Statistical Challenges in HEP

Eilam Gross

Weizmann Institute of Science

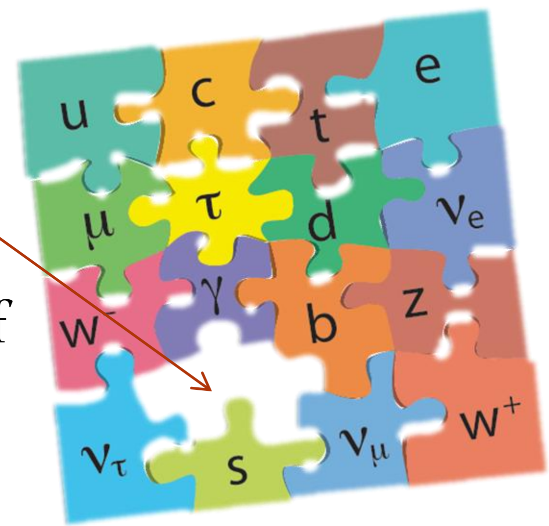
In collaboration with **Ofer Vitells**

Acknowledgements:

Louis Lyons

What is the statistical challenge in HEP?

- High Energy Physicists (HEP) have an hypothesis: The Standard Model.
- This model breaks unless there exists its only one ingredient, yet to be discovered: the Higgs Boson
- The minimal content of the Standard Model includes the Higgs Boson , but extensions of the Model include other particles which are yet to be discovered
- The challenge of HEP is to generate tons of data and to develop powerful analyses to tell if the data contains evidence for new particles, perhaps expected, but yet to be discovered

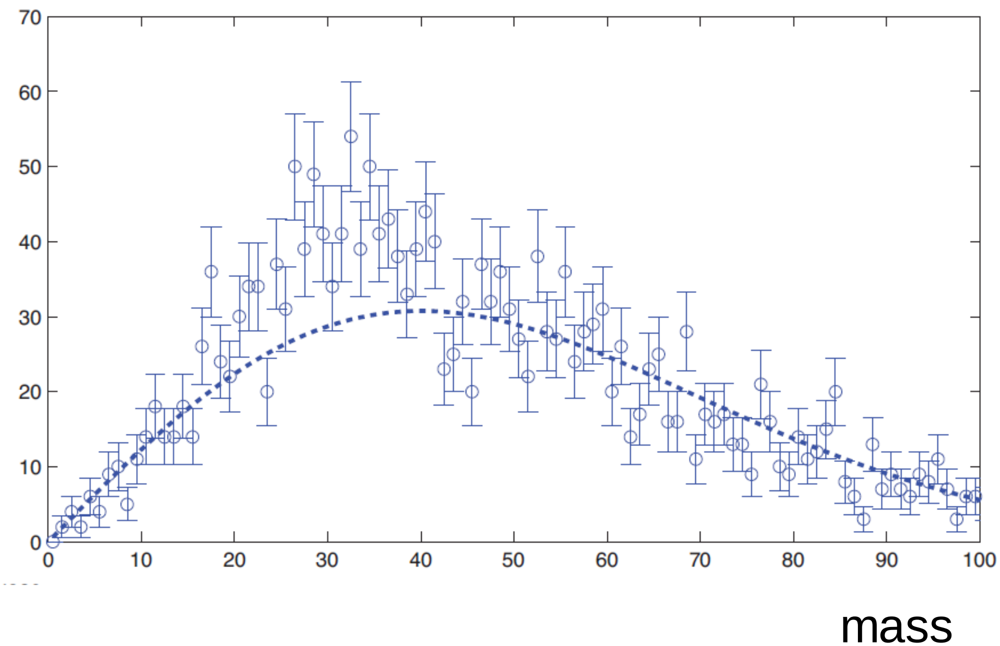


The Large Hadron Collider (LHC)

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

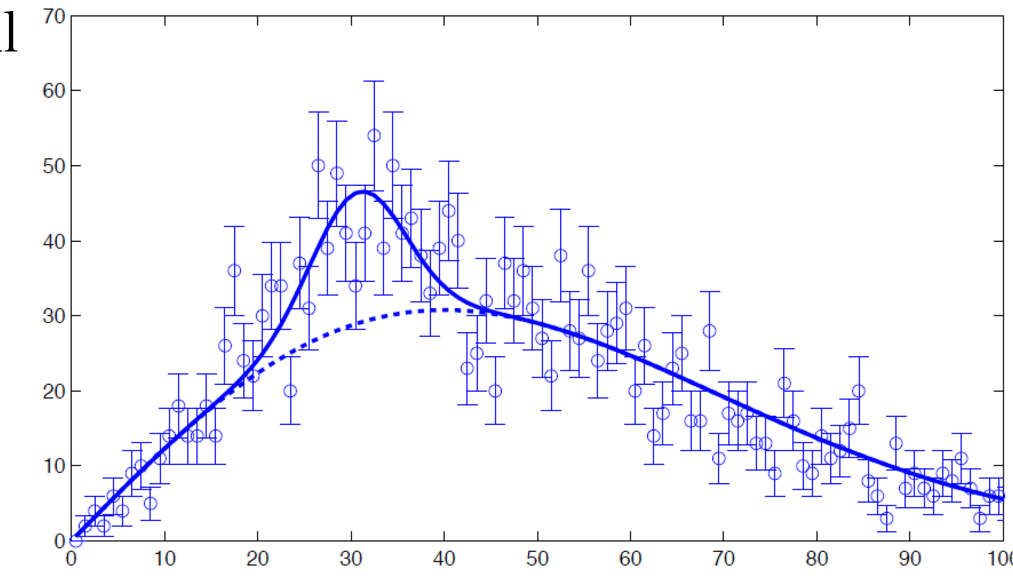
The Statistical Challenge of HEP

- The DATA: Billions of Proton-Proton collisions which could be visualized with histograms
- The Higgs mass is unknown
- In this TOY example, we ask if the expected background (the Standard Model WITHOUT the Higgs Boson) contains a Higgs Boson, which would manifest itself as a peak in the distribution



The Statistical Challenge of HEP

- So the statistical challenge is obvious:
- To tell in the most powerful way, and to the best of our current scientific knowledge, if there is new physics, beyond what is already known, in our data
- The complexity of the apparatus and the background physics suffer from large systematic errors that should be treated in an appropriate way.



mass

An outline of the talk

Statistical methods, multivariate analysis and pattern recognition Teodorescu Liliana

Pattern recognition and estimation methods for track and vertex reconstruction Fruhwirth Rudolf

I could talk about **linear classifiers, neural networks, boosted decision trees and support vector machines.**

An outline of the talk

Statistical methods, multivariate analysis and pattern recognition Teodorescu Liliana

Patte **ack**
and

This talk takes

a different route

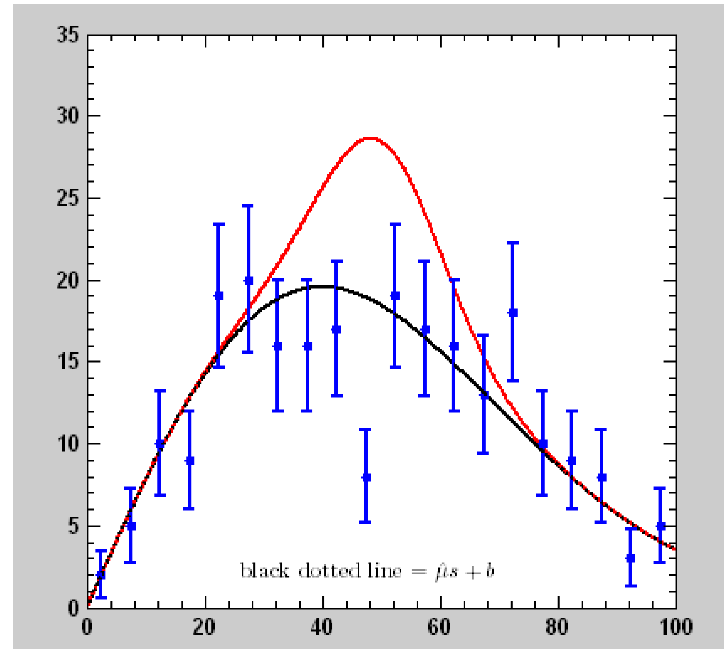
I cou **s,**
boos **es.**

An outline of the talk: **Hypothesis Testing**

- Definition of the null and alternate hypotheses
- Definition of discovery and exclusion
- The two "top" Likelihood Ratio (**LR**) test statistic in the HEP market:
  the Neyman-Pearson (**NP**) LR and the Profile-Likelihood
- Systematics: the treatment of nuisance parameters:
  - Profiling in a frequentist way: Profiled NP, Profile Likelihood
  - Integrating in a Bayesian way
  - The Cousins-Highland hybrid way
- The Look Elsewhere Effect
- Exclusions and **CL** (Confidence or Credibility intervals)

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 Null Hypothesis

- The Standard Model without the Higgs is an hypothesis, (BG only hypothesis) many times referred to as the null hypothesis and is denoted by H_0

- In the absence of an alternate hypothesis, one would like to test the compatibility of the data with H_0

- This is actually a **goodness of fit test**

A Tale of Two Hypotheses

NULL

ALTERNATE

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

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

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

- Reject H_0 in favor of H_1 – A DISCOVERY

A Tale of Two Hypotheses

NULL

H_1- SM with Higgs

ALTERNATE

H_0- SM w/o Higgs

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

- Reject H_0 in favor of H_1 – A DISCOVERY

- Reject H_1 in favor of H_0 – Excluding H_1

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, T, under the null hypothesis
- Compute from the observations the observed value t_{obs} *of the test statistic* T.
- Decide (based on t_{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?**

DISCOVERY

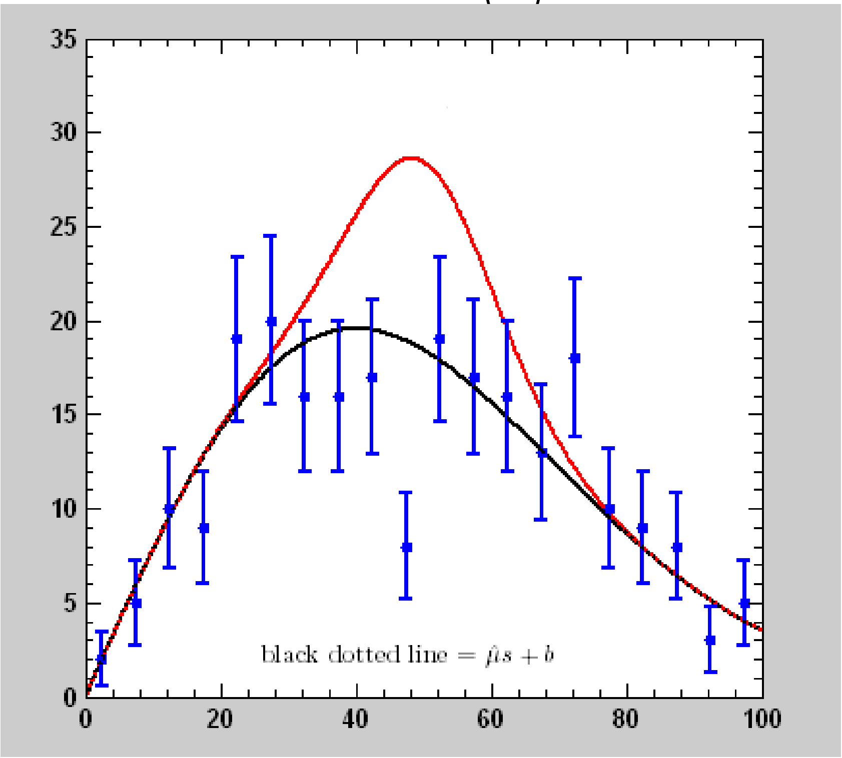
Test Statistic

- To construct a test statistic one needs a model
- $L(H_0) \sim Prob(data | H_0)$
- $L(H_1) \sim Prob(data | H_1)$
- Note: The Likelihood as indicated by its name, is the compatibility of a **given** data set with an hypothesis. If the data changes, so is the Likelihood!

The Toy Physics Model

- The NULL hypothesis H_0: SM without Higgs Background Only

$$\langle n \rangle = b$$

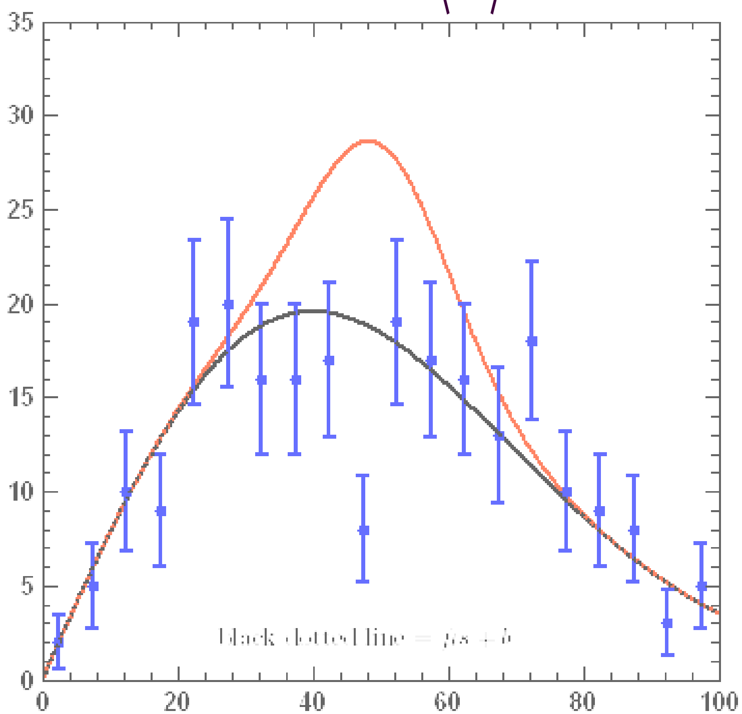


mass

The Toy Physics Model

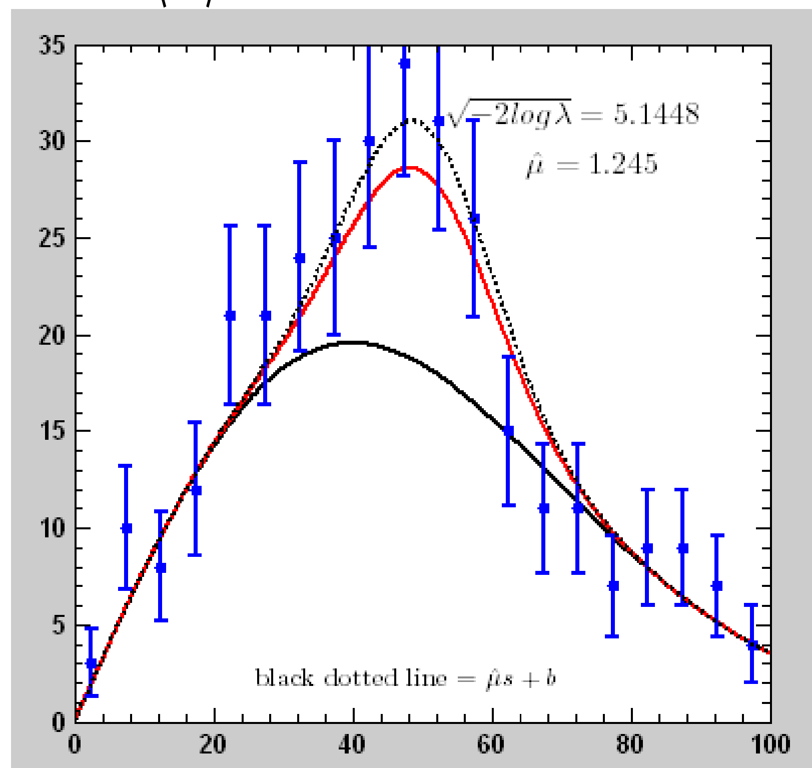
- The NULL hypothesis H_0: SM without Higgs Background Only

$$\langle n \rangle = b$$



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

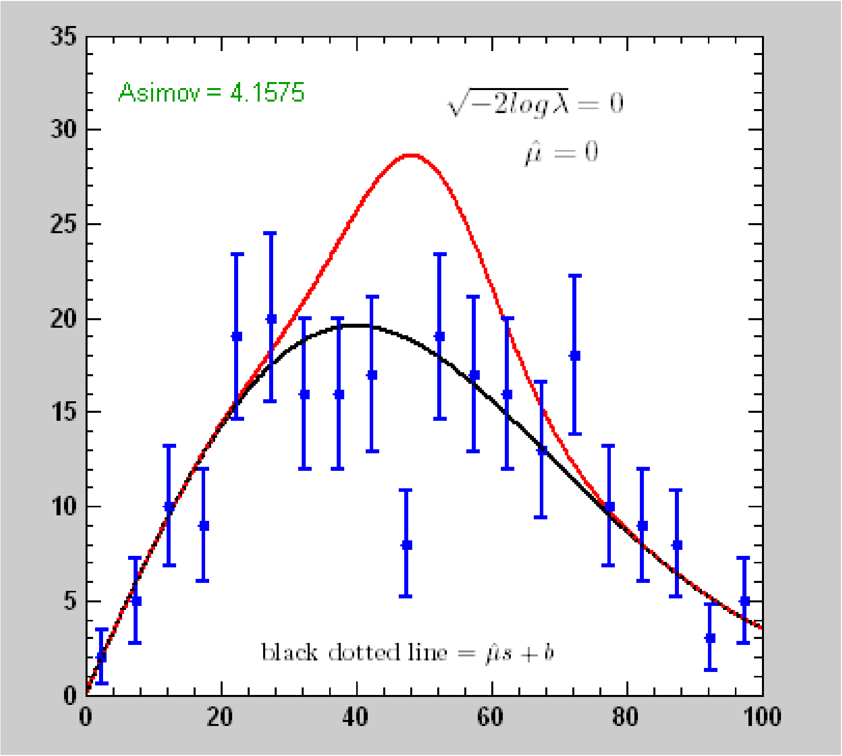
$$\langle n \rangle = s(m_H) + b$$



mass

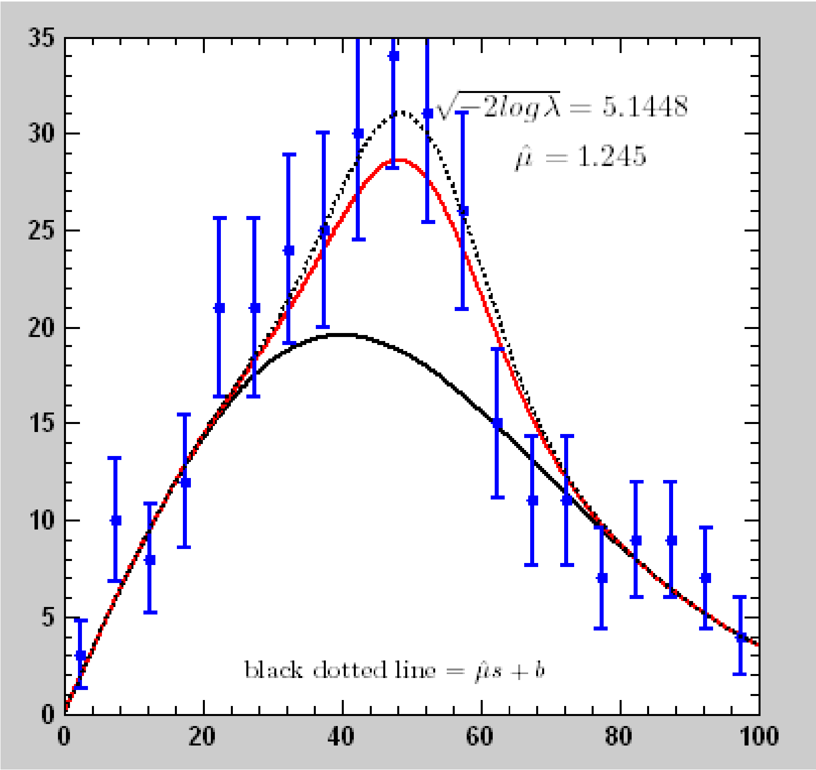
The Toy Physics Model

$$n = \mu s + b$$
$$MLE \quad \hat{\mu}$$
$$\langle \hat{\mu} \rangle = 0 \ \ under \ \ H_0$$



mass

$$n = \mu s + b$$
$$MLE \quad \hat{\mu}$$
$$\langle \hat{\mu} \rangle = 1 \ \ under \ \ H_1$$



mass

The Profile Likelihood ("PL")

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

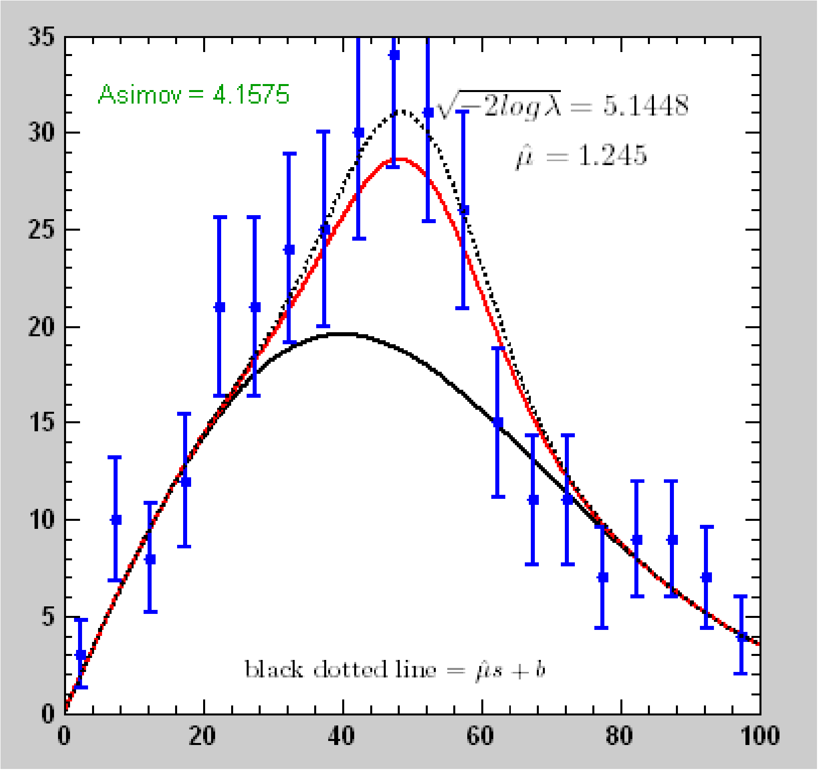
$$t_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s+b)}$$

- For $\hat{\mu} \sim 0,\ t$ small

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

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

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



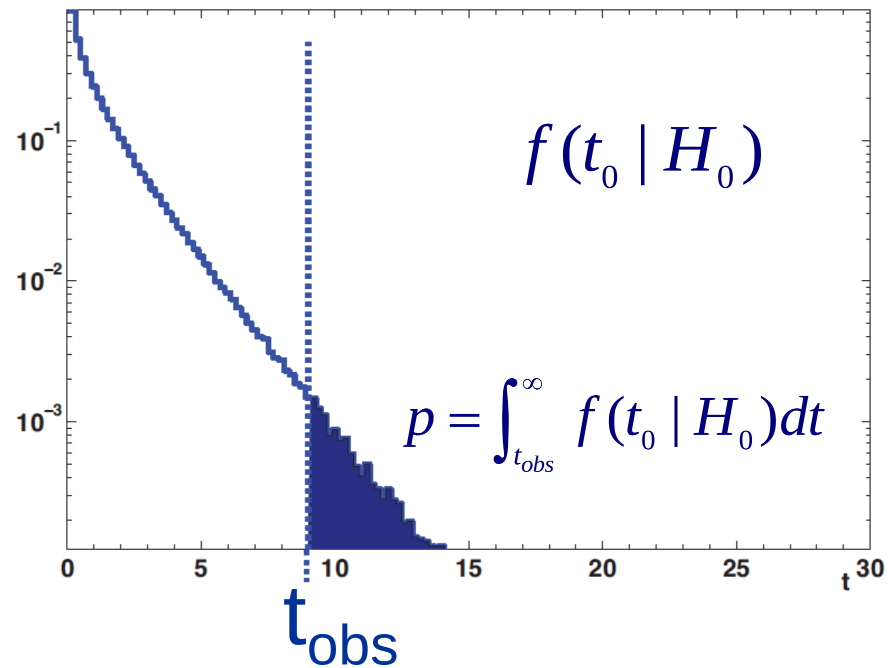
The PDF of the test statistic

$$t_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s + b)}$$

- No, not the **P**arton **D**istribution **F**unction
- Not a **P**ortable **D**ocument **F**ormat
- We need to know the **P**robability **D**istribution **F**unction of the test statistic under the null hypothesis $f(t_0 \mid H_0)$

Significance & p-value

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

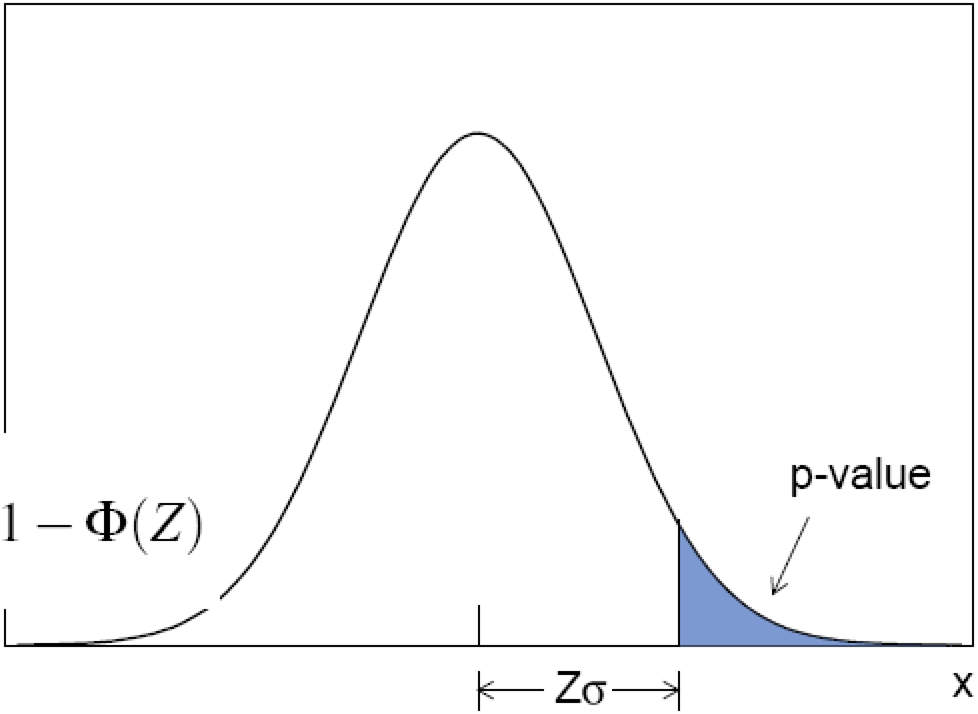


$$p = \int_{t_{obs}}^{\infty} f(t_0 \mid H_0)dt$$

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

From p-values to Gaussian Significance

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



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

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

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

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

The Profile Likelihood ("PL")

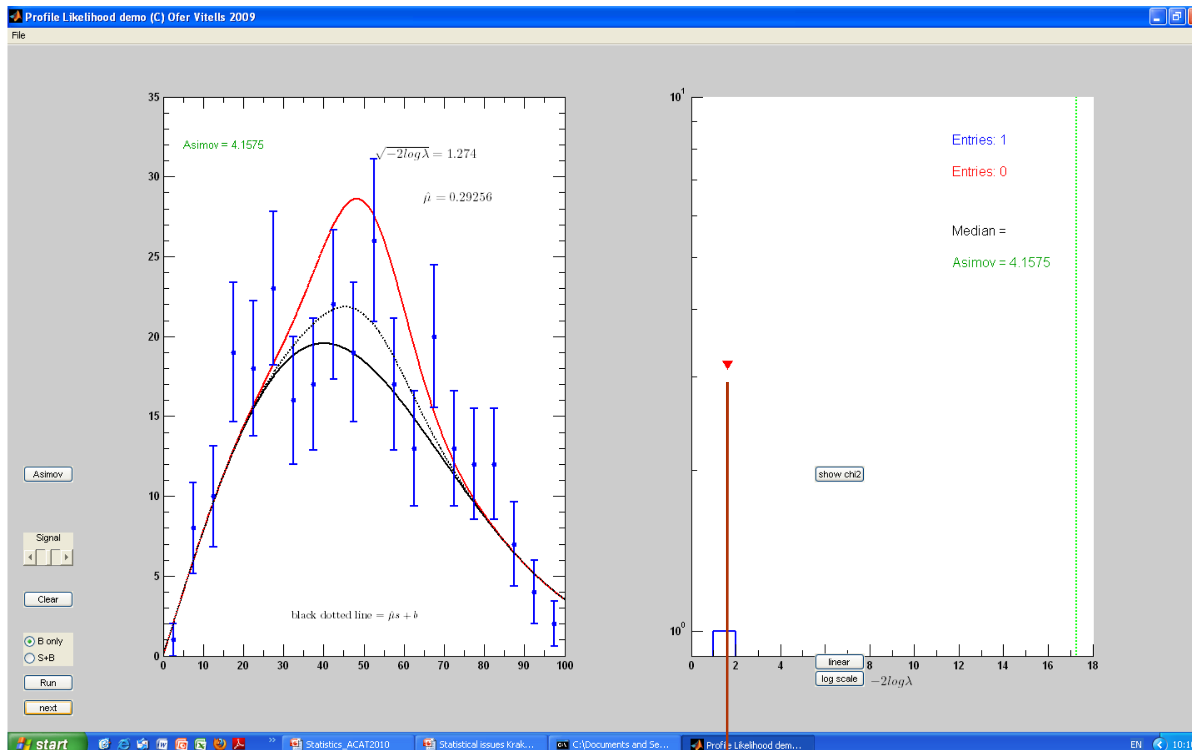
$$\langle \hat{\mu} \rangle = 0 \;\; under \;\; H_0$$

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

$$t_0 = -2\ln \frac{L(b)}{L(\hat{\mu}s + b)}$$

$$\hat{\mu} \sim 0, \;\; t \text{ small}$$

$$\hat{\mu} \sim 1, \;\; t \text{ large}$$



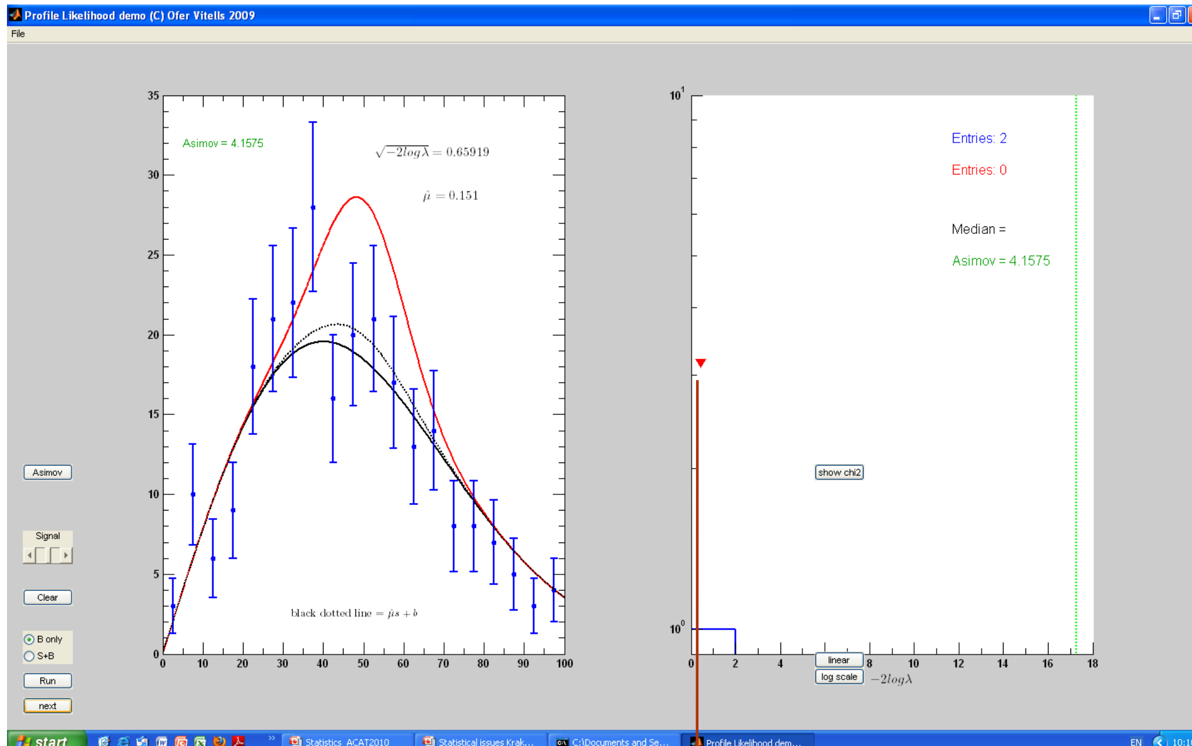
$$t = 1.6 \rightarrow Z = \sqrt{1.6} = 1.27$$

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

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

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

$$t_0 = -2\ln \frac{L(b)}{L(\hat{\mu}s + b)}$$



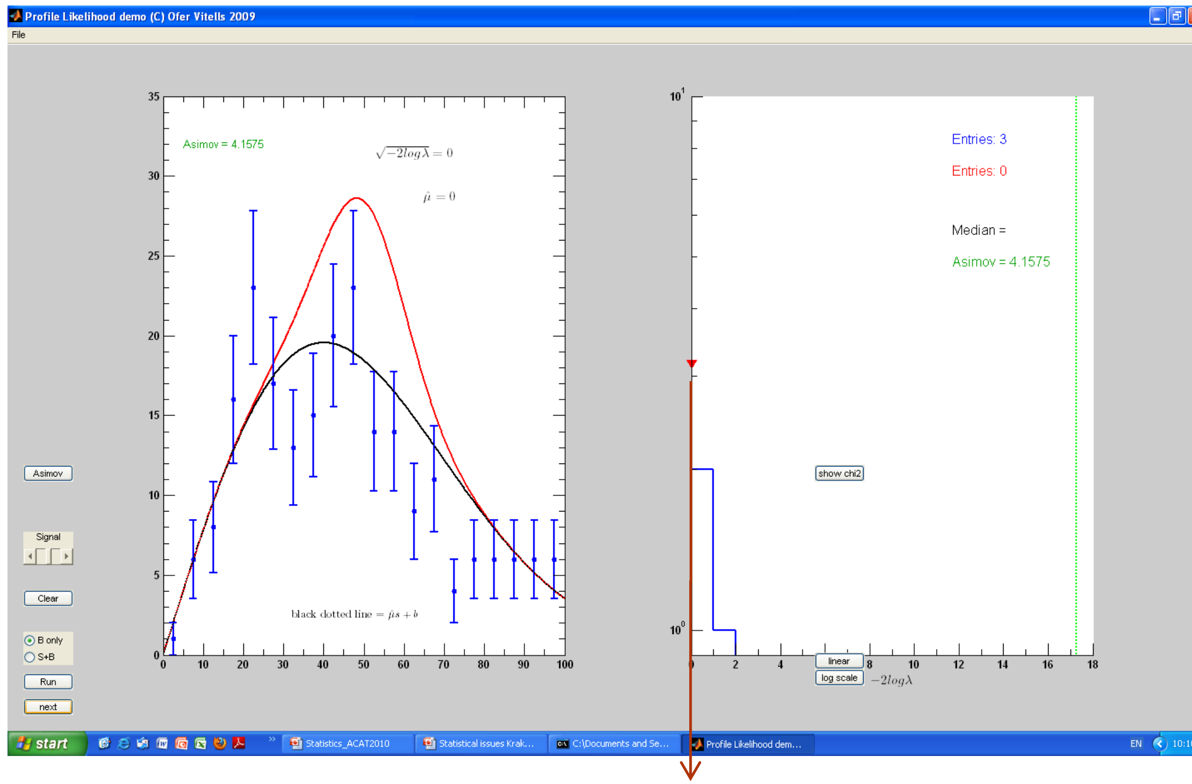
$$t = 0.43 \rightarrow Z = 0.66\sigma$$

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

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

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

$$t_0 = -2\ln \frac{L(b)}{L(\hat{\mu}s + b)}$$



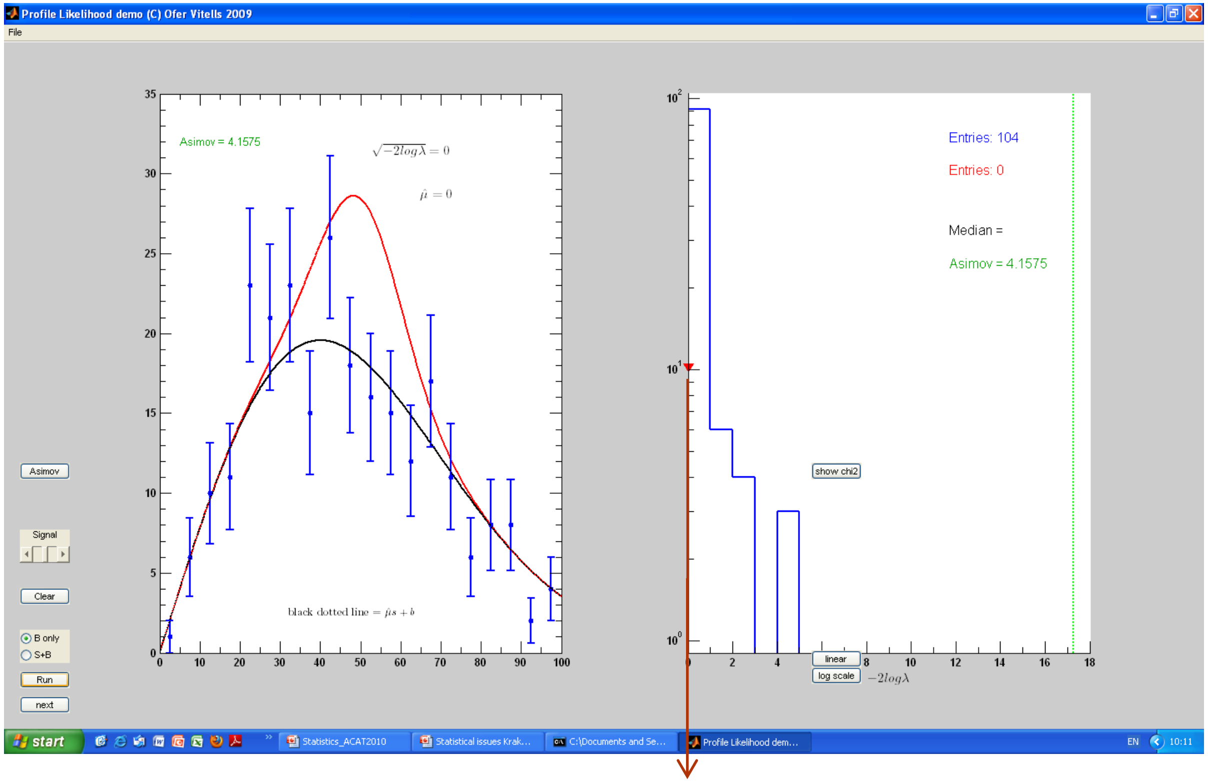
$$t = 0$$

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

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

$$t_0 = -2 \ln \frac{L(b)}{L(\hat{\mu} s + b)}$$

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



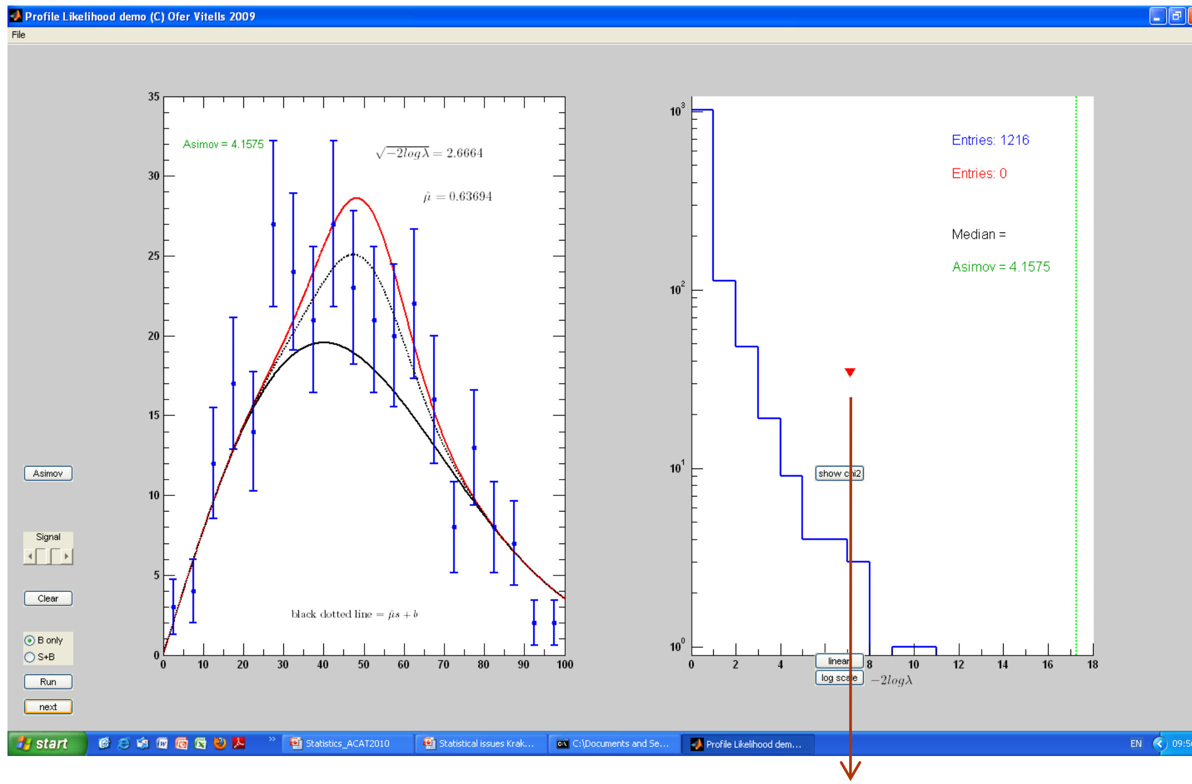
$$t = 0$$

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

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

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

$$t_0 = -2\ln \frac{L(b)}{L(\hat{\mu}s + b)}$$



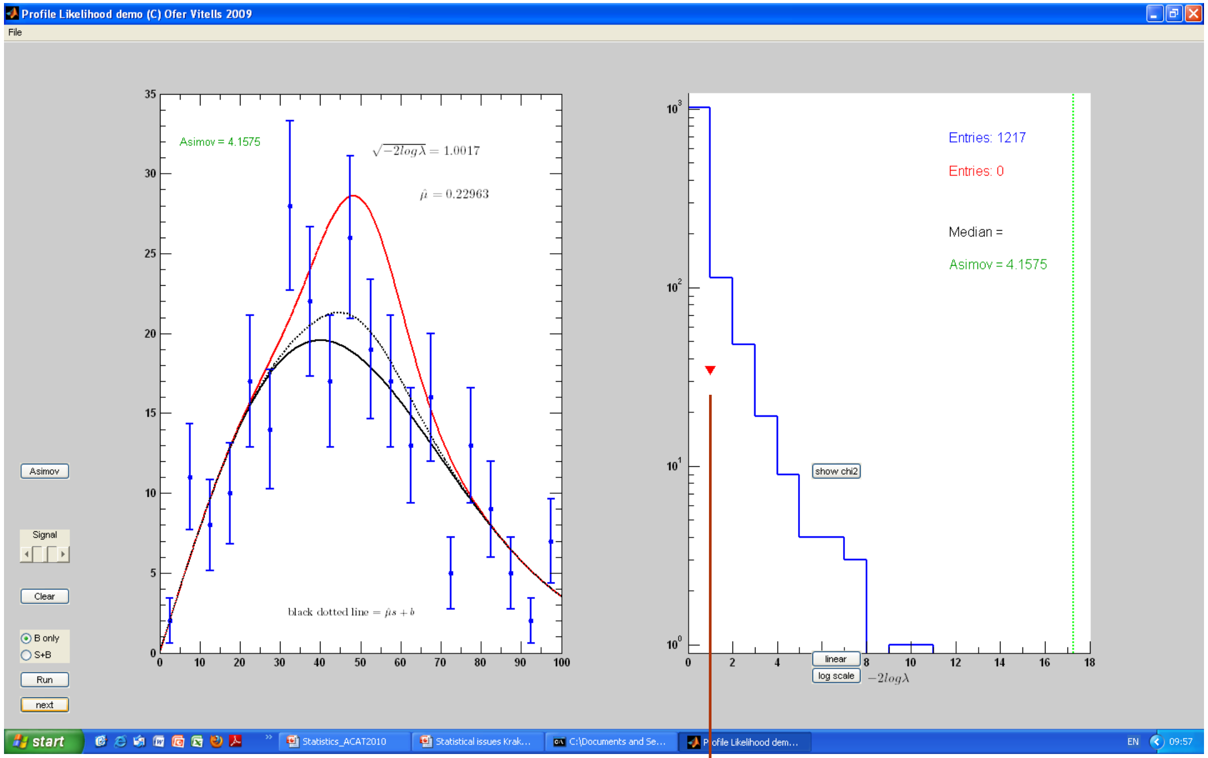
$$t = 6.76 \rightarrow Z = 2.6\sigma$$

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

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

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

$$t_0 = -2\ln \frac{L(b)}{L(\hat{\mu}s + b)}$$



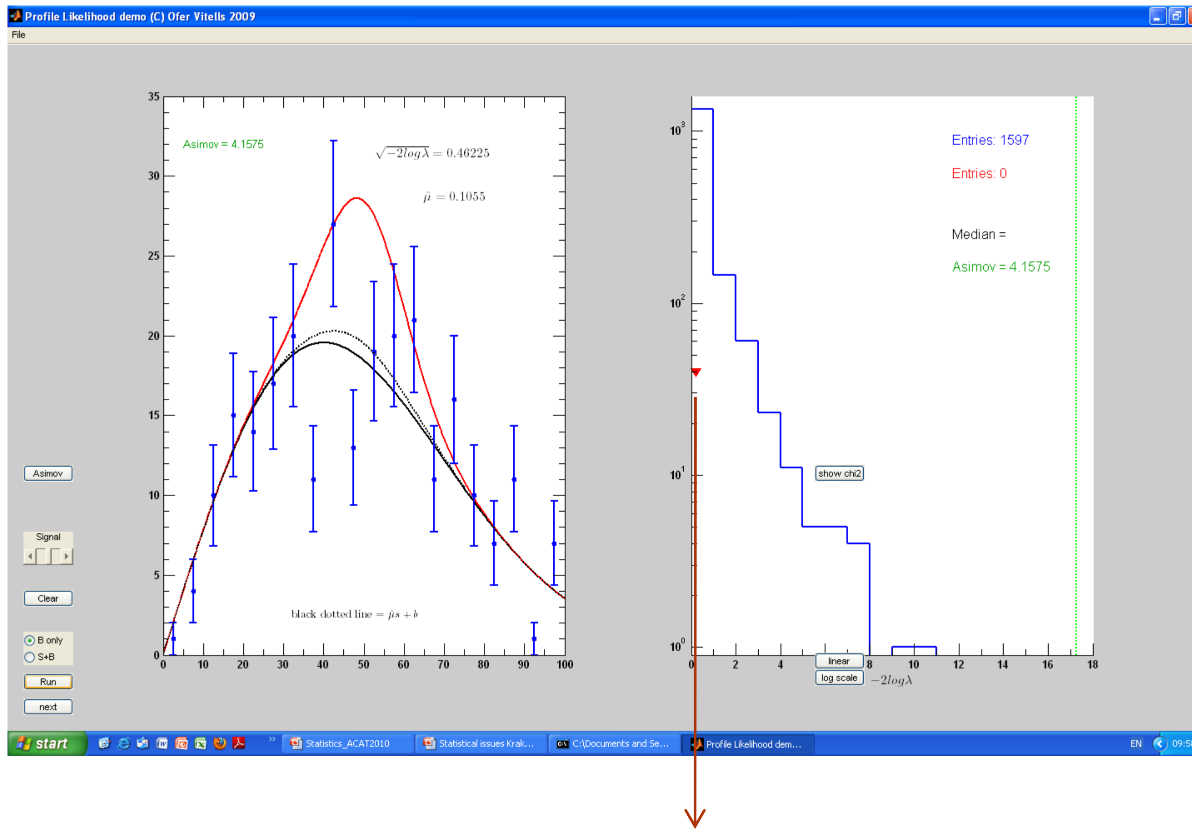
$$t = 1.2 \rightarrow Z = 1.1\sigma$$

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

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

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

$$t_0 = -2\ln \frac{L(b)}{L(\hat{\mu}s + b)}$$



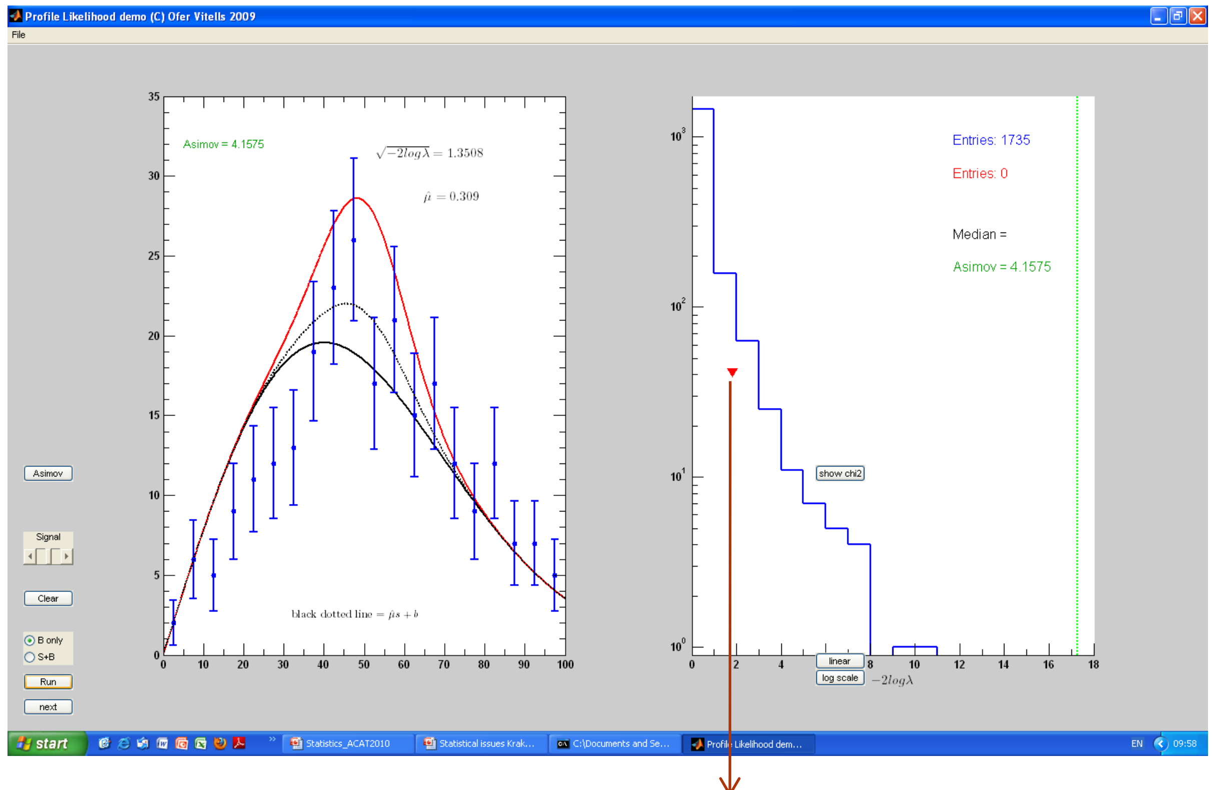
$$t = 0.16 \rightarrow Z = 0.4\sigma$$

PL: test t under BG only ; f(t$_0$ | H$_0$)

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

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

$$t_0 = -2\ln \frac{L(b)}{L(\hat{\mu}s + b)}$$



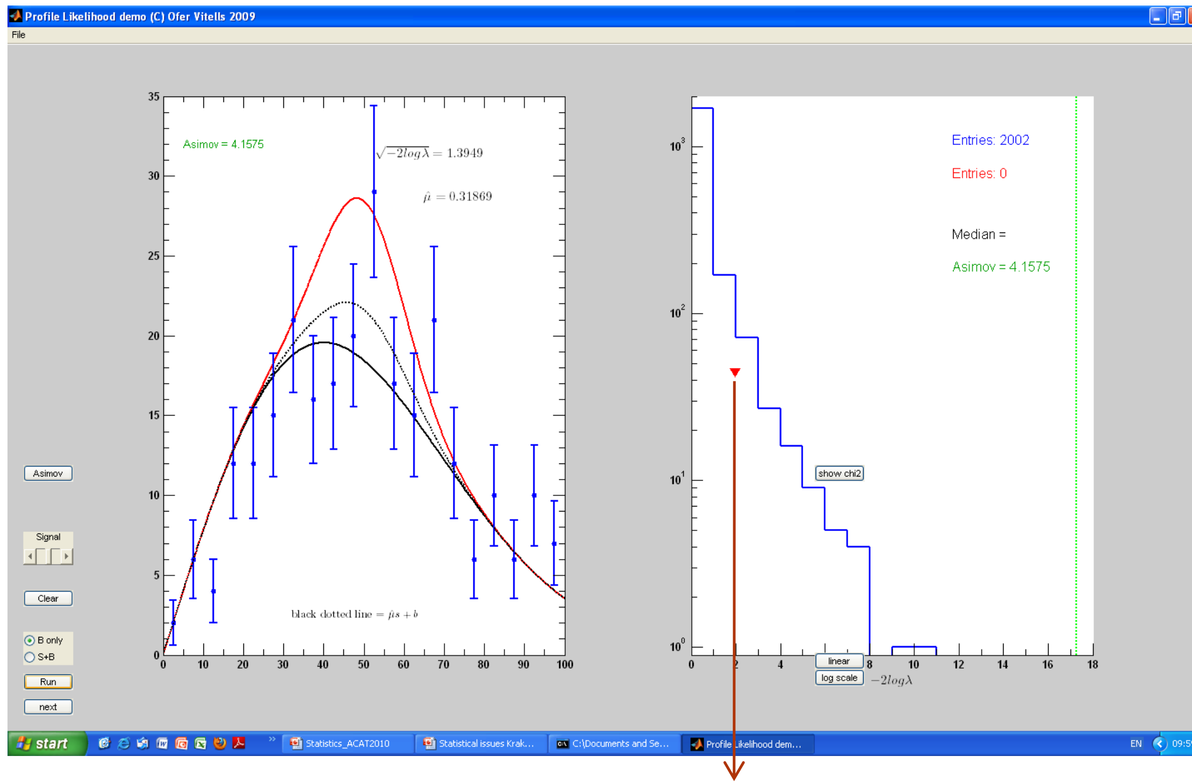
$$t = 1.8 \rightarrow Z = 1.35\sigma$$

PL: test t under BG only ; $f(t_0 \mid H_0)$

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

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

$$t_0 = -2\ln \frac{L(b)}{L(\hat{\mu}s + b)}$$



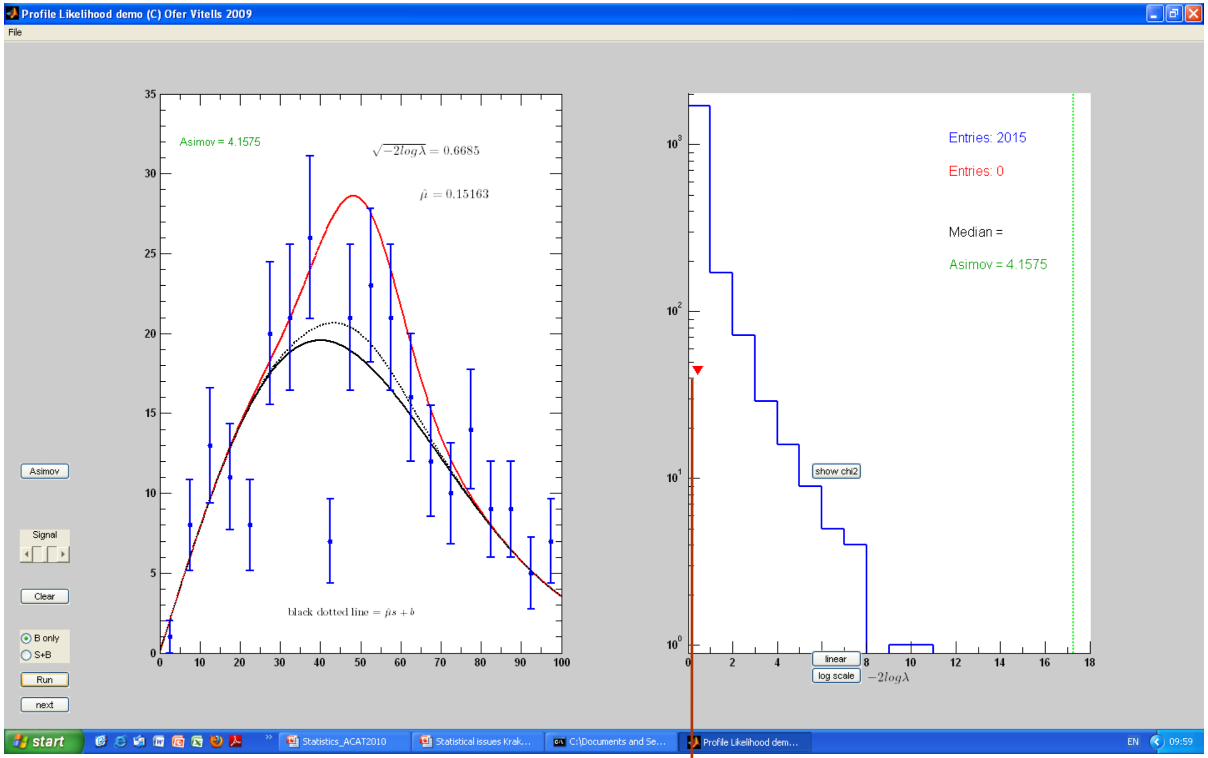
$$t = 1.9 \rightarrow Z = 1.39\sigma$$

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

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

$$t_0 = -2\ln \frac{L(b)}{L(\hat{\mu}s + b)}$$

next: Wilks theorem $\quad \hat{\mu} = 0.15 \rightarrow 0.66\sigma$



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

Wilks Theorem

S.S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Statist. **9** (1938) 60-2.

$$t_0 = -2\ln\frac{L(b)}{L(\hat{\mu}\cdot s+b)}$$

- Under a set of regularity conditions and for a sufficiently large data sample, *Wilks' theorem says that* the pdf of the statistic t under the null hypothesis approaches a chi-square PDF for one degree of freedom $f(t_0\,|\,H_0)=\chi_1^2$
- Same token $t_1 = -2\ln\frac{L(s+b)}{L(\hat{\mu}\cdot s+b)}$ $f(t_1\,|\,H_1)\sim\chi_1^2$

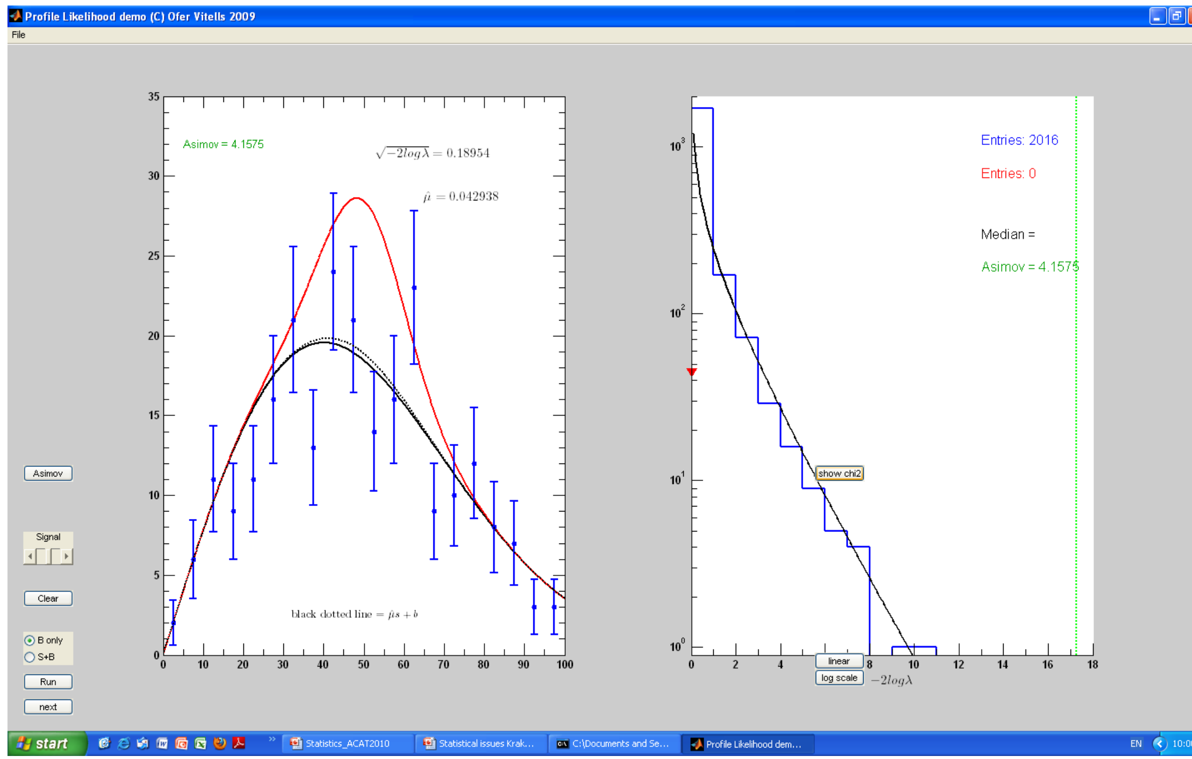
Wilks Theorem

- For the test statistic $t_\mu = -2\ln\frac{L(\mu)}{L(\hat{\mu})}; \quad t_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s+b)}$

$$t_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s+b)}$$

$$f(t_0 \mid H_0) = \chi_1^2$$

$$f(t_\mu \mid H_\mu) = \chi_1^2$$



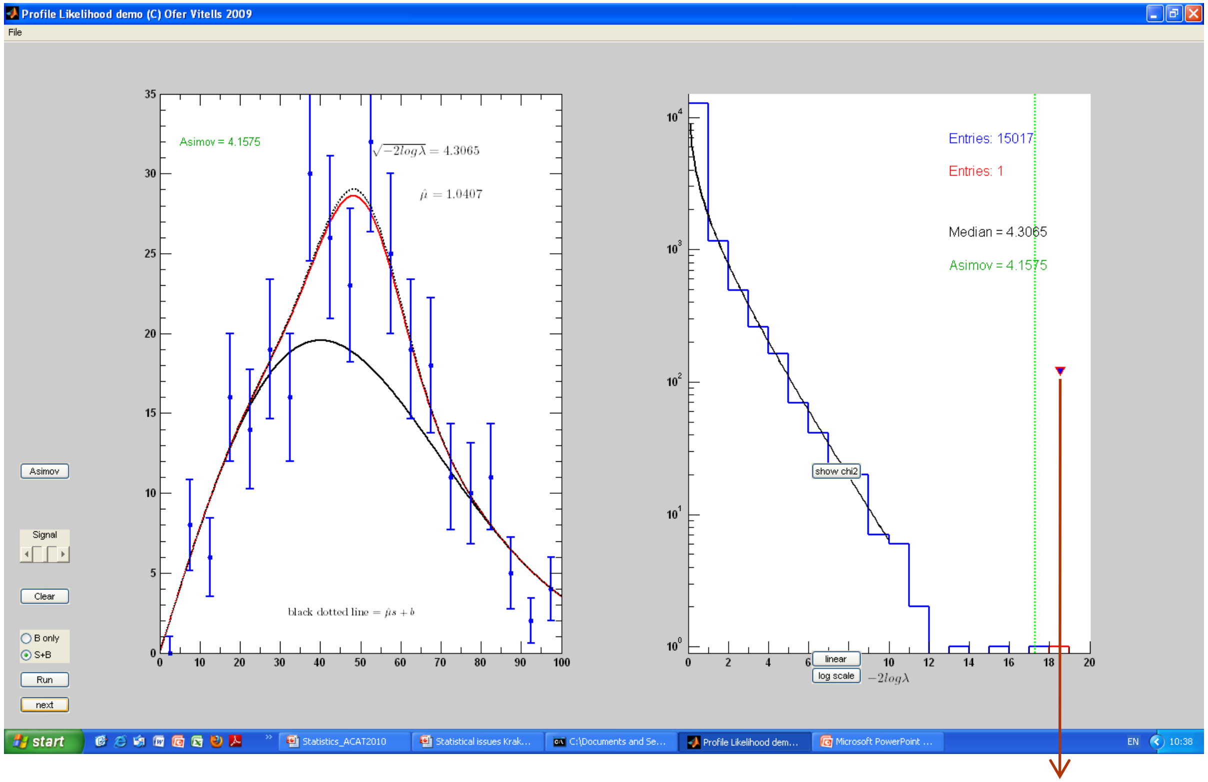
next: s+b experiments

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

$$t_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s+b)} = -2\ln\frac{L(b|H_1)}{L(\hat{\mu}s+b|H_1)}$$

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

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



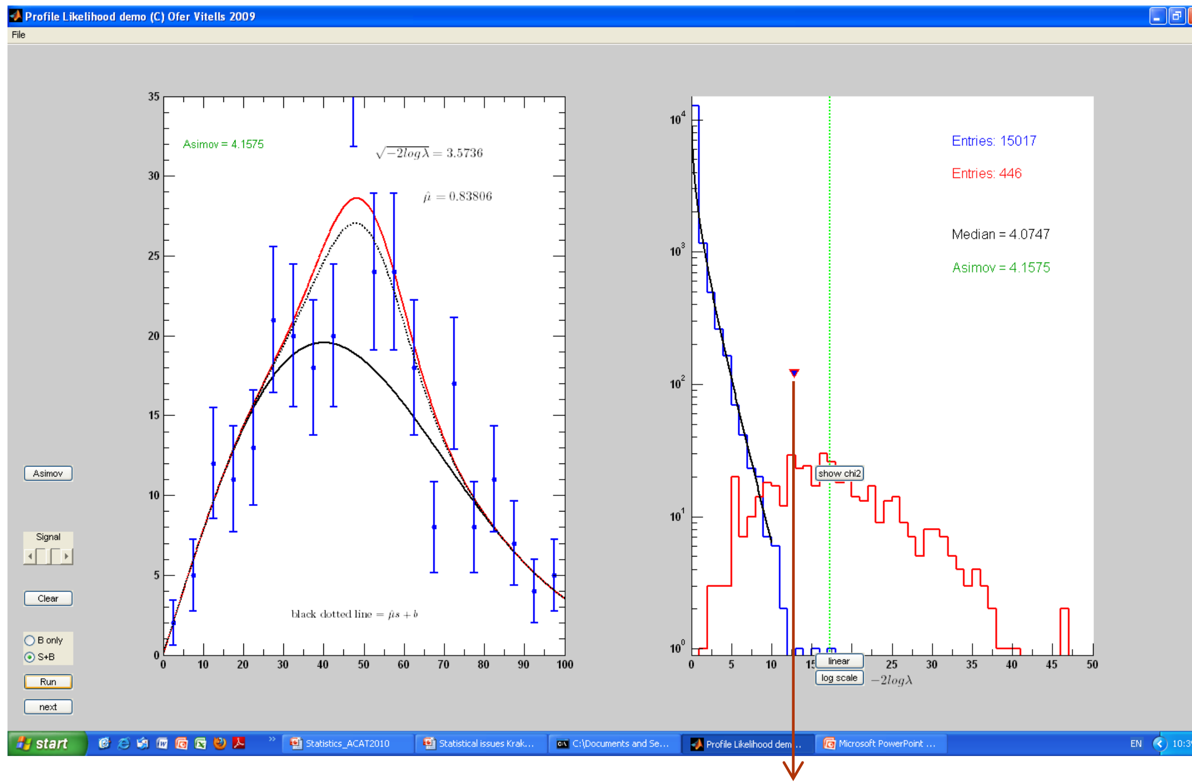
$$t = 18.5 \rightarrow Z = 4.3\sigma$$

PL: test t under s+b; $f(t_0 | H_1)$

$$t_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s+b)} = -2\ln\frac{L(b\,|\,H_1)}{L(\hat{\mu}s+b\,|\,H_1)}$$

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

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



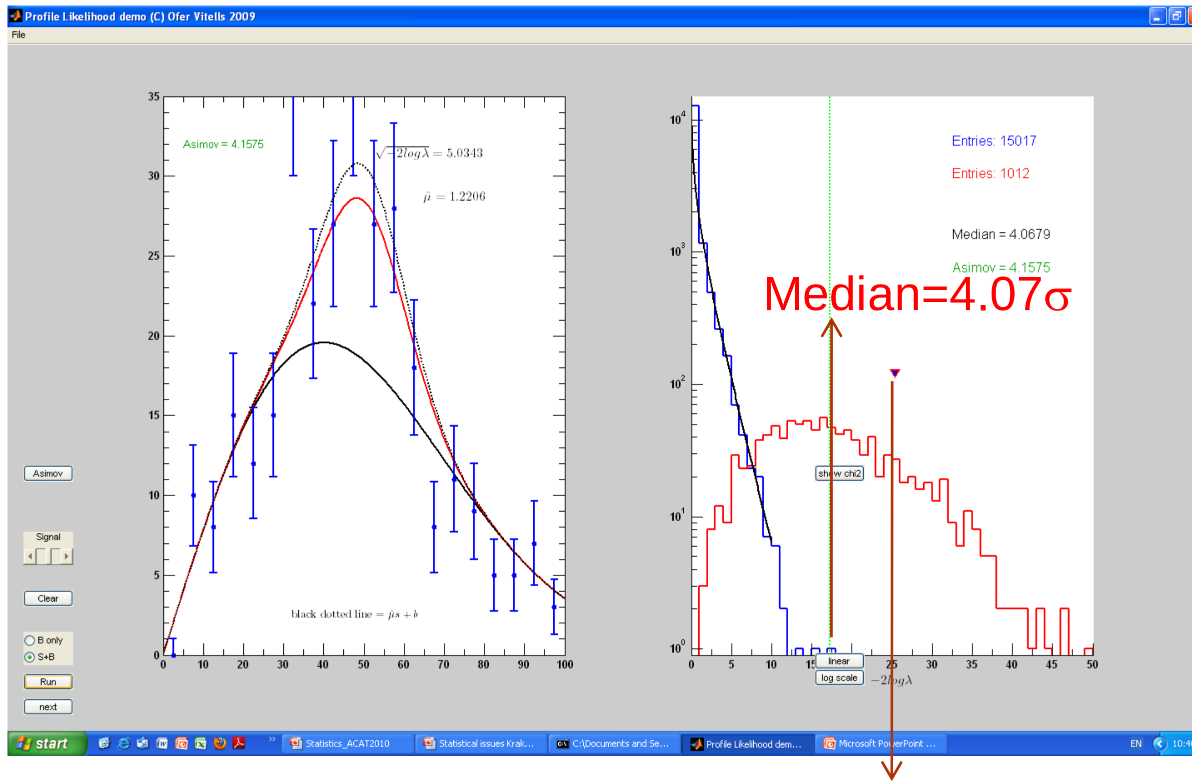
$$t = 12.9 \rightarrow Z = 3.6\sigma$$

Expected Discovery Sensitivity

$$t_0 = -2\ln\frac{L(b)}{L(\hat{\mu}s+b)} = -2\ln\frac{L(b\,|\,H_1)}{L(\hat{\mu}s+b\,|\,H_1)}$$

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

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

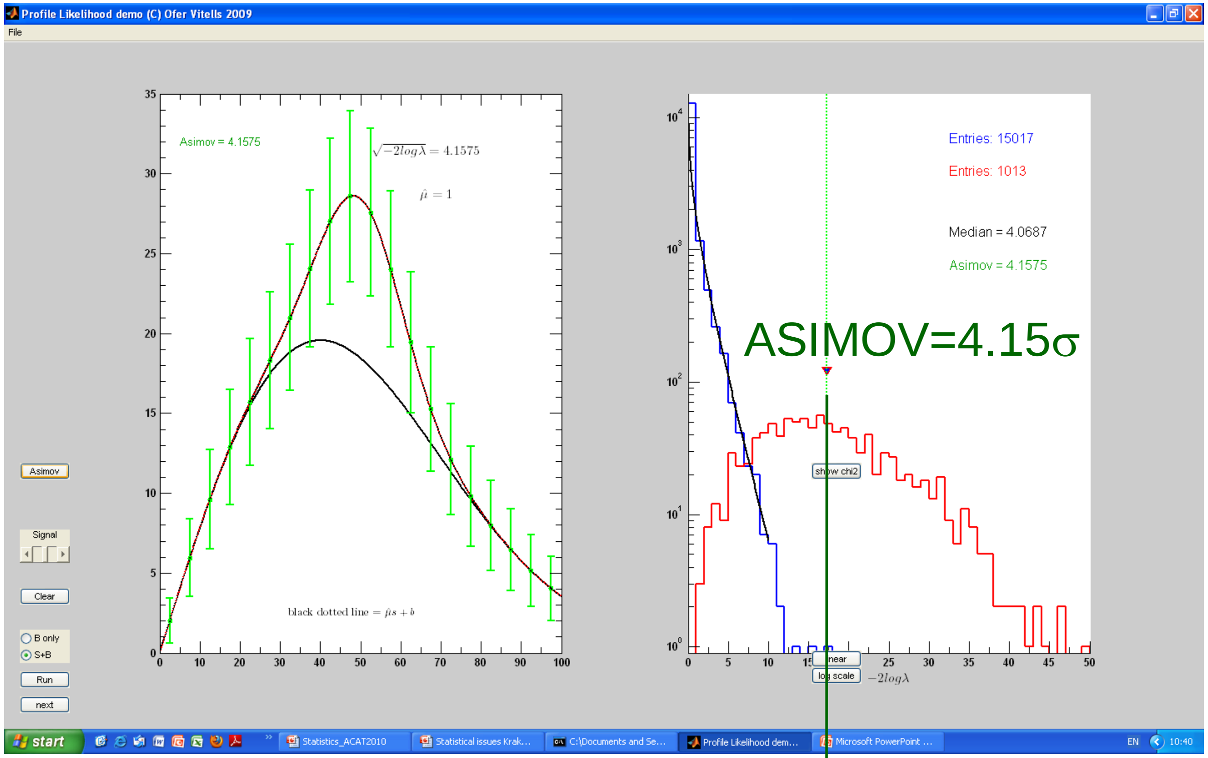


$$t = 25 \rightarrow Z = 5.0\sigma$$

The Median Sensitivity (via ASIMOV)

- To estimate the median sensitivity of an experiment (**before looking at the data**), one can either perform lots of s+b experiments and estimate the median $t_{o,med}$ or evaluate t_0 with respect to a representative data set, the ASIMOV data set with $\mu=1$, i.e. x=s+b

$$\langle \hat{\mu} \rangle = 1 \ under \ H_1 \qquad \hat{\mu} = 1.00 \rightarrow 4.15\sigma$$



$$t_A = 17.22 \rightarrow Z_A = 4.15$$

$$t_{o,med} \simeq t_0(\hat{\mu} = \mu_A = 1) = -2\ln \frac{L(b \mid x = x_A = s+b)}{L(\hat{\mu}s + b \mid x = x_A = s+b)} = -2\ln \frac{L(b)}{L(1 \cdot s + b)}$$

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** $$t = -2\ln\frac{L(H_0)}{L(H_1)}$$
- 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 $t = -2\ln\frac{L(H_0)}{L(H_1)}$
  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: $t = t(\hat{\mu})$

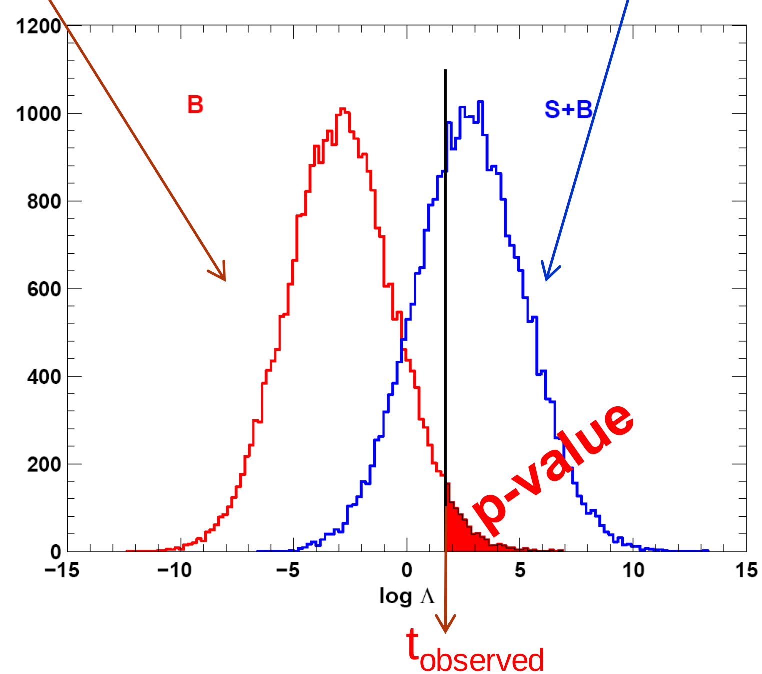
The Neyman Pearson Test Statistic

- Define the test statistic

$$t^{NP} = -2\ln\frac{L(H_0)}{L(H_1)} = -2\ln\frac{L(b)}{L(s+b)}$$

- Generate the PDF of T under the null hypothesis H_0, $f(t\,|\,H_0)$; under H_1, $f(t\,|\,H_0)$
- Let t_{obs} be the result of one experiment (Millions to Billions of collisions)
- Calculate the significance via the p-value under the null hypothesis(H_0)

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



The Neyman Pearson Test Statistic

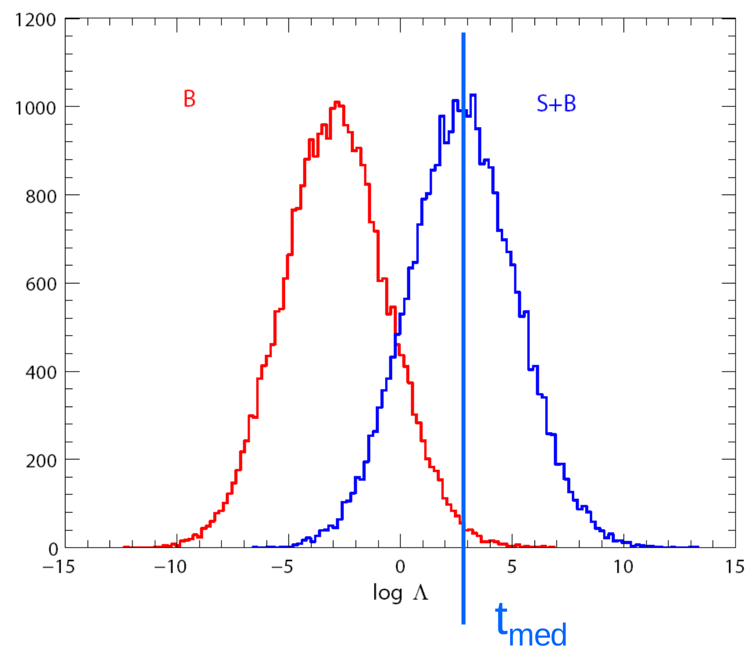
- Reminder, in the PL $t_0 = -2\ln\frac{L(H_0)}{L(best)} = -2\ln\frac{L(b)}{L(\hat{\mu}s+b)} = -2\ln\frac{L(\mu=0)}{L(\hat{\mu})}$

NP $t^{NP} = -2\ln\frac{L(H_0)}{L(H_1)} = -2\ln\frac{L(b)}{L(s+b)} = -2\ln\frac{L(\mu=0)}{L(\mu=1)}$

- The expected median discovery sensitivity (p-value) is

$$p_{med} = \int_{t_{med}}^{\infty} f(t\,|\,H_0)dt,$$

$$0.5 = \int_{t_{med}}^{\infty} f(t\,|\,H_1)dt$$

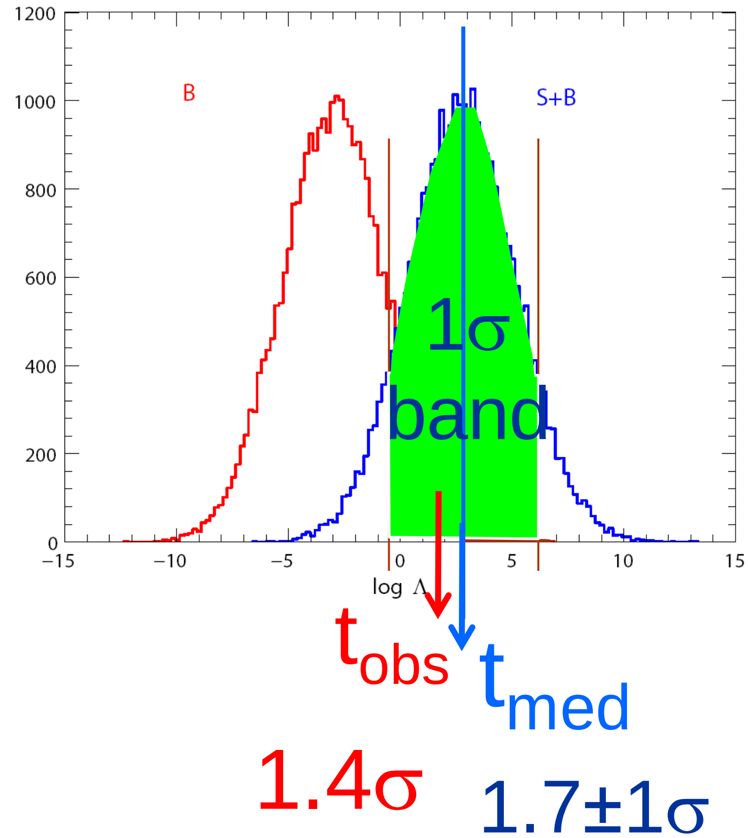


Median Sensitivity

$$p = \int_{t_{med}}^{\infty} f(t|H_0)dt, \quad 0.5 = \int_{t_{med}}^{\infty} f(t|H_1)dt$$

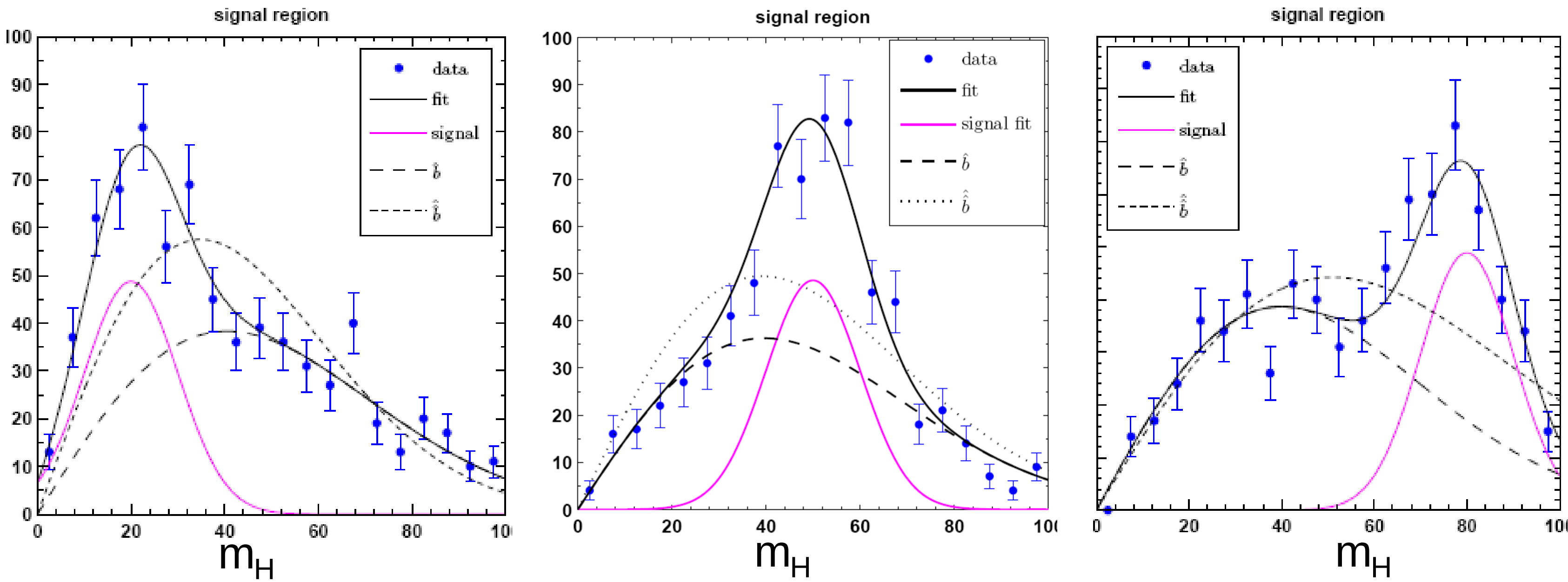
- In this example:
- t_{med}~1.7σ→no discovery sensitivity
- t_{obs} ~1.4σ indicates no discovery, a possible downwards fluctuation of the prospective s+b (within the green band)
- Such an observation is not high enough to reject the H_0 hypothesis (discovery) and not low enough to reject the H_1 hypothesis (and exclude it)

68% of s+b MC experiments give a test statistic , t, in the GREEN band



Let’s Play the Game (it’s a toy ....)

Note, in this example, the signal towards the edges of the background mass distribution (m_H=20,80) is better separated from the signal near the middle (m_H=50).



Nuisance Parameters

- Normally, the background, $b(\theta)$, has an uncertainty which has to be taken into account. In this case θ is called a nuisance parameter (which we associate with background systematics)
- The signal strength μ is a parameter if interest
- How can we take into account the nuisance parameters?

Nuisance Parameters (Systematisc)

- NP Likelihood Ratio:

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

- Either Integrate the Nuisance parameters

prior

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

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

- Or profile them

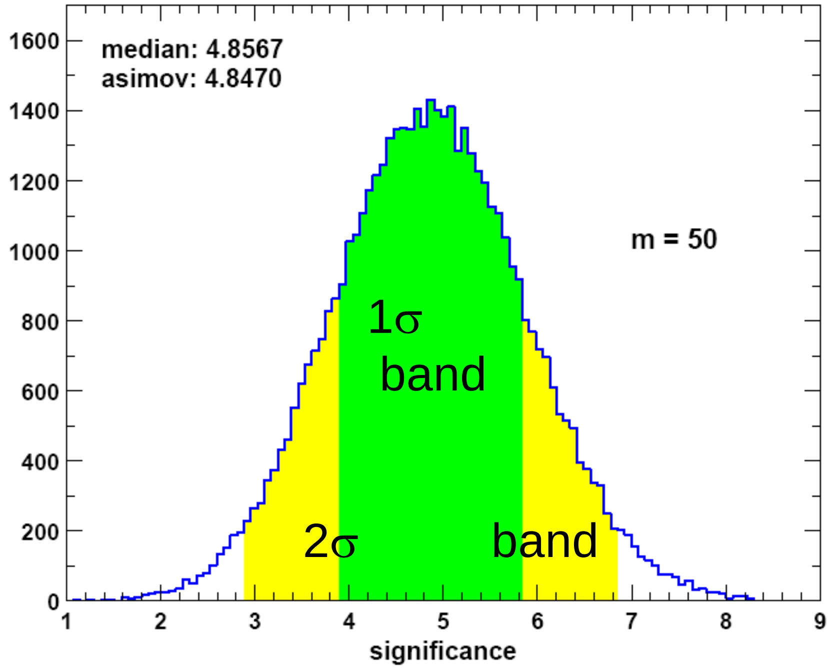
$$t^{NP} = -2\ln\frac{L\left(b(\hat{\hat{\theta}}_b)\right)}{L\left(s+b(\hat{\hat{\theta}}_{s+b})\right)}$$

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

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

The Profiled NP way

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

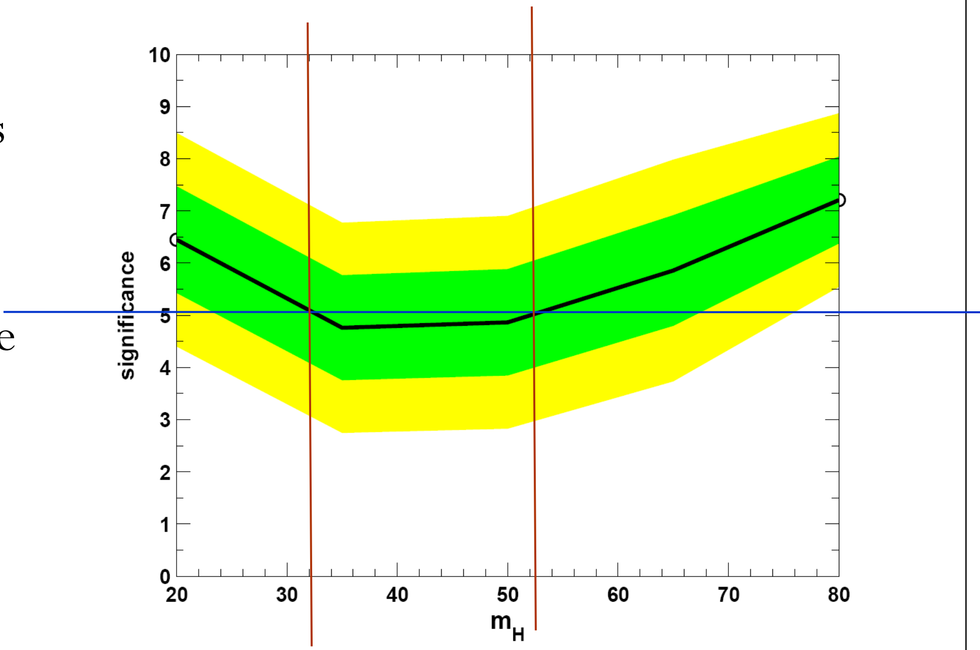


ATLAS, CERN – Open 2008-029
Cowan, Cranmer, E.G., Vitells, in preparation

The Profiled NP way

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

- In this example a Higgs with a mass $m_H<32$ or $m_H>52$ is expected to be discovered, i.e.
- if the Higgs exists in this mass range it will be discovered >50% of hypothetical LHC experiments



The Profile Likelihood vs NP LR

- NP Likelihood Ratio:

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

- PL Ratio: Test the null H_0 hypothesis

$$t_0^{PL} = -2\ln\frac{L(b)}{L(\hat{\mu}s+b)}$$

- Either Integrate the Nuisance parameters

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

- Or profile them

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

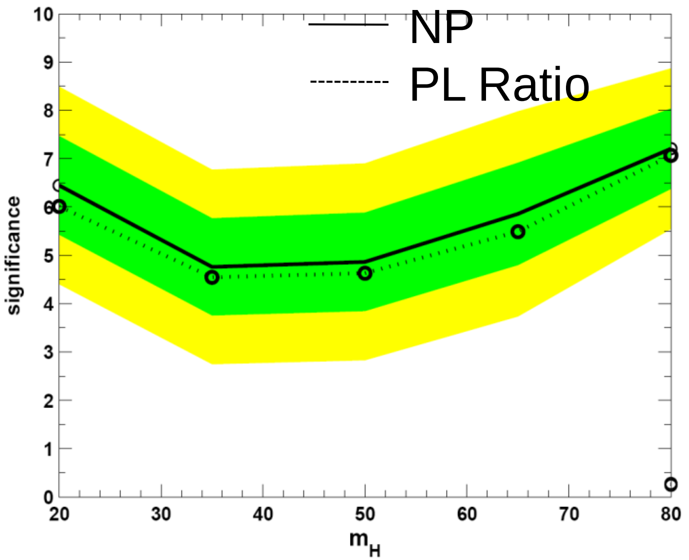
- Profile the Nuisance parameters

$$t_0^{PL} = -2\ln\frac{L\left(b(\hat{\hat{\theta}}_b)\right)}{L\left(\hat{\mu}s+b(\hat{\theta})\right)}$$

$\hat{\hat{\theta}}_b, \hat{\hat{\theta}}_{s+b} : MLE\ of\ \ L(b(\theta)), L(s+b(\theta));\quad \hat{\mu}, \hat{\theta} : MLE\ of\ \ L(\mu s+b(\theta))$

The frequentist NP vs PL methods

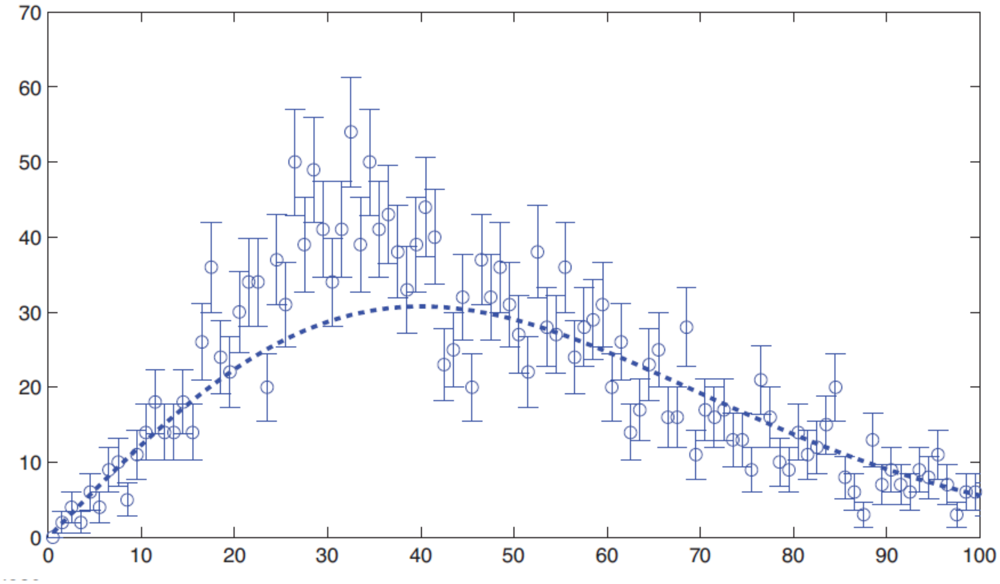
- Both methods have similar sensitivities
- The PL have the advantage that due to the Wilks theorem one can tell the significances without performing even one MC experiment



The Look Elsewhere Effect

Look Elsewhere Effect

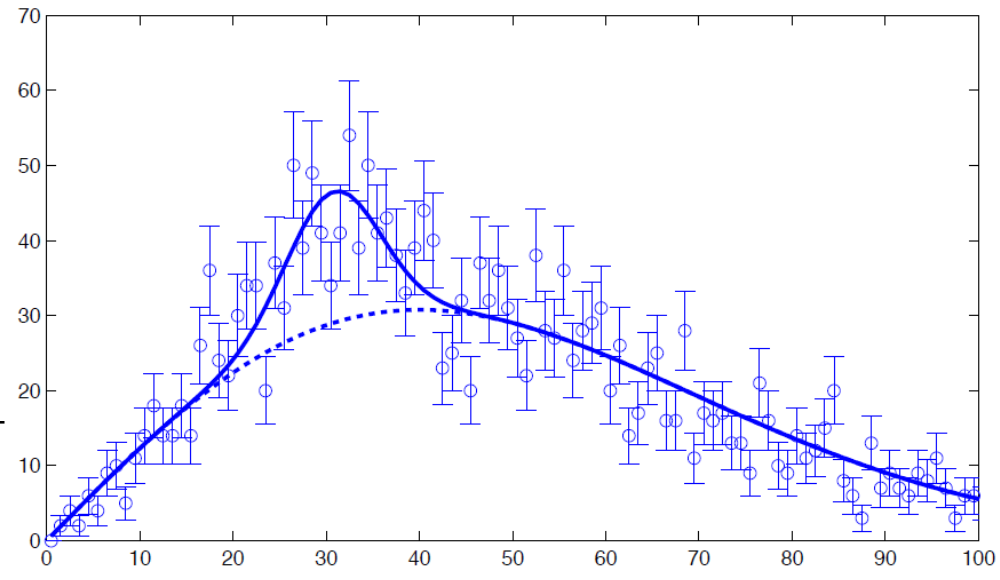
- Is there a signal here?



Look Elsewhere Effect

- Obviously @ m=30
- What is its significance?
- What is your test statistic?

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



Look Elsewhere Effect

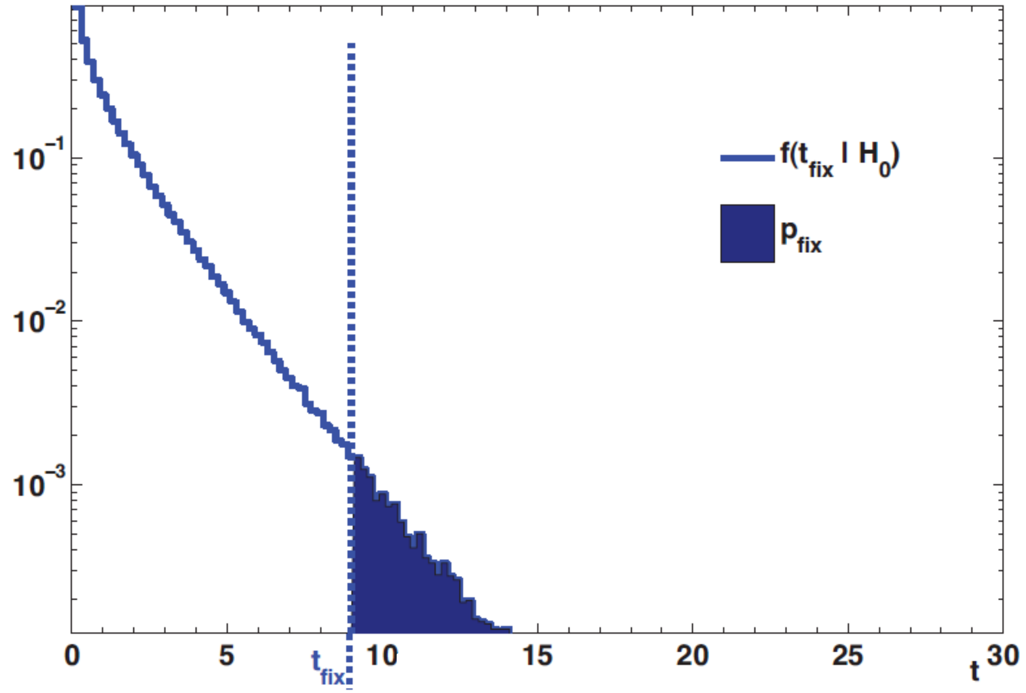
- Test statistic

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

- What is the p-value?
- generate the PDF

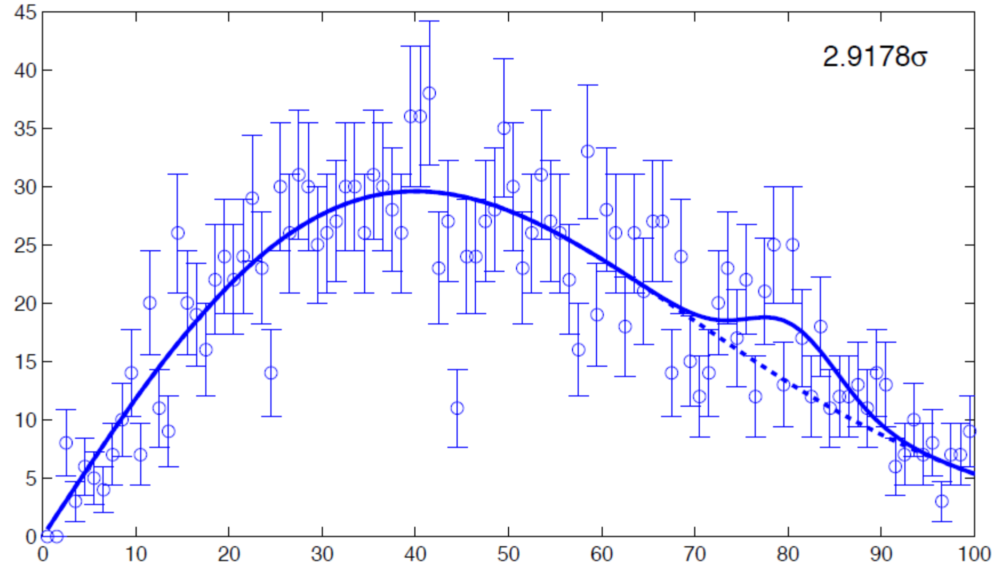
$$f(t_{fix} \mid H_0)$$

and find the **p-value**



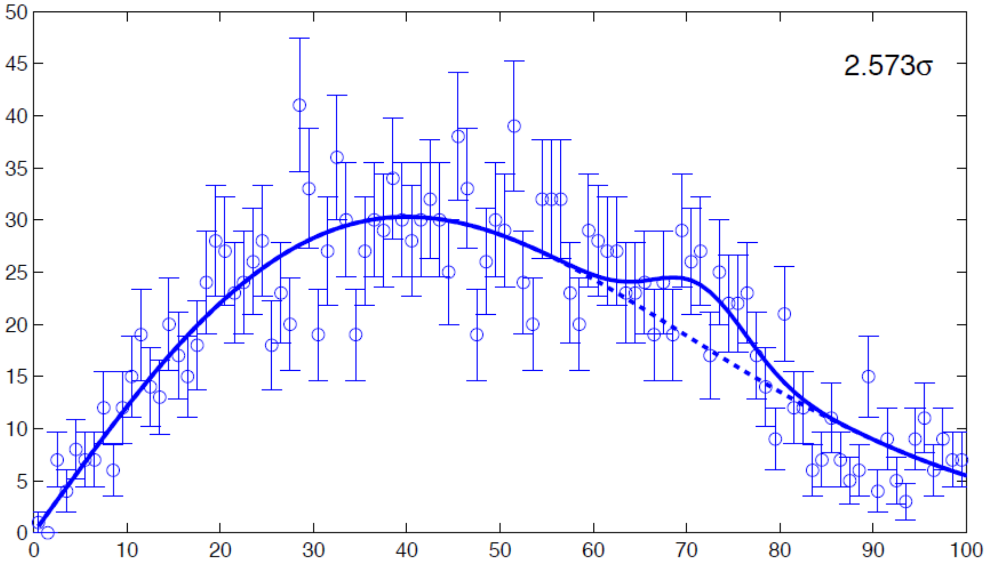
Look Elsewhere Effect

- Would you ignore this signal, had you seen it?



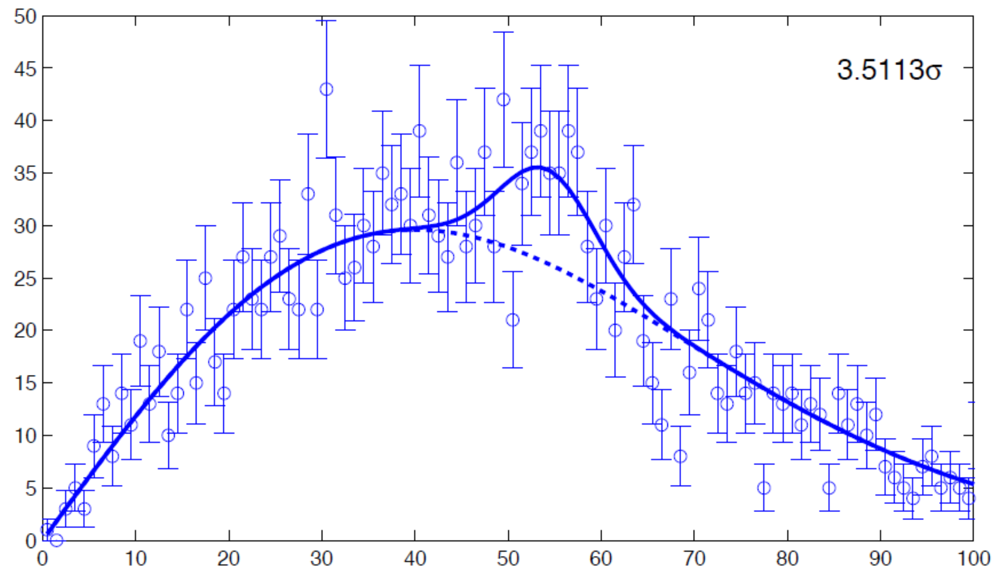
Look Elsewhere Effect

- Or this?



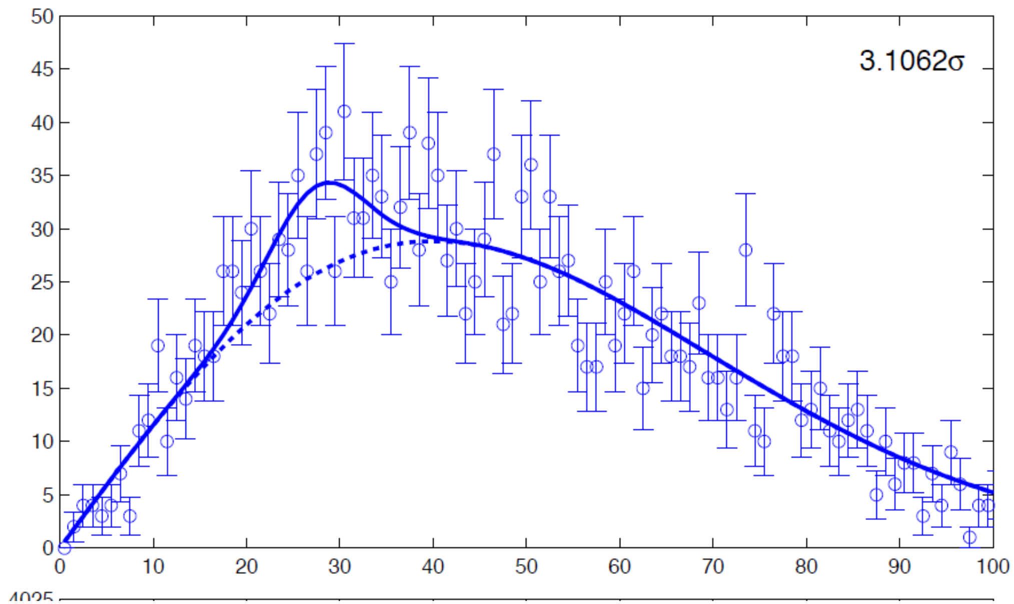
Look Elsewhere Effect

- Or this?



Look Elsewhere Effect

- Or this?
- Obviously NOT!

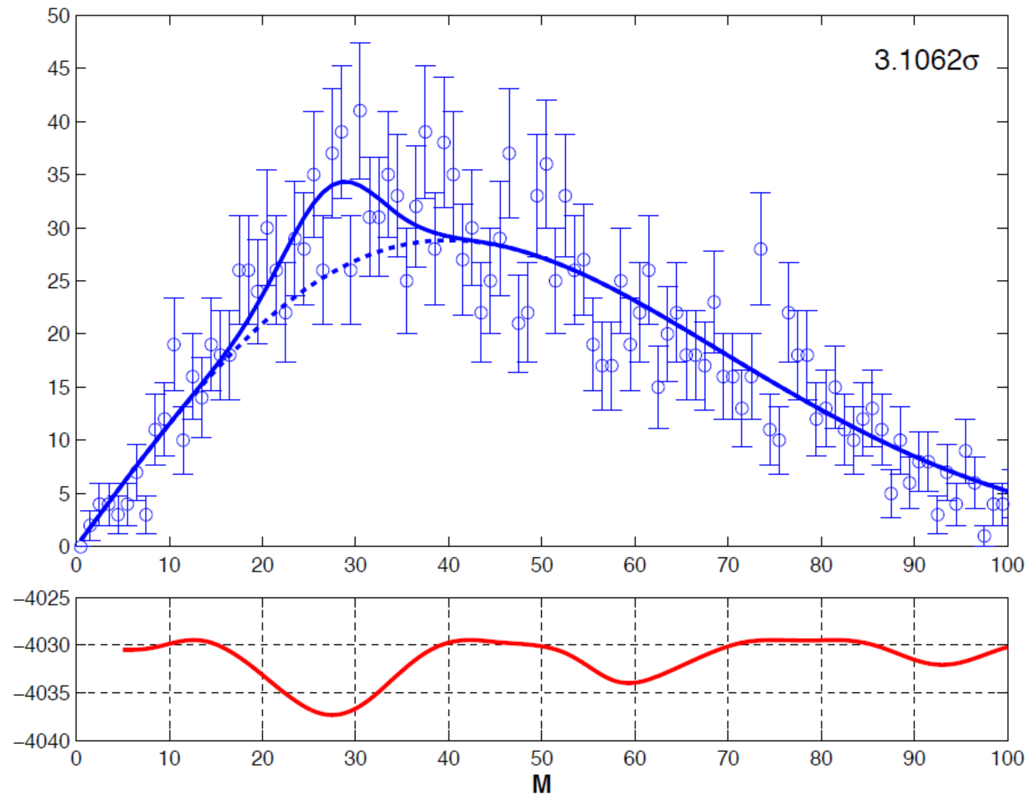


Look Elsewhere Effect

- Having no idea where the signal might be you would allow the signal to be anywhere in the **search range** and use a modified test statistic

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

- The p-value increases because more possibilities are opened



Look Elsewhere Effect

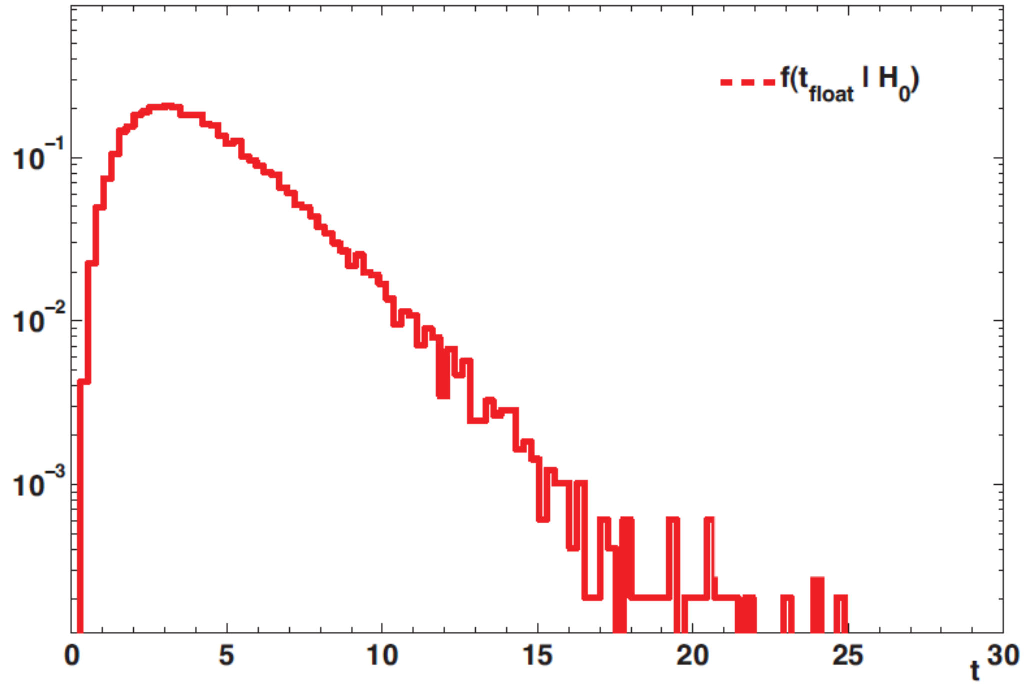
- the test statistis

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

- The null hypothesis PDF

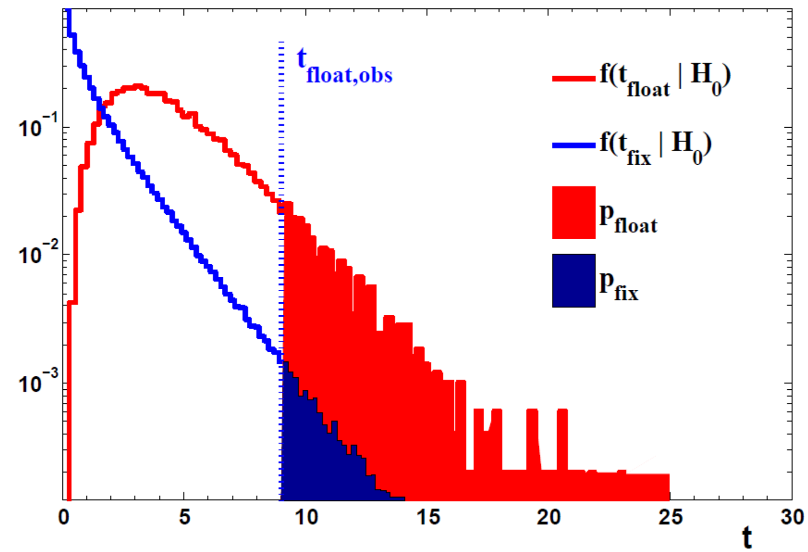
$$f(t_{float} \mid H_0)$$

does not follow a chi-squared with 2dof because there are multiple minima depending on the size of the search range



Look Elsewhere Effect

- We can now ask the question: Assume the Higgs is observed at some mass $\hat{m}$ what is the probability for the background to fluctuate locally $@ m_H = \hat{m}$ at the observed level (or more)



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

- We can calculate the following p-value

$$p_{fix} = \int_{t_{obs}} f(t_{fix} | H_0) dt_{fix} < p_{float} = \int_{t_{obs}} f(t_{float} | H_0) dt_{float}$$

$$trial\,\# = \frac{\int_{t_{obs}} f(t_{float} | H_0) dt_{float}}{\int_{t_{obs}} f(t_{fix} | H_0) dt_{fix}} = \frac{p_{float}}{p_{fix}}$$

Look Elsewhere Effect

- We find a thumb rule:

$$trial\,\# = \frac{\int_{t_{obs}} f(t_{float} \mid H_0)\,dt_{float}}{\int_{t_{obs}} f(t_{fix} \mid H_0)\,dt_{fix}} = \frac{p_{float}}{p_{fix}}$$

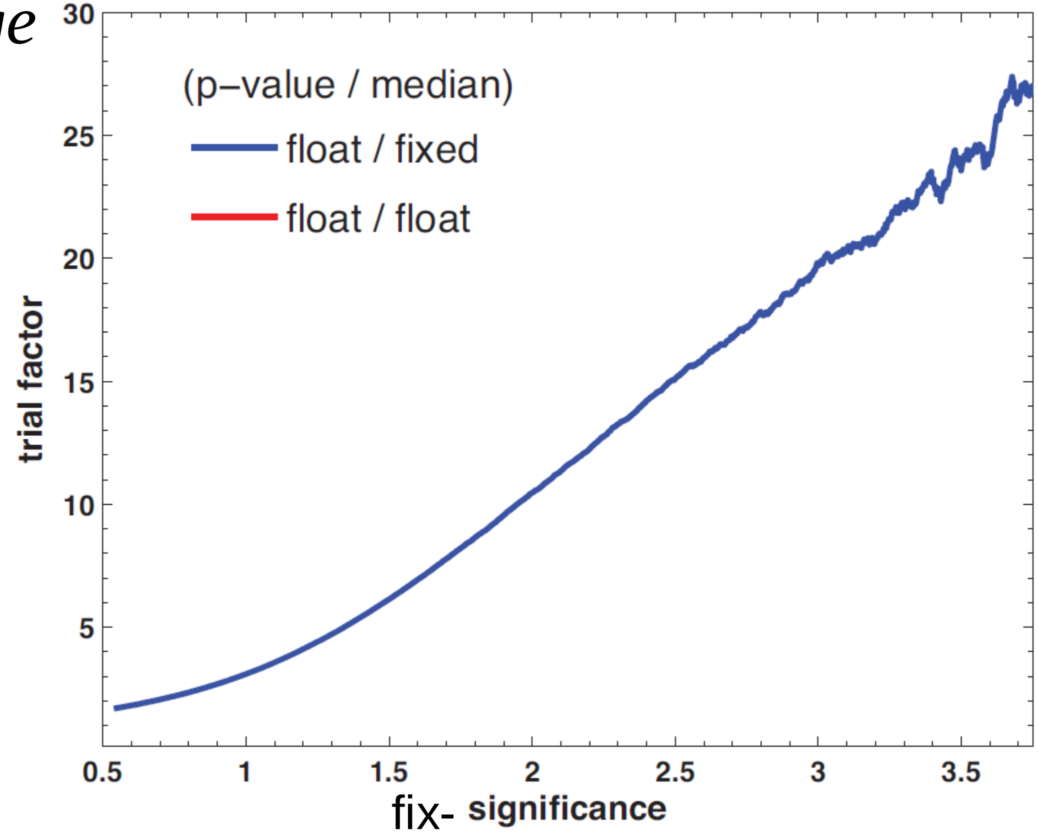
$$\Delta_m = mass-search-range$$

$$\sigma_m = mass-resolution$$

$$Z_{fix} = significance$$

$$trial\,\# \approx \frac{\Delta_m}{3\sigma_m} Z_{fix}$$

E. Gross and O. Vitells



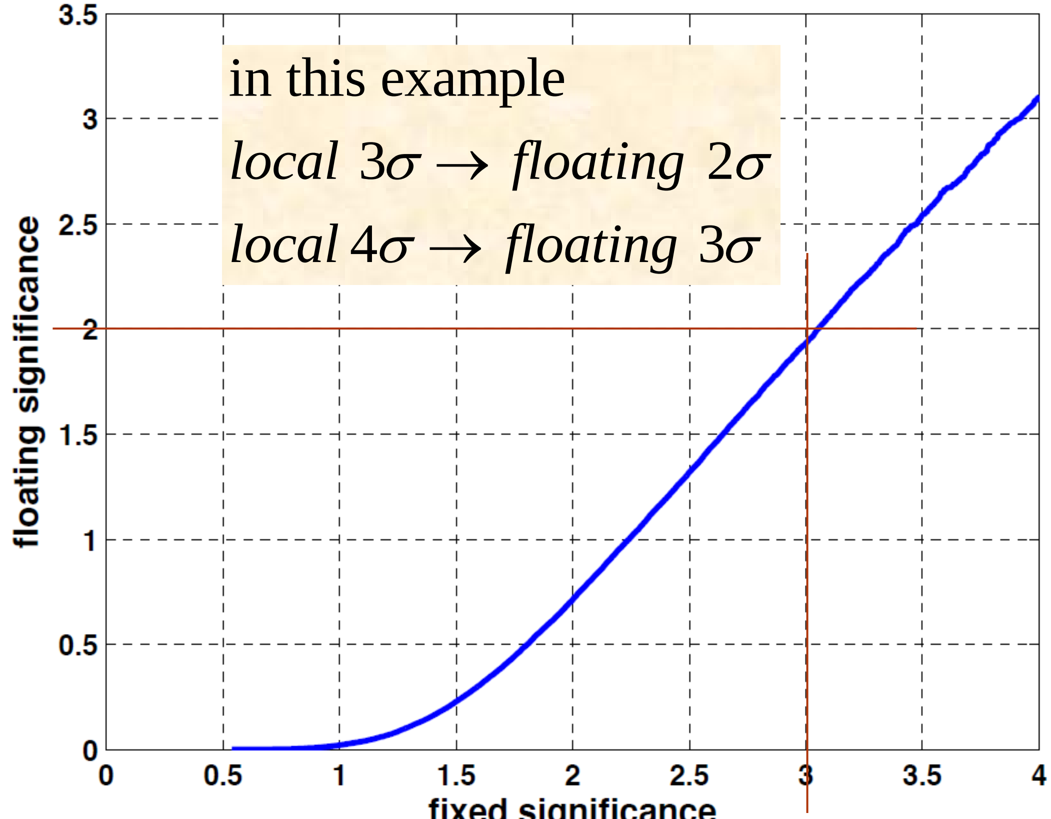
Look Elsewhere Effect

- Conclusion: The Look Elsewhere Effect reduces the apparent significance

- It addresses the alternate hypothesis:

  A Higgs at some mass in the search-range

$$p_{float} = \int_{t_{float}} f(t_{float} \mid H_0)\, dt_{float}$$



$$p_{fix} = \int_{t_{fix}} f(t_{fix} \mid H_0)\, dt_{fix}$$

Discovery Bayes Factors

A new entry in the PDG

The Bayes Way

- Derive the posterior probability of the hypothesis H_1 based on Bayes theorem.

$$P(H_1 | x) = \frac{P(x | H_1)\pi(H_1)}{\pi(x)}$$

- To claim a strong evidence of H_1 over H_0 (a discovery) define the Bayes factor B_{10} as the ratio of the posterior to prior odds

$$B_{10} = \frac{P(H_1 | x) / \pi(H_1)}{P(H_0 | x) / \pi(H_0)} = \frac{L(H_1)}{L(H_0)}$$

Frequentist ~ Bayesian ?

- Using the saddle point approximation we get the relationship between the Bayes factors and the frequentist median sensitivities

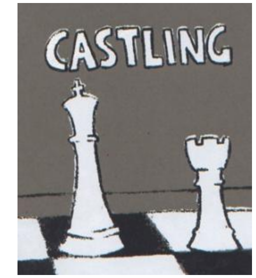
$$\ln B_{10} = \frac{1}{2} z_{med}^2$$

E. Gross and O. Vitells

Z	~B_{10}	
1	1.6	No evidence
2	7.3	Weak evidence
3	90	Evidence
5	26800	Discovery

EXCLUSION

Castling the Hypotheses



NULL

H_0- SM w/o Higgs

ALTERNATE

H_1- SM with Higgs

- Reject H_0 in favor of H_1 – A DISCOVERY

$$t^{NP} = \frac{L(H_0)}{L(H_1)} = -2\ln\frac{L\left(b(\hat{\hat{\theta}}_b)\right)}{L\left(s+b(\hat{\hat{\theta}}_{s+b})\right)}$$

$$t_0^{PL} = -2\ln\frac{L\left(b(\hat{\hat{\theta}}_b)\right)}{L\left(\hat{\mu}s+b(\hat{\theta})\right)}$$

$$\hat{\hat{\theta}}_b, \hat{\hat{\theta}}_{s+b} : MLE\ of\ \ L(b(\theta)), L(s+b(\theta));\quad \hat{\mu},\hat{\theta} : MLE\ of\ \ L(\mu s+b(\theta))$$

Castling the Hypotheses

NULL

H_1- SM with Higgs

ALTERNATE

H_0- SM w/o Higgs

- Reject H_1 in favor of H_0 – Excluding H_1

$$t^{NP}=\frac{L(H_0)}{L(H_1)}=-2\ln\frac{L\left(b(\hat{\hat{\theta}}_b)\right)}{L\left(s+b(\hat{\hat{\theta}}_{s+b})\right)}$$

$$t_0^{PL}=-2\ln\frac{L\left(b(\hat{\hat{\theta}}_b)\right)}{L\left(\hat{\mu}s+b(\hat{\theta})\right)}\quad\Rightarrow\quad t_1^{PL}=-2\ln\frac{L\left(s+b(\hat{\hat{\theta}}_{s+b})\right)}{L\left(\hat{\mu}s+b(\hat{\theta})\right)}$$

$$\hat{\hat{\theta}}_b,\hat{\hat{\theta}}_{s+b}: MLE\ of\ \ L(b(\theta)),L(s+b(\theta));\quad \hat{\mu},\hat{\theta}: MLE\ of\ \ L(\mu s+b(\theta))$$

Exclusion

- Test the H_μ hypothesis, <n>=$\mu s(m_H)+b$
- $\mu = \frac{\sigma}{\sigma_{SM}}$ is the signal strength which is a parameter of interest.

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

$$t_\mu^{PL} = -2\ln \frac{L\left(\mu s + b(\hat{\hat{\theta}}_{s+b})\right)}{L\left(\hat{\mu} s + b(\hat{\theta})\right)}$$

Exclusion

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

- Test the H_μ hypothesis, $<n>=\mu s(m_H)+b$
- $\mu = \frac{\sigma}{\sigma_{SM}}$ is the signal strength which is a parameter of interest.

$$t_\mu^{PL} = -2\ln \frac{L\left(\mu s + b(\hat{\hat{\theta}}_{s+b})\right)}{L\left(\hat{\mu} s + b(\hat{\theta})\right)}$$

- By testing the signal hypothesis (H_μ) we can construct a *95%* confidence (frequentist) or credibility (Bayesian) interval CI: $[0, \mu_{95}]$ (CI: Confidence or Credibility Interval for $\mu = \frac{\sigma}{\sigma_{SM}}$)
- If $\mu_{95}<1$ the SM Higgs (H_1) is excluded at the *95%* CL.

A SUSY Higgs (with a smaller signal strength) can still be hidden there…

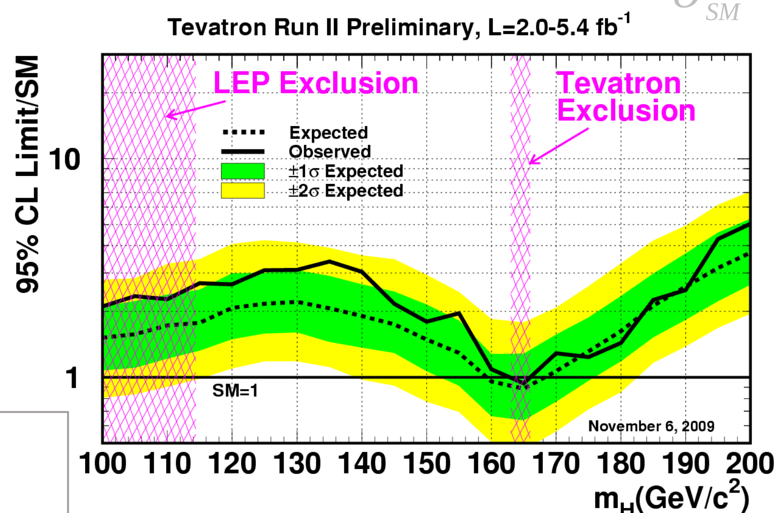
Exclusion

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

$$t_\mu^{PL} = -2\ln \frac{L\left(\mu s + b(\hat{\hat{\theta}}_{s+b})\right)}{L\left(\hat{\mu} s + b(\hat{\theta})\right)}$$

- Test the H_μ hypothesis, $\langle n\rangle = \mu s(m_H)+b$
- $\mu = \frac{\sigma}{\sigma_{SM}}$ is the signal strength which is a parameter of interest.
- By testing the signal hypothesis (H_μ) we can construct a 95% confidence (frequentist) or credibility (Bayesian) interval CI: $[0, \mu_{95}]$ (CI: Confidence or Credibility Interval for $\mu = \frac{\sigma}{\sigma_{SM}}$)
- If $\mu_{95} < 1$ the SM Higgs (H_1) is excluded at the 95% CL.

A SUSY Higgs (with a smaller signal strength) can still be hidden there…



Combined CDF and D0 Upper Limits on Standard Model Higgs-Boson Production with 2.1 - 5.4 fb^{-1} of Data
NOV 2009

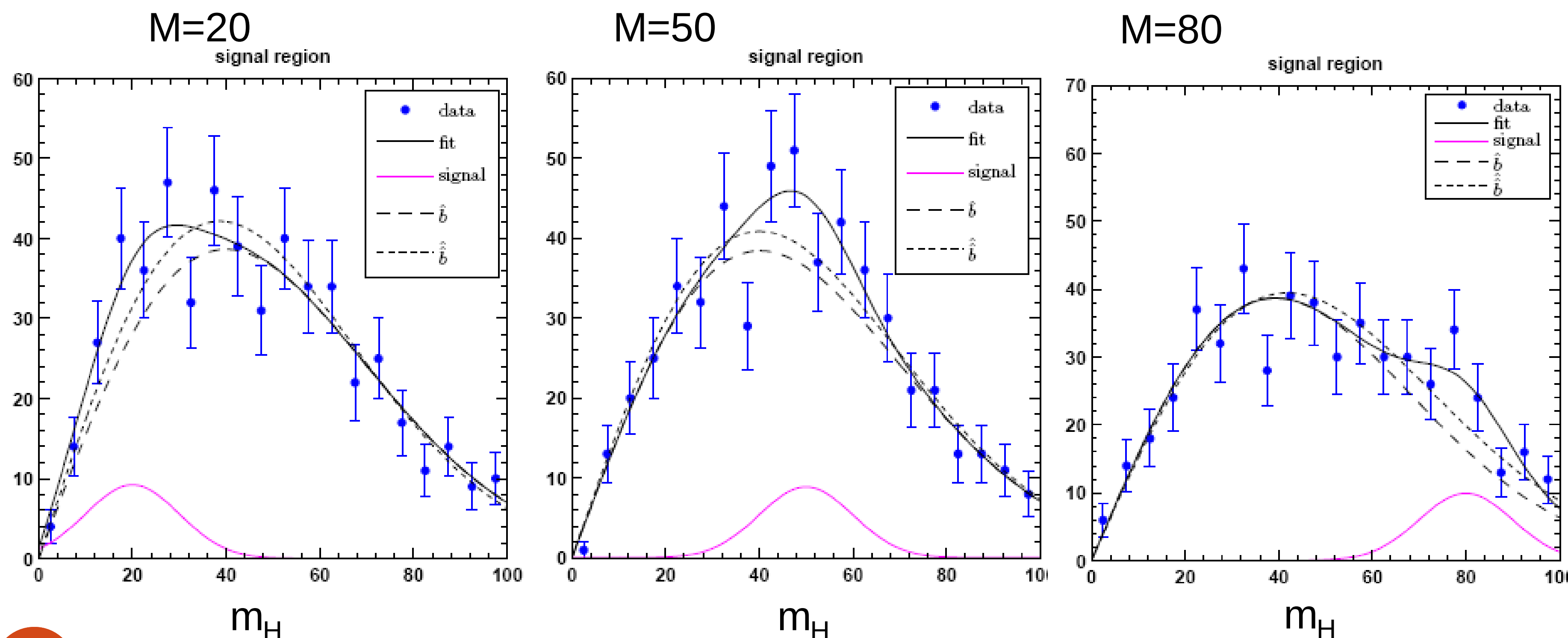
The Equivalence of CL and p-value

- Test the H_μ $(\mu s(m_H)+b)$ hypothesis
- Find the p-value under H_μ $p_\mu(m_H)=\int_{t_{\mu,obs}}^{\infty} f(t_\mu \mid H_\mu)dt_\mu$
- If $p_\mu(m_H)<5\%$ the H_μ hypothesis is rejected
- Find $\mu 95(m_H)$ such that $p_{\mu 95}(m_H)=5\%$
- 95% of the intervals $[0, \mu 95(m_H)]$ could contain a signal with a strength $\mu(m_H)<\mu 95(m_H)$ *(if existed)*
- $\mu 95(m_H)$ is an upper bound on $\mu(m_H)$ *@ 95% CL*
- If $\mu 95(m_H)<1$, a SM Higgs with a mass m_H

  is excluded at >95% CL→ $p_\mu \approx 1 - CL$

Exclusion Case Study

- Strong expected signals are very easy to exclude if your data is BG-only compatible;
- Weak expected signals are more difficult to exclude, unless the background has a strong downward fluctuation
- This leads to a controversy since it allows to exclude extremely small signals for which the experiment might not be sensitive at all



Profile Likelihood Ratio

Test the s(m_H)+b hypothesis
i.e. test the μ=1 hypotheis

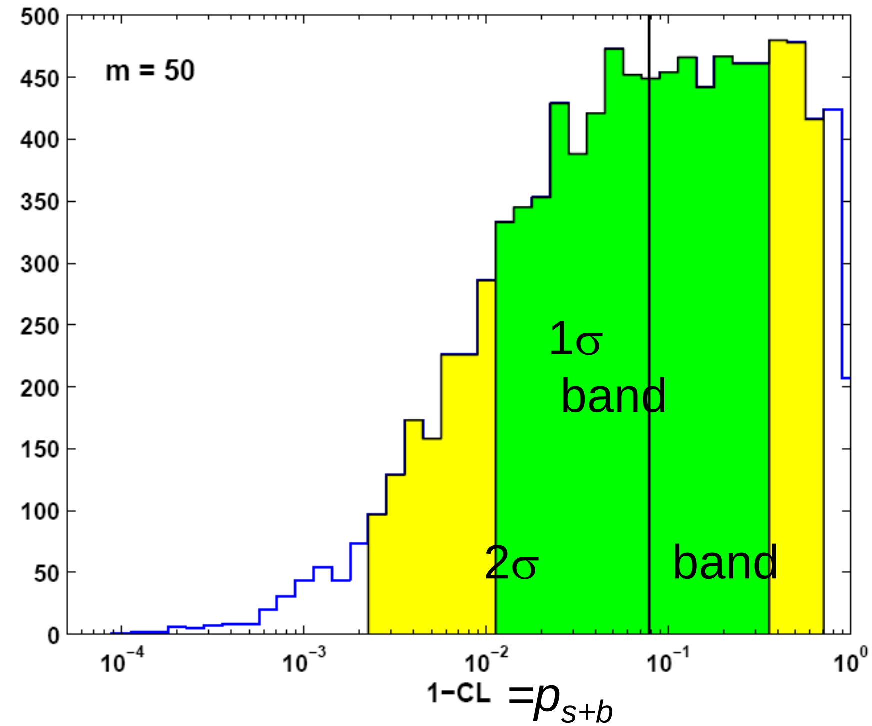
$$t_1 = -2\ln \frac{L\left(s + b(\hat{\hat{\theta}}_{(\mu=1)})\right)}{L\left(\hat{\mu}\cdot s + b(\hat{\theta})\right)}$$

- t_1 distributes as a χ^2 under **s(m_H)+b** experiments (H_1)
- The exclusion significance

$$Z = \sqrt{t_1}$$

can be expressed in terms of an equivalent exclusion CL

$$p_1 = p_{s+b} = 1 - CL$$



If p_{s+b}<5% we say that the signal is excluded at *>95% CL* ***(CL=1-p_{s+b})***

The exclusion sensitivity is the median CL, and using toy MCs one can find the bands

Exclusion Profile Likelihood Ratio

- A Higgs with a specific mass m_H is excluded at the 95% CL if the observed p-value of the ***s(m_H)+b*** hypothesis is below 0.05

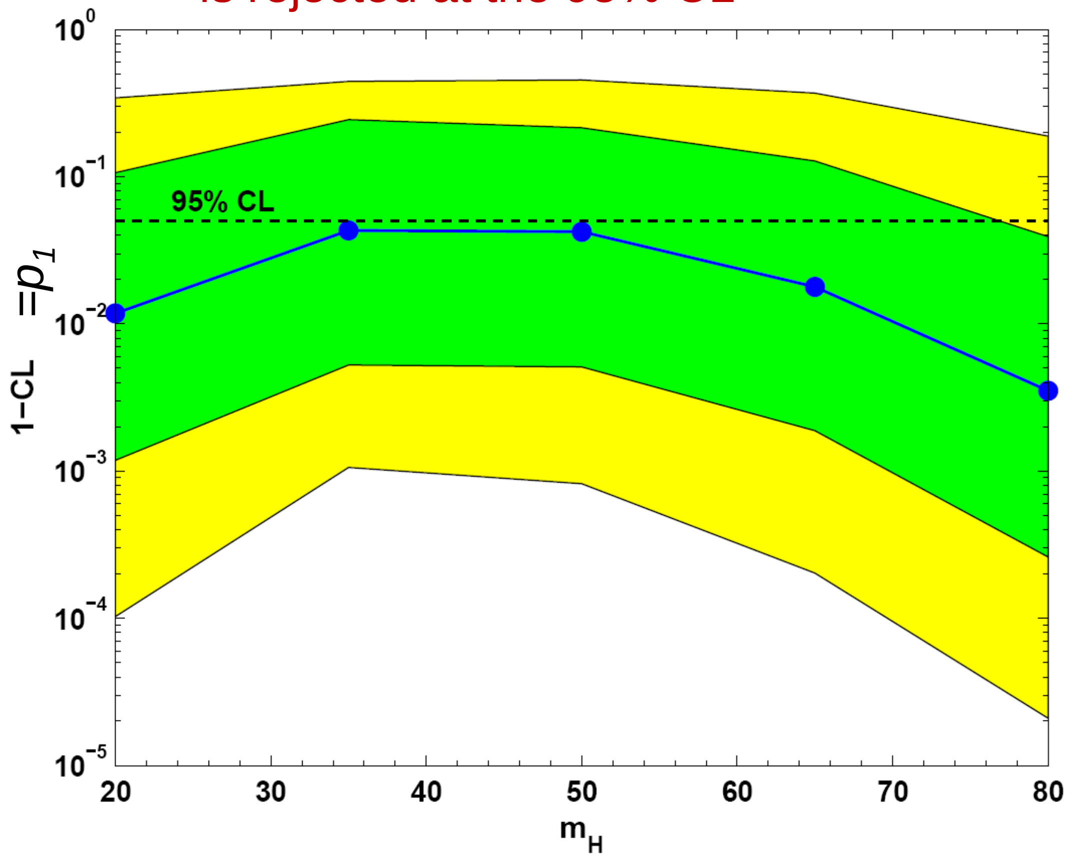
$$p_1 = p_{s+b} = 1 - CL$$

- In this example a Higgs Boson is expected to be excluded

$p_1 < 0.05$ (CL>95%)

in all the mass range

If p_{s+b}<5%, the s(mH)+b hypothesis is rejected at the 95% CL



Exclusion Bayesian

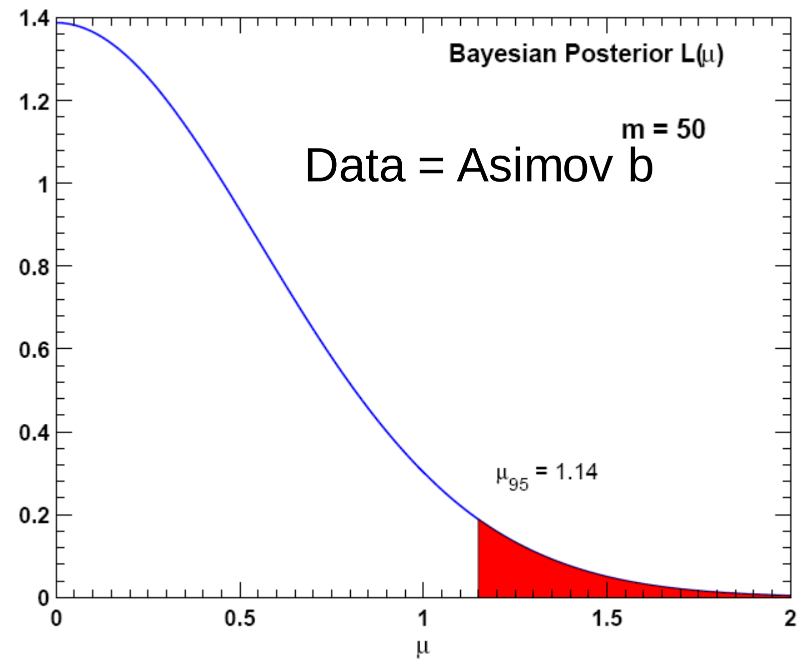
Let $prob(\mu | n)$ be the posterior for μ

$$prob(\mu | n) = \frac{\int L\left(\mu \cdot s + b(\theta)\right)\pi(\mu)\pi(\theta)d\theta}{\iint L\left(\mu \cdot s + b(\theta)\right)\pi(\mu)\pi(\theta)\,d\theta d\mu}$$

! Because there is no experimental information on the production cross section for the Higgs boson, in the Bayesian technique we assign a flat prior for the total number of selected Higgs events **arXiv:0911.3930v1**

- **NOTE:** The PDF of the posterior is based on the **one** observed data set with the likelihood integrated over the nuisance parameters
- It's a function of the hypothesis
- To set an upper limit on the signal strength $\mu = \frac{\sigma}{\sigma_{SM}}$ calculate the credibility interval $[0, \mu_{95}]$

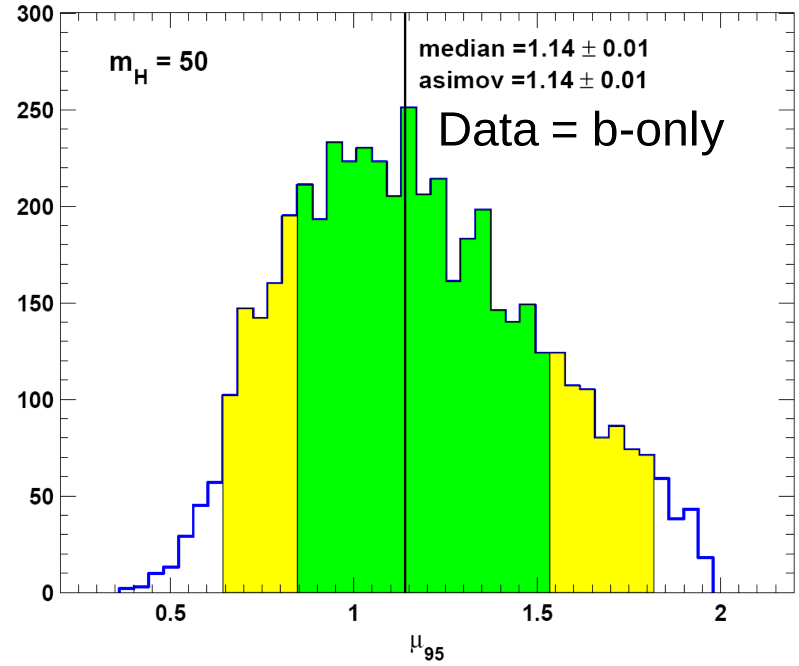
$$0.95 = \int_0^{\mu_{95}} \Pr ob(\mu | n) d\mu$$



Exclusion Bayesian

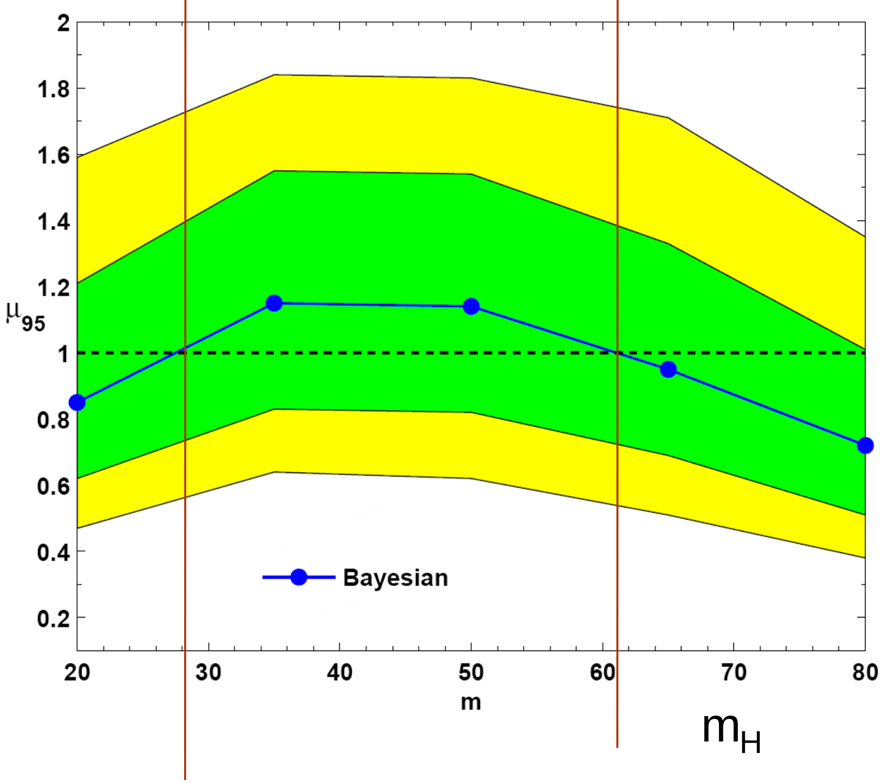
- **NOTE:** The toy MC are needed just to find the sensitivity bands, but once the data is delivered, it is sufficient to determine the upper limit using the posterior integration

$$0.95 = \int_0^{\mu_{95}} \Pr ob(\mu \mid n) d\mu$$



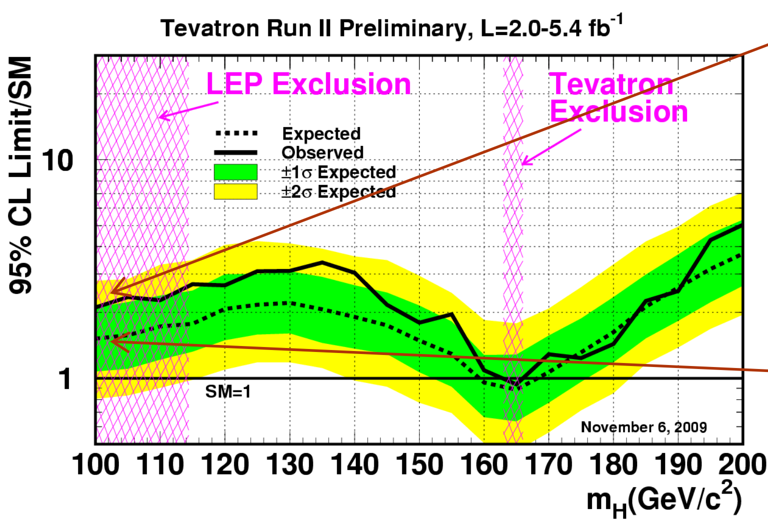
Exclusion Bayesian

- We find that **the credibility interval [0,μ_{95}] does not contain μ_{95}=1** (SM) for m_H<28 or m_H>61
- This is sometimes wrongly expressed as an exclusion at the 95% frequentist Confidence Level



TEVATRON Exclusion

$$\mu = \frac{\sigma}{\sigma_{SM}}$$



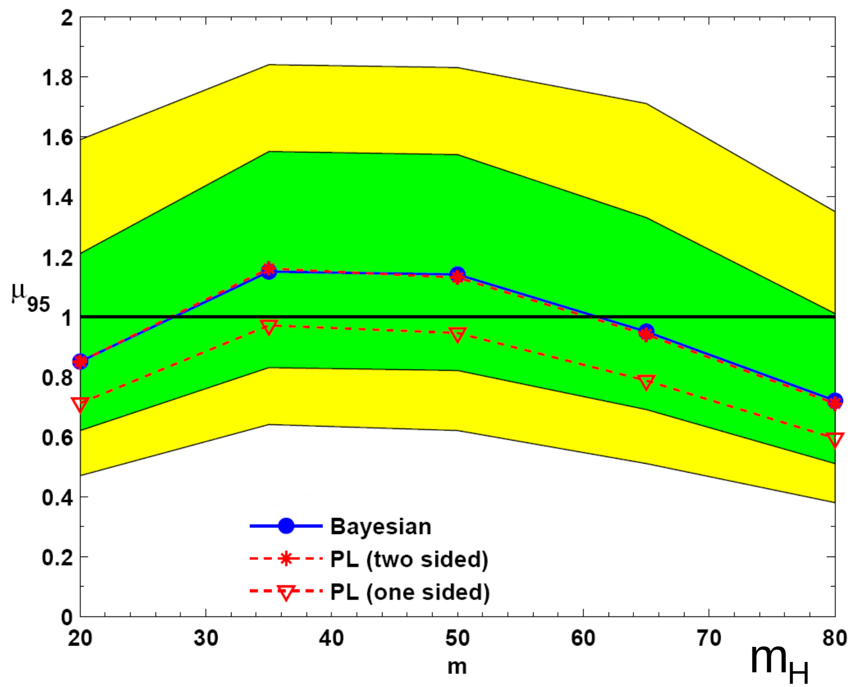
Combined CDF and D0 Upper Limits on Standard Model Higgs-Boson Production with 2.1 - 5.4 fb-1 of Data NOV 2009

- The 95% C.L. upper limits on Higgs boson production are a factor of 2.70 times the SM cross section for a Higgs boson mass of $m_H = 115$

  The corresponding median upper limits expected in the absence of Higgs boson production are 1.78. The mass range excluded at 95% C.L. for a SM Higgs is $163 < mH < 166$ GeV/c2, with an expected exclusion of $159 < mH < 168$ GeV/c2.

Exclusion Bayesian vs PL Ratio

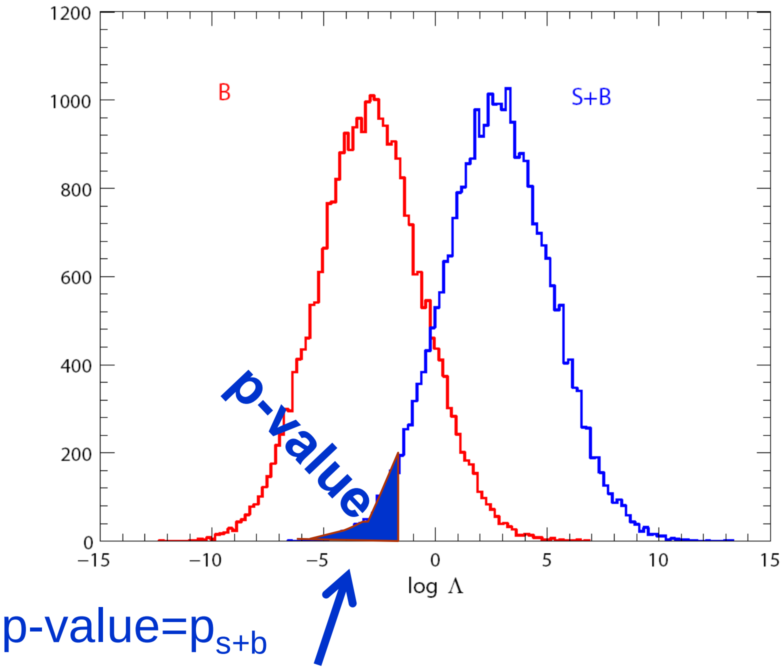
•Comparing a credibility Bayesian interval to 95% frequentist CL is like comparing **oranges** to **apples**….Yet



•**NOTE:** One has to be careful about the 1-sided vs 2-sided significance

The NP Likelihood Ratio method

- Use the LR as a test statistics $$t^{NP} = -2\ln\frac{L(H_0)}{L(H_1)} = \frac{L\big(b(\theta)\big)}{L\big(s+b(\theta)\big)}$$
- To take systematics into account integrate the nuisance parameters or profile them
- The exclusion is given by the $s(m_H)+b$ hypothesis p-value p_{s+b}
- If $p_{s+b}<5\%$, the **$s(m_H)+b$ hypothesis** is rejected at the 95% CL



The modified frequentist CL_s

- A downward fluctuation of the background might lead to an exclusion of a signal to which one is not sensitive (with a very low cross section)
- To protect against such fluctuations, the CL was redefined in a conservative non-frequentist way to be

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

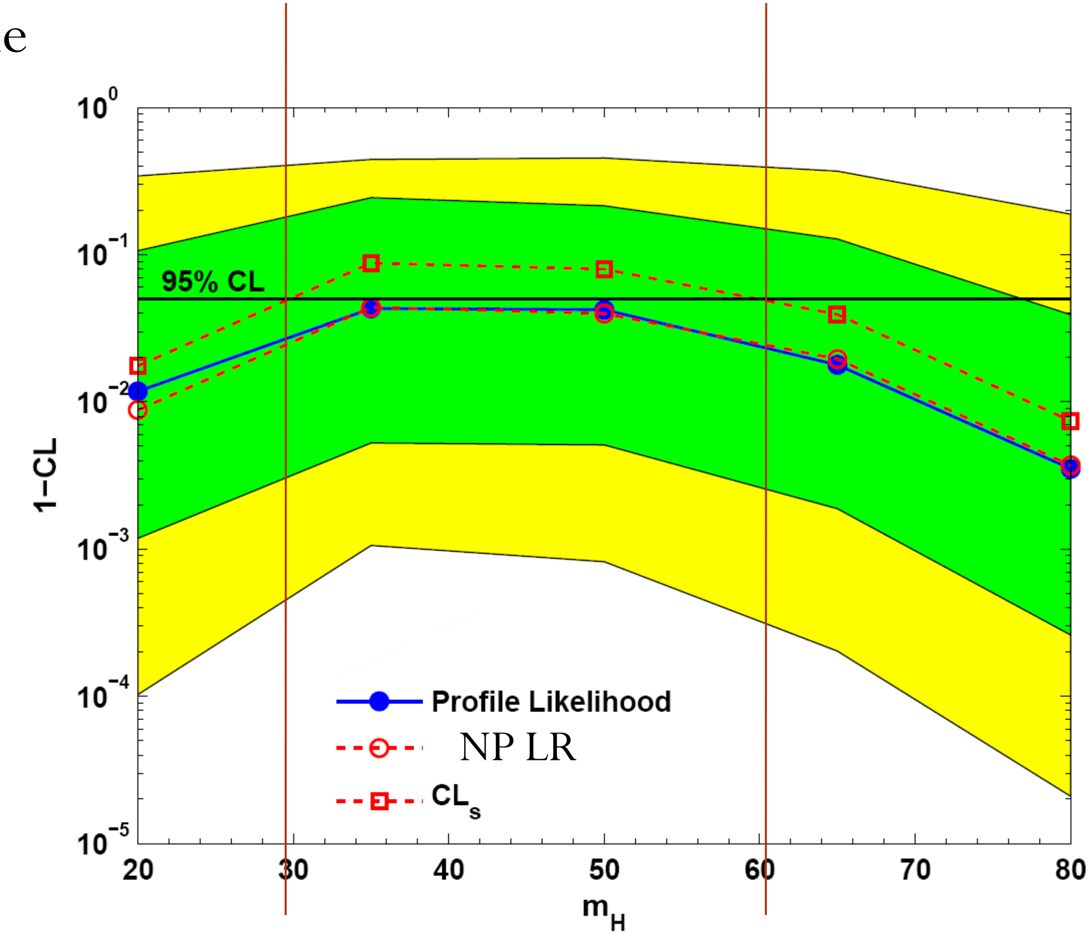
- Statisticians do not like this p-values ratio, yet, physics-wise it is conservative in a sense of coverage.

Alex Read
J.Phys.G28:2693-2704,2002

The modified frequentist CLs

$$CL_s \equiv 1 - CL$$

- In the toy example, while using PL or the NP LR the Higgs is excluded in all the mass range, the CL_s reduces the sensitivity and does not allow to exclude a Higgs with $30 < m_H < 60$



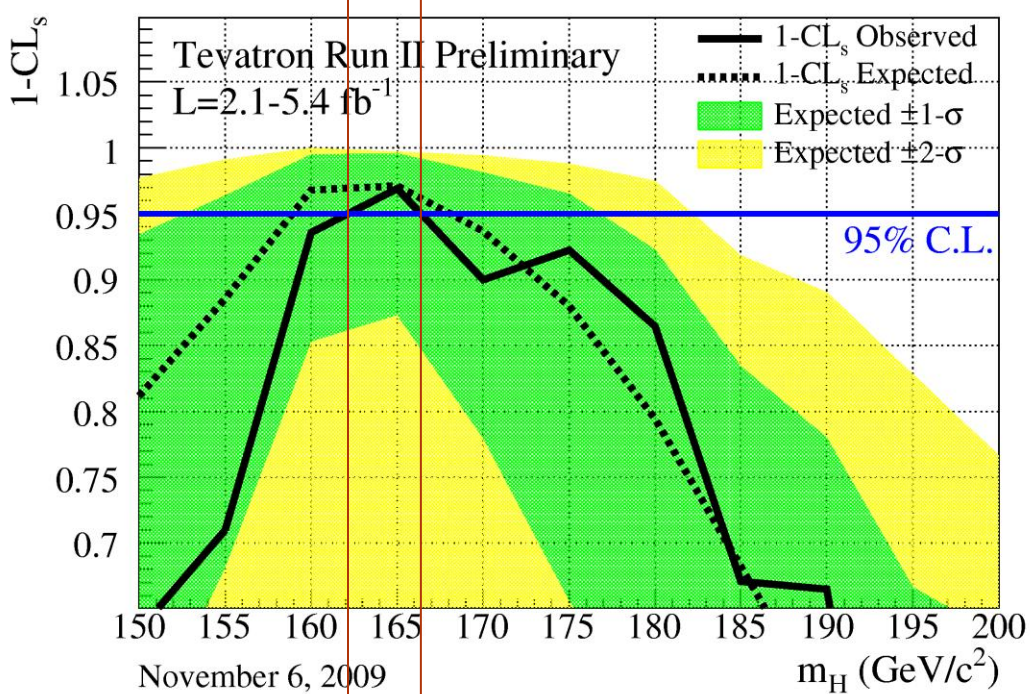
TEVATRON EXCLUSION

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

$$CL \equiv 1 - CL_s$$

- The excluded region obtained by finding the intersections of the linear interpolations of the observed $1-CL_S$ is larger than that obtained with the Bayesian calculation. We choose to quote the exclusion region using the Bayesian calculation.

If CL>95% the Higgs is excluded



Combined CDF and D0 Upper Limits on Standard Model Higgs-Boson Production with 2.1 - 5.4 fb-1 of Data
NOV 2009

Better exclusion than Bayesian $163<m_H<166$

The RooStats Project

- All the hypothesis testing algorithms described in this talk and more (Neyman construction…) are coded in RooStats which is a spin off the Root syatem (Rene Brun)
- RooStats allows to combine search results of experiments in order to increase the sensitivity.
- See talk by Alfio Lazzaro in this conference

Conclusions

- We have explored and compared all the methods to test hypotheses that are currently in use in the High Energy Physics market (Profile LR, NP-LR, NP-CL_s, Bayesian)
- We have shown that all methods tend to give similar results, (for both exclusion and discovery using flat priors) weather one integrates the nuisance parameters or profile them
- We have explained the Look Elsewhere Effect and derived a thumb rule formula for it:

$$trial\# \approx \frac{\Delta_m}{3\sigma_m} Z_{fix}$$

- Even though we have used typical case studies, real life might be different and all available methods should be explored (as done already in the TEVATRON).
- The RooStats project allows the exploration of all methods and easy combination of search results from different experiments

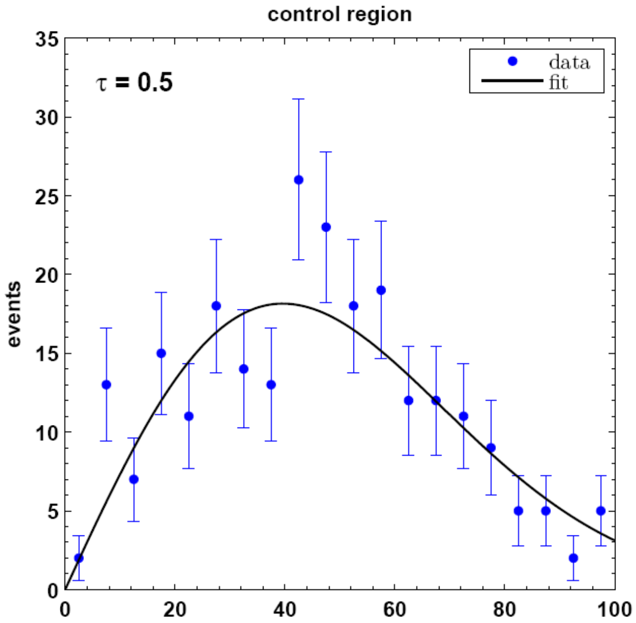
BACKUP

Let's Play the Real Game

We perform 2 measurements

One in a sideband that contains no signal
And constraints the BG

$$b = b(\theta) \qquad \langle n_b \rangle = b$$



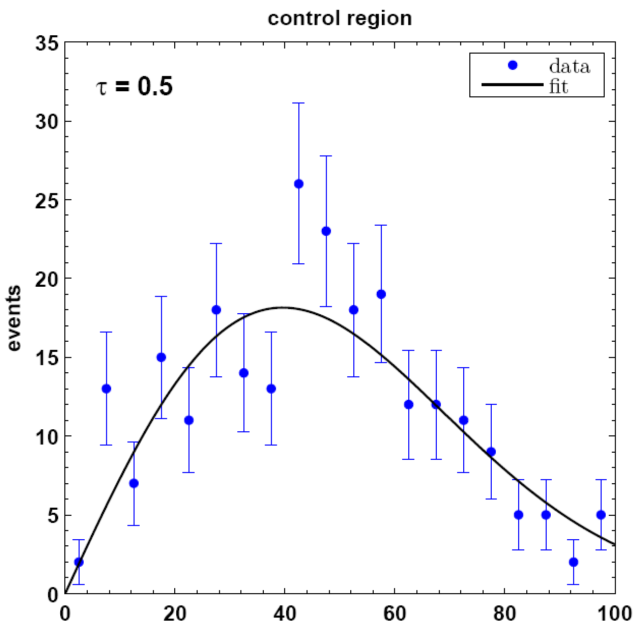
Let's Play the Real Game

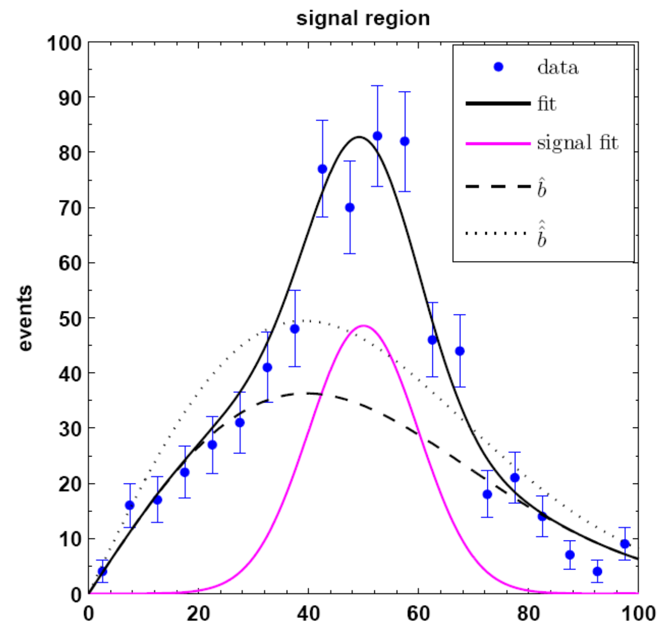
We perform 2 measurements

One in a sideband that contains no signal
And constraints the BG

The other is the main measurement

$b = b(\theta)$ $\quad \langle n_b \rangle = b$ $\quad \langle n \rangle = \mu s(m_H) + b$

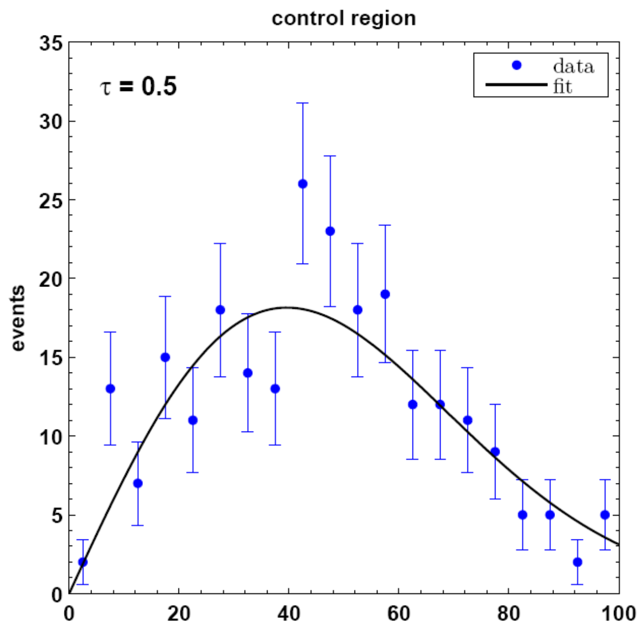




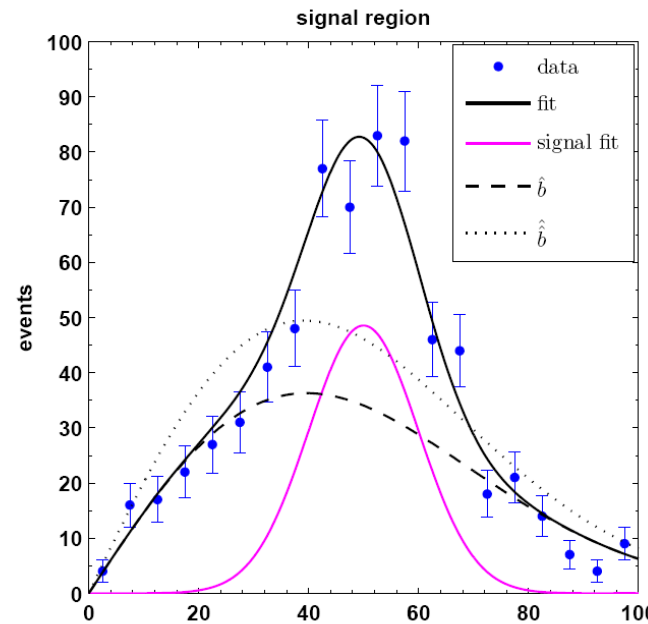
A Simultaneous Fit

Two measurements $n \sim \mu s(m_H) + b$ $n_b \sim b$

$$L(\mu \cdot s + b) = L(\mu \cdot s + b \mid n, n_b) = \prod_{i=1}^{nbins} Poisson(n_i \mid \mu \cdot s_i + b_i) \cdot Poisson(n_{b,i} \mid b_i)$$



m_H



m_H