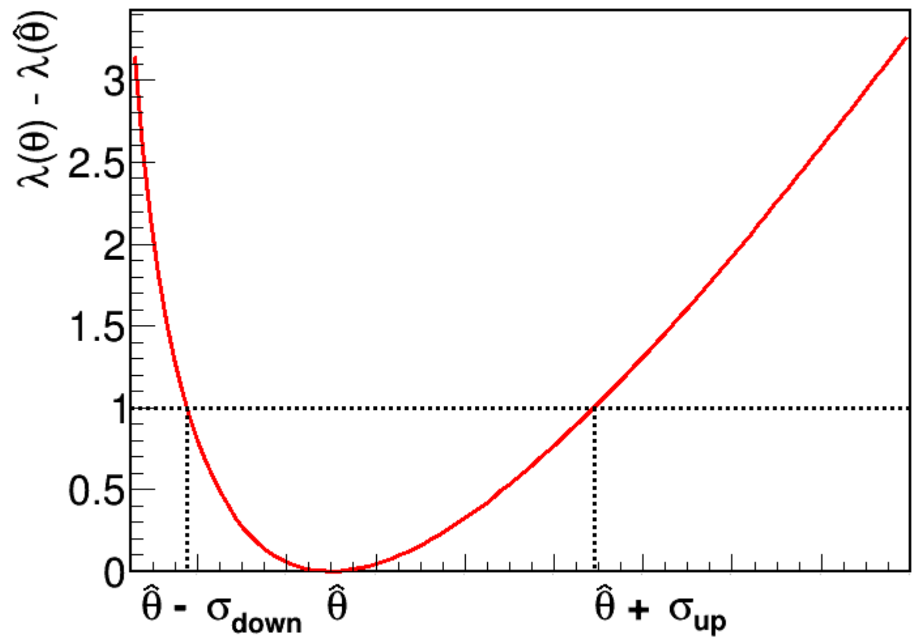
Statistics for HEP, part II

Nicolas Berger (LAPP Annecy)

What we learned so far (previous lecture)

Estimating a parameter

- Build a likelihood **L(θ)** for the measurement



- Compute $\boldsymbol{\lambda(\theta) = -2 \log L_{data}(\theta)}$, as a function of θ.

- Find the minimum of $\lambda(\theta)$
  ⇒ **Minimum is reached for $\hat{\theta}$.**

- Move the parameter up and down to get
  $\lambda(\hat{\theta}+\sigma_{up}) = \lambda(\hat{\theta}) + 1$ and $\lambda(\hat{\theta}-\sigma_{down}) = \lambda(\hat{\theta}) + 1$.
  Then $[\theta-\sigma_{down}, \theta+\sigma_{up}]$ is a **68% confidence interval** for θ : $\boxed{\theta = \hat{\theta}^{+\sigma_{up}}_{-\sigma_{down}}}$

Relation with χ^2

- χ^2: say you measure $\hat{\theta}_1 \ldots \hat{\theta}_n$ with reference values $\theta^*_1 \ldots \theta^*_1$, an uncertainty σ. Then

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{\hat{\theta}_i - \theta_i^*}{\sigma} \right)^2$$

- If good agreement : $\chi^2/n \sim 1$.
- If $\hat{\theta}_i$ are Gaussian (with same θ^*_i and σ as in the χ^2 expression), then χ^2 follows a **χ^2 distribution with n degrees of freedom**, **$\chi^2_n(x)$**
- Now go back to the likelihood picture, assume Gaussian measurements:

$$L = \prod_{i=1}^{n} e^{-\frac{1}{2}\left(\frac{\hat{\theta}_i - \theta_i^*}{\sigma}\right)^2} \qquad \lambda = -2 \log L = \sum_{i=1}^{n} \left(\frac{\hat{\theta}_i - \theta_i^*}{\sigma} \right)^2$$

- So
  - **λ is like a χ^2**
  - L is $\exp(-\chi^2/2)$
  - λ is $\sim \chi^2_n$. Quantiles :
    - for **n=1**, **same as Gaussian**
    - For n>1, look up the values...

N_{sigmas}	χ^2_1	χ^2_2	$1-\alpha$
1	**1**	**2.30**	**0.68**
1.645	**2.71**	**4.61**	**0.90**
1.96	**3.84**	**6.00**	**0.95**

This Lecture

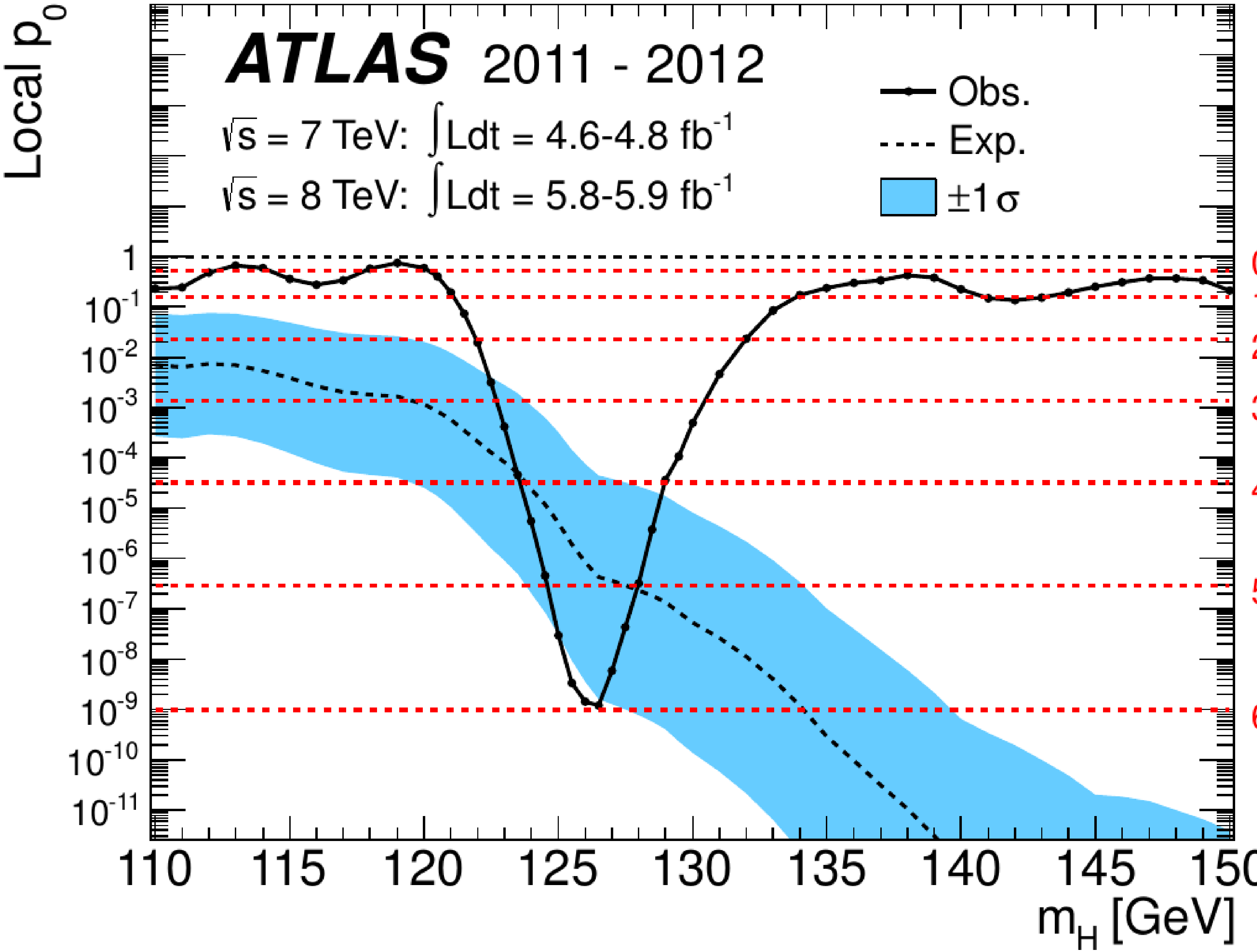
- **“Hypothesis testing”** : determining if a statement is true or not.

- We focus on 2 particular statements:

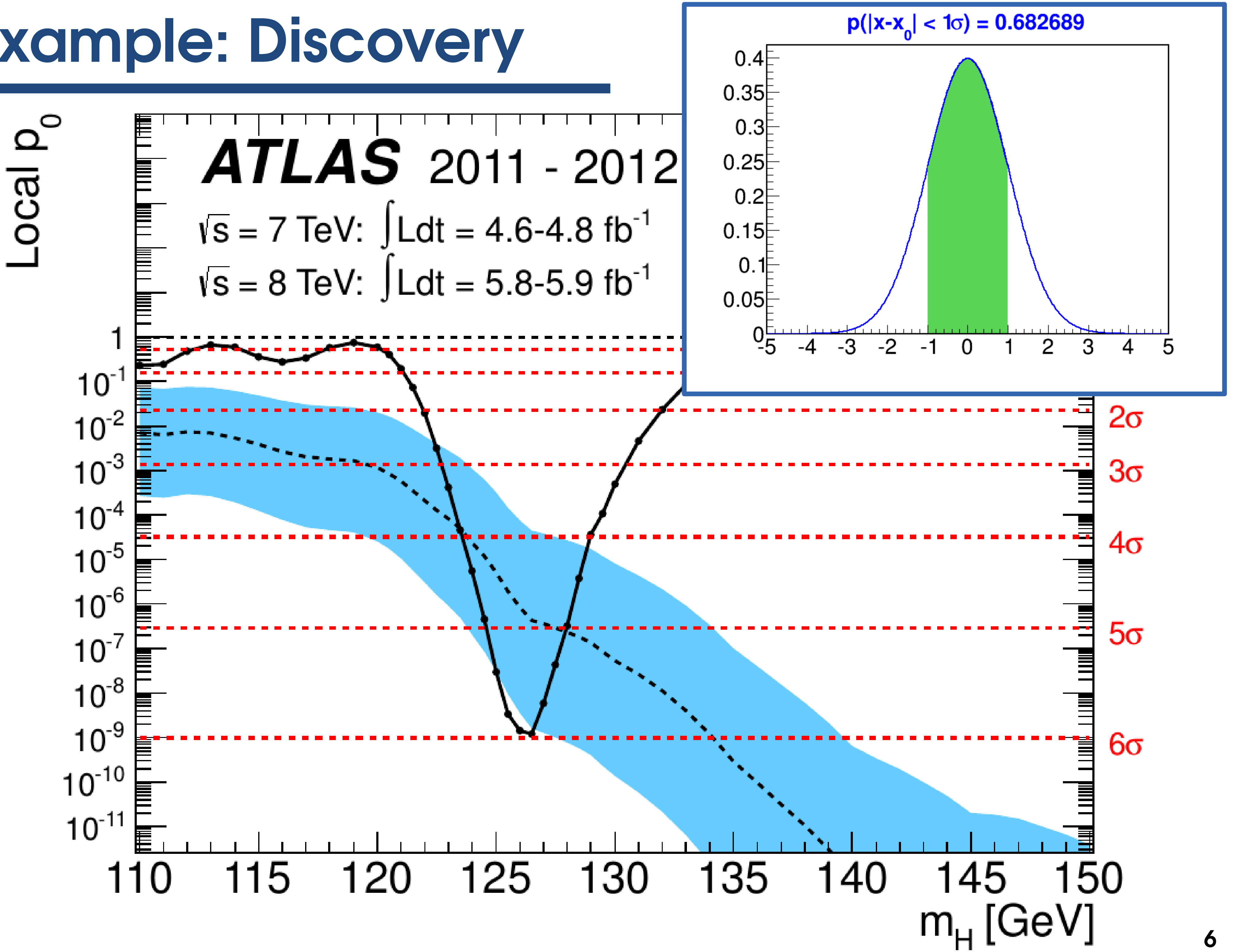
  – **Discovery**
    - Does a new particle exist ?
    - Do we have evidence for a new process ?
    - Useful when we saw something unusual

  – **Upper limits**
    - Useful if we didn't see anything
    - Determine how small the signal must be

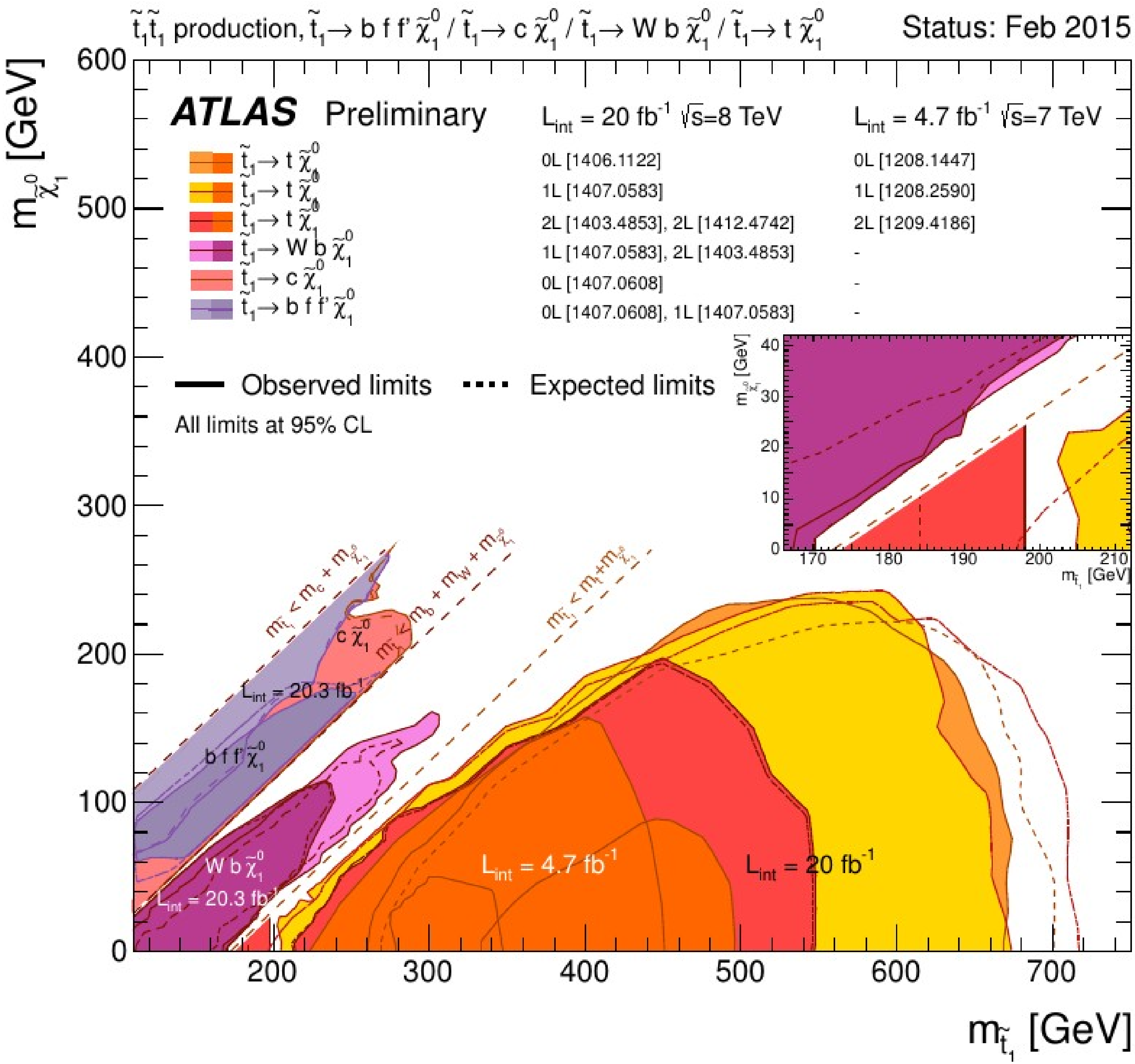
Example: Discovery



Example: Discovery



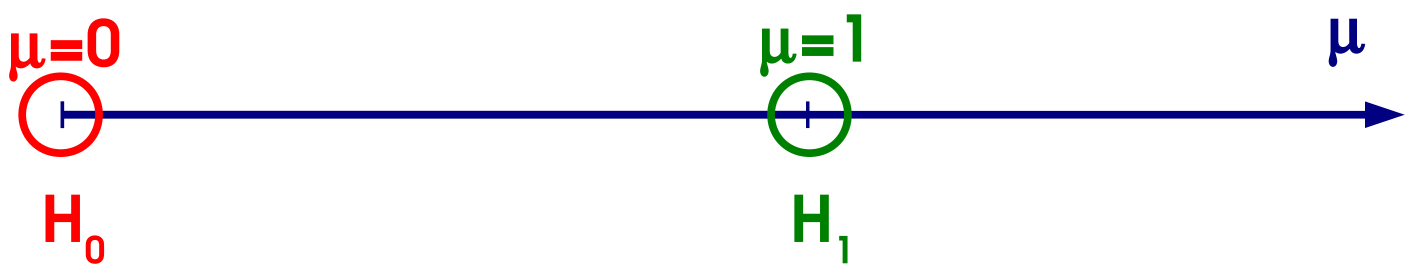
Example: Upper Limits



Hypothesis testing

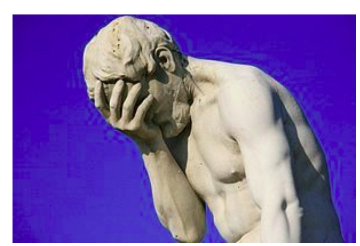
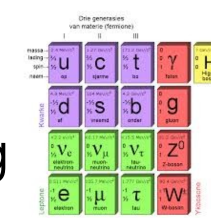
Hypothesis Testing: Setup

- We define 2 mutually exclusive "hypotheses" = regions of parameter space Usually:
  - H_0 ("**null**") = no signal (e.g. no Higgs, μ=0)
  - H_1 ("**alternative**") = **something new** (e.g. Higgs, μ=1)

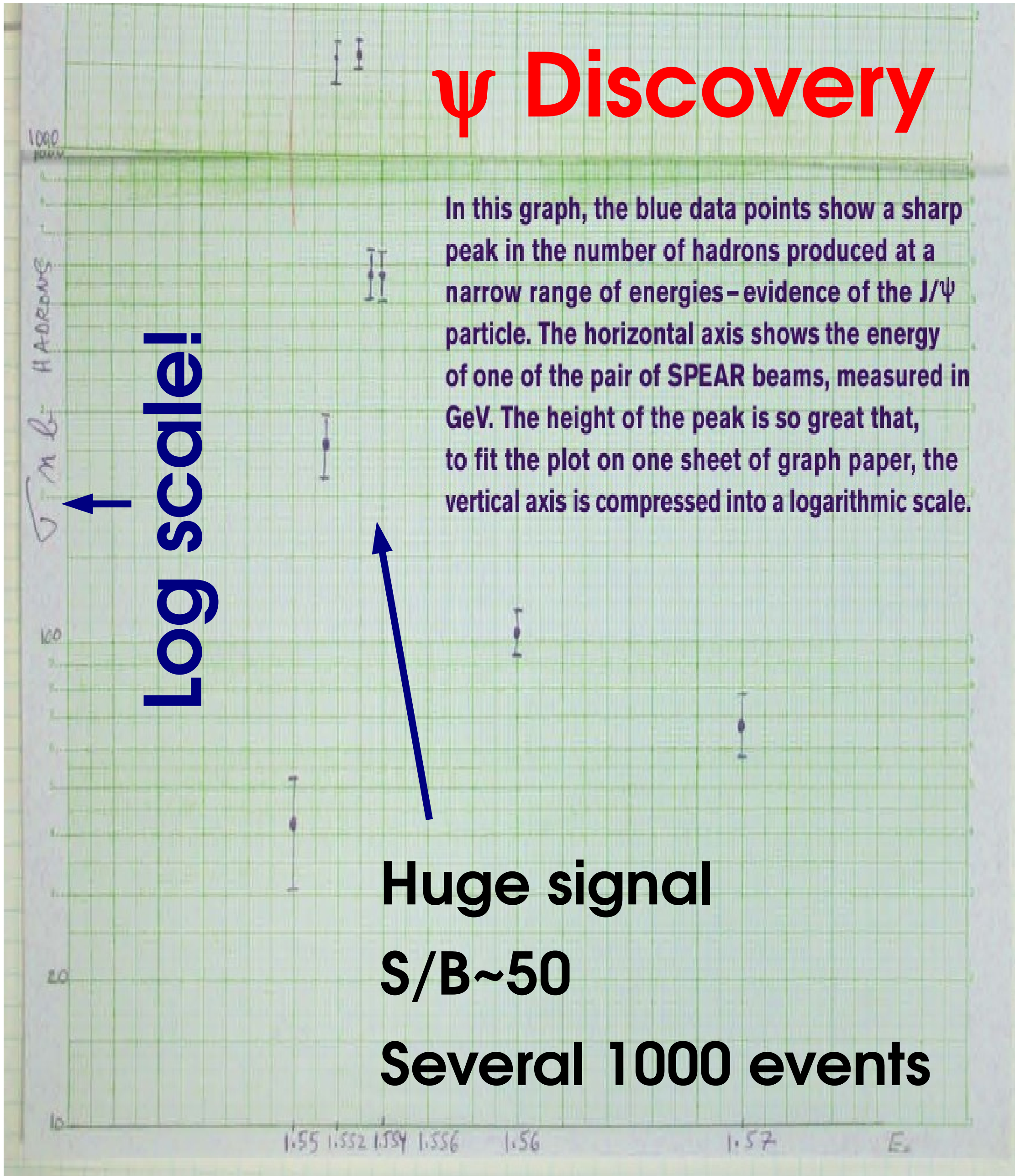


- We want to use data to make a statement on the hypotheses
  - **data favors H_0** : nothing new
  - **data favors H_1** : **discovery!**

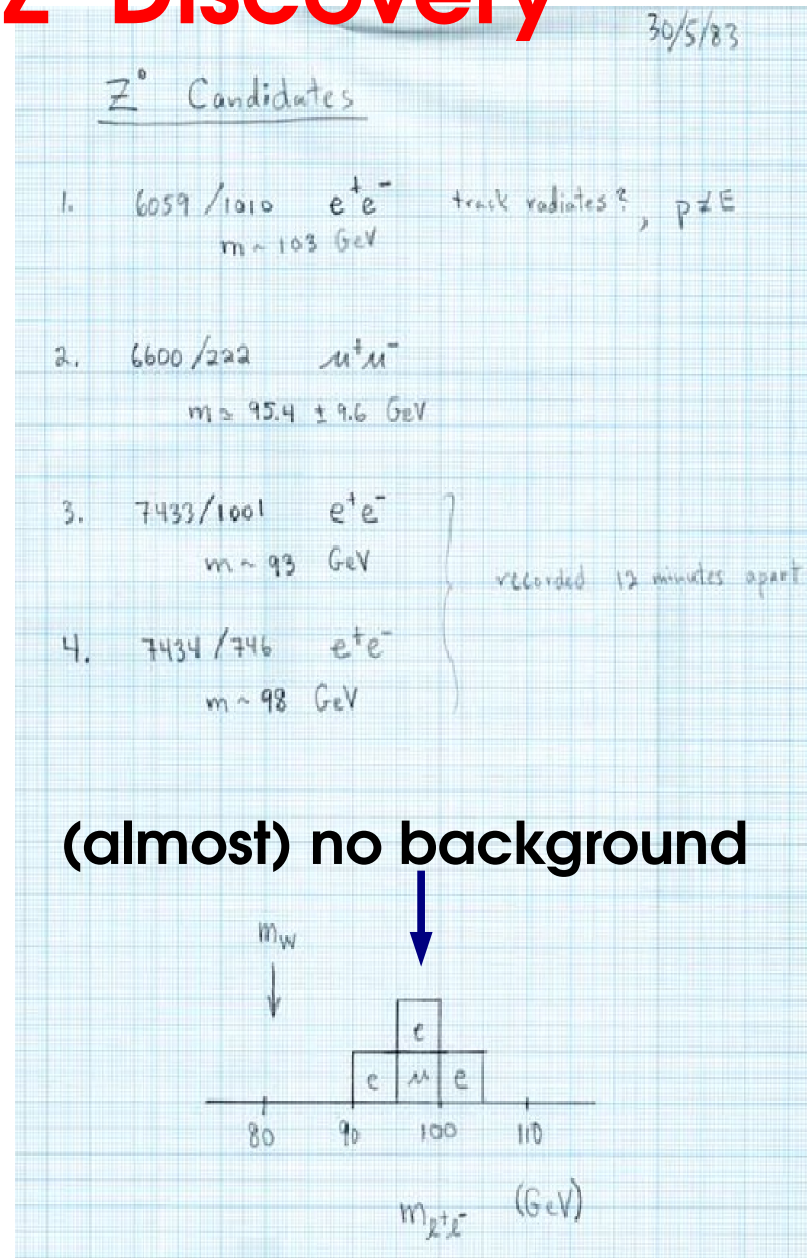
Possible Outcomes in Hypothesis Testing

	Data disfavors H_0 (Discovery claim)	Data favors H_0 (Nothing found)
H_0 is false (New physics!)	Discovery!  	Missed discovery 
H_0 is true (Nothing new)	False discovery claim 	No new physics, none found Not very exciting but correct 

Classic Discoveries (1)

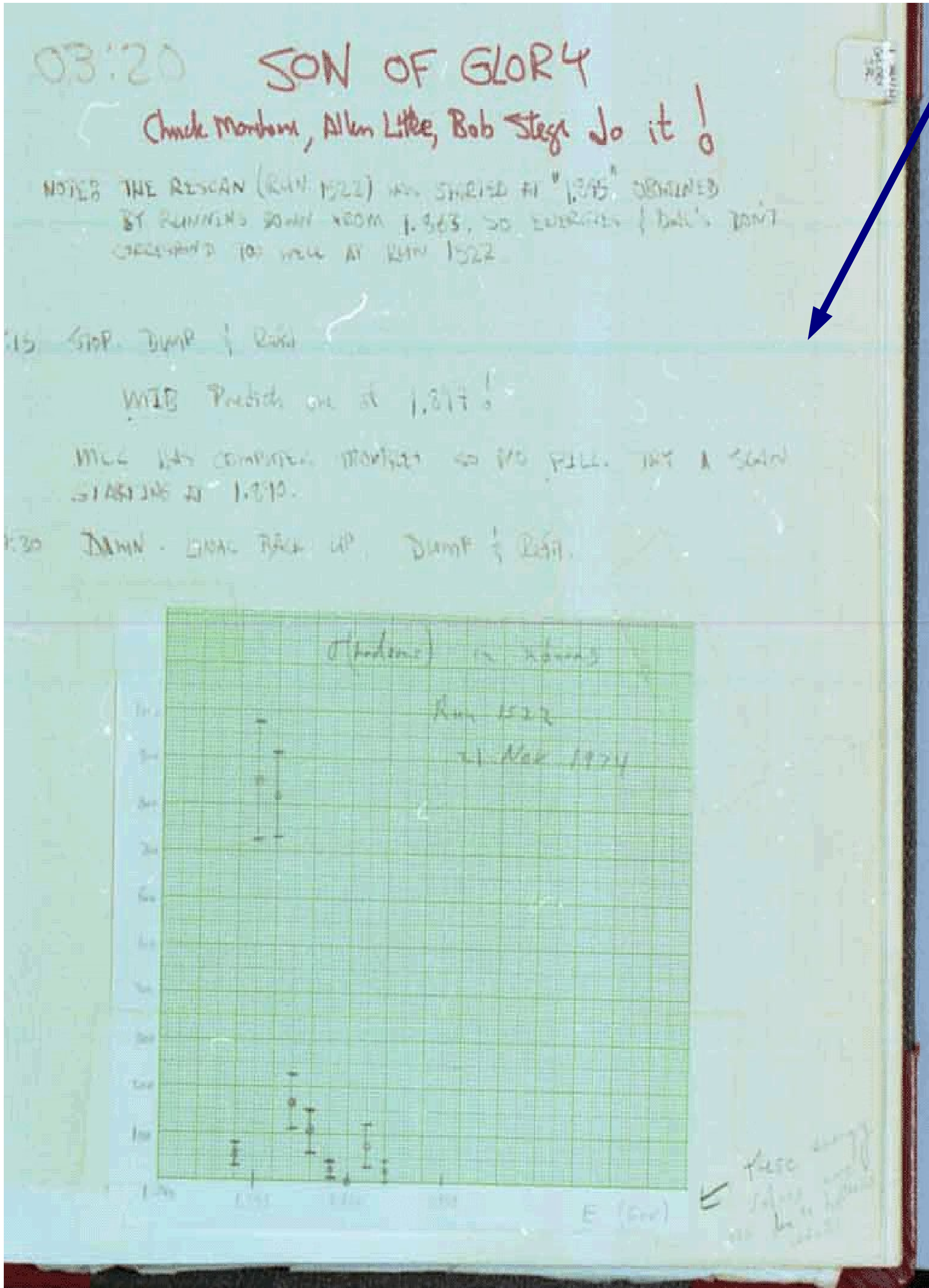


Z^0 Discovery

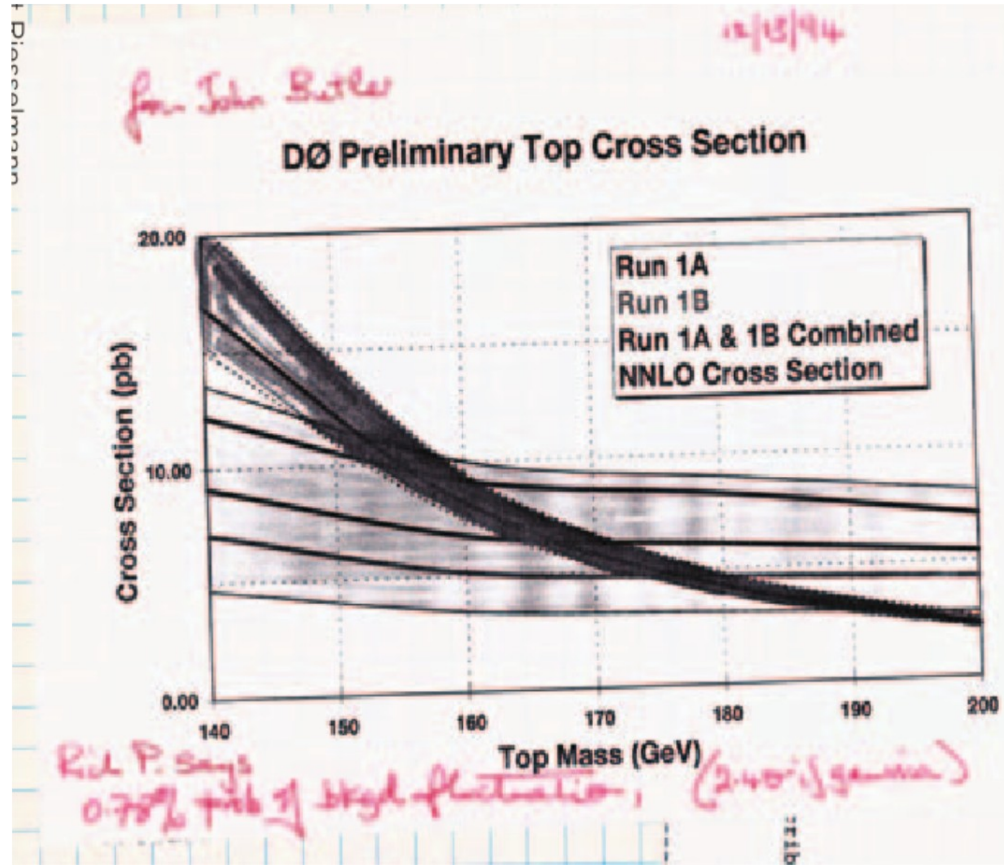


Logbook of J. Rohlf, 1983-05-30

Classic Discoveries (2)



ψ' : discovered online by the (lucky) shifters



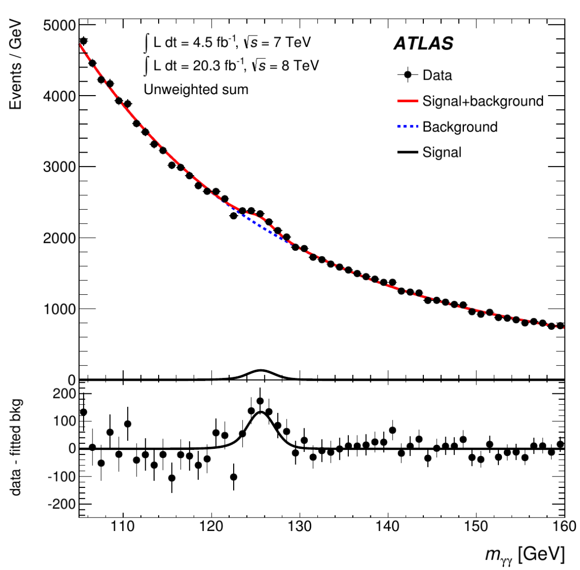
First hints of top at D0:

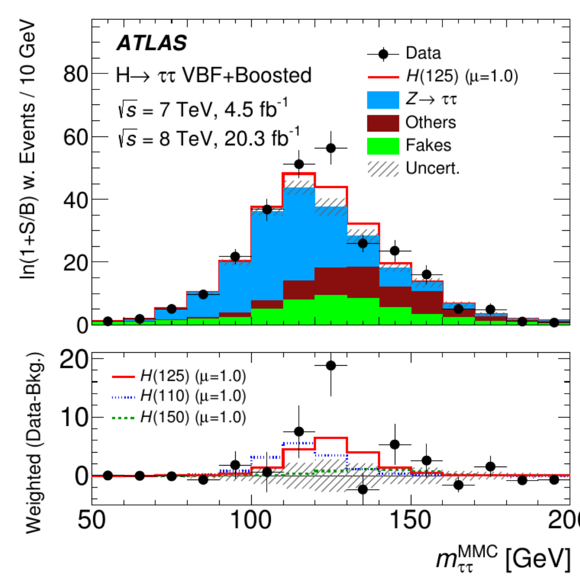
O(10) signal events,

a few bkg events, 2.4σ

Aside : Why we need Statistical Tools now

- The high-signal, low-background experiments have been done already (but a surprise at 13 TeV would be welcome...)
- At LHC:
  - **High background levels**, need precise modeling
  - **Large systematics,** need to be treated correctly
  - **Small signals**: need optimal use of available information :
    - Shape analyses instead of counting
    - Isolation of signal-enriched regions (categories)



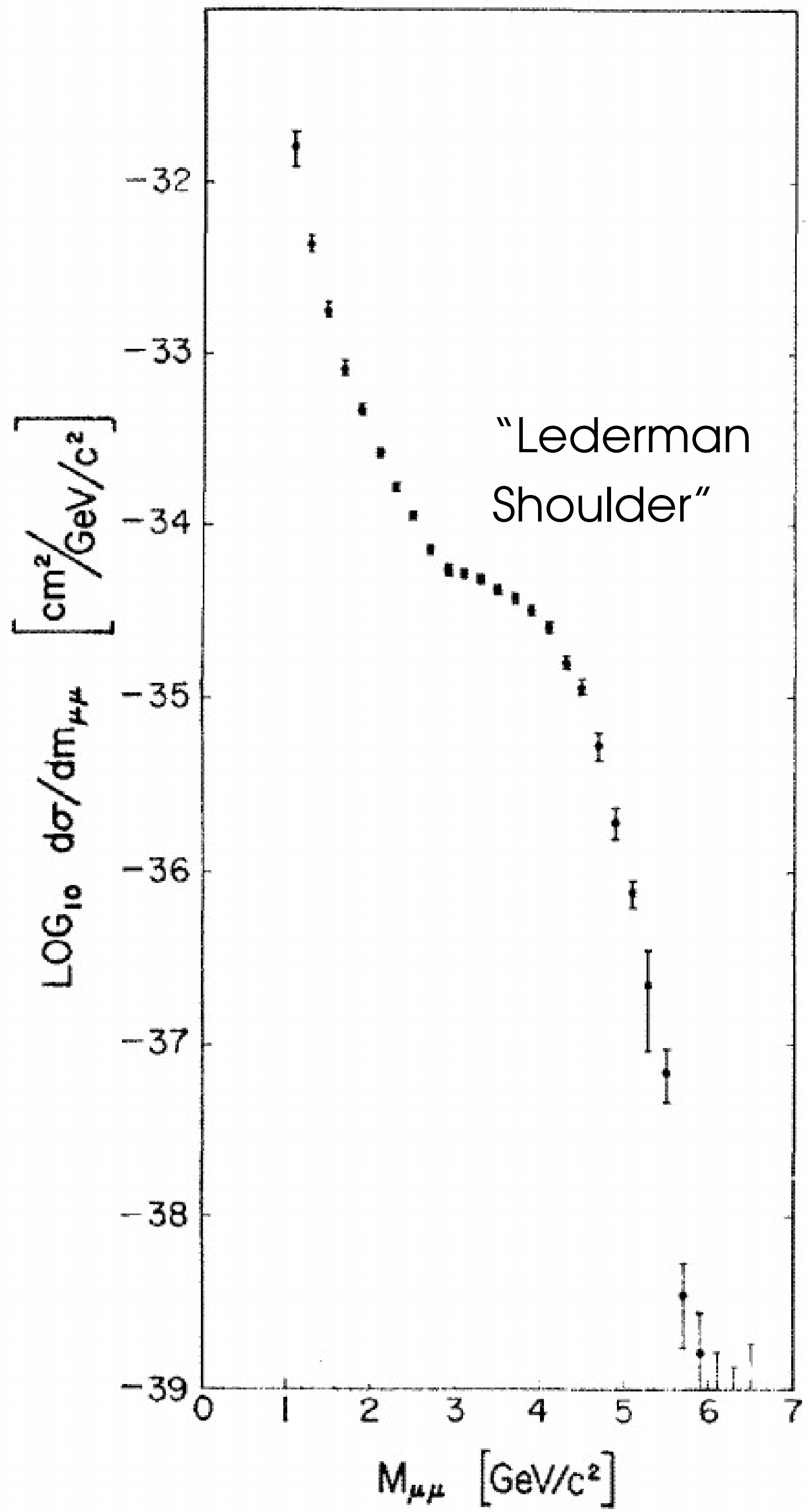


A Classic Miss

H. Christenson, L. Lederman, et al.,

PRL 25(21) 1523 (**1970**)

As seen both in the mass spectrum and the resultant cross section $d\sigma/dm$, there is no forcing evidence of any resonant structure. To obtain limits on the production probability of a vector meson at a given mass, a narrow resonance was introduced in the Monte Carlo production distribution and increased in amplitude until the resulting bump visibly distorted the output spectrum. This procedure properly introduced the single-particle mass resolution and efficiency into the analysis. The sensitivity to vector-meson production is somewhat impaired by the rapidly falling continuum upon which any structure must reside. Indeed, in the mass region near 3.5 GeV/c^2, the observed spectrum may be reproduced by a composite of a resonance and a steeper continuum. These considerations are

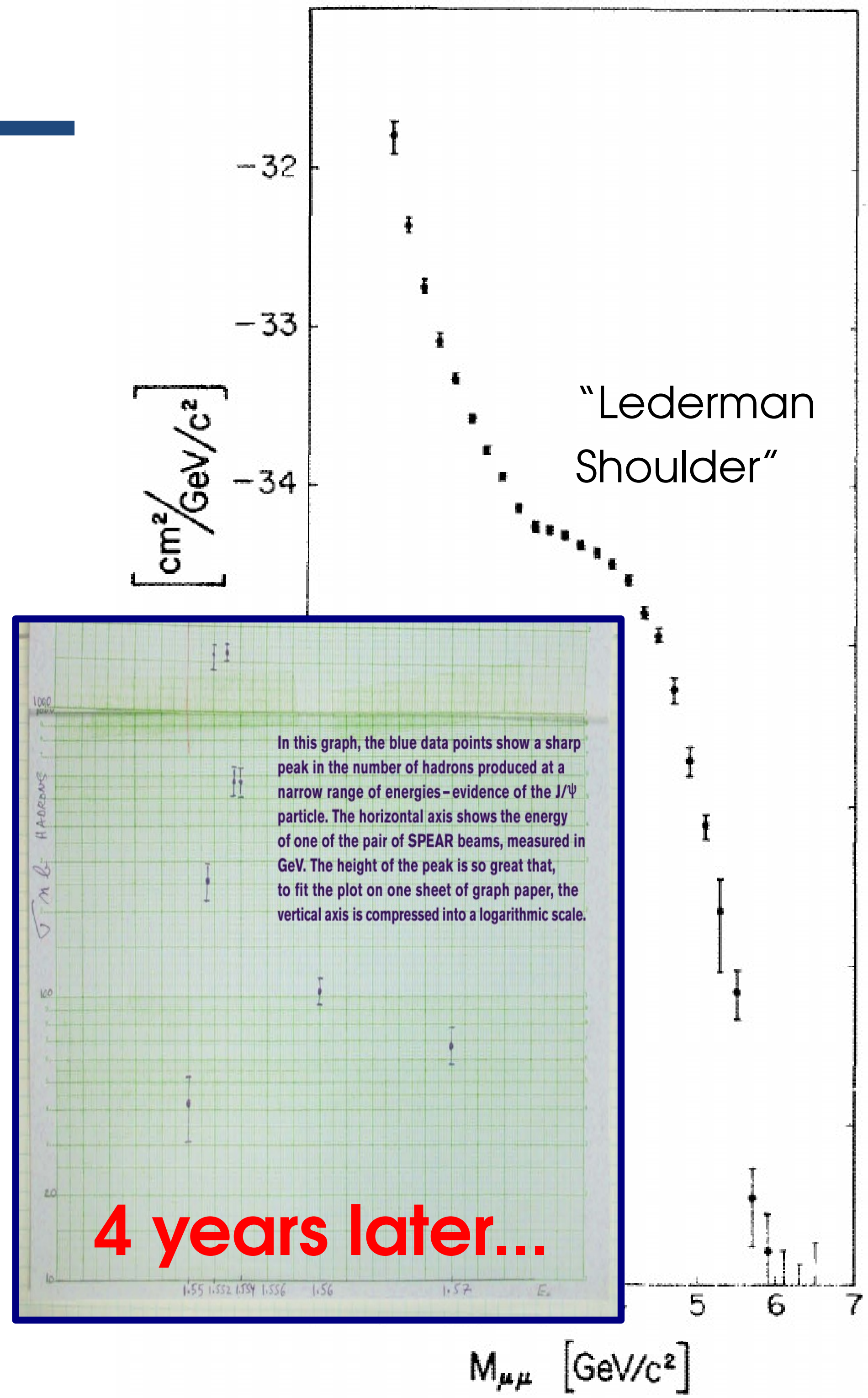


A Classic Miss

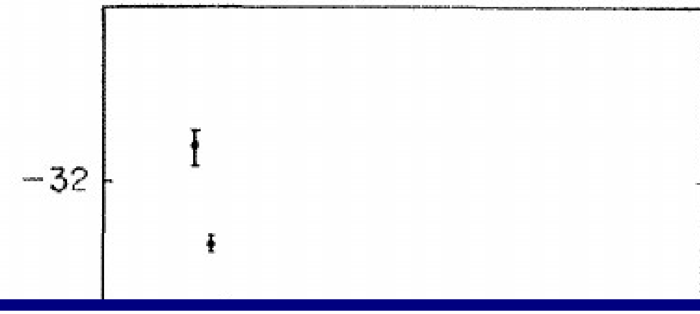
H. Christenson, L. Lederman, et al.,

PRL 25(21) 1523 (**1970**)

As seen both in the mass spectrum and the resultant cross section $d\sigma/dm$, there is no forcing evidence of any resonant structure. To obtain limits on the production probability of a vector meson at a given mass, a narrow resonance was introduced in the Monte Carlo production distribution and increased in amplitude until the resulting bump visibly distorted the output spectrum. This procedure properly introduced the single-particle mass resolution and efficiency into the analysis. The sensitivity to vector-meson production is somewhat impaired by the rapidly falling continuum upon which any structure must reside. Indeed, in the mass region near 3.5 GeV/c^2, the observed spectrum may be reproduced by a composite of a resonance and a steeper continuum. These considerations are



A Classic Miss



H. Christenson, L. Lederman, et al.,

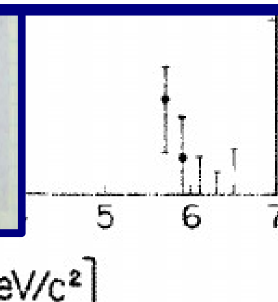
" I think it's important to emphasize that this story is one of missed opportunities, abysmal judgment, monumental blunders, stupid mistakes, and inoperative equipment.

(...)

This experiment, properly carried out, would have produced results that won five Nobel prizes!"

L. Lederman, "The Rise of the Standard Model"

steeper continuum. These considerations are



Discoveries that weren't

UA1 Monojets (1984)

Volume 139B, number 1,2 PHYSICS LETTERS 3 May 1984

EXPERIMENTAL OBSERVATION OF EVENTS WITH LARGE MISSING TRANSVERSE ENERGY ACCOMPANIED BY A JET OR A PHOTON (S) IN $p\bar{p}$ COLLISIONS AT $\sqrt{s}$ = 540 GeV

UA1 Collaboration, CERN, Geneva, Switzerland

At the present time we can only speculate about the origin of this new effect. The missing transverse energy can be due either to:

(i) One or more prompt neutrinos.

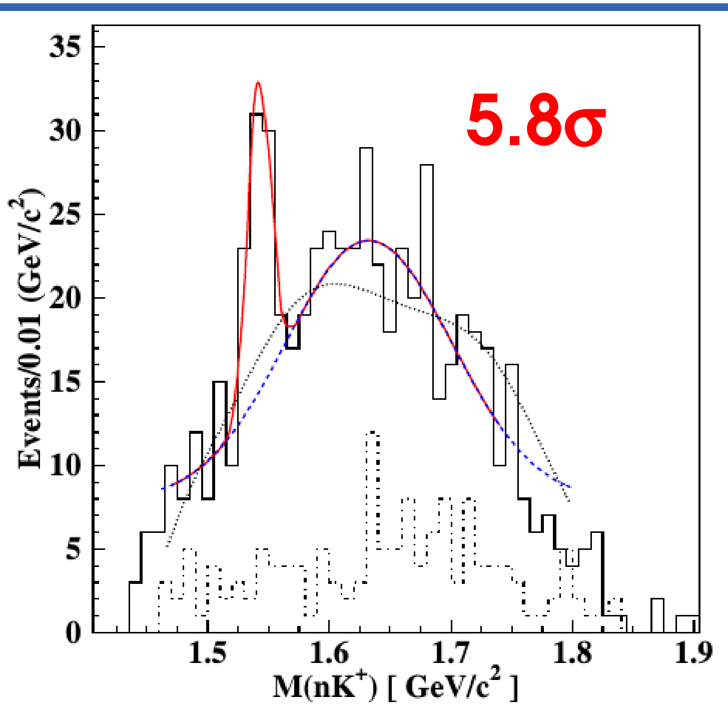
(ii) Any invisible Z^0, such as $Z^0 \to \nu\bar{\nu}$ decay, which is expected to have a large (18%) branching ratio. Note that the corresponding decays into charged lepton pairs $Z^0 \to e^+e^-$, $Z^0 \to \mu^+\mu^-$ have lower branching ratios (~3%) and may not have yet been produced within the present statistics.

(iii) New, non-interacting neutral particles.

The jets appear somewhat narrower and with lower multiplicities than the corresponding QCD jets, although it might be premature to draw conclusions on such limited statistics.

A number of theoretical speculations [9] may be relevant to these results. We mention briefly the possibilities of excited quarks or leptons and of composite or coloured or supersymmetric W's and Higgs. A recent calculation [10] ‡8 has been made in the context of the present collider experiment, on the rate of events with large missing transverse energy from gluino pair production with each gluino decaying into a quark, antiquark, and photino. The non-interacting photinos may produce large apparent missing energy. For instance, the calculation gives an expectation of about 100 single-jet events with $\Delta E_M > 20$ GeV for a gluino mass of 20 GeV/c^2. Taking our excess of 5 events above background as an upper limit for such a process, we deduce that the gluino mass must be greater than about 40 GeV/c^2.

Pentaquarks (2003)



Phys. Rev. Lett. 91, 252001 (2003)

BICEP2 B-mode Polarization (2014)

Selected for a Viewpoint in *Physics*

PRL **112**, 241101 (2014) PHYSICAL REVIEW LETTERS week ending 20 JUNE 2014

Detection of *B*-Mode Polarization at Degree Angular Scales by BICEP2

$$r = 0.20^{+0.07}_{-0.05}, \text{ with } r = 0 \text{ disfavored at } 7.0\sigma$$

Feb. 2015 : arXiv:1502.00612

The final result is expressed as a likelihood curve for r, and yields an upper limit $r < 0.12$ at 95% confidence.

Errors in Hypothesis Testing

	Data disfavors H_0	Data favors H_0
H_0 is false	Discovery, OK	Missed discovery **Type-II error (β)**
H_0 is true	False discovery **Type-I error (α)**	No Discovery, OK

Avoid false discoveries : **lower α** $\Rightarrow$ **make it harder to reject H_0**

$\Rightarrow$ also makes it harder to reject H_0 when it is false : **increase β**

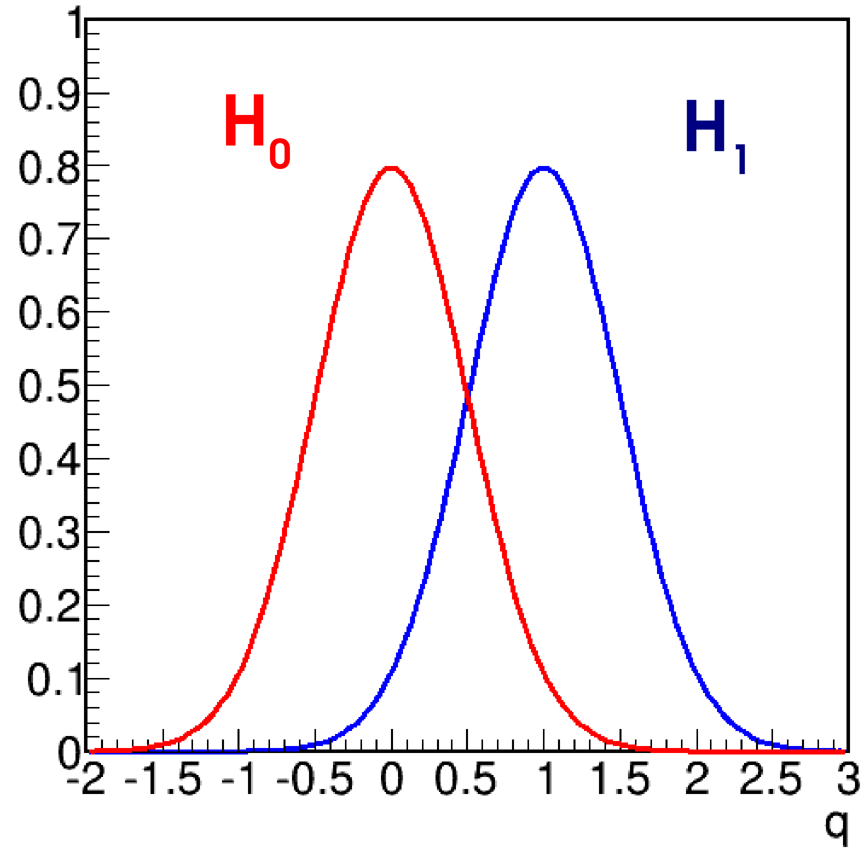
Usually :

- **Fix p-value** α to a small value
- **Goal** find the method giving **the smallest β for this α.**
- Equivalently, maximize **Power** = **1-β**

N_{sigmas}	1-α	p-value
1	0.68	0.32
3	0.997	0.003
5	0.999999	6 10^{-7}

Statistic

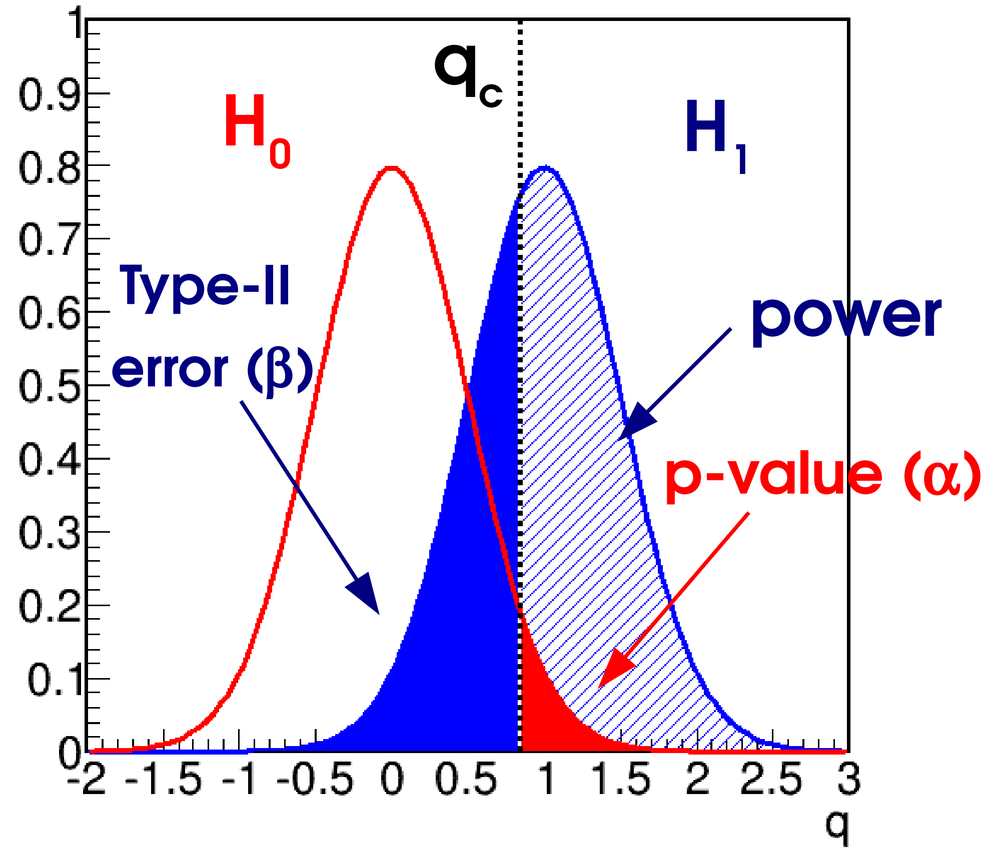
- **Idea**: build some a discriminant **q** from the data to test H_0.
- q has different distributions
  if H_0 is true or H_1 is true
- Looking for q such as
  - **low values of q** ⇒ favor $\mathbf{H_0}$
  - **high values of q** ⇒ favor $\mathbf{H_1}$
  - High separation between the two distributions



- Define a **critical value** $\mathbf{q_c}$.
  Then for the value $\mathbf{q_{obs}}$ computed from data,
  - $q_{obs} < q_c$ ⇒ choose H_0 ("accept H_0") => **Claim no discovery**
  - $q_{obs} > q_c$ ⇒ choose H_1 ("exclude H_0") => **Claim discovery!**

Statistic

- **Idea**: build some a discriminant **q** from the data to test H_0.
- q has different distributions
  if H_0 is true or H_1 is true
- Looking for q such as
  - **low values of q** $\Rightarrow$ favor $\mathbf{H_0}$
  - **high values of q** $\Rightarrow$ favor $\mathbf{H_1}$
  - High separation between the two distributions
- Define a **critical value** q_c.

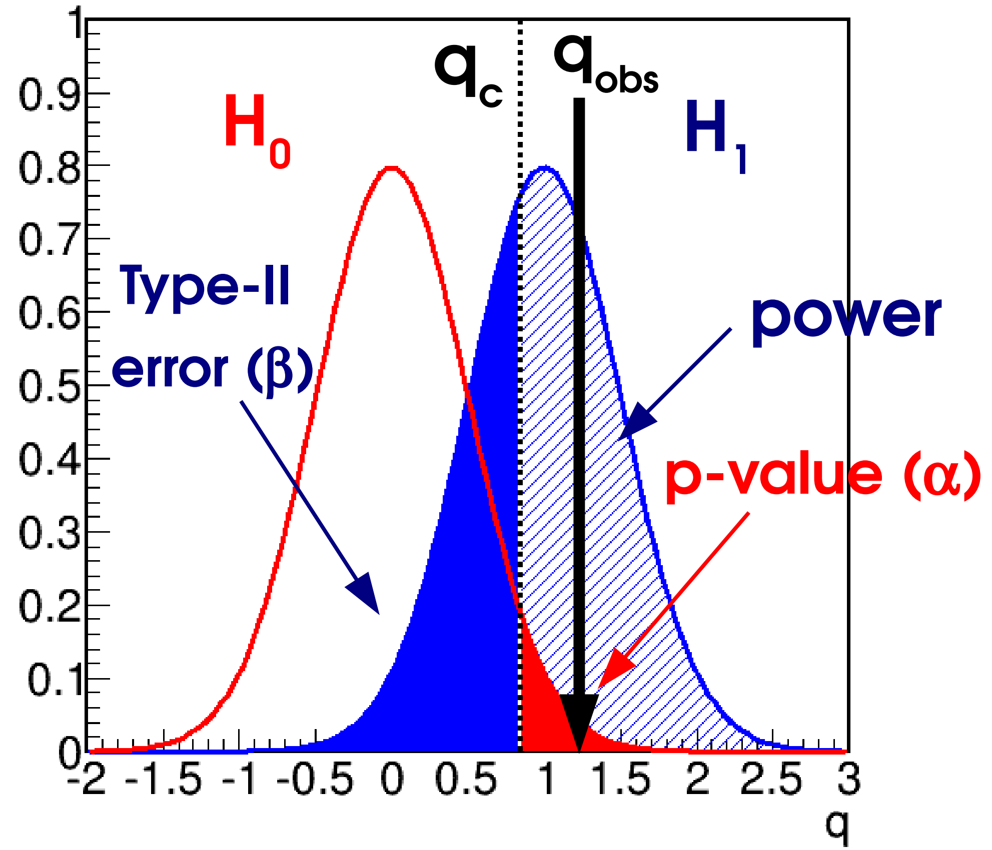


Then for the value q_{obs} computed from data,

- $q_{obs} < q_c$ $\Rightarrow$ choose H_0 ("accept H_0") => **Claim no discovery**
- $q_{obs} > q_c$ $\Rightarrow$ choose H_1 ("exclude H_0") => **Claim discovery!**

Statistic

- **Idea**: build some a discriminant **q** from the data to test H_0.
- q has different distributions
  if H_0 is true or H_1 is true
- Looking for q such as
  - **low values of q** $\Rightarrow$ favor $\mathbf{H_0}$
  - **high values of q** $\Rightarrow$ favor $\mathbf{H_1}$
  - High separation between the two distributions



- Define a **critical value** q_c.
  Then for the value $\mathbf{q_{obs}}$ computed from data,
  - $q_{obs} < q_c$ $\Rightarrow$ choose H_0 ("accept H_0") => **Claim no discovery**
  - $q_{obs} > q_c$ $\Rightarrow$ choose H_1 ("exclude H_0") => **Claim discovery!**

Neyman-Pearson Lemma

- Define q from **likelihoods**:
  - Compute for H_0 : **$L(H_0; data) = L(\theta(H_0); data)$** e.g. $\mu=0$
  - Compute for H_1 : **$L(H_1; data) = L(\theta(H_1); data)$** e.g. $\mu=1$

- Then **$LR = L(H_1; data)/L(H_0; data)$** obviously carries information:
  - Data is **H_0-like** : $L(H_1; data)$ small, **$L(H_0; data)$ large** $\Rightarrow$ **small LR**
  - Data is **H_1-like** : **$L(H_1; data)$ large**, $L(H_0; data)$ small $\Rightarrow$ **large LR**

- **Neyman-Pearson lemma:** The Likelihood ratio is actually **optimal** – it carries the maximum available information

- In practice use :

$$q = -2\log\frac{L(H_0; data)}{L(H_1; data)} = \lambda(H_0; data) - \lambda(H_1; data)$$

What we learned so far (1)

- When testing hypotheses 2 types of potential errors:
  - **False discoveries**
    - probability to happen = **p-value**
  - **Missed discoveries**
    - probability to happen = 1-**power**

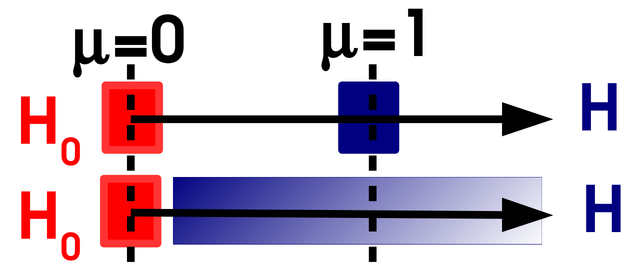
- **Optimal: use likelihood-ratio statistic for decision:**

$$q = -2\log\frac{L(H_0; data)}{L(H_1; data)} = \lambda(H_0; data) - \lambda(H_1; data)$$

- Highest power for a given (small) -value, and vice-versa.
- **High values of q $\Rightarrow$ discovery**

Testing for Discovery

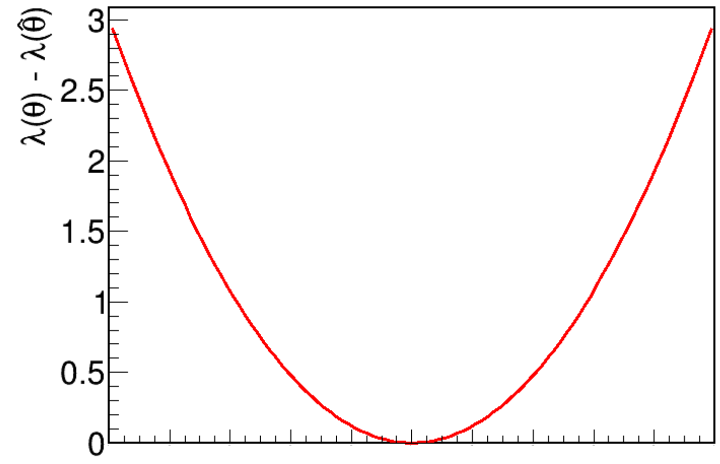
Profile Likelihood Ratio



- Sometimes an hypothesis corresponds to a **range of values**:
  - **"Any Higgs"** : **μ>0** (not just **μ=1**)
  - **Additional parameters**: e.g. m_H can have any value.
- Many values of L(H; data) to consider – which one to take ?
- ⇒ Take the one that **Maximizes the Likelihood within H**.
  - Give the hypothesis its "best chance" (highest L)
- **Discovery test**: **H_0 : μ=0**, **H_1 : μ>0**
  - L(H_0, data) = **L(μ=0; data)**
  - L(H_1, data) = **L($\hat{\mu}$; data)**
  - q_0 gives similar information to $\hat{\mu}$:
    - **$\hat{\mu}$ ~ 0** : L($\hat{\mu}$) ~ L(0) ⇒ **q_0 ~ 0**
    - **|$\hat{\mu}$| ≫ 1** : L($\hat{\mu}$) ≫ L(0) ⇒ **q_0 ≫ 1**

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

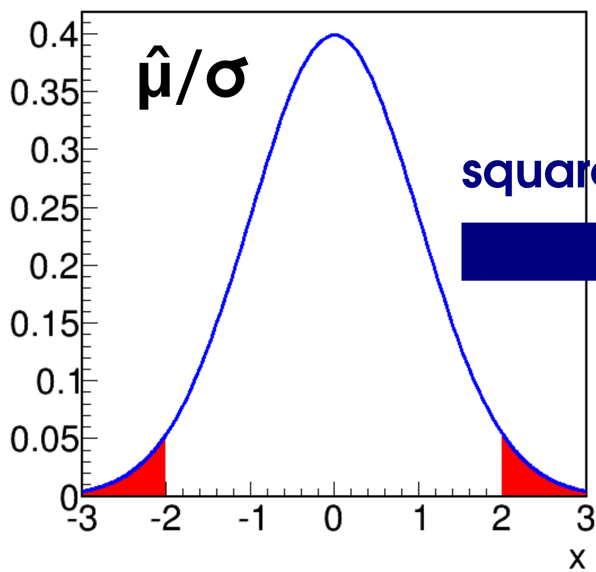
Wilks' Theorem

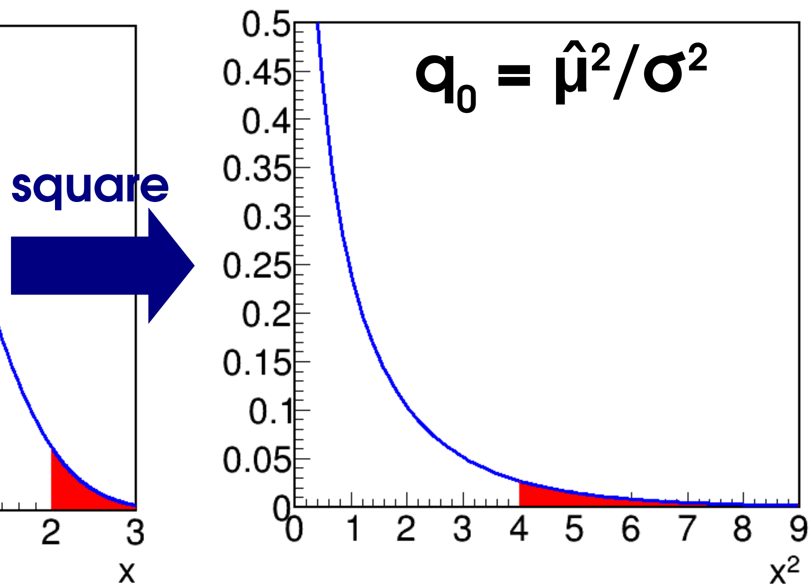


- We have $q_0 = \lambda(0) - \lambda(\hat{\mu})$
- Gaussian likelihood:

  $\boldsymbol{\lambda(\mu) - \lambda(\hat{\mu}) = (\mu - \hat{\mu})^2/\sigma^2}$
- So $\boldsymbol{q_0 = \hat{\mu}^2/\sigma^2}$
- So q_0 = square of a Normal-distributed variable :

  follows a χ^2 distribution So also same quantiles:



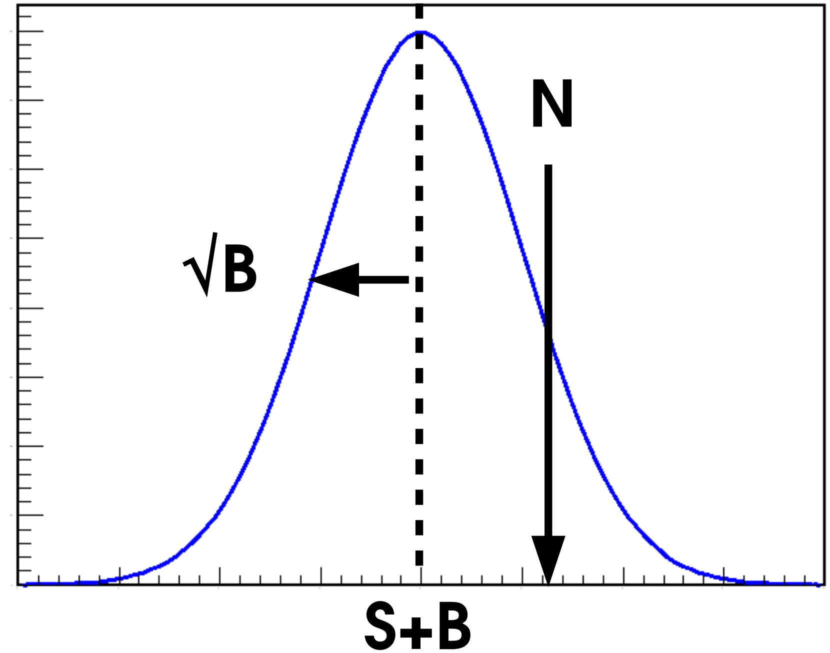


$$P(q_0 > n^2) = P(|x| < n)$$

Z	Region	p-value
1	$q_0 > 1$	**0.32**
3	$q_0 > 9$	**0.003**
5	$q_0 > 25$	**6×10^{-7}**

Significance (p-value in number of sigmas) : $\boxed{Z = \sqrt{q_0}}$

Trivial Examples



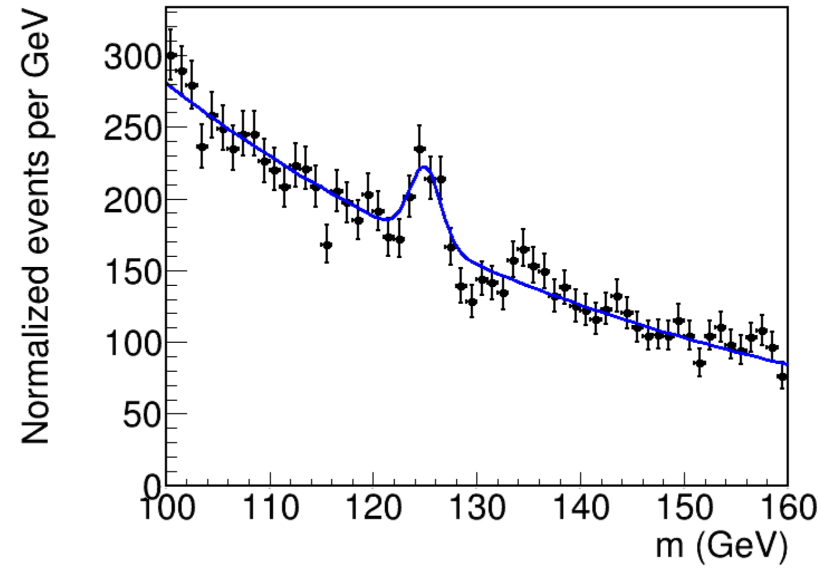
- **Gaussian Counting, fixed B.**
  - $\lambda(S) = (N - (S+B))^2/(\sqrt{B})^2$

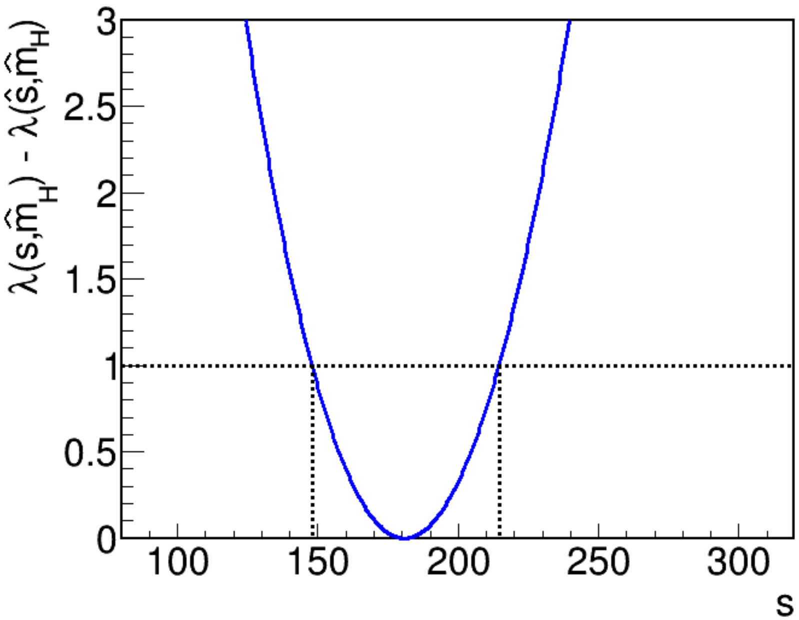
    Reminder:
    - Best fit signal : $\hat{S}$ = N-B
    - 68% CI : $[\hat{S}-\sqrt{B}, \hat{S}+\sqrt{B}]$
  - Now we compute the significance:
    - $q_0 = \lambda(0) - \lambda(\hat{S}) = (N-B)^2/B$
    - So $\mathbf{Z = (N-B)/\sqrt{B} = \hat{S}/\sqrt{B}}$
- **Multiple Gaussian Measurements:**
  - Two indep. measurements, likelihoods L_1 and L_2
  - $L = L_1 L_2 \Rightarrow \lambda = \lambda_1 + \lambda_2 \Rightarrow q_0 = q_{0,1} + q_{0,2}$

$$\mathbf{Z^2 = Z_1^2 + Z_2^2} \textbf{ : significances add in quadrature}$$

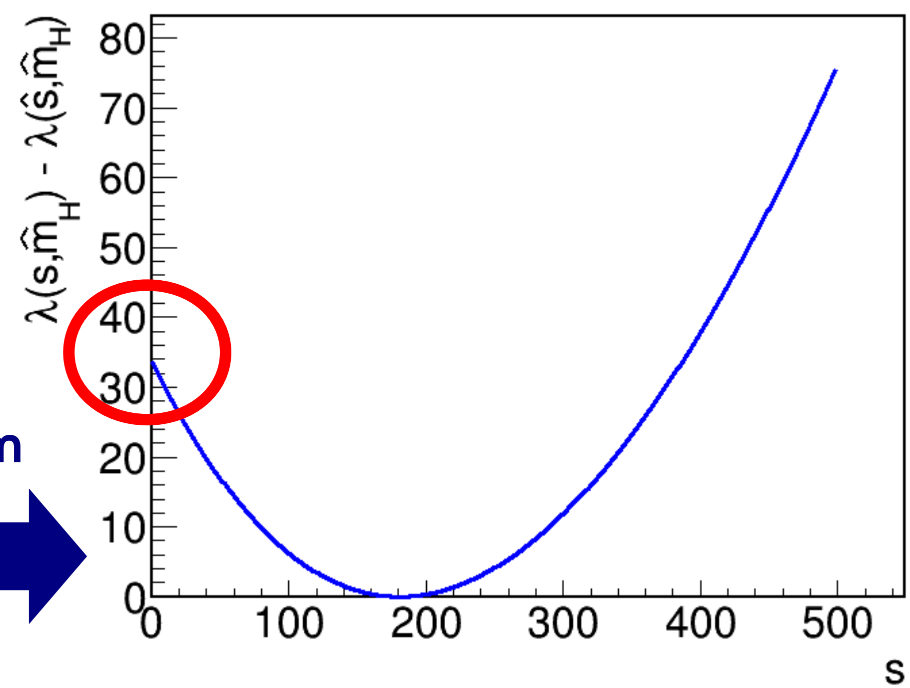
Our Usual Example

- Use pseudo-data generated with **s=200, m_H=125 GeV**
- We had $s = 181^{+34}_{-33}$
- So we expect Z = 181/33 = **5.5**
- $q_0 = \lambda(0) - \lambda(\hat{\mu}) = 33.6 \Rightarrow$ **Z=5.8**
- Small difference due to residual non-Gaussianity





unzoom



Wilks' Theorem (2)

- Testing using q_0 is equivalent to using $\hat{\mu}$, so why use q_0 ?
  - **In case of non-Gaussianity, q_0 is generally more robust**
  - Simple treatment for additional parameters:
- **Wilks' theorem**: if we have $L(\mu, \theta)$, where q can stand for many parameters, then compute q0 as

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

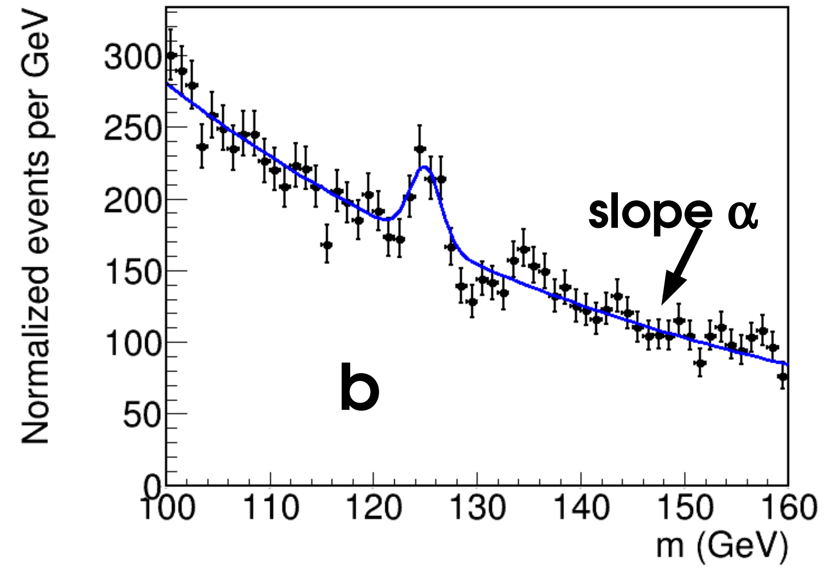
Where

- $\hat{\theta}$ = value of θ at L_{max}, **with μ also free** (within **H_1**)
- $\hat{\theta}_{\mu=0}$ = value of θ at L_{max}, **with fixed $\mu=0$** (within **H_0**)

Then q_0 is still distributed as a χ^2.

In other words, "nuisance parameters" can be "profiled away" and do not need special treatment.

Our Usual Example, Again



- Same as before, but now let the background parameters b and α free

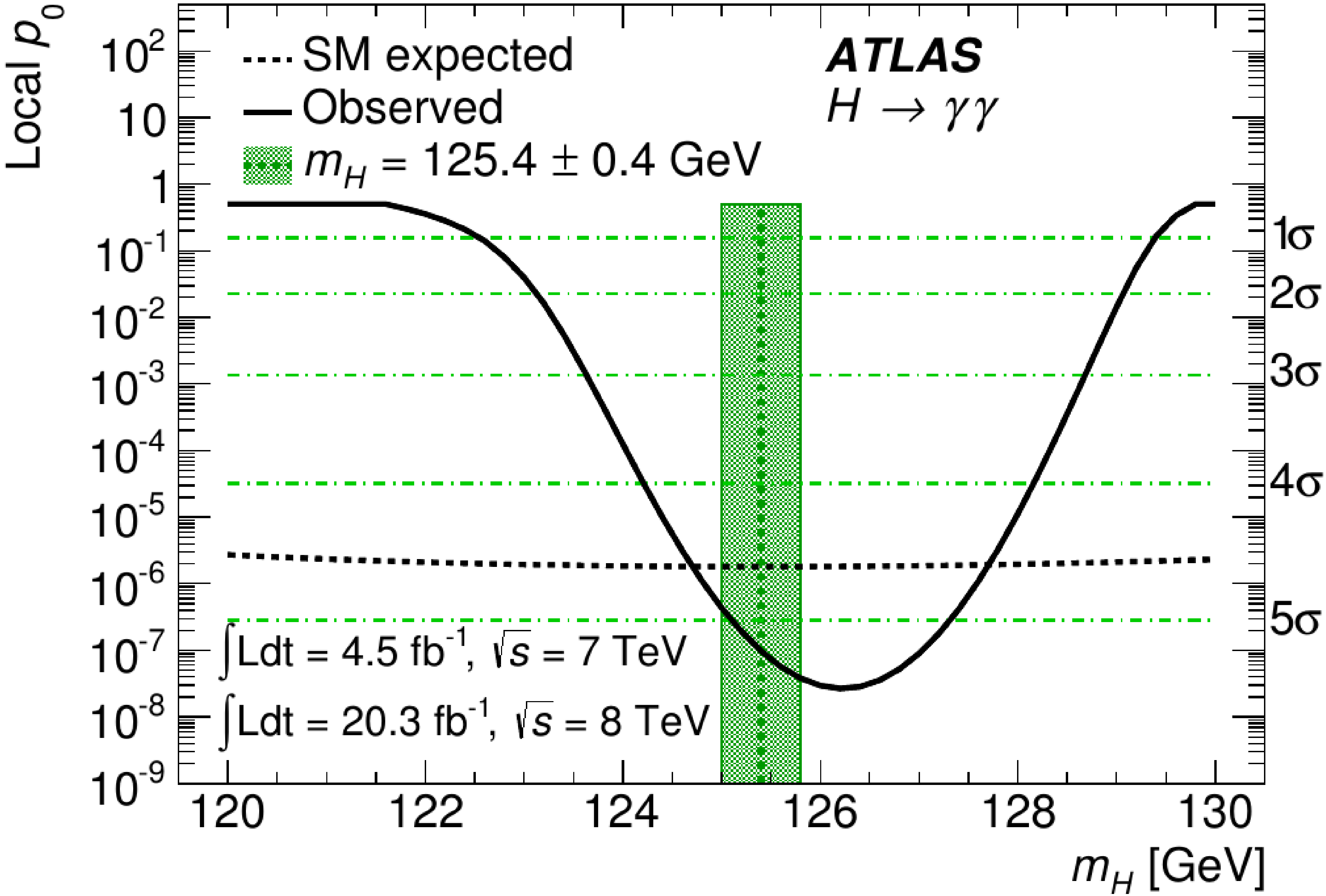
- $q_0 = \lambda(0, \hat{\alpha}_0, \hat{b}_0) - \lambda(\hat{\mu}, \hat{\alpha}, \hat{b})$

- Now b can vary: $\hat{b}_0$ = 10k but $\hat{b}$ = 9800 due to some events going into the signal for $\mu>0$.

- Slope parameter α can also vary

- $q_0 = 37.0$, $Z = 6.1$.

Real-Life: p_0 vs. m_H for $H \rightarrow \gamma\gamma$



What we learned so far (2)

Testing for discovery

- As usual, start with a likelihood **L(θ)** for the measurement
- Compute $\boldsymbol{\lambda(\theta) = -2 \log L_{data}(\theta)}$, as a function of θ.
- Find the minima of λ
  - For no signal
  - In the presence of signal
  - Compute q_0 :

Z	Region	p-value
1	$q_0 > 1$	**0.32**
3	$q_0 > 9$	**0.003**
5	$q_0 > 25$	**6×10^{-7}**

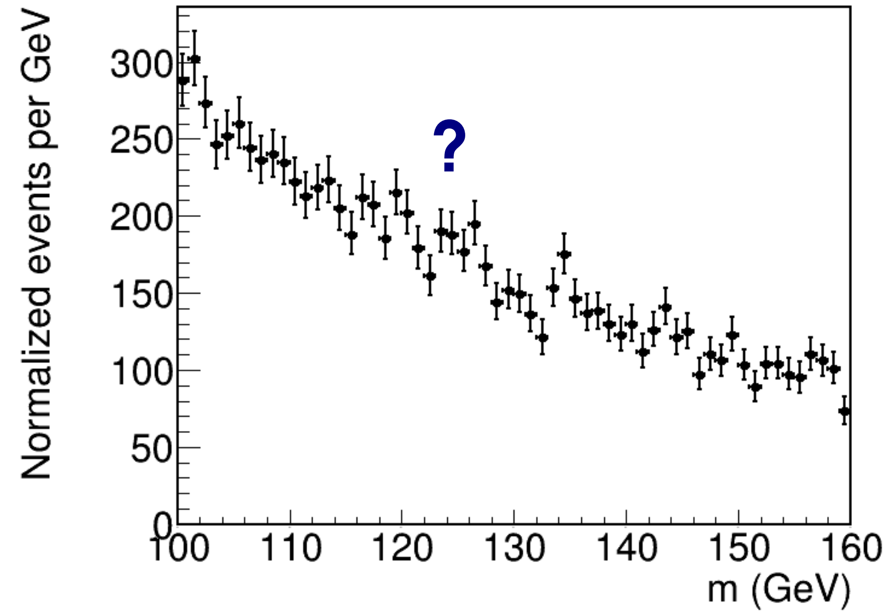
$$q_0 = -2 \log \frac{L(\mu=0)}{L(\hat{\mu})}$$

- Compute the p-value for discovery assuming a χ^2 distribution. For a single signal parameter just use

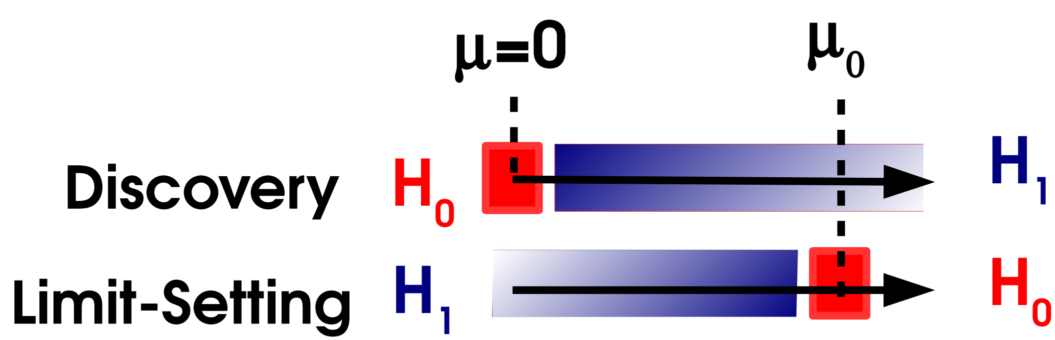
$$Z = \sqrt{q_0}$$

Setting Limits

Hypothesis tests for Limits

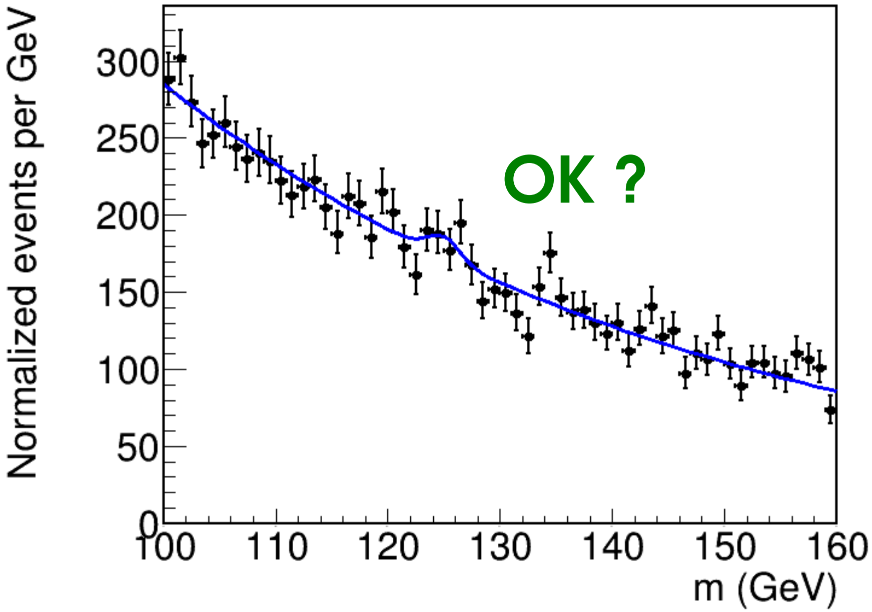


- If no signal in data, testing for discovery is not very relevant (report 0.2σ excess ?)
- More interesting to **exclude large values of μ**
- For **discovery**, hypotheses were:
  – Try to exclude **H_0 : $\mu=0$**
  – Alternative : **H_1 : $\mu>0$**
- For **limit-setting**:
  – Try to exclude **H_0 : $\mu=\mu_0$**
  – Alternative : **H_1 : $\mu<\mu_0$**

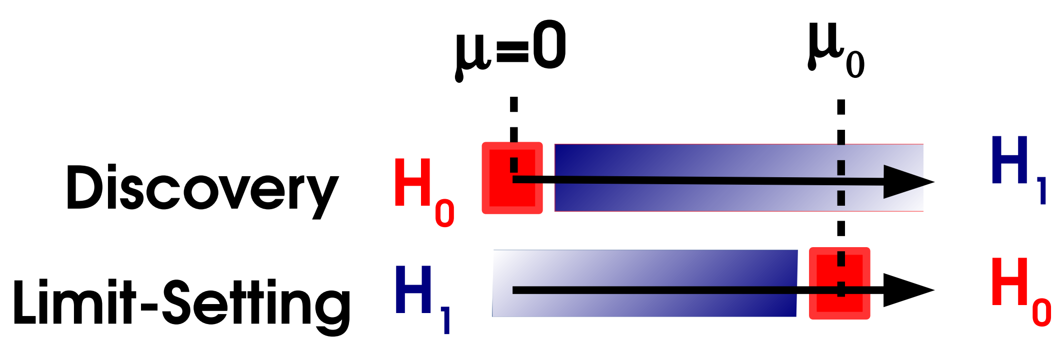


- Usually, **adjust μ_0 until a given p-value is reached** (typically **95%**)
- Interesting p-values ("Confidence Levels", CL) typically lower than for discovery (95% $\Leftrightarrow$ **1.96σ**)

Hypothesis tests for Limits

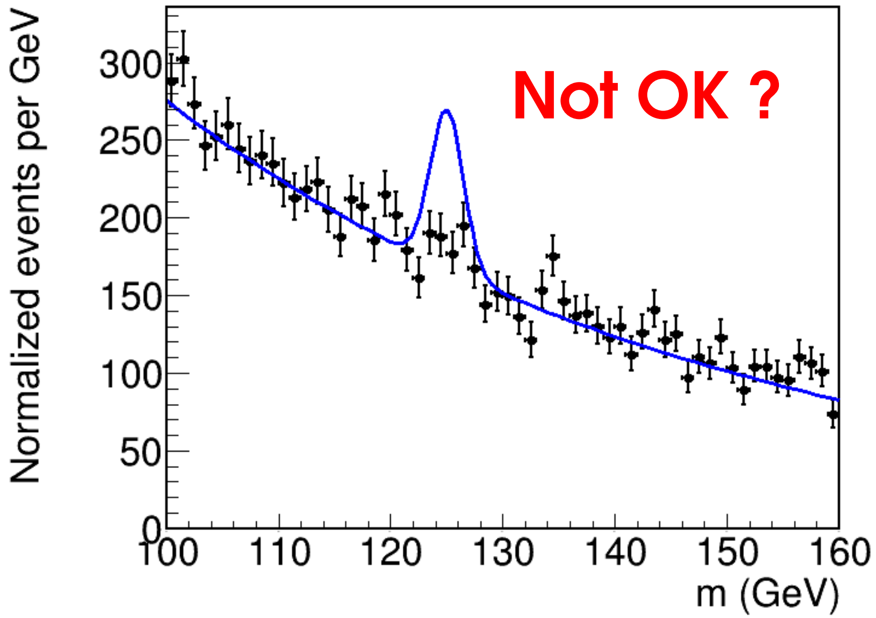


- If no signal in data, testing for discovery is not very relevant (report 0.2σ excess ?)
- More interesting to **exclude large values of μ**
- For **discovery**, hypotheses were:
  - Try to exclude **H_0 : μ=0**
  - Alternative : **H_1 : μ>0**
- For **limit-setting**:
  - Try to exclude **H_0 : $\mu=\mu_0$**
  - Alternative : **H_1 : $\mu<\mu_0$**

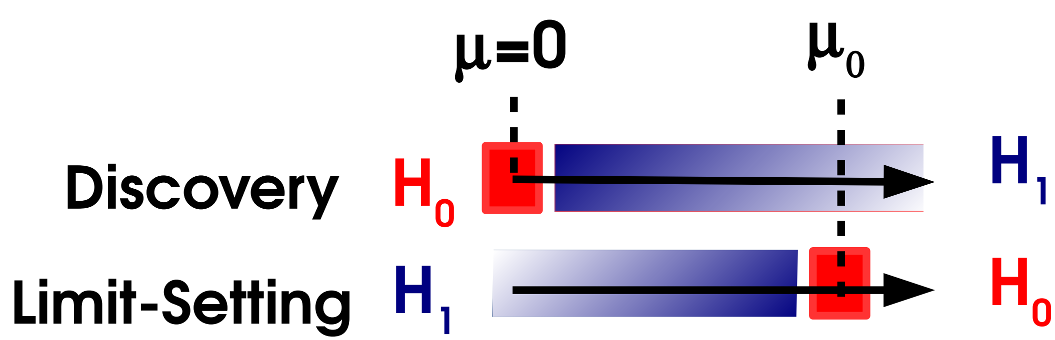


- Usually, **adjust μ_0 until a given p-value is reached** (typically **95%**)
- Interesting p-values ("Confidence Levels", CL) typically lower than for discovery (95% ⇔ **1.96σ**)

Hypothesis tests for Limits



- If no signal in data, testing for discovery is not very relevant (report 0.2σ excess ?)
- More interesting to **exclude large values of μ**
- For **discovery**, hypotheses were:
  - Try to exclude **H_0 : μ=0**
  - Alternative : **H_1 : μ>0**
- For **limit-setting**:
  - Try to exclude **H_0 : μ=μ_0**
  - Alternative : **H_1 : μ<μ_0**



- Usually, **adjust μ_0 until a given p-value is reached** (typically **95%**)
- Interesting p-values ("Confidence Levels", CL) typically lower than for discovery (95% ⇔ **1.96σ**)

Statistic for Limit-Setting

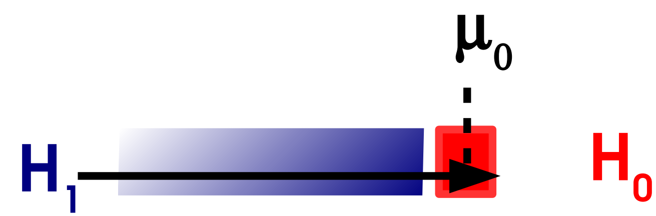
- **For discovery :**
  - H_0 : μ=0
  - H_1 : μ>0

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

Compare

Likelihood of H_0

Likelihood of H_1



- **For limit-setting**
  - H_0 : $\mu=\mu_0$
  - H_1 : $\mu<\mu_0$

$$q_{\mu_0} = -2\log\frac{L(\mu_{0,}\hat{\hat{\theta}}_{\mu_0})}{L(\hat{\mu},\hat{\theta})}$$

Compare

Likelihood of H_0

Likelihood of H_1

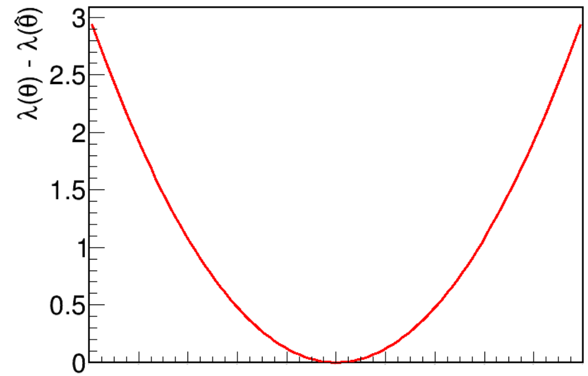
$\hat{\mu} \sim \mu_0$ (no exclusion) : $q_{\mu 0} \sim 0$

$\hat{\mu} \ll \mu_0$ (good exclusion) : $q_{\mu 0} \gg 1$

Same as q_0 : large values

$\Rightarrow$ good rejection of H_0.

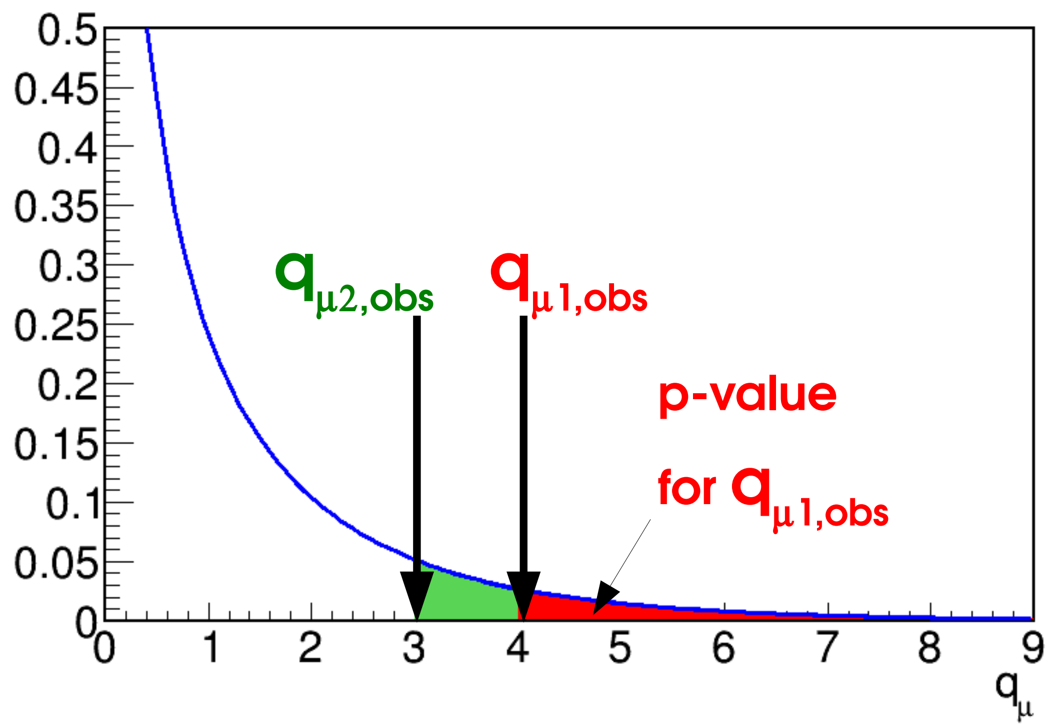
Distribution for q_μ



- **Gaussian case: $\lambda(\mu) - \lambda(\hat{\mu}) = (\mu - \hat{\mu})^2/\sigma^2$**
- We have $q_{\mu 0} = \lambda(\mu_0) - \lambda(\hat{\mu}) = (\hat{\mu}-\mu_0)^2/\sigma^2$
  - Under $H_0(\mu=\mu_0)$, $\hat{\mu} \sim G(\mu_0,\sigma)$,

    $\Rightarrow (\hat{\mu}-\mu_0)/\sigma$ is Normal-distributed $\Rightarrow$ **$q_{\mu 0}$ has a χ^2 distribution**
- **All the q_μ have the same distribution.**
- The **observed** $q_{\mu,obs}$ are different : $q_{\mu,obs}$ higher for larger μ

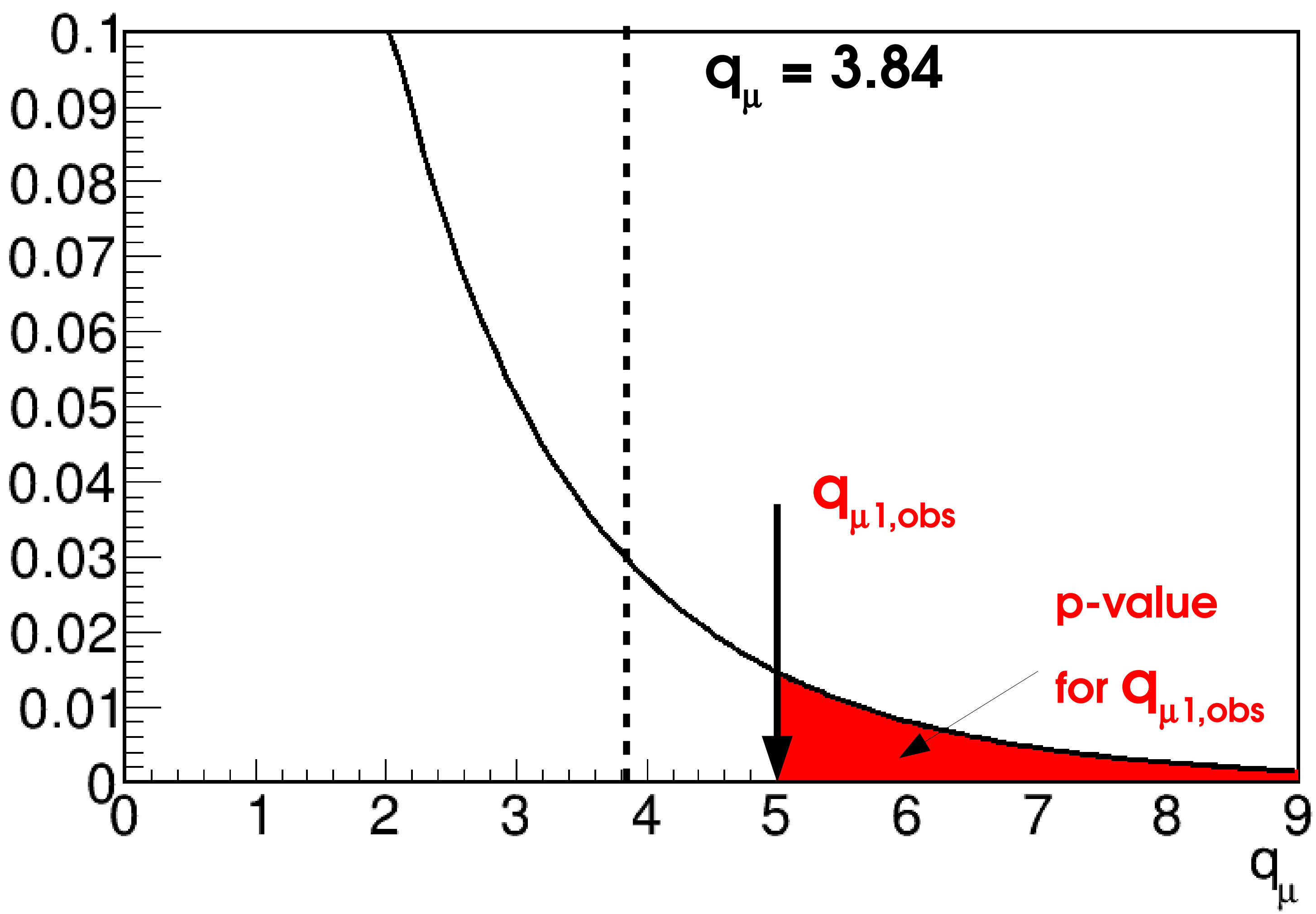
Z	Region	CL
1.64	$q_\mu > 2.70$	**0.90**
1.96	$q_\mu > 3.84$	**0.95**
2.58	$q_\mu > 6.63$	**0.99**



Finding the limit

- Compute $q_{\mu,obs}$ for various μ.
- Find the μ for which $q_{\mu,obs}$ gives the required p-value (e.g. 5% for 95% CL)

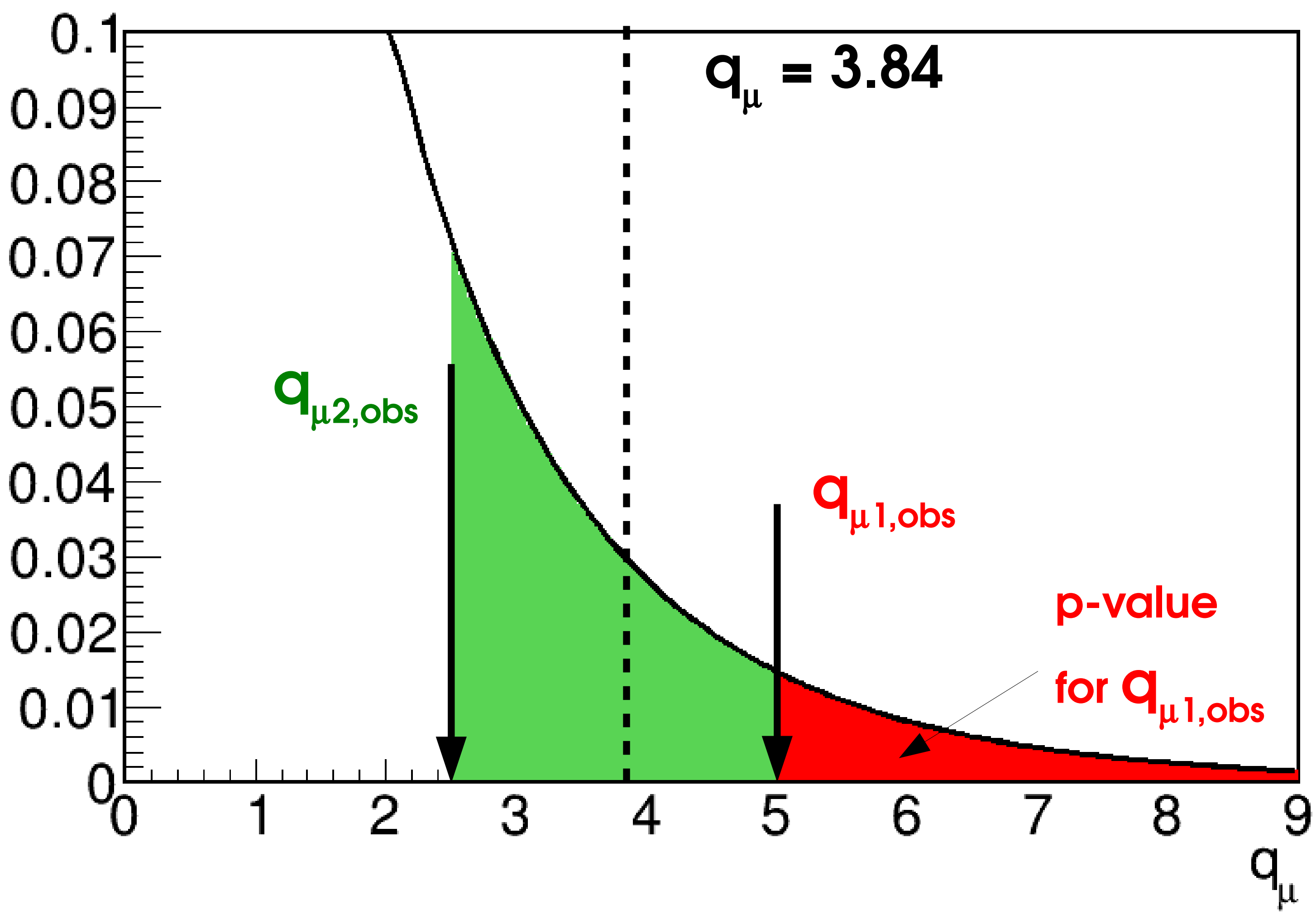
Z	Region	CL
1.64	$q_\mu > 2.70$	**0.90**
1.96	$q_\mu > 3.84$	**0.95**
2.58	$q_\mu > 6.63$	**0.99**



Finding the limit

- Compute $q_{\mu,obs}$ for various μ.
- Find the μ for which $q_{\mu,obs}$ gives the required p-value (e.g. 5% for 95% CL)

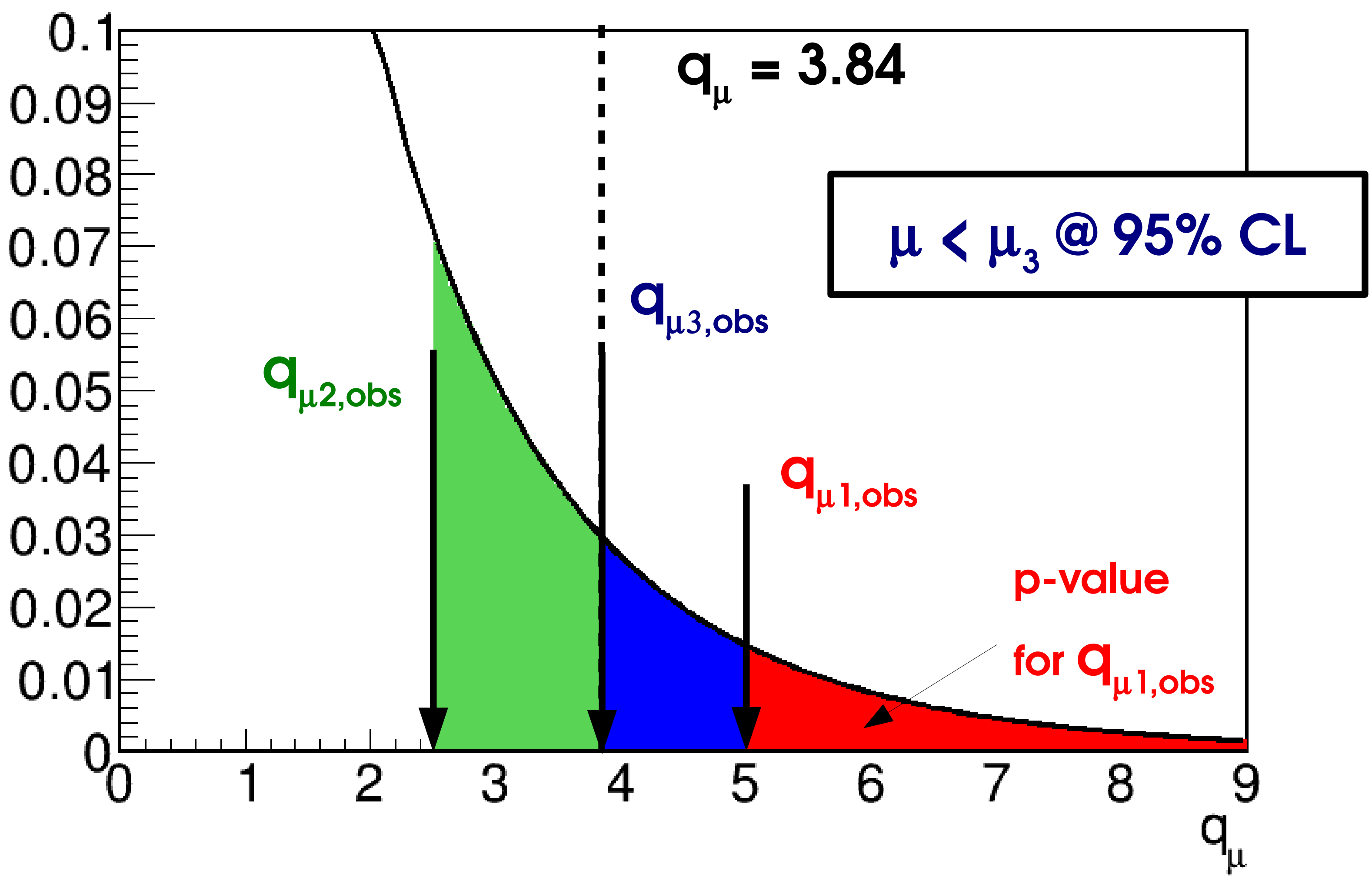
Z	Region	CL
1.64	$q_\mu > 2.70$	**0.90**
1.96	$q_\mu > 3.84$	**0.95**
2.58	$q_\mu > 6.63$	**0.99**



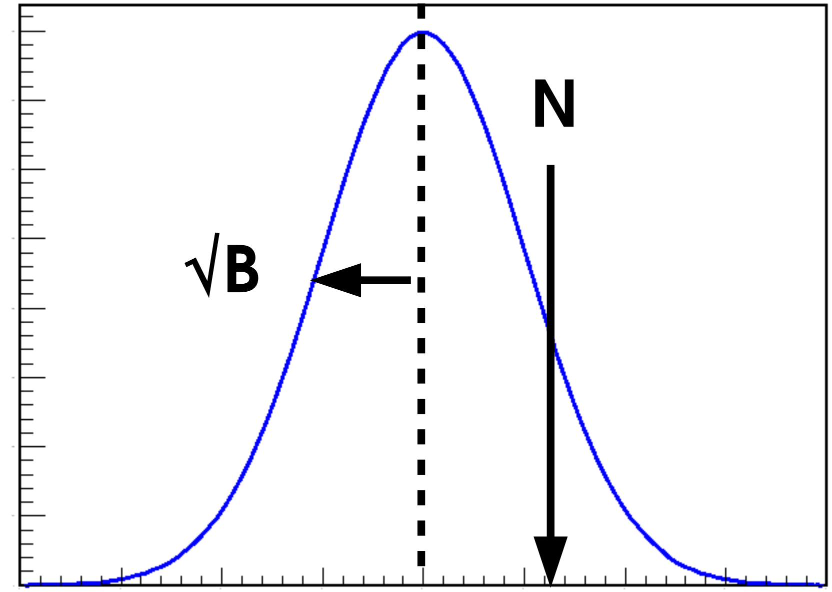
Finding the limit

- Compute $q_{\mu,obs}$ for various μ.
- Find the μ for which $q_{\mu,obs}$ gives the required p-value (e.g. 5% for 95% CL)

Z	Region	CL
1.64	$q_\mu > 2.70$	**0.90**
1.96	$q_\mu > 3.84$	**0.95**
2.58	$q_\mu > 6.63$	**0.99**



Gaussian Example



Gaussian Counting, fixed B.

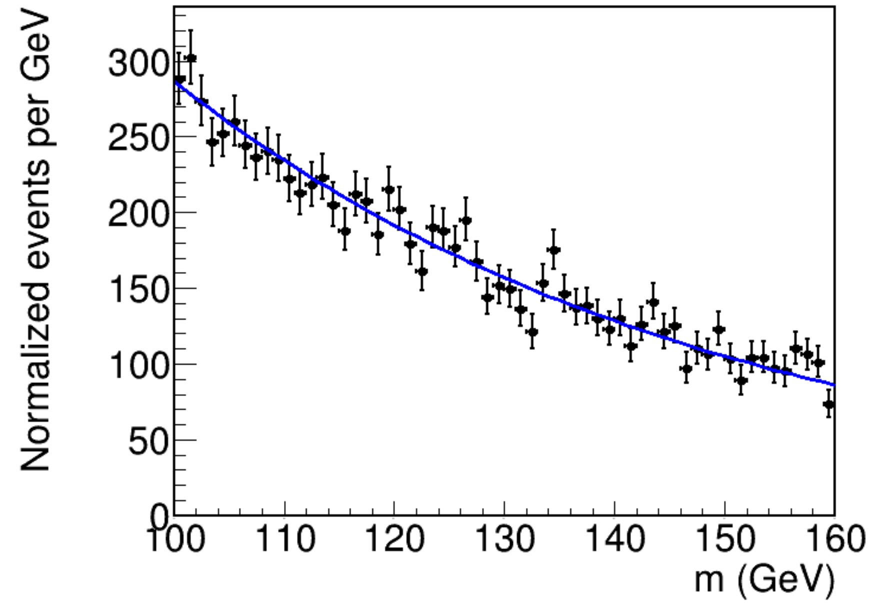
- $\lambda(S) = (N - (S+B))^2/(\sqrt{B})^2$
- Reminder:
  - **Best fit signal** : $\hat{S} = N-B$
  - **68% CI** : $[\hat{S}-\sqrt{B}, \hat{S}+\sqrt{B}]$
  - **Significance**: $\hat{S}/\sqrt{B}$

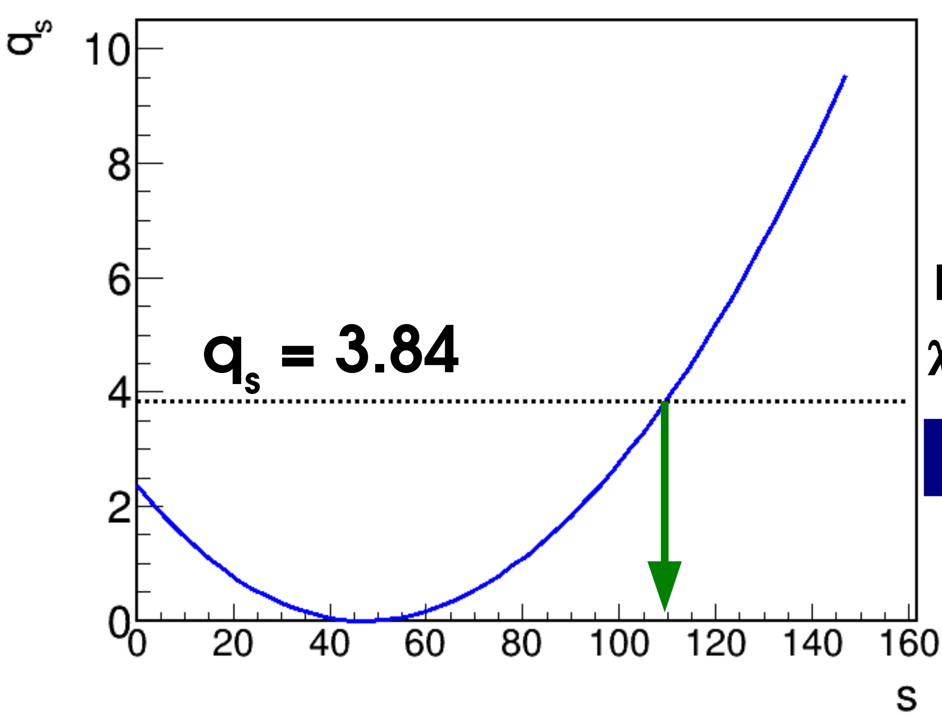
- Now we compute the 95% CL upper limit on S:
  - $q_S = \lambda(S) - \lambda(\hat{S}) = (N - (S+B))^2/B = (S - \hat{S})^2/B$
  - $q_S=3.84 \Rightarrow (S - \hat{S})/\sqrt{B} = 1.96 \Rightarrow S = \hat{S} + 1.96\sqrt{B}$

So at 95% C.L., $S < \hat{S} + 1.96\sqrt{B}$

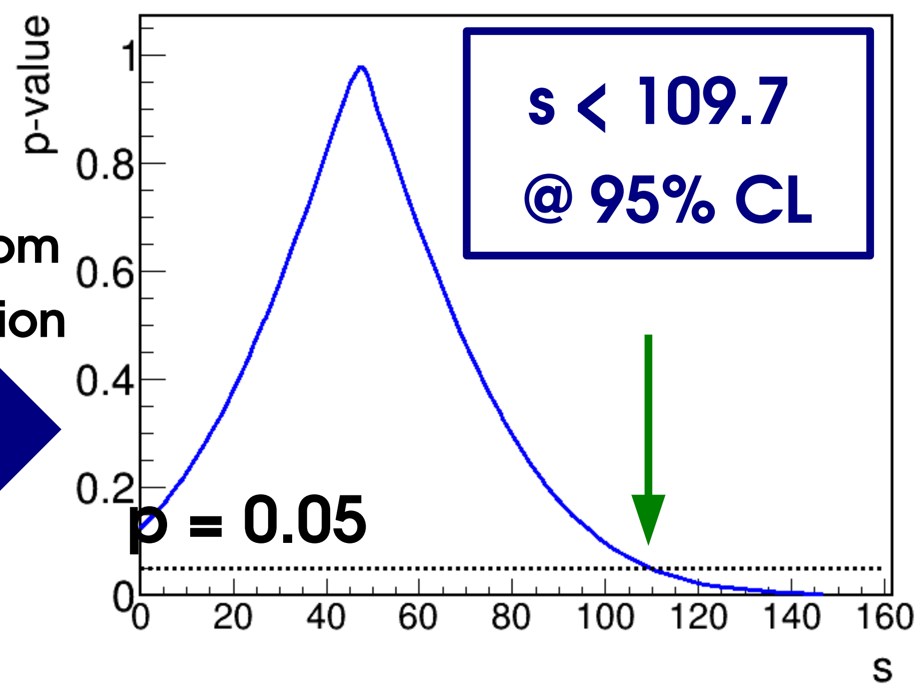
Our Usual Example

- Same as before, but now generate a dataset with **s=0**
- Best-fit **s=47±31**
- 95% CL limit on s in Gaussian approximation: 47+1.96*31 = **107.8**
- Compute q_s using the usual extended likelihood: $L(s,b,\theta;m_1 \ldots m_{n_{obs}}) = e^{-(s+b)} \prod_{i=1}^{n_{obs}} s\,P_{signal}(m) + b\,P_{bkg}(m)$

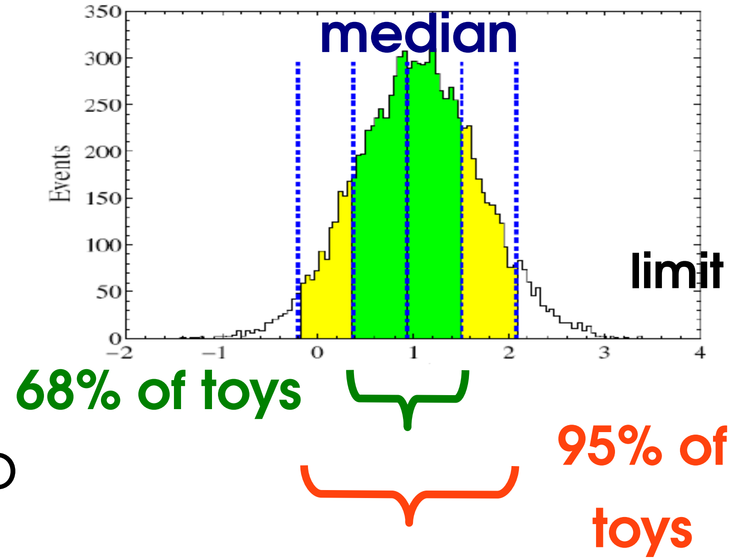




p-value from χ^2 distribution



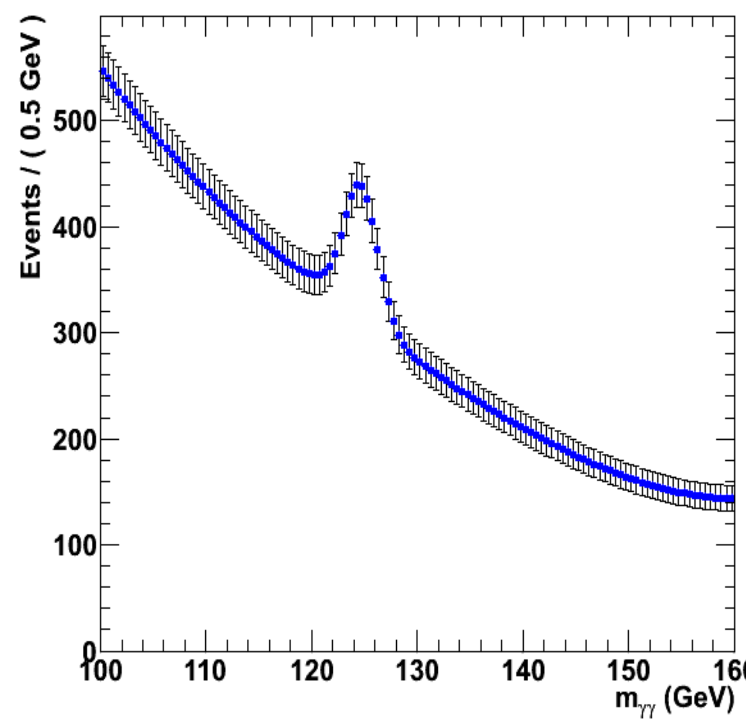
Expected Limits



- "Expected" results: median outcome under given hypothesis, usually B-only
- **From toys:**
  - Generate pseudo-data in B-only hypo
  - Compute limit
  - Repeat and histogram the results, report median & quantiles
- **From Asimov Datasets**
  - Alternative: generate a "perfect dataset" without fluctuations (by setting bin contents carefully)
  - Gives the median immediately
  - Bands from Gaussian approximation (see Asimov paper)

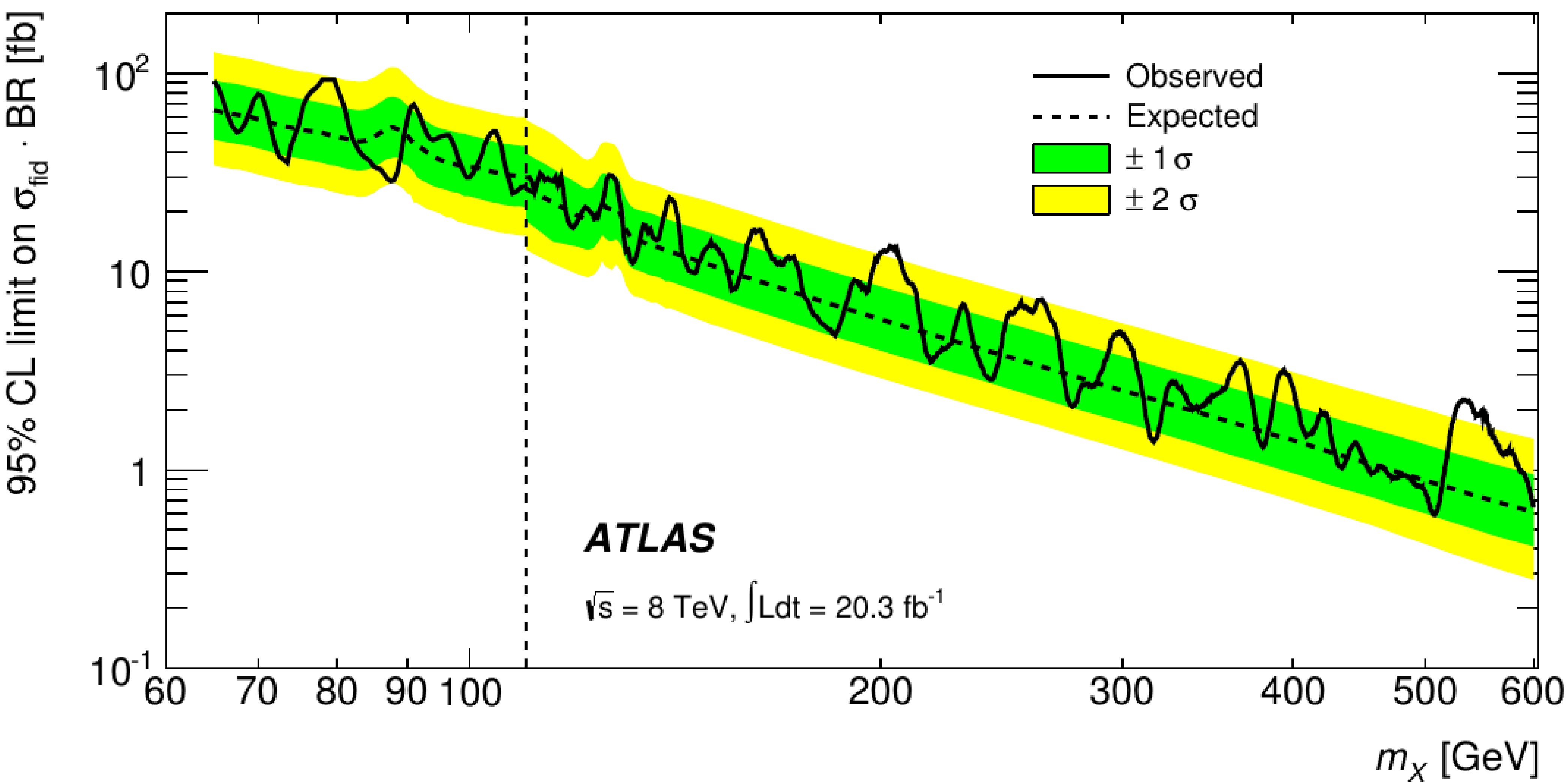
    ⊖ relies on Gaussian approximation

    ⊕ Much faster (1 "toy")



Real-Life Example: Limits on $X \to \gamma\gamma$

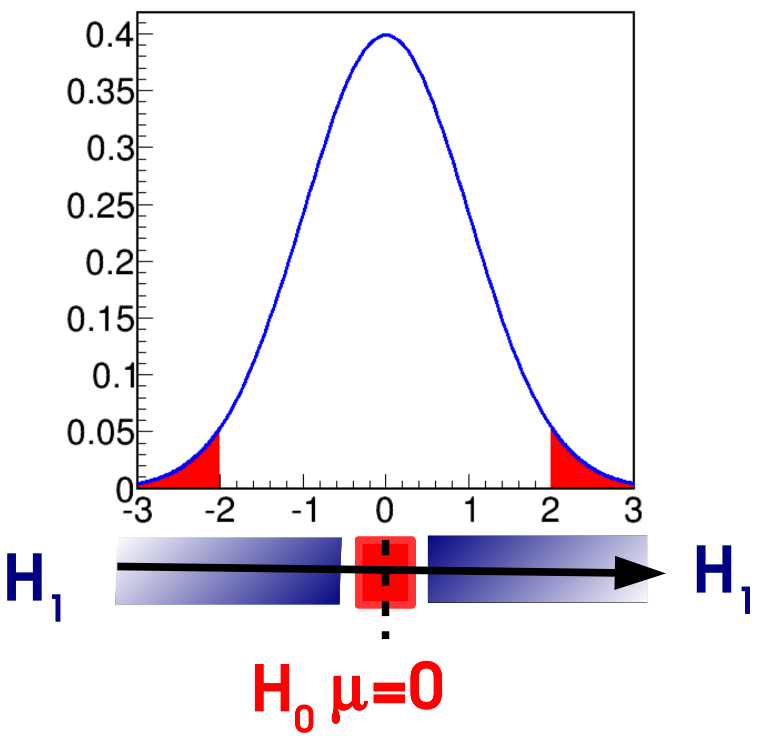
Phys. Rev. Lett. 113, 171801



One-sided vs. Two-Sided : Discovery

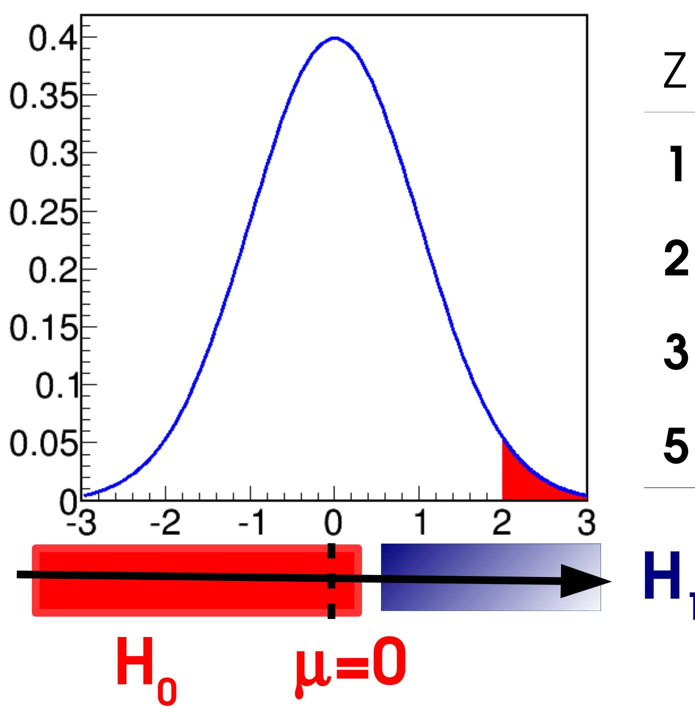
- If $|\mu| \gg 1$ but $\mu<0$ is it a "discovery" (rejects $\mu=0$ hypothesis) ?
- Both treatments are possible, but usually no (A signal is positive!)
- Change statistic so that $\mu<0 \Rightarrow q_0=0$ (perfect agreement)

Yes : Two-sided treatment



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

No : One-sided treatment



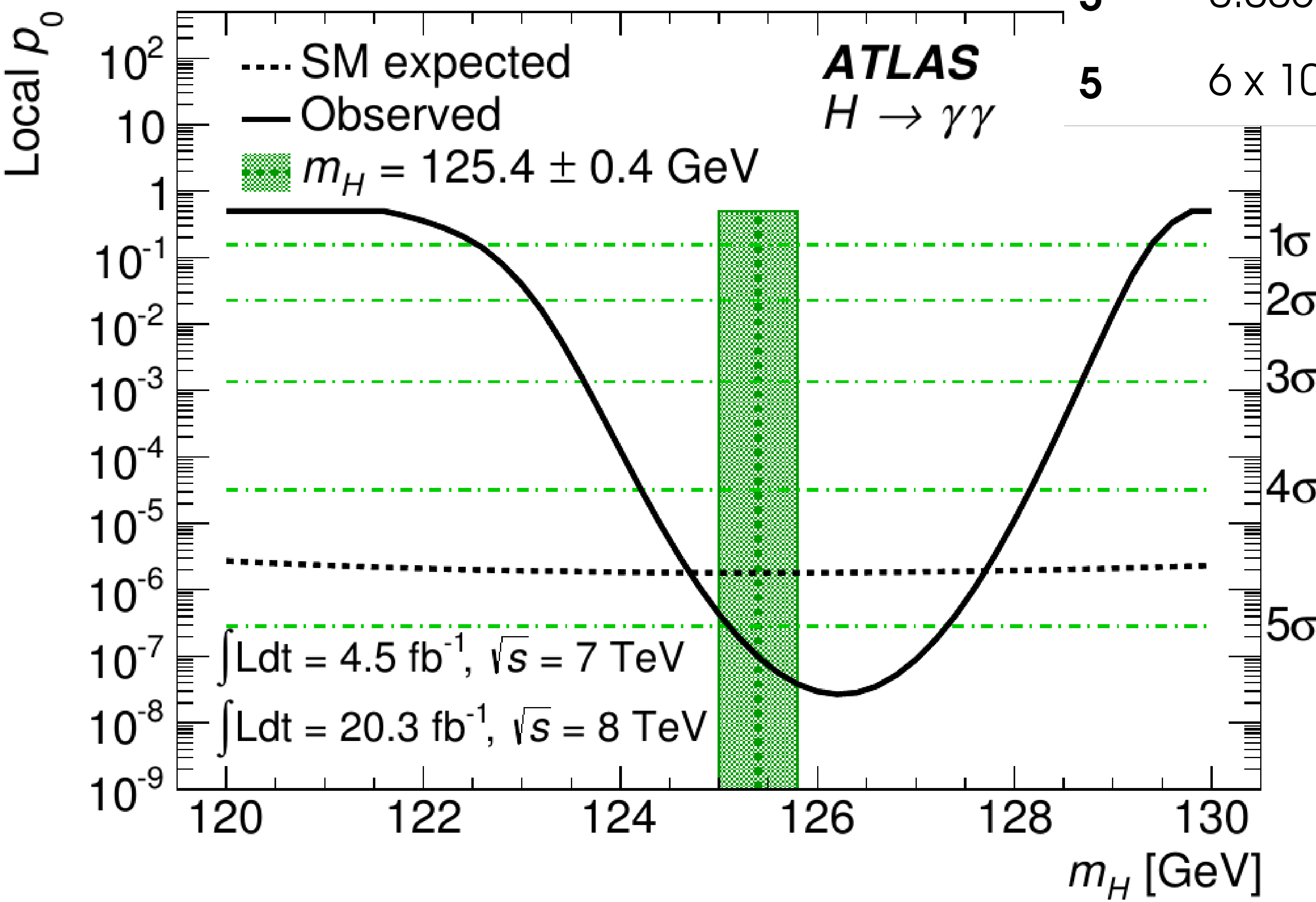
$$q_0 = \begin{cases} -2\log\frac{L(\mu=0)}{L(\hat{\mu})} & \hat{\mu}\geq 0 \\ 0 & \hat{\mu}<0 \end{cases}$$

Z	2-sided	1-sided
1	0.32	0.16
2	0.045	0.022
3	0.003	0.0015
5	6×10^{-7}	6×10^{-7}

Factor of 2 in p-values for a given Z

Real-Life H→γγ

Z	2-sided	1-sided
1	0.32	**0.16**
2	0.045	**0.022**
3	0.003	**1.5 x 10^{-7}**
5	6 x 10^{-7}	**3 x 10^{-7}**

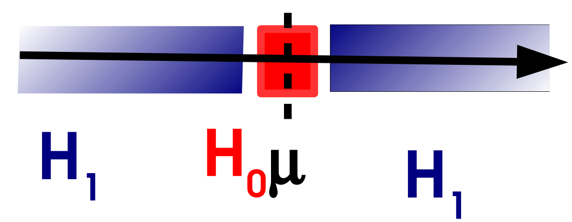


One-Sided vs. Two-Sided limits

- Same issue for limits: if $\hat{\mu} > \mu$, does that help reject the μ hypo ?
- **Usually no**: only reject μ if signal is too low ("**upper**" limit)
- Again, p-values divided by 2
- "Magic number" for 95% CL limits is 1.64 for 1-sided case (recall: 1.96 for 2-sided)

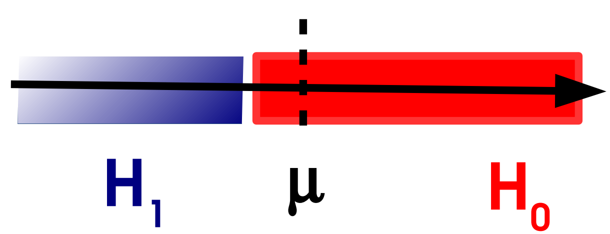
Z	2-sided	1-sided
1.64	**0.90**	**0.95**
1.96	**0.95**	**0.975**

- **2-sided**



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

- **1-sided**



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

What we learned so far (3)

Setting Limits

Z	2-sided	1-sided
1.64	**0.90**	**0.95**
1.96	**0.95**	**0.975**

- As usual, start with a likelihood **L(θ)** for the measurement
- Compute **$\lambda(\theta)$ = -2 log $L_{data}(\theta)$,**
  - For the best-fit signal $\hat{\mu}$
  - For fixed signal hypotheses μ_0
  - Compute $q_{\mu 0}$:

$$q_{\mu_0} = -2\log\frac{L(\mu_0)}{L(\hat{\mu})}$$

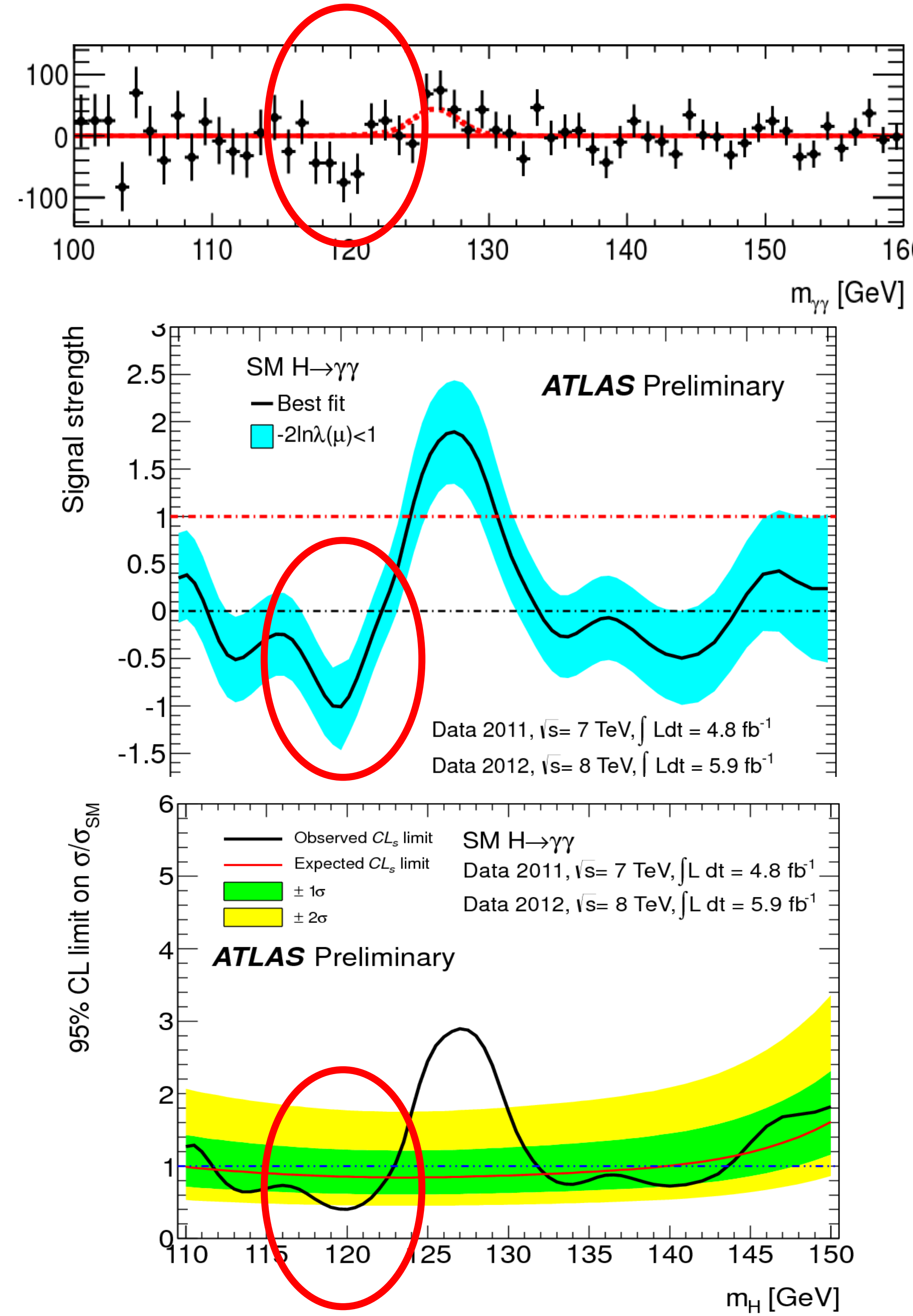
- Compute p-values in data assuming a χ^2 distribution. Adjust the hypothesis μ_0 so that the p-value is 5% (for a 95% exclusion)
  - This corresponds to adjusting μ_0 so that **$q_{\mu 0} = 1.64^2$ or 1.96^2**

Additional Topics

A Limit-Setting Issue

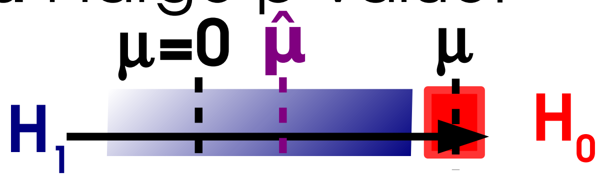
- Limit ~ $\hat{\mu} + 1.96\sigma_{\mu}$
- **Problem**:
  - for negative $\hat{\mu}$, get very good (**too good**) limits.
  - For μ sufficiently negative, can have limit < 0!
- How can this be ?
  - This is a **95%** limit
    $\Rightarrow$ **5% of the time, the limit wrongly excludes the true value.**
  - If we assume μ must be >0, we know these are "wrong" cases

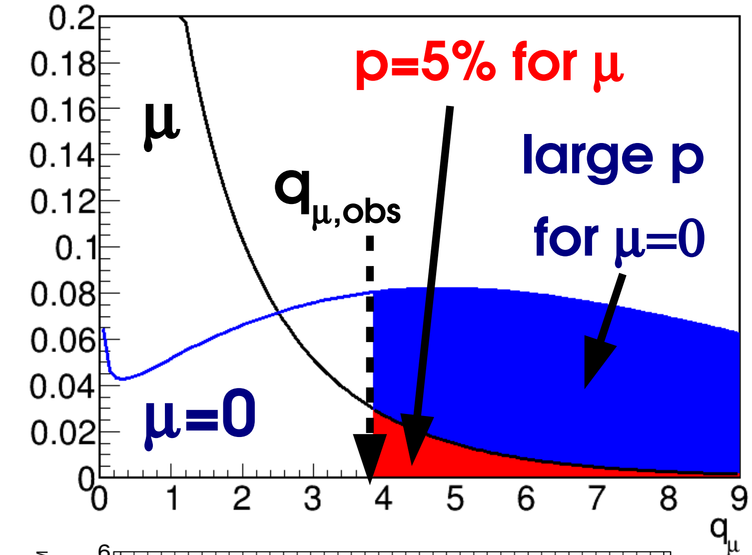
$\Rightarrow$ **Special procedure for $\hat{\mu}$<0**



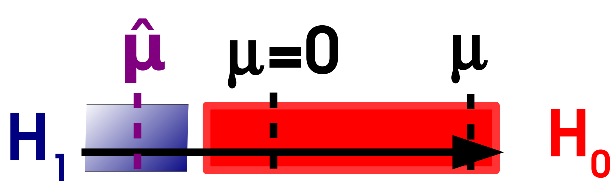
Sensitivity to $\mu>0$

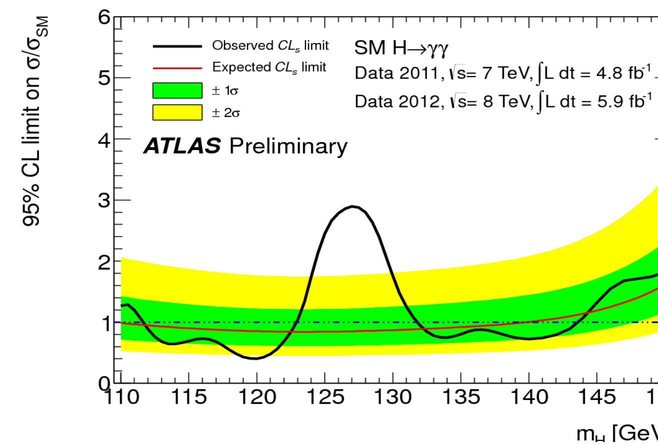
- When setting limits, goal is to exclude large μ, **to indicate that μ~0.**
- Investigate the μ=0 hypothesis:
  - **Normal case**: $\hat{\mu}$~0, so **μ=0 not excluded** : large p-value.

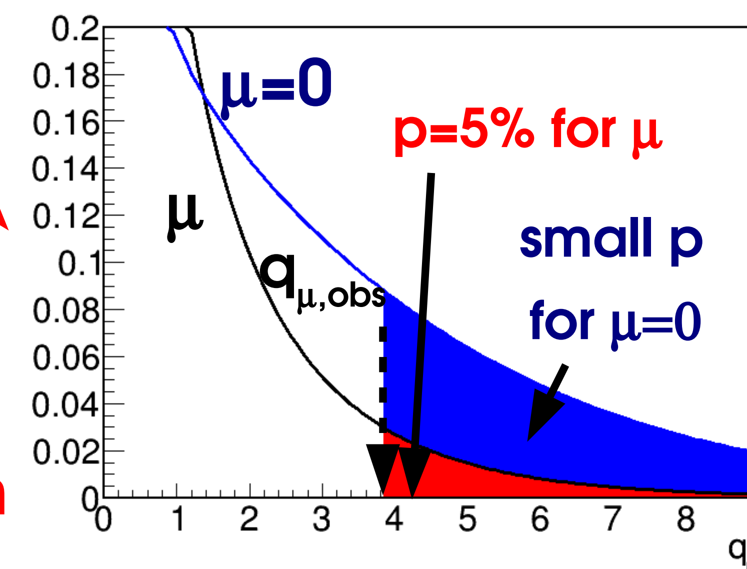




  - **Pathological case**, $\hat{\mu}$<0, so not compatible with μ=0, which is also excluded : p-value for μ=0 also small



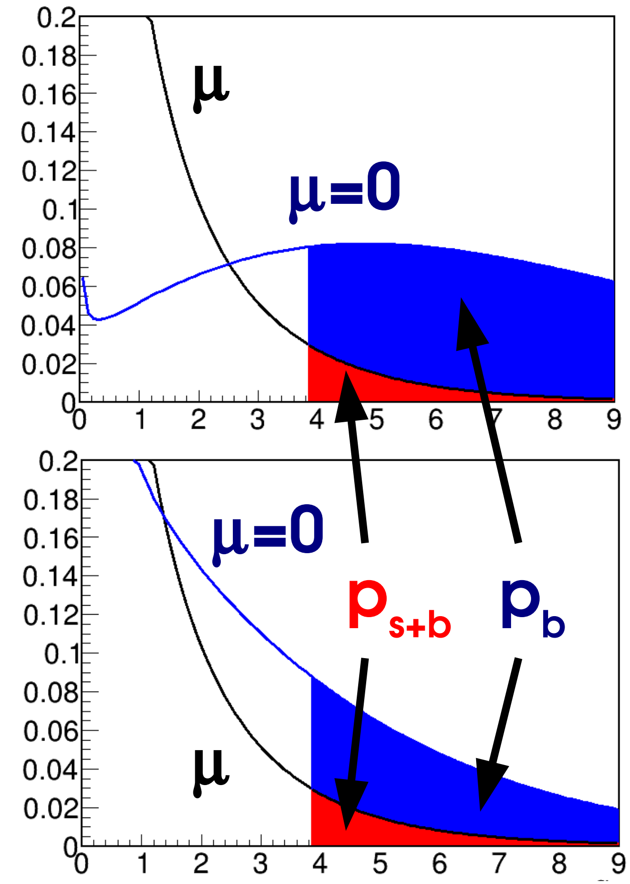


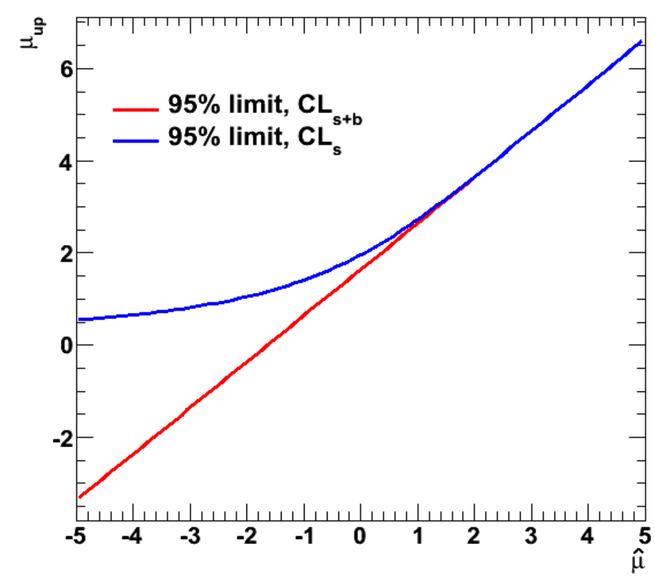


Bad case: **large μ and μ~0 both excluded** : **no sensitivity in $\mu>0$ region !**

CL_s

- Fix method to avoid unphysical limits.
  Usual solution in HEP: **CL_s**.
- Compute modified p-value **p_{s+b} / p_b**
  - **p_{s+b}** is the usual exclusion CL (5%)
  - **p_b** is the p-value for $\mu=0$
  - **Rescale the p_{s+b} by the p-value for $\mu=0$: use $\mu=0$ exclusion as reference**
- **"Good case"** : $p_b \sim 1$
  $CL_s \sim p_{s+b}$ **~ 5%, no change**.
- **"Pathological case"** : $p_b \ll 1$
  $CL_s \sim p_{s+b}/p_b$ **≫ 5%**
- So worse limit, as we wanted
- Drawback: overcoverage (e.g. 98% limit)

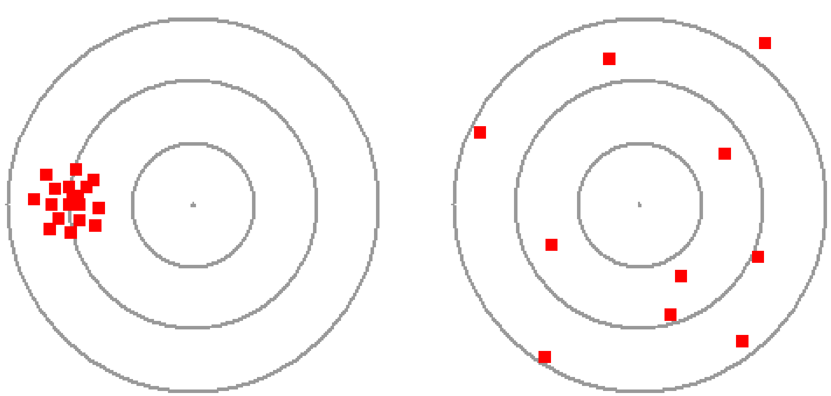




Systematic Uncertainties

- The uncertainties we have dealt with so far are **statistical uncertainties**
  - "Random noise", not correlated between events
  - Decrease with n, usually as $1/\sqrt{n}$.
- **Systematics**
  - Can have anderlying bias in the measurement
    - If we can correct for it, we do
    - If not possible (small, hard to measure,...), treat it as an uncertainty : **systematic uncertainty**
  - Same for all the events : does not improve with more data

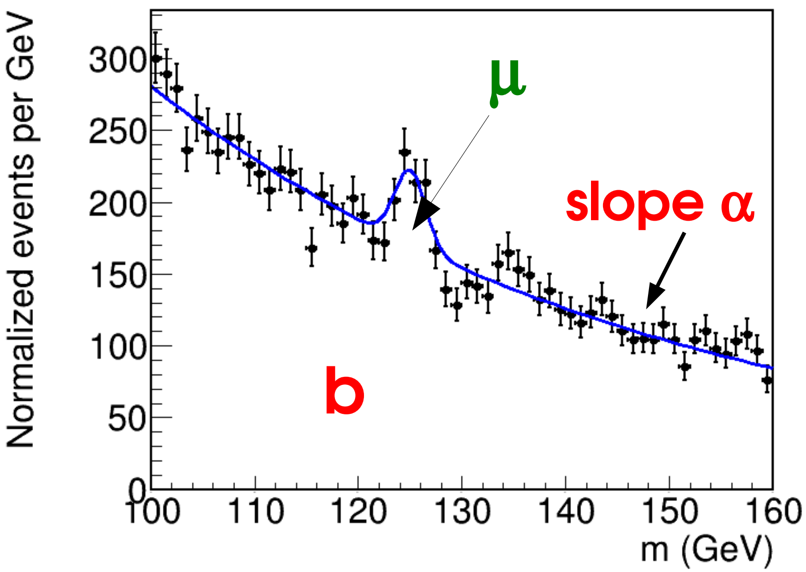
Systematic



Statistical

Nuisance Parameters and Systematics

- Models usually have "**Nuisance parameters**" : not useful but needed to describe the data.
  - e.g. background yield & shape parameters
  - NPs can often be fitted in data
    - **What if not ?**



- **Related issue: systematics**
  - **Example**: signal efficiency in Hgg

    Signal yield : $\mathbf{s = \mu s_{SM} \varepsilon}$ with ε = signal eff, s_{SM} total SM yield

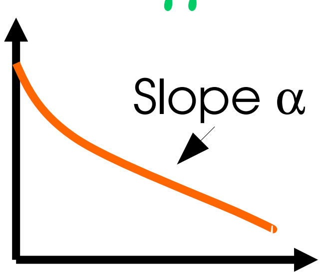
    Get ε from signal MC, but how to account for uncertainty ?
    - **Systematic Uncertainty**, not related to the data itself.
    - Solution : promote ε to a free parameter
    - But ε cannot be fitted from data (degenerate with μ)
- **Each systematic ⇔ a NP – but how to fit their values ?**

Auxiliary measurement

- Solution: get more information!
- Back to efficiency:
  - Often estimated from a "control sample" e.g. $Z\rightarrow ee$ for electrons or photons
  - Separate measurement, but relevant for this result

    => Combine the measurements:
    - $Z\rightarrow ee$ "auxiliary measurement" determines ε
    - $H\rightarrow\gamma\gamma$ main measurement determines the rest
  - Use likelihood combination

$H\rightarrow\gamma\gamma$



$L(\mu, \alpha, \varepsilon; \gamma\gamma \text{ data})$

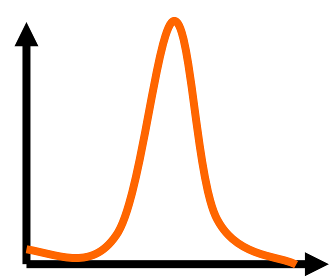
$$L(\gamma\gamma \text{ data}; \mu, \alpha, \varepsilon)\ L(Z \text{ data}; \varepsilon)$$

$$=$$

$$L(\mu, \alpha, \varepsilon; \gamma\gamma\text{+Z data})$$

From $H\rightarrow\gamma\gamma$ From $Z\rightarrow ee$

$Z\rightarrow ee$



$L(\varepsilon; Z \text{ data})$

Systematics In Practice

- **Not practical to include a full auxiliary measurement for each systematic!**
- Include a **simplified** likelihood: $\mathbf{L(\varepsilon;\ Z\ data) = G(\varepsilon_{data};\ \varepsilon,\ \sigma_{\varepsilon})}$

  Central value Uncertainty

- No necessary to know the details of the auxiliary measurement: **central value and uncertainty are usually sufficient**
- Almost always use **Gaussian PDFs** or similar (log-normal), although often not well-motivated
- Can also be used in cases where the auxiliary measurement is hard to define – e.g. **theory uncertainties**

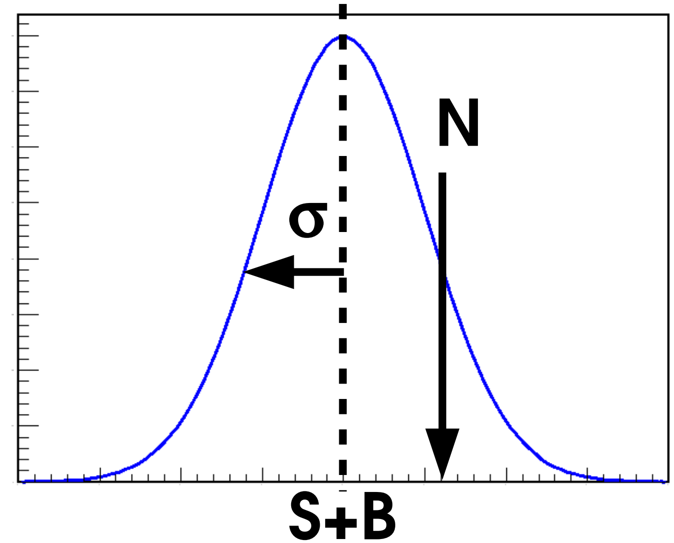
Trivial Example

- Back to our counting experiment with background:

  $L(S, B; n) = G(n; S+B, \sigma_S)$ $\sigma_S \sim \sqrt{B}$

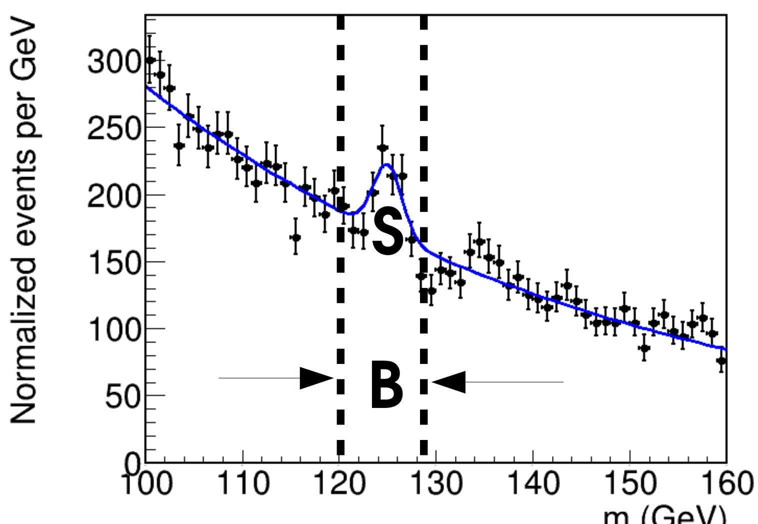
- But assume B is not known a priori, but measured in the sideband

  $L_{SB}(B; n_{SB}) = G(n_{SB}; B, \sigma_{SB})$



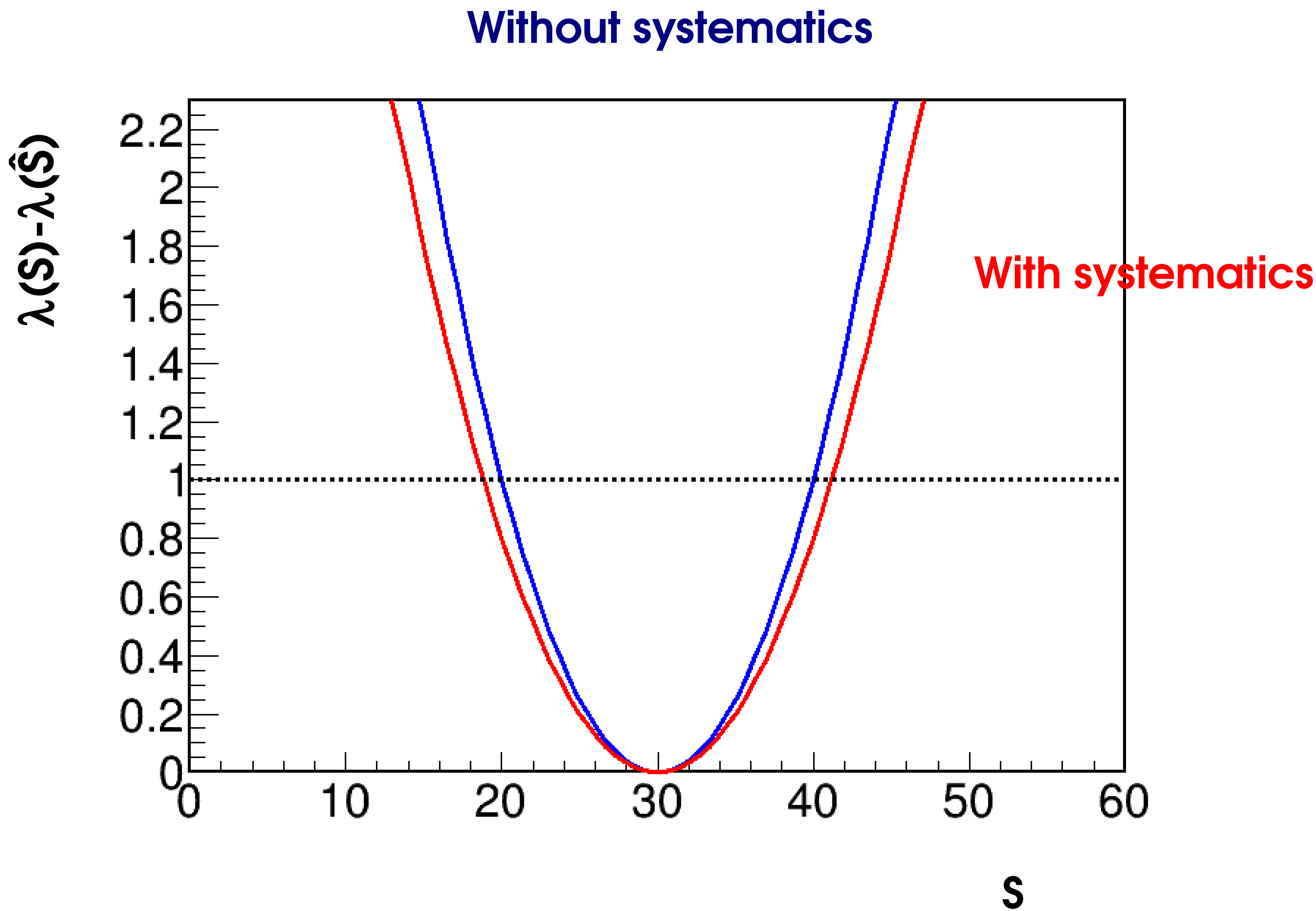
- Combine: $\lambda(S, B; n, n_{SB}) = (n-S-B)^2/\sigma_S^2 + (B-n_{SB})^2/\sigma_{SB}^2$
- $\hat{S}=(n - n_{SB})$, $\hat{B}=n_{SB}$
- Profile B: $$\hat{B}_S = \frac{\sigma_{SB}^2(n-S)+\sigma_S^2 n_{SB}}{\sigma_{SB}^2+\sigma_{SB}^2}$$
- So $\lambda(S)-\lambda(\hat{S}) = (S-\hat{S})^2/\sigma^2$

with $\sigma^2 = \sigma_S^2 + \sigma_{SB}^2$

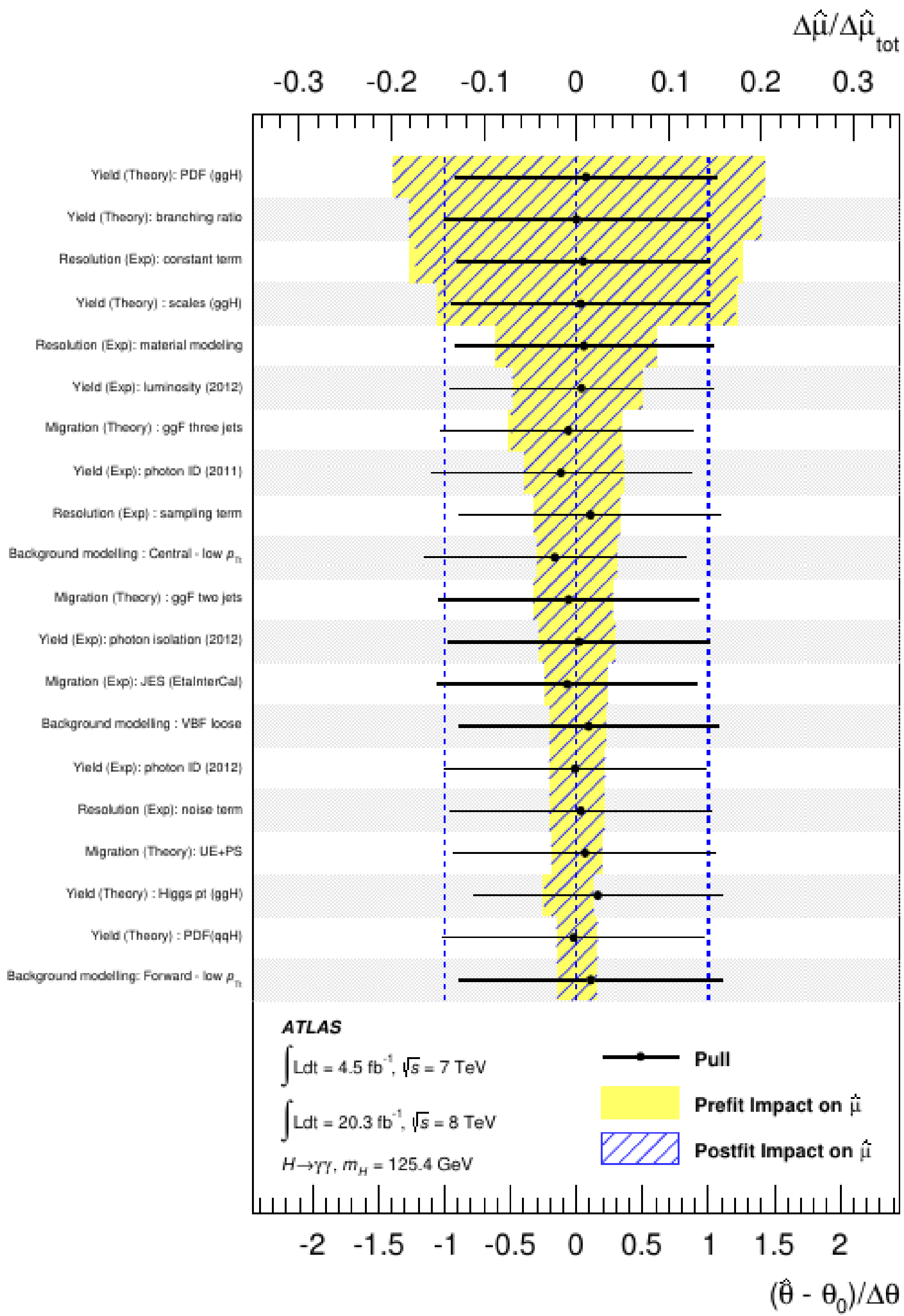


Statistical Uncertainty Systematic Uncertainty

Trivial Example



Real-Life: Systematics from H→γγ



		Syst. source	$N_{\rm NP}$	Implementation
Yield	Theory	Scales	7	$N_{\rm S}^p\, F_{\rm LN}(\sigma_i,\theta_i)$
		PDF+$\alpha_{\rm S}$	2	$N_{\rm S}^p\, F_{\rm LN}(\sigma_i,\theta_i)$
		Br. ratio	1	$N_{\rm S}^{\rm tot}\, F_{\rm LN}(\sigma_i,\theta_i)$
	Exp.	Luminosity	2	$N_{\rm S}^{\rm tot}\, F_{\rm LN}(\sigma_i,\theta_i)$
		Trigger	2	$N_{\rm S}^{\rm tot}\, F_{\rm LN}(\sigma_i,\theta_i)$
		Photon ID	2	$N_{\rm S}^p\, F_{\rm LN}(\sigma_i,\theta_i)$
		Isolation	2	$N_{\rm S}^p\, F_{\rm LN}(\sigma_i,\theta_i)$
MC		MC stats.	14	$N_{\rm S}^p\, F_{\rm G}(\sigma_i^p,\theta_i)$
Migrations	Theory	Jet-bin	2	$N_{\rm S}^{\rm ggF}\, F_{\rm LN}(\sigma_i^{\rm ggF},\theta_i^{\rm ggF})$
		UE+PS	1	$N_{\rm S}^p\, F_{\rm G}(\sigma_i^p,\theta_i)$
		Higgs $p_{\rm T}$	1	$N_{\rm S}^{\rm ggF}\, F_{\rm G}(\sigma_i^{\rm ggF},\theta_i^{\rm ggF})$
		$\Delta\phi_{jj}$	1	$N_{\rm S}^{\rm ggF}\, F_{\rm LN}(\sigma_i^{\rm ggF},\theta_i^{\rm ggF})$
		η^*	1	$N_{\rm S}^{\rm ggF}\, F_{\rm LN}(\sigma_i^{\rm ggF},\theta_i^{\rm ggF})$
		$t\bar{t}H$ model	2	$N_{\rm S}^{t\bar{t}H}\, F_{\rm LN}(\sigma_i^{t\bar{t}H},\theta_i^{t\bar{t}H})$
		HF content	1	$N_{\rm S}^p\, F_{\rm LN}(\sigma_i^p,\theta_i)$
		Scale ($t\bar{t}H$ cat.)	4	$N_{\rm S}^p\, F_{\rm LN}(\sigma_i^{t\bar{t}H},\theta_i^{t\bar{t}H})$
	Exp.	Jet reco.	20	$N_{\rm S}^p\, F_{\rm G}(\sigma_i^p,\theta_i)$
		$E_{\rm T}^{\rm miss}$	5	$N_{\rm S}^p\, F_{\rm G}(\sigma_i^p,\theta_i)$
		b-tagging	13	$N_{\rm S}^p\, F_{\rm G}(\sigma_i^p,\theta_i)$
		Lepton ID+isol.	2	$N_{\rm S}^p\, F_{\rm G}(\sigma_i^p,\theta_i)$
		Lepton isolation	2	$N_{\rm S}^p\, F_{\rm G}(\sigma_i^p,\theta_i)$
Mass		Resolution	4	$\sigma_{\rm CB}\, F_{\rm LN}(\sigma_i,\theta_i)$ $\sigma_{\rm GA}\, F_{\rm LN}(\sigma_i,\theta_i)$
		Scale	43	$\mu_{\rm CB}\, F_{\rm G}(\sigma_i,\theta_i)$ $\mu_{\rm GA}\, F_{\rm G}(\sigma_i,\theta_i)$
Back.		Spurious signal	12	$N_{{\rm spur},c}\,\theta_{{\rm spur},c}$

Log-Normal

Gaussian

Conclusions

- We have reviewed most of the important topics for HEP:
  - **Estimating a parameter**
  - **Computing discovery significances**
  - **Computing Limits**

- Will be put into practice at tutorials on these topics:
  - **Today 16:00 – 18:00**
  - **Tomorrow, same time**