Introduction to Probabilistic Reasoning

– inference, forecasting, decision –

Giulio D'Agostini

giulio.dagostini@roma1.infn.it

Dipartimento di Fisica

Università di Roma La Sapienza

– Part 1 –

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

(Feynman)

Preamble

- What can we say in just a few hours?
  (The course I give in Rome to PhD students on
  "Probability and Uncertainty in Physics" takes 40 hours!)

Preamble

- What can we say in just a few hours?
  (The course I give in Rome to PhD students on *"Probability and Uncertainty in Physics"* takes 40 hours!)
- 'Statistics' is often felt as a boring collection of prescriptions...

Preamble

- What can we say in just a few hours?
  (The course I give in Rome to PhD students on *"Probability and Uncertainty in Physics"* takes 40 hours!)
- 'Statistics' is often felt as a boring collection of prescriptions...
  (note that the mentioned course is not about 'statistics'!)

Preamble

- What can we say in just a few hours?
  (The course I give in Rome to PhD students on *"Probability and Uncertainty in Physics"* takes 40 hours!)
- 'Statistics' is often felt as a boring collection of prescriptions...
  (note that the mentioned course is not about 'statistics'!)

  . . . in contrast to the nice Physics courses, in which the issues are related and stem from some foundamendal 'principles'

Preamble

- What can we say in just a few hours?
  (The course I give in Rome to PhD students on
  "Probability and Uncertainty in Physics" takes 40 hours!)
- 'Statistics' is often felt as a boring collection of prescriptions...
  (note that the mentioned course is not about 'statistics'!)

  ... in contrast to the nice Physics courses, in which the issues are related and stem from some foundamendal 'principles'

$\Rightarrow$ No 'prescriptions', but general ideas

Preamble

- What can we say in just a few hours?
  (The course I give in Rome to PhD students on *"Probability and Uncertainty in Physics"* takes 40 hours!)
- 'Statistics' is often felt as a boring collection of prescriptions...
  (note that the mentioned course is not about 'statistics'!)

  . . . in contrast to the nice Physics courses, in which the issues are related and stem from some foundamendal 'principles'

⇒ No 'prescriptions', but general ideas

... possibly arising from
'first principles' (as we physicists like).

Preamble

- What can we say in just a few hours?
  (The course I give in Rome to PhD students on *"Probability and Uncertainty in Physics"* takes 40 hours!)
- 'Statistics' is often felt as a boring collection of prescriptions...
  (note that the mentioned course is not about 'statistics'!)

  . . . in contrast to the nice Physics courses, in which the issues are related and stem from some foundamendal 'principles'

⇒ No 'prescriptions', but general ideas

... possibly arising from
'first principles' (as we physicists like).

⇒ Probabilistic approach

An invitation to (re-)think
on foundamental aspects
of data analysis.

Contents

Lectio est omnis divisa in partes tres. . .

(not strictly related to the individuel ‘lectures’ and to the ‘tutorial)

Contents

Lectio est omnis divisa in partes tres. . .

(not strictly related to the individuel 'lectures' and to the 'tutorial)

1. *Claims of discoveries based on 'sigmas'*
2. *Basic of probabilistic inference/forcasting*
   (and related topics)
3. *Basic applications in data analysis*

Contents

Lectio est omnis divisa in partes tres...

(not strictly related to the individuel 'lectures' and to the 'tutorial)

1. *Claims of discoveries based on 'sigmas'*
2. *Basic of probabilistic inference/forcasting*
   (and related topics)
3. *Basic applications in data analysis*

More 'professional' applications, including computational issues will be presented in the next days

- Allen Caldwell
- Dan
- Roberto Trotta

Software support

- My *extended pocket calulator*:
  - R

  (a 'kind of' Mathlab/Octave; opensource, multiplatform, very easy to start with; numeric computation, probability/statistics functions, graphics; tons of packages. ≈ Nr 1 in statistics, maching learning, etc.)

Software support

- My *extended pocket calulator*:
  - R

  (a ‘kind of’ Mathlab/Octave; opensource, multiplatform, very easy to start with; numeric computation, probability/statistics functions, graphics; tons of packages. ≈ Nr 1 in statistics, maching learning, etc.)
- Hugin (Lite)
- OpenBUGS (mentioning JAGS/rjags)

Software support

- My *extended pocket calulator*:
  - R

  (a 'kind of' Mathlab/Octave; opensource, multiplatform, very easy to start with; numeric computation, probability/statistics functions, graphics; tons of packages. $\approx$ Nr 1 in statistics, maching learning, etc.)
- Hugin (Lite)
- OpenBUGS (mentioning JAGS/rjags)

Information/examples in a web page:
`http://www.roma1.infn.it/~dagos/stellenbosch`

Software support

- My *extended pocket calulator*:
  - R

  (a 'kind of' Mathlab/Octave; opensource, multiplatform, very easy to start with; numeric computation, probability/statistics functions, graphics; tons of packages. ≈ Nr 1 in statistics, maching learning, etc.)
- Hugin (Lite)
- OpenBUGS (mentioning JAGS/rjags)

Information/examples in a web page:

`http://www.roma1.infn.it/~dagos/stellenbosch`

⇒ Don't be distracted during the lectures with installations or playing with them (→ better later, or even back home)

Part 1

Claims of discoveries based on 'sigmas'

2011: non only Opera...

- April, CDF: absolutely unexpected excess at about 150 GeV

$$\approx 3.2\,\sigma$$

- September, Opera: neutrinos faster than light

$$\approx 6\,\sigma$$

- December, ATLAS e CMS at LHC: signal compatible with the Higgs at about 125 GeV:

$$\approx 3\sigma$$

2011: non only Opera. . .

- April, CDF: absolutely unexpected excess at about 150 GeV

$$\approx 3.2\,\sigma$$

- September, Opera: neutrinos faster than light

$$\approx 6\,\sigma$$

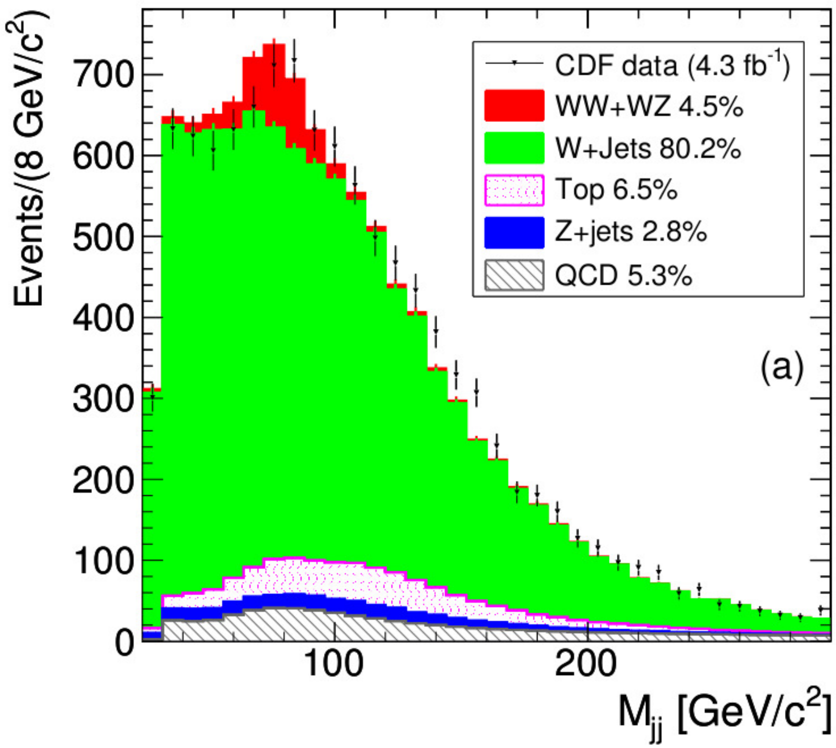
- December, ATLAS e CMS at LHC: signal compatible with the Higgs at about 125 GeV:

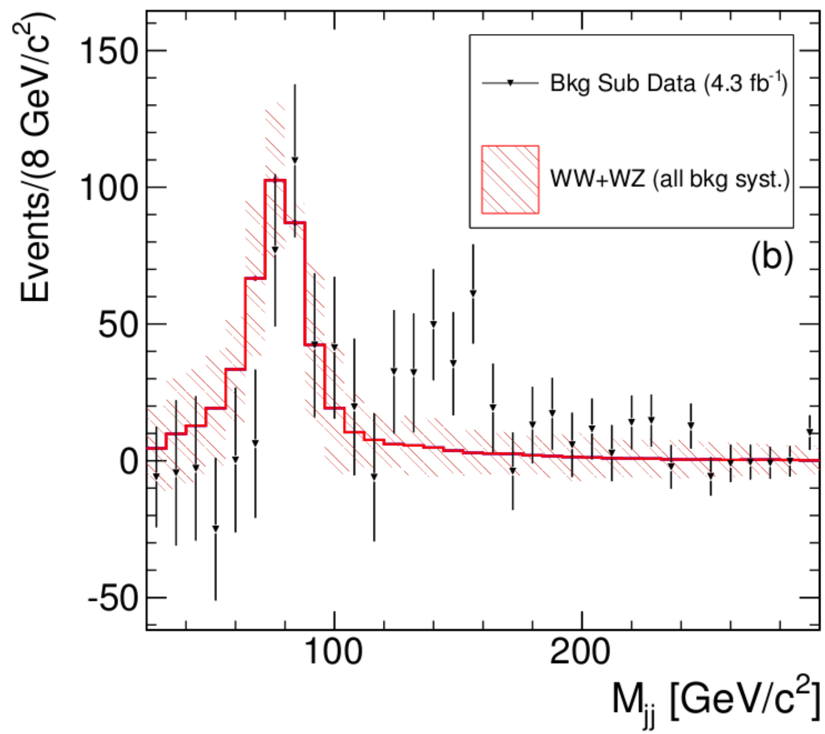
$$\approx 3\sigma$$

Why there was substancial scepticism towards the first two anouncements, in constrast with a cautious/pronounced optimism towards the third one?

April 2011

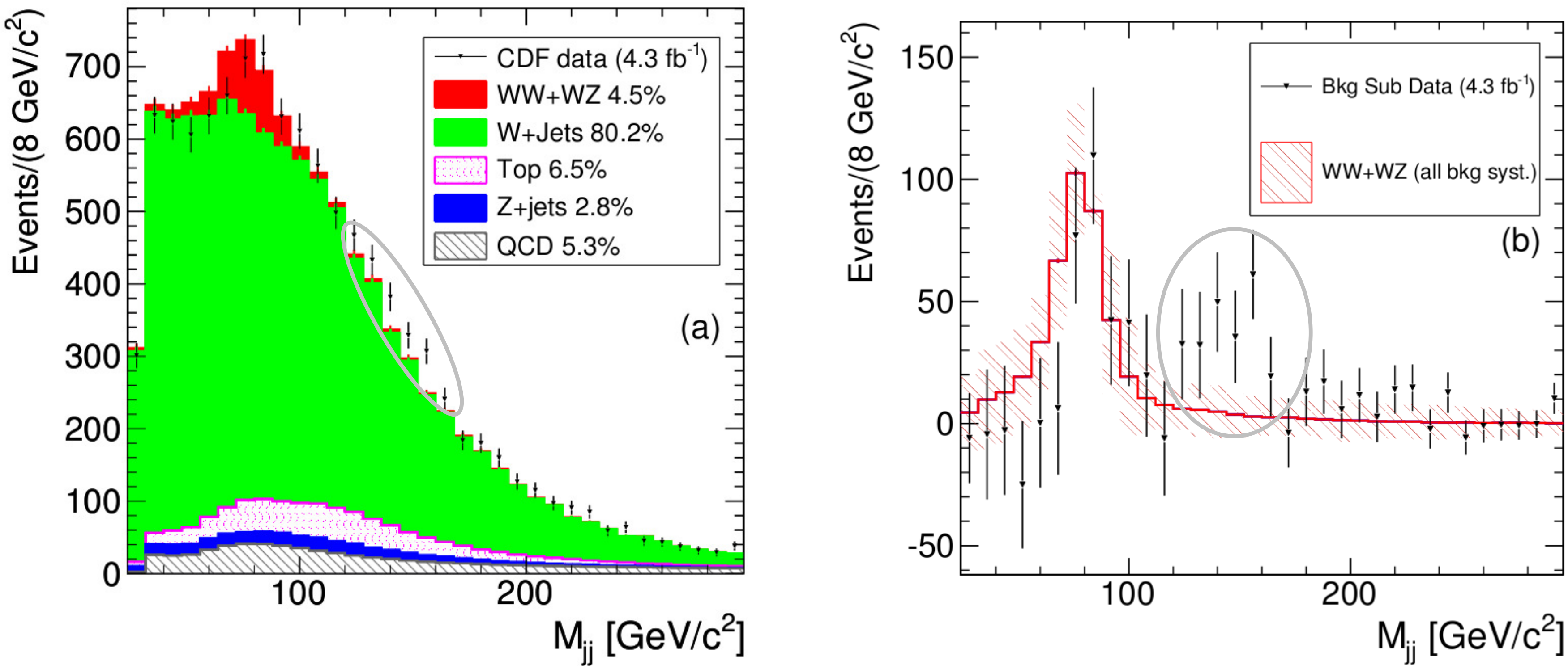
CDF Collaboration at the Tevatron





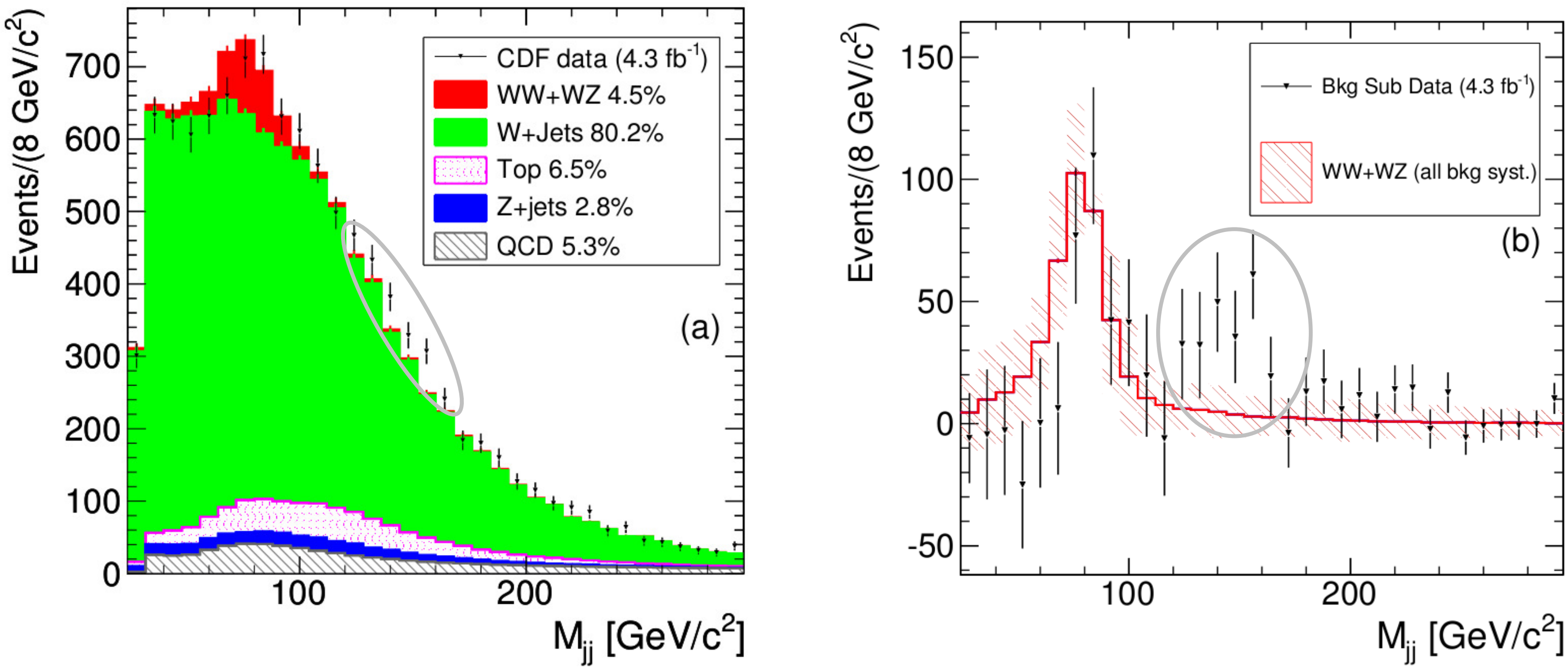
April 2011

CDF Collaboration at the Tevatron



April 2011

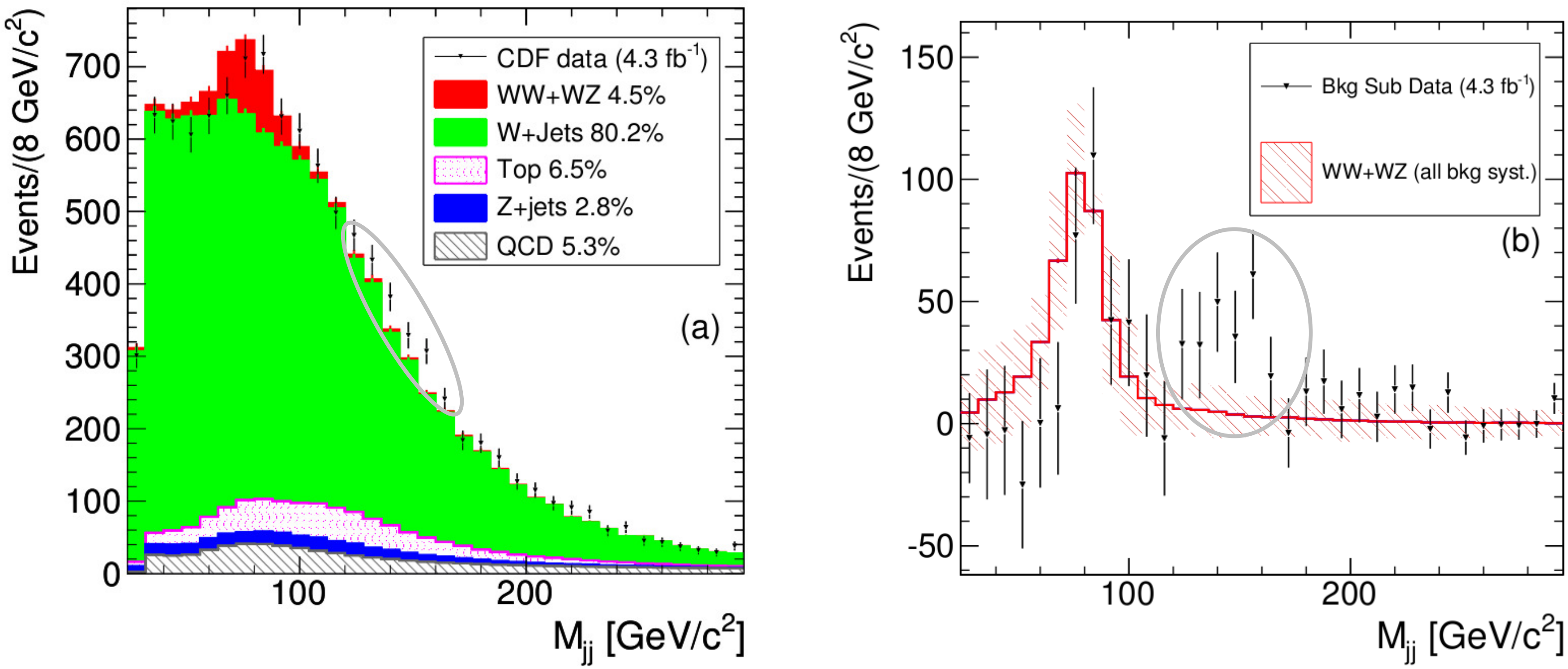
CDF Collaboration at the Tevatron



"we obtain a p-value of 7.6×10^{-4}, corresponding to a significance of 3.2 standard deviations"

April 2011

CDF Collaboration at the Tevatron

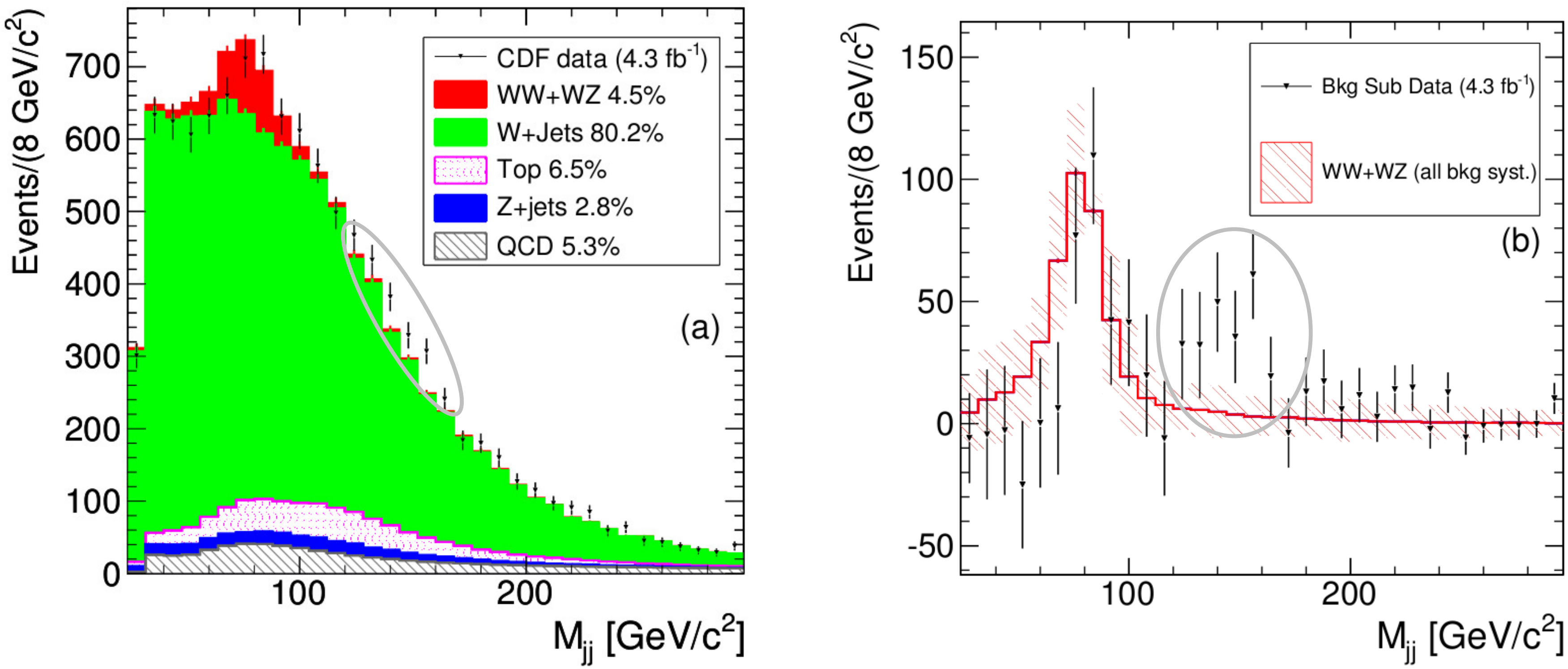


"we obtain a p-value of 7.6×10^{-4}, corresponding to a significance of 3.2 standard deviations"

$3.2\,\sigma\,!$

April 2011

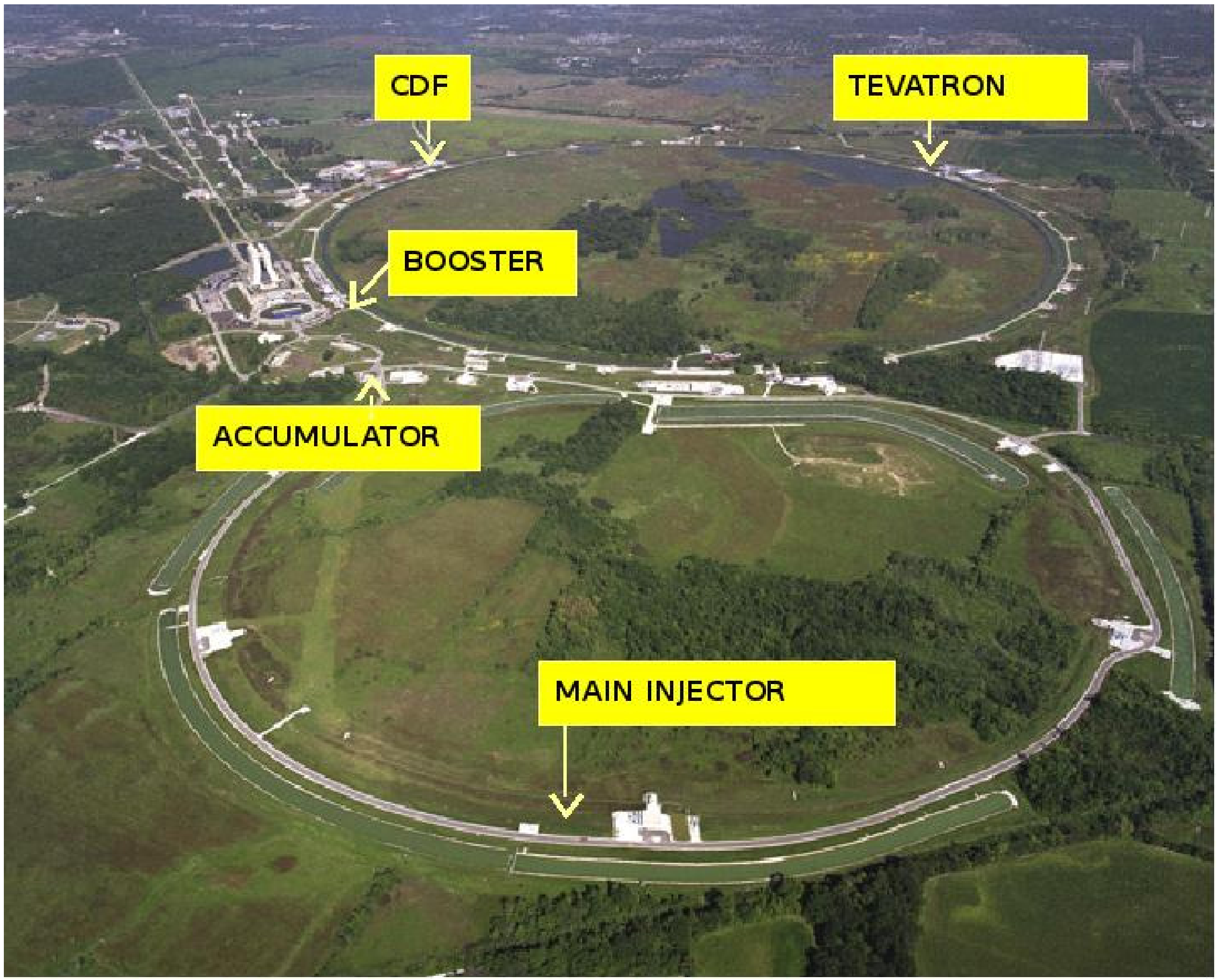
CDF Collaboration at the Tevatron



What does it mean?

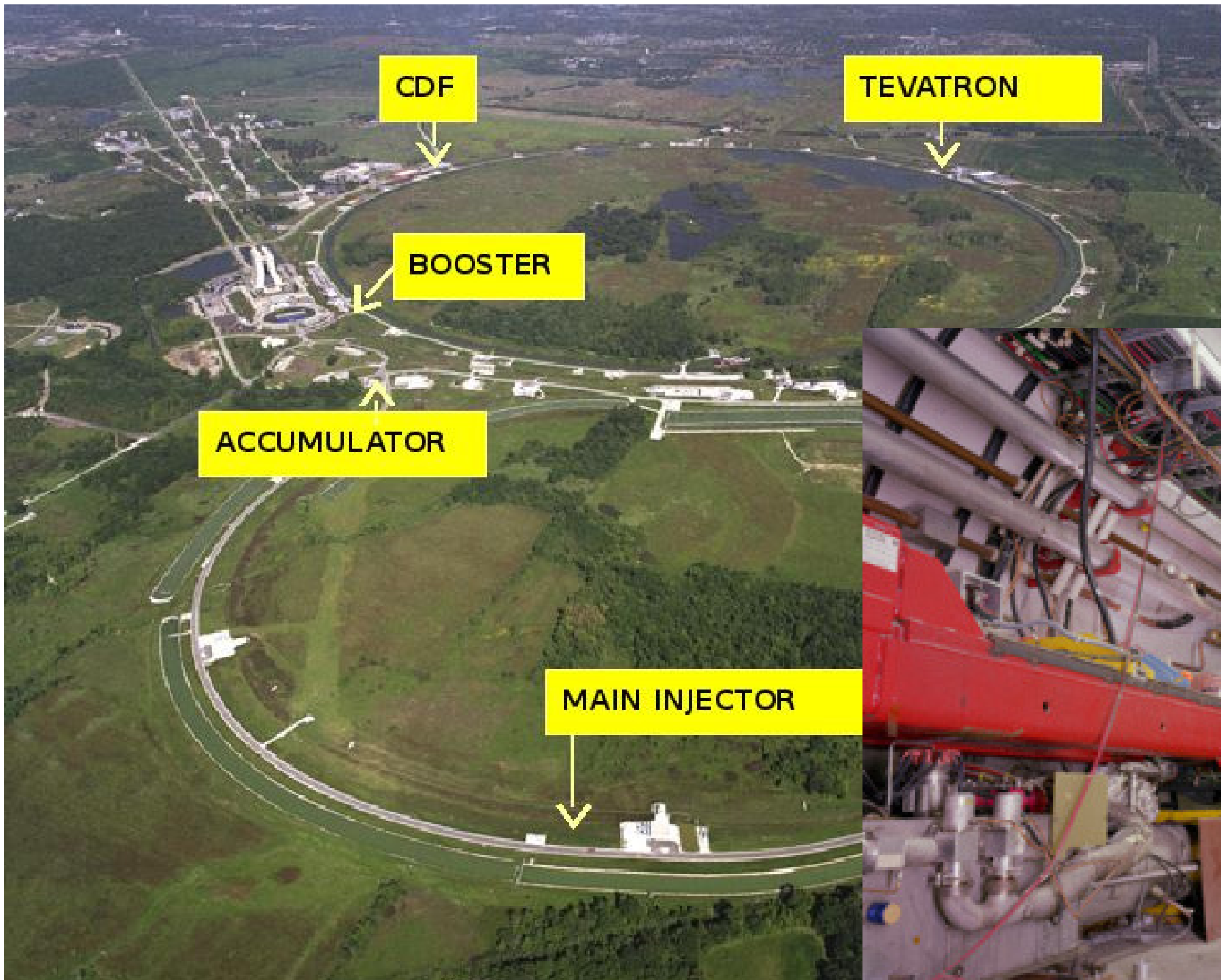
Tevatron and CDF

6.28 km, near Chicago



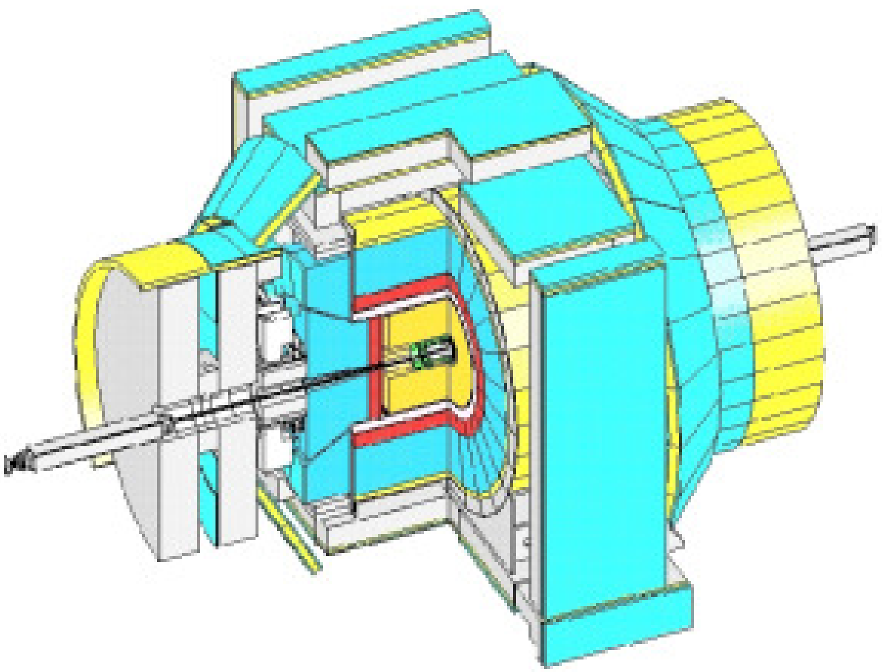
Tevatron and CDF

$p \rightarrow \cdot \leftarrow \bar{p}$ $[\approx 1\,\text{TeV} + 1\,\text{TeV}]$



Tevatron and CDF

CDF: a multipurpose ('hermetic') detector

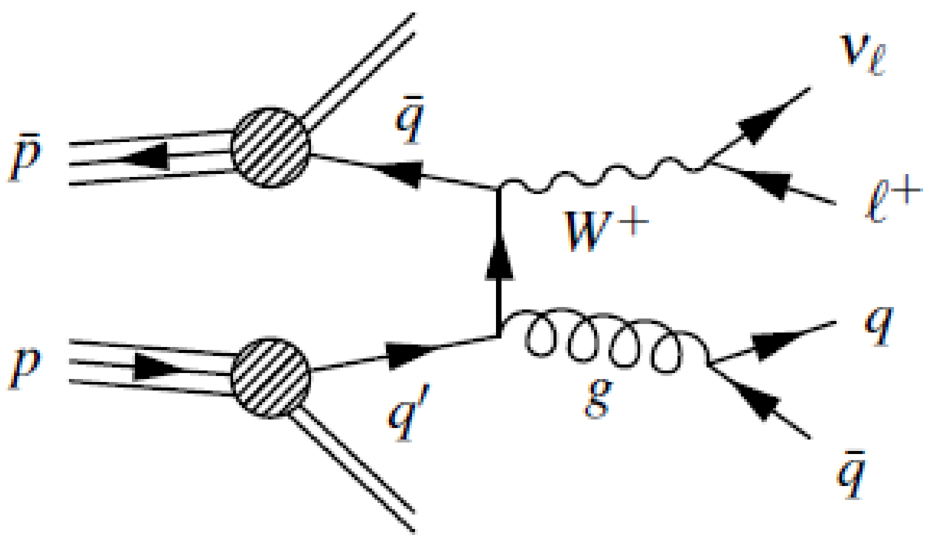


Tevatron and CDF

. . . a large, very sophisticated detector!

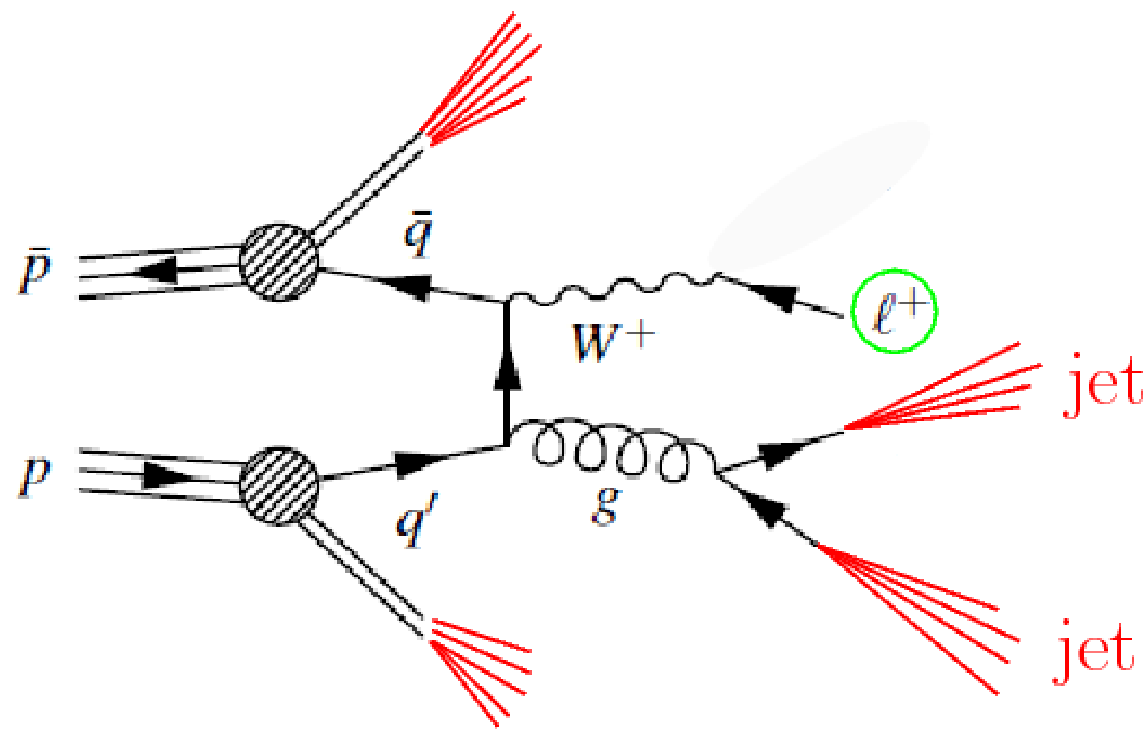
Jet-jet + W

$W + (q\overline{q})$ [+ 'remnants']



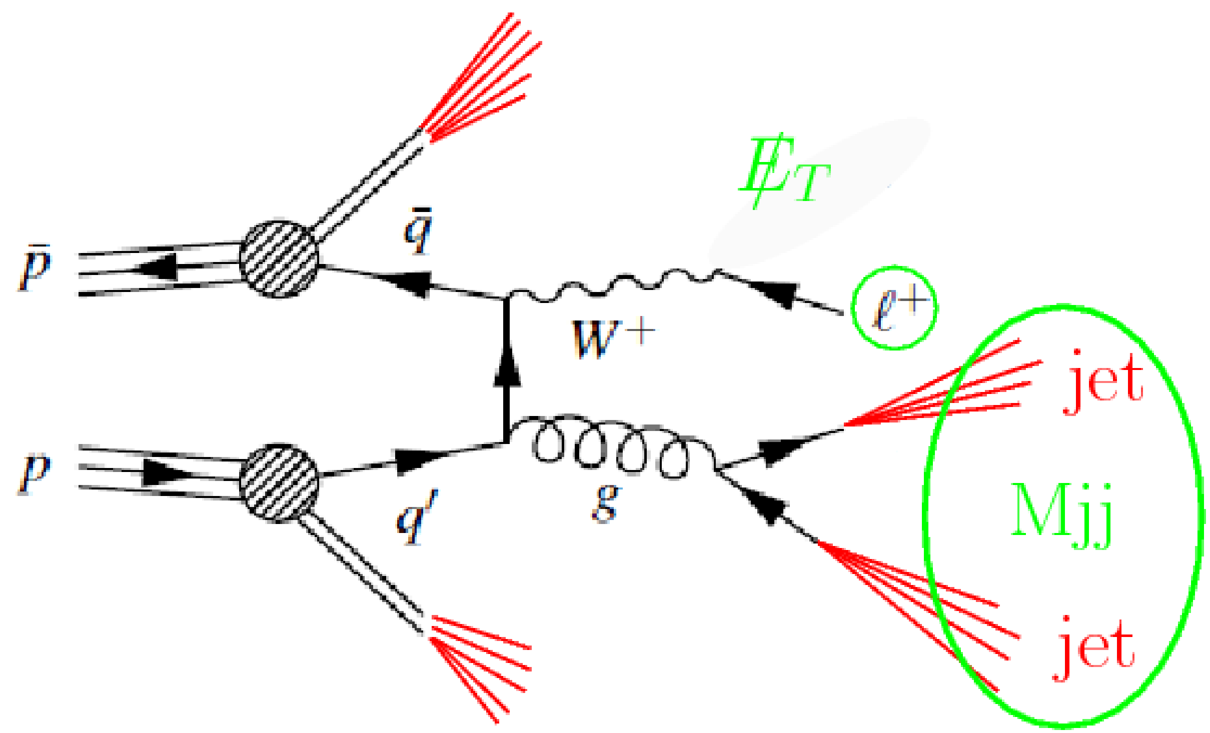
Jet-jet + W

$W + 2$jet [+ much more]



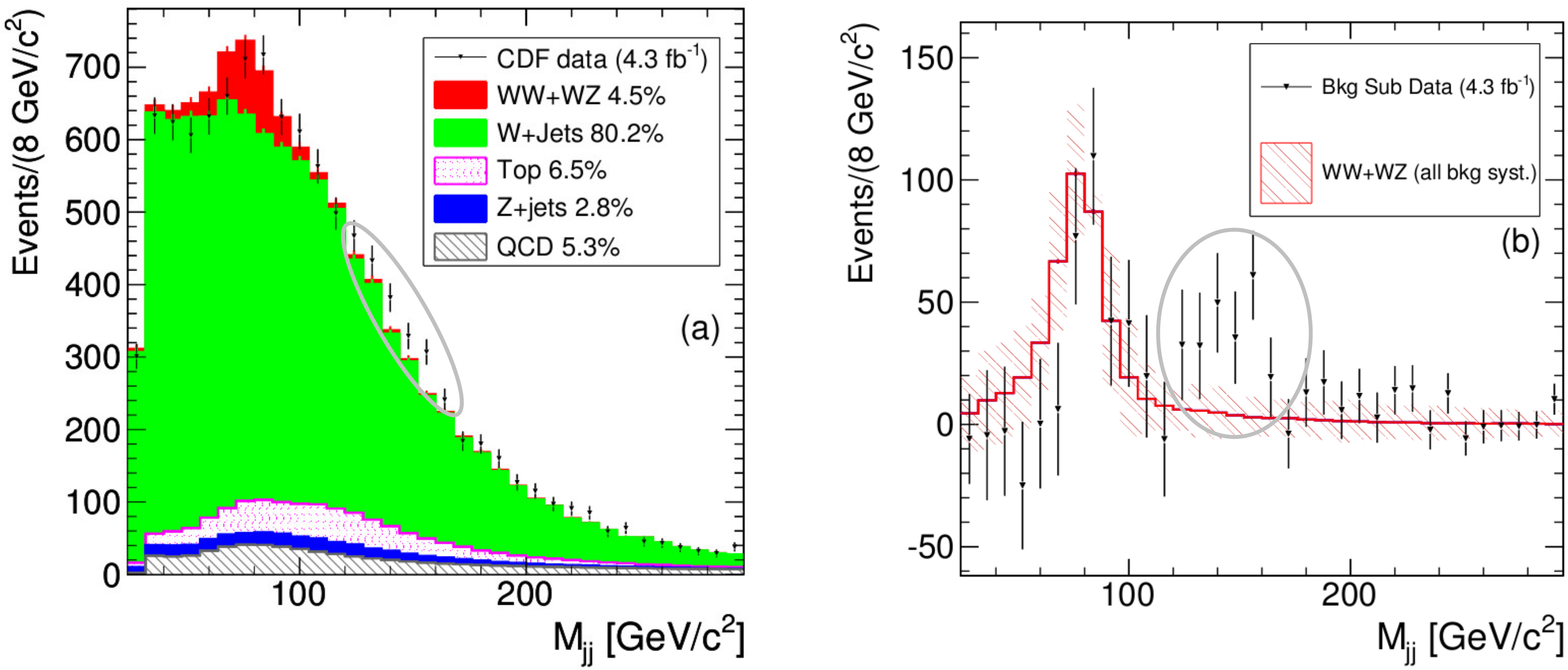
Jet-jet + W

$\Rightarrow M_{jj} + W + \ldots$



The ‘bump’!

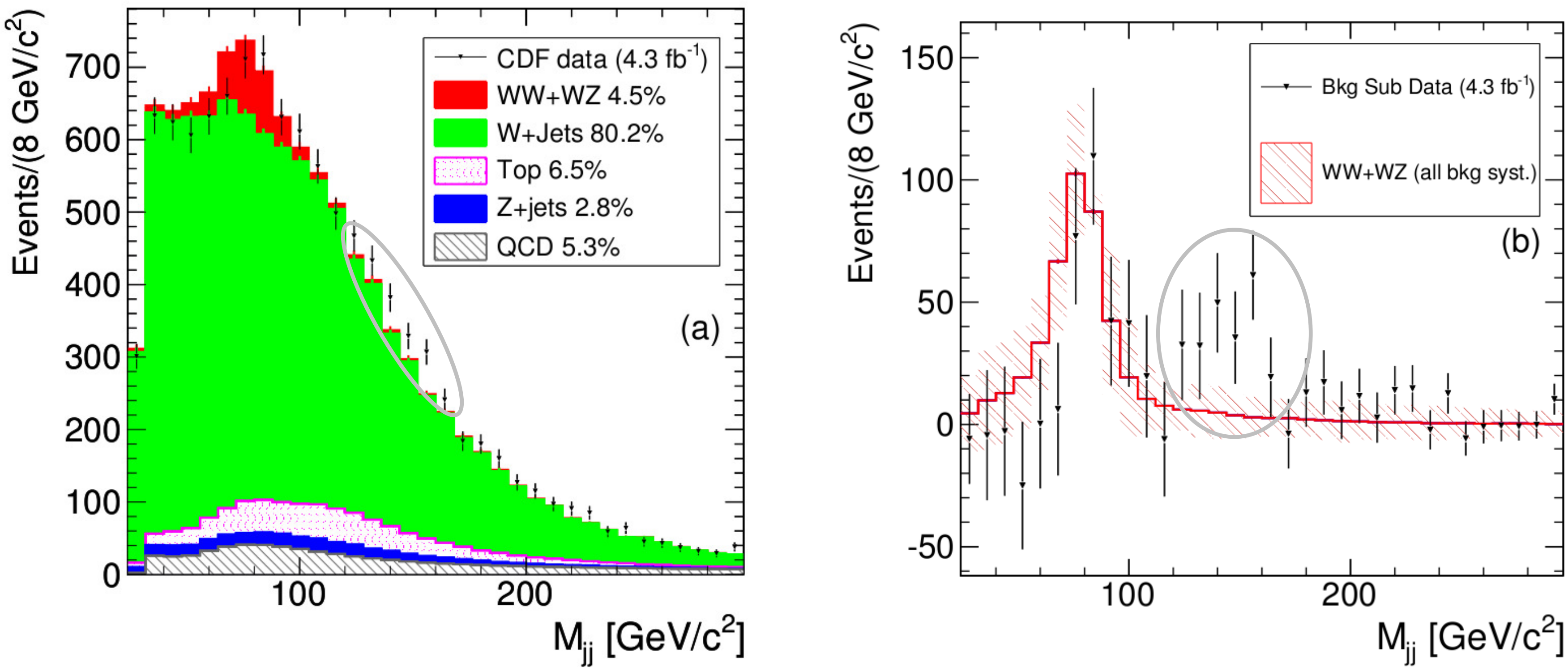
Invariant Mass Distribution of Jet Pairs Produced in Association with a W boson in $p\overline{p}$ Collisions at $\sqrt{s} = 1.96$ TeV”, (CDF, 4 aprile 2011)



“we obtain a p-value of 7.6×10^{-4}, corresponding to a significance of 3.2 standard deviations” [“$3.2\,\sigma$”]

The ‘bump’!

Invariant Mass Distribution of Jet Pairs Produced in Association with a W boson in $p\overline{p}$ Collisions at $\sqrt{s} = 1.96$ TeV”, (CDF, 4 aprile 2011)



What does it mean?

Sigma and gaussian distribution

Sigma and gaussian distribution

Sigma e probability [gaussian!]

If the random number X is described by a gaussian pdf

$$
\begin{aligned}
P(-\sigma \le X \le +\sigma) &= 68.3\% \\
P(-2\,\sigma \le X \le +2\,\sigma) &= 95.4\% \\
P(-3\,\sigma \le X \le +3\,\sigma) &= 99.73\% \\
1 - P(-3\,\sigma \le X \le +3\,\sigma) &= 0.27\% \\
1 - P(-4\,\sigma \le X \le +4\,\sigma) &= 6.3 \times 10^{-5} \\
\ldots &= \ldots \\
1 - P(-6\,\sigma \le X \le +6\,\sigma) &= 2.0 \times 10^{-9}
\end{aligned}
$$

$$
\begin{aligned}
1 - P(-3.2\,\sigma \le X \le +3.2\,\sigma) &= 1.4 \times 10^{-3} \\
P(X \ge +3.17\,\sigma) &= 7.6 \times 10^{-4} \quad \surd
\end{aligned}
$$

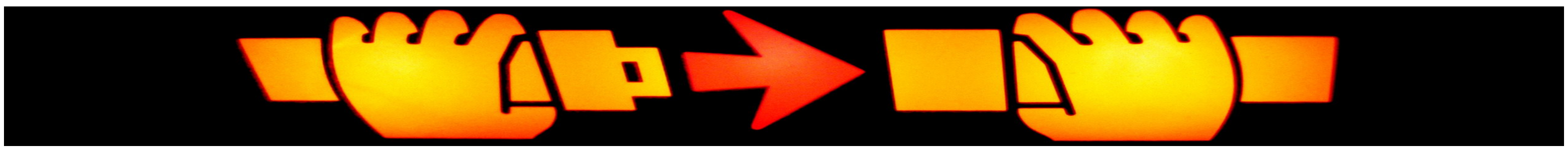
p-value, significance and sigma

"we obtain a p-value of 7.6×10^{-4}, corresponding to a significance of 3.2 standard deviations" ["$3.2\,\sigma$"]

p-value, significance and sigma

"we obtain a p-value of 7.6×10^{-4}, corresponding to a significance of 3.2 standard deviations" [“$3.2\,\sigma$”]

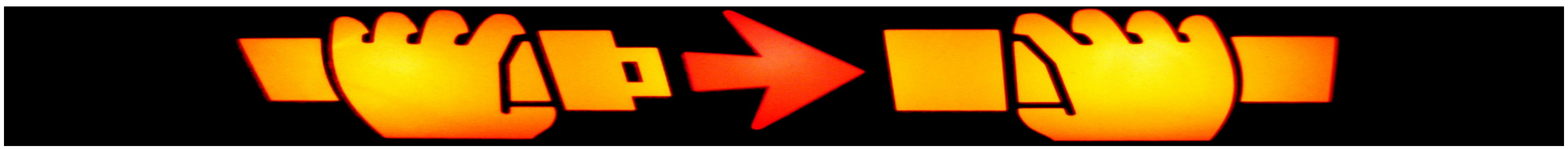
Begin to fasten seat belts!



p-value, significance and sigma

"we obtain a p-value of 7.6×10^{-4}, corresponding to a significance of 3.2 standard deviations" ["$3.2\,\sigma$"]

Begin to fasten seat belts!

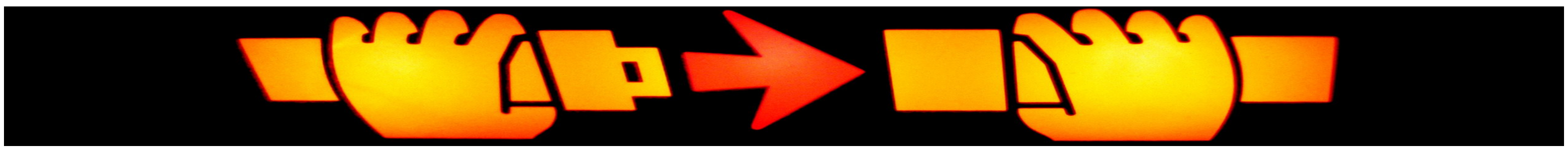


- <u>What is</u> a **p-value**?
- <u>In so far</u> does it provides us a '**significance**'?

p-value, significance and sigma

"we obtain a p-value of 7.6×10^{-4}, corresponding to a significance of 3.2 standard deviations" ["$3.2\,\sigma$"]

Begin to fasten seat belts!



- What is a **p-value**?
- In so far does it provides us a '**significance**'?

In short,

- Is 7.6×10^{-4} a **probability**?
- **of what**?

Aprile 2011, the 'bump' explodes

The New York Times, Tuesday, April 5:

"Physicists at the Fermi National Accelerator Laboratory are planning to announce Wednesday that they have found a suspicious bump in their data that could be evidence of a new elementary particle or even, some say, a new force of nature.

. . .

The experimenters estimate that **there is a less than a quarter of 1 percent chance their bump is a statistical fluctuation***"*

Aprile 2011, the ‘bump’ explodes

The New York Times, Tuesday, April 5:

“Physicists at the Fermi National Accelerator Laboratory are planning to announce Wednesday that they have found a suspicious bump in their data that could be evidence of a new elementary particle or even, some say, a new force of nature.

. . .

The experimenters estimate that **there is a less than a quarter of 1 percent chance their bump is a statistical fluctuation***”*

$$P(\text{Statistical fluctuation}) \le 0.25\%!$$

Aprile 2011, the 'bump' explodes

The New York Times, Tuesday, April 5:

> *"Physicists at the Fermi National Accelerator Laboratory are planning to announce Wednesday that they have found a suspicious bump in their data that could be evidence of a new elementary particle or even, some say, a new force of nature.*
>
> *...*
>
> *The experimenters estimate that* **there is a less than a quarter of 1 percent chance their bump is a statistical fluctuation***"*

$$P(\text{Statistical fluctuation}) \leq 0.25\%!$$

$$P(\text{True Signal}) \geq 99.75\%!!$$

Aprile 2011, the 'bump' explodes

The New York Times, Tuesday, April 5:

"Physicists at the Fermi National Accelerator Laboratory are planning to announce Wednesday that they have found a suspicious bump in their data that could be evidence of a new elementary particle or even, some say, a new force of nature.

...

The experimenters estimate that **there is a less than a quarter of 1 percent chance their bump is a statistical fluctuation***"*

$$P(\text{Statistical fluctuation}) \leq 0.25\%!$$

$$P(\text{True Signal}) \geq 99.75\%!!$$

Eureka!!

Aprile 2011, the ‘bump’ explodes

The New York Times, Tuesday April 5:

> *“the most significant in physics in half a century”*

Aprile 2011, the ‘bump’ explodes

The New York Times, Tuesday April 5:

> *“the most significant in physics in half a century”*

[Do not ask me how 7.6×10^{-4} becomes $< 2.5 \times 10^{-3}$ (but this can be considere a minor detail. . .)]

Aprile 2011, the ‘bump’ explodes

The New York Times, Tuesday April 5:

> *“the most significant in physics in half a century”*

Much more important the unusual fact that an ArXiV appeared one day was commented by NYT the day after!

Aprile 2011, the ‘bump’ explodes

The New York Times, Tuesday April 5:

> *“the most significant in physics in half a century”*

Much more important the unusual fact that an ArXiV appeared one day was commented by NYT the day after!

Who believed it was – at 99.75%! – a discover?

- the journalist who reported the news?
- the CDF contactperson and/or the Fermilab PR’s who contacted him?

Aprile 2011, the 'bump' explodes

The New York Times, Tuesday April 5:

> *"the most significant in physics in half a century"*

Much more important the unusual fact that an ArXiV appeared one day was commented by NYT the day after!

Who believed it was – at 99.75%! – a discover?

- the journalist who reported the news?
- the CDF contactperson and/or the Fermilab PR's who contacted him?

From my experience, journalists might make imprecisions, bad they do not invent pieces of news [. . . at least scientific ones. . . :-)]

Aprile 2011, the 'bump' explodes

Fermilab Today, April 7:

> *"Wednesday afternoon, the CDF collaboration announced that it has evidence of a peak in a specific sample of its data. The peak is an excess of particle collision events that produce a W boson accompanied by two hadronic jets. This peak showed up in a mass region where we did not expect one.*
>
> *. . .*

Aprile 2011, the 'bump' explodes

Fermilab Today, April 7:

"Wednesday afternoon, the CDF collaboration announced that it has evidence of a peak in a specific sample of its data. The peak is an excess of particle collision events that produce a W boson accompanied by two hadronic jets. This peak showed up in a mass region where we did not expect one.

. . .

The significance of this excess was determined to be 3.2 sigma, after accounting for the effect of systematic uncertainties. This means that **there is less than a 1 in 1375 chance that the effect is mimicked by a statistical fluctuation***."*

Aprile 2011, the 'bump' explodes

Fermilab Today, April 7:

> *"Wednesday afternoon, the CDF collaboration announced that it has evidence of a peak in a specific sample of its data. The peak is an excess of particle collision events that produce a W boson accompanied by two hadronic jets. This peak showed up in a mass region where we did not expect one.*
>
> *...*
>
> *The significance of this excess was determined to be 3.2 sigma, after accounting for the effect of systematic uncertainties. This means that* **there is less than a 1 in 1375 chance that the effect is mimicked by a statistical fluctuation**.*"*

$1/1375 = 7.3 \times 10^{-4} \;\Rightarrow\; P(\text{No stat. fluct.}) = 99.93\%$!

Aprile 2011, the 'bump' explodes

Discovery News, April 7:

> *This is a big week for particle physicists, and even they will be having many sleepless nights over the coming months trying to grasp what it all means.*
> *That's what happens when physicists come forward, with observational evidence, of what they believe represents something we've never seen before. Even bigger than that: something we never even expected to see.*
> *...*

Aprile 2011, the ‘bump’ explodes

Discovery News, April 7:

> *This is a big week for particle physicists, and even they will be having many sleepless nights over the coming months trying to grasp what it all means.*
> *That’s what happens when physicists come forward, with observational evidence, of what they believe represents something we’ve never seen before. Even bigger than that: something we never even expected to see.*
> *. . .*
> *It is what is known as a ”three-sigma event,” and this refers to the statistical certainty of a given result. In this case,* **this result has a 99.7 percent chance of being correct (and a 0.3 percent chance of being wrong)***.”*

Aprile 2011, the 'bump' explodes

Discovery News, April 7:

> *This is a big week for particle physicists, and even they will be having many sleepless nights over the coming months trying to grasp what it all means.*
> *That's what happens when physicists come forward, with observational evidence, of what they believe represents something we've never seen before. Even bigger than that: something we never even expected to see.*
>
> *. . .*
>
> *It is what is known as a "three-sigma event," and this refers to the statistical certainty of a given result. In this case,* **this result has a 99.7 percent chance of being correct (and a 0.3 percent chance of being wrong)**."

It seems we are understanding well, besides the fact of how 99.9% becomes 99.7%. . .

Aprile 2011, the ‘bump’ explodes

Jon Butterworth’s blob on the Guardian, April 9:

> *“The last and greatest breakthrough from a fantastic machine, or a false alarm on the frontiers of physics?*
>
> *. . .*
>
> *If the histograms and data are exactly right,* **the paper quotes a one-in-ten-thousand (0.0001) chance that this bump is a fluke***.”*

Aprile 2011, the ‘bump’ explodes

Jon Butterworth’s blob on the Guardian, April 9:

> *“The last and greatest breakthrough from a fantastic machine, or a false alarm on the frontiers of physics?*
>
> *. . .*
>
> *If the histograms and data are exactly right,* **the paper quotes a one-in-ten-thousand (0.0001) chance that this bump is a fluke**.*”* $\Rightarrow P(\text{Not Fluke}) = P(\text{“Genuine”}) = 99.99\%$

Aprile 2011, the 'bump' explodes

Jon Butterworth's blob on the Guardian, April 9:

> *"The last and greatest breakthrough from a fantastic machine, or a false alarm on the frontiers of physics?*
>
> *...*
>
> *If the histograms and data are exactly right,* **the paper quotes a one-in-ten-thousand (0.0001) chance that this bump is a fluke**.*"* $\Rightarrow P(\text{Not Fluke}) = P(\text{“Genuine”}) = 99.99\%$

But, at the end of the post:

1. "My money is on the false alarm at the moment,. . . "
2. ". . . but I would be very happy to lose it."
3. "And I reserve the right to change my mind rapidly as more data come in!"

Aprile 2011, the 'bump' explodes

Jon Butterworth's blob on the Guardian, April 9:

> *"The last and greatest breakthrough from a fantastic machine, or a false alarm on the frontiers of physics?*
>
> *...*
>
> *If the histograms and data are exactly right,* **the paper quotes a one-in-ten-thousand (0.0001) chance that this bump is a fluke**.*"* $\Rightarrow P(\text{Not Fluke}) = P(\text{"Genuine"}) = 99.99\%$

But, at the end of the post:

1. "My money is on the false alarm at the moment,..."
2. "...but I would be very happy to lose it."
3. "And I reserve the right to change my mind rapidly as more data come in!"

Assolutetly meaningful! (A part from the initial mismatch)

A materpieceof good reasoning

Jon Butterworth's blob on the Guardian, April 9:

1. "My money is on the false alarm at the moment,. . . "

A materpieceof good reasoning

Jon Butterworth's blob on the Guardian, April 9:

1. "My money is on the false alarm at the moment,. . . "

   "I don't believe it!"

A materpieceof good reasoning

Jon Butterworth's blob on the Guardian, April 9:

1. "My money is on the false alarm at the moment,. . . "

   "I don't believe it!"

2. ". . . but I would be very happy to lose it."

   "What I wish" $\neq$ "What I would like"

A materpieceof good reasoning

Jon Butterworth's blob on the Guardian, April 9:

1. "My money is on the false alarm at the moment,. . . "

   "I don't believe it!"

2. "... but I would be very happy to lose it."

   "What I wish" $\neq$ "What I would like"

3. "And I reserve the right to change my mind rapidly as more data come in!"

   "Learning from the experience!"

   $\Rightarrow$ A physicist should never be dogmatic

A materpieceof good reasoning

Jon Butterworth's blob on the Guardian, April 9:

1. "My money is on the false alarm at the moment,. . . "

   "I don't believe it!"

2. ". . . but I would be very happy to lose it."

   "What I wish" $\neq$ "What I would like"

3. "And I reserve the right to change my mind rapidly as more data come in!"

   "Learning from the experience!"

   $\Rightarrow$ A physicist should never be dogmatic

But how <u>must</u> our convictions <u>rationally</u> change on the light of new experimental data? Is there a logical rule?

‘Significant’, but not believable!. . .

Jon Butterworth was not the only one to disbelieve the result.
Indeed, the largest majority of physicists disbelieve it.

‘Significant’, but not believable!. . .

Jon Butterworth was not the only one to disbelieve the result.
Indeed, the largest majority of physicists disbelieve it.
$\Rightarrow$ More or less like in the better known case of Opera’s neutrinos faster than light. . . (6σ!)

‘Significant’, but not believable!. . .

Jon Butterworth was not the only one to disbelieve the result.
Indeed, the largest majority of physicists disbelieve it.
⇒ More or less like in the better known case of Opera’s neutrinos faster than light. . . (6σ!)

But, then, what the hell do “significances” mean?

‘Significant’, but not believable!. . .

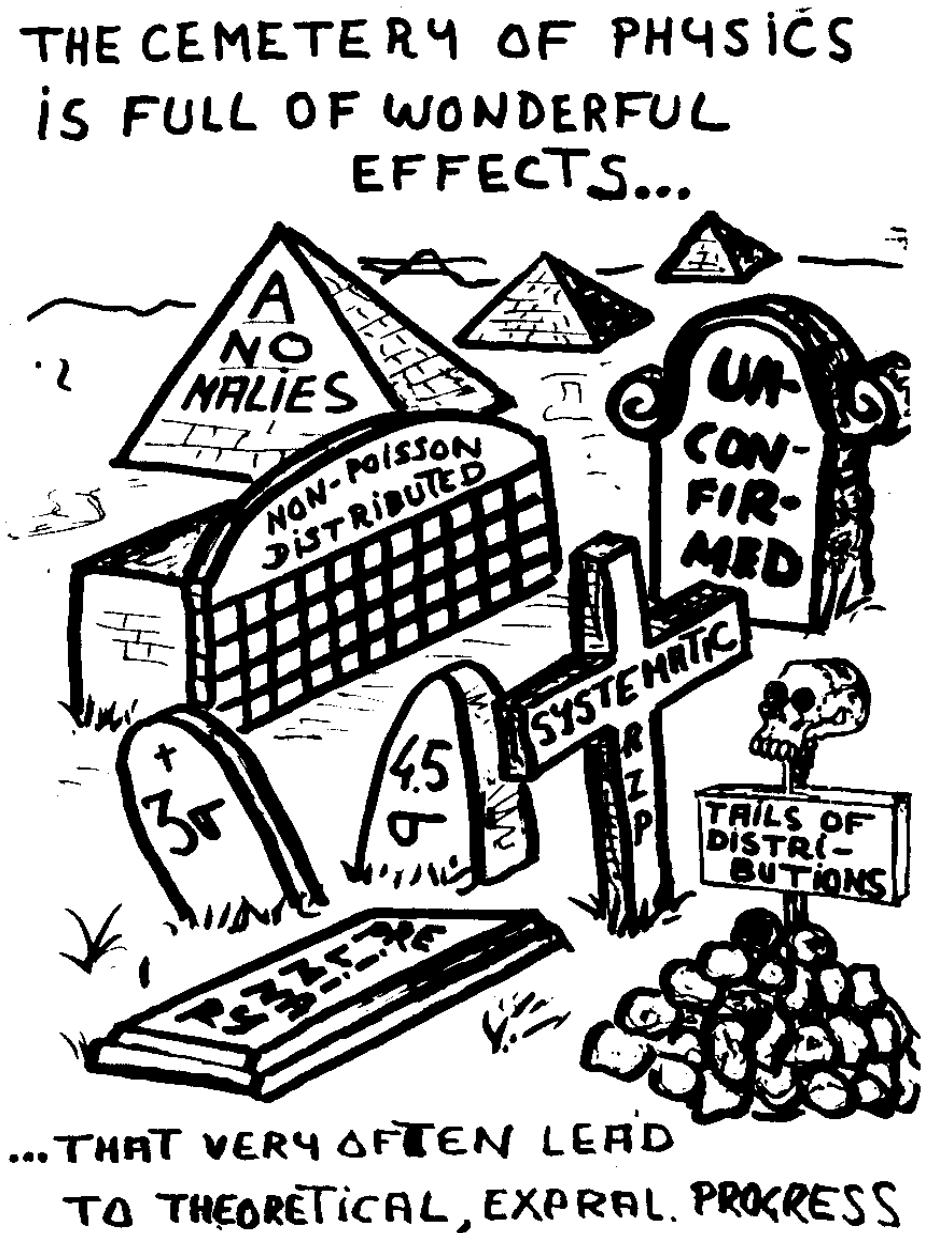
Jon Butterworth was not the only one to disbelieve the result.
Indeed, the largest majority of physicists disbelieve it.
⇒ More or less like in the better known case of Opera’s neutrinos faster than light. . . (6σ!)

But, then, what the hell do “significances” mean?

“de Rujula’s paradox”:

> *“If you disbelieve every result presented as having a 3 sigma – or ”equivalently” a 99.7% chance – of being correct. . . You will turn out to be right 99.7% of the times.”*
>
> (Alvaro de Rujula, private communication)

The cemetery of Physics



Alvaro de Rujula

Testing one hypothesis

- Basic Idea:
  - let's start from a 'conventional' model
    [Standard Modell, rather 'extablished theory', etc:]
    → "H_0" ("null hypothesis")

Testing <u>one</u> hypothesis

- Basic Idea:
  - let's start from a 'conventional' model
    [Standard Modell, rather 'extablished theory', etc:]
    $\rightarrow$ "H_0" ("null hypothesis")

  $\Rightarrow$ search for violations of H_0

Testing one hypothesis

- Basic Idea:
  - let's start from a 'conventional' model
    [Standard Modell, rather 'extablished theory', etc:]
    $\rightarrow$ "H_0" ("null hypothesis")

  $\Rightarrow$ search for violations of H_0
- Ideally

  $\rightarrow$ 'falsify' H_0

Testing <u>one</u> hypothesis

- Basic Idea:
  - let's start from a 'conventional' model
    [Standard Modell, rather 'extablished theory', etc:]
    $\rightarrow$ "H_0" ("null hypothesis")

  $\Rightarrow$ search for violations of H_0
- Ideally

  $\rightarrow$ 'falsify' H_0
- In practice:

  $\rightarrow$ does it make sense?

  $\rightarrow$ how is it done?

Testing one hypothesis

- Basic Idea:
  - let's start from a 'conventional' model
    [Standard Modell, rather 'extablished theory', etc:]
    $\rightarrow$ "H_0" ("null hypothesis")

  $\Rightarrow$ search for violations of H_0
- Ideally

  $\rightarrow$ 'falsify' H_0
- In practice:

  $\rightarrow$ does it make sense?

  $\rightarrow$ how is it done?

Let's review the practice and what is behind it $\Rightarrow$

Falsificationism

Usually referred to Popper

and still considered by many as
the *key of scientific progress*.

Falsificationism

Usually referred to Popper

and still considered by many as the *key of scientific progress*.

$$\textbf{if } C_i \not\longrightarrow E_0, \textbf{ then } E_0^{\textbf{(mis)}} \not\longrightarrow C_i$$

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

Falsificationism

Usually referred to Popper

and still considered by many as the *key of scientific progress*.

$$\textbf{if } C_i \not\longrightarrow E_0, \textbf{ then } E_0^{\textbf{(mis)}} \not\longrightarrow C_i$$

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

It seems OK – *'obvious'*! – but it is indeed naïve for several aspects.

Proof by contradiction . . . ‘extended’. . .

Falsification rule: to what is ‘inspired’?

Proof by contradiction . . . ‘extended’. . .

Falsification rule: to what is ‘inspired’?

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 rejected.

Proof by contradiction . . . ‘extended’. . .

Falsification rule: to what is ‘inspired’?

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 rejected.

Popperian falsificationism

extends the reasoning to experimental sciences

Proof by contradiction . . . ‘extended’. . .

Falsification rule: to what is ‘inspired’?

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 rejected.

Popperian falsificationism

extends the reasoning to experimental sciences

is this extension legitimate?

Falsificationism? OK, but. . .

- What shall we do of all hypotheses not yet falsified? (Limbus? How should we progress?)

Falsificationism? OK, but. . .

- What shall we do of all hypotheses not yet falsified? (Limbus? How should we progress?)
- What to do is nothing of what can be observed is incompatible with the hypothesis (or with many hypotheses)?

Falsificationism? OK, but. . .

- What shall we do of all hypotheses not yet falsified? (Limbus? How should we progress?)
- 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 $\{\mu_i, \sigma_i\}$ (i.e. $\forall H_i$), all values of x from $-\infty$ to $+\infty$ are possible.

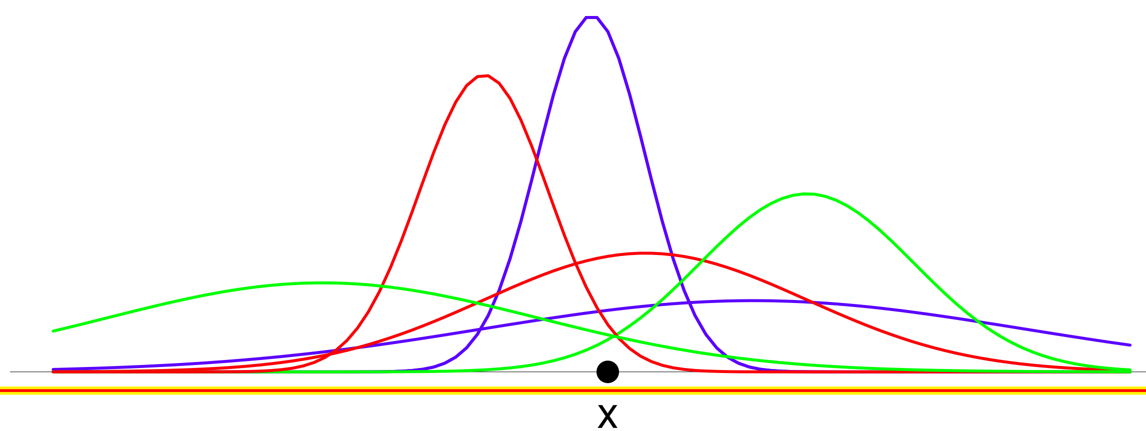
Falsificationism? OK, but. . .

- What shall we do of all hypotheses not yet falsified? (Limbus? How should we progress?)
- 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 $\{\mu_i, \sigma_i\}$ (i.e. $\forall H_i$), all values of x from $-\infty$ to $+\infty$ are possible.

  $\Rightarrow$ Having observed any value of x, none of H_i can be, strictly speaking, falsified.



Falsificationism in action. . .

Obviously, this does not means that falsificationism never works,

Falsificationism in action. . .

Obviously, this does not means that falsificationism never works, as long as no stochastic processes are involved (randomness inherent to the physical processes, or due to ‘errors’ in measurement).

Falsificationism in action. . .

Obviously, this does not means that falsificationism never works, as long as no stochastic processes are involved (randomness inherent to the physical processes, or due to ‘errors’ in measurement).

⇒ Practically never in the experimental sciences!

Falsificationism in action. . .

Obviously, this does not means that falsificationism never works, as long as no stochastic processes are involved (randomness inherent to the physical processes, or due to 'errors' in measurement). Certainly it works against itself:

- Science proceeds, in practice, rather differently:

  The natural development of Science shows that researches are carried along the directions that seem more *credibile* (and hopefully fruitful) at a given moment. A behaviour *"179 degrees or so out of phase from Popper's idea that we make progress by falsificating theories"* (Wilczek, `http://arxiv.org/abs/physics/0403115`)

Falsificationism in action. . .

Obviously, this does not means that falsificationism never works, as long as no stochastic processes are involved (randomness inherent to the physical processes, or due to 'errors' in measurement). Certainly it works against itself:

$\Rightarrow$ logically speaking, falsificationism has to be considered . . . falsified!

Falsificationism and statistics

. . . then, statisticians have invented the “hypothesis tests”

Falsificationism and statistics

... then, statisticians have invented the "hypothesis tests", in which the impossible is replaced by the improbable!

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.

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 dangerous because is deeply rooted in most scientists, due to education and custom, although 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.*]

In summary

A) **if** $C_i \not\rightarrow E$, 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 — OK

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

B) ~~**if** $C_i \xrightarrow[\text{small probability}]{} E$, and **we observe** E~~ — NO

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

~~"most likely false"~~

But it is behind the rational behind the statistical hypothesis tests!

Example

An Italian citizen is chosen at random and sent to take an AIDS test (test is not perfect, as it is the case in practice). *Simplified model*:

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

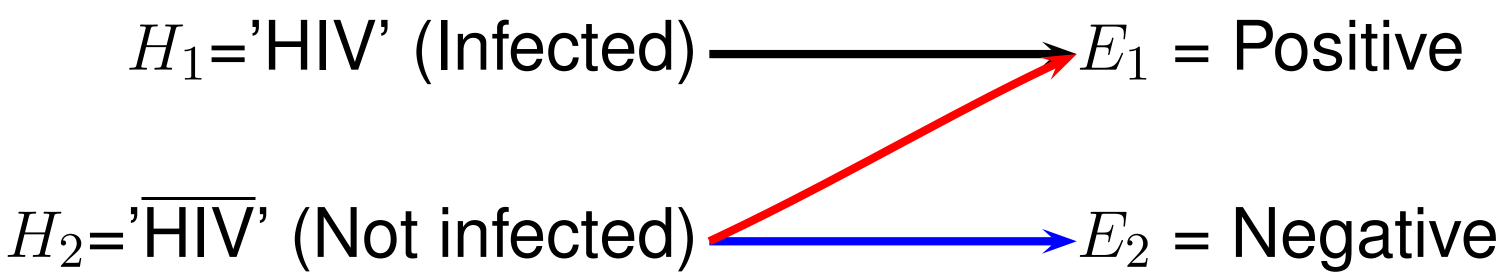
H_1='HIV' (Infected) E_1 = Positive

H_2='$\overline{\text{HIV}}$' (Not infected) E_2 = Negative

Example

An Italian citizen is chosen at random and sent to take an AIDS test (test is not perfect, as it is the case in practice).
Simplified model:

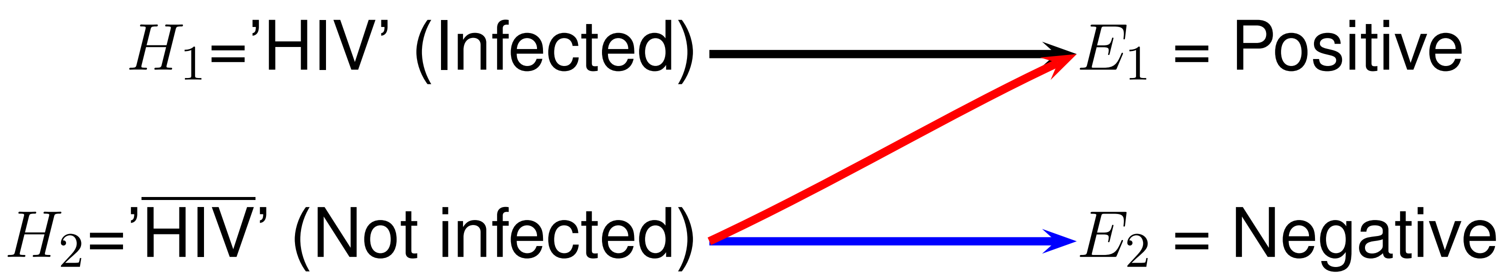
$$P(\text{Pos}\,|\,\text{HIV}) = 100\%$$
$$P(\text{Pos}\,|\,\overline{\text{HIV}}) = 0.2\%$$
$$P(\text{Neg}\,|\,\overline{\text{HIV}}) = 99.8\%$$



Example

An Italian citizen is chosen at random and sent to take an AIDS test (test is not perfect, as it is the case in practice). *Simplified model*:

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



Result: $\Rightarrow$ Positive

Example

An Italian citizen is chosen at random and sent to take an AIDS test (test is not perfect, as it is the case in practice). *Simplified model*:

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

? H_1='HIV' (Infected) ← E_1 = Positive

? H_2='$\overline{\text{HIV}}$' (Not infected) E_2 = Negative

Result: ⇒ Positive

HIV or not HIV?

What shall we conclude?

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"?

What shall we conclude?

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” ?

What shall we conclude?

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”?

What shall we conclude?

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”
- “Hypothesis H_1=Healthy is ruled out with 99.8% C.L.”

?

What shall we conclude?

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"~~
- ~~"Hypothesis H_1=Healthy is ruled out with 99.8% C.L."~~

? NO

Instead, $P(\text{HIV}\,|\,\text{Pos, randomly chosen Italian}) \approx 45\%$

Think about it (a crucial information is missing!)

What shall we conclude?

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"~~
- ~~"Hypothesis H_1=Healthy is ruled out with 99.8% C.L."~~

?   NO

Instead, $P(\text{HIV}\,|\,\text{Pos, randomly chosen Italian}) \approx 45\%$

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

$P(A \,|\, B) \leftrightarrow P(B \,|\, A)$

Pay attention no to arbitrary revert conditional probabilities:

In general $P(A \,|\, B) \neq P(B \,|\, A)$

$P(A\,|\,B) \leftrightarrow P(B\,|\,A)$

Pay attention no to arbitrary revert conditional probabilities:

In general $P(A\,|\,B) \neq P(B\,|\,A)$

- $P(\text{Positive}\,|\,\overline{HIV}) \neq P(\overline{HIV}\,|\,\text{Positive})$

$P(A \,|\, B) \leftrightarrow P(B \,|\, A)$

Pay attention no to arbitrary revert conditional probabilities:

In general $P(A \,|\, B) \neq P(B \,|\, A)$

- $P(\text{Positive} \,|\, \overline{HIV}) \neq P(\overline{HIV} \,|\, \text{Positive})$
- $P(\text{Win} \,|\, \text{Play}) \neq P(\text{Play} \,|\, \text{Win})$ [Lotto]

$P(A\,|\,B) \leftrightarrow P(B\,|\,A)$

Pay attention no to arbitrary revert conditional probabilities:

In general $P(A\,|\,B) \neq P(B\,|\,A)$

- $P(\text{Positive}\,|\,\overline{HIV}) \neq P(\overline{HIV}\,|\,\text{Positive})$
- $P(\text{Win}\,|\,\text{Play}) \neq P(\text{Play}\,|\,\text{Win})$ [Lotto]
- $P(\text{Pregnant}\,|\,\text{Woman}) \neq P(\text{Woman}\,|\,\text{Pregnant})$

$P(A \,|\, B) \leftrightarrow P(B \,|\, A)$

Pay attention no to arbitrary revert conditional probabilities:

In general $P(A \,|\, B) \neq P(B \,|\, A)$

- $P(\text{Positive} \,|\, \overline{HIV}) \neq P(\overline{HIV} \,|\, \text{Positive})$
- $P(\text{Win} \,|\, \text{Play}) \neq P(\text{Play} \,|\, \text{Win})$ [Lotto]
- $P(\text{Pregnant} \,|\, \text{Woman}) \neq P(\text{Woman} \,|\, \text{Pregnant})$

In particular

- A cause might produce a given effect with very low probability, and nevertheless could be the most probable cause of that effect, often the only one!

‘Low probability’ events

Tipical values of statistical practice to reject a hypothesis are 5%, 1%, . . . (see ’AIDS test’)

‘Low probability’ events

Tipical values of statistical practice to reject a hypothesis are 5%, 1%, . . . (see ’AIDS test’)

BUT the greatest majority of the events of interest have very low probability (before occurring!).

‘Low probability’ events

Tipical values of statistical practice to reject a hypothesis are 5%, 1%, . . . (see ’AIDS test’)

BUT the greatest majority of the events of interest have very low probability (before occurring!).

For example, imagine a Gaussian random generator (H_0, with $\mu = 3, \sigma = 1$) gives us $X = 3.1416$.

‘Low probability’ events

Tipical values of statistical practice to reject a hypothesis are 5%, 1%, ... (see ’AIDS test’)

BUT the greatest majority of the events of interest have very low probability (before occurring!).

For example, imagine a Gaussian random generator (H_0, with $\mu = 3, \sigma = 1$) gives us $X = 3.1416$.

→ What was the probability to give exactly that number?:

$$\begin{aligned} P(X = 3.1416 \,|\, H_0) &= \int_{3.14155}^{3.14165} f_{\mathcal{G}}(x \,|\, \mu, \sigma) dx \\ &\approx f_{\mathcal{G}}(3.1416 \,|\, \mu, \sigma) \times \Delta x \\ &\approx f_{\mathcal{G}}(3.1416 \,|\, \mu, \sigma) \times 0.0001 \\ &\approx 39 \times 10^{-6} \end{aligned}$$

‘Low probability’ events

Tipical values of statistical practice to reject a hypothesis are 5%, 1%, . . . (see ’AIDS test’)

BUT the greatest majority of the events of interest have very low probability (before occurring!).

For example, imagine a Gaussian random generator (H_0, with $\mu = 3, \sigma = 1$) gives us $X = 3.1416$.

→ What is the probability that X comes from H_0?

‘Low probability’ events

Tipical values of statistical practice to reject a hypothesis are 5%, 1%, ... (see ’AIDS test’)

BUT the greatest majority of the events of interest have very low probability (before occurring!).

For example, imagine a Gaussian random generator (H_0, with $\mu = 3, \sigma = 1$) gives us $X = 3.1416$.

$\rightarrow$ What is the probability that X comes from H_0?

- Certainly NOT $\approx 39 \times 10^{-6}$;

‘Low probability’ events

Tipical values of statistical practice to reject a hypothesis are 5%, 1%, . . . (see ’AIDS test’)

BUT the greatest majority of the events of interest have very low probability (before occurring!).

For example, imagine a Gaussian random generator (H_0, with $\mu = 3, \sigma = 1$) gives us $X = 3.1416$.

→ What is the probability that X comes from H_0?

- Certainly NOT $\approx 39 \times 10^{-6}$;
- Indeed, it is exactly 1, since H_0 is the only cause which can produce that effect:

$$\begin{aligned} P(X = 3.1416 \,|\, H_0) &\approx 39 \times 10^{-6} \\ P(H_0 \,|\, X = 3.1416) &= 1\,. \end{aligned}$$

Probability of something else. . .

Besides the fact that the reasoning based only on the probability of the event given the cause is logically flawed, the ‘techical issue’ of low probability events which would lead to reject any hypothesis forces the statistician to rethink the question. . .

Probability of something else. . .

Besides the fact that the reasoning based only on the probability of the event given the cause is logically flawed, the ‘techical issue’ of low probability events which would lead to reject any hypothesis forces the statistician to rethink the question. . .

> but, instead of repent, throw everything away and finally start to read Laplace (yes, ‘our’ Laplace!)
> ’he’ makes a new invention:

Probability of something else…

Besides the fact that the reasoning based only on the probability of the event given the cause is logically flawed, the ‘techical issue’ of low probability events which would lead to reject any hypothesis forces the statistician to rethink the question…

but, instead of repent, throw everything away and finally start to read Laplace (yes, ‘our’ Laplace!)
’he’ makes a new invention:
$\rightarrow$ what matter is not the probability of the X, but rather the probability of X or of any other less probable number (or a number farther than X from the expected value – the story is a bit longer…):

$$P(X \geq 3.1416) \;=\; \int_{3.14155}^{+\infty} f_{\mathcal{G}}(x\,|\,\mu,\sigma)dx \approx 44\%$$

Probability of something else...

Besides the fact that the reasoning based only on the probability of the event given the cause is logically flawed, the 'techical issue' of low probability events which would lead to reject any hypothesis forces the statistician to rethink the question...

> but, instead of repent, throw everything away and finally start to read Laplace (yes, 'our' Laplace!)
> 'he' makes a new invention:
> $\rightarrow$ what matter is not the probability of the X, but rather the probability of X or of any other less probable number (or a number farther than X from the expected value – the story is a bit longer...):

$$P(X \geq 3.1416) \; [= P(X \geq x_{obs})] \;\; \Rightarrow \;\; \text{`p-value'}$$

Probability of something else...

Besides the fact that the reasoning based only on the probability of the event given the cause is logically flawed, the 'techical issue' of low probability events which would lead to reject any hypothesis forces the statistician to rethink the question...

- ⇒ Magically the result 'becomes' rather probable! Why, we, silly, worried about it?
- ⇒ The statisticians are happy...

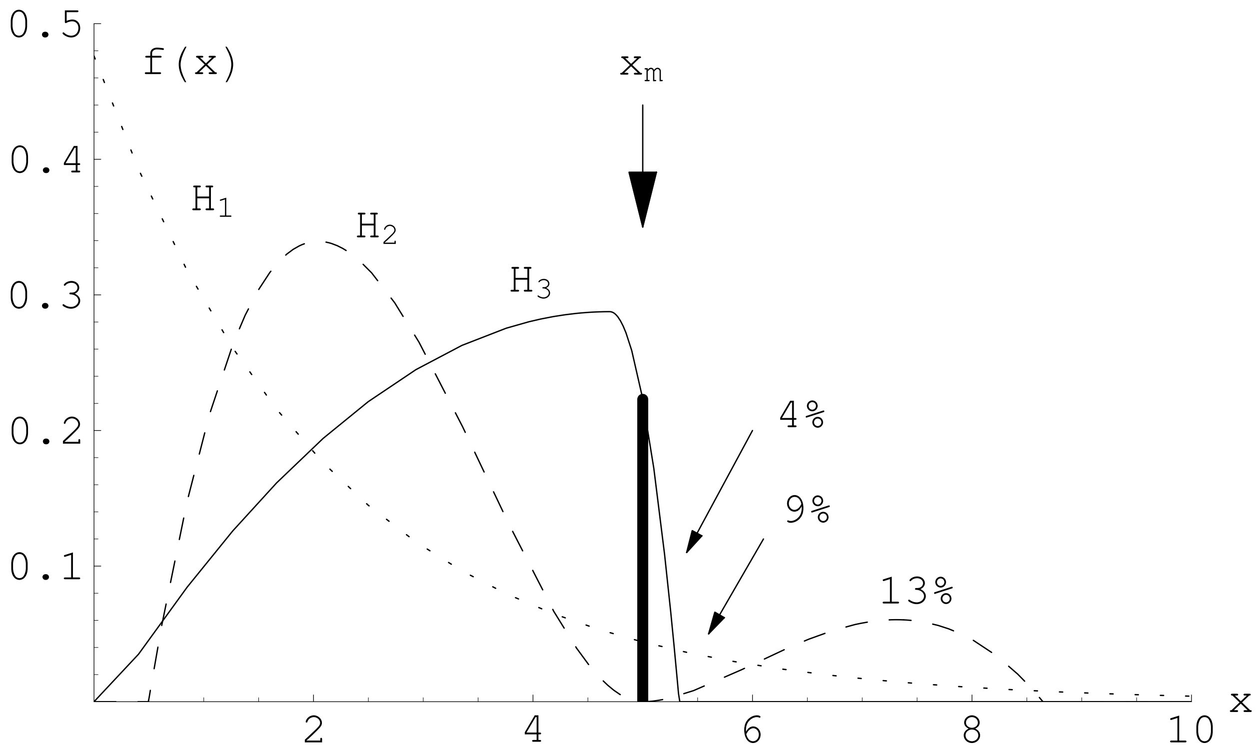
Probability of something else...

Besides the fact that the reasoning based only on the probability of the event given the cause is logically flawed, the ‘techical issue’ of low probability events which would lead to reject any hypothesis forces the statistician to rethink the question...

- ⇒ Magically the result ‘becomes’ rather probable! Why, we, silly, worried about it?
- ⇒ The statisticians are happy... scientists and general public cheated...

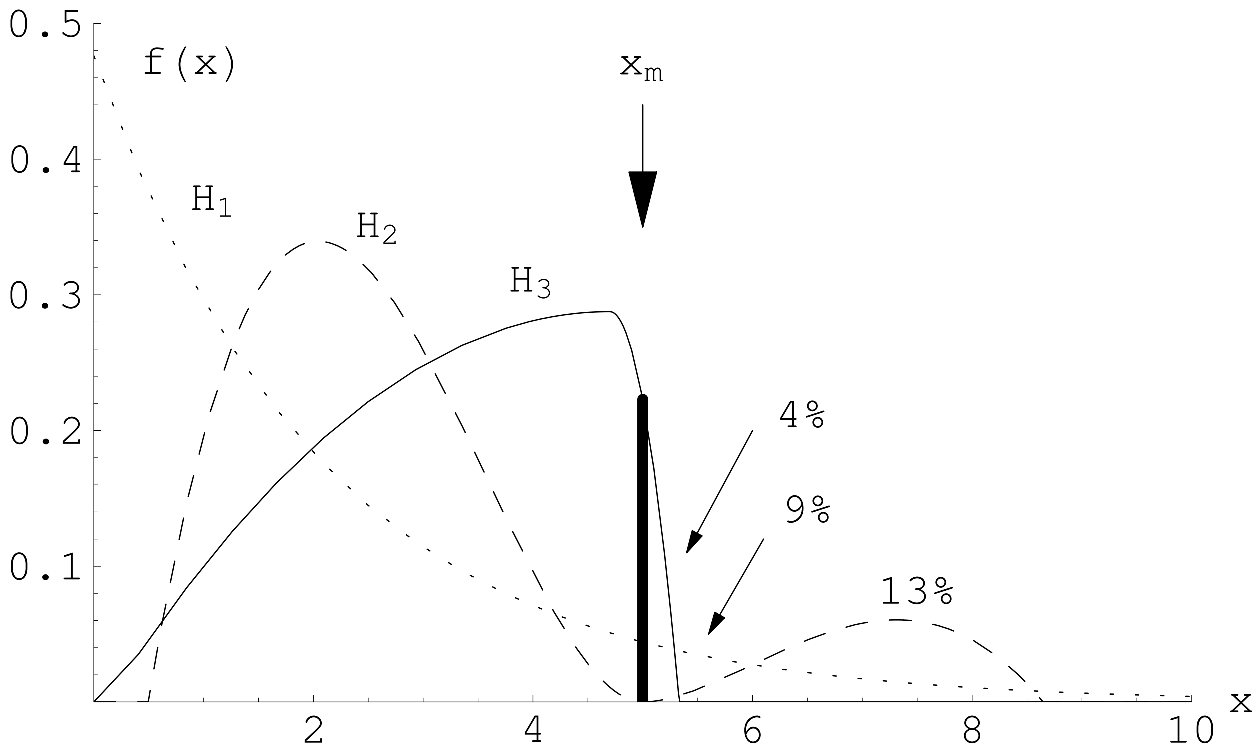
Comparing three hypotheses

Which hypothesis is favored by the experimental observation x_m?



Comparing three hypotheses

Which hypothesis is favored by the experimental observation x_m?

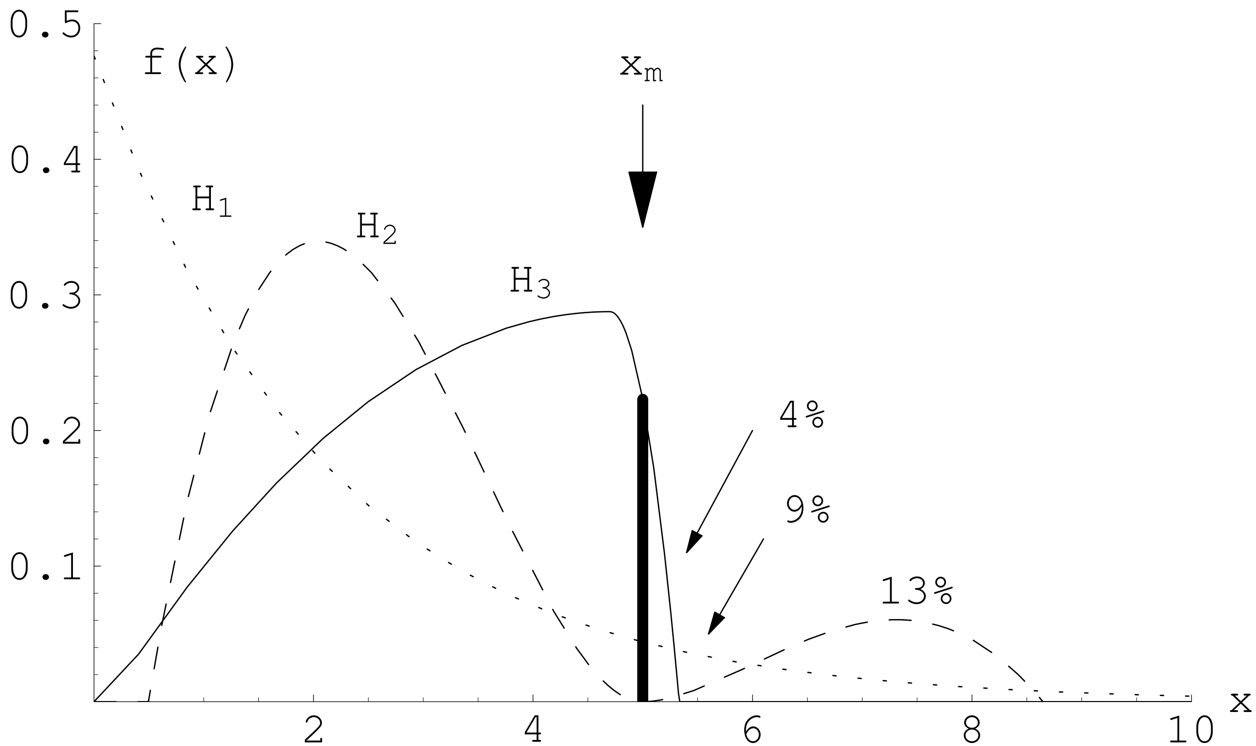


$$P(x_m \,|\, H_3) > P(x_m \,|\, H_1) > P(x_m \,|\, H_2) = 0 \quad (\,!\,)$$

Even if $P(x_m \,|\, H_i) \rightarrow 0$ (it depends on resolution)

Comparing three hypotheses

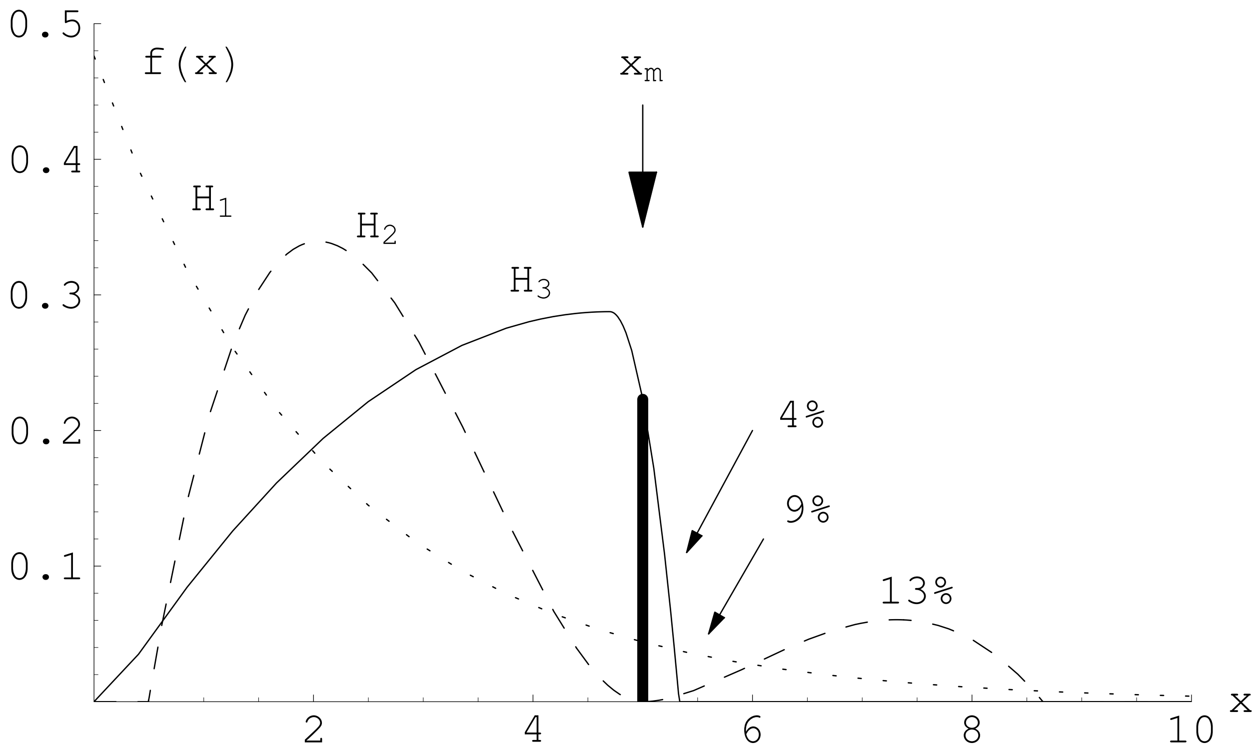
Which hypothesis is favored by the experimental observation x_m?



In particular, the hypothesis H_2 is (<u>truly</u>) falsified (impossible!), although it yields the largest 'p-value'

Comparing three hypotheses

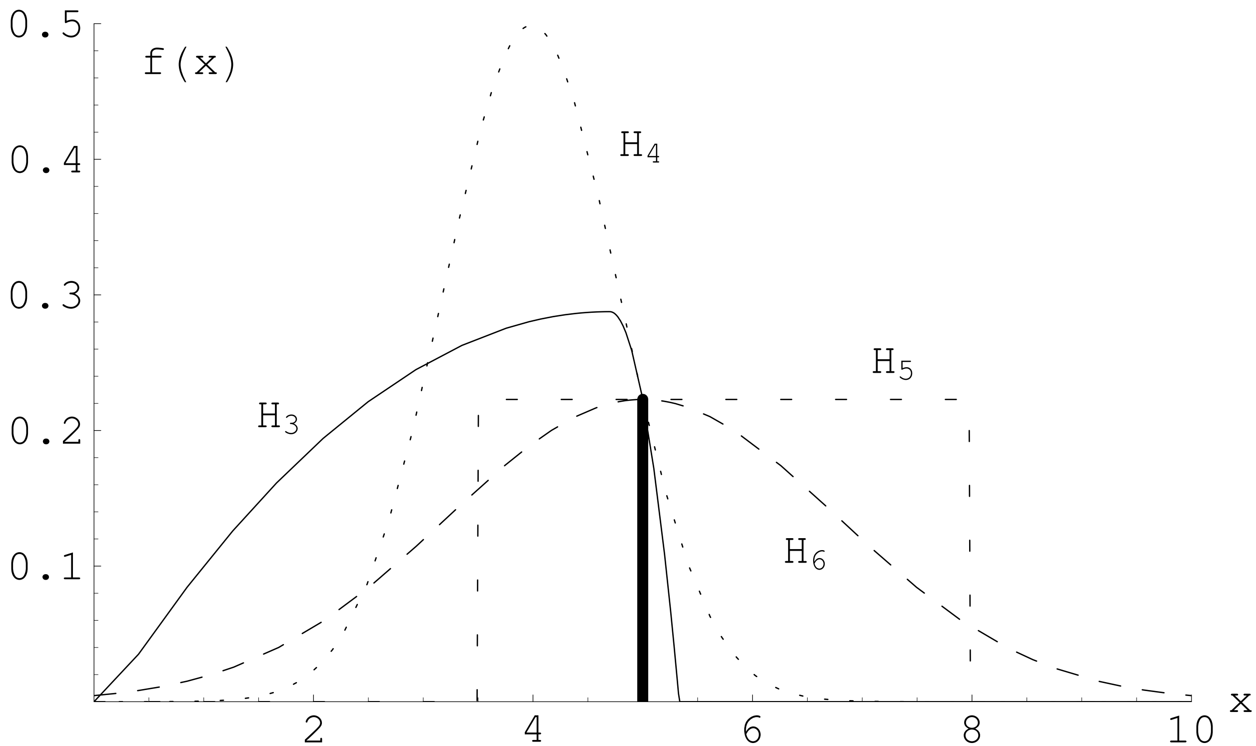
Which hypothesis is favored by the experimental observation x_m?



In particular, the hypothesis H_2 is (truly) falsified (impossible!), although it yields the largest 'p-value', or 'probability of the tail(s)'

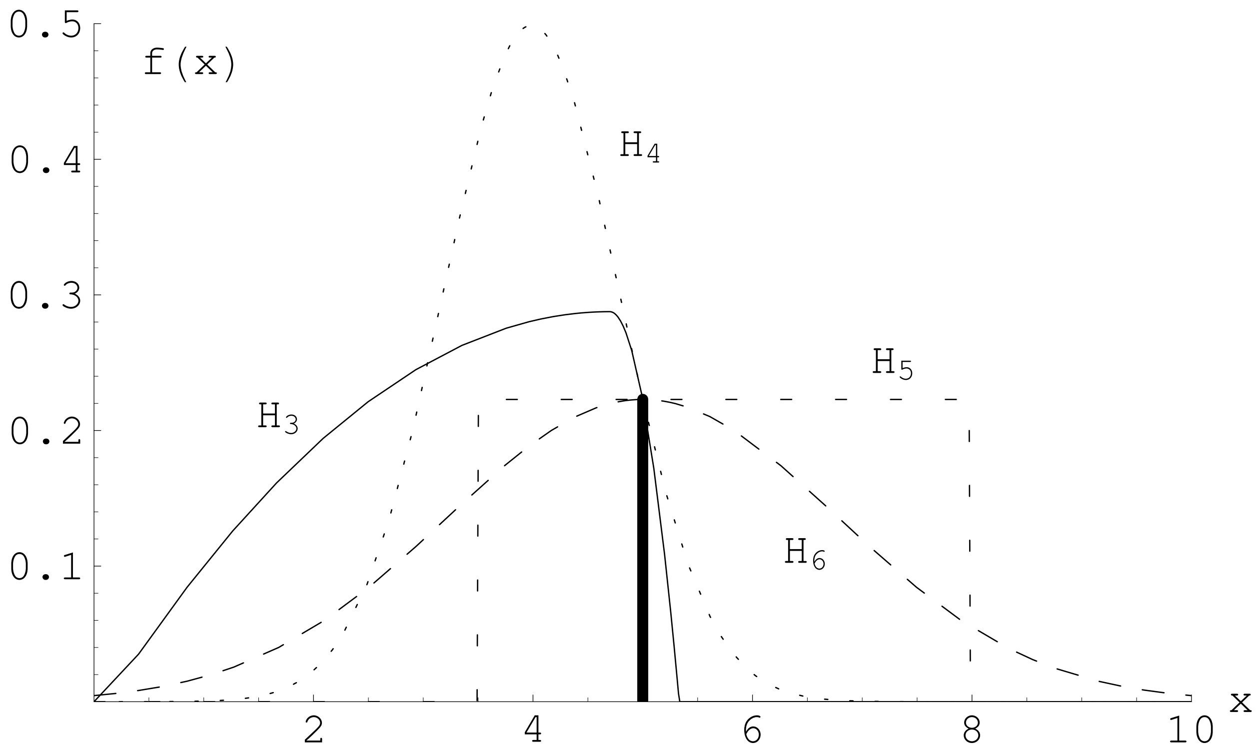
An irrilevant experiment

Which hypothesis is favored by the experimental observation x_m?



An irrilevant experiment

Which hypothesis is favored by the experimental observation x_m?

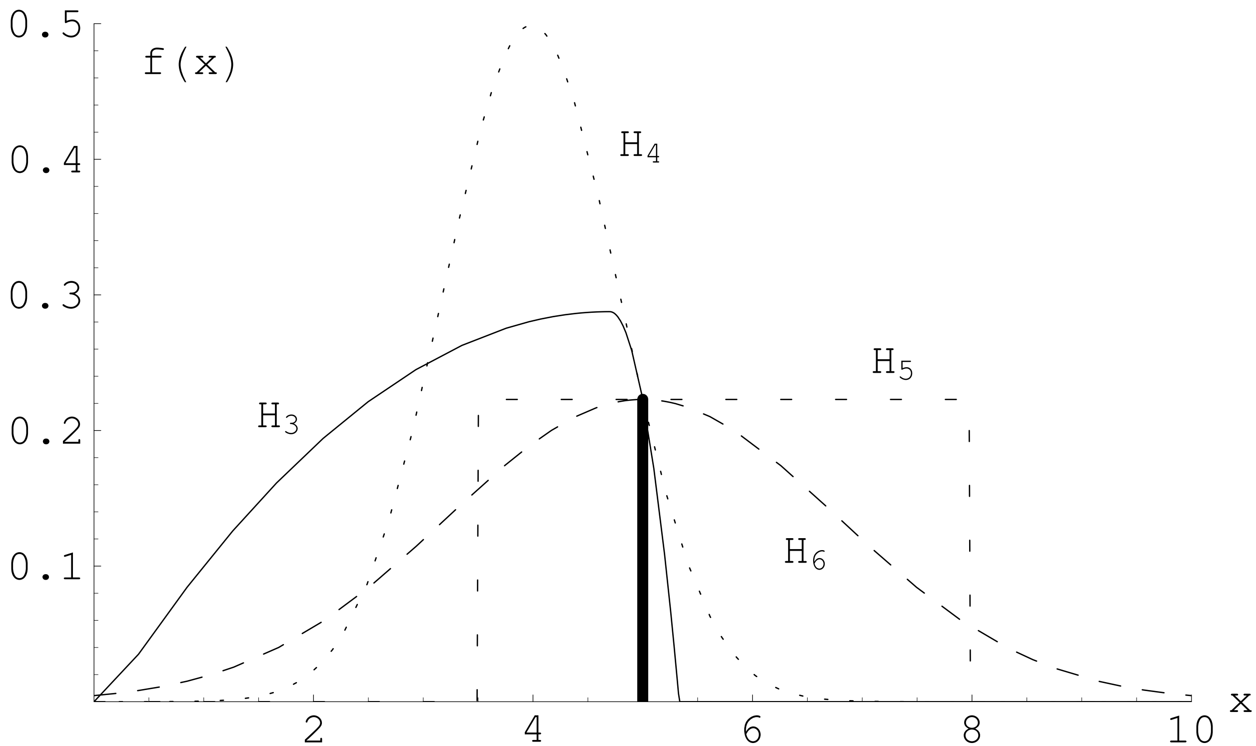


$$P(x_m \,|\, H_3) = P(x_m \,|\, H_4) = P(x_m \,|\, H_5) = P(x_m \,|\, H_6)$$

$\Rightarrow$ *The experimental result is irrelevant!*

An irrilevant experiment

Which hypothesis is favored by the experimental observation x_m?

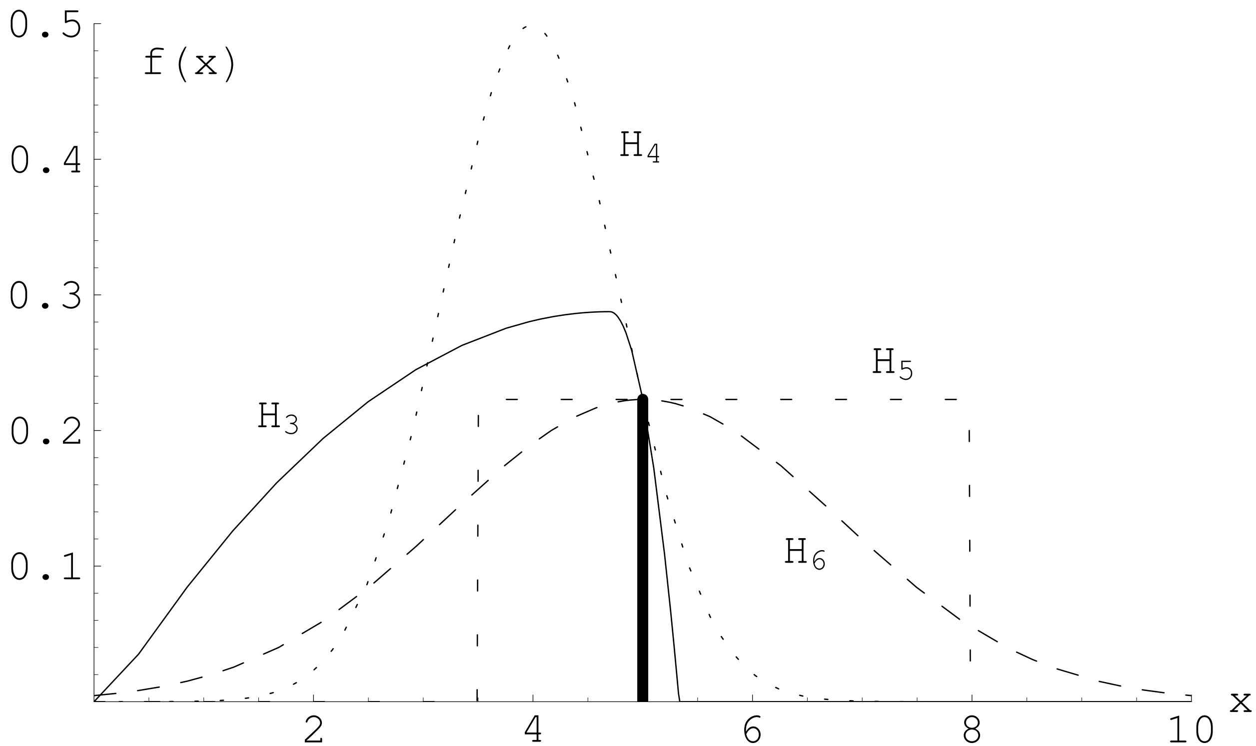


$$P(x_m \,|\, H_3) = P(x_m \,|\, H_4) = P(x_m \,|\, H_5) = P(x_m \,|\, H_6)$$

$\Rightarrow$ *The experimental result is irrelevant!*
$\rightarrow$ we mantain our opinions about H_i

An irrilevant experiment

Which hypothesis is favored by the experimental observation x_m?



$$P(x_m \,|\, H_3) = P(x_m \,|\, H_4) = P(x_m \,|\, H_5) = P(x_m \,|\, H_6)$$

⇒ *The experimental result is irrelevant!*

⇒ *...no matter what the different the p-values are!*

Which p-value?. . .

'p-value' = 'probability of the tail(s)'

Which p-value?. . .

’p-value’ = ‘probability of the tail(s)’

Of what?

Which p-value?...

'p-value' = 'probability of the tail(s)'

Of what?

$\longrightarrow$ the test variable ('θ') is absolutely arbitrary:

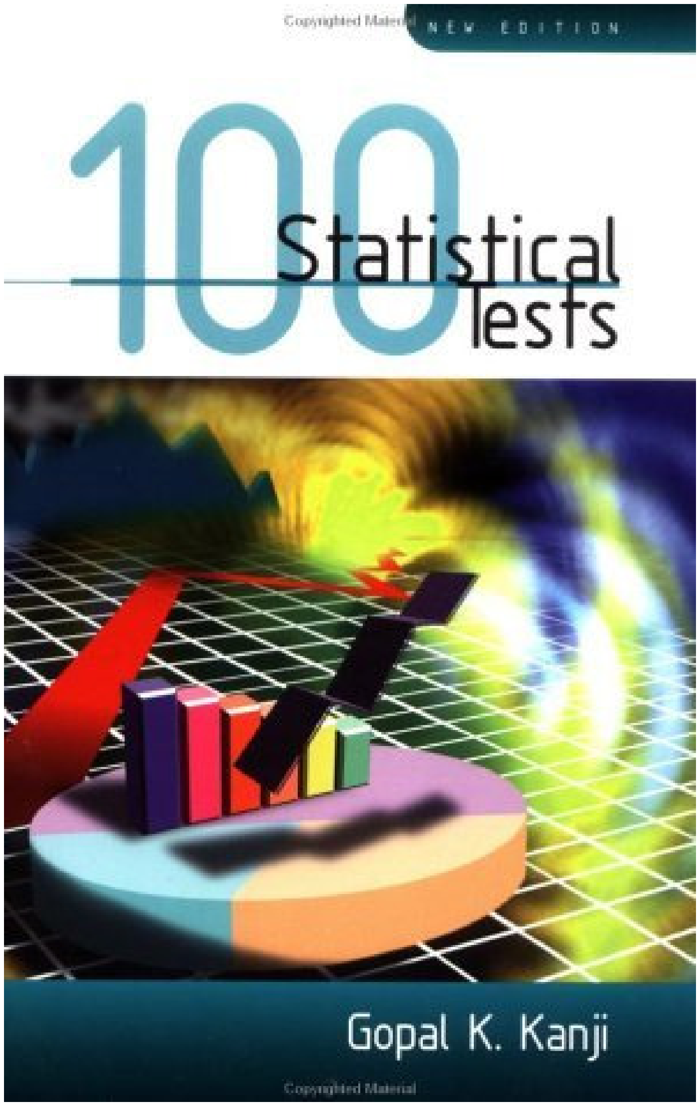
$$\theta = \theta(\mathbf{x})$$

$$\rightarrow f(\theta) \quad \text{[p.d.f]}$$

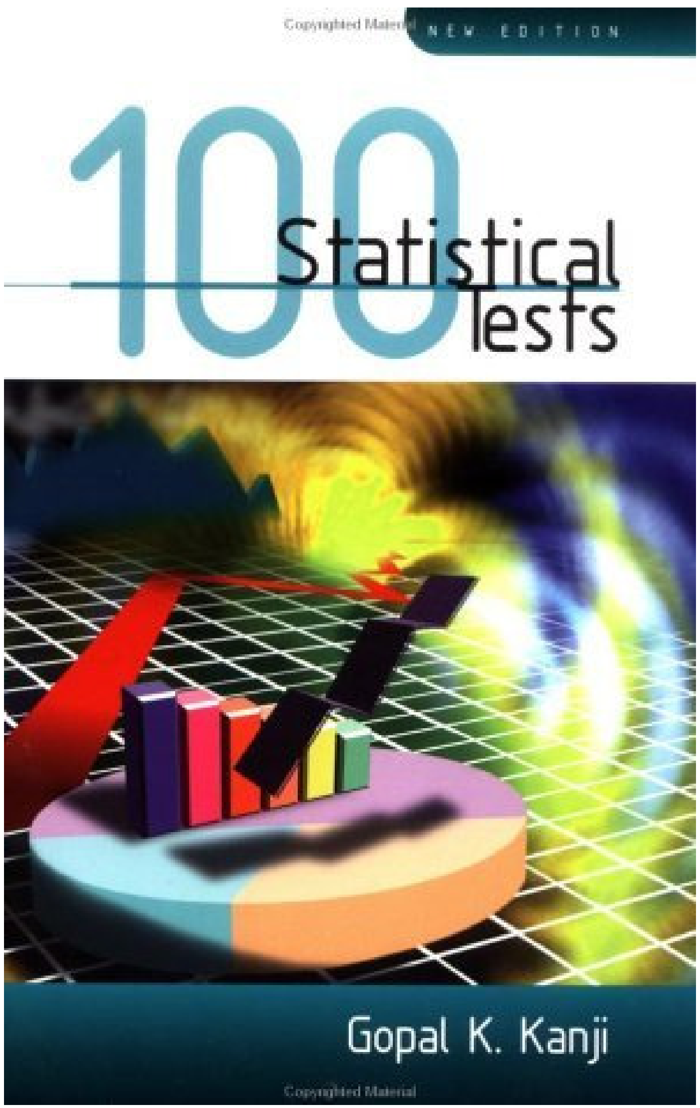
Experiment: $\rightarrow \theta_{mis} = \theta(\mathbf{x}_{mis})$

p-value $= P(\theta \geq \theta_{mis})$ ('one tail')

Which p-value?...

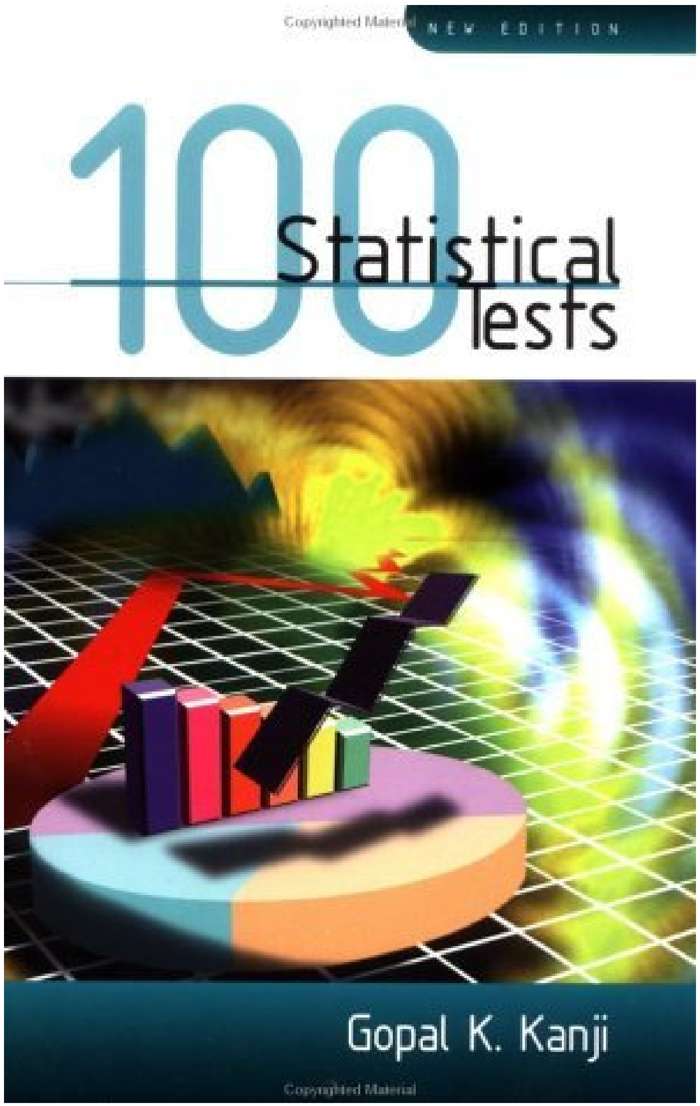


Which p-value?. . .



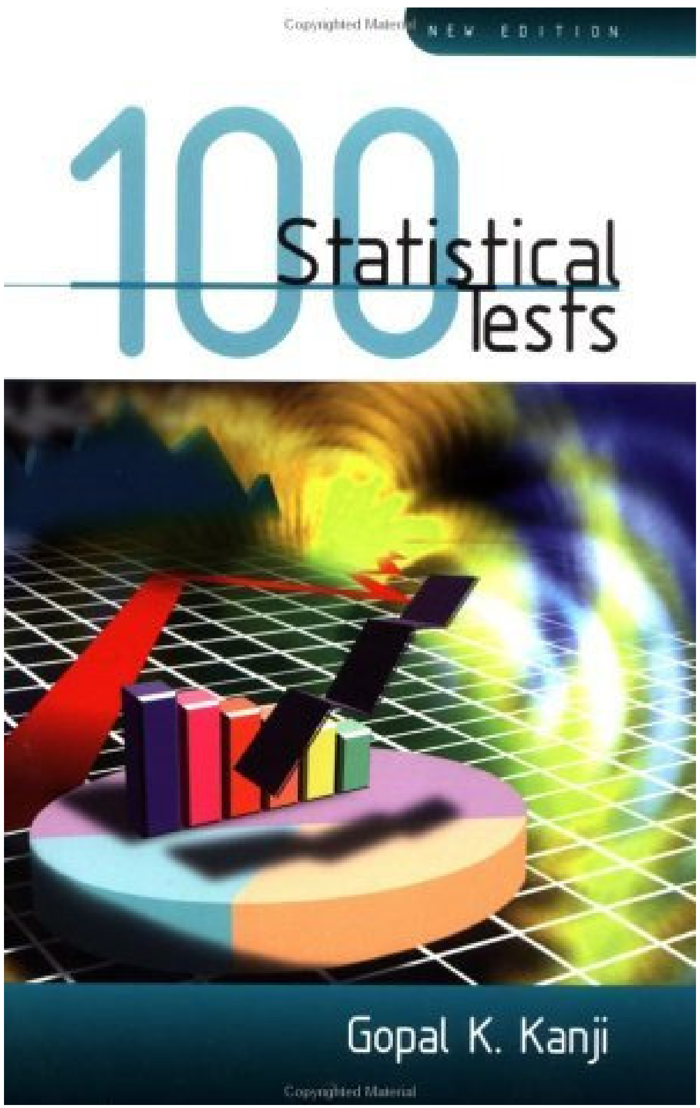
- far from exhaustive list,

Which p-value?. . .



- far from exhaustive list,
- with arbitrary variants:

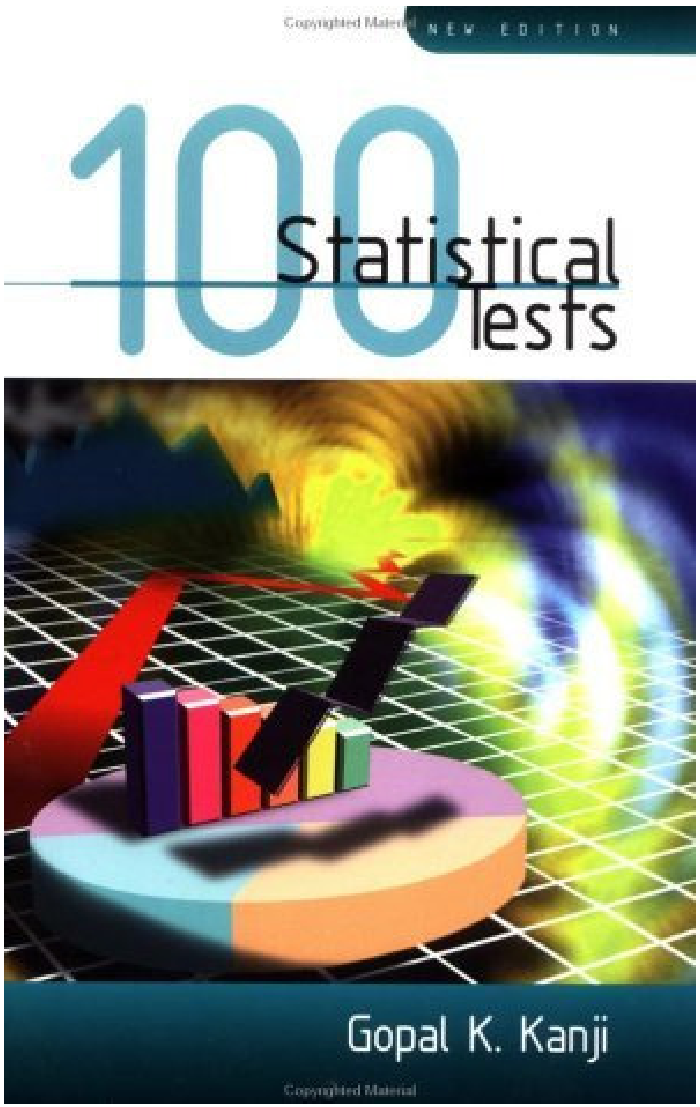
Which p-value?. . .



- far from exhaustive list,
- with arbitrary variants:

$\Rightarrow$ practitioner chose the one that provide the result they like better:
$\rightarrow$ *like if you go around until "someone agrees with you"*

Which p-value?...



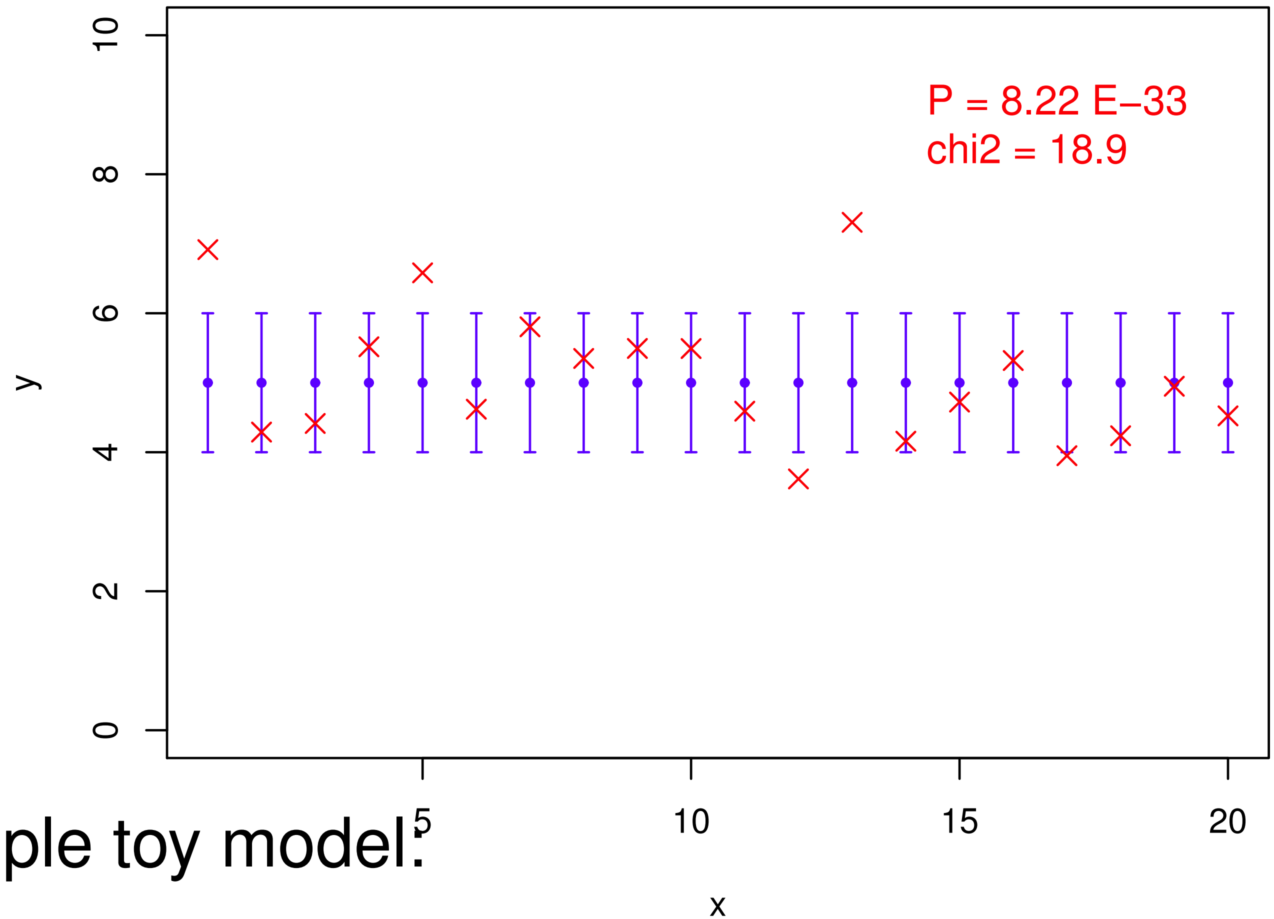
- far from exhaustive list,
- with arbitrary variants:

⇒ practitioner chose the one that provide the result they like better:
→ *like if you go around until "someone agrees with you"*

- personal **'golden rule'**: "the more exotic is the name of the test, the less I believe the result", because I'm pretty shure that several 'normal' tests have been descarded in the meanwhile...

χ^2... the mother of all p-values

Theory Vs experiment *(bars: expectation uncertainty)*:



Very simple toy model:

- True value of y: 5, independently of x (a.u.);
- Gaussian instrumental error with $\sigma = 1$.

Probability of the data sample

$P = 8.22 \times 10^{-33}$ is the probability of the 'configuration' of experimental points:

- obtained multiplying the probability of each point (independent measurements):

$$P = \prod_i P_i$$

where

$$P_i = \int_{y_{m_i}-\Delta y/2}^{y_{m_i}+\Delta y/2} f(y)dy$$

- as seen, P_i depends on the 'resoluzion' Δy (instrumental 'discretization'):

$$\rightarrow \text{we use} \quad \Delta y = \frac{1}{10}\,\sigma$$

‘Distance’ Experiment-theory: χ^2

The costruction of the χ^2 is very popular (usually in first lab. courses – ‘Fisichetta’):

$$\chi^2 = \sum_i \left(\frac{y_{m_i} - y_{th_i}}{\sigma_i}\right)^2 \rightarrow \sum_i \left(\frac{y_{m_i} - y_0}{\sigma}\right)^2$$

$$\begin{aligned}
\chi^2 &\sim \Gamma(\nu/2, 1/2) && [\rightarrow \nu = 20] \\
\text{E}[\chi^2] &= \nu && [\rightarrow 20] \\
\text{Var}[\chi^2] &= 2\nu && [\rightarrow 40] \\
\text{Std}[\chi^2] &= \sqrt{2\nu} && [\rightarrow 6.3] \\
\Rightarrow & \quad \boxed{\chi^2 = 20 \pm 6}
\end{aligned}$$

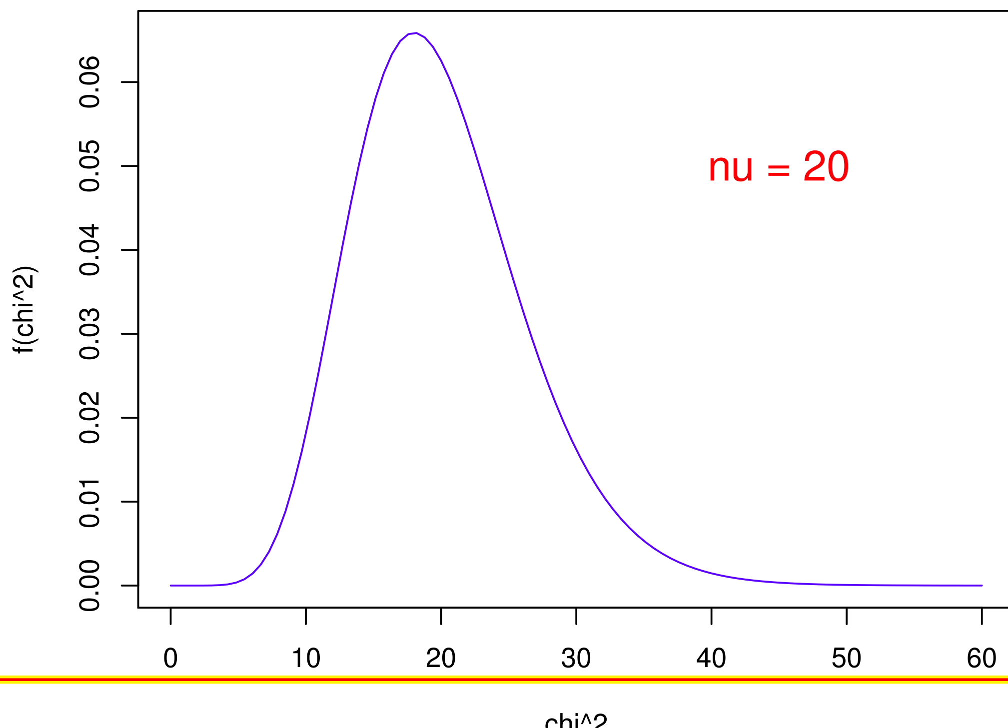
Our expectations about χ^2

$$\mathrm{E}[\chi^2] = \nu \qquad [\to 20]$$

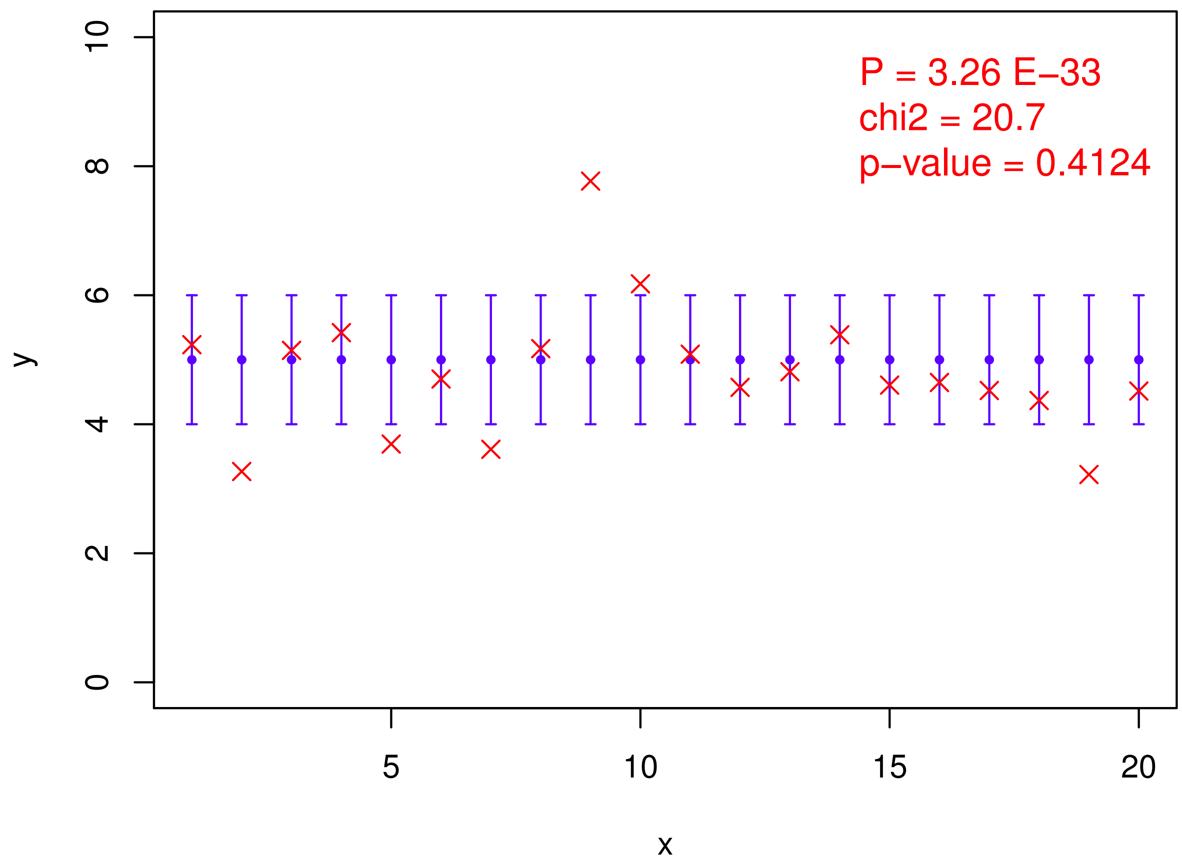
$$\mathrm{Std}[\chi^2] = \sqrt{2\nu} \qquad [\to 6.3]$$

$$\Rightarrow \quad \boxed{\chi^2 = 20 \pm 6}$$

[mode: 18]



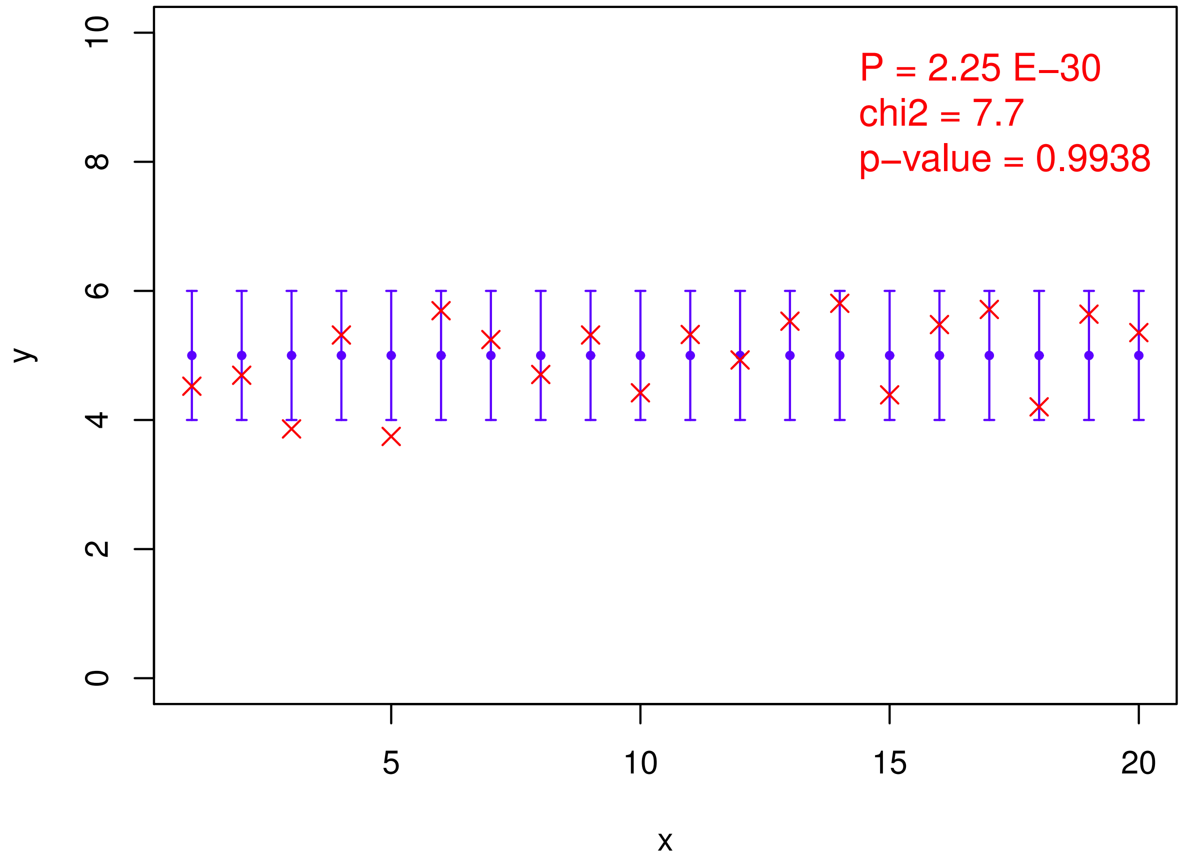
Some examples



In the average.

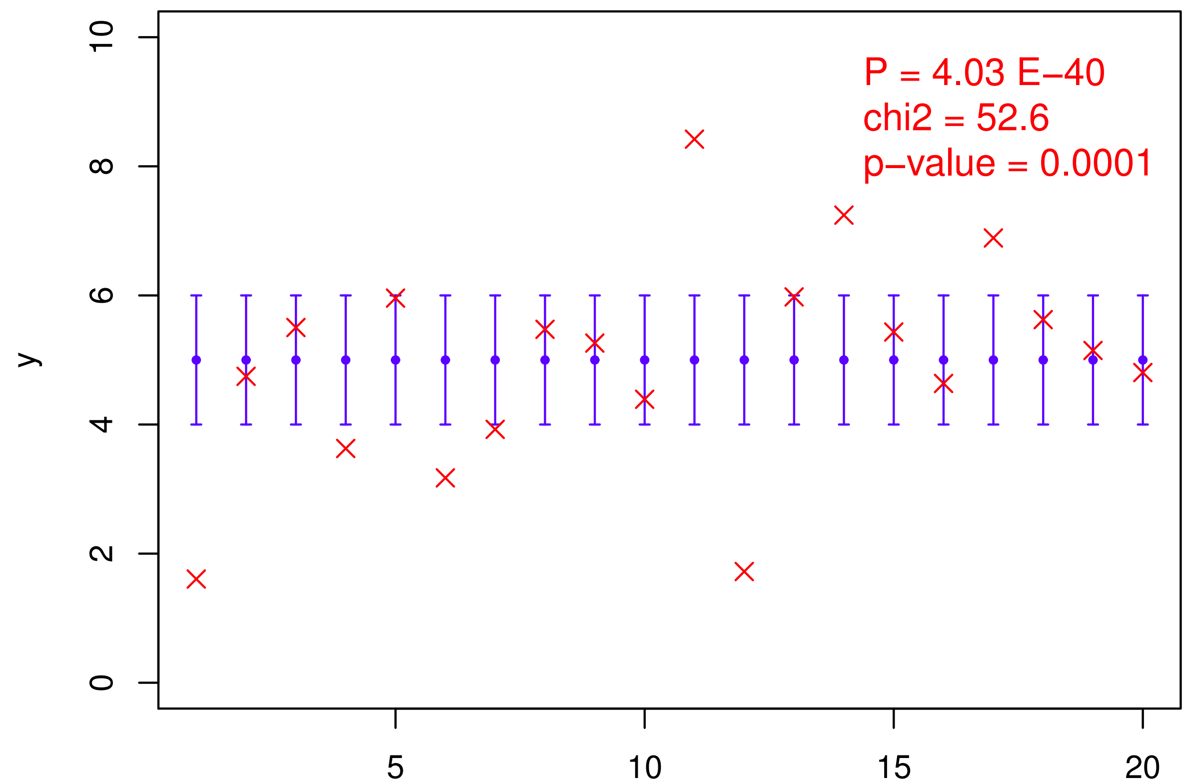
(but someone could see the points forming a 'constellation'. . .)

Some examples



Too good?

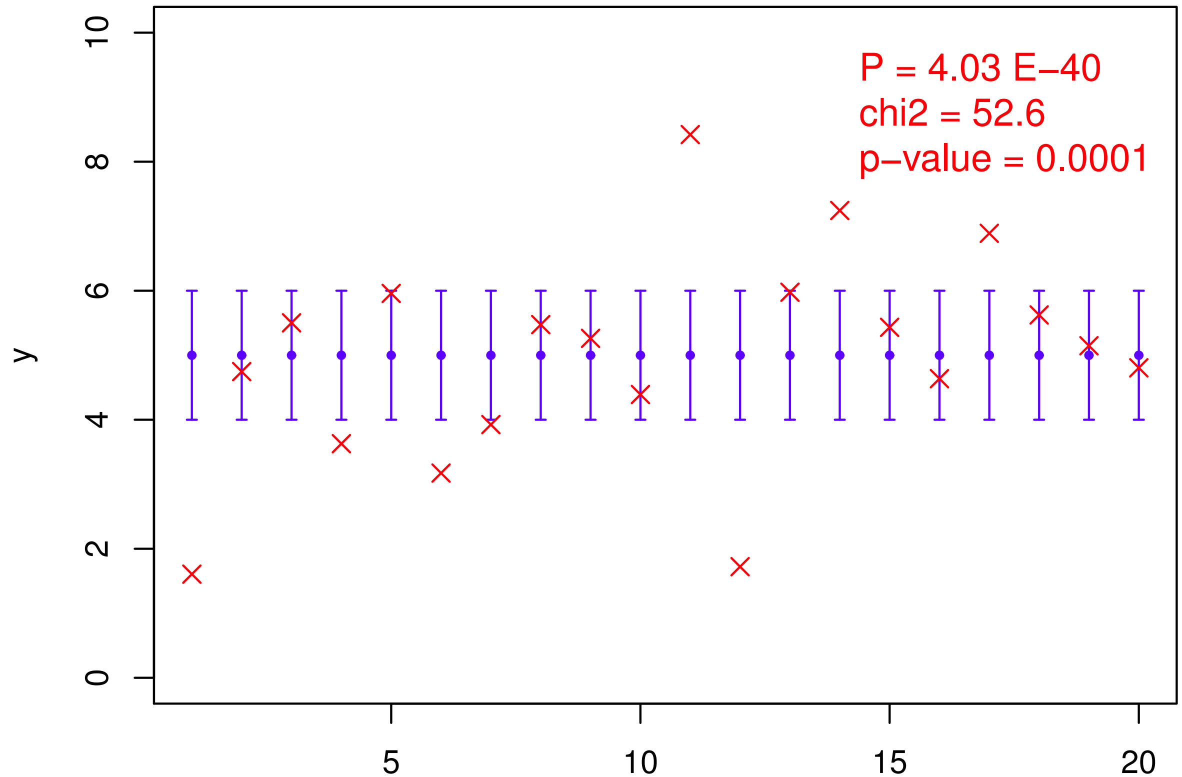
Some examples



$\chi^2 = 52.6$, with a p-value $= 0.93 \times 10^{-4}$

At limit?

Some examples

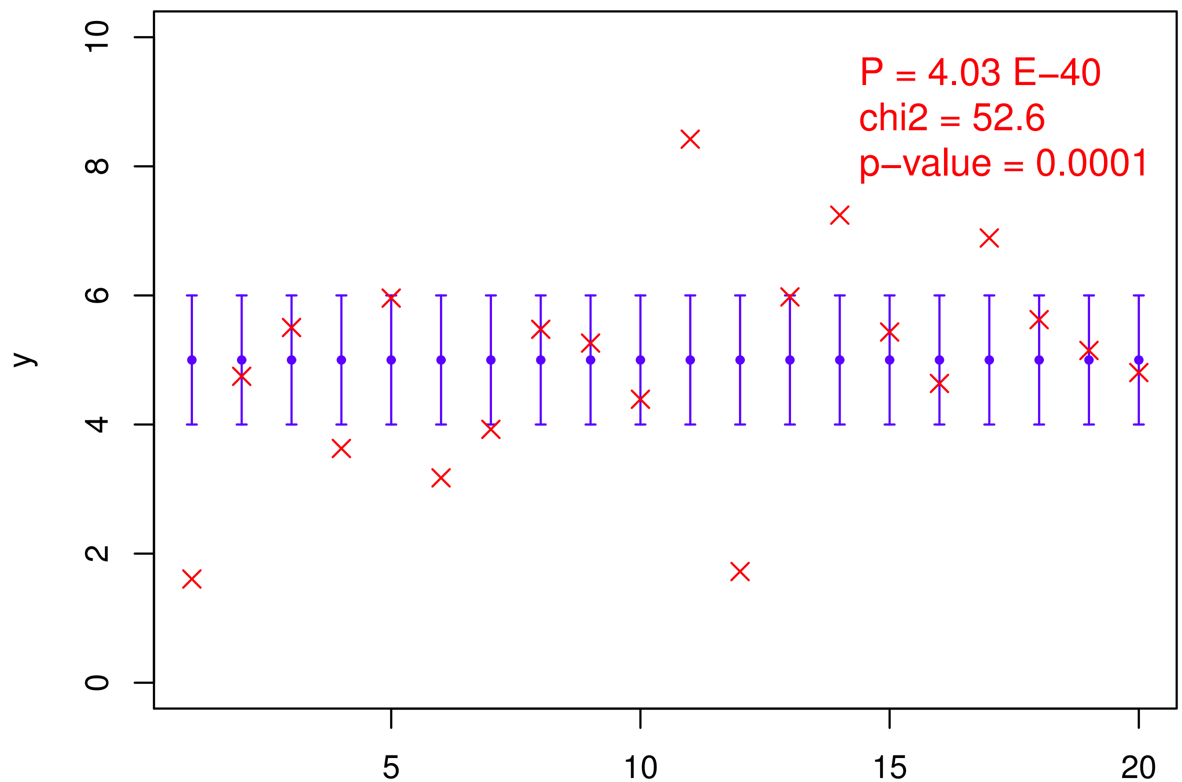


$\chi^2 = 52.6$, with a p-value $= 0.93 \times 10^{-4}$

At limit? Just come out at the first time (October 9, 13:01)

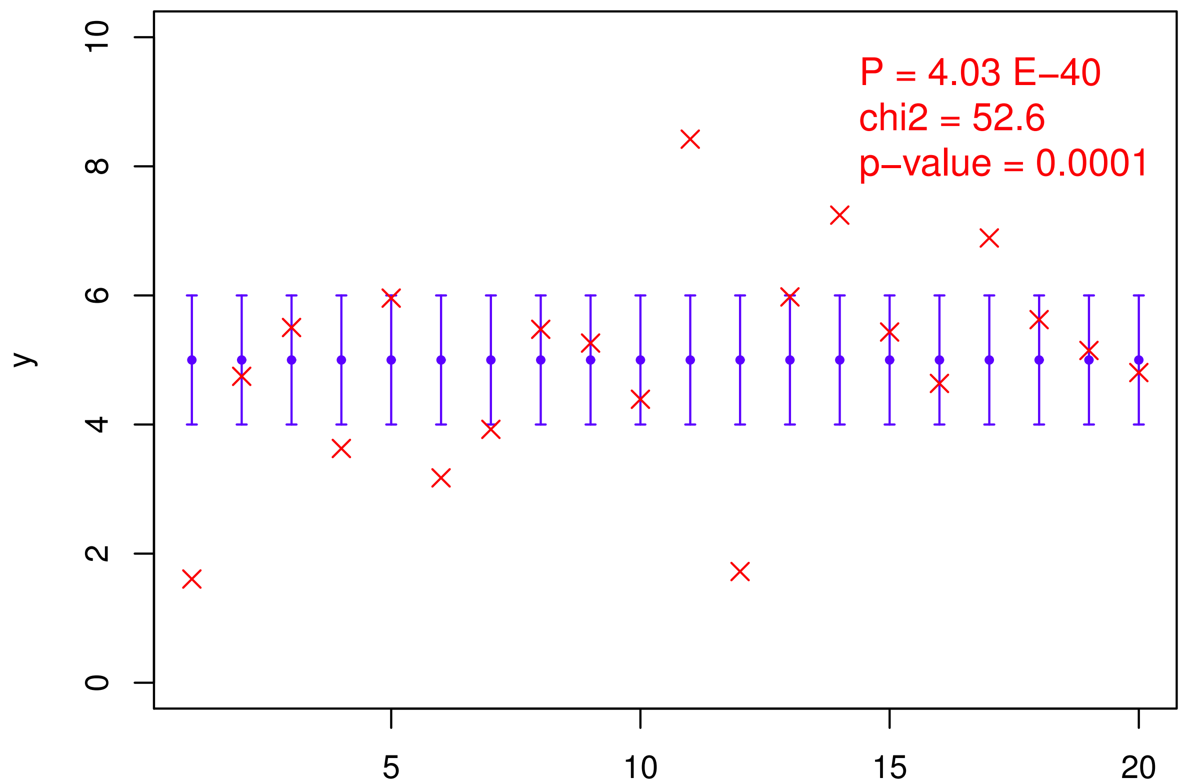
```
while(chi2.ym() < 38) source("chi2_1.R")
```

Some examples



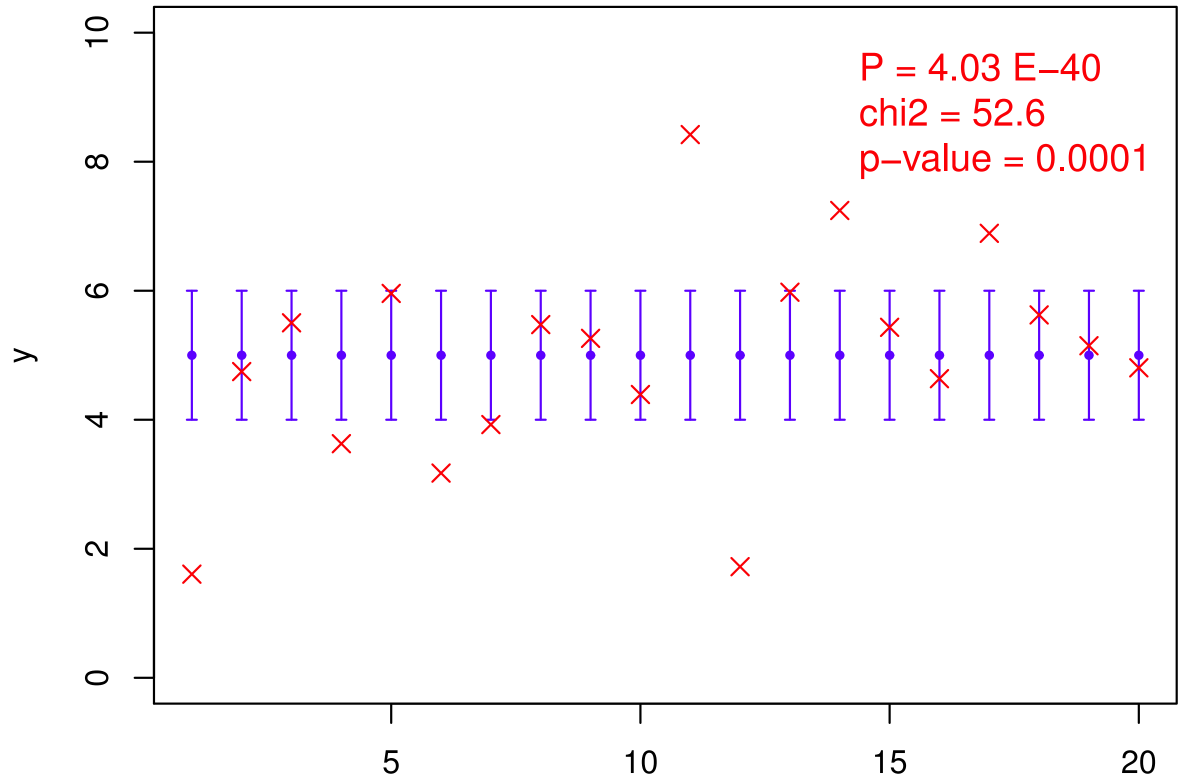
Note: χ^2_{mis} 52.6 is 5.1σ from its expectation [$\frac{52.6-20}{\sqrt{40}} = 5.1$]

Some examples



Note: χ^2_{mis} 52.6 is 5.1σ from its expectation [$\frac{52.6-20}{\sqrt{40}} = 5.1$] , but the p-value is comunicated as "3.7 σ", referring to the probability of the tail above 3.7 σ of an 'equivalent Gaussian'.

Some examples

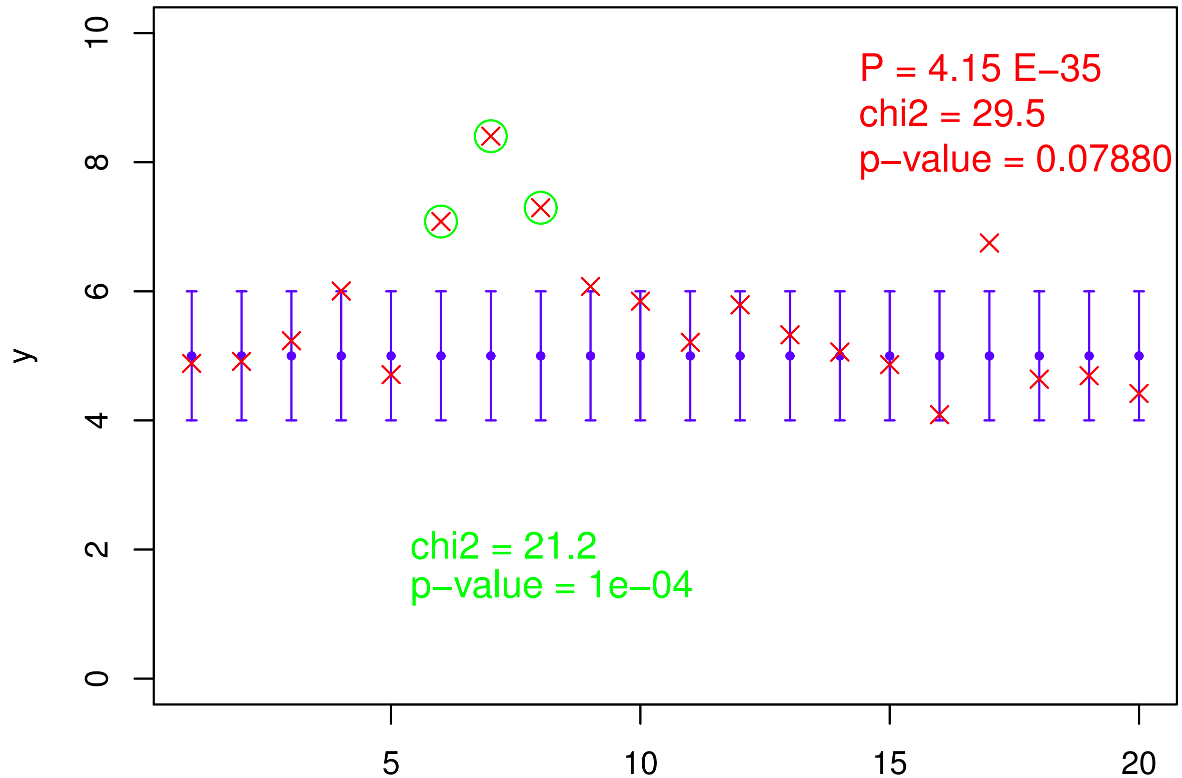


Note: χ^2_{mis} 52.6 is 5.1σ from its expectation [$\frac{52.6-20}{\sqrt{40}} = 5.1$] , but the p-value is comunicated as "3.7 σ", referring to the probability of the tail above 3.7 σ of an 'equivalent Gaussian'.

(as if there were already not enough confusion. . .)

The art of χ^2

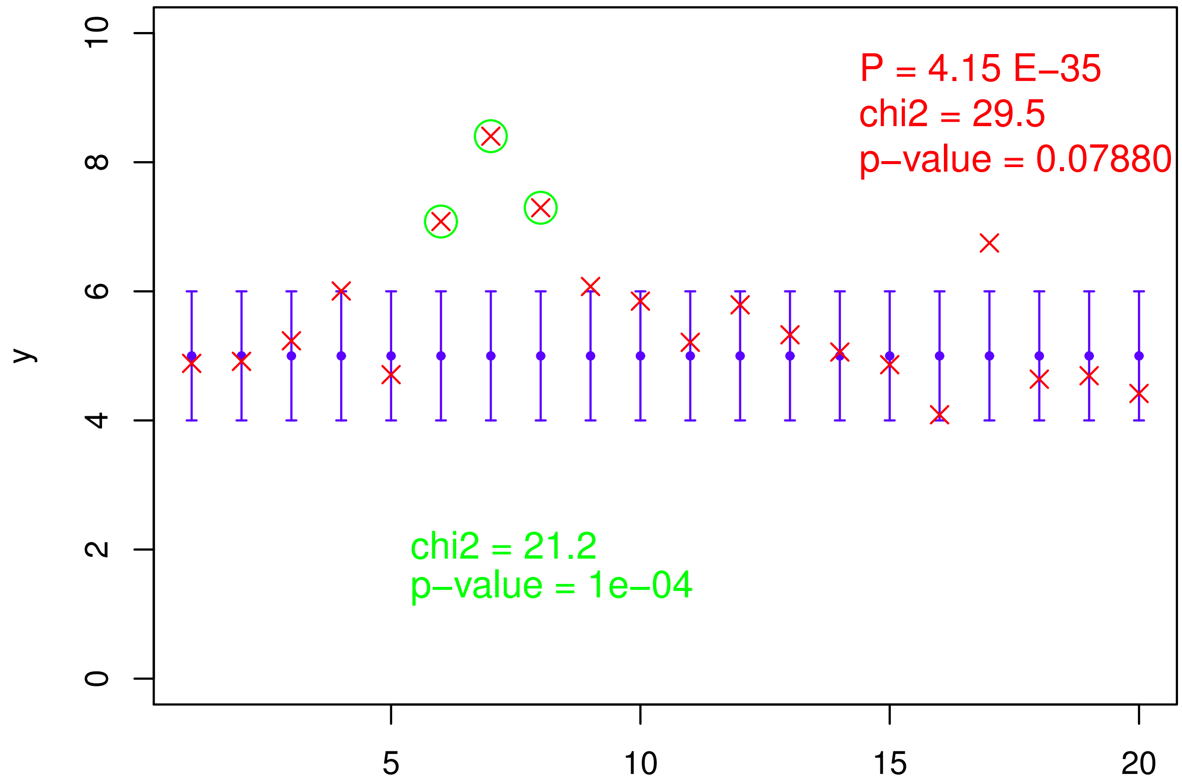
Sometimes the χ^2 test does not give "the wished result"



Then it is calculated in the 'suspicious region'

The art of χ^2

Sometimes the χ^2 test does not give "the wished result"



Then it is calculated in the 'suspicious region'

$\Rightarrow$ If we add the two side points, χ^2 becomes 22.2.

$\Rightarrow$ But with 5 points we had got a p-value of 5×10^{-4}

p-value: what they are

p-value:

- Probability of the tail(s) of a ‘test variable’ (a “statistic”):

$$P(\theta \geq \theta_{mis}) = \int_{\theta_{mis}}^{\infty} f(\theta \,|\, H_0)\, d\theta$$

$$P[(\theta \geq \theta_{mis}) \cap (\theta \leq (\theta^c)_{mis})] = 1 - \int_{(\theta^c)_{mis}}^{\theta_{mis}} f(\theta \,|\, H_0)\, d\theta$$

- θ is an arbitrary function of the data.
- …and often of a subsample of the data.
- $f(\theta \,|\, H_0)$ is obtained ‘somehow’, analitically, numerically, or by Monte Carlo methods.

p-value: what they <u>are not</u>

- What we wanted:
  - falsify the hypothesis H_0:
    $\Rightarrow$ impossible, from the logical point of view (as long as there are stochastic effects).

p-value: what they <u>are not</u>

- What we wanted:
  - falsify the hypothesis H_0:
    $\Rightarrow$ impossible, from the logical point of view (as long as there are stochastic effects).
- Therefore we content ourself with
  - updating our confidence about H_0
    in the light of the experimental data:

$$P(H_0\,|\,\text{data})$$

p-value: what they are not

- What we wanted:
  - falsify the hypothesis H_0:
    $\Rightarrow$ impossible, from the logical point of view (as long as there are stochastic effects).
- Therefore we content ourself with
  - updating our confidence about H_0
    in the light of the experimental data:

$$P(H_0 \,|\, \text{data})$$

$\Rightarrow$ BUT the p-value do not provide this:

$$P(\theta \geq \theta_{mis} \,|\, H_0) \not\Longleftrightarrow P(H_0 \,|\, \theta_{mis})$$

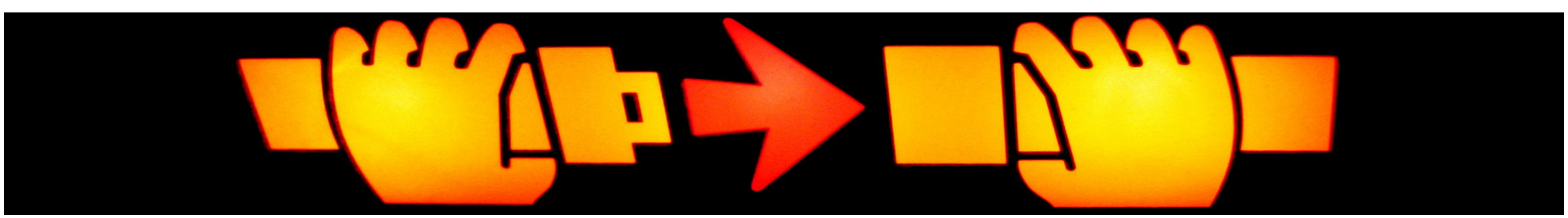
$\Rightarrow$ Although they are erroneously confused with this!

p-value: what they are not

- What we wanted:
  - falsify the hypothesis H_0:
    $\Rightarrow$ impossible, from the logical point of view (as long as there are stochastic effects).
- Therefore we content ourself with
  - updating our confidence about H_0
    in the light of the experimental data:

$$P(H_0 \,|\, \text{data})$$

Tight seat belts!



Misunderstandings p-values

http://en.wikipedia.org/wiki/P-value#Misundersta

Misunderstandings p-values

`http://en.wikipedia.org/wiki/P-value#Misunderst`

1. **The p-value is not the probability that the null hypothesis is true.**

Misunderstandings p-values

http://en.wikipedia.org/wiki/P-value#Misundersta

1. **The p-value is not the probability that the null hypothesis is true.** In fact, frequentist statistics does not, and cannot, attach probabilities to hypotheses. . . .

Misunderstandings p-values

http://en.wikipedia.org/wiki/P-value#Misundersta

1. **The p-value is not the probability that the null hypothesis is true.** In fact, frequentist statistics does not, and cannot, attach probabilities to hypotheses. . . .
2. **The p-value is not the probability that a finding is "merely a fluke."**. . .

Misunderstandings p-values

http://en.wikipedia.org/wiki/P-value#Misundersta

1. **The p-value is not the probability that the null hypothesis is true.** In fact, frequentist statistics does not, and cannot, attach probabilities to hypotheses. . . .
2. **The p-value is not the probability that a finding is "merely a fluke."**. . .
3. The p-value is not the probability of falsely rejecting the null hypothesis.

   . . .

7. . . .

The 5 sigma Higgs!

July 2012

- “The data confirm the 5 sigma threshould, i.e. a probability of discovery of 99.99994%” (one of the many claims you could read on the web).

The 5 sigma Higgs!

July 2012

- “The data confirm the 5 sigma threshould, i.e. a probability of discovery of 99.99994%” (one of the many claims you could read on the web).
- “Ahead of the expected announcement, the journal Nature reported ‘pure elation’ Monday among physicists searching for the Higgs boson. *One team saw only ”a 0.00006% chance of being wrong*, the journal said.” (USA Today, 2 July 2012).

The 5 sigma Higgs!

July 2012

- “The data confirm the 5 sigma threshould, i.e. a probability of discovery of 99.99994%” (one of the many claims you could read on the web).
- “Ahead of the expected announcement, the journal Nature reported ‘pure elation’ Monday among physicists searching for the Higgs boson. *One team saw only ”a 0.00006% chance of being wrong*, the journal said.” (USA Today, 2 July 2012).
- Etc. etc. $\Rightarrow$ Google
  - “higgs cern 0.00006 chance”: $\approx 1.6 \times 10^4$ results

The 5 sigma Higgs!

July 2012

- “The data confirm the 5 sigma threshould, i.e. a probability of discovery of 99.99994%” (one of the many claims you could read on the web).
- “Ahead of the expected announcement, the journal Nature reported ‘pure elation’ Monday among physicists searching for the Higgs boson. *One team saw only ”a 0.00006% chance of being wrong*, the journal said.” (USA Today, 2 July 2012).
- Etc. etc. $\Rightarrow$ Google
  - “higgs cern 0.00006 chance”: $\approx 1.6 \times 10^4$ results
  - “higgs cern ’99.99994%’”: $\approx 1.5 \times 10^6$ results

`http://www.roma1.infn.it/~dagos/badmath/#added`

Probabilistic reasoning

Are we then really stuck?

Probabilistic reasoning

Are we then really stuck?

Fortunatly not, at some conditions . . .

- When the game becomes probabilistic. . .
  ... probability theory has to enter the game.

Probabilistic reasoning

Are we then really stuck?

Fortunatly not, at some conditions . . .

- When the game becomes probabilistic. . .
  ... probability theory has to enter the game.

  ??

  But weren't already Gaussians, χ^2, σ's, etc.?

Probabilistic reasoning

Are we then really stuck?

Fortunatly not, at some conditions . . .

- When the game becomes probabilistic. . .
  ... probability theory has to enter the game.

  ??

  But weren't already Gaussians, χ^2, σ's, etc.?

- The 'classical' framework of hypothesis tests misses – because explicitly forbitten! – the foundamental thing we need in our game:

Probabilistic reasoning

Are we then really stuck?

Fortunatly not, at some conditions . . .

- When the game becomes probabilistic. . .
  ... probability theory has to enter the game.

  ??

  But weren't already Gaussians, χ^2, σ's, etc.?

- The 'classical' framework of hypothesis tests misses – because explicitly forbitten! – the foundamental thing we need in our game:

  probability of hypotheses.

Probabilistic reasoning

Are we then really stuck?

Fortunatly not, at some conditions . . .

- When the game becomes probabilistic. . .
  ... probability theory has to enter the game.

  ??

  But weren't already Gaussians, χ^2, σ's, etc.?

- The 'classical' framework of hypothesis tests misses – because explicitly forbitten! – the foundamental thing we need in our game:

  probability of hypotheses.

  - 'Mismatch' between our natural way of thinking and the statistics theory:
  - $P(H_0 \,|\, \text{data}) \longleftrightarrow P(\theta \geq \theta_{mis} \,|\, H_0)$

Probabilistic reasoning

Are we then really stuck?

Fortunatly not, at some conditions . . .

- When the game becomes probabilistic. . .
  ... probability theory has to enter the game.

  ??

  But weren't already Gaussians, χ^2, σ's, etc.?
- The 'classical' framework of hypothesis tests misses – because explicitly forbitten! – the foundamental thing we need in our game:
- It is enough get rid of '900 statisticians (the 'frequentists') and reload 'serious guys',
  $\rightarrow$ restart from Laplace, together with Gauss, Bayes, etc.,

Beliefs and bets

Recover the natural concept of probability

- “how much I am confident in something”
- “how much I believe something”

Beliefs and bets

Recover the natural concept of probability

- "how much I am confident in something"
- "how much I believe something"

"The usual touchstone, whether that which someone asserts is merely his persuasion – or at least his subjective conviction, that is, his firm belief – is betting. It often happens that someone propounds his views with such positive and uncompromising assurance that he seems to have entirely set aside all thought of possible error. A bet disconcerts him. Sometimes it turns out that he has a conviction which can be estimated at a value of one ducat, but not of ten. For he is very willing to venture one ducat, but when it is a question of ten he becomes aware, as he had not previously been, that it may very well be that he is in error." (Kant)

Beliefs and bets

Recover the natural concept of probability

- “how much I am confident in something”
- “how much I believe something”
- “the more I believe, more money I can bet”

Beliefs and bets

Recover the natural concept of probability

- “how much I am confident in something”
- “how much I believe something”
- “the more I believe, more money I can bet”
- “my degree of belief depends on the information I have got (stored in my brain!)”

Beliefs and bets

Recover the natural concept of probability

- "how much I am confident in something"
- "how much I believe something"
- "the more I believe, more money I can bet"
- "my degree of belief depends on the information I have got (stored in my brain!)"
- "it seems natural – I would be terrified by the contrary! – that other brains store different 'information'" [*'subjective nature of probability'*]

Beliefs and bets

Recover the natural concept of probability

- “how much I am confident in something”
- “how much I believe something”
- “the more I believe, more money I can bet”
- “my degree of belief depends on the information I have got (stored in my brain!)”
- “it seems natural – I would be terrified by the contrary! – that other brains store different ‘information’” [*‘subjective nature of probability’*]
- “I am rationally ready to change my opinion”

Beliefs and bets

Recover the natural concept of probability

- “how much I am confident in something”
- “how much I believe something”
- “the more I believe, more money I can bet”
- “my degree of belief depends on the information I have got (stored in my brain!)”
- “it seems natural – I would be terrified by the contrary! – that other brains store different ‘information’” [*‘subjective nature of probability’*]
- “I am rationally ready to change my opinion”
- “. . . but more unlikelly hypotheses initially were, the stronger evidence is needed, possible provided (independently) by several persons I trust”

Laplace's "Bayes Theorem"

"The greater the probability of an observed event given any one of a number of causes to which that event may be attributed, the greater the likelihood of that cause {given that event}.

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

Laplace's "Bayes Theorem"

"The greater the probability of an observed event given any one of a number of causes to which that event may be attributed, the greater the likelihood of that cause {given that event}. The probability of the existence of any one of these causes {given the event} is thus a fraction whose numerator is the probability of the event given the cause, and whose denominator is the sum of similar probabilities, summed over all causes.

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

Laplace's "Bayes Theorem"

"The greater the probability of an observed event given any one of a number of causes to which that event may be attributed, the greater the likelihood of that cause {given that event}. The probability of the existence of any one of these causes {given the event} is thus a fraction whose numerator is the probability of the event given the cause, and whose denominator is the sum of similar probabilities, summed over all causes. If the various causes are not equally probable *a priory*, it is necessary, instead of the probability of the event given each cause, to use the product of this probability and the *possibility of the cause itself*."

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

Laplace's "Bayes Theorem"

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

"This is the fundamental principle (*) of that branch of the analysis of chance that consists of reasoning *a posteriori* from events to causes"

(*) In his "Philosophical essay" Laplace calls 'principles' the 'fondamental rules'.

Laplace's "Bayes Theorem"

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

"This is the **fundamental principle** (*) of that branch of the analysis of chance that consists of reasoning *a posteriori* **from events to causes**"

(*) In his "Philosophical essay" Laplace calls 'principles' the 'fondamental rules'.

Note: denominator is just a normalization factor.

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

Most convenient way to remember Bayes theorem

Laplace's teaching

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

- We should possible use the data, rather then the test variables 'θ' (χ^2 etc);
  [although in some case 'sufficient summaries' exist]

Laplace's teaching

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

- We should possible use the data, rather then the test variables 'θ' (χ^2 etc);
  [although in some case 'sufficient summaries' exist]
- At least two hypotheses are needed!

Laplace's teaching

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

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

Laplace's teaching

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

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

Laplace's teaching

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

- We should possible use the data, rather then the test variables 'θ' (χ^2 etc);
  [although in some case 'sufficient summaries' exist]
- At least two hypotheses are needed!
- ...and how they appear initially likelly!
- If $P(\text{data}\,|\,H_i) = 0$, it follows $P(H_i\,|\,\text{data}) = 0$:
  $\Rightarrow$ falsification (the 'serious' one) is a corollary of the theorem, rather than a principle.
- There is no conceptual problem with the fact that $P(\text{dati}\,|\,H_1) \to 0$ (e.g. 10^{-37}), provided the ratio $P(\text{dati}\,|\,H_0)/P(\text{dati}\,|\,H_1)$ is not undefined.

But statistical tests do work!

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

But statistical tests do work!

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

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

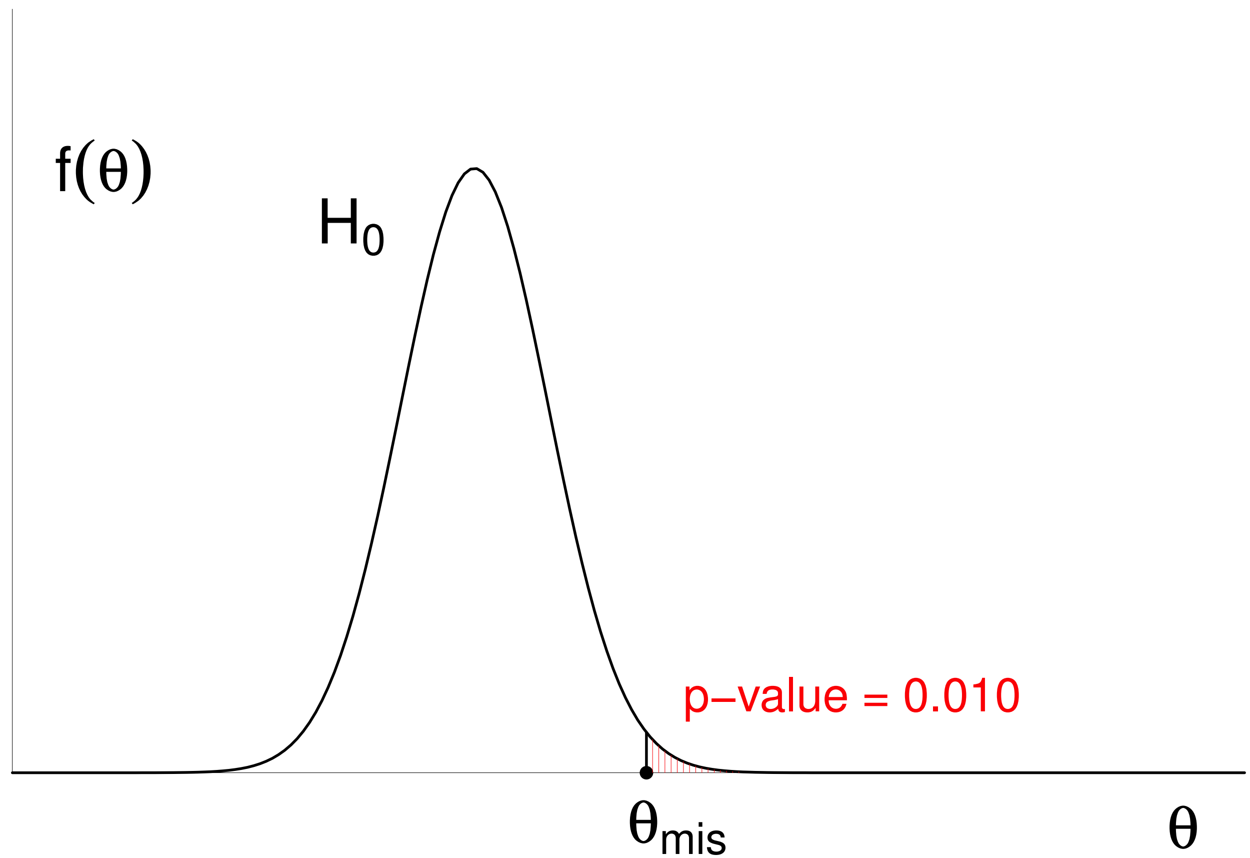
But statistical tests do work!

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

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

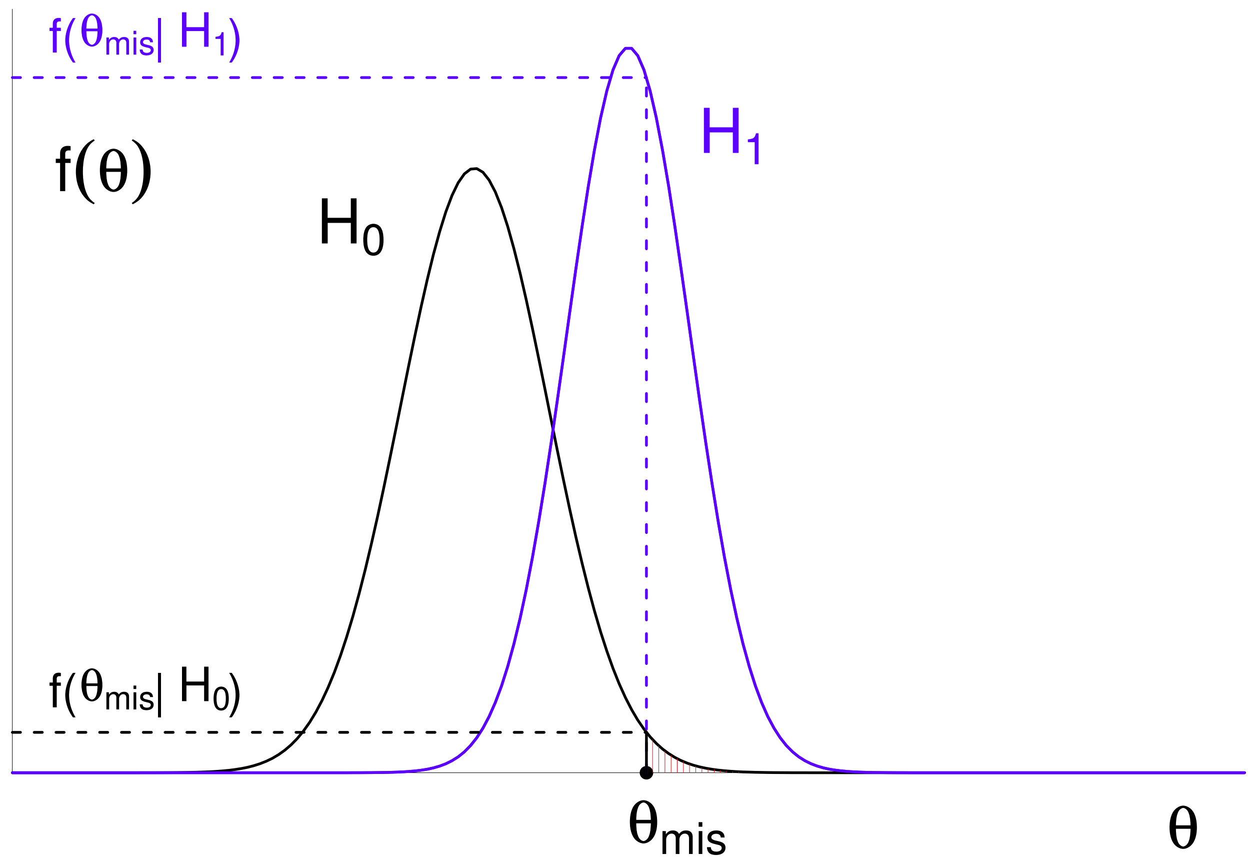
- But now we are also able to explain the reason.

But statistical tests do work!



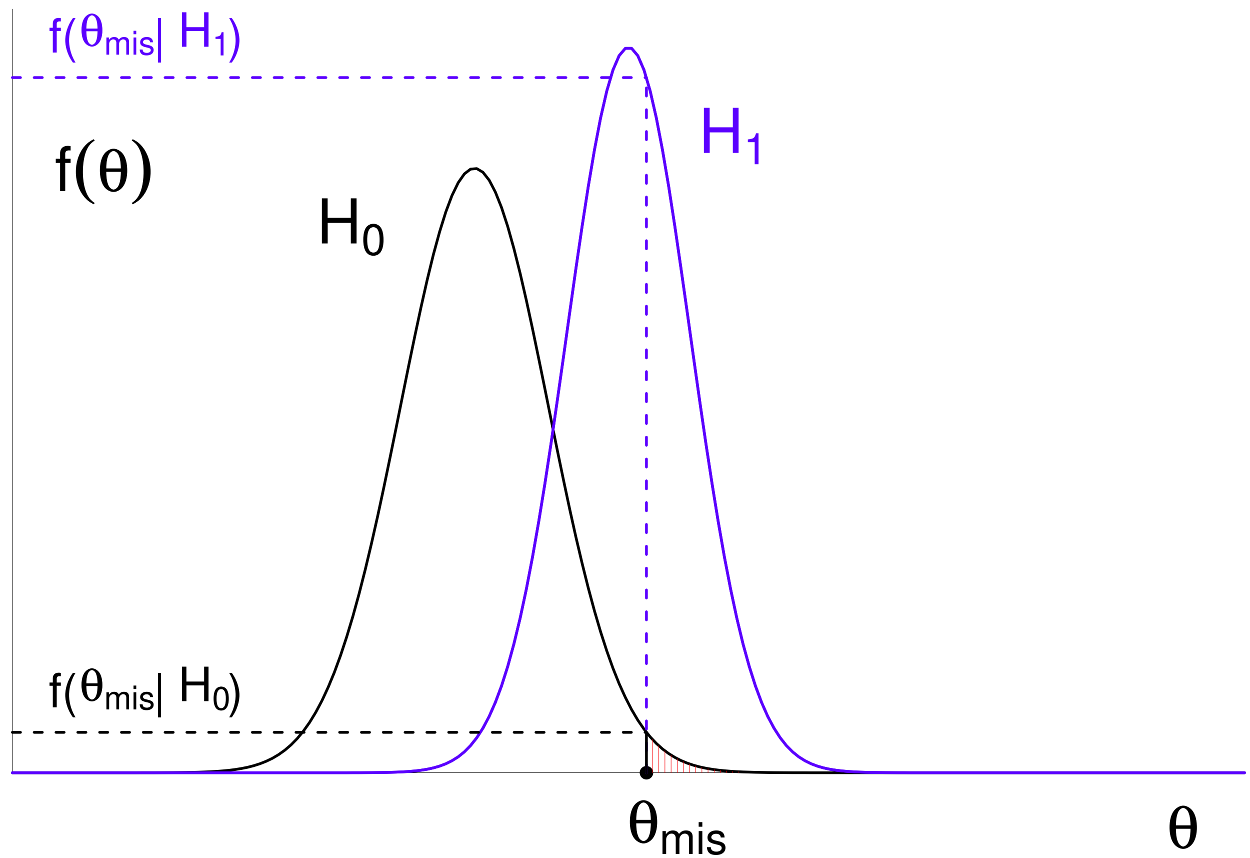
Why should the observation of θ_{mis} should diminish our confidence on H_0?

But statistical tests do work!



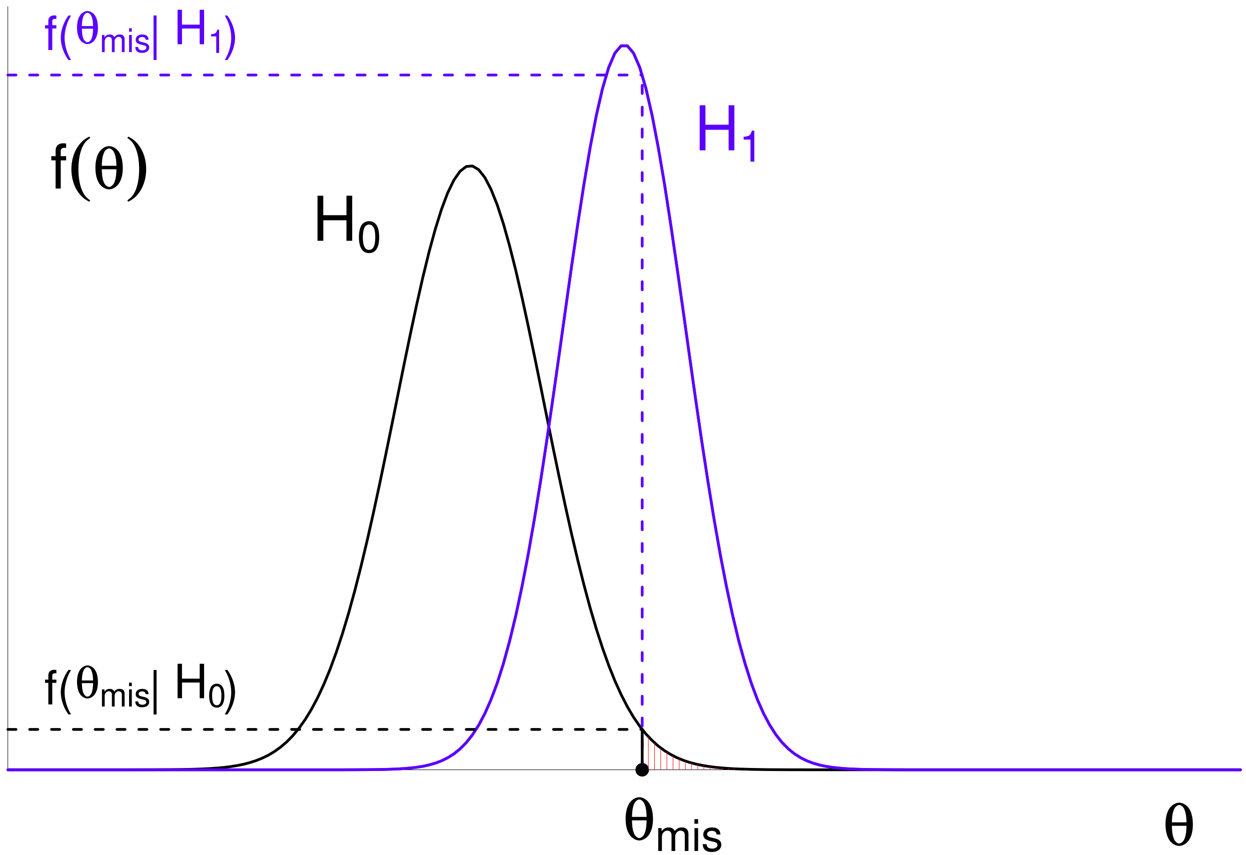
Because *often* we give *some chance* to a possible alternative hypothesis H_1, even if we are not able to exactly formulate it.

But statistical tests do work!



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

But statistical tests do work!



But is the alternative hypothesis H_1 is unconcievable, or hardly believable, the 'smalleness' of the area is irrelevant

Sensational announcements Vs sound Physics

At this point it is rather clear why most physicists disbelieved the 2011 anouncements by CDF and Opera

Sensational announcements Vs sound Physics

At this point it is rather clear why most physicists disbelieved the 2011 anouncements by CDF and Opera

As it was quite obvious that what the LHC experiments were glipsing at the end of 2011 was the 30 years searched for Higgs boson

Sensational announcements Vs sound Physics

At this point it is rather clear why most physicists disbelieved the 2011 anouncements by CDF and Opera

As it was quite obvious that what the LHC experiments were glipsing at the end of 2011 was the 30 years searched for Higgs boson

Sensational announcements Vs sound Physics

At this point it is rather clear why most physicists disbelieved the 2011 anouncements by CDF and Opera

As it was quite obvious that what the LHC experiments were glipsing at the end of 2011 was the 30 years searched for Higgs boson (Also becaause in that case the great discovery would have been not to find it!)

Sensational announcements Vs sound Physics

At this point it is rather clear why most physicists disbelieved the 2011 anouncements by CDF and Opera

As it was quite obvious that what the LHC experiments were glipsing at the end of 2011 was the 30 years searched for Higgs boson (Also becaause in that case the great discovery would have been not to find it!)

Don't get confused by sigma's and 'strange significances' that do not tell you how how much to believe in the claim.

“Is the ‘new particle’ the Higgs?”

We have often listened in the past year the following statement:

> “We have discovered at CERN a new particle.
> We have to understand if it is the Higgs boson”

"Is the 'new particle' the Higgs?"

We have often listened in the past year the following statement:

"We have discovered at CERN a new particle.
We have to understand if it is the Higgs boson"

???

“Is the ‘new particle’ the Higgs?”

We have often listened in the past year the following statement:

> “We have discovered at CERN a new particle.
> We have to understand if it is the Higgs boson”

This statement implies that our confidence that the ≈ 126 GeV ‘excess’ is a new particle is due from the 5 sigmas alone.

“Is the ‘new particle’ <u>the</u> Higgs?”

We have often listened in the past year the following statement:

> “We have discovered at CERN a new particle.
> We have to understand if it is the Higgs boson”

This statement implies that our confidence that the ≈ 126 GeV ‘excess’ is a new particle is due from the 5 sigmas alone.

But we have just seen that this is not logically defendable!

“Is the ‘new particle’ the Higgs?”

We have often listened in the past year the following statement:

> “We have discovered at CERN a new particle.
> We have to understand if it is the Higgs boson”

This statement implies that our confidence that the ≈ 126 GeV ‘excess’ is a new particle is due from the 5 sigmas alone.

But we have just seen that this is not logically defendable!

$\rightarrow$ The excess is surely a particle only if it is the Higgs!

“Is the ‘new particle’ <u>the</u> Higgs?”

We have often listened in the past year the following statement:

> “We have discovered at CERN a new particle.
> We have to understand if it is the Higgs boson”

This statement implies that our confidence that the ≈ 126 GeV ‘excess’ is a new particle is due from the 5 sigmas alone.

It is a question of Physics not (only) of statistics:

- success of standard model;
- radiative corrections
  (the diagrams entering R.C. are essentially the same the produce the Higgs in the final state!)

“Is the ‘new particle’ the Higgs?”

We have often listened in the past year the following statement:

> “We have discovered at CERN a new particle.
> We have to understand if it is the Higgs boson”

This statement implies that our confidence that the ≈ 126 GeV ‘excess’ is a new particle is due from the 5 sigmas alone.

It is a question of Physics not (only) of statistics:

- success of standard model;
- radiative corrections
  (the diagrams entering R.C. are essentially the same the produce the Higgs in the final state!)
- Physics is something SERIOUS! (not a statistician’s toy)

Conclusions of Part 1

Philip Ball (Guardian, 23 dicembre 2011)
(`http://www.guardian.co.uk/commentisfree/2011/de`

"So D'Agostini recommends that, instead of heeding impressive-sounding statistics, we should ask what scientists actually believe. Better, we should find out if they had put money on it – and how much. After all, that is a tactic endorsed by none other than Kant."

Conclusions of Part 1

Philip Ball (Guardian, 23 dicembre 2011)
(http://www.guardian.co.uk/commentisfree/2011/de

"So D'Agostini recommends that, instead of heeding impressive-sounding statistics, we should ask what scientists actually believe. Better, we should find out if they had put money on it – and how much. After all, that is a tactic endorsed by none other than Kant."

Which is why I'm only being scientific when I say screw the sigmas: I'd place a tenner (but not a ton) on the Higgs, while offering to join Jim Al-Khalili in eating my shorts if neutrinos defy relativity."

Conclusions of Part 1

Philip Ball (Guardian, 23 dicembre 2011)
(`http://www.guardian.co.uk/commentisfree/2011/de`

"So D'Agostini recommends that, instead of heeding impressive-sounding statistics, we should ask what scientists actually believe. Better, we should find out if they had put money on it – and how much. After all, that is a tactic endorsed by none other than Kant."

Which is why I'm only being scientific when I say screw the sigmas: I'd place a tenner (but not a ton) on the Higgs, while offering to join Jim Al-Khalili in eating my shorts if neutrinos defy relativity."

⇒ *He has finally won both bets!*