Statistical issues for Higgs Physics

Eilam Gross,
Weizmann Institute of Science

What is the statistical challenge in HEP?

- High Energy Physicists (HEP) have an hypothesis: The Standard Model.
- This model relies on the existence of the probably recently discovered, the Higgs Boson
- The minimal content of the Standard Model includes the Higgs Boson , but extensions of the Model include other particles which are yet to be discovered
- The challenge of HEP is to generate tons of data and to develop powerful analyses to tell if the data indeed contains evidence for the new particle, and confirm it's the expected Higgs Boson (Mass, Spin, CP)

The Large Hadron Collider (LHC)

The LHC is a very powerful accelerator aims to produce 10^9 proton-proton collisions per sec aiming to hunt a Higgs with a 10^{-12} production probability

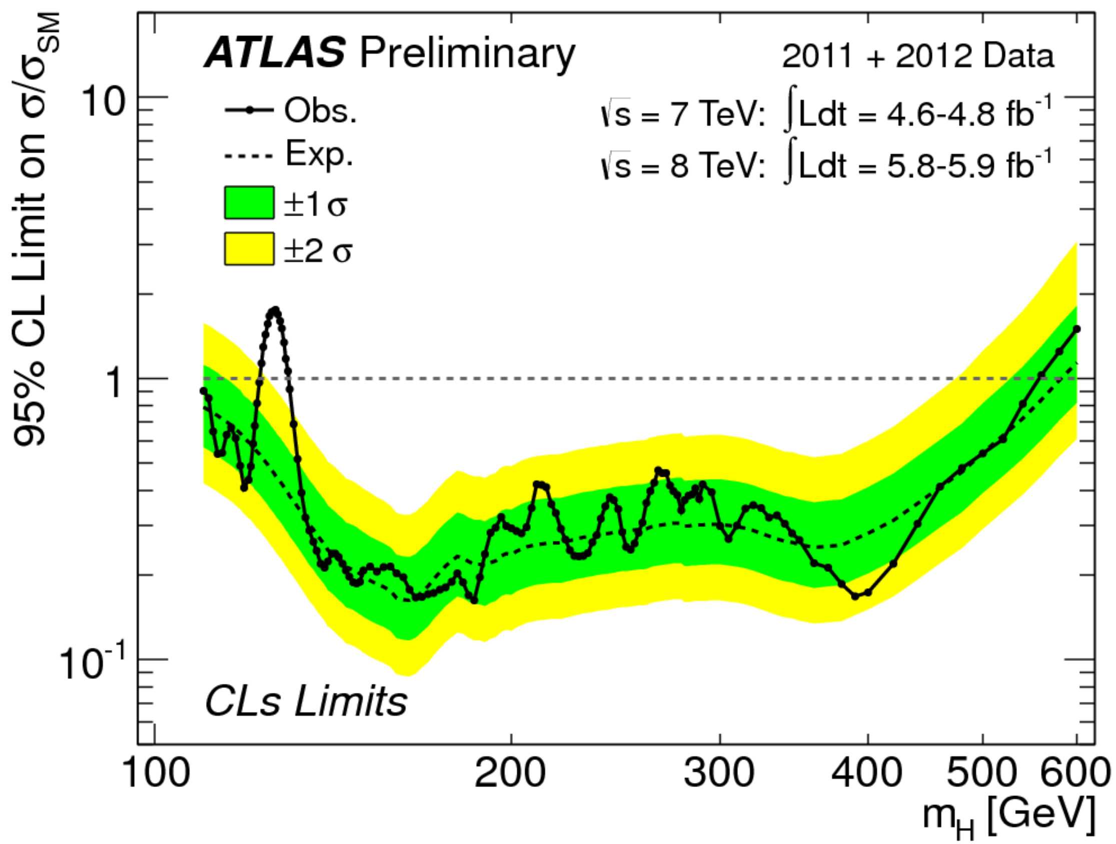
Higgs Hunter's Independence Day July 4th 2012

The Brazil Plot, what does it mean?

Observed Limit

Bands

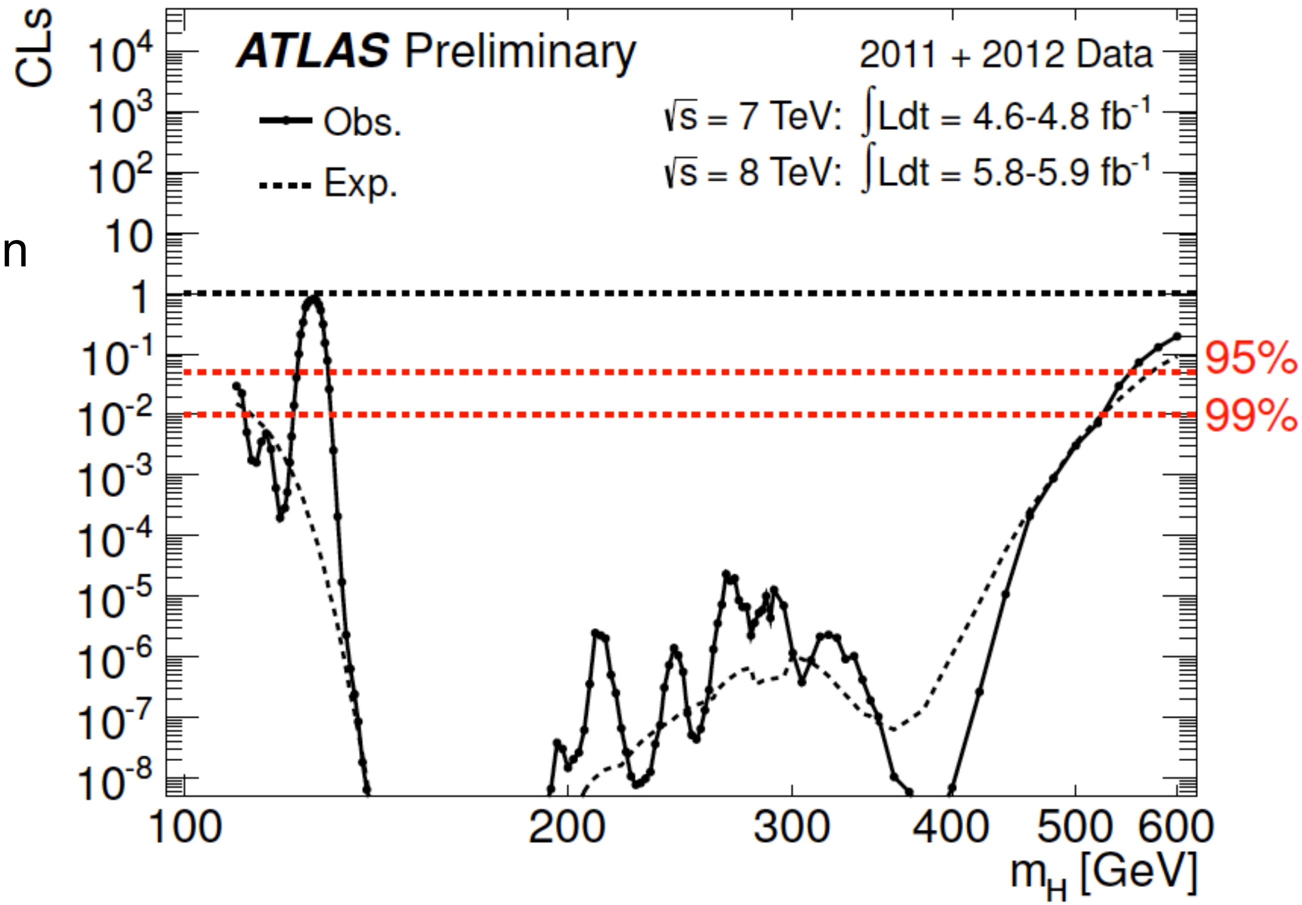
Expected Limit



What the ---- is CLs?

What is exclusion at the 95% CL?

99% CL?



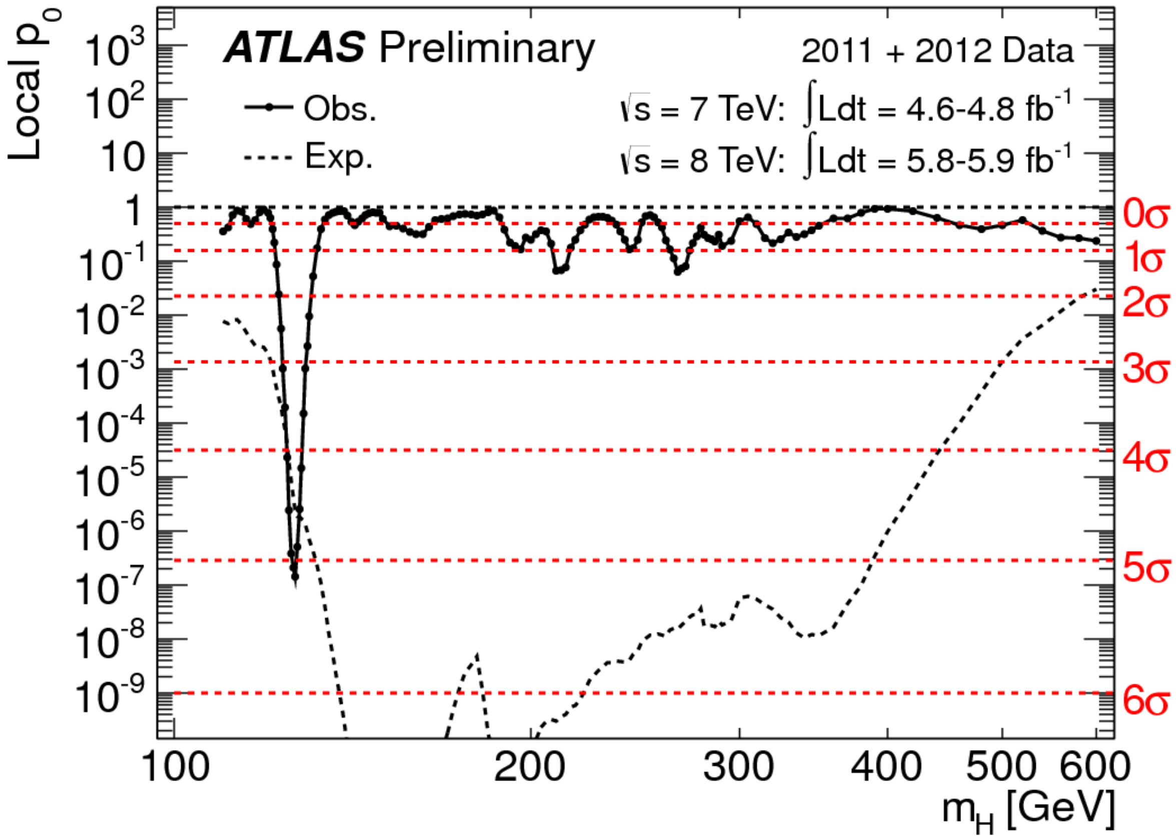
The p0 discovery plot, how to read it?

p-value

Local p0

Expected p0

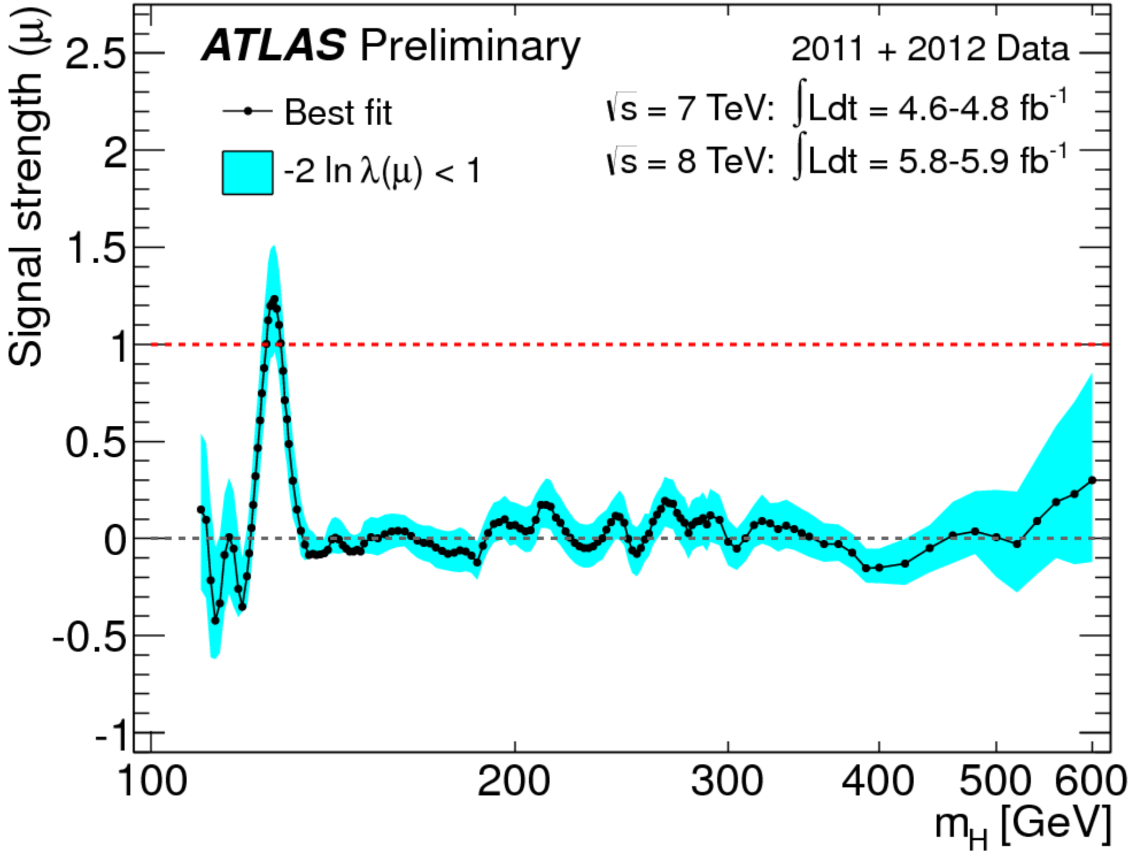
Observed local p0



Global p0 and the Look Elsewhere Effect

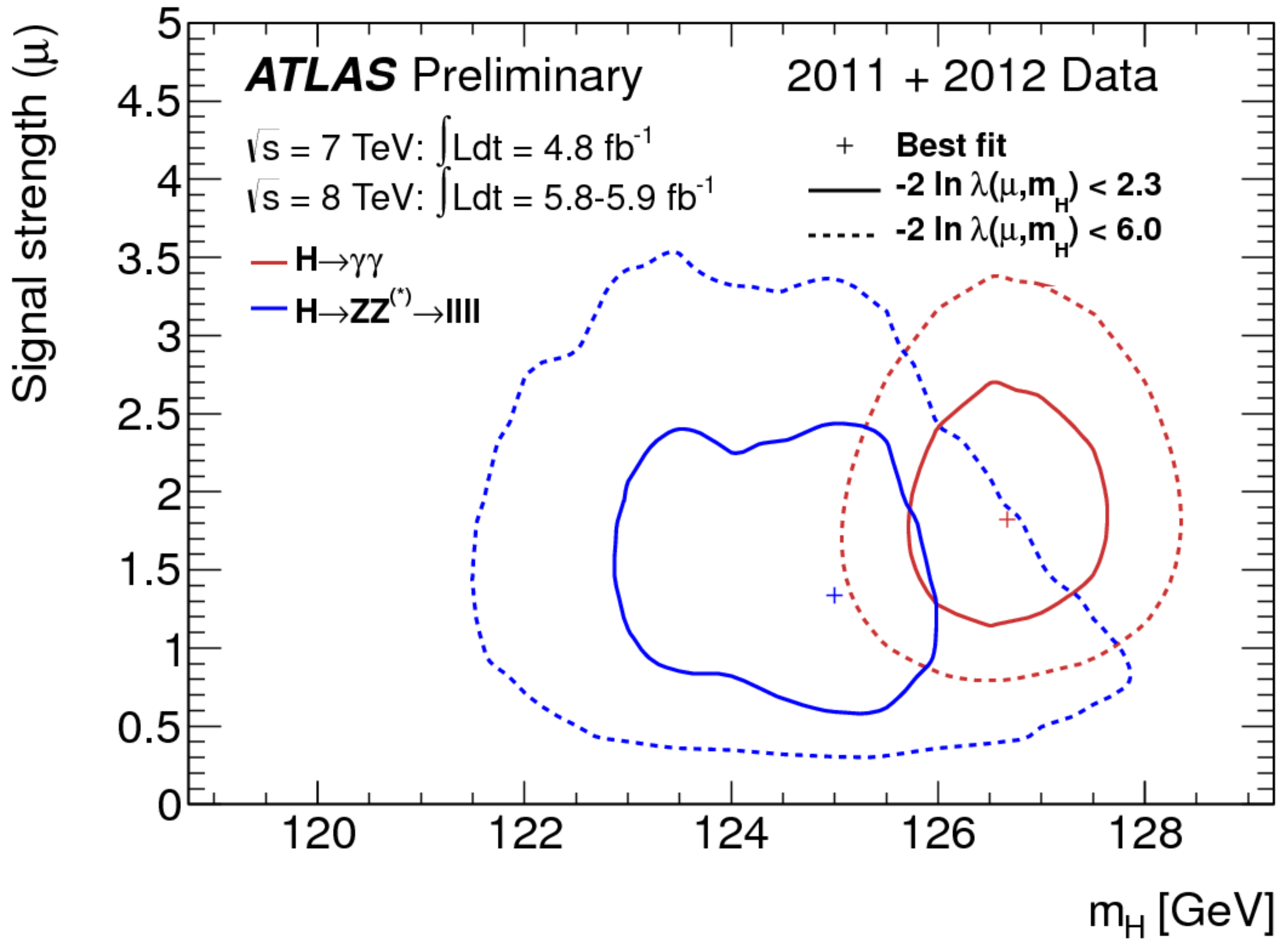
The cyan band plot, what is it?

What is mu hat? $\hat{\mu}$



Towards a measurement

2-D Likelihoods



References

ATLAS

PL [26] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** (2011) 1554. CCGV

CLs [27] A. L. Read, *Presentation of search results: The CL(s) technique*, J. Phys. **G28** (2002) 2693–2704.

LEE [28] E. Gross and O. Vitells, *Trial factors for the look elsewhere effect in high energy physics*, Eur. Phys. J. **C70** (2010) 525–530.

CMS

PL [90] G. Cowan et al., "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1–19, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. CCGV

[91] Moneta, L. et al., "The RooStats Project", in *13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2010)*. SISSA, 2010. arXiv:1009.1003. PoS(ACAT2010)057.

CLs [92] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth. A* **434** (1999) 435–443, doi:10.1016/S0168-9002(99)00498-2.

[93] A. L. Read, "Presentation of search results: the CLs technique", *J. Phys. G: Nucl. Part. Phys.* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.

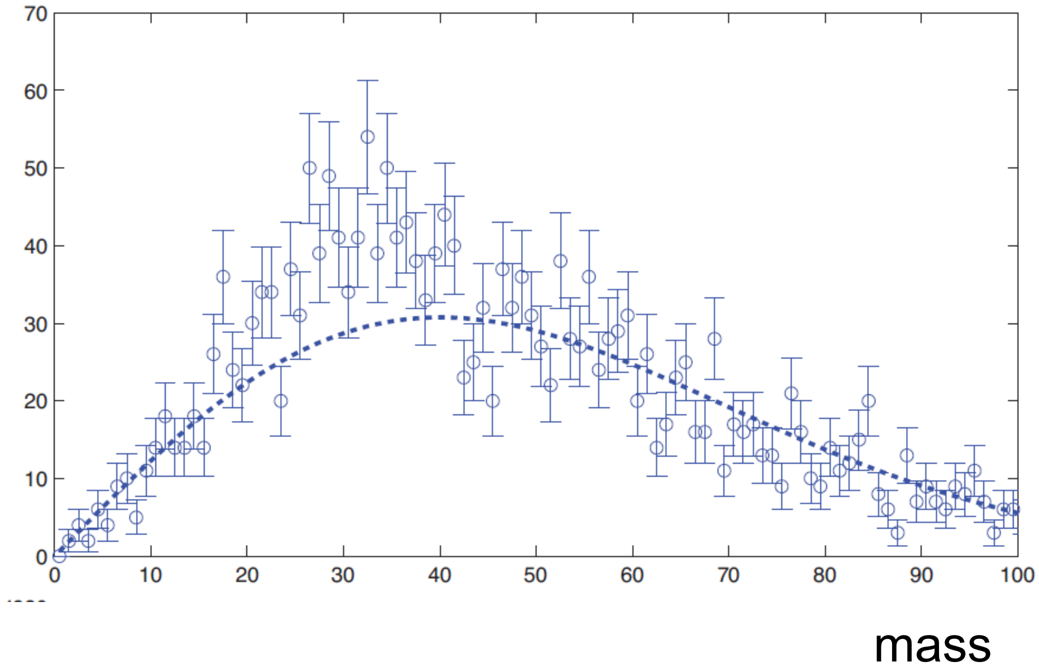
LEE [94] Gross, E. and Vitells, O., "Trial factors for the look elsewhere effect in high energy physics", *Eur. Phys. J. C* **70** (2010) 525–530, doi:10.1140/epjc/s10052-010-1470-8, arXiv:1005.1891.

The Statistical Challenge of HEP

The DATA: Billions of Proton-Proton collisions
which could be visualized with histograms

The Higgs mass is unknown

In this TOY example, we ask if the expected background (the Standard Model WITHOUT the Higgs Boson) contains a Higgs Boson, which would manifest itself as a peak in the distribution

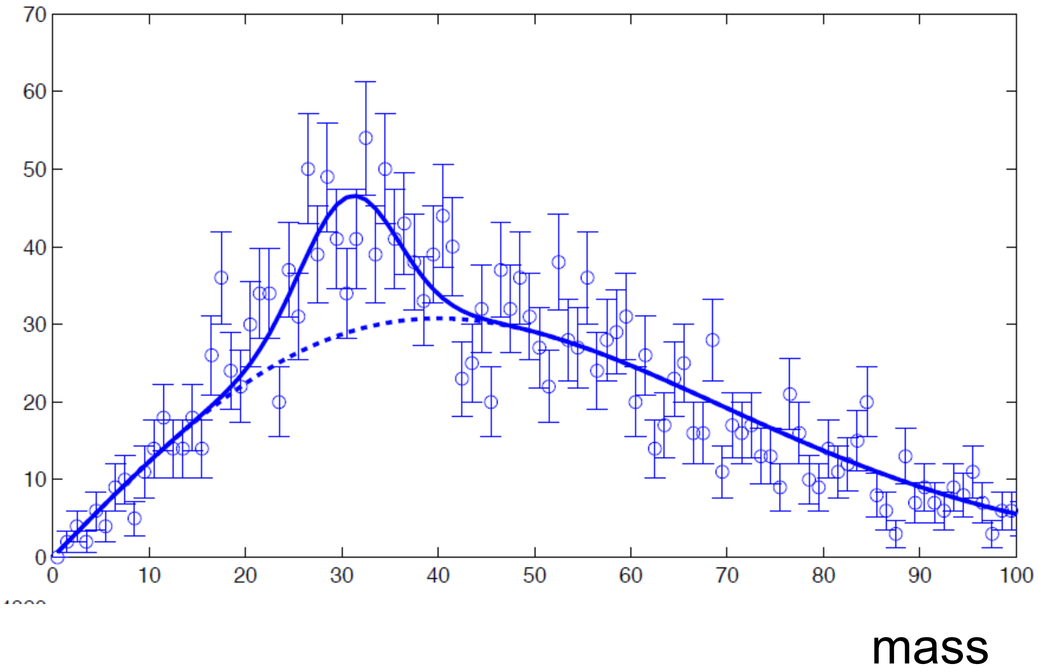


The Statistical Challenge of HEP

So the statistical challenge is obvious:

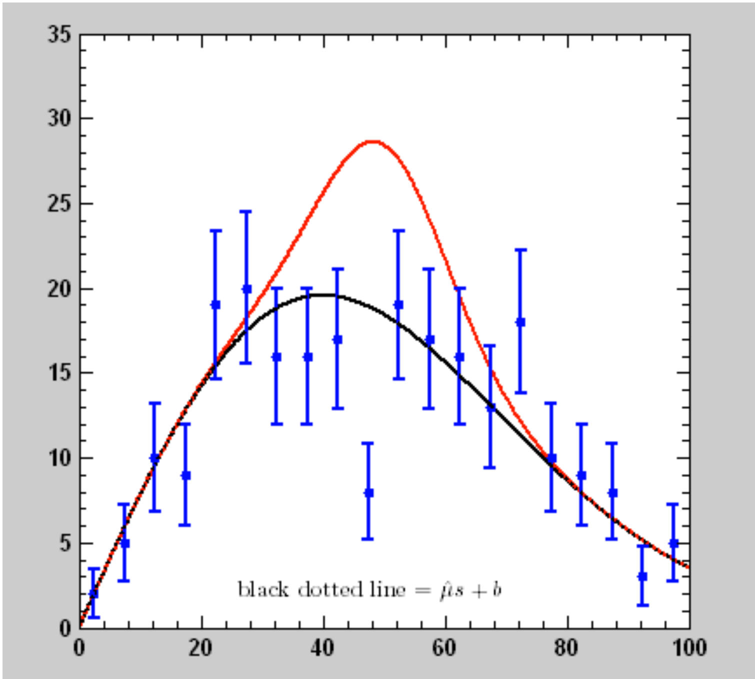
To tell in the most powerful way, and to the best of our current scientific knowledge, if there is new physics, beyond what is already known, in our data

The complexity of the apparatus and the background physics suffer from large systematic errors that should be treated in an appropriate way.



What is the statistical challenge?

- The black line represents the Standard Model (**SM**) expectation (Background only),
- How compatible is the **data** (**blue**) with the **SM expectation** (**black**)?
- Is there a signal hidden in this data?
- What is its statistical significance?
- What is the most powerful test statistic that can tell the SM (**black**) from an **hypothesized signal (red)**?



The Model

- The Higgs hypothesis is that of signal $s(m_H)$

$$s(m_H) = L \cdot \sigma_{SM}(m_H)$$

- In a counting experiment

$$n = \mu \cdot s(m_H) + b$$

$$\mu = \frac{L \cdot \sigma(m_H)}{L \cdot \sigma_{SM}(m_H)} = \frac{\sigma(m_H)}{\sigma_{SM}(m_H)}$$

- μ is the strength of the signal (with respect to the expected Standard Model one
- The hypotheses are therefore denoted by H_μ
- H_1 is the SM with a Higgs, H_0 is the background only model

A Frequentist Tale of Two Hypotheses

NULL

ALTERNATE

- Test the Null hypothesis and try to reject it
- Fail to reject it OR reject it in favor of the Alternate hypothesis

The Null Hypothesis

- The Standard Model without the Higgs is an hypothesis, (BG only hypothesis) many times referred to as the null hypothesis and is denoted by H_0
- In the absence of an alternate hypothesis, one would like to test the compatibility of the data with H_0
- This is actually a **goodness of fit test**

A Tale of Two Hypotheses

NULL

H_0- SM w/o Higgs

ALTERNATE

H_1- SM with Higgs

- Test the Null hypothesis and try to reject it
- Fail to reject it OR reject it in favor of the Alternate hypothesis

The Alternate Hypothesis?

- Let's zoom on

H_1 - SM with Higgs

- Higgs with a specific mass m_H

  OR

- Higgs anywhere in a specific mass-range

  → The look elsewhere effect

A Tale of Two Hypotheses

NULL	ALTERNATE
H_0- SM w/o Higgs	H_1- SM with Higgs

- Test the Null hypothesis and try to reject it
- Fail to reject it OR reject it in favor of the Alternate hypothesis

- Reject H_0 in favor of H_1 – A DISCOVERY

A Tale of Two Hypotheses

NULL

H_1- SM with Higgs

ALTERNATE

H_0- SM w/o Higgs

- Test the Null hypothesis and try to reject it
- Fail to reject it OR reject it in favor of the Alternate hypothesis

- Reject H_0 in favor of H_1 – A DISCOVERY

- Reject H_1 in favor of H_0 – Excluding H_1 (m_H)→Excluding the Higgs with a mass m_H

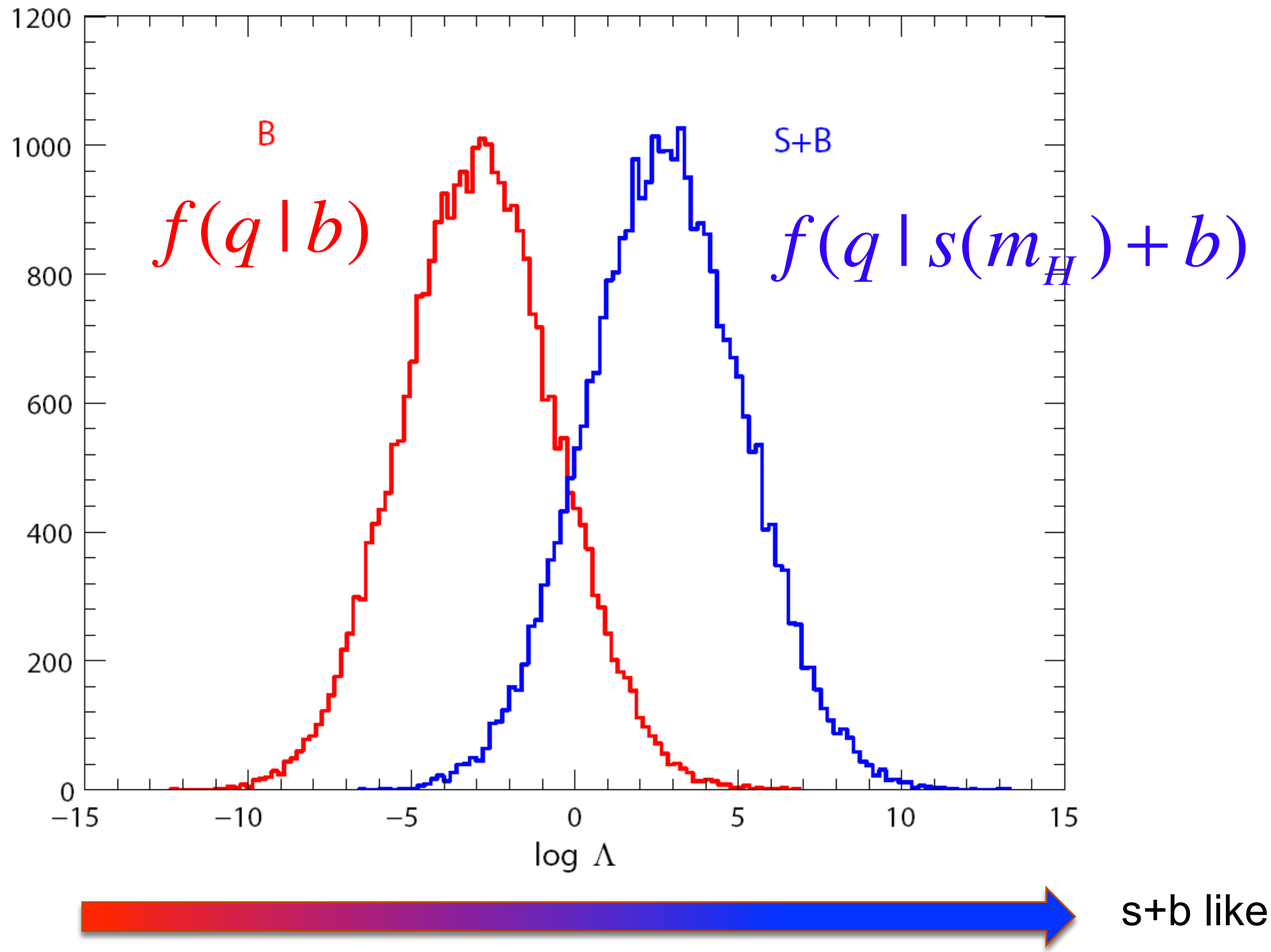
Testing an Hypothesis (wikipedia...)

- The first step in any hypothesis testing is to state the relevant **null,** H_0 and **alternative hypotheses,** say, H_1
- The next step is to define a test statistic, q, under the null hypothesis
- Compute from the observations the observed value q_{obs} *of the test statistic* q.
- Decide (based on q_{obs}) to either
  fail to reject the null hypothesis or
  reject it in favor of an alternative hypothesis

- **next: How to construct a test statistic, how to decide?**

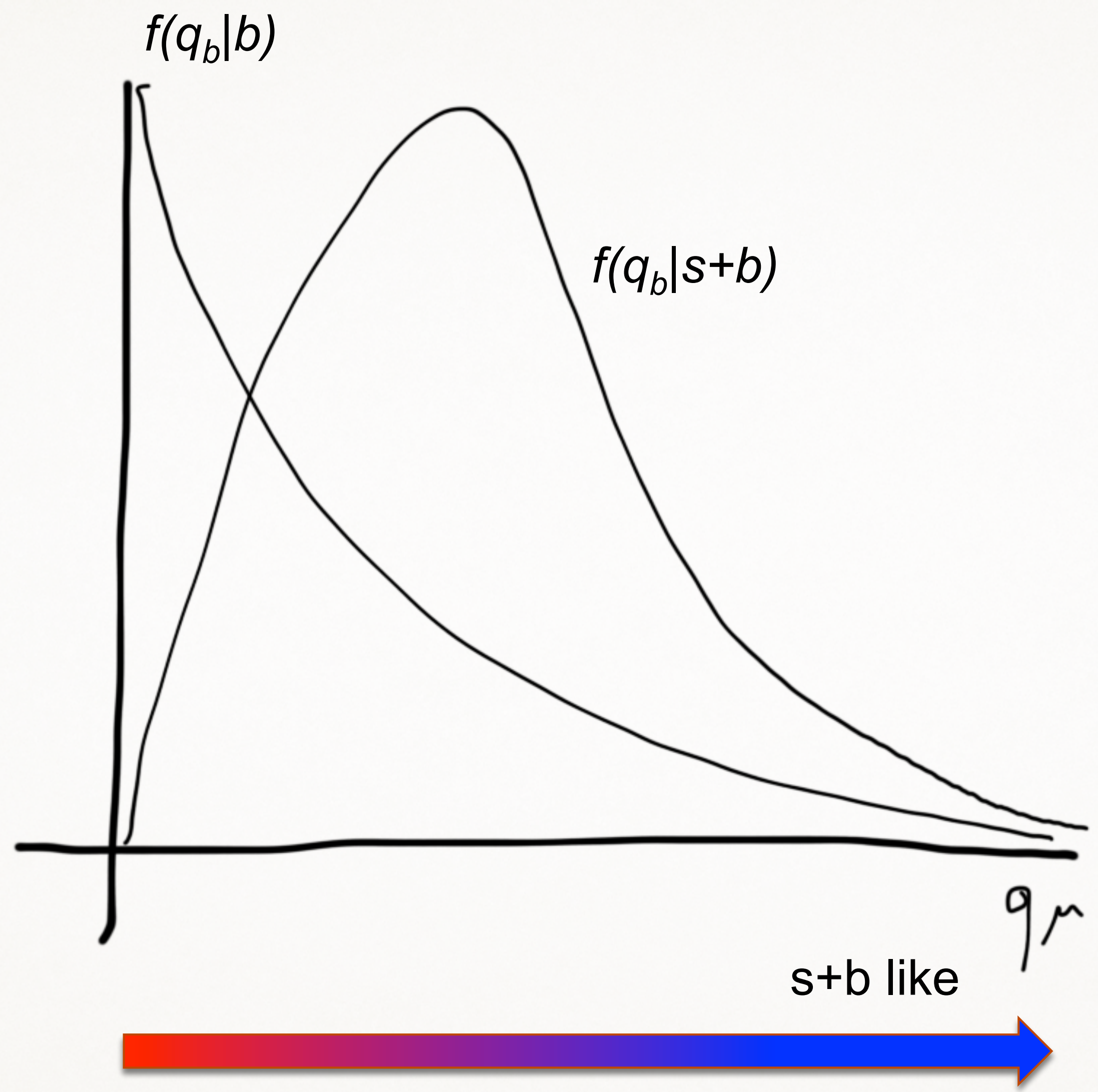
Test statistic and p-value

PDF of a test statistic



Test statistic

- The pdf $f(q\,|\,b)$ or $f(q\,|\,s{+}b)$ might be different depended on the chosen test statistic.
- Some might be powerful than others in distinguishing between the null and alternate hypothesis ($s(m_H){+}b$ and b)



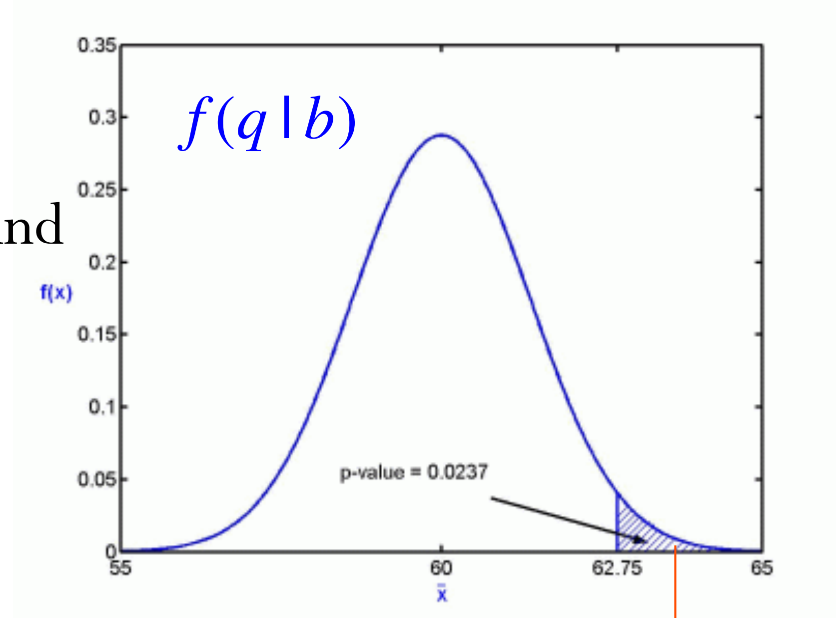
p-Value

- Discovery…. A deviation from the SM - from the background only hypothesis…
- When will one reject an hypothesis?
- **p-value** = probability that result is as or less compatible with the background only hypothesis (->more signal like)
- Define a-priori a control region **α**
- For discovery it is a custom to choose **$\alpha=2.87\times10^{-7}$**
- If result falls within the control region, i.e.

  $p < \alpha$ the BG only hypothesis is rejected

  $\rightarrow$ A discovery

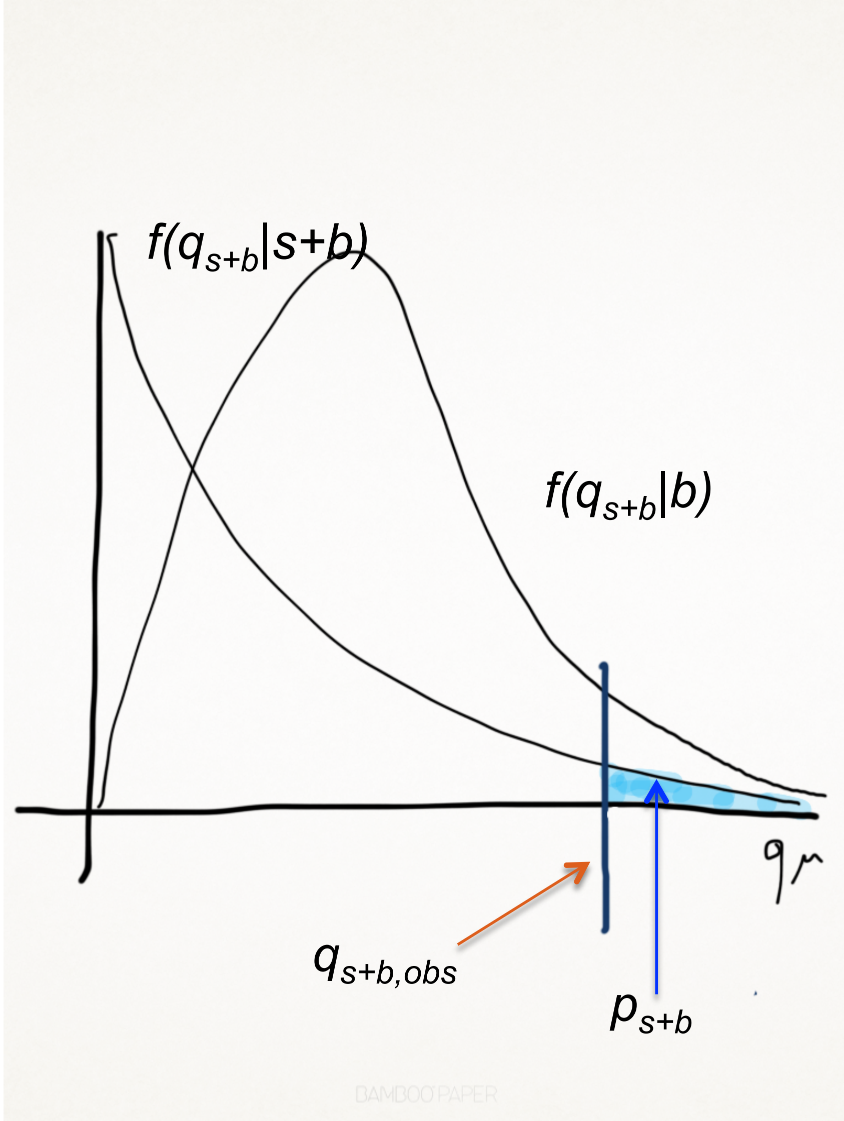
- The pdf of **q**….



Control region
Of size α

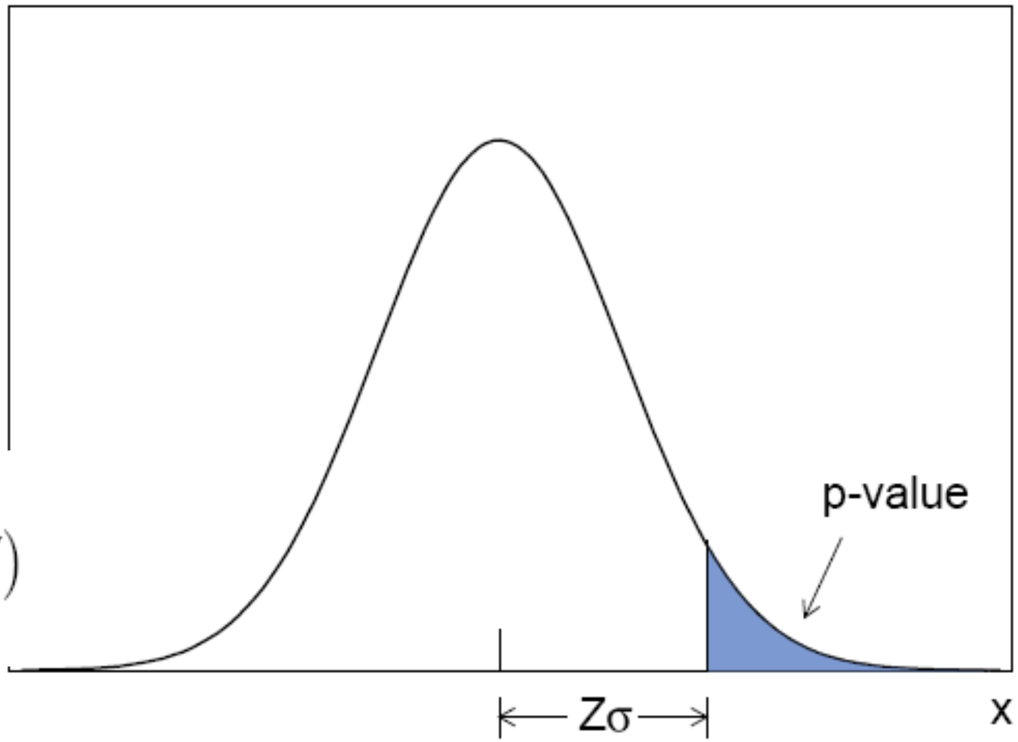
p-value – testing the signal hypothesis

- When testing the signal hypothesis, the p-value is the probability that the observation is less compatible with the signal hypothesis (mor background like) than the observed one
- We denote it by p_{s+b}
- It is custom to say that if p_{s+b}<5% the signal hypothesis is rejected
  → Exclusion



From p-values to Gaussian significance

It is a custom to express the p-value as the significance associated to it, had the pdf were Gaussians



$$p=\int_{Z}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx=1-\Phi(Z)$$

$$Z=\Phi^{-1}(1-p)$$

$\boxed{\text{A significance of } Z=5 \text{ corresponds to } p=2.87\times10^{-7}.}$

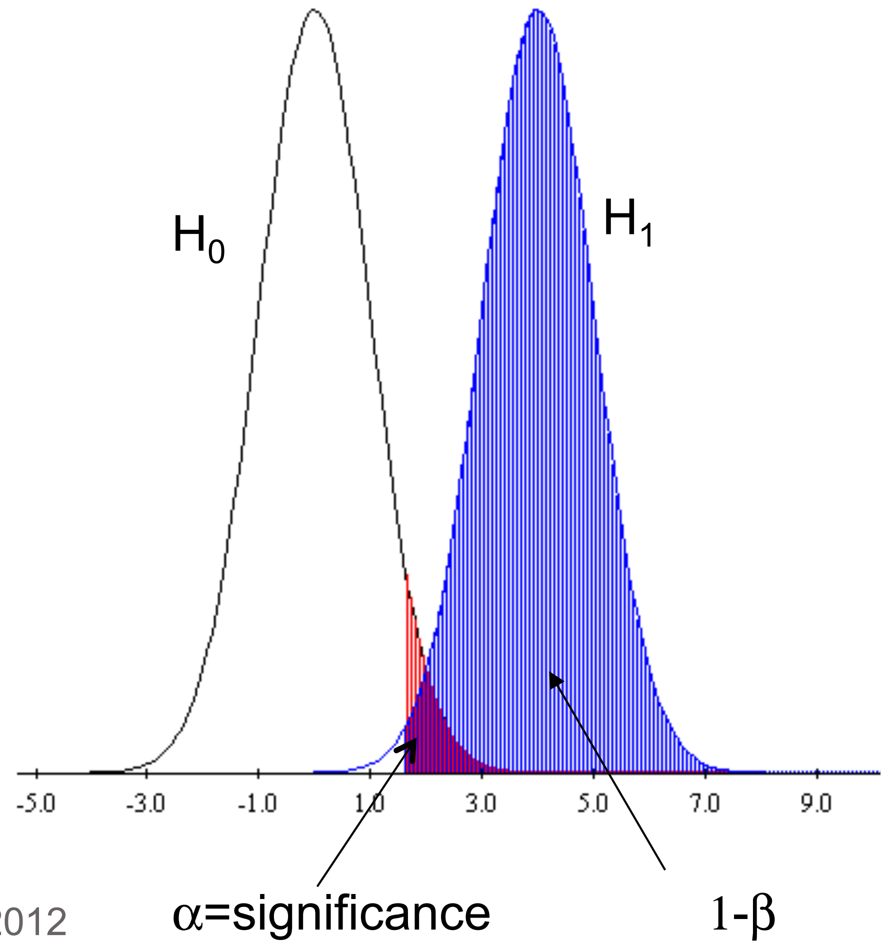
Beware of 1 vs 2-sided definitions!

Basic Definitions: type I-II errors

- By defining α you determine your accepted level of
  type-I error: the probability to reject the tested (null) hypothesis (H_0) when it is true
- $$\alpha = \Pr ob(reject\ H_0 \mid H_0)$$
  $$\alpha = typeI\ error$$
- **Type II**: The probability to accept the null hypothesis when it is wrong

$$\beta = \Pr ob(accept\ H_0 \mid \overline{H}_0)$$
$$= \Pr ob(reject\ H_1 \mid H_1)$$
$$\beta = typeII\ error$$

- The pdf of q….

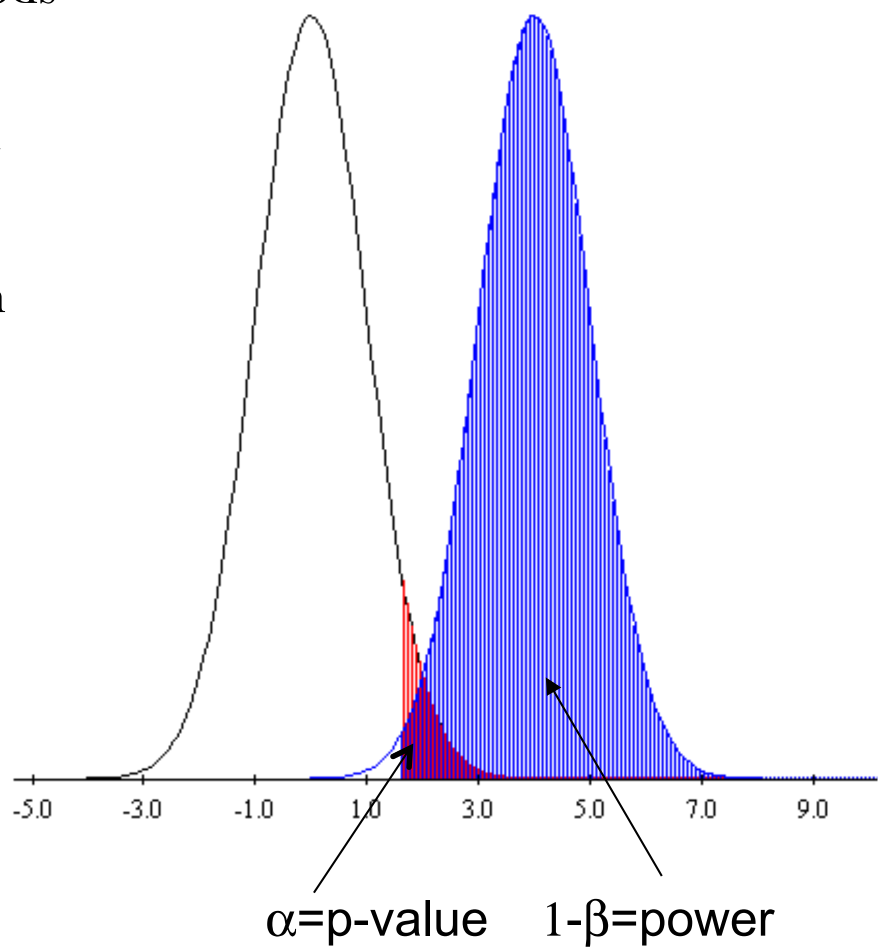


Basic Definitions: POWER

- $\alpha = \Pr ob(reject\ H_0 \mid H_0)$
- The POWER of an hypothesis test is the probability to reject the null hypothesis when the alternate analysis is true!
- $$POWER = \text{Prob}(reject\ H_0 \mid H_1)$$
  $$\beta = \Pr ob(reject\ H_1 \mid H_1) \Rightarrow$$
  $$1-\beta = \Pr ob(accept\ H_1 \mid H_1) \Rightarrow$$
  $$1-\beta = \Pr ob(reject\ H_0 \mid H_1) \Rightarrow$$
  $$POWER = 1-\beta$$
- The power of a test increases as the rate of type II error decreases

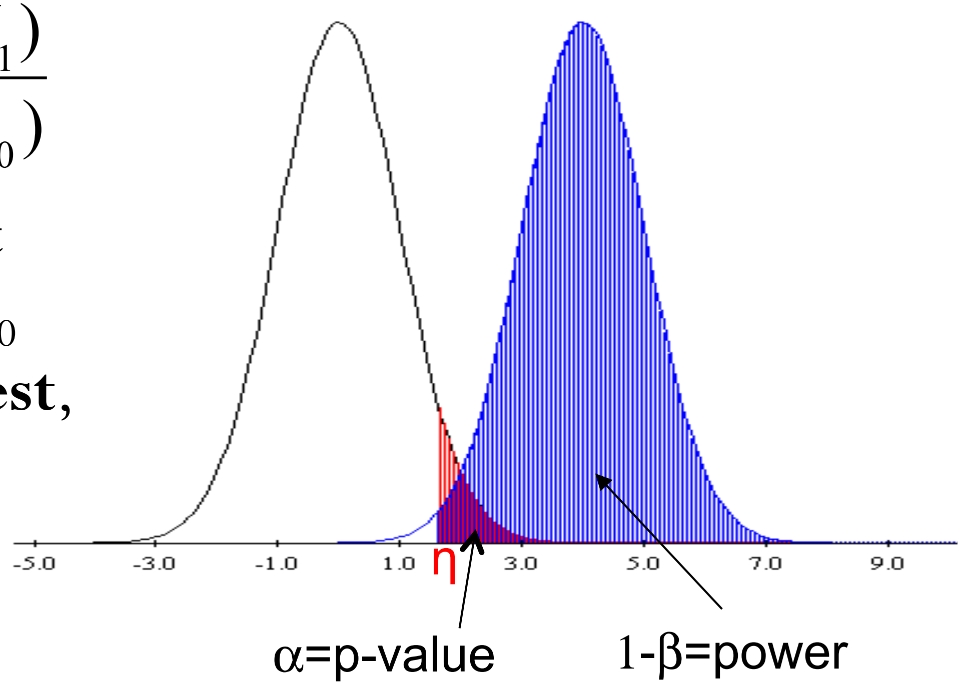
Which Analysis is Better

- To find out which of two methods is better plot the p-value vs the power for each analysis method
- Given the p-value, the one with the higher power is better
- p-value~significance



The Neyman-Pearson Lemma

- Define a **test statistic** $\lambda = \frac{L(H_1)}{L(H_0)}$
- When performing a hypothesis test between two simple hypotheses, H_0 and H_1, **the Likelihood Ratio test**, which rejects H_0 in favor of H_1, **is the most powerful test** of size α for a threshold η
- **Note:** Likelihoods are functions of the data, even though we often not specify it explicitly $\lambda(x) = \frac{L(H_1 \mid x)}{L(H_0 \mid x)}$



The Profile Likelihood

The Profile Likelihood ("PL")

For discovery we test the H_0 null hypothesis and try to reject it

$$q_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s+b)}$$

For $\hat{\mu}\sim 0,\ q$ small

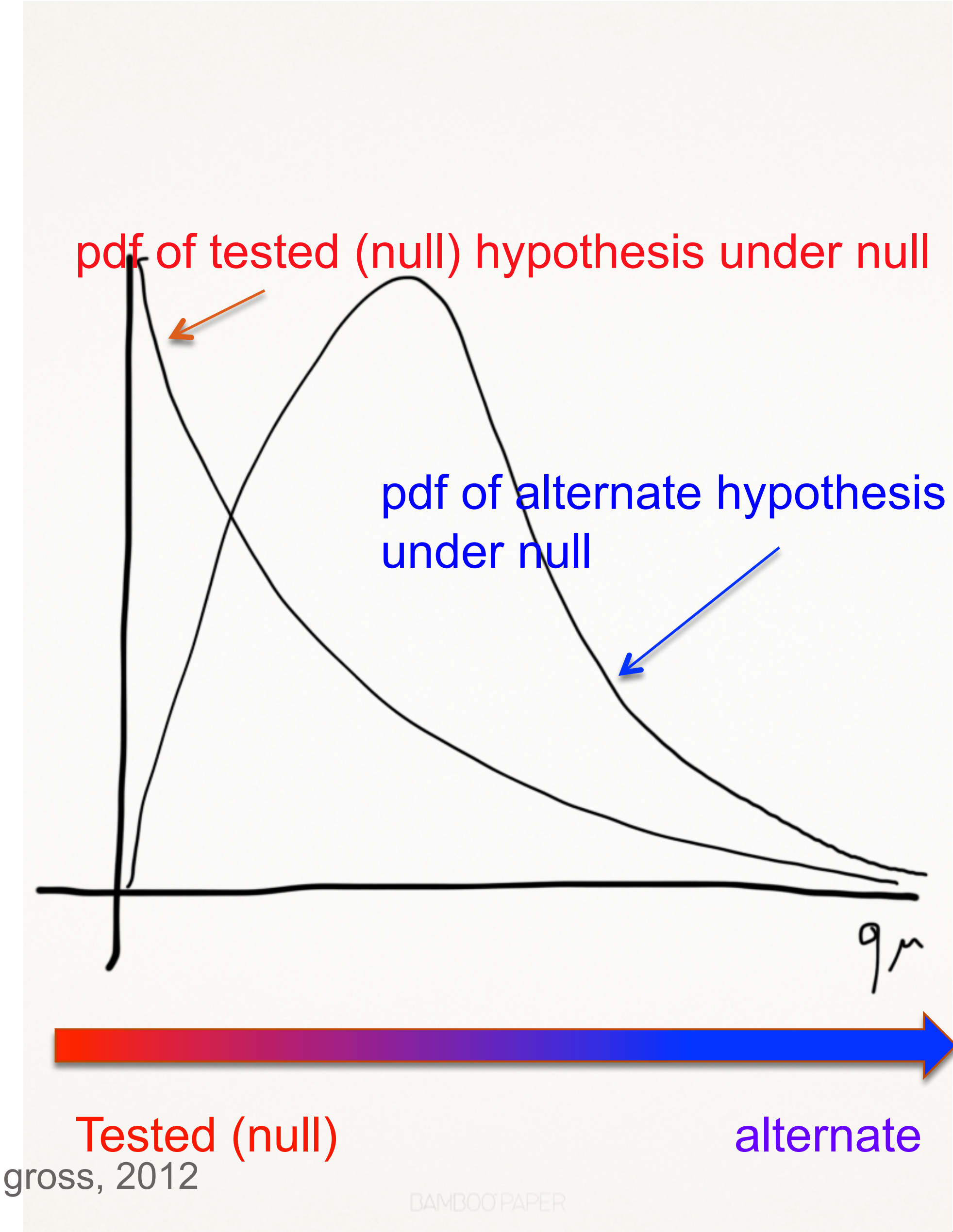
$\hat{\mu}\sim 1,\ q$ large

For exclusion we test the signal hypothesis and try to reject it

$$q_\mu = -2\ln\frac{L(\mu)}{L(\hat{\mu})}$$

$\hat{\mu}\sim\mu,\ q$ small

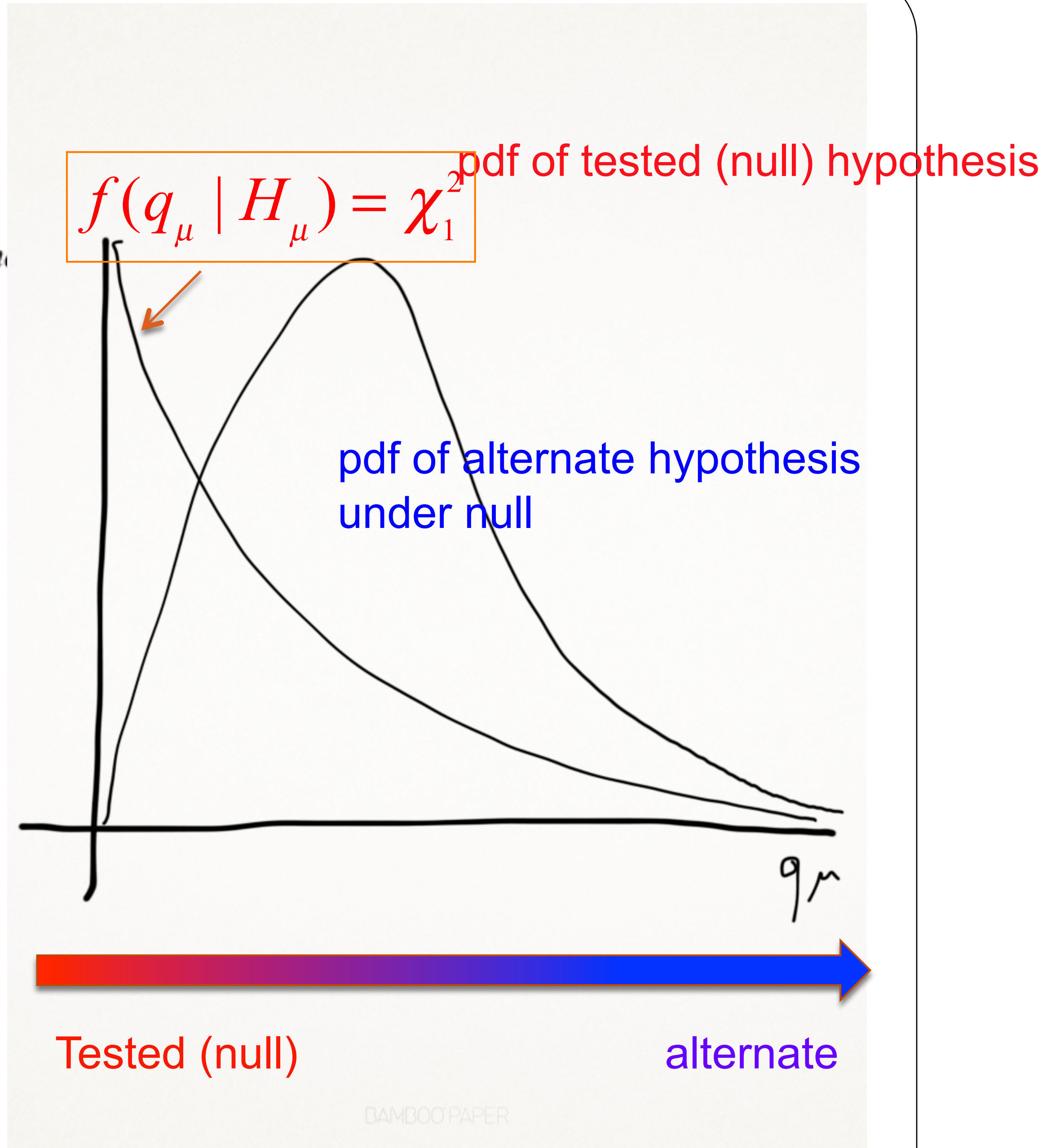
$\hat{\mu}\sim 0,\ q$ large



Wilks Theorem

S.S. Wilks, *The large-sample distribution of th*
Ann. Math. Statist. **9** (1938) 60-2.

- Under a set of regularity conditions and for a sufficiently large data sample, *Wilks' theorem says that* the pdf of the statistic q under the null hypothesis approaches a chi-square PDF for one degree of freedom



$f(q_0 \mid H_0) = \chi_1^2$ $\quad$ $f(q_\mu \mid H_\mu) \sim \chi_1^2$

Nuisance Parameter

Nuisance Parameters

- Normally, the background, b(θ), has an uncertainty which has to be taken into account. In this case θ is called a nuisance parameter (which we associate with background systematics)
- The signal strength μ is a parameter if interest
- Usually the nuisance parameters are auxiliary parameters and their values are constrained by auxiliary measurements
- Example $n \sim \mu s(m_H) + b$ $\langle n \rangle = \mu s + b$

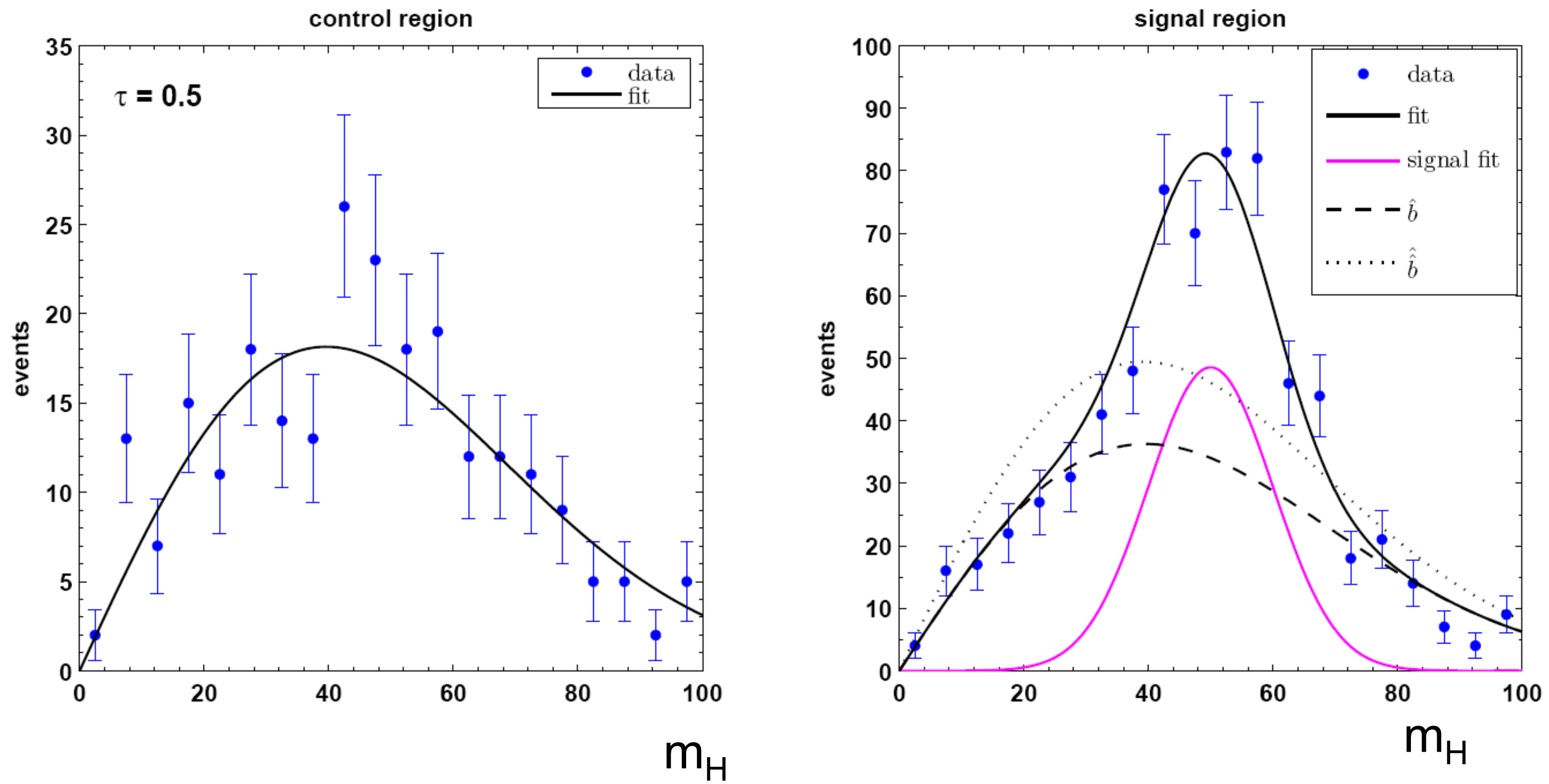
$$m = \tau b$$

$$\boxed{L\left(\mu \cdot s + b(\theta)\right) = Poisson\left(n; \mu \cdot s + b(\theta)\right) \cdot Poisson\left(m; \tau b(\theta)\right)}$$

Mass shape as a discriminator

$$n \sim \mu s(m_H) + b \qquad m \sim \tau b$$

$$L\left(\mu \cdot s + b(\theta)\right) = \prod_{i=1}^{nbins} Poisson\left(n_i; \mu \cdot s_i + b_i(\theta)\right) \cdot Poisson\left(m_i; \tau b_i(\theta)\right)$$



Profile Likelihood with Nuisance Parameters

$$q_\mu = -2ln \frac{L(\mu s + \hat{\hat{b}}_\mu)}{L(\hat{\mu} s + \hat{b})}$$

$$q_\mu = -2ln \frac{\max_b L(\mu s + b)}{\max_{\mu,b} L(\mu s + b)}$$

$$q_\mu = q_\mu(\hat{\mu}) = -2ln \frac{L(\mu s + \hat{\hat{b}}_\mu)}{L(\hat{\mu} s + \hat{b})}$$

$\hat{\mu}$ MLE of μ

$\hat{b}$ MLE of b

$\hat{\hat{b}}_\mu$ MLE of b fixing μ

$\hat{\hat{\theta}}_\mu$ MLE of θ fixing μ

Confidence Interval and Confidence Level (CL)

CL & CI - Wikipedia

- A **confidence interval** (**CI**) is a particular kind of interval estimate of a population parameter. Instead of estimating the parameter by a single value, an interval likely to include the parameter is given. Thus, confidence intervals are used to indicate the reliability of an estimate. How likely the interval is to contain the parameter is determined by the **confidence level** or confidence coefficient. Increasing the desired confidence level will widen the confidence interval.

Confidence Interval & Coverage

- Say you have a measurement μ_{meas} of μ with μ_{true} being the unknown true value of μ
- Assume you know the probability dostribution function of $p(\mu_{meas} | \mu)$
- Given the measurement you deduce somehow (based on your statistical model) that there is a 95% Confidence interval $[\mu_1, \mu_2]$.
- The correct statement: In an ensemble of experiments 95% of the obtained confidence intervals will contain the true value of μ.

Upper limit

- Given the measurement you deduce somehow (based on your statistical model) that there is a 95% Confidence interval $[0, \mu_{up}]$.
- This means: In an ensemble of experiments 95% of the obtained confidence intervals will contain the true value of μ, including $\mu=0$ (no Higgs)
- We therefore deduce that $\mu < \mu_{up}$ at the 95% Confidence Level (CL)
- μ_{up} is therefore an upper limit on μ
- If $\mu_{up}<1 \rightarrow$
  $\sigma(m_H)<\sigma_{SM}(m_H) \rightarrow$
  a SM Higgs with a mass m_H is excluded at the 95% CL

Confidence Interval & Coverage

- Confidence Level: A CL of 95% means that in an ensemble of experiments, each producing a confidence interval, 95% of the confidence intervals will contain the true value of μ
- Normally, we make one experiment and try to estimate from this one experiment the confidence interval at a specified CL
- If in an ensemble of (MC) experiments our estimated Confidence Interval fail to contain the true value of μ 95% of the cases (for every possible μ) we claim that our method undercover
- If in an ensemble of (MC) experiments the true value of μ is covered within the estimated confidence interval , we claim a coverage

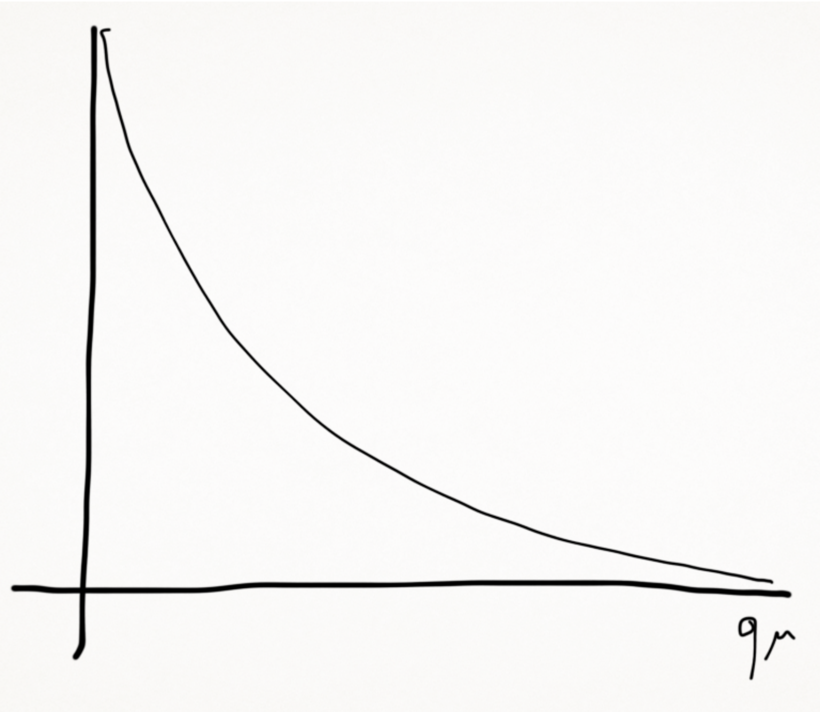
Exclusion of a Higgs with mass m_H

$$\tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \hat{\mu} \le \mu \\ 0 & \hat{\mu} > \mu \end{cases} \quad = \quad \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0 \, , \\ -2\ln\frac{L(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(\hat{\mu},\hat{\boldsymbol{\theta}})} & 0 \le \hat{\mu} \le \mu \, , \\ 0 & \hat{\mu} > \mu \, . \end{cases}$$

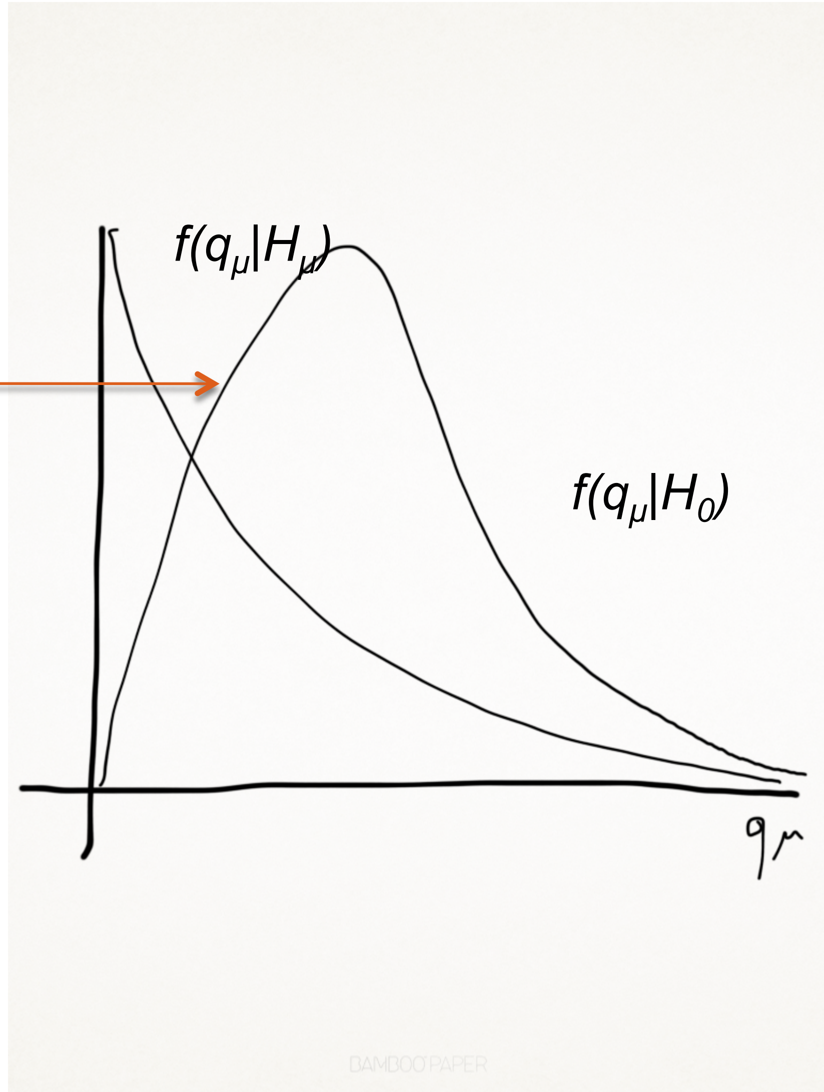
Signal upward fluctuations do not serve as evidence against the signal hypothesis

$$f(q_\mu \mid H_\mu) \sim \chi_1^2$$

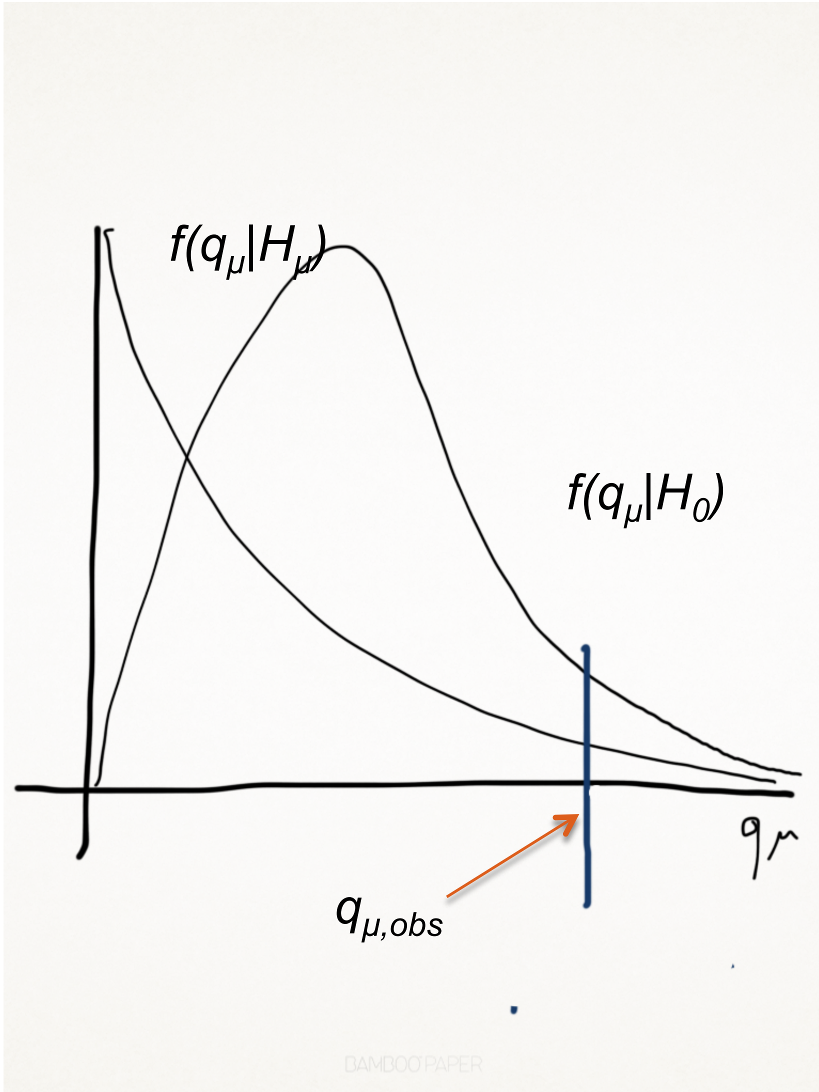
This is a real ap[proximation, exact formulae in the CCGV paper



$f(q_\mu \mid H_0)$ see CCGV



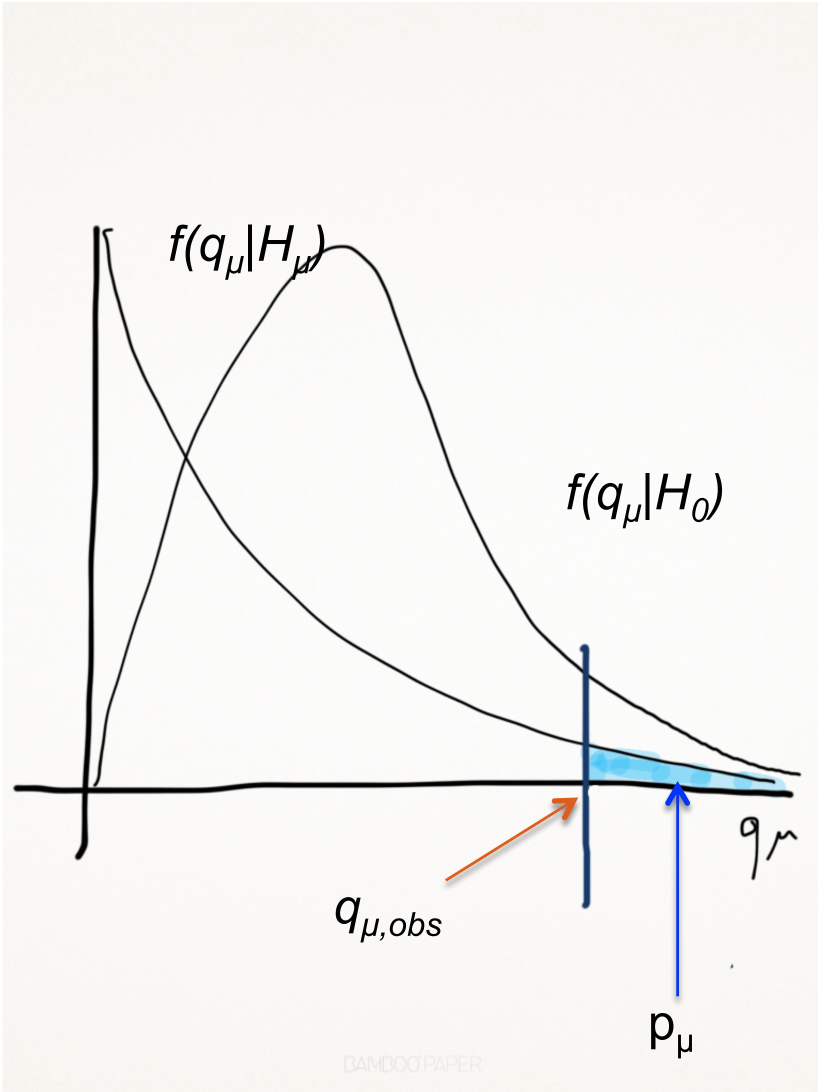
- We test hypothesis H_μ
- We calculate the PL (profile likelihoos) ratio with the one observed data
- $q_{\mu,obs}$



- Find the p-value of the signal hypothesis H_μ

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu \mid H_\mu)\, dq_\mu$$

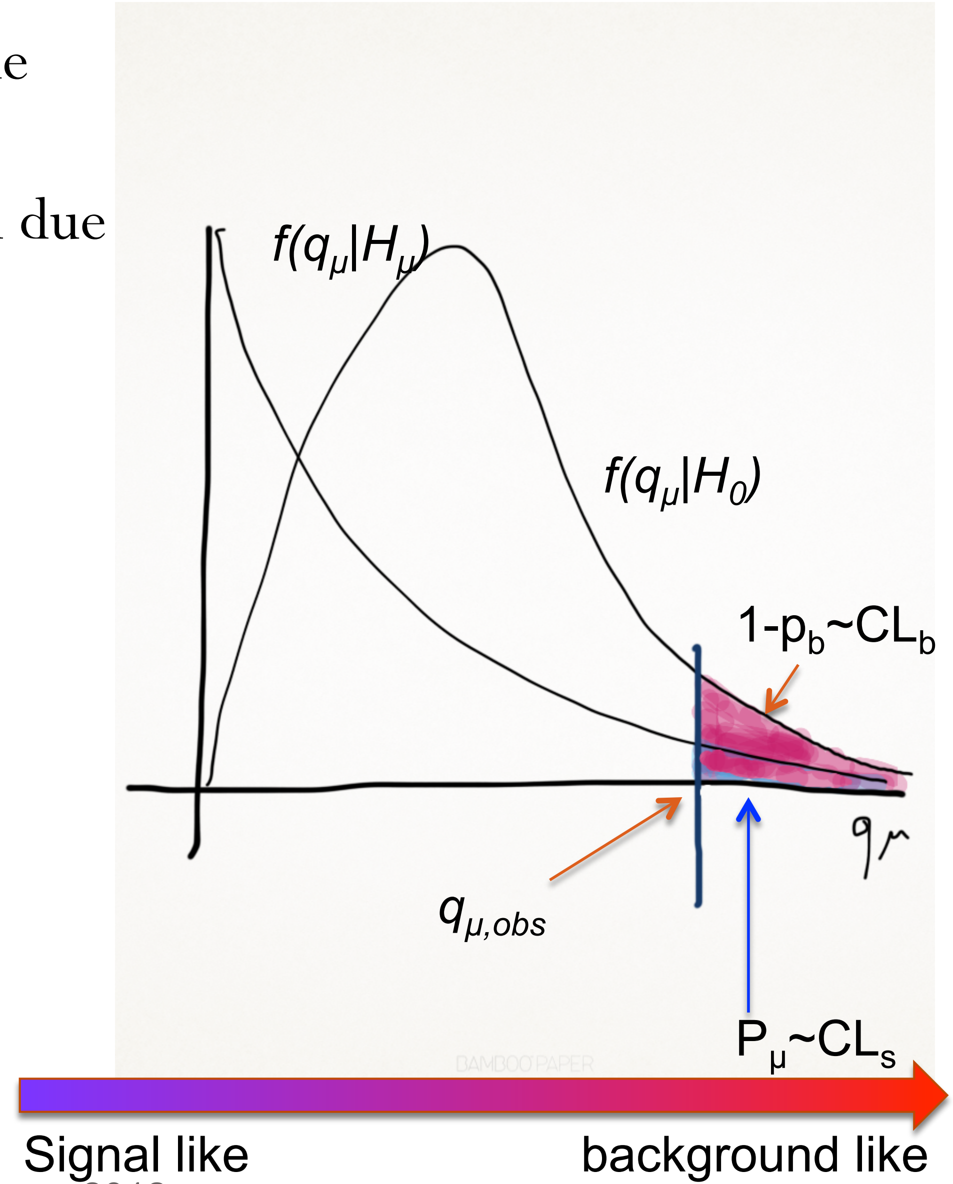
- In principle if $p_\mu < 5\%$, H_μ hypothesis is excluded at the 95% CL
- Note that H_μ is for a given Higgs mass m_H



CLs

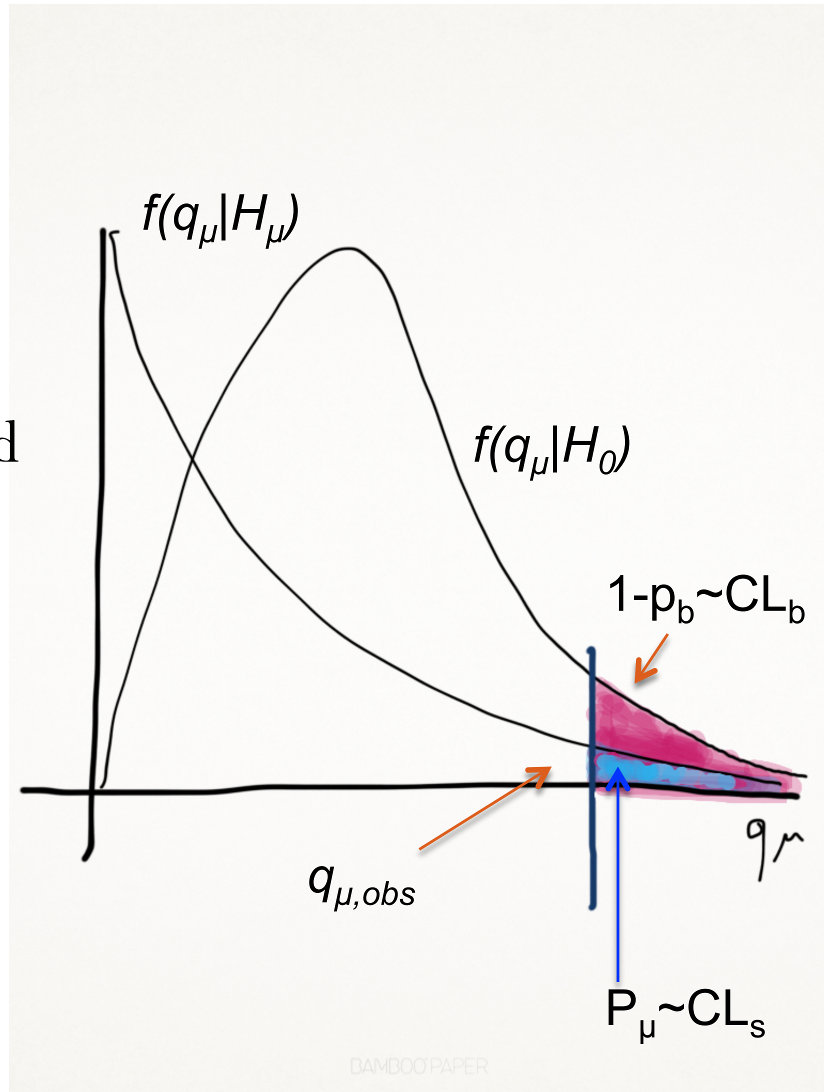
CLb

- $CL_b \sim 1\text{-}p_b$ is the compatibility of the background with the background hypothesis and might be very small due to downward fluctuations of the background



CLs

- A complication arises when μ s+b~b
- When the signal cross section is very small the s(mH)+b hypothesis can be rejected but at the same time the background hypothesis is almost rejected as well due to downward fluctuations of the background
- These downward fluctuations allow the exclusion of a signal the experiment is not sensitive to



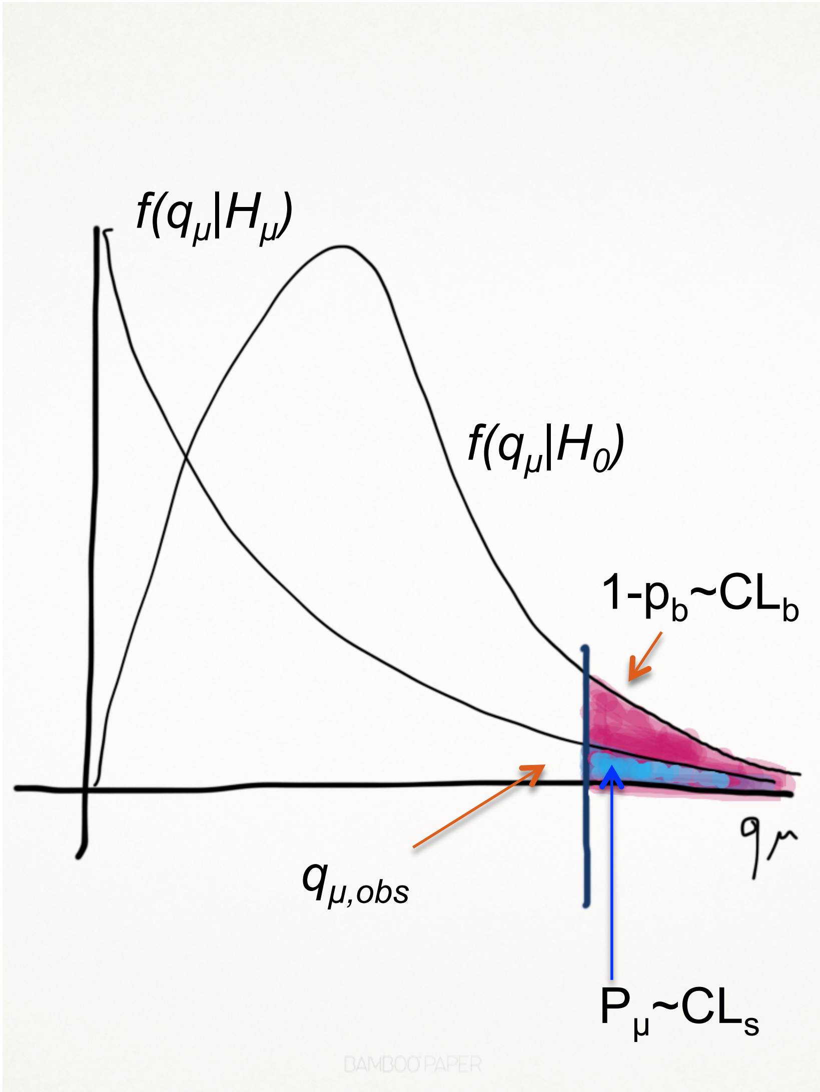
- Inspired by Zech(Roe and Woodroofe)'s derivation for counting experiments

$$P(n \le n_o | n_b \le n_o, s+b) = \frac{P(n \le n_o \mid s+b)}{P(n \le n_o \mid b)}$$

- A. Read suggested the CL_s method with

$$\boxed{CL_s = \frac{CL_{s+b}}{CL_b} = \frac{p_{s+b}}{1-p_b}}$$

- This means that you will never be able to exclude a signal with a tiny cross section (to which you are not sensitive)



CLs

$$P(n_o \le n_{s+b} \,|\, n_b \le n_o, s+b) = \frac{P(n \le n_o \mid s+b)}{P(n \le n_o \mid b)}$$

- Suppose $\langle n_b \rangle = 100$
- $s(m_{H1}) = 30$
- Suppose $n_{obs} = 102$
- $s+b = 130$
- $\mathrm{Prob}(n_{obs} \le 102 \,|\, 130) < 5\%$, m_{H1} is excluded at $>95\%$ CL

- Now suppose $s(m_{H2}) = 1$, can we exclude m_{H2}?
- If $n_{obs} = 102$, obviously we cannot exclude m_{H2}
- Now suppose $n_{obs} = 80$, $\mathrm{prob}(n_{obs} \le 80 \,|\, 101) < 5\%$, we looks like we can exclude m_{H2}… but this is dangerous, because what we exclude is $(s(m_{H2})+b)$ and not s……
- With this logic we could also exclude b (expected b=100)
- To protect we calculate a modofoed p-value
- We cannot exclude m_{H2}

$$\frac{\mathrm{Prob}(\mathrm{nobs} \le 80 \mid 101)}{\mathrm{Prob}(\mathrm{nobs} \le 80 \mid 100)} \sim 1$$

The Modified CLs with the PL test statistic

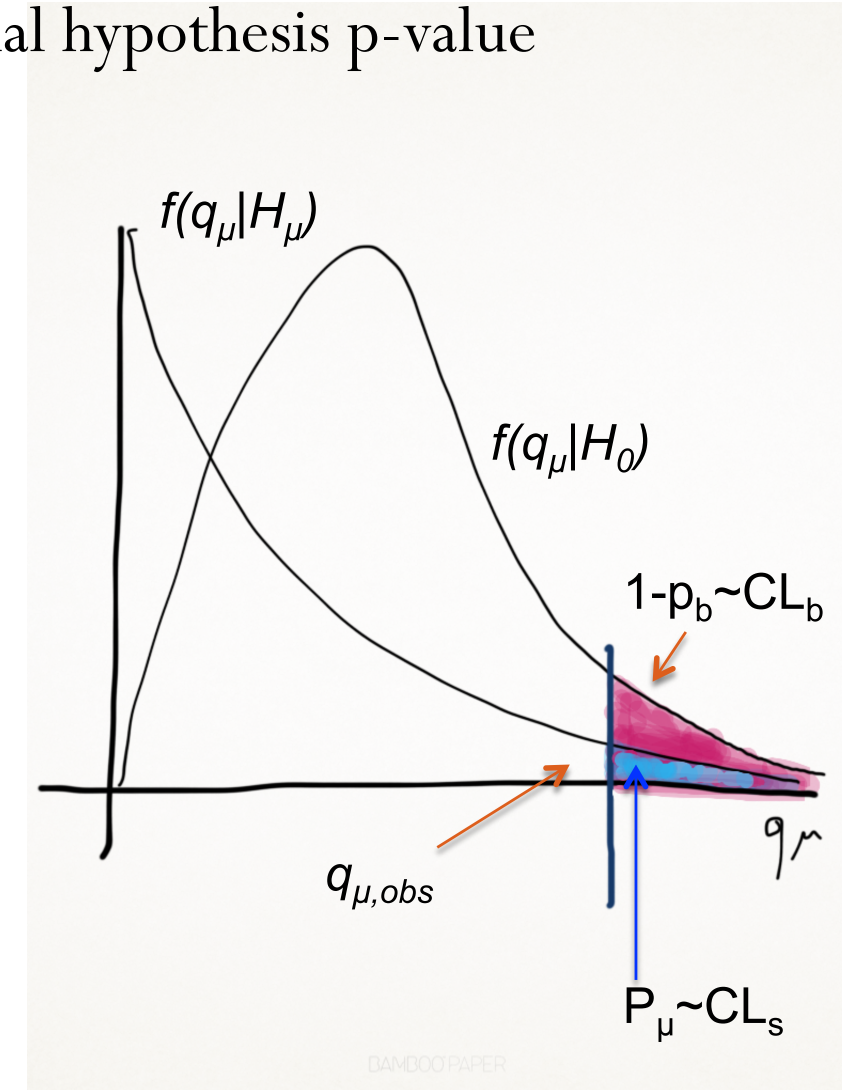
- The CLs method means that the signal hypothesis p-value

p_μ is modified to

$$p_\mu \rightarrow p'_\mu = \frac{p_\mu}{1-p_b}$$

$$p_\mu = \int_{\tilde{q}_{\mu,obs}}^{\infty} f(\tilde{q}_\mu|\mu)d\tilde{q}_\mu$$

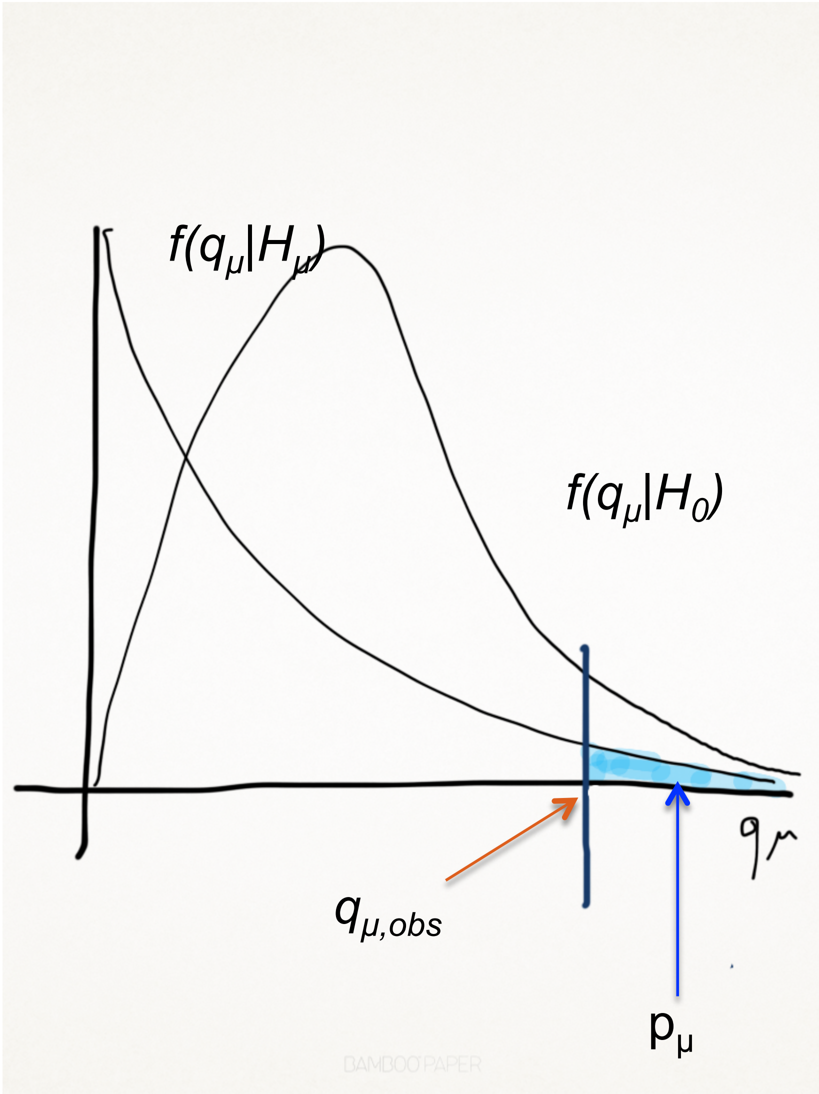
$$p_b = 1 - \int_{\tilde{q}_{\mu,obs}}^{\infty} f(\tilde{q}_\mu|0)d\tilde{q}_\mu$$



- Find the p-value of the signal hypothesis H_μ

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu \mid H_\mu) dq_\mu$$

- In principle if $p_\mu < 5\%$, H_μ hypothesis is excluded at the 95% CL
- Note that H_μ is for a given Higgs mass m_H



- Find the p-value of the signal hypothesis H $_{\mu}$

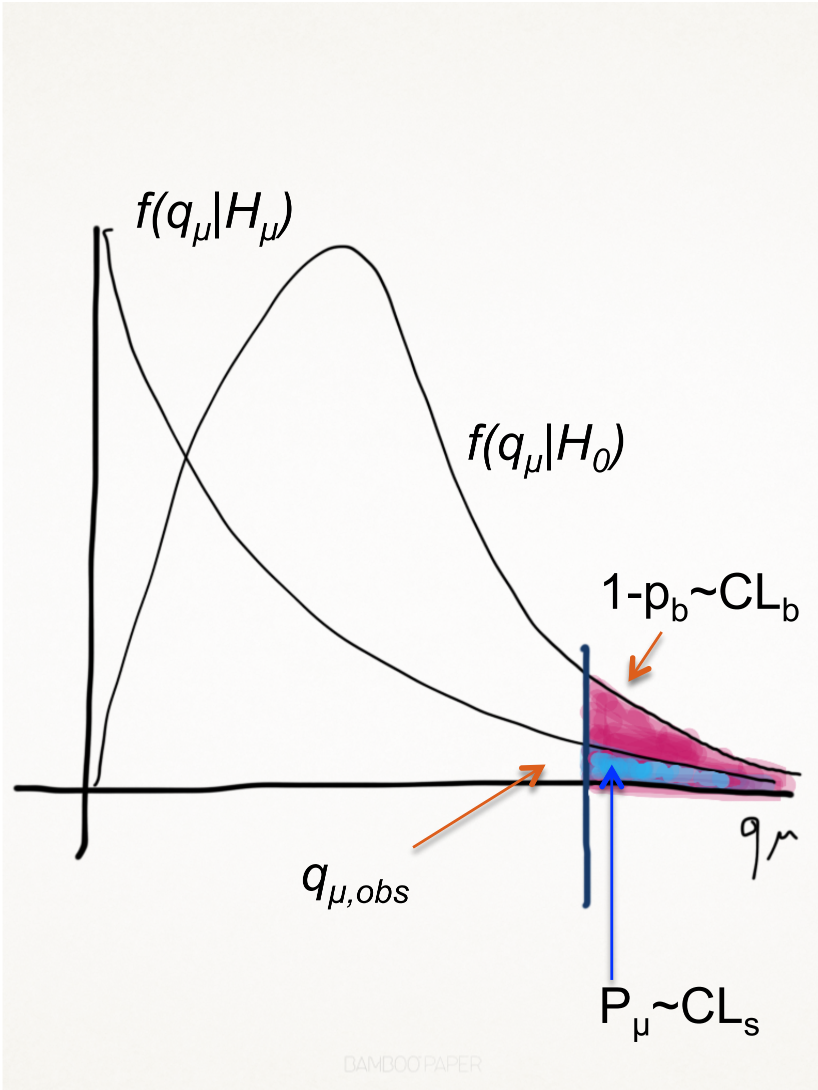
$$p_{\mu} = \int_{q_{\mu,obs}}^{\infty} f(q_{\mu} \mid H_{\mu}) dq_{\mu}$$

- Find the modified p-value

$$p'_{\mu}(m_H) = \frac{p_{\mu}}{1 - p_b}$$

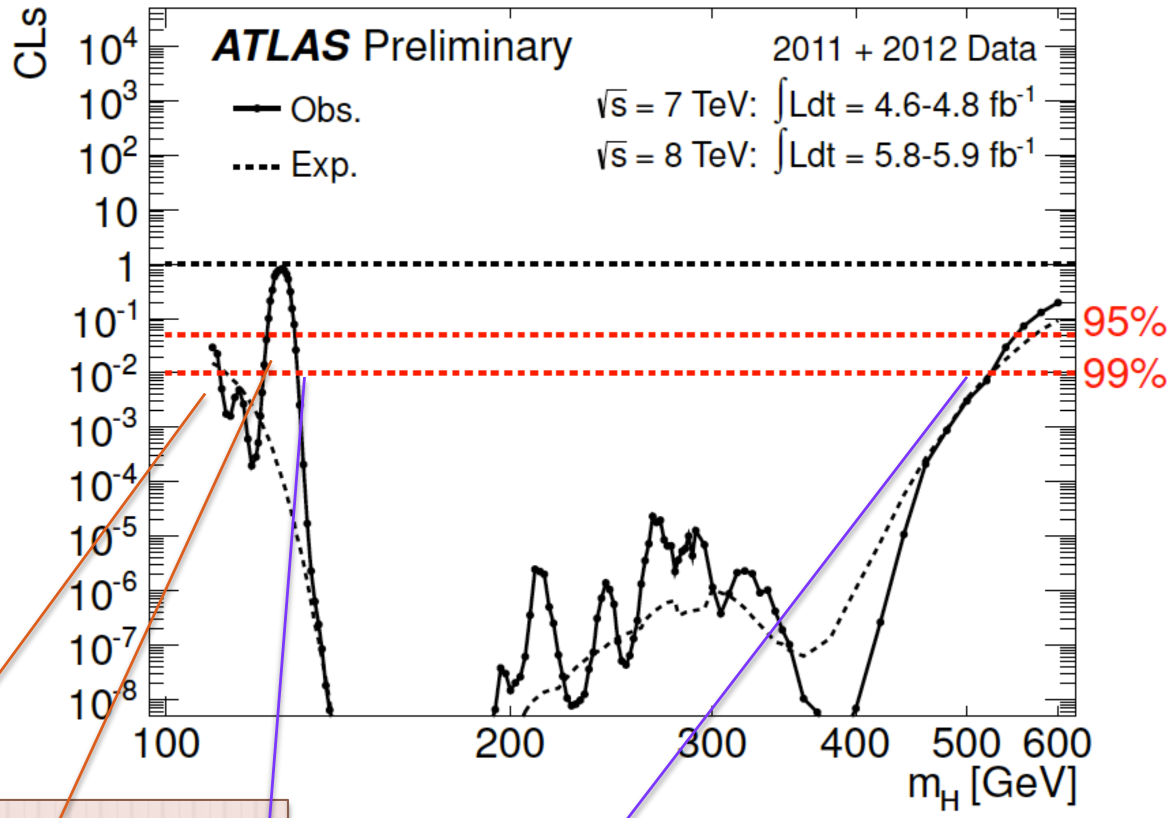
  - Option1: set $\mu = 1$ and find

$$p'_1(m_H) = \frac{p_{\mu}}{1 - p_b} \equiv CLs(m_H)$$



Understanding the CLs plot

- Here, for each Higgs mass m_H, one finds the observed p'_s value, i.e. p'_μ, $\mu = 1$
- This modified p-value, p'_s, is by definition CLs

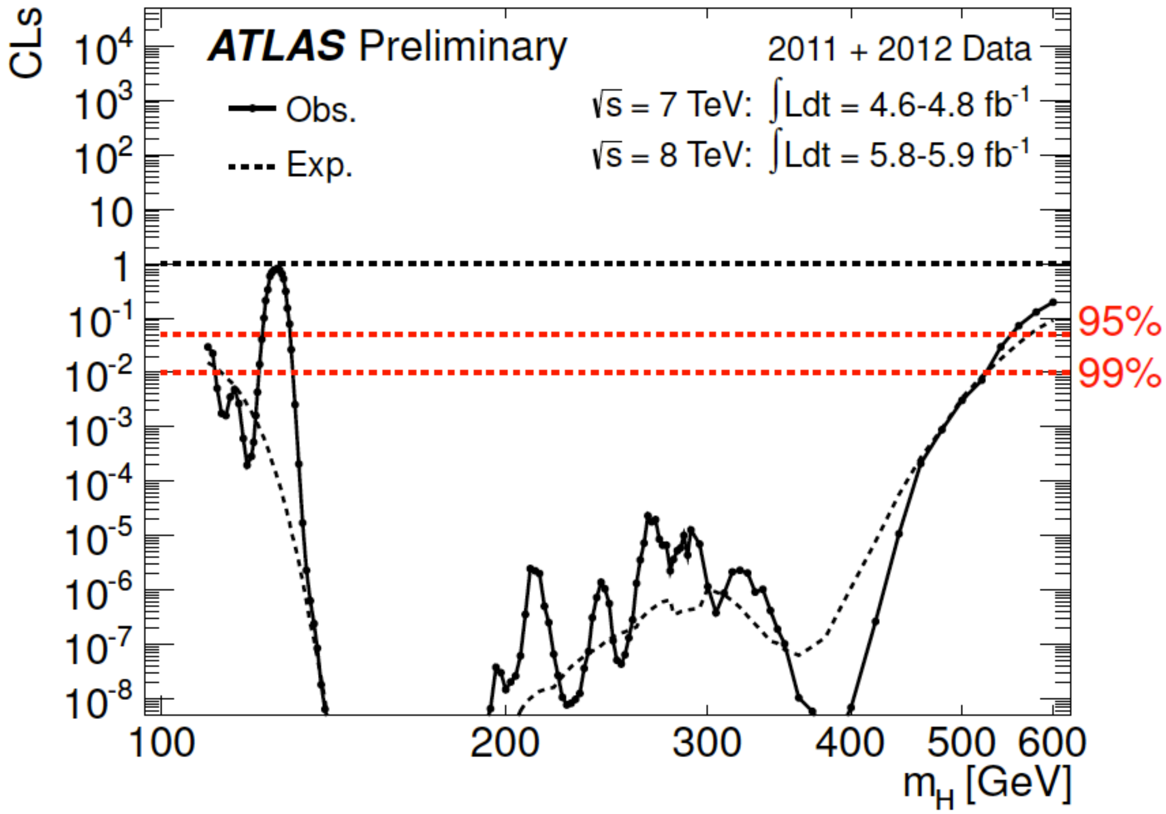


The smaller CLs, the deeper is the exclusion,
Exclusion CL=1-CLs=1-p'_s

to the previous combined search [1]. Figure 2 shows the CL_s values for $\mu = 1$, where it can be seen that the regions between 111.7 GeV to 121.8 GeV and 130.7 GeV to 523 GeV are excluded at the 99% CL.

Understanding the CLs plot

- CLs is the compatibility of the data with the signal hypothesis
- The smaller the CLs, the less compatible the data with the prospective signal



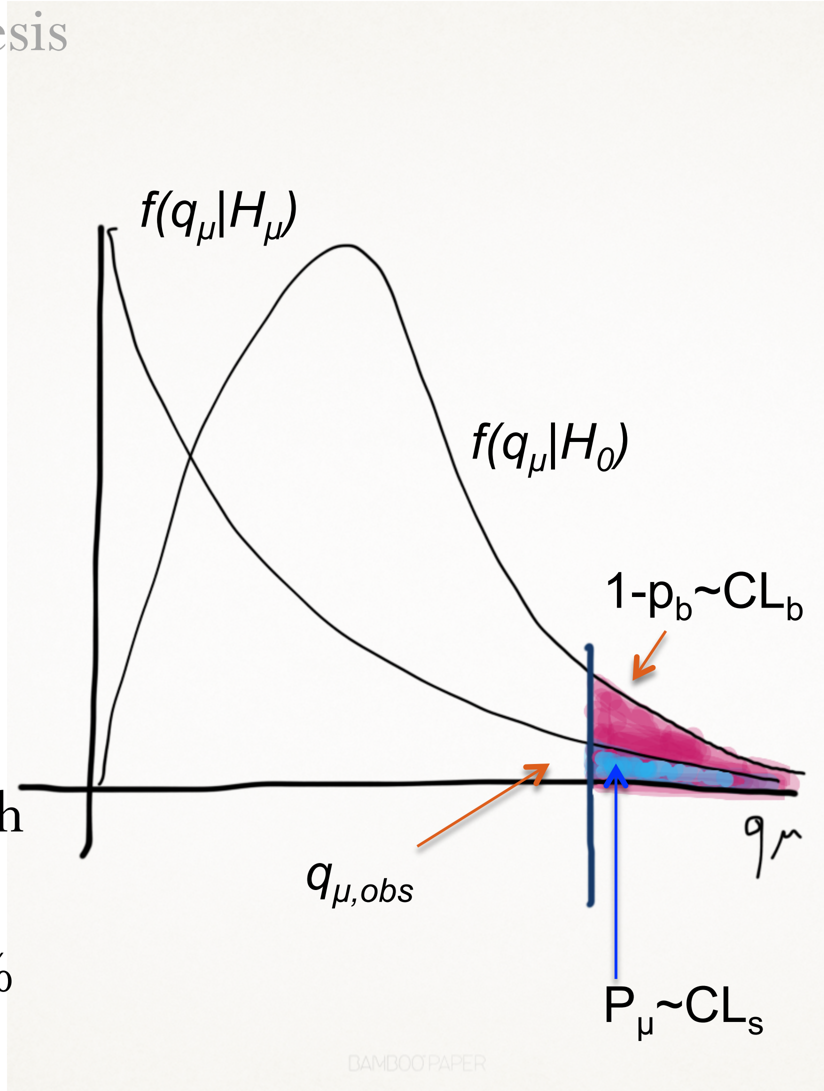
- Find the p-value of the signal hypothesis H_μ

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu \mid H_\mu) dq_\mu$$

- Find the modified p-value

$$p'_\mu(m_H) = \frac{p_\mu}{1 - p_b}$$

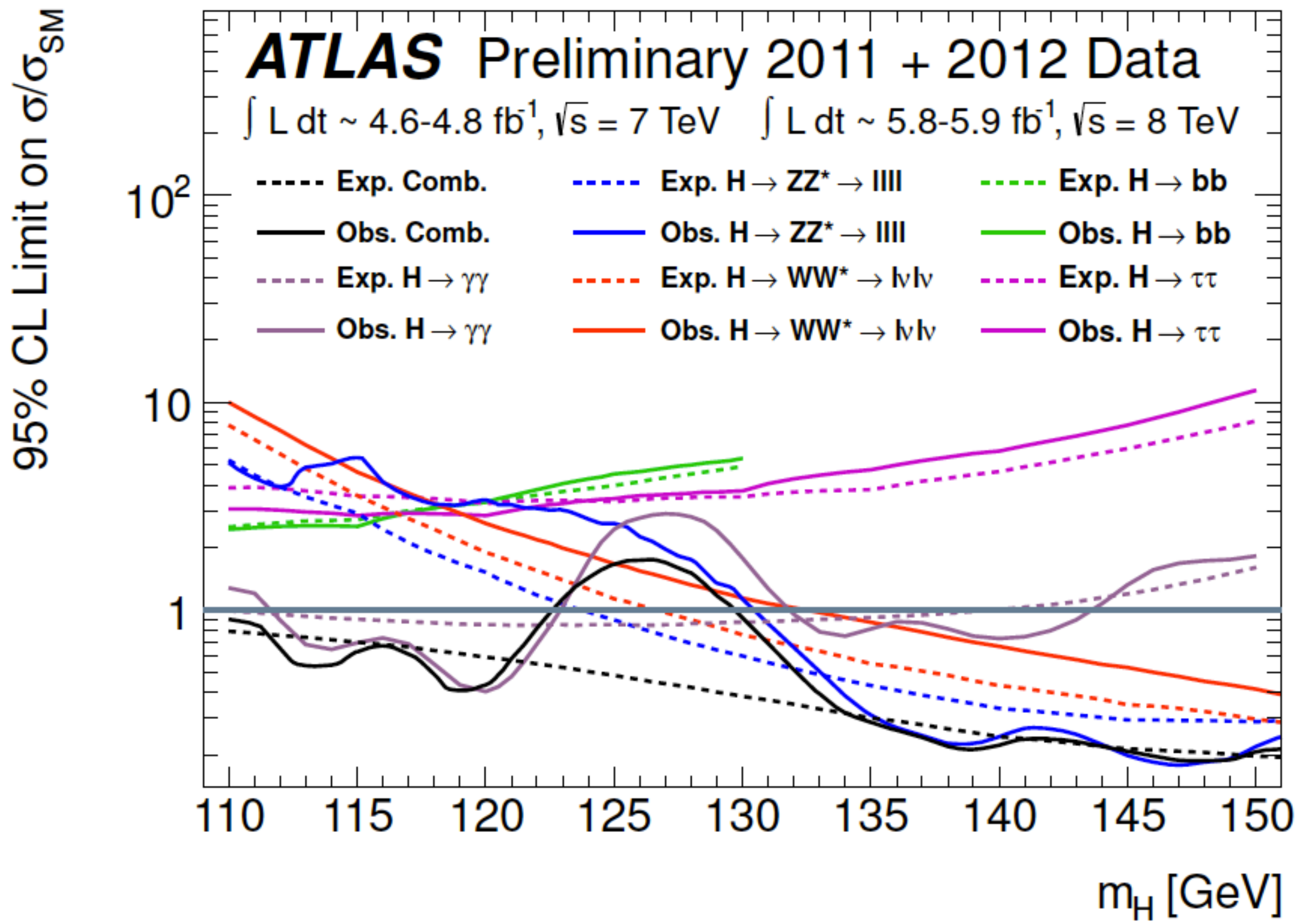
  - Option2: Iterate and find μ for which $p'_\mu(m_H)=5\%$ → $\mu = \mu$ up → If μ up<1, m_H is excluded at the 95%



Exclusion a Higgs with a mass m_H

- First we fix the hypothesized mass to m_H
- We then test the H_μ [$\mu\, s(m_H)+b$] hypothesis
- We find μ_{up}, for which $p'_{\mu\,up}=5\%$-> the $H_{\mu\,up}$ hypothesis is rejected at the 95% CL
- This means that the Confidence Interval of μ is $\mu \in [0, \mu_{up}]$
- If $\mu_{up}=\sigma(mH)/\sigma SM(mH)<1$, we claim that a SM Higgs with a mass m_H is excluded at the 95% CL
- A Higgs with a mass m_H, such that $\mu(m_H)<1$ is excluded at the 95% CL

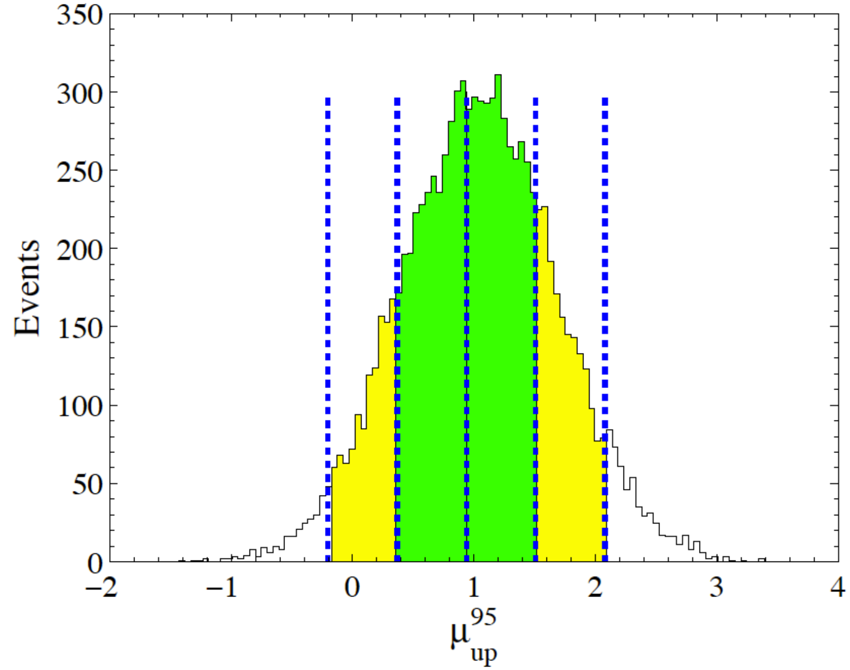
Upper Limit – $\mu_{up}(m_H)$



Sensitivity

- The sensitivity of an experiment to excude a Higgs with a mass m_H is tl median upper limit
- $$\mu_{up}^{med} = med\{\mu_{up} \mid H_0\}$$
- The 68% (green) and 95% (yellow) are the
  1 and 2 σ bands
- The median and the bands can be derived with the Asimov background only dataset n=<n>=b

Distribution of the upper limit with background only experiments



The Asimov data set is n=b
-> median upper limit

CCGV Useful Formulae – The Bands

$$\mu_{up}^{med} = \sigma \Phi^{-1}(1-0.5\alpha) = \sigma \Phi^{-1}(0.975)$$

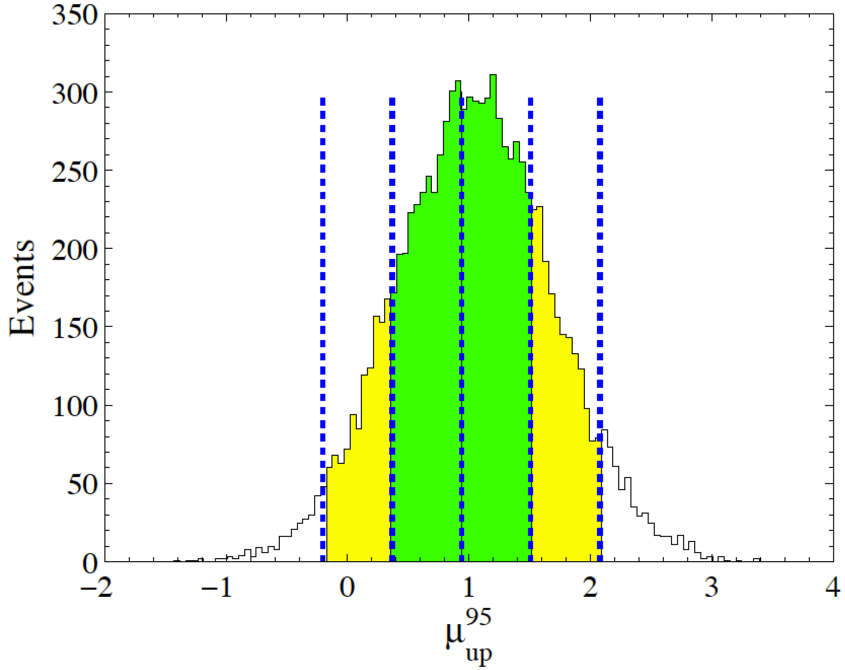
$$\sigma_{\hat{\mu}}^2 = Var[\hat{\mu}]$$

$$\mu_{up+N} = N\sigma_0 + \sigma_{\mu_{up+N}} (\Phi^{-1}(1-\alpha\Phi(N))$$

$$\alpha = 0.05$$

$$\sigma_{\mu_{up+N}}^2 = \frac{\mu_{up+N}^2}{q_{\mu_{up+N},A}}$$

Distribution of the upper limit with background only experiments



The Asimov data set is n=b
-> median upper limit

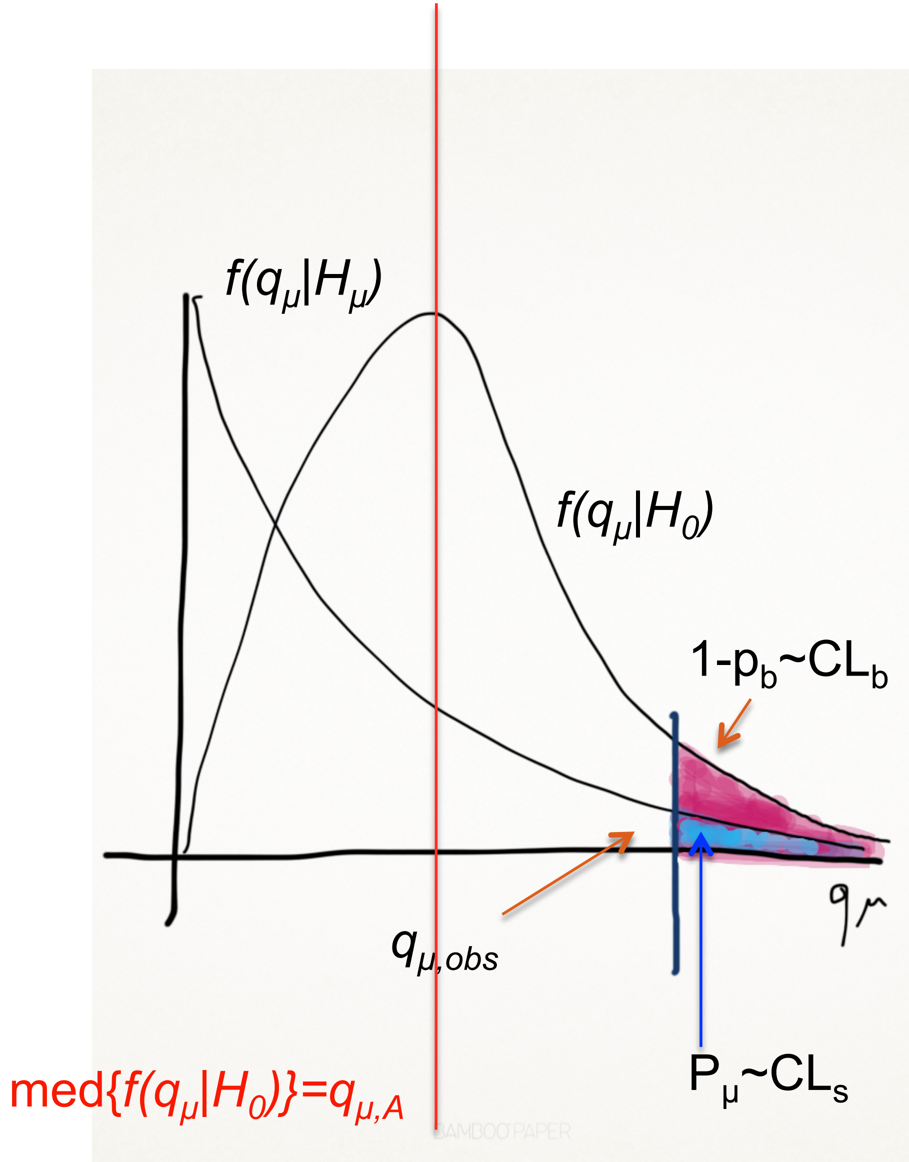
The Asimov data set

- The median of $f(q_{\mu} \,|\, H_0)$

Can be found by plugging in the unique Asimov data set representing the H_0 hypothesis, background only

n=<n>=b

- The sesnitivity f the experiment for searching the Higgs at mass m_H with a signal strength μ , is given by p'_{μ} evaluated at $q_{\mu,A}$



The ASIMOV data sets

- The name of the Asimov data set is inspired by the short story *Franchise, by Isaac Asimov [1]. In it, elections are held by* selecting a single voter to represent the entire electorate.

- The "Asimov" Representative Data-set for Estimating Median Sensitivities with the Profile Likelihood
  G. Cowan, K. Cranmer, E. Gross , O. Vitells

- [1] Isaac Asimov, *Franchise, in Isaac Asimov: The Complete Stories, Vol. 1, Broadway Books, 1990.*

The Asimov Data Set

Franchise (short story)

From Wikipedia, the free encyclopedia

This article **needs additional citations for verification**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. *(December 2009)*

"Franchise"	
Author	Isaac Asimc
Country	United State
Language	English
Series	Multivac
Genre(s)	science ficti
Published in	*If*
Publisher	Quinn Publis
Media type	Magazine
Publication date	August 195
Preceded by	"Question"
Followed by	"The Dead I

Franchise is a science fiction short story by Isaac Asimov. It first appeared in the August 1955 issue of the magazine *If: Worlds of Science Fiction*, and was reprinted in the collections *Earth Is Room Enough* (1957) and *Robot Dreams* (1986). It is one of a loosely connected series of stories concerning a fictional computer called Multivac. It is the first story in which Asimov dealt with computers *as computers* and not as immobile robots.

Plot summary [edit]

In the future, the United States has converted to an "electronic democracy" where the computer Multivac selects a single person to answer a number of questions. Multivac will then use the answers and other data to determine what the results of an election would be, avoiding the need for an actual election to be held.

The story centers around Norman Muller, the man chosen as "Voter of the Year" in 2008. Although the law requires him to accept the dubious honour, he is not sure that he wants the responsibility of representing the entire electorate, worrying that the result will be unfavorable and he will be blamed.

However, after 'voting', he is very proud that the citizens of the United States had, through him, "exercised once again their free, untrammeled franchise" - a statement that is somewhat ironic as the citizens didn't actually get to vote.

The idea of a computer predicting whom the electorate would vote for instead of actually holding an election was probably inspired by the UNIVAC I's correct prediction of the result of the 1952 election.

Influence

The use of a single representative individual to stand in for the entire population can help in evaluating the sensitivity of a statistical method. *Franchise* was cited as the inspiration of the data set", where an ensemble of simulated experiments can be replaced by a single representative one. [1]

References

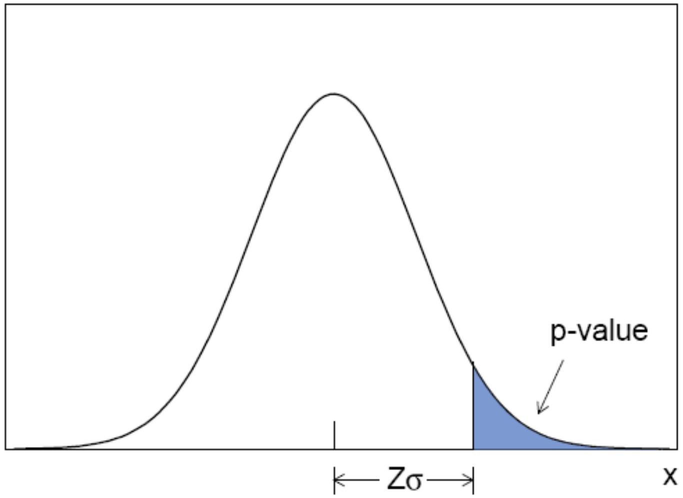
1. ^ G. Cowan, K. Cranmer, E. Gross, and O. Vitells (2011). "Asymptotic formulae for likelihood-based tests of new physics". *Eur.Phys.J.* **C71**: 1554. DOI:10.1140/epjc/s10052-011-1554-0.

Useful Formulae

$$p'_{\mu_{95}} = \frac{1-\Phi(\sqrt{q_{\mu_{95}}})}{\Phi(\sqrt{q_{\mu_{95},A}}-\sqrt{q_{\mu_{95}}})} = 0.05$$

Φ is the cumulative distribution of the standard (zero mean, unit variance) Gaussian.

$q_{\mu_{95},A}$ Is evaluated with the Asimov data set (background only)

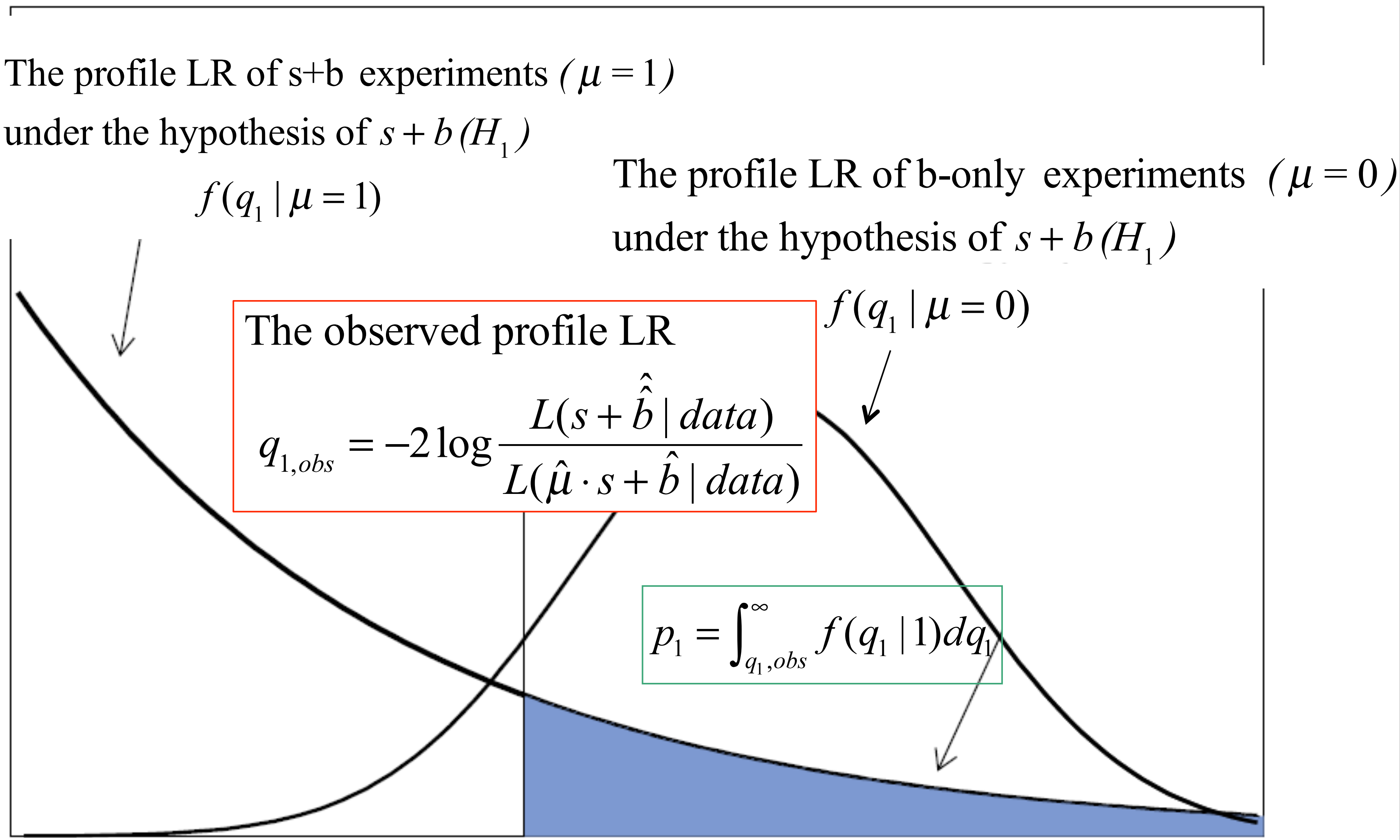


$$p = \int^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 1-\Phi(Z)$$

$$Z = \Phi^{-1}(1-p)$$

Exclusion – Illustrated

$$\lambda(\mu=1)=\frac{L(s+\hat{\hat{b}}\,|\,data)}{L(\hat{\mu}\cdot s+\hat{b}\,|\,data)},\quad q_1=-2\log\lambda(\mu=1)$$



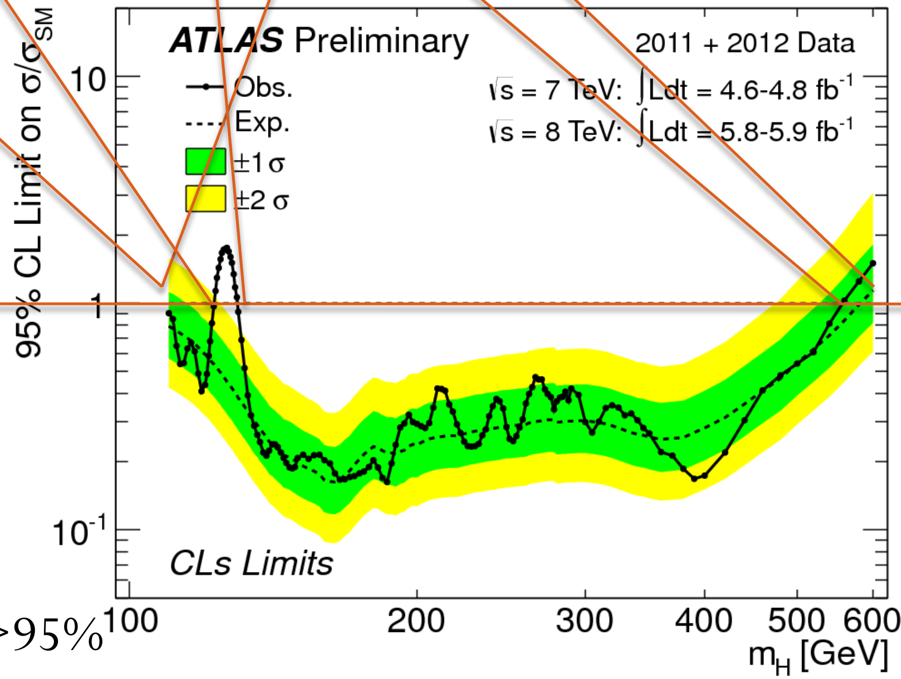
p_1 is the level of compatibility between the data and the Higgs hypothesis
If p_1 is smaller than 0.05 we claim an exclusion at the 95% CL

Understanding the Brasil Plot

to within a few percent.

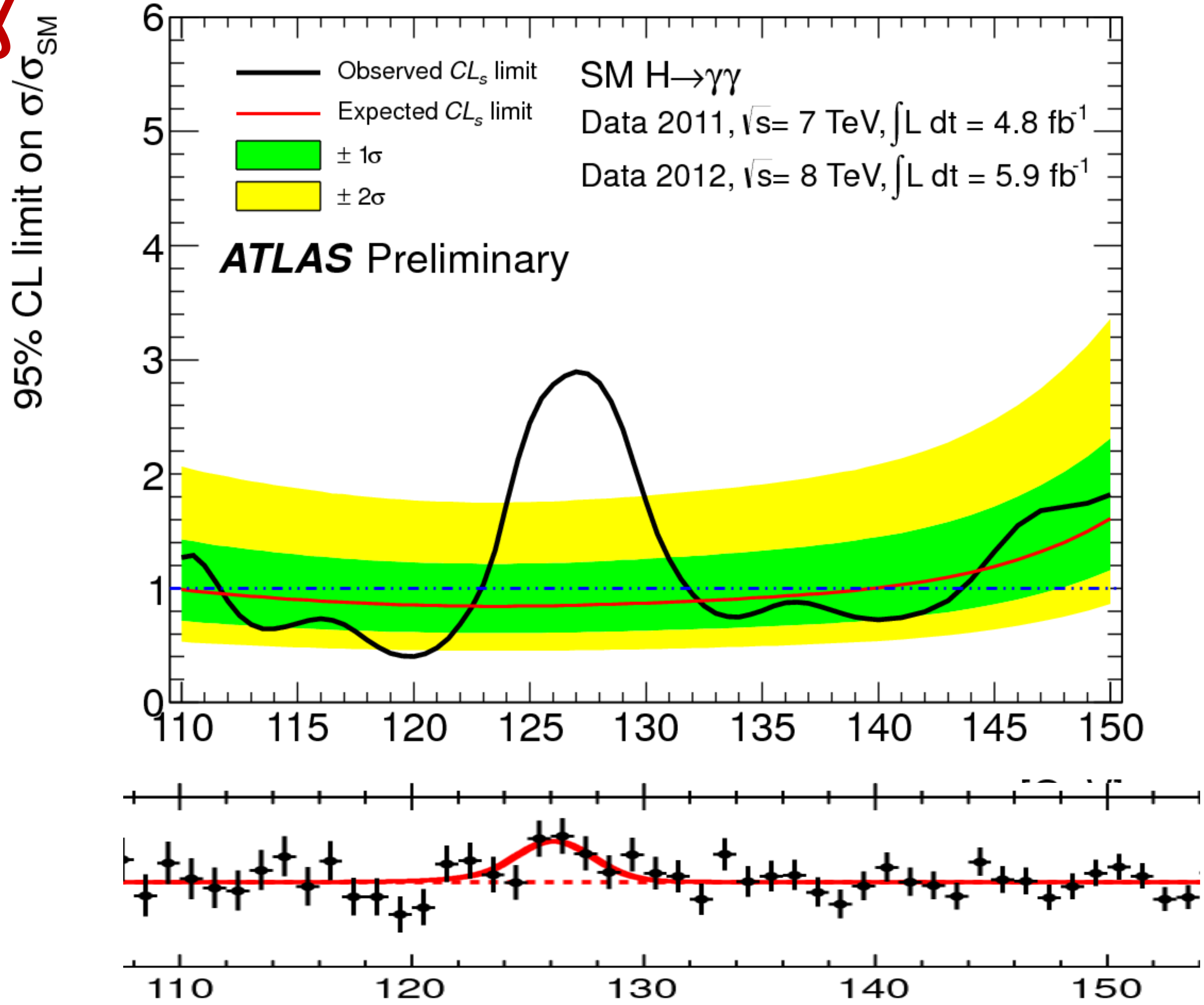
The expected 95% CL exclusion region covers the m_H range from 110 GeV to 582 GeV. The observed 95% CL exclusion regions are from 110 GeV to 122.6 GeV and 129.7 GeV to 558 GeV. The addition of

- $\mu_{up} = \sigma(m_H)/\sigma_{SM}(m_H) < 1 \rightarrow$
  $\sigma(m_H) < \sigma_{SM}(m_H) \rightarrow$ SM m_H excluded



- The line μ up=1 corresponds to CLs=5% (p'_s=5%)
- If μ up<1 the exclusion of a SM Higgs is deeper$\rightarrow$p's<5%, p's=CLs$\rightarrow$CL=1-p's>95%

H→γγ

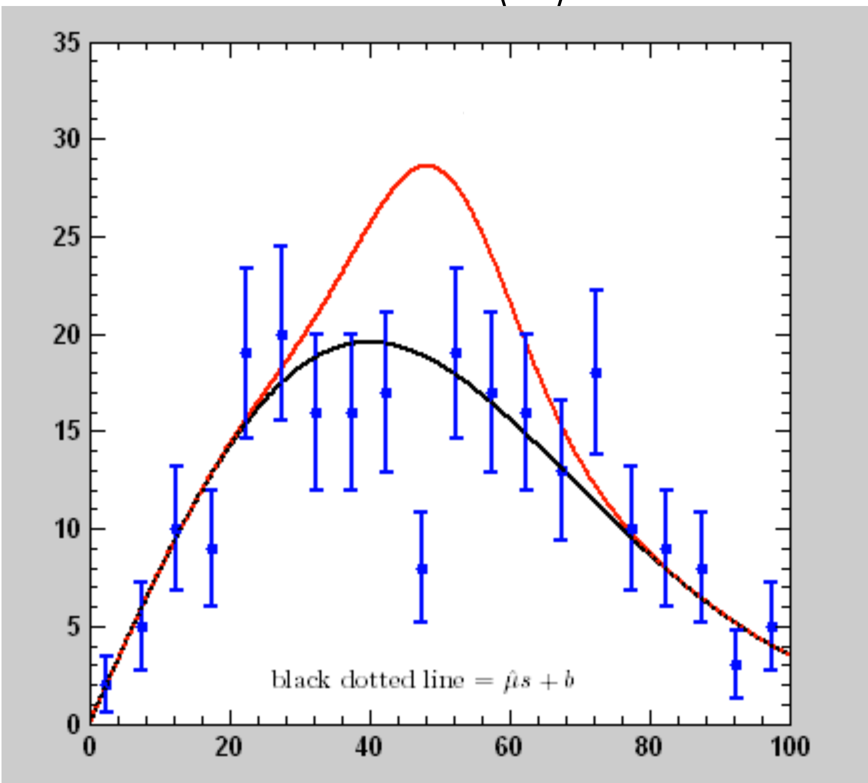

95% CL limit on σ/σ_SM
Observed CL_s limit
Expected CL_s limit
± 1σ
± 2σ
SM H→γγ
Data 2011, √s= 7 TeV, ∫L dt = 4.8 fb⁻¹
Data 2012, √s= 8 TeV, ∫L dt = 5.9 fb⁻¹
ATLAS Preliminary
0
1
2
3
4
5
6
110
115
120
125
130
135
140
145
150
110
120
130
140
150

DISCOVERY

The Toy Physics Model

- The NULL hypothesis H_0: SM without Higgs Background Only

$$\langle n \rangle = b$$

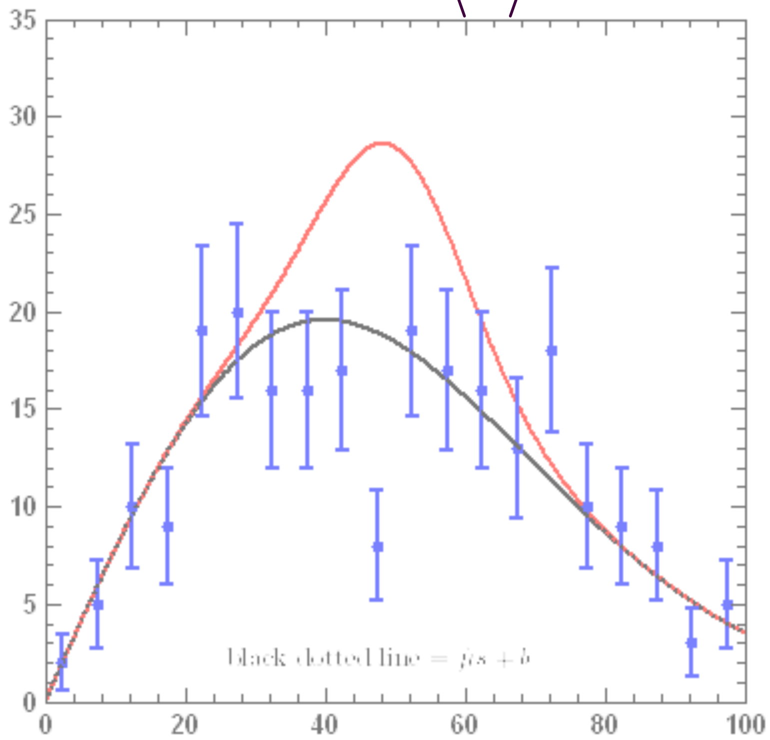


mass

The Toy Physics Model

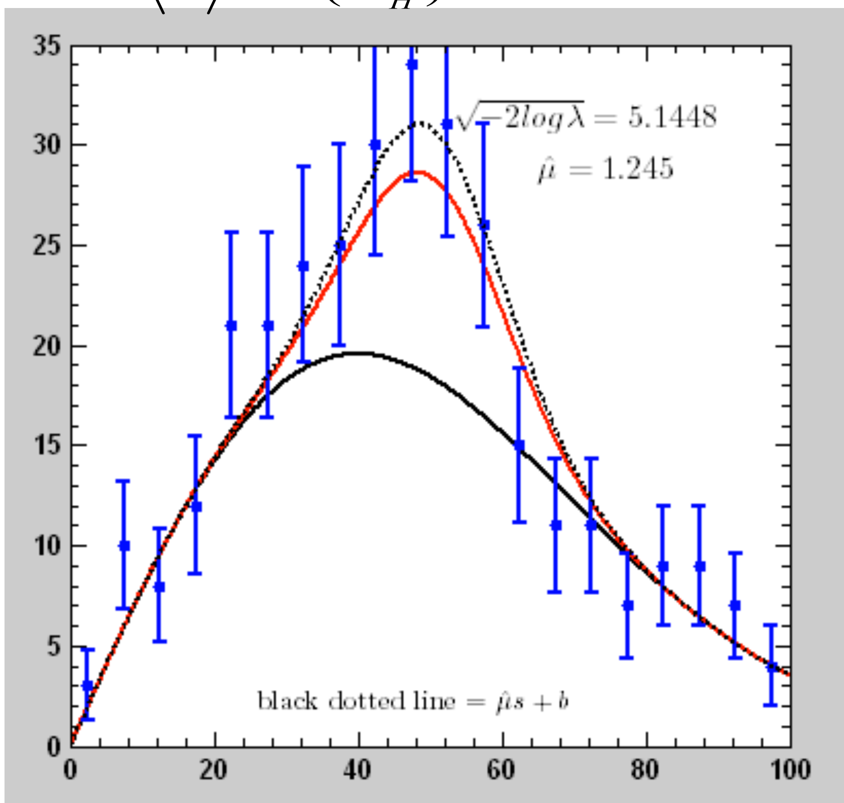
- The NULL hypothesis H_0: SM without Higgs Background Only

$$\langle n \rangle = b$$



- The alternate Hypothesis H_1:
- SM with a Higgs with a mass m_H

$$\langle n \rangle = s(m_H) + b$$

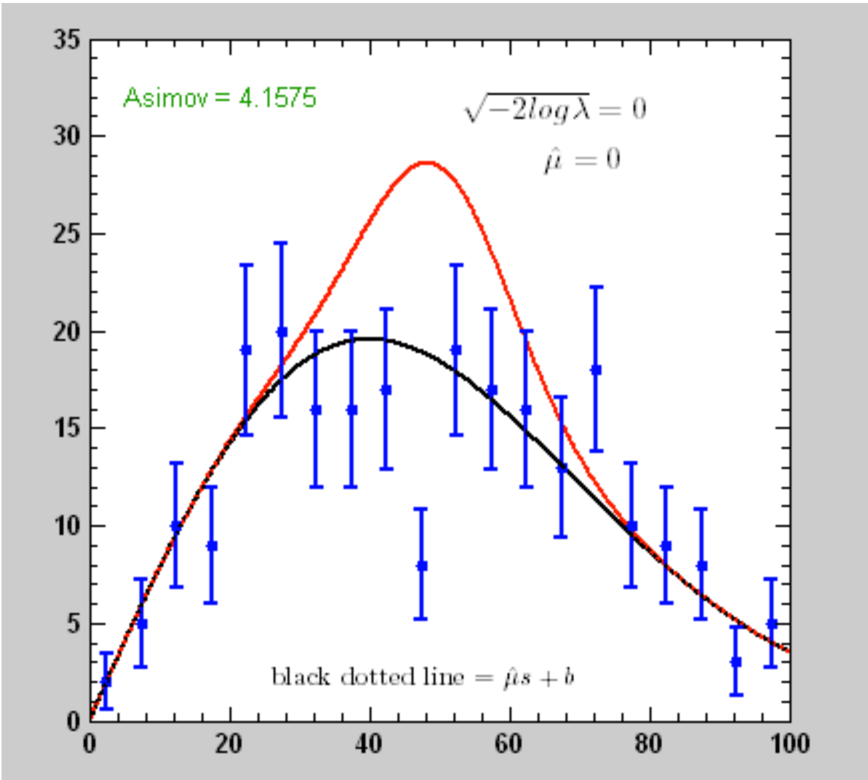


The Toy Physics Model

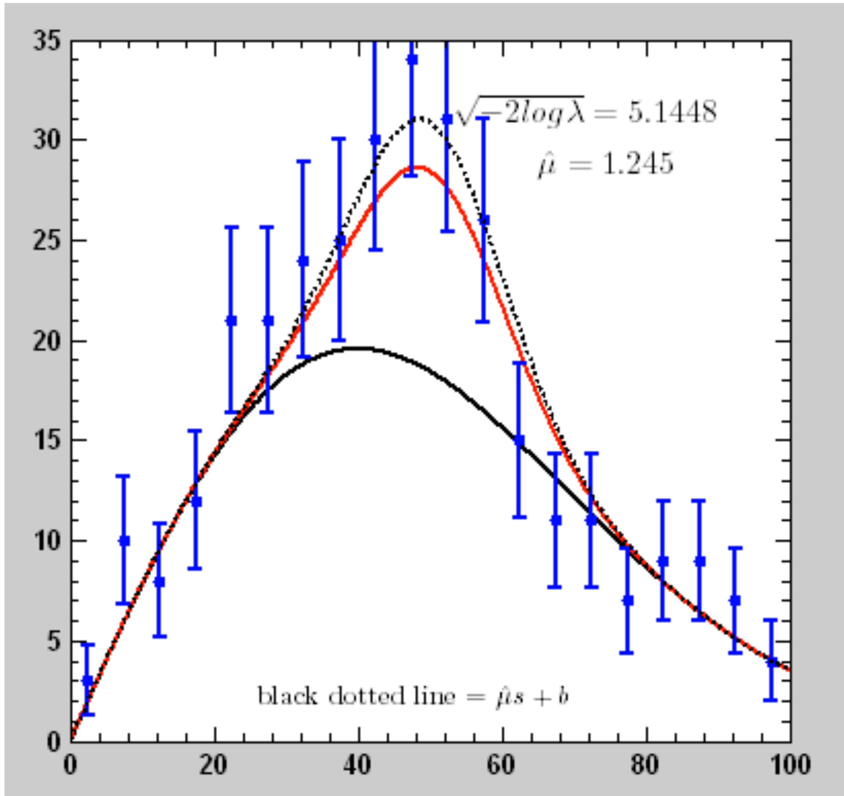
$n = \mu s + b$ $\qquad$ $n = \mu s + b$

$MLE \ \ \hat{\mu}$ $\qquad$ $MLE \ \ \hat{\mu}$

$\langle \hat{\mu} \rangle = 0 \ \ under \ \ H_0$ $\qquad$ $\langle \hat{\mu} \rangle = 1 \ \ under \ \ H_1$



mass



mass

The Profile Likelihood ("PL")

For discovery we test the H_0 null hypothesis and try to reject it

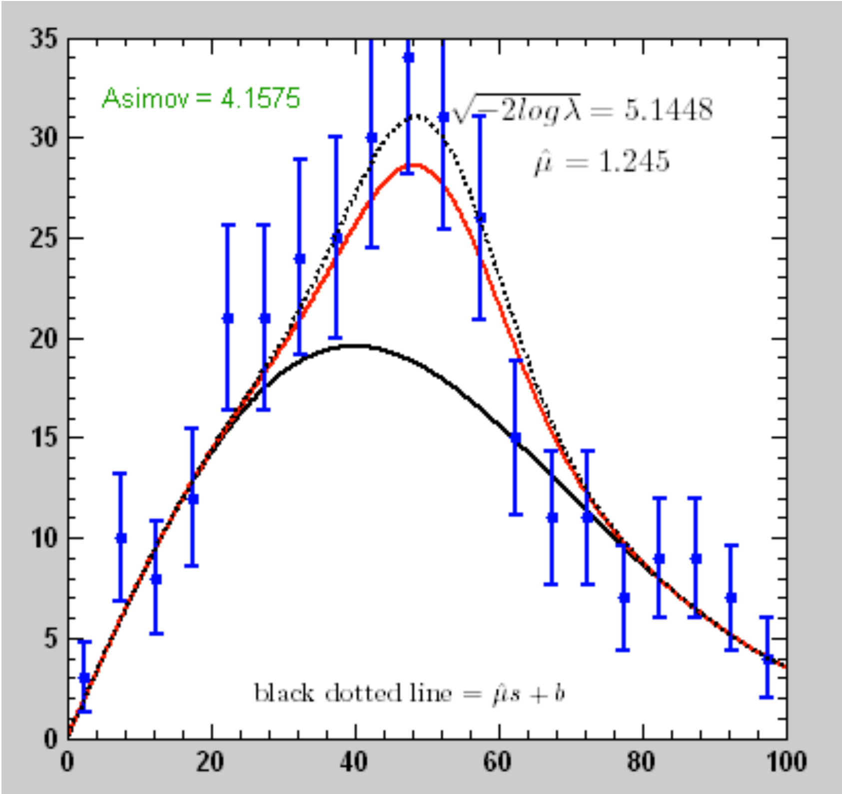
$$q_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s+b)}$$

For
$$\hat{\mu} \sim 0,\ q \text{ small}$$
$$\hat{\mu} \sim 1,\ q \text{ large}$$

In general: testing the H_μ hypothesis i.e., a SM with a signal of strength μ,

$$q_\mu = -2\ln\frac{L(\mu)}{L(\hat{\mu})}$$

$$\boxed{q_{0,obs} = Z^2}$$

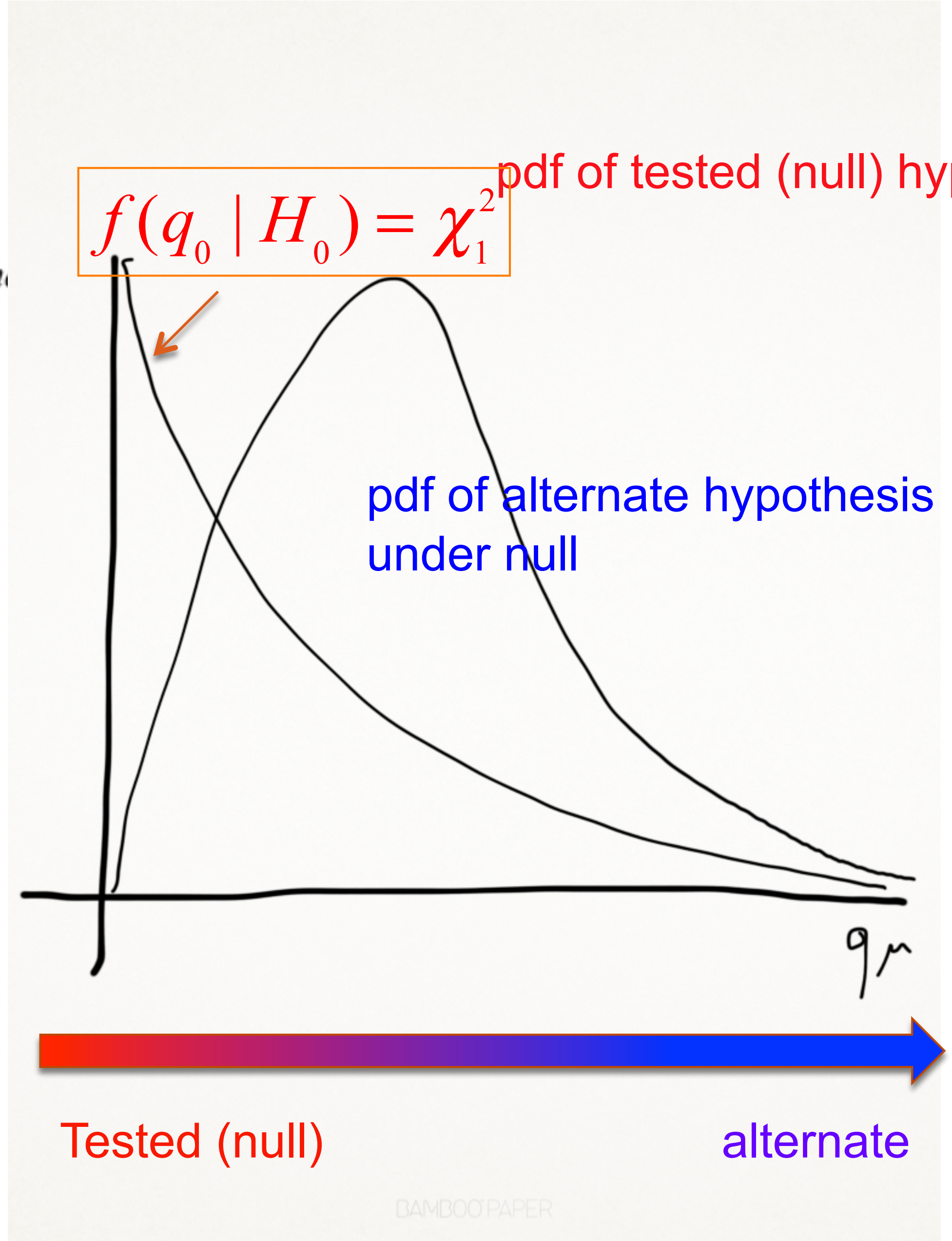


Wilks Theorem

S.S. Wilks, *The large-sample distribution of th*
Ann. Math. Statist. **9** (1938) 60-2.

- Under a set of regularity conditions and for a sufficiently large data sample, *Wilks' theorem says that* the pdf of the statistic q under the null hypothesis approaches a chi-square PDF for one degree of freedom

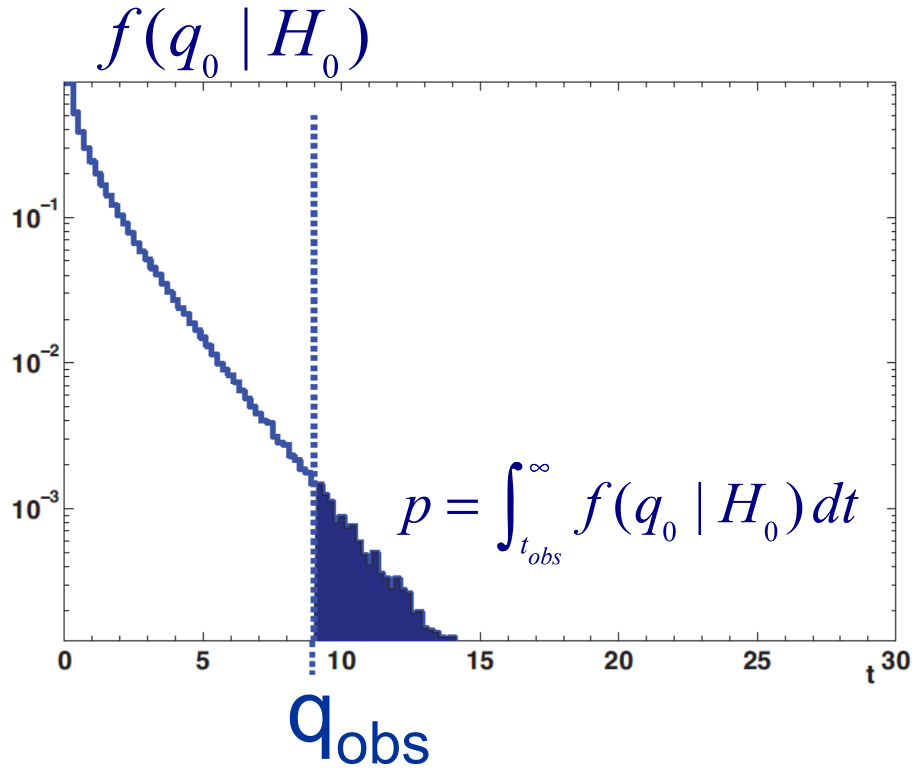
$$\boxed{f(q_0 \mid H_0) = \chi_1^2} \quad \boxed{f(q_\mu \mid H_\mu) \sim \chi_1^2}$$



Significance & p-value

- Calculate the test statistic based on the observed experimental result (after taking tons of data), q_{obs}
- Calculate the probability that the observation is as or less compatible with the background only hypothesis (p-value)

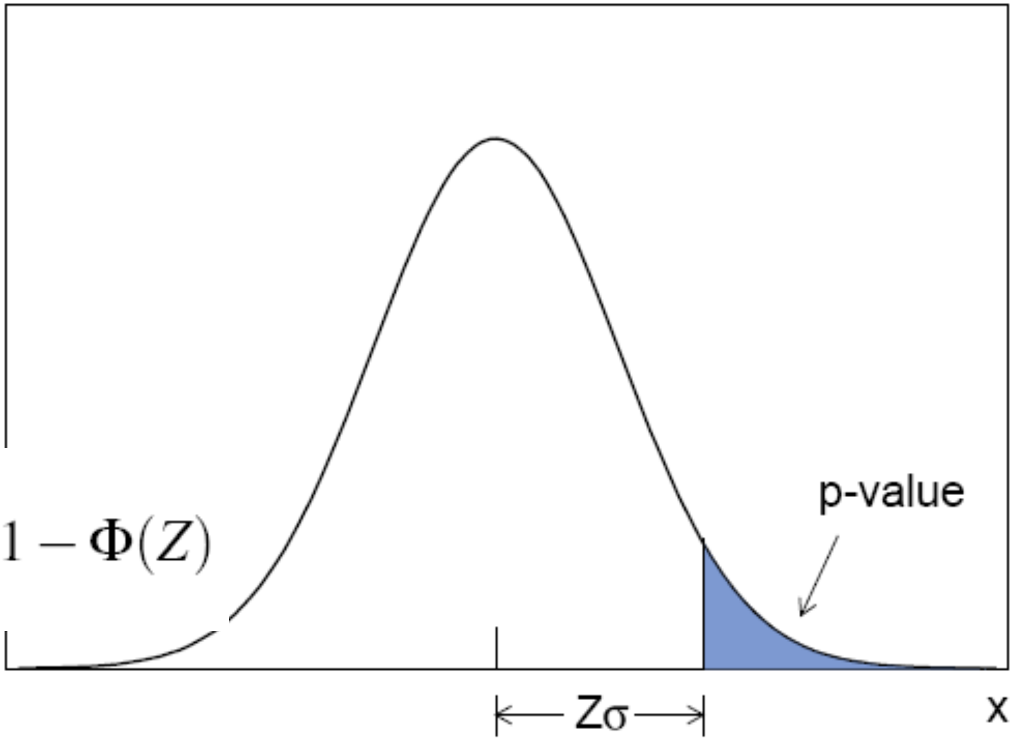
$$p = \int_{q_{obs}}^{\infty} f(q_0 \mid H_0)\, dt$$



If p-value< $2.8 \cdot 10^{-7}$, we claim a 5σ discovery

From p-values to Gaussian Significance

- It is a custom to express the p-value as the significance associated to it, had the PDF been Gaussians



$$p = \int_Z^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 1 - \Phi(Z)$$

$$Z = \Phi^{-1}(1-p)$$

A significance of $Z = 5$ corresponds to $p = 2.87 \times 10^{-7}$.

A significance of Z=1.64 corresponds to p=5%

The Profile Likelihood ("PL")

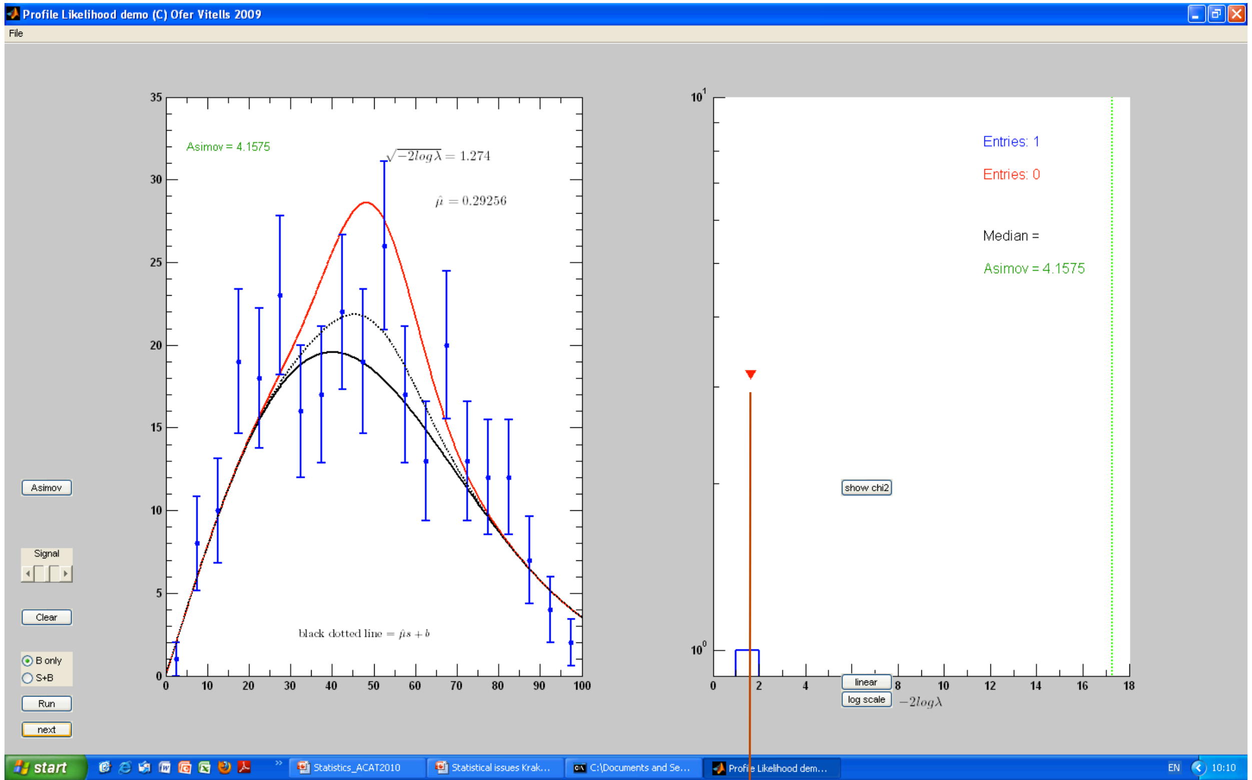
The best signal $\hat{\mu} = 0.3 \rightarrow 1.27\sigma$

$$q_{0,obs} = Z^2$$

$$q_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s + b)}$$

$$\hat{\mu} \sim 0,\ q \text{ small}$$

$$\hat{\mu} \sim 1,\ q \text{ large}$$

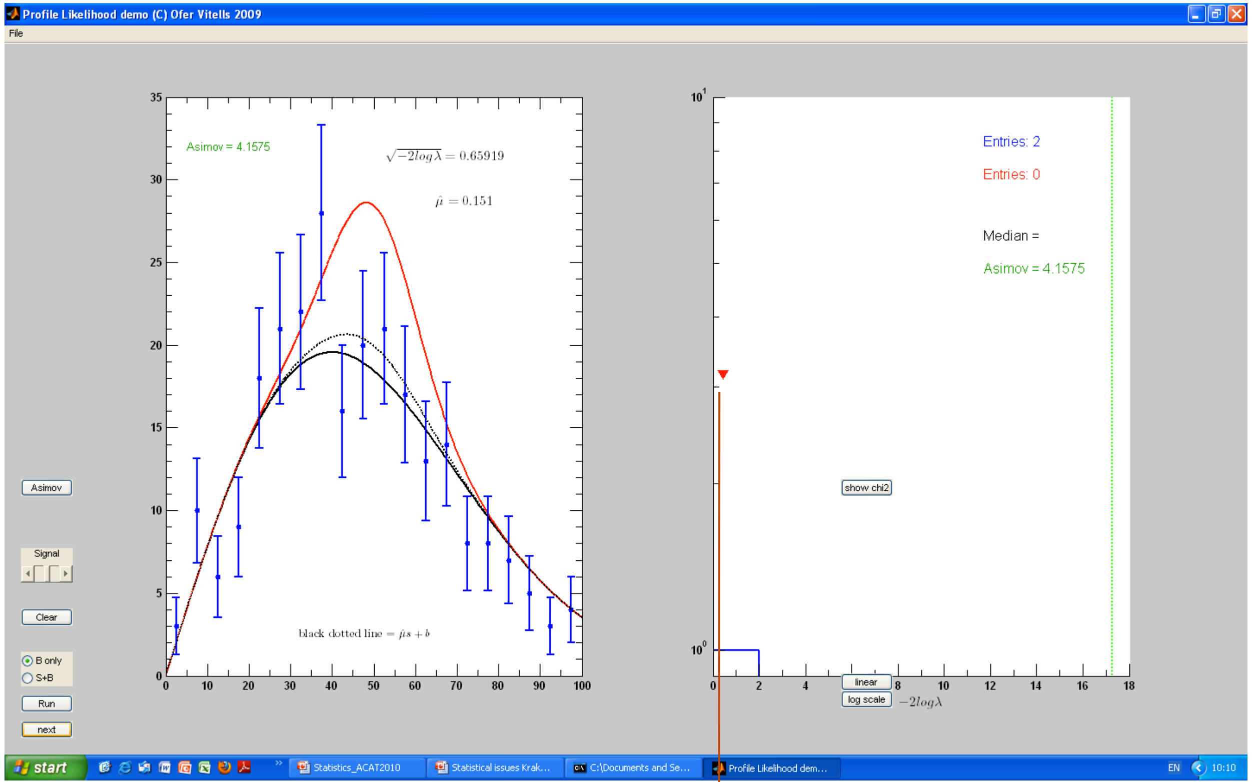


$$q = 1.6 \rightarrow Z = \sqrt{1.6} = 1.27$$

PL: test q_0 under BG only ; $f(q_0 | H_0)$

$$\hat{\mu} = 0.15 \rightarrow 0.6\sigma$$

$$q_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s + b)}$$

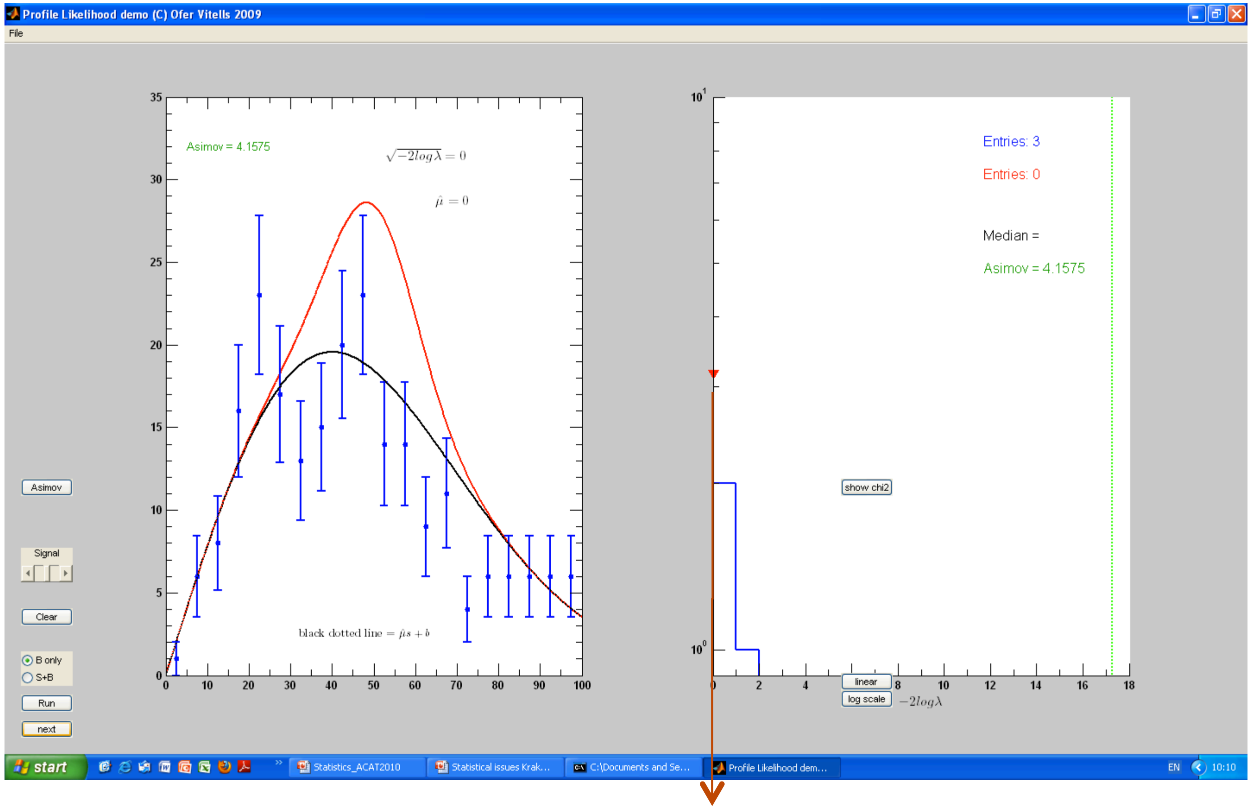


$$q = 0.43 \rightarrow Z = 0.66\sigma$$

PL: test q_0 under BG only ; $f(q_0 | H_0)$

$\hat{\mu} = 0$

$$q_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s + b)}$$

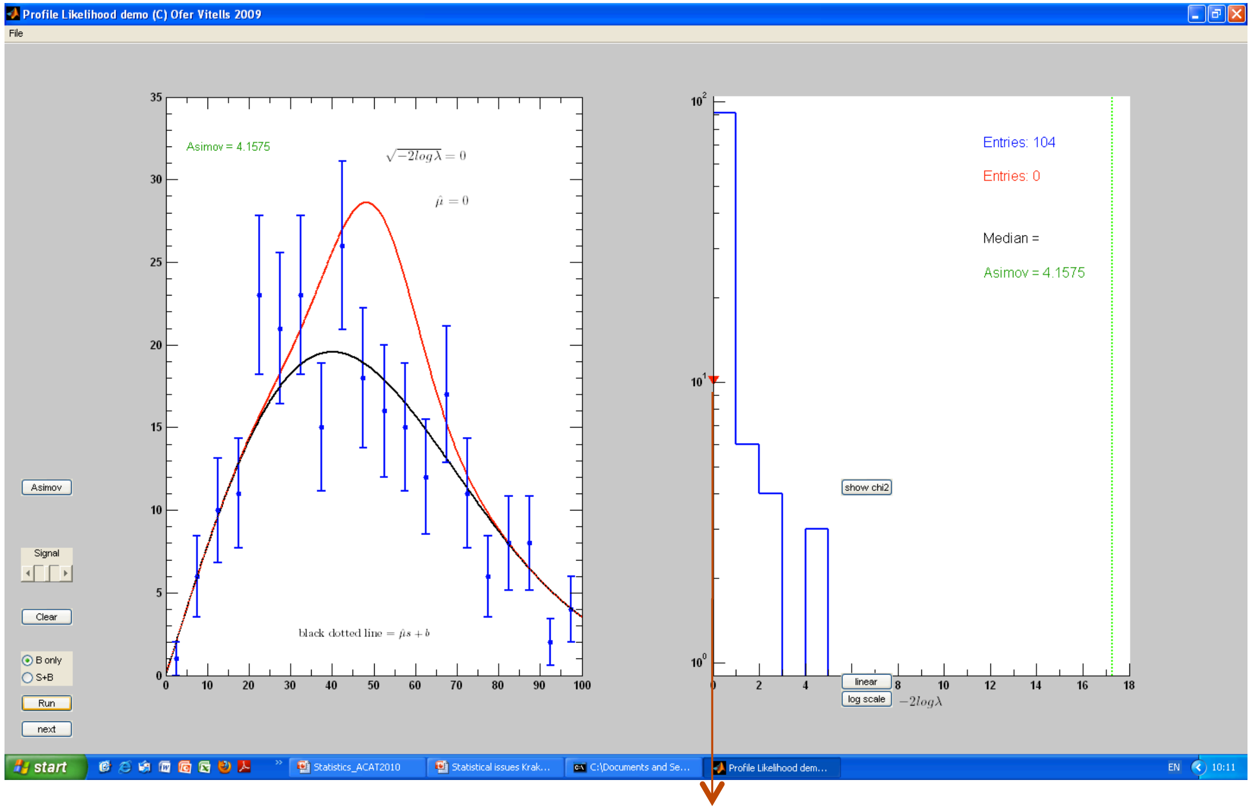


$q = 0$

PL: test q_0 under BG only ; $f(q_0 | H_0)$

$$\hat{\mu} = 0$$

$$q_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s + b)}$$

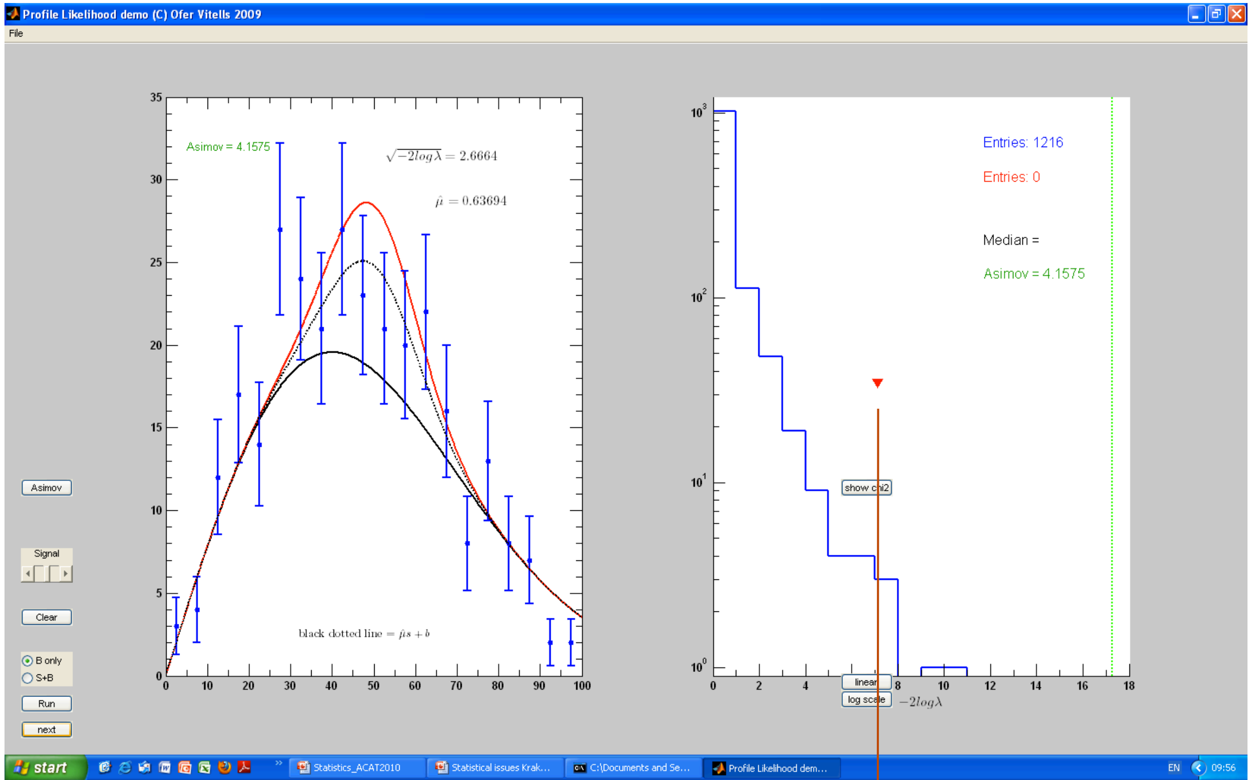


$$q = 0$$

PL: test q_0 under BG only ; $f(q_0 | H_0)$

$$\hat{\mu} = 0.6 \rightarrow 2.6\sigma$$

$$q_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s + b)}$$

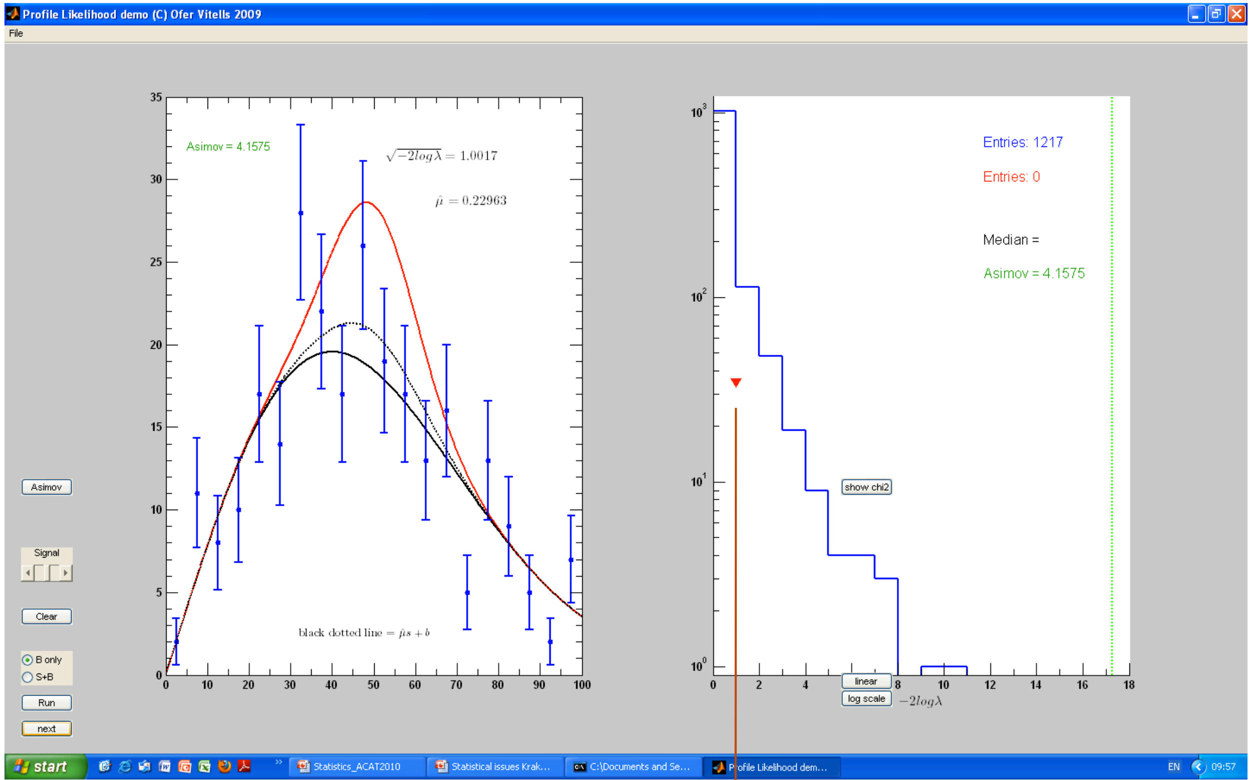


$$q = 6.76 \rightarrow Z = 2.6\sigma$$

PL: test q_0 under BG only ; $f(q_0|H_0)$

$$\hat{\mu} = 0.22 \rightarrow 1.1\sigma$$

$$q_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s+b)}$$

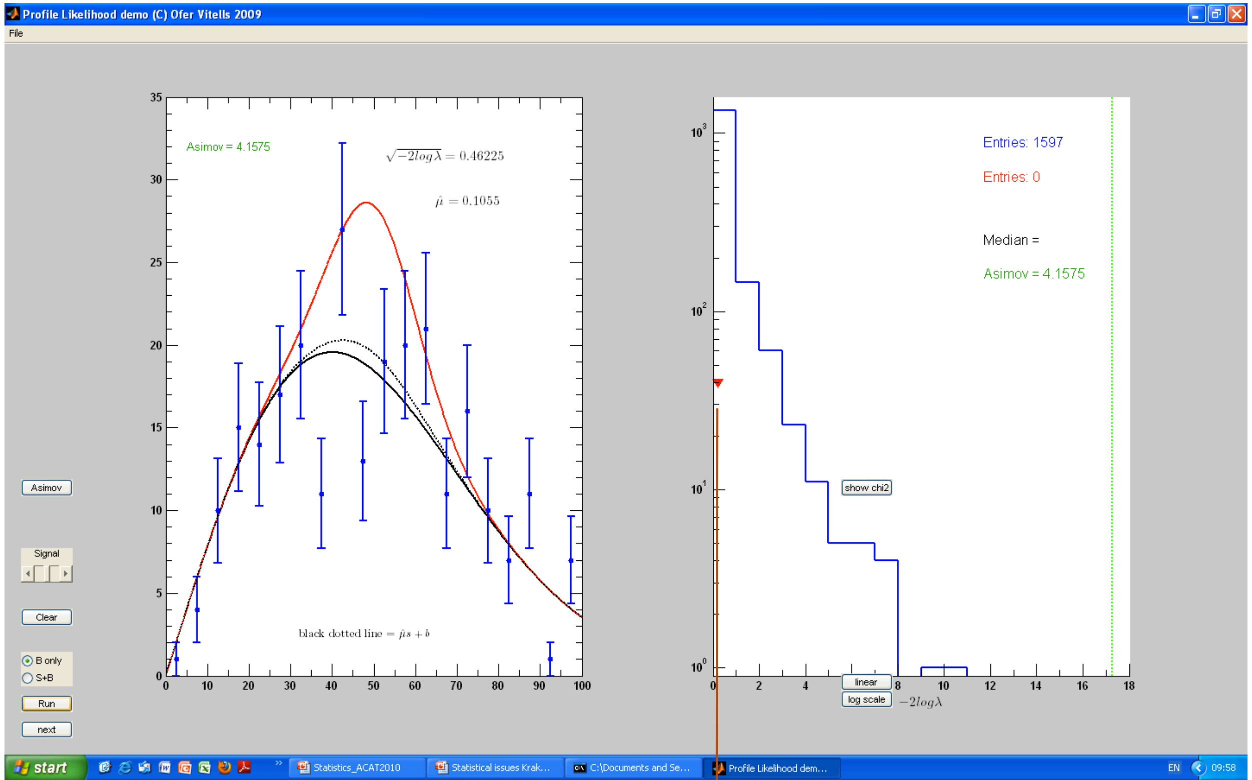


$$q = 1.2 \rightarrow Z = 1.1\sigma$$

PL: test q_0 under BG only ; $f(q_0|H_0)$

$$\hat{\mu} = 0.11 \to 0.4\sigma$$

$$q_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s+b)}$$

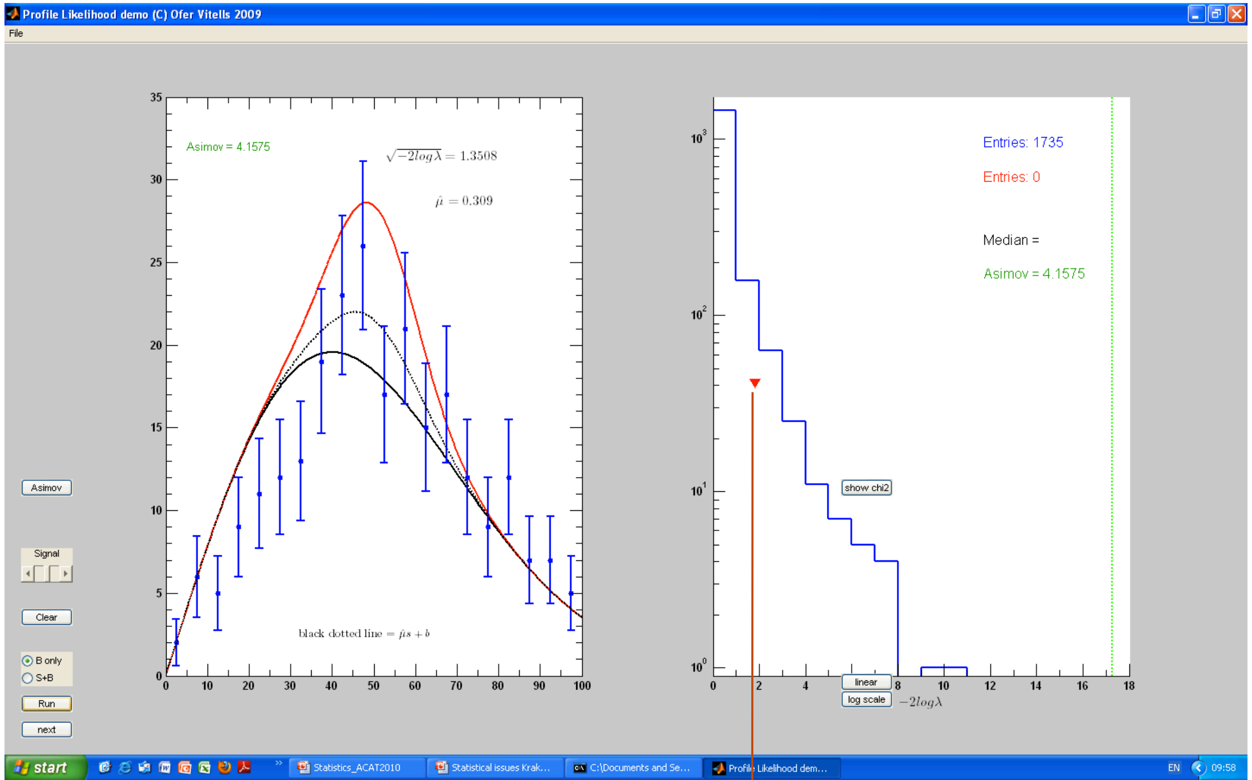


$$q = 0.16 \to Z = 0.4\sigma$$

PL: test q_0 under BG only ; $f(q_0|H_0)$

$$\hat{\mu} = 0.31 \rightarrow 1.35\sigma$$

$$q_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s + b)}$$

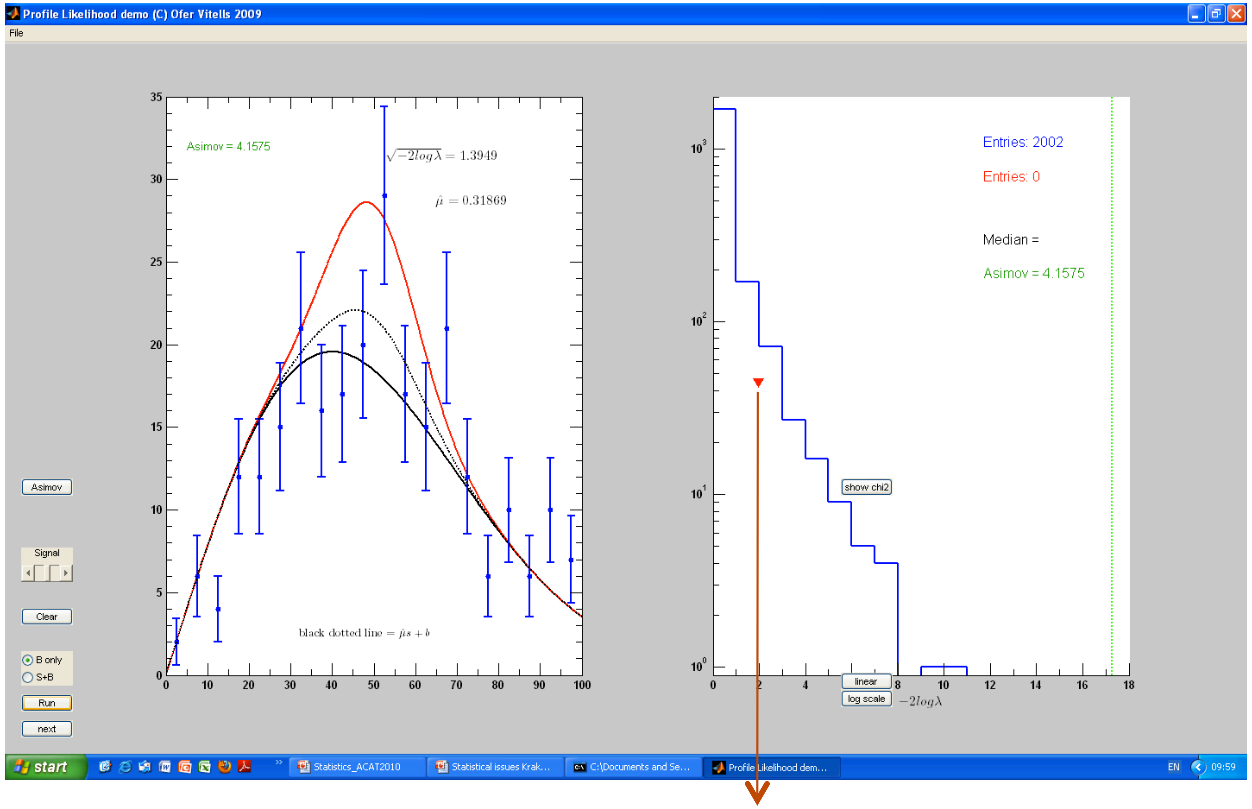


$$q = 1.8 \rightarrow Z = 1.35\sigma$$

PL: test q_0 under BG only ; $f(q_0|H_0)$

$$\hat{\mu} = 0.32 \rightarrow 1.39\sigma$$

$$q_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s+b)}$$

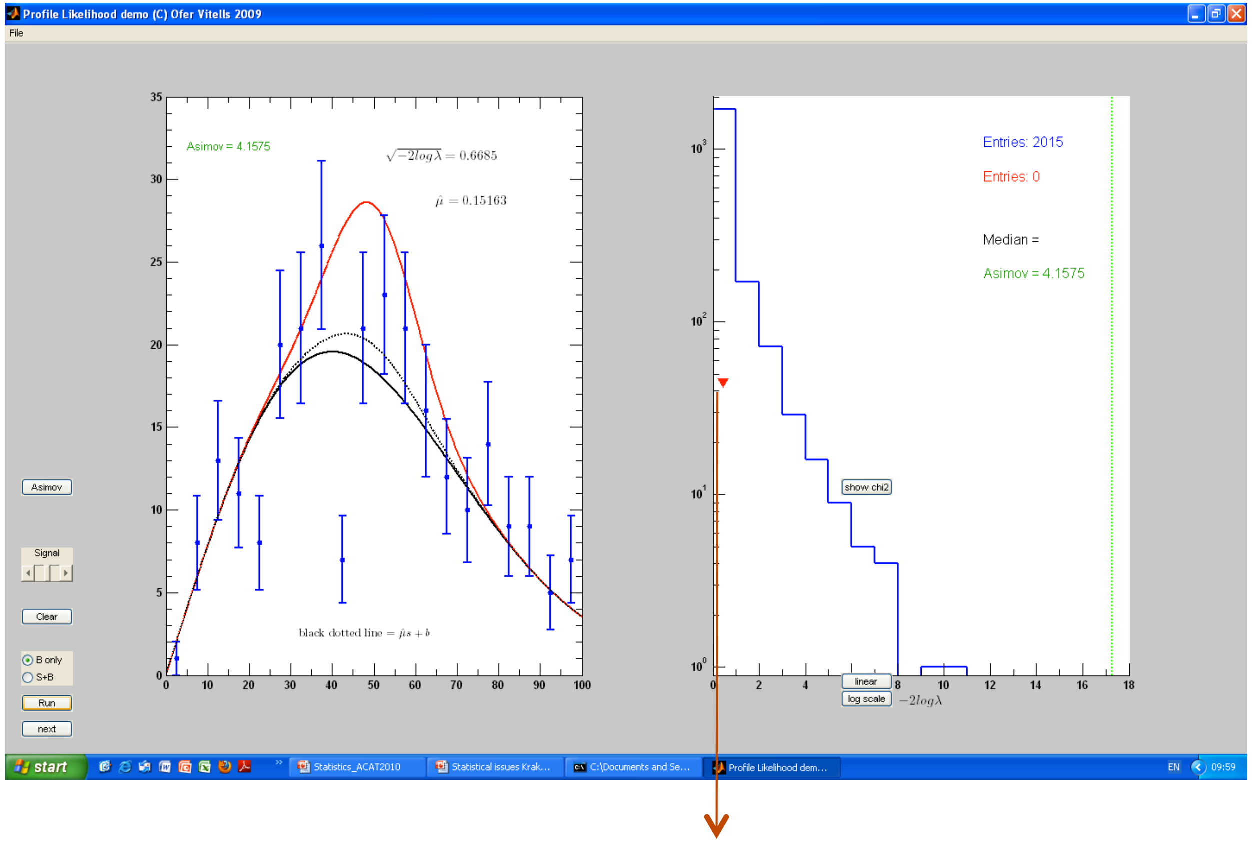


$$q = 1.9 \rightarrow Z = 1.39\sigma$$

PL: test q_0 under BG only ; $f(q_0 | H_0)$

$$\hat{\mu} = 0.15 \rightarrow 0.66\sigma$$

$$q_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s + b)}$$

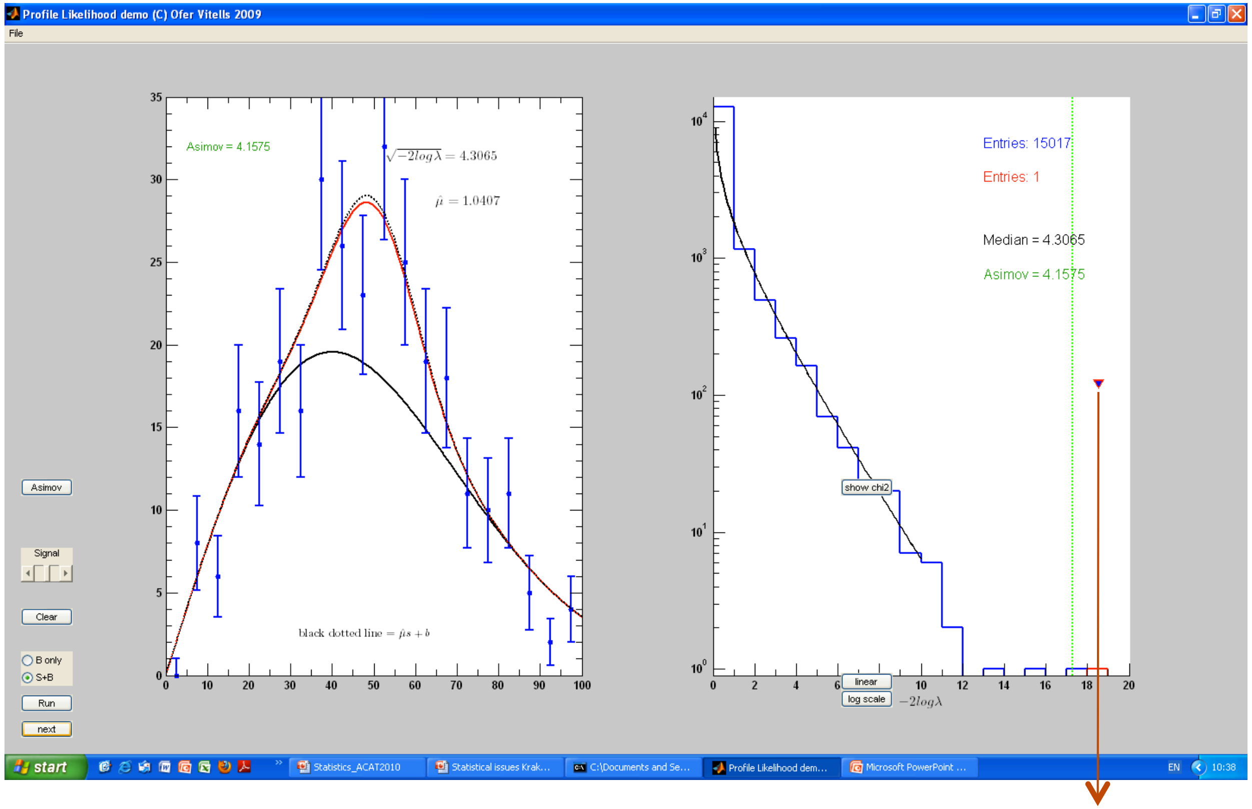


$$q = 0.43 \rightarrow Z = 0.66\sigma$$

The PDF of q_0 under s+b experiments (H_1)

$$q_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s+b)} = -2\ln\frac{L(b\,|\,H_1)}{L(\hat{\mu}s+b\,|\,H_1)}$$

$$\hat{\mu} = 1.04 \rightarrow 4.3\sigma$$

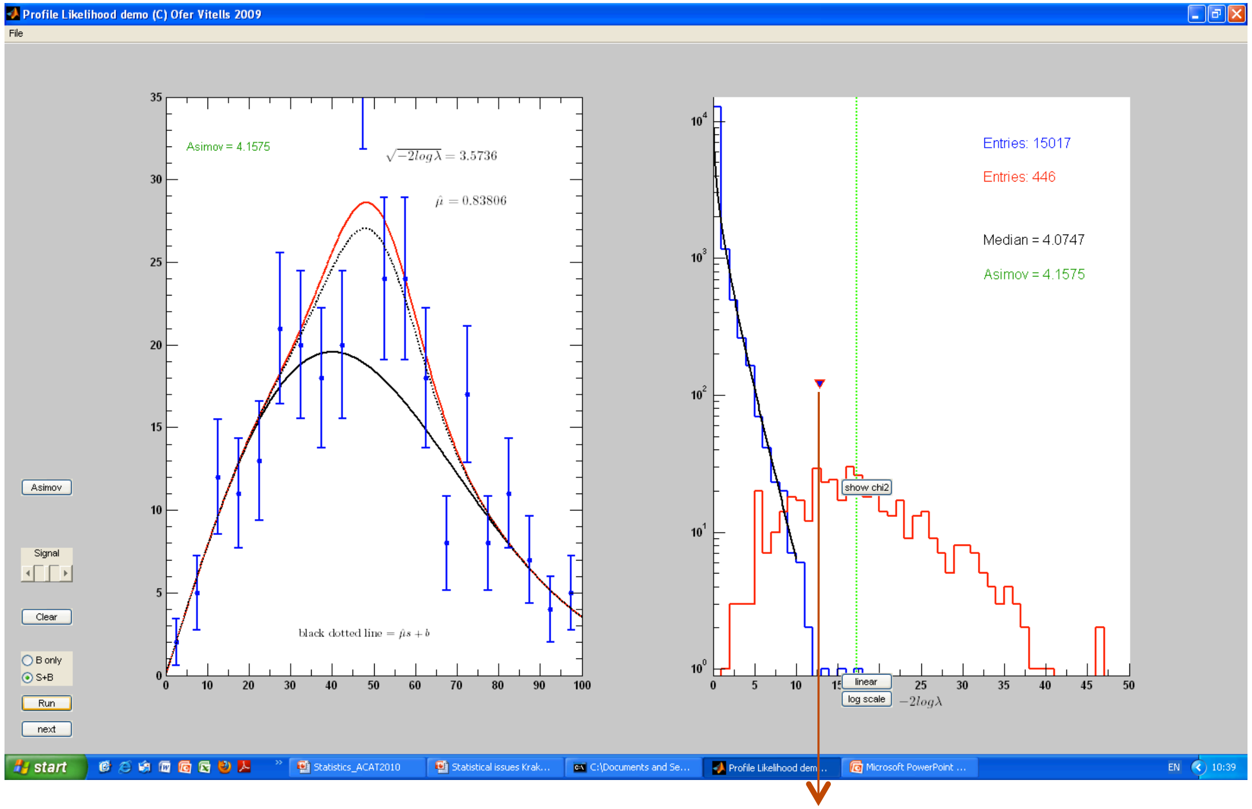


$$q = 18.5 \rightarrow Z = 4.3\sigma$$

The PDF of q_0 under s+b experiments (H_1)

$$q_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s+b)} = -2\ln\frac{L(b\,|\,H_1)}{L(\hat{\mu}s+b\,|\,H_1)}$$

$$\hat{\mu} = 0.83 \rightarrow 3.6\sigma$$

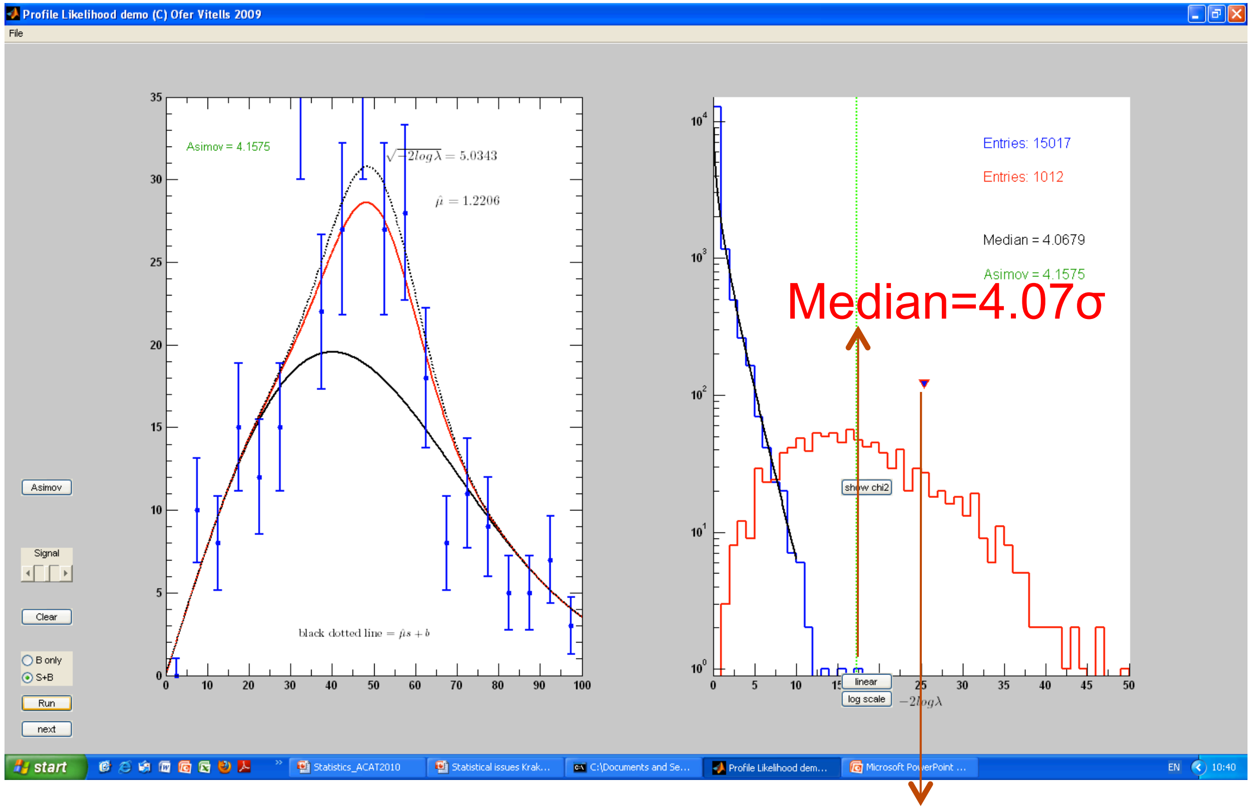


$$q = 12.9 \rightarrow Z = 3.6\sigma$$

Expected Discovery Sensitivity

$$q_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s+b)} = -2\ln\frac{L(b\,|\,H_1)}{L(\hat{\mu}s+b\,|\,H_1)}$$

$$\hat{\mu} = 1.22 \rightarrow 5.0\sigma$$

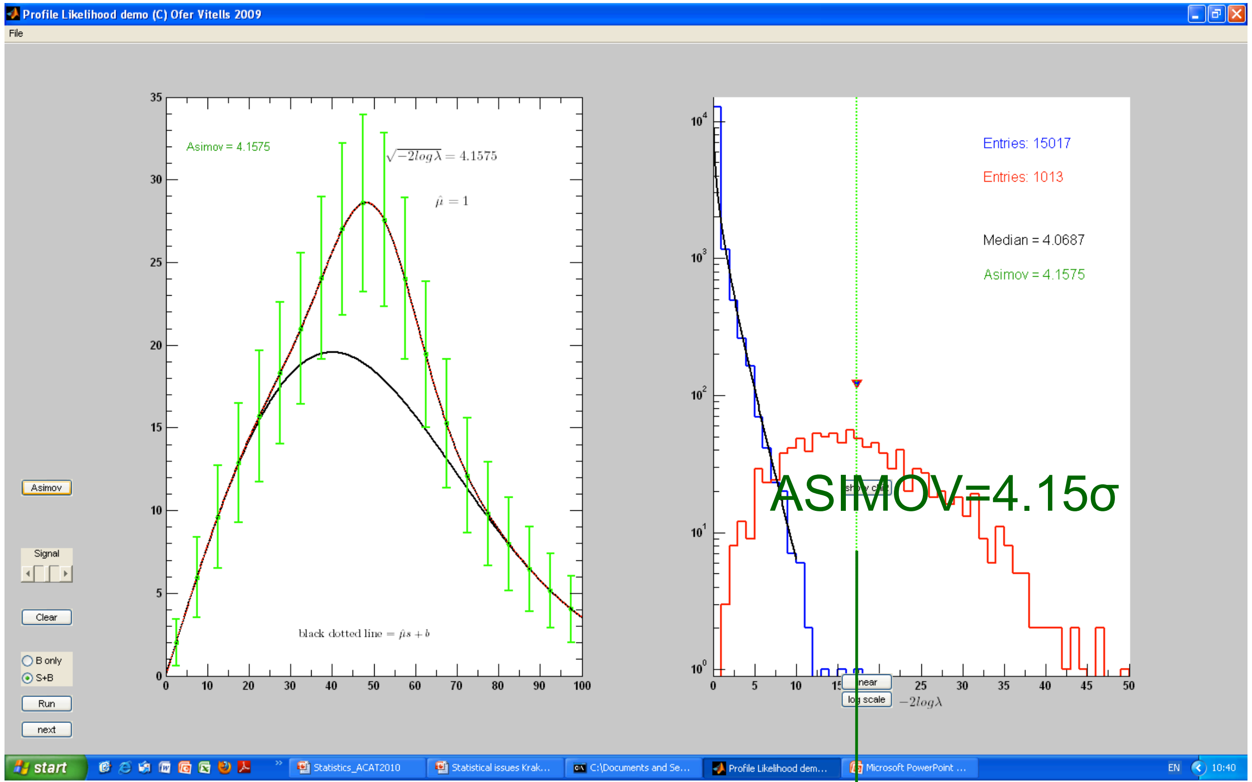


$$q = 25 \rightarrow Z = 5.0\sigma$$

The Median Sensitivity (via ASIMOV)

To estimate the median sensitivity of an experiment (**before looking at the data**), one can either perform lots of s+b experiments and estimate the median $q_{o,med}$ or evaluate q_0 with respect to a representative data set, the ASIMOV data set with $\mu=1$, i.e. x=s+b

$$\hat{\mu} = 1.00 \rightarrow 4.15\sigma$$



$$\boxed{q_A = 17.22 \rightarrow Z_A = 4.15}$$

$$q_{o,med} \simeq q_0(\hat{\mu} = \mu_A = 1) = -2\ln\frac{L(b\,|\,x = x_A = s+b)}{L(\hat{\mu}s + b\,|\,x = x_A = s+b)} = -2\ln\frac{L(b)}{L(1\cdot s+b)}$$

Basic Definition: Signal Strength

- We normally relate the signal strength to its expected Standard Model value, i.e.

$$\mu(m_H) = \frac{\sigma(m_H)}{\sigma_{SM}(m_H)}$$

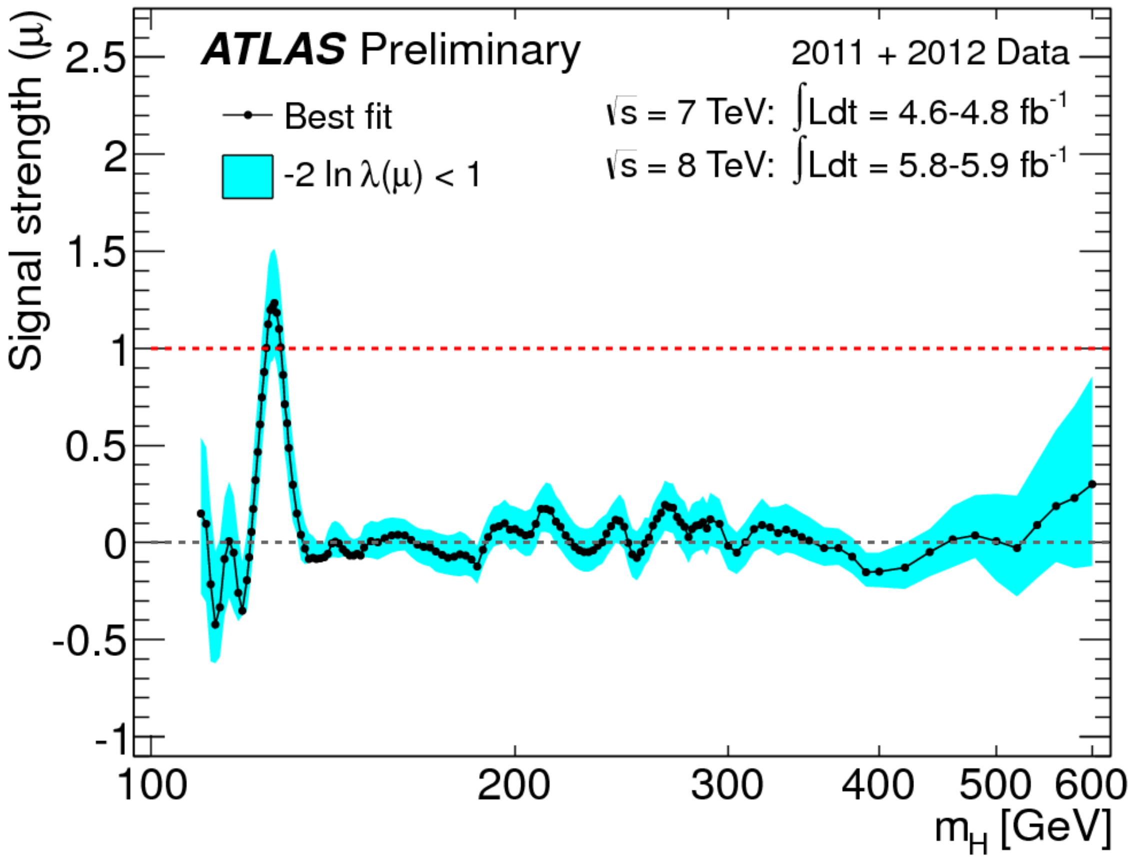
$$\hat{\mu}(m_H) = \text{MLE of } \mu$$

The cyan band plot, what is it?

$$\mu(m_H) = \frac{\sigma(m_H)}{\sigma_{SM}(m_H)}$$

$$\hat{\mu}(m_H) = \text{MLE of } \mu$$

Here we find a possible
 signal which is 1 sigma
away for the SM rcpectation
Of μ=1.

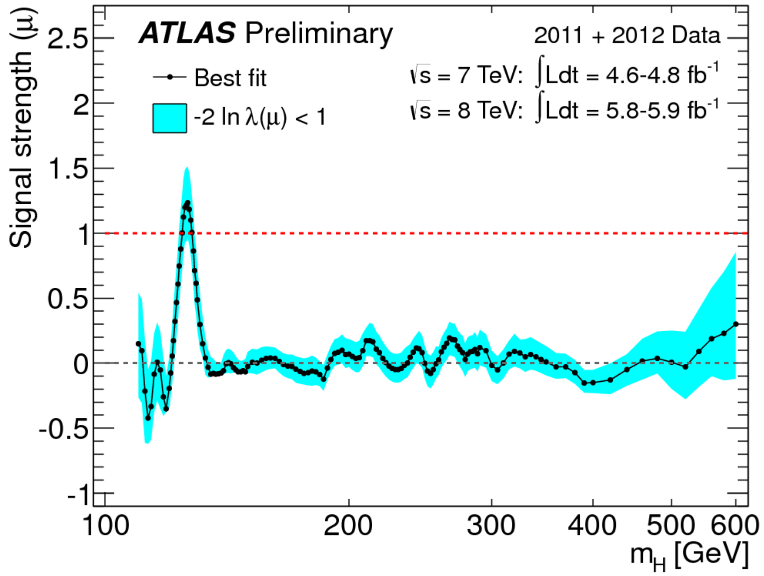


Approximate distribution of the PL ratio

- $-2\ln\lambda(\mu)$ a parabola, with $\hat{\mu}$ being the MLE of μ

$$-2\ln\lambda(\mu) = \frac{(\mu-\hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N}) .$$

$$-2\ln\lambda(\mu) = 1 \rightarrow |\mu - \hat{\mu}| = \sigma$$



$$-2\ln\lambda(0) \;=\; \hat{\mu}^2/\sigma^2.$$

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0 , \\ 0 & \hat{\mu} < 0 , \end{cases}$$

$$q_0 = \begin{cases} \hat{\mu}^2/\sigma^2 & \hat{\mu} \geq 0 , \\ 0 & \hat{\mu} < 0 , \end{cases}$$

$$q_0 = Z^2$$

Nuisance Parameters (Systematics)

- There are two kinds of parameters:
  - Parameters of interest (signal strength… cross section… μ)
  - Nuisance parameters (background cross section, b, signal efficiency)
- The nuisance parameters carry systematic uncertainties
- There are two related issues:
  - Classifying and estimating the systematic uncertainties
  - Implementing them in the analysis
- The physicist must make the difference between cross checks and identifying the sources of the systematic uncertainty.
  - Shifting cuts around and measure the effect on the observable… Very often the observed variation is dominated by the statistical uncertainty in the measurement.

Implementation of Nuisance Parameters

- Implement by marginalizing or profiling

- Marginalization (Integrating) (The C&H Hybrid)
  - Integrate L over possible values of nuisance parameters (weighted by their prior belief functions -- Gaussian,gamma, others…)
  - Consistent Bayesian interpretation of uncertainty on nuisance parameters

- Note that in that sense MC "statistical" uncertainties (like background statistical uncertainty) are systematic uncertainties

Integrating Out The Nuisance Parameters (Marginalization)

$$p(\mu,\theta \mid x)=\frac{L(\mu,\theta)\pi(\mu,\theta)}{\int L(\mu,\theta)\pi(\mu,\theta)d\mu d\theta}=\frac{L(\mu,\theta)\pi(\mu,\theta)}{Normalization}$$

- Our degree of belief in μ is the sum of our degree of belief in μ given θ(nuisance parameter), over "all" possible values of θ
- That's a Bayesian way

$$p(\mu \mid x)=\int p(\mu,\theta \mid x)\pi(\theta)d\theta$$

Nuisance Parameters (Systematisc)

- Neyman Pearson Likelihood Ratio:

$$q^{NP} = -2\ln\frac{L(b)}{L(s+b)}$$

- Either Integrate the Nuisance parameters

$$\boxed{q^{NP}_{Hybrid} = \frac{\int L\left(s+b(\theta)\right)\pi(\theta)\,d\theta}{\int L\left(b(\theta)\right)\pi(\theta)\,d\theta}}$$

prior

R.D. Cousins and V.L. Highland. Incorporating systematic uncertainties into an upper limit. *Nucl. Instrum. Meth.*, A320:331–335, 1992.

- Or profile them

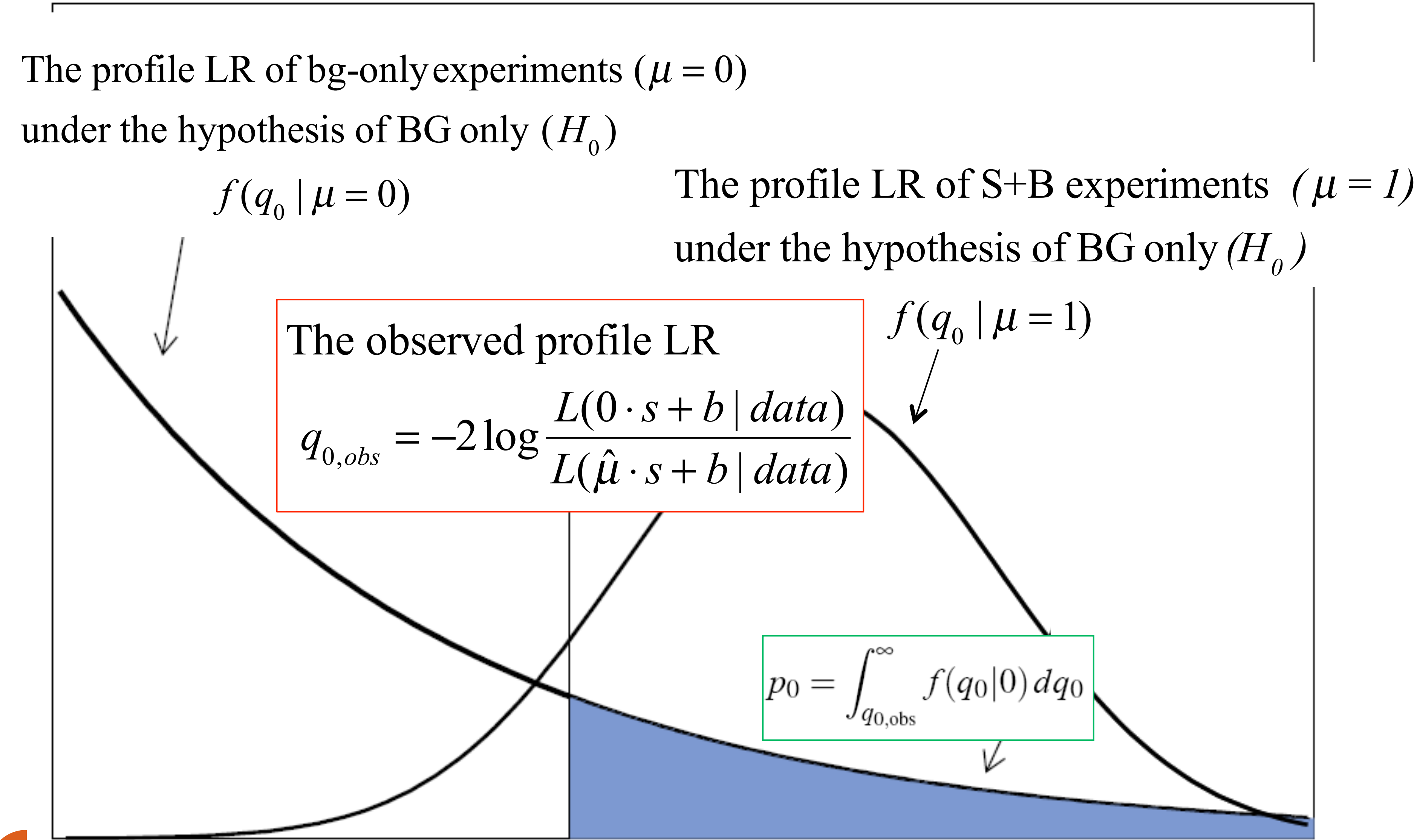
$$q^{NP} = -2\ln\frac{L\left(b(\hat{\hat{\theta}}_b)\right)}{L\left(s+b(\hat{\hat{\theta}}_{s+b})\right)}$$

$$\hat{\hat{\theta}}_b = MLE\ of\ L(b(\theta))$$

$$\hat{\hat{\theta}}_{s+b} = MLE\ of\ L(s+b(\theta))$$

Discovery - Illustrated

$$\lambda(\mu=0)=\frac{L(0\cdot s+b\,|\,data)}{L(\hat{\mu}\cdot s+b\,|\,data)},\quad q_0=-2\log\lambda(\mu=0)$$

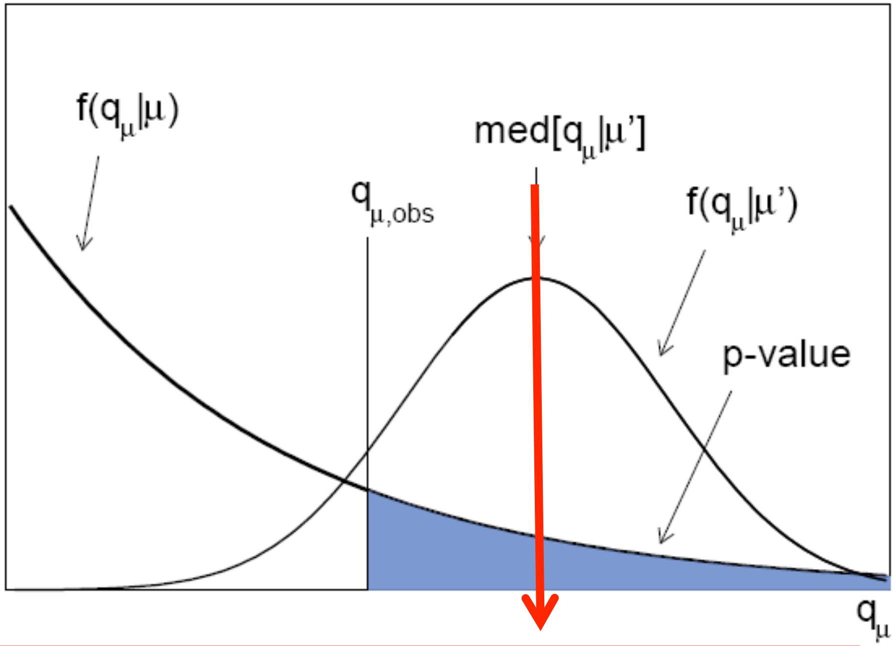


p_0 is the level of compatibility between the data and the no-Higgs hypothesis

If p_0 is smaller than ~$2.8\cdot10^{-7}$ we claim a 5s discovery

Median Sensitivity

- To estimate the median sensitivity of an experiment (before looking at the data), one can either perform lots o s+b experiments and estimate the median $q_{0,med}$ or evaluate q_0 with respect to a representative data set, the ASIMOV data set with $\mu=1$, i.e. n=s+b



$$\boxed{Z_{\text{med}} = \Phi^{-1}(1 - p_{0_{\text{med}}}) = \Phi^{-1}(1 - p_0(q_{0_{\text{med}}}))}$$

$$Z_{med} = \sqrt{-2\ln\lambda_A(0)}$$

$$\lambda_A(0) = \frac{L(\mu = 0 \mid ASIMOV\ data = s + b)}{L(\hat{\mu}_A = 1 \mid ASIMOV\ data = s + b)}$$

p0 and the expected p0

$$p_0 = \int_{q_{0,\mathrm{obs}}}^{\infty} f(q_0|0)\,dq_0$$

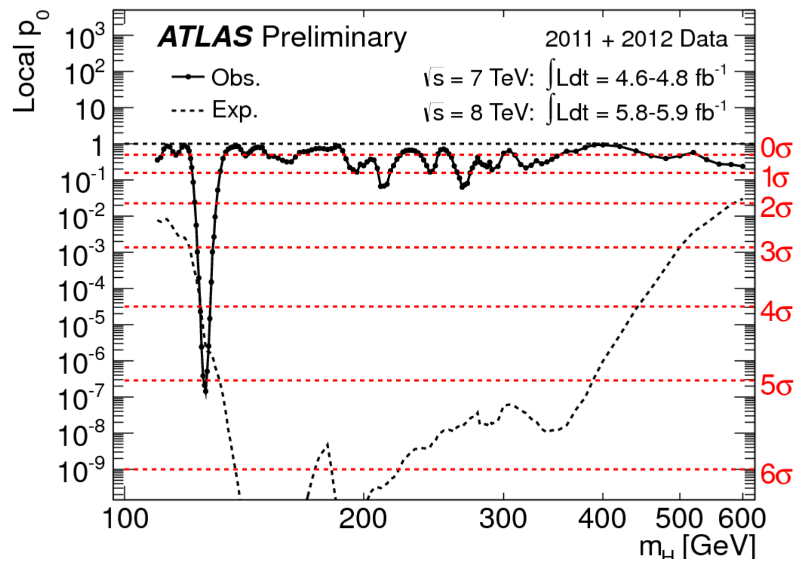
p_0 is the probability to observe a less BG like result (more signal like) than the observed one

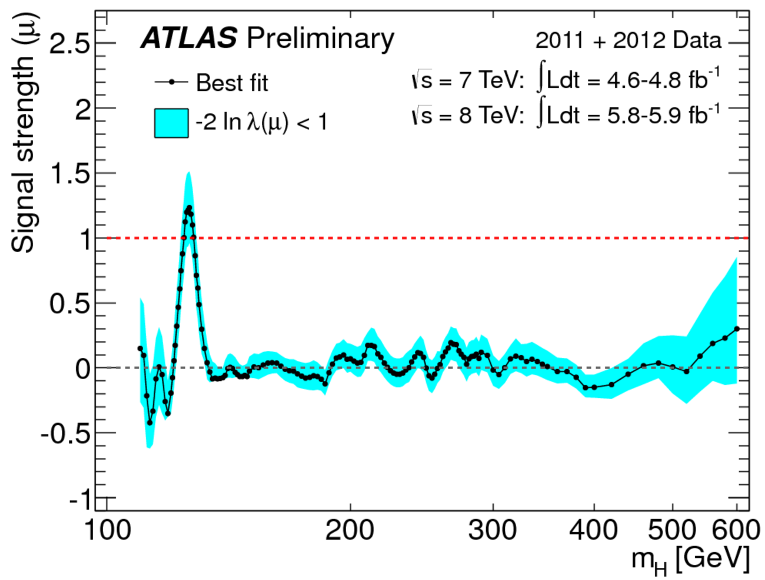
Small p0 leads to an observation

A tiny p0 leads to a discovery

$$p = \int_Z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = 1 - \Phi(Z)$$

$$Z = \Phi^{-1}(1-p)$$

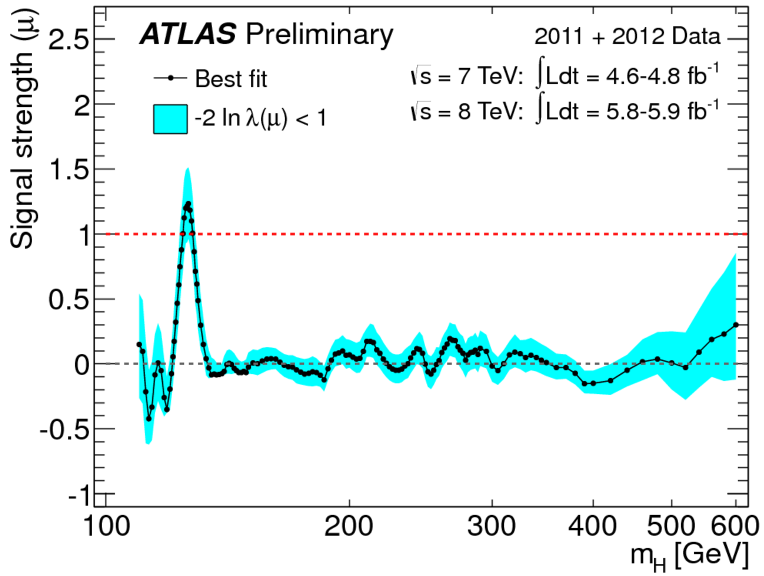




Introducing the Heartbeat

Having a normal heartbeat is an important indication of a healthy lifestyle.
"*Life is one of those precious fleeting gifts, and everything can change in a heartbeat.*"

Having a (normal) scalar is an important indication of a healthy model
"Mass is one of those precious gifts and everything can change in the absence of a scalar "



Physics Complicates Things

- A negative signal is not Physical
- Downward fluctuations of the background do not serve as an evidence against the background
- Upward fluctuations of the signal do not serve as an evidence against the signal

Discovery

- Test statistics

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0\,, \\ 0 & \hat{\mu} < 0\,, \end{cases}$$

Background downward fluctuations do not
serve as an evidence against the background hypothesis

Discovery

- Test statistics

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0\,, \\ 0 & \hat{\mu} < 0\,, \end{cases}$$

Background downward fluctuations do not
serve as an evidence against the background hypothesis

Distribution of q0 (discovery)

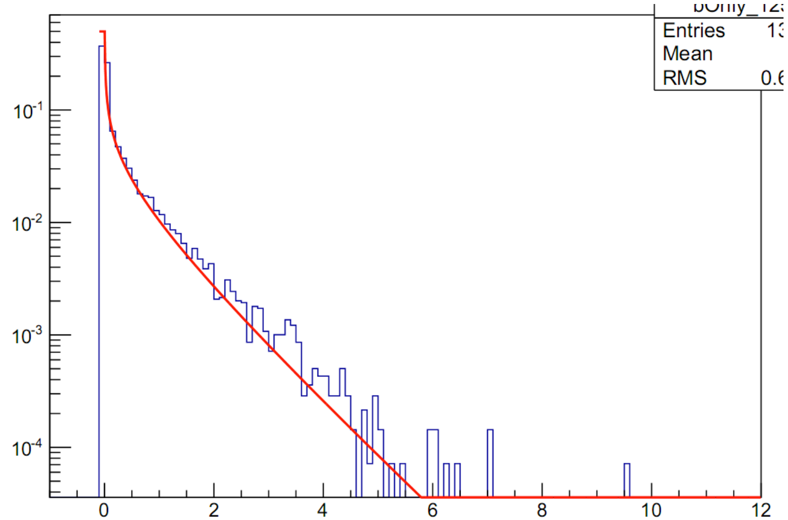
- We find

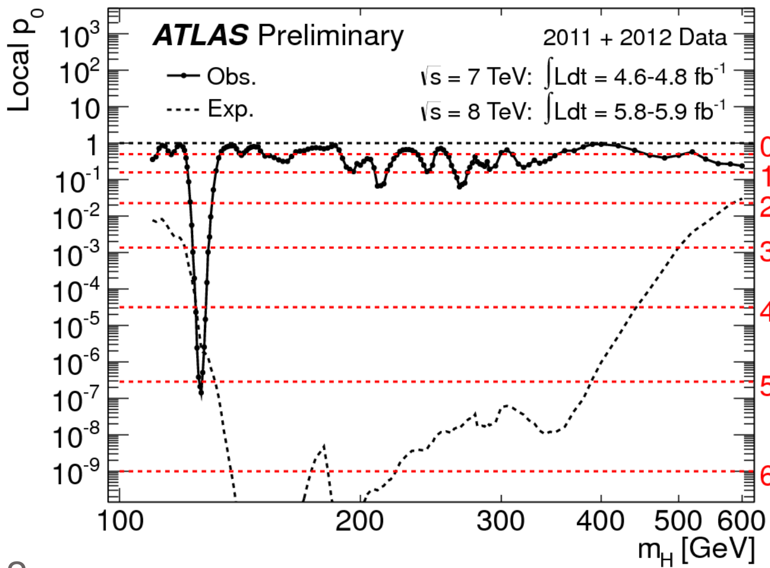
$$p_0 = \int_{q_{0,\mathrm{obs}}}^{\infty} f(q_0|0)\, dq_0 \,.$$

$$f(q_0|0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}e^{-q_0/2} \,.$$

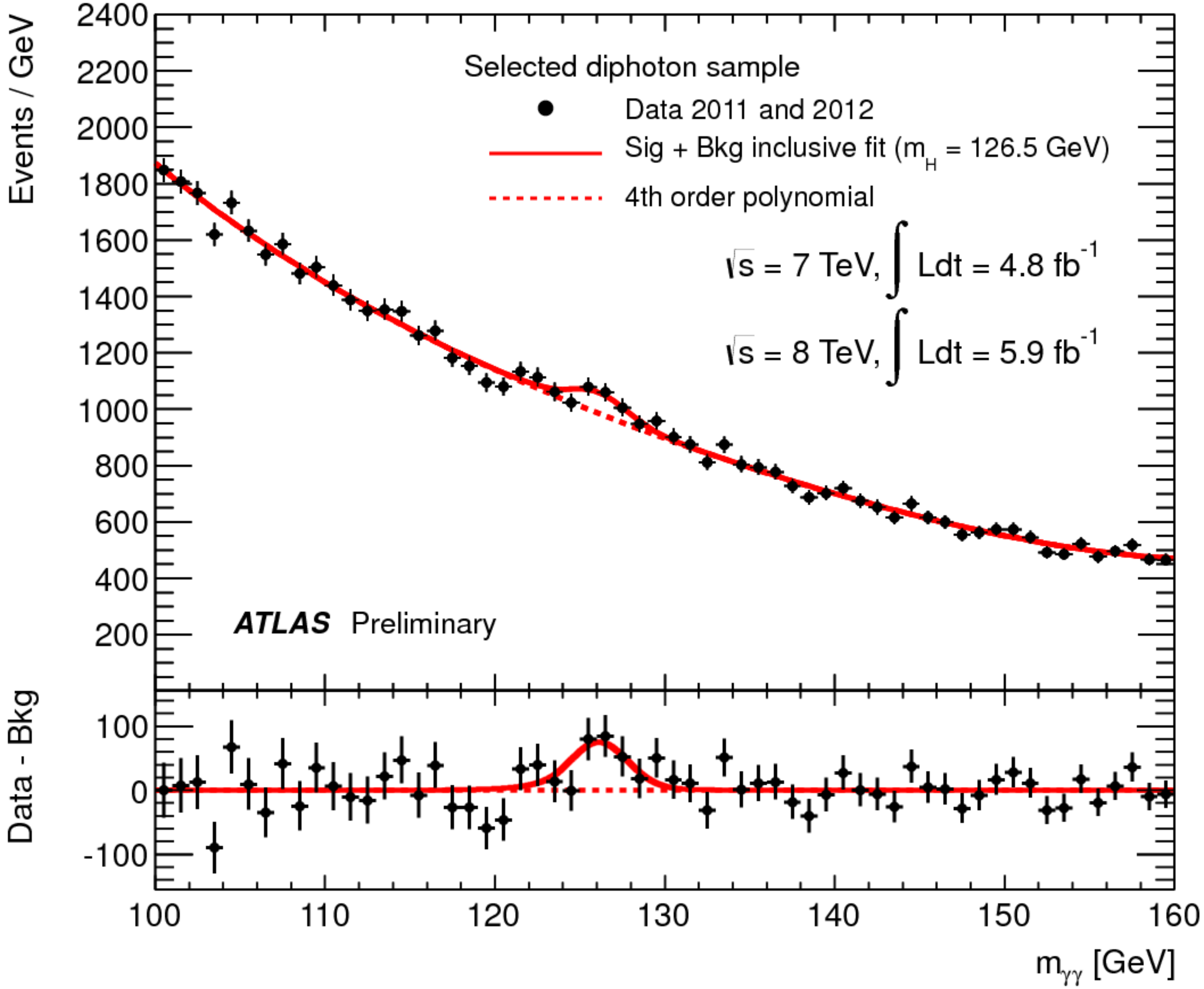
$$Z_0 = \Phi^{-1}(1 - p_0) = \sqrt{q_0} \,.$$

- q_0 distribute as half a delta function at zero and half a chi squared. $q_{0,obs} = q_{0,obs}(m_H) \to p_0 = p_0(m_H)$

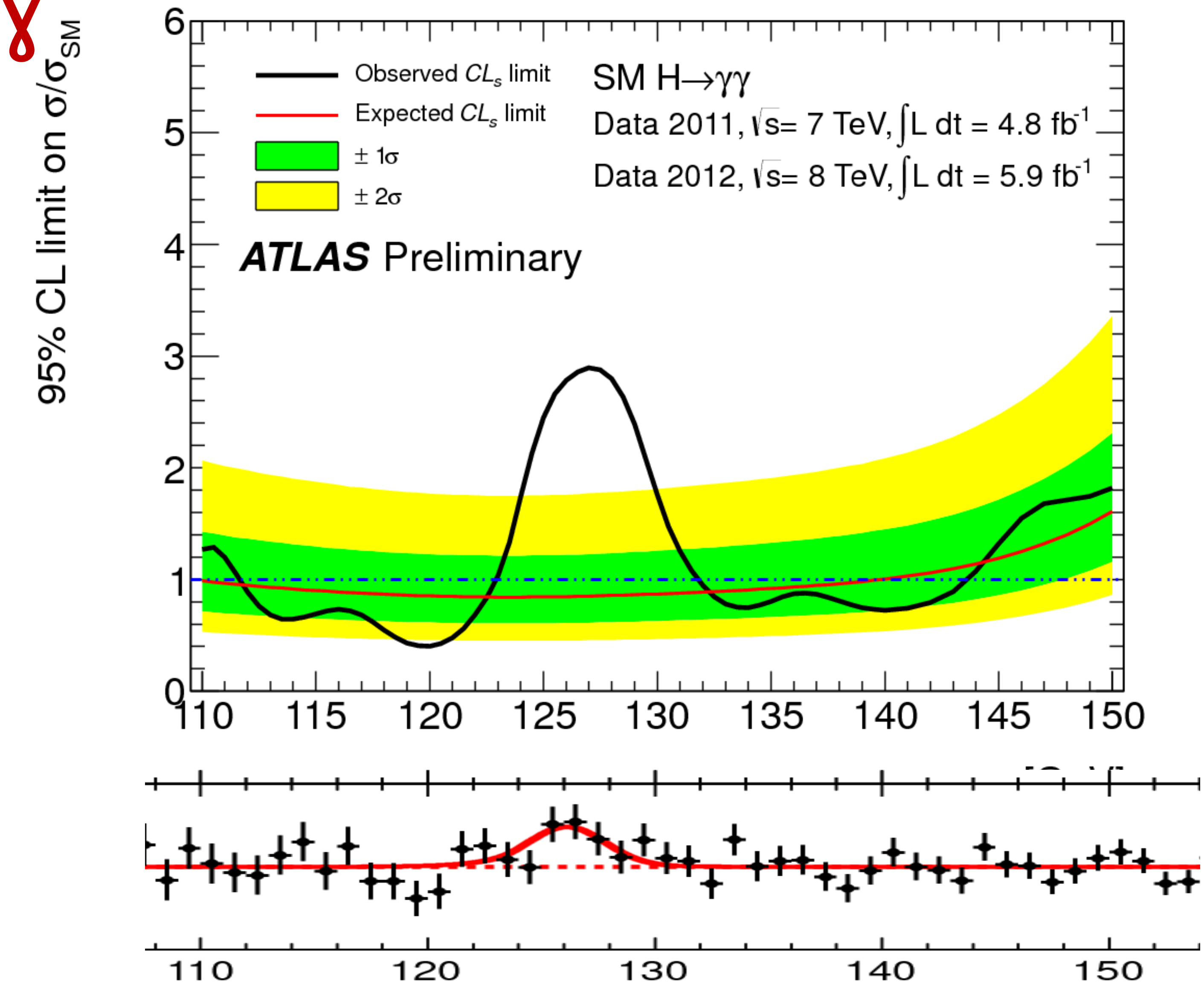




Example: H→γγ

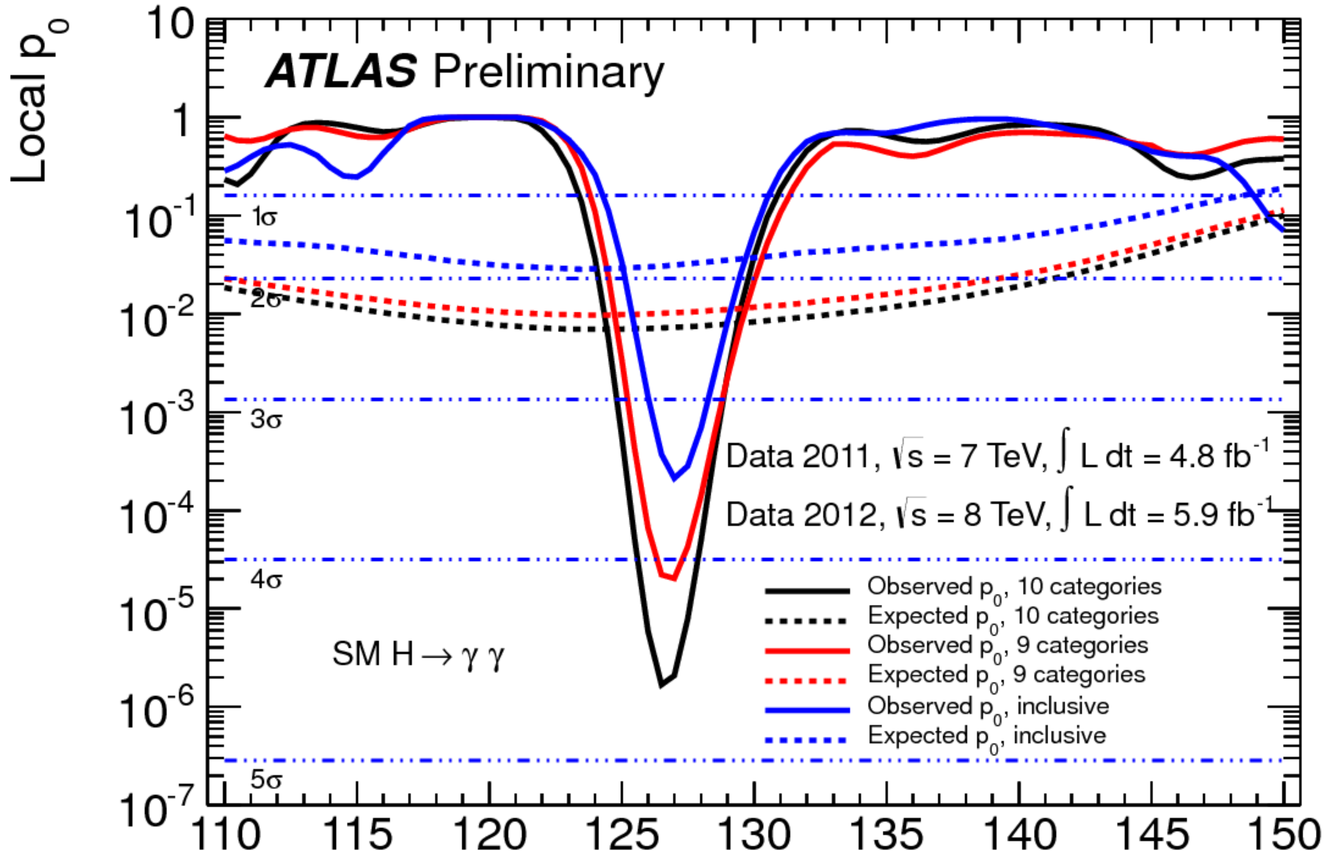

Selected diphoton sample
Data 2011 and 2012
Sig + Bkg inclusive fit (m_H = 126.5 GeV)
4th order polynomial
$\sqrt{s}$ = 7 TeV, $\int$ Ldt = 4.8 fb^{-1}
$\sqrt{s}$ = 8 TeV, $\int$ Ldt = 5.9 fb^{-1}
ATLAS Preliminary
Events / GeV
Data - Bkg
$m_{\gamma\gamma}$ [GeV]
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
100
0
-100
100
110
120
130
140
150
160

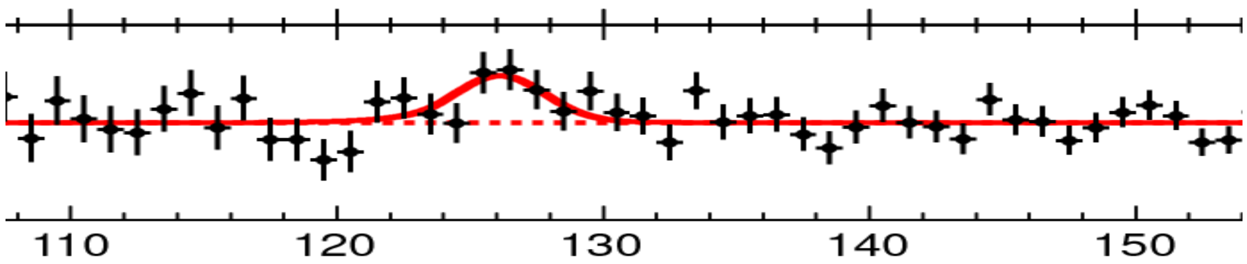
H→γγ



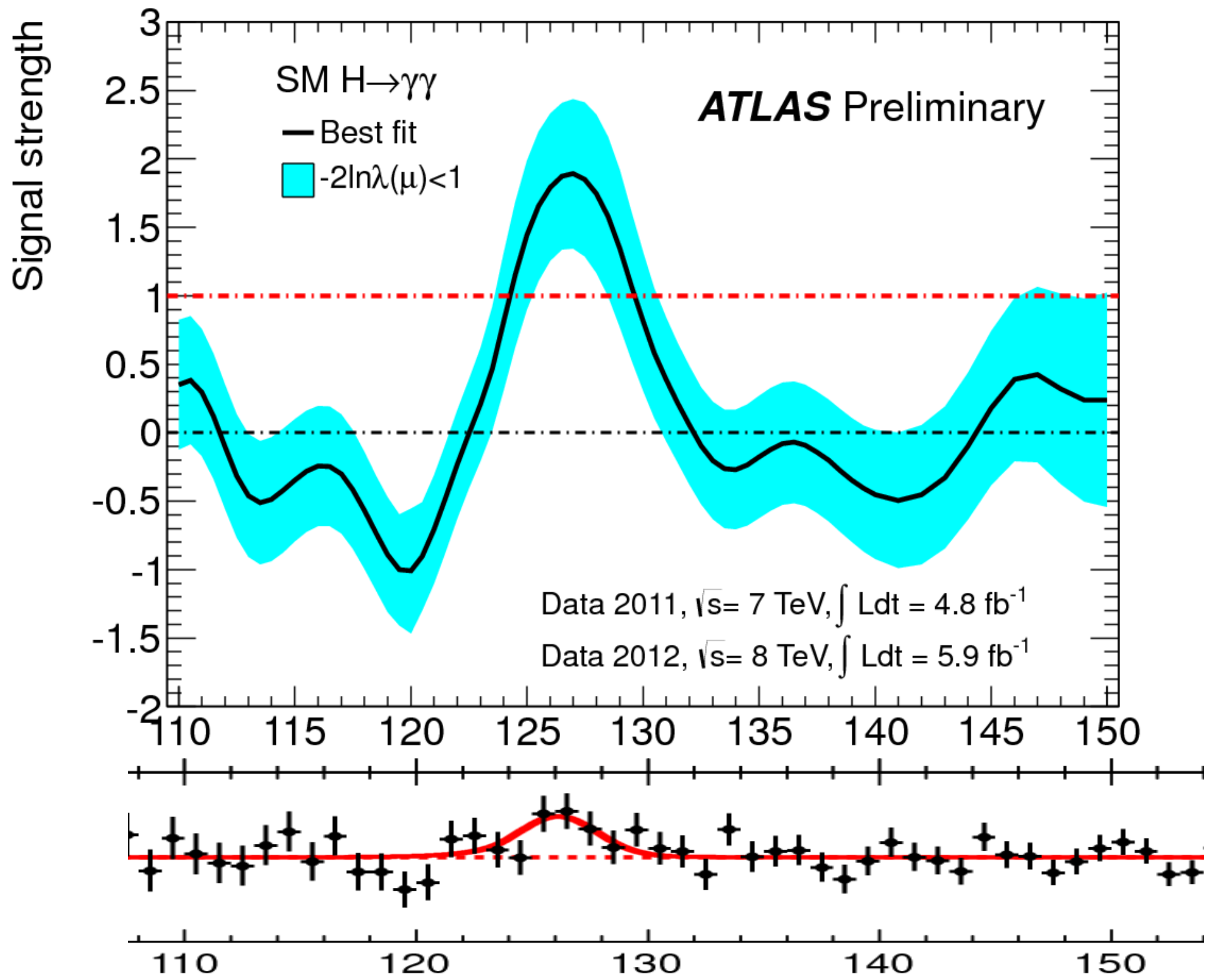
95% CL limit on σ/σ_SM
Observed CL_s limit
Expected CL_s limit
± 1σ
± 2σ
SM H→γγ
Data 2011, √s= 7 TeV, ∫L dt = 4.8 fb⁻¹
Data 2012, √s= 8 TeV, ∫L dt = 5.9 fb⁻¹
ATLAS Preliminary
0
1
2
3
4
5
6
110
115
120
125
130
135
140
145
150
110
120
130
140
150

H→γγ


ATLAS Preliminary
Local p_0
Data 2011, $\sqrt{s}$ = 7 TeV, $\int L\,dt$ = 4.8 fb^{-1}
Data 2012, $\sqrt{s}$ = 8 TeV, $\int L\,dt$ = 5.9 fb^{-1}
SM H → γ γ
Observed p_0, 10 categories
Expected p_0, 10 categories
Observed p_0, 9 categories
Expected p_0, 9 categories
Observed p_0, inclusive
Expected p_0, inclusive
1σ
2σ
3σ
4σ
5σ
10
1
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
110 115 120 125 130 135 140 145 150

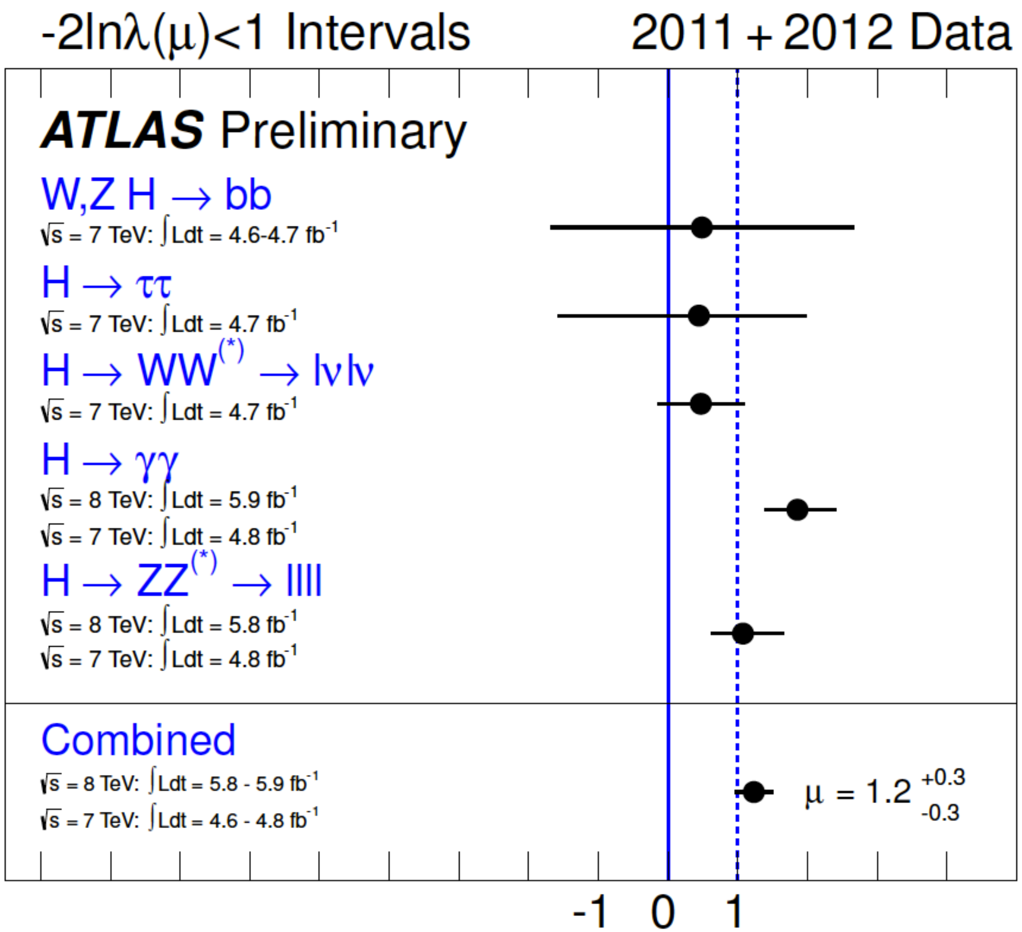

110 120 130 140 150

H→γγ



Signal strength
SM H→γγ
Best fit
-2lnλ(μ)<1
ATLAS Preliminary
Data 2011, √s= 7 TeV, ∫ Ldt = 4.8 fb⁻¹
Data 2012, √s= 8 TeV, ∫ Ldt = 5.9 fb⁻¹
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
110
115
120
125
130
135
140
145
150
110
120
130
140
150

From the signal strength MLE plot one gets



Towards Measurements of the Higgs Boson Properties

To establish the signal we first want to measure the resonance mass and its cross section, next measure its spin and CP.

- In order to address the values of the signal strength and mass of a potential signal that are simultaneously consistent with the data, the following profile likelihood ratio is used:

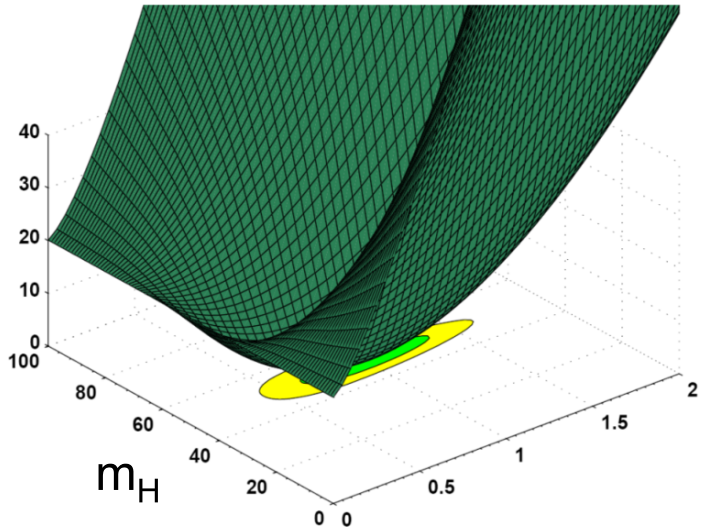
$$\lambda(\mu, m_H) = \frac{L(\mu, m_H, \hat{\hat{\theta}}(\mu, m_H))}{L(\hat{\mu}, \hat{m}_H, \hat{\theta})}$$

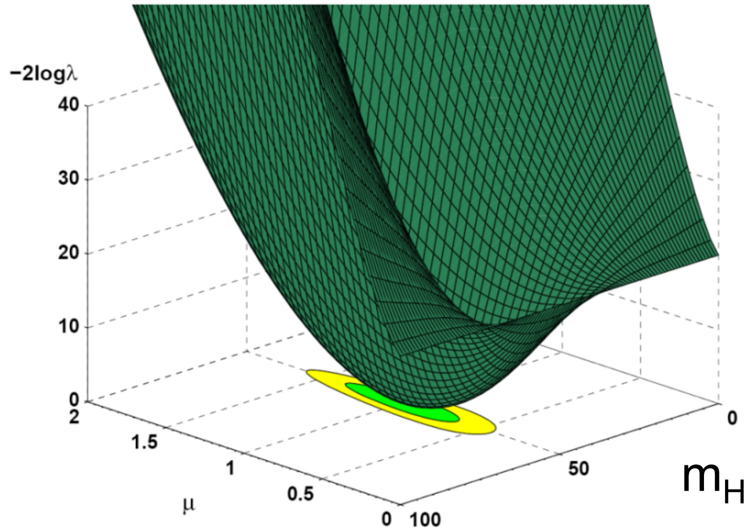
- In the presence of a strong signal, this test statistic will produce closed contours about the best fit point $(\hat{\mu}, \hat{m}_H)$;
- The 2D LR behaves asymptotically as a Chis squared with 2 DOF (Wilks' theorem) so the derivation of 68% and 95% CL cintours is easy

Measuring the signal strength and mass

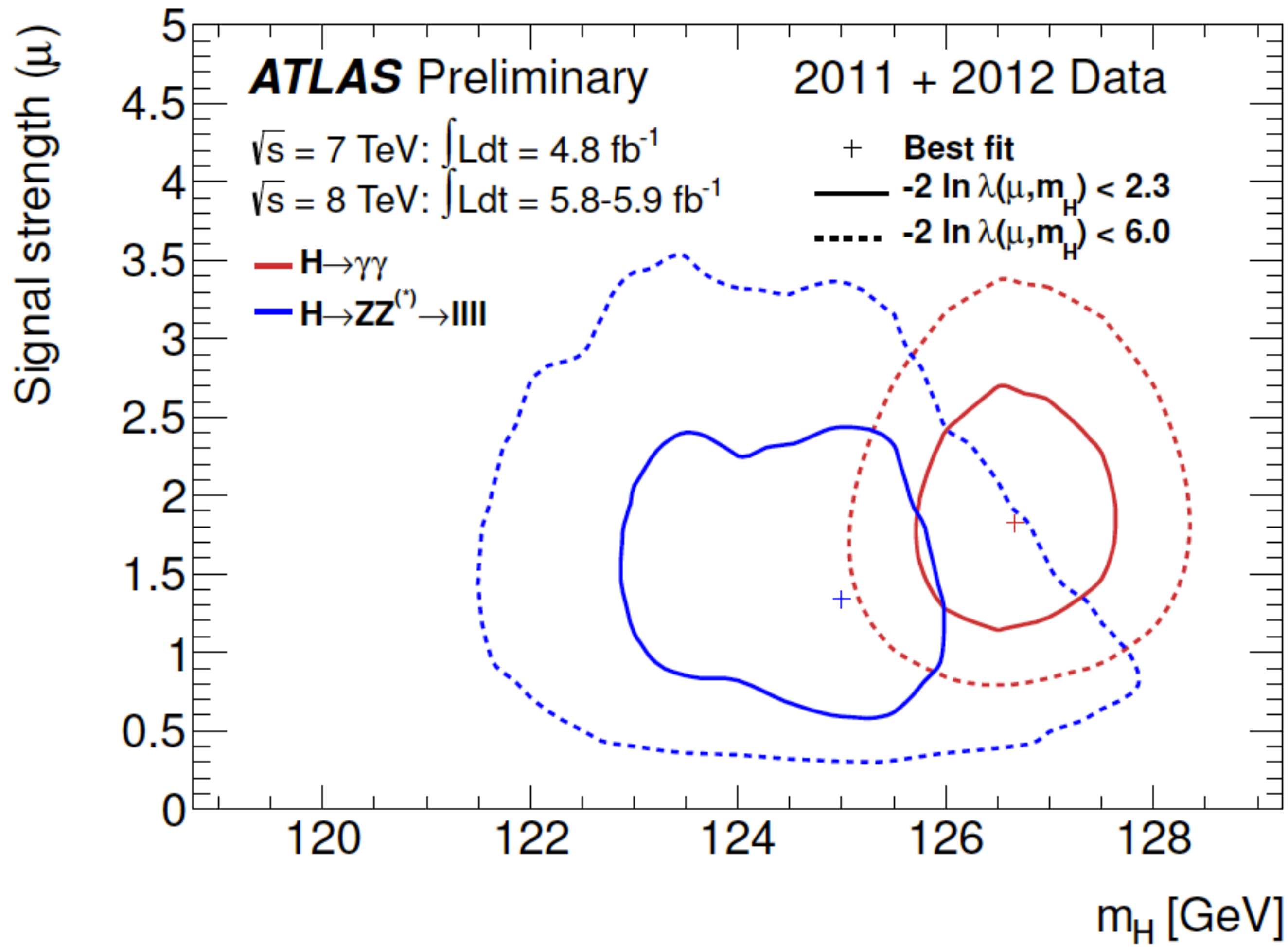
2 parameters of interest: the signal strength μ and the Higgs mass m_H

$$q(\mu, m_H) = -2ln\lambda(\mu, m_H) = -2\ln\frac{L(\mu, m_H, \hat{\hat{b}})}{L(\hat{\mu}, \hat{m}_H, \hat{b})}$$

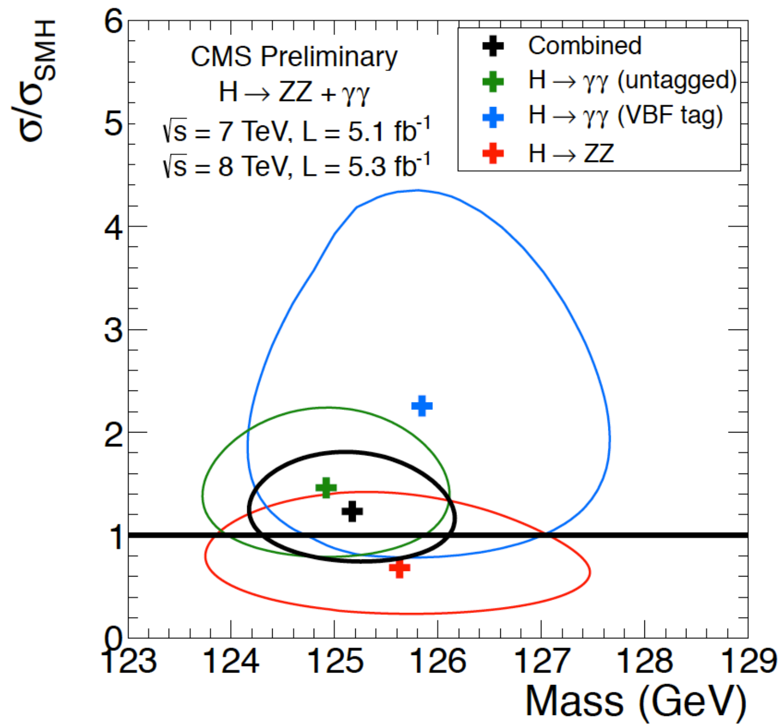


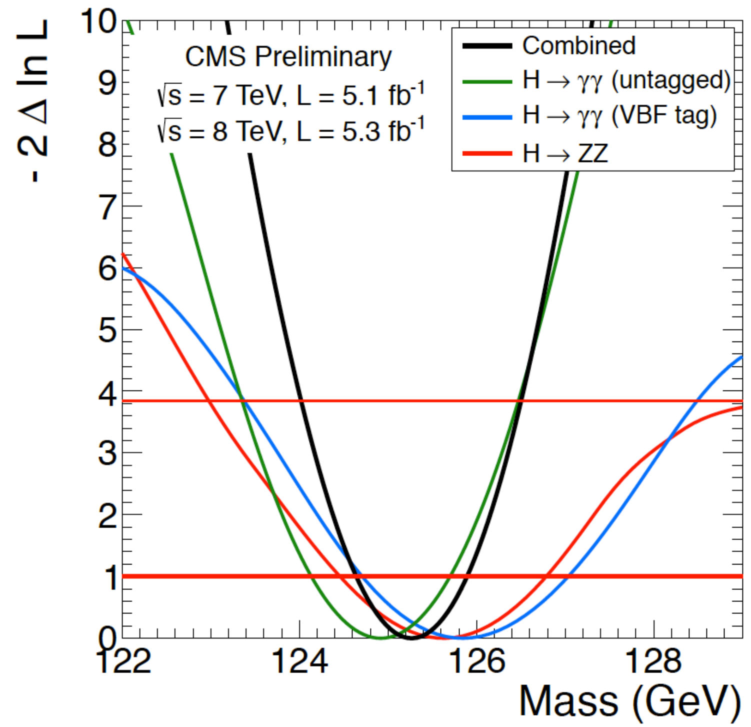


Mass measurement



More of the same





The Look Elsewhere Effect

Look Elsewhere Effect

- To establish a discovery we try to reject the background only hypothesis H_0 against the alternate hypothesis H_1
- H_1 could be
  - A Higgs Boson with a specified mass m_H
  - A Higgs Boson at some mass m_H in the search mass range
- The look elsewhere effect deals with the floating mass case

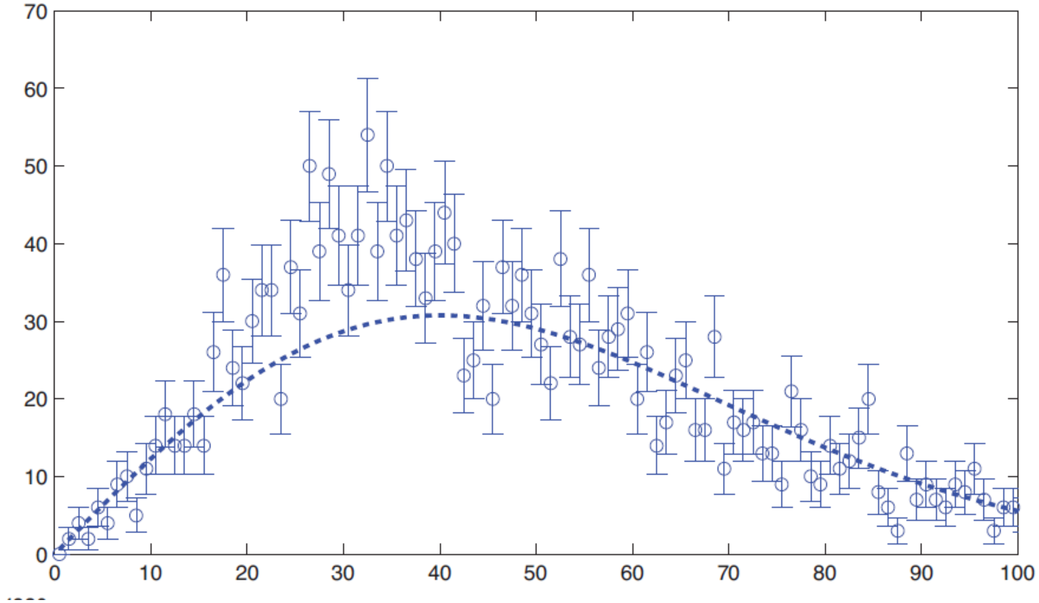
Let the Higgs mass, m_H, and the
signal strength μ
be 2 parameters of interest

$$\lambda(\mu, m_H) = \frac{L(\mu, m_H, \hat{\hat{b}})}{L(\hat{\mu}, \hat{m}_H, \hat{b})}$$

The problem is that m_H is not defined under the null H_0 hypothesis

Look Elsewhere Effect

Is there a signal here?



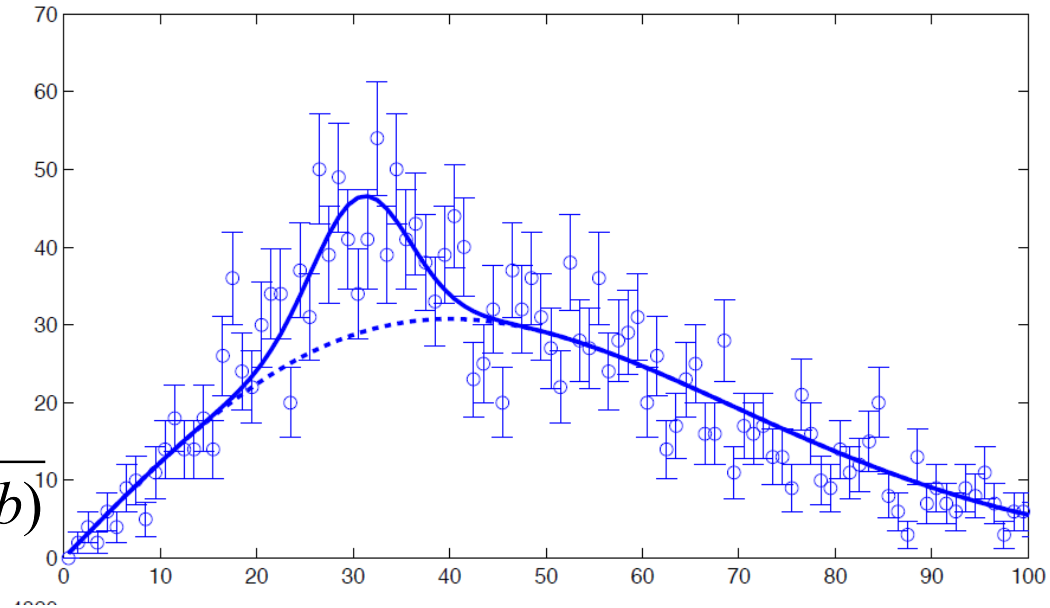
Look Elsewhere Effect

Obviously
@ m=30

What is its significance?

What is your test statistic?

$$q_{fix,obs} = -2\ln\frac{L(b)}{L(\hat{\mu}s(m=30)+b)}$$



Look Elsewhere Effect

Test statistic

$$q_{fix,obs} = -2\ln\frac{L(b)}{L(\hat{\mu}s(m=30)+b)}$$

What is the p-value?

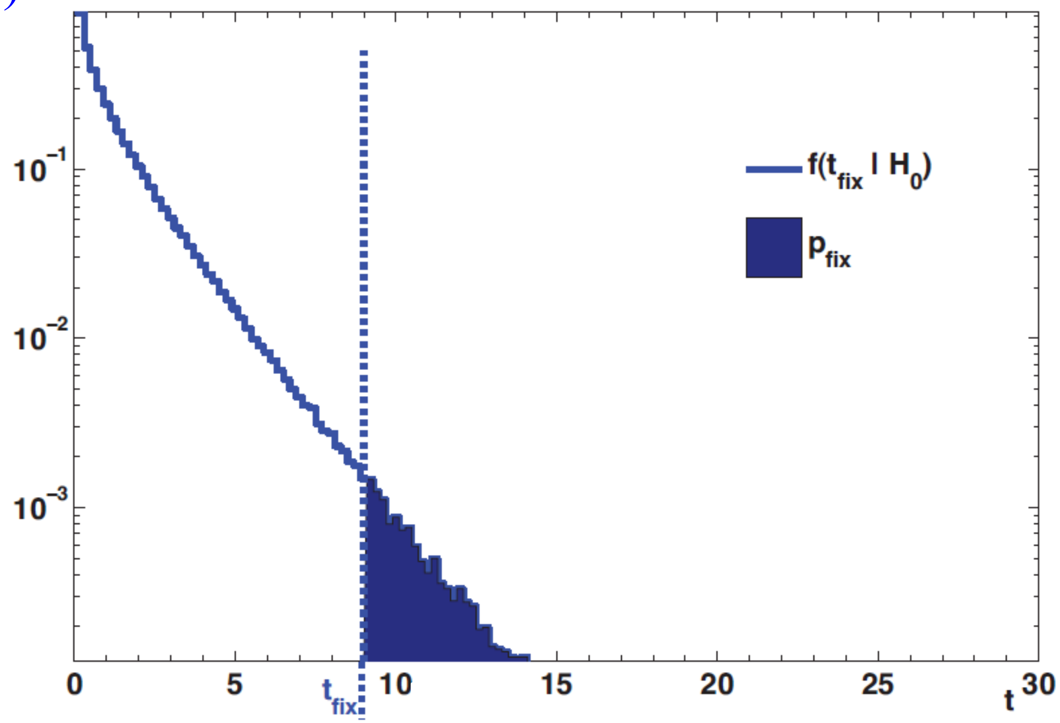
generate the PDF

$$f(q_{fix} \mid H_0)$$

and find the **p-value**

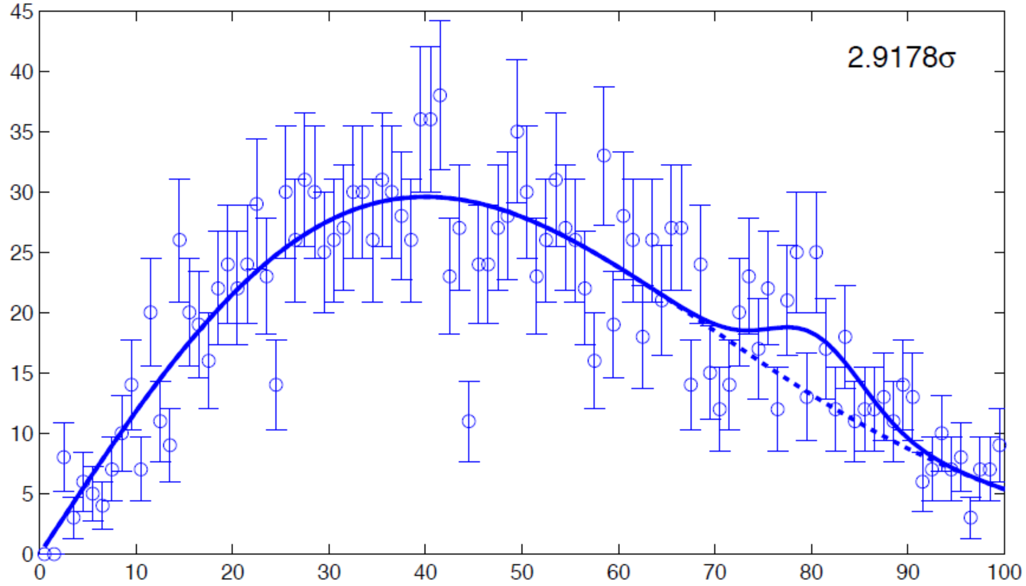
Wilks theorem:

$$f(q_{fix} \mid H_0) \sim \chi_1^2$$



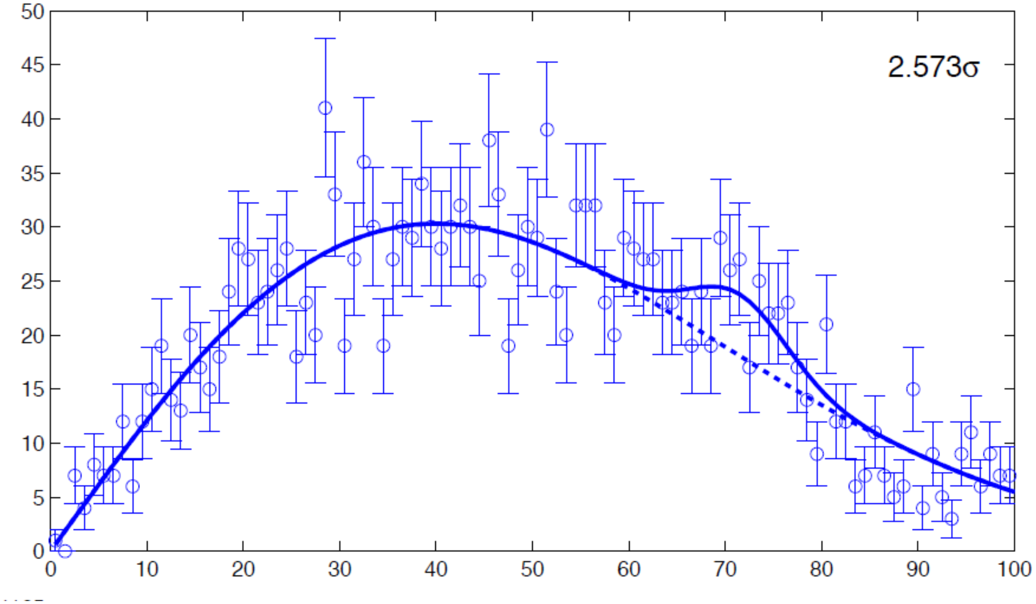
Look Elsewhere Effect

Would you ignore this signal, had you seen it?



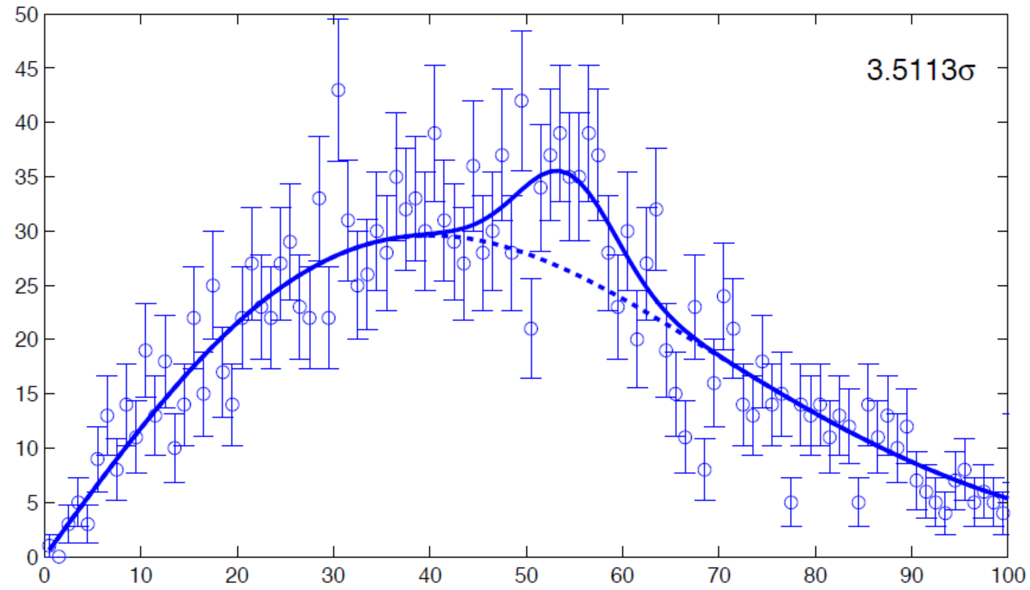
Look Elsewhere Effect

Or this?



Look Elsewhere Effect

Or this?

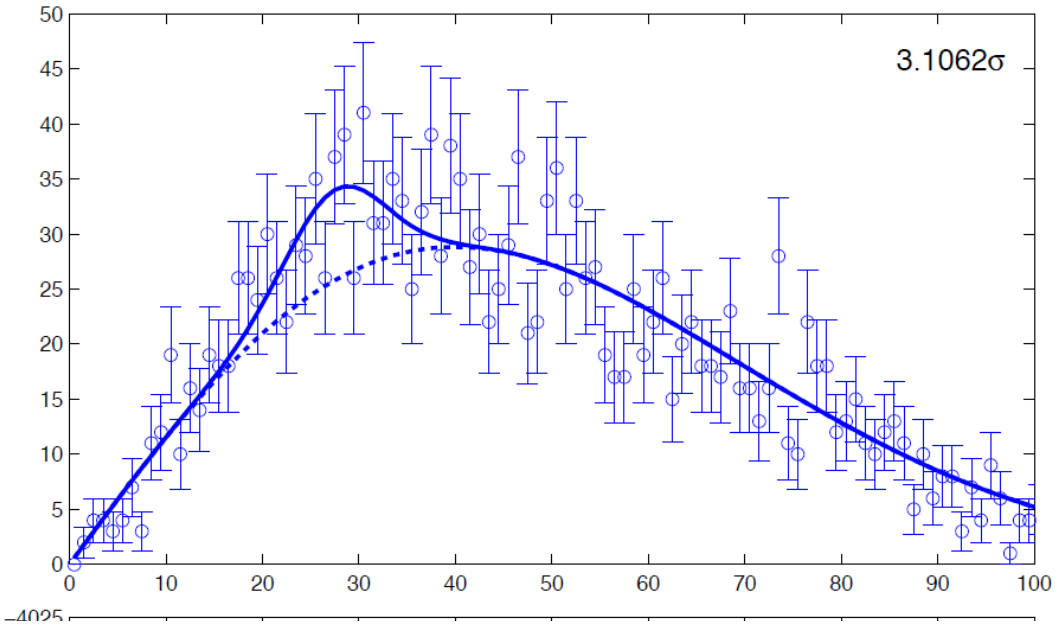


Look Elsewhere Effect

Or this?

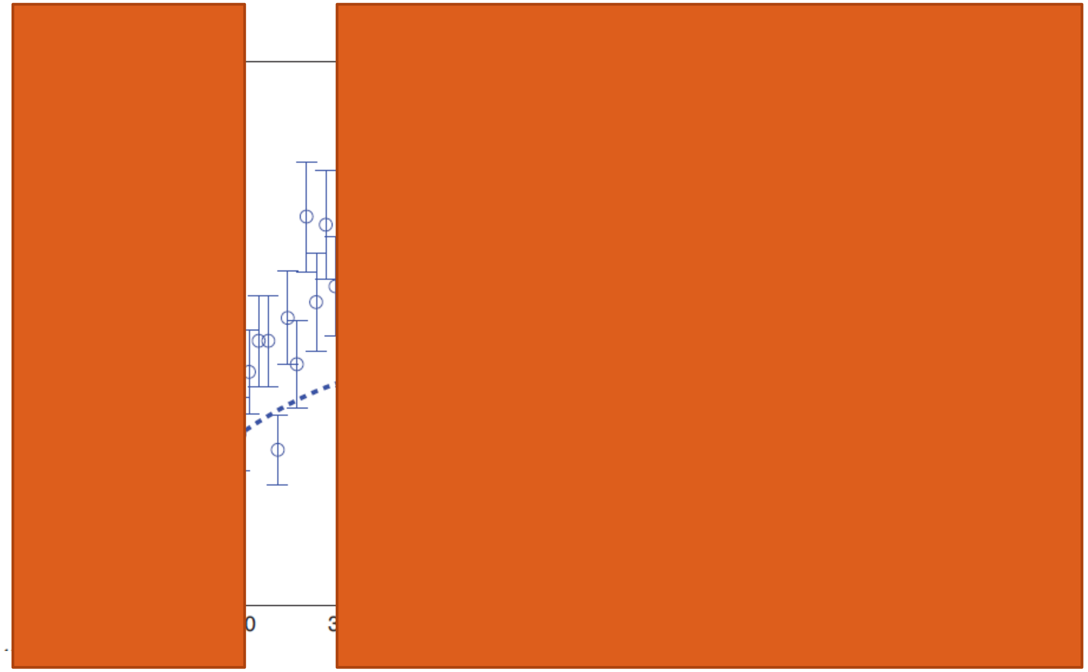
Obviously NOT!

ALL THESE
"SIGNALS" ARE
BG
FLUCTUATIONS



Look Elsewhere Effect

- Having no idea where the signal might be there are two options
- OPTION I: scan the mass range in pre-defined steps and test any disturbing fluctuations

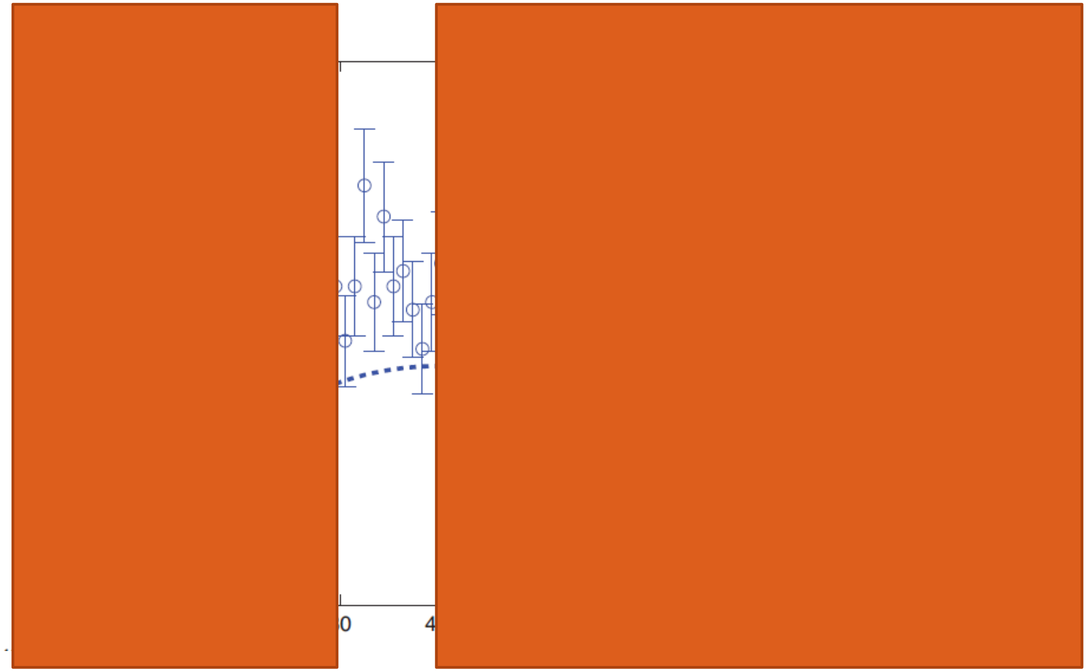


$$q_{fix,obs}(\hat{\mu}) = -2\ln\frac{L(b)}{L(\hat{\mu}s(m)+b)}$$

- Perform a fixed mass analysis at

Look Elsewhere Effect

- Having no idea where the signal might be there are two options
- OPTION I: scan the mass range in pre-defined steps and test any disturbing fluctuations

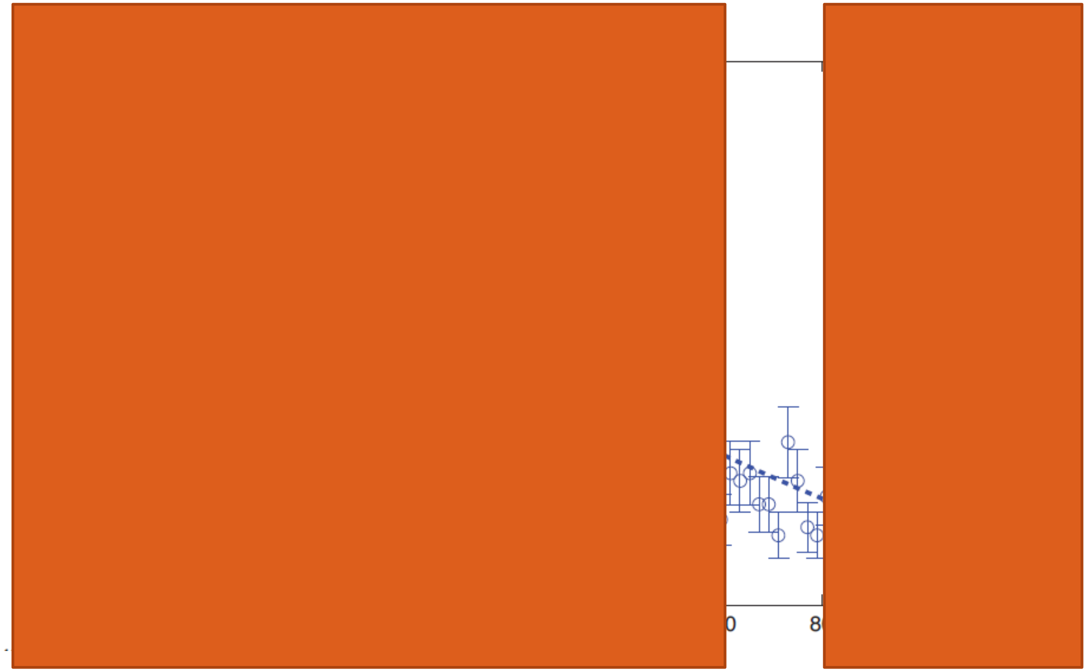


$$q_{fix,obs}(\hat{\mu}) = -2\ln\frac{L(b)}{L(\hat{\mu}s(m)+b)}$$

- Perform a fixed mass analysis at

Look Elsewhere Effect

The scan resolution must be less than the signal mass resolution



Assuming the signal can be only at one place, pick the one with the smallest p-value (maximum significance)

$$q_{fix,obs}(\hat{\mu}) = -2\ln\frac{L(b)}{L(\hat{\mu}s(m)+b)}$$

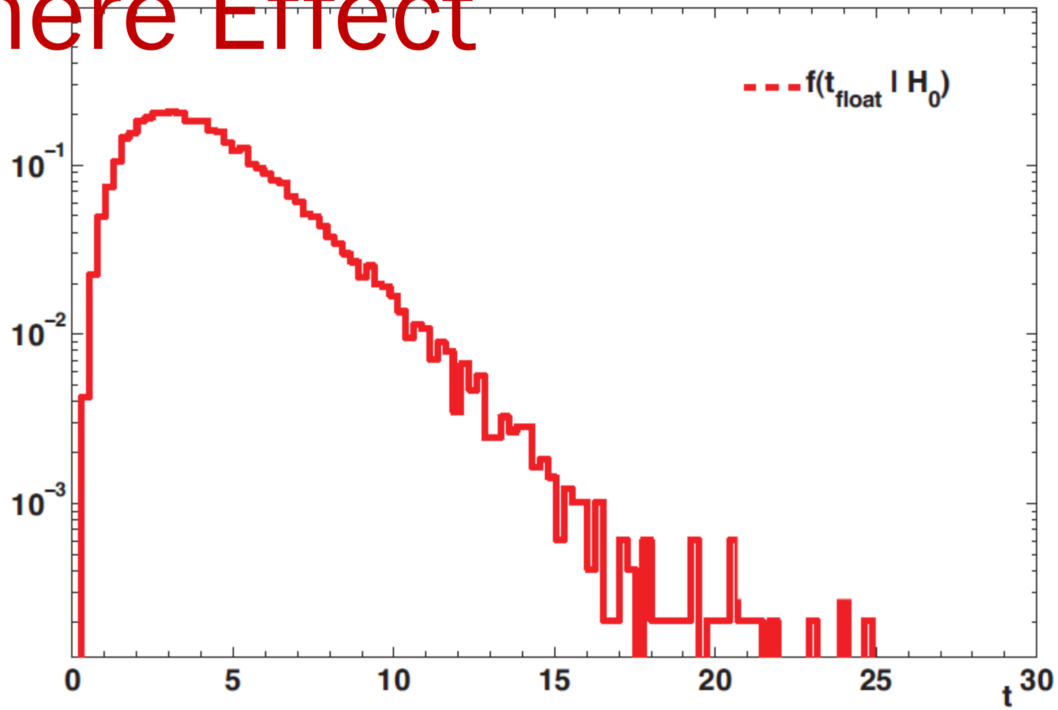
Look Elsewhere Effect

The scan resolution must be less than the signal mass resolution

Assuming the signal can be only at one place, pick the one with the smallest p-value (maximum significance)

This is equivalent to **OPTION II**:

leave the mass floating



$$q_{fix,obs}(\hat{\mu}) = -2\ln\frac{L(b)}{L(\hat{\mu}s(m)+b)}$$

$$q_{float,obs}(\hat{\mu}) = \hat{q}(\hat{\mu}) = \max_m\left\{-2\ln\frac{L(b)}{L(\hat{\mu}s(m)+b)}\right\}$$

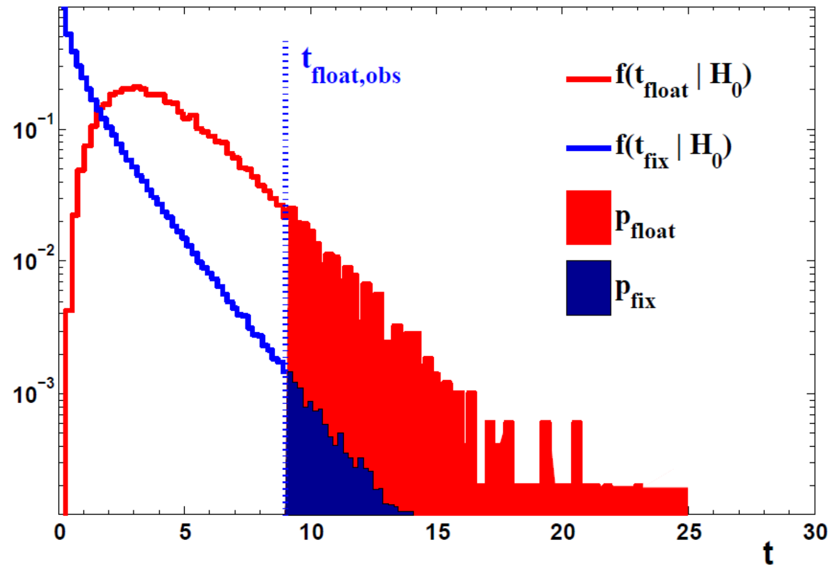
The Thumb Rule

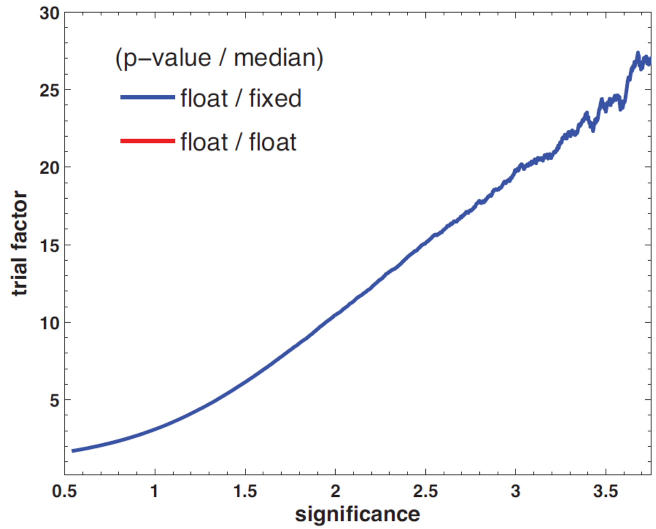
$$\textbf{trial factor} = \frac{p_{float}}{p_{fix}}$$

$$\textbf{trial factor} \stackrel{?}{=} \frac{\textbf{range}}{\textbf{resolution}} = \frac{\Gamma_m}{\sigma_m}$$

This turned out to be wrong, that was a big surprise

$$\textbf{trial factor} \propto \frac{\textbf{range}}{\textbf{resolution}} Z_{local} \propto \frac{\Gamma_m}{\sigma_m} Z_{local}$$





The profile-likelihood test statistic

with a nuisance parameter that is not defined under the Null hypothesis, such as the mass)

Let θ be a nuisance parameter undefined under the null hypothesis, e.g. θ=m

μ="signal strength"

$$H_0 : \mu = 0$$
$$H_1 : \mu > 0$$

- Consider the test statistic:

$$q_0(\theta) = -2\log\frac{\mathcal{L}(\mu=0)}{\mathcal{L}(\hat{\mu},\theta)}$$

- For some fixed θ, $q_0(\theta)$ has (asymptotically) a chi^2 distribution with one degree of freedom by Wilks' theorem.
- $q_0(\theta)$ is a *chi^2 random field* over the space of θ (a random variable indexed by a continuous parameter(s)). we are interested in

$$\hat{q}_0 \equiv q_0(\hat{\theta}) = -2\ln\frac{\mathcal{L}(\mu=0)}{\mathcal{L}(\hat{\mu},\hat{\theta})} = \max_{\theta}[q_0(\theta)]$$

$\hat{\theta}$ is the **global** maximum point

- For which we want to know what is the p-value

$$\text{p-value=P}(\max_{\theta}[q_0(\theta)] \geq u)$$

A small modification

- Usually we only look for 'positive' signals

$$q_0(\theta) = \begin{cases} -2\log\dfrac{\mathcal{L}(\mu=0)}{\mathcal{L}(\hat{\mu},\theta)} & \hat{\mu} > 0 \\ 0 & \hat{\mu} \le 0 \end{cases}$$

$q_0(\theta)$ is 'half chi^2'

[H. Chernoff, Ann. Math. Stat. 25, 573578 (1954)]

The p-value just get divided by $1/2$

- Or equivalently consider $\hat{\mu}$ as a gaussian field

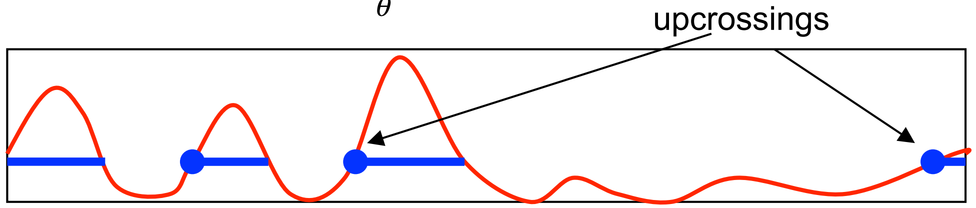
since

$$q_0(\theta) = \left(\frac{\hat{\mu}(\theta)}{\sigma}\right)^2$$

Random fields (1D)

- In 1 dimension: points where the field values become larger then *u* are called *upcrossings*.

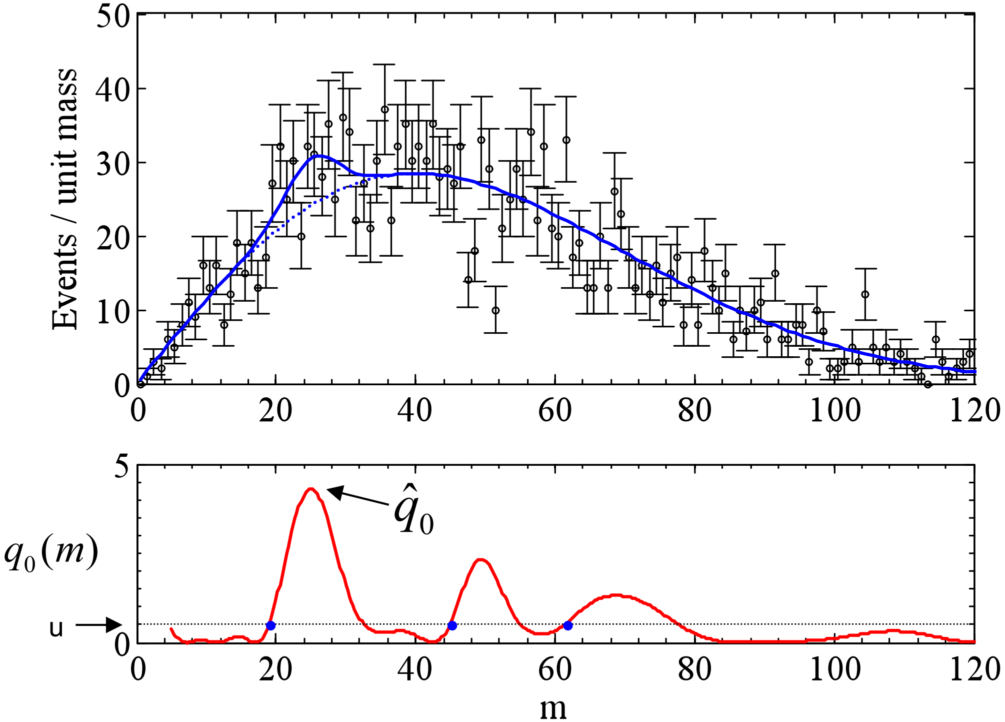
$$\max_{\theta}[q_0(\theta)]$$



- The probability that the global maximum is above the level u is called *exceedance probability*. (p-value of $\hat{q}_0$)

$$P(\max_{\theta}[q_0(\theta)] \geq u)$$

The 1-dimensional case



For a chi² random field, the expected number of *upcrossings* of a level *u* is given by: [Davies,1987]

$$E[N_u] = \mathcal{N}_1 e^{-u/2}$$

To have the global maximum above a level *u:*

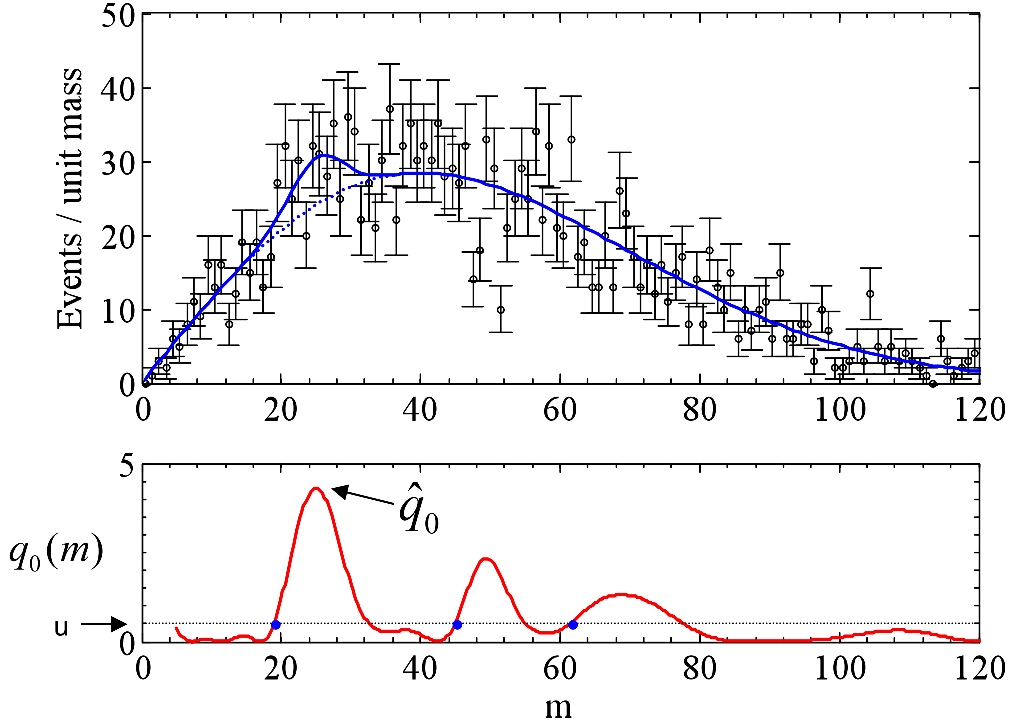
- Either have at least one upcrossing (*N_u>0*) ***or*** have *q_0>u* at the origin *(q_0(0)>u)* :

$$\boxed{\begin{aligned} P(\hat{q}_0 > u) &\le P(N_u > 0) + P(q_0(0) > u) \\ &\le E[N_u] + P(q_0(0) > u) \end{aligned}}$$

[R.B. Davies, Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika **74**, 33–43 (1987)]

Becomes an equality for large *u*

The 1-dimensional case



$$E[N_u] = \mathcal{N}_1 e^{-u/2}$$

The only unknown is $\mathcal{N}_1$ which can be estimated from the average number of upcrossings at some low reference level

$$E[N_u] = N_1 e^{-u/2}$$

$$E[N_{u_0}] = N_1 e^{-u_0/2}$$

$$N_1 = E[N_{u_0}] e^{u_0/2}$$

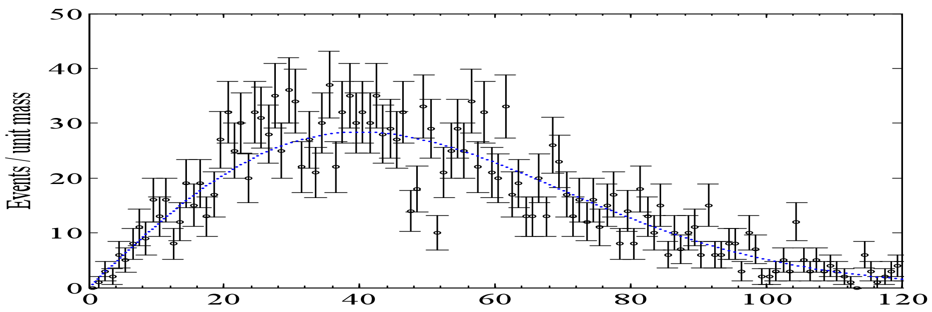
$$E[N_u] = E[N_{u_0}] e^{(u_0-u)/2}$$

$$\mathrm{P}(q_0 > u) \leq E[N_u] + \mathrm{P}(q_0(0) > u)$$

$$= \mathcal{N}_1 e^{-u/2} + \frac{1}{2} P(\chi_1^2 > u) = E[N_{u_0}] e^{(u_0-u)/2} + \frac{1}{2} P(\chi_1^2 > u)$$

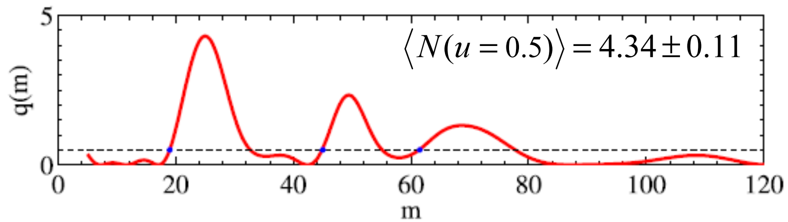
$$p_{global} = E[N_{u_0}] e^{(u_0-u)/2} + p_{local}$$

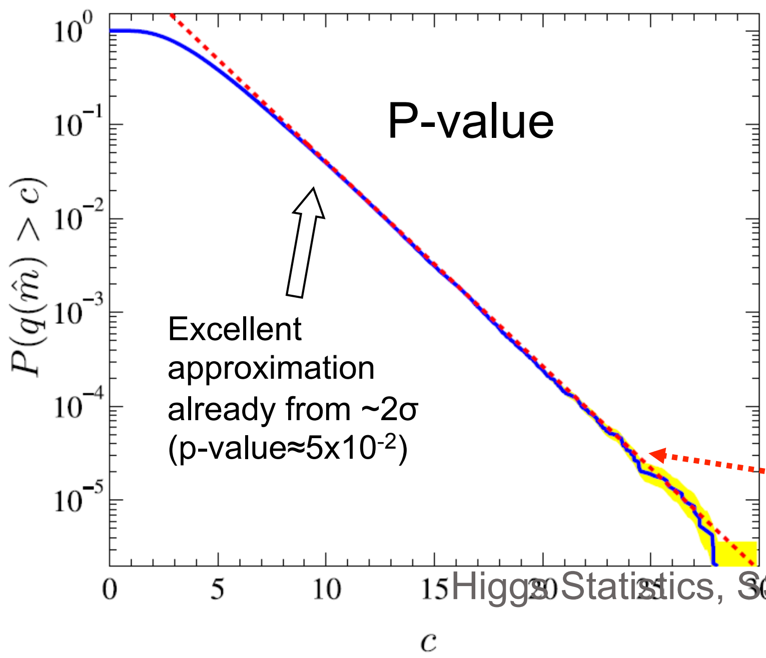
1-D example: resonance search



The model is a gaussian signal (with unknown location m) on top of a continuous background (Rayleigh distribution)

$$\mathcal{L} = \prod_i Poiss(n_i | \mu s_i(m) + \beta b_i)$$



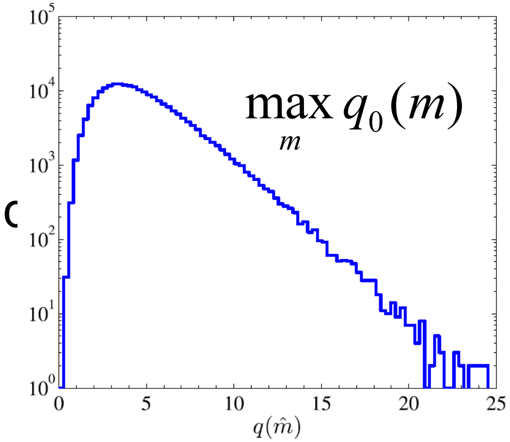


In this example we find

$$\mathcal{N}_1 = 5.58 \pm 0.14$$

[from 100 random background simualtions]

$$\mathcal{N}_1 e^{-u/2} + \frac{1}{2} P(\chi_1^2 > u)$$



[(E. Gross and O. Vitells, Eur. Phys. J. C, 70, 1-2, (2010) , arXiv:1005.1891]

A real life example

$$P(q_0 > u) \le E[N_u] + P(q_0(0) > u)$$

$$E[N_u] = \mathcal{N}_1 e^{-u/2}$$

$$\mathcal{N}_1 \cong \left\langle N_{u_0} \right\rangle e^{u_0/2}$$

$$P(q_0 > u) = \mathcal{N}_1 e^{-u/2} + \frac{1}{2} P(\chi_1^2 > u)$$

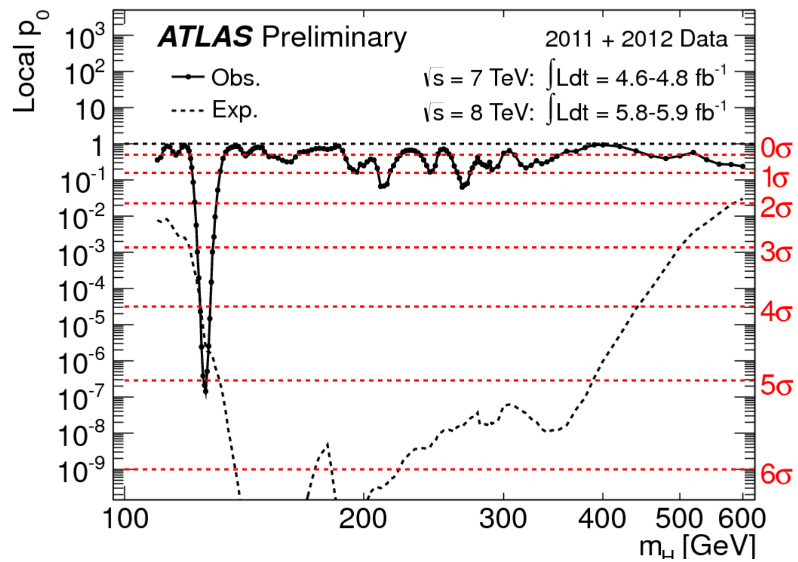
$$p_{global} = \mathcal{N}_1 e^{-u/2} + p_{local}$$

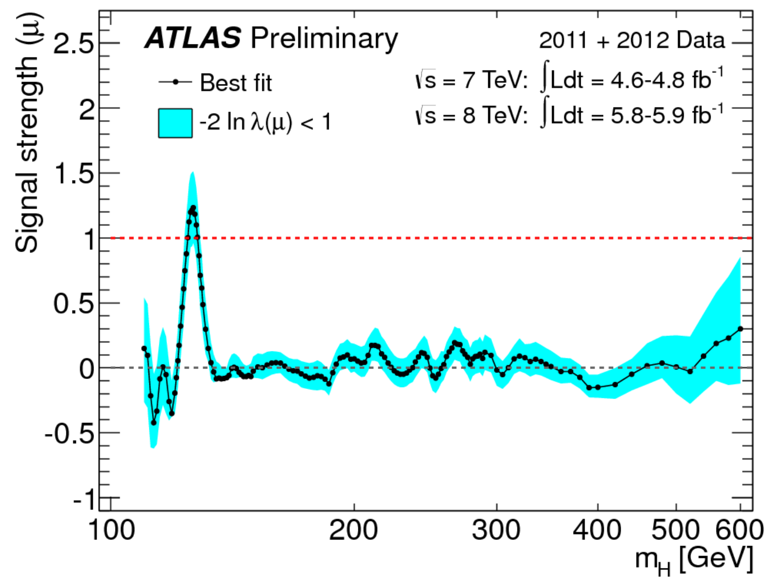
$$p_{global} = \left\langle N_{u_0} \right\rangle e^{\frac{u_0 - u}{2}} + p_{local}$$

$$N_{u_0=0} = 9 \pm 3$$

$$p_{global} = 9 \cdot \mathrm{e}^{-25/2} + O(10^{-7}) = 3.3 \cdot 10^{-5}$$

$$5\sigma \rightarrow 4\sigma \quad \text{trial\#}{\sim}100$$





Bloggers Spot

A combination on a back of an envelope

An exercise in combining experiments (or channels)

- We assume two channels and ignore correlated systematics

$$\mathcal{L} = \mathcal{L}_1(\mu, \theta_1)\mathcal{L}_2(\mu, \theta_2)$$

- We have

$$-2\log \mathcal{L}_i(\mu, \hat{\hat{\theta}}_i) = \left(\frac{\mu - \hat{\mu}_i}{\sigma_i}\right)^2 + const.$$

- It follows that

$$\hat{\mu} = \frac{\hat{\mu}_1 \sigma_1^{-2} + \hat{\mu}_2 \sigma_2^{-2}}{\sigma_1^{-2} + \sigma_2^{-2}}$$

- Variance of $\hat{\mu}$ is is given by $\sigma^{-2} = \sigma_1^{-2} + \sigma_2^{-2}.$

An exercise in combining experiments (or channels)

- The combined limit at CL 1- α is given by

$$\mu_{up} = \hat{\mu} + \sigma\Phi^{-1}(1 - \alpha\Phi(\frac{\hat{\mu}}{\sigma}))$$

- The combined discovery p-value is given by

$$p_0 = 1 - \Phi(\hat{\mu}/\sigma)$$

- Median upper limit

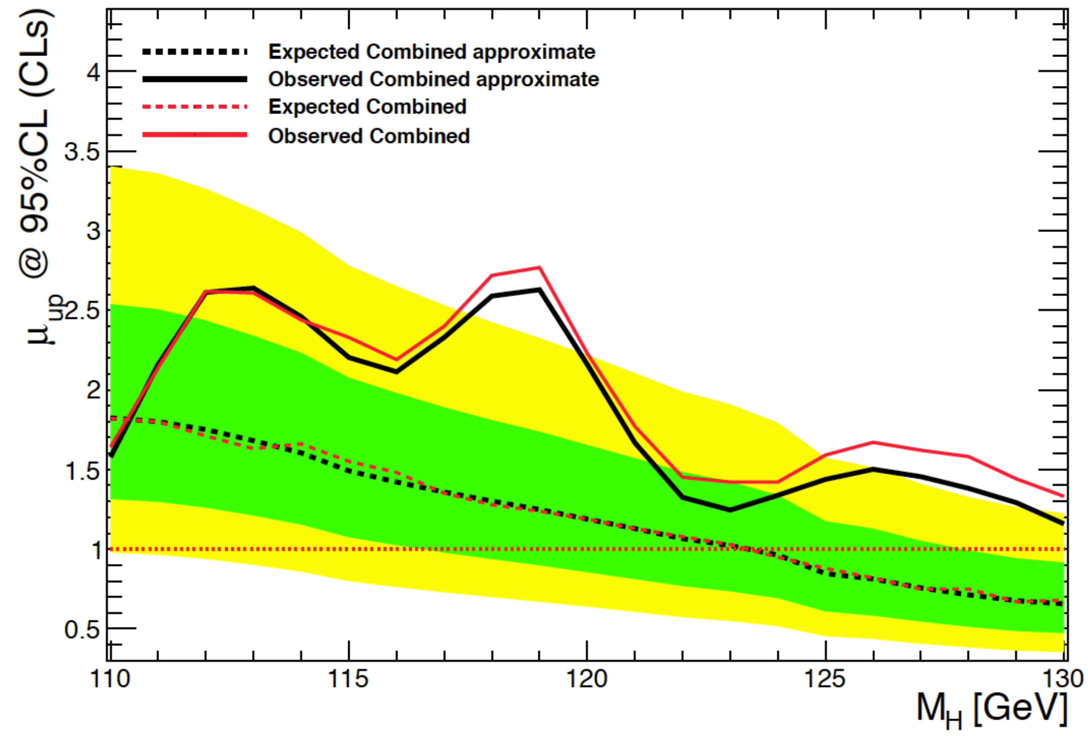
$$\mu_{up}^{med} = \sigma\Phi^{-1}(1 - \alpha/2)$$

- Which gives

$$\frac{1}{(\mu_{up}^{med})^2} = \frac{1}{(\mu_{up,1}^{med})^2} + \frac{1}{(\mu_{up,2}^{med})^2}$$

An exercise in combining experiments (or channels)

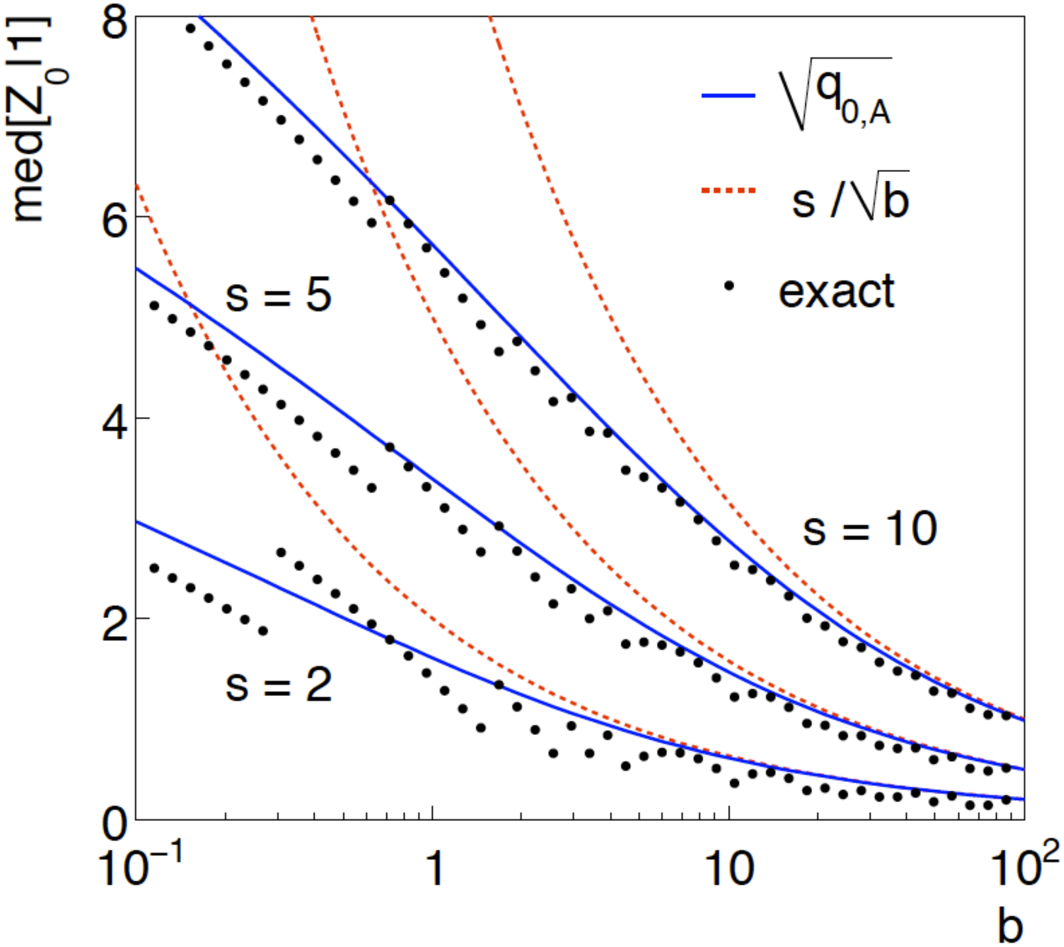
- This combination takes onto account fluctuations of the observed limit



Some Profile Likelihood Useful Spinoffs

Counting on the back of the envelope

- $s/\sqrt{b}$?



$$\mathrm{med}[Z_0|1] = \sqrt{q_{0,\mathrm{A}}} = \sqrt{2\left((s+b)\ln(1+s/b) - s\right)}$$

- The intuitive explanation of $s/\sqrt{b}$ is that it compares the signal, s, to the standard deviation of n assuming no signal, $\sqrt{b}$.
- Now suppose the value of b is uncertain, characterized by a standard deviation σ_b.
- A reasonable guess is to replace $\sqrt{b}$ by the quadratic sum of $\sqrt{b}$ and σ_b, i.e.,

$$\mathrm{med}[Z|s] = \frac{s}{\sqrt{b+\sigma_b^2}}$$

Systematics is Important

- An analysis might be killed by systematics

$$b \pm \Delta \cdot b \Rightarrow \sigma_b = \sqrt{\left(\sqrt{b}\right)^2 + (\Delta \cdot b)^2} = \sqrt{b + \Delta^2 b^2}$$

$$s / \sqrt{b} \Rightarrow s / \sqrt{b(1 + b\Delta^2)} \xrightarrow{L \to \infty} \frac{s/b}{\Delta}$$

$$\frac{s/b}{\Delta} \geq 5 \rightarrow s/b \geq 0.5 \text{ for } \Delta \sim 10\%$$

We can do better

Significance with systematics

- We find (G. Cowan)

$$Z_{\mathrm{A}} = \left[2 \left((s+b) \ln \left[\frac{(s+b)(b+\sigma_b^2)}{b^2 + (s+b)\sigma_b^2} \right] - \frac{b^2}{\sigma_b^2} \ln \left[1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)} \right] \right) \right]^{1/2}$$

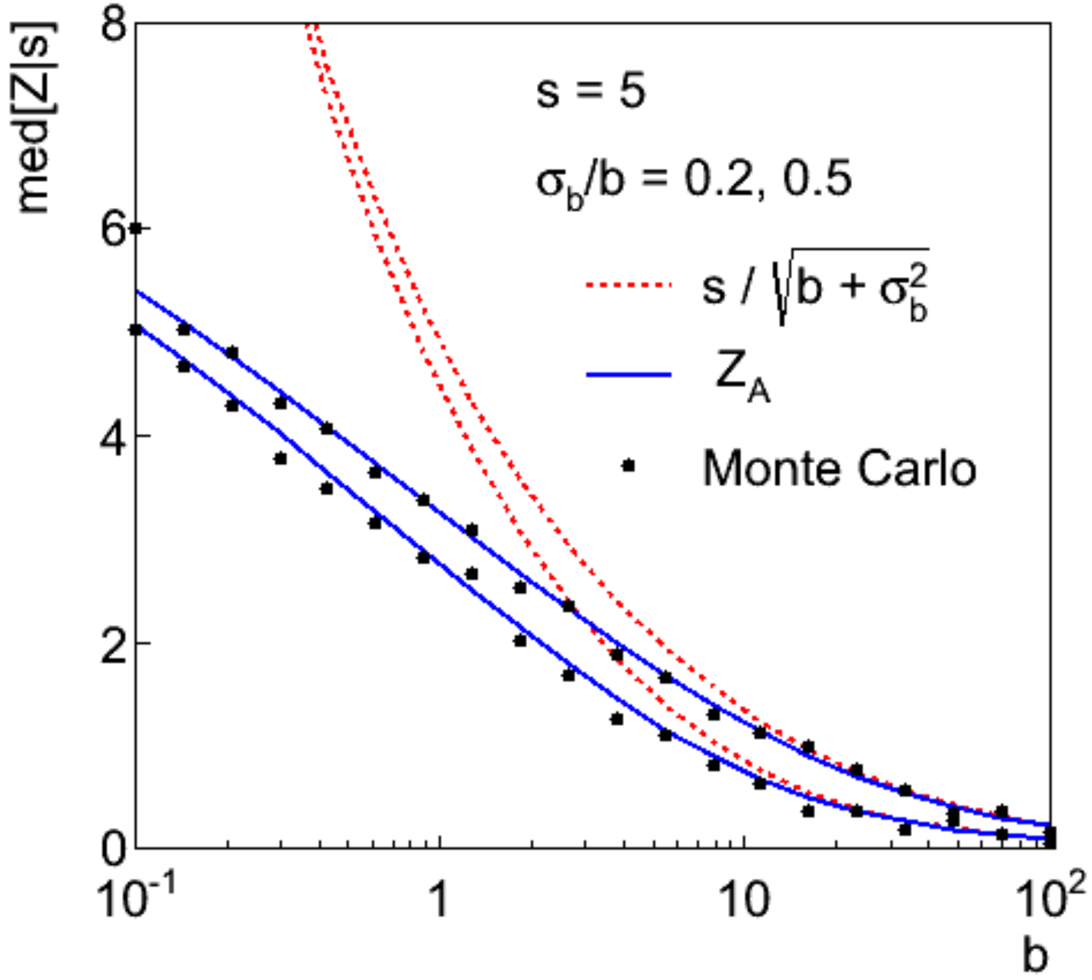
$\sigma_b{}^2/b$ gives

$$Z_{\mathrm{A}} = \frac{s}{\sqrt{b+\sigma_b^2}} \left(1 + \mathcal{O}(s/b) + \mathcal{O}(\sigma_b^2/b) \right)$$

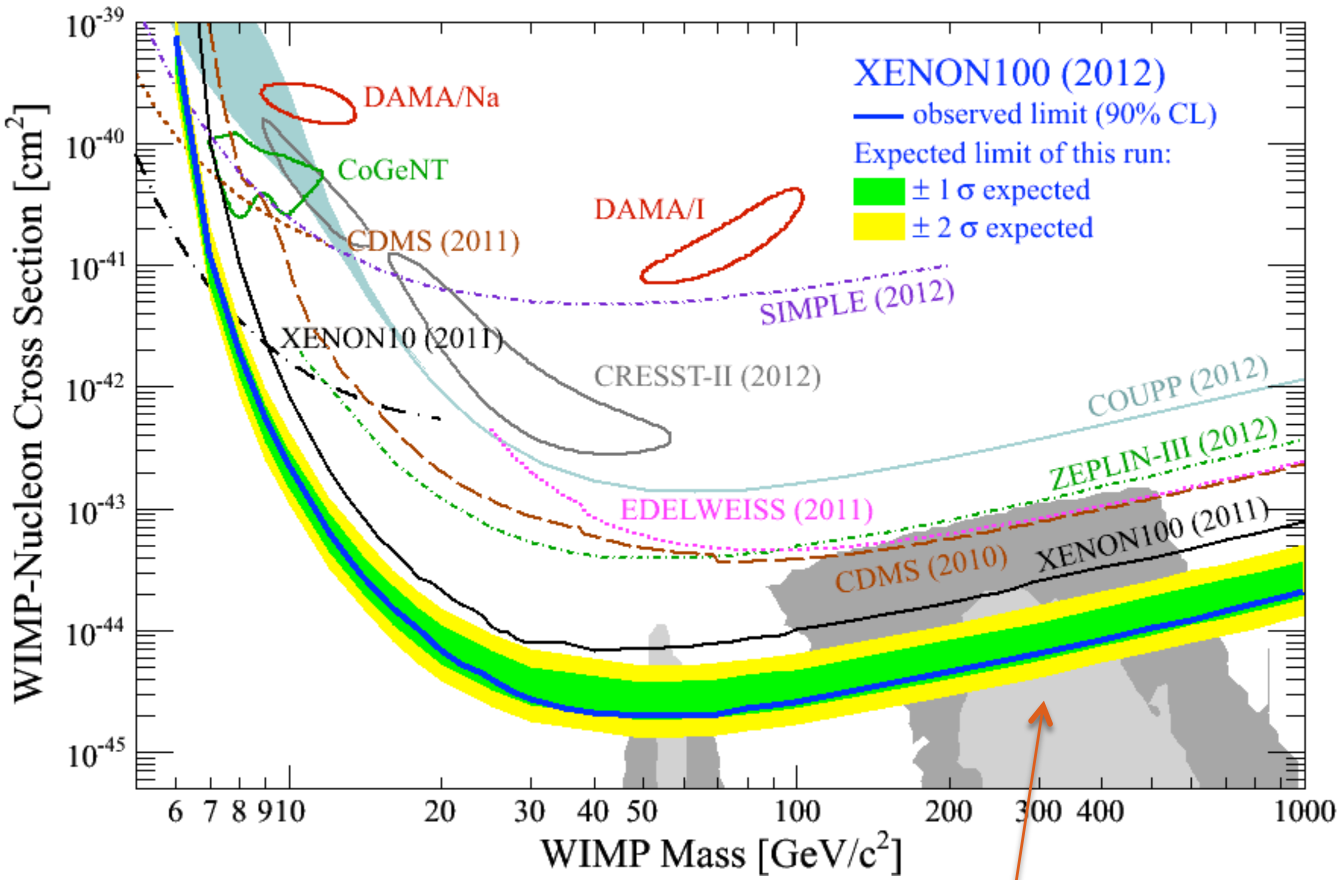
- So the "intuitive" formula can be justified as a limiting case

of the significance from the profile likelihood ratio test evaluated with the Asimov data set.

Significance with systematics



Implications in Astro-Particle Physics



The lack of events in spite of an expected background allows us to set a better limit than the expected

Conclusions

- 10 years of statistics research in HEP as of LEP in ~2000 brought the community to a robust and sensitive method to extract signals
- CLs , LEE, Asimov…all the new jargon became public property
- The method was used to fish the Higgs signal from a $\gamma\gamma$ rare signal
- It was successfully applied in ATLAS and CMS to discover the Higgs Boson
- It was also successfully use by the XENON collaboration in its search for dark matter.

What does my daughter study in school?

Subjective Bayesian is Good for YOU

Thomas Bayes (b 1702)
a British mathematician and Presbyterian minister

What is the Right Question

- Is there a Higgs Boson? What do you mean?
  Given the data , is there a Higgs Boson?
- Can you really answer that without any a priori knowledge of the Higgs Boson?
  Change your question: What is your degree of belief in the Higgs Boson given the data…
  Need a prior degree of belief regarding the Higgs Boson itself…

$$P(Higgs \mid Data) = \frac{P(Datas \mid Higgs)P(Higgs)}{P(Data)} = \frac{L(Higgs)\pi(Higgs)}{\int L(Higgs)\pi(Higgs)d(Higgs)}$$

$$L(\text{Higgs}) \equiv P(\text{Data}|\text{Higgs})$$

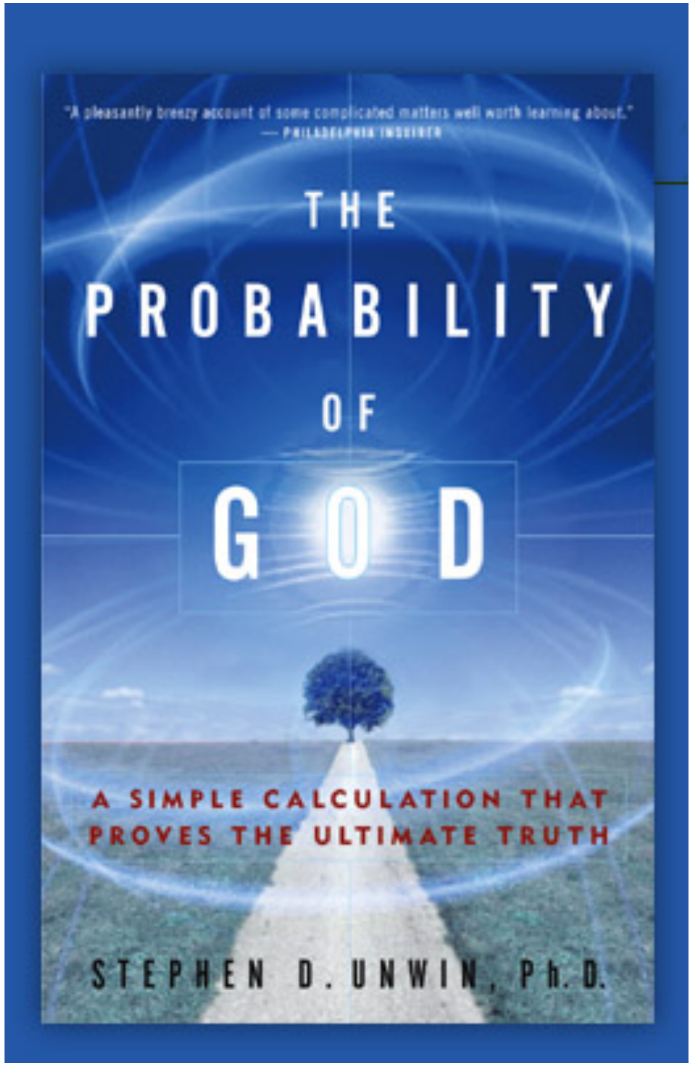
- Make sure that when you quote your answer you also quote your prior assumption!
- Can we assign a probability to a model P(Higgs)? Of course not, we can only assign to it a degree of belief!
- The most refined question is:
  - Assuming there is a Higgs Boson with some mass m_H, how well the data agrees with that?
  - But even then the answer relies on the way you measured the data (i.e. measurement uncertainties), and that might include some pre-assumptions, priors!

What is the Right Answer?

- The Question is:
- Is there a Higgs Boson?
- Is there a God?

$$P(God \mid Earth) = \frac{P(Earth \mid God)P(God)}{P(Earth)}$$

- In the book the author uses
  - "divine factors" to estimate the P(Earth | God),
  - a prior for God of 50%
- He "calculates" a 67% probability for God's existence given earth…
- In Scientific American July 2004, playing a bit with the "divine factors" the probability drops to 2%…



Basic Definitions: The Bayesian Way

- Can the model have a probability (what is Prob(SM)?) ?
- We assign a degree of belief in models parameterized by μ
- $$n = \mu \cdot s(m_H) + b \qquad \begin{array}{l} H_0 : \mu = 0 \\ H_1 : \mu = 1 \end{array}$$
- $$p(\mu \mid x) = \frac{L(\mu)\pi(\mu)}{\int L(\mu)\pi(\mu)d\mu}$$
- Instead of talking about confidence intervals we talk about credible intervals, where $p(\mu \mid x)$ is the credibility of μ given the data.

Basic Definitions: Priors

$$P(\mu \mid data) \sim \int L(\mu,\theta)\pi(\theta)\,d\mu\,d\theta$$

- A prior probability is interpreted as a description of what we believe about a parameter preceding the current experiment
  - **Informative Priors**: When you have some information about θ, the prior might be informative (Gaussian or Truncated Gaussians…)
    - Most would say that subjective informative priors about the parameters of interest should be avoided ("….what's wrong with assuming that there is a Higgs in the mass range [115,140] with equal probability for each mass point?")
    - Subjective informative priors about the Nuisance parameters are more difficult to argue with
      - These Priors can come from our assumed model (Pythia, Herwig etc…)
      - These priors can come from subsidiary measurements of the response of the detector to the photon energy, for example.
      - Some priors come from subjective assumptions (theoretical, prejudice symmetries, ….) of our model

Uninformative Priors

- **Uninformative Priors**: All priors on the parameter of interest are usually uninformative….

- Therefore flat uninformative priors are most common in HEP.
  - When taking a uniform prior for the Higgs mass [115,130]… is it really uninformative? do uninformative priors exist?
  - When constructing an uninformative prior you actually put some information in it…

- **But** a prior flat in the coupling g will not be flat in $\sigma \sim g^2$ Depends on the metric!

- Note, flat priors are improper and might lead to serious problems of undercoverage (when one deals with >1 channel, i.e. beyond counting, one should AVOID them)