30 years of Bayesian unfolding
— underlying ideas and personal recollections —

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

30 years of Bayesian unfolding

— underlying ideas and personal recollections —

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

"... 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)

30 years of Bayesian unfolding
— underlying ideas and personal recollections —

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

*". . . 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)

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

30 years of Bayesian unfolding

— underlying ideas and personal recollections —

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

*"... 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)

"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)

30 years of Bayesian unfolding

— underlying ideas and personal recollections —

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

"... 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)

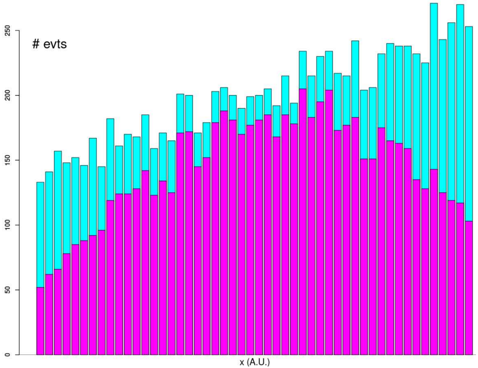
"It is scientific only to say what is more likely and what is less likely"
(R. Feynman)

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

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

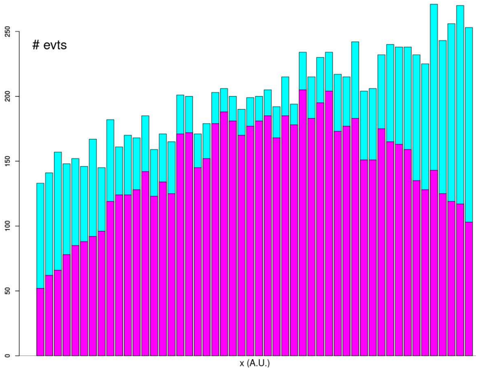
Subject of the talk

'Observed' spectra are often distorted for 'the reasons we know'



Subject of the talk

'Observed' spectra are often distorted for 'the reasons we know'

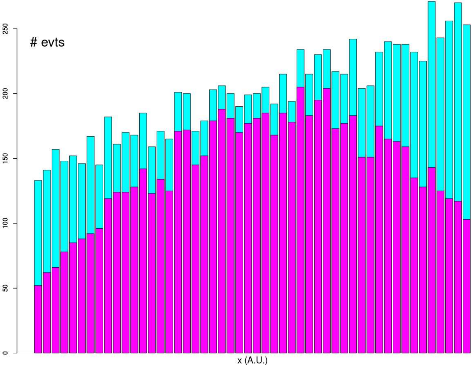


Our aim

- from the **observed one** try to get what we *could have* got with an **ideal detector**

Subject of the talk

'Observed' spectra are often distorted for 'the reasons we know'



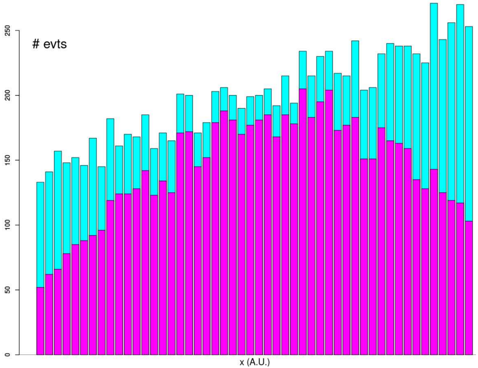
Our aim

- from the **observed one** try to get what we *could have* got with an **ideal detector**

But, obviously, not this cyan histogram from this magenta one**!**

Subject of the talk

'Observed' spectra are often distorted for 'the reasons we know'



Our aim

- from the **observed one** try to get what we *could have* got with an **ideal detector**

But, obviously, not this cyan histogram from this magenta one!

$\Rightarrow$ We have to deal with uncertainty and probability

Although other ‘methods’ might be more fashionable

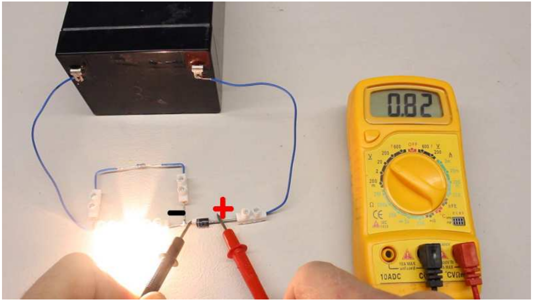
Although other 'methods' might be more fashionable

[Plus other **prescriptions** you might imagine. . .]

Let's start

What is measurement?

What is measurement?

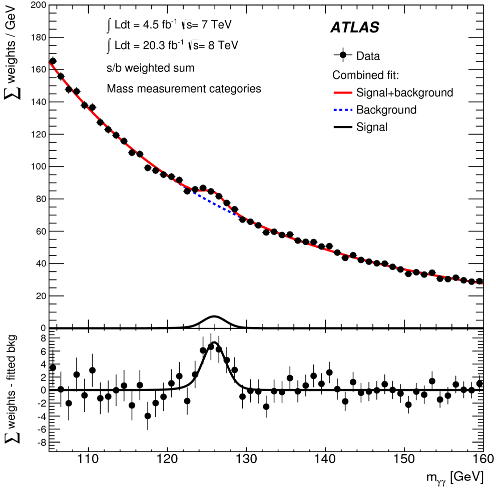


What is measurement?

What is measurement?

What is measurement?

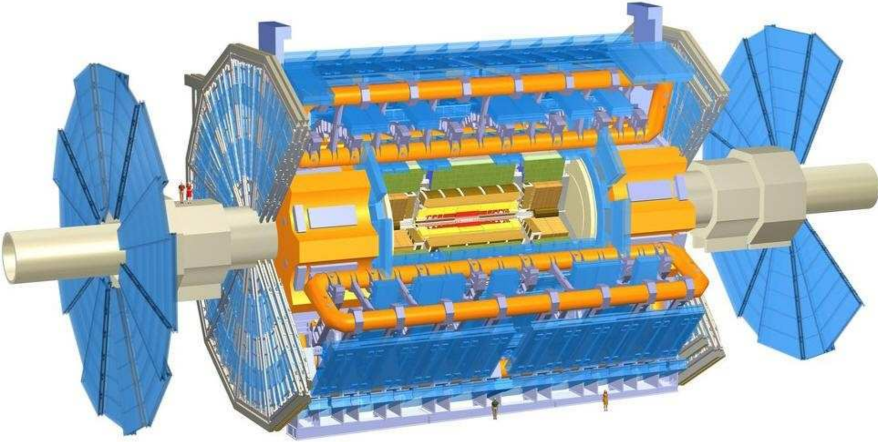
Higgs $\to \gamma\gamma$ (2012)



Two-photon *invariant mass*

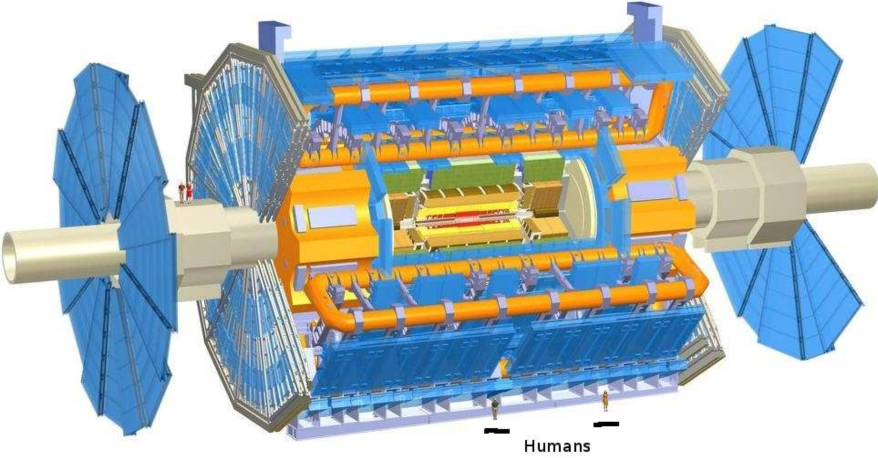
What is measurement?

ATLAS Experiment at LHC (CERN, Geneva)



What is measurement?

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

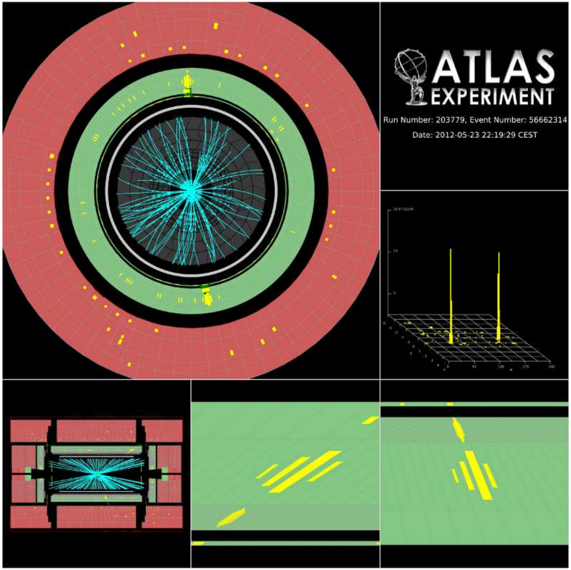


≈ 3000 km cables

≈ 7000 tonnes

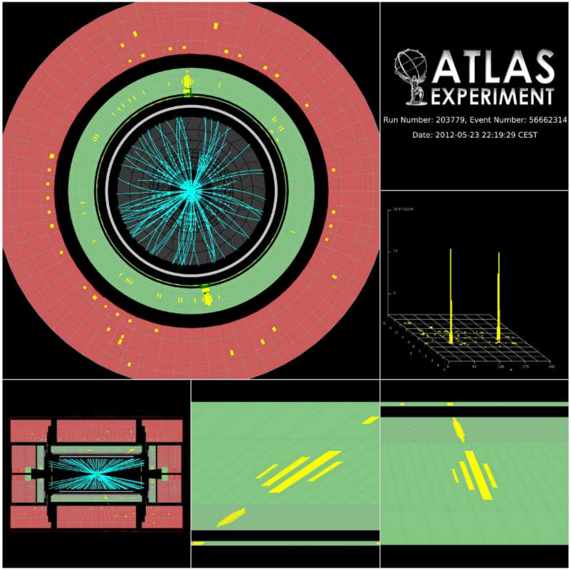
≈ 100 millions electronic channels

What is measurement?



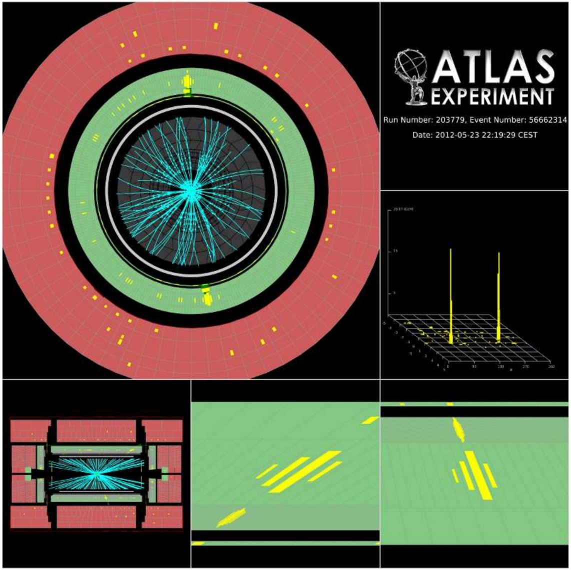
Two flashes of ‘light’ (2 γ’s) in a ‘noisy’ environment.

What is measurement?



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

What is measurement?

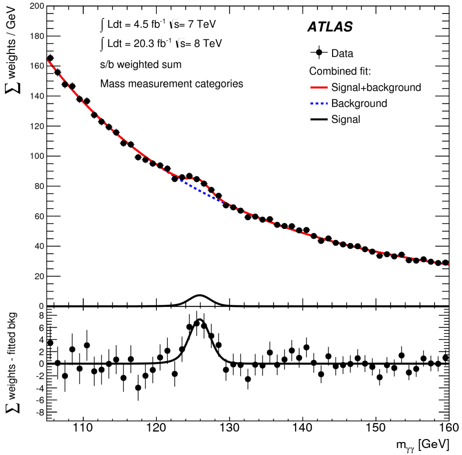


Two flashes of 'light' (2 γ's) in a 'noisy' environment.

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

What is measurement?

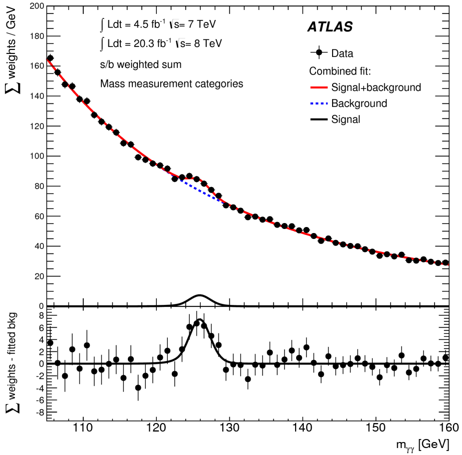
Higgs $\to \gamma\gamma$



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

What is measurement?

Higgs $\to \gamma\gamma$

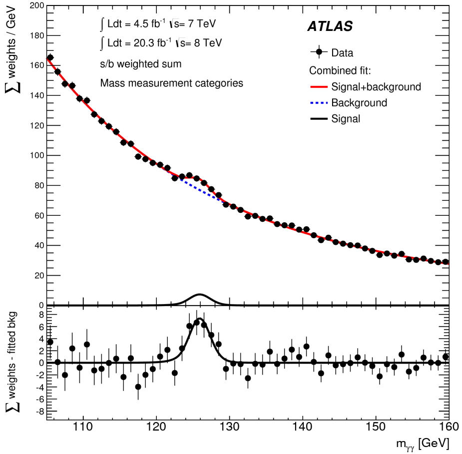


$\Rightarrow \left\{ \begin{array}{l} \text{Mass value} \\ \text{Production rate} \end{array} \right.$

(with uncertainties)

What is measurement?

Higgs $\to \gamma\gamma$



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

(with uncertainties)

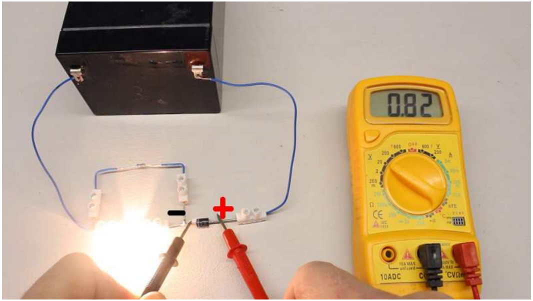
Quite indirect measurements of something we do not "see"!

Can we “see” physics quantities?

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

Can we "see" physics quantities?

. . . or a voltage?



Can we “see” physics quantities?

…or our blood pressure?

Can we "see" physics quantities?

Certainly not!

Can we “see” physics quantities?

Certainly not!

...although for some quantities we can have

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

Measuring a mass on a scale

Equilibrium ('physical principle of the measurement'):

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

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

Measuring a mass on a scale

Equilibrium ('physical principle of the measurement'):

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

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

From the reading to the value of the mass:

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

Measuring a mass on a balance

$$\text{scale reading} \quad \xrightarrow[\textit{given } g,\ k,\ \text{“etc.”}\ldots]{\hspace{6cm}} \quad m$$

Measuring a mass on a balance

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

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

Measuring a mass on a balance

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

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

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

Measuring a mass on a balance

$$\text{scale reading} \quad \xrightarrow[\textit{given } g,\ k,\ \textit{“etc.”}\ldots]{\hspace{6em}} \quad m$$

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

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

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

- Position is usually <u>not</u> at “$R_{\oplus}$” from the Earth center;
- Earth not spherical...
- ...not even ellipsoidal...

Measuring a mass on a balance

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

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

- Position is usually not at "$R_{\oplus}$" from the Earth center;
- Earth not spherical...
- ...not even ellipsoidal...
- ...and not even homogeneous.

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 not at "$R_\oplus$" from the Earth center;
- Earth not spherical...
- ...not even ellipsoidal...
- ...and not even homogeneous.
- Moreover we have to consider centrifugal effects

Measuring a mass on a balance

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

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

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

Measuring a mass on a balance

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

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

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

Certainly not to watch our weight 

But think about it!

Measuring a mass on a balance

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

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
- ...

$\boldsymbol{\Delta}$x $\rightarrow \theta \rightarrow$ 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
- ...

$\boldsymbol{\Delta}$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
- ...

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

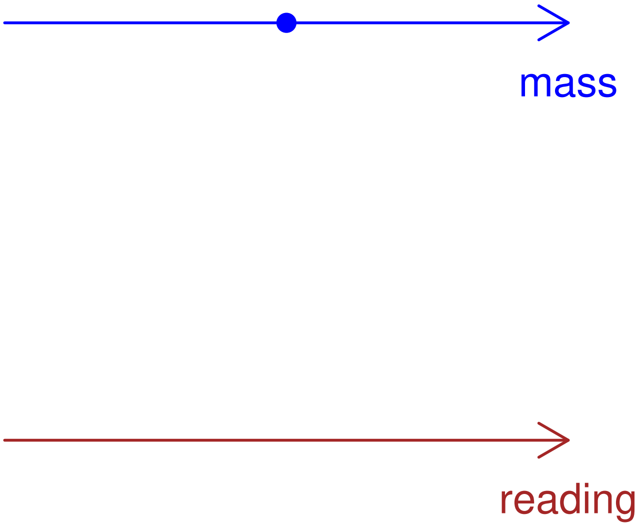
- left to your imagination...

$+$ **randomic effects:**

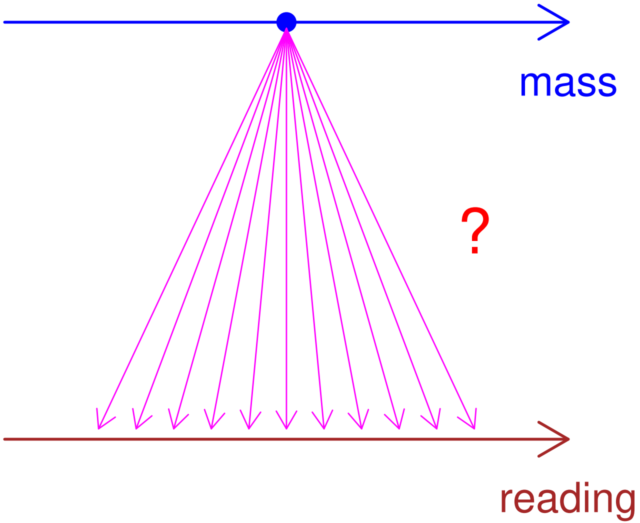
- stopping position of damped oscillation;
- variability of all quantities of influence (in the ISO-GUM sense);
- reading of analog scale.

$\Rightarrow \;\; m$??

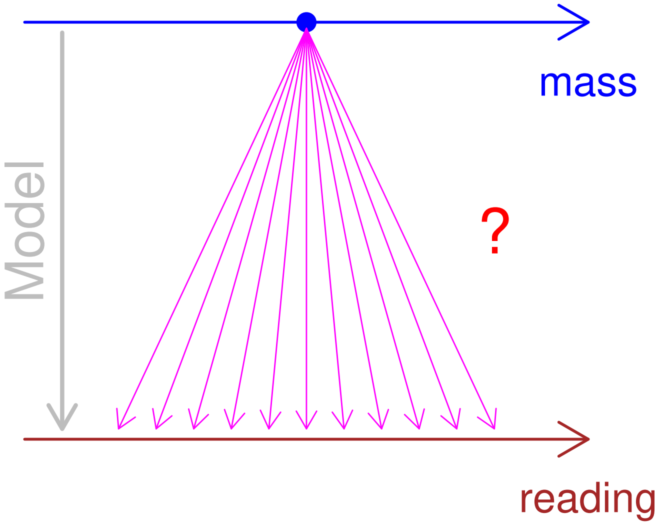
Mass $\longrightarrow$ Reading



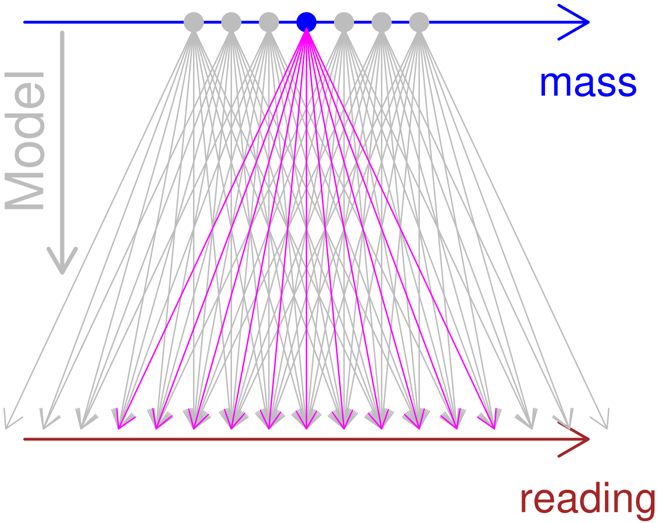
Mass $\longrightarrow$ Reading



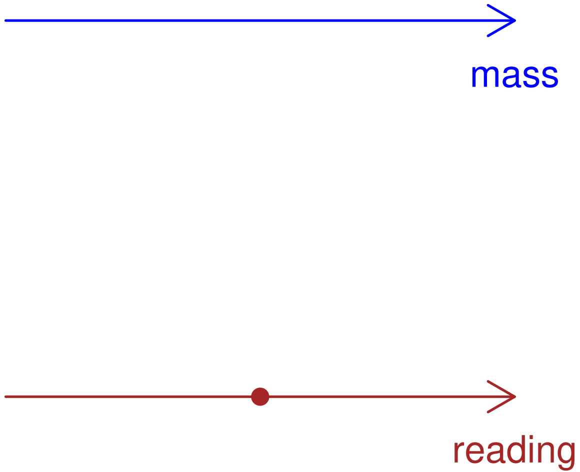
Mass ⟶ Reading



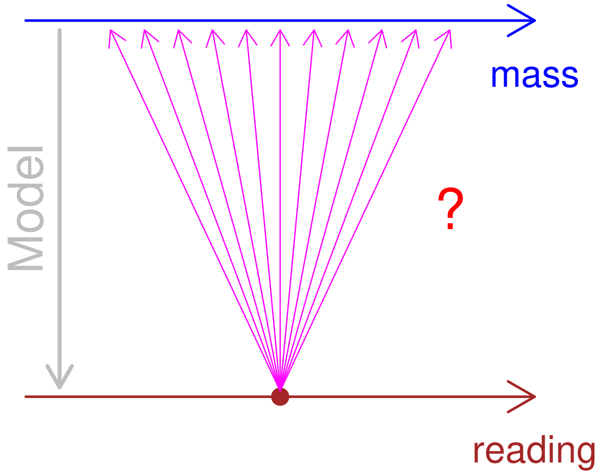
Mass $\longrightarrow$ Reading



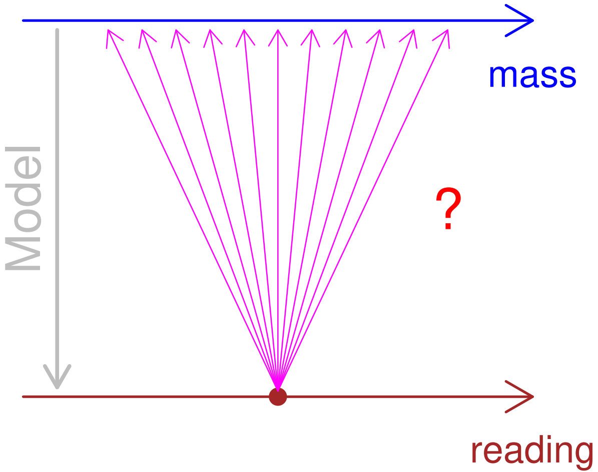
Reading $\longrightarrow$ ‘true’ mass



Reading $\longrightarrow$ 'true' mass



Reading $\longrightarrow$ 'true' mass



Measurement is nothing but

inferring a model parameter

Physical world $\leftrightarrow$ Science (from Latin '*scio*' – to know)

An Einstein's quote (from his *Autobiography*) might help:

> *"Physical concepts are free creations of the human mind,*
> *and are not, however it may seem,*
> *uniquely determined by the external world"*

Physical world $\leftrightarrow$ Science (from Latin '*scio*' – to know)

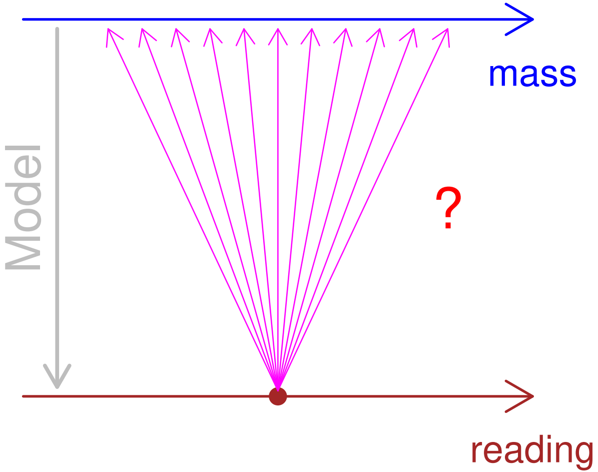
An Einstein's quote (from his *Autobiography*) might help:

> *"Physical concepts are free creations of the human mind, and are not, however it may seem, uniquely determined by the external world"*

And, again there, referring to his revolutianary ideas:

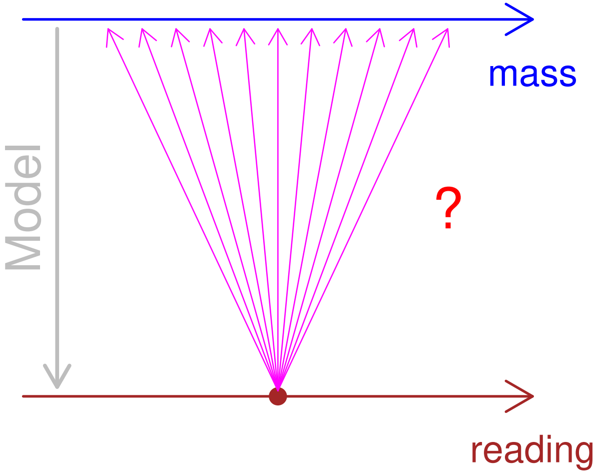
> *"The type of critical reasoning which was required for the discovery of this central point was decisively furthered, in my case, especially by the reading of David Hume's and Ernst Mach's philosophical writings"*

Data $\longrightarrow$ value of a physical quantity ('true value')



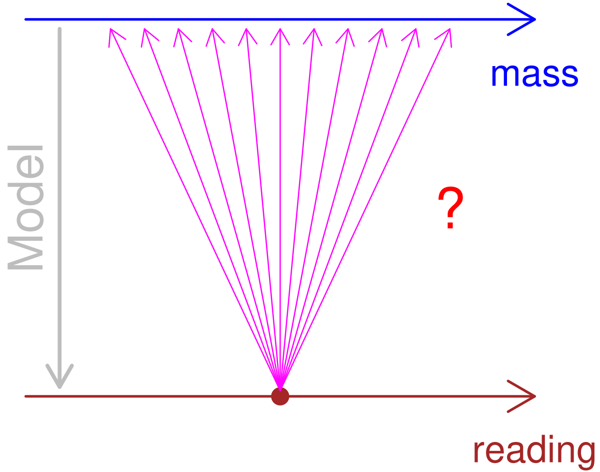
The <u>observed 'data' is certain</u>:

Data $\longrightarrow$ value of a physical quantity ('true value')



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

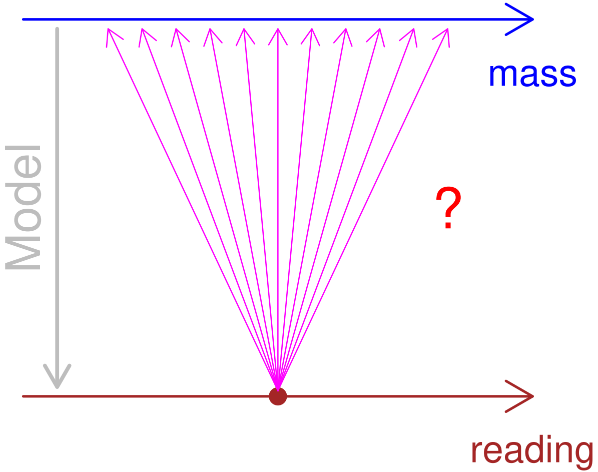
Data $\longrightarrow$ value of a physical quantity ('true value')



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

"Data uncertainty"?

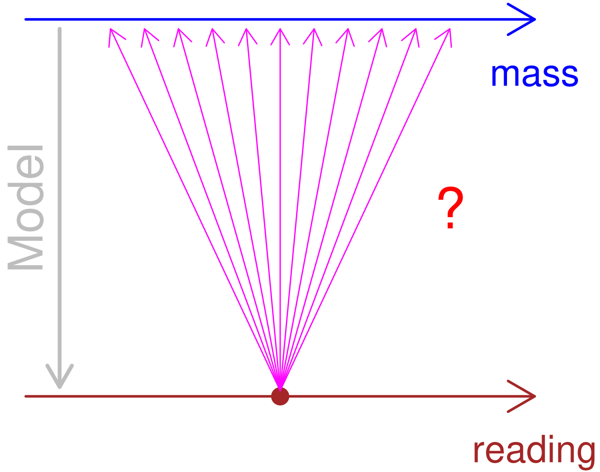
Data $\longrightarrow$ value of a physical quantity ('true value')



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

"Data uncertainty"? **???**

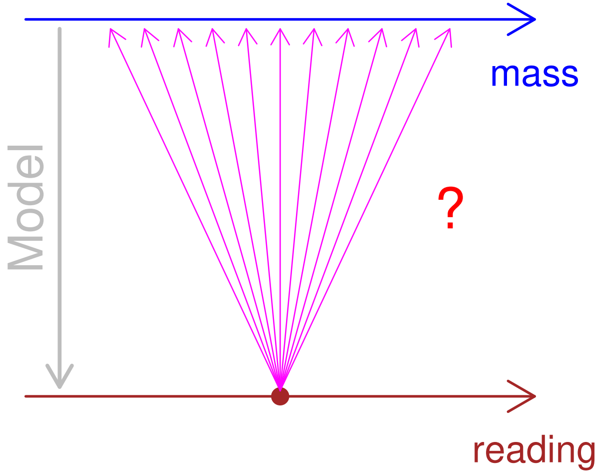
Data $\longrightarrow$ value of a physical quantity ('true value')



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

"Data uncertainty"? **???** **Data corrupted?**

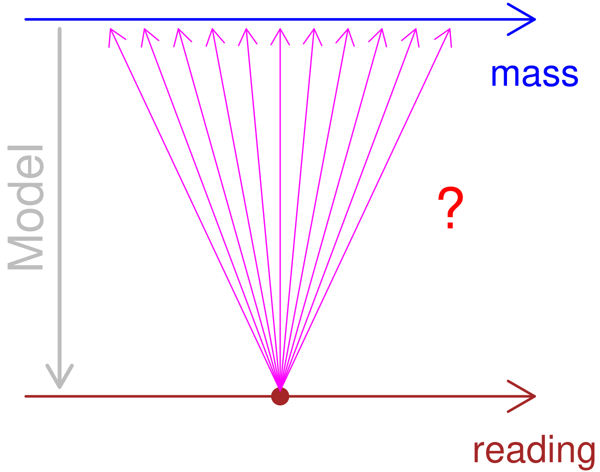
Data $\longrightarrow$ value of a physical quantity ('true value')



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

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

Data $\longrightarrow$ value of a physical quantity ('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**!!** . . .

[Unless we are talking of 'future data' or of 'somebody else data' . . .]

Observation $\rightarrow$ value of a quantity

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

Observation $\rightarrow$ value of a quantity

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

$\Rightarrow$ Measurement is not simply 'reading' a value on an instrument

Observation $\rightarrow$ value of a quantity

$$\text{scale reading} \quad \xrightarrow[\textit{given } g,\ k,\ \text{“etc.”}\ldots]{\hspace{4cm}} \quad m$$

$\Rightarrow$ Measurement is not simply ‘reading’ a value on an instrument
(A reading, without proper contextualization, is . . . just a number)

Observation → value of a quantity

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

⇒ Measurement is not simply 'reading' a value on an instrument
(A reading, without proper contextualization, is . . . just a number)

Mistrust the

"dogma of the Immaculate Observation"!

Observations $\rightarrow$ hypotheses

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

Observations $\rightarrow$ hypotheses

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

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

Observations $\rightarrow$ hypotheses

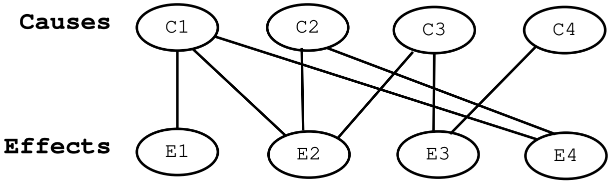
This problem occurs not only "determining" *the* value of a physical quantity.

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

(But logically no substantial difference.)

Causes $\rightarrow$ effects

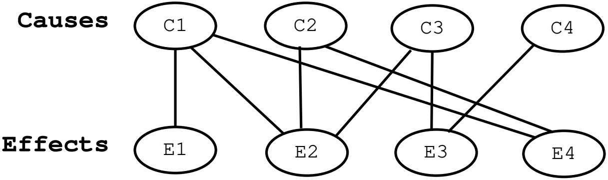
The same *apparent* cause might produce several,different effects



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

Causes → effects

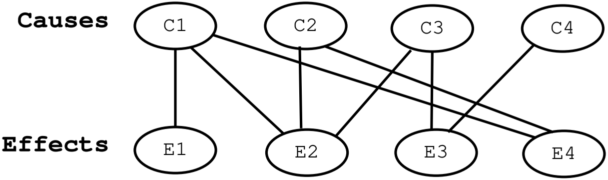
The same *apparent* cause might produce several,different effects



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

Causes $\rightarrow$ effects

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



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

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

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.

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?

“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.

The "essential problem" of the Sciences

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

I play with a gentleman whom I do not know. He has dealt ten times, and he has turned the king up six times. What is the chance that he is a sharper?

The "essential problem" of the Sciences

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

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

(H. Poincaré – *Science and Hypothesis*)

The "essential problem" of the Sciences

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

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

(H. Poincaré – *Science and Hypothesis*)

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

Basic rules of probability $\surd$

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 $\surd$

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. <u>does not</u> define conditional probability.

(Probability is <u>always</u> conditional probability!)

Basic rules of probability $\surd$

1. $0 \le P(A \,|\, I) \le 1$
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!)

$\Rightarrow$ 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.

Misconceptions based on the confusion between the following terms:

- belief
- immagination

} see analysis by Hume

- subjective
- arbitrary

} role of "coherence" (de Finetti)

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

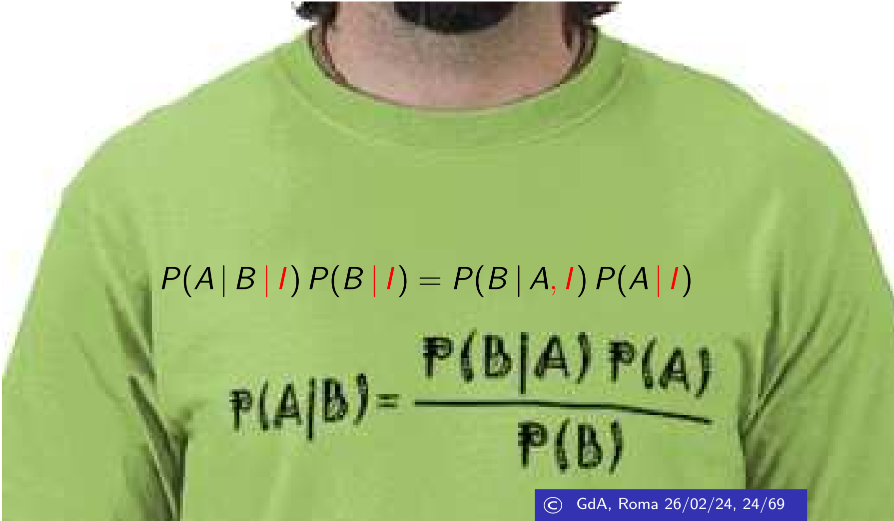
An even better news:

The fourth basic rule
can be fully exploited!

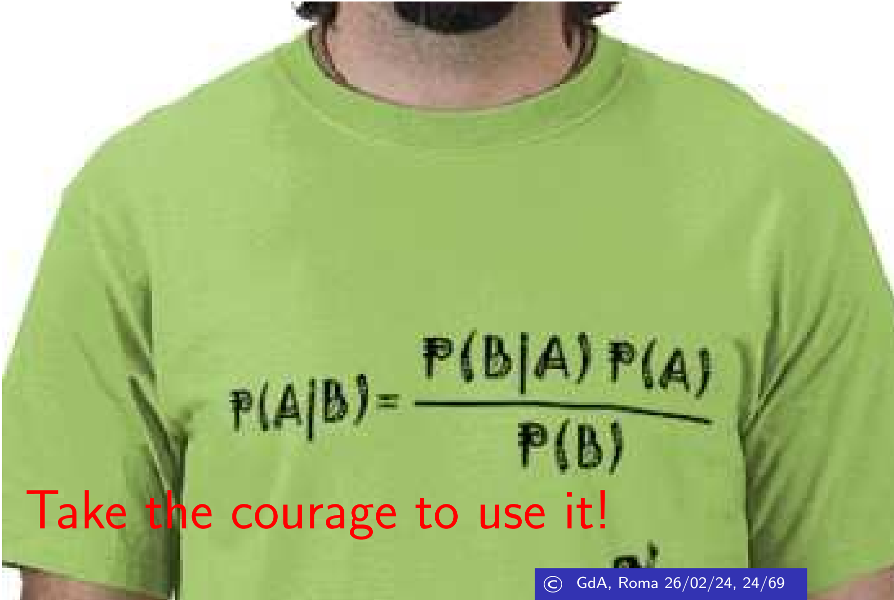
(Liberated by a curious ideology that forbids its use)

A simple, powerful formula

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

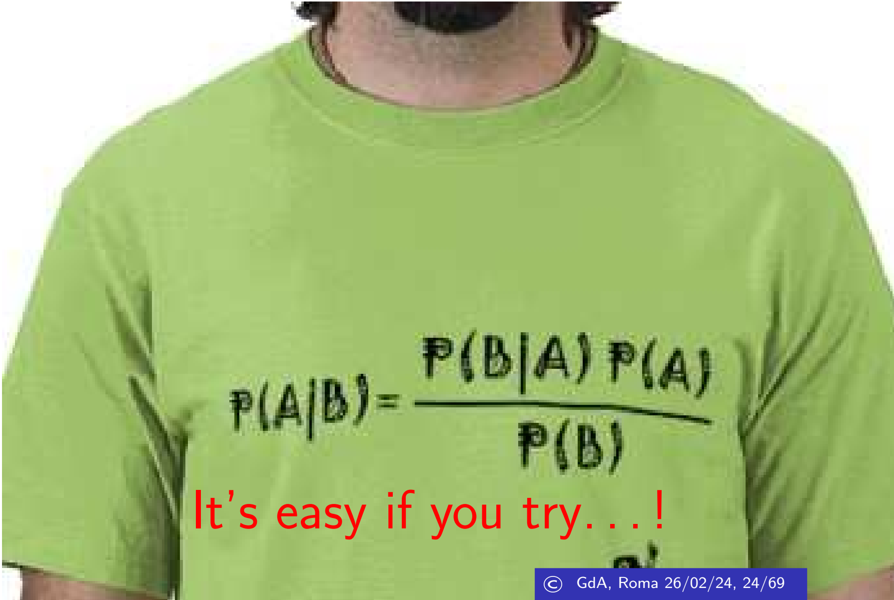


A simple, powerful formula



Take the courage to use it!

A simple, powerful formula



It's easy if you try. . . !

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"

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

"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 '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"

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

"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 'fundamental 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

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 'his' error function

1. $f(\mu \,|\, \{x\}) \propto f(\{x\} \,|\, \mu) \cdot f_0(\mu)$
2. $f_0(\mu)$ 'flat': all values 'a priori' equally possible ("... aeque probabilia fuisse")
3. 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 'his' error function

1. $f(\mu \,|\, \{x\}) \propto f(\{x\} \,|\, \mu) \cdot f_0(\mu)$
2. $f_0(\mu)$ 'flat': all values 'a priori' equally possible
   ("... aeque probabilia fuisse")
3. posterior maximized at $\mu = \overline{x}$

Indeed Gauss had clear ideas about the role of the priors and also of the fact the, strictly speaking, the 'Gaussian' "cannot express the probability of the errors" **(!!)** [cfr e.g. arXiv:2003.10878]
("All models are wrong...")

Personal recollections

(Thanks for patience and compassion. . .)

First impact with a new, simple but intriguing 'theorem'

- Until March 1993 I had **not even heard** about Bayes theorem (as most/'all' colleagues...)

First impact with a new, simple but intriguing ‘theorem’

- Until March 1993 I had **<u>not even heard</u>** about Bayes theorem (as most/’all’ colleagues. . .)
- In March 1993 I started teaching a course of ‘Physics Laboratory’ for Chemestry students (2nd semester, 2nd year)

First impact with a new, simple but intriguing ‘theorem’

- Until March 1993 I had **not even heard** about Bayes theorem (as most/’all’ colleagues. . .)
- In March 1993 I started teaching a course of ‘Physics Laboratory’ for Chemestry students (2nd semester, 2nd year)
  - just teaching what I had learned as student;

First impact with a new, simple but intriguing ‘theorem’

- Until March 1993 I had **<u>not even heard</u>** about Bayes theorem (as most/’all’ colleagues. . .)
- In March 1993 I started teaching a course of ‘Physics Laboratory’ for Chemestry students (2nd semester, 2nd year)
  - just teaching what I had learned as student;
  - ok, as far as laboratory experiences were concerned;

First impact with a new, simple but intriguing 'theorem'

- Until March 1993 I had **not even heard** about Bayes theorem (as most/'all' colleagues. . .)
- In March 1993 I started teaching a course of 'Physics Laboratory' for Chemestry students (2nd semester, 2nd year)
  - just teaching what I had learned as student;
  - ok, as far as laboratory experiences were concerned;
  - **deep crisis** at the moment of introducing probability → see;

First impact with a new, simple but intriguing 'theorem'

- Until March 1993 I had **<u>not even heard</u>** about Bayes theorem (as most/'all' colleagues. . .)
- In March 1993 I started teaching a course of 'Physics Laboratory' for Chemestry students (2nd semester, 2nd year)
  - just teaching what I had learned as student;
  - ok, as far as laboratory experiences were concerned;
  - **deep crisis** at the moment of introducing probability → see;
    - no web → went to libraries → Renzo Orsi's book:

First impact with a new, simple but intriguing ‘theorem’

- Until March 1993 I had **not even heard** about Bayes theorem
  (as most/’all’ colleagues. . .)
- In March 1993 I started teaching a course of ‘Physics Laboratory’ for Chemestry students (2nd semester, 2nd year)
  - just teaching what I had learned as student;
  - ok, as far as laboratory experiences were concerned;
  - **deep crisis** at the moment of introducing probability → see;
    - no web → went to libraries → Renzo Orsi’s book:
      ⇒ intrigued by the Bayes’ theorem
      (indeed so simple that in my opinion it hardly deserved the name of ‘theorem’)

First impact with a new, simple but intriguing ‘theorem’

- Until March 1993 I had **<u>not even heard</u>** about Bayes theorem (as most/’all’ colleagues. . .)
- In March 1993 I started teaching a course of ‘Physics Laboratory’ for Chemestry students (2nd semester, 2nd year)
  - just teaching what I had learned as student;
  - ok, as far as laboratory experiences were concerned;
  - **deep crisis** at the moment of introducing probability → see;
    - no web → went to libraries → Renzo Orsi’s book:
      ⇒ intrigued by the Bayes’ theorem
      (indeed so simple that in my opinion it hardly deserved the name of ‘theorem’)
- Months of work followed (March-June):
  - theorem extended to continuous variables;

First impact with a new, simple but intriguing ‘theorem’

- Until March 1993 I had **<u>not even heard</u>** about Bayes theorem
  (as most/’all’ colleagues...)
- In March 1993 I started teaching a course of ‘Physics Laboratory’ for Chemestry students (2nd semester, 2nd year)
  - just teaching what I had learned as student;
  - ok, as far as laboratory experiences were concerned;
  - **deep crisis** at the moment of introducing probability → see;
    - no web → went to libraries → Renzo Orsi’s book:
      ⇒ intrigued by the Bayes’ theorem
      (indeed so simple that in my opinion it hardly deserved the name of ‘theorem’)
- Months of work followed (March-June):
  - theorem extended to continuous variables;
  - inference of binomial p (reproducing Laplace results), of Poisson λ and of Gaussian μ;

First impact with a new, simple but intriguing ‘theorem’

- Until March 1993 I had **not even heard** about Bayes theorem (as most/’all’ colleagues. . .)
- In March 1993 I started teaching a course of ‘Physics Laboratory’ for Chemestry students (2nd semester, 2nd year)
  - just teaching what I had learned as student;
  - ok, as far as laboratory experiences were concerned;
  - **deep crisis** at the moment of introducing probability $\rightarrow$ see;
    - no web $\rightarrow$ went to libraries $\rightarrow$ Renzo Orsi’s book:
      $\Rightarrow$ intrigued by the Bayes’ theorem
      (indeed so simple that in my opinion it hardly deserved the name of ‘theorem’)
- Months of work followed (March-June):
  - theorem extended to continuous variables;
  - inference of binomial p (reproducing Laplace results), of Poisson λ and of Gaussian μ;
  - + combining measurements, handling systematics, and much more.

First impact with a new, simple but intriguing ‘theorem’

- Until March 1993 I had **not even heard** about Bayes theorem (as most/’all’ colleagues. . .)
- In March 1993 I started teaching a course of ‘Physics Laboratory’ for Chemestry students (2nd semester, 2nd year)
  - just teaching what I had learned as student;
  - ok, as far as laboratory experiences were concerned;
  - **deep crisis** at the moment of introducing probability → see;
    - no web → went to libraries → Renzo Orsi’s book:

      ⇒ intrigued by the Bayes’ theorem

      (indeed so simple that in my opinion it hardly deserved the name of ‘theorem’)
- Months of work followed (March-June):
  - theorem extended to continuous variables;
  - inference of binomial p (reproducing Laplace results),(*) of Poisson λ and of Gaussian μ;
  - + combining measurements, handling systematics, and much more.

[(*)I was a bit disapponted, but glad to be with such a good company. . .]

First impact with a new, simple but intriguing 'theorem'

- Until March 1993 I had **not even heard** about Bayes theorem
  (as most/'all' colleagues...)
- In March 1993 I started teaching a course of 'Physics Laboratory' for Chemestry students (2nd semester, 2nd year)
  - just teaching what I had learned as student;
  - ok, as far as laboratory experiences were concerned;
  - **deep crisis** at the moment of introducing probability → see;
    - no web → went to libraries → Renzo Orsi's book:
      ⇒ intrigued by the Bayes' theorem
      (indeed so simple that in my opinion it hardly deserved the name of 'theorem')
- Months of work followed (March-June):
  - theorem extended to continuous variables;
  - inference of binomial p (reproducing Laplace results),$^{(*)}$
    of Poisson λ and of Gaussian μ;
  - + combining measurements, handling systematics,
    and much more.

  [$^{(*)}$I was a bit disapponted, but glad to be with such a good company...]
- Astonished by how sensible the results were!

First impact with a new, simple but intriguing 'theorem'

- Until March 1993 I had **not even heard** about Bayes theorem (as most/'all' colleagues. . .)
- In March 1993 I started teaching a course of 'Physics Laboratory' for Chemestry students (2nd semester, 2nd year)
  - just teaching what I had learned as student;
  - ok, as far as laboratory experiences were concerned;
  - **deep crisis** at the moment of introducing probability → see;
    - no web → went to libraries → Renzo Orsi's book:

      ⇒ intrigued by the Bayes' theorem

      (indeed so simple that in my opinion it hardly deserved the name of 'theorem')
- Months of work followed (March-June):
  - theorem extended to continuous variables;
  - inference of binomial p (reproducing Laplace results),$^{(*)}$ of Poisson λ and of Gaussian μ;
  - + combining measurements, handling systematics, and much more.

  [$^{(*)}$I was a bit disapponted, but glad to be with such a good company. . .]
- Astonished by how sensible the results were!
- Later I discovered the Bayesian world.

Why the unfolding? The Physics case

- At that time I was working in the ZEUS experiment at the *e*-*p* collider HERA in Hamburg.

Why the unfolding? The Physics case

- At that time I was working in the ZEUS experiment at the e-p collider HERA in Hamburg.
  - Measuring the DIS quantities x and Q^2 was challenging in some physical regions, expecially when charged currents were involved, i.e. with ν_e in the final state.

Why the unfolding? The Physics case

- At that time I was working in the ZEUS experiment at the e-p collider HERA in Hamburg.
  - Measuring the DIS quantities x and Q^2 was challenging in some physical regions, expecially when charged currents were involved, i.e. with ν_e in the final state.

  $\rightarrow$ Not simply Gaussian 'smearing'
  of the reconstructed quantities wrt the physical ones

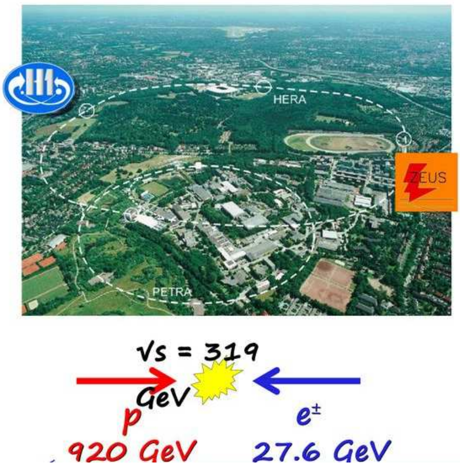
Why the unfolding? The Physics case

- At that time I was working in the ZEUS experiment at the e-p collider HERA in Hamburg.
  - Measuring the DIS quantities x and Q^2 was challenging in some physical regions, expecially when charged currents were involved, i.e. with ν_e in the final state.

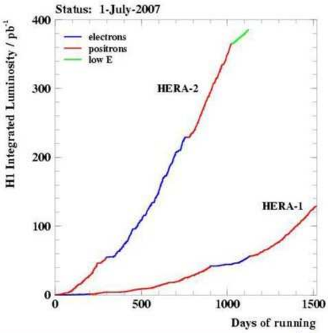
  $\rightarrow$ Not simply Gaussian 'smearing' of the reconstructed quantities wrt the physical ones

$\Rightarrow$ Short reminder

HERA Physics

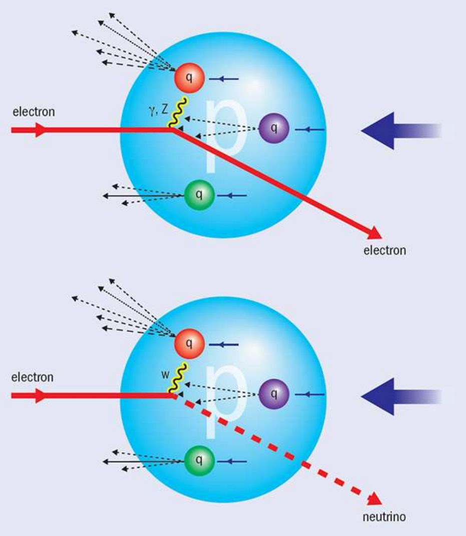


- ep Collider
- Collider experiments H1 and ZEUS
- Luminosity: ~200 pb^{-1} in e^-p
  ~300 pb^{-1} in e^+p



Main physics goal: proton structure (+. . . ..., including searches)

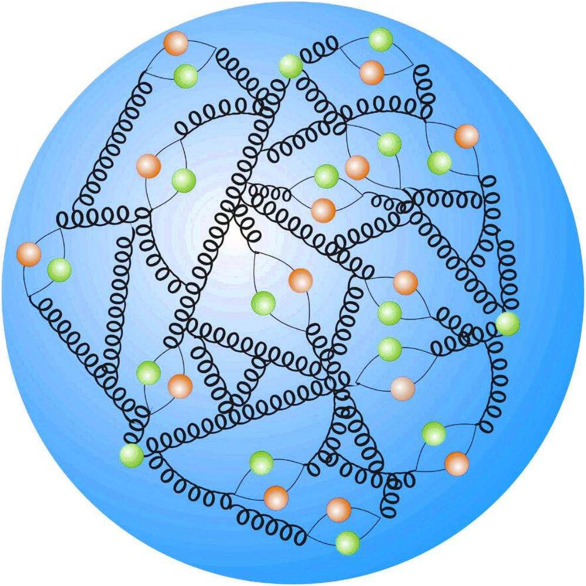
Deep inelastic scattering at HERA



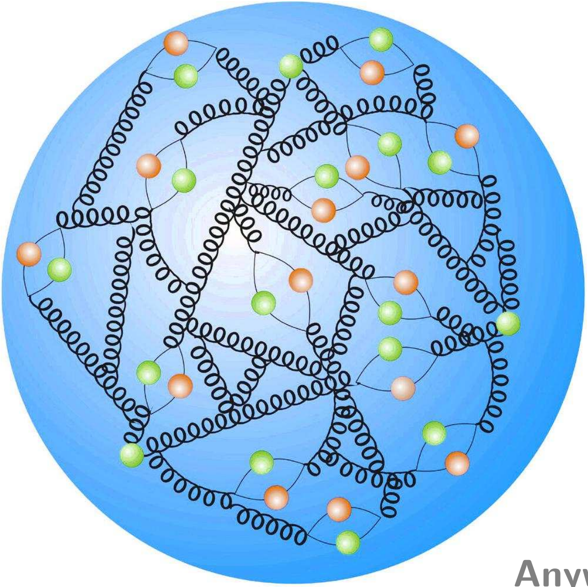
[OVERSIMPLIFIED$^{(*)}$ diagram, from CERN Courier, 2015/08]

$^{(*)}$A shame: it confirms my adversion to popularization of Science

A proton is a complicated ‘structured’ dynamical object. . .

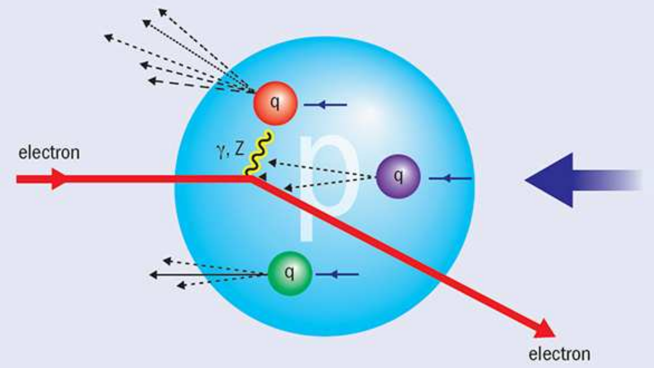


A proton is a complicated ‘structured’ dynamical object...

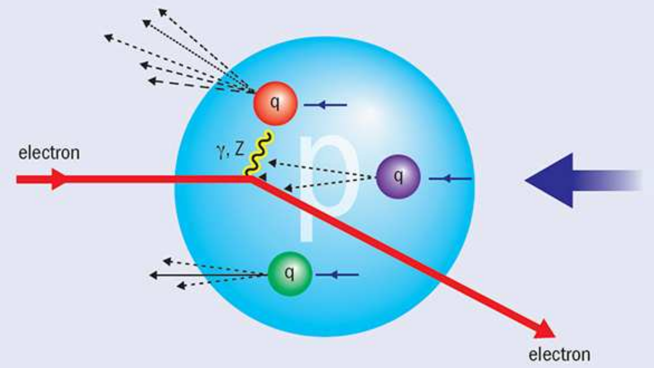


Anyway...

Neutral current: electron in the final state

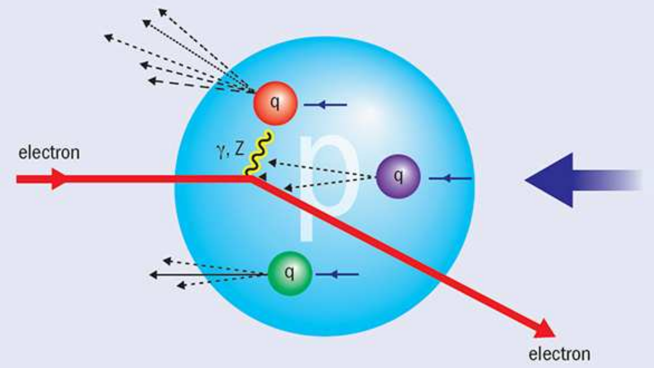


Neutral current: electron in the final state



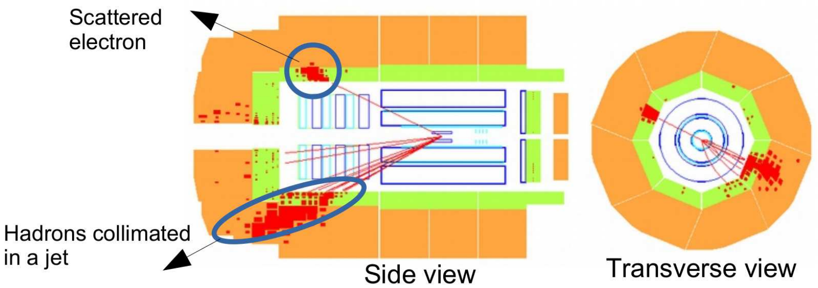
- electron in the final state (+ hadronic jet from scattered q);

Neutral current: electron in the final state



- electron in the final state (+ hadronic jet from scattered q);
- kinematical quantities x and Q^2 'easily' measured (in principle ...).

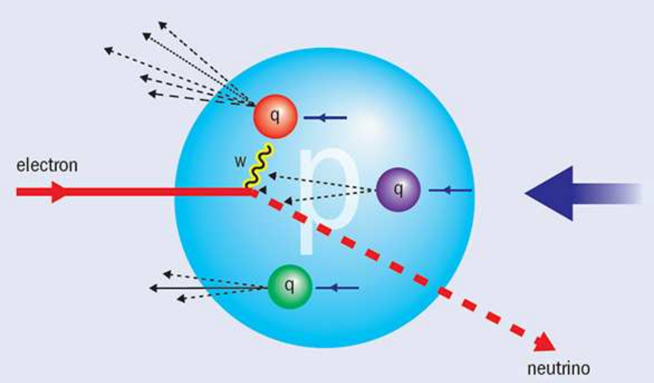
Neutral current: electron in the final state



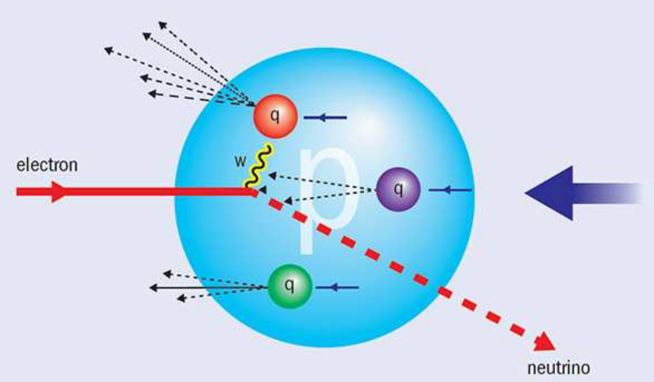
Large electron scattering angle (rare event)

VERY HARD scattering (rare event)

Charged current: neutrino ‘in the final state’

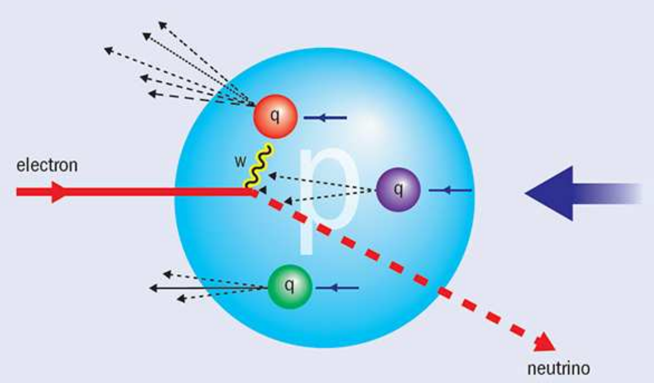


Charged current: neutrino ‘in the final state’



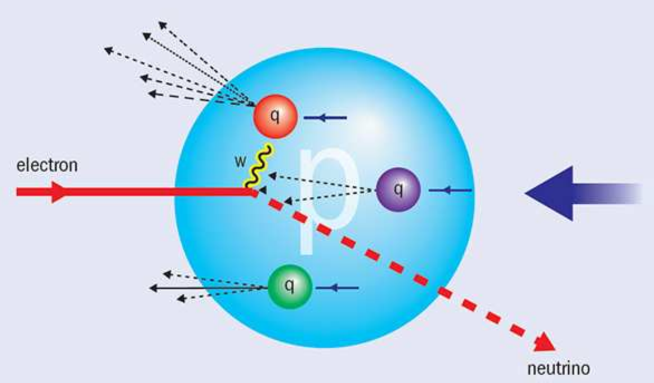
- final state neutrino not directly observable;
- x and Q^2 have to be measured only from (‘current’) jet hadrons. . .

Charged current: neutrino 'in the final state'



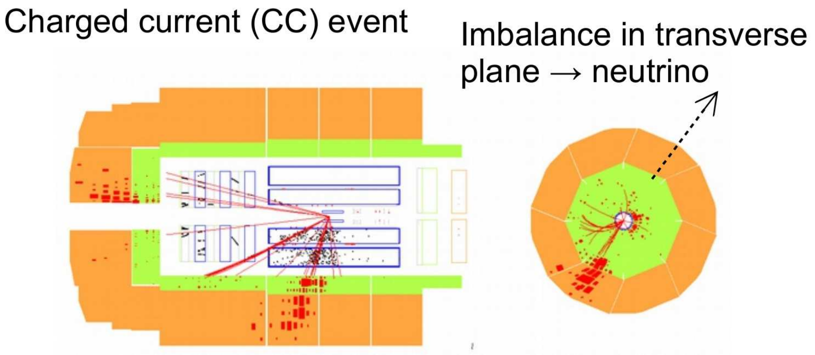
- final state neutrino not directly observable;
- x and Q^2 have to be measured only from ('current') jet hadrons...
- ...many of which are lost in the beam pipe, depending on x and Q^2.

Charged current: neutrino 'in the final state'



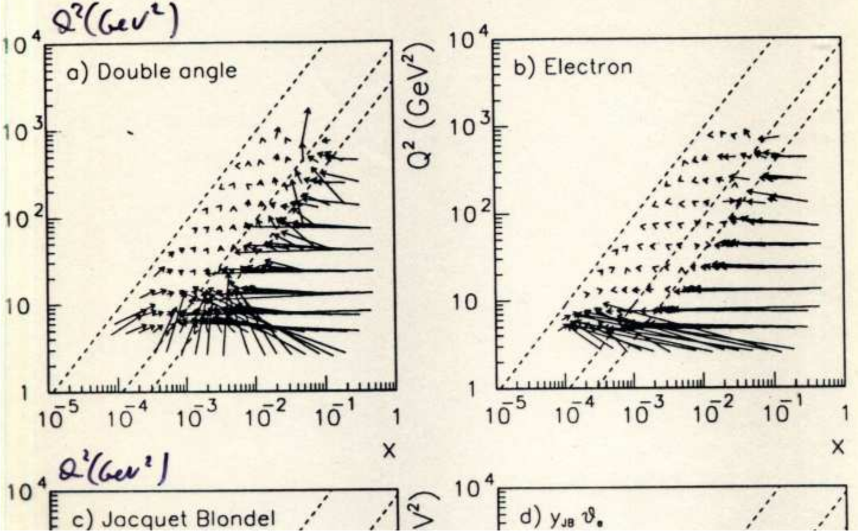
- final state neutrino not directly observable;
- x and Q^2 have to be measured only from ('current') jet hadrons...
- ...many of which are lost in the beam pipe, depending on x and Q^2.
- Measurement of x and Q^2 becomes challenging!!

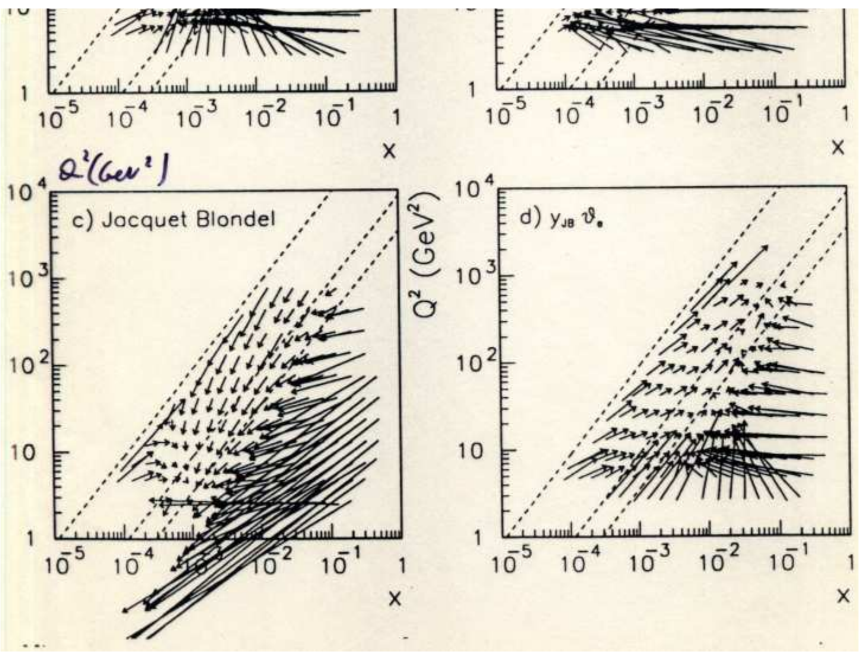
Charged current: neutrino 'in the final state'



VERY HARD scattering (rare event)

Esempio di migrazioni di $\{\Delta x, \Delta Q^2\}$
a seconda dei metodi di ricostruzione





Q²(GeV²)
c) Jacquet Blondel
d) y_{JB} ϑ_e
Q² (GeV²)
x

Why the unfolding? The Physics case

- At that time I was working in the ZEUS experiment at the e-p collider HERA in Hamburg.
  - Measuring the DIS quantities x and Q^2 was challenging in some physical regions, expecially when charged currents were involved, i.e. with ν_e in the final state.

  $\rightarrow$ Not simply Gaussian 'smearing' of the reconstructed quantities wrt the physical ones

How to go back

from reconstructed quantities to physical quantities?

Why the unfolding? The Physics case

- At that time I was working in the ZEUS experiment at the e-p collider HERA in Hamburg.
  - Measuring the DIS quantities x and Q^2 was challenging in some physical regions, expecially when charged currents were involved, i.e. with ν_e in the final state.

  $\rightarrow$ Not simply Gaussian 'smearing'
  of the reconstructed quantities wrt the physical ones

How to go back

from reconstructed quantities
to physical quantities?

Note: 2D smearing: $\rightarrow$ 2D unfolding!

Why the unfolding? The Physics case

- At that time I was working in the ZEUS experiment at the e-p collider HERA in Hamburg.
  - Measuring the DIS quantities x and Q^2 was challenging in some physical regions, expecially when charged currents were involved, i.e. with ν_e in the final state.

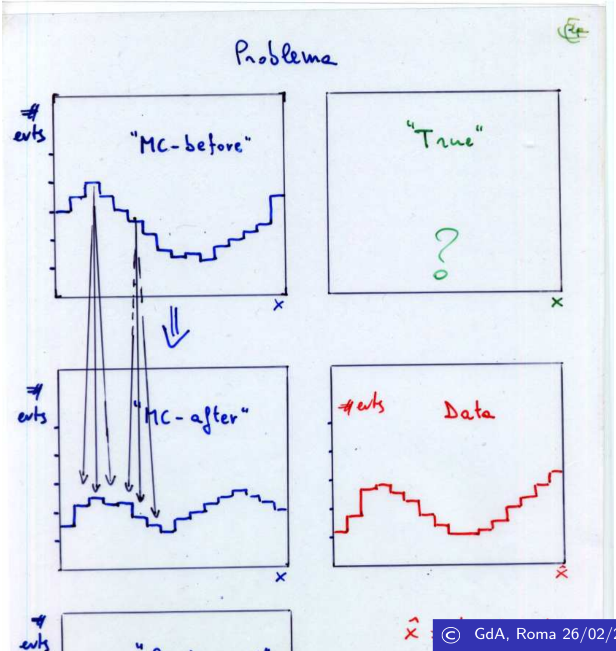
  $\rightarrow$ Not simply Gaussian 'smearing' of the reconstructed quantities wrt the physical ones

How to go back

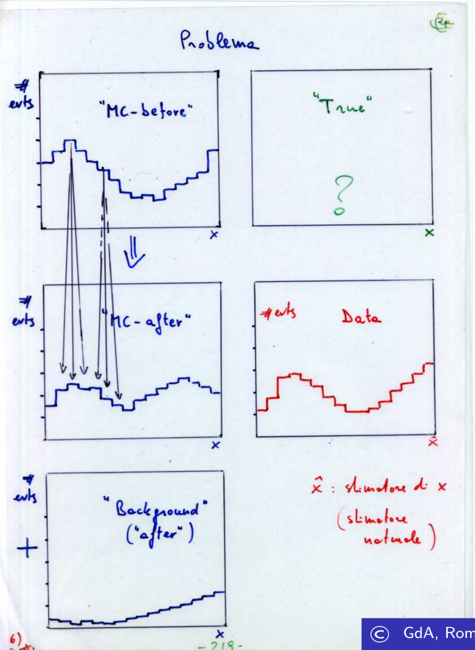
from reconstructed quantities
to physical quantities?

Note: 2D smearing: $\rightarrow$ 2D unfolding!
$\rightarrow$ **Multidimensional unfolding**

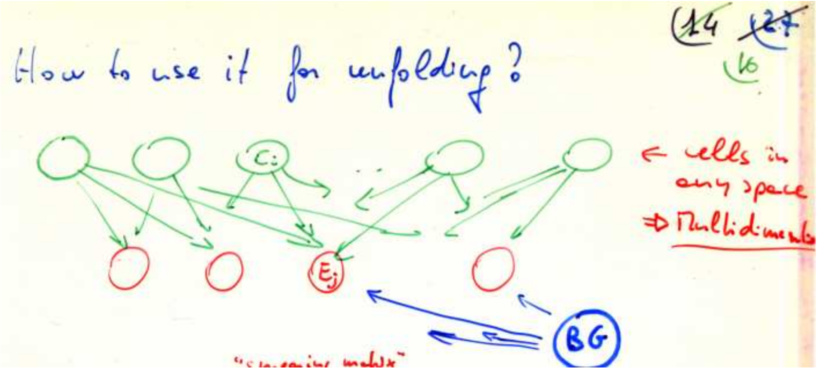
The basic idea — 1D for clarity



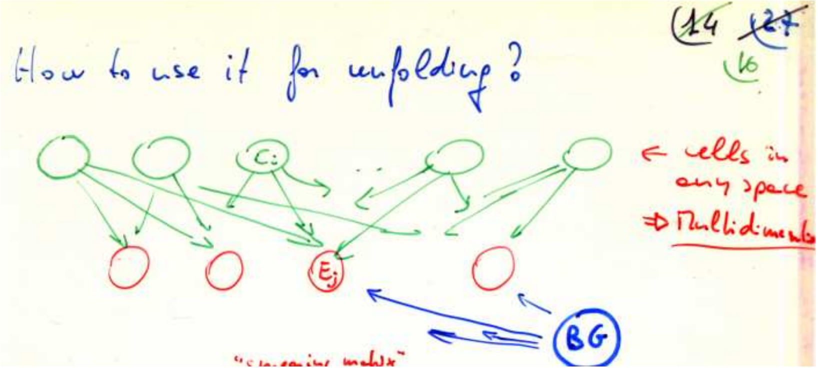
The basic idea — 1D for clarity



Causes-effects diagram

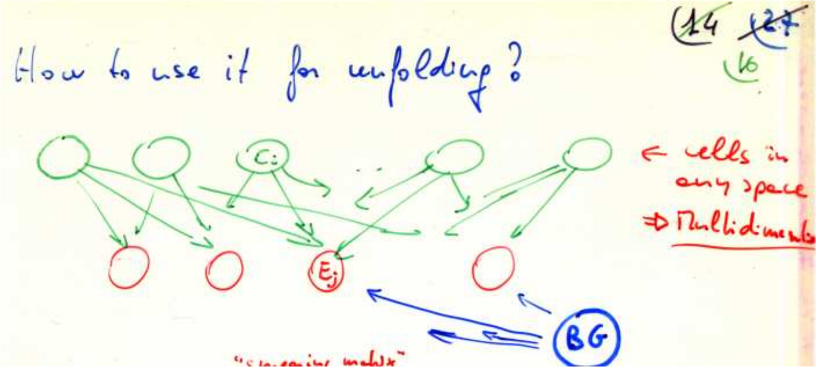


Causes-effects diagram



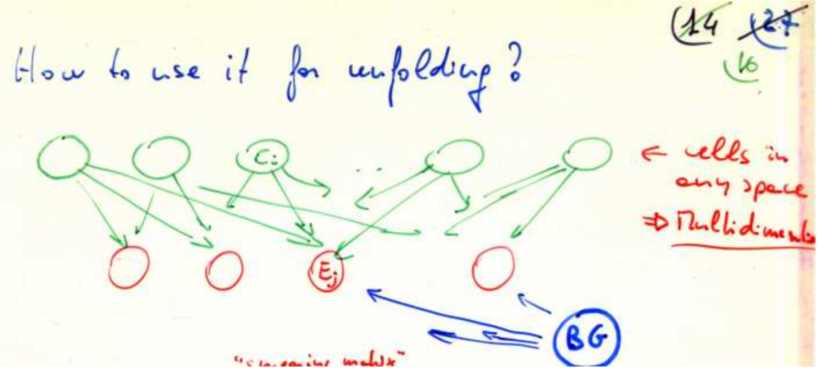
- ‘Cause bin’ C_i $\rightarrow$ ‘Effect bin’ E_j

Causes-effects diagram



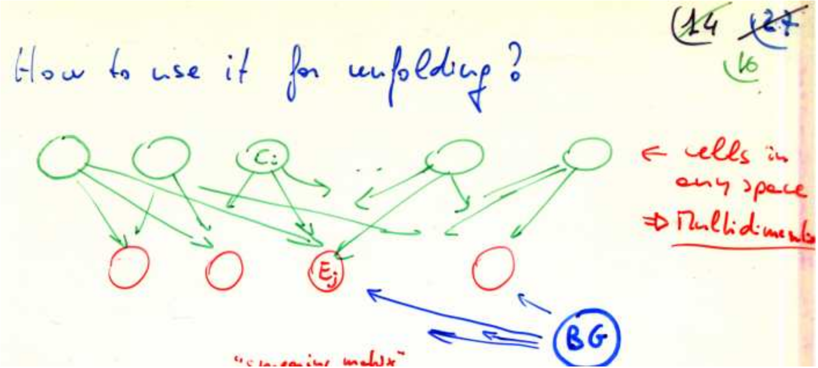
- 'Cause bin' C_i $\rightarrow$ 'Effect bin' E_j
- The C_i can be defined in any space:
  $\Rightarrow$ **multidimensional**

Causes-effects diagram



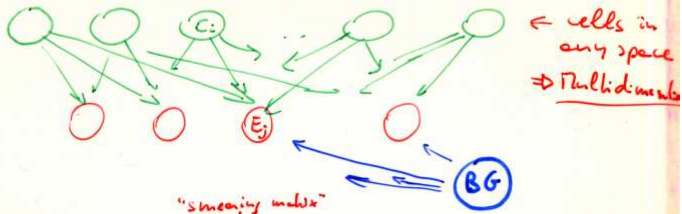
- 'Cause bin' C_i $\rightarrow$ 'Effect bin' E_j
- The C_i can be defined in any space:
  $\Rightarrow$ **multidimensional**
- Inefficiencies are described by adding to the effect cells a Trash cell (it will be clearer in a while)

Causes-effects diagram



- **Background** can also be naturally included, by just adding an extra cause bin:
  - obviously we have to 'know' (by MC) how it will contribute (see figure in the previous slide)
  - also several sources of background can be included.

Sharing the observed events: 1. evaluate $P(C_i|E_j)$



"smearing matrix"

$$P(C_i|E_j) = \frac{P(E_j|C_i)\cdot P_o(C_i)}{\sum_{l=1}^{n_c} P(E_j|C_l)\cdot P_o(C_l)}$$

- $P_o(C_i)$ – initial probabilities $\left(\sum_i P_o(C_i) = 1\right)$
  - wich? best guess
  - if $P_o(C_k)=0$ it will never be updated: [insensitivity of detector!]
- $\sum_i^{n_c} P(E_j|C_i) = 1$ each effect must come from one of the causes (eventually also background!)

Sharing the observed events: 2. evaluate $n(E_j) \rightarrow n(C_i)$

• $\varepsilon_i = \sum_{j=1}^{n_E} P(E_j\,|\,C_i)$ $\quad 0 \le \varepsilon_i \le 1$: efficiency for C_i to be observed as one of the mode E_j

$\boxed{\varepsilon_i \sim 0 : \text{"hic sunt leones"}}$

N_{obs} : $\underline{n}(E) = \{n(E_1), \dots, n(E_{n_E})\}$

$n(C_1)$ $\quad n(C_2)$ $\quad \dots \quad n(C_n)$

$$\hat{n}(C_i)\Big|_{observed} = \sum_{j=1}^{n_E} n(E_j)\cdot P(C_i\,|\,E_j)$$

x) ~~224~~)

-225-

(d19

Sharing the observed events: 2. evaluate $n(E_j) \rightarrow n(C_i)$

- $\epsilon_i = \sum_{j=1}^{n_E} P(E_j\,|\,C_i)$ $\quad 0 \le \epsilon_i \le 1$: ~~efficiency~~ for C_i to be observed as one of the mode E_j

$\boxed{\epsilon_i \sim 0 : \text{“hic sunt leones”}}$

N_{obs} : $\underline{n}(E) = \{n(E_1), \ldots\; n(E_{n_E})\}$

$n(C_1) \quad n(C_2) \quad n(C_n)$

$$\hat{n}(C_i)\Big|_{\text{observed}} = \sum_{j=1}^{n_E} n(E_j)\cdot P(C_i\,|\,E_j)$$

n) 22~~4~~)

-225-

. . . also taking into account the inefficiencies
(events which went to ‘Trash’)

$$\boxed{\hat{n}(C_i) = \frac{1}{\epsilon_i}\sum_{j=1}^{n_E} n(E_j)\cdot P(C_i\,|\,E_j)}$$

Continuing with the recollections

At this point you might have several doubts and questions (e.g. about priors) $\rightarrow$ Please wait a while.

Continuing with the recollections

At this point you might have several doubts and questions (e.g. about priors) $\rightarrow$ Please wait a while.

From the first attempt to the NIM paper

Continuing with the recollections

At this point you might have several doubts and questions (e.g. about priors) $\rightarrow$ Please wait a while.

From the first attempt to the NIM paper

- First Fortran version written a Saturday afternoon in Hamburg

Continuing with the recollections

At this point you might have several doubts and questions (e.g. about priors) $\rightarrow$ Please wait a while.

From the first attempt to the NIM paper

- First Fortran version written a Saturday afternoon in Hamburg
  - 'Obviously', *just to start*, a 'flat prior' on $P_0(C_i)$ was used

Continuing with the recollections

At this point you might have several doubts and questions (e.g. about priors) $\rightarrow$ Please wait a while.

From the first attempt to the NIM paper

- First Fortran version written a Saturday afternoon in Hamburg
  - 'Obviously', *just to start*, a 'flat prior' on $P_0(C_i)$ was used
  - ... but the code worked immediately almost too well

Continuing with the recollections

At this point you might have several doubts and questions (e.g. about priors) $\rightarrow$ Please wait a while.

From the first attempt to the NIM paper

- First Fortran version written a Saturday afternoon in Hamburg
  - 'Obviously', *just to start*, a 'flat prior' on $P_0(C_i)$ was used
  - ... but the code worked immediately almost too well (contrary to what usually happens, when we write a piece of code and then we spend some time to debug it, the *problem* was to understand **why it was working so good!**)

Continuing with the recollections

At this point you might have several doubts and questions (e.g. about priors) $\rightarrow$ Please wait a while.

From the first attempt to the NIM paper

- First Fortran version written a Saturday afternoon in Hamburg
  - 'Obviously', *just to start*, a 'flat prior' on $P_0(C_i)$ was used
  - ... but the code worked immediately almost too well (contrary to what usually happens, when we write a piece of code and then we spend some time to debug it, the *problem* was to understand **why it was working so good!**)
  - Playing with simulated events, the unfolded spectrum was between the prior and the 'true distribution' (of MC events):

Continuing with the recollections

At this point you might have several doubts and questions (e.g. about priors) $\rightarrow$ Please wait a while.

From the first attempt to the NIM paper

- First Fortran version written a Saturday afternoon in Hamburg
  - 'Obviously', *just to start*, a 'flat prior' on $P_0(C_i)$ was used
  - ... but the code worked immediately almost too well (contrary to what usually happens, when we write a piece of code and then we spend some time to debug it, the *problem* was to understand **why it was working so good!**)
  - Playing with simulated events, the unfolded spectrum was between the prior and the 'true distribution' (of MC events): $\rightarrow$ iterations, just as a pragmatic solution (please wait)

Continuing with the recollections

At this point you might have several doubts and questions (e.g. about priors) $\rightarrow$ Please wait a while.

From the first attempt to the NIM paper

- First Fortran version written a Saturday afternoon in Hamburg
  - 'Obviously', *just to start*, a 'flat prior' on $P_0(C_i)$ was used
  - ... but the code worked immediately almost too well (contrary to what usually happens, when we write a piece of code and then we spend some time to debug it, the *problem* was to understand **why it was working so good!**)
  - Playing with simulated events, the unfolded spectrum was between the prior and the 'true distribution' (of MC events): $\rightarrow$ iterations, just as a pragmatic solution (please wait)
- First paper: ZEUS-Note 93-127, December 1993.

Continuing with the recollections

At this point you might have several doubts and questions (e.g. about priors) → Please wait a while.

From the first attempt to the NIM paper

- First Fortran version written a Saturday afternoon in Hamburg
  - 'Obviously', *just to start*, a 'flat prior' on $P_0(C_i)$ was used
  - ... but the code worked immediately almost too well (contrary to what usually happens, when we write a piece of code and then we spend some time to debug it, the *problem* was to understand **why it was working so good!**)
  - Playing with simulated events, the unfolded spectrum was between the prior and the 'true distribution' (of MC events): → iterations, just as a pragmatic solution (please wait)
- First paper: ZEUS-Note 93-127, December 1993.
- Several months needed until I was convinced about the strategies and the evualuation of uncertainties, including correlations (technical details not shown here):

Continuing with the recollections

At this point you might have several doubts and questions (e.g. about priors) $\rightarrow$ Please wait a while.

From the first attempt to the NIM paper

- First Fortran version written a Saturday afternoon in Hamburg
  - 'Obviously', *just to start*, a 'flat prior' on $P_0(C_i)$ was used
  - ... but the code worked immediately almost too well (contrary to what usually happens, when we write a piece of code and then we spend some time to debug it, the *problem* was to understand **why it was working so good!**)
  - Playing with simulated events, the unfolded spectrum was between the prior and the 'true distribution' (of MC events): $\rightarrow$ iterations, just as a pragmatic solution (please wait)
- First paper: ZEUS-Note 93-127, December 1993.
- Several months needed until I was convinced about the strategies and the evualuation of uncertainties, including correlations (technical details not shown here):
  $\rightarrow$ DESY 94-099, June 1994, later submitted to NIM:

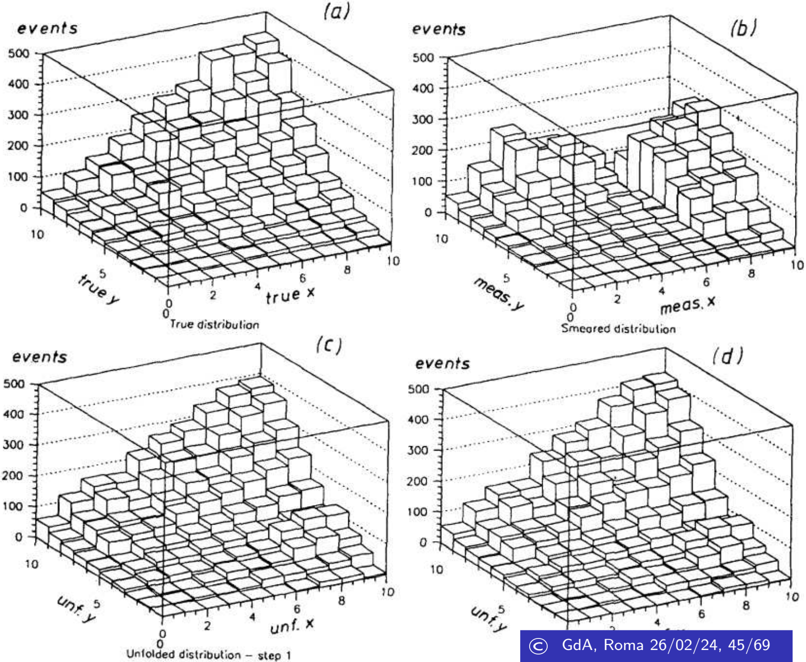
Continuing with the recollections

At this point you might have several doubts and questions (e.g. about priors) → Please wait a while.

From the first attempt to the NIM paper

- First Fortran version written a Saturday afternoon in Hamburg
  - 'Obviously', *just to start*, a 'flat prior' on $P_0(C_i)$ was used
  - ... but the code worked immediately almost too well (contrary to what usually happens, when we write a piece of code and then we spend some time to debug it, the *problem* was to understand **why it was working so good!**)
  - Playing with simulated events, the unfolded spectrum was between the prior and the 'true distribution' (of MC events): → iterations, just as a pragmatic solution (please wait)
- First paper: ZEUS-Note 93-127, December 1993.
- Several months needed until I was convinced about the strategies and the evuaulation of uncertainties, including correlations (technical details not shown here):
  → DESY 94-099, June 1994, later submitted to NIM:
  → 'NIM' A362 (1995) 487

Example of 2D unfolding (from NIM paper)



Initial **prior anxiety**

A multidimensional unfolding method based on Bayes' theorem

G. D'Agostini *

Università "La Sapienza" and INFN, Roma, Italy

Received 5 August 1994; revised form received 2 March 1995

Abstract

Bayes' theorem offers a natural way to unfold experimental distributions in order to get the best estimates of the true ones. The weak point of the Bayes approach, namely the need of the knowledge of the initial distribution, can be overcome by an iterative procedure. Since the method proposed here does not make use of continuous variables, but simply of cells in the spaces of the true and of the measured quantities, it can be applied in multidimensional problems.

Initial **prior anxiety**

A multidimensional unfolding method based on Bayes' theorem

G. D'Agostini *

Università "La Sapienza" and INFN, Roma, Italy

Received 5 August 1994; revised form received 2 March 1995

Abstract

Bayes' theorem offers a natural way to unfold experimental distributions in order to get the best estimates of the true ones. The weak point of the Bayes approach, namely the need of the knowledge of the initial distribution, can be overcome by an iterative procedure. Since the method proposed here does not make use of continuous variables, but simply of cells in the spaces of the true and of the measured quantities, it can be applied in multidimensional problems.

- "The weak point of the Bayes approach, namely the need of the knowledge of the initial distribution, . . ."

Initial **prior anxiety**

A multidimensional unfolding method based on Bayes' theorem

G. D'Agostini *

Università "La Sapienza" and INFN, Roma, Italy

Received 5 August 1994; revised form received 2 March 1995

Abstract

Bayes' theorem offers a natural way to unfold experimental distributions in order to get the best estimates of the true ones. The weak point of the Bayes approach, namely the need of the knowledge of the initial distribution, can be overcome by an iterative procedure. Since the method proposed here does not make use of continuous variables, but simply of cells in the spaces of the true and of the measured quantities, it can be applied in multidimensional problems.

- "The weak point of the Bayes approach, namely the need of the knowledge of the initial distribution, . . ."
  ??
  - **"Week point"?**

Initial **prior anxiety**

A multidimensional unfolding method based on Bayes' theorem

G. D'Agostini *

Università "La Sapienza" and INFN, Roma, Italy

Received 5 August 1994; revised form received 2 March 1995

Abstract

Bayes' theorem offers a natural way to unfold experimental distributions in order to get the best estimates of the true ones. The weak point of the Bayes approach, namely the need of the knowledge of the initial distribution, can be overcome by an iterative procedure. Since the method proposed here does not make use of continuous variables, but simply of cells in the spaces of the true and of the measured quantities, it can be applied in multidimensional problems.

- "The weak point of the Bayes approach, namely the need of the knowledge of the initial distribution, ..."
  - **"Week point"?**
    - **STRONG POINT**, because it allows to combine new pieces of information with prior knowledge.
    - It is an **ESSENTIAL INGREDIENT** of probability theory if we want to reason from effects to causes (see e.g. Laplace and Gauss).

Initial **prior anxiety**

A multidimensional unfolding method based on Bayes' theorem

G. D'Agostini *

Università "La Sapienza" and INFN, Roma, Italy

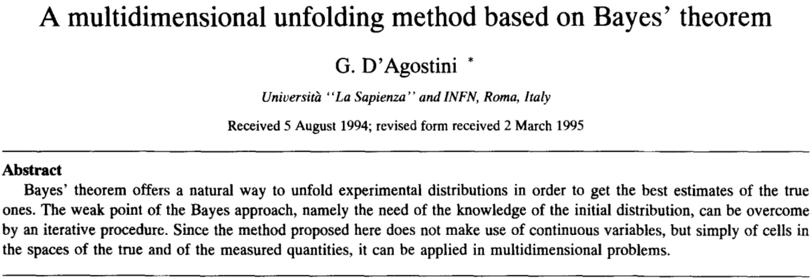
Received 5 August 1994; revised form received 2 March 1995

Abstract

Bayes' theorem offers a natural way to unfold experimental distributions in order to get the best estimates of the true ones. The weak point of the Bayes approach, namely the need of the knowledge of the initial distribution, can be overcome by an iterative procedure. Since the method proposed here does not make use of continuous variables, but simply of cells in the spaces of the true and of the measured quantities, it can be applied in multidimensional problems.

- "The weak point of the Bayes approach, namely the need of the knowledge of the initial distribution, ..."
  - **"Week point"?**
    - **STRONG POINT**, because it allows to combine new pieces of information with prior knowledge.
    - It is an **ESSENTIAL INGREDIENT** of probability theory if we want to reason from effects to causes (see e.g. Laplace and Gauss).
  - **"the knowledge of the initial distribution"**
    - there is not such an ABSTRACT initial distribution that we have to know before we tackle an inferential problem.

Initial **prior anxiety**


A multidimensional unfolding method based on Bayes' theorem

G. D'Agostini *

Università "La Sapienza" and INFN, Roma, Italy

Received 5 August 1994; revised form received 2 March 1995

Abstract

Bayes' theorem offers a natural way to unfold experimental distributions in order to get the best estimates of the true ones. The weak point of the Bayes approach, namely the need of the knowledge of the initial distribution, can be overcome by an iterative procedure. Since the method proposed here does not make use of continuous variables, but simply of cells in the spaces of the true and of the measured quantities, it can be applied in multidimensional problems.

- "The weak point of the Bayes approach, namely the need of the knowledge of the initial distribution, ..."
  - **"Week point"?**
    - **STRONG POINT**, because it allows to combine new pieces of information with prior knowledge.
    - It is an **ESSENTIAL INGREDIENT** of probability theory if we want to reason from effects to causes (see e.g. Laplace and Gauss).
  - **"the knowledge of the initial distribution"**
    - there is not such an ABSTRACT initial distribution that we have to know before we tackle an inferential problem.

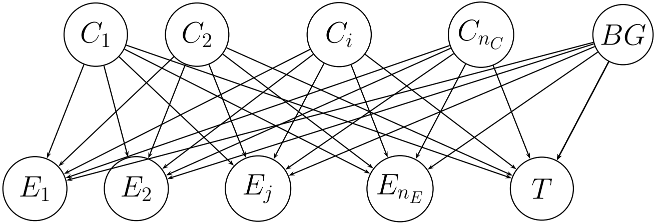
→ see more

A kind of technical recap

now that basic ideas shoud be received

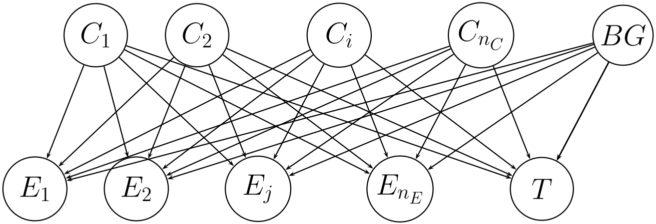
(focusing on 1D unfolding, with life examples)

Discretized unfolding



(T: 'trash')

Discretized unfolding



(T: 'trash')

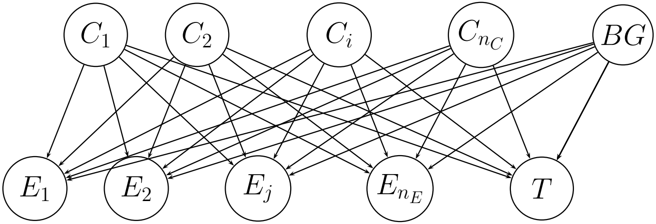
Background ('known') is just a an extra cell

$$n_C \to n_C + 1$$

(Hereafter just included among the causes – several BG's can be included)

Vectors of interest:

Discretized unfolding



(T: 'trash')

Background ('known') is just a an extra cell

$$n_C \to n_C + 1$$

(Hereafter just included among the causes – several BG's can be included)

Vectors of interest:

$\boldsymbol{x}_C$: *true spectrum* (nr of events in cause bins)

$\boldsymbol{x}_E$: *observed spectrum* (nr of events in effect bins)

Remark: parametric inference Vs unfolding

$f(\boldsymbol{\theta} \,|\, \{\boldsymbol{x}, \boldsymbol{y}\})$:

Remark: parametric inference Vs unfolding

$f(\boldsymbol{\theta}\,|\,\{\boldsymbol{x}, \boldsymbol{y}\})$:

$\rightarrow$ probabilistic parametric inference

$\Rightarrow$ it relies on the kind of functions parametrized by $\boldsymbol{\theta}$

$$\boldsymbol{\mu}_y = \boldsymbol{\mu}_y(\boldsymbol{\mu}_x; \boldsymbol{\theta})$$

with (because of errors)

$$\begin{array}{ccc} \mu_{x_i} & \longrightarrow & x_i \\ \mu_{y_i} & \longrightarrow & y_i \end{array}$$

Remark: parametric inference Vs unfolding

$f(\boldsymbol{\theta}\,|\,\{\boldsymbol{x},\boldsymbol{y}\})$:

$\rightarrow$ probabilistic parametric inference

$\Rightarrow$ it relies on the kind of functions parametrized by $\boldsymbol{\theta}$

$$\boldsymbol{\mu}_y = \boldsymbol{\mu}_y(\boldsymbol{\mu}_x;\boldsymbol{\theta})$$

with (because of errors)

$$\begin{array}{ccc} \mu_{x_i} & \longrightarrow & x_i \\ \mu_{y_i} & \longrightarrow & y_i \end{array}$$

$\Rightarrow$ data distilled into $\boldsymbol{\theta}$: $\rightarrow$ $f(\boldsymbol{\theta}\,|\,\text{data})$

Remark: parametric inference Vs unfolding

$f(\boldsymbol{\theta}\,|\,\{\boldsymbol{x},\boldsymbol{y}\})$:

$\rightarrow$ probabilistic parametric inference

$\Rightarrow$ it relies on the kind of functions parametrized by $\boldsymbol{\theta}$

$$\boldsymbol{\mu}_y = \boldsymbol{\mu}_y(\boldsymbol{\mu}_x;\boldsymbol{\theta})$$

with (because of errors)

$$\begin{array}{ccc} \mu_{x_i} & \longrightarrow & x_i \\ \mu_{y_i} & \longrightarrow & y_i \end{array}$$

$\Rightarrow$ data distilled into $\boldsymbol{\theta}$: $\rightarrow f(\boldsymbol{\theta}\,|\,\text{data})$

BUT sometimes we wish to interpret the data as little as possible

$\Rightarrow$ just public 'something equivalent' to an experimental distribution, with the bin contents fluctuating according to an underlying multinomial distribution, but having possibly got rid of physical and instrumental distortions, as well as of background.

Remark: parametric inference Vs unfolding

$f(\boldsymbol{\theta}\,|\,\{\boldsymbol{x},\boldsymbol{y}\})$:

$\rightarrow$ probabilistic parametric inference

$\Rightarrow$ it relies on the kind of functions parametrized by $\boldsymbol{\theta}$

$$\boldsymbol{\mu}_y = \boldsymbol{\mu}_y(\boldsymbol{\mu}_x;\boldsymbol{\theta})$$

with (because of errors)

$$\begin{array}{ccc} \mu_{x_i} & \longrightarrow & x_i \\ \mu_{y_i} & \longrightarrow & y_i \end{array}$$

$\Rightarrow$ data distilled into $\boldsymbol{\theta}$: $\rightarrow f(\boldsymbol{\theta}\,|\,\text{data})$

BUT sometimes we wish to interpret the data as little as possible

$\Rightarrow$ just public 'something equivalent' to an experimental distribution, with the bin contents fluctuating according to an underlying multinomial distribution, but having possibly got rid of physical and instrumental distortions, as well as of background.

$\Rightarrow$ **Unfolding** (deconvolution)

Smearing matrix $\rightarrow$ unfolding matrix

Invert smearing matrix?

Smearing matrix $\rightarrow$ unfolding matrix

Invert smearing matrix?

In general is a bad idea:

not a rotational problem
but an inferential problem!

Smearing matrix $\rightarrow$ unfolding matrix

Imagine $S = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}$: $\rightarrow U = S^{-1} = \begin{pmatrix} 1.33 & -0.33 \\ -0.33 & 1.33 \end{pmatrix}$

Let the true be $s_t = \begin{pmatrix} 10 \\ 0 \end{pmatrix}$: $\rightarrow s_m = S \cdot s_t = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$;

If we measure $s_m = \begin{pmatrix} 8 \\ 2 \end{pmatrix} \rightarrow S^{-1} \cdot s_m = \begin{pmatrix} 10 \\ 0 \end{pmatrix}$ $\surd$

Smearing matrix $\rightarrow$ unfolding matrix

Imagine $S = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}$: $\rightarrow U = S^{-1} = \begin{pmatrix} 1.33 & -0.33 \\ -0.33 & 1.33 \end{pmatrix}$

Let the true be $s_t = \begin{pmatrix} 10 \\ 0 \end{pmatrix}$: $\rightarrow s_m = S \cdot s_t = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$;

If we measure $s_m = \begin{pmatrix} 8 \\ 2 \end{pmatrix} \rightarrow S^{-1} \cdot s_m = \begin{pmatrix} 10 \\ 0 \end{pmatrix}$ $\surd$

BUT

if we had measured $\begin{pmatrix} 9 \\ 1 \end{pmatrix} \rightarrow S^{-1} \cdot s_m = \begin{pmatrix} 11.7 \\ -1.7 \end{pmatrix}$

if we had measured $\begin{pmatrix} 10 \\ 0 \end{pmatrix} \rightarrow S^{-1} \cdot s_m = \begin{pmatrix} 13.3 \\ -3.3 \end{pmatrix}$

Smearing matrix $\rightarrow$ unfolding matrix

Imagine $S = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}$: $\rightarrow U = S^{-1} = \begin{pmatrix} 1.33 & -0.33 \\ -0.33 & 1.33 \end{pmatrix}$

Let the true be $s_t = \begin{pmatrix} 10 \\ 0 \end{pmatrix}$: $\rightarrow s_m = S \cdot s_t = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$;

If we measure $s_m = \begin{pmatrix} 8 \\ 2 \end{pmatrix} \rightarrow S^{-1} \cdot s_m = \begin{pmatrix} 10 \\ 0 \end{pmatrix}$ $\surd$

BUT

if we had measured $\begin{pmatrix} 9 \\ 1 \end{pmatrix} \rightarrow S^{-1} \cdot s_m = \begin{pmatrix} 11.7 \\ -1.7 \end{pmatrix}$

if we had measured $\begin{pmatrix} 10 \\ 0 \end{pmatrix} \rightarrow S^{-1} \cdot s_m = \begin{pmatrix} 13.3 \\ -3.3 \end{pmatrix}$

Indeed, matrix inversion is recognized to producing 'crazy spectra' and even negative values (unless there are so large numbers of events in bins such that fluctuations around expectations are negligeable)

Bin to bin analysis?

En passant:

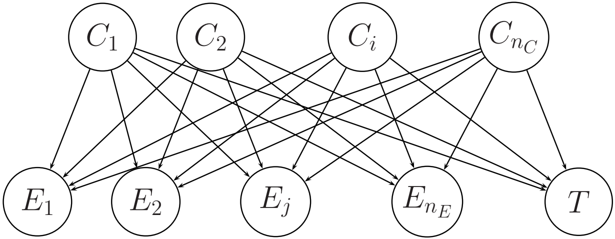
- OK if the are no migrations:
  $\rightarrow$ each bin is an 'independent issue',
  treated with a binomial process, given some efficiencies.

Bin to bin analysis?

En passant:

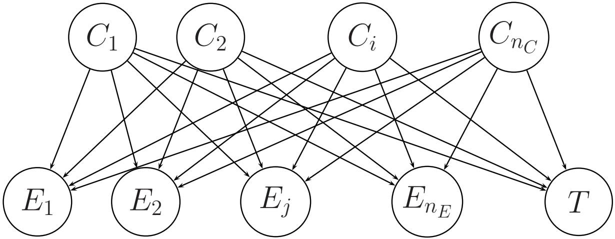
- OK if the are no migrations:
  $\rightarrow$ each bin is an 'independent issue',
  treated with a binomial process, given some efficiencies.
- Otherwise
  - 'error analysis' troublesome
    (just imagine e.g. that a bin has an 'efficiency' > 1,
    because of migrations from other bins);
  - iteration is important
    (efficiencies depend on 'true distribution')

Discretized unfolding



(T: 'trash')

Discretized unfolding

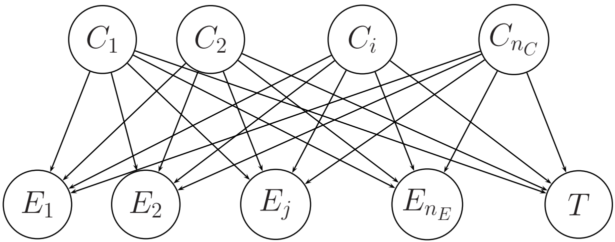


(T: 'trash')

$\boldsymbol{x}_C$: *true spectrum* (nr of events in cause bins)

$\boldsymbol{x}_E$: *observed spectrum* (nr of events in effect bins)

Discretized unfolding



(T: 'trash')

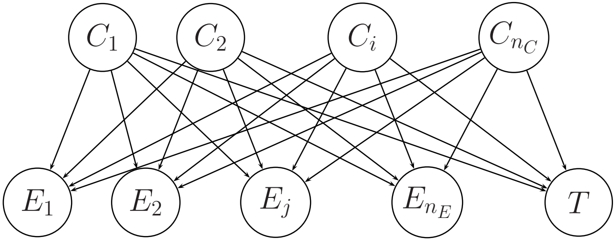
$\boldsymbol{x}_C$: *true spectrum* (nr of events in cause bins)

$\boldsymbol{x}_E$: *observed spectrum* (nr of events in effect bins)

Our aim:

- **not** to find *the* true spectrum
- but, more modestly, rank in beliefs all possible spectra that might have caused the observed one:
  $\Rightarrow P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, I)$

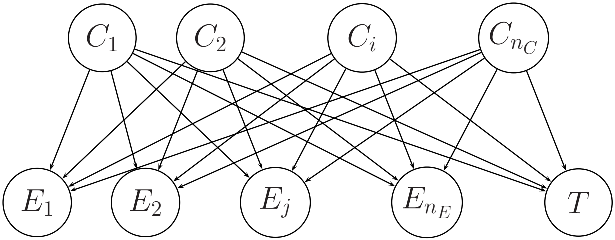
Discretized unfolding



(T: 'trash')

- $P(\boldsymbol{x}_C \mid \boldsymbol{x}_E, I)$ depends on the knowledge of *smearing matrix* Λ,
  with $\lambda_{ji} \equiv P(E_j \mid C_i,\, I)$.

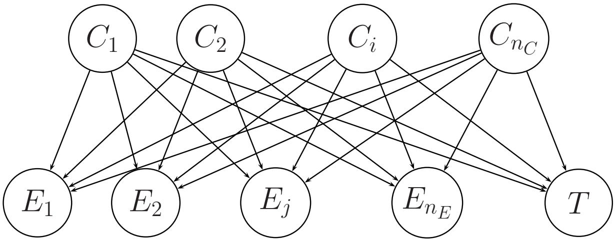
Discretized unfolding



(T: 'trash')

- $P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, I)$ depends on the knowledge of *smearing matrix* Λ, with $\lambda_{ji} \equiv P(E_j \,|\, C_i,\, I)$.
- but Λ is itself uncertain, because inferred from MC simulation: $\Rightarrow f(\Lambda \,|\, I)$

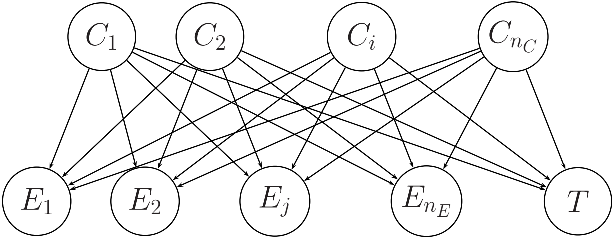
Discretized unfolding



(T: 'trash')

- $P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, I)$ depends on the knowledge of *smearing matrix* Λ,
  with $\lambda_{ji} \equiv P(E_j \,|\, C_i,\, I)$.
- but Λ is itself uncertain, because inferred from MC simulation:
  $\Rightarrow f(\Lambda \,|\, I)$
- for each possible Λ we have a pdf of spectra:
  $\rightarrow P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, \Lambda, I)$

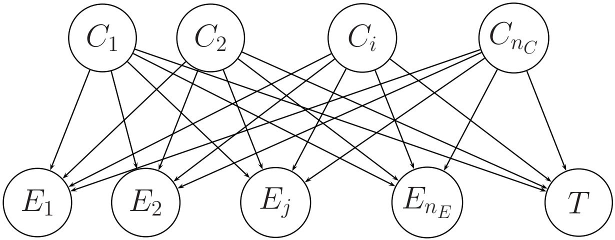
Discretized unfolding



(T: 'trash')

- $P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, I)$ depends on the knowledge of *smearing matrix* Λ,
  with $\lambda_{ji} \equiv P(E_j \,|\, C_i,\, I)$.
- but Λ is itself uncertain, because inferred from MC simulation:
  $\Rightarrow f(\Lambda \,|\, I)$
- for each possible Λ we have a pdf of spectra:
  $\rightarrow P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, \Lambda, I)$
  $\Rightarrow P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, I) = \int P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, \Lambda,\, I)\, f(\Lambda \,|\, I)\, \mathrm{d}\Lambda$ [by MC!]

Discretized unfolding

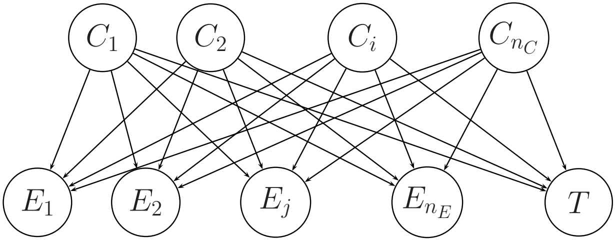


(T: 'trash')

- Bayes theorem:

$$P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, \Lambda, I) \;\; \propto \;\; P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C, \Lambda, I) \cdot P(\boldsymbol{x}_C \,|\, I) \,.$$

Discretized unfolding



(T: 'trash')

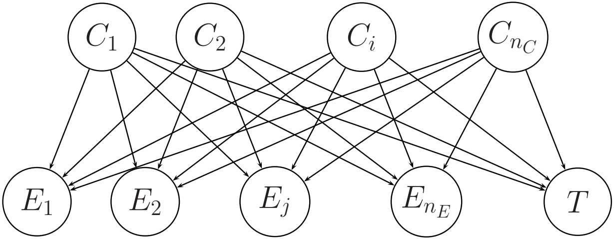
- Bayes theorem:

$$P(\boldsymbol{x}_C\,|\,\boldsymbol{x}_E, \Lambda, I) \;\; \propto \;\; P(\boldsymbol{x}_E\,|\,\boldsymbol{x}_C, \Lambda, I)\cdot P(\boldsymbol{x}_C\,|\,I)\,.$$

- Indifference w.r.t. all possible spectra

$$P(\boldsymbol{x}_C\,|\,\boldsymbol{x}_E, \Lambda, I) \;\; \propto \;\; P(\boldsymbol{x}_E\,|\,\boldsymbol{x}_C, \Lambda, I)$$

$P(\boldsymbol{x}_E \,|\, x_{C_i},\, \Lambda,\, I)$



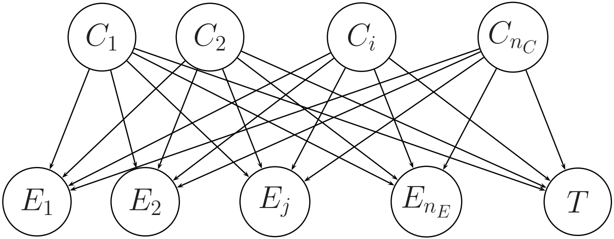
Given a certain number of events in a cause-bin $x(C_i)$, the number of events in the effect-bins, included the 'trash' one, is described by a multinomial distribution:

$$\boldsymbol{x}_E\big|_{x(C_i)} \;\sim\; \mathrm{Mult}[x(C_i), \boldsymbol{\lambda}_i]\,,$$

with

$$\begin{aligned}
\boldsymbol{\lambda}_i &= \{\lambda_{1,i},\, \lambda_{2,i}, \ldots,\, \lambda_{n_E+1,i}\} \\
&= \{P(E_1 \,|\, C_i,\, I),\, P(E_2 \,|\, C_i,\, I), \ldots,\, P(E_{n_E+1,i} \,|\, C_i,\, I)\}
\end{aligned}$$

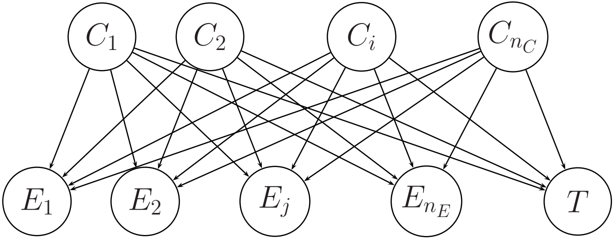
$P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C,\, \Lambda,\, I)$



$\boldsymbol{x}_E\big|_{x(C_i)}$ multinomial random vector,

$\Rightarrow$ $\boldsymbol{x}_E\big|_{\boldsymbol{x}(C)}$ sum of several multinomials.

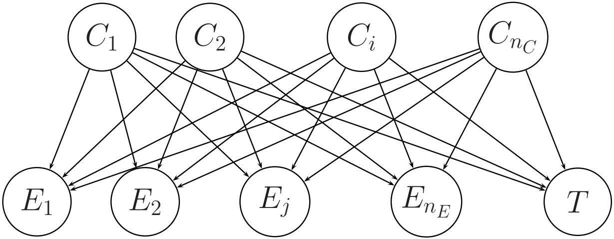
$P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C,\, \Lambda,\, I)$



$\boldsymbol{x}_E\big|_{x(C_i)}$ multinomial random vector,
$\Rightarrow$ $\boldsymbol{x}_E\big|_{\boldsymbol{x}(C)}$ sum of several multinomials.
BUT

no 'easy' expression for $P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C, \Lambda, I)$

$P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C,\, \Lambda,\, I)$



$\boldsymbol{x}_E\big|_{x(C_i)}$ multinomial random vector,

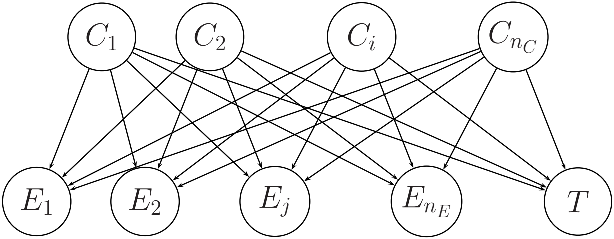
$\Rightarrow$ $\boldsymbol{x}_E\big|_{\boldsymbol{x}(C)}$ sum of several multinomials.

BUT

no 'easy' expression for $P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C, \Lambda, I)$

$\Rightarrow$ STUCK!

$P(\mathbf{x}_E \,|\, \mathbf{x}_C,\, \Lambda,\, I)$



$\mathbf{x}_E\big|_{x(C_i)}$ multinomial random vector,

$\Rightarrow$ $\mathbf{x}_E\big|_{\mathbf{x}(C)}$ sum of several multinomials.

BUT

no 'easy' expression for $P(\mathbf{x}_E \,|\, \mathbf{x}_C, \Lambda, I)$

$\Rightarrow$ STUCK!

$\Rightarrow$ Change strategy

The rescue trick

Instead of using the original probability inversion (applied directly) to spectra

$$P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, \Lambda, I) \;\; \propto \;\; P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C, \Lambda, I) \cdot P(\boldsymbol{x}_C \,|\, I)\,,$$

we restart from

$$P(C_i \,|E_j, I) \;\; \propto \;\; P(E_j \,|\, C_i, I) \cdot P(C_i \,|\, I).$$

The rescue trick

Instead of using the original probability inversion (applied directly) to spectra

$$P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, \Lambda, I) \;\;\propto\;\; P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C, \Lambda, I) \cdot P(\boldsymbol{x}_C \,|\, I)\,,$$

we restart from

$$P(C_i \,|E_j, I) \;\;\propto\;\; P(E_j \,|\, C_i, I) \cdot P(C_i \,|\, I).$$

Consequences:

1. the sharing of observed events among the cause bins needs to be performed 'by hand';

The rescue trick

Instead of using the original probability inversion (applied directly) to spectra

$$P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, \Lambda, I) \;\;\propto\;\; P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C, \Lambda, I) \cdot P(\boldsymbol{x}_C \,|\, I)\,,$$

we restart from

$$P(C_i \,|E_j, I) \;\;\propto\;\; P(E_j \,|\, C_i, I) \cdot P(C_i \,|\, I).$$

Consequences:

1. the sharing of observed events among the cause bins needs to be performed 'by hand';
2. a uniform prior $P(C_i \,|\, I) = k$ <u>does not</u> mean indifference over all possible spectra.
   - $\Rightarrow$ $P(C_i \,|\, I) = k$ is a well precise spectrum (in most cases far from the physical one)
   - $\Rightarrow$ VERY STRONG prior that biases the result!

The rescue trick

Instead of using the original probability inversion (applied directly) to spectra

$$P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, \Lambda, I) \;\; \propto \;\; P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C, \Lambda, I) \cdot P(\boldsymbol{x}_C \,|\, I)\,,$$

we restart from

$$P(C_i \,|E_j, I) \;\; \propto \;\; P(E_j \,|\, C_i, I) \cdot P(C_i \,|\, I).$$

Consequences:

1. the sharing of observed events among the cause bins needs to be performed 'by hand';
2. a uniform prior $P(C_i \,|\, I) = k$ does not mean indifference over all possible spectra.
   - $\Rightarrow$ $P(C_i \,|\, I) = k$ is a well precise spectrum (in most cases far from the physical one)
   - $\Rightarrow$ VERY STRONG prior that biases the result! $\rightarrow$ iterations

Old $\rightarrow$ improved algorithm (technical issues for self study)

1. [$*$] λ_{ij} estimated by MC simulation as

$$\lambda_{ji} \approx x(E_j)^{MC}/x(C_i)^{MC}\,;$$

Old $\rightarrow$ improved algorithm (technical issues for self study)

1. [$*$] λ_{ij} estimated by MC simulation as

$$\lambda_{ji} \approx x(E_j)^{MC}/x(C_i)^{MC} ;$$

2. $P(C_i \,|\, E_j, I)$ from Bayes theorem; $[\theta_{ij} \equiv P(C_i \,|\, E_j, I)]$

$$P(C_i \,|\, E_j, I) = \frac{P(E_j \,|\, C_i, I) \cdot P(C_i \,|\, I)}{\sum_i P(E_j \,|\, C_i, I) \cdot P(C_i \,|\, I)} ,$$

or

$$\theta_{ij} = \frac{\lambda_{ji} \cdot P(C_i \,|\, I)}{\sum_i \lambda_{ji} \cdot P(C_i \,|\, I)} ,$$

Old $\to$ improved algorithm (technical issues for self study)

1. [$*$] λ_{ij} estimated by MC simulation as
$$\lambda_{ji} \approx x(E_j)^{MC}/x(C_i)^{MC} \,;$$
2. $P(C_i\,|E_j,\,I)$ from Bayes theorem; $\quad [\theta_{ij} \equiv P(C_i\,|E_j,\,I)]$
3. [$*$] Assignement of events to cause bins:
$$\begin{aligned}
x(C_i)|_{x(E_j)} &\approx P(C_i\,|E_j,\,I)\cdot x(E_j) \\
x(C_i)|_{\boldsymbol{x}_E} &\approx \sum_{j=1}^{n_E} P(C_i\,|E_j,\,I)\cdot x(E_j) \\
x(C_i) &\approx \frac{1}{\epsilon_i}\, x(C_i)|_{\boldsymbol{x}_E} \,,
\end{aligned}$$

with $\epsilon_i = \sum_{j=1}^{n_E} P(E_j\,|\,C_i,\,I)$

Old $\rightarrow$ improved algorithm (technical issues for self study)

1. [$*$] λ_{ij} estimated by MC simulation as

$$\lambda_{ji} \approx x(E_j)^{MC}/x(C_i)^{MC}\,;$$

2. $P(C_i\,|\,E_j,\,I)$ from Bayes theorem; $\quad[\theta_{ij} \equiv P(C_i\,|\,E_j,\,I)]$
3. [$*$] Assignement of events to cause bins:

$$\begin{aligned} x(C_i)|_{x(E_j)} &\approx P(C_i\,|\,E_j,\,I)\cdot x(E_j) \\ x(C_i)|_{\boldsymbol{x}_E} &\approx \sum_{j=1}^{n_E} P(C_i\,|\,E_j,\,I)\cdot x(E_j) \\ x(C_i) &\approx \frac{1}{\epsilon_i}\, x(C_i)|_{\boldsymbol{x}_E}\,, \end{aligned}$$

with $\epsilon_i = \sum_{j=1}^{n_E} P(E_j\,|\,C_i,\,I)$

4. [$*$] Uncertainty by ‘standard error propagation’

Improvements

1. $\boldsymbol{\lambda}_i$: having each element λ_{ji} the meaning of "p_j" of a Multinomial distribution, their distribution can easily (and conveniently and realistically) modelled by a Dirichlet:

$$\boldsymbol{\lambda}_i \;\sim\; \text{Dir}[\boldsymbol{\alpha}_{prior} + \boldsymbol{x}_E^{MC}\Big|_{x(C_i)^{MC}}]\,,$$

(The Dirichlet is the prior conjugate of the Multinomial)

Improvements

1. $\boldsymbol{\lambda}_i$:

$$\boldsymbol{\lambda}_i \;\sim\; \mathrm{Dir}[\boldsymbol{\alpha}_{prior} + \boldsymbol{x}_E^{MC}\Big|_{x(C_i)^{MC}}]\,,$$

2. uncertainty on $\boldsymbol{\lambda}_i$:
   taken into account by sampling $\Rightarrow$ equivalent to integration

$$\Rightarrow P(\boldsymbol{x}_C\,|\,\boldsymbol{x}_E, I) \;=\; \int P(\boldsymbol{x}_C\,|\,\boldsymbol{x}_E, \Lambda,\, I)\, f(\Lambda\,|\,I)\, \mathrm{d}\Lambda$$

Improvements

1. $\boldsymbol{\lambda}_i$:

$$\boldsymbol{\lambda}_i \;\sim\; \mathrm{Dir}[\boldsymbol{\alpha}_{prior} + \boldsymbol{x}_E^{MC}\Big|_{x(C_i)^{MC}}]\,,$$

2. uncertainty on $\boldsymbol{\lambda}_i$:
   taken into account by sampling
3. sharing $x_{E_j} \rightarrow \boldsymbol{x}_C$: done by a Multinomial:

$$\boldsymbol{x}_C\big|_{x(E_j)} \;\sim\; \mathrm{Mult}[x(E_j),\, \boldsymbol{\theta}_j]\,,$$

Improvements

1. $\boldsymbol{\lambda}_i$:

$$\boldsymbol{\lambda}_i \;\sim\; \mathrm{Dir}[\boldsymbol{\alpha}_{prior} + \boldsymbol{x}_E^{MC}\Big|_{x(C_i)^{MC}}]\,,$$

2. uncertainty on $\boldsymbol{\lambda}_i$:
   taken into account by sampling
3. sharing $x_{E_j} \to \boldsymbol{x}_C$: done by a Multinomial:

$$\boldsymbol{x}_C\big|_{x(E_j)} \;\sim\; \mathrm{Mult}[x(E_j),\, \boldsymbol{\theta}_j]\,,$$

4. $x(E_j) \to \mu_j$: what needs to be shared is not the observed number $x(E_j)$, but rather the estimated true value μ_j: remember $x(E_j) \sim \mathrm{Poisson}[\mu_j]$

$$\mu_j \;\sim\; \mathrm{Gamma}[c_j + x(E_j),\, r_j + 1]\,,$$

(Gamma is prior conjugate of Poisson)

Improvements

1. $\boldsymbol{\lambda}_i$:

$$\boldsymbol{\lambda}_i \;\sim\; \mathrm{Dir}[\boldsymbol{\alpha}_{prior} + \boldsymbol{x}_E^{MC}\Big|_{x(C_i)^{MC}}]\,,$$

2. uncertainty on $\boldsymbol{\lambda}_i$:
   taken into account by sampling
3. sharing $x_{E_j} \to \boldsymbol{x}_C$: done by a Multinomial:

$$\boldsymbol{x}_C\big|_{x(E_j)} \;\sim\; \mathrm{Mult}[x(E_j),\, \boldsymbol{\theta}_j]\,,$$

4. $x(E_j) \to \mu_j$:

$$\mu_j \;\sim\; \mathrm{Gamma}[c_j + x(E_j),\, r_j + 1]\,,$$

   BUT μ_i is real, while the the number of event parameter of a multinomial must be integer $\Rightarrow$ solved with interpolation
5. uncertainty on μ_i: taken into account by sampling

Improvements

1. $\boldsymbol{\lambda}_i$:
$$\boldsymbol{\lambda}_i \;\sim\; \text{Dir}[\boldsymbol{\alpha}_{prior} + \boldsymbol{x}_E^{MC}\Big|_{x(C_i)^{MC}}]\,,$$

2. uncertainty on $\boldsymbol{\lambda}_i$:
taken into account by sampling

3. sharing $x_{E_j} \rightarrow \boldsymbol{x}_C$: done by a Multinomial:
$$\boldsymbol{x}_C\big|_{x(E_j)} \;\sim\; \text{Mult}[x(E_j),\, \boldsymbol{\theta}_j]\,,$$

4. $x(E_j) \rightarrow \mu_j$:
$$\mu_j \;\sim\; \text{Gamma}[c_j + x(E_j),\, r_j + 1]\,,$$

5. uncertainty on μ_i: taken into account by sampling

Iteration and (intermediate) smoothing

$$\boxed{\begin{array}{l}\text{instead of using a flat prior over the possible spectra}\\ \text{we are using a particular (flat) spectrum as prior}\end{array}}$$

Iteration and (intermediate) smoothing

instead of using a flat prior over the possible spectra
we are using a particular (flat) spectrum as prior

$\Rightarrow$ the posterior [i.e. the ensemble of $\boldsymbol{x}_C^{(t)}$ obtained by sampling] is affected by this quite strong assumption, that seldom holds in real cases.

Iteration and (intermediate) smoothing

> instead of using a flat prior over the possible spectra
> we are using a particular (flat) spectrum as prior

$\Rightarrow$ the posterior [i.e. the ensemble of $\boldsymbol{x}_C^{(t)}$ obtained by sampling] is affected by this quite strong assumption, that seldom holds in real cases.

$\Rightarrow$ problem worked around by ITERATIONS

$\Rightarrow$ posterior becomes prior of next iteration

Iteration and (intermediate) smoothing

> instead of using a flat prior over the possible spectra
> we are using a particular (flat) spectrum as prior

$\Rightarrow$ the posterior [i.e. the ensemble of $\boldsymbol{x}_C^{(t)}$ obtained by sampling] is affected by this quite strong assumption, that seldom holds in real cases.

$\Rightarrow$ problem worked around by ITERATIONS

$\Rightarrow$ posterior becomes prior of next iteration

$\Rightarrow$ Usque tandem?

Iteration and (intermediate) smoothing

instead of using a flat prior over the possible spectra
we are using a particular (flat) spectrum as prior

$\Rightarrow$ the posterior [i.e. the ensemble of $\boldsymbol{x}_C^{(t)}$ obtained by sampling] is affected by this quite strong assumption, that seldom holds in real cases.

$\Rightarrow$ problem worked around by ITERATIONS

$\Rightarrow$ posterior becomes prior of next iteration

$\Rightarrow$ Usque tandem?

- Empirical approach (with help of simulation):
  - 'True spectrum' recovered in a couple of steps
  - Then the solution starts to diverge towards a wildy oscillating spectrum (any unavoidable fluctuation is believed more and more. . .)

    $\Rightarrow$ find empirically an optimum

Iteration and (intermediate) smoothing

instead of using a flat prior over the possible spectra we are using a particular (flat) spectrum as prior

$\Rightarrow$ the posterior [i.e. the ensemble of $\boldsymbol{x}_C^{(t)}$ obtained by sampling] is affected by this quite strong assumption, that seldom holds in real cases.

$\Rightarrow$ problem worked around by ITERATIONS

$\Rightarrow$ posterior becomes prior of next iteration

$\Rightarrow$ Usque tandem?

- **regularization** (a subject by itself)
  my preferred approach
  - regularize the posterior before using as next prior

Iteration and (intermediate) smoothing

> instead of using a flat prior over the possible spectra
> we are using a particular (flat) spectrum as prior

$\Rightarrow$ the posterior [i.e. the ensemble of $\boldsymbol{x}_C^{(t)}$ obtained by sampling] is affected by this quite strong assumption, that seldom holds in real cases.

$\Rightarrow$ problem worked around by ITERATIONS

$\Rightarrow$ posterior becomes prior of next iteration

$\Rightarrow$ Usque tandem?

- **regularization** (a subject by itself)
  my preferred approach
  - regularize the posterior before using as next prior
  - intermediate smoothing $\Rightarrow$ we belief physics is 'smooth'
  - ...but 'irregularities' of the data are not washed out

    ($\Rightarrow$ unfolding Vs parametric inference)

Iteration and (intermediate) smoothing

instead of using a flat prior over the possible spectra
we are using a particular (flat) spectrum as prior

$\Rightarrow$ the posterior [i.e. the ensemble of $\boldsymbol{x}_C^{(t)}$ obtained by sampling] is affected by this quite strong assumption, that seldom holds in real cases.

$\Rightarrow$ problem worked around by ITERATIONS

$\Rightarrow$ posterior becomes prior of next iteration

$\Rightarrow$ Usque tandem?

- **regularization** (a subject by itself)
  my preferred approach
  - regularize the posterior before using as next prior

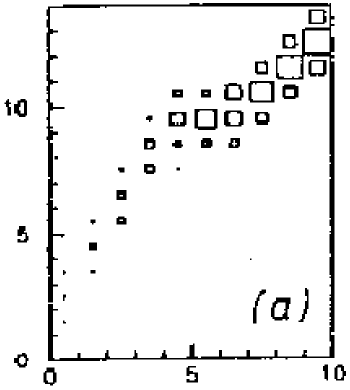
  $\Rightarrow$ Good compromize and good results

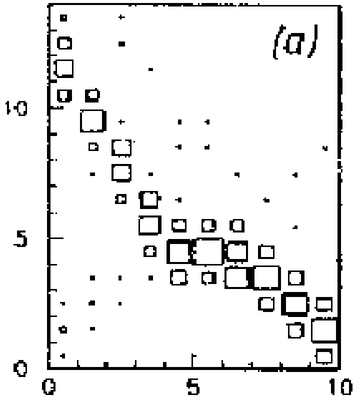
  $\Rightarrow$ Very 'Bayesian'

  $\Rightarrow$ No oscillations for $n_{steps} \to \infty$

Examples

Smearing matrices (from 1995 NIM paper)

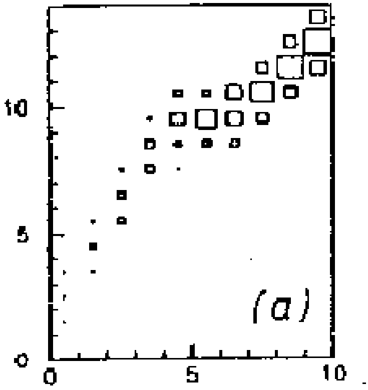


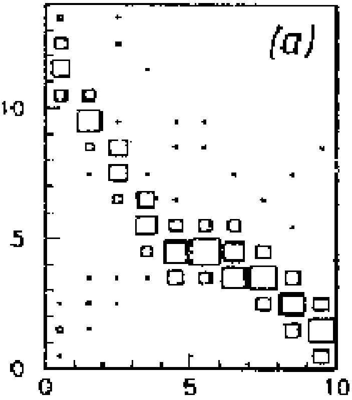


quite bad! (real cases are usually more gentle)

Examples

Smearing matrices (from 1995 NIM paper)





quite bad! (real cases are usually more gentle)

⇒ watch DEMO

It is possible to do better?

- The 2010 paper (arXiv:1010.0632) essentially improves the evaluation of uncertainties;
- paper recommended because it (hopefully) clarifies several issues.

It is possible to do better?

- The 2010 paper (arXiv:1010.0632) essentially improves the evaluation of uncertainties;
- paper recommended because it (hopefully) clarifies several issues.

In reality it is possible nowday to follow the more rigorous approach abandoned ≈ 18 years ago (indeed the 2010 paper was written almost completely in 2006) using

- Bayesian networks;

It is possible to do better?

- The 2010 paper (arXiv:1010.0632) essentially improves the evaluation of uncertainties;
- paper recommended because it (hopefully) clarifies several issues.

In reality it is possible nowday to follow the more rigorous approach abandoned ≈ 18 years ago (indeed the 2010 paper was written almost completely in 2006) using

- Bayesian networks;
- MCMC for integration;

It is possible to do better?

- The 2010 paper (arXiv:1010.0632) essentially improves the evaluation of uncertainties;
- paper recommended because it (hopefully) clarifies several issues.

In reality it is possible nowday to follow the more rigorous approach abandoned $\approx$18 years ago (indeed the 2010 paper was written almost completely in 2006) using

- Bayesian networks;
- MCMC for integration;
- JAGS code to perform the practical work

It is possible to do better?

Draft 16 October 2012

It is possible to do better?

Draft 16 October 2012

Straight Bayesian unfolding performed by MCMC techniques using JAGS/rjags

G. D'Agostini[1] and E. Franco[2]

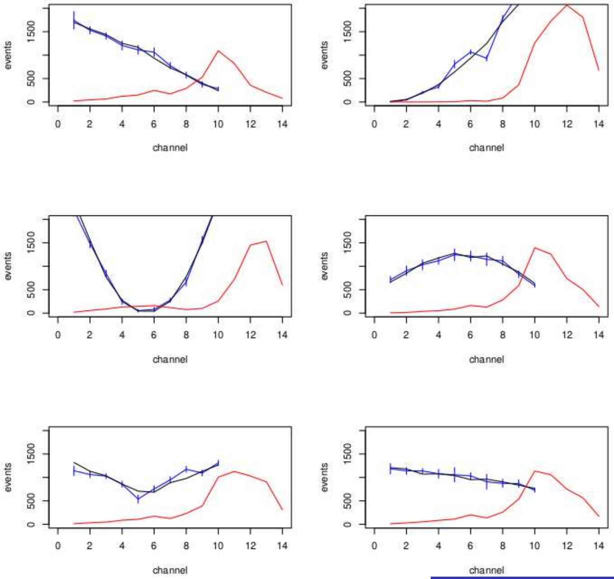
[1] Università "La Sapienza" and INFN, Rome, Italy
(giulio.dagostini@roma1.infn.it, http://www.roma1.infn.it/~dagos)
[2] INFN, Sezione di Roma 1, Rome, Italy
(enrico.franco@roma1.infn.it)

Abstract

The observation that the probability 'vector' of the *observed bins* can be expressed as product of the *smearing matrix* and the probability 'vector' of the *cause bins* avoids to formulate the discretized unfolding problem in terms of sum of multinomial distributions. As a result, the joint distribution of all uncertain quantities entering the problem can be factorized by as a product of conditional probabilities, that can be viewed as a *Probabilistic* ('*Bayesian*') *Graphical Model* and solved using modern *Markov Chain Monte Carlo* methods, finally avoiding the 'dirty trick' of the iterations. In particular, we have implemented these ideas in the *JAGS* program, run under *R* using *rjags*

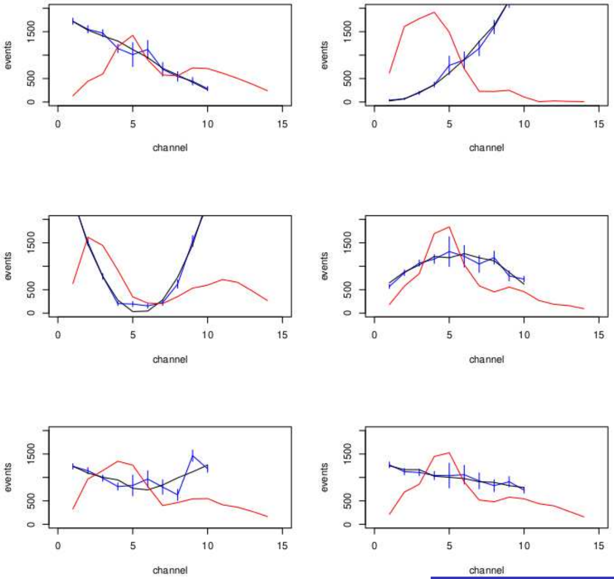
It is possible to do better?

Results on the same toy models (Oct 2012) — No iterations!



It is possible to do better?

Results on the same toy models (Oct 2012) — No iterations!



Stay tuned!

The End

For references, also related to the discussion:

- Lectures to Phd students (lickable link)

Linked notes

Teaching Laplace's *'definition'* of probability

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

Teaching Laplace's *'definition'* of probability

$$p \;=\; \frac{\#\,\text{favorable cases}}{\#\,\text{possible equiprobable cases}} \quad \Rightarrow \textbf{Loop!}$$

Teaching Laplace's *'definition'* of probability

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

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

Teaching Laplace's *'definition'* of probability

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

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

The formula cannot define probability,
but it can be used to evaluate its value p
if we already know what we are talking about

Teaching Laplace's *'definition'* of probability

Cheating students...

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

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

The formula <u>cannot</u> define probability,
but it can be used to evaluate its value p
<u>if we already know what we are talking about</u>

- *Replacing 'equi-probable' by 'equi-possible' is just cheating students!*

Teaching Laplace's *'definition'* of probability

Cheating students...

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

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

The formula <u>cannot</u> define probability,
but it can be used to evaluate its value p
<u>if we already know what we are talking about</u>

- *Replacing 'equi-probable' by 'equi-possible' is just cheating students!*
- *I proposed quickly some simple problems in order to disctract them*

Teaching Laplace's *'definition'* of probability

Cheating students...

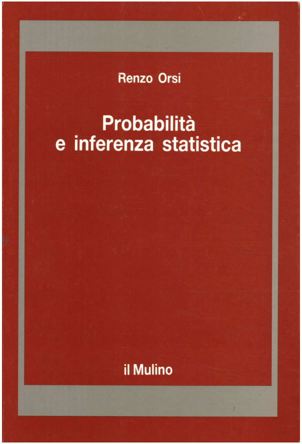
$$p = \frac{\#\,\text{favorable cases}}{\#\,\text{possible equally possible cases}}$$

Note: "*... lorsque rien ne porte à croire que l'un de ces cas doit arriver plutot que les autres*" (Laplace)
The formula cannot define probability,
but it can be used to evaluate its value p
if we already know what we are talking about

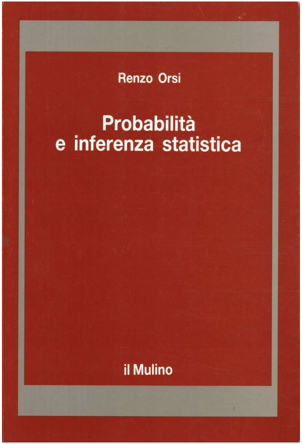
- *Replacing 'equi-probable' by 'equi-possible' is just cheating students!*
- *I proposed quickly some simple problems in order to disctract them*

But for me it was a serious shock that
induced me to rithink the probabilistic issues

Go back

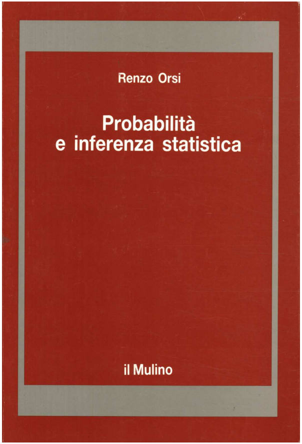


A comprehensive standard 'classical' book, but at least it mentioned Bayes' theorem, providing a couple of examples, of which I read the first one (p. 47):



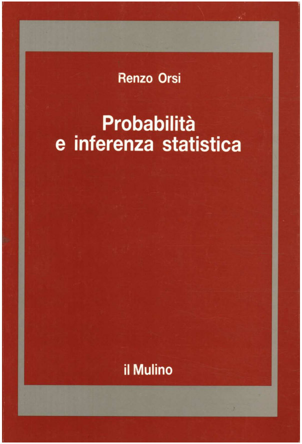
A comprehensive standard 'classical' book, but at least it mentioned Bayes' theorem, providing a couple of examples, of which I read the first one (p. 47):

- Machines M_1, M_2 and M_3 share the producion of the <u>same</u> object, with respectively 30%, 25% and 45% shares,



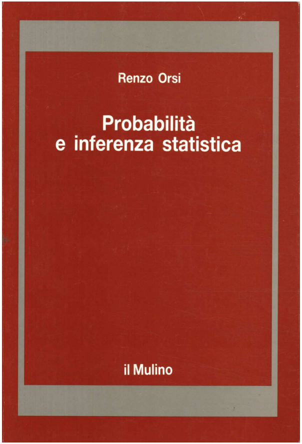
A comprehensive standard 'classical' book, but at least it mentioned Bayes' theorem, providing a couple of examples, of which I read the first one (p. 47):

- Machines M_1, M_2 and M_3 share the producion of the same object, with respectively 30%, 25% and 45% shares, but with faulty rates of 1%, 1.2% and 2%, respectively.



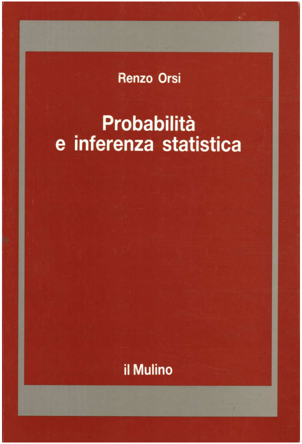
A comprehensive standard ‘classical’ book, but at least it mentioned Bayes’ theorem, providing a couple of examples, of which I read the first one (p. 47):

- Machines M_1, M_2 and M_3 share the producion of the same object, with respectively 30%, 25% and 45% shares, but with faulty rates of 1%, 1.2% and 2%, respectively. The objects are then collected together.



A comprehensive standard 'classical' book, but at least it mentioned Bayes' theorem, providing a couple of examples, of which I read the first one (p. 47):

- Machines M_1, M_2 and M_3 share the producion of the same object, with respectively 30%, 25% and 45% shares, but with faulty rates of 1%, 1.2% and 2%, respectively. The objects are then collected together.
- Having in hand a faulty item, evaluate the probability that it was produced by each of the three machines:



A comprehensive standard 'classical' book, but at least it mentioned Bayes' theorem, providing a couple of examples, of which I read the first one (p. 47):

- Machines M_1, M_2 and M_3 share the producion of the same object, with respectively 30%, 25% and 45% shares, but with faulty rates of 1%, 1.2% and 2%, respectively. The objects are then collected together.
- Having in hand a faulty item, evaluate the probability that it was produced by each of the three machines: $\rightarrow P(M_i \,|\, \text{faulty item})$

Several years of study in order to clarify some ideas, reaching the following conclusions:

Several years of study in order to clarify some ideas, reaching the following conclusions:

- ‘objective’, abstract mathematical priors often lead to nonsense;

Several years of study in order to clarify some ideas, reaching the following conclusions:

- ‘objective’, abstract mathematical priors often lead to nonsense;
- “probability is good sense reduced to a calculus” (Laplace);

Several years of study in order to clarify some ideas, reaching the following conclusions:

- 'objective', abstract mathematical priors often lead to nonsense;
- "probability is good sense reduced to a calculus" (Laplace);
- "all models are wrong, but some are useful" (Cox);

Several years of study in order to clarify some ideas, reaching the following conclusions:

- ‘objective’, abstract mathematical priors often lead to nonsense;
- “probability is good sense reduced to a calculus” (Laplace);
- “all models are wrong, but some are useful” (Cox);
- de Finetti’s logic of ‘coherent bet’ (although virtual) helps.

Several years of study in order to clarify some ideas, reaching the following conclusions:

- ‘objective’, abstract mathematical priors often lead to nonsense;
- “probability is good sense reduced to a calculus” (Laplace);
- “all models are wrong, but some are useful” (Cox);
- de Finetti’s logic of ‘coherent bet’ (although virtual) helps.

Some contributions on the subject:

- *Jeffreys Priors versus Experienced Physicist Priors – Arguments against Objective Bayesian Theory*, arXiv:physics/9811045
- *Overcoming prior anxiety*, arXiv:physics/9906048
- *Constraints on the Higgs Boson Mass from Direct Searches and Precision Measurements* [with G. Degrassi], arXiv:hep-ph/9902226
- *Inferring the intensity of Poisson processes at limit of detector sensitivity (with a case study on gravitational wave burst search)* [with P. Astone], arXiv:hep-ex/9909047
- *Bayesian reasoning in data analysis – A critical introduction*, 2003

Go back