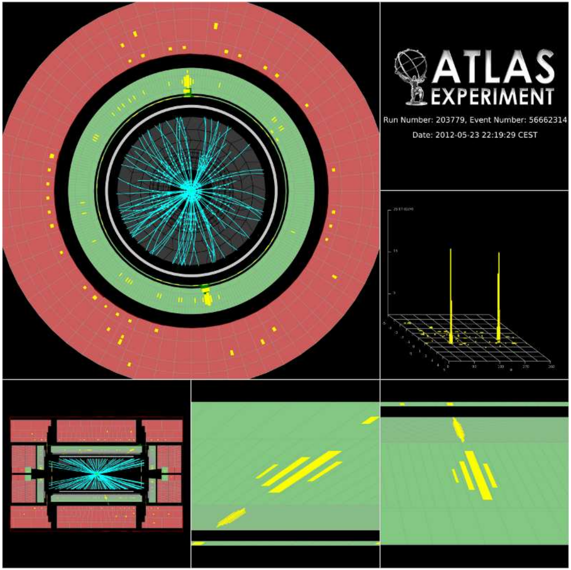
≈ 100 millions electronic channels

What is measurement?



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

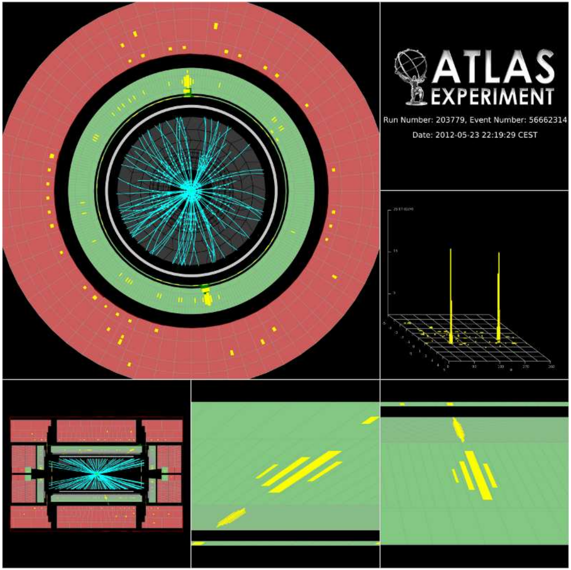
What is measurement?



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

Higgs $\to \gamma\gamma$?

What is measurement?

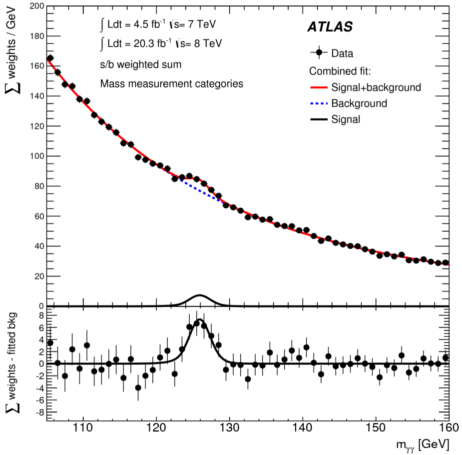


Two flashes of 'light' (2 γ's) in a 'noisy' environment.

Higgs $\rightarrow \gamma\gamma$? Probably not...

What is measurement?

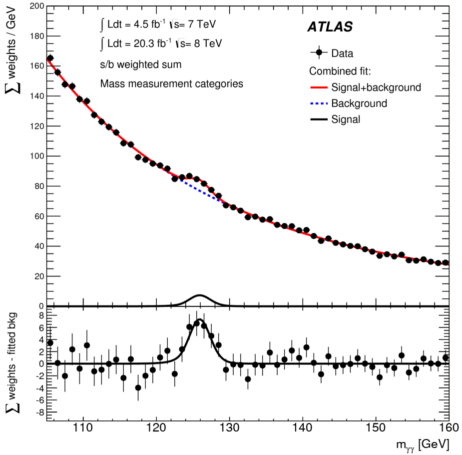
Higgs $\to \gamma\gamma$



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

What is measurement?

Higgs $\to \gamma\gamma$

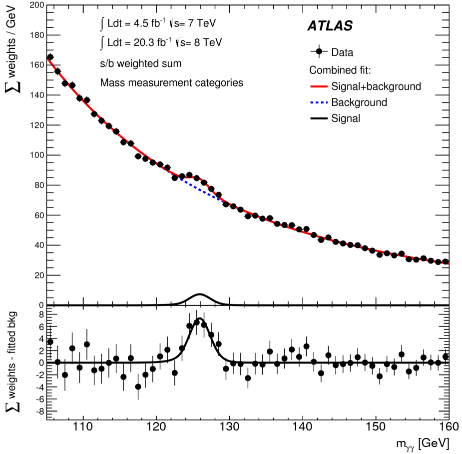


$\Rightarrow \left\{ \begin{array}{l} \text{Mass value} \\ \text{Production rate} \end{array} \right.$

(with uncertainties)

What is measurement?

Higgs $\rightarrow \gamma\gamma$



$\Rightarrow \left\{ \begin{array}{l} \text{Mass value} \\ \text{Production rate} \end{array} \right.$

(with uncertainties)

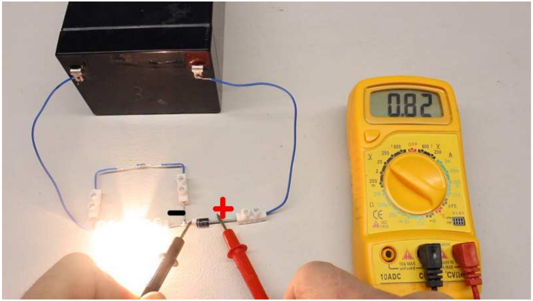
Quite indirect measurements of something we do not "see"!

Can we “see” physics quantities?

But, can we see 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:

$$\color{red}{m}\color{black}{g - k}\color{blue}{\Delta x}\color{black}{\quad = \quad 0}$$

$$\color{blue}{\Delta x}\color{black}{\quad \rightarrow \quad}\color{blue}{\theta}\color{black}{\rightarrow}\color{blue}{\text{scale reading}}$$

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

Measuring a mass on a scale

Equilibrium:

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

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

From the reading to the value of the mass:

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

Measuring a mass on a balance

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

Measuring a mass on a balance

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

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

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 not at "$R_\oplus$" from the Earth center;

Measuring a mass on a balance

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

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...

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. . .

Measuring a mass on a balance

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

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.

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

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 not 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

$$\text{scale reading} \xrightarrow[\textit{given } g,\ k,\ \text{“etc.”}\ldots]{\hspace{5cm}} m$$

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

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

Dependence on 'k':

- temperature
- non linearity
- ...

Measuring a mass on a balance

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

Dependence on 'k':

- temperature
- non linearity
- ...

$\Delta x \rightarrow \theta \rightarrow$ **scale reading:**

- left to your imagination...

Measuring a mass on a balance

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

Dependence on 'k':

- temperature
- non linearity
- ...

$\Delta x \rightarrow \theta \rightarrow$ **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.

Measuring a mass on a balance

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

Dependence on 'k':

- temperature
- non linearity
- ...

$\Delta x \rightarrow \theta \rightarrow$ **scale reading:**

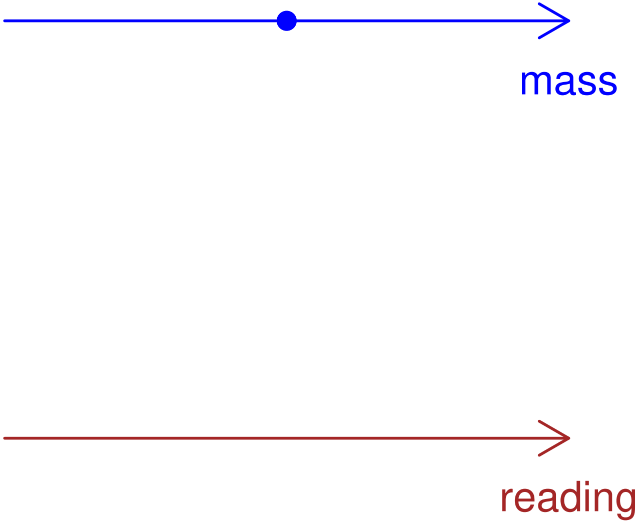
- left to your imagination...

$+$ **randomic effects:**

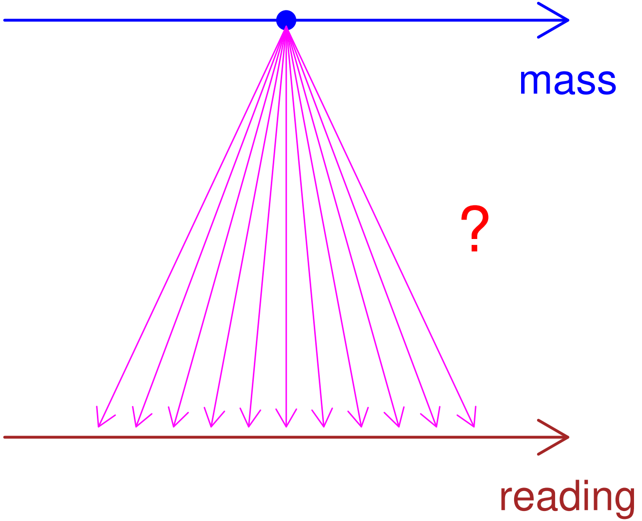
- stopping position of damped oscillation;
- variability of all quantities of influence (in the ISO-GUM sense);
- reading of analog scale.

$\Rightarrow\ m$??

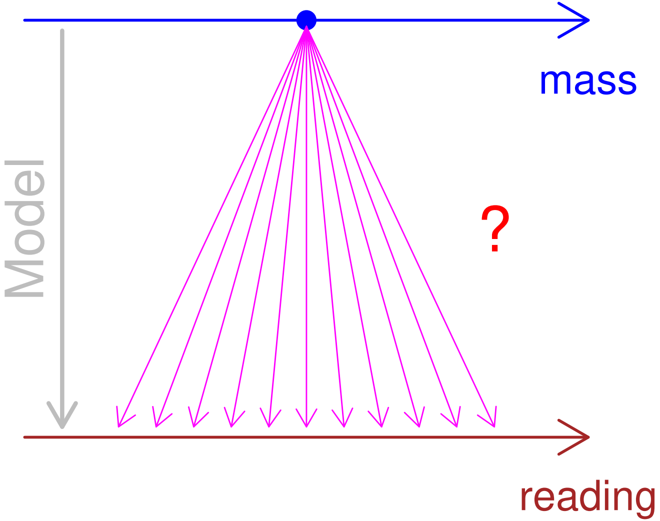
Mass $\longrightarrow$ Reading



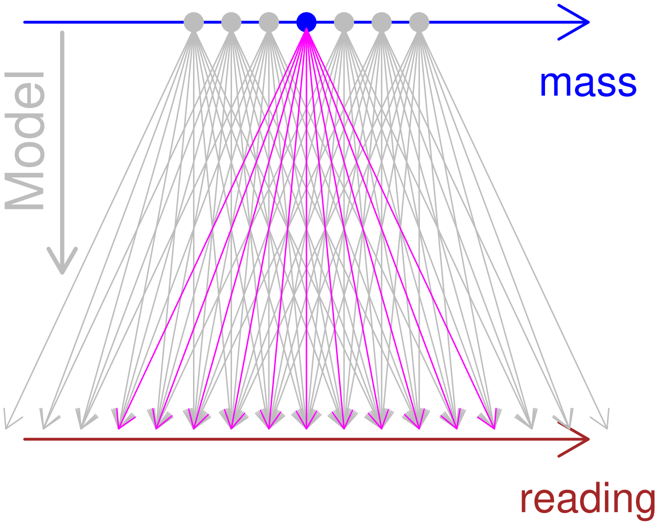
Mass $\longrightarrow$ Reading



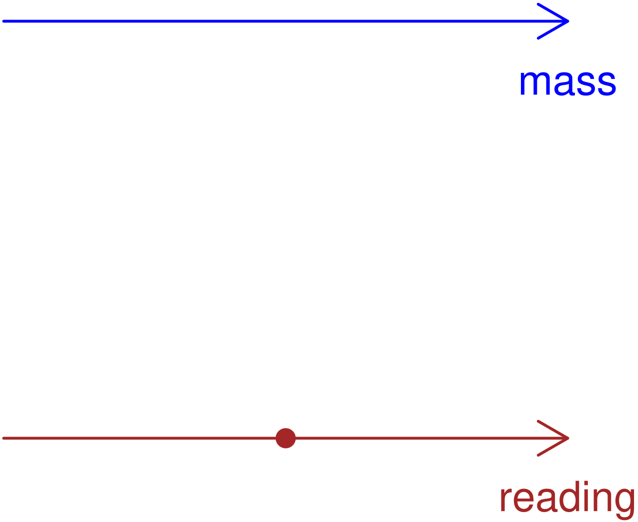
Mass $\longrightarrow$ Reading



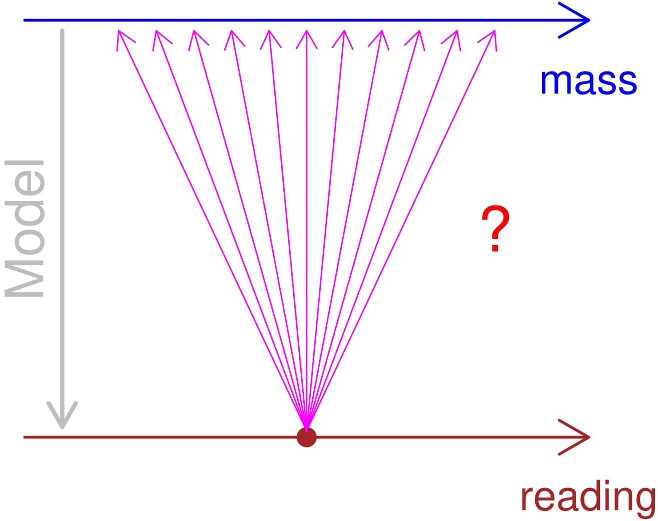
Mass $\longrightarrow$ reading



Reading $\longrightarrow$ ‘true’ mass



Reading $\longrightarrow$ '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*

$\rightarrow$ g

$\rightarrow$where?

$\rightarrow$inertial effects subtracted?

2 *imperfect realization of the definition of the measurand*

$\rightarrow$ scattering on neutron

$\rightarrow$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)

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."*

$\Rightarrow$ *"…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;*

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."*

Type A and Type B uncertainties

Remarks:

- "Type A is due to random ('statistical') effects":

Type A and Type B uncertainties

Remarks:

- "Type A is due to random ('statistical') effects":

$\surd$

Type A and Type B uncertainties

Remarks:

- “Type A is due to random (‘statistical’) effects”:

  $\surd$

- “Type B is due to all other effects (‘systematical’)”

  not necessarily

Type A and Type B uncertainties

Remarks:

- "Type A is due to random ('statistical') effects":

  $\surd$

- "Type B is due to all other effects ('systematical')"

  not necessarily

  In general yes,

Type A and Type B uncertainties

Remarks:

- "Type A is due to random ('statistical') effects":

  $\surd$

- "Type B is due to all other effects ('systematical')"

  not necessarily

  In general yes, but imagine the case in which

  - a single measurement has been performed on a particular quantity;

Type A and Type B uncertainties

Remarks:

- "Type A is due to random ('statistical') effects":

  $\surd$

- "Type B is due to all other effects ('systematical')"

  not necessarily

  In general yes, but imagine the case in which
  - a single measurement has been performed on a particular quantity;
  - previous experiments on <u>other</u> particular quantities, on the same condition, have provided a repeatability standard deviation σ_r;

Type A and Type B uncertainties

Remarks:

- "Type A is due to random ('statistical') effects":

$$\surd$$

- "Type B is due to all other effects ('systematical')"

  not necessarily

  In general yes, but imagine the case in which
  - a single measurement has been performed on a particular quantity;
  - previous experiments on <u>other</u> particular quantities, on the same condition, have provided a repeatability standard deviation σ_r;
  - at the best of our knowledge

$$u = \sigma_r\,.$$

Type A and Type B uncertainties

Remarks:

- "Type A is due to random ('statistical') effects":

$$\surd$$

- "Type B is due to all other effects ('systematical')"

not necessarily

In general yes, but imagine the case in which

- a single measurement has been performed on a particular quantity;
- previous experiments on <u>other</u> particular quantities, on the same condition, have provided a repeatability standard deviation σ_r;
- at the best of our knowledge

$$u = \sigma_r \, .$$

$\rightarrow$ Type B uncertainty due to 'statistical effects'.

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 including 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 including 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.

$\rightarrow$ Right in most cases!

$\rightarrow$ Good sense of physicists $\Longleftrightarrow$ 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.

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

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 scientists
   $P(\overline{x} - \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + \frac{\sigma}{\sqrt{n}}) = 68\%$
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 scientists
   $P(\overline{x} - \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + \frac{\sigma}{\sqrt{n}}) = 68\%$
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 $\left[\overline{x} - \frac{\sigma}{\sqrt{n}},\ \overline{x} + \frac{\sigma}{\sqrt{n}}\right]$ 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 scientists
   $P(\overline{x} - \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + \frac{\sigma}{\sqrt{n}}) = 68\%$
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 $\left[\overline{x} - \frac{\sigma}{\sqrt{n}},\ \overline{x} + \frac{\sigma}{\sqrt{n}}\right]$ 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 most 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 most of you

   - But it depends on what we mean by probability

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

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

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

- But it depends on what we mean by probability
- If probability is the "limit of the frequency", this statement is meaningless,

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 most 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

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

(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* $\left[\overline{x} - \frac{\sigma}{\sqrt{n}},\ \overline{x} + \frac{\sigma}{\sqrt{n}}\right]$*."*

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* $\left[\overline{x} - \frac{\sigma}{\sqrt{n}},\ \overline{x} + \frac{\sigma}{\sqrt{n}}\right]$*."*
   - 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* $\left[\overline{x} - \frac{\sigma}{\sqrt{n}},\ \overline{x} + \frac{\sigma}{\sqrt{n}}\right]$*."*

- 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* $\left[\overline{x} - \frac{\sigma}{\sqrt{n}},\ \overline{x} + \frac{\sigma}{\sqrt{n}}\right]$*."*

- Nothing wrong in principle (in my opinion)
- but a $\sqrt{2}$ mistake in the width of the interval (or a $\approx -24\%$ mistake in the level of probability):

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* $\left[\overline{x} - \frac{\sigma}{\sqrt{n}},\ \overline{x} + \frac{\sigma}{\sqrt{n}}\right]$*."*

- Nothing wrong in principle (in my opinion)

- but a $\sqrt{2}$ mistake in the width of the interval (or a $\approx -24\%$ mistake in the level of probability):

$$\rightarrow\ P\left(\overline{x} - \frac{\sigma}{\sqrt{n}}\ \leq\ \overline{x}_f\ \leq\ \overline{x} + \frac{\sigma}{\sqrt{n}}\right) \ = \ 52\%\,,$$

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* $\left[\overline{x} - \frac{\sigma}{\sqrt{n}},\ \overline{x} + \frac{\sigma}{\sqrt{n}}\right]$*."*

- Nothing wrong in principle (in my opinion)
- but a $\sqrt{2}$ mistake in the width of the interval (or a $\approx -24\%$ mistake in the level of probability):

$$\rightarrow\ P\left(\overline{x} - \frac{\sigma}{\sqrt{n}} \ \leq\ \overline{x}_f\ \leq\ \overline{x} + \frac{\sigma}{\sqrt{n}}\right) \ = \ 52\%\,,$$

where $\overline{x}_f$ stands for future averages;
or

$$P\left(\overline{x} - \sqrt{2}\,\frac{\sigma}{\sqrt{n}} \ \leq\ \overline{x}_f\ \leq\ \overline{x} + \sqrt{2}\,\frac{\sigma}{\sqrt{n}}\right) = 68\%\,,$$

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* $\left[\overline{x} - \frac{\sigma}{\sqrt{n}},\ \overline{x} + \frac{\sigma}{\sqrt{n}}\right]$*."*

- Nothing wrong in principle (in my opinion)
- but a $\sqrt{2}$ mistake in the width of the interval
  (or a $\approx -24\%$ mistake in the level of probability):

$$\rightarrow\ P\left(\overline{x} - \frac{\sigma}{\sqrt{n}}\ \leq\ \overline{x}_f\ \leq\ \overline{x} + \frac{\sigma}{\sqrt{n}}\right)\ =\ 52\%\,,$$

where $\overline{x}_f$ stands for future averages;
or

$$P\left(\overline{x} - \sqrt{2}\,\frac{\sigma}{\sqrt{n}}\ \leq\ \overline{x}_f\ \leq\ \overline{x} + \sqrt{2}\,\frac{\sigma}{\sqrt{n}}\right) = 68\%\,,$$

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:

  → `https://www.roma1.infn.it/~dagos/ci_calc.html`

‘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:

  → `https://www.roma1.infn.it/~dagos/ci_calc.html`

  → 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:

  → `https://www.roma1.infn.it/~dagos/ci_calc.html`

  → 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:

  → `https://www.roma1.infn.it/~dagos/ci_calc.html`

  → 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 scientist who care about your quantity of interest, 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:

  → `https://www.roma1.infn.it/~dagos/ci_calc.html`

  → 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 scientist who care about your quantity of interest, think in terms of ‘confidence’:

  ⇒ 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:

  → `https://www.roma1.infn.it/~dagos/ci_calc.html`

  → 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 scientist who care about your quantity of interest, think in terms of ‘confidence’:

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

For a more argued criticism on how confidence intervals technically derive (trictly following the frequentistic prescription):
⇒ `arXiv:physics/0605140 [physics.data-an]`

Arbitrary probability inversions

How do we turn, just 'intuitively'

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

into

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

Arbitrary probability inversions

How do we turn, just 'intuitively'

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

into

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

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

$\Rightarrow$ 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:

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

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

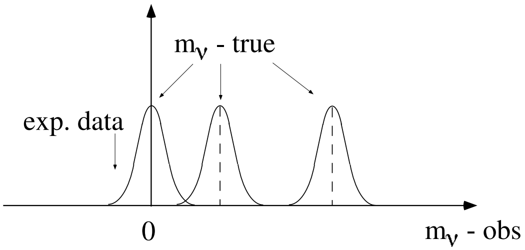
Easy to understand that this conclusion is based on some tacit assumptions:

- 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!

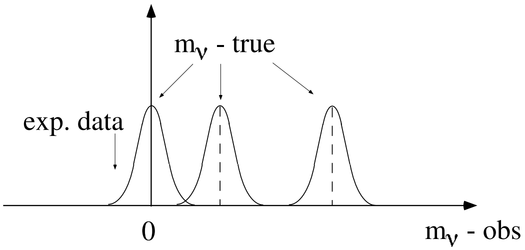
Measurement at the edge of a physical region

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



Measurement at the edge of a physical region

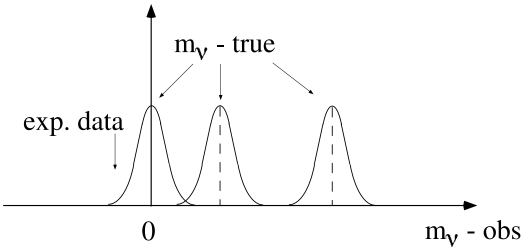
Electron-neutrino experiment, mass resolution $\sigma = 2\,\mathrm{eV}$, independent of m_ν.



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?

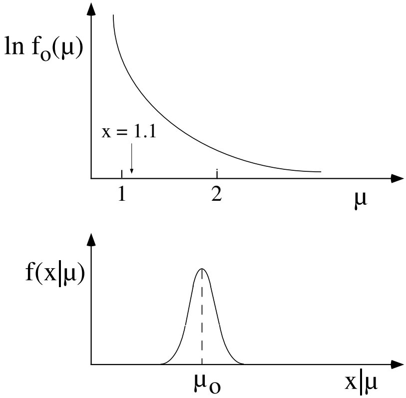
$m_\nu = -4 \pm 2$ eV ?

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

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

Non-flat distribution of a physical quantity

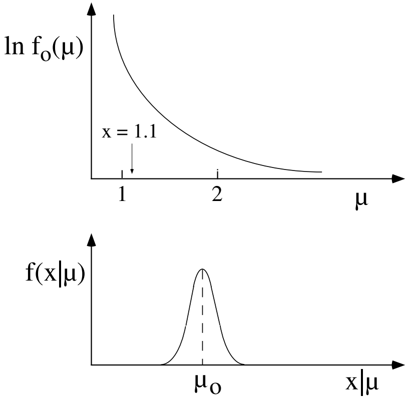
Imagine a cosmic ray particle or a bremsstrahlung γ.



Observed $x = 1.1$.

Non-flat distribution of a physical quantity

Imagine a cosmic ray particle or a bremsstrahlung γ.

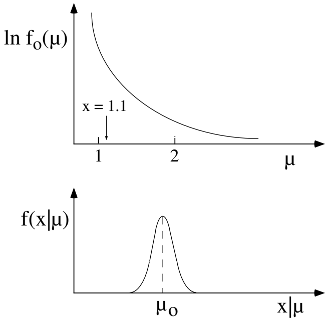


Observed $x = 1.1$.

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

Non-flat distribution of a physical quantity

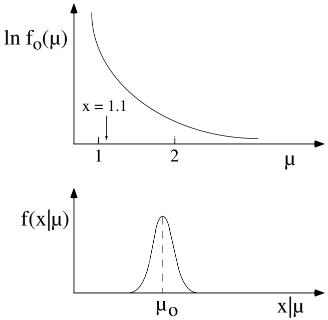
Imagine a cosmic ray particle or a bremsstrahlung γ.



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.

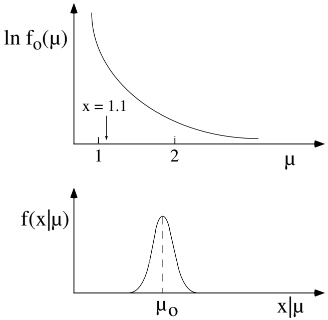


Also in this case the formal definition of the confidence interval does not work.

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 $f(x \mid \mu)$ 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!

Let's restart

Observation $\rightarrow$ value of a quantity

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

Observations → hypotheses

This problem occurs not only "determining" *the* value of a physical quantity.

Observations $\rightarrow$ hypotheses

This problem occurs not only "determining" *the* value of a physical quantity.

- Experimental observation ('data') $\rightarrow$ responsible cause.

Observations $\rightarrow$ hypotheses

This problem occurs not only “determining” *the* value of a physical quantity.

- Experimental observation (‘data’) $\rightarrow$ responsible cause.

(But logically no substantial difference.)

Human ancestral problem

???

Human ancestral problem

???

→ Chase?

→ Run away?

Observation → hypotheses

Dependence from the context

Chase o Run away?

Dependence from the context

Chase o Run away?

...or simply stay quite

Dependence from the context

Chase o Run away?

. . . or simply stay quite

if it is a mold in a museum,

or an artificial track in a school garden,

. . .

Dependence from the context

Chase o Run away?

. . . or simply stay quite

if it is a mold in a museum,

or an artificial track in a school garden,

. . .

(. . . or we are just sated tourists, with no interest in chasing, well protected inside our safari minibus)

Contemporary anthropology (and technology)

???

Effect and possible causes

Effect: car broken down

Effect and possible causes

Effect: car broken down

- Causes:
  - no gasoline
  - broken pump
  - electrical failure

Effect and possible causes

Effect: car broken down

- Causes:
  - no gasoline
  - broken pump
  - electrical failure
  - other (I am not a mechanic. . .)

Effect and possible causes

Effect: car broken down

- Causes:
  - no gasoline
  - broken pump
  - electrical failure
  - other (I am not a mechanic...)
- Guess of the expert:
  - he looks for (or ask about) collateral effects (noise, ...)
  - he has his own ideas about most likely causes.

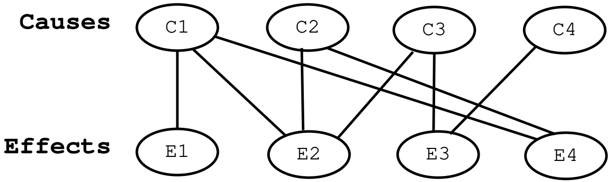
Effect and possible causes

Effect: car broken down

- Causes:
  - no gasoline
  - broken pump
  - electrical failure
  - other (I am not a mechanic. . .)
- Guess of the expert:
  - he looks for (or ask about) collateral effects (noise, . . .)
  - he has his own ideas about most likely causes.
- Action: balance between probability of the several hypotheses, costs and times.

Causes $\rightarrow$ effects

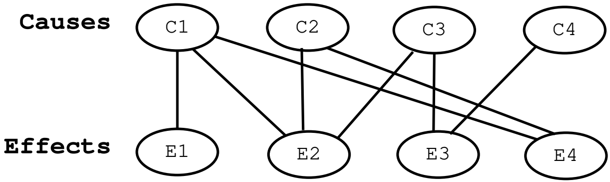
The same *apparent* cause might produce several,different effects



Given an observed effect, we are not sure about the exact cause that has produced it.

Causes $\rightarrow$ effects

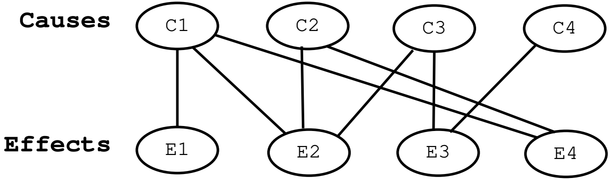
The same *apparent* cause might produce several,different effects



Given an observed effect, we are not sure about the exact cause that has produced it.

Causes $\rightarrow$ effects

The same *apparent* cause might produce several,different effects



Given an observed effect, we are not sure about the exact cause that has produced it.

$$E_2 \Rightarrow \{C_1,\ C_2,\ C_3\}?$$

The "essential problem" of the Sciences

"Now, these problems are classified as *probability of causes*, and are most interesting of all for their scientific applications. I play at *écarté* with a gentleman whom I know to be perfectly honest. What is the chance that he turns up the king? It is $1/8$. This is a problem of the probability of effects.

"Now, these problems are classified as *probability of causes*, and are most interesting of all for their scientific applications. I play at *écarté* with a gentleman whom I know to be perfectly honest. What is the chance that he turns up the king? It is $1/8$. This is a problem of the probability of effects.

I play with a gentleman whom I do not know. He has dealt ten times, and he has turned the king up six times. What is the chance that he is a sharper? This is a problem in the probability of causes. It may be said that it is the essential problem of the experimental method."

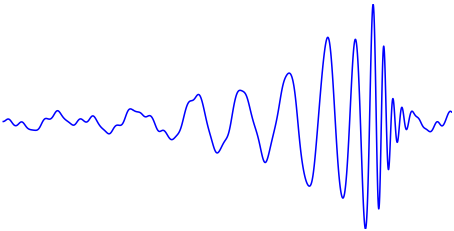
(H. Poincaré – *Science and Hypothesis*)

The "essential problem" of the Sciences

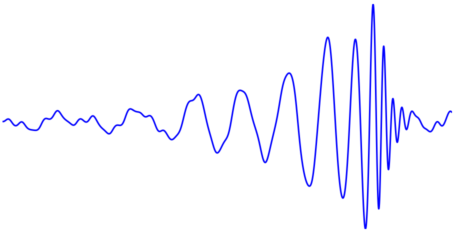
> "Now, these problems are classified as *probability of causes*, and are most interesting of all for their scientific applications. I play at *écarté* with a gentleman whom I know to be perfectly honest. What is the chance that he turns up the king? It is $1/8$. This is a problem of the probability of effects.
>
> I play with a gentleman whom I do not know. He has dealt ten times, and he has turned the king up six times. What is the chance that he is a sharper? This is a problem in the probability of causes. It may be said that it is the essential problem of the experimental method."
>
> (H. Poincaré – *Science and Hypothesis*)

Why we (or most of us) have not been taught how to tackle this kind of problems?

Who has done this 'scribble'?

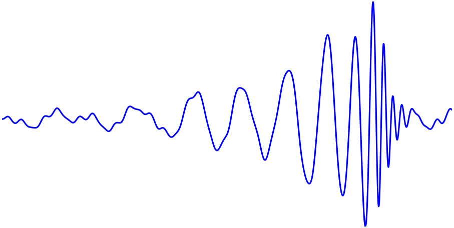


Who has done this ‘scribble’?

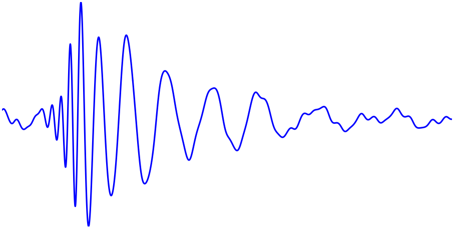


???

Who has done this 'scribble'?

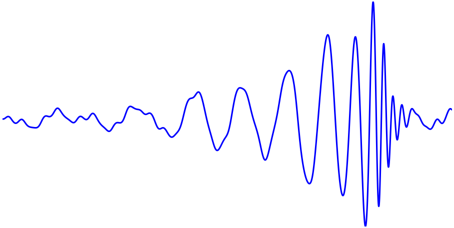


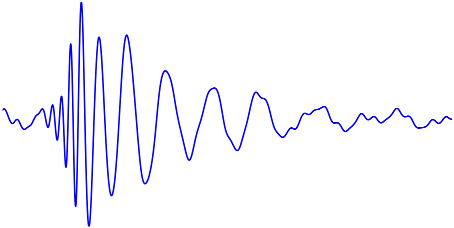
???



???

Who has done this 'scribble'?





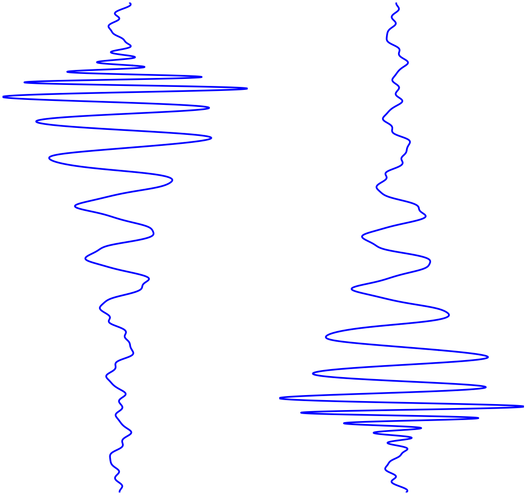
???

???

- Cardiogram?
- Signature?
- Sound?
- Earthquake?

Let's change orientation

(pure despair. . .)



???

Contextualization

Such an information, lacking details about

- what the points mean;

Contextualization

Such an information, lacking details about

- what the points mean;
- how it has been obtained;

Contextualization

Such an information, lacking details about

- what the points mean;
- how it has been obtained;
- with which device;

Contextualization

Such an information, lacking details about

- what the points mean;
- how it has been obtained;
- with which device;
- by **whom**;

Contextualization

Such an information, lacking details about

- what the points mean;
- how it has been obtained;
- with which device;
- by **whom**;
- etc. etc.

does not represent Scientific Knowledge!

Contextualization

Such an information, lacking details about

- what the points mean;
- how it has been obtained;
- with which device;
- by **whom**;
- etc. etc.

does not represent Scientific Knowledge!

It is simply a *scribble*!

Contextualization

Such an information, lacking details about

- what the points mean;
- how it has been obtained;
- with which device;
- by **whom**;
- etc. etc.

does not represent Scientific Knowledge!

It is simply a *scribble*!

Distrust the

Dogma of the Immaculate Observation

Context + further details

Things change completely when we get informed that

- it comes from an GW interferometer;

Context + further details

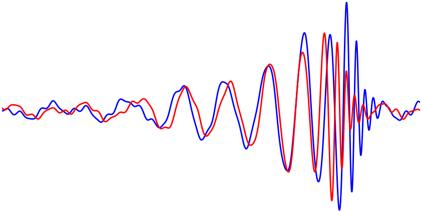
Things change completely when we get informed that

- it comes from an GW interferometer;
  [**Excellent microsysmographs**!!]

Context + further details

Things change completely when we get informed that

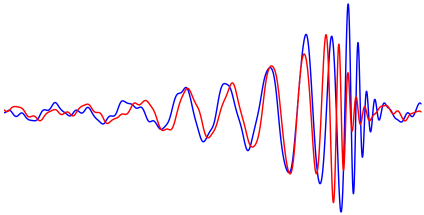
- it comes from an GW interferometer;
  [**Excellent microsysmographs**!!]
- a second 'signature', practically identical, was detected almost simultaneously by another interferometer c.a 3000 km apart!



Context + further details

Things change completely when we get informed that

- it comes from an GW interferometer;
  [**Excellent microsysmographs**!!]
- a second 'signature', practically identical, was detected almost simultaneously by another interferometer c.a 3000 km apart!

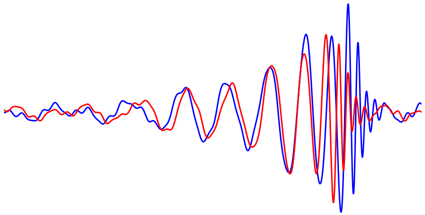


- the 'signer' is 'someone' well known to experts of the field.

Context + further details

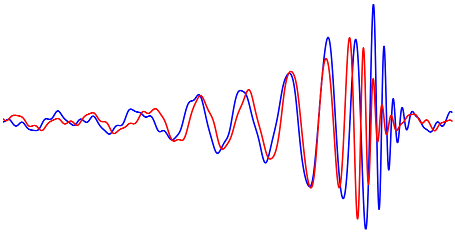
Things change completely when we get informed that

- it comes from an GW interferometer;
  [**Excellent microsysmographs**!!]
- a second 'signature', practically identical, was detected almost simultaneously by another interferometer c.a 3000 km apart!



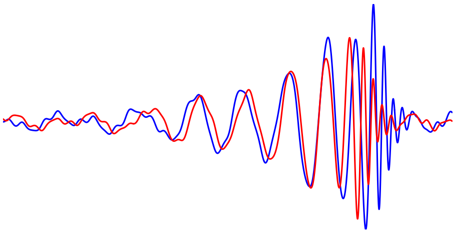
- the 'signer' is 'someone' well known to experts of the field.
  [We tend to believe what trusted people believe]

Effect $\rightarrow$ cause



? ? ?

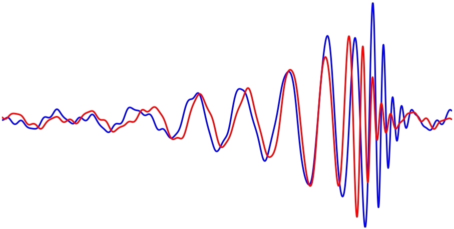
Effect $\rightarrow$ cause



? ? ?

- The expected 'signer'?

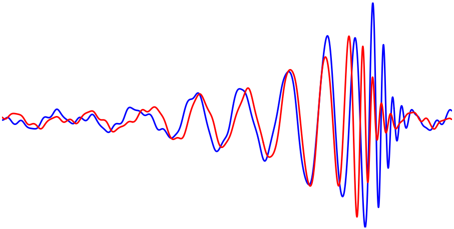
Effect → cause



???

- The expected 'signer'?
- Coherent micro-earthquake? (3000 km apart?)

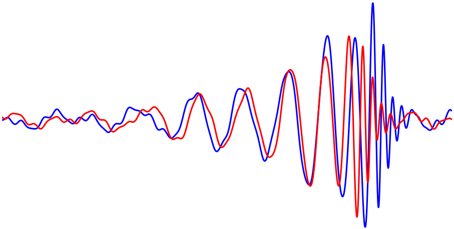
Effect $\rightarrow$ cause



? ? ?

- The expected 'signer'?
- Coherent micro-earthquake? (3000 km apart?)
- Pure coincidence of local trembles?

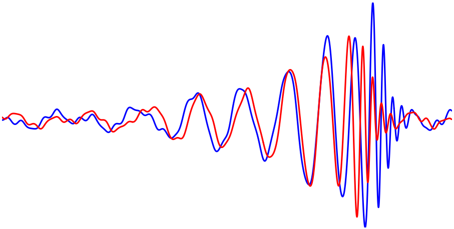
Effect → cause



???

- The expected 'signer'?
- Coherent micro-earthquake? (3000 km apart?)
- Pure coincidence of local trembles?
- A false 'signature' made to control the system?

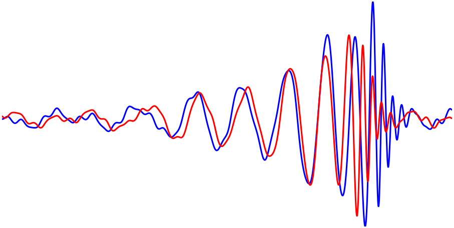
Effect $\rightarrow$ cause



? ? ?

- The expected 'signer'?
- Coherent micro-earthquake? (3000 km apart?)
- Pure coincidence of local trembles?
- A false 'signature' made to control the system?
- A 'sabotage' to discredit the collaboration?

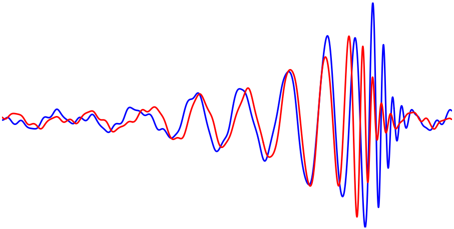
Effect → cause



? ? ?

- The expected 'signer'?
- Coherent micro-earthquake? (3000 km apart?)
- Pure coincidence of local trembles?
- A false 'signature' made to control the system?
- A 'sabotage' to discredit the collaboration?
- A tampering to favor the theorist of the 'signer'?

Effect → cause



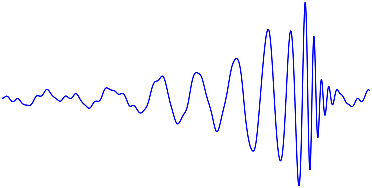
? ? ?

- The expected ‘signer’?
- Coherent micro-earthquake? (3000 km apart?)
- Pure coincidence of local trembles?
- A false ‘signature’ made to control the system?
- A ‘sabotage’ to discredit the collaboration?
- A tampering to favor the theorist of the ‘signer’?

(The last two causes are not just amenities!)

Effect $\rightarrow$ cause

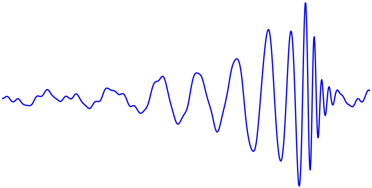
On the basis of the best knowledges about the possible causes



$\Rightarrow$ Gravitational wave

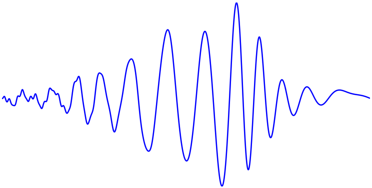
Effect $\rightarrow$ cause

On the basis of the best knowledges about the possible causes



$\Rightarrow$ Gravitational wave

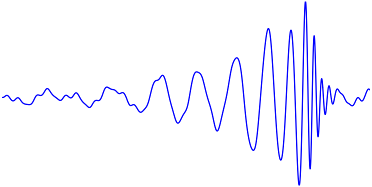
And if instead would have been this other 'scribble'



? ? ?

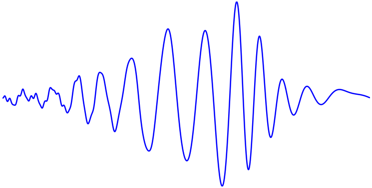
Effect $\rightarrow$ cause

On the basis of the best knowledges about the possible causes



$\Rightarrow$ Gravitational wave

And if instead would have been this other 'scribble'



? ? ?

Perhaps more likely a local random tremble...

$\{$Effect $\rightarrow$ cause$\}$ $\rightarrow$ 'concomitant causes' $\rightarrow$ corroboration

Our beliefs are strengthened by the fact that we can extract from the shape of the signal several parameters (masses, distance, etc...).

$\{$Effect $\rightarrow$ cause$\}$ $\rightarrow$ 'concomitant causes' $\rightarrow$ corroboration

Our beliefs are strengthened by the fact that we can extract from the shape of the signal several parameters (masses, distance, etc...).

$\Rightarrow$ An impressive number of parameters

$\{$Effect $\rightarrow$ cause$\}$ $\rightarrow$ 'concomitant causes' $\rightarrow$ corroboration

Our beliefs are strengthened by the fact that we can extract from the shape of the signal several parameters (masses, distance, etc...).

$\Rightarrow$ An impressive number of parameters although with unavoidable uncertainties

$\{$Effect $\rightarrow$ cause$\}$ $\rightarrow$ 'concomitant causes' $\rightarrow$ corroboration

Our beliefs are strengthened by the fact that we can extract from the shape of the signal several parameters (masses, distance, etc...).

$\Rightarrow$ An impressive number of parameters although with unavoidable uncertainties ... whose reasonabless strengthens our belief on the supposed phenomenon

$\{$Effect $\rightarrow$ cause$\}$ $\rightarrow$ 'concomitant causes' $\rightarrow$ corroboration

Our beliefs are strengthened by the fact that we can extract from the shape of the signal several parameters (masses, distance, etc...).

$\Rightarrow$ An impressive number of parameters although with unavoidable uncertainties ... whose reasonabless strengthens our belief on the supposed phenomenon

Note the *logical thread*:

- We **cannot** say (at least for the very first event) to have observed a gravitational wave, and then we search for the phenomenon which has produced it.

$\{$Effect $\rightarrow$ cause$\}$ $\rightarrow$ 'concomitant causes' $\rightarrow$ corroboration

Our beliefs are strengthened by the fact that we can extract from the shape of the signal several parameters (masses, distance, etc...).

$\Rightarrow$ An impressive number of parameters although with unavoidable uncertainties . . . whose reasonabless strengthens our belief on the supposed phenomenon

Note the *logical thread*:

- We **cannot** say (at least for the very first event) to have observed a gravitational wave, and then we search for the phenomenon which has produced it.

  On the contrary, **we believe it is gravitational wave** because of the overall consistency of the scenario.

$\{$Effect $\rightarrow$ cause$\}$ $\rightarrow$ 'concomitant causes' $\rightarrow$ corroboration

Our beliefs are strengthened by the fact that we can extract from the shape of the signal several parameters (masses, distance, etc...).

$\Rightarrow$ An impressive number of parameters although with unavoidable uncertainties ... whose reasonabless strengthens our belief on the supposed phenomenon

Note the *logical thread*:

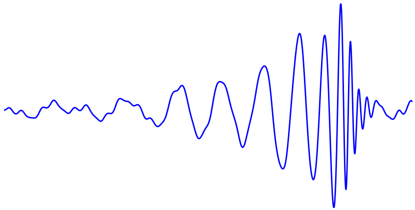
- We **cannot** say (at least for the very first event) to have observed a gravitational wave, and then we search for the phenomenon which has produced it.

  On the contrary, **we believe it is gravitational wave** because of the overall consistency of the scenario.

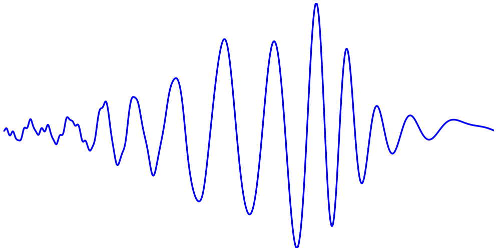
Despite of the 'sigmas'...

What is the difference between the two “scribbles”?

A)

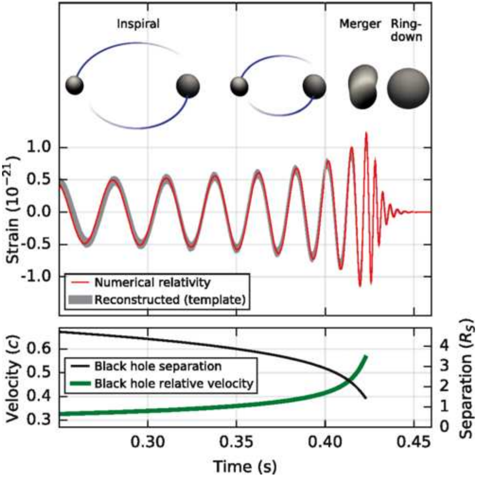


B)



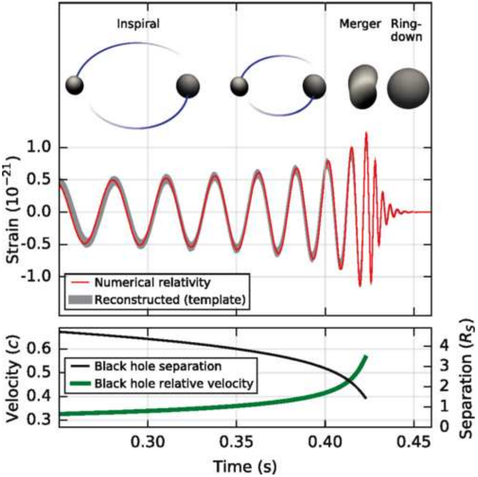
What is the difference between the two “scribbles”?

A)



What is the difference between the two "scribbles"?

A)



B)

NOTHING (as far as we understand it <u>now</u>. . .)

From the cosmic space down to problems of common mortals

An example easy to understand:

- two causes;
- two effects;

From the cosmic space down to problems of common mortals

An example easy to understand:

- two causes;
- two effects;
- medical diagnostics helps to clarify the issues:
  - easier to reach intuitive answers

From the cosmic space down to problems of common mortals

An example easy to understand:

- two causes;
- two effects;
- medical diagnostics helps to clarify the issues:
  - easier to reach intuitive answers
  - ... although if someone might have fallacious intuitions

From the cosmic space down to problems of common mortals

An example easy to understand:

- two causes;
- two effects;
- medical diagnostics helps to clarify the issues:
  - easier to reach intuitive answers
  - ... although if someone might have fallacious intuitions
    $\Rightarrow$ a formal guide helps us avoiding errors
    - $\Rightarrow$ **logics of the uncertain** (theory of probabilities)

AIDS test

An Italian citizen is selected at random
to undergo an AIDS test.
$\rightarrow$ Performance of clinical trial is not perfect, as customary:

$$\begin{aligned} P(\text{Pos}\,|\,\text{HIV}) &= 100\% \\ P(\text{Pos}\,|\,\overline{\text{HIV}}) &= 0.2\% \\ P(\text{Neg}\,|\,\overline{\text{HIV}}) &= 99.8\% \end{aligned}$$

H_1='HIV' (Infected) $\qquad E_1$ = Positive

H_2='$\overline{\text{HIV}}$' (Not infected) $\qquad E_2$ = Negative

AIDS test

An Italian citizen is selected at random
to undergo an AIDS test.
$\rightarrow$ Performance of clinical trial is not perfect, as customary:

$$\begin{aligned}
P(\text{Pos}\,|\,\text{HIV}) &= 100\% \\
P(\text{Pos}\,|\,\overline{\text{HIV}}) &= 0.2\% \\
P(\text{Neg}\,|\,\overline{\text{HIV}}) &= 99.8\%
\end{aligned}$$

$$\begin{tikzcd}
H_1\!=\!\text{'HIV' (Infected)} \arrow[rr] & & E_1 = \text{Positive} \\
H_2\!=\!\text{'}\overline{\text{HIV}}\text{' (Not infected)} \arrow[rr] \arrow[urr] & & E_2 = \text{Negative}
\end{tikzcd}$$

AIDS test

An Italian citizen is selected at random
to undergo an AIDS test.
$\rightarrow$ Performance of clinical trial is not perfect, as customary:

$$
\begin{aligned}
P(\text{Pos}\,|\,\text{HIV}) &= 100\% \\
P(\text{Pos}\,|\,\overline{\text{HIV}}) &= 0.2\% \\
P(\text{Neg}\,|\,\overline{\text{HIV}}) &= 99.8\%
\end{aligned}
$$

$$
\begin{tikzcd}
H_1=\text{'HIV' (Infected)} \arrow[r] & E_1 = \text{Positive} \\
H_2=\text{'}\overline{\text{HIV}}\text{' (Not infected)} \arrow[r] \arrow[ur] & E_2 = \text{Negative}
\end{tikzcd}
$$

Result: $\Rightarrow$ Positive

AIDS test

An Italian citizen is selected at random
to undergo an AIDS test.
$\rightarrow$ Performance of clinical trial is not perfect, as customary:

$$\begin{aligned} P(\mathrm{Pos}\,|\,\mathrm{HIV}) &= 100\% \\ P(\mathrm{Pos}\,|\,\overline{\mathrm{HIV}}) &= 0.2\% \\ P(\mathrm{Neg}\,|\,\overline{\mathrm{HIV}}) &= 99.8\% \end{aligned}$$

? H_1='HIV' (Infected) $\longleftarrow$ E_1 = Positive

? H_2='$\overline{\mathrm{HIV}}$' (Not infected) $\qquad$ E_2 = Negative

Result: $\Rightarrow$ <u>Positive</u>

Infected or not infected?

AIDS test: how to interpret the result?

Being $P(\text{Pos} \,|\, \overline{\text{HIV}}) = 0.2\%$ and having observed 'Positive', can we say?

- "It is practically impossible that the person is not infected, since it was practically impossible that a non infected person would result positive"

AIDS test: how to interpret the result?

Being $P(\text{Pos} \,|\, \overline{\text{HIV}}) = 0.2\%$ and having observed 'Positive', can we say?

- "It is practically impossible that the person is not infected, since it was practically impossible that a non infected person would result positive"
- "There is only 0.2% probability that the person has no HIV"

AIDS test: how to interpret the result?

Being $P(\text{Pos} \,|\, \overline{\text{HIV}}) = 0.2\%$ and having observed 'Positive', can we say?

- "It is practically impossible that the person is not infected, since it was practically impossible that a non infected person would result positive"
- "There is only 0.2% probability that the person has no HIV"
- "We are 99.8% confident that the person is infected"

AIDS test: how to interpret the result?

Being $P(\text{Pos} \,|\, \overline{\text{HIV}}) = 0.2\%$ and having observed 'Positive',
can we say

- "It is practically impossible that the person is not infected, since it was practically impossible that a non infected person would result positive"
- "There is only 0.2% probability that the person has no HIV"
- "We are 99.8% confident that the person is infected"
- "The hypothesis H_1='no HIV' is ruled out with 99.8% C.L."

?

AIDS test: how to interpret the result?

Being $P(\text{Pos}\,|\,\overline{\text{HIV}}) = 0.2\%$ and having observed 'Positive', can we say

- "It is practically impossible that the person is not infected, since it was practically impossible that a non infected person would result positive"
- "There is only 0.2% probability that the person has no HIV"
- "We are 99.8% confident that the person is infected"
- "The hypothesis H_1='no HIV' is ruled out with 99.8% C.L."

NO

AIDS test: how to interpret the result?

Being $P(\text{Pos} \,|\, \overline{\text{HIV}}) = 0.2\%$ and having observed 'Positive', can we say

- ~~"It is practically impossible that the person is not infected, since it was practically impossible that a non infected person would result positive"~~
- ~~"There is only 0.2% probability that the person has no HIV"~~
- ~~"We are 99.8% confident that the person is infected"~~
- ~~"The hypothesis H_1='no HIV' is ruled out with 99.8% C.L."~~

NO

Instead, $P(\text{HIV} \,|\, \text{Pos, random Italian}) \approx 45\%$

AIDS test: how to interpret the result?

Being $P(\text{Pos}\,|\,\overline{\text{HIV}}) = 0.2\%$ and having observed 'Positive', can we say

- "It is practically impossible that the person is not infected, since it was practically impossible that a non infected person would result positive"
- "There is only 0.2% probability that the person has no HIV"
- "We are 99.8% confident that the person is infected"
- "The hypothesis H_1='no HIV' is ruled out with 99.8% C.L."

NO

Instead, $P(\text{HIV}\,|\,\text{Pos, random Italian}) \approx 45\%$

(We will learn in the sequel how to evaluate it correctly)

AIDS test: how to interpret the result?

Being $P(\text{Pos} \,|\, \overline{\text{HIV}}) = 0.2\%$ and having observed 'Positive',
can we say

- "It is practically impossible that the person is not infected, since it was practically impossible that a non infected person would result positive"
- "There is only 0.2% probability that the person has no HIV"
- "We are 99.8% confident that the person is infected"
- "The hypothesis H_1='no HIV' is ruled out with 99.8% C.L."

NO

Instead, $P(\text{HIV} \,|\, \text{Pos, random Italian}) \approx 45\%$

$\Rightarrow$ Serious mistake! (not just 99.8% instead of 98.3% or so)

AIDS test: how to interpret the result?

Being $P(\text{Pos}\,|\,\overline{\text{HIV}}) = 0.2\%$ and having observed 'Positive', can we say

- ~~"It is practically impossible that the person is not infected, since it was practically impossible that a non infected person would result positive"~~
- ~~"There is only 0.2% probability that the person has no HIV"~~
- ~~"We are 99.8% confident that the person is infected"~~
- ~~"The hypothesis H_1='no HIV' is ruled out with 99.8% C.L."~~

NO

Instead, $P(\text{HIV}\,|\,\text{Pos, random Italian}) \approx 45\%$

$\Rightarrow$ Serious mistake! (not just 99.8% instead of 98.3% or so)

...from which bad decisions might result!

AIDS test

???

Where is the problem?

AIDS test

???

Where is the problem?

The previous statements, although dealing with probabilistic issues, **are not ground** on probability theory

AIDS test

???

Where is the problem?

The previous statements, although dealing with probabilistic issues, **are not ground** on probability theory

. . . and in these issues intuition can be fallacious!

AIDS test

???

Where is the problem?

The previous statements, although dealing with probabilistic issues, **are not ground** on probability theory

... and in these issues intuition can be fallacious!

$\Rightarrow$ A sound formal guidance can rescue us

$P(A \mid B) \leftrightarrow P(B \mid A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \mid B) \neq P(B \mid A)$

$P(A \mid B) \leftrightarrow P(B \mid A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \mid B) \neq P(B \mid A)$

- $P(\text{Positive} \mid \overline{HIV}) \neq P(\overline{HIV} \mid \text{Positive})$

$P(A \mid B) \leftrightarrow P(B \mid A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \mid B) \neq P(B \mid A)$

- $P(\text{Positive} \mid \overline{HIV}) \neq P(\overline{HIV} \mid \text{Positive})$
- $P(\text{Win} \mid \text{Play}) \neq P(\text{Play} \mid \text{Win})$ [Lotto]

$P(A \,|\, B) \leftrightarrow P(B \,|\, A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \,|\, B) \neq P(B \,|\, A)$

- $P(\text{Positive} \,|\, \overline{HIV}) \neq P(\overline{HIV} \,|\, \text{Positive})$
- $P(\text{Win} \,|\, \text{Play}) \neq P(\text{Play} \,|\, \text{Win})$ [Lotto]
- $P(\text{Pregnant} \,|\, \text{Woman}) \neq P(\text{Woman} \,|\, \text{Pregnant})$

$P(A \mid B) \leftrightarrow P(B \mid A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \mid B) \neq P(B \mid A)$

- $P(\text{Positive} \mid \overline{HIV}) \neq P(\overline{HIV} \mid \text{Positive})$
- $P(\text{Win} \mid \text{Play}) \neq P(\text{Play} \mid \text{Win})$ [Lotto]
- $P(\text{Pregnant} \mid \text{Woman}) \neq P(\text{Woman} \mid \text{Pregnant})$

In particular

- A cause might produce a given effect with very low probability, and nevertheless could be the most probable cause of that effect

$P(A\,|\,B) \leftrightarrow P(B\,|\,A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A\,|\,B) \neq P(B\,|\,A)$

- $P(\text{Positive}\,|\,\overline{HIV}) \neq P(\overline{HIV}\,|\,\text{Positive})$
- $P(\text{Win}\,|\,\text{Play}) \neq P(\text{Play}\,|\,\text{Win})$ [Lotto]
- $P(\text{Pregnant}\,|\,\text{Woman}) \neq P(\text{Woman}\,|\,\text{Pregnant})$

In particular

- A cause might produce a given effect with very low probability, and nevertheless could be the most probable cause of that effect, **often the only one!**

$P(A \,|\, B) \leftrightarrow P(B \,|\, A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \,|\, B) \neq P(B \,|\, A)$

- $P(\text{Positive} \,|\, \overline{HIV}) \neq P(\overline{HIV} \,|\, \text{Positive})$
- $P(\text{Win} \,|\, \text{Play}) \neq P(\text{Play} \,|\, \text{Win})$ [Lotto]
- $P(\text{Pregnant} \,|\, \text{Woman}) \neq P(\text{Woman} \,|\, \text{Pregnant})$

In particular

- A cause might produce a given effect with very low probability, and nevertheless could be the most probable cause of that effect, **often the only one!**

In particular

$$P(E \,|\, H) \lll 1 \quad \underline{\text{does not}} \text{ imply} \quad P(H \,|\, E) \lll 1$$

$P(A \mid B) \leftrightarrow P(B \mid A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \mid B) \neq P(B \mid A)$

- $P(\text{Positive} \mid \overline{HIV}) \neq P(\overline{HIV} \mid \text{Positive})$
- $P(\text{Win} \mid \text{Play}) \neq P(\text{Play} \mid \text{Win})$ [Lotto]
- $P(\text{Pregnant} \mid \text{Woman}) \neq P(\text{Woman} \mid \text{Pregnant})$

In particular

- A cause might produce a given effect with very low probability, and nevertheless could be the most probable cause of that effect, **often the only one!**

In particular

$P(E \mid H) \lll 1$ <u>does not</u> imply $P(H \mid E) \lll 1$

and '<u>hence</u>' $P(\overline{H} \mid E) \approx 1$

$P(A \,|\, B) \leftrightarrow P(B \,|\, A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \,|\, B) \neq P(B \,|\, A)$

- $P(\text{Positive} \,|\, \overline{HIV}) \neq P(\overline{HIV} \,|\, \text{Positive})$
- $P(\text{Win} \,|\, \text{Play}) \neq P(\text{Play} \,|\, \text{Win})$ [Lotto]
- $P(\text{Pregnant} \,|\, \text{Woman}) \neq P(\text{Woman} \,|\, \text{Pregnant})$

In particular

- A cause might produce a given effect with very low probability, and nevertheless could be the most probable cause of that effect, **often the only one!**

In particular

$$P(E \,|\, H) \lll 1 \quad \underline{\text{does not}} \text{ imply} \quad P(H \,|\, E) \lll 1$$

$$\text{and `}\underline{\text{hence}}\text{'} \quad P(\overline{H} \,|\, E) \approx 1$$

$\Rightarrow$ Prosecutor's fallacy

Most events ‘had’ very small probability to occur!

A practical example:

- I shut a picture with my faithful pocket camera.

Most events 'had' very small probability to occur!

A practical example:

- I shut a picture with my faithful pocket camera.
- What is the probability of every configuration of the three RGB codes of the 20MB pixels, given this scene?

$$P(\text{Picture} \equiv \text{X}_{recorded} \,|\, \text{This scene}) \lll 1$$

Most events 'had' very small probability to occur!

A practical example:

- I shut a picture with my faithful pocket camera.
- What is the probability of every configuration of the three RGB codes of the 20MB pixels, given this scene?

$$P(\text{Picture} \equiv \text{X}_{recorded} \,|\, \text{This scene}) \lll 1$$

- But

$$P(\text{This scene} \,|\, \text{Picture}) = 1$$

Most events 'had' very small probability to occur!

A practical example:

- I shut a picture with my faithful pocket camera.
- What is the probability of every configuration of the three RGB codes of the 20MB pixels, given this scene?

$$P(\text{Picture} \equiv \text{X}_{recorded} \,|\, \text{This scene}) \lll 1$$

- But

$$P(\text{This scene} \,|\, \text{Picture}) = 1$$

What else?

An so on...

Most events 'had' very small probability to occur!

Another example, with which we introduce R.

- Extract a number 'a random' integer between 1 and 1 billion (indeed pseudo-random, but it is the same for the purpose)

Most events ‘had’ very small probability to occur!

Another example, with which we introduce R.

- Extract a number ‘a random’ integer between 1 and 1 billion (indeed pseudo-random, but it is the same for the purpose)

```
ceiling( runif(1, 0, 1e9) )
```

Most events ‘had’ very small probability to occur!

Another example, with which we introduce R.

- Extract a number ‘a random’ integer between 1 and 1 billion (indeed pseudo-random, but it is the same for the purpose)

```
ceiling( runif(1, 0, 1e9) )
```

- Each generated number 'had' probability 10^{-9} to occur:

$$P(x \mid H_0) = 10^{-9}$$

with $H_0 =$ “this random number generator”.

Most events ‘had’ very small probability to occur!

Another example, with which we introduce R.

- Extract a number ‘a random’ integer between 1 and 1 billion (indeed pseudo-random, but it is the same for the purpose)

  ```
  ceiling( runif(1, 0, 1e9) )
  ```

- Each generated number 'had' probability 10^{-9} to occur:

  $$P(x\,|\,H_0) = 10^{-9}$$

  with $H_0 =$ “this random number generator”.

- But

  $$P(H_0\,|\,x) \;\; \neq \;\; 10^{-9}$$

Most events 'had' very small probability to occur!

Another example, with which we introduce R.

- Extract a number 'a random' integer between 1 and 1 billion (indeed pseudo-random, but it is the same for the purpose)

```
ceiling( runif(1, 0, 1e9) )
```

- Each generated number 'had' probability 10^{-9} to occur:

$$P(x \,|\, H_0) = 10^{-9}$$

  with H_0 = "this random number generator".

- But

$$\begin{aligned} P(H_0 \,|\, x) &\neq 10^{-9} \\ P(H_0 \,|\, x) &= 1 \end{aligned}$$

Most events 'had' very small probability to occur!

Another example, with which we introduce R.

- Extract a number 'a random' integer between 1 and 1 billion (indeed pseudo-random, but it is the same for the purpose)

```
ceiling( runif(1, 0, 1e9) )
```

- Each generated number 'had' probability 10^{-9} to occur:

$$P(x \,|\, H_0) = 10^{-9}$$

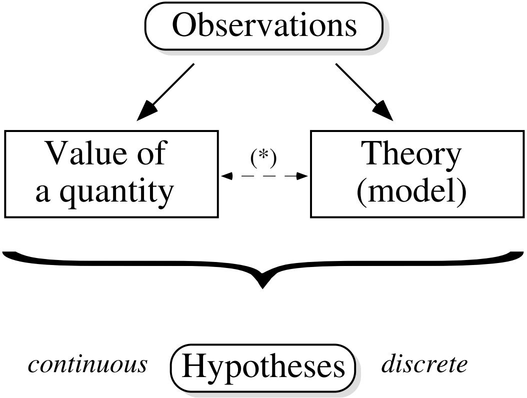
  with H_0 = "this random number generator".

- <u>But</u>

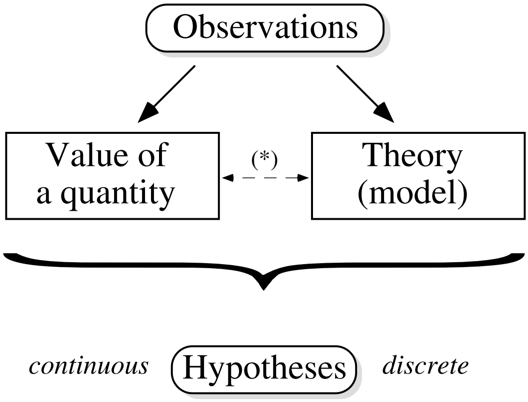
$$\begin{aligned} P(H_0 \,|\, x) &\neq 10^{-9} \\ P(H_0 \,|\, x) &= 1 \end{aligned}$$

What else?

Learning from data

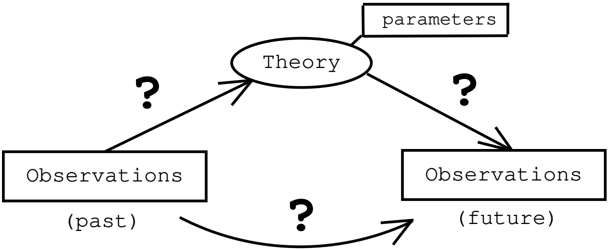


Learning from data



(*) A quantity might be meaningful only within a theory/model

From past to future



Our task:

- Describe/understand the 'physical world'

  $\Rightarrow$ inference of laws ('models') and their parameters $[\Theta]$

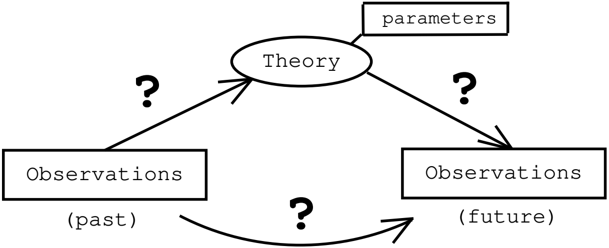
  $\Rightarrow$ $[\Theta \,|\, X_{past}]$

- Predict observations $[X]$

  $\Rightarrow$ forecasting

  $\Rightarrow$ $[X_{future} \,|\, \Theta]$ $\quad\rightarrow\quad$ $[X_{future} \,|\, X_{past}]$

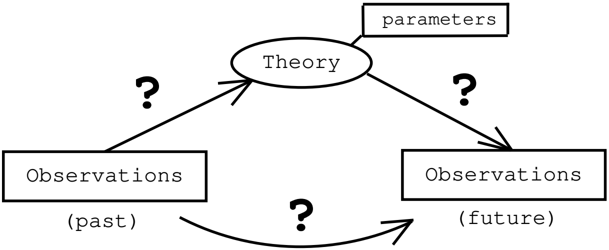
From past to future



Process

- neither automatic
- nor purely contemplative
  - → 'scientific method'
  - → planned experiments ('actions') ⇒ decision.

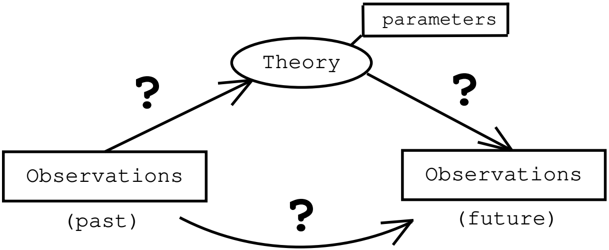
From past to future



$\Rightarrow$ The role of theories/models

- a theory and its parameters are the 'distillate' of all our knowledge about the 'universe' of interest;
- empirical analogical thinking is <u>in most cases not usable</u>:
  - A theory can predict effects never observed before
  - Example: shooting a bullet

From past to future



⇒ The role of theories/models

- a theory and its parameters are the 'distillate' of all our knowledge about the 'universe' of interest;
- empirical analogical thinking is in most cases not usable:

"La cognizione d'un solo effetto acquistata per le sue cause ci apre l'intelletto a 'ntendere ed assicurarci d'altri effetti senza bisogno di ricorrere alle esperienze" (Galileo)

Model thinking

"The scientific method is based on repeated experiments"
(or some like that)

Model thinking

"The scientific method is based on repeated experiments"
(or some like that)

- Are then astrophysics, cosmology etc. Science?

Model thinking

"The scientific method is based on repeated experiments"
(or some like that)

- Are then astrophysics, cosmology etc. Science?
- Who told so?

Model thinking

"The scientific method is based on repeated experiments"
(or some like that)

- Are then astrophysics, cosmology etc. Science?
- Who told so?
  … Galileo…

Model thinking

"The scientific method is based on repeated experiments"
(or some like that)

- Are then astrophysics, cosmology etc. Science?
- Who told so?
  ... Galileo...
  "The knowledge of a single effect acquired by its causes open our mind to understand and ensure us of other effects without the need of making experiments"
  (Galileo)

Model thinking

"The scientific method is based on repeated experiments"
(or some like that)

- Are then astrophysics, cosmology etc. Science?
- Who told so?
  ... Galileo...
  "The knowledge of a single effect acquired by its causes open our mind to understand and ensure us of other effects without the need of making experiments"
  (Galileo)

What really matters is to have a Model which links parameters to observations

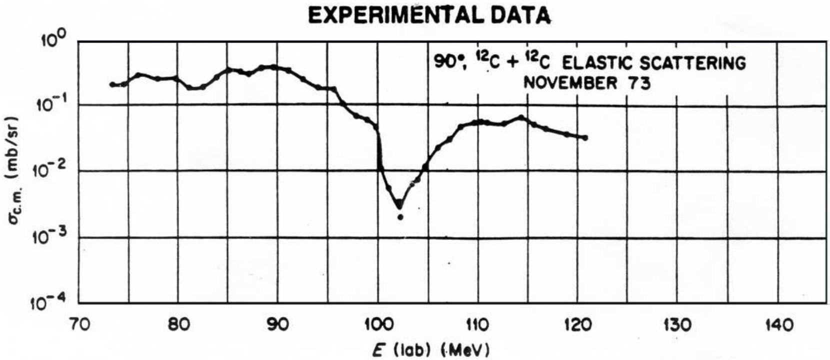
Model thinking

"The scientific method is based on repeated experiments"
(or some like that)

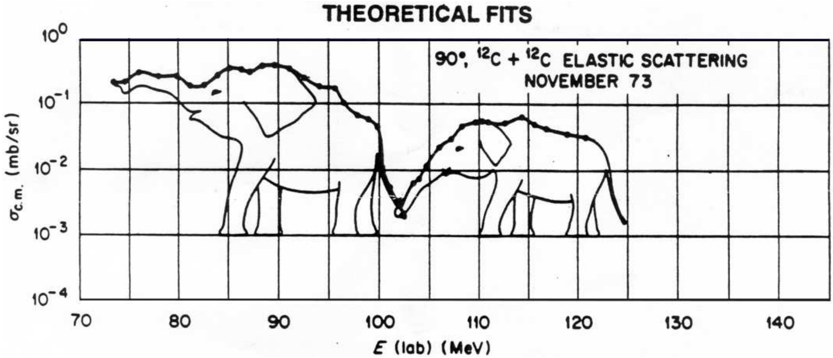
- Are then astrophysics, cosmology etc. Science?
- Who told so?
  … Galileo…
  "The knowledge of a single effect acquired by its causes open our mind to understand and ensure us of other effects without the need of making experiments"
  (Galileo)

What really matters is to have a Model which links parameters to observations
But remind that "all models are wrong, some are useful"…

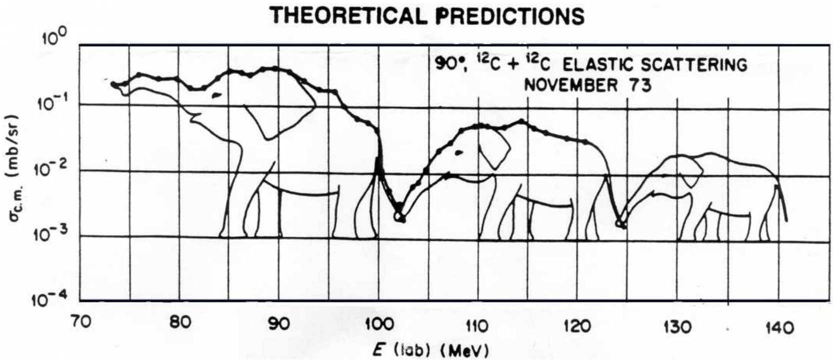
Inferential-predictive process



Inferential-predictive process

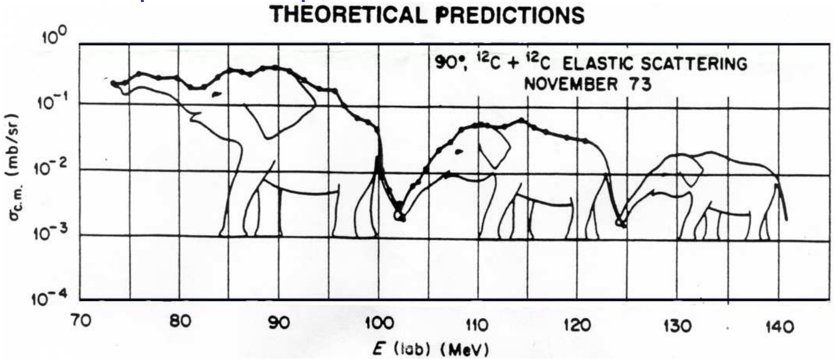


Inferential-predictive process



(S. Raman, *Science with a smile*)

Inferential-predictive process



(S. Raman, *Science with a smile*)

Even if the (*ad hoc*) model fits perfectly the data,
we do not believe the predictions
because we don't trust the model!

[Many 'good' models are *ad hoc* models!]

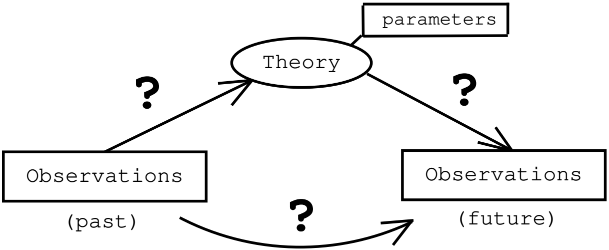
2011 IgNobel prize in Mathematics

- D. Martin of USA (who predicted the world would end in 1954)
- P. Robertson of USA (who predicted the world would end in 1982)
- E. Clare Prophet of the USA (who predicted the world would end in 1990)
- L.J. Rim of KOREA (who predicted the world would end in 1992)
- C. Mwerinde of UGANDA (who predicted the world would end in 1999)
- H. Camping of the USA (who predicted the world would end on September 6, 1994 and later predicted that the world will end on October 21, 2011)

2011 IgNobel prize in Mathematics

"For teaching the world to be careful when making mathematical assumptions and calculations"

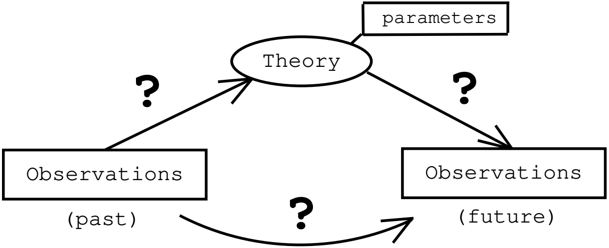
Uncertainty



⇒ Uncertainty:

1. Given the past observations, in general we are not sure about the parameters of the model (and/or the model itself)
2. Even if we were sure about theory and parameters, there could be internal ("noise", variables out of our control) or external effects (initial/boundary conditions, 'errors', etc) that make the forecasting uncertain.

Uncertainty

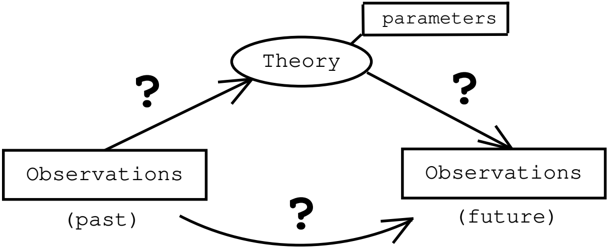


⇒ Uncertainty:

- No certainties, only probabilities

- $P(\Theta \,|\, X_{past})$
- $P(X_{future} \,|\, \Theta)$
- $P(X_{future} \,|\, X_{past})$

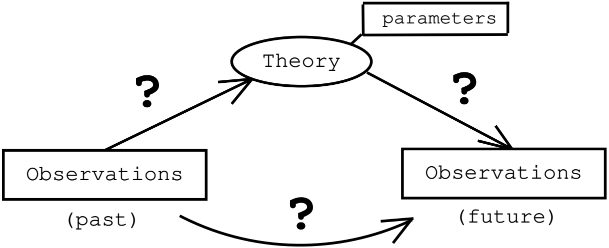
Deep source of uncertainty



Uncertainty:

Theory	— ? ⟶	Future observations
Past observations	— ? ⟶	Theory
Theory	— ? ⟶	Future observations

Deep source of uncertainty



Uncertainty:

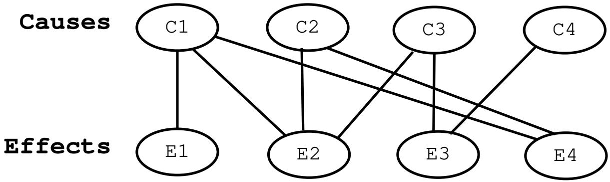
Theory	— ? $\longrightarrow$	Future observations
Past observations	— ? $\longrightarrow$	Theory
Theory	— ? $\longrightarrow$	Future observations

$\Longrightarrow$ **Uncertainty about causal connections**

CAUSE $\Longleftrightarrow$ **EFFECT**

Causes $\rightarrow$ effects

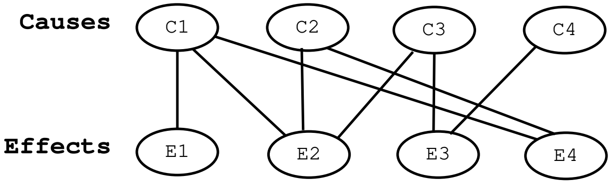
The same *apparent* cause might produce several,different effects



Given an observed effect, we are not sure about the exact cause that has produced it.

Causes → effects

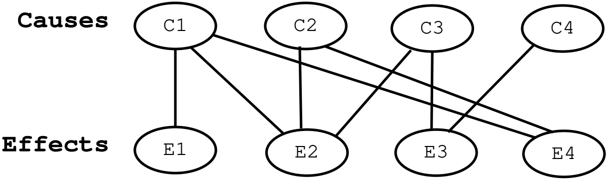
The same *apparent* cause might produce several,different effects



Given an observed effect, we are not sure about the exact cause that has produced it.

Causes $\rightarrow$ effects

The same *apparent* cause might produce several,different effects



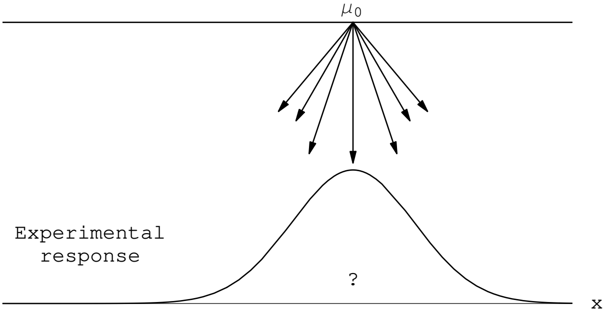
Given an observed effect, we are not sure about the exact cause that has produced it.

$$E_2 \Rightarrow \{C_1,\ C_2,\ C_3\}?$$

$\rightarrow$ Probability of causes

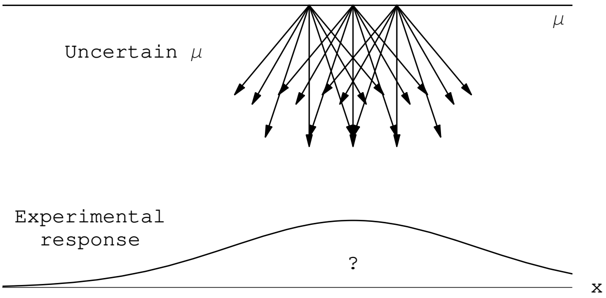
"the essential problem of the experimental method"

From 'true value' to observations



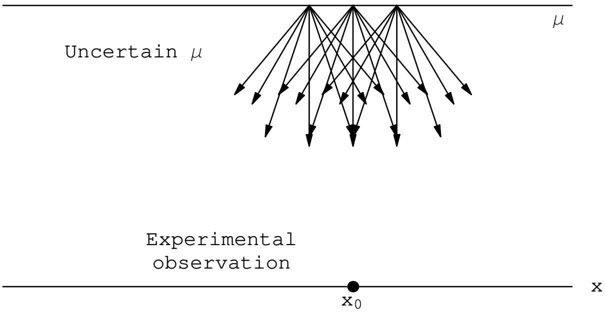
Given μ (exactly known) we are uncertain about x

From ‘true value’ to observations



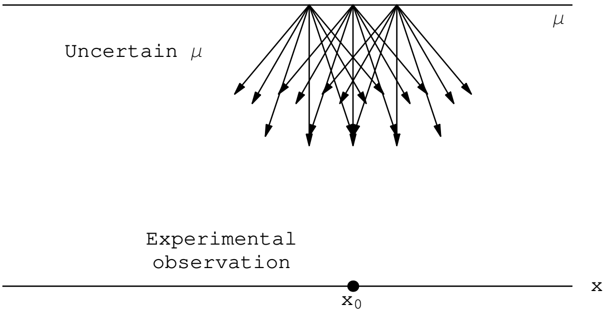
Uncertainty about μ makes us more uncertain about x

. . . and back: Inferring a true value



The observed data is <u>certain</u>: $\rightarrow$ 'true value' uncertain.

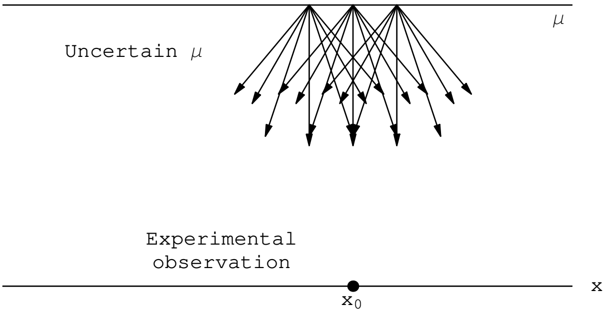
. . . and back: Inferring a true value



The observed data is certain: $\rightarrow$ 'true value' uncertain.

"data uncertainty" ?

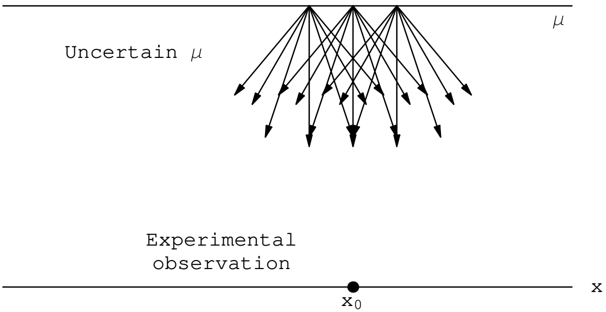
…and back: Inferring a true value



The observed data is <u>certain</u>: $\rightarrow$ ‘true value’ uncertain.

“data uncertainty” ? Data corrupted?

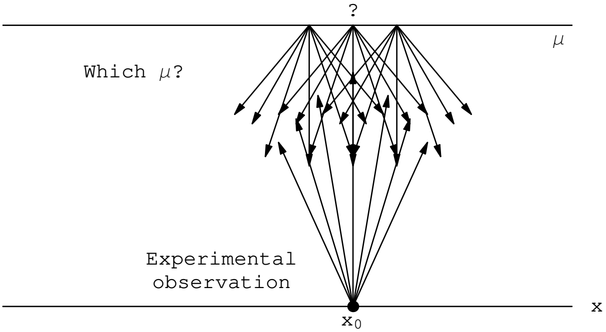
...and back: Inferring a true value



The observed data is <u>certain</u>: $\rightarrow$ 'true value' uncertain.

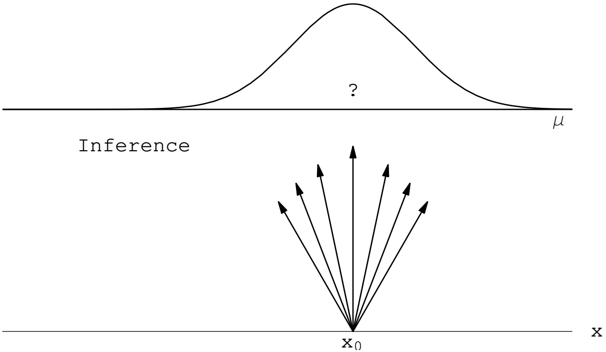
"data uncertainty" ? Data corrupted?
Even if the data were corrupted, the <u>data</u> were the corrupted data**!!**...

. . . and back: Inferring a true value



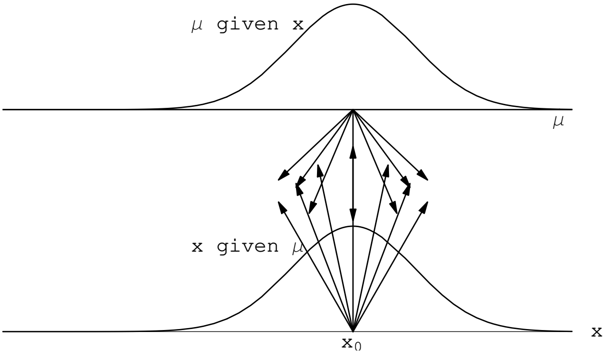
Where does the observed value of x comes from?

. . . and back: Inferring a true value



We are now uncertain about μ, given x.

. . . and back: Inferring a true value



Note the symmetry in reasoning.

A very simple experiment

Let's make an experiment

A very simple experiment

Let's make an experiment

- Here
- Now

A very simple experiment

Let's make an experiment

- Here
- Now

For simplicity

- μ can assume only six possibilities:

$$0, 1, \ldots, 5$$

- x is binary:

$$0, 1$$

$$[\,(1,2);\ \text{Black/White};\ \text{Yes/Not};\ \ldots\,]$$

A very simple experiment

Let's make an experiment

- Here
- Now

For simplicity

- μ can assume only six possibilities:

$$0, 1, \ldots, 5$$

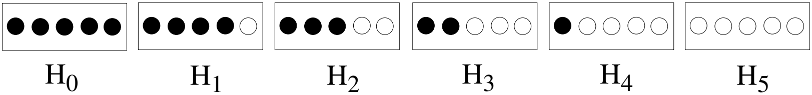
- x is binary:

$$0, 1$$

$$[\,(1, 2);\ \text{Black/White};\ \text{Yes/Not};\ \ldots\,]$$

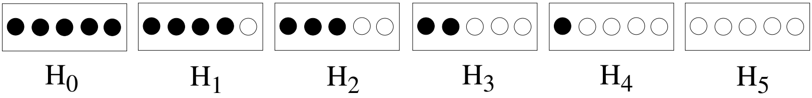
$\Rightarrow$ Later we shall make μ continuous.

Which box? Which ball?



Let us take randomly one of the boxes.

Which box? Which ball?



Let us take randomly one of the boxes.

We are in a state of uncertainty concerning several *events*, the most important of which correspond to the following questions:

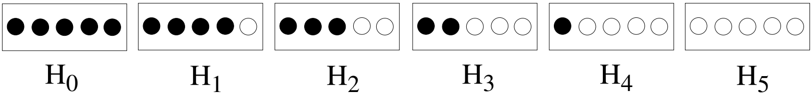
(a) Which box have we chosen, H_0, H_1, ..., H_5?

(b) If we extract randomly a ball from the chosen box, will we observe a white ($E_W \equiv E_1$) or black ($E_B \equiv E_2$) ball?

Our certainties:

$$\begin{aligned} \cup_{j=0}^{5} H_j &= \Omega \\ \cup_{i=1}^{2} E_i &= \Omega\,. \end{aligned}$$

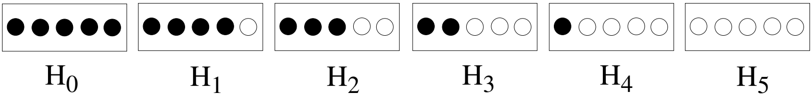
Which box? Which ball?



Let us take randomly one of the boxes.

- What happens after we have extracted one ball and looked its color?
  - Intuitively feel *how to* roughly *change* our opinion about
    - the possible cause
    - a future observation

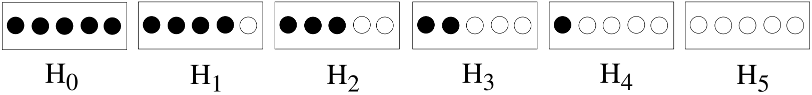
Which box? Which ball?



Let us take randomly one of the boxes.

- What happens after we have extracted one ball and looked its color?
  - Intuitively feel *how to* roughly *change* our opinion about
    - the possible cause
    - a future observation
  - Can we do it *quantitatively*, in an 'objective way'?

Which box? Which ball?



Let us take randomly one of the boxes.

- What happens after we have extracted one ball and looked its color?
  - Intuitively feel *how to* roughly *change* our opinion about
    - the possible cause
    - a future observation
  - Can we do it *quantitatively*, in an 'objective way'?
- And after a sequence of extractions?

The toy inferential experiment

The aim of the experiment will be to guess the content of the box without looking inside it, only extracting a ball, record its color and reintroducing in the box

The toy inferential experiment

The aim of the experiment will be to guess the content of the box without looking inside it, only extracting a ball, record its color and reintroducing in the box

This toy experiment is conceptually very close to what we do in the pure and applied sciences

- ⇒ try to guess what we cannot see (the electron mass, a magnetic field, etc)

  . . . from what we can see (somehow) with our senses.

The rule of the game is that we are not allowed to watch inside the box! (As we cannot open and electron and read its properties, unlike we read the MAC address of a PC interface.)

Where is probability?

We all agree that the experimental results change

- the probabilities of the box compositions;
- the probabilities of a future outcomes,

Where is probability?

We all agree that the experimental results change

- the probabilities of the box compositions;
- the probabilities of a future outcomes,

although the box composition remains unchanged ('extractions followed by re-introduction').

Where is probability?

We all agree that the experimental results change

- the probabilities of the box compositions;
- the probabilities of a future outcomes,

although the box composition remains unchanged ('extractions followed by re-introduction').

Where is probability?

Where is probability?

We all agree that the experimental results change

- the probabilities of the box compositions;
- the probabilities of a future outcomes,

although the box composition remains unchanged ('extractions followed by re-introduction').

Where is probability?

Certainly not *in* the box!

The End