Which ball? Which box? Where did the ball stop?

— A playful introduction to Bayesian reasoning —

Giulio D'Agostini

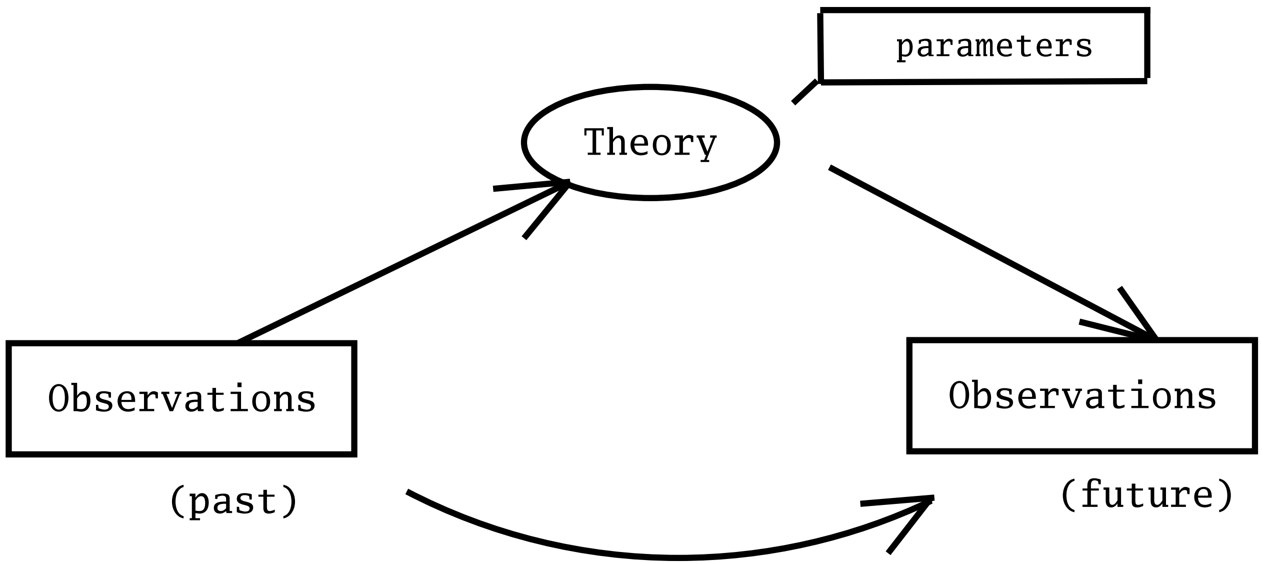
giulio.dagostini@roma1.infn.it

Dipartimento di Fisica

Università di Roma La Sapienza

"Probability is good sense reduced to a calculus" (Laplace)

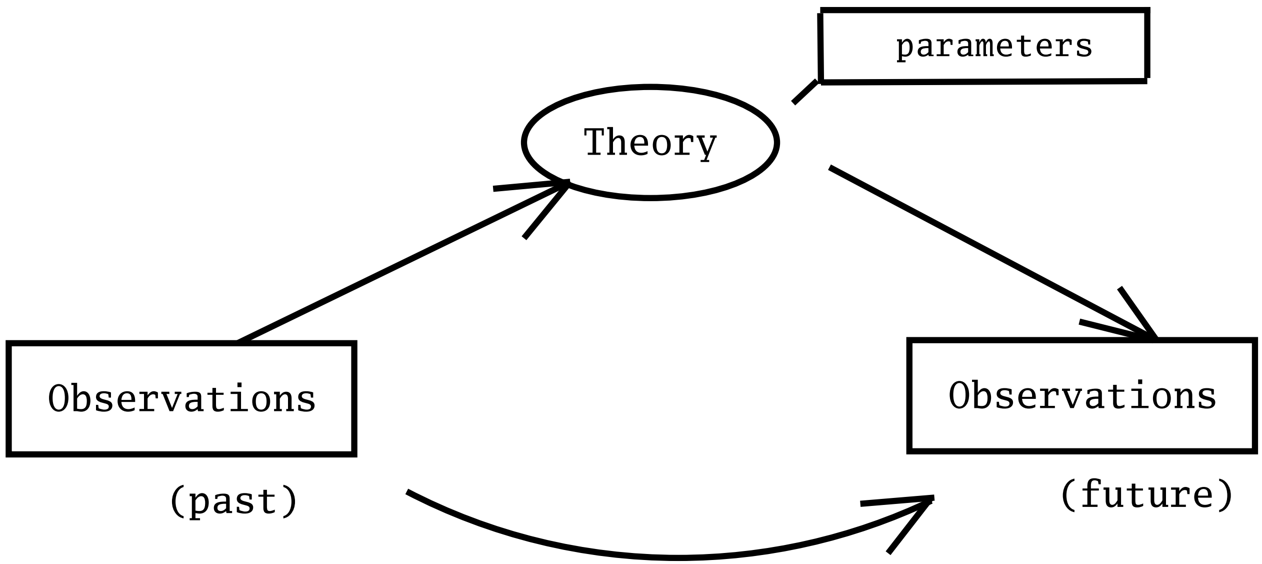
Doing Physics [Science in general]



Task of physicists:

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

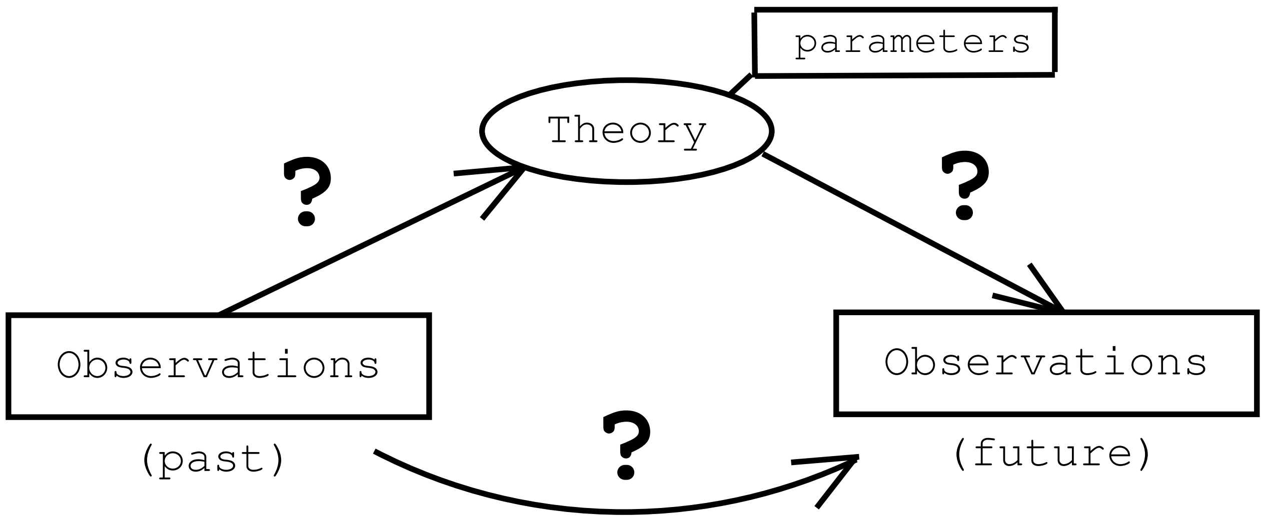
Doing Physics [Science in general]



Process

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

Doing Physics [Science in general]

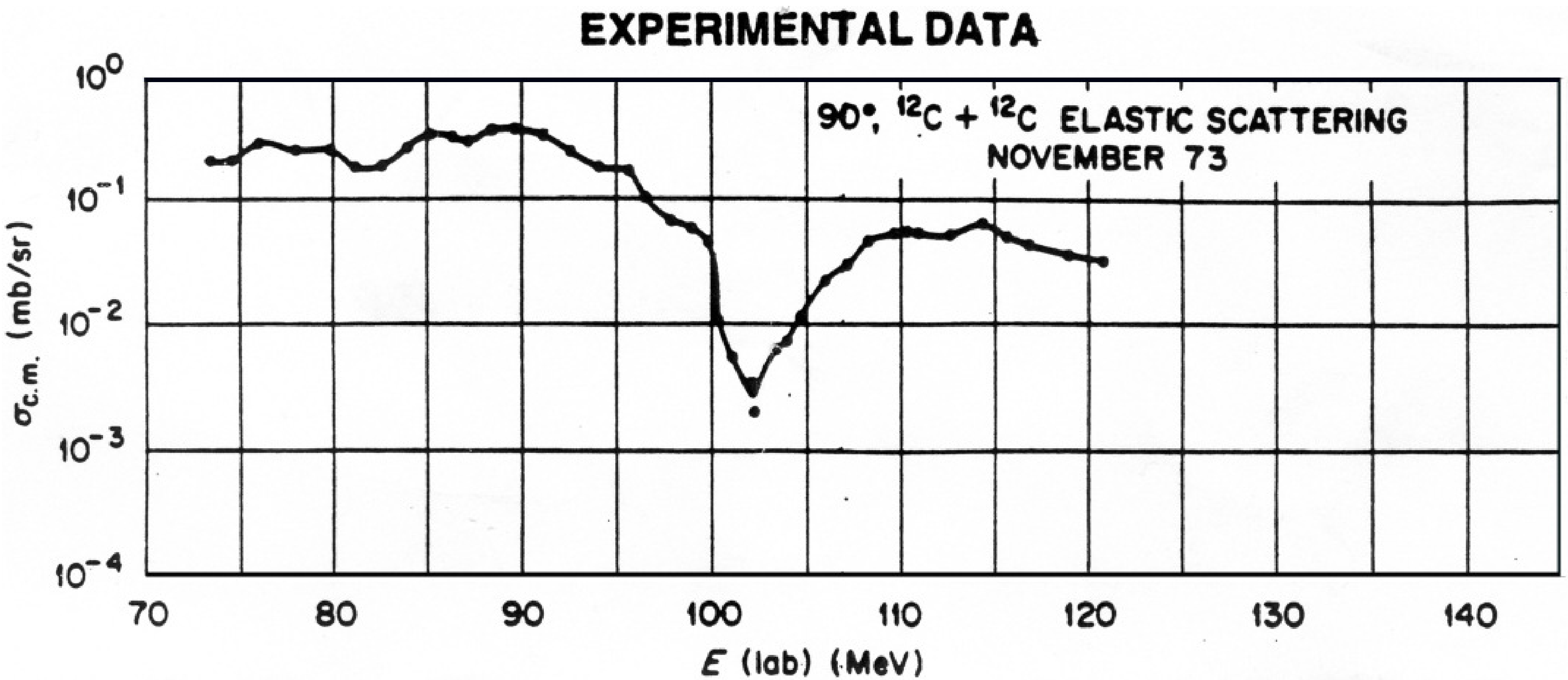


⇒ Uncertainty:

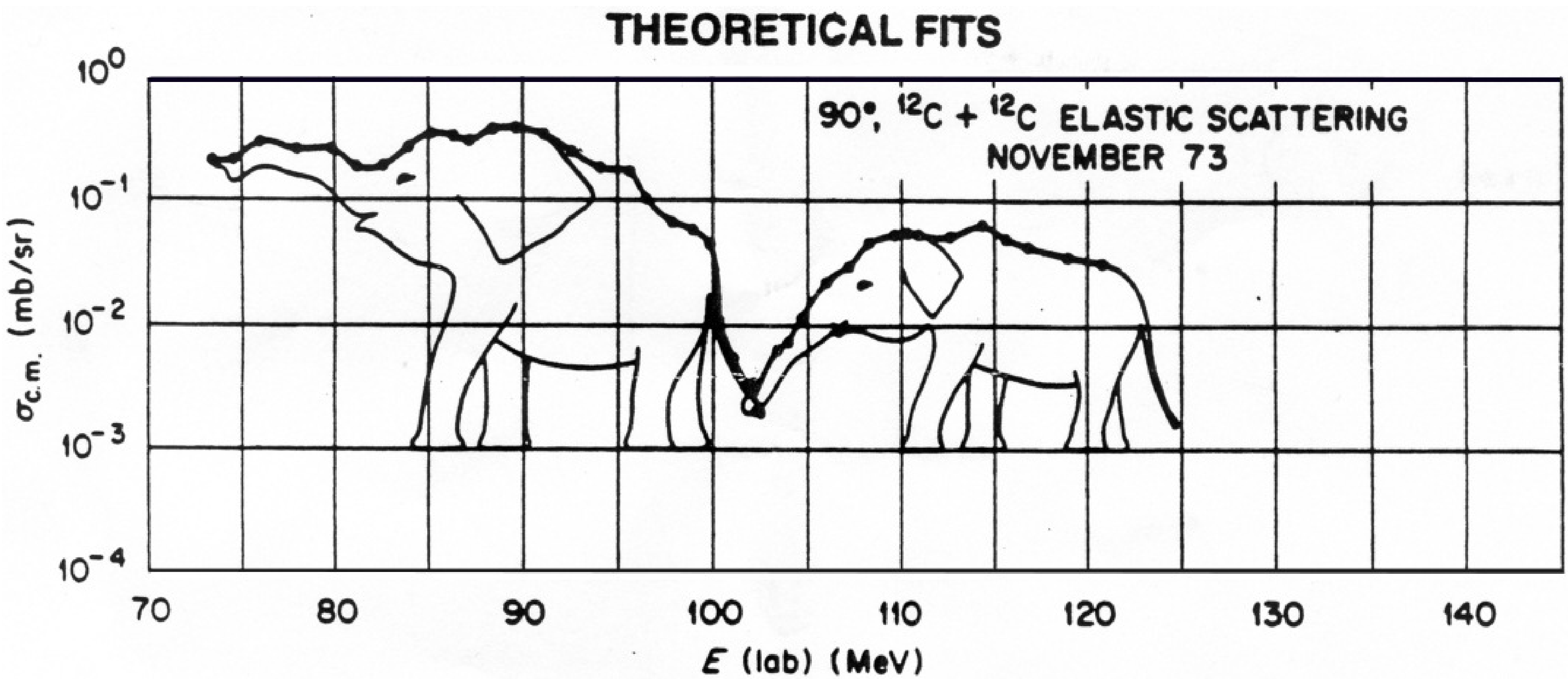
1. Given the past observations, in general we are not sure about the theory parameters (and/or the theory itself)

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.

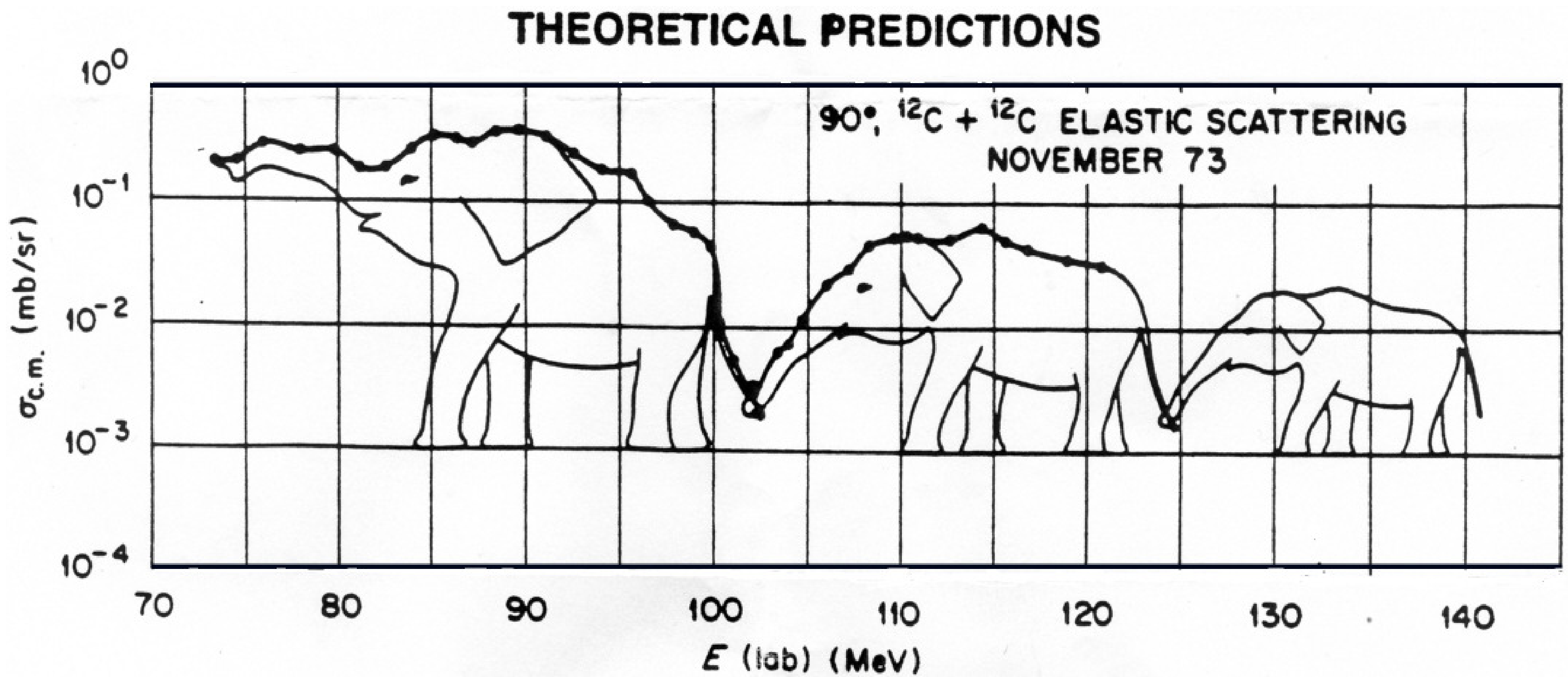
Inferential-predictive process



Inferential-predictive process

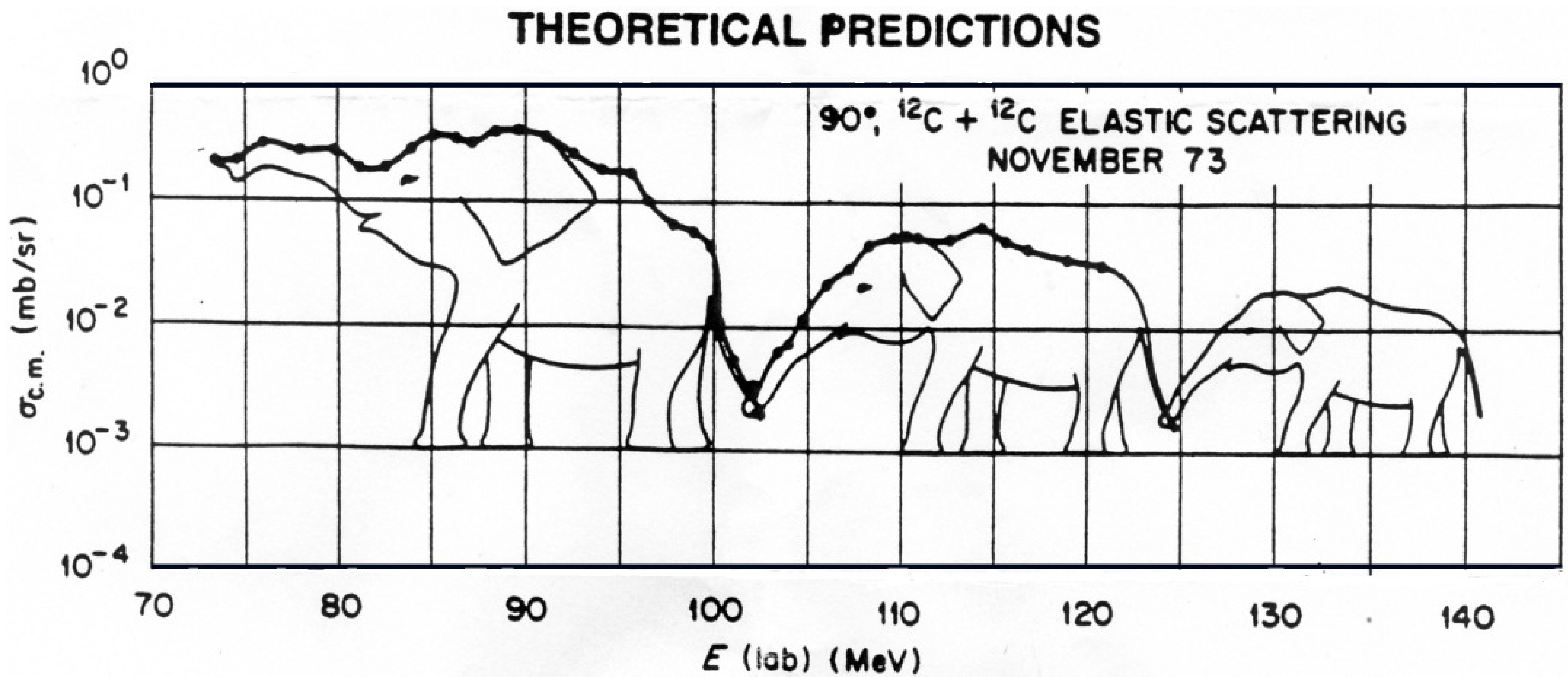


Inferential-predictive process



(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!]

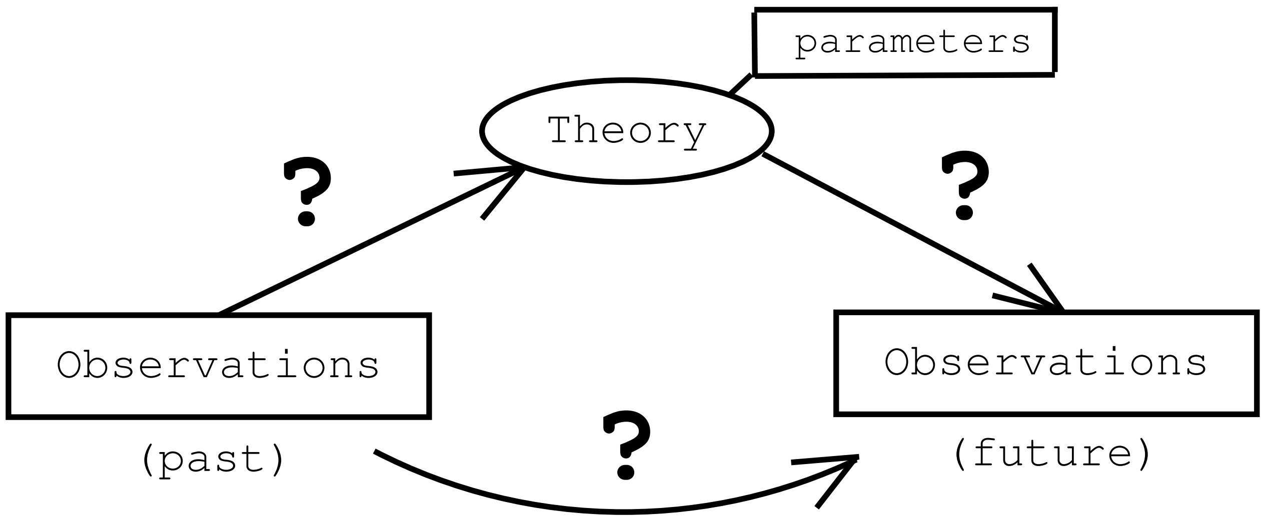
2011 IgNobel prize in Mathematics

- 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"

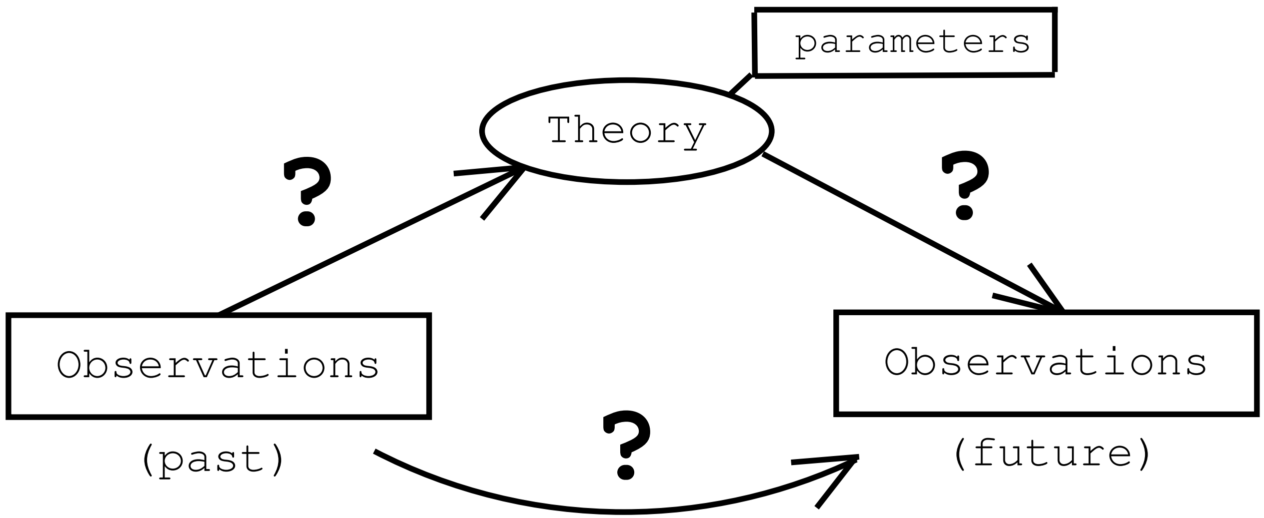
Deep source of uncertainty



Uncertainty:

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

Deep source of uncertainty



Uncertainty:

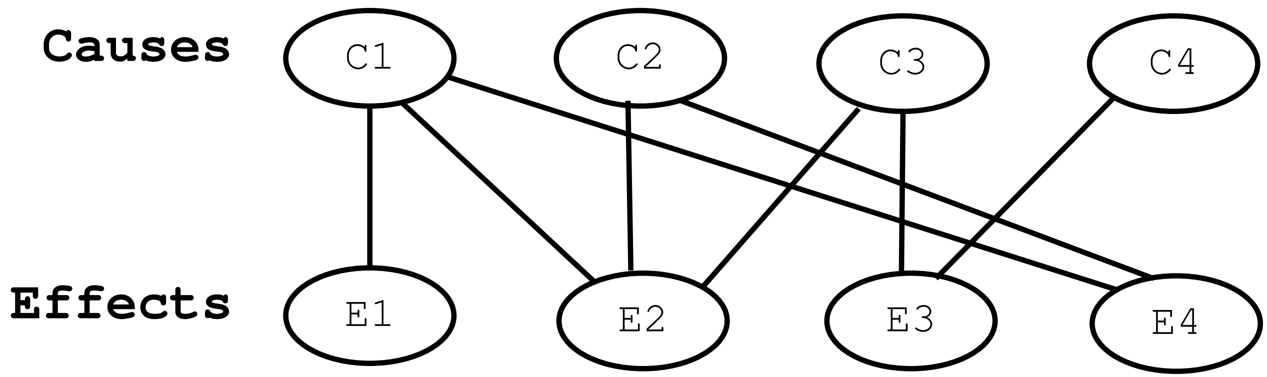
Theory	— ? ⟶	Future observations
Past observations	— ? ⟶	Theory
Theory	— ? ⟶	Future observations

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

CAUSE $\Longleftrightarrow$ **EFFECT**

Causes → effects

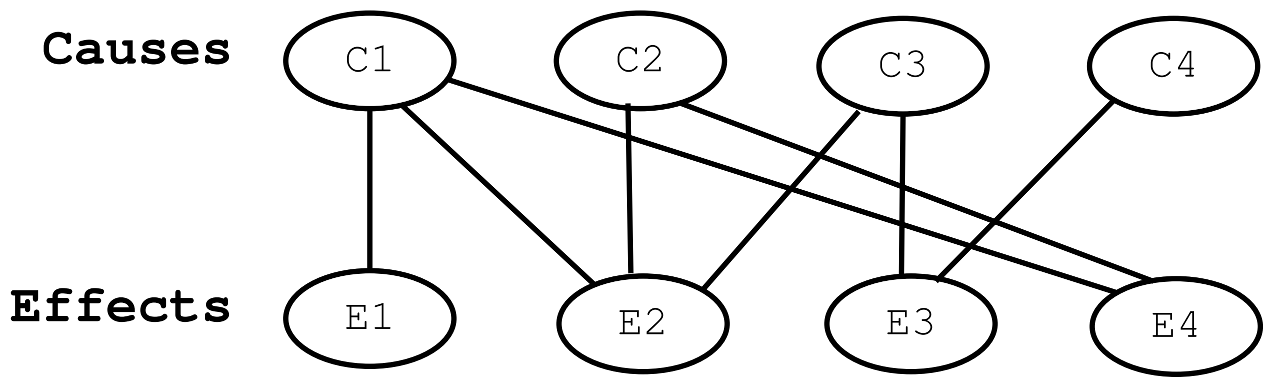
The same *apparent* cause might produce several,different effects



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

Causes → effects

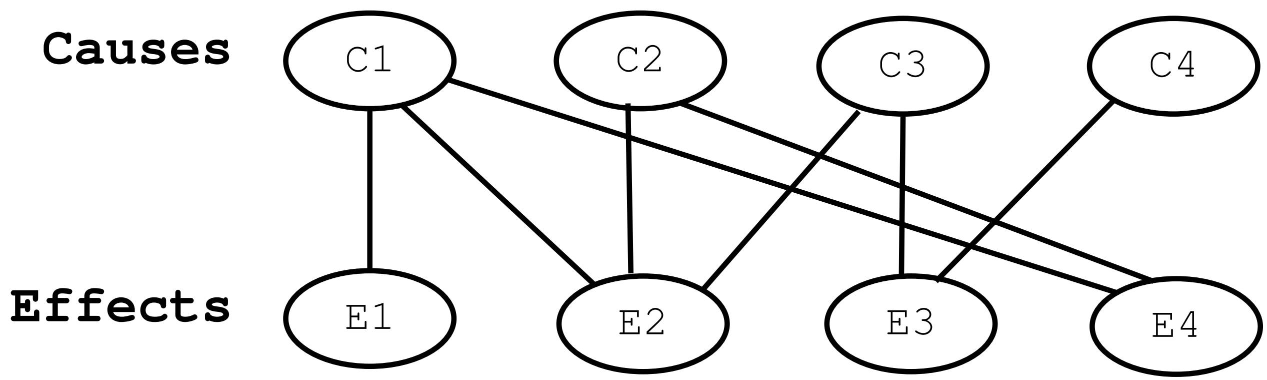
The same *apparent* cause might produce several,different effects



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 for their scientific applications.

The "essential problem" of the Sciences

"Now, these problems are classified as *probability of causes*, and are most interesting of all for 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 for 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 for 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

- $P(-10 < \epsilon'/\epsilon \times 10^4 < 50) >> P(\epsilon'/\epsilon \times 10^4 > 100)$
- $P(172 \le m_{top}/\text{GeV} \le 174) \approx 70\%$
- $P(M_H < 125\,\text{GeV}) > P(M_H > 125\,\text{GeV})$

Uncertainty and probability

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

- $P(-10 < \epsilon'/\epsilon \times 10^4 < 50) >> P(\epsilon'/\epsilon \times 10^4 > 100)$
- $P(172 \le m_{top}/\text{GeV} \le 174) \approx 70\%$
- $P(M_H < 125\,\text{GeV}) > P(M_H > 125\,\text{GeV})$

…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

- $P(-10 < \epsilon'/\epsilon \times 10^4 < 50) >> P(\epsilon'/\epsilon \times 10^4 > 100)$
- $P(172 \le m_{top}/\text{GeV} \le 174) \approx 70\%$
- $P(M_H < 125\,\text{GeV}) > P(M_H > 125\,\text{GeV})$

…although, such statements are considered
blaspheme to statistics gurus

[The fact that for several 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)

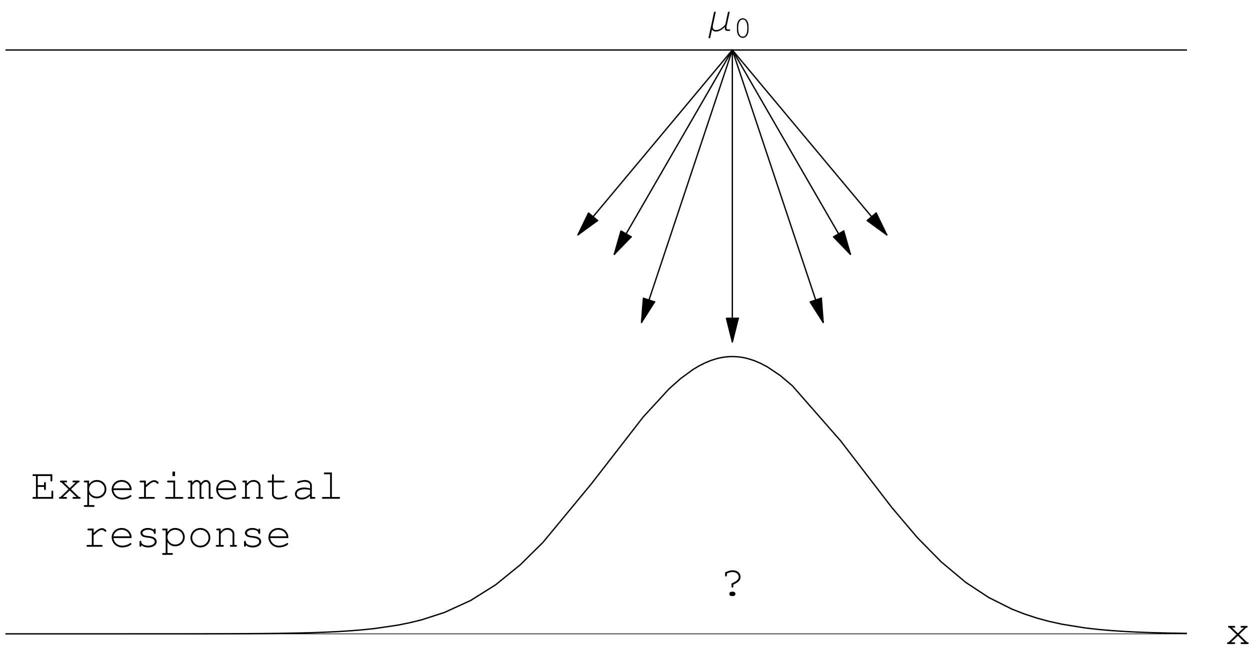
Doing Science in conditions of uncertainty

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

Indeed

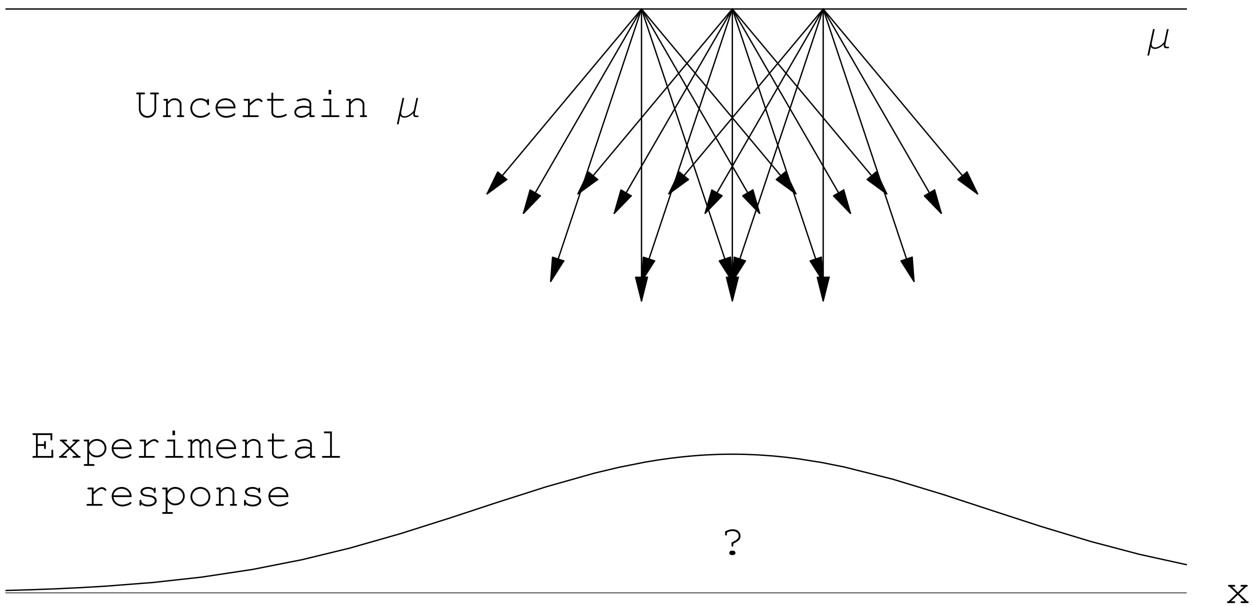
"It is scientific only to say what is more likely and what is less likely" (Feynman)

From 'true value' to observations



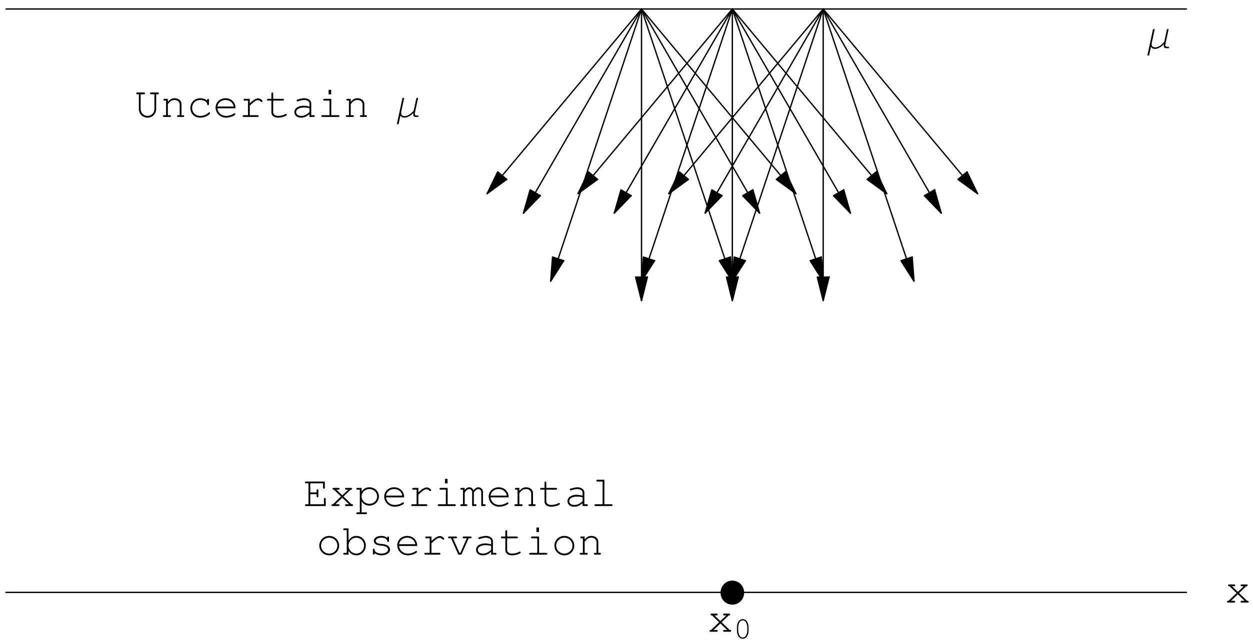
Given μ (exactly known) we are uncertain about x

From ‘true value’ to observations



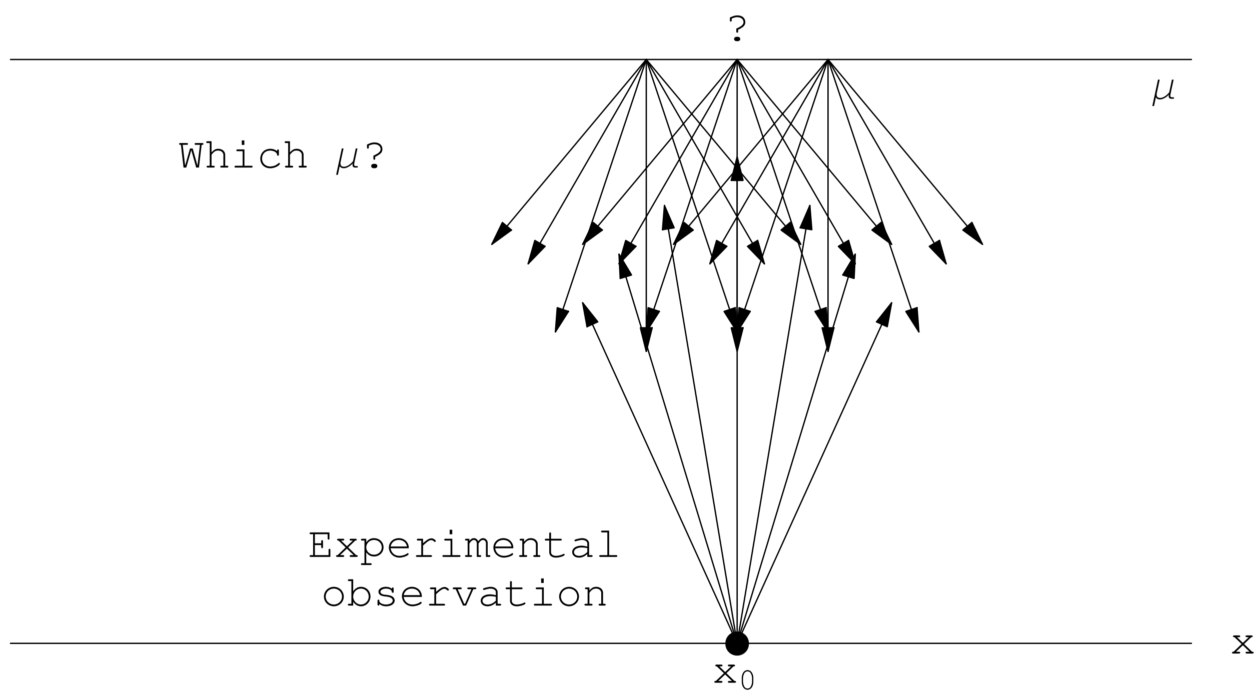
Uncertainty about μ makes us more uncertain about x

. . . and back: Inferring a true value



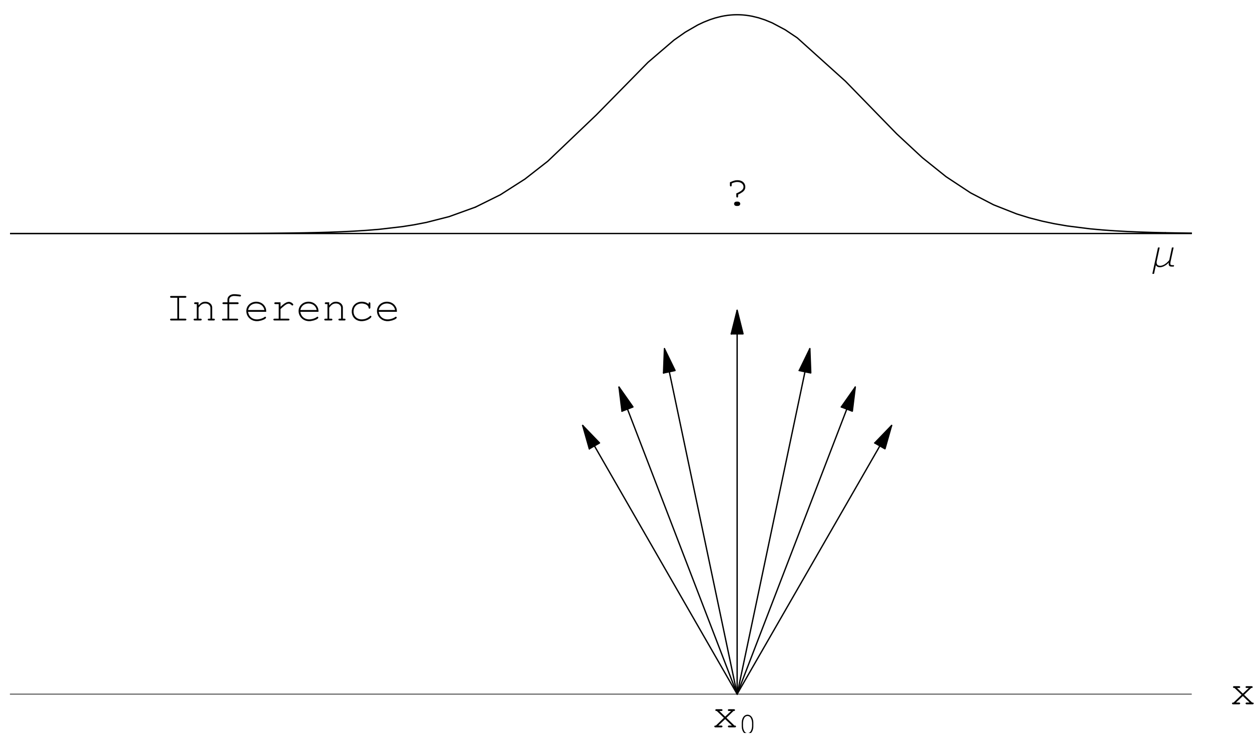
The observed data is certain: $\rightarrow$ ‘true value’ uncertain.

. . . 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.

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:

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

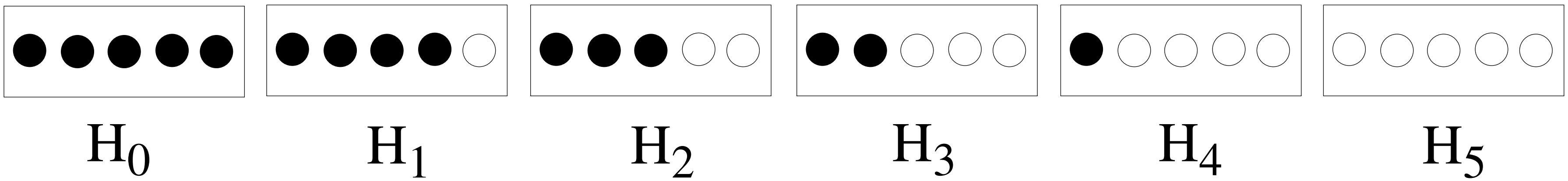
- x is binary:

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

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

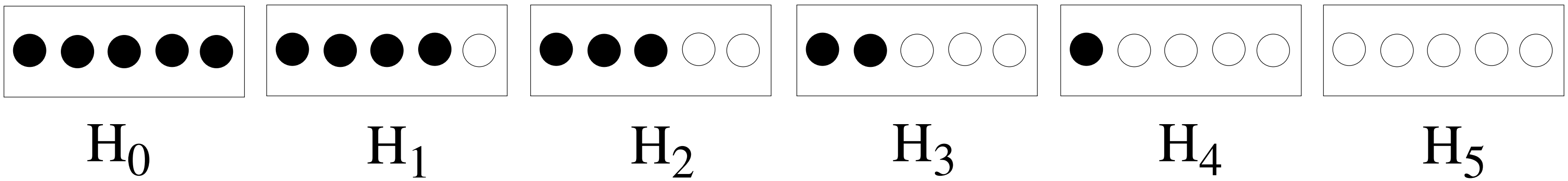
$\Rightarrow$ Later we shall make μ continous.

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:

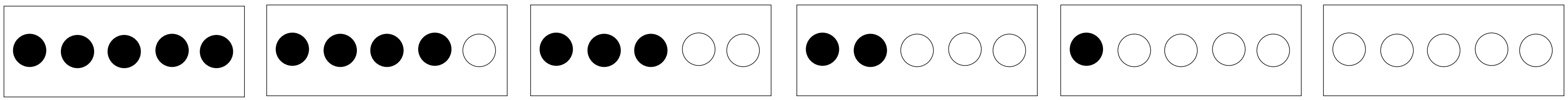
(a) Which box have we chosen, $H_0, H_1, \ldots, H_5$?

(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:

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

Which box? Which ball?

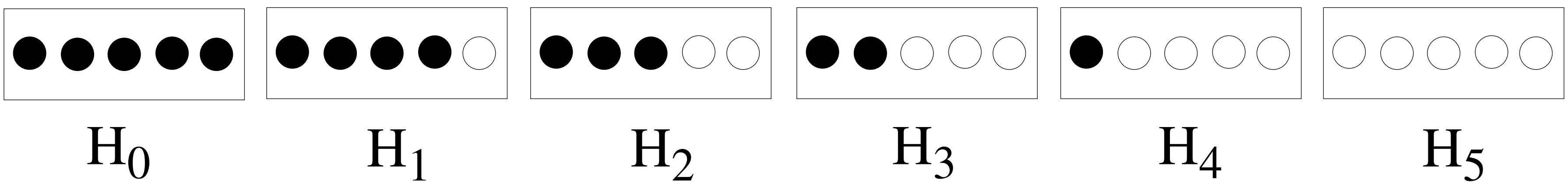


H_0 H_1 H_2 H_3 H_4 H_5

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

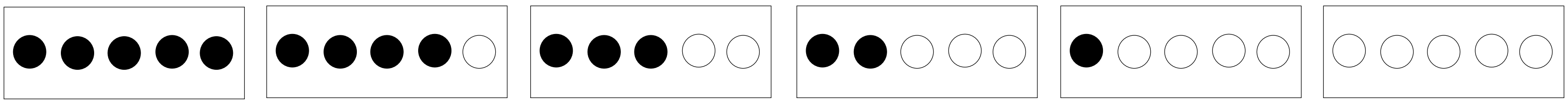
Which box? Which ball?



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?



H_0 H_1 H_2 H_3 H_4 H_5

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 toy inferential experiment

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

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

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

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

The rule of the game is that we are not allowed to watch inside the box! (As we cannot whatch a real movie showing the beginning of the Universe and compare it with our speculations.)

Where <u>is</u> probability?

We all agree that the experimental results change

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

Where *<u>is</u>* **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 <u>is</u>* 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 not *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 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"

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)

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)

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

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

$\Rightarrow$ Three box game

(Box with white ball wins)

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?
    - Coherent 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.

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 <u>does not mean</u> 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)

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

Beliefs and ‘coherent’ bets

Remarks:

- Subjective <u>does not mean</u> 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.

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

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

NO!

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.

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

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

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

For more on the subject
see `http://arxiv.org/abs/1112.3620`
and references therein.

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.
  - $P(\text{Rain next Saturday in Paris}) = 68\%$
  - $P(\text{Germany wins European soccer cup}) = 68\%$
  - $P(M_H \leq 130\,\text{GeV}) = 68\%$
  - $P(\text{Free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$

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.
  - $P(\text{Rain next Saturday in Paris}) = 68\%$
  - $P(\text{Germany wins European soccer cup}) = 68\%$
  - $P(M_H \leq 130\,\text{GeV}) = 68\%$
  - $P(\text{Free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$

  They all convey unambiguously the same confidence on something.

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.
  - $P(\text{Rain next Saturday in Paris}) = 68\%$
  - $P(\text{Germany wins European soccer cup}) = 68\%$
  - $P(M_H \leq 130\,\text{GeV}) = 68\%$
  - $P(\text{Free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$
- 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.
  - $P(\text{Rain next Saturday in Paris}) = 68\%$
  - $P(\text{Germany wins European soccer cup}) = 68\%$
  - $P(M_H \leq 130\,\text{GeV}) = 68\%$
  - $P(\text{Free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$
- You might agree or disagree, but at least You know what this person has in his mind. (NOT TRUE with "C.L.'s"!)
- 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.
  - $P(\text{Rain next Saturday in Paris}) = 68\%$
  - $P(\text{Germany wins European soccer cup}) = 68\%$
  - $P(M_H \leq 130\,\text{GeV}) = 68\%$
  - $P(\text{Free neutron decays before 17 s}) = 68\%$
  - $P(\text{White ball from a box with 68W+32B}) = 68\%$

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

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

Remember that probability is always conditional probability!

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

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

Basic rules of probability

1. $0 \leq P(A \,|\, I) \leq 1$
2. $P(\Omega \,|\, I) = 1$
3. $P(A \cup B \,|\, I) = P(A \,|\, I) + P(B \,|\, I)$ [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 of '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!

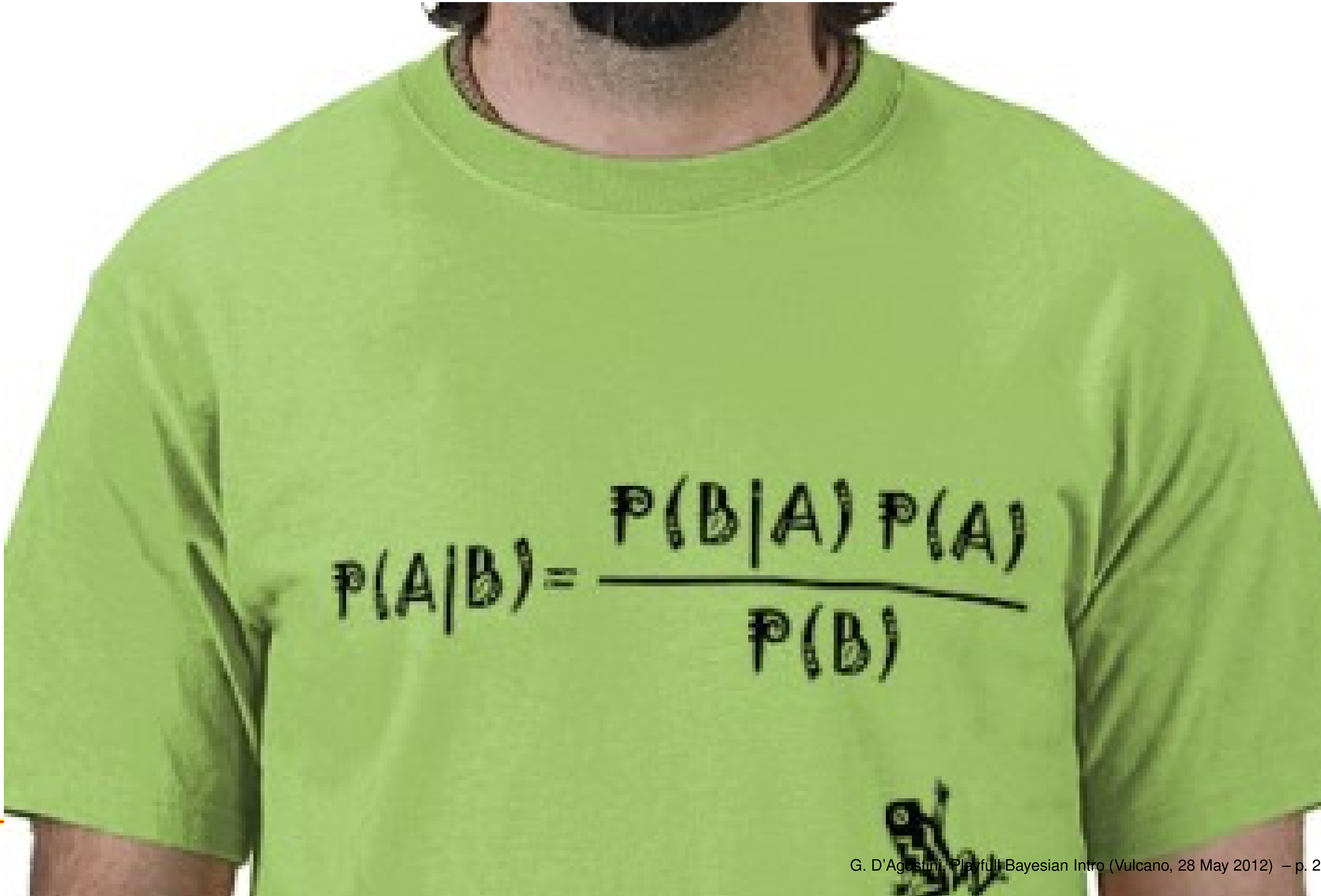
Mathematics of beliefs

An even better news:

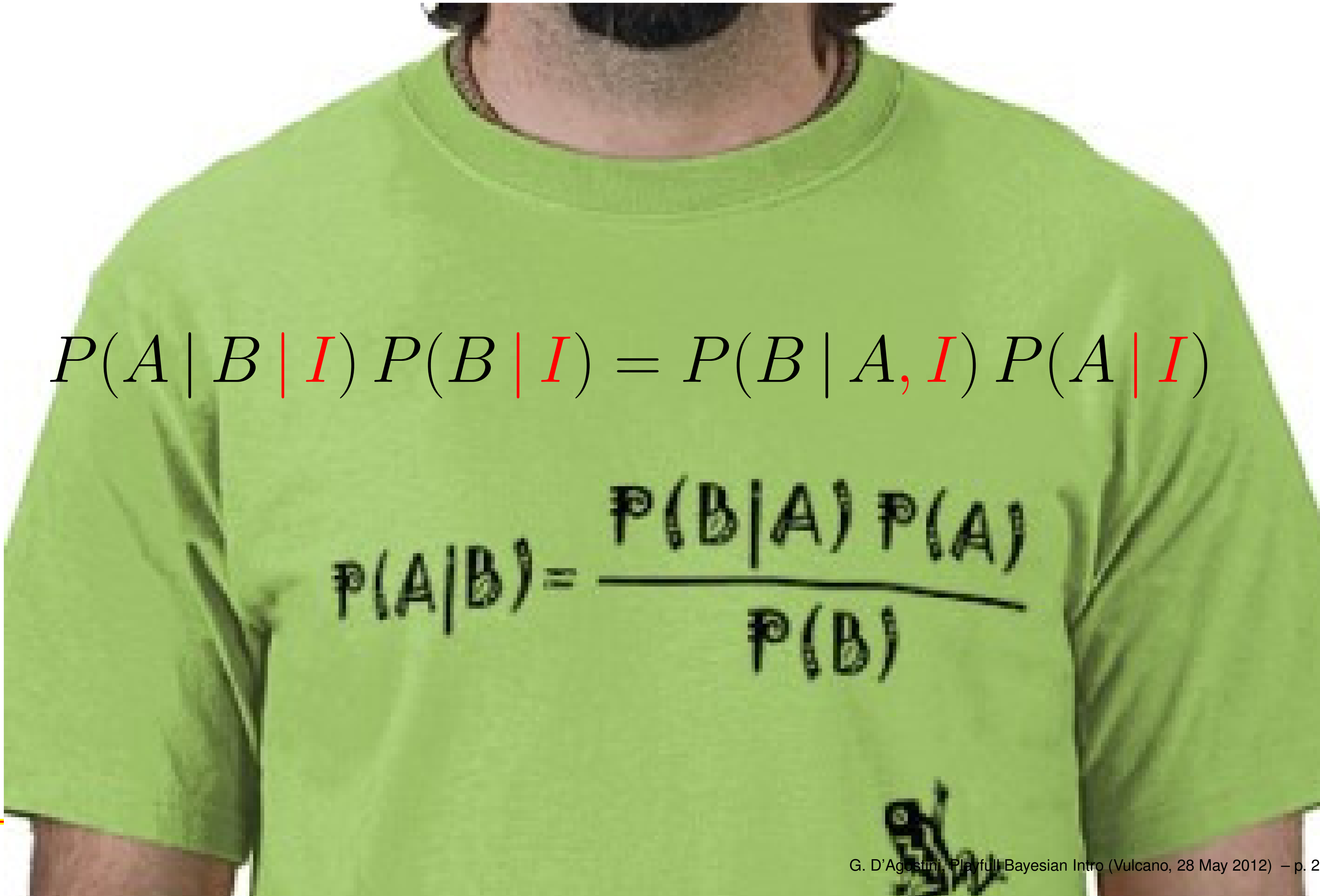
The fourth basic rule
can be fully exploided!

(Liberated by a curious ideology that forbits its use)

A simple, powerful formula

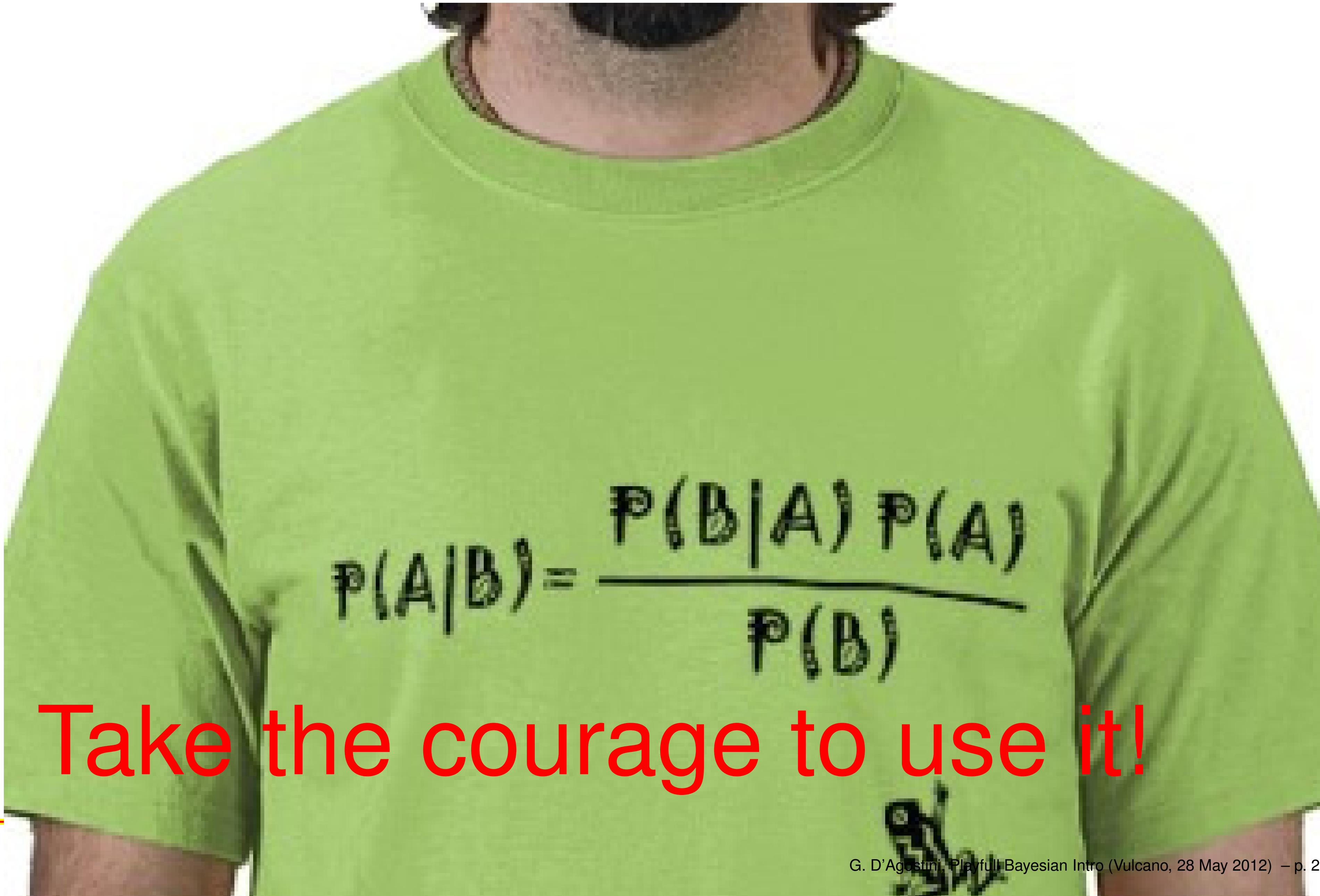


A simple, powerful formula



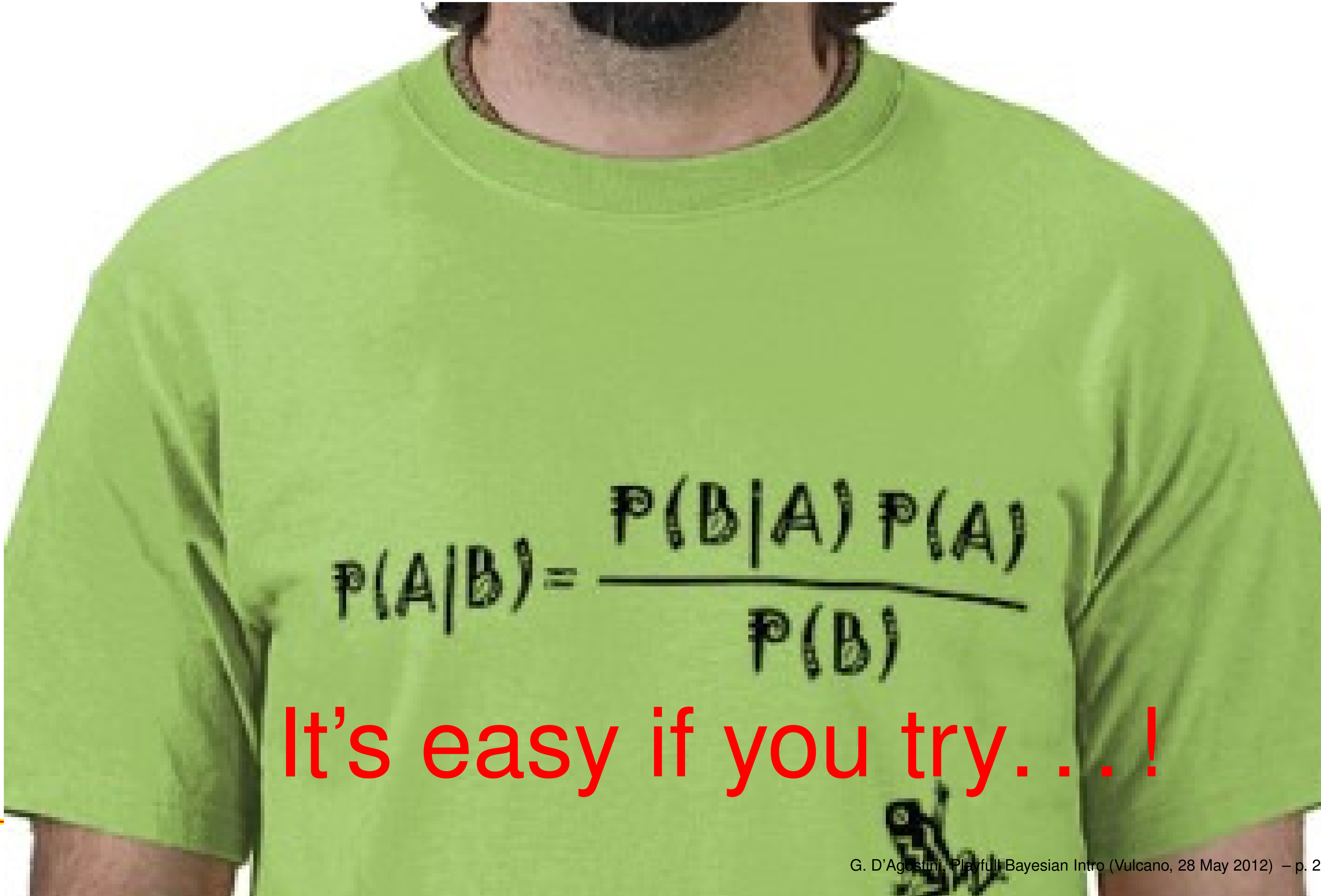
$$P(A \,|\, B \,|\, I)\, P(B \,|\, I) = P(B \,|\, A, I)\, P(A \,|\, I)$$

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'.

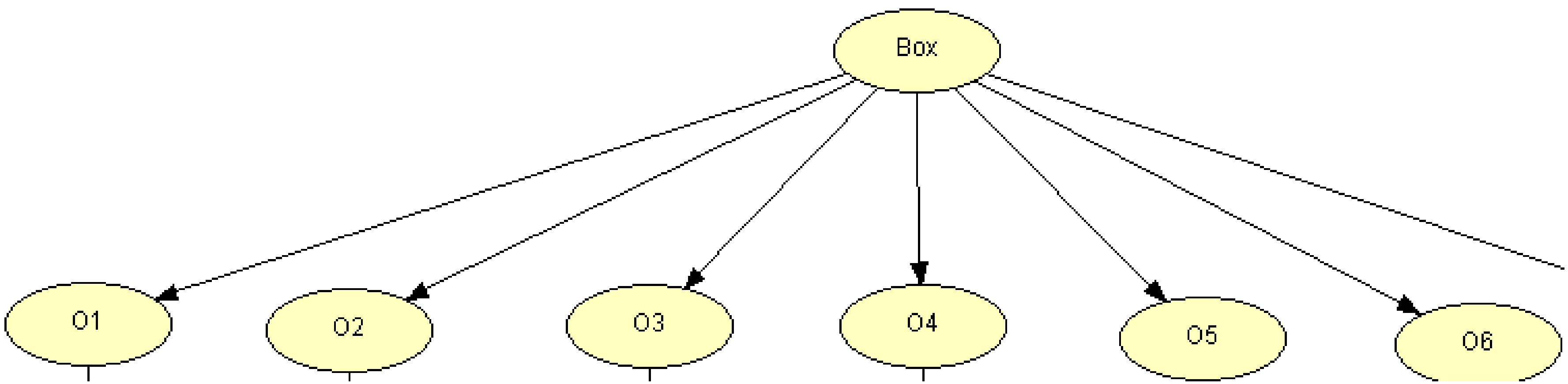
Note: denominator is just a normalization factor.

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

Most convenient way to remember Bayes theorem

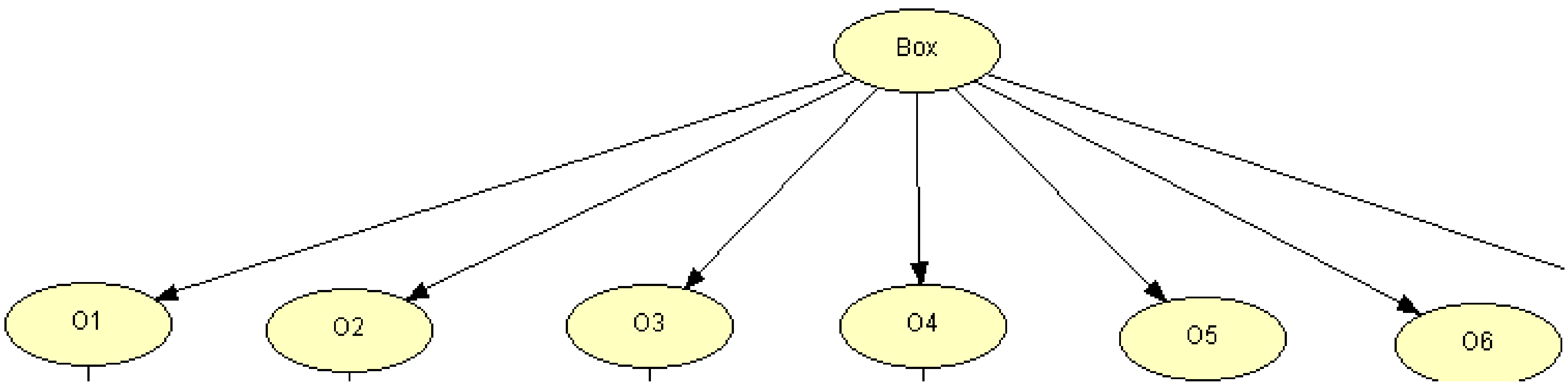
Cause-effect representation

box content → observed color



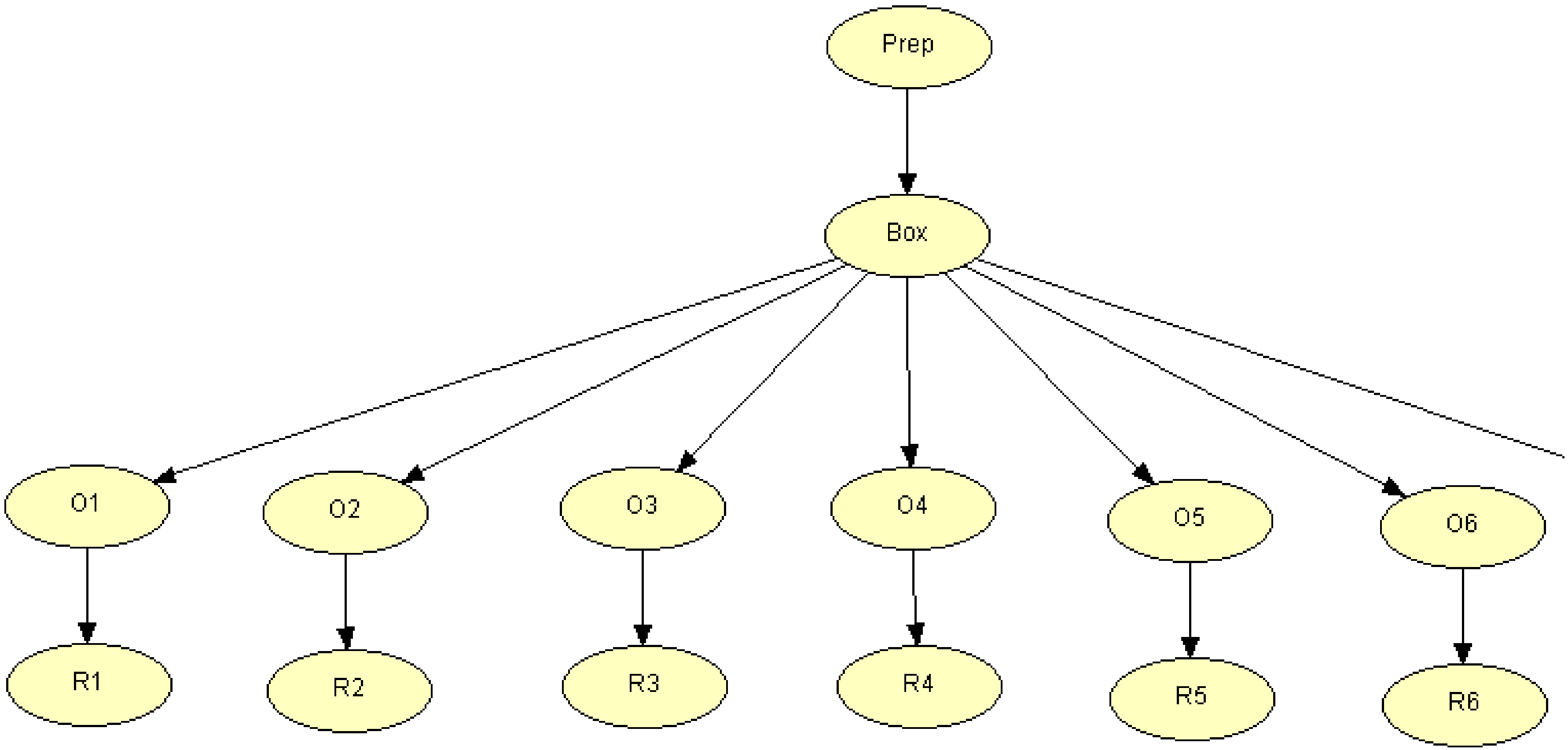
Cause-effect representation

box content $\rightarrow$ observed color

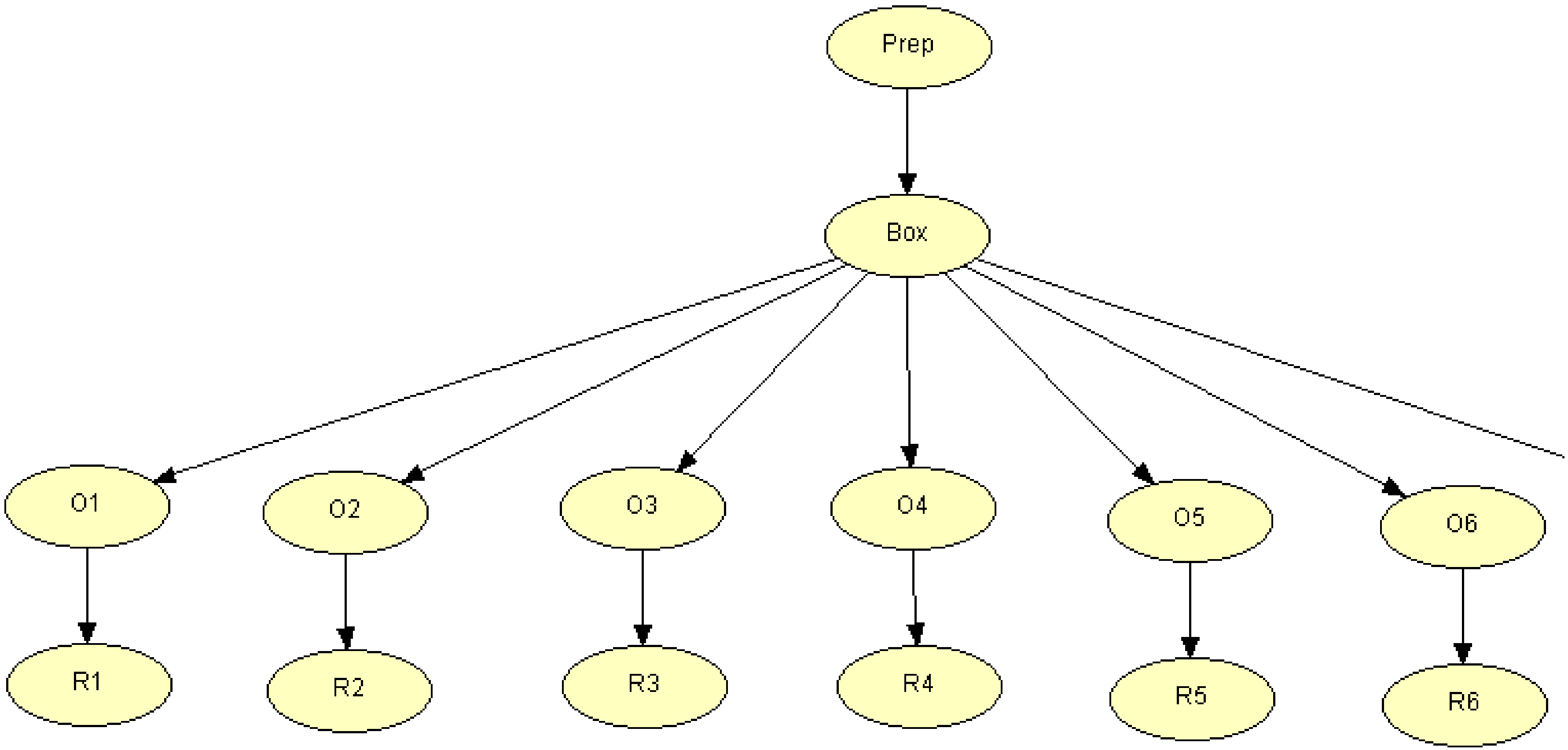


An effect might be the cause of another effect $\Longrightarrow$

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

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

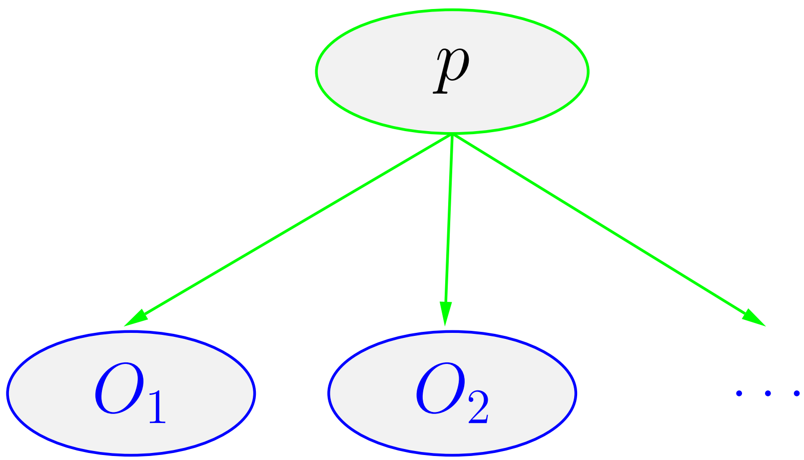
$\Rightarrow$ p continous in $[0, 1]$

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

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

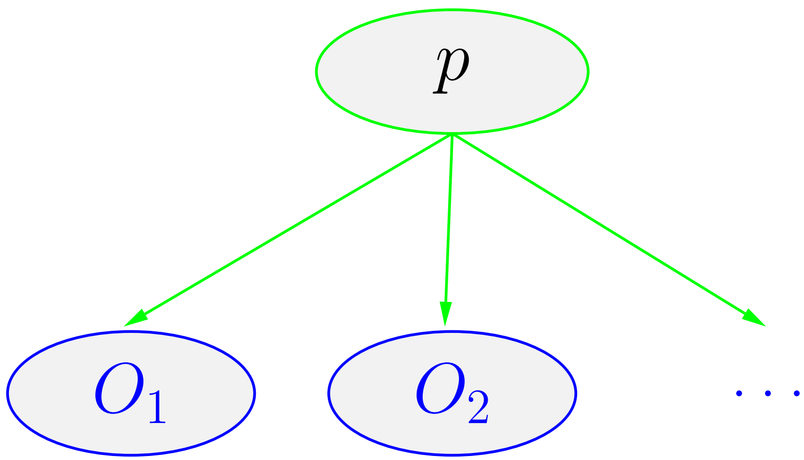
Inferring Bernoulli's trial parameter p

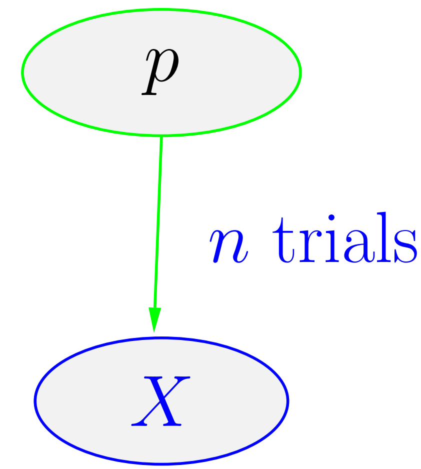
Making several independent trials *assuming* the same p



Inferring Bernoulli's trial parameter p

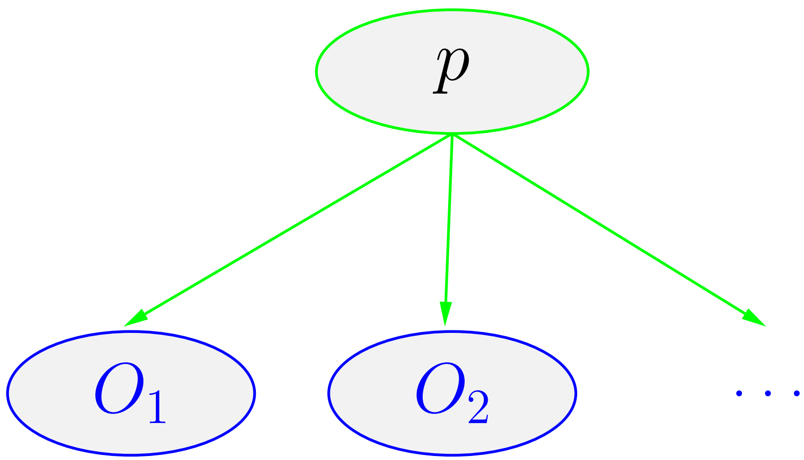
Making several independent trials *assuming* the same p



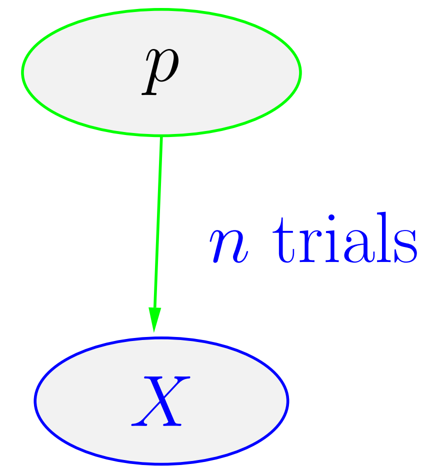


Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p



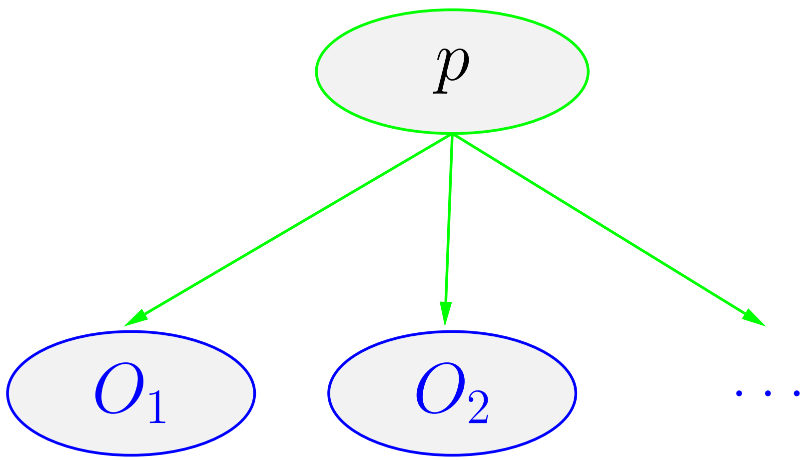
"independent Bernoulli trials"



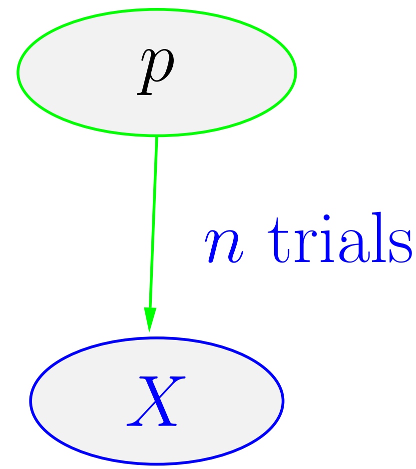
"binomial distribution"

Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p



"independent Bernoulli trials"

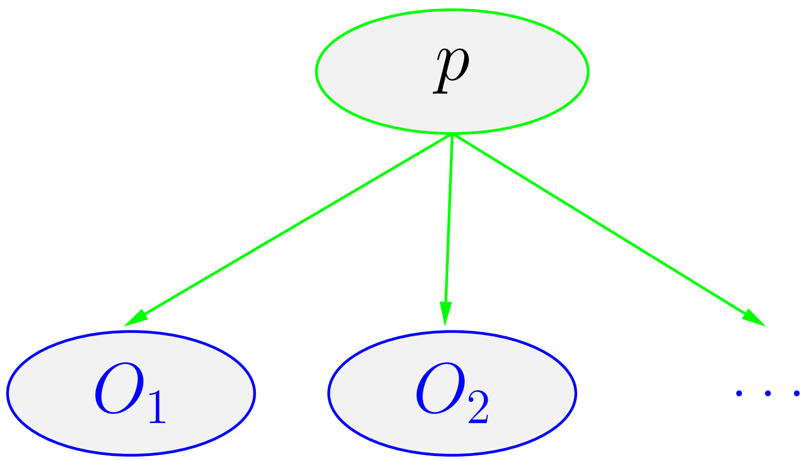


"binomial distribution"

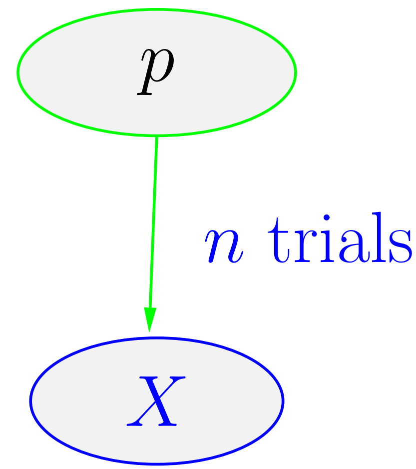
$\Rightarrow$ In the light of the experimental information there will be values of p we shall believe more, and others we shall believe less.

Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p



"independent Bernoulli trials"



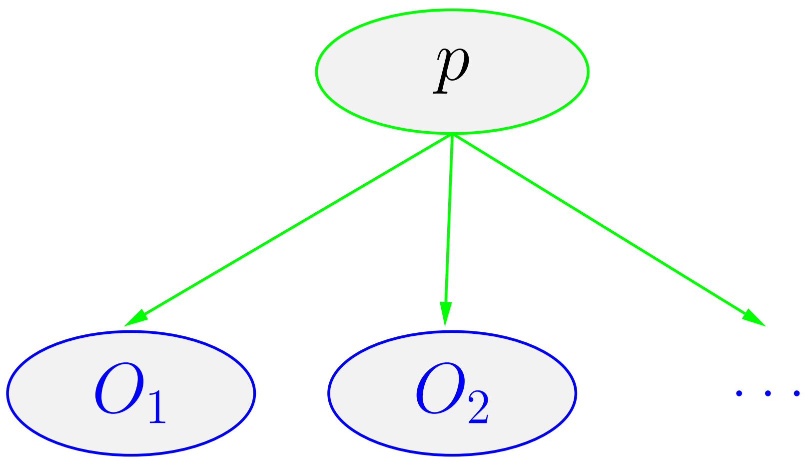
"binomial distribution"

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

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

Inferring Bernoulli's trial parameter p

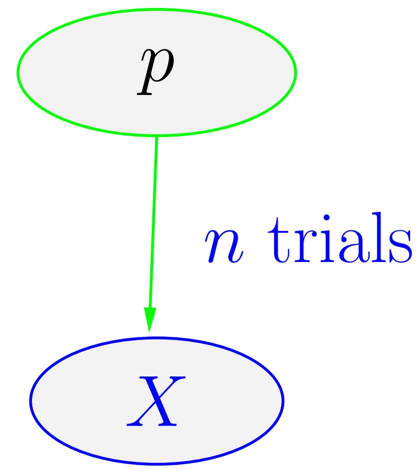
Making several independent trials *assuming* the same p



"independent Bernoulli trials"

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

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



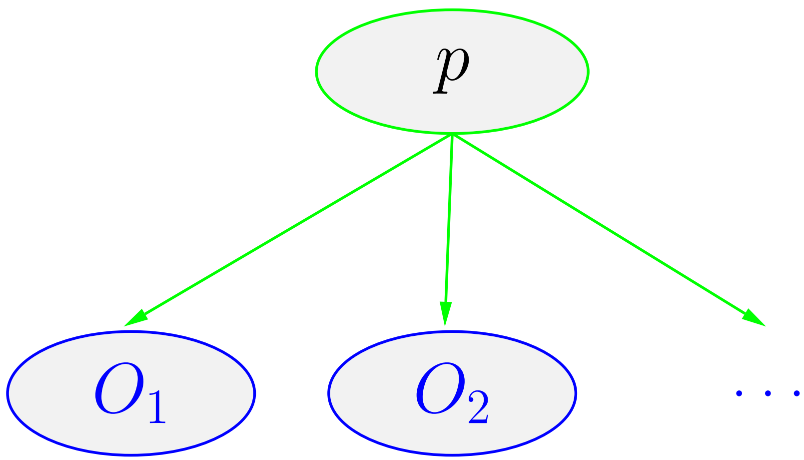
"binomial distribution"

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

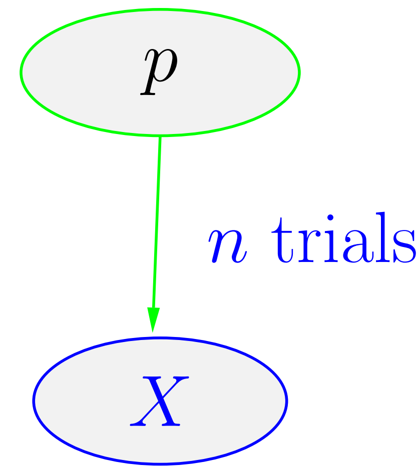
$$\propto f(X \,|\, n, p) \cdot f_0(p)$$

Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p



"independent Bernoulli trials"



"binomial distribution"

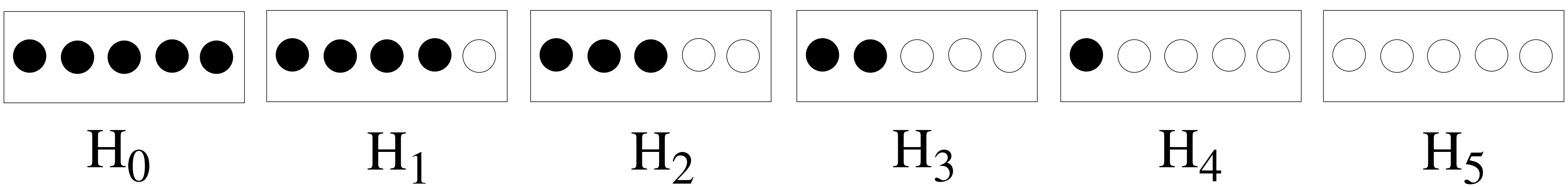
$$P(p_i \,|\, O_1, O_2, \ldots) \qquad\qquad P(p_i \,|\, X, n)$$

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

Are the two inferences the same?

(not obvious in principle)

Application to the six box problem



Remind:

- E_1 = White
- E_2 = Black

Collecting the pieces of information we need

Our tool:

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

Collecting the pieces of information we need

Our tool:

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

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

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.
→ 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.
We can rewrite it as
$P(E_i \,|\, I) = \sum_j P(E_i \,|\, H_j, I) \cdot P(H_j \,|\, I)$

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)

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, then
  - $l/L \leftrightarrow p$ 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, then
  - $l/L \leftrightarrow p$ 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, then
  - $l/L \leftrightarrow p$ 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, then
  - $l/L \leftrightarrow p$ 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, then
  - $l/L \leftrightarrow p$ 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, then
  - $l/L \leftrightarrow p$ 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, then
  - $l/L \leftrightarrow p$ of binomial (Bernoulli trials)

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

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

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 phycisists (as well as of 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 phycisists (as well as of 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 phycisists (as well as of 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 phycisists (as well as of 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 phycisists (as well as of 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’.