Modifications to original lectures:
- How to determine σ (p. 107)
- Plotting Pulls and Impact of NPs (p.200-p.225)
- Combining Results on a back of an Envelope (p. 226-p.229)

Search and Discovery Statistics in HEP

Eilam Gross, Weizmann Institute of Science

This presentation would have not been possible without the tremendous help of the following people throughout many years

Louis Lyons, Alex Read, Glen Cowan ,Kyle Cranmer
Ofer Vitells & Bob Cousins

Search and Discovery Statistics in HEP
Lecture 1: INTRODUCTION

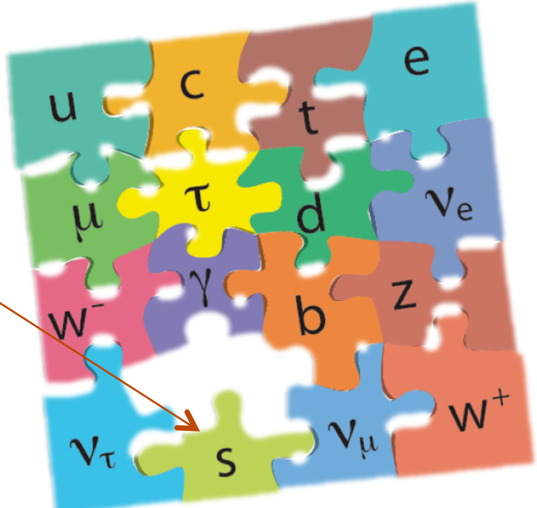
Eilam Gross, Weizmann Institute of Science

This presentation would have not been possible without the tremendous help of the following people throughout many years

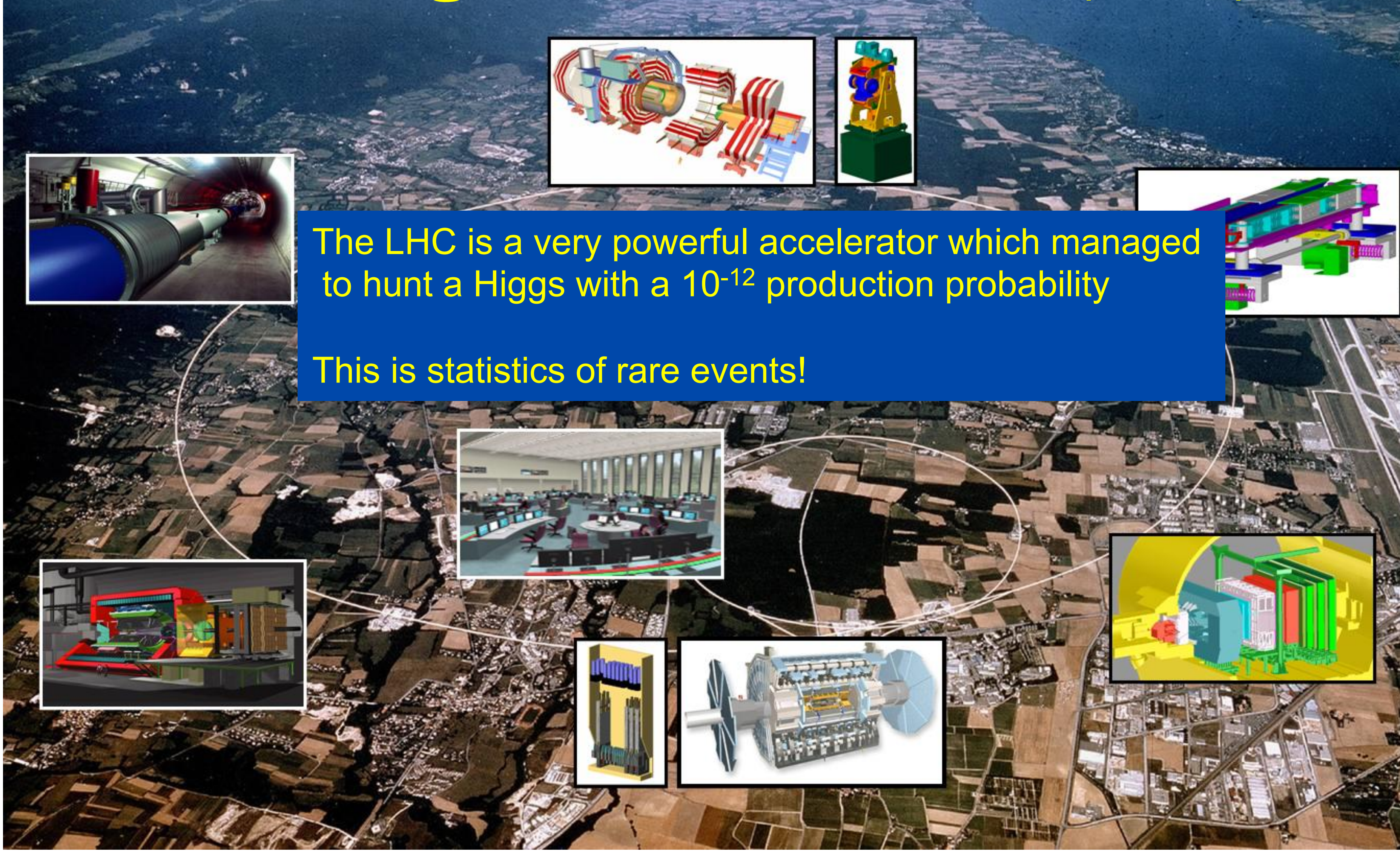
Louis Lyons, Alex Read, Glen Cowan ,Kyle Cranmer ,
Ofer Vitells & Bob Cousins

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 2012 discovery of 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 if it is the expected Higgs Boson (Mass, Spin, CP) or a member of a family of Scalar Bosons



The Large Hadron Collider (LHC)



The LHC is a very powerful accelerator which managed to hunt a Higgs with a 10^{-12} production probability

This is statistics of rare events!

Higgs Hunter's Independence Day July 4th 2012

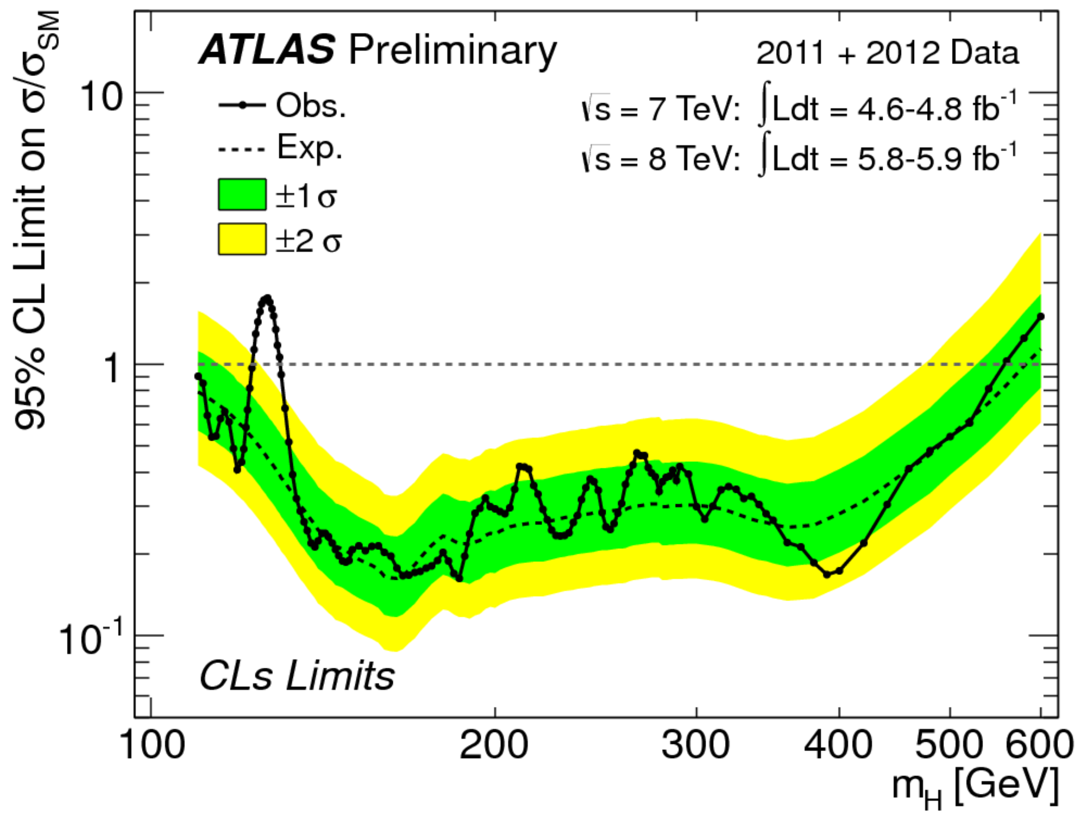
The Charge of the Lectures

The Brazil Plot, what does it mean?

Observed Limit

Bands

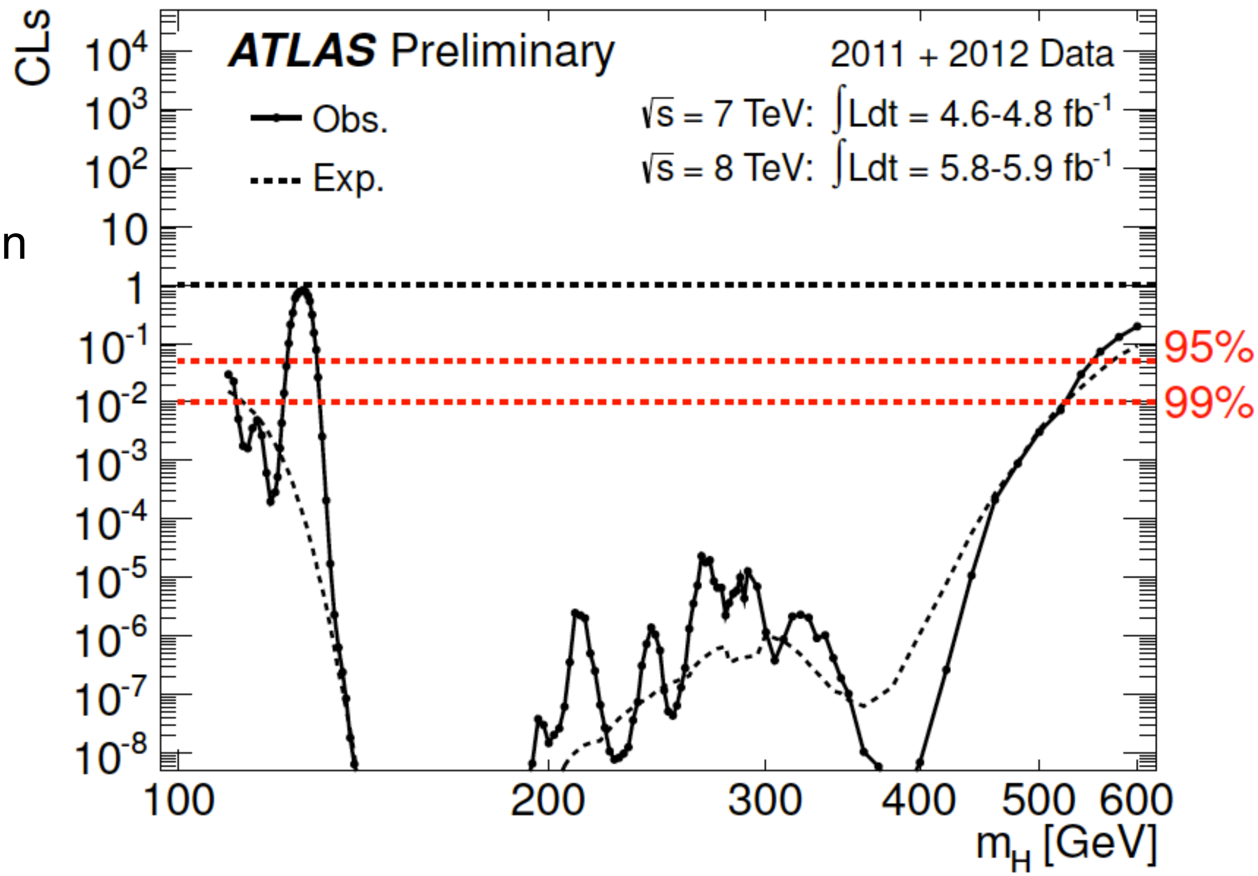
Expected Limit



What the ---- CLs?

What is exclusion at the 95% CL?

99% CL?



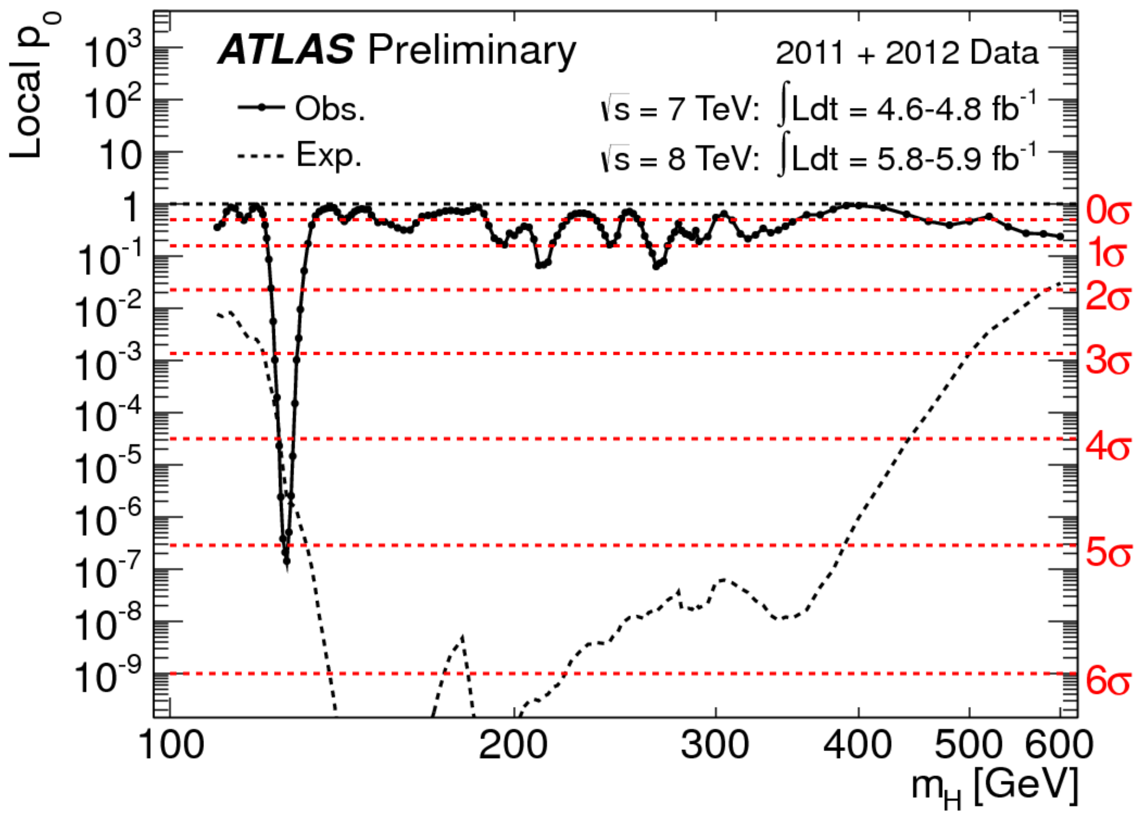
The p0 discovery plot, how to read it?

p-value

Local p0

Expected p0

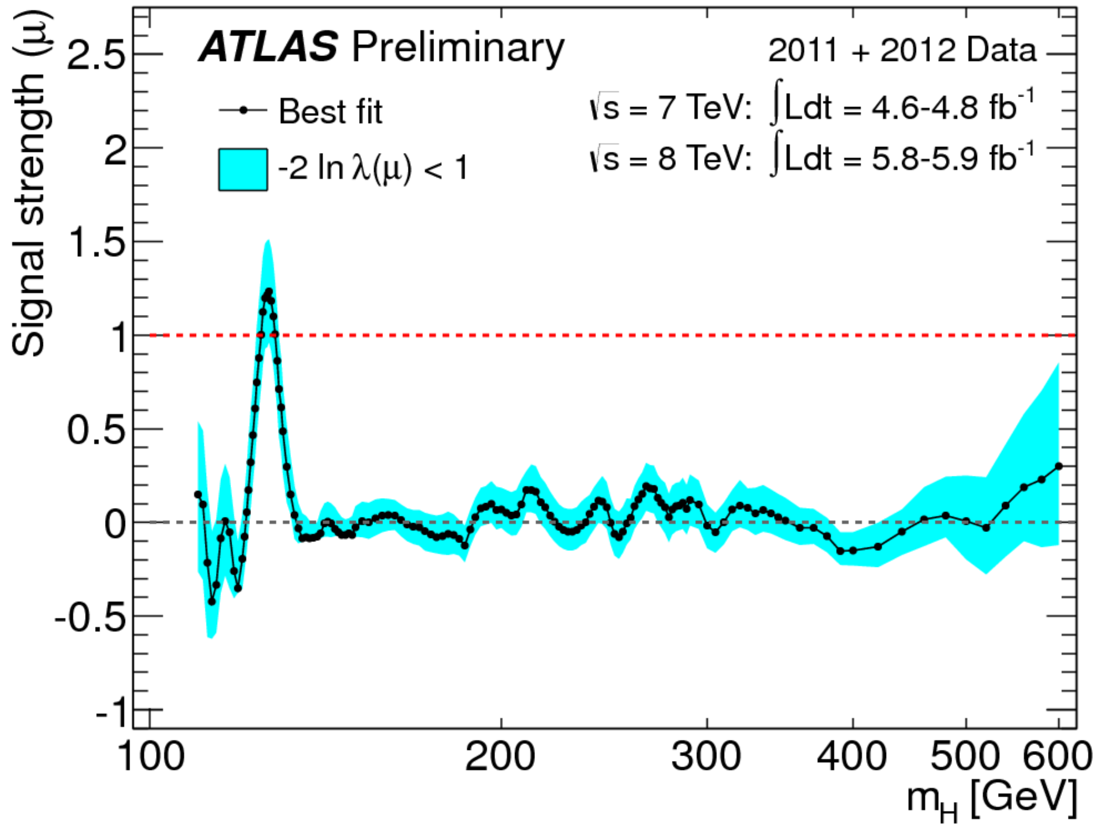
Observed local p0



Global p0 and the Look Elsewhere Effect

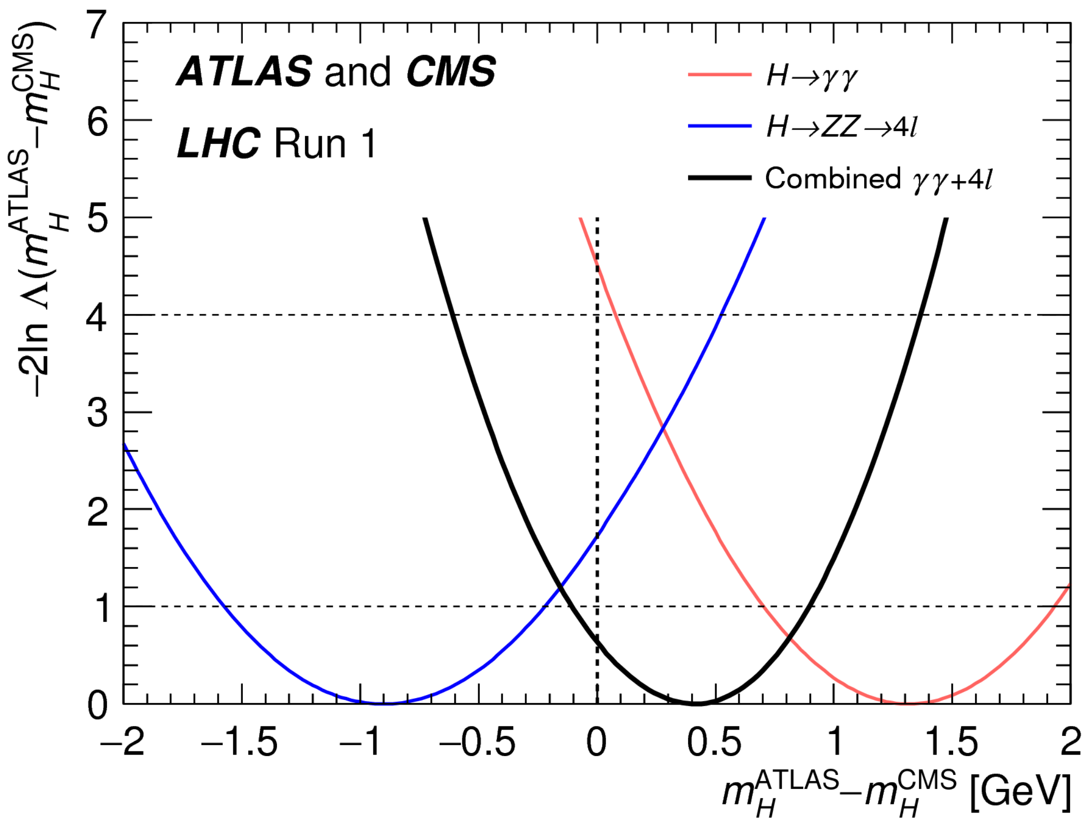
The cyan band plot, what is it?

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



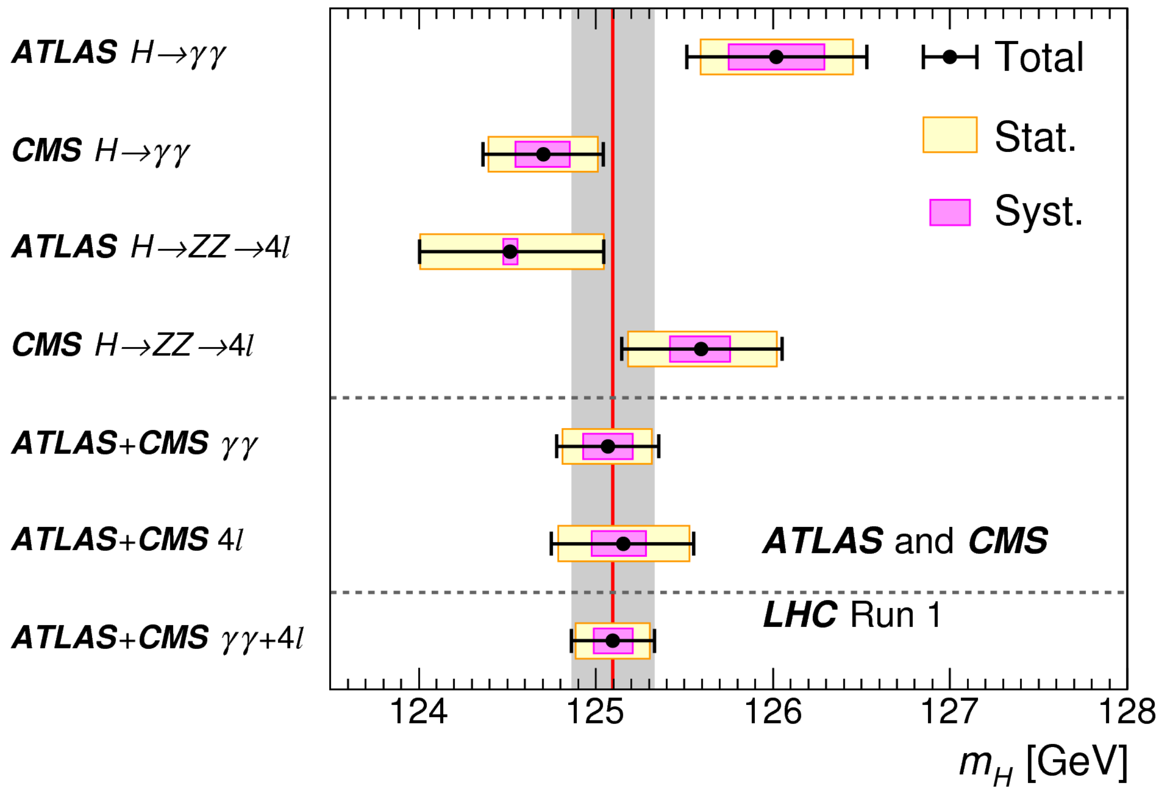
Towards a measurement

Likelihoods Scans



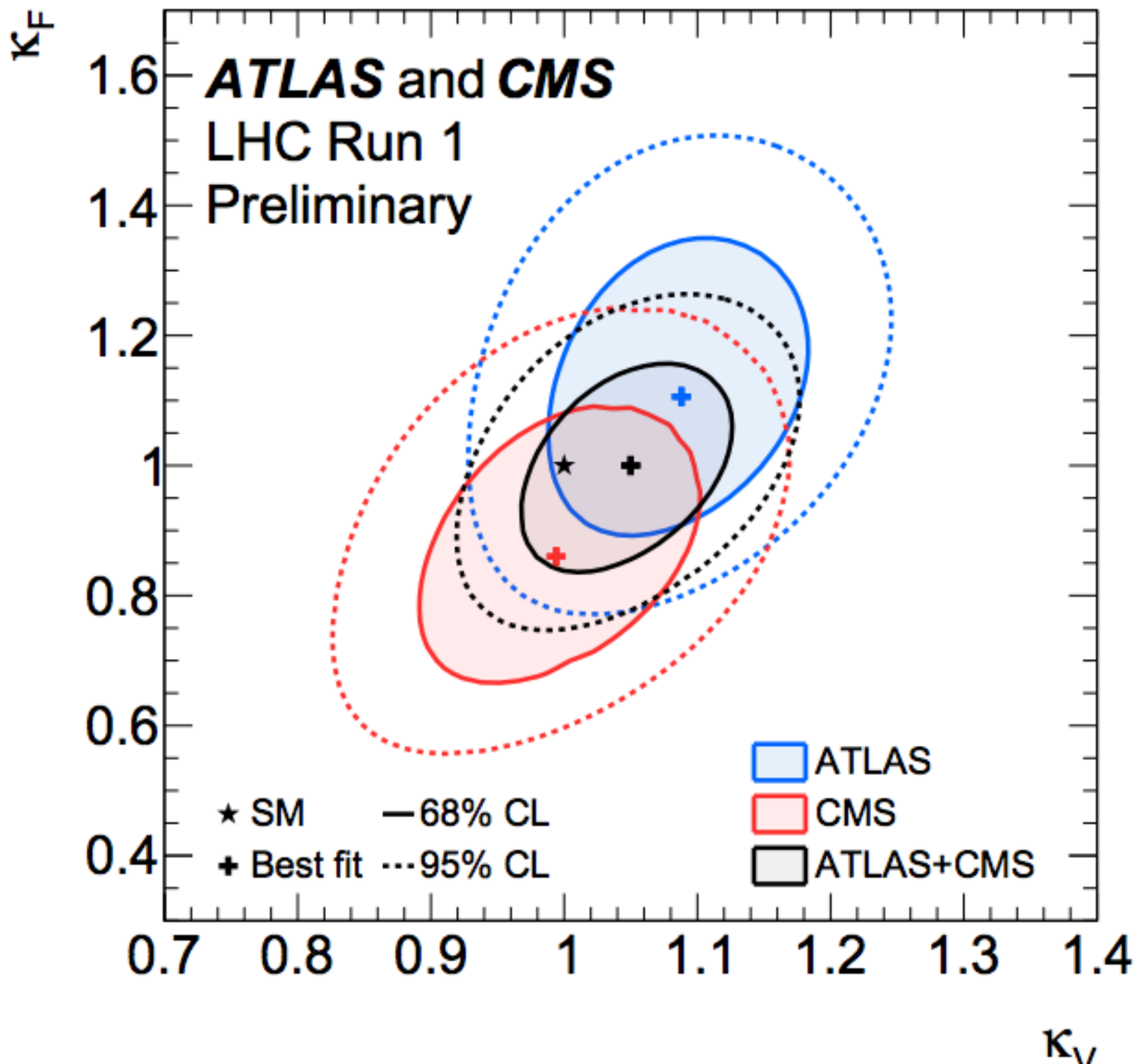
Towards a measurement

Measurements & Systematics vs Stat errors

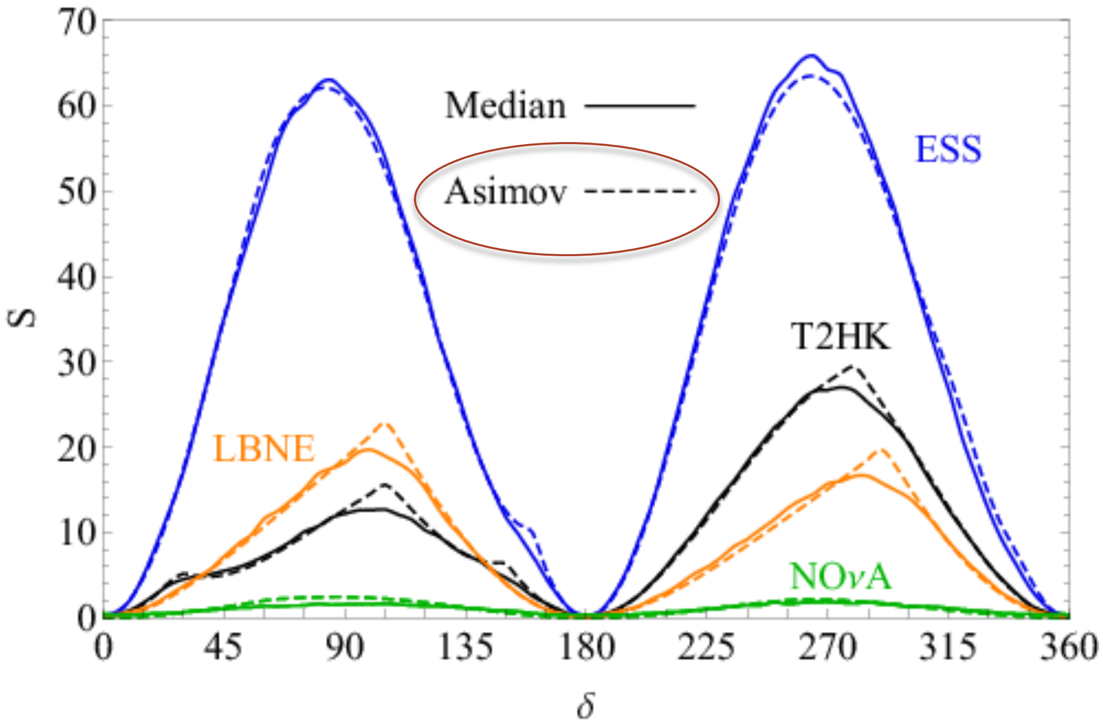


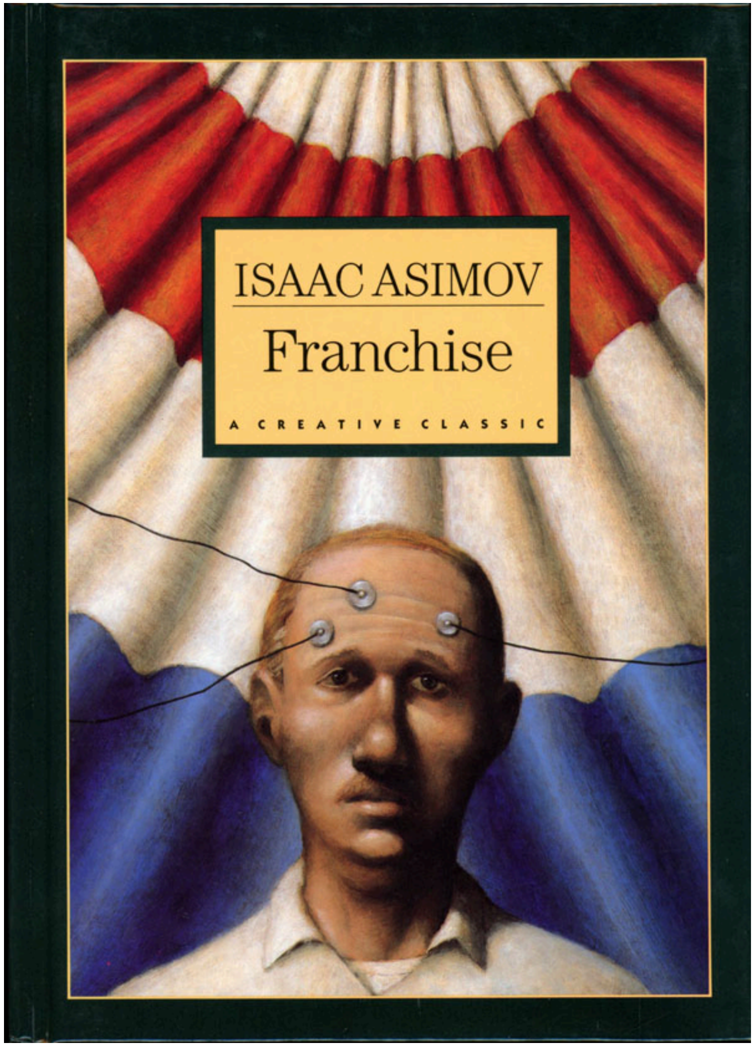
Towards a measurement

2-D Likelihoods



The Asimov Data Set





References in the Discovery Papers

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

RooStats [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.

CLs [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.

More Refs (taken from CMS legacy Run 1 Paper)

Wilks Approximation

[186] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses", *Ann. Math. Statist.* **9** (1938) 60, doi:10.1214/aoms/1177732360.

Wald Approximation

[187] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large", *Trans. Amer. Math. Soc.* **54** (1943) 426, doi:10.1090/S0002-9947-1943-0012401-3.

Wald Approximation

[188] R. F. Engle, "Chapter 13 Wald, likelihood ratio, and Lagrange multiplier tests in econometrics", in *Handbook of Econometrics*, Z. Griliches and M. D. Intriligator, eds., volume 2, p. 775. Elsevier, 1984. doi:10.1016/S1573-4412(84)02005-5.

[189] G. J. Feldman and R. D. Cousins, "Unified approach to the classical statistical analysis of small signals", *Phys. Rev. D* **57** (1998) 3873, doi:10.1103/PhysRevD.57.3873, arXiv:physics/9711021.

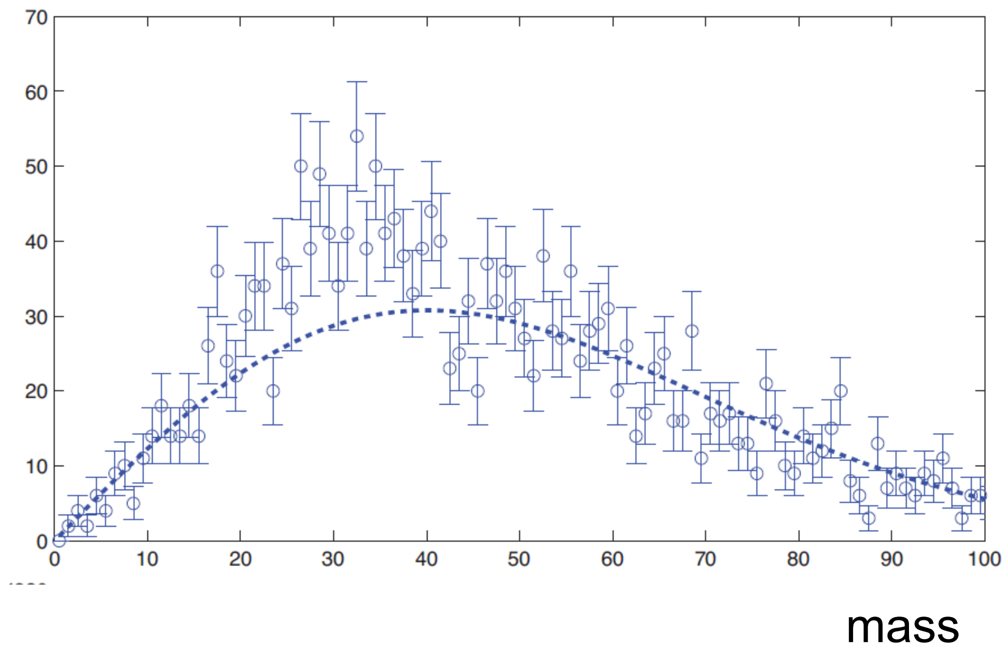
Feldman-Cousins

The Statistical Challenge of HEP

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

The searched particle mass is unknown (for the sake of this lecture)

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



WE NEED TO KNOW WHAT WE SEARCH FOR…..

We need to have a model

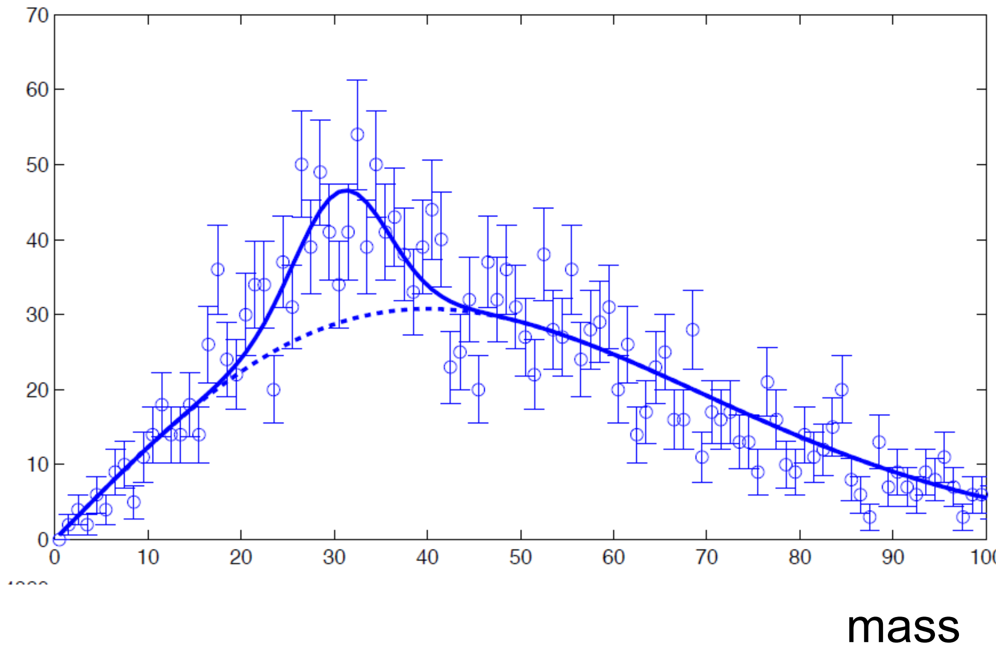
We need to have two hypotheses if we want a powerful test

The Statistical Challenge of HEP

So the statistical challenge is obvious:

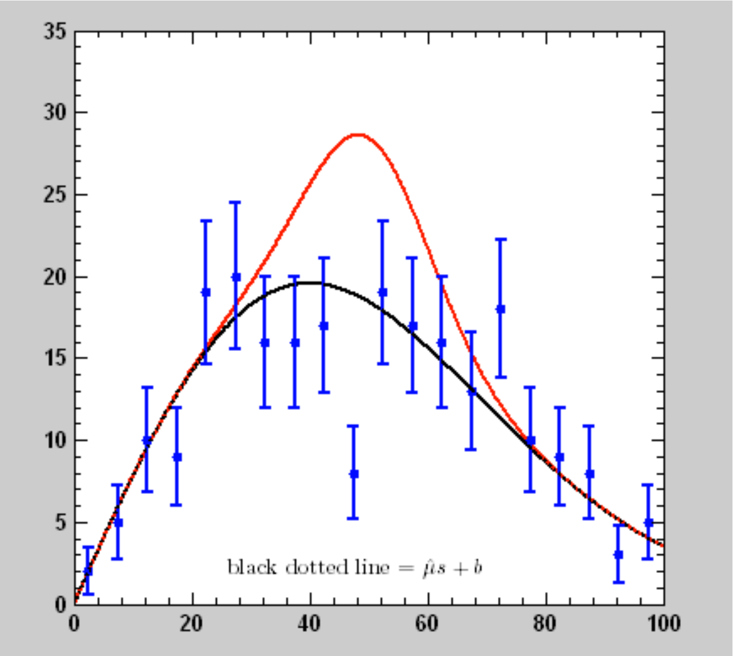
To tell in the most powerful way, and to the best of our current scientific knowledge, if there is new physics, beyond what is already known, in our data

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 $s(m_H)$

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

For simplicity unless otherwise noted $s(m_H) = L \cdot \sigma_{SM}(m_H)$

- 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 (remember that it is the null hypothesis ONLY if we aim at a discovery)

- 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**, NOT an hypothesis 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

H_0- SM w/o Higgs

ALTERNATE

H_1- SM with Higgs

- Reject H_0 in favor of H_1 – A DISCOVERY

Swapping Hypotheses → exclusion

NULL

ALTERNATE

H_0- SM w/o Higgs

H_1- SM with Higgs

- Reject H_1 in favor of H_0

Excluding H_1 (m_H) → Excluding the Higgs with a mass m_H

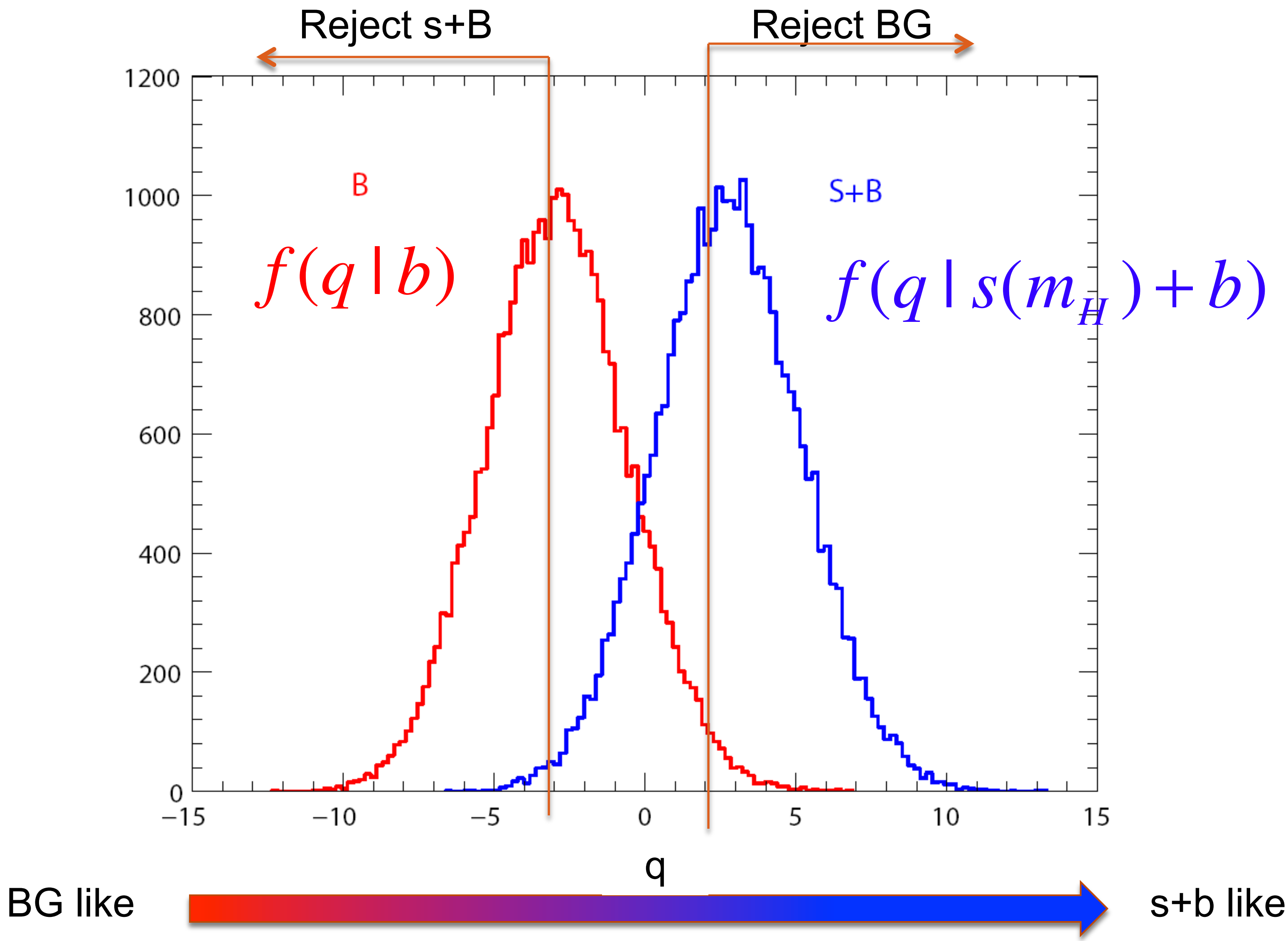
Testing an Hypothesis (wikipedia…)

- 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?**

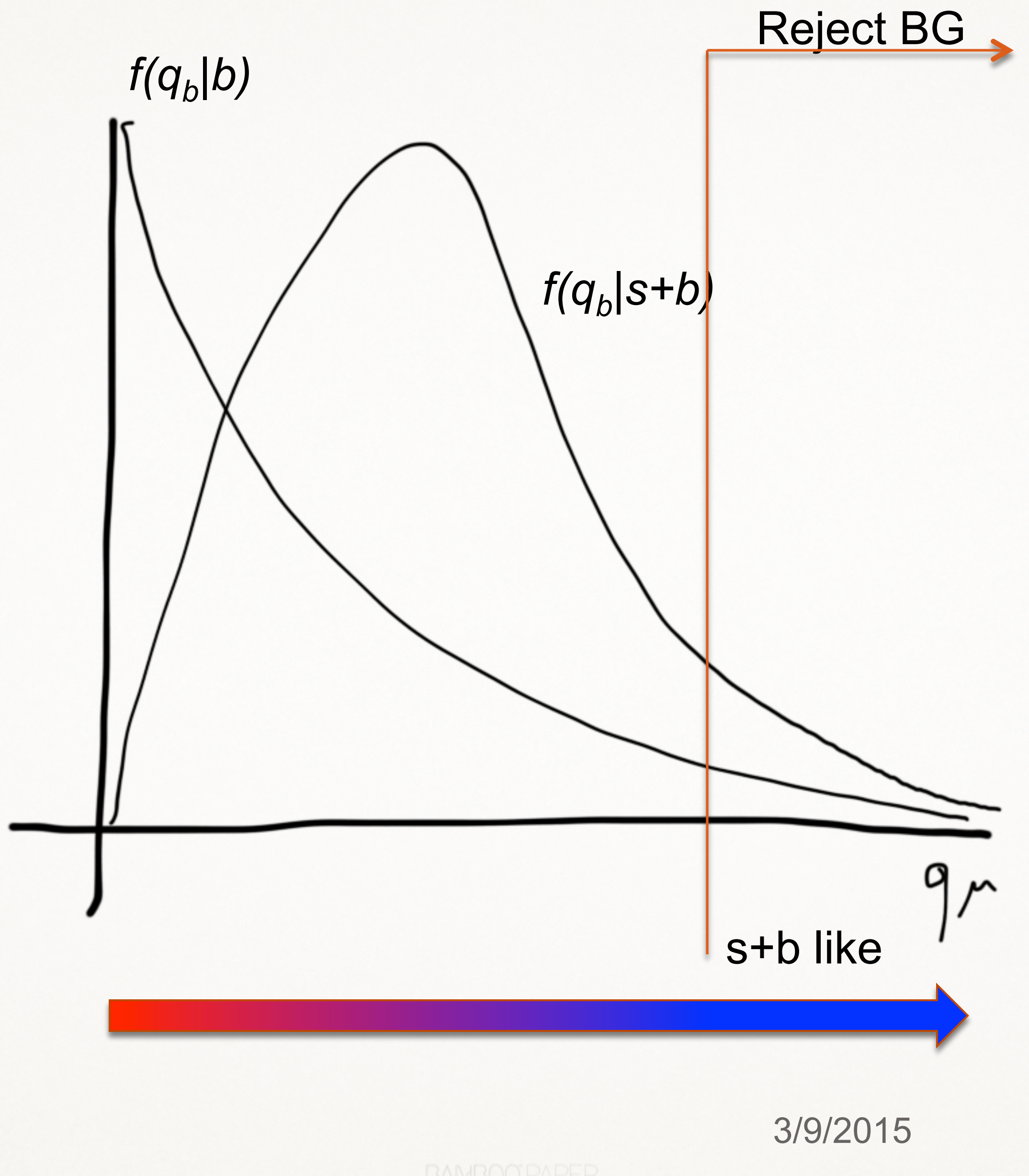
Test statistic and p-value

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 ($s(m_H){+}b$ vs b)



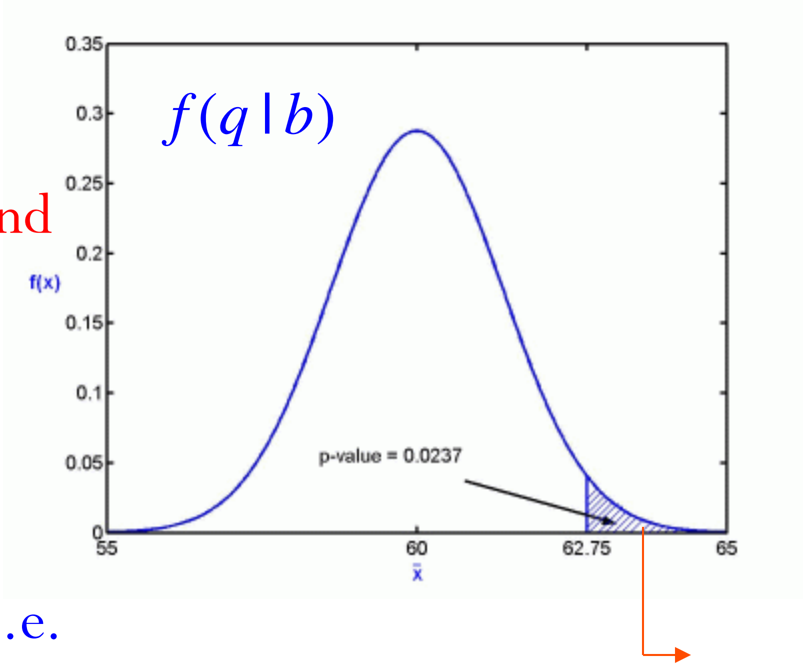
p-Value

- 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 critical region, i.e.

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

→A discovery

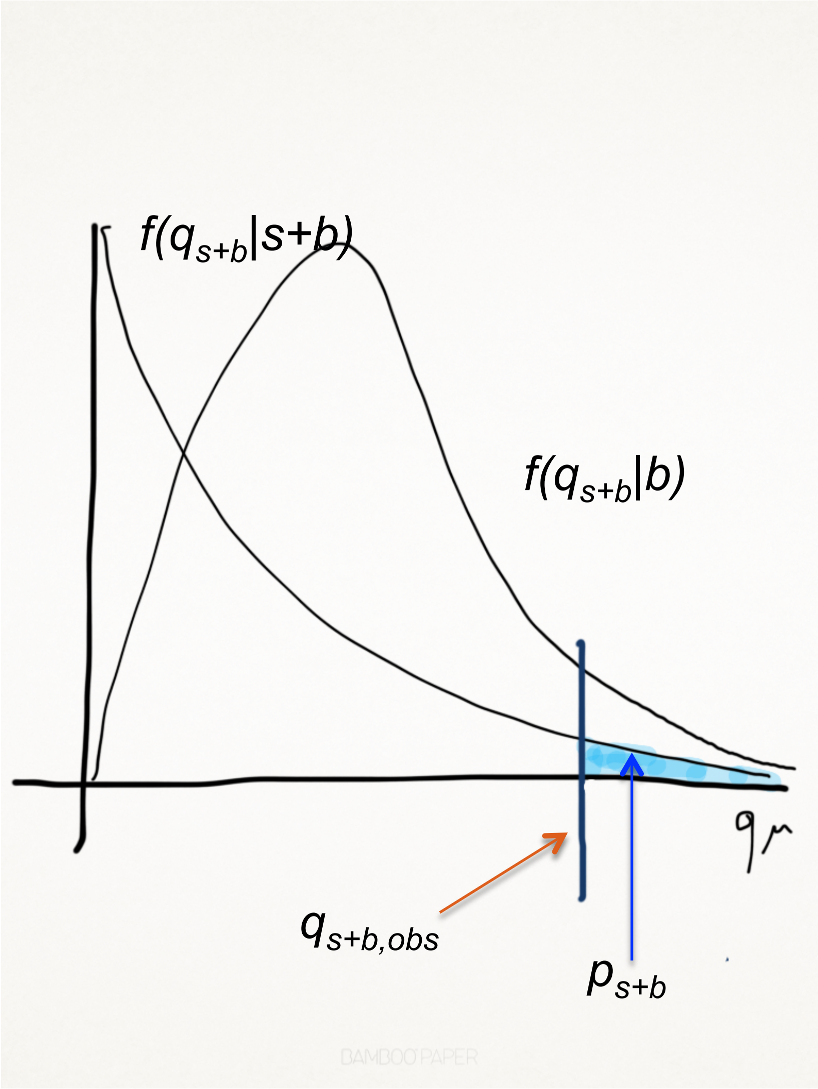
- The pdf of **q**….



Critical region
Of size α

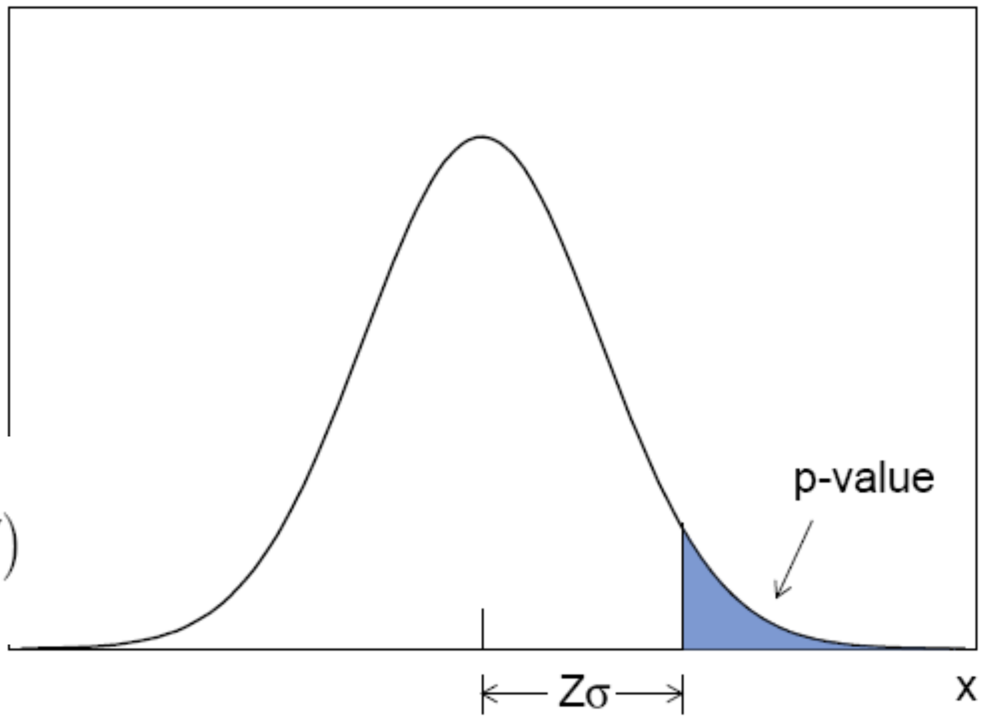
p-value – testing the signal hypothesis

- When testing the signal hypothesis, the p-value is the probability that the observation is less compatible with the signal hypothesis (more 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 at the 95% Confidence Level (CL) → 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



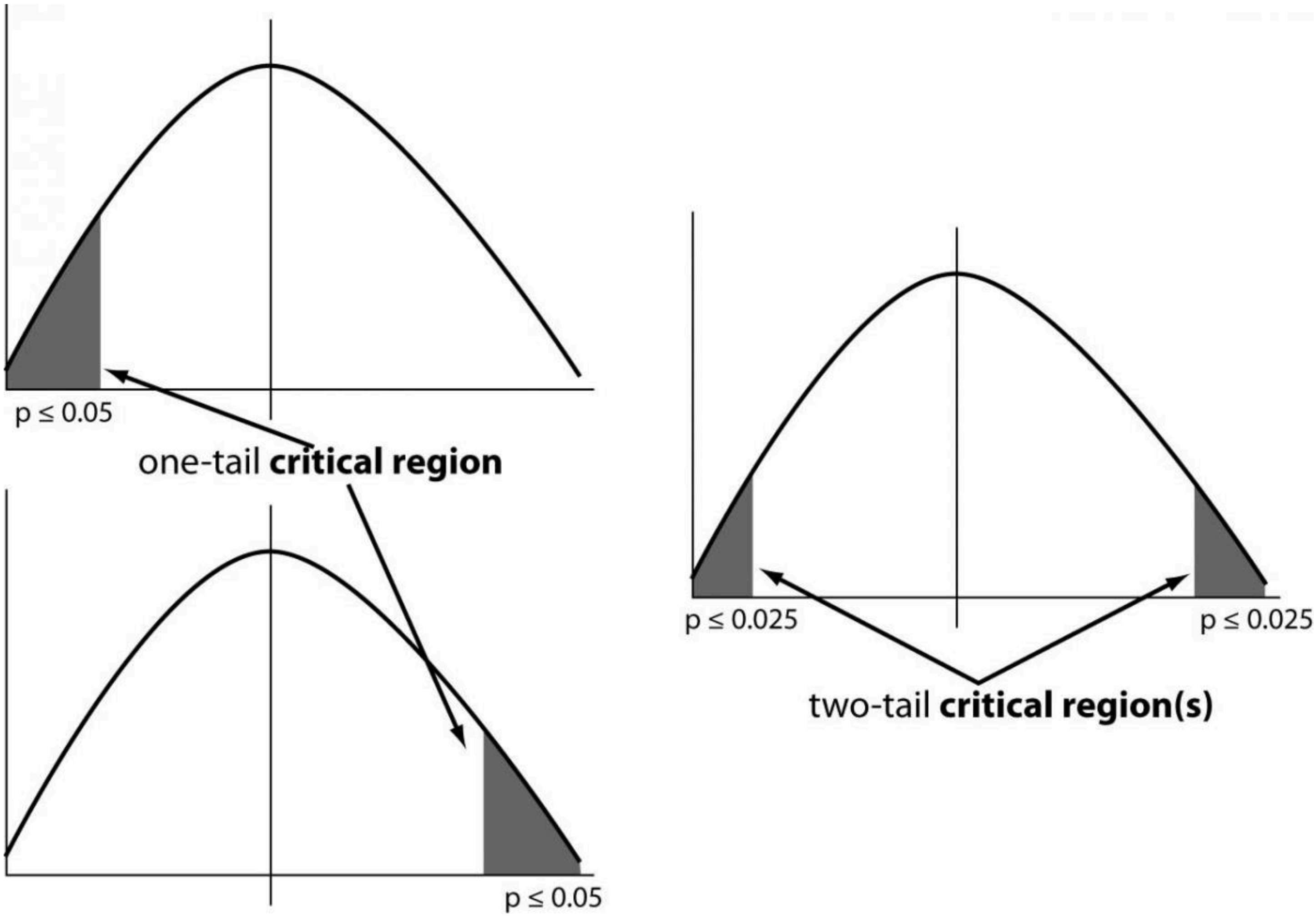
$$p = \int_Z^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 1 - \Phi(Z)$$

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

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

Beware of 1 vs 2-sided definitions!

1 sided vs 2 sided



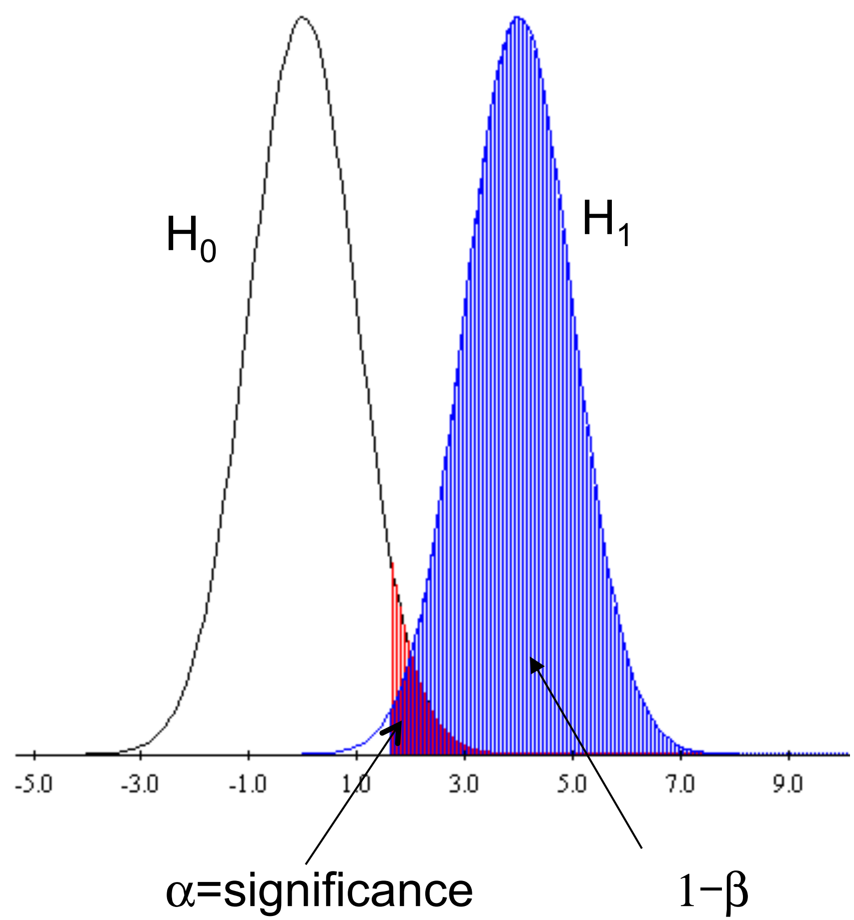
To determine a 1 sided 95% CL,
we sometimes need to set the critical region to 10% 2 sided

Basic Definitions: type I-II errors

- By defining α you determine your tolerance towards mistakes… (accepted mistakes frequency)
- **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

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

- The pdf of q….

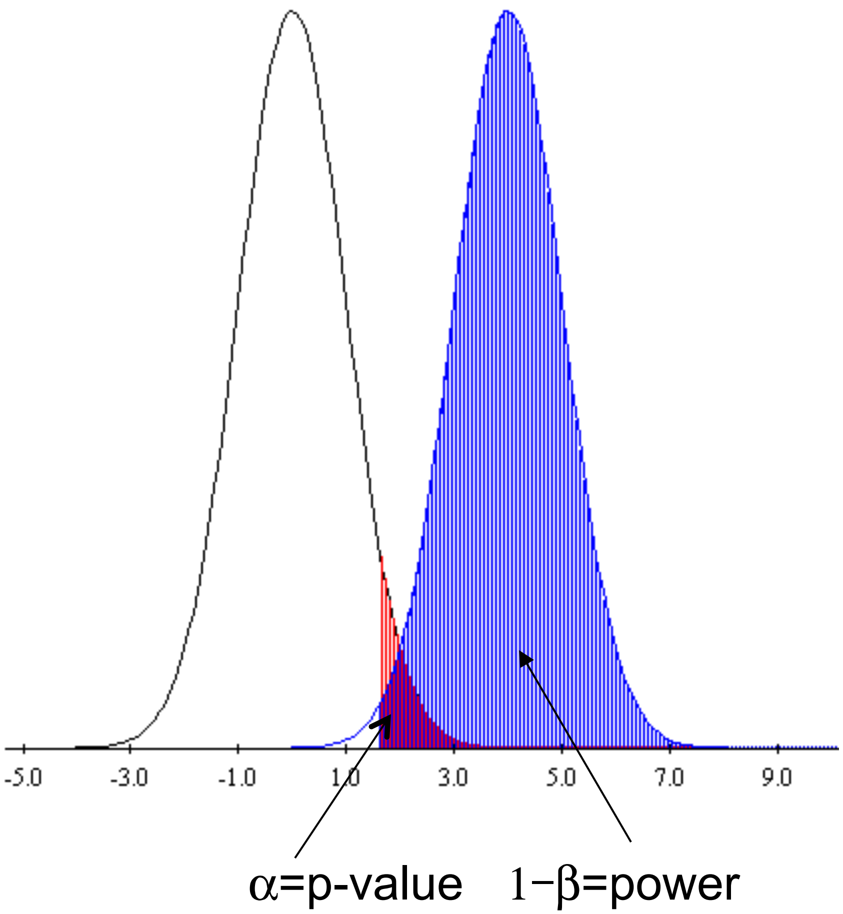


Basic Definitions: POWER

- $\alpha = \Pr ob(reject\ H_0 \mid H_0)$
- 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

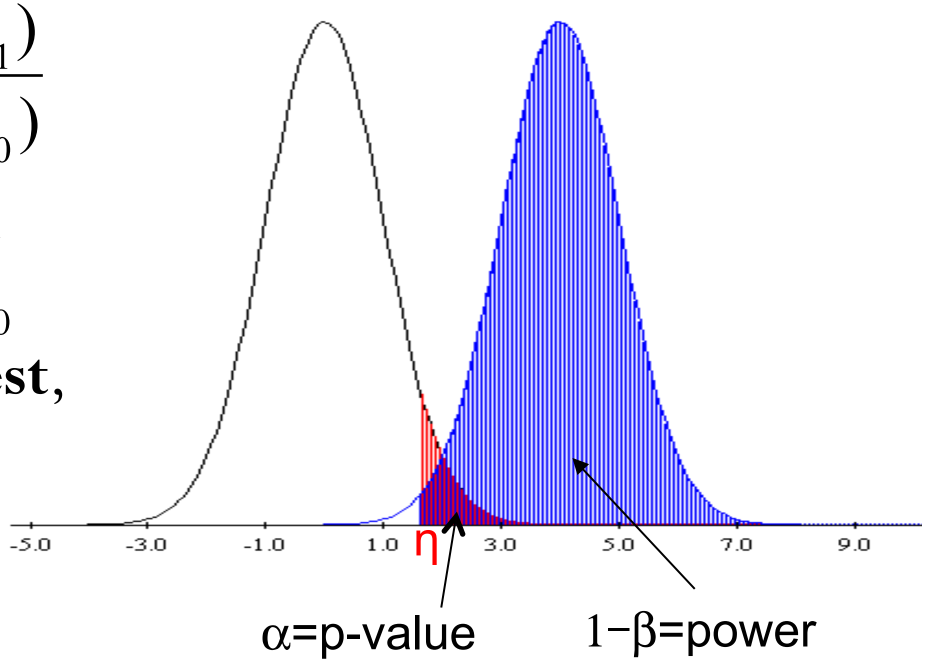
Which Analysis is Better

- 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



Likelihood

- Likelihood is a function of the data

$$L(H) = L(H|x) = f(x)$$

$$L(H|x) = P(x|H)$$

$$\lambda(x) = \frac{L(H_1|x)}{L(H_0|x)}$$

Bayes Theorem

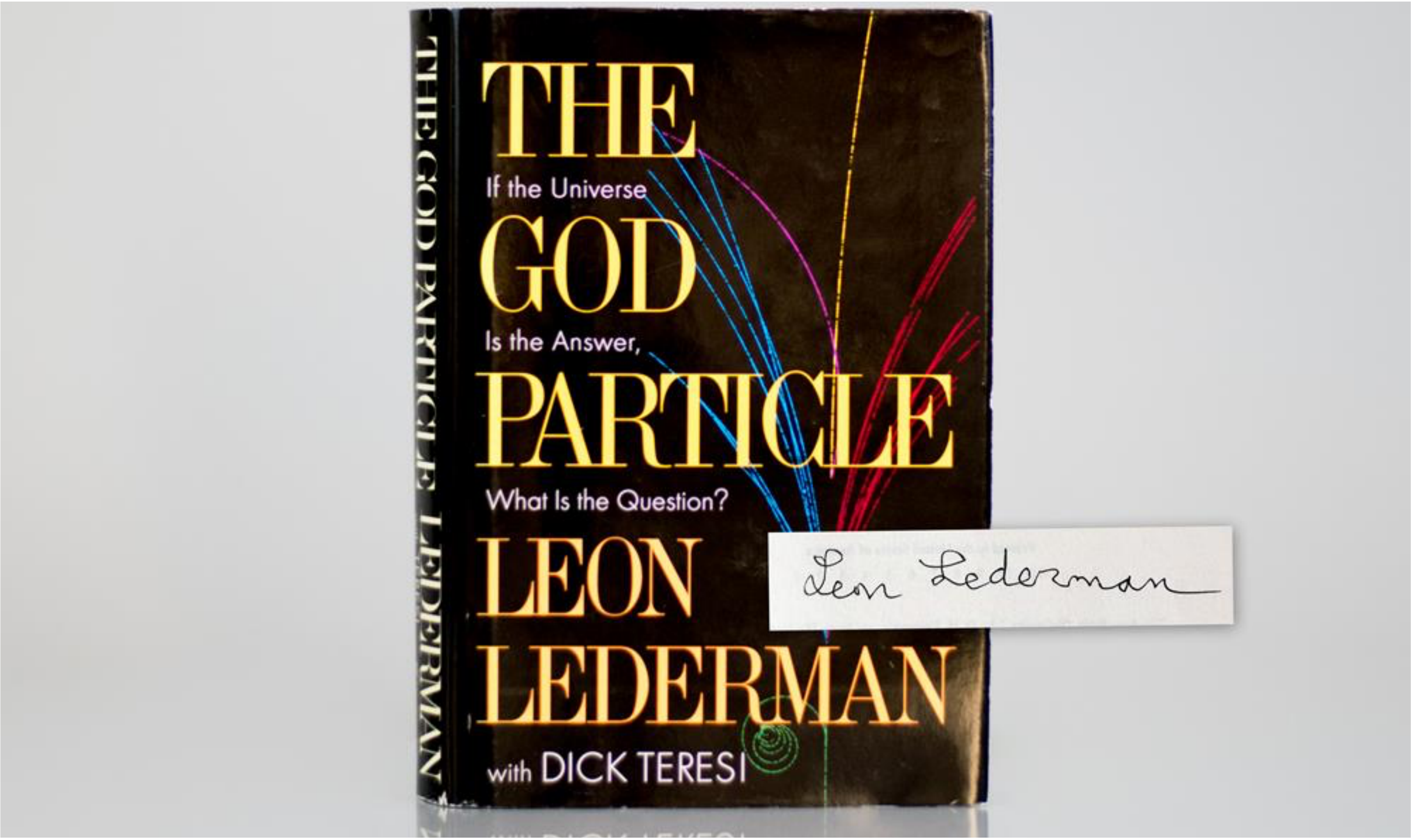
- Likelihood is not the probability of the hypothesis given the data

$$P(H|x) = \frac{P(x|H) \cdot P(H)}{\Sigma_H P(x|H) P(H)}$$

$$P(H|x) \approx P(x|H) \cdot P(H)$$

Prior

If the Universe is the answer, what is the question?

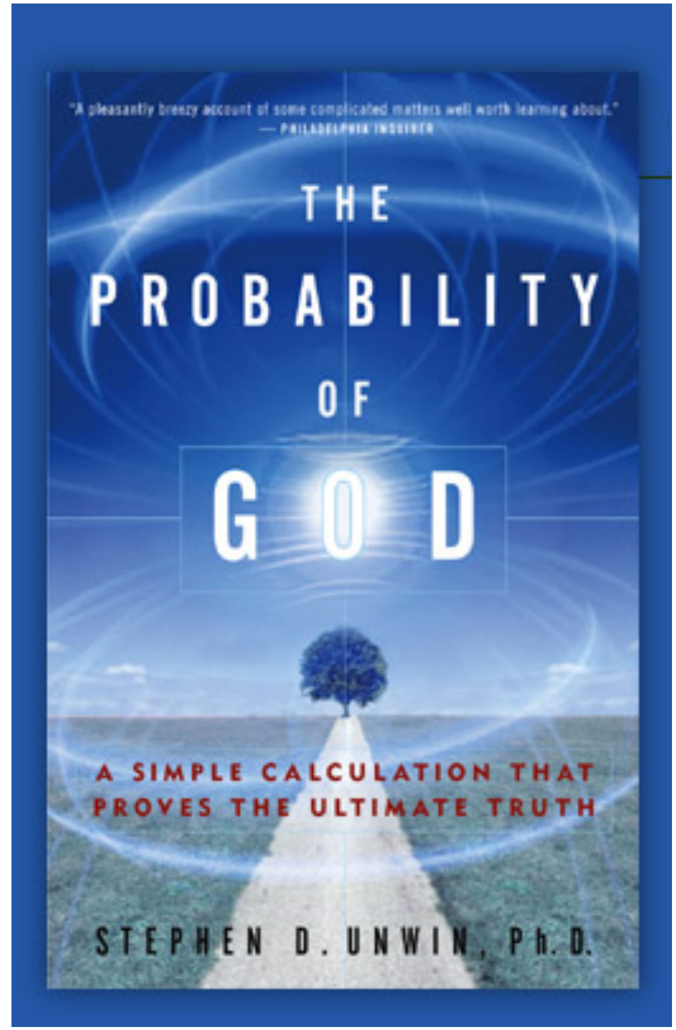


What is the Right Answer?

- 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%…



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…

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

- Make sure that when you quote your answer you also quote your prior assumption!
- 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!

$$L(Higgs(m_H)) = P(Data \mid Higgs)$$

Priors

$$P(\theta|data) \sim \int L(data|\theta,\lambda)\pi(\lambda)d\theta d\lambda$$

- 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 the parameter, 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 usually come from subsidiary measurements of the response of the detector to the photon energy, for example.
      - One should try to get priors by subsidiary measurements
      - Some priors come from subjective assumptions (theoretical, prejudice symmetries….) of our model

Priors – Uninformative Priors

- **Uninformative Priors**: All priors on the parameter of interest should be uninformative….
  IS THAT SO?
  Therefore flat uninformative priors are most common in HEP.
  - When taking a uniform prior for the Higgs mass [115,250]… 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 g^2
  Depends on the metric!
  Moreover, flat priors are improper and lead to serious problems of undercoverage (when one deals with >1 channel, i.e. beyond counting, one should AVOID them

Frequentist vs Bayesian

- The Bayesian infers from the data using priors

posterior $P(H \mid x) \approx P(x \mid H) \cdot P(H)$

- Priors is a science on its own.
  Are they objective? Are they subjective?
- The Frequentist calculates the probability of an hypothesis to be inferred from the data based on a large set of hypothetical experiments
  Ideally, the frequentist does not need priors, or any degree of belief while the Baseian posterior based inference **is** a "Degree of Belief".

- However, NPs inject a Bayesian flavour to any Frequentist analysis

Confidence Interval and Confidence Level (CL)

CL & CI - Wikipedia

$$\mu = 1.1 \pm 0.3$$

$$\mu = [0.8, 1.4] \ @ \ 68\% \ CL$$

CI=[0.8,1.4]

what does it mean?

- 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.
- 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 distribution function $p(\mu_{meas} | \mu)$
- Given the measurement you deduce somehow (based on your statistical model) that there is a 95% Confidence interval $[\mu_1, \mu_2]$.

(it is 95% likely that the μ_{true} is in the quoted interval)

The correct statement:

- In an ensemble of experiments 95% of the obtained confidence intervals will contain the true value of μ.

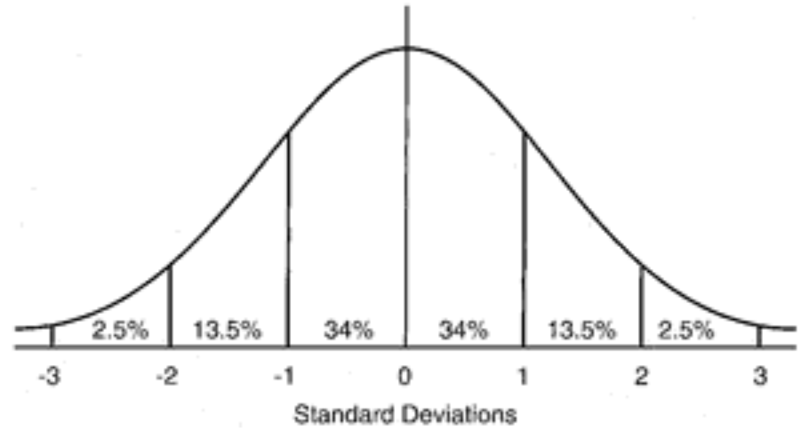
Upper limit

- 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 $\mu_{up} < 1$ → $\sigma(m_H) < \sigma_{SM}(m_H)$ → a SM Higgs with a mass m_H is excluded at the 95% CL

Confidence Interval & Coverage

- Confidence Level: A CL of (e.g.) 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 our estimated Confidence Interval contains the true value of μ more than 95% of the cases (for every possible μ) we claim that our method overcover (being conservative)
- If in an ensemble of (MC) experiments the true value of μ is covered within the estimated confidence interval , we claim a coverage

How to deduce a CI?



- One can show that if the data is distributed normal around the average i.e. P(data | μ)=normal

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

then one can construct a 68% CI around the estimator of μ to be

$$\hat{x} \pm \sigma$$

However, not all distributions are normal, many distributions are even unknown and coverage might be a real issue

How to deduce a CI?

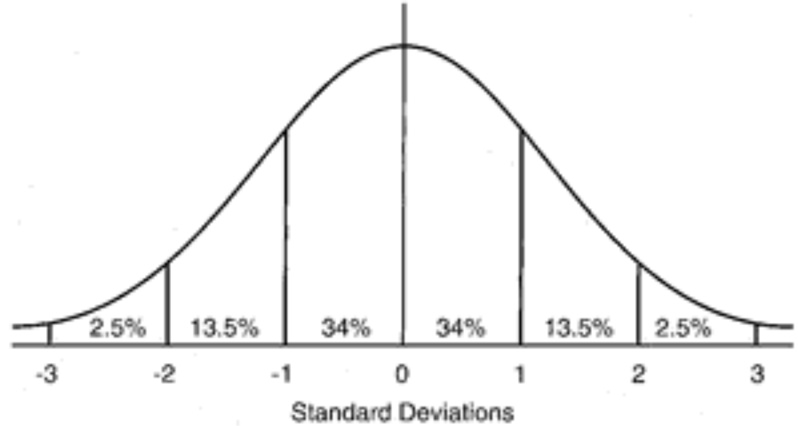
- One can show that if the data is distributed normal around the average i.e. P(data | μ)=normal

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

then one can construct a 68% CI around the estimator of μ to be

$$\hat{x} \pm \sigma$$

However, not all distributions are normal, many distributions are even unknown and coverage might be a real issue



- One may construct many 68% intervals…. $CI = [\mu_L, \mu_U]$

$$\int_{\mu_L}^{\mu_U} f(x \mid \hat{x})dx = 68\%$$

- Which one has a full coverage?
- How can we guarantee a coverage

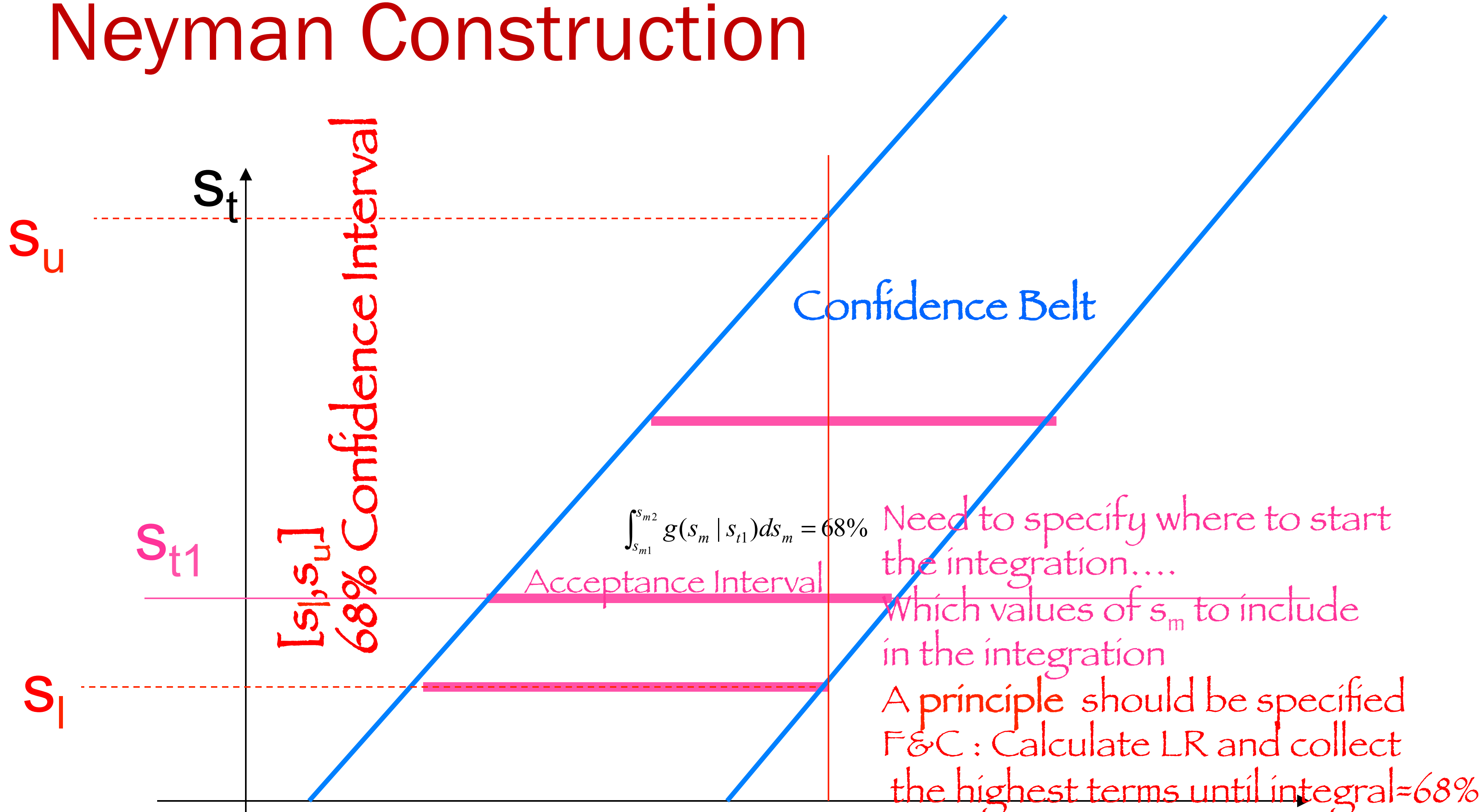
- The QUESTION is NOT how to construct a CI, it is
- HOW TO CONSTRUCT A CI WHICH HAS A COVERAGE @ THE 68% CL

The Frequentist Game a ’la Feldman & Cousins

Or

How to ensure a Coverage

Neyman Construction



$[s_l, s_u]$ 68% Confidence Interval
In **68%** of the experiments the derived **C.I. contains** the **unknown true value of s**

- With Neyman Construction we guarantee a coverage via construction, i.e. for any value of the unknown true s, the Construction Confidence Interval will cover s with the correct rate.

The Flip Flop Way of an Experiment

- The most intuitive way to analyze the results of an experiment would be
  - Construct a test statistics
    e.g. $Q(x) \sim L(x|H_1)/L(x|H_0)$
  - If the significance of the measured $Q(x_{obs})$, is less than 3 sigma, derive an upper limit (just looking at tables), if the result is >5 sigma (and some minimum number of events is observed….), derive a discovery central confidence interval for the measured parameter (cross section, mass….) …..
- **This Flip Flopping policy leads to undercoverage:**
  Is that really a problem for Physicists?
  Some physicists say, for each experiment quote always two results, an upper limit, and a (central?) discovery confidence interval

Frequentist Paradise – F&C Unified with Full Coverage

- Frequentist Paradise is certainly made up of an interpretation by constructing a confidence interval in brute force ensuring a coverage!
- This is the Neyman confidence interval adopted by F&C….

- The motivation:
  - Ensures Coverage
  - Avoid Flip-Flopping – an ordering rule determines the nature of the interval (1-sided or 2-sided depending on your observed data)
  - Ensures Physical Intervals

- Let the test statistics be

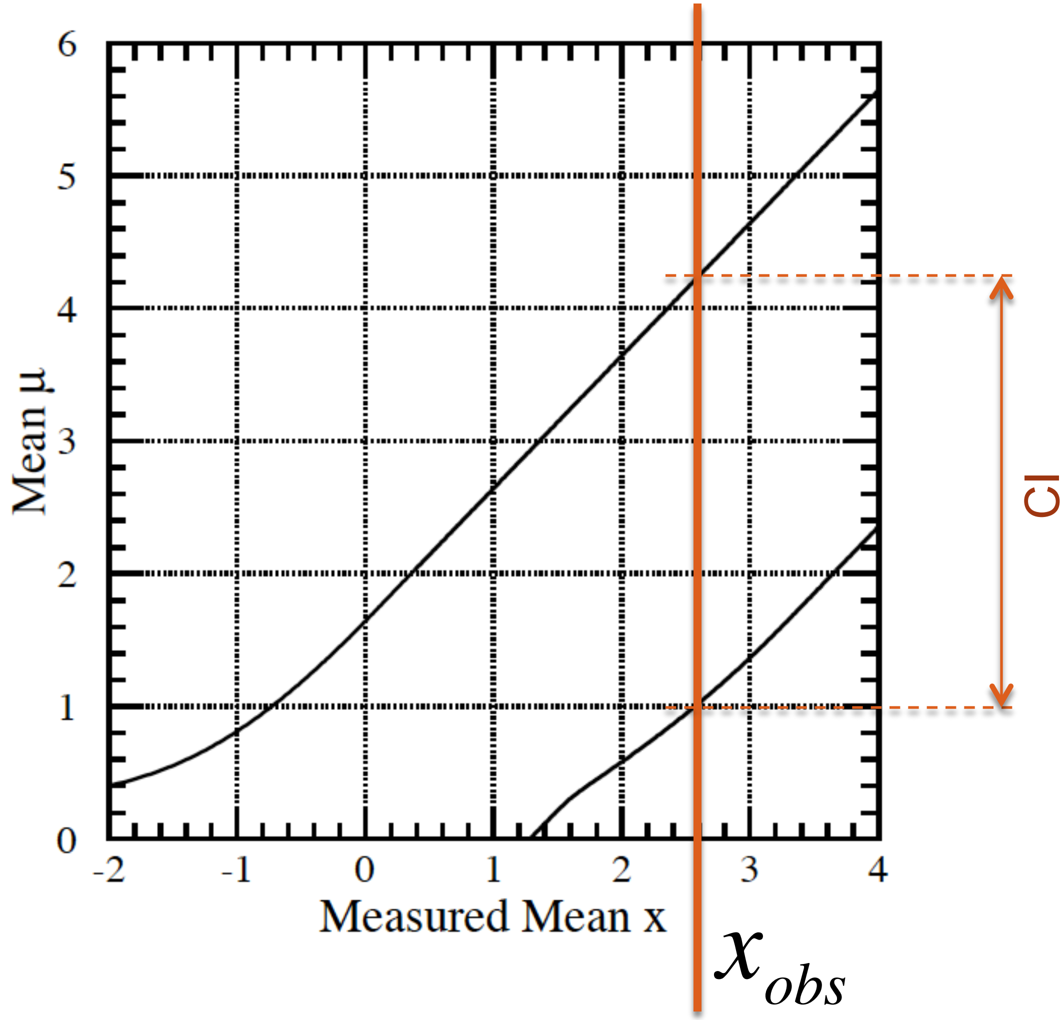
$$Q = \frac{L(s+b)}{L(\hat{s}+b)} = \frac{P(n \mid s+b)}{P(n \mid \hat{s}+b)}$$

where $\hat{s}$ is the
physically allowed mean s that maximizes L($\hat{s}$+b)
(**protect a downward fluctuation of the background, n_{obs}>b ;** $\hat{s}$>0)

- Order by taking the 68% highest Qs

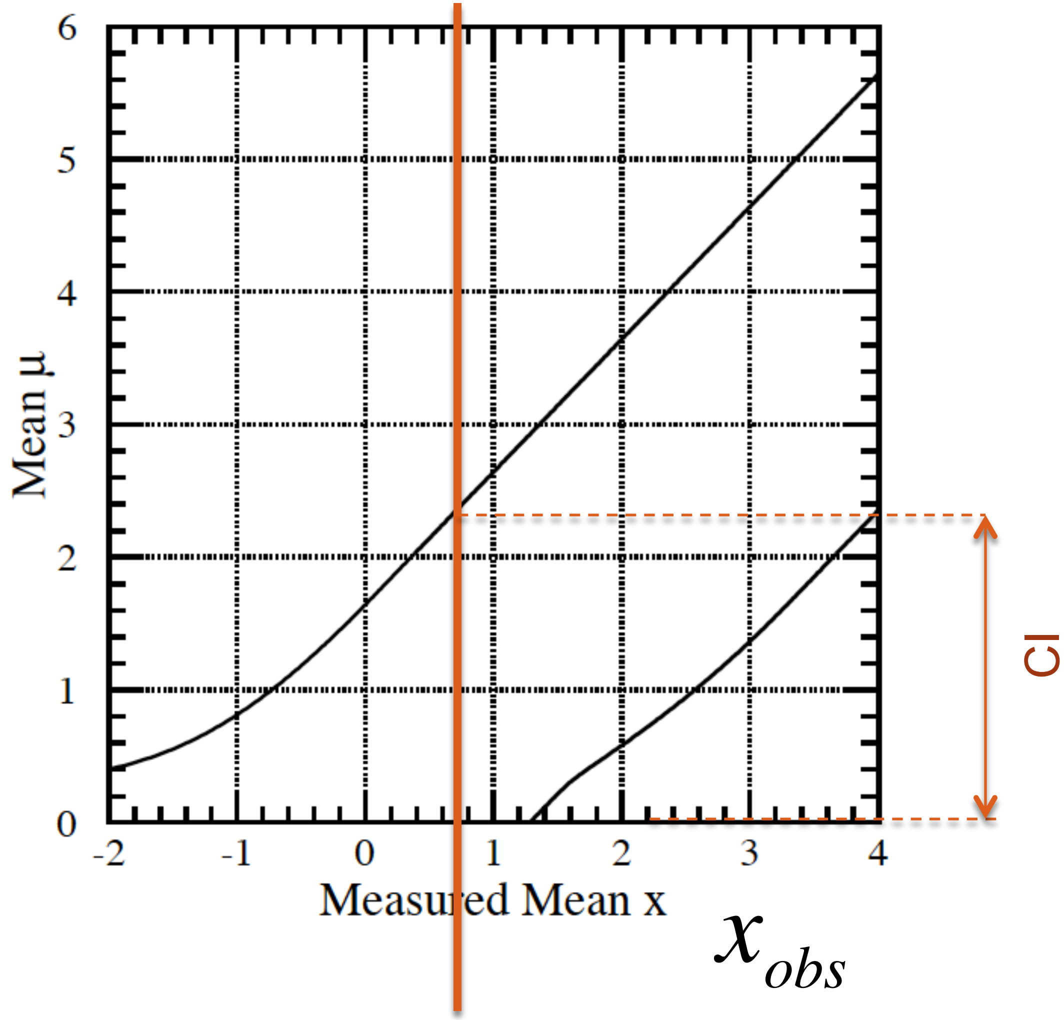
How to tell an Upper limit from a Measurement without Flip Flopping

- A measurement (2 sided)



How to tell an Upper limit from a Measurement without Flip Flopping

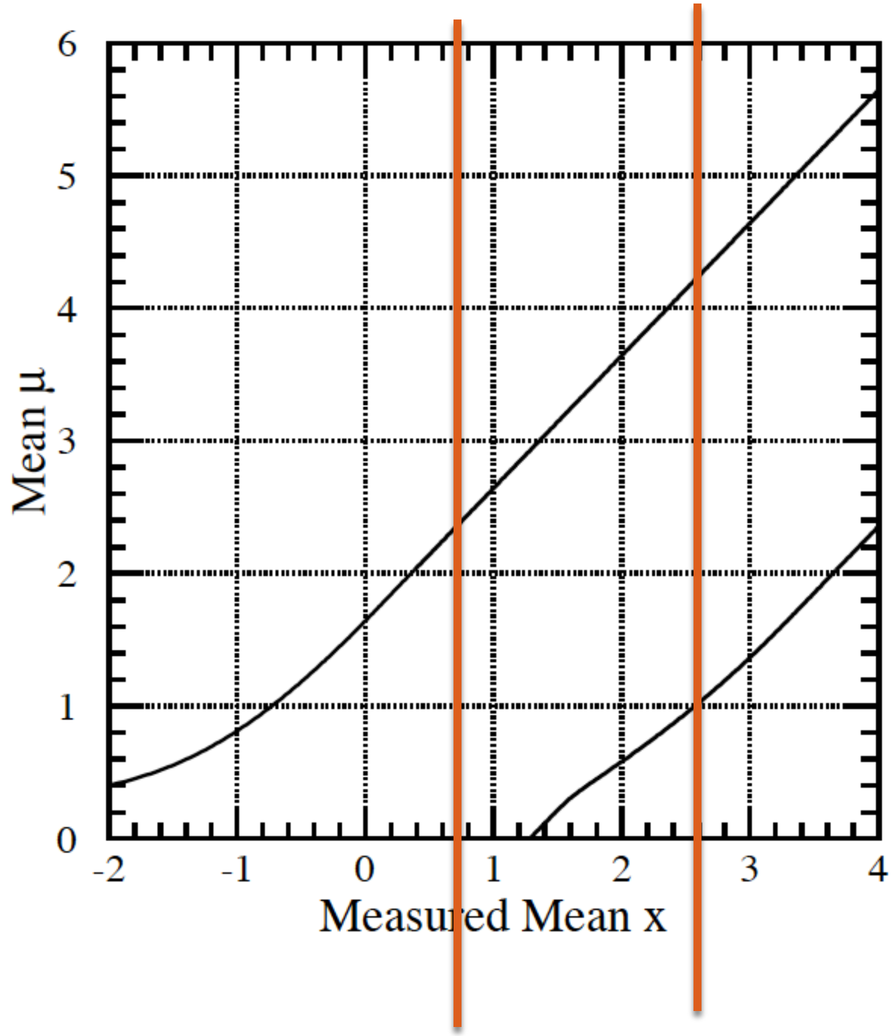
- An upper limit (1 sided)



How to tell an Upper limit from a Measurement without Flip Flopping

- Your observed result will tell you if it's a measurement or an upper limit

- But how to deal with systematics?



Search and Discovery Statistics in HEP
Lecture 2: PL, Asymptotic Distributions Exclusion & CLs

Eilam Gross, Weizmann Institute of Science

This presentation would have not been possible without the tremendous help of the following people throughout many years

Louis Lyons, Alex Read, Glen Cowan ,Kyle Cranmer ,
Ofer Vitells & Bob Cousins

The Profile Likelihood

The choice of the LHC for hypothesis inference in Higgs search

$$n = \mu s + b$$

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

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

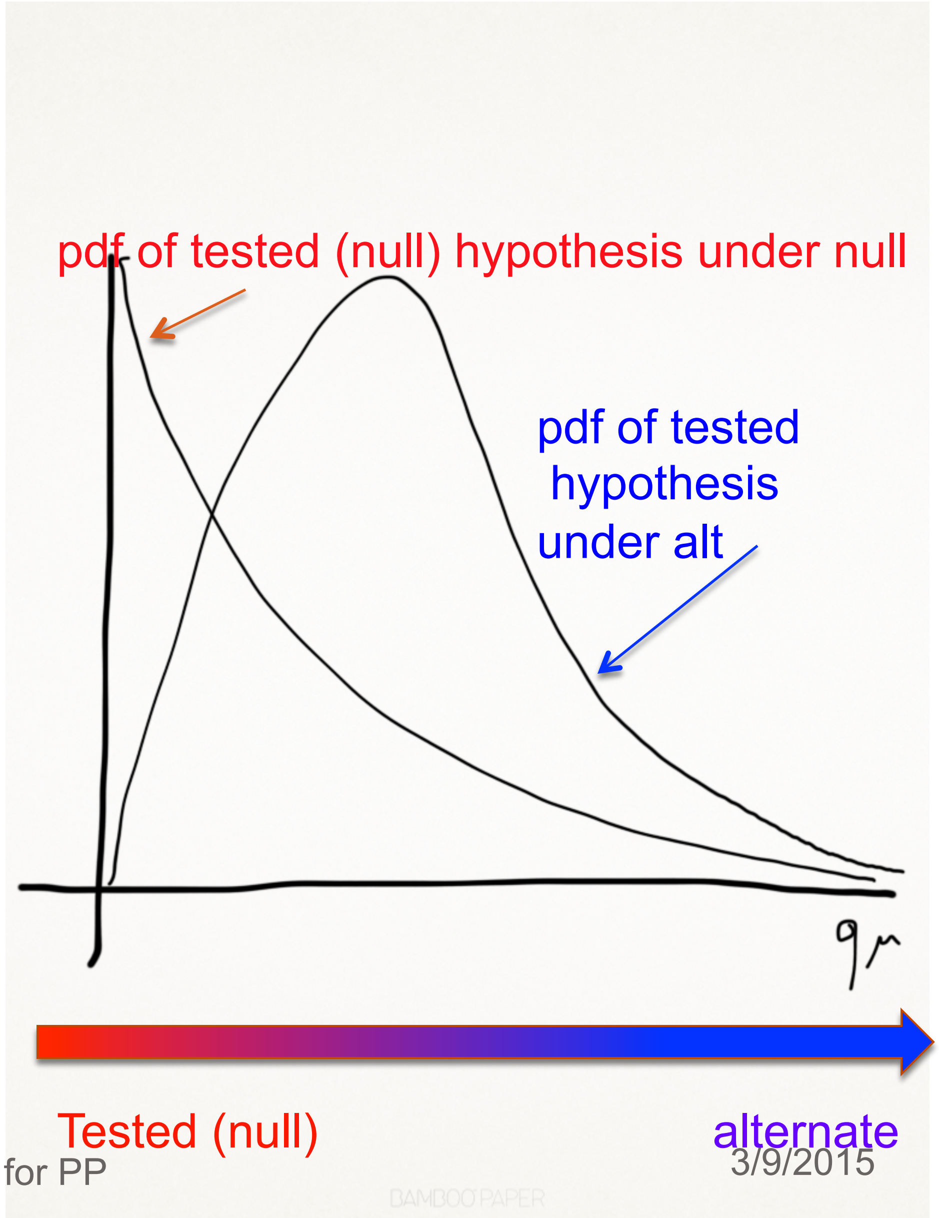
$\hat{\mu} \sim 1,\ q$ large

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

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

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

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



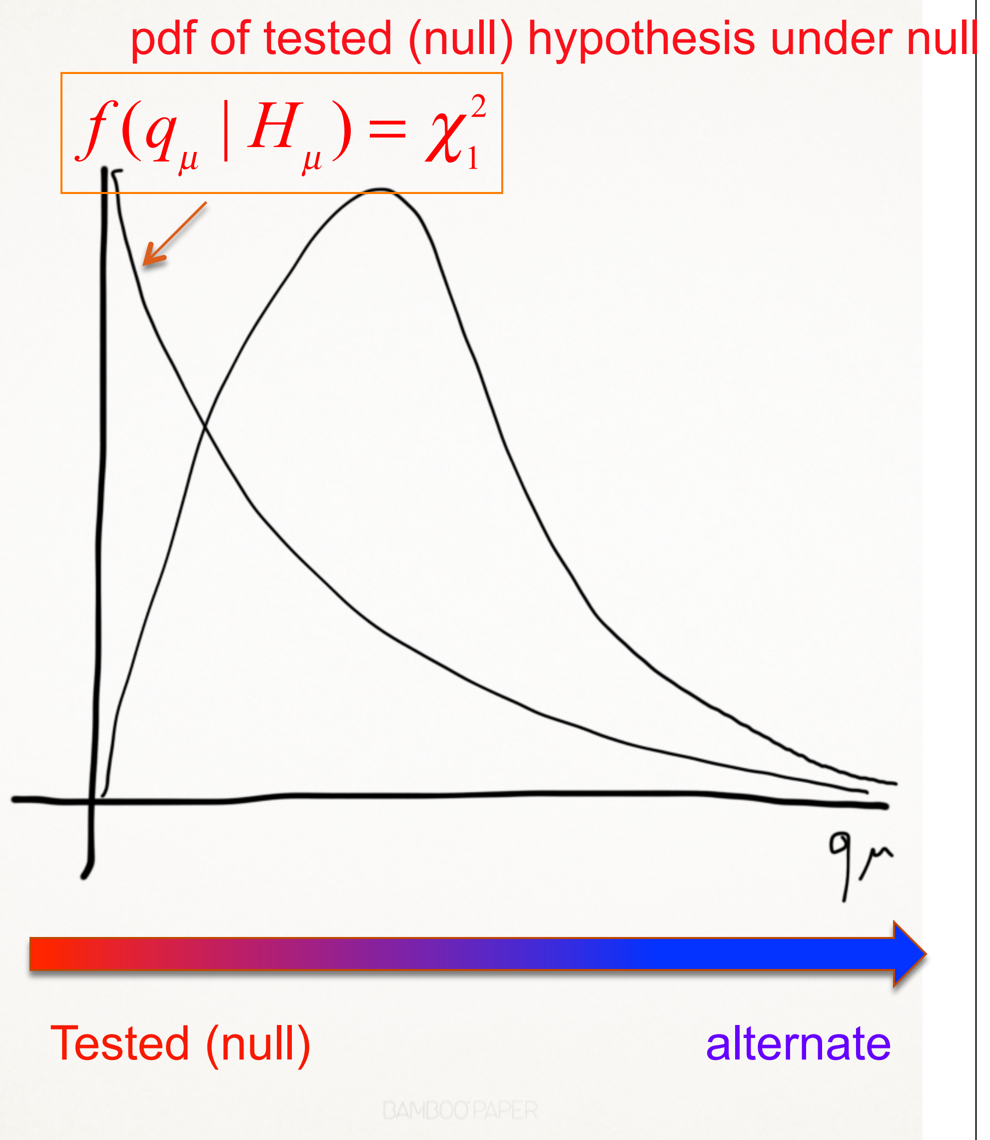
Wilks Theorem

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

$$f(q_0 \mid H_0) = \chi_1^2$$

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



Nuisance Parameters

or Systematics

Nuisance Parameters (Systematics)

- There are two kinds of parameters:
  - Parameters of interest (signal strength… cross section… μ)
  - Nuisance parameters (background (b), signal efficiency, resolution, energy scale,…)
- 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 (Bayesian) or profiling (Frequentist)
  - One can also use a frequentist test statistics (PL) while treating the NPs via marginalization (Hybrid, Cousins & Highland way)
- Marginalization (Integrating))
  - Integrate the Likelihood, L, over possible values of nuisance parameters (weighted by their prior belief functions -- Gaussian,gamma, others…)
  - Consistent Bayesian interpretation of uncertainty on nuisance parameters

Integrating Out The Nuisance Parameters (Marginalization)

- 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

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

Credible Interval $CI = [0, \mu_{up}]$

$$0.95 = \int_0^{\mu_{up}} p(\mu \mid x)\,d\mu$$

Nuisance Parameters (Systematisc)

- Neyman Pearson (NP) Likelihood Ratio:

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

- Either Integrate the Nuisance parameters

prior

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

Cousins & Highland

- Or profile them

$$q^{NP} = -2\ln\frac{L\left(b(\hat{\hat{\theta}}_0)\right)}{L\left(s+b(\hat{\hat{\theta}}_1)\right)}$$

$$\hat{\hat{\theta}}_0 = MLE_{\mu=0}\ of\ L(b(\theta))$$

$$\hat{\hat{\theta}}_1 = MLE_{\mu=1}\ of\ L(s+b(\theta))$$

Nuisance Parameters and Subsidiary Measurements

- Usually the nuisance parameters are auxiliary parameters and their values are constrained by auxiliary measurements
- Example

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

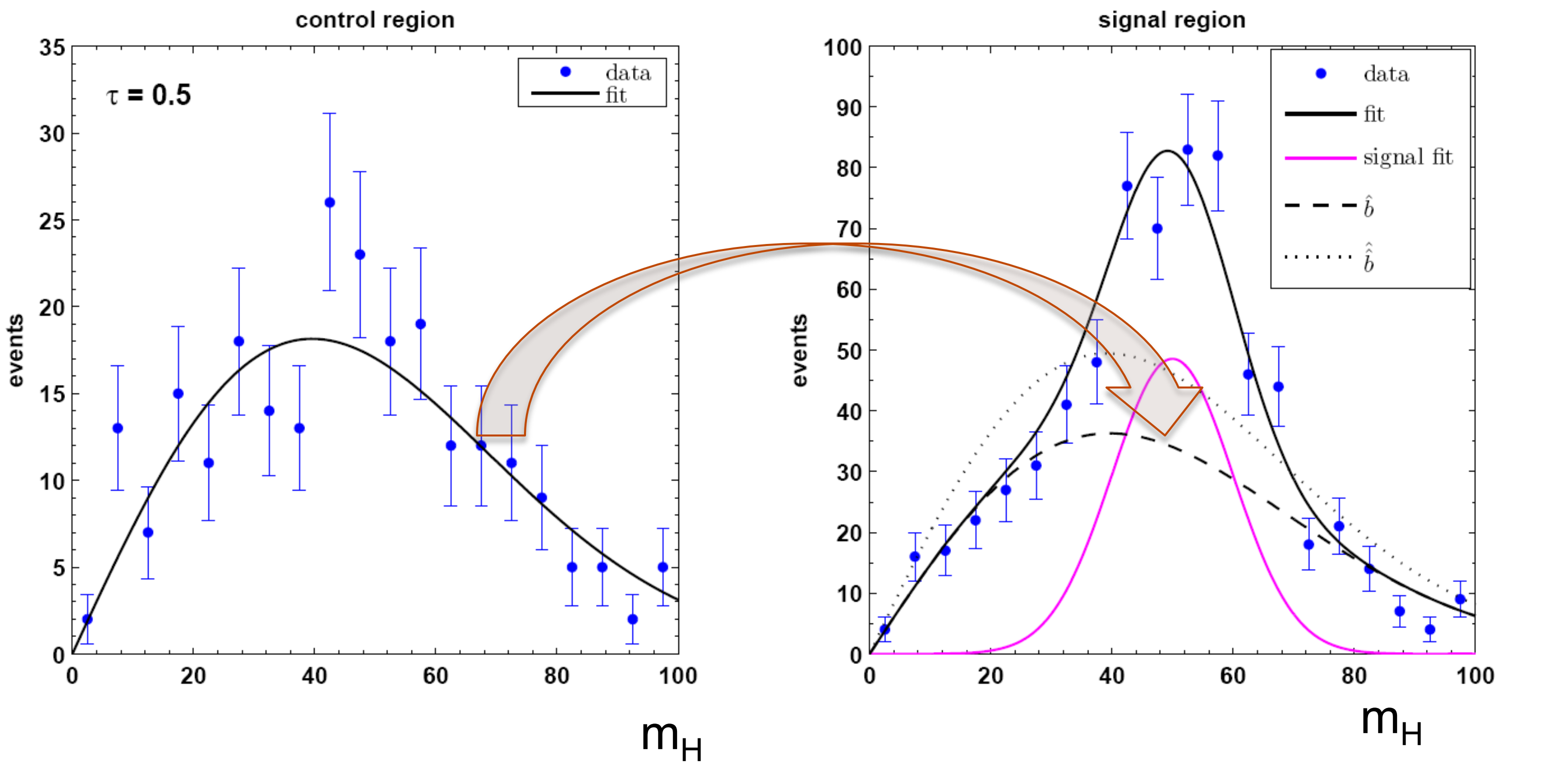
$$m = \tau b$$

$$L\big(\mu \cdot s + b(\theta)\big) = Poisson\big(n; \mu \cdot s + b(\theta)\big) \cdot Poisson\big(m; \tau b(\theta)\big)$$

Mass shape as a discriminator

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

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



Profile Likelihood with Nuisance Parameters

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

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

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

$\hat{\mu}$ 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

Wilks theorem in the presence of NPs

- Given n parameters of interest and any number of NPs, then

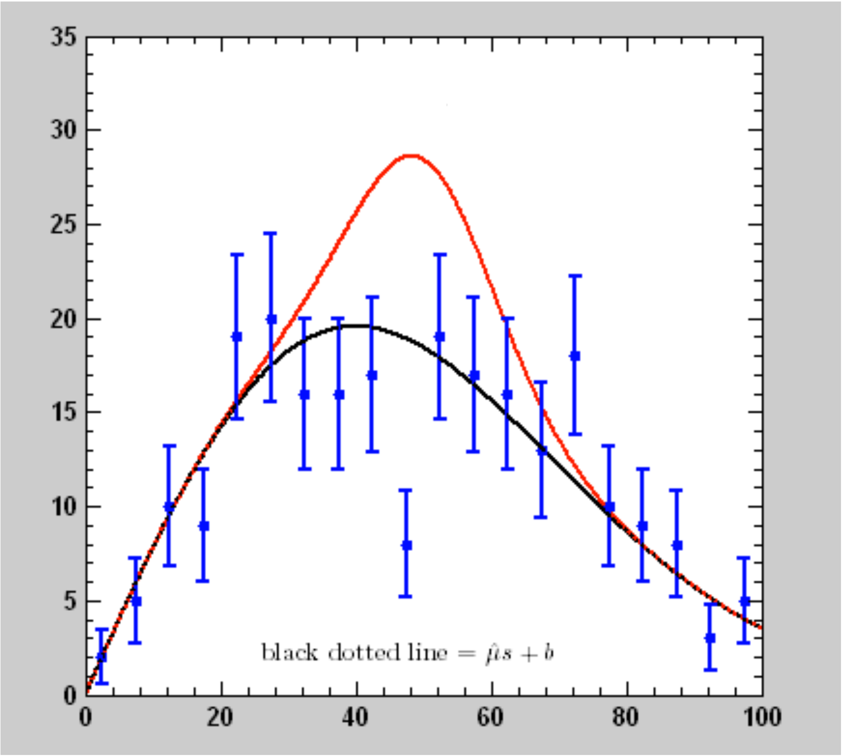
$$\lambda(\alpha_i) = \frac{L(\alpha_i, \hat{\hat{\theta}}_j)}{L(\hat{\alpha}_i, \hat{\theta}_j)}$$

$$q(\alpha_i) \equiv -2\log\lambda(\alpha_i) \sim \chi_n^2$$

Tossing Toys

Understanding the Basic Concepts

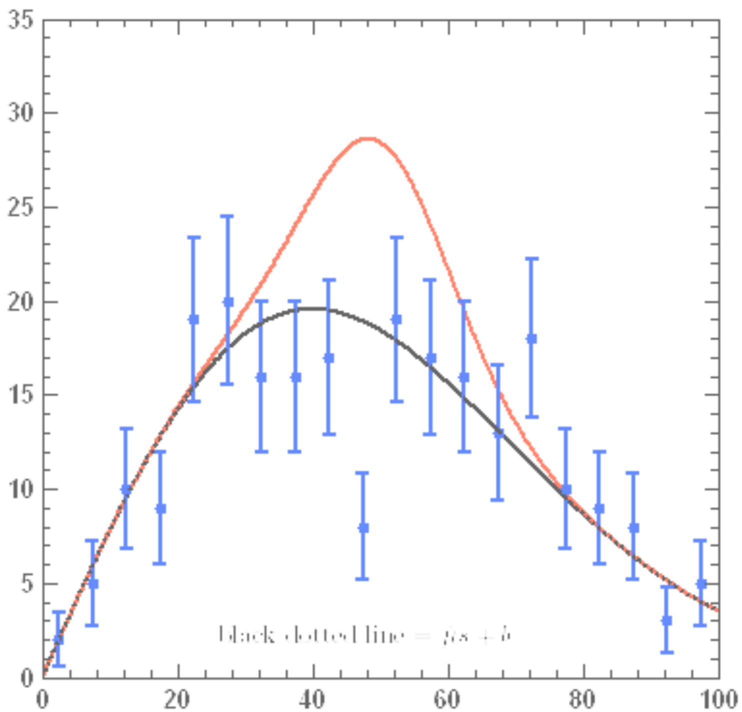
The Physics Model

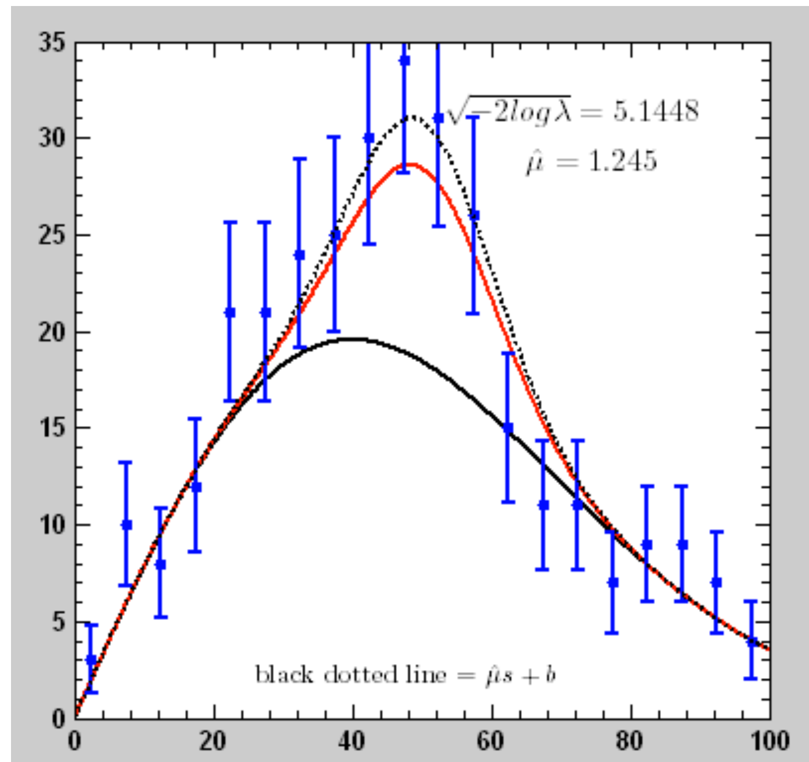
- SM without Higgs Background
  No signal $\langle n \rangle = b$
- 

mass

The Physics Model

- SM without Higgs Background Only $\langle n \rangle = b$
- SM with a Higgs Boson with a mass m_H $\langle n \rangle = s(m_H) + b$
-



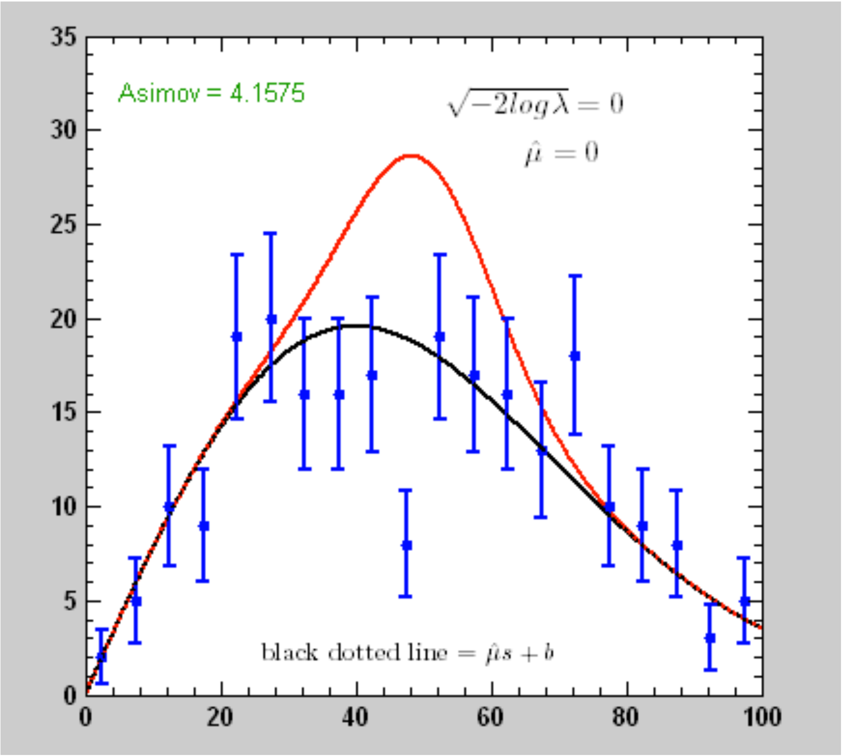


mass

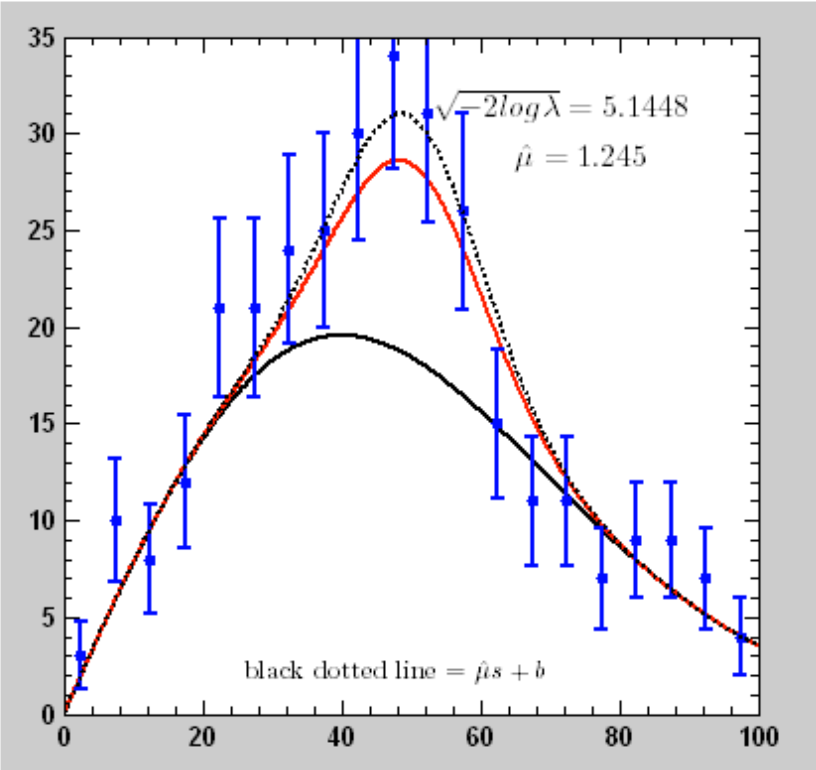
The Physics Model

$$n = \mu s + b \qquad MLE \quad \hat{\mu}$$

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



mass



mass

The Profile Likelihood ("PL")

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

For discovery we test the H_0 null hypothesis

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

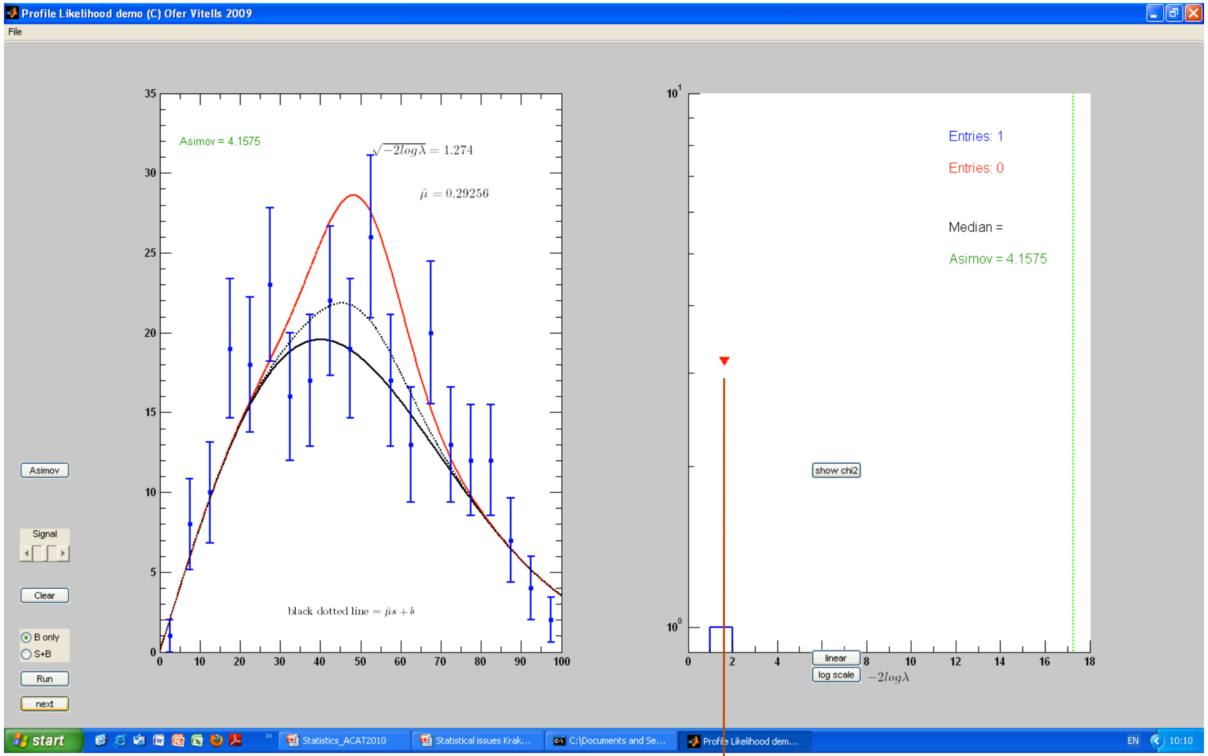
For

$$\hat{\mu} \sim 0,\ q_0 \text{ small}$$

$$\hat{\mu} \sim 1,\ q_0 \text{ large}$$

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

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



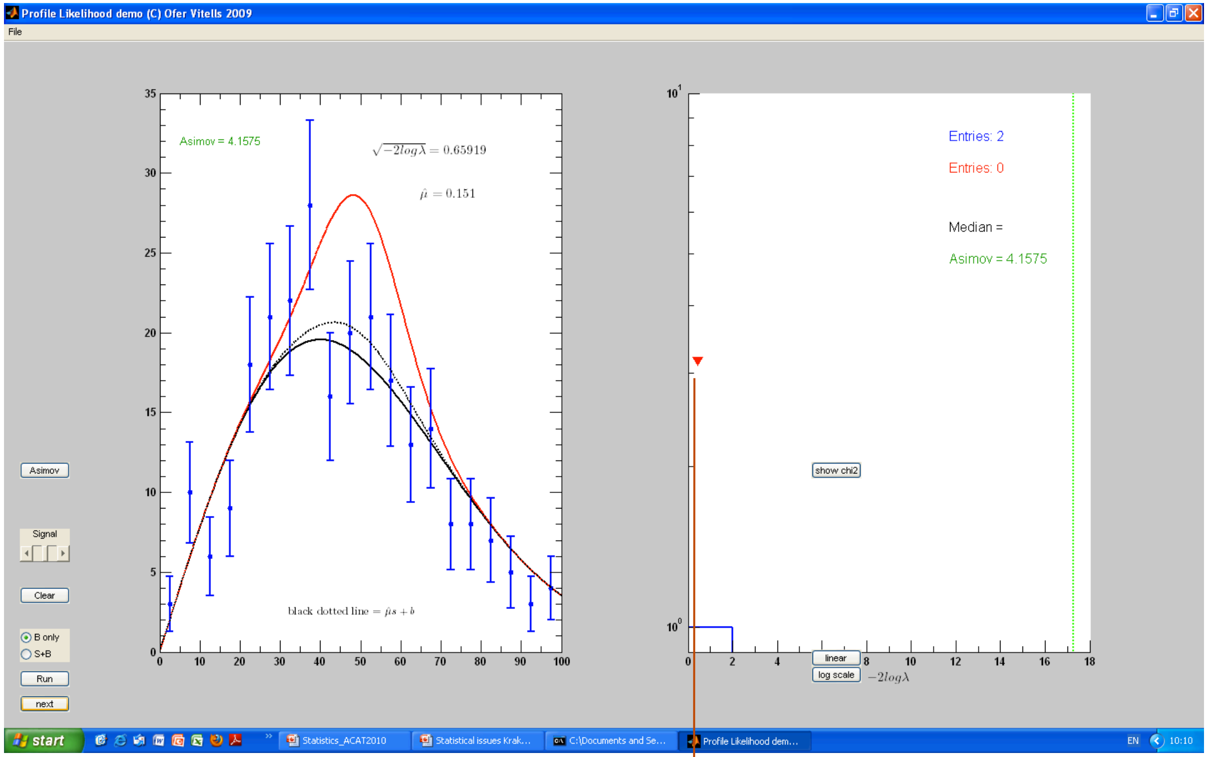
$$Z_{obs} = \sqrt{q_{0,obs}}$$

$$q_0 = 1.6 \rightarrow Z = \sqrt{1.6} = 1.27$$

PL: test t_0 under BG only ; $f(q_0|H_0)$

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

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



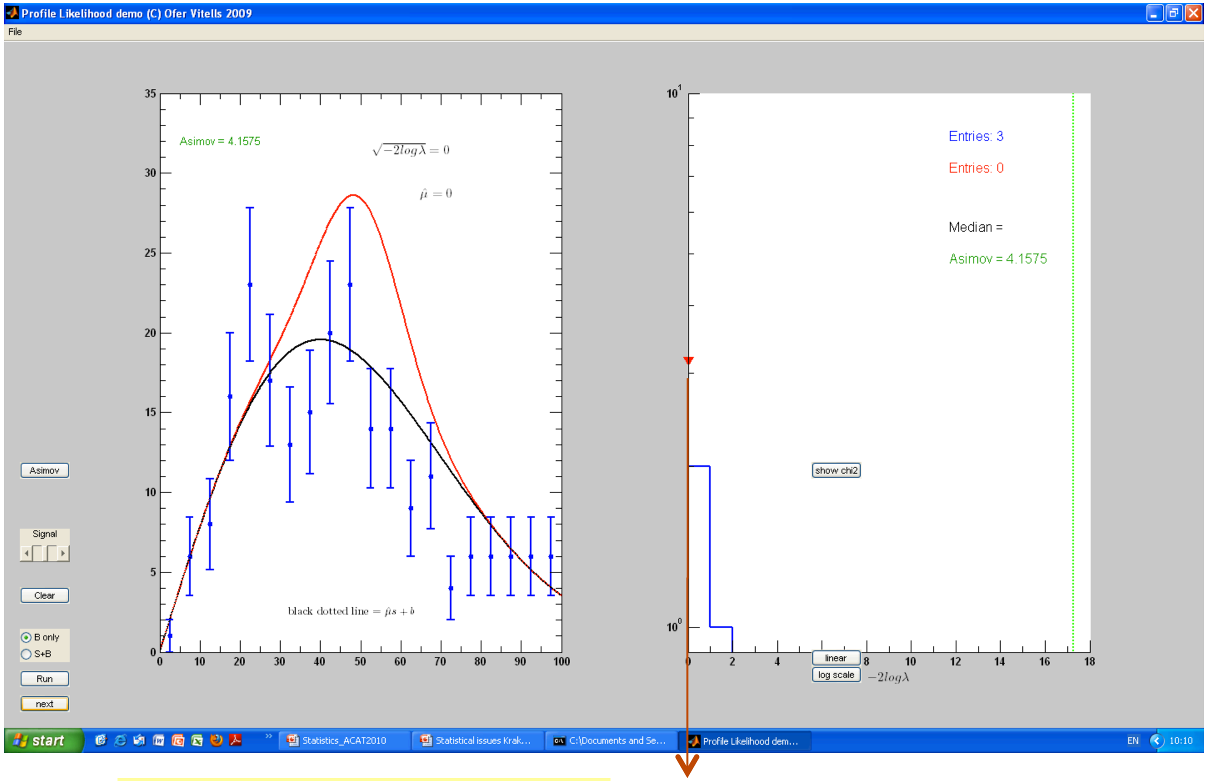
$$Z_{obs} = \sqrt{q_{0,obs}}$$

$$q_0 = 0.43 \rightarrow Z = 0.66\sigma$$

PL: test t_0 under BG only ; $f(q_0|H_0)$

$\hat{\mu}=0$

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



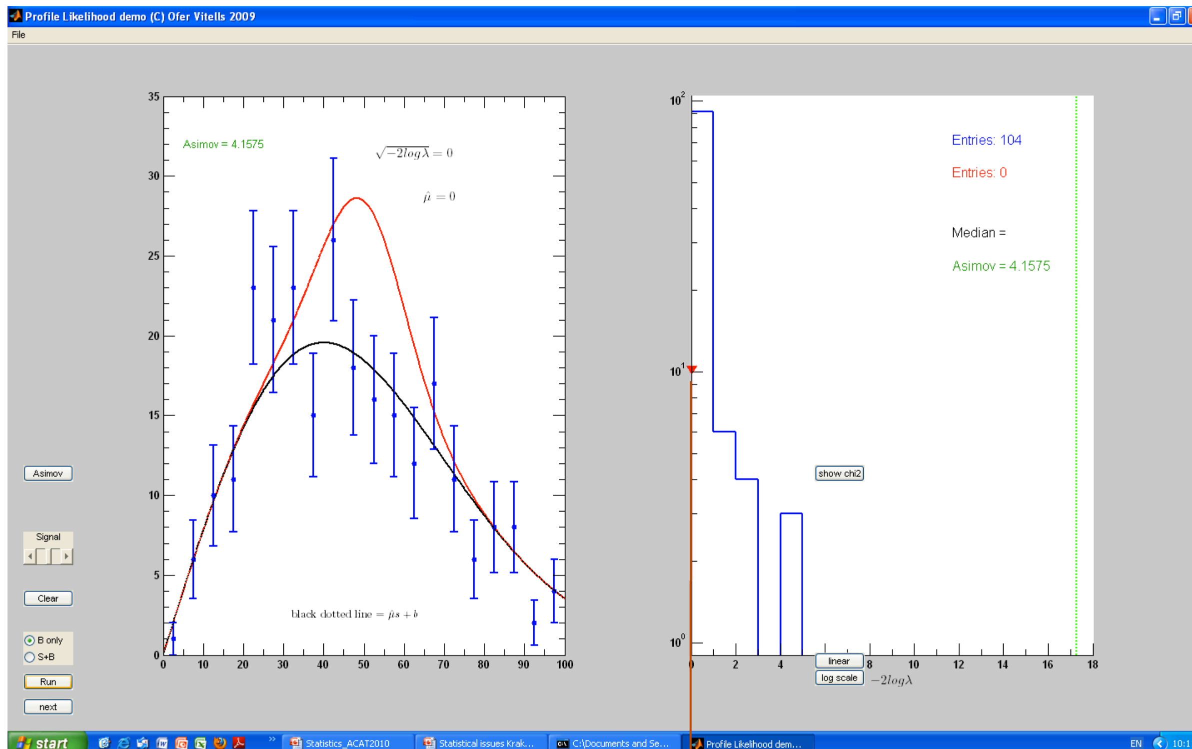
$$Z_{obs} = \sqrt{q_{0,obs}}$$

$q_0 = 0$

PL: test t_0 under BG only ; $f(q_0 | H_0)$

$\hat{\mu} = 0$

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



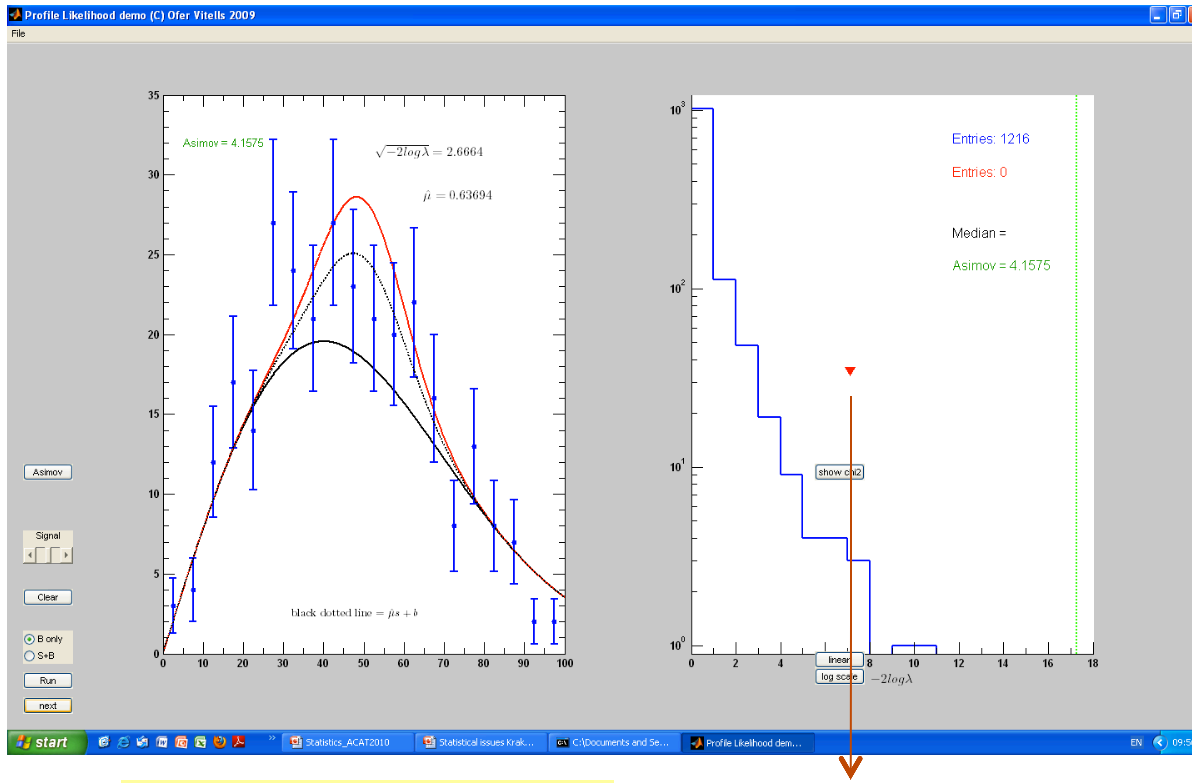
$$Z_{obs} = \sqrt{q_{0,obs}}$$

$q_0 = 0$

PL: test t_0 under BG only ; $f(q_0 | H_0)$

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

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



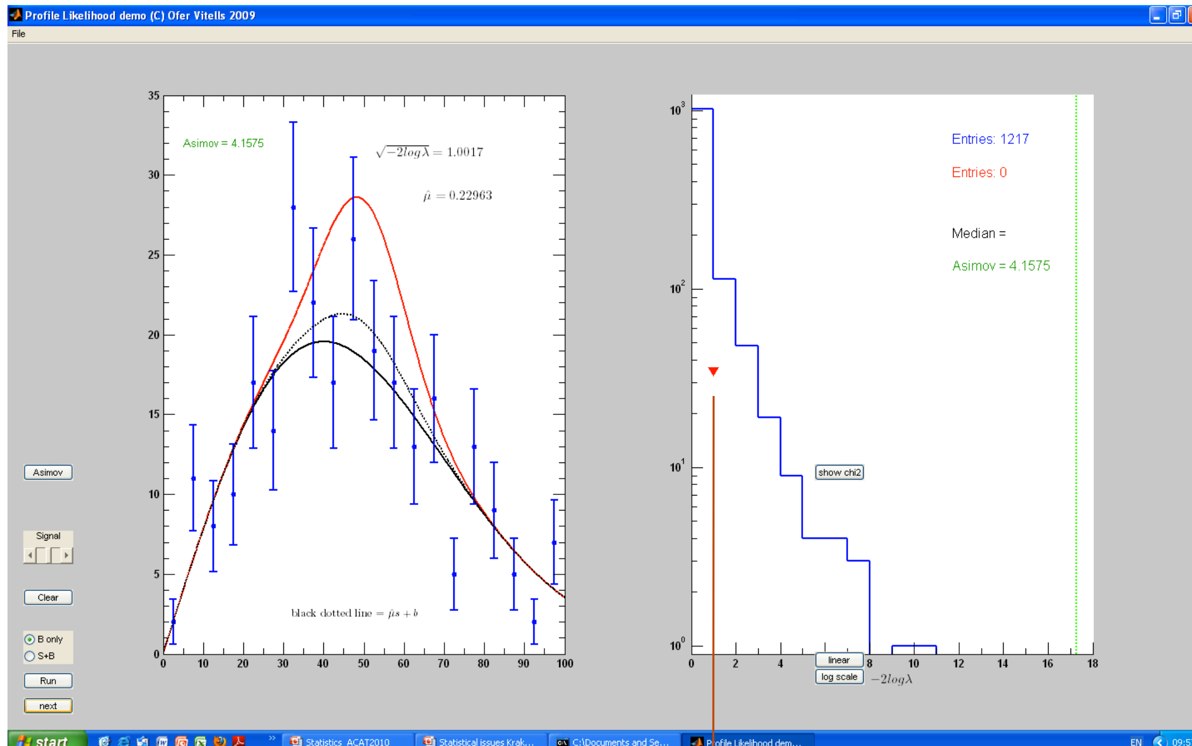
$$Z_{obs} = \sqrt{q_{0,obs}}$$

$$q_0 = 6.76 \rightarrow Z = 2.6\sigma$$

PL: test t_0 under BG only ; $f(q_0|H_0)$

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

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



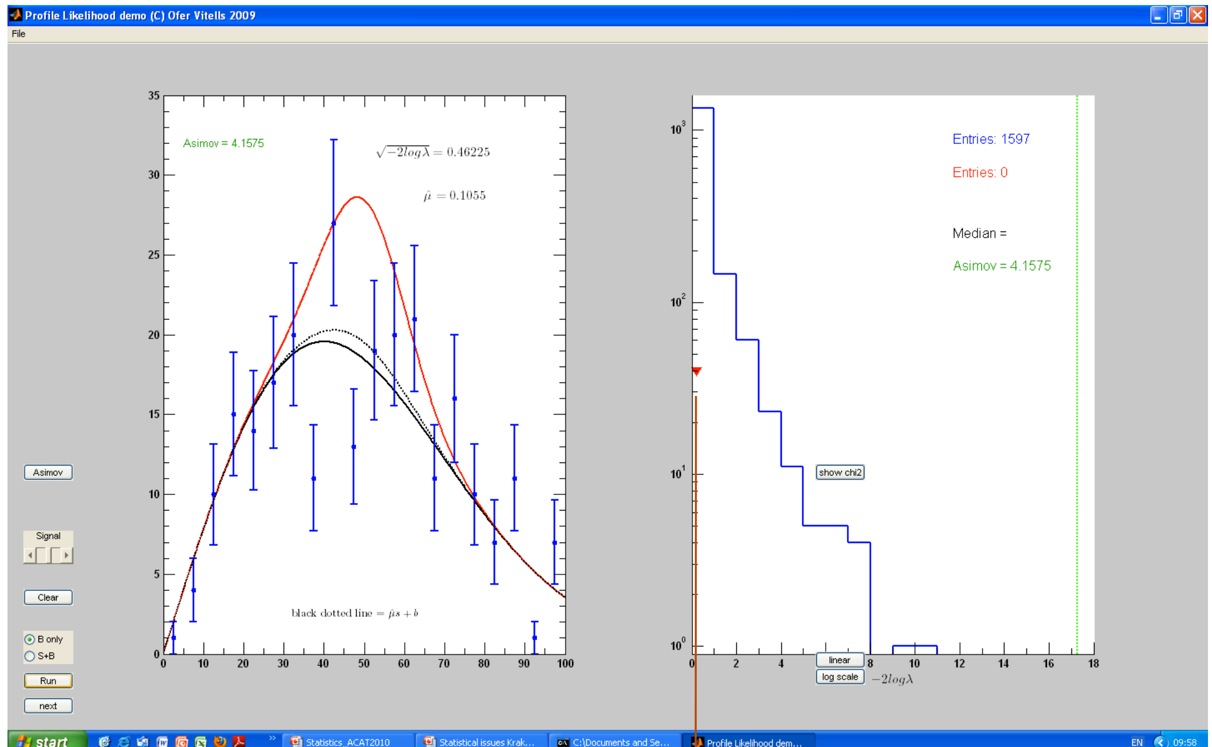
$$Z_{obs} = \sqrt{q_{0,obs}}$$

$$q_0 = 1.2 \rightarrow Z = 1.1\sigma$$

PL: test t_0 under BG only ; $f(q_0|H_0)$

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

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



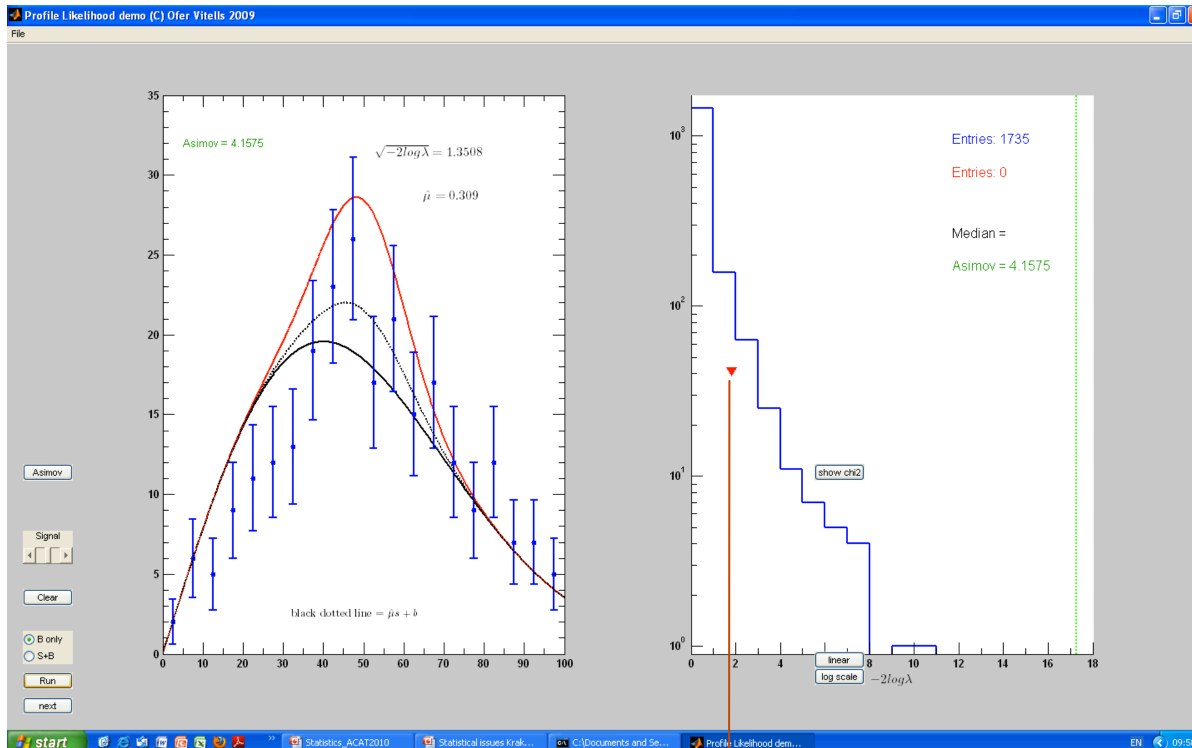
$$Z_{obs} = \sqrt{q_{0,obs}}$$

$$q_0 = 0.16 \rightarrow Z = 0.4\sigma$$

PL: test t_0 under BG only ; $f(q_0|H_0)$

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

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



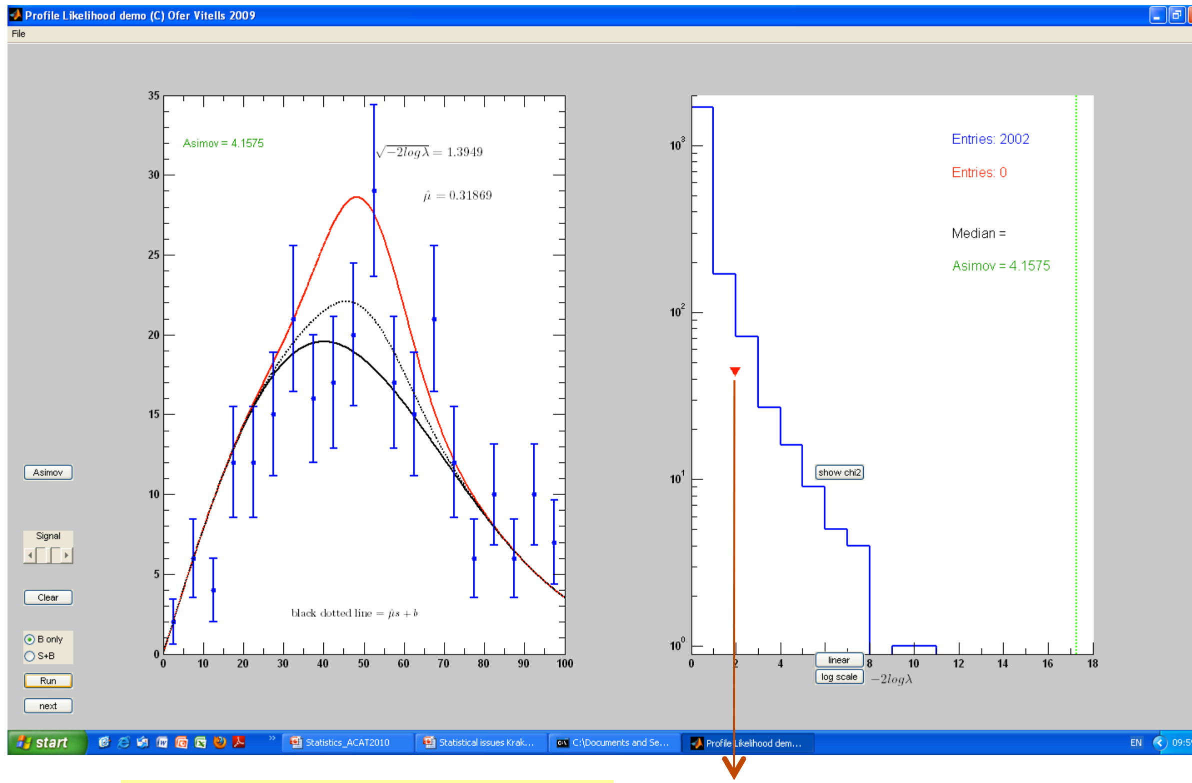
$$q_0 = 1.8 \rightarrow Z = 1.35\sigma$$

$$Z_{obs} = \sqrt{q_{0,obs}}$$

PL: test t_0 under BG only ; $f(q_0|H_0)$

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

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



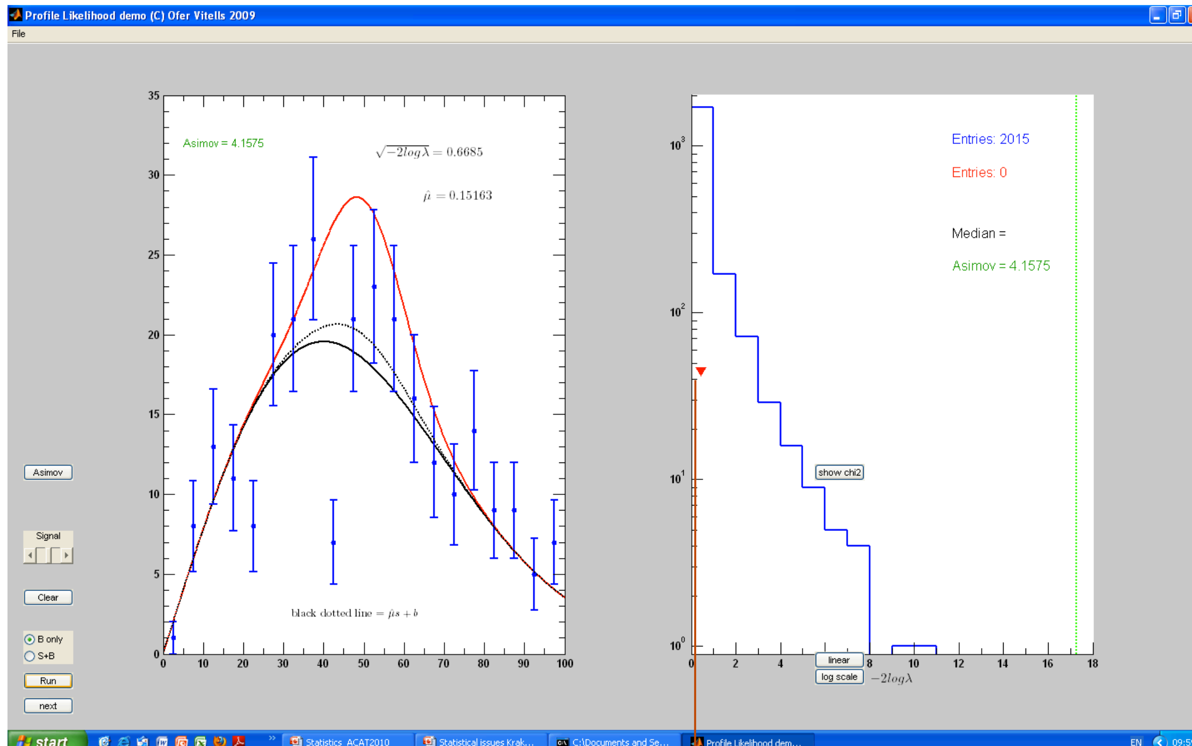
$$Z_{obs} = \sqrt{q_{0,obs}}$$

$$q_0 = 1.9 \rightarrow Z = 1.39\sigma$$

PL: test t_0 under BG only ; $f(q_0|H_0)$

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

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



$$Z_{obs} = \sqrt{q_{0,obs}} \quad q_0 = 0.43 \rightarrow Z = 0.66\sigma$$

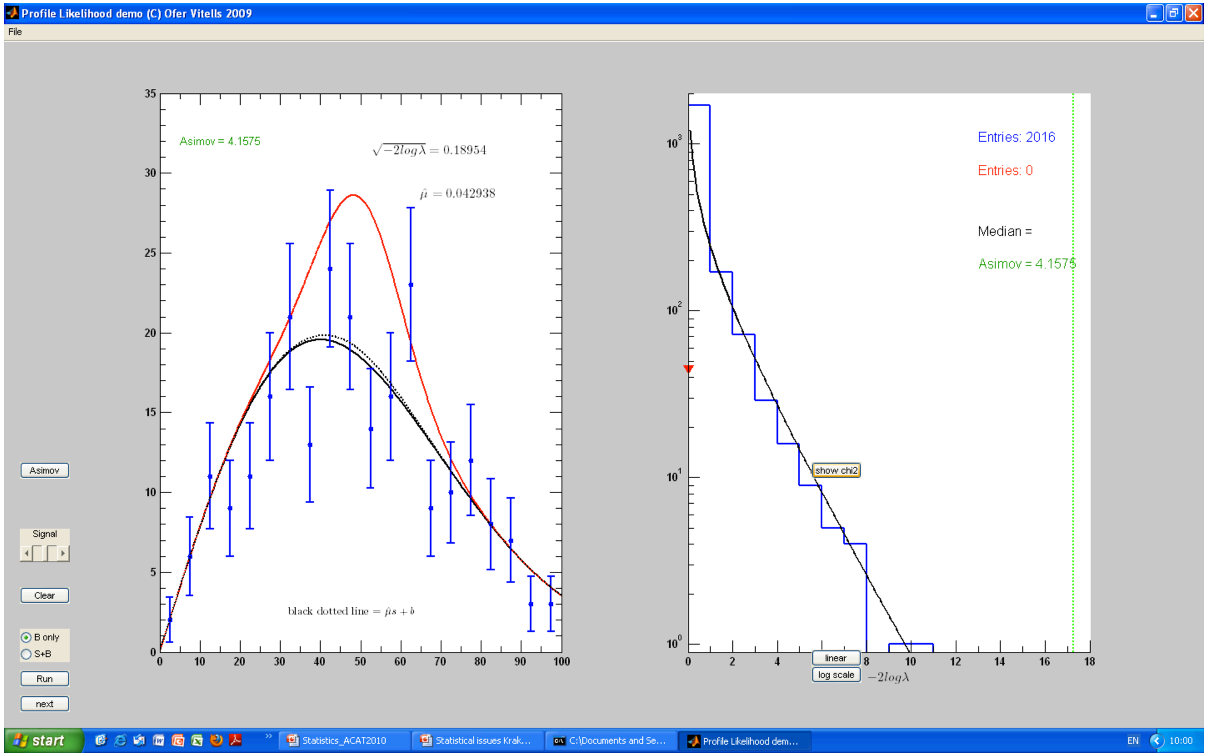
Confirm Wilks Theorem

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

For the test statistic

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

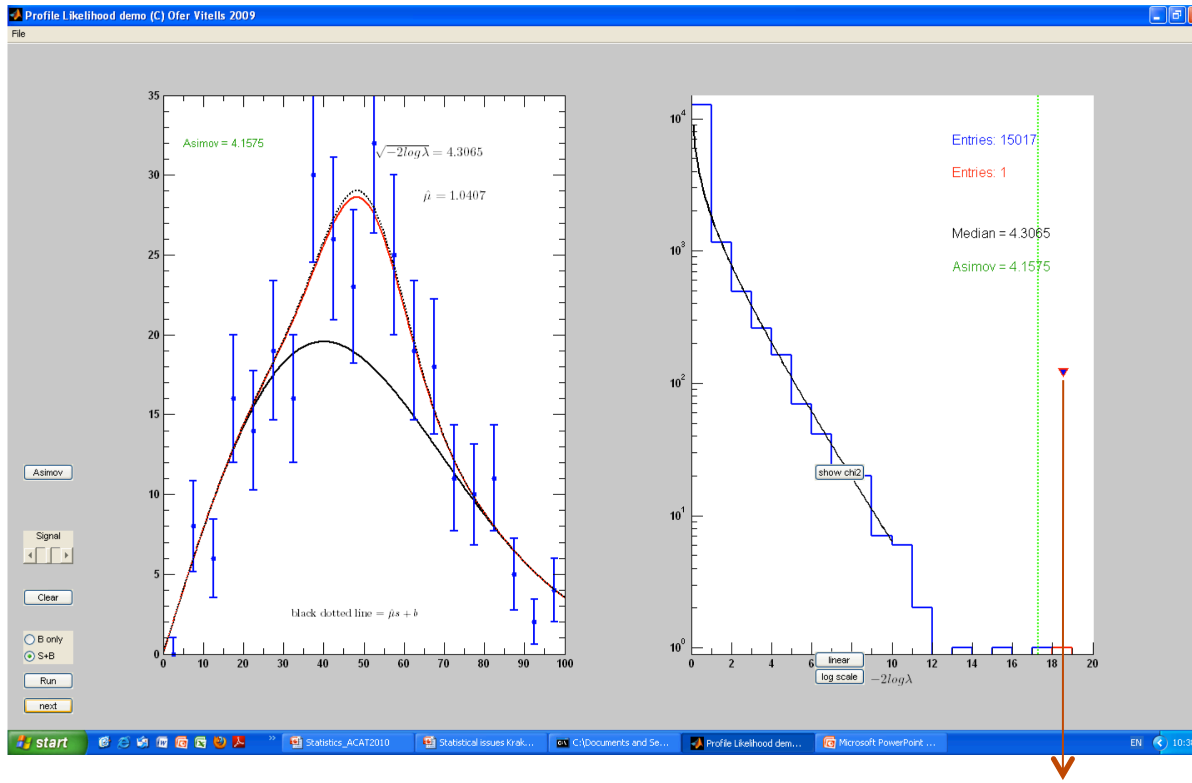
$$f(q_0 \mid H_0) = \chi_1^2$$



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

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

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

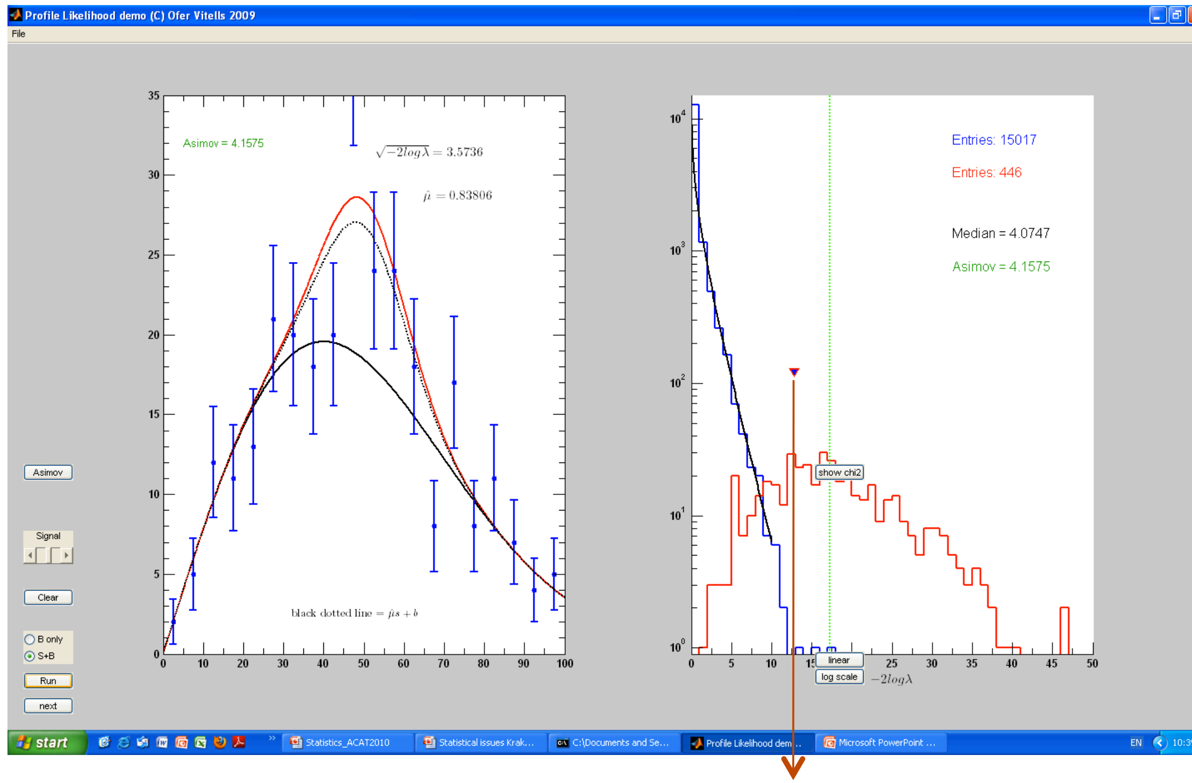


$$q_0 = 18.5 \rightarrow Z = 4.3\sigma$$

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

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

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

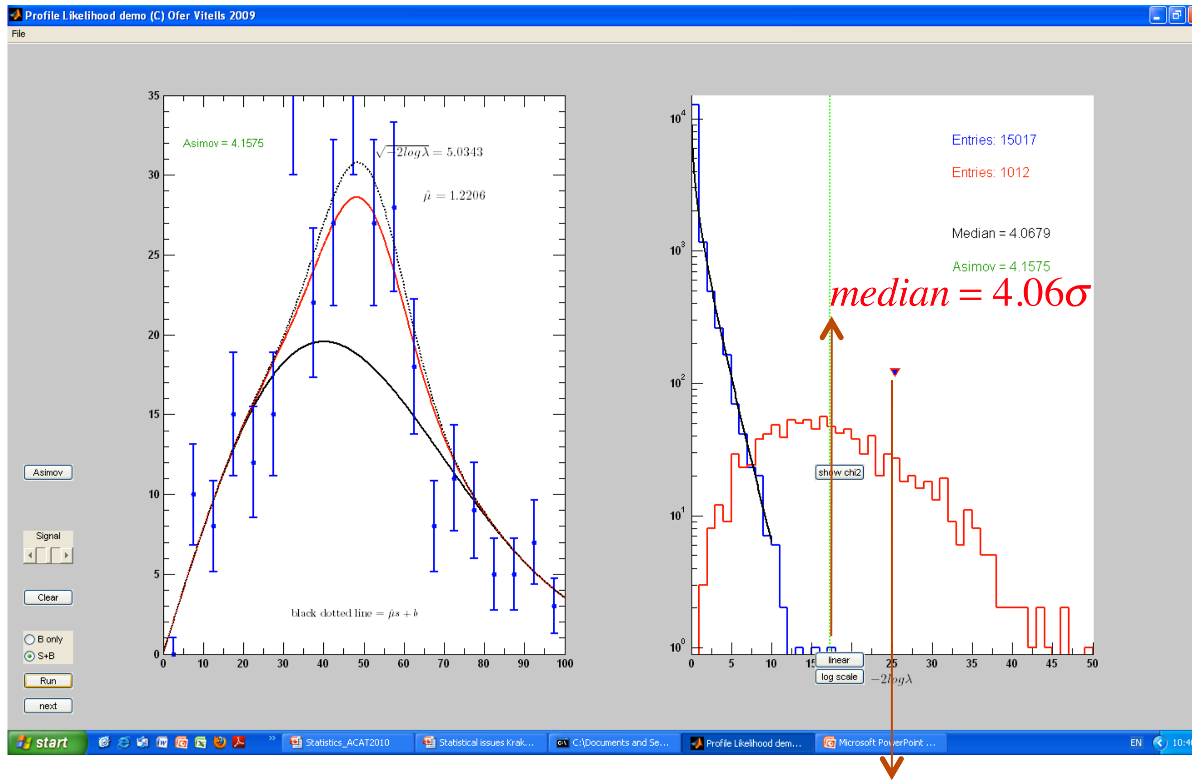


$$q_0 = 12.9 \rightarrow Z = 3.6\sigma$$

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

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

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



$$q_0 = 25 \rightarrow Z = 5.0\sigma$$

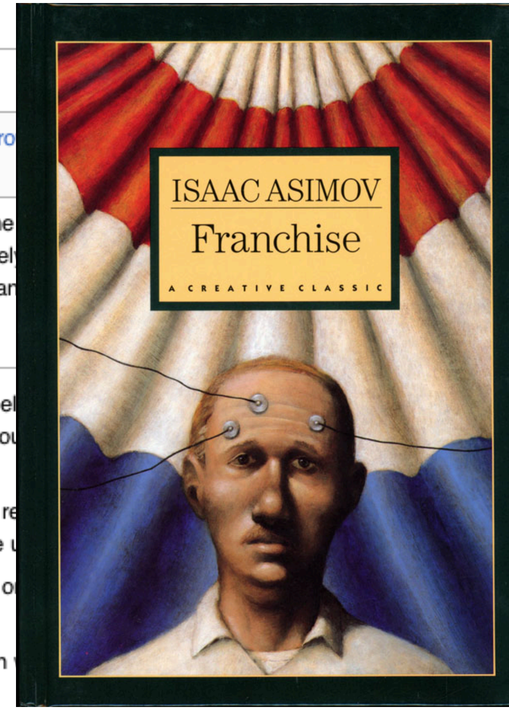
Median sensitivity in a Click (Asimov)

Franchise (short story)

From Wikipedia, the free encyclopedia

This article **needs additional citations for verification**. Please help impro ... s. Unsourced material may be challenged and removed. *(December 2009)*

Franchise is a science fiction short story by Isaac Asimov. It first appeared in the August 1955 issue of the ... was reprinted in the collections *Earth Is Room Enough* (1957) and *Robot Dreams* (1986). It is one of a loosel... fictional computer called Multivac. It is the first story in which Asimov dealt with computers *as computers* an...



"Franchise"	
uthor	Isaac Asimo
ountry	United State
anguage	English
eries	Multivac
enre(s)	science ficti
ublished in	*If*
ublisher	Quinn Publis
ledia type	Magazine
ublication date	August 195
receded by	"Question"
ollowed by	"The Dead F

Plot summary

In the future, the United States has converted to an "electronic democracy" where the computer Multivac sel... questions. Multivac will then use the answers and other data to determine what the results of an election wou... to be held.

The story centers around Norman Muller, the man chosen as "Voter of the Year" in 2008. Although the law re... not sure that he wants the responsibility of representing the entire electorate, worrying that the result will be ...

However, after 'voting', he is very proud that the citizens of the United States had, through him, "exercised o... 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 ... 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

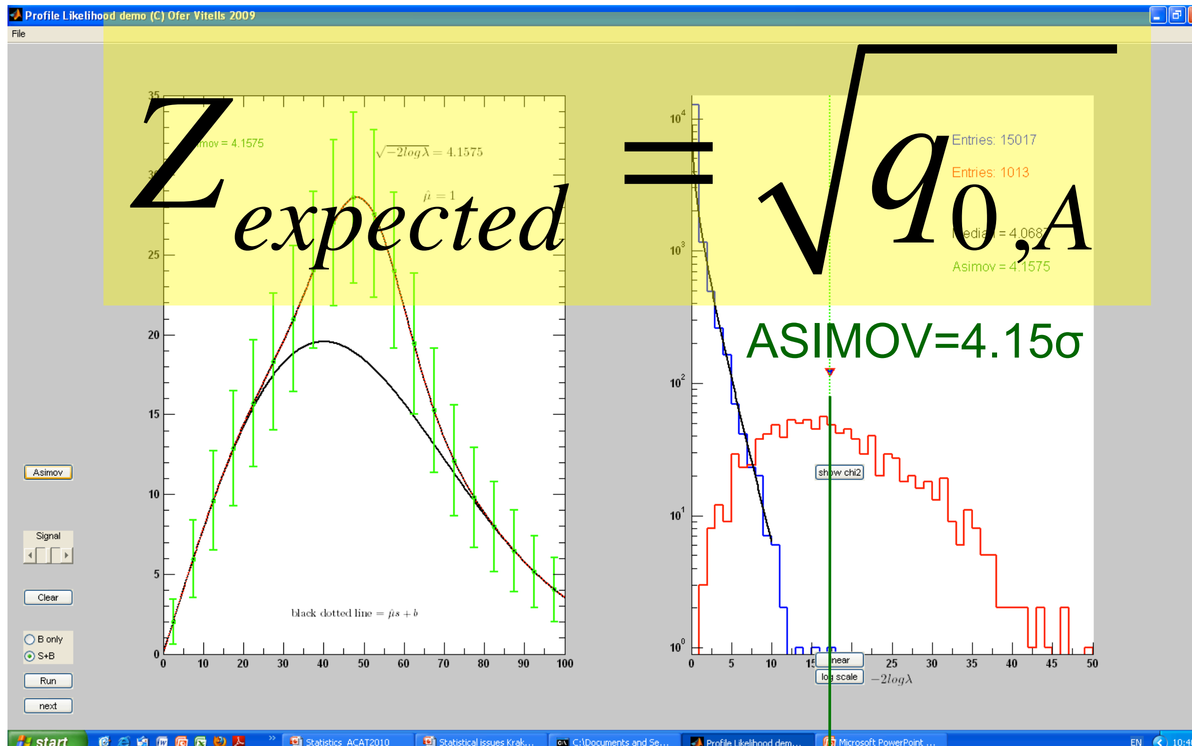
1. ^ G. Cowan, K. Cranmer, E. Gross, and O. Vitells (2011). "Asymptotic formulae for likelihood-based tests of new physics". *Eur.Phys.J.* **C71**: 1554. DOI:10.1140/epjc/s10052-011-1554-0.

CCGV ref

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 $t_{o,med}$ or evaluate t_0 with respect to a representative data set, the ASIMOV data set with $\mu=1$, i.e. x=s+b

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



$$Z_{expected} = \sqrt{q_{0,A}}$$

$$q_A = 17.22 \rightarrow Z_A = 4.15$$

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

Asymptotic Distributions

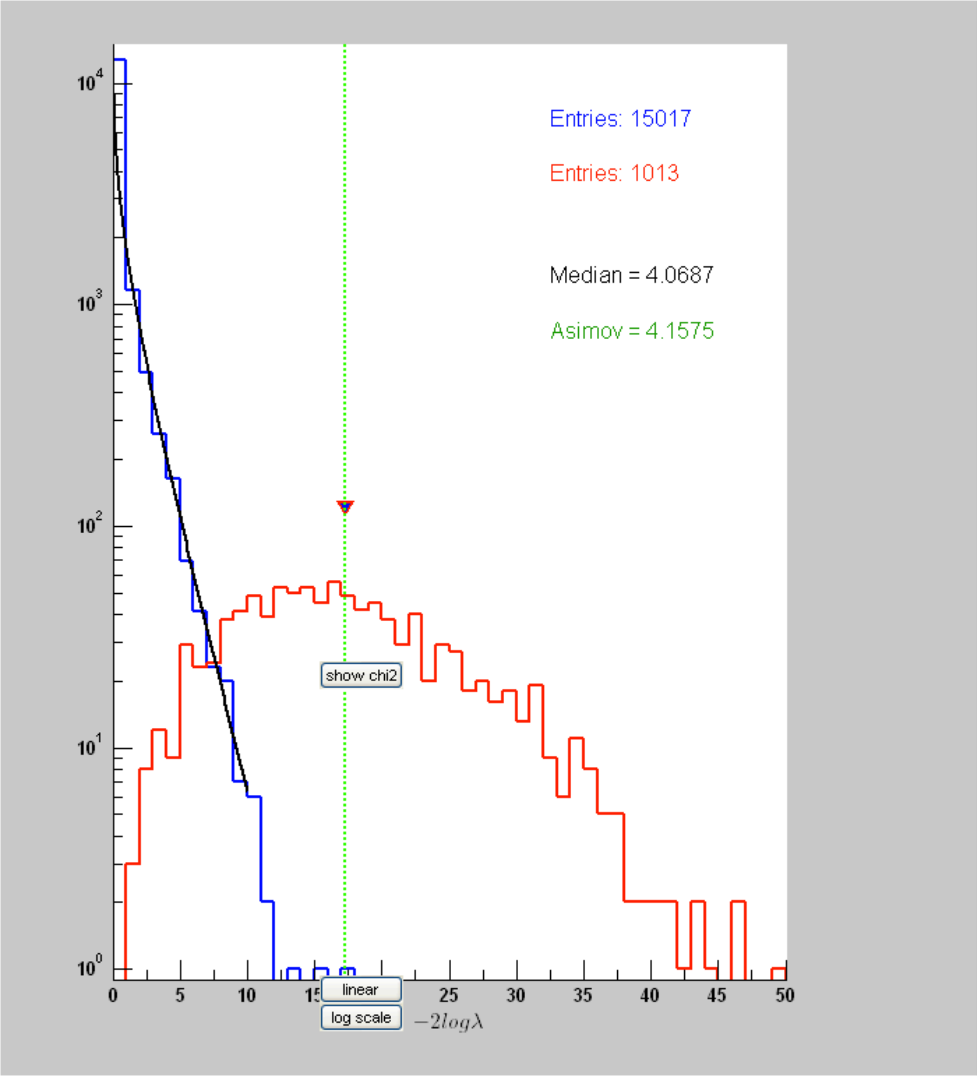
Tossing Monte Carlos to get the test statistic distribution functions (PDF) is sometimes beyong the experiment technical capability.

Knowing both PDF

$$f(q_{null} \mid H_{null})$$

$$f(q_{null} \mid H_{alternate})$$

enables calculating both the observed and expected significance (or exclusion) without a single toy….



Asymptotic Distributions

CCGV

q_{null}

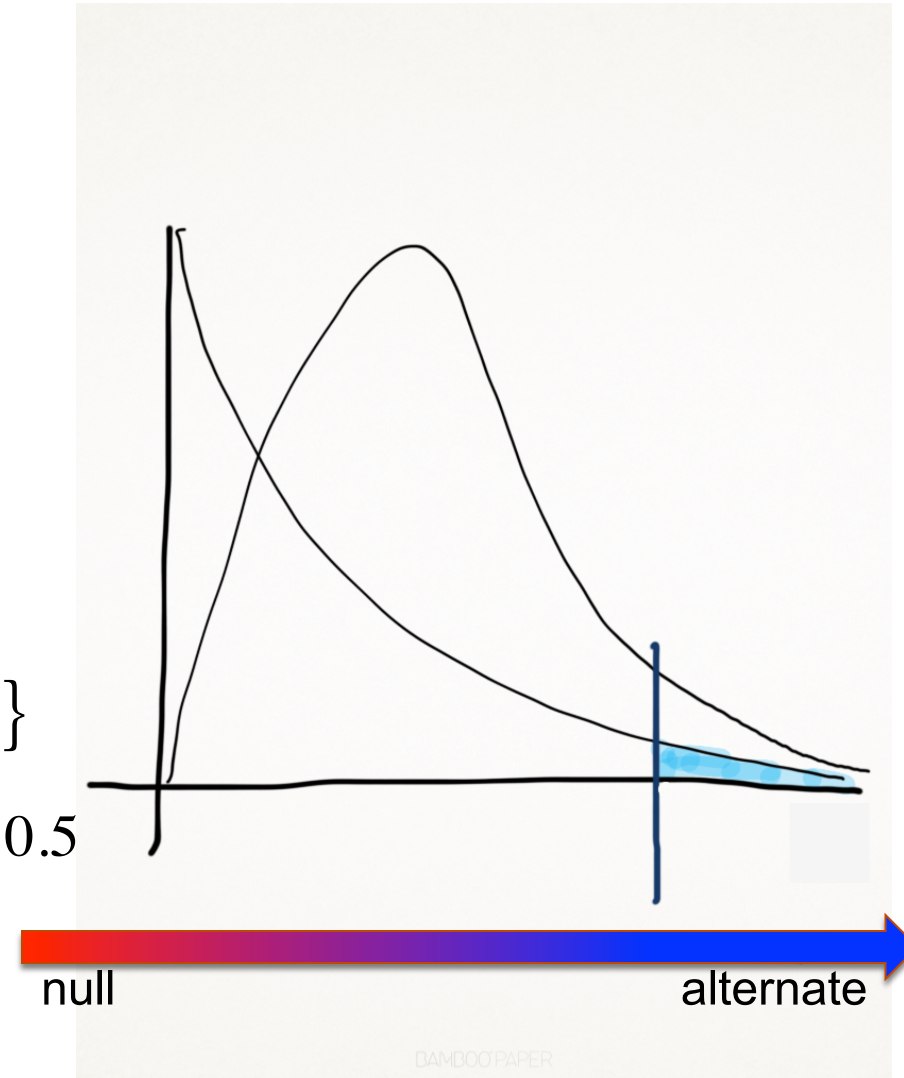
$f(q_{null} \mid H_{null})$

$q_{obs} \equiv q_{null,obs}$

$$p = \int_{q_{obs}}^{\infty} f(q_{null} \mid H_{null}) dq_{null}$$

$f(q_{null} \mid H_{alt})$

$$q_A \equiv q_{null,A} = \begin{matrix} \{q \mid med\{f(q_{null} \mid H_{alt})\}\} \\ \int_{q_{null,A}}^{\infty} f(q_{null} \mid H_{null}) dq_{null} = 0.5 \end{matrix}$$



q_{null}

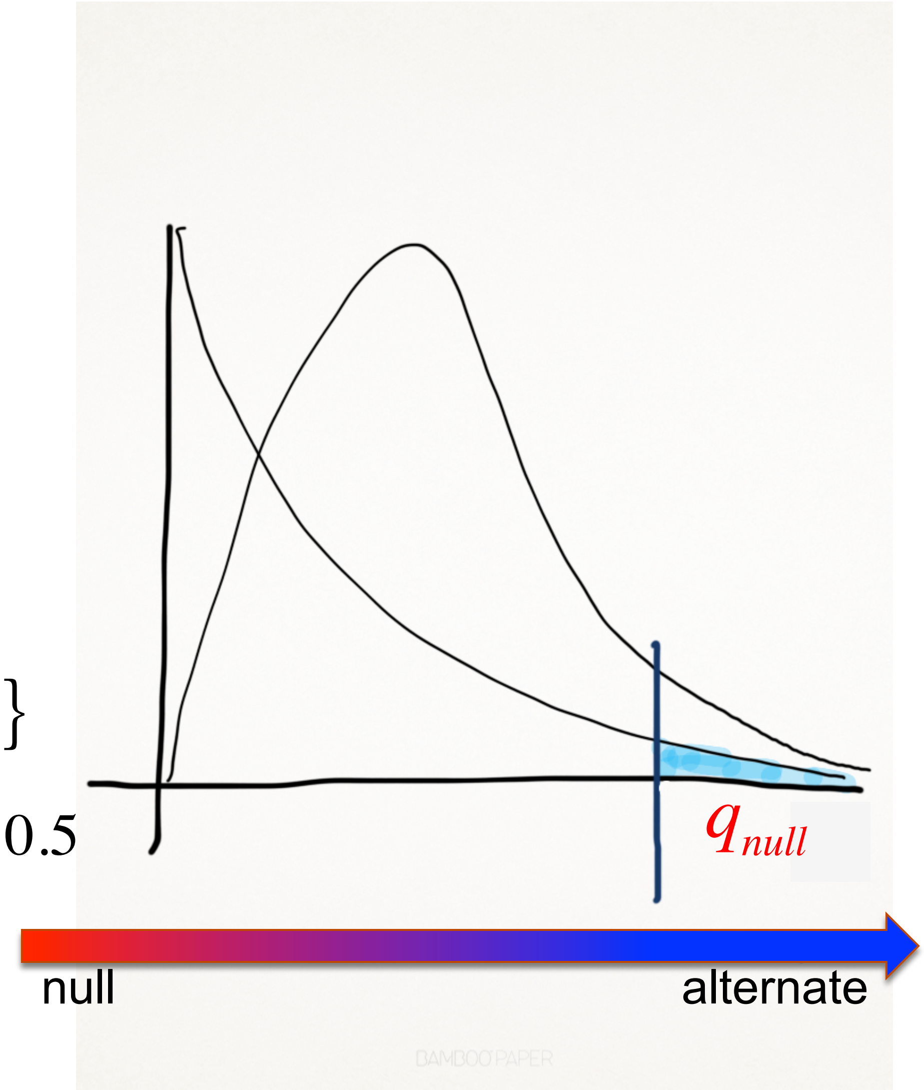
$f(q_{null} \mid H_{null})$

$q_{obs} \equiv q_{null,obs}$

$$p = \int_{q_{obs}}^{\infty} f(q_{null} \mid H_{null}) dq_{null}$$

$f(q_{null} \mid H_{alt})$

$$q_A \equiv q_{null,A} = \begin{array}{l} \{q \mid med\{f(q_{null} \mid H_{alt})\}\} \\ \int_{q_{null,A}}^{\infty} f(q_{null} \mid H_{null}) dq_{null} = 0.5 \end{array}$$



q_{null}

$f(q_{null} | H_{null})$

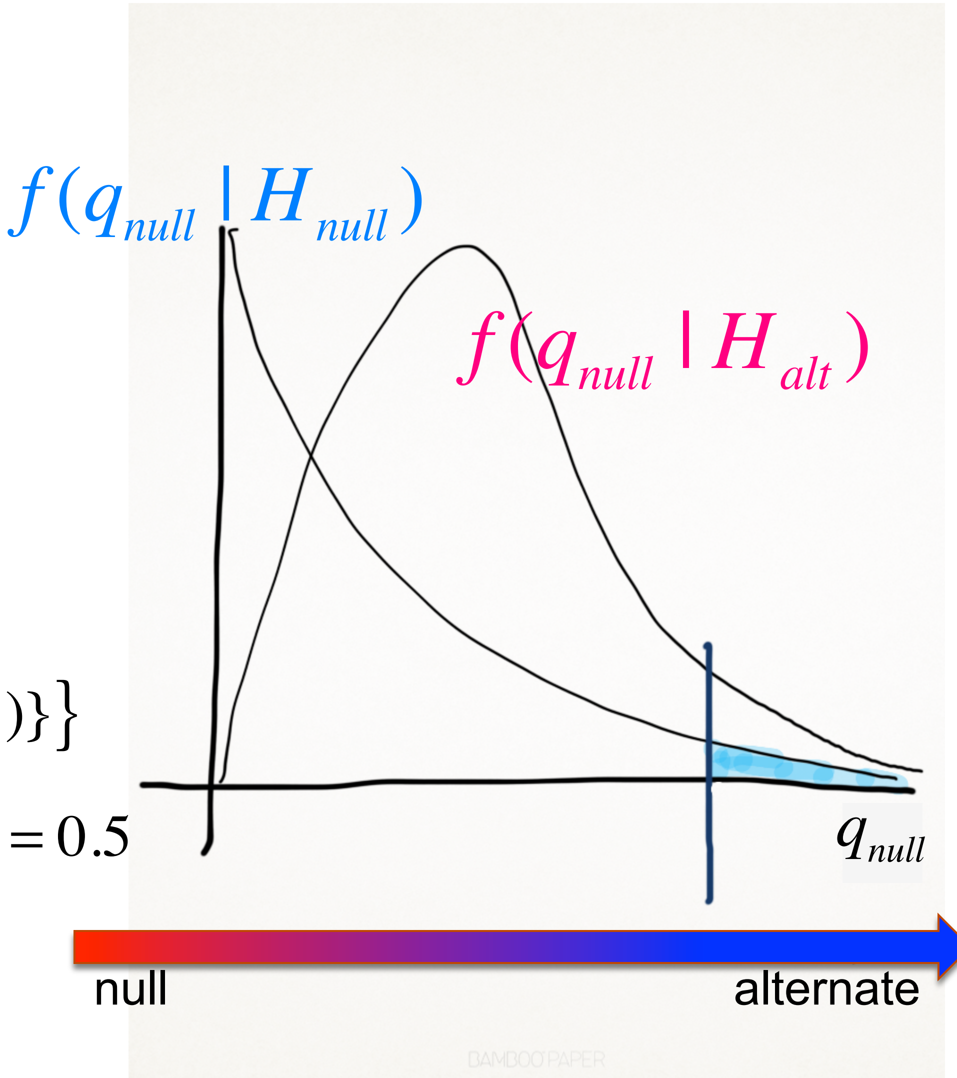
$q_{obs} \equiv q_{null,obs}$

$$p = \int_{q_{obs}}^{\infty} f(q_{null} | H_{null}) dq_{null}$$

$f(q_{null} | H_{alt})$

$$q_A \equiv q_{null,A} = \{q | med\{f(q_{null} | H_{alt})\}\}$$

$$\int_{q_{null,A}}^{\infty} f(q_{null} | H_{null}) dq_{null} = 0.5$$



q_{null}

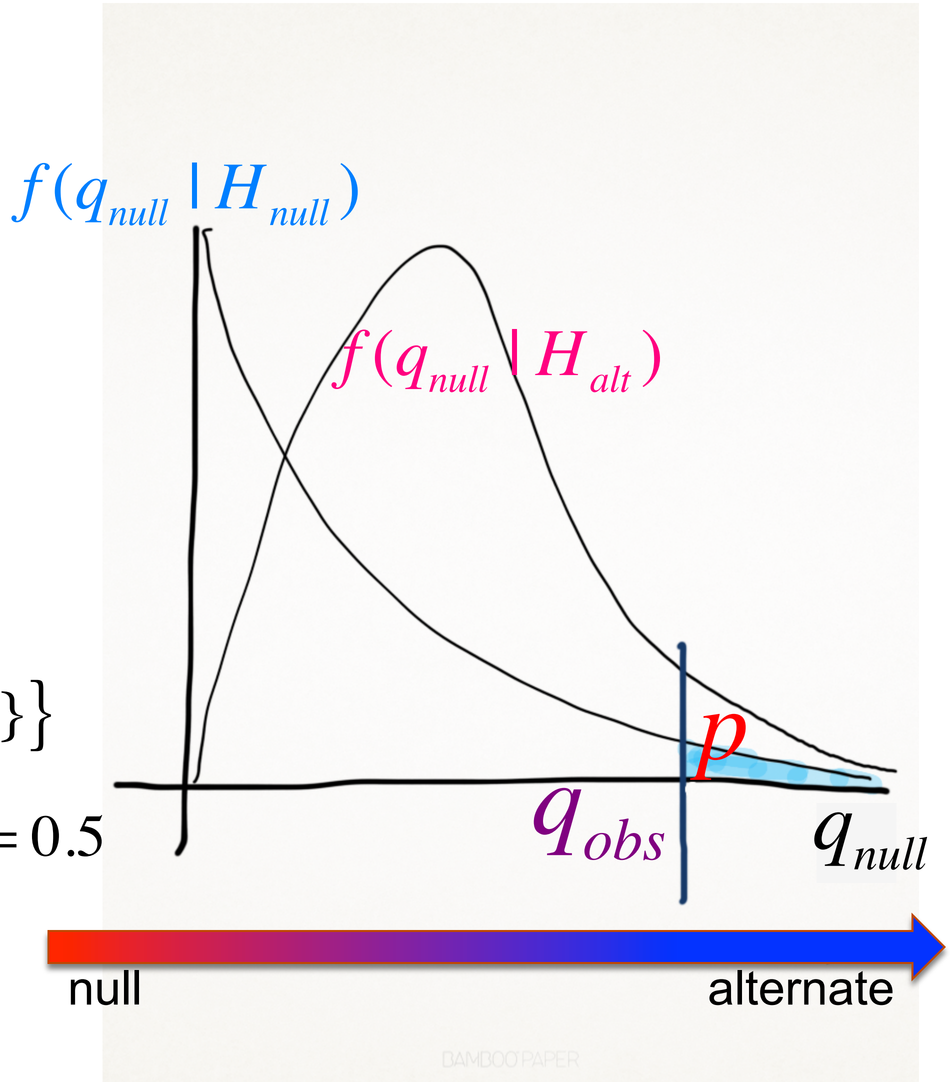
$f(q_{null} \mid H_{null})$

$q_{obs} \equiv q_{null,obs}$

$$p = \int_{q_{obs}}^{\infty} f(q_{null} \mid H_{null}) dq_{null}$$

$f(q_{null} \mid H_{alt})$

$$q_A \equiv q_{null,A} = \begin{matrix} \{q \mid med\{f(q_{null} \mid H_{alt})\}\} \\ \int_{q_{null,A}}^{\infty} f(q_{null} \mid H_{null}) dq_{null} = 0.5 \end{matrix}$$



q_{null}

$f(q_{null} \mid H_{null})$

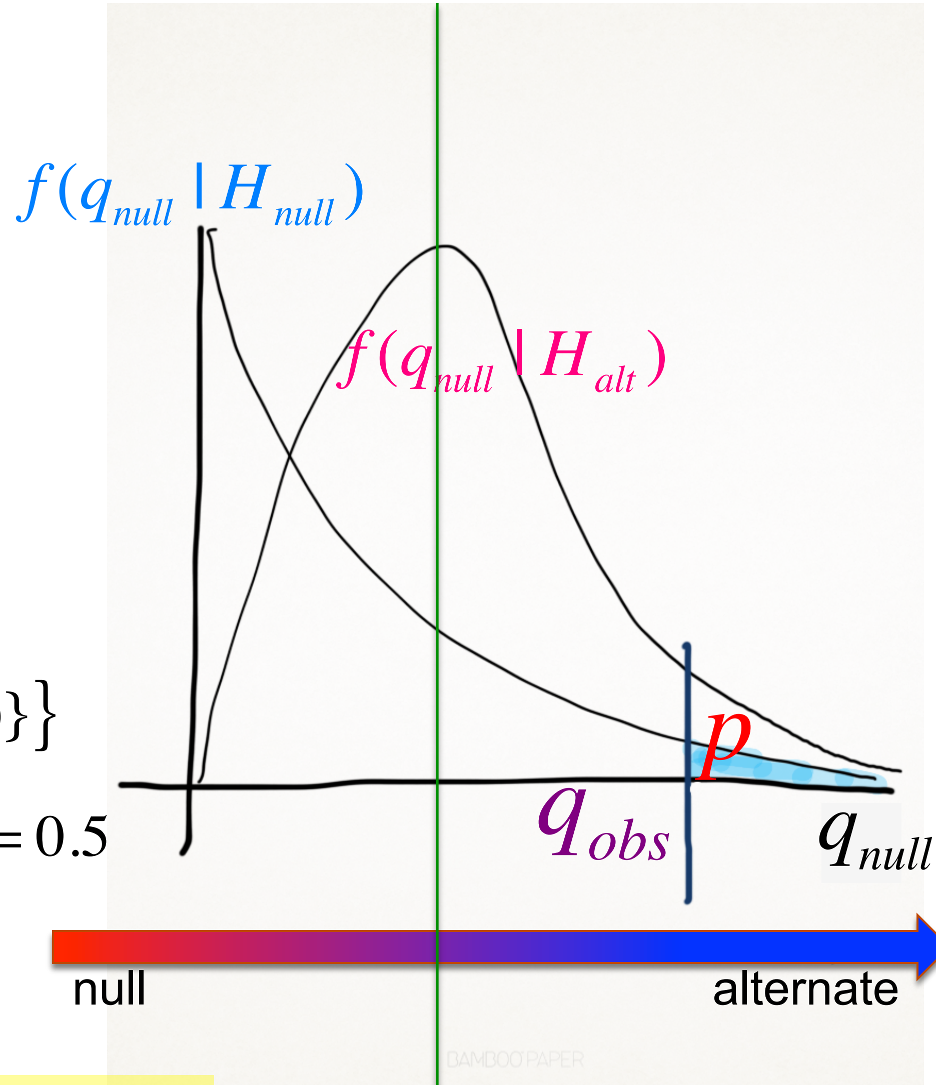
$q_{obs} \equiv q_{null,obs}$

$p = \int_{q_{obs}}^{\infty} f(q_{null} \mid H_{null}) dq_{null}$

$f(q_{null} \mid H_{alt})$

$$q_A \equiv q_{null,A} = \{q \mid med\{f(q_{null} \mid H_{alt})\}\}$$

$$\int_{q_{null,A}}^{\infty} f(q_{null} \mid H_{null}) dq_{null} = 0.5$$



$$Z_{expected} = \sqrt{q_{null,A}}$$

$q_A \equiv q_{null,A}$

Test Statistics	Purpose	Experession	LR
q_0	discovery of positive signal	$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}$	$\lambda(0) = \dfrac{L(0,\hat{\hat{\theta}}_0)}{L(\hat{\mu},\hat{\theta})}$
t_μ	2-sided measurement	$t_\mu = -2\ln\lambda(\mu)$	$\lambda(\mu) = \dfrac{L(\mu,\hat{\hat{\theta}}_\mu)}{L(\hat{\mu},\hat{\theta})}$
$\tilde{t}_\mu$	avoid negative signal (FC)	$\tilde{t}_\mu = -2\ln\tilde{\lambda}(\mu)$	$\tilde{\lambda}(\mu) = \begin{cases} \dfrac{L(\mu,\hat{\hat{\theta}}_\mu)}{L(\hat{\mu},\hat{\theta})} & \hat{\mu} \geq 0 \\ \dfrac{L(\mu,\hat{\hat{\theta}}_\mu)}{L(0,\hat{\hat{\theta}}_0)} & \hat{\mu} < 0 \end{cases}$
q_μ	exclusion	$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$	
$\tilde{q}_\mu$	exclusion of positive signal	$\tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$	

Resolving $f(q_{null}|H_{alt})$

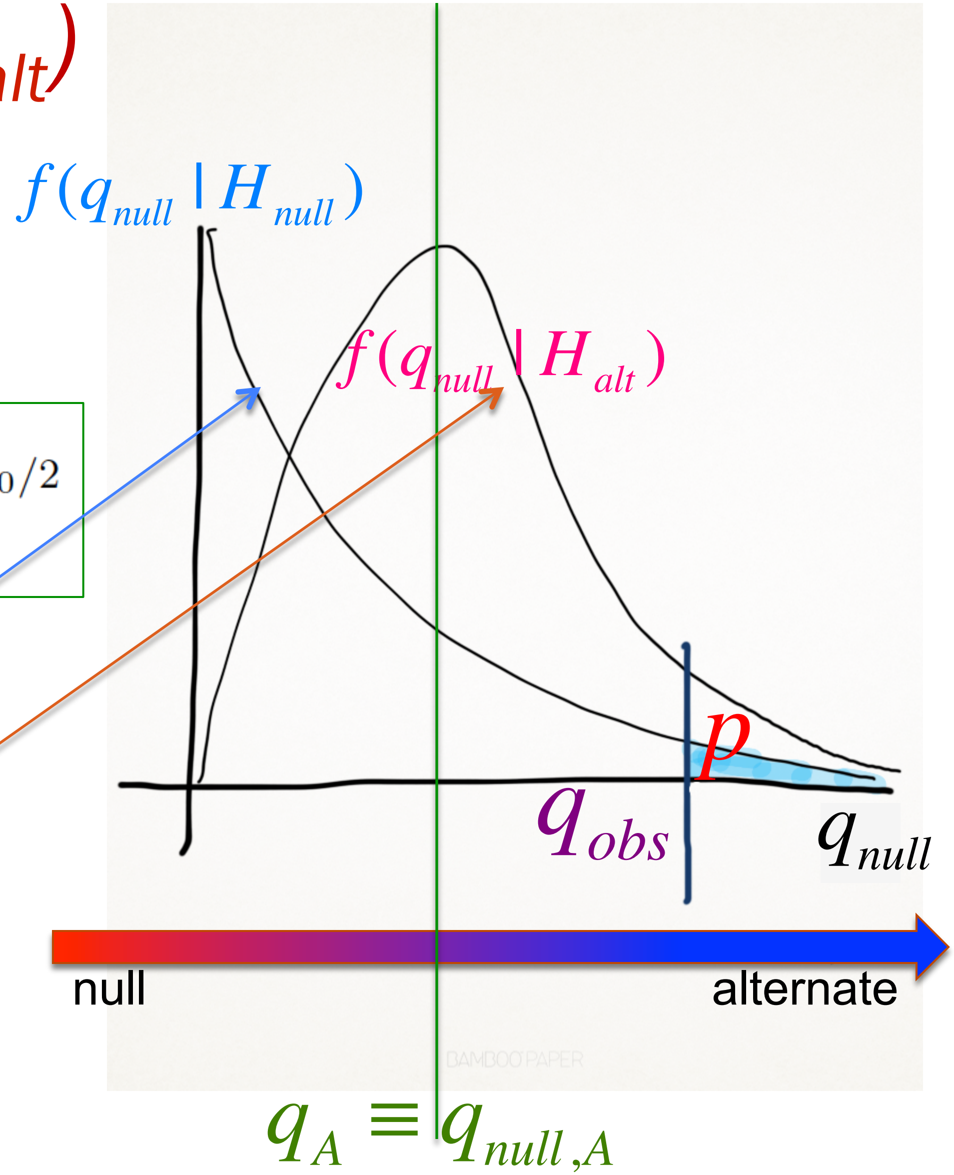
$$n = \mu s + b(\theta)$$

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

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

$$f(q_0|0) \sim \frac{1}{2}\chi^2$$

$$f(q_0|\mu') \sim ?$$



Wald Theorem

- Consider a test of the strength parameter μ , which here can either be zero (for discovery) or nonzero (for an upper limit), and suppose the data are distributed according to a strength parameter μ'
- The desired distribution $f(q_\mu | \mu')$ can be found using a result due to Wald [1946], who showed that for the case of a single parameter of interest,

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

$$\lambda(\mu) = \frac{L(\mu,\hat{\hat{\theta}}_\mu)}{L(\hat{\mu},\hat{\theta})}$$

$$\langle\hat{\mu}\rangle = \mu'$$

Wald Theorem

- Following the Wald Theorem we find that the 2-sided $t_\mu = -2\ln\lambda(\mu)$ distributes like a non-central chi squared

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}}_\mu)}{L(\hat{\mu}, \hat{\theta})}$$

$$f(t_\mu; \Lambda) = \frac{1}{2\sqrt{t_\mu}} \frac{1}{\sqrt{2\pi}} \left[\exp\left(-\frac{1}{2} \left(\sqrt{t_\mu} + \sqrt{\Lambda} \right)^2 \right) + \exp\left(-\frac{1}{2} \left(\sqrt{t_\mu} - \sqrt{\Lambda} \right)^2 \right) \right]$$

2 sided CI

μ is the tested hypothesis while $\langle \hat{\mu} \rangle = \mu'$

$$\Lambda = \frac{(\mu - \mu')^2}{\sigma^2}$$

under H_μ, if $\mu' = \mu$

we get Wilks theorem

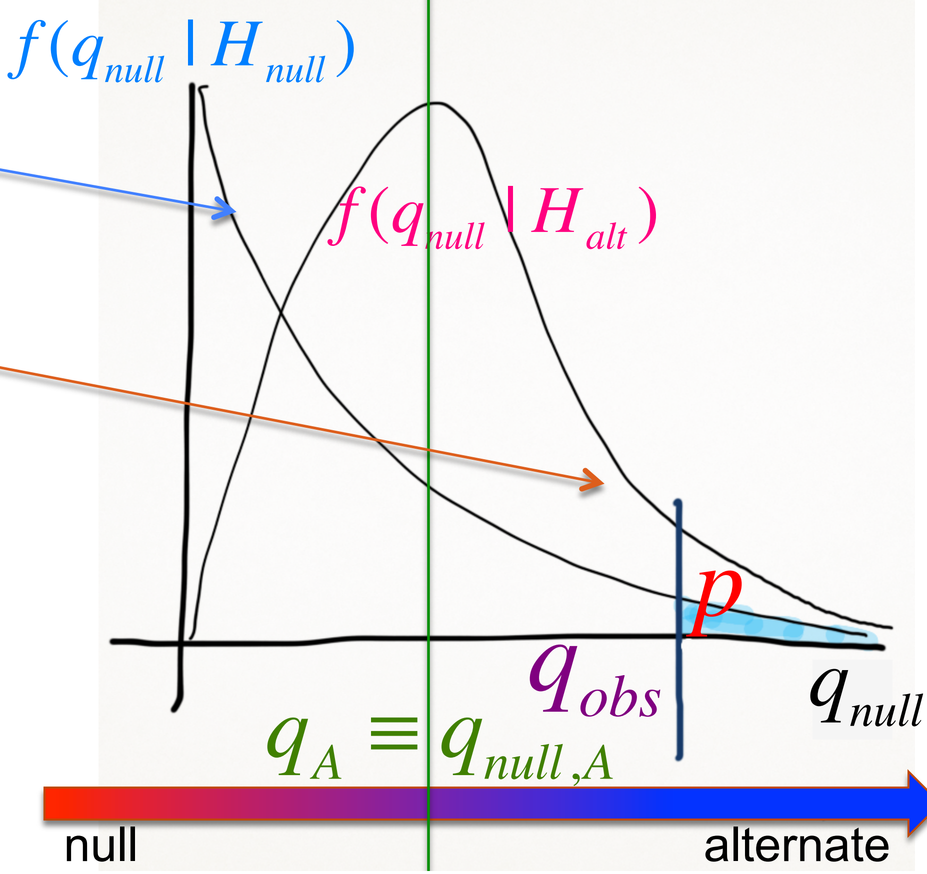
$$f(t_\mu|\mu) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{t_\mu}} e^{-t_\mu/2}$$

The rediscovery Wald theorem helped us to find the asymptotic distributions of all PL test Statistics, including the Neyman Pearson one, calculate the CLs modified p-values the expected sensitivity and save months if not years of computing

Asymptotic Distribution for Discovery

$$f(q_0 | 0) \sim \frac{1}{2}\chi^2$$

$$f(q_0 | \mu') \sim ?$$



1 sided CI

$$f(q_0|\mu') = \left(1 - \Phi\left(\frac{\mu'}{\sigma}\right)\right)\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}\exp\left[-\frac{1}{2}\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right)^2\right]$$

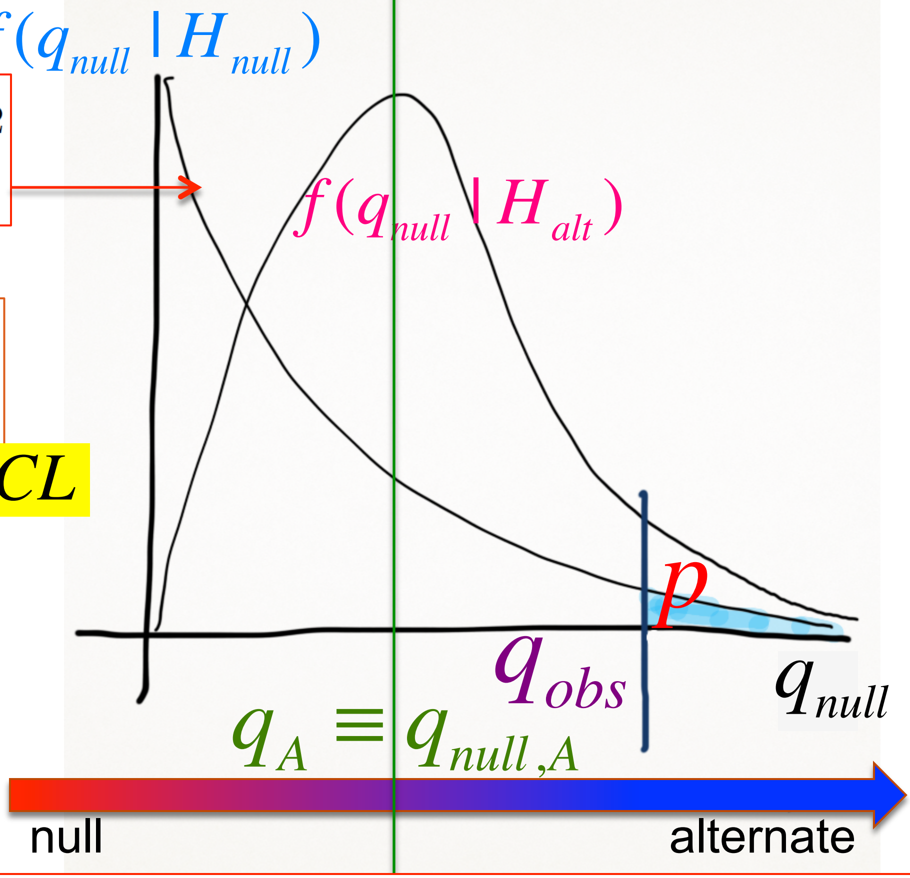
$$\Phi(Z) = 1 - \int_Z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

Asymptotic Distribution for Exclusion

$$f(q_\mu|\mu) = \frac{1}{2}\delta(q_\mu) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}e^{-q_\mu/2}$$

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

$\alpha = 1 - CL$



1 sided CI

$$f(q_\mu|\mu') = \Phi\left(\frac{\mu'-\mu}{\sigma}\right)\delta(q_\mu) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}\exp\left[-\frac{1}{2}\left(\sqrt{q_\mu} - \frac{\mu-\mu'}{\sigma}\right)^2\right]$$

Asymptotic Distribution for FC

Depends on the observation
one might get 1-sided or 2-sided CI

3.4 Distribution of $\tilde{t}_\mu$

Assuming the Wald approximation, the statistic t_μ as defined by Eq. (11) can be written

$$\tilde{t}_\mu = \begin{cases} \frac{\mu^2}{\sigma^2} - \frac{2\mu\hat{\mu}}{\sigma^2} & \hat{\mu} < 0 \,, \\ \frac{(\mu-\hat{\mu})^2}{\sigma^2} & \hat{\mu} \geq 0 \,. \end{cases} \tag{40}$$

From this the pdf $f(\tilde{t}_\mu|\mu')$ is found to be

$$f(\tilde{t}_\mu|\mu') = \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\tilde{t}_\mu}}\exp\left[-\frac{1}{2}\left(\sqrt{\tilde{t}_\mu}+\frac{\mu-\mu'}{\sigma}\right)^2\right] \tag{41}$$

$$+ \begin{cases} \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\tilde{t}_\mu}}\exp\left[-\frac{1}{2}\left(\sqrt{\tilde{t}_\mu}-\frac{\mu-\mu'}{\sigma}\right)^2\right] & \tilde{t}_\mu \leq \mu^2/\sigma^2 \,, \\ \frac{1}{\sqrt{2\pi}(2\mu/\sigma)}\exp\left[-\frac{1}{2}\frac{\left(\tilde{t}_\mu-\frac{\mu^2-2\mu\mu'}{\sigma^2}\right)^2}{(2\mu/\sigma)^2}\right] & \tilde{t}_\mu > \mu^2/\sigma^2 \end{cases} \,. \tag{42}$$

The special case $\mu = \mu'$ is therefore

$$f(\tilde{t}_\mu|\mu') = \begin{cases} \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\tilde{t}_\mu}}e^{-\tilde{t}_\mu/2} & \tilde{t}_\mu \leq \mu^2/\sigma^2 \,, \\ \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\tilde{t}_\mu}}e^{-\tilde{t}_\mu/2} + \frac{1}{\sqrt{2\pi}(2\mu/\sigma)}\exp\left[-\frac{1}{2}\frac{(\tilde{t}_\mu+\mu^2/\sigma^2)^2}{(2\mu/\sigma)^2}\right] & \tilde{t}_\mu > \mu^2/\sigma^2 \,. \end{cases} \tag{43}$$

How to determine σ

- To estimate the uncertainty σ there are a few possibilities
  - Given the asymptotic formulae, fit the distribution of

$$f(q_{null} \mid H_{alt}) = f(q_\mu \mid \mu')$$ and extract σ

  - Implement the Wald formula to the Asimov data set and find

$$\sigma_A^2 = \frac{(\mu - \mu')^2}{q_{\mu,A}}$$

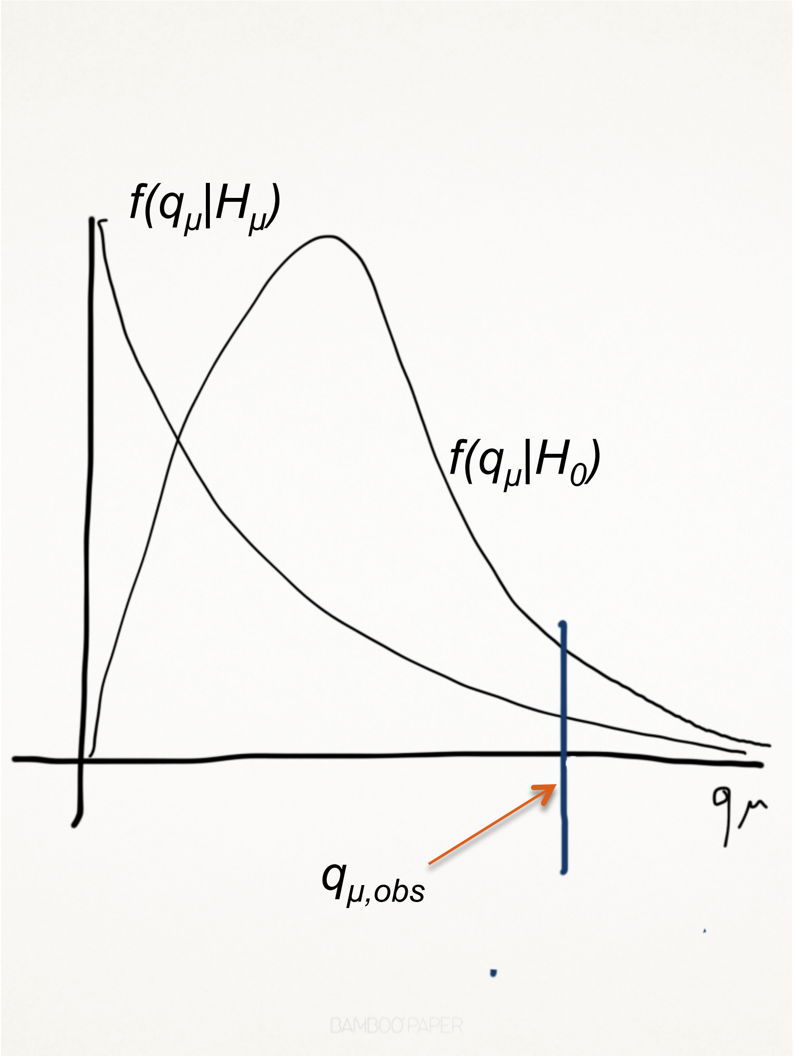
where μ is the tested (null) hypothesis and μ' is the alt hypothesis.
For discovery, μ =0 while for exclusion μ '=0.

Exclusion

Case Study:

Exclusion of a Higgs with mass m_H

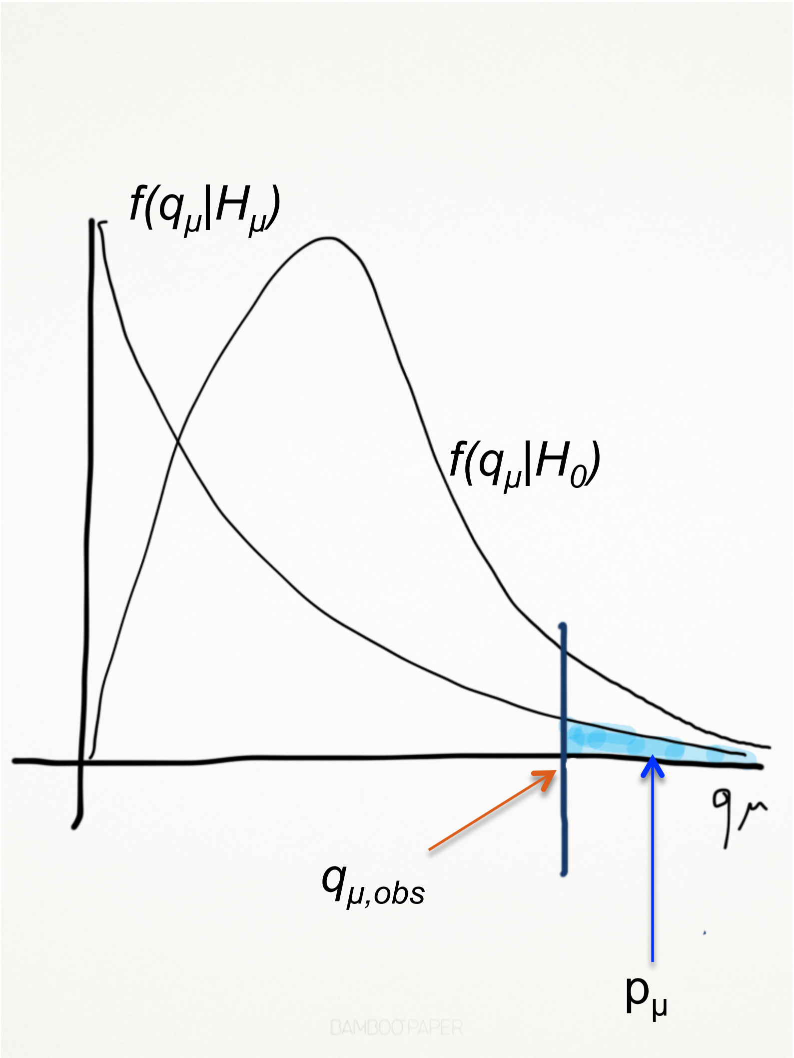
- We test hypothesis H_μ
- We calculate the PL (profile likelihood) ratio with the one observed data
- $q_{\mu,obs}$



- 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

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

- Now suppose $s(m_{H2})=1$, can we exclude m_{H2}?
- Suppose $n_{obs}=80$, $\text{prob}(n_{obs} \leq 80 | 102) < 5\%$, it 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 modified p-value
- We cannot exclude m_{H2}

$$\frac{\text{Prob}(\text{nobs} \leq 80 | 101)}{\text{Prob}(\text{nobs} \leq 80 | 100)} \sim 1$$

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

CLs

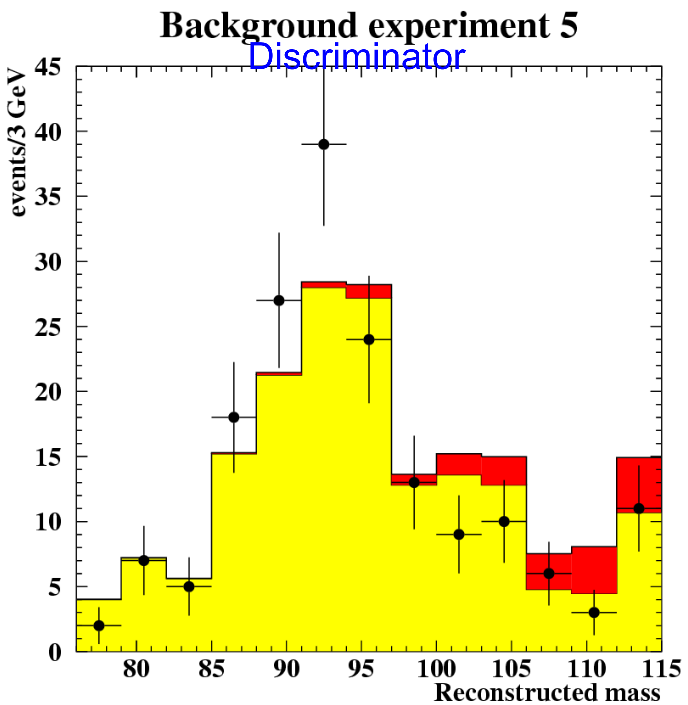
The Neyman-Pearson Lemma (lite version)

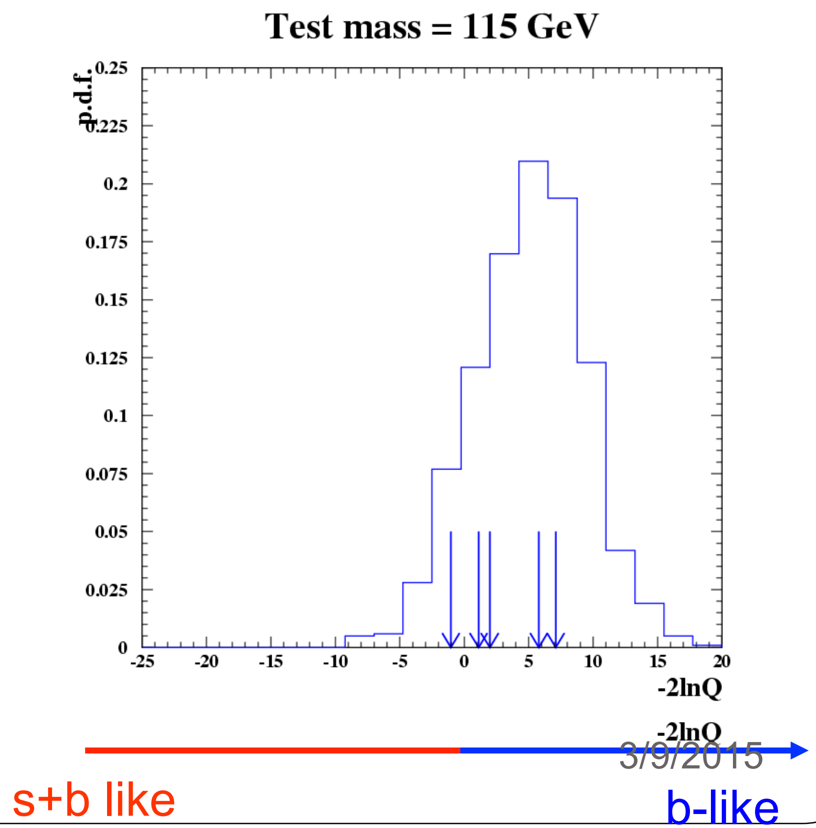
- When performing a hypothesis test between two simple hypotheses, H_0 and H_1, **the Likelihood Ratio test**, which rejects H_0 in favor of H_1, **is the most powerful test** …..
- Define a **test statistic** $Q = -2\ln\frac{L(H_0)}{L(H_1)}$
- Then for a given $\alpha = Prob(reject\ H_0 \mid H_0)$
  the probability $Prob(reject\ H_0 \mid \bar{H}_0) = Prob(reject\ H_0 \mid H_1)$
  is the highest, i.e.
  The Likelihood Ratio $Q = -2\ln\frac{L(H_0)}{L(H_1)}$
  is the most powerful test
- (The POWER of an hypothesis test is the probability to reject the null hypothesis when the alternate hypothesis is true!)
  NOTE: $Q = Q(\hat{\mu})$

Example: Simulating BG Only Experiments

$$Q(m) = \frac{L(H_1)}{L(H_0)} = \frac{L(s(m)+b)}{L(b)}$$

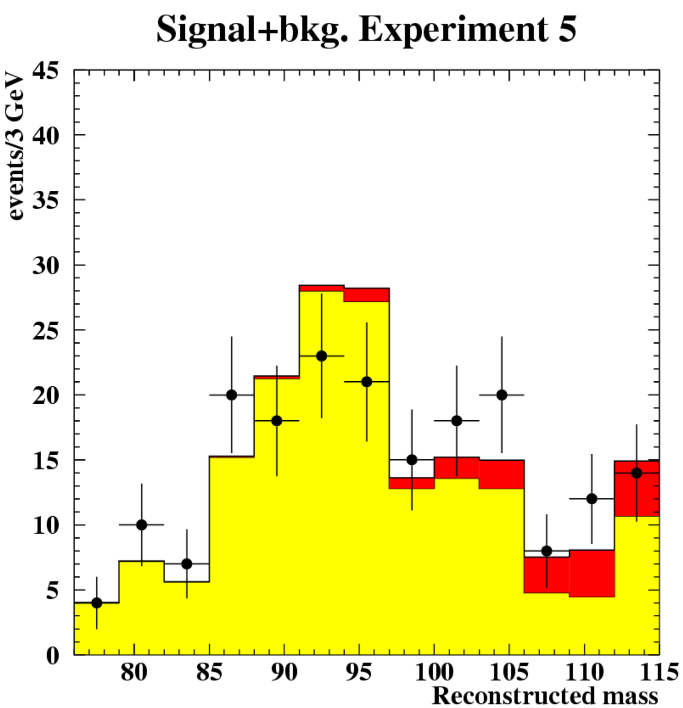
- The likelihood ratio, $-2\ln Q(m_H)$ tells us how much the outcome of an experiment is signal-like
- NOTE, here the s+b pdf is plotted to the left (it's the null hypothesis)!

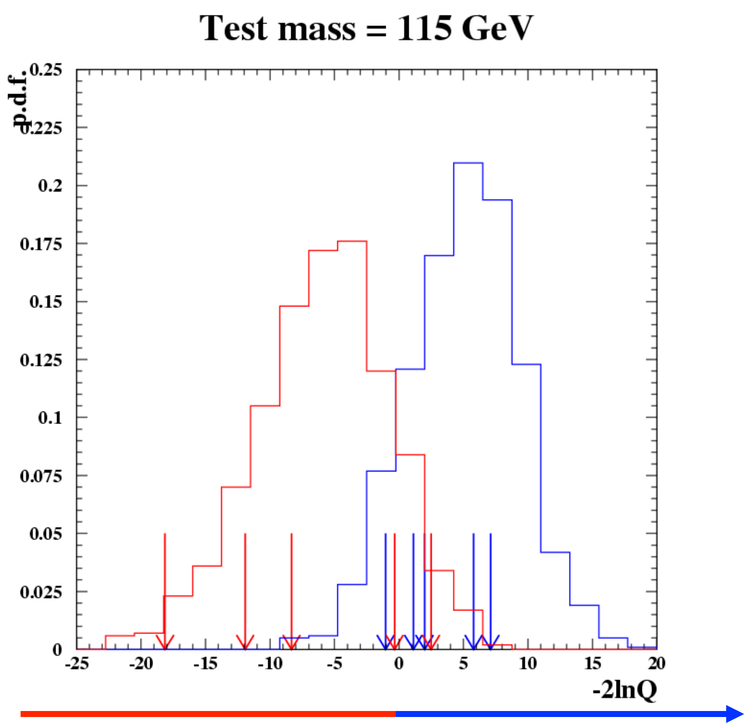




Example:

Simulating S(m_H)+b Experiments



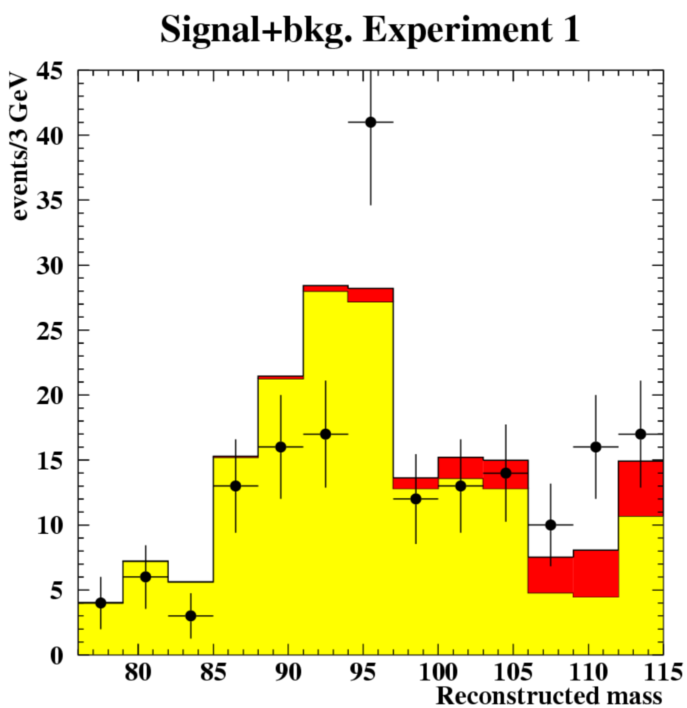


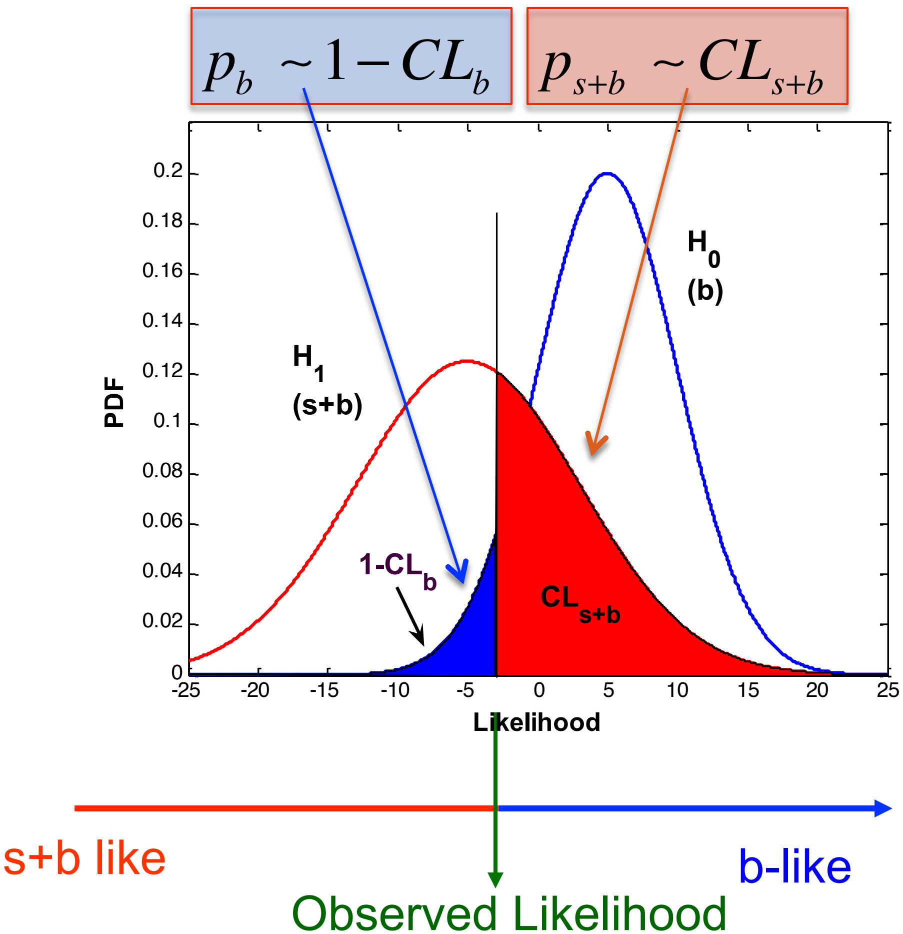
s+b like

b-like

Example:

Simulating $S(m_H)$+b Experiments

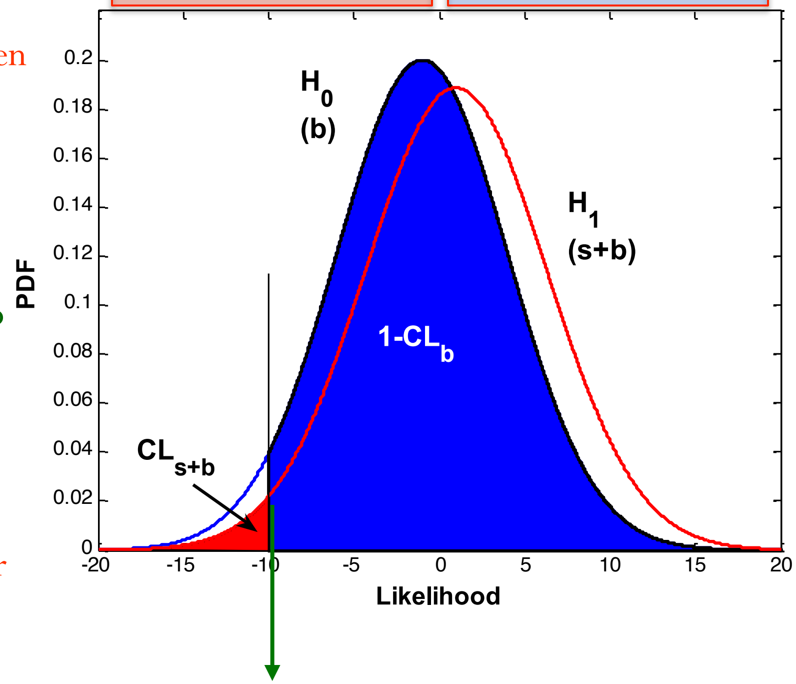




The Problem of Small Signal

- $\langle N_{obs} \rangle = s+b$ leads to the physical requirement that $N_{obs} > b$
- A very small expected s might lead to an anomaly when N_{obs} fluctuates far below the expected background, b.
- At one point DELPHI alone had $CL_{s+b}=0.03$ for $m_H=116$ GeV
- However, the cross section for 116 GeV Higgs at LEP was too small and Delphi actually had no sensitivity to observe it
- The frequntist would say: Suppose there is a 116 GeV Higgs….
  In 3% of the experiments the true signal would be rejected… (one would obtain a result incompatible or more so with m=116)
  i.e. a 116 GeV Higgs is excluded at the 97% CL…..

$p_{s+b} \sim CL_{s+b}$ $\quad$ $p_b \sim 1 - CL_b$



Observed Likelihood

The CLs Method for Upper Limits

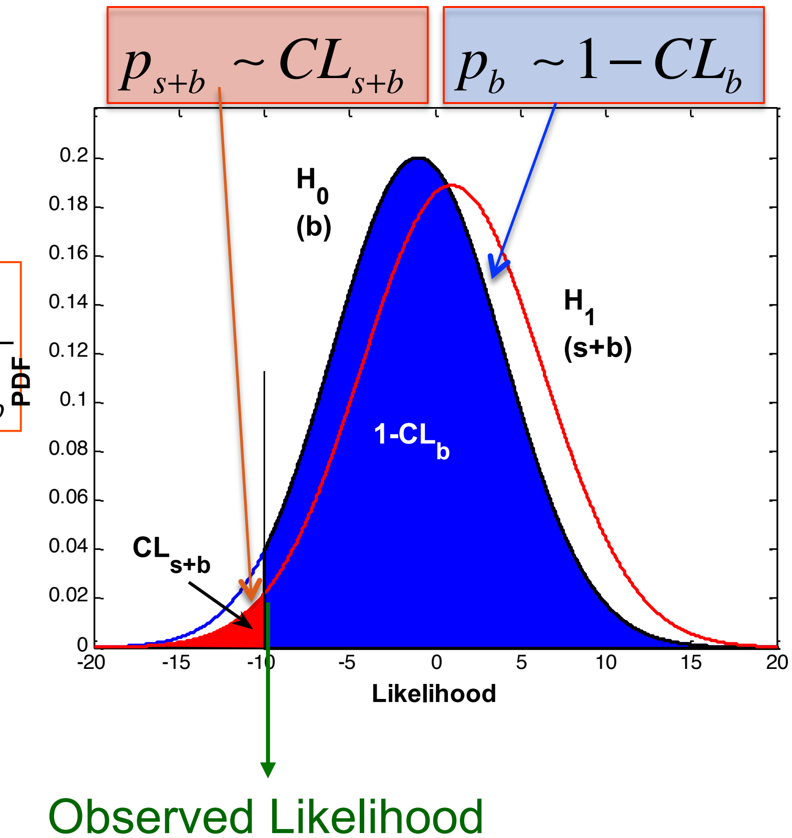
- Inspired by Zech(Roe and Woodroofe)' s derivation for counting experiments

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

- A. Read suggested the CL_s method with

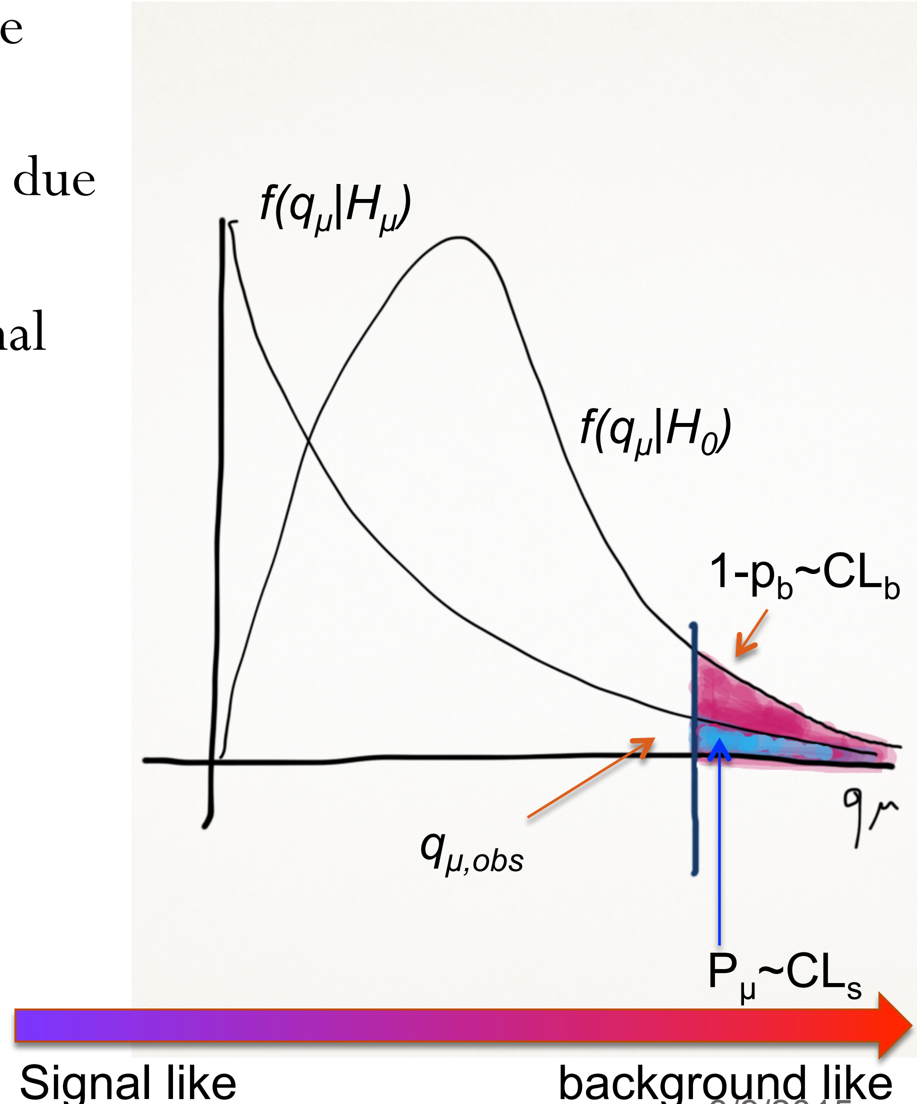
$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{p_{s+b}}{1-p_b}$$

- In the DELPHI example, CLs=0.03/0.13=0.26, i.e. a 116 GeV could not be excluded at the 97% CL anymore….. (p_b=1-CL_b=0.87)



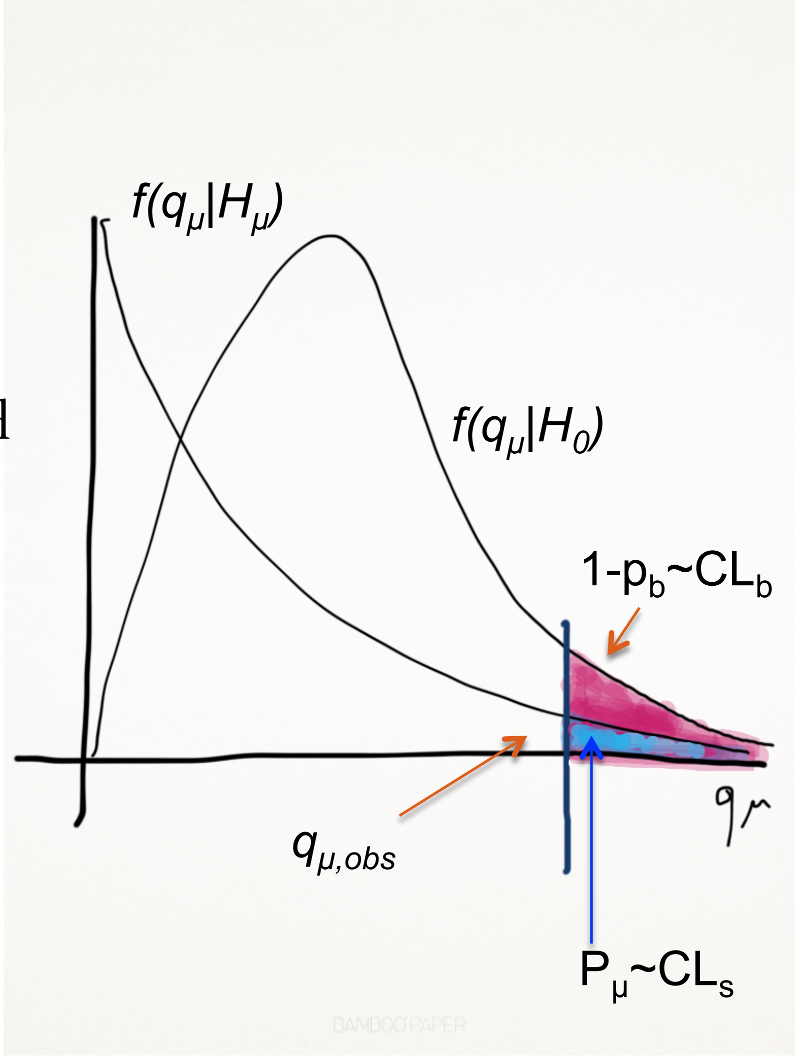
Second Verse- Same as the First (PL) CLb

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



CLs

- A complication arises when $\mu s+b \sim b$
- When the signal cross section is very small the $s(m_H)+b$ 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



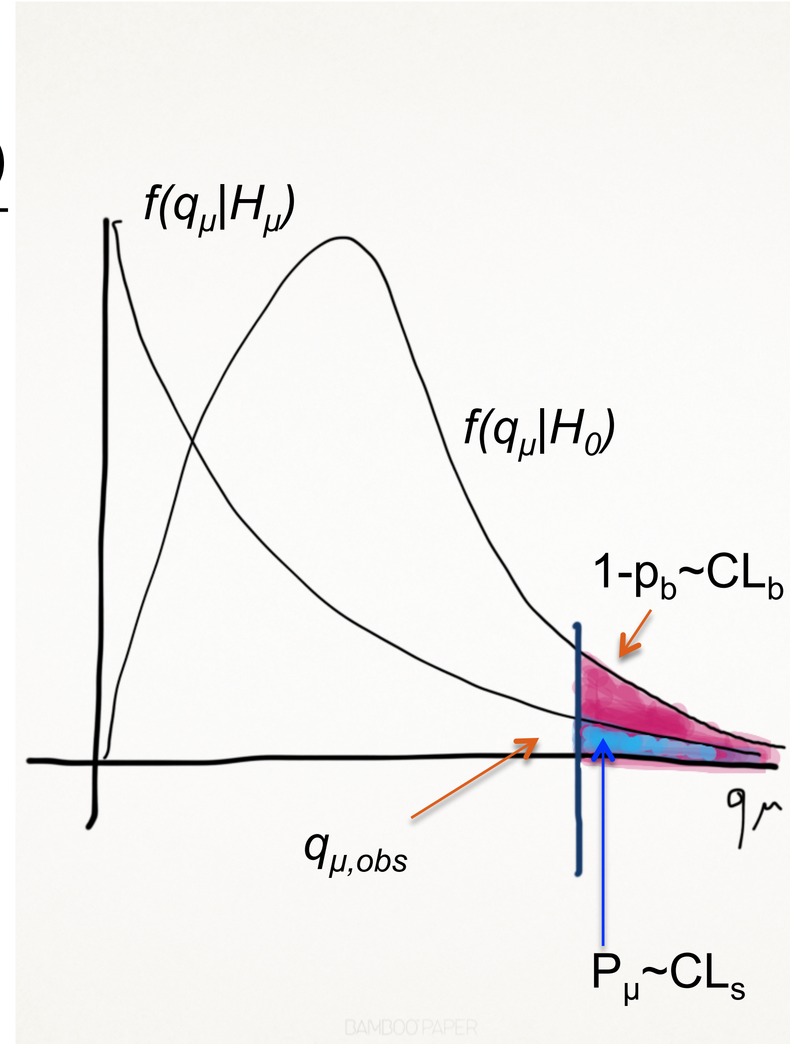
- Inspired by Zech derivation for counting experiments

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

- A. Read suggested the CL_s method with

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

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



The Modified CLs with the PL test statistic

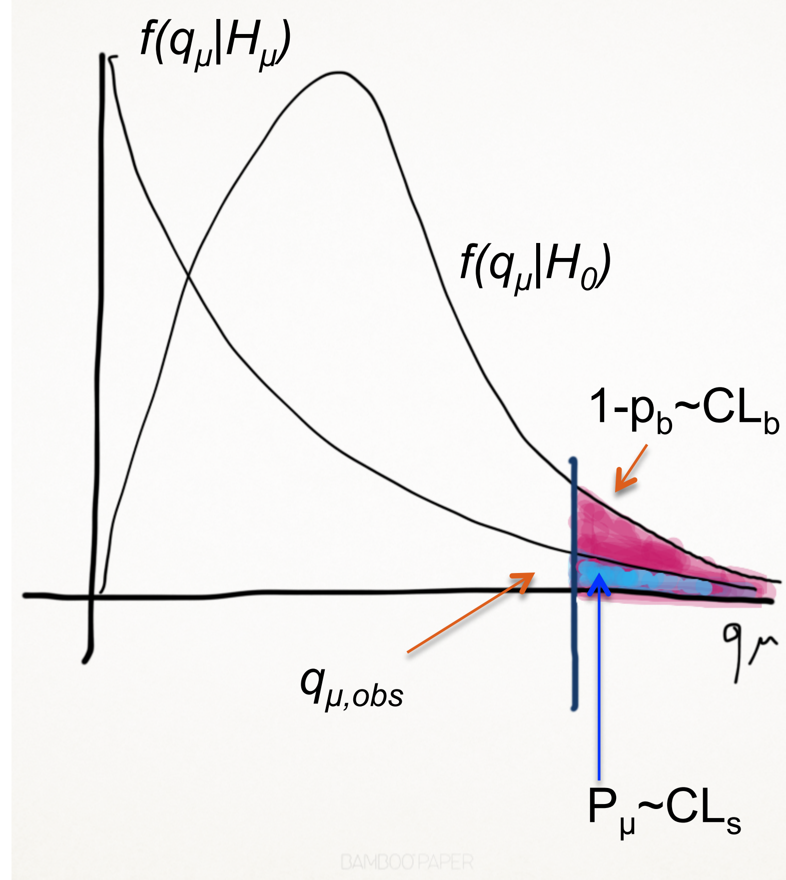
- The CLs method means that the signal hypothesis p-value

p_μ is modified to

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

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

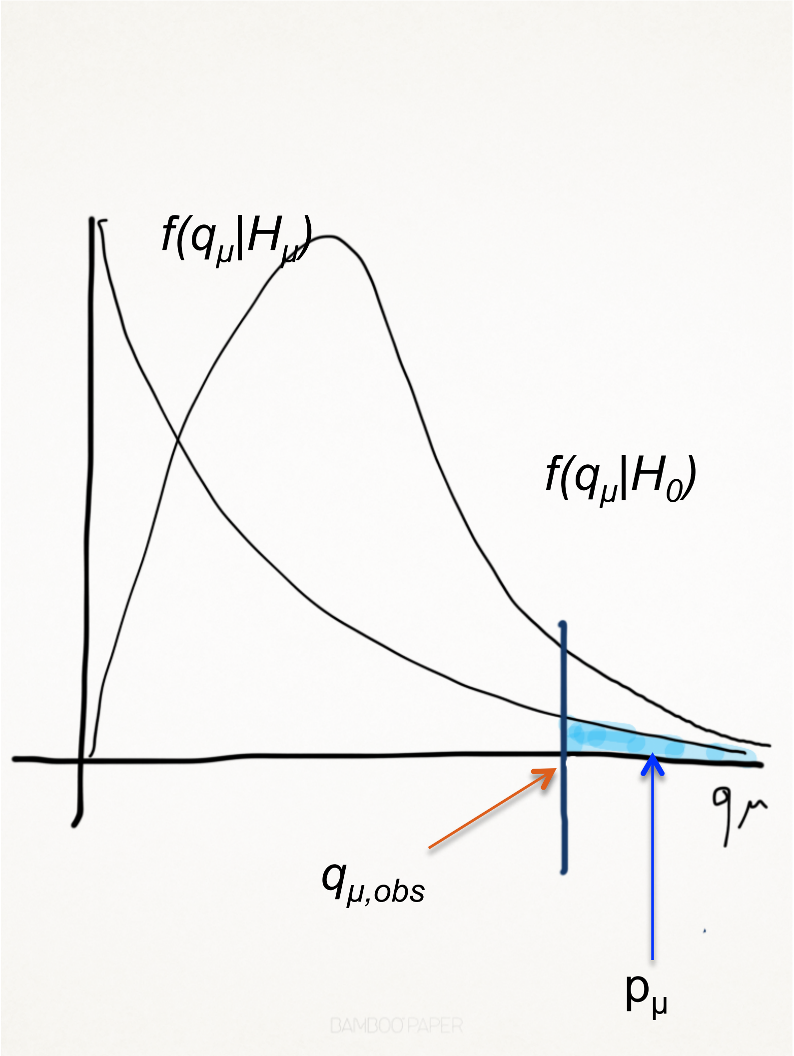
$$p_b = 1 - \int_{\tilde{q}_{\mu,obs}}^{\infty} f(\tilde{q}_\mu|0)d\tilde{q}_\mu$$



- 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_μ

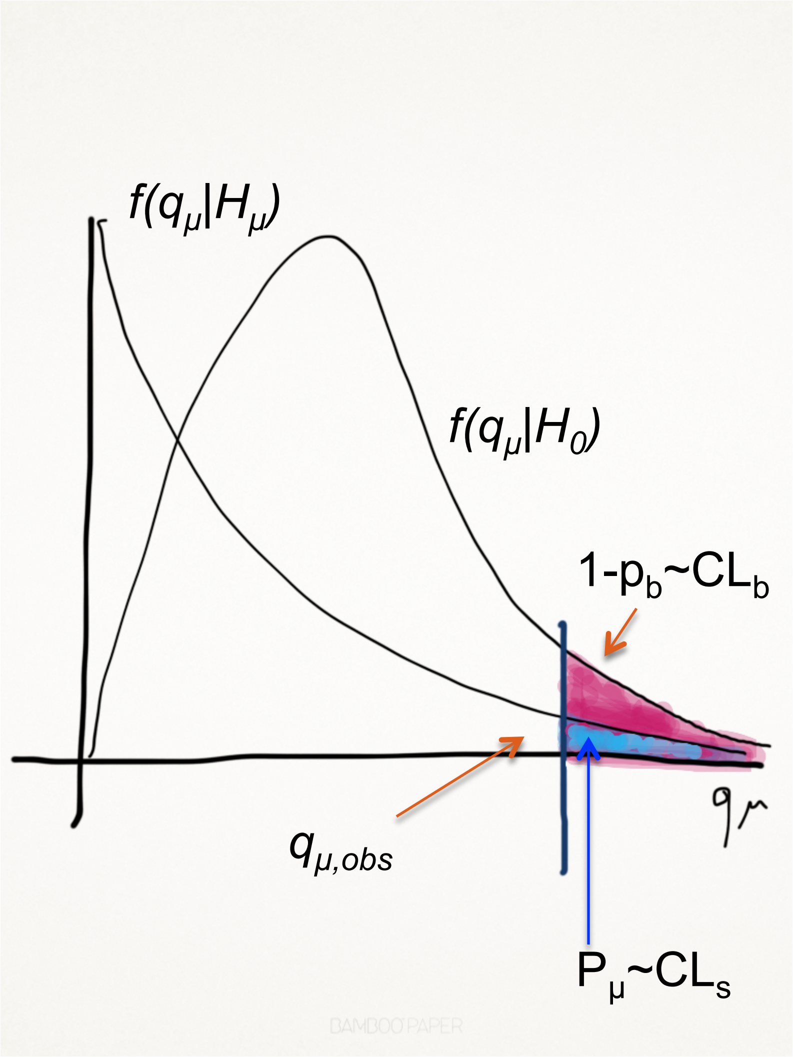
$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu \mid H_\mu)\, dq_\mu$$

- Find the modified p-value

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

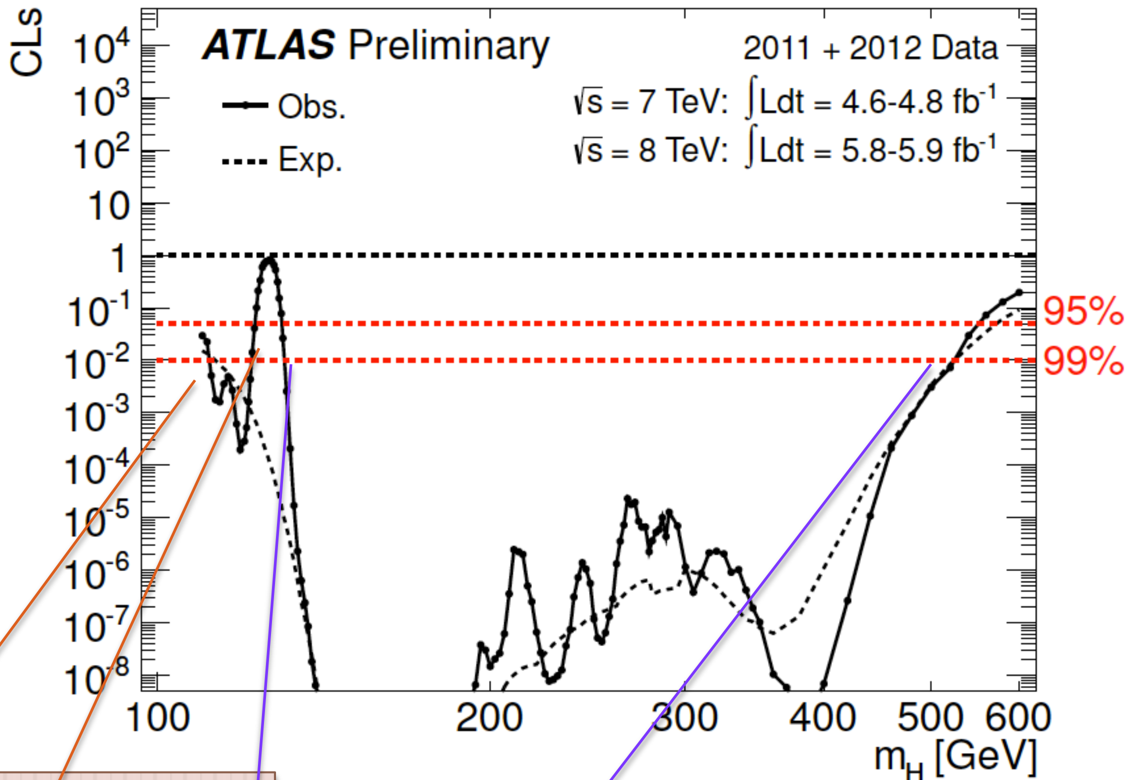
  - To tell if s is excluded, set $\mu = 1$ and find

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



Understanding the CLs plot

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

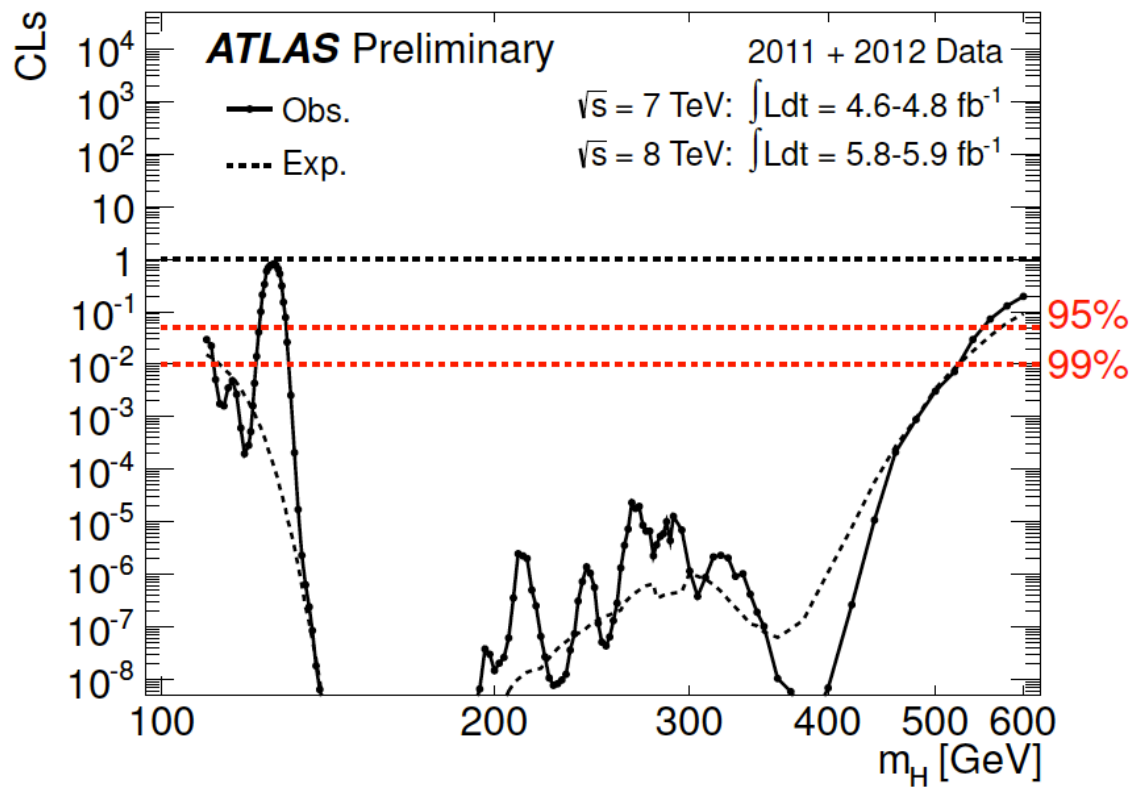


The smaller CLs, the deeper is the exclusion, Exclusion CL=1-CLs=1-p'_s

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



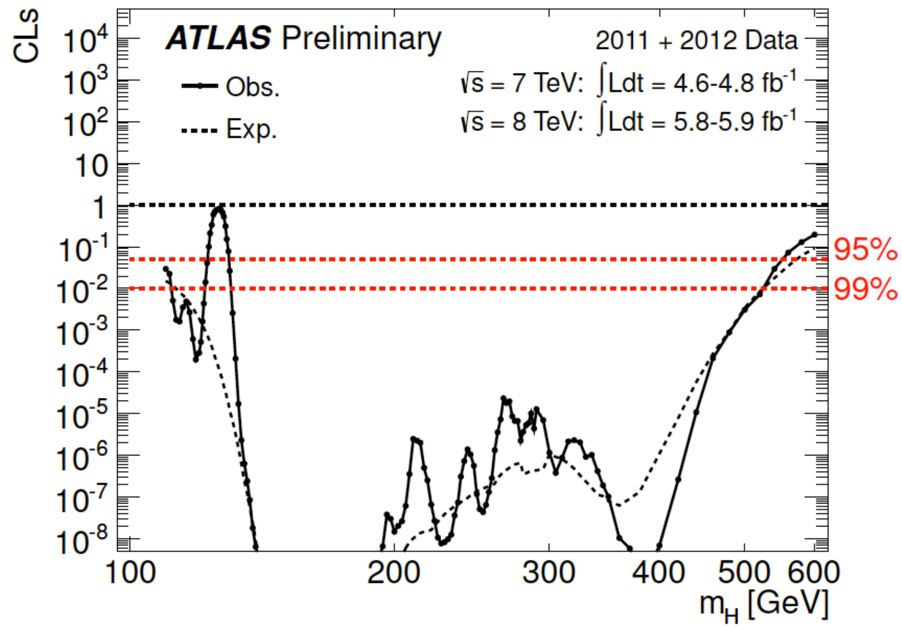
- Find the p-value of the signal h H_{μ}

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

- 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 $\mu_{up}<1$, m_H is excluded at the 95%

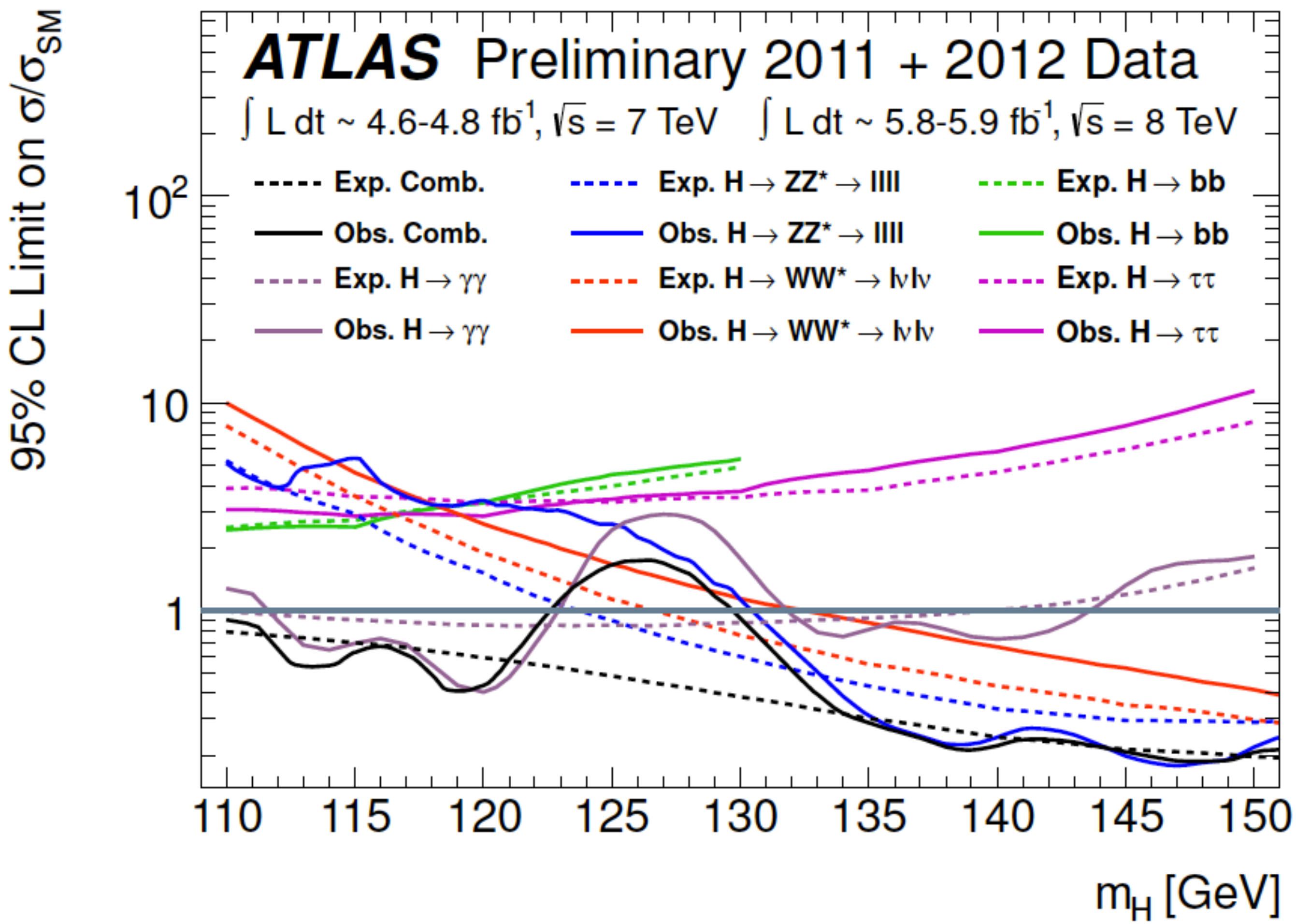


For a given data set, in the absence of a signal, the bigger the tested μ is the exclusion is deeper i.e. p'_{μ} is smaller

Exclusion a Higgs with a mass m_H

- First we fix the hypothesized mass to m_H
- We then test the H_μ [μ $s(m_H)+b$] 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

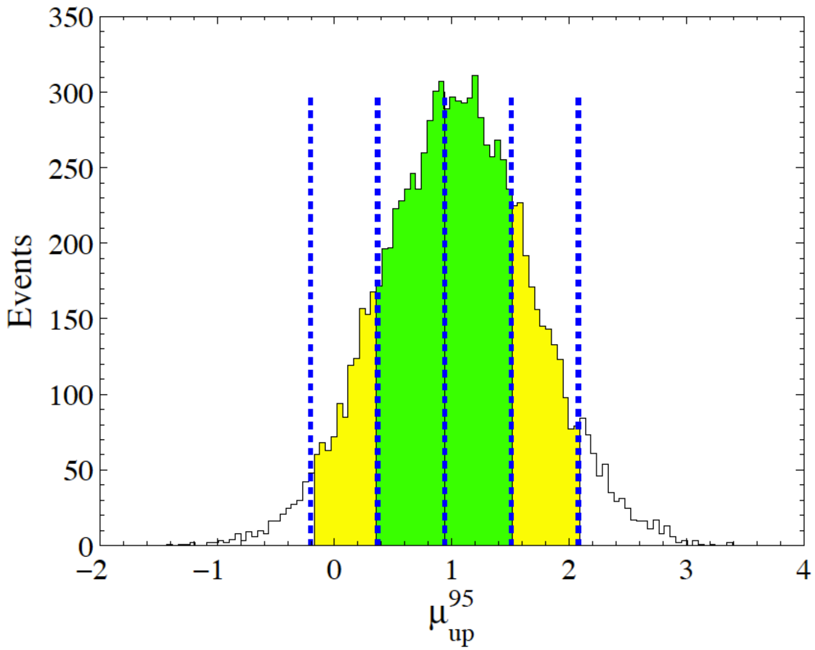
Upper Limit – $\mu_{up}(m_H)$



Sensitivity

- The sensitivity of an experiment to exclude a Higgs with a mass m_H is the 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=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

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

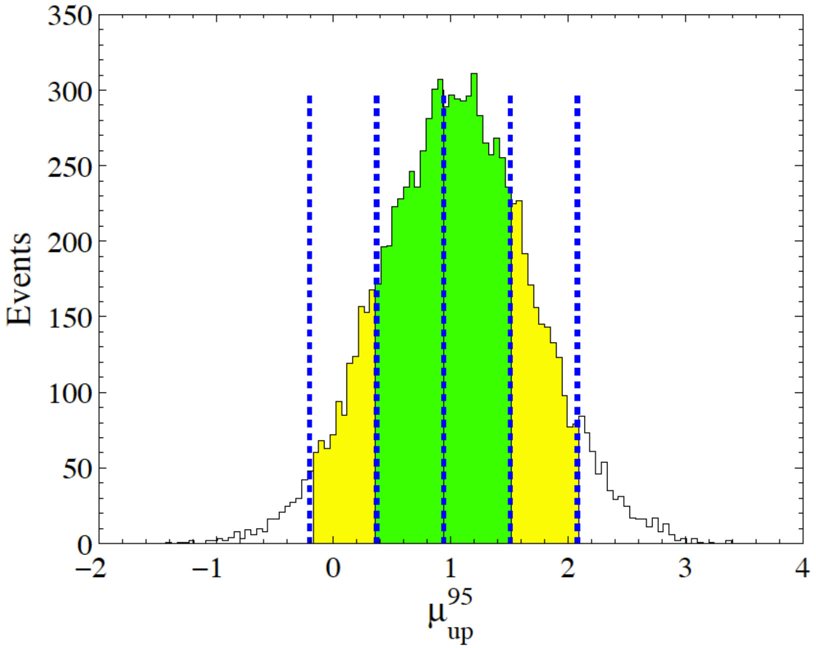
$$\sigma_{\hat{\mu}}^2 = Var[\hat{\mu}]$$

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

$$\alpha = 0.05$$

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

Distribution of the upper limit with background only experiments



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

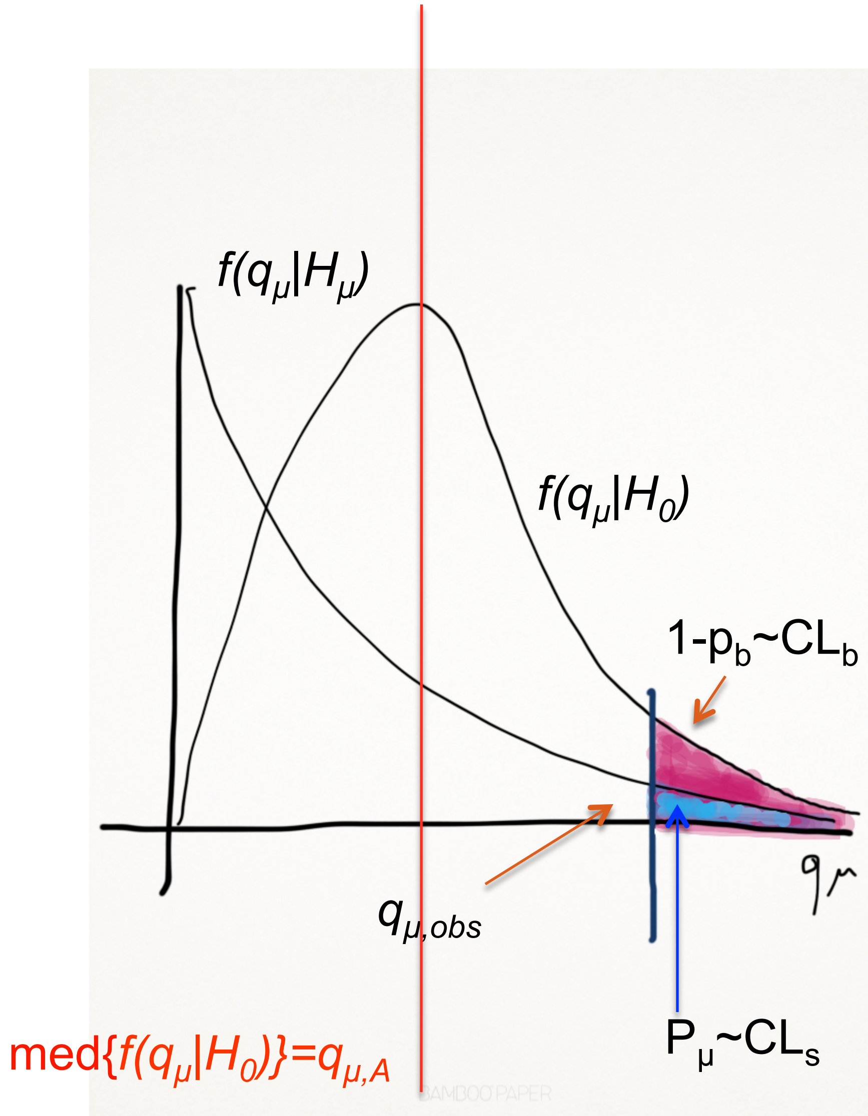
The Asimov data set

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

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

n=b

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



Eilam Gross, WIS, Statistics for PP

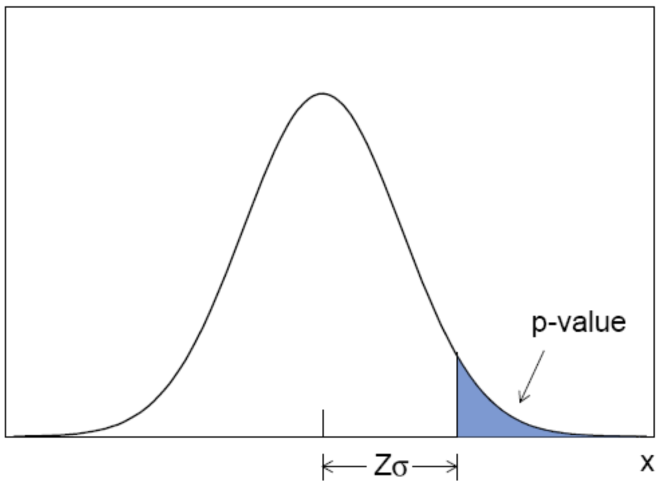
3/9/2015

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)

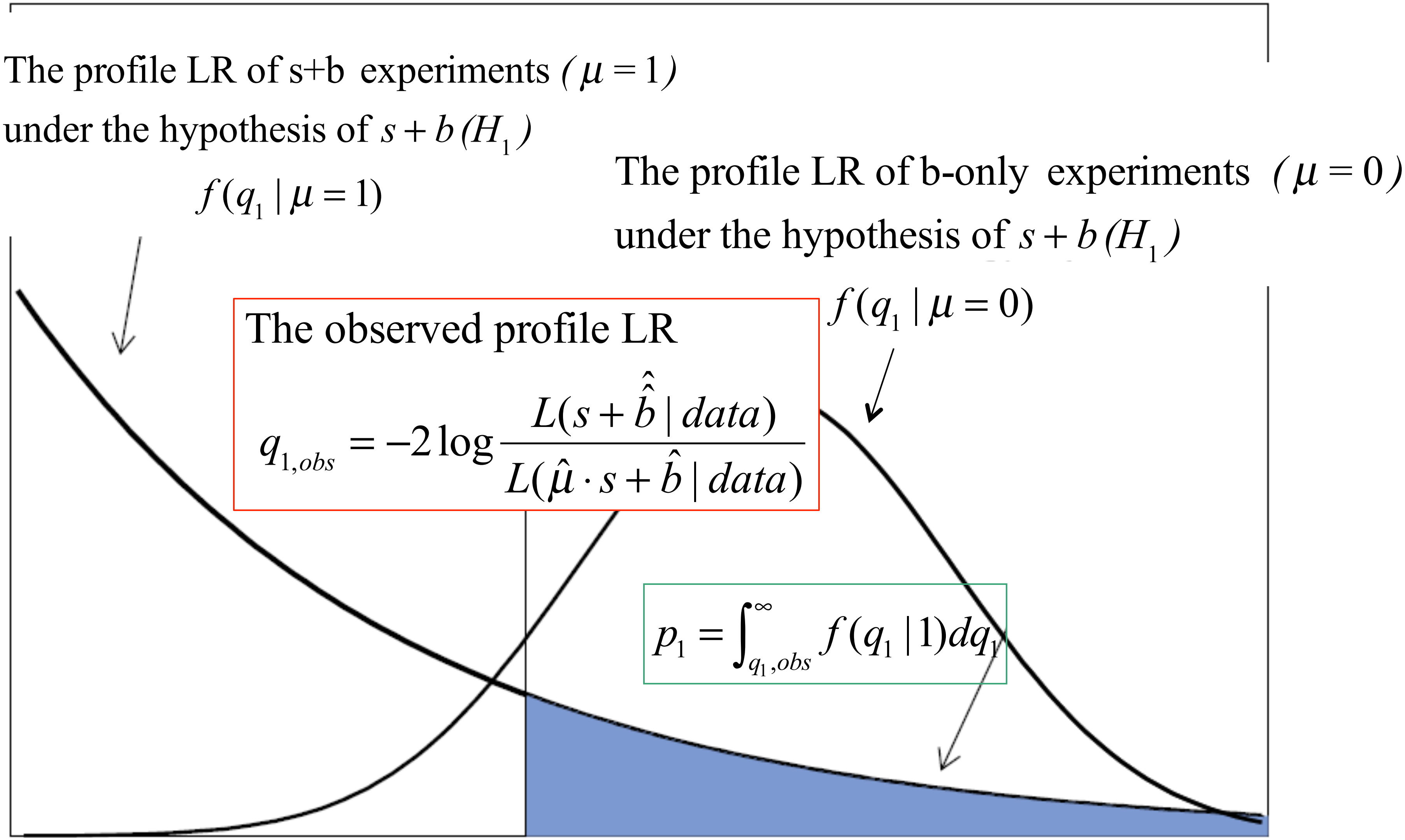


$$p = \int_Z^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 1 - \Phi(Z)$$

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

Exclusion – Illustrated

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



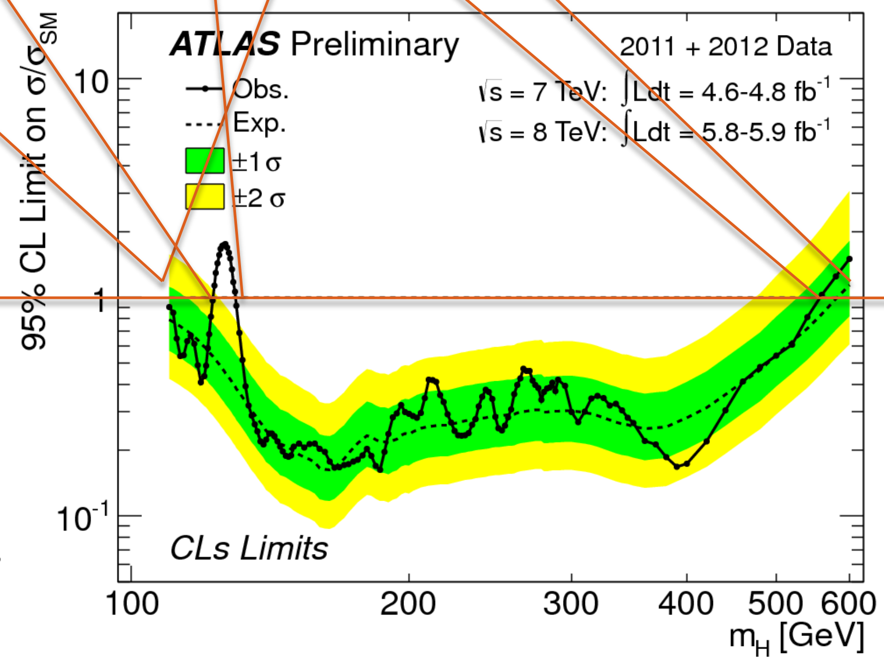
p_1 is the level of compatibility between the data and the Higgs hypothesis

If p_1 is smaller than 0.05 we claim an exclusion at the 95% CL

Understanding the Brazil Plot

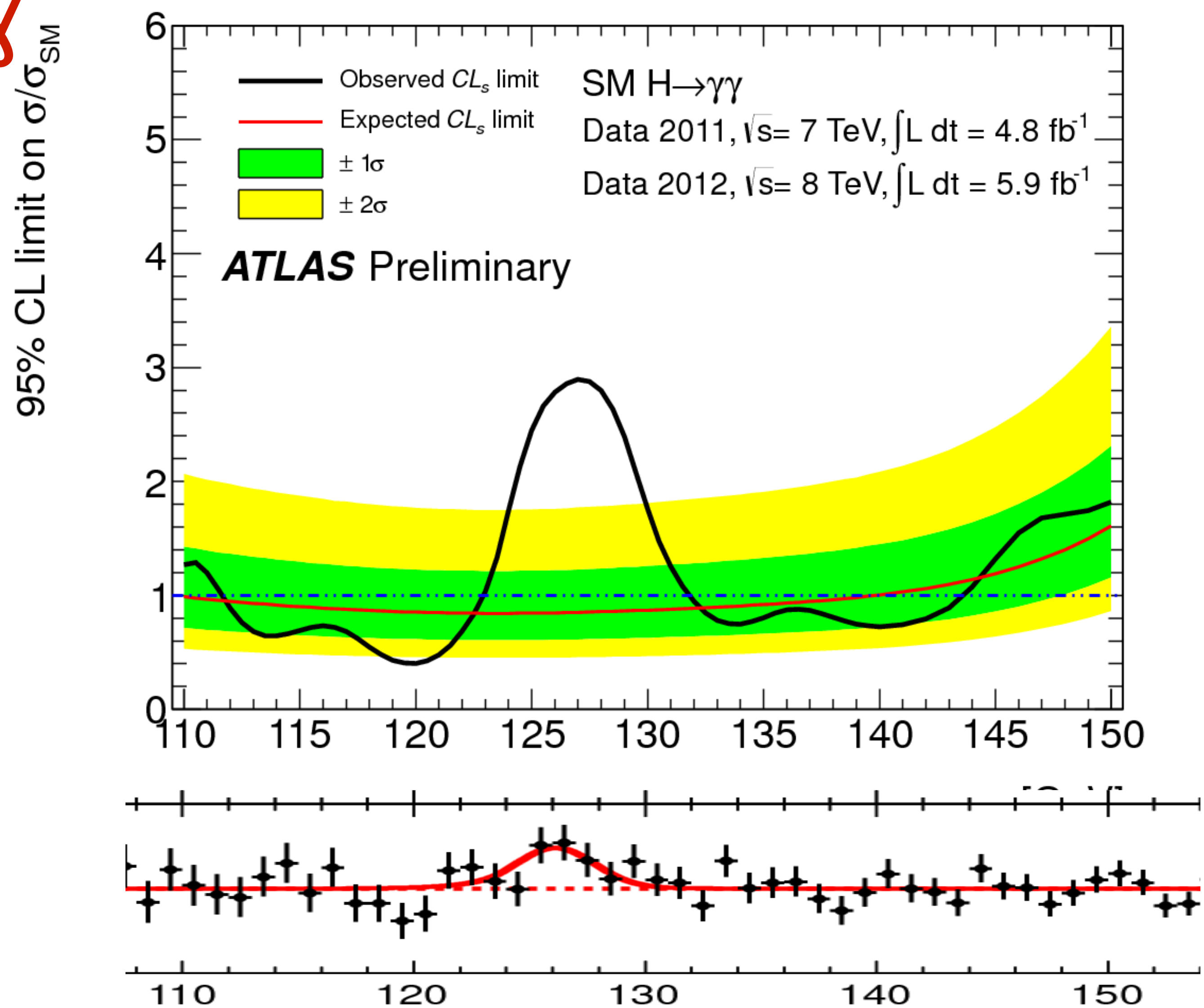
The expected 95% CL exclusion region covers the m_H range from 110 GeV to 582 GeV. The observed 95% CL exclusion regions are from 110 GeV to 122.6 GeV and 129.7 GeV to 558 GeV. The addition of

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

Search and Discovery Statistics in HEP
Lecture 3: p0, Discovery and the LEE, Multidimensional PL & Measurements

Eilam Gross, Weizmann Institute of Science

This presentation would have not been possible without the tremendous help of the following people throughout many years

Louis Lyons, Alex Read, Glen Cowan ,Kyle Cranmer , Yonatan Shlomi
Ofer Vitells & Bob Cousins

DISCOVERY

Case Study:
Higgs Discovery

Basic Definition: Signal Strength

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

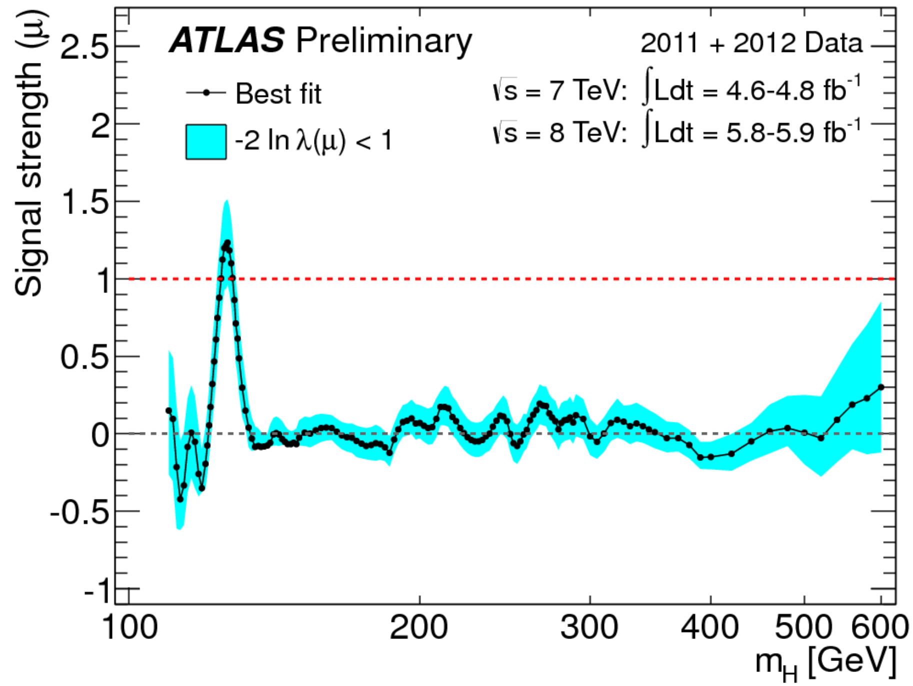
$$\mu(m_H) = \frac{\sigma(m_H)}{\sigma_{SM}(m_H)}$$

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

Introducing the Heartbeat

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

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



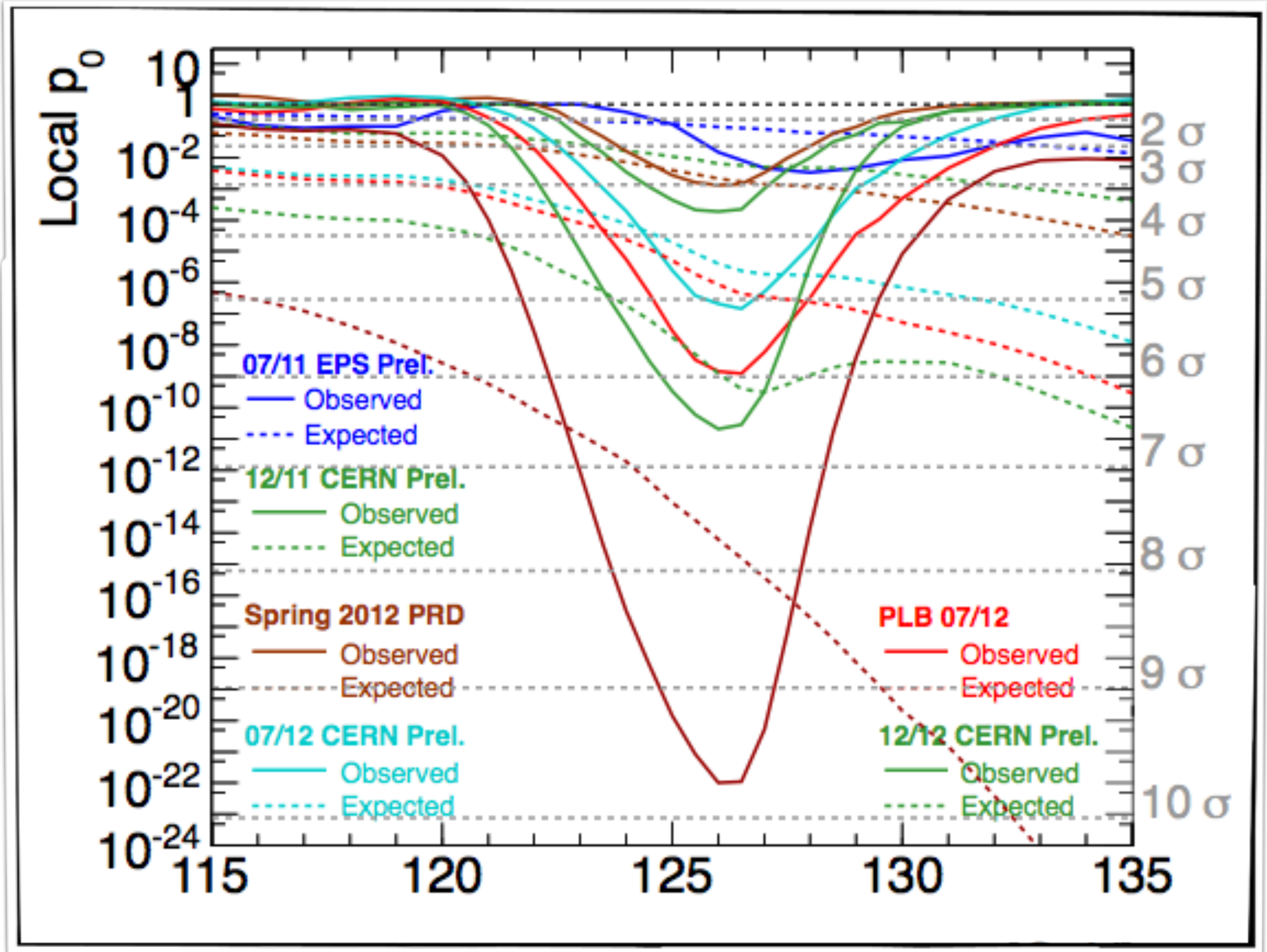
Reminder: The test statistic

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

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

- 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

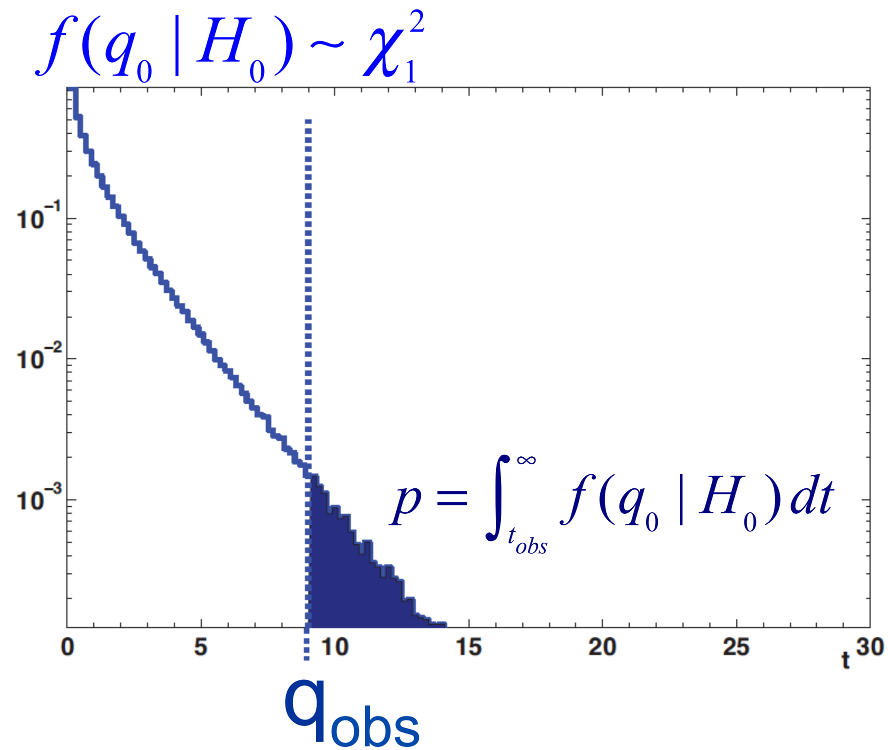
p0



Local p_0
10
1
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
10^{-20}
10^{-22}
10^{-24}
115
120
125
130
135
2 σ
3 σ
4 σ
5 σ
6 σ
7 σ
8 σ
9 σ
10 σ
07/11 EPS Prel.
Observed
Expected
12/11 CERN Prel.
Observed
Expected
Spring 2012 PRD
Observed
Expected
07/12 CERN Prel.
Observed
Expected
PLB 07/12
Observed
Expected
12/12 CERN Prel.
Observed
Expected

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)



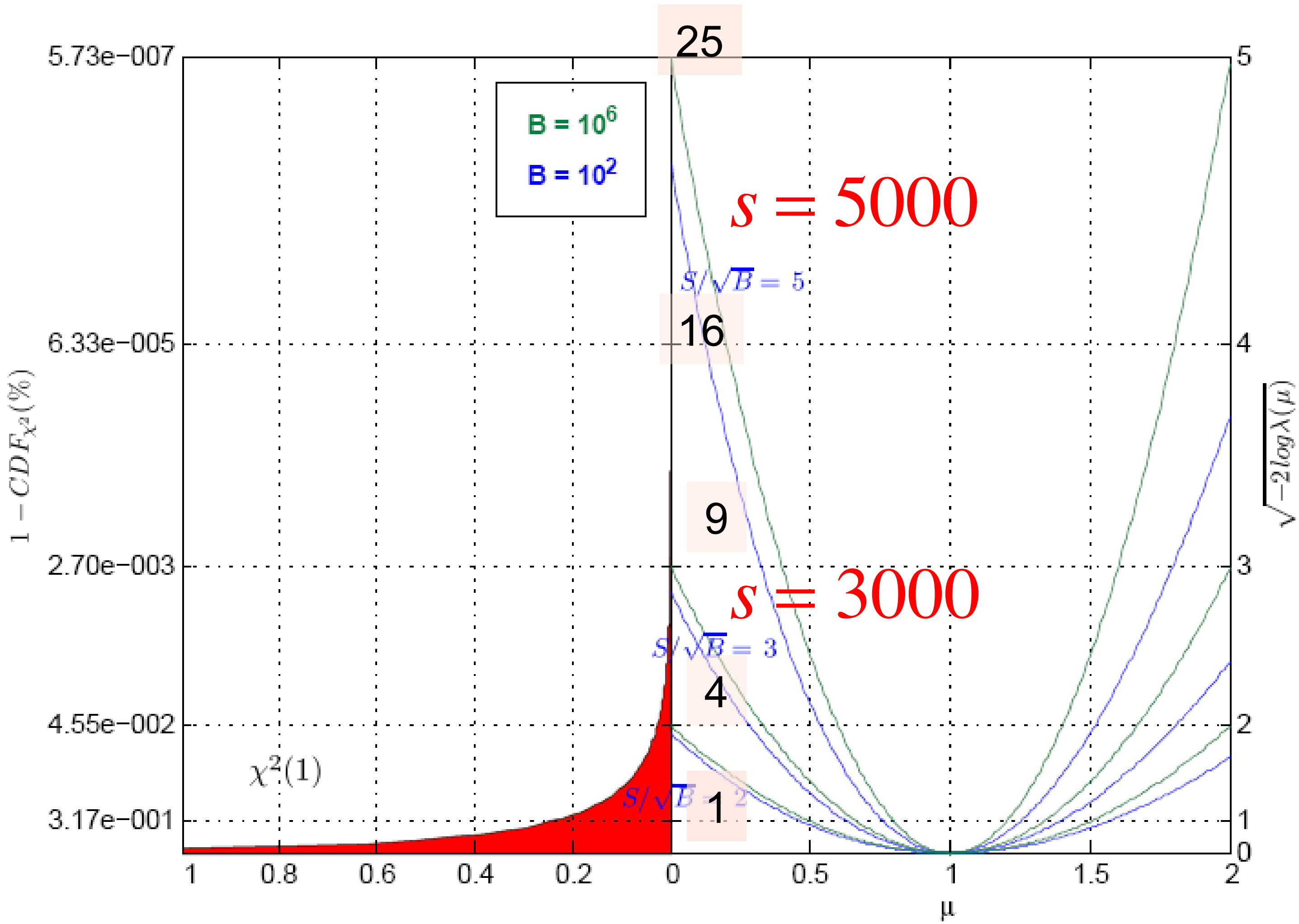
$$p = \int_{q_{obs}}^{\infty} f(q_0 | H_0)\,dt$$

If p-value< $2.8\cdot10^{-7}$, we claim a 5σ discovery

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

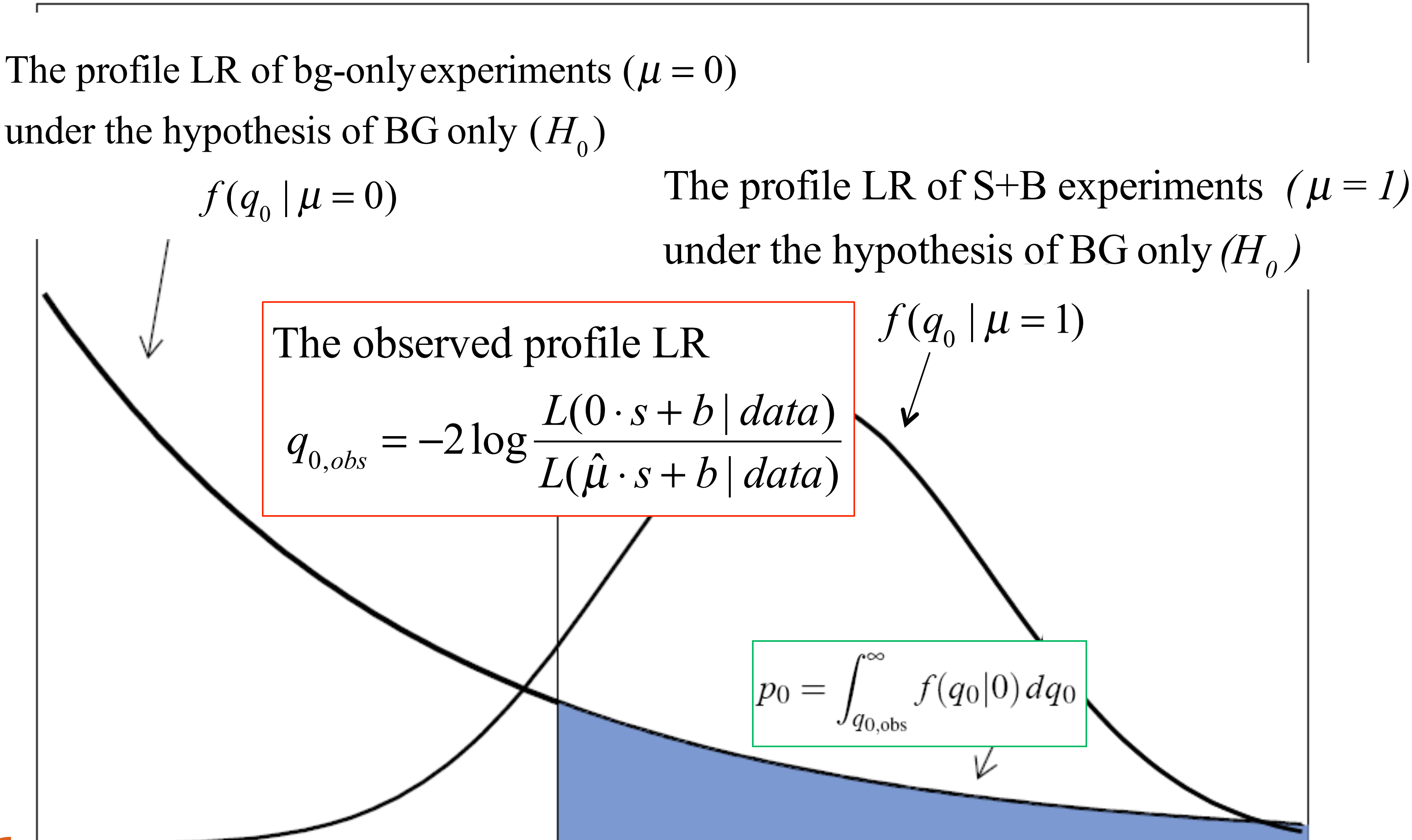
$$Z_{obs} = \sqrt{q_0} = \sqrt{q_0(\hat{\mu})}$$

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



Discovery - Illustrated

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

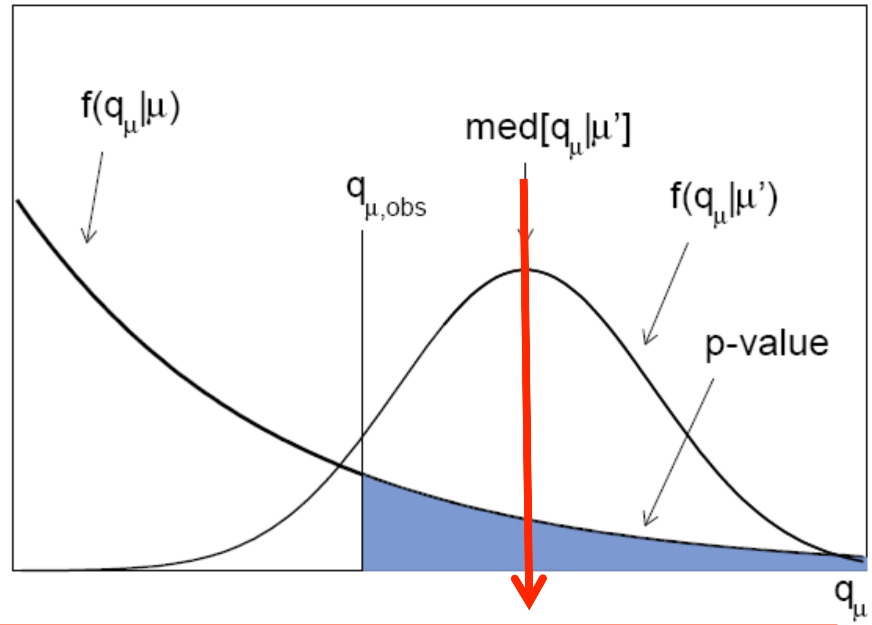


p_0 is the level of compatibility between the data and the no-Higgs hypothesis

If p_0 is smaller than $\sim 2.8\cdot 10^{-7}$ 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



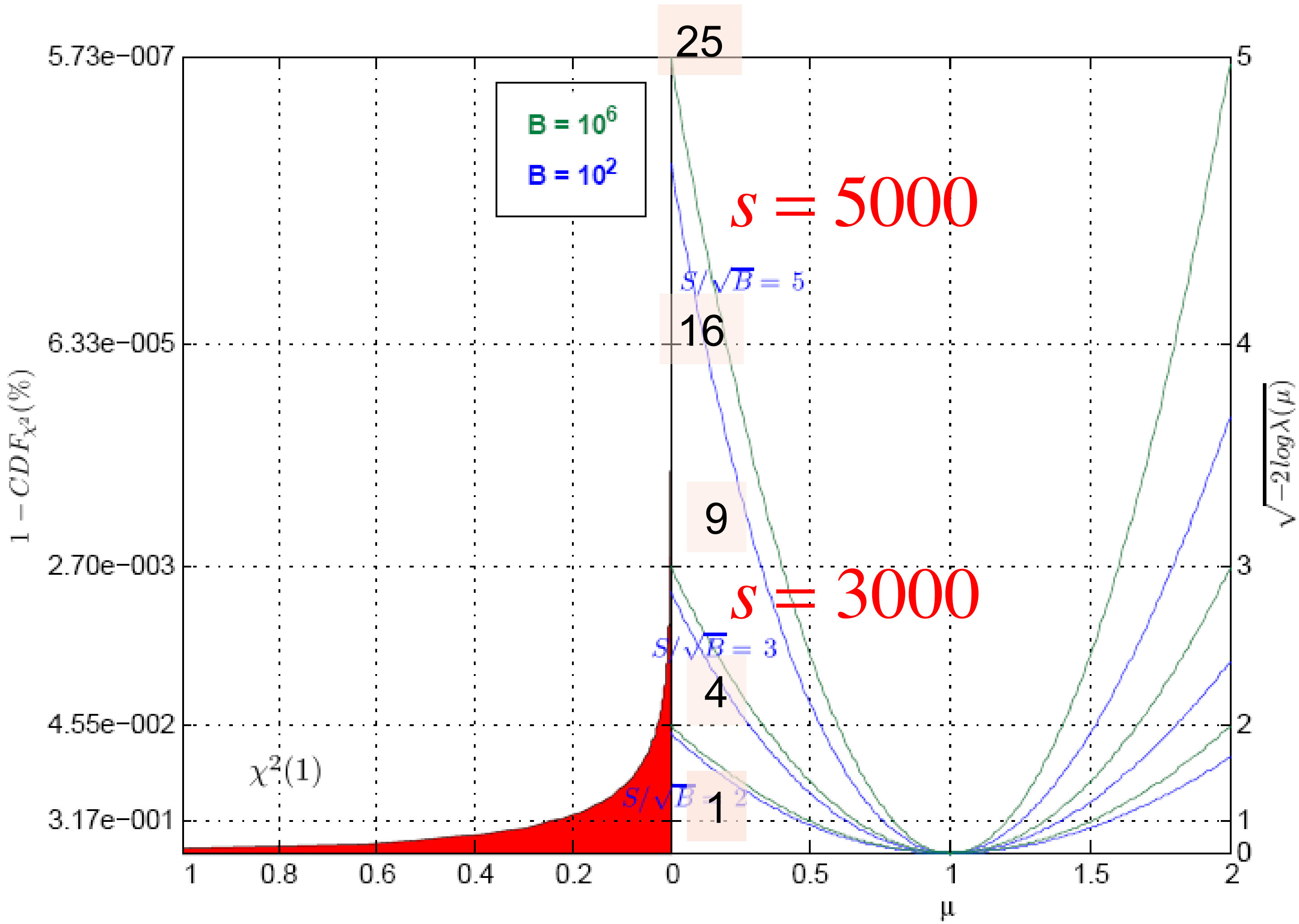
$$Z_{\mathrm{med}} = \Phi^{-1}(1 - p_{0_{\mathrm{med}}}) = \Phi^{-1}(1 - p_0(q_{0_{\mathrm{med}}}))$$

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

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

$$Z_{obs} = \sqrt{q_0} = \sqrt{q_0(\hat{\mu})}$$

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



The New $s/\sqrt{b}$

The new $s/\sqrt{b}$

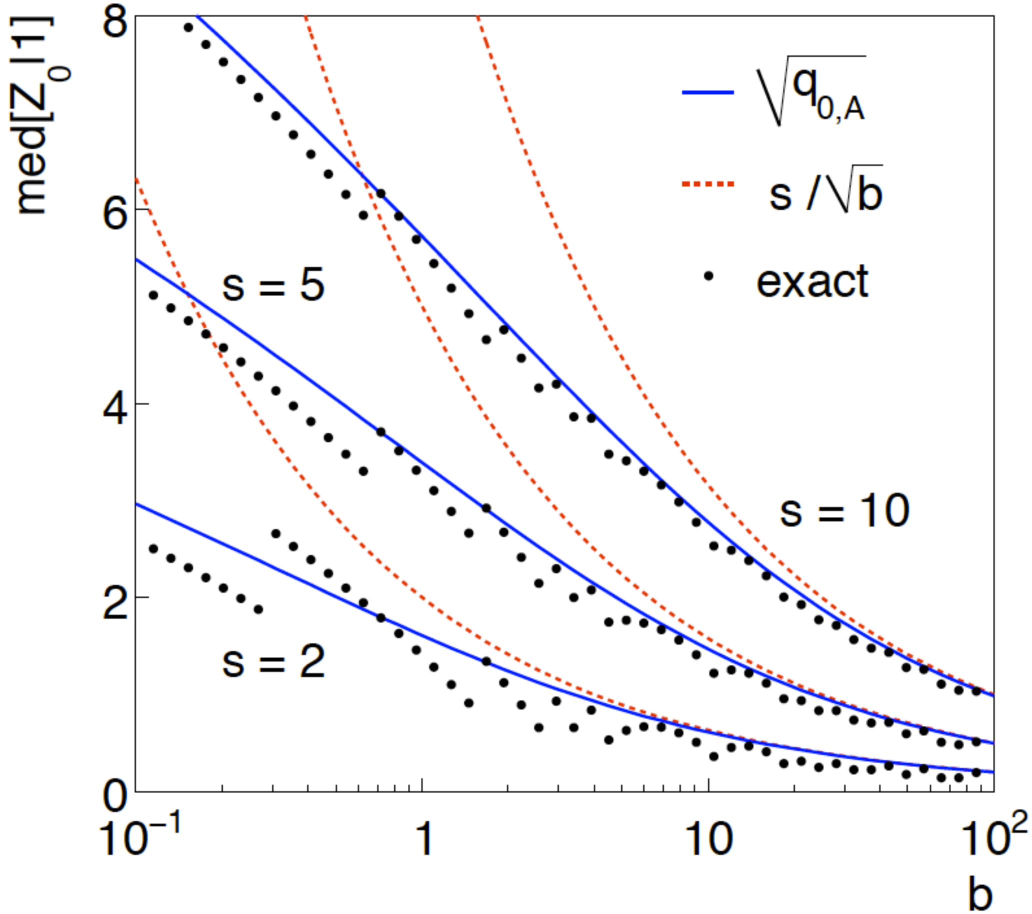
$$Z_A = \sqrt{q_{0,A}}$$

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

$$Z_A = \sqrt{q_{0,A}} \xrightarrow{s/b \ll 1} \frac{s}{\sqrt{b}} + O(s/b)$$

The New $s/\sqrt{b}$

$s/\sqrt{b}$?



The new $s/\sqrt{b}$

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

Taking Background Systematics into Account

- 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.,

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

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

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

If s/b<0.5 we will never be able to make a discovery

But even that formula can be omproved using the Asimov formalism

Significance with systematics

- We find (G. Cowan)

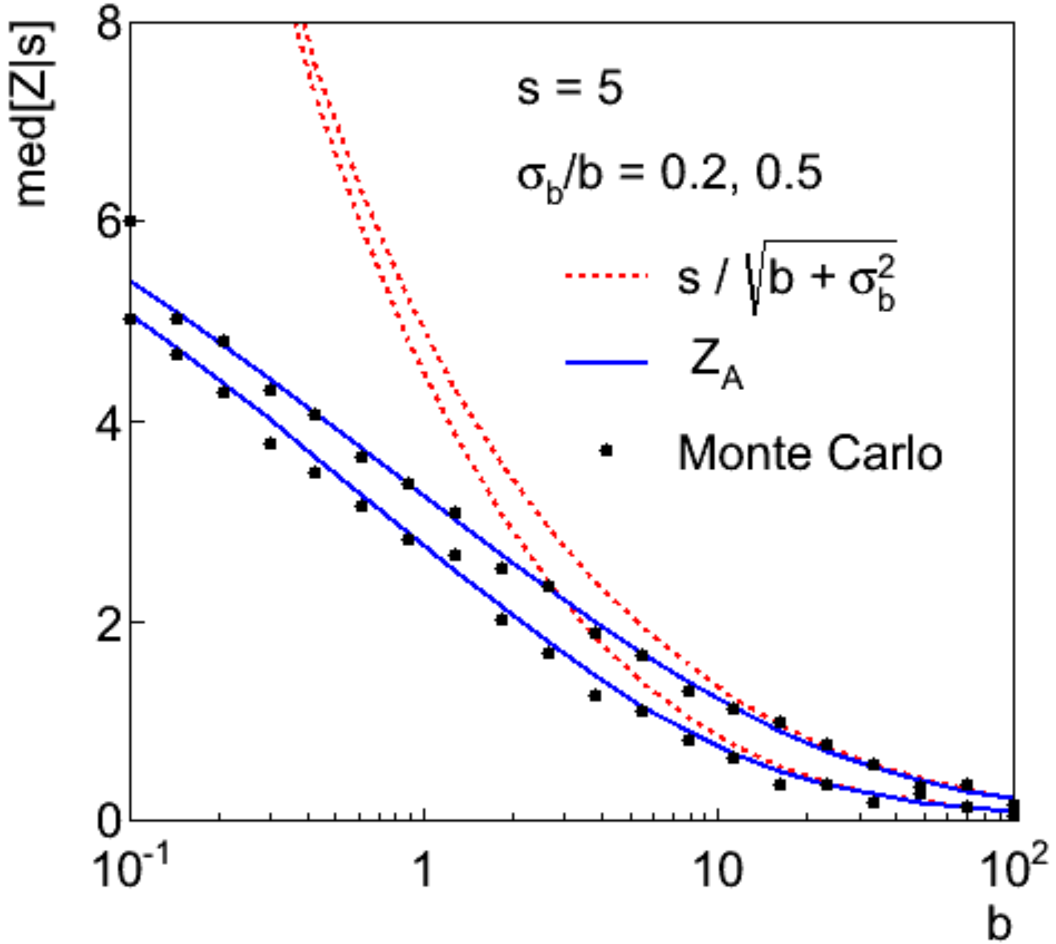
$$Z_{\mathrm{A}} = \left[2 \left((s+b) \ln \left[\frac{(s+b)(b+\sigma_b^2)}{b^2 + (s+b)\sigma_b^2} \right] - \frac{b^2}{\sigma_b^2} \ln \left[1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)} \right] \right) \right]^{1/2}$$

Expanding the Asimov formula in powers of s/b and $\sigma_b{}^2/b$ 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



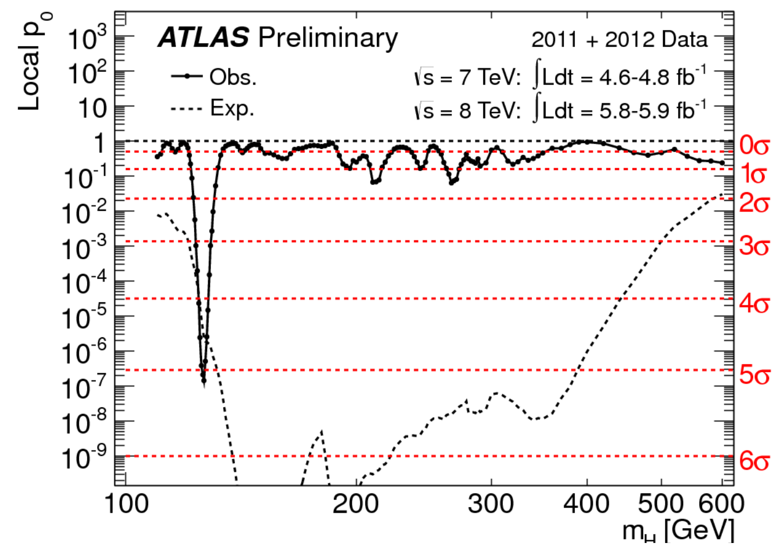
p0 and the expected p0

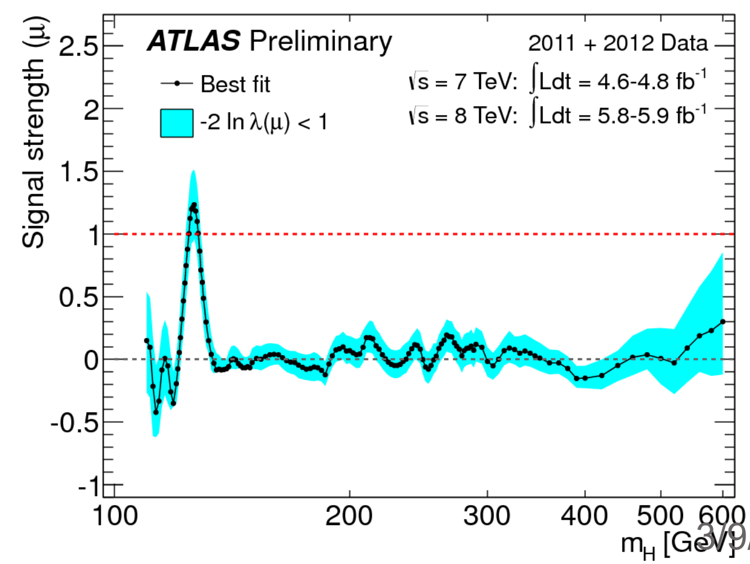
$$p_0 = \int_{q_{0,\text{obs}}}^{\infty} f(q_0|0)\, dq_0$$

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





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

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

Distribution of q0 (discovery)

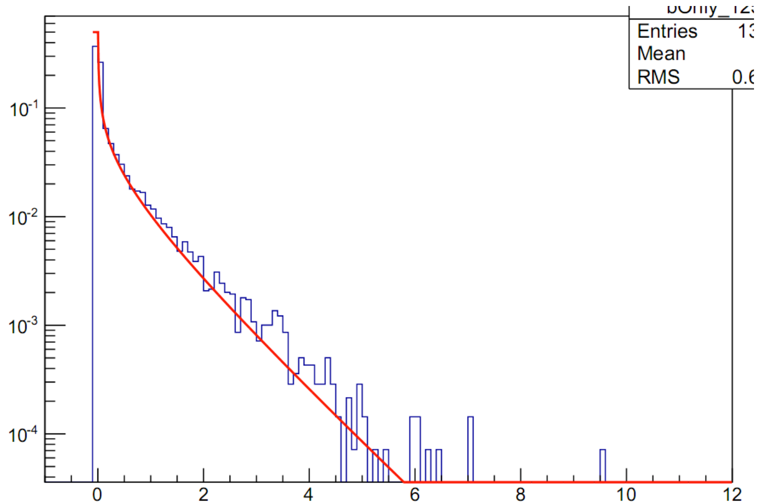
- We find

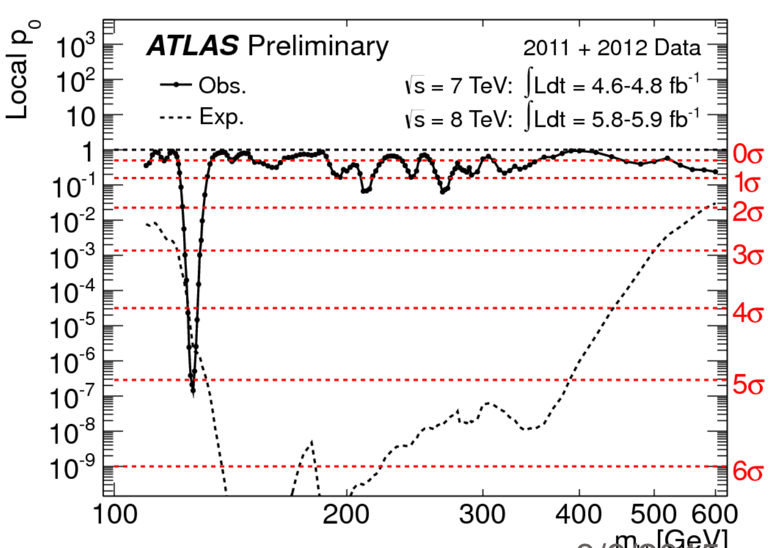
$$f(q_0|0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}e^{-q_0/2} .$$

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

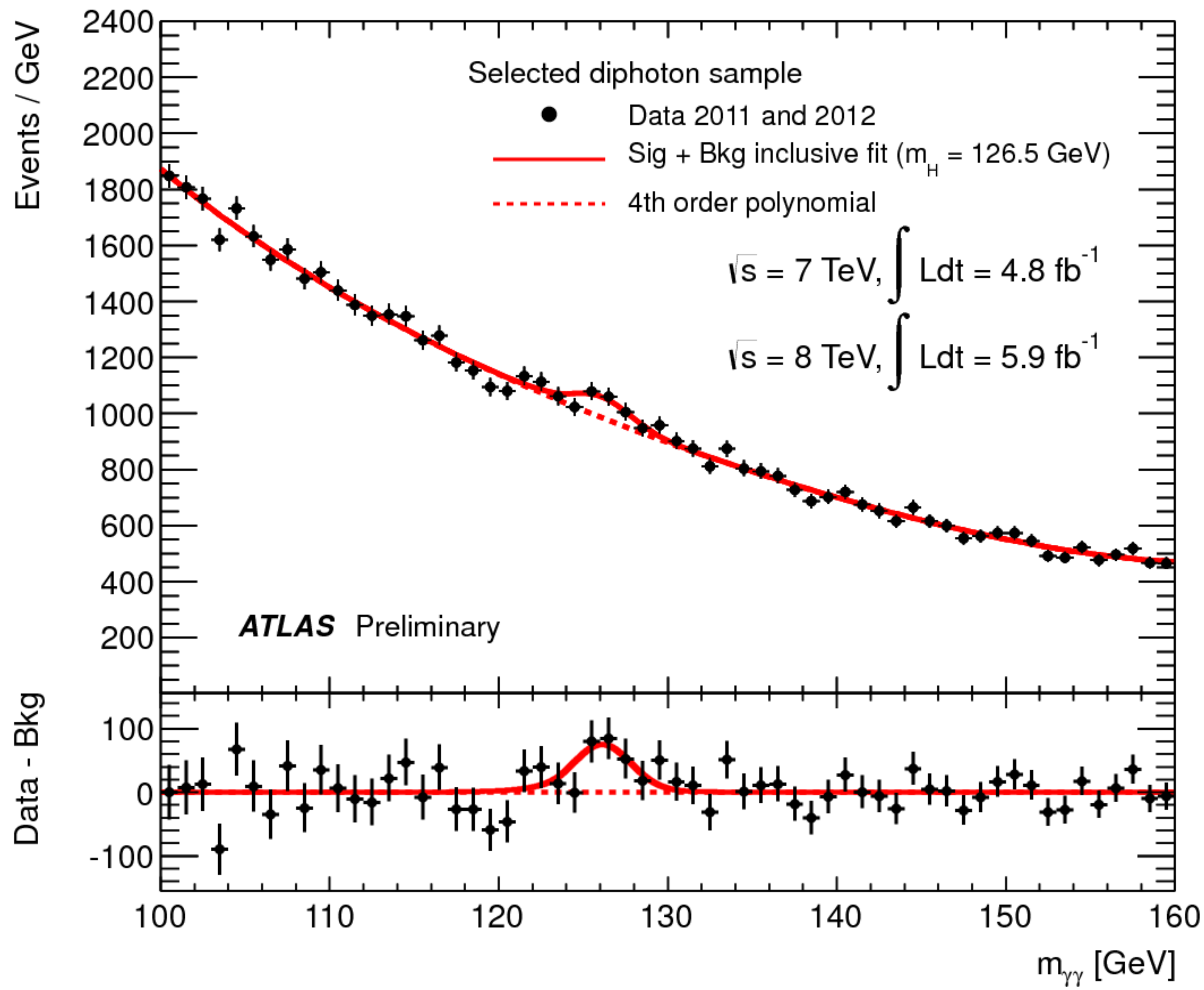
$$Z_0 = \Phi^{-1}(1 - p_0) = \sqrt{q_0} .$$

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



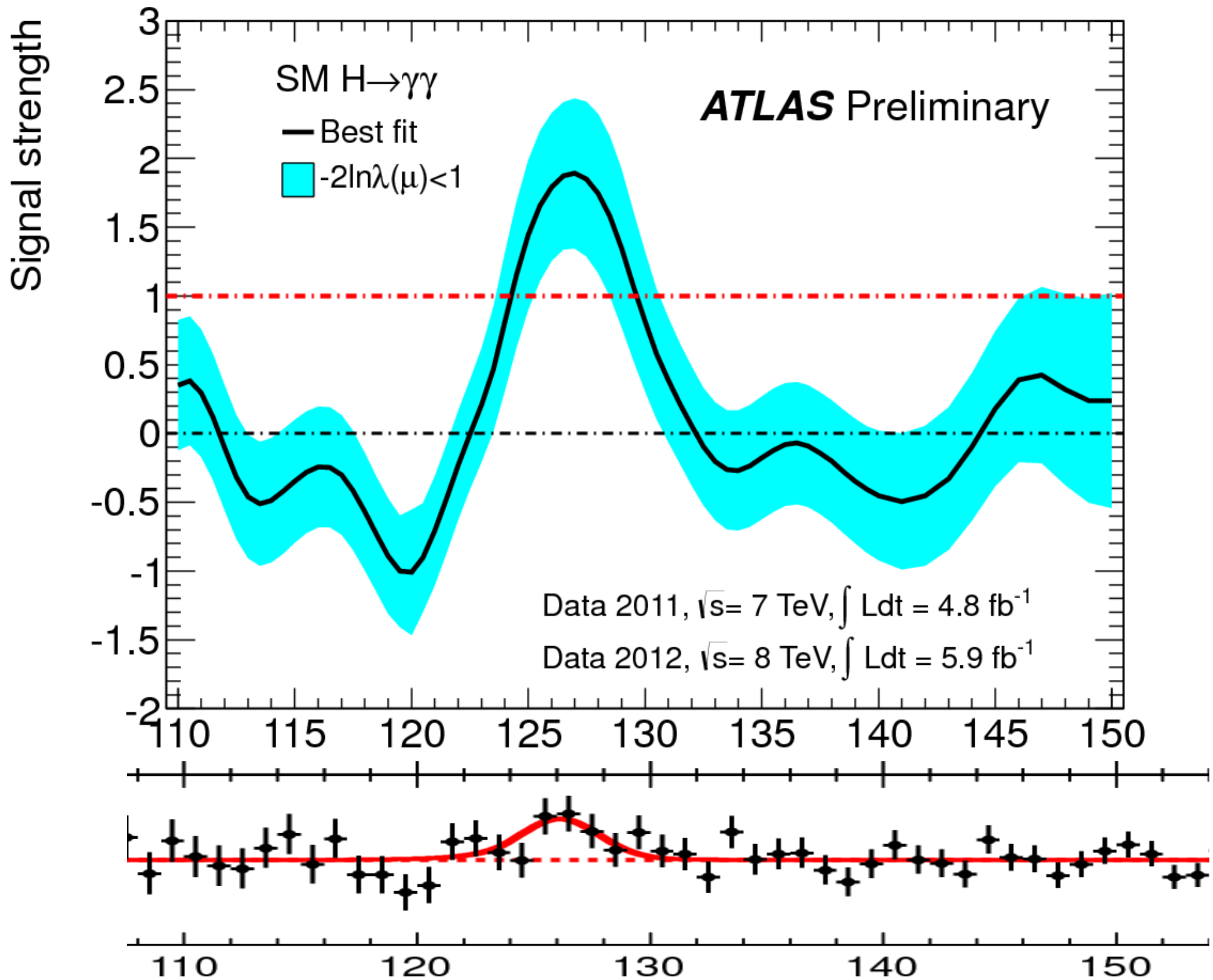


Example: H→γγ



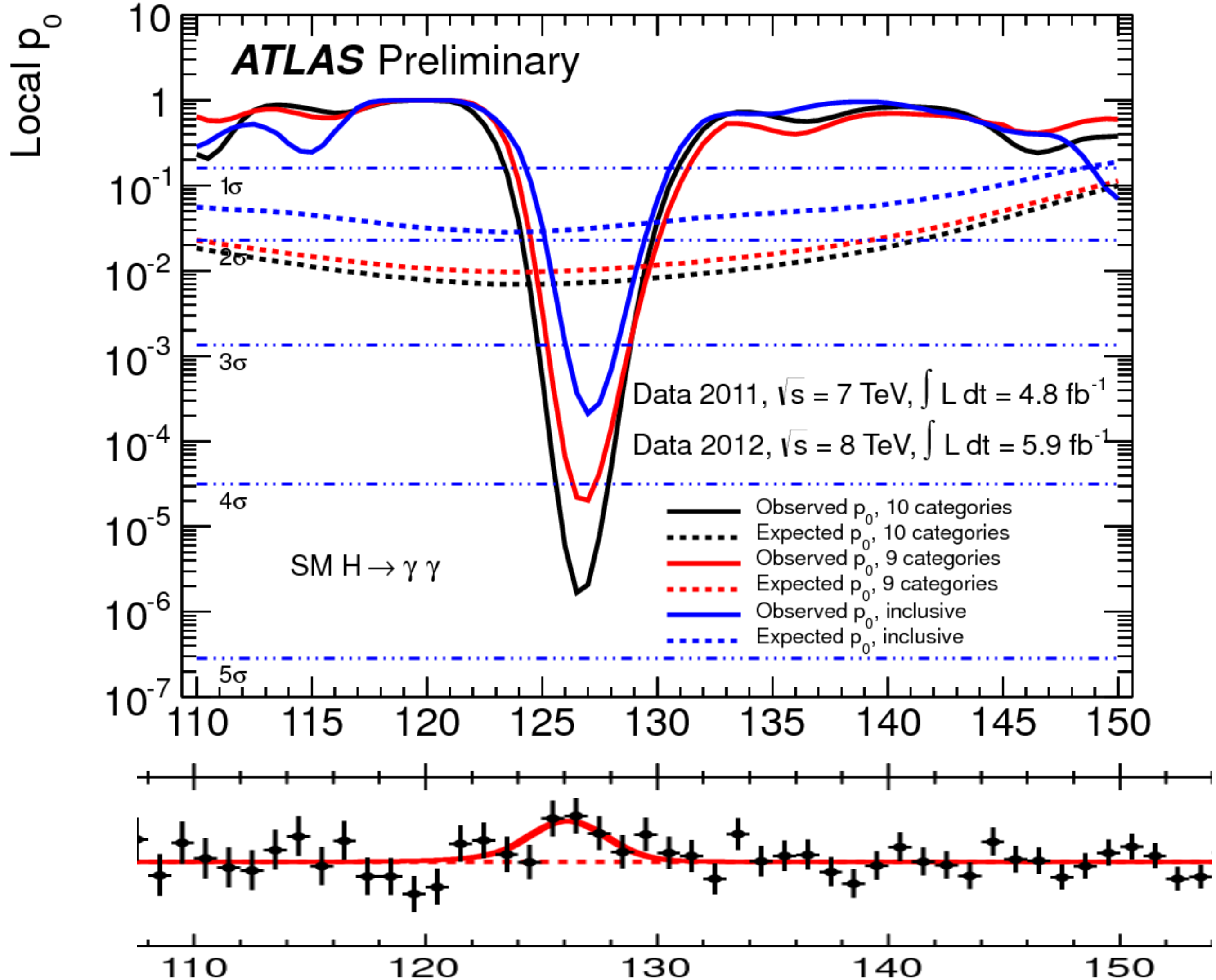
Events / GeV
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
Selected diphoton sample
Data 2011 and 2012
Sig + Bkg inclusive fit (m_H = 126.5 GeV)
4th order polynomial
$\sqrt{s}$ = 7 TeV, $\int$ Ldt = 4.8 fb^{-1}
$\sqrt{s}$ = 8 TeV, $\int$ Ldt = 5.9 fb^{-1}
ATLAS Preliminary
Data - Bkg
100
0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

H→γγ

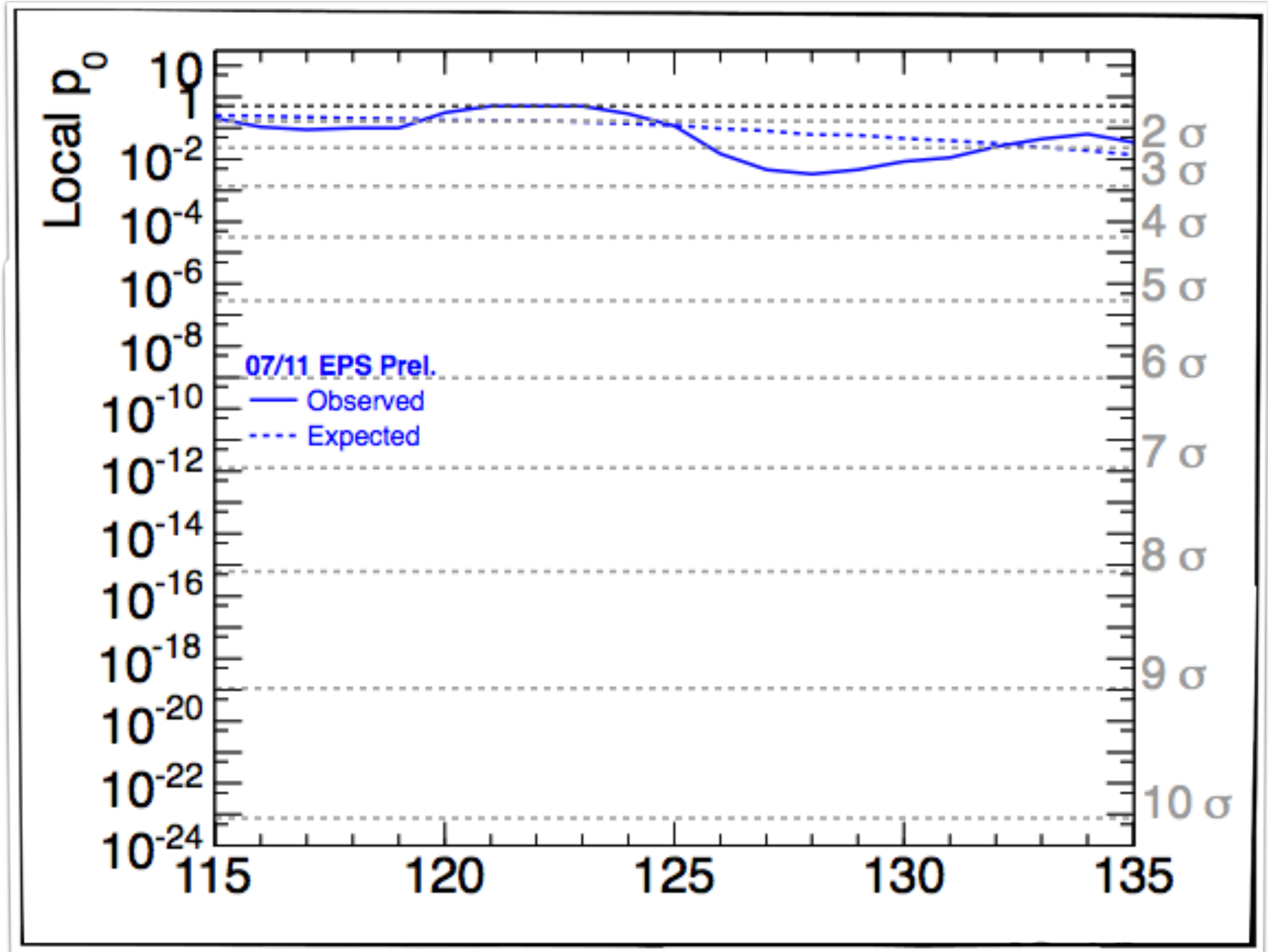


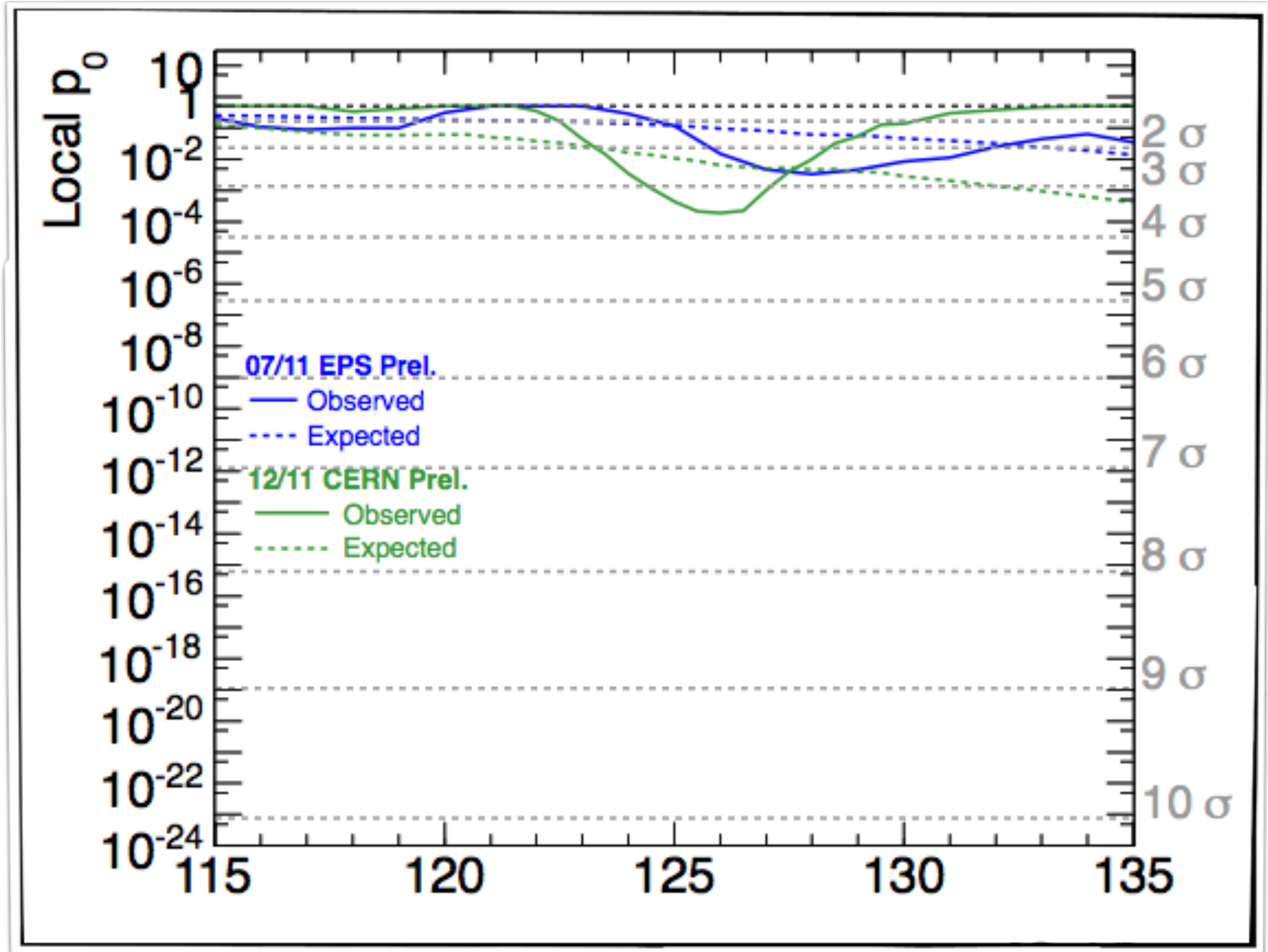
Signal strength
SM H→γγ
Best fit
-2lnλ(μ)<1
ATLAS Preliminary
Data 2011, √s= 7 TeV, ∫ Ldt = 4.8 fb⁻¹
Data 2012, √s= 8 TeV, ∫ Ldt = 5.9 fb⁻¹
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
110
115
120
125
130
135
140
145
150

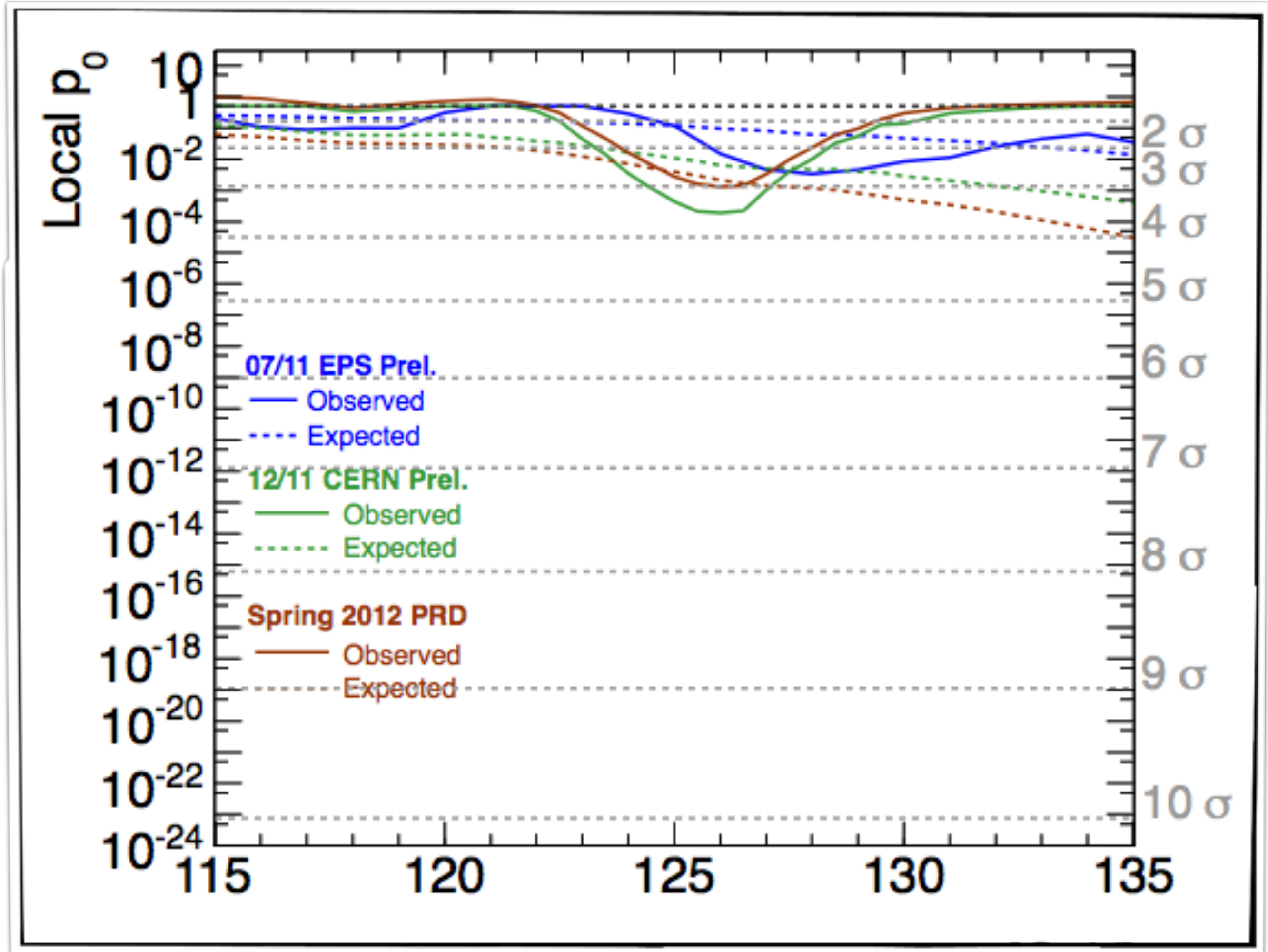
H→γγ

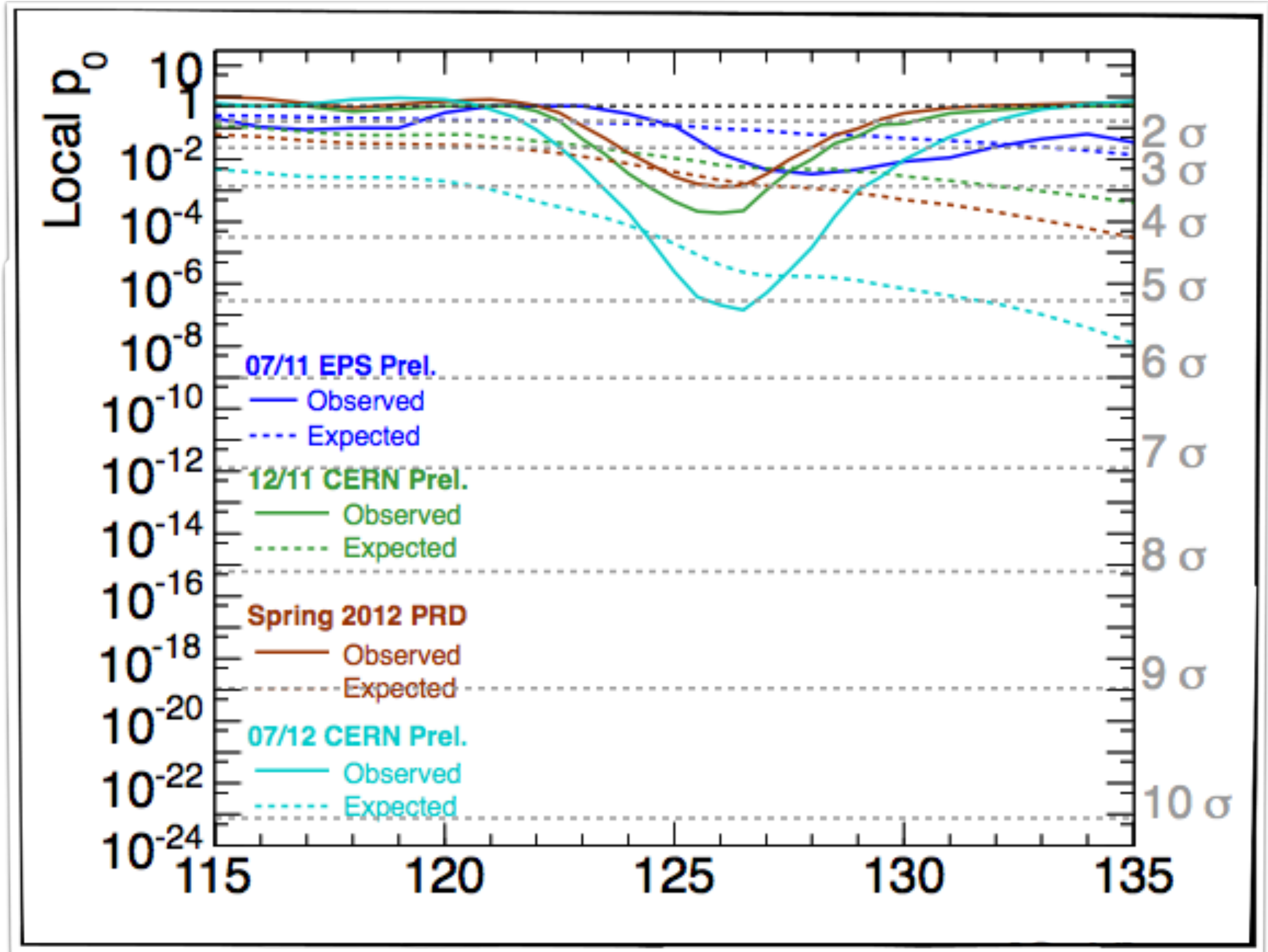


ATLAS Preliminary
Local p_0
Data 2011, $\sqrt{s}$ = 7 TeV, $\int L\,dt$ = 4.8 fb^{-1}
Data 2012, $\sqrt{s}$ = 8 TeV, $\int L\,dt$ = 5.9 fb^{-1}
SM H → γ γ
Observed p_0, 10 categories
Expected p_0, 10 categories
Observed p_0, 9 categories
Expected p_0, 9 categories
Observed p_0, inclusive
Expected p_0, 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

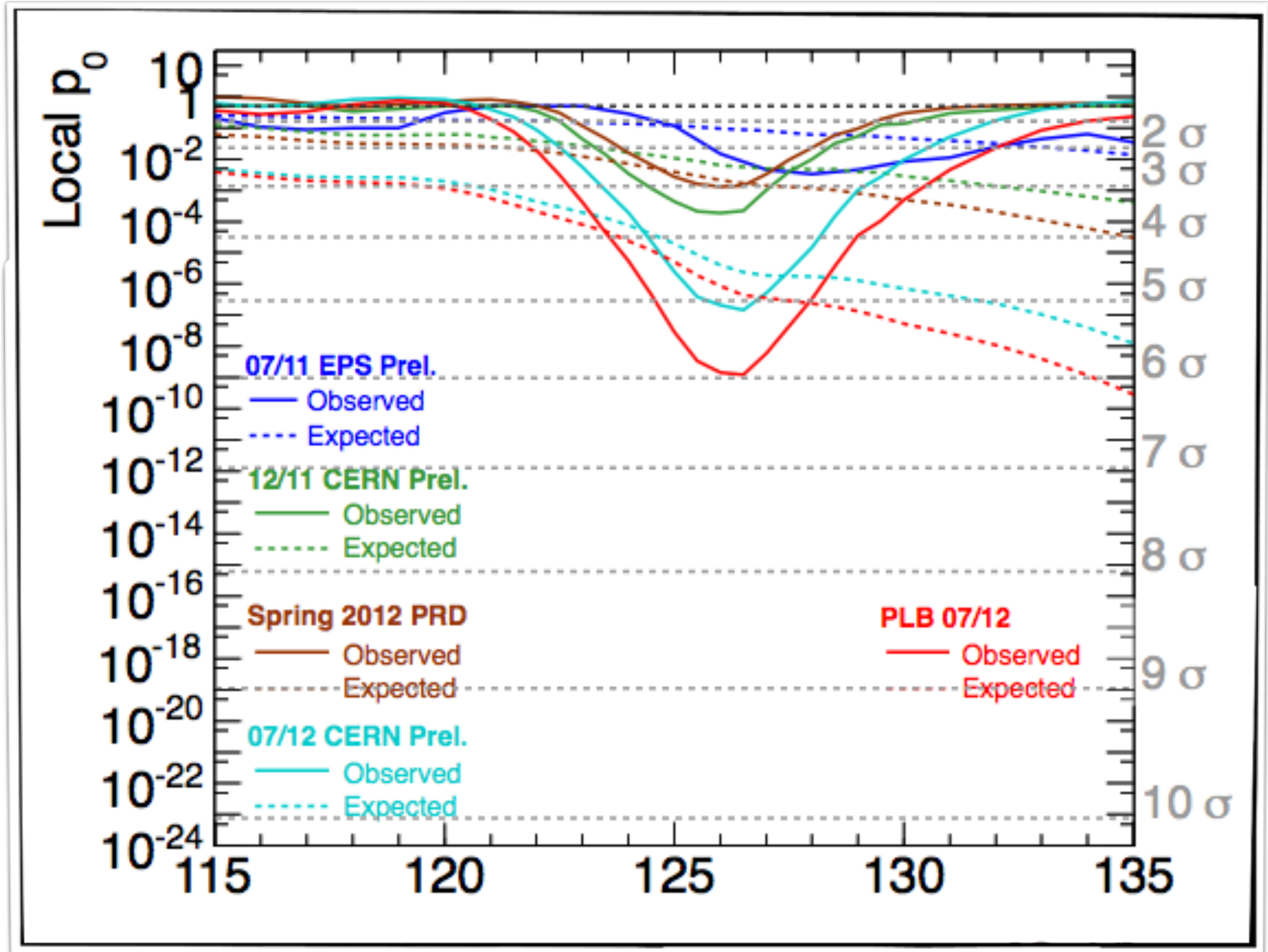

Local p0
10
1
10-2
10-4
10-6
10-8
10-10
10-12
10-14
10-16
10-18
10-20
10-22
10-24
115
120
125
130
135
2 σ
3 σ
4 σ
5 σ
6 σ
7 σ
8 σ
9 σ
10 σ
07/11 EPS Prel.
Observed
Expected

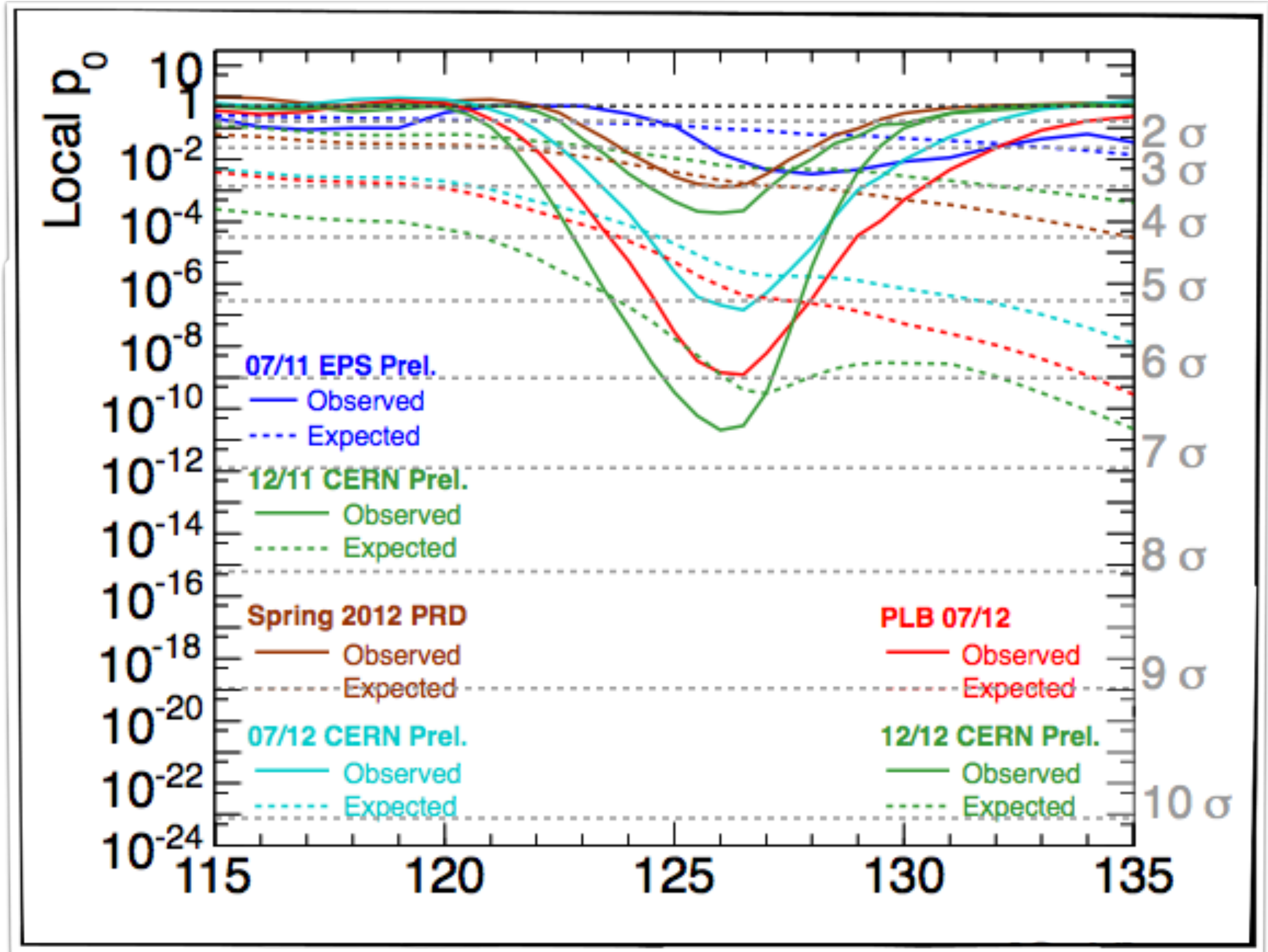

Local p0
10
1
10-2
10-4
10-6
10-8
10-10
10-12
10-14
10-16
10-18
10-20
10-22
10-24
115
120
125
130
135
2 σ
3 σ
4 σ
5 σ
6 σ
7 σ
8 σ
9 σ
10 σ
07/11 EPS Prel.
Observed
Expected
12/11 CERN Prel.
Observed
Expected

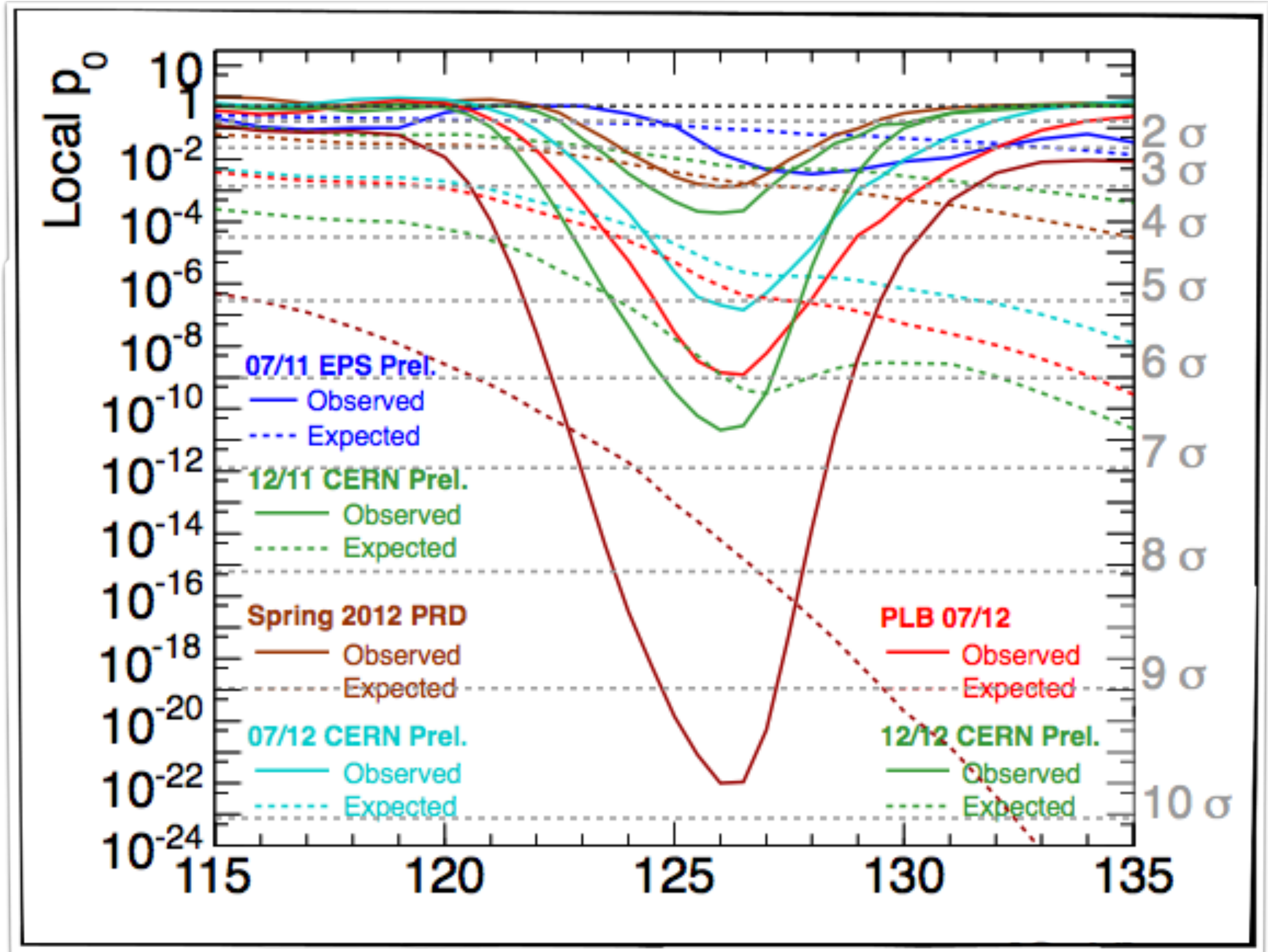

Local p_0
10
1
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
10^{-20}
10^{-22}
10^{-24}
115
120
125
130
135
2 σ
3 σ
4 σ
5 σ
6 σ
7 σ
8 σ
9 σ
10 σ
07/11 EPS Prel.
Observed
Expected
12/11 CERN Prel.
Observed
Expected
Spring 2012 PRD
Observed
Expected



Local p_0
1
10
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
10^{-20}
10^{-22}
10^{-24}
115
120
125
130
135
2 σ
3 σ
4 σ
5 σ
6 σ
7 σ
8 σ
9 σ
10 σ
07/11 EPS Prel.
Observed
Expected
12/11 CERN Prel.
Observed
Expected
Spring 2012 PRD
Observed
Expected
07/12 CERN Prel.
Observed
Expected


Local p_0
10
1
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
10^{-20}
10^{-22}
10^{-24}
115
120
125
130
135
2 σ
3 σ
4 σ
5 σ
6 σ
7 σ
8 σ
9 σ
10 σ
07/11 EPS Prel.
Observed
Expected
12/11 CERN Prel.
Observed
Expected
Spring 2012 PRD
Observed
Expected
07/12 CERN Prel.
Observed
Expected
PLB 07/12
Observed
Expected


Local p_0
10
1
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
10^{-20}
10^{-22}
10^{-24}
115
120
125
130
135
2 σ
3 σ
4 σ
5 σ
6 σ
7 σ
8 σ
9 σ
10 σ
07/11 EPS Prel.
Observed
Expected
12/11 CERN Prel.
Observed
Expected
Spring 2012 PRD
Observed
Expected
07/12 CERN Prel.
Observed
Expected
PLB 07/12
Observed
Expected
12/12 CERN Prel.
Observed
Expected


Local p_0
10
1
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
10^{-20}
10^{-22}
10^{-24}
115
120
125
130
135
2 σ
3 σ
4 σ
5 σ
6 σ
7 σ
8 σ
9 σ
10 σ
07/11 EPS Prel.
Observed
Expected
12/11 CERN Prel.
Observed
Expected
Spring 2012 PRD
Observed
Expected
07/12 CERN Prel.
Observed
Expected
PLB 07/12
Observed
Expected
12/12 CERN Prel.
Observed
Expected

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

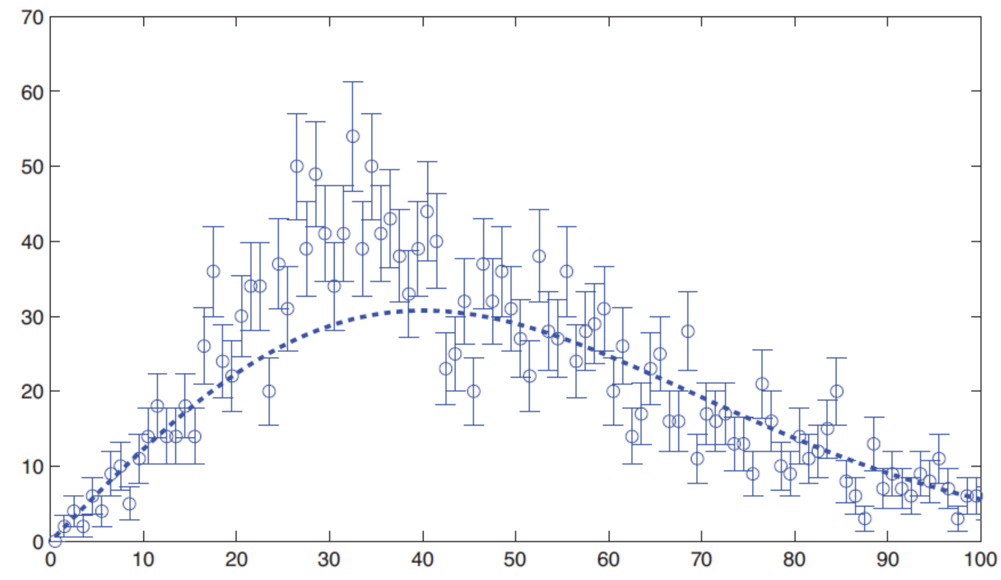
Let the Higgs mass, m_H, and the signal strength μ be 2 parameters of interest

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

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

Look Elsewhere Effect

Is there a signal here?



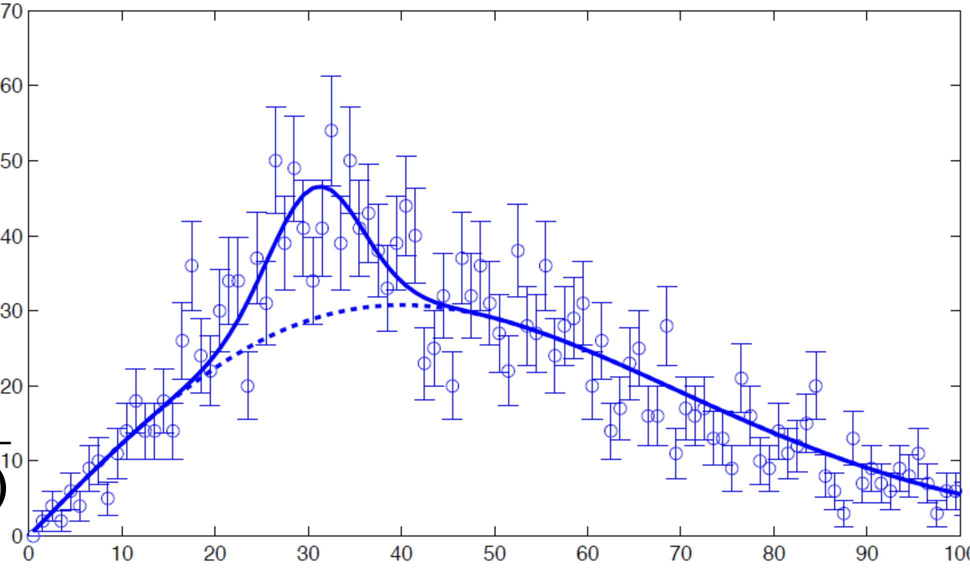
Look Elsewhere Effect

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?

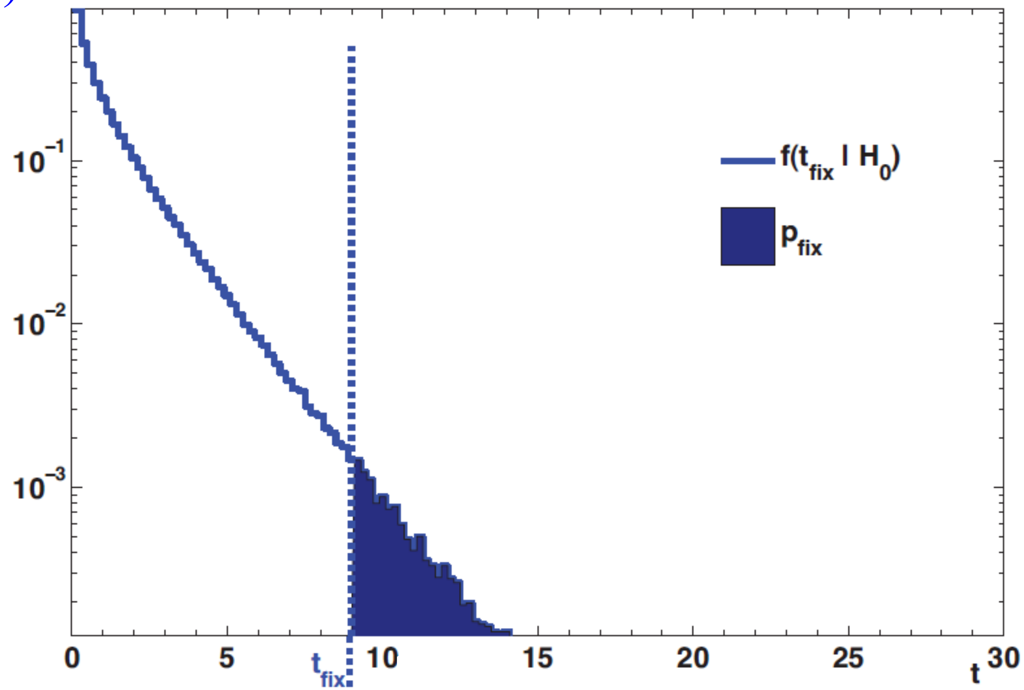
generate the PDF

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

and find the **p-value**

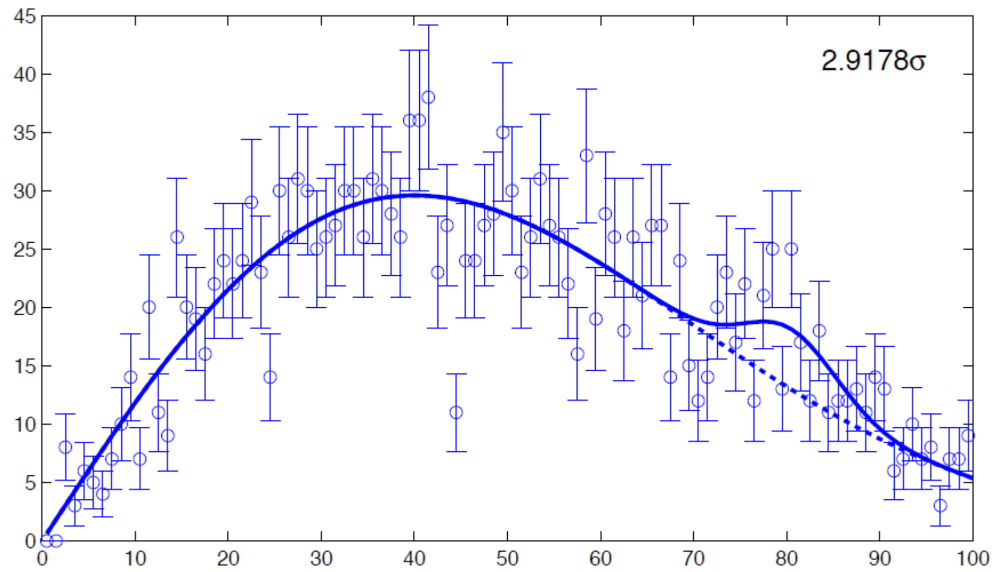
Wilks theorem:

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



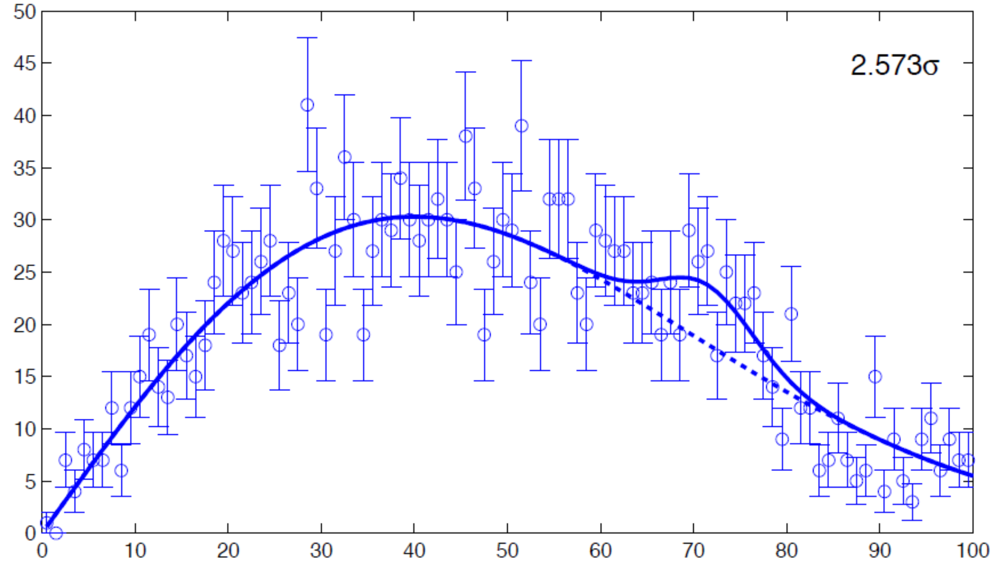
Look Elsewhere Effect

Would you ignore this signal, had you seen it?



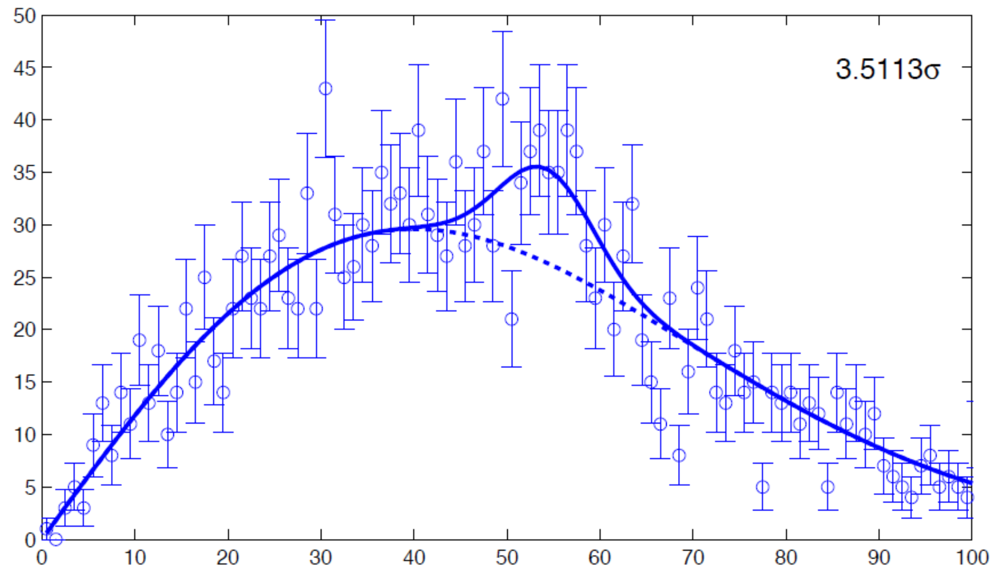
Look Elsewhere Effect

Or this?



Look Elsewhere Effect

Or this?

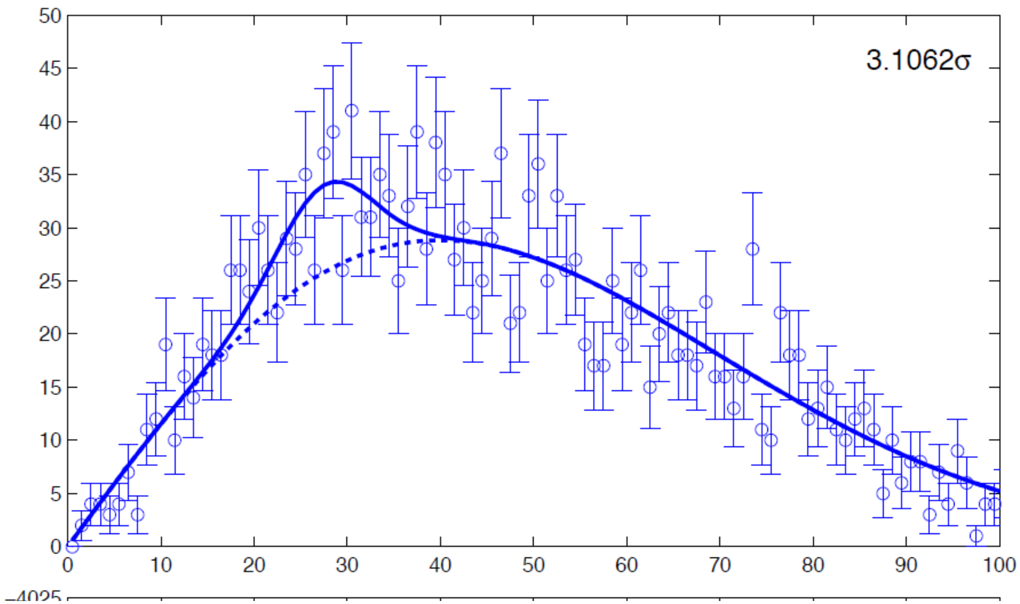


Look Elsewhere Effect

Or this?

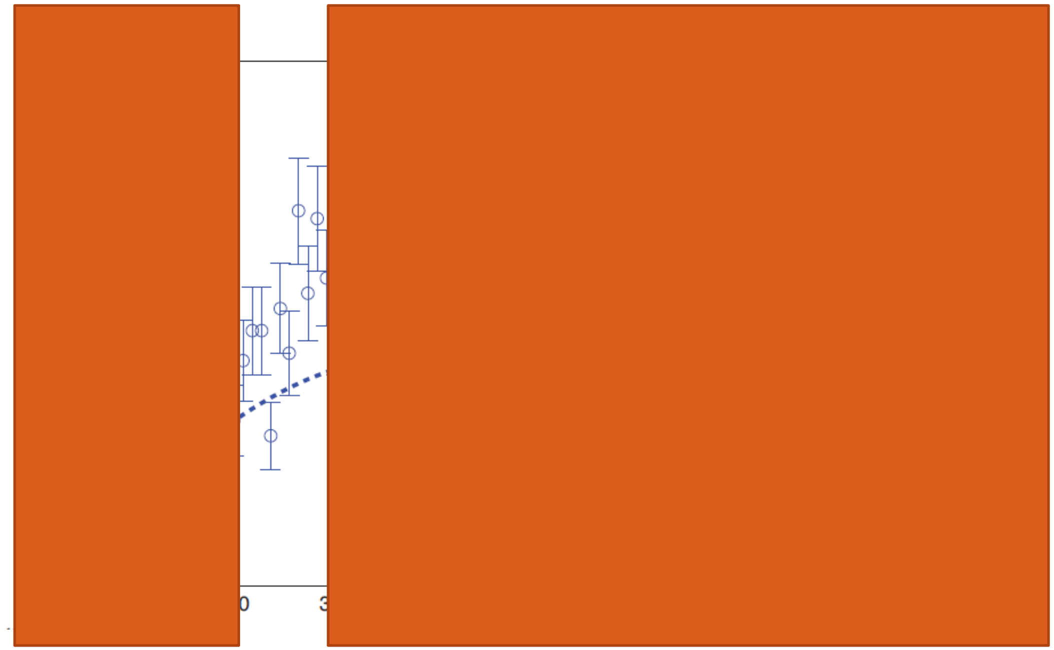
Obviously NOT!

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

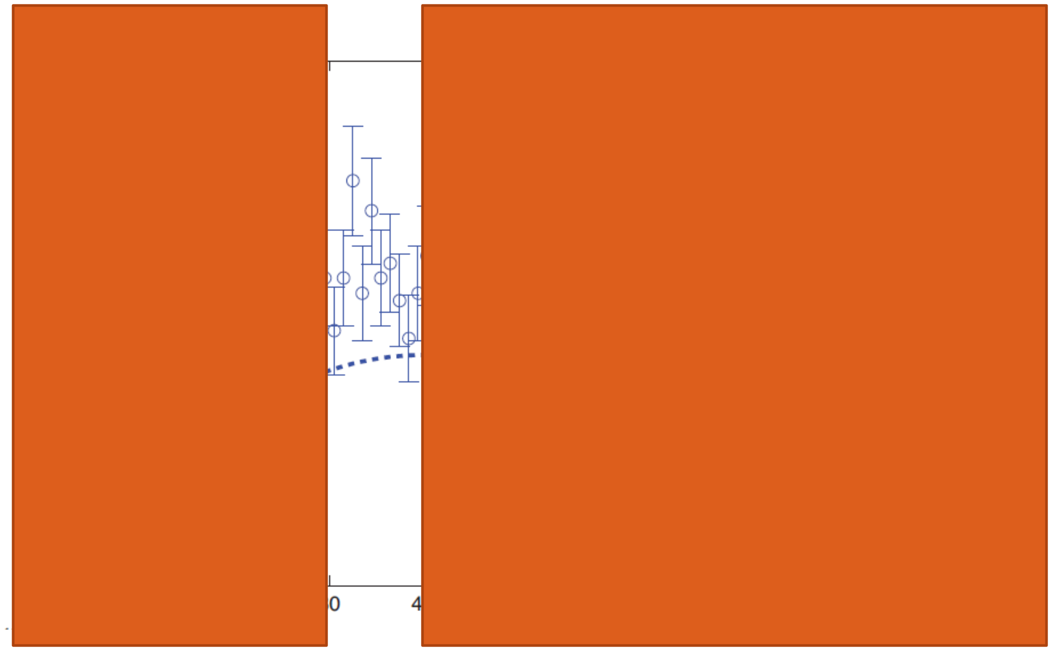


$$q_{fix,obs}(\hat{\mu}) = -2\ln\frac{L(b)}{L(\hat{\mu}s(m)+b)}$$

- Perform a fixed

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



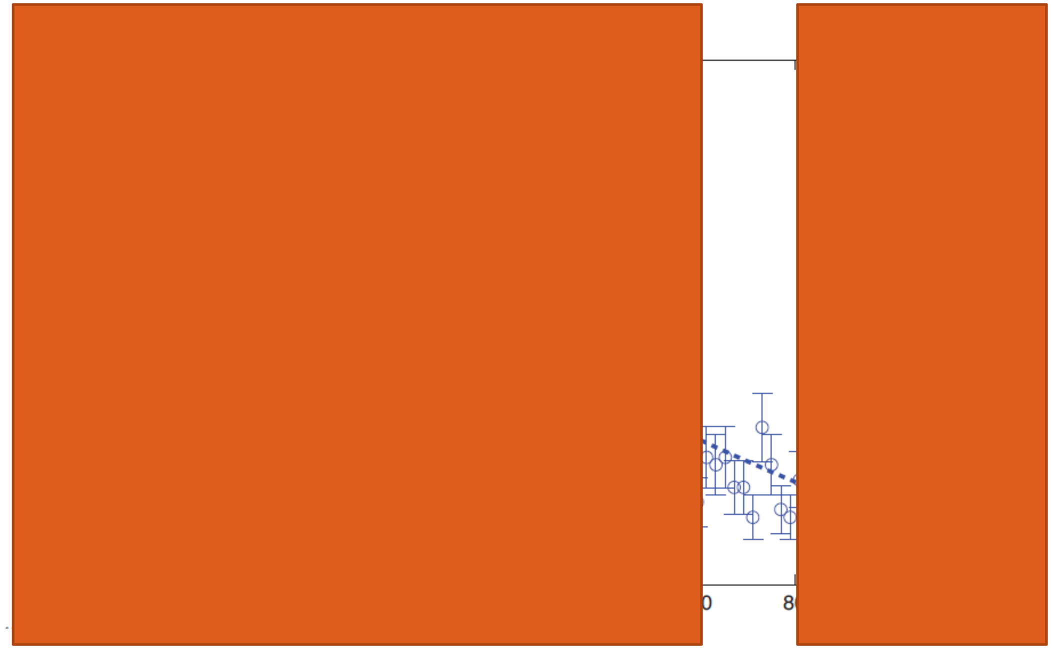
$$q_{fix,obs}(\hat{\mu}) = -2\ln\frac{L(b)}{L(\hat{\mu}s(m)+b)}$$

- Perform a fixed

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)



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

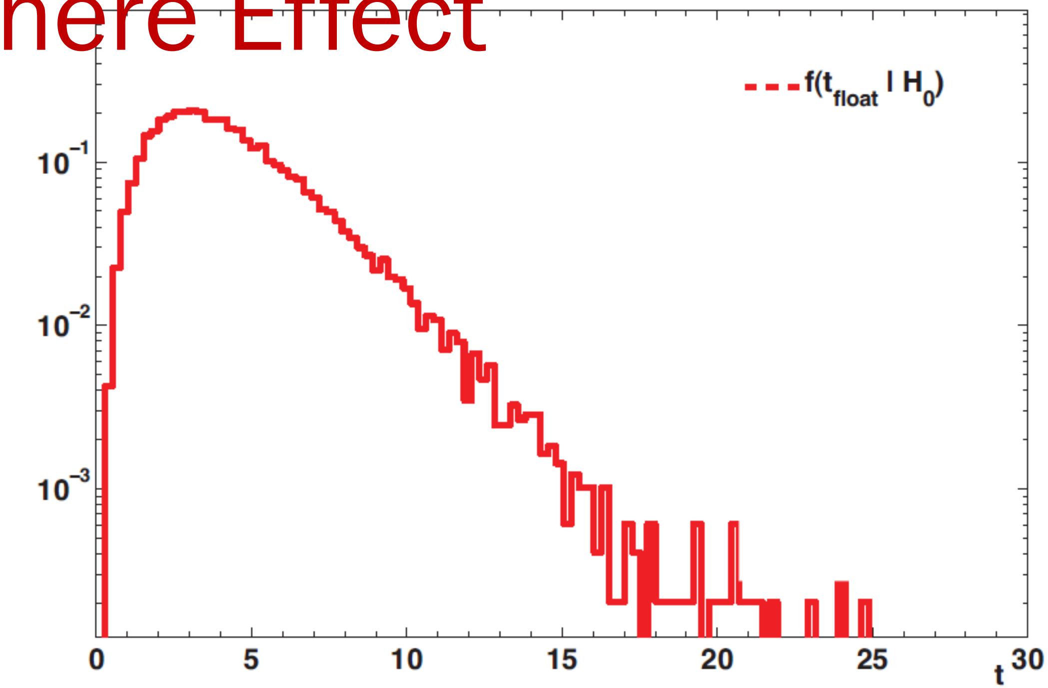
Look Elsewhere Effect

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



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

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

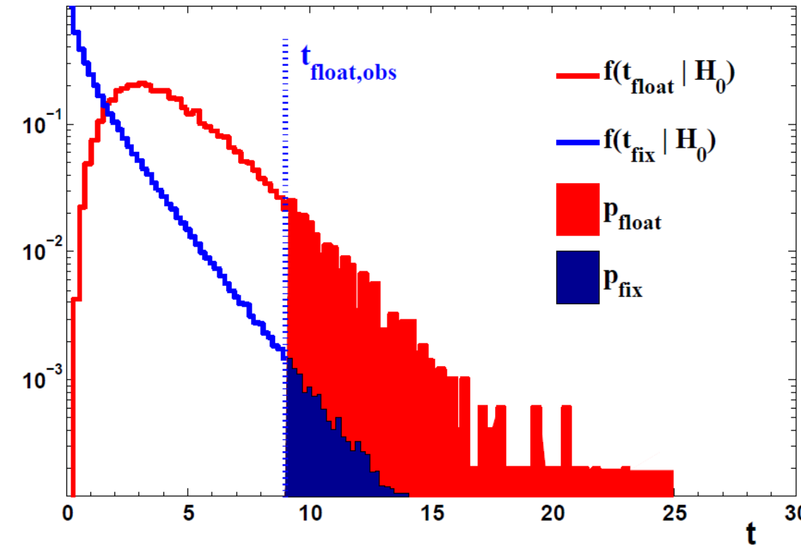
The Thumb Rule

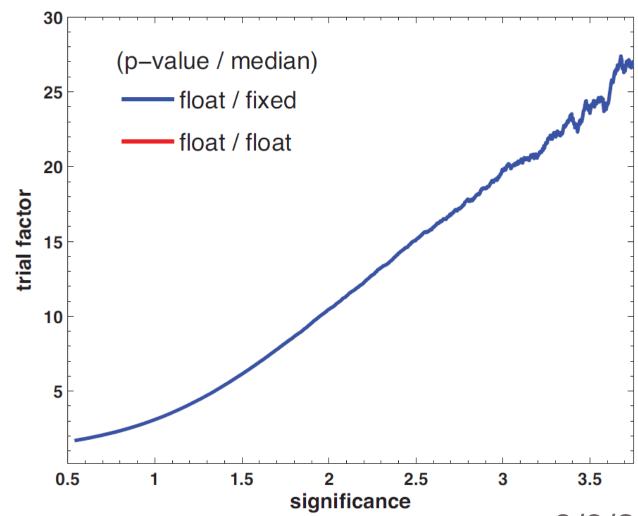
$$\textbf{trial factor} = \frac{p_{float}}{p_{fix}}$$

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

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 <u>fixed</u> θ, $q_0(\theta)$ has (asymptotically) a chi^2 distribution with one degree of freedom by Wilks' theorem.
- $q_0(\theta)$ is a <u>*chi^2 random field*</u> 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

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

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

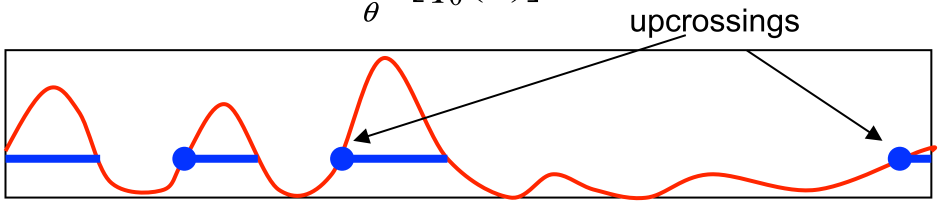
since

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

Random fields (1D)

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

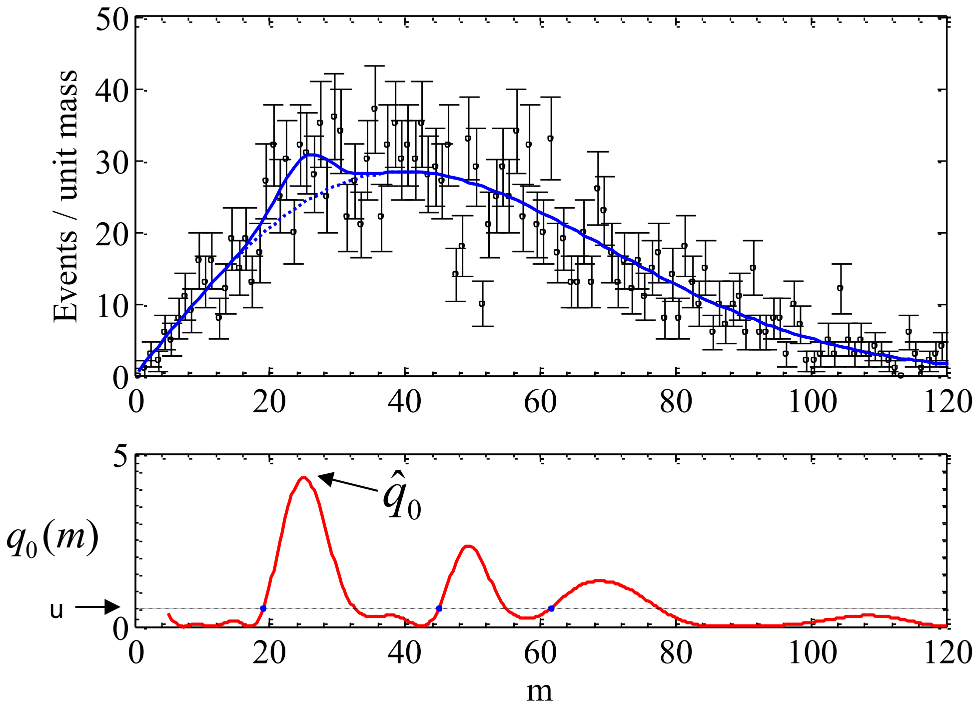
$$\max_\theta[q_0(\theta)]$$



- The probability that the global maximum is above the level u is called *exceedance probability*. (p-value of $\hat{q}_0 \equiv q_0(\hat{\theta}) = \max_\theta[q_0(\theta)]$)

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

The 1-dimensional case



For a chi[2] 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:*

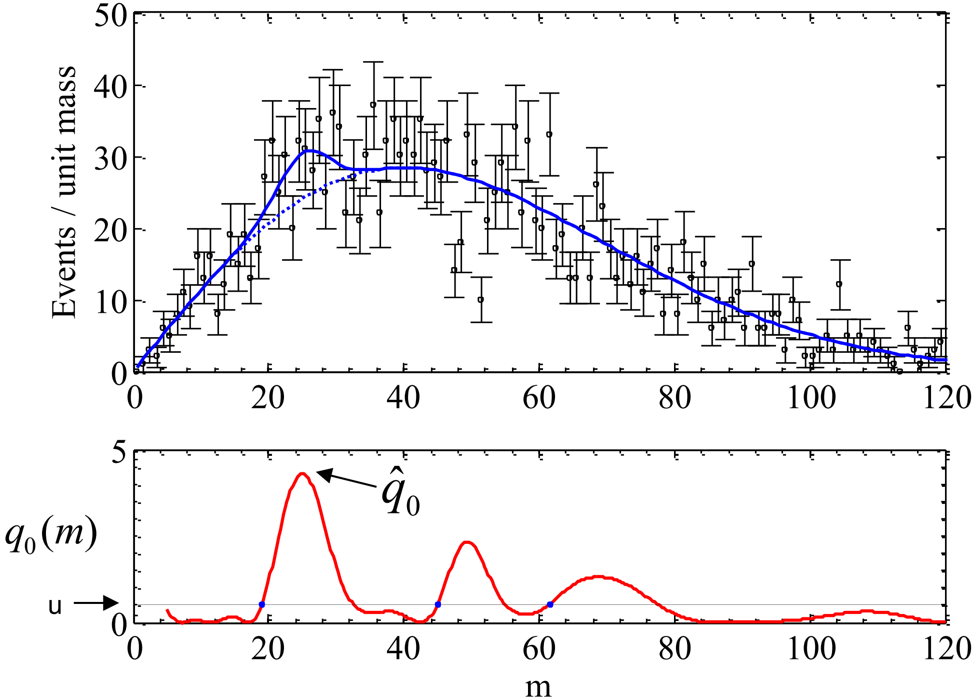
- Either have at least one upcrossing ($N_u>0$) ***or*** have $q_0>u$ at the origin ($q_0(0)>u$) :

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

[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

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

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

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

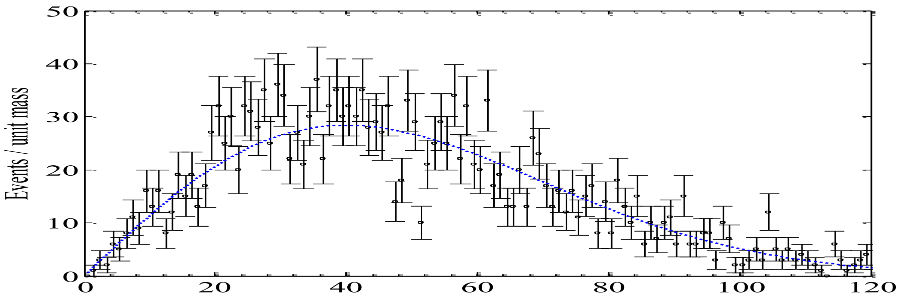
$$E[N_u] = E[N_{u_0}] e^{(u_0-u)/2}$$

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

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

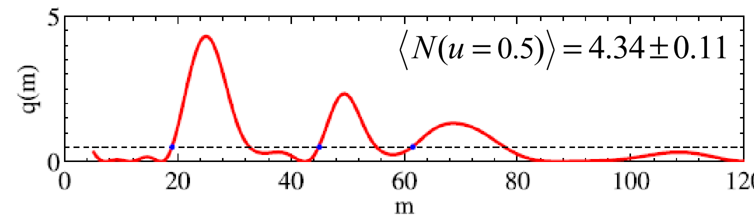
$$p_{global} = E[N_{u_0}] e^{(u_0-u)/2} + p_{local}$$

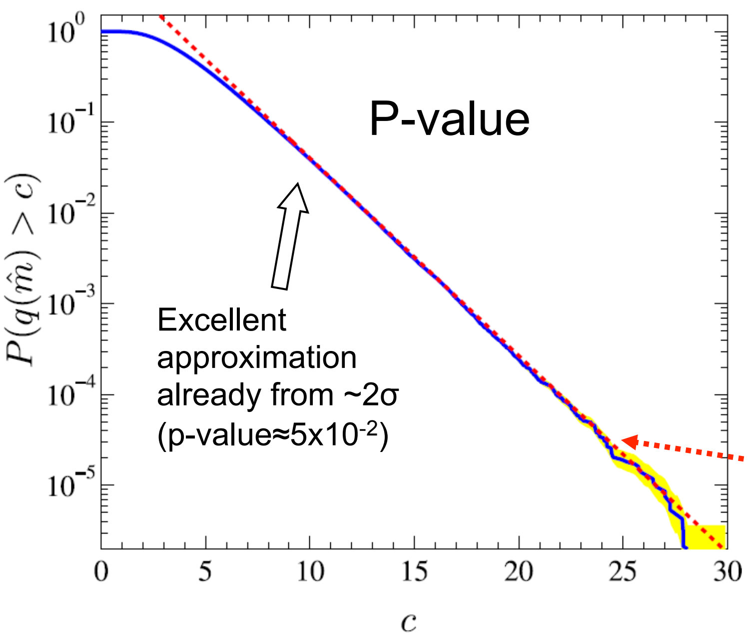
1-D example: resonance search



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

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



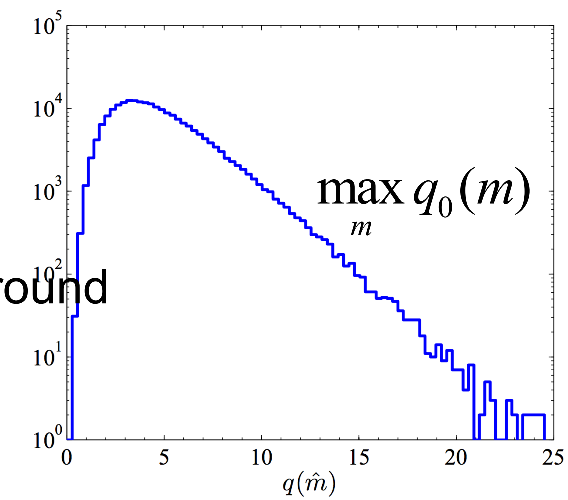


In this example we find

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

[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

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

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

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

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

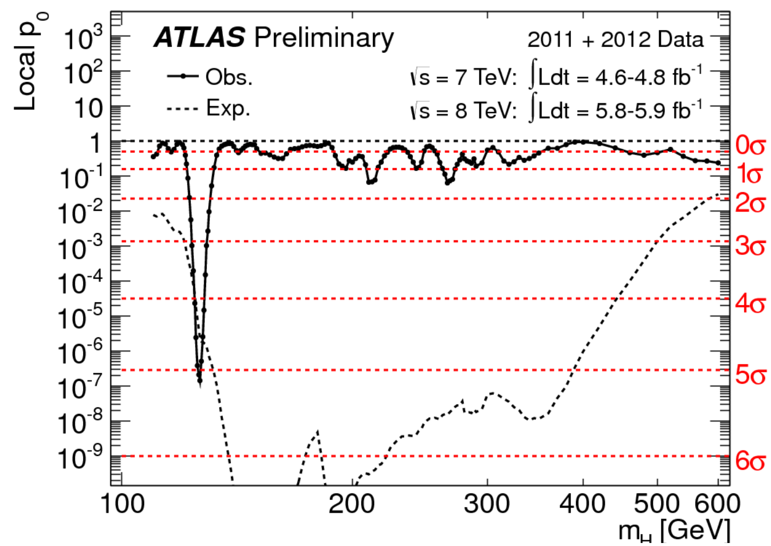
$$p_{global} = \mathcal{N}_1 e^{-u/2} + p_{local}$$

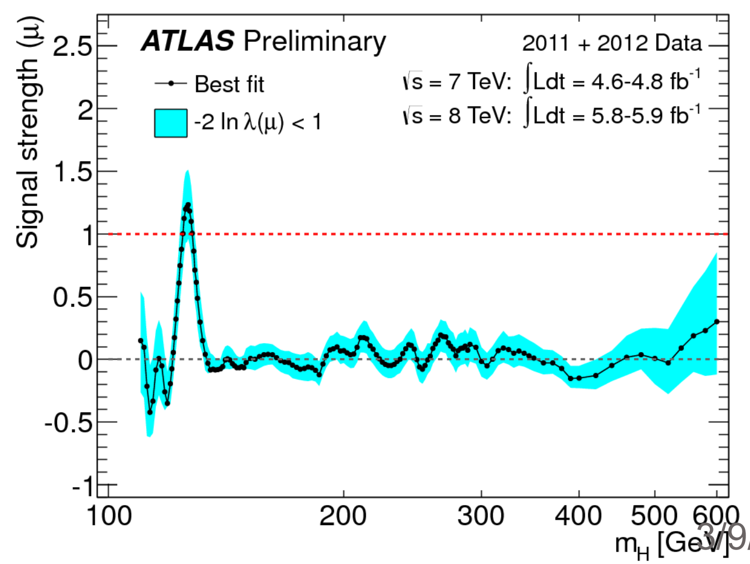
$$p_{global} = \langle N_{u_0} \rangle e^{\frac{u_0 - u}{2}} + p_{local}$$

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

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

$5\sigma \rightarrow 4\sigma$ trial#~100





Measurements

Case studies: ATLAS and CMS mass and coupling combinations

PL in obtaining the mass

$$\Lambda(\boldsymbol{\alpha}) = \frac{L(\boldsymbol{\alpha}\,,\,\hat{\hat{\boldsymbol{\theta}}}(\boldsymbol{\alpha}))}{L(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\theta}})} \qquad t_\alpha = -2ln\Lambda(\alpha)$$

$$\Lambda(m_H) = \frac{L(m_H\,,\,\hat{\hat{\mu}}^{\gamma\gamma}_{ggF+t\bar{t}H(+b\bar{b}H)}(m_H)\,,\hat{\hat{\mu}}^{\gamma\gamma}_{VBF+VH}(m_H)\,,\,\hat{\hat{\mu}}^{ZZ}(m_H)\,,\,\hat{\hat{\boldsymbol{\theta}}}(m_H))}{L(\hat{m}_H,\hat{\mu}^{\gamma\gamma}_{ggF+t\bar{t}H(+b\bar{b}H)}\,,\hat{\mu}^{\gamma\gamma}_{VBF+VH},\hat{\mu}^{ZZ},\hat{\boldsymbol{\theta}})}$$

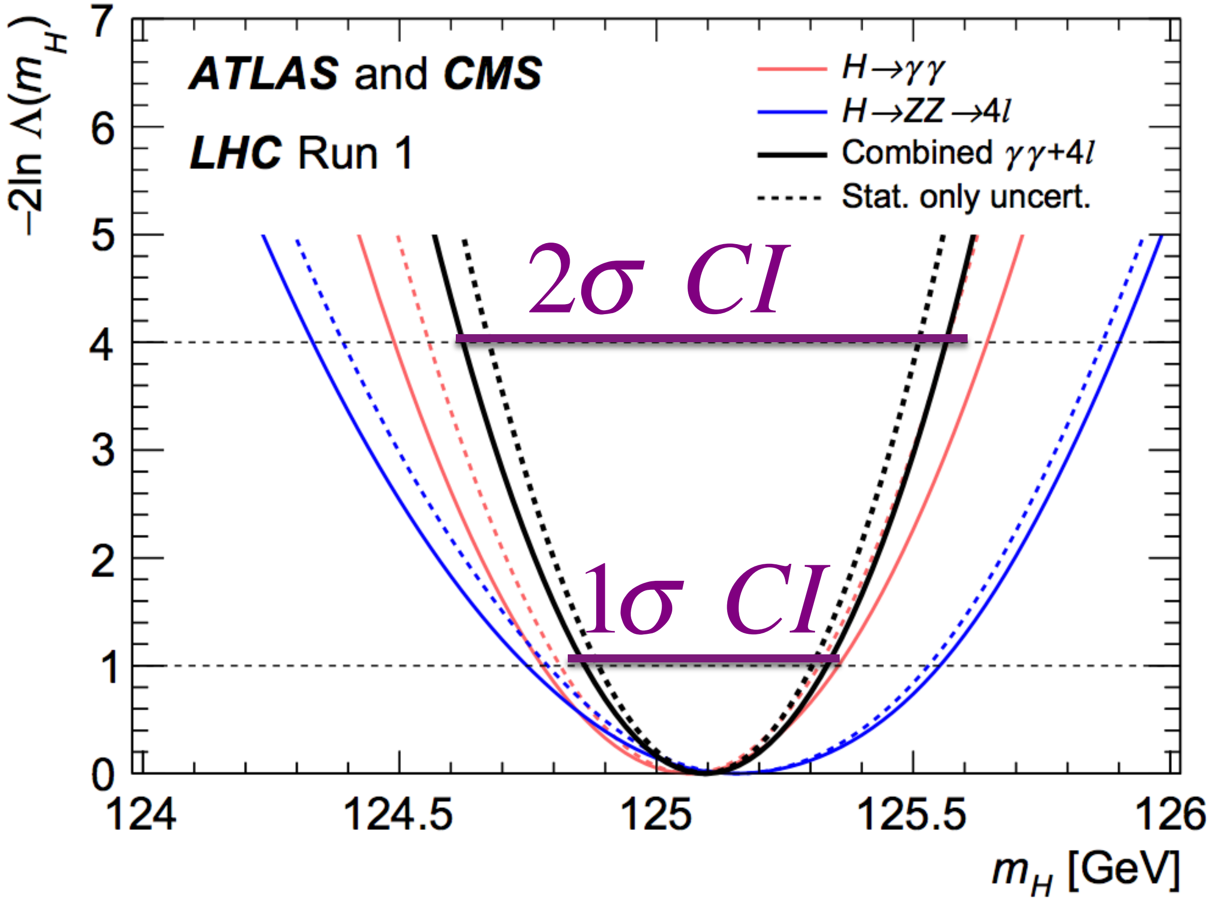
Scan the test statistic $t_\alpha = t(\alpha)$

find $\hat{\alpha}$

$t(\hat{\alpha} \pm N\sigma_{\hat{\alpha}}) = N^2$

Obtaining the Syst Error

$$\sigma_{syst} = \sqrt{\sigma_{tot}^2 - \sigma_{stat}^2}$$



PL in obtaining the mass

$$\Lambda(\boldsymbol{\alpha}) = \frac{L(\boldsymbol{\alpha}\,,\hat{\hat{\boldsymbol{\theta}}}(\boldsymbol{\alpha}))}{L(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\theta}})}$$

$$t_\alpha = -2ln\Lambda(\alpha)$$

$$\Delta m_{\gamma Z} = m_H^{\gamma\gamma} - m_H^{4\ell}$$

$$\Lambda(\Delta m_{\gamma Z}) = \frac{L(\Delta m_{\gamma Z}\,,\hat{\hat{m}}_H,\,\hat{\hat{\mu}}^{\gamma\gamma}_{ggF+t\bar{t}H(+b\bar{b}H)}\,,\hat{\hat{\mu}}^{\gamma\gamma}_{VBF+VH}\,,\,\hat{\hat{\mu}}^{ZZ}\,,\,\hat{\hat{\boldsymbol{\theta}}})}{L(\Delta\hat{m}_{\gamma Z},\hat{m}_H,\hat{\mu}^{\gamma\gamma}_{ggF+t\bar{t}H(+b\bar{b}H)}\,,\hat{\mu}^{\gamma\gamma}_{VBF+VH},\hat{\mu}^{ZZ},\hat{\boldsymbol{\theta}})}$$

2nd verse same as the first

A case of 2 poi

- 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:

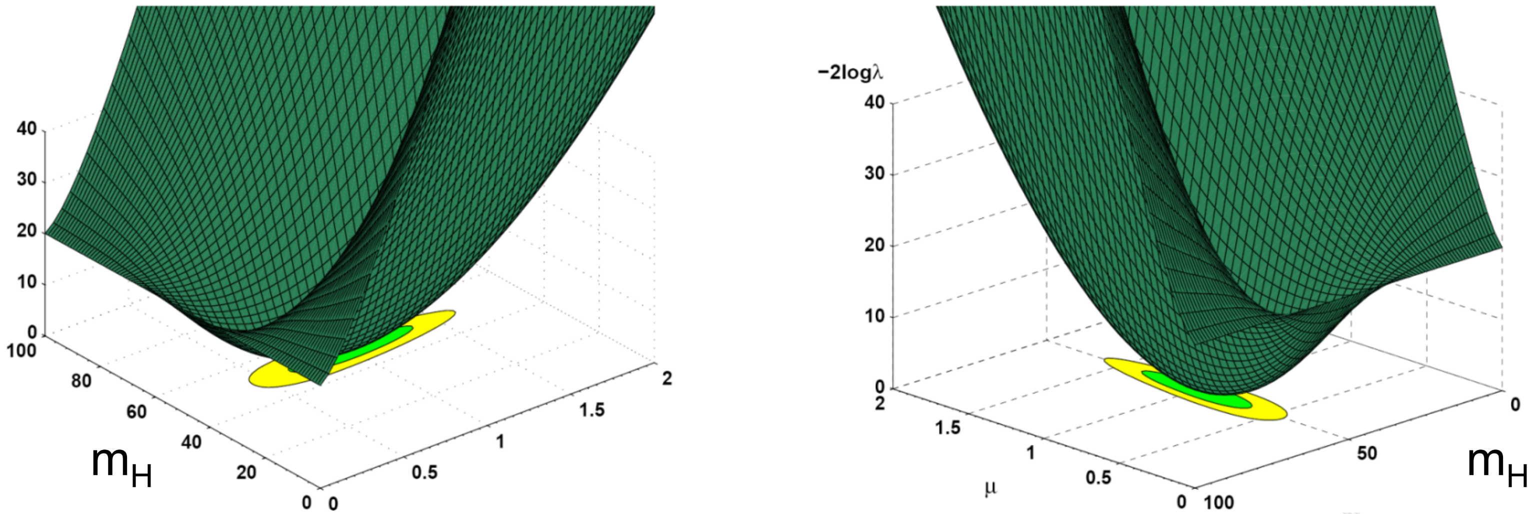
$$\lambda(\mu, m_H) = \frac{L(\mu, m_H, \hat{\hat{\theta}}(\mu, m_H))}{L(\hat{\mu}, \hat{m}_H, \hat{\theta})}$$

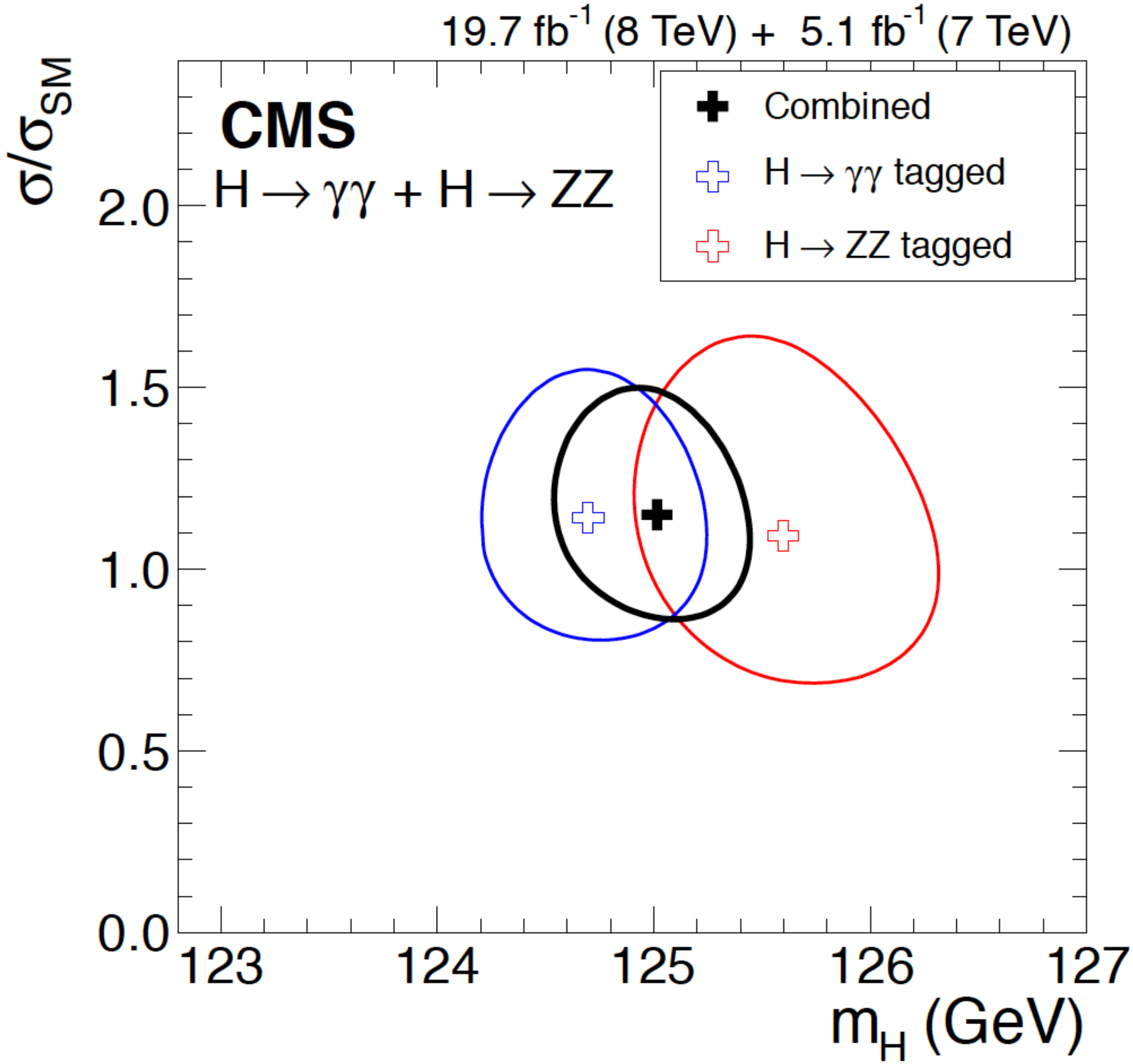
- In the presence of a signal, this test statistic will produce closed contours about the best fit point $(\hat{\mu}, \hat{m}_H)$;
- The 2D LR behaves asymptotically as a Chis squared with 2 DOF (Wilks' theorem) so the derivation of 68% and 95% CL cintours is easy, but care must be taken; The projection of 2D CI are not 1D CI!

Measuring the signal strength and mass, a 2D Scan

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

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

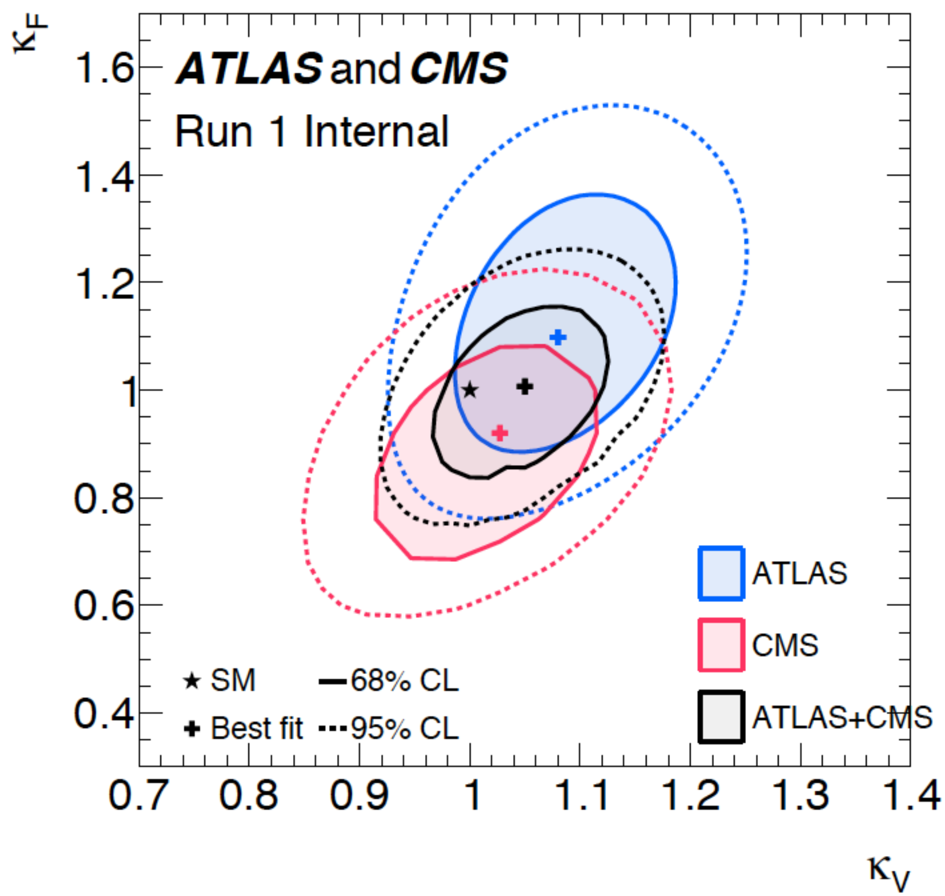




19.7 fb-1 (8 TeV) + 5.1 fb-1 (7 TeV)
CMS
H → γγ + H → ZZ
Combined
H → γγ tagged
H → ZZ tagged
σ/σSM
mH (GeV)
2.0
1.5
1.0
0.5
0.0
123
124
125
126
127

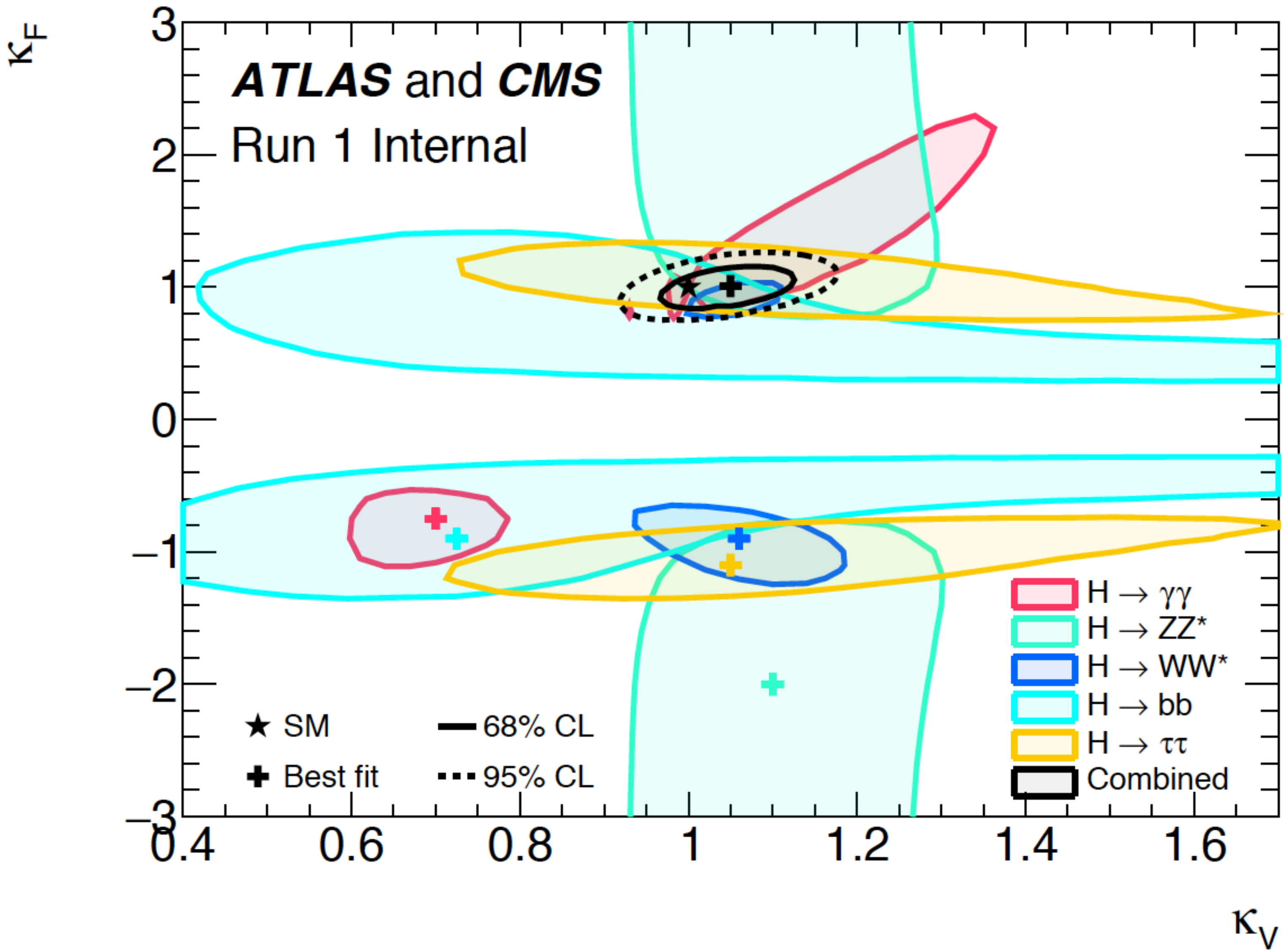
PL in obtaining the Couplings

$$\Lambda(\kappa_F,\kappa_V)=\frac{L(\kappa_F,\kappa_V,\hat{\hat{\vec{\theta}}}(\kappa_F,\kappa_V))}{L(\hat{\kappa}_F,\hat{\kappa}_V,\hat{\vec{\theta}})}\ .$$



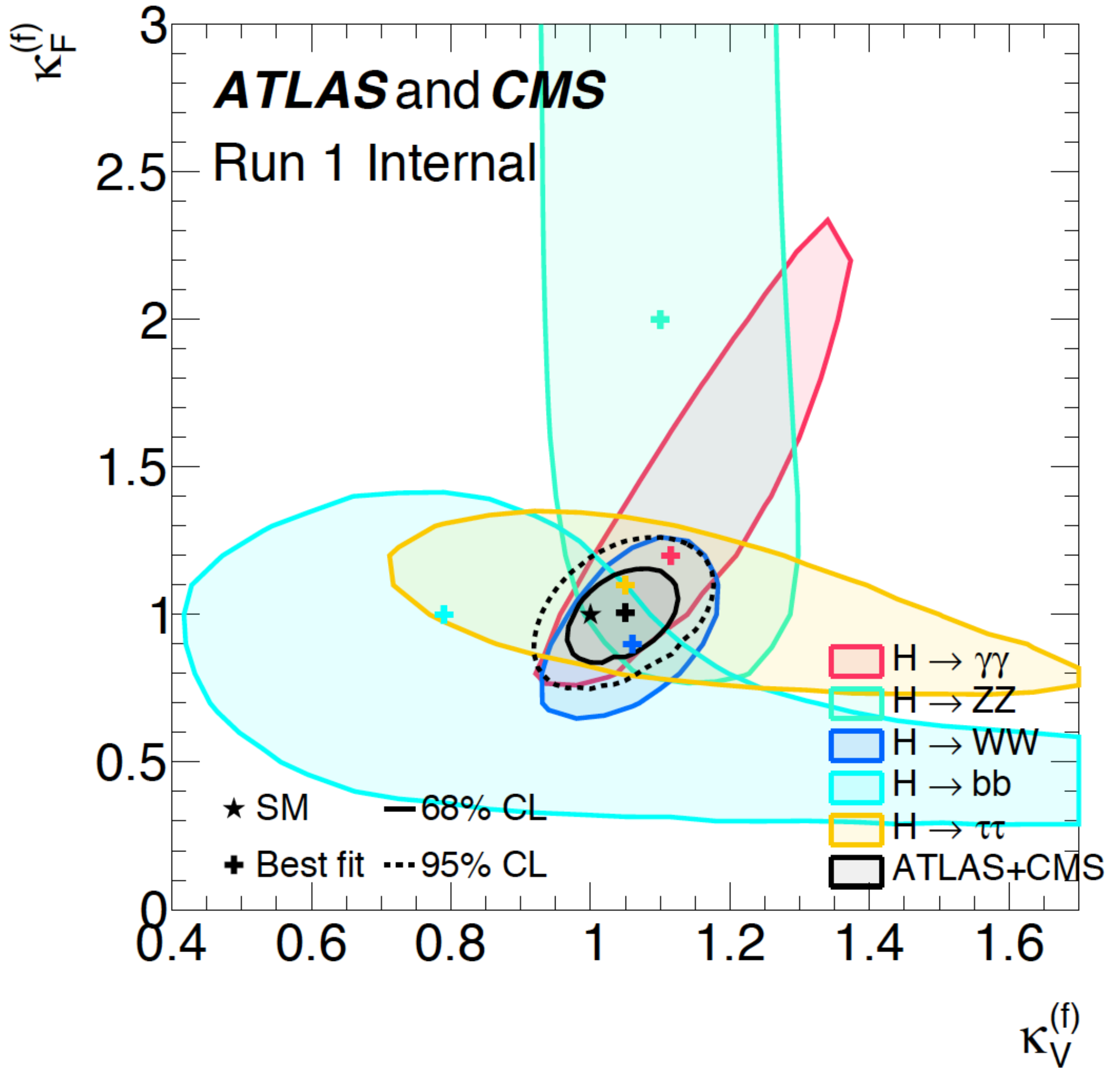
68% CI is a tricky issue

Is the WW a better measurement than the combination?



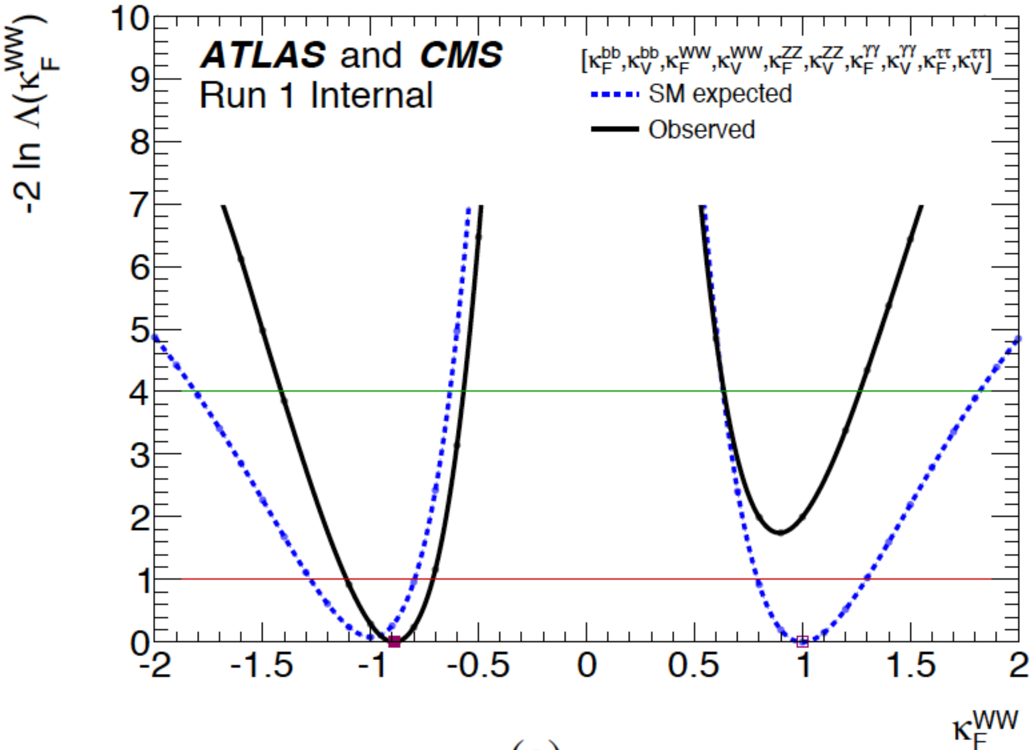
68% CI is a tricky issue

When constraining to positive couplings, the WW gains the full CI



68% CI is a tricky issue

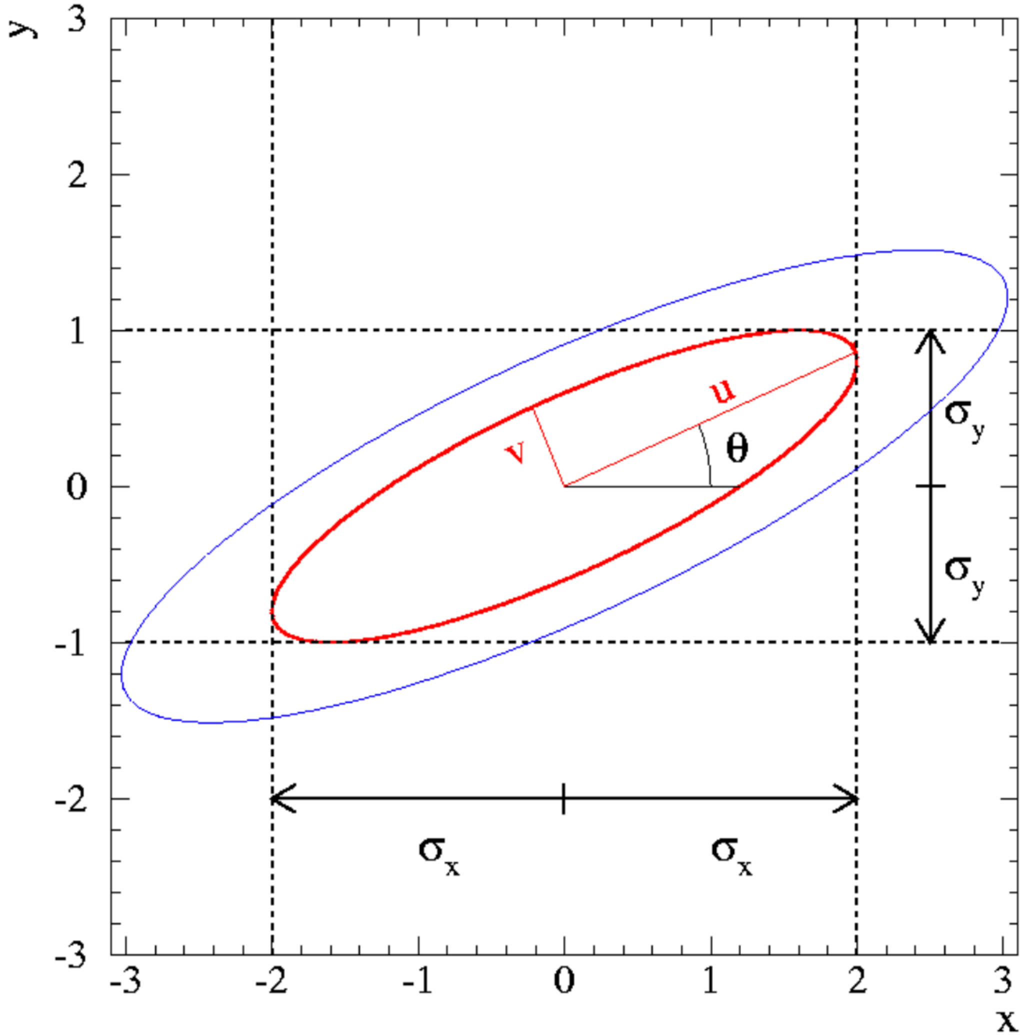
Is the WW a better measurement than the combination?



(c)

1D CI
Is not
2D CI

1D vs 2D Confidence Interval



$$\Delta\chi^2 = 1$$

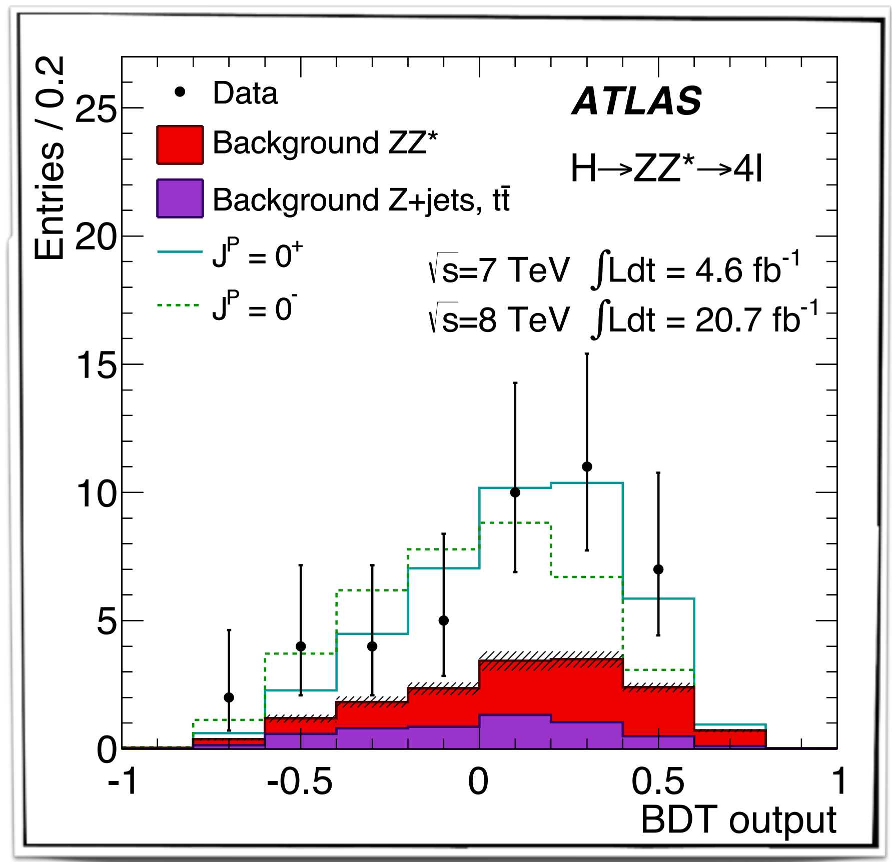
$$\Delta\chi^2 = 2.3 \quad (68\%\ CL)$$

Application of CIs and q^{NP} test statistic

Can you tell O^+ from 0^-?

$$q^{NP} = -2\ln\frac{L(H_0)}{L(H_1)} = \sum_{bins} -2\ln\frac{L_i(0^+)}{L_i(0^-)}$$

$$L_i(0^+) = Pois(n_i; n_i^{0^+}) = \frac{\left(n_i^{0^+}\right)^{n_i} e^{-n_i^{0^+}}}{n_i!}$$



Test Spin 0 parity

$$H_0 = 0^+$$

$$H_1 = 0^-$$

$$q^{NP} = -2\ln\frac{L(H_0)}{L(H_1)}$$

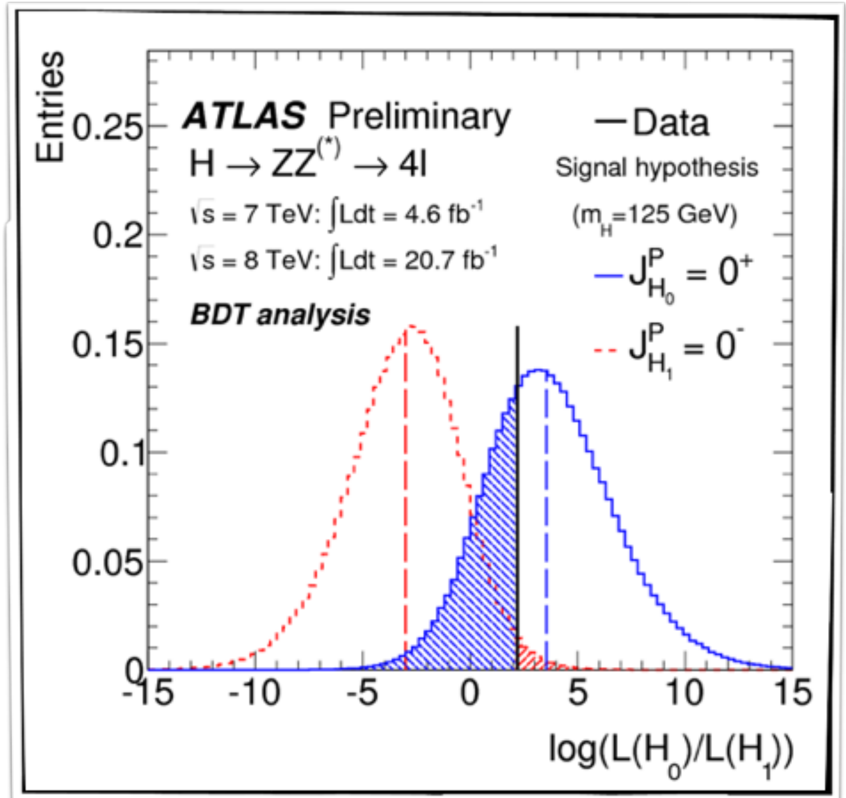
$$p_{H_1}(\exp | H_0) = 0.37\%,$$

$$p_{H_1}(obs) = 1.5\%$$

$$p_{H_0}(obs) = 31\%$$

$$p_{H_1}^{CL_s}(obs) = 2.2\%$$

$$p_{H_1}^{CL_s} = \frac{p_{H_1}}{1-p_{H_0}} = \frac{1.5\%}{1-0.31} = 2.2\%$$



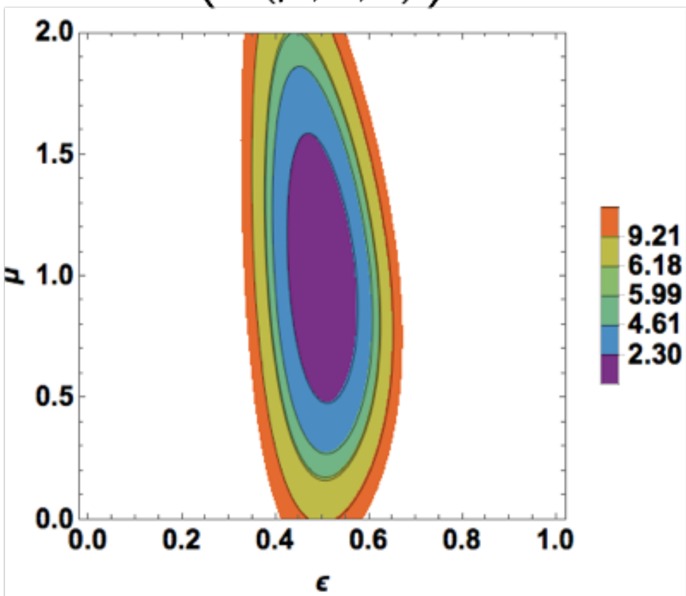
Which means
$J^p=0^-$ is excluded at the
97.8% CL in favour of $J^p=0^+$

$H_1 like$ ← $H_0 like$

Multidimensional PL

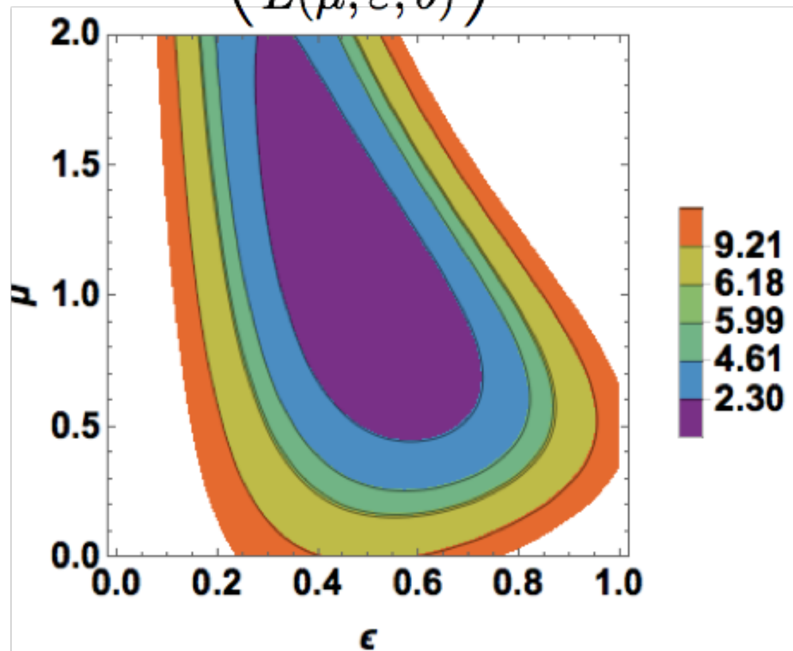
A toy case with 2 poi

$$-2log\left(\frac{L(\mu,\varepsilon,\hat{\hat{b}})}{L(\hat{\mu},\hat{\varepsilon},\hat{b})}\right)$$



background = 100
signal = 90
ε = 0.5
σ_ε **= 0.05**
σ_b = 20

$$-2log\left(\frac{L(\mu,\varepsilon,\hat{\hat{b}})}{L(\hat{\mu},\hat{\varepsilon},\hat{b})}\right)$$



background = 100
signal = 90
ε = 0.5
σ_ε **= 0.15**
σ_b = 10

A toy case with 1-3 poi

3 cases studied

$1poi: \mu$ while ϵ, A, b *profiled*

$2poi: \mu, \epsilon$ *profile* A and b

3poi: μ, ϵ, A profile b

$$n = \mu\epsilon A s + b$$

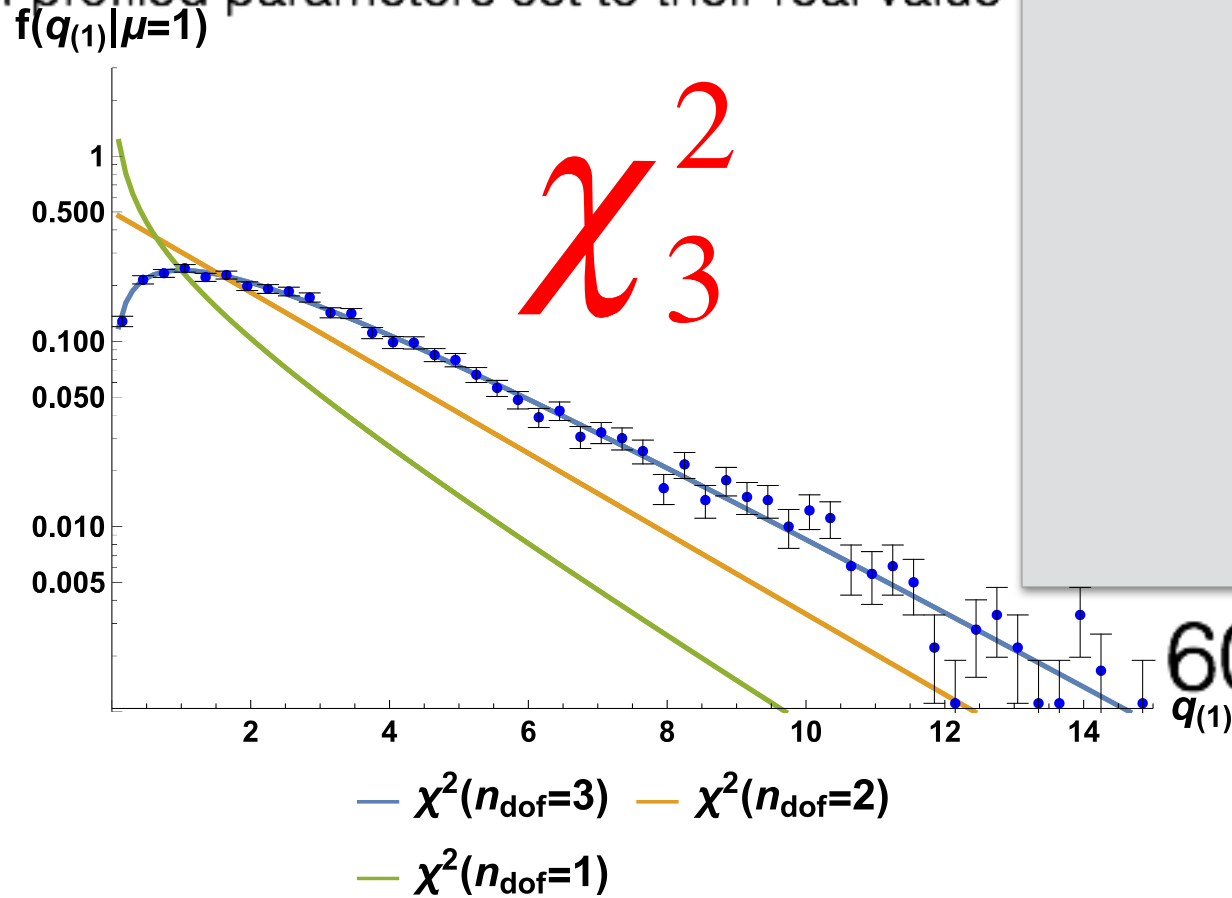
$$L = L(\mu, \epsilon, A, b)$$

$$L(\mu, \varepsilon, A) = \frac{(\mu\varepsilon A s + b)^n}{n!} e^{-(\mu\varepsilon A s + b)} \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} e^{-(\varepsilon_{meas}-\varepsilon)^2/2\sigma_\varepsilon^2} \frac{1}{\sigma_b\sqrt{2\pi}} e^{-(b_{meas}-b)^2/2\sigma_b^2} \frac{1}{\sigma_A\sqrt{2\pi}} e^{-(A_{meas}-A)^2/2\sigma_A^2}$$

A toy case with 3 poi

$$L(\mu,\varepsilon,A)=\frac{(\mu\varepsilon As+b)^n}{n!}e^{-(\mu\varepsilon As+b)}\frac{1}{\sigma_\varepsilon\sqrt{2\pi}}e^{-(\varepsilon_{meas}-\varepsilon)^2/2\sigma_\varepsilon^2}\frac{1}{\sigma_b\sqrt{2\pi}}e^{-(b_{meas}-b)^2/2\sigma_b^2}\frac{1}{\sigma_A\sqrt{2\pi}}e^{-(A_{meas}-A)^2/2\sigma_A^2}$$

three parameters of interest (profiling only b)
non-profiled parameters set to their real value



background = 100
signal = 90
ε = 0.5
A=0.7
σ_ε **= 0.05**
σ_b = 10
σ_A **= 0.2**

6000 events

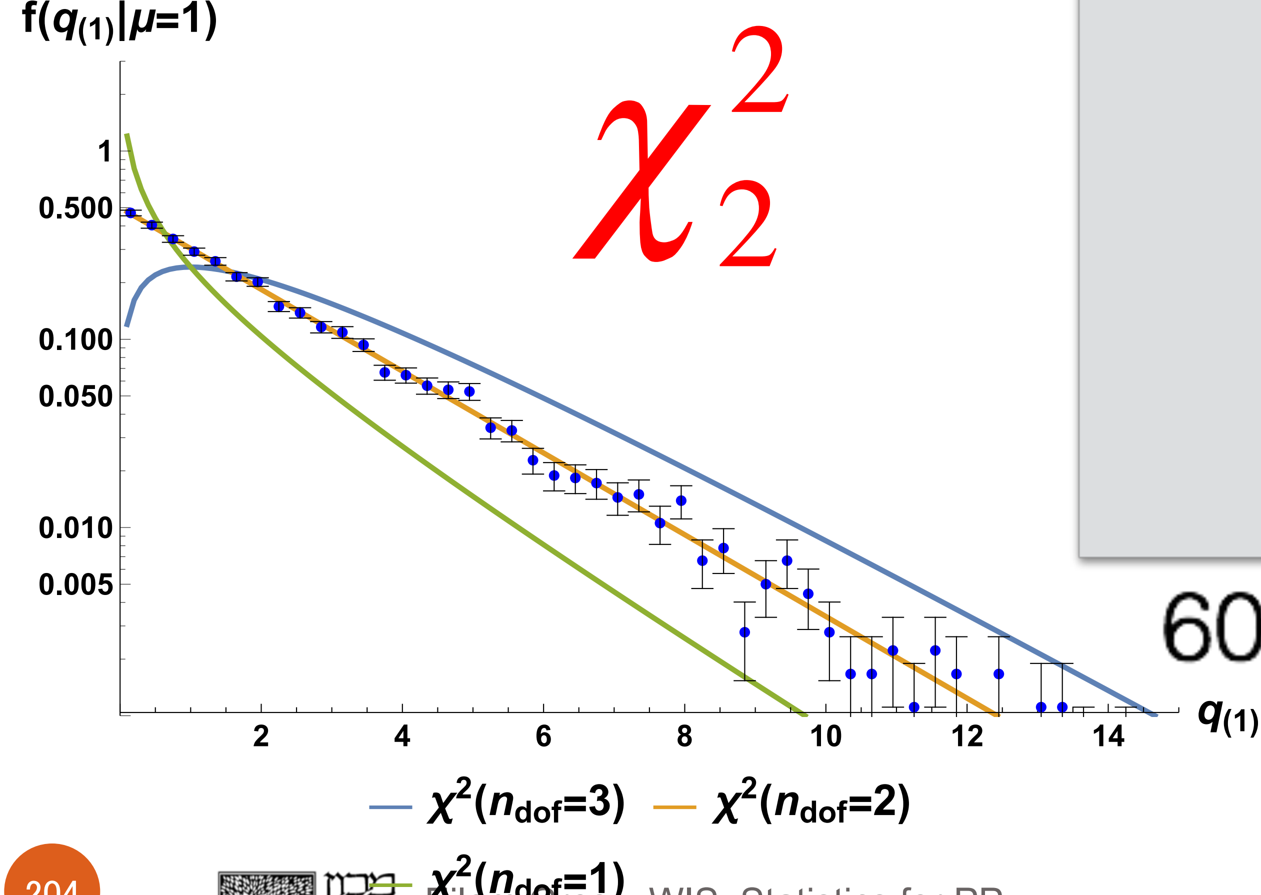
A toy case with 2 poi

$$L(\mu,\varepsilon,A) = \frac{(\mu\varepsilon A s + b)^n}{n!} e^{-(\mu\varepsilon A s + b)} \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} e^{-(\varepsilon_{meas}-\varepsilon)^2/2\sigma_\varepsilon^2} \frac{1}{\sigma_b\sqrt{2\pi}} e^{-(b_{meas}-b)^2/2\sigma_b^2} \frac{1}{\sigma_A\sqrt{2\pi}} e^{-(A_{meas}-A)^2/2\sigma_A^2}$$

two parameters of interest (profiling A and b)
non-profiled parameters set to their real value

background = 100
signal = 90
ε = 0.5
A=0.7
= 0.05
= 10
= 0.2

χ^2_2

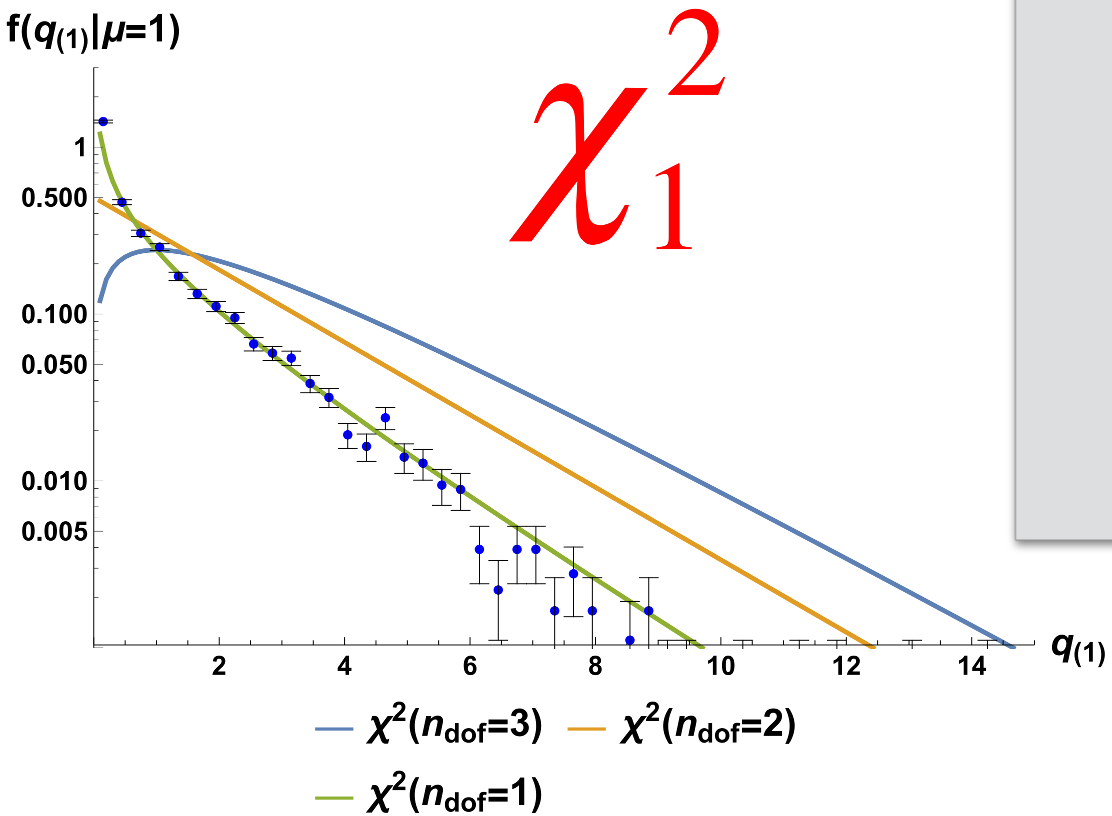


6000 events

A toy case with 1 poi

$$L(\mu,\varepsilon,A)=\frac{(\mu\varepsilon As+b)^n}{n!}e^{-(\mu\varepsilon As+b)}\frac{1}{\sigma_\varepsilon\sqrt{2\pi}}e^{-(\varepsilon_{meas}-\varepsilon)^2/2\sigma_\varepsilon^2}\frac{1}{\sigma_b\sqrt{2\pi}}e^{-(b_{meas}-b)^2/2\sigma_b^2}\frac{1}{\sigma_A\sqrt{2\pi}}e^{-(A_{meas}-A)^2/2\sigma_A^2}$$

one parameter of interest (profiling ε A and b)
non-profiled parameters set to their real value



$$\chi^2_1$$

background = 100
signal = 90
ε = 0.5
A=0.7
σ_ε **= 0.05**
σ_b = 10
σ_A **= 0.2**

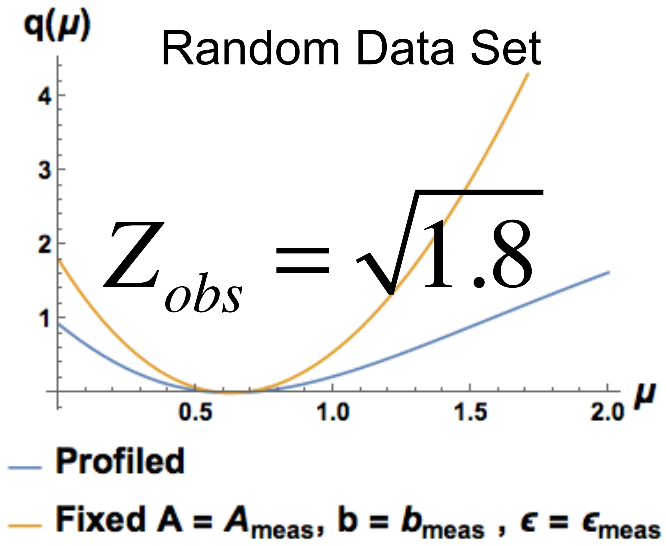
6000 events

Significance

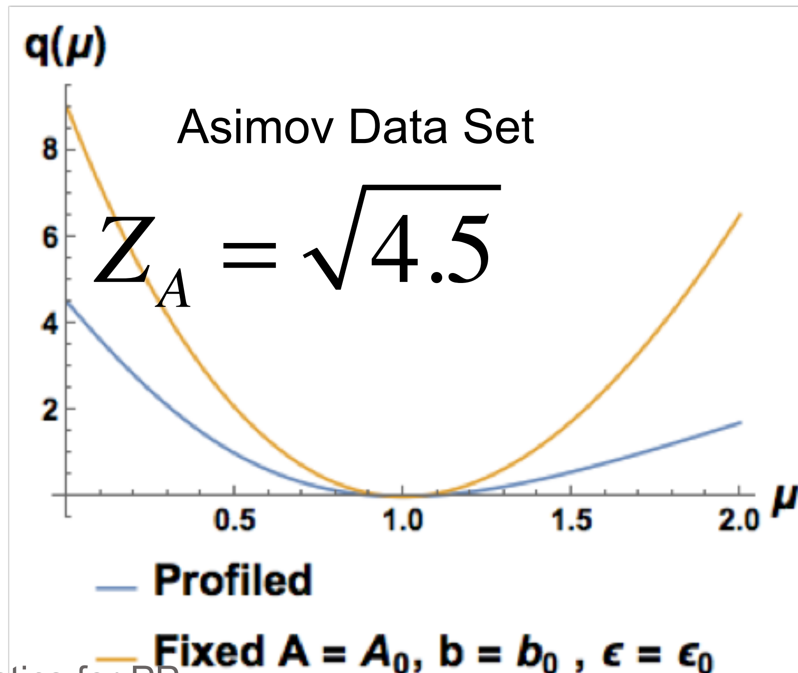
random data set

$b_{meas} = 106.84$
$\varepsilon_{meas} = 0.523$
$A_{meas} = 0.477$
$\mu_{meas} = 0.629$
$n_{meas} = 121$

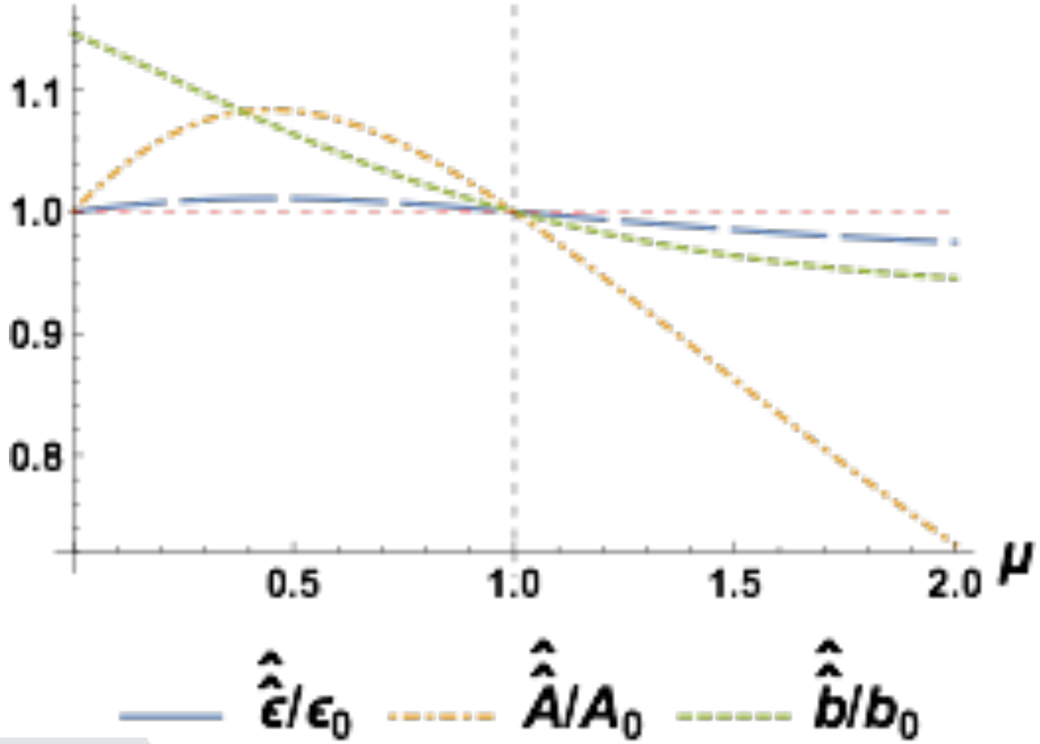
For the fixed data set
The Nuisance Parameters
Are fixed to their nominal values.
The likelihood are more parabolic,
yet, never symmetric
The asymptotocs hold!



background = 100
signal = 90
ε = 0.5
A=0.7
σ_ε = **0.05**
σ_b = 10
σ_A = **0.2**



Asimov Data Set



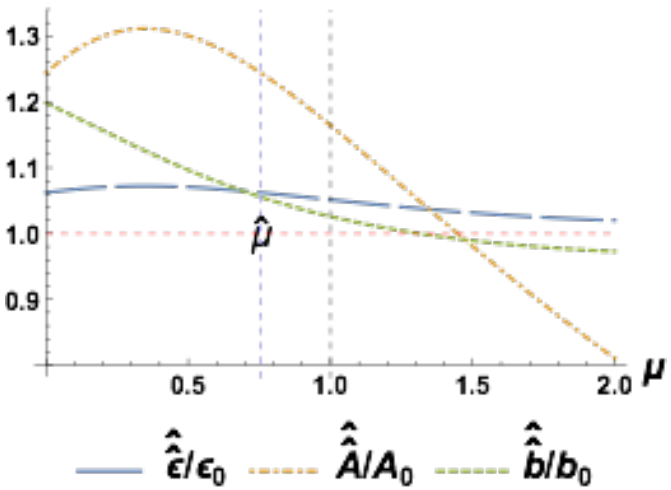
background = 100
signal = 90
ε = 0.5
A=0.7
σ_ε = 0.05
σ_b = 10
σ_A = 0.2

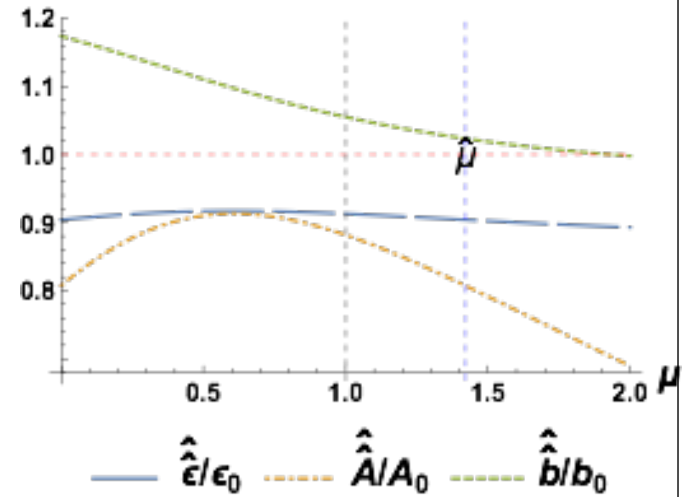
Random Data Set (with signal)

n_{meas} =	137
b_{meas} =	105.533
ϵ_{meas} =	0.531025
A_{meas} =	0.870554
μ_{meas} =	0.756304

background = 100
signal = 90
ε = 0.5
A=0.7
σ_ε = 0.05
σ_b = 10
σ_A = 0.2

n_{meas} =	135
b_{meas} =	102.337
ϵ_{meas} =	0.452067
A_{meas} =	0.565271
μ_{meas} =	1.42021



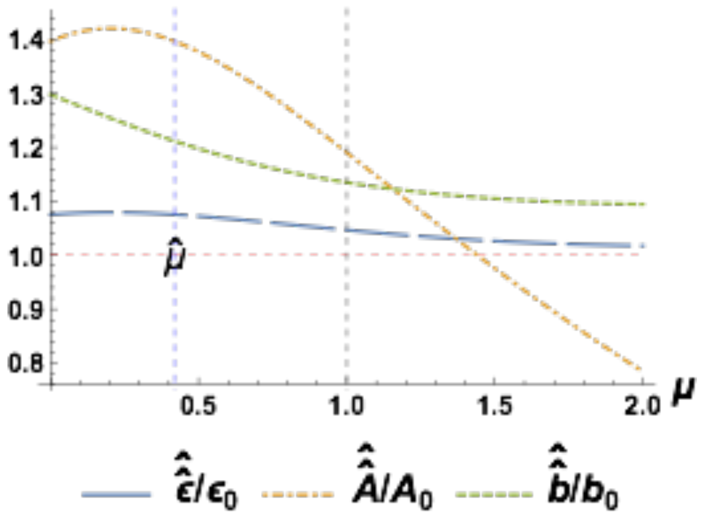


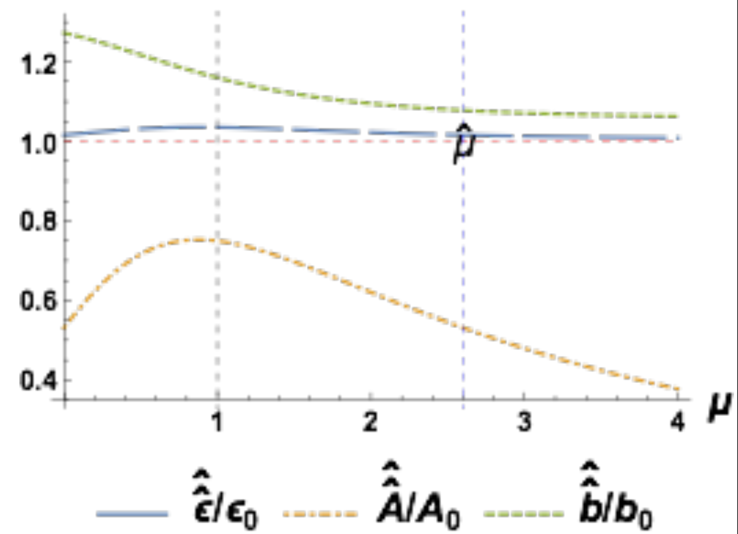
Random Data Set (with signal)

n_{meas} =	141
b_{meas} =	121.143
ϵ_{meas} =	0.53765
A_{meas} =	0.977535
μ_{meas} =	0.419804

background = 100
signal = 90
ε = 0.5
A=0.7
σ_ε = 0.05
σ_b = 10
σ_A = 0.2

n_{meas} =	152
b_{meas} =	107.781
ϵ_{meas} =	0.507957
A_{meas} =	0.371606
μ_{meas} =	2.60291





Pulls and Ranking of NPs

The pull of θ_i is given by $\dfrac{\hat{\theta}_i - \theta_{0,i}}{\sigma_0}$

without constraint $\sigma\left(\dfrac{\hat{\theta}_i - \theta_{0,i}}{\sigma_0}\right) = 1 \quad \left\langle \dfrac{\hat{\theta}_i - \theta_{0,i}}{\sigma_0} \right\rangle = 0$

It's a good habit to look at the pulls of the NPs and make sure that
Nothing irregular is seen

In particular one would like to guarantee that the fits do not over constrain
A NP in a non sensible way

Asimov

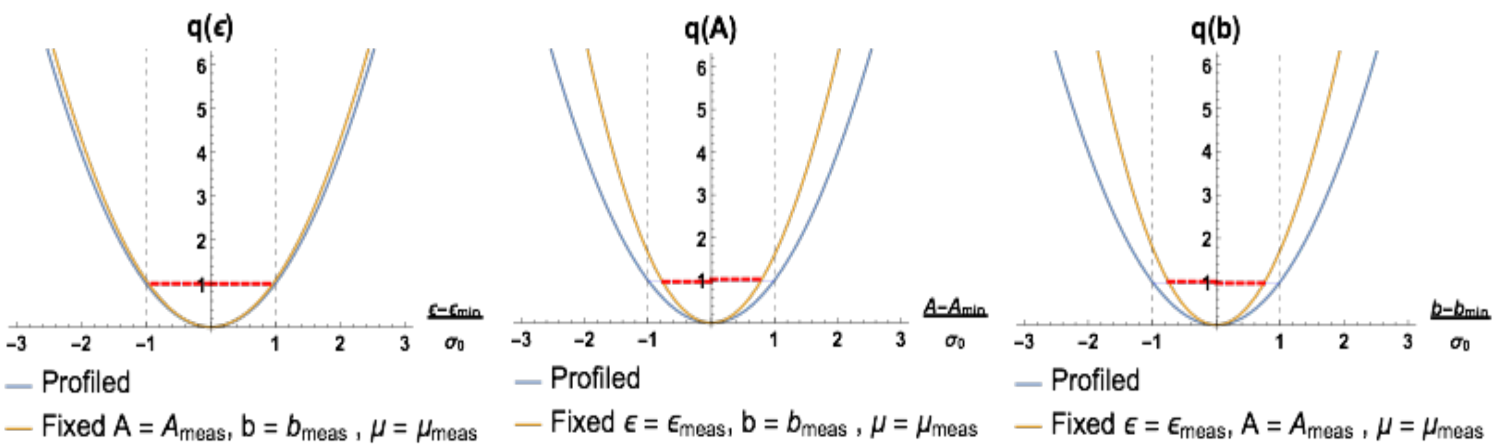
b_{meas} = 100
ε_{meas} = 0.5
A_{meas} =0.7
μ_{meas} = 1
$n_{meas}=\mu s\varepsilon A+b=131.5$

→

reminder:
b_0 = 100
ε_0 = 0.5
A_0 =0.7
μ_0 = 1
n_0=131.5
signal =90

σ
0
σ_ε = 0.05
σ_b = 10
σ_A = 0.2

To get the pulls:
–scan $q(\epsilon)$
–Find $\hat{\epsilon}$
–Find σ_ϵ^+ and σ_ϵ^- i.e. the poitive and negative error bar substituting $q(\epsilon)=1$



With the Asimov data sets we find perfect pulls for the profiled scans
But not for the fix scans!

Random Data Set

n_{meas} = 132

b_{meas} = 103.208

ϵ_{meas} = 0.465459

A_{meas} = 0.487107

μ_{meas} = 1.41099

reminder:
$b_0 = 100$
$\varepsilon_0 = 0.5$
$A_0 = 0.7$
$\mu_0 = 1$
$n_0 = 131.5$
signal =90

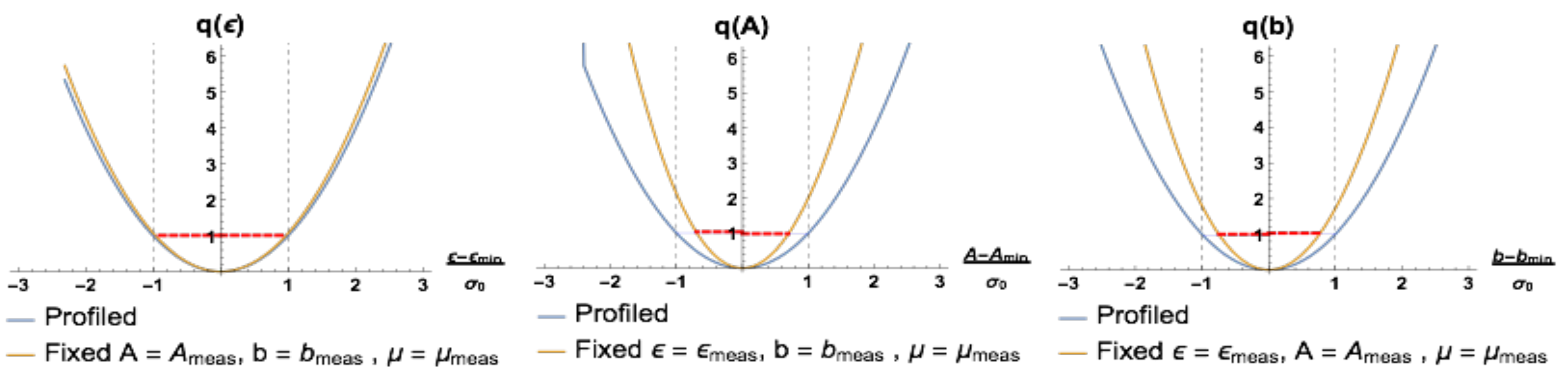
σ
0

$\sigma_\varepsilon = 0.05$

$\sigma_b = 10$

$\sigma_A = 0.2$

To get the pulls:
–scan $q(\epsilon)$
–Find $\hat{\epsilon}$
–Find σ_ϵ^+ and σ_ϵ^- i.e. the poitive and negative error bar substituting $q(\epsilon) = 1$



With the random data sets we find perfect pulls for the profiled scans
But not for the fix scans!

Back to Asimov: Find the Impact of a NP

b_{meas} = 100
ε_{meas} = 0.5
A_{meas} =0.7
μ_{meas} = 1
$n_{meas}=\mu s\varepsilon A+b=131.5$

reminder:
b_0 = 100
ε_0 = 0.5
A_0 =0.7
μ_0 = 1
n_0=131.5
signal =90

σ
0
σ_ε = 0.05
σ_b = 10
σ_A = 0.2

To get the impact of a Nuisance Parameter in order to rank them:

Say we want the impact of ϵ

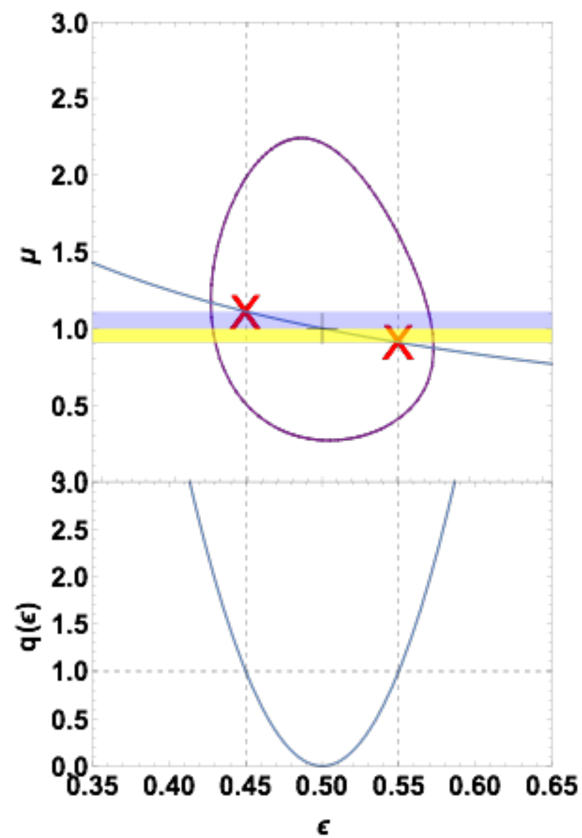
–Scan q(ϵ), profiling all other NPs

–Find $\hat{\epsilon}$

–(note that $\hat{\mu}_{\hat{\epsilon}} = \hat{\mu}$)

–Find $\hat{\mu}_{\hat{\epsilon}\pm\sigma_\epsilon^\pm} = \hat{\hat{\mu}}_{\hat{\epsilon}\pm\sigma_\epsilon^\pm}$

–The impact is given by $\Delta\mu^\pm = \hat{\hat{\mu}}_{\hat{\epsilon}\pm\sigma_\epsilon^\pm} - \hat{\mu}$



Random Data Set: Find the Impact of NP

n_{meas} = 132
b_{meas} = 103.208
ϵ_{meas} = 0.465459
A_{meas} = 0.487107
μ_{meas} = 1.41099

reminder:
b_0 = 100
ε_0 = 0.5
A_0 =0.7
μ_0 = 1
n_0=131.5
signal =90

σ
0
σ_ε = 0.05
σ_b = 10
σ_A = 0.2

To get the impact of a Nuisance Parameter in order to rank them:

Say we want the impact of ϵ

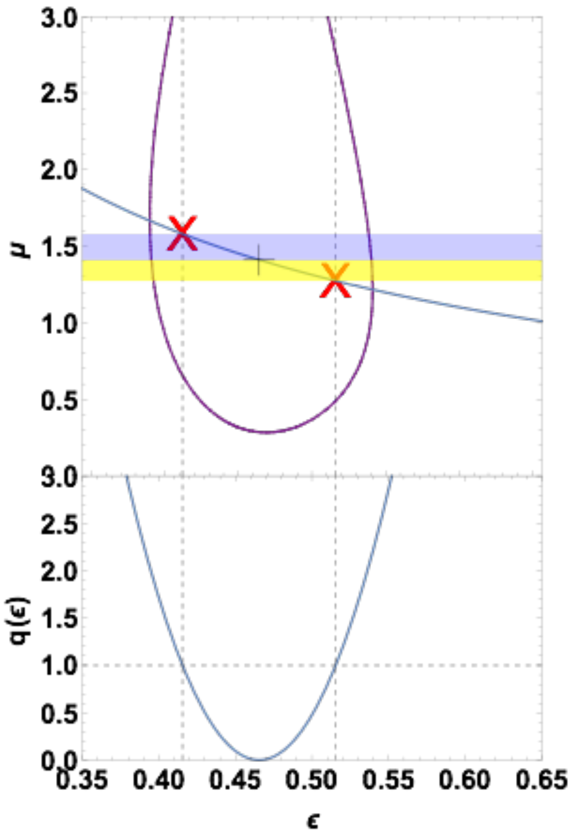
–Scan q(ϵ), profiling all other NPs

–Find $\hat{\epsilon}$

–(note that $\hat{\mu}_{\hat{\epsilon}} = \hat{\mu}$)

–Find $\hat{\mu}_{\hat{\epsilon}\pm\sigma_\epsilon^\pm} = \hat{\hat{\mu}}_{\hat{\epsilon}\pm\sigma_\epsilon^\pm}$

–The impact is given by $\Delta\mu^\pm = \hat{\hat{\mu}}_{\hat{\epsilon}\pm\sigma_\epsilon^\pm} - \hat{\mu}$



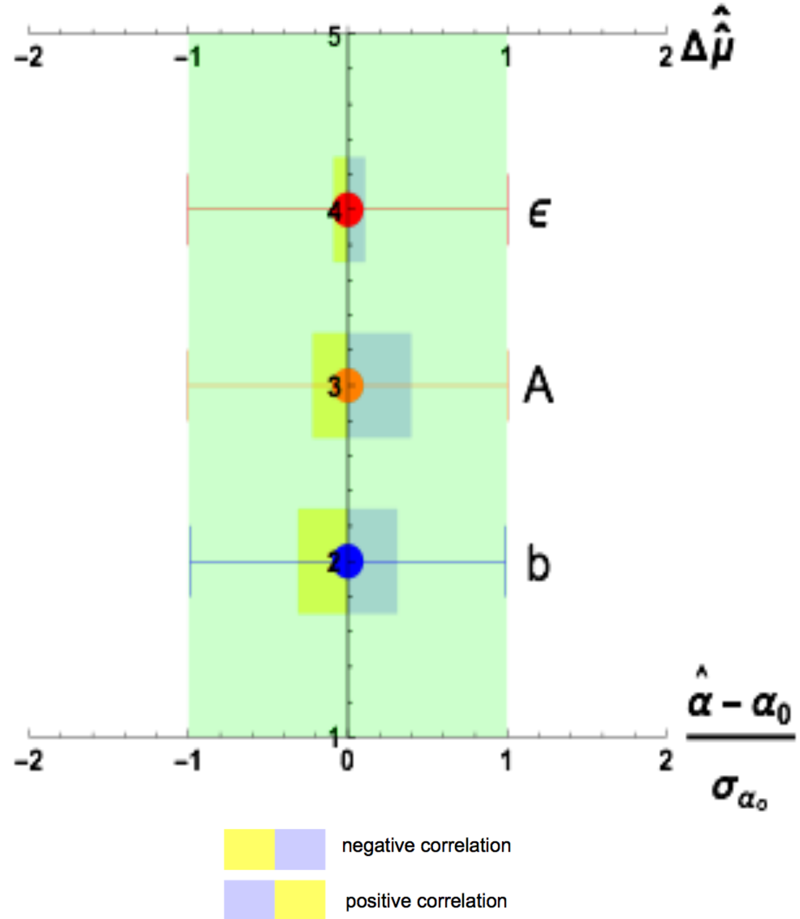
Asimov: SUMMARY of Pulls and Impact

b_{meas} = 100
ε_{meas} = 0.5
A_{meas} =0.7
μ_{meas} = 1
$n_{meas}=\mu s\varepsilon A+b=131.5$

reminder:
b_0 = 100
ε_0 = 0.5
A_0 =0.7
μ_0 = 1
n_0=131.5
signal =90

σ 0

σ_ε = 0.05
σ_b = 10
σ_A = 0.2



Random Data Set: SUMMARY of Pulls and Impact

n_{meas} = 132

b_{meas} = 103.208

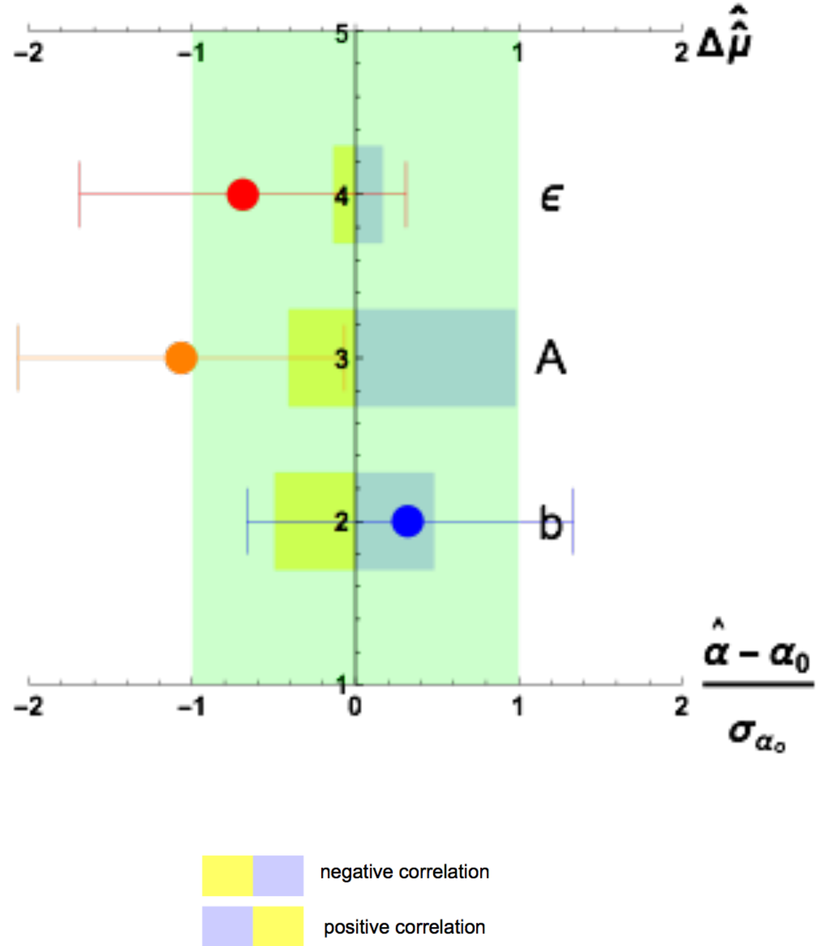
ϵ_{meas} = 0.465459

A_{meas} = 0.487107

μ_{meas} = 1.41099

reminder:
b_0 = 100
ε_0 = 0.5
A_0 =0.7
μ_0 = 1
n_0=131.5
signal =90

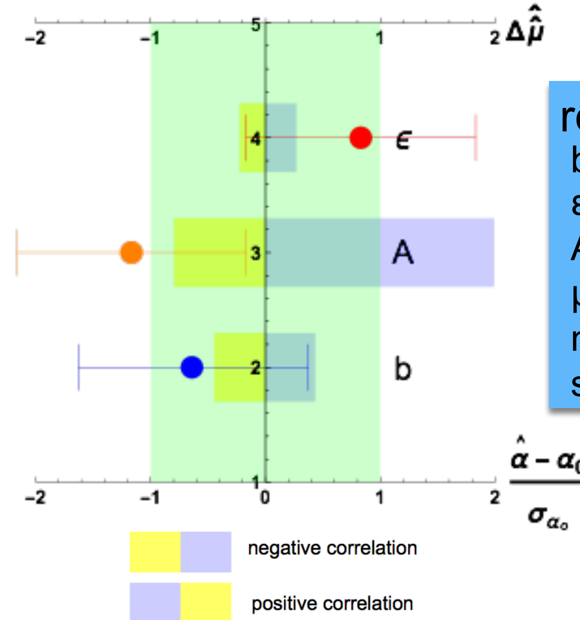
σ
0
σ_ε = 0.05
σ_b = 10
σ_A = 0.2



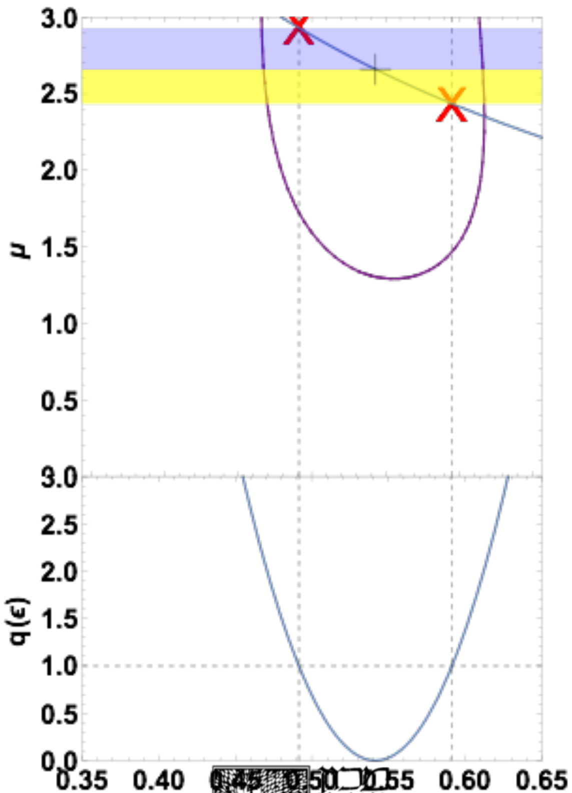
Pulls and Impacts: More examples

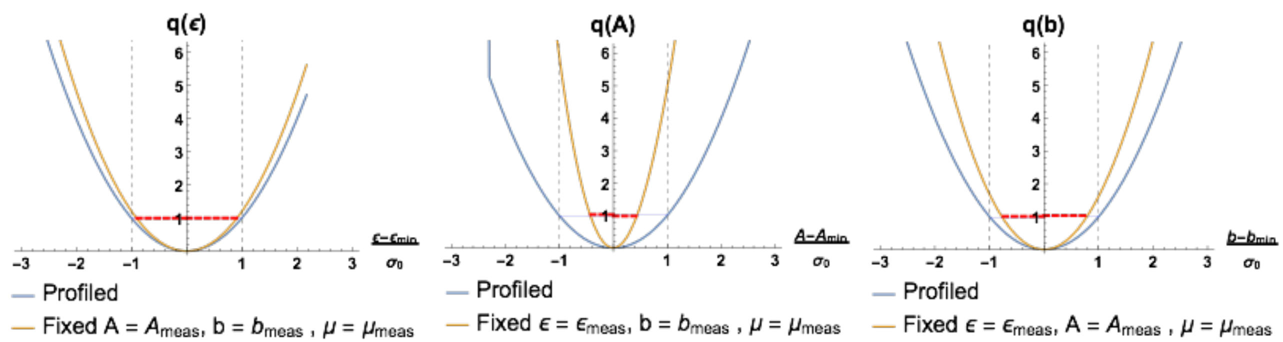
n_{meas} = 154
b_{meas} = 93.6307
ϵ_{meas} = 0.541389
A_{meas} = 0.465922
μ_{meas} = 2.6592

σ
0
σ_ε = 0.05
σ_b = 10
σ_A = 0.2



reminder:
b_0 = 100
ε_0 = 0.5
A_0 =0.7
μ_0 = 1
n_0=131.5
signal =90





$n_{meas} =$ 120

$b_{meas} =$ 104.334

$\epsilon_{meas} =$ 0.487497

$A_{meas} =$ 0.666568

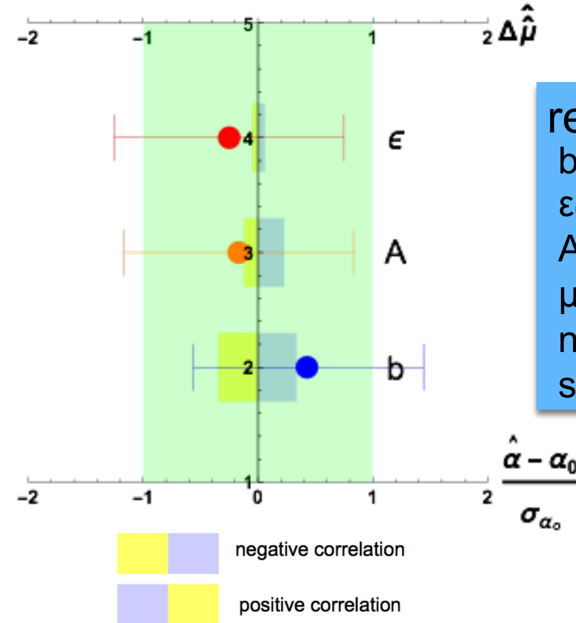
$\mu_{meas} =$ 0.535663

σ
0

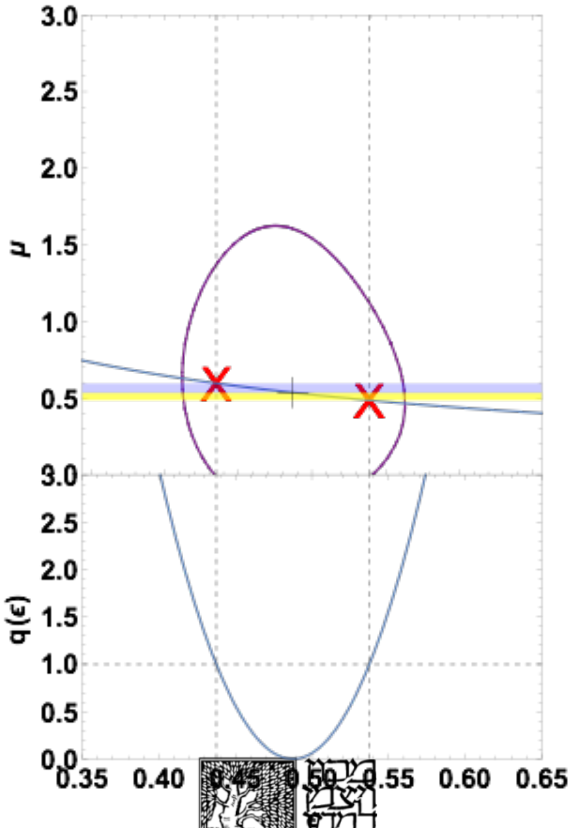
σ_ϵ = 0.05

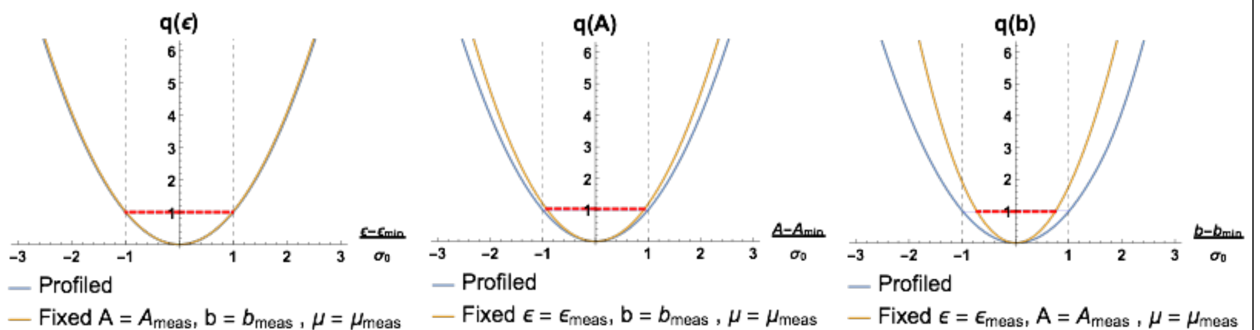
σ_b = 10

σ_A = 0.2



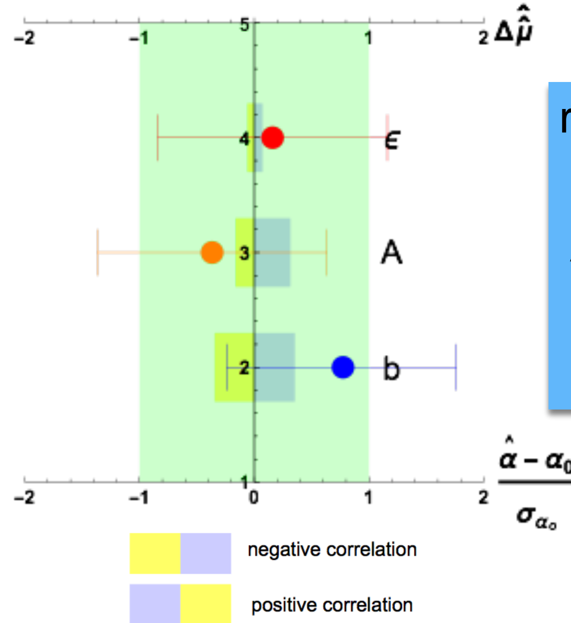
reminder:
b_0 = 100
ε_0 = 0.5
A_0 =0.7
μ_0 = 1
n_0=131.5
signal =90



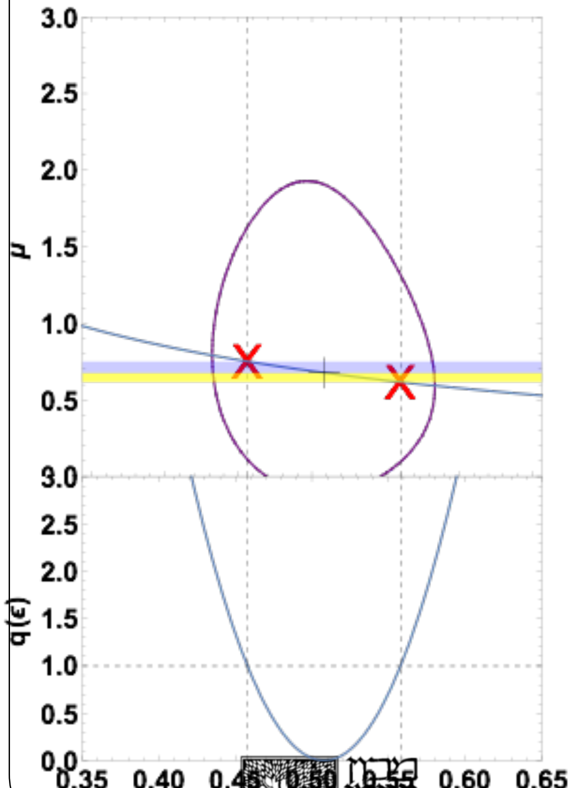


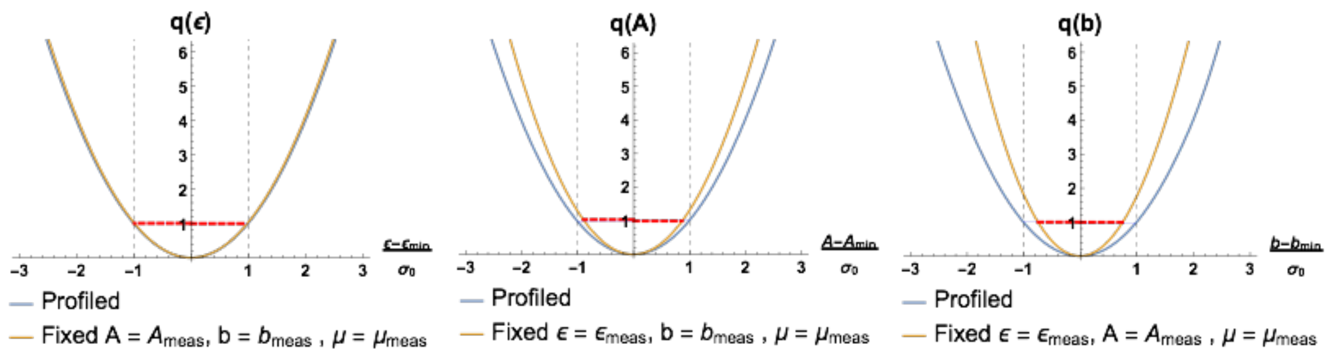
n_{meas} = 127
b_{meas} = 107.675
ϵ_{meas} = 0.507863
A_{meas} = 0.62459
μ_{meas} = 0.676915

σ_0
σ_ε = 0.05
σ_b = 10
σ_A = 0.2



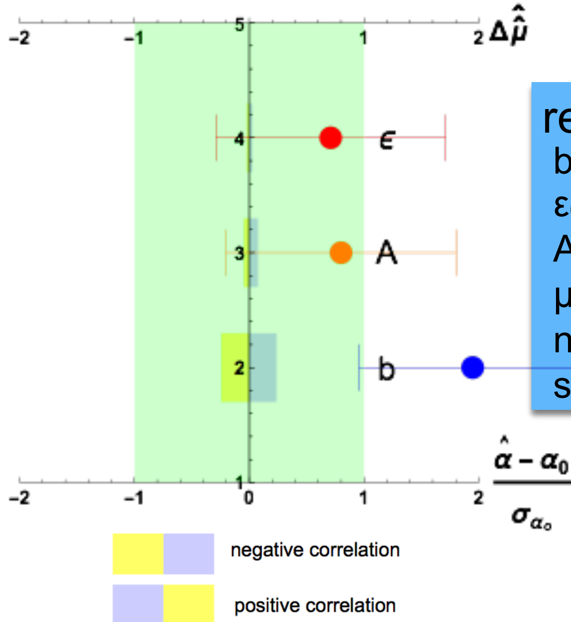
reminder:
b_0 = 100
ε_0 = 0.5
A_0 =0.7
μ_0 = 1
n_0=131.5
signal =90



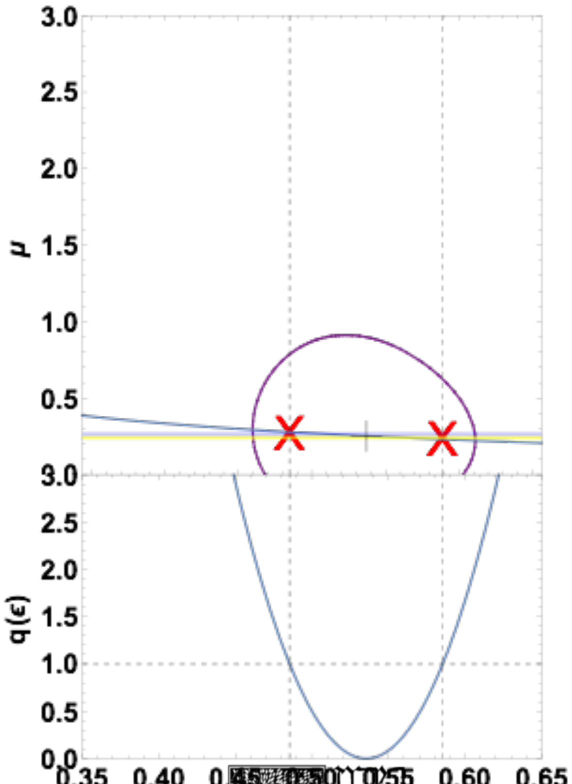


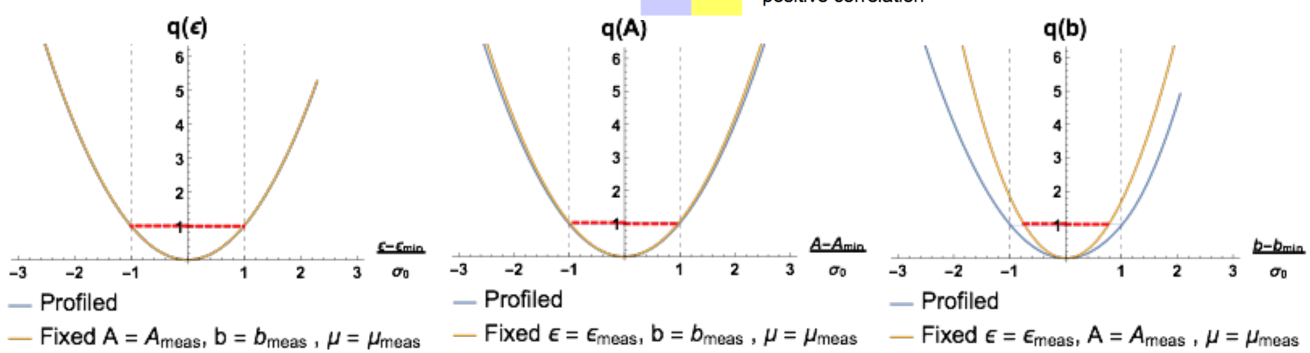
n_{meas} = 130
b_{meas} = 119.561
ϵ_{meas} = 0.535367
A_{meas} = 0.858264
μ_{meas} = 0.252421

σ
0
σ_ε = 0.05
σ_b = 10
σ_A = 0.2

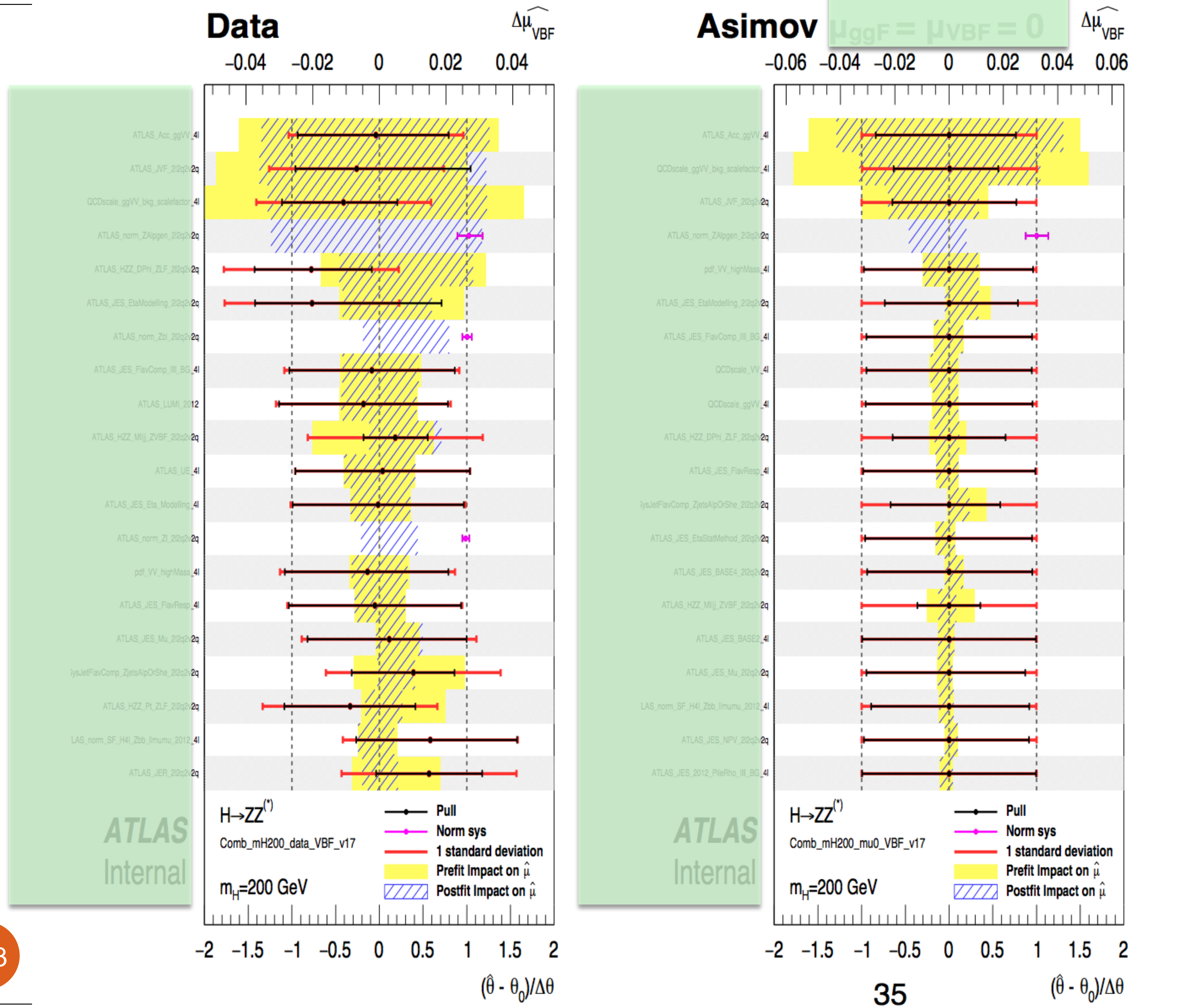


reminder:
b_0 = 100
ε_0 = 0.5
A_0 =0.7
μ_0 = 1
n_0=131.5
signal =90





Real Examples

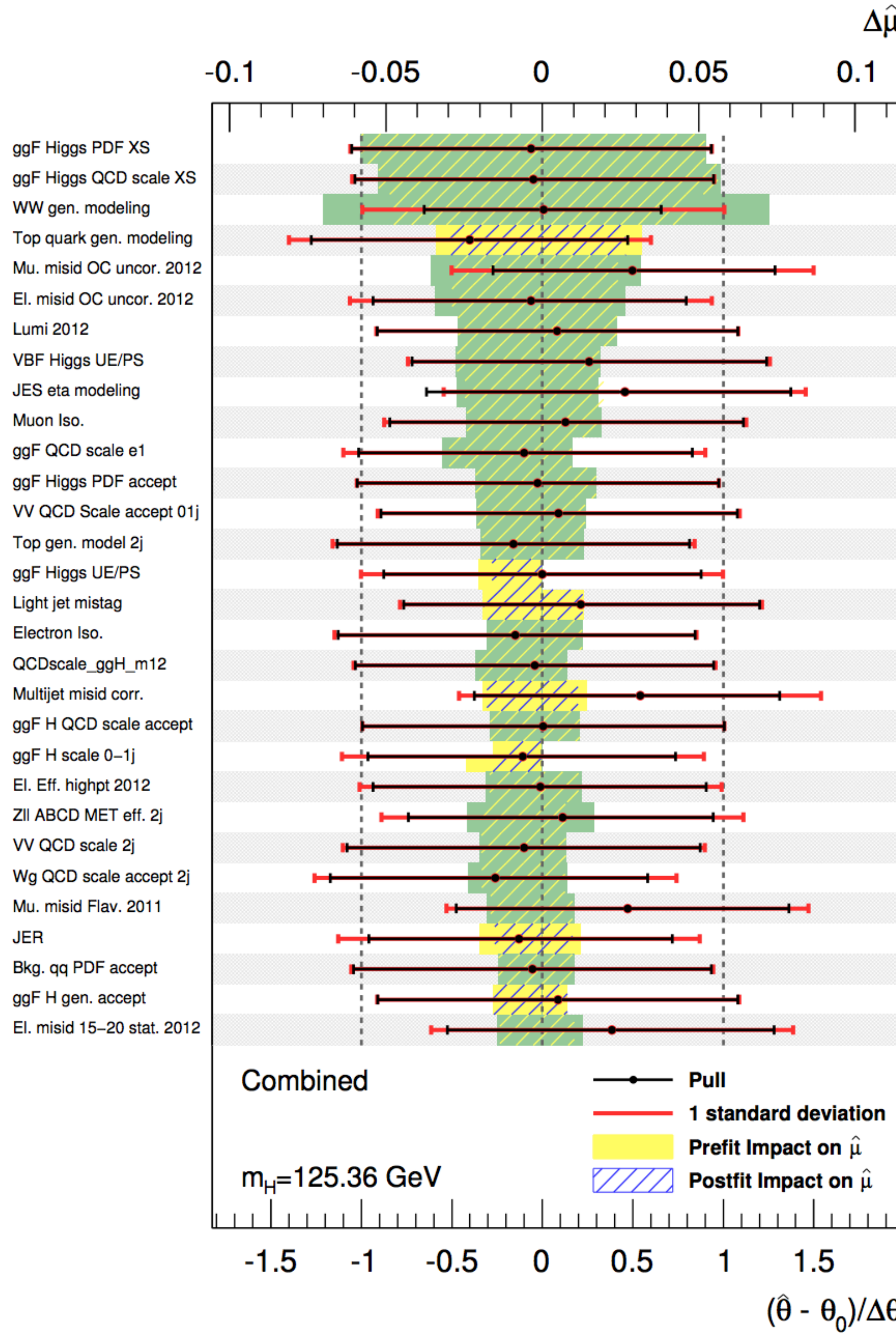

Data
Asimov
μggF = μVBF = 0
$\Delta\widehat{\mu}_{VBF}$
ATLAS_Acc_ggVV_4l
ATLAS_JVF_2l2q2ν2q
QCDscale_ggVV_bkg_scalefactor_4l
ATLAS_norm_ZAlpgen_2l2q2ν2q
ATLAS_HZZ_DPhi_ZLF_2l2q2ν2q
ATLAS_JES_EtaModelling_2l2q2ν2q
ATLAS_norm_Zbl_2l2q2ν2q
ATLAS_JES_FlavComp_lll_BG_4l
ATLAS_LUMI_2012
ATLAS_HZZ_Mljj_ZVBF_2l2q2ν2q
ATLAS_UE_4l
ATLAS_JES_Eta_Modelling_4l
ATLAS_norm_Zl_2l2q2ν2q
pdf_VV_highMass_4l
ATLAS_JES_FlavResp_4l
ATLAS_JES_Mu_2l2q2ν2q
QCDscale_ggVV_4l
QCDscale_VV_4l
ATLAS_JES_EtaStatMethod_2l2q2ν2q
ATLAS_JES_BASE4_2l2q2ν2q
ATLAS_JES_BASE2_4l
ATLAS_JES_NPV_2l2q2ν2q
ATLAS_JES_2012_PileRho_lll_BG_4l
ATLAS_HZZ_Pt_ZLF_2l2q2ν2q
ATLAS_JER_2l2q2ν2q
ATLAS Internal
H→ZZ(*)
Comb_mH200_data_VBF_v17
Comb_mH200_mu0_VBF_v17
m_H=200 GeV
Pull
Norm sys
1 standard deviation
Prefit Impact on $\hat{\mu}$
Postfit Impact on $\hat{\mu}$
$(\hat{\theta} - \theta_0)/\Delta\theta$

Pulls and Ranking

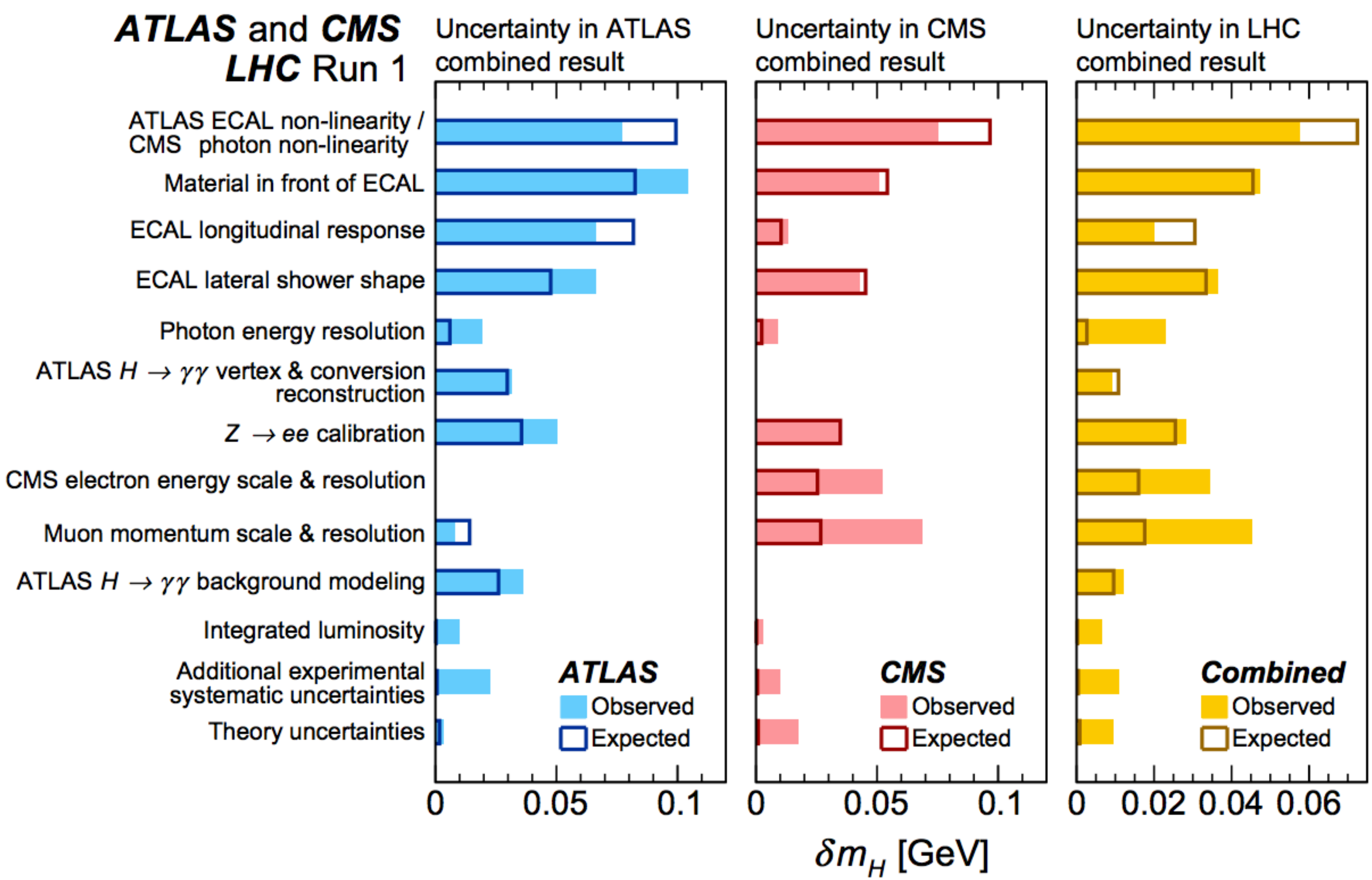
Ranking θ_i by its effect in the NP

$$\Delta\mu^{\pm} = \hat{\hat{\mu}}_{\hat{\epsilon}\pm\sigma_{\epsilon}^{\pm}} - \hat{\mu}$$

By ranking we can tell which NPs are the important ones and which can be pruned



The Higgs Mass Paper



Bloggers Spot

A combination on a back of an envelope

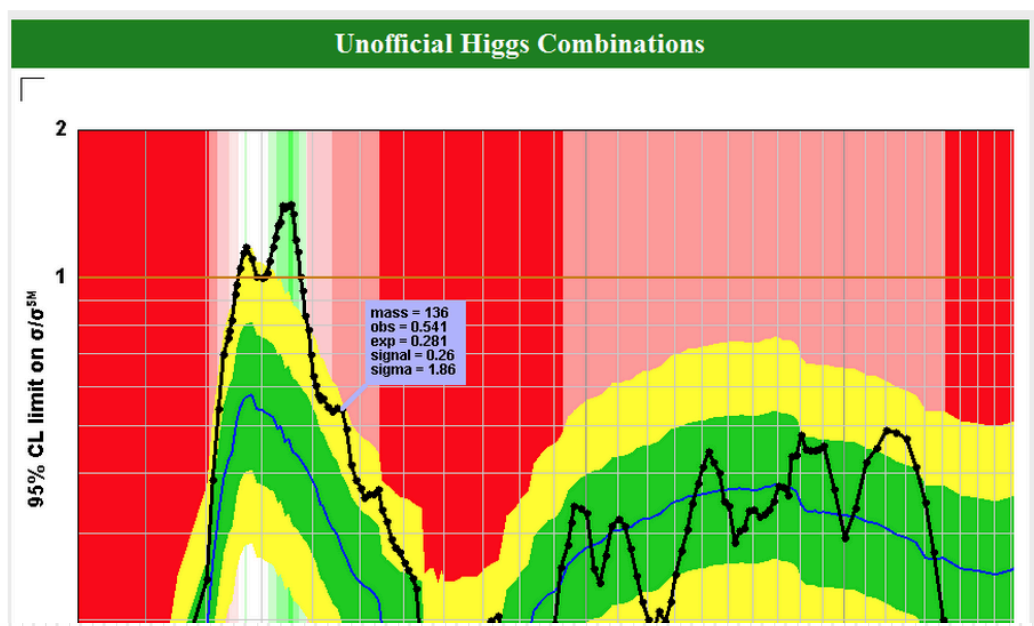
viXra log

Home | About | Science Calendar

Higgs Combination Applet

I have been showing unofficial Higgs combinations here for the last year or so but maybe you want to try some unusual combinations of your own. Now you can using the viXra unofficial Higgs combination Java applet. It is armed with most of the plots published by the experiments CDF, D0, CMS, ATLAS and LEP. You just have to choose how to combine them. I am hoping it is self-explanatory but ask some questions and you may get some good tips. You may need to update your Java plug-in.

Disclaimer: The results are approximate, unofficial and not endorsed by the experiments.



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

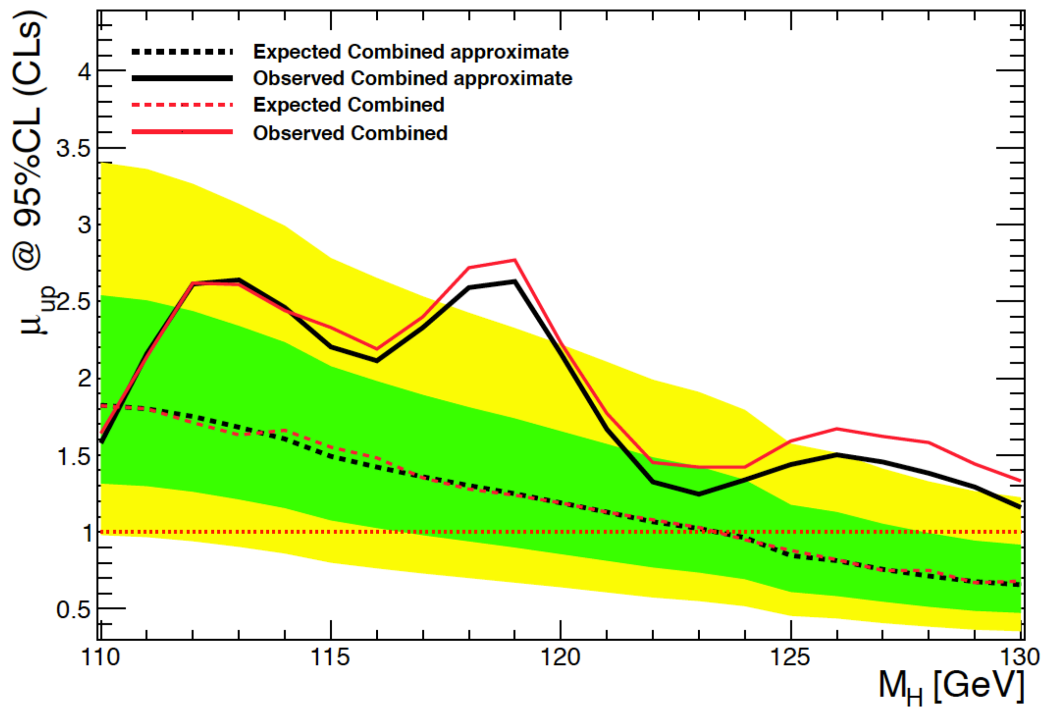
$$\mu_{up}^{med} = \sigma\Phi^{-1}(1 - \alpha/2)$$

- Which gives

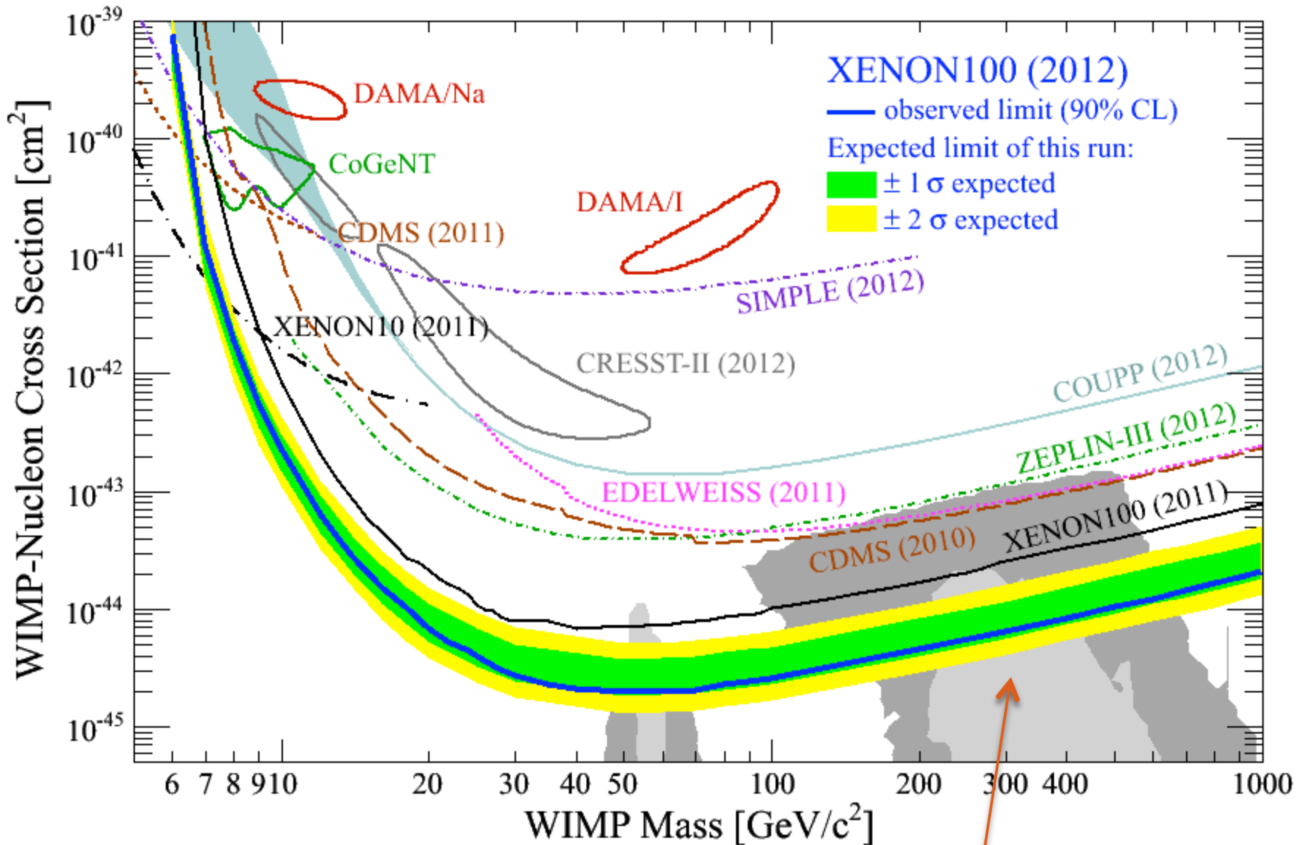
$$\frac{1}{(\mu_{up}^{med})^2} = \frac{1}{(\mu_{up,1}^{med})^2} + \frac{1}{(\mu_{up,2}^{med})^2}$$

An exercise in combining experiments (or channels)

- This combination takes onto account fluctuations of the observed limit



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