Introduction to Probabilistic Reasoning

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

A invitation

to (re-)think

on foundamental aspects

of data analysis.

Uncertainty: some examples

Roll a die:

1, 2, 3, 4, 5, 6: ?

Toss a coin:

Head/Tail: ?

Having to perform a measurement:

Which numbers shall come out from our device ?

Having performed a measurement:

What have we learned about the value of the quantity of interest ?

Many other examples from real life:

Football, weather, tests/examinations, ...

$\rightarrow$ events and their consequences in Risk Management

Rolling a die

Let us consider three outcomes:

$E_1 =$ '1'

$E_2 =$ '2 or 3'

$E_3 =$ '$\geq$ 4'

Rolling a die

Let us consider three outcomes:

$E_1 =$ ‘1’

$E_2 =$ ‘2 or 3‘

$E_3 =$ ‘$\geq$ 4’

We are not uncertain in the same way about E_1, E_2 and E_3:

Rolling a die

Let us consider three outcomes:

$E_1 =$ ‘1’

$E_2 =$ ‘2 or 3‘

$E_3 =$ ‘$\geq$ 4’

We are not uncertain in the same way about E_1, E_2 and E_3:

- Which event do you consider more likely, possible, credible, believable, plausible?

Rolling a die

Let us consider three outcomes:

$$E_1 = \text{‘1’}$$

$$E_2 = \text{‘2 or 3‘}$$

$$E_3 = \text{‘}\geq 4\text{’}$$

We are not uncertain in the same way about E_1, E_2 and E_3:

- Which event do you consider more likely, possible, credible, believable, plausible?
- You will get a prize if the event you chose will occur.
  On which event would you bet?

Rolling a die

Let us consider three outcomes:

$E_1 =$ ‘1’

$E_2 =$ ‘2 or 3‘

$E_3 =$ ‘$\geq$ 4’

We are not uncertain in the same way about E_1, E_2 and E_3:

- Which event do you consider more likely, possible, credible, believable, plausible?
- You will get a prize if the event you chose will occur.
On which event would you bet?
- On which event are you more confident? Which event you trust more, you believe more? etc

Rolling a die

Let us consider three outcomes:

$E_1 =$ '1'

$E_2 =$ '2 or 3'

$E_3 =$ '$\geq$ 4'

We are not uncertain in the same way about E_1, E_2 and E_3:

- Which event do you consider more likely, possible, credible, believable, plausible?
- You will get a prize if the event you chose will occur.
On which event would you bet?
- On which event are you more confident? Which event you trust more, you believe more? etc
- Imagine to repeat the experiment: which event do you expect to occur mostly (more frequently)?

Rolling a die

Let us consider three outcomes:

$E_1 =$ '1'

$E_2 =$ '2 or 3'

$E_3 =$ '$\geq$ 4'

$\Rightarrow$ Many expressions to state our preference

Rolling a die

Let us consider three outcomes:

$E_1 =$ '1'

$E_2 =$ '2 or 3'

$E_3 =$ '$\geq$ 4'

$\Rightarrow$ Many expressions to state our preference

Which reasoning have we applied to prefer E_3?

Rolling a die

Let us consider three outcomes:

$E_1 =$ ‘1’

$E_2 =$ ‘2 or 3‘

$E_3 =$ ‘$\geq$ 4’

$\Rightarrow$ Many expressions to state our preference

Which reasoning have we applied to prefer E_3?
Can we use it for all other events of our interest?

Rolling a die

Let us consider three outcomes:

$E_1 =$ '1'

$E_2 =$ '2 or 3'

$E_3 =$ '$\geq$ 4'

$\Rightarrow$ Many expressions to state our preference

Which reasoning have we applied to prefer E_3?
Can we use it for all other events of our interest?

$\rightarrow$ two envelop 'paradox'

A counting experiment

Imagine a small scintillation counter, with suitable threshold, placed here and now.
We set the measuring time (e.g. 1 s each) and perform some measurements.

A counting experiment

Imagine a small scintillation counter, with suitable threshold, placed here and now.
We set the measuring time (e.g. 1 s each) and perform some measurements.

Remark:
This device can be seen as the quintessence of any ‘counter’:

- nr of alarms per day received by a control station;
- nr of failures per month in a plant;
- or even the nr of holes per km in a road; etc.

A counting experiment

Imagine a small scintillation counter, with suitable threshold, placed here and now.

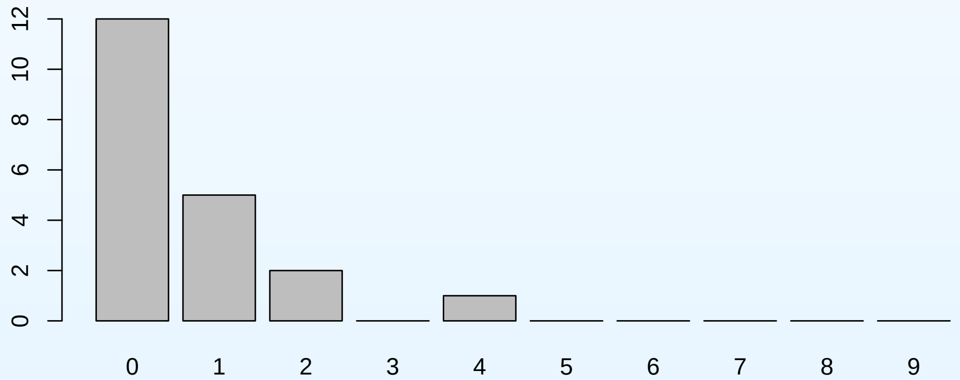
We set the measuring time (e.g. 1 s each) and perform some measurements.

The first 20 outcomes ('reports') are:

0, 0, 1, 0, 0, 0, 1, 2, 0, 0, 1, 1, 0, 4, 2, 0, 0, 0, 0, 1.

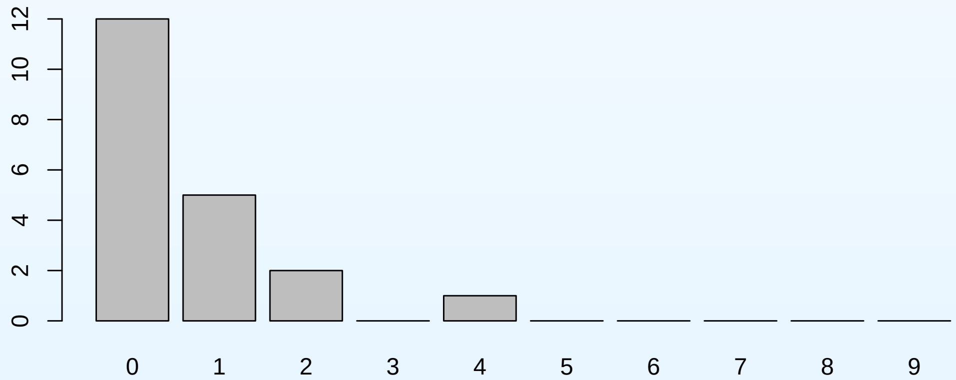
A counting experiment

Imagine a small scintillation counter, with suitable threshold, placed here and now.
We set the measuring time (e.g. 1 s each) and perform some measurements.
The first 20 outcomes ('reports') are:
0, 0, 1, 0, 0, 0, 1, 2, 0, 0, 1, 1, 0, 4, 2, 0, 0, 0, 0, 1.



A counting experiment

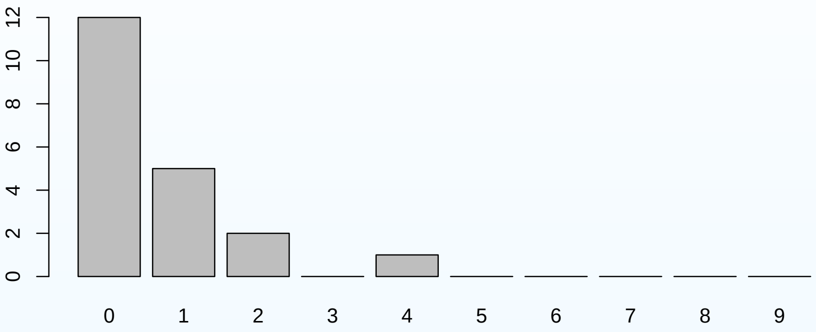
Imagine a small scintillation counter, with suitable threshold, placed here and now.
We set the measuring time (e.g. 1 s each) and perform some measurements.
The first 20 outcomes (‘reports’) are:
$0, 0, 1, 0, 0, 0, 1, 2, 0, 0, 1, 1, 0, 4, 2, 0, 0, 0, 0, 1$.



Think at the 21^{st} measurement/report:

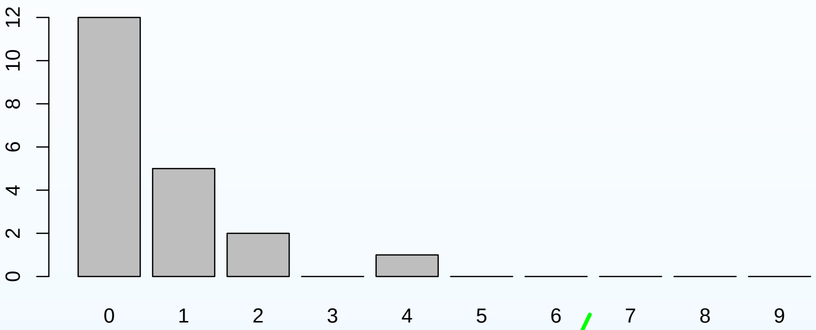
- Which outcome do you consider more likely? (0, 1, 2, 3, . . .)
- Why?

A counting experiment



$\Rightarrow$ Next ?

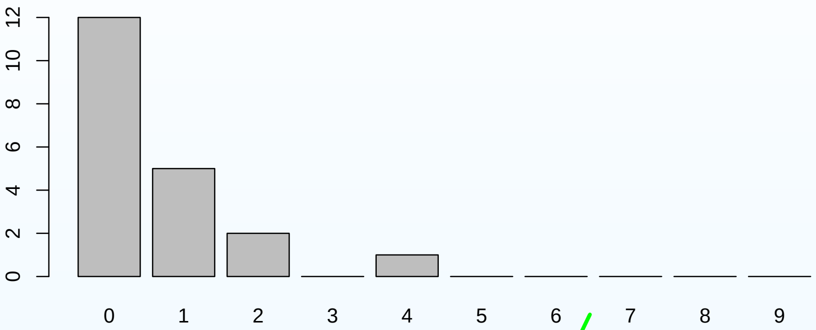
A counting experiment



$\Rightarrow$ Next ?

$P(0) > P(1) > P(2)$ $\surd$

A counting experiment

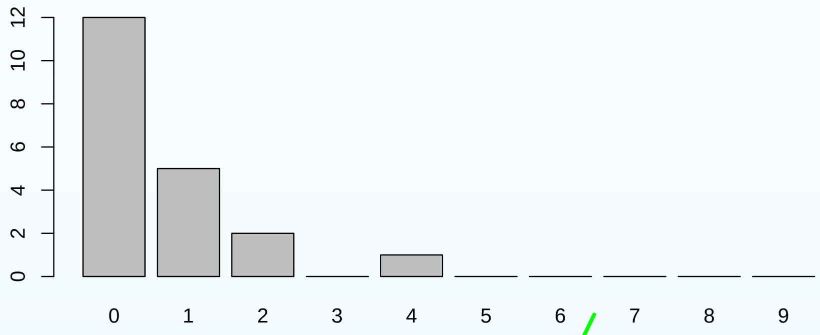


$\Rightarrow$ Next ?

$P(0) > P(1) > P(2)$ $\surd$

$P(3) < P(4)$, or $P(3) \geq P(4)$?

A counting experiment



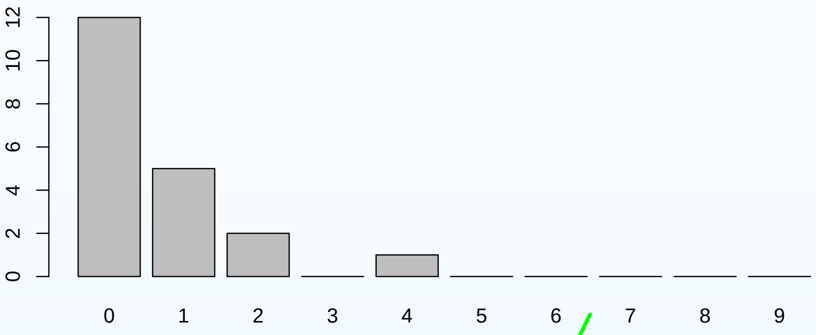
$\Rightarrow$ Next ?

$P(0) > P(1) > P(2)$ $\surd$

$P(3) < P(4)$, or $P(3) \geq P(4)$?

$P(3) = 0$, or $P(5) = 0$?

A counting experiment



$\Rightarrow$ Next ?

$P(0) > P(1) > P(2)$ ✓

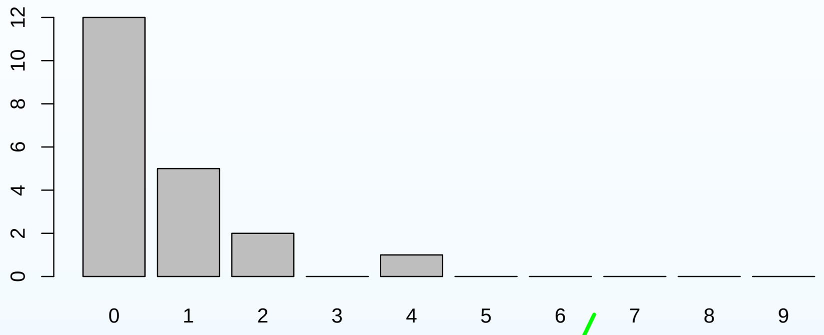
$P(3) < P(4)$, or $P(3) \geq P(4)$?

$P(3) = 0$, or $P(5) = 0$?

Not correct to say *"we cannot do it"*, or *"let us do other measurements and see"*:

> In real life we are asked to make assessments (and take decisions) with the information we have in hand NOW.
> If, later, the information changes, we can (must!) use the update one (and perhaps update our opinion).

A counting experiment



$\Rightarrow$ Next ?

$P(0) > P(1) > P(2)$ $\surd$

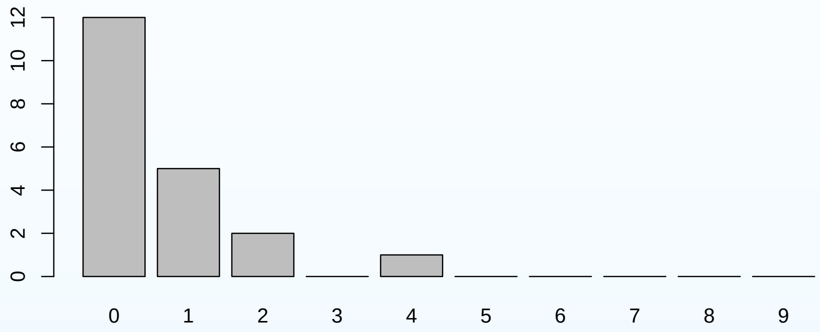
$P(3) < P(4)$, or $P(3) \geq P(4)$?

$P(3) = 0$, or $P(5) = 0$?

Not correct to say *"we cannot do it"*, or *"let us do other measurements and see"*:

$\Rightarrow$ But, obviously, IF we have time and money, we can take *other data*, gather other pieces of information about the system, plan other 'experiments' to understand it better. Or we wish to delay the decision, and so on, BUT this is not always the case

A counting experiment



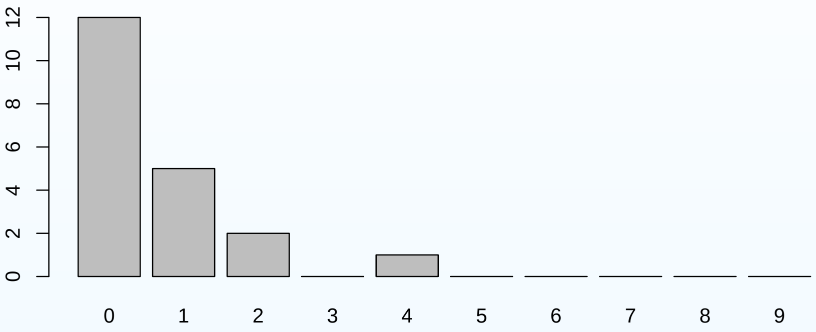
$\Rightarrow$ Next ?

We, as ‘experts’, tend to assess that

$$P(3) > P(4) \text{ and } P(5) > 0$$

Why? Is this arbitrary? Should we only stick to ‘data’?

A counting experiment



$\Rightarrow$ Next ?

We, as 'experts', tend to assess that

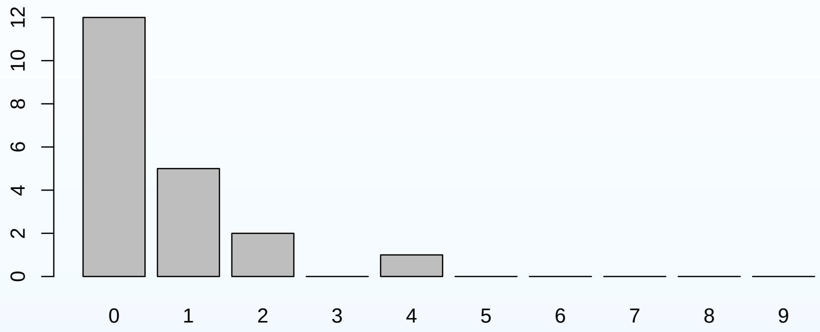
$$P(3) > P(4) \text{ and } P(5) > 0$$

Why? Is this arbitrary? Should we only stick to 'data'?

Given our 'experience', 'education', 'mentality' (...)

We 'know' / 'assume' / 'hope' / 'guess' / 'postulate' regularity of nature

A counting experiment



$\Rightarrow$ Next ?

We, as 'experts', tend to assess that

$$P(3) > P(4) \text{ and } P(5) > 0$$

Why? Is this arbitrary? Should we only stick to 'data'?

Given our 'experience', 'education', 'mentality' (...)

'know'

'assume'

We 'hope' regularity of nature

'guess'

'postulate'

'believe'

A philosopher, physicist and mathematician joke

A philosopher, a physicist and a mathematician travel by train through Scotland.

The train is going slowly and they see a cow walking along a country road parallel to the railway.

A philosopher, physicist and mathematician joke

A philosopher, a physicist and a mathematician travel by train through Scotland.

The train is going slowly and they see a cow walking along a country road parallel to the railway.

- Philosopher: *"In Scotland cows are black"*

A philosopher, physicist and mathematician joke

A philosopher, a physicist and a mathematician travel by train through Scotland.

The train is going slowly and they see a cow walking along a country road parallel to the railway.

- Philosopher: *“In Scotland cows are black”*
- Physicist: *“In Scotland there is at least a black cow”*

A philosopher, physicist and mathematician joke

A philosopher, a physicist and a mathematician travel by train through Scotland.

The train is going slowly and they see a cow walking along a country road parallel to the railway.

- Philosopher: *“In Scotland cows are black”*
- Physicist: *“In Scotland there is at least a black cow”*
- Mathematician: *“In Scotland at least a cow has a black side”*

A philosopher, physicist and mathematician joke

A philosopher, a physicist and a mathematician travel by train through Scotland.

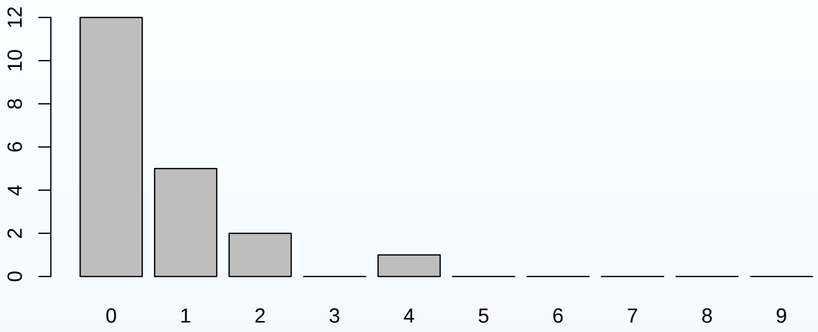
The train is going slowly and they see a cow walking along a country road parallel to the railway.

- Philosopher: *“In Scotland cows are black”*
- Physicist: *“In Scotland there is at least a black cow”*
- Mathematician: *“In Scotland at least a cow has a black side”*

Statements about reality have plenty of tacit – mostly very reasonable! — assumptions that derive from experience and rationality.

$\Rightarrow$ We constantly use theory/models to link past and future!.

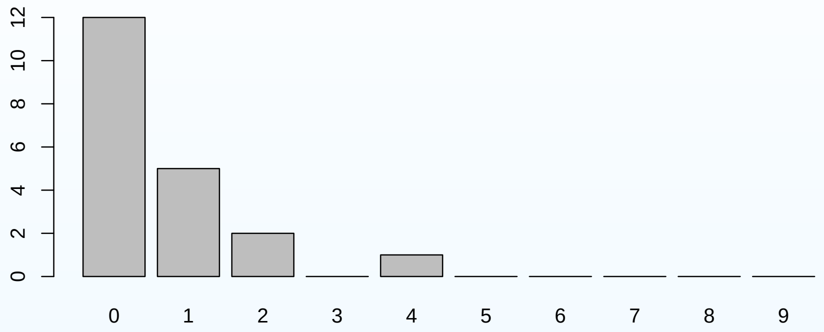
Transferring past to future



$\Rightarrow$ Next ?

Basic reasoning: assuming regularity of nature and a regular flow from the past to the future, we tend to believe that the effects that happened more frequently in the past will also occur more likely in the future.

Transferring past to future



$\Rightarrow$ Next ?

Basic reasoning: assuming regularity of nature and a regular flow from the past to the future, we tend to believe that the effects that happened more frequently in the past will also occur more likely in the future.

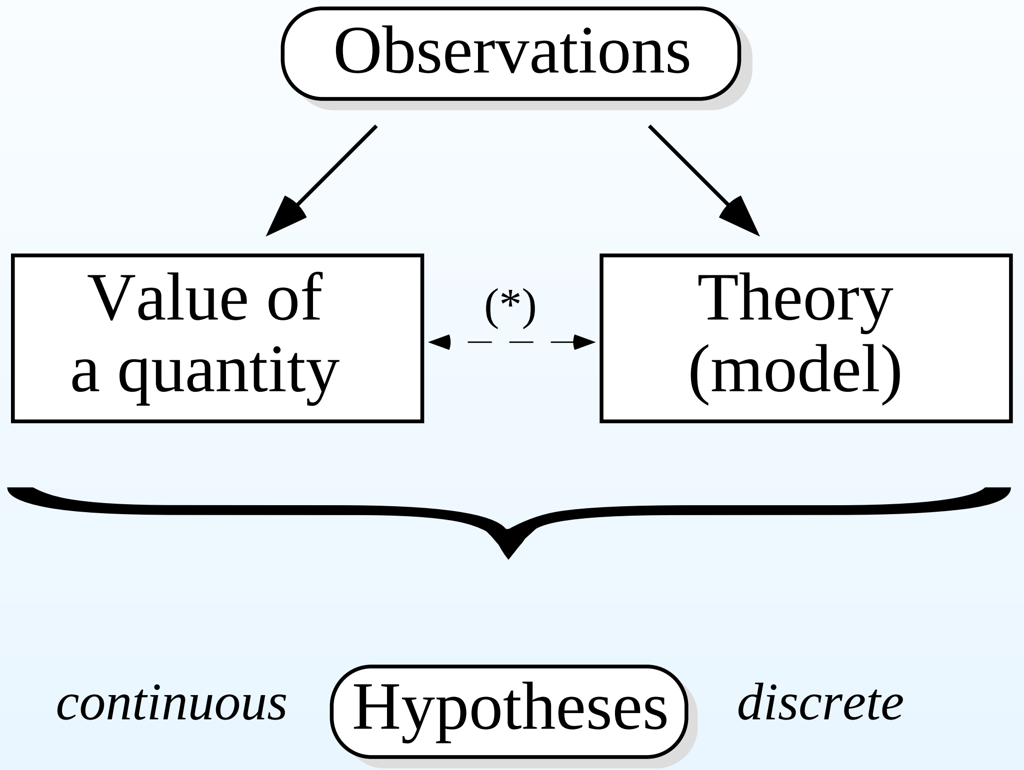
We ‘physicists’ (all experts about matters of fact) tend to filter the process of transferring the past to the future by ’laws’.

$\Rightarrow$ an experimental histogram shows a relative-frequency distribution, and not a probability distribution!

Relative frequencies *might* become probabilities, but only after they have been processed by our mind:

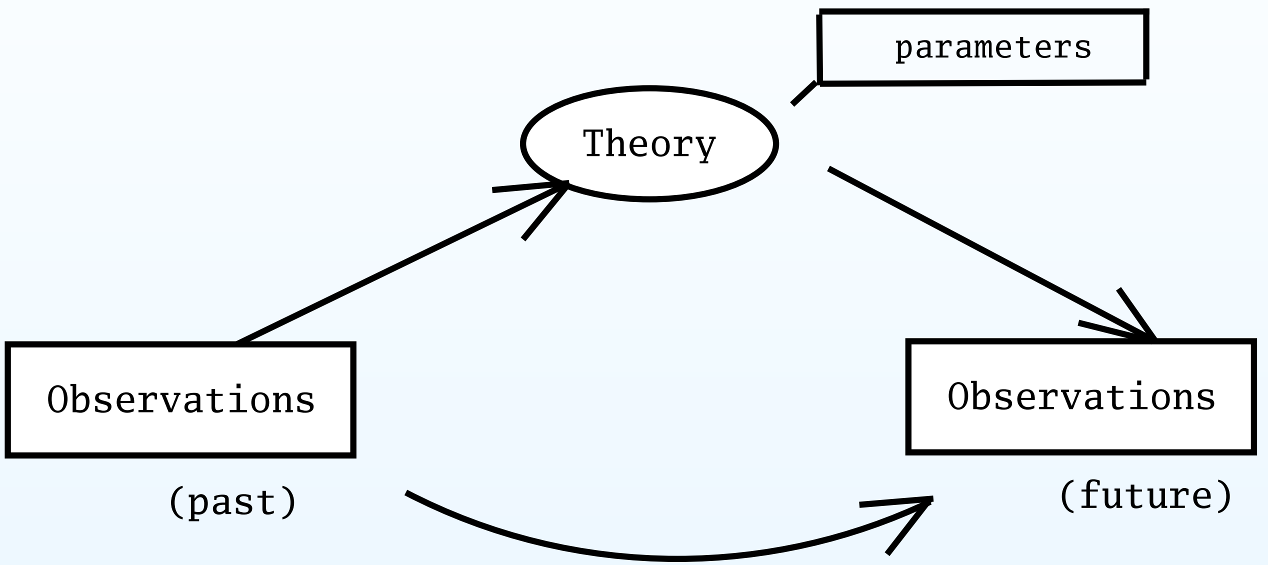
$\Rightarrow$ models, prior knowledge, analogy, etc.

Physics



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

From past to future



Task of the physicist:

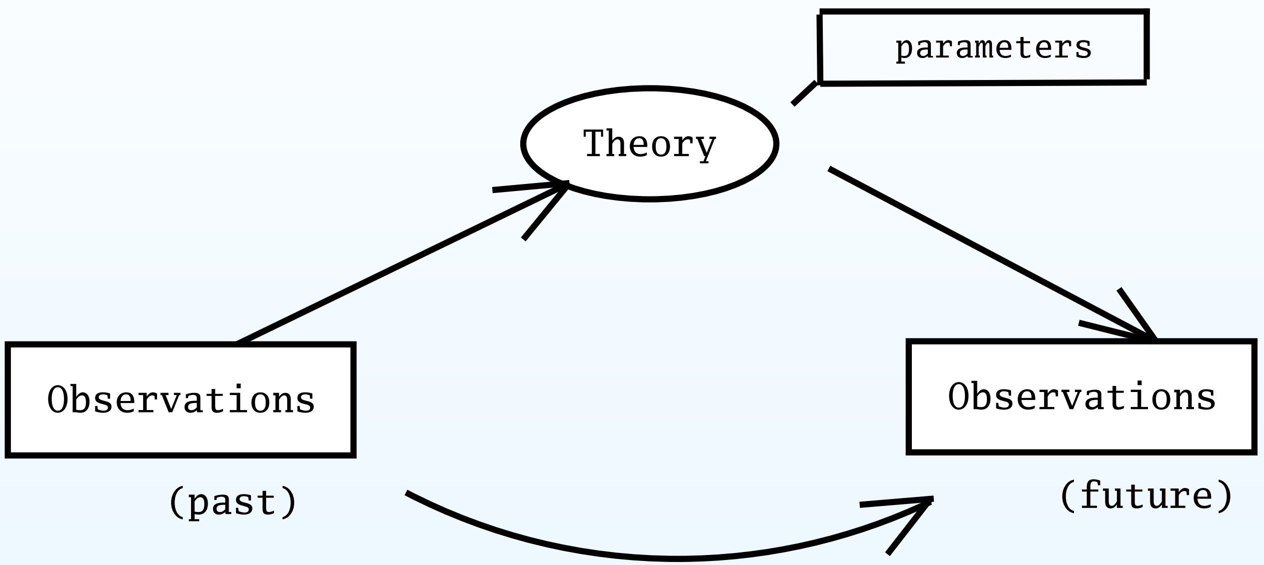
- Describe/understand the physical world

  $\Rightarrow$ inference of laws and their parameters

- Predict observations

  $\Rightarrow$ forecasting

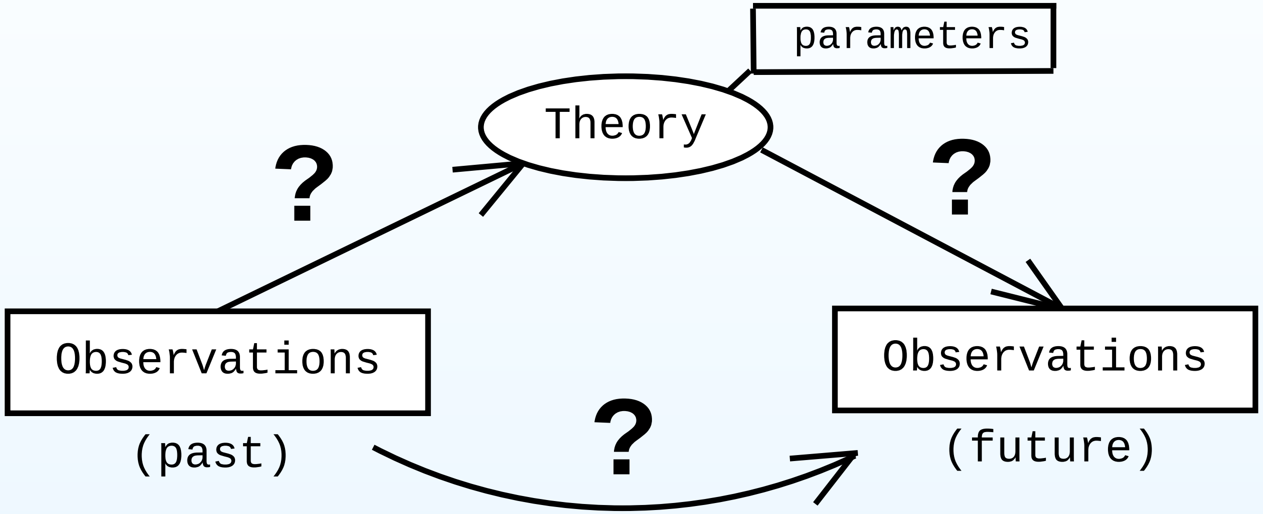
From past to future



Process

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

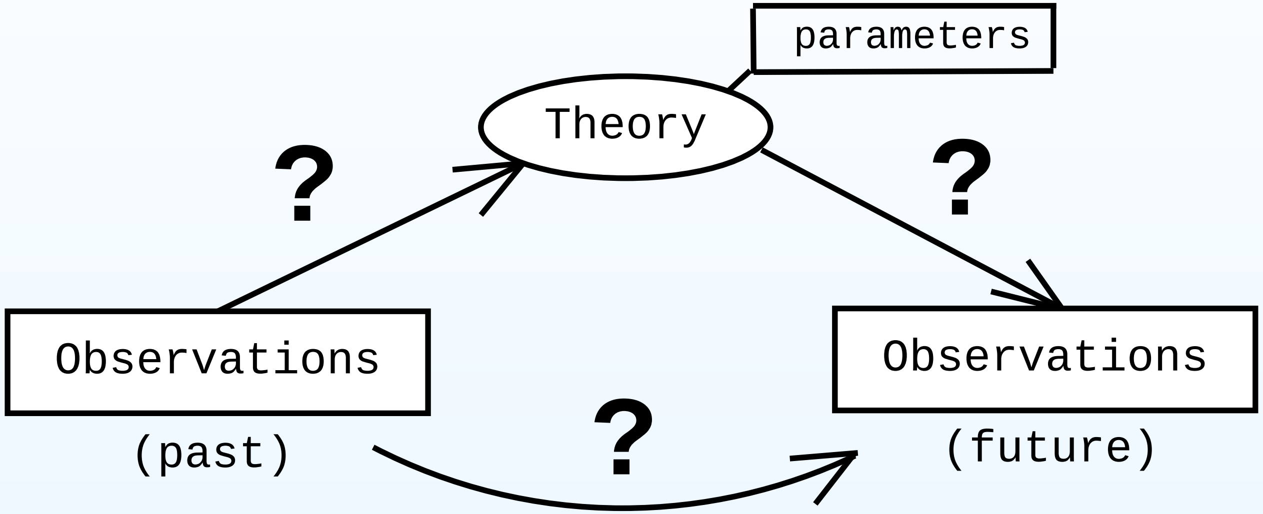
From past to future



⇒ Uncertainty:

1. Given the past observations, in general we are not sure about the theory 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.

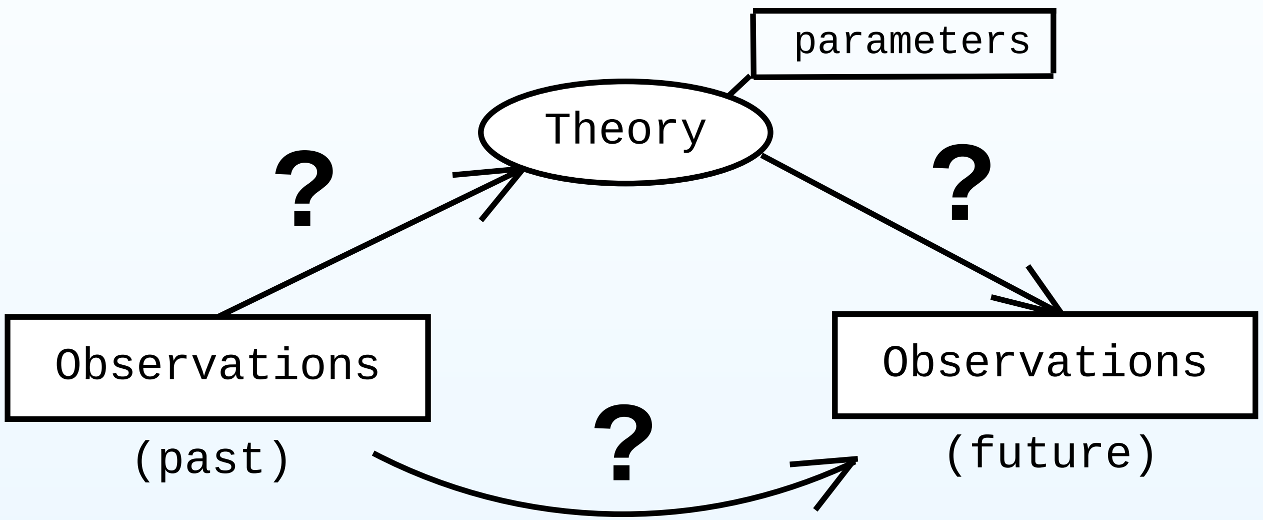
From past to future



$\Rightarrow$ 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

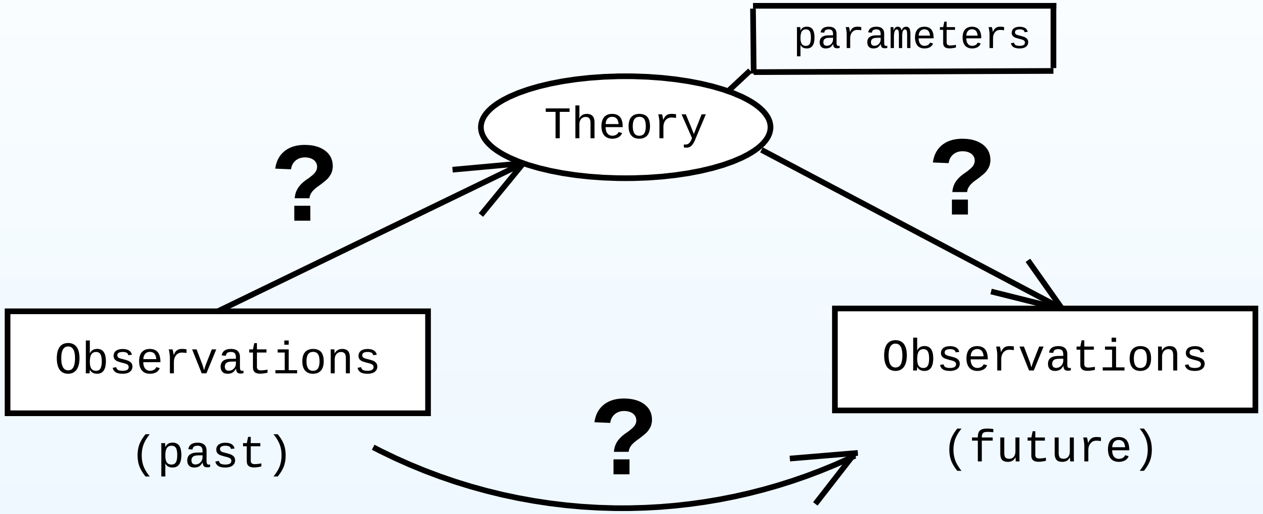
Remember:

“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)

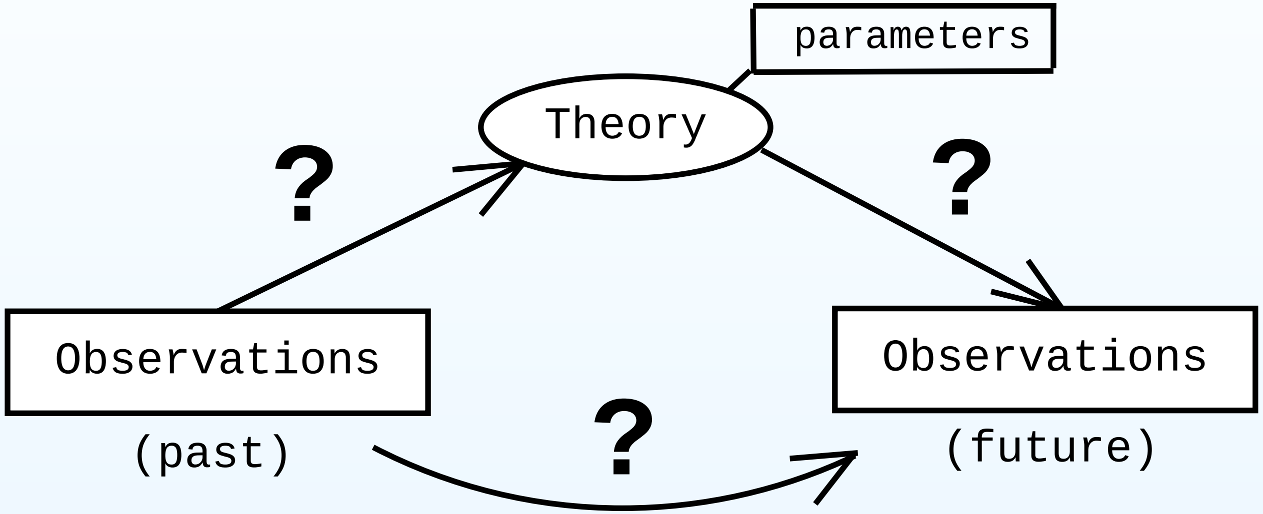
Deep source of uncertainty



Uncertainty:

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

Deep source of uncertainty



Uncertainty:

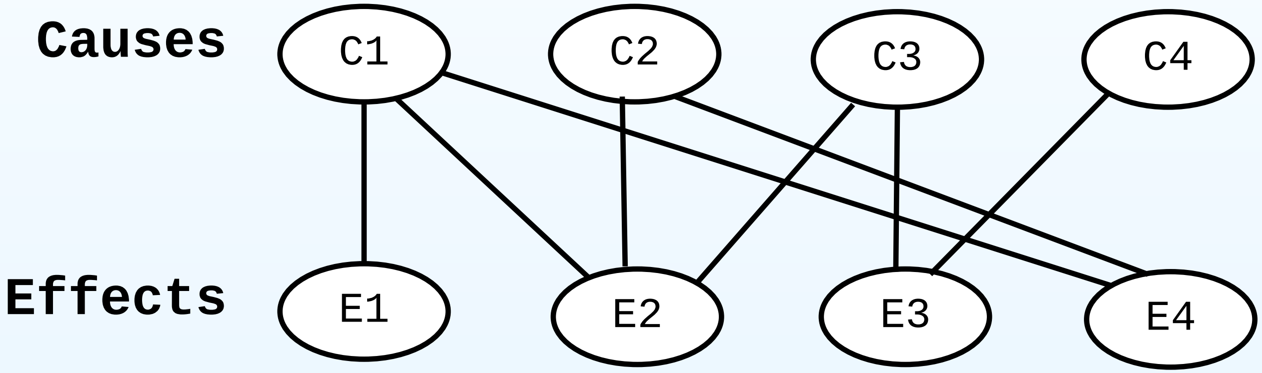
Theory	— ? $\longrightarrow$	Future observations
Past observations	— ? $\longrightarrow$	Theory
Theory	— ? $\longrightarrow$	Future observations

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

CAUSE $\Longleftrightarrow$ EFFECT

Causes $\longrightarrow$ effects

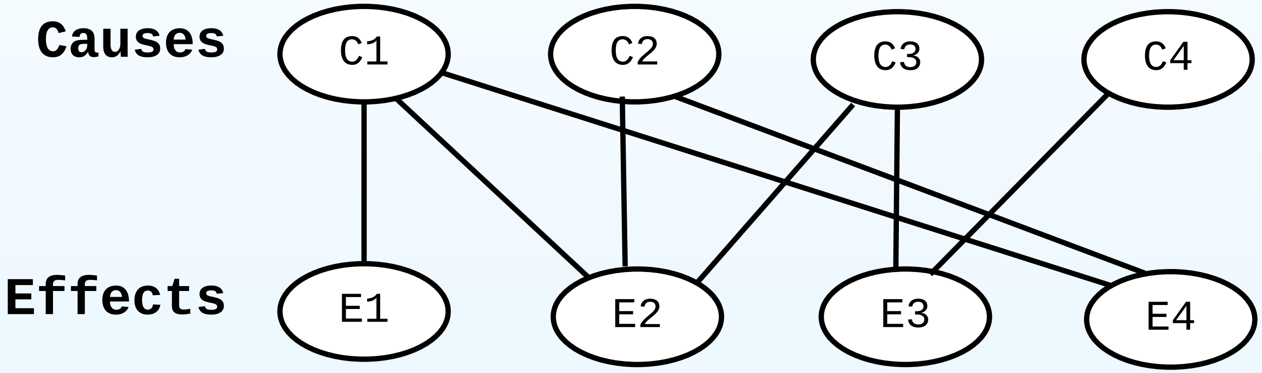
The same *apparent* cause might produce several,different effects



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

Causes $\longrightarrow$ effects

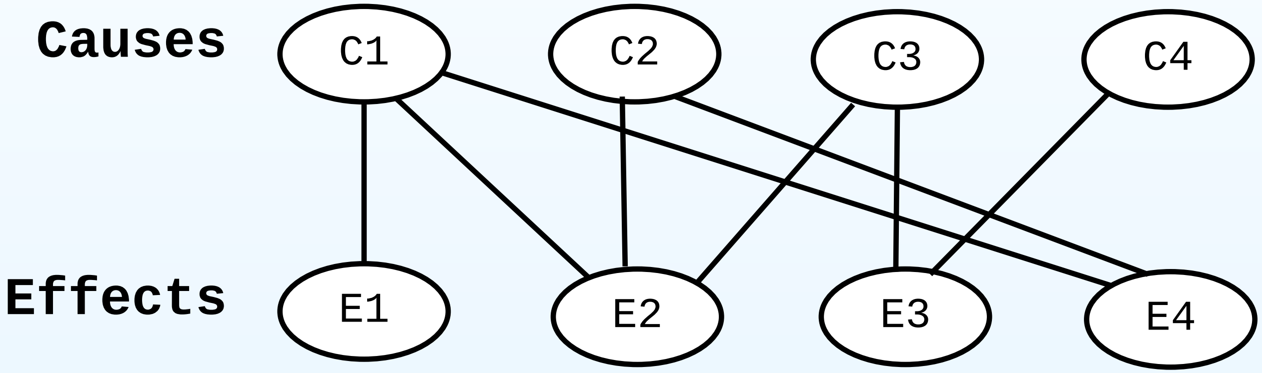
The same *apparent* cause might produce several,different effects



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."

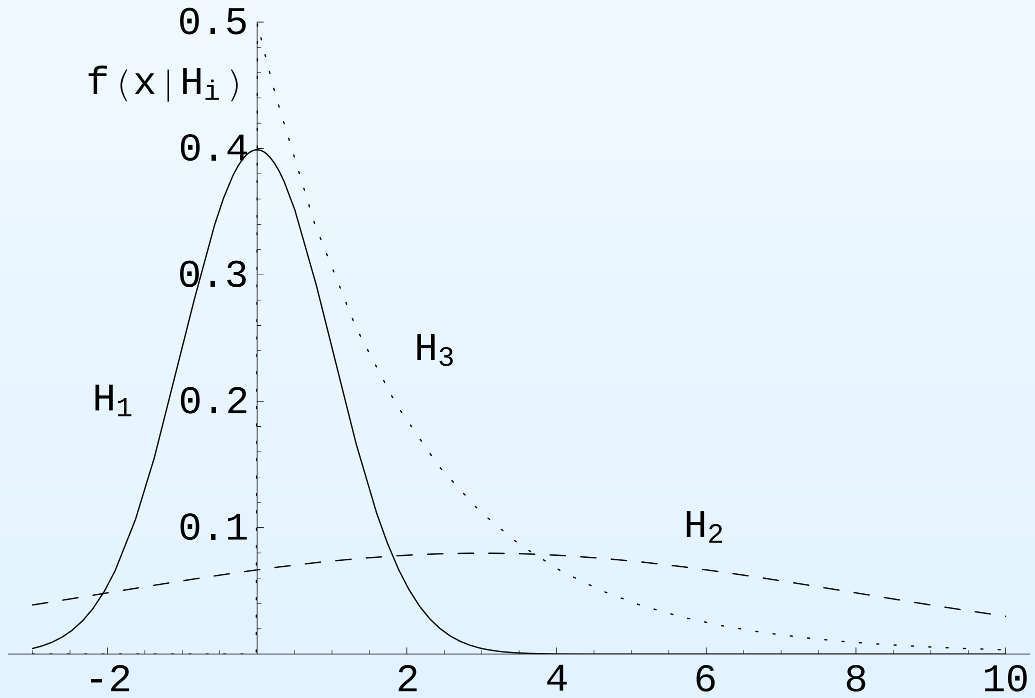
(H. Poincaré – *Science and Hypothesis*)

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"

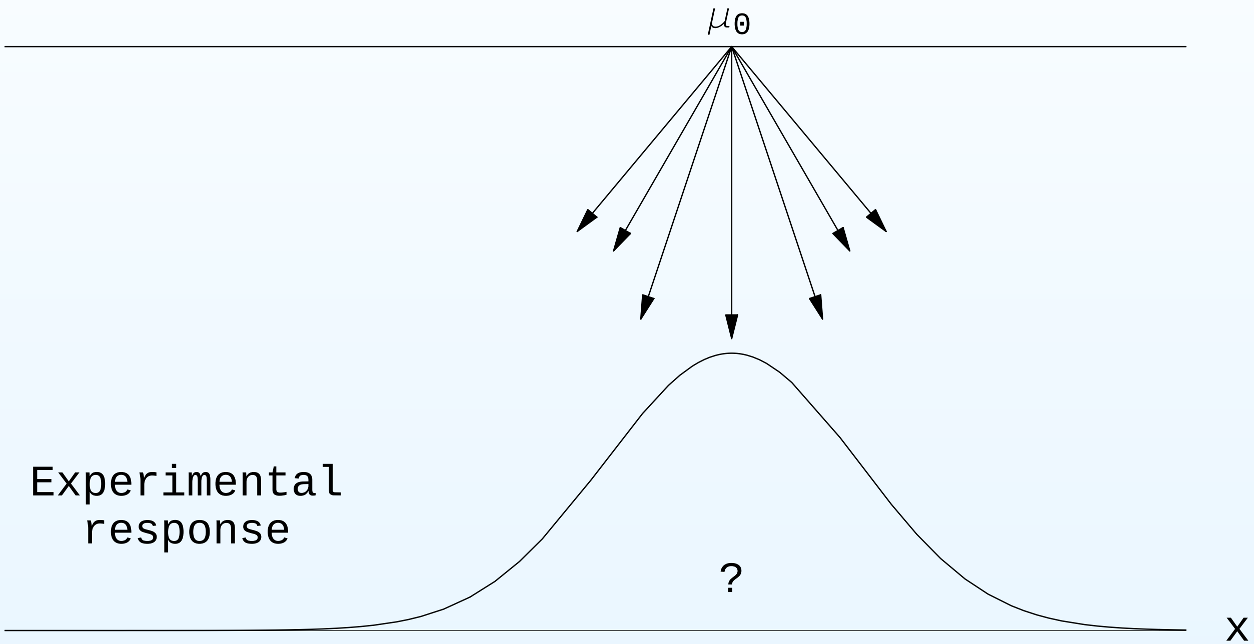
A numerical example

- 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

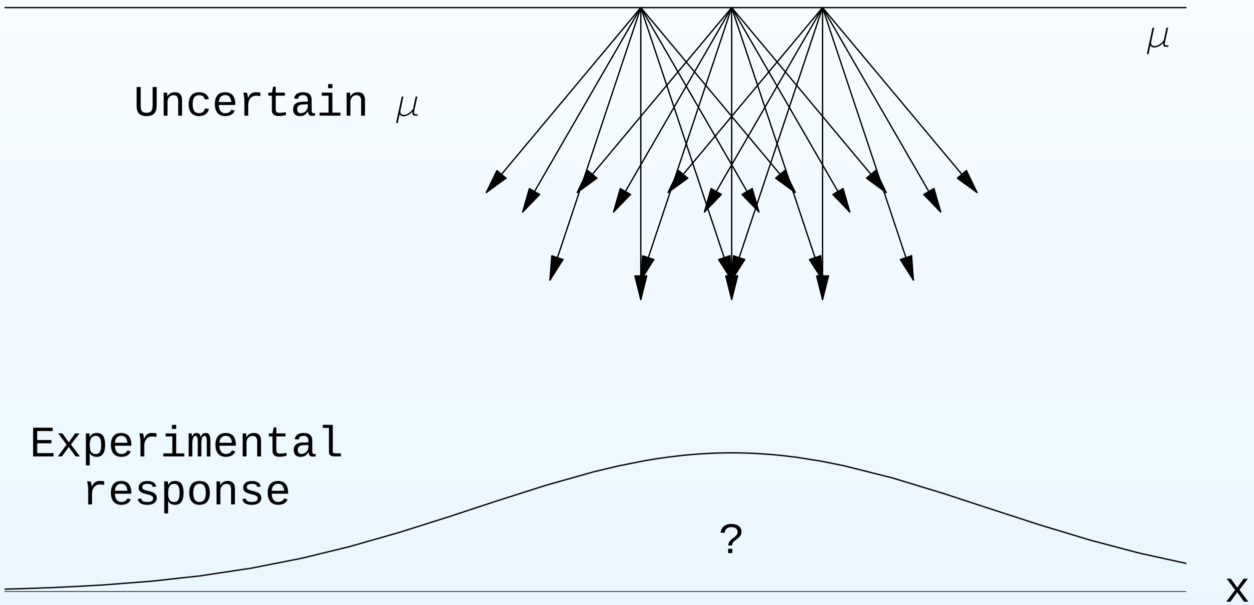
probability.

From ‘true value’ to observations



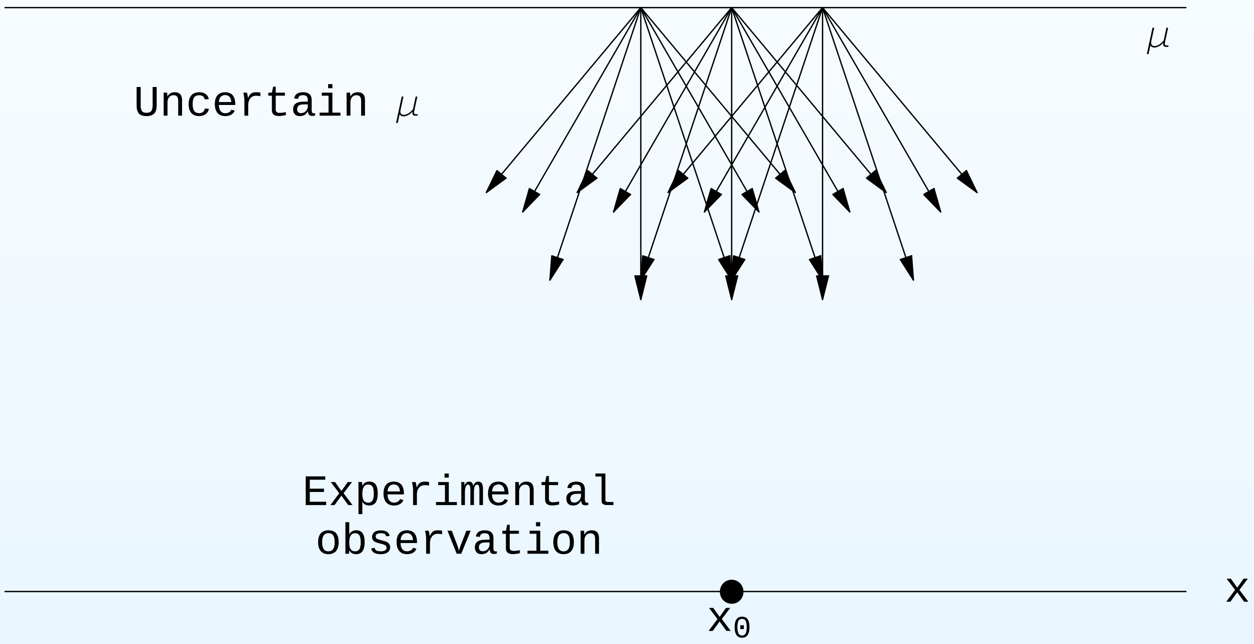
Given μ (exactly known) we are uncertain about x

From ‘true value’ to observations



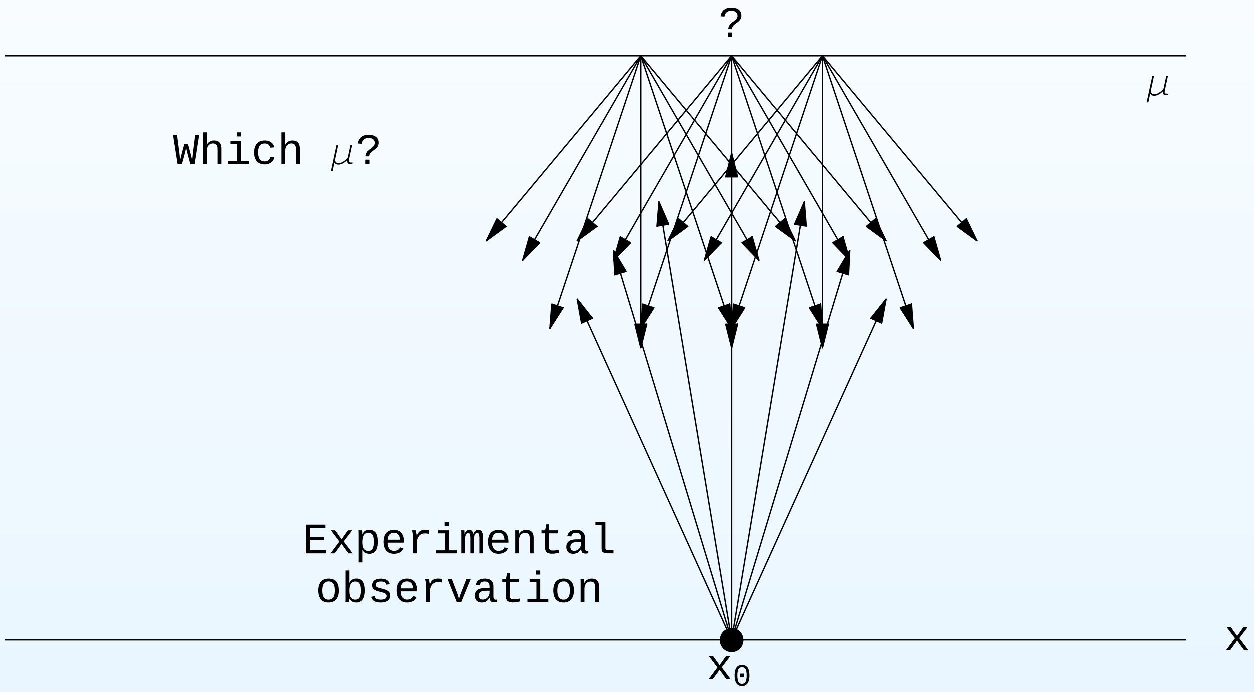
Uncertainty about μ makes us more uncertain about x

Inferring a true value



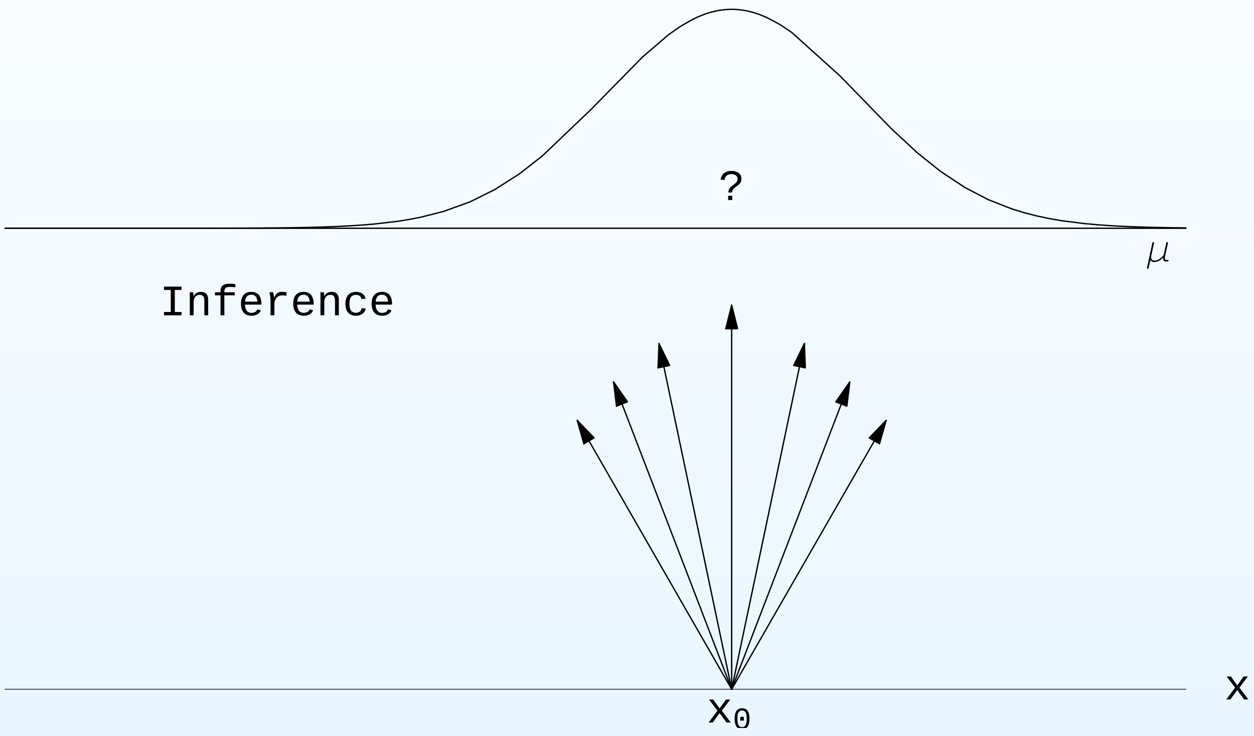
The observed data is <u>certain</u>: $\rightarrow$ ‘true value’ uncertain.

Inferring a true value



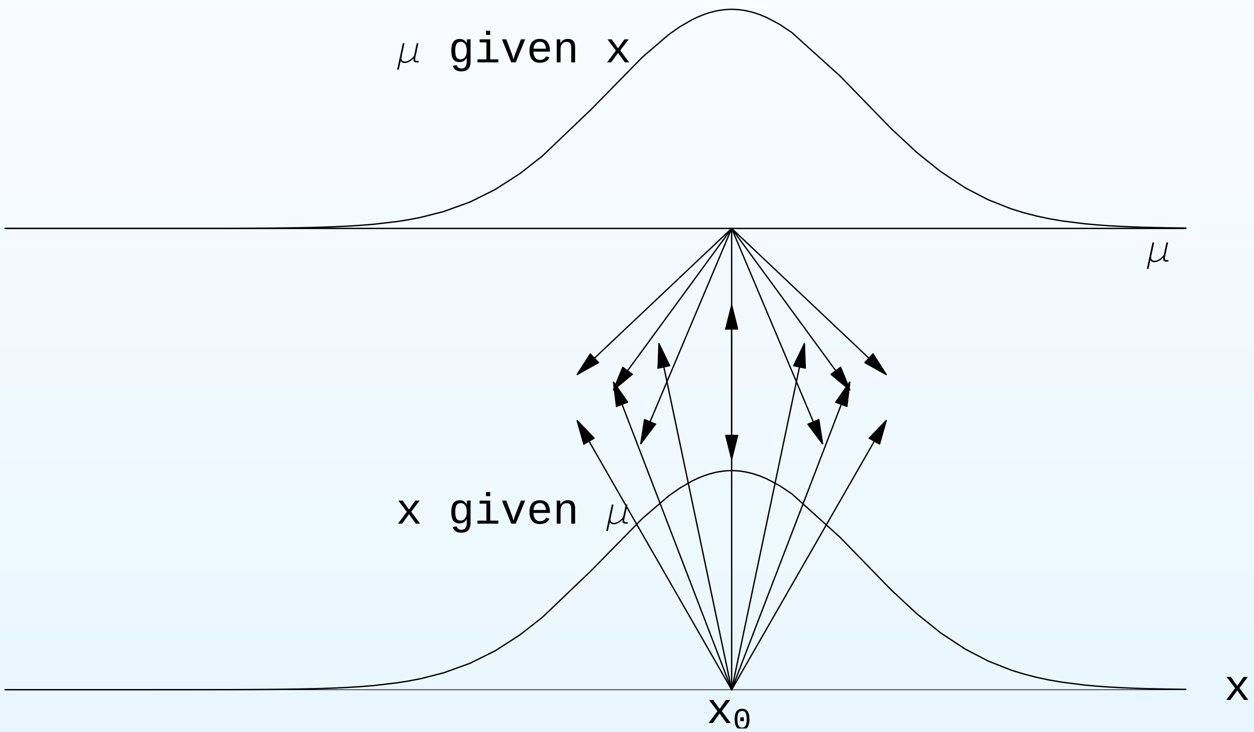
Where does the observed value of x comes from?

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

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)

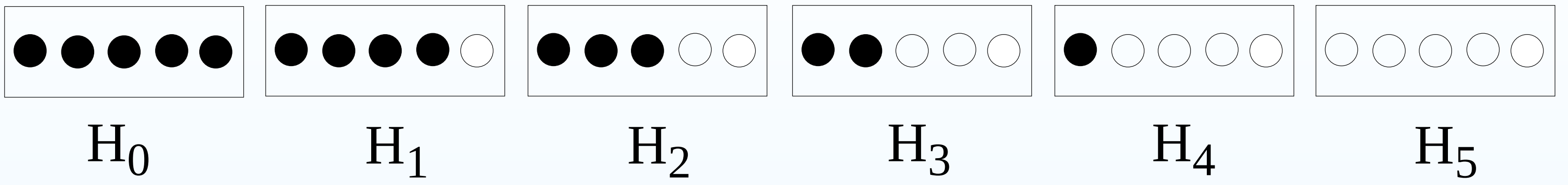
Doing Science in conditions of uncertainty

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

Indeed

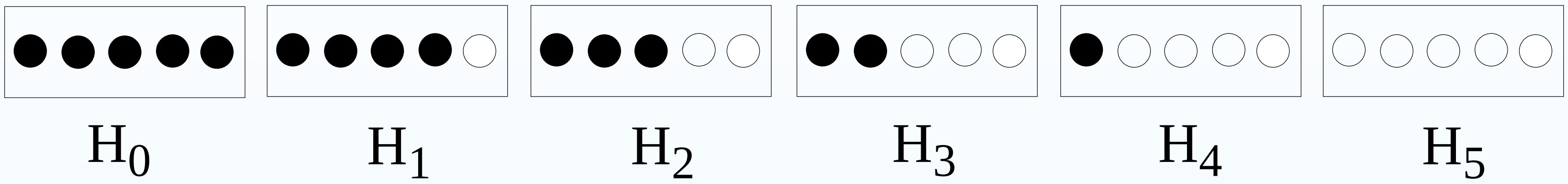
"It is scientific only to say what is more likely and what is less likely" (Feynman)

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:

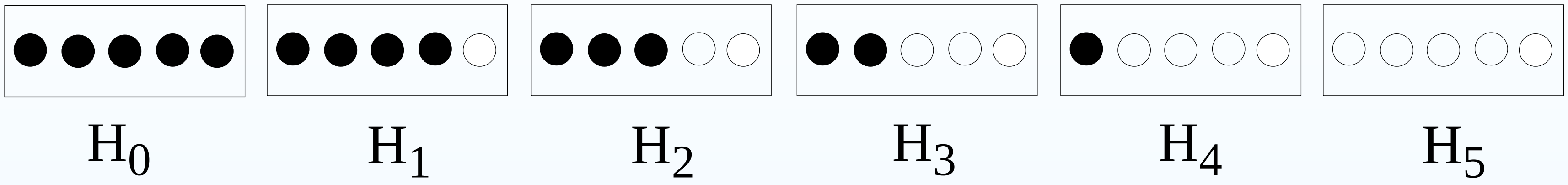
(a) Which box have we chosen, $H_0, H_1, \ldots, H_5$?

(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



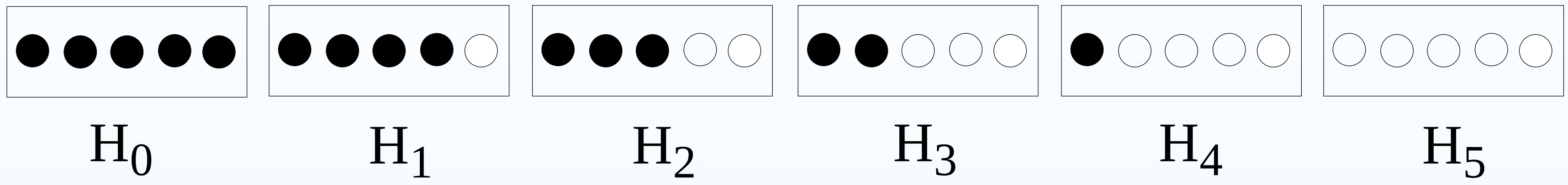
Let us take randomly one of the boxes.
We are in a state of uncertainty concerning several *events*, the most important of which correspond to the following questions:

(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)

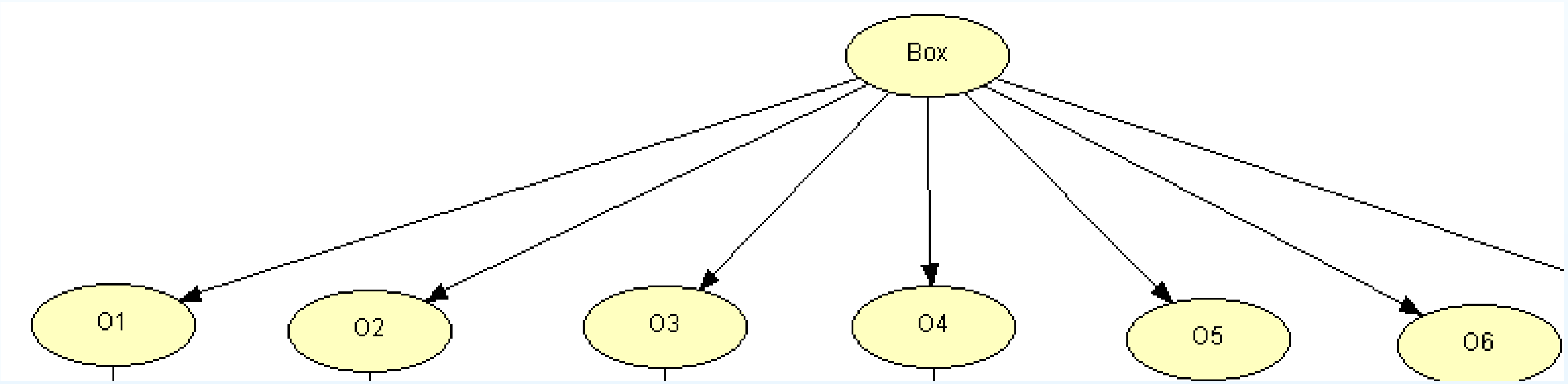
An interesting exercise

Probabilities of the 4 sequences from the first 3 extractions from the box of unknow composition:

- WW
- WB
- BW
- BB

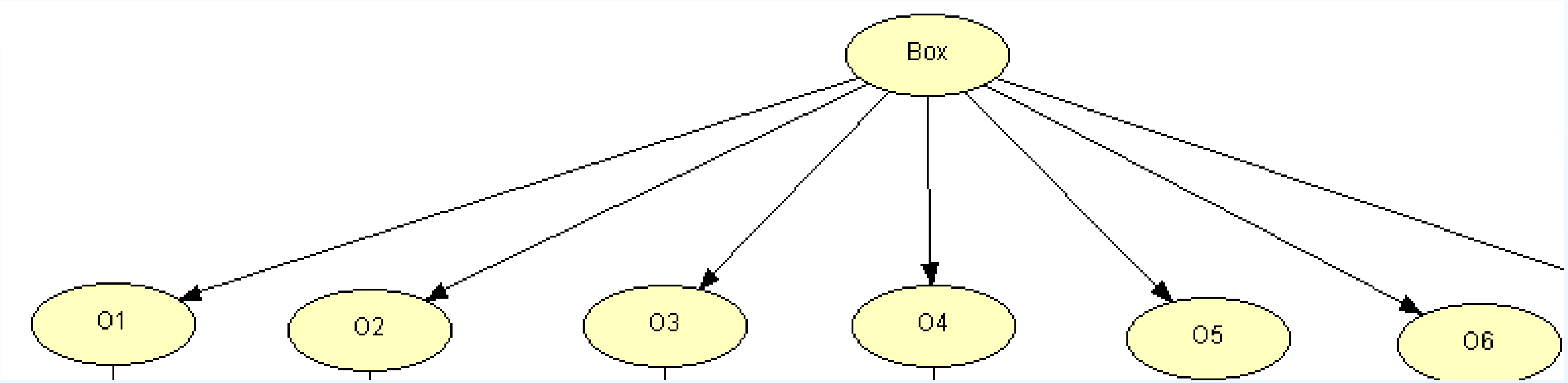
Cause-effect representation

box content $\rightarrow$ observed color



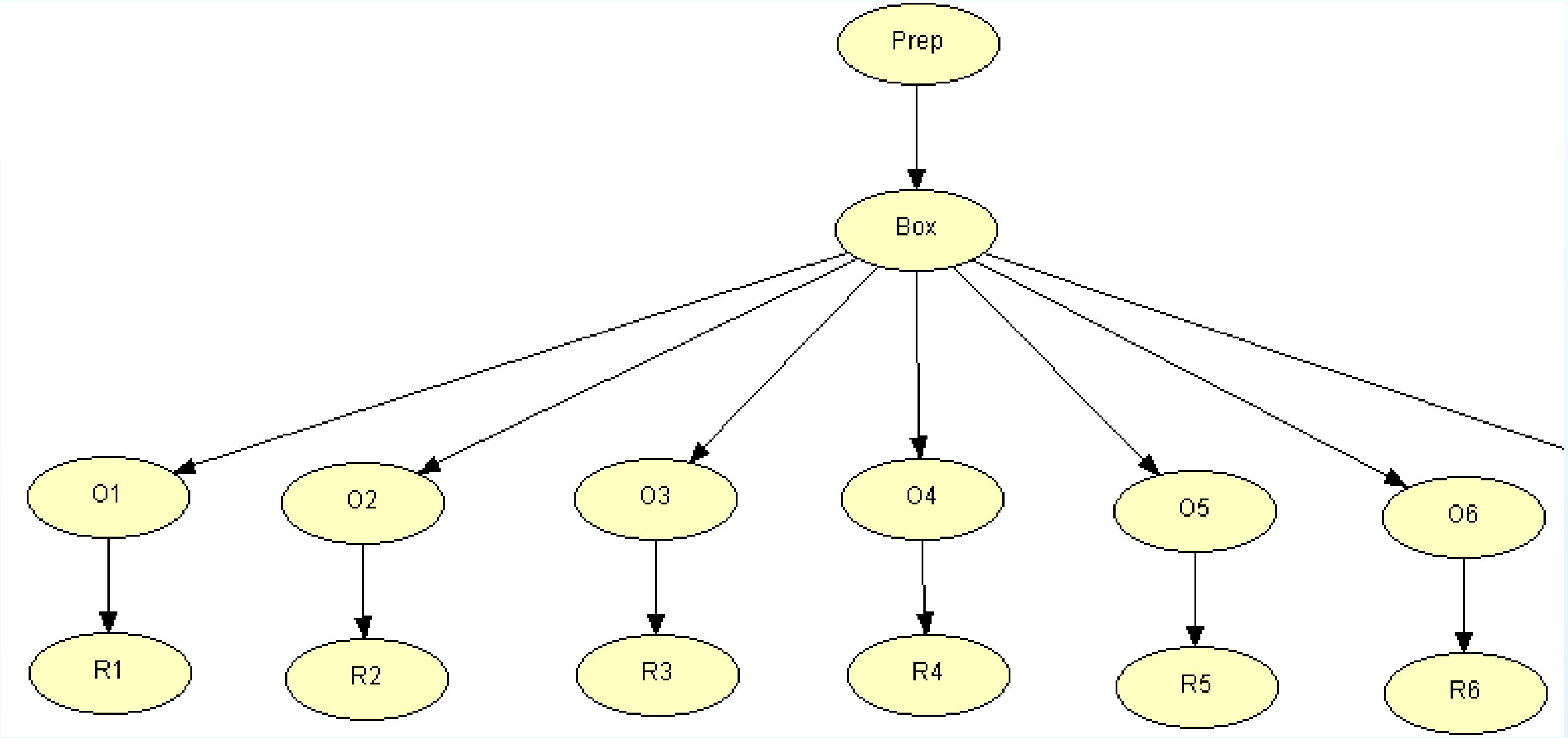
Cause-effect representation

box content $\rightarrow$ observed color

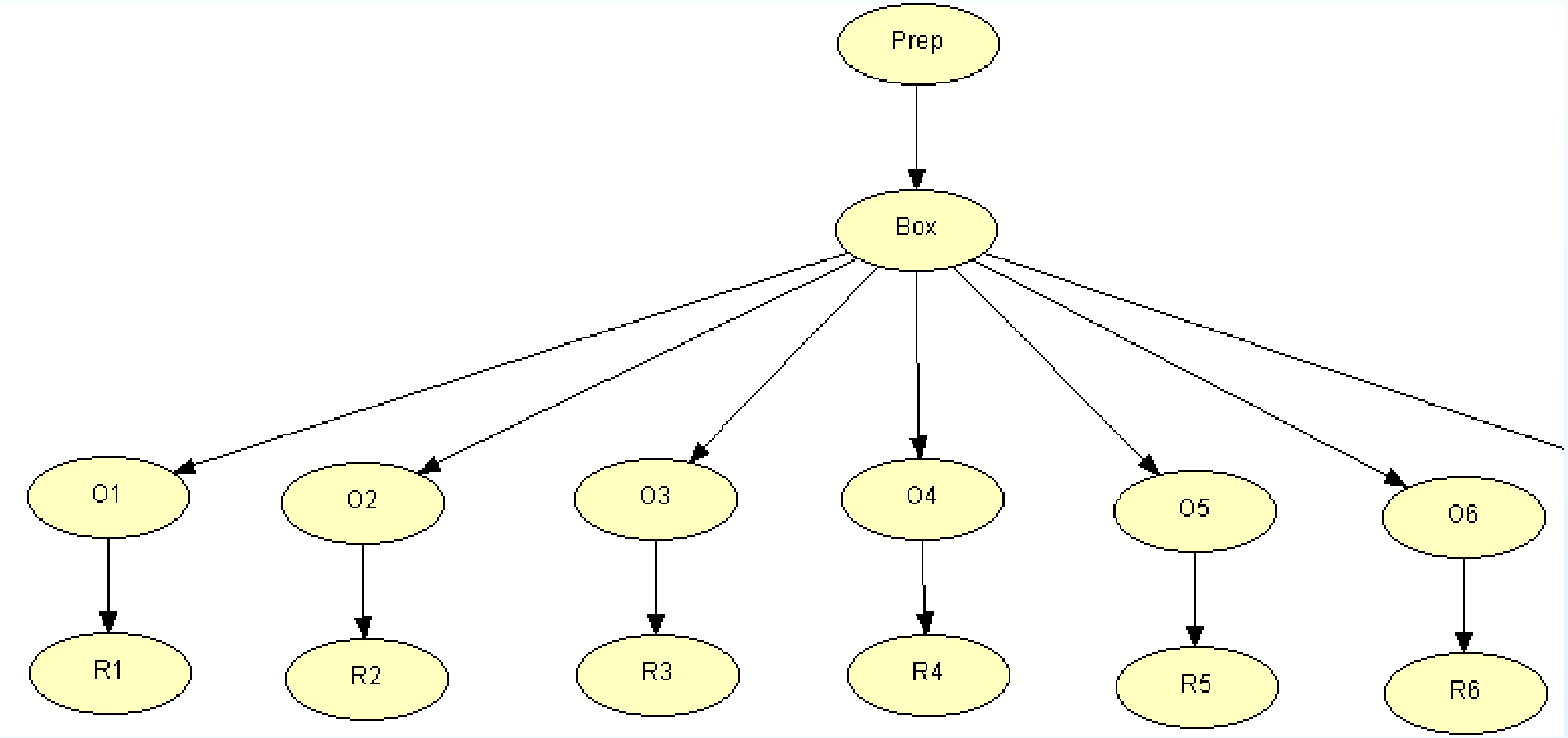


An effect might be the cause of another effect $\longrightarrow$

A network of causes and effects

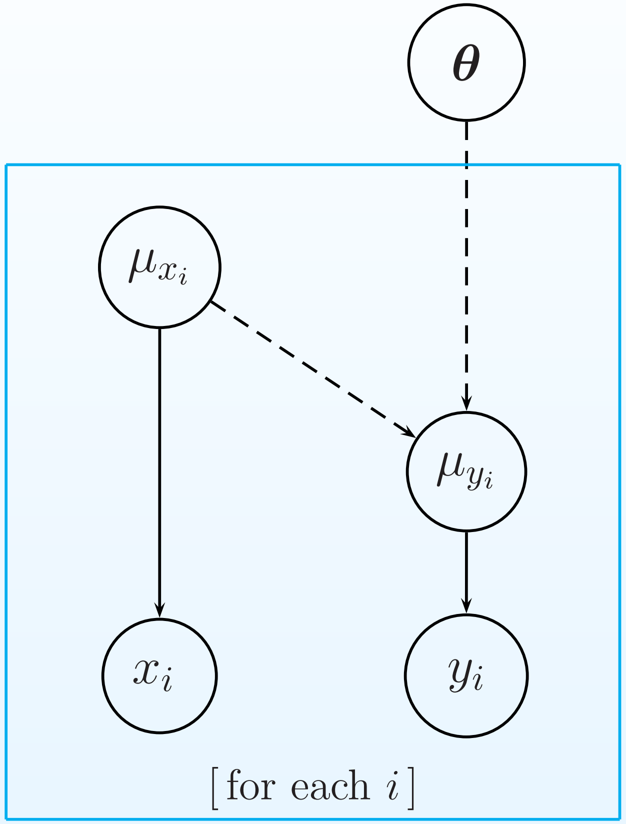


A network of causes and effects



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

A different way to view fit issues



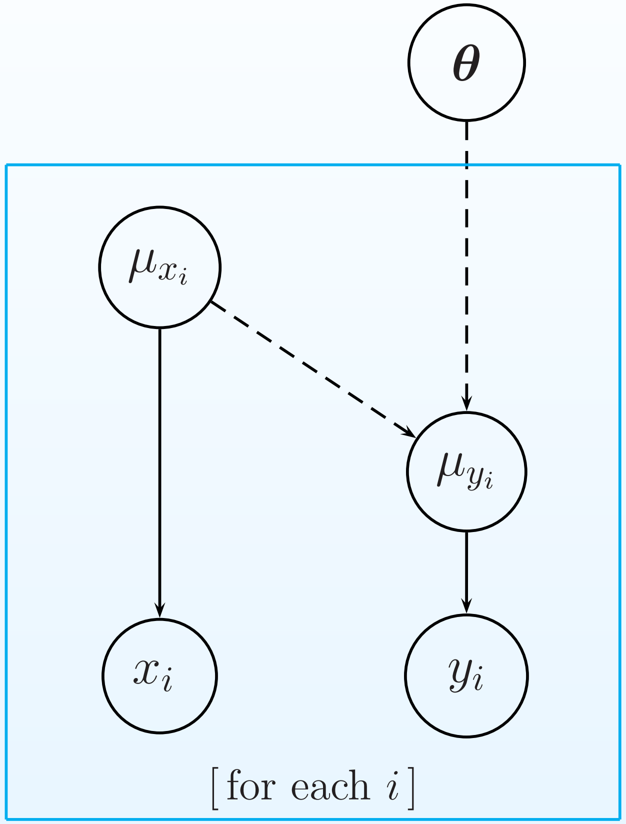
Determistic link μ_x's to μ_y's

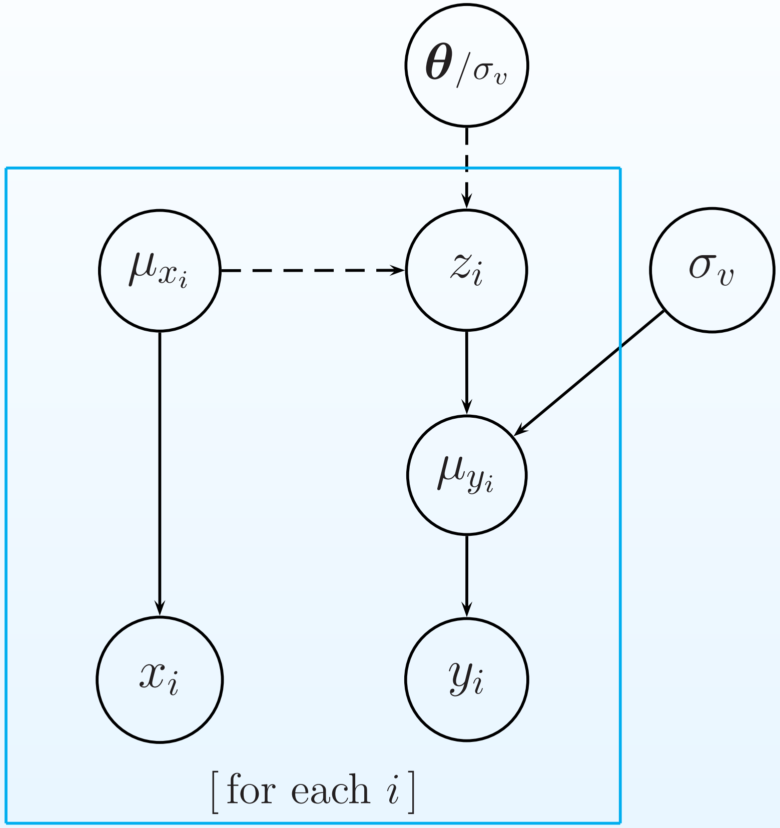
Probabilistic links $\mu_x \to x$, $\mu_y \to y$

(errors on both axes!)

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

A different way to view fit issues





Determistic link μ_x's to μ_y's

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

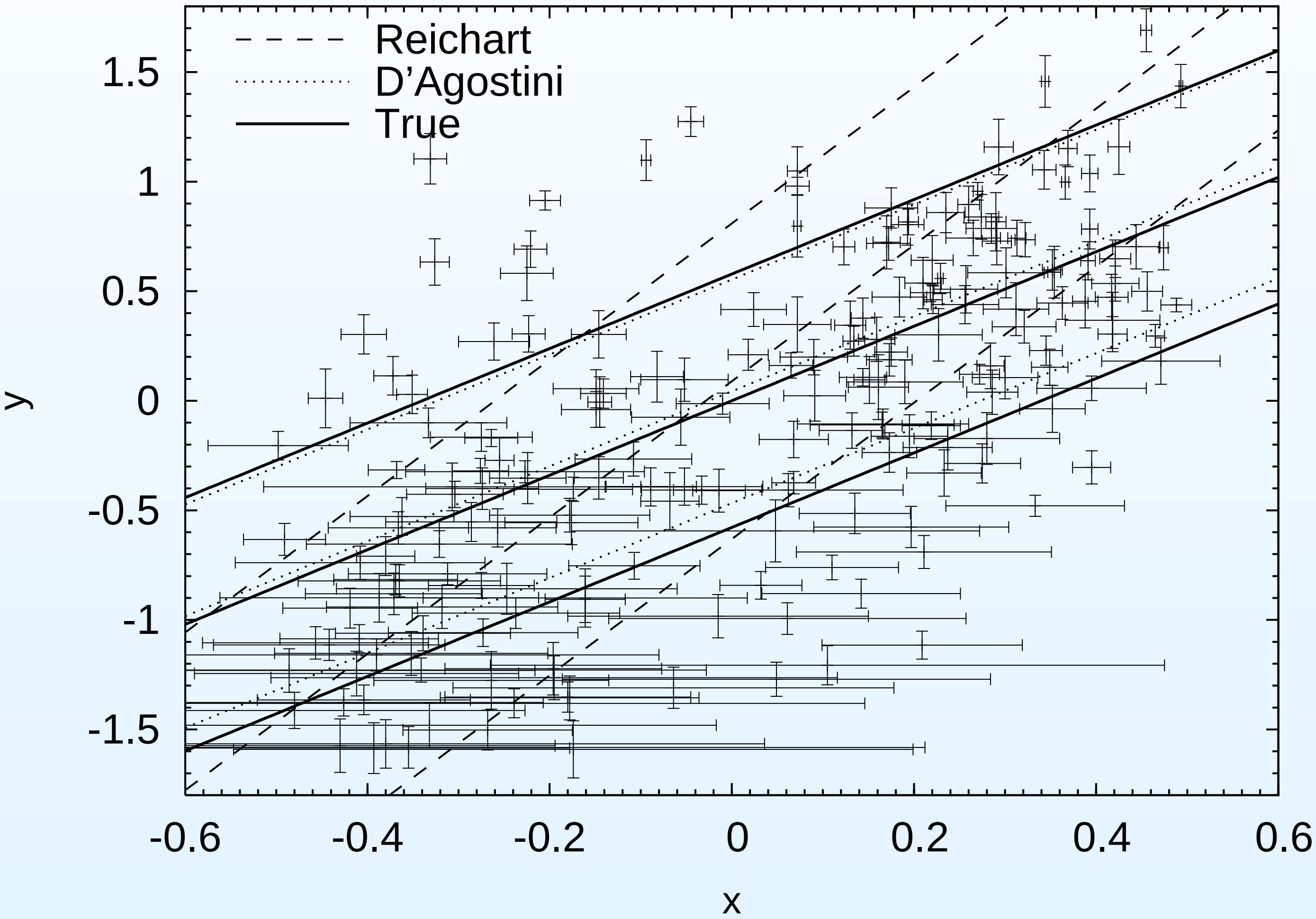
(errors on both axes!)

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

Extra spread of the data points

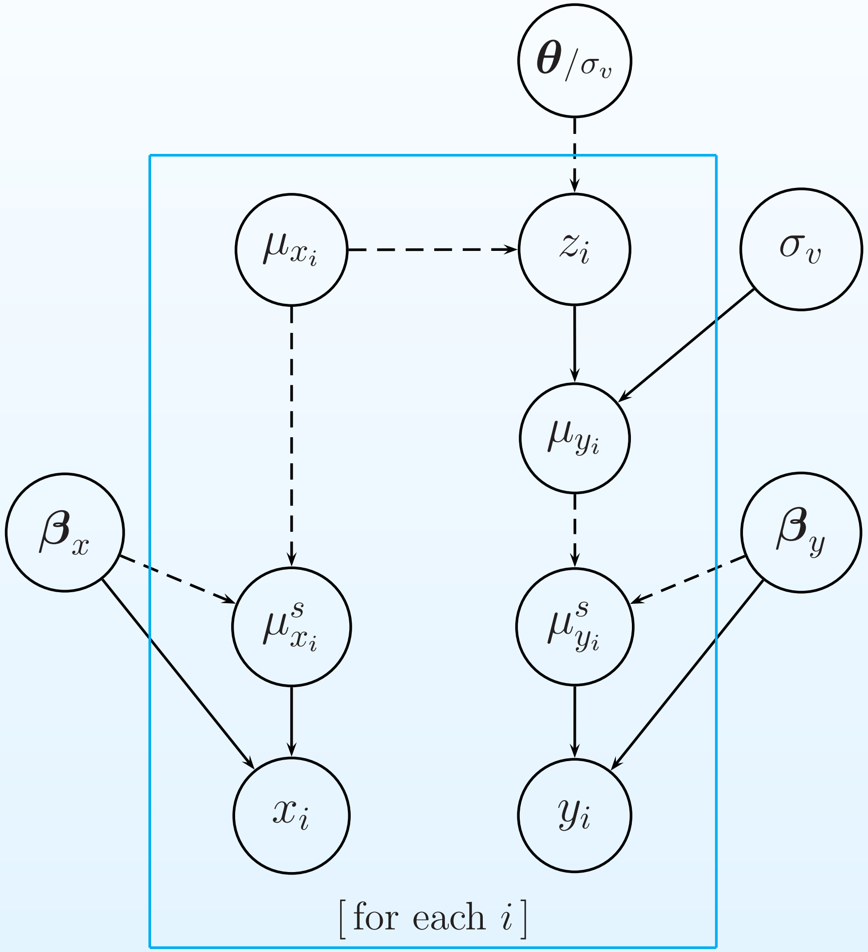
A different way to view fit issues

A physics case (from Gamma ray burts):



(Guidorzi et al., 2006)

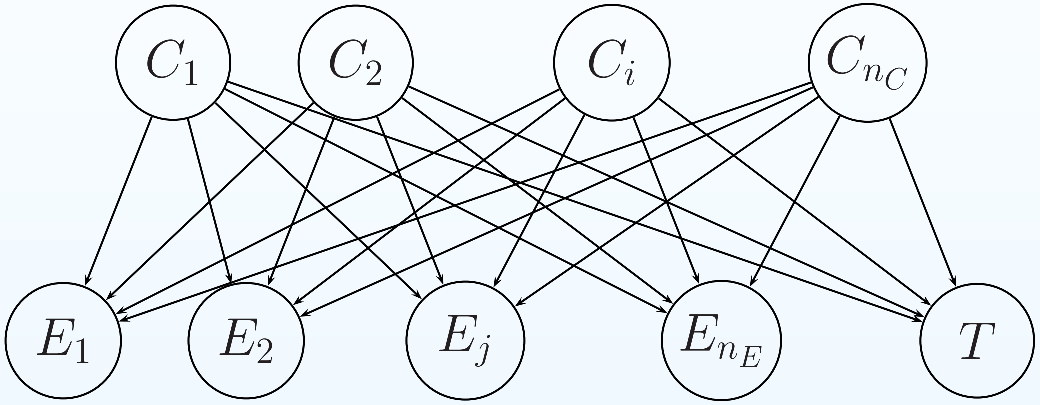
A different way to view fit issues



Adding systematics

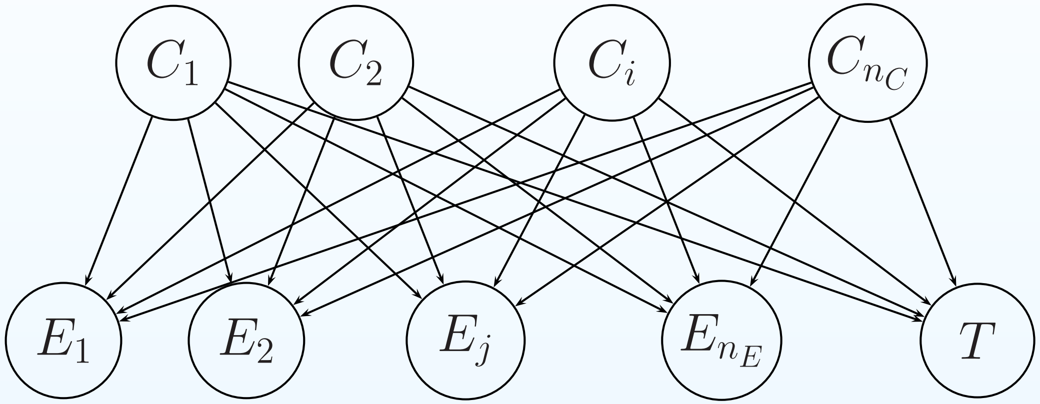
Unfolding a discretized spectrum

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

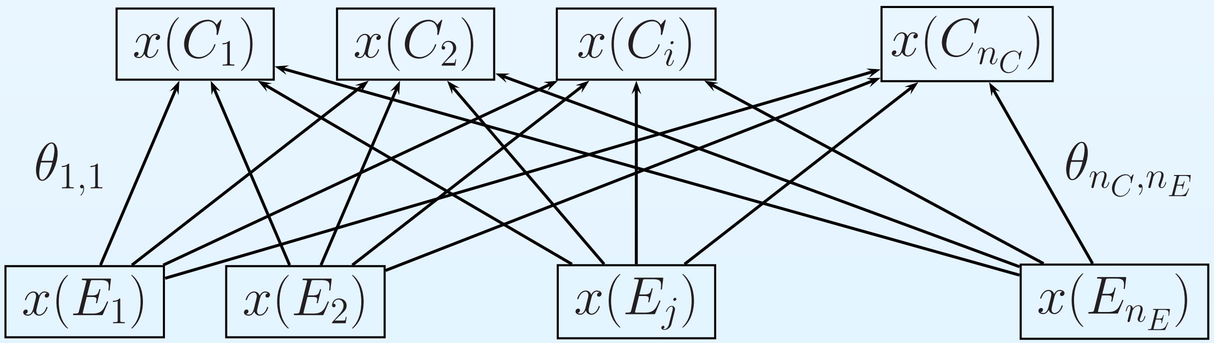


Unfolding a discretized spectrum

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

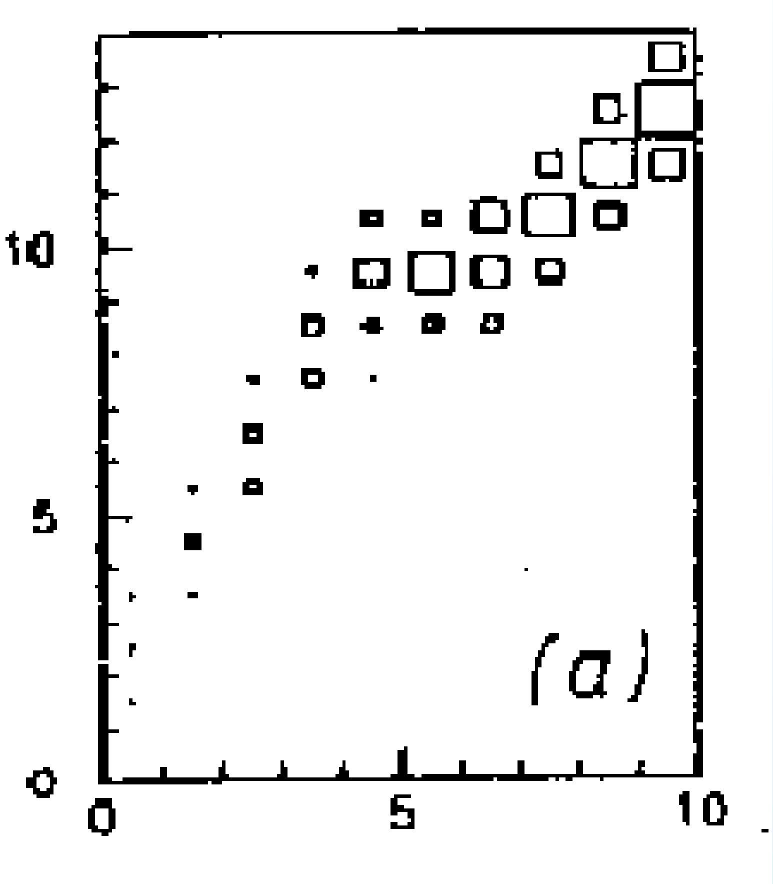


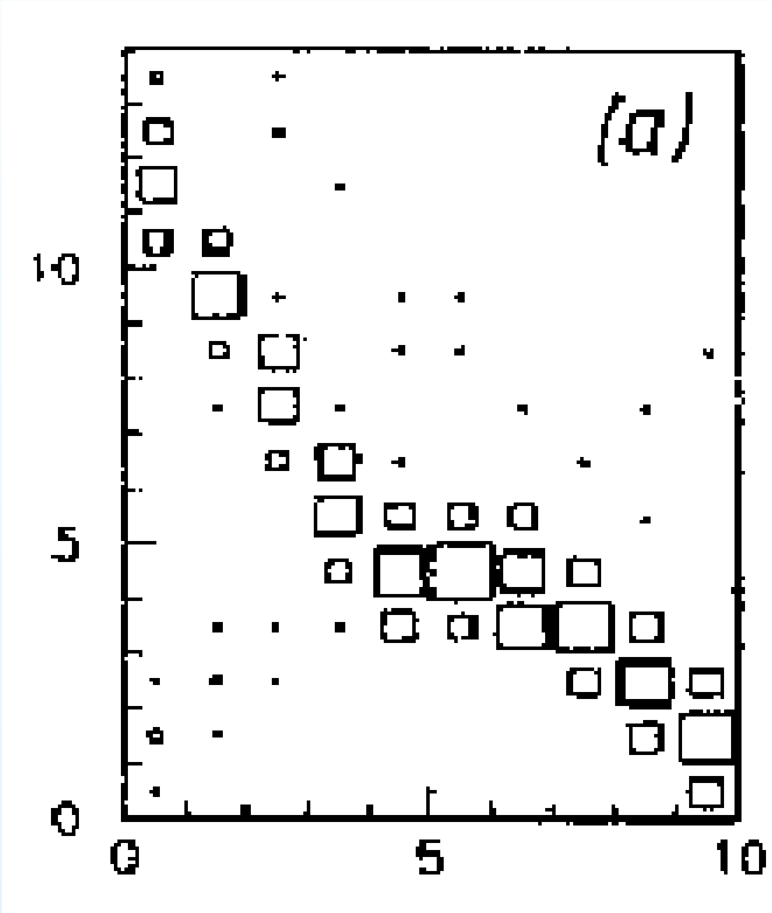
Sharing the observed events among the cause-bins



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** $C_i \not\rightarrow \; E$, 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:

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

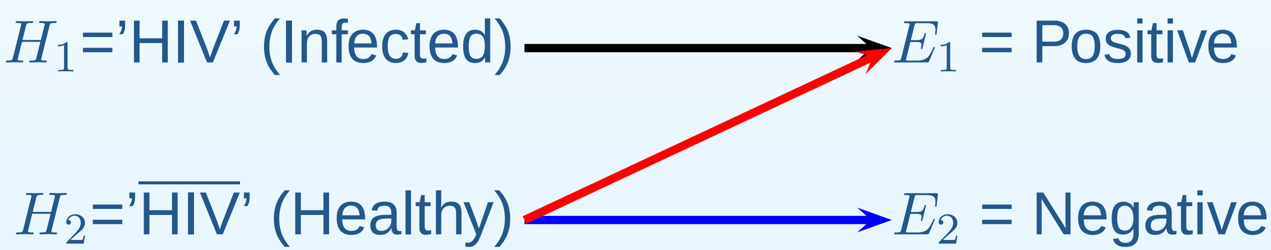
H_1='HIV' (Infected)	E_1 = Positive
H_2='$\overline{\text{HIV}}$' (Healthy)	E_2 = Negative

Example

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

Simplified model:

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



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...)

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 ⇔ 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!

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)

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

$\rightarrow$ 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)

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

  $\rightarrow$ no longer an excuse!

$\Rightarrow$ Use consistently probability theory

What to do?

$\Rightarrow$ 'Forward to past'

But benefitting of

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

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

  $\rightarrow$ no longer an excuse!

$\Rightarrow$ 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?

Standard textbook definitions

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

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

Standard textbook definitions

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

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

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

Standard textbook definitions

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

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

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

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

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

Standard textbook definitions

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

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

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

Future $\Leftrightarrow$ Past (believed so)

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

Definitions $\rightarrow$ evaluation rules

Very useful evaluation rules

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

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

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

Definitions $\longrightarrow$ evaluation rules

Very useful evaluation rules

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

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

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

BUT they cannot define the concept of probability!

Definitions $\longrightarrow$ evaluation rules

Very useful evaluation rules

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

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

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

In the probabilistic approach we are going to see

- Rule A will be recovered immediately (under the assumption of equiprobability, when it applies).
- Rule B will result from a theorem (under well defined assumptions).

What is probability?

What is probability?

It is what everybody knows what it is before going at school

What <u>is</u> probability?

It is what everybody knows what it is before going at school

$\rightarrow$ how much we are confident that something is true

What <u>is</u> probability?

It is what everybody knows what it is before going at school

- $\rightarrow$ how much we are confident that something is true
- $\rightarrow$ how much we believe something

What <u>is</u> probability?

It is what everybody knows what it is before going at school

→ how much we are confident that something is true

→ how much we believe something

→ "A measure of the degree of belief that an event *will* occur"

[Remark: 'will' does not imply future, but only uncertainty.]

Or perhaps you prefer this way. . .

"Given the state of our knowledge about everything that could possible have any bearing on the coming true[1]*. . . ,*

Or perhaps you prefer this way…

“Given the state of our knowledge about everything that could possible have any bearing on the coming true[1]*…,*

Or perhaps you prefer this way…

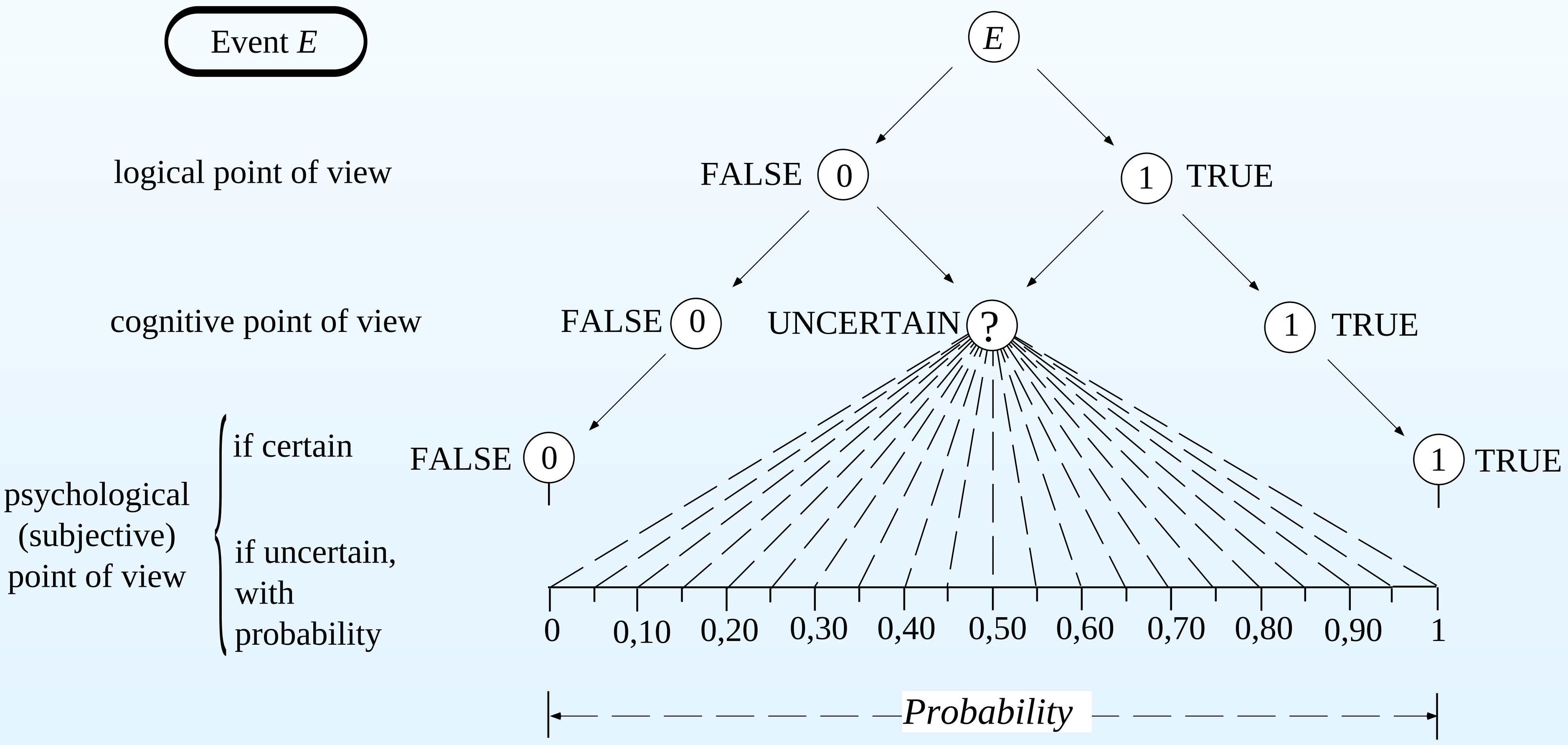
"Given the state of our knowledge about everything that could possible have any bearing on the coming true[1]*. . . , the numerical probability p of this event is to be a real number by the indication of which we try* in some cases *to setup a quantitative measure of the strength of our conjecture or anticipation, founded on the said knowledge, that the event comes true"*

Or perhaps you prefer this way...

"Given the state of our knowledge about everything that could possible have any bearing on the coming true[1]*..., the numerical probability* p *of this event is to be a real number by the indication of which we try* in some cases *to setup a quantitative measure of the strength of our conjecture or anticipation, founded on the said knowledge, that the event comes true"*
(E. Schrödinger, *The foundation of the theory of probability - I*, Proc. R. Irish Acad. 51A (1947) 51)

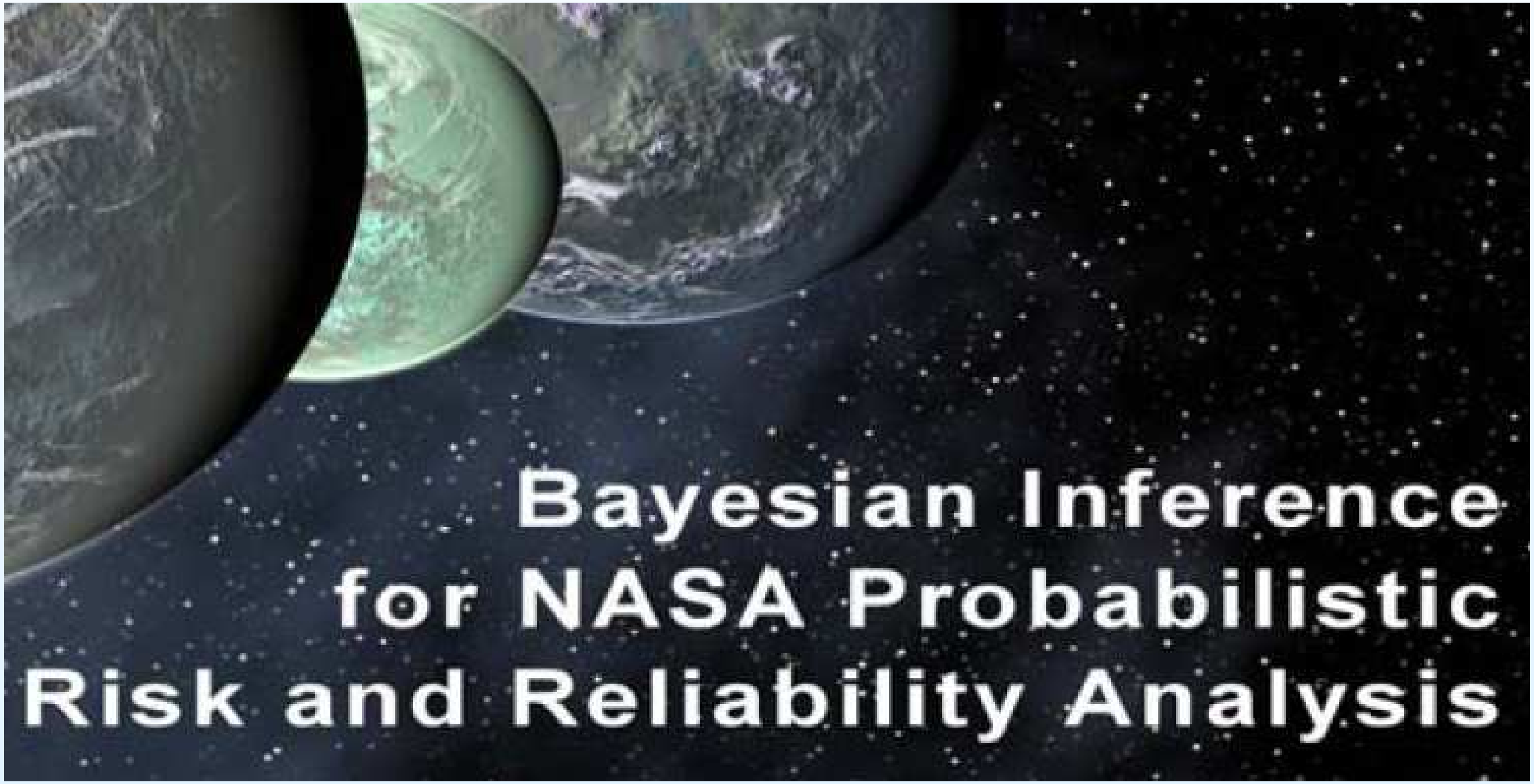
[1]*While in ordinary speech "to come true" usually refers to an event that is envisaged before it has happened, we use it here in the general sense, that the verbal description turns out to agree with actual facts.*

False, True and probable

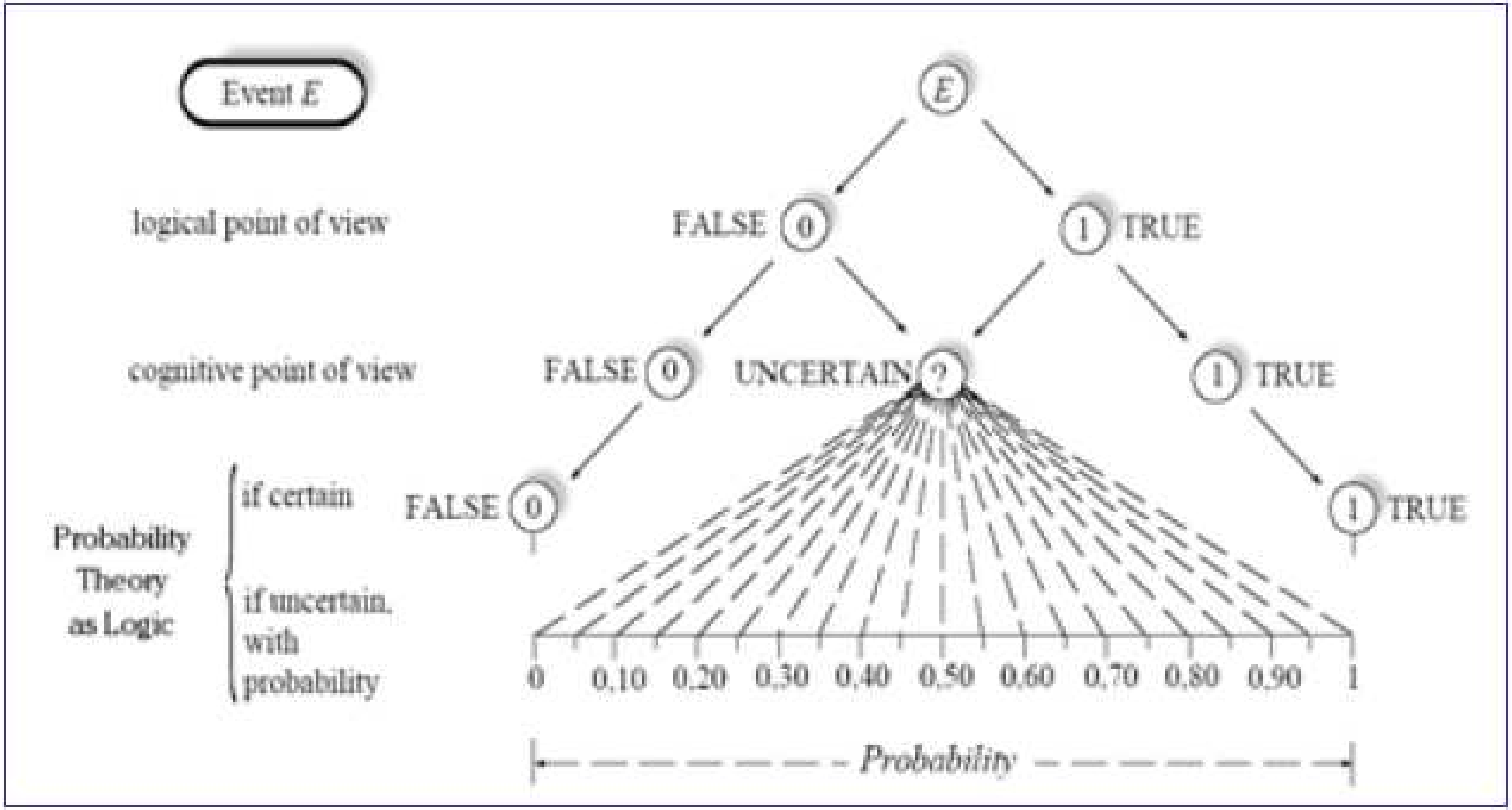


An helpful diagram

The previous diagram seems to help the understanding of the concept of probability



An helpful diagram



• Figure 2-1. Graphical abstraction of probability as a measure of information (adapted from "Probability and Measurement Uncertainty in Physics" by D'Agostini, [1995]).

(. . . but NASA guys are afraid of 'subjective', or 'psychological')

Uncertainty $\longrightarrow$ probability

Probability is related to uncertainty and not (only) to the results of repeated experiments

Uncertainty $\longrightarrow$ probability

Probability is related to uncertainty and not (only) to the results of repeated experiments

"If we were not ignorant there would be no probability, there could only be certainty. But our ignorance cannot be absolute, for then there would be no longer any probability at all. Thus the problems of probability may be classed according to the greater or less depth of our ignorance." (Poincaré)

Uncertainty $\rightarrow$ probability

Probability is related to uncertainty and not (only) to the results of repeated experiments

The state of information can be different from subject to subject

$\Rightarrow$ intrinsic subjective nature.

- No negative meaning: only an acknowledgment that several persons might have different information and, therefore, necessarily different opinions.

Uncertainty $\longrightarrow$ probability

Probability is related to uncertainty and not (only) to the results of repeated experiments

The state of information can be different from subject to subject

$\Rightarrow$ intrinsic subjective nature.

- No negative meaning: only an acknowledgment that several persons might have different information and, therefore, necessarily different opinions.
- *"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)

Uncertainty $\rightarrow$ probability

Probability is related to uncertainty and not (only) to the results of repeated experiments

Probability is always conditional probability

$$\text{`}P(E)\text{'} \longrightarrow P(E\,|\,I) \longrightarrow P(E\,|\,I(t))$$

Uncertainty $\longrightarrow$ probability

Probability is related to uncertainty and not (only) to the results of repeated experiments

Probability is always conditional probability

$$\text{`}P(E)\text{'} \quad \longrightarrow \quad P(E\,|\,I) \quad \longrightarrow \quad P(E\,|\,I(t))$$

- *"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)

Unifying role of subjective probability

- Wide range of applicability

Unifying role of subjective probability

- Wide range of applicability
- Probability statements all have the same meaning no matter to what they refer and how the number has been evaluated.
  - $P(\text{rain next Saturday}) = 68\%$
  - $P(\text{Milan will win Italian champion league}) = 68\%$
  - $P(\text{free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$

Unifying role of subjective probability

- Wide range of applicability
- Probability statements all have the same meaning no matter to what they refer and how the number has been evaluated.
  - $P(\text{rain next Saturday}) = 68\%$
  - $P(\text{Milan will win Italian champion league}) = 68\%$
  - $P(\text{free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$

  They all convey unambiguously the <u>same confidence</u> on something.

Unifying role of subjective probability

- Wide range of applicability
- Probability statements all have the same meaning no matter to what they refer and how the number has been evaluated.
  - $P(\text{rain next Saturday}) = 68\%$
  - $P(\text{Milan will win Italian champion league}) = 68\%$
  - $P(\text{free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$

  They all convey unambiguously the same confidence on something.
- You might agree or disagree, but at least You know what this person has in his mind. (NOT TRUE with “C.L.’s”!)

Unifying role of subjective probability

- Wide range of applicability
- Probability statements all have the same meaning no matter to what they refer and how the number has been evaluated.
  - $P(\text{rain next Saturday}) = 68\%$
  - $P(\text{Milan will win Italian champion league}) = 68\%$
  - $P(\text{free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$

  They all convey unambiguously the <u>same confidence</u> on something.
- You might agree or disagree, but at least You know what this person has in his mind. (<u>NOT TRUE with "C.L.'s"!</u>)
- If a person has these beliefs and he/she has the chance to win a rich prize bound to one of these events, he/she has no rational reason to chose an event instead than the others.

Unifying role of subjective probability

- Wide range of applicability
- Probability statements all have the same meaning no matter to what they refer and how the number has been evaluated.
  - $P(\text{rain next Saturday}) = 68\%$
  - $P(\text{Milan will win Italian champion league}) = 68\%$
  - $P(\text{free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$
- Probability not bound to a single evaluation rule.

Unifying role of subjective probability

- Wide range of applicability
- Probability statements all have the same meaning no matter to what they refer and how the number has been evaluated.
  - $P(\text{rain next Saturday}) = 68\%$
  - $P(\text{Milan will win Italian champion league}) = 68\%$
  - $P(\text{free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$

- Probability not bound to a single evaluation rule.
- In particular, combinatorial and frequency based 'definitions' are easily recovered as evaluation rules under well defined hypotheses.
- Keep separate concept from evaluation rule.

Confidence on the Higgs mass fron direct searches

PDG: $m_H > 114.4$ GeV at 95% C.L.

What does it mean?

Confidence on the Higgs mass fron direct searches

PDG: $m_H > 114.4$ GeV at 95% C.L.

What does it mean?

given only this piece of information from our LEP colleagues:

- What is $P(m_H \geq 114.4$ GeV)?
- What is $P(m_H \leq 114.4$ GeV)?

Confidence on the Higgs mass fron direct searches

PDG: $m_H > 114.4$ GeV at 95% C.L.

What does it mean?

given only this piece of information from our LEP colleagues:

- What is $P(m_H \geq 114.4$ GeV)?
- What is $P(m_H \leq 114.4$ GeV)?

Definitely not 95% and 5%! (. . . ??)

Confidence on the Higgs mass fron direct searches

PDG: $m_H > 114.4$ GeV at 95% C.L.

What does it mean?

given only this piece of information from our LEP colleagues:

- What is $P(m_H \geq 114.4$ GeV$)$?
- What is $P(m_H \leq 114.4$ GeV$)$?

Definitely not 95% and 5%! (. . . ??)

But, nevertheless, the 95% upper limit from radiative corrections gives a 95% probability. . .

Confidence on the Higgs mass fron direct searches

PDG: $m_H > 114.4$ GeV at 95% C.L.

What does it mean?

given only this piece of information from our LEP colleagues:

- What is $P(m_H \geq 114.4$ GeV)?
- What is $P(m_H \leq 114.4$ GeV)?

Definitely not 95% and 5%! (...??)

But, nevertheless, the 95% upper limit from radiative corrections gives a 95% probability...

Siamo uomini o caporali?

From the concept of probability to the probability theory

Ok, it looks nice, . . . but “how do we deal with ‘numbers’?”

From the concept of probability to the probability theory

- Formal structure: we need a mathematical structure in order to ‘propagate’ probability values to other, logically connected events:
  - basic rules
  - logic (mathematics)

From the concept of probability to the probability theory

- Formal structure: we need a mathematical structure in order to ‘propagate’ probability values to other, logically connected events:
  - basic rules
  - logic (mathematics)
- Assess probability: The formal structure is an empty box, in which we have to insert some numbers.
  - Is there a very general rule?

From the concept of probability to the probability theory

- Formal structure: we need a mathematical structure in order to 'propagate' probability values to other, logically connected events:
  - basic rules
  - logic (mathematics)
- Assess probability: The formal structure is an empty box, in which we have to insert some numbers.
  - Is there a very general rule?

Coherent bet (de Finetti, Ramsey - 'Dutch book argument')

It is well understood that bet odds can express confidence†

From the concept of probability to the probability theory

- Formal structure: we need a mathematical structure in order to ‘propagate’ probability values to other, logically connected events:
  - basic rules
  - logic (mathematics)
- Assess probability: The formal structure is an empty box, in which we have to insert some numbers.
  - Is there a very general rule?

Coherent bet $\rightarrow$ A bet acceptable in both directions:

- You state your confidence fixing the bet odds
- . . . but somebody else chooses the direction of the bet
- best way to honestly assess beliefs.

From the concept of probability to the probability theory

- Formal structure: we need a mathematical structure in order to ‘propagate’ probability values to other, logically connected events:
  - basic rules
  - logic (mathematics)
- Assess probability: The formal structure is an empty box, in which we have to insert some numbers.
  - Is there a very general rule?

Consistency arguments (Cox, + Good, Lucas)

From the concept of probability to the probability theory

- Formal structure: we need a mathematical structure in order to 'propagate' probability values to other, logically connected events:
  - basic rules
  - logic (mathematics)
- Assess probability: The formal structure is an empty box, in which we have to insert some numbers.
  - Is there a very general rule?

  Consistency arguments (Cox, + Good, Lucas)
- Similar approach by Schrödinger (much less known)
- Supported by Jaynes and Maximum Entropy school

From the concept of probability to the probability theory

- Formal structure: we need a mathematical structure in order to 'propagate' probability values to other, logically connected events:
  - basic rules
  - logic (mathematics)
- Assess probability: The formal structure is an empty box, in which we have to insert some numbers.
  - Is there a very general rule?

Consistency arguments (Cox, + Good, Lucas)

- Similar approach by Schrödinger (much less known)
- Supported by Jaynes and Maximum Entropy school

From the concept of probability to the probability theory

- Formal structure: we need a mathematical structure in order to 'propagate' probability values to other, logically connected events:
  - basic rules
  - logic (mathematics)
- Assess probability: The formal structure is an empty box, in which we have to insert some numbers.
  - Is there a very general rule?

Lindley's 'calibration' against 'standards'

$\rightarrow$ analogy to measures (we need to measure 'befiefs')

From the concept of probability to the probability theory

- Formal structure: we need a mathematical structure in order to 'propagate' probability values to other, logically connected events:
  - basic rules
  - logic (mathematics)
- Assess probability: The formal structure is an empty box, in which we have to insert some numbers.
  - Is there a very general rule?

Lindley's 'calibration' against 'standards'

$\rightarrow$ analogy to measures (we need to measure 'befiefs')

$\Rightarrow$ reference probabilities provided by simple cases in which equiprobability applies (coins, dice, turning wheels,. . .).

- Example: You are offered to options to receive a price: a) if E happens, b) if a coin will show head. Etc. . . .

From the concept of probability to the probability theory

- Formal structure: we need a mathematical structure in order to 'propagate' probability values to other, logically connected events:
  - basic rules
  - logic (mathematics)
- Assess probability: The formal structure is an empty box, in which we have to insert some numbers.
  - Is there a very general rule?

Lindley's 'calibration' against 'standards'

$\rightarrow$ Rational under everedays expressions like "there are 90 possibilities in 100" to state beliefs in situations in which the real possibilities are indeed only 2 (e.g. dead or alive)

- Example: a question to a student that has to pass an exam: a) normal test; b) pass it is a uniform random x will be ≤ 0.8.

From the concept of probability to the probability theory

- Formal structure: we need a mathematical structure in order to ‘propagate’ probability values to other, logically connected events:
  - basic rules
  - logic (mathematics)
- Assess probability: The formal structure is an empty box, in which we have to insert some numbers.
  - Is there a very general rule?

  Lindley’s ‘calibration’ against ‘standards’
- Also based on coherence, but it avoids the ‘repulsion’ of several person when they are asked to think directly in terms of bet (it is proved that many persons have reluctance to bet money).

Basic rules of probability

Coherence leads to

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

where

- Ω stands for ‘tautology’ (a proposition that is certainly true $\rightarrow$ referring to an event that is certainly true) and $\emptyset = \overline{\Omega}$.
- $A \cap B$ is true only when both A and B are true (logical AND) (shorthands ‘A, B’ or $A\,B$ often used $\rightarrow$ logical product)
- $A \cup B$ is true when at least one of the two propositions is true (logical OR)

Basic rules of probability

Remember that probability is always conditional probability!

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

I is the background condition (related to information I)

$\rightarrow$ usually implicit (we only care on ‘re-conditioning’)

Meaning of the basic rules

Have we recovered the famous axioms?

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

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

Meaning of the basic rules

More or less yes, at least formally

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

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

Meaning of the basic rules

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

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

- In the axiomatic approach
    - ‘probability’ is just a real number that satisfies 1-3
    - rule 4 comes straight from the definition of conditional probability as

$$P(A \,|\, B) = \frac{P(A \cap B)}{P(B)} \qquad [\text{if } P(B) > 0\,]$$

Meaning of the basic rules

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

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

- In the subjective approach
  - the intuitive meaning of ‘probability’ is recovered
  - rules 1-4 derive from more basic assumptions (e.g. the coherent bet)
  - $P(A \,|\, B) = P(A \cap B)/P(B)$ does not define $P(A \,|\, B)$
    $\rightarrow$ conditional probability is an intuitive concept!
    (Remember Schrödinger quote)

Meaning of the basic rules

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

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

- In the subjective approach
  - the intuitive meaning of ‘probability’ is recovered
  - rules 1-4 derive from more basic assumptions (e.g. the coherent bet)
  - $P(A \,|\, B) = P(A \cap B)/P(B)$ does not define $P(A \,|\, B)$
    $\rightarrow$ conditional probability is an intuitive concept!

$\Rightarrow$ As we actually use it! $\rightarrow$

About the ‘conditional probability formula’

$$4. \qquad P(E \cap H) = P(E \,|\, H) \cdot P(H) = P(H \,|\, E) \cdot P(E)$$

$$4a. \qquad P(E \,|\, H) = \frac{P(E \cap H)}{P(H)} \quad [P(H) > 0]$$

About the ‘conditional probability formula’

$$4. \qquad P(E\cap H) = P(E\,|\,H)\cdot P(H) = P(H\,|\,E)\cdot P(E)$$

$$4a. \qquad P(E\,|\,H) = \frac{P(E\cap H)}{P(H)} \quad [P(H)>0]$$

In the subjective approach the meaning is clear:

- Depending on the information we have, we can assess any of the three probabilities that enter the formula: $P(H)$, $P(E\,|\,H)$ or $P(E\cap H)$.
- But, once two of the three have been assessed, the third one is constraint!

  (otherwise, one can prove it is possible to imagine a set of bets, such that one certainly gains or loses – incoherent)
- 4 is more general than 4.a, valid also if $P(H)=0$

About the 'conditional probability formula'

$$4. \qquad P(E\cap H) = P(E\,|\,H)\cdot P(H) = P(H\,|\,E)\cdot P(E)$$

$$4a. \qquad P(E\,|\,H) = \frac{P(E\cap H)}{P(H)} \qquad [P(H)>0]$$

What is the chance that a 95 GeV Higgs is detected by ATLAS?

- H = "Higgs mass 95 GeV"
- E = "Decay products observed in ATLAS"

$\Rightarrow$ $P(E\,|\,H)$ is a routine task: $\rightarrow$ set $M_H = 95$ GeV in the physics generator $\rightarrow$ run the events through the full simulation chain $\rightarrow$ run analysis program $\rightarrow$ estimate $P(E\,|\,H)$ from percentage of reconstructed events.

- None would use definition 4a [what is $P(E\cap H)$?]
- Note: $P(E\,|\,H)$ is meaningful even if $P(H)=0$ (why not?).

Subjective $\neq$ arbitrary

Crucial role of the coherent bet

- You claim that <u>this</u> coin has 70% to give head?
  No problem with me: you place 70€ on head, I 30€ on tail and who wins take 100€.
  $\Rightarrow$ If OK with you, let's start.

Subjective $\neq$ arbitrary

Crucial role of the coherent bet

- You claim that this coin has 70% to give head?
  No problem with me: you place 70€ on head, I 30€ on tail and who wins take 100€.
  $\Rightarrow$ If OK with you, let's start.
- You claim that this coin has 30% to give head?
  $\Rightarrow$ Just reverse the bet

  (Like sharing goods, e.g. a cake with a child)

Subjective $\neq$ arbitrary

Crucial role of the coherent bet

- You claim that <u>this</u> coin has 70% to give head?
  No problem with me: you place 70€ on head, I 30€ on tail and who wins take 100€.
  $\Rightarrow$ If OK with you, let's start.
- You claim that <u>this</u> coin has 30% to give head?
  $\Rightarrow$ Just reverse the bet

  (Like sharing goods, e.g. a cake with a child)

$\Rightarrow$ Take into account all available information *in the most 'objective way'*
(Even that someone has a different opinion!)

$\Rightarrow$ It might seem paradoxically, but the 'subjectivist' is much more 'objective' than those who blindly use so-called objective methods.

Summary on probabilistic approach

- Probability means how much we believe something
- Probability values obey the following basic rules

$$
\begin{array}{ll}
1. & 0 \le P(A) \le 1 \\
2. & P(\Omega) = 1 \\
3. & P(A \cup B) = P(A) + P(B) \qquad [\,\text{if } P(A \cap B) = \emptyset\,] \\
4. & P(A \cap B) = P(A \,|\, B) \cdot P(B) = P(B \,|\, A) \cdot P(A)\,,
\end{array}
$$

- All the rest by logic

$\rightarrow$ And, please, be coherent!

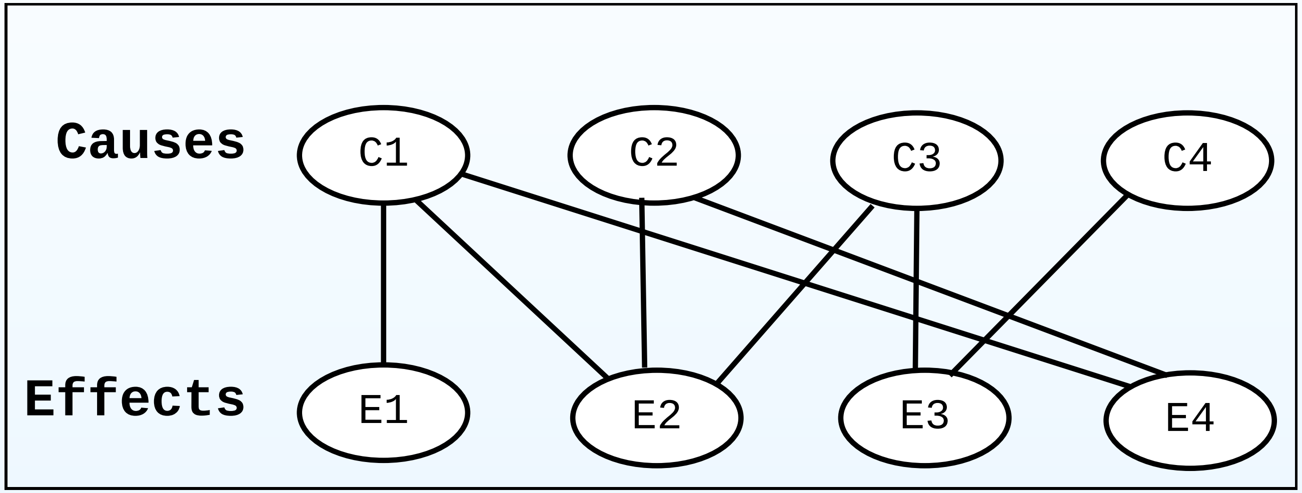
Inference

Inference

$\Rightarrow$ How do we learn from data in a probabilistic framework?

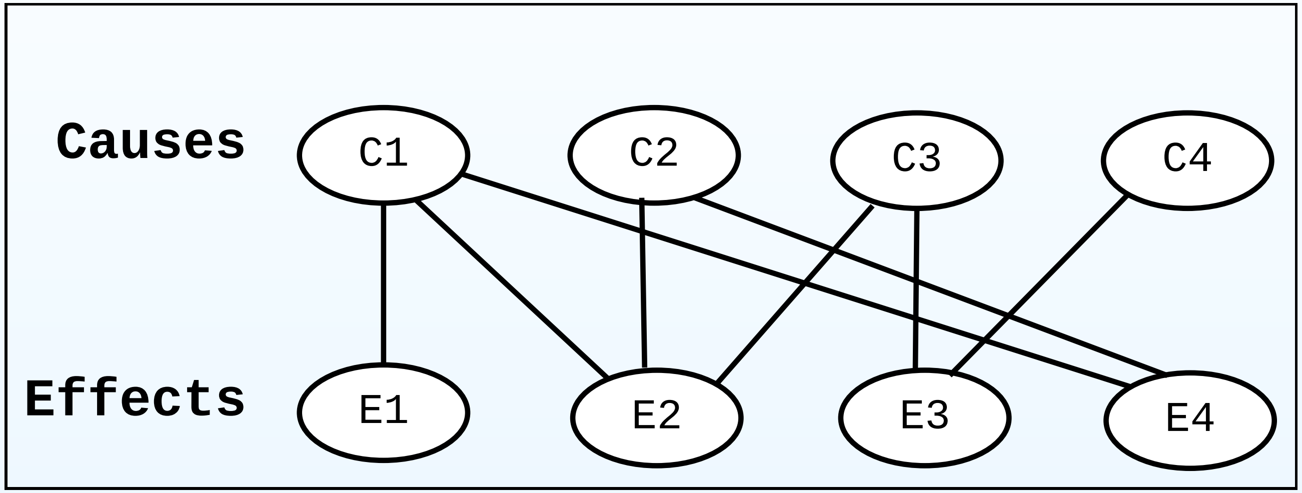
From causes to effects and back

Our original problem:



From causes to effects and back

Our original problem:

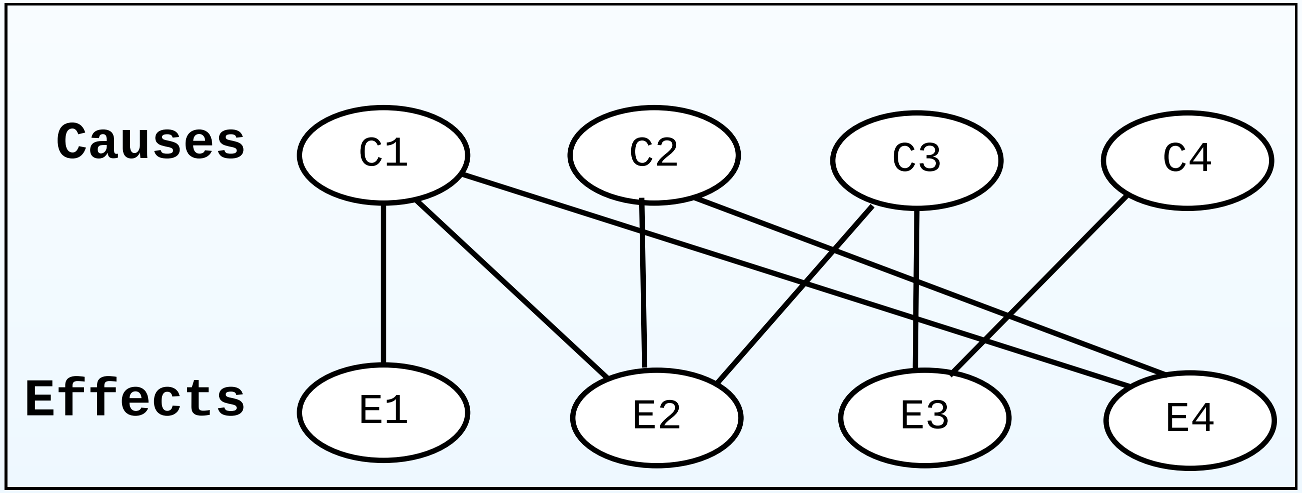


Our conditional view of probabilistic causation

$$\boxed{P(E_i \,|\, C_j)}$$

From causes to effects and back

Our original problem:



Our conditional view of probabilistic causation

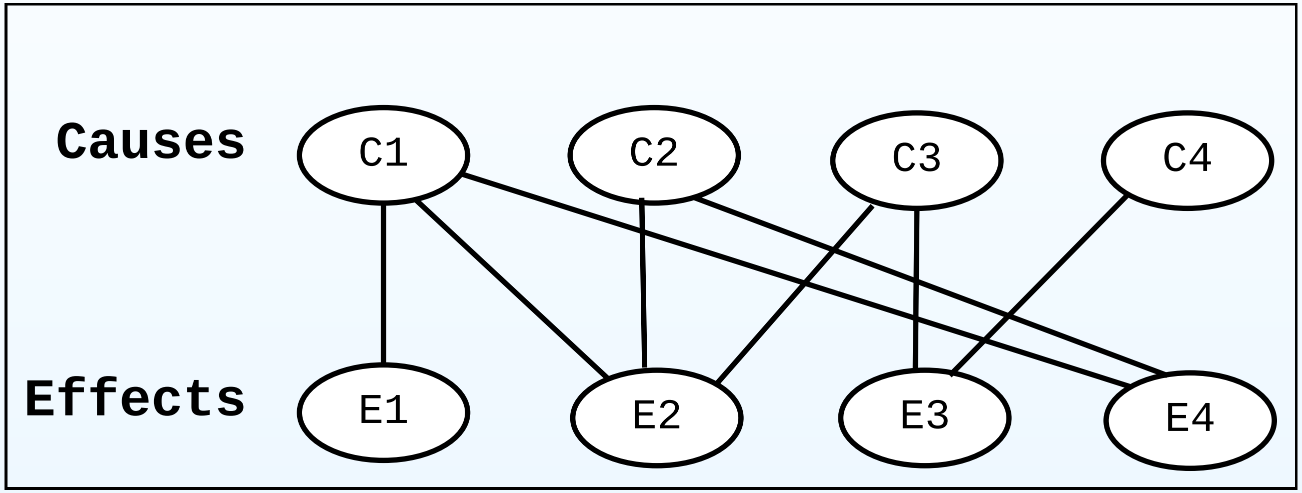
$$\boxed{P(E_i\,|\,C_j)}$$

Our conditional view of probabilistic inference

$$\boxed{P(C_j\,|\,E_i)}$$

From causes to effects and back

Our original problem:



Our conditional view of probabilistic causation

$$\boxed{P(E_i\,|\,C_j)}$$

Our conditional view of probabilistic inference

$$\boxed{P(C_j\,|\,E_i)}$$

The fourth basic rule of probability:

$$\boxed{P(C_j, E_i) = P(E_i\,|\,C_j)\,P(C_j) = P(C_j\,|\,E_i)\,P(E_i)}$$

Symmetric conditioning

Let us take basic rule 4, written in terms of hypotheses H_j and effects E_i, and rewrite it this way:

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

"The condition on E_i changes in percentage the probability of H_j as the probability of E_i is changed in percentage by the condition H_j."

Symmetric conditioning

Let us take basic rule 4, written in terms of hypotheses H_j and effects E_i, and rewrite it this way:

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

"The condition on E_i changes in percentage the probability of H_j as the probability of E_i is changed in percentage by the condition H_j."

It follows

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

Symmetric conditioning

Let us take basic rule 4, written in terms of hypotheses H_j and effects E_i, and rewrite it this way:

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

"The condition on E_i changes in percentage the probability of H_j as the probability of E_i is changed in percentage by the condition H_j."

It follows

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

Got 'after' Calculated 'before'

(where 'before' and 'after' refer to the knowledge that E_i is true.)

Symmetric conditioning

Let us take basic rule 4, written in terms of hypotheses H_j and effects E_i, and rewrite it this way:

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

"The condition on E_i changes in percentage the probability of H_j as the probability of E_i is changed in percentage by the condition H_j."

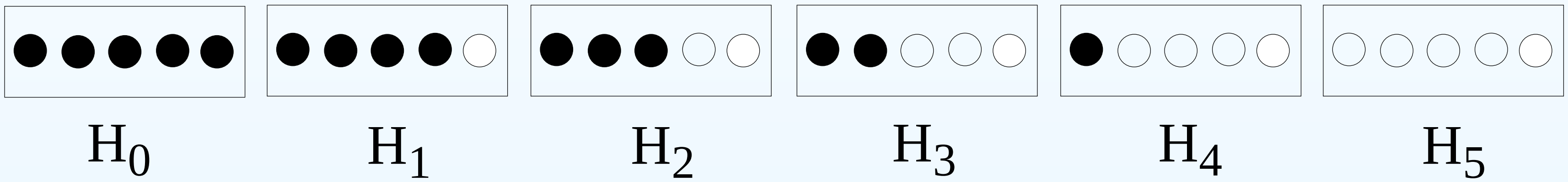
It follows

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

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

(Gauss)

Application to the six box problem



Remind:

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

Collecting the pieces of information we need

Our tool:

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

Collecting the pieces of information we need

Our tool:

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

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

Collecting the pieces of information we need

Our tool:

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

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

Collecting the pieces of information we need

Our tool:

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

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

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

Collecting the pieces of information we need

Our tool:

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

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

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

Our prior belief about H_j

Collecting the pieces of information we need

Our tool:

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

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

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

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

Collecting the pieces of information we need

Our tool:

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

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

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

Probability of E_i taking account all possible H_j
$\rightarrow$ How much we are confident that E_i will occur.

Collecting the pieces of information we need

Our tool:

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

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

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

Probability of E_i taking account all possible H_j
$\rightarrow$ How much we are confident that E_i will occur.
Easy in this case, because of the symmetry of the problem.
But already after the first extraction of a ball our opinion about the box content will change, and symmetry will break.

Collecting the pieces of information we need

Our tool:

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

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

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

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

Collecting the pieces of information we need

Our tool:

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

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

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

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

'decomposition law': $P(E_i \,|\, I) = \sum_j P(E_i \,|\, H_j,\, I) \cdot P(H_j \,|\, I)$

($\rightarrow$ Easy to check that it gives $P(E_i \,|\, I) = 1/2$ in our case).

Collecting the pieces of information we need

Our tool:

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

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

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

We are ready!

$\longrightarrow$ Let's play with our toy

Naming the method

Some ‘remarks’ on formalism and notation.

(But nothing deep!)

Naming the method

Some ‘remarks’ on formalism and notation.

(But nothing deep!)

From now on it is only a question of

- experience and good sense to model the problem;
- patience;
- math skill;
- computer skill.

Naming the method

Some 'remarks' on formalism and notation.

(But nothing deep!)

From now on it is only a question of

- experience and good sense to model the problem;
- patience;
- math skill;
- computer skill.

Moving to continuous quantities:

- transitions discrete$\rightarrow$continuous rather simple;
- prob. functions $\rightarrow$ pdf
- learn to summarize the result in *'a couple of meaningful numbers'* (but remembering that the full answer is in the *final pdf*).

Bayes theorem

The formulae used to *infer* H_i and
to *predict* $E_j^{(2)}$ are related to the name of Bayes

Bayes theorem

The formulae used to *infer* H_i and
to *predict* $E_j^{(2)}$ are related to the name of Bayes

Neglecting the background state of information I:

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

Bayes theorem

The formulae used to *infer* H_i and
to *predict* $E_j^{(2)}$ are related to the name of Bayes

Neglecting the background state of information I:

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

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

Bayes theorem

The formulae used to *infer* H_i and
to *predict* $E_j^{(2)}$ are related to the name of Bayes

Neglecting the background state of information I:

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

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

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

Bayes theorem

The formulae used to *infer* H_i and to *predict* $E_j^{(2)}$ are related to the name of Bayes

Neglecting the background state of information I:

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

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

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

$$P(H_j\,|\,E_i) \propto P(E_i\,|\,H_j)\cdot P(H_j)$$

Bayes theorem

The formulae used to *infer* H_i and
to *predict* $E_j^{(2)}$ are related to the name of Bayes

Neglecting the background state of information I:

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

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

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

$$P(H_j \,|\, E_i) \propto P(E_i \,|\, H_j) \cdot P(H_j)$$

Different ways to write the

Bayes' Theorem

Updating the knowledge by new observations

Let us repeat the experiment:

$\boxed{\text{Sequential use of Bayes theorem}}$

Old posterior becomes new prior, and so on

Updating the knowledge by new observations

Let us repeat the experiment:

Sequential use of Bayes theorem

Old posterior becomes new prior, and so on

$$P(H_j \,|\, E^{(1)}, E^{(2)}) \;\;\propto\;\; P(E^{(2)} \,|\, H_j, E^{(1)}) \cdot P(H_j \,|\, E^{(1)})$$

Updating the knowledge by new observations

Let us repeat the experiment:

$\boxed{\text{Sequential use of Bayes theorem}}$

Old posterior becomes new prior, and so on

$$\begin{aligned} P(H_j \,|\, E^{(1)}, E^{(2)}) &\propto P(E^{(2)} \,|\, H_j, E^{(1)}) \cdot P(H_j \,|\, E^{(1)}) \\ &\propto P(E^{(2)} \,|\, H_j) \cdot P(H_j \,|\, E^{(1)}) \end{aligned}$$

Updating the knowledge by new observations

Let us repeat the experiment:

Sequential use of Bayes theorem

Old posterior becomes new prior, and so on

$$
\begin{aligned}
P(H_j \,|\, E^{(1)}, E^{(2)}) &\propto P(E^{(2)} \,|\, H_j, E^{(1)}) \cdot P(H_j \,|\, E^{(1)}) \\
&\propto P(E^{(2)} \,|\, H_j) \cdot P(H_j \,|\, E^{(1)}) \\
&\propto P(E^{(2)} \,|\, H_j) \cdot P(E^{(1)} \,|\, H_j) \cdot P_0(H_j)
\end{aligned}
$$

Updating the knowledge by new observations

Let us repeat the experiment:

Sequential use of Bayes theorem

Old posterior becomes new prior, and so on

$$
\begin{aligned}
P(H_j \,|\, E^{(1)}, E^{(2)}) &\propto P(E^{(2)} \,|\, H_j, E^{(1)}) \cdot P(H_j \,|\, E^{(1)}) \\
&\propto P(E^{(2)} \,|\, H_j) \cdot P(H_j \,|\, E^{(1)}) \\
&\propto P(E^{(2)} \,|\, H_j) \cdot P(E^{(1)} \,|\, H_j) \cdot P_0(H_j) \\
&\propto P(E^{(1)}, E^{(1)} \,|\, H_j) \cdot P_0(H_j)
\end{aligned}
$$

Updating the knowledge by new observations

Let us repeat the experiment:

$\boxed{\text{Sequential use of Bayes theorem}}$

Old posterior becomes new prior, and so on

$$
\begin{aligned}
P(H_j \,|\, E^{(1)}, E^{(2)}) &\propto P(E^{(2)} \,|\, H_j, E^{(1)}) \cdot P(H_j \,|\, E^{(1)}) \\
&\propto P(E^{(2)} \,|\, H_j) \cdot P(H_j \,|\, E^{(1)}) \\
&\propto P(E^{(2)} \,|\, H_j) \cdot P(E^{(1)} \,|\, H_j) \cdot P_0(H_j) \\
&\propto P(E^{(1)}, E^{(1)} \,|\, H_j) \cdot P_0(H_j) \\
P(H_j \,|\, \text{data}) &\propto P(\text{data} \,|\, H_j) \cdot P_0(H_j)
\end{aligned}
$$

Updating the knowledge by new observations

Let us repeat the experiment:

$\boxed{\text{Sequential use of Bayes theorem}}$

Old posterior becomes new prior, and so on

$$
\begin{aligned}
P(H_j \,|\, E^{(1)}, E^{(2)}) &\propto P(E^{(2)} \,|\, H_j, E^{(1)}) \cdot P(H_j \,|\, E^{(1)}) \\
&\propto P(E^{(2)} \,|\, H_j) \cdot P(H_j \,|\, E^{(1)}) \\
&\propto P(E^{(2)} \,|\, H_j) \cdot P(E^{(1)} \,|\, H_j) \cdot P_0(H_j) \\
&\propto P(E^{(1)}, E^{(1)} \,|\, H_j) \cdot P_0(H_j) \\
P(H_j \,|\, \text{data}) &\propto P(\text{data} \,|\, H_j) \cdot P_0(H_j)
\end{aligned}
$$

$\boxed{\text{Bayesian inference}}$

Updating the knowledge by new observations

Let us repeat the experiment:

Sequential use of Bayes theorem

Old posterior becomes new prior, and so on

$$
\begin{aligned}
P(H_j \,|\, E^{(1)}, E^{(2)}) &\propto P(E^{(2)} \,|\, H_j, E^{(1)}) \cdot P(H_j \,|\, E^{(1)}) \\
&\propto P(E^{(2)} \,|\, H_j) \cdot P(H_j \,|\, E^{(1)}) \\
&\propto P(E^{(2)} \,|\, H_j) \cdot P(E^{(1)} \,|\, H_j) \cdot P_0(H_j) \\
&\propto P(E^{(1)}, E^{(1)} \,|\, H_j) \cdot P_0(H_j) \\
P(H_j \,|\, \text{data}) &\propto P(\text{data} \,|\, H_j) \cdot P_0(H_j)
\end{aligned}
$$

Learning from data using probability theory

Exercises and discussions

- Continue with six box problem [$\rightarrow$ *AJP* 67 (1999) 1260]
  $\rightarrow$ Slides
- <u>Home work 1</u>: AIDS problem $\rightarrow P(\text{HIV}\,|\,\text{Pos})$?

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

- <u>Home work 2</u>: Particle identification:
  A particle detector has a μ identification efficiency of $95\,\%$, and a probability of identifying a π as a μ of $2\,\%$. If a particle is identified as a μ, then a trigger is fired. Knowing that the particle beam is a mixture of $90\,\%$ π and $10\,\%$ μ, what is the probability that a trigger is really fired by a μ? What is the signal-to-noise (S/N) ratio?

Odd ratios and Bayes factor

$$
\begin{aligned}
\frac{P(\mathsf{HIV}\,|\,\mathsf{Pos})}{P(\overline{\mathsf{HIV}}\,|\,\mathsf{Pos})} &= \frac{P(\mathsf{Pos}\,|\,\mathsf{HIV})}{P(\mathsf{Pos}\,|\,\overline{\mathsf{HIV}})} \cdot \frac{P_\circ(\mathsf{HIV})}{P(\overline{\mathsf{HIV}})} \\
&= \frac{\approx 1}{0.002} \times \frac{0.1/60}{\approx 1} = 500 \times \frac{1}{600} = \frac{1}{1.2} \\
\Rightarrow P(\mathsf{HIV}\,|\,\mathsf{Pos}) &= 45.5\%\,.
\end{aligned}
$$

Odd ratios and Bayes factor

$$\begin{aligned}
\frac{P(\text{HIV}\,|\,\text{Pos})}{P(\overline{\text{HIV}}\,|\,\text{Pos})} &= \frac{P(\text{Pos}\,|\,\text{HIV})}{P(\text{Pos}\,|\,\overline{\text{HIV}})}\cdot\frac{P_\circ(\text{HIV})}{P(\overline{\text{HIV}})} \\
&= \frac{\approx 1}{0.002}\times\frac{0.1/60}{\approx 1} = 500\times\frac{1}{600} = \frac{1}{1.2} \\
\Rightarrow P(\text{HIV}\,|\,\text{Pos}) &= 45.5\%\,.
\end{aligned}$$

There are some advantages in expressing Bayes theorem in terms of odd ratios:

- There is no need to consider all possible hypotheses (how can we be sure?)
  We just make a comparison of any couple of hypotheses!

Odd ratios and Bayes factor

$$\frac{P(\text{HIV}\,|\,\text{Pos})}{P(\overline{\text{HIV}}\,|\,\text{Pos})} = \frac{P(\text{Pos}\,|\,\text{HIV})}{P(\text{Pos}\,|\,\overline{\text{HIV}})} \cdot \frac{P_\circ(\text{HIV})}{P(\overline{\text{HIV}})}$$

$$= \frac{\approx 1}{0.002} \times \frac{0.1/60}{\approx 1} = 500 \times \frac{1}{600} = \frac{1}{1.2}$$

$$\Rightarrow P(\text{HIV}\,|\,\text{Pos}) = 45.5\%\,.$$

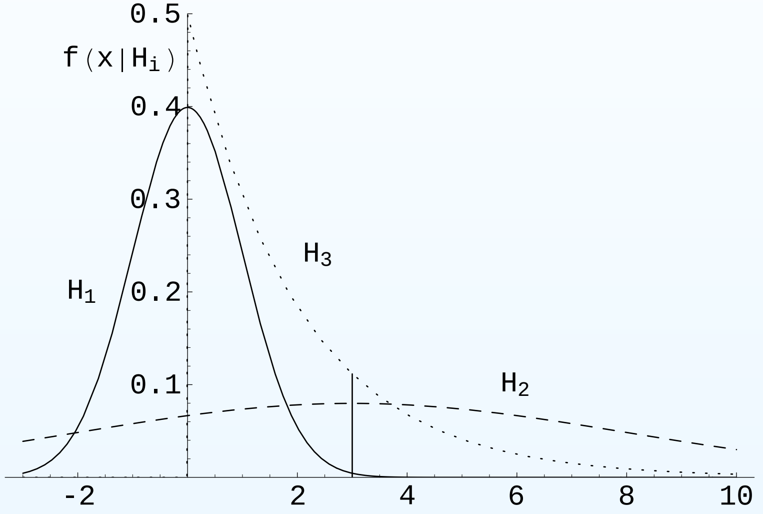
There are some advantages in expressing Bayes theorem in terms of odd ratios:

- There is no need to consider all possible hypotheses (how can we be sure?)
  We just make a comparison of any couple of hypotheses!
- Bayes factor is usually much more inter-subjective, and it is often considered an 'objective' way to report how much the data favor each hypothesis.

The three models example

Choose among H_1, H_2 and H_3 having observed $x = 3$:

In case of 'likelihoods' given by pdf's, the same formulae apply: "$P(\text{data} \,|\, H_j)$" $\longleftrightarrow$ "$f(\text{data} \,|\, H_j)$".



$$BF_{j,k} = \frac{f(x\!=\!3 \,|\, H_j)}{f(x\!=\!3 \,|\, H_k)}$$

$BF_{2,1} = 18$, $BF_{3,1} = 25$ and $BF_{3,2} = 1.4$ $\rightarrow$ data favor model H_3 (as we can see from figure!), but if we want to state how much we believe to each model we need to 'filter' them with priors.

Assuming the three models initially equally likely, we get final probabilities of 2.3%, 41% and 57% for the three models.

A last remark

A last remark on model comparisons

- for a ‘serious’ probabilistic model comparisons,
  at least two well defined models are needed

A last remark

A last remark on model comparisons

- for a ‘serious’ probabilistic model comparisons, **at least two well defined models are needed**
- p-values (e.g. ‘χ^2 tests) have to be considered very useful starting points to understand if further investigation is worth [Yes, I also use χ^2 to get an idea of the “distance” between a model and the experimental data – but not more than that].

A last remark

A last remark on model comparisons

- for a ‘serious’ probabilistic model comparisons, **at least two well defined models are needed**
- p-values (e.g. ‘χ^2 tests) have to be considered very useful starting points to understand if further investigation is worth [Yes, I also use χ^2 to get an idea of the “distance” between a model and the experimental data – but not more than that].
- But until you don’t have an alternative and credible model to explain the data, there is little to say about the “chance that the data come from the model”, unless the data are really impossible.

A last remark

A last remark on model comparisons

- for a ‘serious’ probabilistic model comparisons, **at least two well defined models are needed**
- p-values (e.g. ‘χ^2 tests) have to be considered very useful starting points to understand if further investigation is worth [Yes, I also use χ^2 to get an idea of the “distance” between a model and the experimental data – but not more than that].
- But until you don’t have an alternative and credible model to explain the data, there is little to say about the “chance that the data come from the model”, unless the data are really impossible.
- Why do frequentistic test often work? $\rightarrow$ Think about...
  (Just by chance – no logical necessity)

End

FINE