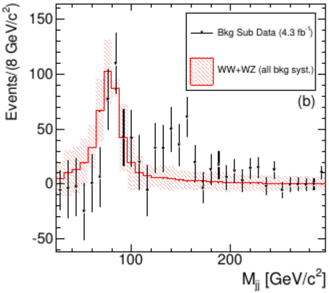
Misunderstanding probabilities - how to make a discovery out of nothing

Zuzana Kučerová & Georgios Kassaras

July 30, 2014

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 experiment at Fermilab, 4. 4. 2011, arXiv:1104.0699
- Excess in the region 120-160 GeV, figure shows the data after subtraction of background.
- Paper reports:
  - $\rightarrow$ p-value = 7.6 $\times 10^{-4}$
  - $\rightarrow$ significance = 3.2σ



Press reactions

- **New York Times**
  ... there is a less than a quarter of 1 % chance that the bump is a statistical fluctuation ...
- **Fermilab Today**
  ... there is less than 1 in 1375 chance that the effect is mimicked by a statistical fluctuation ...
- **Discovery News**
  ... this result has a 99.7 % chance of being correct (and a 0.3 % chance of being wrong) ...
- **Guardian**
  ... the paper quotes a 1 in 10 000 chance that this bump is a fluke ...

Misunderstanding p-value

p-value *IS NOT*:

- probability that the null hypothesis is true
- probability that the excess is only a fluctuation
- probability of falsely rejecting the null hypothesis
- probability that replicating the experiment would lead to the same conclusion
- a value according to which we should determine the significance level α
- something that would indicate the size or importance of the observed effect

Definition (p-value)

p-value is the probability of obtaining a test statistic at least as extreme as the one observed, assuming the null hypothesis is true. The null hypothesis is (in this case) only Standard Model (no new physics).

Frequentist statistics

- People are highly confused about the meaning of p-values.
- It is forbidden to speak of probability of causes or probability of true values.
- "There is a 1 in 10 000 chance that this bump is a fluke." $\rightarrow$ misinterpretation of p-value.

Bayesian statistics

- There must be more hypotheses, to each of them we assign the prior probability.
- If there is only one hypothesis H, probability, that observed data come from H is 1 (there is no other option).
- Probability of the hypothesis represents the state of knowledge and belief, therefore it might depend on person and time.
- If probability of an event given a hypothesis is 0, then if the event is directly observed, the hypothesis is ruled out.
- Probability can change in time, if some new relevant information appears.

Bayesian theory of probabilities

Probability of an event depends on person (subject) and time:

$$P\left[E \mid I_s\left(t\right)\right]$$

Updating probability of the model (or cause of the event) in the light of new information *(Bayes' theory)*:

$$P\left[C_i \mid E, I_s\left(t\right)\right] = \frac{P\left[E \mid C_i, I_s\left(t\right)\right] \times P\left[C_i \mid I_s\left(t\right)\right]}{\sum_j P\left[E \mid C_j, I_s\left(t\right)\right] \times P\left[C_j \mid I_s\left(t\right)\right]}$$

where

C_i is i-th **c**ause

E is an **e**vent leading to some new information

$I_s\left(t\right)$ is all the relevant **i**nformation available to the **s**ubject at time **t**

$P\left[C_i \mid I_s\left(t\right)\right]$ is the prior probability of the i-th cause set up according to the actual knowledge of the subject

$P\left[C_i \mid E, I_s\left(t\right)\right]$ is the posterior probability of the cause given the event

Comparing two models, Bayes' factor

We can compare probabilities of any two models, how they change after introducing new relevant information:

$$\frac{P\left[C_i \mid E\right]}{P\left[C_j \mid E\right]} = \frac{P\left[E \mid C_i\right]}{P\left[E \mid C_j\right]} \times \frac{P\left[C_i\right]}{P\left[C_j\right]}$$

The change in probability ratio is expressed by Bayes' factor

BF:

$$BF = \frac{P\left[E \mid C_i\right]}{P\left[E \mid C_j\right]}$$

Note

The updating of odds of the hypotheses depends only on the Bayes' factor *BF*, which is given only by that event, that has been actually observed.

Continuous case

In continuous case it is the probability density function (PDF) of a continuous parameter of the theory - θ - that may change. Therefore the Bayes' formula becomes:

$$\pi(\theta \mid x) = \frac{f(x \mid \theta)\,\pi(\theta)}{\int_{\Theta} f(x \mid \theta)\,\pi(\theta)\; d\theta}$$

where

$\pi(\theta)$ is prior PDF describing parameter θ

$f(x \mid \theta)$ is PDF to obtain the data given the parameter θ

In the denominator, we integrate over all possible values of parameter $\rightarrow$ normalization.

Higgs

Why did so many physicists believe, that the excess at 125 GeV was Higgs?

Higgs

Why did so many physicists believe, that the excess at 125 GeV was Higgs?   Because of the **prior** probabilities.

Higgs

Why did so many physicists believe, that the excess at 125 GeV was Higgs? Because of the **prior** probabilities.

- Higgs boson was the only particle missing in Standard Model.
- SM was (at that time and is now) the best theory with no disagreement with experiment.
- The excess was seen in two different experiments (ATLAS, CMS).
- The mass was compatible with all previous measurements and excluded regions.

All of these facts contributed to the prior probability, that Higgs boson exists (or that the SM prediction about the existence of Higgs is correct).

What to remember...?

What to remember...?

Definition (p-value)

p-value is the probability of obtaining a test statistic at least as extreme as the one observed, assuming the null hypothesis is true. The null hypothesis is (in this case) only Standard Model (no new physics).

What to remember...?

Definition (p-value)

p-value is the probability of obtaining a test statistic at least as extreme as the one observed, assuming the null hypothesis is true. The null hypothesis is (in this case) only Standard Model (no new physics).

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

What to remember...?

Definition (p-value)

p-value is the probability of obtaining a test statistic at least as extreme as the one observed, assuming the null hypothesis is true. The null hypothesis is (in this case) only Standard Model (no new physics).

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

If you calculate as *a frequentist*, do interpret the results as *a frequentist*!

D'Agostini, Giulio: *Probably a discovery: Bad mathematics means rough scientific communication.* 2011. arXiv:1112.3620v2