Telling the Truth with Statistics

Giulio D'Agostini

Università di Roma La Sapienza e INFN

Roma, Italy

Kind of lectures

As the title implies:

- mainly on fundamental aspects
- and on HEP applications where the fundamental aspects play a key role (comparing hypotheses, frontier type measurements, propagation of uncertainty, not-Gaussian likelihoods, systematics).

But also simple ‘routine‘ cases, where approximated methods are satisfactory.

Overview of the contents

1st part Review of the process of learning from data

Mainly based on

- *"From observations to hypotheses: Probabilistic reasoning versus falsificationism and its statistical variations"* (Vulcano 2004, physics/0412148)
- Chapter 1 of *"Bayesian reasoning in high energy physics. Principles and applications"* (CERN Yellow Report 99-03)

Overview of the contents

1st part Review of the process of learning from data
Mainly based on

- *“From observations to hypotheses: Probabilistic reasoning versus falsificationism and its statistical variations”* (Vulcano 2004, physics/0412148)
- Chapter 1 of *“Bayesian reasoning in high energy physics. Principles and applications”* (CERN Yellow Report 99-03)

2nd part Review of the probability and ‘direct probability’ problems, including ‘propagation of uncertainties.
Partially covered in

- First 3 sections of Chapter 3 of YR 99-03
- Chapter 4 of YR 99-03
- *"Asymmetric uncertainties: sources, treatment and possible dangers"* (physics/0403086)

Overview of the contents

3th part Probabilistic inference and applications to HEP

Much material and references in my web page. In particular, I recommend a quite concise review

- *"Bayesian inference in processing experimental data: principles and basic applications"*, Rep.Progr.Phys. 66 (2003)1383 [physics/0304102]

For a more extensive treatment:,

- *“Bayesian reasoning in data analysis – A critical introduction”*, World Scientific Publishing, 2003

  (CERN Yellow Report 99-03 updated and $\approx$ doubled in contents)

Preamble

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

Preamble

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

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

Preamble

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

- proposed by organizers $\longrightarrow$ 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 $\longrightarrow$ 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 <u>the Truth</u>"?
     - What is <u>the true</u> value of a quantity?
     - What is <u>the true</u> theory that describes the world?
  2. "Tell the truth" $\Longleftrightarrow$ "to lie"?

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"?

Preamble

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

- proposed by organizers $\longrightarrow$ 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")

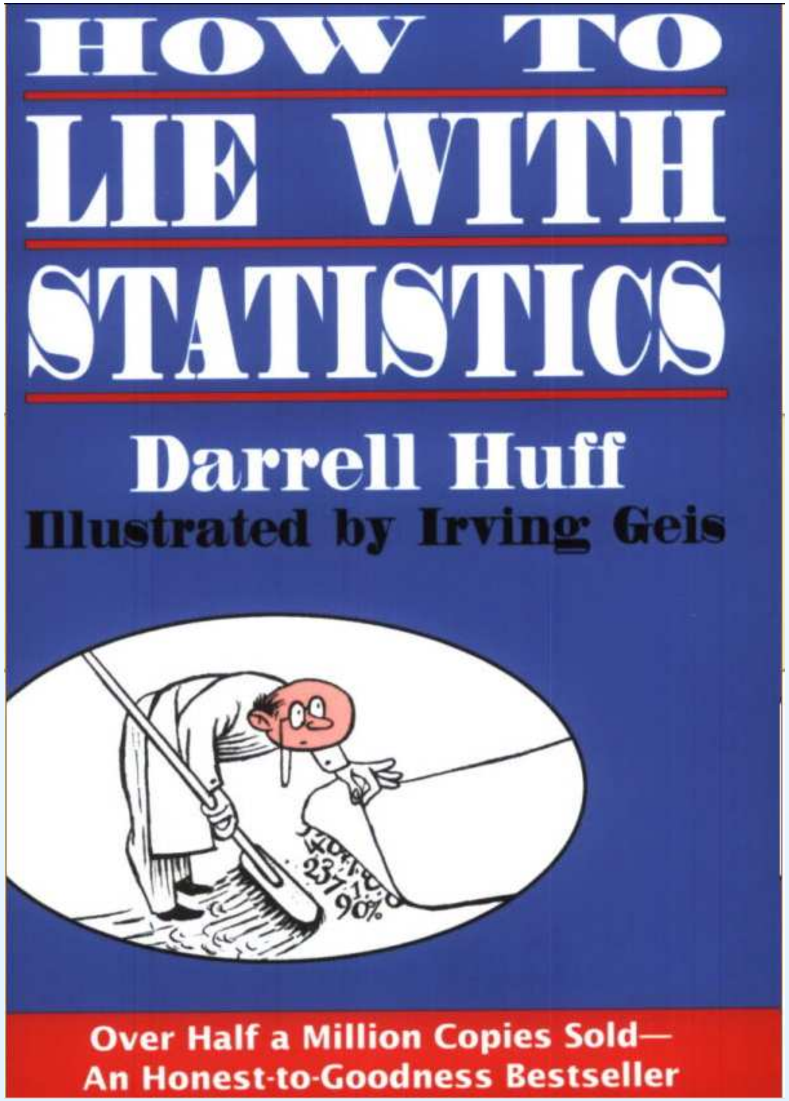
- proposed by organizers → accepted...
- I interpret it as a direct question, to which I will try to give my best answer , quite frankly.
- How to interpret the question?
  1. ~~"Tell the Truth"?~~ ⇒ Question to God
     - ~~What is the true value of a quantity?~~
     - ~~What is the true theory that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

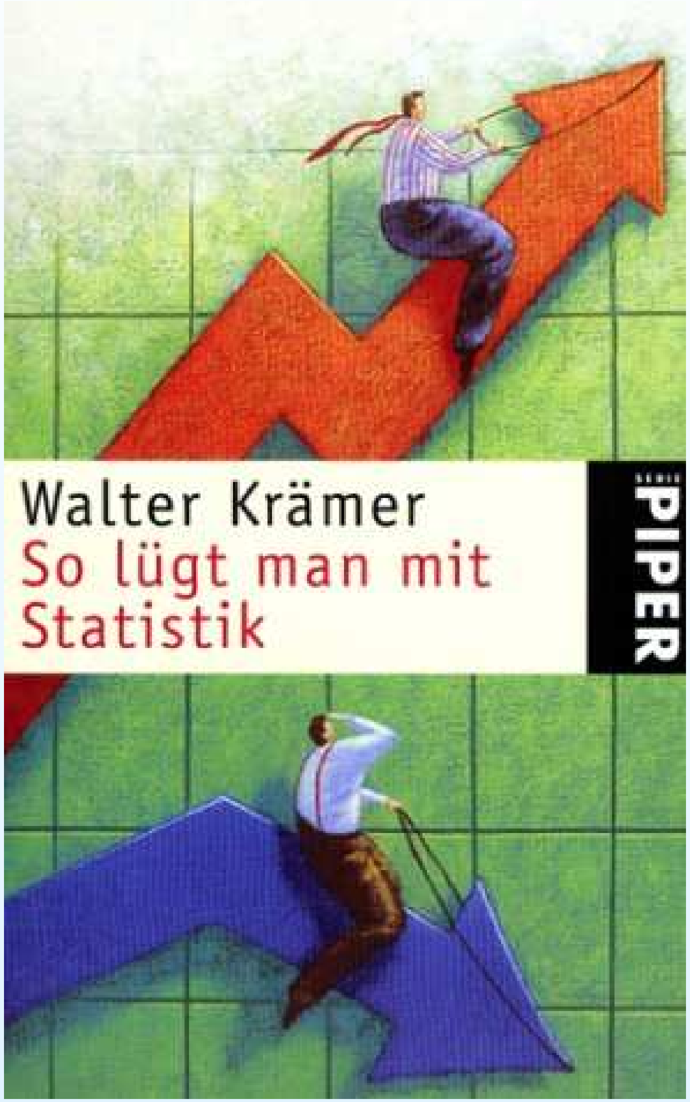
"There are three kinds of lies: lies, damn lies, and statistics"

(Benjamin Disraeli/Mark Twain)

Damned lies and statistics

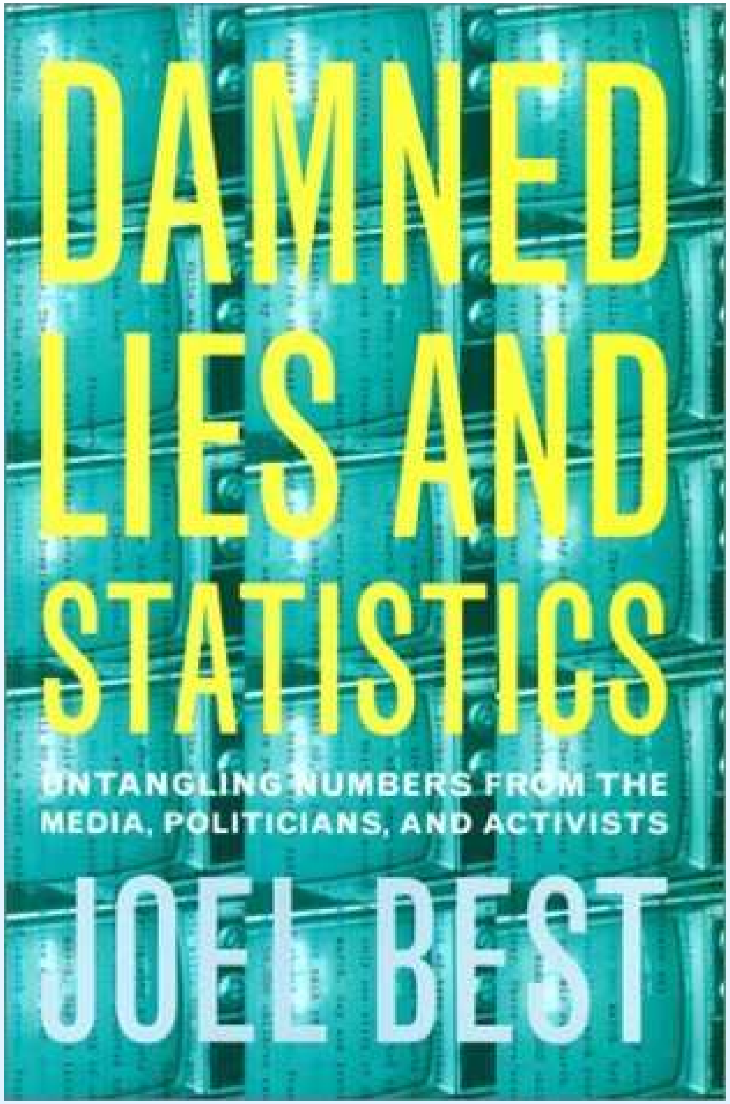
Well known subject

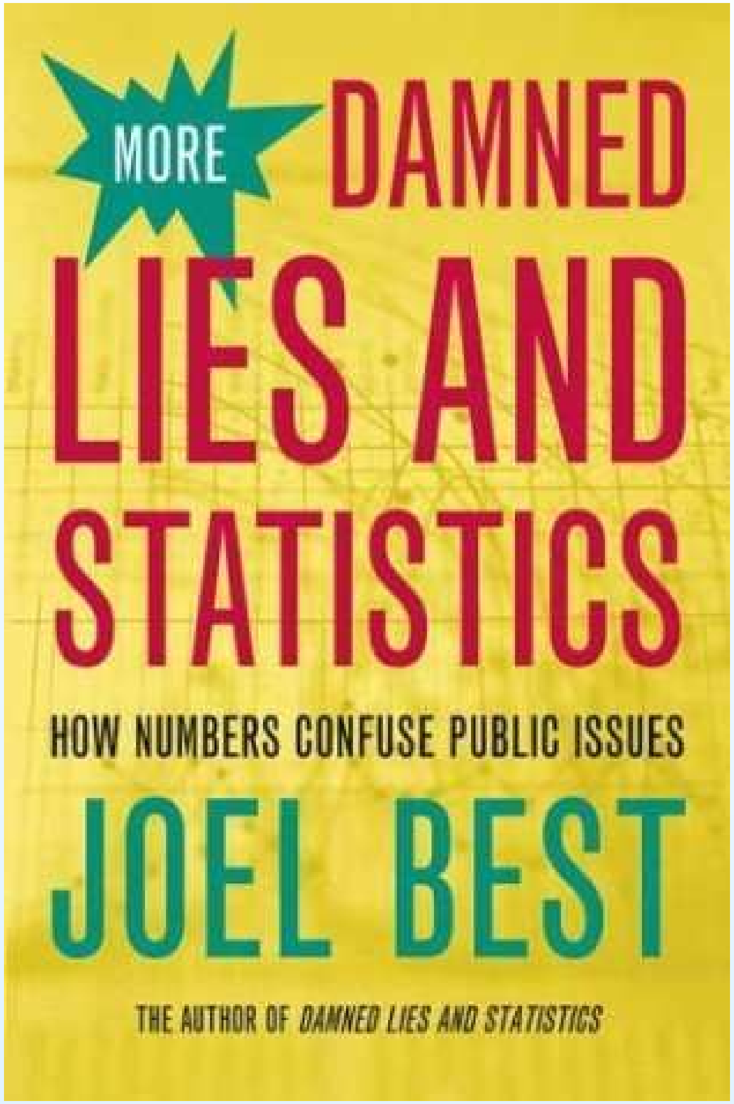




Damned lies and statistics

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."
    $\Rightarrow$ extended to scientific data.
- ***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.[†]

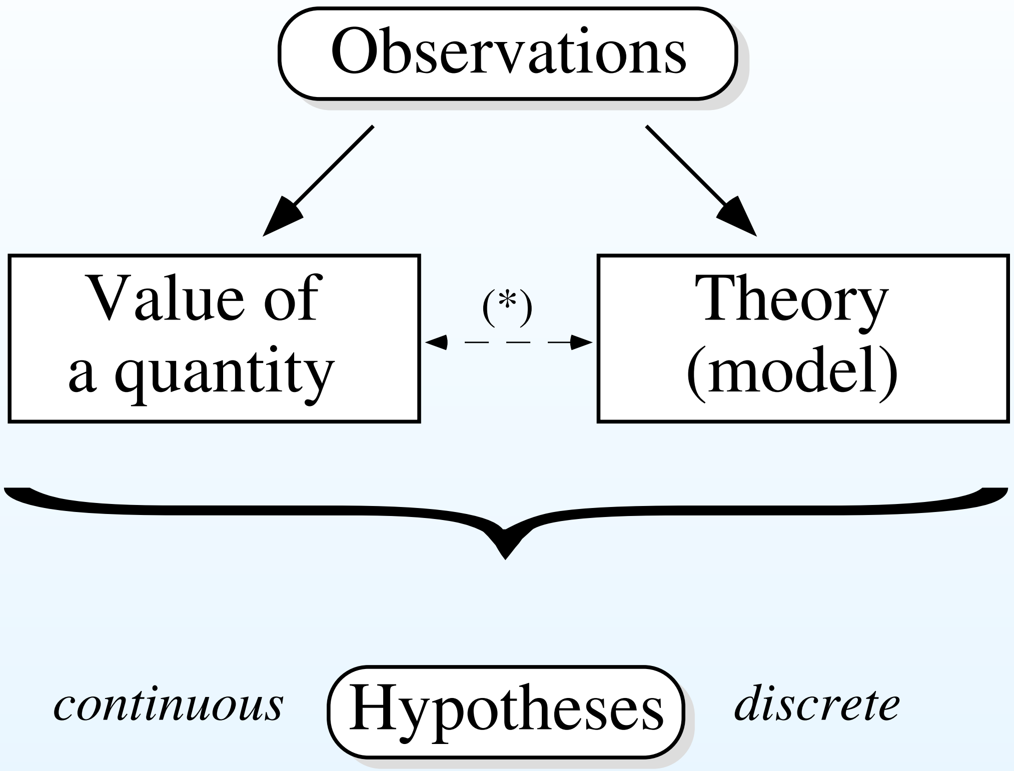
Probability theory Essentially OK, if we only consider the mathematical apparatus.

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)

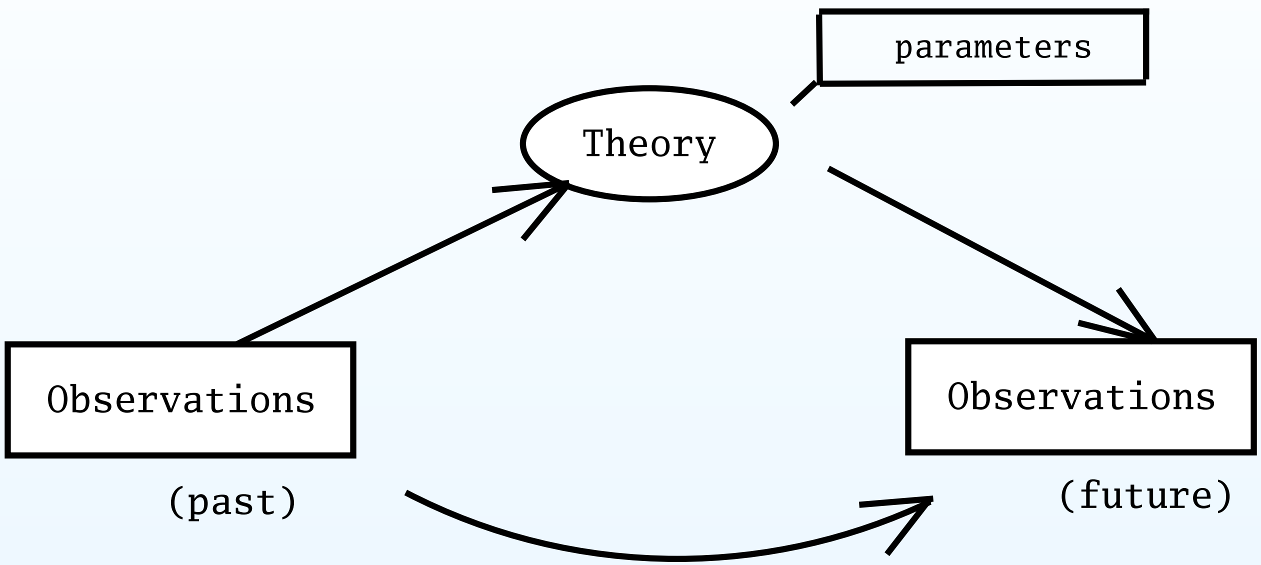
What do we really need?

Physics



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

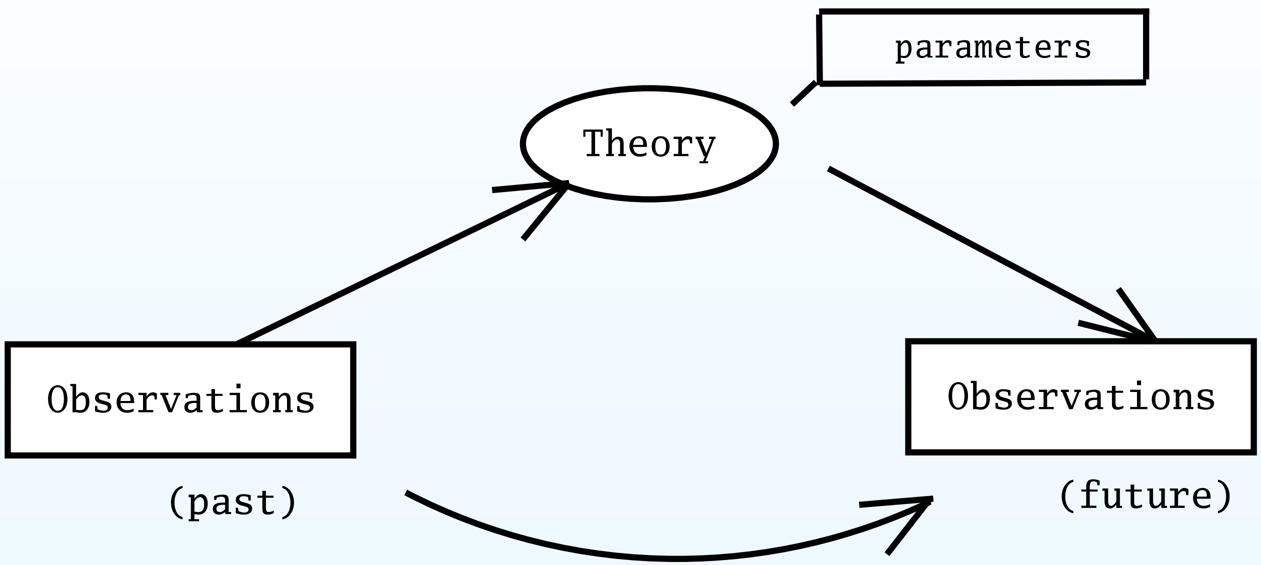
From the past to the future



The task of the physicists is to

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

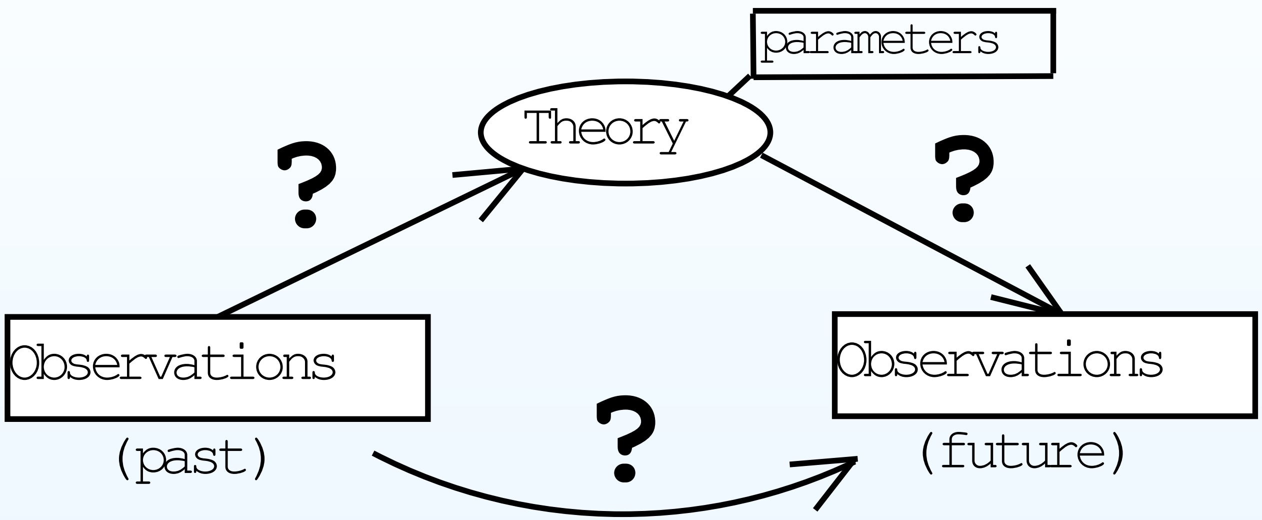
From the past to the future



Process

- neither automatic
- nor purely contemplative
  - → 'scientific method'
  - → planned experiments

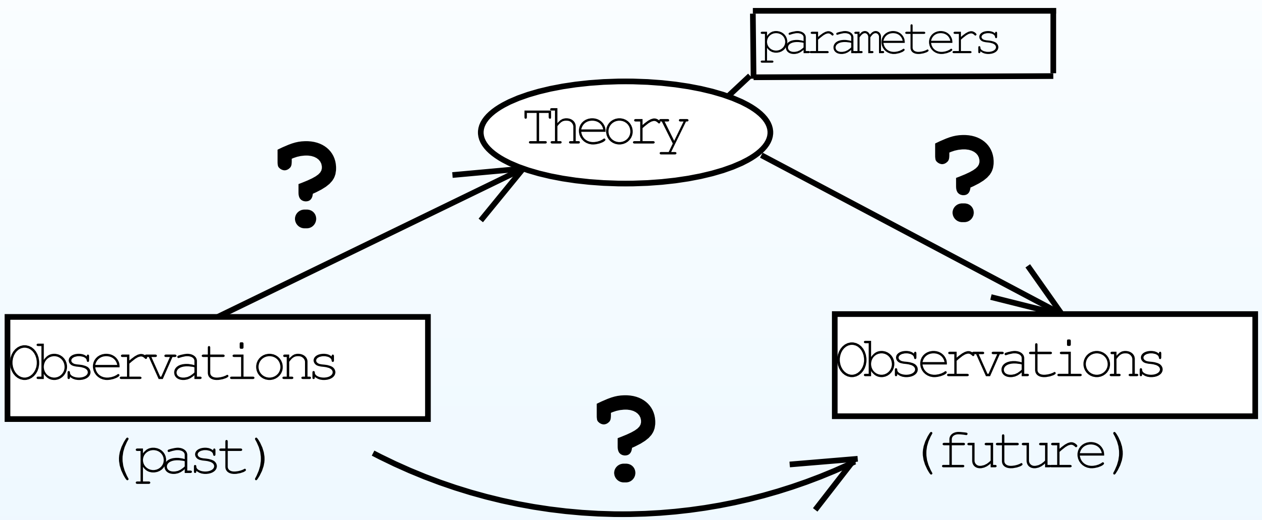
From the past to the future



⇒ Uncertainty:

1. Given the past observations, in general we are not sure about the theory parameter (and/or the theory itself)
2. Even if we were sure about theory and parameters, there could be internal (e.g. Q.M.) or external effects (initial/boundary conditions, 'errors', etc) that make the forecasting uncertain.

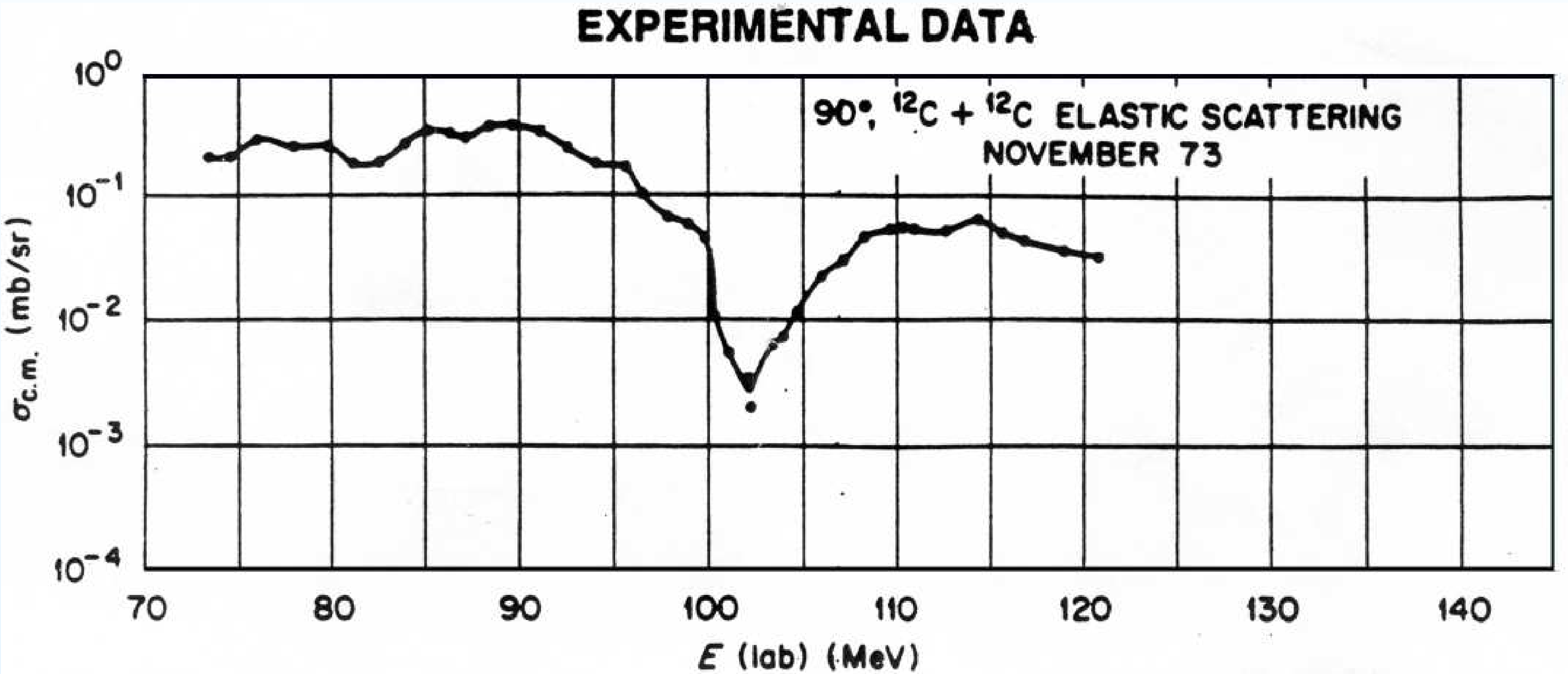
From the past to the future



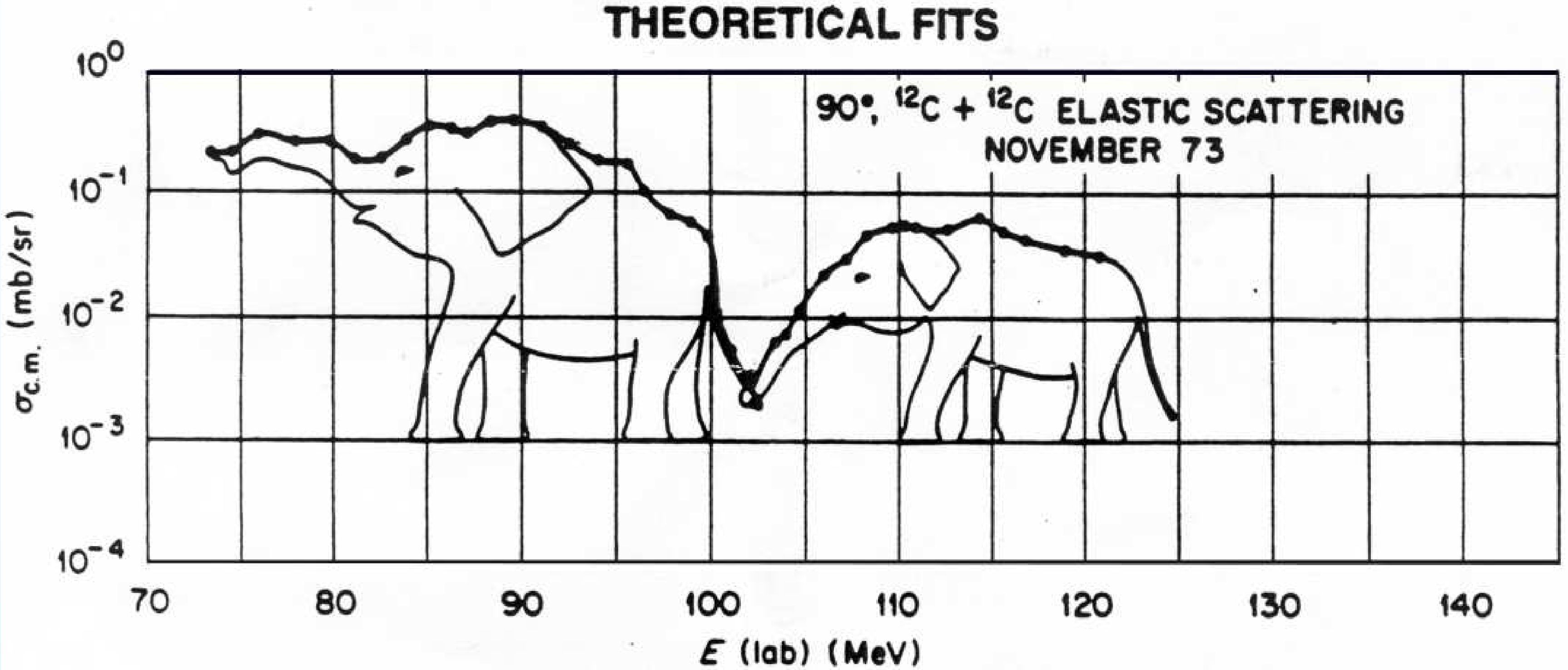
Uncertainty:

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

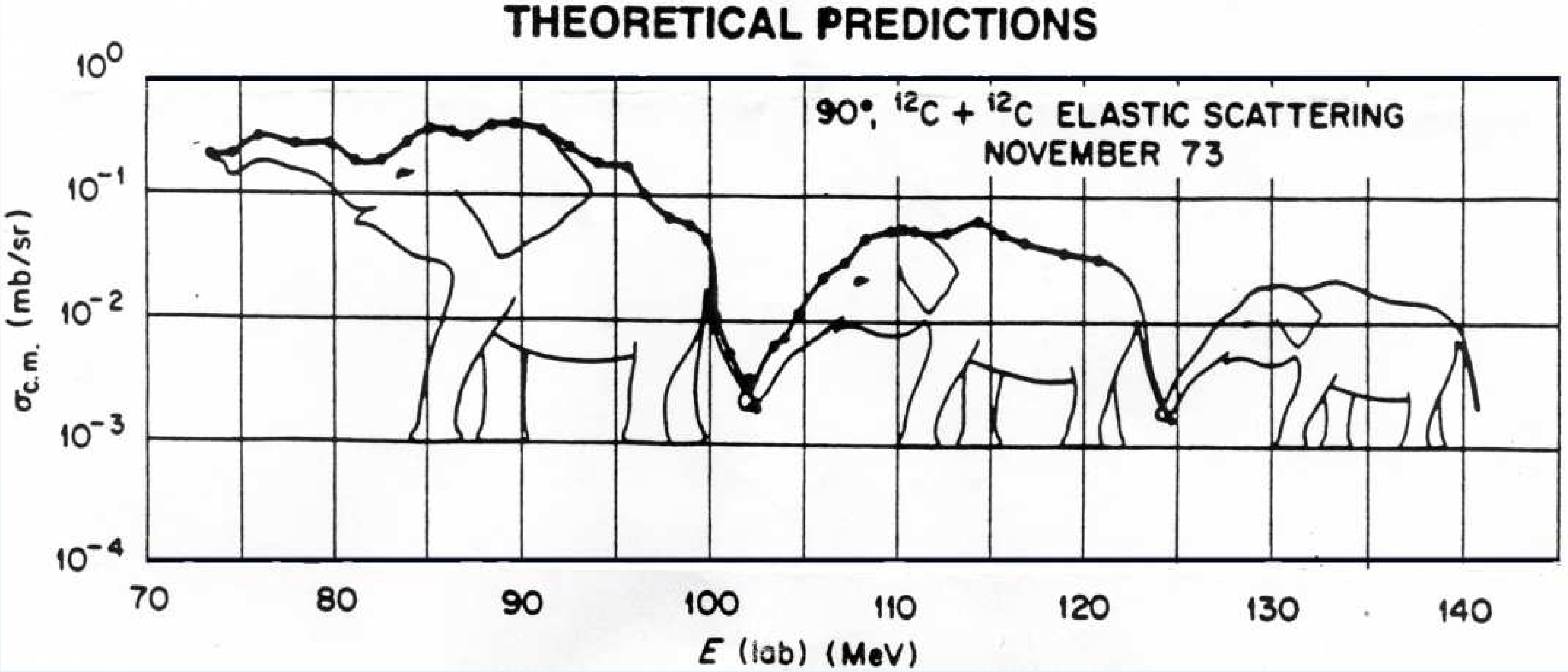
Inferential process



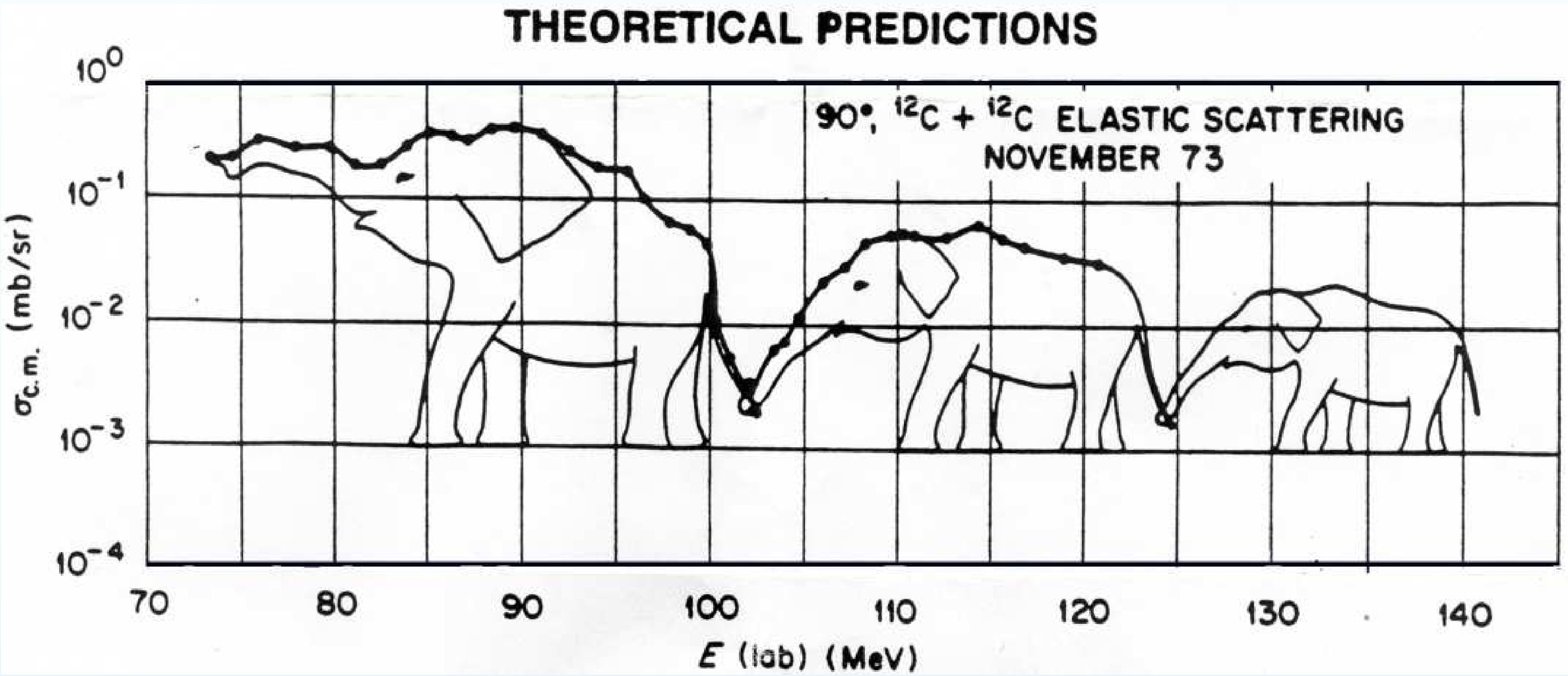
Inferential process



Inferential process



Inferential process



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

About predictions

Remember:

"Prediction is very difficult, especially if it's about the future" (Bohr)

About predictions

Remember:

“Prediction is very difficult, especially if it’s about the future” (Bohr)

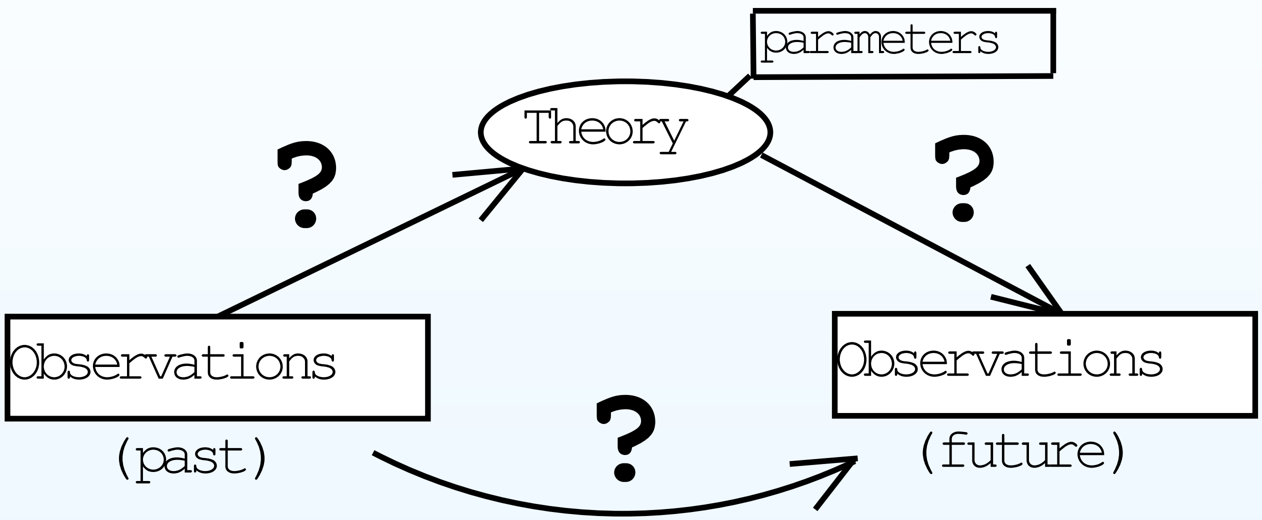
But, anyway:

“It is far better to foresee

even without certainty

than not to foresee at all” (Poincaré)

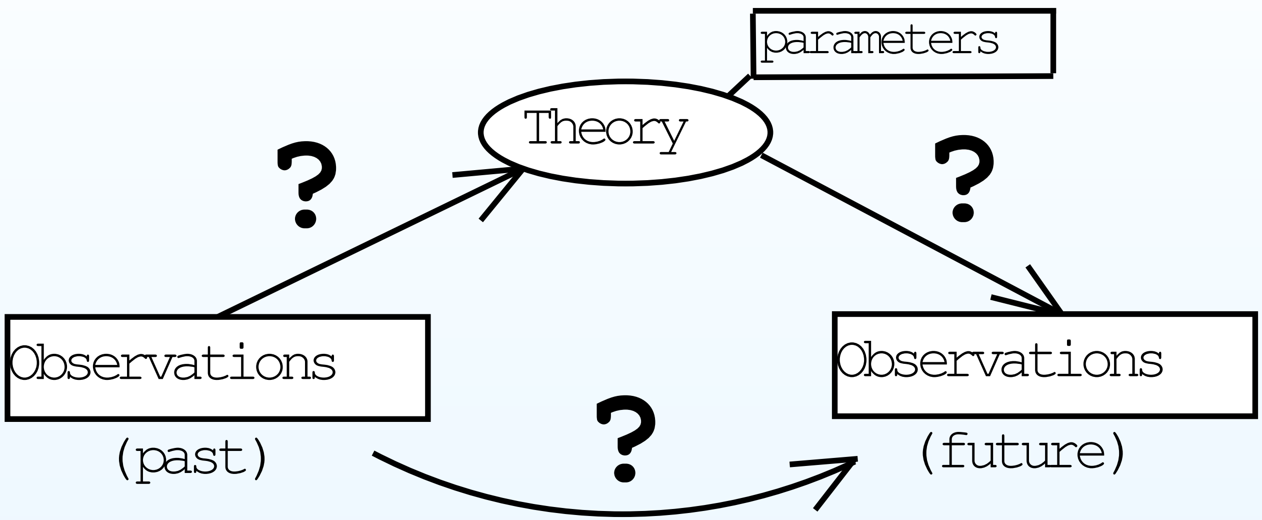
From the past to the future



Uncertainty:

Theory	$-\,?\longrightarrow$	
Past observations	$-\,?\longrightarrow$	
Theory	$-\,?\longrightarrow$	Future observations

From the past to the future



Uncertainty:

Theory — **?** $\longrightarrow$

Past observations — **?** $\longrightarrow$

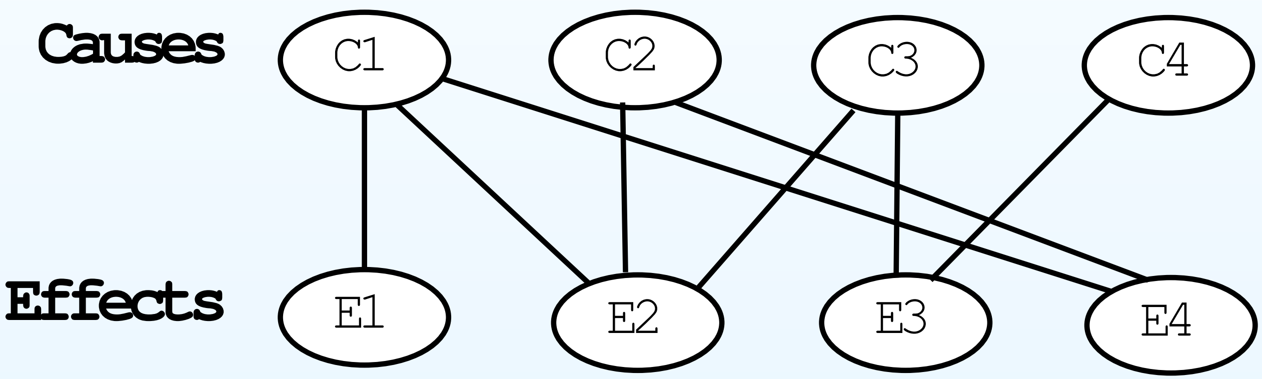
Theory — **?** $\longrightarrow$ Future observations

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

CAUSE $\Longleftrightarrow$ **EFFECT**

Causes $\longrightarrow$ effects

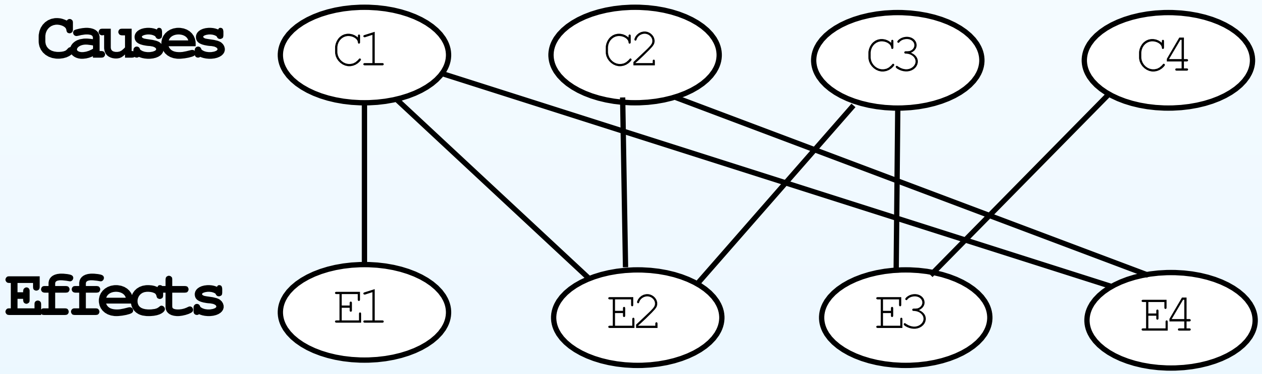
The same *apparent*† cause might produce several,different effects



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

Causes $\longrightarrow$ effects

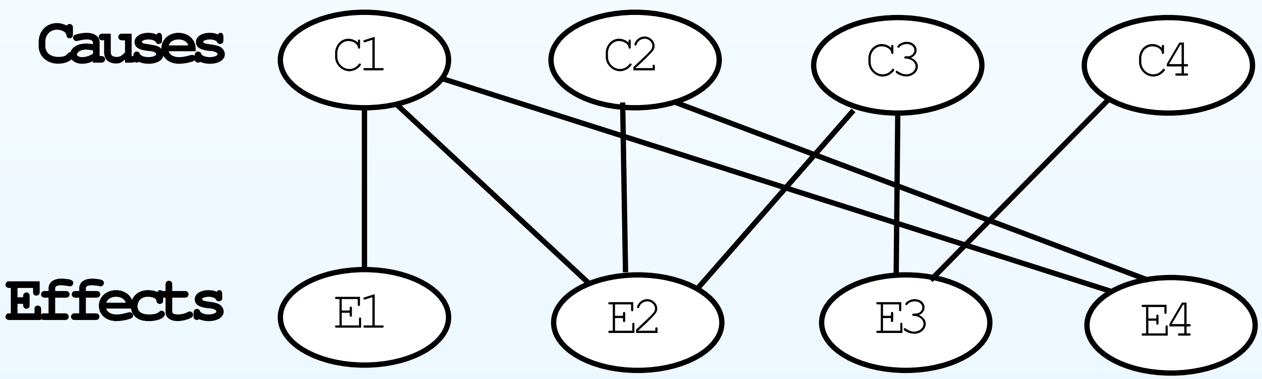
The same *apparent*† cause might produce several,different effects



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

Causes $\longrightarrow$ effects

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



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

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

The essential problem of the experimental method

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

The essential problem of the experimental method

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

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

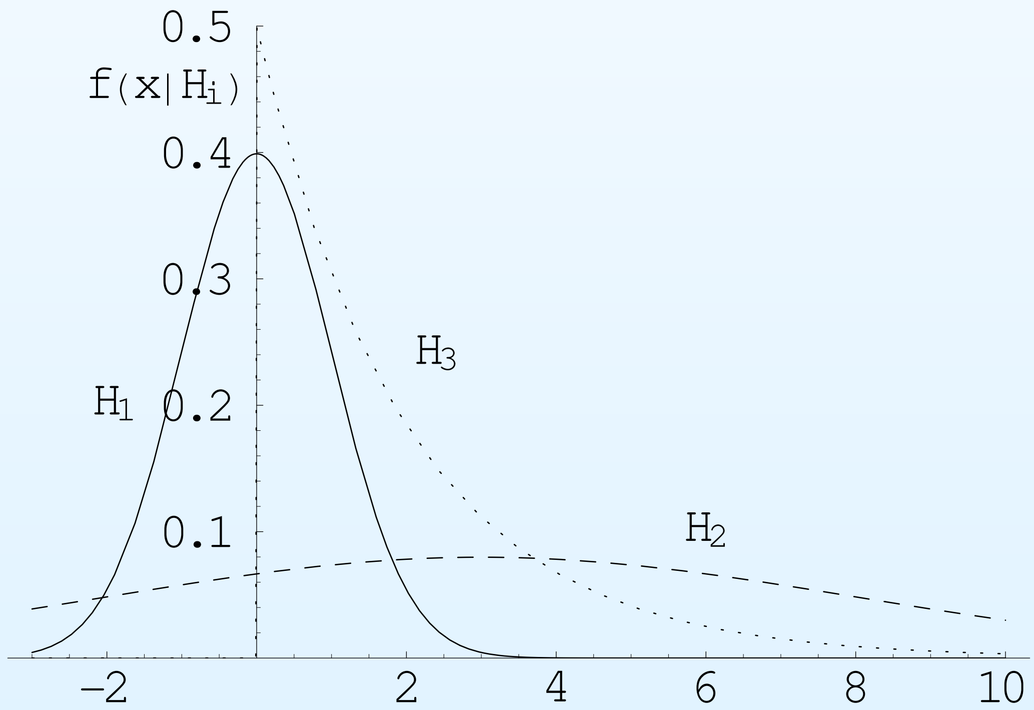
(H. Poincaré – *Science and Hypothesis*)

A numerical example

- Effect: number $x = 3$ extracted 'at random'
- Hypotheses: one of the following random generators:
  - H_1 Gaussian, with $\mu = 0$ and $\sigma = 1$
  - H_2 Gaussian, with $\mu = 3$ and $\sigma = 5$
  - H_3 Exponential, with $\tau = 2$

A numerical example

- Effect: number $x = 3$ extracted 'at random'
- Hypotheses: one of the following random generators:
  - H_1 Gaussian, with $\mu = 0$ and $\sigma = 1$
  - H_2 Gaussian, with $\mu = 3$ and $\sigma = 5$
  - H_3 Exponential, with $\tau = 2$



A numerical example

- Effect: number $x = 3$ extracted 'at random'
- Hypotheses: one of the following random generators:
  - H_1 Gaussian, with $\mu = 0$ and $\sigma = 1$
  - H_2 Gaussian, with $\mu = 3$ and $\sigma = 5$
  - H_3 Exponential, with $\tau = 2$

$\Rightarrow$ Which one to prefer?

Note: $\Rightarrow$ none of the hypotheses of this example can be excluded and, therefore, there is no way to reach a boolean conclusion. We can only state, somehow, our *rational preference*, based on the experimental result and our best knowledge of the behavior of each *model*.

A numerical example

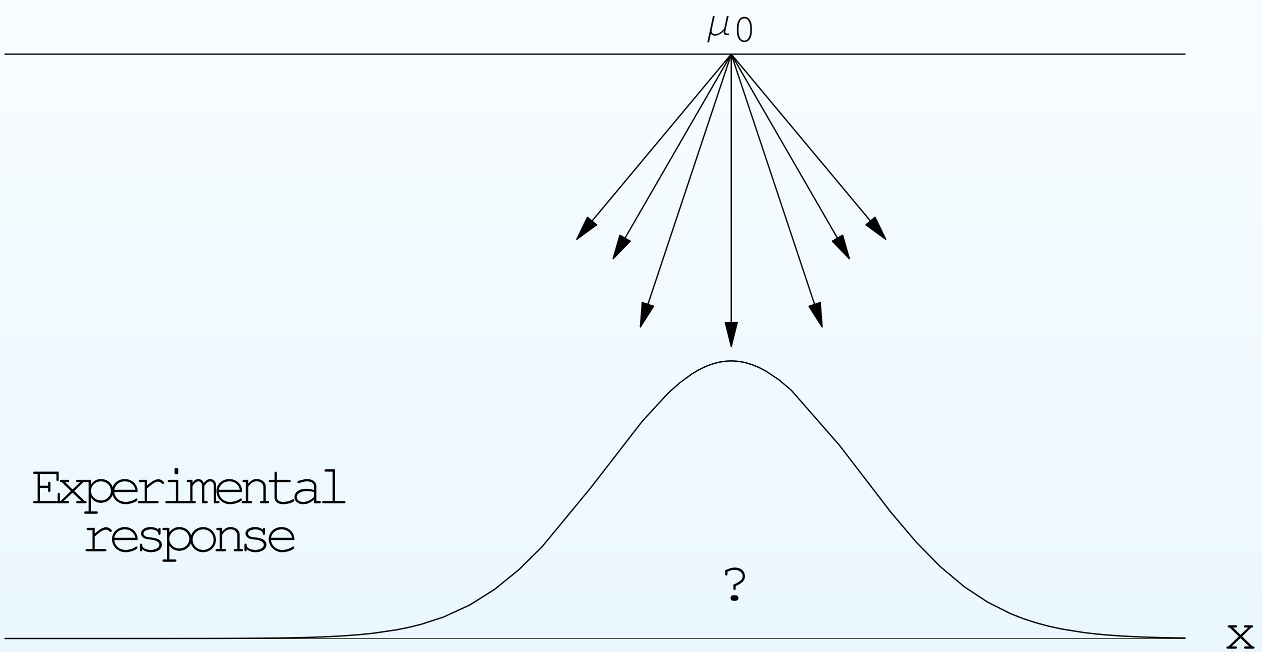
- Effect: number $x = 3$ extracted 'at random'
- Hypotheses: one of the following random generators:
  - H_1 Gaussian, with $\mu = 0$ and $\sigma = 1$
  - H_2 Gaussian, with $\mu = 3$ and $\sigma = 5$
  - H_3 Exponential, with $\tau = 2$

$\Rightarrow$ Which one to prefer?

Note: $\Rightarrow$ none of the hypotheses of this example can be excluded and, therefore, there is no way to reach a boolean conclusion. We can only state, somehow, our *rational preference*, based on the experimental result and our best knowledge of the behavior of each *model*.

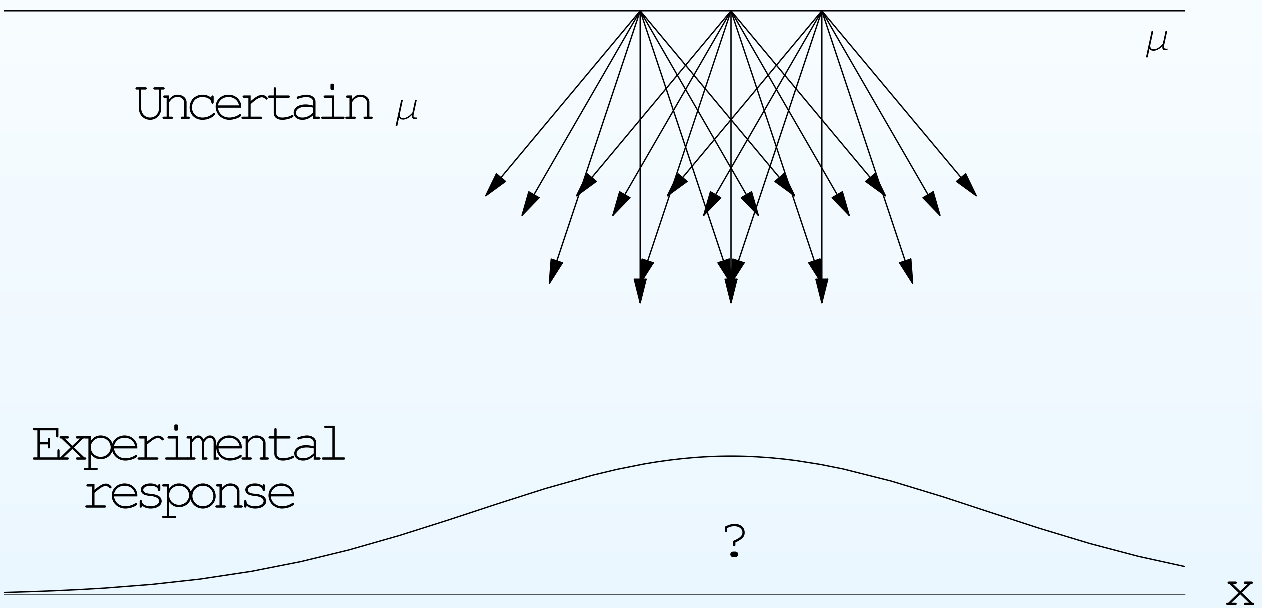
We shall come back to this example
$\rightarrow$ Let's now move to 'measuring true values'

From ‘true value’ to observations



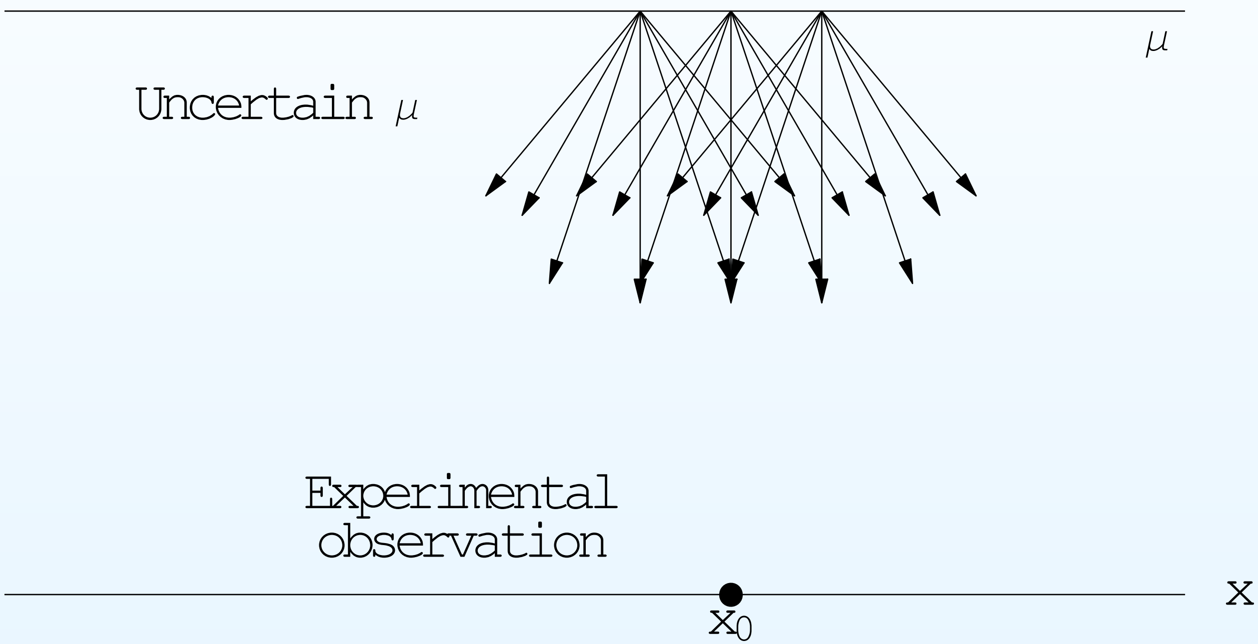
Given μ (exactly known) we are uncertain about x

From ‘true value’ to observations



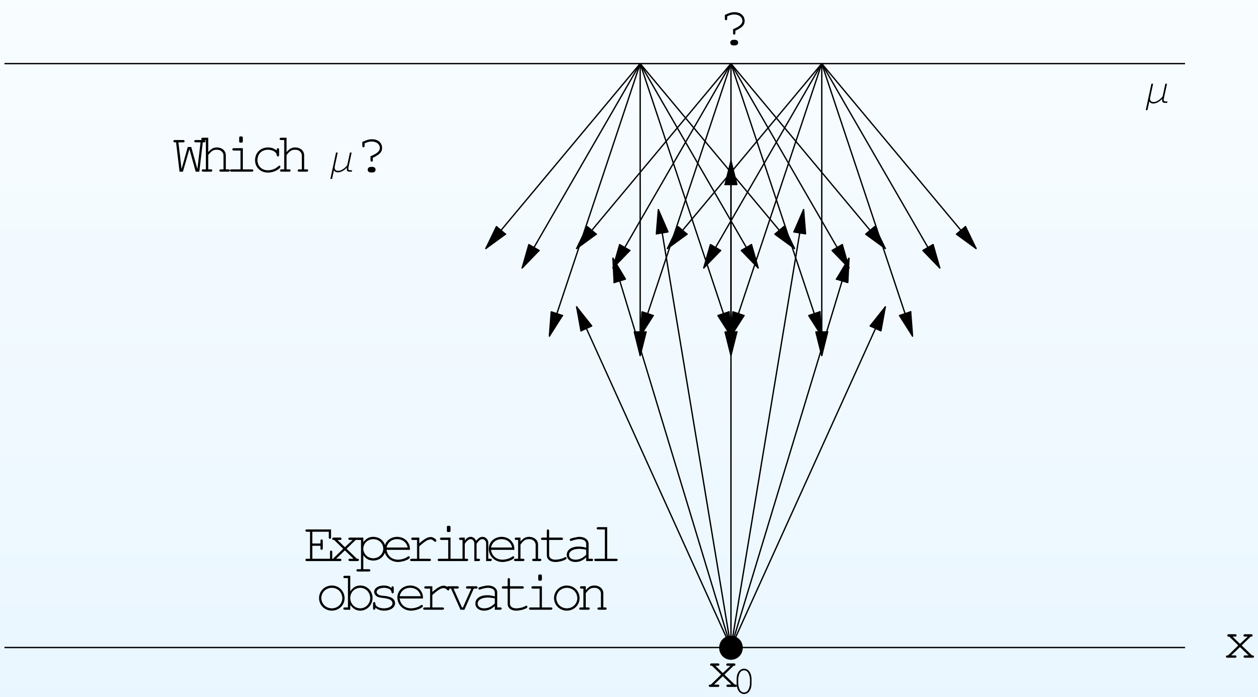
Uncertainty about μ makes us more uncertain about x

Inferring a true value



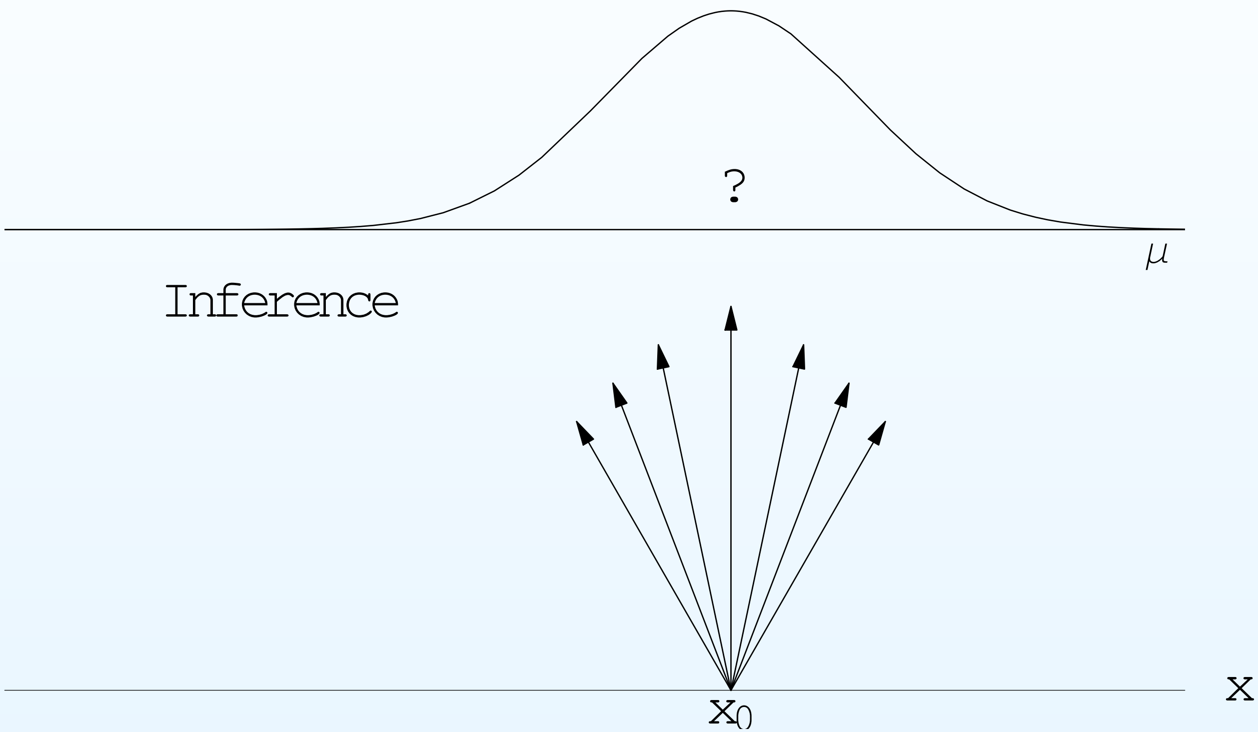
The observed data is certain: → 'true value' uncertain.

Inferring a true value



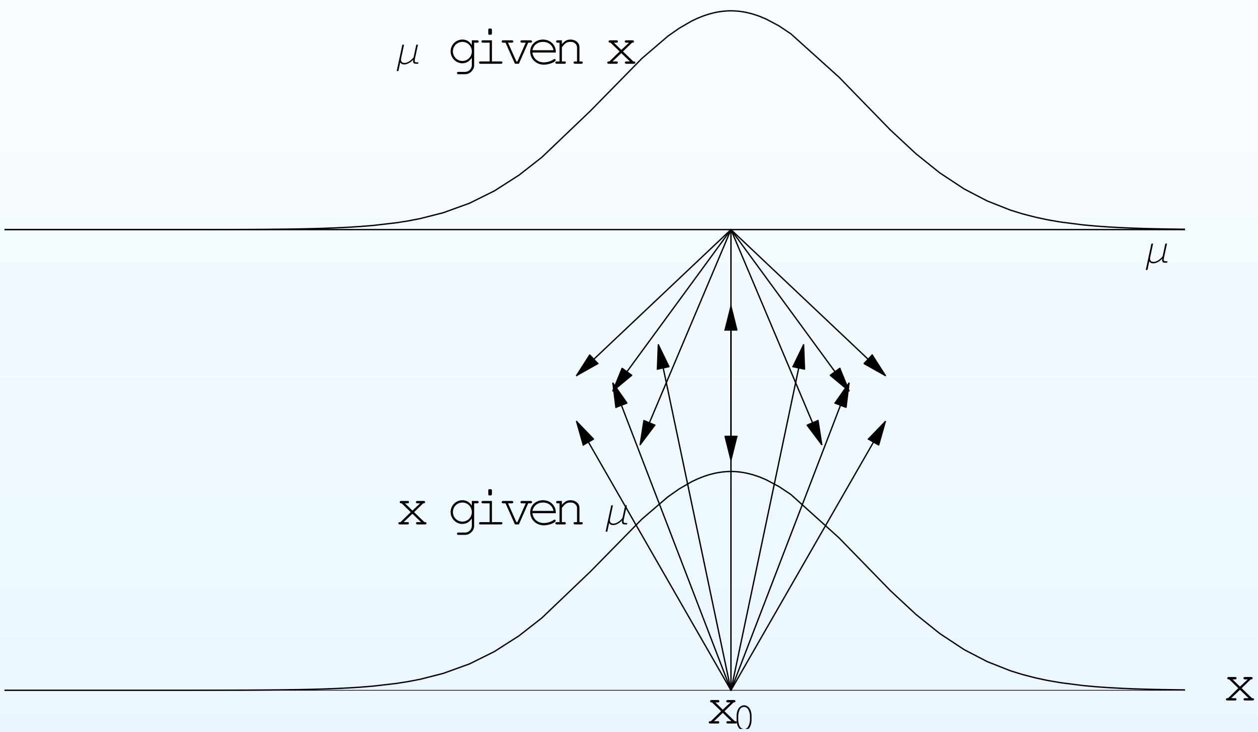
Where does the observed value of x comes from?

Inferring a true value



We are now uncertain about μ, given x.

Inferring a true value



Note the symmetry in reasoning.

Uncertainty

The human mind is used to live — and survive — in conditions of uncertainty and has developed mental categories to handle it.

Uncertainty

The human mind is used to live — and survive — in conditions of uncertainty and has developed mental categories to handle it.

As a matter of fact, although we are in a constant state of uncertainty about many events which might or might not occur,

- we can be "more or less *sure* — or *confident* — on something than on something else";
- "we consider something more or less *probable* (or *likely*)";
- or "we *believe* something more or less than something else".

Uncertainty

The human mind is used to live — and survive — in conditions of uncertainty and has developed mental categories to handle it.

As a matter of fact, although we are in a constant state of uncertainty about many events which might or might not occur,

- we can be "more or less *sure* — or *confident* — on something than on something else";
- "we consider something more or less *probable* (or *likely*)";
- or "we *believe* something more or less than something else".

We can use similar expressions, all referring to the intuitive idea of probability.

Uncertainty and probability

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

- $P(-10 < \epsilon'/\epsilon \times 10^4 < 50) >> P(\epsilon'/\epsilon \times 10^4 > 100)$
- $P(170 \le m_{top}/\text{GeV} \le 180) \approx 70\%$
- $P(M_H < 200\,\text{GeV}) > P(M_H > 200\,\text{GeV})$

Uncertainty and probability

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

- $P(-10 < \epsilon'/\epsilon \times 10^4 < 50) >> P(\epsilon'/\epsilon \times 10^4 > 100)$
- $P(170 \le m_{top}/\text{GeV} \le 180) \approx 70\%$
- $P(M_H < 200\,\text{GeV}) > P(M_H > 200\,\text{GeV})$

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

Doing Natural Science in conditions of uncertainty

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

Doing Natural Science in conditions of uncertainty

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

Indeed

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

How to quantify all that?

(Let us start with the issue of 'hypothesis tests')

How to quantify all that?

(Let us start with the issue of ‘hypothesis tests’)

- **Falsificationist approach**

  [and statistical variations over the theme].

How to quantify all that?

(Let us start with the issue of ‘hypothesis tests’)

- **Falsificationist approach**

  [and statistical variations over the theme].

- **Probabilistic approach**

  [In the sense that probability theory is used throughly]

Falsificationism

Usually associated to the name of Popper

Considered by many scientists the *key to scientific progress*.

Falsificationism

Usually associated to the name of Popper

Considered by many scientists the *key to scientific progress*.

$$\textbf{if } C_i \not\longrightarrow E, \textbf{ then } E_{\textbf{obs}} \not\longrightarrow C_i$$

$\Rightarrow$ Causes that cannot produce observed effects are ruled out ('falsified').

Falsificationism

Usually associated to the name of Popper

Considered by many scientists the *key to scientific progress*.

$$\textbf{if } C_i \not\longrightarrow E, \textbf{ then } E_{\textbf{obs}} \not\longrightarrow C_i$$

$\Rightarrow$ Causes that cannot produce observed effects are ruled out ('falsified').

It seems OK, but it is naive for several aspects.

Falsificationism

Usually associated to the name of Popper

Considered by many scientists the *key to scientific progress*.

$$\textbf{if } C_i \not\longrightarrow E, \textbf{ then } E_{\textbf{obs}} \not\longrightarrow C_i$$

$\Rightarrow$ Causes that cannot produce observed effects are ruled out ('falsified').

It seems OK, but it is naive for several aspects.

Let start realizing that the method is analogous with method of the proof by contradiction of classical, deductive logic.

Falsificationism

Usually associated to the name of Popper

Considered by many scientists the *key to scientific progress*.

$$\text{if } C_i \not\longrightarrow E, \text{ then } E_{\text{obs}} \not\longrightarrow C_i$$

⇒ Causes that cannot produce observed effects are ruled out ('falsified').

It seems OK, but it is naive for several aspects.

Let start realizing that the method is analogous with method of the proof by contradiction of classical, deductive logic.

- Assume that a hypothesis is true
- Derive 'all' logical consequence
- If (at least) one of the consequences is known to be false, then the hypothesis is declared false.

Falsificationism? OK, but...

- What to do of all hypotheses that are not falsified? (Limbus? Get stuck?)

Falsificationism? OK, but. . .

- What to do of all hypotheses that are not falsified? (Limbus? Get stuck?)
  ". . . This little episode, it seems to me, is 179 degrees or so out of phase from Popper's idea that we make progress by falsifying theories"
  [F. Wilczek - *From "Not Wrong" to (Maybe) Right,* Physics-0403115]

Falsificationism? OK, but...

- What to do of all hypotheses that are not falsified? (Limbus? Get stuck?)
  "... This little episode, it seems to me, is 179 degrees or so out of phase from Popper's idea that we make progress by falsifying theories"
  [F. Wilczek - *From "Not Wrong" to (Maybe) Right*, Physics-0403115]
- What to do is nothing of what can be observed is incompatible with the hypothesis (or with many hypotheses)?
  E.g. H_i being a Gaussian $f(x \,|\, \mu_i, \sigma_i)$
  $\Rightarrow$ Given any pair or parameters $\{\mu_i, \sigma_i\}$, all values of x between $-\infty$ and $+\infty$ are <u>possible</u>.
  $\Rightarrow$ Having observed any value of x, <u>none</u> of H_i can be, strictly speaking, <u>falsified</u>.

Falsificationism and statistics

…then, statisticians have invented the “hypothesis tests”

in which the impossible is replaced by the improbable!

But from the impossible to the improbable there is not just a question of quantity, but a question of quality.

This mechanism, logically flawed, is particularly perverse, because deeply rooted in most people, due to education, but is not supported by logic.

$\Rightarrow$ Basically responsible of all fake claims of discoveries in the past decades.

[I am particularly worried about claims concerning our health, or the status of the planet, of which I have no control of the experimental data.]

Falsificationism and statistics

…then, statisticians have invented the "hypothesis tests"

in which the impossible is replaced by the improbable!

But from the impossible to the improbable there is not just a question of quantity, but a question of quality.

This mechanism, logically flawed, is particularly perverse, because deeply rooted in most people, due to education, but is not supported by logic.

⇒ Basically responsible of all fake claims of discoveries in the past decades.

[I am particularly worried about claims concerning our health, or the status of the planet, of which I have no control of the experimental data.]

- Last minute example from epidemiology

In summary

A) **if** $C_i \not\longrightarrow E$, and **we observe** E

$\Rightarrow C_i$ is impossible ('false')

In summary

A) **if** $C_i \not\longrightarrow E$, and **we observe** E

$\Rightarrow C_i$ is impossible ('false')

B) **if** $C_i \xrightarrow[\text{small probability}]{} E$, and **we observe** E

$\Rightarrow C_i$ has small probability to be true

"most likely false"

In summary

A) **if** $C_i \not\longrightarrow E$, and **we observe** E OK

$\Rightarrow C_i$ is impossible ('false')

B) **if** $C_i \xrightarrow[\text{small probability}]{} E$, and **we observe** E

$\Rightarrow C_i$ has small probability to be true

"most likely false"

In summary

A) **if** $C_i \not\longrightarrow E$, and **we observe** E

$\Rightarrow C_i$ is impossible ('false')

OK

B) ~~**if** $C_i \xrightarrow[\text{small probability}]{} E$, and **we observe** E~~

~~$\Rightarrow C_i$ has small probability to be true~~

~~"most likely false"~~

NO

Example 1

Playing lotto

H: "I play honestly at lotto, betting on a rare combination"

E: "I win"

$$H \xrightarrow[\text{“practically impossible”}]{\hspace{6cm}} E$$

Example 1

Playing lotto

H: "I play honestly at lotto, betting on a rare combination"

E: "I win"

$$\begin{tikzcd}[column sep=huge] H \arrow[r, "\text{“practically impossible”}"'] & E \arrow[l, bend left=60] \end{tikzcd}$$

"practically to exclude"

Example 1

Playing lotto

H: "I play honestly at lotto, betting on a rare combination"

E: "I win"

$$\begin{tikzcd} H \arrow[rr, "\text{“practically impossible”}"'] & & E \arrow[ll, bend left=90] \end{tikzcd}$$

"practically to exclude"

$\Rightarrow$ almost certainly I have cheated. . .

(or it is false that I won. . .)

Example 2

An Italian citizen is selected at random to undergo an AIDS test. Performance of clinical trial is not perfect, as customary:

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

H_1='HIV' (Infected) E_1 = Positive

H_2='$\overline{\text{HIV}}$' (Healthy) E_2 = Negative

Example 2

An Italian citizen is selected at random to undergo an AIDS test. Performance of clinical trial is not perfect, as customary:

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

$$
\begin{tikzcd}
H_1 = \text{'HIV' (Infected)} \arrow[r] & E_1 = \text{Positive} \\
H_2 = \text{'}\overline{\text{HIV}}\text{' (Healthy)} \arrow[r] \arrow[ur] & E_2 = \text{Negative}
\end{tikzcd}
$$

Example 2

An Italian citizen is selected at random to undergo an AIDS test. Performance of clinical trial is not perfect, as customary:

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

$$
\begin{array}{lcl}
H_1\text{='HIV' (Infected)} & \longrightarrow & E_1 = \text{Positive} \\
H_2\text{='}\overline{\text{HIV}}\text{' (Healthy)} & \longrightarrow & E_2 = \text{Negative}
\end{array}
$$

Result: $\Rightarrow$ Positive

Example 2

An Italian citizen is selected at random to undergo an AIDS test. Performance of clinical trial is not perfect, as customary:

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

? H_1='HIV' (Infected) $\longleftarrow$ E_1 = Positive

? H_2='$\overline{\text{HIV}}$' (Healthy) E_2 = Negative

Result: $\Rightarrow$ Positive

Infected or healthy?

Example 2

Being $P(\text{Pos} \,|\, \overline{\text{HIV}}) = 0.2\%$ and having observed ‘Positive’, can we say?

- ”It is practically impossible that the person is healthy, since it was practically impossible that an healthy person would result positive”

Example 2

Being $P(\text{Pos}\,|\,\overline{\text{HIV}}) = 0.2\%$ and having observed ‘Positive’, can we say?

- ”It is practically impossible that the person is healthy, since it was practically impossible that an healthy person would result positive”
- “There is only 0.2% probability that the person has no HIV”

Example 2

Being $P(\text{Pos} \,|\, \overline{\text{HIV}}) = 0.2\%$ and having observed ‘Positive’, can we say?

- ”It is practically impossible that the person is healthy, since it was practically impossible that an healthy person would result positive”
- “There is only 0.2% probability that the person has no HIV”
- “We are 99.8% confident that the person is infected?”

Example 2

Being $P(\text{Pos} \,|\, \overline{\text{HIV}}) = 0.2\%$ and having observed 'Positive', can we say

- "It is practically impossible that the person is healthy, since it was practically impossible that an healthy person would result positive"
- "There is only 0.2% probability that the person has no HIV"
- "We are 99.8% confident that the person is infected?"
- "The hypothesis H_1=Healthy is ruled out with 99.8% C.L."

?

Example 2

Being $P(\text{Pos} \,|\, \overline{\text{HIV}}) = 0.2\%$ and having observed ‘Positive’,
can we say

- ~~”It is practically impossible that the person is healthy, since it was practically impossible that an healthy person would result positive”~~
- ~~“There is only 0.2% probability that the person has no HIV”~~
- ~~“We are 99.8% confident that the person is infected?”~~
- ~~“The hypothesis H_1=Healthy is ruled out with 99.8% C.L.”~~

? NO

Instead, $P(\overline{\text{HIV}} \,|\, \text{Pos, random Italian}) \approx 45\%$
(We will see in the sequel how to evaluate it correctly)

Example 2

Being $P(\text{Pos} \mid \overline{\text{HIV}}) = 0.2\%$ and having observed 'Positive', can we say

- ~~"It is practically impossible that the person is healthy, since it was practically impossible that an healthy person would result positive"~~
- ~~"There is only 0.2% probability that the person has no HIV"~~
- ~~"We are 99.8% confident that the person is infected?"~~
- ~~"The hypothesis H_1=Healthy is ruled out with 99.8% C.L."~~

? NO

Instead, $P(\overline{\text{HIV}} \mid \text{Pos, random Italian}) \approx 45\%$

$\Rightarrow$ Serious mistake! (not just 99.8% instead of 98.3% or so)

Confidence arrows

$P(\text{Pos}\,|\,\text{HIV})$ and $P(\text{Pos}\,|\,\overline{\text{HIV}})$ express our confidence that the analysis will give ‘Positive’ if we are sure about the health of the patient.

H_1 = ’HIV’ (Infected) $\longrightarrow$ E_1 = Positive **?**

H_2 = ’$\overline{\text{HIV}}$’ (Healthy)

Confidence arrows

? H_1 = 'HIV' (Infected) $\longleftarrow$ E_1 = Positive

? H_2 = '$\overline{\text{HIV}}$' (Healthy)

Our confidence about the status of health of the patient, *given the observation 'Positive'*, is expressed by $P(\text{HIV}\,|\,\text{Pos})$ and $P(\overline{\text{HIV}}\,|\,\text{Pos})$.

- $\Rightarrow$ In the general case the *intensity* of the confidence arrows changes if the arrow direction is reversed!
- $\Rightarrow$ We need to learn how they do.

Similar arbitrary inversion in upper limits

Imagine we have done a counting experiment, believed to be described by a Poisson distribution.

- Result $x = 0$

$\Rightarrow$ What can we tell about λ?
(Remember that the physical parameter is $r = \lambda/\Delta T$)

Similar arbitrary inversion in upper limits

Imagine we have done a counting experiment, believed to be described by a Poisson distribution.

- Result $x = 0$

⇒ What can we tell about λ?
(Remember that the physical parameter is $r = \lambda/\Delta T$)

- All values of λ (or r) are in principle possible
- …though, we do not believe them equally likely.

Similar arbitrary inversion in upper limits

Imagine we have done a counting experiment, believed to be described by a Poisson distribution.

- Result $x = 0$

$\Rightarrow$ What can we tell about λ?
(Remember that the physical parameter is $r = \lambda/\Delta T$)

- All values of λ (or r) are in principle possible
- ... though, we do not believe them equally likely.
- Standard way to report the result: 95% C.L. upper limit:

$$\lambda \leq 3 \text{ @ } 95\%\text{C.L.}$$

- Why?

Similar arbitrary inversion in upper limits

“Because if I repeat a large number of experiments,
I get $x = 0$ *in 5% of the cases”*

Similar arbitrary inversion in upper limits

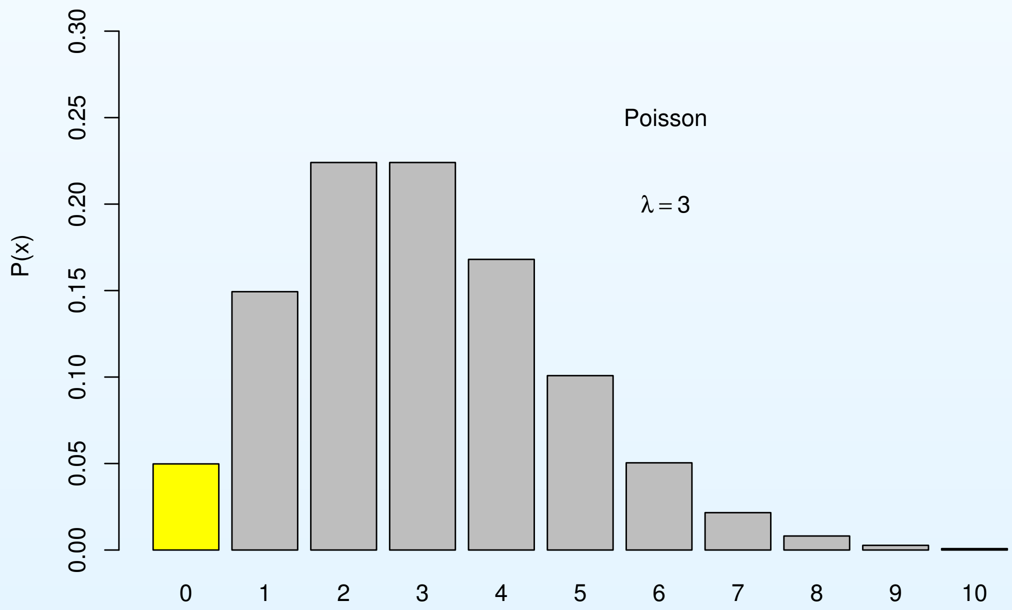
"Because if I repeat a large number of experiments, I get $x = 0$ in 5% of the cases"

$$\Rightarrow \; P(x = 0 \,|\, \mathcal{P}_{\lambda=3}) = 5\%.$$

Similar arbitrary inversion in upper limits

"Because if I repeat a large number of experiments, I get $x = 0$ in 5% of the cases"

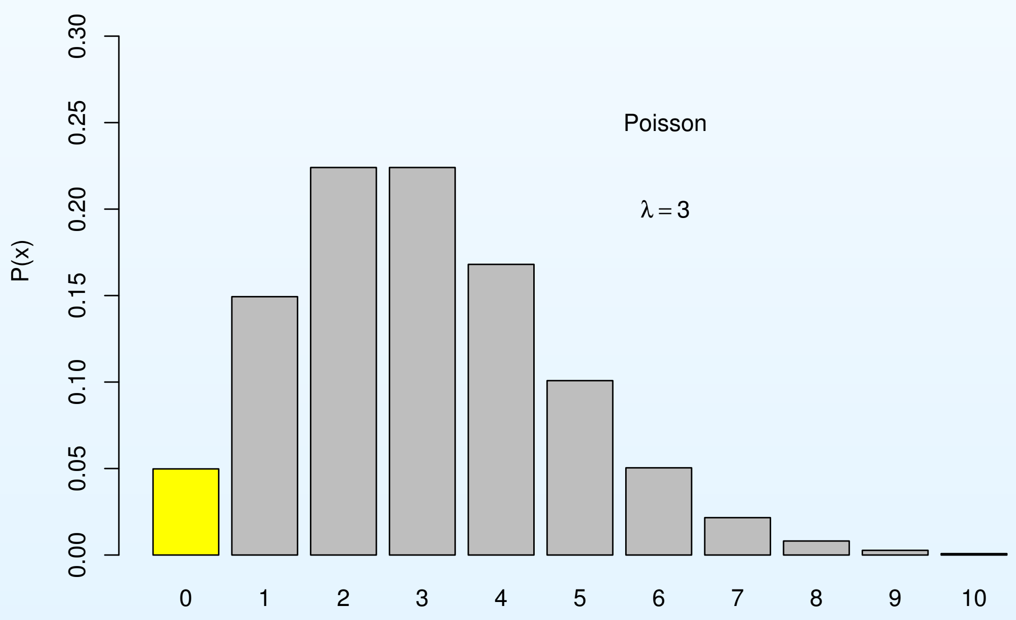
$$\Rightarrow \; P(x = 0 \,|\, \mathcal{P}_{\lambda=3}) = 5\%.$$



Similar arbitrary inversion in upper limits

"Because if I repeat a large number of experiments,
I get $x = 0$ in 5% of the cases"

$$\Rightarrow \; P(x = 0 \,|\, \mathcal{P}_{\lambda=3}) = 5\%.$$



But what has this to do with our confidence that $\lambda \geq 3$?

A little Socratic dialog

- Why?

  "Because it is defined so: The value of λ such that, if I repeat a large number of experiments, I get $x = 0$ in 5% of the cases"

A little Socratic dialog

- Why?

  "Because it is <u>defined so</u>: The value of λ such that, if I repeat a large number of experiments, I get $x = 0$ in 5% of the cases"

- But why?

  "Because this prescription defines the interval [0,3] that covers with 95% the true value of λ"

A little Socratic dialog

- Why?

  "Because it is <u>defined so</u>: The value of λ such that, if I repeat a large number of experiments, I get $x = 0$ in 5% of the cases"

- But why?

  "Because this prescription defines the interval [0,3] that covers with 95% the true value of λ"

- What does it mean?

  "If I repeat many times the experiment, in 95% of the cases the interval [0,3] 'covers' the true λ"

A little Socratic dialog

- Why?

  "Because it is <u>defined so</u>: The value of λ such that, if I repeat a large number of experiments, I get $x = 0$ in 5% of the cases"

- But why?

  "Because this prescription defines the interval [0,3] that covers with 95% the true value of λ"

- What does it mean?

  "If I repeat many times the experiment, in 95% of the cases the interval [0,3] 'covers' the true λ"

- In other words, there is 95% probability that the true λ of this physics case is in this interval, isn't it?

- *"[. . .] more or less so, though you are not allowed to talk about probability of λ. . . "*

A little Socratic dialog

- Sorry. . . . But how you know that?

  “Because this is a classical, exact method, reported in all classical books of statistics.”

A little Socratic dialog

- Sorry.... But how you know that?

  "Because this is a classical, exact method, reported in all classical books of statistics."

- Than it must be true!
  But let me understand. Does the method covers also in the other side, i.e. there is 5% probability that $[3, \infty]$ covers the true λ?

  "Right"

A little Socratic dialog

- Sorry. . . . But how you know that?

  “Because this is a classical, exact method, reported in all classical books of statistics.”

- Than it must be true!
  But let me understand. Does the method covers also in the other side, i.e. there is 5% probability that $[3, \infty]$ covers the true λ?

  “Right”

- I imagine that this kind of statement does not depend on the C.L., therefore, let us calculate the 50% C.L. upper limit on λ, i.e. 0.7

$$\lambda \geq 0.7 \quad @ \quad 50\%\text{C.L.}$$

$$\lambda \leq 0.7 \quad @ \quad 50\%\text{C.L.}$$

  “Right”

A little Socratic dialog

- Than, if, given $\lambda_B = 0.7$, such that such that $P(x=0\,|\,\mathcal{P}_{\lambda_B}) = 50\%$, and you say that this implies

$$\lambda \leq 0.7 \quad @ \quad 50\%\text{C.L.}\,,$$

  then the intervals $[0, 0.7]$ and $[0.7, \infty]$ cover with the same probability the true λ, i.e. the statements

  A: "The interval $[0, 0.7]$ includes λ"

  B: "The interval $[0.7, \infty]$ includes λ"

  have the same probability of being true.

- Yes

A little Socratic dialog

- Than, if, given $\lambda_B = 0.7$, such that such that $P(x=0\,|\,\mathcal{P}_{\lambda_B}) = 50\%$, and you say that this implies

$$\lambda \le 0.7 \quad @ \quad 50\%\text{C.L.}\,,$$

  then the intervals $[0, 0.7]$ and $[0.7, \infty]$ cover with the same probability the true λ, i.e. the statements

  A: "The interval $[0, 0.7]$ includes λ"

  B: "The interval $[0.7, \infty]$ includes λ"

  have the same probability of being true.
- Yes
- Or, that if we make a large number of experiments, in 50% of the case A will be true, in the other 50% of the cases B.
- Precisely so.

A little Socratic dialog

- Than, if, given $\lambda_B = 0.7$, such that such that $P(x = 0 \,|\, \mathcal{P}_{\lambda_B}) = 50\%$, and you say that this implies

$$\lambda \le 0.7 \quad @ \quad 50\%\text{C.L.}\,,$$

  then the intervals $[0, 0.7]$ and $[0.7, \infty]$ cover with the same probability the true λ, i.e. the statements

  A: "The interval $[0, 0.7]$ includes λ"

  B: "The interval $[0.7, \infty]$ includes λ"

  have the same probability of being true.
- Yes
- Or, that if we make a large number of experiments, in 50% of the case A will be true, in the other 50% of the cases B.
- Precisely so.
- Or, in 50% of the cases the true λ should be in the upper interval. Very interesting!

A little Socratic dialog

But people never give 50% upper limits

A little Socratic dialog

But people never give 50% upper limits

- I start to think that's what they should do, at least as an mental exercise, do get a feeling of what they state. Anyway, since many thousands of upper limit are given about rare processes, we should be prepared to see several dozens, or even hundreds of quantities to show up in the 5% side!

Uhm. . .

A little Socratic dialog

But people never give 50% upper limits

- I start to think that's what they should do, at least as an mental exercise, do get a feeling of what they state. Anyway, since many thousands of upper limit are given about rare processes, we should be prepared to see several dozens, or even hundreds of quantities to show up in the 5% side!

Uhm. . .

- But it could be even better: since many 95% C.L. limits are given more or less from zero events observed, we can easily roughly rescale upper 95% C.L. bounds into 50% C.L. bounds, just dividing the bounds by four.

A little Socratic dialog

But people never give 50% upper limits

- I start to think that's what they should do, at least as an mental exercise, do get a feeling of what they state. Anyway, since many thousands of upper limit are given about rare processes, we should be prepared to see several dozens, or even hundreds of quantities to show up in the 5% side!

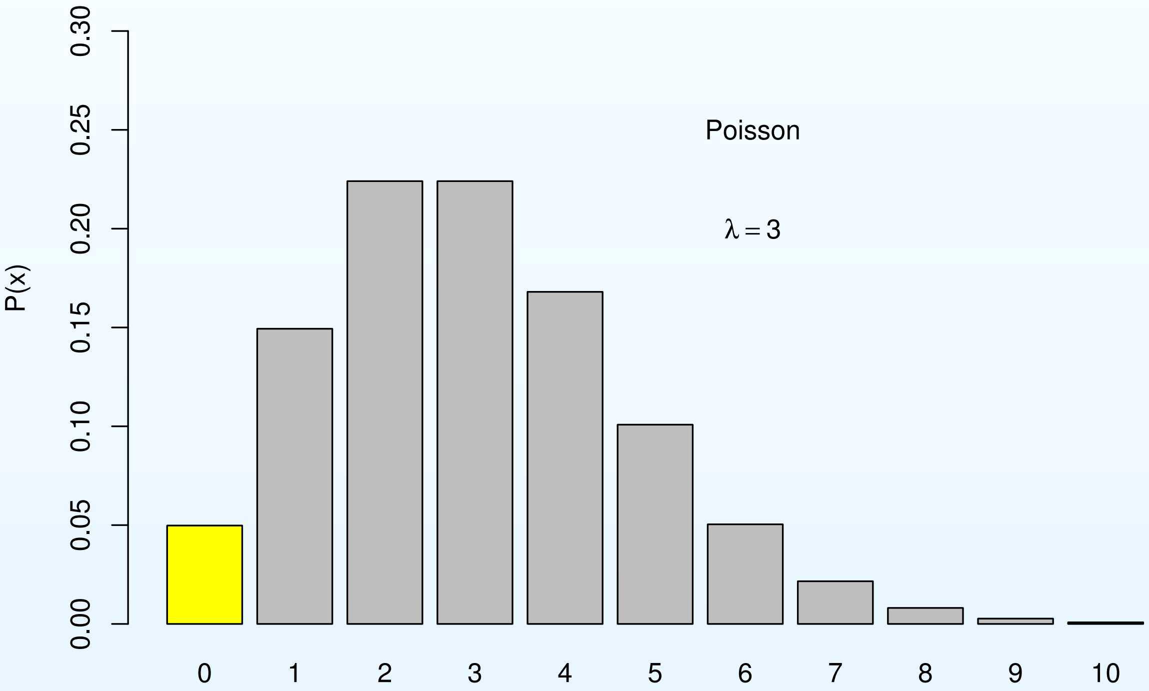
Uhm. . .

- But it could be even better: since many 95% C.L. limits are given more or less from zero events observed, we can easily roughly rescale upper 95% C.L. bounds into 50% C.L. bounds, just dividing the bounds by four.

  And 'coverage' tells that in 50% of the cases of the rescaled results the true value is in the upper side of the rescaled upper limit. Great!

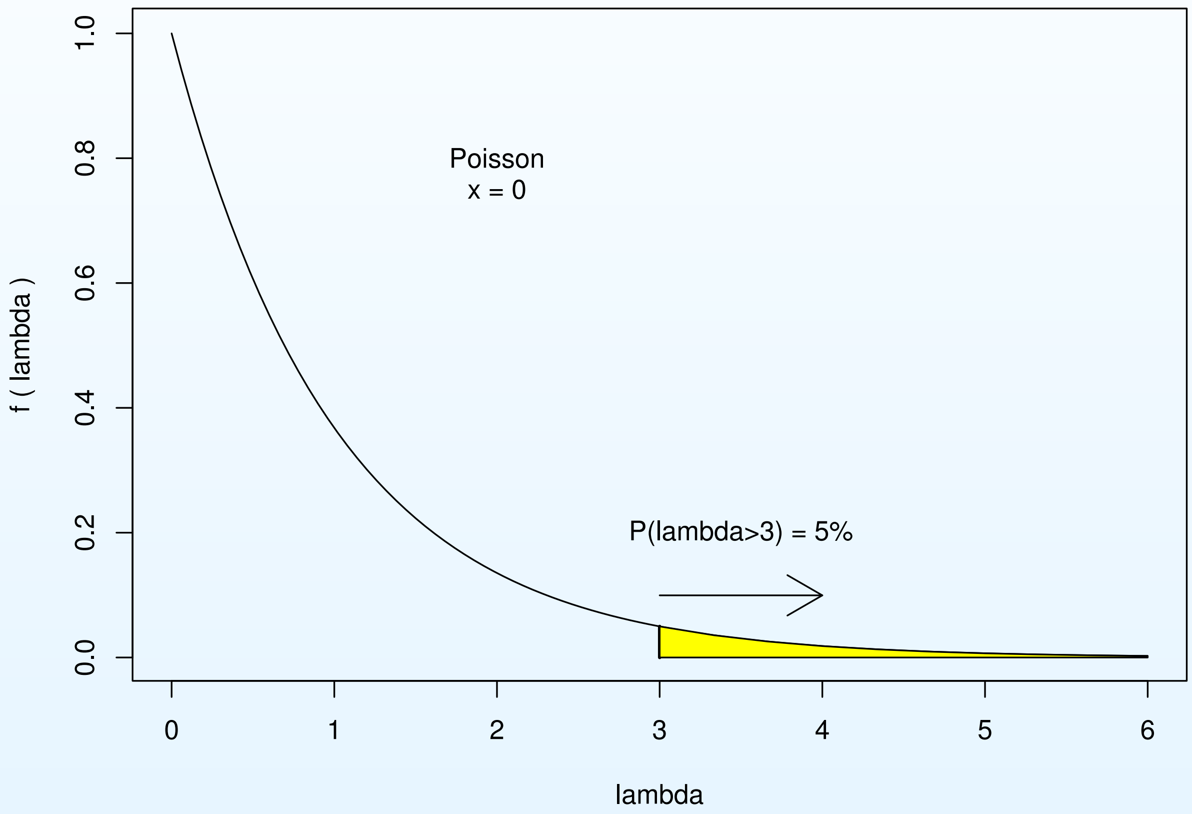
Special case of the Poisson with observed $x = 0$

Probability function of x given $\lambda = 3$



Special case of the Poisson with observed $x = 0$

Probability density function of λ given $x = 0$



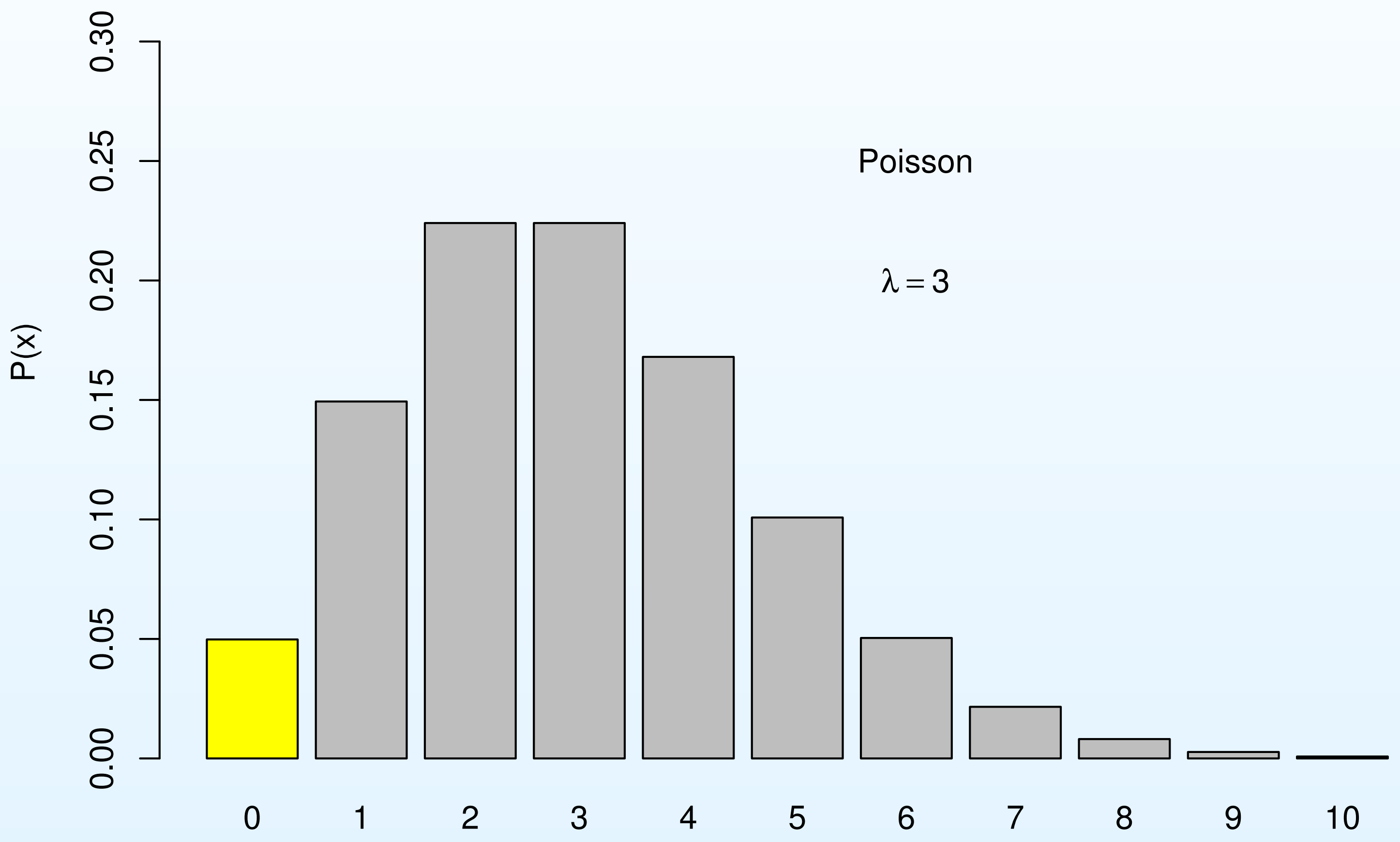
(We shall came later to the details of the calculation)

. . . but

It is not a general property

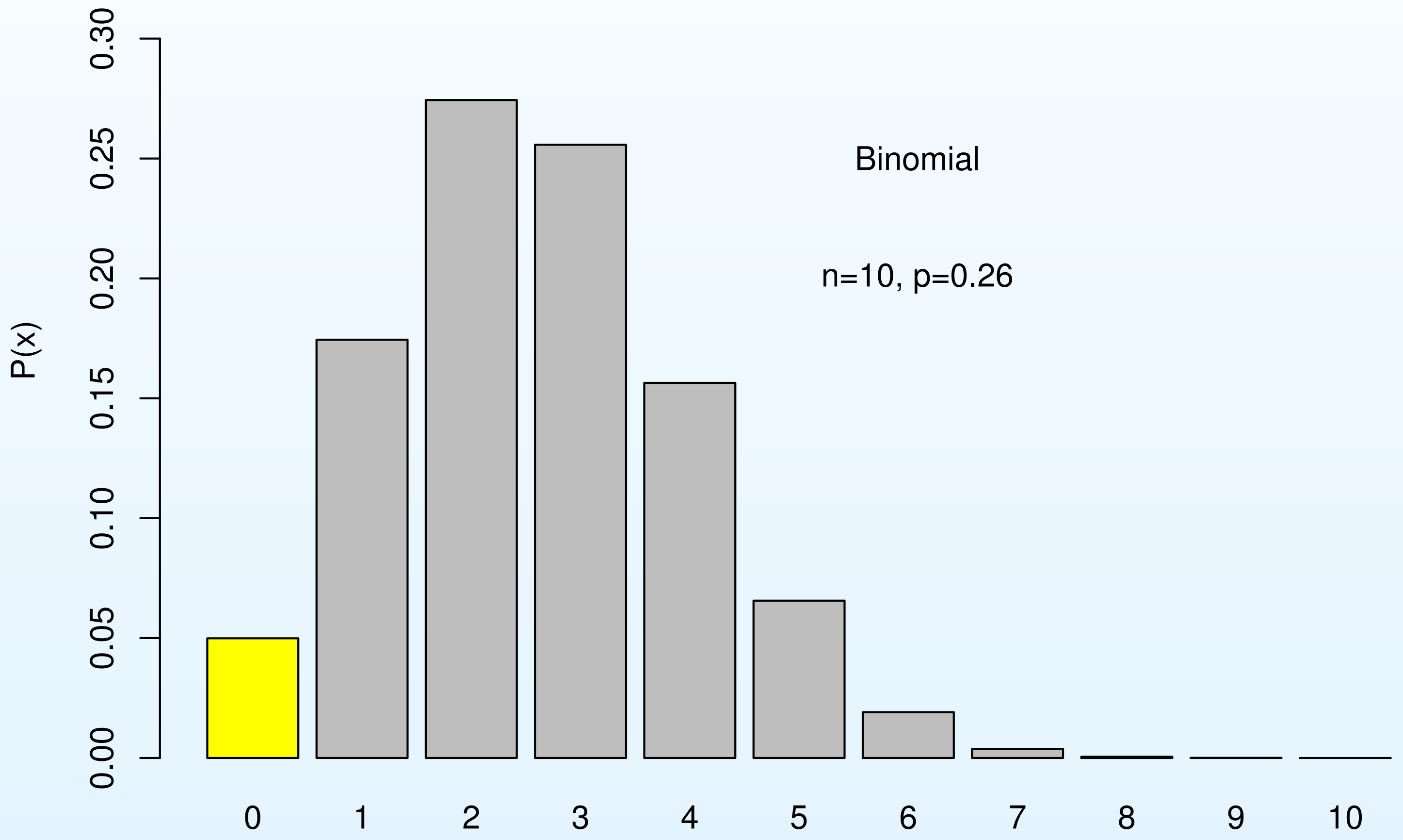
Let us check with other simple cases

A Poisson distribution with $\lambda = 3$



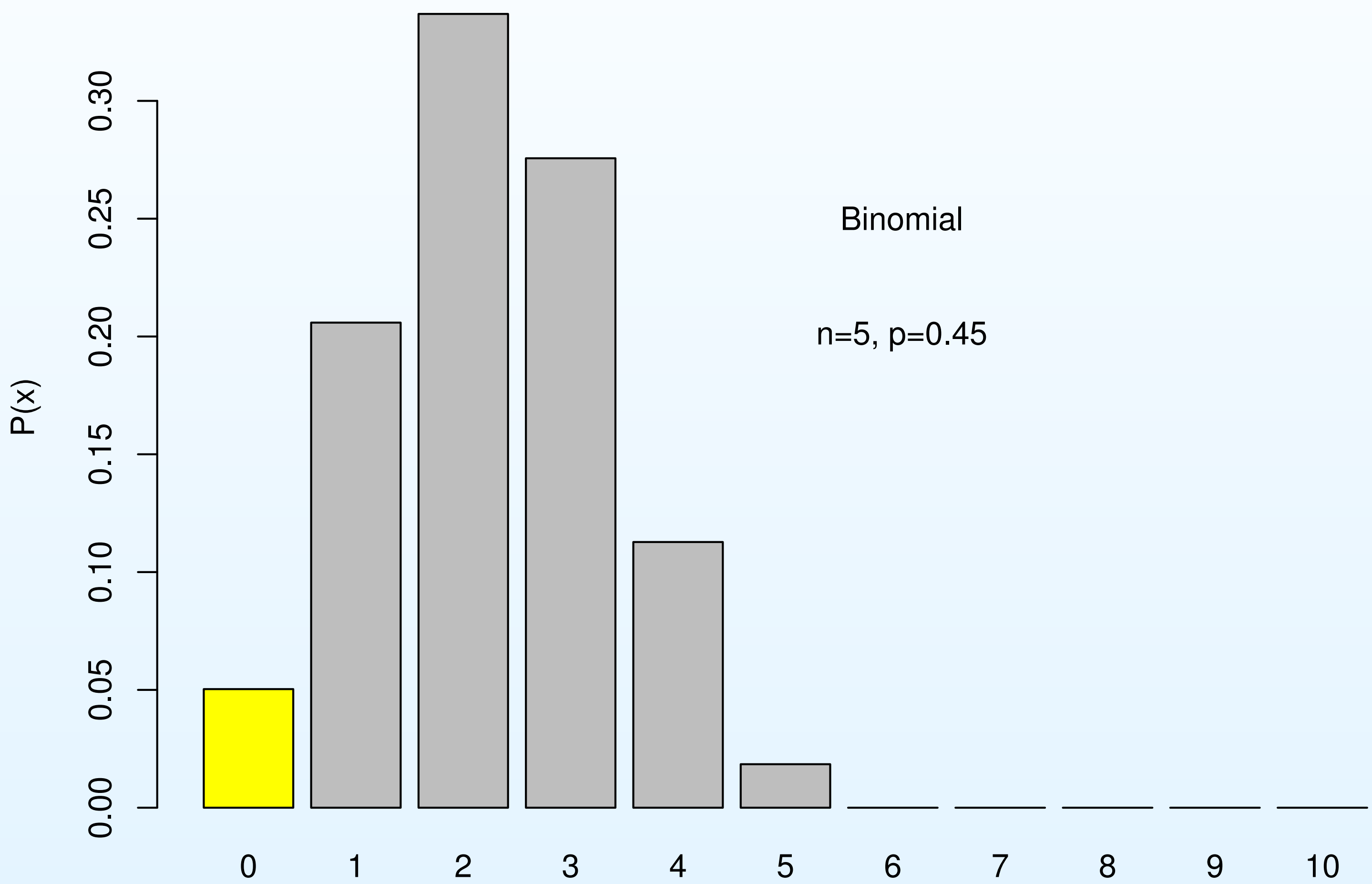
Let us check with other simple cases

A binomial distribution with $n = 10$ and $p = 0.26$



Let us check with other simple cases

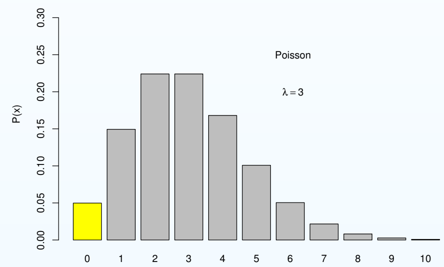
A binomial distribution with $n = 5$ and $p = 0.45$

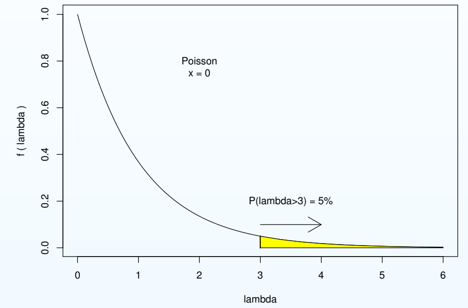


All give 5% to observe $x = 0$ $\Rightarrow$ apply probability inversion $\longrightarrow$

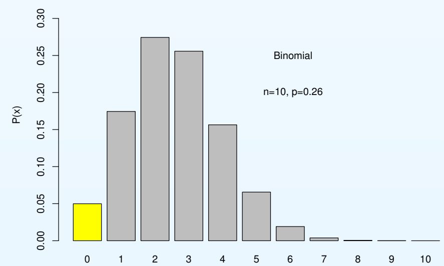
The game does not work already with the binomial

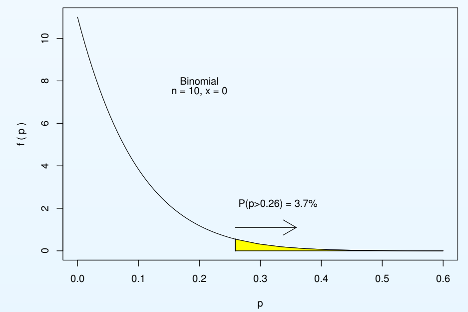
$$\text{`}\lambda_L = 3\text{'}: \; P(x=0\,|\,\lambda_L) = 5\% \qquad\qquad P(\lambda \geq \lambda_L) = 5\% \; \surd$$



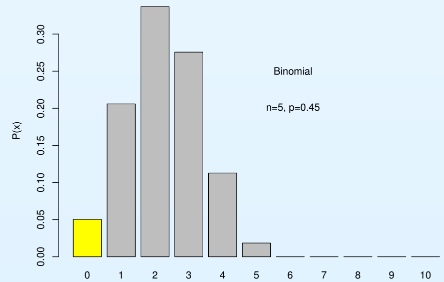


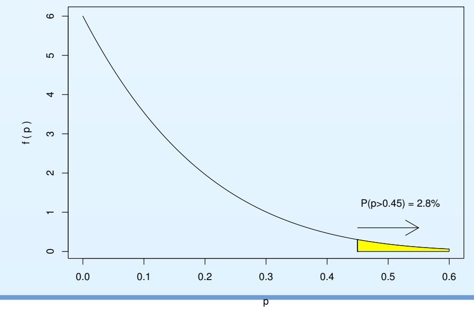
$$\text{`}p_L = 0.26\text{'}: \; P(x=0\,|\,p_L) = 5\% \qquad\qquad \text{but } P(p \geq p_L) = 3.7\%!$$



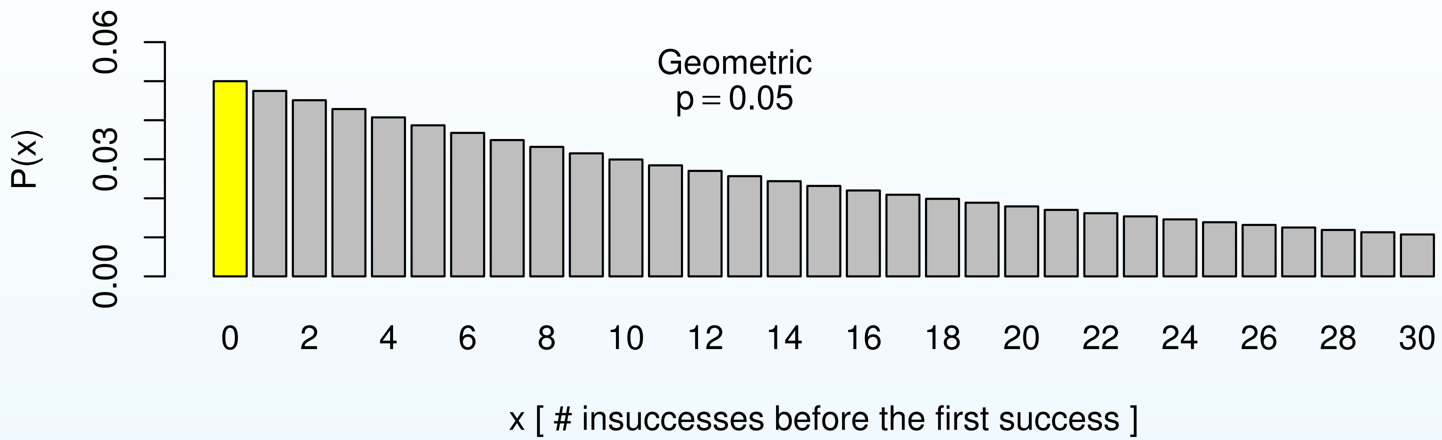


$$\text{`}p_L = 0.45\text{'}: \; P(x=0\,|\,p_L) = 5\% \qquad\qquad \text{but } P(p \geq p_L) = 2.8\%!$$

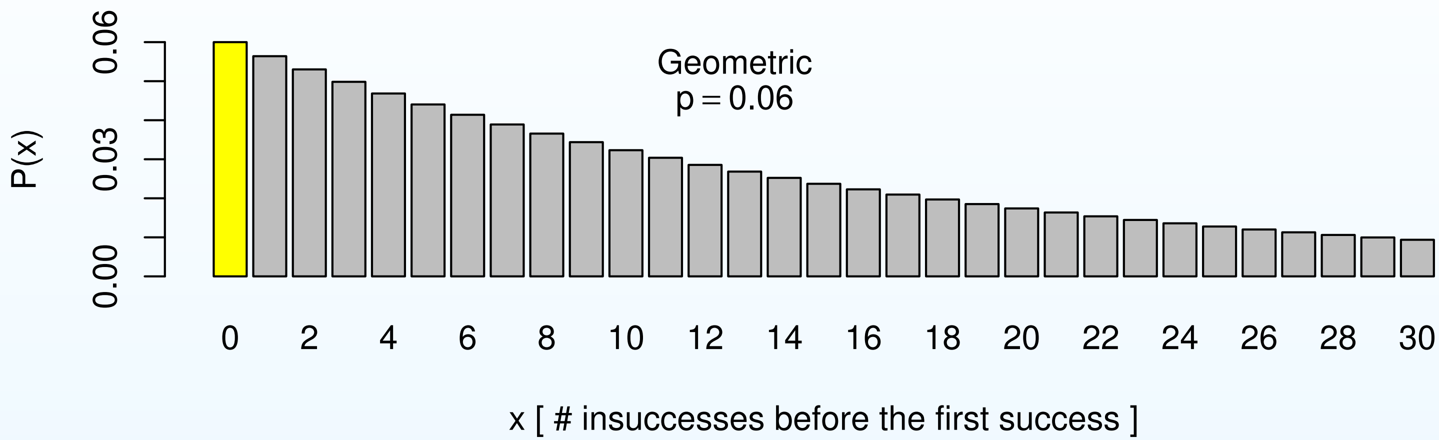




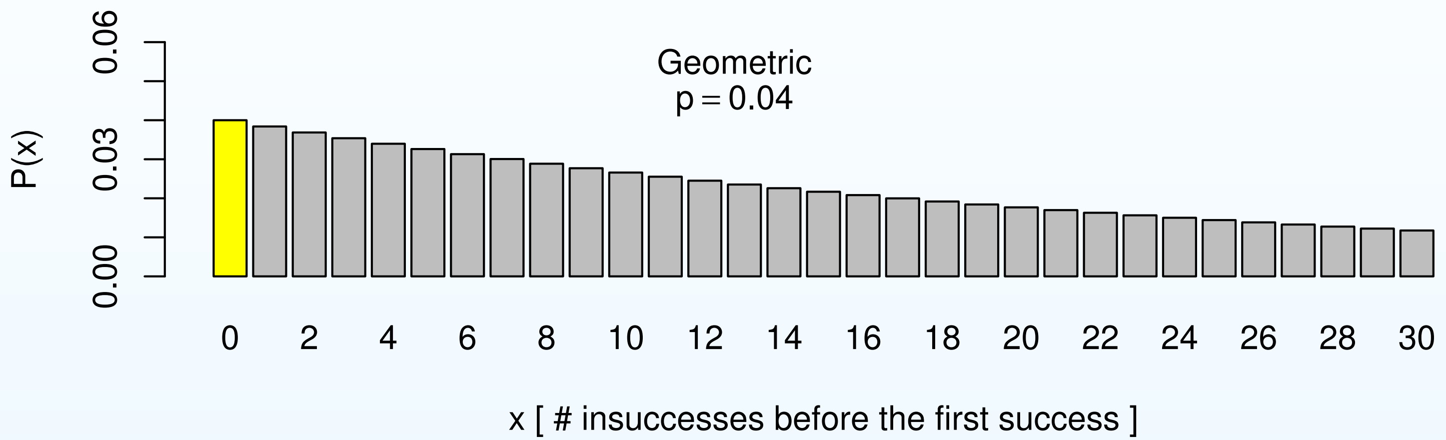
or with a geometric



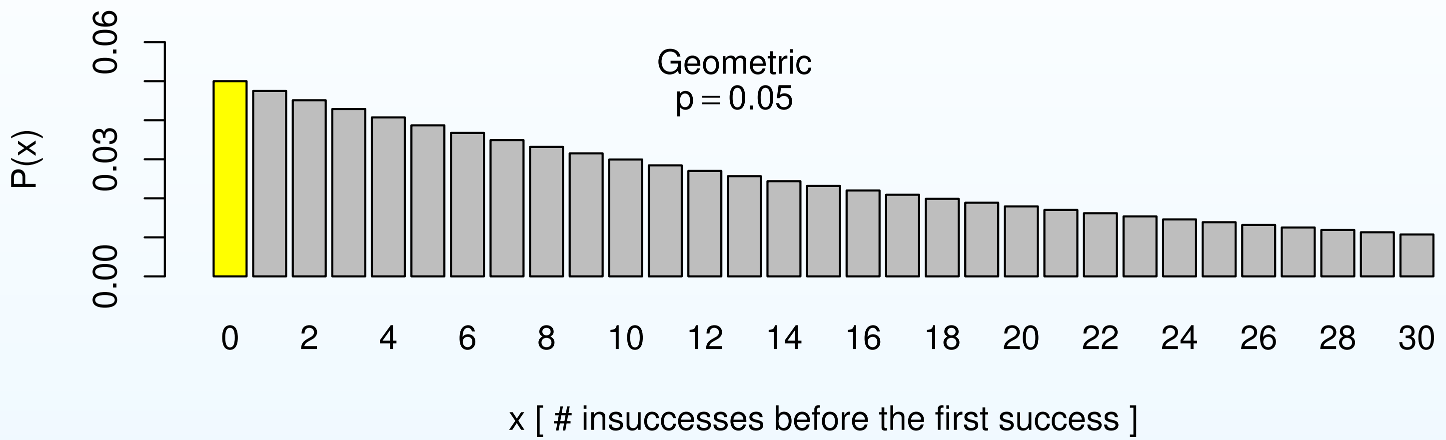
or with a geometric



or with a geometric

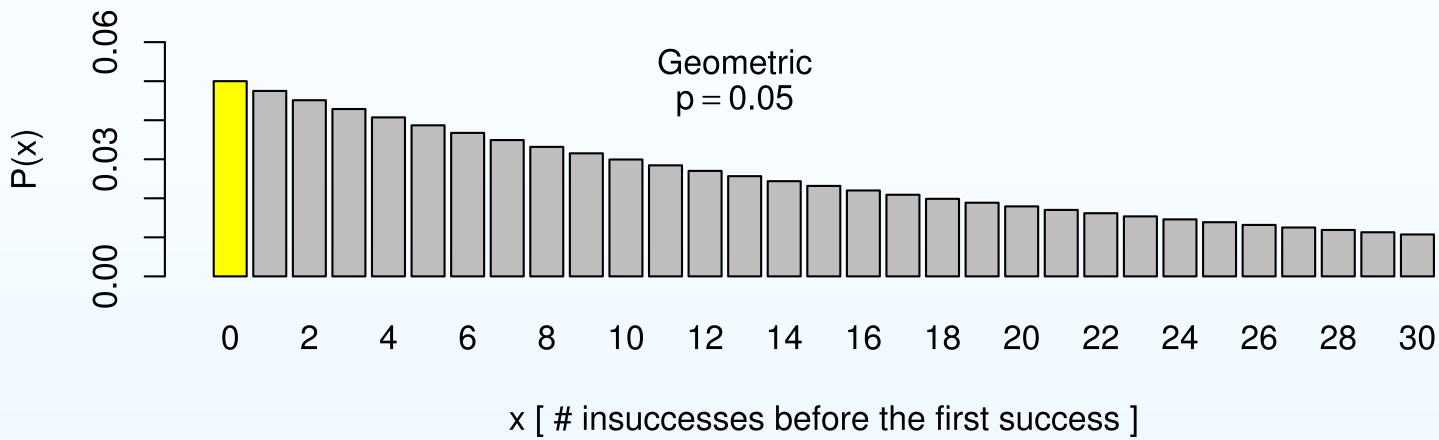


or with a geometric



Observe $x = 0$: take $p_L = 0.05$ as 'limit'? ("lower 95% C.L."?)

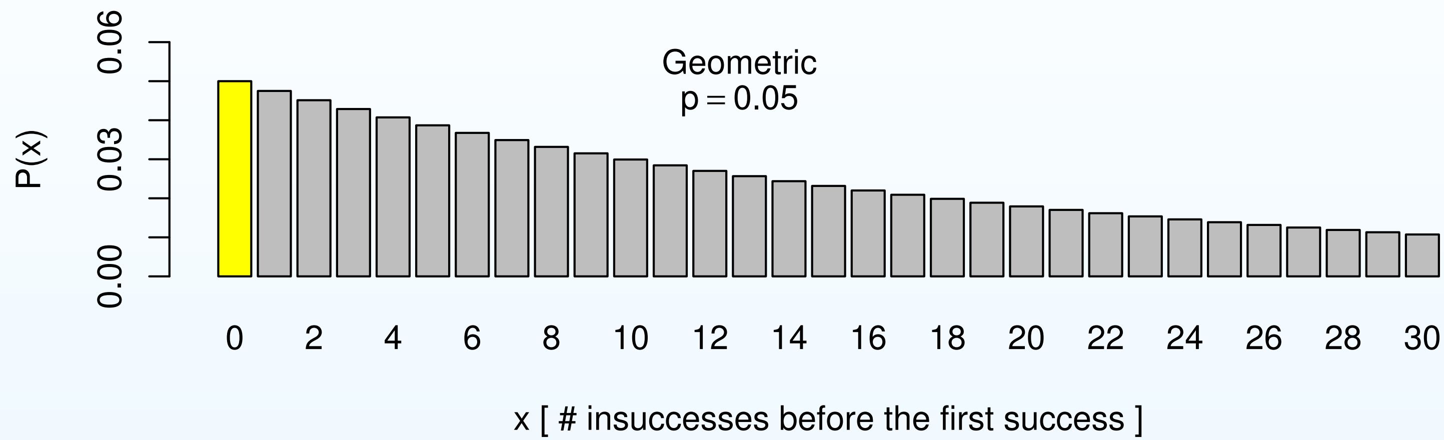
or with a geometric



Observe $x = 0$: take $p_L = 0.05$ as 'limit'? ("lower 95% C.L."?)

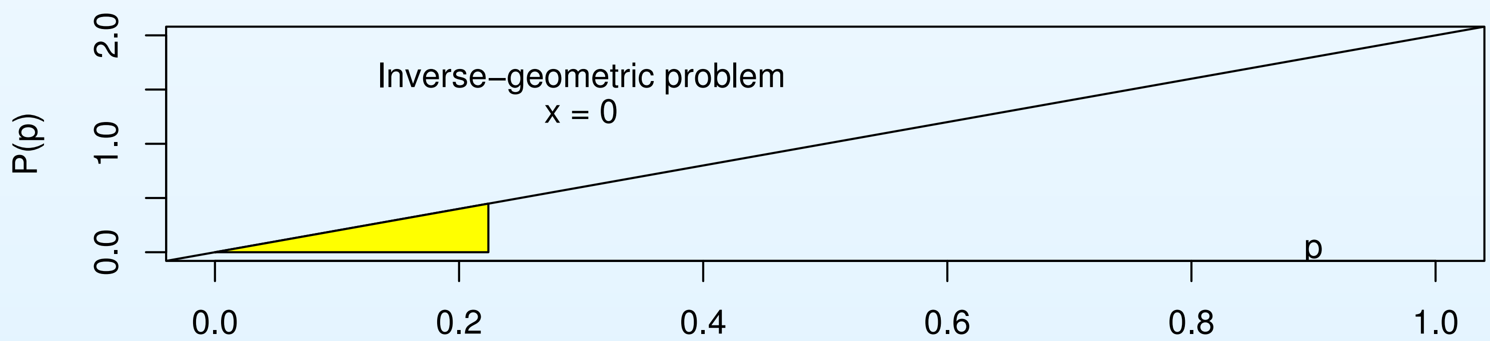
But

or with a geometric



Observe $x = 0$: take $p_L = 0.05$ as 'limit'? ("lower 95% C.L."?)

But



$P(p \leq 0.05) = 0.25\%$, while $P(p \leq 0.224) = 5\%$.

Statisticians are clever!

This is not yet the end of the story.

Statisticians are clever!

This is not yet the end of the story.
In many cases: $H_i \longrightarrow$ large number of $\{E_j\}$:

Statisticians are clever!

This is not yet the end of the story.
In many cases: $H_i \longrightarrow$ large number of $\{E_j\}$:

- Each effect has little probability $\rightarrow$ 'practically improbable'

$\Rightarrow$ whatever we observe is an evidence against the hypothesis

Statisticians are clever!

This is not yet the end of the story.
In many cases: $H_i \longrightarrow$ large number of $\{E_j\}$:

- Each effect has little probability $\rightarrow$ 'practically improbable'

$\Rightarrow$ whatever we observe is an evidence against the hypothesis

- Even those who trust the (flawed) reasoning based on the small probability of effects have to realize that the reasoning fails in these cases.

Statisticians are clever!

This is not yet the end of the story.
In many cases: $H_i \longrightarrow$ large number of $\{E_j\}$:

- Each effect has little probability $\rightarrow$ 'practically improbable'

$\Rightarrow$ whatever we observe is an evidence against the hypothesis

- Even those who trust the (flawed) reasoning based on the small probability of effects have to realize that the reasoning fails in these cases.

$\Rightarrow$ **statistician 'way out'**: individual observable effects are replaced by two *sets of effects*, one of high chance to happen, the other of low chance ('the tail(s) of the distribution')
$\rightarrow$ the reasoning is extended to these two sets of effects

Statisticians are clever!

This is not yet the end of the story.
In many cases: $H_i \longrightarrow$ large number of $\{E_j\}$:

- Each effect has little probability $\rightarrow$ 'practically improbable'

$\Rightarrow$ whatever we observe is an evidence against the hypothesis

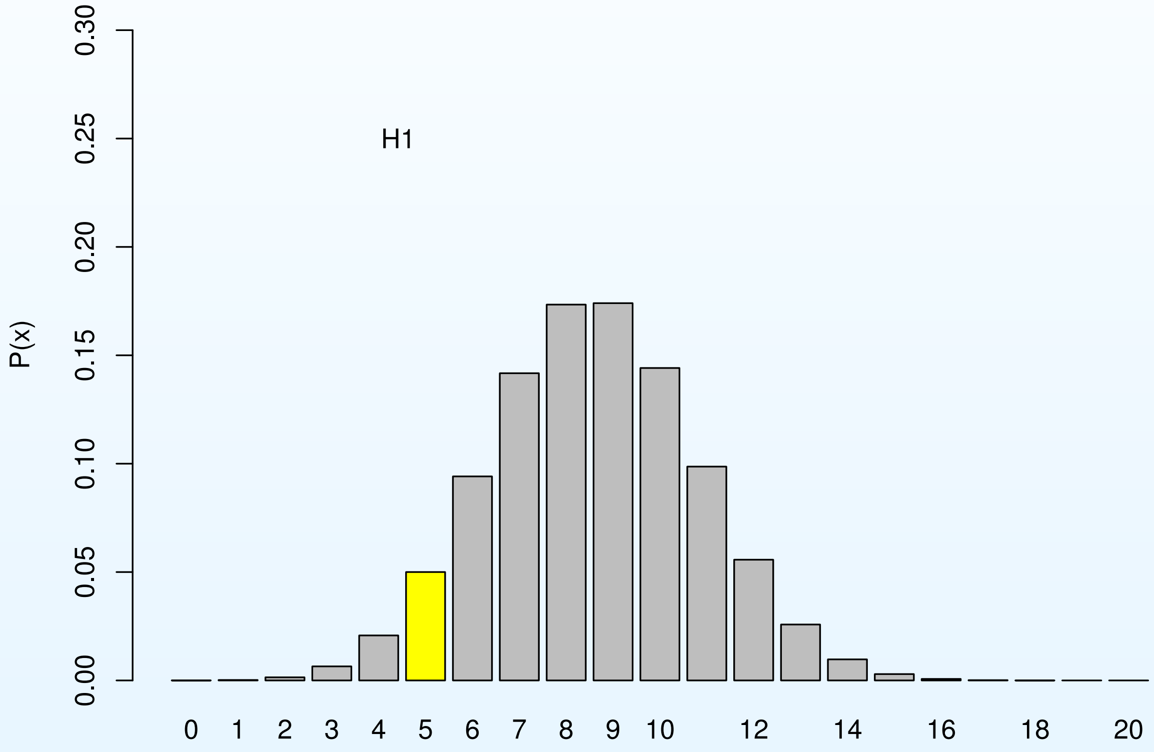
- Even those who trust the (flawed) reasoning based on the small probability of effects have to realize that the reasoning fails in these cases.

$\Rightarrow$ **statistician 'way out'**: individual observable effects are replaced by two *sets of effects*, one of high chance to happen, the other of low chance ('the tail(s) of the distribution')
$\rightarrow$ the reasoning is extended to these two sets of effects

$\Rightarrow$ Logically, the situation worsens:
$\rightarrow$ conclusions depend not only on on observed effects, but also on non-observed effects!

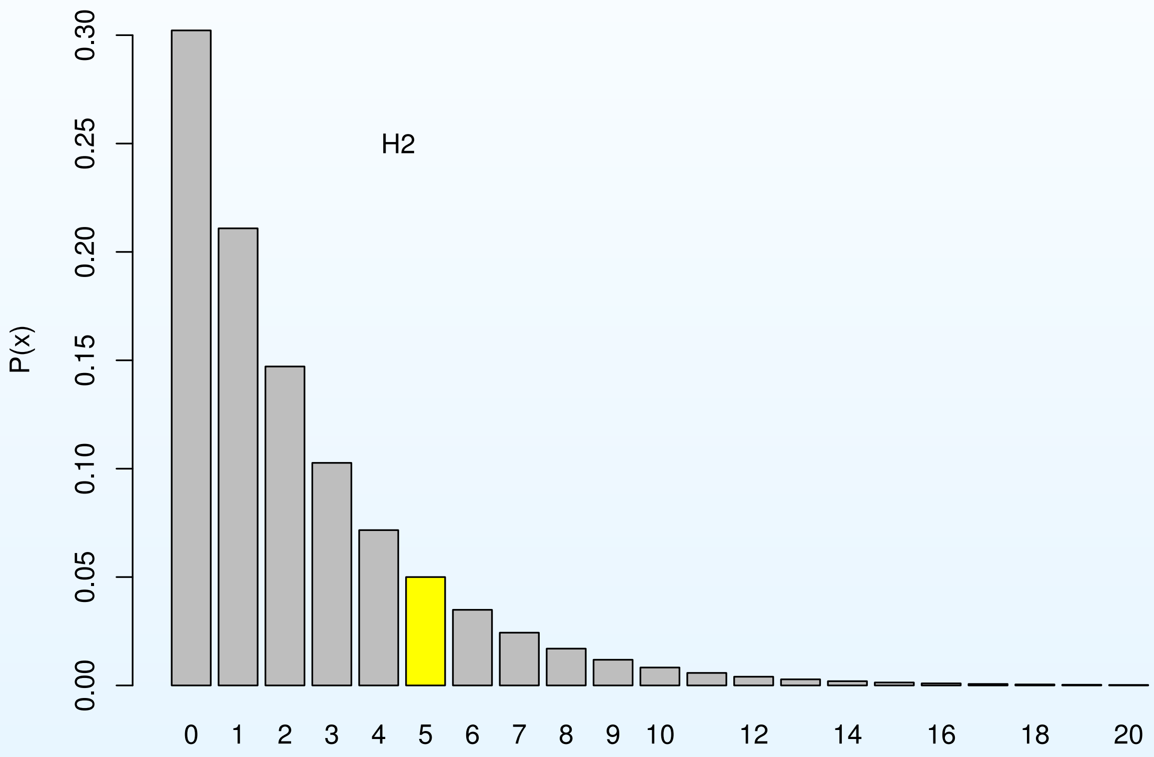
Observed value and tails

Several hypotheses to be tested against the observation $x = 5$



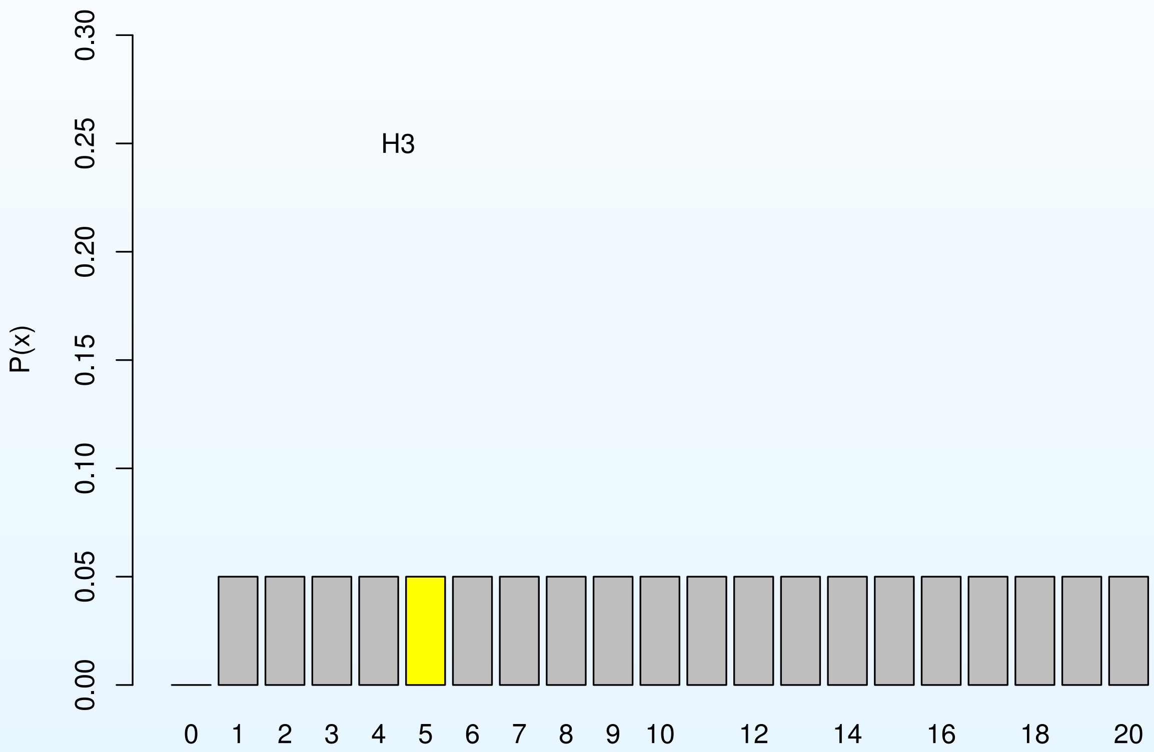
Observed value and tails

Several hypotheses to be tested against the observation $x = 5$



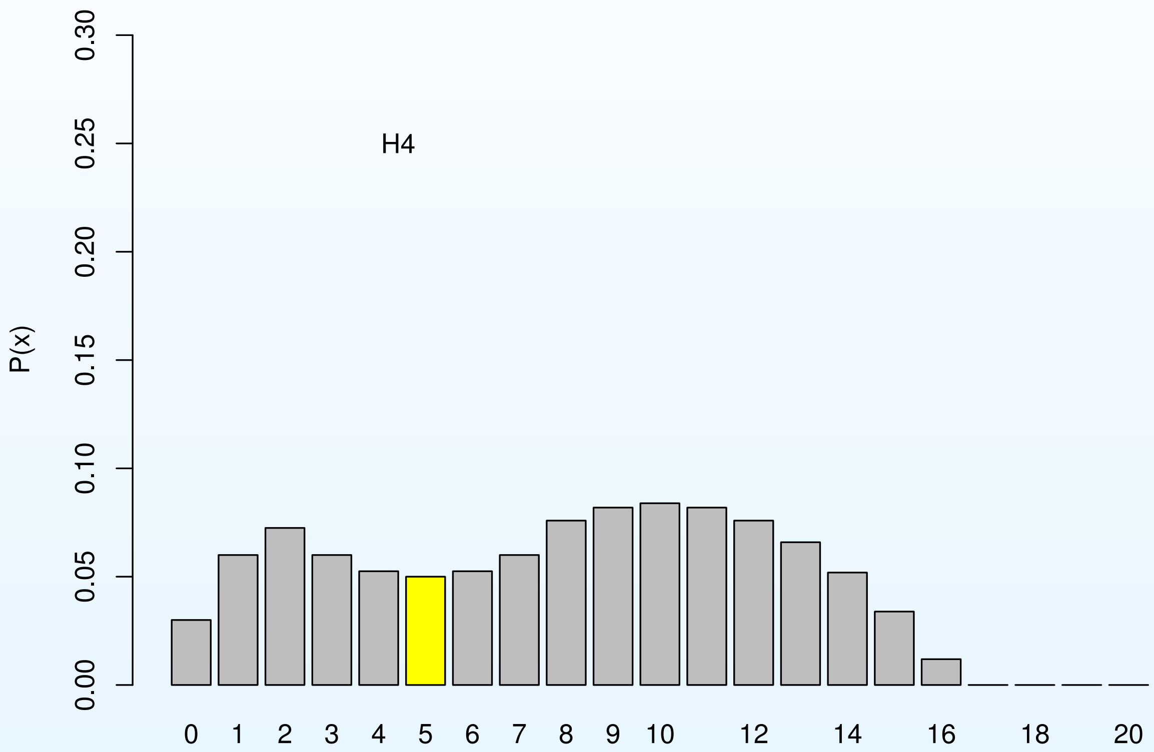
Observed value and tails

Several hypotheses to be tested against the observation $x = 5$



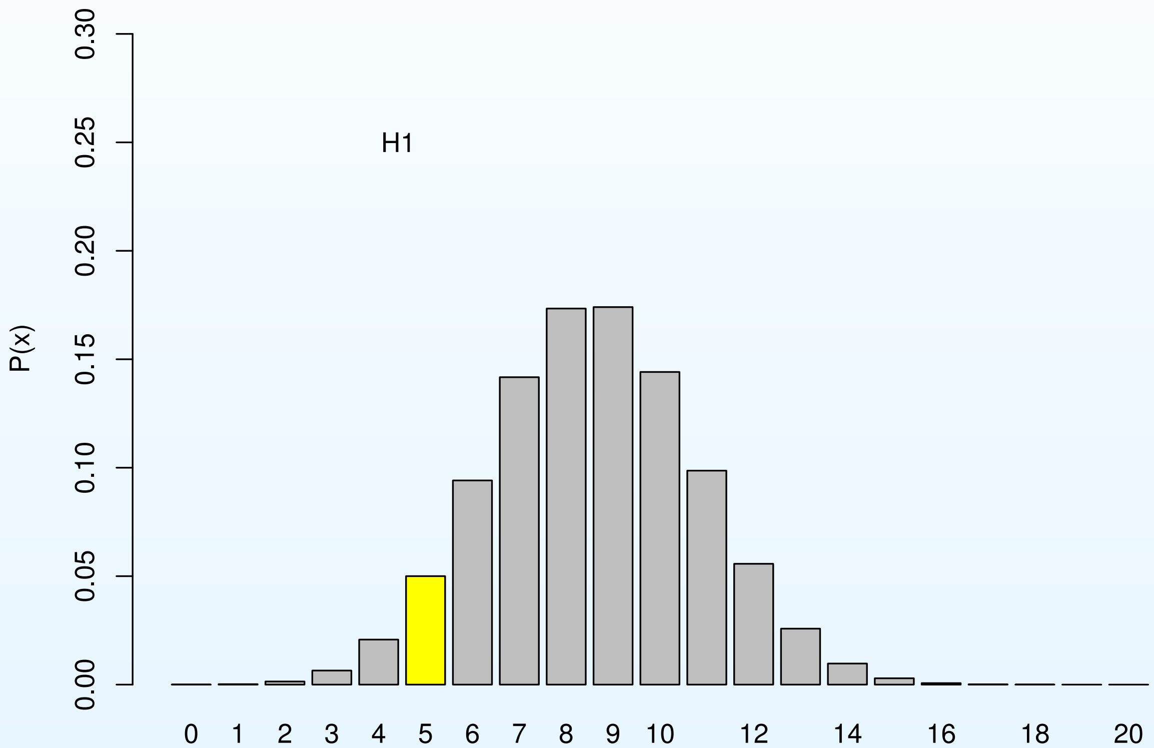
Observed value and tails

Several hypotheses to be tested against the observation $x = 5$



Observed value and tails

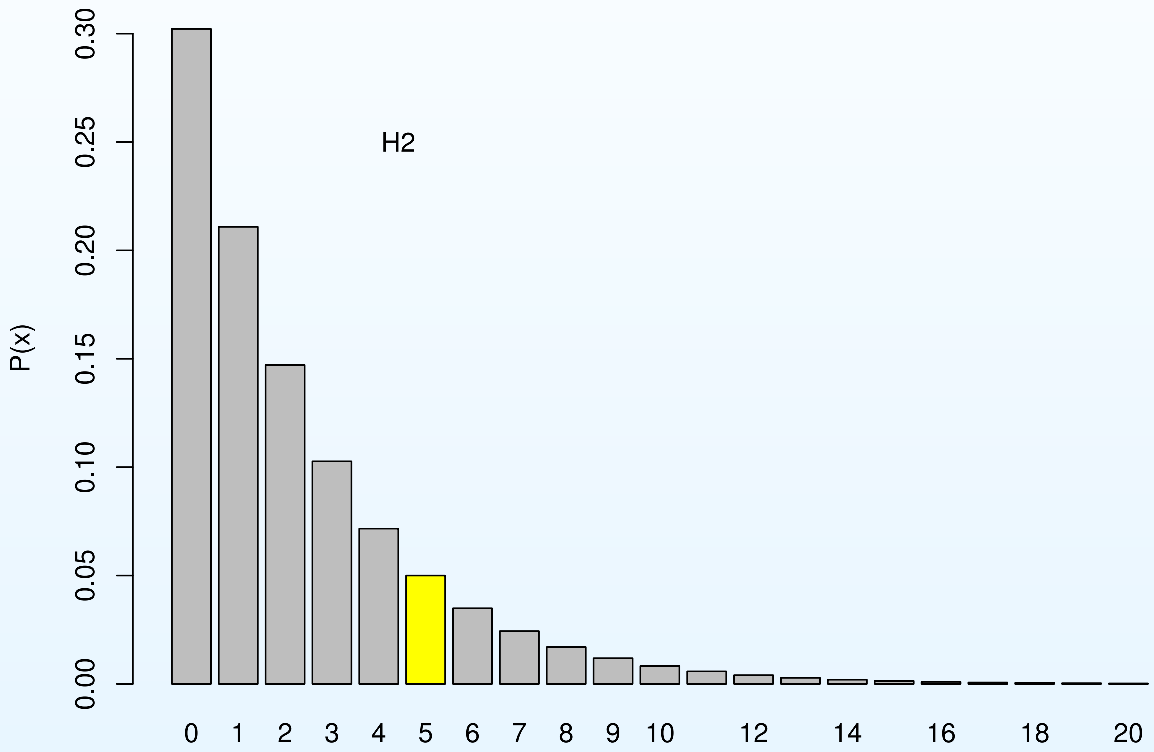
All have the same probability to give $x = 5$



Natural that our conclusions depend ‘somehow’ on $P(x = 5 \,|\, H_i)$

Observed value and tails

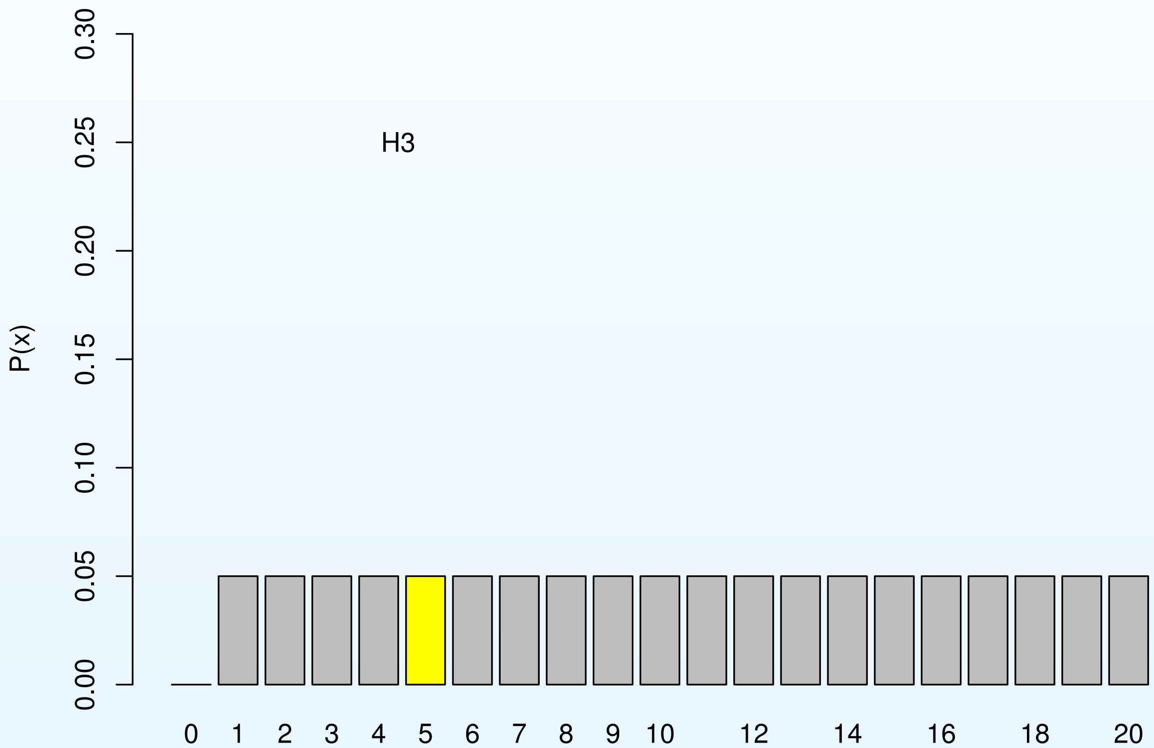
All have the same probability to give $x = 5$



Natural that our conclusions depend 'somehow' on $P(x = 5 \,|\, H_i)$

Observed value and tails

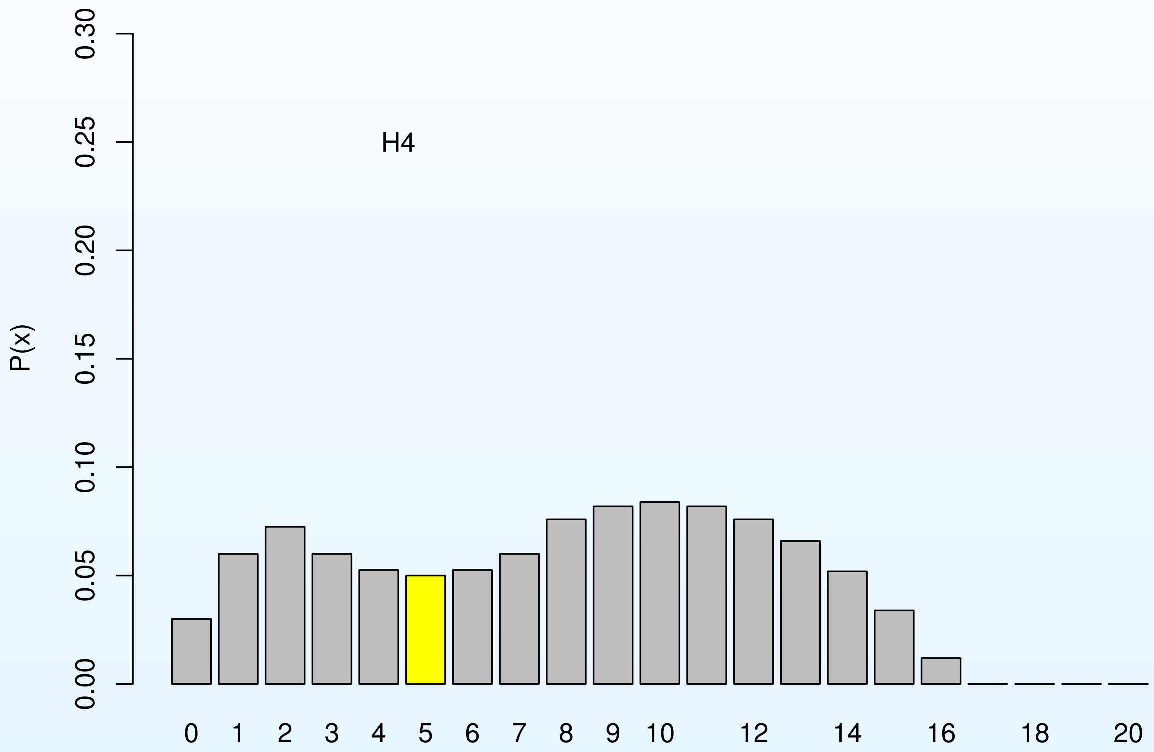
All have the same probability to give $x = 5$



Natural that our conclusions depend ‘somehow’ on $P(x = 5 \mid H_i)$

Observed value and tails

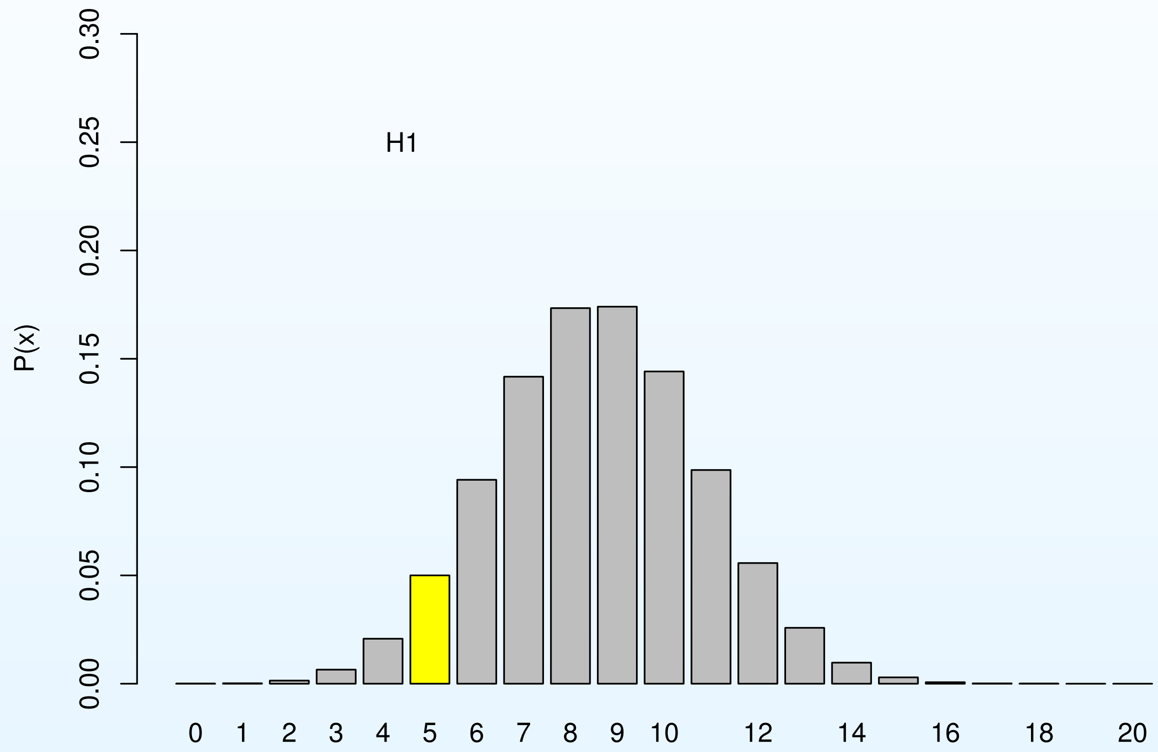
All have the same probability to give $x = 5$



Natural that our conclusions depend ‘somehow’ on $P(x = 5 \,|\, H_i)$

Observed value and tails

All have the same probability to give $x = 5$

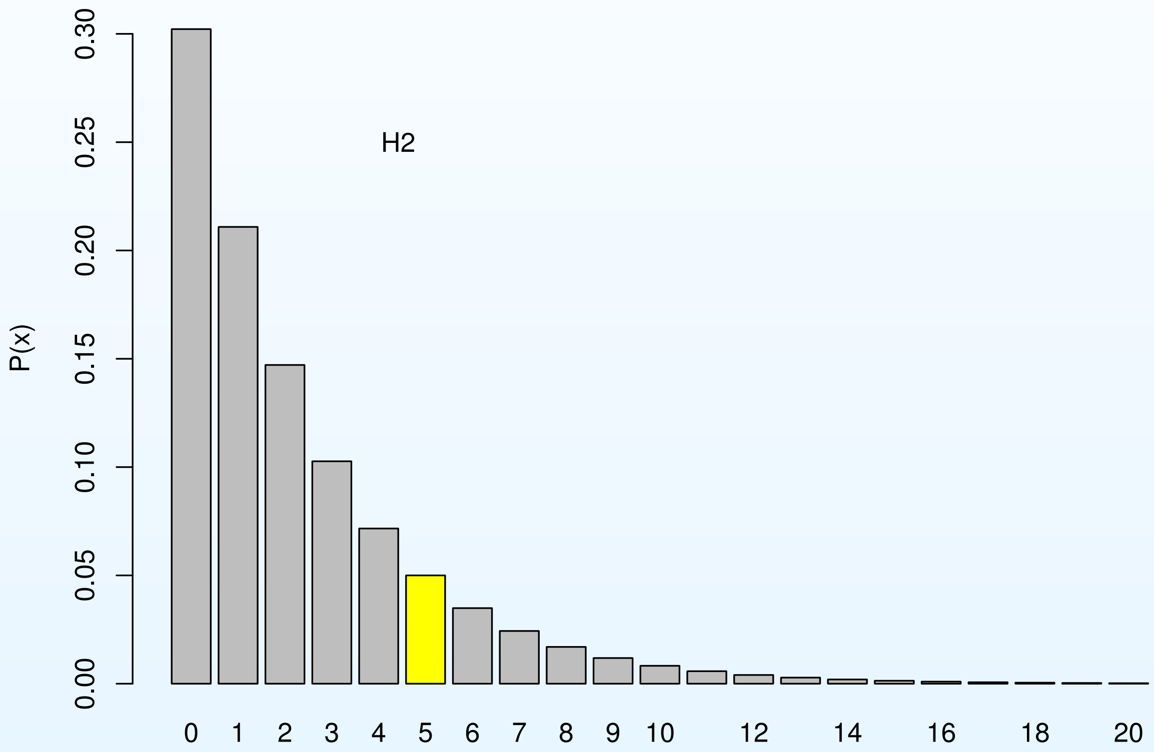


Natural that our conclusions depend ‘somehow’ on $P(x = 5 \mid H_i)$

(and in this case we learn nothing!)

Observed value and tails

All have the same probability to give $x = 5$

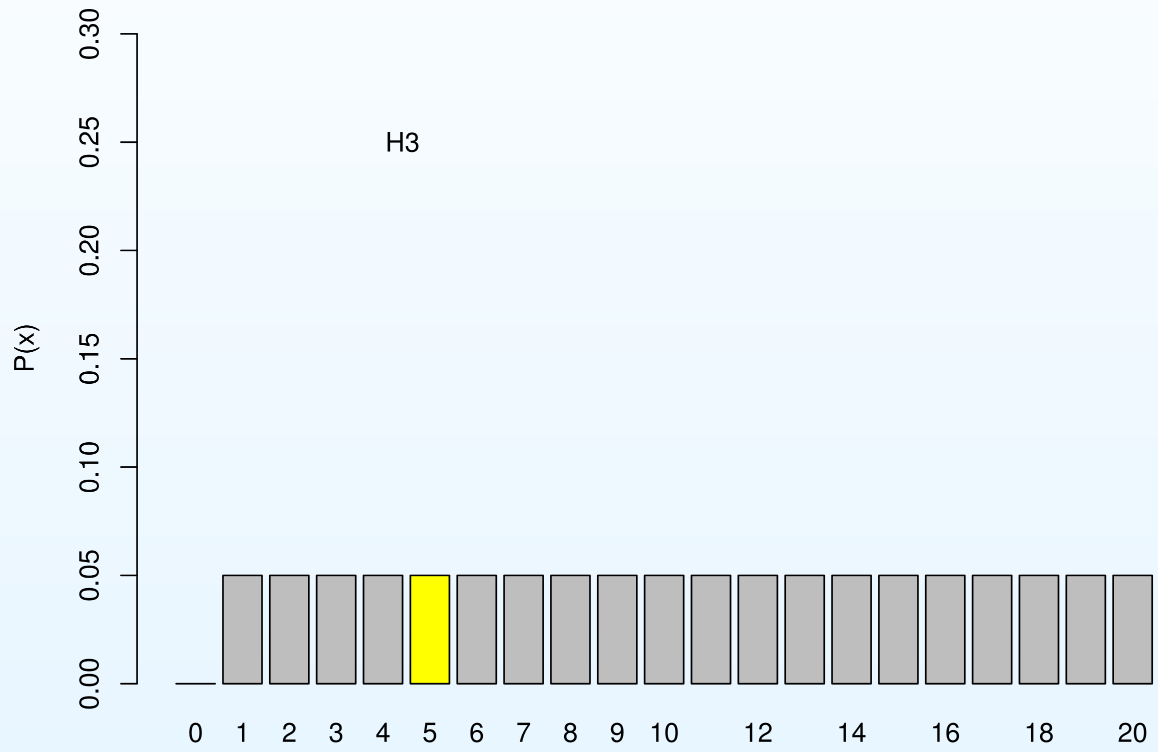


Natural that our conclusions depend ‘somehow’ on $P(x = 5 \,|\, H_i)$

(and in this case we learn nothing!)

Observed value and tails

All have the same probability to give $x = 5$

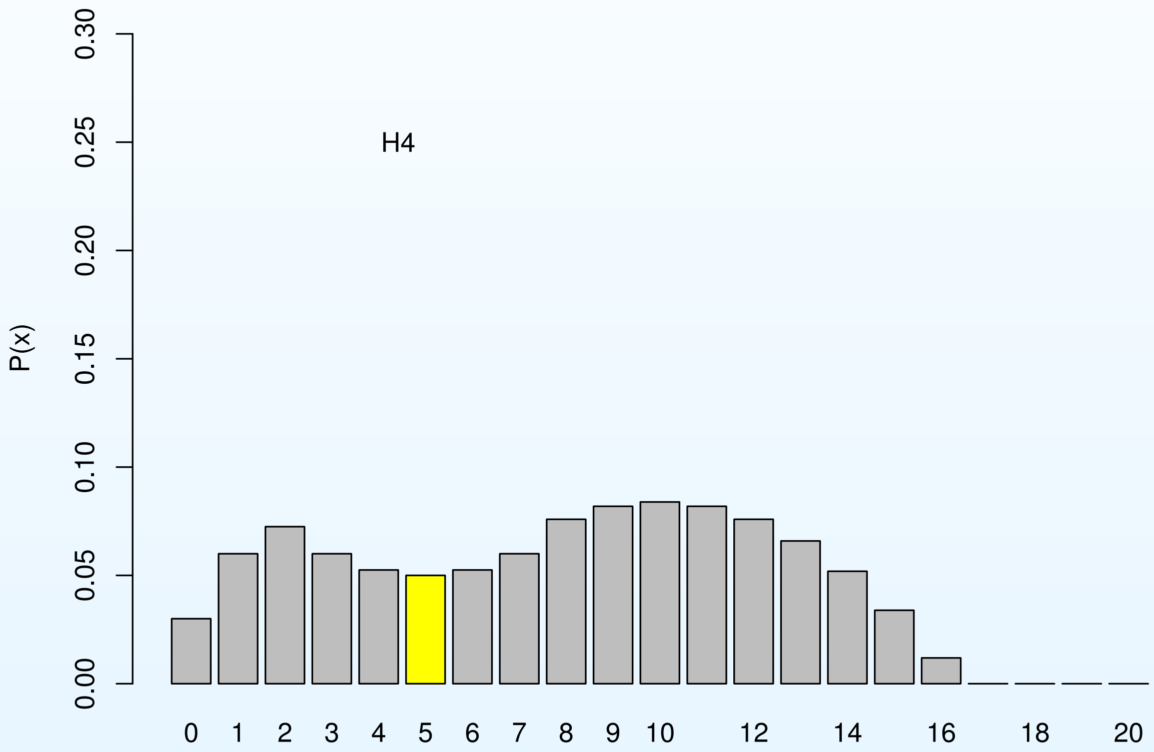


Natural that our conclusions depend ‘somehow’ on $P(x = 5 \,|\, H_i)$

(and in this case we learn nothing!)

Observed value and tails

All have the same probability to give $x = 5$

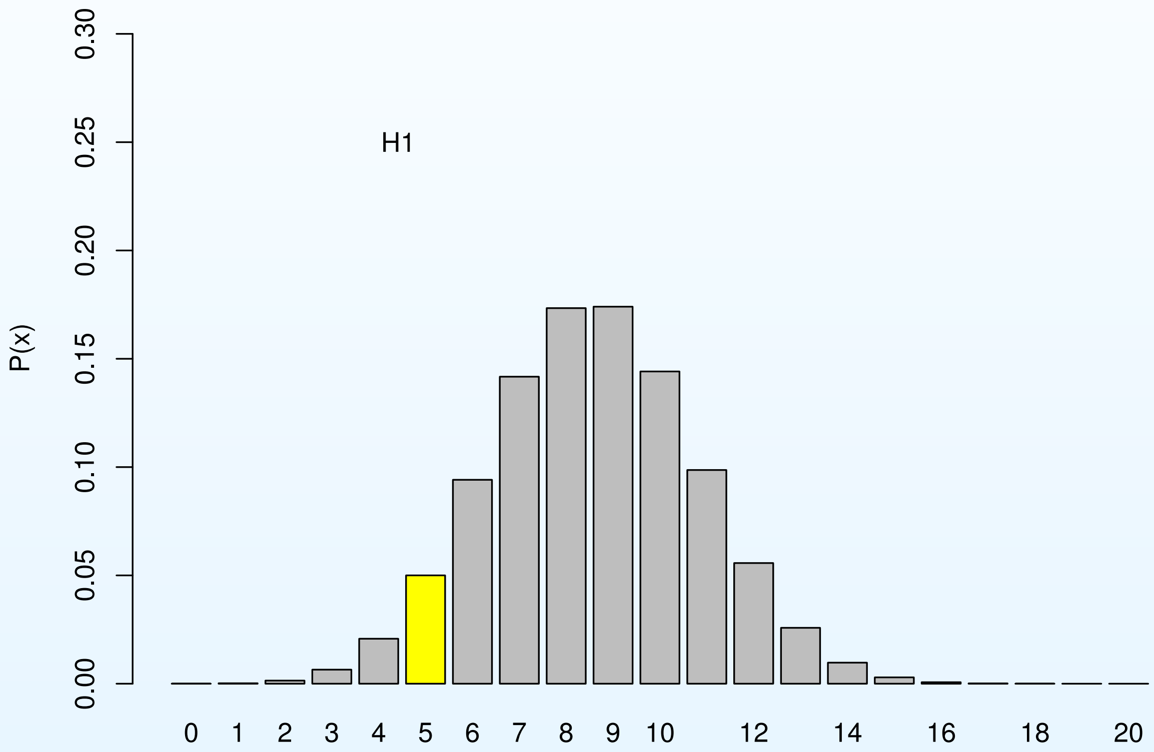


Natural that our conclusions depend ‘somehow’ on $P(x = 5 \,|\, H_i)$

(and in this case we learn nothing!)

Observed value and tails

All have the same probability to give $x = 5$

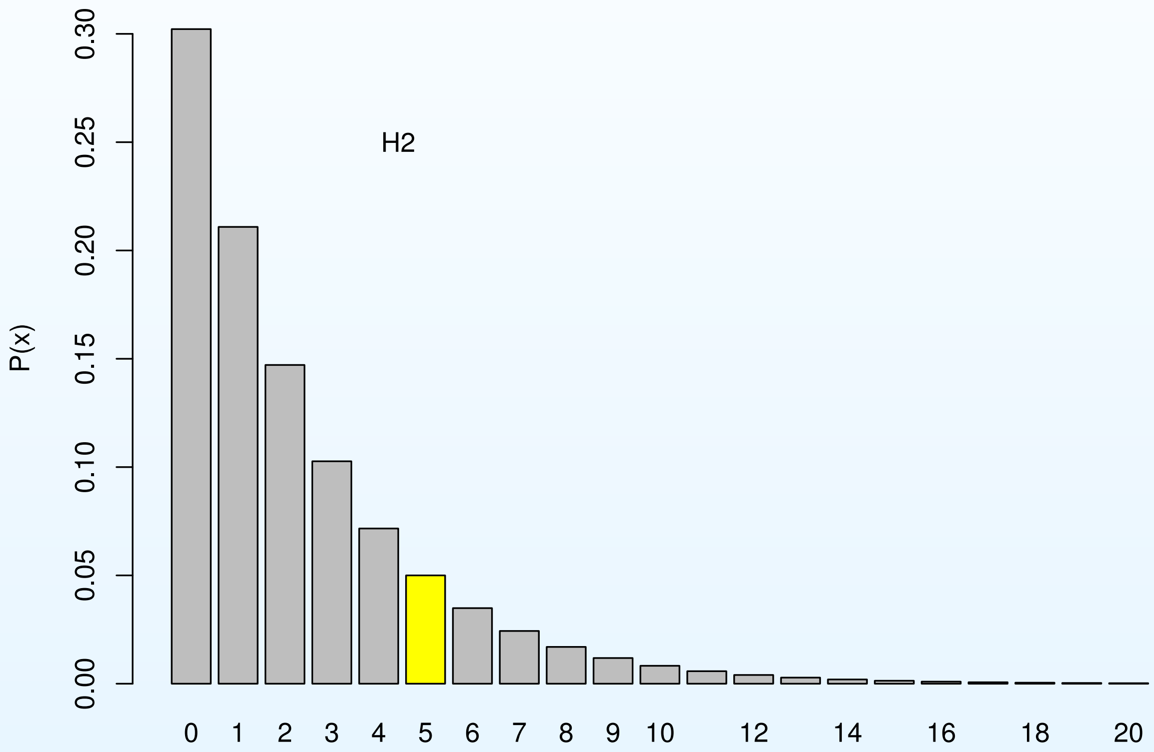


Natural that our conclusions depend 'somehow' on $P(x = 5 \,|\, H_i)$

but why also on $P(x \neq 5 \,|\, H_i)$?

Observed value and tails

All have the same probability to give $x = 5$

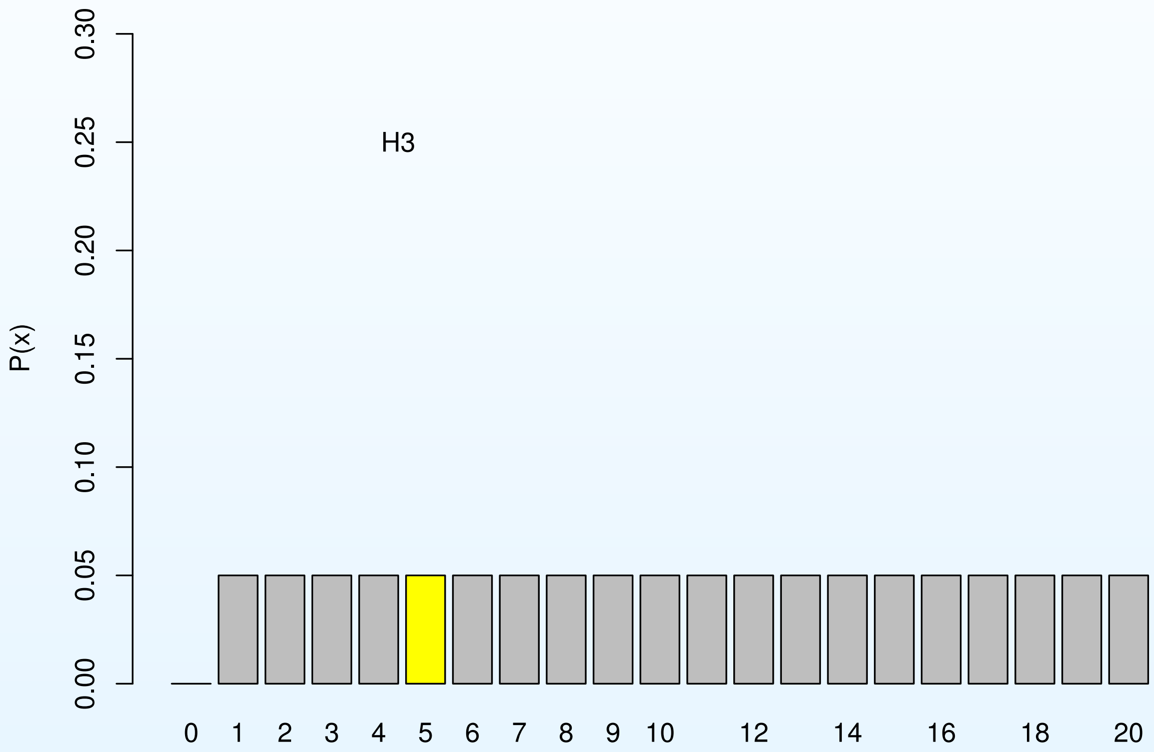


Natural that our conclusions depend ‘somehow’ on $P(x = 5 \mid H_i)$

but why also on $P(x \neq 5 \mid H_i)$?

Observed value and tails

All have the same probability to give $x = 5$

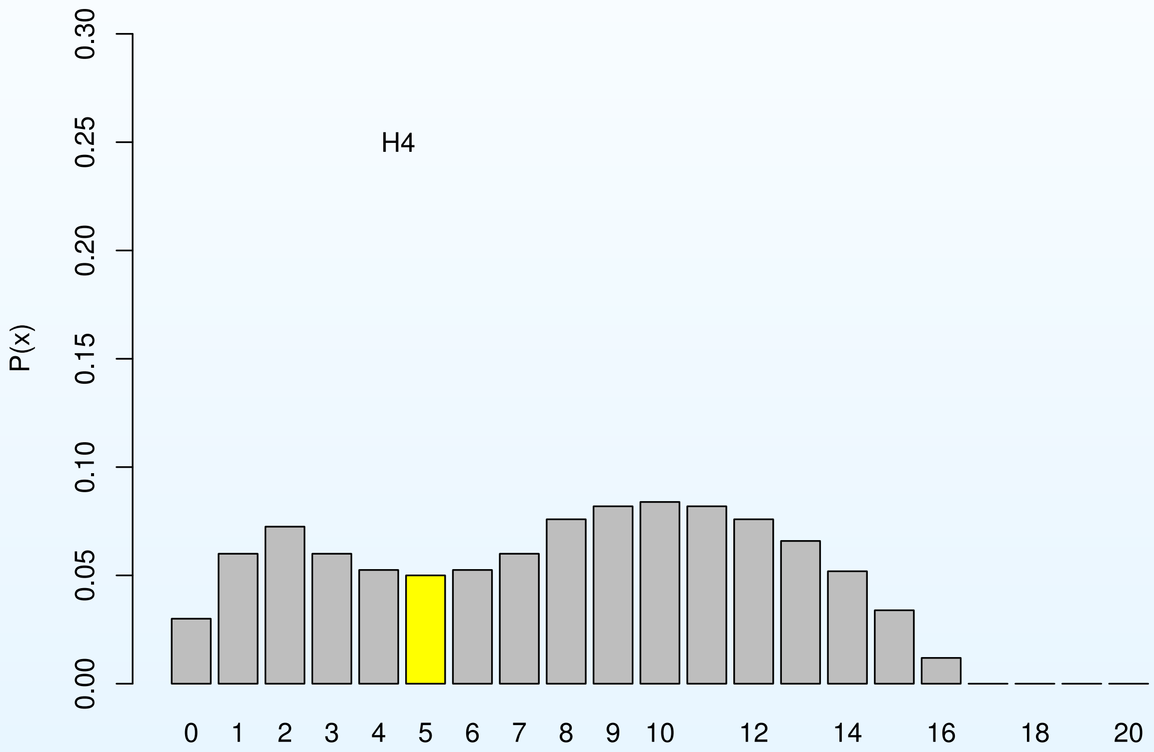


Natural that our conclusions depend ‘somehow’ on $P(x = 5 \,|\, H_i)$

but why also on $P(x \neq 5 \,|\, H_i)$?

Observed value and tails

All have the same probability to give $x = 5$



Natural that our conclusions depend ‘somehow’ on $P(x = 5 \,|\, H_i)$

but why also on $P(x \neq 5 \,|\, H_i)$?

P-values

But this is what we do when we draw scientific conclusions based on the probability of *'what we have really observed, or something rarer than that'*,

P-values

But this is what we do when we draw scientific conclusions based on the probability of *'what we have really observed, or something rarer than that'*,

what statisticians call p-values

(But physicists are more used with 'χ^2 probabilities', or something similar).

P-values

But this is what we do when we draw scientific conclusions based on the probability of *'what we have really observed, or something rarer than that'*,

> what statisticians call p-values
>
> (But physicists are more used with 'χ^2 probabilities', or something similar).

$\Rightarrow$ Nothing to do with the interpretation that we cannot use Monte Carlo 'data' (in the sense of non observed data)

$\rightarrow$ see Holy Inquisition style question $\longrightarrow$ Slide

P-values

But this is what we do when we draw scientific conclusions based on the probability of *'what we have really observed, or something rarer than that'*,

what statisticians call p-values

(But physicists are more used with 'χ^2 probabilities', or something similar).

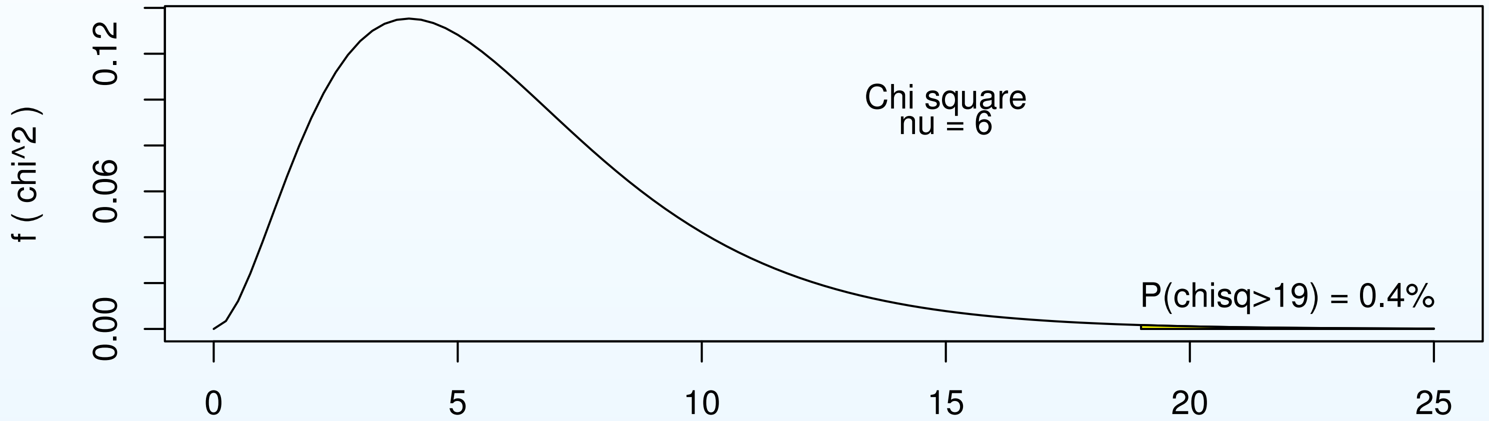
$\Rightarrow$ Nothing to do with the interpretation that we cannot use Monte Carlo 'data' (in the sense of non observed data)

$\longrightarrow$ see Holy Inquisition style question $\longrightarrow$ Slide

- Indeed, MC's are a summary of all our beliefs!
- What is the meaning of the probabilities we put in and take out from MC's? (and attach later to physics processes?)

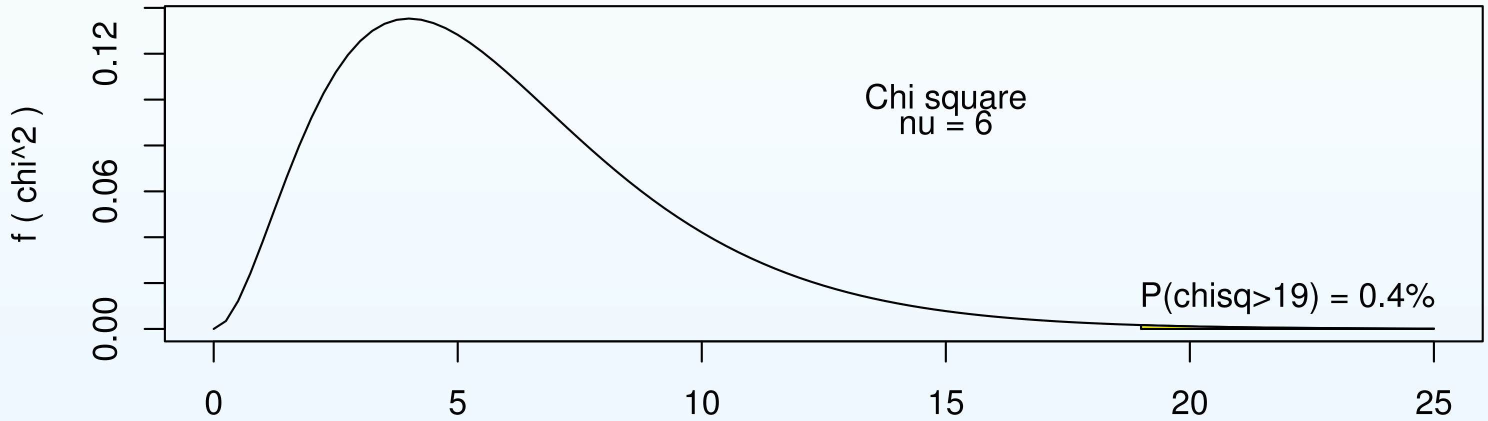
P-values

Ex.: χ^2, $\nu = 6$, $\chi^2_{obs} = 19$: $\rightarrow$ p-value $= \int_{\chi^2_{obs}}^{\infty} f(\chi^2_6)\, \mathrm{d}\chi^2_6$ = 0.4%.



P-values

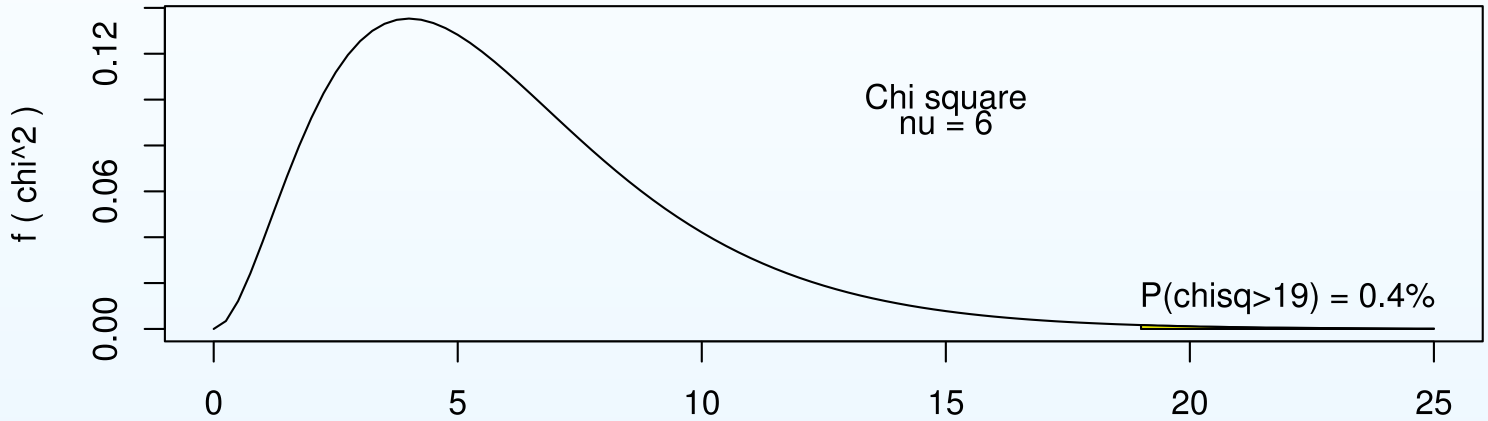
Ex.: χ^2, $\nu = 6$, $\chi^2_{obs} = 19$: $\longrightarrow$ p-value $= \int_{\chi^2_{obs}}^{\infty} f(\chi^2_6)\, d\chi^2_6$ = 0.4%.



What do you conclude?

P-values

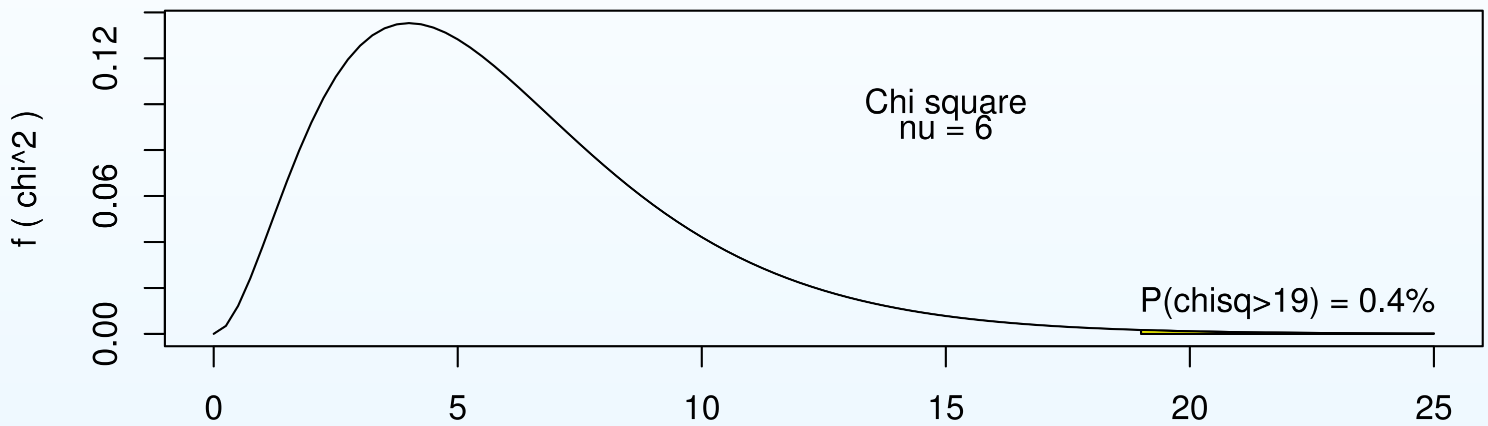
Ex.: χ^2, $\nu = 6$, $\chi^2_{obs} = 19$: $\;\longrightarrow$ p-value $= \int_{\chi^2_{obs}}^{\infty} f(\chi^2_6)\, \mathrm{d}\chi^2_6$ = 0.4%.



What do you conclude? (We shall come back later on this point) Note for the moment:

P-values

Ex.: χ^2, $\nu = 6$, $\chi^2_{obs} = 19$: $\longrightarrow$ p-value $= \int_{\chi^2_{obs}}^{\infty} f(\chi^2_6)\, \mathrm{d}\chi^2_6$ = 0.4%.

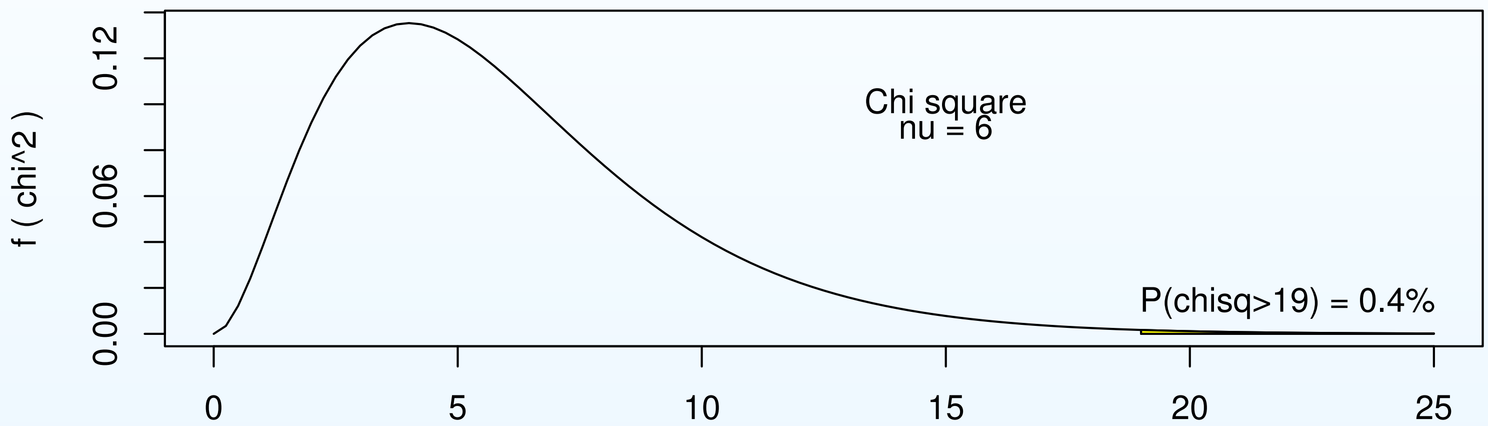


What do you conclude? (We shall come back later on this point) Note for the moment: Whatever your conclusion is, based on this information, be aware:

- It does not depend directly on the observed data, but on the 'statistical summary' χ^2.

P-values

Ex.: χ^2, $\nu = 6$, $\chi^2_{obs} = 19$: $\rightarrow$ p-value $= \int_{\chi^2_{obs}}^{\infty} f(\chi^2_6)\,\mathrm{d}\chi^2_6$ = 0.4%.



What do you conclude? (We shall come back later on this point) Note for the moment: Whatever your conclusion is, based on this information, be aware:

- It does not depend directly on the observed data, but on the 'statistical summary' χ^2.
- Indeed, it does not even depend precisely on the 'observed summary' alone (χ^2_{obs}), but on all other values of the summary that are less likely than the observed one.

P-values

Rationale?

P-values

Rationale?

As most of these kind of prescriptions, they are not based on solid principles but only on authority and use.

P-values

Rationale?

As most of these kind of prescriptions, they are not based on solid principles but only on authority and use.

But then it must work, otherwise it should have been realized!

P-values

Rationale?

As most of these kind of prescriptions, they are not based on solid principles but only on authority and use.

But then it must work, otherwise it should have been realized!

- Yes!

Rationale?

As most of these kind of prescriptions, they are not based on solid principles but only on authority and use.

But then it must work, otherwise it should have been realized!

- Yes! 'It does often work',
  but this has little to do with the 'probability of the tail', as we shall see later.

Example: Has the student made a mistake?

Homework: calculate the average of 300 random numbers, uniformly distributed between 0 and 1.

Example: Has the student made a mistake?

Homework: calculate the average of 300 random numbers, uniformly distributed between 0 and 1.

- Teacher expectation:

$$\mathrm{E}\left[\overline{X}_{300}\right] = \frac{1}{2}$$

$$\sigma\left[\overline{X}_{300}\right] = \frac{1}{\sqrt{12}} \cdot \frac{1}{\sqrt{300}} = 0.017\,,$$

Example: Has the student made a mistake?

Homework: calculate the average of 300 random numbers, uniformly distributed between 0 and 1.

- Teacher expectation:

$$\begin{aligned} \mathsf{E}\left[\overline{X}_{300}\right] &= \frac{1}{2} \\ \sigma\left[\overline{X}_{300}\right] &= \frac{1}{\sqrt{12}} \cdot \frac{1}{\sqrt{300}} = 0.017\,, \end{aligned}$$

- 99% probability interval

$$P(0.456 \le \overline{X}_{300} \le 0.544) = 99\%\,.$$

Example: Has the student made a mistake?

Homework: calculate the average of 300 random numbers, uniformly distributed between 0 and 1.

- Teacher expectation:

$$\mathrm{E}\left[\overline{X}_{300}\right] = \frac{1}{2}$$

$$\sigma\left[\overline{X}_{300}\right] = \frac{1}{\sqrt{12}} \cdot \frac{1}{\sqrt{300}} = 0.017\,,$$

- 99% probability interval

$$P(0.456 \le \overline{X}_{300} \le 0.544) = 99\%\,.$$

- Student gets a value outside the interval, e.g. $\overline{x} = 0.550$.

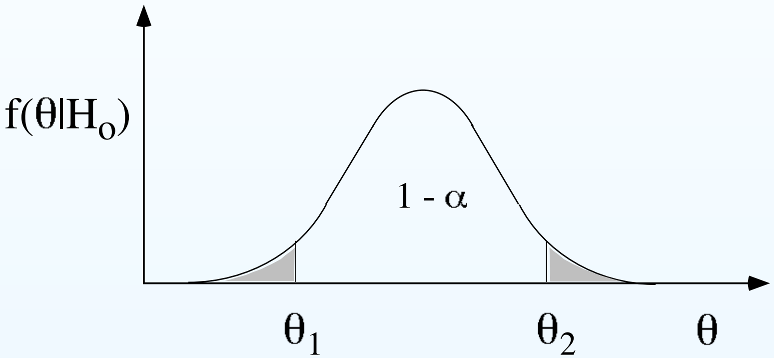
$\Rightarrow$ Has the student made a mistake?

Example: Has the student made a mistake?

Conventional statistician solution:
$\Rightarrow$ test the hypothesis $H_0 =$ 'no mistakes'

Example: Has the student made a mistake?

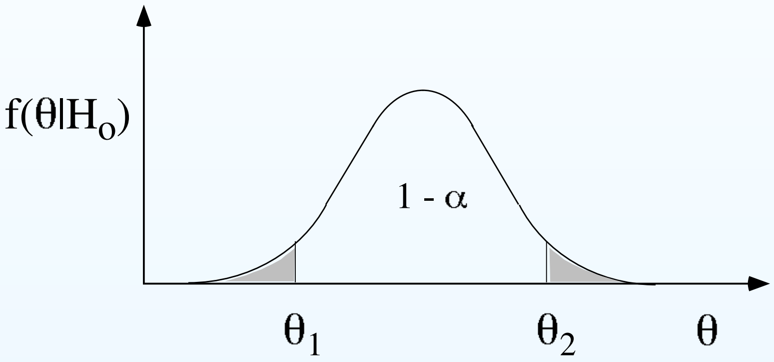
Conventional statistician solution:
$\Rightarrow$ test the hypothesis $H_0 =$ 'no mistakes'



- Test variable θ is $\overline{X}_{300}$.
- Acceptance interval $[\theta_1, \theta_2]$ is $[0.456, 0.544]$.
  We are 99% confident that $\overline{X}_{300}$ will fall inside it:
  $\rightarrow \alpha = 1\%$.

Example: Has the student made a mistake?

Conventional statistician solution:
$\Rightarrow$ test the hypothesis $H_0 =$ 'no mistakes'

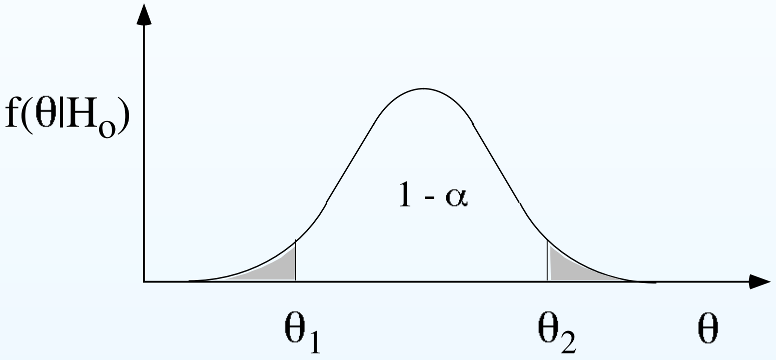


- Test variable θ is $\overline{X}_{300}$.
- Acceptance interval $[\theta_1, \theta_2]$ is $[0.456, 0.544]$.
  We are 99% confident that $\overline{X}_{300}$ will fall inside it:
  $\rightarrow \alpha = 1\%$.
- $\overline{x} = 0.550$ lies outside the acceptance interval

$\Rightarrow$ Hypothesis H_0 is rejected at 1% significance.

Example: Has the student made a mistake?

Conventional statistician solution:
$\Rightarrow$ test the hypothesis $H_0 =$ 'no mistakes'



- Test variable θ is $\overline{X}_{300}$.
- Acceptance interval $[\theta_1, \theta_2]$ is $[0.456, 0.544]$.
  We are 99% confident that $\overline{X}_{300}$ will fall inside it:
  $\rightarrow \alpha = 1\%$.
- $\overline{x} = 0.550$ lies outside the acceptance interval

$\Rightarrow$ Hypothesis H_0 is rejected at 1% significance.

$\Rightarrow$ What does it mean?

Meaning of the hypothesis test

Conclusion from test:

"the hypothesis $H_\circ$ *= 'no mistakes' is rejected at the 1% level of significance".*

Meaning of the hypothesis test

Conclusion from test:

"the hypothesis $H_\circ$ = 'no mistakes' is rejected at the 1% level of significance".

What does it mean?

"there is only a 1% probability that the average falls outside the selected interval, if the calculations were done correctly".

Meaning of the hypothesis test

Conclusion from test:

> *"the hypothesis $H_\circ$ = 'no mistakes' is rejected at the 1% level of significance".*

What does it mean?

> *"there is only a 1% probability that the average falls outside the selected interval, if the calculations were done correctly".*

So what?

Meaning of the hypothesis test

Conclusion from test:

> *"the hypothesis $H_\circ$ = 'no mistakes' is rejected at the 1% level of significance".*

What does it mean?

> *"there is only a 1% probability that the average falls outside the selected interval, if the calculations were done correctly".*

So what?

- It does not reply our natural question, i.e. that concerning the probability of mistake – quite impolite, by the way.
- The statement sounds as if one would be 99% sure that the student has made a mistake! (Mostly interpreted in this way).

$\Rightarrow$ Highly misleading!

Something is missing in the reasoning

If you ask the students (before they take a standard course in hypothesis tests) you will realize of a crucial ingredient extraneous to the logic of hypothesis tests:

Something is missing in the reasoning

If you ask the students (before they take a standard course in hypothesis tests) you will realize of a crucial ingredient extraneous to the logic of hypothesis tests:

“It all depends on whom has made the calculation!”

Something is missing in the reasoning

If you ask the students (before they take a standard course in hypothesis tests) you will realize of a crucial ingredient extraneous to the logic of hypothesis tests:

> *"It all depends on whom has made the calculation!"*

In fact, if the calculation was done by a well-tested program, the probability of mistake would be zero.
And students know rather well their tendency to do or not mistakes.

‘Something is missing’: another example

The value $x = 3.01$ is extracted from a Gaussian random number generator having $\mu = 0$ and $\sigma = 1$.

‘Something is missing’: another example

The value $x = 3.01$ is extracted from a Gaussian random number generator having $\mu = 0$ and $\sigma = 1$.
It is well known that $P(|X| > 3) = 0.27\%$, <u>but</u>

‘Something is missing’: another example

The value $x = 3.01$ is extracted from a Gaussian random number generator having $\mu = 0$ and $\sigma = 1$.
It is well known that $P(|X| > 3) = 0.27\%$, but

we cannot say

- “the value X has 0.27% probability of coming from that generator”

‘Something is missing’: another example

The value $x = 3.01$ is extracted from a Gaussian random number generator having $\mu = 0$ and $\sigma = 1$.
It is well known that $P(|X| > 3) = 0.27\%$, <u>but</u>

we cannot say

- “the value X has 0.27% probability of coming from that generator”
- “the probability that the observation is a statistical fluctuation is 0.27%”

'Something is missing': another example

The value $x = 3.01$ is extracted from a Gaussian random number generator having $\mu = 0$ and $\sigma = 1$.
It is well known that $P(|X| > 3) = 0.27\%$, but

we cannot say

- "the value X has 0.27% probability of coming from that generator"
- "the probability that the observation is a statistical fluctuation is 0.27%"

$\Rightarrow$ the value comes with 100% probability from that generator!

‘Something is missing’: another example

The value $x = 3.01$ is extracted from a Gaussian random number generator having $\mu = 0$ and $\sigma = 1$.
It is well known that $P(|X| > 3) = 0.27\%$, but

we cannot say

- “the value X has 0.27% probability of coming from that generator”
- “the probability that the observation is a statistical fluctuation is 0.27%”

$\Rightarrow$ the value comes with 100% probability from that generator!

$\Rightarrow$ it is at 100% a statistical fluctuation

‘Something is missing’: another example

The value $x = 3.01$ is extracted from a Gaussian random number generator having $\mu = 0$ and $\sigma = 1$.
It is well known that $P(|X| > 3) = 0.27\%$, <u>but</u>

we cannot say

- “the value X has 0.27% probability of coming from that generator”
- “the probability that the observation is a statistical fluctuation is 0.27%”

$\Rightarrow$ the value comes with 100% probability from that generator!

$\Rightarrow$ it is at 100% a statistical fluctuation

Logical bug of the reasoning:

$\Rightarrow$ One cannot tell how much one is confident in generator A only if another generator B is not taken into account.

'Something is missing': another example

The value $x = 3.01$ is extracted from a Gaussian random number generator having $\mu = 0$ and $\sigma = 1$.
It is well known that $P(|X| > 3) = 0.27\%$, but

we cannot say

- "the value X has 0.27% probability of coming from that generator"
- "the probability that the observation is a statistical fluctuation is 0.27%"

$\Rightarrow$ the value comes with 100% probability from that generator!

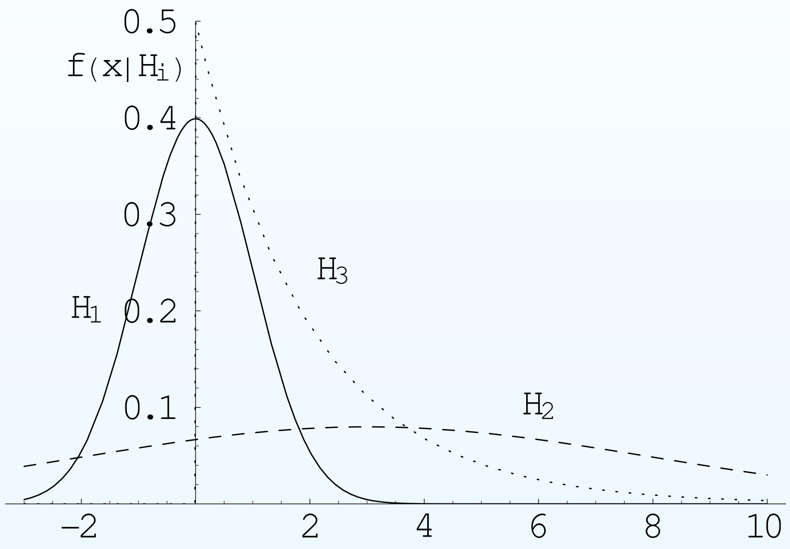
$\Rightarrow$ it is at 100% a statistical fluctuation

Logical bug of the reasoning:

$\Rightarrow$ This is the original sin of conventional hypothesis test methods

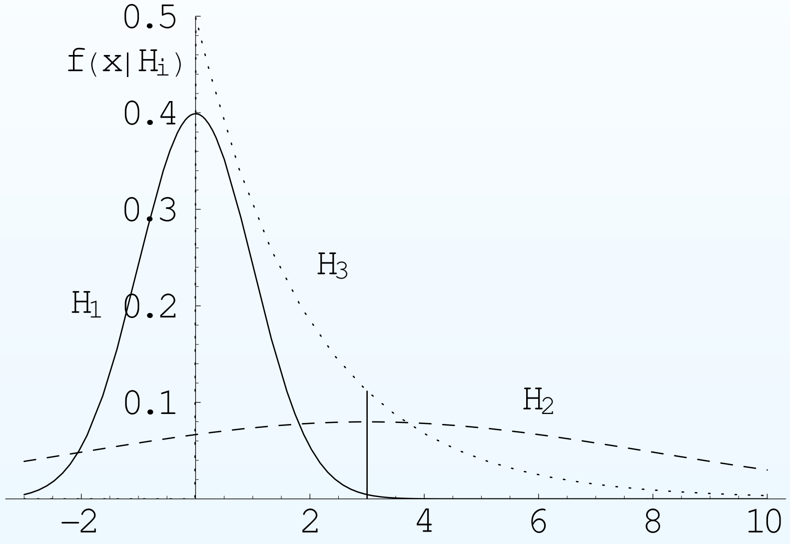
Well posed problem

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



Well posed problem

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

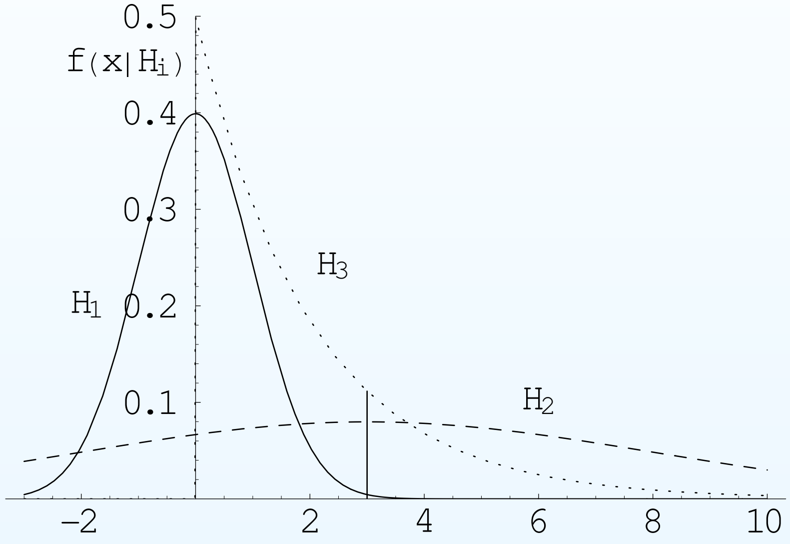


The statistics-uneducated student would suggest:

- our preference should depend on how likely each model might yield $x = 3$

Well posed problem

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

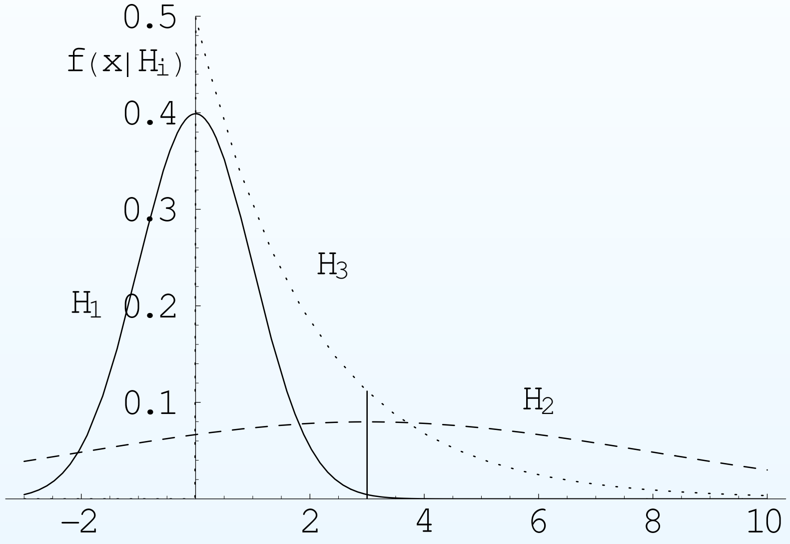


The statistics-uneducated student would suggest:

- our preference should depend on how likely each model might yield $x = 3$
- ... but perhaps also on 'how reasonable' each model is, given the physical situation under study

Well posed problem

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



The statistics-uneducated student would suggest:

- our preference should depend on how likely each model might yield $x = 3$
- . . . but perhaps also on 'how reasonable' each model is, given the physical situation under study

$\Rightarrow$ Right!

Objections

“These are chosen academic examples.”

Objections

"These are chosen academic examples."

$\Rightarrow$ logic is logic!

Objections

"These are chosen academic examples."

- $\Rightarrow$ logic is logic!

  How can we use a reasoning in frontier physics
  if it fails in simple cases?

- $\Rightarrow$ All fake claims of discoveries are due to
  the criticized reasoning (examples in a while $\longrightarrow$)

Objections

“These are chosen academic examples.”

$\Rightarrow$ logic is logic!

How can we use a reasoning in frontier physics
if it fails in simple cases?

$\Rightarrow$ All fake claims of discoveries are due to
the criticized reasoning (examples in a while $\longrightarrow$)

“Hypotheses tests are well proved to work”

Objections

"These are chosen academic examples."

$\Rightarrow$ logic is logic!

How can we use a reasoning in frontier physics
if it fails in simple cases?

$\Rightarrow$ All fake claims of discoveries are due to
the criticized reasoning (examples in a while $\longrightarrow$)

"Hypotheses tests are well proved to work"

Yes and not. . .

$\Rightarrow$ They 'often work' due to reasons external to their logic, but which are not always satisfied, especially in the frontier cases that mostly concern us.

$\longrightarrow$ we shall come back to this point

Examples from particle physics

$\Rightarrow$ *See transparencies*

End of lecture

End of lecture 1

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

  $\Rightarrow$ 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.

. . . 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.

  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 steadly, 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 steadly, 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

Notes

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

Determinism/indeterminism

Pragmatically, as far as uncertainty and inference matter, it doesn't really matter.

"Though there be no such thing as Chance in the world; our ignorance of the real cause of any event has the same influence on the understanding, and begets a like species of belief or opinion" (Hume)

Go Back

Processo di Biscardi

A single quote gives an idea of the talk show:

> *"Please, don't speak more than two or three at the same time!"*

Go Back

Hume’s view about ‘combinatoric evaluation’

“There is certainly a probability, which arises from a superiority of chances on any side; and according as this superiority increases, and surpasses the opposite chances, the probability receives a proportionable increase, and begets still a higher degree of belief or assent to that side, in which we discover the superiority.”

Hume's view about 'combinatoric evaluation'

"There is certainly ***a probability, which arises from a superiority of chances*** *on any side; and according as this superiority increases, and surpasses the opposite chances, the probability receives a proportionable increase, and begets still a higher degree of belief or assent to that side, in which we discover the superiority. If a dye were marked with one figure or number of spots on four sides, and with another figure or number of spots on the two remaining sides, it would be more probable, that the former would turn up than the latter; though, if it had a thousand sides marked in the same manner, and only one side different,* ***the probability would be much higher, and our belief or expectation of the event more steady and secure****."* (David Hume)

Go Back

Hume’s view about ‘frequency based evaluation’

“Being determined by custom to transfer the past to the future, in all our inferences; where the past has been entirely regular and uniform, we expect the event with the greatest assurance, and leave no room for any contrary supposition.”

Hume's view about 'frequency based evaluation'

> *"Being determined by custom to transfer the* past to the future*, in all our inferences; where the past has been entirely regular and uniform, we expect the event with the greatest assurance, and leave no room for any contrary supposition. But where* different effects *have been found to* follow from causes*, which are to appearance exactly similar,* all these various effects must occur to the mind *in transferring the past to the future, and enter into our consideration, when we determine the probability of the event."*

Hume's view about 'frequency based evaluation'

"Being determined by custom to transfer the ***past to the future****, in all our inferences; where the past has been entirely regular and uniform, we expect the event with the greatest assurance, and leave no room for any contrary supposition. But where* ***different effects*** *have been found to* ***follow from causes****, which are to appearance exactly similar,* ***all these various effects must occur to the mind*** *in transferring the past to the future, and enter into our consideration, when we determine the probability of the event." Though we give the preference to that which has been found most usual, and believe that this effect will exist, we must not overlook the other effects, but must* ***assign to each of them a particular weight and authority, in proportion as we have found it to be more or less frequent****."* (David Hume)

Go Back

Bet odds to express confidence

“The best way to explain it is, I’ll bet you fifty to one that you don’t find anything”
(Feynman)

Bet odds to express confidence

"The best way to explain it is, I'll bet you fifty to one that you don't find anything" (Feynman)

"It is a bet of 11,000 to 1 that the error on this result (the mass of Saturn) *is not 1/100th of its value"* (Laplace)

Bet odds to express confidence

"The best way to explain it is, I'll bet you fifty to one that you don't find anything" (Feynman)

"It is a bet of 11,000 to 1 that the error on this result (the mass of Saturn) *is not 1/100th of its value"* (Laplace)

$\longrightarrow$ 99.99% confidence on the result

Bet odds to express confidence

“The best way to explain it is, I’ll bet you fifty to one that you don’t find anything” (Feynman)

“It is a bet of 11,000 to 1 that the error on this result (the mass of Saturn) *is not 1/100th of its value”* (Laplace)

$\longrightarrow$ 99.99% confidence on the result

$\Rightarrow$ Is a 95% C.L. upper/lower limit a ‘19 to 1 bet’?

Go Back