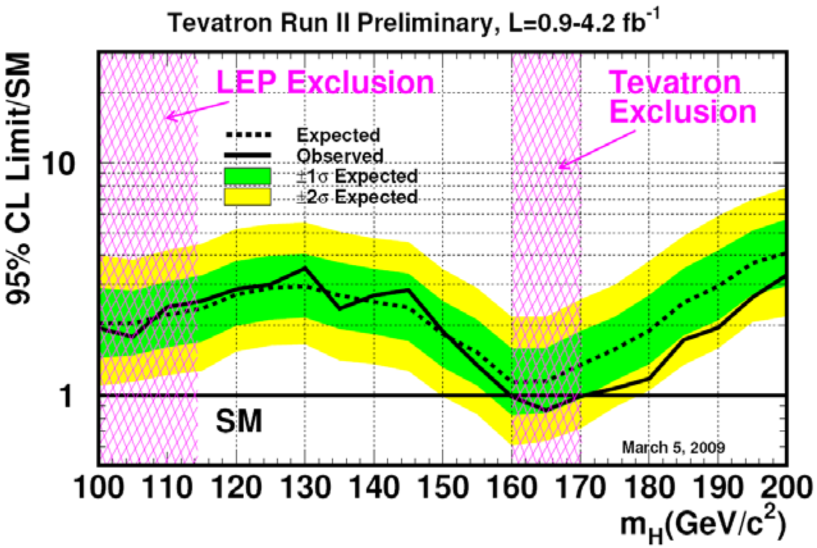
Statistical issues for Higgs Physics

Eilam Gross, Weizmann Institute of Science

Acknowledgements: Louis Lyons & Ofer Vitells

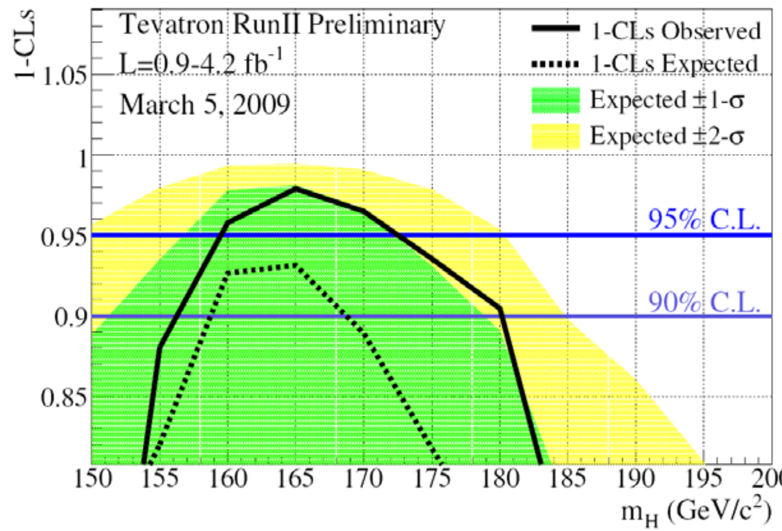
Combined CDF and DØ Upper Limits on Standard Model Higgs-Boson Production with up to 4.2 fb^{-1} of Data

Bayesian



FERMILAB-PUB-09-060-E
CDF Note 9713
DØ Note 5889

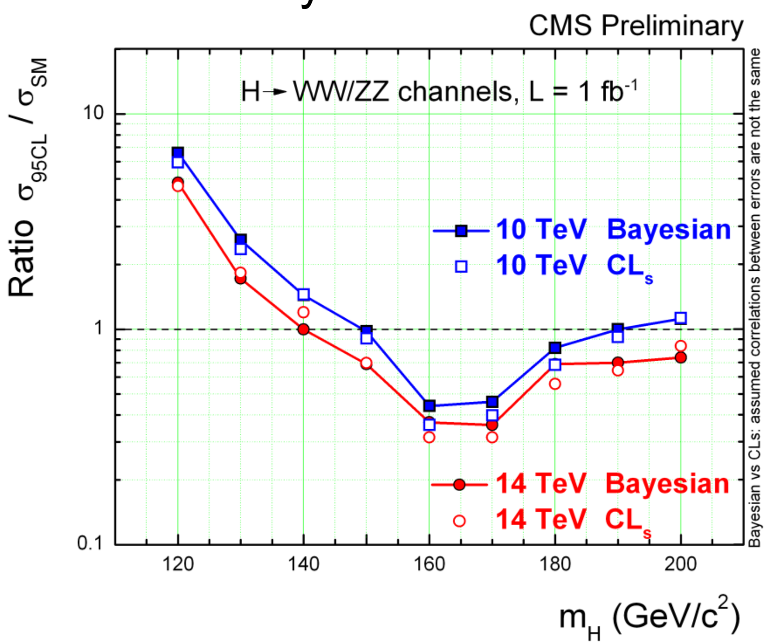
Modified Frequentist - CLs



FERMILAB-PUB-09-060-E
CDF Note 9713
DØ Note 5889

CMS & ATLAS Higgs Prospects

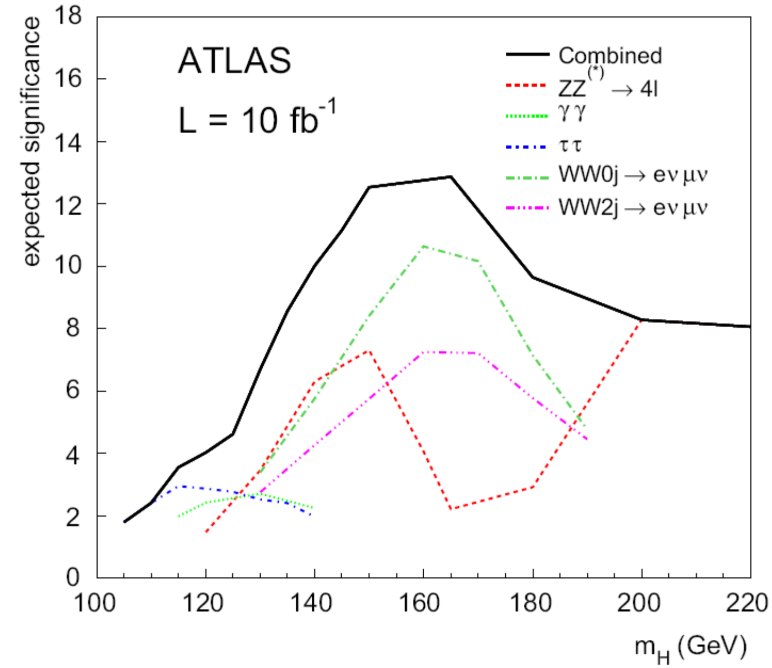
Bayesian



CMS

Gr'egory Schott SUSY 2009

Profile Likelihood



ATLAS

CERN-OPEN-2008-020

December 2008

Discovery vs Exclusion

- Higgs statistics is about testing one hypothesis against another hypothesis
  - One hypothesis is the Standard Model with no Higgs Boson (H_0)
  - Another hypothesis is the SM with a Higgs boson with a specific mass m_H (H_1)
- Rejecting the No-Higgs (H_0) hypothesis →DISCOVERY
- Rejecting the Higgs hypothesis (H_1) →EXCLUDING the Higgs

DISCOVERY

The Discovery Case Study

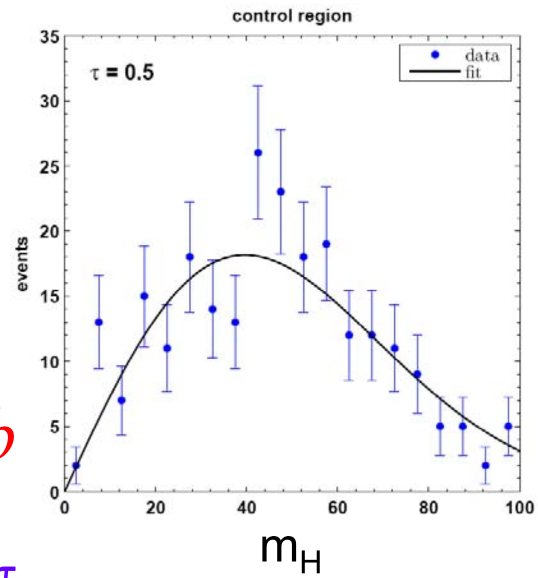
- We assume a Gaussian "Higgs" signal (**s**) on top of a Rayleigh shaped background (**b**)
- The signal strength is μ

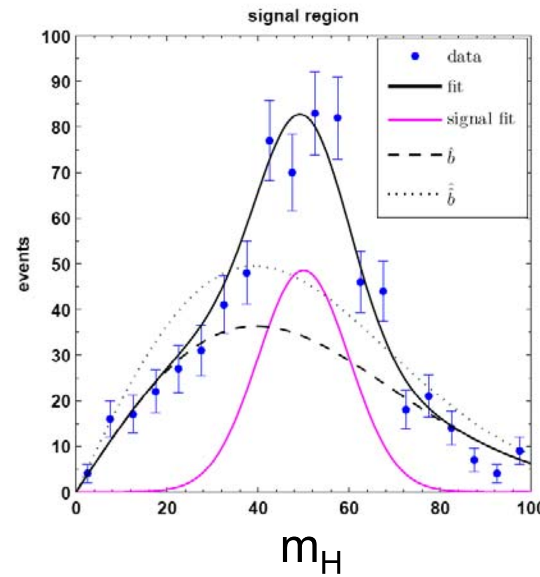
$$\langle n \rangle = \mu s + b$$

  - μ=1, SM Higgs
  - μ=0, SM without Higgs
- **Two hypothetical measurements**
- **Data** $n \sim \mu s(m_H) + b$
- **BG control sample scaled to the expected BG via a factor τ** $m \sim \tau b$

NOTE: $b=b(\theta)$

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

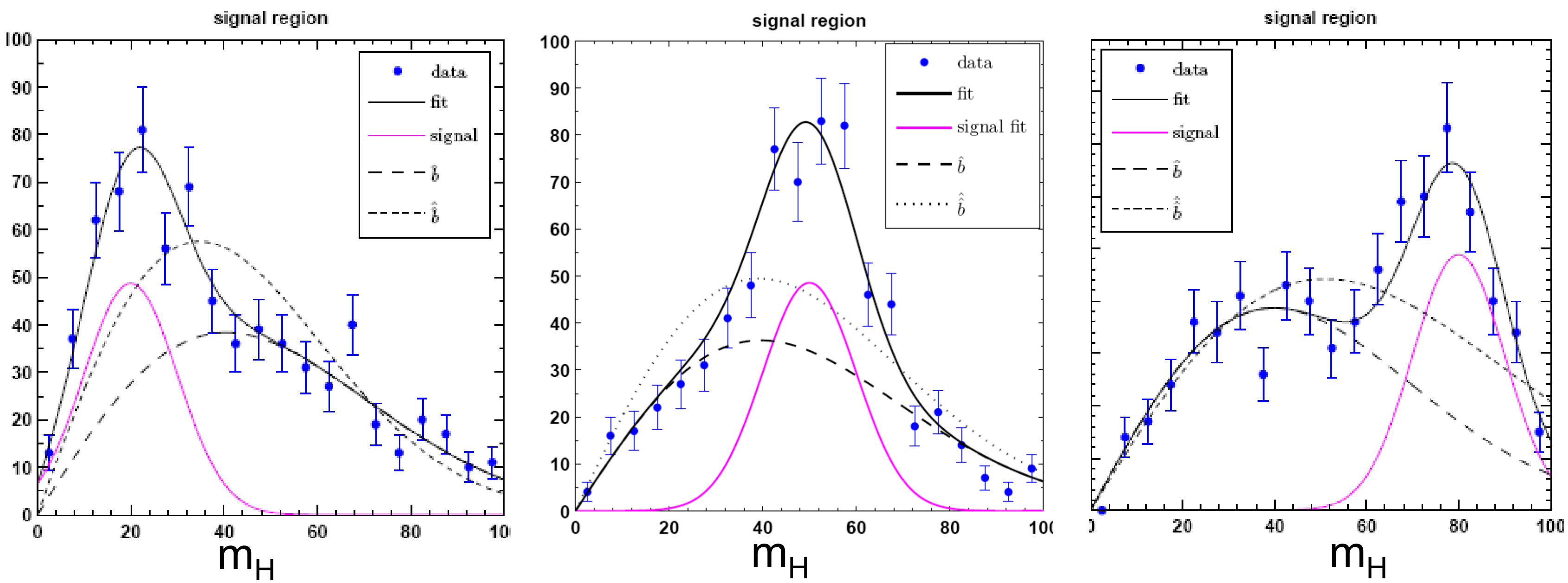




$$L(n,m \mid s+b\ (\theta)) = \prod_{i=1}^{nbins} Poisson(n_i; s_i + b_i) \cdot Poisson(m_i; \tau b_i)$$

The Discovery Case Study

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



CL motivation-
The Neyman-Pearson Lemma

- 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 $\Lambda(x) = \frac{L(H_1 \mid x)}{L(H_0 \mid x)} \leq \eta, \; P(\Lambda(x) \leq \eta \mid H_0) = \alpha$
- Define a **test statistic** $$\Lambda = \frac{L(H_1)}{L(H_0)}$$
- **Note:** Likelihoods are functions of the data, even though we often not specify it explicitly $$\Lambda(x) = \frac{L(H_1 \mid x)}{L(H_0 \mid x)}$$

The frequentist LR (CL) method

- Define a test statistics

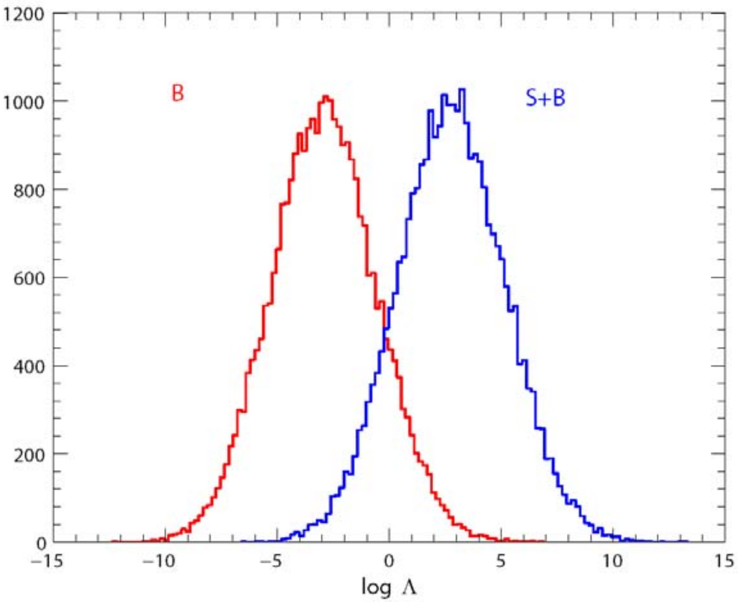
$$\Lambda = \frac{L(H_1)}{L(H_0)} = \frac{L\left(n, m \mid s + b(\theta)\right)}{L\left(n, m \mid b(\theta)\right)}$$

The frequentist LR (CL) method

- Define a test statistics

$$\Lambda = \frac{L(H_1)}{L(H_0)} = \frac{L\left(n,m \mid s+b(\theta)\right)}{L\left(n,m \mid b(\theta)\right)}$$

- Use MC to generate the pdf of Λ under H_0(B only) and H_1 (S+B)

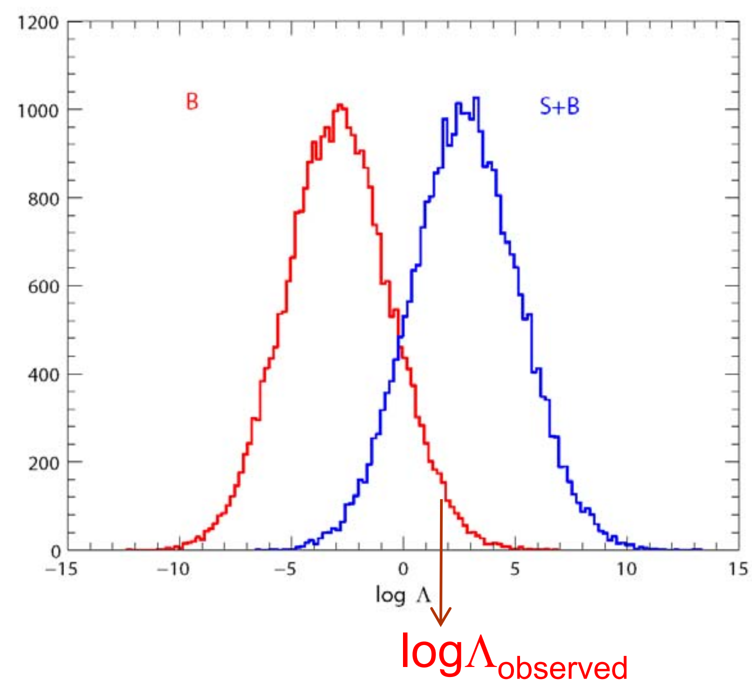


The frequentist LR (CL) method

- Define a test statistics

$$\Lambda = \frac{L(H_1)}{L(H_0)} = \frac{L\left(n, m \mid s + b(\theta)\right)}{L\left(n, m \mid b(\theta)\right)}$$

- Use MC to generate the pdf of Λ under H_0(B only) and H_1 (S+B)
- Let Λ_{obs} be a result of **one experiment** (LHC)

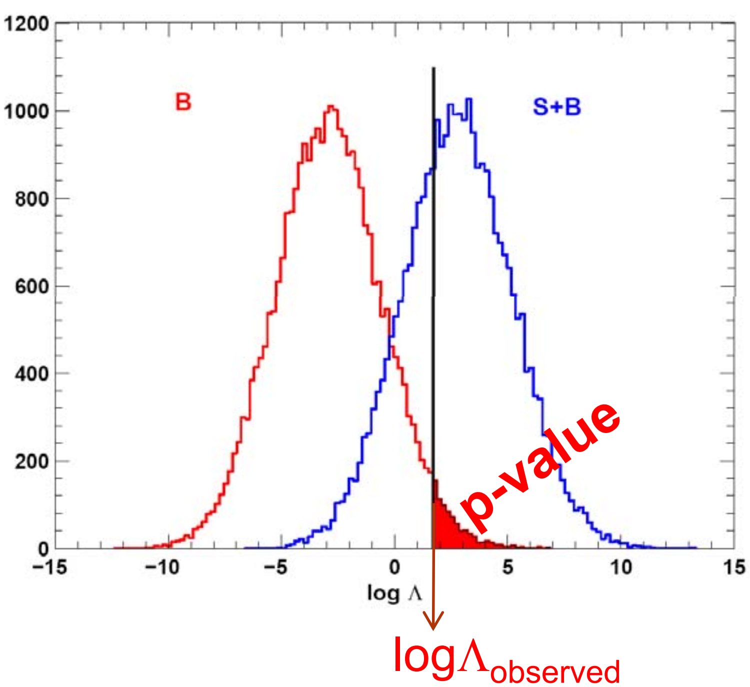


The frequentist LR (CL) method

- Define a test statistics

$$\Lambda = \frac{L(H_1)}{L(H_0)} = \frac{L\left(n,m \mid s+b(\theta)\right)}{L\left(n,m \mid b(\theta)\right)}$$

- Use MC to generate the pdf of Λ under H_0(B only) and H_1 (S+B)
- Let Λ be a result of **one experiment** (LHC)
- The p-value is the probability to get an observation which is less B-like than the observed one
- If the result of the experiment (LHC) yields a p-value< $2.8\cdot10^{-7}$ a 5σ discovery is claimed
- NOTE: the p-value can be interpreted as a frequency → this is a **frequentist** approach

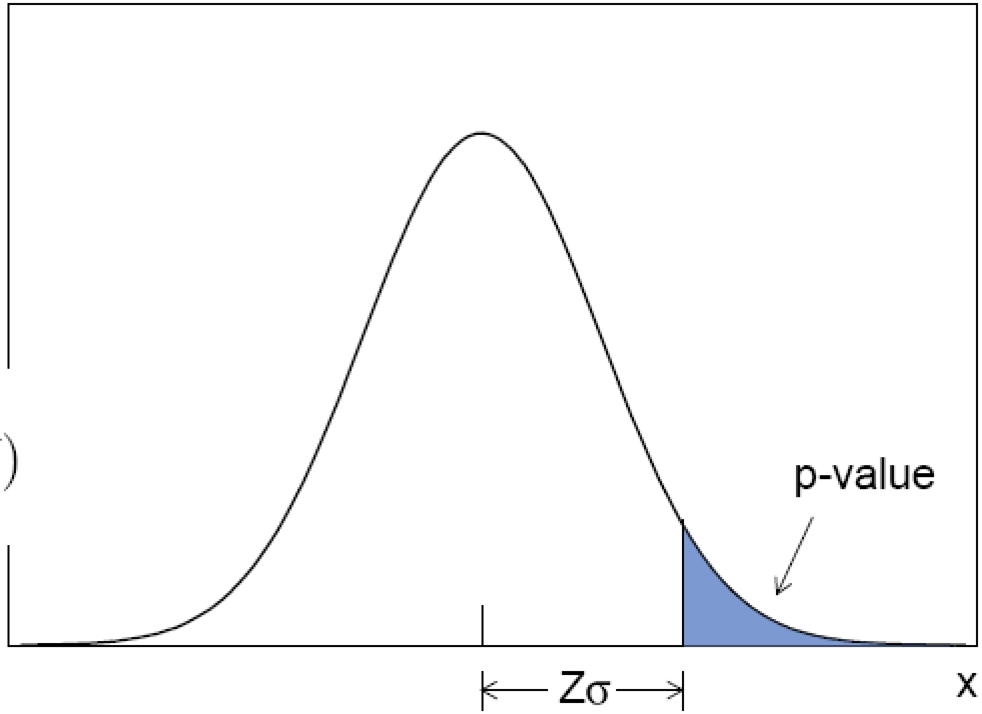


From p-values to Gaussian significance

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 $p = 2.87 \times 10^{-7}$.

Systematics

- Normally, the background, b(θ), has an uncertainty which has to be taken into account. In this case θ is called a nuisance parameter (which we associate with background systematics)
- How can we take into account the nuisance parameters?
- One way: marginalize them (integrate them out using priors)
  → the Hybrid CL (mix frequentist and Bayesian approach)

prior

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

$$\Lambda_{Hybrid} = \frac{\int L\left(n,m \mid s+b(\theta)\right)\pi(\theta)d\theta}{\int L\left(n,m \mid b(\theta)\right)\pi(\theta)\,d\theta}$$

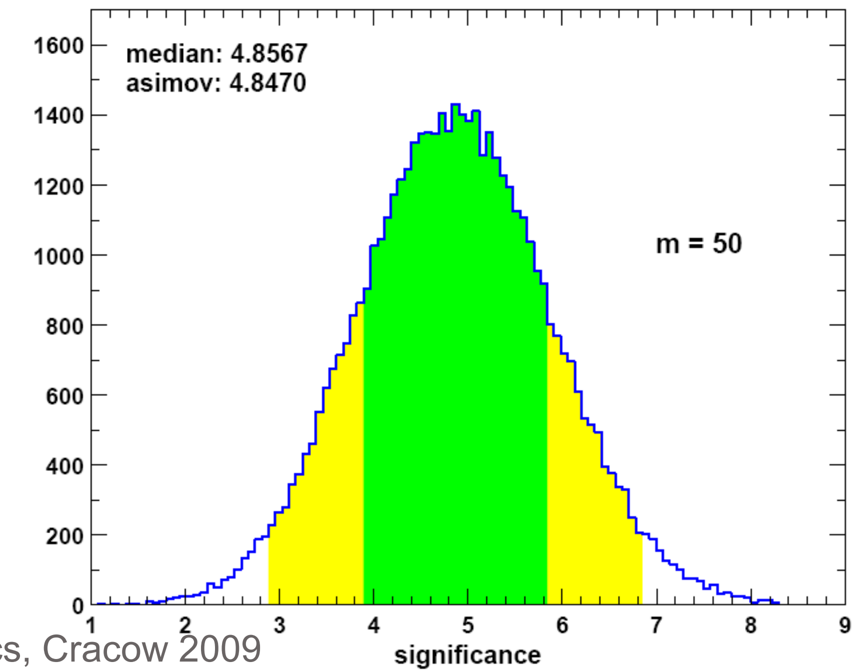
- Another way is profiling via the MLEs:

$$\Lambda_{PL} = \frac{L(H_1)}{L(H_0)} = \frac{L\left(n,m \mid s+b(\hat{\hat{\theta}}_{s+b})\right)}{L\left(n,m \mid b(\hat{\theta}_b)\right)}$$

The Profiled CL way

$$\Lambda_{PL} = \frac{L(H_1)}{L(H_0)} = \frac{L\left(n,m \mid s+b(\hat{\hat{\theta}}_{s+b})\right)}{L\left(n,m \mid b(\hat{\theta}_b)\right)}$$

- The median significance can be obtained with the one Asimov data set
  n~s+b, m~b



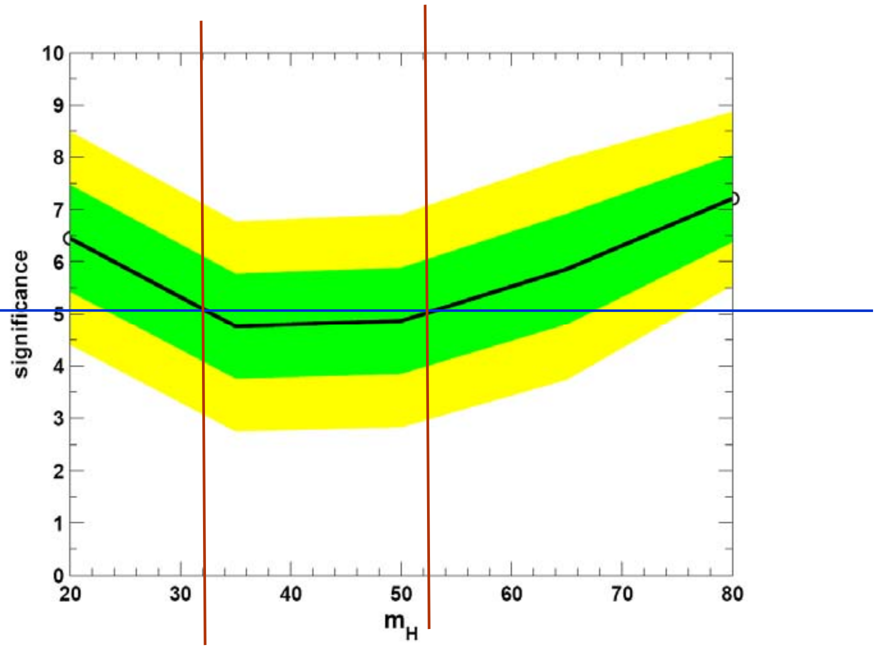
ATLAS, CERN – Open 2008-029
Cowan, Cranmer, E.G., Vitells, in preparation

The Profiled CL way

$$\Lambda_{PL} = \frac{L(H_1)}{L(H_0)} = \frac{L\left(n,m\,|\,s+b(\hat{\hat{\theta}}_{s+b})\right)}{L\left(n,m\,|\,b(\hat{\theta}_b)\right)}$$

In this example a Higgs with a mass m_H<32 or m_H>52 is expected to be discovered, i.e.

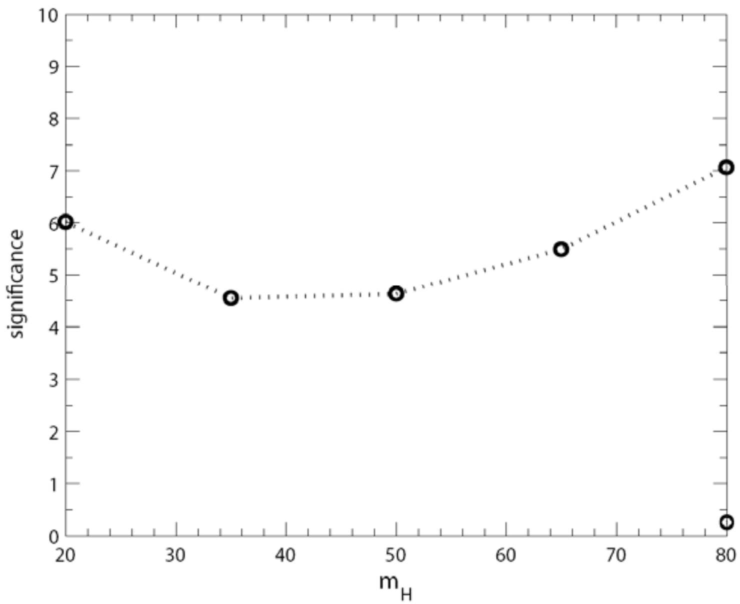
if the Higgs exists in this mass range it will be discovered >50% of hypothetical LHC experiments



The Profile Likelihood Ratio

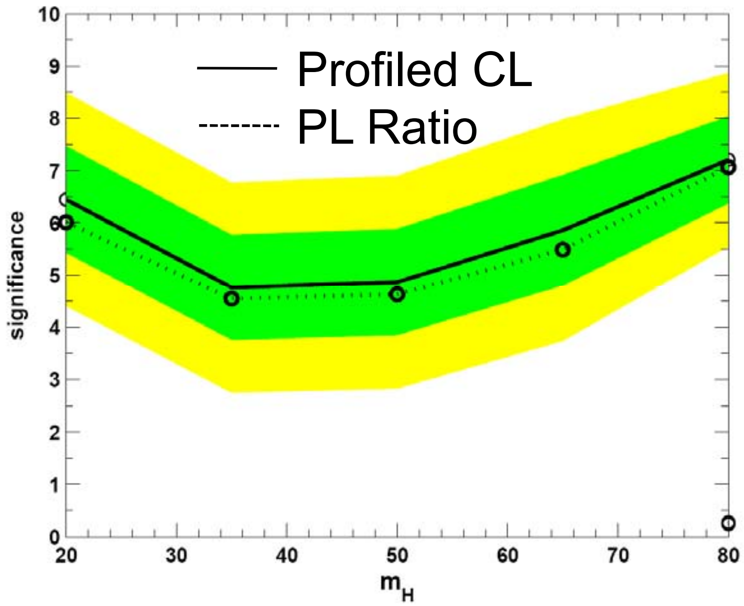
- PL Ratio: $\lambda(\mu) = \dfrac{L\left(\mu \cdot s + b(\hat{\hat{\theta}}_{(\mu=1)})\right)}{L\left(\hat{\mu} \cdot s + b(\hat{\theta})\right)}$

- Test the null H_0 hypothesis $\lambda(0) = \dfrac{L\left(b(\hat{\hat{\theta}}_{(\mu=1)})\right)}{L\left(\hat{\mu} \cdot s + b(\hat{\theta})\right)}$



The frequentist
Profile Likelihood Ratio vs Profiled CL

- Why using a method with a slightly lower sensitivity?



The Profile Likelihood Ratio & Wilks theorem

- PL Ratio: $\lambda(\mu) = \dfrac{L\left(\mu \cdot s + b(\hat{\hat{\theta}}_{(\mu=1)})\right)}{L\left(\hat{\mu} \cdot s + b(\hat{\theta})\right)}$

- Test the null H_0 hypothesis

$$\lambda(0) = \frac{L\left(b(\hat{\hat{\theta}}_{(\mu=1)})\right)}{L\left(\hat{\mu} \cdot s + b(\hat{\theta})\right)}$$

- Test statistics $q_0 = -2 \ln \lambda(0)$

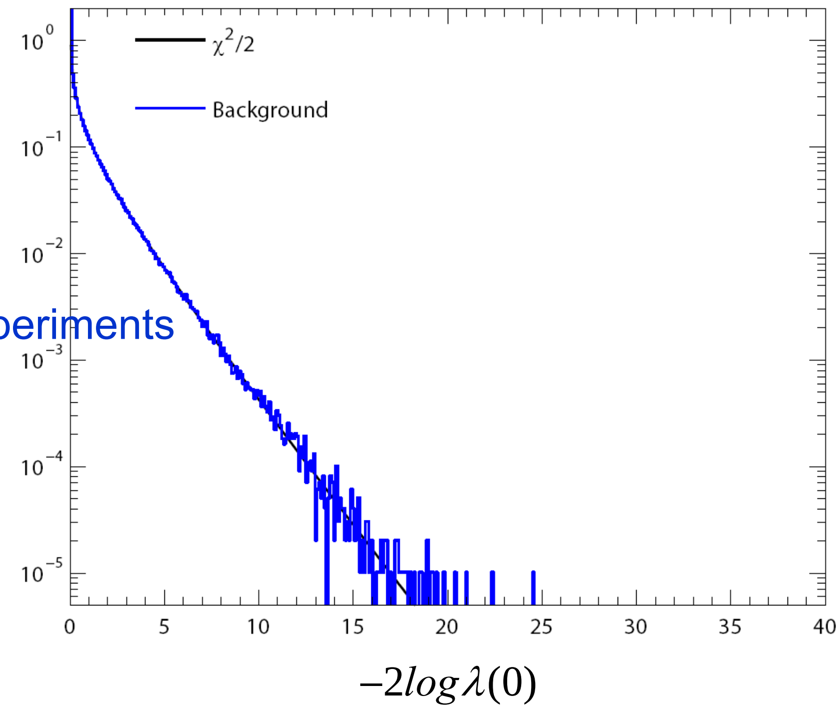
- For B-only experiments

$$\hat{\mu} \sim 0, \; q_0 \sim -2 \ln 1 \sim 0$$

Wilks theorem:

$q_0 = -2 \log \lambda(0)$ distributes as χ^2 for b-only experiments

$$\left(\hat{\mu} > 0 \rightarrow \frac{1}{2}\chi^2\right)$$



Using Wilks theorem there is no need to generate a Billion of events to get the BG pdf and test a 5σ sensitivity

The Profile Likelihood Ratio

- PL Ratio: $\lambda(\mu) = \dfrac{L\left(\mu \cdot s + b(\hat{\hat{\theta}}_{(\mu=1)})\right)}{L\left(\hat{\mu} \cdot s + b(\hat{\theta})\right)}$
- Test the null H_0 hypothesis $\lambda(0) = \dfrac{L\left(b(\hat{\hat{\theta}}_{(\mu=1)})\right)}{L\left(\hat{\mu} \cdot s + b(\hat{\theta})\right)}$
- Test statistics $q_0 = -2 \ln \lambda(0)$
- For B-only experiments

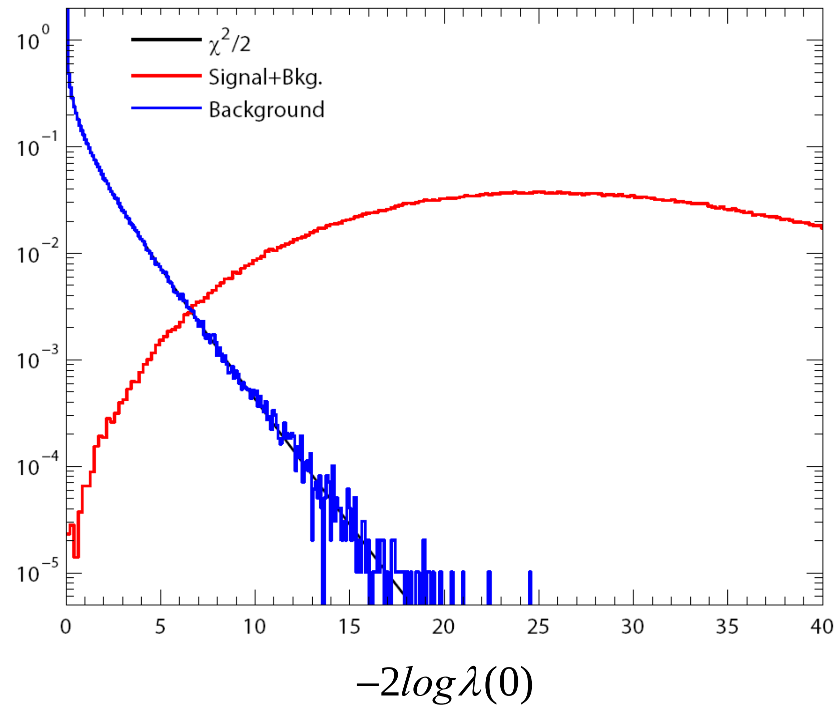
  $\hat{\mu} \sim 0,\ q_0 = -2 \ln 1 \sim 0$

  and distributes as χ^2
- For S+B experiments

  $\hat{\mu} \sim 1,\ q_0 = -2 \ln 0 >> 1$
- The significance is given to a very good approximation by

  $Z(x) = \sqrt{q_0(x)} = \sqrt{-2 \ln \lambda(0 \mid x)}$

  and is dependent on the data x



Eilam Gross, Higgs Statistics, Cracow 2009

The Profile Likelihood Ratio

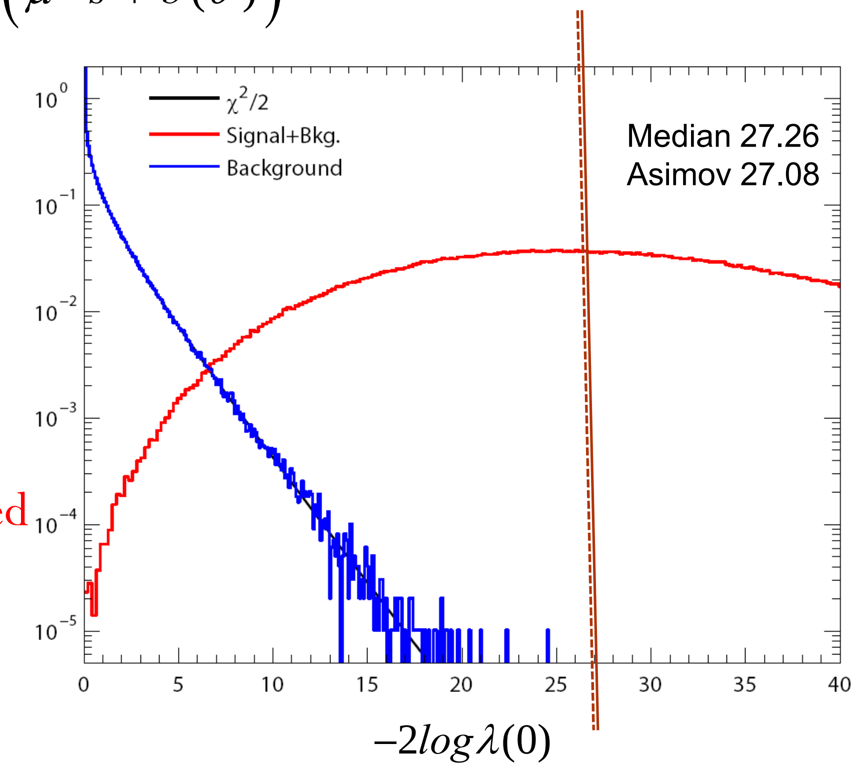
- PL Ratio: $\lambda(\mu) = \dfrac{L\left(\mu \cdot s + b(\hat{\hat{\theta}}_{(\mu=1)})\right)}{L\left(\hat{\mu} \cdot s + b(\hat{\theta})\right)}$
- Test the null H_0 hypothesis $\lambda(0) = \dfrac{L\left(b(\hat{\hat{\theta}}_{(\mu=1)})\right)}{L\left(\hat{\mu} \cdot s + b(\hat{\theta})\right)}$
- Test statistics $q_0 = -2\ln\lambda(0)$
- For B-only experiments

  $\hat{\mu} \sim 0,\ q_0 = -2\ln 1 \sim 0$

  and distributes as χ^2
- For S+B experiments

  $\hat{\mu} \sim 1,\ q_0 = -2\ln 0 >> 1$
- The median significance can be obtained by the median of the $q_0 = -2\ln\lambda(0)$ distribution or

  using the Asimov data (x~S+B)



Eilam Gross, Higgs Statistics, Cracow 2009

The Bayes way

- Use the observed single data set *x* with priors to derive the posterior probability $P(H_1 | x)$ based on Bayes' theorem

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

- 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

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

$$B_{10}(\mu) = \frac{\int L\left(n,m | \mu \cdot s + b(\theta)\right)\pi(\theta)d\theta}{\int L\left(n,m | (\mu = 0) \cdot s + b(\theta)\right)\pi(\theta)\,d\theta}$$

$$B_{10} = \frac{\iint L\left(n,m | \mu \cdot s + b(\theta)\right)\pi(\mu)\pi(\theta)d\theta d\mu}{\int L\left(n,m | (\mu = 0) \cdot s + b(\theta)\right)\pi(\theta)\,d\theta}$$

PL vs Bayesian

- Using the saddle point approximation one can approximate (see appendix)

$$\log B_{10} \approx \frac{-2\log\lambda(0)}{2} + C; \qquad \log B_{10} \approx \frac{Z^2}{2} + C$$

and get a **rough estimation** for $\mathbf{B_{10}}$ (neglecting C)

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

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

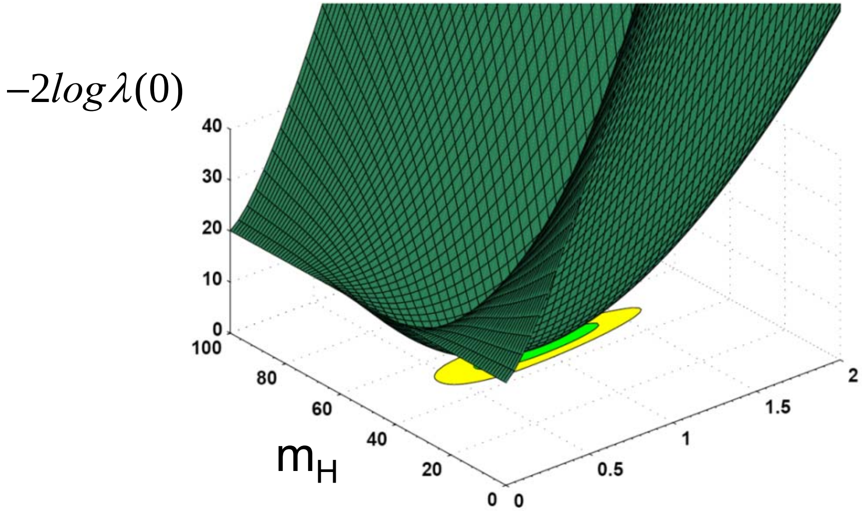
Let the Higgs mass, m_H, and the signal strength μ be 2 parameters of interest

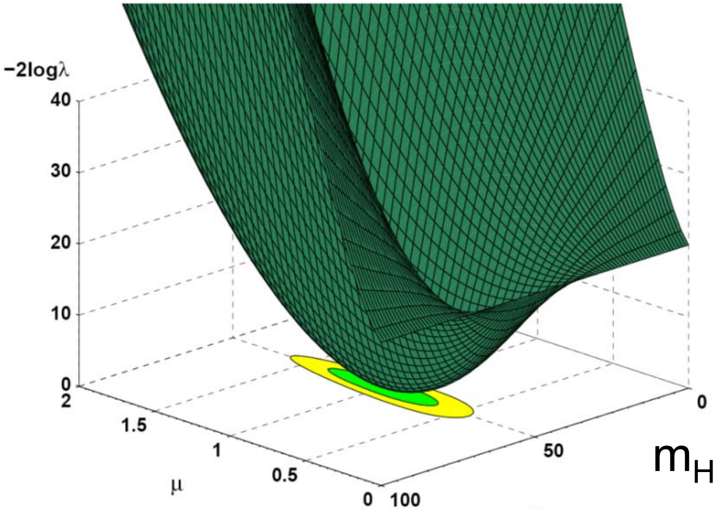
$$\lambda(\mu, m_H) = \frac{L(\mu, m_H, \hat{\hat{b}})}{L(\hat{\mu}, \hat{m}_H, \hat{b})}$$

Look Elsewhere Effect

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

$$\lambda(\mu, m_H \mid n = s_{(m_H=50)} + b, m = \tau b) = \frac{L(\mu, m_H, \hat{\hat{b}})}{L(\hat{\mu}, \hat{m}_H, \hat{b})}$$

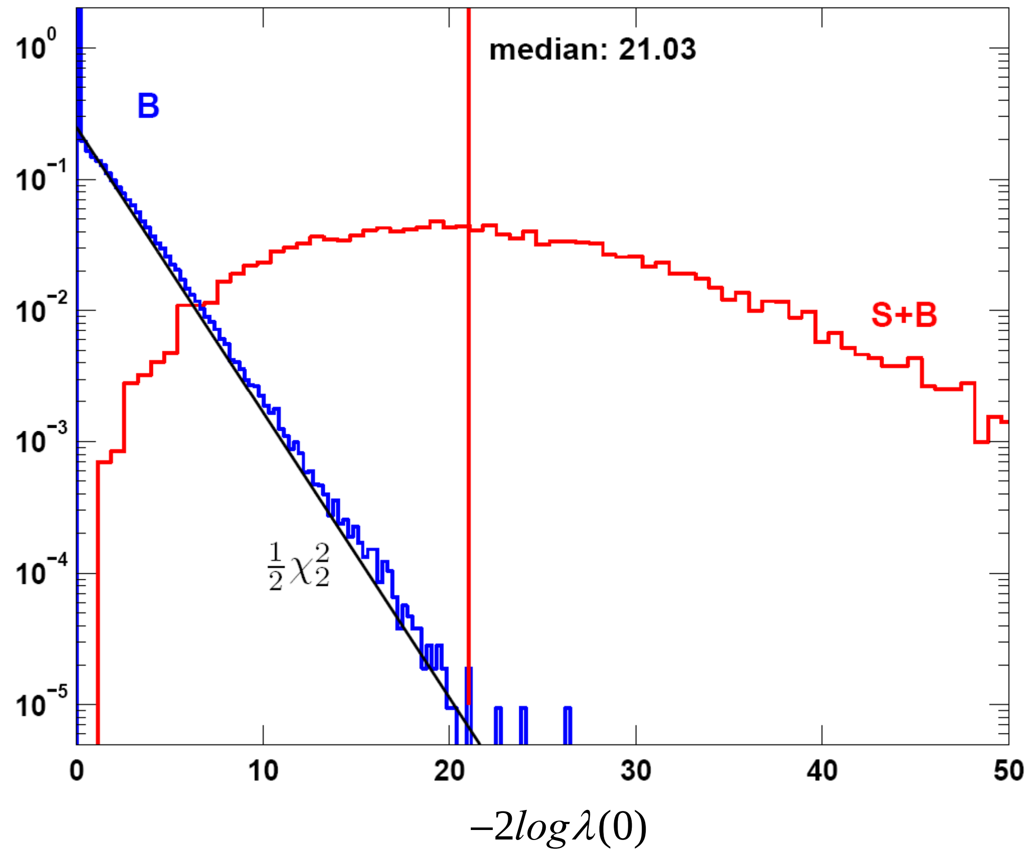




Look Elsewhere Effect

- Letting the Higgs mass float Wilks' theorem tells us that the background-only experiments will distribute as a χ_2^2

- The median sensitivity is given by the corresponding p-value

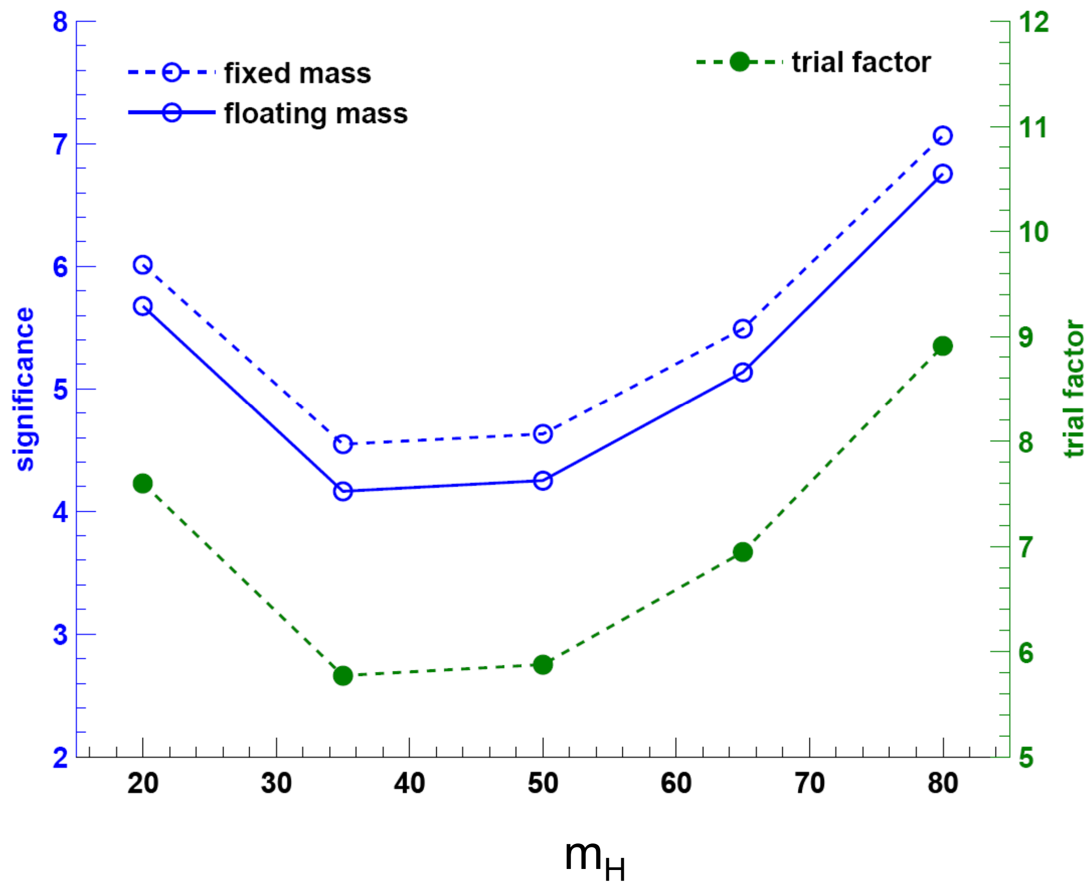


Look Elsewhere Effect

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

Back of the envelope:

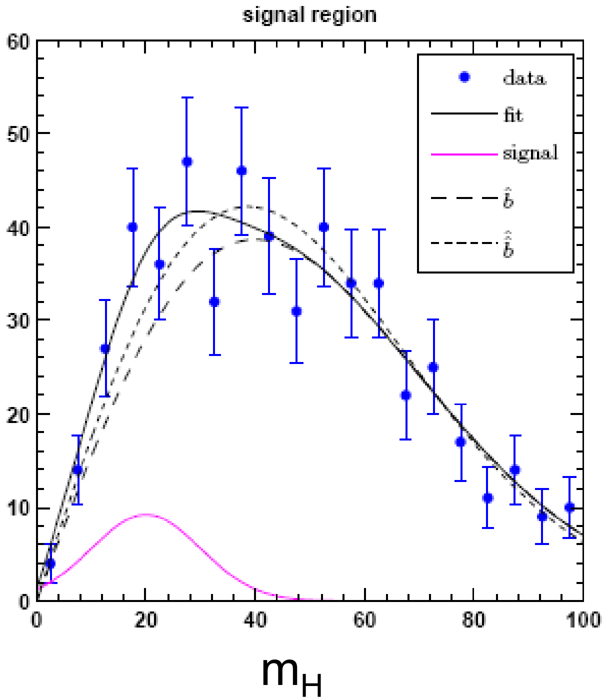
$$trial\ factor = \frac{range}{resolution} = \frac{\Gamma_m}{\sigma_m}$$



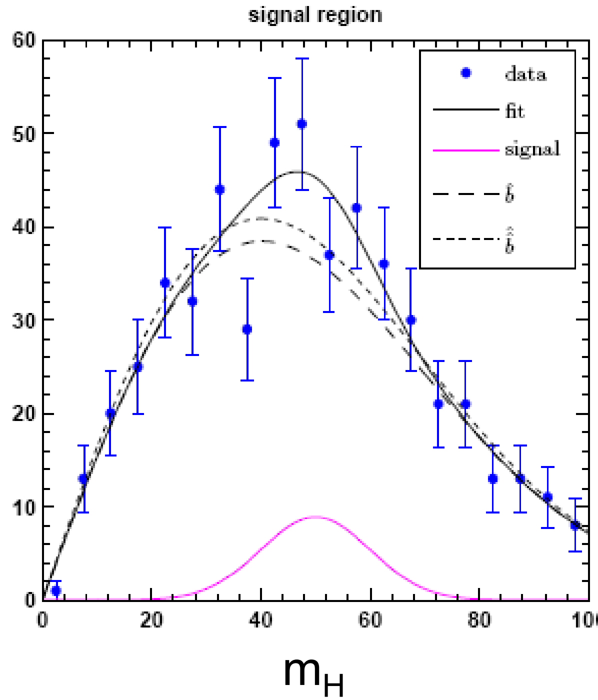
EXCLUSION

Exclusion Case Study

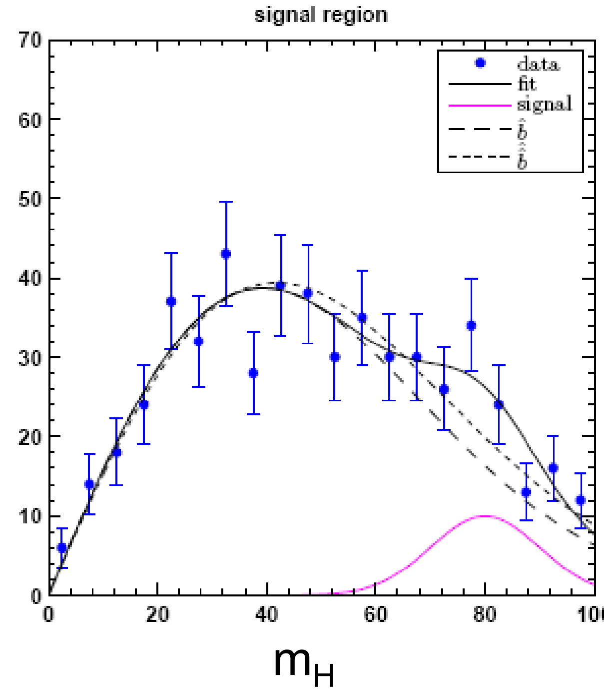
M=20



M=50



M=80



Profile Likelihood Ratio

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

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

$$q_1=-2\ln\lambda(\mu=1)$$

- q_1 distributes as a χ^2 under **$s(m_H)$+b** experiments (H_1)

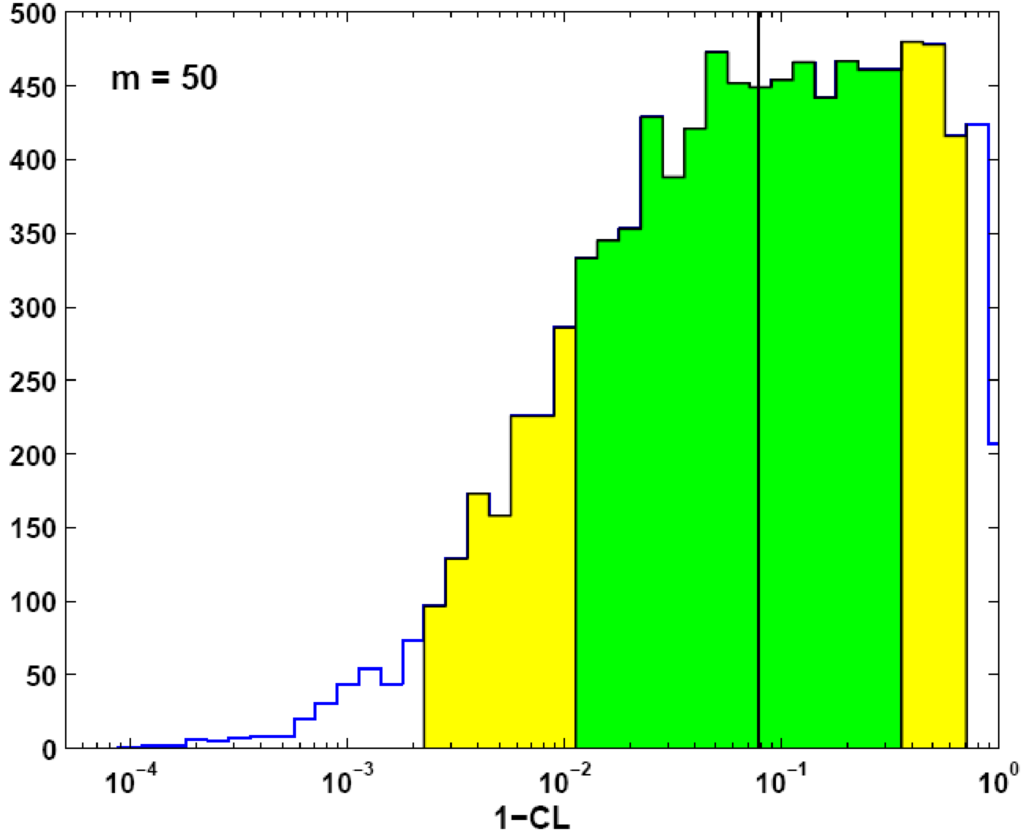
- The exclusion significance

$$Z=\sqrt{q_1}=\sqrt{-2\ln\lambda(1)}$$

can be expressed in terms of an equivalent exclusion CL

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

- The exclusion sensitivity is the median CL, and using toy MCs one can find the 1 and 2 σ bands



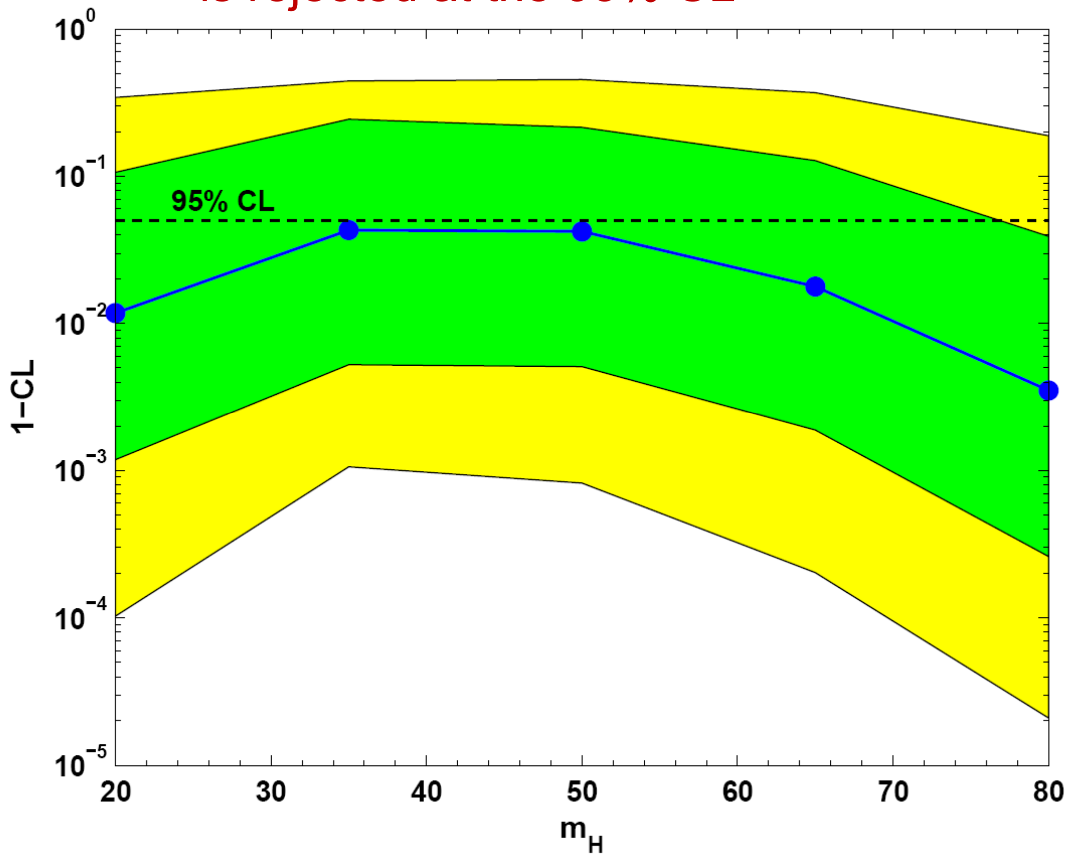
Exclusion Profile Likelihood Ratio

- 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

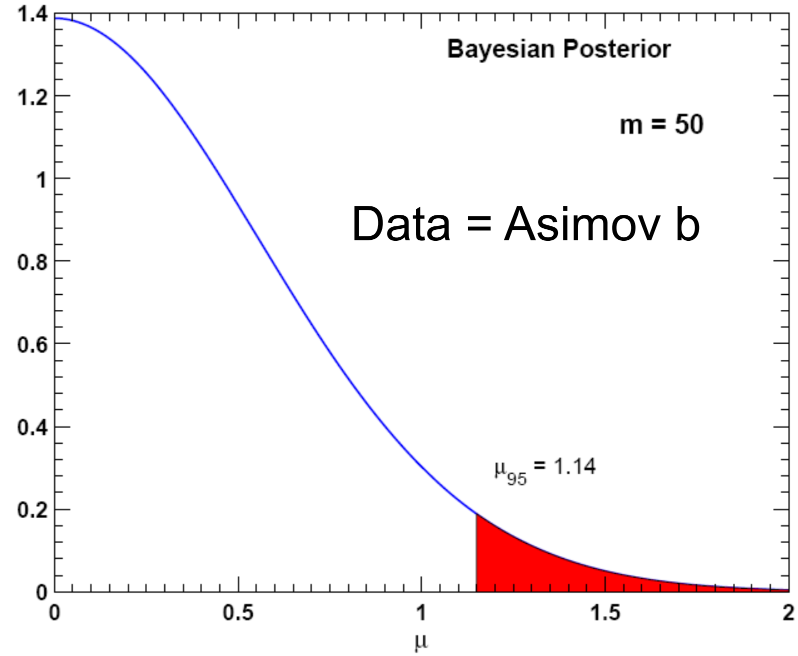
Let $prob(\mu | n,m)$ be the posterior for μ

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

NOTE: The pdf of the posterior is based on the **one** observed data event with the likelihood integrated over the nuisance parameters

To set an upper limit on the signal strength $\mu = \frac{\sigma}{\sigma_{SM}}$ calculate the credibility interval $[0, \mu_{95}]$

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

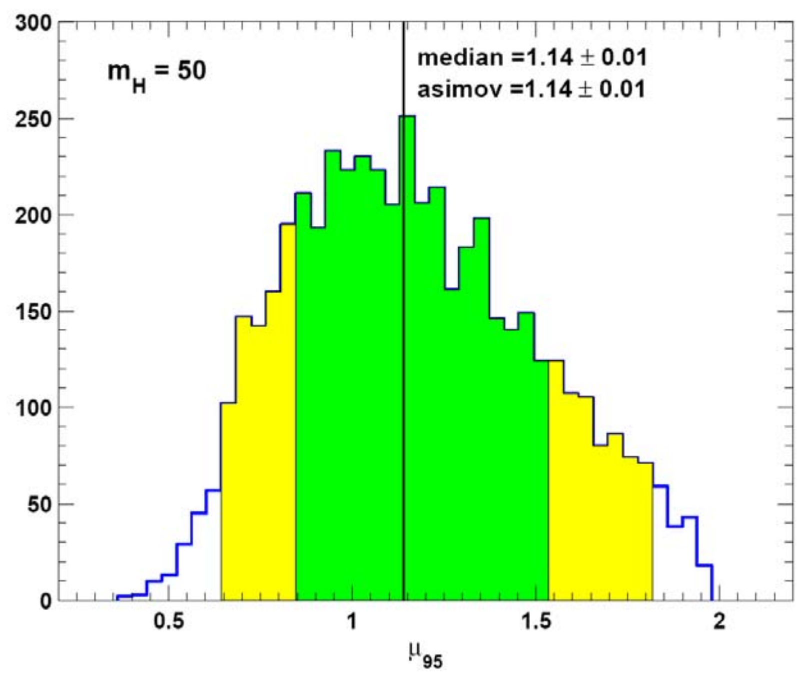


Exclusion Bayesian

Let $prob(\mu \mid n, m)$ be the posterior for μ

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

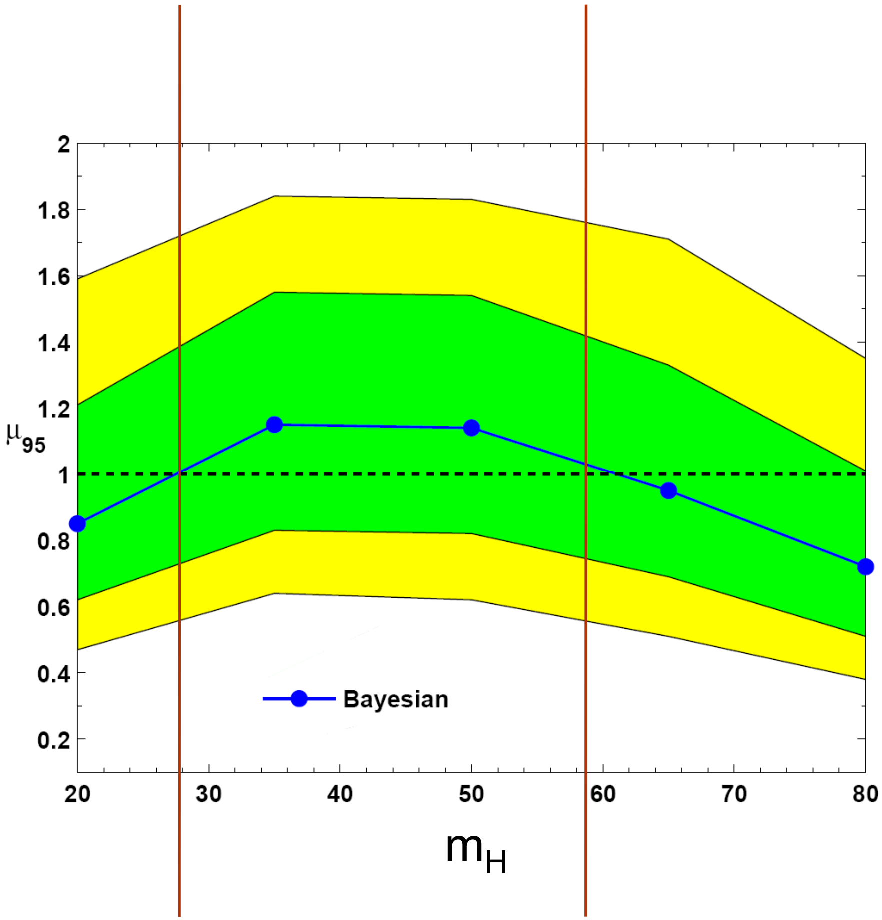
Data = b-only



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

Exclusion Bayesian

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



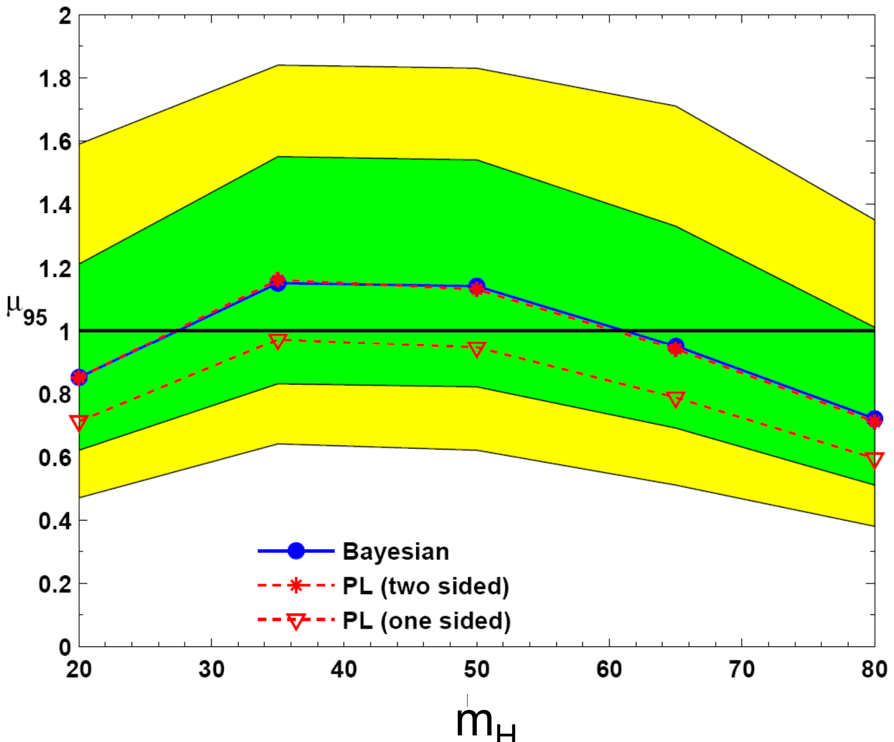
Exclusion Bayesian vs PL Ratio

- Comparing a credibility Bayesian interval to 95% frequentist CL is like comparing **oranges** to **apples**…. yet
- The saddle point approximation

$$prob(\mu \mid x) \sim \lambda(\mu) \sim e^{-\frac{\mu^2}{2\sigma^2}}$$

ensures that the Profile Likelihood Ratio and the marginalized Bayes limits are equivalent in this flat priors example

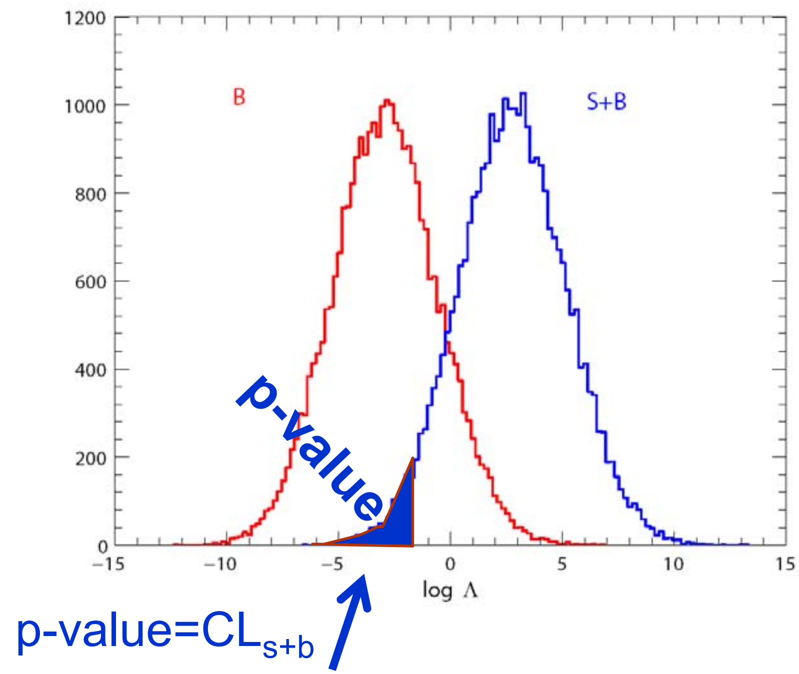
- See appendix for the proof



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

The frequentist CL_{s+b} method

- Use the LR as a test statistics $$\Lambda = \frac{L(H_1)}{L(H_0)} = \frac{L\left(n,m \mid s+b(\theta)\right)}{L\left(n,m \mid b(\theta)\right)}$$
- To take systematics unto account integrate the nuisance parameters or profile them
- The exclusion is given by the $s(m_H)+b$ hypothesis p-value $p_{s+b}=CL_{s+b}$
- If $p_{s+b}<5\%$, the $s(m_H)+b$ hypothesis is rejected at the 95% CL



The modified frequentist CLs

- CL_{s+b} enables the exclusion of the $s(m_H)+b$ hypothesis, 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 flustuations, the CL was redefined in a non-frequentist way to be

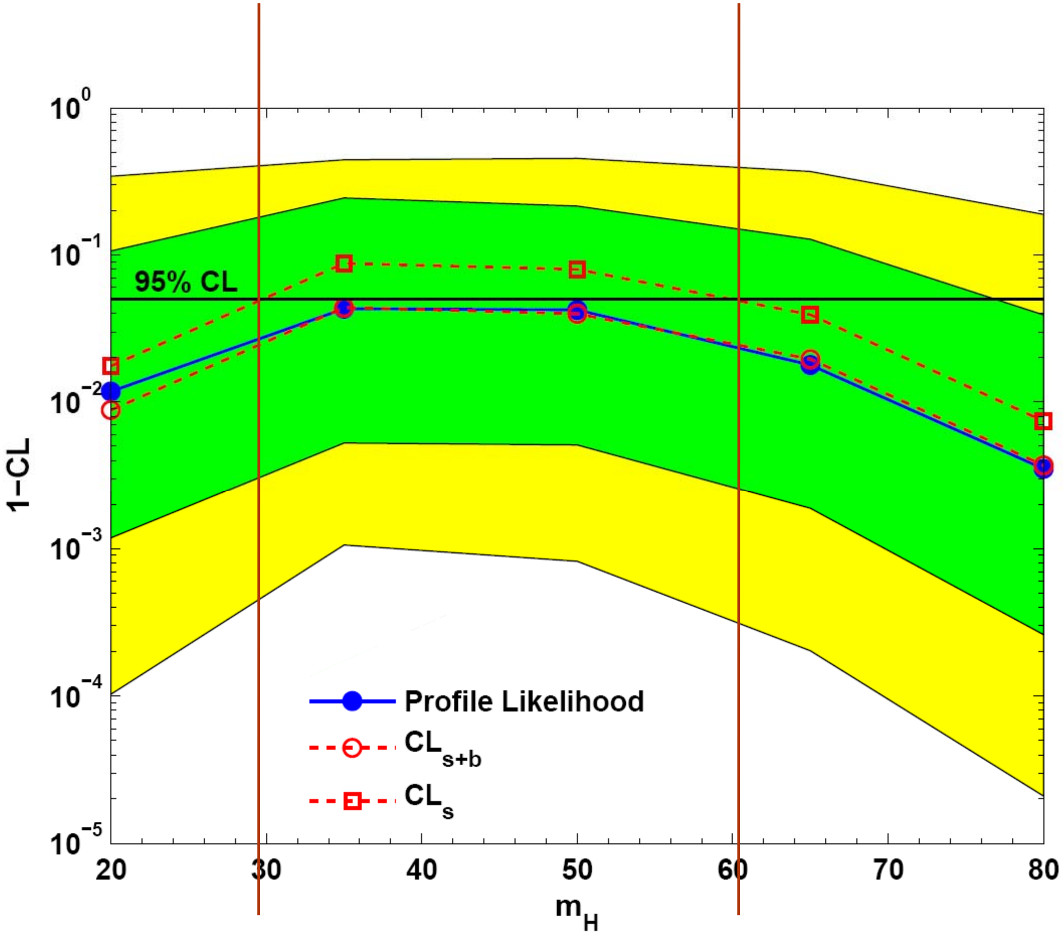
$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{p_{s+b}}{1-p_b} \sim \frac{p(n_{obs} \le s+b)}{p(n_{obs} \le b)}$$

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

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

The modified frequentist CLs

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



Conclusions

- We have explored and compared all the methods to test hypotheses that are currently in use in the High Energy Physics market (PLR, CL_{s+b}, CL_s, Bayesian)

- We have shown a way to appreciate the Bayes factor by comparing it to the PLR

- 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

- Even though we have used typical case studies, real life might be different and all available methods should be explored

APPENDIX & BACKUP

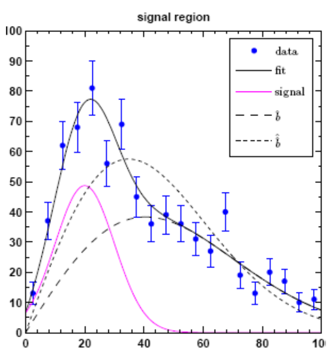
Discovery vs Exclusion

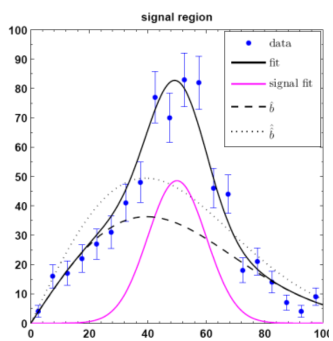
- Higgs statistics is about testing one hypothesis against another hypothesis
  - One hypothesis is the Standard Model with no Higgs Boson (H_0)
  - Another hypothesis is the SM with a Higgs boson with a specific mass m_H (H_1)
- Rejecting the No-Higgs (H0) hypothesis →DISCOVERY
- Rejecting the Higgs hypothesis (H_1) →EXCLUDING the Higgs
- **The null hypothesis** is the hypothesis we try to reject in favor of **the alternate hypothesis**.
  - DISCOVERY: Null=H_0, Alternate=H_1
  - EXCLUSION: Null=H_1, Alternate=H_0

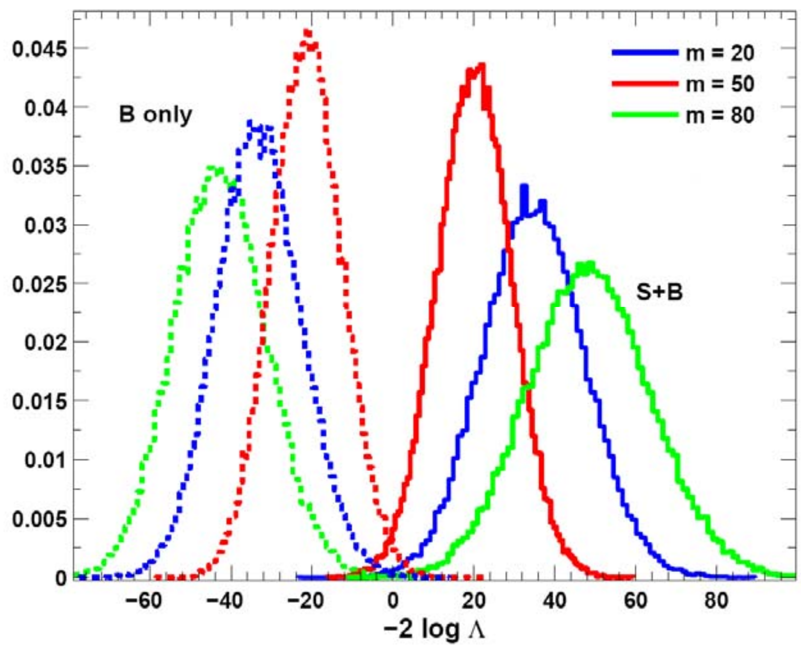
The Profiled CL way

- The signal close to the end of the background mass spectrum is better separated than the signal in the middle

$$\Lambda_{PL} = \frac{L(H_1)}{L(H_0)} = \frac{L\left(n,m \mid s+b(\hat{\hat{\theta}}_{s+b})\right)}{L\left(n,m \mid b(\hat{\theta}_b)\right)}$$







PL vs Bayesian

$$B_{10}(\mu) = \frac{\int L(\mu \cdot s + b)\pi(\mu)\pi(b)db}{\int L((\mu = 0) \cdot s + b)\pi(b)\,db}$$

$$= \frac{\int e^{\log L(\mu \cdot s + b)\pi(\mu)\pi(b)} db}{\int e^{\log L((\mu=0)\cdot s+b)\pi(b)} db}$$

$$\approx \frac{e^{\log L\left(\mu \cdot s + \hat{\hat{b}}(\mu)\right)}}{e^{\log L\left((\mu=0)\cdot s + \hat{\hat{b}}(0)\right)}} = \frac{\lambda(\mu)}{\lambda(0)}$$

↓

saddle-point approximation (for flat priors)

http://en.wikipedia.org/wiki/Method_of_steepest_descent

PL vs Bayesian

$$B_{10}(\mu) = \frac{\int L(\mu \cdot s + b)\pi(\mu)\pi(b)db}{\int L((\mu = 0) \cdot s + b)\pi(b)\,db}$$

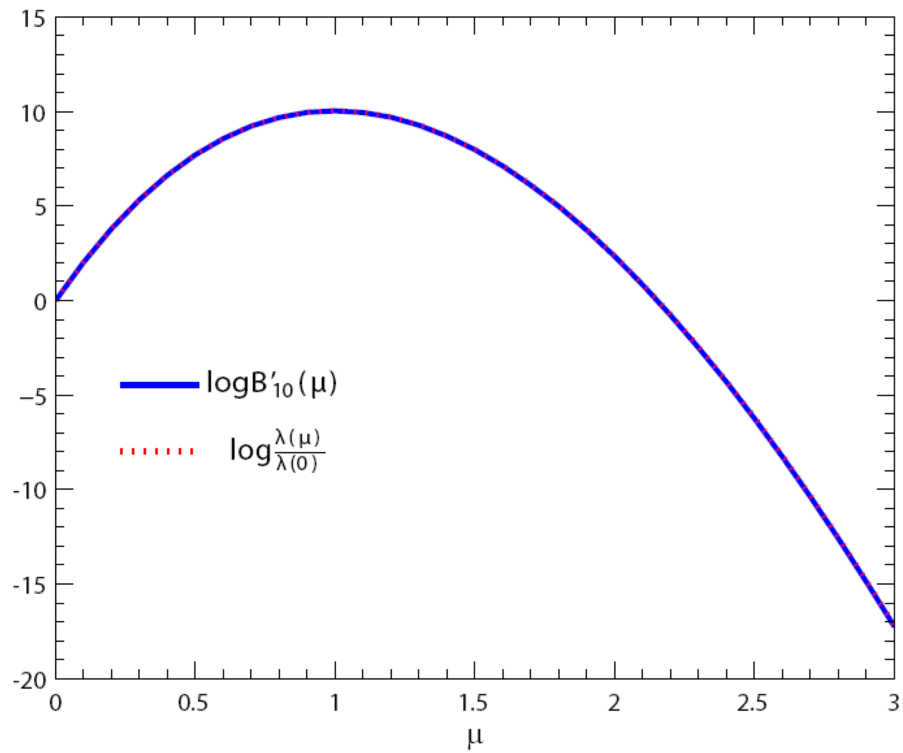
$$= \frac{\int e^{\log L(\mu \cdot s + b)\pi(\mu)\pi(b)} db}{\int e^{\log L((\mu=0)\cdot s + b)\pi(b)} db}$$

$$\approx \frac{e^{\log L\left(\mu \cdot s + \hat{\hat{b}}(\mu)\right)}}{e^{\log L\left((\mu=0)\cdot s + \hat{\hat{b}}(0)\right)}} = \frac{\lambda(\mu)}{\lambda(0)}$$

↓

saddle-point approximation

http://en.wikipedia.org/wiki/Method_of_steepest_descent



Asimov Data = S+B

PL vs Bayesian ?

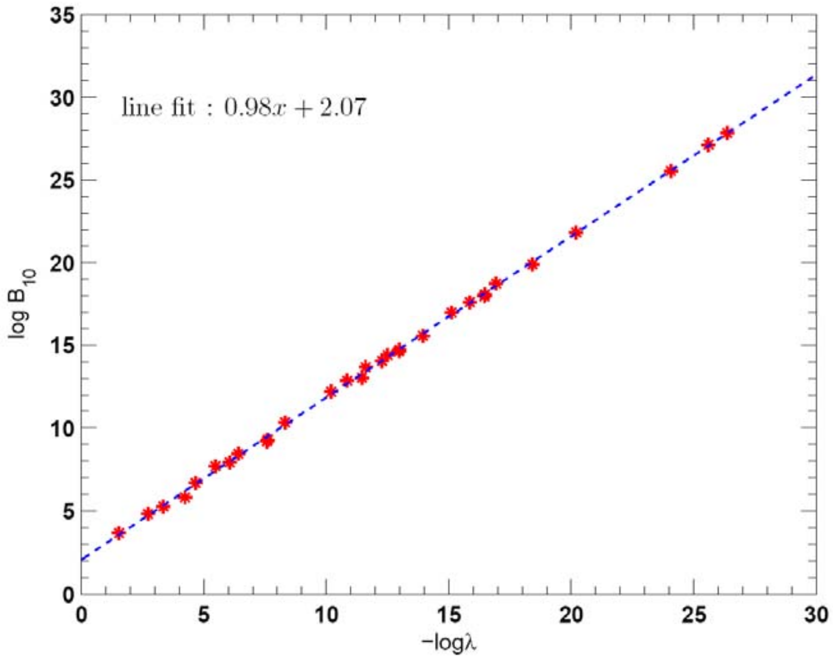
$$B_{10}(\mu) \approx \frac{\lambda(\mu)}{\lambda(0)}$$

$$B_{10} = \int B_{10}(\mu) d\mu$$

$$B_{10} \approx \frac{1}{\lambda(0)} \int \lambda(\mu) d\mu$$

$$\log B_{10} \approx -\log \lambda(0) + \log \int \lambda(\mu) d\mu$$

$$\log B_{10} \approx -\log \lambda(0) + C$$



PL vs Bayesian ?

$$B'_{10}(\mu) \approx \frac{\lambda(\mu)}{\lambda(0)}$$

$$B_{10} = \int B'_{10}(\mu) d\mu$$

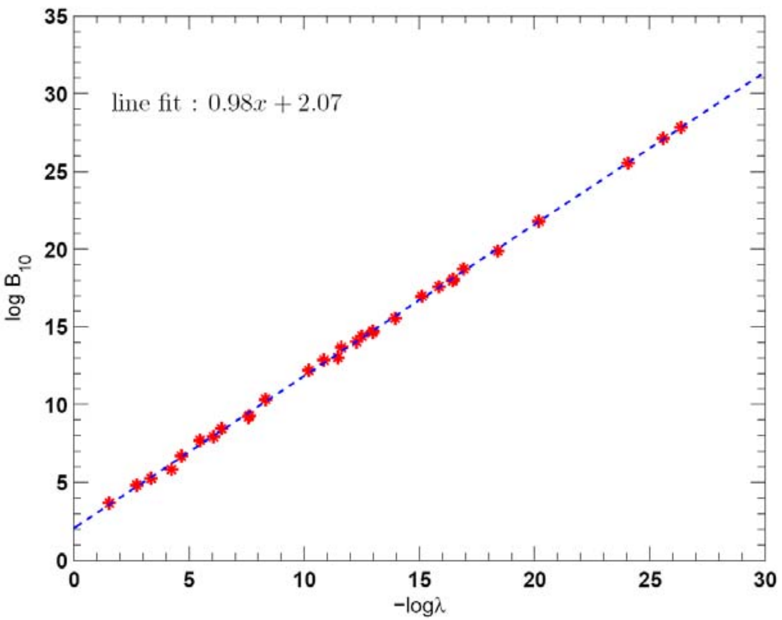
$$B_{10} \approx \frac{1}{\lambda(0)} \int \lambda(\mu) d\mu$$

$$\log B_{10} \approx -\log \lambda(0) + \log \int \lambda(\mu) d\mu$$

$$\log B_{10} \approx -\log \lambda(0) + C$$

$$Z = \sqrt{-2 \log \lambda(0)} \Rightarrow \frac{Z^2}{2} = -\log \lambda(0)$$

$$\log B_{10} \approx \frac{Z^2}{2} + C$$



Comparing Bayesian to Frequentist PL

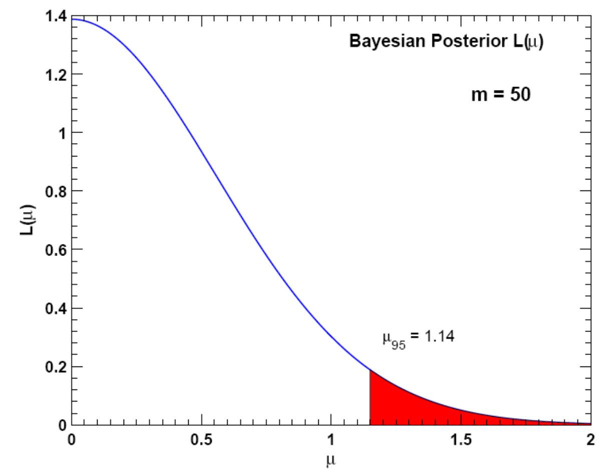
$$prob(\mu | \vec{n},\vec{m}) = \frac{\int L\big(n,m \,|\, \mu\cdot s + b(\theta)\big)\pi(\mu)\pi(\theta)d\theta}{\iint L\big(n,m \,|\, \mu\cdot s + b(\theta)\big)\pi(\mu)\pi(\theta)\,d\theta d\mu} =$$

saddle-point approximation (for flat priors)

$$= \frac{e^{\log L(\mu s + \hat{\hat{b}})}}{e^{\log L(\hat{\mu} s + \hat{b})}} = e^{\log \frac{L(\mu s + \hat{\hat{b}})}{L(\hat{\mu} s + \hat{b})}} = \lambda(\mu)$$

$$-2\log\lambda(\mu) \sim \frac{\mu^2}{\sigma^2}$$

$$prob(\mu | x) \sim \lambda(\mu) \sim e^{-\frac{\mu^2}{2\sigma^2}}$$



Note: 95% Credibility Interval ~Gaussian 1.96σ

The 95% CL Profile Likelihood limit is based on 1.64σ (2-sideded 10%~1 Sided 5%)

When comparing the expected limits

(though it is like comparing oranges to apples)

one has to be careful about the definitions

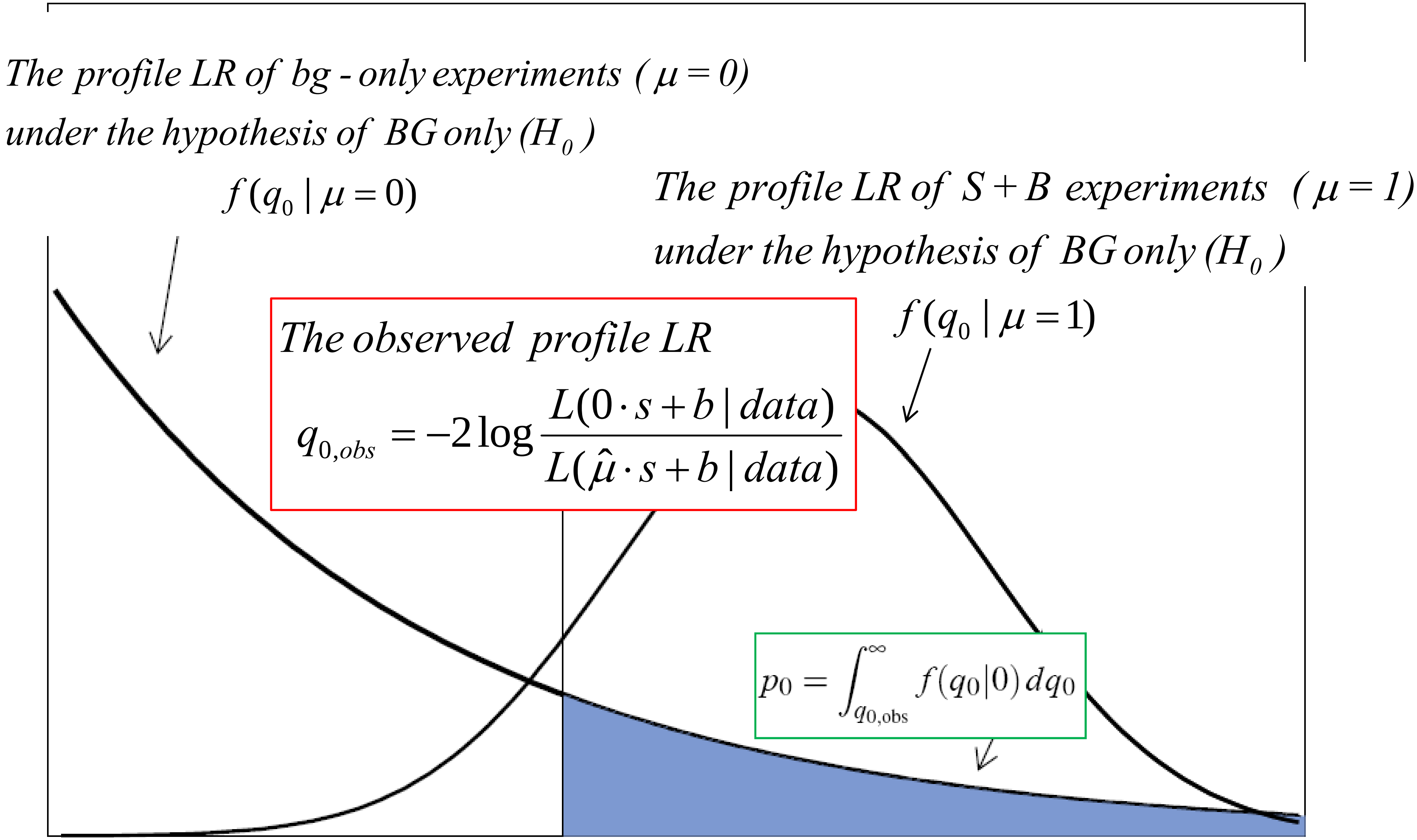
Wilks Theorem

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

$$\lambda(\mu=0)=\frac{L(0\cdot s(m_H)+\hat{\hat{b}}\,|\,x)}{L(\hat{\mu}\cdot s(m_H)+\hat{b}\,|\,x)}$$

- Under a set of regularity conditions and for a sufficiently large data sample, Wilks' theorem says that for a hypothesized value of $\mu=0$, the pdf of the statistic $q_0=-2\ln\lambda\ (\mu=0)$ approaches the chi-square pdf for one degree of freedom

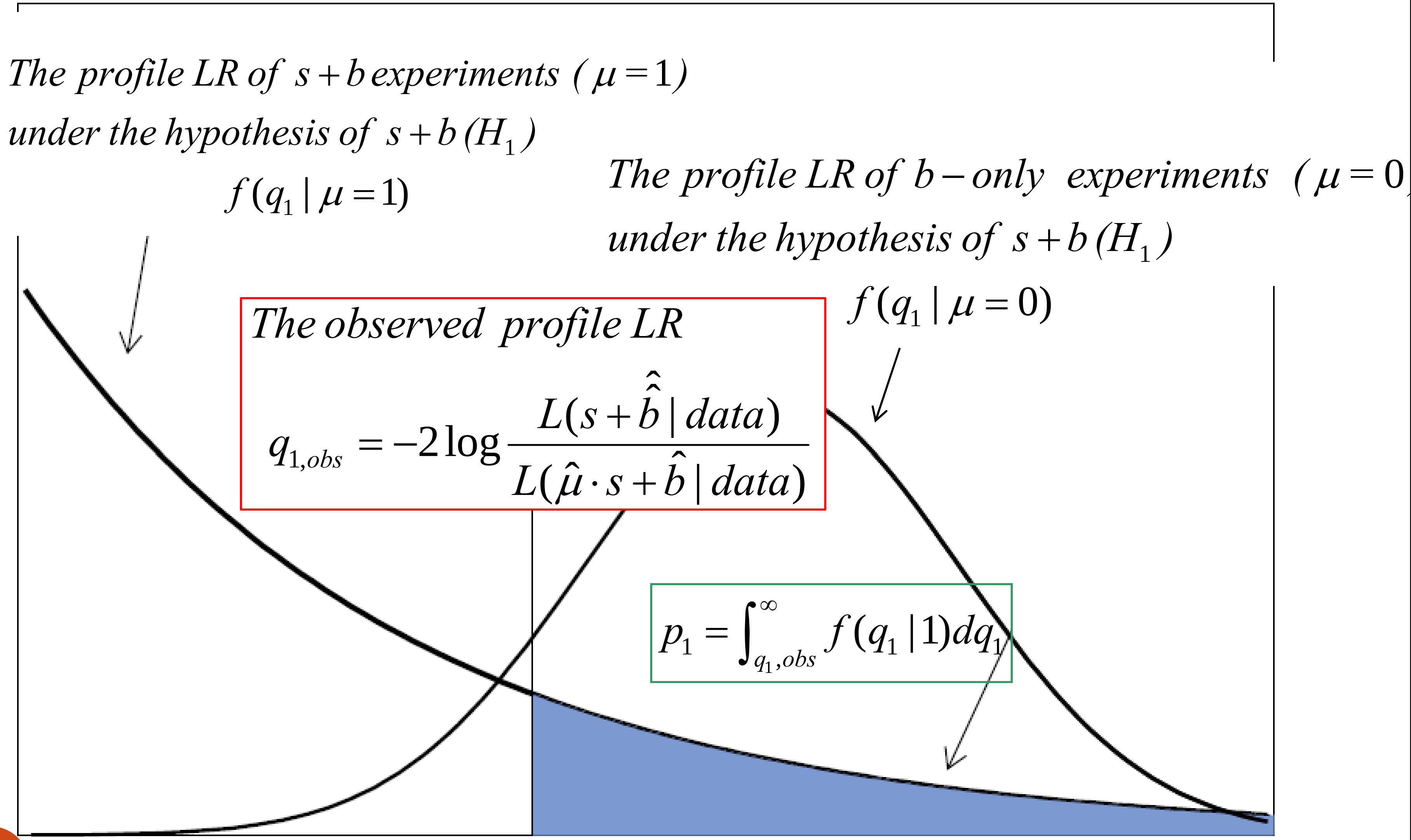
PL Discovery - Illustrated



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 5σ discovery

PL Exclusion - Illustrated



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