Probabilistic inference in Physics introduced with a Toy Experiment

Giulio D'Agostini

giulio.dagostini@roma1.infn.it
http://www.roma1.infn.it/~dagos/

Università La Sapienza e INFN, Roma, Italy

"Probability is the very guide of life" (*Digest* of Cicero's thought)

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

"All models are wrong but some are useful" (G. Box)

Outline

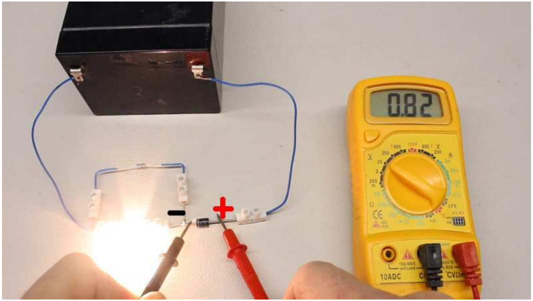
- "Science and hypothesis" (Poincaré)
- Uncertainty, probability, decision.
- Causes $\longleftrightarrow$ Effects

  *"The essential problem of the experimental method" **(Poincaré)**.*
- A toy model and its physics analogy: the six box game

  *"Probability is either referred to real cases or it is nothing" **(de Finetti)**.*
- Probabilistic approach [but . . . What is probability?]
- Basic rules of probability and Bayes rule.
- Bayesian inference and its graphical representation:
  $\Rightarrow$ Bayesian networks
- From ball and boxes to real measurements
- Conclusions

What is measurement?

What is measurement?

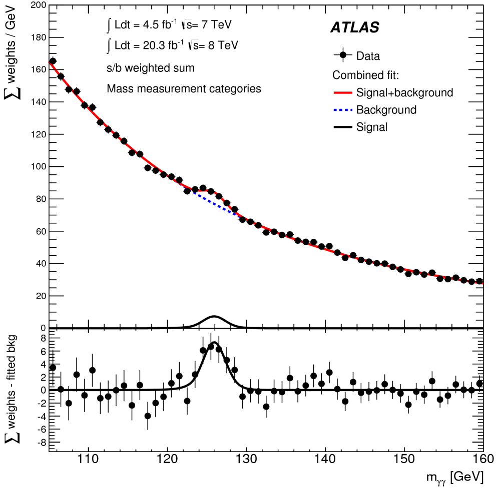


What is measurement?

What is measurement?

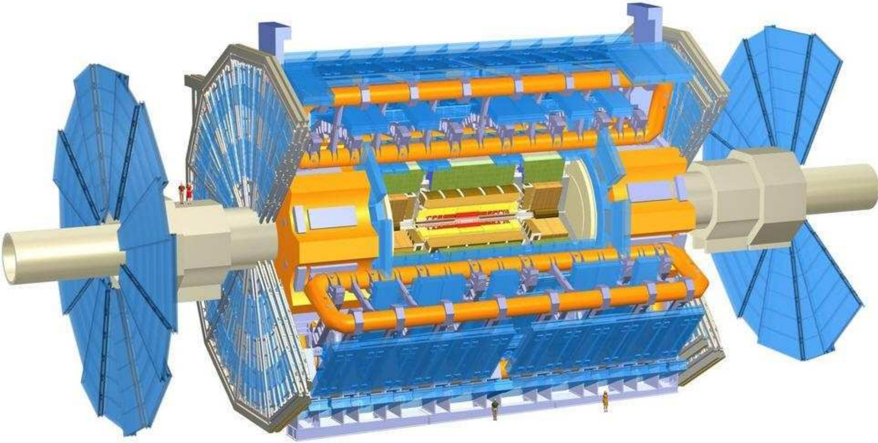
What is measurement?

Higgs $\to \gamma\gamma$



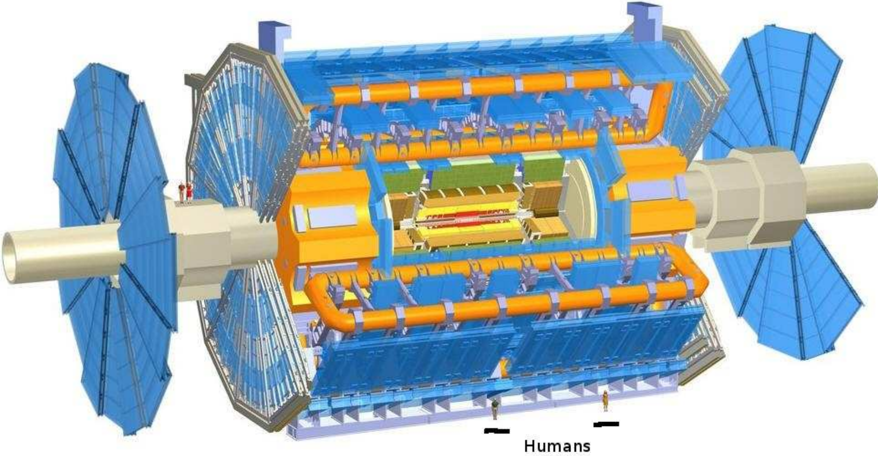
What is measurement?

ATLAS Experiment at LHC



What is measurement?

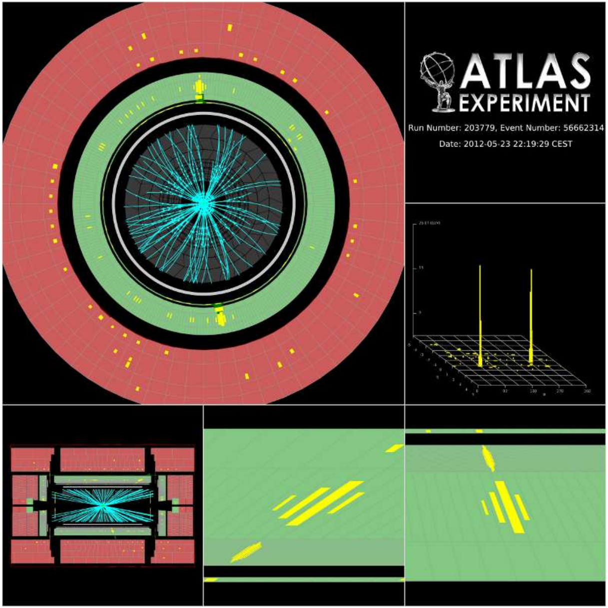
ATLAS Experiment at LHC [length: 46 m; ⌀ 25 m]



$\approx$ 3000 km cables

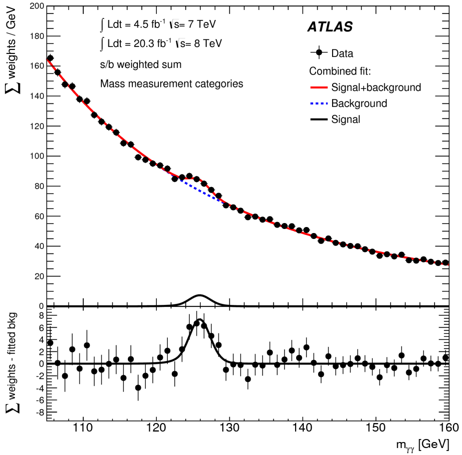
$\approx$ 7000 tonnes $\approx$ 100 millions electronic channels

What is measurement?



What is measurement?

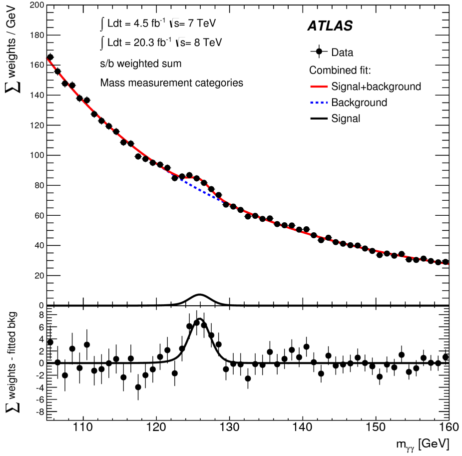
Higgs $\to \gamma\gamma$



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

What is measurement?

Higgs $\to \gamma\gamma$



$\Rightarrow$ { Mass
Production rate

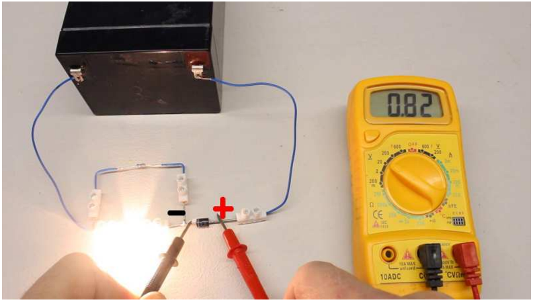
Quite indirect measurements of something we do not "see"!

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 balance

Equilibrium:

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

From the reading to the value of the mass:

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

Measuring a mass on a balance

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

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

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 homogenous.
- Moreover we have to consider centrifugal effects
- ...and even the effect from the Moon

Measuring a mass on a balance

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

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

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

Certainly not to watch our weight 

Measuring a mass on a balance

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

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

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

Certainly not to watch our weight

But think about it!

Measuring a mass on a balance

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

Dependence on 'k':

- temperature
- non linearity
- ...

Measuring a mass on a balance

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

Dependence on 'k':

- temperature
- non linearity
- ...

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

- left to your imagination...

Measuring a mass on a balance

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

Dependence on 'k':

- temperature
- non linearity
- ...

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

- left to your imagination...

$+$ **randomic effects:**

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

Measuring a mass on a balance

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

Dependence on 'k':

- temperature
- non linearity
- ...

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

- left to your imagination...

$+$ **randomic effects:**

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

$\Rightarrow$ m??

Sources of uncertainties (from ISO GUM)

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

Sources of uncertainties (from ISO GUM)

6. *finite instrument resolution or discrimination threshold;*
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.*

    $\rightarrow$ *"statistical errors"*

Note

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

Pure empirical information?

A number, outside a contest, and denuted of all information the physicist or engineer has about its ‘production’ provides little (or zero) information: it is not a measurement.

Pure empirical information?

A number, outside a contest, and denuted of all information the physicist or engineer has about its 'production' provides little (or zero) information: it is not a measurement.

mistrust the

Dogma of the Immaculate Observation!

Comparing hypotheses

We do measurements not only to 'estimate' the numeric value of a quantity.

Experimental observations are also used in order to

- "check hypotheses"
  (a generic expression that needs clarification. . .)

Comparing hypotheses

We do measurements not only to 'estimate' the numeric value of a quantity.

Experimental observations are also used in order to

- "check hypotheses"
  (a generic expression that needs clarification. . .)
- make decisions accordingly

Comparing hypotheses

We do measurements not only to 'estimate' the numeric value of a quantity.

Experimental observations are also used in order to

- "check hypotheses"
  (a generic expression that needs clarification...)
- make decisions accordingly

Diagnostics, reliability, etc.

Comparing hypotheses

We do measurements not only to 'estimate' the numeric value of a quantity.

Experimental observations are also used in order to

- "check hypotheses"
  (a generic expression that needs clarification. . .)
- make decisions accordingly

Diagnostics, reliability, etc.

Diagnostics concerning health helps to clarify the issues ⇒

AIDS test

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

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

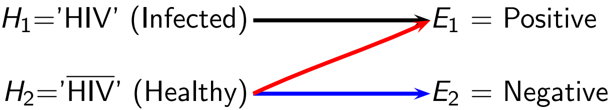
H_1='HIV' (Infected) $\quad$ E_1 = Positive

H_2='$\overline{\text{HIV}}$' (Healthy) $\quad$ E_2 = Negative

AIDS test

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

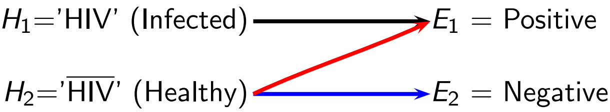
$$
\begin{aligned}
P(\text{Pos} \,|\, \text{HIV}) &= 100\% \\
P(\text{Pos} \,|\, \overline{\text{HIV}}) &= 0.2\% \\
P(\text{Neg} \,|\, \overline{\text{HIV}}) &= 99.8\%
\end{aligned}
$$



AIDS test

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

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



Result: ⇒ Positive

AIDS test

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

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

? H_1='HIV' (Infected) ⟵ E_1 = Positive

? H_2='$\overline{\text{HIV}}$' (Healthy) E_2 = Negative

Result: $\Rightarrow$ Positive
Infected or healthy?

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 healthy, since it was practically impossible that an healthy 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 healthy, since it was practically impossible that an healthy 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 healthy, since it was practically impossible that an healthy 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 healthy, since it was practically impossible that an healthy 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=Healthy is ruled out with 99.8% C.L."

?

AIDS test: how to interpret the result?

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

- "It is practically impossible that the person is healthy, since it was practically impossible that an healthy 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=Healthy is ruled out with 99.8% C.L."

NO

Instead, $P(\overline{\text{HIV}} \mid \text{Pos, random Italian}) \approx 45\%$
(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 healthy, since it was practically impossible that an healthy 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=Healthy is ruled out with 99.8% C.L."

NO

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

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

AIDS test

???

Where is the problem?

AIDS test

???

Where is the problem?

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

AIDS test

???

Where is the problem?

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

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

AIDS test

???

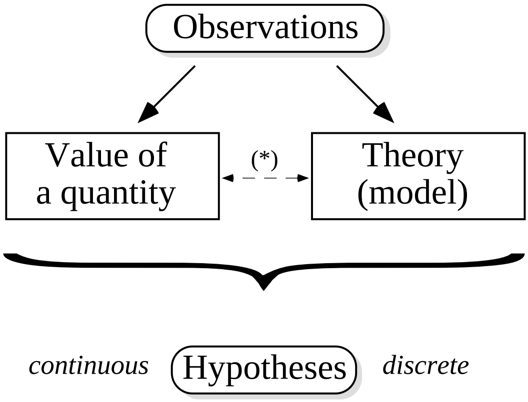
Where is the problem?

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

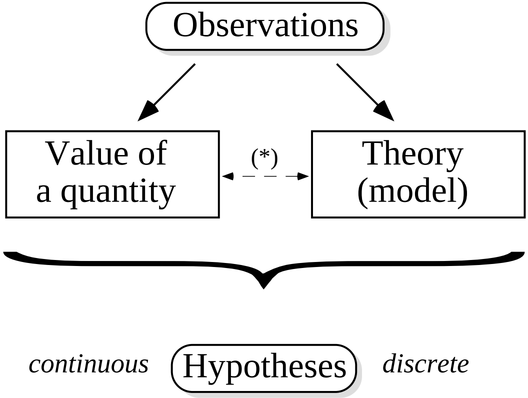
. . . and in these issues intuition can be fallacious!

⇒ A sound formal guidance can rescue us

Learning from data

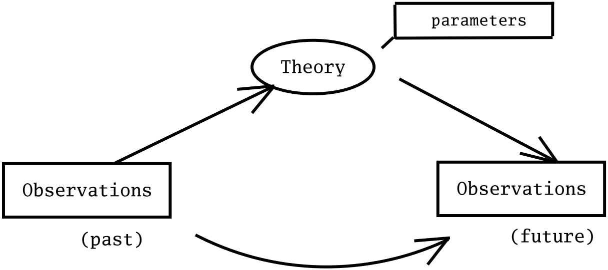


Learning from data



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

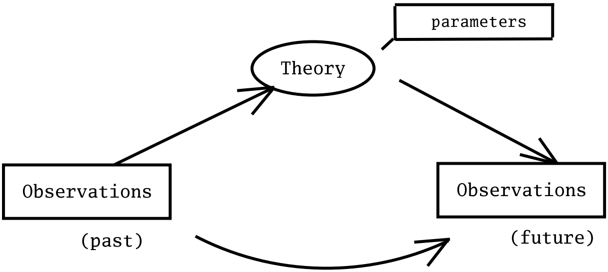
From past to future



Our task:

- Describe/understand the physical world
  ⇒ inference of laws and their parameters
- Predict observations
  ⇒ forecasting

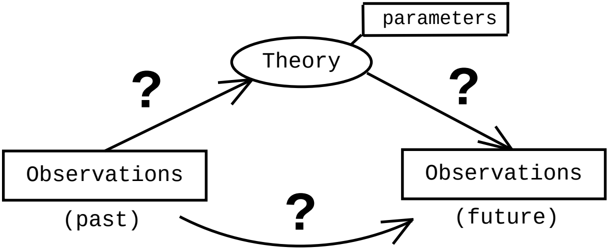
From past to future



Process

- neither automatic
- nor purely contemplative
  - $\rightarrow$ 'scientific method'
  - $\rightarrow$ planned experiments ('actions') $\Rightarrow$ decision.

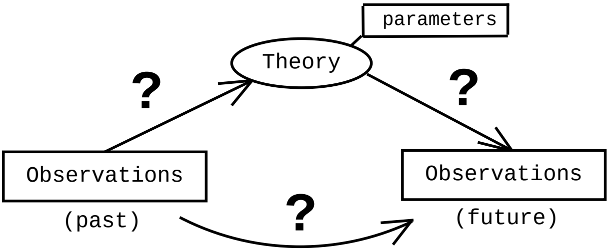
From past to future



⇒ Uncertainty:

1. Given the past observations, in general we are not sure about the theory parameters (and/or the theory itself)
2. Even if we were sure about theory and parameters, there could be internal (e.g. Q.M.) or external effects (initial/boundary conditions, 'errors', etc) that make the forecasting uncertain.

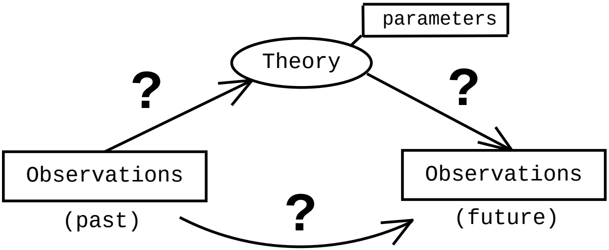
From past to future



⇒ Decision

- What is be best action ('experiment') to take in order 'to be confident' that what we would like will occur?
  (Decision issues always assume uncertainty about future outcomes.)
- Before tackling problems of decision we need to learn to reason about uncertainty, possibly in a quantitative way.

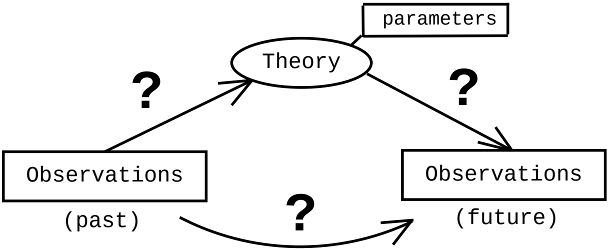
From past to future



Deep reason of uncertainty

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

From past to future



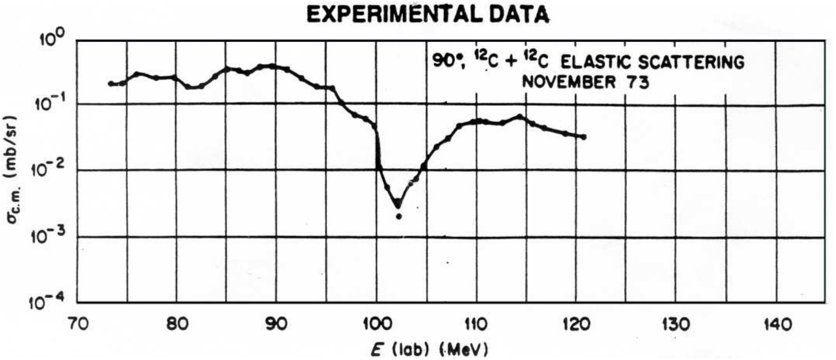
Deep reason of uncertainty

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

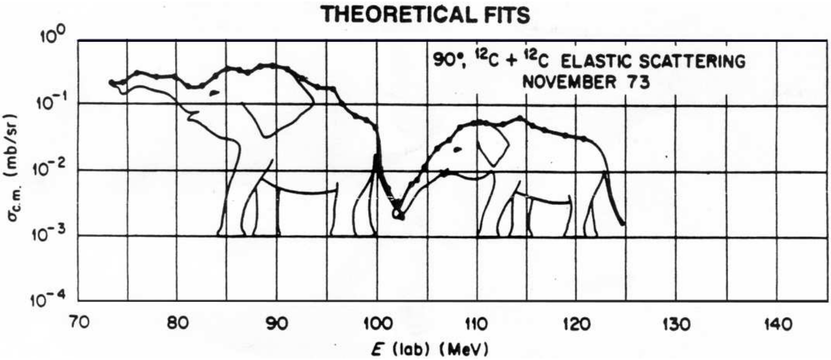
$\Longrightarrow$ **Uncertainty about causal connections**

CAUSE $\Longleftrightarrow$ **EFFECT**

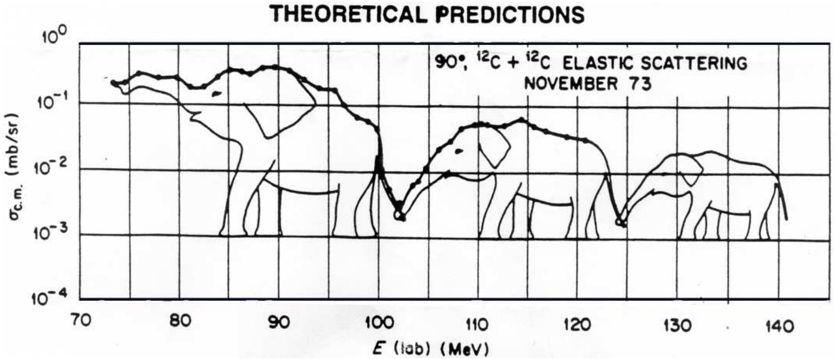
Inferential-predictive process



Inferential-predictive process

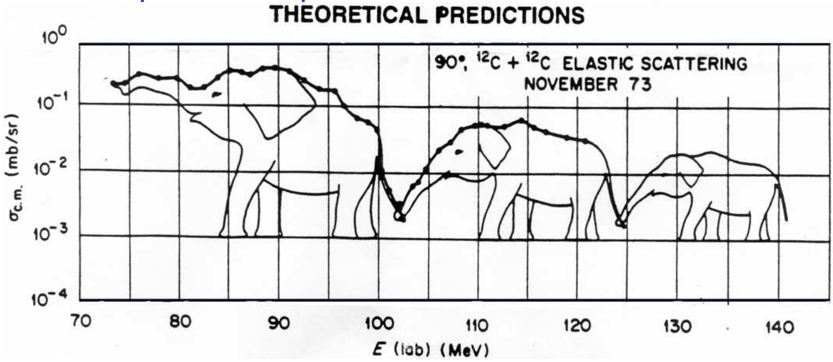


Inferential-predictive process



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

Inferential-predictive process



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

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

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

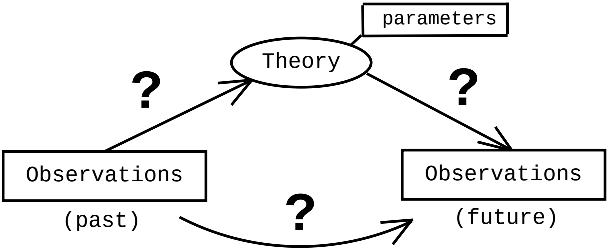
2011 IgNobel prize in Mathematics

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

2011 IgNobel prize in Mathematics

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

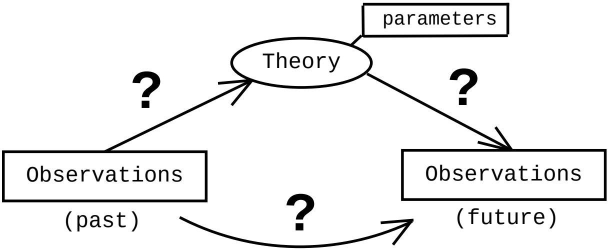
Deep source of uncertainty



Uncertainty:

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

Deep source of uncertainty



Uncertainty:

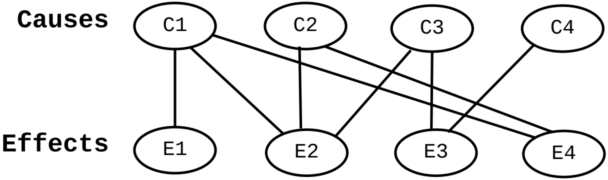
Theory — **?** ⟶ Future observations
Past observations — **?** ⟶ Theory
Theory — **?** ⟶ Future observations

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

CAUSE $\Longleftrightarrow$ **EFFECT**

Causes → effects

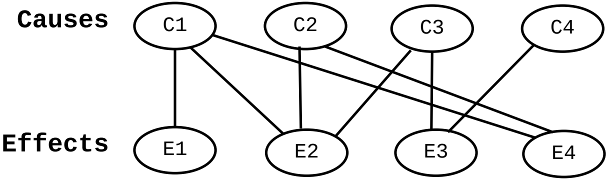
The same *apparent* cause might produce several,different effects



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

Causes → effects

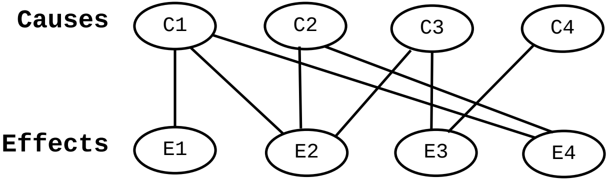
The same *apparent* cause might produce several,different effects



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

Causes → 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.

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

"Now, these problems are classified as *probability of causes*, and are most interesting of all 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.

The "essential problem" of the Sciences

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

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

(H. Poincaré – *Science and Hypothesis*)

The “essential problem” of the Sciences

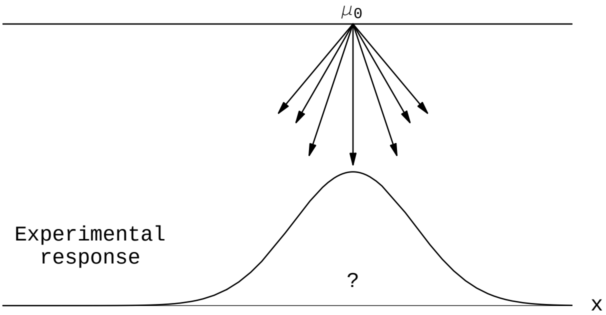
“Now, these problems are classified as *probability of causes*, and are most interesting of all their scientific applications. I play at *écarté* with a gentleman whom I know to be perfectly honest. What is the chance that he turns up the king? It is $1/8$. This is a problem of the probability of effects.

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*)

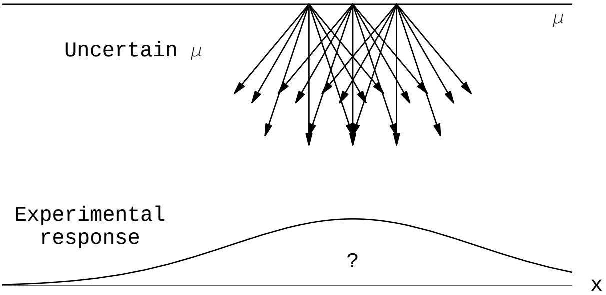
Why we (or most of us) have not been taught how to tackle this kind of problems?

From ‘true value’ to observations



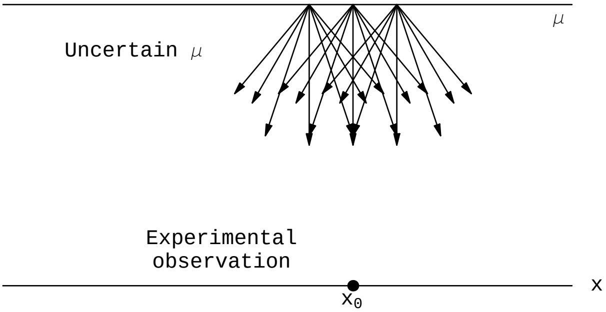
Given μ (exactly known) we are uncertain about x

From 'true value' to observations



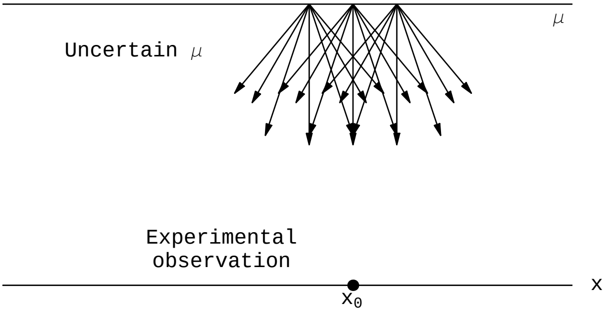
Uncertainty about μ makes us more uncertain about x

. . . and back: Inferring a true value



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

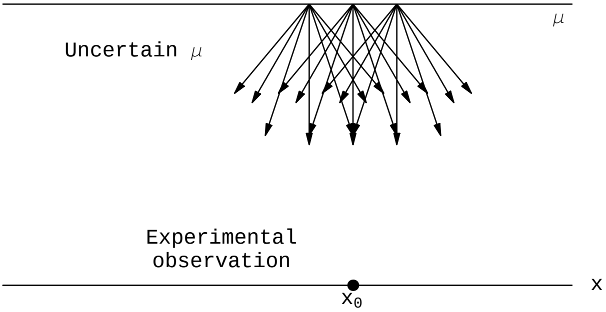
. . . and back: Inferring a true value



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

"data uncertainty" ?

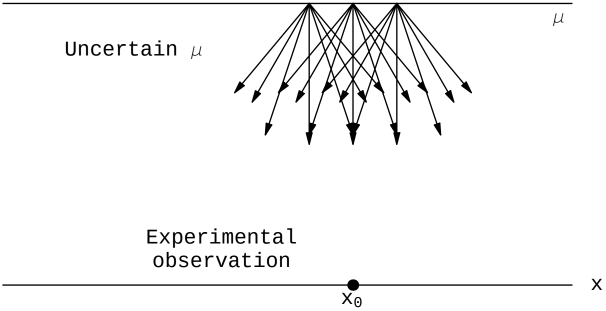
... and back: Inferring a true value



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

"data uncertainty" ? Data corrupted?

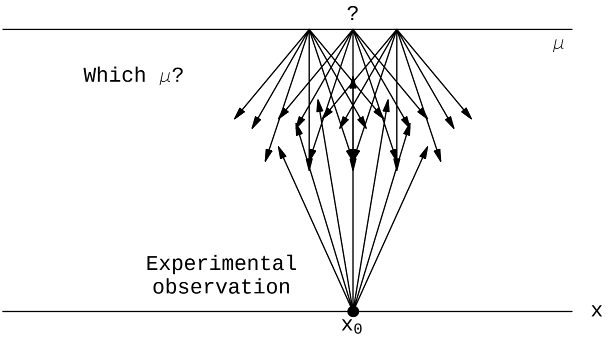
. . . and back: Inferring a true value



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

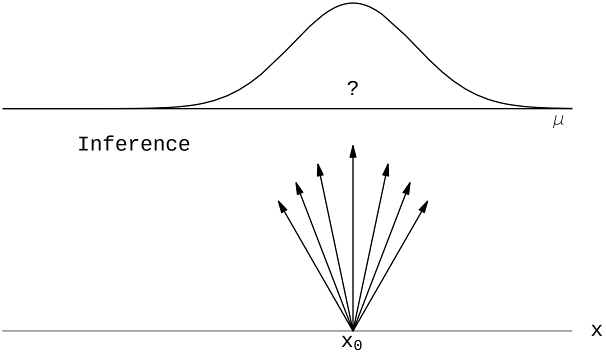
“data uncertainty” ? Data corrupted?
Even if the data were corrupted, the data were the corrupted data**!!** . . .

. . . and back: Inferring a true value



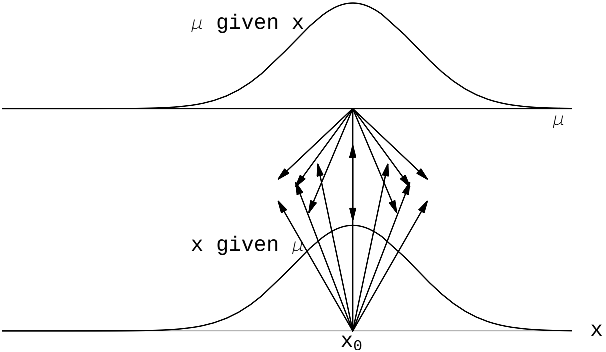
Where does the observed value of x comes from?

. . . 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.

A very simple experiment

Let's make an experiment

A very simple experiment

Let's make an experiment

- Here
- Now

A very simple experiment

Let's make an experiment

- Here
- Now

For simplicity

- μ can assume only six possibilities:

$$\mathbf{0}, \mathbf{1}, \ldots, \mathbf{5}$$

- x is binary:

$$\mathbf{0}, \mathbf{1}$$

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

A very simple experiment

Let's make an experiment

- Here
- Now

For simplicity

- μ can assume only six possibilities:

$$\mathbf{0}, \mathbf{1}, \ldots, \mathbf{5}$$

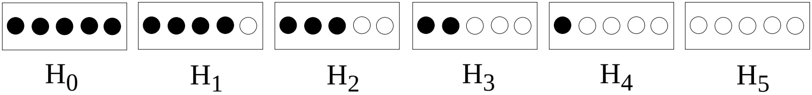
- x is binary:

$$\mathbf{0}, \mathbf{1}$$

[(1, 2); Black/White; Yes/Not; ...]

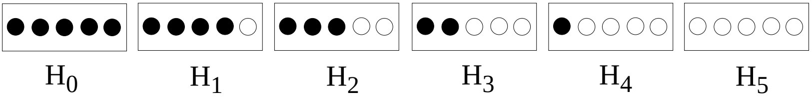
⇒ Later we shall make μ continous.

Which box? Which ball?



Let us take randomly one of the boxes.

Which box? Which ball?



Let us take randomly one of the boxes.

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

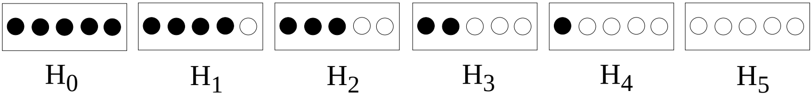
(a) Which box have we chosen, $H_0, H_1, \ldots, H_5$?

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

Our certainties:

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

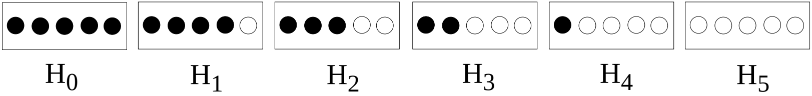
Which box? Which ball?



Let us take randomly one of the boxes.

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

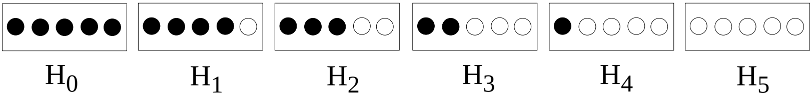
Which box? Which ball?



Let us take randomly one of the boxes.

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

Which box? Which ball?



Let us take randomly one of the boxes.

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

The toy inferential experiment

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

The toy inferential experiment

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

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

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

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

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

Where is probability?

We all agree that the experimental results change

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

Where is probability?

We all agree that the experimental results change

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

although the box composition remains unchanged ('extractions followed by reintroduction').

Where is probability?

We all agree that the experimental results change

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

although the box composition remains unchanged ('extractions followed by reintroduction').

Where is the probability?

Where is probability?

We all agree that the experimental results change

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

although the box composition remains unchanged ('extractions followed by reintroduction').

Where is the probability?

Certainly not *in* the box!

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)

$$P(E) \longrightarrow P(E \mid I_s)$$

where I_s is the information available to *subject* s.

What are we talking about?

"Given the state of our knowledge about everything that could possible have any bearing on the coming true...

What are we talking about?

"Given the state of our knowledge about everything that could possible have any bearing on the coming true. . . the numerical probability P of this event is to be a real number by the indication of which we try in some cases to setup a quantitative measure of the strength of our conjecture or anticipation, founded on the said knowledge, that the event comes true"

(Schrödinger, 1947)

What are we talking about?

"Given the state of our knowledge about everything that could possible have any bearing on the coming true... the numerical probability P of this event is to be a real number by the indication of which we try in some cases to setup a quantitative measure of the strength of our conjecture or anticipation, founded on the said knowledge, that the event comes true"

$\Rightarrow$ How much we believe something

What are we talking about?

"Given the state of our knowledge about everything that could possible have any bearing on the coming true. . . the numerical probability P of this event is to be a real number by the indication of which we try in some cases to setup a quantitative measure of the strength of our conjecture or anticipation, founded on the said knowledge, that the event comes true"

$\rightarrow$ 'Degree of belief' $\leftarrow$

Beliefs and 'coherent' bets

Remarks:

- Subjective does not mean arbitrary!

Beliefs and 'coherent' bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?

Beliefs and 'coherent' bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet

Beliefs and 'coherent' bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet:
  - you state the odds according on your beliefs;
  - somebody else will choose the direction of the bet.

Beliefs and ‘coherent’ bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet:
  - you state the odds according on your beliefs;
  - somebody else will choose the direction of the bet.

“His [Bouvard] calculations give him the mass of Saturn as 3,512th part of that of the sun. Applying my probabilistic formulae to these observations, I find that the odds are 11,000 to 1 that the error in this result is not a hundredth of its value.” *(Laplace)*

Beliefs and 'coherent' bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet:
  - you state the odds according on your beliefs;
  - somebody else will choose the direction of the bet.

"His [Bouvard] calculations give him the mass of Saturn as 3,512th part of that of the sun. Applying my probabilistic formulae to these observations, I find that the odds are 11,000 to 1 that the error in this result is not a hundredth of its value." *(Laplace)*

Beliefs and 'coherent' bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet:
  - you state the odds according on your beliefs;
  - somebody else will choose the direction of the bet.

"His [Bouvard] calculations give him the mass of Saturn as 3,512th part of that of the sun. Applying my probabilistic formulae to these observations, I find that the odds are 11,000 to 1 that the error in this result is not a hundredth of its value." (Laplace)

$\rightarrow P(3477 \leq M_{Sun}/M_{Sat} \leq 3547 \,|\, I(\text{Laplace})) = 99.99\%$

Beliefs and 'coherent' bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet:
  - you state the odds according on your beliefs;
  - somebody else will choose the direction of the bet.

Is a 'conventional' 95% C.L. lower/upper bound a 19 to 1 bet?

Beliefs and 'coherent' bets

Remarks:

- Subjective <u>does not mean</u> arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet:
  - you state the odds according on your beliefs;
  - somebody else will choose the direction of the bet.

Is a 'conventional' 95% C.L. lower/upper bound a 19 to 1 bet?

NO!

Beliefs and 'coherent' bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet:
  - you state the odds according on your beliefs;
  - somebody else will choose the direction of the bet.

Is a 'conventional' 95% C.L. lower/upper bound a 19 to 1 bet?

- It does not imply one has to be 95% confident on something!
- If you do so you are going to make a bad bet!

Beliefs and 'coherent' bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet:
  - you state the odds according on your beliefs;
  - somebody else will choose the direction of the bet.

Is a 'conventional' 95% C.L. lower/upper bound a 19 to 1 bet?

For more on the subject
see `http://arxiv.org/abs/1112.3620`
and references therein.

Standard textbook definitions

$$p = \frac{\#\,\text{favorable cases}}{\#\,\text{possible equiprobable cases}}$$

$$p = \frac{\#\,\text{times the event has occurred}}{\#\,\text{independent trials under same conditions}}$$

Standard textbook definitions

It is easy to check that 'scientific' definitions suffer of circularity

$$p = \frac{\#\,\text{favorable cases}}{\#\,\text{possible equiprobable cases}}$$

$$p = \frac{\#\,\text{times the event has occurred}}{\#\,\text{independent trials under same conditions}}$$

Standard textbook definitions

It is easy to check that 'scientific' definitions suffer of circularity

$$p = \frac{\#\,\text{favorable cases}}{\#\,\text{possible equally possible cases}}$$

$$p = \frac{\#\,\text{times the event has occurred}}{\#\,\text{independent trials under same conditions}}$$

Note!: *"lorsque rien ne porte à croire que l'un de ces cas doit arriver plutot que les autres" (Laplace)*

Replacing 'equi-probable' by 'equi-possible' is just cheating students (as I did in my first lecture on the subject. . .).

Standard textbook definitions

It is easy to check that 'scientific' definitions suffer of circularity, plus other problems

$$p = \frac{\#\,\text{favorable cases}}{\#\,\text{possible } \textcolor{red}{\text{equiprobable}} \text{ cases}}$$

$$p = \lim_{n\to\infty} \frac{\#\,\text{times the event has occurred}}{\#\,\textcolor{red}{\text{independent}} \text{ trials under } \textcolor{red}{\text{same condition}}}$$

Future ⇔ Past (belief!)

$n \to \infty$: → *"usque tandem?"*
→ *"in the long run we are all dead"*
→ It limits the range of applications

‘Definitions’ → evaluation rules

Very useful evaluation rules

$$A) \quad p = \frac{\#\,\text{favorable cases}}{\#\,\text{possible equiprobable cases}}$$

$$B) \quad p = \frac{\#\,\text{times the event has occurred}}{\#\text{independent trials under same condition}}$$

If the implicit beliefs are well suited for each case of application.

‘Definitions’ $\rightarrow$ evaluation rules

Very useful evaluation rules

$$A) \quad p = \frac{\#\,\text{favorable cases}}{\#\,\text{possible equiprobable cases}}$$

$$B) \quad p = \frac{\#\,\text{times the event has occurred}}{\#\text{independent trials under same condition}}$$

If the implicit beliefs are well suited for each case of application.

BUT they cannot define the concept of probability!

‘Definitions’ $\rightarrow$ evaluation rules

Very useful evaluation rules

$$A) \quad p = \frac{\#\,\text{favorable cases}}{\#\,\text{possible equiprobable cases}}$$

$$B) \quad p = \frac{\#\,\text{times the event has occurred}}{\#\text{independent trials under same condition}}$$

In the probabilistic approach we are following

- Rule A is recovered immediately (under the assumption of equiprobability, when it applies).
- Rule B results from a theorem of Probability Theory (under well defined assumptions).

'Definitions' $\rightarrow$ evaluation rules

Very useful evaluation rules

$$A) \quad p = \frac{\#\,\text{favorable cases}}{\#\,\text{possible equiprobable cases}}$$

$$B) \quad p = \frac{\#\,\text{times the event has occurred}}{\#\text{independent trials under same condition}}$$

In the probabilistic approach we are following

- Rule A is recovered immediately (under the assumption of equiprobability, when it applies).
- Rule B results from a theorem of Probability Theory (under well defined assumptions): $\Rightarrow$ Laplace's rule of succession

Mathematics of beliefs

The good news:

> *The basic laws of degrees of belief are the same we get from the inventory of favorable and possible cases, or from events occurred in the past.*

It can be proved that

> *the requirement of **coherence** leads to the famous 4 basic rules* $\Longrightarrow$

[Details skipped. . .]

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 on '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')

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

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

Mathematics of beliefs

An even better news:

The fourth basic rule
can be fully exploided!

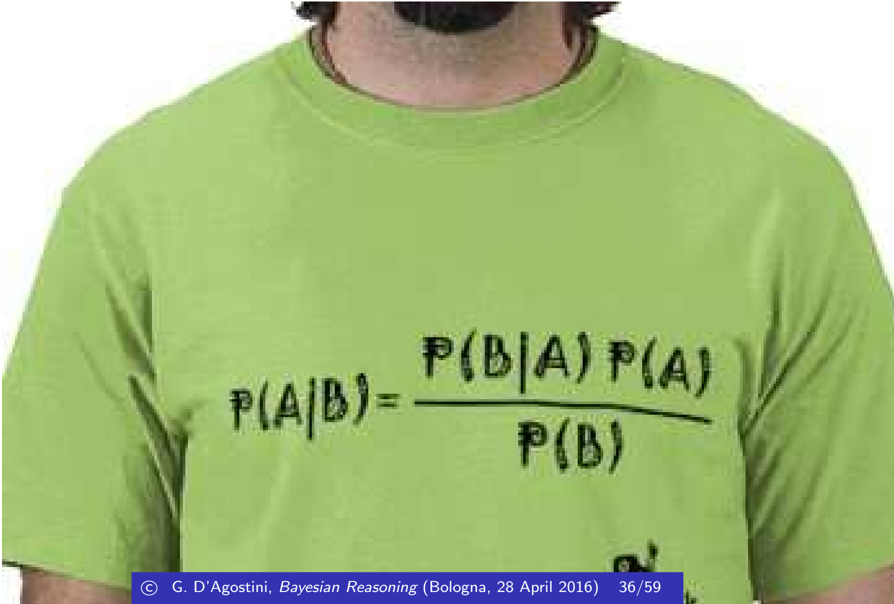
Mathematics of beliefs

An even better news:

The fourth basic rule
can be fully exploided!

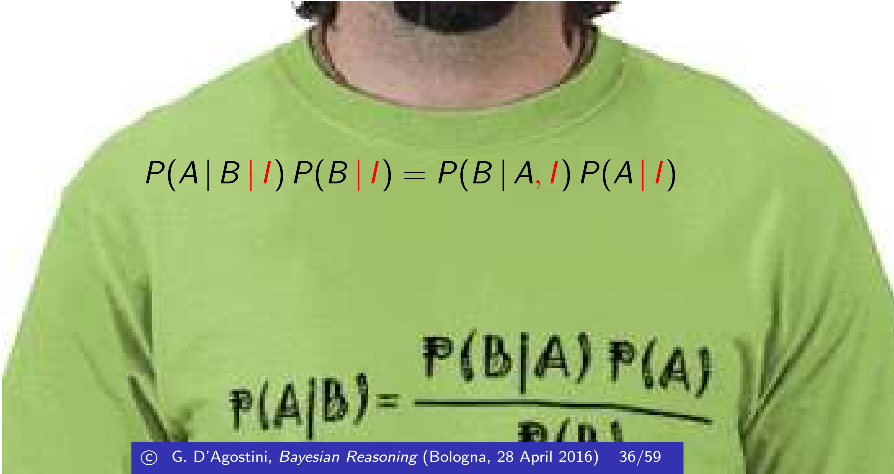
(Liberated by a curious ideology that forbits its use)

A simple, powerful formula

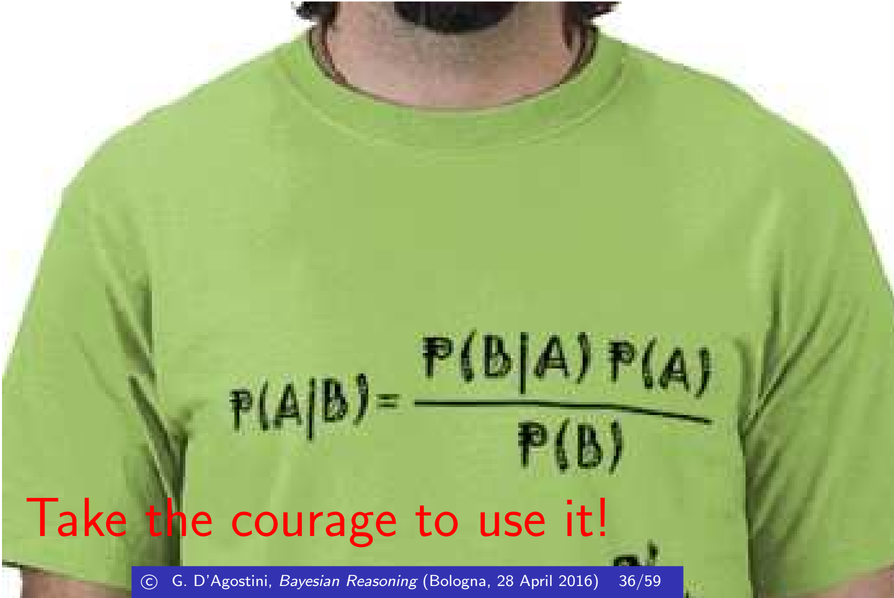


A simple, powerful formula

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

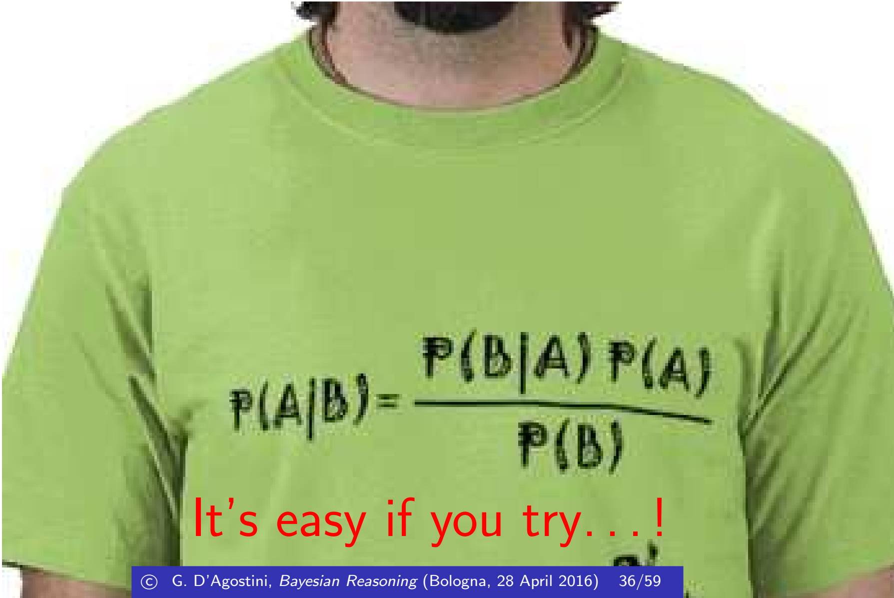


A simple, powerful formula



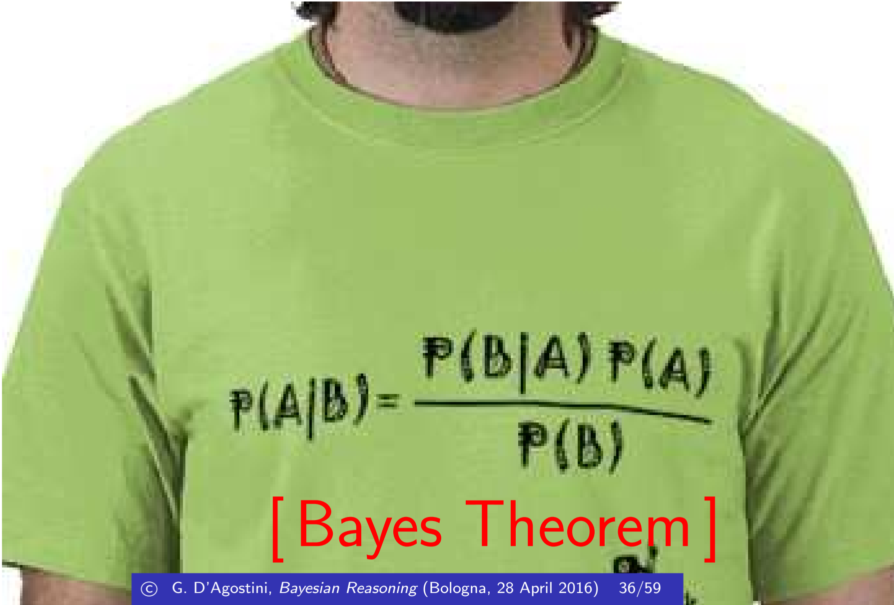
Take the courage to use it!

A simple, powerful formula



It's easy if you try. . . !

A simple, powerful formula



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*."

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

(*Philosophical Essai on Probabilities*)

Laplace's "Bayes Theorem"

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

"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 'fondamental rules'.

Laplace's "Bayes Theorem"

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

"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 'fondamental 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"

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

"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 'fondamental rules'.

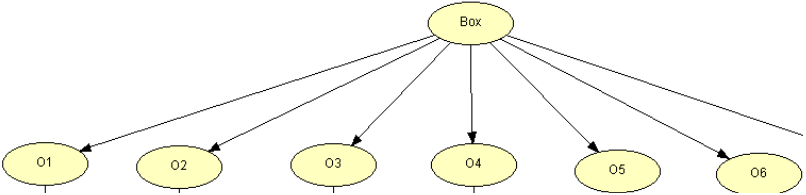
Note: denominator is just a normalization factor.

$$\Rightarrow \qquad P(C_i \mid E) \propto P(E \mid C_i)\, P(C_i)$$

Most convenient way to remember Bayes theorem

Cause-effect representation

box content $\rightarrow$ observed color

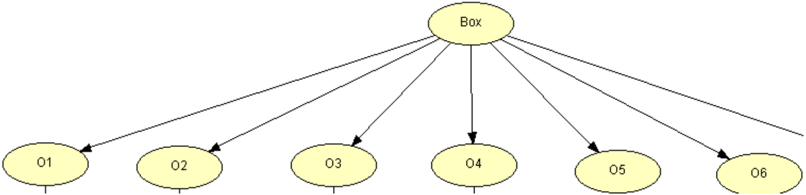


$$P(B^{(1)} \mid H_j), \quad P(B^{(2)} \mid H_j), \ldots$$

$$P(W^{(1)} \mid H_j), \quad P(W^{(2)} \mid H_j), \ldots$$

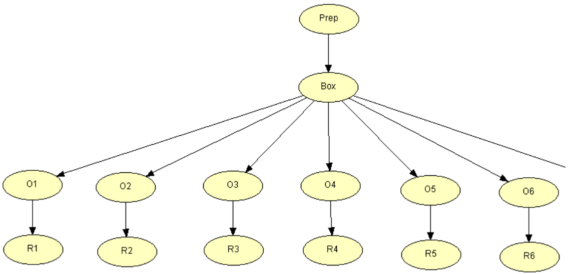
Cause-effect representation

box content $\rightarrow$ observed color



An effect might be the cause of another effect $\Longrightarrow$

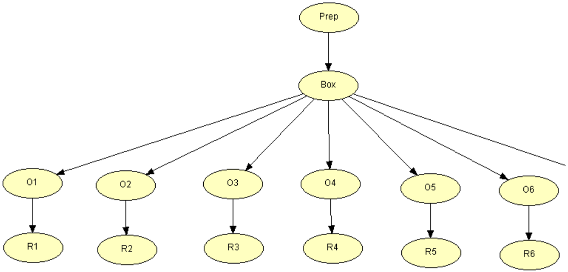
A network of causes and effects



A network of causes and effects

Preparation 'node' models prior knowledge about Box.

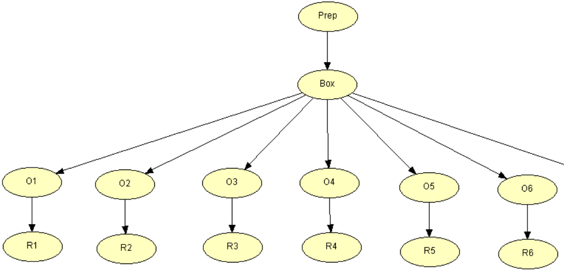
$\Rightarrow P(H_j \,|\, \text{Prep}_k)$



A network of causes and effects

Preparation 'node' models prior knowledge about Box.

$\Rightarrow P(H_j \,|\, \text{Prep}_k)$



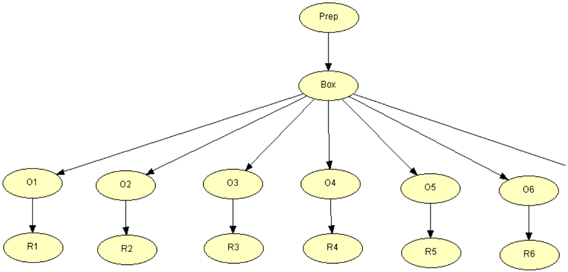
R_i model extra uncertainty in cascade.

$\Rightarrow P(W_R \,|\, W)$, $P(B_R \,|\, W)$, etc.

A network of causes and effects

Preparation 'node' models prior knowledge about Box.

$\Rightarrow P(H_j \,|\, \text{Prep}_k)$



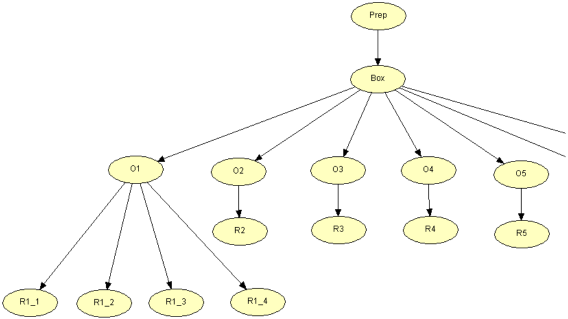
R_i model extra uncertainty in cascade.

$\Rightarrow P(W_R \,|\, W)$, $P(B_R \,|\, W)$, etc.

We shall also include multi-reporters and systematic effects

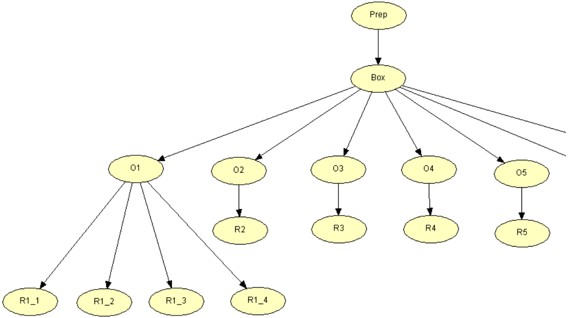
Multi-reporters

Multiple ‘testimonies’ of the same empirical fact.



Multi-reporters

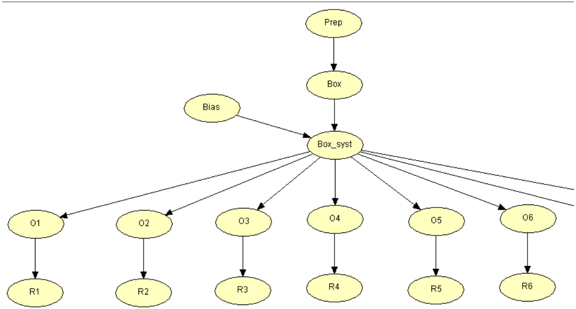
Multiple 'testimonies' of the same empirical fact.



⇒ Our belief on O_1 being Black or White will depend on the consistencies of the 'testimonies'

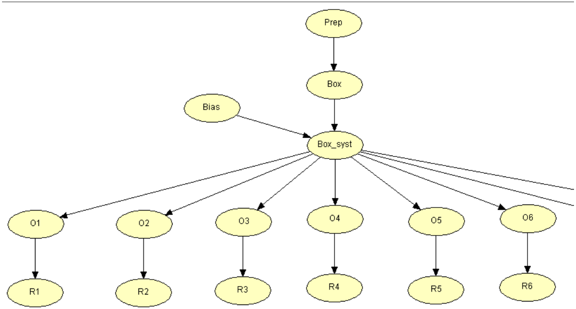
Systematic effects

The box content could be biased. . .



Systematic effects

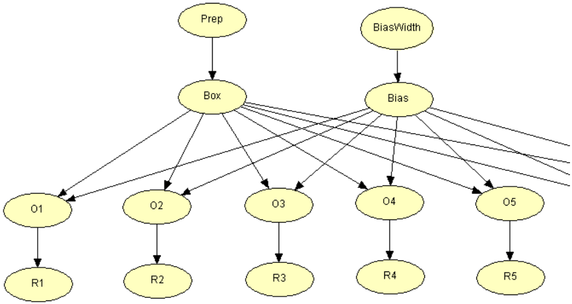
The box content could be biased. . .



. . . if one or more balls of either color might be added to the original box content

Systematic effects

The box content could be biased. . .



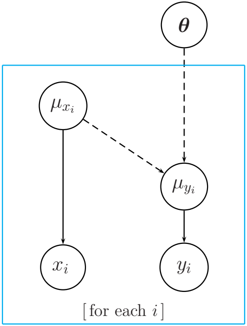
[technical implementation of the bias – logically equivalent]

Graphical models

The importance of graphical models is that

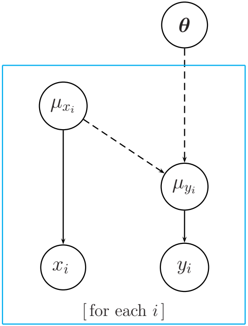
⇒ Nowadays, thanks to progresses in mathematics and computing, drawing the problem as a 'belief network' is more than 1/2 step towards its solution!

A different way to view fit issues

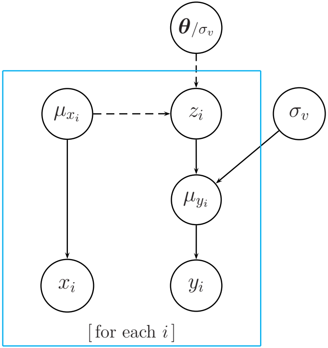


Determistic link μ_x's to μ_y's
Probabilistic links $\mu_x \rightarrow x$, $\mu_y \rightarrow y$
(errors on both axes!)
$\Rightarrow$ aim of fit: $\{\mathbf{x}, \mathbf{y}\} \rightarrow \boldsymbol{\theta}$

A different way to view fit issues



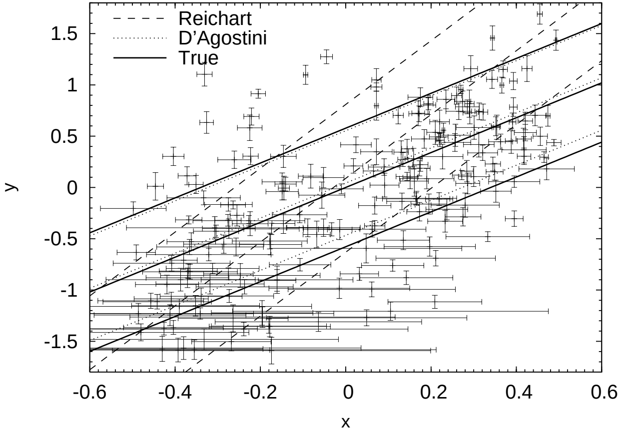
Determistic link μ_x's to μ_y's
Probabilistic links $\mu_x \rightarrow x$, $\mu_y \rightarrow y$
(errors on both axes!)
$\Rightarrow$ aim of fit: $\{\mathbf{x}, \mathbf{y}\} \rightarrow \boldsymbol{\theta}$



Extra spread
of the data points

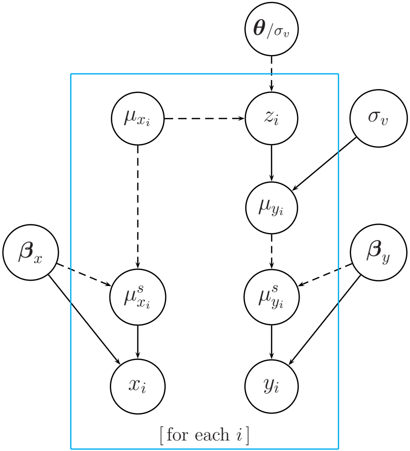
A different way to view fit issues

A physics case (from Gamma ray burts):



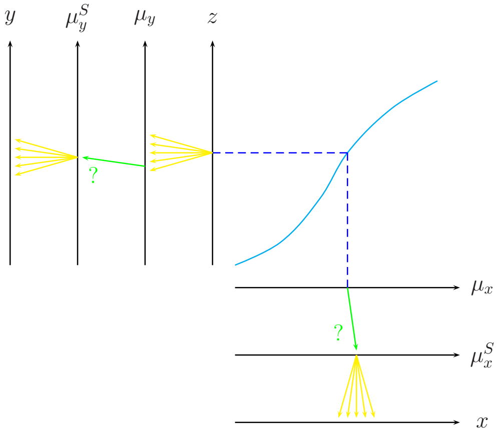
(Guidorzi et al., 2006)

A different way to view fit issues

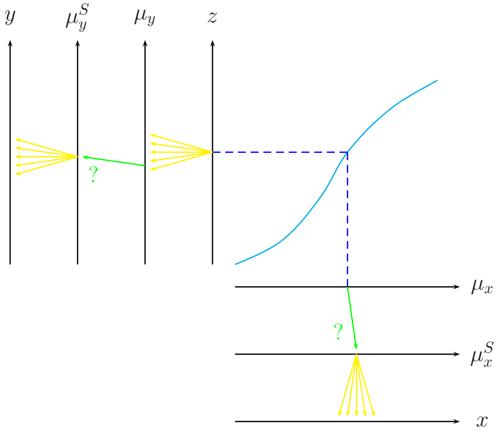


Adding systematics

A different way to view fit issues

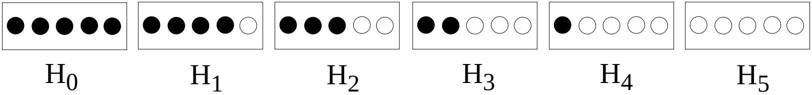


A different way to view fit issues



⇒ the mathematical function relating, generally speaking, "*y* to *x*" relates the true values, not the observations!

Application to the six box problem



Remind:

- $E_1 =$ White
- $E_2 =$ Black

Collecting the pieces of information we need

Our tool:

$$\boxed{P(H_j \mid E_i, I) = \frac{P(E_i \mid H_j, I)}{P(E_i \mid I)}\, P(H_j \mid I)}$$

Collecting the pieces of information we need

Our tool:

$$P(H_j \mid E_i, I) = \frac{P(E_i \mid H_j, I)}{P(E_i \mid I)}\, P(H_j \mid I)$$

- $P(H_j \mid I) = 1/6$

Collecting the pieces of information we need

Our tool:

$$P(H_j \mid E_i, I) = \frac{P(E_i \mid H_j, I)}{P(E_i \mid I)} P(H_j \mid I)$$

- $P(H_j \mid I) = 1/6$
- $P(E_i \mid I) = 1/2$

Collecting the pieces of information we need

Our tool:

$$P(H_j \mid E_i, I) = \frac{P(E_i \mid H_j, I)}{P(E_i \mid I)} P(H_j \mid I)$$

- $P(H_j \mid I) = 1/6$
- $P(E_i \mid I) = 1/2$
- $P(E_i \mid H_j, I)$:

$$\begin{aligned} P(E_1 \mid H_j, I) &= j/5 \\ P(E_2 \mid H_j, I) &= (5-j)/5 \end{aligned}$$

Collecting the pieces of information we need

Our tool:

$$P(H_j \,|\, E_i, I) = \frac{P(E_i \,|\, H_j, I)}{P(E_i \,|\, I)}\, P(H_j \,|\, I)$$

- $P(H_j \,|\, I) = 1/6$
- $P(E_i \,|\, I) = 1/2$
- $P(E_i \,|\, H_j, I)$:

$$\begin{aligned} P(E_1 \,|\, H_j, I) &= j/5 \\ P(E_2 \,|\, H_j, I) &= (5-j)/5 \end{aligned}$$

Our prior belief about H_j

Collecting the pieces of information we need

Our tool:

$$P(H_j \,|\, E_i, I) = \frac{P(E_i \,|\, H_j, I)}{P(E_i \,|\, I)}\, P(H_j \,|\, I)$$

- $P(H_j \,|\, I) = 1/6$
- $P(E_i \,|\, I) = 1/2$
- $P(E_i \,|\, H_j,\, I)$:

$$\begin{aligned} P(E_1 \,|\, H_j,\, I) &= j/5 \\ P(E_2 \,|\, H_j,\, I) &= (5-j)/5 \end{aligned}$$

Probability of E_i under a well defined hypothesis H_j
It corresponds to the 'response of the apparatus in measurements.
$\rightarrow$ likelihood (traditional, rather confusing name!)

Collecting the pieces of information we need

Our tool:

$$P(H_j \mid E_i, I) = \frac{P(E_i \mid H_j, I)}{P(E_i \mid I)} \, P(H_j \mid I)$$

- $P(H_j \mid I) = 1/6$
- $P(E_i \mid I) = 1/2$
- $P(E_i \mid H_j, I)$:

$$P(E_1 \mid H_j, I) = j/5$$
$$P(E_2 \mid H_j, I) = (5 - j)/5$$

Probability of E_i taking account all possible H_j

$\rightarrow$ How much we are confident that E_i will occur.

Collecting the pieces of information we need

Our tool:

$$P(H_j \,|\, E_i, I) = \frac{P(E_i \,|\, H_j, I)}{P(E_i \,|\, I)} \, P(H_j \,|\, I)$$

- $P(H_j \,|\, I) = 1/6$
- $P(E_i \,|\, I) = 1/2$
- $P(E_i \,|\, H_j, I)$:

$$\begin{aligned} P(E_1 \,|\, H_j, I) &= j/5 \\ P(E_2 \,|\, H_j, I) &= (5-j)/5 \end{aligned}$$

Probability of E_i taking account all possible H_j
→ How much we are confident that E_i will occur.
We can rewrite it as $P(E_i \,|\, I) = \sum_j P(E_i \,|\, H_j, I) \cdot P(H_j \,|\, I)$

Collecting the pieces of information we need

Our tool:

$$P(H_j \mid E_i, I) = \frac{P(E_i \mid H_j, I)}{P(E_i \mid I)} P(H_j \mid I)$$

- $P(H_j \mid I) = 1/6$
- $P(E_i \mid I) = 1/2$
- $P(E_i \mid H_j, I)$:

$$P(E_1 \mid H_j, I) = j/5$$
$$P(E_2 \mid H_j, I) = (5 - j)/5$$

But it easy to prove that $P(E_i \mid I)$ is related to the other ingredients, usually easier to 'measure' or to assess somehow, though vaguely

Collecting the pieces of information we need

Our tool:

$$\boxed{P(H_j \mid E_i, I) = \frac{P(E_i \mid H_j, I)}{P(E_i \mid I)}\, P(H_j \mid I)}$$

- $P(H_j \mid I) = 1/6$
- $P(E_i \mid I) = 1/2$
- $P(E_i \mid H_j, I)$:

$$\begin{aligned} P(E_1 \mid H_j, I) &= j/5 \\ P(E_2 \mid H_j, I) &= (5-j)/5 \end{aligned}$$

But it easy to prove that $P(E_i \mid I)$ is related to the other ingredients, usually easier to 'measure' or to assess somehow, though vaguely

'decomposition law': $P(E_i \mid I) = \sum_j P(E_i \mid H_j, I) \cdot P(H_j \mid I)$
($\rightarrow$ Easy to check that it gives $P(E_i \mid I) = 1/2$ in our case).

Collecting the pieces of information we need

Our tool:

$$P(H_j \mid E_i, I) = \frac{P(E_i \mid H_j, I) \cdot P(H_j \mid I)}{\sum_j P(E_i \mid H_j, I) \cdot P(H_j \mid I)}$$

- $P(H_j \mid I) = 1/6$
- $P(E_i \mid I) = \sum_j P(E_i \mid H_j, I) \cdot P(H_j \mid I)$
- $P(E_i \mid H_j, I)$:

$$\begin{aligned} P(E_1 \mid H_j, I) &= j/5 \\ P(E_2 \mid H_j, I) &= (5-j)/5 \end{aligned}$$

We are ready!

$\longrightarrow$ Let's play with our toy

We are ready

Now that we have set up our formalism, let's play a little

- analyse real data
- some simulations
- make variations

Let's play!

- Hugin Expert (Lite – demo version);
- R scripts

How does it work?

Simply – and nothing more! – Probability Theory

How does it work?

Simply – and nothing more! – Probability Theory

Given n variables X_i (each node), each of which can assume several values,

- build the joint 'pdf' using the 'chain rule'

$$f(x_1, x_2, \ldots, x_n \,|\, I)$$

How does it work?

Simply – and nothing more! – Probability Theory

Given n variables X_i (each node), each of which can assume several values,

- build the joint 'pdf' using the 'chain rule'

$$f(x_1, x_2, \ldots, x_n \,|\, I)$$

$\Rightarrow$ marginalize to get $f(x_i \,|\, I)$;

How does it work?

Simply – and nothing more! – Probability Theory

Given n variables X_i (each node), each of which can assume several values,

- build the joint 'pdf' using the 'chain rule'

$$f(x_1, x_2, \ldots, x_n \,|\, I)$$

⇒ marginalize to get $f(x_i \,|\, I)$;

⇒ condition on what is assumed to get the distribution of all the others.
E.g. $f(x_1, x_2, \ldots, x_{n-1} \,|\, I, x_n) = \frac{f(x_1, x_2, \ldots, x_n \,|\, I)}{f(x_n \,|\, I)}$.

How does it work?

Simply – and nothing more! – Probability Theory

Given n variables X_i (each node), each of which can assume several values,

- build the joint 'pdf' using the 'chain rule'

$$f(x_1, x_2, \ldots, x_n \,|\, I)$$

$\Rightarrow$ marginalize to get $f(x_i \,|\, I)$;

$\Rightarrow$ condition on what is assumed to get the distribution of all the others.
E.g. $f(x_1, x_2, \ldots, x_{n-1} \,|\, I, x_n) = \frac{f(x_1, x_2, \ldots, x_n \,|\, I)}{f(x_n \,|\, I)}$.

$\Rightarrow$ marginalize to get $f(x_i \,|\, I, x_n)$

How does it work?

Simply – and nothing more! – Probability Theory

Given n variables X_i (each node), each of which can assume several values,

- build the joint 'pdf' using the 'chain rule'

$$f(x_1, x_2, \ldots, x_n \mid I)$$

$\Rightarrow$ marginalize to get $f(x_i \mid I)$;

$\Rightarrow$ condition on what is assumed to get the distribution of all the others.
E.g. $f(x_1, x_2, \ldots, x_{n-1} \mid I, x_n) = \frac{f(x_1, x_2, \ldots, x_n \mid I)}{f(x_n \mid I)}$.

$\Rightarrow$ marginalize to get $f(x_i \mid I, x_n)$

(Only some 'technical tricks' to factorize the problem when the number of 'states' becomes very large)

OK, ...but the priors?

Priors are an important ingredient of the framework:

OK, . . . but the priors?

Priors are an important ingredient of the framework:

- They are crucial in the Bayes theorem:
  - there is no other way to perform a probabilistic inference without passing through priors
    . . . although they can be often so vague to be ignored.

OK, ...but the priors?

Priors are an important ingredient of the framework:

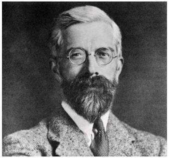
- They are crucial in the Bayes theorem:
  - there is no other way to perform a probabilistic inference without passing through priors
    ...although they can be often so vague to be ignored.
- They allow us to use consistently all pieces of prior information. And we all have much prior information in our job!
  Only the perfect idiot hase no priors

OK, ...but the priors?

Priors are an important ingredient of the framework:

- They are crucial in the Bayes theorem:
  - there is no other way to perform a probabilistic inference without passing through priors
    ...although they can be often so vague to be ignored.
- They allow us to use consistently all pieces of prior information. And we all have much prior information in our job!
  Only the perfect idiot hase no priors
- Mistrust all prior-free methods that pretend to provide numbers that should mean how you have to be confident on something.

Prescriprions?



Objective prescriptions?

Mistrust those who promise you 'objective' methods to form up your confidence about the physical world!

Principles?

Too many unnecessary ‘principles’ on the market.

Principles?

Too many unnecessary ‘principles’ on the market.

“These are my principles.
If you don’t like them,
I have others.”

(Groucho Marx)

Summarizing

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

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

Summarizing

- The probabilistic framework basically set up by Laplace(*) in his monumental work is healthy and grows up well (browse e.g. Amazon.com)
- It is very close to the natural way of reasoning.

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

Summarizing

- The probabilistic framework basically set up by Laplace[(*)] in his monumental work is healthy and grows up well (browse e.g. Amazon.com)
- It is very close to the natural way of reasoning.
- Its consistent application in small-complex problems was prohibitive many years ago.

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

Summarizing

- The probabilistic framework basically set up by Laplace[(*)] in his monumental work is healthy and grows up well (browse e.g. Amazon.com)
- It is very close to the natural way of reasoning.
- 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]

Summarizing

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

Summarizing

- The probabilistic framework basically set up by Laplace$^{(*)}$ in his monumental work is healthy and grows up well (browse e.g. Amazon.com)
- It is very close to the natural way of reasoning.
- 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.
- It makes little sense to stick to old 'ah hoc' methods that had their *raison d'être* in the computational barrier.
- Mistrust all results that sound as 'confidence', 'probability' etc about physics quantities, if they are obtained by methods that do not contemplate 'beliefs'.

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

The End

FINE

Notes

The following slides should be reached by hyper-links, clicking on highlighted words marked by the symbol †

ISO dictionary

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

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 measurement."*

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

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

Type A and Type B uncertainties →

Go back

ISO dictionary

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

ISO dictionary

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

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

ISO dictionary

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

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

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

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

 ta taken

ISO dictionary

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

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

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

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

ta taken

Solution of the AIDS test problem

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

We miss something: $P_\circ(\text{HIV})$ and $P_\circ(\overline{\text{HIV}})$: Yes! We need some input from our best knowledge of the problem. Let us take $P_\circ(\text{HIV}) = 1/600$ and $P_\circ(\overline{\text{HIV}}) \approx 1$ (the result is rather stable against *reasonable* variations of the inputs!)

$$
\begin{aligned}
\frac{P(\text{HIV}\,|\,\text{Pos})}{P(\overline{\text{HIV}}\,|\,\text{Pos})} &= \frac{P(\text{Pos}\,|\,\text{HIV})}{P(\text{Pos}\,|\,\overline{\text{HIV}})} \cdot \frac{P_\circ(\text{HIV})}{P_\circ(\overline{\text{HIV}})} \\
&= \frac{\approx 1}{0.002} \times \frac{0.1/60}{\approx 1} = 500 \times \frac{1}{600} = \frac{1}{1.2}
\end{aligned}
$$

Go back