Claims of discoveries based on sigmas

Giulio D'Agostini

giulio.dagostini@roma1.infn.it
http://www.roma1.infn.it/~dagos/

Università La Sapienza e INFN, Roma, Italy

Preamble

Someone might have come here to hear about statistics...

Preamble

Someone might have come here to hear about statistics. . .

Indeed I am often invited to give talks, tutorials or courses on *statistics* (for physicists),

Preamble

Someone might have come here to hear about statistics. . .

Indeed I am often invited to give talks, tutorials or courses on *statistics* (for physicists), although I dislike "statistics" . . . and (with exceptions) *statisticians*.

Statistics lectures?

If I insist on probability, rather than speaking, very generally, about statistics, it is because I have good reasons.

Statistics lectures?

As far as the laws of mathematics refer to reality,
they are not certain,
and as far as they are certain,
they do not refer to reality.

(Einstein)

Statistics lectures?

"If we were not ignorant there would be no probability, there could only be certainty.

Statistics lectures?

"If we were not ignorant there would be no probability, there could only be certainty. But our ignorance cannot be absolute, for then there would be no longer any probability at all.

Statistics lectures?

"If we were not ignorant there would be no probability, there could only be certainty. But our ignorance cannot be absolute, for then there would be no longer any probability at all. Thus the problems of probability may be classed according to the greater or less depth of our ignorance."

(Poincaré)

Statistics lectures?

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

(Feynman)

Statistics and truth (from lectures at CERN)

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

Statistics and truth (from lectures at CERN)

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

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

Statistics and truth (from lectures at CERN)

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

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

Statistics and truth (from lectures at CERN)

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

- proposed by organizers → accepted...
- I interpret the title 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"?
     - What is the true value of a quantity?
     - What is the true theory that describes the world?
  2. "Tell the truth" $\Longleftrightarrow$ "to lie"?

Statistics and truth (from lectures at CERN)

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

- proposed by organizers → accepted...
- I interpret the title 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"?

Statistics and truth (from lectures at CERN)

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

- proposed by organizers → accepted...
- I interpret the title 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

Statistics and truth (from lectures at CERN)

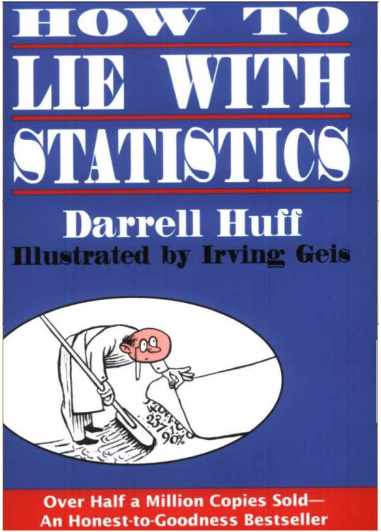
Title of the lectures ("Telling the truth with statistics")

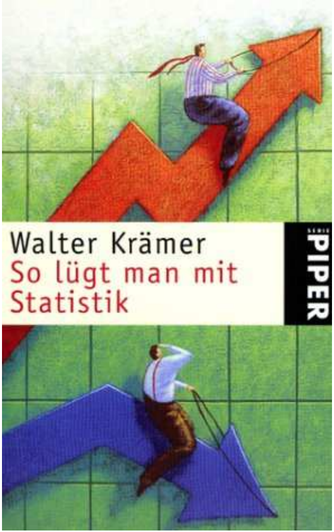
- proposed by organizers → accepted...
- I interpret the title as a direct question, to which I will try to give my best answer, quite frankly.
- How to interpret the question?
  1. ~~"Tell the Truth"?~~ ⇒ Question to God
     - ~~What is the true value of a quantity?~~
     - ~~What is the true theory 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

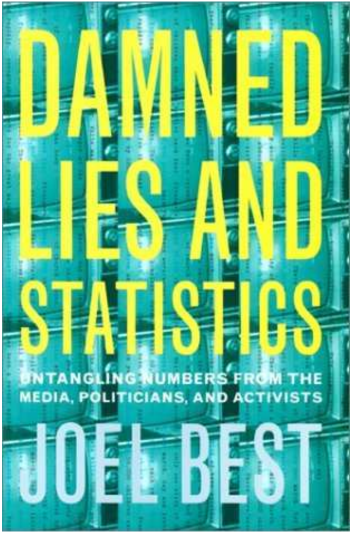
Well known subject

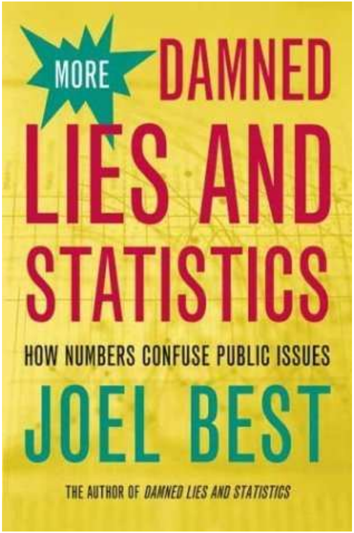




Damned lies and statistics

Well known subject, especially in marketing and politics





...

BloombergView

HOME | EDITORIAL BOARD | COLUMNISTS | TOPICS

SCIENCE

Lies, Damned Lies and Physics

16 DEC 30, 2015 9:30 AM EST

By Faye Flam

To most of us, 93-to-1 odds would make for a clear-cut bet. To physicists? Not so much.

On Dec. 15, the New York Times reported that Santa may have brought physics a new subatomic particle, a hitherto unknown entity materializing in the giant collider at CERN, near Geneva. It wasn't a sure thing, but according to the Times, the odds are in the scientists' favor, with only a 1-in-93 chance that the data pointing to the particle represent a statistical fluke.

...

The New York Times   SUBSCRIBE

SCIENCE

Physicists in Europe Find Tantalizing Hints of a Mysterious New Particle

By **DENNIS OVERBYE** DEC. 15, 2015

Email

Share

Tweet

Save

More

Does the Higgs boson have a cousin?

Two teams of physicists working independently at the Large Hadron Collider at CERN, the European Organization for Nuclear Research, reported on Tuesday that they had seen traces of what could be a new fundamental particle of nature.

One possibility, out of a gaggle of

Researchers at the Large Hadron Collider at CERN are smashing together protons to search for new particles and forces.

Fabrice Coffrini/Agence France-Presse — Getty Images

New York Times, 15 December 2015

"I don't think there is anyone around who thinks this is conclusive," said Kyle Cranmer, a physicist from New York University who works on one of the CERN teams, known as Atlas. "But it would be huge if true," he said, noting that many theorists had put their other work aside to study the new result.

When all the statistical effects are taken into consideration, Dr. Cranmer said, the bump in the Atlas data had about a 1-in-93 chance of being a fluke — far stronger than the 1-in-3.5-million odds of mere chance, known as five-sigma, considered the gold standard for a discovery. That might not be enough to bother presenting in a talk except for the fact that the competing CERN team, named C.M.S., found a bump in the same place.

Le Scienze, 19 dicembre 2015

Le Scienze
EDIZIONE ITALIANA DI SCIENTIFIC AMERICAN

ZOOM SU ⊙ optogenetica ⊙ epidemiologia ⊙ longevità ⊙ Internet ⊙ visione

19 dicembre 2015

Qualcosa di nuovo da LHC? Solo il tempo lo dirà

(Cortesia Maximilien Brice/CERN)

Nuovi dati degli esperimenti ATLAS e CMS del Large Hadron Collider del CERN di Ginevra hanno mostrato un eccesso nella produzione di coppie di fotoni, localizzato a una massa particolare. Ma è ancora troppo presto per dire se sia un primo segno di una nuova era per la fisica delle particelle oppure solo una fluttuazione del rumore di fondo *di Marco Delmastro*

Le Scienze, 19 dicembre 2015

Nel caso dell'eccesso sullo spettro delle coppie di fotoni, se uno prende il grafico di ATLAS in cui la montagnola è più prominente, la probabilità che questa sia dovuta a una casualità è due su 10.000, dunque piuttosto piccola. Quando però consideriamo il fatto di aver cercato montagnole un po' dappertutto, allora questa probabilità aumenta a due su 100. I numeri di CMS sono persino più grandi, indicando una probabilità ancora più grande che si tratti solo di una fluttuazione del rumore di fondo.

"In the case of the excess in the two-photon spectrum, if one takes the ATLAS plot, where the bump is more prominent, the probability that this is due to randomness is 2 in 10,000, then rather small.

Le Scienze, 19 dicembre 2015

Nel caso dell'eccesso sullo spettro delle coppie di fotoni, se uno prende il grafico di ATLAS in cui la montagnola è più prominente, la probabilità che questa sia dovuta a una casualità è due su 10.000, dunque piuttosto piccola. Quando però consideriamo il fatto di aver cercato montagnole un po' dappertutto, allora questa probabilità aumenta a due su 100. I numeri di CMS sono persino più grandi, indicando una probabilità ancora più grande che si tratti solo di una fluttuazione del rumore di fondo.

"In the case of the excess in the two-photon spectrum, if one takes the ATLAS plot, where the bump is more prominent, the probability that this is due to randomness is 2 in 10,000, then rather small. When instead we consider the fact that we have been looking bumps everywhere, this probability increases to 2 in 100. CMS' numbers are even larger, indicating an even larger probability that it is just a fluctuation of the background."

Le Scienze, 19 dicembre 2015

Nel caso dell'eccesso sullo spettro delle coppie di fotoni, se uno prende il grafico di ATLAS in cui la montagnola è più prominente, la probabilità che questa sia dovuta a una casualità è due su 10.000, dunque piuttosto piccola. Quando però consideriamo il fatto di aver cercato montagnole un po' dappertutto, allora questa probabilità aumenta a due su 100. I numeri di CMS sono persino più grandi, indicando una probabilità ancora più grande che si tratti solo di una fluttuazione del rumore di fondo.

"In the case of the excess in the two-photon spectrum, if one takes the ATLAS plot, where the bump is more prominent, the probability that this is due to randomness is 2 in 10,000, then rather small. When instead we consider the fact that we have been looking bumps everywhere, this probability increases to 2 in 100. CMS' numbers are even larger, indicating an even larger probability that it is just a fluctuation of the background."

A surreal dialogue of a friend of mine with the author

Amico: Nell'articolo è scritto: "... la probabilità che questa sia dovuta a una casualità è due su 10.000, dunque piuttosto piccola. Quando però consideriamo il fatto di aver cercato montagnole un po' dappertutto, allora questa probabilità aumenta a due su 100."

Se capisco bene, lei stima a $(1 - 0.02) = 0.98$ la probabilità che NON si tratti di una fluttuazione casuale nell'ipotesi peggiore.

Cioè ne siamo praticamente certi?

A surreal dialogue of a friend of mine with the author

Friend: In the article there is written "...the probability that is due to randomness is two in 10000, hence rather low. When however we take into account the fact that we have been searching for bumps everywhere, this probability rises to two in 100."

If I understand well, you estimate in $(1 - 0.02) = 0.98$ the probability that it is NOT a random fluctuation, in the worst hypothesis.

Does it mean we are almost certain of it?

A surreal dialogue of a friend of mine with the author

Autore: Ciao,

Due commenti:

1) non puoi trasformare la probabilità dell'ipotesi nulla in quella dell'ipotesi di scoperta così. Che ci sia il 2% di probabilità che l'eccesso sia dovuto alla fluttuazione del fondo non vuol dire che c'è il 98% di probabilità che l'eccesso sia generato da un segnale genuino. I p-valori sono complicati ;-)

2) il 2% che si tratti di una fluttuazione non è una probabilità piccola!

A surreal dialogue of a friend of mine with the author

Author: Ciao,

Two comments:

1) you cannot transform so the probability of the null hypothesis in that of the hypothesis of discovery. The fact that there is 2% probability that the excess is due to a fluctuation of the background does not mean that there is 98% probability that the excess is generated by a genuine signal. P-values are complicate ;-)

2) 2% of being a fluctuation is not a small probability!

A surreal dialogue of a friend of mine with the author

Amico: Perdonami, non è questione di p-value, [...]
Ma del senso letterale di quello che scrivi:

Se A è l'affermazione "questa sia dovuta a una casualità", tu dici che $P(\text{A}) = 2\%$

Ergo $P(\text{non-A}) = 98\%$ perché $P(\text{A}) + P(\text{non-A}) = 1$ sta negli assiomi della probabilità.
O no?

A surreal dialogue of a friend of mine with the author

Friend: Excuse me, it isn't a matter of p-values, [. . .]
but of the literal meaning of what you wrote:

If A is the statement "this is due to randomness", you state that
$P(\text{A}) = 2\%$

Therefore $P(\text{non-A}) = 98\%$ because $P(\text{A}) + P(\text{non-A}) = 1$ is in
the axioms of probability.
Or not?

A surreal dialogue of a friend of mine with the author

Autore: Ciao,

No, purtroppo si tratta proprio di p-value, e del confronto tra probabilità condizionali e non condizionali tra due ipotesi. Tutto questo nell'articolo per le Scienze ovviamente non c'è, e li ho dovuto "tradurre" per il pubblico non-tecnico in termini (approssimati) di probabilità tradizionale una trattazione in realtà più complessa. Se però ti interessa fare una discussione formale, allora mi spiace ma non è quell'articolo a cui devi fare referenza, ma questo:

https://cds.cern.ch/record/2114853
(vedi per esempio la sezione 8 e le sue referenze).

Buona lettura, M.

A surreal dialogue of a friend of mine with the author

Author: Ciao,

No, unfortunately it is indeed about p-values, and the comparison between conditional and non conditional probabilities of two hypotheses. All this in the Le Scienze article is obviously missing, and I had to "translate" a treatment in reality much more complex for the general public in (approximated) terms of traditional probability . If however your are interested in a formal discussion, then I am sorry but it is not that article that you have to take as reference, but this one:

`https://cds.cern.ch/record/2114853`
(see for example section 8 and references therein).

Have a nice reading, M.

(Personal mails omitted)

Interacting with Kyle Cranmer (→ NYT 15/12/2015)

To Cranmer (23/12/2016 15:16)

```
According to the journalist you state that
"the bump in the Atlas data had about a 1-in-93 chance
of being a fluke", THAT IS 92-in-93 of NOT being a fluke.

In other words, FAIR bet odds are 1 to 92, right?
If this is you opinion, you should be ready to accept
the bet in either direction.

For my reasons, I choose to bet 10 CHF on Fluke,
asking you to bet 920 CHF on non-Fluke.

To be more clear (its is a question of money!):
- I pay 10 CHF and you pay 920 CHF;
- if the present excess will result to be something
  a real new particle, you will get the 930 CHF;
- if the present excess will turn out to be just a fluke,
  I will get the 930 CHF.
```

Using bets to assess/check beliefs

Even Emmanuel Kant would agree with my ‘provocation’.

Using bets to assess/check beliefs

Even Emmanuel Kant would agree with my ‘provocation’.

> *“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.*

Using bets to assess/check beliefs

Even Emmanuel Kant would agree with my ‘provocation’.

> *“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.*

Using bets to assess/check beliefs

Even Emmanuel Kant would agree with my 'provocation'.

> *"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.*

Using bets to assess/check beliefs

Even Emmanuel Kant would agree with my 'provocation'.

> *"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)

Using bets to assess/check beliefs

Even Emmanuel Kant would agree with my ‘provocation’.

> *“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)*

Interacting with Kyle Cranmer ($\rightarrow$ NYT 15/12/2015)

From Cranmer (23/12/2016 19:08)

```
I understand the betting odds, but that wasn't my quote.
I provided the p-value number and he wrote the part
about being a "fluke".

That phrase is not precise and I can interpret either
as a classic probability inversion (mistake)
or as a colloquial way of saying
"a bump at least this big assuming there is no signal"
(i.e. a p-value.)

My odds are more like 1/3 that this is real.
I'll bet you 30CHF if you want.
```

Interacting with Kyle Cranmer (→ NYT 15/12/2015)

To Cranmer (23/12/2016 19:47)

```
Thanks a lot for your prompt reply, Kyle!

This is what want I wanted to hear, although I can ensure
you that in other cases similar statements have been
provided _verbatim_ to journalists by our colleagues,
or they have been directly written by them.

(And also in this case, an Italian physicist of ATLAS
has WRITTEN something similar, so that he cannot blame
the journalist)

Anyway, I accept the bet you propose (10CHF Vs 30CHF),
and I do not think we need a kind of notary :-)
```

Interacting with Kyle Cranmer ($\rightarrow$ NYT 15/12/2015)

From Cranmer (23/12/2016 22:38)

```
I agree and appreciate your interest in these matters.
I took an extended interview trying to break down
these points of confusion.

I'll take the bet, and I agree, no notary is needed.
I would hope that by this time next year it will be clear.

All the best,
```

Statistics $\leftrightarrow$ probability

The fact that statistical results are often "misinterpreted" is rather well known.

Statistics $\leftrightarrow$ probability

The fact that statistical results are often "misinterpreted" is rather well known.

But not because the general public is made of idiots!

Statistics $\leftrightarrow$ probability

The fact that statistical results are often "misinterpreted" is rather well known.

But not because the general public is made of idiots!

It is just because the 'conventional' statistical school misuses words and convey wrong messages
(also among expert practitioners, as most physicists).

Statistics $\leftrightarrow$ probability

The fact that statistical results are often "misinterpreted" is rather well known.

But not because the general public is made of idiots!

It is just because the 'conventional' statistical school misuses words and convey wrong messages
(also among expert practitioners, as most physicists).

$\Longrightarrow$

2011: not 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: not only Opera...

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

$$\approx 3.2\,\sigma$$

- September, Opera: neutrinos faster than light

$$\approx 6\,\sigma$$

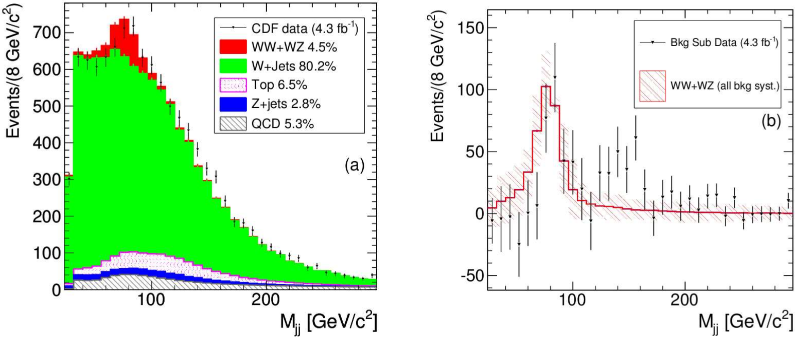
- December, ATLAS e CMS at LHC: signal compatible with the Higgs at about 125 GeV:

$$\approx 3\sigma$$

Why there was substantial scepticism towards the first two announcements, in contrast with a cautious/pronounced optimism towards the third one?

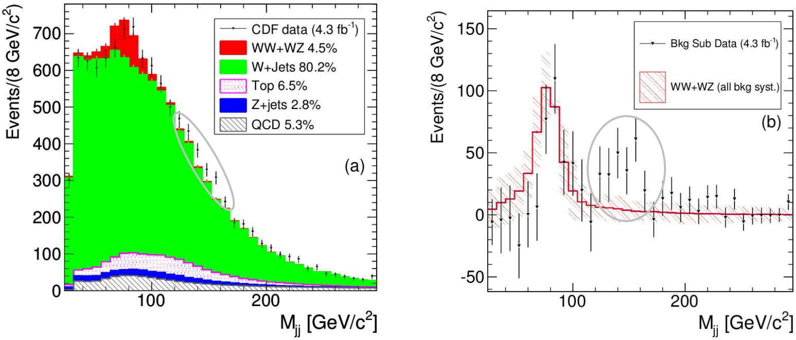
April 2011

CDF Collaboration at the Tevatron



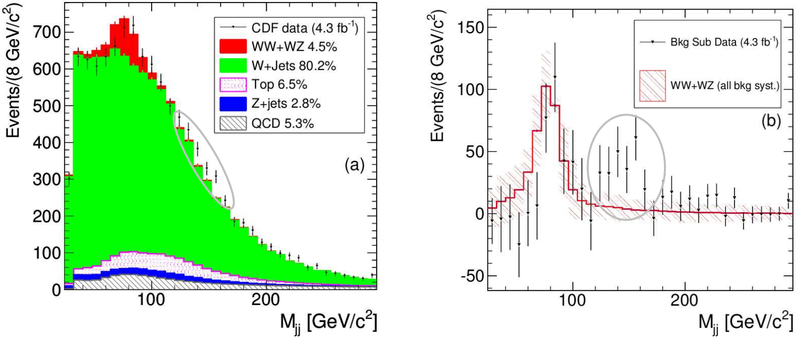
April 2011

CDF Collaboration at the Tevatron



April 2011

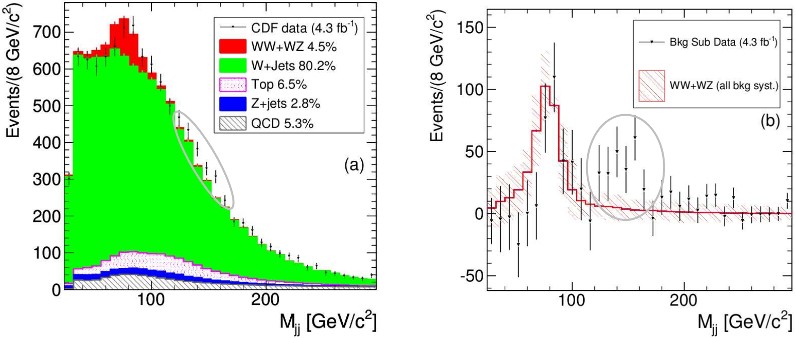
CDF Collaboration at the Tevatron



"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

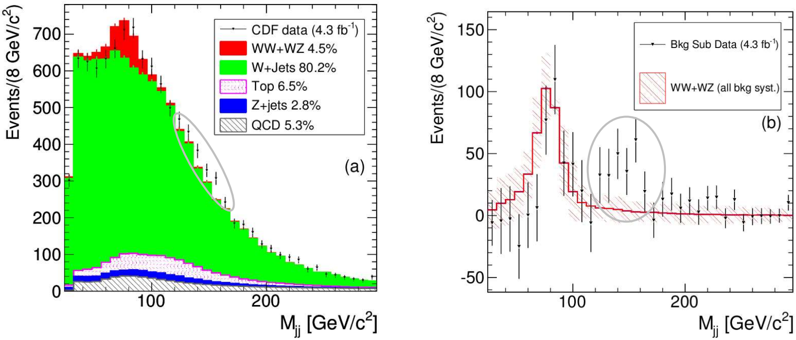


"we obtain a p-value of 7.6×10^{-4}, corresponding to a significance of 3.2 standard deviations"

$3.2\,\sigma$!

April 2011

CDF Collaboration at the Tevatron



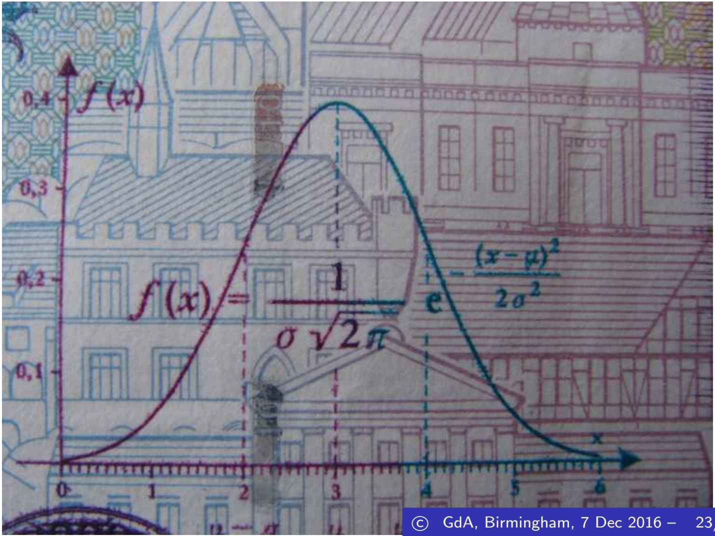
What does it mean?

Sigma and gaussian distribution

Princeps mathematicorum

Sigma and gaussian distribution

“Functio nostra fiet. . .”



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Sigma e probability [gaussian!]

If the random number X is described by a gaussian pdf

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

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

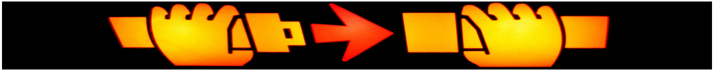
p-value, significance and sigma

"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" ["$3.2\,\sigma$"]

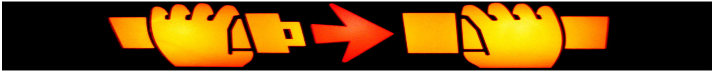
Begin to fasten seat belts!



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!

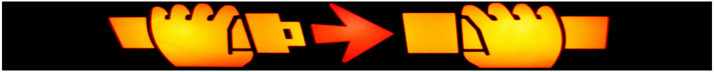


- What is a **p-value**?
- In so far does it provides us a **‘significance’**?

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 σ"]

Begin to fasten seat belts!



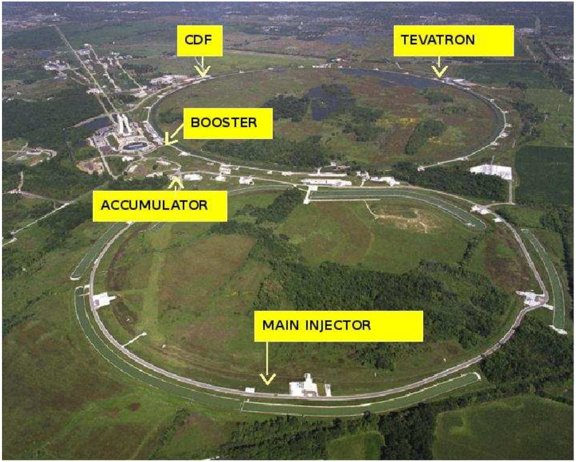
- What is a **p-value**?
- In so far does it provides us a **'significance'**?

In short,

- Is $\mathbf{7.6 \times 10^{-4}}$ a **probability**?
- **of what**?

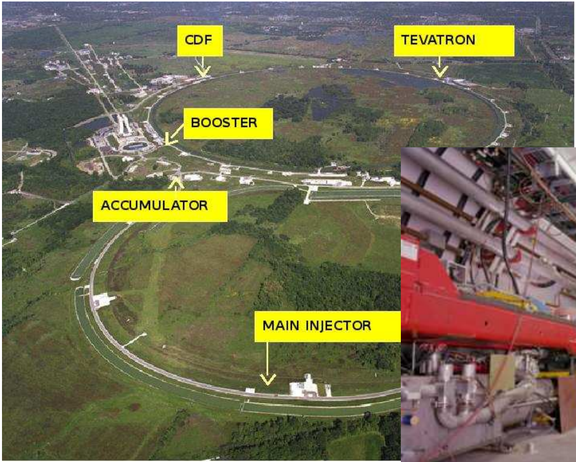
Tevatron and CDF

6.28 km, near Chicago



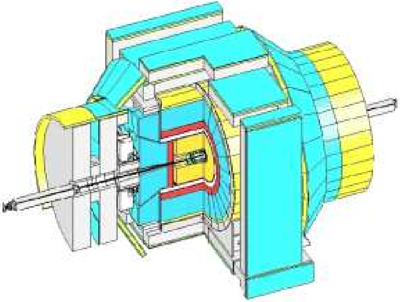
Tevatron and CDF

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



Tevatron and CDF

CDF: a multipurpose ('hermetic') detector

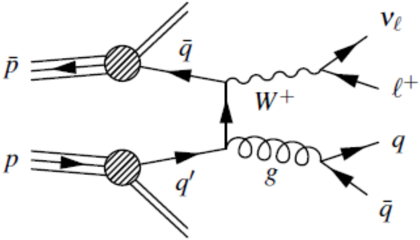


Tevatron and CDF

. . . a large, very sophisticated detector!

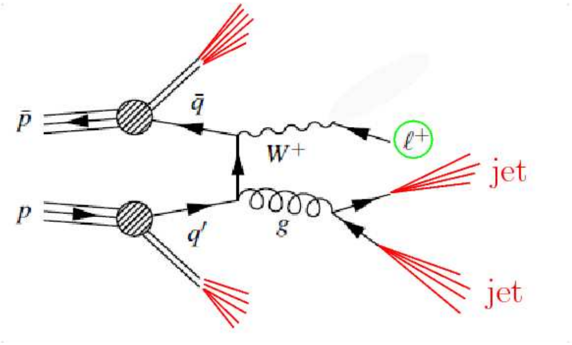
Jet-jet + W

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



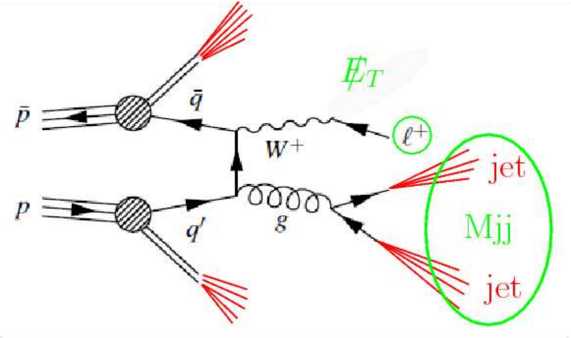
Jet-jet + W

$W + 2\text{jet}$ [+ much more]



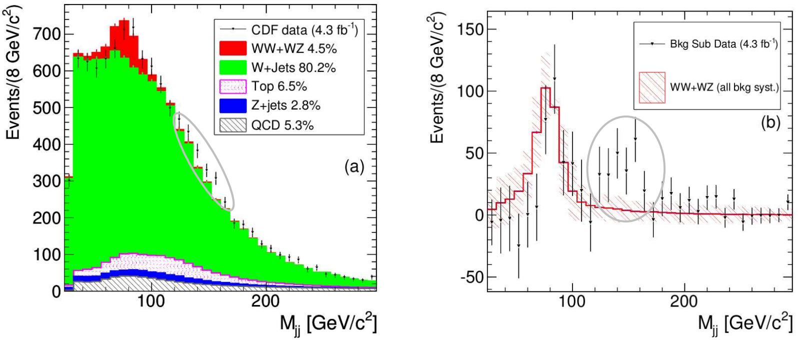
Jet-jet + W

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



The ‘bump’!

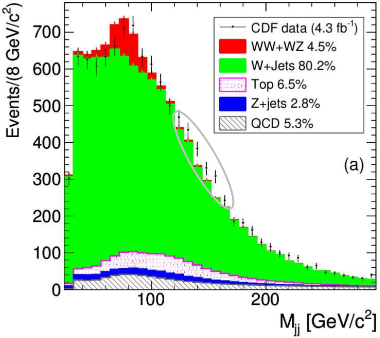
Invariant Mass Distribution of Jet Pairs Produced in Association with a W boson in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV”, (CDF, 4 April 2011)

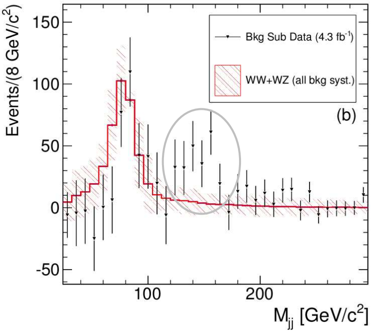


“we obtain a p-value of 7.6×10^{-4}, corresponding to a significance of 3.2 standard deviations” [“$3.2\,\sigma$”]

The 'bump'!

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





What does it mean?

April 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**"

April 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***”*

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

April 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(Statistical fluctuation) $\leq 0.25\%$!

P(True Signal) $\geq 99.75\%$!!

April 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(Statistical fluctuation) $\leq$ 0.25%!

P(True Signal) $\geq$ 99.75%!!

Eureka!!

April 2011, the ‘bump’ explodes

The New York Times, Tuesday April 5:

“the most significant in physics in half a century”

April 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 considered a minor detail. . .)]

April 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!

April 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 contact-person and/or the Fermilab PR’s who contacted him?

April 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 contact-person and/or the Fermilab PR's who contacted him?

From my experience, journalists might make imprecisions, but they do not invent pieces of news [. . . at least the scientific ones. . . 🙂]

April 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.*
>
> *. . .*

April 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***."*

April 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**."

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

April 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.*
> ...

April 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)**."

April 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%. . .

April 2011, the ‘bump’ explodes

Jon Butterworth’s blog 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**.”

April 2011, the 'bump' explodes

Jon Butterworth's blog 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**.*"*
>
> $\Rightarrow P(\textit{Not Fluke}) = P(\textit{"Genuine"}) = 99.99\%$

April 2011, the 'bump' explodes

Jon Butterworth's blog 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**."
>
> $\Rightarrow P(\textit{Not Fluke}) = P(\textit{"Genuine"}) = 99.99\%$

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!"

April 2011, the ‘bump’ explodes

Jon Butterworth’s blog 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**.”
>
> $\Rightarrow P(\mathit{Not\ Fluke}) = P(\text{“}\mathit{Genuine}\text{”}) = 99.99\%$

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!”

Absolutely meaningful! (A part from the initial mismatch)

A masterpiece of good reasoning

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

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

A masterpiece of 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 masterpiece of 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 believe"

A masterpiece of 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 believe"
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 masterpiece of 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 believe"

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?

A masterpiece of 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 believe"

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?

A masterpiece of 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 believe"

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? $\Rightarrow$ "Bayes"

A masterpiece of 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 believe"

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

   "Learning from the experience!"

   ⇒ A physicist should never be dogmatic

But how must our convictions rationally change on the light of new experimental data? Is there a logical rule? ⇒ "Bayes"

⇒ Intuition might fail – it does often fail! –

A masterpiece of 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 believe"

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? $\Rightarrow$ "Bayes"

$\Rightarrow$ Intuition might fail – it <u>does</u> often fail! – and the sigmas are not (always) a good guidance!

A masterpiece of 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" ≠ "What I believe"

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

   "Learning from the experience!"

   ⇒ A physicist should never be dogmatic

But how must our convictions rationally change on the light of new experimental data? Is there a logical rule? ⇒ "Bayes"

⇒ Intuition might fail [*] – it does often fail! – and the sigmas are not (always) a good guidance!

BUT the intuition of experienced scientists is in most cases far superior than the aseptic/pedantic rules of statisticians.

A masterpiece of 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 believe"

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

   "Learning from the experience!"

   ⇒ A physicist should never be dogmatic

But how must our convictions rationally change on the light of new experimental data? Is there a logical rule? ⇒ "Bayes"

⇒ Intuition might fail [*] – it does often fail! – and the sigmas are not (always) a good guidance!

BUT the intuition of experienced scientists is in most cases far superior than the aseptic/pedantic rules of statisticians.

⇒ Informative priors!

‘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 “significance” 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 “significance” 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 (1985)

Testing one hypothesis

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

Testing one hypothesis

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

⇒ search for violations of H_0

Testing one hypothesis

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

  ⇒ search for violations of H_0
- Ideally

  → 'falsify' H_0

Testing one hypothesis

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

  ⇒ search for violations of H_0
- Ideally

  → 'falsify' H_0
- In practice:

  → does it make sense?

  → how is it done?

Testing one hypothesis

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

  ⇒ search for violations of H_0
- Ideally
  → 'falsify' H_0
- In practice:
  → does it make sense?
  → how is it done?

Let's review the practice and what is behind it ⇒

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 if 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 if 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.

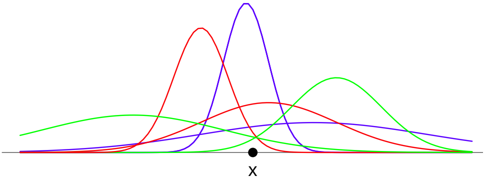
Falsificationism? OK, but. . .

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

$\Rightarrow$ 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 credible (and hopefully fruitful) at a given moment. A behavior "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:

⇒ 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** $C_i \not\rightarrow E$, 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:

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

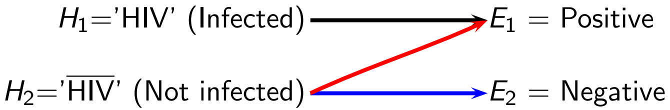
H_1='HIV' (Infected) $\qquad E_1$ = Positive

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

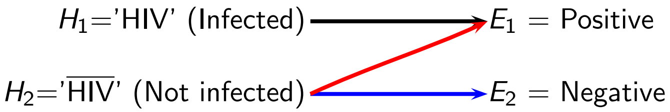
$$\begin{aligned} P(\text{Pos}\,|\,\text{HIV}) &= 100\% \\ P(\text{Pos}\,|\,\overline{\text{HIV}}) &= 0.2\% \\ P(\text{Neg}\,|\,\overline{\text{HIV}}) &= 99.8\% \end{aligned}$$



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 <u>at random</u> and sent to take an AIDS test (test is not perfect, as it is the case in practice).
Simplified model:

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

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

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

Result: ⇒ <u>Positive</u>

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 $P(\text{Pos} \mid \overline{\text{HIV}}) = 0.2\%$ 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"?

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 $P(\text{Pos} \mid \overline{\text{HIV}}) = 0.2\%$ 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"
- "Hypothesis H_1=Healthy is ruled out with 99.8% C.L."

? NO

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

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%)

$P(A \mid B) \leftrightarrow P(B \mid A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \mid B) \neq P(B \mid A)$

$P(A \mid B) \leftrightarrow P(B \mid A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \mid B) \neq P(B \mid A)$

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

$P(A \mid B) \leftrightarrow P(B \mid A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \mid B) \neq P(B \mid A)$

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

$P(A \mid B) \leftrightarrow P(B \mid A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \mid B) \neq P(B \mid A)$

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

$P(A \mid B) \leftrightarrow P(B \mid A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \mid B) \neq P(B \mid A)$

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

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

Typical values of statistical practice to reject a hypothesis are 5%, 1%, . . .

'Low probability' events

Typical values of statistical practice to reject a hypothesis are 5%, 1%, ...

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

'Low probability' events

Typical values of statistical practice to reject a hypothesis are 5%, 1%, ...

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

Typical values of statistical practice to reject a hypothesis are 5%, 1%, ...

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

Typical values of statistical practice to reject a hypothesis are 5%, 1%, . . .

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

Typical values of statistical practice to reject a hypothesis are 5%, 1%, ...

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 $\approx 39 \times 10^{-6}$;

'Low probability' events

Typical values of statistical practice to reject a hypothesis are 5%, 1%, ...

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 $\approx 39 \times 10^{-6}$;
- 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 'technical 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 'technical 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 'technical 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 matters 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 'technical 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:
> → what matters 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. . .):

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

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 'technical 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 'technical 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...

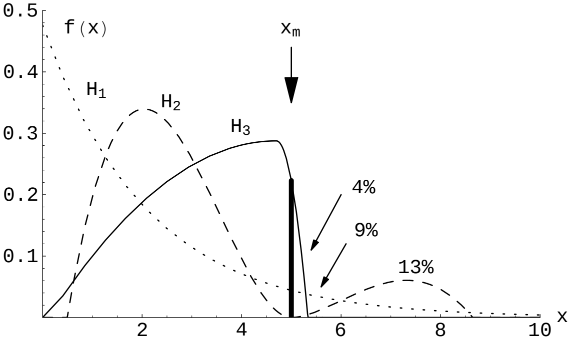
Probability of something else...

Besides the fact that the reasoning based only on the probability of the event given the cause is logically flawed, the 'technical 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...
- ⇒ From the logical point of view the situation has worsened:
  → our conclusions do not depend on what we have observed, but also from rarer events not actually observed!

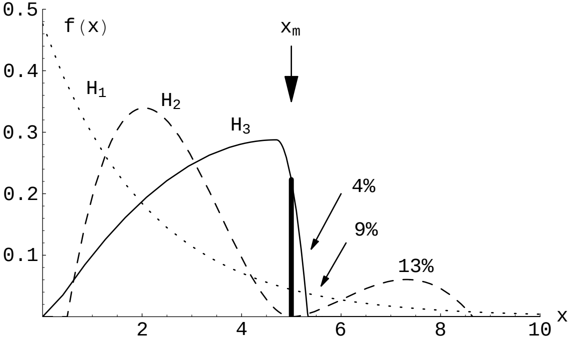
Comparing three hypotheses

Which hypothesis is favored by the experimental observation x_m?



Comparing three hypotheses

Which hypothesis is favored by the experimental observation x_m?

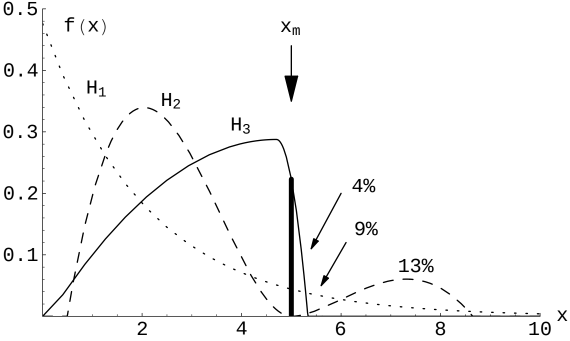


$$P(x_m \mid H_3) > P(x_m \mid H_1) > P(x_m \mid H_2) = 0 \quad (\,!\,)$$

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

Comparing three hypotheses

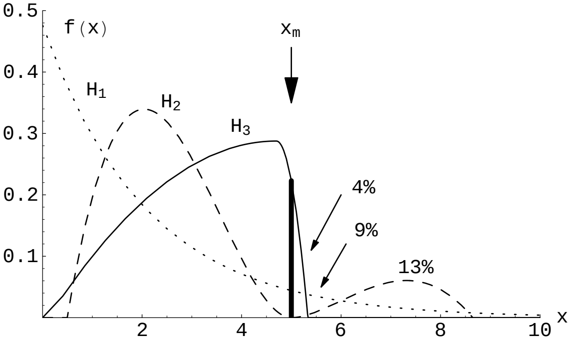
Which hypothesis is favored by the experimental observation x_m?



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

Comparing three hypotheses

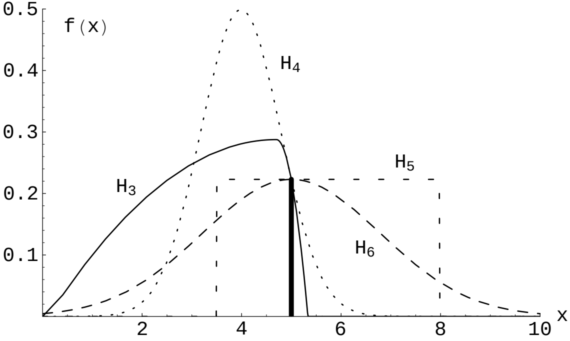
Which hypothesis is favored by the experimental observation x_m?



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

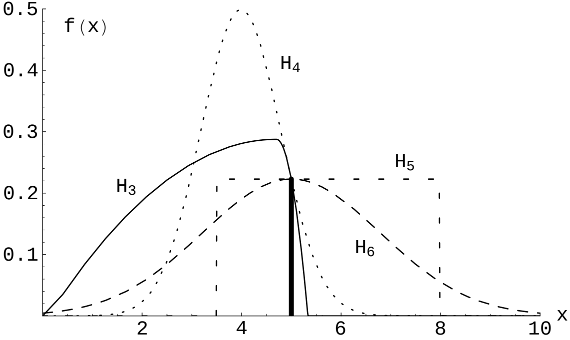
An irrelevant experiment

Which hypothesis is favored by the experimental observation x_m?



An irrelevant experiment

Which hypothesis is favored by the experimental observation x_m?

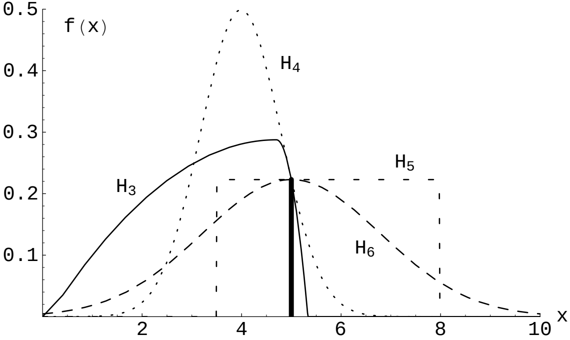


$$P(x_m \,|\, H_3) = P(x_m \,|\, H_4) = P(x_m \,|\, H_5) = P(x_m \,|\, H_6)$$

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

An irrelevant experiment

Which hypothesis is favored by the experimental observation x_m?

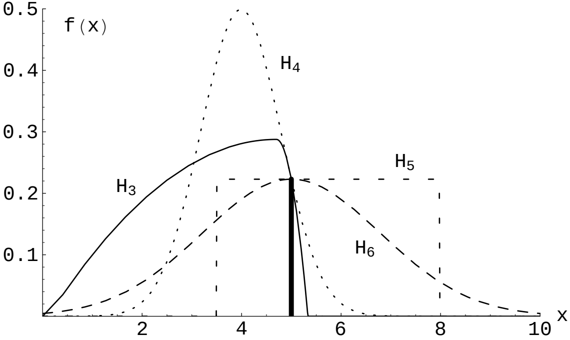


$$P(x_m \mid H_3) = P(x_m \mid H_4) = P(x_m \mid H_5) = P(x_m \mid H_6)$$

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

An irrelevant experiment

Which hypothesis is favored by the experimental observation x_m?



$$P(x_m \mid H_3) = P(x_m \mid H_4) = P(x_m \mid H_5) = P(x_m \mid H_6)$$

⇒ *The experimental result is irrelevant!*
⇒ *...no matter what the different 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?

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

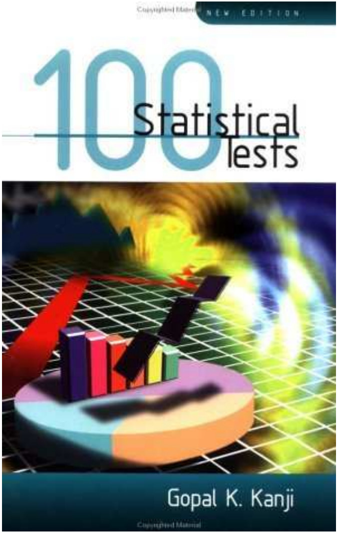
$$\theta = \theta(\mathbf{x})$$

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

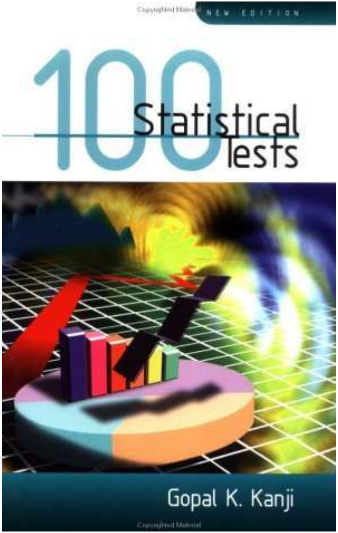
$$\text{Experiment:} \rightarrow \theta_{mis} = \theta(\mathbf{x}_{mis})$$

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

Which p-value?...

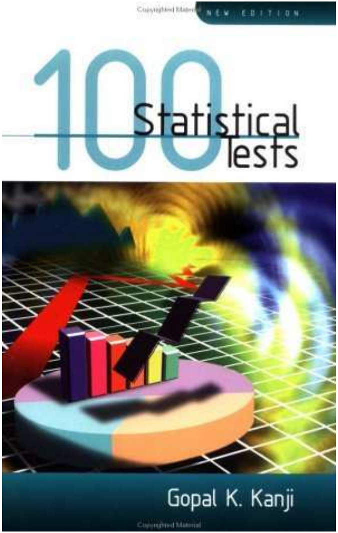


Which p-value?...



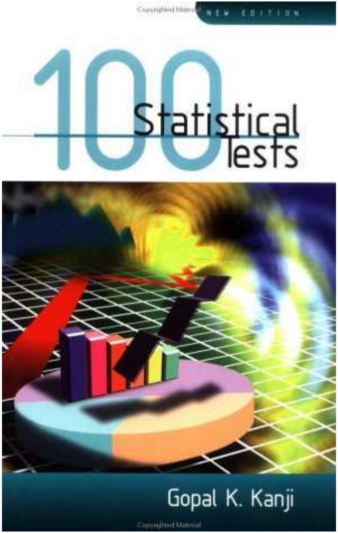
- far from exhaustive list,

Which p-value?...



- far from exhaustive list,
- with arbitrary variants:

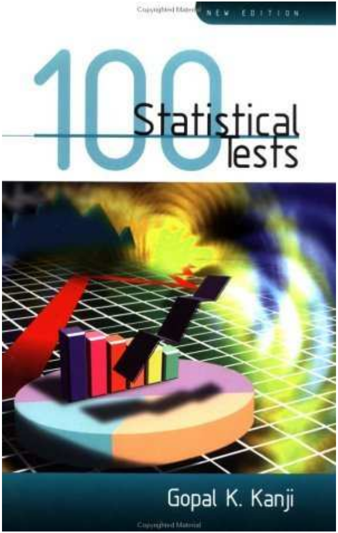
Which p-value?...



- far from exhaustive list,
- with arbitrary variants:

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

Which p-value?...



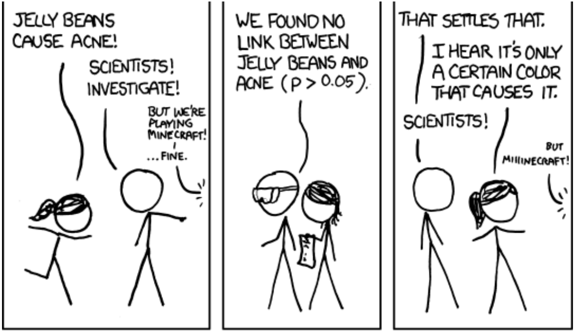
- far from exhaustive list,
- with arbitrary variants:

⇒ practitioners 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 sure that several 'normal' tests have been discarded in the meanwhile...

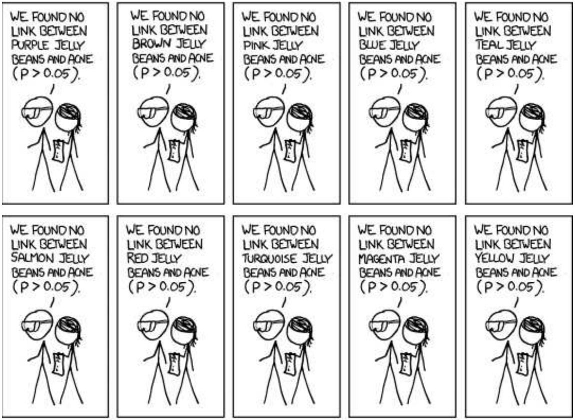
Or look around, searching for 'significance'

If changing the test does not help, change hypotheses...



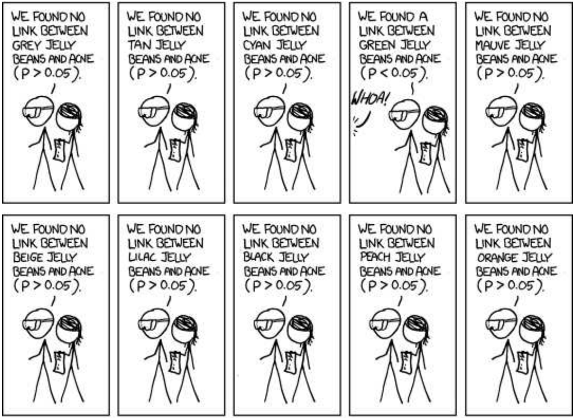
Or look around, searching for 'significance'

If changing the test does not help, change hypotheses...



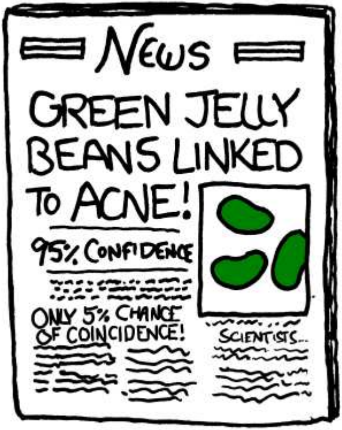
Or look around, searching for 'significance'

If changing the test does not help, change hypotheses...



Or look around, searching for‘significance’

If changing the test does not help, change hypotheses...



P-hacking ("p-value hacking")

The 'science' of inventing significant results. . .

p-hacking, or cheating on a p-value

June 11, 2015

By arthur charpentier

Share

(This article was first published on **Freakonometrics » R-english**, and kindly contributed to R-bloggers)

Yesterday evening, I discovered some interesting slides on False-Positives, p-Hacking, Statistical Power, and Evidential Value, via @UCBITSS 's post on Twitter. More precisely, there was this slide on how cheating (because that's basically what it is) to get a 'good' model (by targeting the *p*-value)

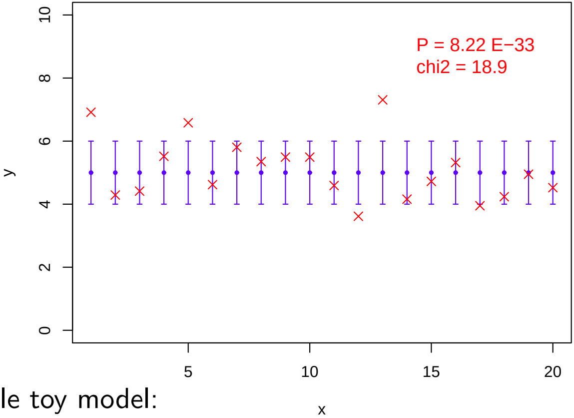
1. Stop collecting data once $p<.05$
2. Analyze many measures, but report only those with $p<.05$.
3. Collect and analyze many conditions, but only report those with $p<.05$.
4. Use covariates to get $p<.05$.
5. Exclude participants to get $p<.05$.
6. Transform the data to get $p<.05$.

http://www.r-bloggers.com/p-hacking-or-cheating-on-a-p-value/

- Google for "p-hacking"

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 'resolution' Δy (instrumental 'discretization'):

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

'Distance' Experiment-theory: χ^2

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

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

$$\chi^2 \sim \Gamma(\nu/2, 1/2) \qquad [\rightarrow \nu = 20]$$

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

$$\mathrm{Var}[\chi^2] = 2\nu \qquad [\rightarrow 40]$$

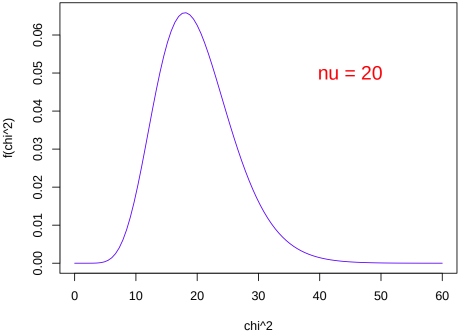
$$\mathrm{Std}[\chi^2] = \sqrt{2\nu} \qquad [\rightarrow 6.3]$$

$$\mathrm{Mode}[\chi^2] = \begin{cases} 0 & \text{if } \nu \leq 2 \\ \nu - 2 & \text{if } \nu > 2 \end{cases} \qquad [\rightarrow 18]$$

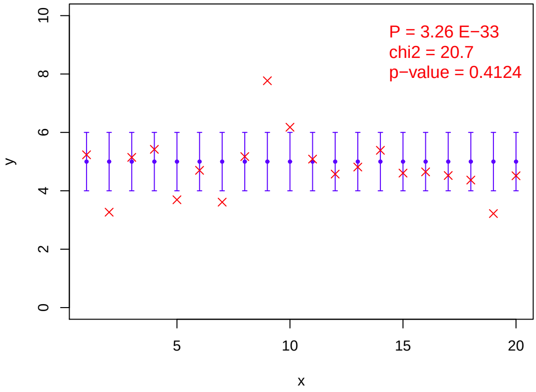
$$\Rightarrow \quad \boxed{\chi^2 = 20 \pm 6}$$

Our expectations about χ^2

$$
\begin{aligned}
\mathrm{E}[\chi^2] &= \nu & [\rightarrow 20] \\
\mathrm{Std}[\chi^2] &= \sqrt{2\nu} & [\rightarrow 6.3] \\
\Rightarrow & \quad \boxed{\chi^2 = 20 \pm 6} \\
& \quad [\,\text{mode: } 18\,]
\end{aligned}
$$

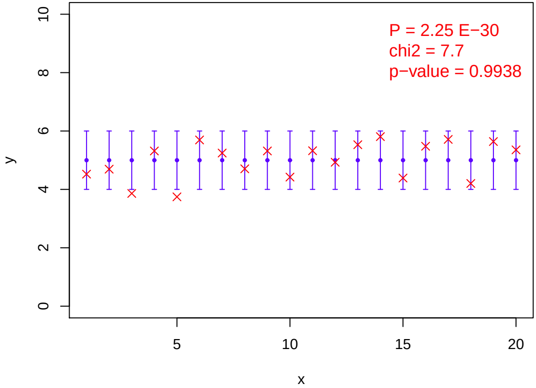


Some examples



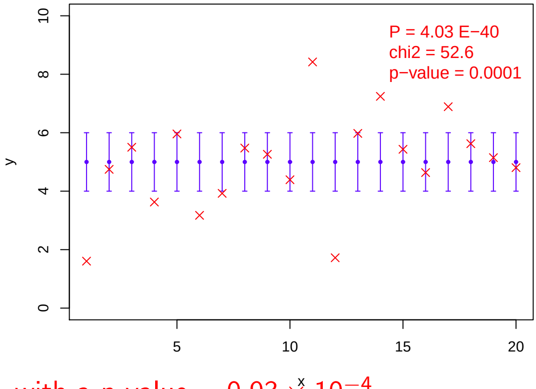
In the average.
(but someone could see the points forming a 'constellation'. . .)

Some examples



Too good?

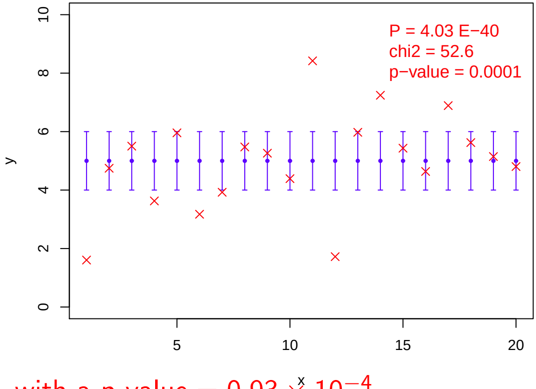
Some examples



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

At limit?

Some examples

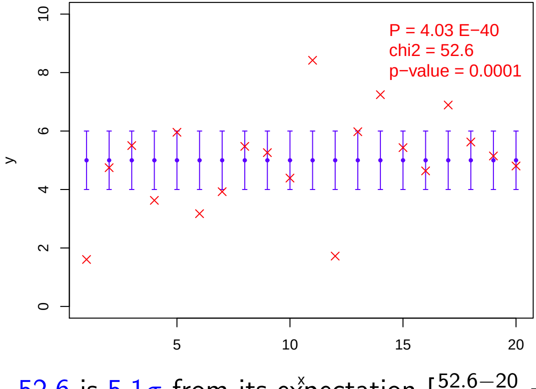


$\chi^2 = 52.6$, with a p-value $= 0.93 \times 10^{-4}$

At limit? Just come out at the first time (9 Oct. 2012, 13:01)

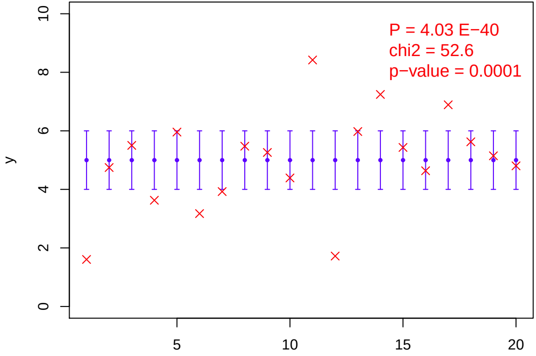
```
while(chi2.ym() < 38) source("chi2_1.R")
```

Some examples



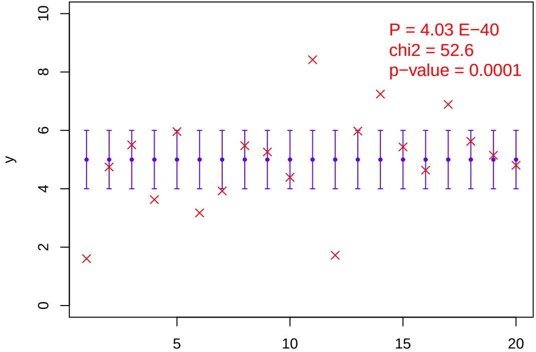
Note: χ^2_{mis} 52.6 is 5.1σ from its expectation $[\frac{52.6-20}{\sqrt{40}} = 5.1]$

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 communicated as "$3.7\,\sigma$", referring to the probability of the tail above $3.7\,\sigma$ of an 'equivalent Gaussian'.

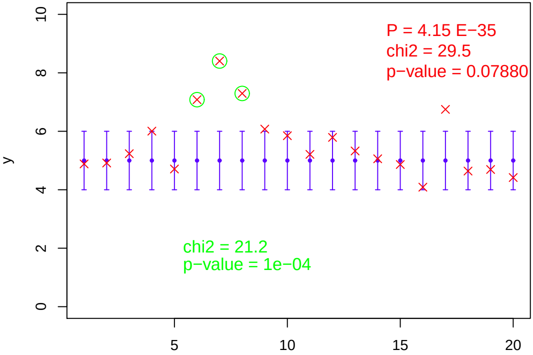
Some examples



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 communicated as "$3.7\,\sigma$", referring to the probability of the tail above $3.7\,\sigma$ of an 'equivalent Gaussian'.
(as if there were already not enough confusion. . .)

The art of χ^2

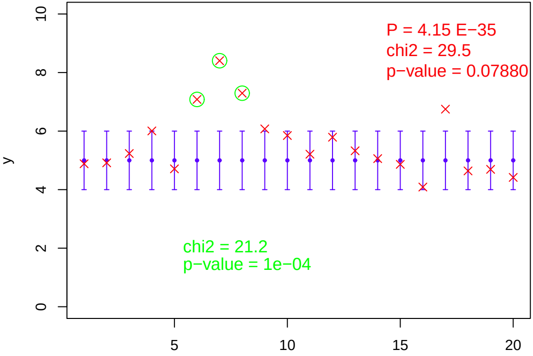
Sometimes the χ^2 test does not give "the wished result"



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'

⇒ If we add the two side points, χ^2 becomes 22.2.

⇒ 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"):

$$
\begin{aligned}
P(\theta \geq \theta_{mis}) &= \int_{\theta_{mis}}^{\infty} f(\theta \,|\, H_0)\, d\theta \\
P[(\theta \geq \theta_{mis}) \cup (\theta \leq (\theta^c)_{mis})] &= 1 - \int_{(\theta^c)_{mis}}^{\theta_{mis}} f(\theta \,|\, H_0)\, d\theta
\end{aligned}
$$

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

p-value: what they are

p-value:

- Probability of the tail(s) of a 'test variable' (a "statistic"):

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

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

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

What they are not ⇒

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:

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

Example: Has the student made a mistake?

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

- Teacher expectation:

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

- 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

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

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

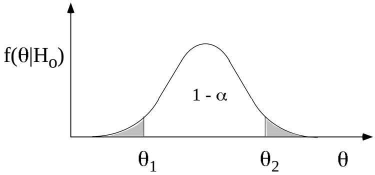
$\Rightarrow$ Has the student made a mistake?

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?

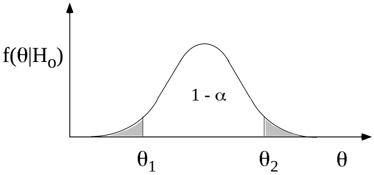
Conventional statistician solution:
$\Rightarrow$ test the hypothesis $H_0 =$ 'no mistakes'



- 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'

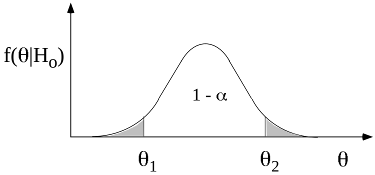


- Test variable θ is $\overline{X}_{300}$.
- Acceptance interval $[\theta_1, \theta_2]$ is $[0.456, 0.544]$.
  We are 99% confident that $\overline{X}_{300}$ will fall inside it:
  $\rightarrow \alpha = 1\%$.
- $\overline{x} = 0.550$ lies outside the acceptance interval

$\Rightarrow$ 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).

⇒ 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 $P(|X| > 3) = 0.27\%$, but

'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 $P(|X| > 3) = 0.27\%$, 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%”

'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%"

⇒ 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 $P(|X| > 3) = 0.27\%$, 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%"

⇒ the value comes with 100% probability from that generator!

⇒ it is at 100% a statistical fluctuation

Logical bug of the reasoning:

⇒ 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:

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

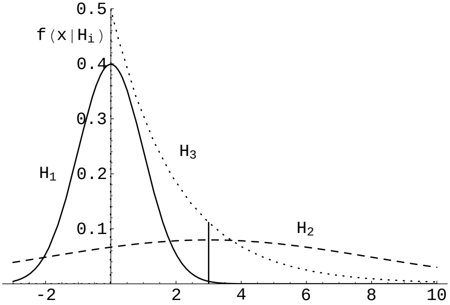
$\Rightarrow$ This is the original sin of conventional hypothesis test methods

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 $x = 3$:

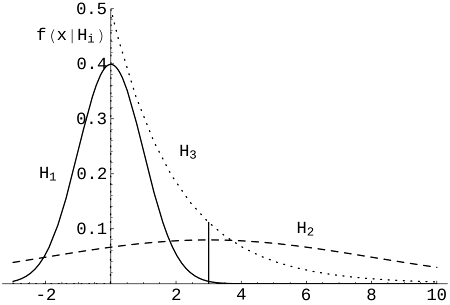


The statistics-uneducated student would suggest:

- 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 $x = 3$:

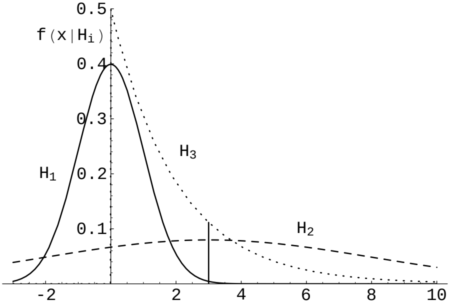


The statistics-uneducated student would suggest:

- 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."

⇒ logic is logic!

Objections

"These are chosen academic examples."

⇒ logic is logic!

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

⇒ All fake claims of discoveries are due to the criticized reasoning

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

"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

"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

Many, too many, unfortunately... 

I case I lived in first person was that of the (in)famous HERA events

⇒ see slides at `http://www.roma1.infn.it/~dagos/cernAT05_scanned/`

Examples from particle physics

Many, too many, unfortunately... 

I case I lived in first person was that of the (in)famous HERA events

⇒ see slides at
`http://www.roma1.infn.it/~dagos/cernAT05_scanned/`

(And the logical error happens not only in the case of fake discoveries, but also when a highly expected particle is finally found – wait for a while. . .)

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).

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:

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

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

- What we wanted:
  - falsify the hypothesis H_0:
    ⇒ 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:

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

⇒ BUT the p-value do not provide this:

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

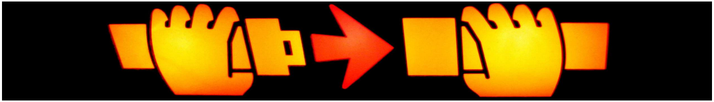
⇒ Although they are erroneously confused with this!

p-value: what they are not

- What we wanted:
  - falsify the hypothesis H_0:
    ⇒ 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#Misunderstandings

Misunderstandings p-values

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

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

Misunderstandings p-values

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

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#Misunderstandings

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#Misunderstandings

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 threshold, i.e. a probability of discovery of 99.99994%" (one of the many claims you could read on the web).
- "I dati confermano la soglia dei 5 sigma, vale a dire una probabilità di scoperta pari al 99,99994 per cento" spiega Gian Francesco Giudice, teorico del CERN (corriere.it, 3 luglio)

The 5 sigma Higgs!

July 2012

- “The data confirm the 5 sigma threshold, i.e. a probability of discovery of 99.99994%” (one of the many claims you could read on the web).
- “I dati confermano la soglia dei 5 sigma, vale a dire una probabilità di scoperta pari al 99,99994 per cento” spiega Gian Francesco Giudice, teorico del CERN (corriere.it, 3 luglio)
- “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 threshold, i.e. a probability of discovery of 99.99994%" (one of the many claims you could read on the web).
- "I dati confermano la soglia dei 5 sigma, vale a dire una probabilità di scoperta pari al 99,99994 per cento" spiega Gian Francesco Giudice, teorico del CERN (corriere.it, 3 luglio)
- "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 (July 2014)
  - "higgs cern 0.00006 chance": $\approx 1.6 \times 10^4$ results

The 5 sigma Higgs!

July 2012

- "The data confirm the 5 sigma threshold, i.e. a probability of discovery of 99.99994%" (one of the many claims you could read on the web).
- "I dati confermano la soglia dei 5 sigma, vale a dire una probabilità di scoperta pari al 99,99994 per cento" spiega Gian Francesco Giudice, teorico del CERN (corriere.it, 3 luglio)
- "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 (July 2014)
  - "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`

“Is the ‘new particle’ the Higgs?”

We have often listened after July 2012 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 after July 2012 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 after July 2012 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 $\approx 126\,\mathrm{GeV}$ 'excess' is a new particle comes from the 5 sigmas alone.

"Is the 'new particle' the Higgs?"

We have often listened after July 2012 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 $\approx 126\,\text{GeV}$ 'excess' is a new particle comes 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 after July 2012 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 $\approx 126\,\text{GeV}$ ‘excess’ is a new particle comes 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’ the Higgs?”

We have often listened after July 2012 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 $\approx 126\,\text{GeV}$ ‘excess’ is a new particle comes from the 5 sigmas alone.

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

- success of standard model;
- radiative corrections
  - the diagrams entering radiative corrections are essentially the same that produce the Higgs in the final state!
  - the mass found ($\approx 126\,\text{GeV}$) falls right in the middle of that inferred from indirect processes! (GdA & Degrassi, 1999)

"Is the 'new particle' the Higgs?"

We have often listened after July 2012 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 $\approx 126\,\mathrm{GeV}$ 'excess' is a new particle comes from the 5 sigmas alone.

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

- success of standard model;
- radiative corrections
  - the diagrams entering radiative corrections are essentially the same that produce the Higgs in the final state!
  - the mass found ($\approx 126\,\mathrm{GeV}$) falls right in the middle of that inferred from indirect processes! (GdA & Degrassi, 1999)
- Physics is something SERIOUS! (not a toy for statisticians)

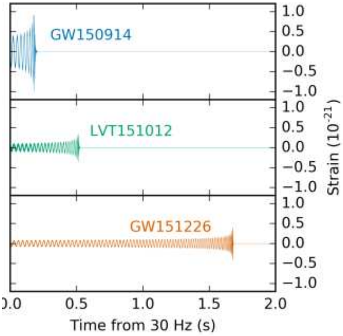
2011 $\rightarrow$ 2016: remarkable events during this year

(From a personally biased point of view. . .)

2011 → 2016: remarkable events during this year

(From a personally biased point of view. . .)

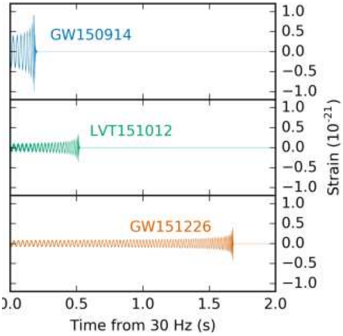
- Announcement(s) of Gravitational Wave detection



2011 → 2016: remarkable events during this year

(From a personally biased point of view. . .)

- Announcement(s) of Gravitational Wave detection



- American Statistical Association's statement on p-values

The waves and the sigmas

Ironically the last two events are in odds with each other.

The waves and the sigmas

Ironically the last two events are in odds with each other.

- The February 11 announcement by LIGO-Virgo puts great emphasis on the "5.1 σ's" as a figure of evidence.

The waves and the sigmas

Ironically the last two events are in odds with each other.

- The February 11 announcement by LIGO-Virgo puts great emphasis on the "5.1 σ's" as a figure of evidence.
  [The *desired* number of sigmas was achieved using a kind of frequentistic stopping rule, after the September 14 event was observed.]

The waves and the sigmas

Ironically the last two events are in odds with each other.

- The February 11 announcement by LIGO-Virgo puts great emphasis on the "5.1 σ's" as a figure of evidence.
  [The *desired* number of sigmas was achieved using a kind of frequentistic stopping rule, after the September 14 event was observed.]
- Less than four weeks later (March 7) the American Statistical Association came out with a strong statement warning scientists about interpretation and misuse of p-values (more or less what it has been in the Wiki since years).

The ASA statement on p-values

→ For details please read the paper, very easy to find and freely downloadable.

The ASA statement on p-values

→ For details please read the paper, very easy to find and freely downloadable.

- I report here just some a quote from Nature to stress its importance:

The ASA statement on p-values

→ For details please read the paper, very easy to find and freely downloadable.

- I report here just some a quote from Nature to stress its importance:


nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | Fo

Archive > Volume 531 > Issue 7593 > News > Article

NATURE | NEWS

Statisticians issue warning over misuse of *P* values

The ASA statement on p-values

→ For details please read the paper, very easy to find and freely downloadable.

- I report here just some a quote from Nature to stress its importance:


nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | Fo

Archive > Volume 531 > Issue 7593 > News > Article

NATURE | NEWS

Statisticians issue warning over misuse of *P* values

> *"This is the first time that the 177-year-old ASA has made explicit recommendations on such a foundational matter in statistics, says executive director Ron Wasserstein. The society's members had become increasingly concerned that the P value was being misapplied in ways that cast doubt on statistics generally, he adds." (March 7 2016)*

How believable is the 14 September event?

How believable is the 14 September event?

Very very much

How believable is the 14 September event?

Very very much

. . . but not because of the sigmas. . .

How believable is the 14 September event?

Very very much

. . . but not because of the sigmas. . .

1. Ask the opinion of reliable experienced people of the LIGO-Virgo Collaboration.

How believable is the 14 September event?

Very very much

...but not because of the sigmas...

1. Ask the opinion of reliable experienced people of the LIGO-Virgo Collaboration.
   - The rumor I had got *after two months* from the detection was they had seen a Monster!
   - After the announcement we have all learned that background was 'ruled out' and the only serious alternative hypothesis had been for some time that of a sabotage – really!

How believable is the 14 September event?

Very very much

. . . but not because of the sigmas. . . 🙂

1. Ask the opinion of reliable experienced people of the LIGO-Virgo Collaboration.
   - The rumor I had got *after two months* from the detection was they had seen a Monster!
   - After the announcement we have all learned that background was 'ruled out' and the only serious alternative hypothesis had been for some time that of a sabotage – really!
2. Yes, you might argue, but what does it means 'ruled out'?

How believable is the 14 September event?

Very very much

...but not because of the sigmas... ☺

1. Ask the opinion of reliable experienced people of the LIGO-Virgo Collaboration.
   - The rumor I had got *after two months* from the detection was they had seen a Monster!
   - After the announcement we have all learned that background was 'ruled out' and the only serious alternative hypothesis had been for some time that of a sabotage – really!
2. Yes, you might argue, but what does it means 'ruled out'?

   → The Collaboration provided a number
   - based on probability theory

How believable is the 14 September event?

Very very much

. . . but not because of the sigmas. . .

1. Ask the opinion of reliable experienced people of the LIGO-Virgo Collaboration.
   - The rumor I had got *after two months* from the detection was they had seen a Monster!
   - After the announcement we have all learned that background was 'ruled out' and the only serious alternative hypothesis had been for some time that of a sabotage – really!
2. Yes, you might argue, but what does it means 'ruled out'?

   → The Collaboration provided a number
   - based on probability theory. . .
   - that tells us how we must update our rational belief on an hypothesis Vs its complementary one.

How believable is the 14 September event?

Very very much

. . . but not because of the sigmas. . . ☺

1. Ask the opinion of reliable experienced people of the LIGO-Virgo Collaboration.
   - The rumor I had got *after two months* from the detection was they had seen a Monster!
   - After the announcement we have all learned that background was 'ruled out' and the only serious alternative hypothesis had been for some time that of a sabotage – really!
2. Yes, you might argue, but what does it means 'ruled out'?

   → The Collaboration provided a number
   - based on probability theory. . .
   - that tells us how we <u>must</u> update our <u>rational belief</u> on an hypothesis Vs its complementary one.
   - Bayes factor

Interpretation of Bayes factors

(Details on how the arise → arXiv:1609.01668)

- If you consider two alternative hypotheses, H and $\overline{H}$, equally likely, for all <u>other</u> reasons you might have in your mind, the odds you assign to the two hypotheses (H Vs $\overline{H}$) are 1 to 1, or simply 1.

Interpretation of Bayes factors

(Details on how the arise → arXiv:1609.01668)

- If you consider two alternative hypotheses, H and $\overline{H}$, equally likely, for all other reasons you might have in your mind, the odds you assign to the two hypotheses (H Vs $\overline{H}$) are 1 to 1, or simply 1.
- A Bayes factor (**BF**) (H Vs $\overline{H}$) based on the piece of evidence E simply multiplies the 'prior odds' in order to provide 'posterior odds' (subject to extra piece of evidence E)

Interpretation of Bayes factors

(Details on how the arise → arXiv:1609.01668)

- If you consider two alternative hypotheses, H and $\overline{H}$, equally likely, for all <u>other</u> reasons you might have in your mind, the odds you assign to the two hypotheses (H Vs $\overline{H}$) are 1 to 1, or simply 1.
- A Bayes factor (**BF**) (H Vs $\overline{H}$) based on the piece of evidence E simply multiplies the 'prior odds' in order to provide 'posterior odds' (subject to extra piece of evidence E)

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

Interpretation of Bayes factors

(Details on how the arise → arXiv:1609.01668)

- If you consider two alternative hypotheses, H and $\overline{H}$, equally likely, for all <u>other</u> reasons you might have in your mind, the odds you assign to the two hypotheses (H Vs $\overline{H}$) are 1 to 1, or simply 1.
- A Bayes factor (**BF**) (H Vs $\overline{H}$) based on the piece of evidence E simply multiplies the 'prior odds' in order to provide 'posterior odds' (subject to extra piece of evidence E)

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

- Examples
  1. Prior odds 1 (H and $\overline{H}$ equally likely); $BF = 100$

Interpretation of Bayes factors

(Details on how the arise → arXiv:1609.01668)

- If you consider two alternative hypotheses, H and $\overline{H}$, equally likely, for all <u>other</u> reasons you might have in your mind, the odds you assign to the two hypotheses (H Vs $\overline{H}$) are 1 to 1, or simply 1.
- A Bayes factor (**BF**) (H Vs $\overline{H}$) based on the piece of evidence E simply multiplies the 'prior odds' in order to provide 'posterior odds' (subject to extra piece of evidence E)

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

- Examples
  1. Prior odds 1 (H and $\overline{H}$ equally likely); $BF = 100$
     - posterior odds 100, i.e. 100:1;
     - H 100 times more probable than $\overline{H}$

Interpretation of Bayes factors

(Details on how the arise → arXiv:1609.01668)

- If you consider two alternative hypotheses, H and $\overline{H}$, equally likely, for all <u>other</u> reasons you might have in your mind, the odds you assign to the two hypotheses (H Vs $\overline{H}$) are 1 to 1, or simply 1.
- A Bayes factor (**BF**) (H Vs $\overline{H}$) based on the piece of evidence E simply multiplies the 'prior odds' in order to provide 'posterior odds' (subject to extra piece of evidence E)

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

- Examples
  1. Prior odds 1 (H and $\overline{H}$ equally likely); $BF = 100$
     - posterior odds 100, i.e. 100:1;
     - H 100 times more probable than $\overline{H}$
  2. Prior odds $1/1000$ (H 10^3 times <u>less</u> likely than $\overline{H}$); $BF = 10^6$

Interpretation of Bayes factors

(Details on how the arise → arXiv:1609.01668)

- If you consider two alternative hypotheses, H and $\overline{H}$, equally likely, for all <u>other</u> reasons you might have in your mind, the odds you assign to the two hypotheses (H Vs $\overline{H}$) are 1 to 1, or simply 1.
- A Bayes factor (**BF**) (H Vs $\overline{H}$) based on the piece of evidence E simply multiplies the 'prior odds' in order to provide 'posterior odds' (subject to extra piece of evidence E)

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

- Examples
  1. Prior odds 1 (H and $\overline{H}$ equally likely); $BF = 100$
     - posterior odds 100, i.e. 100:1;
     - H 100 times more probable than $\overline{H}$
  2. Prior odds $1/1000$ (H 10^3 times <u>less</u> likely than $\overline{H}$); $BF = 10^6$
     - posterior odds 1000, i.e. 1000:1;
     - H 10^3 times <u>more</u> probable than $\overline{H}$

The Bayes factor of a Monster

Now that we have learned *how to use* BF's,

The Bayes factor of a Monster

Now that we have learned *how to use* BF's, relax, breathe deeply

The Bayes factor of a Monster

Now that we have learned *how to use* BF's, relax, breathe deeply and than read the number

The Bayes factor of a Monster

Now that we have learned *how to use* BF's, relax, breathe deeply and than read the number

$$BF(\text{BBH merger Vs Noise}) \approx 5 \times 10^{125}$$

The Bayes factor of a Monster

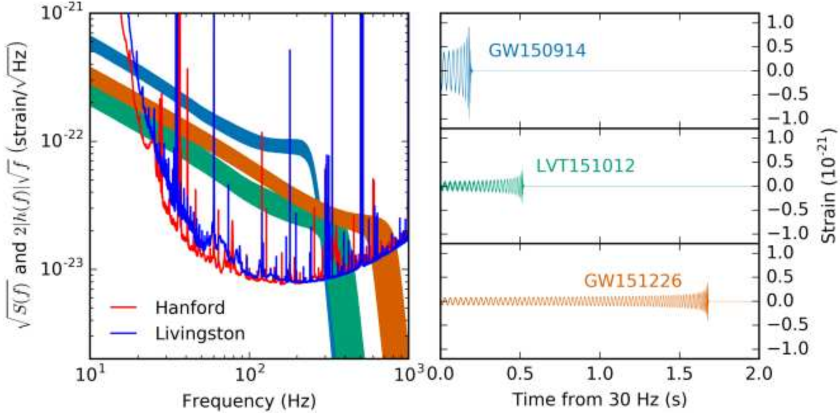
Now that we have learned *how to use* BF's, relax, breathe deeply and than read the number

$$BF(\text{BBH merger Vs Noise}) \approx 5 \times 10^{125}$$

(Definitely a **Monster**!)

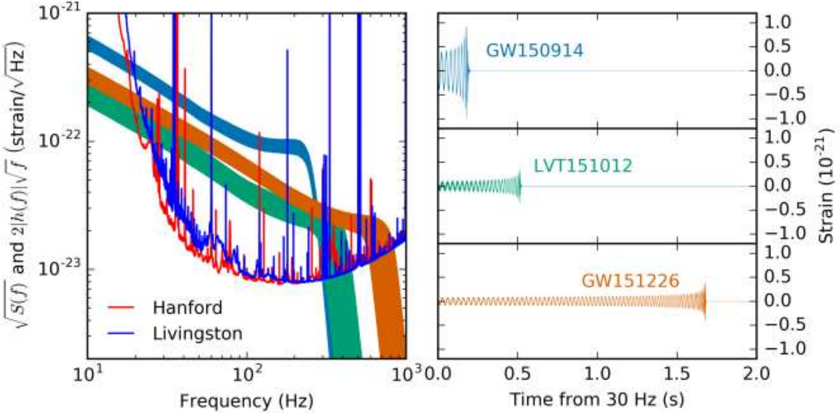
Cinderella and her two sisters

Here is the (official) status of the search in date June 8:



Cinderella and her two sisters

Here is the (official) status of the search in date June 8:



Note how the second event doesn't look bad, but nevertheless it does not deserve the 'title' GW, and it simply ranked as LVT, for "LIGO-Virgo Trigger".

What is the problem with Cinderella?

The October 12 event cannot wear a "$> 5\sigma$'s dress" to go to the sumptuous ball of the Establishment.

What is the problem with Cinderella?

The October 12 event cannot wear a "$> 5\sigma$'s dress" to go to the sumptuous ball of the Establishment.

Indeed, the poor event has only $1.7\,\sigma$'s, getting therefore no respect by particle physicists and those who control money and scientific journals.

What is the problem with Cinderella?

The October 12 event cannot wear a "$> 5\sigma$'s dress" to go to the sumptuous ball of the Establishment.
Indeed, the poor event has only $1.7\,\sigma$'s, getting therefore no respect by particle physicists and those who control money and scientific journals.

- But LIGO-Virgo experts did and <u>do believe</u> it is a genuine gravitational wave.

What is the problem with Cinderella?

The October 12 event cannot wear a "$> 5\sigma$'s dress" to go to the sumptuous ball of the Establishment.

Indeed, the poor event has only 1.7 σ's, getting therefore no respect by particle physicists and those who control money and scientific journals.

- But LIGO-Virgo experts did and do believe it is a genuine gravitational wave.
- And the Collaboration has even the courage of writing it in the paper (arXiv:1606.04856v2 – 'BBH': Binary Black Hole) :

What is the problem with Cinderella?

The October 12 event cannot wear a "$> 5\sigma$'s dress" to go to the sumptuous ball of the Establishment.

Indeed, the poor event has only 1.7 σ's, getting therefore no respect by particle physicists and those who control money and scientific journals.

- But LIGO-Virgo experts did and do believe it is a genuine gravitational wave.
- And the Collaboration has even the courage of writing it in the paper (arXiv:1606.04856v2 – 'BBH': Binary Black Hole) :
  - "it is more likely to have resulted from a gravitational-wave signal than from an instrumental or environmental noise transient."
  - "During its first observing run Advanced LIGO has observed gravitational waves from the coalescence of two stellar-mass BBHs GW150914 and GW151226 with a third candidate LVT151012 also likely to be a BBH system."

What is the problem with Cinderella?

The October 12 event cannot wear a "$>5\sigma$'s dress" to go to the sumptuous ball of the Establishment.

Indeed, the poor event has only 1.7 σ's, getting therefore no respect by particle physicists and those who control money and scientific journals.

- But LIGO-Virgo experts did and do believe it is a genuine gravitational wave.
- And the Collaboration has even the courage of writing it in the paper (arXiv:1606.04856v2 – 'BBH': Binary Black Hole) :
  - "it is more likely to have resulted from a gravitational-wave signal than from an instrumental or environmental noise transient."
  - "During its first observing run Advanced LIGO has observed gravitational waves from the coalescence of two stellar-mass BBHs GW150914 and GW151226 with a third candidate LVT151012 also likely to be a BBH system."
- How much likely? It will certainly depend on your (informed!) priors,

What is the problem with Cinderella?

The October 12 event cannot wear a "$>5\sigma$'s dress" to go to the sumptuous ball of the Establishment.

Indeed, the poor event has only $1.7\,\sigma$'s, getting therefore no respect by particle physicists and those who control money and scientific journals.

- But LIGO-Virgo experts did and do believe it is a genuine gravitational wave.
- And the Collaboration has even the courage of writing it in the paper (arXiv:1606.04856v2 – 'BBH': Binary Black Hole) :
  - "it is more likely to have resulted from a gravitational-wave signal than from an instrumental or environmental noise transient."
  - "During its first observing run Advanced LIGO has observed gravitational waves from the coalescence of two stellar-mass BBHs GW150914 and GW151226 with a third candidate LVT151012 also likely to be a BBH system."
- How much likely? It will certainly depend on your (informed!) priors, but a Bayes Factor $\approx 10^{10}$ is VERY impressive!

Sigmas vs beliefs

Particle physicists will be highly amazed to learn that their colleagues from GW community take seriously a $1.7\,\sigma$ event,

Sigmas vs beliefs

Particle physicists will be highly amazed to learn that their colleagues from GW community take seriously a $1.7\,\sigma$ event, while they are presently rather sceptical if their data show a $3\,\sigma$ 'bump', several decades ago considered the threshold to declared it a physical effect.

Sigmas vs beliefs

Particle physicists will be highly amazed to learn that their colleagues from GW community take seriously a $1.7\,\sigma$ event, while they are presently rather sceptical if their data show a $3\,\sigma$ 'bump', several decades ago considered the threshold to declared it a physical effect.

Why are our LIGO-Virgo colleagues essentially right?

Sigmas vs beliefs

Particle physicists will be highly amazed to learn that their colleagues from GW community take seriously a $1.7\,\sigma$ event, while they are presently rather sceptical if <u>their</u> data show a $3\,\sigma$ 'bump', several decades ago considered the threshold to declared it a physical effect.

Why are our LIGO-Virgo colleagues essentially right?

$\Rightarrow$ Physics!

Sigmas vs beliefs

Particle physicists will be highly amazed to learn that their colleagues from GW community take seriously a $1.7\,\sigma$ event, while they are presently rather sceptical if their data show a $3\,\sigma$ 'bump', several decades ago considered the threshold to declared it a physical effect.

Why are our LIGO-Virgo colleagues essentially right?

$\Rightarrow$ Physics!

- Detecting something that has good reason to exist, because of our understanding of the Physical World (related to a network of other experimental facts and theories connecting them!), is quite different than just observing an unexpected bump, possibly due to background (even if with small probability)!

Sigmas vs beliefs

Particle physicists will be highly amazed to learn that their colleagues from GW community take seriously a $1.7\,\sigma$ event, while they are presently rather sceptical if their data show a $3\,\sigma$ 'bump', several decades ago considered the threshold to declared it a physical effect.

Why are our LIGO-Virgo colleagues essentially right?

⇒ Physics!

- Detecting something that has good reason to exist, because of our understanding of the Physical World (related to a network of other experimental facts and theories connecting them!), is quite different than just observing an unexpected bump, possibly due to background (even if with small probability)!
  [Remember that all observed events of real life, if seen with high enough resolution in the N-dimensional phase space, had VERY VERY VERY small probability to occur!]

Sigmas Vs Bayes Factors

LVT151012 (Cinderella) has only 1.7σ, but a BF of $\approx 10^{10}$

Sigmas Vs Bayes Factors

LVT151012 (Cinderella) has only 1.7σ, but a BF of $\approx 10^{10}$

- How can it be?
- Which figure should be we use in order to make up our judgement if the observation is most likely due to BBH or to noise?

Sigmas Vs Bayes Factors

LVT151012 (Cinderella) has only 1.7σ, but a BF of $\approx 10^{10}$

- How can it be?
- Which figure should be we use in order to make up our judgement if the observation is most likely due to BBH or to noise?

Moreover

- Are Cinderella and her sisters independent?
(I mean *independent in probability*)

Sigmas Vs Bayes Factors

LVT151012 (Cinderella) has only 1.7σ, but a BF of $\approx 10^{10}$

- How can it be?
- Which figure should be we use in order to make up our judgement if the observation is most likely due to BBH or to noise?

Moreover

- Are Cinderella and her sisters independent?
  (I mean *independent in probability*)

More on arXiv:1609.01668 (*The sigmas and the waves*)

Fine

Addendum – 1

Bibliography

- GdA, Probably a discovery: Bad mathematics means rough scientific communication, arXiv:1112.3620;
- GdA, The Waves and the Sigmas (To Say Nothing of the 750 GeV Mirage), arXiv:1609.01668.

(Much more on `http://www.roma1.infn.it/~dagos/prob+stat.html`)

Addendum – 2

Solution to the AIDS test problem

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

We miss something: $P_\circ(\text{HIV})$ and $P_\circ(\overline{\text{HIV}})$: Yes! We need some input from our best knowledge of the problem. Let us take $P_\circ(\text{HIV}) = 1/600$ and $P_\circ(\overline{\text{HIV}}) \approx 1$ (the result is rather stable against *reasonable* variations of the inputs!)

$$\begin{aligned} \frac{P(\text{HIV} \mid \text{Pos})}{P(\overline{\text{HIV}} \mid \text{Pos})} &= \frac{P(\text{Pos} \mid \text{HIV})}{P(\text{Pos} \mid \overline{\text{HIV}})} \cdot \frac{P_\circ(\text{HIV})}{P_\circ(\overline{\text{HIV}})} \\ &= \frac{\approx 1}{0.002} \times \frac{0.1/60}{\approx 1} = 500 \times \frac{1}{600} = \frac{1}{1.2} \end{aligned}$$

Odd ratios and Bayes factor

$$
\begin{aligned}
\frac{P(\mathrm{HIV}\,|\,\mathrm{Pos})}{P(\overline{\mathrm{HIV}}\,|\,\mathrm{Pos})} &= \frac{P(\mathrm{Pos}\,|\,\mathrm{HIV})}{P(\mathrm{Pos}\,|\,\overline{\mathrm{HIV}})} \cdot \frac{P_\circ(\mathrm{HIV})}{P(\overline{\mathrm{HIV}})} \\
&= \frac{\approx 1}{0.002} \times \frac{0.1/60}{\approx 1} = 500 \times \frac{1}{600} = \frac{1}{1.2} \\
\Rightarrow P(\mathrm{HIV}\,|\,\mathrm{Pos}) &= 45.5\%\,.
\end{aligned}
$$

Odd ratios and Bayes factor

$$\begin{aligned}
\frac{P(\mathrm{HIV}\,|\,\mathrm{Pos})}{P(\overline{\mathrm{HIV}}\,|\,\mathrm{Pos})} &= \frac{P(\mathrm{Pos}\,|\,\mathrm{HIV})}{P(\mathrm{Pos}\,|\,\overline{\mathrm{HIV}})} \cdot \frac{P_\circ(\mathrm{HIV})}{P(\overline{\mathrm{HIV}})} \\
&= \frac{\approx 1}{0.002} \times \frac{0.1/60}{\approx 1} = 500 \times \frac{1}{600} = \frac{1}{1.2} \\
\Rightarrow P(\mathrm{HIV}\,|\,\mathrm{Pos}) &= 45.5\%\,.
\end{aligned}$$

There are some advantages in expressing Bayes theorem in terms of odd ratios:

- There is no need to consider all possible hypotheses (how can we be sure?)
  We just make a comparison of any couple of hypotheses!

Odd ratios and Bayes factor

$$\frac{P(\text{HIV}\,|\,\text{Pos})}{P(\overline{\text{HIV}}\,|\,\text{Pos})} = \frac{P(\text{Pos}\,|\,\text{HIV})}{P(\text{Pos}\,|\,\overline{\text{HIV}})} \cdot \frac{P_\circ(\text{HIV})}{P(\overline{\text{HIV}})}$$

$$= \frac{\approx 1}{0.002} \times \frac{0.1/60}{\approx 1} = 500 \times \frac{1}{600} = \frac{1}{1.2}$$

$$\Rightarrow P(\text{HIV}\,|\,\text{Pos}) = 45.5\%\,.$$

There are some advantages in expressing Bayes theorem in terms of odd ratios:

- There is no need to consider all possible hypotheses (how can we be sure?)
  We just make a comparison of any couple of hypotheses!
- Bayes factor is usually much more inter-subjective, and it is often considered an 'objective' way to report how much the data favor each hypothesis.

Addendum – 3

Numbers from LIGO-VIRGO

During the seminar it was pointed out by members of the LIGO-VIRGO collaborators that I was presenting

Wrong Bayes factors

This is an important statement, presented publically, which deserves public comments:

→ `http://www.roma1.infn.it/~dagos/prob+stat.html`