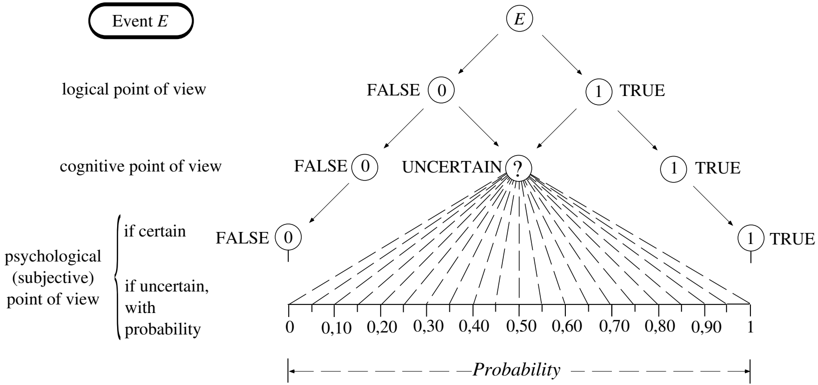
Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

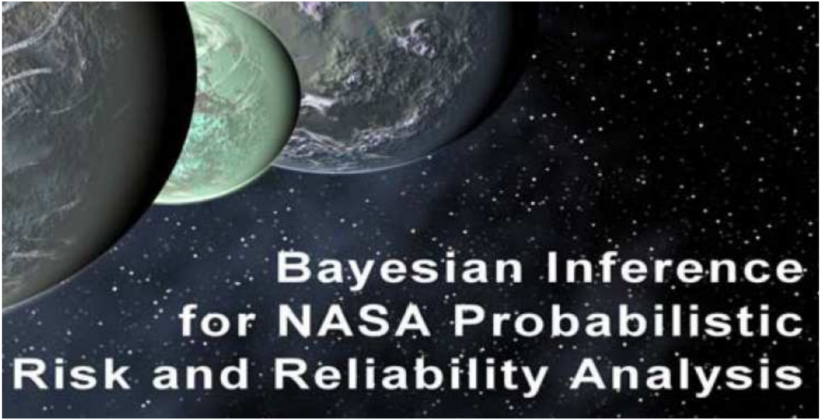
Università di Roma La Sapienza e INFN
Roma, Italy

False, True and probable

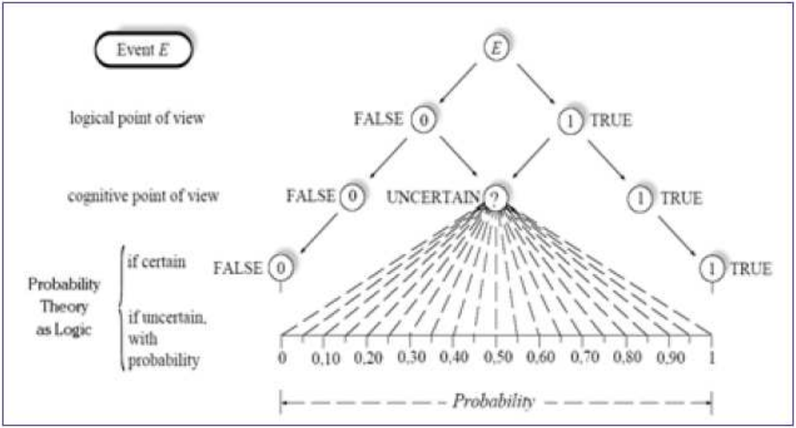


An helpful diagram

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 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 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 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)*

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

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.

> *"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)*

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

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!

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)*

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

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 think so, then you could make a bad bet!

A clear statement from a honest expert

MODIFIED FREQUENTIST ANALYSIS OF SEARCH RESULTS (THE CL_s METHOD)

A. L. Read
University of Oslo, Department of Physics, P.O. Box 1048, Blindern, 0316 Oslo 3, Norway

CONCLUSION

It neither adheres to the frequentist principle of coverage (it overcovers *by design* as the experimental sensitivity to the hypothetical signal vanishes) nor does it indicate the bounds of a Bayesian subjective probability distribution.

Let me try to make an important point about the previous sentence as clearly and simply as possible (even my friends claim I got it wrong all the three times I tried to explain this in my presentation): The lower bounds on the Higgs mass that are quoted for the direct Higgs searches at LEP say *absolutely nothing* about the *probability of the Higgs mass* being higher or lower than some value.

To make such a statement the direct search results must be first folded with a prior probability distribution for the Higgs mass [15].

Beliefs and bet

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

Beliefs and bet

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

"Although I felt sure that parity would not be violated, there was a possibility that it would be, and it was important to find out.

Beliefs and bet

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

"Although I felt sure that parity would not be violated, there was a possibility that it would be, and it was important to find out.
'Would you bet a hundred dollars a dollar that parity is not violated?' he asked.

Beliefs and bet

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

> *"Although I felt sure that parity would not be violated, there was a possibility that it would be, and it was important to find out.*
> *'Would you bet a hundred dollars a dollar that parity is not violated?' he asked.*
> *'No. But fifty dollars I will'"*

(Feynman)

Beliefs and bet

Even Kant sponsors it...

"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.

Beliefs and bet

Even Kant sponsors it...

"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.

Beliefs and bet

Even Kant sponsors it...

"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.

Beliefs and bet

Even Kant sponsors it. . .

"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."

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

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

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

Future ⇔ Past (belief!)

$n \to \infty$: → *“usque tandem?”*

→ *“in the long run we are all dead”*

→ It limits the range of applications

Standard textbook definitions

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

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

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

Future ⇔ Past (belief!)

$n\to\infty$: → *"usque tandem?"*

→ *"in the long run we are all dead"*

→ It limits the range of applications

Future ⇔ Past: avoid the end of the *inductivist turkey*!

‘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

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

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

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 of Probability Theory (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 of Probability Theory (under well defined assumptions):
  $\Rightarrow$ Laplace’s rule of succession (see later)

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.*

It can be proved that

> *the requirement of* ***coherence*** *leads to the famous 4 basic rules* $\Longrightarrow$

[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')

→ usually implicit (we only care about 're-conditioning')

Basic rules of probability

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

Remember that probability is always conditional probability!

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

→ usually implicit (we only care about 're-conditioning')

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

Mathematics of beliefs

An even <u>better news</u>:

The fourth basic rule
can be fully exploited!

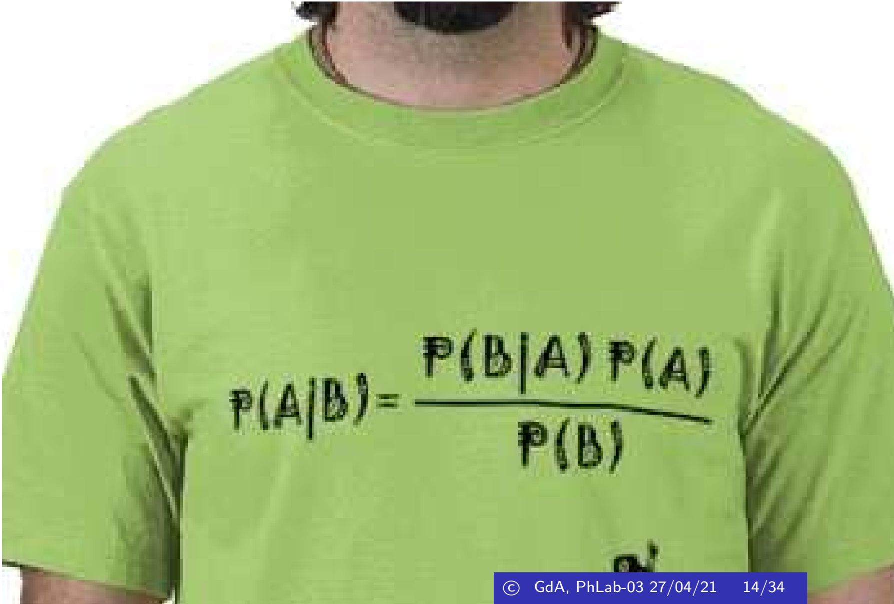
Mathematics of beliefs

An even better news:

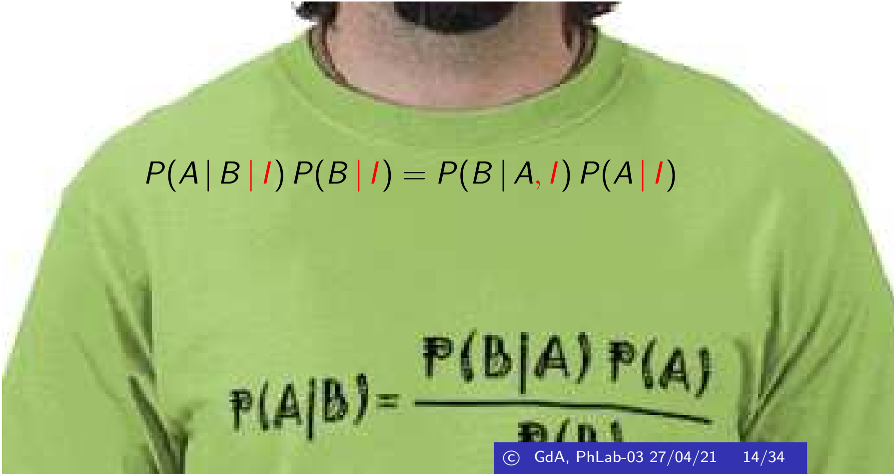
The fourth basic rule
can be fully exploited!

(Liberated by a curious ideology that forbids its use)

A simple, powerful formula

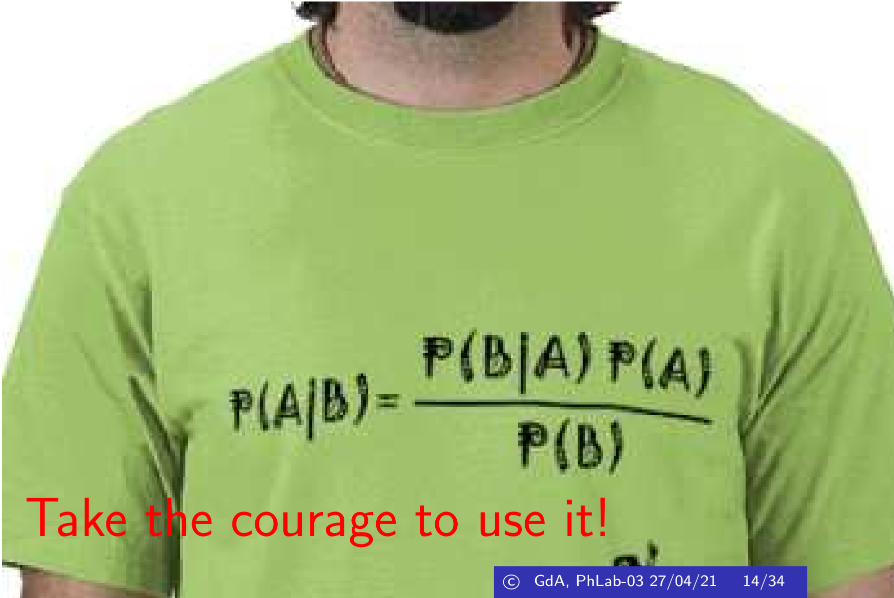


A simple, powerful formula



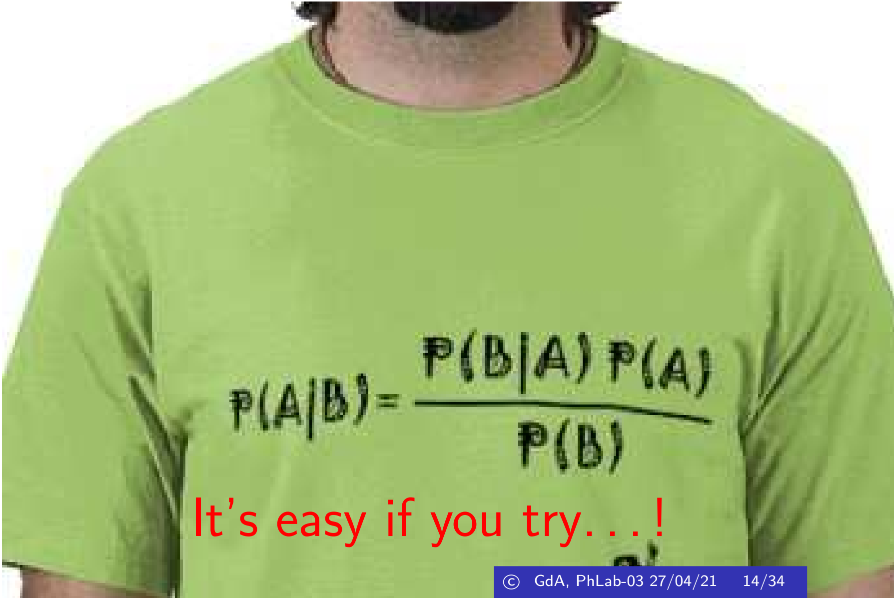
$$P(A \mid B \mid I)\, P(B \mid I) = P(B \mid A, I)\, P(A \mid I)$$

A simple, powerful formula



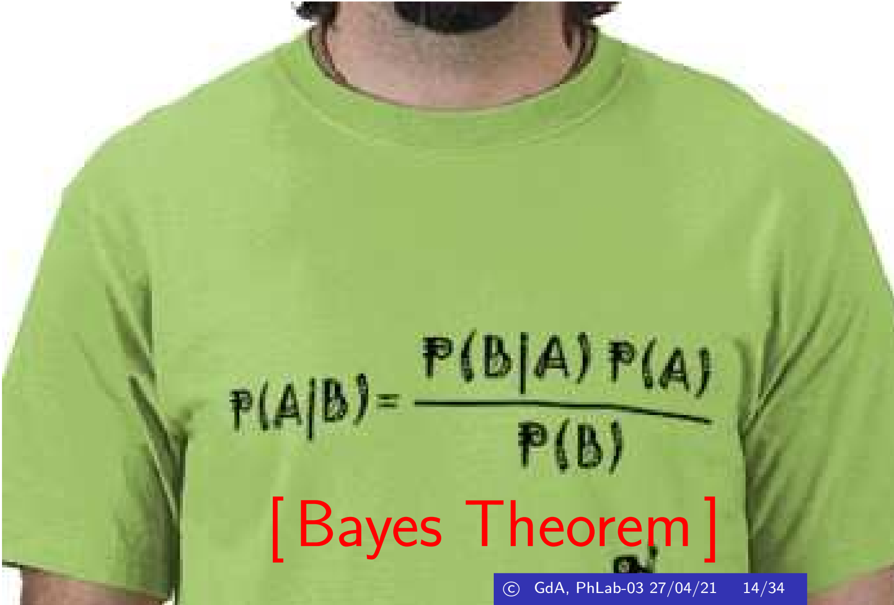
Take the courage to use it!

A simple, powerful formula



It's easy if you try. . . !

A simple, powerful formula



[Bayes Theorem]

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.

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

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*."

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

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*."

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

(*Philosophical Essai on Probabilities*)

[In general $P(E) = \sum_j P(E \,|\, C_j)\, P(C_j)$ (*weighted average*, with weigths being the probabilities of the conditions) if C_j form a *complete class of hypotheses*]

Laplace's "Bayes Theorem"

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

"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 'fundamental rules'.

Laplace's "Bayes Theorem"

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

"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 'fundamental rules'.

Note: denominator is just a normalization factor.

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

Laplace's "Bayes Theorem"

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

"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 'fundamental rules'.

Note: denominator is just a normalization factor.

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

Most convenient way to remember Bayes theorem

Laplace's teaching

$$\frac{P(H_0 \,|\, \text{data})}{P(H_1 \,|\, \text{data})} = \frac{P(\text{data} \,|\, H_0)}{P(\text{data} \,|\, H_1)} \times \frac{P(H_0)}{P(H_1)}$$

- We should possibly use the data, rather then the test variables 'θ' (χ^2 etc);

  [although in some case 'sufficient summaries' do exist]

Laplace's teaching

$$\frac{P(H_0 \,|\, \text{data})}{P(H_1 \,|\, \text{data})} \;=\; \frac{P(\text{data} \,|\, H_0)}{P(\text{data} \,|\, H_1)} \times \frac{P(H_0)}{P(H_1)}$$

- We should possibly use the data, rather then the test variables 'θ' (χ^2 etc);

  [although in some case 'sufficient summaries' do exist]

- At least two hypotheses are needed!

Laplace's teaching

$$\frac{P(H_0 \,|\, \text{data})}{P(H_1 \,|\, \text{data})} = \frac{P(\text{data} \,|\, H_0)}{P(\text{data} \,|\, H_1)} \times \frac{P(H_0)}{P(H_1)}$$

- We should possibly use the data, rather then the test variables 'θ' (χ^2 etc);
  [although in some case 'sufficient summaries' do exist]
- At least two hypotheses are needed!
- ...and also how they appear belivable *a priori*!

Laplace's teaching

$$\frac{P(H_0 \mid \text{data})}{P(H_1 \mid \text{data})} = \frac{P(\text{data} \mid H_0)}{P(\text{data} \mid H_1)} \times \frac{P(H_0)}{P(H_1)}$$

- We should possibly use the data, rather then the test variables 'θ' (χ^2 etc);
  [although in some case 'sufficient summaries' do exist]
- At least two hypotheses are needed!
- ... and also how they appear belivable *a priori*!
- If $P(\text{data} \mid H_i) = 0$, it follows $P(H_i \mid \text{data}) = 0$:
  $\Rightarrow$ falsification (the 'serious' one) is a corollary of the theorem, rather than a principle.

Laplace's teaching

$$\frac{P(H_0\,|\,\text{data})}{P(H_1\,|\,\text{data})} \;=\; \frac{P(\text{data}\,|\,H_0)}{P(\text{data}\,|\,H_1)} \times \frac{P(H_0)}{P(H_1)}$$

- We should possibly use the data, rather then the test variables 'θ' (χ^2 etc);

  [although in some case 'sufficient summaries' do exist]
- At least two hypotheses are needed!
- ...and also how they appear belivable *a priori*!
- If $P(\text{data}\,|\,H_i) = 0$, it follows $P(H_i\,|\,\text{data}) = 0$:
  $\Rightarrow$ falsification (the 'serious' one) is a corollary of the theorem, rather than a principle.
- There is no conceptual problem with the fact that $P(\text{data}\,|\,H_1) \to 0$ (e.g. 10^{-37}), provided the ratio $P(\text{data}\,|\,H_0)/P(\text{data}\,|\,H_1)$ is not undefined.

Bayes factor ('likelihood ratio')

$$\frac{P(H_0 \,|\, \text{data})}{P(H_1 \,|\, \text{data})} \;=\; \frac{P(\text{data} \,|\, H_0)}{P(\text{data} \,|\, H_1)} \times \frac{P(H_0)}{P(H_1)}$$

Bayes factor ('likelihood ratio')

$$\frac{P(H_0 \,|\, \text{data})}{P(H_1 \,|\, \text{data})} \;=\; \frac{P(\text{data} \,|\, H_0)}{P(\text{data} \,|\, H_1)} \times \frac{P(H_0)}{P(H_1)}$$

$$\text{Prob. ratio}\big|_{\textit{posterior}} \;=\; \text{Bayes factor} \times \text{Prob. ratio}\big|_{\textit{prior}}$$

(*prior*/*posterior* w.r.t. data)

Bayes factor ('likelihood ratio')

$$\frac{P(H_0\,|\,\text{data})}{P(H_1\,|\,\text{data})} \;=\; \frac{P(\text{data}\,|\,H_0)}{P(\text{data}\,|\,H_1)} \times \frac{P(H_0)}{P(H_1)}$$

$$\text{Prob. ratio}\big|_{posterior} \;=\; \text{Bayes factor} \times \text{Prob. ratio}\big|_{prior}$$

(*prior*/*posterior* w.r.t. data)

If H_0 and H_1 are 'complementary', that is $H_1 = \overline{H}_0$, then

$$\textbf{posterior odds} \;=\; \textbf{Bayes factor} \times \textbf{prior odds}$$

But statistical tests do work!

Someone would object that p-values and, in general, ‘hypothesis tests’ *usually* do work!

But statistical tests do work!

Someone would object that p-values and, in general, 'hypothesis tests' *usually* do work!

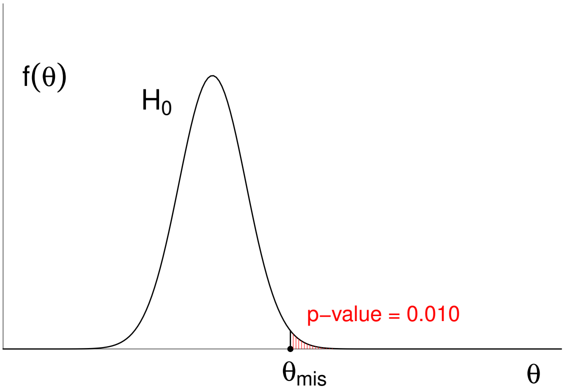
- Certainly! I agree!
  As it *usually* work overtakes in curve
  on remote mountain road!

But statistical tests do work!

Someone would object that p-values and, in general, 'hypothesis tests' *usually* do work!

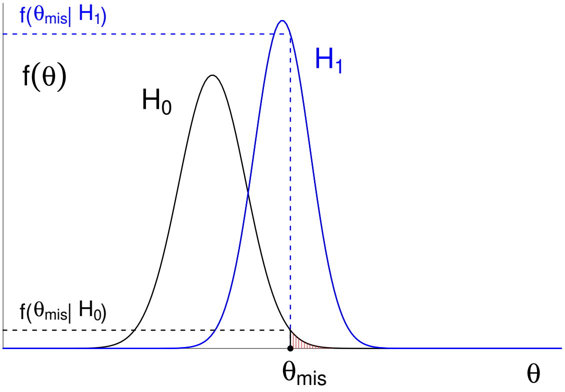
- Certainly! I agree!
  As it *usually* work overtakes in curve
  on remote mountain road!
- But now we are also able to explain the reason.

But statistical tests do work!



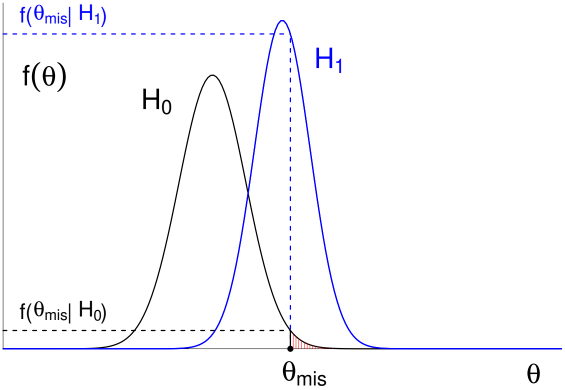
Why should the observation of θ_{mis} should diminish our confidence on H_0?

But statistical tests do work!



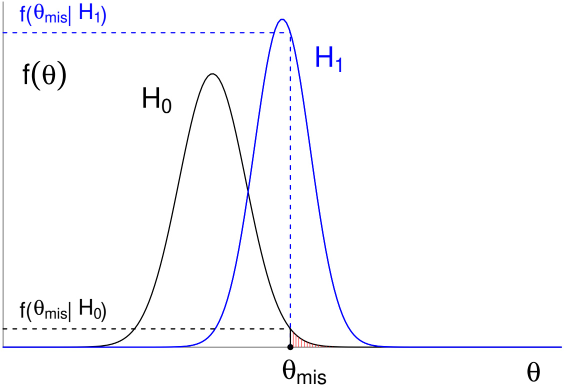
Because *often* we give *some chance* to a possible alternative hypothesis H_1, even if we are not able to exactly formulate it.

But statistical tests do work!



Indeed, what really matters is not the area to the right of θ_{mis}.
What matters is the ratio of $f(\theta_{mis} \,|\, H_1)$ to $f(\theta_{mis} \,|\, H_0)$!
$\Rightarrow$ to a 'small' area it corresponds a 'small' $f(\theta_{mis} \,|\, H_0)$.

But statistical tests do work!



But is the alternative hypothesis H_1 is unconceivable, or hardly believable, the 'smallness' of the area is irrelevant

Telling it with Gauss' words

A quote from the Princeps Mathematicorum
(Prince of Mathematicians) is a must.

Telling it with Gauss' words

A quote from the Princeps Mathematicorum (Prince of Mathematicians) is a must.

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

Telling it with Gauss' words

A quote from the Princeps Mathematicorum (Prince of Mathematicians) is a must.

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

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

(Gauss)

Telling it with Gauss' words

A quote from the Princeps Mathematicorum
(Prince of Mathematicians) is a must.

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

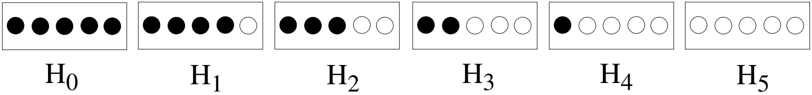
"post illa observationes" *"ante illa observationes"*

(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 $\mu = \overline{x}$

Application to the six box problem



Remind:

- E_1 = White
- E_2 = Black

Collecting the pieces of information we need

Our tool:

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

Collecting the pieces of information we need

Our tool:

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

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

Collecting the pieces of information we need

Our tool:

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

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

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:

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

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

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

Probability of E_i taking account all possible H_j
$\rightarrow$ How much we are confident that E_i will occur.
(taking into account all possible hypotheses H_j)

Collecting the pieces of information we need

Our tool:

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

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

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

But it easy to prove that $P(E_i \,|\, I)$ is related to the other ingredients, usually easier to 'measure' or to assess somehow, though vaguely

Collecting the pieces of information we need

Our tool:

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

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

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

But it easy to prove that $P(E_i \,|\, I)$ is related to the other ingredients, usually easier to 'measure' or to assess somehow, though vaguely

'decomposition law': $P(E_i \,|\, I) = \sum_j P(E_i \,|\, H_j, I) \cdot P(H_j \,|\, I)$
(→ Easy to check that it gives $P(E_i \,|\, I) = 1/2$ in our case).

Collecting the pieces of information we need

Our tool:

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

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

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

We are ready!

$\longrightarrow$ Let's play with our toy

We are ready

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

- analyse real data
- some simulations
- make variations

We are ready

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

- analyse real data
- some simulations
- make variations

Let's play!

- Hugin Expert (Lite – demo version);
- R scripts

Playing with the six boxes

Learning by simulations

- History of $P(H_j \,|\, \text{obs. sequence})$.

Playing with the six boxes

Learning by simulations

- History of $P(H_j \,|\, \text{obs. sequence})$.
- History of $P(B/W \,|\, \text{obs. sequence})$.

Playing with the six boxes

Learning by simulations

- History of $P(H_j\,|\,\text{obs. sequence})$.
- History of $P(B/W\,|\,\text{obs. sequence})$.
- Comparison of the $P(B/W\,|\,\text{obs. sequence})$ with the relative frequency with the color has occurred in the past ("probability evaluated by relative frequency").

Playing with the six boxes

Learning by simulations

- History of $P(H_j \,|\, \text{obs. sequence})$.
- History of $P(B/W \,|\, \text{obs. sequence})$.
- Comparison of the $P(B/W \,|\, \text{obs. sequence})$ with the relative frequency with the color has occurred in the past ("probability evaluated by relative frequency").
  - Why does the Bayesian solution performs better?

Playing with the six boxes

Learning by simulations

- History of $P(H_j \,|\, \text{obs. sequence})$.
- History of $P(B/W \,|\, \text{obs. sequence})$.
- Comparison of the $P(B/W \,|\, \text{obs. sequence})$ with the relative frequency with the color has occurred in the past ("probability evaluated by relative frequency").
  - Why does the Bayesian solution performs better?
  - → It takes into account at the best all available information. (The frequency based answer is, at most, the solution to a different problem...)

Playing with the six boxes

Learning by simulations

- History of $P(H_j \,|\, \text{obs. sequence})$.
- History of $P(B/W \,|\, \text{obs. sequence})$.
- Comparison of the $P(B/W \,|\, \text{obs. sequence})$ with the relative frequency with the color has occurred in the past ("probability evaluated by relative frequency").
  - Why does the Bayesian solution performs better?
  - → It takes into account at the best all available information. (The frequency based answer is, at most, the solution to a different problem...)
- Comparison of $P(H_j \,|\, \text{obs. sequence})$ with frequentistic methods?

Playing with the six boxes

Learning by simulations

- History of $P(H_j \,|\, \text{obs. sequence})$.
- History of $P(B/W \,|\, \text{obs. sequence})$.
- Comparison of the $P(B/W \,|\, \text{obs. sequence})$ with the relative frequency with the color has occurred in the past ("probability evaluated by relative frequency").
  - Why does the Bayesian solution performs better?
  - → It takes into account at the best all available information. (The frequency based answer is, at most, the solution to a different problem...)
- Comparison of $P(H_j \,|\, \text{obs. sequence})$ with frequentistic methods?

NO!

Playing with the six boxes

Learning by simulations

- History of $P(H_j \,|\, \text{obs. sequence})$.
- History of $P(B/W \,|\, \text{obs. sequence})$.
- Comparison of the $P(B/W \,|\, \text{obs. sequence})$ with the relative frequency with the color has occurred in the past ("probability evaluated by relative frequency").
  - Why does the Bayesian solution performs better?
  - → It takes into account at the best all available information. (The frequency based answer is, at most, the solution to a different problem...)
- Comparison of $P(H_j \,|\, \text{obs. sequence})$ with frequentistic methods?

NO!

- Don't even think: frequentists refuse to assign probabilities to hypotheses (in general), to causes, to true values, etc. (And you have seen the results. . .)

How does it work?

Simple case (no reporter/composition/etc. complications)

How does it work?

Simple case (no reporter/composition/etc. complications)

- Update probabilities of hypotheses (cause, Box): *inference*:

$$P^{(n)}(B_j) \propto P(E_i^{(n)} \,|\, B_j) \cdot P^{(n-1)}(B_j)$$

How does it work?

Simple case (no reporter/composition/etc. complications)

- Update probabilities of hypotheses (cause, Box): *inference*:

$$P^{(n)}(B_j) \propto P(E_i^{(n)} \mid B_j) \cdot P^{(n-1)}(B_j)$$

- Update probabilities of next extraction: *prediction*:

$$P^{(n+1)}(E_i) = \sum_j P(E_i \mid B_j) \cdot P^{(n)}(B_j)$$

How does it work?

General case (more complicate 'network')

How does it work?

General case (more complicate 'network')
for example including uncertain Preparation ('C' not to confuse it with 'Probability') and a Reporter for each estraction ($\underline{R}$):

How does it work?

General case (more complicate 'network')
for example including uncertain Preparation ('C' not to confuse it with 'Probability') and a Reporter for each estraction ($\underline{R}$):

- Write down the joint distribution of all variables in the game:

$$P(C, B, \underline{E}, \underline{R})$$

How does it work?

General case (more complicate 'network')
for example including uncertain Preparation ('C' not to confuse it with 'Probability') and a Reporter for each estraction ($\underline{R}$):

- Write down the joint distribution of all variables in the game:

$$P(C, B, \underline{E}, \underline{R})$$

$$\begin{array}{ll} \underline{E}: & E_i^{(1)}, E_i^{(2)}, E_i^{(3)}, \ldots \\ \underline{R}: & R_i^{(1)}, R_i^{(2)}, R_i^{(3)}, \ldots \end{array}$$

How does it work?

General case (more complicate 'network')
for example including uncertain Preparation ('C' not to confuse it with 'Probability') and a Reporter for each estraction ($\underline{R}$):

- Write down the joint distribution of all variables in the game:

$$P(C, B, \underline{E}, \underline{R})$$

$$\underline{E}: \; E_i^{(1)}, E_i^{(2)}, E_i^{(3)}, \ldots$$
$$\underline{R}: \; R_i^{(1)}, R_i^{(2)}, R_i^{(3)}, \ldots$$

- Condition on the 'observations':

$$P(C, B, \underline{E}, \underline{R}^{(k>n)} \,|\, \underline{R}^{(k\leq n)}) = \frac{P(C, B, \underline{E}, \underline{R})}{P(\underline{R}^{(k\leq n)})}$$

How does it work?

General case (more complicate 'network')
for example including uncertain Preparation ('C' not to confuse it with 'Probability') and a Reporter for each estraction ($\underline{R}$):

- Write down the joint distribution of all variables in the game:

$$P(C, B, \underline{E}, \underline{R})$$

$$\begin{array}{ll} \underline{E}: & E_i^{(1)}, E_i^{(2)}, E_i^{(3)}, \ldots \\ \underline{R}: & R_i^{(1)}, R_i^{(2)}, R_i^{(3)}, \ldots \end{array}$$

- Condition on the 'observations':

$$P(C, B, \underline{E}, \underline{R}^{(k>n)} \,|\, \underline{R}^{(k\leq n)}) = \frac{P(C, B, \underline{E}, \underline{R})}{P(\underline{R}^{(k\leq n)})}$$

No real distinction between inference and prediction

How does it work?

General case (more complicate 'network')
for example including uncertain Preparation ('C' not to confuse it with 'Probability') and a Reporter for each estraction ($\underline{R}$):

- Write down the joint distribution of all variables in the game:

$$P(C, B, \underline{E}, \underline{R})$$

$$\begin{array}{ll} \underline{E}: & E_i^{(1)}, E_i^{(2)}, E_i^{(3)}, \ldots \\ \underline{R}: & R_i^{(1)}, R_i^{(2)}, R_i^{(3)}, \ldots \end{array}$$

- Condition on the 'observations':

$$P(C, B, \underline{E}, \underline{R}^{(k>n)} \,|\, \underline{R}^{(k\leq n)}) = \frac{P(C, B, \underline{E}, \underline{R})}{P(\underline{R}^{(k\leq n)})}$$

No real distinction between inference and prediction

(We shall see it later in the case of *continuous distributions*)

Bayes' *billiard*

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

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;

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

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

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;

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.

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 many balls stopped left and how many stopped right

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 many balls stopped left and how many stopped right

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 many balls stopped left and how many stopped right

Although Bayes never mentioned a *billiard*

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 many balls stopped left and how many stopped right

Although Bayes never mentioned a *billiard*

Bayes' billiard and Bernoulli trials

It is easy to recognize 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 recognize 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 recognize 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]:

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

Bayes' billiard and Bernoulli trials

It is easy to recognize 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]:

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

Bayes' billiard and Bernoulli trials

It is easy to recognize 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 recognize 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]:

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

Bayes' billiard and Bernoulli trials

It is easy to recognize 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]:

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

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 continuous 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

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

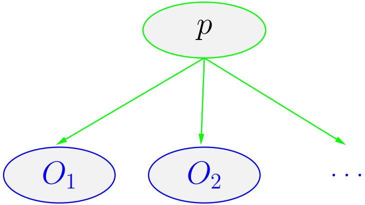
$\Rightarrow$ p continuous in $[0, 1]$

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

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

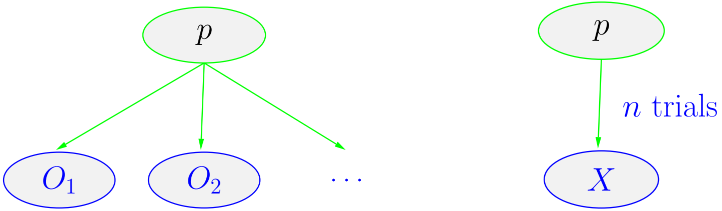
Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p



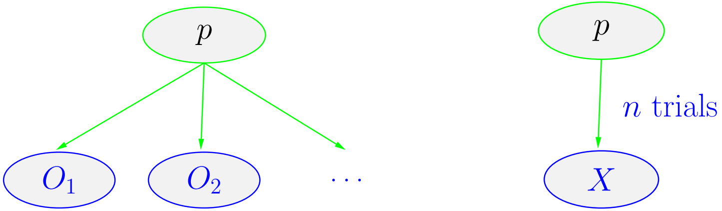
Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p



Inferring Bernoulli's trial parameter p

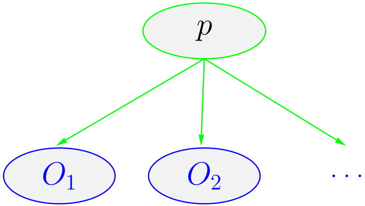
Making several independent trials *assuming* the same p



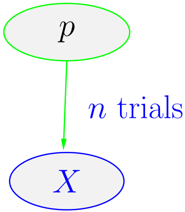
"independent Bernoulli trials" "binomial distribution"

Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p



"independent Bernoulli trials"

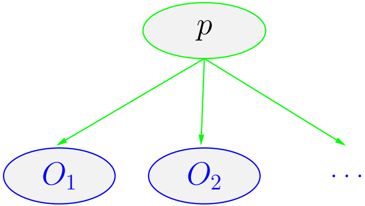


"binomial distribution"

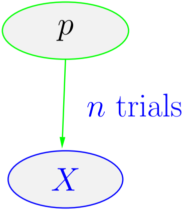
⇒ In the light of the experimental information
there will be values of p we shall believe more,
and others we shall believe less.

Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p



"independent Bernoulli trials"



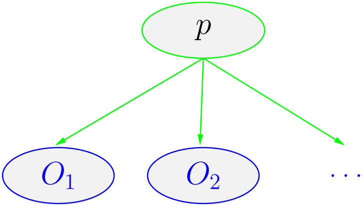
"binomial distribution"

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

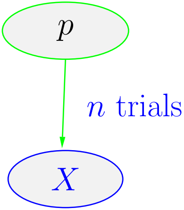
$P(p_i \,|\, X, n)$
$f(p \,|\, X, n)$

Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p



"independent Bernoulli trials"



"binomial distribution"

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

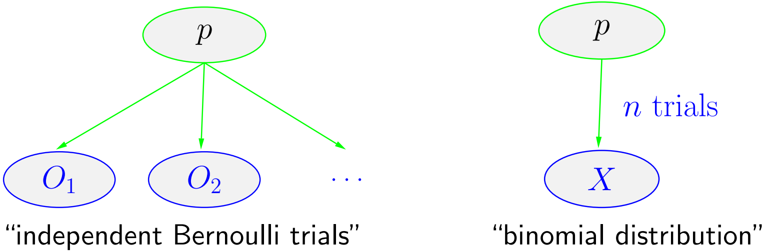
$$\propto f(O_1, O_2, \ldots \,|\, p) \cdot f_0(p)$$

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

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

Inferring Bernoulli's trial parameter p

Making several independent trials *assuming* the same p



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

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

Are the two inferences the same?

(not obvious in principle)

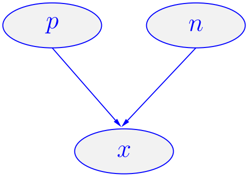
n independent Bernoulli processes

General case

n independent Bernoulli processes

General case

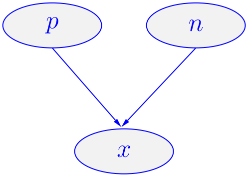
Model



n independent Bernoulli processes

General case

Model



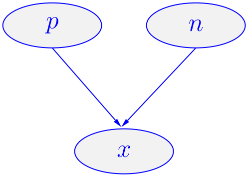
Joint pdf (omitting background condition I):

$$f(x, p, n) = f(x \mid p, n)$$

n independent Bernoulli processes

General case

Model



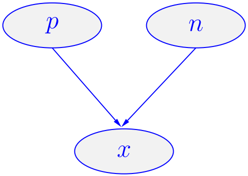
Joint pdf (omitting background condition I):

$$f(x, p, n) \quad = \quad f(x \,|\, p, n) \cdot f(p, n)$$

n independent Bernoulli processes

General case

Model



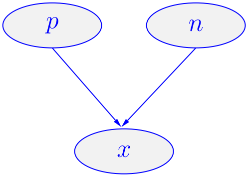
Joint pdf (omitting background condition I):

$$
\begin{aligned}
f(x, p, n) &= f(x \,|\, p, n) \cdot f(p, n) \\
&= f(x \,|\, p, n) \cdot f(p \,|\, n) \cdot f(n)
\end{aligned}
$$

n independent Bernoulli processes

General case

Model



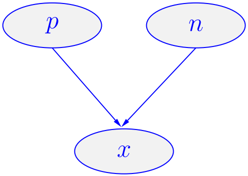
Joint pdf (omitting background condition I):

$$
\begin{aligned}
f(x, p, n) &= f(x \,|\, p, n) \cdot f(p, n) \\
&= f(x \,|\, p, n) \cdot f(p \,|\, n) \cdot f(n) \\
&= f(x \,|\, p, n) \cdot f(n \,|\, p) \cdot f(p)
\end{aligned}
$$

n independent Bernoulli processes

General case

Model



Joint pdf (omitting background condition I):

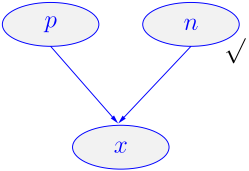
$$
\begin{aligned}
f(x, p, n) &= f(x \,|\, p, n) \cdot f(p, n) \\
&= f(x \,|\, p, n) \cdot f(p \,|\, n) \cdot f(n) \\
&= f(x \,|\, p, n) \cdot f(n \,|\, p) \cdot f(p) \\
&= f(x \,|\, p, n) \cdot f(p) \cdot f(n)
\end{aligned}
$$

(n and p are independent)

n independent Bernoulli processes

Usual case → *n* fixed (for the moment)

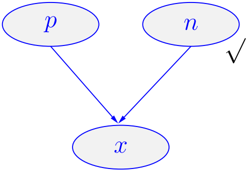
Model



n independent Bernoulli processes

Usual case → n fixed (for the moment)

Model



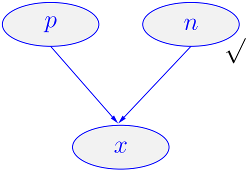
Joint pdf

$$f(x, p \mid n) \;=\; f(x \mid p, n) \cdot f(p)$$

n independent Bernoulli processes

Usual case → n fixed (for the moment)

Model



Joint pdf

$$f(x, p \,|\, n) \;=\; f(x \,|\, p, n) \cdot f(p)$$

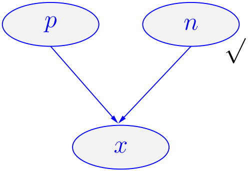
Typical problems

- p is assumed → interested in $f(x \,|\, n, p)$
  → well known binomial;

n independent Bernoulli processes

Usual case → n fixed (for the moment)

Model



Joint pdf

$$f(x, p \mid n) \;=\; f(x \mid p, n) \cdot f(p)$$

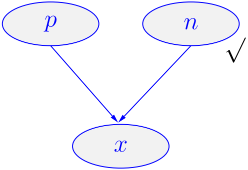
Typical problems

- p is assumed → interested in $f(x \mid n, p)$
  → well known binomial;
- x is assumed ('observed')

n independent Bernoulli processes

Usual case → n fixed (for the moment)

Model



Joint pdf

$$f(x, p \mid n) \;=\; f(x \mid p, n) \cdot f(p)$$

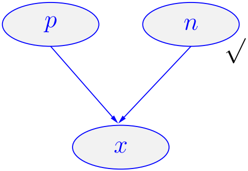
Typical problems

- p is assumed → interested in $f(x \mid n, p)$
  → well known binomial;
- x is assumed ('observed') → $f(p \mid n, x)$

n independent Bernoulli processes

Usual case → n fixed (for the moment)

Model



Joint pdf

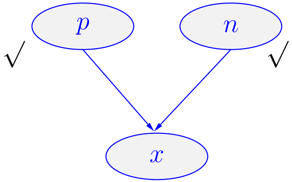
$$f(x, p \mid n) \;=\; f(x \mid p, n) \cdot f(p)$$

Typical problems

- p is assumed → interested in $f(x \mid n, p)$
  → well known binomial;
- x is assumed ('observed') → $f(p \mid n, x)$:
  → ?

n independent Bernoulli processes

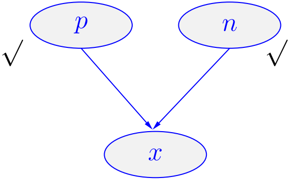
Graphical models of the typical problems



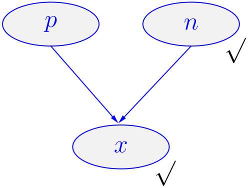
$\rightarrow\; f(x \mid n, p)$

n independent Bernoulli processes

Graphical models of the typical problems



$\rightarrow f(x \,|\, n, p)$



$\rightarrow f(p \,|\, n, x)$

The End