Probabilistic Inference and Applications to Frontier Physics – Part 1 –

Giulio D'Agostini

Dipartimento di Fisica

Università di Roma La Sapienza

Preamble

No ‘prescriptions’, but general ideas

Preamble

No ‘prescriptions’, but general ideas

... possibly arising from

’first principles’ (as we physicists like).

Preamble

No ‘prescriptions’, but general ideas

... possibly arising from

’first principles’ (as we physicists like).

$\Rightarrow$ Probabilistic approach

Preamble

No ‘prescriptions’, but general ideas

... possibly arising from

’first principles’ (as we physicists like).

$\Rightarrow$ Probabilistic approach

- Mostly on basic concepts

Preamble

No ‘prescriptions’, but general ideas

... possibly arising from

’first principles’ (as we physicists like).

$\Rightarrow$ Probabilistic approach

- ◦ Mostly on basic concepts
- ◦ Extension to applications

“easy if you try” (at least conceptually)

Preamble

NO Exotic tests

“with russian names”

Preamble

A invitation

to (re-)think

on foundamental aspects

of data analysis.

Outline

- A short introduction from a physicist's point of view.
- Uncertainty, probability, decision.
- Causes $\longleftrightarrow$ Effects

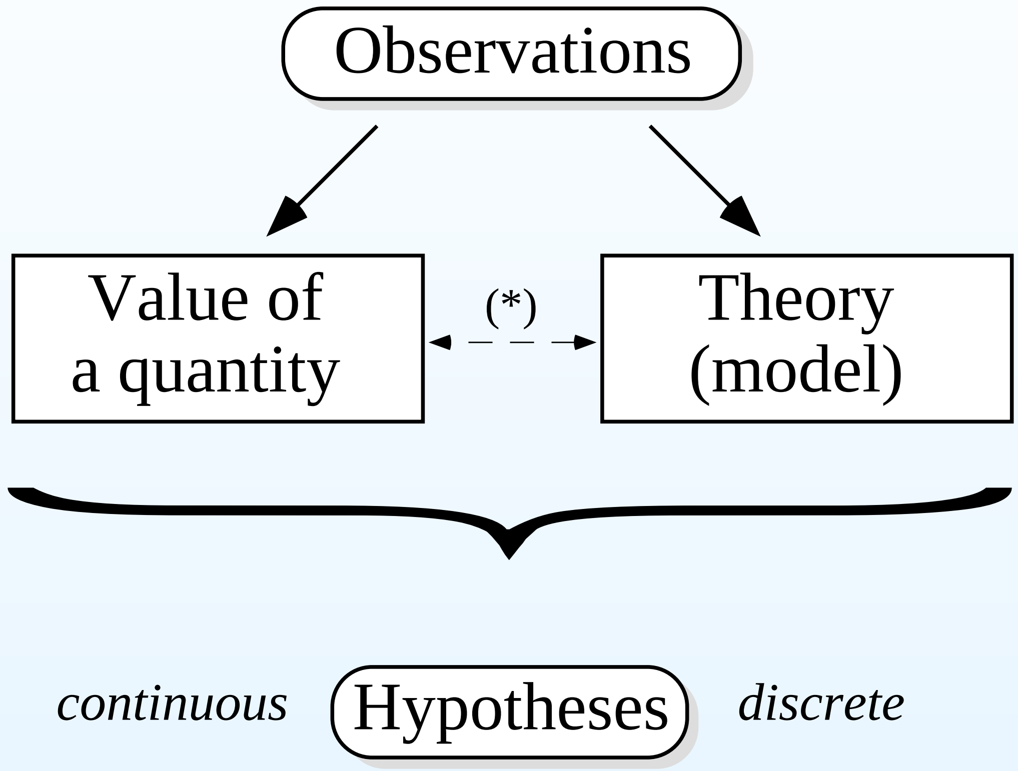
  "The essential problem of the experimental method" (Poincaré).
- A toy model and its physics analogy: the six box game

  "Probability is either referred to real cases or it is nothing" (de Finetti).
- Probabilistic approach, but What is probability?
- Basic rules of probability and Bayes rule.
- Bayesian inference and its graphical representation:
  $\Rightarrow$ Bayesian networks
- Let us play a while with the toy
- Some examples of applications in Physics
- Conclusions

Applications

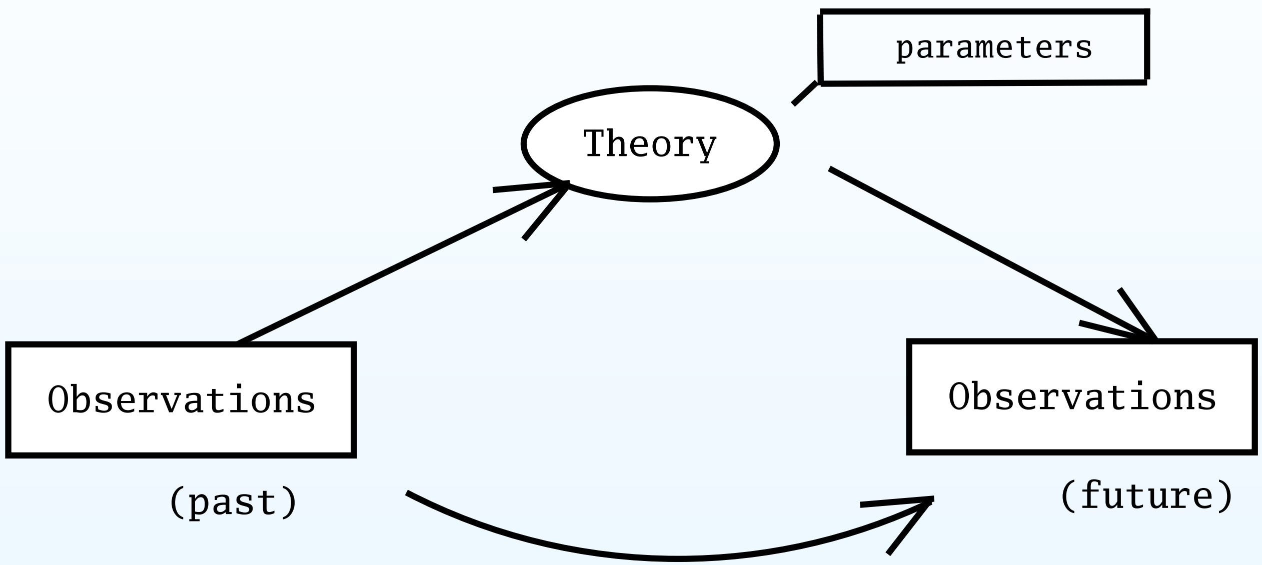
- Inferring a quantity and predicting a future observable
- Fits, including ‘extra variablity’ of data and systematics
- Unfolding
- Setting limits ($\longrightarrow$ understand the role of the likelihood!)

Physics



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

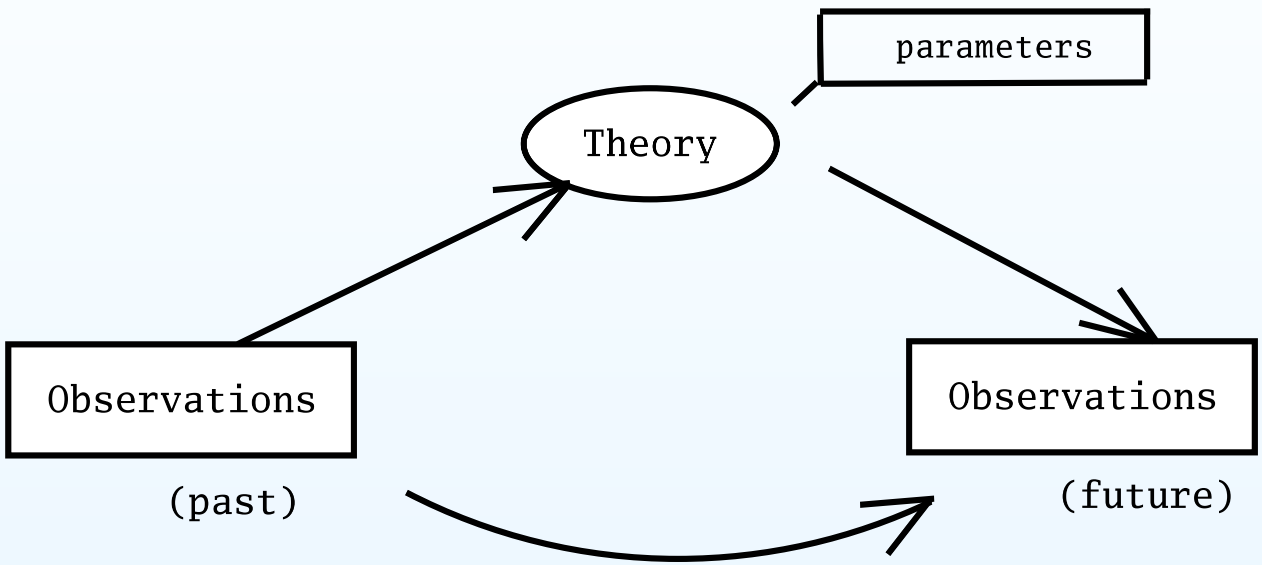
From past to future



Task of the physicist:

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

From past to future



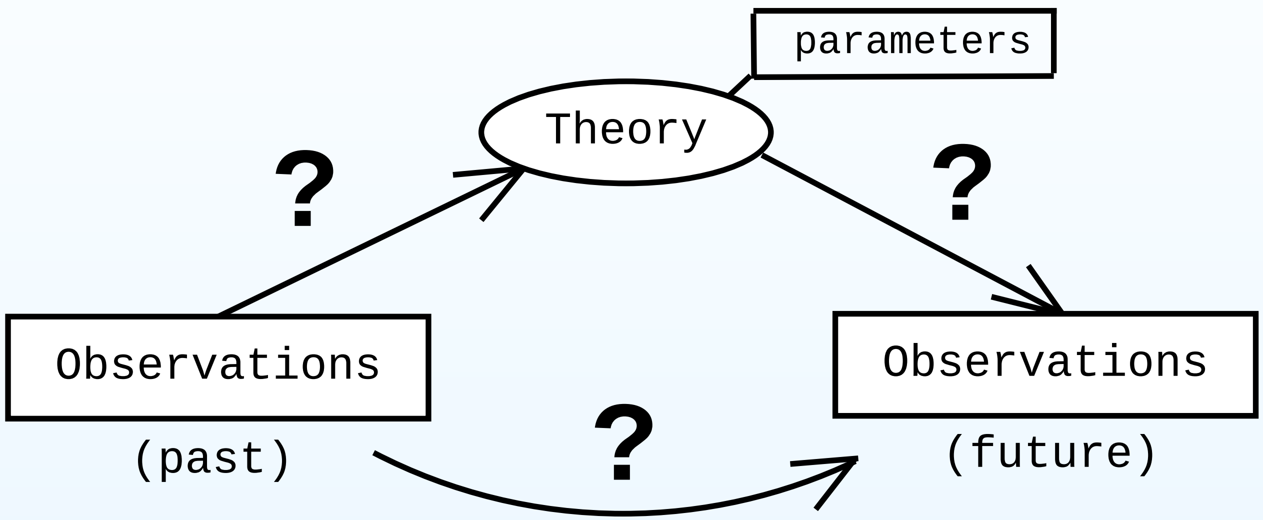
Process

- neither automatic
- nor purely contemplative

  $\longrightarrow$ 'scientific method'

  $\longrightarrow$ planned experiments ('actions') $\Rightarrow$ decision.

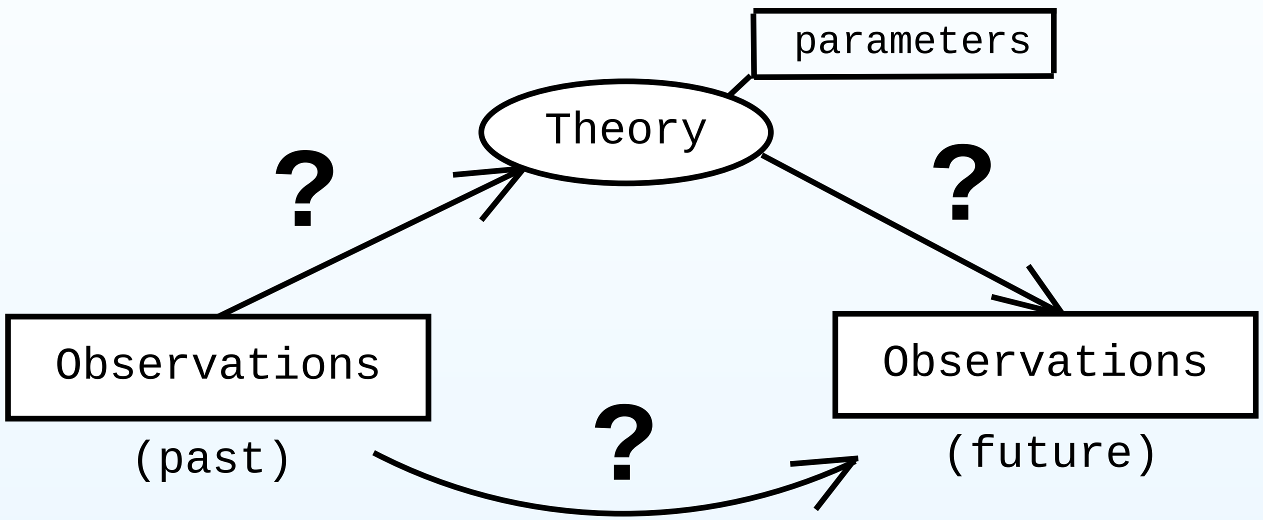
From past to future



⇒ Uncertainty:

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

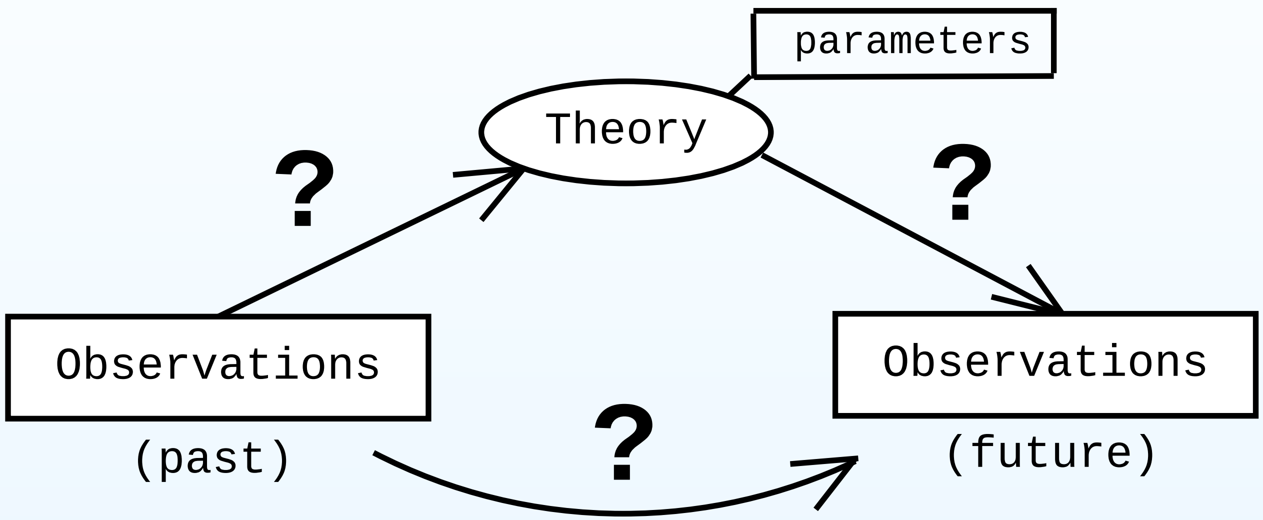
From past to future



⇒ Decision

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

From past to future



Deep reason of uncertainty

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

About predictions

Remember:

“Prediction is very difficult,
especially if it’s about the future” (Bohr)

About predictions

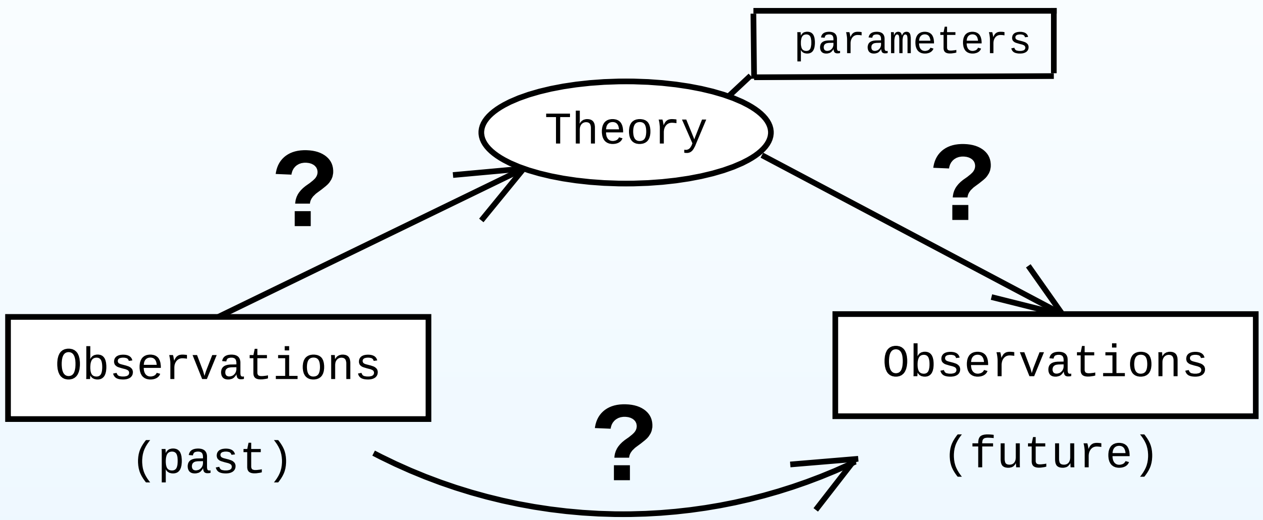
Remember:

"Prediction is very difficult, especially if it's about the future" (Bohr)

But, anyway:

"It is far better to foresee even without certainty than not to foresee at all" (Poincaré)

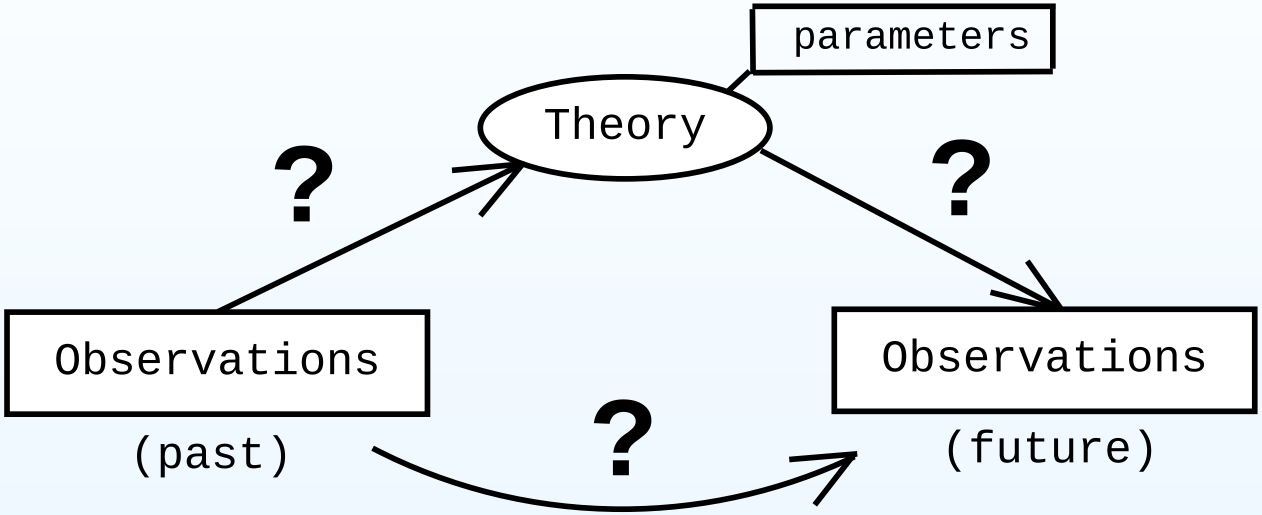
Deep source of uncertainty



Uncertainty:

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

Deep source of uncertainty



Uncertainty:

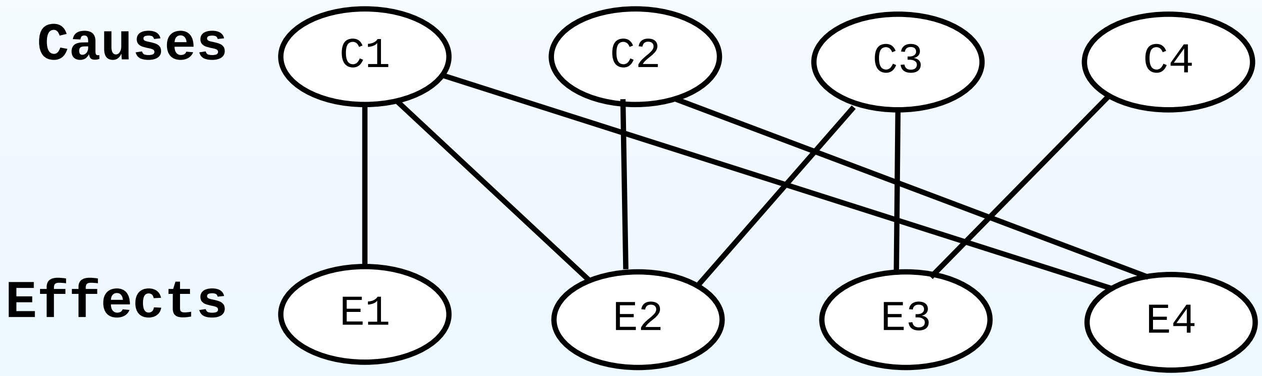
Theory	—?⟶	Future observations
Past observations	—?⟶	Theory
Theory	—?⟶	Future observations

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

CAUSE $\Longleftrightarrow$ EFFECT

Causes $\longrightarrow$ effects

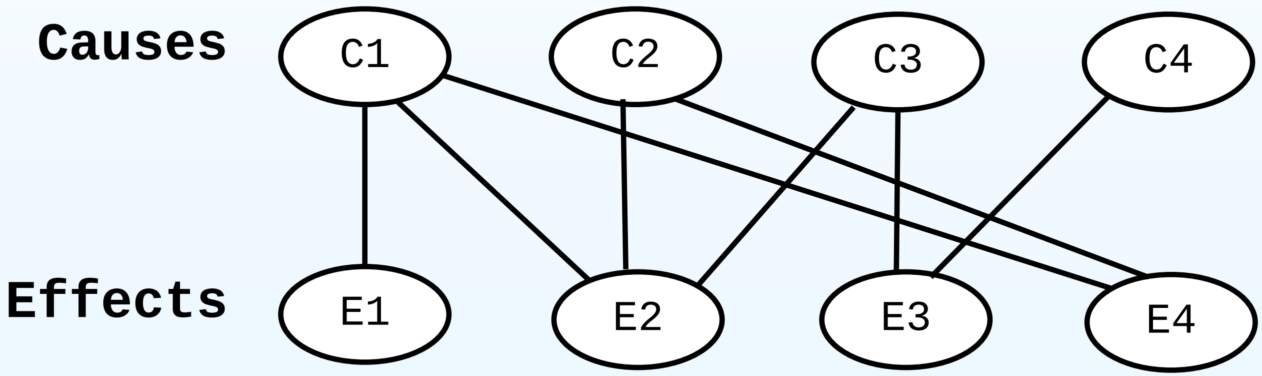
The same *apparent* cause might produce several,different effects



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

Causes $\longrightarrow$ effects

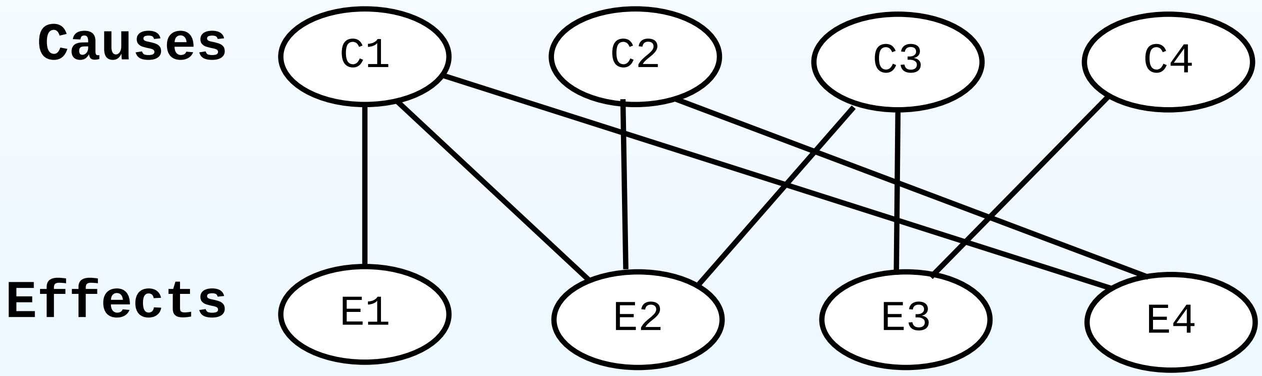
The same *apparent* cause might produce several,different effects



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

Causes $\longrightarrow$ 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 experimental method

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

The essential problem of the experimental method

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

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

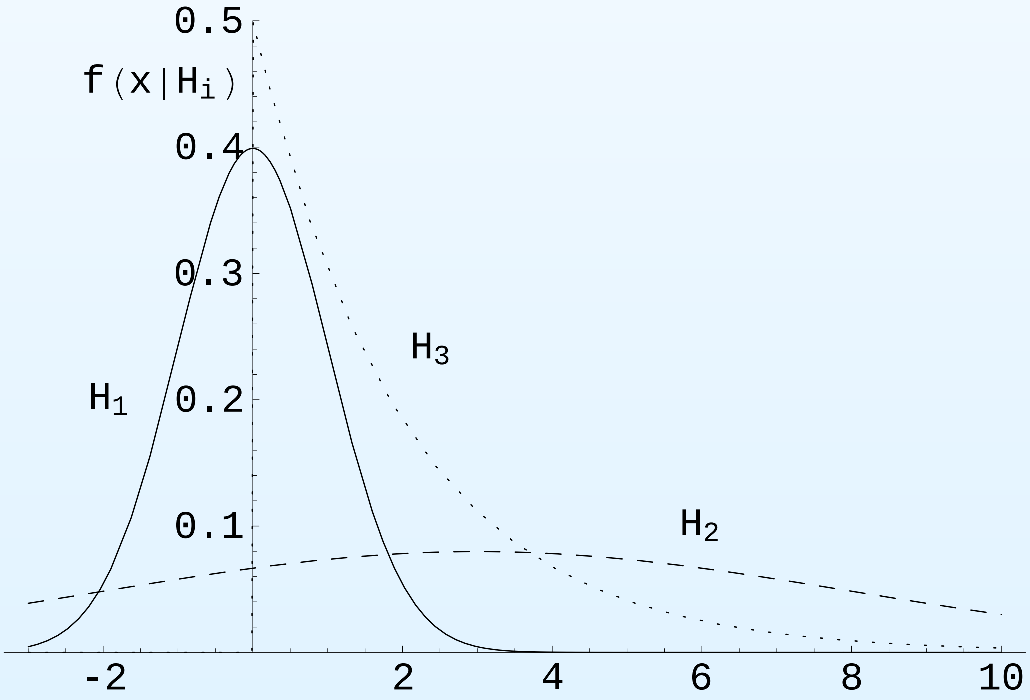
(H. Poincaré – *Science and Hypothesis*)

A numerical example

- Effect: number $x = 3$ extracted 'at random'
- Hypotheses: one of the following random generators:
  - H_1 Gaussian, with $\mu = 0$ and $\sigma = 1$
  - H_2 Gaussian, with $\mu = 3$ and $\sigma = 5$
  - H_3 Exponential, with $\tau = 2$

A numerical example

- Effect: number $x = 3$ extracted 'at random'
- Hypotheses: one of the following random generators:
  - H_1 Gaussian, with $\mu = 0$ and $\sigma = 1$
  - H_2 Gaussian, with $\mu = 3$ and $\sigma = 5$
  - H_3 Exponential, with $\tau = 2$



A numerical example

- Effect: number $x = 3$ extracted ‘at random’
- Hypotheses: one of the following random generators:
  - H_1 Gaussian, with $\mu = 0$ and $\sigma = 1$
  - H_2 Gaussian, with $\mu = 3$ and $\sigma = 5$
  - H_3 Exponential, with $\tau = 2$

$\Rightarrow$ Which one to prefer?

Note: $\Rightarrow$ none of the hypotheses of this example can be excluded and, therefore, there is no way to reach a boolean conclusion. We can only state, somehow, our *rational preference*, based on the experimental result and our best knowledge of the behavior of each *model*.

A numerical example

- Effect: number $x = 3$ extracted 'at random'
- Hypotheses: one of the following random generators:
  - H_1 Gaussian, with $\mu = 0$ and $\sigma = 1$
  - H_2 Gaussian, with $\mu = 3$ and $\sigma = 5$
  - H_3 Exponential, with $\tau = 2$
- we can only state how much *we are* *sure* — or *confident* — on each of them;
- or "we consider each of them more or less *probable* (or *likely*)";
- or "we *believe* each of them more or less than onother one"

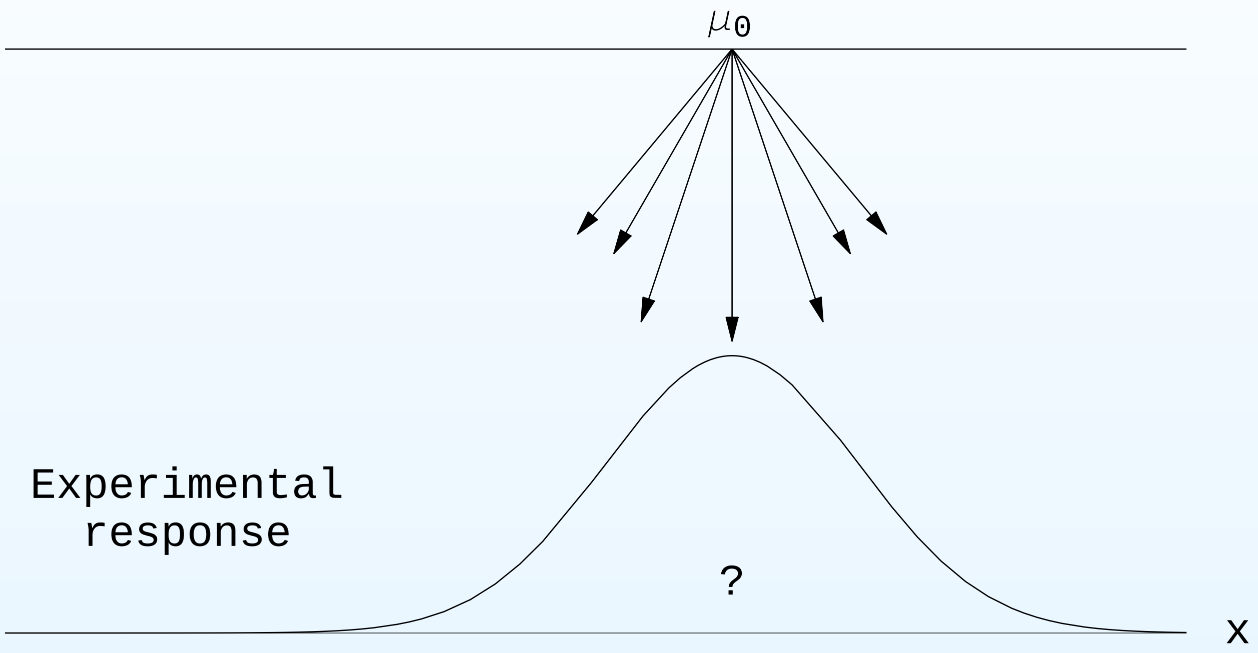
A numerical example

- Effect: number $x = 3$ extracted 'at random'
- Hypotheses: one of the following random generators:
  - H_1 Gaussian, with $\mu = 0$ and $\sigma = 1$
  - H_2 Gaussian, with $\mu = 3$ and $\sigma = 5$
  - H_3 Exponential, with $\tau = 2$
- we can only state how much *we are* *sure* — or *confident* — on each of them;
- or "we consider each of them more or less *probable* (or *likely*)";
- or "we *believe* each of them more or less than onother one"

or similar expressions, all referring to the intuitive concept of

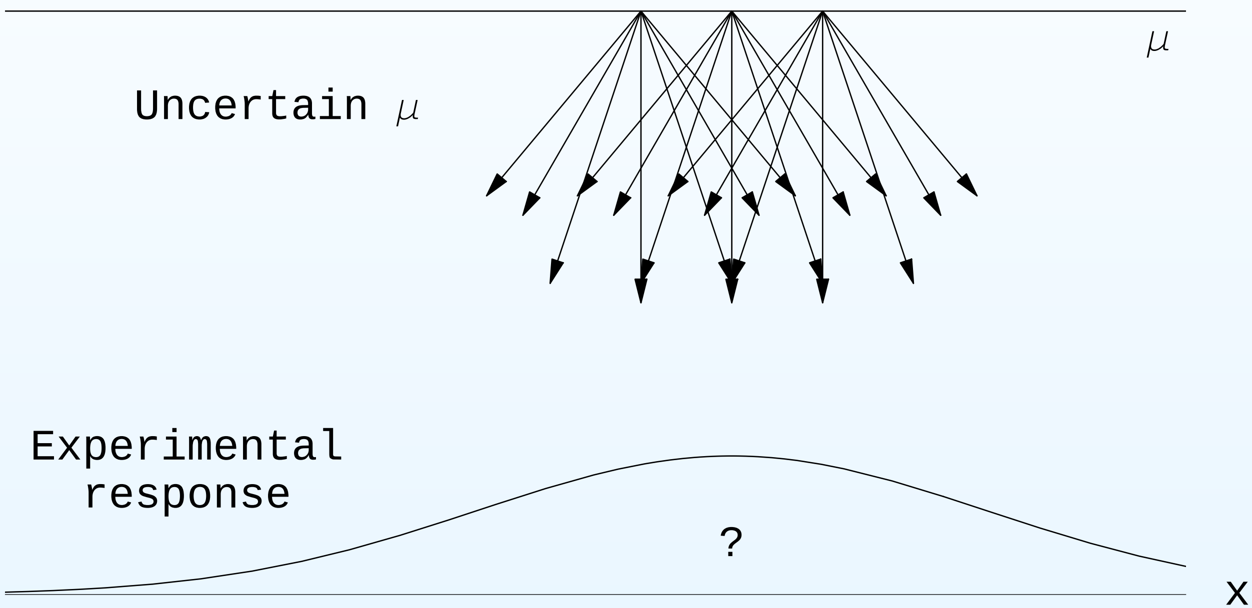
probability.

From ‘true value’ to observations



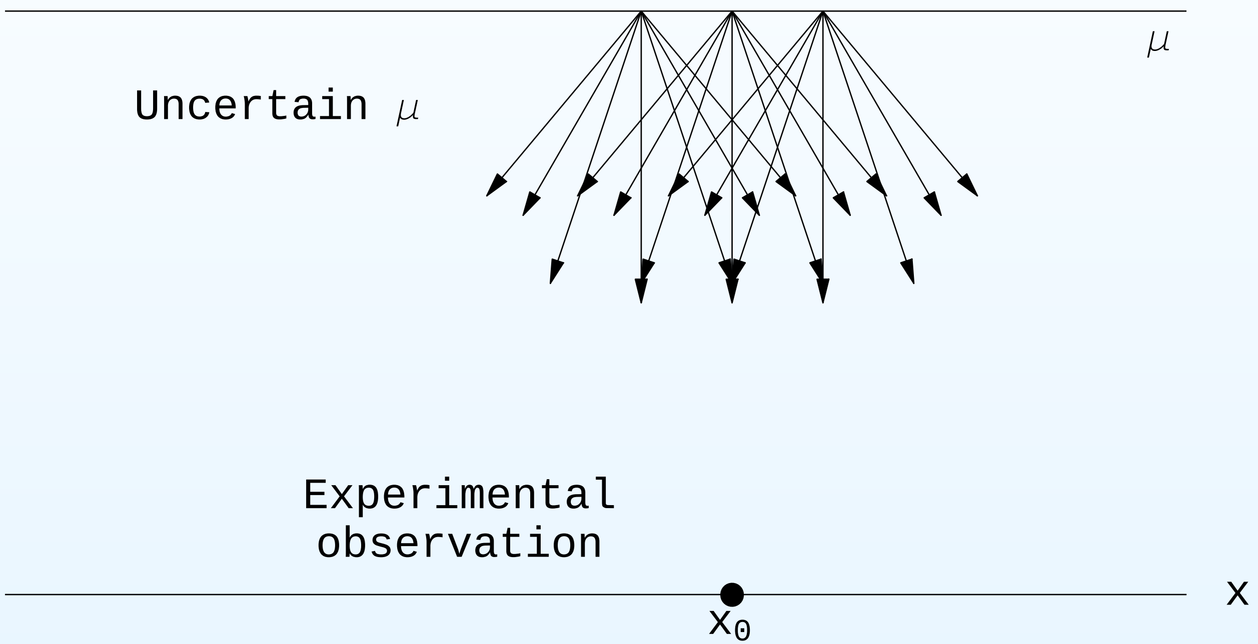
Given μ (exactly known) we are uncertain about x

From ‘true value’ to observations



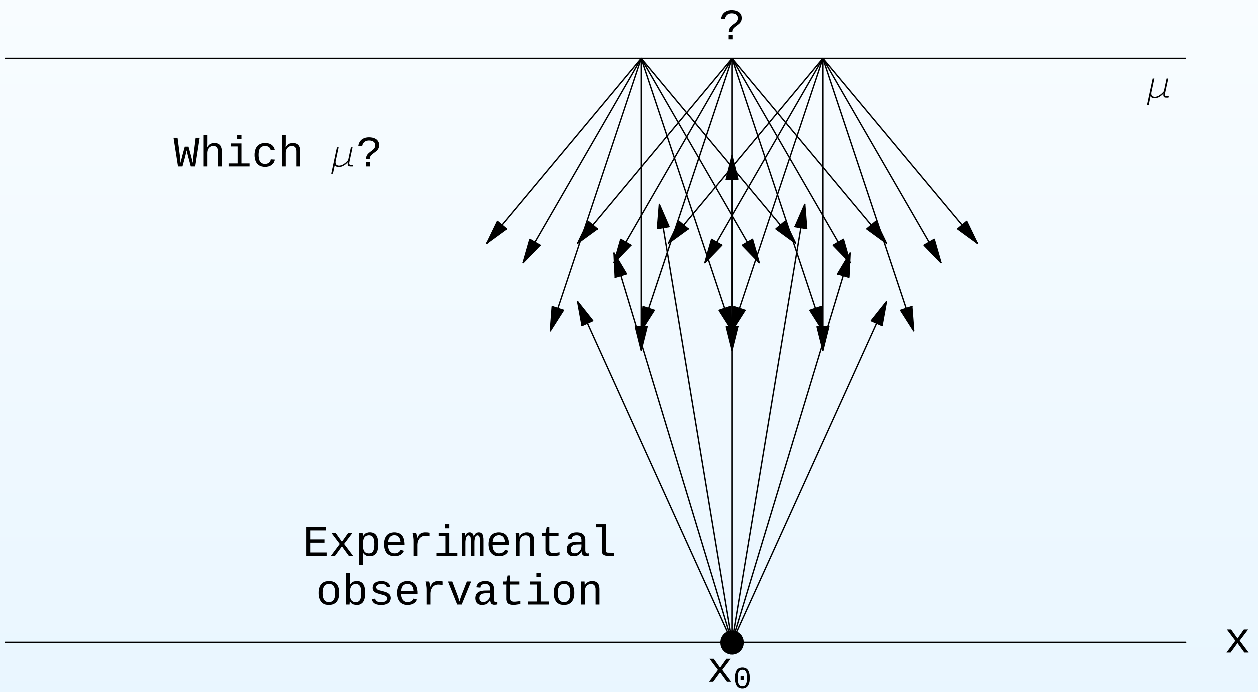
Uncertainty about μ makes us more uncertain about x

Inferring a true value



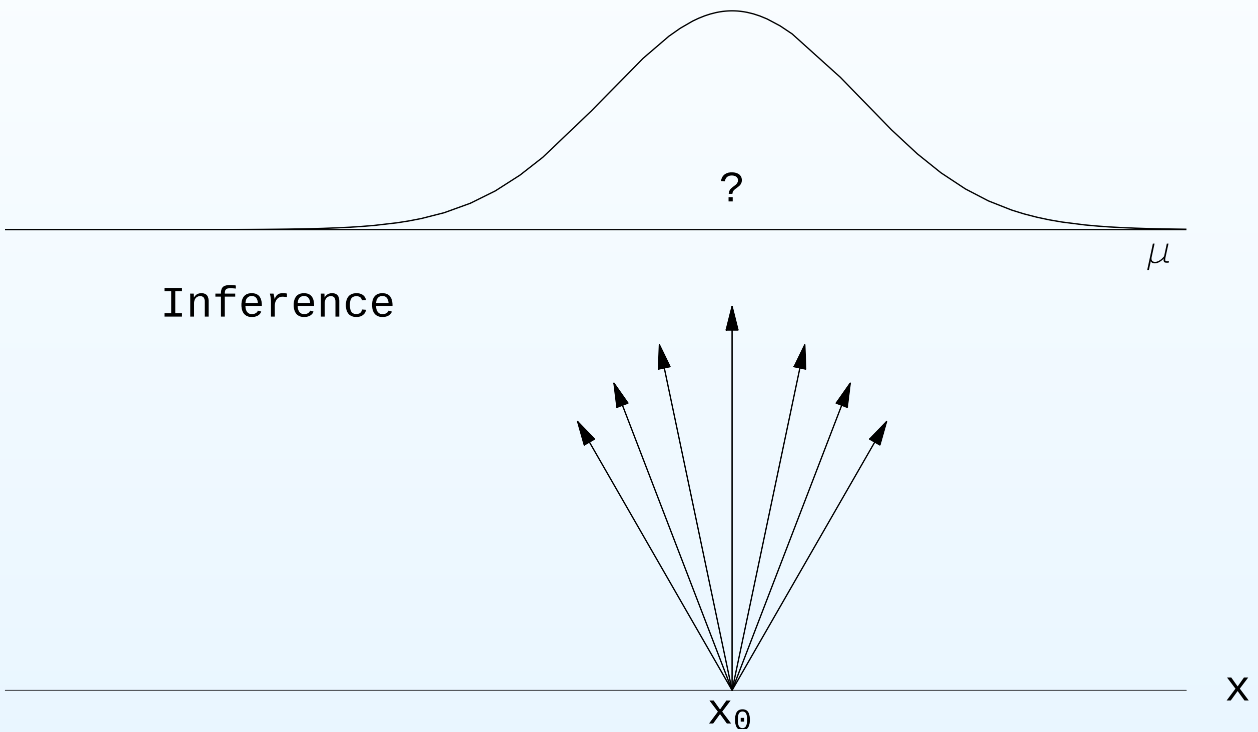
The observed data is <u>certain</u>: $\rightarrow$ 'true value' uncertain.

Inferring a true value



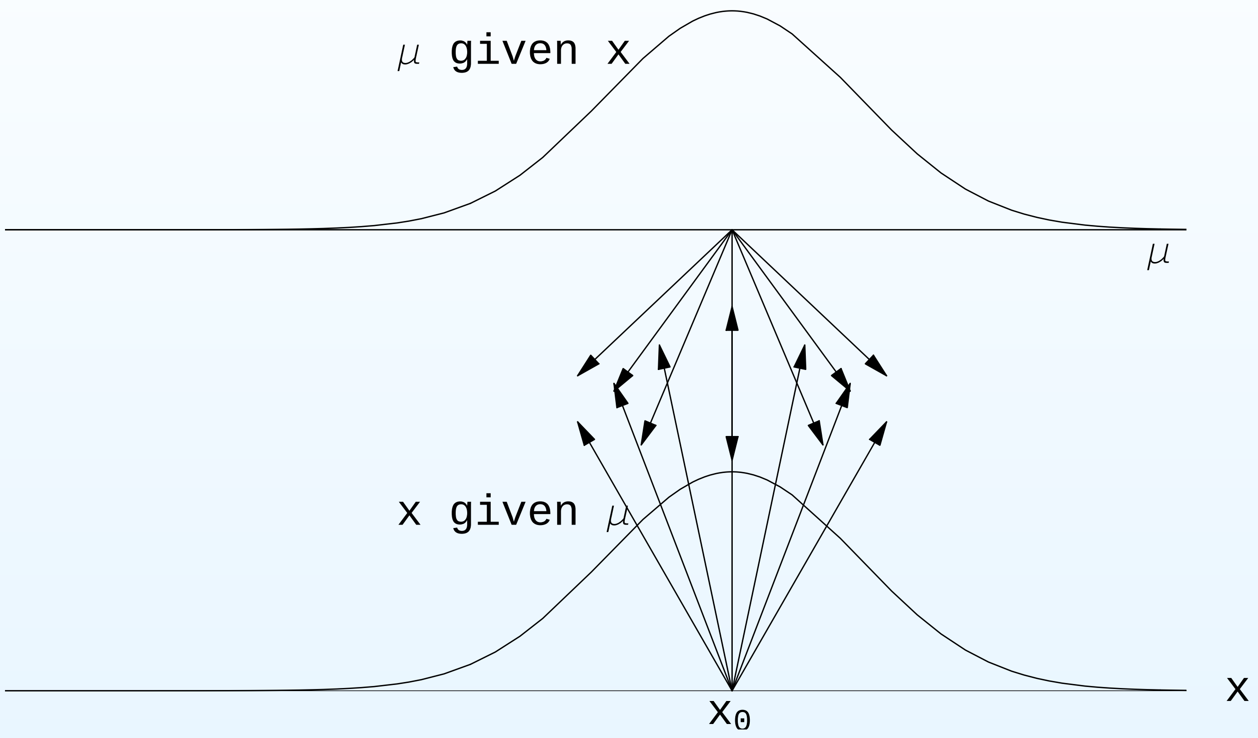
Where does the observed value of x comes from?

Inferring a true value



We are now uncertain about μ, given x.

Inferring a true value



Note the symmetry in reasoning.

Uncertainty and probability

We, as physicists, consider absolutely natural and meaningful statements of the following kind

- $P(-10 < \epsilon'/\epsilon \times 10^4 < 50) >> P(\epsilon'/\epsilon \times 10^4 > 100)$
- $P(170 \le m_{top}/\text{GeV} \le 180) \approx 70\%$
- $P(M_H < 200\,\text{GeV}) > P(M_H > 200\,\text{GeV})$

Uncertainty and probability

We, as physicists, consider absolutely natural and meaningful statements of the following kind

- $P(-10 < \epsilon'/\epsilon \times 10^4 < 50) >> P(\epsilon'/\epsilon \times 10^4 > 100)$
- $P(170 \le m_{top}/\text{GeV} \le 180) \approx 70\%$
- $P(M_H < 200\,\text{GeV}) > P(M_H > 200\,\text{GeV})$

. . . although, such statements are considered
blaspheme to statistics gurus

Uncertainty and probability

We, as physicists, consider absolutely natural and meaningful statements of the following kind

- $P(-10 < \epsilon'/\epsilon \times 10^4 < 50) >> P(\epsilon'/\epsilon \times 10^4 > 100)$
- $P(170 \le m_{top}/\text{GeV} \le 180) \approx 70\%$
- $P(M_H < 200\,\text{GeV}) > P(M_H > 200\,\text{GeV})$

…although, such statements are considered
blaspheme to statistics gurus

[The fact that for several people in this audience this sentence is misterious is a clear indication of the confusion concerning this matter]

Doing Science in conditions of uncertainty

The constant status of uncertainty does not prevent us from doing Science (in the sense of Natural Science and not just Mathematics)

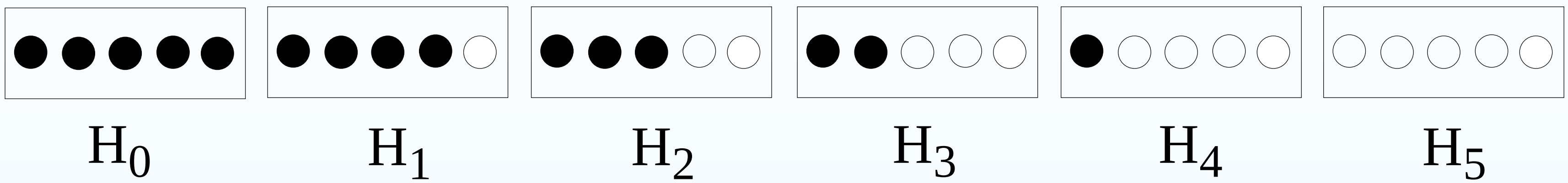
Doing Science in conditions of uncertainty

The constant status of uncertainty does not prevent us from doing Science (in the sense of Natural Science and not just Mathematics)

Indeed

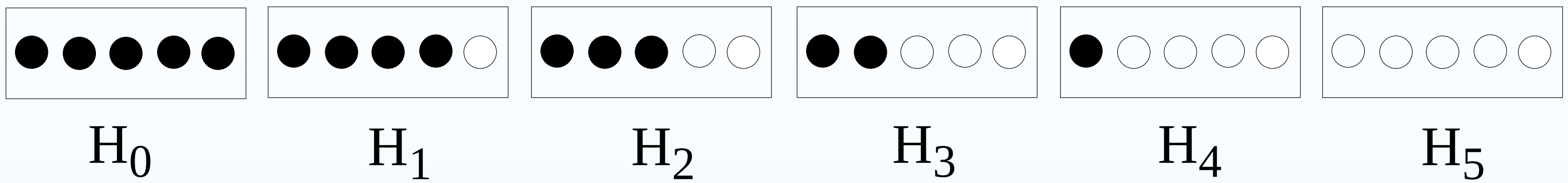
"It is scientific only to say what is more likely and what is less likely" (Feynman)

The six box problem



Let us take randomly one of the boxes.

The six box problem



Let us take randomly one of the boxes.
We are in a state of uncertainty concerning several *events*, the most important of which correspond to the following questions:

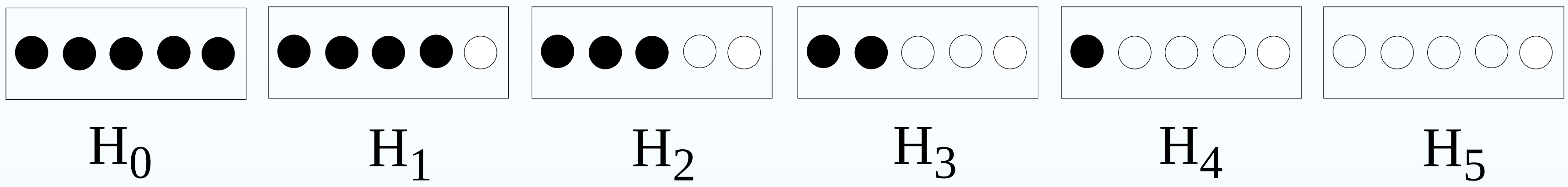
(a) Which box have we chosen, $H_0, H_1, \ldots, H_5$?

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

Our certainty:

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

The six box problem



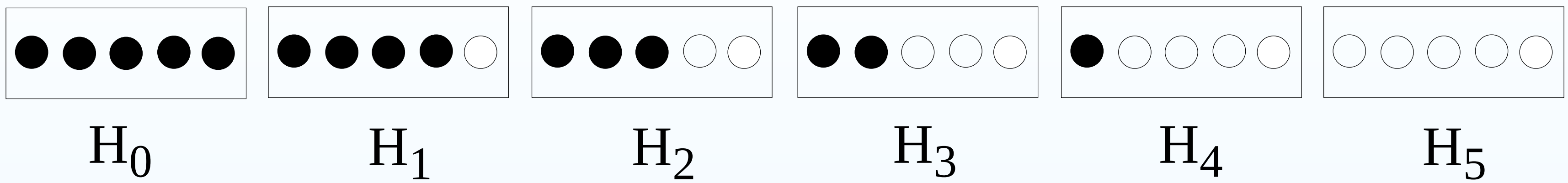
Let us take randomly one of the boxes.
We are in a state of uncertainty concerning several *events*, the most important of which correspond to the following questions:

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

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

- What happens after we have extracted one ball and looked its color?
  - Intuitively we now how to roughly change our opinion.
  - Can we do it quantitatively, in an objective way?

The six box problem



Let us take randomly one of the boxes.
We are in a state of uncertainty concerning several *events*, the most important of which correspond to the following questions:

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

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

- What happens after we have extracted one ball and looked its color?
  - Intuitively we now how to roughly change our opinion.
  - Can we do it quantitatively, in an objective way?
- And after a sequence of extractions?

The toy inferential experiment

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

The toy inferential experiment

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

This toy experiment is conceptually very close to what we do in Physics

- try to guess what we cannot see (the electron mass, a branching ratio, etc)

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

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

An interesting exercise

Probabilities of the 4 sequences from the first 2 extractions (with reintroduction) from the box of unknow composition:

- WW
- WB
- BW
- BB

An interesting exercise

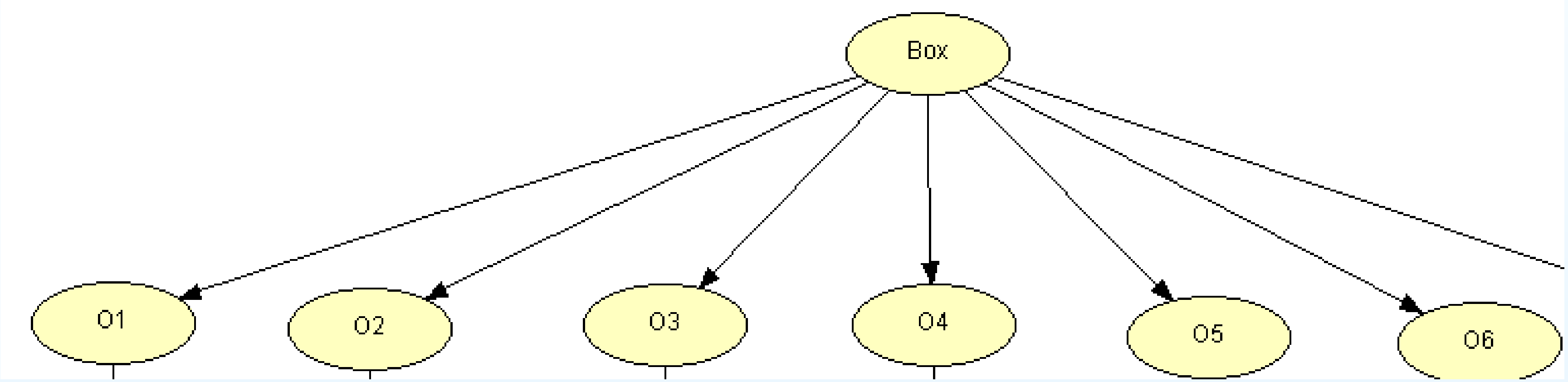
Probabilities of the 4 sequences from the first 2 extractions (with reintroduction) from the box of unknow composition:

- WW
- WB
- BW
- BB

If you have the possibility to win a prize if you predict the right sequence, on which one would you put your money?

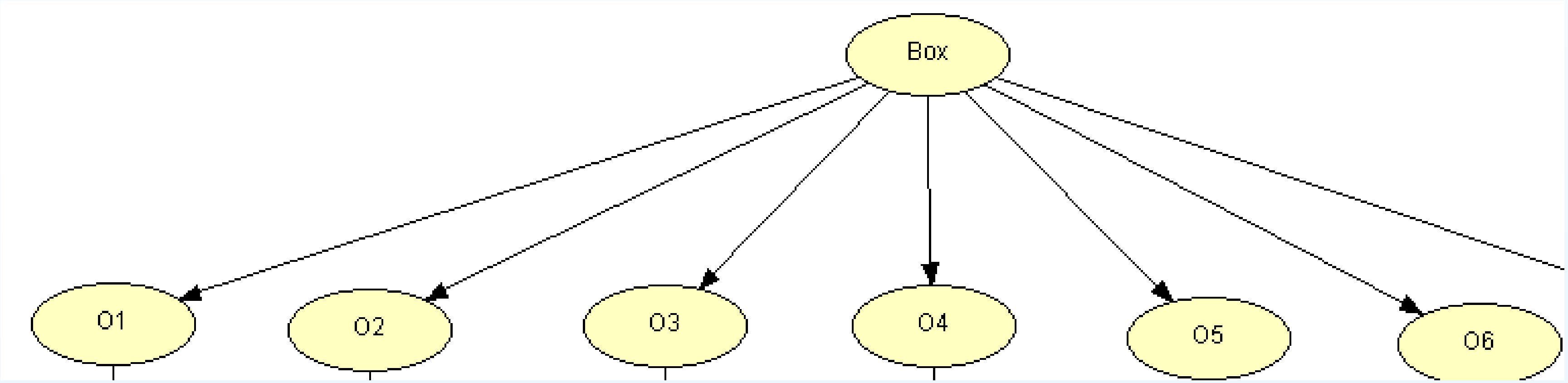
Cause-effect representation

box content $\longrightarrow$ observed color



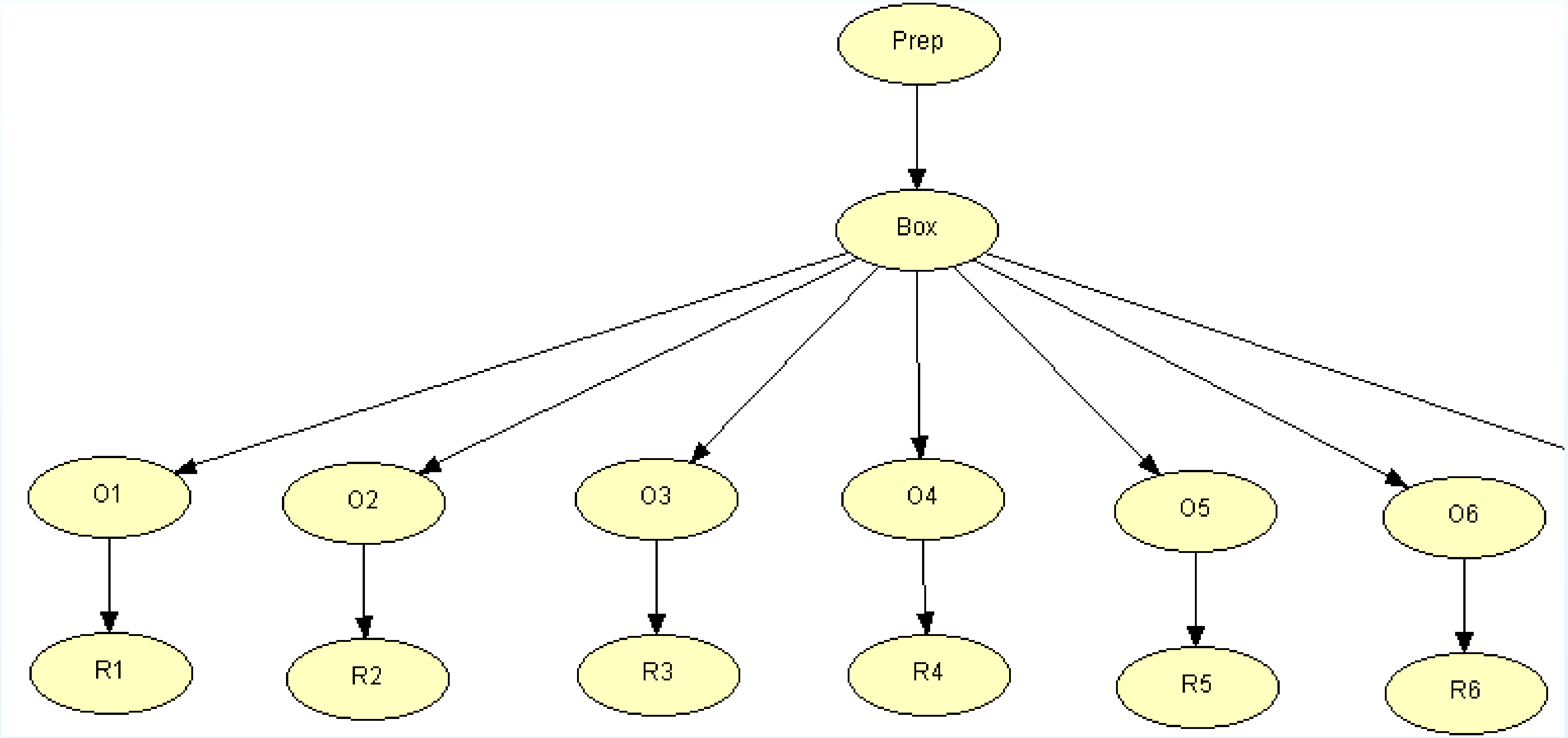
Cause-effect representation

box content $\longrightarrow$ observed color

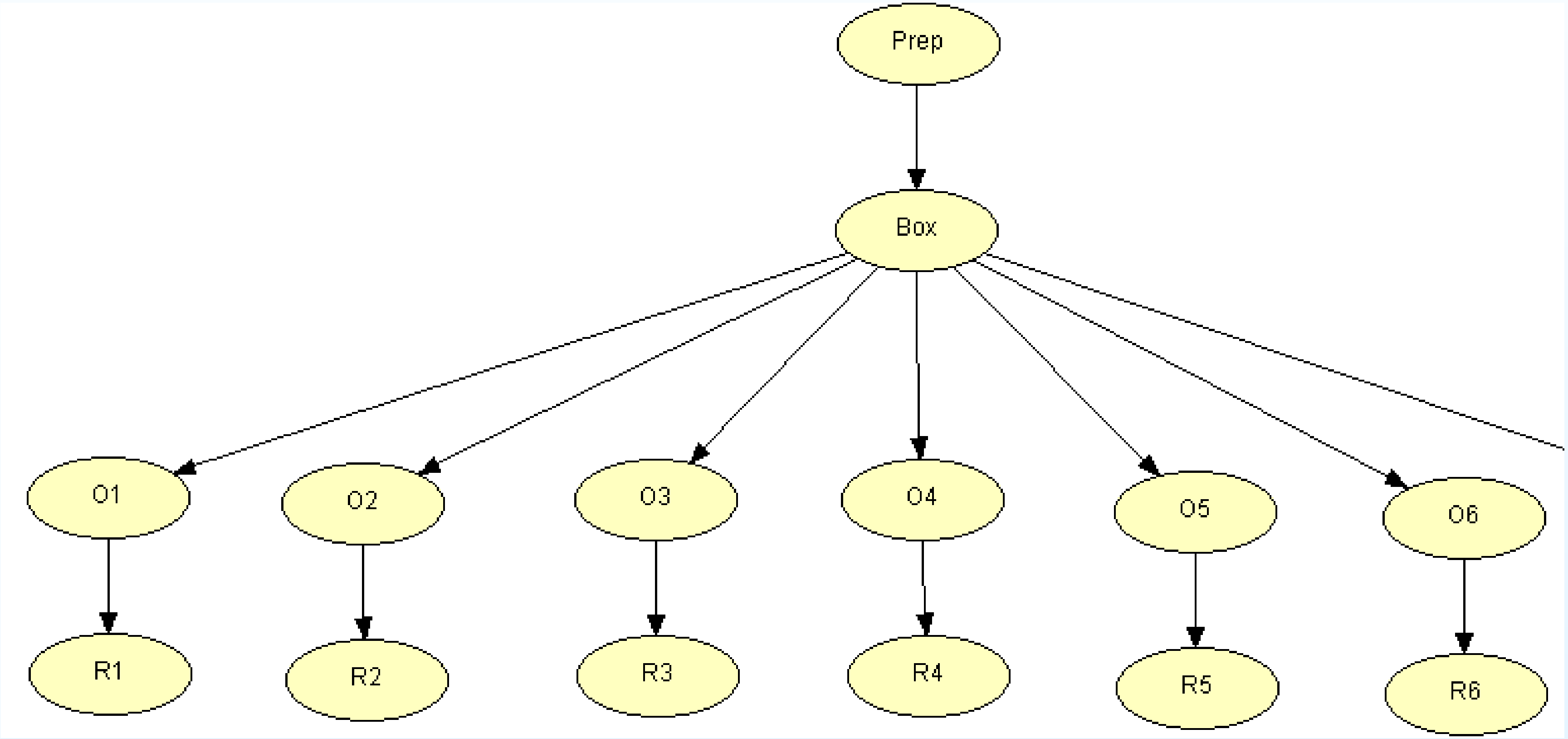


An effect might be the cause of another effect $\longrightarrow$

A network of causes and effects

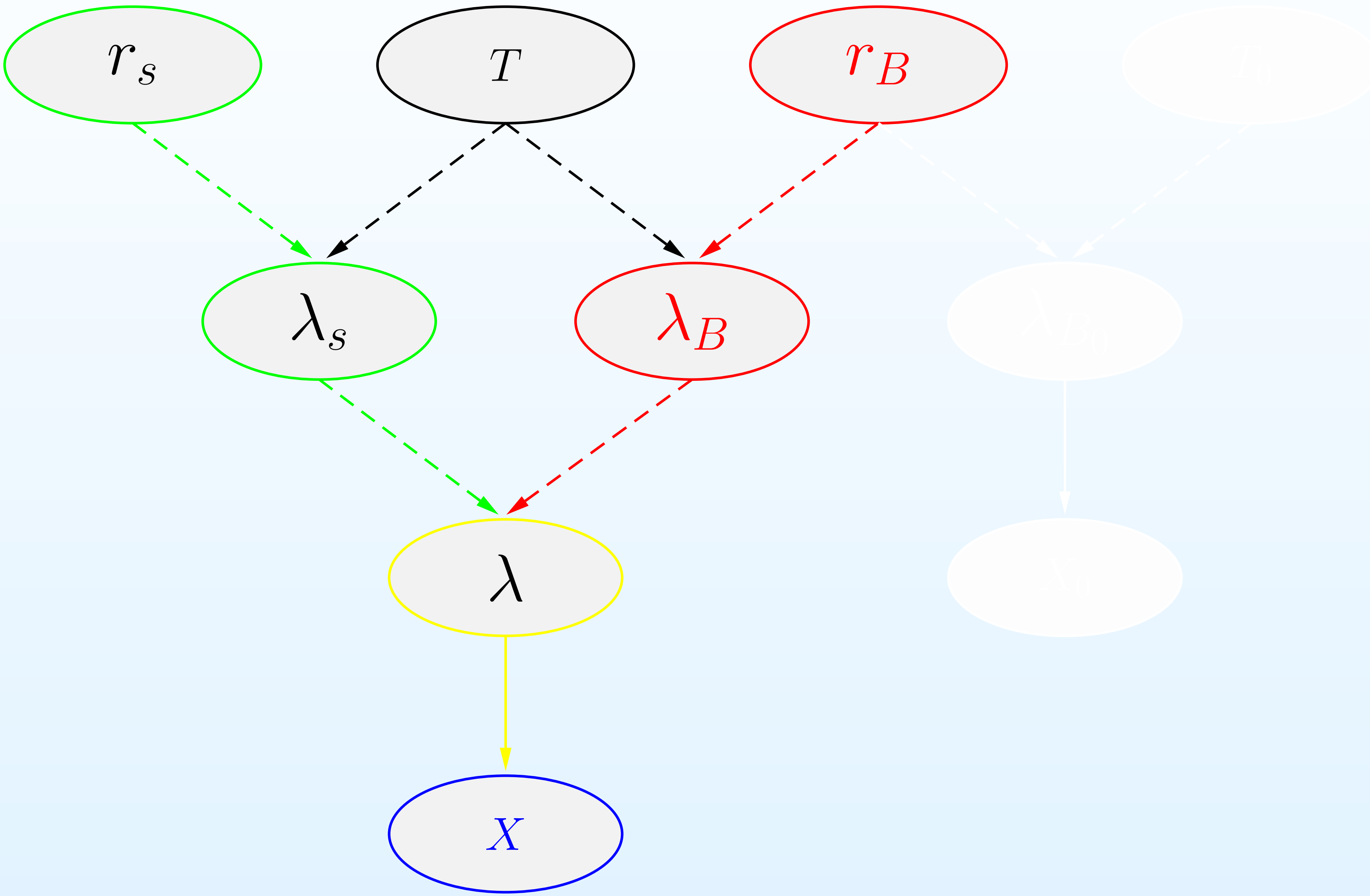


A network of causes and effects

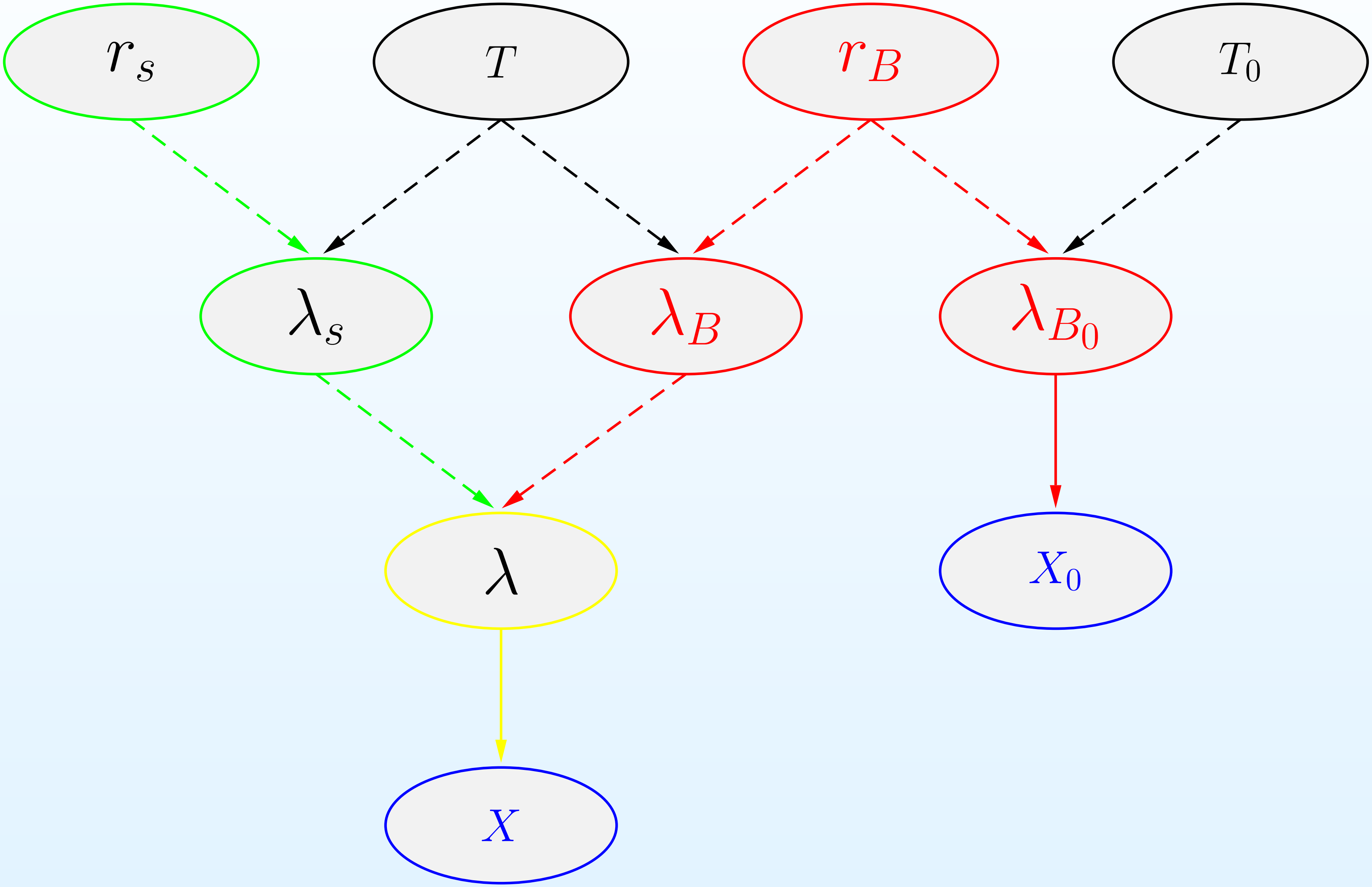


and so on. . . $\Rightarrow$ Physics applications

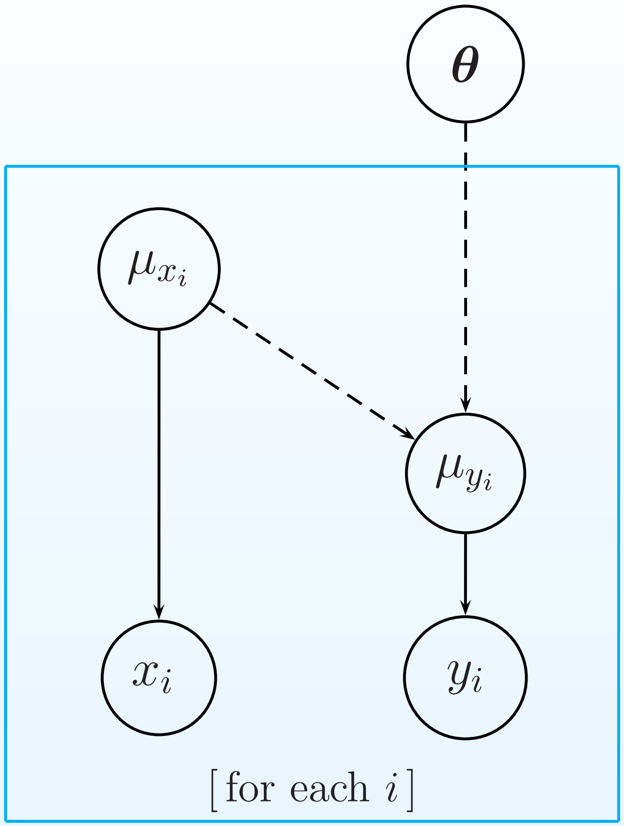
Signal and background



Signal and background



A different way to view fit issues



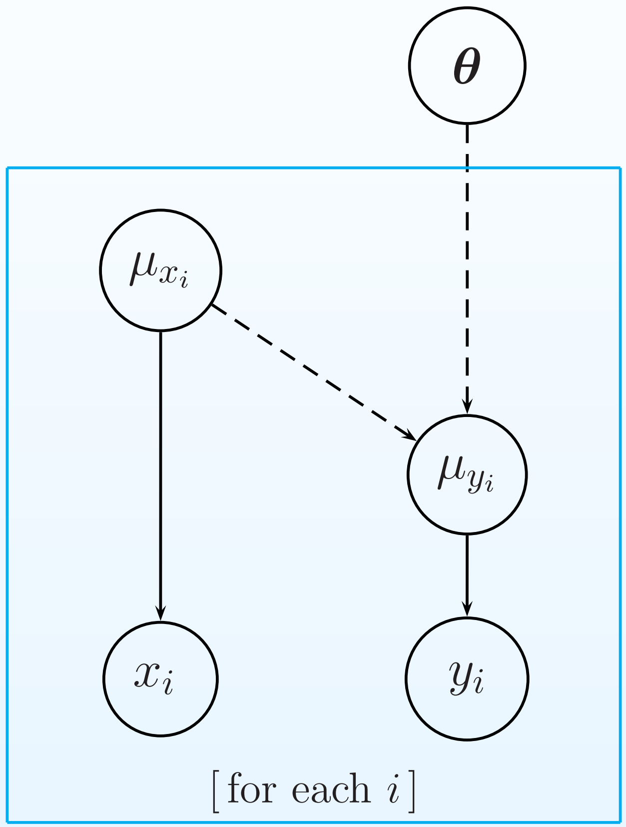
Determistic link μ_x's to μ_y's

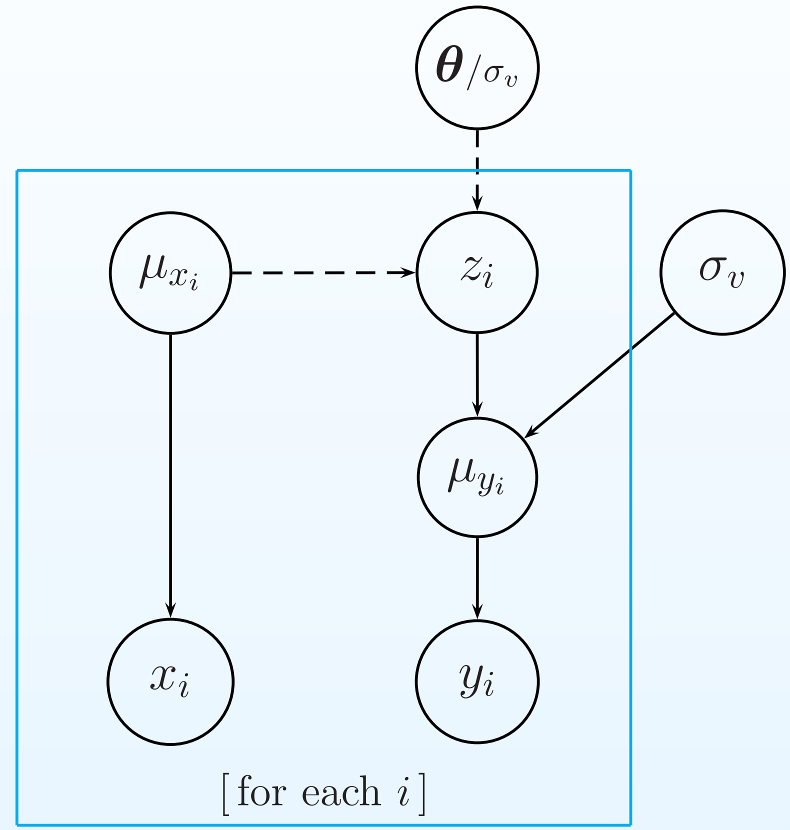
Probabilistic links $\mu_x \longrightarrow x$, $\mu_y \longrightarrow y$

(errors on both axes!)

$\Rightarrow$ aim of fit: $\{\boldsymbol{x}, \boldsymbol{y}\} \longrightarrow \boldsymbol{\theta}$

A different way to view fit issues





Determistic link μ_x's to μ_y's

Probabilistic links $\mu_x \longrightarrow x$, $\mu_y \longrightarrow y$

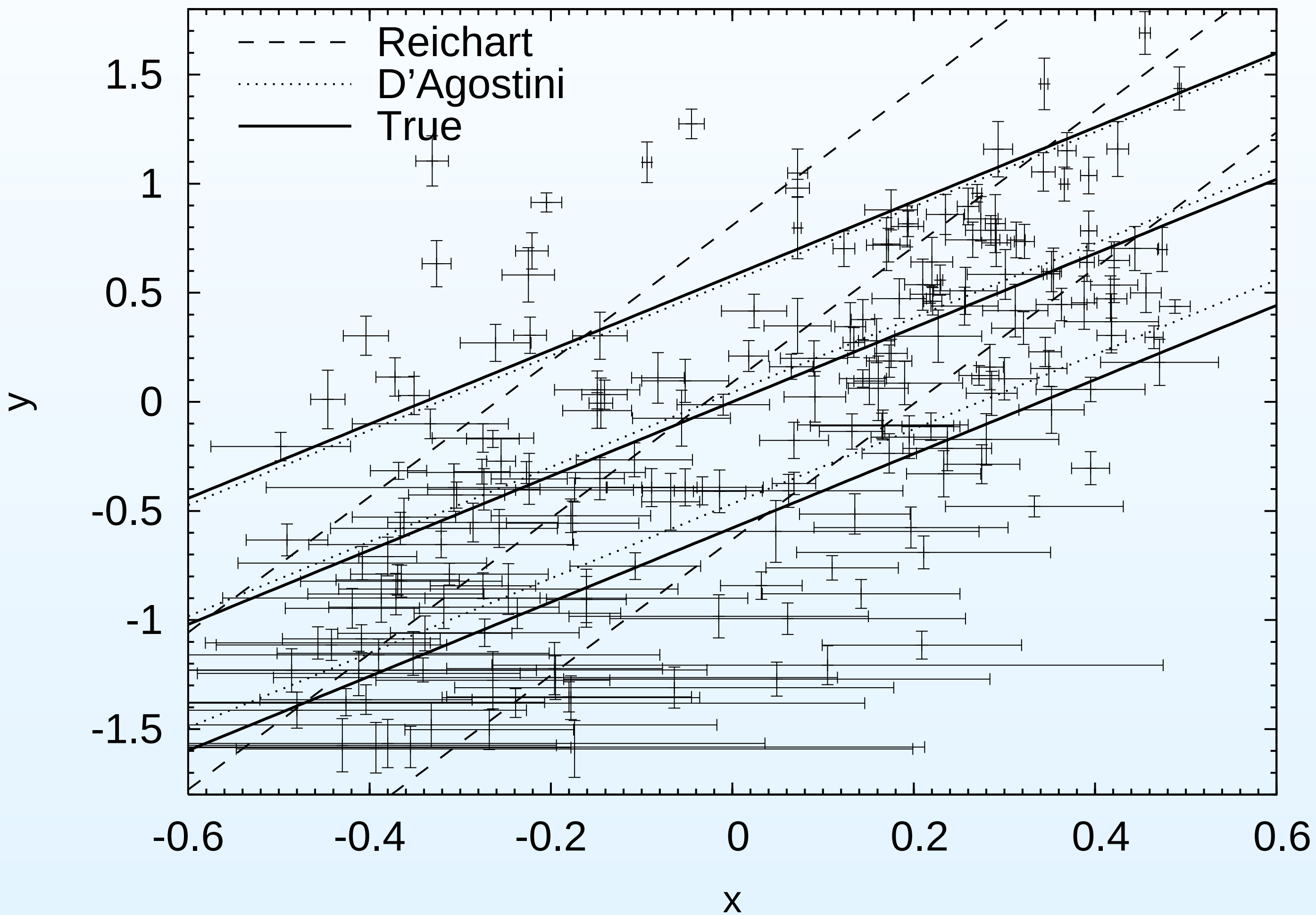
(errors on both axes!)

$\Rightarrow$ aim of fit: $\{\boldsymbol{x}, \boldsymbol{y}\} \rightarrow \boldsymbol{\theta}$

Extra spread of the data points

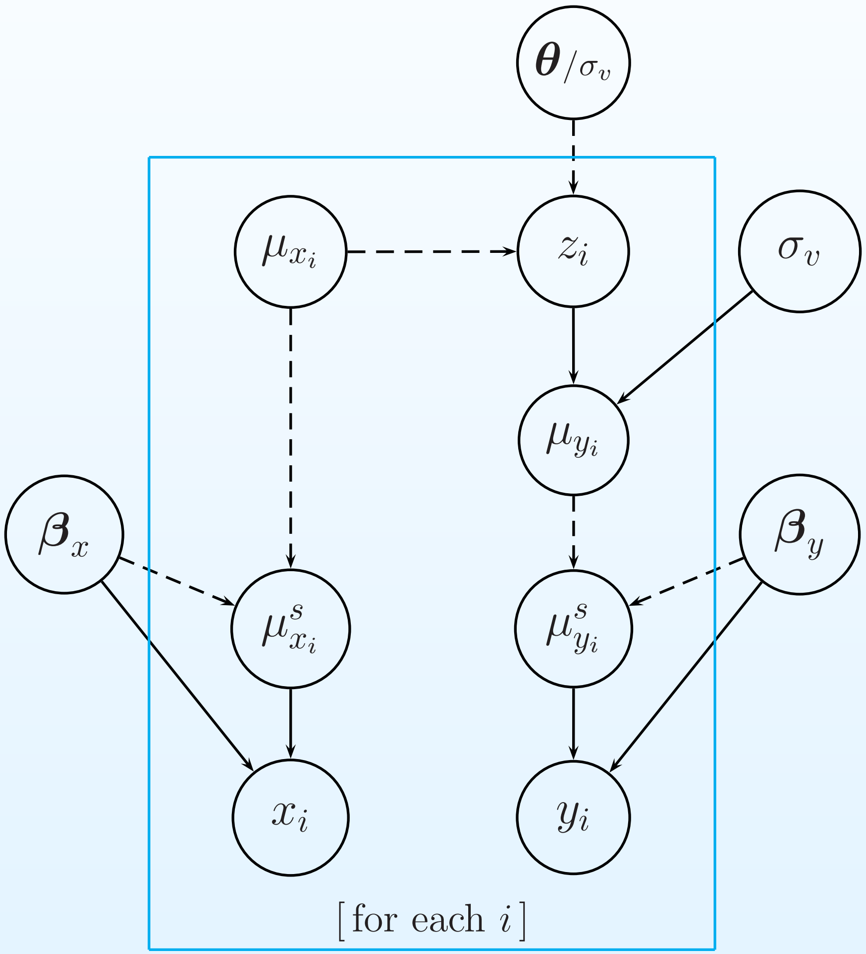
A different way to view fit issues

A physics case (from Gamma ray burts):



(Guidorzi et al., 2006)

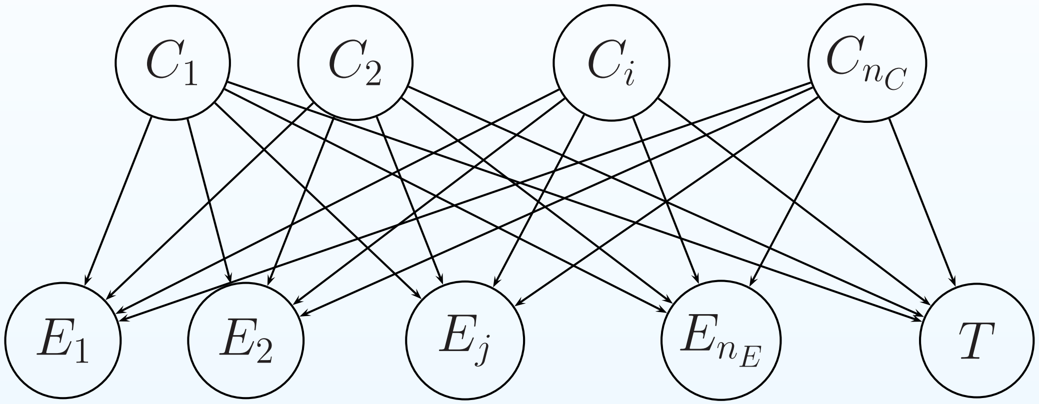
A different way to view fit issues



Adding systematics

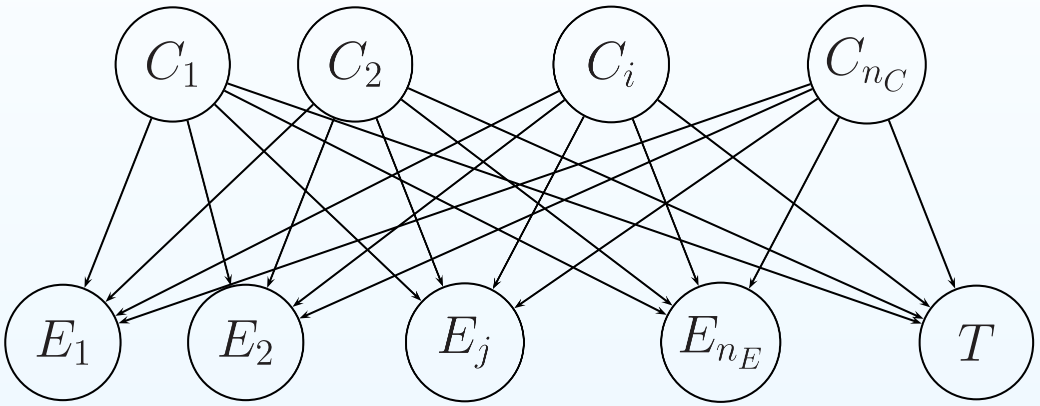
Unfolding a discretized spectrum

Probabilistic links: Cause-bins $\leftrightarrow$ effect-bins

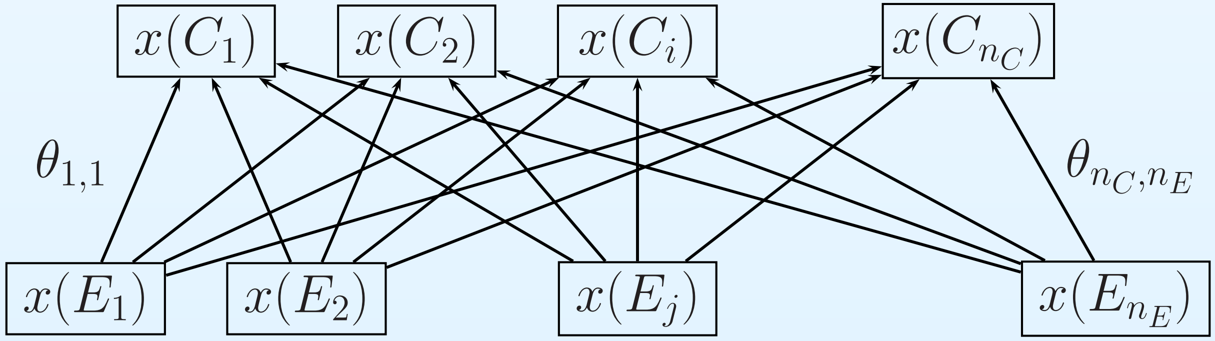


Unfolding a discretized spectrum

Probabilistic links: Cause-bins $\leftrightarrow$ effect-bins

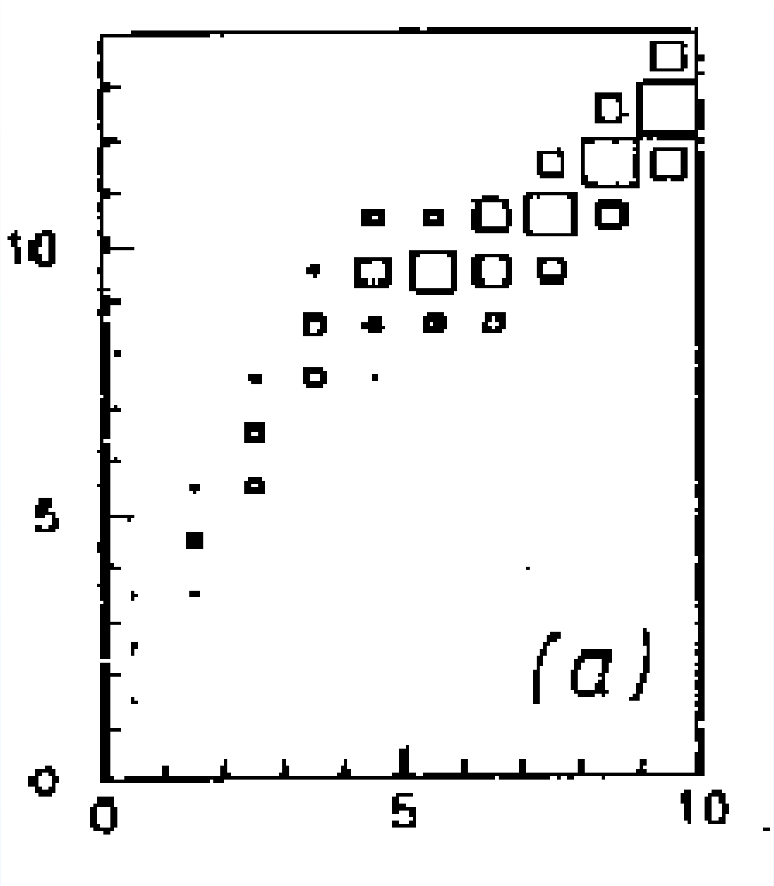


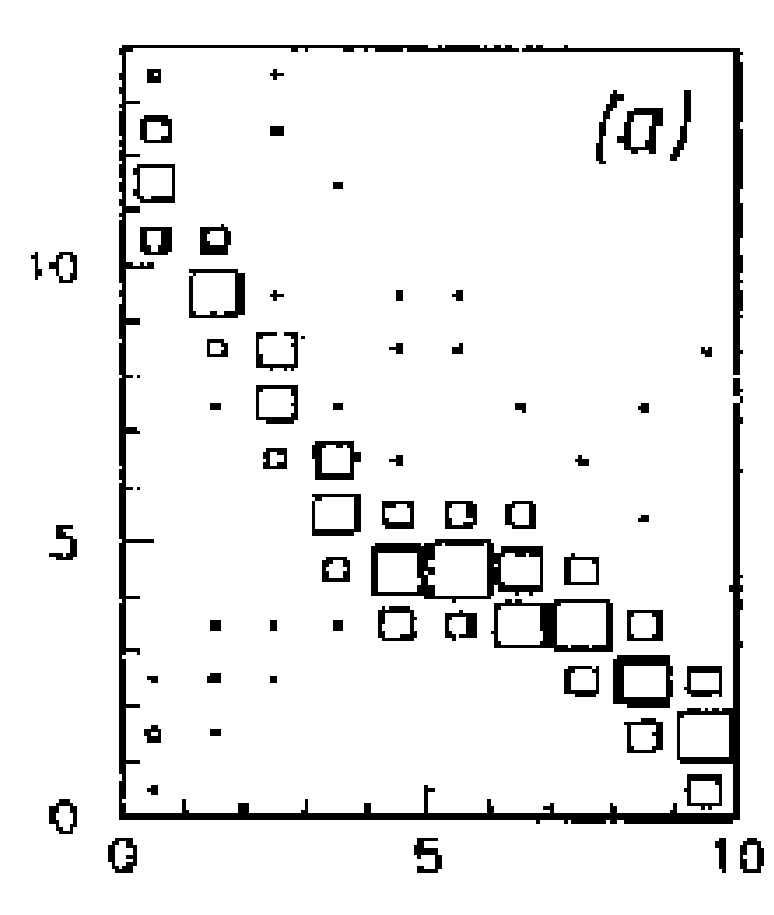
Sharing the observed events among the cause-bins



Unfolding a discretized spectrum

Academic smearing matrices:





Two main streams of reasoning

Two main streams of reasoning

- **Falsificationist approach**
  [and statistical variations over the theme].

Two main streams of reasoning

- **Falsificationist approach**
  [and statistical variations over the theme].

- **Probabilistic approach**
  [In the sense that probability theory is used throughly]

Summary about 'falsificationism/statistics'

A) **if** $C_i \not\longrightarrow E$, and **we observe** E
$\Rightarrow C_i$ is impossible ('false')

Summary about ‘falsificationism/statistics’

A) **if** $C_i \not\longrightarrow E$, and **we observe** E

$\Rightarrow C_i$ is impossible (‘false’)

B) **if** $C_i \xrightarrow[\text{small probability}]{} E$, and **we observe** E

$\Rightarrow C_i$ has small probability to be true

“most likely false”

Summary about ‘falsificationism/statistics’

A) **if** $C_i \not\longrightarrow E$, and **we observe** E OK

$\Rightarrow C_i$ is impossible (‘false’)

B) **if** $C_i \xrightarrow[\text{small probability}]{} E$, and **we observe** E

$\Rightarrow C_i$ has small probability to be true

“most likely false”

Summary about ‘falsificationism/statistics’

A) **if** $C_i \not\longrightarrow E$, and **we observe** E OK

$\Rightarrow C_i$ is impossible (‘false’)

B) **if** $C_i \xrightarrow[\text{small probability}]{} E$, and **we observe** E NO

$\Rightarrow C_i$ has small probability to be true

“most likely false”

(The base of tests, p-values, etc.)

Example

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

Simplified model:

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

H_1='HIV' (Infected) E_1 = Positive

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

Example

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

Simplified model:

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

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

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

Example

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

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

$$\begin{tikzcd} H_1\text{='HIV' (Infected)} \arrow[r] & E_1 \text{ = Positive} \\ H_2\text{='}\overline{\text{HIV}}\text{' (Healthy)} \arrow[r] \arrow[ur] & E_2 \text{ = Negative} \end{tikzcd}$$

Result: $\Rightarrow$ Positive

Example

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

Simplified model:

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

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

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

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

Infected or healthy?

What to conclude?

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

- "It is practically impossible that the person is healthy, since it was practically impossible that an healthy person would result positive"?

What to conclude?

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

- "It is practically impossible that the person is healthy, since it was practically impossible that an healthy person would result positive"
- "There is only 0.2% probability that the person has no HIV" ?

What to conclude?

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

- ”It is practically impossible that the person is healthy, since it was practically impossible that an healthy person would result positive”
- “There is only 0.2% probability that the person has no HIV”
- “We are 99.8% confident that the person is infected”?

What to conclude?

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

- ”It is practically impossible that the person is healthy, since it was practically impossible that an healthy person would result positive”
- “There is only 0.2% probability that the person has no HIV”
- “We are 99.8% confident that the person is infected”
- “The hypothesis H_1=Healthy is ruled out with 99.8% C.L.”

?

What to conclude?

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

- ~~”It is practically impossible that the person is healthy, since it was practically impossible that an healthy person would result positive”~~
- ~~“There is only 0.2% probability that the person has no HIV”~~
- ~~“We are 99.8% confident that the person is infected”~~
- ~~“The hypothesis H_1=Healthy is ruled out with 99.8% C.L.”~~

? NO

Instead, $P(\text{HIV} \,|\, \text{Pos, random Italian}) \approx 45\%$
(We will see in the sequel how to evaluate it correctly)

What to conclude?

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

- ~~"It is practically impossible that the person is healthy, since it was practically impossible that an healthy person would result positive"~~
- ~~"There is only 0.2% probability that the person has no HIV"~~
- ~~"We are 99.8% confident that the person is infected"~~
- ~~"The hypothesis H_1=Healthy is ruled out with 99.8% C.L."~~

? **NO**

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

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

‘Standard’ statistical tests, p-values, etc

- This kind of logical mistake is quite common.
  “Si sbaglia da professionisti” (P. Conte)

‘Standard’ statistical tests, p-values, etc

- This kind of logical mistake is quite common.
  “Si sbaglia da professionisti” (P. Conte)
- Yes, statisticians have invented p-values (something like ‘probability of the tail(s)’ – I cannot enter into details) to overcome the problem that often the probability of any observation is always very small and the null hypotheses would always be rejected.

‘Standard’ statistical tests, p-values, etc

- This kind of logical mistake is quite common.
  “Si sbaglia da professionisti” (P. Conte)
- Yes, statisticians have invented p-values (something like ‘probability of the tail(s)’ – I cannot enter into details) to overcome the problem that often the probability of any observation is always very small and the null hypotheses would always be rejected.
  But
  - as far as logic is concerned, the situation is worsened (. . . although p-values ‘often, by chance work’).

‘Standard’ statistical tests, p-values, etc

- This kind of logical mistake is quite common.
  “Si sbaglia da professionisti” (P. Conte)
- Yes, statisticians have invented p-values (something like ‘probability of the tail(s)’ – I cannot enter into details) to overcome the problem that often the probability of any observation is always very small and the null hypotheses would always be rejected.
  But
  - as far as logic is concerned, the situation is worsened (. . . although p-values ‘often, by chance work’).
- Mistrust statistical tests, unless you know the details of what it has been done.

  $\longrightarrow$ You might take bad decisions!

Example from particle/event classification

A discrimination analysis can find a ‘discriminator’ d related to a particle p_i, or to a certain event of interest (e.g. as a result from neural networks or whatever).

Example from particle/event classification

A discrimination analysis can find a 'discriminator' d related to a particle p_i, or to a certain event of interest (e.g. as a result from neural networks or whatever).

OK, but, in general

$$P(d \geq d_{cut} \,|\, p_i) \neq P(p_i \,|\, d \geq d_{cut}) \quad \textbf{!}$$

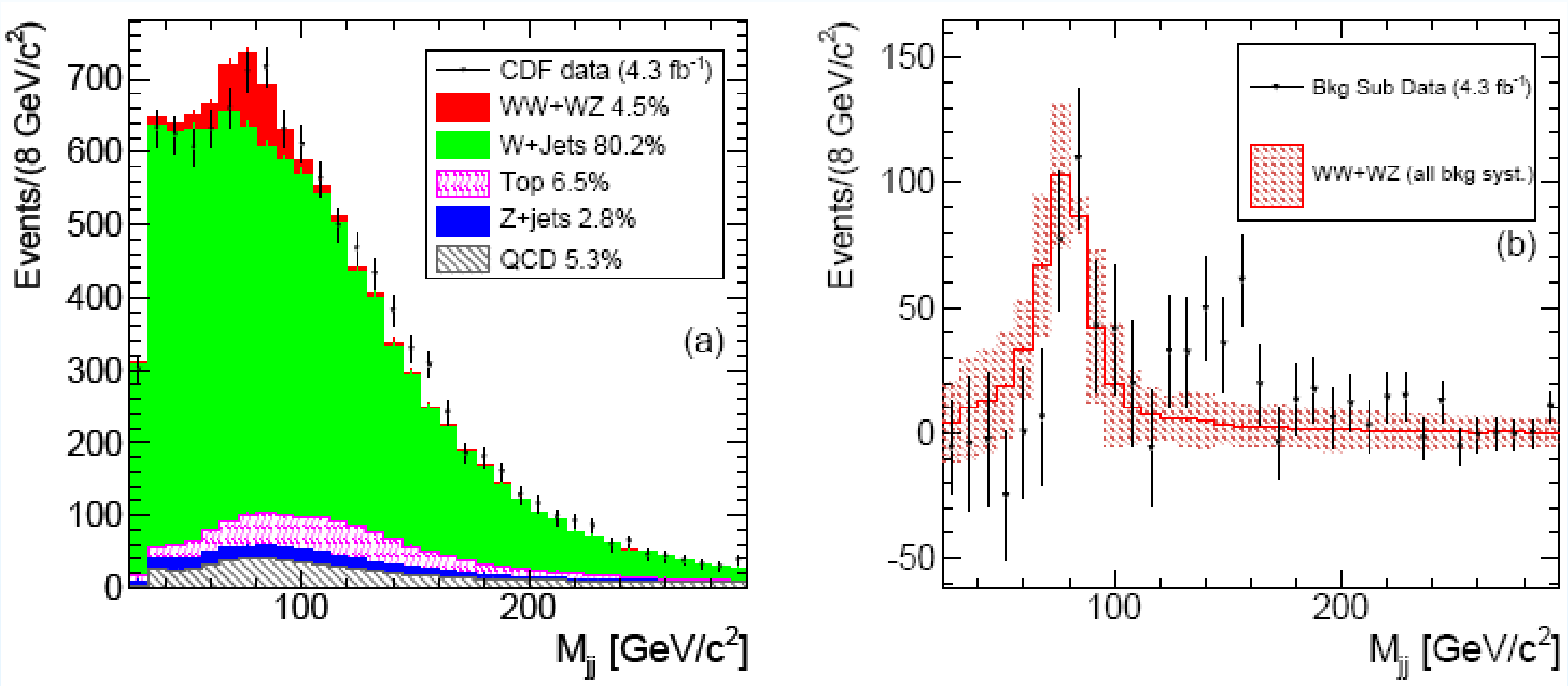
(I am pretty sure that often what is called a probability of a particle, or an event, of being something is not really that probability...)

Claims of discoveries based on p-values

But, amazingly, there are ‘claims’ of discoveried based on logical mistakes of this kind a p-value misunderstood as probability of the hypothesis to test.

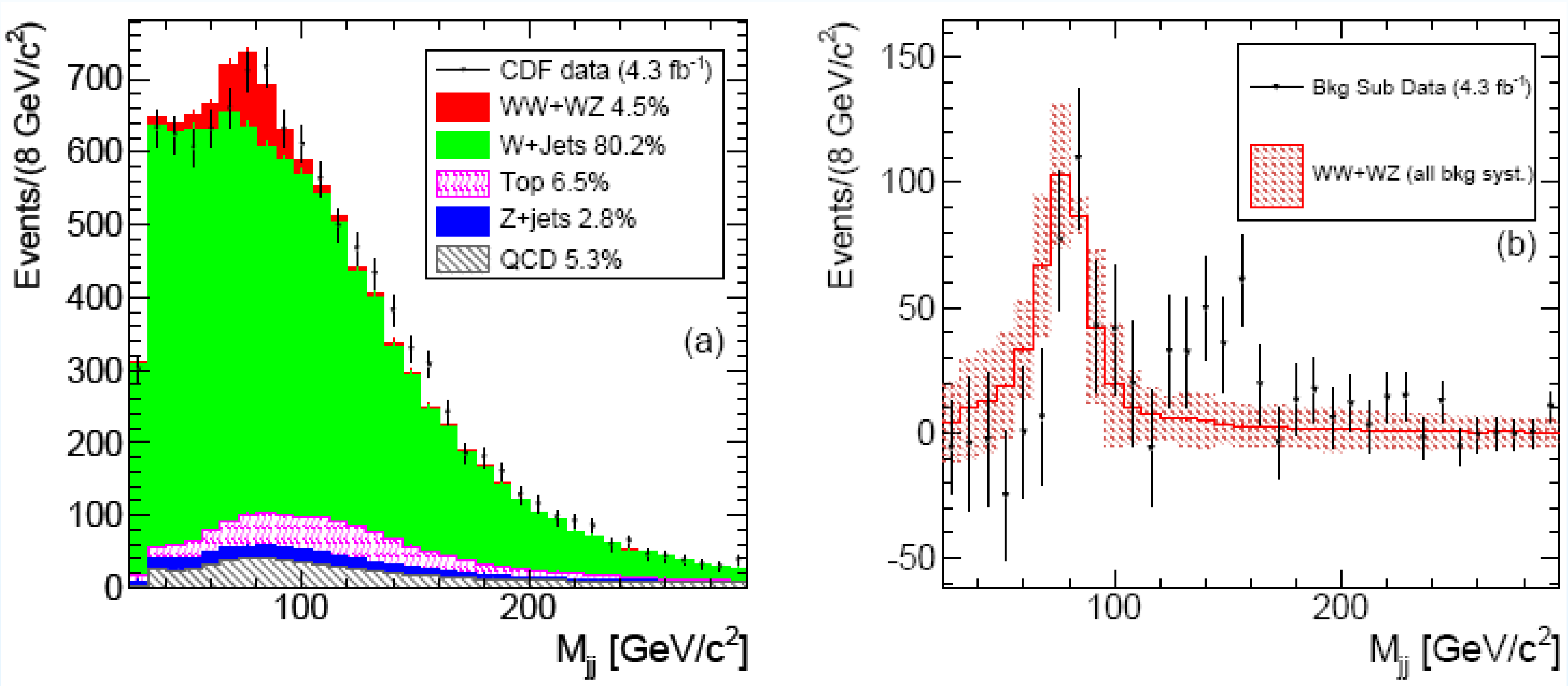
Claims of discoveries based on p-values

→ Last case from particle physics:
CDF, Fermilab, April 2011



Claims of discoveries based on p-values

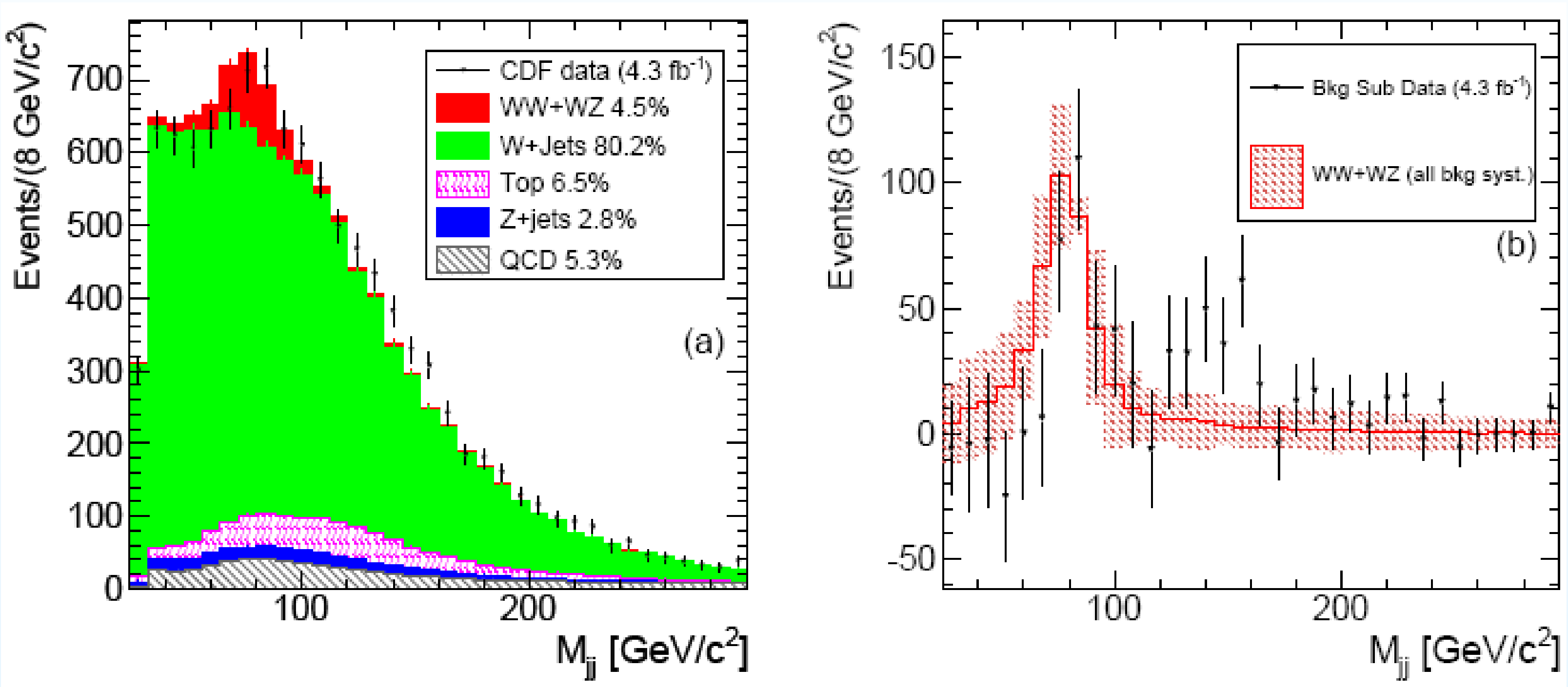
→ Last case from particle physics:
CDF, Fermilab, April 2011



p-value 0.8×10^{-3}: probability of observing an eccess equal or larger than the observed one, given the best understanding of Standard Model, detector, etc.

Claims of discoveries based on p-values

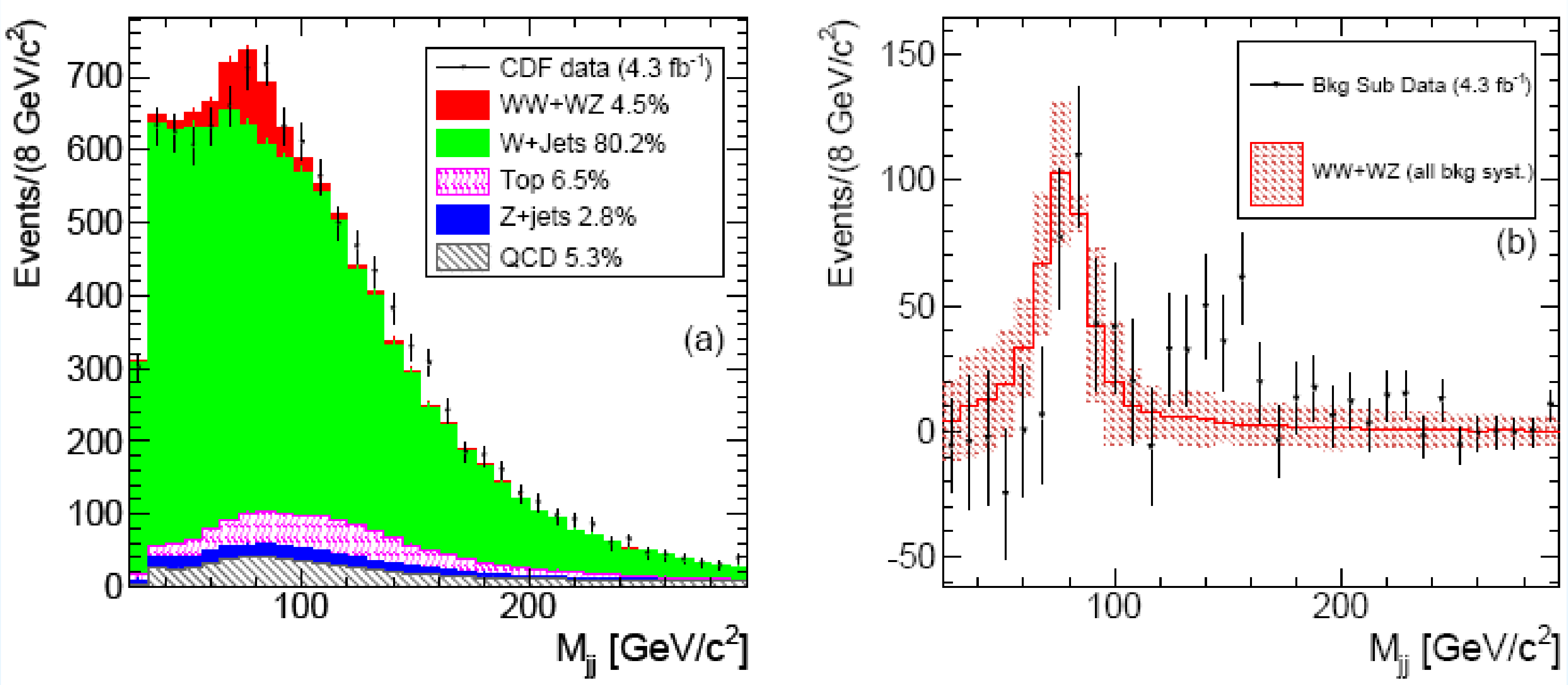
→ Last case from particle physics:
CDF, Fermilab, April 2011 (p-value 0.8×10^{-3})



So, what?

Claims of discoveries based on p-values

→ Last case from particle physics:
CDF, Fermilab, April 2011 (p-value 0.8×10^{-3})

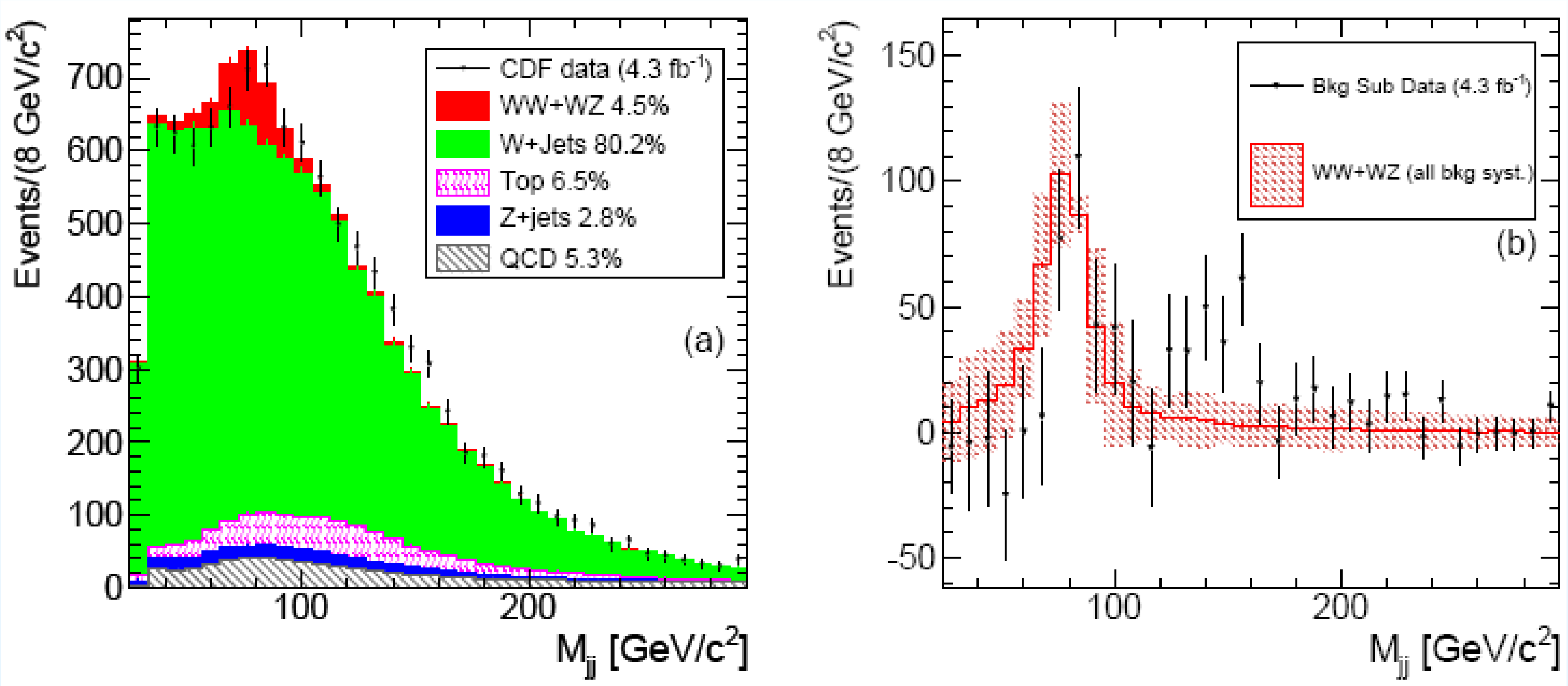


So, what?

What is the probability that the first two speakers of this school meet twice, the same morning, within meters on Gran Sasso mauntain?

Claims of discoveries based on p-values

→ Last case from particle physics:
CDF, Fermilab, April 2011 (p-value 0.8×10^{-3})

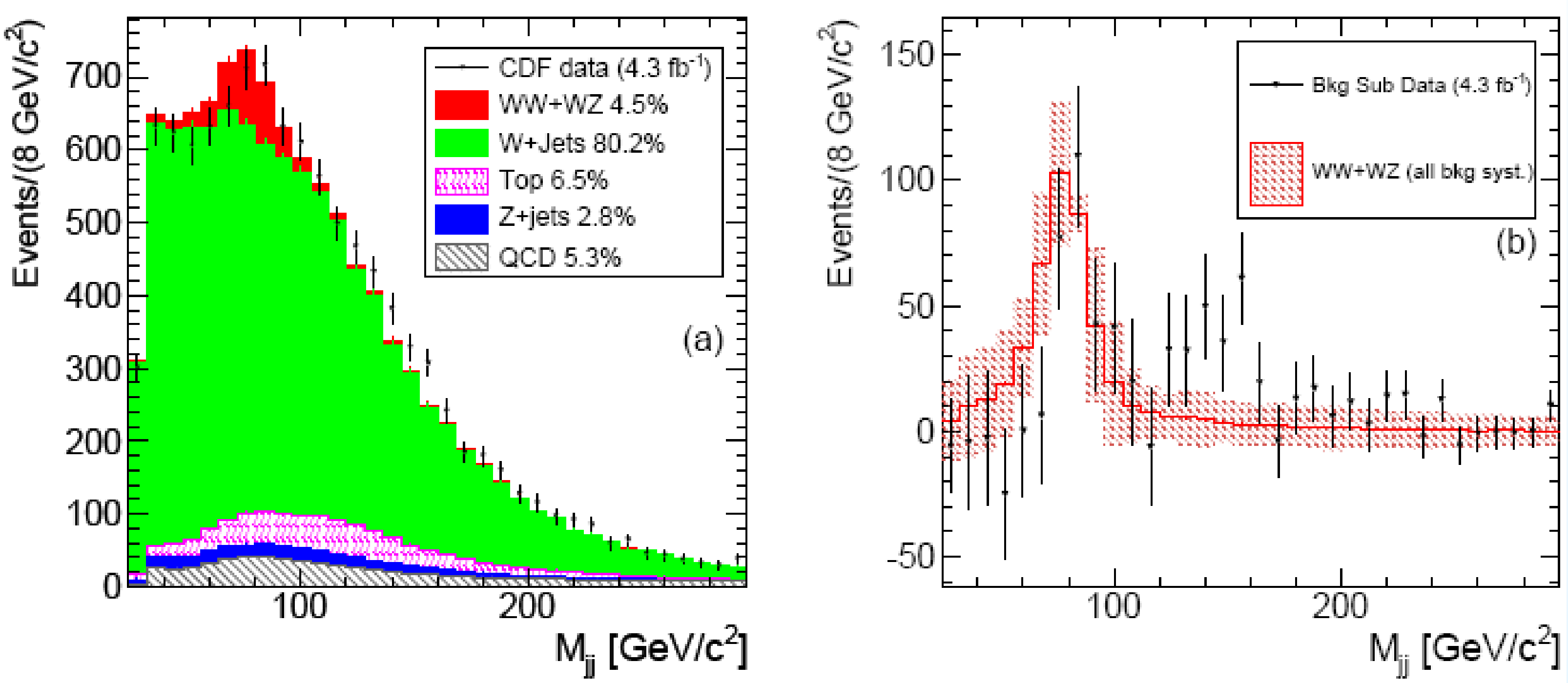


So, what?

If such an event happenes, should we be logically compelled to think that it was not 'by chance', but there must be a particular cause to cause it?
(A sign of God?)

Claims of discoveries based on p-values

→ Last case from particle physics:
CDF, Fermilab, April 2011 (p-value 0.8×10^{-3})



No problem about evaluation of p-values, but about their meaning, and how they are perceived by scientists and are spread to the media:
→ unjustified excitement and expectations

From p-values to beliefs (?!)

- Fermilab Today:
  "This means that there is less than a 1 in 1375 chance that the effect is mimicked by a statistical fluctuation"

From p-values to beliefs (?!)

- Fermilab Today:
  “This means that there is less than a 1 in 1375 chance that the effect is mimicked by a statistical fluctuation”
- Press releases, press conference and ‘solemn seminar’

From p-values to beliefs (?!)

- Fermilab Today:
  “This means that there is less than a 1 in 1375 chance that the effect is mimicked by a statistical fluctuation”
- Press releases, press conference and ‘solemn seminar’
- New York Times:
  “The experimenters estimate that there is a less than a quarter of 1 percent chance their bump is a statistical fluctuation”

From p-values to beliefs (?!)

- Fermilab Today:
  “This means that there is less than a 1 in 1375 chance that the effect is mimicked by a statistical fluctuation”
- Press releases, press conference and ‘solemn seminar’
- New York Times:
  “The experimenters estimate that there is a less than a quarter of 1 percent chance their bump is a statistical fluctuation”
- Discovery News:
  “This result has a 99.7 percent chance of being correct (and a 0.3 percent chance of being wrong)”

From p-values to beliefs (?!)

- Guardian (blob by Jon Butterworth, brillant physicist and CDF collaborator:)
  - “The paper quotes a one-in-ten-thousand (0.0001) chance that this bump is a fluke”

From p-values to beliefs (?!)

- Guardian (blob by Jon Butterworth, brillant physicist and CDF collaborator:)
  - “The paper quotes a one-in-ten-thousand (0.0001) chance that this bump is a fluke”
  - But he concludes with very wise and sharable statements:
    - “My money is on the false alarm at the moment”
    - “... but I would be very happy to lose it.”
    - “... And I reserve the right to change my mind rapidly as more data come in”

From p-values to beliefs (?!)

- Guardian (blob by Jon Butterworth, brillant physicist and CDF collaborator:)
  - “The paper quotes a one-in-ten-thousand (0.0001) chance that this bump is a fluke”
  - But he concludes with very wise and sharable statements:
    - “My money is on the false alarm at the moment”
    - “... but I would be very happy to lose it.”
    - “... And I reserve the right to change my mind rapidly as more data come in”

$\Rightarrow$ His beliefs are in clear contradictions with the way he tried to explain to the general public the meaning of the p-value!

Conflict: natural thinking $\Leftrightarrow$ cultural superstructure

Why? 'Who' is responsible?

- Since beginning of '900 it is dominant an unnatural approach to probability, in contrast to that of the founding fathers (Poisson, Bernoulli, Bayes, Laplace, Gauss, . . .).

Conflict: natural thinking $\Leftrightarrow$ cultural superstructure

Why? ‘Who’ is responsible?

- Since beginning of ’900 it is dominant an unnatural approach to probability, in contrast to that of the founding fathers (Poisson, Bernoulli, Bayes, Laplace, Gauss, . . .).
- In this, still dominant approach (frequentism) it is forbidden to speak about probability of hypotheses, probability of causes, probability of values of physical quantities, etc.

Conflict: natural thinking ⇔ cultural superstructure

Why? ‘Who’ is responsible?

- Since beginning of ’900 it is dominant an unnatural approach to probability, in contrast to that of the founding fathers (Poisson, Bernoulli, Bayes, Laplace, Gauss, . . .).
- In this, still dominant approach (frequentism) it is forbidden to speak about probability of hypotheses, probability of causes, probability of values of physical quantities, etc.
- The concept of probability of causes [“*The essential problem of the experimental method*” (Poincaré)] has been surrogated by the mechanism of hypothesis test and ‘p-values’. (And of ‘confidence intervals’ in parametric inference)

Conflict: natural thinking $\Leftrightarrow$ cultural superstructure

Why? 'Who' is responsible?

- Since beginning of '900 it is dominant an unnatural approach to probability, in contrast to that of the founding fathers (Poisson, Bernoulli, Bayes, Laplace, Gauss, . . .).
- In this, still dominant approach (frequentism) it is forbidden to speak about probability of hypotheses, probability of causes, probability of values of physical quantities, etc.
- The concept of probability of causes [*"The essential problem of the experimental method"* (Poincaré)] has been surrogated by the mechanism of hypothesis test and 'p-values'. (And of 'confidence intervals' in parametric inference)

$\Rightarrow$ BUT people think naturally in terms of probability of causes, and use p-values as if they were probabilities of null hypotheses.

Conflict: natural thinking $\Leftrightarrow$ cultural superstructure

Why? ‘Who’ is responsible?

- Since beginning of ’900 it is dominant an unnatural approach to probability, in contrast to that of the founding fathers (Poisson, Bernoulli, Bayes, Laplace, Gauss, . . .).
- In this, still dominant approach (frequentism) it is forbidden to speak about probability of hypotheses, probability of causes, probability of values of physical quantities, etc.
- The concept of probability of causes [*“The essential problem of the experimental method”* (Poincaré)] has been surrogated by the mechanism of hypothesis test and ‘p-values’. (And of ‘confidence intervals’ in parametric inference)

$\Rightarrow$ BUT people think naturally in terms of probability of causes, and use p-values as if they were probabilities of null hypotheses. $\Rightarrow$ Terrible mistakes!

“Probability” Vs probability…

Errors on ratios of small numbers of events
F. James$^{(*)}$ and M. Roos

When the result of the measurement of a physical quantity is published as $R=R_0\pm\sigma_0$ without further explanation, it is implied that R is a Gaussian-distributed measurement with mean R_0 and variance $\sigma_0{}^2$. This allows one to calculate various confidence intervals of given "probability", i.e. the "probability" P that the true value of R is within a given interval. P is given by the area under the corresponding part of the Gaussian curve, and is the basis of well-known rules-of-thumb such as "the probability of exceeding two standard deviations is 5%".

$^{(*)}$ Influential CERN ’frequentistic guru’ of HEP community

“Probability” Vs probability. . .

Errors on ratios of small numbers of events
F. James$^{(*)}$ and M. Roos

When the result of the measurement of a physical quantity is published as $R=R_0\pm\sigma_0$ without further explanation, it is implied that R is a Gaussian-distributed measurement with mean R_0 and variance $\sigma_0{}^2$. This allows one to calculate various confidence intervals of given "probability", i.e. the "probability" P that the true value of R is within a given interval. P is given by the area under the corresponding part of the Gaussian curve, and is the basis of well-known rules-of-thumb such as "the probability of exceeding two standard deviations is 5%".

$^{(*)}$ Influential CERN ’frequentistic guru’ of HEP community

Nowhere in the article is clarified why “probability” is in quote marks! $\Rightarrow$ they know they are not allowed to speak about “probability of true values”

What to do?

$\Rightarrow$ ‘Forward to past’

What to do?

$\Rightarrow$ 'Forward to past'

But benefitting of

- Theoretical progresses in probability theory
- Advance in computation (both symbolic and numeric)

  $\longrightarrow$ many frequentistic ideas had their *raison d'être* in the computational barrier (and many simplified – often simplistic – methods were ingeniously worked out)

  $\longrightarrow$ no longer an excuse!

What to do?

$\Rightarrow$ ‘Forward to past’

But benefitting of

- Theoretical progresses in probability theory
- Advance in computation (both symbolic and numeric)

  $\longrightarrow$ many frequentistic ideas had their *raison d’être* in the computational barrier (and many simplified – often simplistic – methods were ingeniously worked out)

  $\longrightarrow$ no longer an excuse!

$\Rightarrow$ Use consistently probability theory

What to do?

⇒ ‘Forward to past’

But benefitting of

- Theoretical progresses in probability theory
- Advance in computation (both symbolic and numeric)
  - → many frequentistic ideas had their *raison d’être* in the computational barrier (and many simplified – often simplistic – methods were ingeniously worked out)
    - → no longer an excuse!

⇒ Use consistently probability theory

- “It’s easy if you try”
- But first you have to recover the intuitive concept of probability.

Probability

What is probability?

End

FINE