Introduction to Probabilistic Reasoning

2. From uncertain events to uncertain numbers

Giulio D'Agostini

Dipartimento di Fisica

Università di Roma La Sapienza

Preambolo

Ched'è la statistica?
(Trilussa)

$\Rightarrow$ preambolo a lezioni al CERN

Preamble

Title of the lectures (“Telling the truth with statistics”)

Preamble

Title of the lectures (“Telling the truth with statistics”)

- proposed by organizers $\rightarrow$ accepted...

Preamble

Title of the lectures ("Telling the truth with statistics")

- proposed by organizers $\rightarrow$ accepted...
- I interpret it as a direct question, to which I will try to give my best answer

Preamble

Title of the lectures (“Telling the truth with statistics”)

- proposed by organizers $\rightarrow$ accepted...
- I interpret it as a direct question, to which I will try to give my best answer, quite frankly.
- How to interpret the question?

    1.

Preamble

Title of the lectures ("Telling the truth with statistics")

- proposed by organizers $\to$ accepted...
- I interpret it as a direct question, to which I will try to give my best answer, quite frankly.
- How to interpret the question?

1. ~~"Tell the Truth"?~~ $\Rightarrow$ Question to God
   - ~~What is the true value of a quantity?~~
   - ~~What is the true theory that describes the world?~~
2.

Preamble

Title of the lectures ("Telling the truth with statistics")

- proposed by organizers $\rightarrow$ accepted...
- I interpret it as a direct question, to which I will try to give my best answer, quite frankly.
- How to interpret the question?
  1. ~~"Tell the Truth"?~~ $\Rightarrow$ Question to God
     - ~~What is the true value of a quantity?~~
     - ~~What is the true theory that describes the world?~~
  2. ~~"Tell the truth" $\Longleftrightarrow$ "to lie"?~~ $\Rightarrow$ Not fair

Preamble

Title of the lectures ("Telling the truth with statistics")

- proposed by organizers → accepted...
- I interpret it as a direct question, to which I will try to give my best answer, quite frankly.
- How to interpret the question?
  1. ~~"Tell the Truth"?~~ ⇒ Question to God
     - ~~What is the true value of a quantity?~~
     - ~~What is the true theory that describes the world?~~
  2. ~~"Tell the truth" ⟺ "to lie"?~~ ⇒ Not fair, though

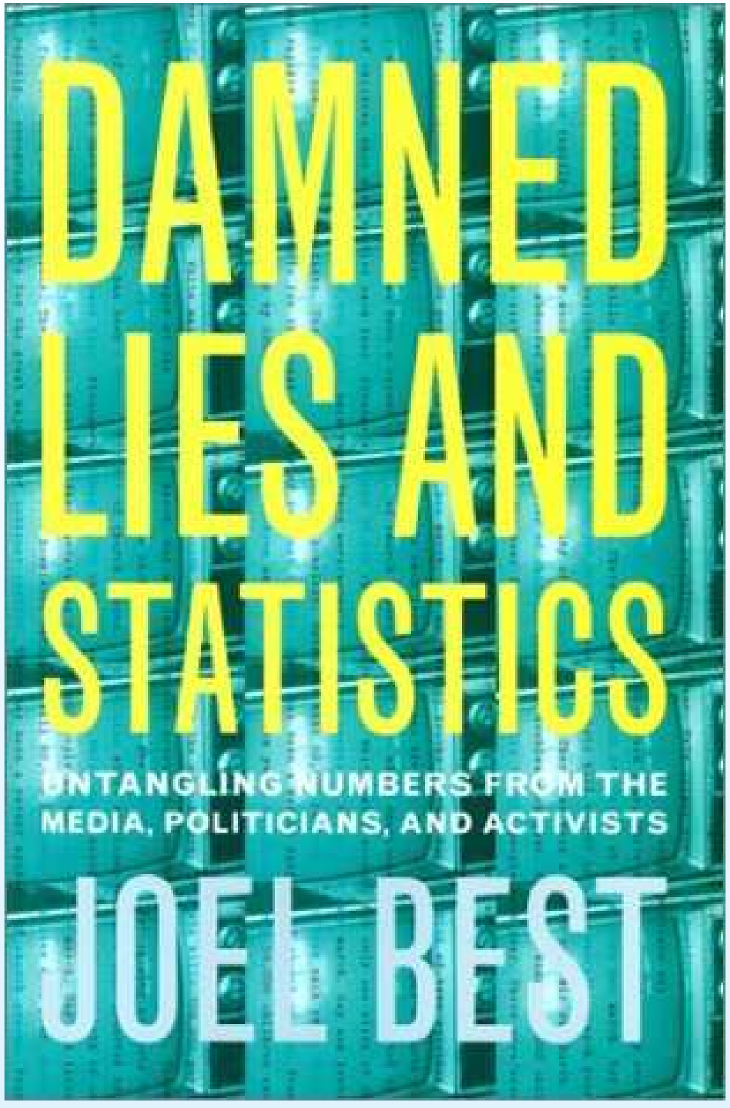
"There are three kinds of lies: lies, damn lies, and statistics"
(Benjamin Disraeli/Mark Twain)

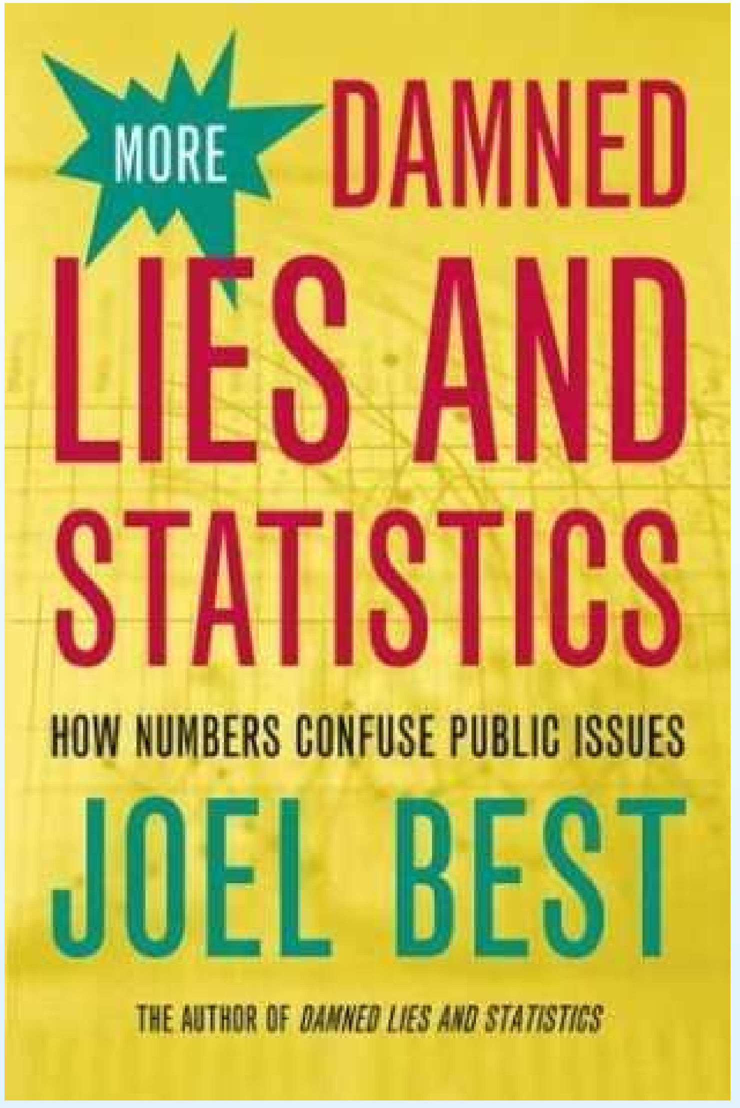
Damned lies and statistics

Well known subject

Damned lies and statistics

Well known subject, especially in marketing and politics





Defining the issue

What do we mean by “statistics”?

Defining the issue

What do we mean by “statistics”?

Usually several things:

- ***descriptive*** statistics [e.g. Webster’s (Kdict)]
  - “The science which has to do with the collection and classification of certain facts respecting the condition of the people in a state.”
  - “(pl.) Classified facts respecting the condition of the people in a state, their health, their longevity, . . . especially, those facts which can be stated in numbers, or in tables of numbers, or in any tabular and classified arrangement.”
    $\Rightarrow$ extended to scientific data.

Defining the issue

What do we mean by “statistics”?

Usually several things:

- ***descriptive*** statistics [e.g. Webster’s (Kdict)]
  - “The science which has to do with the collection and classification of certain facts respecting the condition of the people in a state.”
  - “(pl.) Classified facts respecting the condition of the people in a state, their health, their longevity, . . . especially, those facts which can be stated in numbers, or in tables of numbers, or in any tabular and classified arrangement.”
    $\Rightarrow$ <u>extended to scientific data</u>.
- ***Probability theory***
- ***Inference***

Defining the issue

What do we mean by “statistics”?

Usually several things:

- ***descriptive*** statistics [e.g. Webster’s (Kdict)]
  - “The science which has to do with the collection and classification of certain facts respecting the condition of the people in a state.”
  - “(pl.) Classified facts respecting the condition of the people in a state, their health, their longevity, . . . especially, those facts which can be stated in numbers, or in tables of numbers, or in any tabular and classified arrangement.”
    $\Rightarrow$ extended to scientific data.
- ***Probability theory***
- ***Inference*** $\Rightarrow$ primary interest to physicists

Defining the issue

What do we mean by "statistics"?

. . . and all together:

"A branch of applied mathematics concerned with the collection and interpretation of quantitative data and the use of probability theory to estimate population parameters" [WordNet (Kdict)]

Defining the issue

What do we mean by “statistics”?

. . . and all together:

> “A branch of applied mathematics concerned with the collection and interpretation of quantitative data and the use of probability theory to estimate population parameters” [WordNet (Kdict)]

$\Rightarrow$ inferential aspect enhanced

Defining the issue

What do we mean by “statistics”?

. . . and all together:

“A branch of applied mathematics concerned with the collection and interpretation of quantitative data and the use of probability theory to ”estimate population parameters [WordNet (Kdict)]

$\Rightarrow$ inferential aspect enhanced

Though we physicists are usually not interested in population parameters, but rather on physics quantities, theories, and so on.

Defining the issue

What do we mean by “statistics”?

. . . and all together:

“A branch of applied mathematics concerned with the collection and interpretation of quantitative data and the use of probability theory to ”estimate population parameters [WordNet (Kdict)]

$\Rightarrow$ inferential aspect enhanced

Though we physicists are usually not interested in population parameters, but rather on physics quantities, theories, and so on.

Inference: learning about theoretical objects from experimental observations (see later)

Where are the problems?

Descriptive statistics Little to comment, apart that the process of summarizing '*a State*' in a few numbers, in a diagram or in a table causes an enormous loss of detailed information, and this might lead to misunderstandings or even 'lies'.

$\Rightarrow$ the famous 'half chicken' joke.[†]

Where are the problems?

Descriptive statistics Little to comment, apart that the process of summarizing '*a State*' in a few numbers, in a diagram or in a table causes an enormous loss of detailed information, and this might lead to misunderstandings or even 'lies'.

$\Rightarrow$ the famous 'half chicken' joke.†

Probability theory Essentially OK, if we only consider the mathematical apparatus.

Where are the problems?

Descriptive statistics Little to comment, apart that the process of summarizing '*a State*' in a few numbers, in a diagram or in a table causes an enormous loss of detailed information, and this might lead to misunderstandings or even 'lies'.

$\Rightarrow$ the famous 'half chicken' joke.†

Probability theory Essentially OK, if we only consider the mathematical apparatus.

Inference Messy:

- Traditionally, a collection of *ad hoc* prescriptions

  ...accepted more by authority than by full awareness of what they mean

$\Rightarrow$ The physicist is confused† between good sense and statistics education

Where are the problems?

Descriptive statistics Little to comment, apart that the process of summarizing '*a State*' in a few numbers, in a diagram or in a table causes an enormous loss of detailed information, and this might lead to misunderstandings or even 'lies'.

$\Rightarrow$ the famous 'half chicken' joke.†

Probability theory Essentially OK, if we only consider the mathematical apparatus.

Inference Do better?

- Much improvement is gained if inference is grounded on probability theory

Where are the problems?

Descriptive statistics Little to comment, apart that the process of summarizing '*a State*' in a few numbers, in a diagram or in a table causes an enormous loss of detailed information, and this might lead to misunderstandings or even 'lies'.

$\Rightarrow$ the famous 'half chicken' joke.†

Probability theory Essentially OK, if we only consider the mathematical apparatus.

Inference Do better?

- Much improvement is gained if inference is grounded on probability theory
- Summaries of descriptive statistics can be used in those cases in which *statistical sufficiency* holds (e.g. when we use the sample arithmetic average and standard deviation, instead of the n data points)

Torniamo a noi

Punto della situazione

Outline of first meeting

- Brainstorm on ‘standard’ teaching of data analysis methods. Problems with confidence intervals and p-values
- Uncertainty, probability, decision.
- Causes $\longleftrightarrow$ Effects

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

  “Probability is either referred to real cases or it is nothing” (de Finetti).
- Probabilistic approach, but What is probability?
- Basic rules of probability and Bayes rule.
- Bayesian inference and its graphical representation:
  $\Rightarrow$ Bayesian networks
- Let us play for a while with the toy

Summary on probabilistic approach

- Probability means how much we believe something
- Probability depends on available information
  $\rightarrow$ subjective
- Probability values obey the following basic rules

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

- All the rest by logic

$\rightarrow$ And, please, be coherent!

Summary on probabilistic approach

- Probability means how much we believe something
- Probability depends on available information
  $\rightarrow$ subjective
- Probability values obey the following basic rules

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

- All the rest by logic

$\rightarrow$ And, please, be coherent!

$\Rightarrow$ more comments on $P(E \,|\, I)$ $\rightarrow$

Three boxes ‘paradox’

1. The guest and two contestants
2. The guest and one contestant

Three boxes ‘paradox’

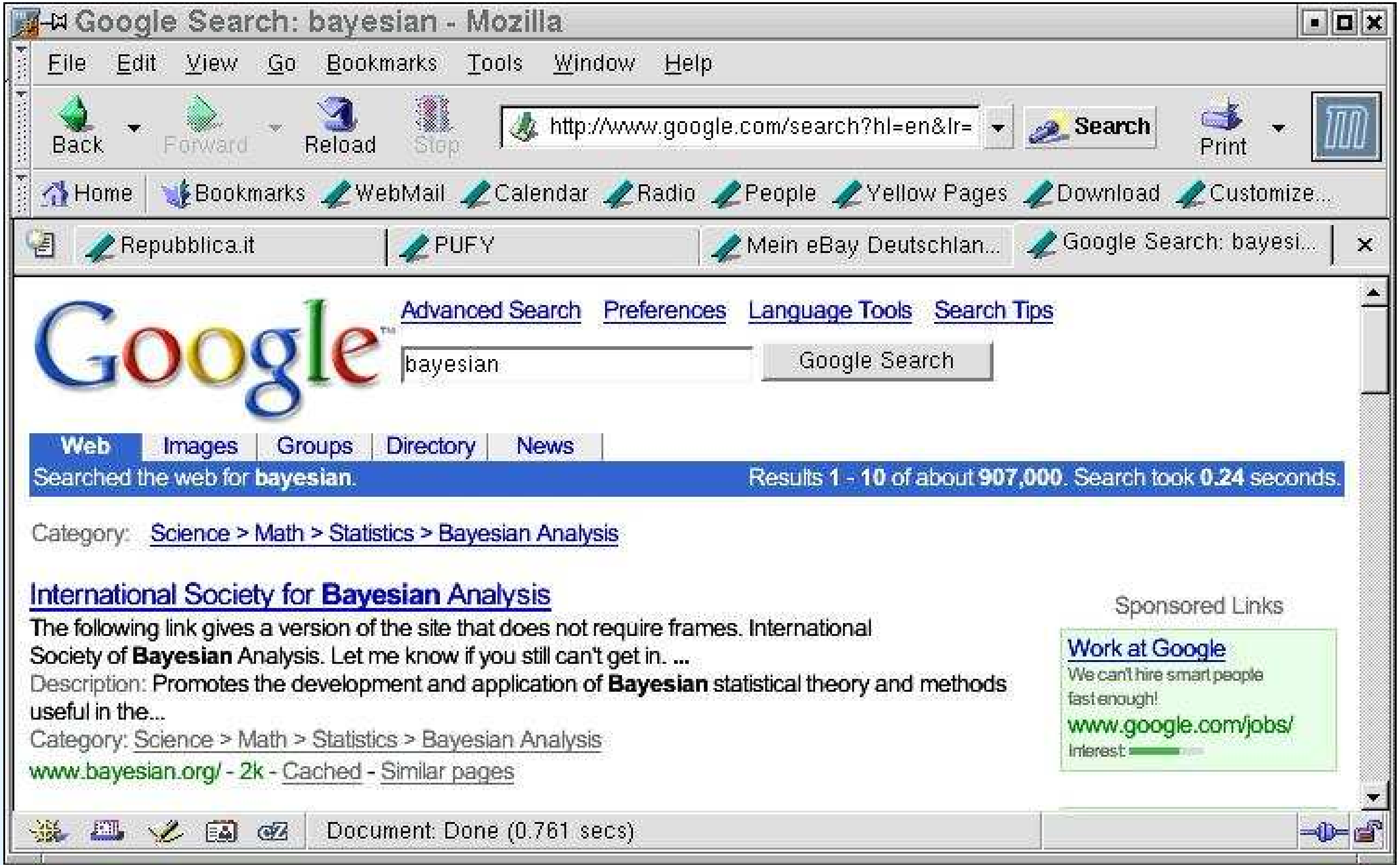
1. The guest and two contestants
2. The guest and one contestant

Nr. 2 $\rightarrow$ Monty Hall problem

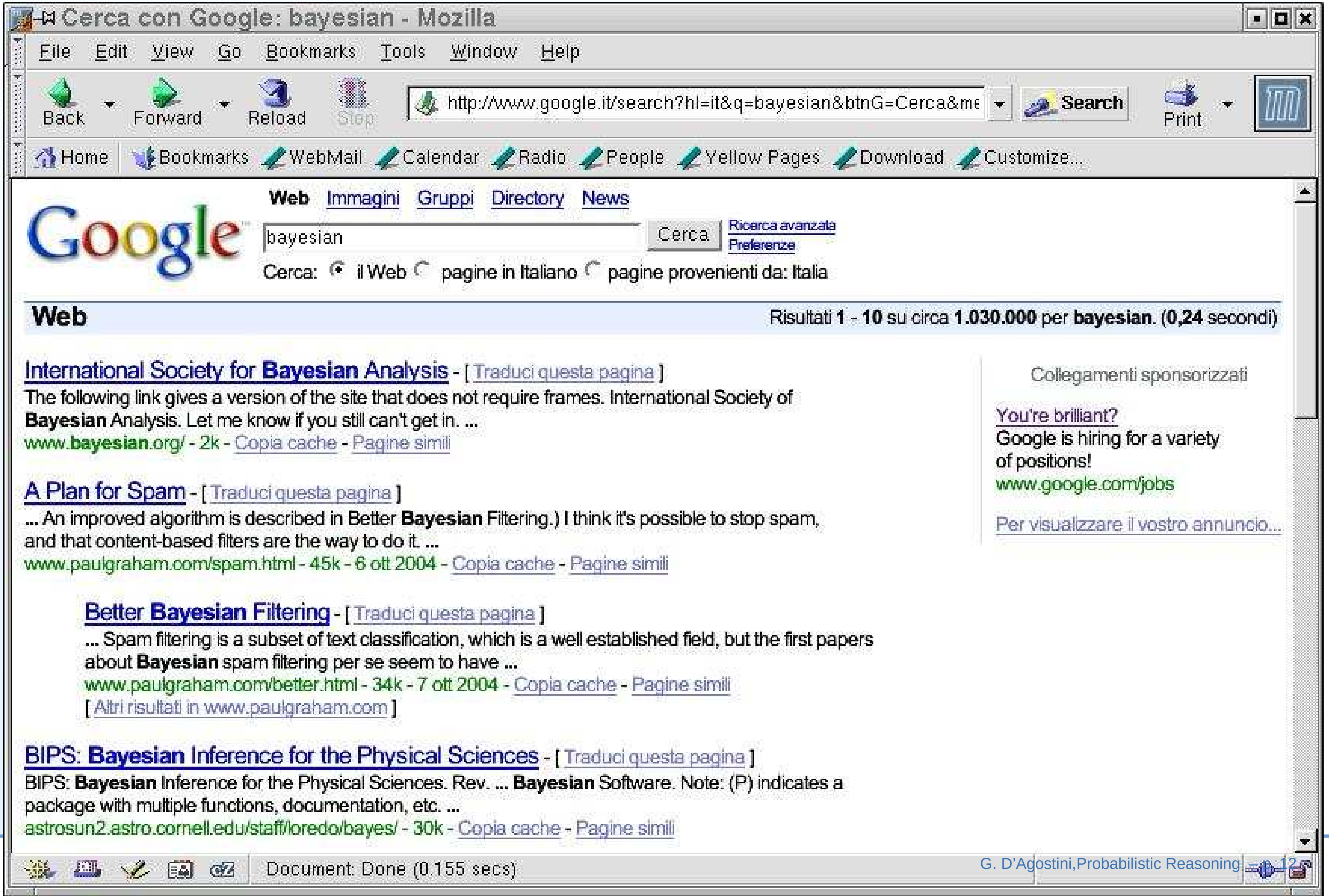
Conclusions on intro to probabilistic reasoning

- Subjective probability recovers intuitive idea of probability.
- Nothing negative in the adjective ’subjective’. Just recognize, honestly, that probability depends on the status of knowledge, different from person to person.
- Most general concept of probability that can be applied to a large variety of cases.
- The adjective Bayesian comes from the intense use of Bayes’ theorem to update probability once new data are acquired.
- Subjective probability is foundamental in decision issues, if you want to base decision on the probability of different events, together with the gain of each of them.
- Bayesian networks are powerful conceptuals/mathematical/ software tools to handle complex problems with variables related by probabilistic links.

Are Bayesians ‘smart’ and ‘brilliant’?



Are Bayesians ‘smart’ and ‘brilliant’?

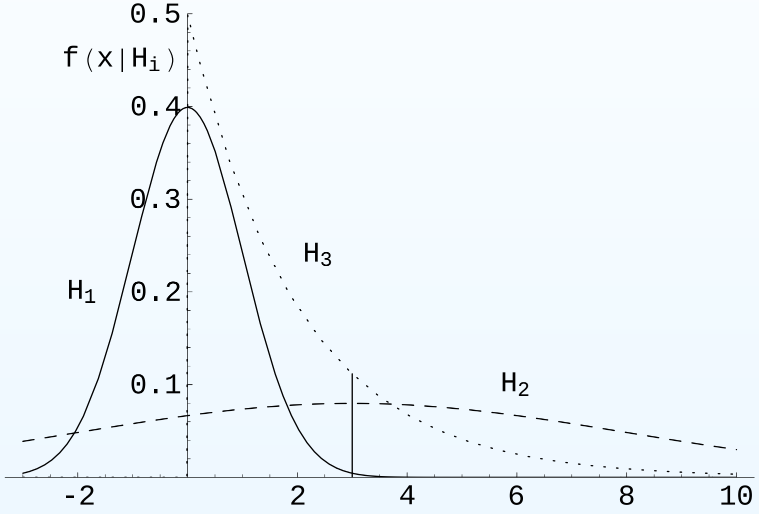


Further comments on first meeting

The three models example

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

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



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

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

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

A last remark

A last remark on model comparisons

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

A last remark

A last remark on model comparisons

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

A last remark

A last remark on model comparisons

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

A last remark

A last remark on model comparisons

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

Exercises and discussions

- Continue with six box problem [$\rightarrow$ *AJP* 67 (1999) 1260]

  $\rightarrow$ Slides

- <u>Home work 1</u>: AIDS problem $\rightarrow P(\text{HIV}\,|\,\text{Pos})$?

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

- <u>Home work 2</u>: Particle identification:

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

The hidden uniform

What was the mistake of people saying $P(\overline{\text{HIV}}\,|\,\text{Pos}) = 0.2$?

We can easily check that this is due to have set $\frac{\text{P}_\circ(\text{HIV})}{\text{P}_\circ(\overline{\text{HIV}})} = 1$,

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

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

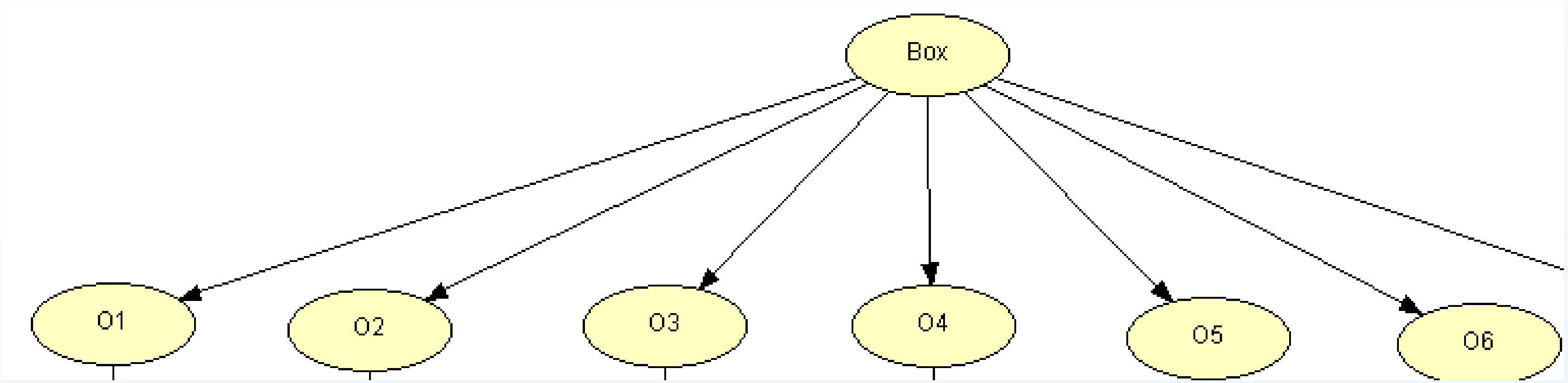
$\rightarrow$ "absence of priors" means in most times uniform priors over the all possible hypotheses

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

$\boxed{\text{Better methods based on ‘sand’ than methods based on nothing!}}$

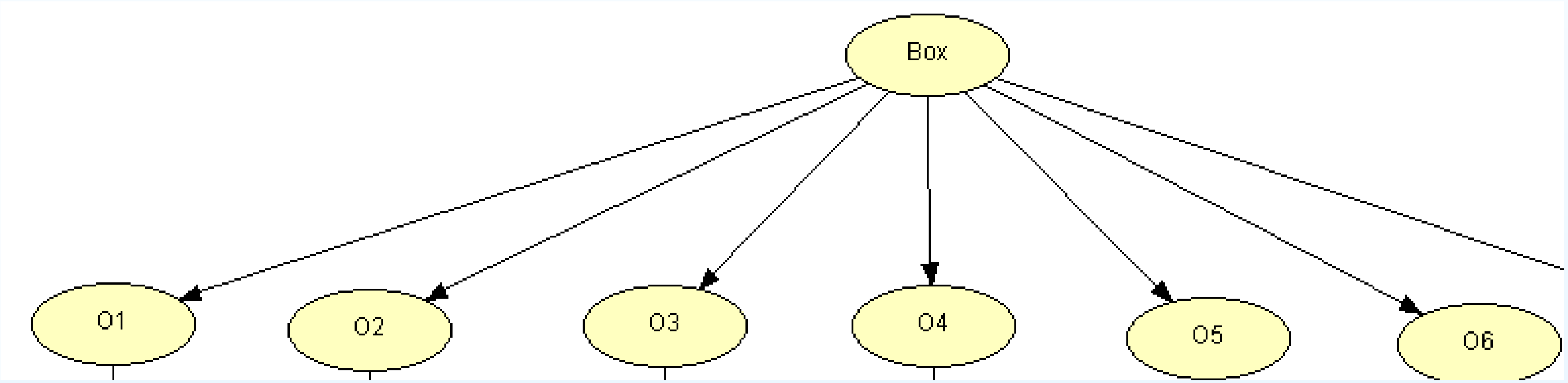
Cause-effect representation

box content $\rightarrow$ observed color



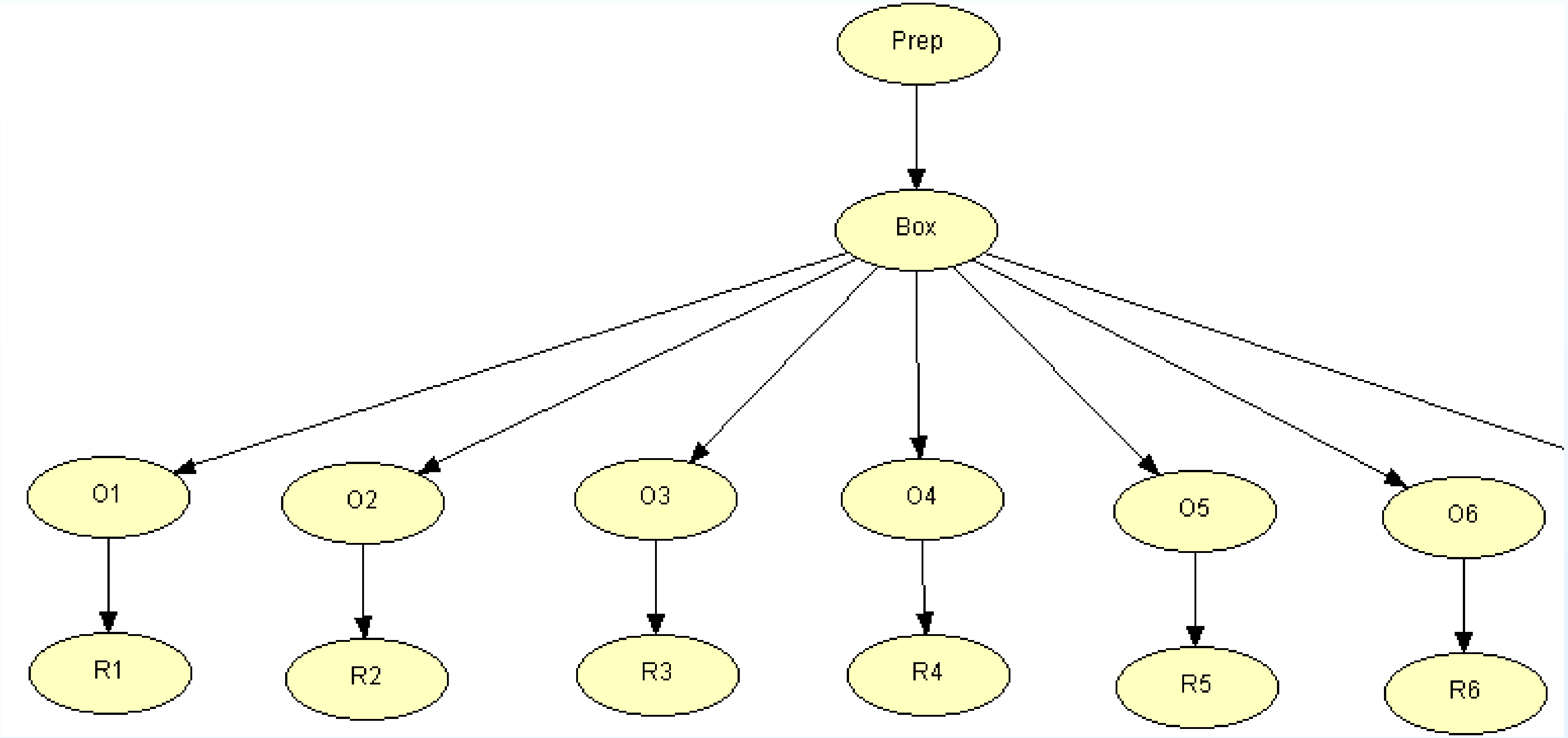
Cause-effect representation

box content $\rightarrow$ observed color

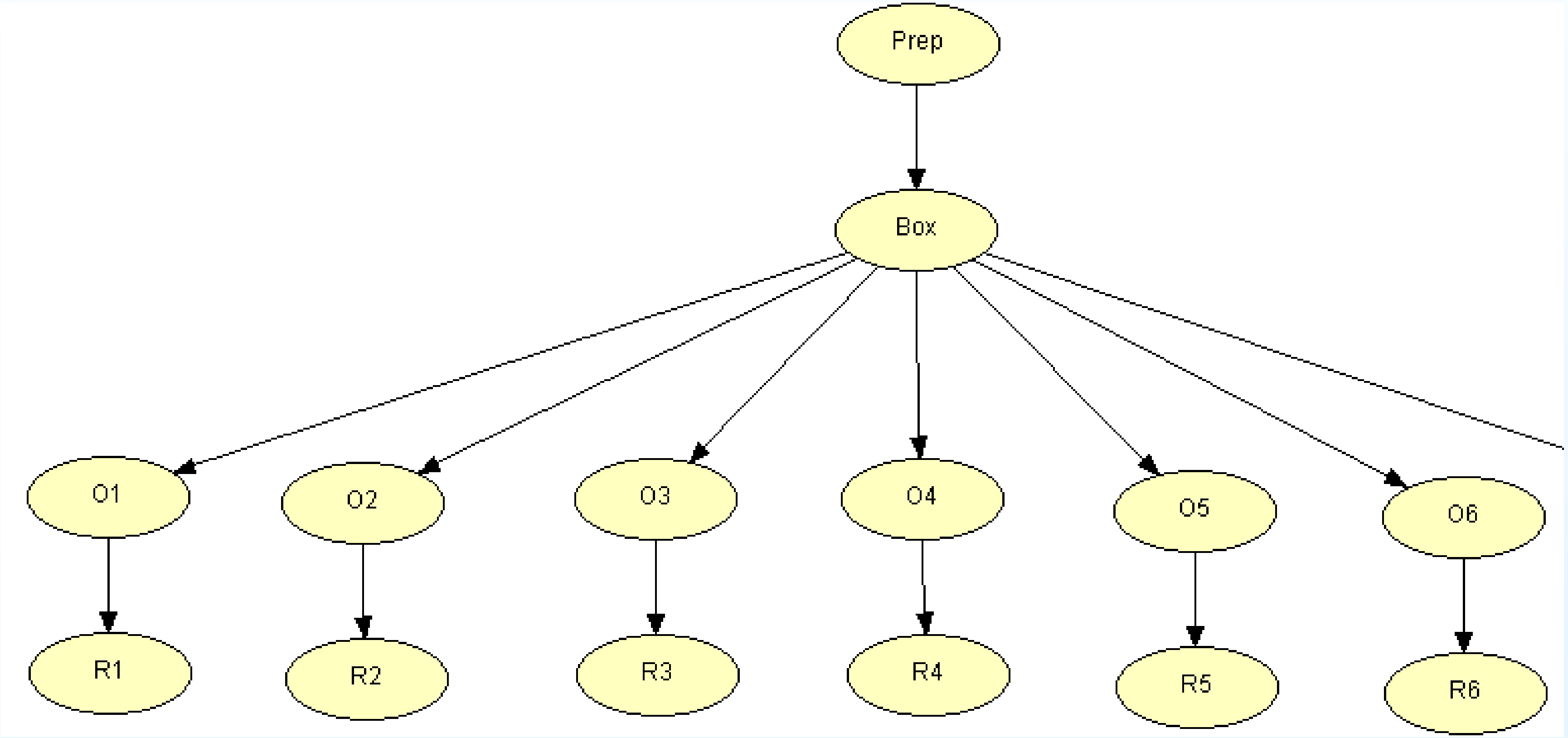


An effect might be the cause of another effect $\longrightarrow$

A network of causes and effects



A network of causes and effects



and so on. . .

⇒ Let's play with **Hugin**

from 6 boxes to 1001 boxes

Overview

- example with 1001 boxes
- from uncertainty on H_j to uncertainty on p_j (proportion):
  $P(H_j) \leftrightarrow P(p_j)$
- Physical meaning of p_j
- Probability Vs ‘Chance’ (‘propension’)
- The discretized Bayes billard.
- The extension to continuous values of p

Uncertain numbers

We are often uncertain in numbers and, consistently, we quantify of belief with probability.

Uncertain number is the more general term for random variable, though the adjective random is more committing, since it rely on the concept of randomness (see von Mises).

Nevertheless, I often use the name ‘random variable’, just to mean ’uncertain number’,

Uncertain numbers

We are often uncertain in numbers and, consistently, we quantify of belief with probability.

Uncertain number is the more general term for random variable, though the adjective random is more committing, since it rely on the concept of randomness (see von Mises).

Nevertheless, I often use the name ‘random variable’, just to mean ’uncertain number’, i.e.

A number respect to which we are in condition of uncertainty

Uncertain numbers

We are often uncertain in numbers and, consistently, we quantify of belief with probability.

Uncertain number is the more general term for random variable, though the adjective random is more committing, since it rely on the concept of randomness (see von Mises).

Nevertheless, I often use the name ‘random variable’, just to mean ’uncertain number’, i.e.

A number respect to which we are in condition of uncertainty

- The first number rolling a die
- The temperature at the Rome airport (FCO) tomorrow at 7:00 am
- The height of the next person who enters this room

Uncertain numbers

We are often uncertain in numbers and, consistently, we quantify of belief with probability.

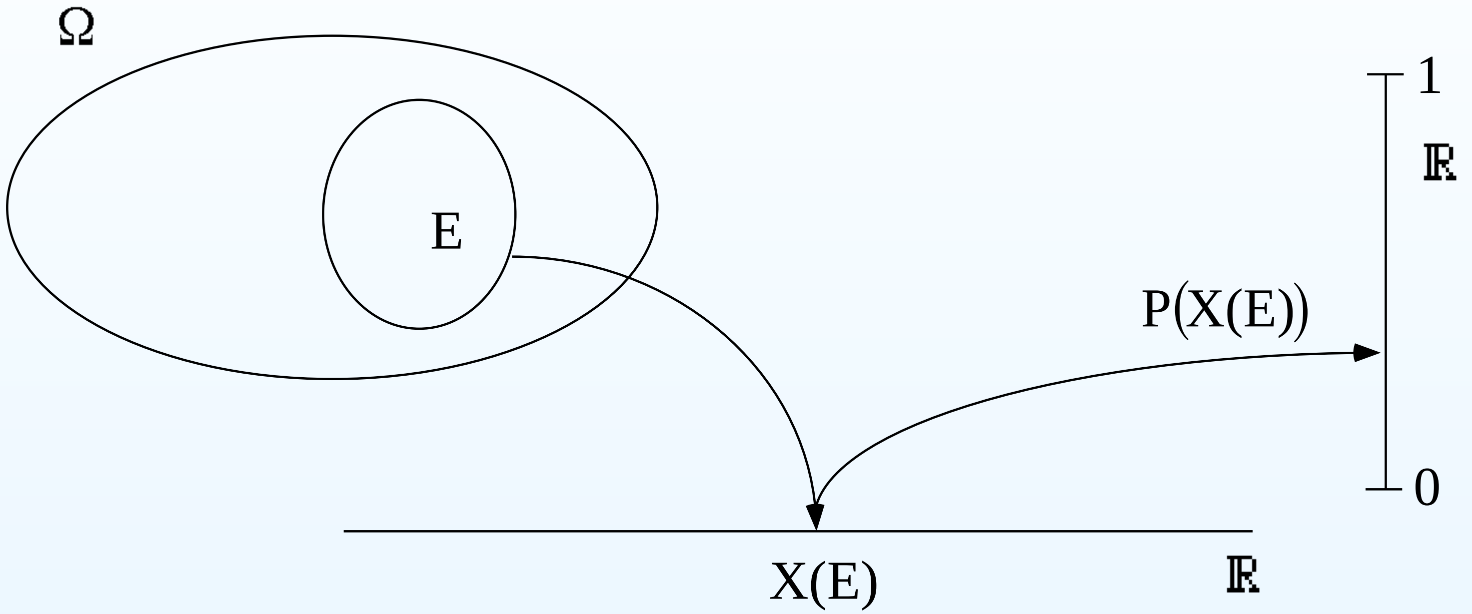
Uncertain number is the more general term for random variable, though the adjective random is more committing, since it rely on the concept of randomness (see von Mises).

Nevertheless, I often use the name ‘random variable’, just to mean ’uncertain number’, i.e.

> *A number respect to which we are in condition of uncertainty*

- No need that the numbers can be framed in a von Mises' *collective*
- But it must be a well defined number (any uncertainty on its definition will increase our uncertainty about it)

From events to uncertain numbers

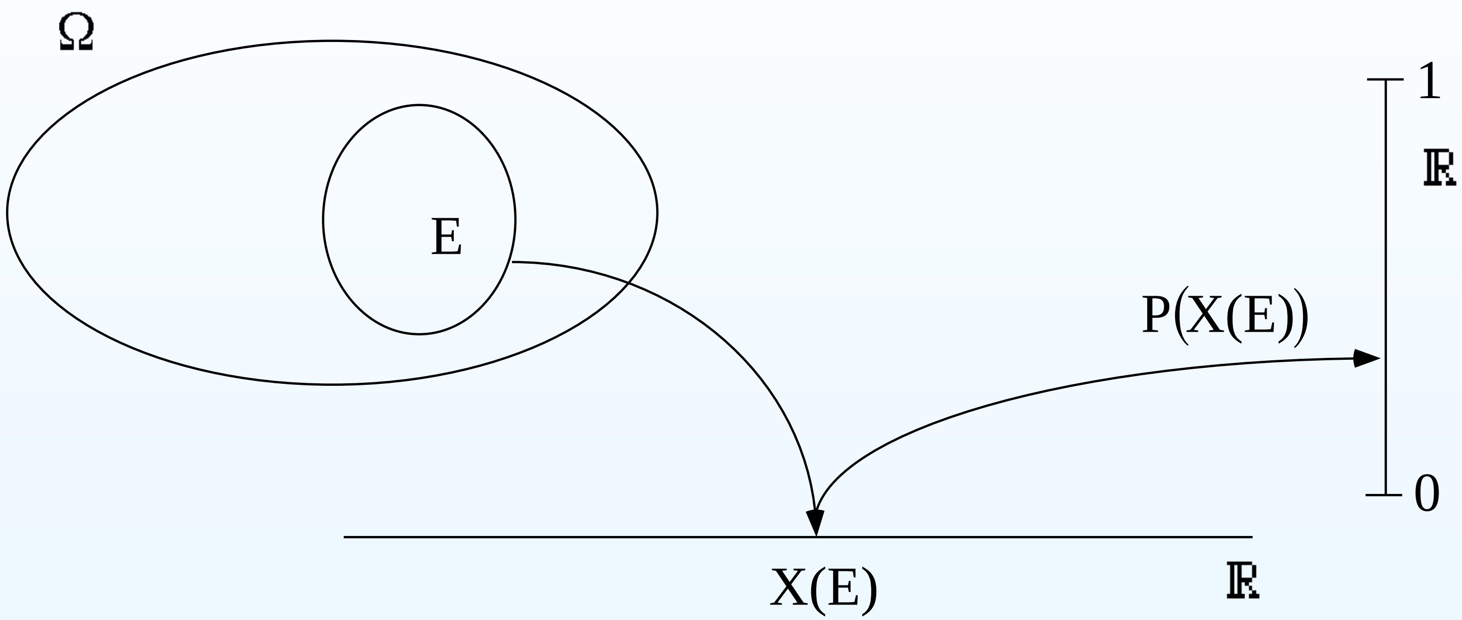


Uncertain numbers are associated to events

- Rolling one die: $X = 4 \leftrightarrow$ ‘face marked with 4’
  (note: no intrinsic order in the numbers associated to a die)

$\rightarrow$ $P(X = 4) = P($‘face marked with 4’$)$

From events to uncertain numbers



Uncertain numbers are associated to events

Event $\rightarrow$ number: univocal, but not bi-univocal

- Rolling two dice, with X 'sum of results'

$\rightarrow$ $P(X = 4) = \sum P($'events giving 4'$)$

Probability function (discrete numbers)

To each possible value of X we associate a degree of belief:

$$f(x) = P(X = x)\,.$$

Probability function (discrete numbers)

To each possible value of X we associate a degree of belief:

$$f(x) = P(X = x)\,.$$

$f(x)$, being a probability, must satisfy the following properties:

$$0 \le f(x_i) \le 1\,,$$

$$P(X = x_i \cup X = x_j) = f(x_i) + f(x_j)\,,$$

$$\sum_i f(x_i) = 1\,.$$

Probability function (discrete numbers)

To each possible value of X we associate a degree of belief:

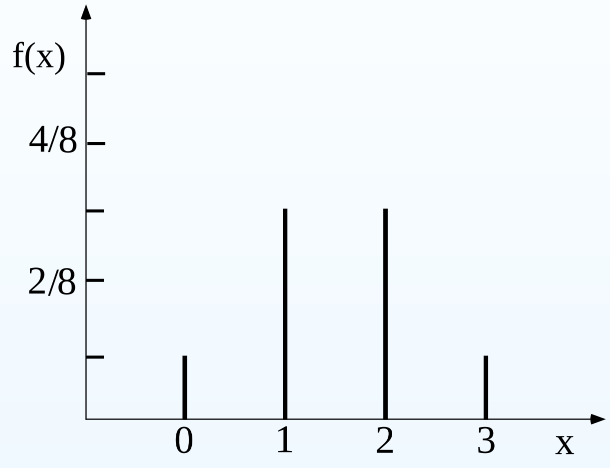
$$f(x) = P(X = x)\,.$$

$f(x)$, being a probability, must satisfy the following properties:

$$0 \leq f(x_i) \leq 1\,,$$

$$P(X = x_i \cup X = x_j) = f(x_i) + f(x_j)\,,$$

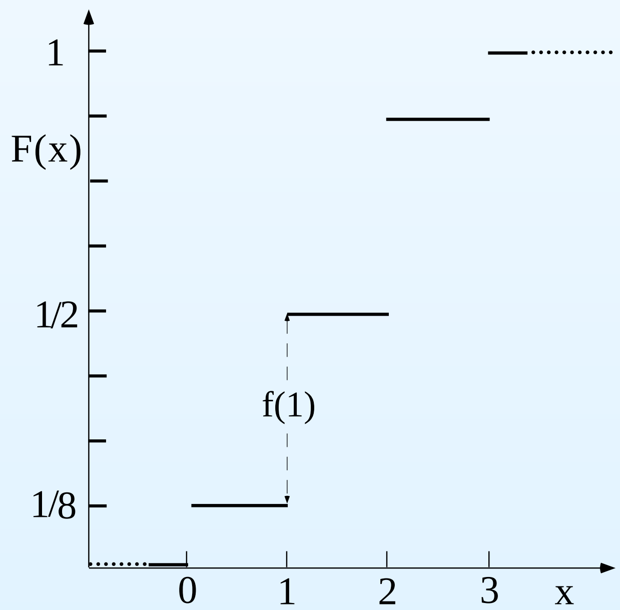
$$\sum_i f(x_i) = 1\,.$$



Cumulative function (defined for all x)

$$F(x_k) \equiv P(X \leq x_k) = \sum_{x_i \leq x_k} f(x_i)\,.$$

$[\,F(-\infty) = 0$; $F(+\infty) = 1$;
$F(x_i) - F(x_{i-1}) = f(x_i)$;
$\lim_{\epsilon\to 0} F(x+\epsilon) = F(x)\,]$



First intro to Monte Carlo

How to generate numbers in a way that their chance of occurring are proportional to $f(x)$?

- simple consideration based on the graphical representation of
  - $f(x)$
  - $F(x)$
  - extention to continuous functions
- a curious game, trowing stones...

Some simple examples

- Discrete uniform, well known $\rightarrow f(x) = 1/n \quad (1 \leq X \leq n)$

Some simple examples

- Discrete uniform, well known $\rightarrow f(x) = 1/n \quad (1 \leq X \leq n)$
- Bernoulli process
  - $X : 0, 1$ (failure/success)

    $f(0) = 1 - p$

    $f(1) = p$
  - it seems of practical irrelevance,
    $\rightarrow$ but of primary importance

Some simple examples

- Discrete uniform, well known $\rightarrow f(x) = 1/n \quad (1 \leq X \leq n)$
- Bernoulli process
  - $X : 0, 1$ (failure/success)
    $f(0) = 1 - p$
    $f(1) = p$
  - it seems of practical irrelevance,
    $\rightarrow$ but of primary importance
- The drunk man problem
  - Six keys (like rolling a die)
  - After each trial he 'loses memory'
  - We watch him and – cynically – bet on the attempt on which he will succeed:
  - $X = 1$, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, ... ?

Some simple examples

- Discrete uniform, well known $\rightarrow f(x) = 1/n \quad (1 \leq X \leq n)$
- Bernoulli process
  - $X : 0, 1$ (failure/success)

    $f(0) = 1 - p$

    $f(1) = p$
  - it seems of practical irrelevance,
    $\rightarrow$ but of primary importance
- The drunk man problem
  - Six keys (like rolling a die)
  - After each trial he 'loses memory'
  - We watch him and – cynically – bet on the attempt on which he will succeed:
  - $X = 1$, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, . . . ?

$\rightarrow$ On which number would you bet?

Propagating probability values

We cannot say "we consider all values of X equally likelly because all attempts are equally likely"...

Propagating probability values

We cannot say "we consider all values of X equally likelly because all attempts are equally likely"...

- $\rightarrow$ what is constant is $P(E_i \,|\, I) = p$,
  where E_i is the success in the i-th attempt.
- $\rightarrow$ instead, "$X = i$" stands for first success in the attempt i, i.e.
  $\overline{E}_1 \cap \overline{E}_2 \cap \cdots \cap \overline{E}_{i-1} \cap E_i$.

Propagating probability values

We cannot say "we consider all values of X equally likelly because all attempts are equally likely"...

- $\rightarrow$ what is constant is $P(E_i \,|\, I) = p$, where E_i is the success in the i-th attempt.
- $\rightarrow$ instead, "$X = i$" stands for first success in the attempt i, i.e. $\overline{E}_1 \cap \overline{E}_2 \cap \cdots \cap \overline{E}_{i-1} \cap E_i$.

How to evaluate $f(i)$, i.e. $P(X = i)$?

Propagating probability values

We cannot say "we consider all values of X equally likelly because all attempts are equally likely"...

→ what is constant is $P(E_i \,|\, I) = p$,
where E_i is the success in the i-th attempt.

→ instead, "$X = i$" stands for first success in the attempt i, i.e.
$\overline{E}_1 \cap \overline{E}_2 \cap \cdots \cap \overline{E}_{i-1} \cap E_i$.

How to evaluate $f(i)$, i.e. $P(X = i)$?

⇒ Beliefs are framed in a network!

- Once we assess something, we are implicitly making an infinite number of assessments concerning logically connected events!
- We only need to make them explicit, using logic:

Propagating probability values

We cannot say "we consider all values of X equally likelly because all attempts are equally likely"...

$\rightarrow$ what is constant is $P(E_i \,|\, I) = p$,
where E_i is the success in the i-th attempt.

$\rightarrow$ instead, "$X = i$" stands for first success in the attempt i, i.e.
$\overline{E}_1 \cap \overline{E}_2 \cap \cdots \cap \overline{E}_{i-1} \cap E_i$.

How to evaluate $f(i)$, i.e. $P(X = i)$?

- In this case, simply chain rule:

$$P(X = 2) = P(\overline{E}_1 \cap E_2) = P(\overline{E}_1) \cdot P(E_2 \,|\, \overline{E}_1);$$

$$P(X = 3) = P(\overline{E}_1) \cdot P(\overline{E}_2 \,|\, \overline{E}_1) \cdot P(E_3 \,|\, \overline{E}_1, \overline{E}_2);$$

etc.

Propagating probability values

We cannot say "we consider all values of X equally likelly because all attempts are equally likely"...

- $\rightarrow$ what is constant is $P(E_i \,|\, I) = p$, where E_i is the success in the i-th attempt.
- $\rightarrow$ instead, "$X = i$" stands for first success in the attempt i, i.e. $\overline{E}_1 \cap \overline{E}_2 \cap \cdots \cap \overline{E}_{i-1} \cap E_i$.

How to evaluate $f(i)$, i.e. $P(X = i)$?

[BUT sometimes the math might be hard:

- $\rightarrow$ fortunatly, nowadays most tough 'direct probability' problems can be easily solved by simulation ("Monte Carlo" methods)]

Building up $f(x)$ of the drunk man problem

Building up $f(x)$ of the drunk man problem

$$f(1) \quad = \quad P(E_1) = p$$

Building up $f(x)$ of the drunk man problem

$$
\begin{aligned}
f(1) &= P(E_1) = p \\
f(2) &= P(\overline{E}_1) \cdot P(E_2 \,|\, \overline{E}_1) = (1-p)\,p
\end{aligned}
$$

Building up $f(x)$ of the drunk man problem

$$
\begin{aligned}
f(1) &= P(E_1) = p \\
f(2) &= P(\overline{E}_1) \cdot P(E_2 \,|\, \overline{E}_1) = (1-p)\,p \\
f(3) &= P(\overline{E}_1) \cdot P(\overline{E}_2 \,|\, \overline{E}_1) \cdot P(E_3 \,|\, \overline{E}_1, \overline{E}_2) = (1-p)^2\,p
\end{aligned}
$$

Building up $f(x)$ of the drunk man problem

$$
\begin{aligned}
f(1) &= P(E_1) = p \\
f(2) &= P(\overline{E}_1)\cdot P(E_2\,|\,\overline{E}_1) = (1-p)\,p \\
f(3) &= P(\overline{E}_1)\cdot P(\overline{E}_2\,|\,\overline{E}_1)\cdot P(E_3\,|\,\overline{E}_1,\overline{E}_2) = (1-p)^2\,p \\
\ldots &\ \ldots\ \ldots \\
f(x) &= p\,(1-p)^{x-1}
\end{aligned}
$$

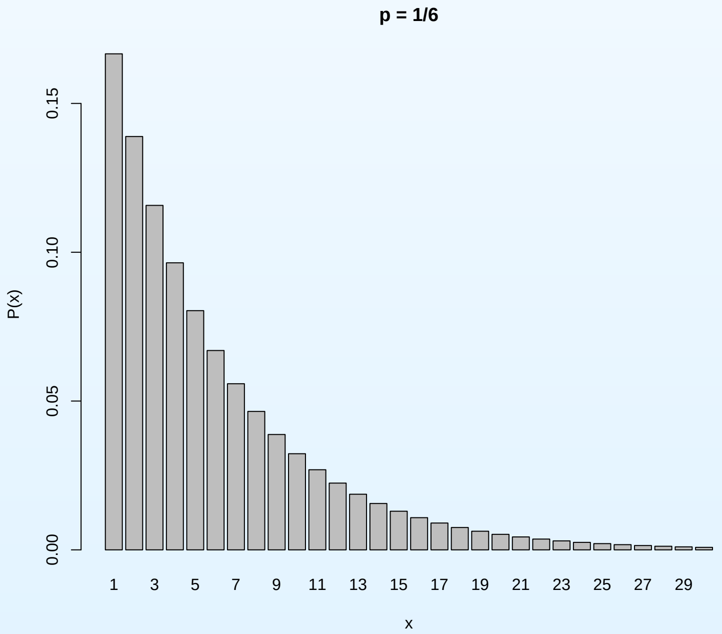
Building up $f(x)$ of the drunk man problem

$$\begin{aligned}
f(1) &= P(E_1) = p \\
f(2) &= P(\overline{E}_1)\cdot P(E_2\,|\,\overline{E}_1) = (1-p)\,p \\
f(3) &= P(\overline{E}_1)\cdot P(\overline{E}_2\,|\,\overline{E}_1)\cdot P(E_3\,|\,\overline{E}_1,\overline{E}_2) = (1-p)^2\,p \\
\dots &\;\dots\; \dots \\
f(x) &= p\,(1-p)^{x-1}
\end{aligned}$$

Beliefs decrease geometrically

$\Rightarrow$ Geometric distribution

$[p = 1/6]$

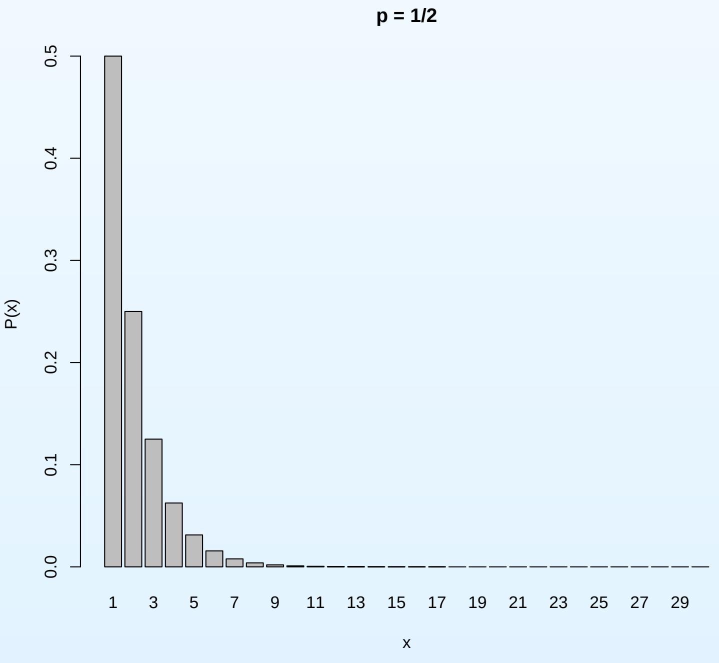


Building up $f(x)$ of the drunk man problem

$$
\begin{aligned}
f(1) &= P(E_1) = p \\
f(2) &= P(\overline{E}_1) \cdot P(E_2 \,|\, \overline{E}_1) = (1-p)\, p \\
f(3) &= P(\overline{E}_1) \cdot P(\overline{E}_2 \,|\, \overline{E}_1) \cdot P(E_3 \,|\, \overline{E}_1, \overline{E}_2) = (1-p)^2\, p \\
\ldots & \ldots \ldots \\
f(x) &= p\,(1-p)^{x-1}
\end{aligned}
$$

$p = 1/2 \rightarrow$ tossing a coin

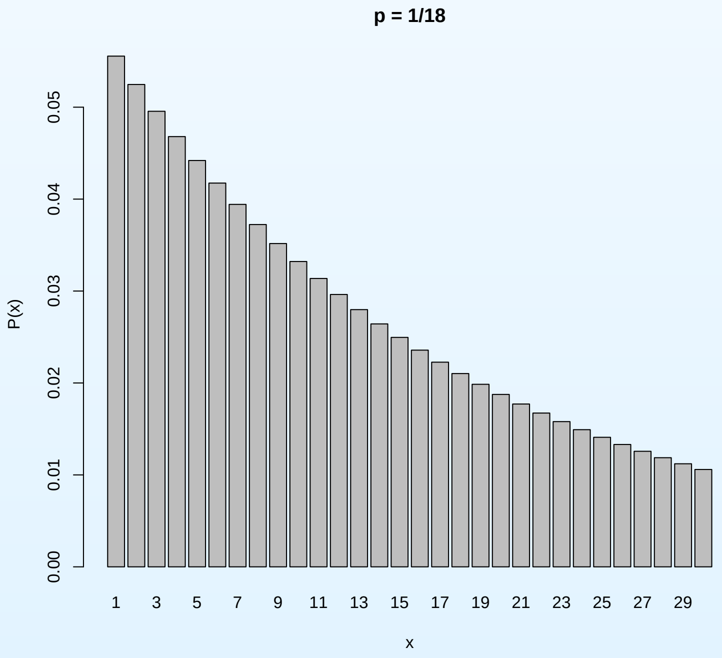
[Note: $f(x) = \left(\frac{1}{2}\right)^x$]



Building up $f(x)$ of the drunk man problem

$$\begin{aligned}
f(1) &= P(E_1) = p \\
f(2) &= P(\overline{E}_1)\cdot P(E_2\,|\,\overline{E}_1) = (1-p)\,p \\
f(3) &= P(\overline{E}_1)\cdot P(\overline{E}_2\,|\,\overline{E}_1)\cdot P(E_3\,|\,\overline{E}_1,\overline{E}_2) = (1-p)^2\,p \\
\ldots &\ \ldots\ \ldots \\
f(x) &= p\,(1-p)^{x-1}
\end{aligned}$$

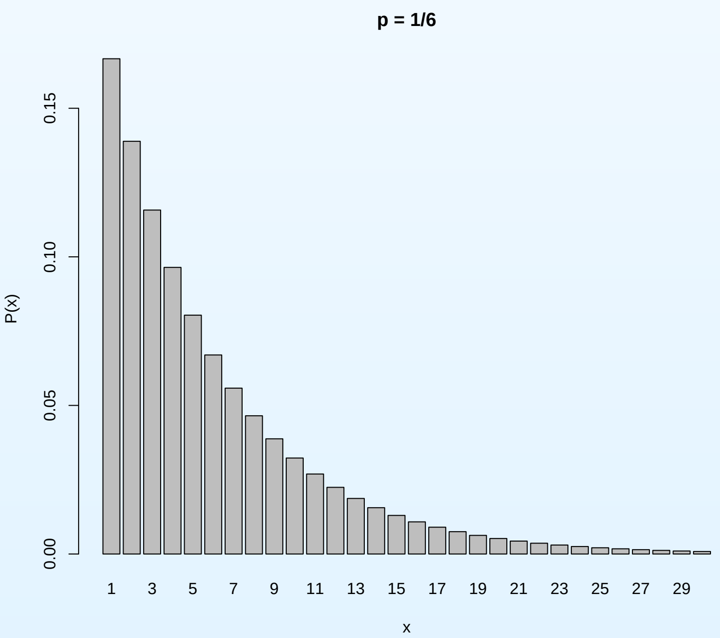
$p = 1/18 \rightarrow$ a particular number at the Italian lotto ($p = 5/90$)



Building up $f(x)$ of the drunk man problem

$$
\begin{aligned}
f(1) &= P(E_1) = p \\
f(2) &= P(\overline{E}_1)\cdot P(E_2\,|\,\overline{E}_1) = (1-p)\,p \\
f(3) &= P(\overline{E}_1)\cdot P(\overline{E}_2\,|\,\overline{E}_1)\cdot P(E_3\,|\,\overline{E}_1,\overline{E}_2) = (1-p)^2\,p \\
\ldots &\quad \ldots \quad \ldots \\
f(x) &= p\,(1-p)^{x-1}
\end{aligned}
$$

Most probable value does not depend on p.
Not a suitable indicator to state our expectation
The same is true for the range of possibilities:
$X : 1, 2, \ldots, \infty$



Prevision and prevision uncertainty

More suitable quantity to summarize in two numbers the our probabilistic ‘expectation’ and its uncertainty:

Prevision and prevision uncertainty

More suitable quantity to summarize in two numbers the our probabilistic ‘expectation’ and its uncertainty:

$$\mathsf{E}[X] \;=\; \sum_x x\, f(x)$$

Prevision and prevision uncertainty

More suitable quantity to summarize in two numbers the our probabilistic ‘expectation’ and its uncertainty:

$$
\begin{aligned}
\mathsf{E}[X] &= \sum_x x\, f(x) \\
\mathsf{Var}(X) &= \sum_x (x - \mathsf{E}[X])^2\, f(x) \;\longrightarrow\; \sigma(X) = \sqrt{\mathsf{Var}(X)}
\end{aligned}
$$

Prevision and prevision uncertainty

More suitable quantity to summarize in two numbers the our probabilistic ‘expectation’ and its uncertainty:

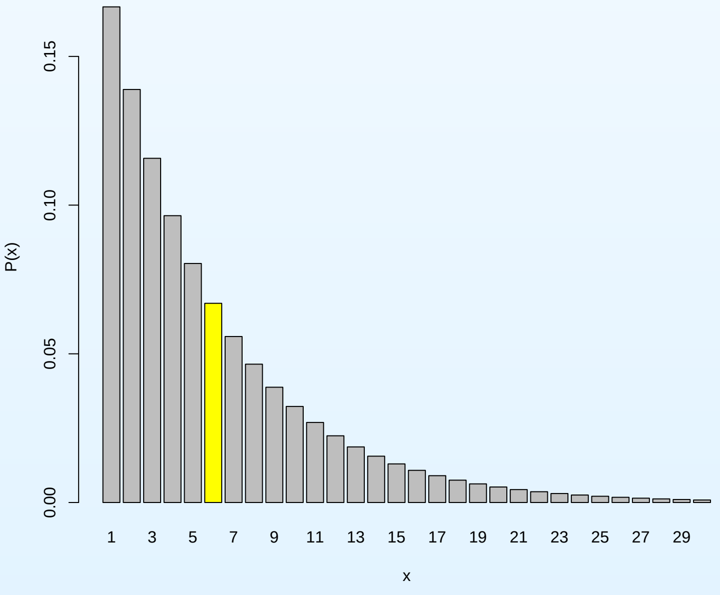
$$\mathsf{E}[X] = \sum_x x\, f(x)$$

$$\mathsf{Var}(X) = \sum_x (x - \mathsf{E}[X])^2\, f(x) \longrightarrow \sigma(X) = \sqrt{\mathsf{Var}(X)}$$

$\mathsf{E}[X] = 1/p$
$\sigma(X) = \sqrt{1-p}/p$

$p = 1/8$:

$$\mathsf{E}[X] = 6$$
$$\sigma(X) = 5.5$$



Prevision and prevision uncertainty

More suitable quantity to summarize in two numbers the our probabilistic ‘expectation’ and its uncertainty:

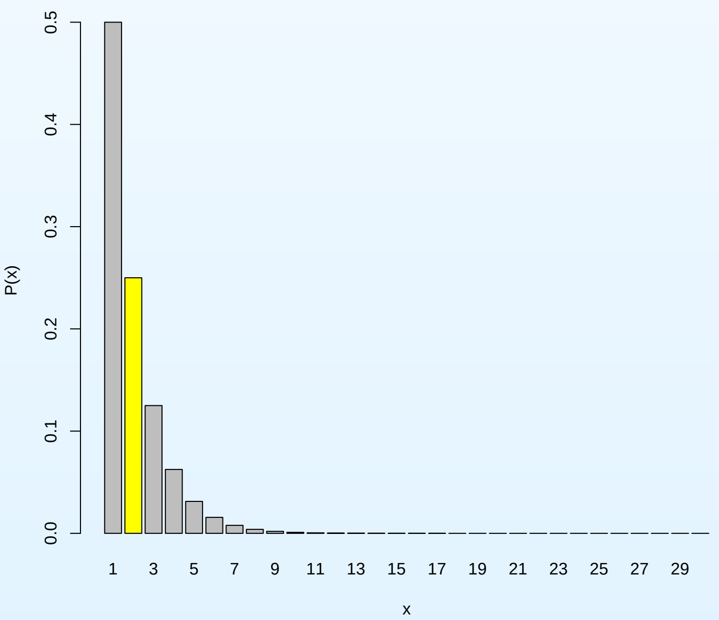
$$\mathsf{E}[X] = \sum_x x\, f(x)$$

$$\mathsf{Var}(X) = \sum_x (x - \mathsf{E}[X])^2\, f(x) \longrightarrow \sigma(X) = \sqrt{\mathsf{Var}(X)}$$

$\mathsf{E}[X] = 1/p$
$\sigma(X) = \sqrt{1-p}/p$

$p = 1/2$:

$$\mathsf{E}[X] = 2$$
$$\sigma(X) = 1.4$$



Prevision and prevision uncertainty

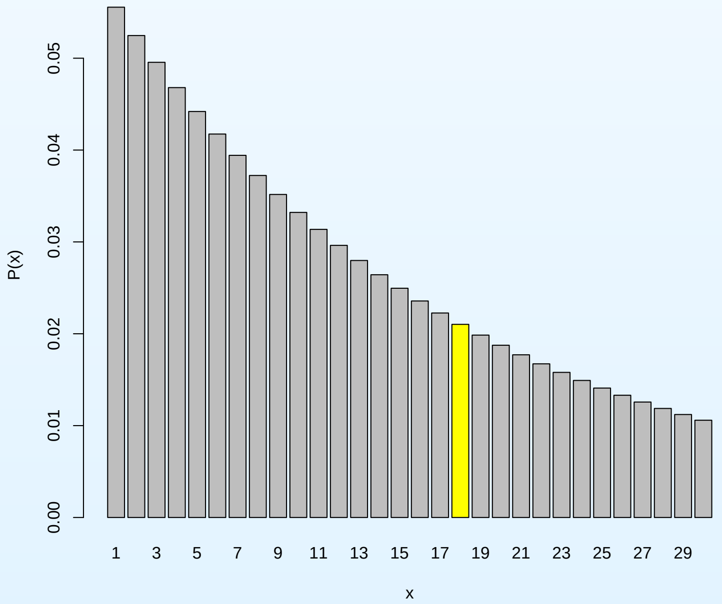
More suitable quantity to summarize in two numbers the our probabilistic ‘expectation’ and its uncertainty:

$$\begin{aligned} \mathsf{E}[X] &= \sum_x x\, f(x) \\ \mathsf{Var}(X) &= \sum_x (x - \mathsf{E}[X])^2\, f(x) \quad \longrightarrow \quad \sigma(X) = \sqrt{\mathsf{Var}(X)} \end{aligned}$$

$\mathsf{E}[X] = 1/p$
$\sigma(X) = \sqrt{1-p}/p$

$p = 1/18$:

$$\begin{aligned} \mathsf{E}[X] &= 18 \\ \sigma(X) &= 17.5 \end{aligned}$$



Prevision and prevision uncertainty

More suitable quantity to summarize in two numbers the our probabilistic 'expectation' and its uncertainty:

$$\mathsf{E}[X] = \sum_x x\, f(x)$$

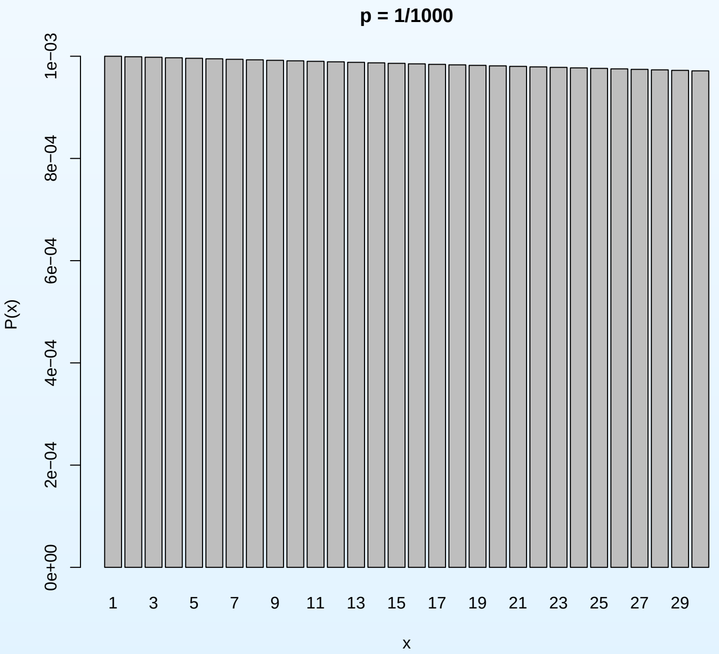
$$\mathsf{Var}(X) = \sum_x (x - \mathsf{E}[X])^2 f(x) \longrightarrow \sigma(X) = \sqrt{\mathsf{Var}(X)}$$

$\mathsf{E}[X] = 1/p$

$\sigma(X) = \sqrt{1-p}/p \xrightarrow[p\to 0]{} 1/p$

→ Rare events might happen at any moment!

(Although they have very small probability to happen in any given very small time interval!)



⇒ be carefull in Risk Management!

Further remarks on first occurrance probability

- If $p \to 0$ in an observational interval $\Delta t \to 0$, then it makes no sense to speak about the probability in the i-th trial:

Further remarks on first occurrance probability

- If $p \to 0$ in an observational interval $\Delta t \to 0$, then it makes no sense to speak about the probability in the i-th trial:

  $\Rightarrow$ It makes only sense of the probability that the 'success' occures between t_1 and t_2;

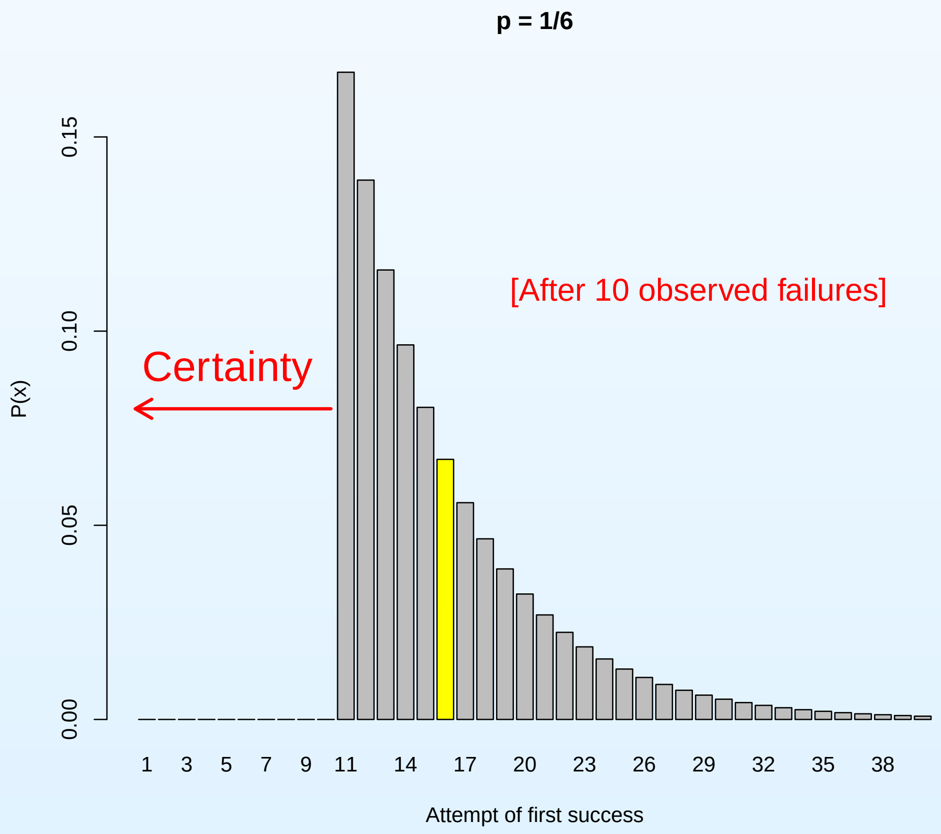
  $\Rightarrow$ $p \to$ intensity of the Poisson process: $r = dp/dt$;

  $\Rightarrow$ Geometric distribution $\to$ exponential distribution;

Further remarks on first occurrance probability

- If $p \rightarrow 0$ in an observational interval $\Delta t \rightarrow 0$, then it makes no sense to speak about the probability in the i-th trial:

  $\Rightarrow$ Geometric distribution $\rightarrow$ exponential distribution;
- No memory property of geometric and exponential:



Further remarks on first occurrance probability

- If $p \to 0$ in an observational interval $\Delta t \to 0$, then it makes no sense to speak about the probability in the i-th trial:

  $\Rightarrow$ Geometric distribution $\to$ exponential distribution;

- No memory property of geometric and exponential;

- Be careful: p (or r) might depend on time (think to aging effects!):

  $\Rightarrow$ geometrical/exponential model might not be any longer suitable models!

Expected value and 'standard uncertainty'

The detail on the uncertainty is provided by $f(x)$.

- E$[X]$ and $\sigma(X)$ are just convenient summaries.
- In the general case they do not convey a precise confidence that X will occur in the range E$[X] \pm \sigma(X)$, though this probability is rather 'high' for typical $f(x)$ of interest.

Expected value and 'standard uncertainty'

The detail on the uncertainty is provided by $f(x)$.

- E$[X]$ and $\sigma(X)$ are just convenient summaries.
- In the general case they do not convey a precise confidence that X will occur in the range E$[X] \pm \sigma(X)$, though this probability is rather 'high' for typical $f(x)$ of interest.
- Another location summary (that statisticians like much) is given by the median, while the 'quantiles' provide (left open) intervals in which the variable is expected to fall with some probability (typically 10%, 20%, etc.).

Expected value and ‘standard uncertainty’

The detail on the uncertainty is provided by $f(x)$.

- E$[X]$ and $\sigma(X)$ are just convenient summaries.
- In the general case they do not convey a precise confidence that X will occur in the range E$[X] \pm \sigma(X)$, though this probability is rather ‘high’ for typical $f(x)$ of interest.
- Another location summary (that statisticians like much) is given by the median, while the ’quantiles’ provide (left open) intervals in which the variable is expected to fall with some probability (typically 10%, 20%, etc.).
- Anyway, it is important to prepared to $f(x)$ of any kind, because – fortunately! – nature is not boring. . .

Expected value and ‘standard uncertainty’

The detail on the uncertainty is provided by $f(x)$.

- E$[X]$ and $\sigma(X)$ are just convenient summaries.
- In the general case they do not convey a precise confidence that X will occur in the range E$[X] \pm \sigma(X)$, though this probability is rather ‘high’ for typical $f(x)$ of interest.
- Another location summary (that statisticians like much) is given by the median, while the ’quantiles’ provide (left open) intervals in which the variable is expected to fall with some probability (typically 10%, 20%, etc.).
- Anyway, it is important to prepared to $f(x)$ of any kind, because – fortunately! – nature is not boring. . .
- In particular, $f(x)$ might be asymmetric or, ‘multinomial’, i.e. with more than one local maximum.

When to bet on the barycenter of the distribution?

When asked about the drunk man problem, most people ask they would bet on the 8-th trial, or something around it.

When to bet on the barycenter of the distribution?

When asked about the drunk man problem, most people ask they would bet on the 8-th trial, or something around it.
…and even when they are told the should bet on the first one, they reply that the first attempt has a little probability...

When to bet on the barycenter of the distribution?

When asked about the drunk man problem, most people ask they would bet on the 8-th trial, or something around it.
. . . and even when they are told the should bet on the first one, they reply that the first attempt has a little probability...

- Yes, $P(1)$ can be small, but it is the largest one (although all others, are all together more probable!)

When to bet on the barycenter of the distribution?

When asked about the drunk man problem, most people ask they would bet on the 8-th trial, or something around it.
...and even when they are told the should bet on the first one, they reply that the first attempt has a little probability...

- Yes, $P(1)$ can be small, but it is the largest one (although all others, are all <u>together</u> more probable!)
- **Bet on the 1-st if you win/lose if you hit/miss the number**

When to bet on the barycenter of the distribution?

When asked about the drunk man problem, most people ask they would bet on the 8-th trial, or something around it.
…and even when they are told the should bet on the first one, they reply that the first attempt has a little probability...

- Yes, $P(1)$ can be small, but it is the largest one (although all others, are all together more probable!)
- **Bet on the 1-st if you win/lose if you hit/miss the number**
- BUT sometimes wins who gets closest.

When to bet on the barycenter of the distribution?

When asked about the drunk man problem, most people ask they would bet on the 8-th trial, or something around it.
. . . and even when they are told the should bet on the first one, they reply that the first attempt has a little probability...

- Yes, $P(1)$ can be small, but it is the largest one (although all others, are all together more probable!)
- **Bet on the 1-st if you win/lose if you hit/miss the number**
- BUT sometimes wins who gets closest.
  - Bet on median if loss is linear with the error.
  - Bet on average if loss is quadratic with the error

Probability distributions Vs ‘statistical’ distributions

It is important to stress the difference between

- Probability distribution
  - To each possible outcome we associate how much we are confident on it:

$$x \longleftrightarrow f(x)$$

- Statistical distribution
  - To each observed outcome we associated its (relative) frequency

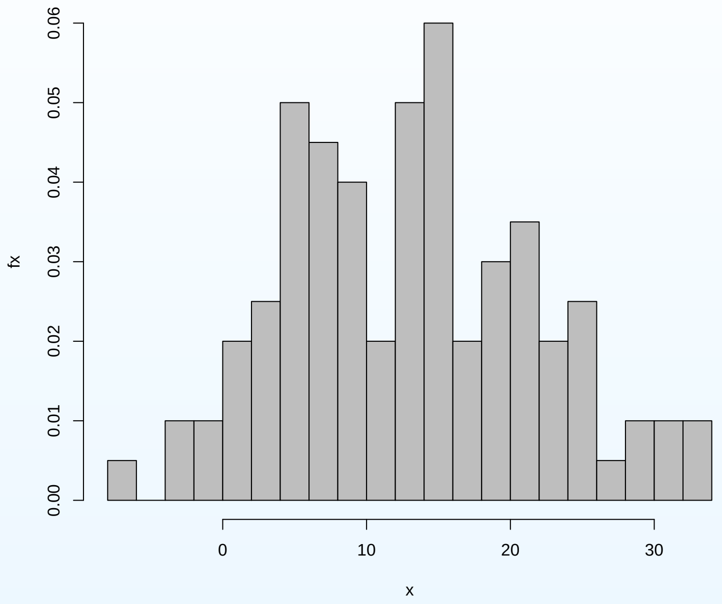
$$x \longleftrightarrow f_x$$

(e.g. an histogram of experimental observations)

Summaries (‘mean’, variance, ’σ’, ’skewness’, etc) have similar names and analogous definitions, but conceptual different meaning.

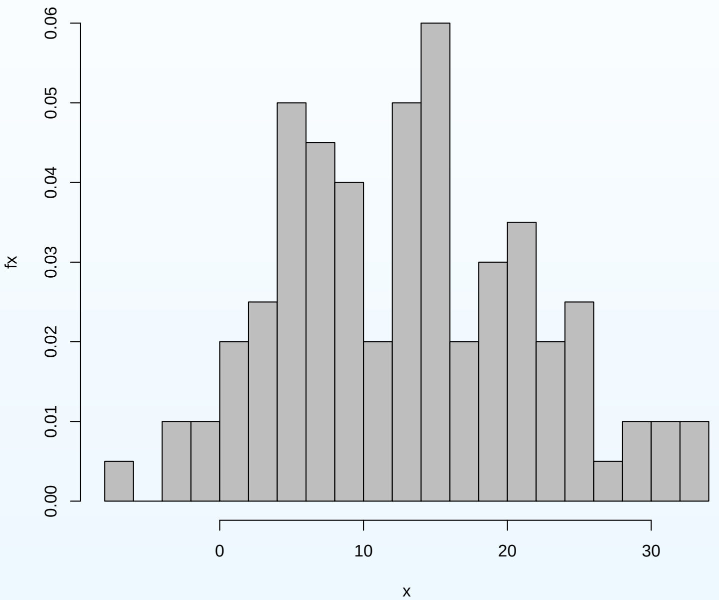
A histogram is not, usually, a probability distribution

In particular a histogram of experimental data is not a probability distribution (unless one reshuffles those events, and extracts one of them at random).



A histogram is not, usually, a probability distribution

In particular a histogram of experimental data <u>is not</u> a probability distribution (unless one reshuffles those events, and extracts one of them at random).



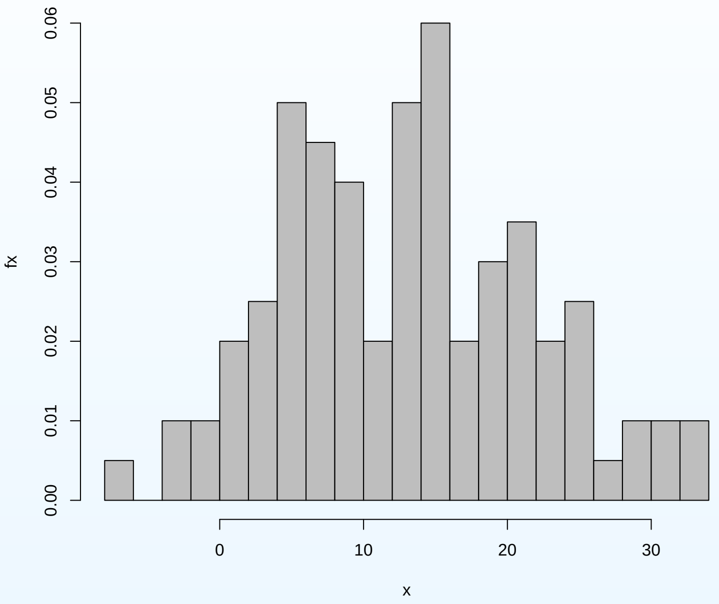
Average and variance

$$\overline{x} = \sum_x x\, f_x$$

$$\sigma^2 = \sum_x (x-\overline{x})^2\, f_x$$

A histogram is not, usually, a probability distribution

In particular a histogram of experimental data is not a probability distribution (unless one reshuffles those events, and extracts one of them at random).



Average and variance

$$\overline{x} = \sum_x x\, f_x$$

$$\sigma^2 = \sum_x (x-\overline{x})^2\, f_x$$

$\rightarrow$ Just a rough empirical description of the shape
$\Rightarrow$ center of mass and momentum of inertia!
(Famous '$n/(n-1)$' correction: interference descriptive $\leftrightarrow$ inferential statistics.)

Distributions derived from the Bernoulli process

Bernoulli

Geometric

(trial of 1st success)

Pascal

(trial of k-th success)

Binomial

(# of successes in ind. n trials)

(Binomial well known. We shall not use the Pascal)

Poisson distribution

One of the best known distributions by physicist.

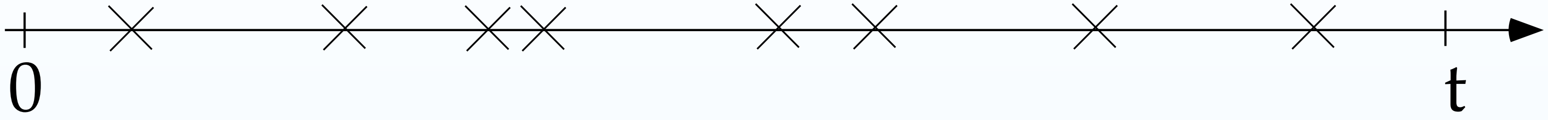
For a while, just take the mathematical approach to the Poisson distribution:

$$f(x \,|\, \mathcal{P}_\lambda) = \frac{\lambda^x}{x!}\, e^{-\lambda} \qquad \left\{ \begin{array}{l} 0 < \lambda < \infty \\ x = 0, 1, \ldots, \infty \end{array} \right. .$$

Reminding also the well known property

$$\mathcal{B}_{n,p} \xrightarrow[\substack{n \to \infty \\ p \to 0 \\ (n\, p = \lambda)}]{} P_\lambda \, .$$

Poisson process



Let us consider some phenomena that might happen at a give instant, such that

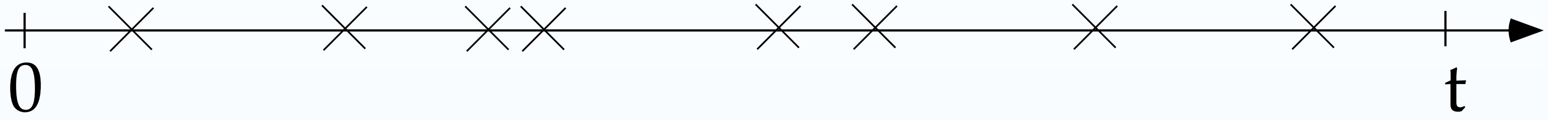
- Probability of 1 count in ΔT is proportional to ΔT, with ΔT 'small'.

  $$p = P(\text{“1 count in } \Delta T'') = r\,\Delta T$$

  where r is the intensity of the process'
- $P(\geq 2\ \text{counts}) \ll P(1\text{count})$ (OK if ΔT is small enough)
- What happens in one interval does not depend on other intervals (if disjoints)

Let us divide a finite interval T in n small intervals, i.e. $T = n\,\Delta T$, and $\Delta T = T/n$.

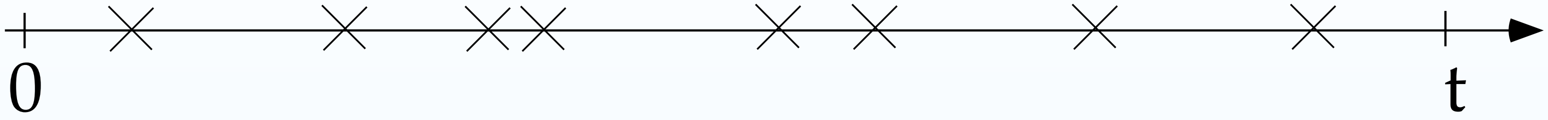
Poisson process $\longrightarrow$ Poisson distribution



Considering the possible occurrence of a count in each small interval ΔT an independent Bernoulli trial, of probability

$$p = r\,\Delta T = r\,T/n$$

Poisson process $\longrightarrow$ Poisson distribution

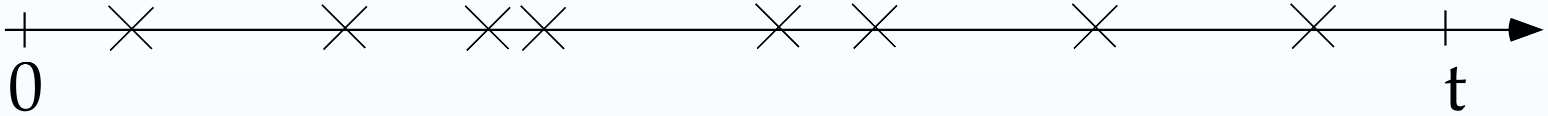


Considering the possible occurrence of a count in each small interval ΔT an independent Bernoulli trial, of probability

$$p = r\,\Delta T = r\,T/n$$

If we are interested in the number of counts in T, independently from the order: $\rightarrow$ Binomial : $\mathcal{B}_{n,p}$

Poisson process $\longrightarrow$ Poisson distribution



Considering the possible occurrence of a count in each small interval ΔT an independent Bernoulli trial, of probability

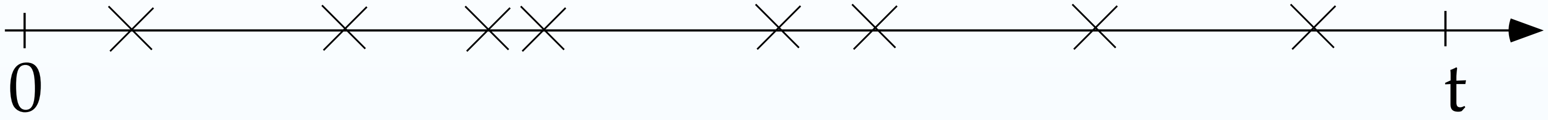
$$p = r\,\Delta T = r\,T/n$$

If we are interested in the number of counts in T, independently from the order: $\rightarrow$ Binomial : $\mathcal{B}_{n,p}$

But $n \rightarrow \infty$ and $p \rightarrow 0$ $\Rightarrow \mathcal{B}_{n,p} \rightarrow \mathcal{P}_\lambda$ where $\lambda = n\,p = r\,T$

$\Rightarrow$ λ depends only on the intensity of the process and on the finite time of observation.

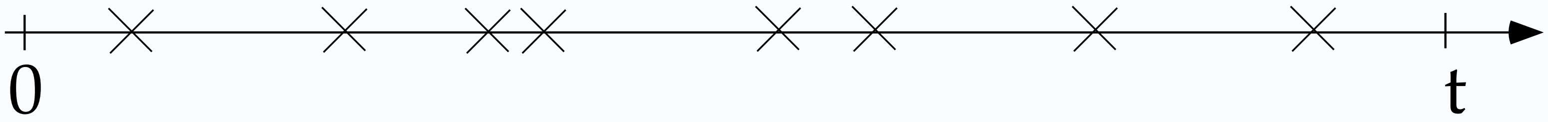
Poisson process $\longrightarrow$ waiting time



Another interesting problem: how long do we have to wait for the first count? (Starting from any arbitrary time)

Problem analogous to the Geometric, but now it makes no sense to ask at which small interval the counts will occur!

Poisson process $\longrightarrow$ waiting time



Another interesting problem: how long do we have to wait for the first count? (Starting from any arbitrary time)

Problem analogous to the Geometric, but now it makes no sense to ask at which small interval the counts will occur!

Let us restart from the Geometric and calculate $P(X > x)$:

$$P(X > x) = \sum_{i>x} f(i \,|\, \mathcal{G}_p) = (1-p)^x$$

(The count will not occur in the first x trials).

In the domain of time, using $p = r\,t/n$ and then making the limit:

$$P(T > t) = (1-p)^n = (1 - r\,t/n)^n \xrightarrow[n\to\infty]{} e^{-r\,t}$$

Poisson process $\longrightarrow$ Exponential distribution

Knowing $P(T > t)$ we get easily the cumulative $F(t)$:

$$F(t) = P(T \leq t) = 1 - P(T > t) = 1 - e^{-r\,t}\,.$$

$F(t)$ is now a continuous function!

Poisson process $\longrightarrow$ Exponential distribution

Knowing $P(T > t)$ we get easily the cumulative $F(t)$:

$$F(t) = P(T \leq t) = 1 - P(T > t) = 1 - e^{-r\,t}\,.$$

$F(t)$ is now a continuous function!

In some region of t there is a concentration of probability more than in other regions.

Poisson process $\longrightarrow$ Exponential distribution

Knowing $P(T > t)$ we get easily the cumulative $F(t)$:

$$F(t) = P(T \leq t) = 1 - P(T > t) = 1 - e^{-r\,t}\,.$$

$F(t)$ is now a continuous function!

In some region of t there is a concentration of probability more than in other regions.

$\rightarrow$ This leads us to define a probability density function (pdf) for continuous variables:

$$f(t) = \frac{d\,F(x)}{d\,t}\,.$$

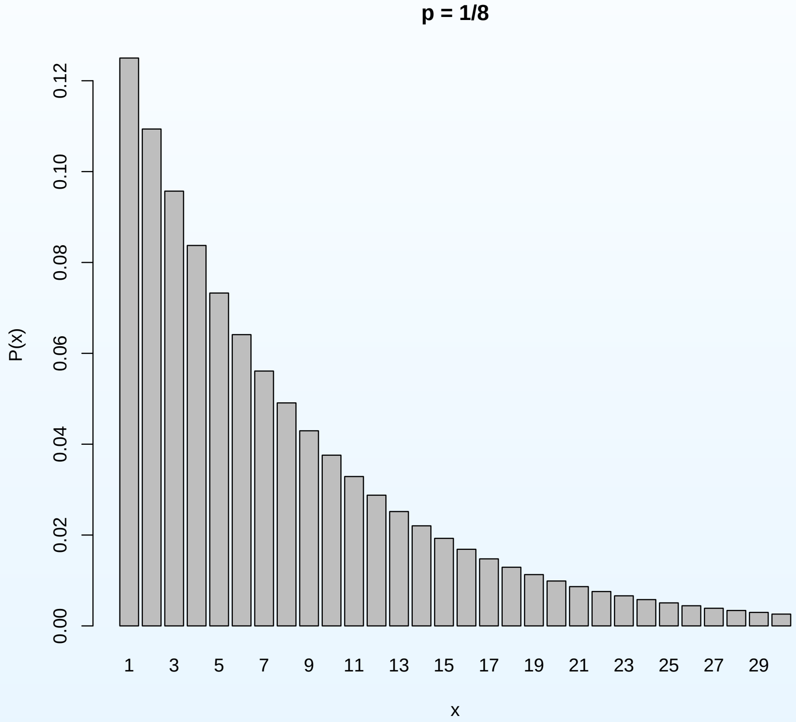
- In this case $f(t) = r\,e^{-r\,t} = \frac{1}{\tau}\,e^{-t/\tau}$

$\rightarrow$ Exponential distribution ($\tau = 1/r$): E$[T] = \sigma(T) = \tau$.

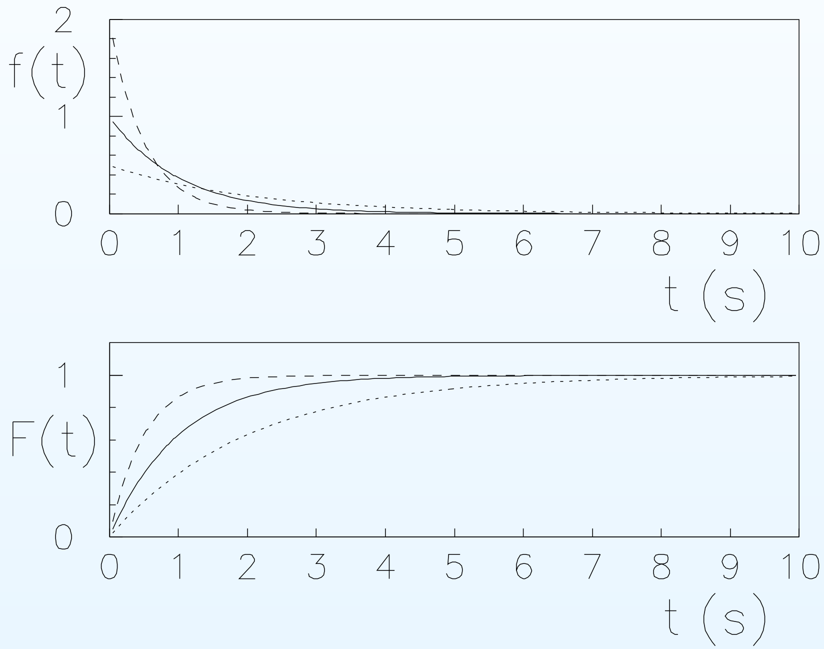
(⇒ Properties of pdf assumed to be well known.)

Geometric ↔ Exponential

Geometric

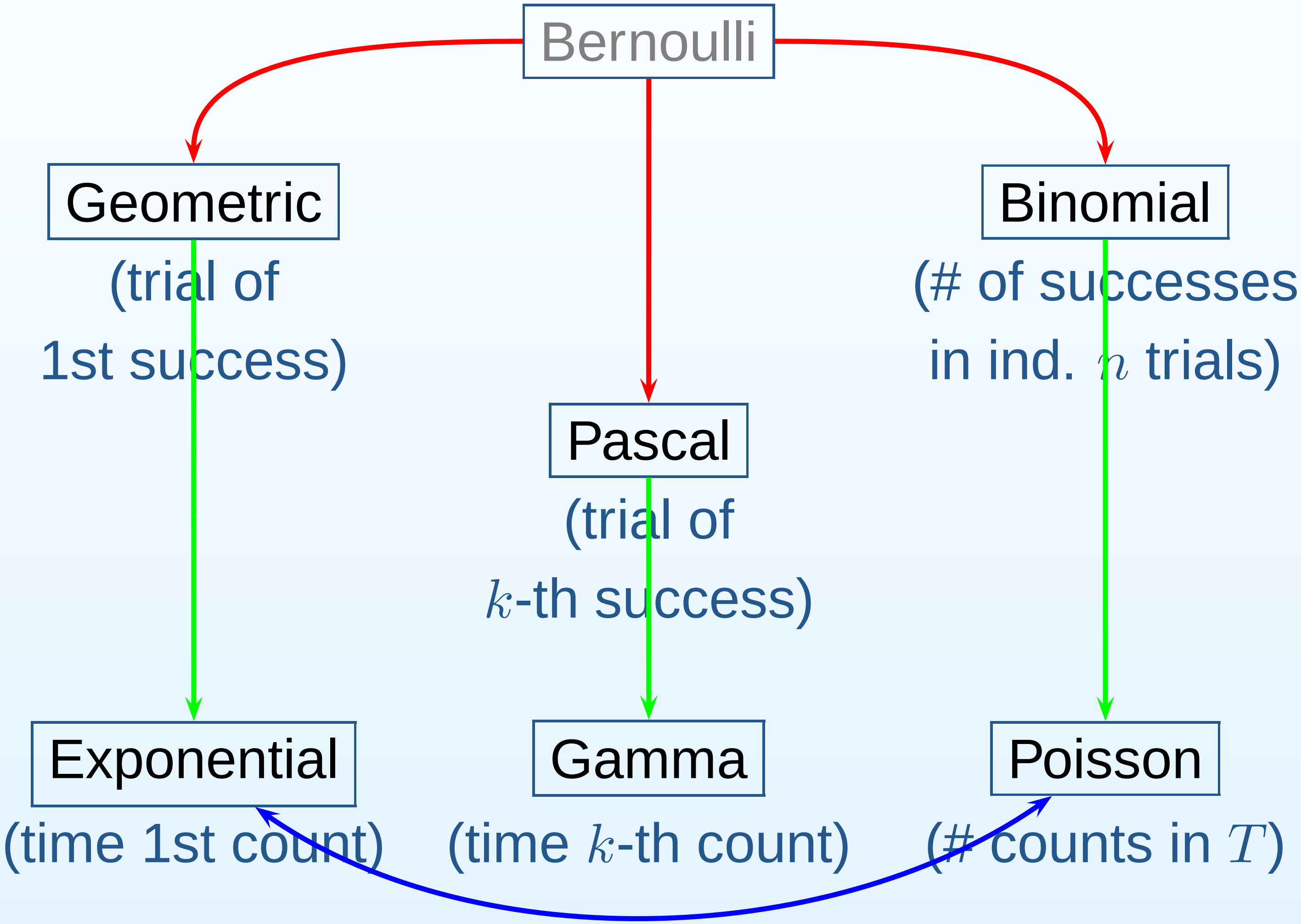


Exponential

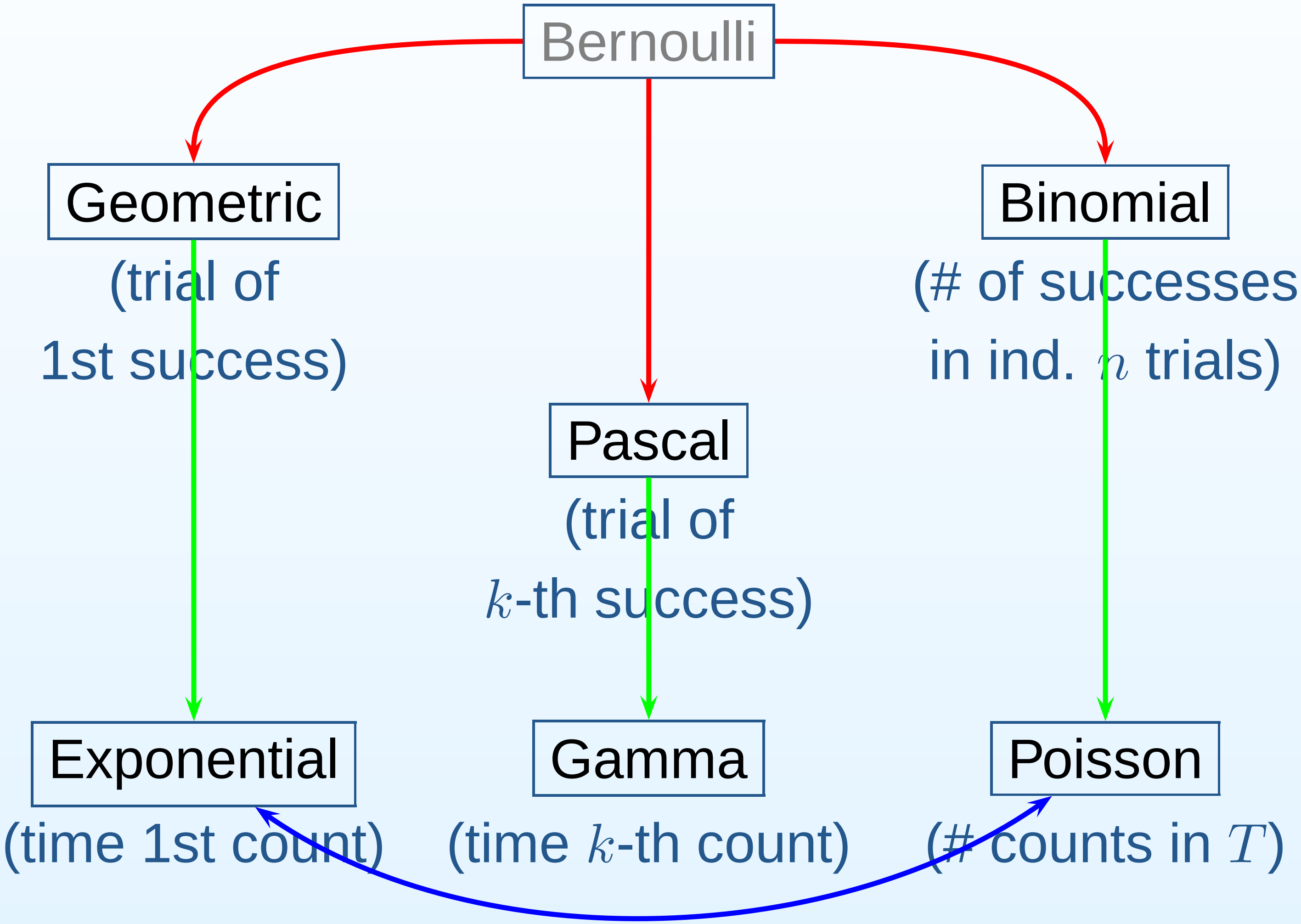


Exponential is just the limit to the continuum of the Geometric.
'No memory' property for both: Assuming a success (or a count) has not happened until a certain trial (or time), the distributions restart from there. No need to know the instant of particle creation to measure 'life time' ($\rightarrow$ the "10^{33} year old" proton!).

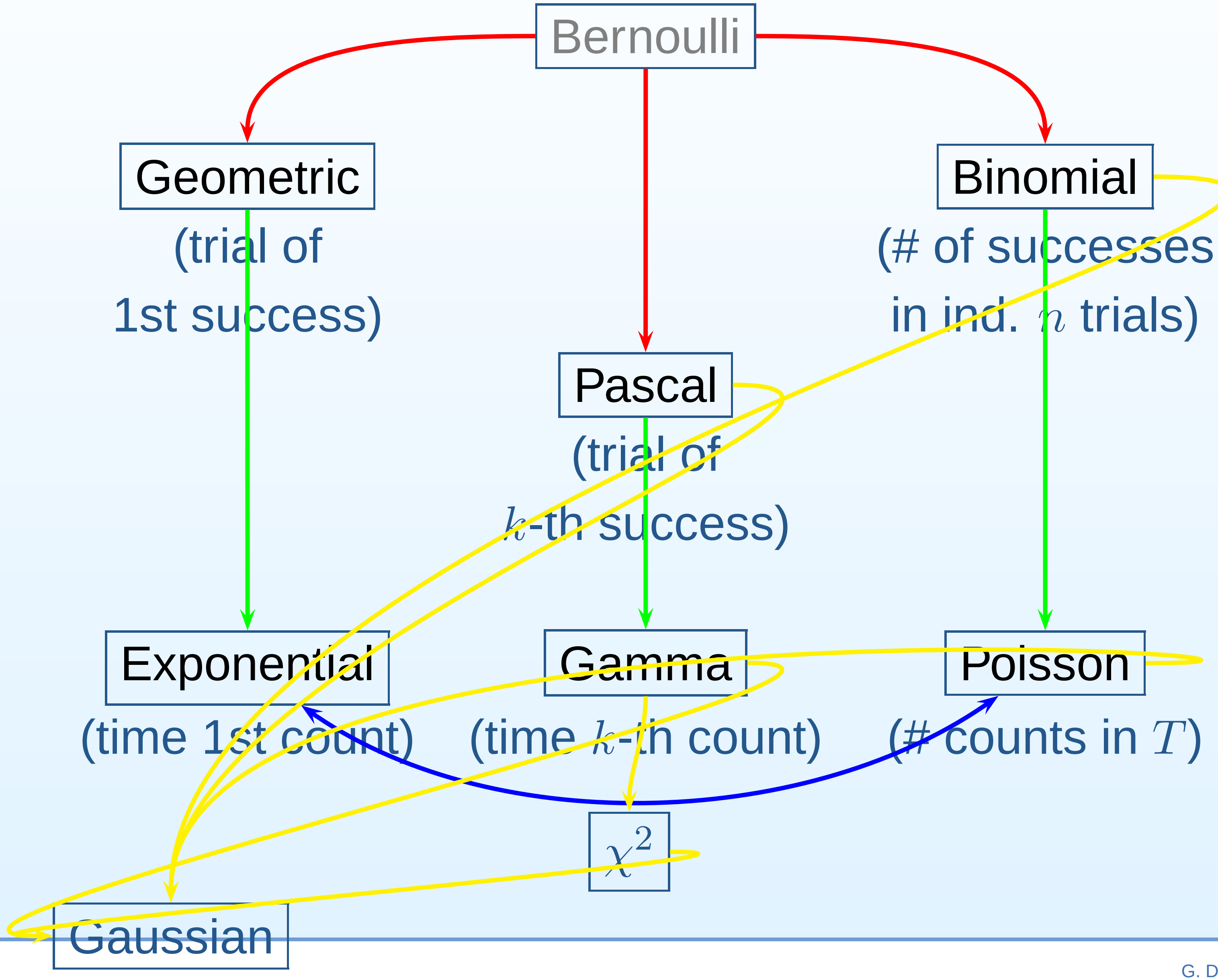
Distributions derived from the Bernoulli process



Distributions derived from the Bernoulli process



Distributions derived from the Bernoulli process



Note

Though we could not go through all technical details, it is important to remark that all these distributions are obtained assuming that each ‘act of observation’, that can be asymptotically associated to a single point, is an independent Bernoulli trial of constant probability p (that might tend to zero).

Important properties of probability distributions

$\mathsf{E}(\cdot)$ is a linear operator:

$$\mathsf{E}(aX + b) = a\,\mathsf{E}(X) + b\,.$$

Transformation properties of variance and standard deviation:

$$\begin{aligned} \mathsf{Var}(aX + b) &= a^2\,\mathsf{Var}(X)\,, \\ \sigma(aX + b) &= |a|\,\sigma(X)\,. \end{aligned}$$

Obviously, I have to assume that most of the basic formalism is well known, e.g. that $P(a \le X \le b) = \int_a^b f(x)\,dx$, etc.

From probability to future frequencies

Let us think to n independent Bernoulli trials that have to be made.

Number of successes $X \sim \mathcal{B}_{n,p}$, with p.

We might be interested to the relative frequency of successes, i.e. $f_n = X/n$: $f_n = 0, 1/n, 2/n, \ldots, 1$

What do we expect for f_n?

From probability to future frequencies

Let us think to n independent Bernoulli trials that have to be made.

Number of successes $X \sim \mathcal{B}_{n,p}$, with p.

We might be interested to the relative frequency of successes, i.e. $f_n = X/n$: $f_n = 0, 1/n, 2/n, \ldots, 1$

What do we expect for f_n? $f(f_n)$ can be obtained from $f(x)$.

$$\begin{aligned} \mathrm{E}(f_n) &\equiv \frac{1}{n}\,\mathrm{E}(X\,|\,\mathcal{B}_{n,p}) = \frac{n\,p}{n} = p \\ \sigma(f_n) &\equiv \frac{1}{n}\,\sigma(X\,|\,\mathcal{B}_{n,p}) = \frac{\sqrt{p\,(1-p)}}{\sqrt{n}} \xrightarrow[n\to\infty]{} 0 \end{aligned}$$

We expect p, with uncertainty that decreases with $\sqrt{n}$:
$\rightarrow$ *Bernoulli's theorem*, the most known, misunderstood and misused probability theory theorem.

From probability to future frequencies

Let us think to n independent Bernoulli trials that <u>have to be made</u>.

Number of successes $X \sim \mathcal{B}_{n,p}$, with p.

We might be interested to the relative frequency of successes, i.e. $f_n = X/n$: $f_n = 0, 1/n, 2/n, \ldots, 1$

What do we expect for f_n? $f(f_n)$ can be obtained from $f(x)$.

$$\begin{aligned} \mathrm{E}(f_n) &\equiv \frac{1}{n}\mathrm{E}(X \,|\, \mathcal{B}_{n,p}) = \frac{n\,p}{n} = p \\ \sigma(f_n) &\equiv \frac{1}{n}\sigma(X \,|\, \mathcal{B}_{n,p}) = \frac{\sqrt{p\,(1-p)}}{\sqrt{n}} \xrightarrow[n\to\infty]{} 0 \end{aligned}$$

In particular, it justifies the increased probability of neither 'late numbers' at lotto, nor frequency based definition of probability (Circular: cannot define probability from probability theorem!)

Note sul teorema di Bernoulli

- La formulazione va intesa in termini di probabilità e non di certezza.

Note sul teorema di Bernoulli

- La formulazione va intesa in termini di probabilità e non di certezza.
- Il teorema non implica assolutamente che, se per un certo N lo scarto $|f_n - p|$ è grande, allora per $n > N$ la frequenza relativa f_n “debba recuperare” per “mettersi in regola con la legge”.

Note sul teorema di Bernoulli

- La formulazione va intesa in termini di probabilità e non di certezza.
- Il teorema non implica assolutamente che, se per un certo N lo scarto $|f_n - p|$ è grande, allora per $n > N$ la frequenza relativa f_n “debba recuperare” per “mettersi in regola con la legge”.
- Esso non giustifica la “definizione” frequentista di probabilità. Affermando infatti che “è molto probabile che la frequenza non differisca molto dalla probabilità” si sta assumendo il concetto di probabilità:

Note sul teorema di Bernoulli

- La formulazione va intesa in termini di probabilità e non di certezza.
- Il teorema non implica assolutamente che, se per un certo N lo scarto $|f_n - p|$ è grande, allora per $n > N$ la frequenza relativa f_n "debba recuperare" per "mettersi in regola con la legge".
- Esso non giustifica la "definizione" frequentista di probabilità. Affermando infatti che "è molto probabile che la frequenza non differisca molto dalla probabilità" si sta assumendo il concetto di probabilità:
  - "Probabilità come propensione" $\approx$ OK
    $\rightarrow \quad P(E\,|\,\text{prop} = p) = p$
    $\rightarrow f_n$ dati eventi analoghi in cui crediamo che la propensione sia la stessa, vale il Th. di Bernoulli.
  - Probabilità <u>come</u> limite della frequenza: NO

Note sul teorema di Bernoulli - 2

- Non si dimentichi che il teorema di Bernoulli . . . è un teorema, basato sulle regole di base dalla probabilità e su tutte le proprietà che ne derivano. Quindi non può definire il concetto di probabilità.

Note sul teorema di Bernoulli - 2

- Non si dimentichi che il teorema di Bernoulli . . . è un teorema, basato sulle regole di base dalla probabilità e su tutte le proprietà che ne derivano. Quindi non può definire il concetto di probabilità.
- Su tale argomento è molto convincente de Finetti

> *"Per quanti tendono a ricollegare il concetto stesso di probabilità alla nozione di frequenza, tali risultati [che f_n "tenda a p"] vengono ad assumere un ruolo di cerniera per convalidare tale avvicinamento o identificazione di nozioni. Logicamente non si sfugge però al dilemma che la stessa cosa non si può assumere prima per definizione e poi dimostrare come teorema, né alla contraddizione di una definizione che assumerebbe una cosa certa mentre il teorema afferma che è soltanto molto probabile.*

Note sul teorema di Bernoulli - 2

- Non si dimentichi che il teorema di Bernoulli . . . è un teorema, basato sulle regole di base dalla probabilità e su tutte le proprietà che ne derivano. Quindi non può definire il concetto di probabilità.
- Su tale argomento è molto convincente de Finetti

> *"Per quanti tendono a ricollegare il concetto stesso di probabilità alla nozione di frequenza, tali risultati [che f_n "tenda a p"] vengono ad assumere un ruolo di cerniera per convalidare tale avvicinamento o identificazione di nozioni. Logicamente non si sfugge però al dilemma che la stessa cosa non si può assumere prima per definizione e poi dimostrare come teorema, né alla contraddizione di una definizione che assumerebbe una cosa certa mentre il teorema afferma che è soltanto molto probabile.*

- Si noti inoltre che la condizione di p costante implica che essa sia prefissata a priori e che anche le valutazioni sui possibili esiti di f_n siano fatte prima di iniziare le prove (o in condizione di incertezza sul loro esito).

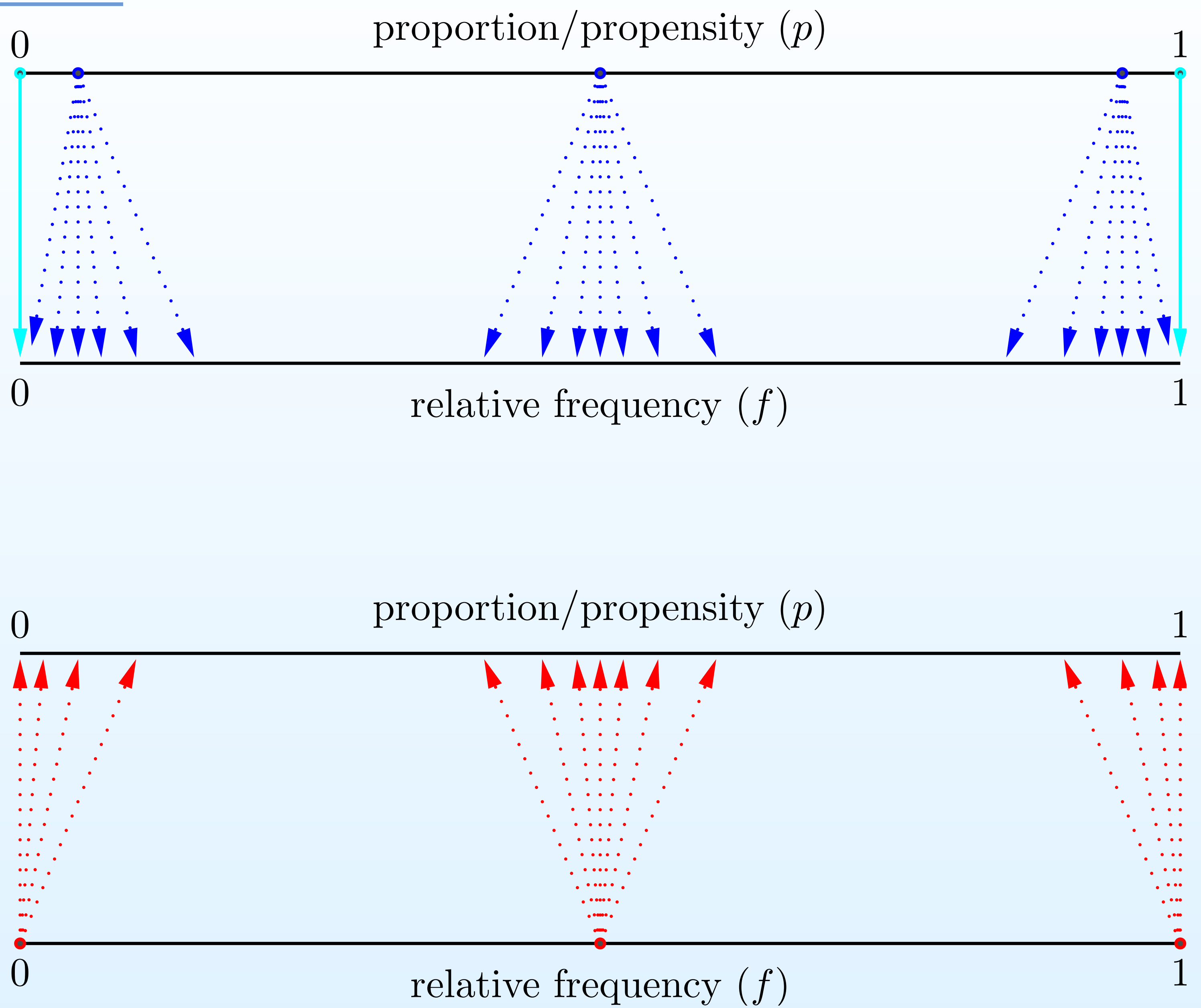
Probabilità Vs frequenza relativa

Concetti da tenere ben distinti anche se esiste un collegamento fra di loro:

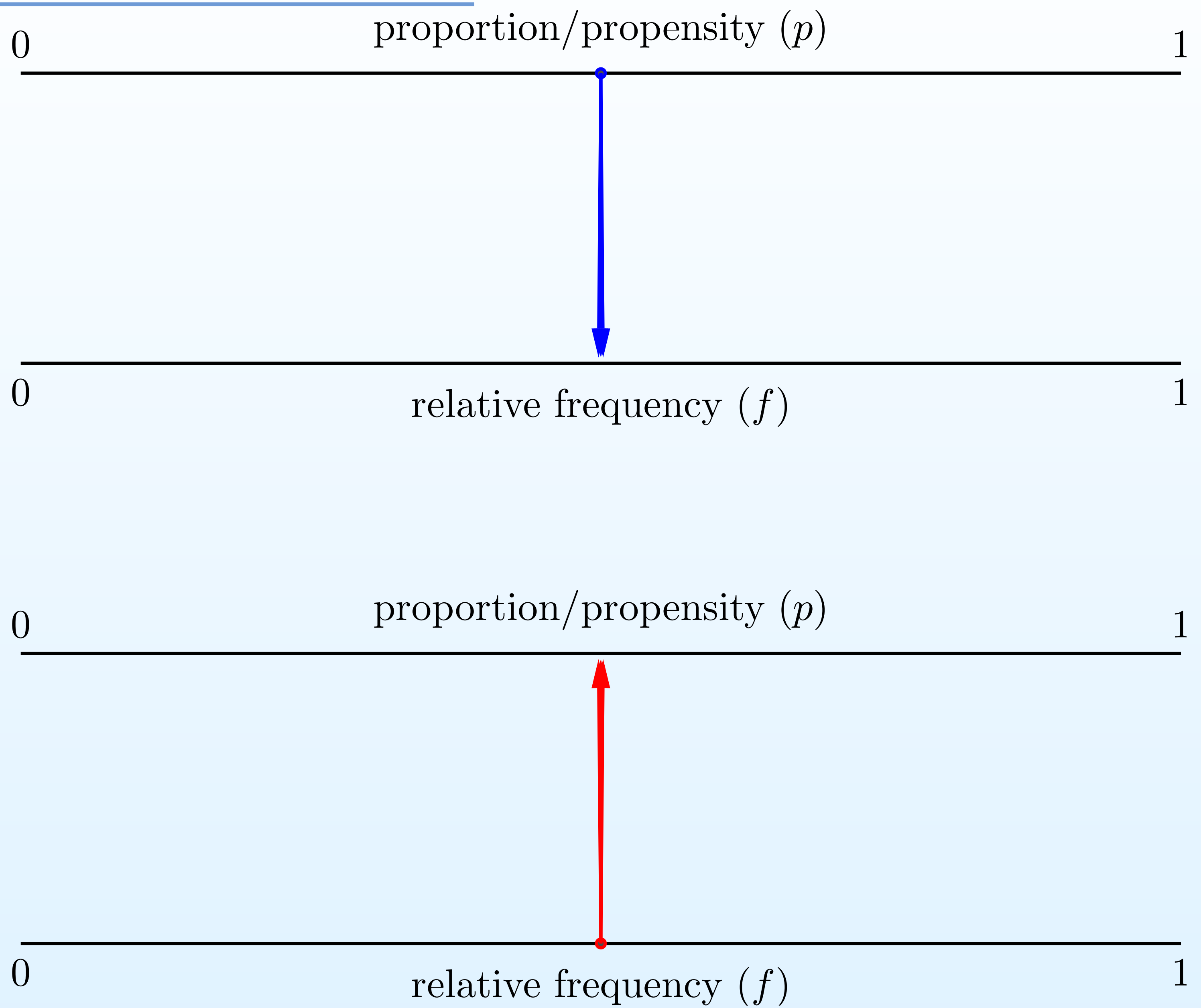
$p \rightarrow f_n$: Teorema di Bernoulli

$f_n \rightarrow p$: Teorema di Bayes sotto precise condizioni

$p \leftrightarrow f$



$p \leftrightarrow f$: Large n limits



Propagation of uncertainties

All we have seen so far in this short review of ‘direct probability’ is how to ‘propagate probability’ to logically connected events or variables.

Propagation of uncertainties

All we have seen so far in this short review of ‘direct probability’ is how to ‘propagate probability’ to logically connected events or variables.

$\Rightarrow$ Therefore, the famous problem of propagation of uncertainty is straightforward in a probabilistic approach: just use probability theory.

[Note that in the frequency based approach one does something similar, but in a ‘strange’ way, because one is not allowed to use probability for physical quantities, but only for estimators.]

Propagation of uncertainties

All we have seen so far in this short review of ‘direct probability’ is how to ‘propagate probability’ to logically connected events or variables.

$\Rightarrow$ Therefore, the famous problem of propagation of uncertainty is straightforward in a probabilistic approach: just use probability theory.

[Note that in the frequency based approach one does something similar, but in a ‘strange’ way, because one is not allowed to use probability for physical quantities, but only for estimators.]

The general problem:

$$f(x_1, x_2, \ldots, x_n) \xrightarrow[Y_j=Y_j(X_1,X_2,\ldots,X_n)]{} f(y_1, y_2, \ldots, y_m)\,.$$

This calculation can be quite challenging, but it can be easily performed by Monte Carlo techniques.

General solution for discrete variables

$Y = Y(X)$, where $Y()$ stands for the mathematical function relating X and Y.

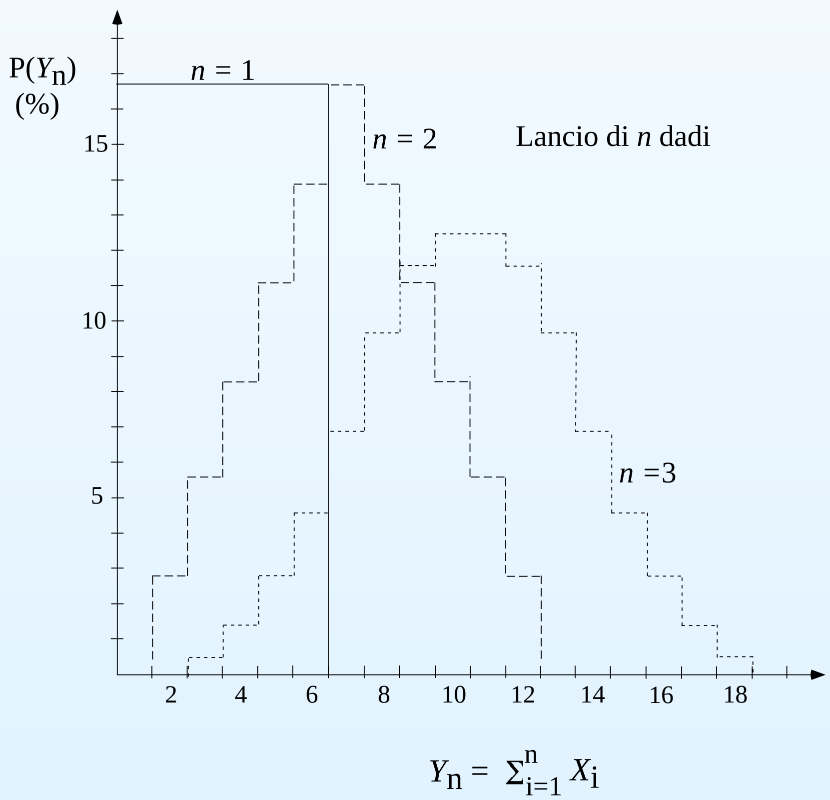
The probability of a given $Y = y$ is equal to the sum of the probability of each X such that $Y(X = x) = y$.

General solution for discrete variables

$Y = Y(X)$, where $Y()$ stands for the mathematical function relating X and Y.

The probability of a given $Y = y$ is equal to the sum of the probability of each X such that $Y(X = x) = y$.

Probability distributions of the sums of the results from n dice.



General solution for discrete variables

$Y = Y(X)$, where $Y()$ stands for the mathematical function relating X and Y.

The probability of a given $Y = y$ is equal to the sum of the probability of each X such that $Y(X = x) = y$.

The extension to many variables is straightforward:
for ex., given two *input* quantities X_1 and X_2, with their probability function $f(x_1, x_2)$, and two *output* quantities Y_1 and Y_2:

$$f(y_1, y_2) = \sum_{\substack{x_1, x_2 \\ \left\{ \begin{array}{l} Y_1(x_1, x_2) = y_1 \\ Y_2(x_1, x_2) = y_2 \, . \end{array} \right.}} f(x_1, x_2)$$

(For each point $\{y_1, y_2\}$ sum up the probability of all points in the $\{X_1, X_2\}$ space that satisfy the constrain.)

General solution for continuous variable

Just extend to the continuum the previous formula:

- replace sums by integrals
- replace constrains by suitable Dirac $\delta()$:

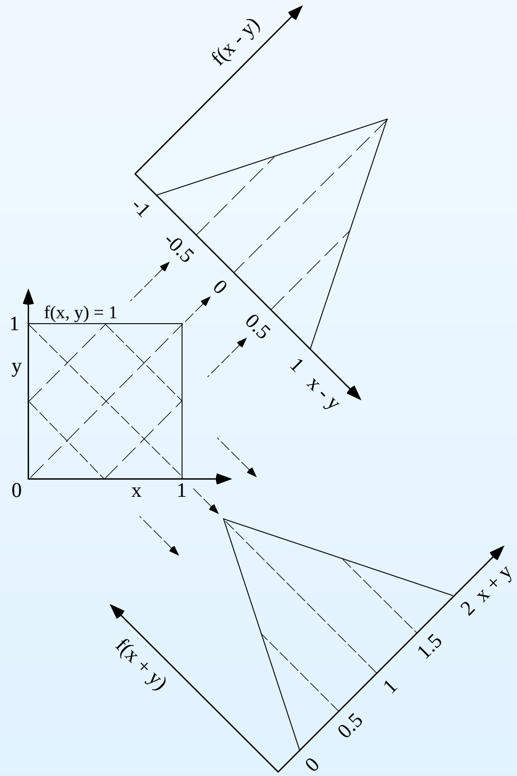
$$f(y_1, y_2) = \int \delta(y_1 - Y_1(x_1, x_2))\, \delta(y_2 - Y_2(x_1, y_2))\, f(x_1, x_2)\, \mathrm{d}x_1 \mathrm{d}x_2 \,.$$

General solution for continuous variable

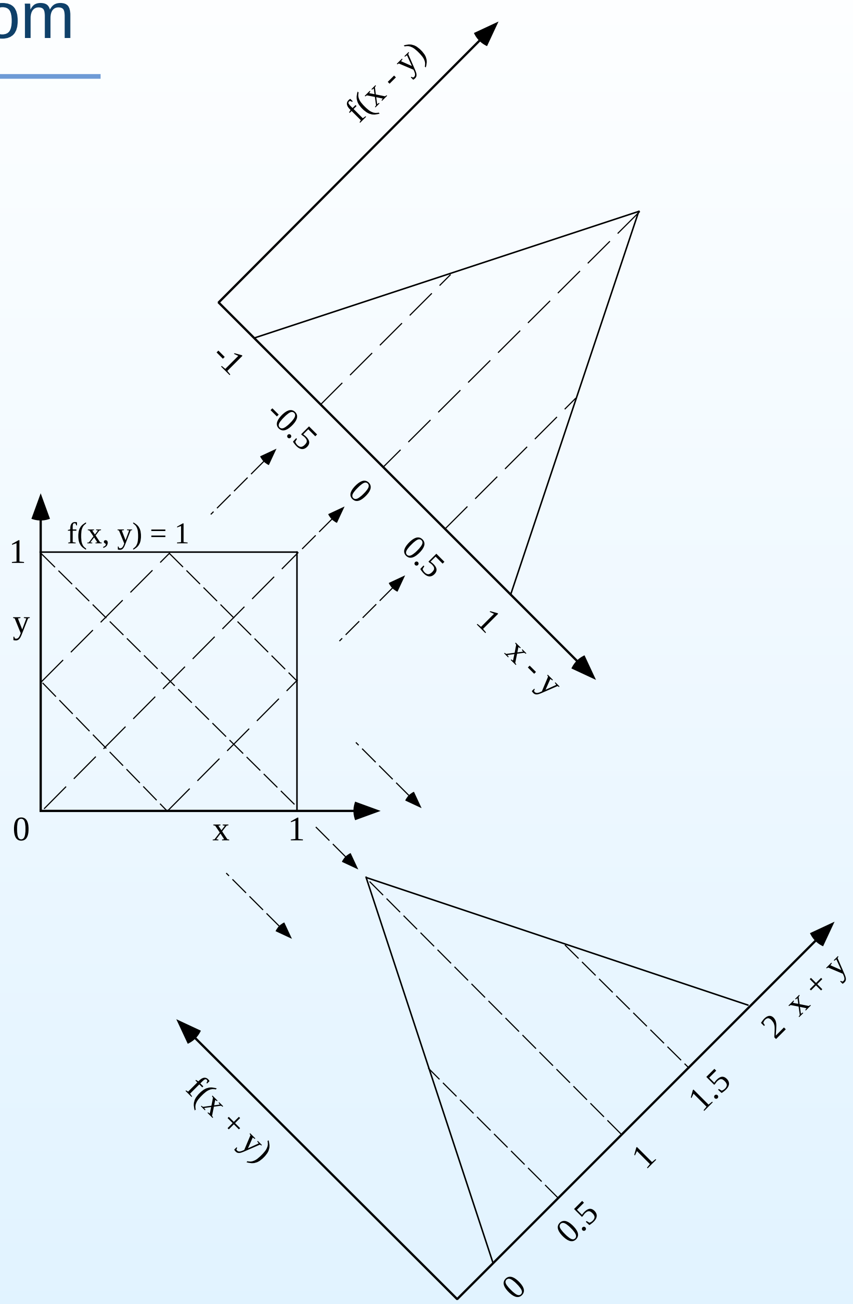
Just extend to the continuum the previous formula:

- replace sums by integrals
- replace constrains by suitable Dirac $\delta()$:

$$f(y_1,y_2) = \int \delta(y_1 - Y_1(x_1,x_2))\,\delta(y_2 - Y_2(x_1,y_2))\,f(x_1,x_2)\,\mathrm{d}x_1\mathrm{d}x_2\,.$$



Zoom

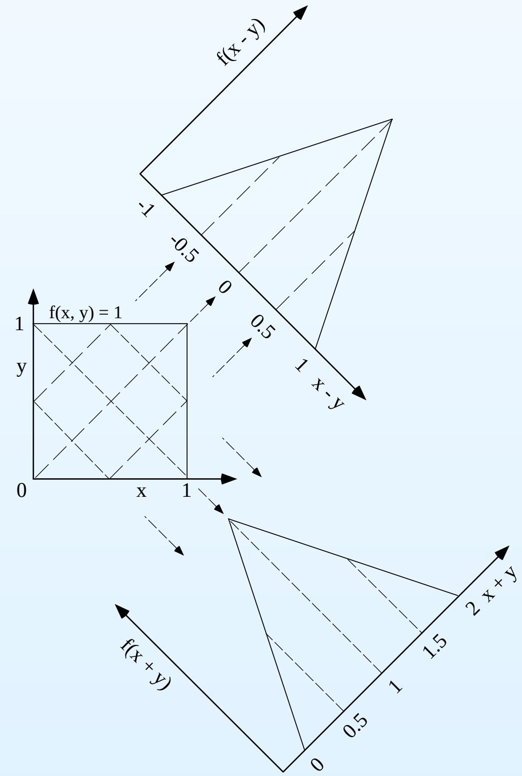


General solution for continuous variable

Just extend to the continuum the previous formula:

- replace sums by integrals
- replace constrains by suitable Dirac $\delta()$:

$$f(y_1, y_2) = \int \delta(y_1 - Y_1(x_1, x_2))\, \delta(y_2 - Y_2(x_1, y_2))\, f(x_1, x_2)\, \mathrm{d}x_1 \mathrm{d}x_2\,.$$



General solution for continuous variable

Just extend to the continuum the previous formula:

- replace sums by integrals
- replace constrains by suitable Dirac $\delta()$:

$$f(y_1, y_2) = \int \delta(y_1 - Y_1(x_1, x_2))\,\delta(y_2 - Y_2(x_1, y_2))\,f(x_1, x_2)\,\mathrm{d}x_1 \mathrm{d}x_2\,.$$

$$\mathrm{E}(Y) = \mathrm{E}(X_1) + \mathrm{E}(X_2)$$
$$\sigma^2(Y) = \sigma^2(X_1) + \sigma^2(X_2)$$

$\text{mode}(Y) \leftrightarrow \text{mode}(X_i)$

$\text{median}(Y) \leftrightarrow \text{median}(X_i)$

?

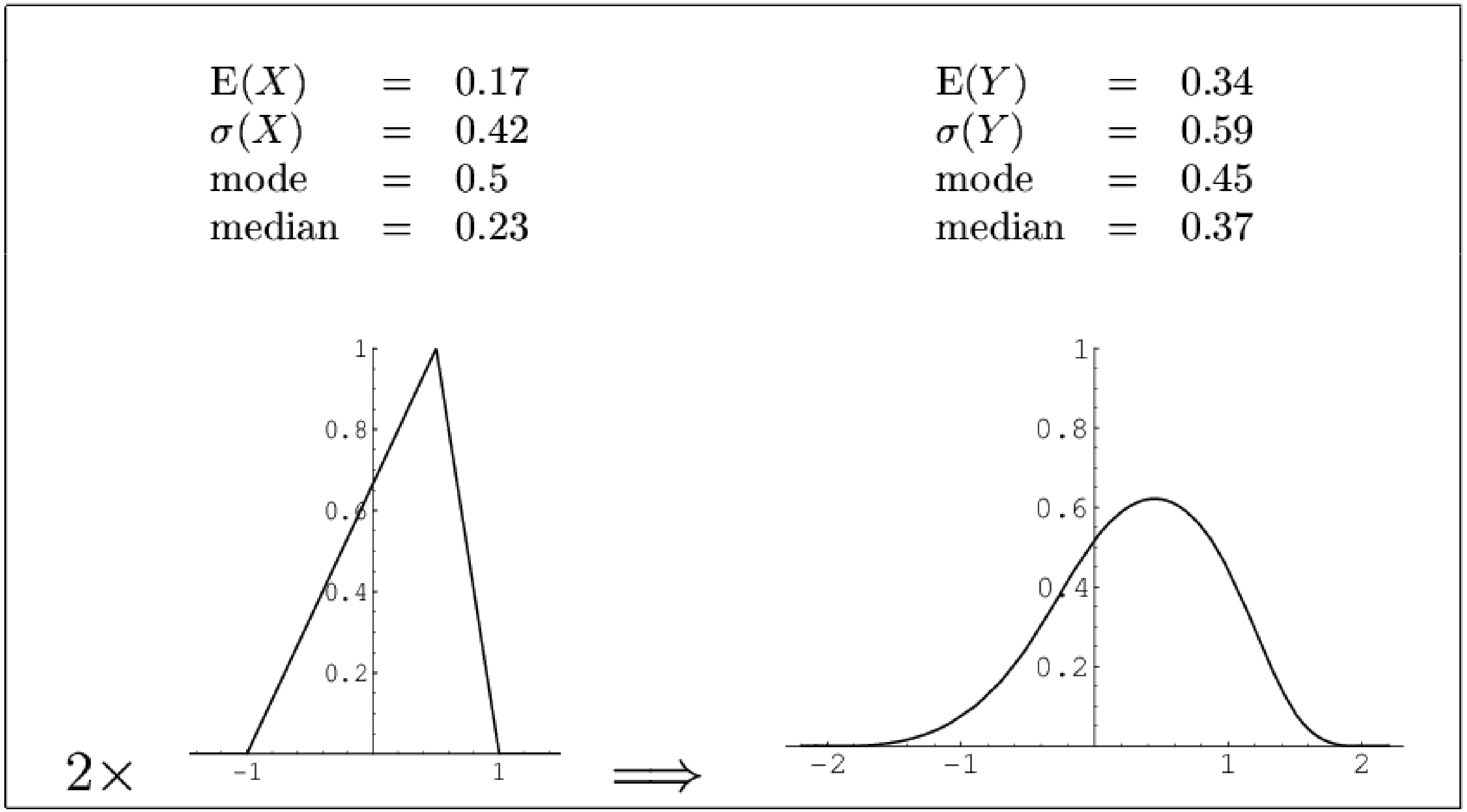
$\mathrm{E}(X)$	=	0.17	$\mathrm{E}(Y)$	=	0.34
$\sigma(X)$	=	0.42	$\sigma(Y)$	=	0.59
mode	=	0.5	mode	=	0.45
median	=	0.23	median	=	0.37

No equivalent rule for the most probable values!

But there is nothing similar for the most probable values

$\boxed{0.5} + \boxed{0.5} = \boxed{1}$ only for nice symmetric distributions

$\boxed{0.5} + \boxed{0.5} = \boxed{0.45}$ in our 'asymmetric' example!



No equivalent rule for the most probable values!

But there is nothing similar for the most probable values

$\boxed{0.5} + \boxed{0.5} = \boxed{1}$ only for nice symmetric distributions

$\boxed{0.5} + \boxed{0.5} = \boxed{0.45}$ in our 'asymmetric' example!

Not just an odd academic example:

- asymmetric uncertainties occur often in HEP
  every time you read 'best value' $^{+\Delta_+}_{-\Delta_-}$!

No equivalent rule for the most probable values!

But there is nothing similar for the most probable values

$\boxed{0.5} + \boxed{0.5} = \boxed{1}$ only for nice symmetric distributions

$\boxed{0.5} + \boxed{0.5} = \boxed{0.45}$ in our 'asymmetric' example!

Not just an odd academic example:

- asymmetric uncertainties occur often in HEP
  every time you read 'best value' $^{+\Delta_+}_{-\Delta_-}$!

→ asymmetric χ^2 or log-likelihoods

→ asymmetry in – well treated! – uncertainty propagations

→ systematics (often related to non linear propagation)

No equivalent rule for the most probable values!

But there is nothing similar for the most probable values

$\boxed{0.5} + \boxed{0.5} = \boxed{1}$ only for nice symmetric distributions

$\boxed{0.5} + \boxed{0.5} = \boxed{0.45}$ in our 'asymmetric' example!

Not just an odd academic example:

- asymmetric uncertainties occur often in HEP
  every time you read 'best value' $^{+\Delta_+}_{-\Delta_-}$!

$\rightarrow$ asymmetric χ^2 or log-likelihoods

$\rightarrow$ asymmetry in – well treated! – uncertainty propagations

$\rightarrow$ systematics (often related to non linear propagation)

And remember that standard methods (χ^2 or ML fits) provide something equivalent to 'most probable values', not to E( )!

(As we shall see.)

End

FINE

Notes

The following slides should be reached by hyper-links, clicking on words with the symbol †

If I eat a chicken and you eat no chicken. . .

. . . for the statistics each of us eats 1/2 chicken.

For the pleasure of Italian readers, this is how Trilussa put is:

La statistica

Sai ched'è la statistica? È 'na cosa
che serve pe' fa' un conto in generale
de la gente che nasce, che sta male,
che more, che va in carcere e che sposa.

Ma pe' me la statistica curiosa
è dove c'entra la percentuale,
pe' via che, lì, la media è sempre eguale
puro co' la persona bisognosa.

(continues on next slide →)

Go Back

La Statistica di Trilussa (continua)

Me spiego, da li conti che se fanno
seconno le statistiche d'adesso
risurta che te tocca un pollo all'anno:

e, se nun entra ne le spese tue,
t'entra ne la statistica lo stesso
perché c'è un antro che se ne magna due.

Go Back

For example:

For example:

- Why should one be allowed to state that
  "the interval 170–180 GeV contains the value of the top quark mass with a given probability",

For example:

- Why should one be allowed to state that
  "the interval 170–180 GeV contains the value of the top quark mass with a given probability",

  ... but not that say that
  "the value of the top quark mass lies in that interval with the same probability"?

For example:

- Why should one be allowed to state that
  "the interval 170–180 GeV contains the value of the top quark mass with a given probability",

  . . . but not that say that
  "the value of the top quark mass lies in that interval with the same probability"?

  $\Rightarrow$ quite an odd ideology about what probability is!
  Aristotle would get mad. . .

For example:

- Why should one be allowed to state that
  "the interval 170–180 GeV contains the value of the top quark mass with a given probability",

  . . . but not that say that
  "the value of the top quark mass lies in that interval with the same probability"?

  $\Rightarrow$ quite an odd ideology about what probability is!
  Aristotle would get mad. . .

  ◦ So unnatural that essentially all teachers teach 'standard confidence intervals' as probability intervals
  (or this is, at least, what remains in the students minds – who will later become teachers, and the circle goes on).

For example:

- Why should one be allowed to state that
  "the interval 170–180 GeV contains the value of the top quark mass with a given probability",

  ... but not that say that
  "the value of the top quark mass lies in that interval with the same probability"?

  $\Rightarrow$ quite an odd ideology about what probability is!
  Aristotle would get mad...

  - So unnatural that essentially all teachers teach 'standard confidence intervals' as probability intervals
    (or this is, at least, what remains in the students minds – who will later become teachers, and the circle goes on).
  - And even statistics experts, when they have to transmit to the rest of the community the meaning of what they do, they have hard time in doing it $\longrightarrow$ Slide

. . . or

- Why a 95% C.L lower bound does not mean that we are 95% confident that the quantity is above this limit?

. . . or

- Why a 95% C.L lower bound does not mean that we are 95% confident that the quantity is above this limit?

  More precisely:

  - If we know that a box contains 95% of white balls, then
    - we can evaluate $P(\text{white}) = 95\%$

    $\Rightarrow$ we feel 95% confident to extract a white ball.

...or

- Why a 95% C.L lower bound does not mean that we are 95% confident that the quantity is above this limit?

  More precisely:

  - If we know that a box contains 95% of white balls, then
    - we can evaluate $P(\text{white}) = 95\%$

    $\Rightarrow$ we feel 95% confident to extract a white ball.
  - 95% C.L lower bounds do no have [in most cases – but somethimes they do(!)] the same meaning:

...or

- Why a 95% C.L lower bound does not mean that we are 95% confident that the quantity is above this limit?

  More precisely:

  - If we know that a box contains 95% of white balls, then
    - we can evaluate $P(\text{white}) = 95\%$

    $\Rightarrow$ we feel 95% confident to extract a white ball.
  - 95% C.L lower bounds do no have [in most cases – but somethimes they do(!)] the same meaning:

  $\Rightarrow$ we are not as confident that the quantity is above the bound as we are confident to extract a white box from the box!

  - great confusion! $\rightarrow$ 1998 survey $\longrightarrow$ Slides
  - At least, clear after 2000 CERN CLW $\longrightarrow$ Slide

. . . or

- Why a 95% C.L lower bound does not mean that we are 95% confident that the quantity is above this limit?

  More precisely:

  - If we know that a box contains 95% of white balls, then
    - we can evaluate $P(\text{white}) = 95\%$

    $\Rightarrow$ we feel 95% confident to extract a white ball.
  - 95% C.L lower bounds do no have [in most cases – but somethimes they do(!)] the same meaning:

  $\Rightarrow$ we are not as confident that the quantity is above the bound as we are confident to extract a white box from the box!

  - great confusion! $\rightarrow$ 1998 survey $\longrightarrow$ Slides
  - At least, clear after 2000 CERN CLW $\longrightarrow$ Slide
    (But I am afraid if I would redo the survey now, I would get similar answers. . .)

. . . or

- Why do we insist in using the 'frequentistic coverage' that, apart the high sounding names and attributes ('exact', 'classical', "guarantees ..", . . .), manifestly does not cover?

. . . or

- Why do we insist in using the 'frequentistic coverage' that, apart the high sounding names and attributes ('exact', 'classical', "guarantees ..", . . .), manifestly does not cover?

  More precisely (and besides the 'philosophical quibbles' of the interval that covers the value with a given probability, and not the value being in the interval with that probability):

  - many thousands C.L. upper/lower bounds have been published in the past years

  ⇒ but never a value has shown up in the 5% or 10% side, that, by complementarity, the method should cover in 5% or 10% of the cases.

...or

- Why do we insist in using the 'frequentistic coverage' that, apart the high sounding names and attributes ('exact', 'classical', "guarantees ..", ...), manifestly does not cover?

  More precisely (and besides the 'philosophical quibbles' of the interval that covers the value with a given probability, and not the value being in the interval with that probability):

  - many thousands C.L. upper/lower bounds have been published in the past years

  $\Rightarrow$ but never a value has shown up in the 5% or 10% side, that, by complementarity, the method should cover in 5% or 10% of the cases.

  Notwithstanding the fact that there is been a lot of activity in the past years by several physicists, convinced that the idea is basically good, but one only needs 'a better prescription'.

. . . or

- Why do we insist in using the ‘frequentistic coverage’ that, apart the high sounding names and attributes (‘exact’, ‘classical’, “guarantees ..”, . . .), manifestly does not cover?

  More precisely (and besides the ‘philosophical quibbles’ of the interval that covers the value with a given probability, and not the value being in the interval with that probability):

  - many thousands C.L. upper/lower bounds have been published in the past years

  $\Rightarrow$ but never a value has shown up in the 5% or 10% side, that, by complementarity, the method should cover in 5% or 10% of the cases.

  If the method guarantees the claimed coverage, who refunds us if it does not work?

. . . or

- Why do we insist in using the ‘frequentistic coverage’ that, apart the high sounding names and attributes (‘exact’, ‘classical’, “guarantees ..”, . . .), manifestly does not cover?
- In January 2000 I was answered that the reason “is because people have been flip-flopping. Had they used a unified approach, this would not have happened” (G. Feldman)

. . . or

- Why do we insist in using the ‘frequentistic coverage’ that, apart the high sounding names and attributes (‘exact’, ‘classical’, “guarantees ..”, . . .), manifestly does not cover?
- In January 2000 I was answered that the reason “is because people have been flip-flopping. Had they used a unified approach, this would not have happened” (G. Feldman)
- After six years the production of 90-95% C.L. bounds has continued steadily, and in many cases the so called ‘unified approach’ has been used, but still coverage does not do its job.

. . . or

- Why do we insist in using the 'frequentistic coverage' that, apart the high sounding names and attributes ('exact', 'classical', "guarantees ..", . . .), manifestly does not cover?
- In January 2000 I was answered that the reason "is because people have been flip-flopping. Had they used a unified approach, this would not have happened" (G. Feldman)
- After six years the production of 90-95% C.L. bounds has continued steadily, and in many cases the so called 'unified approach' has been used, but still coverage does not do its job.
- What will be the next excuse?

⇒ I do not know what the so-called 'flip-plopping' is,

but we can honestly acknowledge the flop of that reasoning.

Go Back