Probabilistic Reasoning in Frontier Science

Giulio D'Agostini

Dipartimento di Fisica

Università di Roma La Sapienza

Preamble

- not *statistical methods for . . . ”*

Preamble

- not *statistical methods for . . . ”*
- but Probabilistic Reasoning in . . .

Preamble

- not *statistical methods for . . . ”*
- but Probabilistic Reasoning in . . .

⇒ Foundamental aspects enhanced
. . . although some usefull *probabilistic methods* will be presented

Preamble

- not *statistical methods for . . . ”*
- but Probabilistic Reasoning in . . .

⇒ Foundamental aspects enhanced
. . . although some usefull *probabilistic methods* will be presented

- Just finishing a “40 hours” course on

Probabilità e Incertezza di Misura

to PhD students in Rome
→ I asked an advice to the students about what to present.

In particular, most things of this first day reflect what they find it is important I tell you.

Outline (today)

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

  "The essential problem of the experimental method" (Poincaré).
- The master example: the six box problem.

  "Probability is either referred to real cases or it is nothing" (de Finetti).
- Falsificationism and *statistics variations* ('test')
- Probabilistic approach.
- *What is probability?*
- Basic rules of probability and Bayes rule.
- Bayesian inference.
- Conclusions

Causes ↔ effects

CAUSAS E CONSEQUÊNCIAS

(p.2 Correiro da Manhã got on TAP)

Tomorrow

Applications of probabilistic inference to physics quantities

(after finishing with pending items from today. . .)

- Parametric inference

Tomorrow

Applications of probabilistic inference to physics quantities

(after finishing with pending items from today. . .)

- Parametric inference
- Application to typical detector responses
  - binomial (efficiencies, branching ratios, ‘proportions’)
  - Poisson (counts following “Poisson process”)
  - Gaussian (‘normal errors’, approximation of other pdf)

Tomorrow

Applications of probabilistic inference to physics quantities

(after finishing with pending items from today. . .)

- Parametric inference
- Application to typical detector responses
  - binomial (efficiencies, branching ratios, ‘proportions’)
  - Poisson (counts following “Poisson process”)
  - Gaussian (‘normal errors’, approximation of other pdf)

. . . with some ‘complications’:

  - systematics
  - background
  - measurements at the limit of the detector sensitivity

Tomorrow

Applications of probabilistic inference to physics quantities

(after finishing with pending items from today. . .)

- Parametric inference
- Application to typical detector responses
  - binomial (efficiencies, branching ratios, ‘proportions’)
  - Poisson (counts following “Poisson process”)
  - Gaussian (‘normal errors’, approximation of other pdf)

. . . with some ‘complications’:

  - systematics
  - background
  - measurements at the limit of the detector sensitivity

- Propagation of uncertainties

Tomorrow

Applications of probabilistic inference to physics quantities

(after finishing with pending items from today. . .)

- Parametric inference
- Application to typical detector responses
  - binomial (efficiencies, branching ratios, ‘proportions’)
  - Poisson (counts following “Poisson process”)
  - Gaussian (‘normal errors’, approximation of other pdf)

. . . with some ‘complications’:

  - systematics
  - background
  - measurements at the limit of the detector sensitivity

- Propagation of uncertainties
- Any special wish?

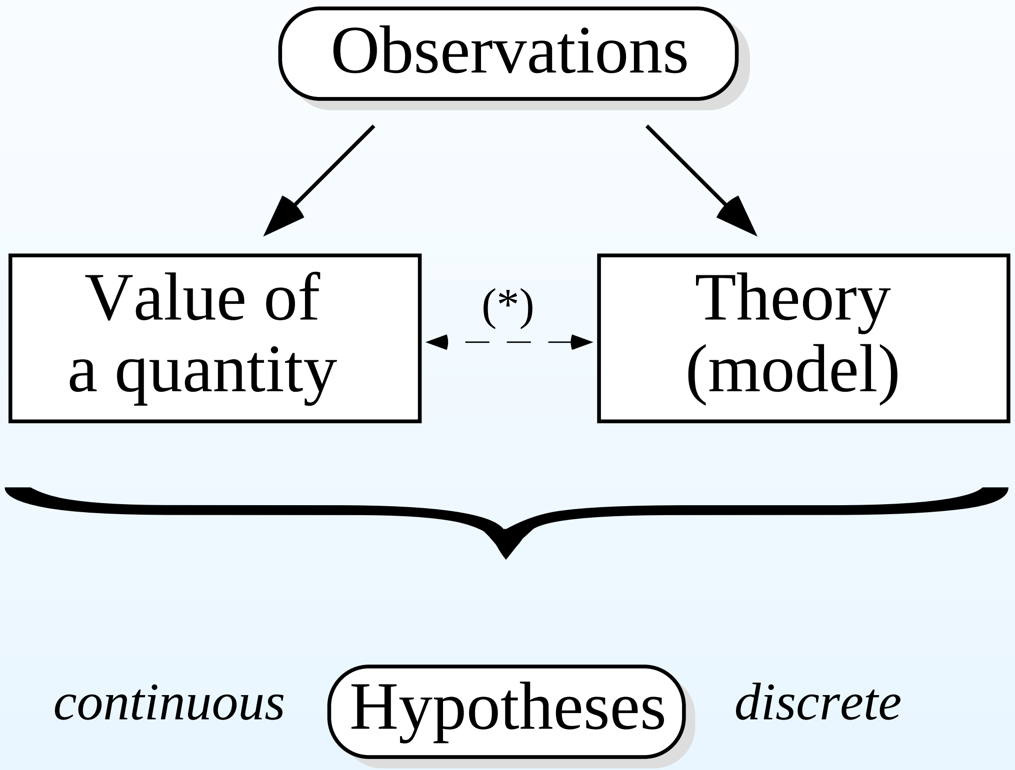
Friday

Unfolding method

based on the probabilistic reasoning illustrated today:

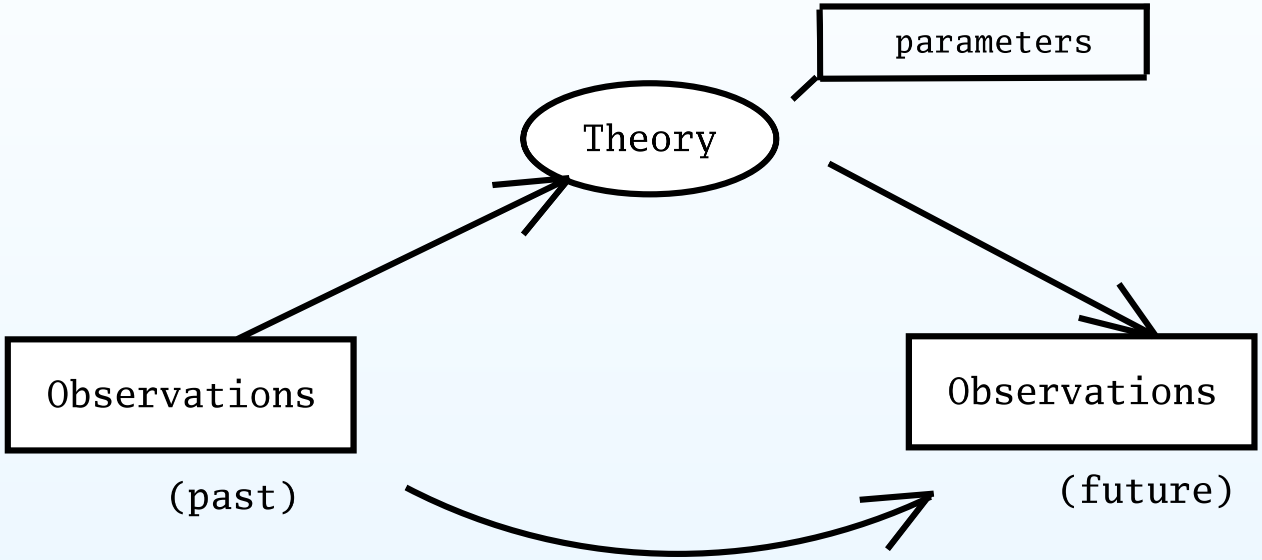
⇒ how to correct an observed spectrum for distorsions of several kinds.

Physics



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

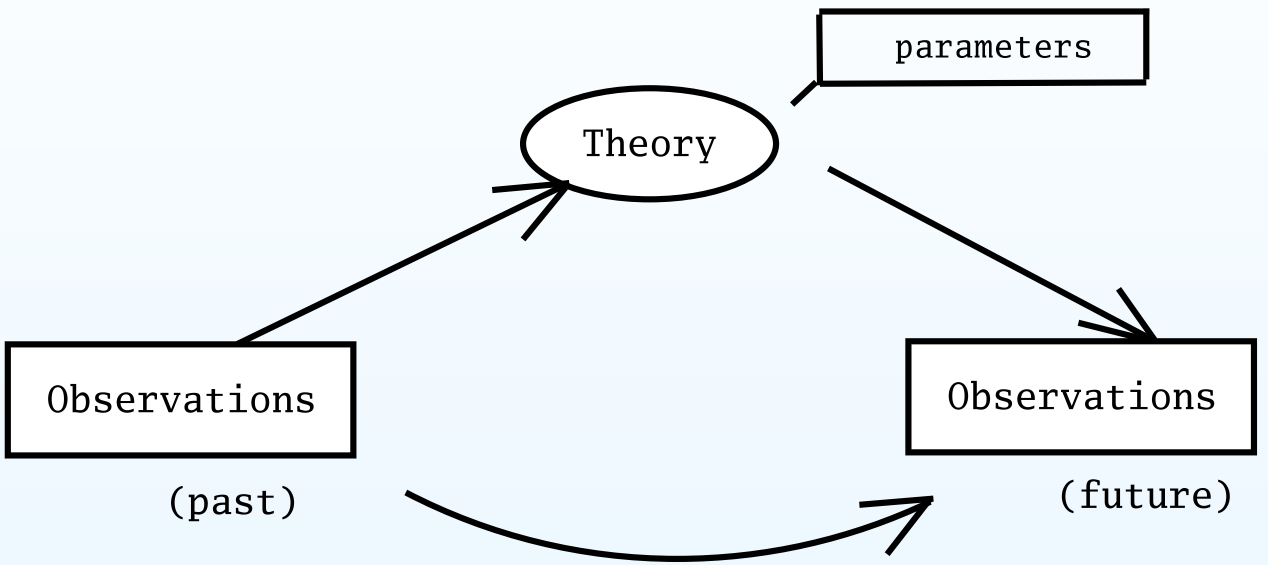
From past to future



Task of the ‘physicist’ (scientist, decision maker):

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

From past to future



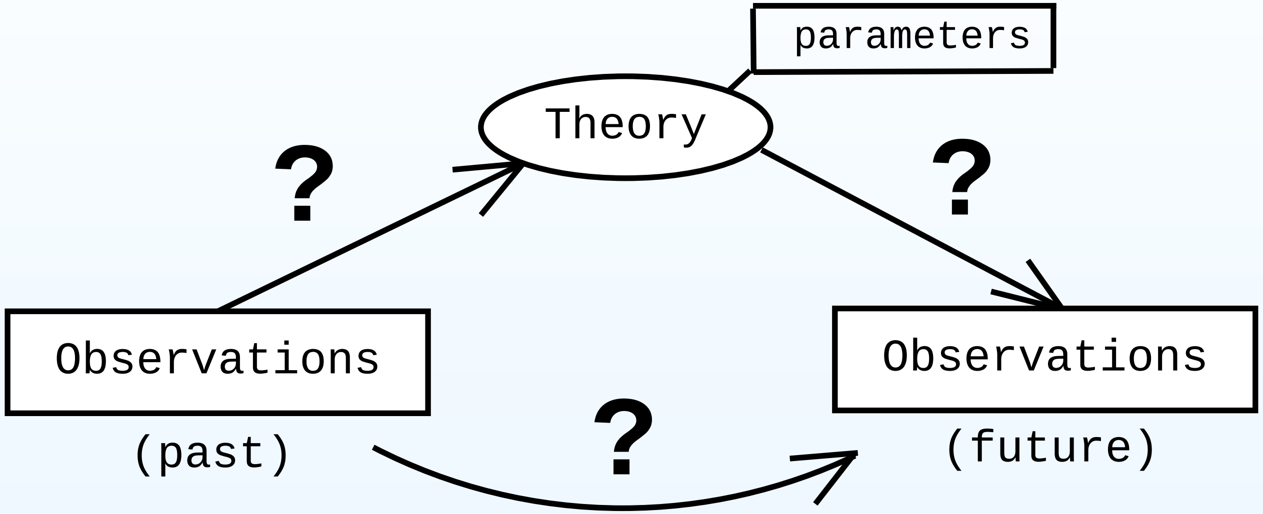
Process

- neither automatic
- nor purely contemplative

  $\longrightarrow$ 'scientific method'

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

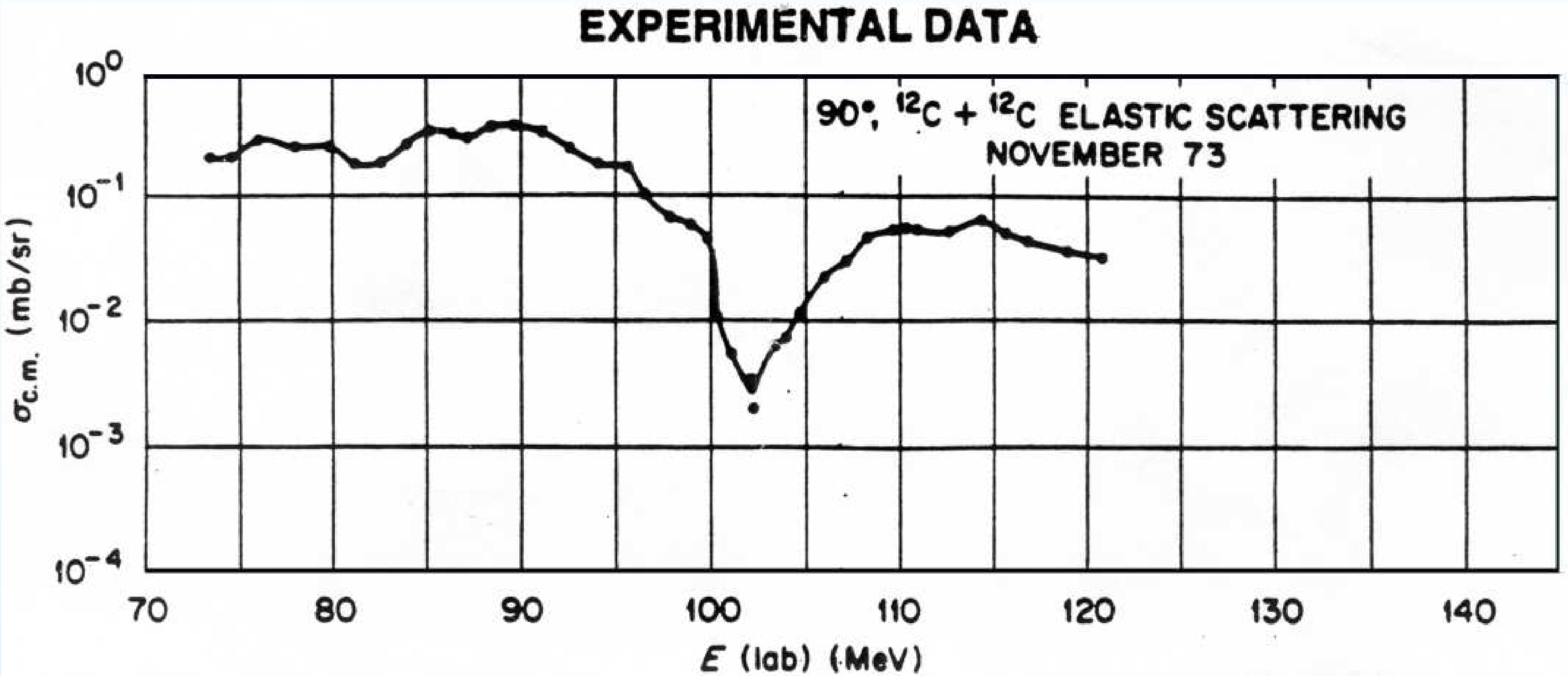
From past to future



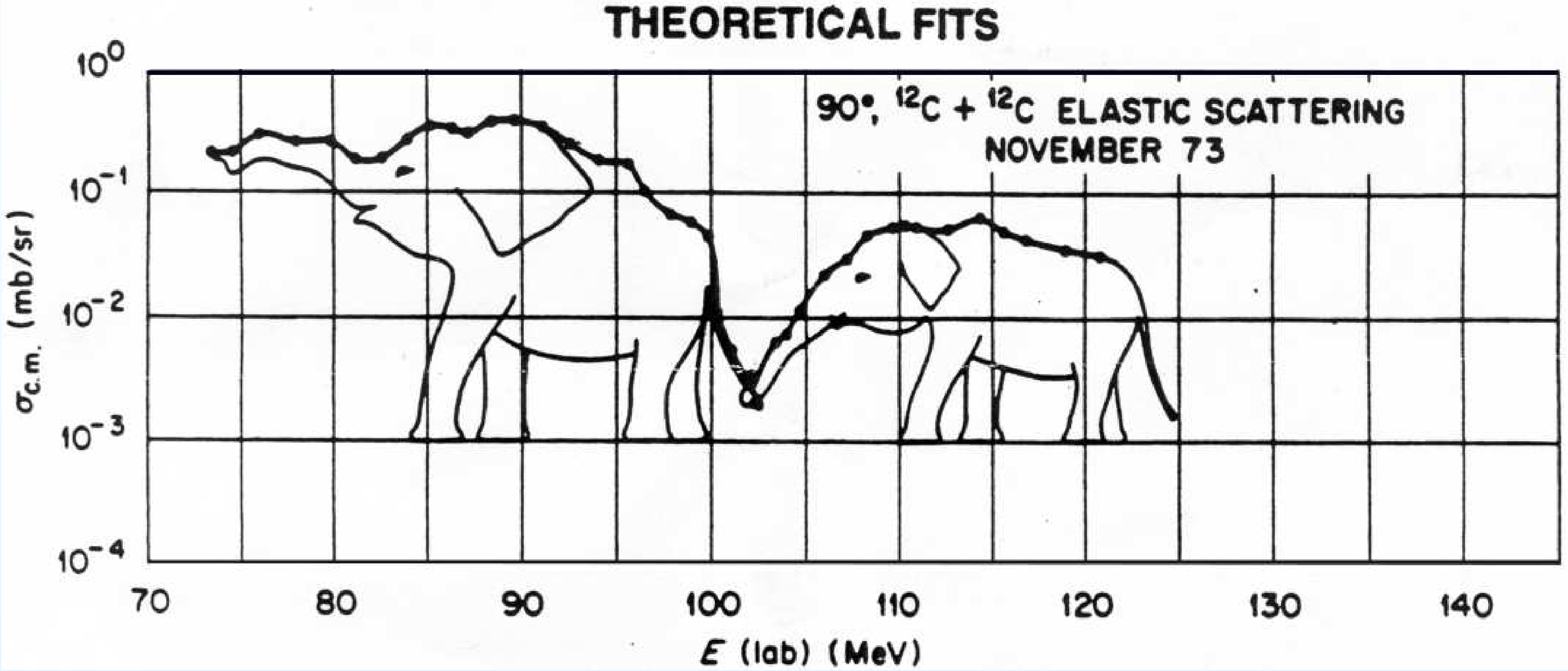
⇒ Uncertainty:

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.

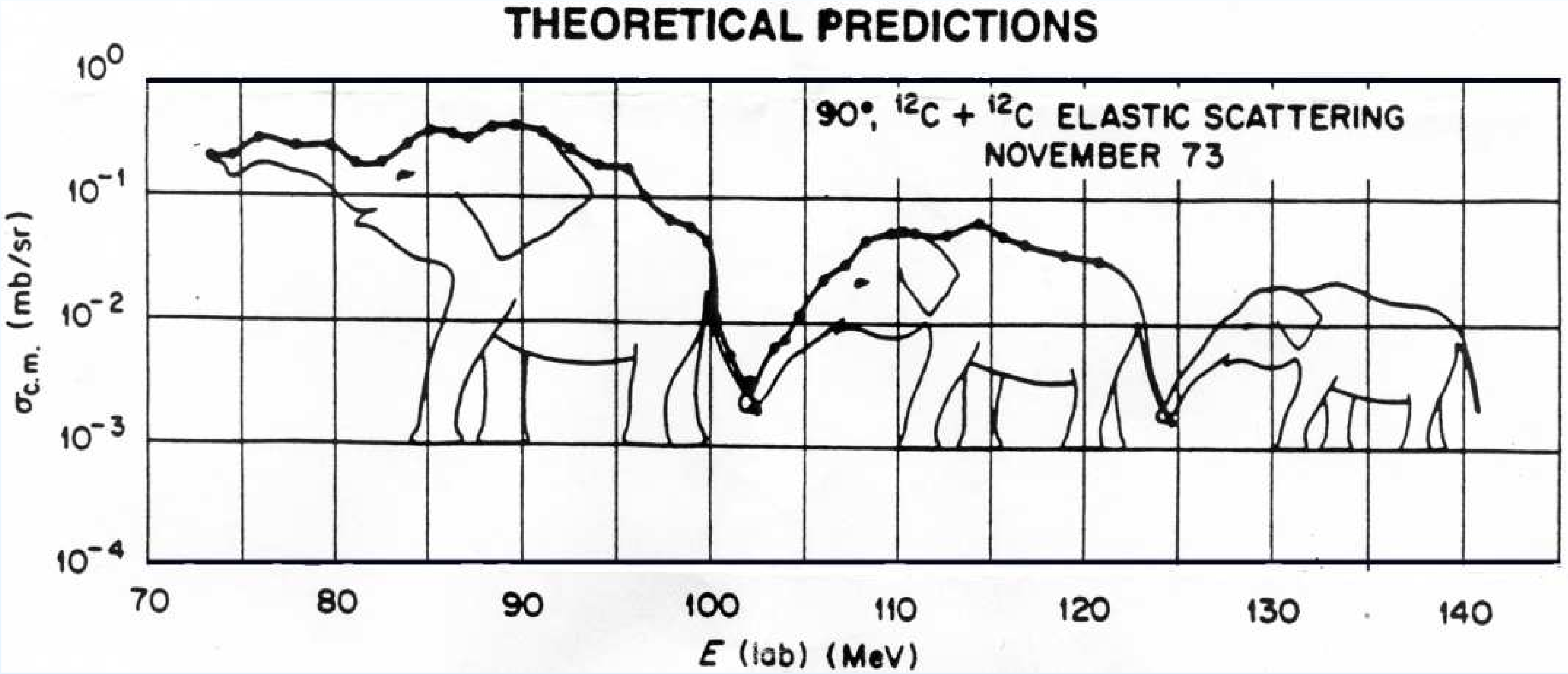
Inferential process



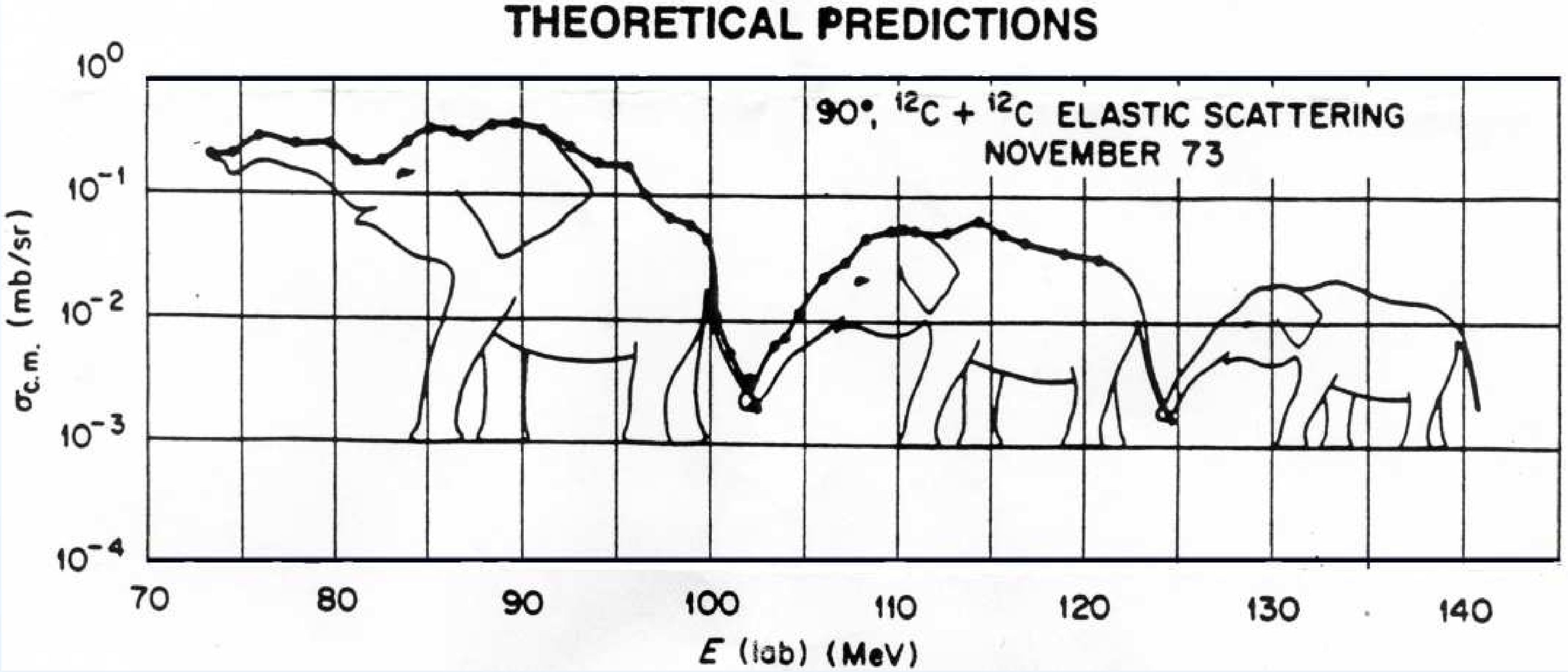
Inferential process



Inferential process



Inferential process



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

About predictions

Remember:

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

About predictions

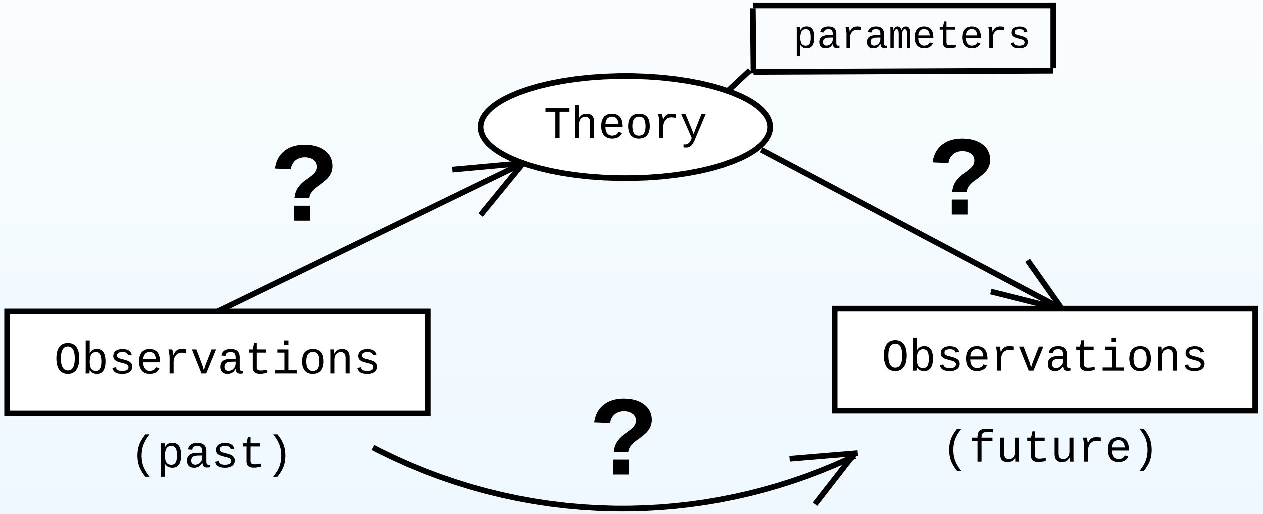
Remember:

"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)

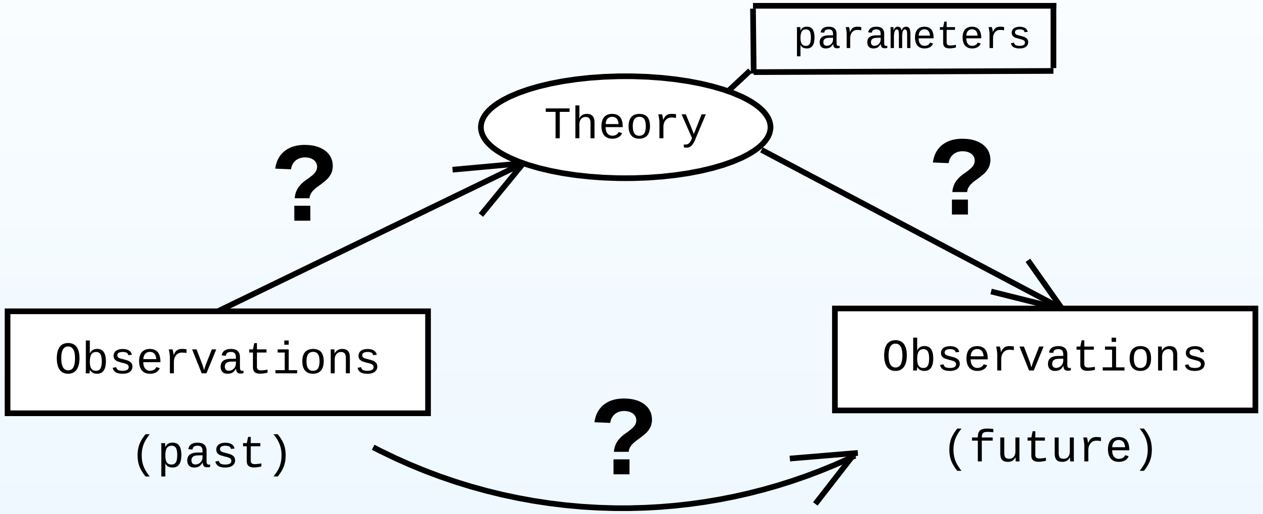
Deep source of uncertainty



Uncertainty:

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

Deep source of uncertainty



Uncertainty:

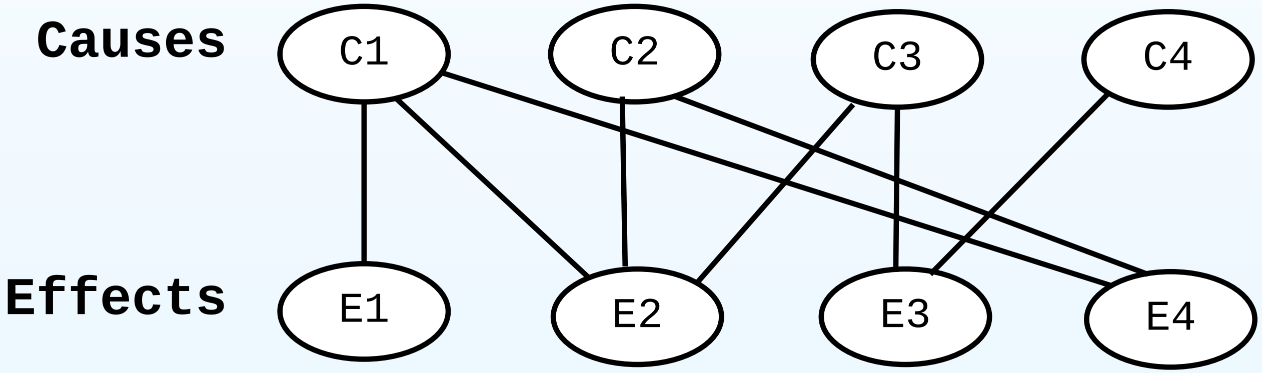
Theory	—?⟶	Future observations
Past observations	—?⟶	Theory
Theory	—?⟶	Future observations

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

CAUSE $\Longleftrightarrow$ EFFECT

Causes $\longrightarrow$ effects

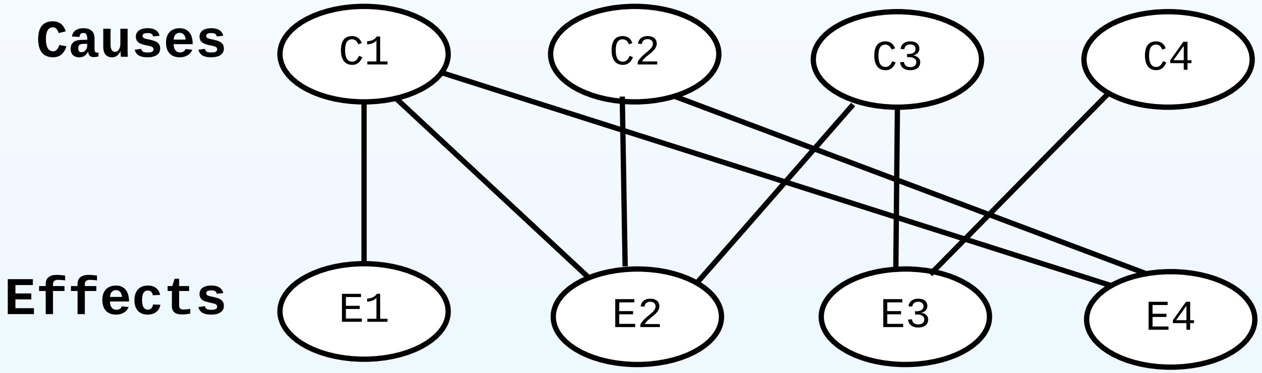
The same *apparent* cause might produce several,different effects



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

Causes ⟶ effects

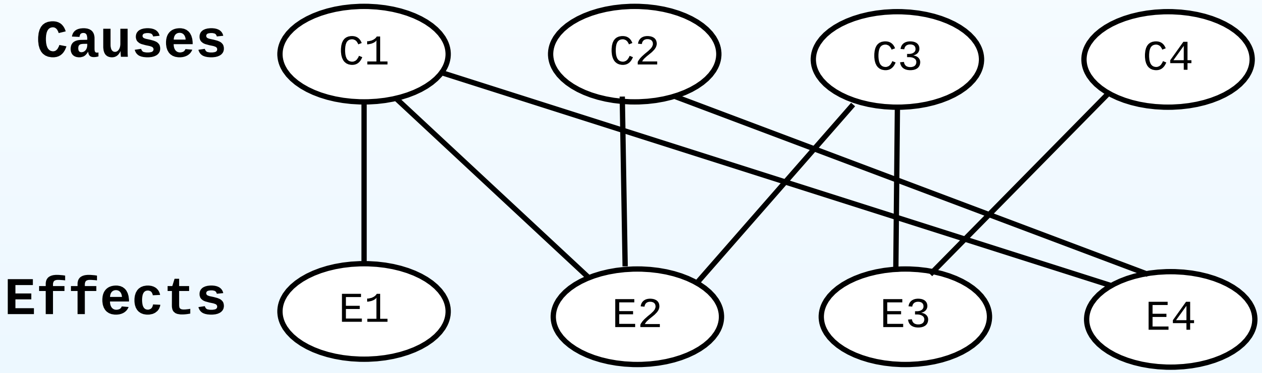
The same *apparent* cause might produce several,different effects



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."

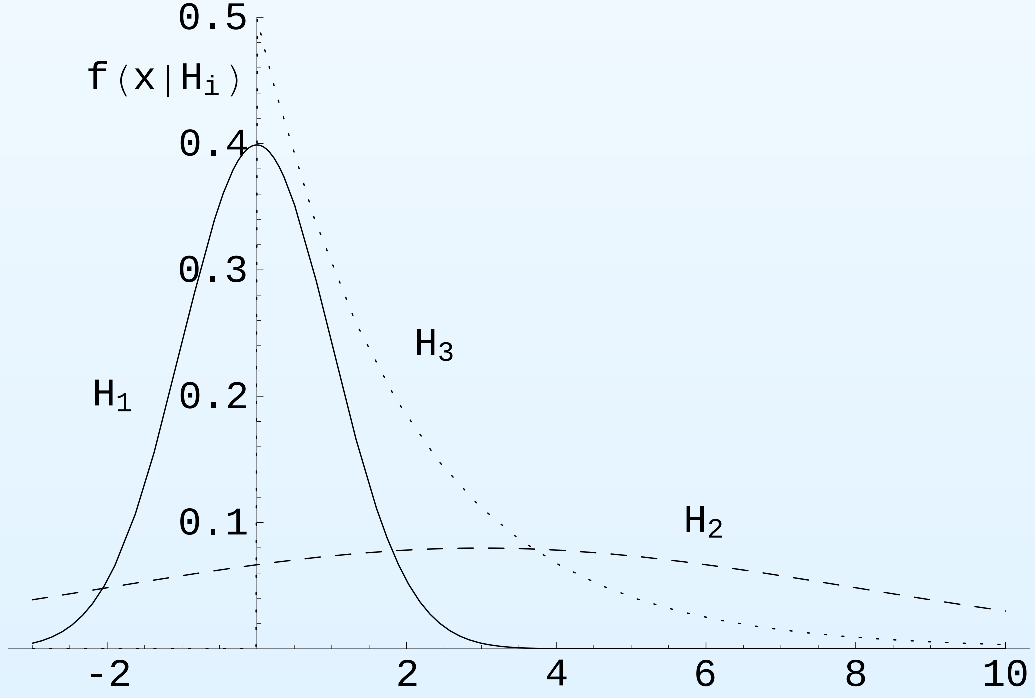
(H. Poincaré – *Science and Hypothesis*)

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 $\tau = 2$

$\Rightarrow$ 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*.

We shall come back to this example

Uncertainty

The human mind is used to live — and survive — in conditions of uncertainty and has developed mental categories to handle it.

Uncertainty

The human mind is used to live — and survive — in conditions of uncertainty and has developed mental categories to handle it.

As a matter of fact, although we are in a constant state of uncertainty about many events which might or might not occur,

- we can be "more or less *sure* — or *confident* — on something than on something else";
- "we consider something more or less *probable* (or *likely*)";
- or "we *believe* something more or less than something else".

Uncertainty

The human mind is used to live — and survive — in conditions of uncertainty and has developed mental categories to handle it.

As a matter of fact, although we are in a constant state of uncertainty about many events which might or might not occur,

- we can be "more or less *sure* — or *confident* — on something than on something else";
- "we consider something more or less *probable* (or *likely*)";
- or "we *believe* something more or less than something else".

We can use similar expressions, all referring to the intuitive idea of probability.

Uncertainty and probability

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

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

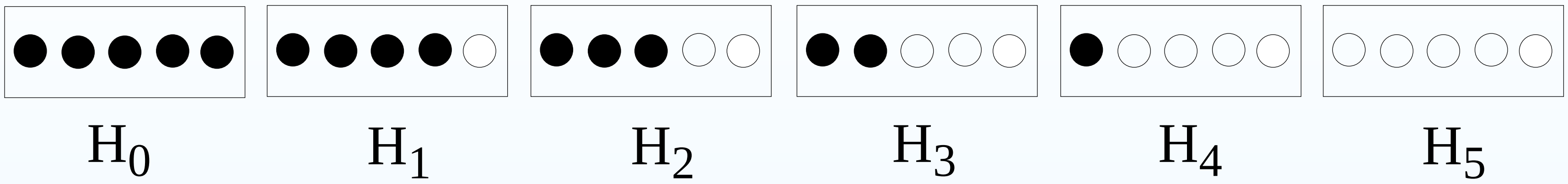
Uncertainty and probability

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

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

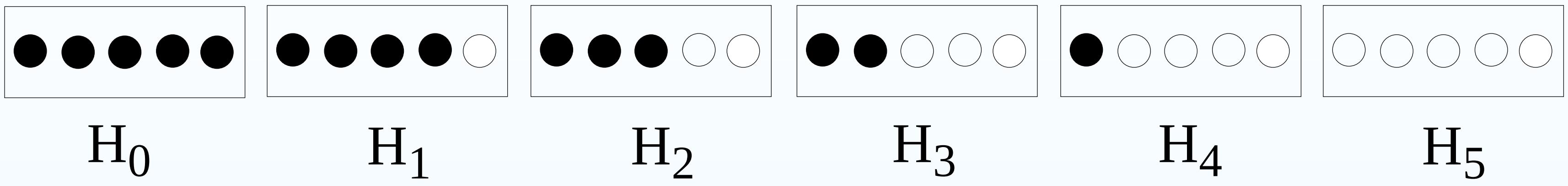
. . . thus, such statements are considered blaspheme to statistics gurus

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:

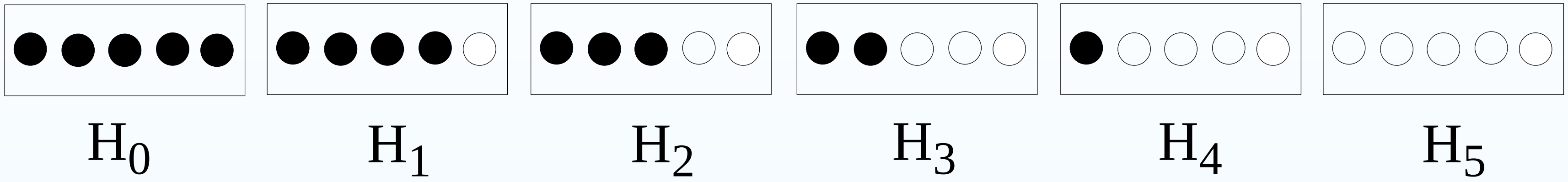
(a) Which box have we chosen, $H_0, H_1, \ldots, H_5$?

(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



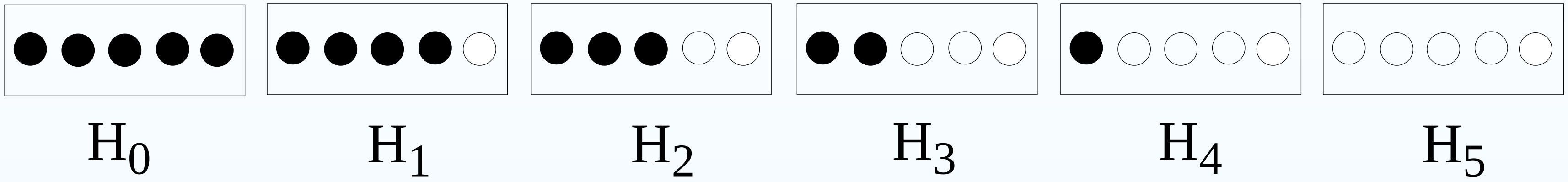
Let us take randomly one of the boxes.
We are in a state of uncertainty concerning several *events*, the most important of which correspond to the following questions:

(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)

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)

How to quantify all that?

How to quantify all that?

- **Falsificationist approach**

  [and statistical variations over the theme].

How to quantify all that?

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

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

Falsificationism

Usually associated to the name of Popper

and considered to be the *key to scientific progress*.

Falsificationism

Usually associated to the name of Popper

and considered to be the *key to scientific progress*.

$$\textbf{if } C_i \not\longrightarrow E, \textbf{ then } E_{\textbf{obs}} \not\longrightarrow C_i$$

⇒ Causes that cannot produce observed effects are ruled out (‘falsified’).

Falsificationism

Usually associated to the name of Popper

and considered to be the *key to scientific progress*.

$$\textbf{if } C_i \not\longrightarrow E, \textbf{ then } E_{\textbf{obs}} \not\longrightarrow C_i$$

⇒ Causes that cannot produce observed effects are ruled out ('falsified').

It seems OK, but it is naive for several aspects.

Falsificationism

Usually associated to the name of Popper

and considered to be the *key to scientific progress*.

$$\textbf{if } C_i \not\longrightarrow E, \textbf{ then } E_{\textbf{obs}} \not\longrightarrow C_i$$

⇒ Causes that cannot produce observed effects are ruled out (‘falsified’).

It seems OK, but it is naive for several aspects.

Let start realizing that the method is analogous with method of the proof by contradiction of classical, deductive logic.

Falsificationism

Usually associated to the name of Popper

and considered to be the *key to scientific progress*.

$$\text{if } C_i \not\longrightarrow E, \ \text{then } E_{\text{obs}} \not\longrightarrow C_i$$

⇒ Causes that cannot produce observed effects are ruled out ('falsified').

It seems OK, but it is naive for several aspects.

Let start realizing that the method is analogous with method of the proof by contradiction of classical, deductive logic.

- Assume that a hypothesis is true
- Derive 'all' logical consequence
- If (at least) one of the consequences is known to be false, then the hypothesis is declared false.

Falsificationism? OK, but. . .

- What to do of all hypotheses that are not falsified? (Limbus? Get stuck?)

Falsificationism? OK, but. . .

- What to do of all hypotheses that are not falsified? (Limbus? Get stuck?)
- What to do is nothing of what can be observed is incompatible with the hypothesis (or with many hypotheses)?

  E.g. H_i being a Gaussian $f(x \,|\, \mu_i, \sigma_i)$

  $\Rightarrow$ Given any pair or parameters $\{\mu_i, \sigma_i\}$, <u>all values</u> of x between $-\infty$ and $+\infty$ <u>are possible</u>.

  $\Rightarrow$ Having observed any value of x, <u>none</u> of H_i can be, strictly speaking, <u>falsified</u>.

Falsificationism and statistics

. . . then, statisticians have invented the "hypothesis tests"

Falsificationism and statistics

. . . then, statisticians have invented the “hypothesis tests”

in which the impossible is replaced by the improbable!

Falsificationism and statistics

. . . then, statisticians have invented the "hypothesis tests"

in which the impossible is replaced by the improbable!

But from the impossible to the improbable there is not just a question of quantity, but a question of quality.

This mechanism, logically flawed, is particularly perverse, because deeply rooted in most people, due to education, but is not supported by logic.

⇒ Basically responsible of all fake claims of discoveries in the past decades.

[I am particularly worried about claims concerning our health, or the status of the planet, of which I have no control of the experimental data.]

In summary

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

$\Rightarrow C_i$ is impossible ('false')

In summary

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"

In summary

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"

In summary

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"

Example 1

Playing lotto

H: "I play honestly at lotto, betting on a rare combination"

E: "I win"

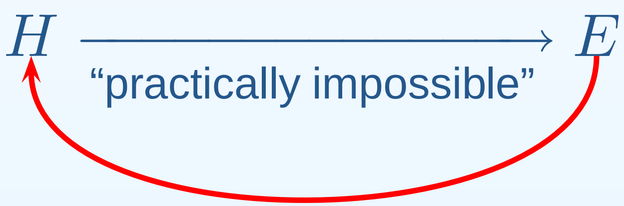
$$H \xrightarrow[\text{"practically impossible"}]{} E$$

Example 1

Playing lotto

H: "I play honestly at lotto, betting on a rare combination"

E: "I win"



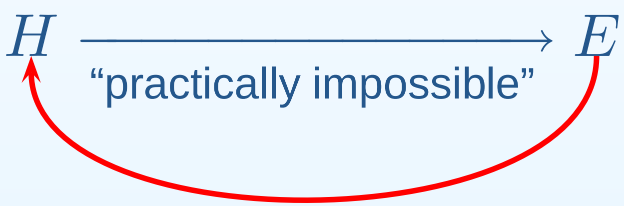
"practically to exclude"

Example 1

Playing lotto

H: "I play honestly at lotto, betting on a rare combination"

E: "I win"



"practically to exclude"

$\Rightarrow$ almost certainly I have cheated. . .

(or it is false that I won. . .)

Example 2

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

Toy 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

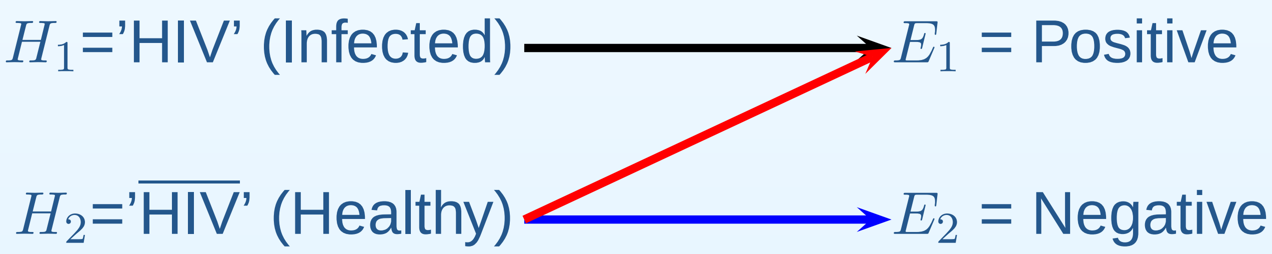
H_2='$\overline{\text{HIV}}$' (Healthy) E_2 = Negative

Example 2

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

Toy 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 2

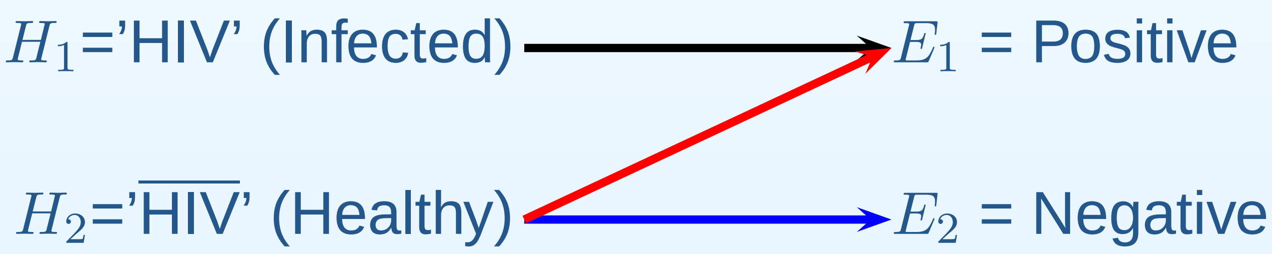
An Italian citizen is selected at random to undergo an AIDS test. Performance of clinical trial is not perfect, as customary.

Toy model:

$$P(\text{Pos} \,|\, \text{HIV}) = 100\%$$

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

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



Result: $\Rightarrow$ Positive

Example 2

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

Toy 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: $\Rightarrow$ Positive

Infected or healthy?

Example 2

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”?

Example 2

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" ?

Example 2

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”?

Example 2

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.”

?

Example 2

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)

Example 2

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)

Example 2

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)

... which might result into very bad decisions!

Similar arbitrary inversion in upper limits

Imagine we have done a counting experiment, believed to be described by a Poisson distribution.

- Result $x = 0$

$\Rightarrow$ What can we tell about λ?
(Remember that the physical parameter is $r = \lambda/\Delta T$)

Similar arbitrary inversion in upper limits

Imagine we have done a counting experiment, believed to be described by a Poisson distribution.

- Result $x = 0$

$\Rightarrow$ What can we tell about λ?
(Remember that the physical parameter is $r = \lambda/\Delta T$)

- All values of λ (or r) are in principle possible
- ...although, we do not believe them equally likely.

Similar arbitrary inversion in upper limits

Imagine we have done a counting experiment, believed to be described by a Poisson distribution.

- Result $x = 0$

$\Rightarrow$ What can we tell about λ?
(Remember that the physical parameter is $r = \lambda/\Delta T$)

- All values of λ (or r) are in principle possible
- ...although, we do not believe them equally likely.
- Standard way to report the result: 95% C.L. upper limit:

$$\lambda \leq 3 \text{ @ } 95\% \text{C.L.}$$

- Why?

Similar arbitrary inversion in upper limits

“Because if I repeat a large number of experiments, I get $x = 0$ *in 5% of the cases”*

Similar arbitrary inversion in upper limits

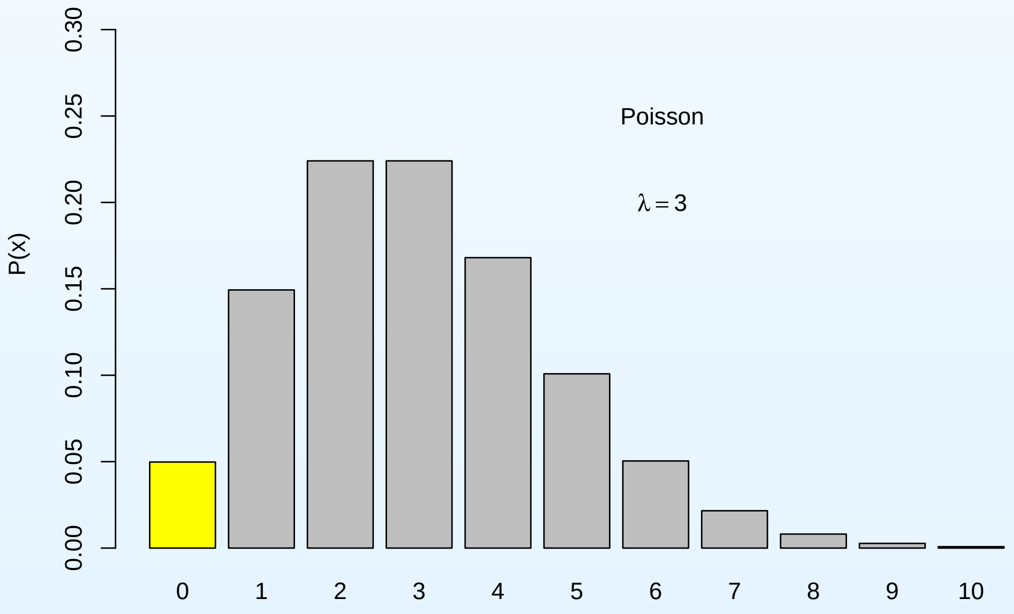
"Because if I repeat a large number of experiments,
I get $x = 0$ in 5% of the cases"

$$\Rightarrow \; P(x = 0 \,|\, \mathcal{P}_{\lambda=3}) = 5\%.$$

Similar arbitrary inversion in upper limits

"Because if I repeat a large number of experiments, I get $x = 0$ in 5% of the cases"

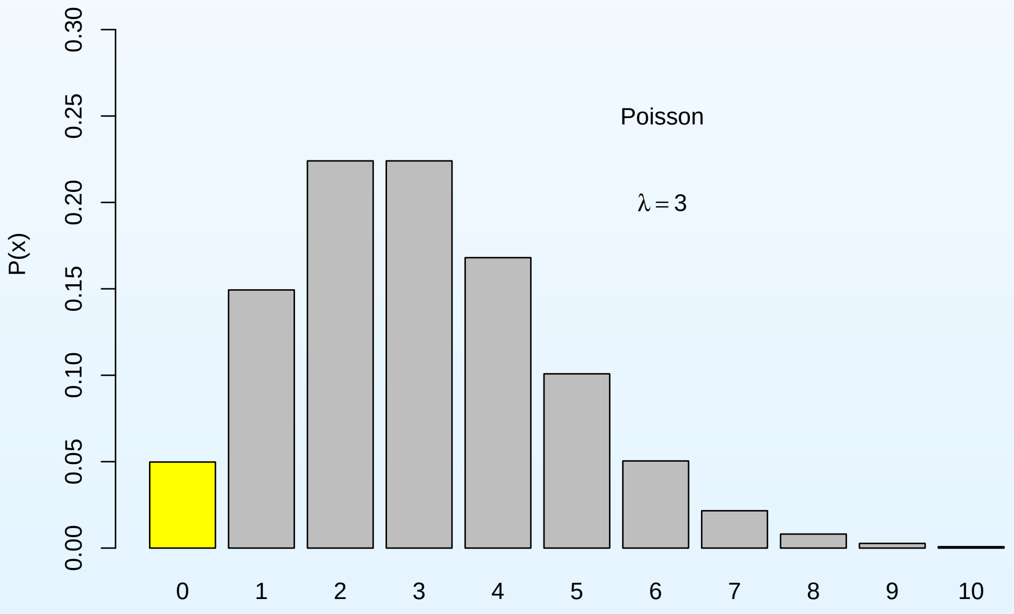
$$\Rightarrow \; P(x = 0 \,|\, \mathcal{P}_{\lambda=3}) = 5\%.$$



Similar arbitrary inversion in upper limits

"Because if I repeat a large number of experiments,
I get $x = 0$ in 5% of the cases"

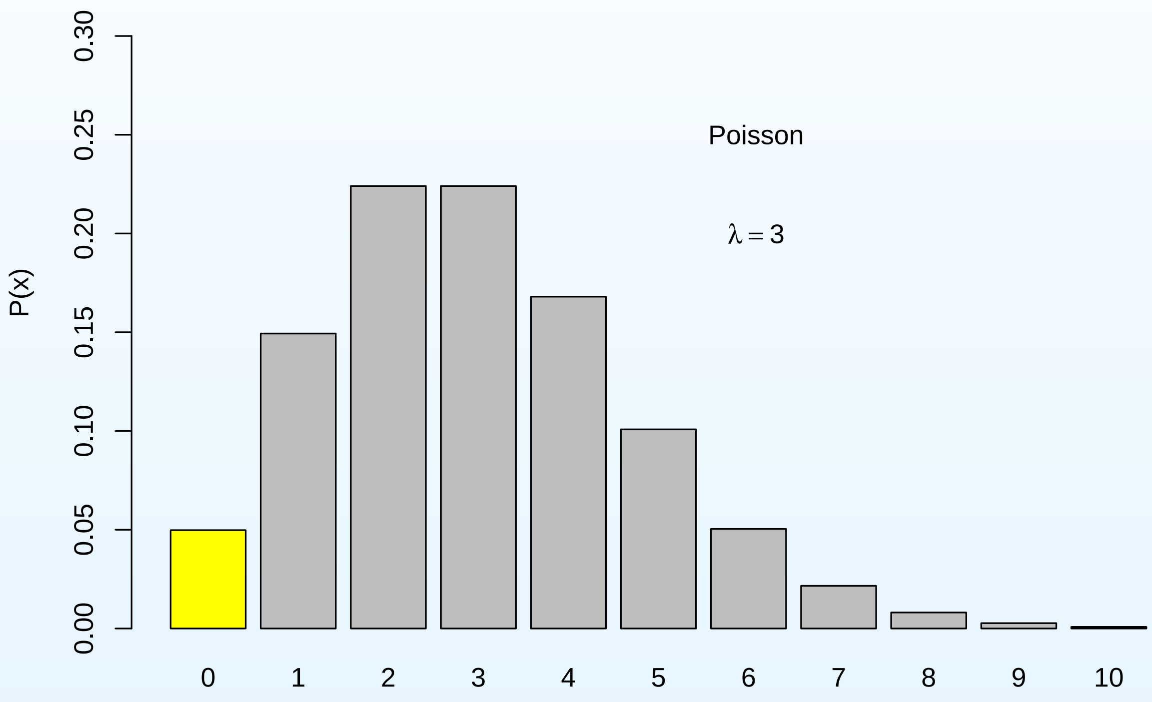
$$\Rightarrow\ P(x = 0 \,|\, \mathcal{P}_{\lambda=3}) = 5\%.$$



But what has this to do with our confidence that $\lambda \geq 3$?

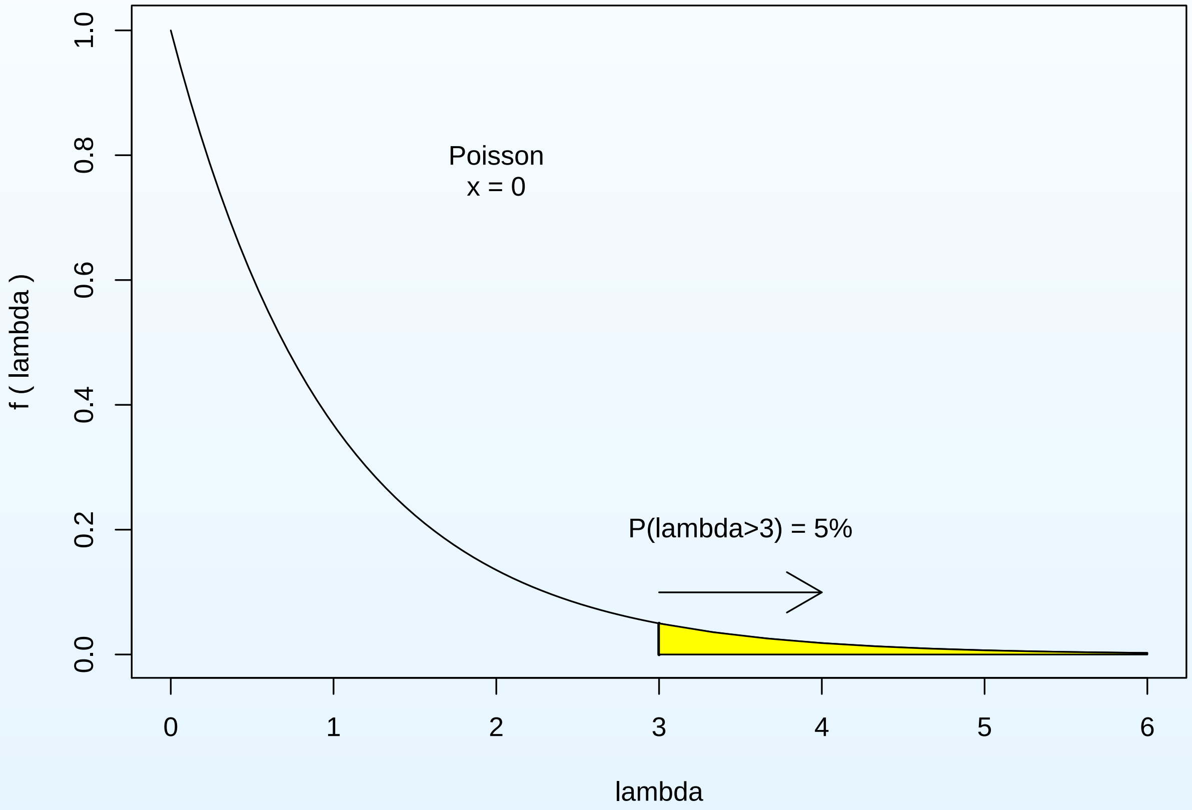
Special case of the Poisson with observed $x = 0$

Probability function of x given $\lambda = 3$



Special case of the Poisson with observed $x = 0$

Probability density function of λ given $x = 0$



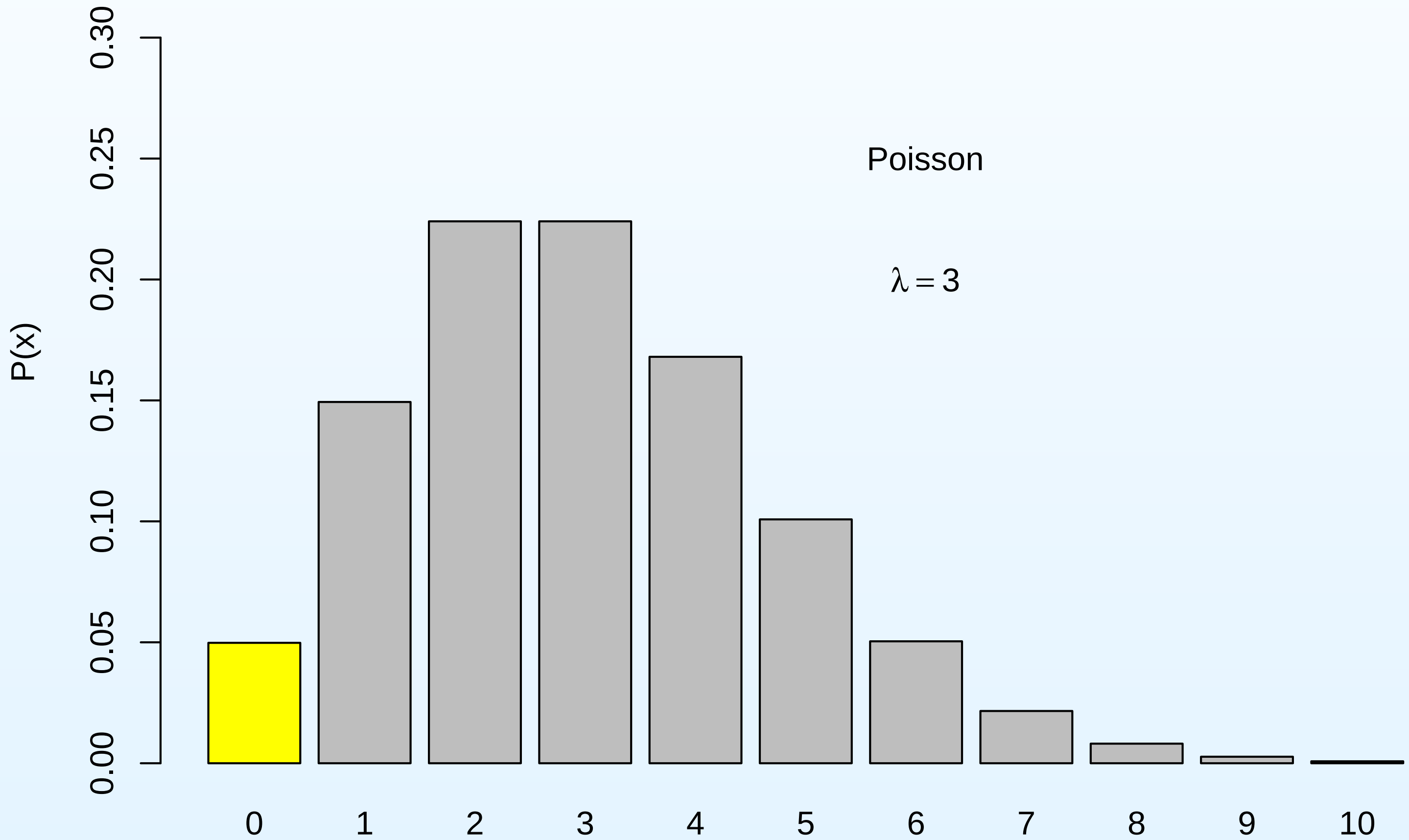
(We shall came later to the details of the calculation)

. . . but

It is not a general property

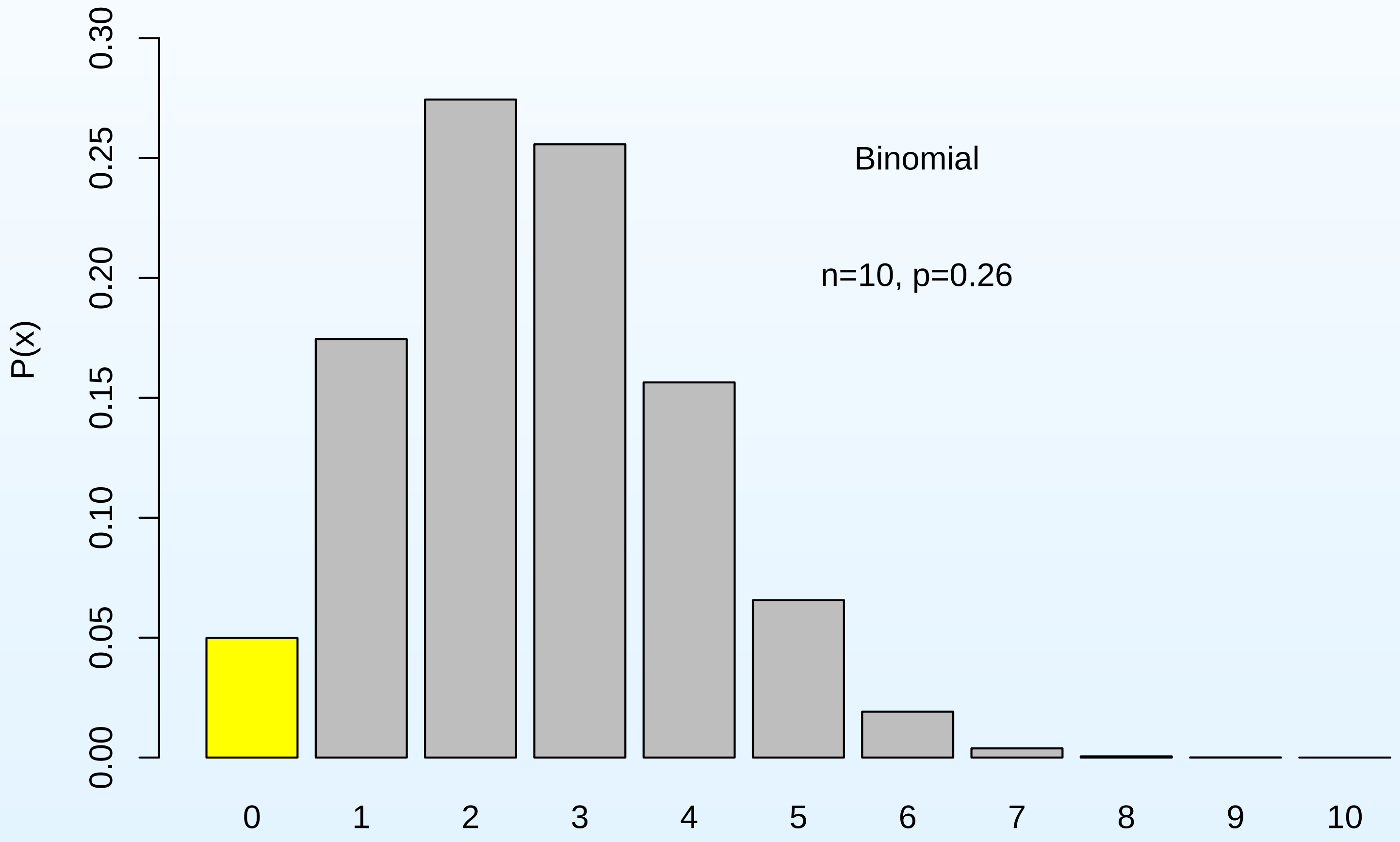
Let us check with other simple cases

A Poisson distribution with $\lambda = 3$



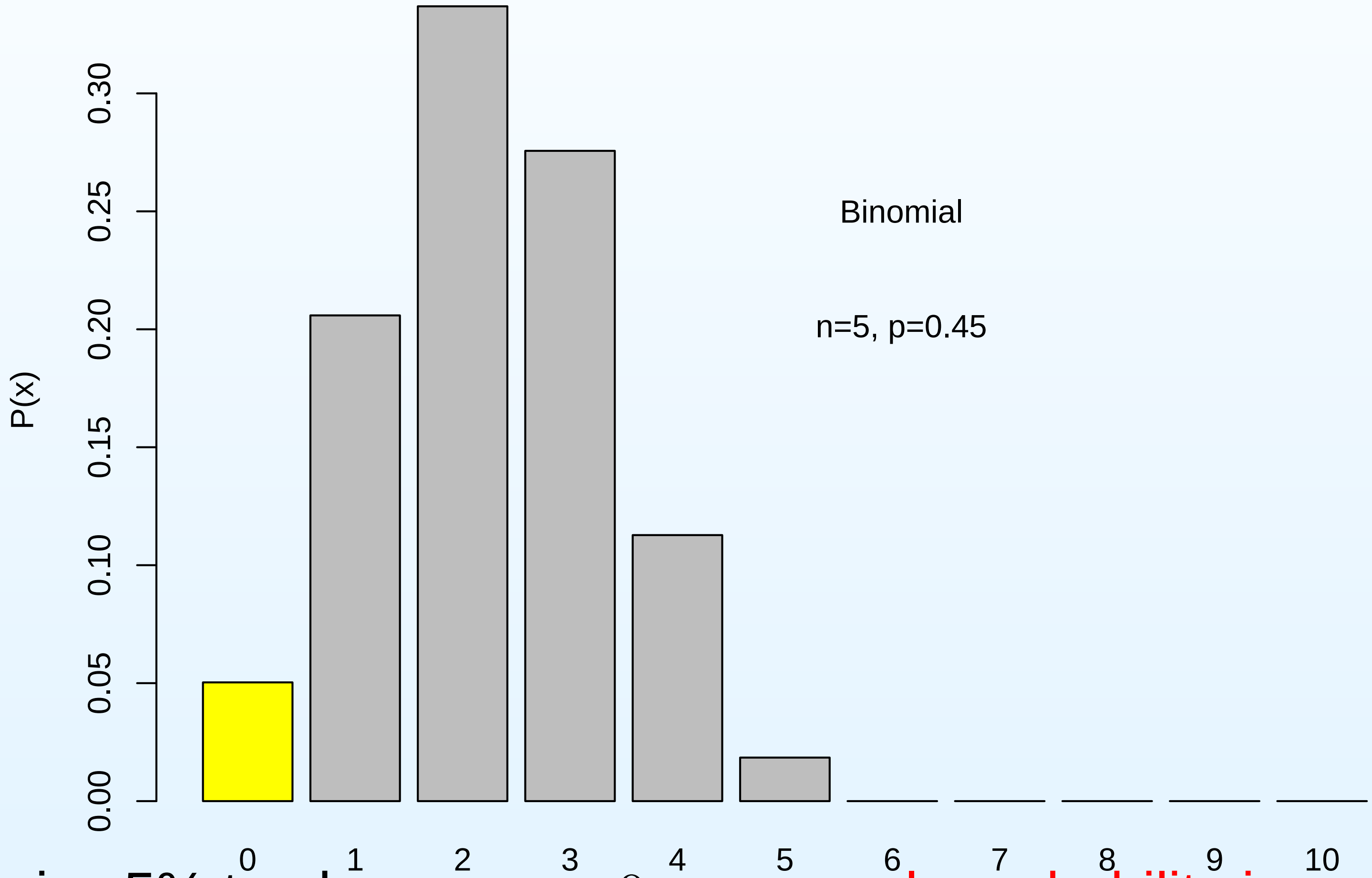
Let us check with other simple cases

A binomial distribution with $n = 10$ and $p = 0.26$



Let us check with other simple cases

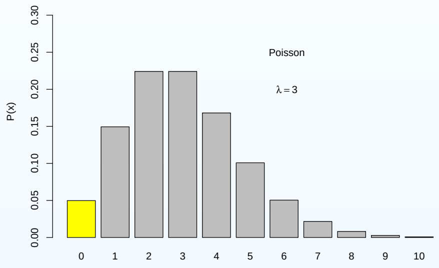
A binomial distribution with $n = 5$ and $p = 0.45$

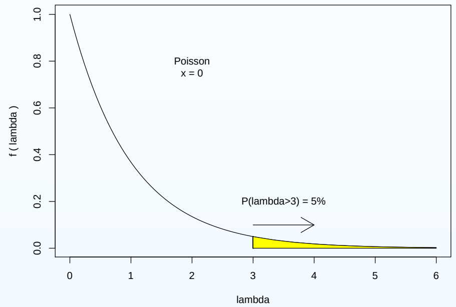


All give 5% to observe $x = 0$ $\Rightarrow$ apply probability inversion $\longrightarrow$

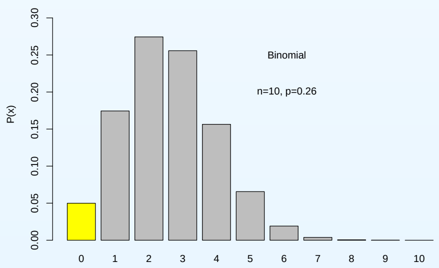
The game does not work already with the binomial

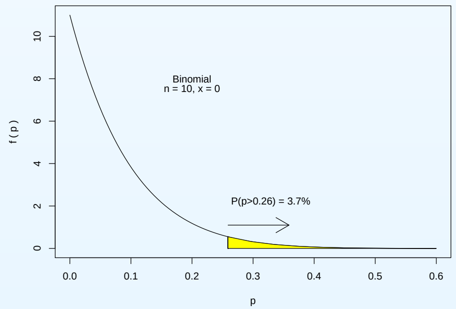
'$\lambda_L = 3$': $P(x=0\,|\,\lambda_L) = 5\%$ $\qquad$ $P(\lambda \geq \lambda_L) = 5\%$ ✓



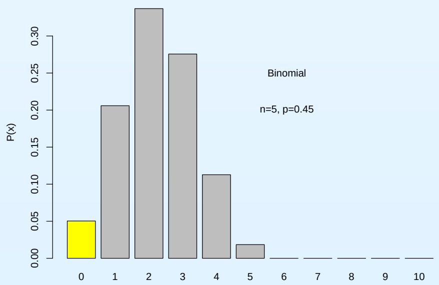


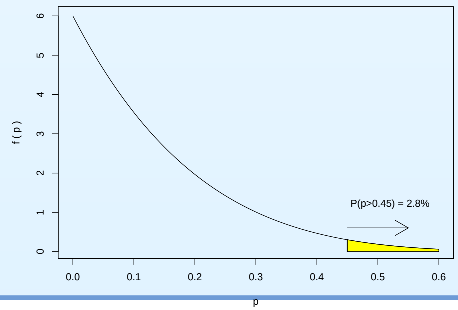
'$p_L = 0.26$': $P(x=0\,|\,p_L) = 5\%$ $\qquad$ but $P(p \geq p_L) = 3.7\%$!





'$p_L = 0.45$': $P(x=0\,|\,p_L) = 5\%$ $\qquad$ but $P(p \geq p_L) = 2.8\%$!





Statisticians are clever!

This is not yet the end of the story.

Statisticians are clever!

This is not yet the end of the story.
In many cases: $H_i \longrightarrow$ large number of $\{E_j\}$:

Statisticians are clever!

This is not yet the end of the story.
In many cases: $H_i \longrightarrow$ large number of $\{E_j\}$:

- Each effect has little probability $\rightarrow$ 'practically improbable'

$\Rightarrow$ whatever we observe is an evidence against the hypothesis

Statisticians are clever!

This is not yet the end of the story.
In many cases: $H_i \longrightarrow$ large number of $\{E_j\}$:

- Each effect has little probability $\rightarrow$ ‘practically improbable’

$\Rightarrow$ whatever we observe is an evidence against the hypothesis

- Even those who trust the (flawed) reasoning based on the small probability of effects have to realize that the reasoning fails in these cases.

Statisticians are clever!

This is not yet the end of the story.
In many cases: $H_i \longrightarrow$ large number of $\{E_j\}$:

- Each effect has little probability $\rightarrow$ 'practically improbable'

$\Rightarrow$ whatever we observe is an evidence against the hypothesis

- Even those who trust the (flawed) reasoning based on the small probability of effects have to realize that the reasoning fails in these cases.

$\Rightarrow$ **statistician 'way out'**: individual observable effects are replaced by two *sets of effects*, one of high chance to happen, the other of low chance ('the tail(s) of the distribution')
$\rightarrow$ the reasoning is extended to these two sets of effects

Statisticians are clever!

This is not yet the end of the story.
In many cases: $H_i \longrightarrow$ large number of $\{E_j\}$:

- Each effect has little probability $\rightarrow$ 'practically improbable'

$\Rightarrow$ whatever we observe is an evidence against the hypothesis

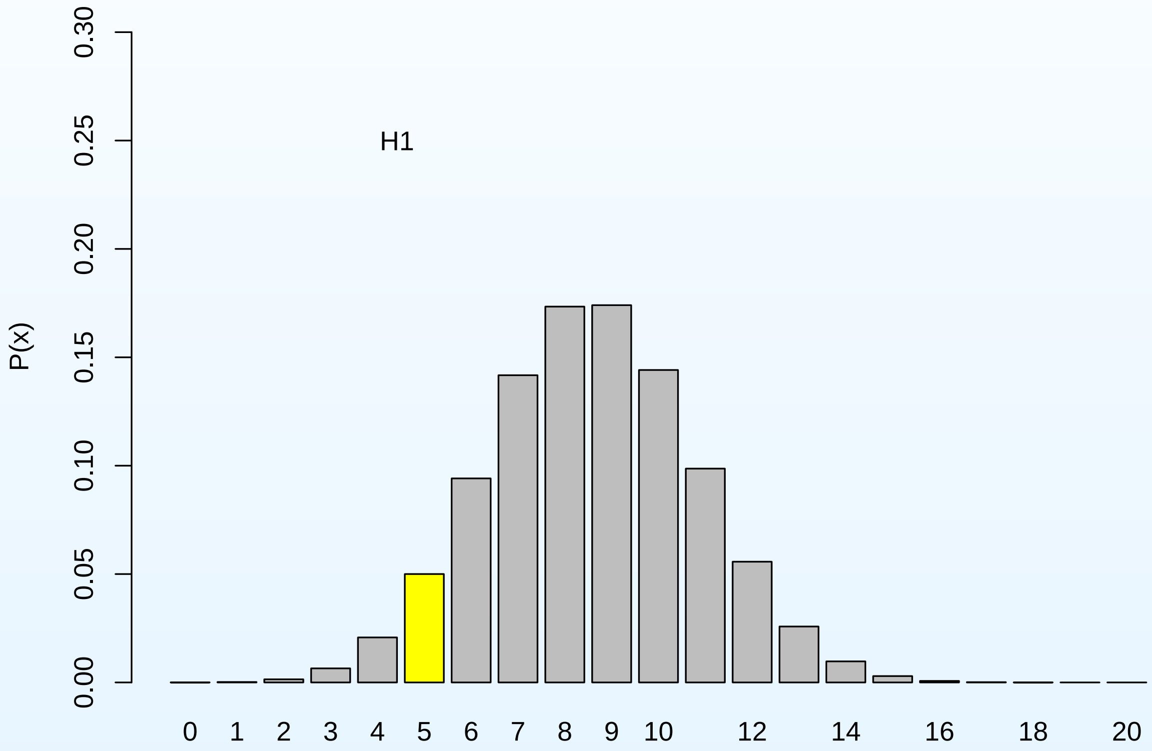
- Even those who trust the (flawed) reasoning based on the small probability of effects have to realize that the reasoning fails in these cases.

$\Rightarrow$ **statistician 'way out'**: individual observable effects are replaced by two *sets of effects*, one of high chance to happen, the other of low chance ('the tail(s) of the distribution')
$\rightarrow$ the reasoning is extended to these two sets of effects

$\Rightarrow$ Logically, the situation worsens:
$\rightarrow$ conclusions depend not only on on observed effects, but also on non-observed effects!

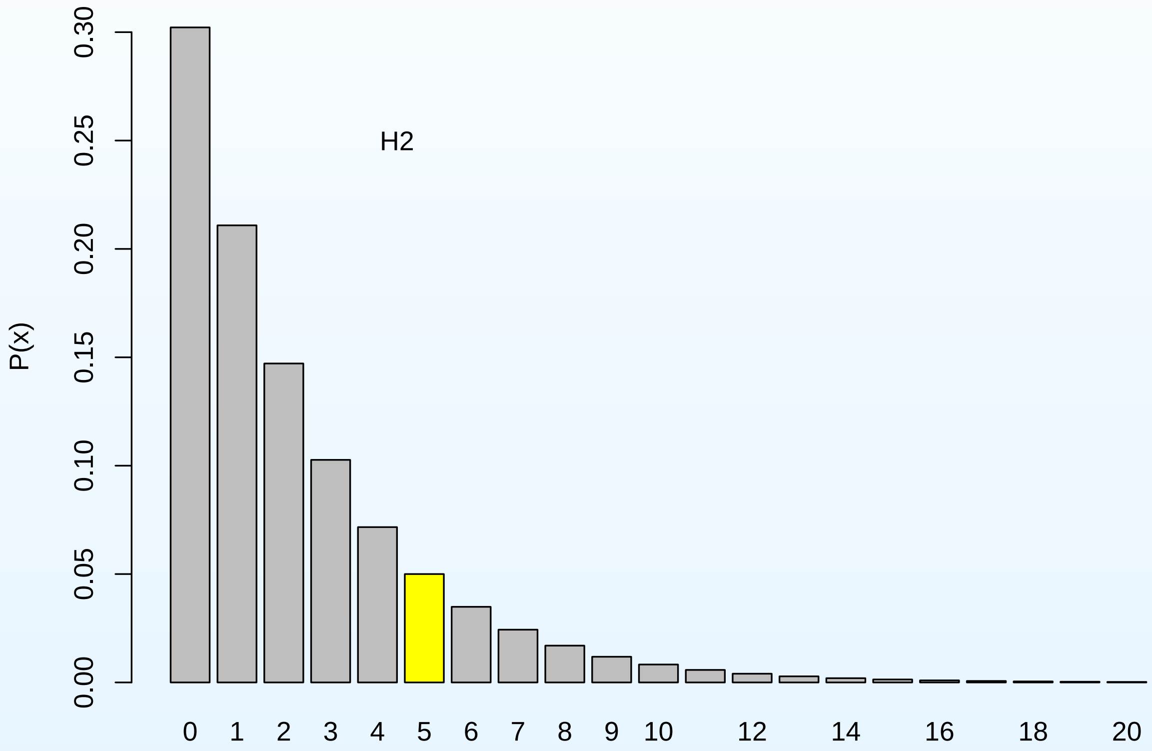
Observed value and tails

Several hypotheses to be tested against the observation $x = 5$



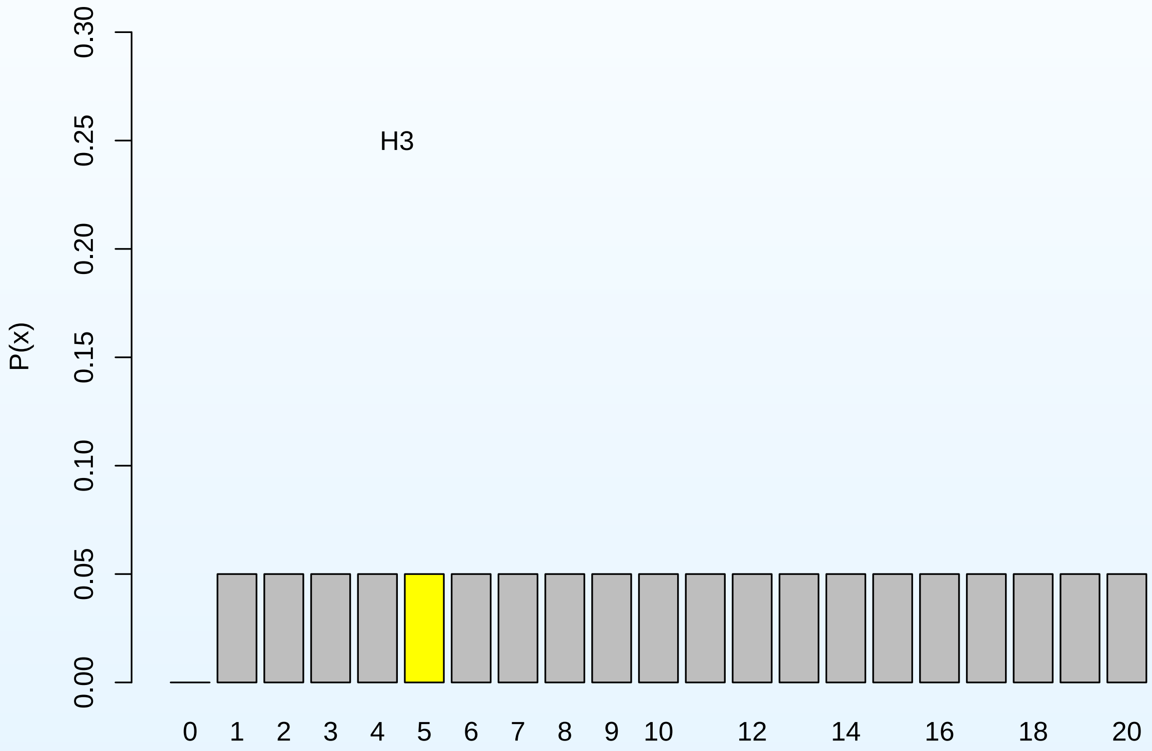
Observed value and tails

Several hypotheses to be tested against the observation $x = 5$



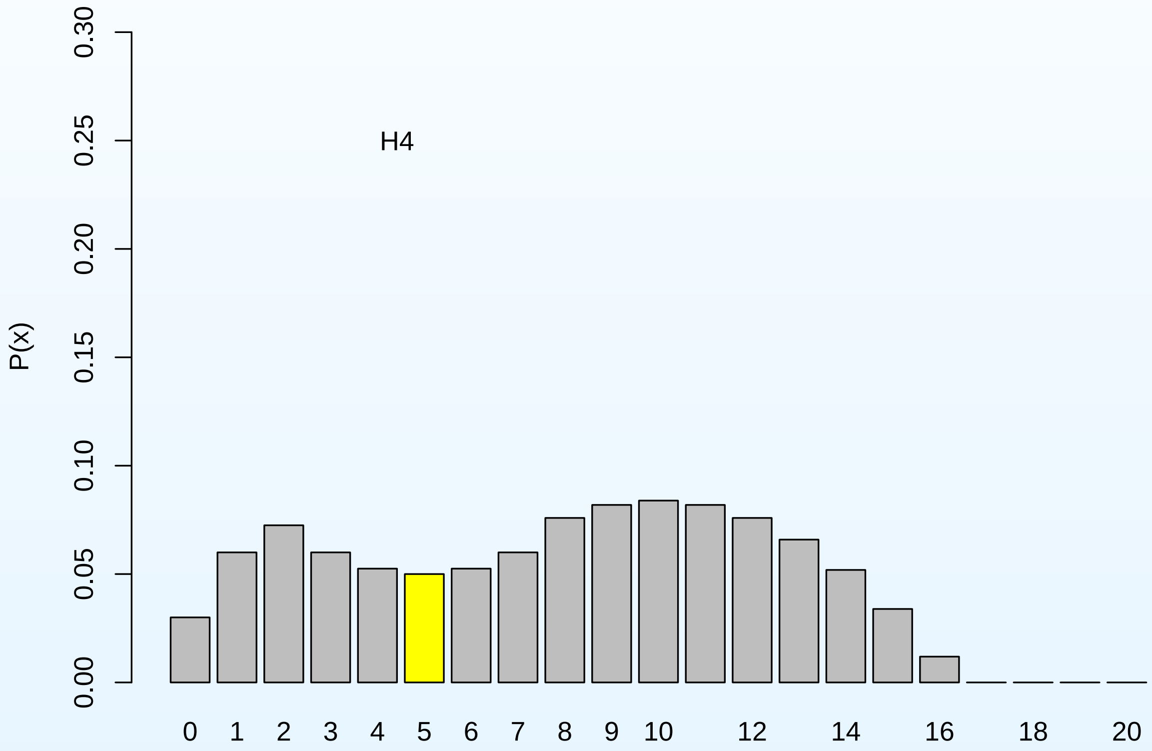
Observed value and tails

Several hypotheses to be tested against the observation $x = 5$



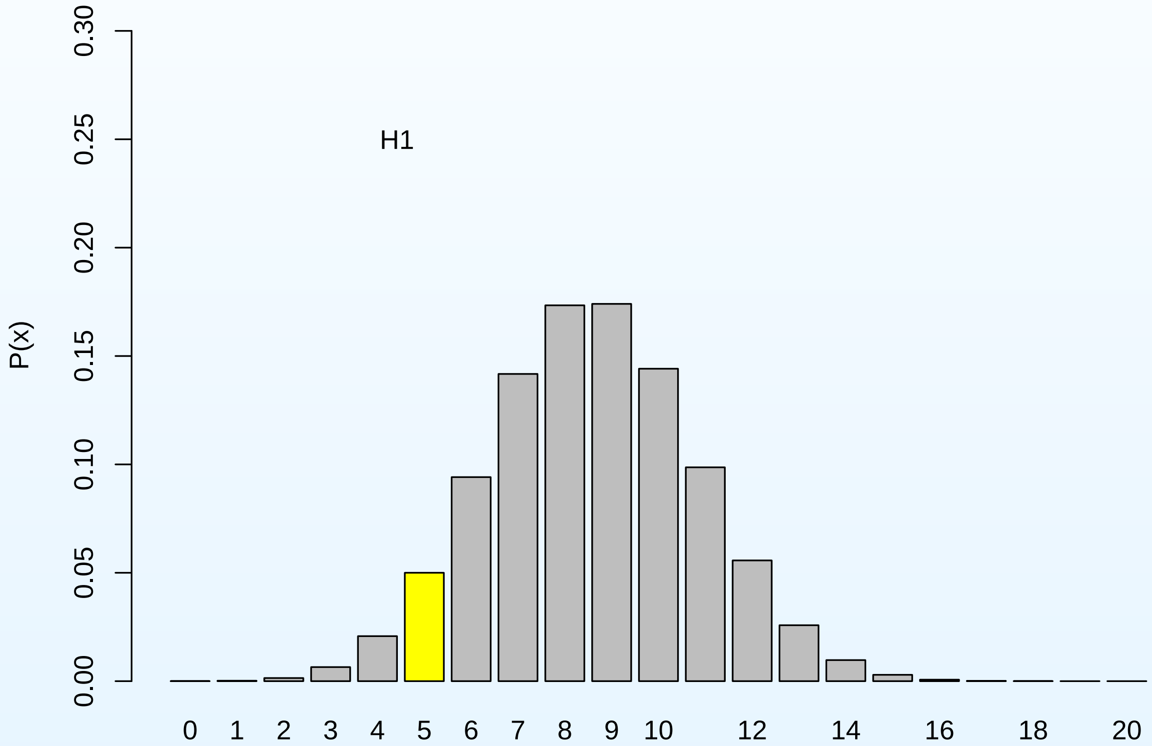
Observed value and tails

Several hypotheses to be tested against the observation $x = 5$



Observed value and tails

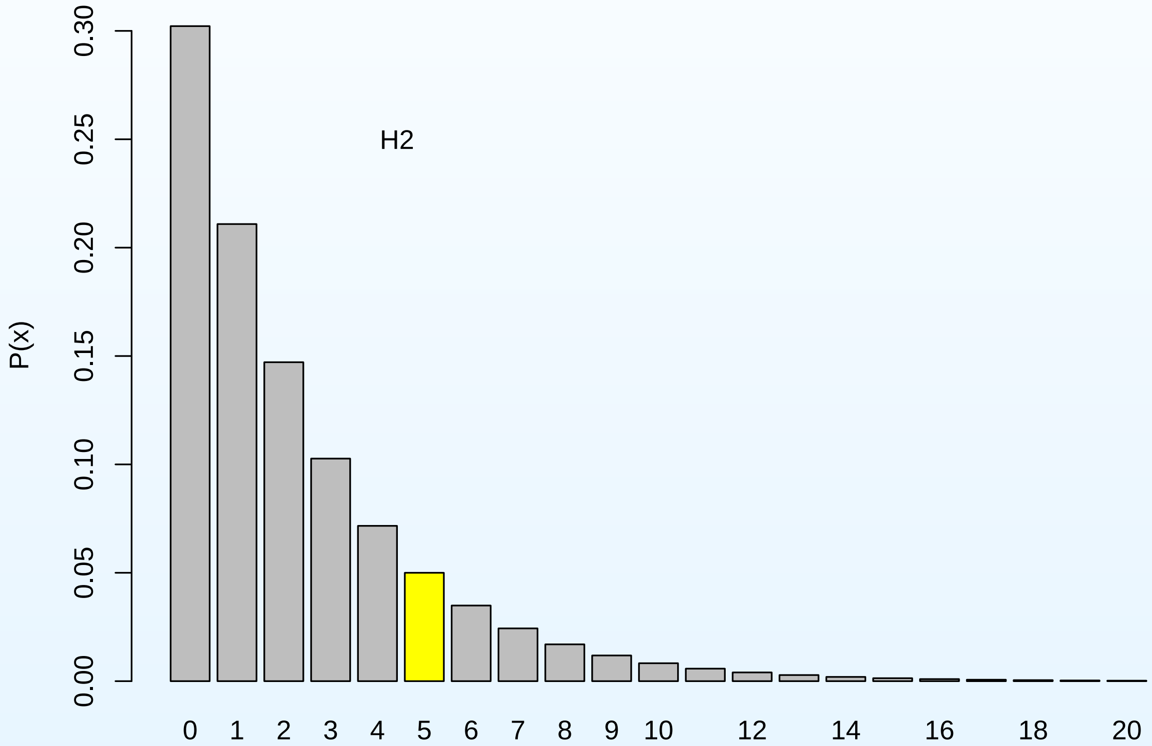
All have the same probability to give $x = 5$



Natural that our conclusions depend 'somehow' on $P(x = 5 \,|\, H_i)$

Observed value and tails

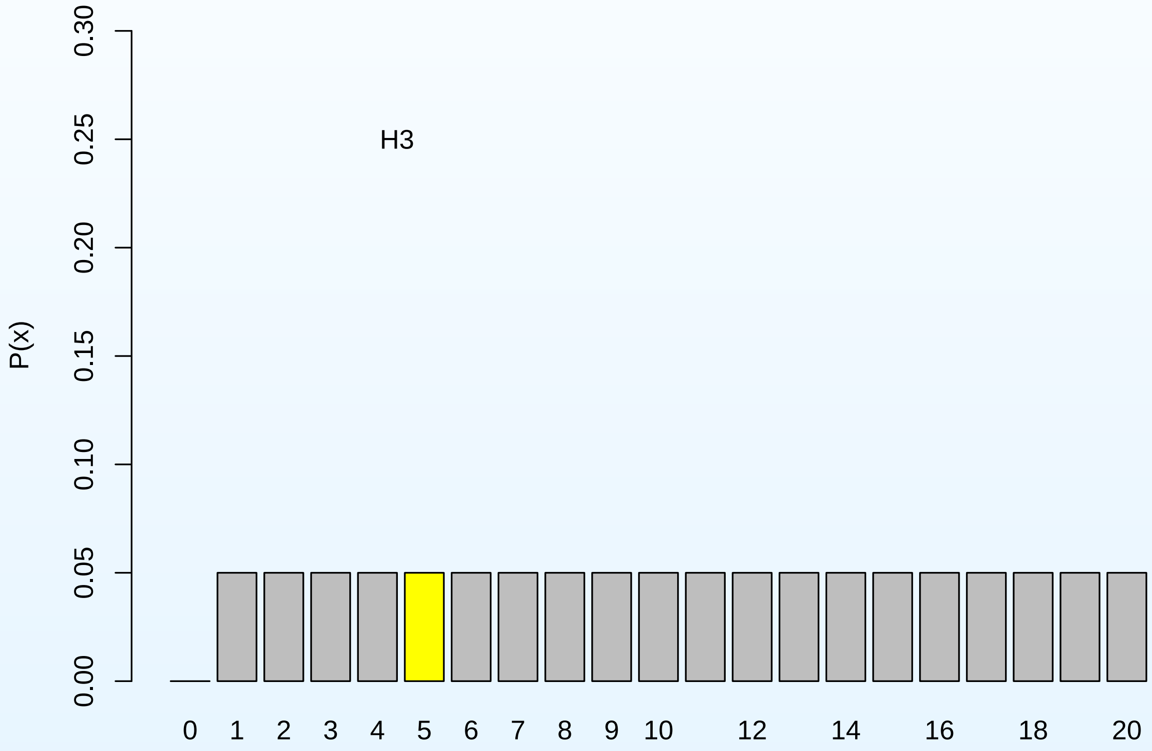
All have the same probability to give $x = 5$



Natural that our conclusions depend ‘somehow’ on $P(x = 5 \mid H_i)$

Observed value and tails

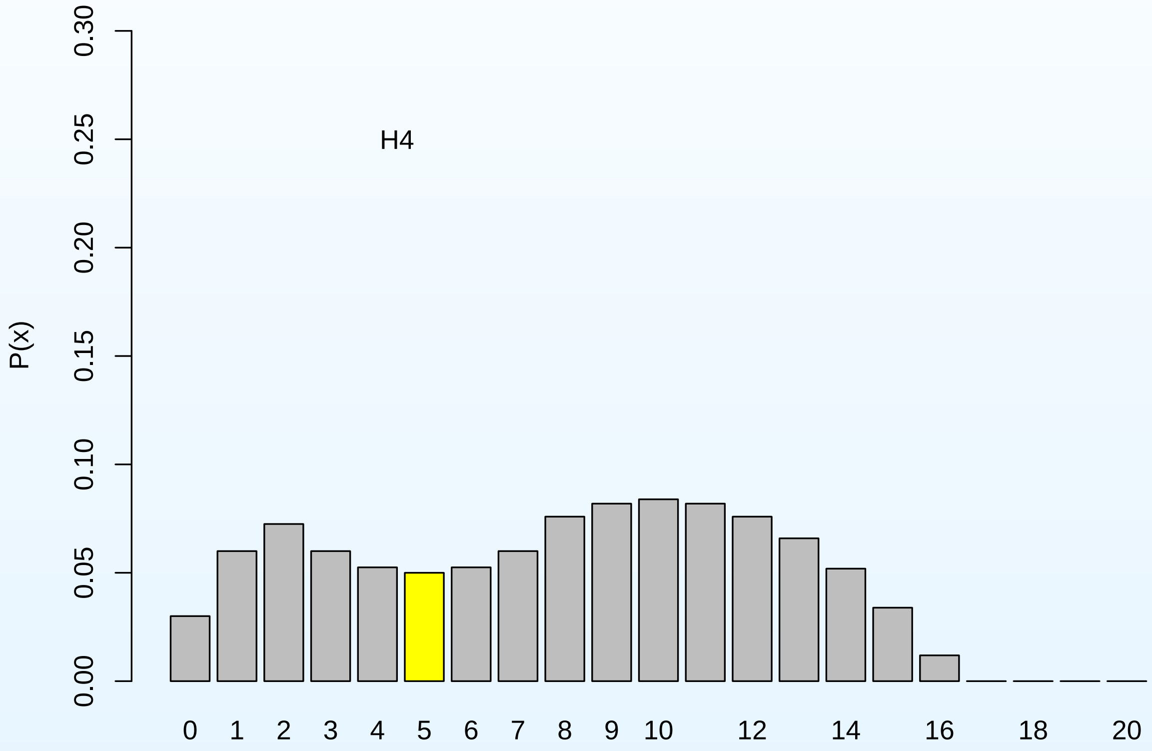
All have the same probability to give $x = 5$



Natural that our conclusions depend 'somehow' on $P(x = 5 \,|\, H_i)$

Observed value and tails

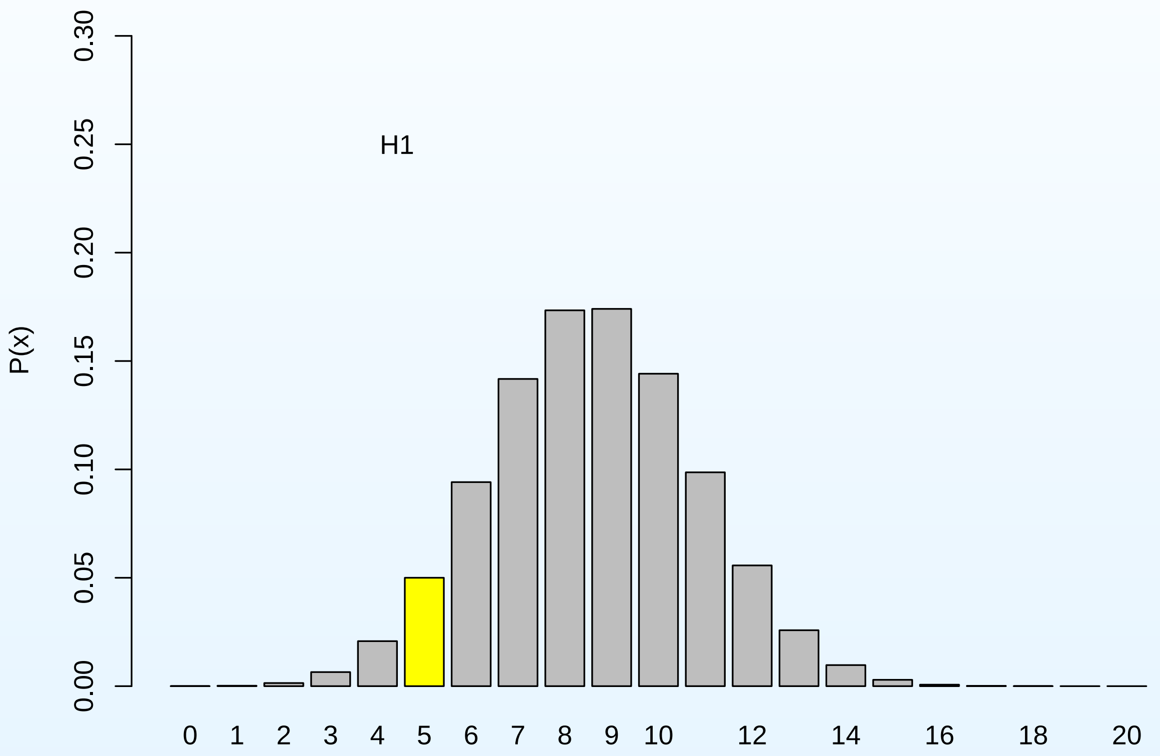
All have the same probability to give $x = 5$



Natural that our conclusions depend 'somehow' on $P(x = 5 \,|\, H_i)$

Observed value and tails

All have the same probability to give $x = 5$

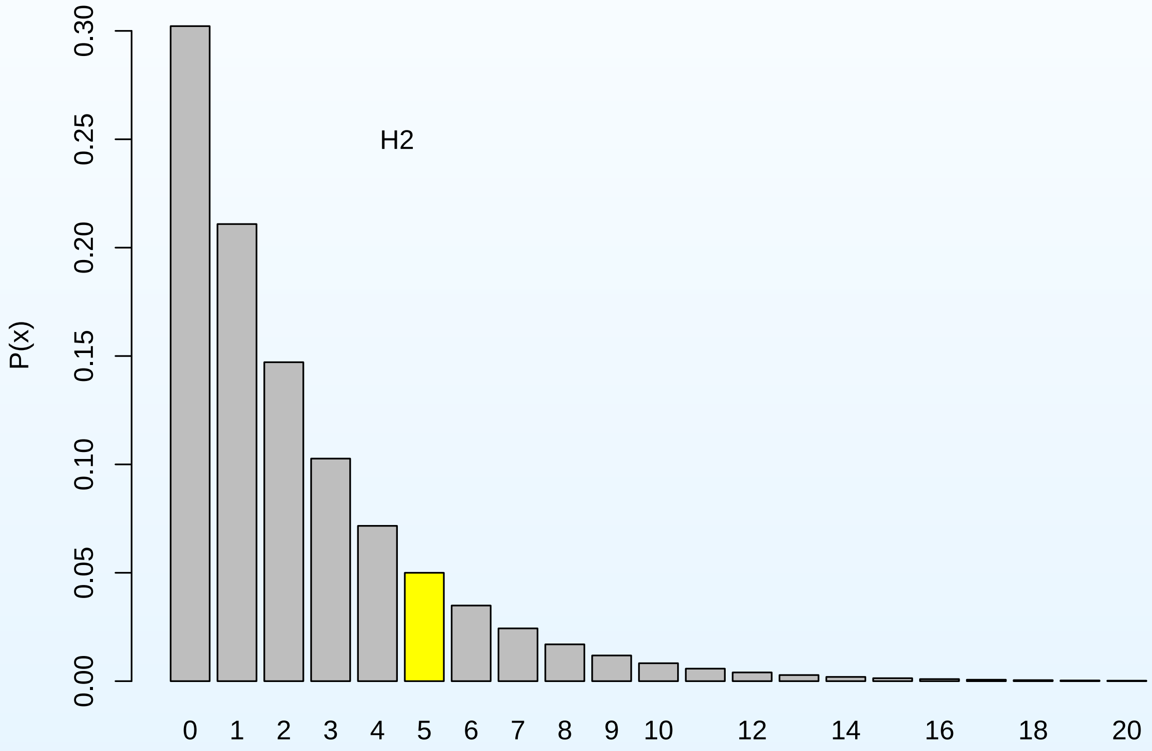


Natural that our conclusions depend ‘somehow’ on $P(x = 5 \mid H_i)$

(and in this case we learn nothing!)

Observed value and tails

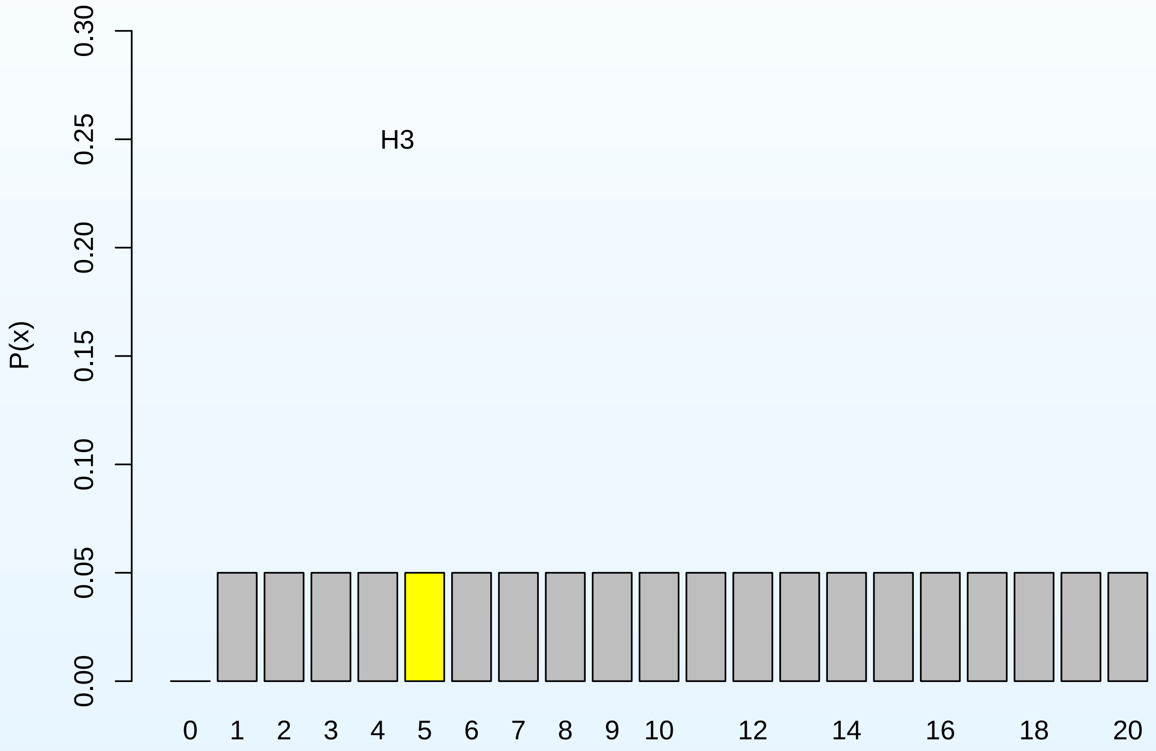
All have the same probability to give $x = 5$



Natural that our conclusions depend 'somehow' on $P(x = 5 \,|\, H_i)$
(and in this case we learn nothing!)

Observed value and tails

All have the same probability to give $x = 5$

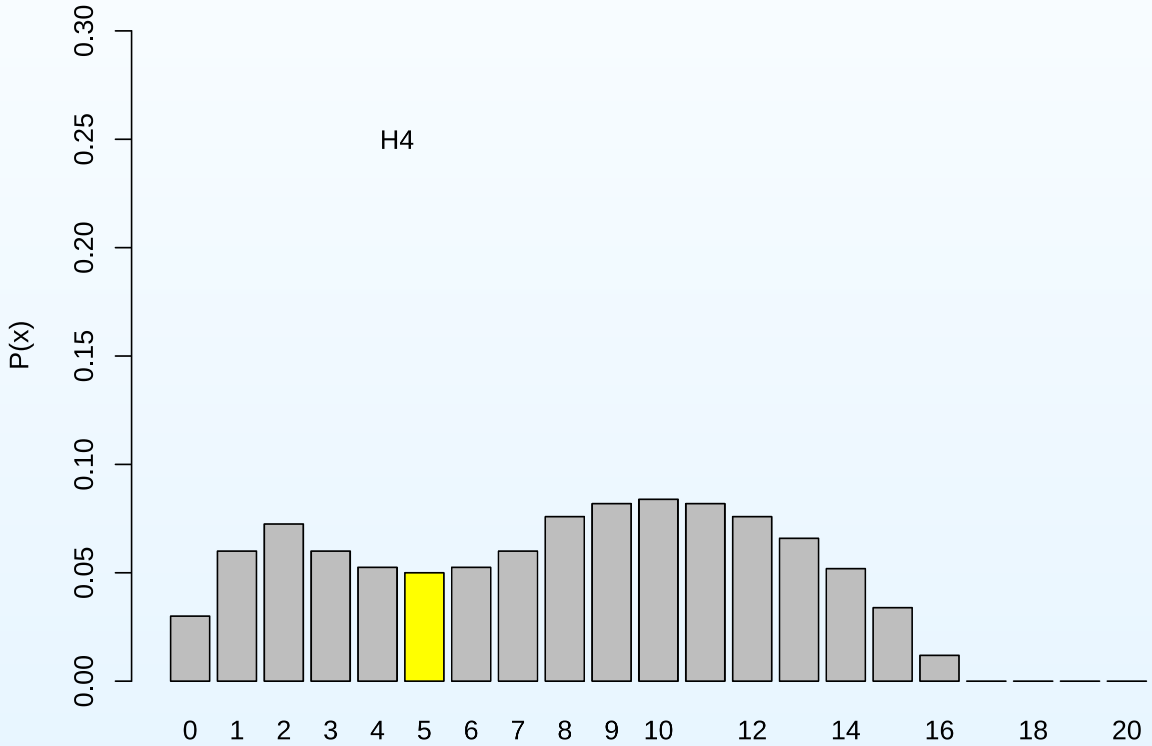


Natural that our conclusions depend ‘somehow’ on $P(x = 5 \,|\, H_i)$

(and in this case we learn nothing!)

Observed value and tails

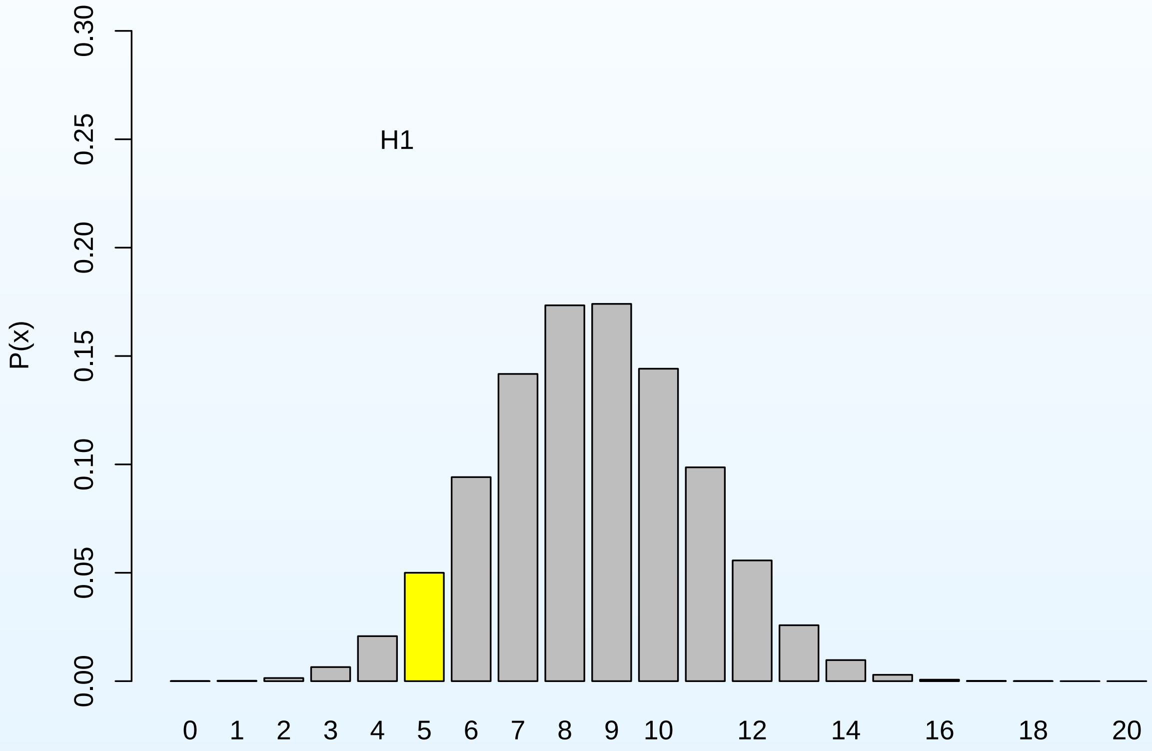
All have the same probability to give $x = 5$



Natural that our conclusions depend 'somehow' on $P(x = 5 \,|\, H_i)$
(and in this case we learn nothing!)

Observed value and tails

All have the same probability to give $x = 5$

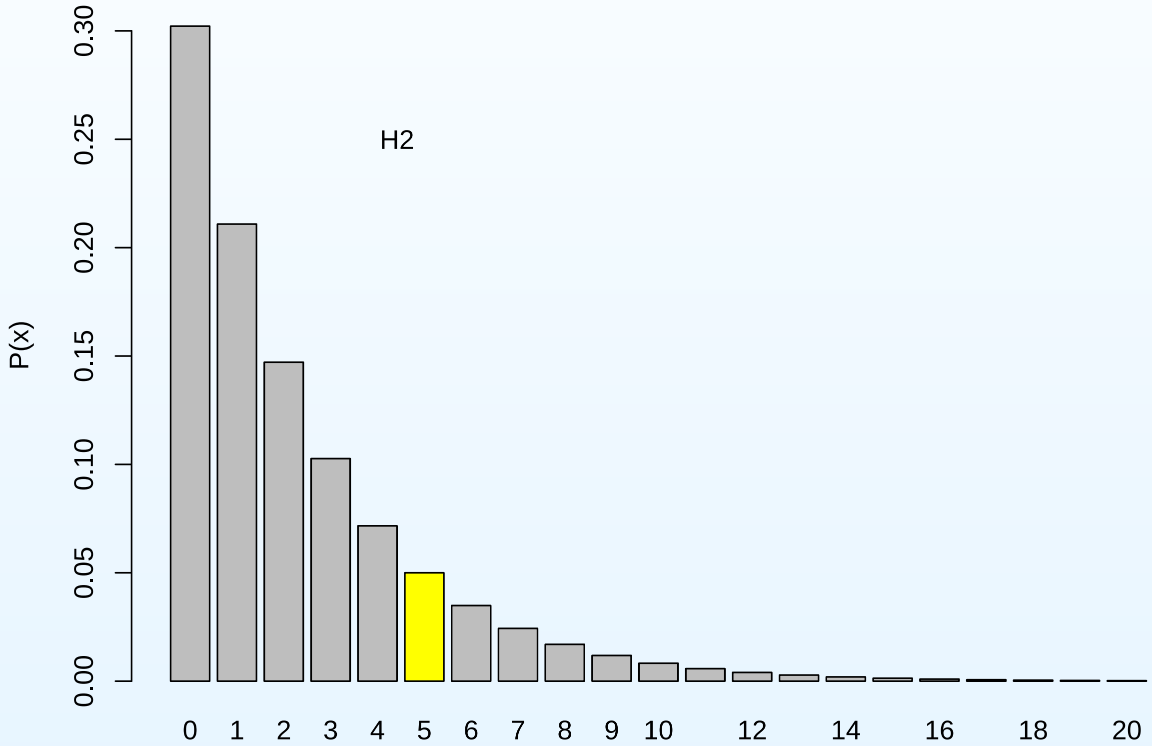


Natural that our conclusions depend ‘somehow’ on $P(x = 5 \,|\, H_i)$

but why also on $P(x \neq 5 \,|\, H_i)$?

Observed value and tails

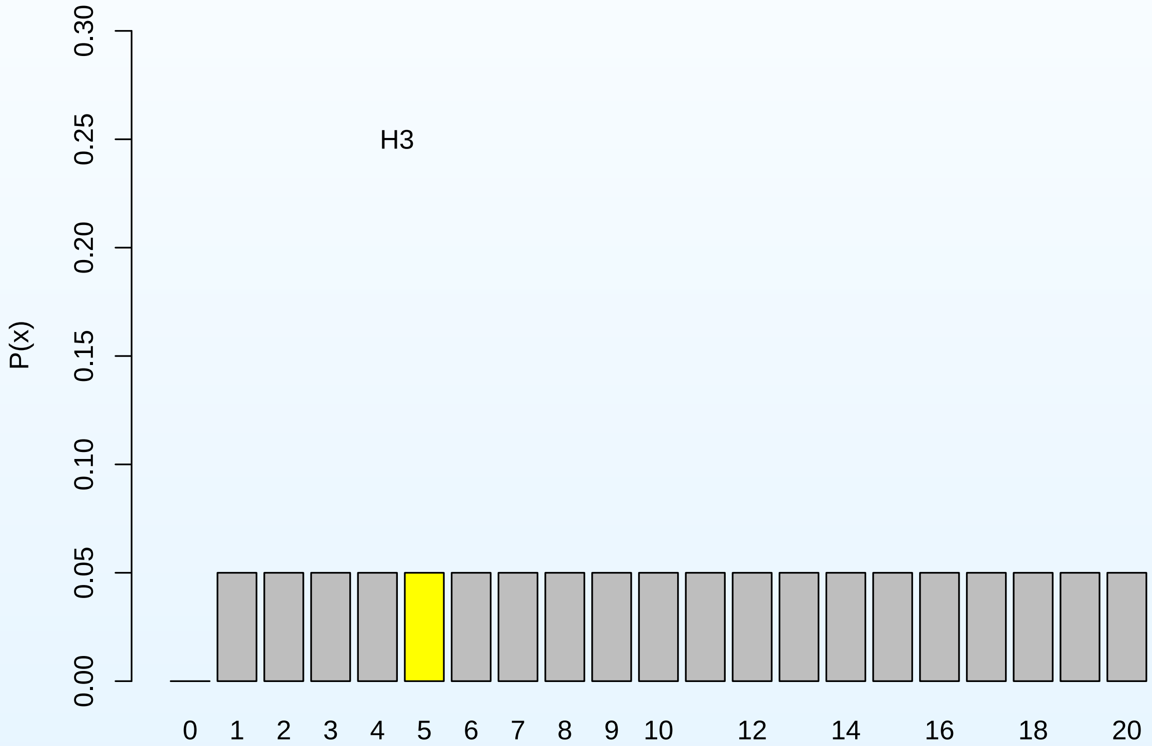
All have the same probability to give $x = 5$



Natural that our conclusions depend 'somehow' on $P(x = 5 \,|\, H_i)$
but why also on $P(x \neq 5 \,|\, H_i)$?

Observed value and tails

All have the same probability to give $x = 5$

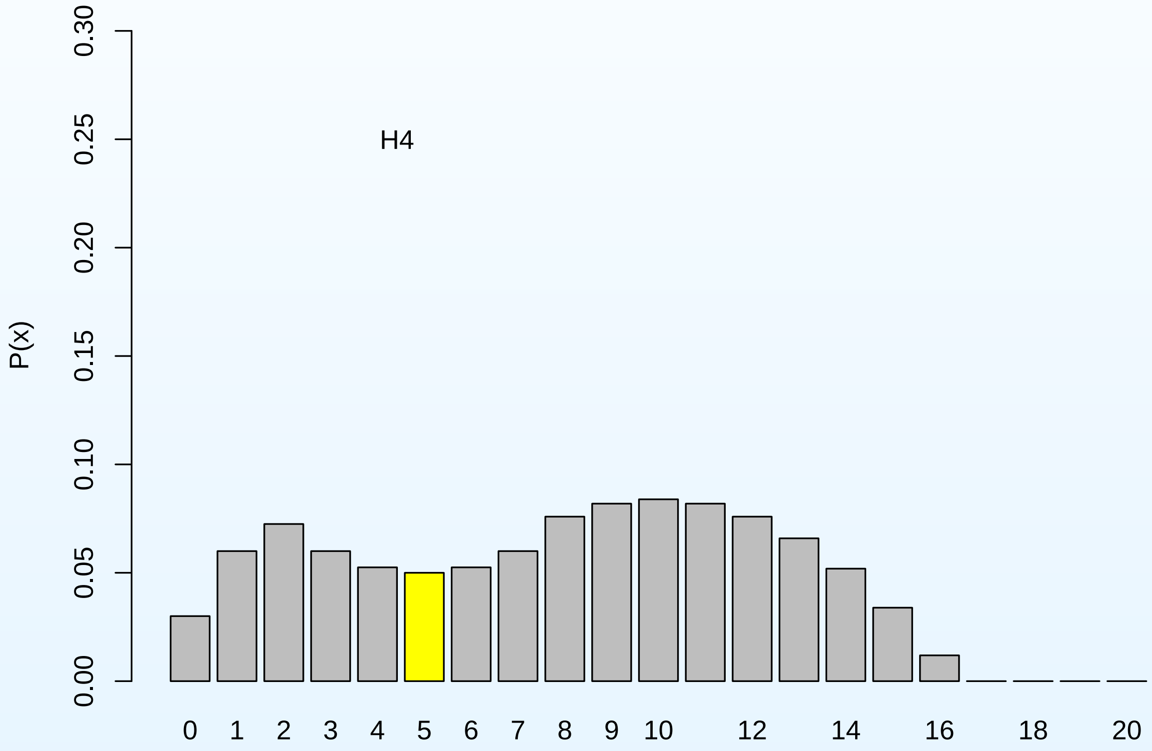


Natural that our conclusions depend 'somehow' on $P(x = 5 \,|\, H_i)$

but why also on $P(x \neq 5 \,|\, H_i)$?

Observed value and tails

All have the same probability to give $x = 5$



Natural that our conclusions depend 'somehow' on $P(x = 5 \,|\, H_i)$
but why also on $P(x \neq 5 \,|\, H_i)$?

P-values

But this is what we do when we draw scientific conclusions based on the probability of *‘what we have really observed, or something rarer than that’*,

P-values

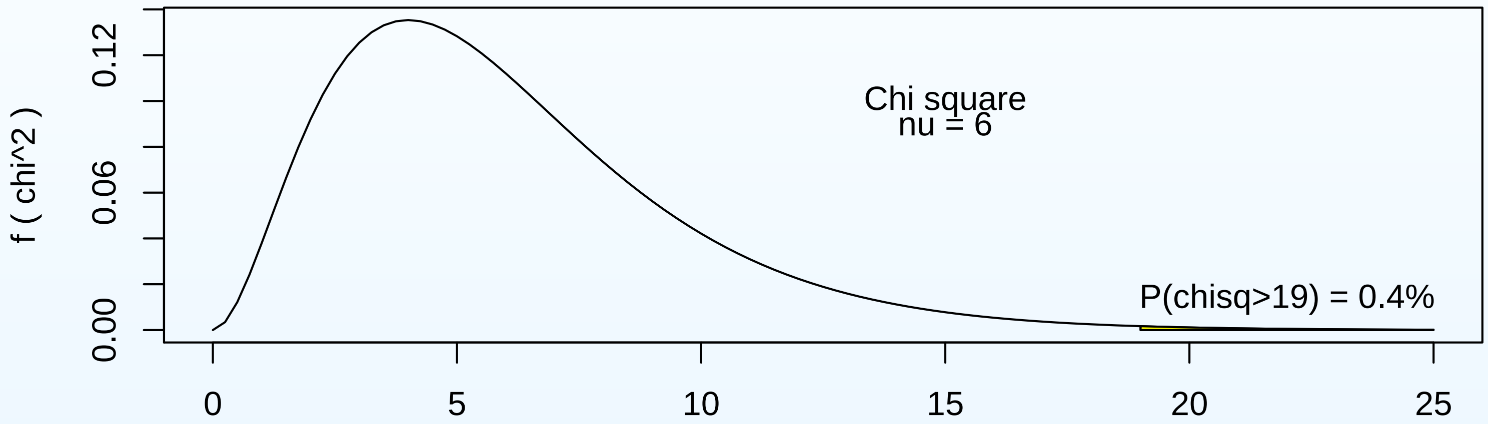
But this is what we do when we draw scientific conclusions based on the probability of *'what we have really observed, or something rarer than that'*,

what statisticians call p-values

(But physicists are more used with 'χ^2 probabilities', or something similar).

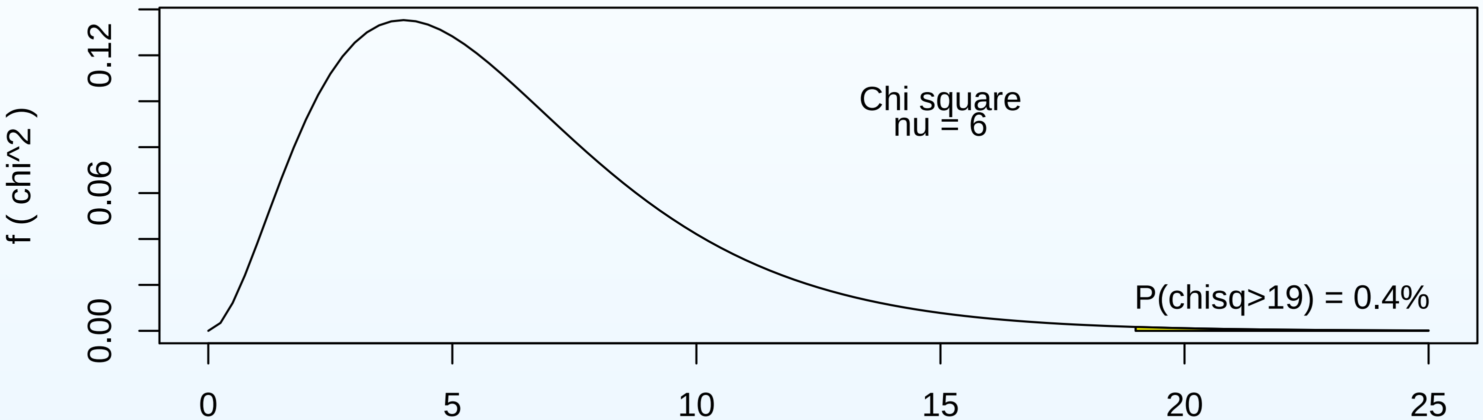
P-values

Ex.: χ^2, $\nu = 6$, $\chi^2_{obs} = 19$: $\rightarrow$ p-value $= \int_{\chi^2_{obs}}^{\infty} f(\chi^2_6)\,\mathrm{d}\chi^2_6$ = 0.4%.



P-values

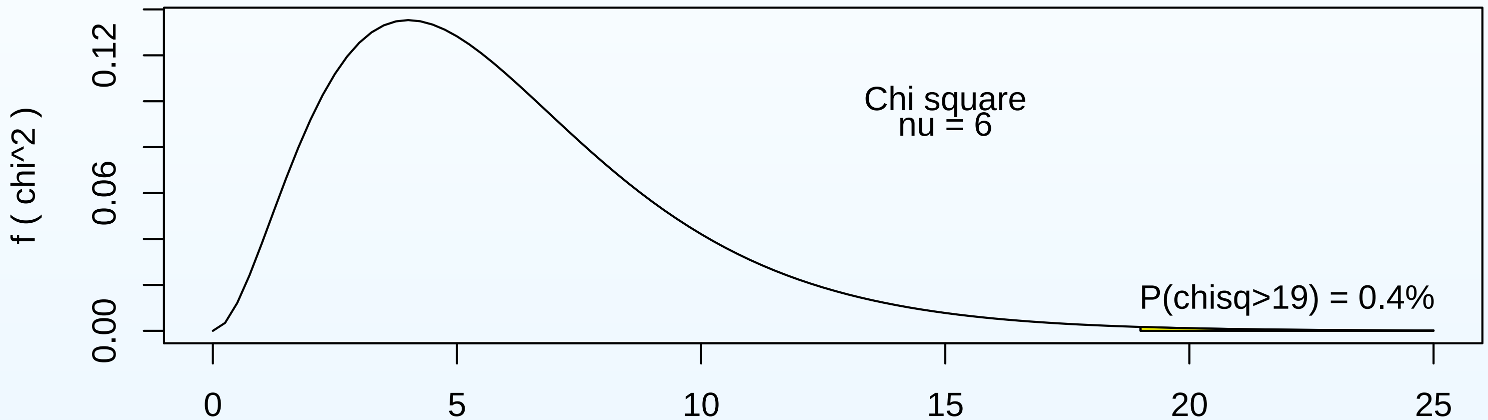
Ex.: χ^2, $\nu = 6$, $\chi^2_{obs} = 19$: $\rightarrow$ p-value $= \int_{\chi^2_{obs}}^{\infty} f(\chi^2_6)\, \mathrm{d}\chi^2_6$ = 0.4%.



What do you conclude?

P-values

Ex.: χ^2, $\nu = 6$, $\chi^2_{obs} = 19$: $\rightarrow$ p-value $= \int_{\chi^2_{obs}}^{\infty} f(\chi^2_6)\, d\chi^2_6$ = 0.4%.

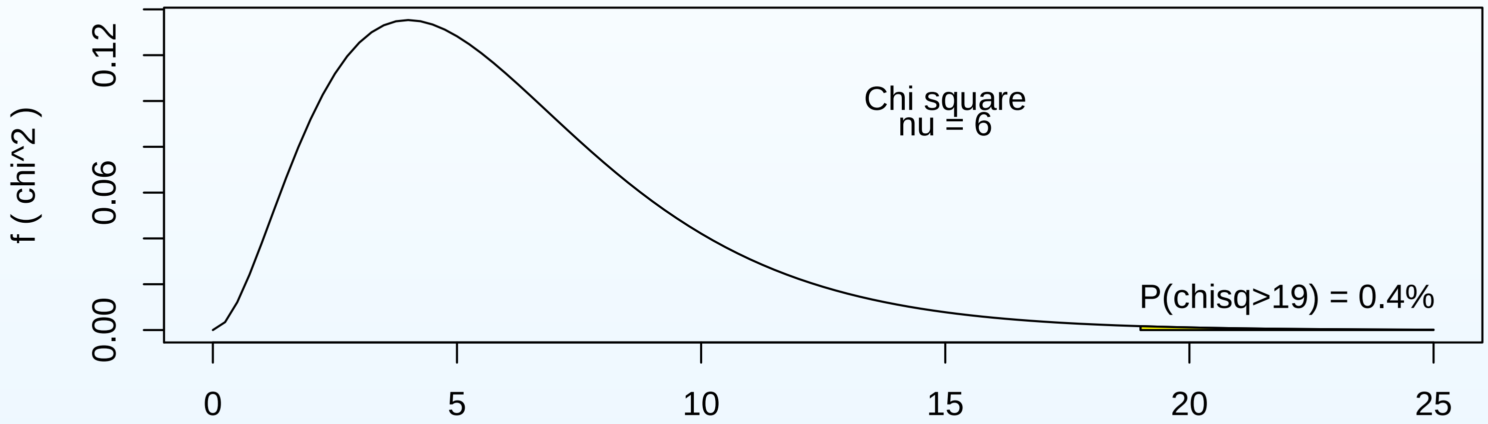


What do you conclude? (We shall come back later on this point)

Note for the moment:

P-values

Ex.: χ^2, $\nu = 6$, $\chi^2_{obs} = 19$: $\rightarrow$ p-value $= \int_{\chi^2_{obs}}^{\infty} f(\chi^2_6)\, \mathrm{d}\chi^2_6$ = 0.4%.



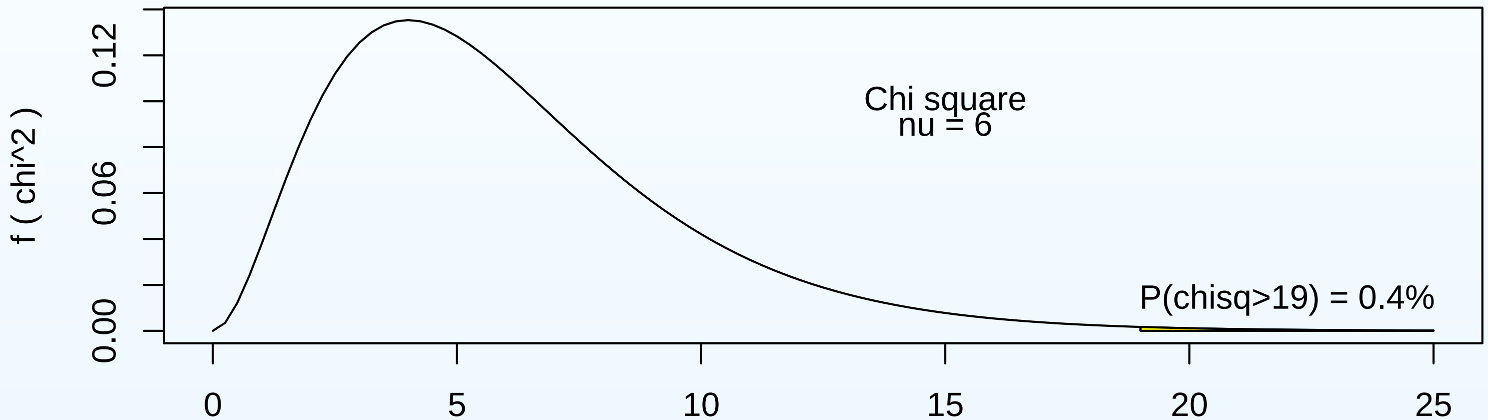
What do you conclude? (We shall come back later on this point)

Note for the moment: Whatever your conclusion is, based on this information, be aware:

- It does not depend directly on the observed data, but on the 'statistical summary' χ^2.

P-values

Ex.: χ^2, $\nu = 6$, $\chi^2_{obs} = 19$: $\rightarrow$ p-value $= \int_{\chi^2_{obs}}^{\infty} f(\chi^2_6)\, \mathrm{d}\chi^2_6$ = 0.4%.



What do you conclude? (We shall come back later on this point)
Note for the moment: Whatever your conclusion is, based on this information, be aware:

- It does not depend directly on the observed data, but on the 'statistical summary' χ^2.
- Indeed, it does not even depend precisely on the 'observed summary' alone (χ^2_{obs}), but on all other values of the summary that are less likely than the observed one.

P-values

Rationale?

Rationale?

As most of these kind of prescriptions, they are not based on solid principles but only on authority and use.

P-values

Rationale?

As most of these kind of prescriptions, they are not based on solid principles but only on authority and use.

But then it must work, otherwise it should have been realized!

P-values

Rationale?

As most of these kind of prescriptions, they are not based on solid principles but only on authority and use.

But then it must work, otherwise it should have been realized!

- Yes!

P-values

Rationale?

As most of these kind of prescriptions, they are not based on solid principles but only on authority and use.

But then it must work, otherwise it should have been realized!

- Yes! 'It does often work',
  but this has little to do with the 'probability of the tail', as we shall see later.

Example: Has the student made a mistake?

Homework: calculate the average of 300 random numbers, uniformly distributed between 0 and 1.

Example: Has the student made a mistake?

Homework: calculate the average of 300 random numbers, uniformly distributed between 0 and 1.

- Teacher expectation:

$$\mathsf{E}\left[\overline{X}_{300}\right] = \frac{1}{2}$$

$$\sigma\left[\overline{X}_{300}\right] = \frac{1}{\sqrt{12}} \cdot \frac{1}{\sqrt{300}} = 0.017\,,$$

Example: Has the student made a mistake?

Homework: calculate the average of 300 random numbers, uniformly distributed between 0 and 1.

- Teacher expectation:

$$
\begin{aligned}
\mathrm{E}\left[\overline{X}_{300}\right] &= \frac{1}{2} \\
\sigma\left[\overline{X}_{300}\right] &= \frac{1}{\sqrt{12}} \cdot \frac{1}{\sqrt{300}} = 0.017\,,
\end{aligned}
$$

- 99% probability interval

$$P(0.456 \leq \overline{X}_{300} \leq 0.544) = 99\%\,.$$

Example: Has the student made a mistake?

Homework: calculate the average of 300 random numbers, uniformly distributed between 0 and 1.

- Teacher expectation:

$$\mathrm{E}\left[\overline{X}_{300}\right] = \frac{1}{2}$$

$$\sigma\left[\overline{X}_{300}\right] = \frac{1}{\sqrt{12}} \cdot \frac{1}{\sqrt{300}} = 0.017\,,$$

- 99% probability interval

$$P(0.456 \le \overline{X}_{300} \le 0.544) = 99\%\,.$$

- Student gets a value outside the interval, e.g. $\overline{x} = 0.550$.

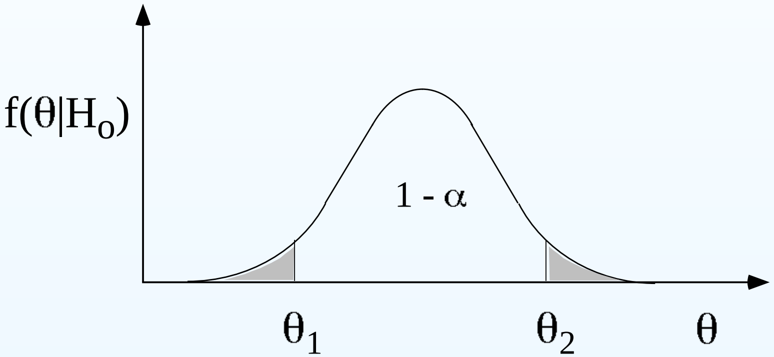
$\Rightarrow$ Has the student made a mistake?

Example: Has the student made a mistake?

Conventional statistician solution:
$\Rightarrow$ test the hypothesis $H_0 =$ 'no mistakes'

Example: Has the student made a mistake?

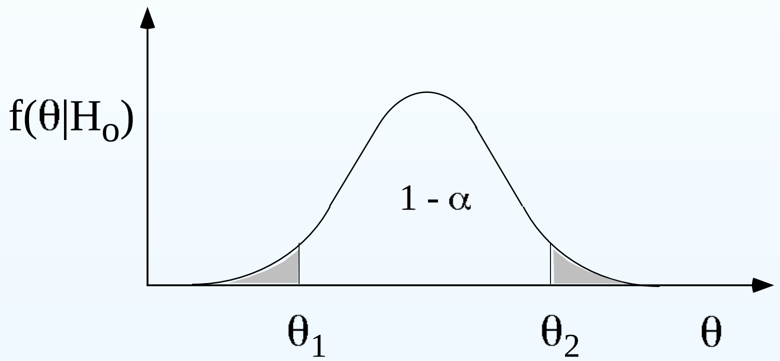
Conventional statistician solution:
$\Rightarrow$ test the hypothesis $H_0 =$ 'no mistakes'



- Test variable θ is $\overline{X}_{300}$.
- Acceptance interval $[\theta_1, \theta_2]$ is $[0.456, 0.544]$.
  We are 99% confident that $\overline{X}_{300}$ will fall inside it:
  $\rightarrow \alpha = 1\%$.

Example: Has the student made a mistake?

Conventional statistician solution:
$\Rightarrow$ test the hypothesis $H_0 =$ 'no mistakes'

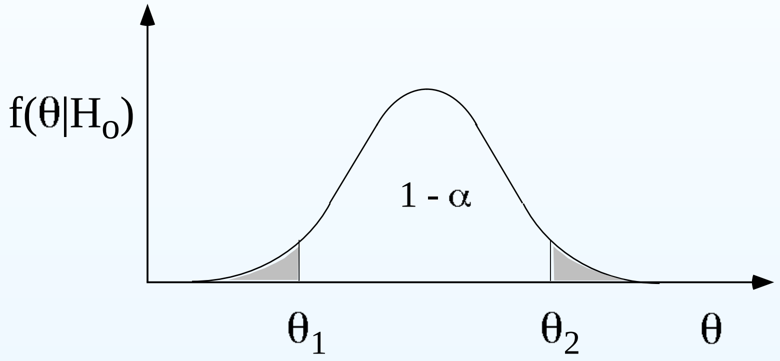


- Test variable θ is $\overline{X}_{300}$.
- Acceptance interval $[\theta_1, \theta_2]$ is $[0.456, 0.544]$.
  We are 99% confident that $\overline{X}_{300}$ will fall inside it:
  $\rightarrow \alpha = 1\%$.
- $\overline{x} = 0.550$ lies outside the acceptance interval

$\Rightarrow$ Hypothesis H_0 is rejected at 1% significance.

Example: Has the student made a mistake?

Conventional statistician solution:
$\Rightarrow$ test the hypothesis $H_0 =$ 'no mistakes'



- Test variable θ is $\overline{X}_{300}$.
- Acceptance interval $[\theta_1, \theta_2]$ is $[0.456, 0.544]$.
  We are 99% confident that $\overline{X}_{300}$ will fall inside it:
  $\longrightarrow \alpha = 1\%$.
- $\overline{x} = 0.550$ lies outside the acceptance interval

$\Rightarrow$ Hypothesis H_0 is rejected at 1% significance.

$\Rightarrow$ What does it mean?

Meaning of the hypothesis test

Conclusion from test:

"the hypothesis $H_\circ$ *= 'no mistakes' is rejected at the 1% level of significance".*

Meaning of the hypothesis test

Conclusion from test:

> *"the hypothesis $H_\circ$ = 'no mistakes' is rejected at the 1% level of significance".*

What does it mean?

> *"there is only a 1% probability that the average falls outside the selected interval, if the calculations were done correctly".*

Meaning of the hypothesis test

Conclusion from test:

> *"the hypothesis $H_\circ$ = 'no mistakes' is rejected at the 1% level of significance".*

What does it mean?

> *"there is only a 1% probability that the average falls outside the selected interval, if the calculations were done correctly".*

So what?

Meaning of the hypothesis test

Conclusion from test:

> *"the hypothesis $H_\circ$ = 'no mistakes' is rejected at the 1% level of significance".*

What does it mean?

> *"there is only a 1% probability that the average falls outside the selected interval, if the calculations were done correctly".*

So what?

- It does not reply our natural question, i.e. that concerning the probability of mistake – quite impolite, by the way.
- The statement sounds as if one would be 99% sure that the student has made a mistake! (Mostly interpreted in this way).

$\Rightarrow$ Highly misleading!

Something is missing in the reasoning

If you ask the students (before they take a standard course in hypothesis tests) you will realize of a crucial ingredient extraneous to the logic of hypothesis tests:

Something is missing in the reasoning

If you ask the students (before they take a standard course in hypothesis tests) you will realize of a crucial ingredient extraneous to the logic of hypothesis tests:

> *"It all depends on whom has made the calculation!"*

Something is missing in the reasoning

If you ask the students (before they take a standard course in hypothesis tests) you will realize of a crucial ingredient extraneous to the logic of hypothesis tests:

“It all depends on whom has made the calculation!”

In fact, if the calculation was done by a well-tested program, the probability of mistake would be zero.
And students know rather well their tendency to do or not mistakes.

‘Something is missing’: another example

The value $x = 3.01$ is extracted from a Gaussian random number generator having $\mu = 0$ and $\sigma = 1$.

‘Something is missing’: another example

The value $x = 3.01$ is extracted from a Gaussian random number generator having $\mu = 0$ and $\sigma = 1$.
It is well known that $P(|X| > 3) = 0.27\%$, but

‘Something is missing’: another example

The value $x = 3.01$ is extracted from a Gaussian random number generator having $\mu = 0$ and $\sigma = 1$.
It is well known that $P(|X| > 3) = 0.27\%$, but

we cannot say

- “the value X has 0.27% probability of coming from that generator”

‘Something is missing’: another example

The value $x = 3.01$ is extracted from a Gaussian random number generator having $\mu = 0$ and $\sigma = 1$.
It is well known that $P(|X| > 3) = 0.27\%$, but

we cannot say

- “the value X has 0.27% probability of coming from that generator”
- “the probability that the observation is a statistical fluctuation is 0.27%”

‘Something is missing’: another example

The value $x = 3.01$ is extracted from a Gaussian random number generator having $\mu = 0$ and $\sigma = 1$.
It is well known that $P(|X| > 3) = 0.27\%$, but

we cannot say

- “the value X has 0.27% probability of coming from that generator”
- “the probability that the observation is a statistical fluctuation is 0.27%”

$\Rightarrow$ the value comes with 100% probability from that generator!

‘Something is missing’: another example

The value $x = 3.01$ is extracted from a Gaussian random number generator having $\mu = 0$ and $\sigma = 1$.
It is well known that $P(|X| > 3) = 0.27\%$, but

we cannot say

- “the value X has 0.27% probability of coming from that generator”
- “the probability that the observation is a statistical fluctuation is 0.27%”

$\Rightarrow$ the value comes with 100% probability from that generator!

$\Rightarrow$ it is at 100% a statistical fluctuation

‘Something is missing’: another example

The value $x = 3.01$ is extracted from a Gaussian random number generator having $\mu = 0$ and $\sigma = 1$.
It is well known that $P(|X| > 3) = 0.27\%$, but

we cannot say

- “the value X has 0.27% probability of coming from that generator”
- “the probability that the observation is a statistical fluctuation is 0.27%”

$\Rightarrow$ the value comes with 100% probability from that generator!

$\Rightarrow$ it is at 100% a statistical fluctuation

Logical bug of the reasoning:

$\Rightarrow$ One cannot tell how much one is confident in generator A only if another generator B is not taken into account.

‘Something is missing’: another example

The value $x = 3.01$ is extracted from a Gaussian random number generator having $\mu = 0$ and $\sigma = 1$.
It is well known that $P(|X| > 3) = 0.27\%$, but

we cannot say

- “the value X has 0.27% probability of coming from that generator”
- “the probability that the observation is a statistical fluctuation is 0.27%”

$\Rightarrow$ the value comes with 100% probability from that generator!

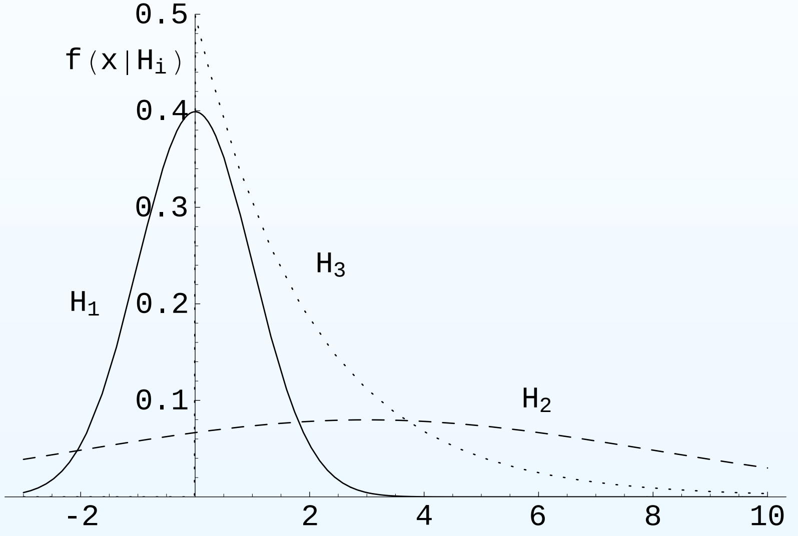
$\Rightarrow$ it is at 100% a statistical fluctuation

Logical bug of the reasoning:

$\Rightarrow$ This is the original sin of conventional hypothesis test methods

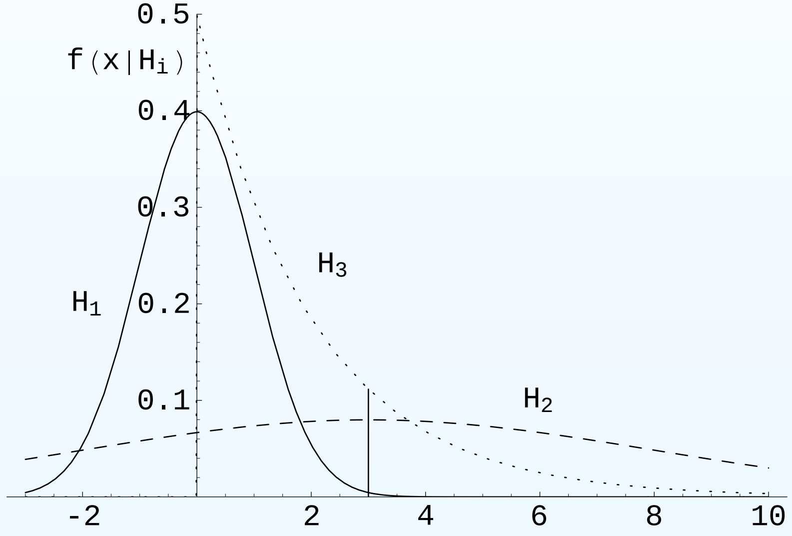
Well posed problem

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



Well posed problem

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

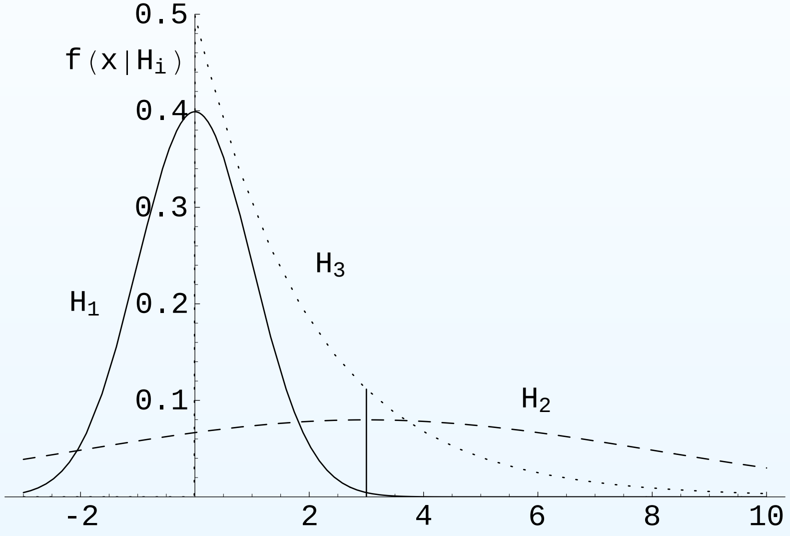


The statistics-uneducated student would suggest:

- our preference should depend on how likely each model might yield $x = 3$

Well posed problem

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

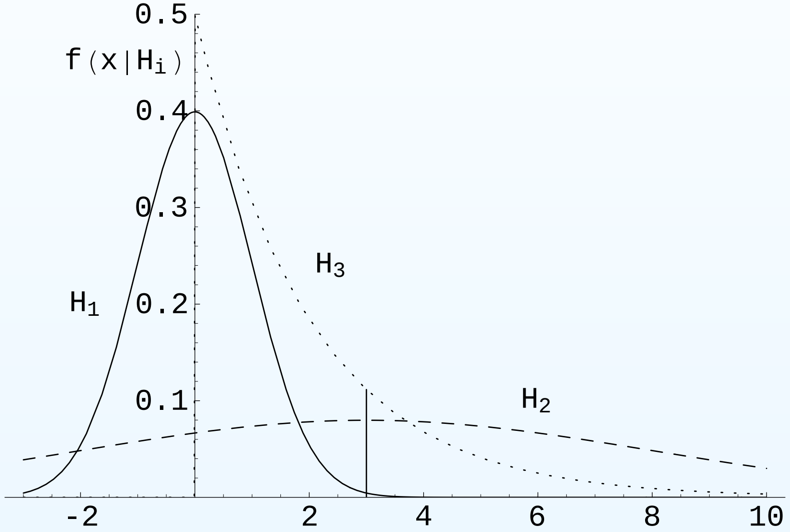


The statistics-uneducated student would suggest:

- our preference should depend on how likely each model might yield $x = 3$
- . . . but perhaps also on 'how reasonable' each model is, given the physical situation under study

Well posed problem

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



The statistics-uneducated student would suggest:

- our preference should depend on how likely each model might yield $x = 3$
- . . . but perhaps also on 'how reasonable' each model is, given the physical situation under study

$\Rightarrow$ Right!

Objections

“These are chosen academic examples.”

Objections

"These are chosen academic examples."

$\Rightarrow$ logic is logic!

Objections

“These are chosen academic examples.”

$\Rightarrow$ logic is logic!

How can we use a reasoning in frontier physics
if it fails in simple cases?

$\Rightarrow$ All fake claims of discoveries are due to
the criticized reasoning (examples in a while $\longrightarrow$)

Objections

“These are chosen academic examples.”

$\Rightarrow$ logic is logic!

How can we use a reasoning in frontier physics
if it fails in simple cases?

$\Rightarrow$ All fake claims of discoveries are due to
the criticized reasoning (examples in a while $\longrightarrow$)

“Hypotheses tests are well proved to work”

Objections

“These are chosen academic examples.”

$\Rightarrow$ logic is logic!

How can we use a reasoning in frontier physics if it fails in simple cases?

$\Rightarrow$ All fake claims of discoveries are due to the criticized reasoning (examples in a while $\longrightarrow$)

“Hypotheses tests are well proved to work”

Yes and not. . .

$\Rightarrow$ They ‘often work’ due to reasons external to their logic, but which are not always satisfied, especially in the frontier cases that mostly concern us.

$\longrightarrow$ we shall come back to this point

Examples from particle physics

$\Rightarrow$ *See transparencies*

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 $\Leftrightarrow$ 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)

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)

⇒ 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)

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

. . . indeed not a very solid superstructure

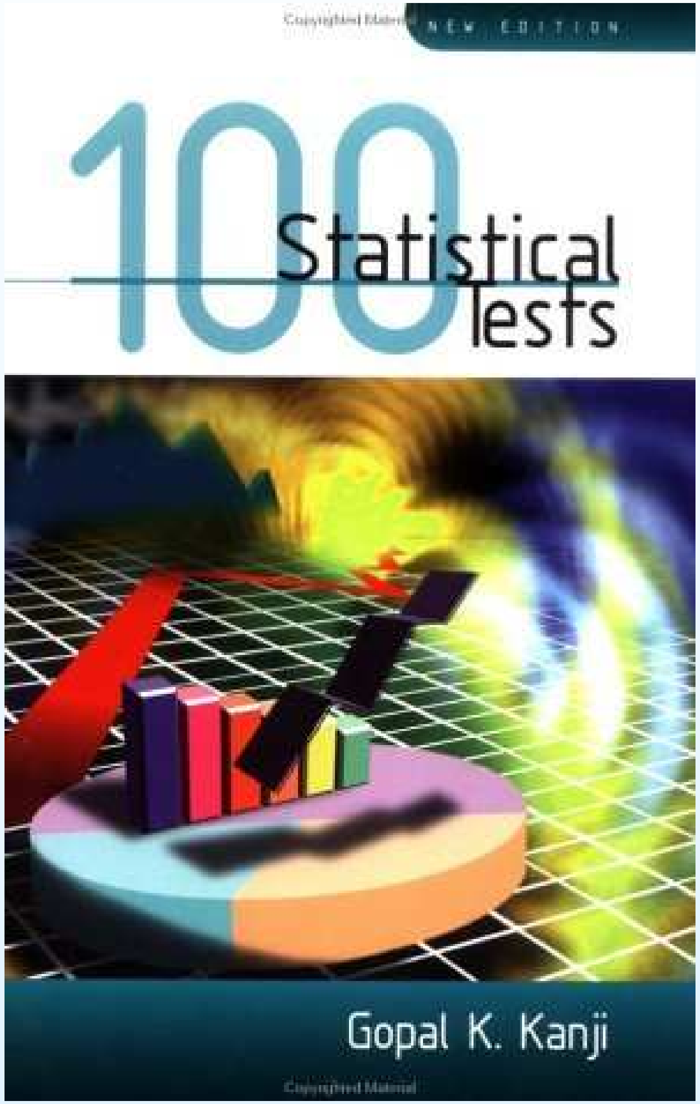
Moreover, a part the ‘philosophical’ problem of interpretation, there are plenty of ‘practical’ problems, since ‘statistical test’ are based on authority principle and not grounded on solid bases (probabilistic ‘first principles’).

. . . indeed not a very solid superstructure

Moreover, a part the ‘philosophical’ problem of interpretation, there are plenty of ‘practical’ problems, since ‘statistical test’ are based on authority principle and not grounded on solid bases (probabilistic ‘first principles’).

- Rich choice

. . . indeed not a very solid superstructure



Not exhaustive compilation. . .

. . . indeed not a very solid superstructure

Moreover, a part the ‘philosophical’ problem of interpretation, there are plenty of ‘practical’ problems, since ‘statistical test’ are based on authority principle and not grounded on solid bases (probabilistic ‘first principles’).

- Rich choice $\longrightarrow$ > ‘100 tests’

. . . indeed not a very solid superstructure

Moreover, a part the 'philosophical' problem of interpretation, there are plenty of 'practical' problems, since 'statistical test' are based on authority principle and not grounded on solid bases (probabilistic 'first principles').

- Rich choice $\longrightarrow$ > '100 tests'
- Discussions about which test to use it and how to use it are not deeper that discussions in pubs among soccer fans (Italians might think to the 'Processo di Biscardi' †)

... indeed not a very solid superstructure

Moreover, a part the 'philosophical' problem of interpretation, there are plenty of 'practical' problems, since 'statistical test' are based on authority principle and not grounded on solid bases (probabilistic 'first principles').

- Rich choice $\longrightarrow$ > '100 tests'
- Discussions about which test to use it and how to use it are not deeper that discussions in pubs among soccer fans (Italians might think to the 'Processo di Biscardi' †)

$\Rightarrow$ Tendency to look for the test that gives the result one wants

. . . indeed not a very solid superstructure

Moreover, a part the ‘philosophical’ problem of interpretation, there are plenty of ‘practical’ problems, since ‘statistical test’ are based on authority principle and not grounded on solid bases (probabilistic ‘first principles’).

- Rich choice $\longrightarrow$ > ‘100 tests’
- Discussions about which test to use it and how to use it are not deeper that discussions in pubs among soccer fans (Italians might think to the ‘Processo di Biscardi’ †)

$\Rightarrow$ Tendency to look for the test that gives the result one wants

- My personal prejudice: *The fancier the name of the test is, the less believable the claim is*, because I am pretty sure that other, more common tests were discarded, because ‘they did not work’.

. . . indeed not a very solid superstructure

Moreover, a part the ‘philosophical’ problem of interpretation, there are plenty of ‘practical’ problems, since ‘statistical test’ are based on authority principle and not grounded on solid bases (probabilistic ‘first principles’).

- Rich choice $\longrightarrow$ > ‘100 tests’
- Discussions about which test to use it and how to use it are not deeper that discussions in pubs among soccer fans (Italians might think to the ‘Processo di Biscardi’ †)

$\Rightarrow$ Tendency to look for the test that gives the result one wants

- My personal prejudice: *The fancier the name of the test is, the less believable the claim is*, because I am pretty sure that other, more common tests were discarded, because ~~‘they did not work’~~ $\longrightarrow$ ‘they did not support what the guy wanted the data to prove’
  $\chi^2 \longrightarrow$ run-test $\longrightarrow$ Kolmogorov $\longrightarrow$ . . . ?. . . $\Rightarrow$ ~~Lourdes~~ Fatima.

.

Is statistics something serious?

Last, but not least, standard statistical methods, essentially a contradictory collection of *ad-hoc-eries*, induce scientists, and physicists in particular, to think that

'statistics' is something 'not scientific'.

$\Rightarrow$ 'creative' behavior is encouraged

What to do?

$\Rightarrow$ Back to the past

What to do?

$\Longrightarrow$ Back to the 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?

⇒ Back to the 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

What to do?

⇒ Back to the 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 idea of 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 ⇔ Past (believed so)

$n \longrightarrow \infty$: $\longrightarrow$ *"usque tandem?"*

$\longrightarrow$ *"in the long run we are all dead"*

$\longrightarrow$ It limits the range of applications

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.

Definitions → evaluation rules

Very useful evaluation rules

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

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

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

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 is probability?

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

→ how much we are confident that something is true

What is 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

What is 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. . .

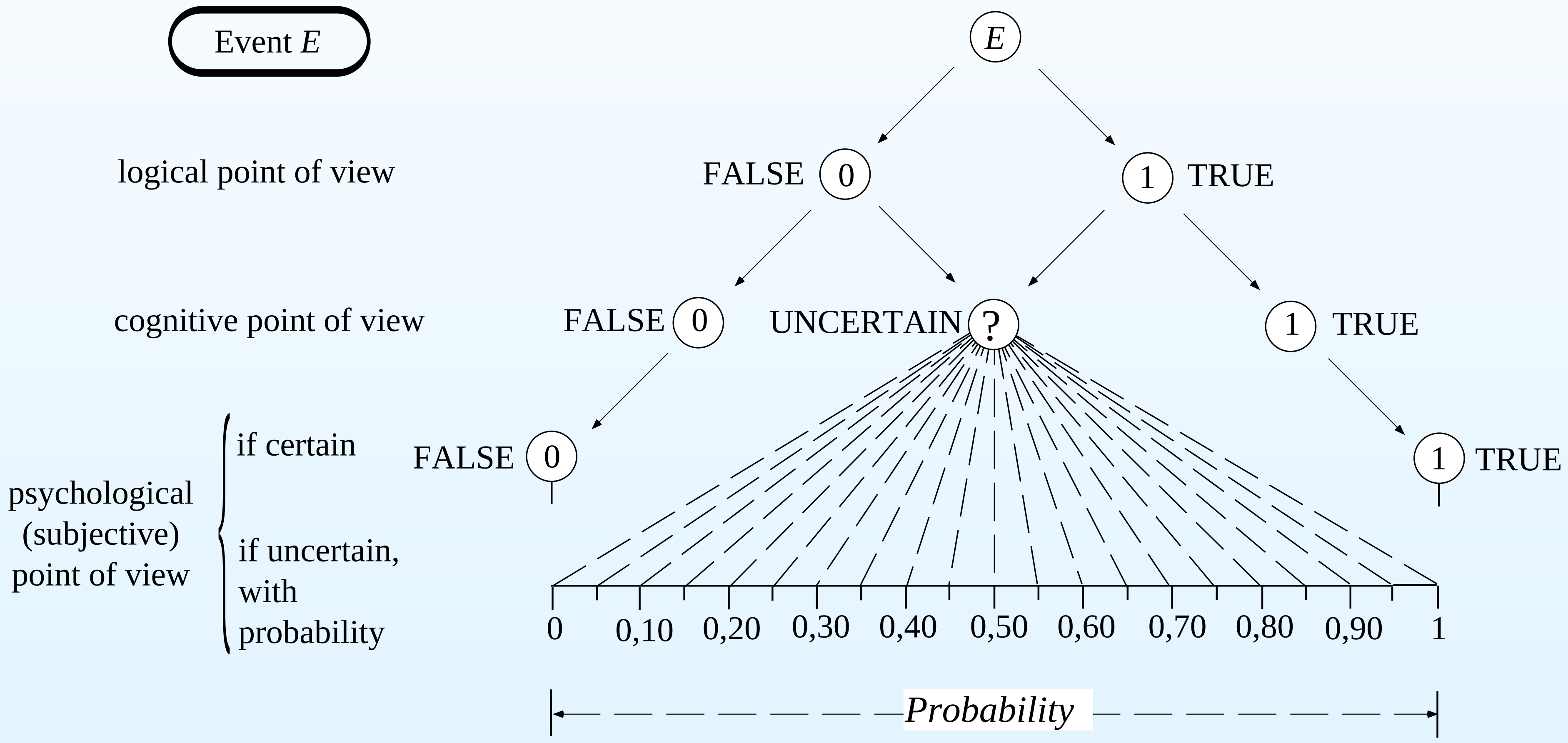
"Given the state of our knowledge about everything that could possible have any bearing on the coming true[1]*. . . , the numerical probability* p *of this event is to be a real number by the indication of which we try* in some cases *to setup a quantitative measure of the strength of our conjecture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



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 $\longrightarrow$ 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.

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 $\longrightarrow$ probability

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

Probability is always conditional probability

$$\text{‘}P(E)\text{’} \;\longrightarrow\; P(E\,|\,I) \;\longrightarrow\; P(E\,|\,I(t))$$

Uncertainty $\longrightarrow$ probability

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

Probability is always conditional probability

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

- *"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)

Uncertainty $\longrightarrow$ probability

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

Probability is always conditional probability

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

- *"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)
- Some examples:
    - tossing a die;
    - 'three box problems';
    - two envelops' paradox.

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.
  - $P(\text{next Saturday}) = 68\%$
  - $P(\text{Inter will win Italian champion league}) = 68\%$
  - $P(\text{free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$

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.
  - $P(\text{next Saturday}) = 68\%$
  - $P(\text{Inter will win Italian champion league}) = 68\%$
  - $P(\text{free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$

  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.
  - $P(\text{next Saturday}) = 68\%$
  - $P(\text{Inter will win Italian champion league}) = 68\%$
  - $P(\text{free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$

  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>)

  $\Rightarrow$ SLIDES (Higgs mass limits)

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.
  - $P(\text{next Saturday}) = 68\%$
  - $P(\text{Inter will win Italian champion league}) = 68\%$
  - $P(\text{free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$

  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"!)
- 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.
  - $P(\text{next Saturday}) = 68\%$
  - $P(\text{Inter will win Italian champion league}) = 68\%$
  - $P(\text{free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$

- 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.
  - $P(\text{next Saturday}) = 68\%$
  - $P(\text{Inter will win Italian champion league}) = 68\%$
  - $P(\text{free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$
- 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.

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 $\longrightarrow$ 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.

$\longrightarrow$ see later for details, examples, objections, 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?

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’

$\longrightarrow$ 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'

→ analogy to measures (we need to measure 'befiefs')

⇒ 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'

$\longrightarrow$ 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

They all lead to

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

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. $P(A \cup B \,|\, I) = P(A \,|\, I) + P(B \,|\, I)$ [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)

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

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€.
  ⇒ If OK with you, let's start.
- You claim that <u>this</u> coin has 30% to give head?
  ⇒ Just reverse the bet

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

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

⇒ 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

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

- All the rest by logic

$\longrightarrow$ And, please, be coherent!

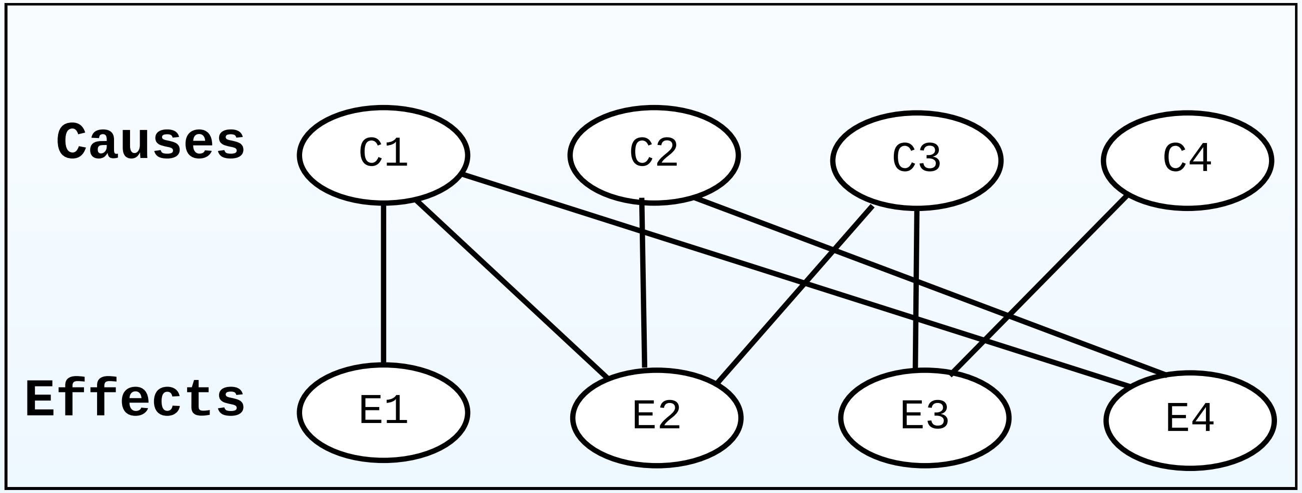
Inference

Inference

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

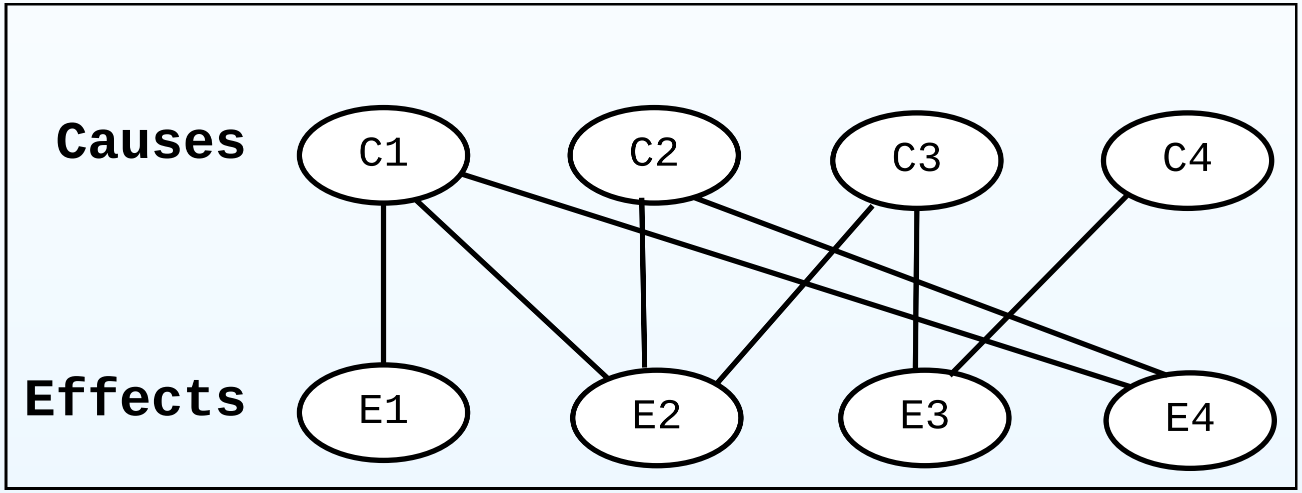
From causes to effects and back

Our original problem:



From causes to effects and back

Our original problem:

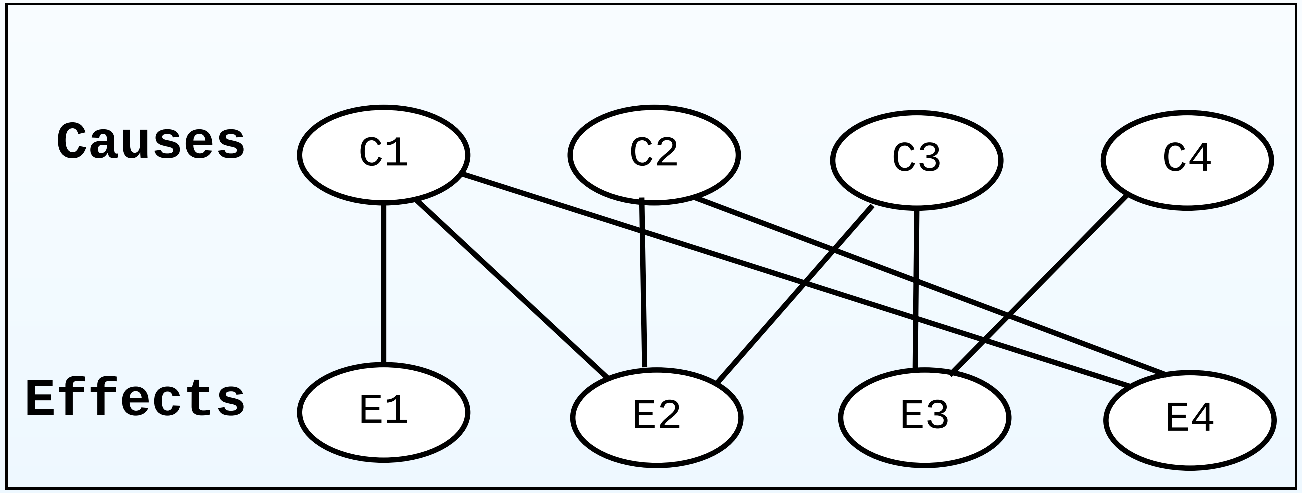


Our conditional view of probabilistic causation

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

From causes to effects and back

Our original problem:



Our conditional view of probabilistic causation

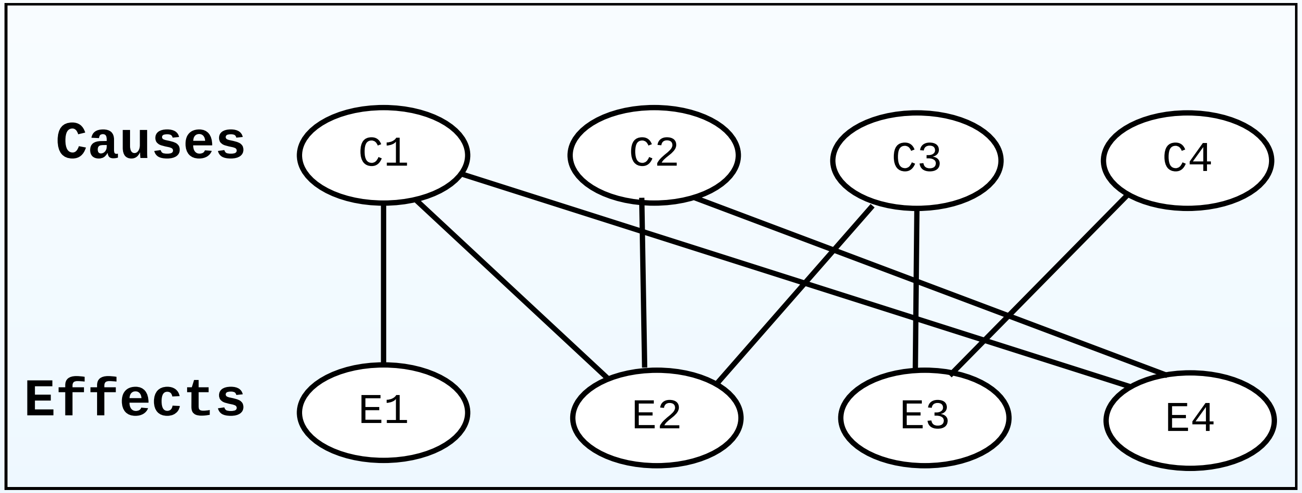
$$\boxed{P(E_i \,|\, C_j)}$$

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

$$P(E_i \,|\, C_j)$$

Our conditional view of probabilistic inference

$$P(C_j \,|\, E_i)$$

The fourth basic rule of probability:

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

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:

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

"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

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

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."

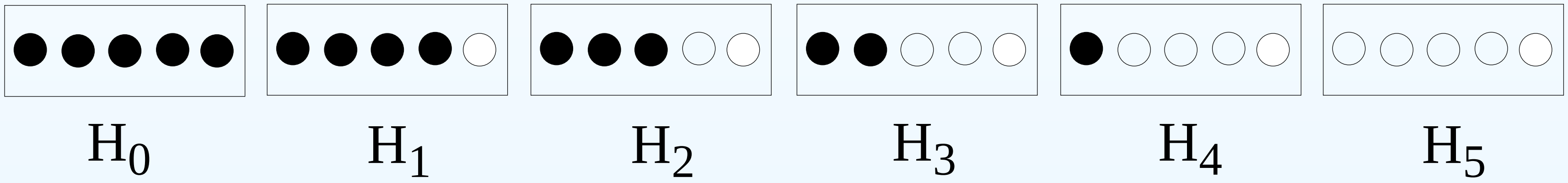
It follows

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

"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:

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

Collecting the pieces of information we need

Our tool:

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

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

Collecting the pieces of information we need

Our tool:

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

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

Collecting the pieces of information we need

Our tool:

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

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

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

Collecting the pieces of information we need

Our tool:

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

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

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

Our prior belief about H_j

Collecting the pieces of information we need

Our tool:

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

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

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

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

Collecting the pieces of information we need

Our tool:

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

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

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

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:

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

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

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

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:

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

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

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

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:

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

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

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

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:

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

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

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

We are ready!

$\longrightarrow$ R program

First extraction

After first extraction (and reintroduction) of the ball:

- $P(H_j)$ changes
- $P(E_j)$ for next extraction changes

Note: The box is exactly in the same status as before

First extraction

After first extraction (and reintroduction) of the ball:

- $P(H_j)$ changes
- $P(E_j)$ for next extraction changes

Note: The box is exactly in the same status as before

Where is probability?

→ Certainly not in the box!

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:

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

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

Updating the knowledge by new observations

Let us repeat the experiment:

Sequential use of Bayes theorem

Old posterior becomes new prior, and so on

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

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:

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:

Sequential use of Bayes theorem

Old posterior becomes new prior, and so on

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

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

Solution of the AIDS test problem

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

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

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

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

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

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!

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 hidden uniform

What was the mistake of people saying $P(\overline{\text{HIV}}\,|\,\text{Pos}) = 0.2$?

We can easily check that this is due to have set $\frac{P_\circ(\text{HIV})}{P(\overline{\text{HIV}})} = 1$,

that, hopefully, does not apply for a randomly selected Italian.

- This is typical in arbitrary inversions, and often also in frequentistic prescriptions that are used by the practitioners to form their confidence on something:

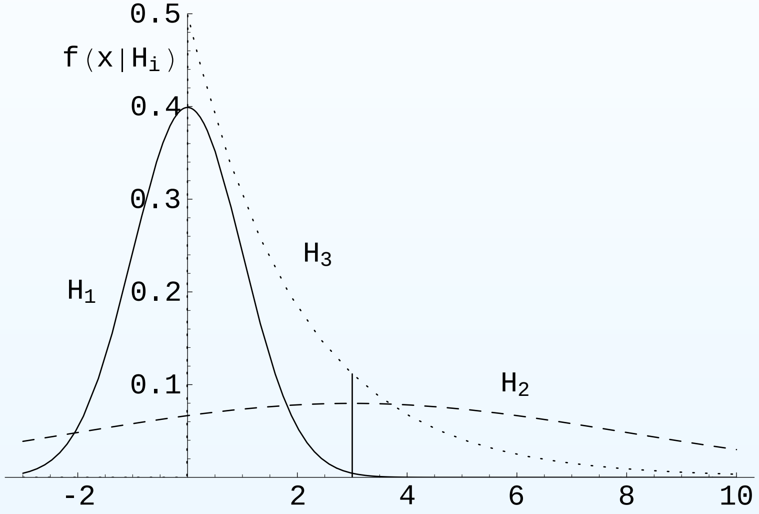
$\longrightarrow$ "absence of priors" means in most times uniform priors over the all possible hypotheses

- but they criticize the Bayesian approach because it takes into account priors explicitly !

Better methods based on 'sand' than methods based on nothing!

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: "$P(\text{data} \mid H_j)$" $\longleftrightarrow$ "$f(\text{data} \mid H_j)$".

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

$BF_{2,1} = 18$, $BF_{3,1} = 25$ 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

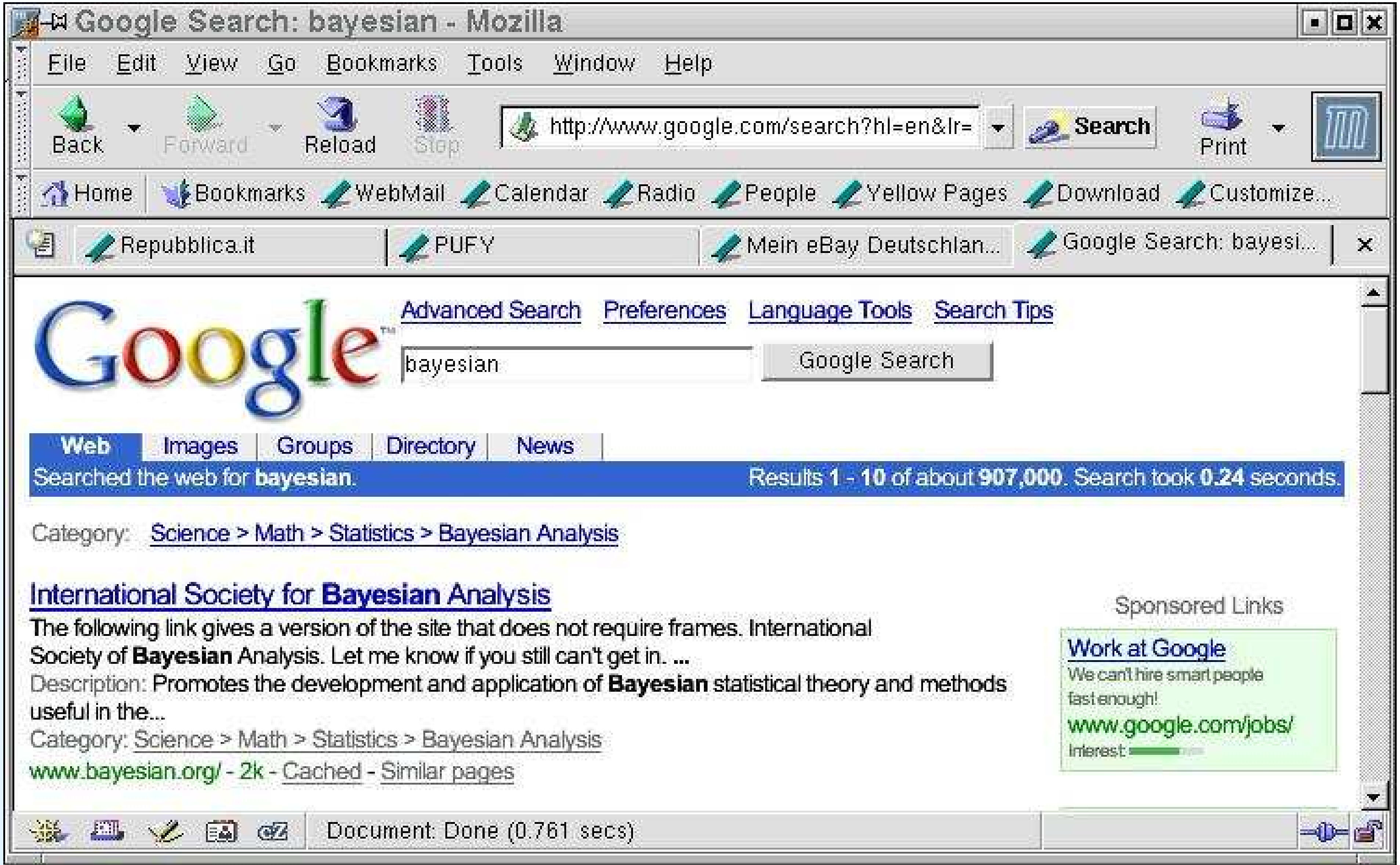
A last remark on model comparisons

- for a ‘serious’ probabilistic model comparisons,
  at least two well defined models are needed
- 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? $\longrightarrow$ Slides

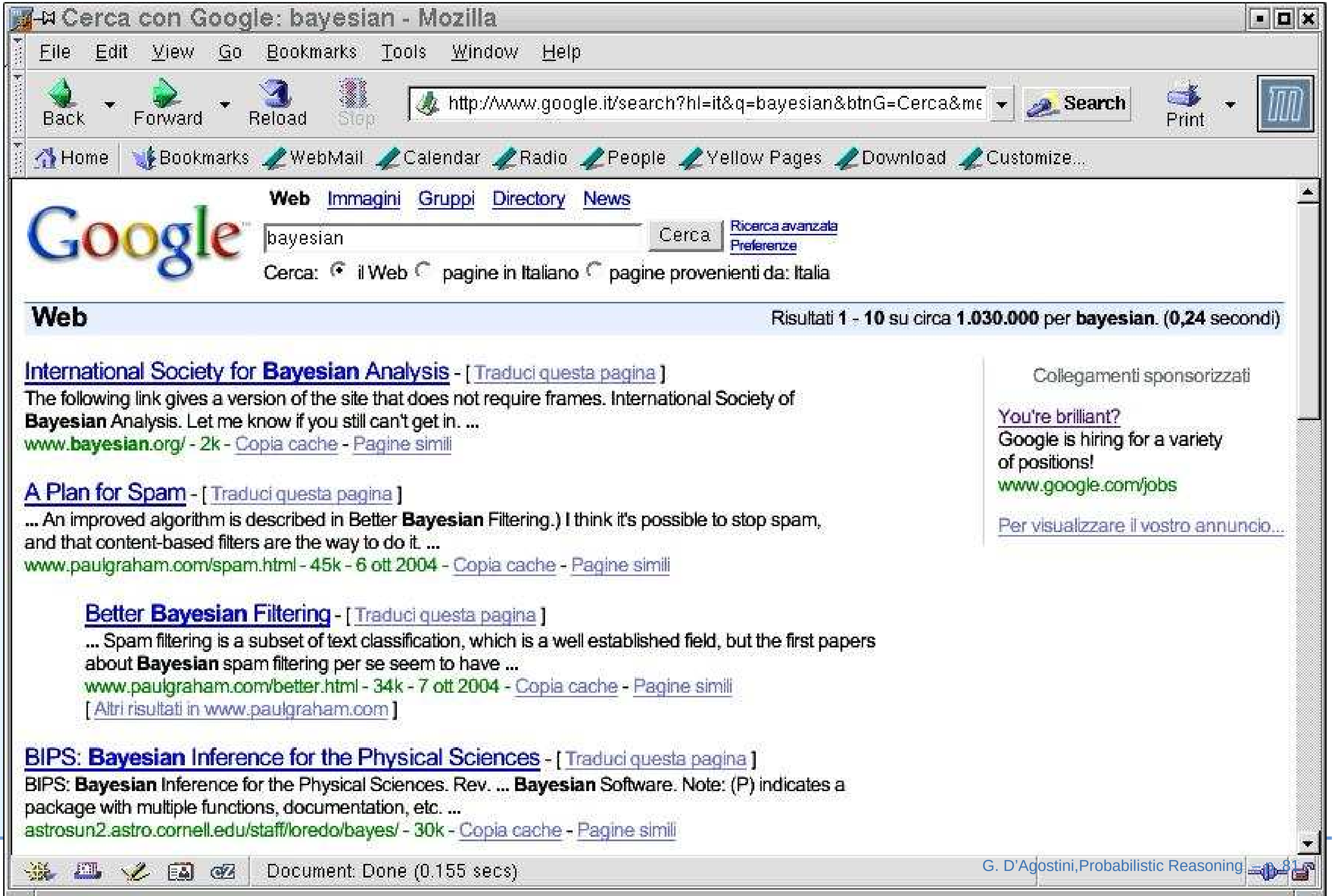
Conclusions

- Subjective probability recovers intuitive idea of probability.
- Nothing negative in the adjective 'subjective'. Just recognize, honestly, that probability depends on the status of knowledge, different from person to person.
- Most general concept of probability that can be applied to a large variety of cases.
- The adjective Bayesian comes from the intense use of Bayes' theorem to update probability once new data are acquired.
- Subjective probability is foundamental in decision issues, if you want to base decision on the probability of different events, together with the gain of each of them.
- Bayesian networks are powerful conceptuals/mathematical/ software tools to handle complex problems with variables related by probabilistic links.

Are Bayesians ‘smart’ and ‘brilliant’?



Are Bayesians ‘smart’ and ‘brilliant’?



End

FINE

Bet odds to express confidence

“The best way to explain it is, I’ll bet you fifty to one that you don’t find anything”
(Feynman)

Bet odds to express confidence

"The best way to explain it is, I'll bet you fifty to one that you don't find anything" (Feynman)

"It is a bet of 11,000 to 1 that the error on this result (the mass of Saturn) *is not 1/100th of its value"* (Laplace)

Bet odds to express confidence

"The best way to explain it is, I'll bet you fifty to one that you don't find anything" (Feynman)

"It is a bet of 11,000 to 1 that the error on this result (the mass of Saturn) *is not 1/100th of its value"* (Laplace)

$\longrightarrow$ 99.99% confidence on the result

Go Back

Processo di Biscardi

A single quote gives an idea of the talk show:

“Please, don’t speak more than two or three at the same time!”

Go Back