Introduction to Probabilistic Reasoning

– inference, forecasting, decision –

Giulio D'Agostini

giulio.dagostini@roma1.infn.it

Dipartimento di Fisica

Università di Roma La Sapienza

– Part 2 –

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

Probability

What is probability?

Standard textbook definitions

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

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

Standard textbook definitions

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

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

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

Standard textbook definitions

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

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

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

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

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

Standard textbook definitions

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

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

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

Future $\Leftrightarrow$ Past (believed so)

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

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

$\rightarrow$ It limits the range of applications

Definitions → 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

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

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

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

In the probabilistic approach we are going to see

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

Probability

What <u>is</u> probability?

Probability

What is probability?

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

Probability

What is probability?

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

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

Probability

What is probability?

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

→ how much we are confident that something is true

→ how much we believe something

Probability

What <u>is</u> probability?

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

→ how much we are confident that something is true

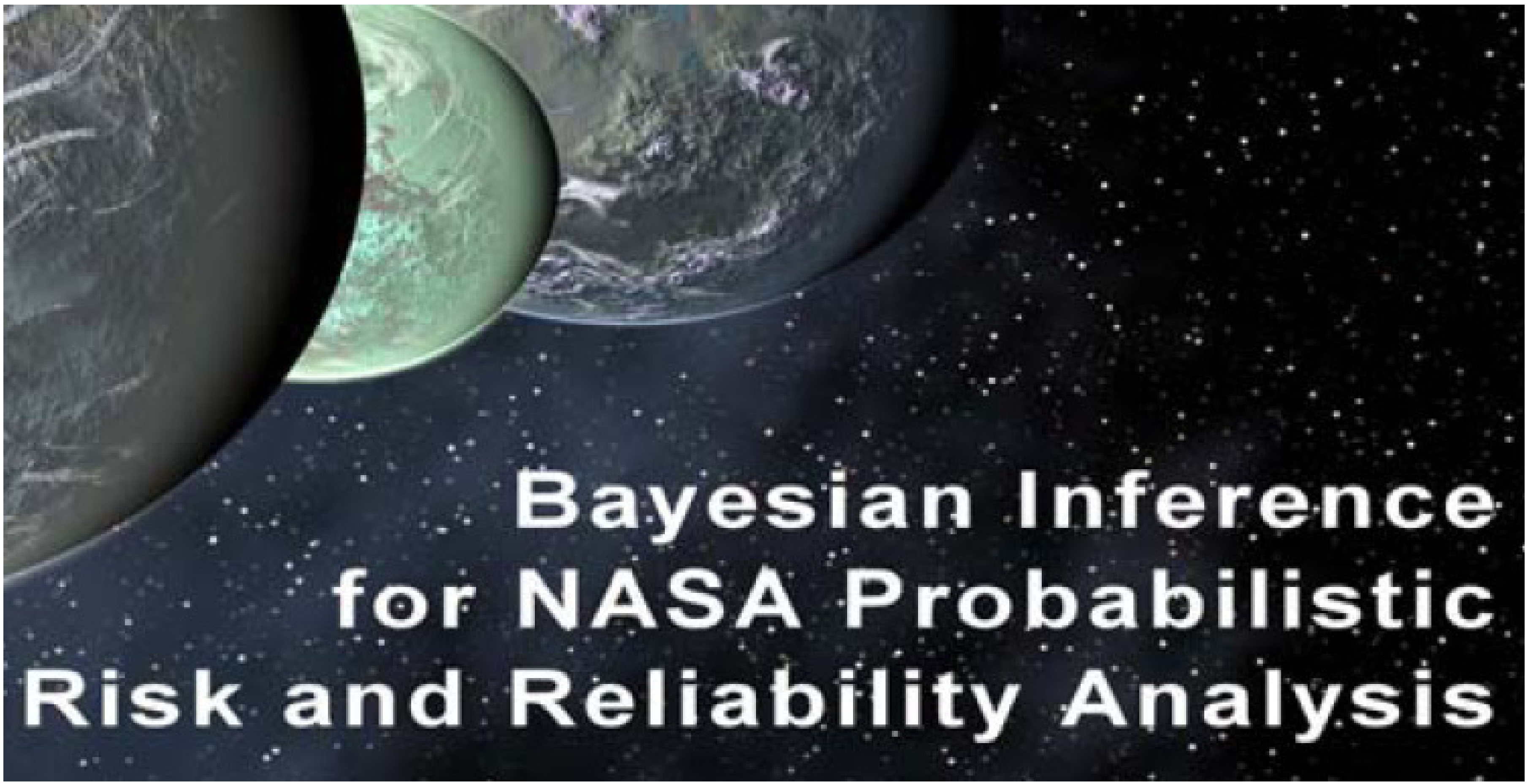
→ how much we believe something

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

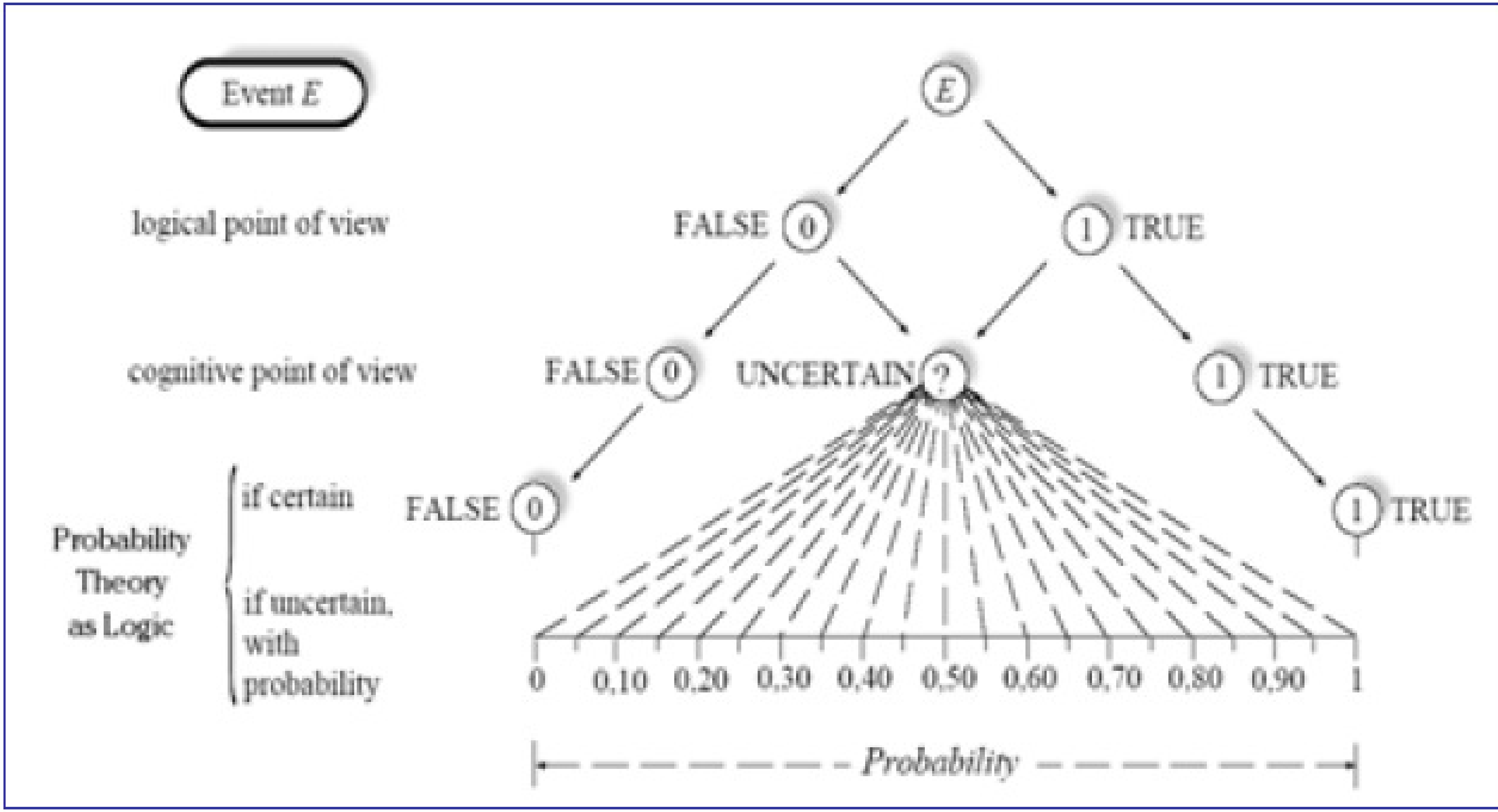
[Remark: 'will' does not imply future, but only uncertainty.]

An helpful diagram

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



An helpful diagram



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

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

Beliefs and ‘coherent’ bets

Remarks:

- Subjective <u>does not mean</u> 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				
11/07 20:30							
VOJVODINA - HIBERNIANS	1,05	10,00	25,00	3,10	1,30	2,55	1,42
GLENTORAN - KR REYKJAV	4,75	3,50	1,65	1,90	1,75	1,75	1,90
HONV BUDAP. - CELIK NIKS.	1,15	7,00	12,00	2,80	1,35	2,00	1,70
GERMANIA - OLANDA	1,15	6,50	13,00	2,50	1,45	2,20	1,57
11/07 20:45							
S PATRICKS - ZALGIRIS	1,90	3,40	3,50	1,75	1,90	1,73	1,95
11/07 21:00							
LIBERTAS - SARAJEVO	22,00	8,00	1,08	3,20	1,28	2,25	1,55
11/07 22:00							
STJARNAN - HAFNARFJOR	2,65	3,40	2,35	2,15	1,60	1,50	2,35

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

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

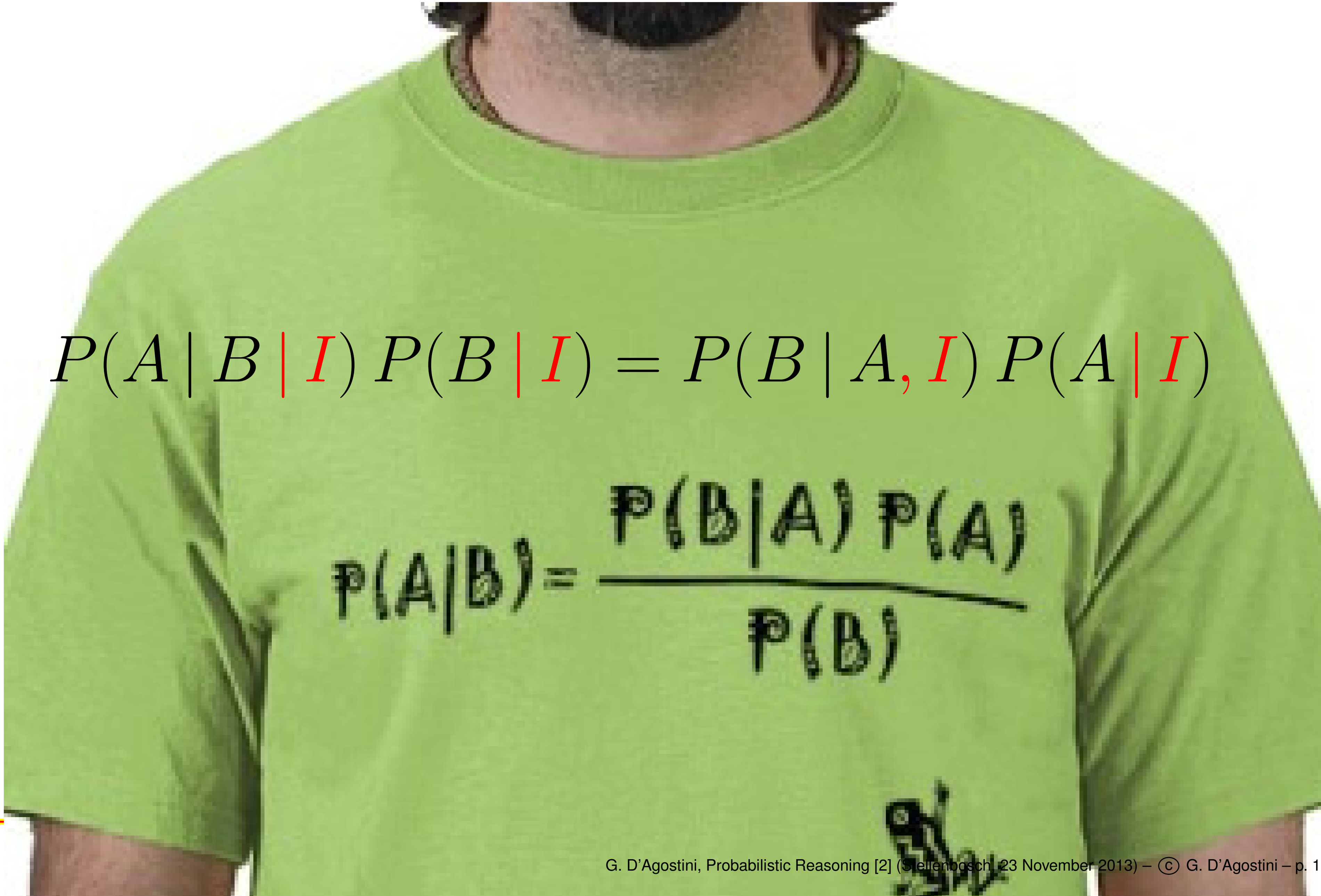
An even better news:

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

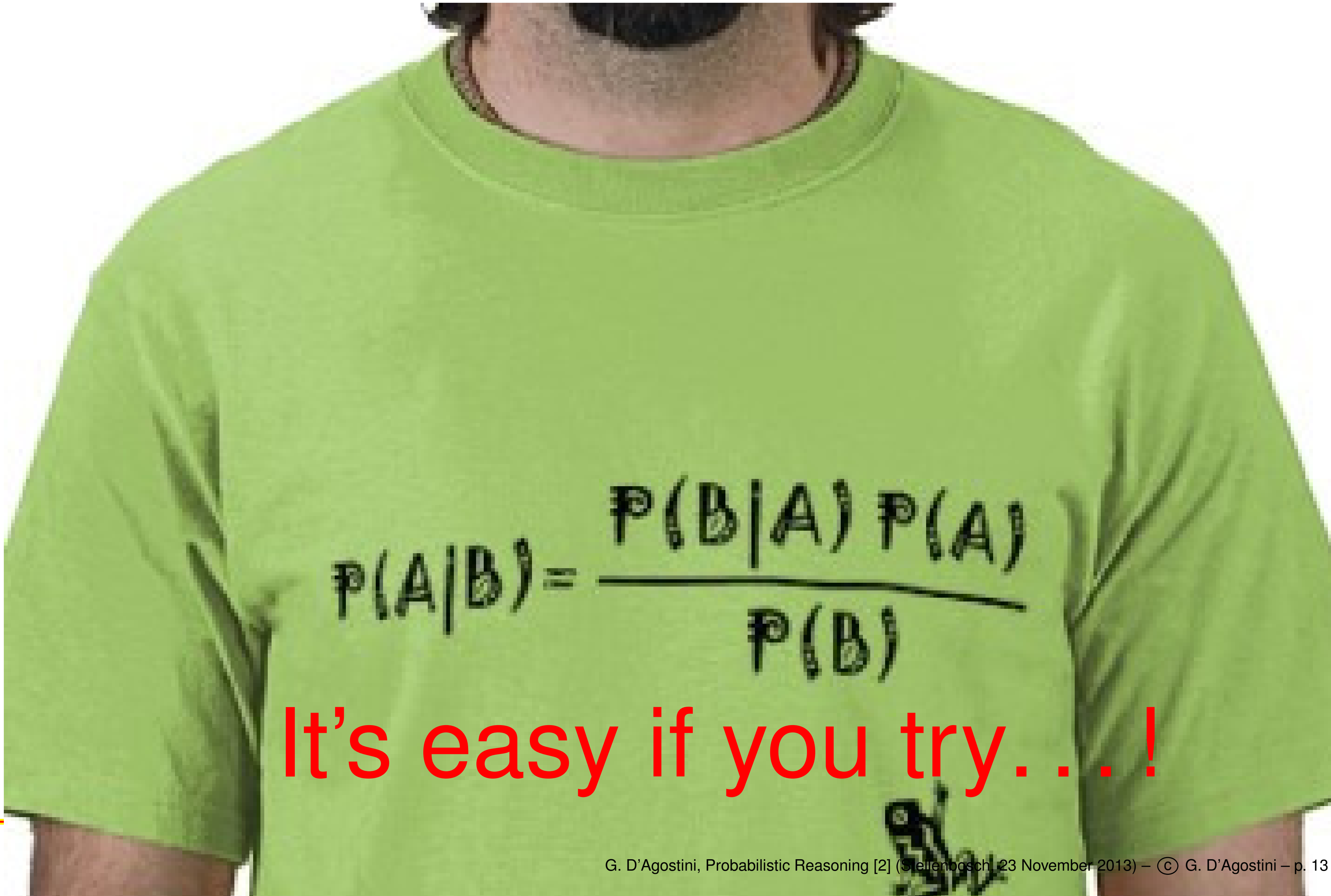


$$P(A\,|\,B\,|\,I)\,P(B\,|\,I) = P(B\,|\,A, I)\,P(A\,|\,I)$$

A simple, powerful formula

Take the courage to use it!

A simple, powerful formula



It's easy if you try. . . !

Telling it with Gauss' words

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

Telling it with Gauss' words

A quote from 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 quote from 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)

Bayes formulae

The essence is all contained in the fourth basic rule of probability theory:

Bayes formulae

The essence is all contained in the fourth basic rule of probability theory:

$$\frac{P(C_i \,|\, E, I)}{P(C_i \,|\, I)} \;=\; \frac{P(E \,|\, C_i, I)}{P(E \,|\, I)}$$

Bayes formulae

The essence is all contained in the fourth basic rule of probability theory:

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

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

Bayes formulae

The essence is all contained in the fourth basic rule of probability theory:

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

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

$$P(C_i \,|\, E \,|\, I) = \frac{P(E \,|\, C_i \,|\, I) \cdot P(C_i \,|\, I)}{\sum_k P(E \,|\, C_k, I) \cdot P(C_k \,|\, I)}$$

Bayes formulae

The essence is all contained in the fourth basic rule of probability theory:

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

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

$$P(C_i \,|\, E \,|\, I) = \frac{P(E \,|\, C_i \,|\, I) \cdot P(C_i \,|\, I)}{\sum_k P(E \,|\, C_k, I) \cdot P(C_k \,|\, I)}$$

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

Bayes formulae

The essence is all contained in the fourth basic rule of probability theory:

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

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

$$P(C_i\,|\,E\,|\,I) = \frac{P(E\,|\,C_i\,|\,I)\cdot P(C_i\,|\,I)}{\sum_k P(E\,|\,C_k,I)\cdot P(C_k\,|\,I)}$$

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

or even (my preferred form to grasp its meaning):

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

Bayesian parametric inference

If we want to infer a continuous parameter p from a set of data

$\rightarrow$ straghtforwad extension to probability density functions (pdf)

Bayesian parametric inference

If we want to infer a continuous parameter *p* from a set of data

→ straghtforwad extension to probability density functions (pdf)

$$f(p \,|\, \text{data}, I) \;\propto\; f(\text{data} \,|\, p, I) \cdot f(p \,|\, I)$$

Bayesian parametric inference

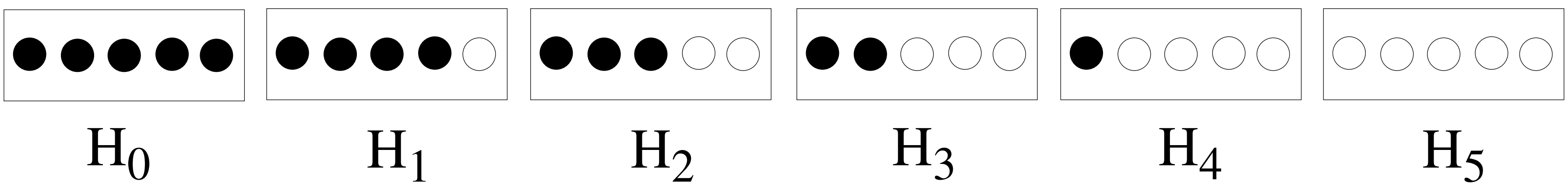
If we want to infer a continuous parameter *p* from a set of data

→ straghtforwad extension to probability density functions (pdf)

$$f(p\,|\,\text{data}, I) \propto f(\text{data}\,|\,p, I)\cdot f(p\,|\,I)$$

$$f(p\,|\,\text{data}, I) = \frac{f(\text{data}\,|\,p, I)\cdot f(p\,|\,I)}{\int_p f(\text{data}\,|\,p, I)\cdot f(p\,|\,I)\,dp}$$

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
→ 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)
- 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:

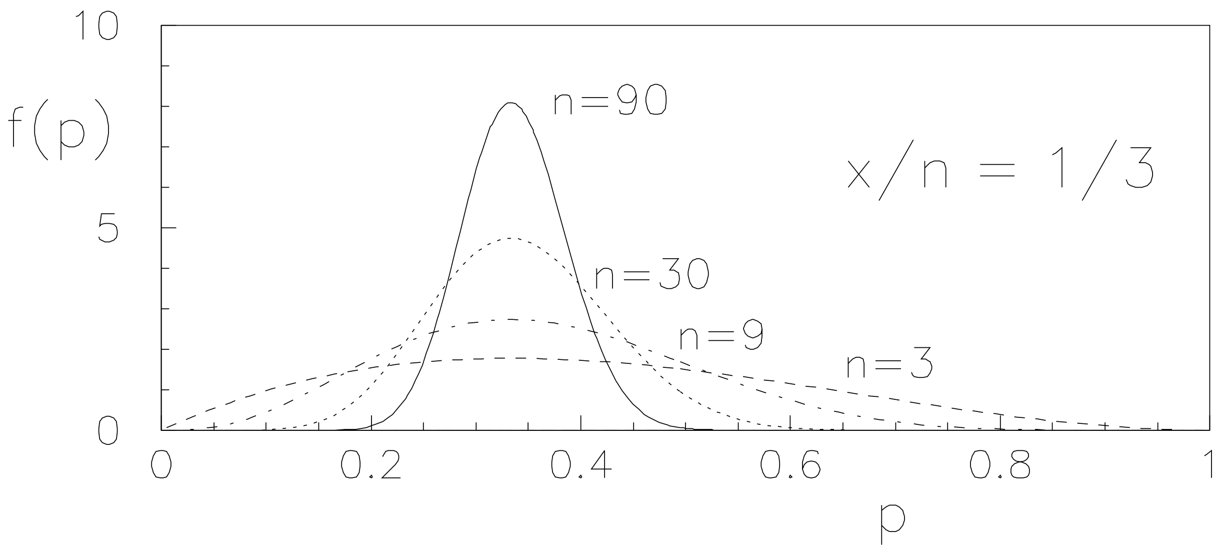
- Left/Right $\rightarrow$ Success/Failure
- if Left $\leftrightarrow$ Success:
  - $l/L \leftrightarrow p$ of binomial (Bernoulli trials)

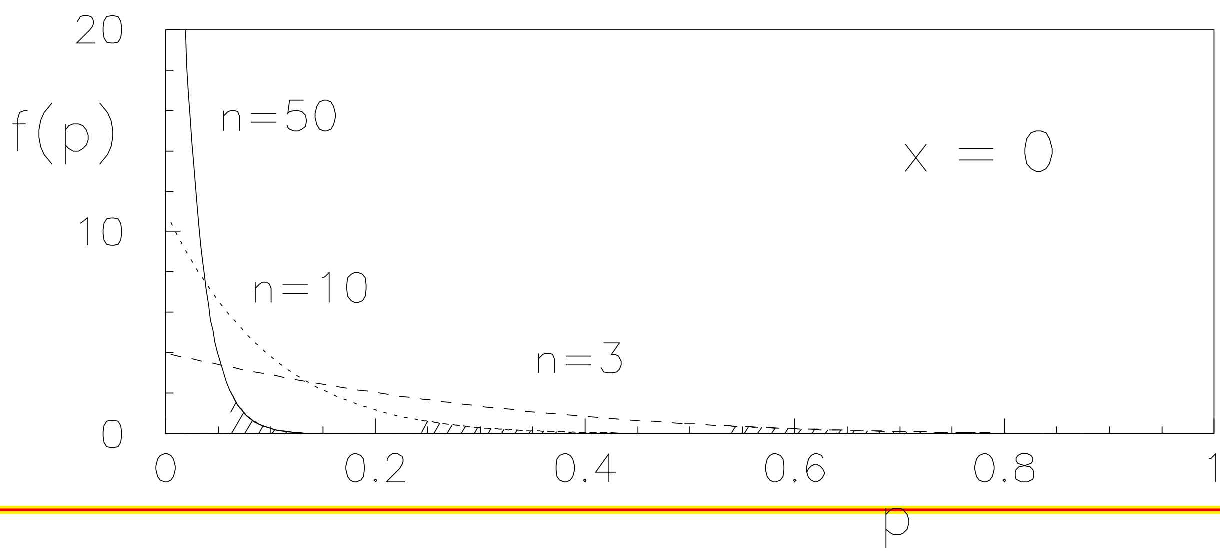
Solution with modern notation:
Imagine a sequence $\{S, S, F, S, \ldots\}$ [f_0 is uniform]:

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

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

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

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

Laplace's rule of successions

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

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

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

Laplace's rule of successions

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

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

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

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

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

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

Special case with $x = 0$

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

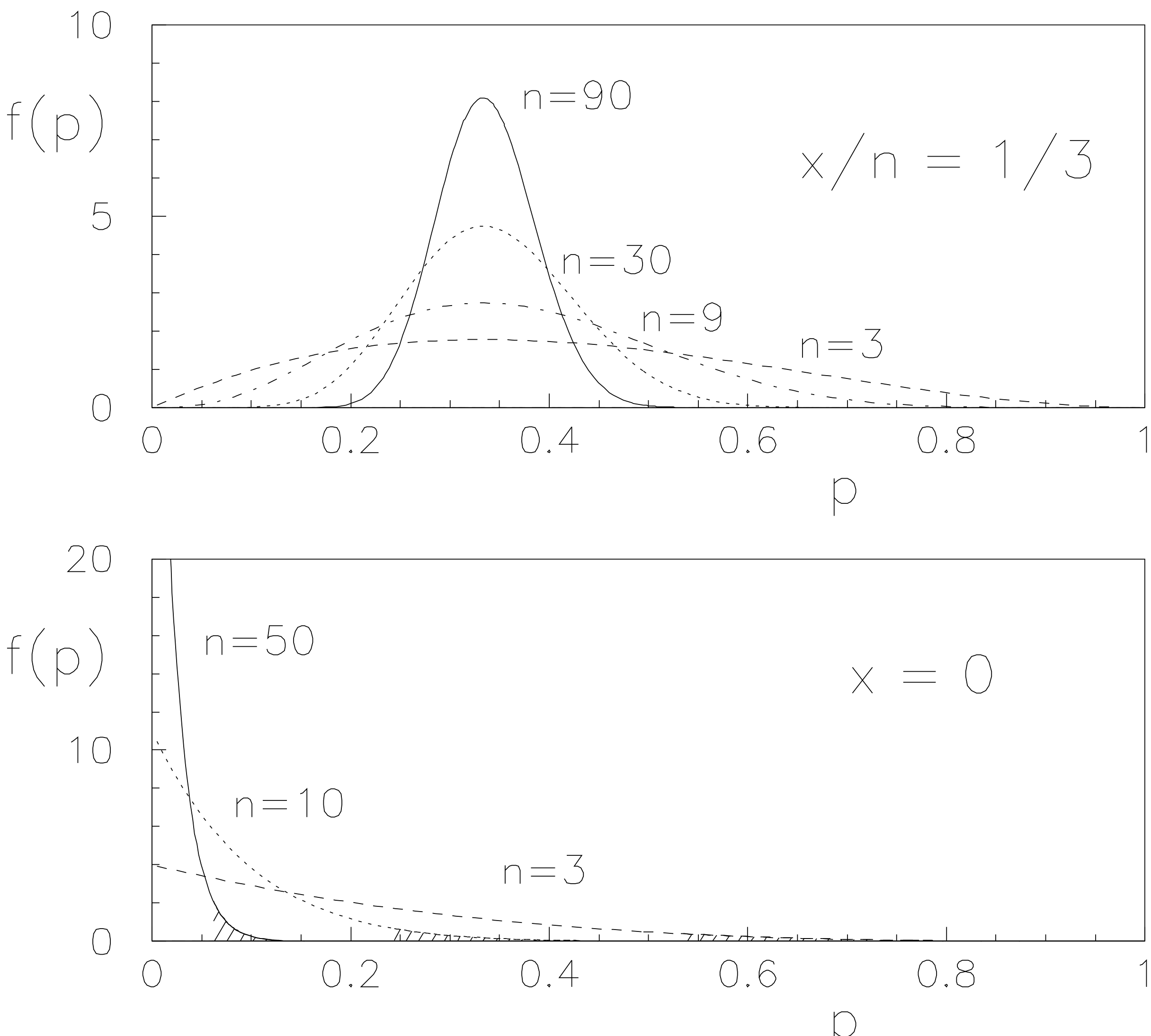
Special case with $x = 0$

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

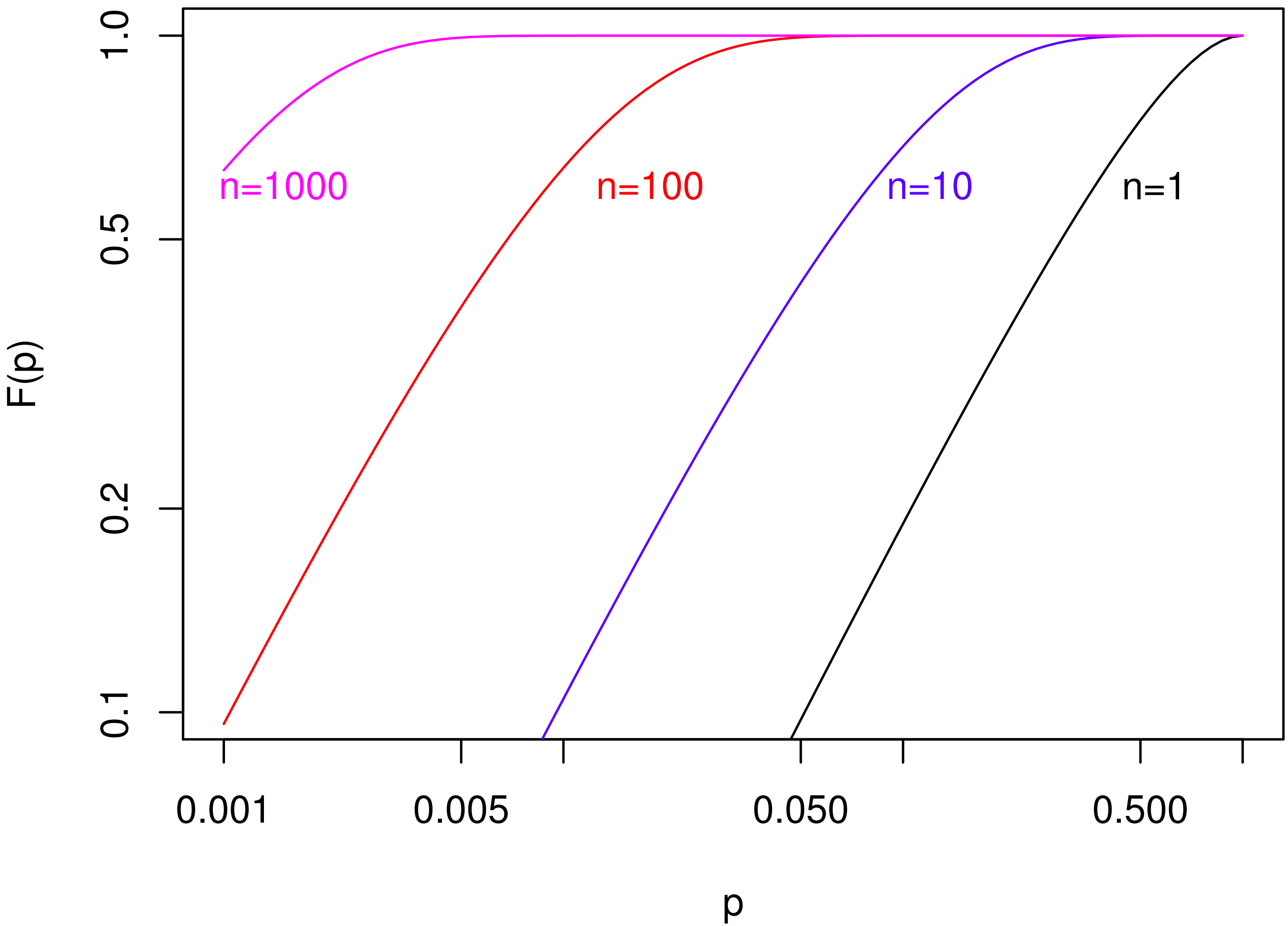
$$\begin{aligned} P(p \le p_u\,|\,0,n,\mathcal{B}) &= 95\% \\ &\Rightarrow \quad p_u = 1 - \sqrt[n+1]{0.05}\ : \end{aligned}$$

Probabilistic upper bound

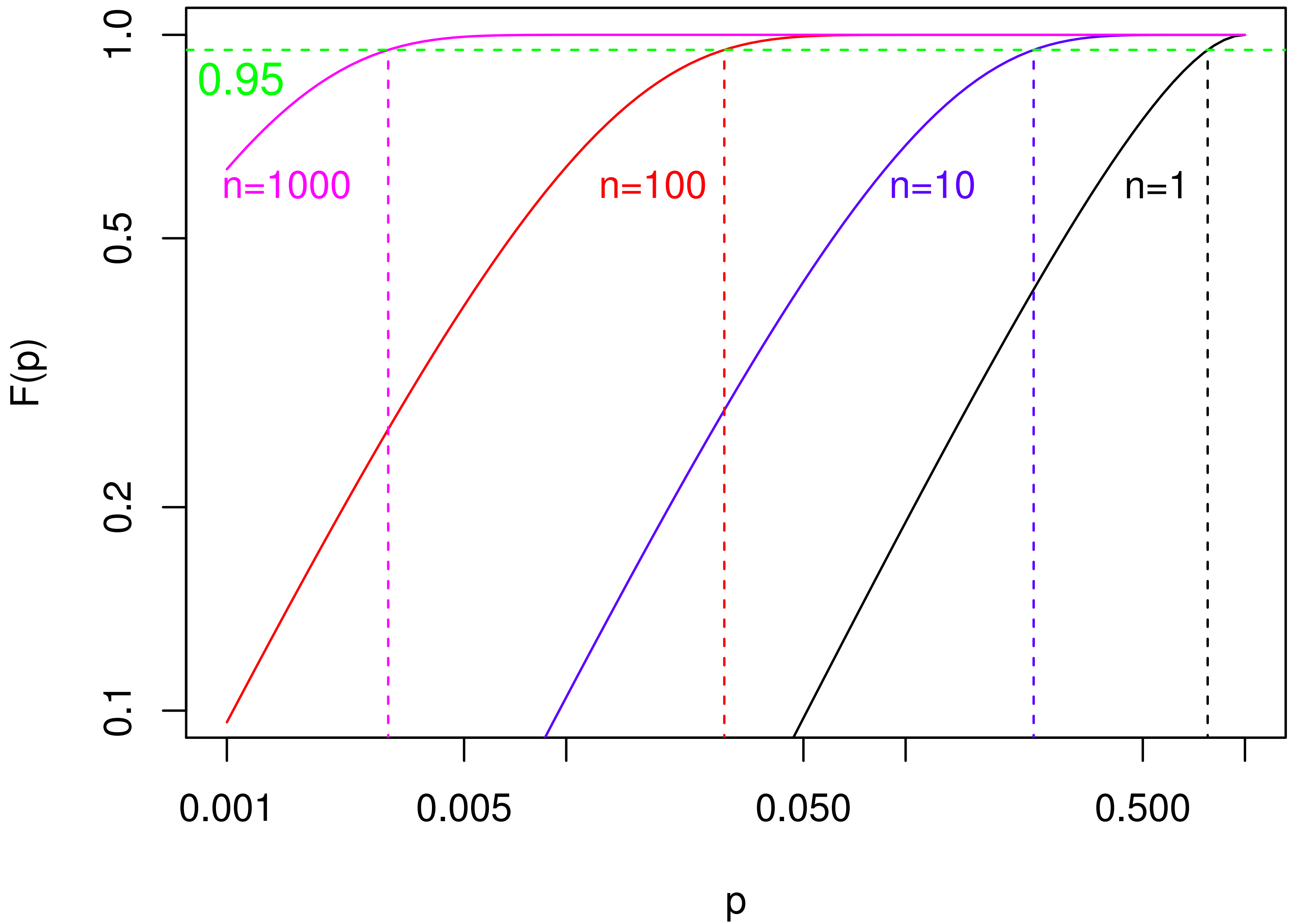
Special case with $x = 0$



Special case with $x = 0$



Special case with $x = 0$



Special case with $x = 0$

For the case $x = n$

(like 'observing' a 100% efficiency):

$\rightarrow$ just reason on the complementary parameter

$$q = 1 - p$$

Continuing the game

We have seen ho to tackle with a single idea problems that are treated difefrently in ‘standard statistics’:

- comparing hypotheses
- parametric inference

Continuing the game

We have seen ho to tackle with a single idea problems that are treated difefrently in ‘standard statistics’:

- comparing hypotheses
- parametric inference

You can continue the game

- playing with other models of $f(\text{data} \,|\, p, I)$

Continuing the game

We have seen ho to tackle with a single idea problems that are treated difefrently in ‘standard statistics’:

- comparing hypotheses
- parametric inference

You can continue the game

- playing with other models of $f(\text{data}\,|\,p, I)$
- make a simultaneous inference on several parameters

$$\rightarrow f(p_1, p_2, \ldots \,|\, \text{data}, I)$$

Continuing the game

We have seen ho to tackle with a single idea problems that are treated difefrently in ‘standard statistics’:

- comparing hypotheses
- parametric inference

You can continue the game

- playing with other models of $f(\text{data}\,|\,p, I)$
- make a simultaneous inference on several parameters

$$\rightarrow f(p_1, p_2, \ldots \,|\, \text{data}, I)$$

- take into account for systematics

$$\rightarrow \text{“integrating over subsamples of } I\text{”}$$

Continuing the game

We have seen ho to tackle with a single idea problems that are treated difefrently in ‘standard statistics’:

- comparing hypotheses
- parametric inference

You can continue the game

- playing with other models of $f(\text{data} \,|\, p, I)$
- make a simultaneous inference on several parameters

$$\rightarrow f(p_1, p_2, \ldots \,|\, \text{data}, I)$$

- take into account for systematics

$$\rightarrow \text{“integrating over subsamples of } I\text{”}$$

- etc.

Continuing the game

We have seen ho to tackle with a single idea problems that are treated difefrently in ‘standard statistics’:

- comparing hypotheses
- parametric inference

You can continue the game
...although at a certain point you need to face
computational issues

Continuing the game

We have seen ho to tackle with a single idea problems that are treated difefrently in ‘standard statistics’:

- comparing hypotheses
- parametric inference

You can continue the game
. . . although at a certain point you need to face computational issues

$\rightarrow$ that was of the main reason why Laplace’s methods were set apart and frequentistics methods fourished

Continuing the game

We have seen ho to tackle with a single idea problems that are treated difefrently in ‘standard statistics’:

- comparing hypotheses
- parametric inference

You can continue the game
...although at a certain point you need to face computational issues

$\rightarrow$ that was of the main reason why Laplace’s methods were set apart and frequentistics methods fourished

But we can now benefit of powerful computers and impressive improvements in computation methods

Continuing the game

We have seen ho to tackle with a single idea problems that are treated difefrently in 'standard statistics':

- comparing hypotheses
- parametric inference

You can continue the game
...although at a certain point you need to face computational issues

$\rightarrow$ that was of the main reason why Laplace's methods were set apart and frequentistics methods fourished

But we can now benefit of powerful computers and impressive improvements in computation methods

We have no longer excuses!!

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 in 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 in something.
  (Diffidate chi vi promette di far germogliar zecchini nel Campo dei Miracoli! – Pinocchio docet)

Good reasoning Vs ’prohibitive calculations’

The main reasons why the so called Bayesian reasoning hab been blooming in the last decades are related to

- computational power
- helped from brilliant ideas on how to use it
  (but sterile without modern computers!)

Good reasoning Vs 'prohibitive calculations'

The main reasons why the so called Bayesian reasoning hab been blooming in the last decades are related to

- computational power
- helped from brilliant ideas on how to use it (but sterile without modern computers!)

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

→ no longer an excuse!

Good reasoning Vs 'prohibitive calculations'

The main reasons why the so called Bayesian reasoning hab been blooming in the last decades are related to

- computational power
- helped from brilliant ideas on how to use it (but sterile without modern computers!)

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

- instead of methods that require integrals, methods have been 'invented' that only require derivatives
  ⇒ Maximum Likelihood

  → no longer an excuse!

Good reasoning Vs 'prohibitive calculations'

The main reasons why the so called Bayesian reasoning hab been blooming in the last decades are related to

- computational power
- helped from brilliant ideas on how to use it (but sterile without modern computers!)

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

- instead of methods that require integrals, methods have been 'invented' that only require derivatives
  ⇒ Maximum Likelihood ...often a good approximated solution <u>under</u> some assumptions usually unkown to practitioners.

  → no longer an excuse!

Conclusions from Part - 2

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

Conclusions from Part - 2

- 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 of everybody else).

Conclusions from Part - 2

- 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 of everybody else).
- Its consistent application in small-complex problems was prohibitive many years ago.

Conclusions from Part - 2

- 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 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 from Part - 2

- 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 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 from Part - 2

- 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 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’.