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

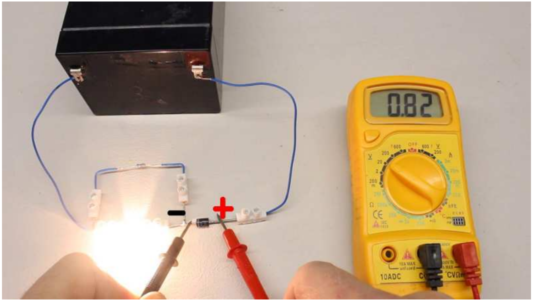
"As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality. "
(A. Einstein)

"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)

What is measurement?

What is measurement?

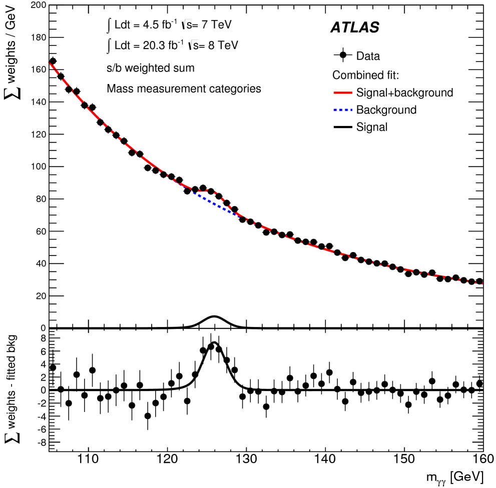


What is measurement?

What is measurement?

What is measurement?

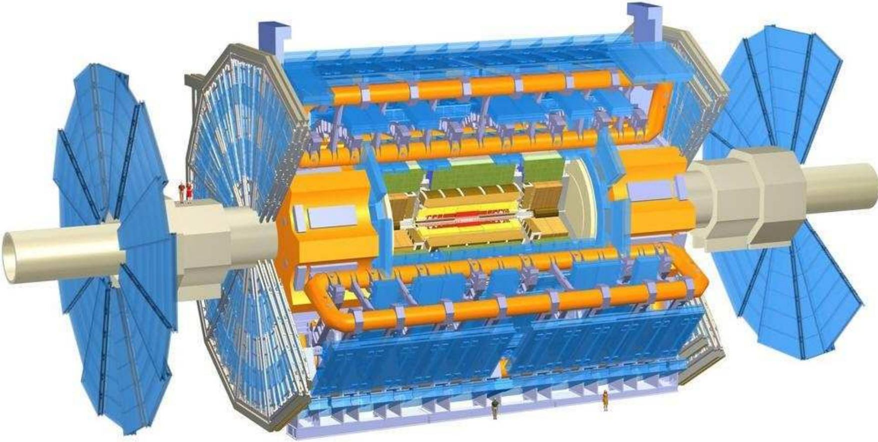
Higgs $\to \gamma\gamma$ (2012)



Two-photon *invariant mass*

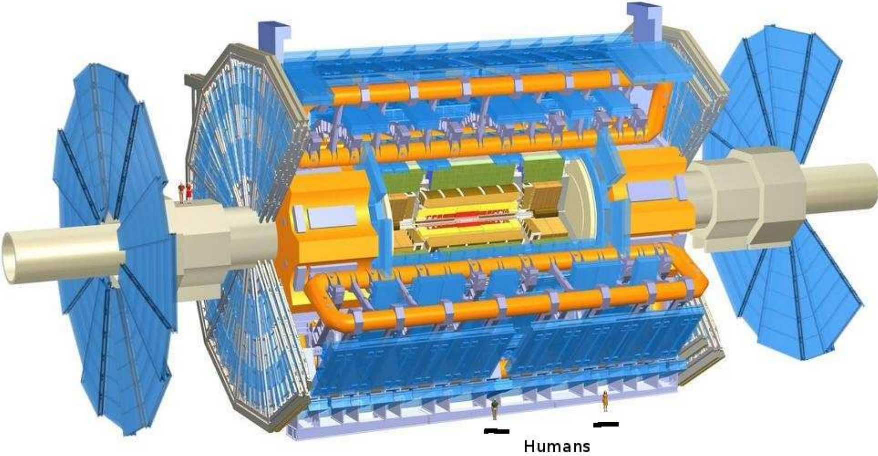
What is measurement?

ATLAS Experiment at LHC (CERN, Geneva)



What is measurement?

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

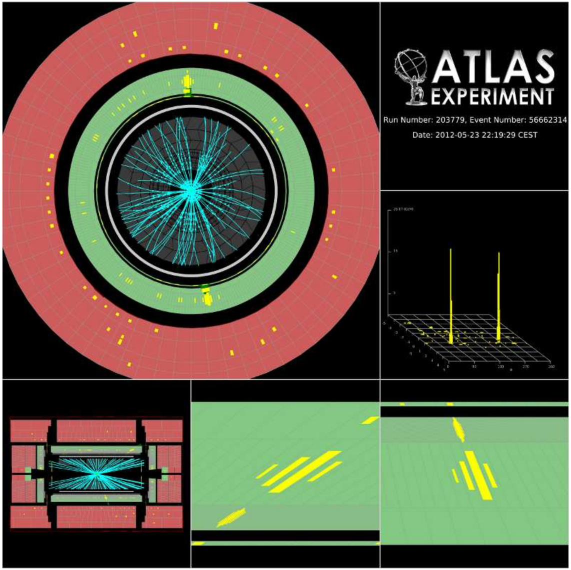


≈ 3000 km cables

≈ 7000 tonnes

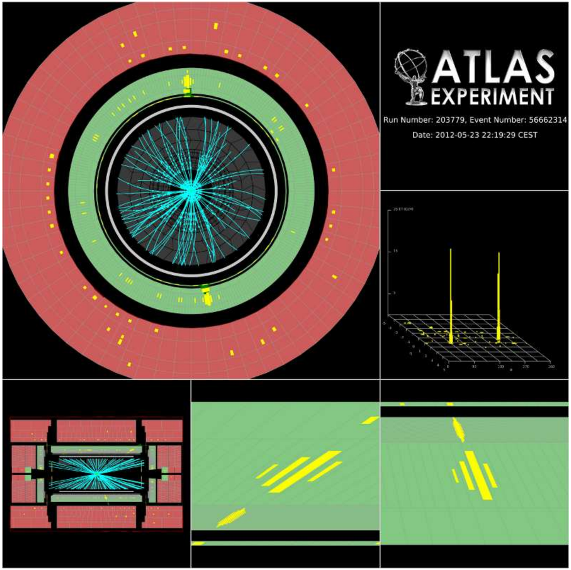
≈ 100 millions electronic channels

What is measurement?



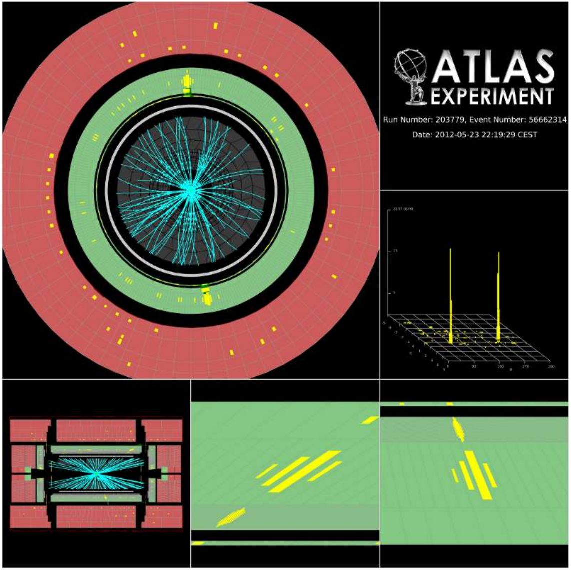
Two flashes of ‘light’ (2 γ’s) in a ‘noisy’ environment.

What is measurement?



Two flashes of 'light' (2 γ's) in a 'noisy' environment.
Higgs $\to \gamma\gamma$?

What is measurement?

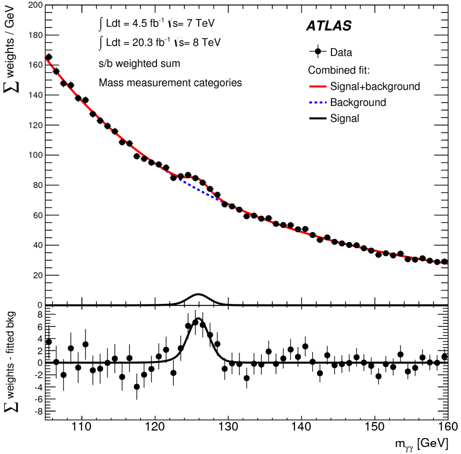


Two flashes of ‘light’ (2 γ’s) in a ‘noisy’ environment.

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

What is measurement?

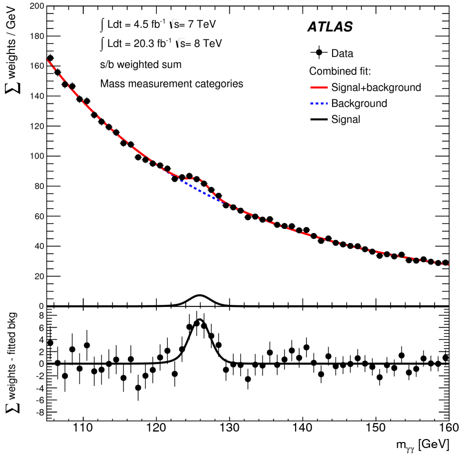
Higgs $\to \gamma\gamma$



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

What is measurement?

Higgs $\to \gamma\gamma$

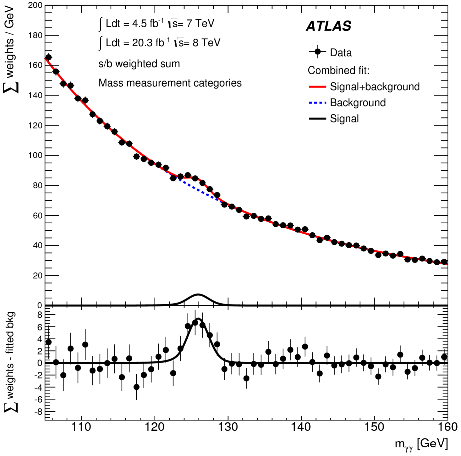


$\Rightarrow \begin{cases} \text{Mass value} \\ \text{Production rate} \end{cases}$

(with uncertainties)

What is measurement?

Higgs $\to \gamma\gamma$



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

(with uncertainties)

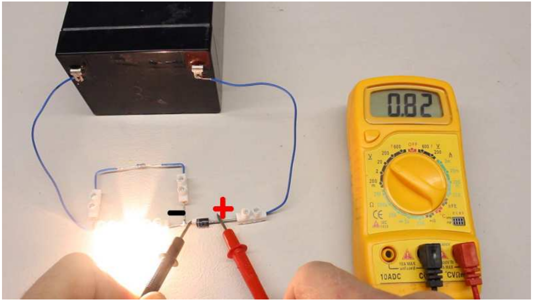
Quite indirect measurements of something we do not "see"!

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:

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

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

Measuring a mass on a scale

Equilibrium:

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

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

From the reading to the value of the mass:

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

Measuring a mass on a balance

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

Measuring a mass on a balance

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

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

Measuring a mass on a balance

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

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

- Position is usually <u>not</u> at "$R_{\oplus}$" from the Earth center;

Measuring a mass on a balance

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

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

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

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

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

Measuring a mass on a balance

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

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

- Position is usually <u>not</u> at “$R_{♁}$” 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

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

Dependence on 'k':

- temperature
- non linearity
- ...

Measuring a mass on a balance

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

Dependence on 'k':

- temperature
- non linearity
- ...

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

- left to your imagination...

Measuring a mass on a balance

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

Dependence on 'k':

- temperature
- non linearity
- ...

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

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

Dependence on 'k':

- temperature
- non linearity
- ...

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

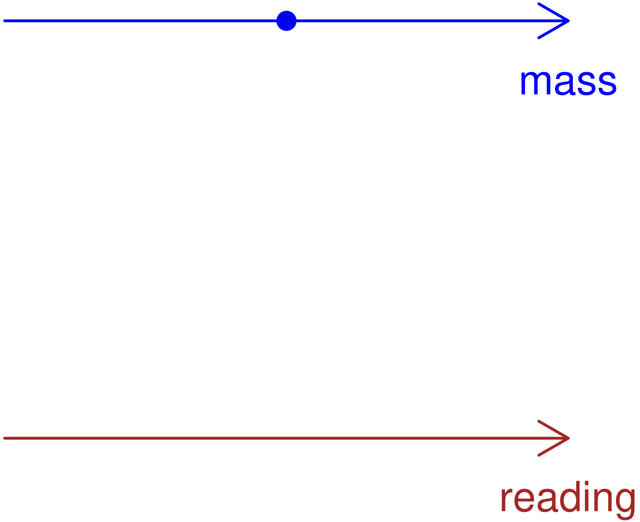
- left to your imagination...

$+$ **randomic effects:**

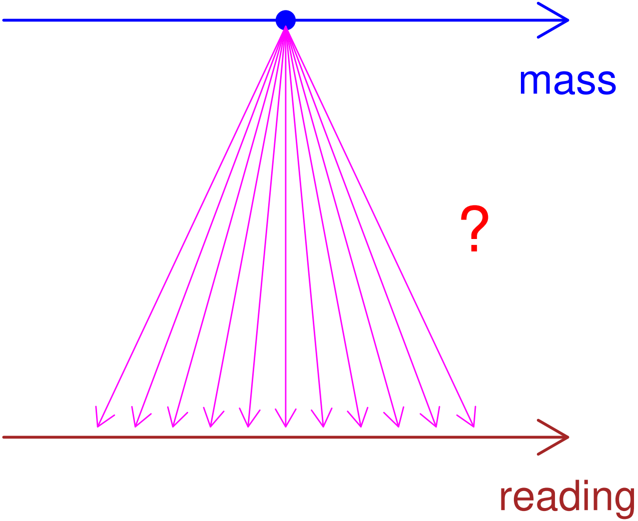
- stopping position of damped oscillation;
- variability of all quantities of influence (in the ISO-GUM sense);
- reading of analog scale.

$$\Rightarrow \ m??$$

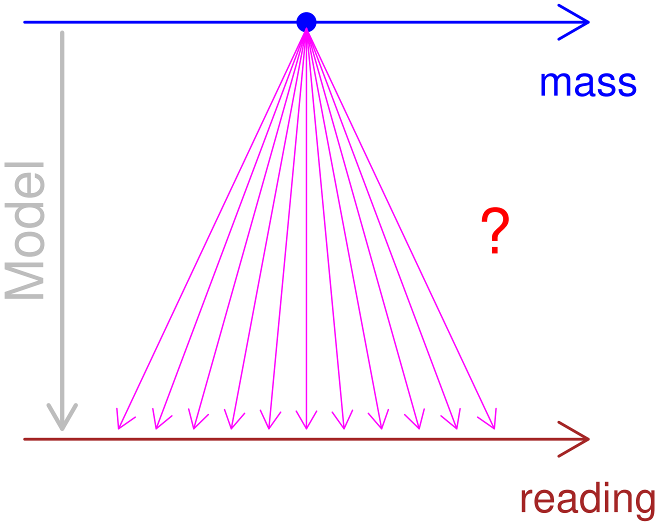
Mass $\longrightarrow$ Reading



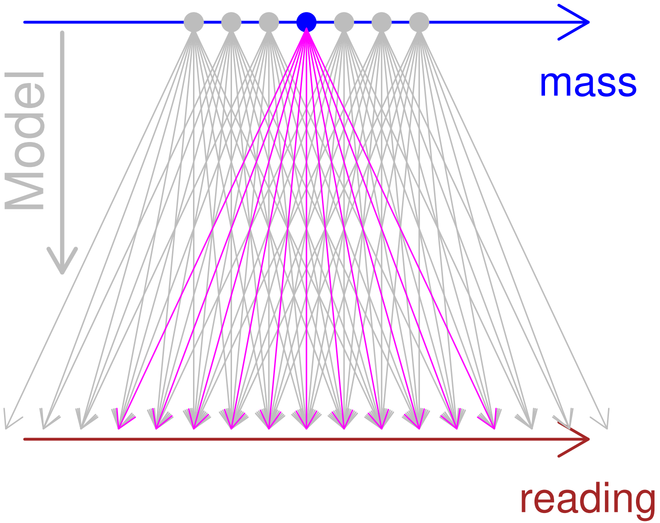
Mass $\longrightarrow$ Reading



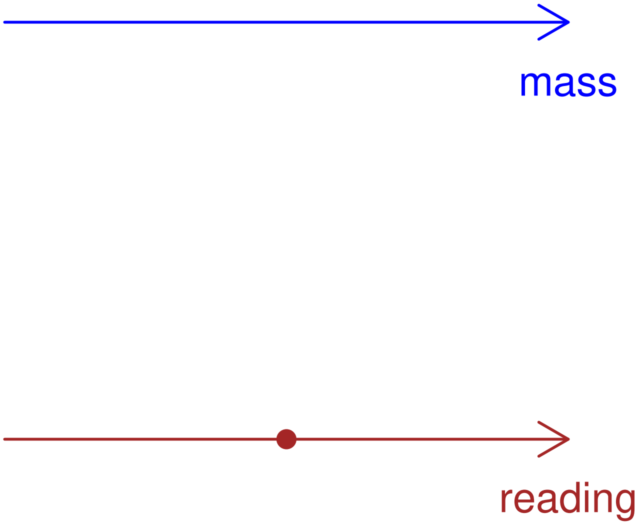
Mass $\longrightarrow$ Reading



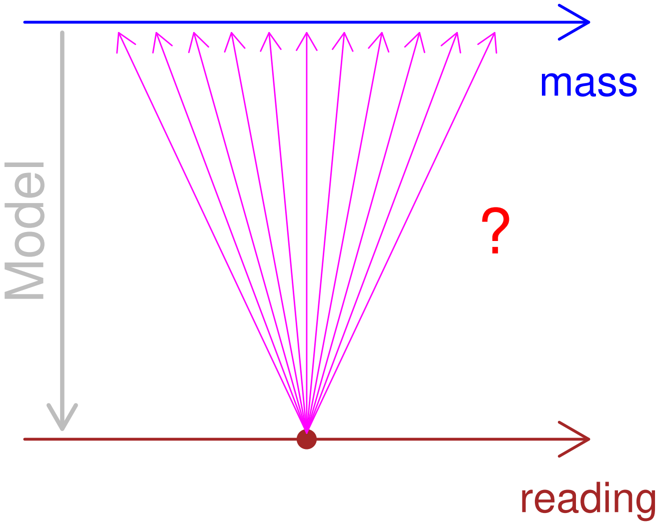
Mass $\longrightarrow$ Reading



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*
   - $\rightarrow$ *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?

Sources of uncertainties (from ISO GUM)

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

Sources of uncertainties (from ISO GUM)

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

Sources of uncertainties (from ISO GUM)

1. *incomplete definition of the measurand*
   - $\rightarrow$ 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;*
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!

Observation $\rightarrow$ value of a quantity

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

Observations $\rightarrow$ hypotheses

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

Observations → hypotheses

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

- Experimental observation ('data') → responsible cause.

Observations $\rightarrow$ hypotheses

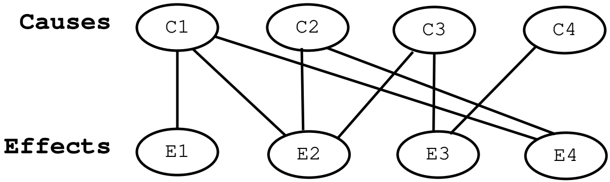
This problem occurs not only “determining” *the* value of a physical quantity.

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

(But logically no substantial difference.)

Causes → effects

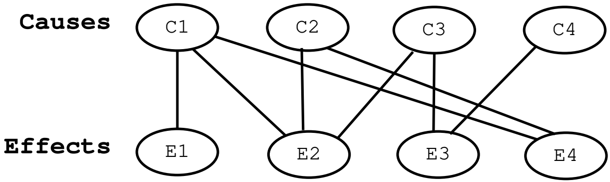
The same *apparent* cause might produce several,different effects



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

Causes → effects

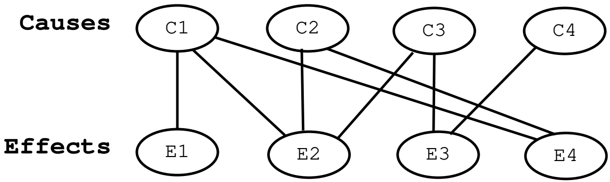
The same *apparent* cause might produce several,different effects



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.

$$\mathrm{E}_2 \Rightarrow \{C_1,\ C_2,\ C_3\}?$$

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*)

> "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?

A minimalist though not trival problem of the kind

An example easy to understand:

- two causes;
- two effects;

A minimalist though not trival problem of the kind

An example easy to understand:

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

A minimalist though not trival problem of the kind

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

A minimalist though not trival problem of the kind

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.
→ 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) E_1 = Positive

H_2='$\overline{\text{HIV}}$' (Not infected) 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:

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

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

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

AIDS test

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

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

$$
\begin{array}{lcl}
?\ H_1 = \text{'HIV' (Infected)} & \longleftarrow & E_1 = \text{Positive} \\
?\ H_2 = \text{'}\overline{\text{HIV}}\text{' (Not infected)} & \swarrow & E_2 = \text{Negative}
\end{array}
$$

Result: $\Rightarrow$ Positive

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 follow!

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

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

Pay attention not to arbitrary revert conditional probabilities:

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

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

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

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]

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

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]
- $P(\text{Pregnant} \mid \text{Woman}) \neq P(\text{Woman} \mid \text{Pregnant})$

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

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!**

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

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

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

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

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

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

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

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

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

$P(E \mid H) \lll 1$ does not imply $P(H \mid E) \lll 1$

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

$\Rightarrow$ Prosecutor's fallacy

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

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

$P(E \mid H) \lll 1$ does not imply $P(H \mid E) \lll 1$

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

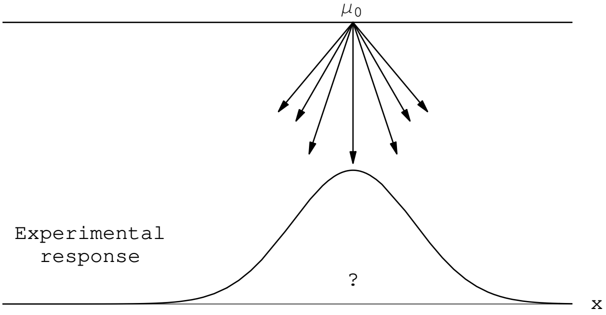
$\Rightarrow$ Prosecutor's fallacy

$\Rightarrow$ Misunderstanding **p-values** (a *related* logical mistake)

$\rightarrow$ Probability of causes

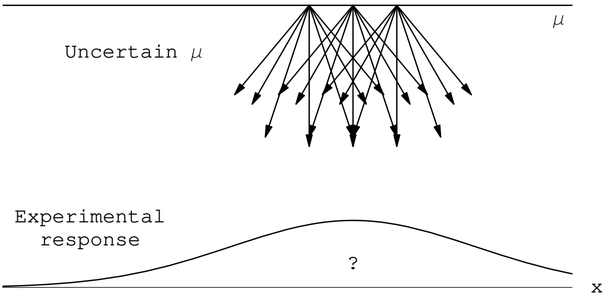
"the essential problem of the experimental method"

From 'true value' to observations



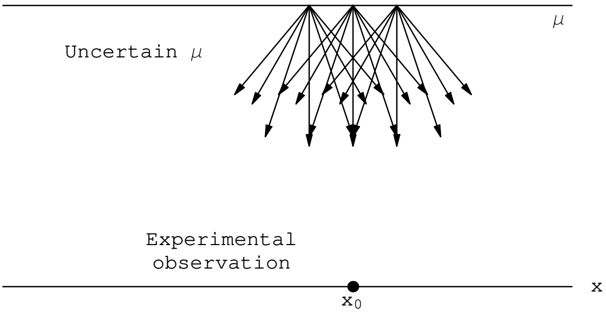
Given μ (exactly known) we are uncertain about x

From 'true value' to observations



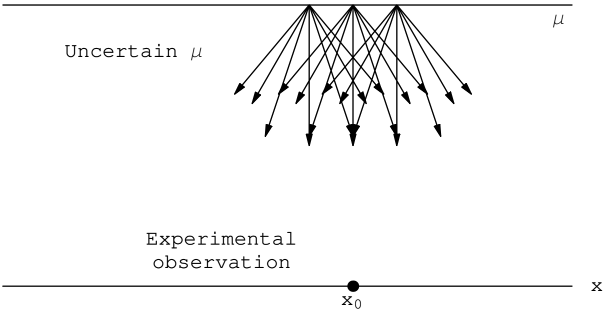
Uncertainty about μ makes us more uncertain about x

. . . and back: Inferring a true value



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

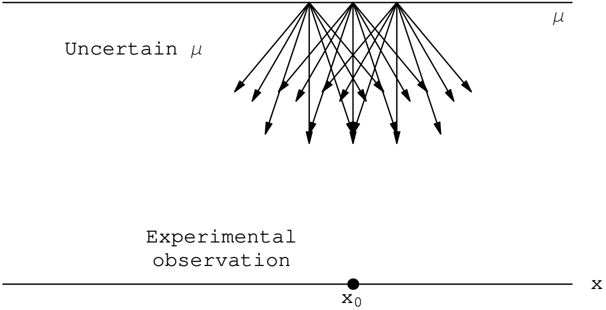
. . . and back: Inferring a true value



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

"data uncertainty" ?

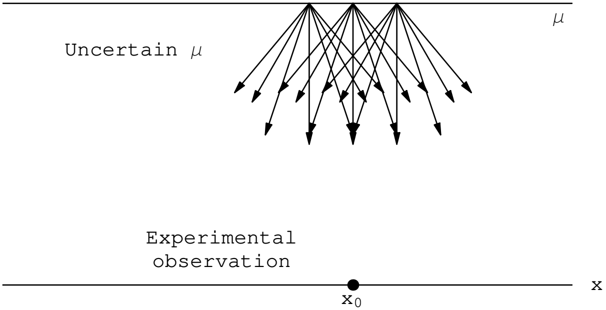
. . . and back: Inferring a true value



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

"data uncertainty" ? Data corrupted?

. . . and back: Inferring a true value

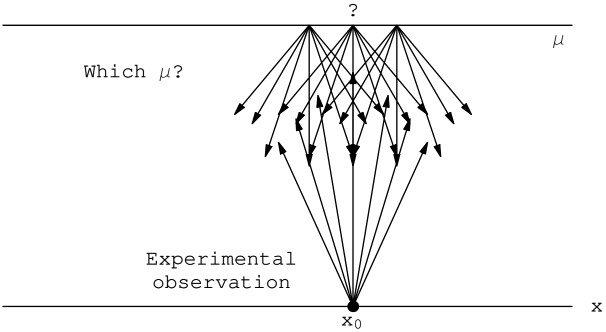


The observed data is certain: $\rightarrow$ 'true value' uncertain.

"data uncertainty" ? Data corrupted?

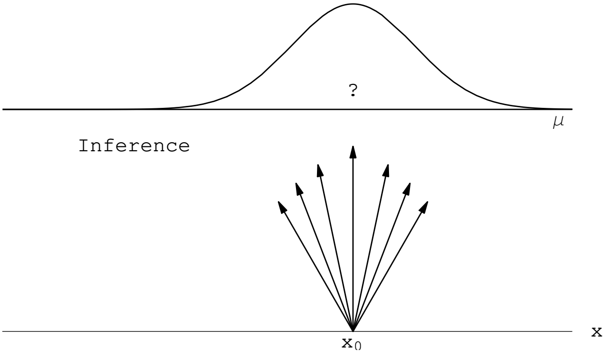
Even if the data were corrupted, the data were the corrupted data**!!** . . .

. . . and back: Inferring a true value



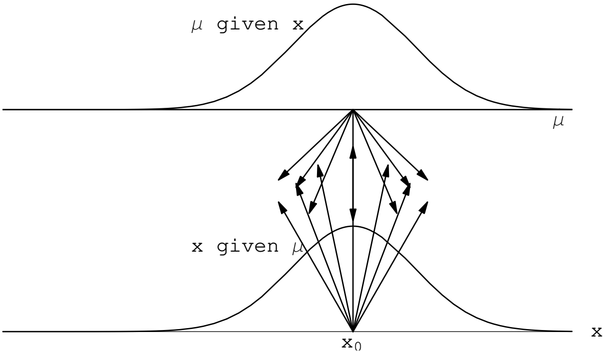
Where does the observed value of x comes from?

. . . 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.

Basic rules of probability

1. $0 \leq P(A \,|\, I) \leq 1$
2. $P(\Omega \,|\, I) = 1$
3. $P(A \cup B \,|\, I) = P(A \,|\, I) + P(B \,|\, I) \quad [\text{if } P(A \cap B \,|\, I) = \emptyset]$
4. $P(A \cap B \,|\, I) = P(A \,|\, B, I) \cdot P(B \,|\, I) = P(B \,|\, A, I) \cdot P(A \,|\, I)$

Remember that probability is always conditional probability!

I is the background condition (related to information 'I_s')

$\rightarrow$ usually implicit (we only care about 're-conditioning')

Basic rules of probability

1. $0 \leq P(A \,|\, I) \leq 1$
2. $P(\Omega \,|\, I) = 1$
3. $P(A \cup B \,|\, I) = P(A \,|\, I) + P(B \,|\, I) \quad [\text{if } P(A \cap B \,|\, I) = \emptyset]$

4. $P(A \cap B \,|\, I) = P(A \,|\, B, I) \cdot P(B \,|\, I) = P(B \,|\, A, I) \cdot P(A \,|\, I)$

Remember that probability is always conditional probability!

I is the background condition (related to information 'I_s'')

$\rightarrow$ usually implicit (we only care about 're-conditioning')

Note: 4. does not define conditional probability.
(Probability is always conditional probability!)

Basic rules of probability

1. $0 \leq P(A \mid I) \leq 1$
2. $P(\Omega \mid I) = 1$
3. $P(A \cup B \mid I) = P(A \mid I) + P(B \mid I) \quad [\text{if } P(A \cap B \mid I) = \emptyset]$

4. $P(A \cap B \mid I) = P(A \mid B, I) \cdot P(B \mid I) = P(B \mid A, I) \cdot P(A \mid I)$

Remember that probability is always conditional probability!

I is the background condition (related to information 'I_s')

→ usually implicit (we only care about 're-conditioning')

Note: 4. does not define conditional probability.
(Probability is always conditional probability!)

⇒ easily extended to uncertain numbers ('*random variables*')

Subjective nature of probability

"Since the knowledge may be different with different persons

Subjective nature of probability

"Since the knowledge may be different with different persons or with the same person at different times,

Subjective nature of probability

"Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence,

Subjective nature of probability

"Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence, and thus different numerical probabilities may be attached to the same event"

Subjective nature of probability

"Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence, and thus different numerical probabilities may be attached to the same event"

(Schrödinger, 1947)

Subjective nature of probability

"Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence, and thus different numerical probabilities may be attached to the same event"

(Schrödinger, 1947)

Probability depends on the status of information of the *subject* who evaluates it.

Probability is always conditional probability

"Thus whenever we speak loosely of 'the probability of an event', it is always to be understood: probability with regard to a certain given state of knowledge"

Probability is always conditional probability

"Thus whenever we speak loosely of 'the probability of an event', it is always to be understood: probability with regard to a certain given state of knowledge"

(Schrödinger, 1947)

Probability is always conditional probability

"Thus whenever we speak loosely of 'the probability of an event', it is always to be understood: probability with regard to a certain given state of knowledge"

(Schrödinger, 1947)

$$P(E) \quad \longrightarrow \quad P(E \,|\, I_s(t))$$

where $I_s(t)$ is the information available to *subject* s at time t.

Mathematics of beliefs

An even better news:

The fourth basic rule
can be fully exploited!

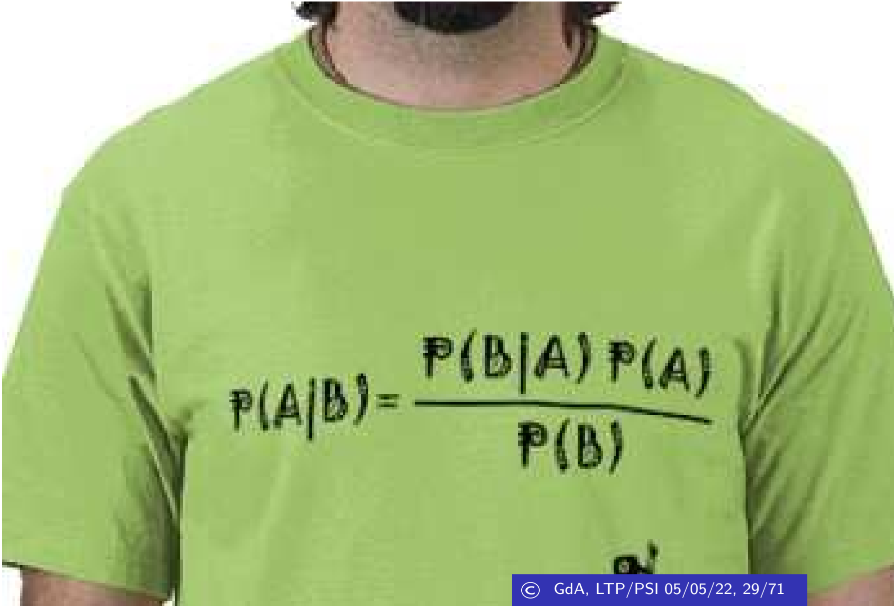
Mathematics of beliefs

An even better news:

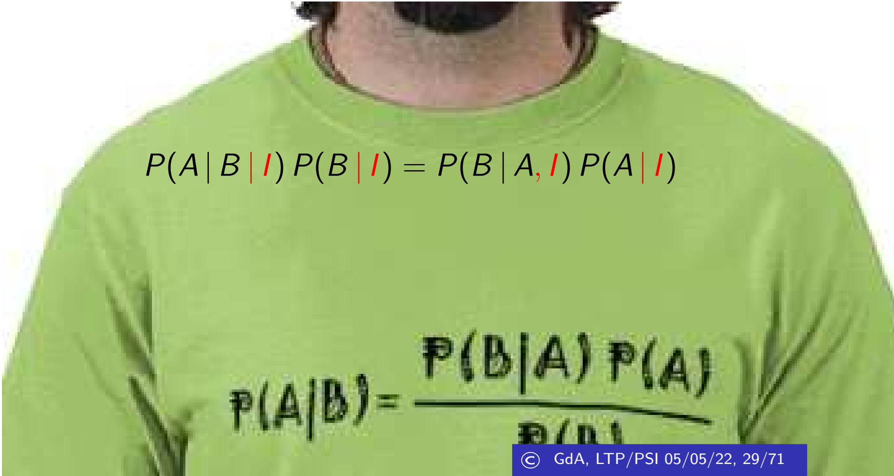
The fourth basic rule
can be fully exploited!

(Liberated by a curious ideology that forbids its use)

A simple, powerful formula

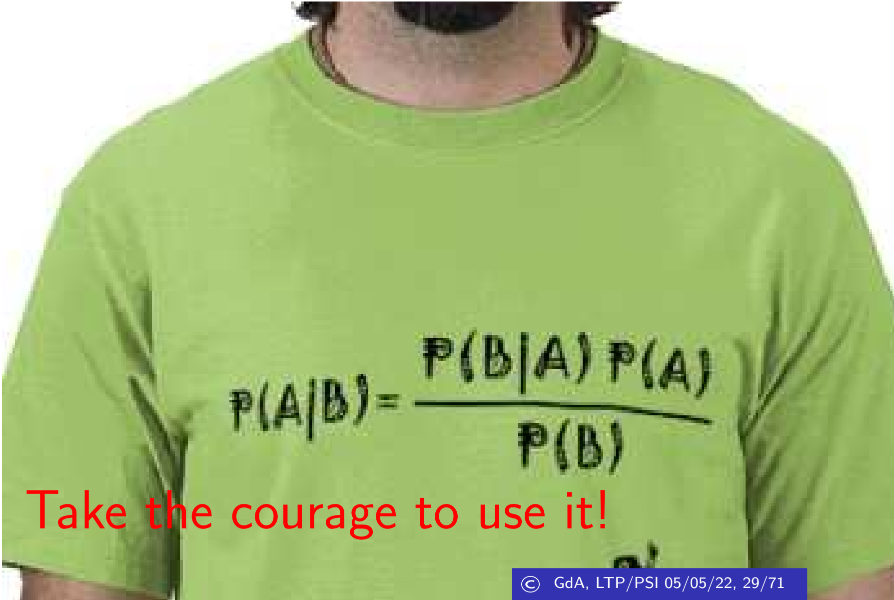


A simple, powerful formula



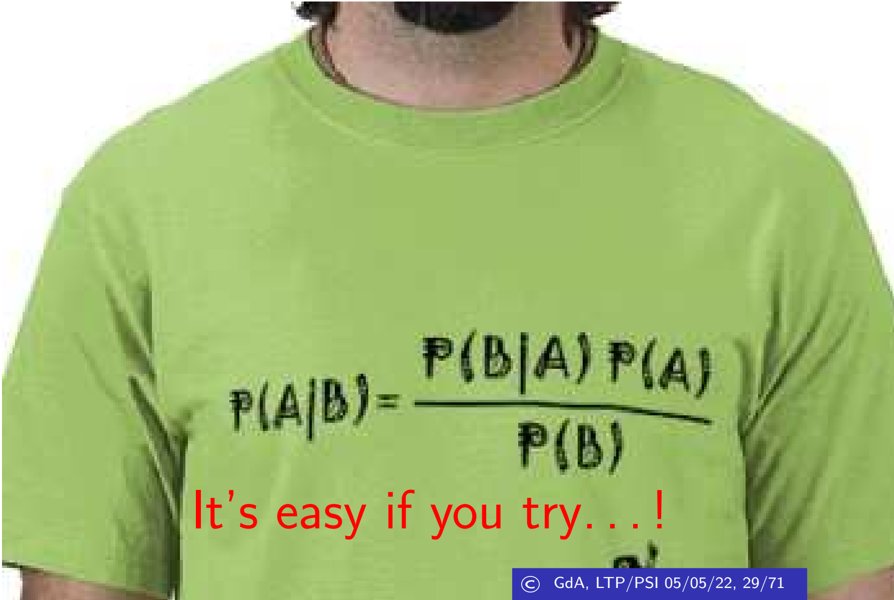
$$P(A \mid B \mid I)\, P(B \mid I) = P(B \mid A, I)\, P(A \mid I)$$

A simple, powerful formula



Take the courage to use it!

A simple, powerful formula



It's easy if you try. . . !

A simple, powerful formula

[Bayes Theorem]

Laplace's "Bayes Theorem"

"The greater the probability of an observed event given any one of a number of causes to which that event may be attributed, the greater the likelihood of that cause {given that event}.

$$P(C_i \mid E) \propto P(E \mid C_i)$$

Laplace's "Bayes Theorem"

"The greater the probability of an observed event given any one of a number of causes to which that event may be attributed, the greater the likelihood of that cause {given that event}. The probability of the existence of any one of these causes {given the event} is thus a fraction whose numerator is the probability of the event given the cause, and whose denominator is the sum of similar probabilities, summed over all causes.

$$P(C_i \mid E) = \frac{P(E \mid C_i)}{\sum_j P(E \mid C_j)}$$

Laplace's "Bayes Theorem"

"The greater the probability of an observed event given any one of a number of causes to which that event may be attributed, the greater the likelihood of that cause {given that event}. The probability of the existence of any one of these causes {given the event} is thus a fraction whose numerator is the probability of the event given the cause, and whose denominator is the sum of similar probabilities, summed over all causes. If the various causes are not equally probable *a priory*, it is necessary, instead of the probability of the event given each cause, to use the product of this probability and the *possibility of the cause itself*."

$$P(C_i \mid E) = \frac{P(E \mid C_i)\, P(C_i)}{\sum_j P(E \mid C_j)\, P(C_j)}$$

Laplace's "Bayes Theorem"

"The greater the probability of an observed event given any one of a number of causes to which that event may be attributed, the greater the likelihood of that cause {given that event}. The probability of the existence of any one of these causes {given the event} is thus a fraction whose numerator is the probability of the event given the cause, and whose denominator is the sum of similar probabilities, summed over all causes. If the various causes are not equally probable *a priory*, it is necessary, instead of the probability of the event given each cause, to use the product of this probability and the *possibility of the cause itself*."

$$P(C_i \,|\, E) = \frac{P(E \,|\, C_i)\, P(C_i)}{P(E)}$$

(*Philosophical Essai on Probabilities*)

[In general $P(E) = \sum_j P(E \,|\, C_j)\, P(C_j)$ (*weighted average*, with weigths being the probabilities of the conditions) if C_j form a *complete class of hypotheses*]

Laplace's "Bayes Theorem"

$$P(C_i \,|\, E) \;=\; \frac{P(E \,|\, C_i)\, P(C_i)}{P(E)} = \frac{P(E \,|\, C_i)\, P(C_i)}{\sum_j P(E \,|\, C_j)\, P(C_j)}$$

"This is the fundamental principle $^{(*)}$ **of that branch of the analysis of chance that consists of reasoning a posteriori from events to causes"**

(*) In his "Philosophical essay" Laplace calls 'principles' the 'fundamental rules'.

Laplace's "Bayes Theorem"

$$P(C_i \,|\, E) \;=\; \frac{P(E \,|\, C_i)\, P(C_i)}{P(E)} = \frac{P(E \,|\, C_i)\, P(C_i)}{\sum_j P(E \,|\, C_j)\, P(C_j)}$$

"This is the fundamental principle $^{(*)}$ of that branch of the analysis of chance that consists of reasoning a posteriori from events to causes"

(*) In his "Philosophical essay" Laplace calls 'principles' the 'fundamental rules'.

Note: denominator is just a normalization factor.

$$\Rightarrow \qquad P(C_i \,|\, E) \;\propto\; P(E \,|\, C_i)\, P(C_i)$$

Laplace's "Bayes Theorem"

$$P(C_i \,|\, E) \;=\; \frac{P(E \,|\, C_i)\, P(C_i)}{P(E)} = \frac{P(E \,|\, C_i)\, P(C_i)}{\sum_j P(E \,|\, C_j)\, P(C_j)}$$

"This is the fundamental principle $^{(*)}$ of that branch of the analysis of chance that consists of reasoning a posteriori from events to causes"

(*) In his "Philosophical essay" Laplace calls 'principles' the 'fundamental rules'.

Note: denominator is just a normalization factor.

$$\Rightarrow \qquad P(C_i \,|\, E) \;\propto\; P(E \,|\, C_i)\, P(C_i)$$

Most convenient way to remember Bayes theorem

Laplace's teaching

$$\frac{P(H_0 \,|\, \text{data})}{P(H_1 \,|\, \text{data})} = \frac{P(\text{data} \,|\, H_0)}{P(\text{data} \,|\, H_1)} \times \frac{P(H_0)}{P(H_1)}$$

- We should possibly use the data, rather then the test variables 'θ' (χ^2 etc);

  [although in some case 'sufficient summaries' do exist]

Laplace's teaching

$$\frac{P(H_0 \,|\, \text{data})}{P(H_1 \,|\, \text{data})} = \frac{P(\text{data} \,|\, H_0)}{P(\text{data} \,|\, H_1)} \times \frac{P(H_0)}{P(H_1)}$$

- We should possibly use the data, rather then the test variables 'θ' (χ^2 etc);

  [although in some case 'sufficient summaries' do exist]
- At least two hypotheses are needed!

Laplace's teaching

$$\frac{P(H_0 \,|\, \text{data})}{P(H_1 \,|\, \text{data})} \;=\; \frac{P(\text{data} \,|\, H_0)}{P(\text{data} \,|\, H_1)} \times \frac{P(H_0)}{P(H_1)}$$

- We should possibly use the data, rather then the test variables 'θ' (χ^2 etc);
  [although in some case 'sufficient summaries' do exist]
- At least two hypotheses are needed!
- ...and also how they appear belivable *a priori*!

Laplace's teaching

$$\frac{P(H_0 \,|\, \text{data})}{P(H_1 \,|\, \text{data})} = \frac{P(\text{data} \,|\, H_0)}{P(\text{data} \,|\, H_1)} \times \frac{P(H_0)}{P(H_1)}$$

- We should possibly use the data, rather then the test variables 'θ' (χ^2 etc);
  [although in some case 'sufficient summaries' do exist]
- At least two hypotheses are needed!
- ... and also how they appear belivable *a priori*!
- If $P(\text{data} \,|\, H_i) = 0$, it follows $P(H_i \,|\, \text{data}) = 0$:
  $\Rightarrow$ falsification (the 'serious' one) is a corollary of the theorem, rather than a principle.

Laplace's teaching

$$\frac{P(H_0 \,|\, \text{data})}{P(H_1 \,|\, \text{data})} \;=\; \frac{P(\text{data} \,|\, H_0)}{P(\text{data} \,|\, H_1)} \times \frac{P(H_0)}{P(H_1)}$$

- We should possibly use the data, rather then the test variables 'θ' (χ^2 etc);

  [although in some case 'sufficient summaries' do exist]
- At least two hypotheses are needed!
- ...and also how they appear belivable *a priori*!
- If $P(\text{data} \,|\, H_i) = 0$, it follows $P(H_i \,|\, \text{data}) = 0$:

  $\Rightarrow$ falsification (the 'serious' one) is a corollary of the theorem, rather than a principle.
- There is no conceptual problem with the fact that $P(\text{data} \,|\, H_1) \rightarrow 0$ (e.g. 10^{-37}), provided the ratio $P(\text{data} \,|\, H_0)/P(\text{data} \,|\, H_1)$ is not undefined.

Bayes factor ('likelihood ratio')

$$\frac{P(H_0 \mid \text{data})}{P(H_1 \mid \text{data})} = \frac{P(\text{data} \mid H_0)}{P(\text{data} \mid H_1)} \times \frac{P(H_0)}{P(H_1)}$$

Bayes factor ('likelihood ratio')

$$\frac{P(H_0\,|\,\text{data})}{P(H_1\,|\,\text{data})} = \frac{P(\text{data}\,|\,H_0)}{P(\text{data}\,|\,H_1)} \times \frac{P(H_0)}{P(H_1)}$$

$$\text{Prob. ratio}\big|_{posterior} = \text{Bayes factor} \times \text{Prob. ratio}\big|_{prior}$$

(*prior*/*posterior* w.r.t. data)

Bayes factor ('likelihood ratio')

$$\frac{P(H_0\,|\,\text{data})}{P(H_1\,|\,\text{data})} \;=\; \frac{P(\text{data}\,|\,H_0)}{P(\text{data}\,|\,H_1)} \times \frac{P(H_0)}{P(H_1)}$$

$$\text{Prob. ratio}|_{posterior} \;=\; \text{Bayes factor} \times \text{Prob. ratio}|_{prior}$$

(*prior*/*posterior* w.r.t. data)

If H_0 and H_1 are 'complementary', that is $H_1 = \overline{H}_0$, then

$$\textbf{posterior odds} \;=\; \textbf{Bayes factor} \times \textbf{prior odds}$$

Telling it with Gauss' words

A quote from the Princeps Mathematicorum
(Prince of Mathematicians) is a must.

Telling it with Gauss' words

A quote from the Princeps Mathematicorum
(Prince of Mathematicians) is a must.

$$P(C_i \,|\, \text{data}) = \frac{P(\text{data} \,|\, C_i)}{P(\text{data})}\, P_0(C_i)$$

Telling it with Gauss' words

A quote from the Princeps Mathematicorum
(Prince of Mathematicians) is a must.

$$P(C_i \,|\, \text{data}) = \frac{P(\text{data} \,|\, C_i)}{P(\text{data})}\, P_0(C_i)$$

"post illa observationes" *"ante illa observationes"*

(Gauss)

Telling it with Gauss' words

A quote from the Princeps Mathematicorum
(Prince of Mathematicians) is a must.

$$P(C_i\,|\,\text{data}) = \frac{P(\text{data}\,|\,C_i)}{P(\text{data})}\,P_0(C_i)$$

"post illa observationes" *"ante illa observationes"*

(Gauss)

Arguments used to derive Gaussian distribution

- $f(\mu\,|\,\{x\}) \propto f(\{x\}\,|\,\mu)\cdot f_0(\mu)$
- $f_0(\mu)$ 'flat' (all values a priory equally possible)
- posterior maximized at $\mu = \overline{x}$

Telling it with Gauss' words

A quote from the Princeps Mathematicorum (Prince of Mathematicians) is a must.

$$P(C_i \,|\, \text{data}) = \frac{P(\text{data} \,|\, C_i)}{P(\text{data})} \, P_0(C_i)$$

"post illa observationes" *"ante illa observationes"*

(Gauss)

Arguments used to derive Gaussian distribution

- $f(\mu \,|\, \{x\}) \propto f(\{x\} \,|\, \mu) \cdot f_0(\mu)$
- $f_0(\mu)$ 'flat' (all values a priory equally possible)
- posterior maximized at $\mu = \overline{x}$

Note: indeed Gauss had also invented the "Bayes Factor"!
(GdA, arXiv:2003.10878 [math.HO])

Probabilistic inference/prediction

applied to the ‘binomial’ case

Namely the original problem tackled by Laplace and Bayes

Probabilistic inference/prediction

applied to the 'binomial' case

Namely the original problem tackled by Laplace and Bayes,
but in modern notation and making use of a graphical model:

Probabilistic inference/prediction

applied to the 'binomial' case

Namely the original problem tackled by Laplace and Bayes,
but in modern notation and making use of a graphical model:

1. draw the graphical model;
2. write down the joint pdf of all variables entering the game;
3. use Bayes theorem in order to condition on what is known/assumed;
4. *marginalize* over all variables on which we are not interesting;
5. do *somehow* the math.

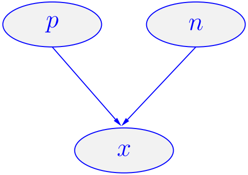
n independent *Bernoulli processes*

General case

n independent *Bernoulli processes*

General case

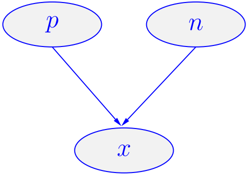
Model



n independent *Bernoulli processes*

General case

Model



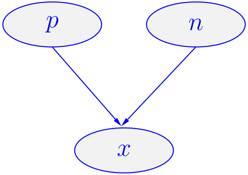
Joint pdf (omitting background condition I)
→ making use of the **chain rule**:

$$f(x, p, n) \;=\; f(x \,|\, p, n) \cdot f(p, n)$$

n independent *Bernoulli processes*

General case

Model



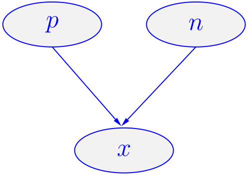
Joint pdf (omitting background condition I)
$\rightarrow$ making use of the **chain rule**:

$$
\begin{aligned}
f(x, p, n) &= f(x \mid p, n) \cdot f(p, n) \\
&= f(x \mid p, n) \cdot f(p \mid n) \cdot f(n)
\end{aligned}
$$

n independent *Bernoulli processes*

General case

Model



Joint pdf (omitting background condition I)

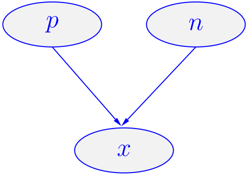
→ making use of the **chain rule**:

$$
\begin{aligned}
f(x, p, n) &= f(x \,|\, p, n) \cdot f(p, n) \\
&= f(x \,|\, p, n) \cdot f(p \,|\, n) \cdot f(n) \\
&= f(x \,|\, p, n) \cdot f(n \,|\, p) \cdot f(p)
\end{aligned}
$$

n independent *Bernoulli processes*

General case

Model



Joint pdf (omitting background condition I)
→ making use of the **chain rule**:

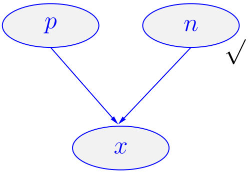
$$
\begin{aligned}
f(x, p, n) &= f(x \mid p, n) \cdot f(p, n) \\
&= f(x \mid p, n) \cdot f(p \mid n) \cdot f(n) \\
&= f(x \mid p, n) \cdot f(n \mid p) \cdot f(p) \\
&= f(x \mid p, n) \cdot f(p) \cdot f(n)
\end{aligned}
$$

(n and p are independent)

n independent Bernoulli processes

Usual case → n fixed (for the moment)

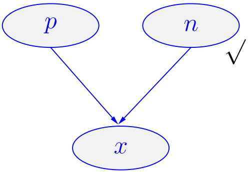
Model



n independent Bernoulli processes

Usual case $\rightarrow$ n fixed (for the moment)

Model

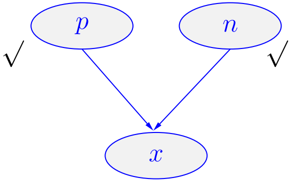


Joint pdf

$$f(x, p \mid n) \quad = \quad f(x \mid p, n) \cdot f(p)$$

n independent Bernoulli processes

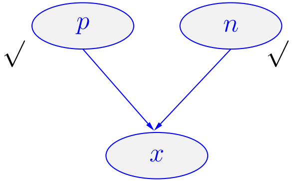
Graphical models of the typical problems



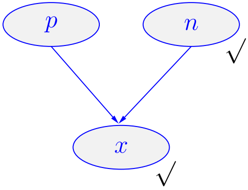
$\rightarrow\ f(x \mid n, p)$

n independent Bernoulli processes

Graphical models of the typical problems



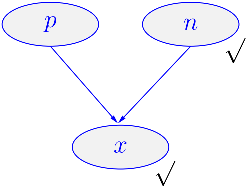
$\rightarrow\ f(x \mid n, p)$



$\rightarrow\ f(p \mid n, x)$

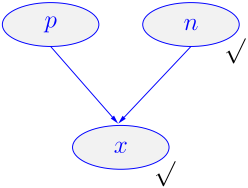
n independent Bernoulli processes

Inferring p



n independent Bernoulli processes

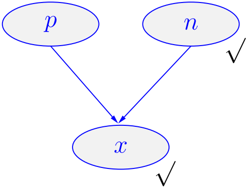
Inferring p



$$f(p\,|\,x,n) \;=\; \frac{f(p,x\,|\,n)}{f(x\,|\,n)}$$

n independent Bernoulli processes

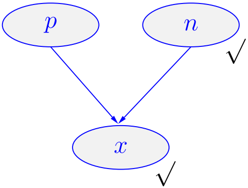
Inferring p



$$
\begin{aligned}
f(p\,|\,x,n) &= \frac{f(p,x\,|\,n)}{f(x\,|\,n)} \\
&= \frac{f(x\,|\,n,p)\cdot f_0(p)}{f(x\,|\,n)}
\end{aligned}
$$

n independent Bernoulli processes

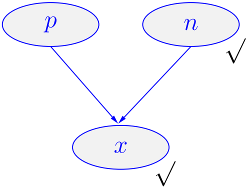
Inferring p



$$
\begin{aligned}
f(p \,|\, x, n) &= \frac{f(p, x \,|\, n)}{f(x \,|\, n)} \\
&= \frac{f(x \,|\, n, p) \cdot f_0(p)}{f(x \,|\, n)} \\
&= \frac{f(x \,|\, n, p) \cdot f_0(p)}{\int_0^1 f(x \,|\, n, p) \cdot f_0(p)\, dp}
\end{aligned}
$$

n independent Bernoulli processes

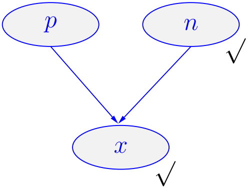
Inferring p



$$
\begin{aligned}
f(p\,|\,x,n) &= \frac{f(p,x\,|\,n)}{f(x\,|\,n)} \\
&= \frac{f(x\,|\,n,p)\cdot f_0(p)}{f(x\,|\,n)} \\
&= \frac{f(x\,|\,n,p)\cdot f_0(p)}{\int_0^1 f(x\,|\,n,p)\cdot f_0(p)\,dp} \\
&\propto f(x\,|\,n,p)\cdot f_0(p)
\end{aligned}
$$

n independent Bernoulli processes

Inferring p



$$
\begin{aligned}
f(p\,|\,x,n) &= \frac{f(p,x\,|\,n)}{f(x\,|\,n)} \\
&= \frac{f(x\,|\,n,p)\cdot f_0(p)}{f(x\,|\,n)} \\
&= \frac{f(x\,|\,n,p)\cdot f_0(p)}{\int_0^1 f(x\,|\,n,p)\cdot f_0(p)\,dp} \\
&\propto f(x\,|\,n,p)\cdot f_0(p)
\end{aligned}
$$

(denominator just normalization!)

Inferring "Bernoulli's p"

We just need to make explicit $f(x \mid n, p)$:

$$f(x \mid n, p) \;=\; \binom{n}{x}\, p^x\, (1-p)^{n-x}$$

Inferring "Bernoulli's p"

We just need to make explicit $f(x\,|\,n,p)$:

$$f(x\,|\,n,p) \;=\; \binom{n}{x}\, p^x\,(1-p)^{n-x} \;=\; \frac{n!}{(n-x)!\,x!}\, p^x\,(1-p)^{n-x}$$

Inferring "Bernoulli's p"

We just need to make explicit $f(x\,|\,n,p)$:

$$f(x\,|\,n,p) \;=\; \binom{n}{x}\,p^x\,(1-p)^{n-x} \;=\; \frac{n!}{(n-x)!\,x!}\,p^x\,(1-p)^{n-x}$$

We get then, including normalization:

$$f(p\,|\,x,n) \;=\; \frac{\frac{n!}{(n-x)!\,x!}\,p^x\,(1-p)^{n-x}\,f_\circ(p)}{\int_0^1 \frac{n!}{(n-x)!\,x!}\,p^x\,(1-p)^{n-x}\,f_\circ(p)\,\mathrm{d}p}$$

Inferring "Bernoulli's p"

We just need to make explicit $f(x \,|\, n, p)$:

$$f(x \,|\, n, p) \;=\; \binom{n}{x} p^x (1-p)^{n-x} \;=\; \frac{n!}{(n-x)!\,x!}\, p^x (1-p)^{n-x}$$

We get then, including normalization:

$$\begin{aligned} f(p \,|\, x, n) &= \frac{\frac{n!}{(n-x)!\,x!}\, p^x (1-p)^{n-x} f_\circ(p)}{\int_0^1 \frac{n!}{(n-x)!\,x!}\, p^x (1-p)^{n-x} f_\circ(p)\, \mathrm{d}p} \\ &= \frac{p^x (1-p)^{n-x} f_\circ(p)}{\int_0^1 p^x (1-p)^{n-x} f_\circ(p)\, \mathrm{d}p} \end{aligned}$$

Inferring "Bernoulli's p"

We just need to make explicit $f(x \,|\, n, p)$:

$$f(x \,|\, n, p) \;=\; \binom{n}{x} p^x (1-p)^{n-x} \;=\; \frac{n!}{(n-x)!\,x!}\, p^x (1-p)^{n-x}$$

We get then, including normalization:

$$\begin{aligned} f(p \,|\, x, n) &= \frac{\frac{n!}{(n-x)!\,x!}\, p^x (1-p)^{n-x} f_\circ(p)}{\int_0^1 \frac{n!}{(n-x)!\,x!}\, p^x (1-p)^{n-x} f_\circ(p)\,\mathrm{d}p} \\ &= \frac{p^x (1-p)^{n-x} f_\circ(p)}{\int_0^1 p^x (1-p)^{n-x} f_\circ(p)\,\mathrm{d}p} \end{aligned}$$

(The binomial coefficient is irrelevant, not depending on p)

Inferring "Bernoulli's p"

$$f(p\,|\,x,n) \;=\; \frac{p^x\,(1-p)^{n-x}\,f_\circ(p)}{\int_0^1 p^x\,(1-p)^{n-x}\,f_\circ(p)\,\mathrm{d}p}$$

Inferring "Bernoulli's p"

$$f(p\,|\,x,n) \;=\; \frac{p^x\,(1-p)^{n-x}\,f_\circ(p)}{\int_0^1 p^x\,(1-p)^{n-x}\,f_\circ(p)\,\mathrm{d}p}$$

For teaching purposes we start from a **uniform prior**, i.e. $\boldsymbol{f_\circ(p)} = 1$:

$$f(p\,|\,x,n) \;=\; \frac{p^x\,(1-p)^{n-x}}{\int_0^1 p^x\,(1-p)^{n-x}\,\mathrm{d}p}$$

Inferring "Bernoulli's p"

$$f(p\,|\,x,n) \;=\; \frac{p^x\,(1-p)^{n-x}\,f_\circ(p)}{\int_0^1 p^x\,(1-p)^{n-x}\,f_\circ(p)\,\mathrm{d}p}$$

For teaching purposes we start from a **uniform prior**, i.e. $\boldsymbol{f_\circ(p) = 1}$:

$$f(p\,|\,x,n) \;=\; \frac{p^x\,(1-p)^{n-x}}{\int_0^1 p^x\,(1-p)^{n-x}\,\mathrm{d}p}$$

- The integral at the denominator is the special function "β" (also defined for real values of x and n).

Inferring "Bernoulli's p"

$$f(p\,|\,x,n) \;=\; \frac{p^x\,(1-p)^{n-x}\,f_\circ(p)}{\int_0^1 p^x\,(1-p)^{n-x}\,f_\circ(p)\,\mathrm{d}p}$$

For teaching purposes we start from a **uniform prior**, i.e. $\boldsymbol{f_\circ(p) = 1}$:

$$f(p\,|\,x,n) \;=\; \frac{p^x\,(1-p)^{n-x}}{\int_0^1 p^x\,(1-p)^{n-x}\,\mathrm{d}p}$$

- The integral at the denominator is the special function "β" (also defined for real values of x and n).
- In our case these two numbers are integer and the integral becomes equal to

$$\frac{x!\,(n-x)!}{(n+1)!}$$

Inferring "Bernoulli's p"

Solution for uniform prior

$$f(p \,|\, x, n) \;=\; \frac{(n+1)!}{x!\,(n-x)!}\, p^x\,(1-p)^{n-x}$$

Inferring "Bernoulli's p"

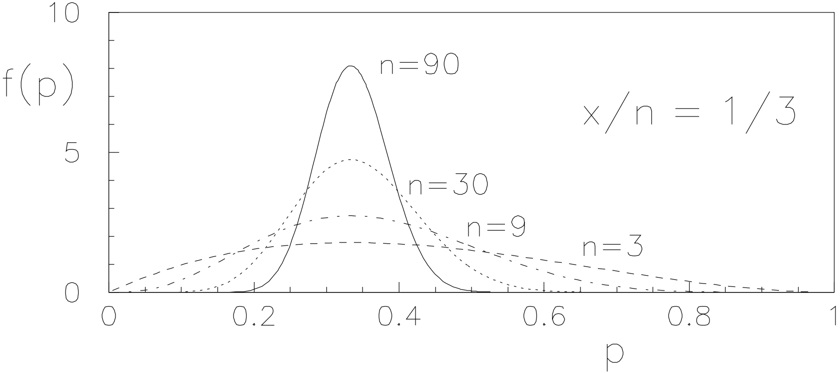
Solution for uniform prior

$$\begin{aligned} f(p\,|\,x,n) &= \frac{(n+1)!}{x!\,(n-x)!}\,p^x\,(1-p)^{n-x} \\ &= (n+1)\cdot\frac{n!}{x!\,(n-x)!}\,p^x\,(1-p)^{n-x} \end{aligned}$$

Inferring "Bernoulli's p"

Solution for uniform prior

$$
\begin{aligned}
f(p\,|\,x,n) &= \frac{(n+1)!}{x!\,(n-x)!}\,p^x\,(1-p)^{n-x}\\
&= (n+1)\cdot\frac{n!}{x!\,(n-x)!}\,p^x\,(1-p)^{n-x}
\end{aligned}
$$



Inferring "Bernoulli's p"

Summaries of the **posterior distribution**

$$p_m = \mathrm{mode}(p) \quad = \quad \frac{x}{n}$$

Inferring "Bernoulli's p"

Summaries of the **posterior distribution**

$$
\begin{aligned}
p_m = \text{mode}(p) &= \frac{x}{n} \\
\mathrm{E}(p) &= \frac{x+1}{n+2}
\end{aligned}
$$

Inferring "Bernoulli's p"

Summaries of the **posterior distribution**

$$p_m = \text{mode}(p) = \frac{x}{n}$$

$$\mathrm{E}(p) = \frac{x+1}{n+2}$$

"recursive Laplace formula"

("Laplace's rule of succession")

Inferring "Bernoulli's p"

Summaries of the **posterior distribution**

$$p_m = \text{mode}(p) = \frac{x}{n}$$

$$\mathrm{E}(p) = \frac{x+1}{n+2}$$

"recursive Laplace formula"

("Laplace's rule of succession")

$$\begin{aligned}
\mathrm{Var}(p) &= \frac{(x+1)(n-x+1)}{(n+3)(n+2)^2} \\
&= \frac{x+1}{n+2}\left(\frac{n+2}{n+2} - \frac{x+1}{n+2}\right)\frac{1}{n+3} \\
&= \mathrm{E}(p)\,(1-\mathrm{E}(p))\,\frac{1}{n+3}
\end{aligned}$$

Inferring the "Bernoulli's p"

About the meaning of E(p)

- We have used the "first"$^{(*)}$ n trials to learn about "p".

  [$^{(*)}$ "First" does not imply time order, but just order in usage.]

Inferring the "Bernoulli's p"

About the meaning of E(p)

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

Inferring the “Bernoulli's p”

About the meaning of E(p)

- We have used the “first”$^{(*)}$ n trials to learn about “p”.
  [$^{(*)}$ “First” does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- If we were sure about p,

Inferring the "Bernoulli's p"

About the meaning of E(p)

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- If we were sure about p, then p would be our probability:

Inferring the "Bernoulli's p"

About the meaning of E(p)

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- If we were sure about p, then p would be our probability:

$$P(E_i \,|\, p) = p$$

Inferring the "Bernoulli's p"

About the meaning of E(p)

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- If we were sure about p, then p would be our probability:

$$P(E_i \mid p) = p$$

- But since we are uncertain about it, we have to take into account all possible values, weighing them with our degree of belief.

Inferring the "Bernoulli's p"

About the meaning of E(p)

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- If we were sure about p, then p would be our probability:

$$P(E_i \,|\, p) = p$$

- But since we are uncertain about it, we have to take into account all possible values, weighing them with our degree of belief.

$$P(E_{i>n} \,|\, x, n) \;=\; \int_0^1 P(E_i \,|\, p) \cdot f(p \,|\, x, n)\, \mathrm{d}p$$

Inferring the "Bernoulli's p"

About the meaning of E(p)

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- If we were sure about p, then p would be our probability:

$$P(E_i \,|\, p) = p$$

- But since we are uncertain about it, we have to take into account all possible values, weighing them with our degree of belief.

$$\begin{aligned} P(E_{i>n}\,|\,x,n) &= \int_0^1 P(E_i\,|\,p)\cdot f(p\,|\,x,n)\,\mathrm{d}p \\ &= \int_0^1 p\cdot f(p\,|\,x,n)\,\mathrm{d}p \end{aligned}$$

Inferring the "Bernoulli's p"

About the meaning of $\mathrm{E}(p)$

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- <u>If</u> we were sure about p, <u>then</u> p would be our probability:

$$P(E_i\,|\,p) = p$$

- But since we are uncertain about it, we have to take into account all possible values, weighing them with our degree of belief.

$$\begin{aligned} P(E_{i>n}\,|\,x,n) &= \int_0^1 P(E_i\,|\,p)\cdot f(p\,|\,x,n)\,\mathrm{d}p \\ &= \int_0^1 p\cdot f(p\,|\,x,n)\,\mathrm{d}p = \mathrm{E}(p) \quad (!!) \end{aligned}$$

Inferring the "Bernoulli's p"

About the meaning of E(p)

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- If we were sure about p, then p would be our probability:

$$P(E_i \,|\, p) = p$$

- But since we are uncertain about it, we have to take into account all possible values, weighing them with our degree of belief.

$$\begin{aligned} P(E_{i>n} \,|\, x, n) &= \int_0^1 P(E_i \,|\, p) \cdot f(p \,|\, x, n)\, \mathrm{d}p \\ &= \int_0^1 p \cdot f(p \,|\, x, n)\, \mathrm{d}p = \mathrm{E}(p) \quad (!!) \end{aligned}$$

E(p) (and not the mode!) is the probability of every 'future' event which is believed to have the **same** $\boldsymbol{p}$ of the 'previous' ones.

Related recent applications

(with many details, including general introduction to the relevant ideas and methods, and program samples)

- GdA and A. Esposito, Checking individuals and sampling populations with imperfect tests, (arXiv:2009.04843 [q-bio.PE])
- GdA and A. Esposito, What is the probability that a vaccinated person is shielded from Covid-19? A Bayesian MCMC based reanalysis of published data with emphasis on what should be reported as 'efficacy', (arXiv:2102.11022 [stat.AP])

Inferring the “Bernoulli's p”

Large number behaviour: summary

When

- n large;
- x large;

Inferring the "Bernoulli's p"

Large number behaviour: summary

When

- n large;
- x large;
- <u>and</u> $(n-x)$ large

Inferring the "Bernoulli's p"

Large number behaviour: summary

When

- n large;
- x large;
- <u>and</u> $(n-x)$ large
  (remember: in the binomial what is 'success' and what is 'failure' is not absolute: $p \longleftrightarrow q = 1-p$)

Inferring the "Bernoulli's p"

Large number behaviour: summary

When

- n large;
- x large;
- <u>and</u> $(n-x)$ large
  (remember: in the binomial what is 'success' and what is 'failure' is not absolute: $p \longleftrightarrow q = 1-p$),

then

$$\mathrm{E}(p) \;\approx\; \frac{x}{n}$$

Inferring the "Bernoulli's p"

Large number behaviour: summary

When

- n large;
- x large;
- and $(n-x)$ large
  (remember: in the binomial what is 'success' and what is 'failure' is not absolute: $p \longleftrightarrow q = 1-p$),

then

$$\begin{aligned} \mathrm{E}(p) &\approx \frac{x}{n} \\ \sigma(p) &\approx \frac{1}{\sqrt{n}}\sqrt{\frac{x}{n}\left(1-\frac{x}{n}\right)} \end{aligned}$$

Inferring the "Bernoulli's p"

Large number behaviour: summary

When

- n large;
- x large;
- and $(n-x)$ large
  (remember: in the binomial what is 'success' and what is 'failure' is not absolute: $p \longleftrightarrow q = 1-p$),

then

$$\begin{aligned} \mathrm{E}(p) &\approx \frac{x}{n} \\ \sigma(p) &\approx \frac{1}{\sqrt{n}}\sqrt{\frac{x}{n}\left(1-\frac{x}{n}\right)} \end{aligned}$$

— $f(p\,|\,x, n)$ tends to Gaussian

Inferring the "Bernoulli's p"

Large number behaviour: summary

When

- n large;
- x large;
- and $(n - x)$ large
  (remember: in the binomial what is 'success' and what is 'failure' is not absolute: $p \longleftrightarrow q = 1 - p$),

then

$$\begin{aligned} \mathrm{E}(p) &\approx \frac{x}{n} \\ \sigma(p) &\approx \frac{1}{\sqrt{n}} \sqrt{\frac{x}{n}\left(1 - \frac{x}{n}\right)} \end{aligned}$$

— $f(p \,|\, x, n)$ tends to Gaussian,
a reflection of the Gaussian limit of $f(x \,|\, p, n)$

Inferring the "Bernoulli's p"

Large number behaviour: summary

When

- n large;
- x large;
- and $(n - x)$ large
  (remember: in the binomial what is 'success' and what is 'failure' is not absolute: $p \longleftrightarrow q = 1 - p$),

then

$$
\begin{aligned}
\mathrm{E}(p) &\approx \frac{x}{n} \\
\sigma(p) &\approx \frac{1}{\sqrt{n}}\sqrt{\frac{x}{n}\left(1 - \frac{x}{n}\right)}
\end{aligned}
$$

— $f(p \,|\, x, n)$ tends to Gaussian,
a reflection of the Gaussian limit of $f(x \,|\, p, n)$

— **The probability of a future events is evaluated from the relative frequency of the past events**

Inferring the "Bernoulli's p"

Large number behaviour: summary

When

- n large;
- x large;
- <u>and</u> $(n-x)$ large
  (remember: in the binomial what is 'success' and what is 'failure' is not absolute: $p \longleftrightarrow q = 1-p$),

then

$$\begin{aligned} \mathrm{E}(p) &\approx \frac{x}{n} \\ \sigma(p) &\approx \frac{1}{\sqrt{n}}\sqrt{\frac{x}{n}\left(1-\frac{x}{n}\right)} \end{aligned}$$

— $f(p\,|\,x, n)$ tends to Gaussian,
a reflection of the Gaussian limit of $f(x\,|\,p, n)$

— **The probability of a future events is evaluated from the relative frequency of the past events**

— **No need of 'frequentistic definition'!**

Inferring the “Bernoulli's p”

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas (“Bayesian”) has always been a severe problem!

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$\rightarrow$ **Computational barrier**

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$\rightarrow$ **Computational barrier**

Some tricks have been invented.

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$\rightarrow$ **Computational barrier**

Some tricks have been invented.
Here is a very elegant one, particularly suitable useful to infer Bernoulli's p, in order to include the prior in an effordable mathematical way

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$\rightarrow$ **Computational barrier**

Some tricks have been invented.
Here is a very elegant one, particularly suitable useful to infer Bernoulli's p, in order to include the prior in an effordable mathematical way

- imagine that we could express $f_0(p)$ in the following form

$$f_0(p) \propto p^a \, (1-p)^b$$

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$$\rightarrow \textbf{Computational barrier}$$

Some tricks have been invented.
Here is a very elegant one, particularly suitable useful to infer Bernoulli's p, in order to include the prior in an effordable mathematical way

- imagine that we could express $f_0(p)$ in the following form

$$f_0(p) \propto p^a\,(1-p)^b$$

- Then the inference becomes

$$f(p\,|\,x,n) \quad \propto \quad p^x\,(1-p)^{n-x}\cdot p^a\,(1-p)^b$$

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$\rightarrow$ **Computational barrier**

Some tricks have been invented.
Here is a very elegant one, particularly suitable useful to infer Bernoulli's p, in order to include the prior in an effordable mathematical way

- imagine that we could express $f_0(p)$ in the following form

$$f_0(p) \propto p^a\,(1-p)^b$$

- Then the inference becomes

$$\begin{aligned} f(p\,|\,x,n) &\propto p^x\,(1-p)^{n-x}\cdot p^a\,(1-p)^b \\ &\propto p^{a+x}\,(1-p)^{b+(n-x)} \end{aligned}$$

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$\rightarrow$ **Computational barrier**

Some tricks have been invented.
Here is a very elegant one, particularly suitable useful to infer Bernoulli's p, in order to include the prior in an effordable mathematical way

- imagine that we could express $f_0(p)$ in the following form

$$f_0(p) \propto p^a\,(1-p)^b$$

- Then the inference becomes

$$\begin{aligned} f(p\,|\,x,n) &\propto p^x\,(1-p)^{n-x}\cdot p^a\,(1-p)^b \\ &\propto p^{a+x}\,(1-p)^{b+(n-x)} \\ &\propto p^{a'}\,(1-p)^{b'} \end{aligned}$$

Conjugate priors

Indeed, for the *binomial problem* such a pdf exists.

Conjugate priors

Indeed, for the *binomial problem* such a pdf exists.

In general, given the generic uncertain number X,

$$f(x \,|\, \text{Beta}(r,s)) = \frac{1}{\beta(r,s)} x^{r-1}(1-x)^{s-1} \qquad \begin{cases} r,\, s > 0 \\ 0 \leq x \leq 1 \end{cases}$$

with $a = r - 1$ and $b = s - 1$

Beta distribution

Conjugate priors

Indeed, for the *binomial problem* such a pdf exists.

In general, given the generic uncertain number X,

$$f(x \mid \text{Beta}(r,s)) = \frac{1}{\beta(r,s)} x^{r-1}(1-x)^{s-1} \qquad \begin{cases} r,\ s > 0 \\ 0 \le x \le 1 \end{cases}$$

with $a = r - 1$ and $b = s - 1$

Beta distribution

The Beta distribution is an example of conjugate prior:

Conjugate priors

Indeed, for the *binomial problem* such a pdf exists.
In general, given the generic uncertain number X,

$$f(x \mid \text{Beta}(r,s)) = \frac{1}{\beta(r,s)} x^{r-1}(1-x)^{s-1} \qquad \begin{cases} r,\ s > 0 \\ 0 \le x \le 1 \end{cases}$$

with $a = r - 1$ and $b = s - 1$

Beta distribution

The Beta distribution is an example of conjugate prior:

- a pdf such that prior and posterior belong to the same family;
- its parameters are updated by the the 'likelihood'.

Conjugate priors

Indeed, for the *binomial problem* such a pdf exists.
In general, given the generic uncertain number X,

$$f(x\,|\,\mathrm{Beta}(r,s)) = \frac{1}{\beta(r,s)} x^{r-1}(1-x)^{s-1} \qquad \begin{cases} r,\ s > 0 \\ 0 \le x \le 1 \end{cases}$$

with $a = r - 1$ and $b = s - 1$

Beta distribution

The Beta distribution is an example of conjugate prior:

- a pdf such that prior and posterior belong to the same family;
- its parameters are updated by the the 'likelihood'.

Note:

- not all *conjugate priors* are as flexible as the Beta.

Conjugate priors

Indeed, for the *binomial problem* such a pdf exists.
In general, given the generic uncertain number X,

$$f(x \,|\, \mathrm{Beta}(r,s)) = \frac{1}{\beta(r,s)} x^{r-1}(1-x)^{s-1} \qquad \begin{cases} r,\ s > 0 \\ 0 \le x \le 1 \end{cases}$$

with $a = r - 1$ and $b = s - 1$

Beta distribution

The Beta distribution is an example of conjugate prior:

- a pdf such that prior and posterior belong to the same family;
- its parameters are updated by the the 'likelihood'.

Note:

- not all *conjugate priors* are as flexible as the Beta.
  (In particular, the Gaussian is self-conjugate,
  which is not so great...)

Predictive distribution

Predicting future nr. of successes and future frequences

- Imagine we have have got 5 successes in 10 trials.

Predictive distribution

Predicting future nr. of successes and future frequences

- Imagine we have have got 5 successes in 10 trials.
- Imagine that we want to make another 10 trials:
  what is the probability to get 0, 1, . . . , 10 successes?

Predictive distribution

Predicting future nr. of successes and future frequences

- Imagine we have have got 5 successes in 10 trials.
- Imagine that we want to make another 10 trials:
  what is the probability to get 0, 1, . . . , 10 successes?
- From the past data (and assuming a flat prior), we 'know' that $p \approx 0.5$.

Predictive distribution

Predicting future nr. of successes and future frequences

- Imagine we have have got 5 successes in 10 trials.
- Imagine that we want to make another 10 trials:
  what is the probability to get 0, 1, . . . , 10 successes?
- From the past data (and assuming a flat prior), we 'know' that $p \approx 0.5$.
- If we were sure that p was $1/2$, then we could simply use $\mathcal{B}_{10,\,1/2}$.

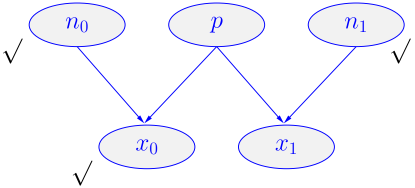
Predictive distribution

Predicting future nr. of successes and future frequences

- Imagine we have have got 5 successes in 10 trials.
- Imagine that we want to make another 10 trials:
  what is the probability to get 0, 1, . . . , 10 successes?
- From the past data (and assuming a flat prior), we 'know' that $p \approx 0.5$.
- If we were sure that p was $1/2$, then we could simply use $\mathcal{B}_{10,\,1/2}$.
- But we are not sure about it: we need to take into account all possible values, each weighted by $f(p)$

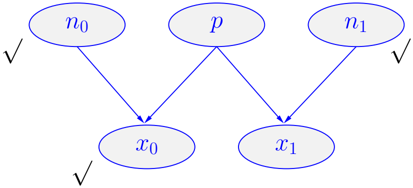
Predictive distribution

Predicting future nr. of successes and future frequences



Predictive distribution

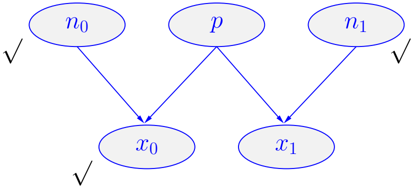
Predicting future nr. of successes and future frequences



- We need to take into account all possible values of p, each weighted by how much we believe it, i.e. by $f(p)$

Predictive distribution

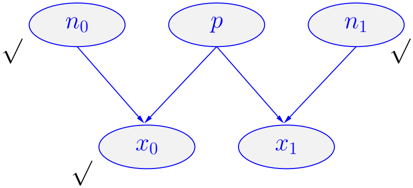
Predicting future nr. of successes and future frequences



- We need to take into account all possible values of p, each weighted by how much we believe it, i.e. by $f(p)$
- $f(x) = \int_0^1 f(x \mid p)\, f(p)\, dp.$

Predictive distribution

Predicting future nr. of successes and future frequences



- We need to take into account all possible values of p, each weighted by how much we believe it, i.e. by $f(p)$
- $f(x) = \int_0^1 f(x \,|\, p)\, f(p)\, dp.$

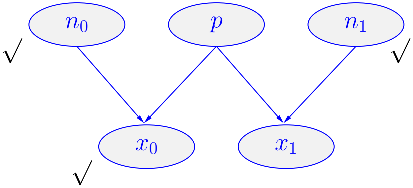
- More precisely,

$$f(x_1 \,|\, n_1, n_0, x_0) = \int_0^1 f(x_1 \,|\, n_1, p)\, f(p \,|\, x_0, n_0)\, dp$$

- $X_1 \to f_1$

Predictive distribution

Predicting future nr. of successes and future frequences



- We need to take into account all possible values of p, each weighted by how much we believe it, i.e. by $f(p)$
- $f(x) = \int_0^1 f(x \mid p)\, f(p)\, dp$.
- More precisely,

$$f(x_1 \mid n_1, n_0, x_0) = \int_0^1 f(x_1 \mid n_1, p)\, f(p \mid x_0, n_0)\, dp$$

- $X_1 \to f_1$ (Predicting a **future frequency from a past frequency**)

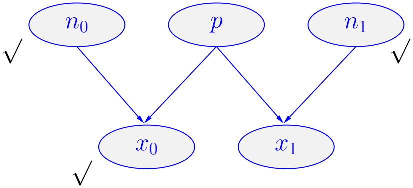
Predictive distribution

Some examples

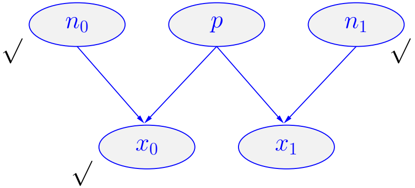
$f(x_1 \mid n_0, x_0, n_1 = 10)$ in %

X_1	$\frac{X_1}{n_1}$	$\begin{cases} x_0 = 1 \\ n_0 = 2 \end{cases}$	$\begin{cases} x_0 = 10 \\ n_0 = 20 \end{cases}$	$\begin{cases} x_0 = 100 \\ n_0 = 200 \end{cases}$	$\begin{cases} x_0 = 1000 \\ n_0 = 2000 \end{cases}$
0	0	3.85	0.42	0.12	0.10
1	0.1	6.99	2.29	1.11	0.99
2	0.2	9.44	6.51	4.67	4.42
3	0.3	11.19	12.54	11.88	11.74
4	0.4	12.24	18.07	20.21	20.48
5	0.5	12.59	20.33	24.02	24.55
6	0.6	12.24	18.07	20.21	20.48
7	0.7	11.19	12.54	11.88	11.74
8	0.8	9.44	6.51	4.67	4.42
9	0.9	6.99	2.29	1.11	0.99
10	1	3.84	0.42	0.12	0.10
$\mathrm{E}(X_1)$		5	5	5	5
$\sigma[X_1]$		2.64	1.87	1.62	1.58

Joint inference and prediction



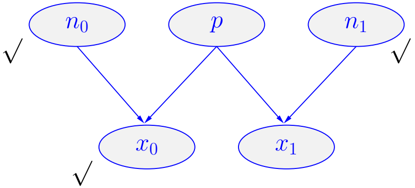
Joint inference and prediction



In reality the general solution starts from

$$f(n_0, p, n_1, x_0, x_1)$$

Joint inference and prediction



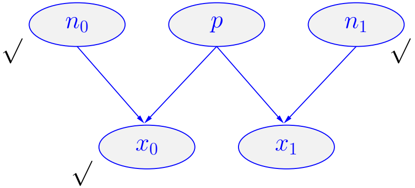
In reality the general solution starts from

$$f(n_0, p, n_1, x_0, x_1)$$

conditioning on what is 'known' (or 'assumed'):

$$f(p, x_1 \,|\, n_0, x_0, n_1) \;=\; \frac{f(p, x_1, n_0, x_0, n_1)}{f(n_0, x_0, n_1)}$$

Joint inference and prediction



In reality the general solution starts from

$$f(n_0, p, n_1, x_0, x_1)$$

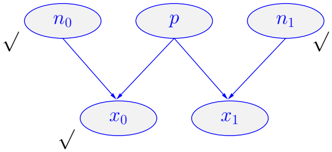
conditioning on what is 'known' (or 'assumed'):

$$\begin{aligned} f(p, x_1 \,|\, n_0, x_0, n_1) &= \frac{f(p, x_1, n_0, x_0, n_1)}{f(n_0, x_0, n_1)} \\ &\propto f(p, x_1, n_0, x_0, n_1) \end{aligned}$$

⇒ The denominator is just a constant.

⇒ **Very important observation** in order to solve the problem numerically or by Monte Carlo methods! (And remember that the numerator can be obtained using the chain rule)

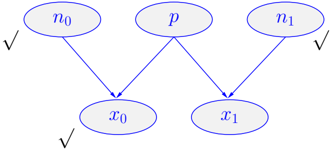
Graphical models



Terminology:

- nodes (observed/unobserved);
- child/childred;
- parent(s).

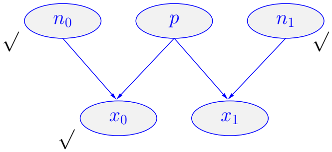
Graphical models



Terminology:

- nodes (observed/unobserved);
- child/childred;
- parent(s).
- An **unobserved node without parents needs a prior** (p in this case)

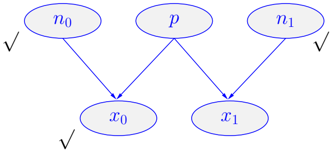
Graphical models



Terminology:

- nodes (observed/unobserved);
- child/childred;
- parent(s).
- An **unobserved node without parents needs a prior** (p in this case)
  No dogma or wickedness of the 'Bayesians'
  $\Rightarrow$ just Probability Theory (chain rule).

Graphical models



Terminology:

- nodes (observed/unobserved);
- child/childred;
- parent(s).
- An **unobserved node without parents needs a prior** (p in this case)
  No dogma or wickedness of the 'Bayesians'
  $\Rightarrow$ just Probability Theory (chain rule).

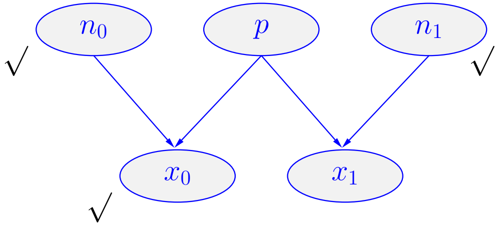
Software to analyse it:

- instructions which remind the description of the model by a suitable chain rule;
- computation performed by Markov Chain Monte Carlo.

Joint inference and prediction in JAGS

JAGS: Just Another Gibbs Sampler

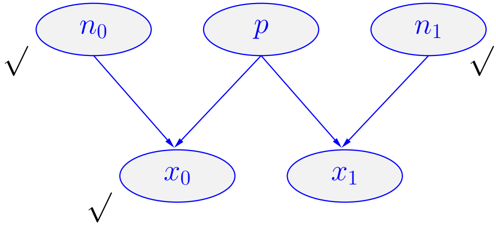
(The Gibbs Sampler is an MCMC algorithm,
but the software also uses Metropolis in *though cases*)



Joint inference and prediction in JAGS

JAGS: Just Another Gibbs Sampler

(The Gibbs Sampler is an MCMC algorithm,
but the software also uses Metropolis in *though cases*)

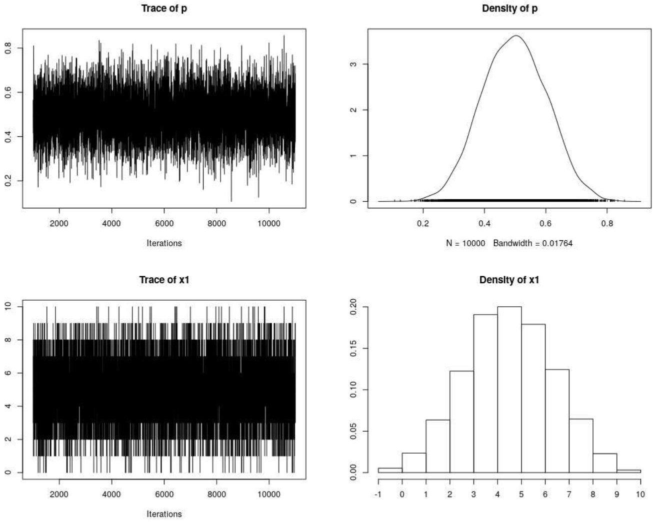


Model:

```
model{
    x0 ~ dbin(p, n0)
    x1 ~ dbin(p, n1)
     p ~ dbeta(1, 1)    # flat prior in termes of a Beta
}
```

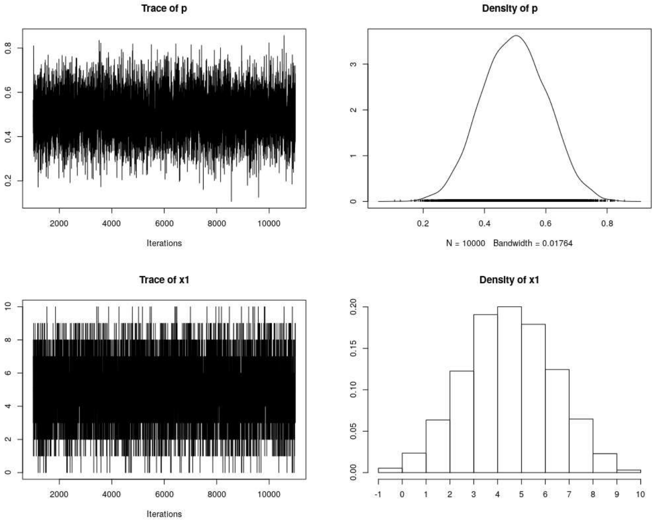
Use of JAGS from R via rjags

$$(n0 = 20,\ x0 = 10,\ n1 = 10)$$



Use of JAGS from R via rjags

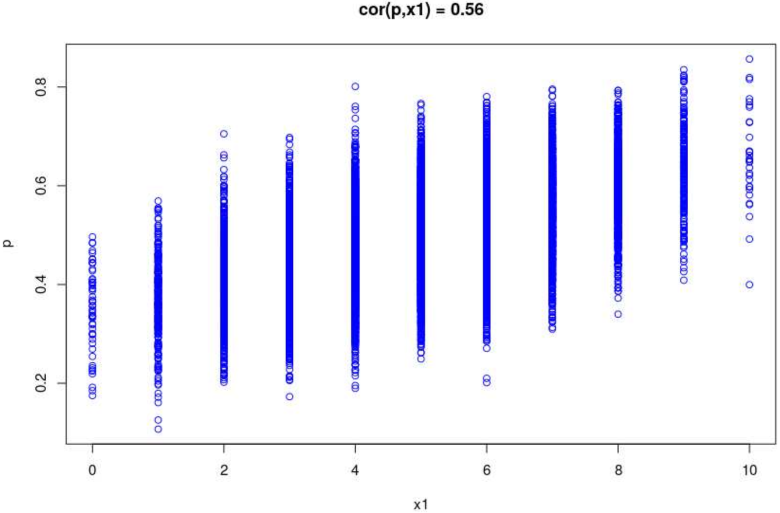
$(n0 = 20,\ x0 = 10,\ n1 = 10)$



$p = 0.498 \pm 0.105$; $x_1 = 4.98 \pm 1.86$ (10000 samples).

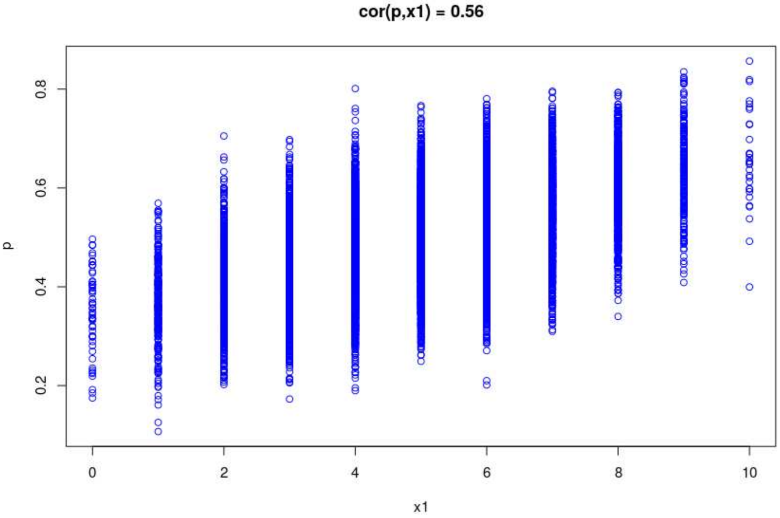
Use of JAGS from R via rjags

Scatter plot of sampled $f(p, x_1 \mid n_0, x_0, n_1)$



Use of JAGS from R via rjags

Scatter plot of sampled $f(p, x_1 \mid n_0, x_0, n_1)$



BAT - the Bayesian Analysis Toolkit

A package more suited for Physics analysis (expecially HEP)

https://bat.mpp.mpg.de/

BAT - the Bayesian Analysis Toolkit

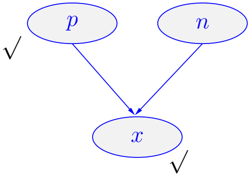
A package more suited for Physics analysis (expecially HEP)

https://bat.mpp.mpg.de/

Presently rewritten in Julia: https://github.com/bat/BAT.jl

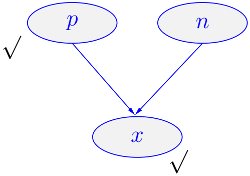
n independent Bernoulli processes

Inferring n



n independent Bernoulli processes

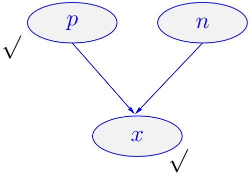
Inferring n



Think at a detector having a *well known efficiency* ($\epsilon \equiv p$)

n independent Bernoulli processes

Inferring *n*

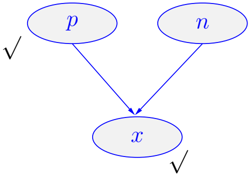


Think at a detector having a *well known efficiency* ($\epsilon \equiv p$):

- we have recorded x 'signals';

n independent Bernoulli processes

Inferring n

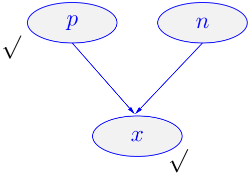


Think at a detector having a *well known efficiency* ($\epsilon \equiv p$):

- we have recorded x 'signals';
- how many particles impinged the detector?

n independent Bernoulli processes

Inferring n



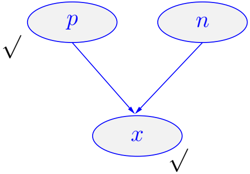
Think at a detector having a *well known efficiency* ($\epsilon \equiv p$):

- we have recorded x 'signals';
- how many particles impinged the detector? $\longrightarrow f(n \,|\, x, p)$ **?**

n independent Bernoulli processes

Extending the model

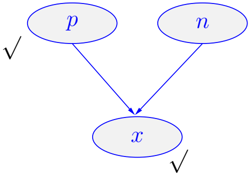
Our problem (but in Physics it is often not so simple)



n independent Bernoulli processes

Extending the model

Our problem (but in Physics it is often not so simple)

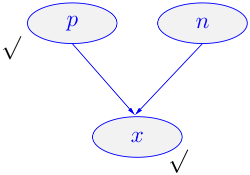


But we need some (usually indirect) knowledge about *p*

n independent Bernoulli processes

Extending the model

Our problem (but in Physics it is often not so simple)

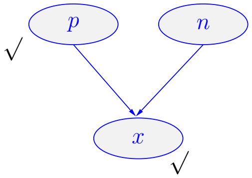


But we need some (usually indirect) knowledge about p
(Usually we do not calculate p from the fraction of white balls!)

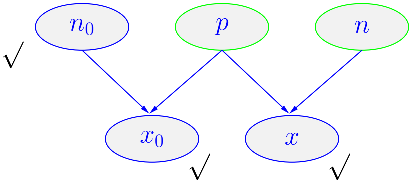
n independent Bernoulli processes

Extending the model

Our problem (but in Physics it is often not so simple)



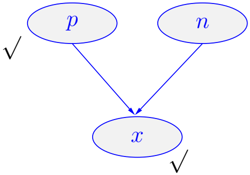
But we need some (usually indirect) knowledge about p
(Usually we <u>do not</u> calculate p from the fraction of white balls!)



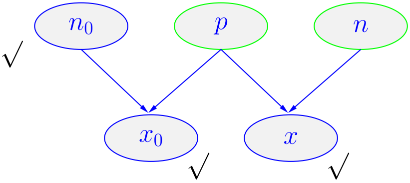
n independent Bernoulli processes

Extending the model

Our problem (but in Physics it is often not so simple)



But we need some (usually indirect) knowledge about p
(Usually we <u>do not</u> calculate p from the fraction of white balls!)



But what is n?

Extending the model

In Physics we are usually not interested in the numbers we do see, but in those which have 'physical meaning'.

Extending the model

In Physics we are usually not interested in the numbers we do see, but in those which have ‘physical meaning’.

- When we say “we are uncertain on numbers”, **we do not mean we are uncertain on the numbers we ‘see’ in our detector, but to ‘other numbers’**.

Extending the model

In Physics we are usually not interested in the numbers we do see, but in those which have 'physical meaning'.

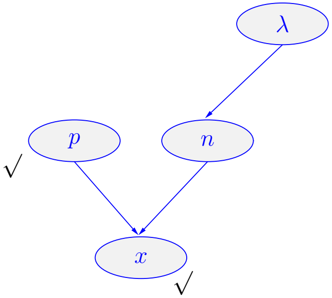
- When we say "we are uncertain on numbers", **we do not mean we are uncertain on the numbers we 'see' in our detector, but to 'other numbers'**.
- Typically $n \longleftrightarrow \lambda$.

Extending the model

In Physics we are usually not interested in the numbers we do see, but in those which have 'physical meaning'.

- When we say "we are uncertain on numbers", **we do not mean we are uncertain on the numbers we 'see' in our detector, but to 'other numbers'**.
- Typically $n \longleftrightarrow \lambda$.

Assuming for a while p well known and focusing on 'n':

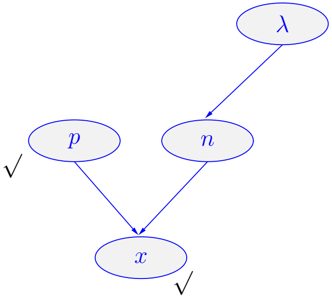


Extending the model

In Physics we are usually not interested in the numbers we do see, but in those which have ‘physical meaning’.

- When we say “we are uncertain on numbers”, **we do not mean we are uncertain on the numbers we ‘see’ in our detector, but to ‘other numbers’**.
- Typically $n \longleftrightarrow \lambda$.

Assuming for a while p well known and focusing on ‘n’:



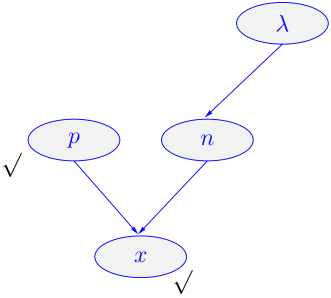
But λ is not really physical

Extending the model

In Physics we are usually not interested in the numbers we do see, but in those which have 'physical meaning'.

- When we say "we are uncertain on numbers", **we do not mean we are uncertain on the numbers we 'see' in our detector, but to 'other numbers'**.
- Typically $n \longleftrightarrow \lambda$.

Assuming for a while p well known and focusing on 'n':

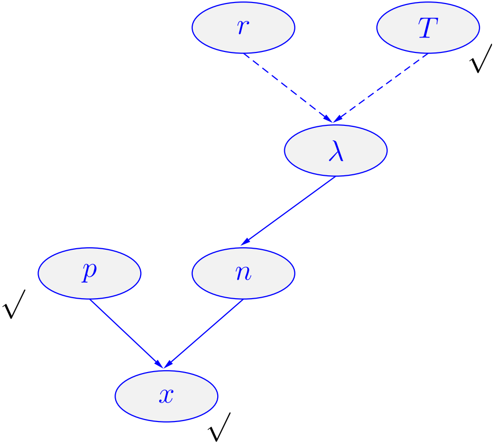


But λ is not really physical $\longrightarrow \lambda = r\,T$

(r: intensity of the Poisson process).

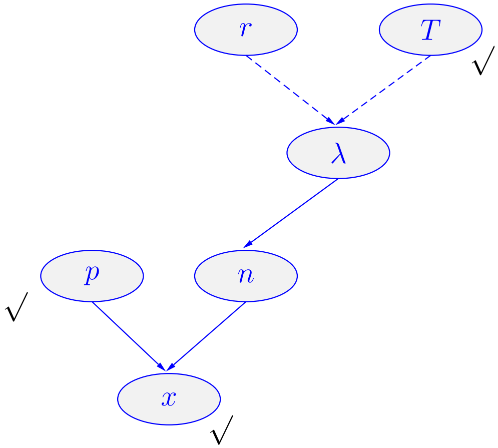
Extending the model

$\boldsymbol{\lambda} = r \cdot T$:



Extending the model

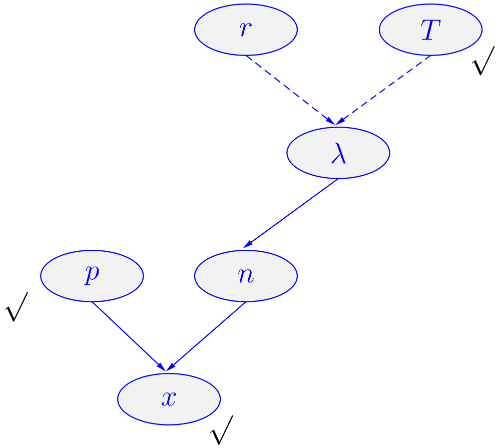
$\boldsymbol{\lambda} = r \cdot T$:



(Dashed arrows used in literature for deterministic links)

Extending the model

$\boldsymbol{\lambda} = r \cdot T$:

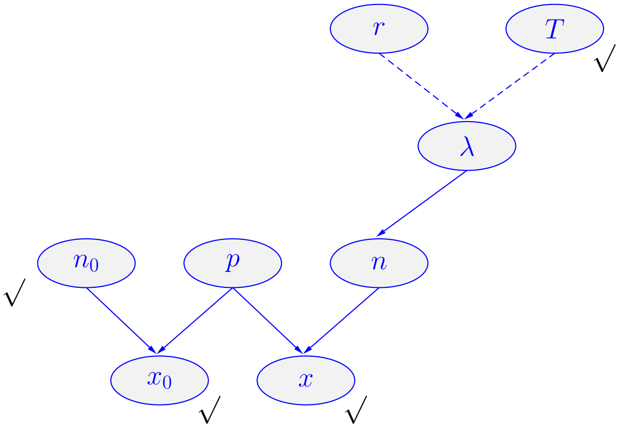


(Dashed arrows used in literature for deterministic links)

In JAGS, e.g., `lambda <- r * T;`

Extending the model

Remembering that p was got from a measurement:

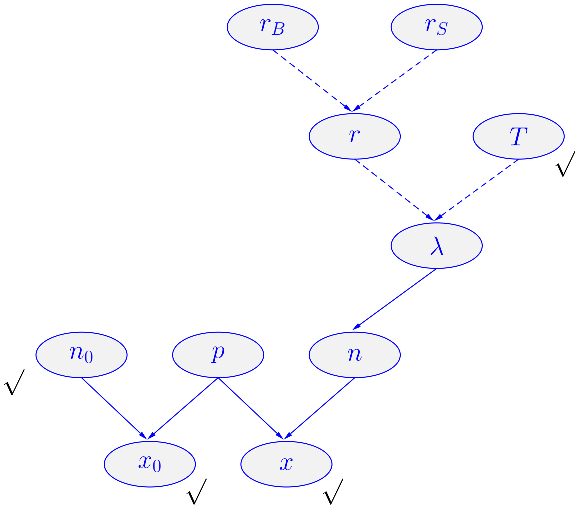


Extending the model

The rate r gets contributions from signal and background

Extending the model

The rate r gets contributions from signal and background



Extending the model

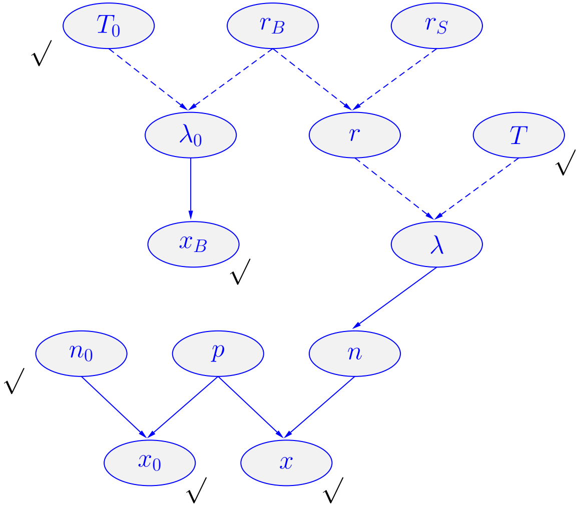
But, since $r = r_S + r_B$,

we need some independent knowledge of the background

Extending the model

But, since $r = r_S + r_B$,

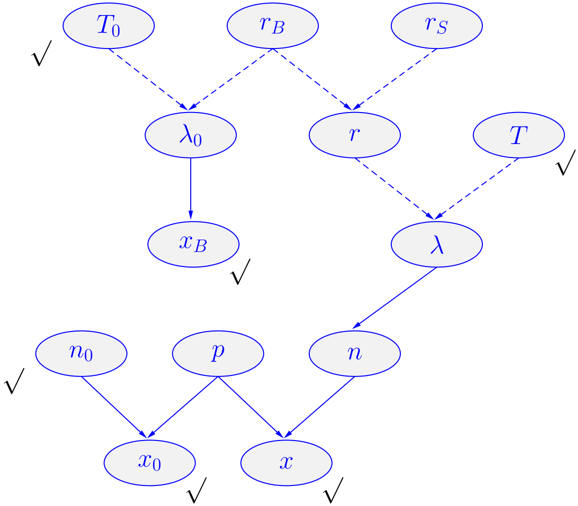
we need some independent knowledge of the background



Extending the model

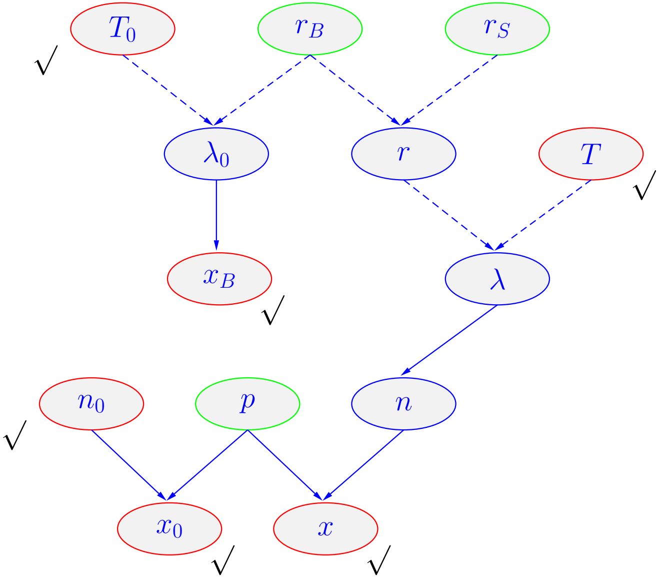
But, since $r = r_S + r_B$,

we need some independent knowledge of the background

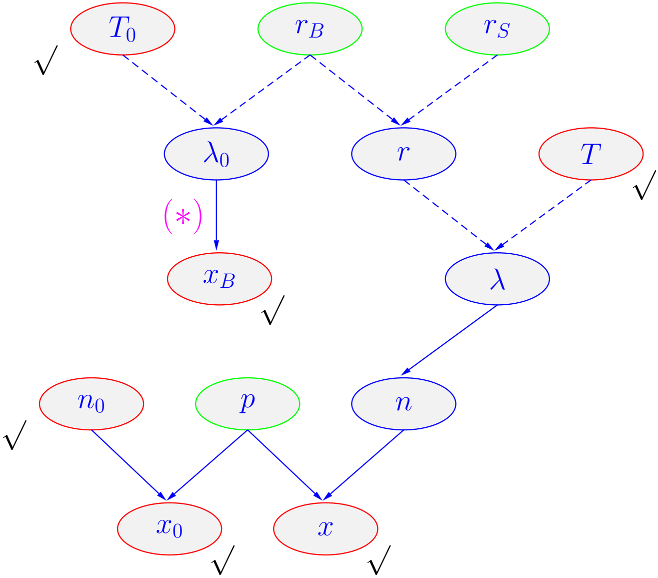


(T_0 and T assumed to be measured with sufficient accuracy)

Extending the model

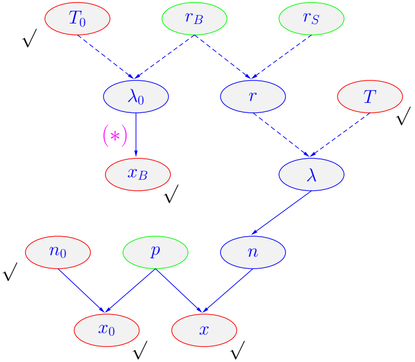


Extending the model



(∗) Assuming unity efficiency

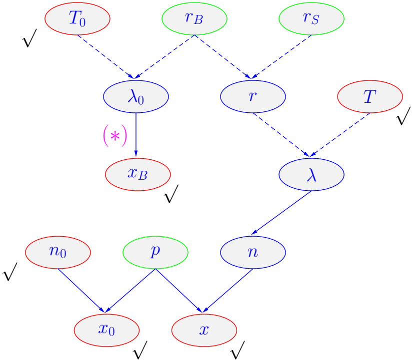
Extending the model



All the rest is a technical question of

- writing down the joint pdf of all variables;
- (re-)conditioning on the assumed/observed quantities;
- marginalize;
- getting suitable summaries, including correlations.

Extending the model



All the rest is a technical question of

- writing down the joint pdf of all variables;
- (re-)conditioning on the assumed/observed quantities;
- marginalize;
- getting suitable summaries, including correlations.

Or, more easily, use software **grounded on probability theory**, like BUGS, JAGS, BAT, etc.

Other interesting models

Once we have understood the basic reasoning, moving to other inferential/predictive problems is just an exercise
(although the math can become a bit more complicate, but this is just a technical issue).

Other interesting models

Once we have understood the basic reasoning, moving to other inferential/predictive problems is just an exercise
(although the math can become a bit more complicate, but this is just a technical issue).
For example:

- Gaussian model;
- including systematics;
- fits;
- ...

Other interesting models

Once we have understood the basic reasoning, moving to other inferential/predictive problems is just an exercise
(although the math can become a bit more complicate, but this is just a technical issue).
For example:

- Gaussian model;
- including systematics;
- fits;
- ...

For some basic examples in JAGS/rjags:

- `https://www.roma1.infn.it/~dagos/JAGS/`

Other interesting models

Once we have understood the basic reasoning, moving to other inferential/predictive problems is just an exercise
(although the math can become a bit more complicate, but this is just a technical issue).
For example:

- Gaussian model;
- including systematics;
- fits;
- ...

For some basic examples in JAGS/rjags:

- `https://www.roma1.infn.it/~dagos/JAGS/`

More detailed applications (including scripts) in

- `https://www.roma1.infn.it/~dagos/prob+stat.html`

[For BAT and BAT.jl see their web pages]

Summing up

- The probabilistic framework basically set up by Laplace$^{(*)}$ in his monumental work is healthy and grows up well.

[$^{(*)}$ See https://www.youtube.com/watch?v=8oD6eBkjF9o and relates book]

Summing up

- The probabilistic framework basically set up by Laplace$^{(*)}$ in his monumental work is healthy and grows up well.
- It is very close to the natural way of reasoning (see e.g. also Gauss).

[$^{(*)}$ See https://www.youtube.com/watch?v=8oD6eBkjF9o and relates book]

Summing up

- The probabilistic framework basically set up by Laplace$^{(*)}$ in his monumental work is healthy and grows up well.
- It is very close to the natural way of reasoning (see e.g. also Gauss).
- Its consistent application in small-complex problems was prohibitive many years ago.

[$^{(*)}$ See `https://www.youtube.com/watch?v=8oD6eBkjF9o` and relates book]

Summing up

- The probabilistic framework basically set up by Laplace$^{(*)}$ in his monumental work is healthy and grows up well.
- It is very close to the natural way of reasoning (see e.g. also Gauss).
- Its consistent application in small-complex problems was prohibitive many years ago.
- But it is now possible, thanks to progresses in applied mathematics and computation.

[$^{(*)}$ See https://www.youtube.com/watch?v=8oD6eBkjF9o and relates book]

Summing up

- The probabilistic framework basically set up by Laplace$^{(*)}$ in his monumental work is healthy and grows up well.
- It is very close to the natural way of reasoning (see e.g. also Gauss).
- Its consistent application in small-complex problems was prohibitive many years ago.
- But it is now possible, thanks to progresses in applied mathematics and computation.

⇒ Paradigmatic shift from the old formulae oriented approach (little sense to stick to old 'ah hoc' methods that had their *raison d'être* in the computational barrier).

[$^{(*)}$ See `https://www.youtube.com/watch?v=8oD6eBkjF9o` and relates book]

Summing up

- The probabilistic framework basically set up by Laplace$^{(*)}$ in his monumental work is healthy and grows up well.
- It is very close to the natural way of reasoning (see e.g. also Gauss).
- Its consistent application in small-complex problems was prohibitive many years ago.
- But it is now possible, thanks to progresses in applied mathematics and computation.

$\Rightarrow$ Paradigmatic shift from the old formulae oriented approach (little sense to stick to old 'ah hoc' methods that had their *raison d'être* in the computational barrier).

- In many cases, convenient formulae can be recovered as approximations.

[$^{(*)}$ See `https://www.youtube.com/watch?v=8oD6eBkjF9o` and relates book]

Summing up

- The probabilistic framework basically set up by Laplace$^{(*)}$ in his monumental work is healthy and grows up well.
- It is very close to the natural way of reasoning (see e.g. also Gauss).
- Its consistent application in small-complex problems was prohibitive many years ago.
- But it is now possible, thanks to progresses in applied mathematics and computation.

⇒ Paradigmatic shift from the old formulae oriented approach (little sense to stick to old 'ah hoc' methods that had their *raison d'être* in the computational barrier).

- In many cases, convenient formulae can be recovered as approximations.
- Nowadays, once you draw the graphical model of the problem,

[$^{(*)}$ See `https://www.youtube.com/watch?v=8oD6eBkjF9o` and relates book]

Summing up

- The probabilistic framework basically set up by Laplace$^{(*)}$ in his monumental work is healthy and grows up well.
- It is very close to the natural way of reasoning (see e.g. also Gauss).
- Its consistent application in small-complex problems was prohibitive many years ago.
- But it is now possible, thanks to progresses in applied mathematics and computation.

$\Rightarrow$ Paradigmatic shift from the old formulae oriented approach (little sense to stick to old 'ah hoc' methods that had their *raison d'être* in the computational barrier).

- In many cases, convenient formulae can be recovered as approximations.
- Nowadays, once you draw the graphical model of the problem,
  - you get a superior awareness of how the variables are related;

[$^{(*)}$ See `https://www.youtube.com/watch?v=8oD6eBkjF9o` and relates book]

Summing up

- The probabilistic framework basically set up by Laplace$^{(*)}$ in his monumental work is healthy and grows up well.
- It is very close to the natural way of reasoning (see e.g. also Gauss).
- Its consistent application in small-complex problems was prohibitive many years ago.
- But it is now possible, thanks to progresses in applied mathematics and computation.

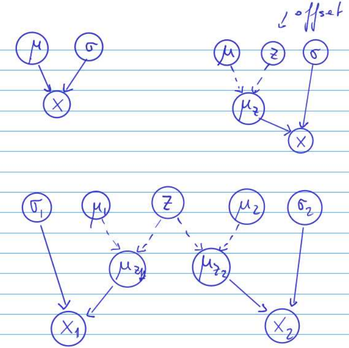
⇒ Paradigmatic shift from the old formulae oriented approach (little sense to stick to old 'ah hoc' methods that had their *raison d'être* in the computational barrier).

- In many cases, convenient formulae can be recovered as approximations.
- Nowadays, once you draw the graphical model of the problem,
  - you get a superior awareness of how the variables are related;
  - you are very close to the solution, although the analitic one is often out of reach (MCMC's rescue us!)

[$^{(*)}$ See `https://www.youtube.com/watch?v=8oD6eBkjF9o` and relates book]

Handling systematics in the probabilistic approach

(Answering to a question: diagrams show the case of uncertain offset systematics, best known to be, after proper calibration, $z = 0 \pm \sigma_z$)



1) Joint pdf : $f(\cdots\cdots)$

2) $\Rightarrow$ $f(\mu_1, \mu_2 \,|\, \ldots\ldots)$

$\Rightarrow$ details in Bayesian Reasoning (and CERN Report)

References

- As starting points, particularly recommended are the last papers, in which graphical models are systematically exploited:
  - arXiv:2001.03466 [physics.data-an]
  - arXiv:2009.04843 [q-bio.PE]
  - arXiv:2012.04455 [stat.ME]
  - arXiv:2102.11022 [stat.AP]
- Much more can be found in
  - `https://www.roma1.infn.it/~dagos/prob+stat.html`
  - `https://www.roma1.infn.it/~dagos/dott-prob_31/`