Probabilistic Reasoning in Physics

– inference, forecasting, decision –

Giulio D’Agostini

giulio.dagostini@roma1.infn.it

Dipartimento di Fisica

Università di Roma La Sapienza

“Probability is good sense reduced to a calculus” (Laplace)

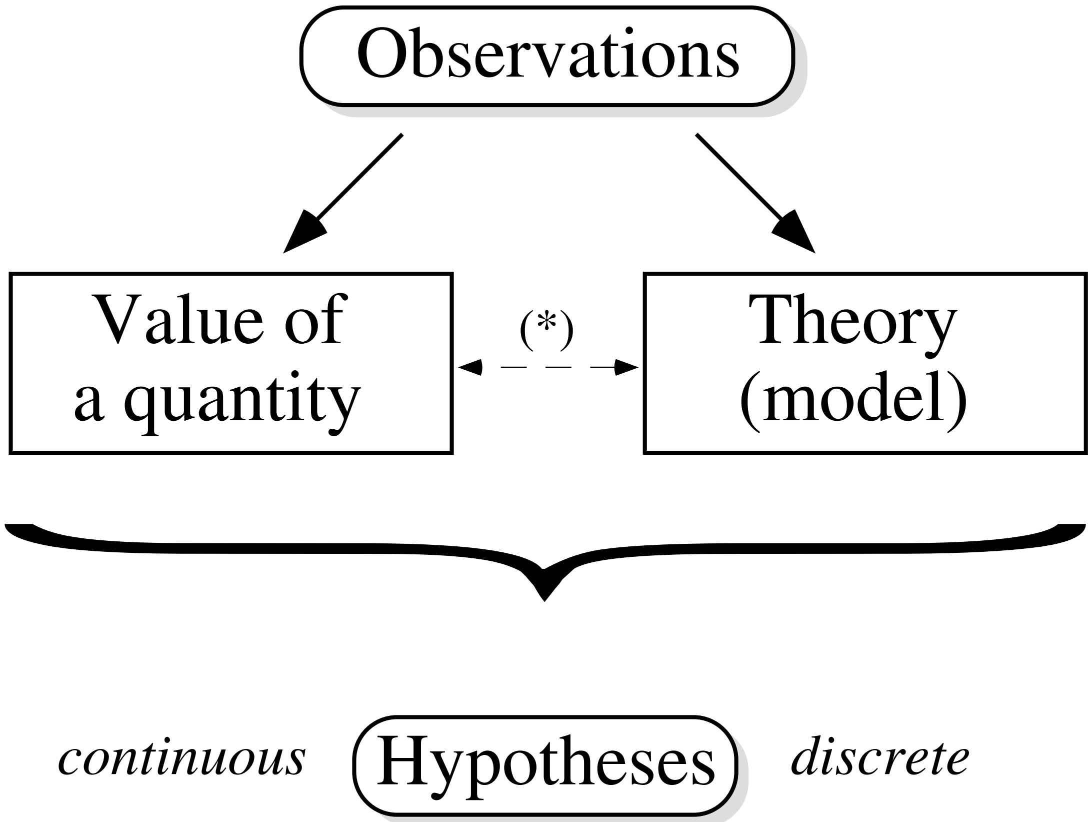
Outline

- "Science and hypothesis" (Poincaré)
- Uncertainty, probability, decision.
- Causes $\longleftrightarrow$ Effects

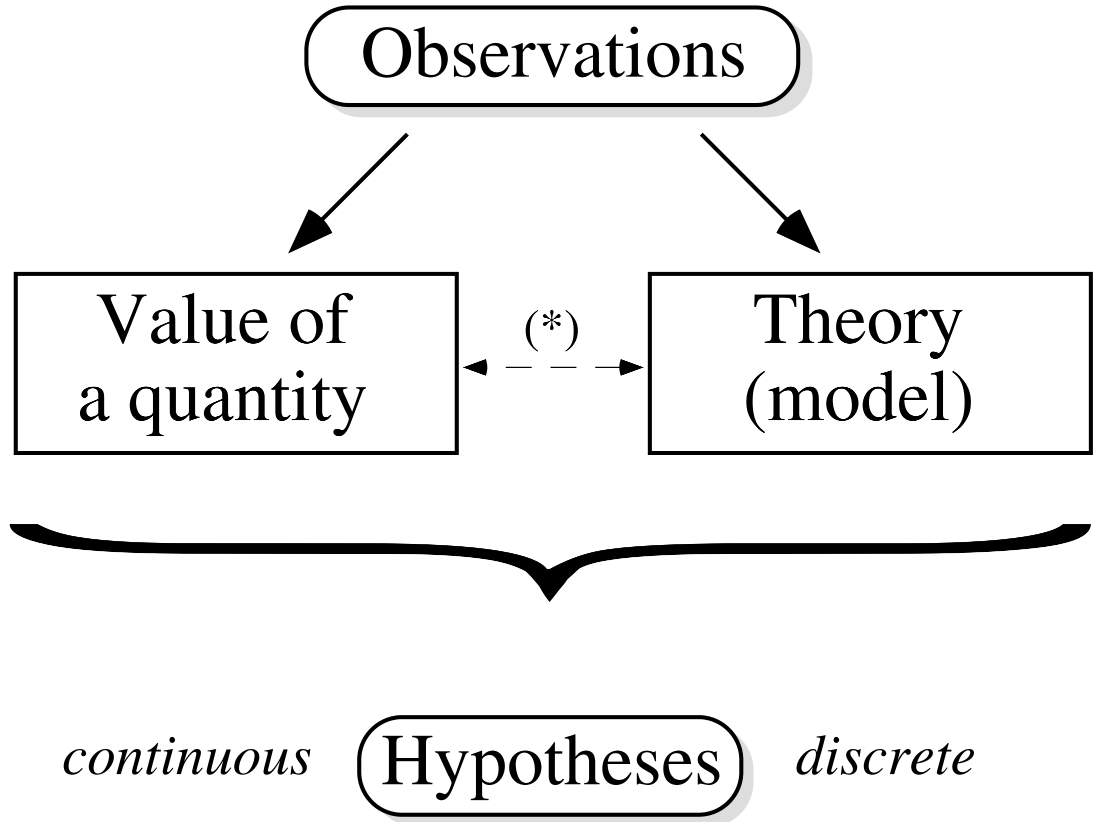
  "The essential problem of the experimental method" (Poincaré).
- A toy model and its physics analogy: the six box game

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

Physics

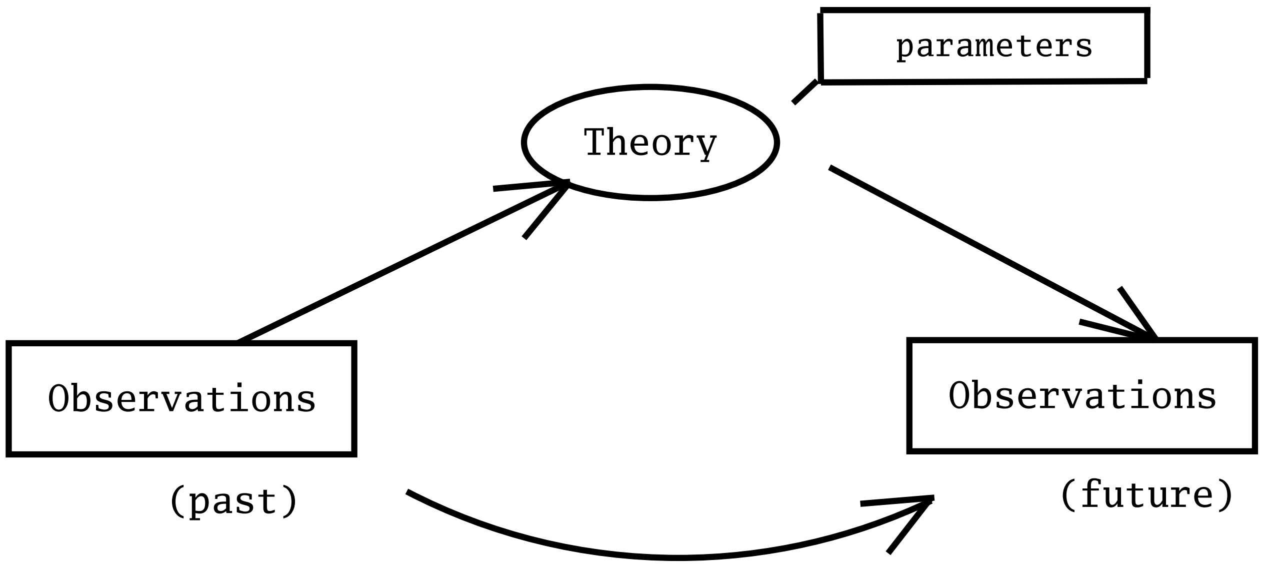

Observations
Value of a quantity
(*)
Theory (model)
continuous
Hypotheses
discrete

Physics



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

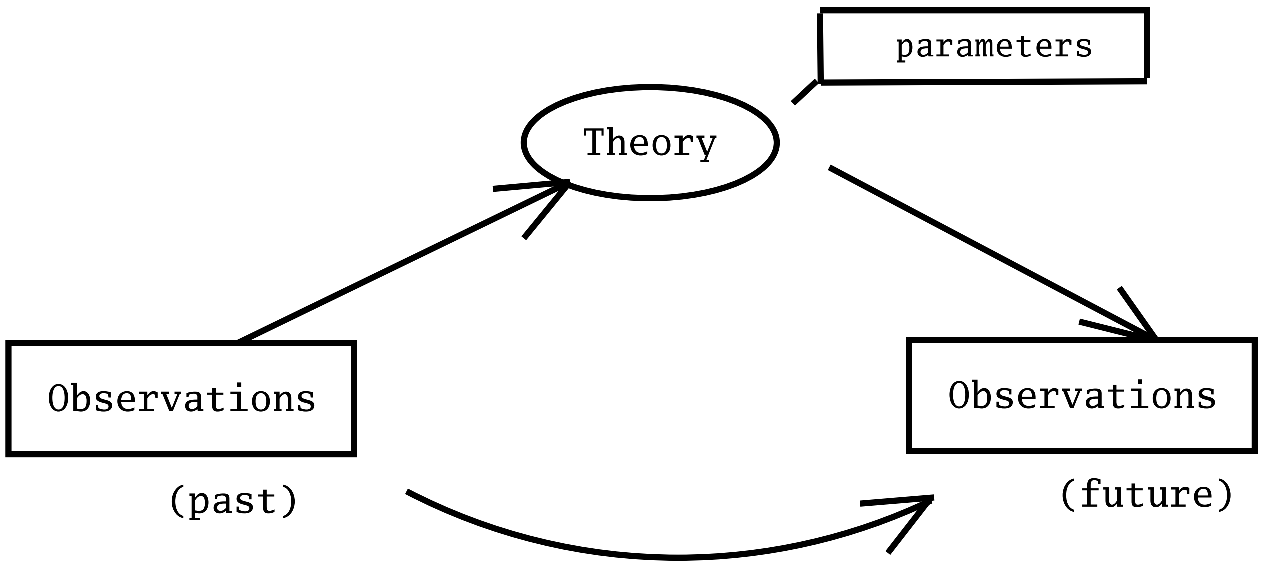
From past to future



Task of physicists:

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

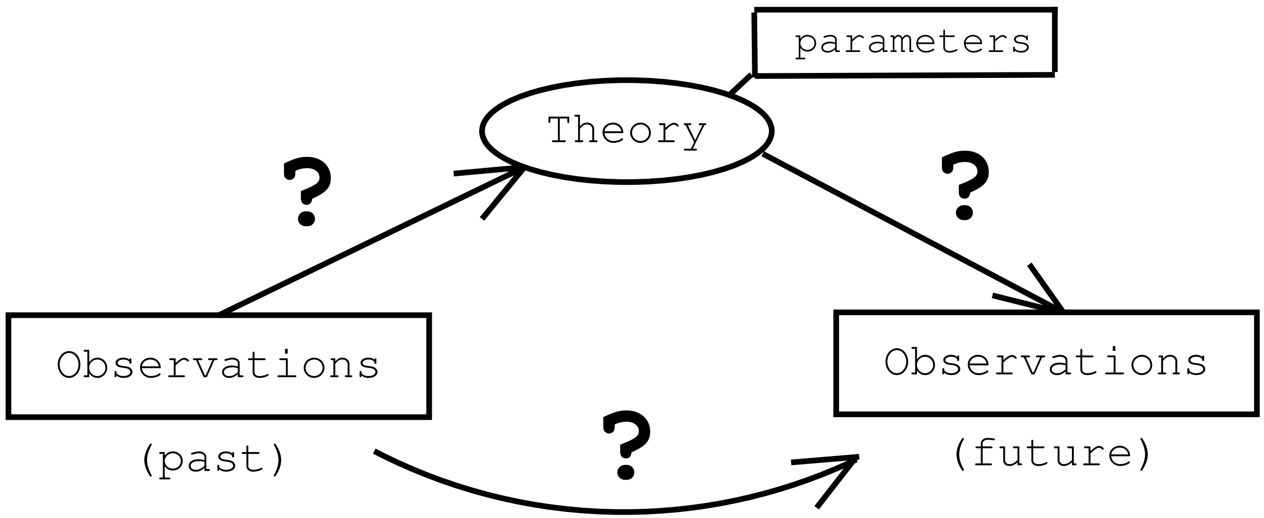
From past to future



Process

- neither automatic
- nor purely contemplative
  - → 'scientific method'
  - → planned experiments ('actions') ⇒ decision.

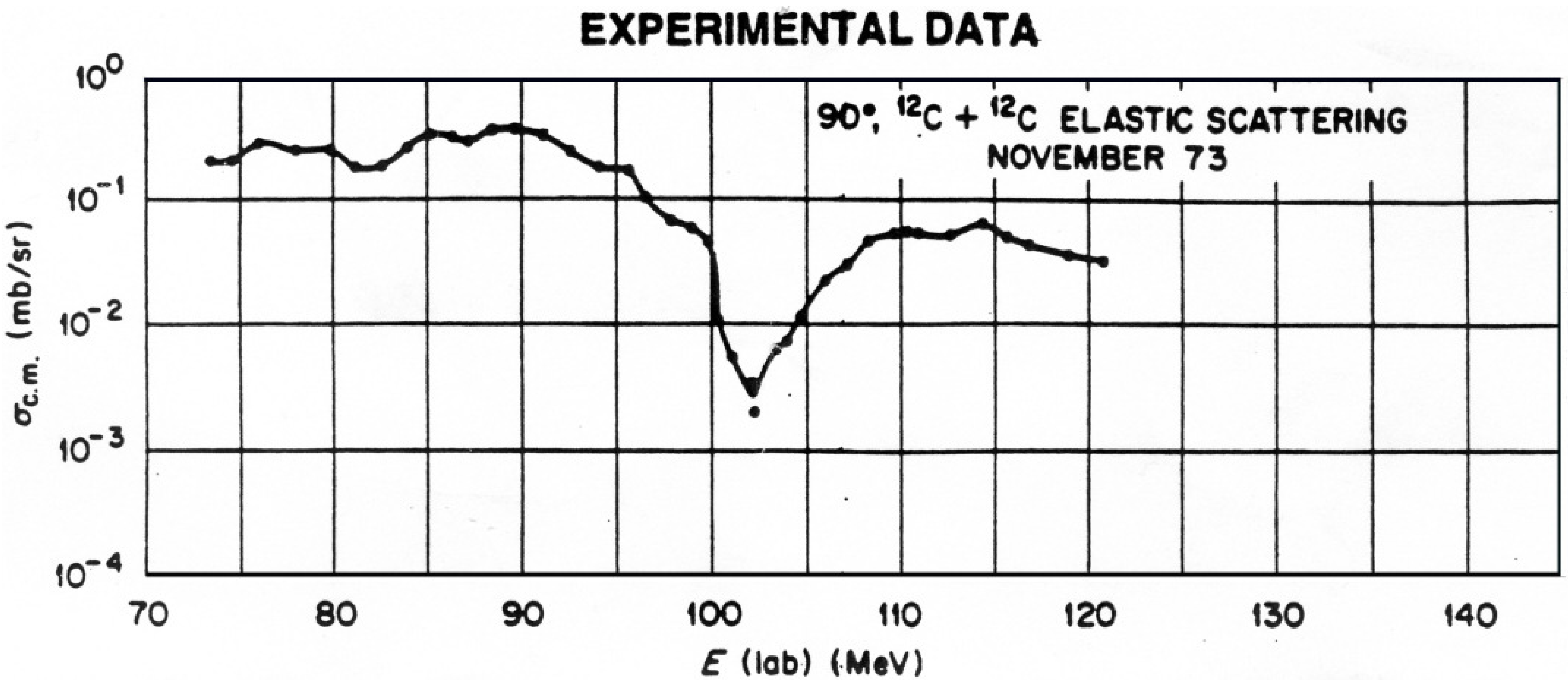
From past to future



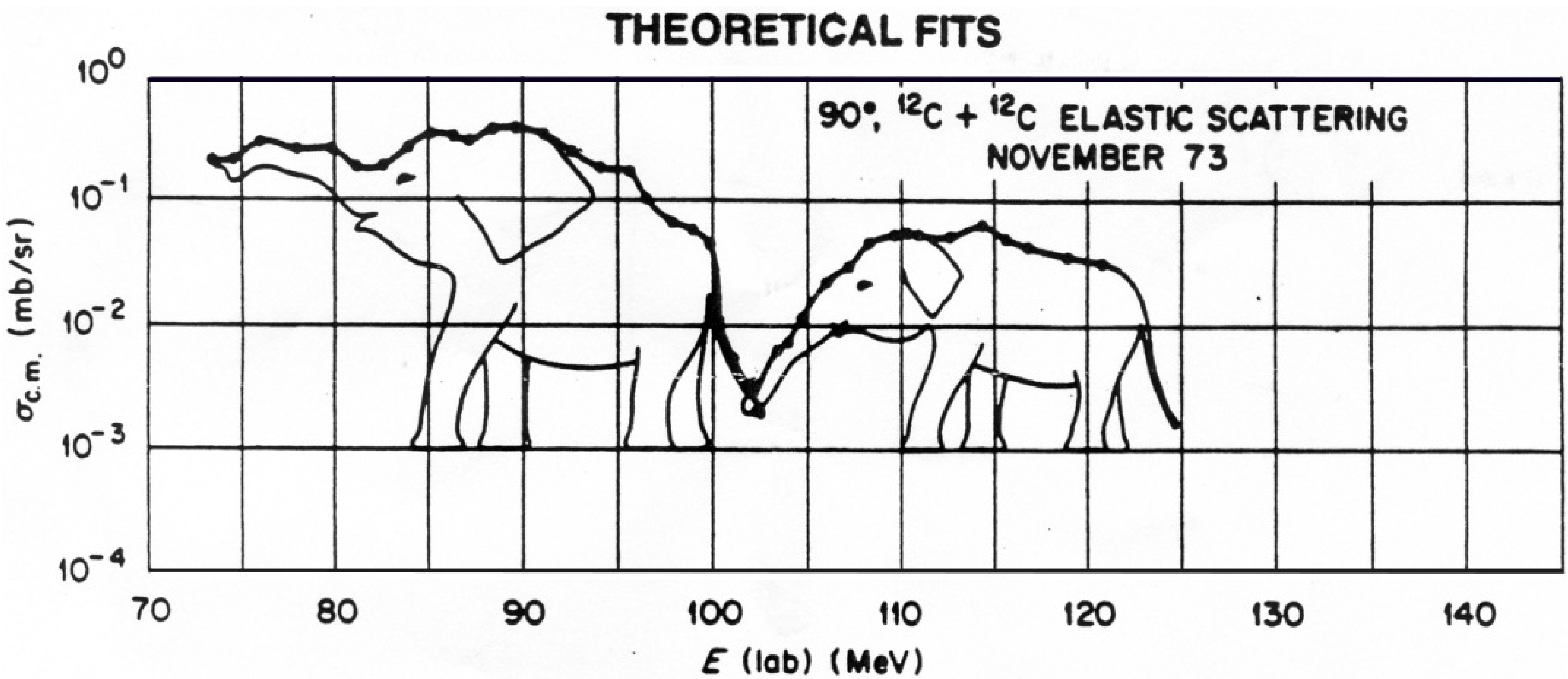
⇒ Uncertainty:

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

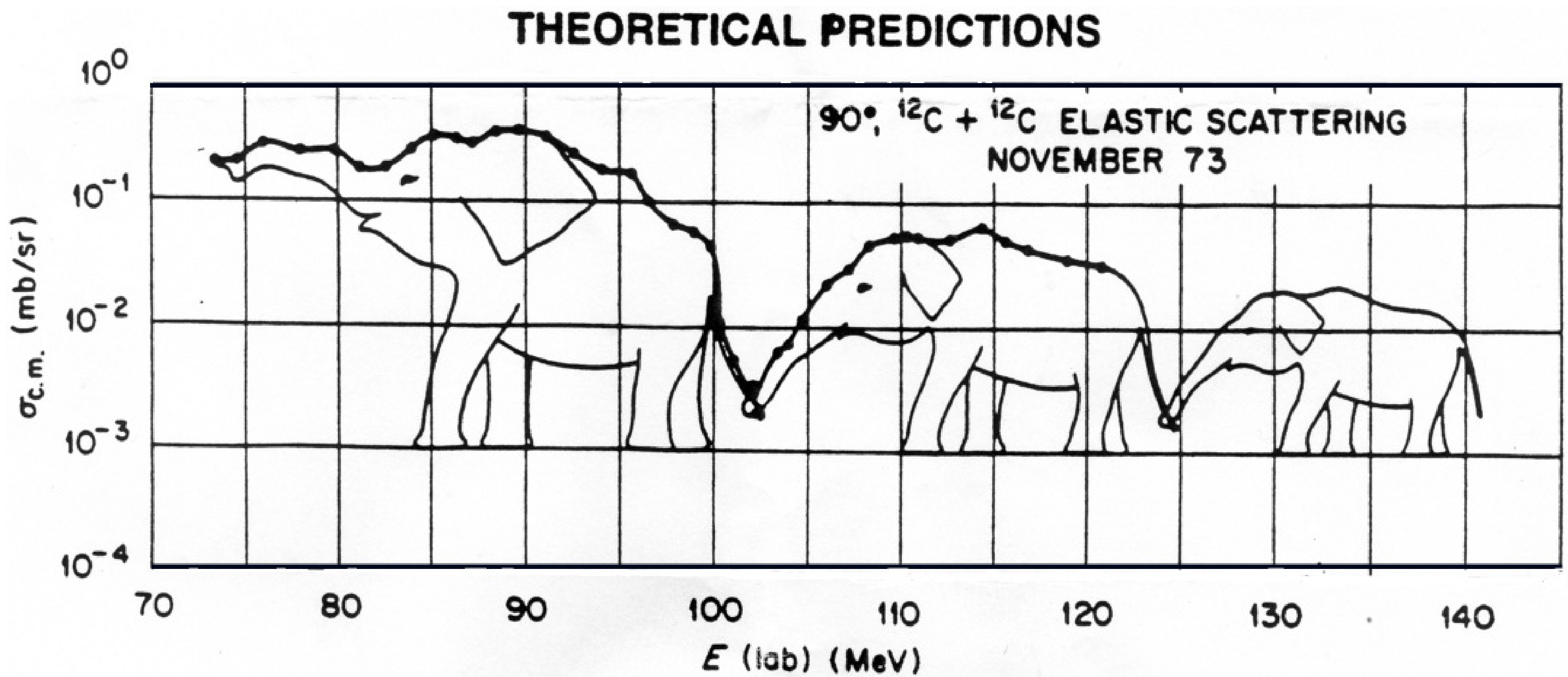
Inferential-predictive process



Inferential-predictive process

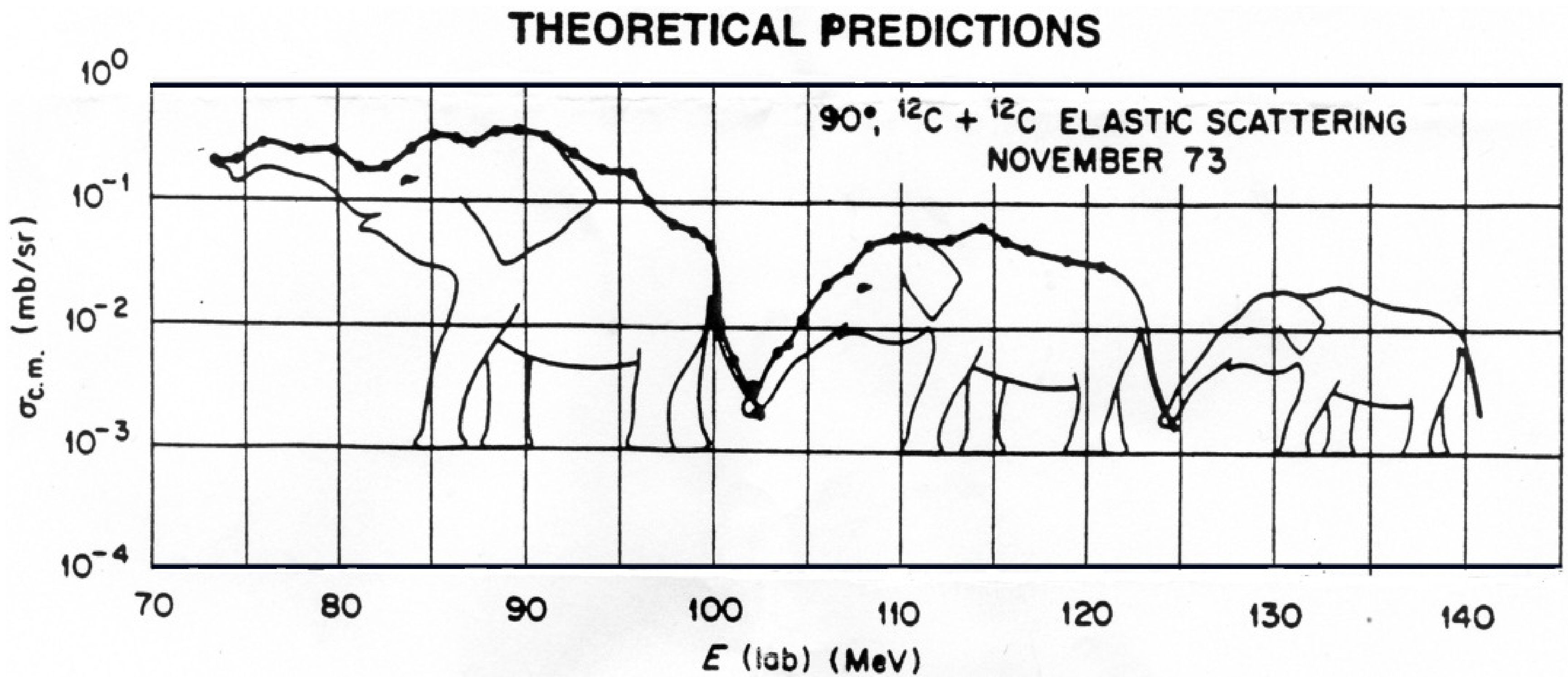


Inferential-predictive process



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

Inferential-predictive process



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

Even if the (*ad hoc*) model fits perfectly the data,
we do not believe the predictions
because we don't trust the model!

[Many 'good' models are *ad hoc* models!]

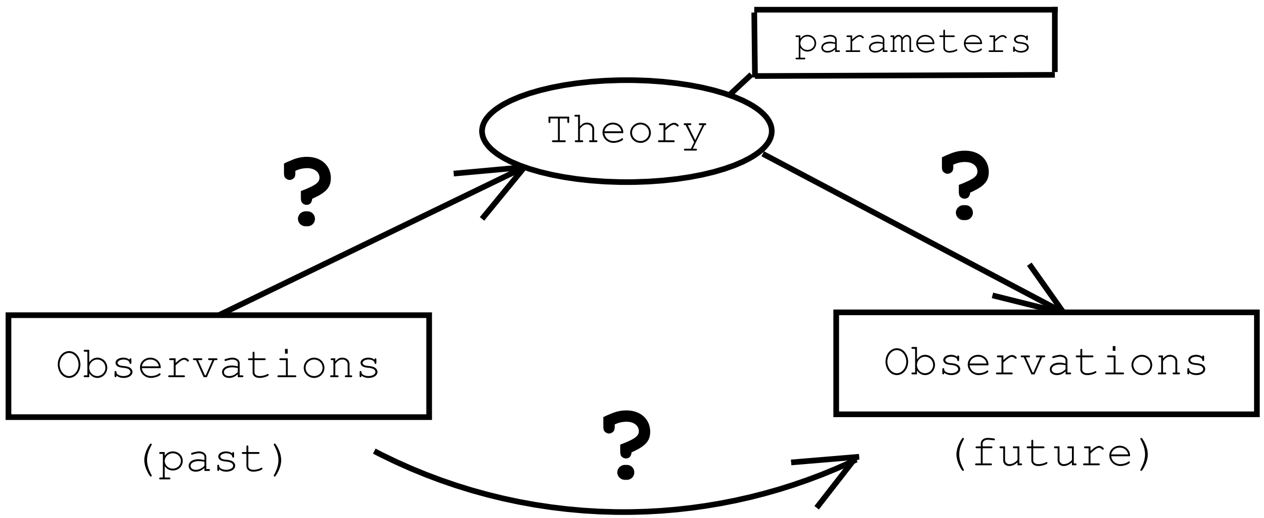
2011 IgNobel prize in Mathematics

- D. Martin of USA (who predicted the world would end in 1954)
- P. Robertson of USA (who predicted the world would end in 1982)
- E. Clare Prophet of the USA (who predicted the world would end in 1990)
- L.J. Rim of KOREA (who predicted the world would end in 1992)
- C. Mwerinde of UGANDA (who predicted the world would end in 1999)
- H. Camping of the USA (who predicted the world would end on September 6, 1994 and later predicted that the world will end on October 21, 2011)

2011 IgNobel prize in Mathematics

"For teaching the world to be careful when making mathematical assumptions and calculations"

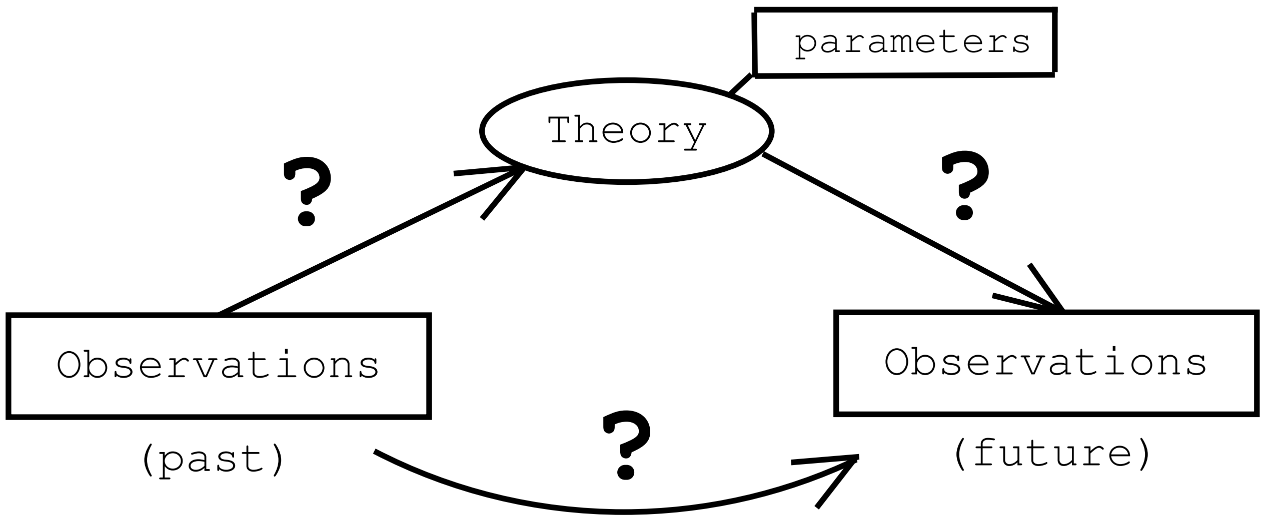
Deep source of uncertainty



Uncertainty:

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

Deep source of uncertainty



Uncertainty:

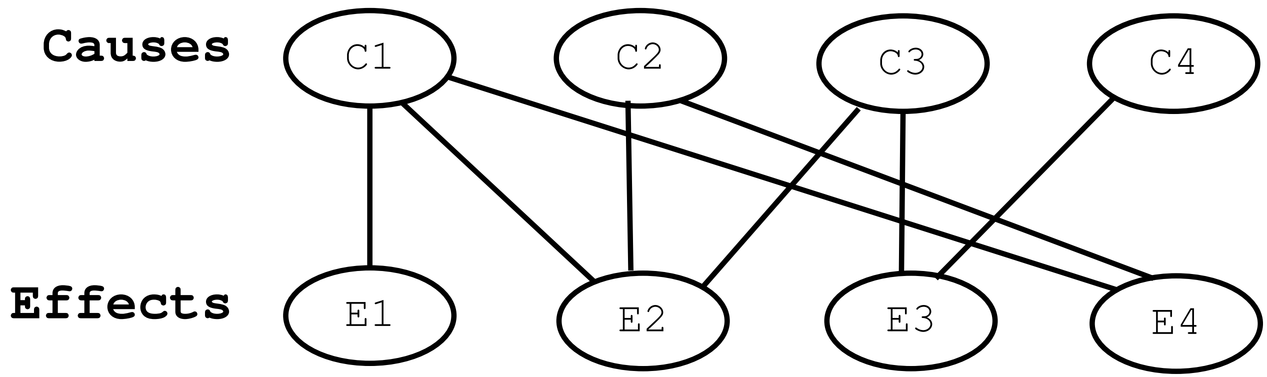
Theory	— ? $\longrightarrow$	Future observations
Past observations	— ? $\longrightarrow$	Theory
Theory	— ? $\longrightarrow$	Future observations

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

CAUSE $\Longleftrightarrow$ **EFFECT**

Causes $\rightarrow$ effects

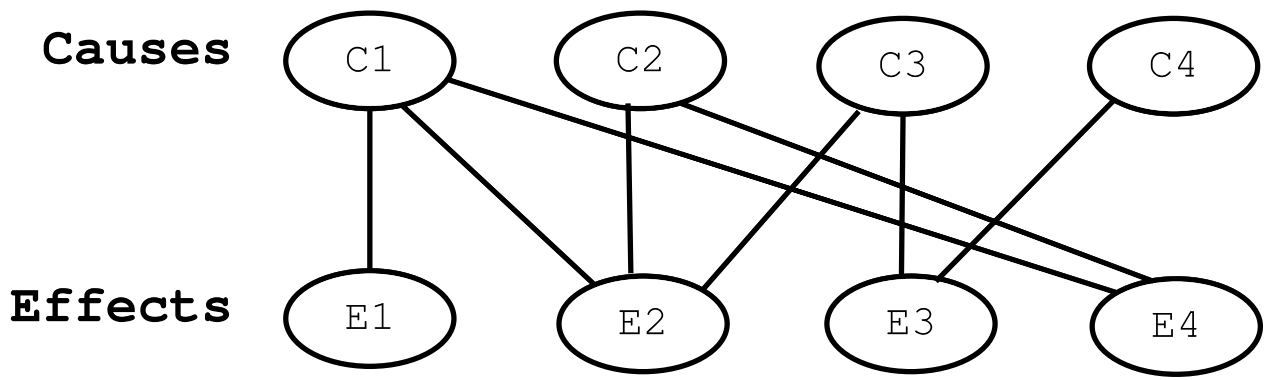
The same *apparent* cause might produce several,different effects



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

Causes → effects

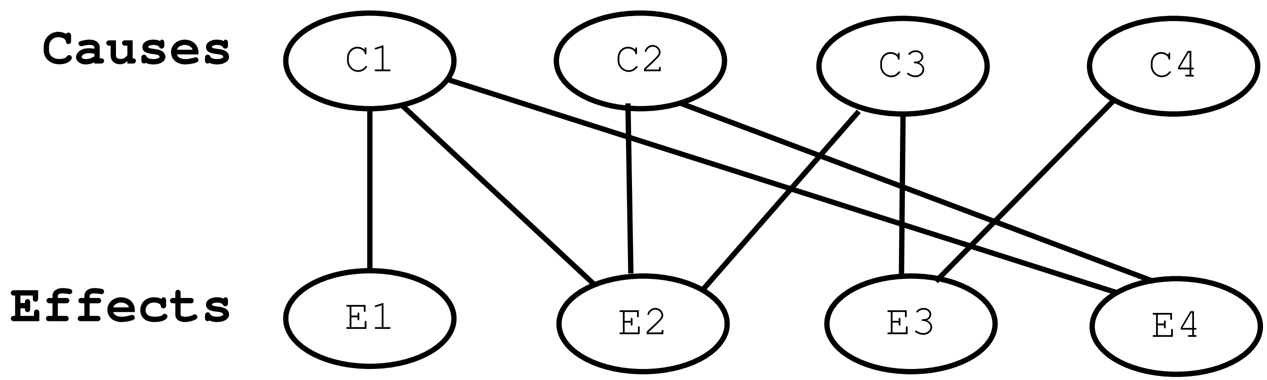
The same *apparent* cause might produce several,different effects



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

Causes $\rightarrow$ effects

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



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

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

The "essential problem" of the Sciences

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

The "essential problem" of the Sciences

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

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

(H. Poincaré – *Science and Hypothesis*)

The "essential problem" of the Sciences

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

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

(H. Poincaré – *Science and Hypothesis*)

Why physics students are not taught how to tackle this kind of problems?

Uncertainty and probability

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

- $P(-10 < \epsilon'/\epsilon \times 10^4 < 50) >> P(\epsilon'/\epsilon \times 10^4 > 100)$
- $P(172 \le m_{top}/\text{GeV} \le 174) \approx 70\%$
- $P(M_H < 125.5\,\text{GeV}) > P(M_H > 125.5\,\text{GeV})$

Uncertainty and probability

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

- $P(-10 < \epsilon'/\epsilon \times 10^4 < 50) >> P(\epsilon'/\epsilon \times 10^4 > 100)$
- $P(172 \le m_{top}/\text{GeV} \le 174) \approx 70\%$
- $P(M_H < 125.5\,\text{GeV}) > P(M_H > 125.5\,\text{GeV})$

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

Uncertainty and probability

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

- $P(-10 < \epsilon'/\epsilon \times 10^4 < 50) >> P(\epsilon'/\epsilon \times 10^4 > 100)$
- $P(172 \le m_{top}/\text{GeV} \le 174) \approx 70\%$
- $P(M_H < 125.5\,\text{GeV}) > P(M_H > 125.5\,\text{GeV})$

…although, such statements are considered
blaspheme to statistics gurus

[The fact that for several (most?) people in this audience this criticism is misterious is a clear indication of the confusion concerning this matter]

Doing Science in conditions of uncertainty

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

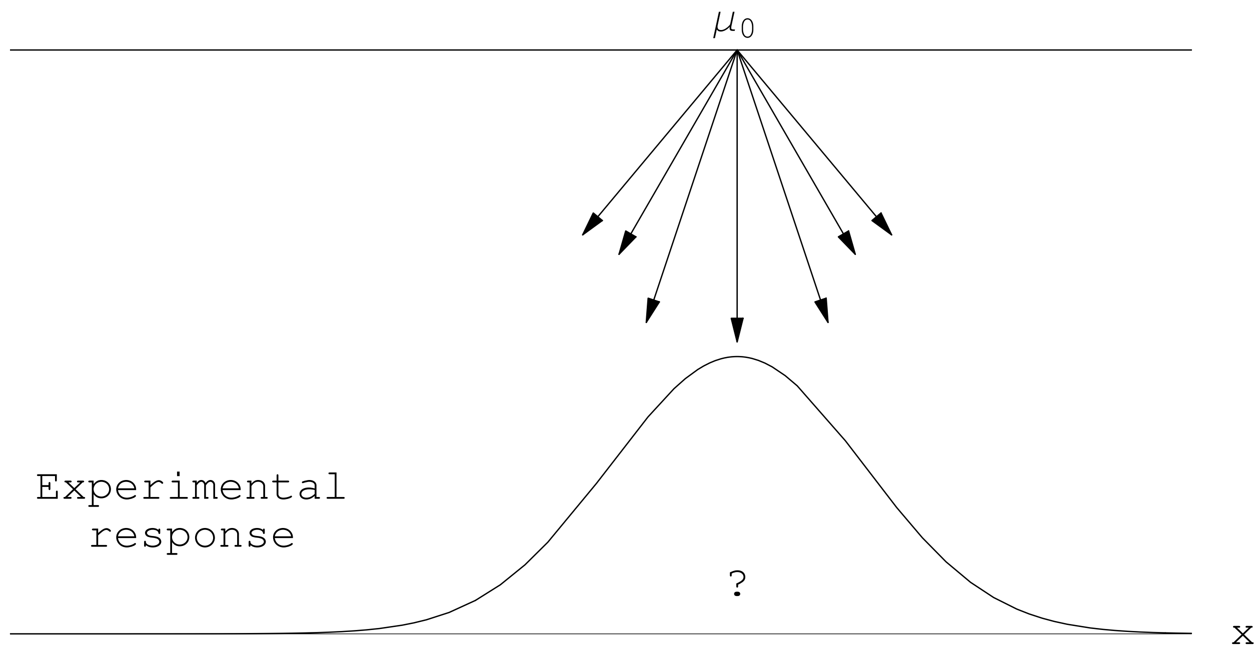
Doing Science in conditions of uncertainty

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

Indeed

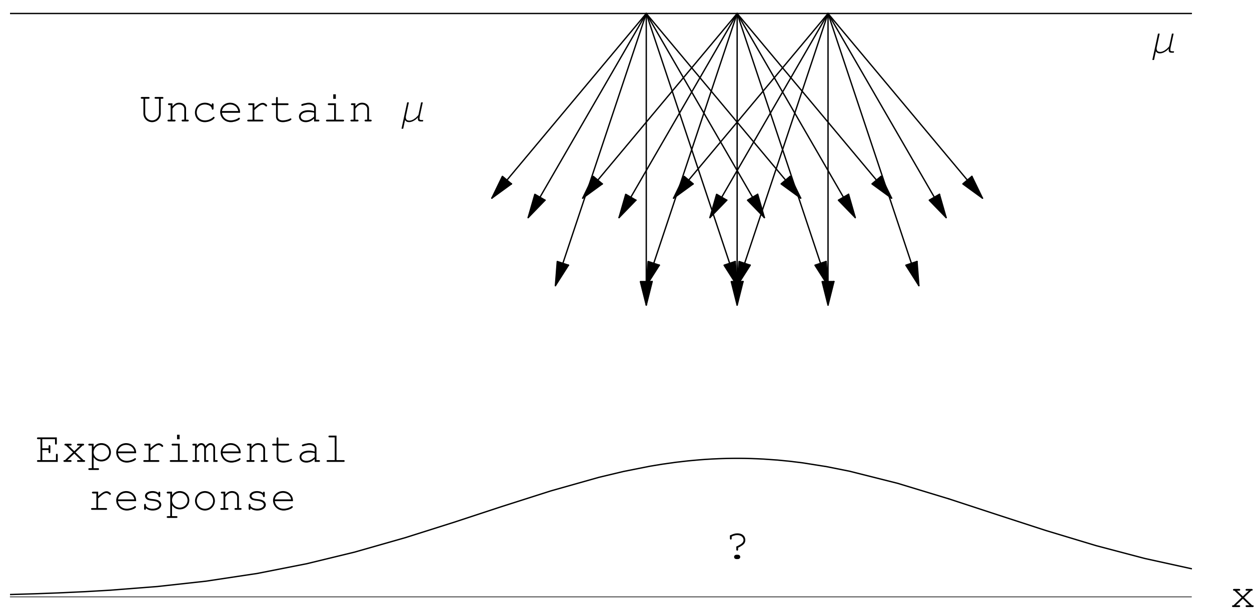
"It is scientific only to say what is more likely and what is less likely" (Feynman)

From ‘true value’ to observations



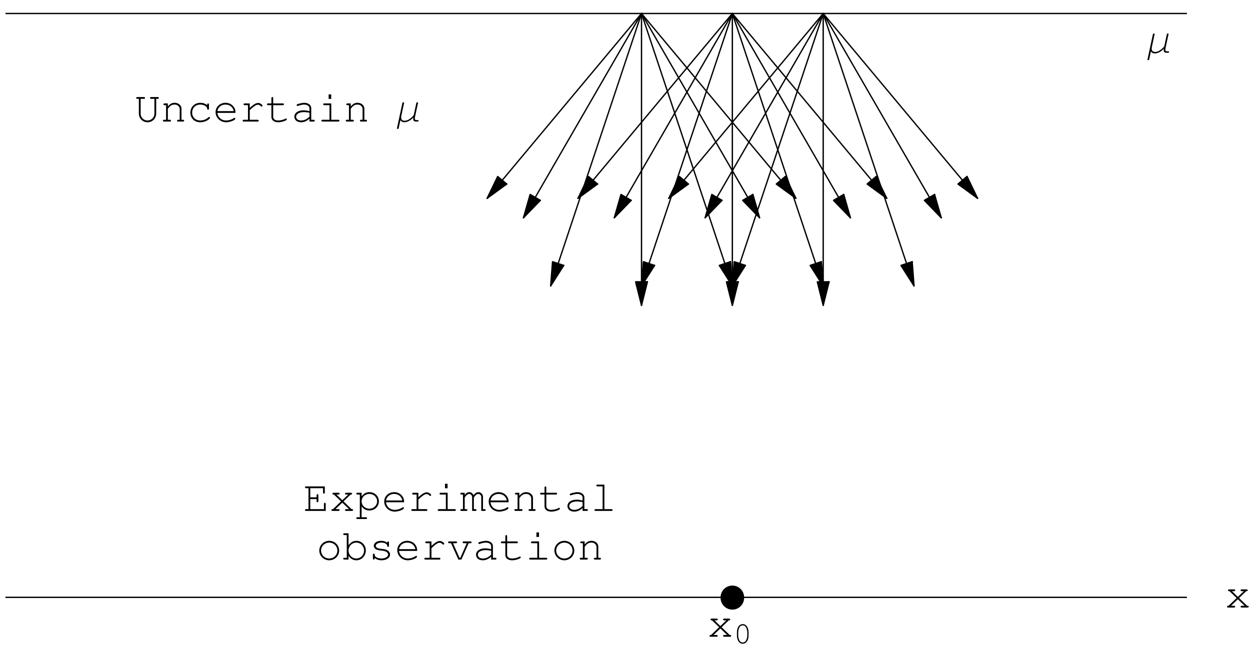
Given μ (exactly known) we are uncertain about x

From ‘true value’ to observations



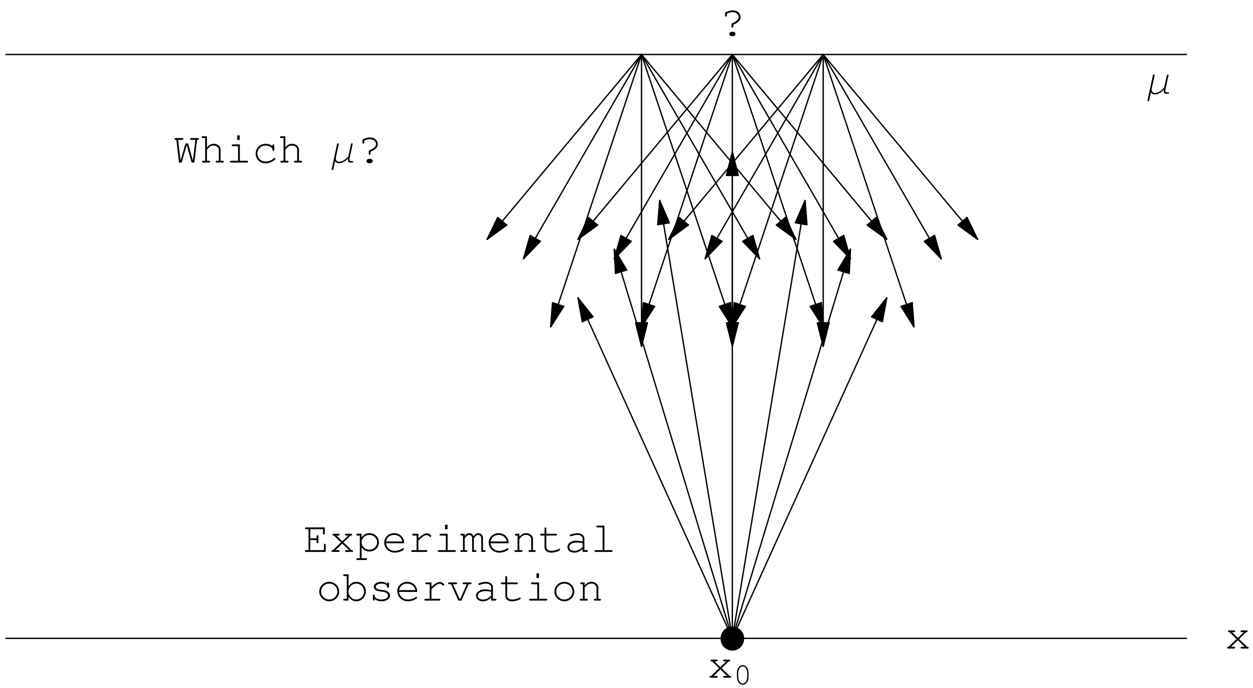
Uncertainty about μ makes us more uncertain about x

…and back: Inferring a true value



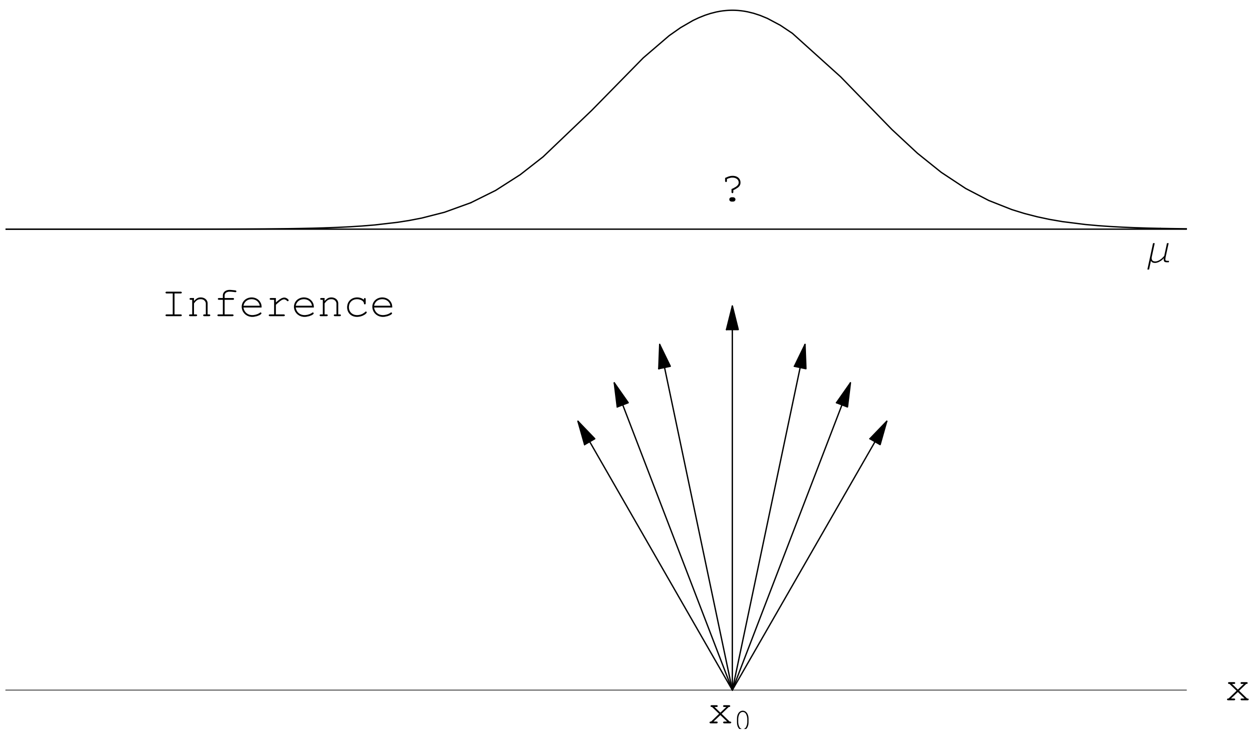
The observed data is certain: $\rightarrow$ 'true value' uncertain.

. . . and back: Inferring a true value



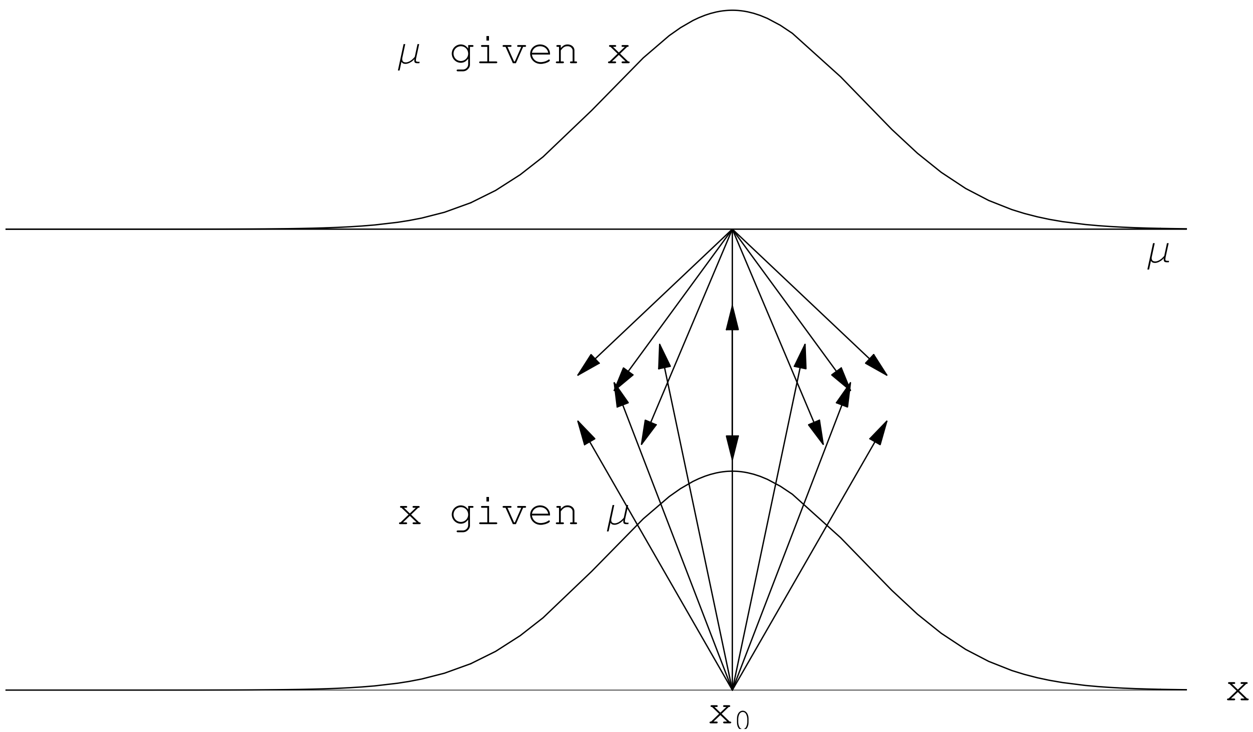
Where does the observed value of x comes from?

. . . and back: Inferring a true value



We are now uncertain about μ, given x.

…and back: Inferring a true value



Note the symmetry in reasoning.

A very simple experiment

Let’s make an experiment

A very simple experiment

Let's make an experiment

- Here
- Now

A very simple experiment

Let's make an experiment

- Here
- Now

For simplicity

- μ can assume only six possibilities:

$$\mathbf{0, 1, \ldots, 5}$$

- x is binary:

$$\mathbf{0, 1}$$

$$[\,(1,2);\ \text{Black/White};\ \text{Yes/Not};\ \ldots\,]$$

A very simple experiment

Let's make an experiment

- Here
- Now

For simplicity

- μ can assume only six possibilities:

$$\mathbf{0, 1, \ldots, 5}$$

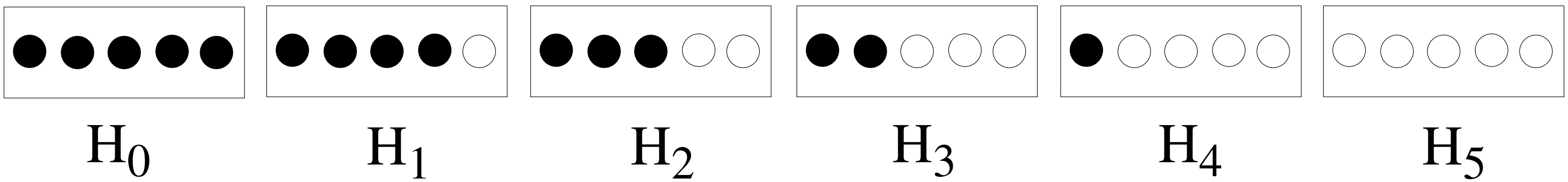
- x is binary:

$$\mathbf{0, 1}$$

$$[\,(1,2);\ \text{Black/White};\ \text{Yes/Not};\ \ldots\,]$$

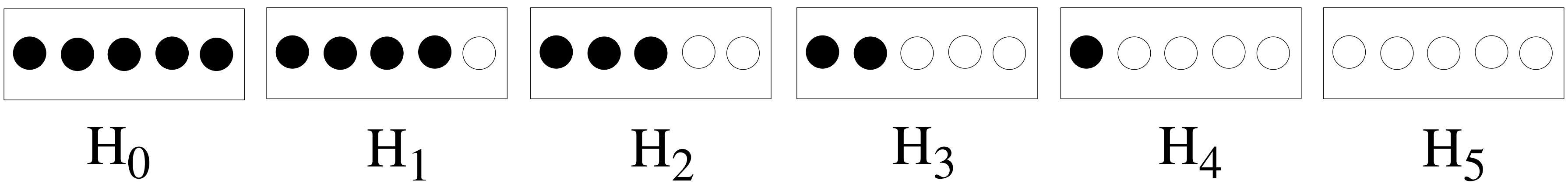
$\Rightarrow$ Later we shall make μ continous.

Which box? Which ball?



Let us take randomly one of the boxes.

Which box? Which ball?



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:

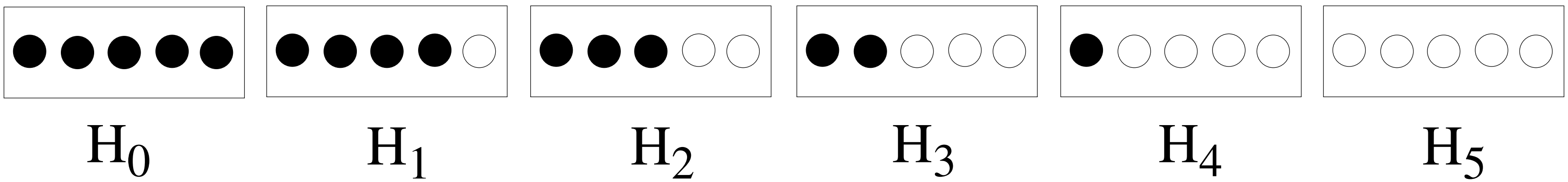
(a) Which box have we chosen, $H_0, H_1, \ldots, H_5$?

(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 certainties:

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

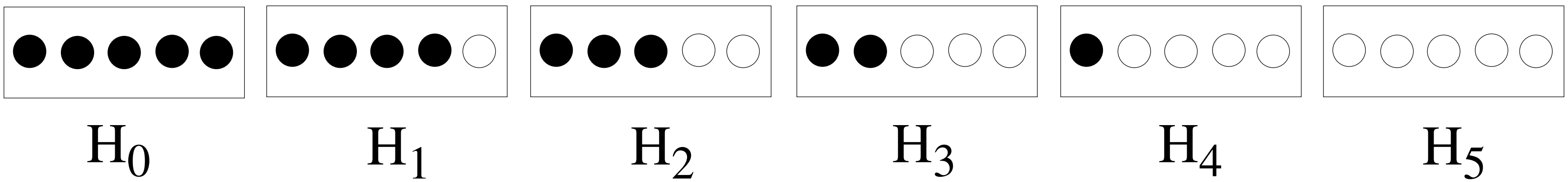
Which box? Which ball?



Let us take randomly one of the boxes.

- What happens after we have extracted one ball and looked its color?
  - Intuitively feel *how to* roughly *change* our opinion about
    - the possible cause
    - a future observation

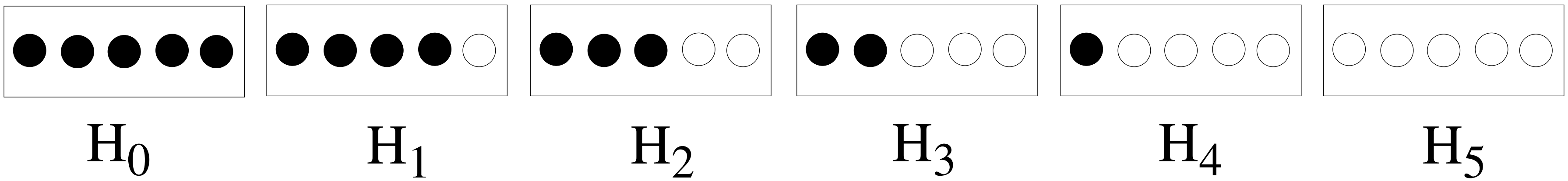
Which box? Which ball?



Let us take randomly one of the boxes.

- What happens after we have extracted one ball and looked its color?
  - Intuitively feel *how to* roughly *change* our opinion about
    - the possible cause
    - a future observation
  - Can we do it *quantitatively*, in an 'objective way'?

Which box? Which ball?



Let us take randomly one of the boxes.

- What happens after we have extracted one ball and looked its color?
  - Intuitively feel *how to* roughly *change* our opinion about
    - the possible cause
    - a future observation
  - 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

$\Rightarrow$ 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 an electron and read its properties, unlike we read the MAC address of a PC interface.)

***Where is* probability?**

We all agree that the experimental results change

- the probabilities of the box compositions;
- the probabilities of a future outcomes,

Where is probability?

We all agree that the experimental results change

- the probabilities of the box compositions;
- the probabilities of a future outcomes,

although the box composition remains unchanged
('extractions followed by reintroduction').

Where is probability?

We all agree that the experimental results change

- the probabilities of the box compositions;
- the probabilities of a future outcomes,

although the box composition remains unchanged ('extractions followed by reintroduction').

Where is the probability?

Where is probability?

We all agree that the experimental results change

- the probabilities of the box compositions;
- the probabilities of a future outcomes,

although the box composition remains unchanged
('extractions followed by reintroduction').

Where is the probability?

Certainly <u>not</u> *in* the box!

Subjective nature of probability

“Since the knowledge may be different with different persons

Subjective nature of probability

“Since the knowledge may be different with different persons or with the same person at different times,

Subjective nature of probability

"Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence,

Subjective nature of probability

"Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence, and thus different numerical probabilities may be attached to the same event"

Subjective nature of probability

"Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence, and thus different numerical probabilities may be attached to the same event"

(Schrödinger, 1947)

Subjective nature of probability

"Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence, and thus different numerical probabilities may be attached to the same event"

(Schrödinger, 1947)

Probability depends on the status of information of the *subject* who evaluates it.

Probability is <u>always</u> conditional probability

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

Probability is <u>always</u> conditional probability

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

(Schrödinger, 1947)

Probability is always conditional probability

“Thus whenever we speak loosely of ‘the probability of an event’, it is always to be understood: probability with regard to a certain given state of knowledge”

(Schrödinger, 1947)

$$P(E) \longrightarrow P(E\,|\,I_s)$$

where I_s is the information available to *subject* s.

Probability is always conditional probability

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

(Schrödinger, 1947)

$$P(E) \longrightarrow P(E\,|\,I_s)$$

where I_s is the information available to *subject* s.

$\longrightarrow$ Three boxes TV contests

What are we talking about?

“Given the state of our knowledge about everything that could possible have any bearing on the coming true. . .

What are we talking about?

"Given the state of our knowledge about everything that could possible have any bearing on the coming true. . . 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"

(Schrödinger, 1947)

What are we talking about?

"Given the state of our knowledge about everything that could possible have any bearing on the coming true. . . 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"

$\Rightarrow$ How much we believe something

What are we talking about?

"Given the state of our knowledge about everything that could possible have any bearing on the coming true... 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"

$\rightarrow$ 'Degree of belief' $\leftarrow$

Beliefs and 'coherent' bets

Remarks:

- Subjective does not mean arbitrary!

Beliefs and ‘coherent’ bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?

Beliefs and 'coherent' bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?

  "The usual touchstone, whether that which someone asserts is merely his persuasion – or at least his subjective conviction, that is, his firm belief – is betting. It often happens that someone propounds his views with such positive and uncompromising assurance that he seems to have entirely set aside all thought of possible error. A bet disconcerts him. Sometimes it turns out that he has a conviction which can be estimated at a value of one ducat, but not of ten. For he is very willing to venture one ducat, but when it is a question of ten he becomes aware, as he had not previously been, that it may very well be that he is in error." (Kant)

Beliefs and ‘coherent’ bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?

		1	X	2
17/07 19:00	SL. LIBEREC - S KARAGANDY	1,20	6,00	10,00
17/07 18:00	HJK HELSINK - KR REYKJAV	1,30	4,75	8,00
17/07 17:45	FLORA TALLINN - BASILEA	9,50	5,00	1,25
17/07 18:30	DUDELANGE - SALISBURGO	10,00	5,00	1,25
17/07 20:45	VALLETTA - PARTIZAN	10,00	5,00	1,25
18/07 18:00	BATE BORISOV - VARDAR	1,15	6,50	13,00

Beliefs and ‘coherent’ bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet:
  - you state the odds according on your beliefs;
  - somebody else will choose the direction of the bet.

Beliefs and 'coherent' bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet:
  - you state the odds according on your beliefs;
  - somebody else will choose the direction of the bet.

"His [Bouvard] calculations give him the mass of Saturn as 3,512th part of that of the sun. Applying my probabilistic formulae to these observations, I find that the odds are 11,000 to 1 that the error in this result is not a hundredth of its value." (Laplace)

Beliefs and 'coherent' bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet:
  - you state the odds according on your beliefs;
  - somebody else will choose the direction of the bet.

"His [Bouvard] calculations give him the mass of Saturn as 3,512th part of that of the sun. Applying my probabilistic formulae to these observations, I find that the odds are 11,000 to 1 that the error in this result is not a hundredth of its value." (Laplace)

Beliefs and ‘coherent’ bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet:
  - you state the odds according on your beliefs;
  - somebody else will choose the direction of the bet.

“His [Bouvard] calculations give him the mass of Saturn as 3,512th part of that of the sun. Applying my probabilistic formulae to these observations, I find that the odds are 11,000 to 1 that the error in this result is not a hundredth of its value.” (Laplace)

$\rightarrow P(3477 \leq M_{Sun}/M_{Sat} \leq 3547 \,|\, I(\text{Laplace})) = 99.99\%$

‘C.L.’ Vs Degree of Confidence

Is a ‘conventional’ 95% C.L. lower/upper bound a 19 to 1 bet?

‘C.L.’ Vs Degree of Confidence

Is a ‘conventional’ 95% C.L. lower/upper bound a 19 to 1 bet?

NO!

‘C.L.’ Vs Degree of Confidence

Is a ‘conventional’ 95% C.L. lower/upper bound a 19 to 1 bet?

NO!

- It does not imply one has to be 95% confident on something!
- If you do so you are going to make a bad bet!

‘C.L.’ Vs Degree of Confidence

Is a ‘conventional’ 95% C.L. lower/upper bound a 19 to 1 bet?

NO!

- It does not imply one has to be 95% confident on something!
- If you do so you are going to make a bad bet!

For more on the subject:

http://arxiv.org/abs/1112.3620

http://www.roma1.infn.it/~dagos/badmath/#added

'C.L.' Vs Degree of Confidence

Is a 'conventional' 95% C.L. lower/upper bound a 19 to 1 bet?

NO!

- It does not imply one has to be 95% confident on something!
- If you do so you are going to make a bad bet!

For more on the subject:
http://arxiv.org/abs/1112.3620
http://www.roma1.infn.it/~dagos/badmath/#added

Hint: $P(\theta \leq \theta_{obs} \,|\, m_0) \neq P(m \geq m_o \,|\, \theta_{obs})$!!

$\Rightarrow$ more in second lecture.

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

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

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

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

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

Standard textbook definitions

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

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

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

Future $\Leftrightarrow$ Past (belief!)

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

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

$\to$ It limits the range of applications

‘Definitions’ $\rightarrow$ evaluation rules

Very useful evaluation rules

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

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

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

‘Definitions’ $\rightarrow$ evaluation rules

Very useful evaluation rules

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

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

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

BUT they cannot define the concept of probability!

‘Definitions’ $\rightarrow$ evaluation rules

Very useful evaluation rules

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

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

In the probabilistic approach we are following

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

‘Definitions’ $\rightarrow$ evaluation rules

Very useful evaluation rules

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

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

In the probabilistic approach we are following

- Rule A is recovered immediately (under the assumption of equiprobability, when it applies).
- Rule B results from a theorem (under well defined assumptions): $\Rightarrow$ Laplace’s rule of succession

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.
  - $P(\text{rain next Saturday in Vienna}) = 68\%$
  - $P(\text{Usain Bolt will win the 100m in London}) = 68\%$
  - $P(M_H \leq 125.5\,\text{GeV}) = 68\%$
  - $P(\text{free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$

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.
  - $P($rain next Saturday in Vienna$) = 68\%$
  - $P($Usain Bolt will win the 100m in London$) = 68\%$
  - $P(M_H \leq 125.5\,\text{GeV}) = 68\%$
  - $P($free neutron decays before 17 s$) = 68\%$
  - $P($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.
  - $P(\text{rain next Saturday in Vienna}) = 68\%$
  - $P(\text{Usain Bolt will win the 100m in London}) = 68\%$
  - $P(M_H \le 125.5\,\text{GeV}) = 68\%$
  - $P(\text{free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$
- You might agree or disagree, but at least You know what this person has in his mind. (NOT TRUE with "C.L.'s"!)

Unifying role of subjective probability

- Wide range of applicability
- Probability statements all have the same meaning no matter to what they refer and how the number has been evaluated.
  - P(rain next Saturday in Vienna) $= 68\%$
  - P(Usain Bolt will win the 100m in London) $= 68\%$
  - $P(M_H \leq 125.5\,\text{GeV}) = 68\%$
  - P(free neutron decays before 17 s) $= 68\%$
  - P(White ball from a box with 68W+32B) $= 68\%$
- 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 is indifferent to the choice.

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.
  - $P(\text{rain next Saturday in Vienna}) = 68\%$
  - $P(\text{Usain Bolt will win the 100m in London}) = 68\%$
  - $P(M_H \leq 125.5\,\text{GeV}) = 68\%$
  - $P(\text{free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$

We can talk very naturally about **probabilities of true values!**

Probability Vs “probability”…

Errors on ratios of small numbers of events
F. James$^{(*)}$ and M. Roos
Nucl. Phys. **B172** (1980) 475
(`http://ccdb4fs.kek.jp/cgi-bin/img_index?8101205`)

When the result of the measurement of a physical quantity is published as $R=R_0\pm\sigma_0$ without further explanation, it is implied that R is a Gaussian-distributed measurement with mean R_0 and variance σ_0^2. This allows one to calculate various confidence intervals of given "probability", i.e. the "probability" P that the true value of R is within a given interval. P is given by the area under the corresponding part of the Gaussian curve, and is the basis of well-known rules-of-thumb such as "the probability of exceeding two standard deviations is 5%".

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

Mathematics of beliefs

The good news:

The basic laws of degrees of belief are the same we get from the inventory of favorable and possible cases, or from events occurred in the past.

[Details skipped. . .]

Basic rules of probability

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

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

Remember that probability is always conditional probability!

I is the background condition (related to information 'I'_s')

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

Basic rules of probability

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

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

Remember that probability is always conditional probability!

I is the background condition (related to information 'I'_s')

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

Note: 4. does not define conditional probability.
(Probability is always conditional probability!)

Mathematics of beliefs

An even better news:

The fourth basic rule
can be fully exploided!

Mathematics of beliefs

An even better news:

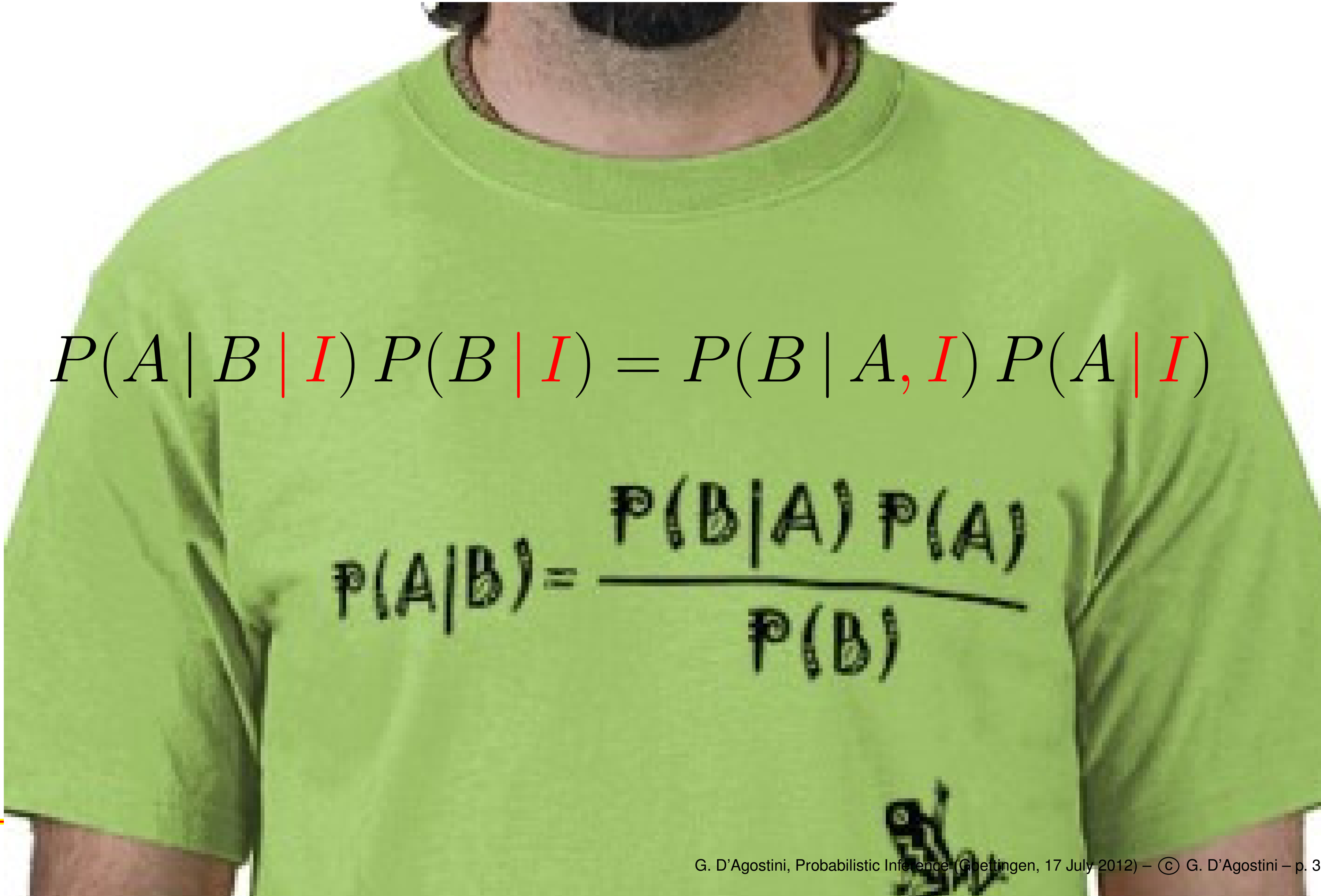
The fourth basic rule
can be fully exploided!

(Liberated by a curious ideology that forbits its use)

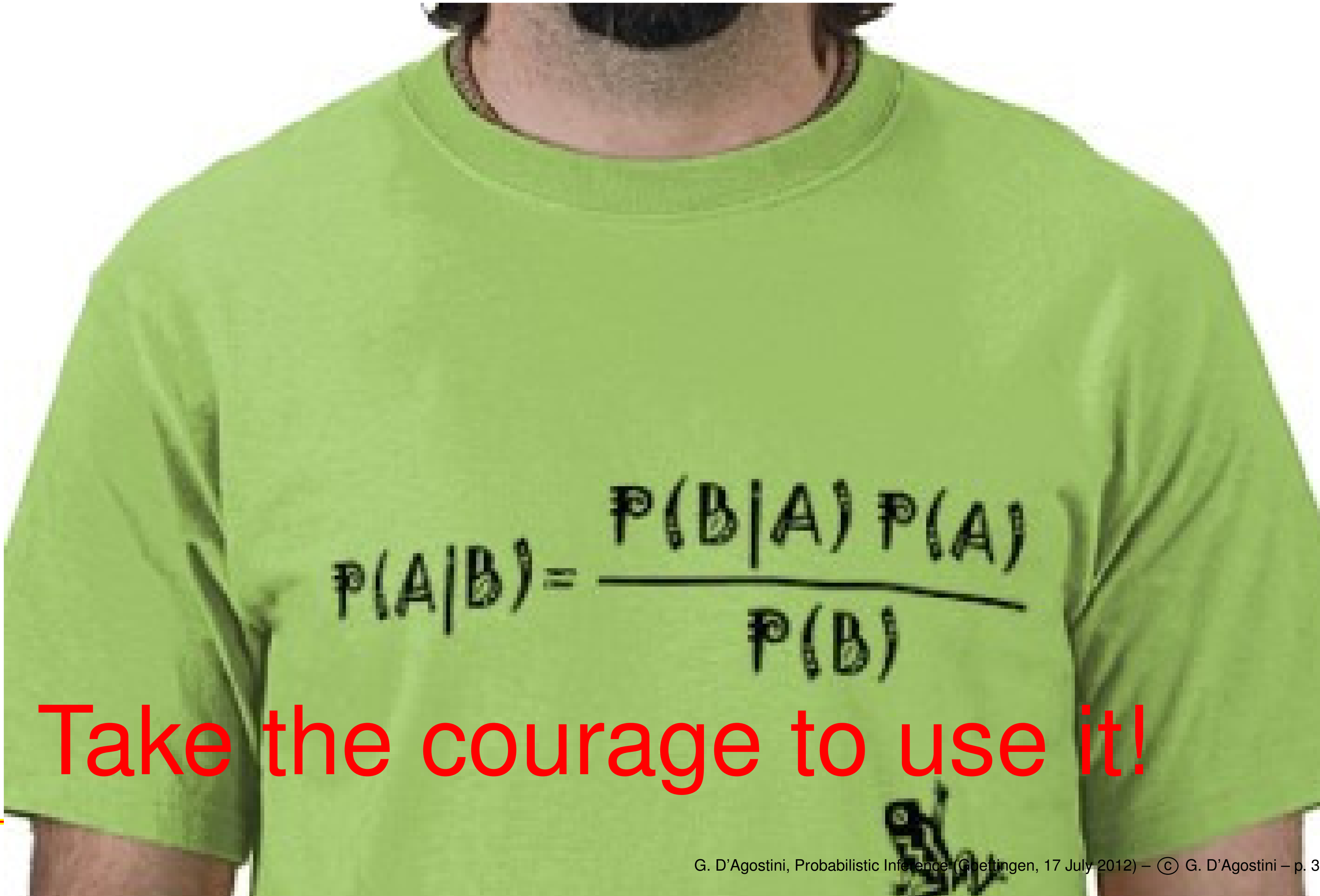
A simple, powerful formula

A simple, powerful formula

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

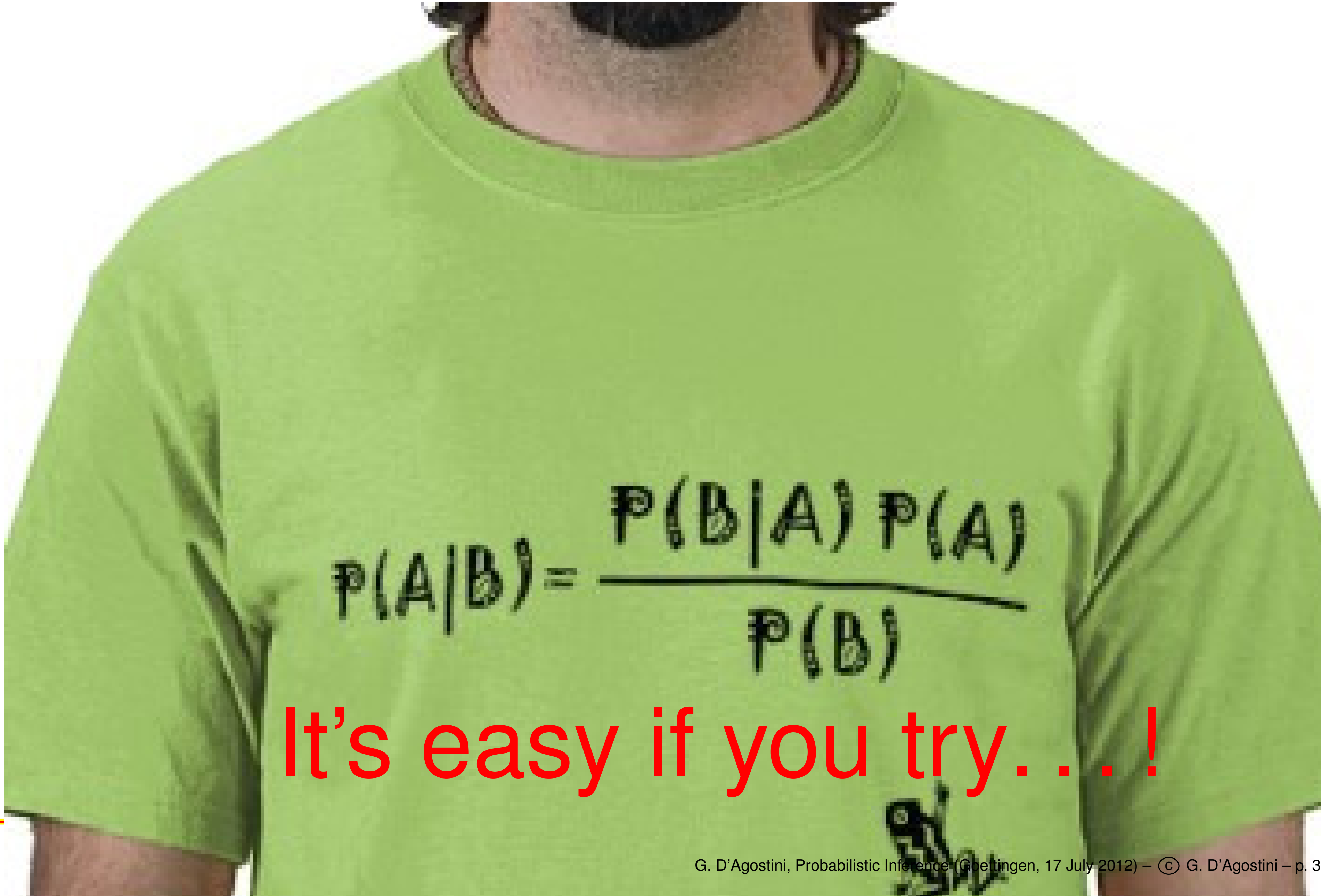


A simple, powerful formula



Take the courage to use it!

A simple, powerful formula



It's easy if you try. . . !

Laplace's "Bayes Theorem"

"The greater the probability of an observed event given any one of a number of causes to which that event may be attributed, the greater the likelihood of that cause {given that event}.

$$P(C_i \,|\, E) \propto P(E \,|\, C_i)$$

Laplace's "Bayes Theorem"

"The greater the probability of an observed event given any one of a number of causes to which that event may be attributed, the greater the likelihood of that cause {given that event}. The probability of the existence of any one of these causes {given the event} is thus a fraction whose numerator is the probability of the event given the cause, and whose denominator is the sum of similar probabilities, summed over all causes.

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

Laplace's "Bayes Theorem"

"The greater the probability of an observed event given any one of a number of causes to which that event may be attributed, the greater the likelihood of that cause {given that event}. The probability of the existence of any one of these causes {given the event} is thus a fraction whose numerator is the probability of the event given the cause, and whose denominator is the sum of similar probabilities, summed over all causes. If the various causes are not equally probable *a priory*, it is necessary, instead of the probability of the event given each cause, to use the product of this probability and the *possibility of the cause itself*."

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

Laplace's "Bayes Theorem"

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

"This is the fundamental principle (*) of that branch of the analysis of chance that consists of reasoning *a posteriori* from events to causes"

(*) In his "Philosophical essay" Laplace calls 'principles' the 'fondamental rules'.

Laplace's "Bayes Theorem"

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

"This is the fundamental principle (*) of that branch of the analysis of chance that consists of reasoning *a posteriori* from events to causes"

(*) In his "Philosophical essay" Laplace calls 'principles' the 'fondamental rules'.

Note: denominator is just a normalization factor.

$$\Rightarrow \qquad P(C_i \,|\, E) \propto P(E \,|\, C_i)\, P(C_i)$$

Most convenient way to remember Bayes theorem

Telling it with Gauss' words

A reference to the Princeps Mathematicorum (Prince of Mathematicians) is a must in this town and in this place.

Telling it with Gauss' words

A reference to the Princeps Mathematicorum (Prince of Mathematicians) is a must in this town and in this place.

$$P(C_i \,|\, \text{data}) = \frac{P(\text{data} \,|\, C_i)}{P(\text{data})} \, P_0(C_i)$$

Telling it with Gauss' words

A reference to the Princeps Mathematicorum (Prince of Mathematicians) is a must in this town and in this place.

$$P(C_i\,|\,\text{data}) = \frac{P(\text{data}\,|\,C_i)}{P(\text{data})}\,P_0(C_i)$$

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

(Gauss)

Telling it with Gauss' words

A reference to the Princeps Mathematicorum (Prince of Mathematicians) is a must in this town and in this place.

$$P(C_i \,|\, \text{data}) = \frac{P(\text{data} \,|\, C_i)}{P(\text{data})} P_0(C_i)$$

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

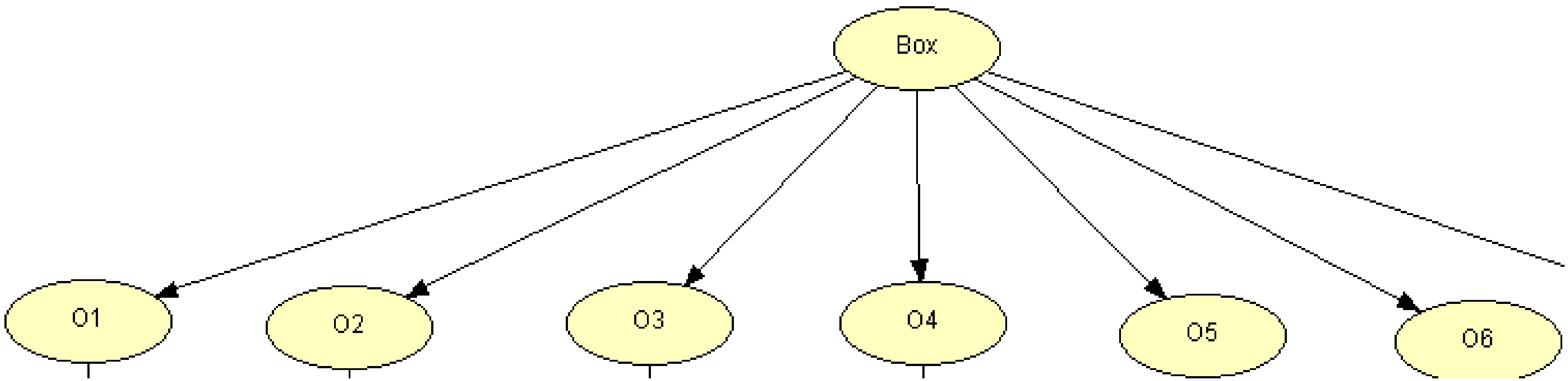
(Gauss)

Arguments used to derive Gaussian distribution

- $f(\mu \,|\, \{x\}) \propto f(\{x\} \,|\, \mu) \cdot f_0(\mu)$
- $f_0(\mu)$ 'flat' (all values a priory equally possible)
- posterior maximized at $\mu = \overline{x}$

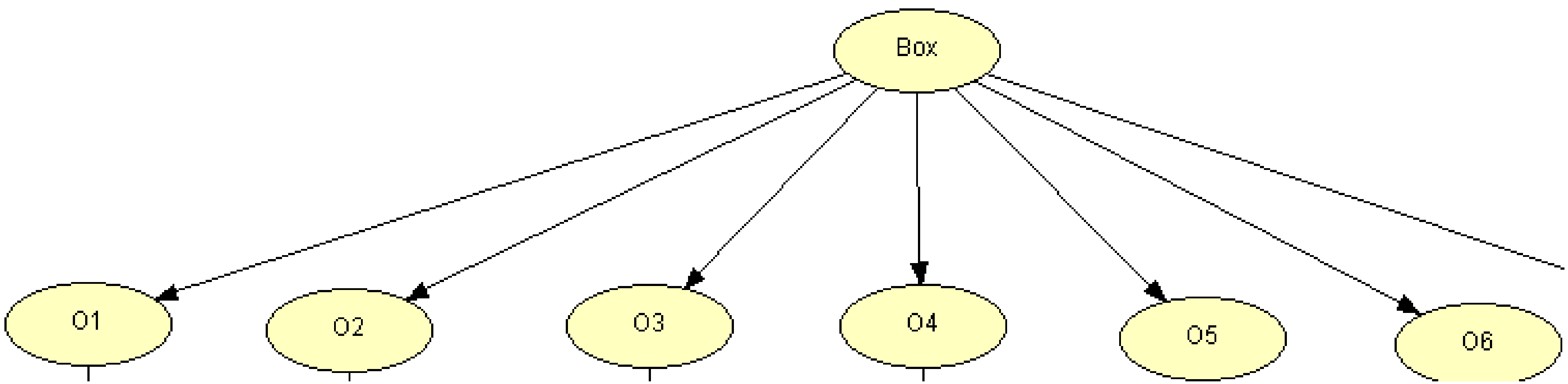
Cause-effect representation

box content → observed color



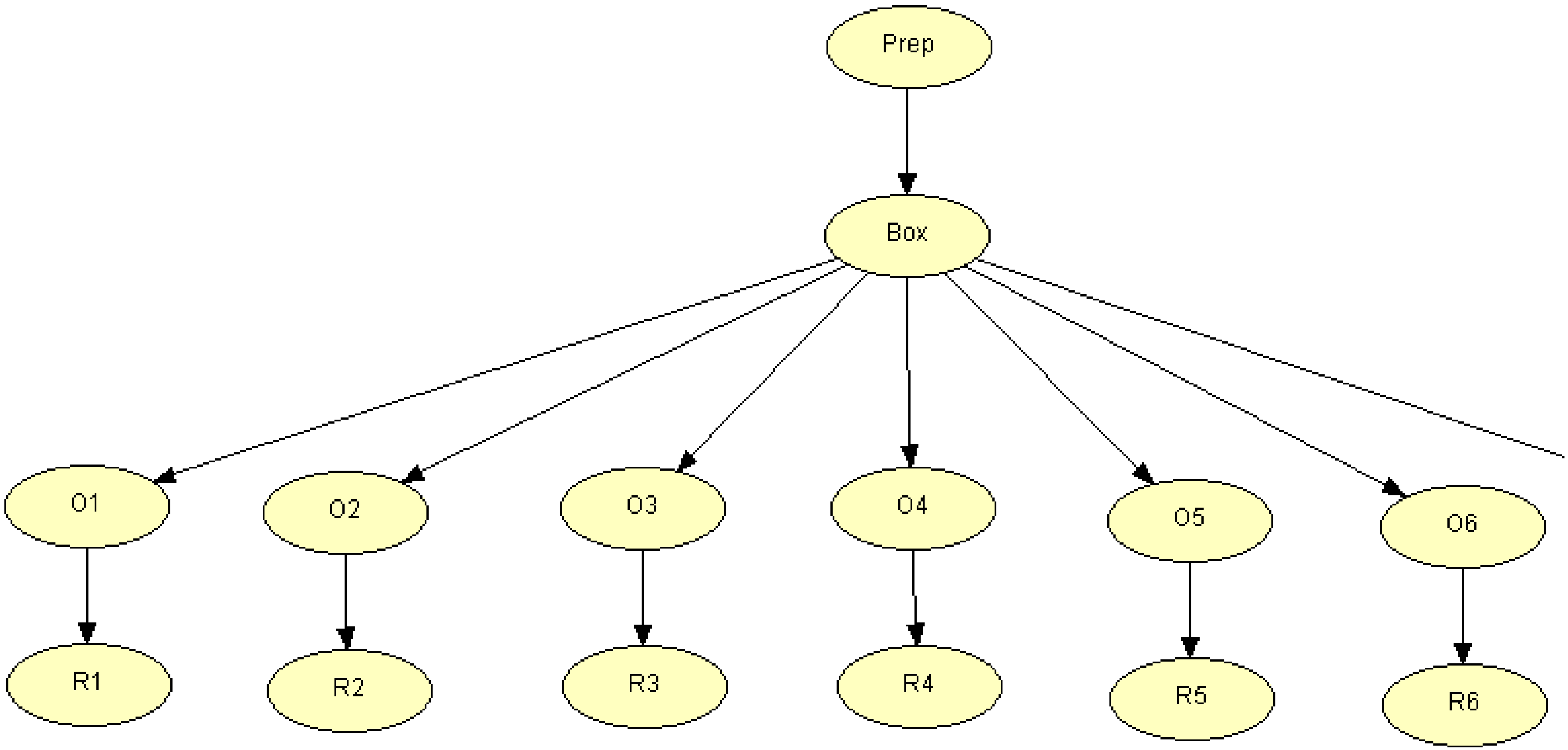
Cause-effect representation

box content $\rightarrow$ observed color

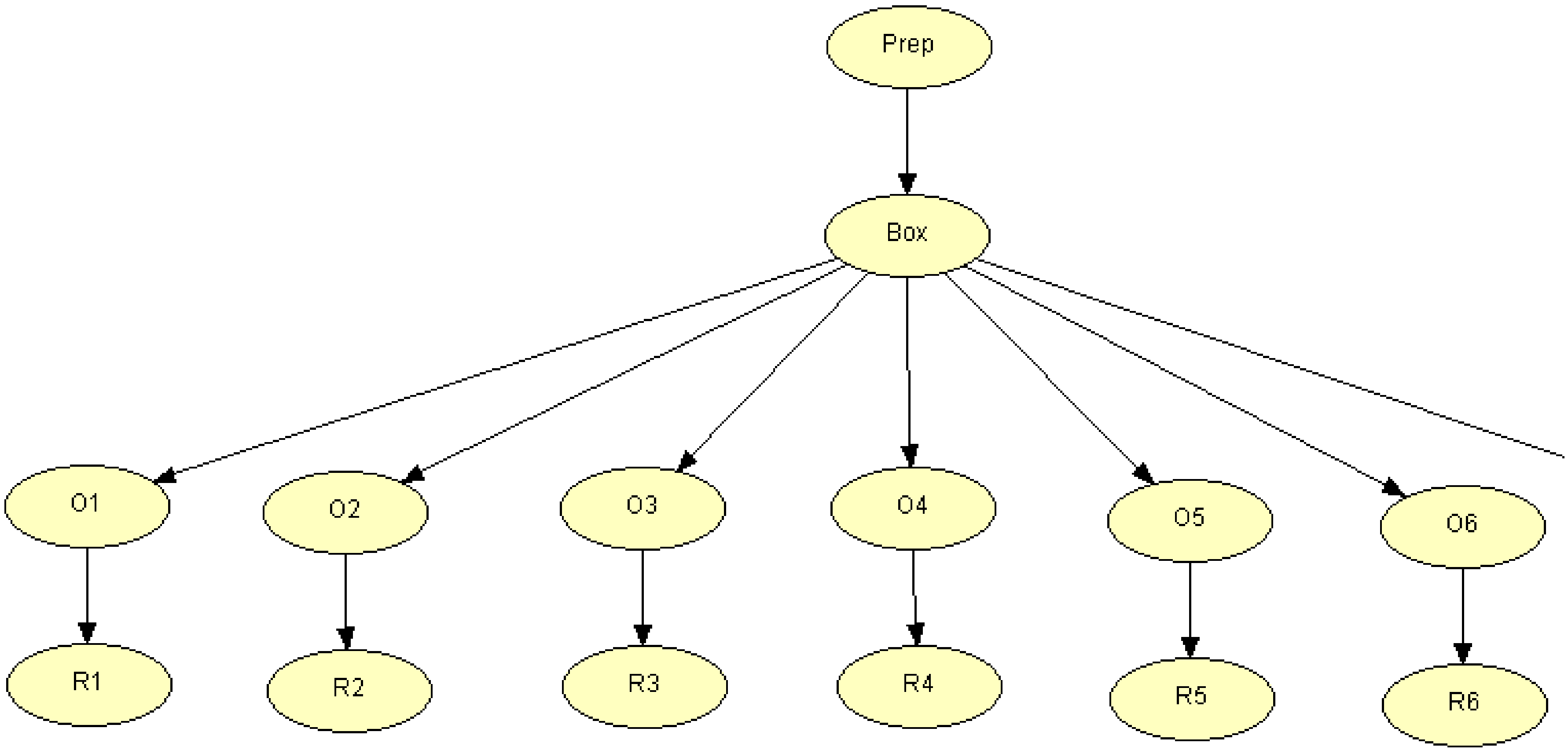


An effect might be the cause of another effect $\Longrightarrow$

A network of causes and effects



A network of causes and effects



and so on. . .

$\Rightarrow$ Physics applications

Inferring ‘proportions’

Let’s turn the toy experiment to a ‘serious’ physics case:

- Inferring H_j is the same as inferring the proportion of white balls:

$$H_j \quad \longleftrightarrow \quad j \quad \longleftrightarrow \quad p = \frac{j}{5}$$

Inferring ‘proportions’

Let’s turn the toy experiment to a ‘serious’ physics case:

- Inferring H_j is the same as inferring the proportion of white balls:

$$H_j \quad \longleftrightarrow \quad j \quad \longleftrightarrow \quad p = \frac{j}{5}$$

- Increase the number of balls

$$n: \quad 6 \rightarrow \infty$$

$\Rightarrow p$ continous in $[0, 1]$

Inferring 'proportions'

Let's turn the toy experiment to a 'serious' physics case:

- Inferring H_j is the same as inferring the proportion of white balls:

$$H_j \quad \longleftrightarrow \quad j \quad \longleftrightarrow \quad p = \frac{j}{5}$$

- Increase the number of balls

$$n: \quad 6 \to \infty$$

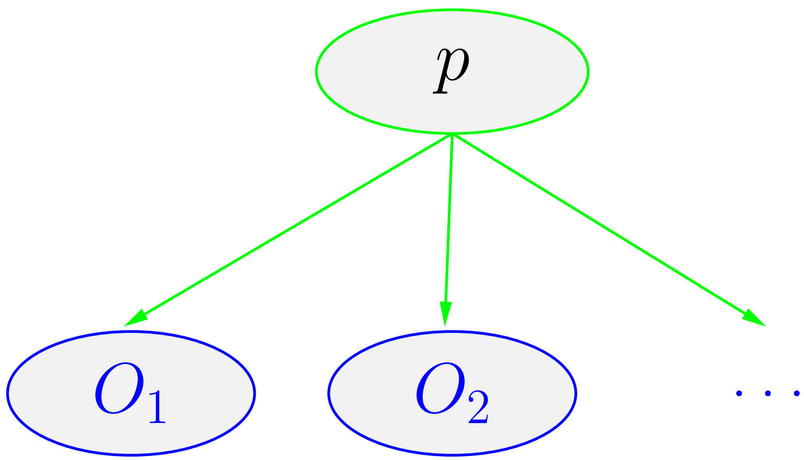
$\Rightarrow$ p continous in $[0, 1]$

- Generalize White/Black $\longrightarrow$ Success/Failure

$\Rightarrow$ efficiencies, branching ratios, ...

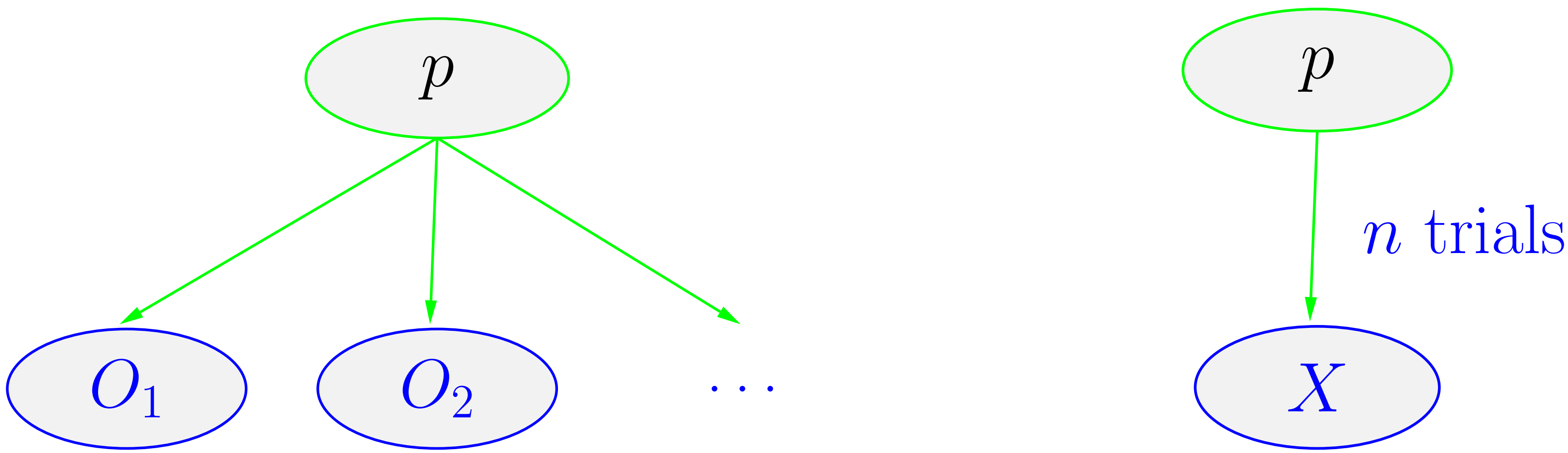
Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p



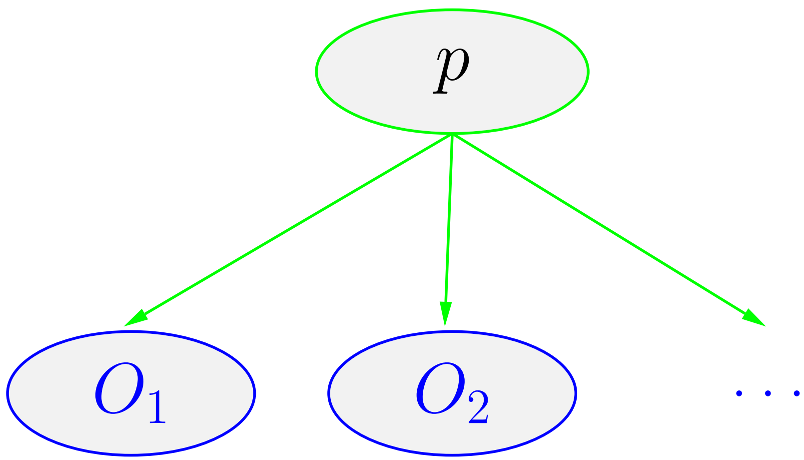
Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p

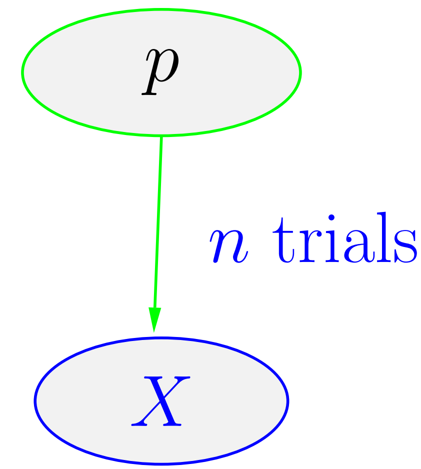


Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p



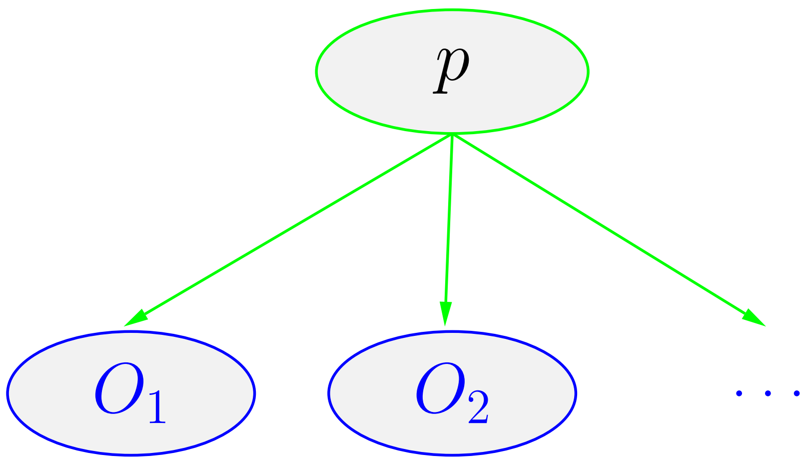
"independent Bernoulli trials"



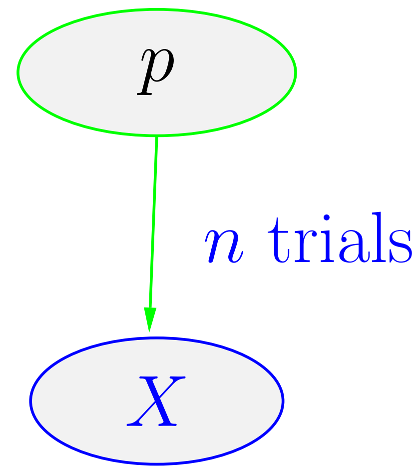
"binomial distribution"

Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p



"independent Bernoulli trials"

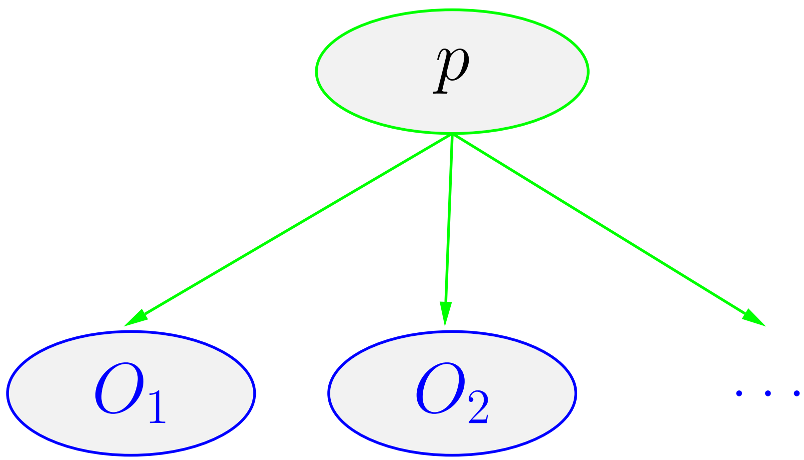


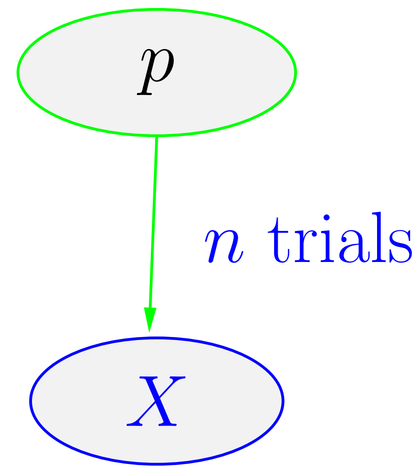
"binomial distribution"

$\Rightarrow$ In the light of the experimental information
there will be values of p we shall believe more,
and others we shall believe less.

Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p





"independent Bernoulli trials"

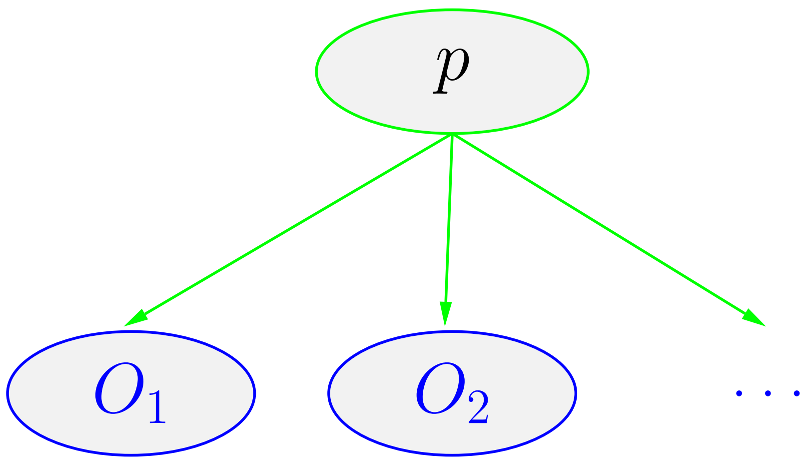
"binomial distribution"

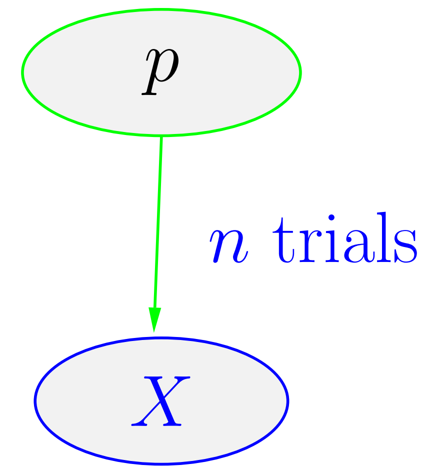
$$P(p_i \,|\, O_1, O_2, \ldots)$$
$$f(p \,|\, O_1, O_2, \ldots)$$

$$P(p_i \,|\, X, n)$$
$$f(p \,|\, X, n)$$

Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p





"independent Bernoulli trials"

$$P(p_i \,|\, O_1, O_2, \ldots)$$
$$f(p \,|\, O_1, O_2, \ldots)$$

$$\propto f(O_1, O_2, \ldots \,|\, p) \cdot f_0(p)$$

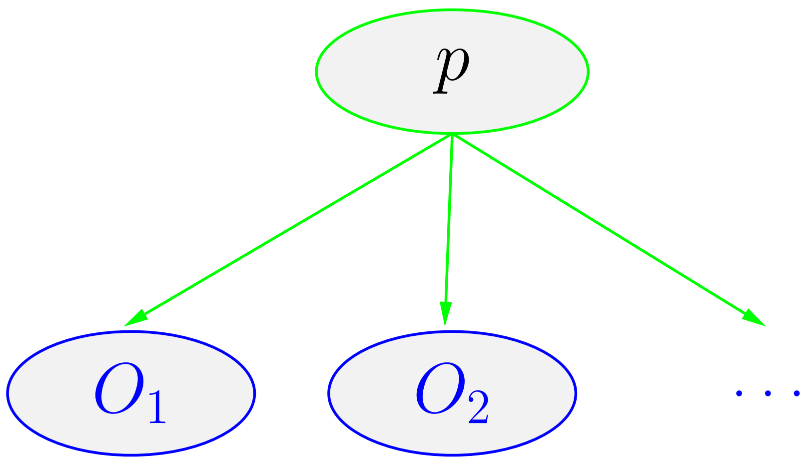
"binomial distribution"

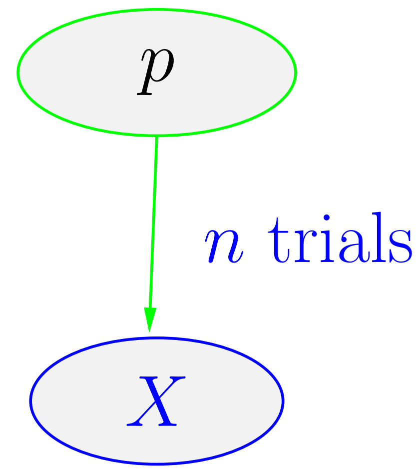
$$P(p_i \,|\, X, n)$$
$$f(p \,|\, X, n)$$

$$\propto f(X \,|\, n, p) \cdot f_0(p)$$

Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p





"independent Bernoulli trials" "binomial distribution"

$$P(p_i\,|\,O_1, O_2, \ldots) \qquad P(p_i\,|\,X, n)$$
$$f(p\,|\,O_1, O_2, \ldots) \qquad f(p\,|\,X, n)$$

Are the two inferences the same?
(not obvious in principle)

Graphical models

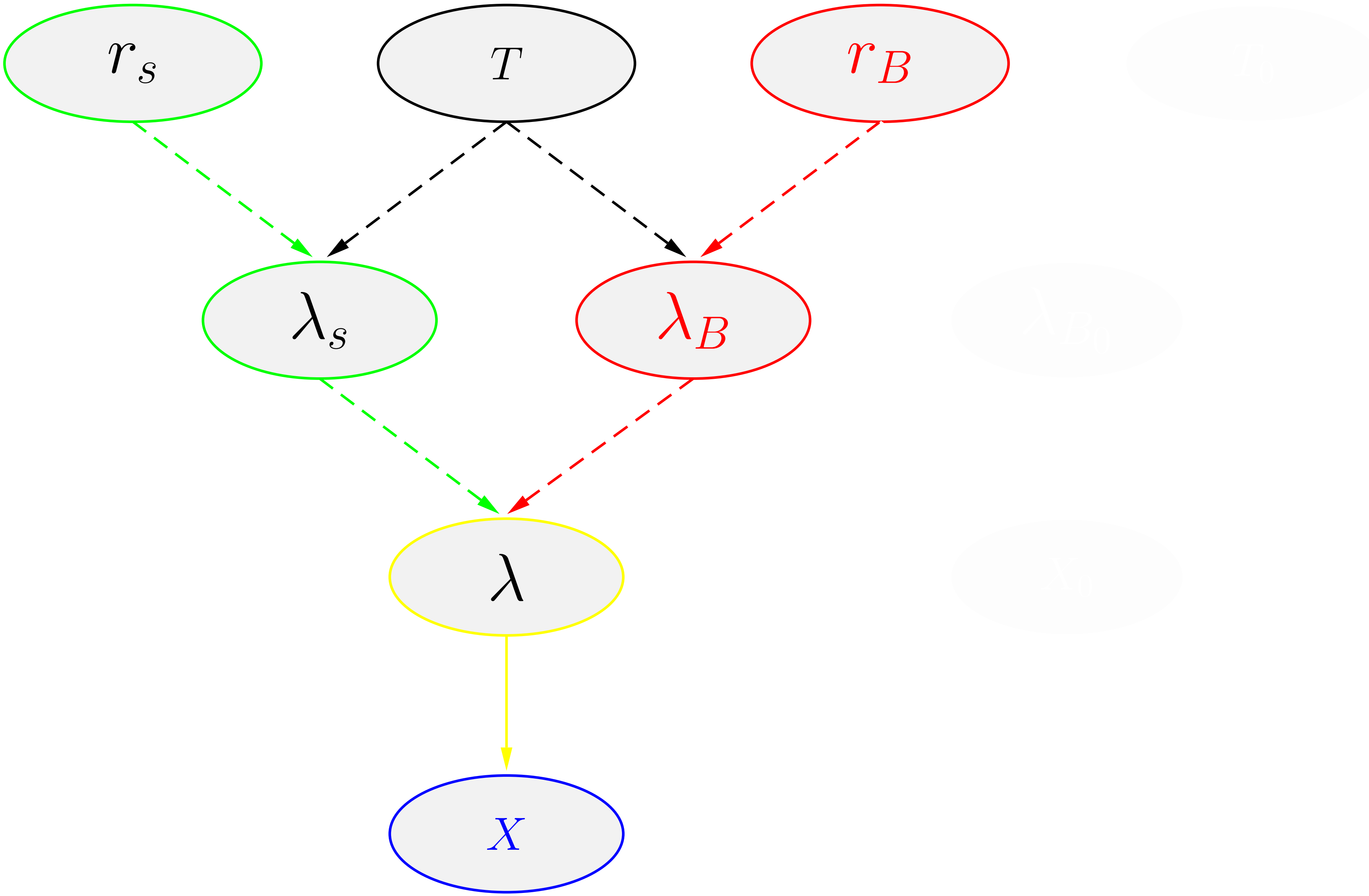
Before analysing in some detail this case let's make an overview of other important cases in physics

Graphical models

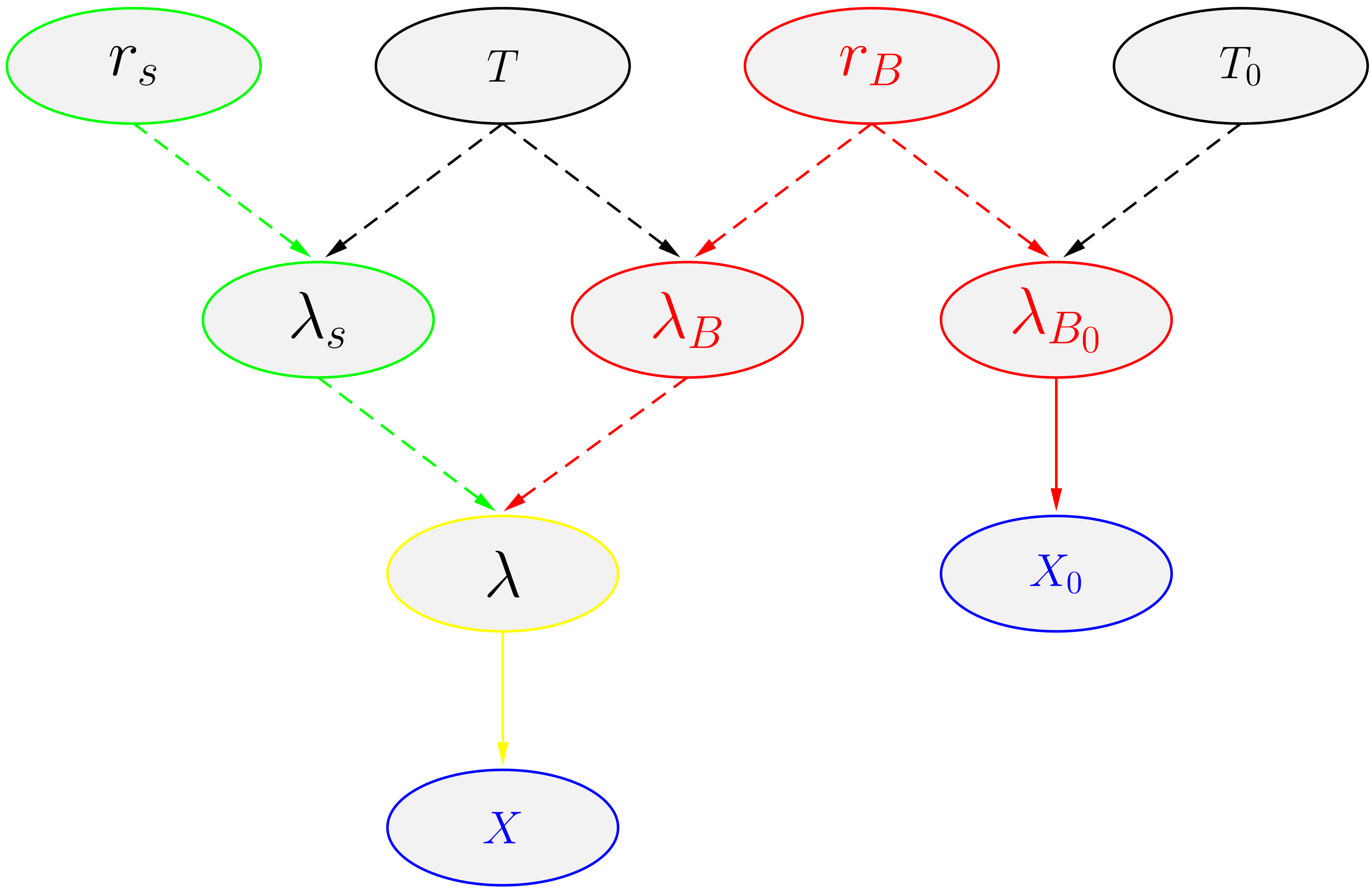
Before analysing in some detail this case let's make an overview of other important cases in physics

⇒ Nowadays, thanks to progresses in mathematics and computing, drawing the problem as a 'belief network' is more than 1/2 step towards its solution!

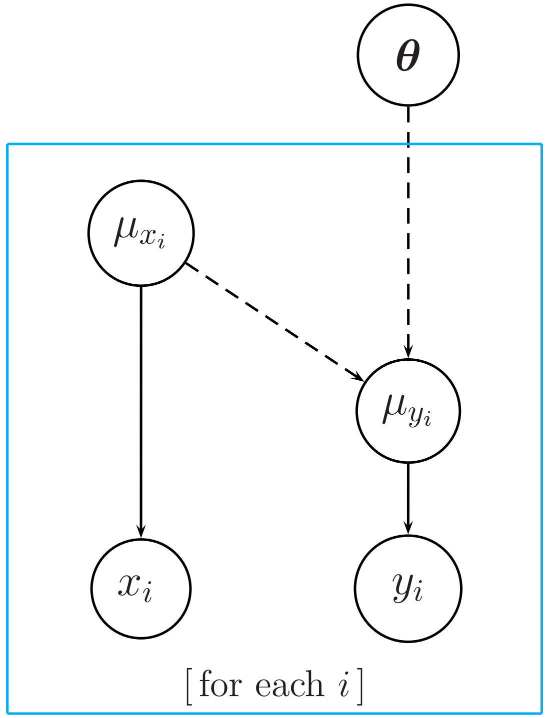
Signal and background



Signal and background



A different way to view fit issues



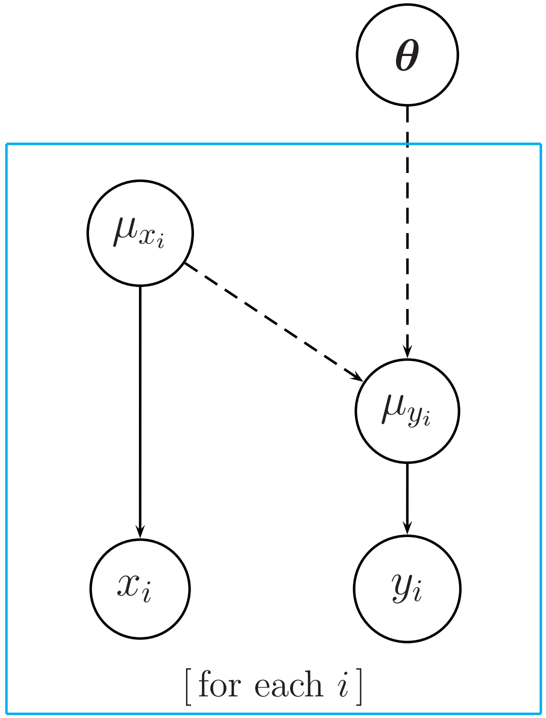
Determistic link μ_x's to μ_y's

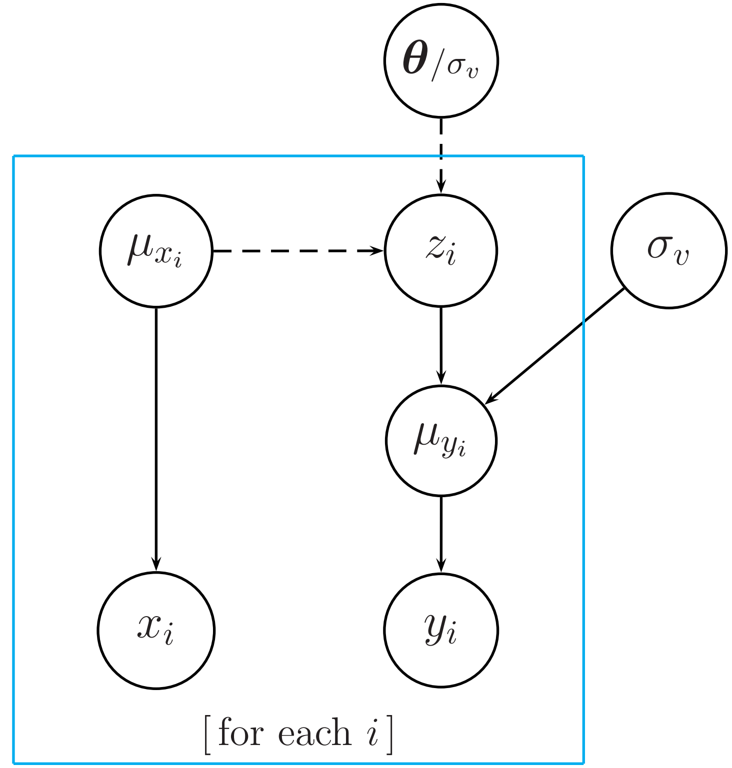
Probabilistic links $\mu_x \rightarrow x$, $\mu_y \rightarrow y$

(errors on both axes!)

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

A different way to view fit issues



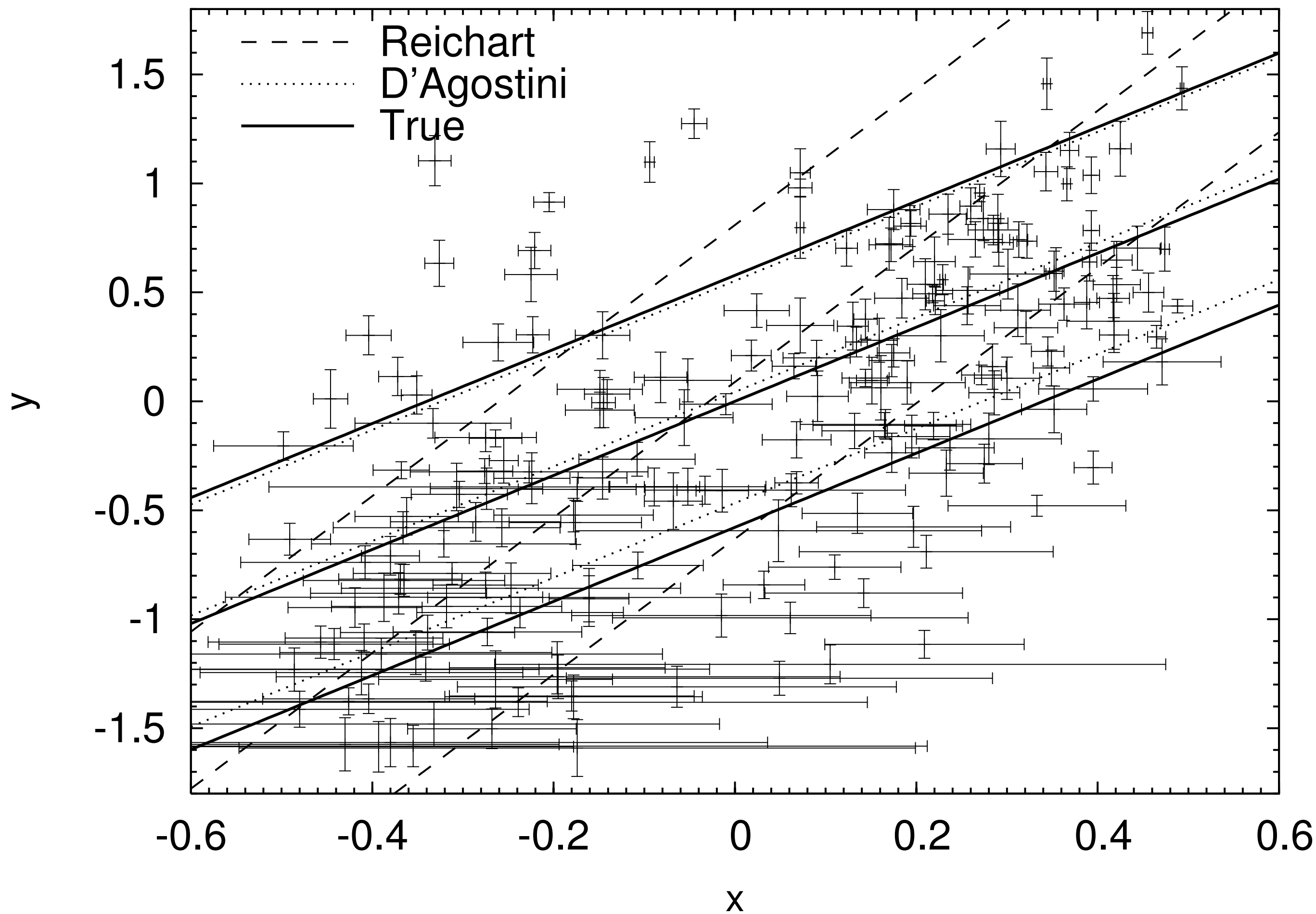


Determistic link μ_x's to μ_y's
Probabilistic links $\mu_x \to x$, $\mu_y \to y$
(errors on both axes!)
$\Rightarrow$ aim of fit: $\{\boldsymbol{x}, \boldsymbol{y}\} \to \boldsymbol{\theta}$

Extra spread
of the data points

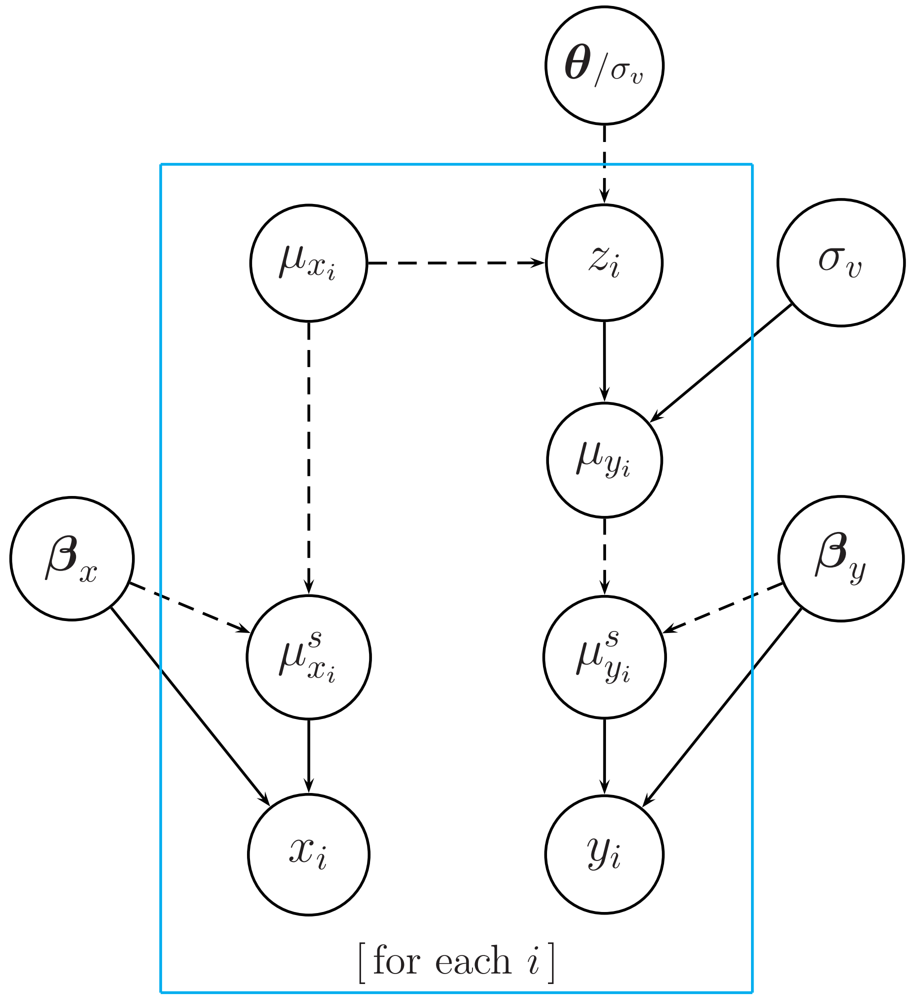
A different way to view fit issues

A physics case (from Gamma ray burts):



(Guidorzi et al., 2006)

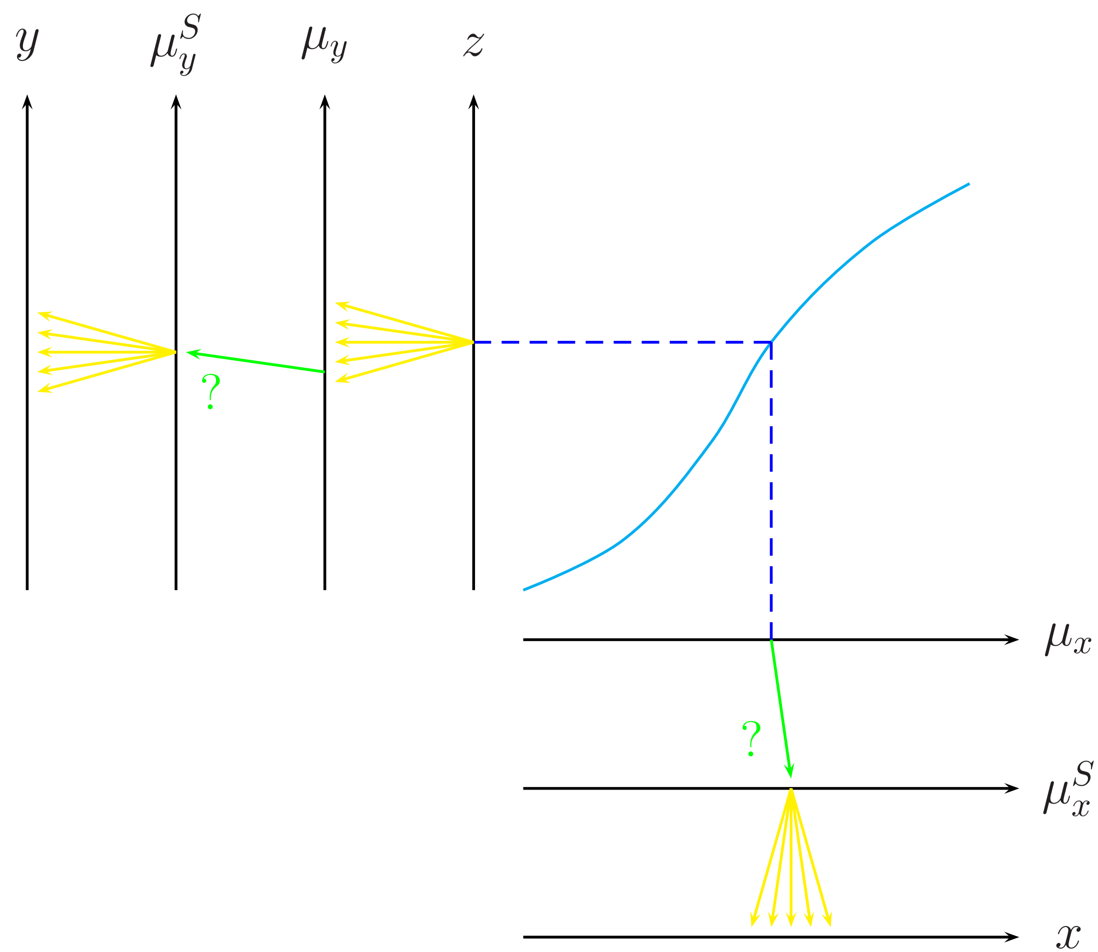
A different way to view fit issues



Adding systematics

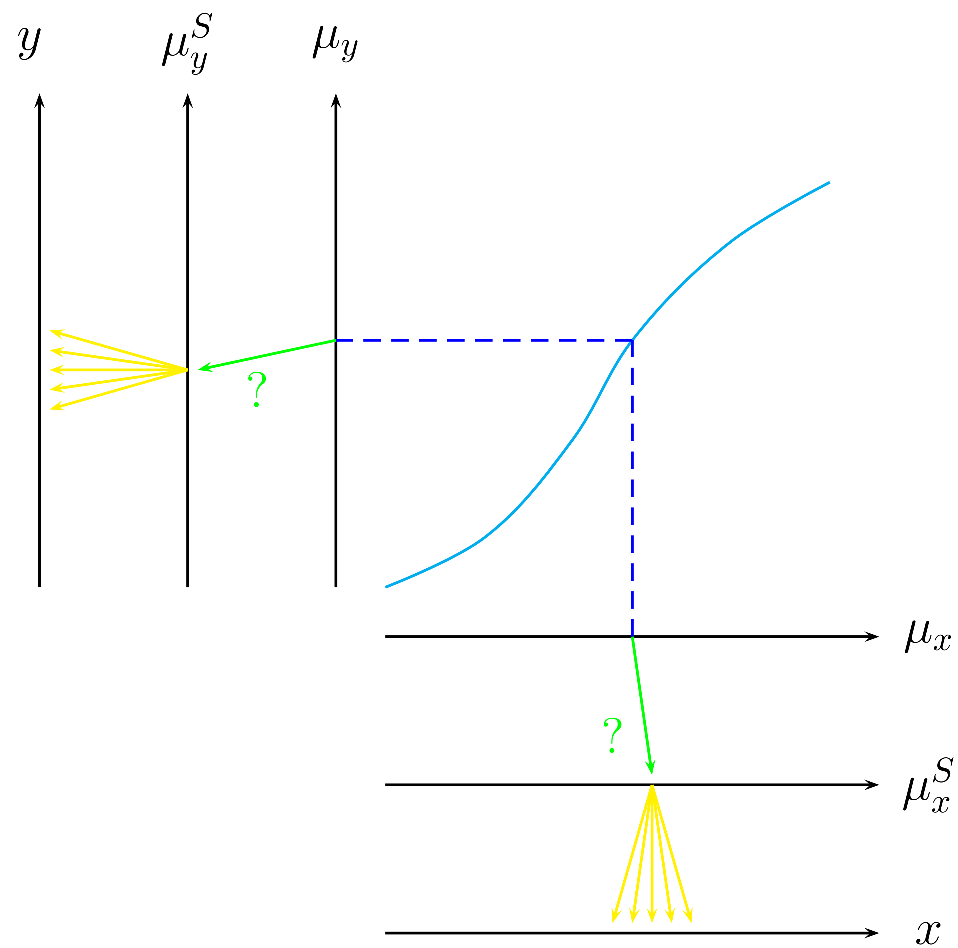
A different way to view fit issues

Stated differently:



A different way to view fit issues

Only systematics (on both axes)



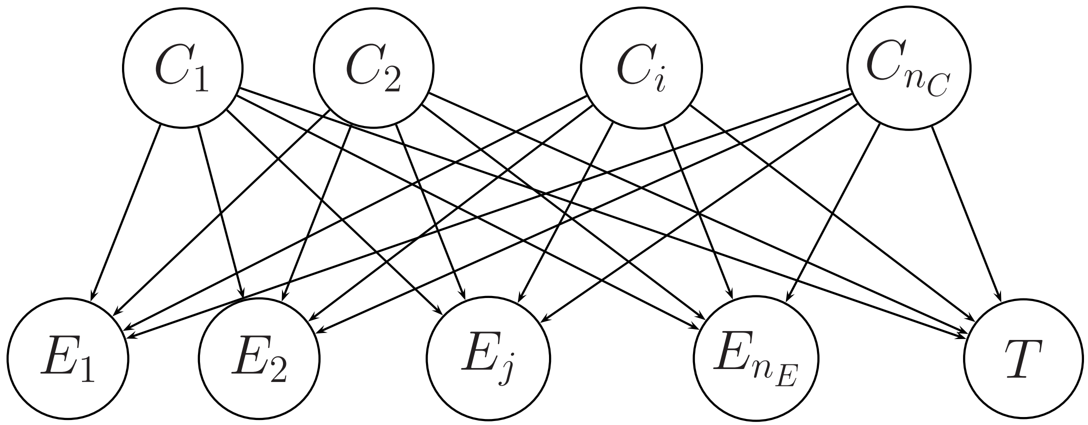
A different way to view fit issues

In this approach systematic effects reflect our uncertainty

$\Rightarrow$ they can be handled rigorously using probability theory!

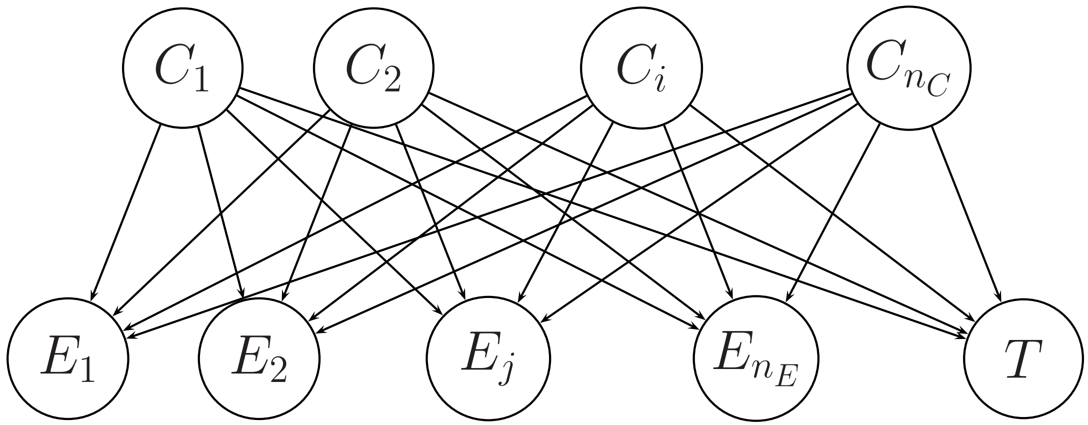
Unfolding a discretized spectrum

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

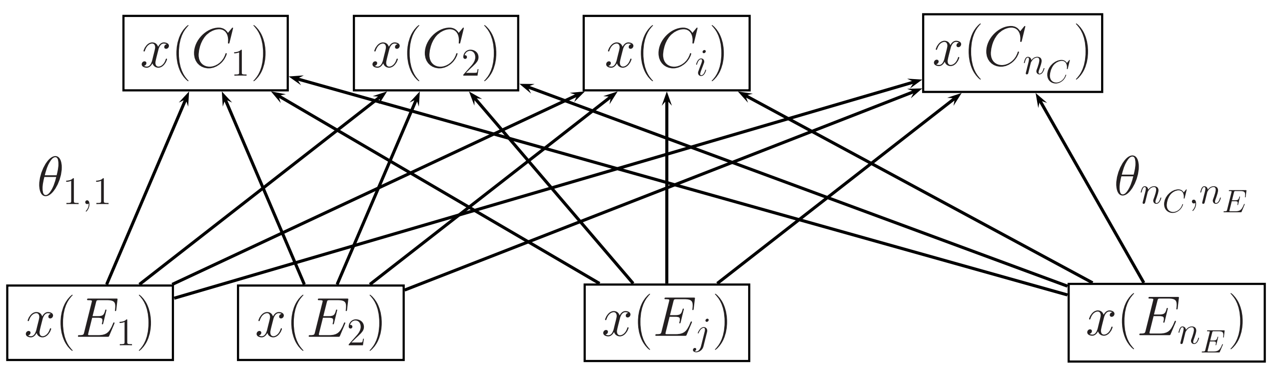


Unfolding a discretized spectrum

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

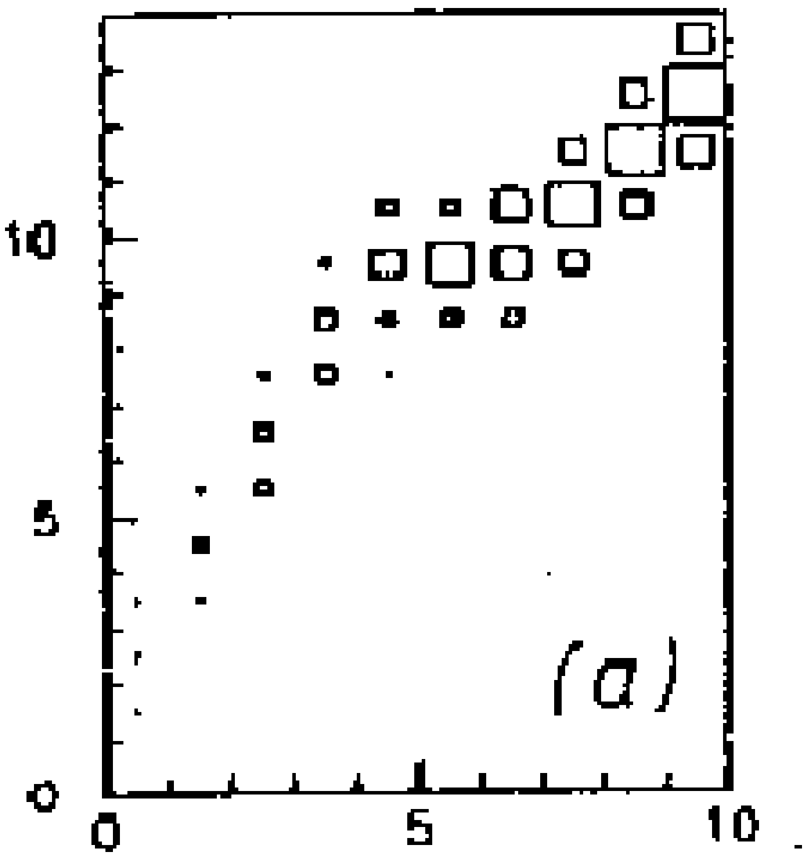


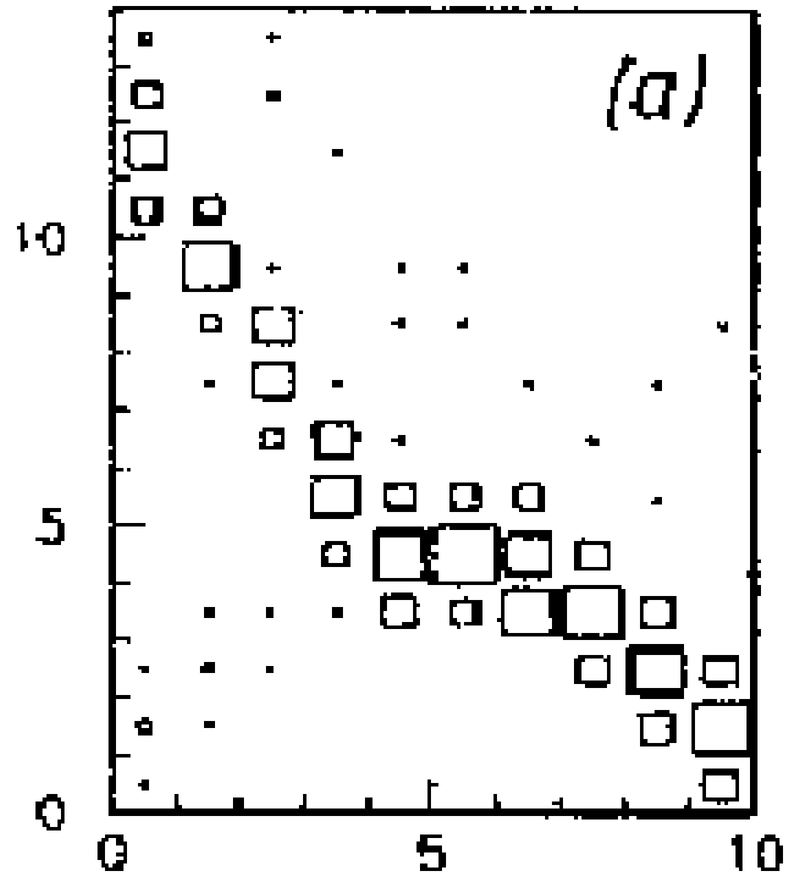
Sharing the observed events among the cause-bins



Unfolding a discretized spectrum

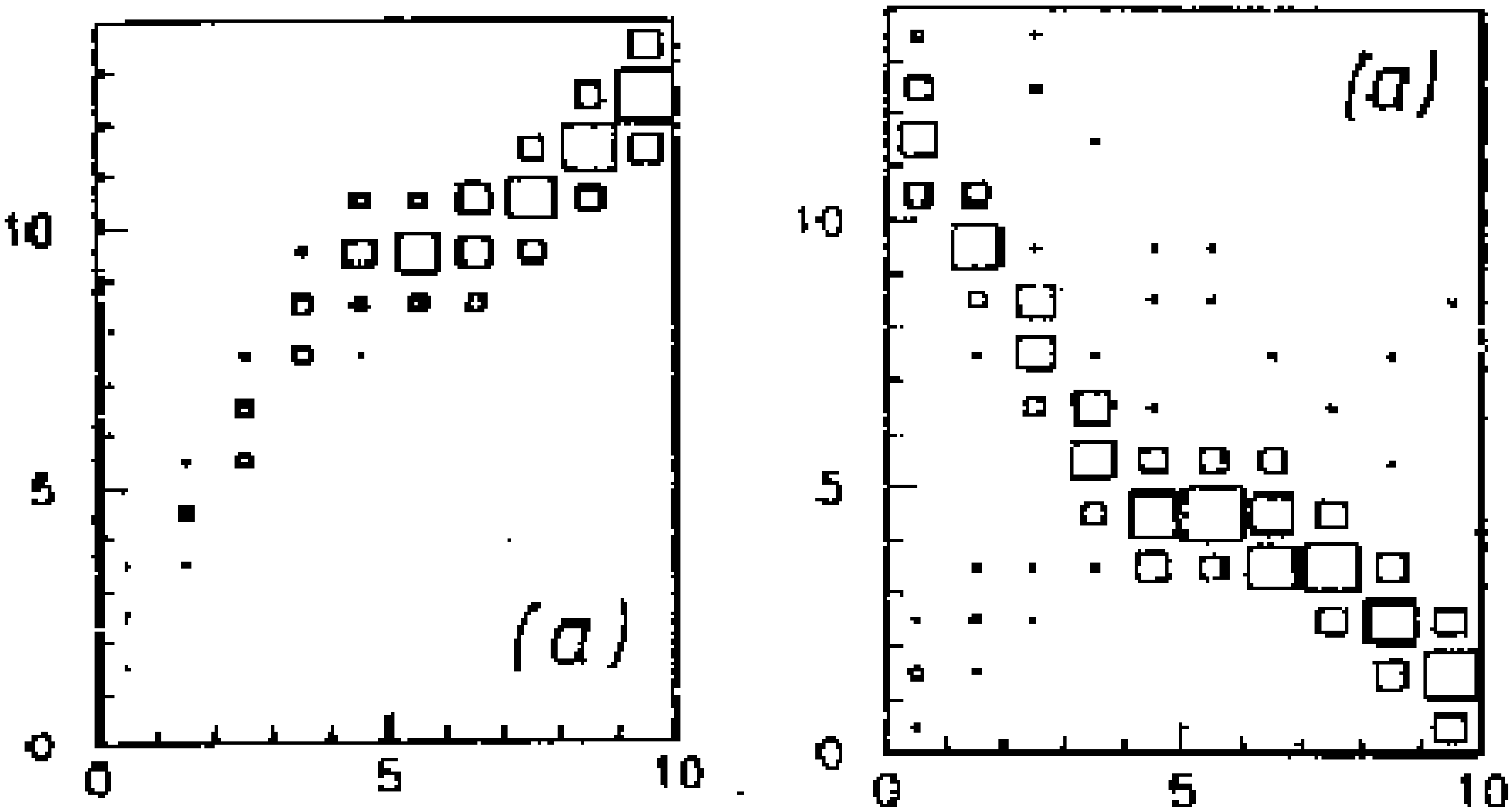
Academic (quite nasty!) smearing matrices:





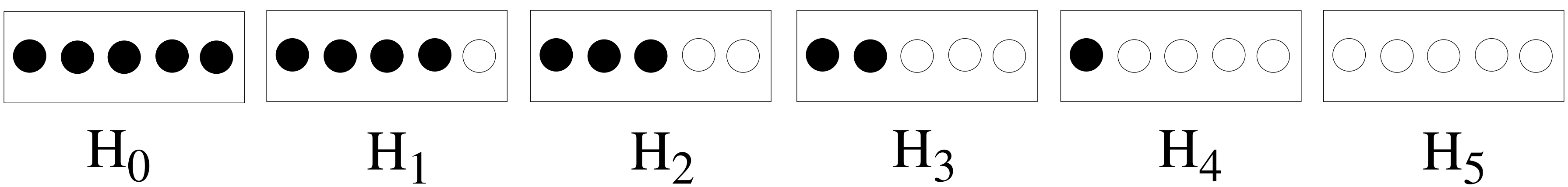
Unfolding a discretized spectrum

Academic (quite nasty!) smearing matrices:



Some demos ⇒

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:

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

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

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

Our prior belief about H_j

Collecting the pieces of information we need

Our tool:

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

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

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

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:

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

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

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

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:

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

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

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

Probability of E_i taking account all possible H_j

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

We can rewrite it as

$P(E_i \mid I) = \sum_j P(E_i \mid H_j, I) \cdot P(H_j \mid I)$

We are ready

Now that we have set up our formalism, let's play a little

- analyse real data
- some simulations

Then

- $H_j \longleftrightarrow j \longleftrightarrow p_j$
- extending p to a continuum:

  $\Rightarrow$ Bayes' billiard

  (prototype for all questions related to efficiencies, branching ratios)
- On the meaning of p

Which box? Which ball?

Inferential/forecasting history:

1. $k = 0$
   $P_0(H_j) = P(H_j \,|\, I_0)$ (priors)
2. `begin loop`:
   $k = k + 1$
   $\Rightarrow E^{(k)}$ (k-th extraction)
3. $P_k(H_j \,|\, I_k) \propto P(E^{(k)} \,|\, H_j) \times P_{k-1}(H_j \,|\, I_k)$

   $P_k(E_i \,|\, I_k) = \sum_j P(E_i \,|\, H_j) \cdot P_k(H_j \,|\, I_k)$
4. $\rightarrow$ `go to 2`

Which box? Which ball?

Inferential/forecasting history:

1. $k = 0$
   $P_0(H_j) = P(H_j \,|\, I_0)$ (priors)
2. `begin loop`:
   $k = k + 1$
   $\Rightarrow E^{(k)}$ (k-th extraction)
3. $P_k(H_j \,|\, I_k) \propto P(E^{(k)} \,|\, H_j) \times P_{k-1}(H_j \,|\, I_k)$

   $P_k(E_i \,|\, I_k) = \sum_j P(E_i \,|\, H_j) \cdot P_k(H_j \,|\, I_k)$
4. $\rightarrow$ `go to 2`

Let's play!

Bayes' billiard

This is the original problem in the theory of chances solved by Thomas Bayes in late '700:

- imagine you roll a ball at random on a billiard;
- you mark the relative position of the ball along the billiard's length (l/L) and remove the ball
- then you roll at random other balls
  - write down if it stopped left or right of the first ball;
  - remove it and go on with n balls.
- Somebody has to guess the position of the first ball knowing only how mane balls stopped left and how many stoppe right

Bayes' billiard and Bernoulli trials

It is easy to recongnize the analogy:

- Left/Right $\rightarrow$ Success/Failure
- if Left $\leftrightarrow$ Success:
  - $l/L \leftrightarrow p$ of binomial (Bernoulli trials)

Bayes' billiard and Bernoulli trials

It is easy to recongnize the analogy:

- Left/Right $\rightarrow$ Success/Failure
- if Left $\leftrightarrow$ Success:
  - $l/L \leftrightarrow p$ of binomial (Bernoulli trials)

Solution with modern notation:
Imagine a sequence $\{S, S, F, S, \ldots\}$ [f_0 is uniform]:

Bayes' billiard and Bernoulli trials

It is easy to recongnize the analogy:

- Left/Right $\rightarrow$ Success/Failure
- if Left $\leftrightarrow$ Success:
  - $l/L \leftrightarrow p$ of binomial (Bernoulli trials)

Solution with modern notation:
Imagine a sequence $\{S, S, F, S, \ldots\}$ [f_0 is uniform]:

$$f(p\,|\,S) \;\;\propto\;\; f(S\,|\,p) = p$$

Bayes' billiard and Bernoulli trials

It is easy to recongnize the analogy:

- Left/Right $\rightarrow$ Success/Failure
- if Left $\leftrightarrow$ Success:
  - $l/L \leftrightarrow p$ of binomial (Bernoulli trials)

Solution with modern notation:
Imagine a sequence $\{S, S, F, S, \ldots\}$ [f_0 is uniform]:

$$f(p\,|\,S) \;\propto\; f(S\,|\,p) = p$$

$$f(p\,|\,S,S) \;\propto\; f(S\,|\,p)\cdot f(p\,|\,S) = p^2$$

Bayes' billiard and Bernoulli trials

It is easy to recongnize the analogy:

- Left/Right $\rightarrow$ Success/Failure
- if Left $\leftrightarrow$ Success:
  - $l/L \leftrightarrow p$ of binomial (Bernoulli trials)

Solution with modern notation:
Imagine a sequence $\{S, S, F, S, \ldots\}$ [f_0 is uniform]:

$$
\begin{aligned}
f(p\,|\,S) &\propto f(S\,|\,p) = p \\
f(p\,|\,S, S) &\propto f(S\,|\,p) \cdot f(p\,|\,S) = p^2 \\
f(p\,|\,S, S, F) &\propto f(F\,|\,p) \cdot f(p\,|\,S, S) = p^2(1-p)
\end{aligned}
$$

Bayes' billiard and Bernoulli trials

It is easy to recongnize the analogy:

- Left/Right $\rightarrow$ Success/Failure
- if Left $\leftrightarrow$ Success:
  - $l/L \leftrightarrow p$ of binomial (Bernoulli trials)

Solution with modern notation:
Imagine a sequence $\{S, S, F, S, \ldots\}$ [f_0 is uniform]:

$$
\begin{aligned}
f(p\,|\,S) &\propto f(S\,|\,p) = p \\
f(p\,|\,S,S) &\propto f(S\,|\,p)\cdot f(p\,|\,S) = p^2 \\
f(p\,|\,S,S,F) &\propto f(F\,|\,p)\cdot f(p\,|\,S,S) = p^2(1-p) \\
\ldots & \quad \ldots \\
f(p\,|\,\#S,\#F) &\propto p^{\#S}(1-p)^{\#F} = p^{\#S}(1-p)^{(1-\#s)}
\end{aligned}
$$

Bayes' billiard and Bernoulli trials

It is easy to recongnize the analogy:

- Left/Right $\rightarrow$ Success/Failure
- if Left $\leftrightarrow$ Success:
  - $l/L \leftrightarrow p$ of binomial (Bernoulli trials)

Solution with modern notation:
Imagine a sequence $\{S, S, F, S, \ldots\}$ [f_0 is uniform]:

$$
\begin{aligned}
f(p\,|\,S) &\propto f(S\,|\,p) = p \\
f(p\,|\,S,S) &\propto f(S\,|\,p)\cdot f(p\,|\,S) = p^2 \\
f(p\,|\,S,S,F) &\propto f(F\,|\,p)\cdot f(p\,|\,S,S) = p^2(1-p) \\
\ldots & \qquad \ldots \\
f(p\,|\,\#S,\#F) &\propto p^{\#S}(1-p)^{\#F} = p^{\#S}(1-p)^{(1-\#s)} \\
f(p\,|\,x,n) &\propto p^{x}(1-p)^{(n-x)} \qquad\qquad [x=\#S]
\end{aligned}
$$

Parametric inference

$\rightarrow$ Choose a model and infer its parameter(s).

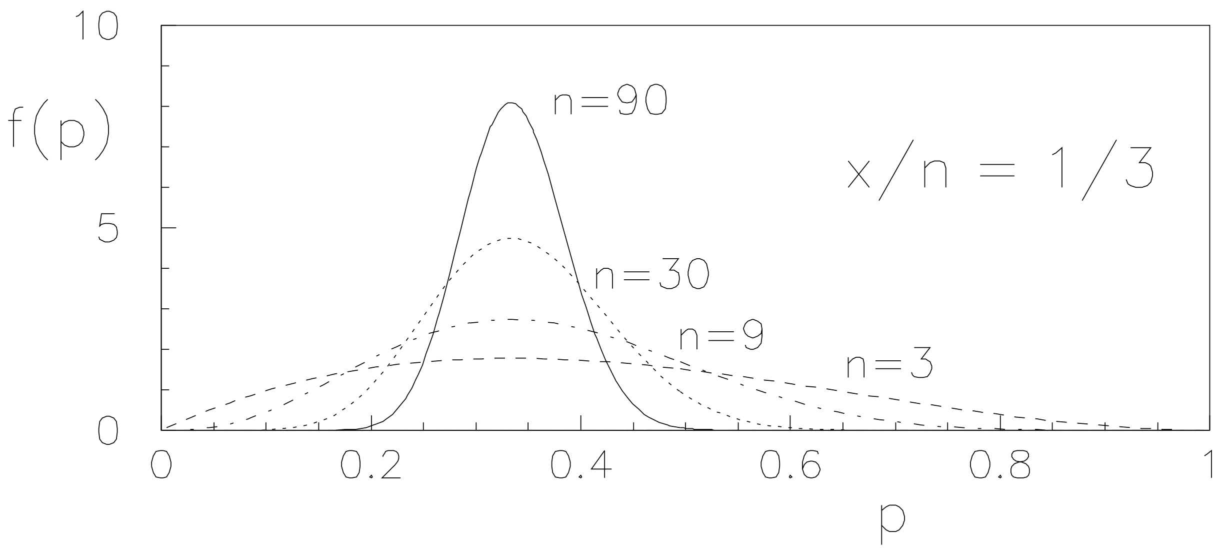
Bayes theorem for continuous variables has following structure

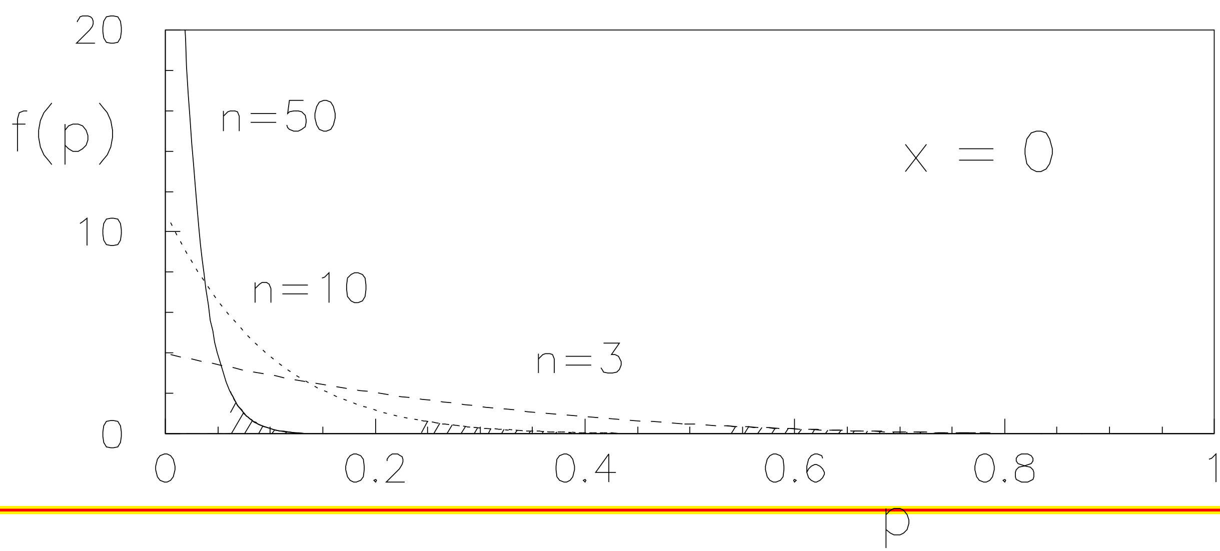
$$f(\theta\,|\,\text{data}) \propto f(\text{data}\,|\,\theta)\, f_0(\theta)$$

$$\begin{aligned}
f(p\,|\,x,n,\mathcal{B}) &= \frac{f(x\,|\,\mathcal{B}_{n,p})\, f_\circ(p)}{\int_0^1 f(x\,|\,\mathcal{B}_{n,p})\, f_\circ(p)\, dp} \\
&= \frac{\frac{n!}{(n-x)!\,x!}\, p^x\,(1-p)^{n-x}\, f_\circ(p)}{\int_0^1 \frac{n!}{(n-x)!\,x!}\, p^x\,(1-p)^{n-x}\, f_\circ(p)\, dp} \\
&= \frac{p^x\,(1-p)^{n-x}}{\int_0^1 p^x\,(1-p)^{n-x}\, dp}\,,
\end{aligned}$$

Inferring the Binomial p

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





Inferring the Binomial p

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

$$\begin{aligned} \mathsf{E}(p) &= \frac{x+1}{n+2} \\ \mathsf{Var}(p) &= \frac{(x+1)(n-x+1)}{(n+3)(n+2)^2} \\ &= \mathsf{E}(p)\;(1-\mathsf{E}(p))\;\frac{1}{n+3}. \end{aligned}$$

Laplace's rule of successions

Interpretation of E(p)

Think at any future event $E_{i>n}$
$\Rightarrow$ if we were sure of p, then our confidence on $E_{i>n}$ will be exactly p, i.e.

$$P(E_i \,|\, p) = p\,.$$

Interpretation of E(p)

Think at any future event $E_{i>n}$
$\Rightarrow$ if we were sure of p, then our confidence on $E_{i>n}$ will be exactly p, i.e.

$$P(E_i \,|\, p) = p\,.$$

But we are uncertain about p.

How much should we believe $E_{i>n}$?.

Interpretation of E(p)

Think at any future event $E_{i>n}$
$\Rightarrow$ if we were sure of p, then our confidence on $E_{i>n}$ will be exactly p, i.e.

$$P(E_i\,|\,p) = p\,.$$

But we are uncertain about p.
How much should we believe $E_{i>n}$?.

$$\begin{aligned} P(E_{i>n}\,|\,x,n,\mathcal{B}) &= \int_0^1 P(E_i\,|\,p)\,f(p\,|\,x,n,\mathcal{B})\,\mathrm{d}p \\ &= \int_0^1 p\,f(p\,|\,x,n,\mathcal{B})\,\mathrm{d}p \\ &= \mathrm{E}(p) \\ &= \frac{x+1}{n+2} \quad \text{(for uniform prior)}\,. \end{aligned}$$

From frequencies to probabilities

$$\mathsf{E}(p) = \frac{x+1}{n+2}$$

Laplace's rule of successions

$$\mathsf{Var}(p) = \mathsf{E}(p)\,(1-\mathsf{E}(p))\,\frac{1}{n+3}.$$

For 'large' n, x and $n-x$: asymptotic behaviors of $f(p)$:

$$\mathsf{E}(p) \approx p_m = \frac{x}{n} \quad [\text{with } p_m \text{ mode of } f(p)]$$

$$\sigma_p \approx \sqrt{\frac{p_m\,(1-p_m)}{n}} \xrightarrow[n\to\infty]{} 0$$

$$p \sim \mathcal{N}(p_m, \sigma_p)\,.$$

Under these conditions the frequentistic "definition" (evaluation rule!) of probability (x/n) is recovered.

Special case with $x = 0$

$$
\begin{aligned}
f(p \,|\, 0, n, \mathcal{B}) &= (n+1)\,(1-p)^n \\
F(p \,|\, 0, n, \mathcal{B}) &= 1 - (1-p)^{n+1} \\
p_m &= 0 \\
\mathrm{E}(p) &= \frac{1}{n+2} \longrightarrow \frac{1}{n} \\
\sigma(p) &= \sqrt{\frac{(n+1)}{(n+3)(n+2)^2}} \longrightarrow \frac{1}{n}
\end{aligned}
$$

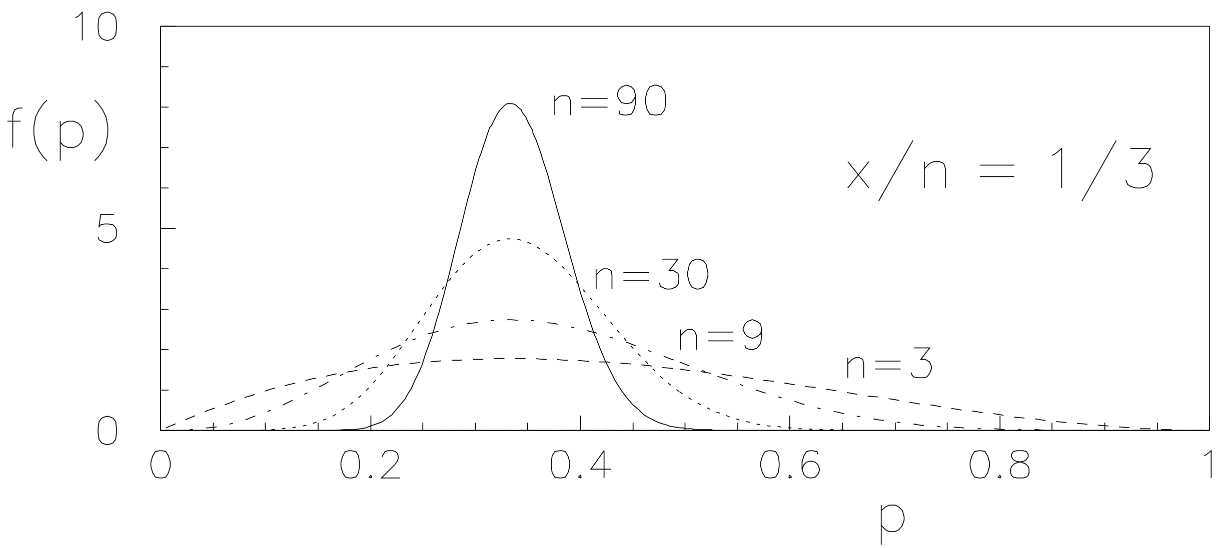
Special case with $x = 0$

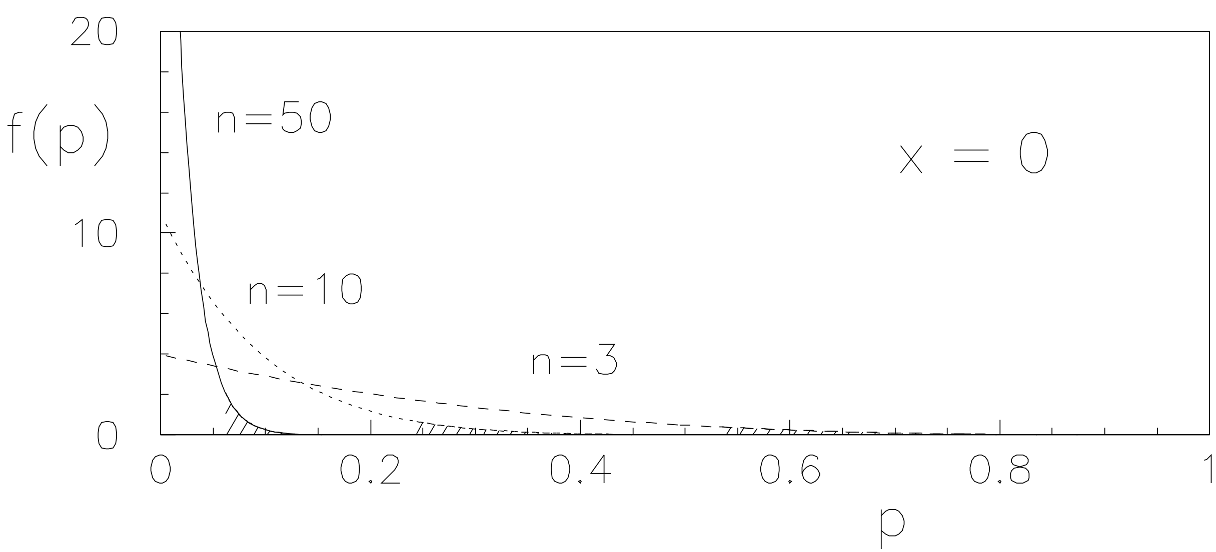
$$f(p\,|\,0,n,\mathcal{B}) = (n+1)\,(1-p)^n$$
$$F(p\,|\,0,n,\mathcal{B}) = 1-(1-p)^{n+1}$$
$$p_m = 0$$
$$\mathrm{E}(p) = \frac{1}{n+2} \longrightarrow \frac{1}{n}$$
$$\sigma(p) = \sqrt{\frac{(n+1)}{(n+3)(n+2)^2}} \longrightarrow \frac{1}{n}$$

$$P(p \le p_u\,|\,0,n,\mathcal{B}) = 95\%$$
$$\Rightarrow \quad p_u = 1 - \sqrt[n+1]{0.05} \;:$$

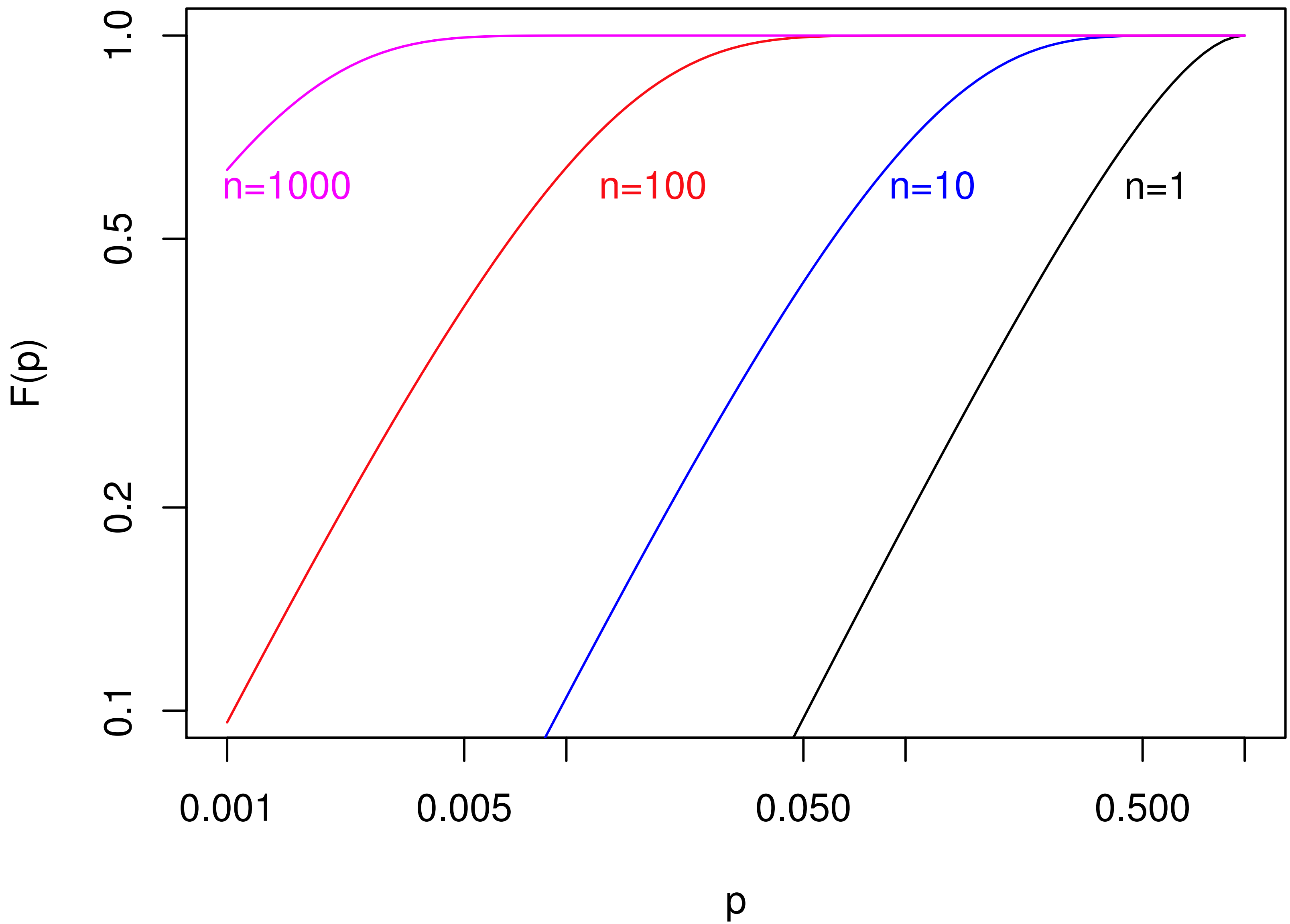
Probabilistic upper bound

Special case with $x = 0$

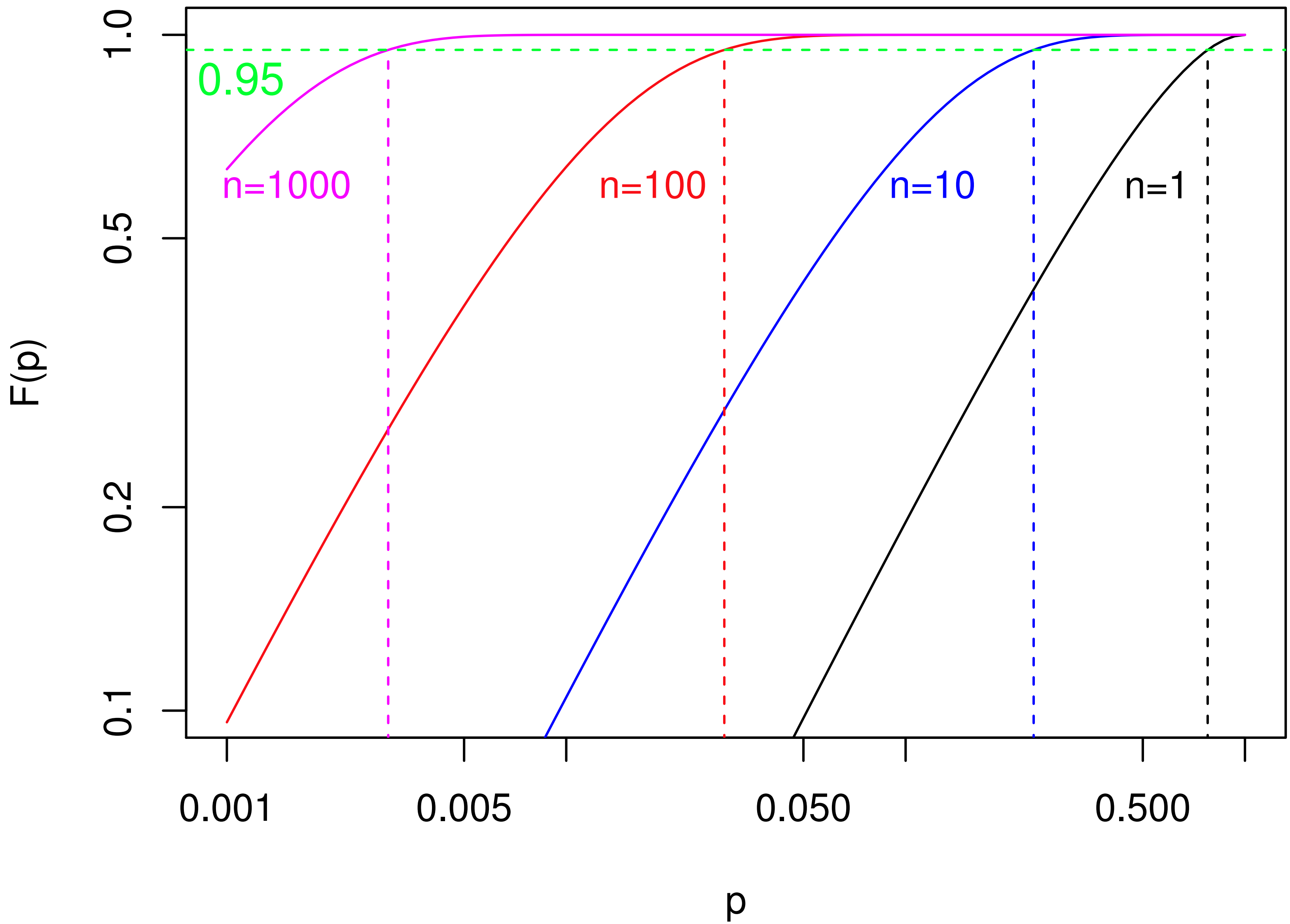




Special case with $x = 0$



Special case with $x = 0$



Special case with $x = n$

As a exercise, let's consider the cases 'all successes', the calculations from general ideas (and assuming a flat prior):

Special case with $x = n$

As a exercise, let's consider the cases 'all successes', the calculations from general ideas (and assuming a flat prior):

- $f(p \,|\, x=n, \mathcal{B}) \propto p^n$ [Bayes Th.]

Special case with $x = n$

As a exercise, let's consider the cases 'all successes', the calculations from general ideas (and assuming a flat prior):

- $f(p \,|\, x = n, \mathcal{B}) \propto p^n$ [Bayes Th.]
- Norm. const.: $k = \int_0^1 p^n = 1/(n+1)$

Special case with $x = n$

As a exercise, let's consider the cases 'all successes', the calculations from general ideas (and assuming a flat prior):

- $f(p\,|\,x=n,\mathcal{B}) \propto p^n$ [Bayes Th.]
- Norm. const.: $k = \int_0^1 p^n = 1/(n+1)$

$\Rightarrow f(p\,|\,x=n,\mathcal{B}) = (n+1)\,p^n$ [posterior]

Special case with $x = n$

As a exercise, let's consider the cases 'all successes', the calculations from general ideas (and assuming a flat prior):

- $f(p \,|\, x = n, \mathcal{B}) \propto p^n$ [Bayes Th.]
- Norm. const.: $k = \int_0^1 p^n = 1/(n+1)$

$\Rightarrow f(p \,|\, x = n, \mathcal{B}) = (n+1)\, p^n$ [posterior]

$\Rightarrow F(p \,|\, x = n, \mathcal{B}) = \int_0^p f(p')dp' = p^{n+1}$ [its cumulative]

Special case with $x = n$

As a exercise, let's consider the cases 'all successes', the calculations from general ideas (and assuming a flat prior):

- $f(p \,|\, x = n, \mathcal{B}) \propto p^n$ [Bayes Th.]
- Norm. const.: $k = \int_0^1 p^n = 1/(n+1)$

$\Rightarrow f(p \,|\, x = n, \mathcal{B}) = (n+1)\, p^n$ [posterior]

$\Rightarrow F(p \,|\, x = n, \mathcal{B}) = \int_0^p f(p')dp' = p^{n+1}$ [its cumulative]

$\Rightarrow p_m = 1$ [mode of posterior]

$\Rightarrow \mathrm{E}[p] = \int_0^1 p f(p)\, dp = \frac{n+1}{n+2}$ [expected value]

Special case with $x = n$

$$\begin{aligned}\sigma^2(p) &= \mathsf{E}[(p-\mathsf{E}[p])^2] = \mathsf{E}[p^2] - \mathsf{E}^2[p] \\ &= \int_0^1 p^2 f(p)\,dp - \left(\frac{n+1}{n+2}\right)^2 \\ &= \frac{n+1}{n+3} - \frac{(n+1)^2}{(n+2)^2} = \frac{n+1}{(n+3)(n+2)^2} \\ &\rightarrow \frac{1}{n^2}\,.\end{aligned}$$

$\Rightarrow$ Asymptotically ($n \rightarrow \infty$) the variance is the same for the two cases $x = 0$ and $x = n$ (just a question of symmetry)

Belief Vs ‘propension’

The main difficulty with probability is that since ever it has embedded two different meanings:

Belief Vs ‘propension’

The main difficulty with probability is that since ever it has embedded two different meanings:

- How much we belief something (‘degree of belief’ – original meaning of probability from its Latin root).

Belief Vs 'propension'

The main difficulty with probability is that since ever it has embedded two different meanings:

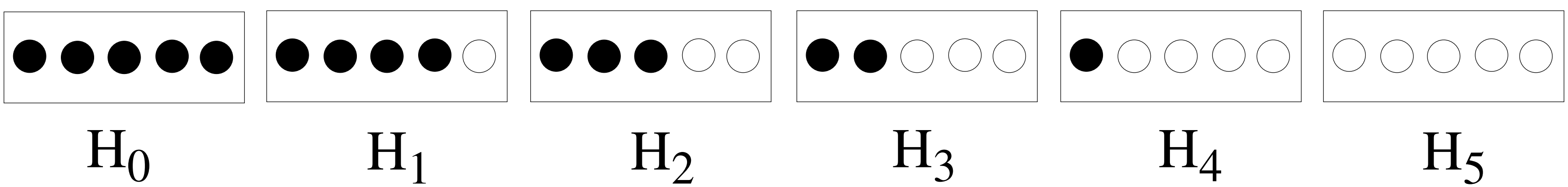
- How much we belief something ('degree of belief' – original meaning of probability from its Latin root).
- A property of a physical system to behave in a certain way ('chance' $\rightarrow$ 'propensity').

Belief Vs 'propension'

The main difficulty with probability is that since ever it has embedded two different meanings:

- How much we belief something ('degree of belief' – original meaning of probability from its Latin root).
- A property of a physical system to behave in a certain way ('chance' $\to$ 'propensity').

The six box model can help to make the question clear.



Degree of belief Vs ‘propension’

- There is no problem to interpret the proportion p of whate balls as a propensity of a box to yield white balls.

Degree of belief Vs 'propension'

- There is no problem to interpret the proportion p of whate balls as a propensity of a box to yield white balls.
- If we know p, this will be our belief to get a white ball (just because of equiprobability to pick up one ball at random):

$$P(W\,|\,p) = p$$

Degree of belief Vs 'propension'

- There is no problem to interpret the proportion p of whate balls as a propensity of a box to yield white balls.
- If we know p, this will be our belief to get a white ball (just because of equiprobability to pick up one ball at random):

$$P(W\,|\,p) = p$$

- If, under this assumption, we imagine a great number of trials, we expect a relative frequency of white equal to $P(W\,|\,p)$ [Bernoulli's Theorem]:

$$''\lim_{n\rightarrow\infty} f_n(W\,|\,p)'' = P(W\,|\,p) = p$$

Degree of belief Vs 'propension'

- There is no problem to interpret the proportion p of whate balls as a propensity of a box to yield white balls.
- If we know p, this will be our belief to get a white ball (just because of equiprobability to pick up one ball at random):

$$P(W\,|\,p) = p$$

- If, under this assumption, we imagine a great number of trials, we expect a relative frequency of white equal to $P(W\,|\,p)$ [Bernoulli's Theorem]:

$$'' \lim_{n\to\infty} f_n(W\,|\,p)'' = P(W\,|\,p) = p$$

There is no need to adhere to the frequentistic ideology to say this

Degree of belief Vs 'propension'

- There is no problem to interpret the proportion p of whate balls as a propensity of a box to yield white balls.
- If we know p, this will be our belief to get a white ball (just because of equiprobability to pick up one ball at random):

$$P(W\,|\,p) = p$$

- If, under this assumption, we imagine a great number of trials, we expect a relative frequency of white equal to $P(W\,|\,p)$ [Bernoulli's Theorem]:

$$''\lim_{n\to\infty} f_n(W\,|\,p)'' = P(W\,|\,p) = p$$

Instead, "probability is the limit of frequency for $n \to \infty$" is not more than an empty statement.

Beliefs about propensions

But the Laplacean ("Bayesian") approach is much more general and allows more possibilities, those which we naturally seek:

Beliefs about propensions

But the Laplacean ("Bayesian") approach is much more general and allows more possibilities, those which we naturally seek:

- Talking about $P(H_j)$ is the same as probability of propensity $p_j = j/5$.

Beliefs about propensions

But the Laplacean ("Bayesian") approach is much more general and allows more possibilities, those which we naturally seek:

- Talking about $P(H_j)$ is the same as probability of propensity $p_j = j/5$.
- But I remind that none had objection that initially the probability of white was $1/2$, although there was no box with propensity 50%!

Beliefs about propensions

But the Laplacean ("Bayesian") approach is much more general and allows more possibilities, those which we naturally seek:

- Talking about $P(H_j)$ is the same as probability of propensity $p_j = j/5$.
- But I remind that none had objection that initially the probability of white was $1/2$, although there was no box with propensity 50%!

⇒ Simple result of probability theory:

$$P(W\,|\,I) = \sum_j P(W\,|\,p_j, I) \cdot P(p_j\,|\,I)$$

Beliefs about propensions

But the Laplacean ("Bayesian") approach is much more general and allows more possibilities, those which we naturally seek:

- Talking about $P(H_j)$ is the same as probability of propensity $p_j = j/5$.
- But I remind that none had objection that initially the probability of white was $1/2$, although there was no box with propensity 50%!

⇒ Simple result of probability theory:

$$P(W \mid I) = \sum_j P(W \mid p_j, I) \cdot P(p_j \mid I)$$

Probability theory (in Laplage's sense) allows to attach probabilities to whatever we feel uncertain about!

Uncertainty about parameters

In other terms, in general, in physics (and in all sciences) p is a parameter of a model, like m in classical mechanics or M_H in the Standard Model.

Uncertainty about parameters

In other terms, in general, in physics (and in all sciences) p is a parameter of a model, like m in classical mechanics or M_H in the Standard Model.

- Defining p as the limit of the relative frequency is more or less the same as defining M_H as the value got by a great number of independent experiments ...

Uncertainty about parameters

In other terms, in general, in physics (and in all sciences) p is a parameter of a model, like m in classical mechanics or M_H in the Standard Model.

- Defining p as the limit of the relative frequency is more or less the same as defining M_H as the value got by a great number of independent experiments ...
  - something is the definition of a parameter in a mathematical model
  - something else is how to evaluate the parameter from real data

Uncertainty about parameters

In other terms, in general, in physics (and in all sciences) p is a parameter of a model, like m in classical mechanics or M_H in the Standard Model.

- Defining p as the limit of the relative frequency is more or less the same as defining M_H as the value got by a great number of independent experiments . . .

  The logically consistent way to estimate p comes from a theorem of probability theory, which in its simplest case leads to Laplace's rule of succession

$$E[p] = \frac{x+1}{n+2} \quad \longrightarrow \frac{x}{n}$$

Uncertainty about parameters

In other terms, in general, in physics (and in all sciences) p is a parameter of a model, like m in classical mechanics or M_H in the Standard Model.

- Defining p as the limit of the relative frequency is more or less the same as defining M_H as the value got by a great number of independent experiments . . .

  The logically consistent way to estimate p comes from a theorem of probability theory, which in its simplest case leads to Laplace's rule of succession

$$E[p] = \frac{x+1}{n+2} \quad \longrightarrow \frac{x}{n}$$

- Other important parameters are related to background, systematics, 'etc.' [arguments not covere here]

OK, ...but the priors?

Priors are an important ingredient of the framework:

OK, . . . but the priors?

Priors are an important ingredient of the framework:

- They are crucial in the Bayes theorem:
  - there is no other way to perform a probabilistic inference without passing through priors

    . . . although they can be often so vague to be ignored.

OK, . . . but the priors?

Priors are an important ingredient of the framework:

- They are crucial in the Bayes theorem:
  - there is no other way to perform a probabilistic inference without passing through priors

    . . . although they can be often so vague to be ignored.
- They allow us to use consistently all pieces of prior information. And we all have much prior information in our job!
  Only the perfect idiot hase no priors

OK, . . . but the priors?

Priors are an important ingredient of the framework:

- They are crucial in the Bayes theorem:
  - there is no other way to perform a probabilistic inference without passing through priors

    . . . although they can be often so vague to be ignored.
- They allow us to use consistently all pieces of prior information. And we all have much prior information in our job!
  Only the perfect idiot hase no priors
- Mistrust all prior-free methods that pretend to provide numbers that should mean how you have to be confident on something.

OK, ... but the priors?

Priors are an important ingredient of the framework:

- They are crucial in the Bayes theorem:
  - there is no other way to perform a probabilistic inference without passing through priors

    ... although they can be often so vague to be ignored.
- They allow us to use consistently all pieces of prior information. And we all have much prior information in our job!
  Only the perfect idiot hase no priors
- Mistrust all prior-free methods that pretend to provide numbers that should mean how you have to be confident on something.
  (Diffidate chi vi promette di far germogliar zecchini nel Campo dei Miracoli! – Collodi docet)

Conclusions

- The probabilistic framework basically set up by Laplace in his monumental work is healthy and grows up well (browse e.g. Amazon.com)

Conclusions

- The probabilistic framework basically set up by Laplace in his monumental work is healthy and grows up well (browse e.g. Amazon.com)
- It is very close to the natural way of reasoning of physicists (as everybody else).

Conclusions

- The probabilistic framework basically set up by Laplace in his monumental work is healthy and grows up well (browse e.g. Amazon.com)
- It is very close to the natural way of reasoning of physicists (as everybody else).
- Its consistent application in small-complex problems was prohibitive many years ago.

Conclusions

- The probabilistic framework basically set up by Laplace in his monumental work is healthy and grows up well (browse e.g. Amazon.com)
- It is very close to the natural way of reasoning of physicists (as everybody else).
- Its consistent application in small-complex problems was prohibitive many years ago.
- But it is now possible thank to progresses in applied mathematics and computation.

Conclusions

- The probabilistic framework basically set up by Laplace in his monumental work is healthy and grows up well (browse e.g. Amazon.com)
- It is very close to the natural way of reasoning of physicists (as everybody else).
- Its consistent application in small-complex problems was prohibitive many years ago.
- But it is now possible thank to progresses in applied mathematics and computation.
- It makes little sense to stick to old 'ah hoc' methods that had their *raison d'être* in the computational barrier.

Conclusions

- The probabilistic framework basically set up by Laplace in his monumental work is healthy and grows up well (browse e.g. Amazon.com)
- It is very close to the natural way of reasoning of physicists (as everybody else).
- Its consistent application in small-complex problems was prohibitive many years ago.
- But it is now possible thank to progresses in applied mathematics and computation.
- It makes little sense to stick to old 'ah hoc' methods that had their *raison d'être* in the computational barrier.
- Mistrust all results that sound as 'confidence', 'probability' etc about physics quantities, if they are obtained by methods that do not contemplate 'beliefs'.

. . . postponed preamble

"The celebrated Monsieur Leibnitz has observed it to be a defect in the common systems of logic, that they are very copious when they explain the operations of the understanding in the forming of demonstrations, but are too concise when they treat of probabilities, and those other measures of evidence on which life and action entirely depend, and which are our guides even in most of our philosophical speculations."

(David Hume)

. . . postponed preamble

"The celebrated Monsieur Leibnitz has observed it to be a defect in the common systems of logic, that they are very copious when they explain the operations of the understanding in the forming of demonstrations, but are too concise when they treat of probabilities, and those other measures of evidence on which life and action entirely depend, and which are our guides even in most of our philosophical speculations."

(David Hume)

$\Rightarrow$ still very true after $\approx$ 300 years!

. . . postponed preamble

And, by the way, for those who

- are not familiar with David Hume
- and/or think that lucid, deep philosofical thinking has little/nothing to do with Physics

. . . postponed preamble

And, by the way, for those who

- are not familiar with David Hume
- and/or think that lucid, deep philosofical thinking has little/nothing to do with Physics

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

. . .

[And, in a different writing,]

. . .

"It is to the immortal credit of D. Hume and E. Mach that they, above all others, introduced this critical conception."

(Albert Einstein)