Probabilistic Inference and Applications to Frontier Physics – Part 2 –

Giulio D'Agostini

Dipartimento di Fisica

Università di Roma La Sapienza

Probability

What is probability?

What is probability?

"How much we believe something"

Versione velocizzata per MAPSES 2011
$\rightarrow$ slide mancanti sulla pagina web dedicata

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…

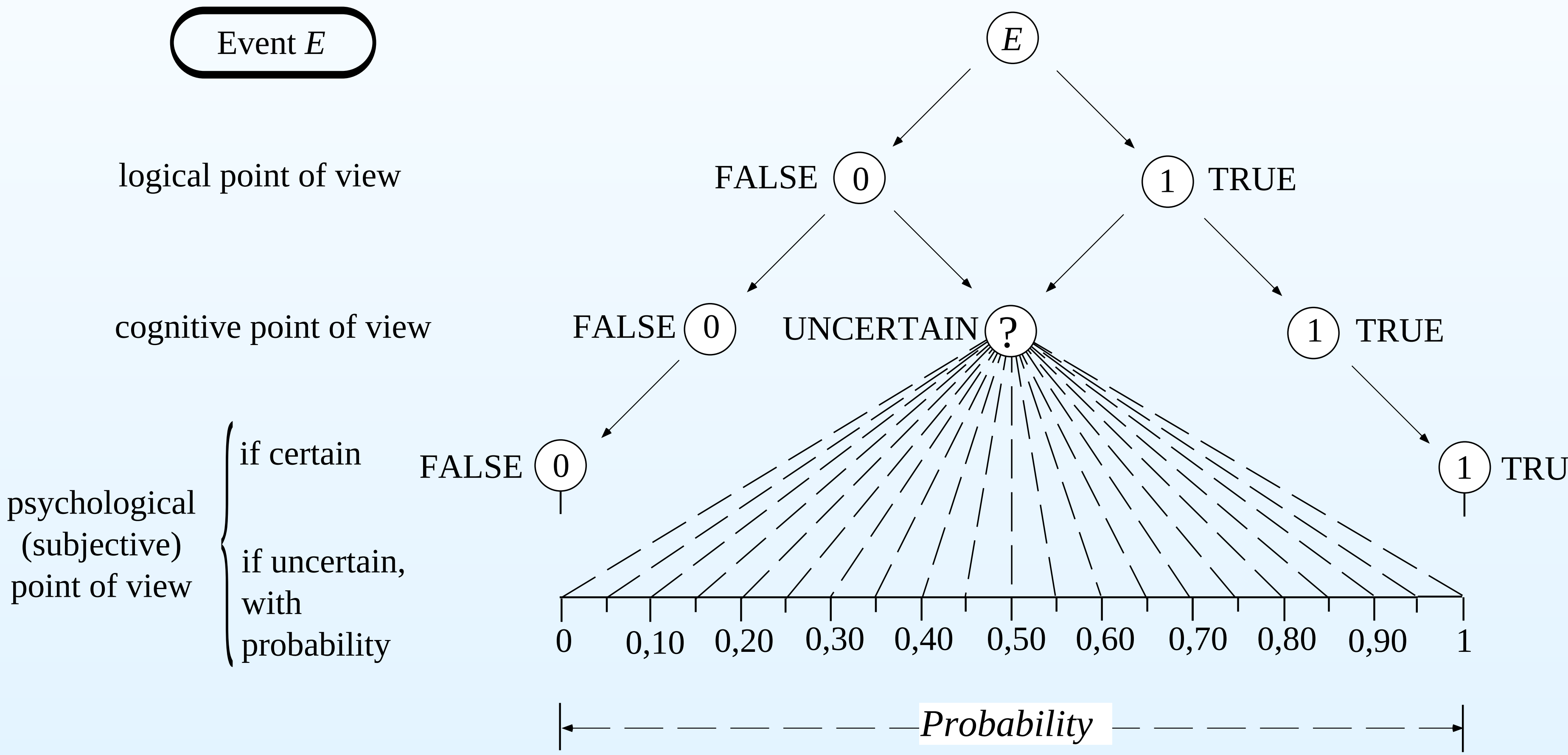
"Given the state of our knowledge about everything that could possible have any bearing on the coming true[1]*…, the numerical probability p of this event is to be a real number by the indication of which we try* in some cases *to setup a quantitative measure of the strength of our conjecture or anticipation,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



A reminder

Forse vale la pena di ricordare la famosa citazione di Einstein

La geometria, quando è certa, non dice

nulla del mondo reale,

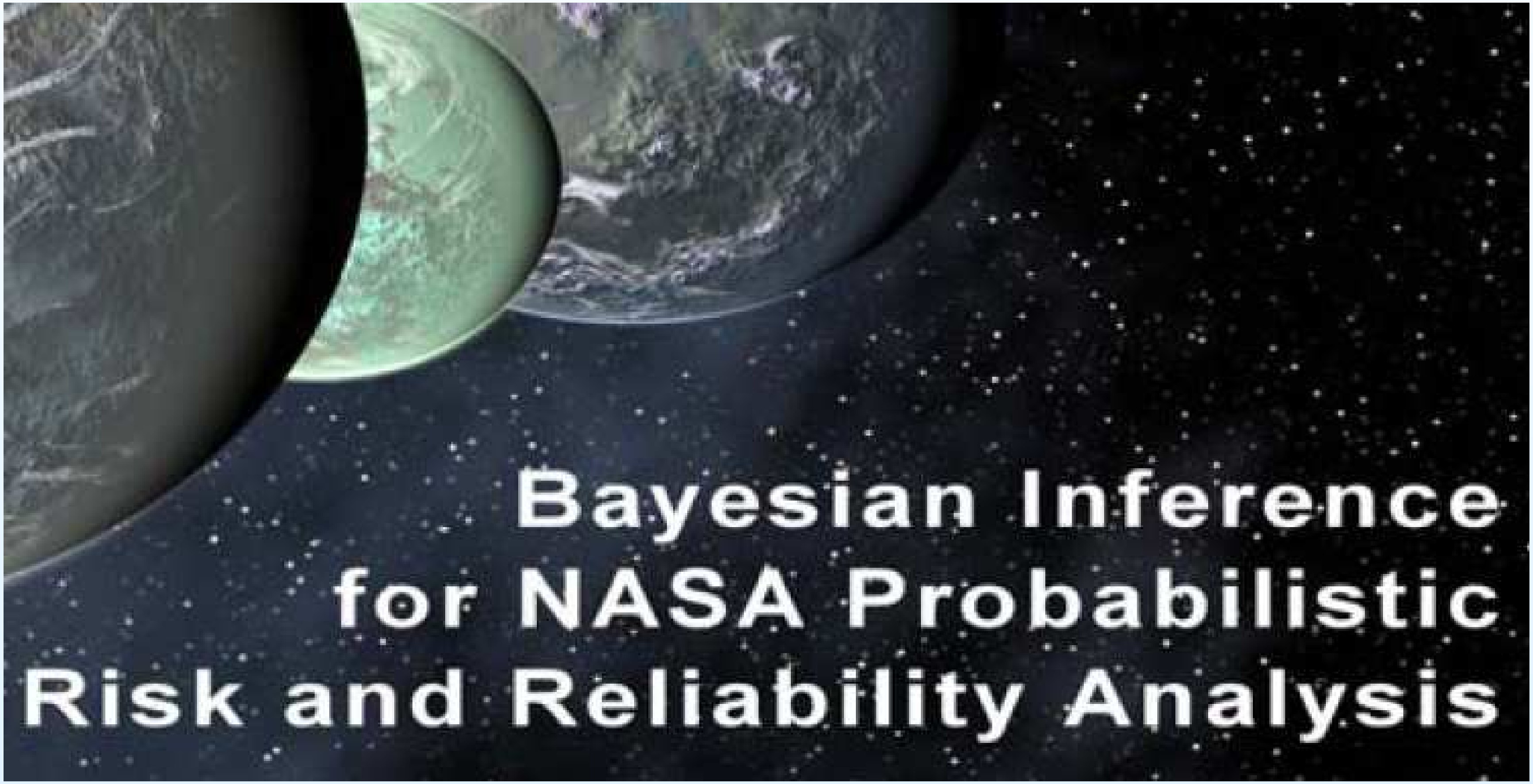
e, quando dice qualcosa a proposito della

nostra esperienza, è incerta.

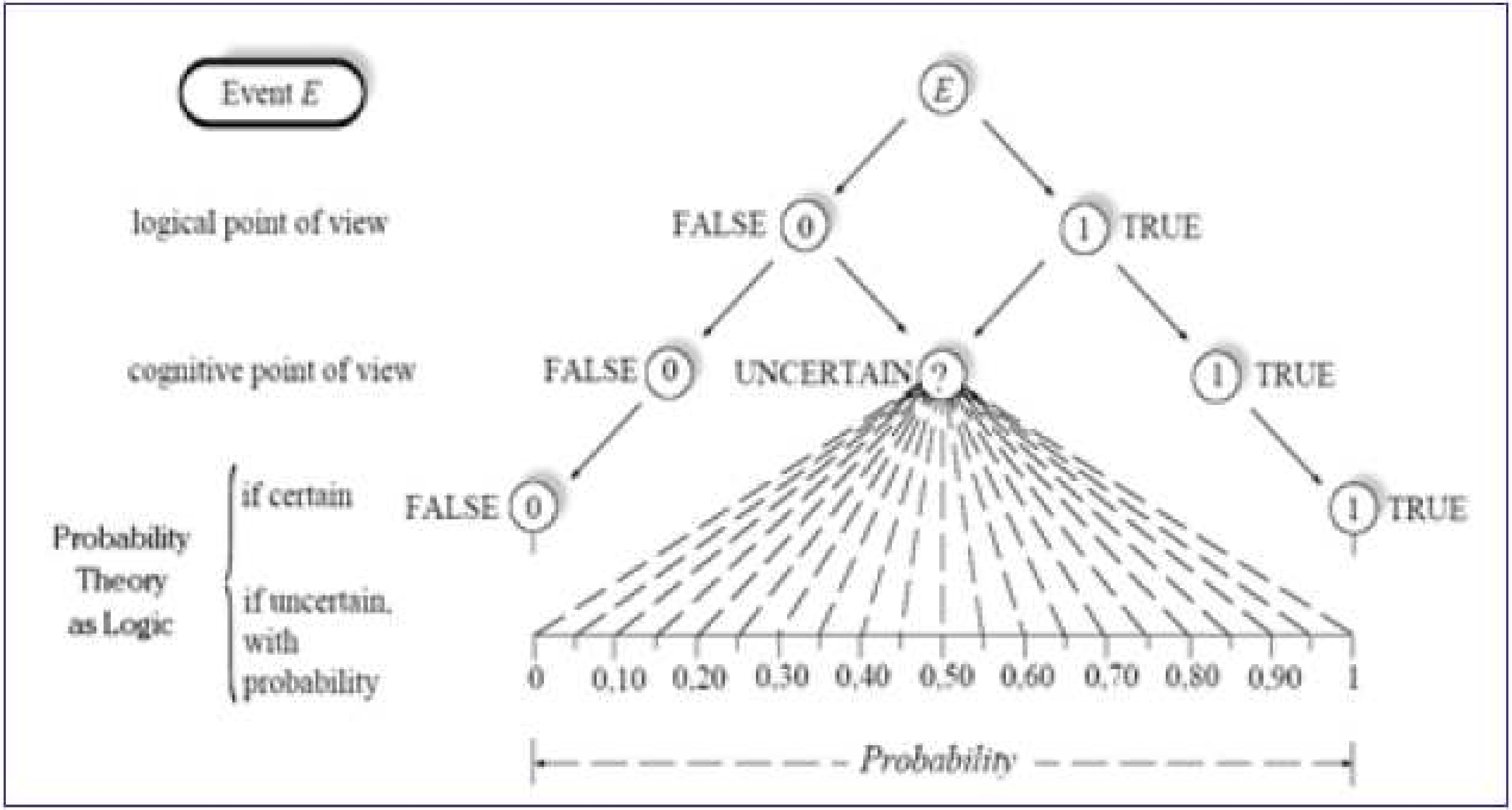
Chi vuole attenersi al regno del certo è meglio che si occupi di matematica che di fisica.

An helpful diagram

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



An helpful diagram



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

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

Uncertainty $\longrightarrow$ probability

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

Uncertainty $\longrightarrow$ probability

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

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

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

- *"Thus whenever we speak loosely of 'the probability of an event,' it is always to be understood: probability with regard to a certain given state of knowledge*" (Schrödinger)

Unifying role of subjective probability

- Wide range of applicability

Unifying role of subjective probability

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

Unifying role of subjective probability

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

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

Unifying role of subjective probability

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

  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>)

Unifying role of subjective probability

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

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

- Probability not bound to a single evaluation rule.

Unifying role of subjective probability

- Wide range of applicability
- Probability statements all have the same meaning no matter to what they refer and how the number has been evaluated.
  - $P(\text{rain next Saturday}) = 68\%$
  - $P(\text{Juventus will win Italian champion league}) = 68\%$
  - $P(\text{free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$
- Probability not bound to a single evaluation rule.
- In particular, combinatorial and frequency based 'definitions' are easily recovered as evaluation rules under well defined hypotheses.
- Keep separate concept from evaluation rule.

Confidence on the Higgs mass fron direct searches

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

What does it mean?

Confidence on the Higgs mass fron direct searches

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

What does it mean?

given only this piece of information from our LEP colleagues:

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

Confidence on the Higgs mass fron direct searches

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

What does it mean?

given only this piece of information from our LEP colleagues:

- What is $P(m_H \geq 114.4\,\text{GeV})$?
- What is $P(m_H \leq 114.4\,\text{GeV})$?

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

Confidence on the Higgs mass fron direct searches

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

What does it mean?

given only this piece of information from our LEP colleagues:

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

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

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

Confidence on the Higgs mass fron direct searches

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

What does it mean?

given only this piece of information from our LEP colleagues:

- What is $P(m_H \geq 114.4\,\text{GeV})$?
- What is $P(m_H \leq 114.4\,\text{GeV})$?

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

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

Siamo uomini o caporali?

Summary on probabilistic approach

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

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

- All the rest by logic

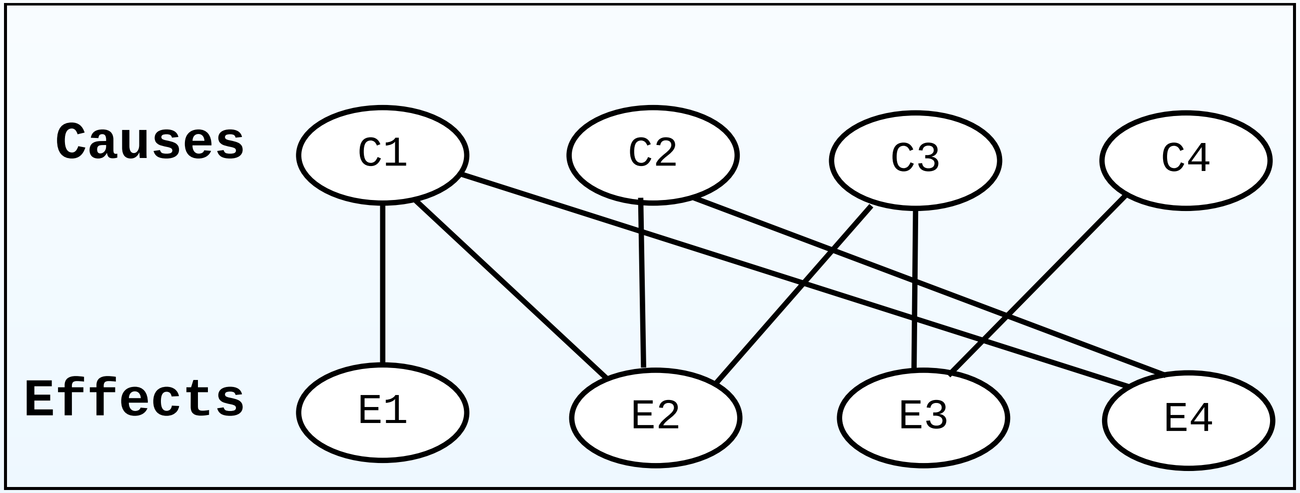
$\longrightarrow$ And, please, be coherent!

Inference

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

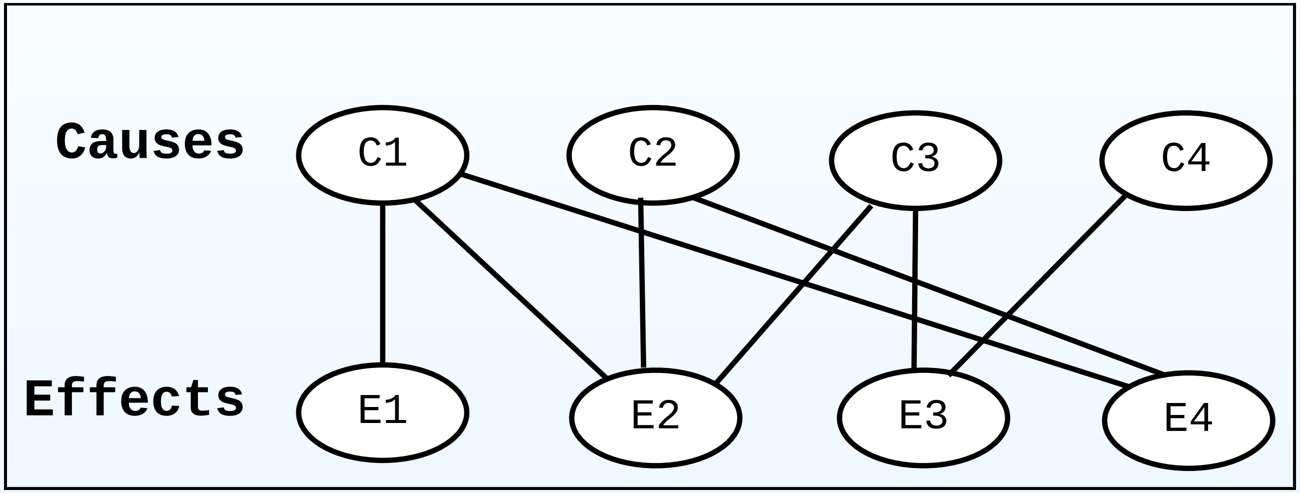
From causes to effects and back

Our original problem:



From causes to effects and back

Our original problem:

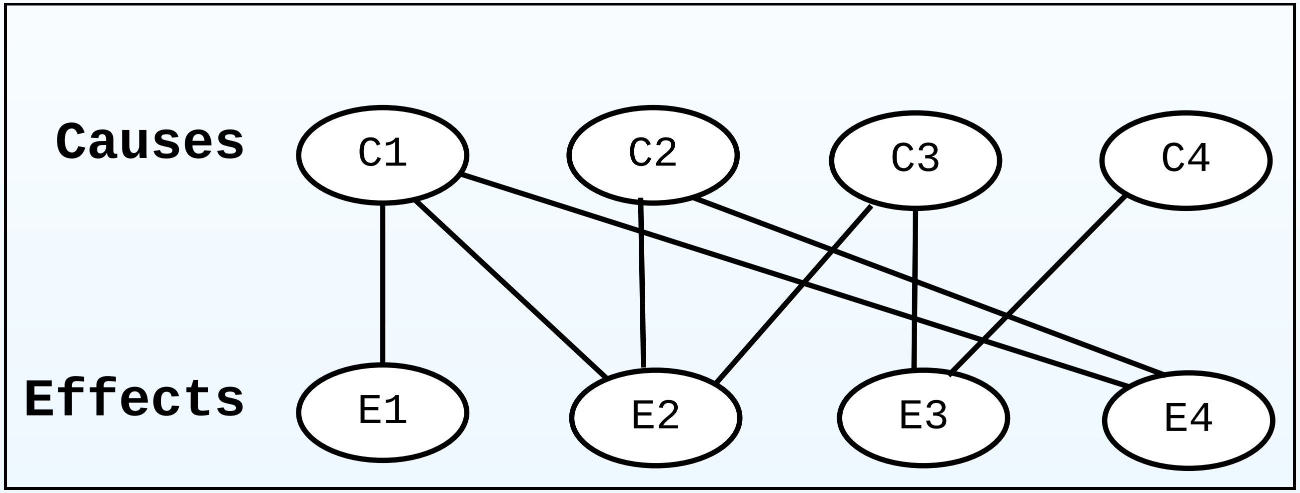


Our conditional view of probabilistic causation

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

From causes to effects and back

Our original problem:



Our conditional view of probabilistic causation

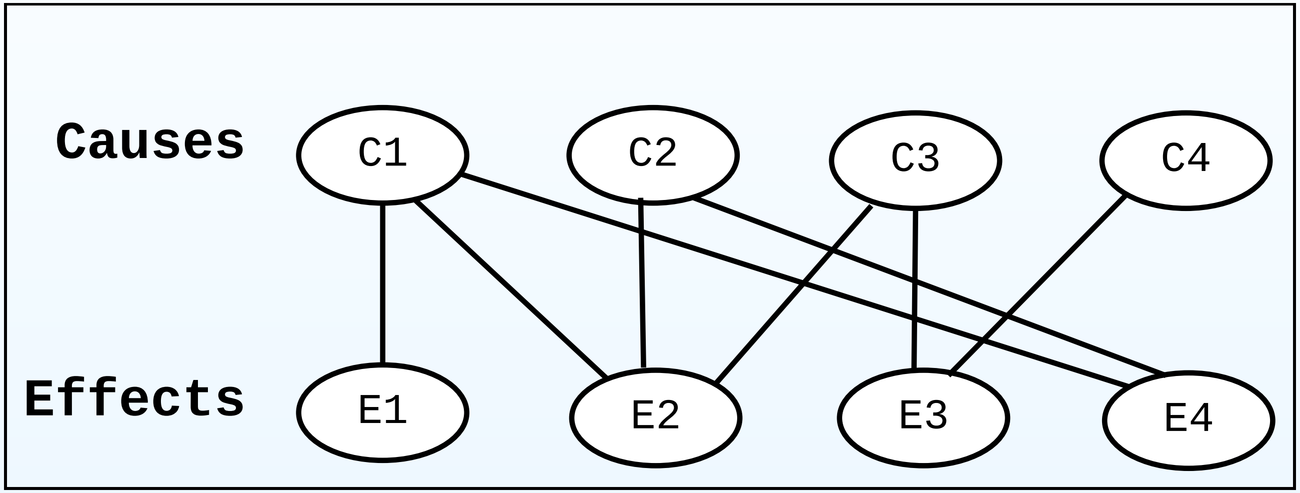
$$\boxed{P(E_i \,|\, C_j)}$$

Our conditional view of probabilistic inference

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

From causes to effects and back

Our original problem:



Our conditional view of probabilistic causation

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

Our conditional view of probabilistic inference

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

The fourth basic rule of probability:

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

Symmetric conditioning

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

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

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

Symmetric conditioning

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

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

"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."

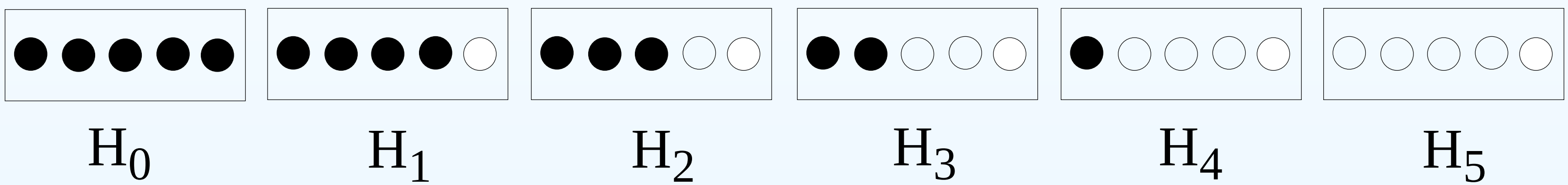
It follows

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

"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:

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

Collecting the pieces of information we need

Our tool:

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

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

Collecting the pieces of information we need

Our tool:

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

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

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:

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

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

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

Our prior belief about H_j

Collecting the pieces of information we need

Our tool:

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

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

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

Probability of E_i 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:

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

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

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

Probability of E_i taking account all possible H_j

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

Collecting the pieces of information we need

Our tool:

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

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

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

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!

$\longrightarrow$ Let's play with our toy

Naming the method

Some ‘remarks’ on formalism and notation.

(But nothing deep!)

Naming the method

Some ‘remarks’ on formalism and notation.

(But nothing deep!)

From now on it is only a question of

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

Naming the method

Some 'remarks' on formalism and notation.

(But nothing deep!)

From now on it is only a question of

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

Moving to continuous quantities:

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

Bayes theorem

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

Bayes theorem

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

Neglecting the background state of information I:

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

Bayes theorem

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

Neglecting the background state of information I:

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

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

Bayes theorem

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

Neglecting the background state of information I:

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

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

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

Bayes theorem

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

Neglecting the background state of information I:

$$
\begin{aligned}
\frac{P(H_j \,|\, E_i)}{P(H_j)} &= \frac{P(E_i \,|\, H_j)}{P(E_i)} \\
P(H_j \,|\, E_i) &= \frac{P(E_i \,|\, H_j)}{P(E_i)}\, P(H_j) \\
P(H_j \,|\, E_i) &= \frac{P(E_i \,|\, H_j) \cdot P(H_j)}{\sum_j P(E_i \,|\, H_j) \cdot P(H_j)} \\
P(H_j \,|\, E_i) &\propto P(E_i \,|\, H_j) \cdot P(H_j)
\end{aligned}
$$

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

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

Updating the knowledge by new observations

Let us repeat the experiment:

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

Exercises and discussions

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

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

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

Odd ratios and Bayes factor

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

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

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

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

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

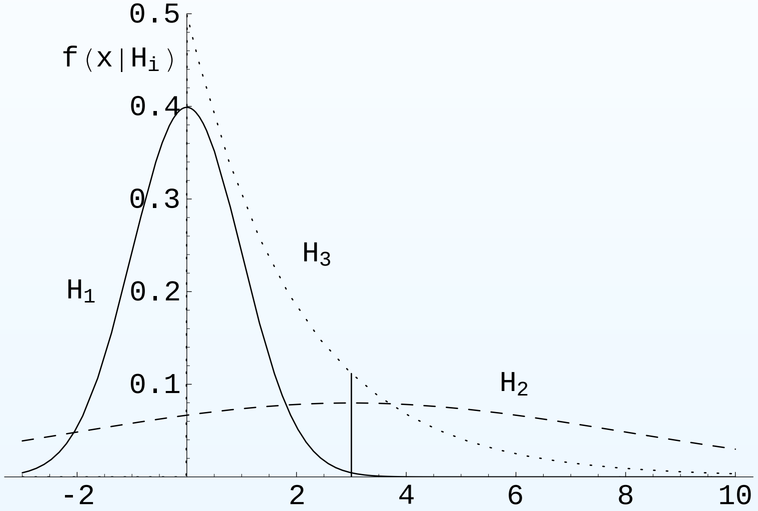
There are some advantages in expressing Bayes theorem in terms of odd ratios:

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

Further comments on first meeting

The three models example

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



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

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

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

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

A last remark

A last remark on model comparisons

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

A last remark

A last remark on model comparisons

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

A last remark

A last remark on model comparisons

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

A last remark

A last remark on model comparisons

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

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 $\frac{\text{P}_\circ(\text{HIV})}{\text{P}_\circ(\overline{\text{HIV}})} = 1$,

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

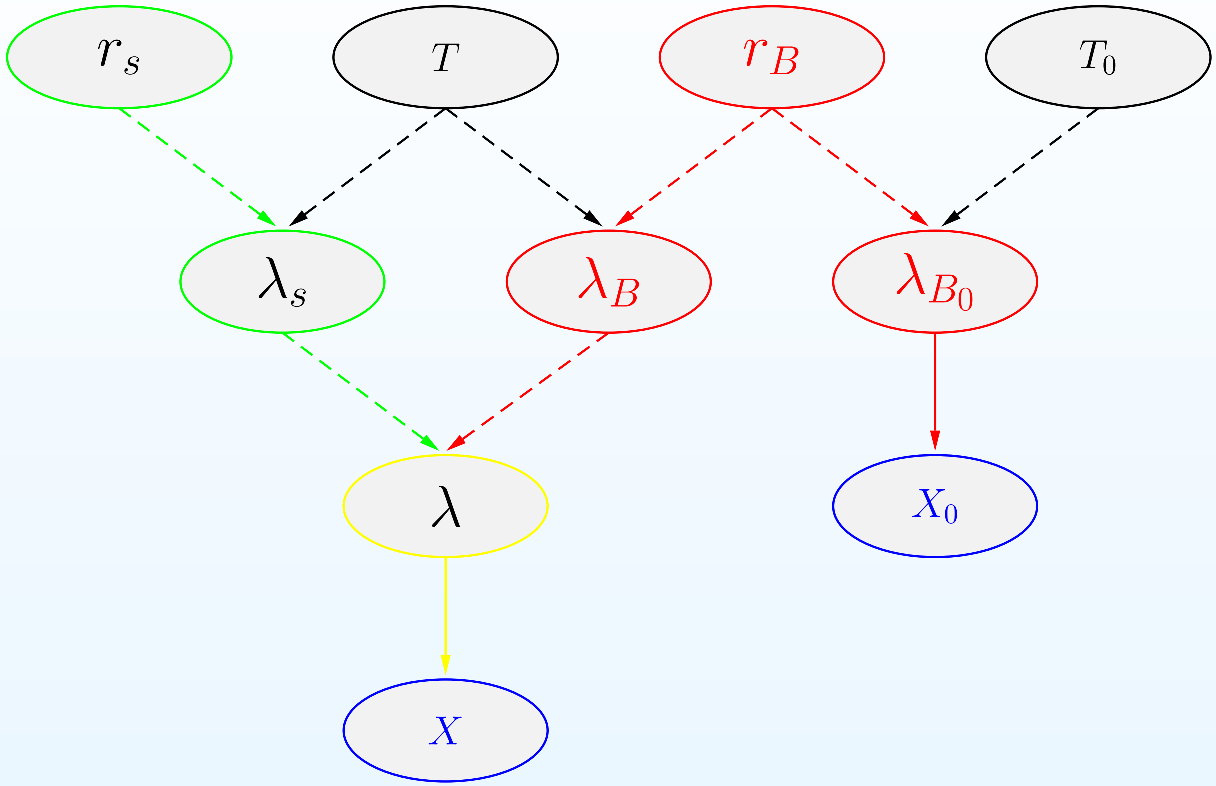
- This is typical in arbitrary inversions, and often also in frequentistic prescriptions that are used by the practitioners to form their confidence on something:

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

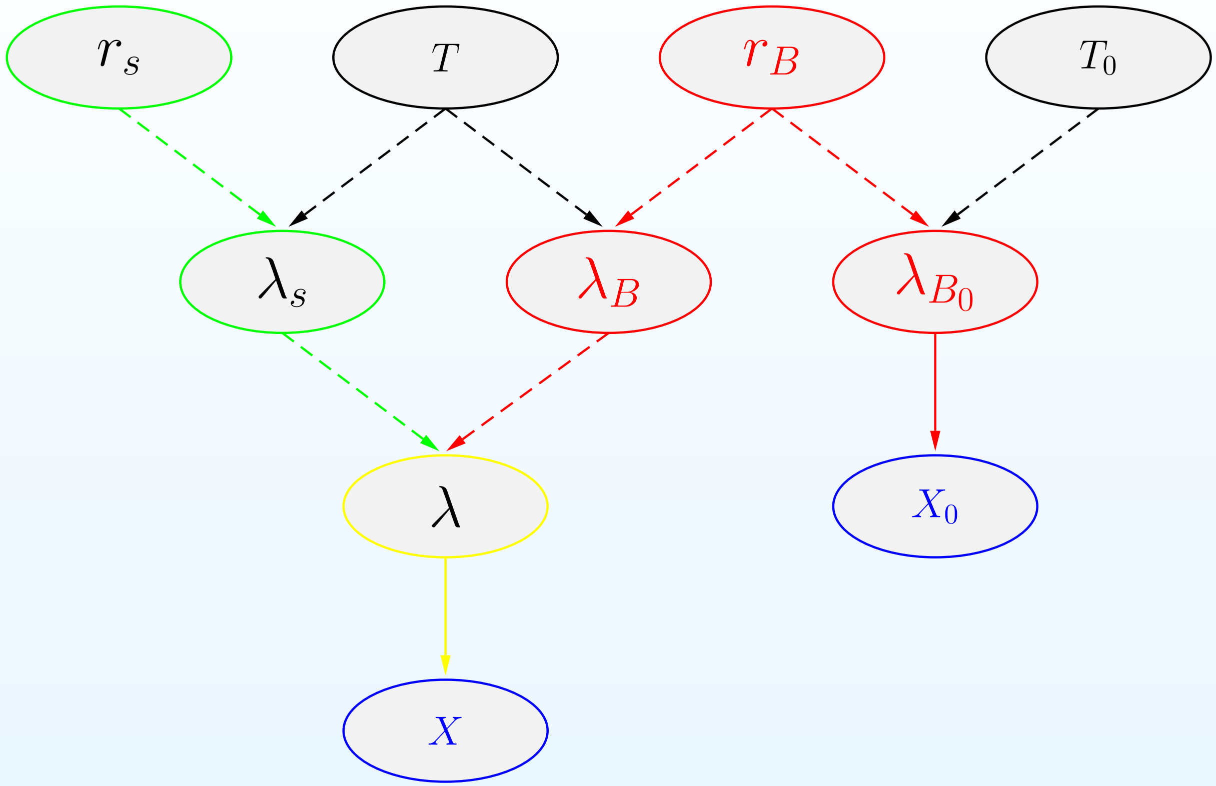
- but they criticize the Bayesian approach because it takes into account priors explicitly !

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

Inferring a rate of a Poisson process



Inferring a rate of a Poisson process

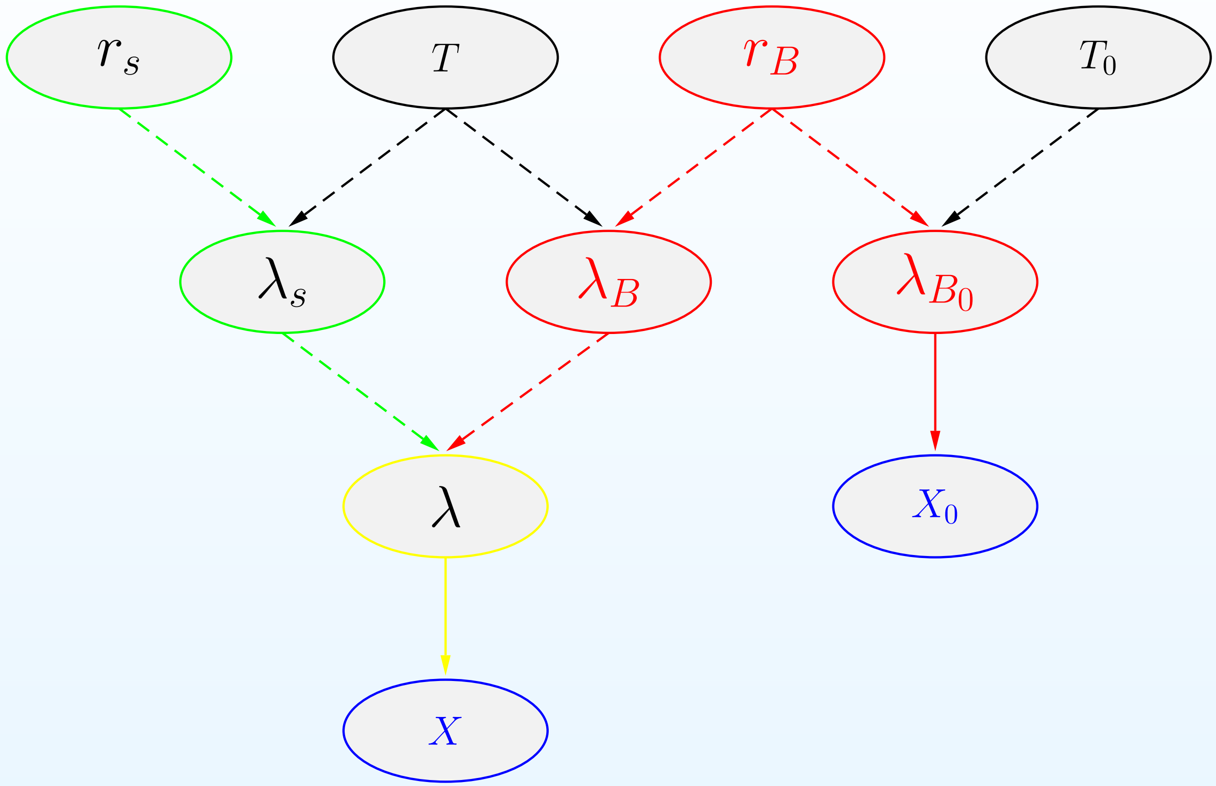


$$f(r_s, r_b \,|\, x, x_0, T, T_0) \;\propto\; f(x, x_0 \,|\, r_s, rb, T, T_0) \cdot f_0(r_s, r_b)$$

$$\propto\; f(x \,|\, (r_s + r_b) \cdot T) \cdot f(x_0 \,|\, r_b \cdot T_0) \cdot f_0(r_s) \cdot f_0(r$$

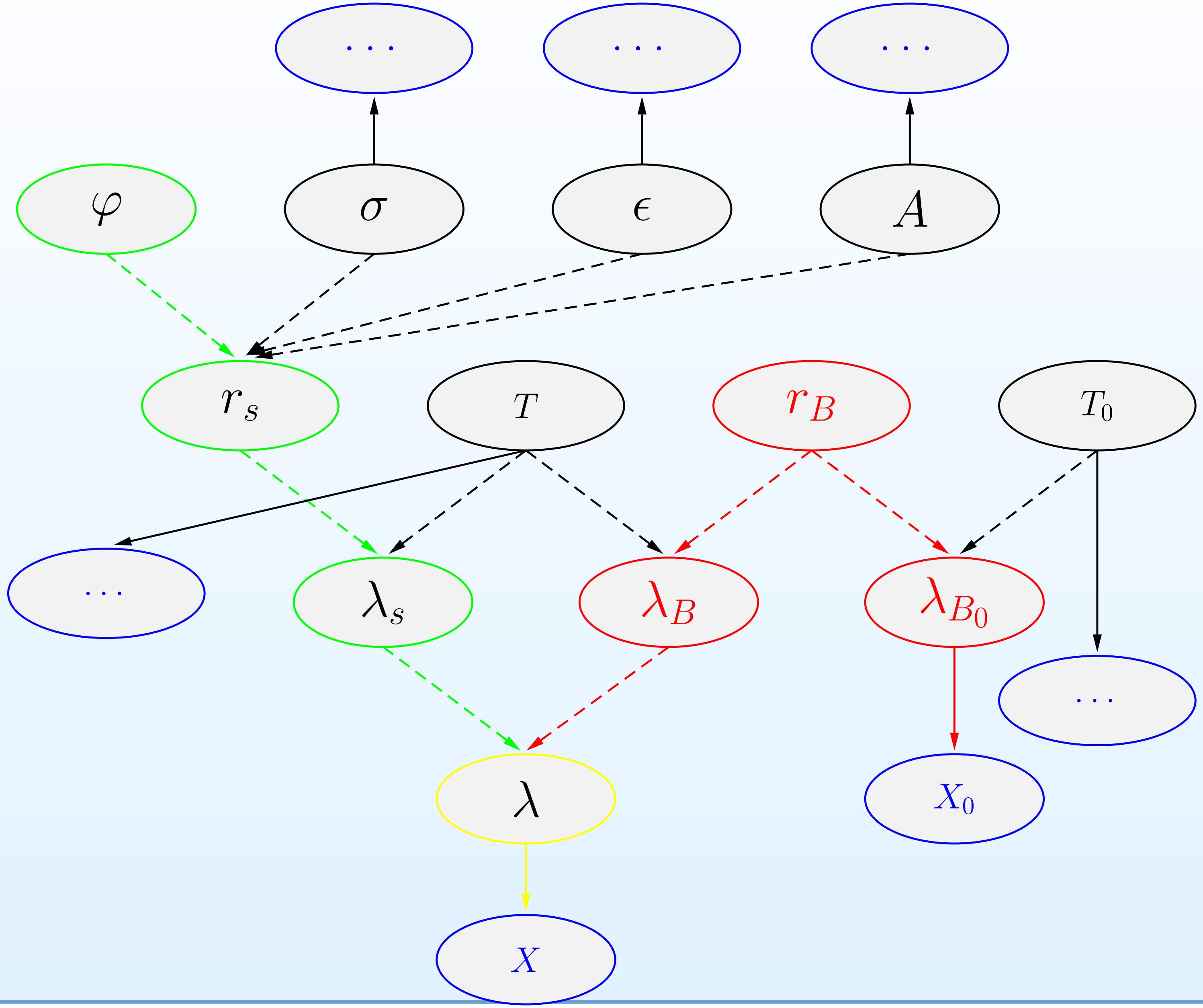
$$f(r_s \,|\, x, x_0, T, T_0) \;\propto\; \int_0^\infty f(r_s, r_b \,|\, x, x_0, T, T_0)\, dr_b$$

Inferring a rate of a Poisson process



$$
\begin{aligned}
f(r_s, r_b \,|\, x, x_0, T, T_0) &\propto f(x, x_0 \,|\, r_s, rb, T, T_0) \cdot f_0(r_s, r_b) \\
&\propto f(x \,|\, (r_s + r_b) \cdot T) \cdot f(x_0 \,|\, r_b \cdot T_0) \cdot f_0(r_s) \cdot f_0(r \\
f(r_s \,|\, x, x_0, T, T_0) &\propto \int_0^{\infty} f(r_s, r_b \,|\, x, x_0, T, T_0)\, dr_b \quad \Rightarrow JAGS
\end{aligned}
$$

Making the model more realistic



Upper/lower limits

"Ogni limite ha una pazienza" (Totò)

Upper/lower limits

"Ogni limite ha una pazienza" (Totò)

A very simple problem:

- counting experiment described by a binomial of unkown p;
- our aim is to 'get' p, in the sense of evaluating $f(p\,|\,\text{data})$;
- we make n trials and get $x = 0$ successes.

Upper/lower limits

"Ogni limite ha una pazienza" (Totò)

A very simple problem:

- counting experiment described by a binomial of unkown p;
- our aim is to 'get' p, in the sense of evaluating $f(p\,|\,\text{data})$;
- we make n trials and get $x = 0$ successes.

Bayes' theorem:

$$f(p\,|\,n, x=0, \mathcal{B}) \;=\; \frac{f(x=0\,|\,n,\mathcal{B})\, f_0(p)}{\int_0^1 f(x=0\,|\,n,\mathcal{B})\, f_0(p)\, dp}$$

with

$$f(x=0\,|\,n,\mathcal{B}) \;=\; (1-p)^n$$

Bernoulli trials $\Rightarrow N$ boxes $\longrightarrow \infty$

Conceptually exactly equivalente to the 6-box problem:

- "success" $\leftrightarrow$ "white ball"
- $p \leftrightarrow$ "proportion of white balls"
- $f(p\,|\,x,n) \;\leftrightarrow\; P(H_i\,|\,x,n)$

Bernoulli trials $\Rightarrow N$ boxes $\longrightarrow \infty$

Conceptually exactly equivalente to the 6-box problem:

- "success" $\leftrightarrow$ "white ball"
- $p \leftrightarrow$ "proportion of white balls"
- $f(p\,|\,x,n) \;\leftrightarrow\; P(H_i\,|\,x,n)$

- as log as we continue to extract only black boxes we get more and more convinced ('confident') that Nature has presented us H_0, although we cannot exclude H_1, a bit less H_2, etc.
  $\Rightarrow$ Rigorously speaking, only H_N gets falsified!

Bernoulli trials $\Rightarrow N$ boxes $\longrightarrow \infty$

Conceptually exactly equivalente to the 6-box problem:

- "success" $\leftrightarrow$ "white ball"
- $p \leftrightarrow$ "proportion of white balls"
- $f(p\,|\,x,n) \;\leftrightarrow\; P(H_i\,|\,x,n)$

- as log as we continue to extract only black boxes we get more and more convinced ('confident') that Nature has presented us H_0, although we cannot exclude H_1, a bit less H_2, etc.
  $\Rightarrow$ Rigorously speaking, only H_N gets falsified!

$$P(H_N\,|\,n, x=0) = 0 \quad \leftrightarrow \quad f(p=1\,|\,n, x=0) = 0$$

Inference about p from 0 counts

Using flat prior, i.e. $f_0(p) = k$

$$
\begin{aligned}
f(p \,|\, n, x = 0, \mathcal{B}) &= (n+1)\,(1-p)^n \\
p_{max} &= 0 \\
E(p) &= \frac{1}{n+2} \longrightarrow \frac{1}{n} \\
\sigma(p) &= \sqrt{\frac{(n+1)}{(n+3)(n+2)^2}} \longrightarrow \frac{1}{n} \\
p_{95\%UL} &= 1 - \sqrt[n+1]{0.05}.
\end{aligned}
$$

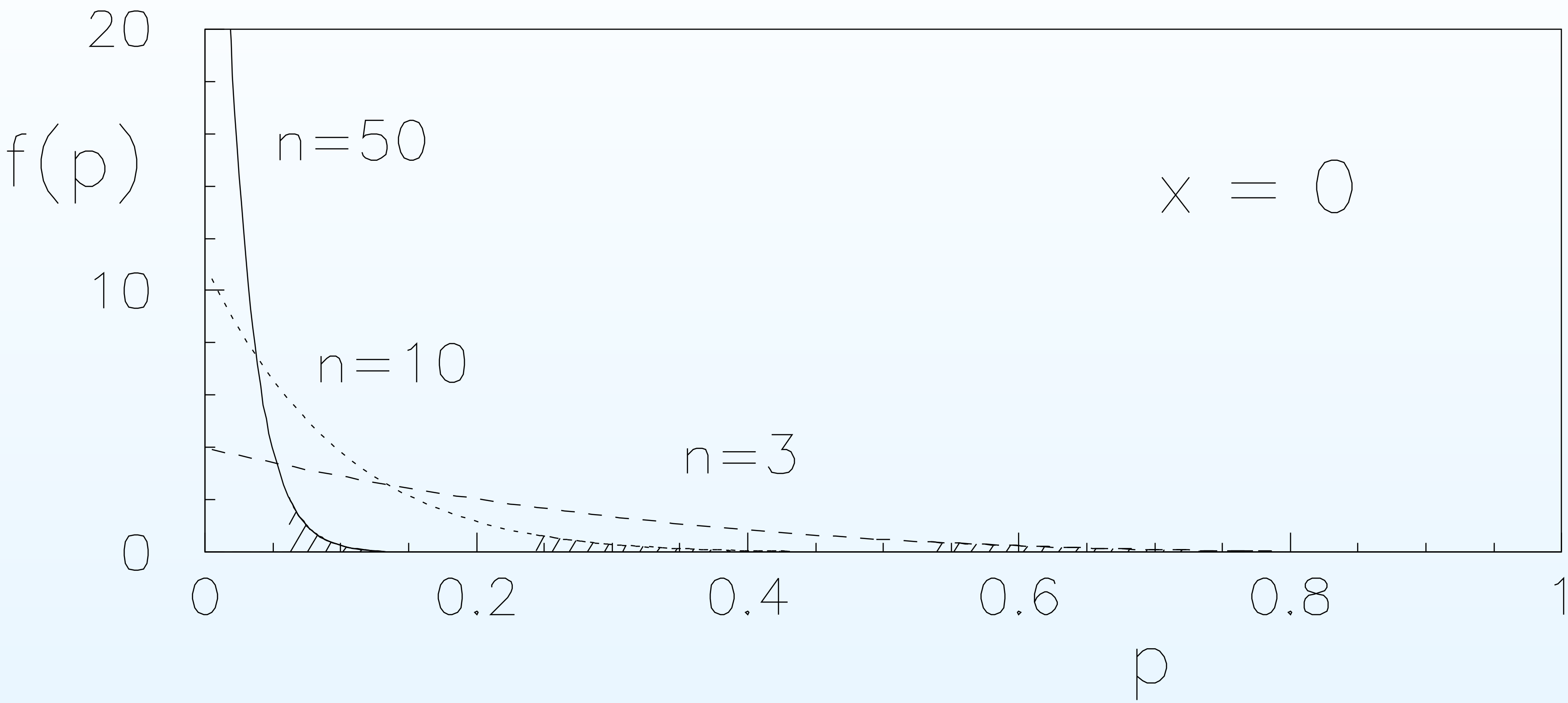
Inference about p from 0 counts

Using flat prior, i.e. $f_0(p) = k$

$$\begin{aligned} f(p\,|\,n, x=0, \mathcal{B}) &= (n+1)\,(1-p)^n \\ p_{max} &= 0 \\ E(p) &= \frac{1}{n+2} \rightarrow \frac{1}{n} \\ \sigma(p) &= \sqrt{\frac{(n+1)}{(n+3)(n+2)^2}} \rightarrow \frac{1}{n} \\ p_{95\%UL} &= 1 - \sqrt[n+1]{0.05}. \end{aligned}$$

As n increases, we get more and more convinced that p has to be very small

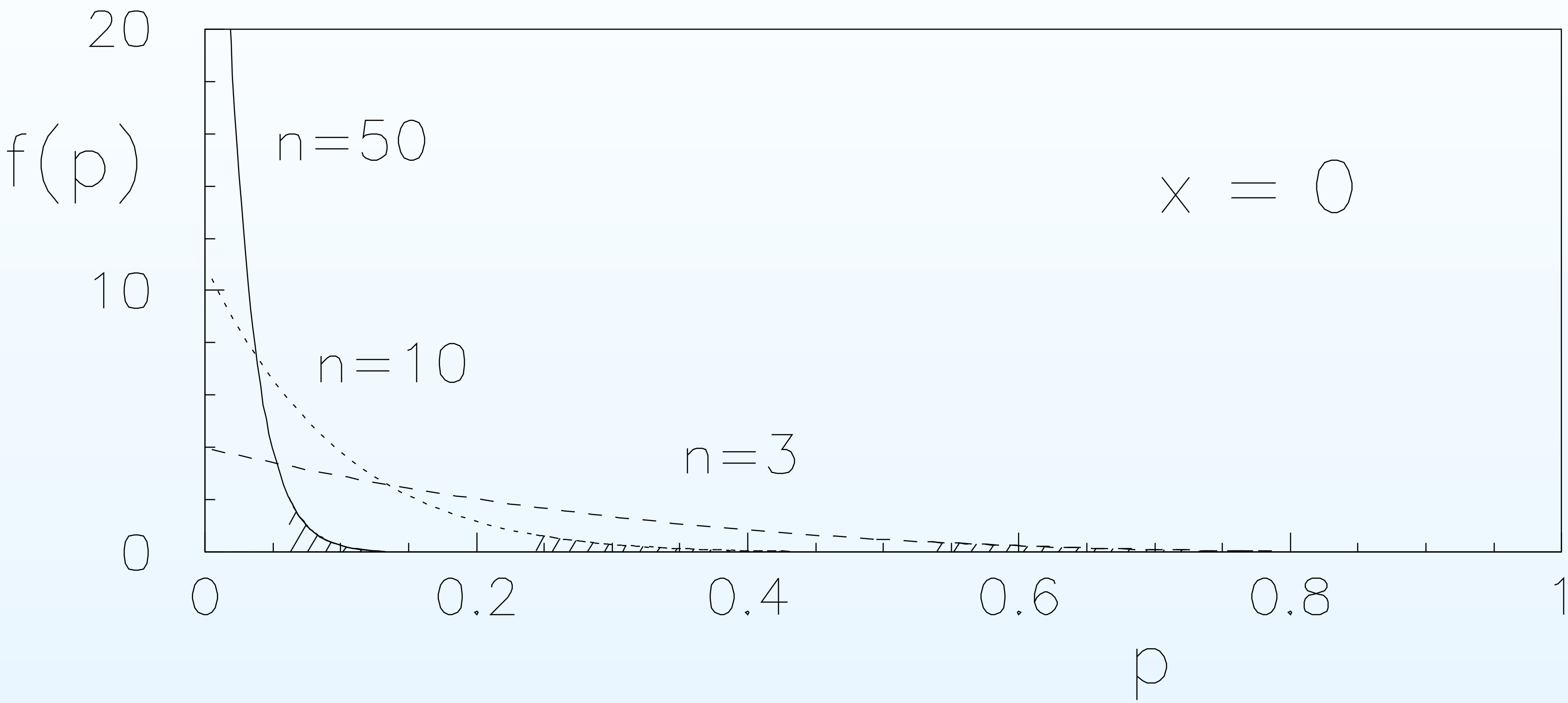
Inference about p from 0 counts



$$f(p \,|\, n, x=0, \mathcal{B}) \quad = \quad (n+1)\,(1-p)^n$$

$$p_{95\%UL} \quad = \quad 1 - \sqrt[n+1]{0.05}.$$

Inference about p from 0 counts



Seems not problematic at all, but we have to remember that it relies on

$$f(x = 0 \,|\, n, \mathcal{B}) = (1-p)^n$$
$$f_0(p) = k$$

When likelihoods are non ‘closed’

Where is the problem? (Flat priors are regulary used, and are often assumed in other approaches, e.g. ML methods)

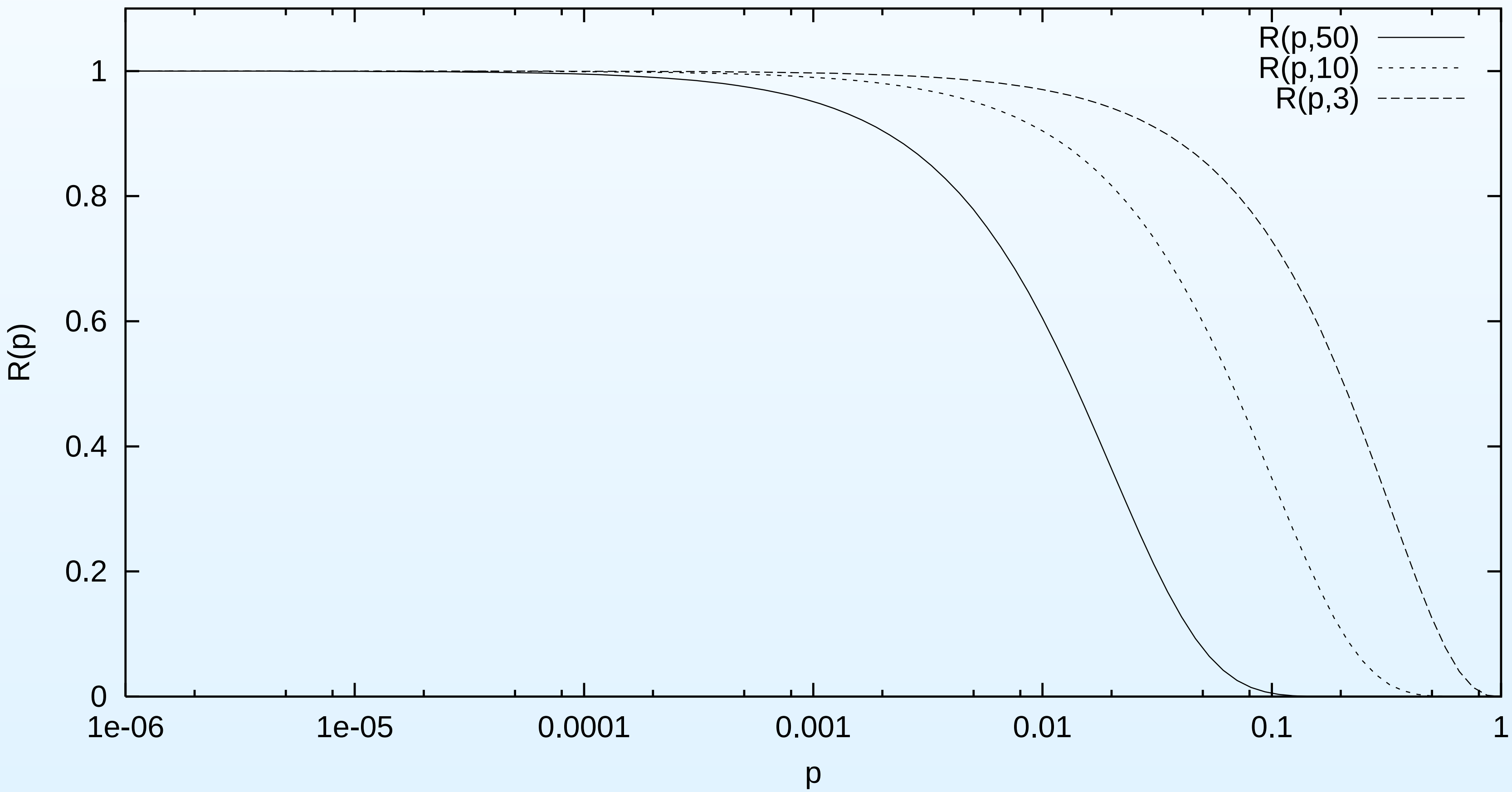
When likelihoods are non ‘closed’

The major problem is not in $f_0(p)$, but rather in the likelihood $f(x = 0, | n, \mathcal{B})$ that does not go to zero on both sides!

When likelihoods are non ‘closed’

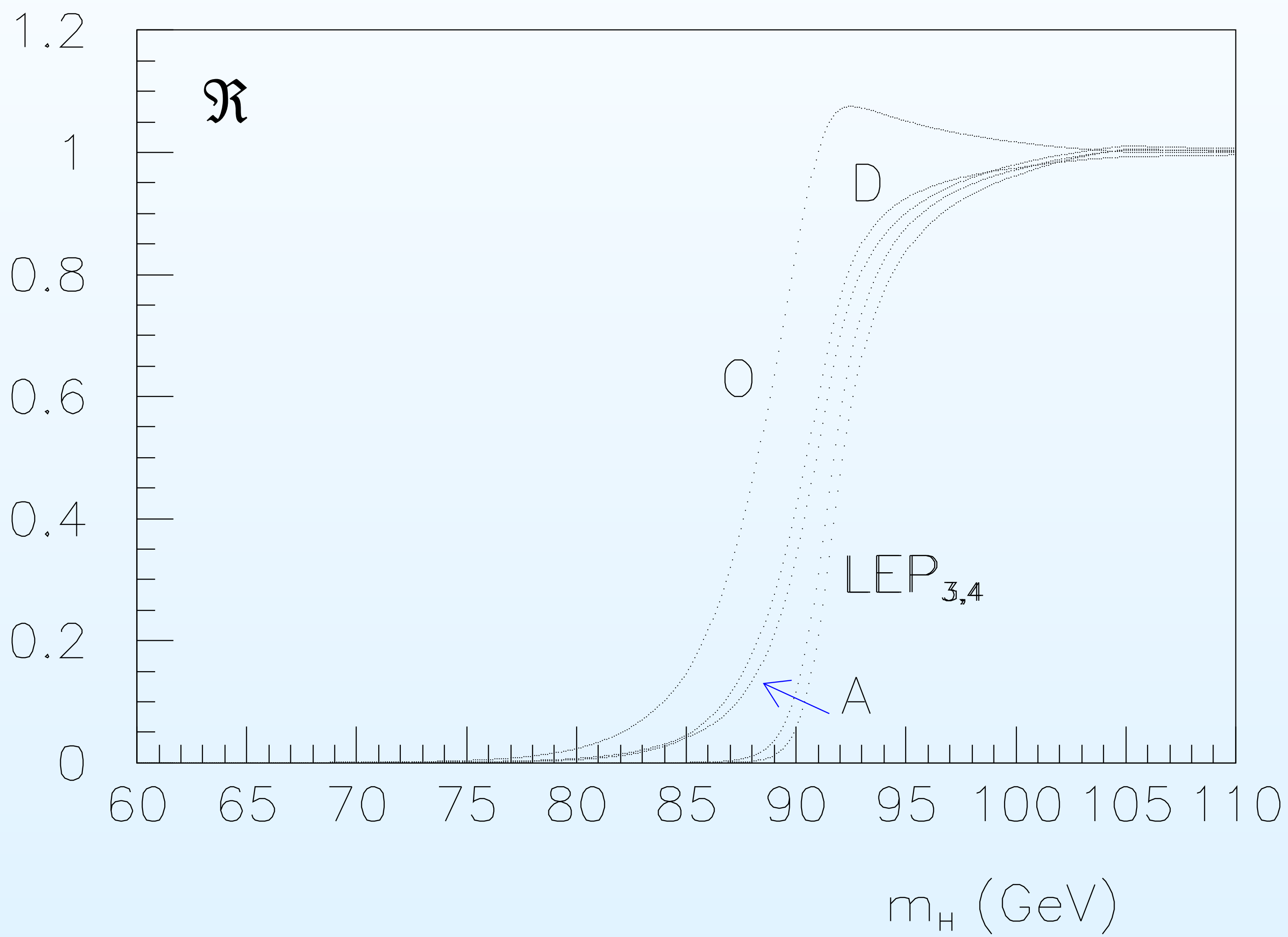
The major problem is not in $f_0(p)$, but rather in the likelihood $f(x=0,\,|\,n,\mathcal{B})$ that does not go to zero on both sides!

A different representation of the likelihood (properly rescaled) helps:



A probabilistic lower bound for the Higgs?

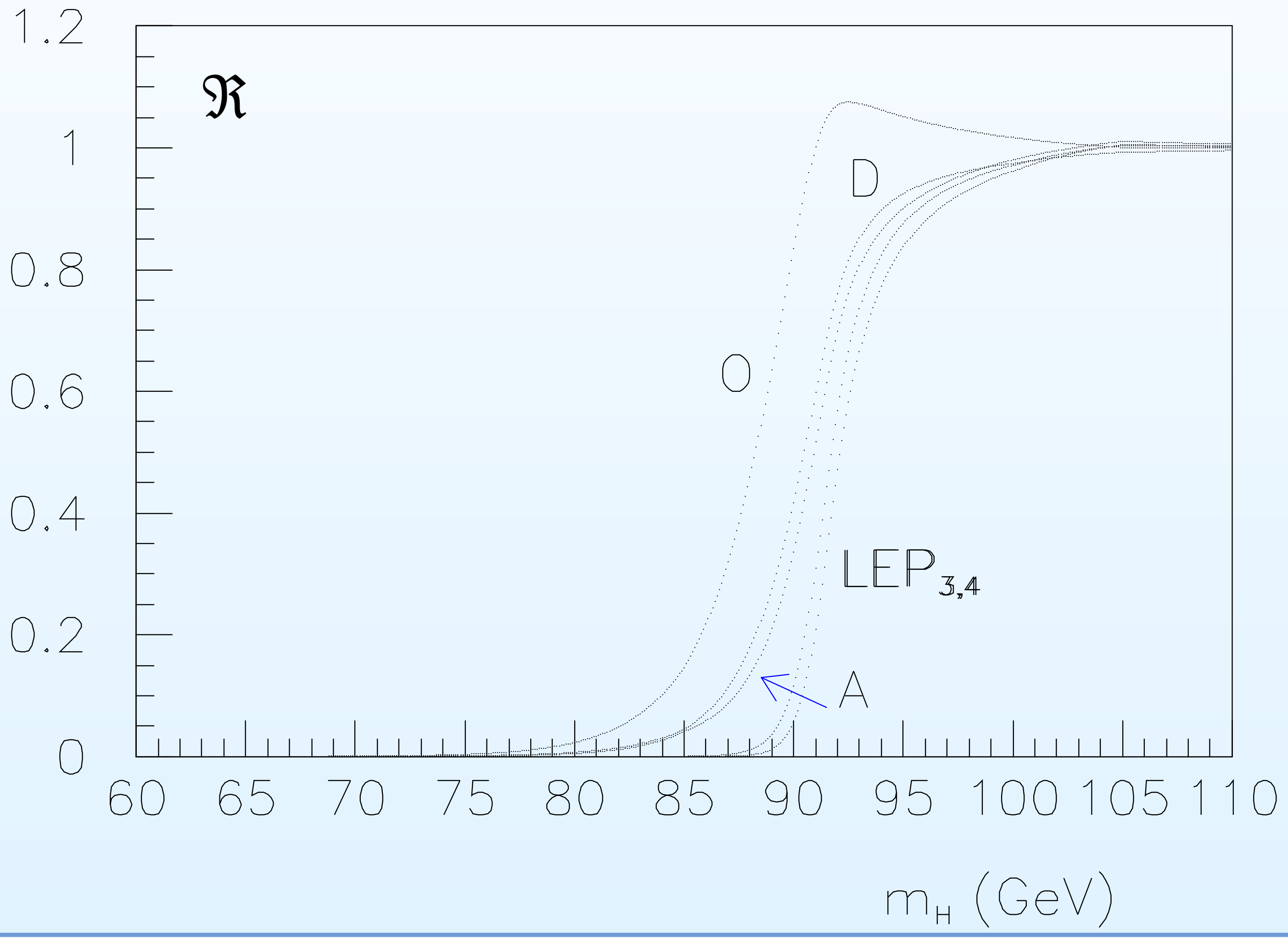
A similar think happens with the direct searches of the Higgs particle at LEP



(1999 figure, but substance unchanged)

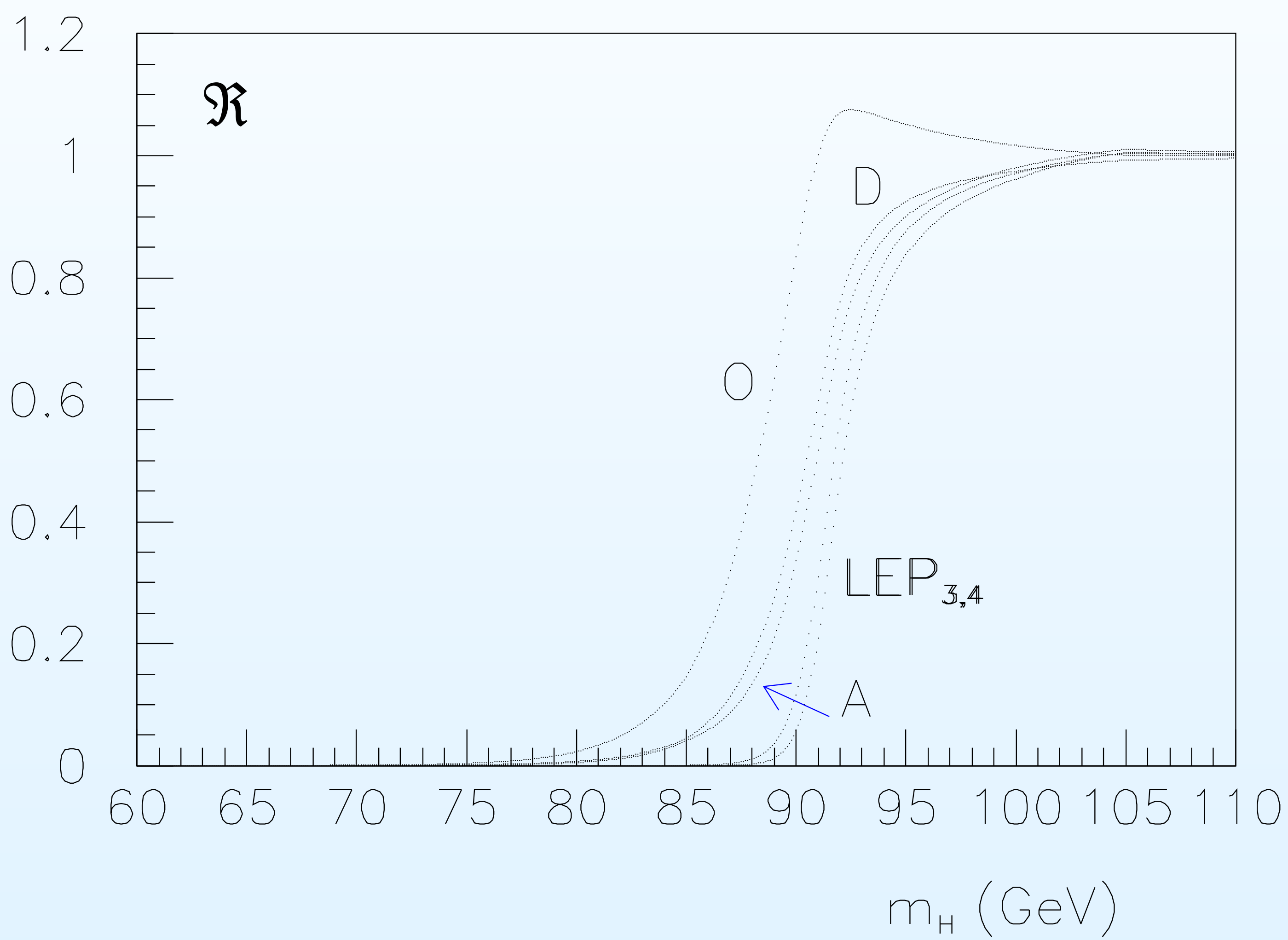
A probabilistic lower bound for the Higgs?

Impossible to express our confidence in probabilistic terms, unless we define an upper cut!



A probabilistic lower bound for the Higgs?

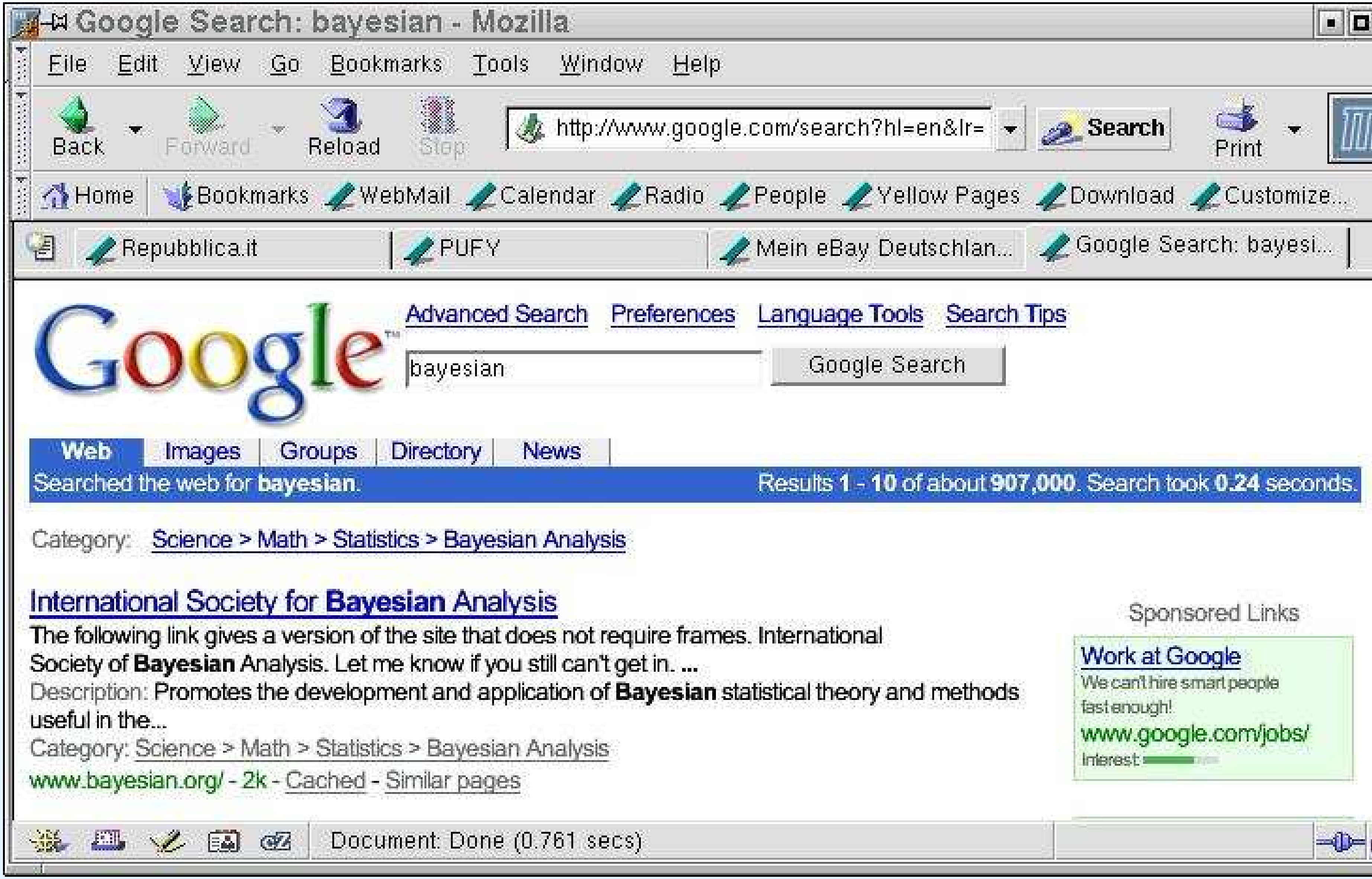
Confidence limit ⇒ Sensitivity bound



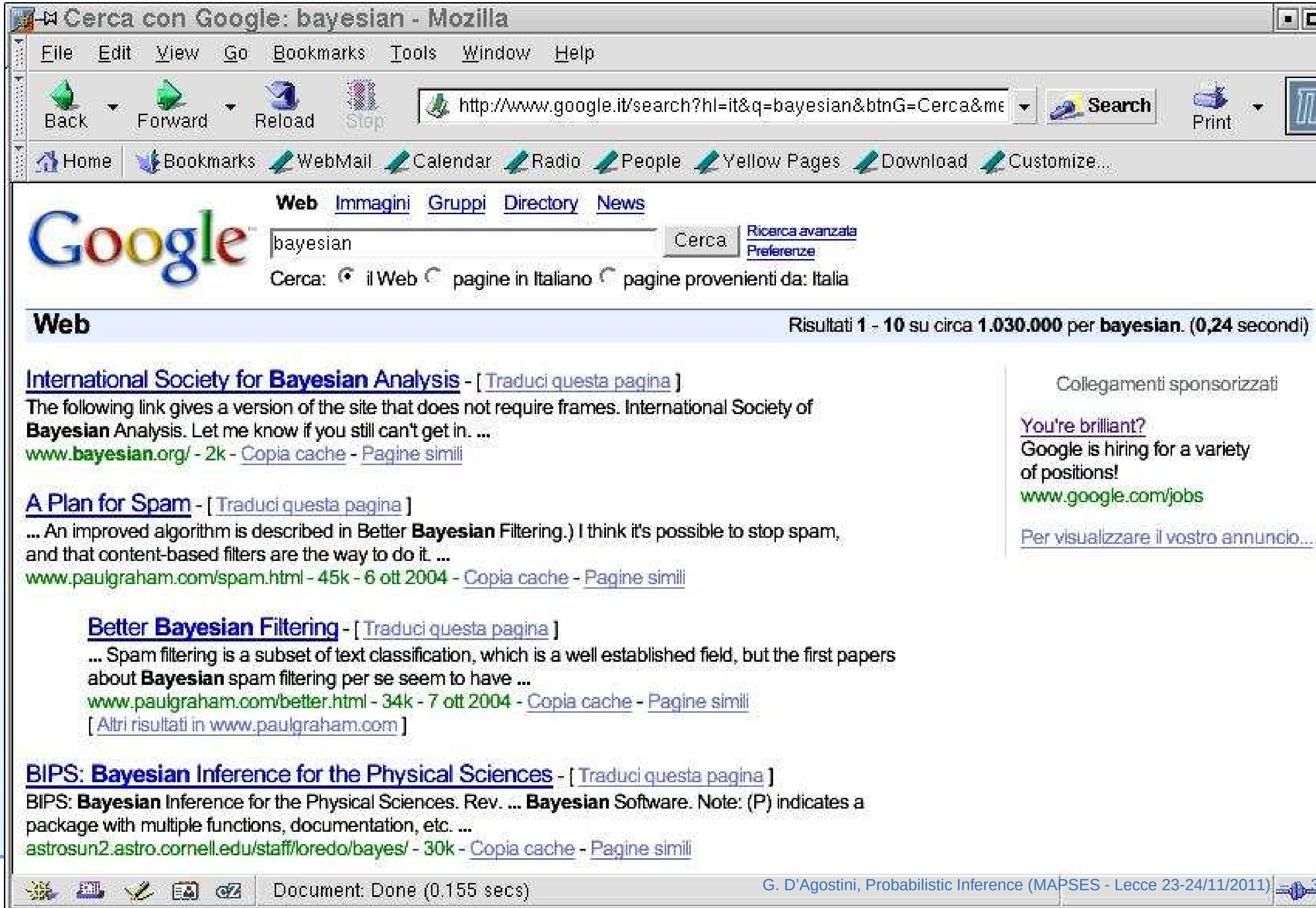
Conclusions

- Probabilistic reasoning helps . . .
  . . . at least to avoid conceptual errors.
- Probabilistic statements can attributed, quantitaively and consistently, to all ‘objects’ respect to which we are in condition of uncertainty
- . . . allowing us to make meaninful statements concerning true values.
- In particular uncertainties due to systematic errors can be easily included
- Several ‘standard’ methods (like Least Square, etc.) can be easily recovered under well defined assumptions.
- But if this is not the case, nowdays there are no longer excuses to avoid the more general approach.
- Bayesian networks are a powerful conceptual and computational tool.

Are Bayesians ‘smart’ and ‘brilliant’?



Are Bayesians 'smart' and 'brilliant'?



End

FINE