Search and Discovery Statistics in HEP Lecture 2

Eilam Gross, Weizmann Institute of Science

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

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

What can you expect from the Lectures

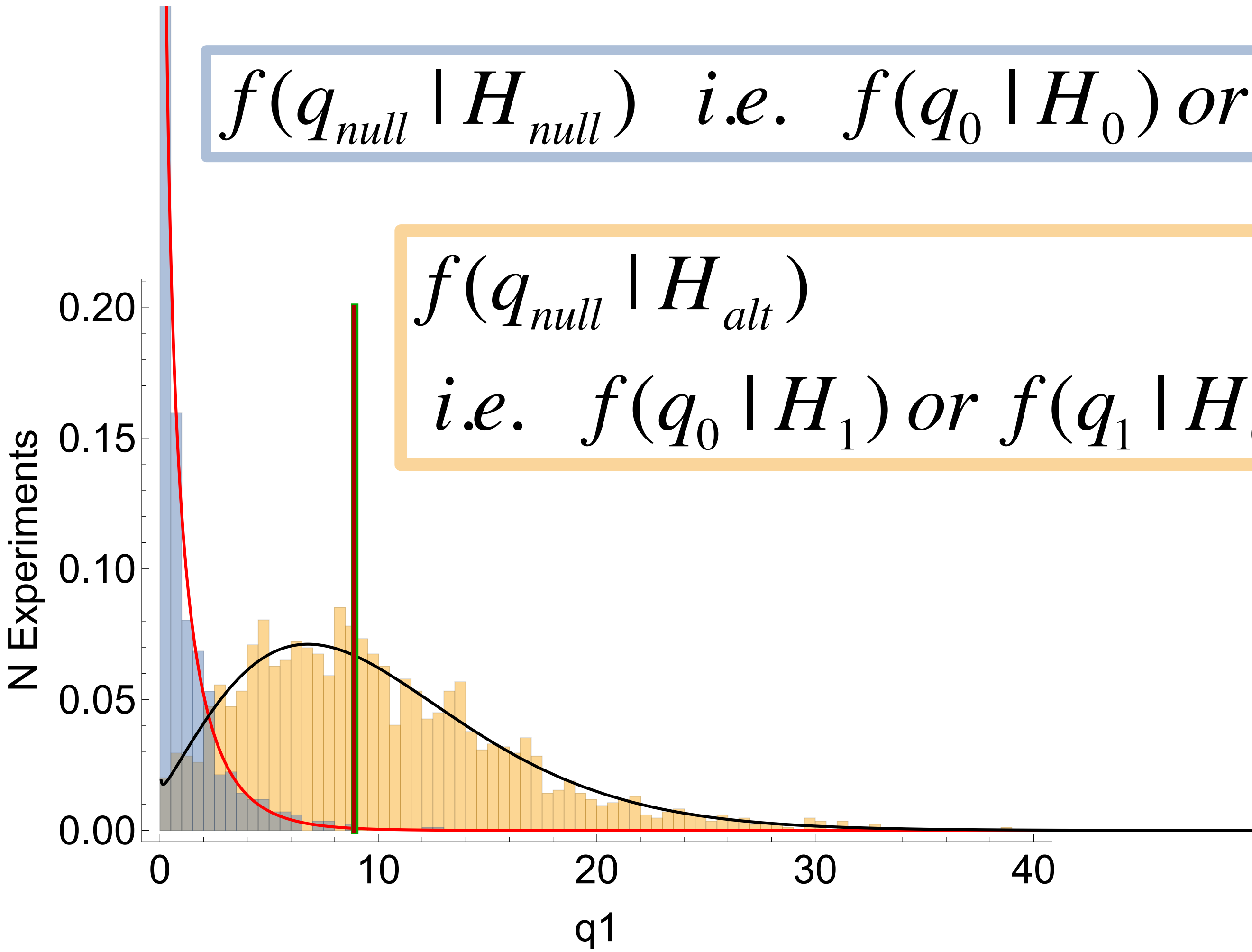
Lecture 1: Basic Concepts
Histograms, PDF, Testing Hypotheses,
LR as a Test Statistics, p-value, POWER, CLs
Measurements

Lecture 2: **Wald Theorem, Asymptotic Formalism, Asimov Data Set, Feldman-Cousins, PL & CLs, Asimov Significance**

Lecture 3: Look Elsewhere Effect
1D LEE the non-intuitive thumb rule
(upcrossings, trial #~Z)
2D LEE (Euler Characteristic)

Lecture 4: Basic Introduction to Deep Learning

This Lecture's Questions



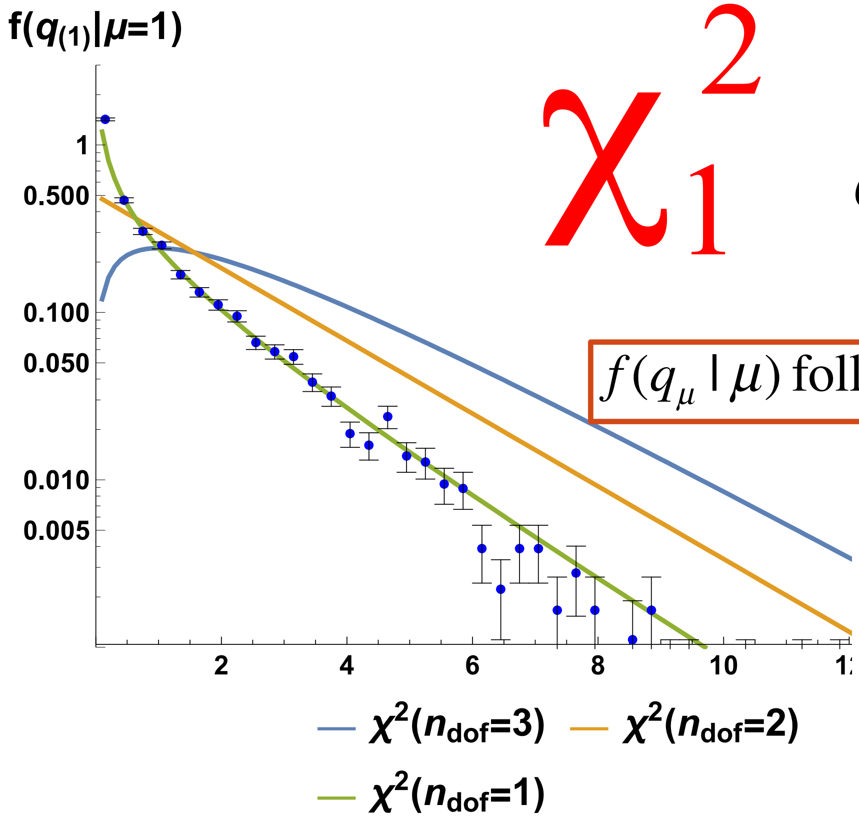
$$f(q_{null} \mid H_{null}) \quad i.e. \quad f(q_0 \mid H_0) \; or \; f(q_1 \mid H_1)$$

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

$$i.e. \quad f(q_0 \mid H_1) \; or \; f(q_1 \mid H_0)$$

Profile Likelihood & Wilks' Theorem

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



$$\chi_1^2$$

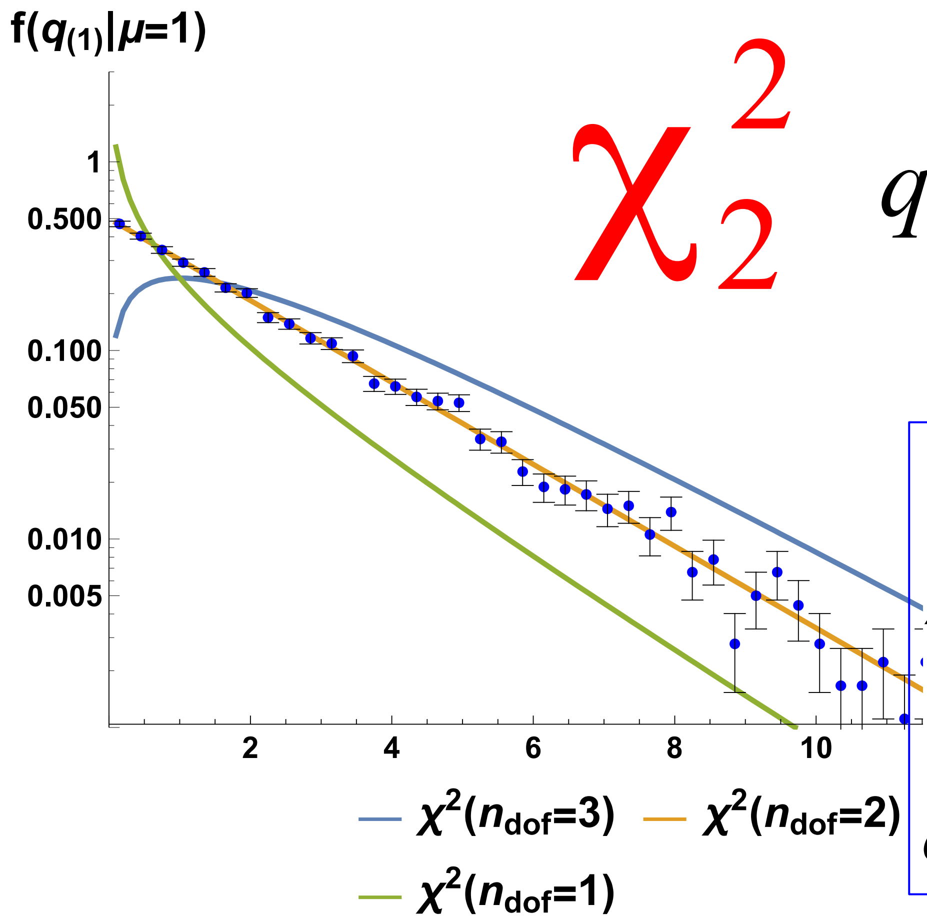
$$q_\mu=-2\ln\frac{L(\mu,\hat{\hat{\epsilon}},\hat{\hat{A}},\hat{\hat{b}})}{L(\hat{\mu},\hat{\epsilon},\hat{A},\hat{b})}$$

$f(q_\mu \mid \mu)$ follows a Chi squared distribution with 1 d.o.f

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

Profile Likelihood & Wilks' Theorem

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



$$\chi^2_2$$

$$q_{\mu,\epsilon}=-2\ln\frac{L(\mu,\epsilon,\hat{\hat{A}},\hat{\hat{b}})}{L(\hat{\mu},\hat{\epsilon},\hat{A},\hat{b})}$$

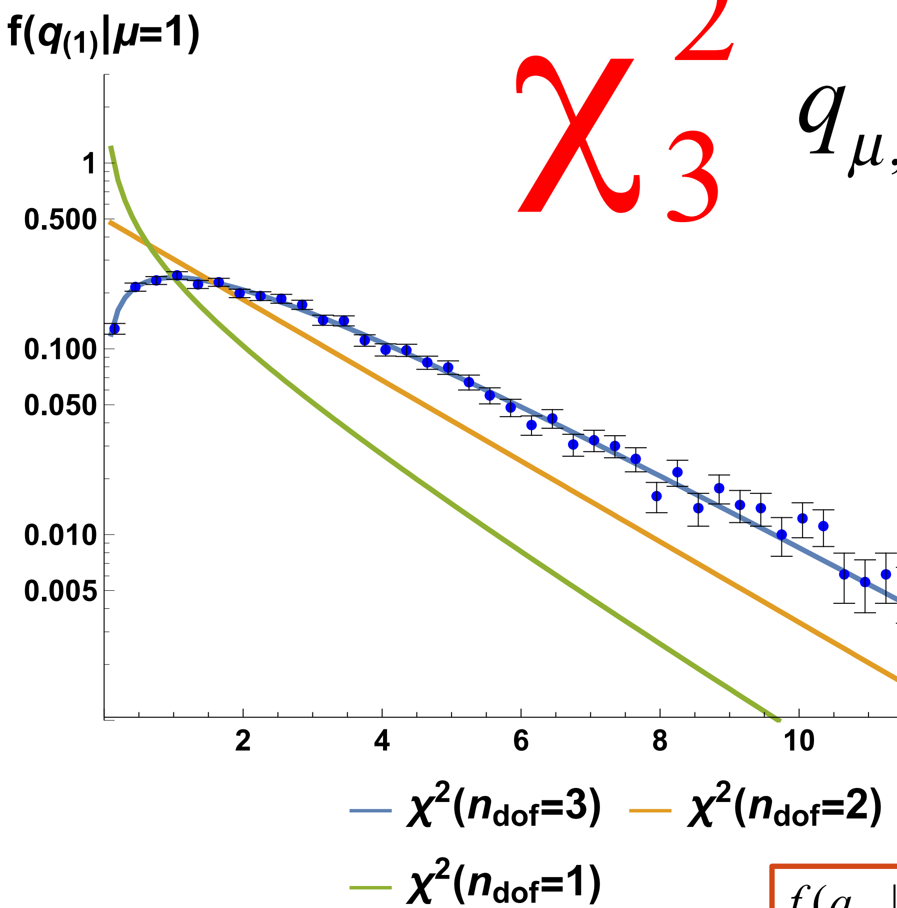
$$\lambda(\alpha_i)=\frac{L(\alpha_i,\hat{\hat{\theta}}_j)}{L(\hat{\alpha}_i,\hat{\theta}_j)}$$

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

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

Profile Likelihood & Wilks' Theorem

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



$$\chi_3^2$$

$$q_{\mu,\epsilon,A}=-2\ln\frac{L(\mu,\epsilon,A,\hat{\hat{b}})}{L(\hat{\mu},\hat{\epsilon},\hat{A},\hat{b})}$$

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

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

$f(q_{\alpha_i}|\alpha_i)$ follows a Chi squared distribution with n d.o.f

$n=\#$ pars of interest

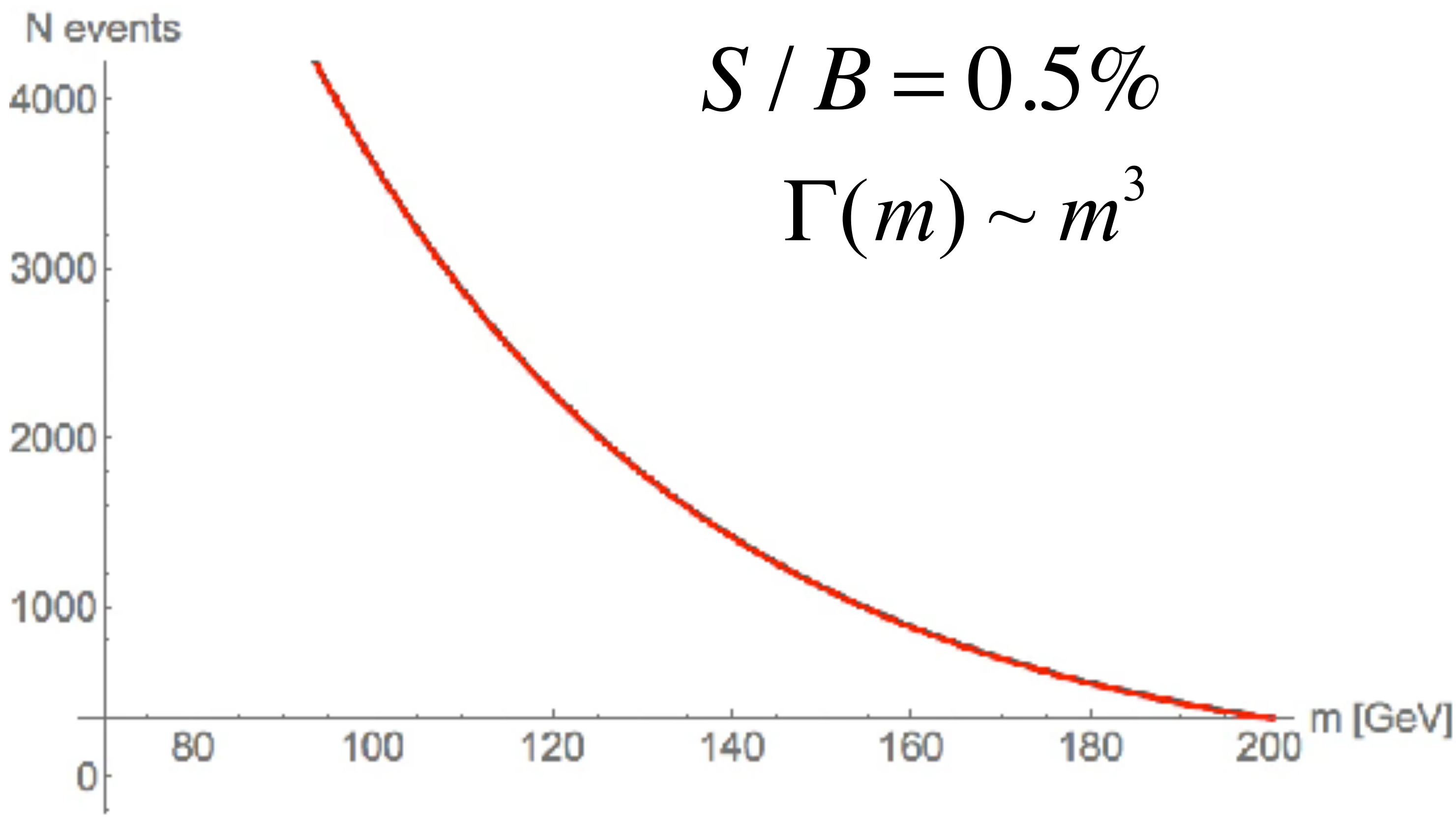
Classification of Test Statistics

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

Study Case 2: Bump Hunt

Bump Hunt

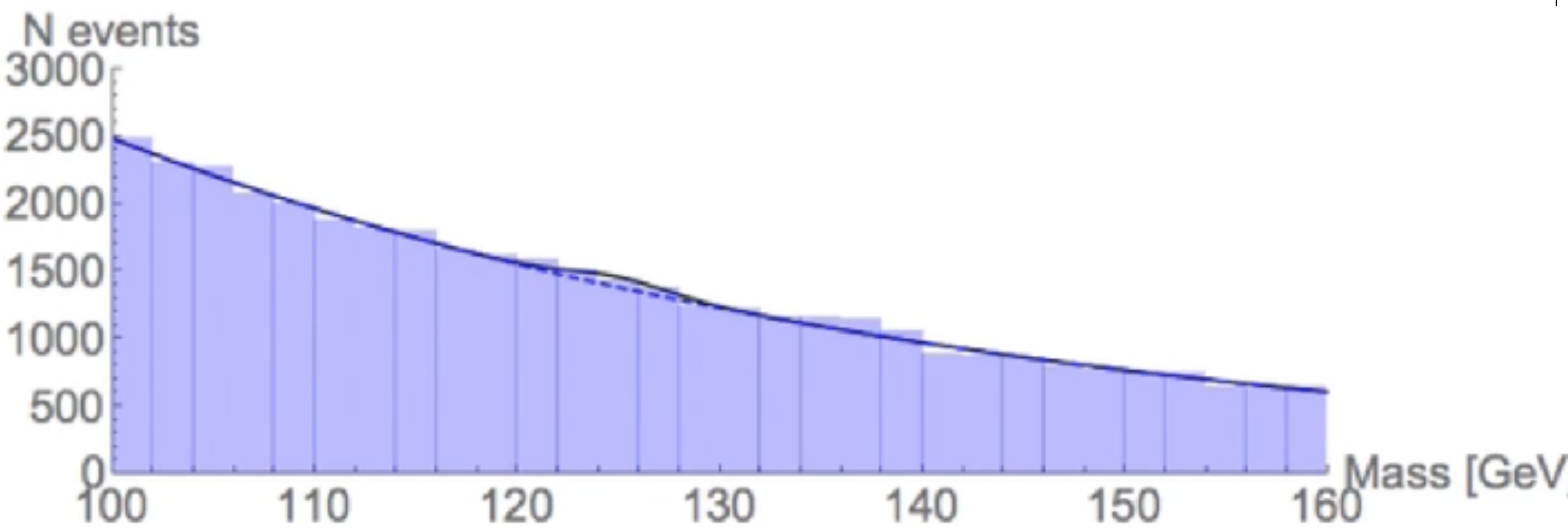
Gamma Gamma like BG and a Gaussian signal on top of it



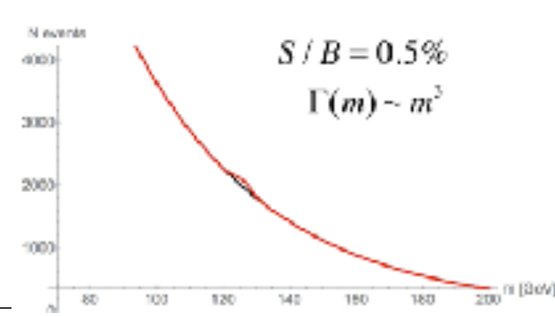
$$S / B = 0.5\%$$

$$\Gamma(m) \sim m^3$$

A GammaGammaLike Signal

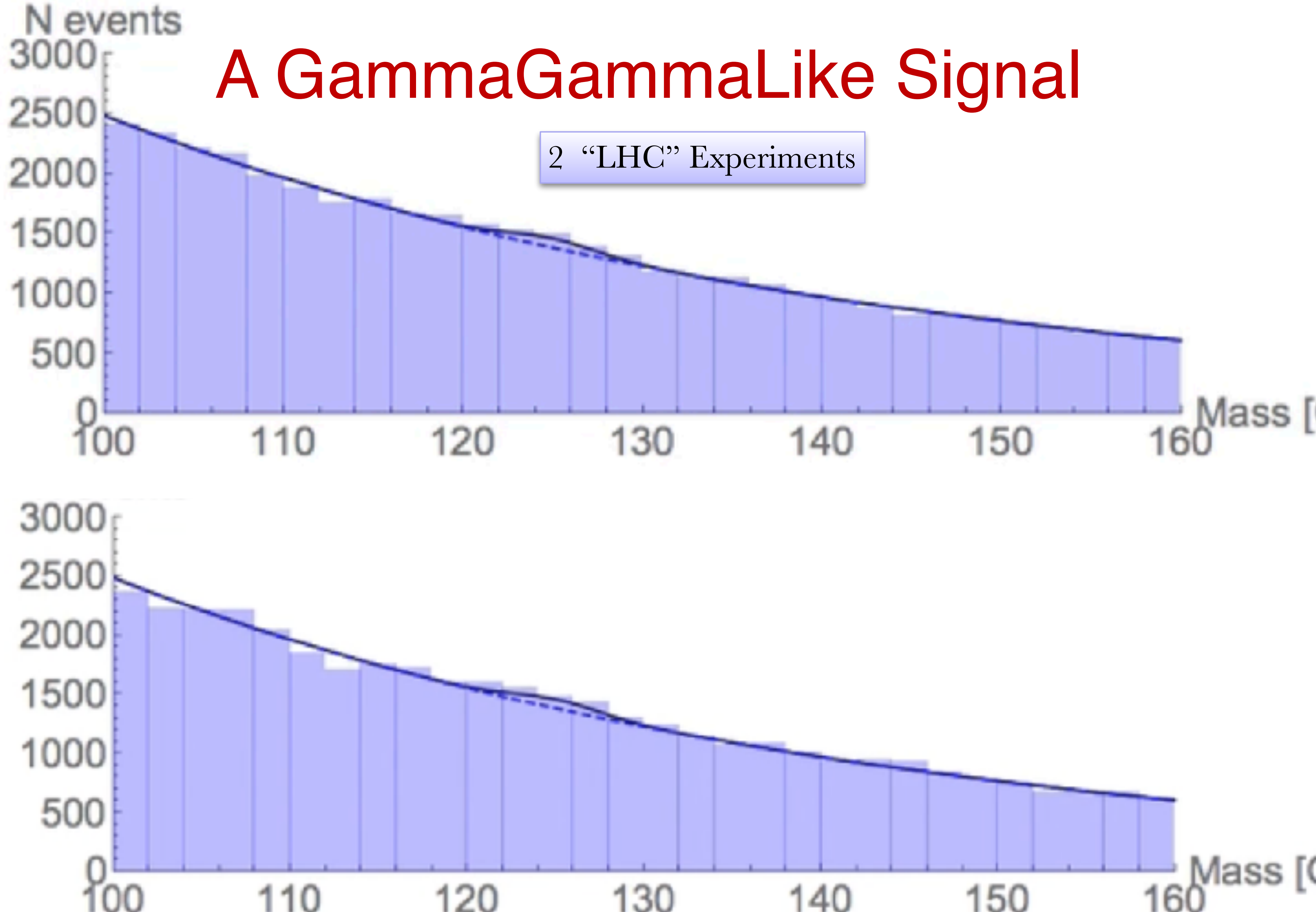


Luminosity is the number of events in the histogram



A GammaGammaLike Signal

2 "LHC" Experiments



Bump Hunt

$$Test\ H_0\ \ with\ q_0,\ Reject\ H_0 \Rightarrow Discovery$$

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases} \qquad \lambda(0) = \frac{L(0,\hat{\hat{\theta}}_0)}{L(\hat{\mu},\hat{\theta})}$$

$$Test\ H_\mu(m_H)\ \ with\ q_\mu\ Reject\ H_\mu(m_H) \Rightarrow$$

$$Exclusion\ of\ a\ Higgs\ with\ m_H \Rightarrow \mu_{up}(m_H)$$

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \qquad \lambda(\mu) = \frac{L(\mu,\hat{\hat{\theta}}_\mu)}{L(\hat{\mu},\hat{\theta})}$$

Asymptotic Approximation

CCGV

Asymptotic formulae for likelihood-based tests of new physics
Glen Cowan (Royal Holloway, U. of London), Kyle Cranmer (New York U.), Eilam Gross, Ofer Vitells (Weizmann Inst.). Jul 10, 2010. 25 pp.
Published in **Eur.Phys.J. C71 (2011) 1554**, Erratum: **Eur.Phys.J. C73 (2013) 2501**

Test Statistic $t_\mu = -2ln\lambda(\mu)$

$$t_\mu = -2ln\lambda(\mu) \qquad \lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}}_\mu)}{L(\hat{\mu}, \hat{\theta})}$$

Higher values of t_μ *correspond to increasing incompatibility between the data and* μ

Wald Theorem

$$\lambda(\mu) = \frac{L(\mu,\hat{\hat{\theta}}_\mu)}{L(\hat{\mu},\hat{\theta})} \qquad t_\mu = -2ln\lambda(\mu) \qquad Wilks \Rightarrow f(t_\mu \mid \mu) \sim \chi_1^2$$

How does t_μ *distributes under* $H_{\mu'}$ $(\mu' \neq \mu)$

A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*, Transactions of the American Mathematical Society, Vol. **54**, No. 3 (Nov., 1943), pp. 426-482.

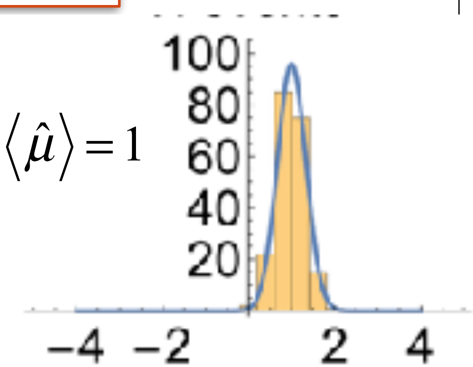
$$t_\mu = -2ln\lambda(\mu) = \frac{(\mu-\hat{\mu})^2}{\sigma_{\hat{\mu}}^2} + O\left(1/\sqrt{N}\right)$$

where $\hat{\mu} \sim G(\mu',\sigma)$

N is the sample size

$(Use\ the\ Asimov\ Dataset\ to\ estimate\ \sigma)$

$\mu' = 1 \Rightarrow \langle\hat{\mu}\rangle = 1$



$f(t_\mu \mid \mu')$ follows a noncentral Chi squared distribution

with non-centrality parameter $\Lambda = \frac{(\mu-\mu')^2}{\sigma^2}$ with 1 d.o.f

Wald Theorem

$$t_\mu = -2ln\lambda(\mu) = \frac{(\mu-\hat{\mu})^2}{\sigma_{\hat{\mu}}^2} + O\left(1/\sqrt{N}\right)$$

$$\hat{\mu} \sim G(\mu',\sigma)$$

N is the sample size

$f(t_\mu \mid \mu')$ follows a noncentral Chi squared distribution

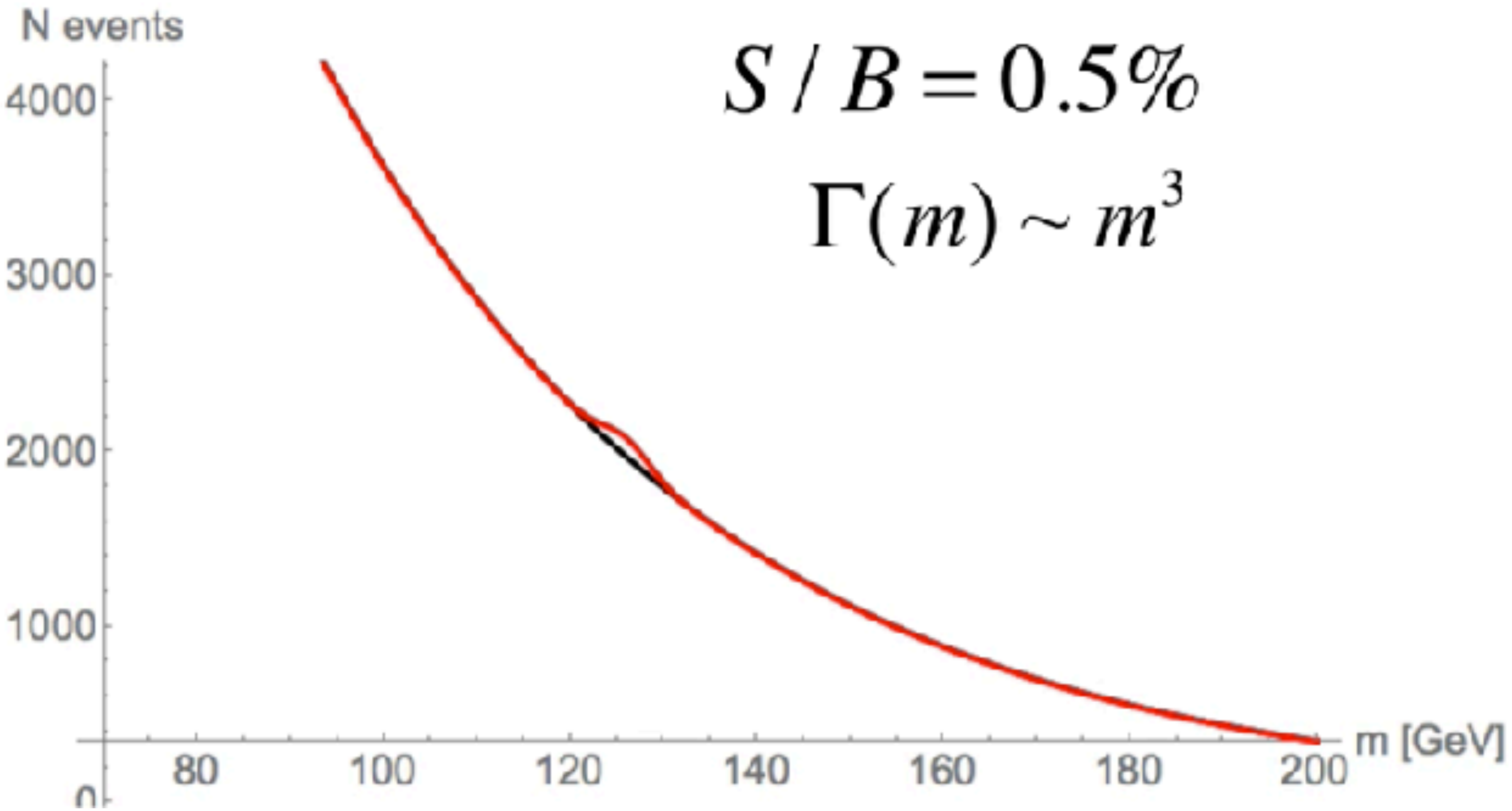
with non-centrality parameter $\Lambda = \frac{(\mu-\mu')^2}{\sigma^2}$ with 1 d.o.f

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

for $\mu' = \mu$ we retrieve Wilks' theorem

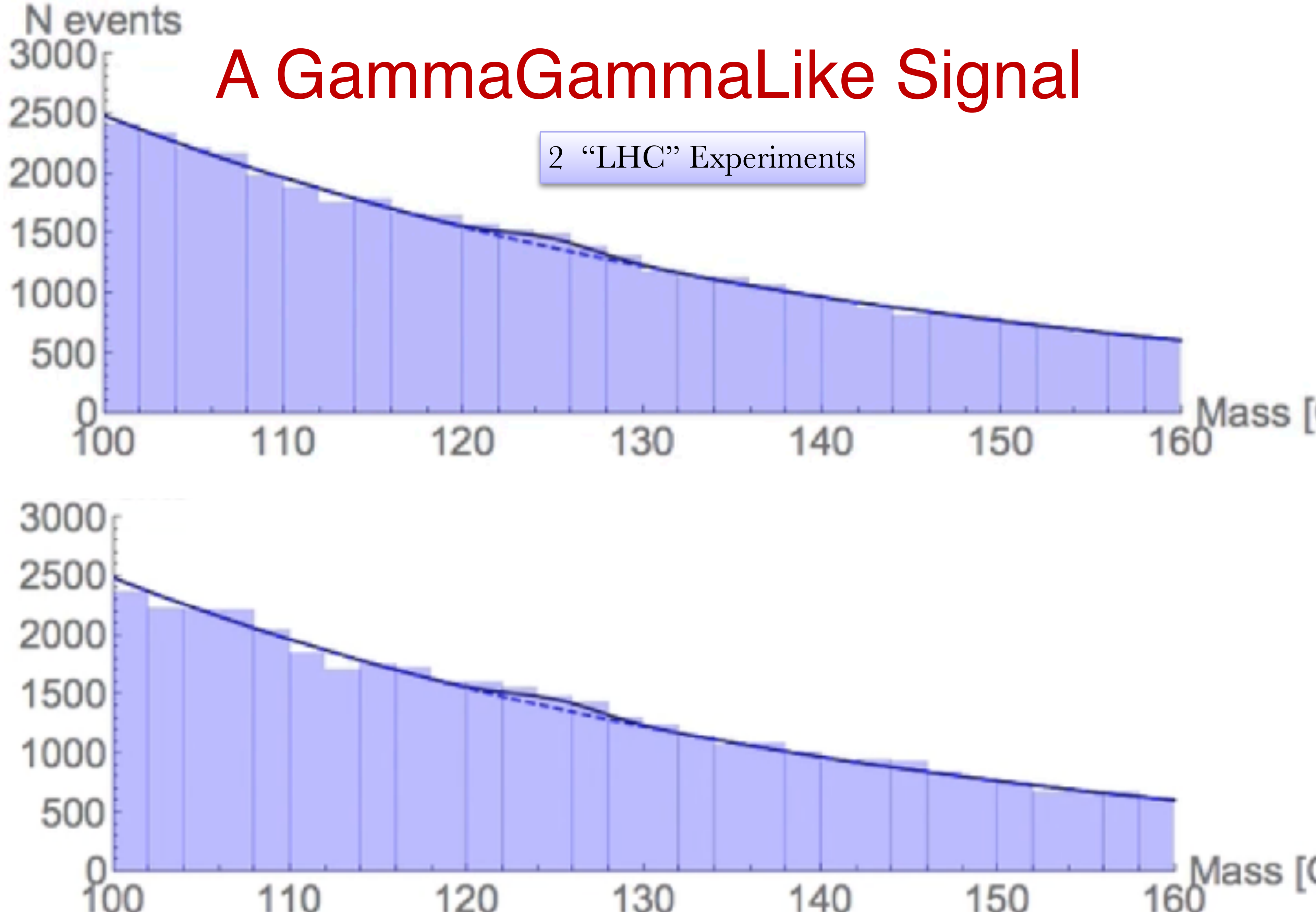
$$f(t_\mu) = \frac{1}{\sqrt{2\pi t_\mu}}e^{-\frac{1}{2}t_\mu} = \chi^2$$

A GammaGammaLike Signal



A GammaGammaLike Signal

2 "LHC" Experiments

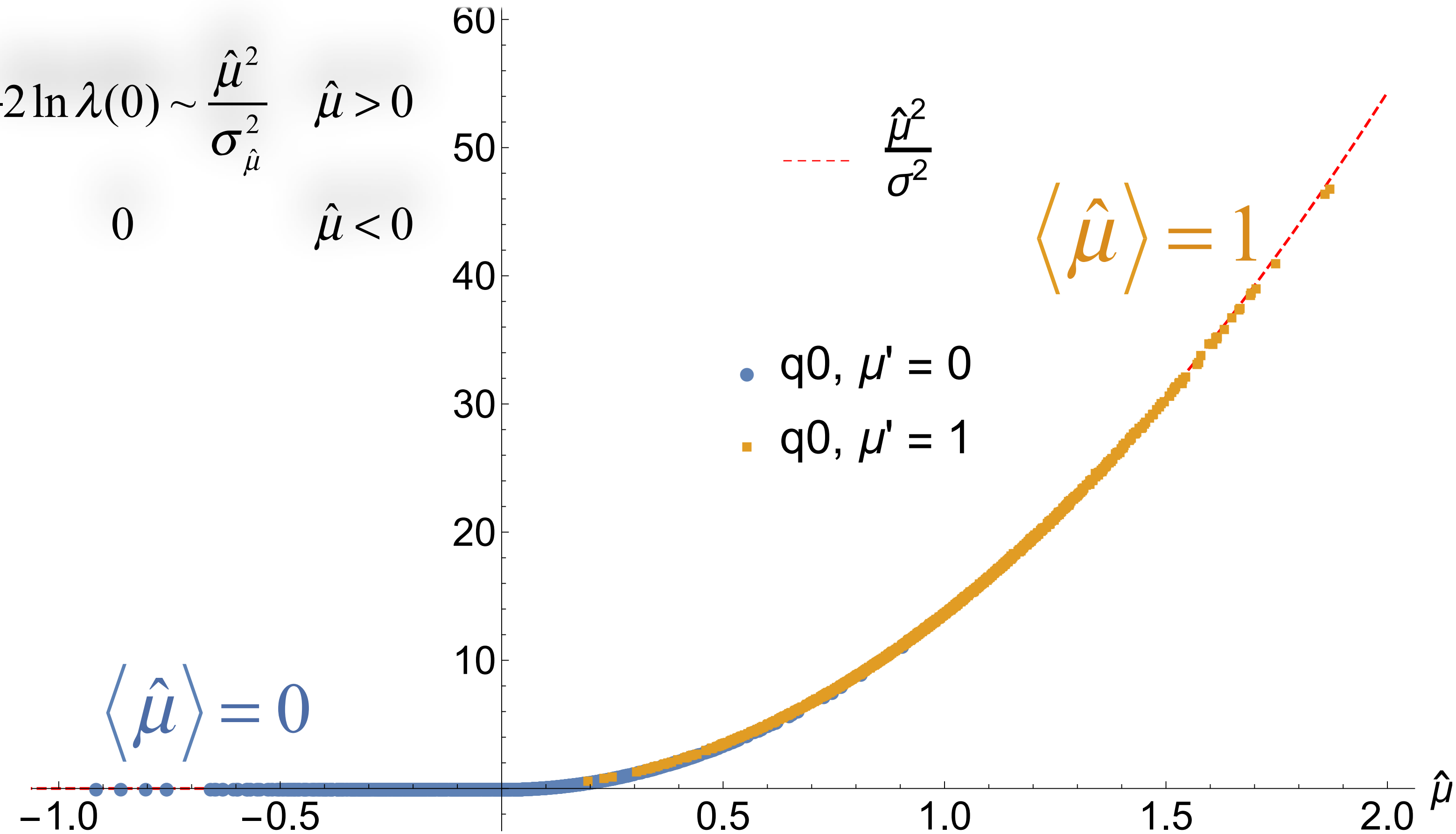


Wald Theorem Demonstration

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

$$\mathcal{L} = 60000$$

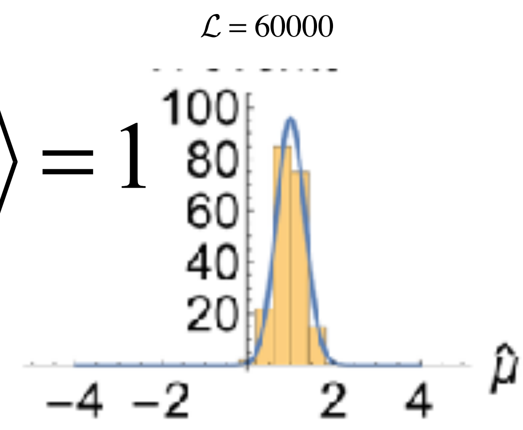
$$q_0(\hat{\mu}) = \begin{cases} -2\ln\lambda(0) \sim \dfrac{\hat{\mu}^2}{\sigma_{\hat{\mu}}^2} & \hat{\mu} > 0 \\ 0 & \hat{\mu} < 0 \end{cases}$$



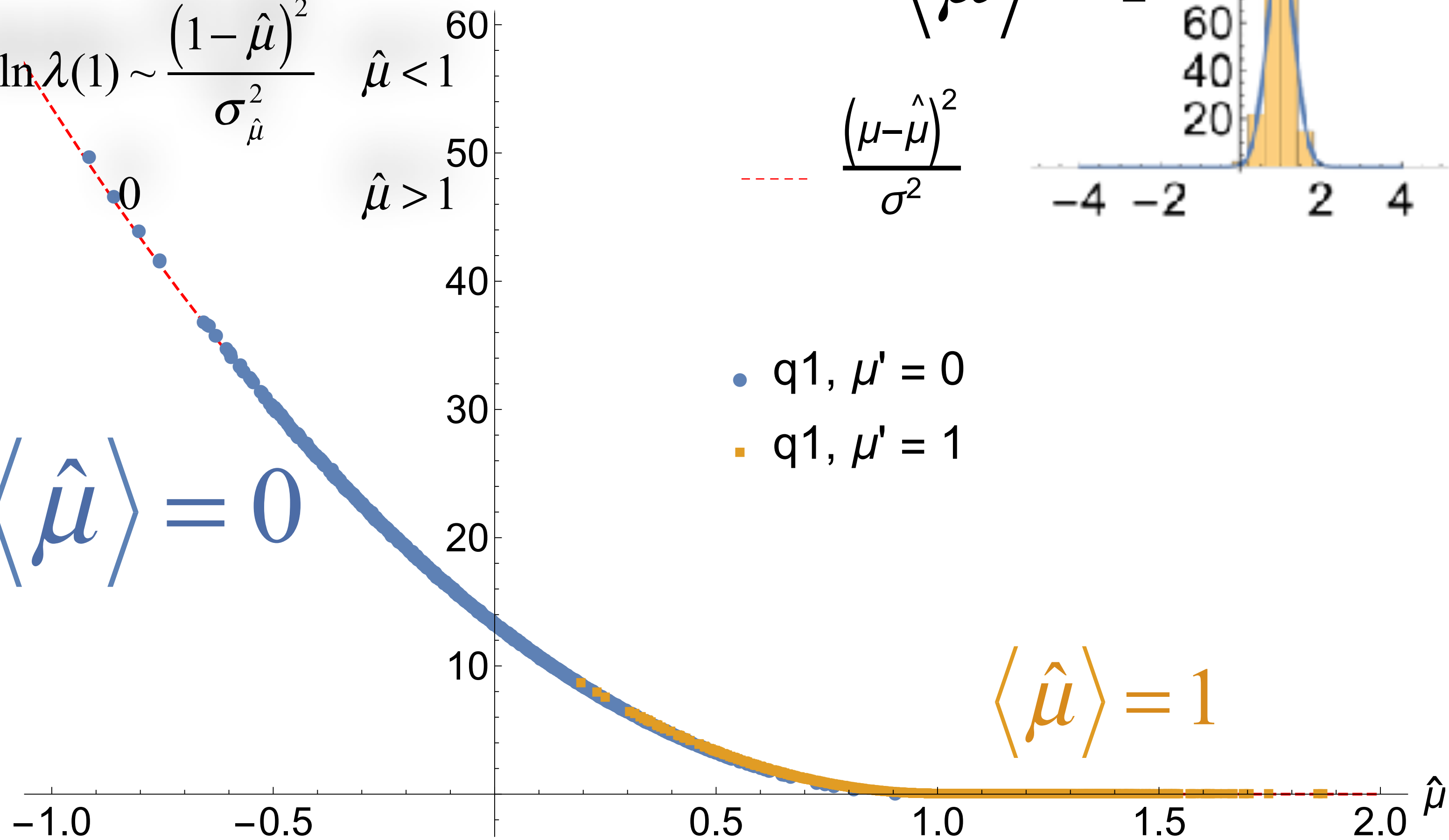
Wald Theorem Demonstration

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

$$q_1(\hat{\mu}) = \begin{cases} -2\ln\lambda(1) \sim \dfrac{(1-\hat{\mu})^2}{\sigma_{\hat{\mu}}^2} & \hat{\mu} < 1 \\ 0 & \hat{\mu} > 1 \end{cases}$$

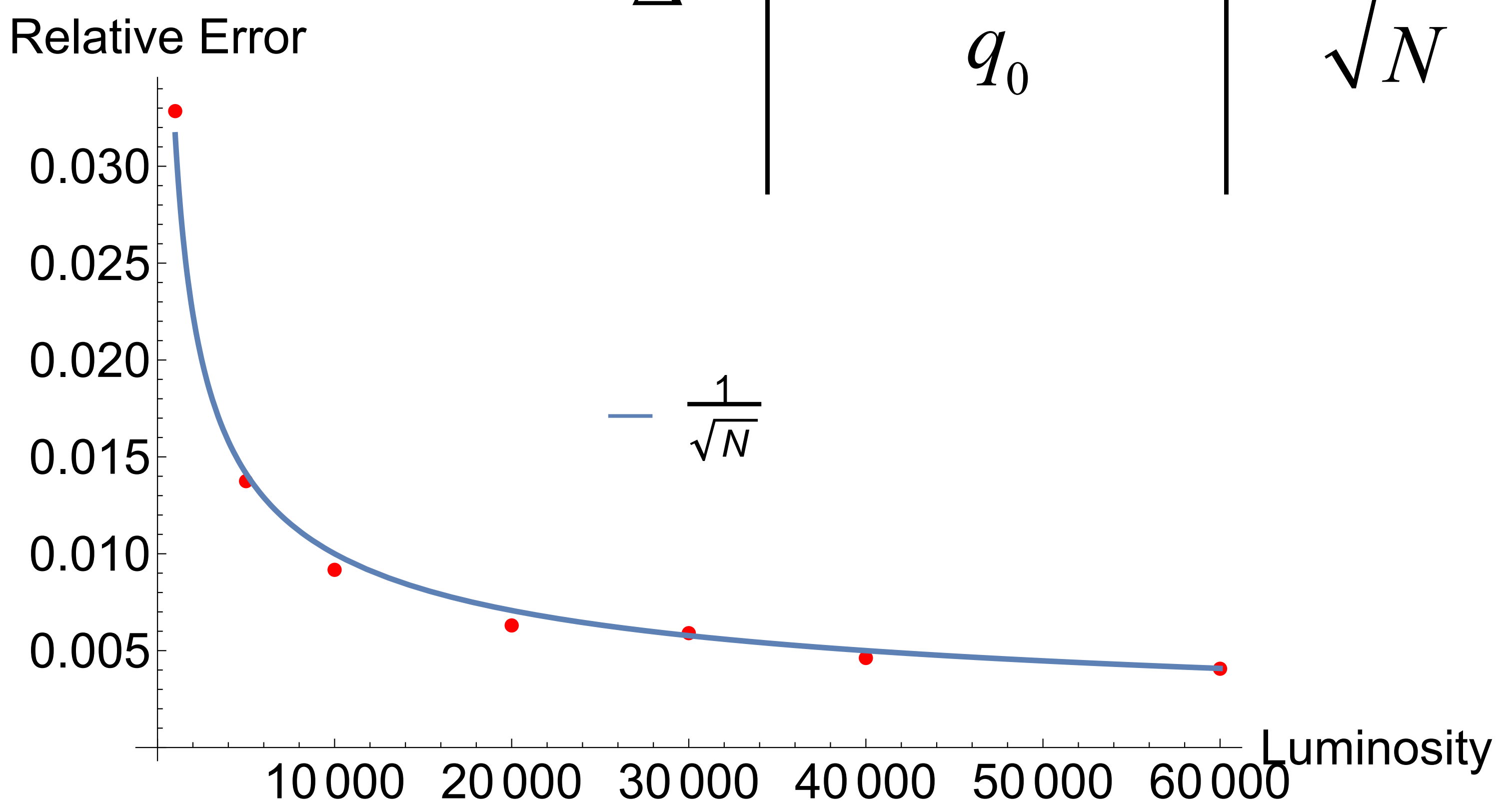


$\langle \hat{\mu} \rangle = 1$

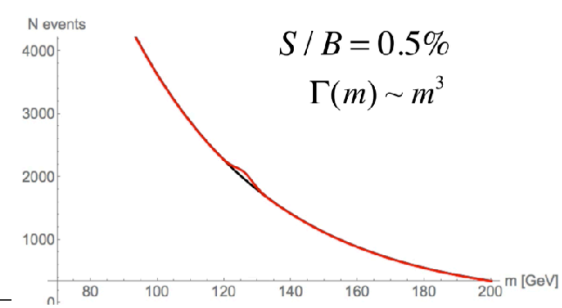


Wald Theorem

$$\Delta = \left| \frac{q_0 - \frac{(\mu - \hat{\mu})^2}{\sigma^2}}{q_0} \right| \sim \frac{1}{\sqrt{N}}$$

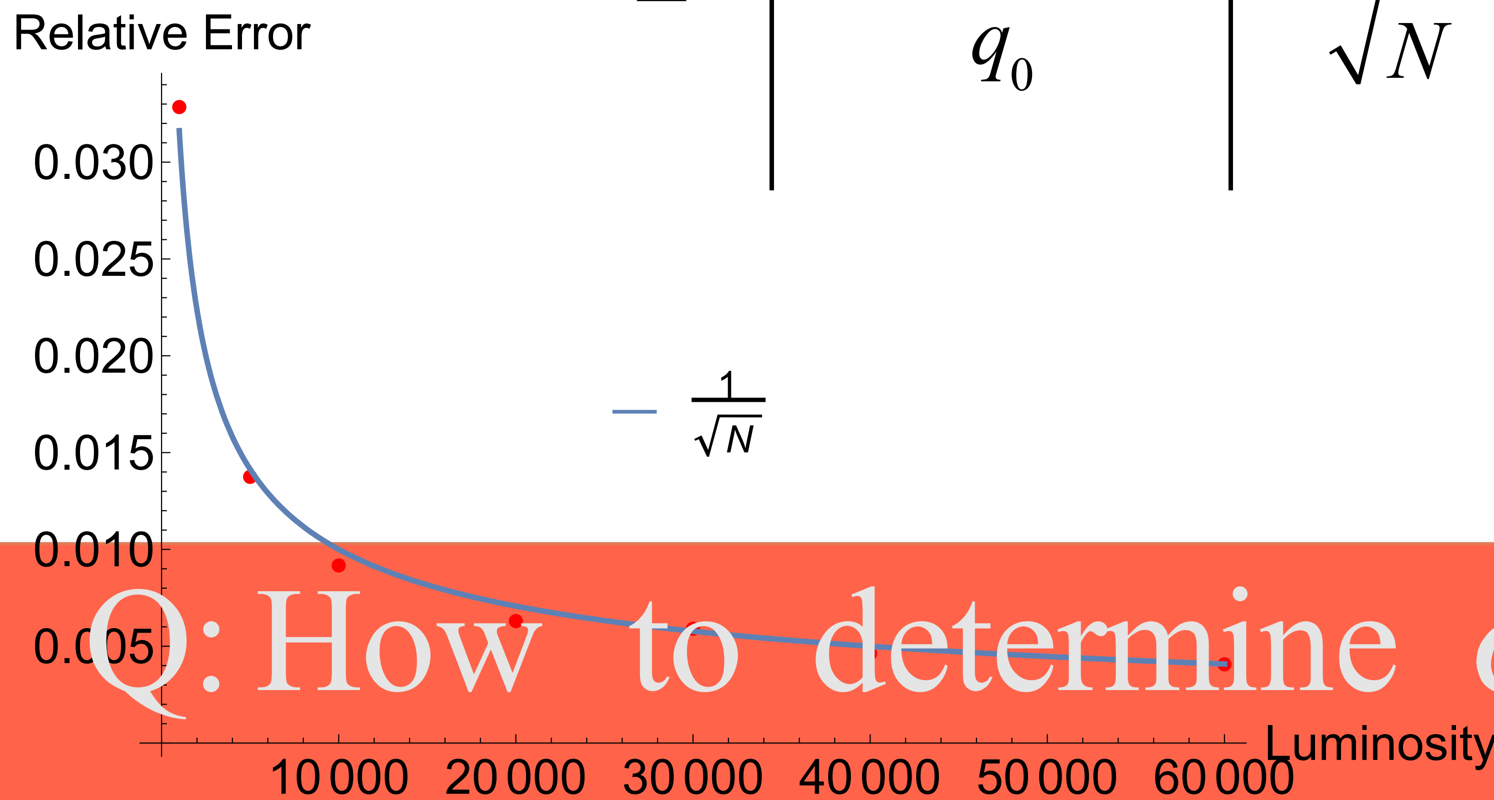


Luminosity is the number of events in the histogram



Wald Theorem

$$\Delta = \left| \frac{q_0 - \frac{(\mu-\hat{\mu})^2}{\sigma^2}}{q_0} \right| \sim \frac{1}{\sqrt{N}}$$

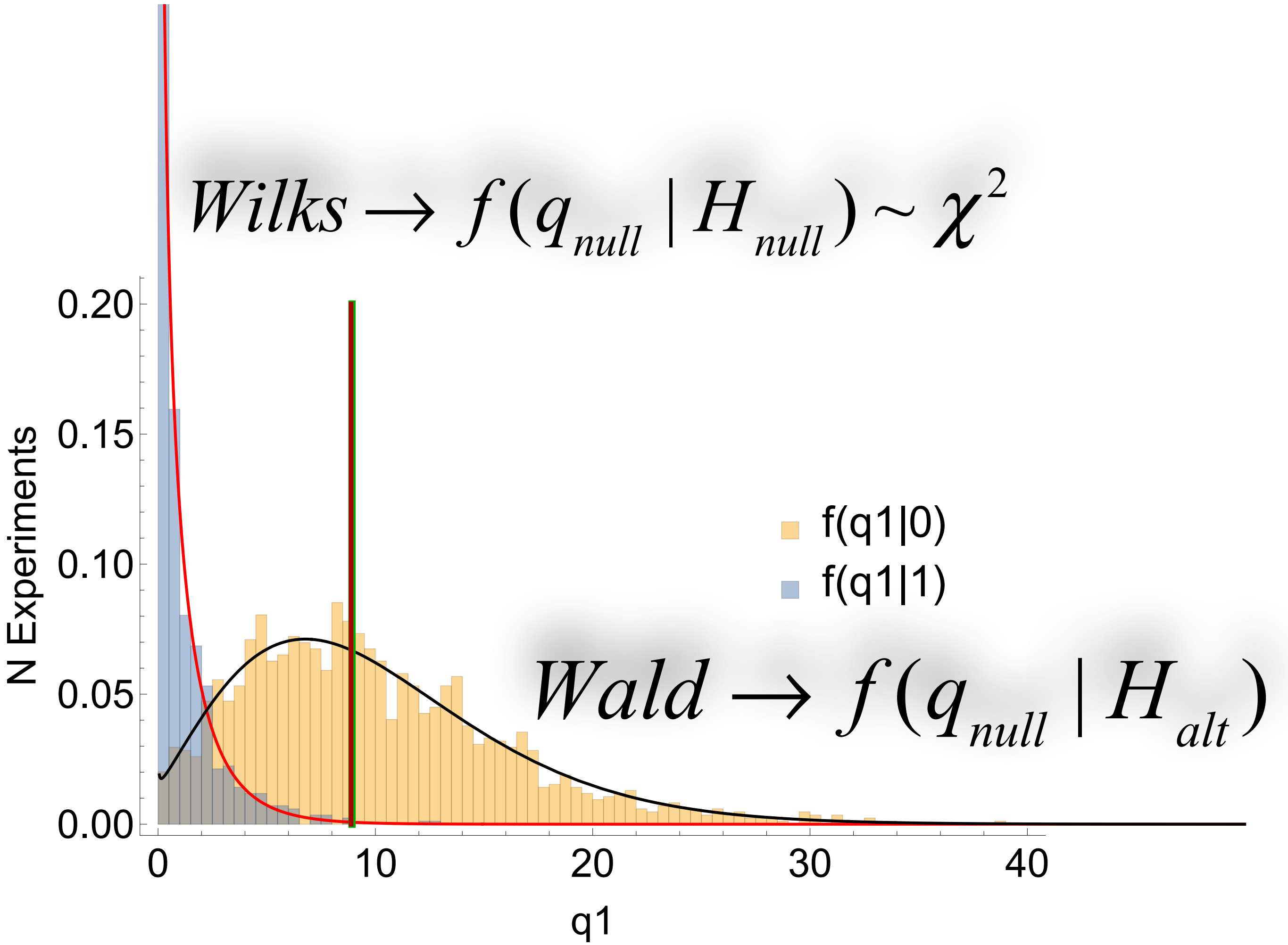


Q: How to determine σ?

Luminosity is the number of events in the histogram

A: With the Asimov DATA

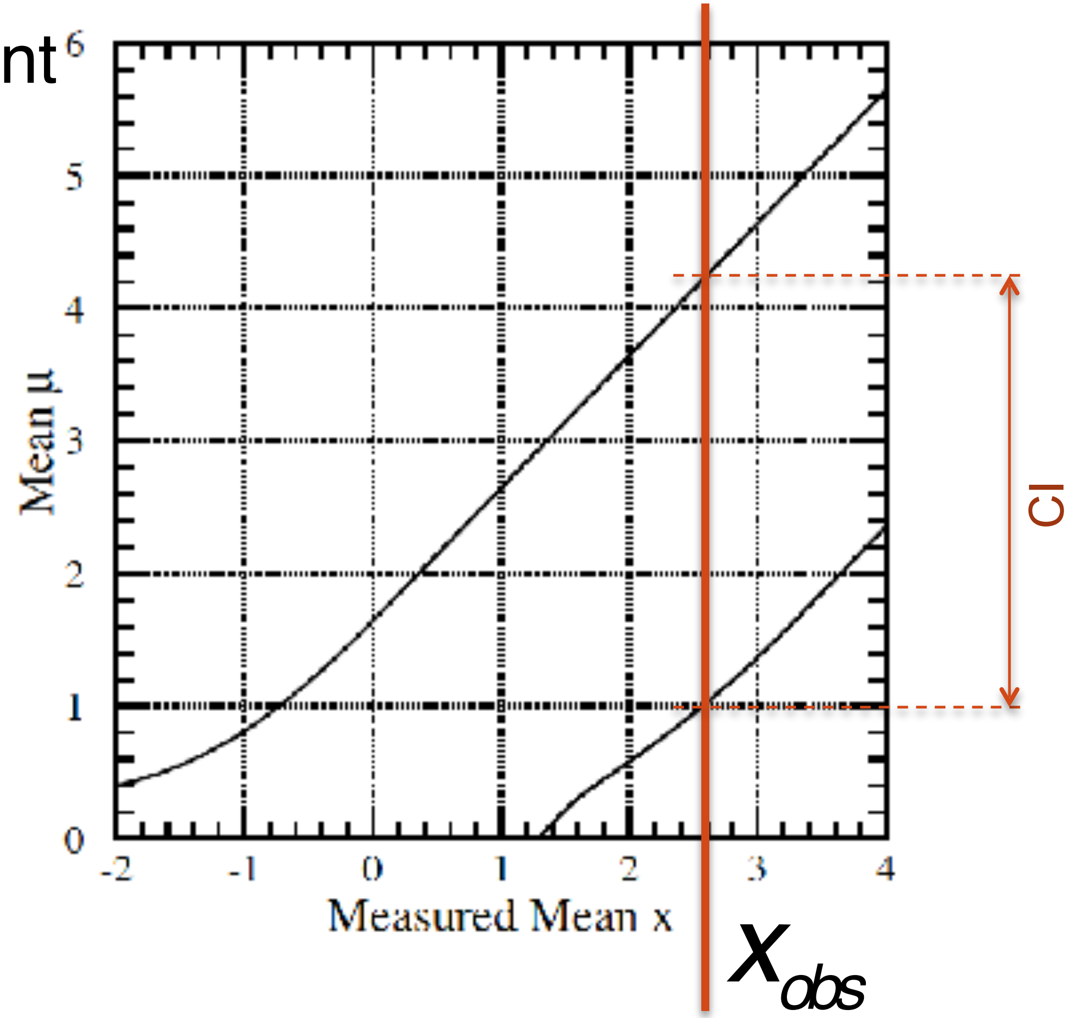
Asymptotics



The Feldman Cousins Unified Method - Take 2

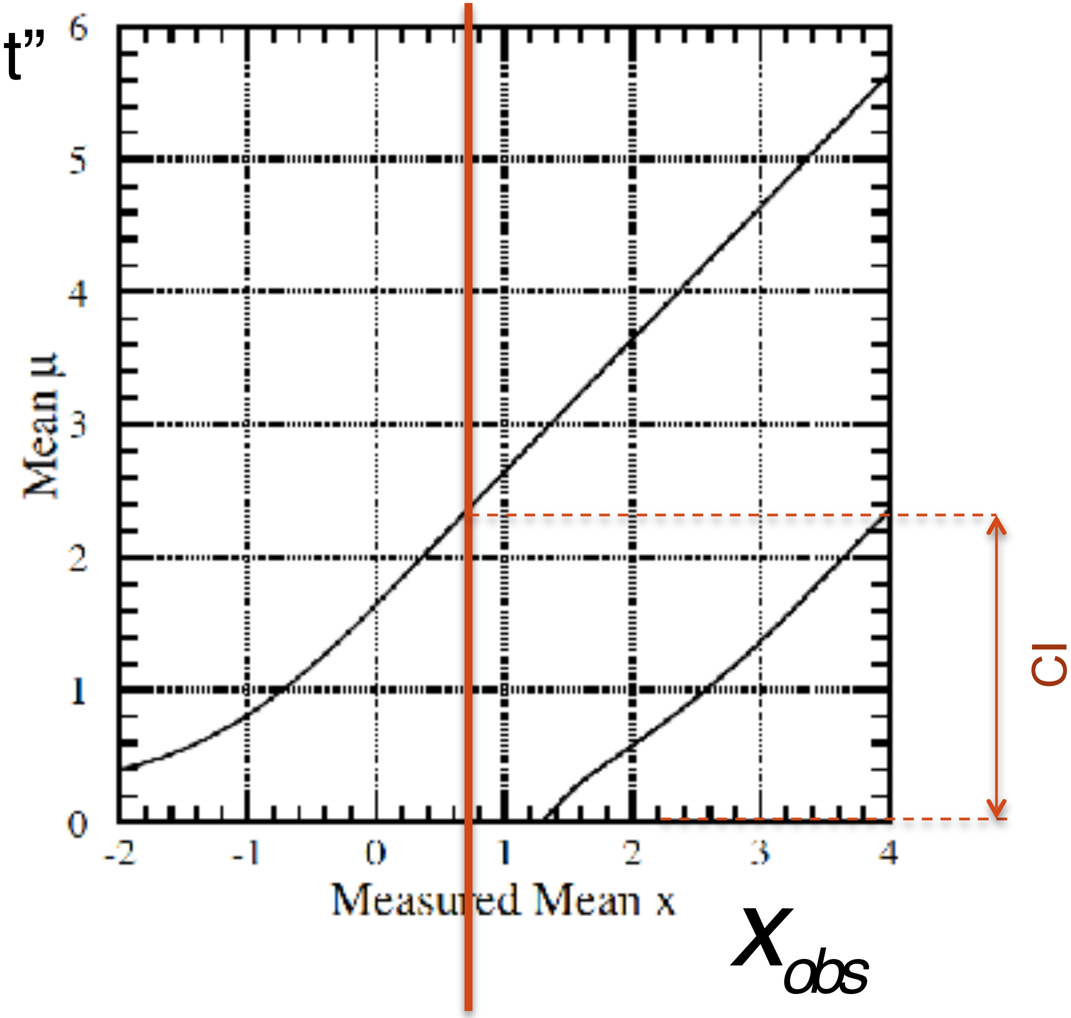
How to tell an Upper limit from a Measurement without Flip Flopping

- A measurement (2 sided)



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

- An "upper limit" (1 sided)



$Asymptotic\ Feldman - Cousins$

$\left[\tilde{t}_{\mu} \text{ for } \mu \geq 0\right]$

CCGV

Feldman Cousins - Asymptotic

If $\mu \geq 0$ *due to physics constraints, for* $\hat{\mu} < 0$ *the best agreement between data and the physical* μ *is* $\hat{\mu} = 0$. *We define*

$$\tilde{t}_\mu \equiv -2log\left(\tilde{\lambda}(\mu)\right) \qquad \tilde{\lambda}(\mu) \equiv \begin{cases} \frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(\hat{\mu},\hat{\theta})} & \hat{\mu} \geq 0 \\ \frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0 \end{cases}$$

$Wald \rightarrow$

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

Feldman Cousins - Asymptotic

$$f(\tilde{t}_\mu|\mu) = \begin{cases} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\tilde{t}_\mu}} e^{-\tilde{t}_\mu/2} & \tilde{t}_\mu \le \frac{\mu^2}{\sigma^2} \\ \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\tilde{t}_\mu}} e^{-\tilde{t}_\mu/2} + \frac{1}{\sqrt{2\pi}(2\mu/\sigma)} exp\left[-\frac{1}{2} \frac{(\tilde{t}_\mu + \mu^2/\sigma^2)^2}{(2\mu/\sigma)^2}\right] & \tilde{t}_\mu > \frac{\mu^2}{\sigma^2} \end{cases}$$

$$p_\mu = 1 - F(\tilde{t}_\mu|\mu)$$

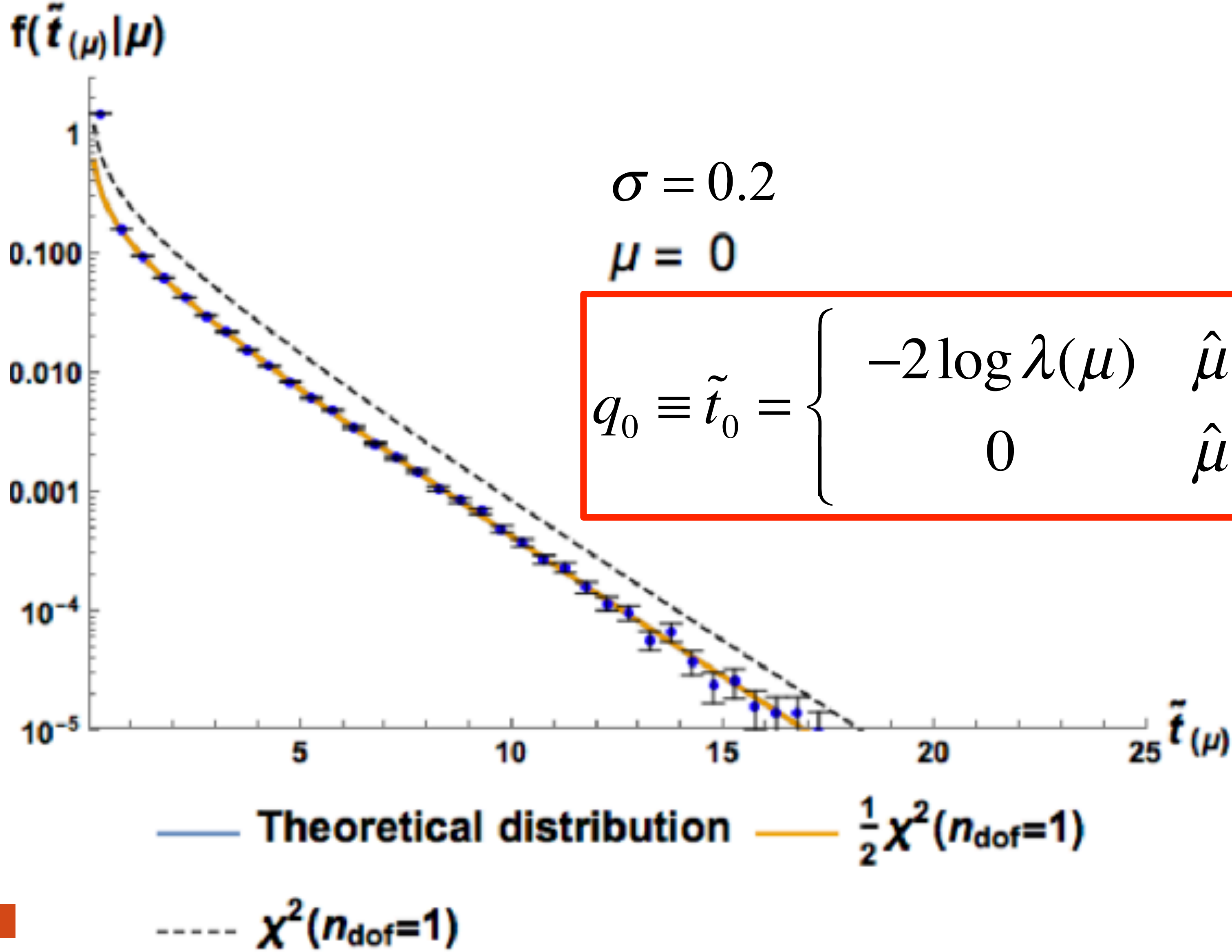
$$\Phi(Z) = \int_{-\infty}^{Z} G(x;0,1)dx$$

$$F(\tilde{t}_\mu|\mu) = \begin{cases} 2\Phi\left(\sqrt{\tilde{t}_\mu}\right) - 1 & \tilde{t}_\mu \le \frac{\mu^2}{\sigma^2} \\ \Phi\left(\sqrt{\tilde{t}_\mu}\right) + \Phi\left(\frac{\tilde{t}_\mu + \mu^2/\sigma^2}{2\mu/\sigma}\right) - 1 & \tilde{t}_\mu > \frac{\mu^2}{\sigma^2} \end{cases}$$

$$CI\ of\ \mu\ at\ the\ (1-\alpha)\ CL = \left\{\mu \middle| p_\mu \ge \alpha\right\}$$

$$CI\ of\ \mu\ at\ the\ 95\%\ CL = \left\{\mu \middle| p_\mu \ge 5\%\right\}$$

Validation of $f(\tilde{t}_\mu \mid \mu)$

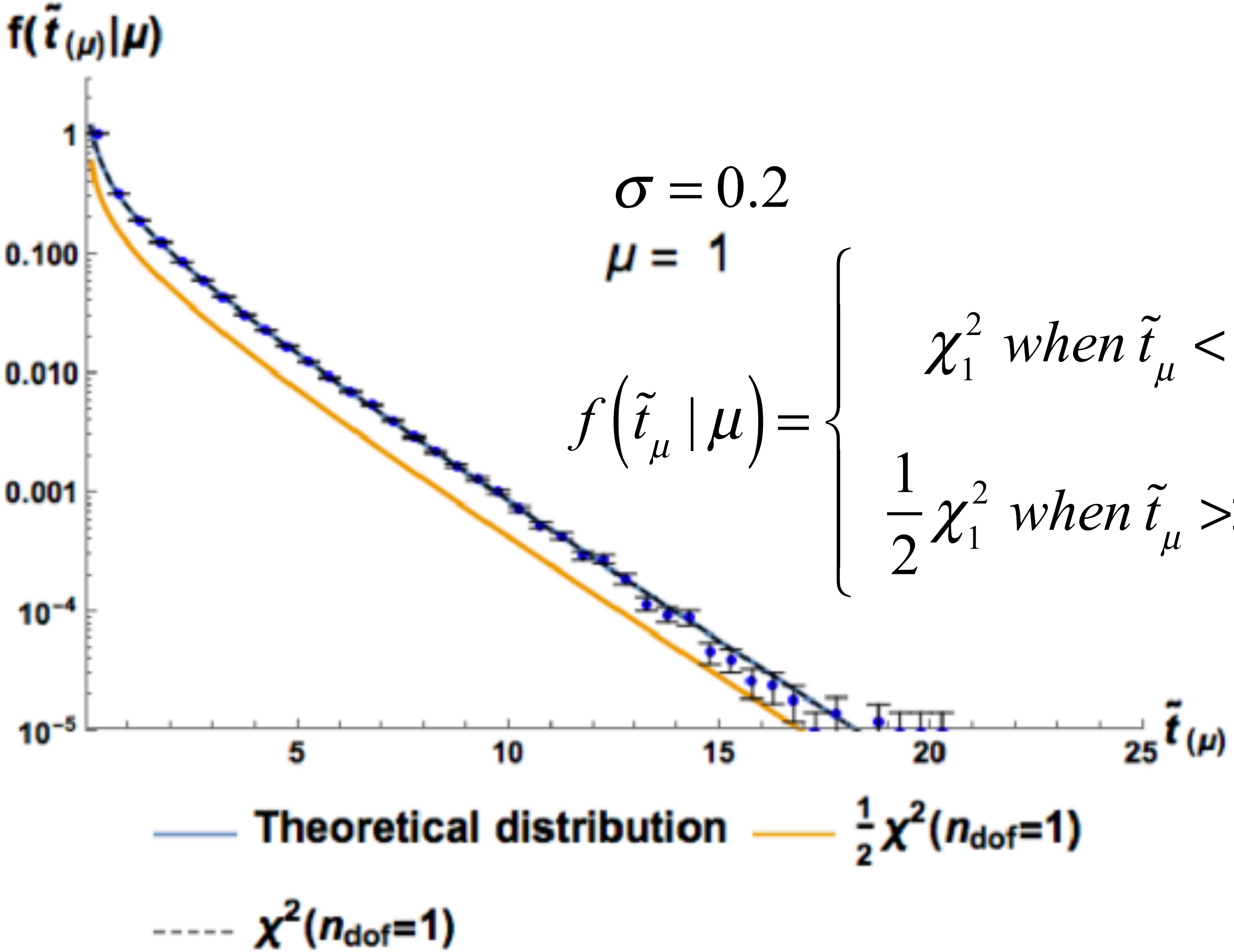


$\sigma = 0.2$

$\mu = 0$

$$q_0 \equiv \tilde{t}_0 = \begin{cases} -2\log\lambda(\mu) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}$$

Validation of $f(\tilde{t}_\mu \mid \mu)$

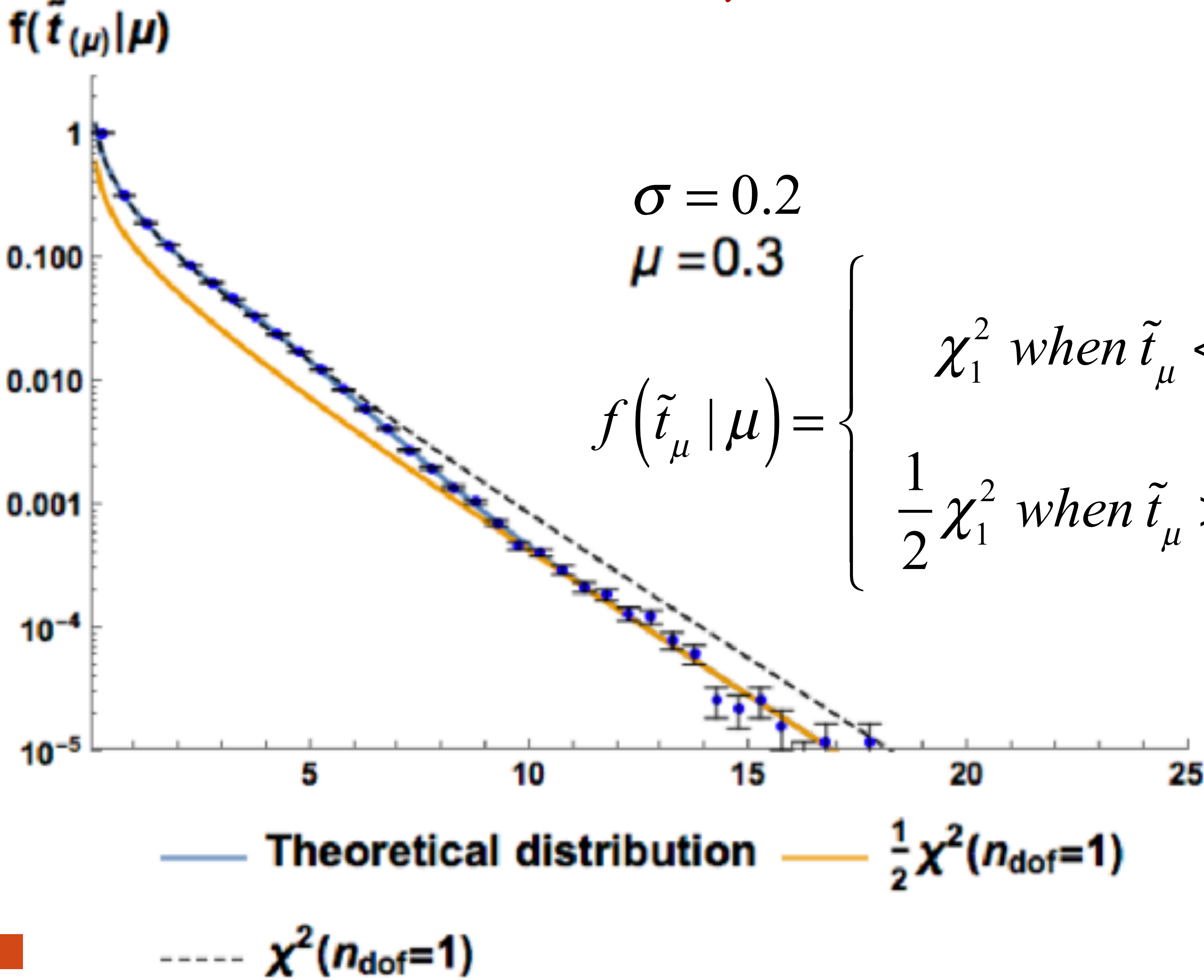


$$\sigma = 0.2$$

$$\mu = 1$$

$$f\left(\tilde{t}_\mu \mid \mu\right) = \begin{cases} \chi_1^2 \text{ when } \tilde{t}_\mu < \dfrac{\mu^2}{\sigma^2} = 25 \\ \dfrac{1}{2}\chi_1^2 \text{ when } \tilde{t}_\mu >> \dfrac{\mu^2}{\sigma^2} = 25 \end{cases}$$

Validation of $f(\tilde{t}_\mu \mid \mu)$



$\sigma = 0.2$

$\mu = 0.3$

$$f\left(\tilde{t}_\mu \mid \mu\right) = \begin{cases} \chi_1^2 \; when \; \tilde{t}_\mu < \dfrac{\mu^2}{\sigma^2} = 2.25 \\ \dfrac{1}{2}\chi_1^2 \; when \; \tilde{t}_\mu >> \dfrac{\mu^2}{\sigma^2} = 2.25 \end{cases}$$

FC confidence belt

Given $\mu_{true} = 0.2$

derive the $\hat{\mu}$ *interval*

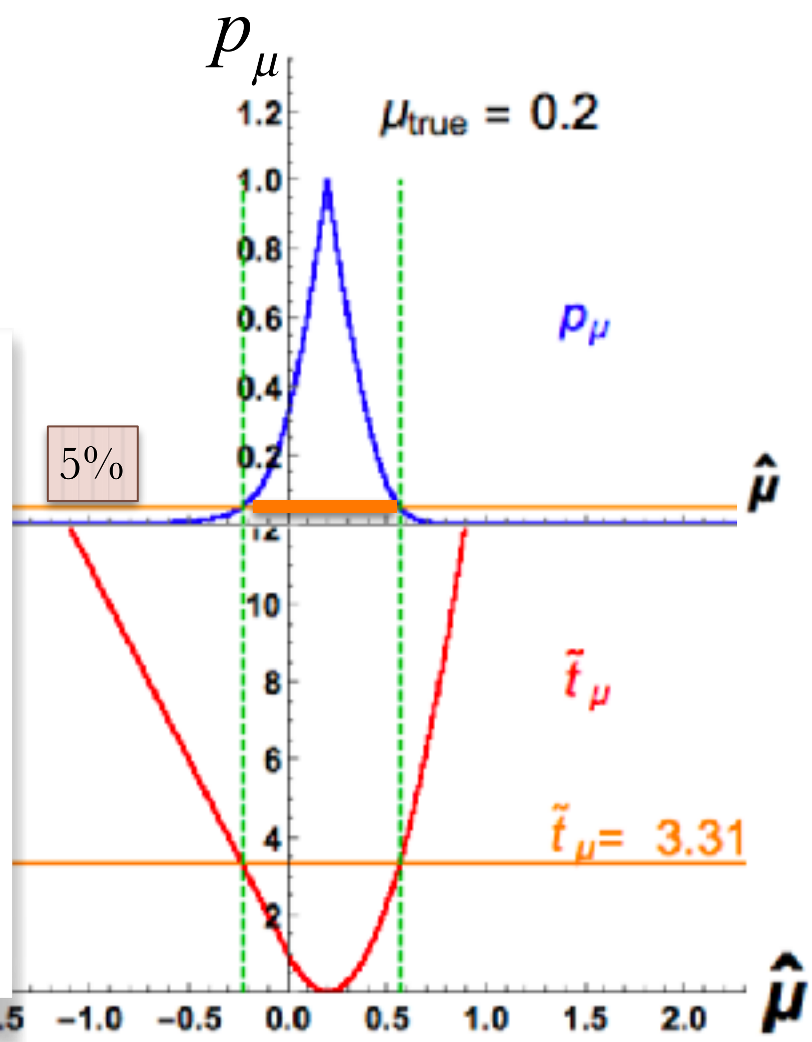
for which $p_{\mu} > 0.05$

set μ_{true} e.g. $\mu_{true} = 0.2$

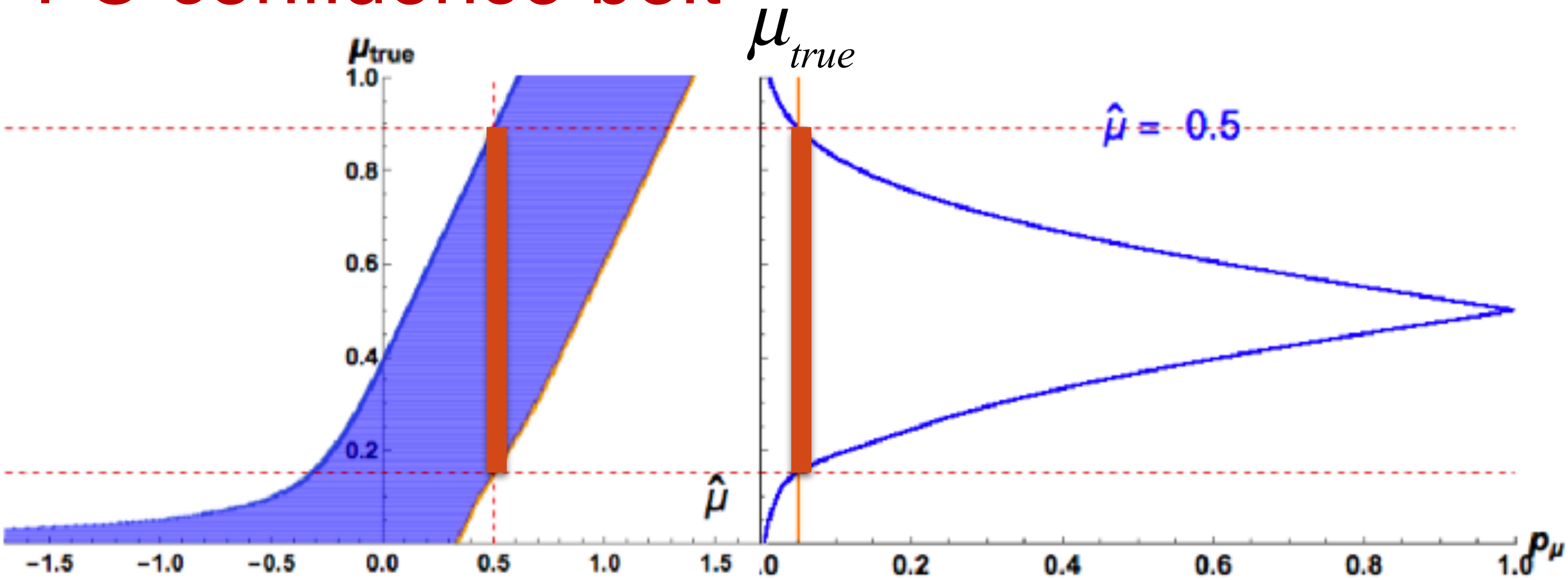
$$p_{\mu}(\hat{\mu}) = 1 - F\left(\tilde{t}_{\mu}(\hat{\mu}) \mid \mu\right) \Rightarrow$$

$$CI_{\hat{\mu}} = \left\{\hat{\mu} \middle| p_{\mu}(\hat{\mu}) \geq 5\%\right\} \Rightarrow$$

$$\tilde{t}_{\mu} = 3.31$$



FC confidence belt



The values of μ_{true} *for which* $p_\mu \geq 0.05$

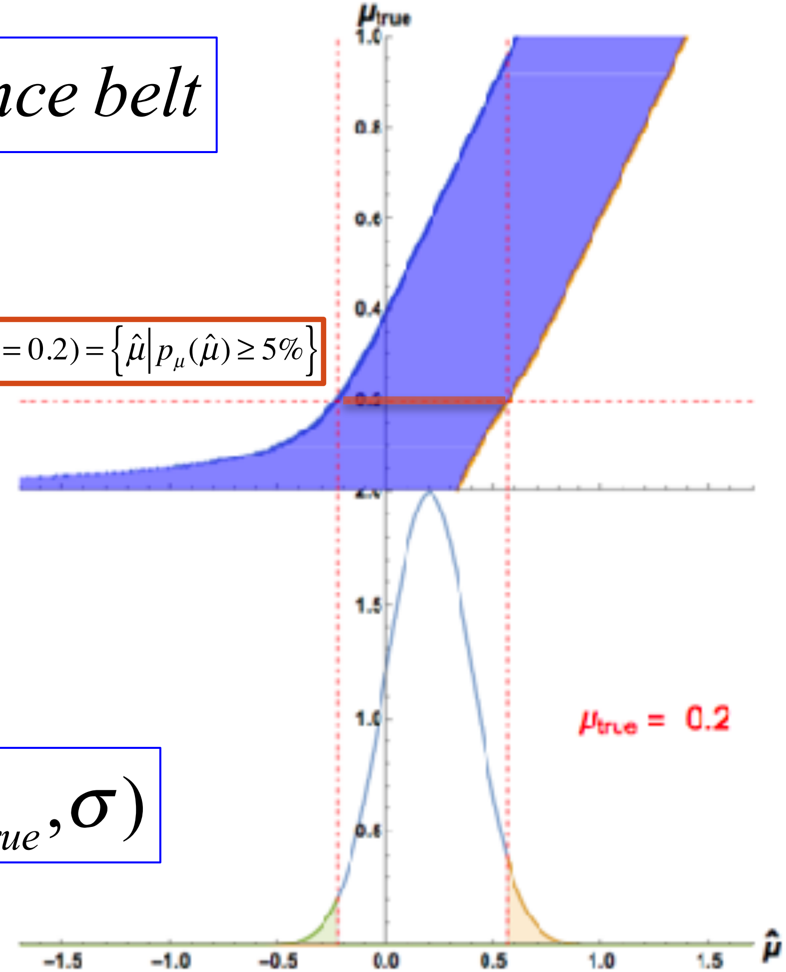
for a given $\hat{\mu}$, *are the CI of* μ

FC confidence belt

$\mu_{true}(\hat{\mu})$

Scan μ_{true} *and build the Confidence belt*

$CI_{\hat{\mu}}(\mu_{true}=0.2)=\left\{\hat{\mu}\,\middle|\,p_{\mu}(\hat{\mu})\geq 5\%\right\}$



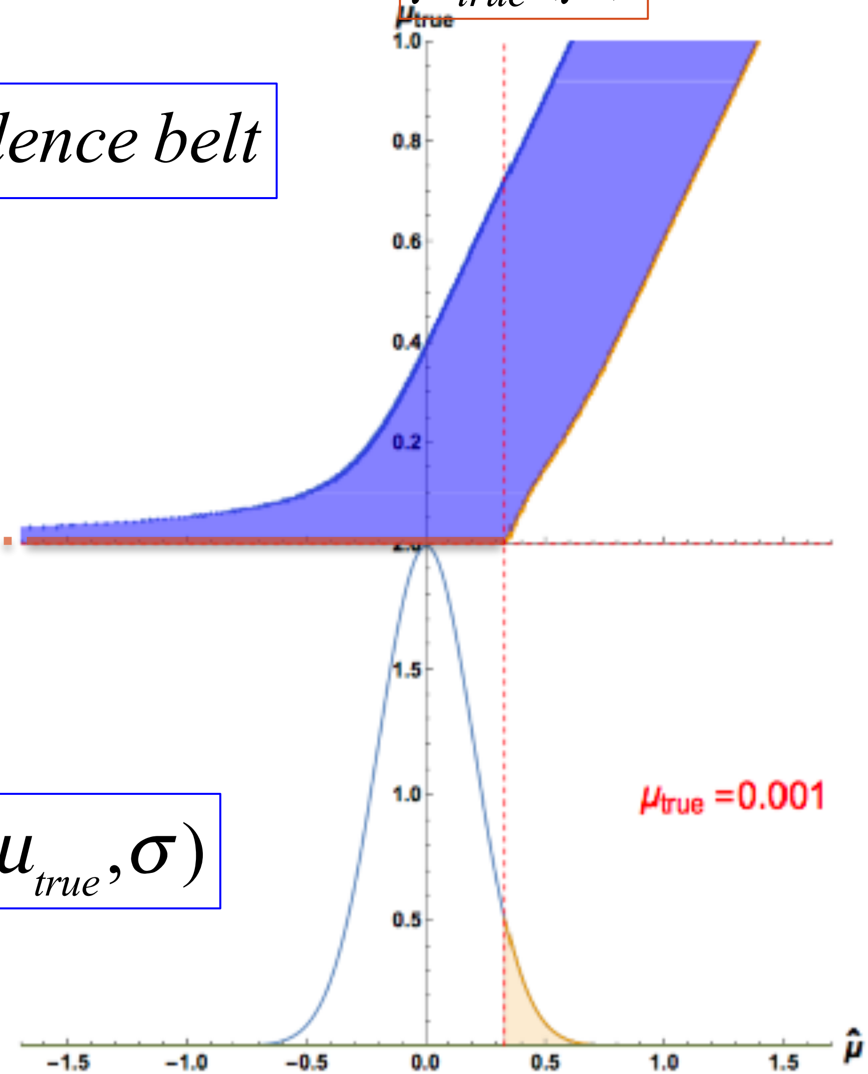
$G(\hat{\mu};\mu_{true},\sigma)$

FC confidence belt

Scan μ_{true} and build the Confidence belt

$\mu_{true}(\hat{\mu})$

$CI_{\hat{\mu}}(\mu_{true} = 0.001) = \left\{ \hat{\mu} \middle| p_{\mu}(\hat{\mu}) \geq 5\% \right\}$



$G(\hat{\mu}; \mu_{true}, \sigma)$

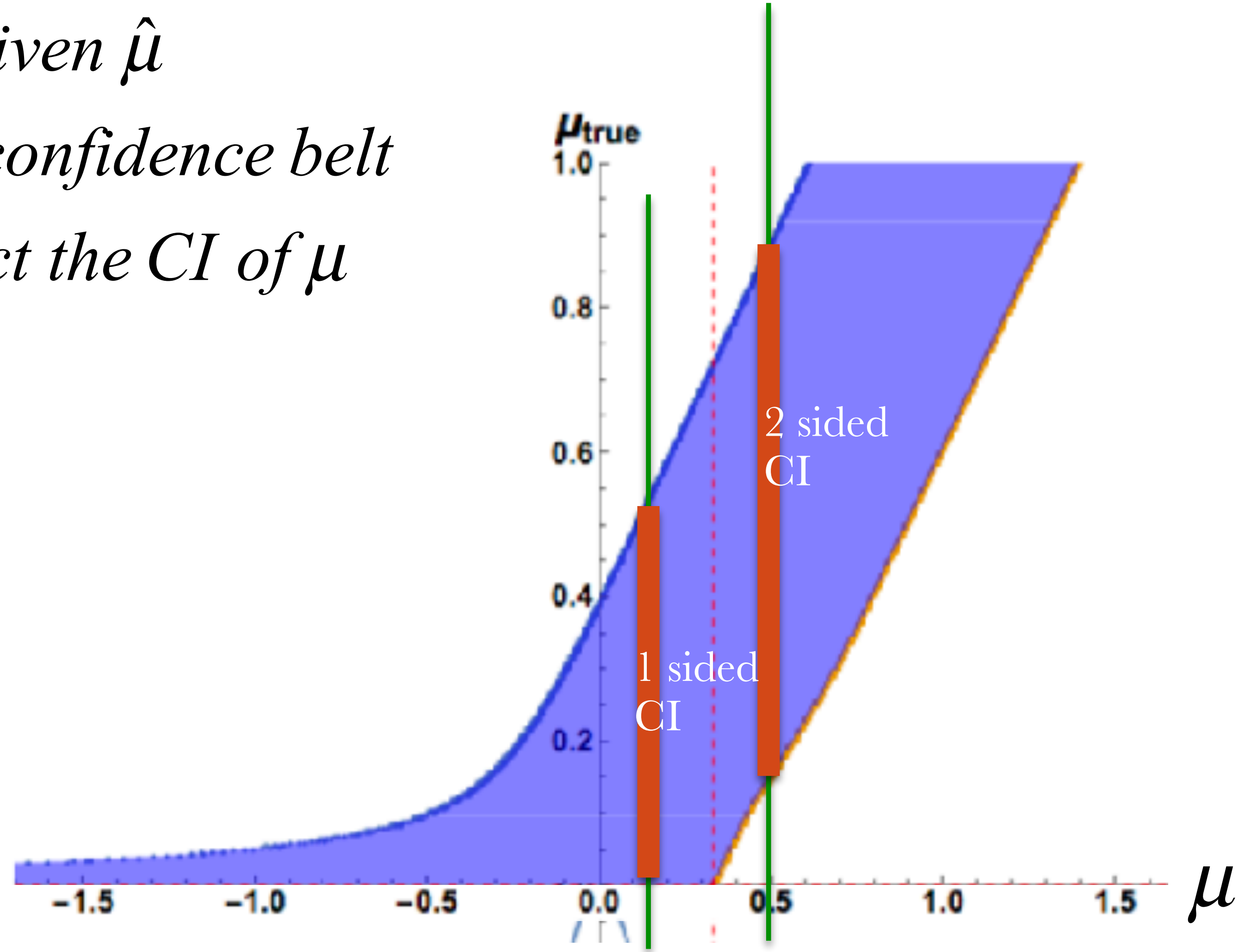
FC confidence belt

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

For a given $\hat{\mu}$

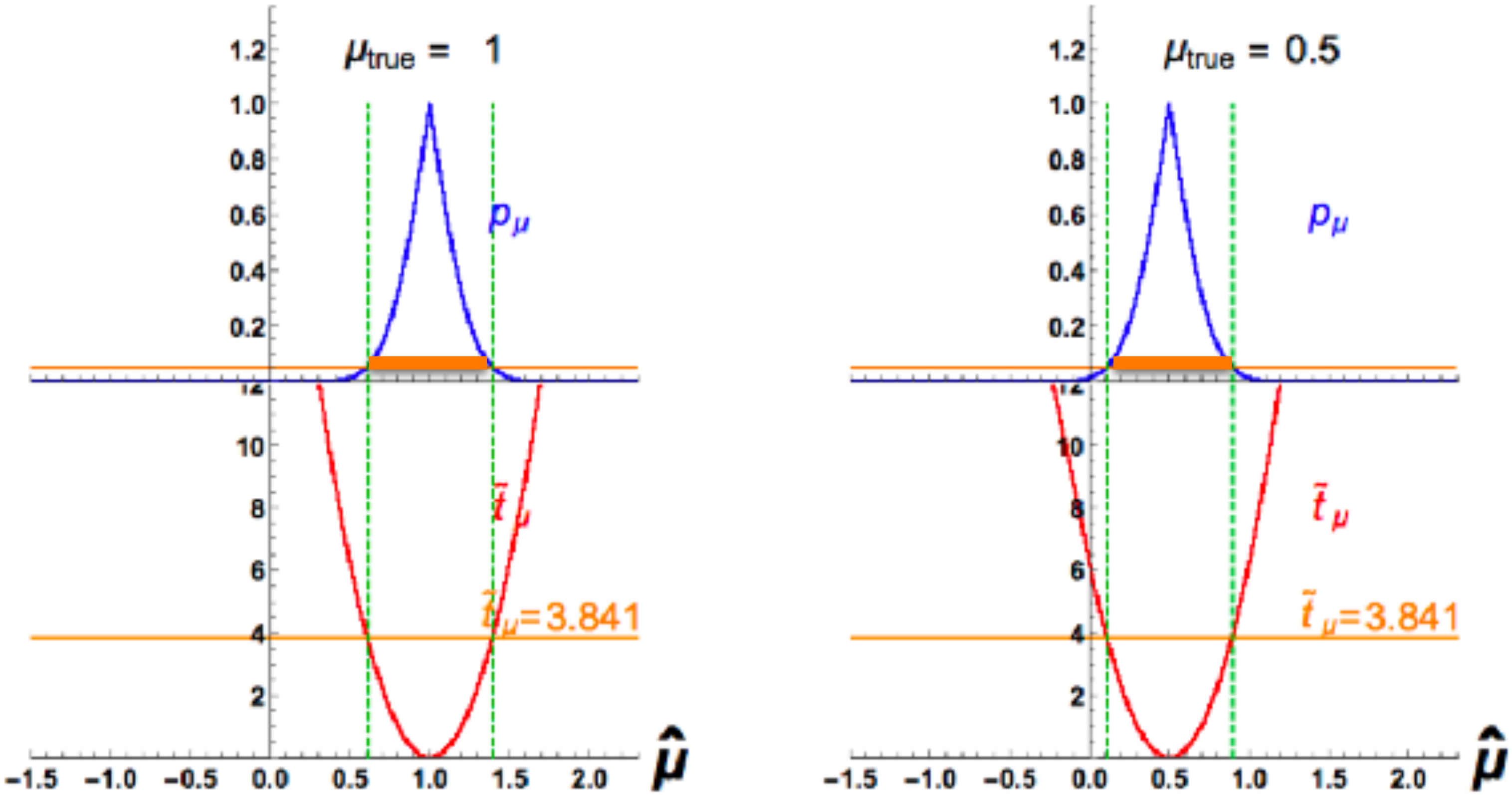
use the confidence belt

to extract the CI of μ



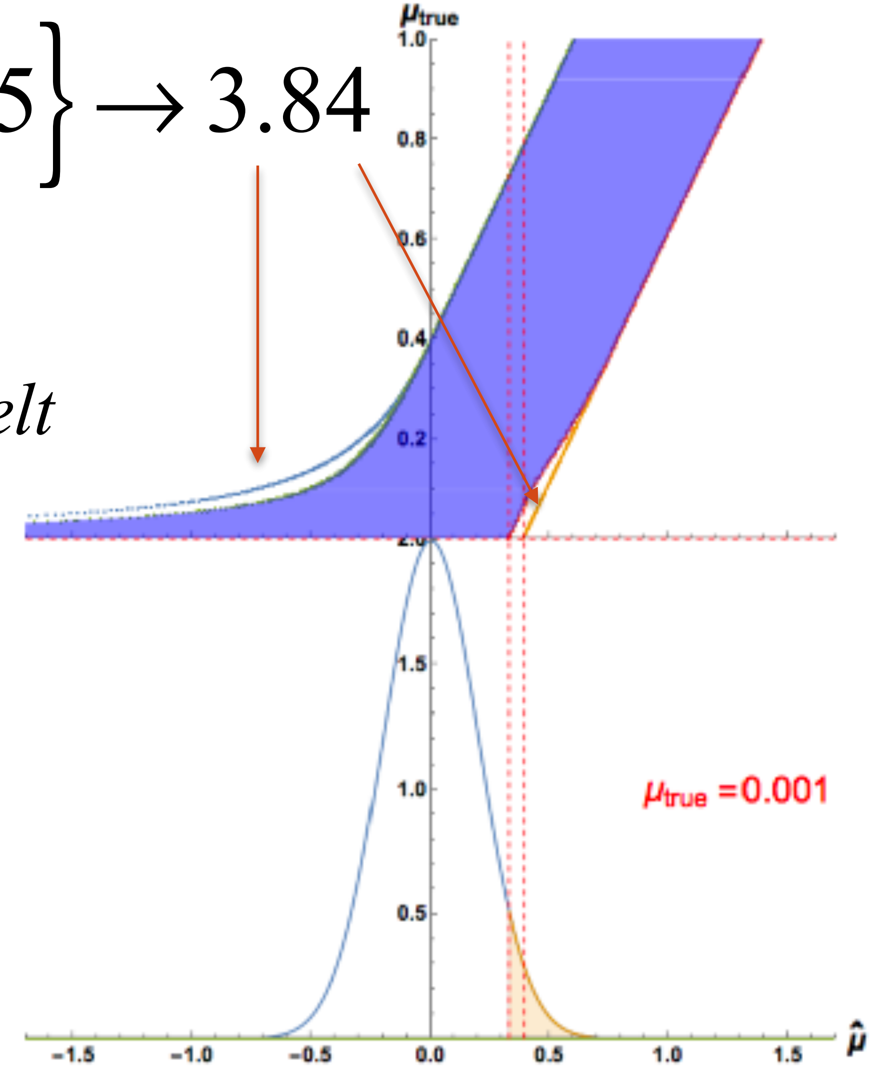
FC confidence belt - A Shortcut

$$\mu_{true} >> 0 \Rightarrow \left\{ \tilde{t}_\mu \middle| p_\mu > 0.95 \right\} \rightarrow 3.84$$



FC confidence belt - A Shortcut

$$\mu_{true} >> 0 \Rightarrow \left\{ \tilde{t}_\mu \middle| p_\mu > 0.95 \right\} \rightarrow 3.84$$



Use $\tilde{t}_\mu = 3.84$ *to construct the belt*

The CI will be at CL $> 95\%$

(Conservative)

$$q_0 \equiv \tilde{t}_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases} \qquad \lambda(0) = \frac{L(\mu=0)}{L(\hat{\mu})} = \frac{L(\hat{\hat{b}}_{\mu=0})}{L(\hat{\mu}s+\hat{b})} = \frac{L(\hat{\hat{b}}_{\mu=0})}{L(\hat{s}+\hat{b})}$$

q_0 for discovery

CCGV

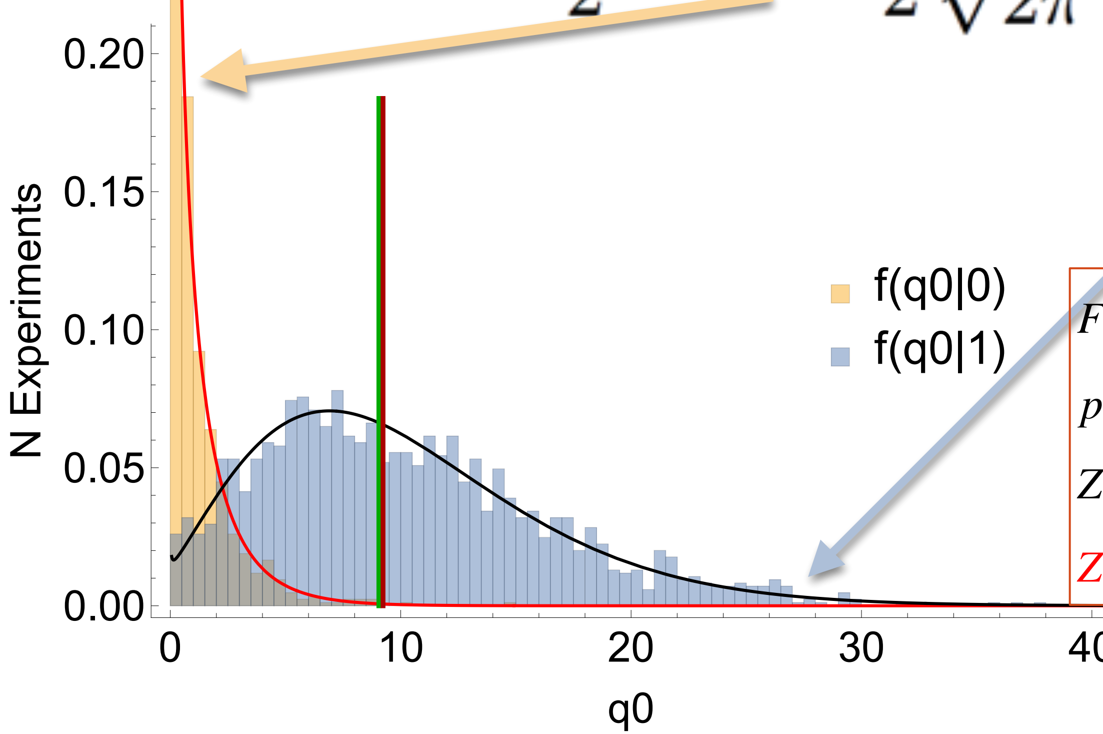
$$q_0 = \begin{cases} -2\log\lambda(\mu) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}$$

Downward fluctuations of the background
do not serve as an evidence against the background

PDF of (q0|0) and (q0|1)

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

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



$$F(q_0|0) = \int_{-\infty}^{q_0} f(q_0|0)dq_0 \equiv \Phi\left(\sqrt{q_0}\right)$$

$$p_0 = 1 - F(q_0|0)$$

$$Z_0 = \Phi^{-1}(1 - p_0) = \Phi^{-1}\left(F(q_0|0)\right) = \sqrt{q_0}$$

$$Z_0 = \sqrt{q_0}$$

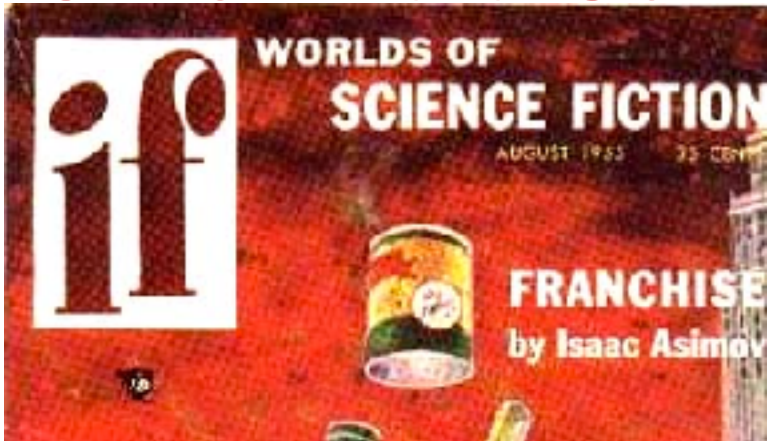
Estimating the Sensitivity of an Experiment

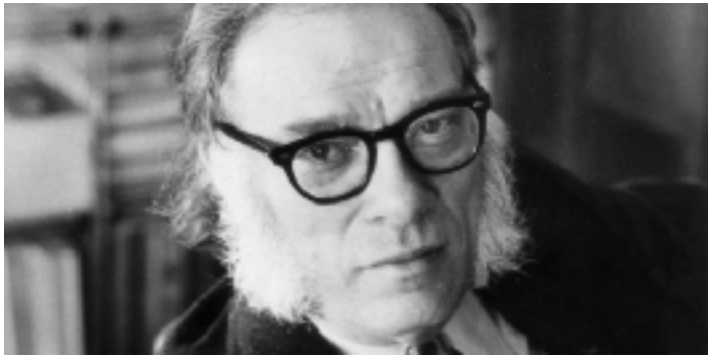
- Estimate the expected significance one could achieve (for discovering the Higgs Boson) with a given analysis, a given Luminosity and CM energy..

- Option 1:
  - Toss, say, 1,000,000 BG only events (null) and derive the BG-only pdf of q, $f(q_{null}|BG)$.
    Toss another 1,000,000 S+BG (alt) events and find the significance for each one of them
    then, find the median significance….
  - This may take ages…, is there a shortcut?

- Option 2:
  - Asymptotics+Asimov Data Set

The Asimov Data Set





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

The Asimov Data Set

- *The use of a single representative individual to stand in for the entire population can help in evaluating the sensitivity of a statistical method.*

- *The "**Asimov data set**":*
an ensemble of simulated experiments can be replaced by a single representative one.

Estimating the Sensitivity of an Experiment

- one can replace each ensemble of the alternate-hypothesis experiments with one data set that represents the typical experiment.

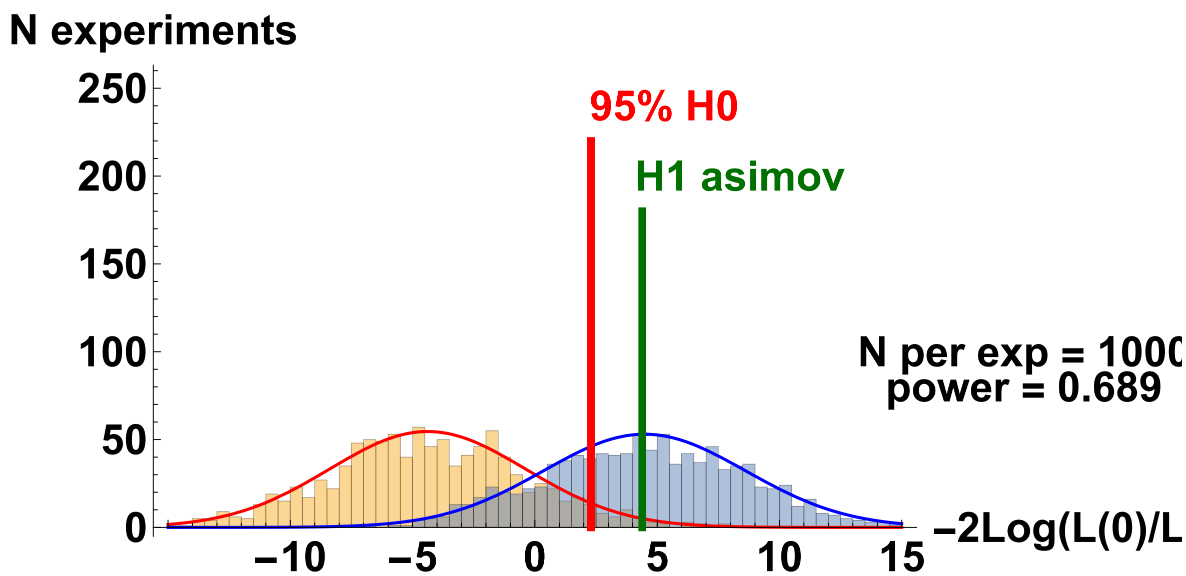
This "Asimov" data set delivers the desired median sensitivity. Hence, one is exempted from the need to perform an ensemble of experiments for each set of parameters.

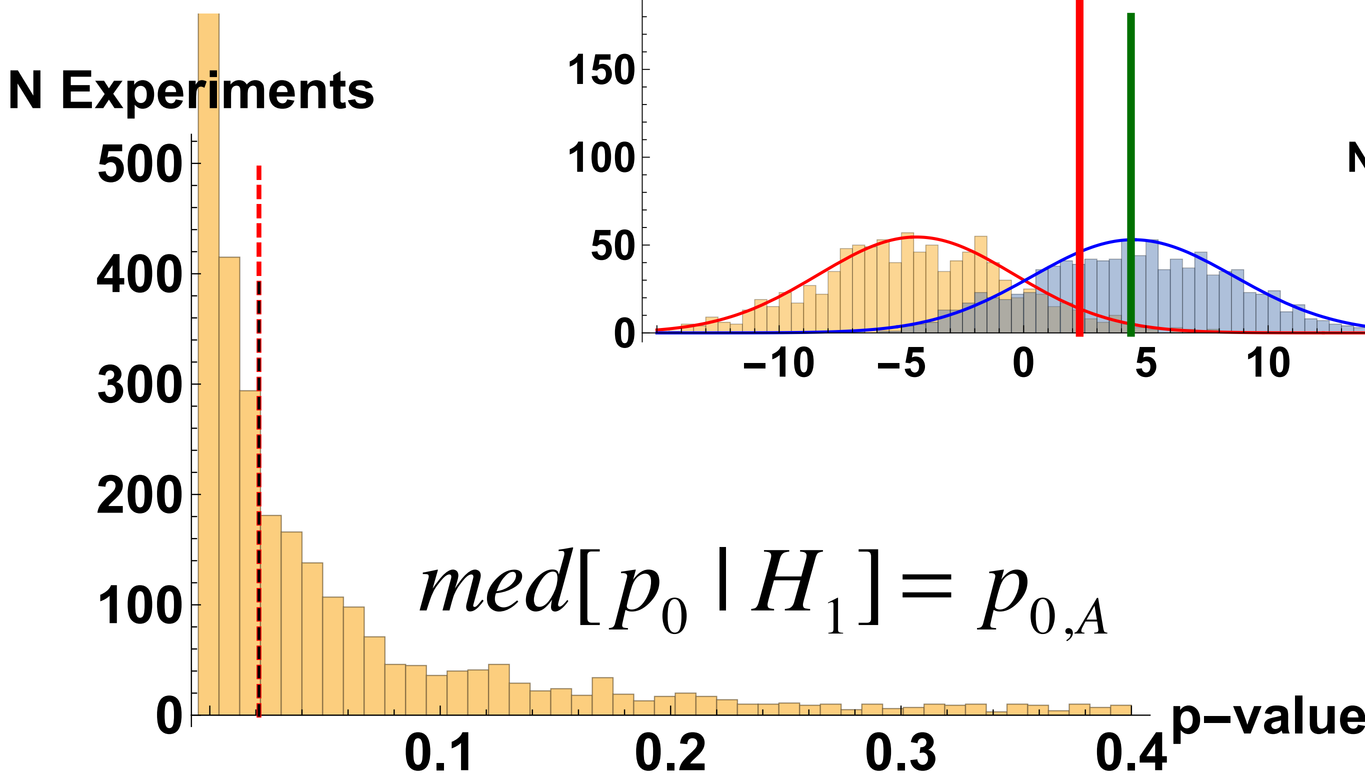
- The Asimov data set is constructed such that when one uses it to evaluate the estimators for all parameters, one obtains the true parameter values.

- the Asimov data set can trivially be constructed from the true parameters values. For example, a set corresponding to the H_1 hypothesis is $n_A = s + b$. and the one correspond to the H_0 hypothesis is $n_A = b$.

- As strange as it reads, the Asimov data set is **not** necessarily an **integer**.

Back to Spin
Distribution of p_0-value under H1

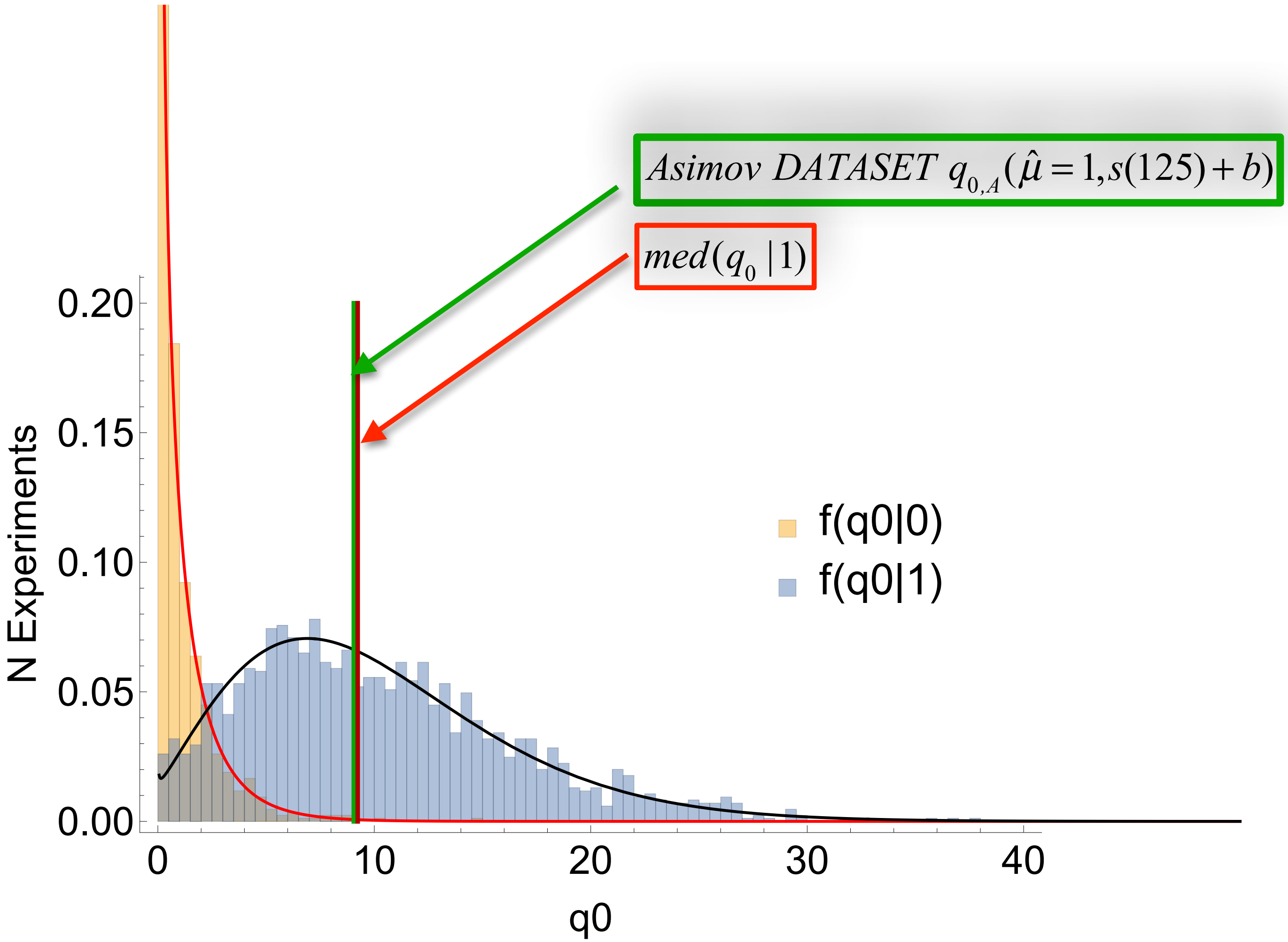




$$med[p_0 \mid H_1] = p_{0,A}$$

$$f(p_0 \mid H_1) = f\left(\left[prob(q \geq q_{obs}) \mid H_0\right] \middle\| H_1\right)$$

The Magic of Asimov



Back to Wald, **what is** $\sigma_{\hat{\mu}}$?

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

For the Asimov $\hat{\mu} = \mu_{true}$

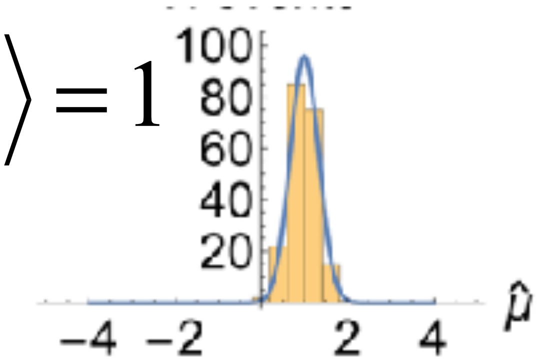
$$q_{\mu,A} = -2\ln\lambda_A(\mu) = \frac{(\mu-\mu_{true})^2}{\sigma_{\hat{\mu}}^2}$$

$$\sigma_{\hat{\mu}}^2 = \frac{(\mu-\mu_{true})^2}{q_{\mu,A}} \qquad \langle\hat{\mu}\rangle = \mu_{true}$$

$$set\ \mu = 0\ \ \mu_{true} = 1 \rightarrow \sigma_{\hat{\mu}}^2 = \frac{1}{q_{0,A}}$$

$$set\ \mu = 1\ \ \mu_{true} = 0 \rightarrow \sigma_{\hat{\mu}}^2 = \frac{1}{q_{1,A}}$$

$$\langle\hat{\mu}\rangle = 1$$



Test μ for exclusion, $\mu_{true} = 0$

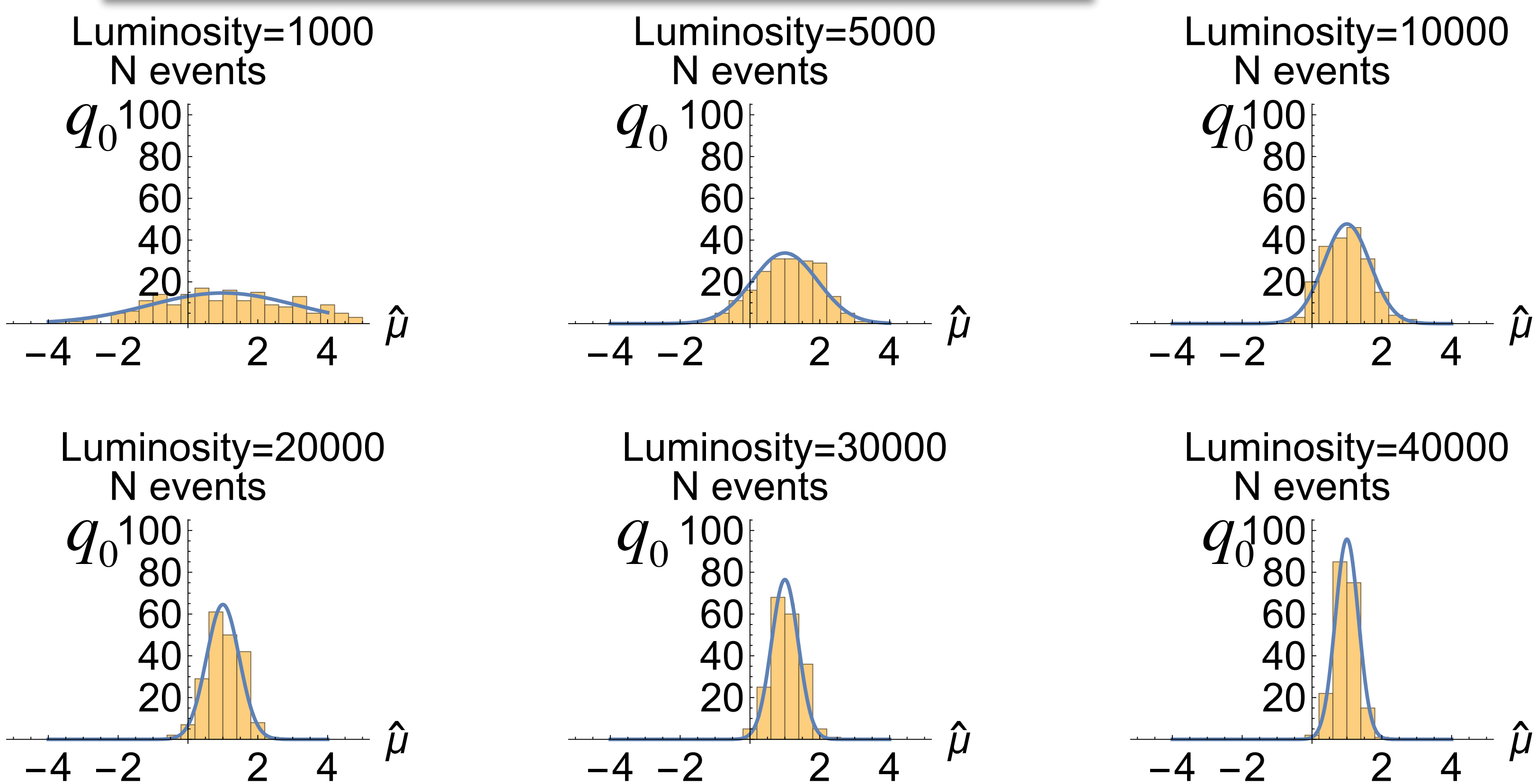
$$\rightarrow \sigma_{\hat{\mu}}^2 = \frac{\mu^2}{q_{\mu,A}}$$

Back to Wald, **what is** $\sigma_{\hat{\mu}}$?

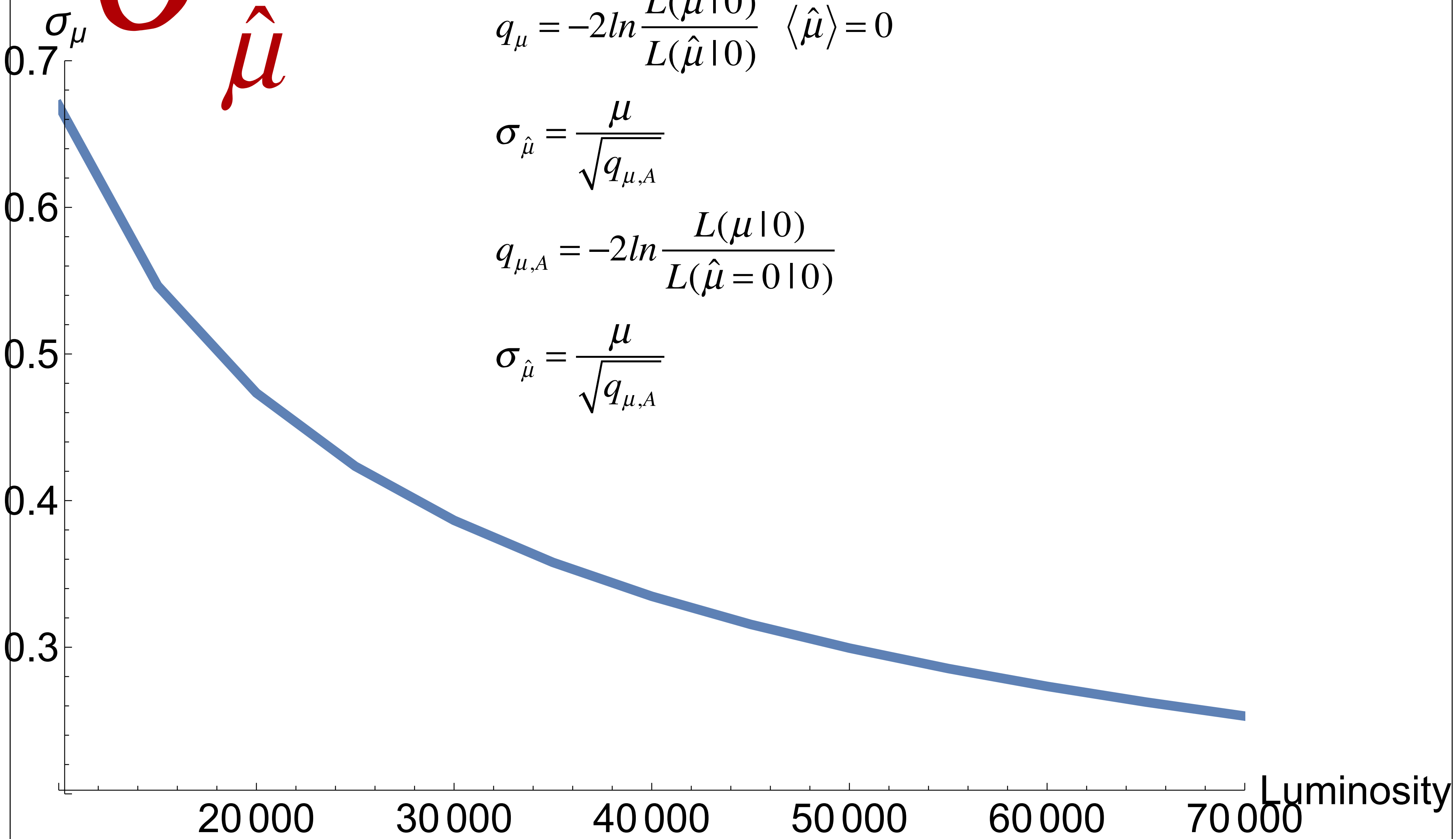
$$-2\ln\lambda(0) = \frac{\hat{\mu}^2}{\sigma_{\hat{\mu}}^2} + O\left(\frac{1}{\sqrt{N}}\right) \quad \langle\hat{\mu}\rangle = 1$$

$$\sigma_{\hat{\mu}}^2 = \frac{1}{q_{0,A}}$$

Fit the distribution with the Asimov calculated Sigma



$\sigma_{\hat{\mu}}$



$$q_\mu = -2\ln\frac{L(\mu|0)}{L(\hat{\mu}|0)} \quad \langle\hat{\mu}\rangle = 0$$

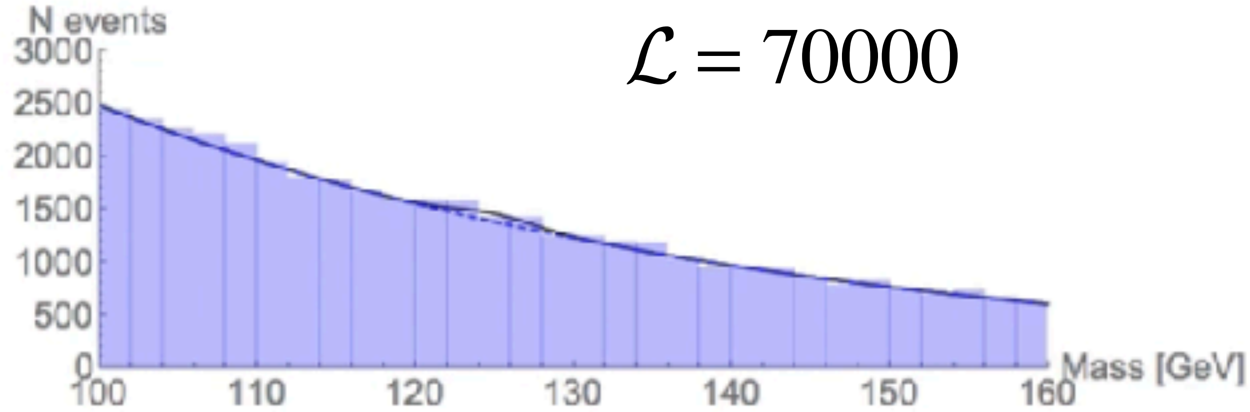
$$\sigma_{\hat{\mu}} = \frac{\mu}{\sqrt{q_{\mu,A}}}$$

$$q_{\mu,A} = -2\ln\frac{L(\mu|0)}{L(\hat{\mu}=0|0)}$$

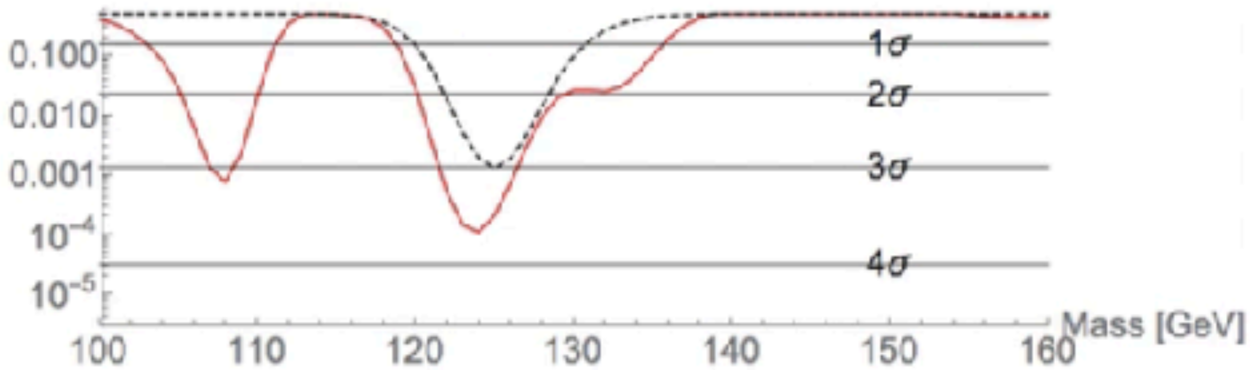
$$\sigma_{\hat{\mu}} = \frac{\mu}{\sqrt{q_{\mu,A}}}$$

p- value

$$\mathcal{L} = 70000$$

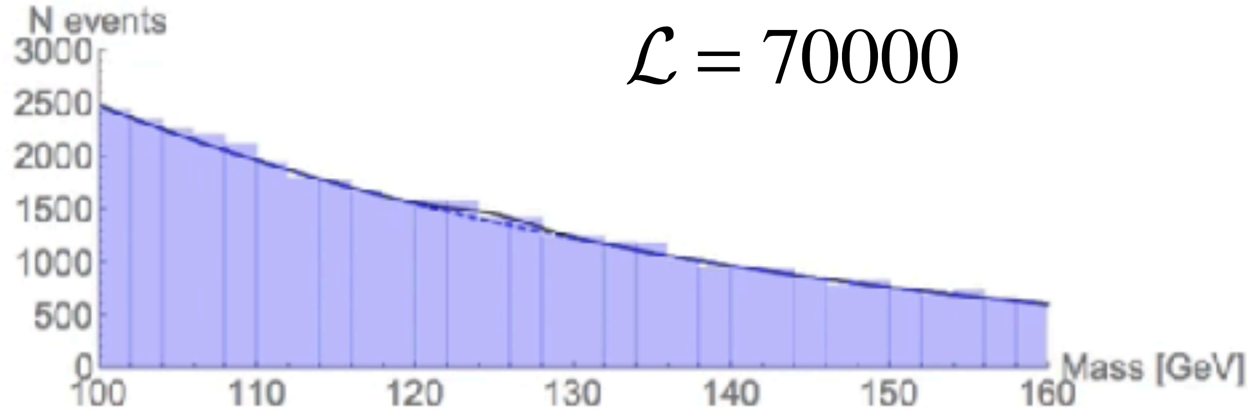


$$p = prob(q_0 \geq q_{0,obs})$$

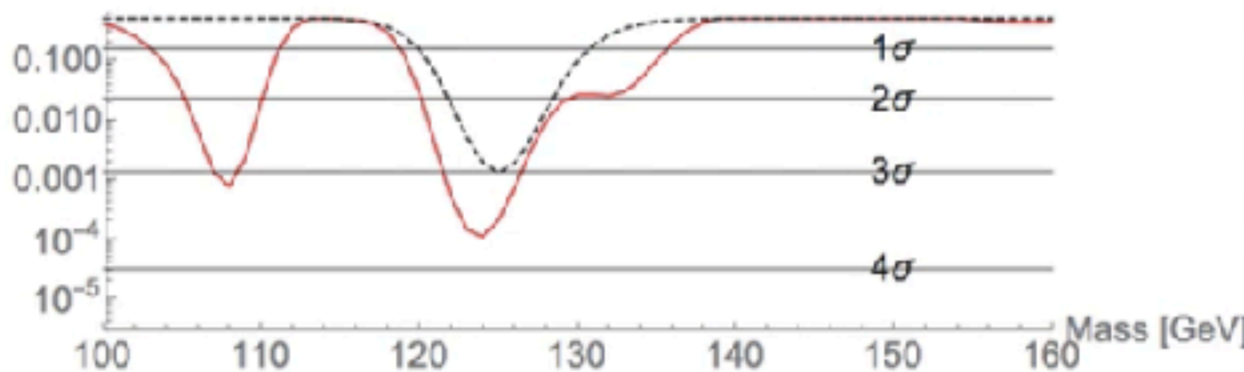


p- value

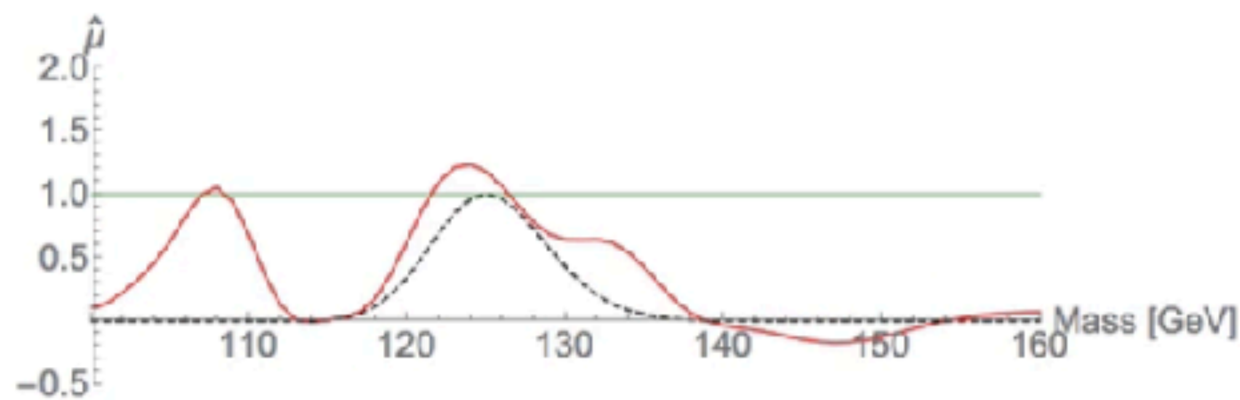
$\mathcal{L} = 70000$



$$p = prob(q_0 \geq q_{0,obs})$$

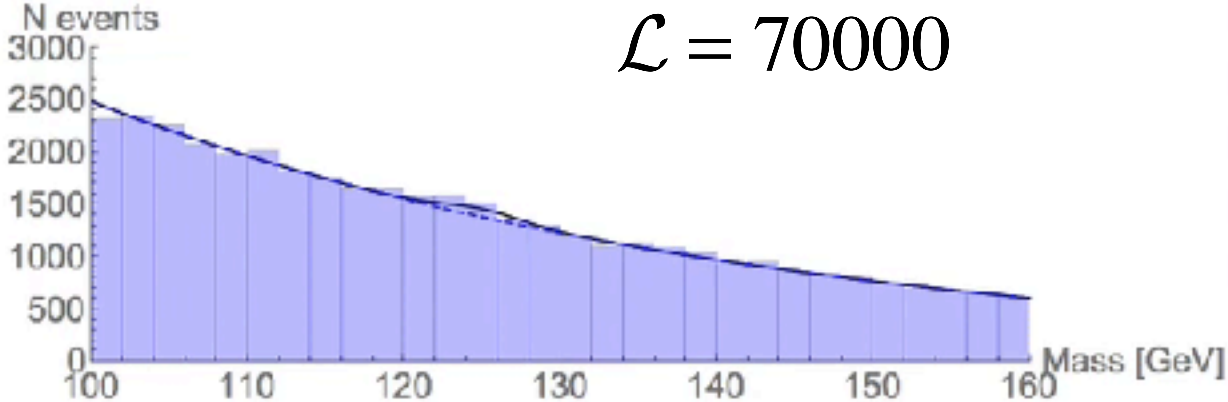


$$\hat{\mu} = \frac{\sigma_{obs}}{\sigma_{exp}}$$

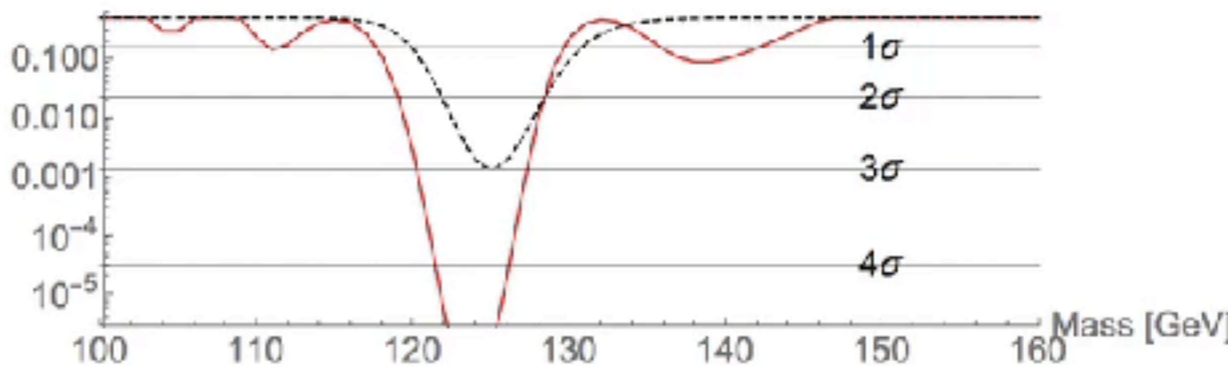


p- value

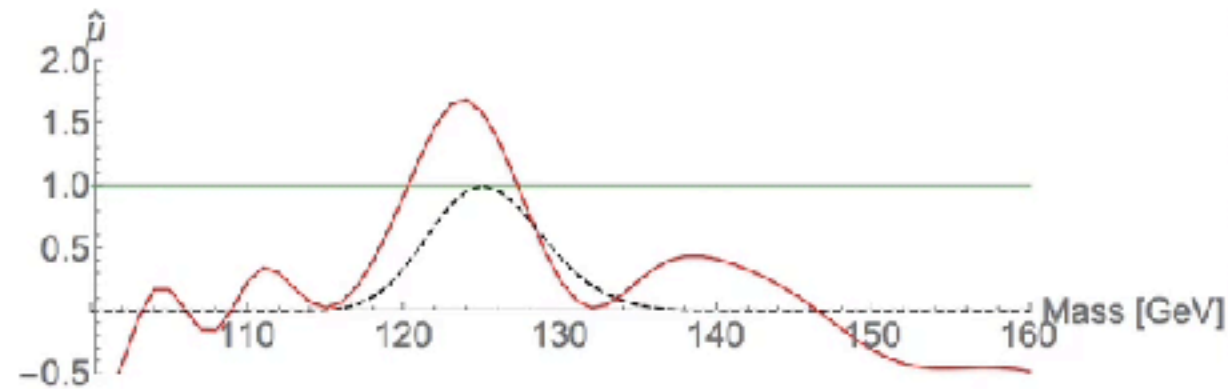
$$\mathcal{L} = 70000$$



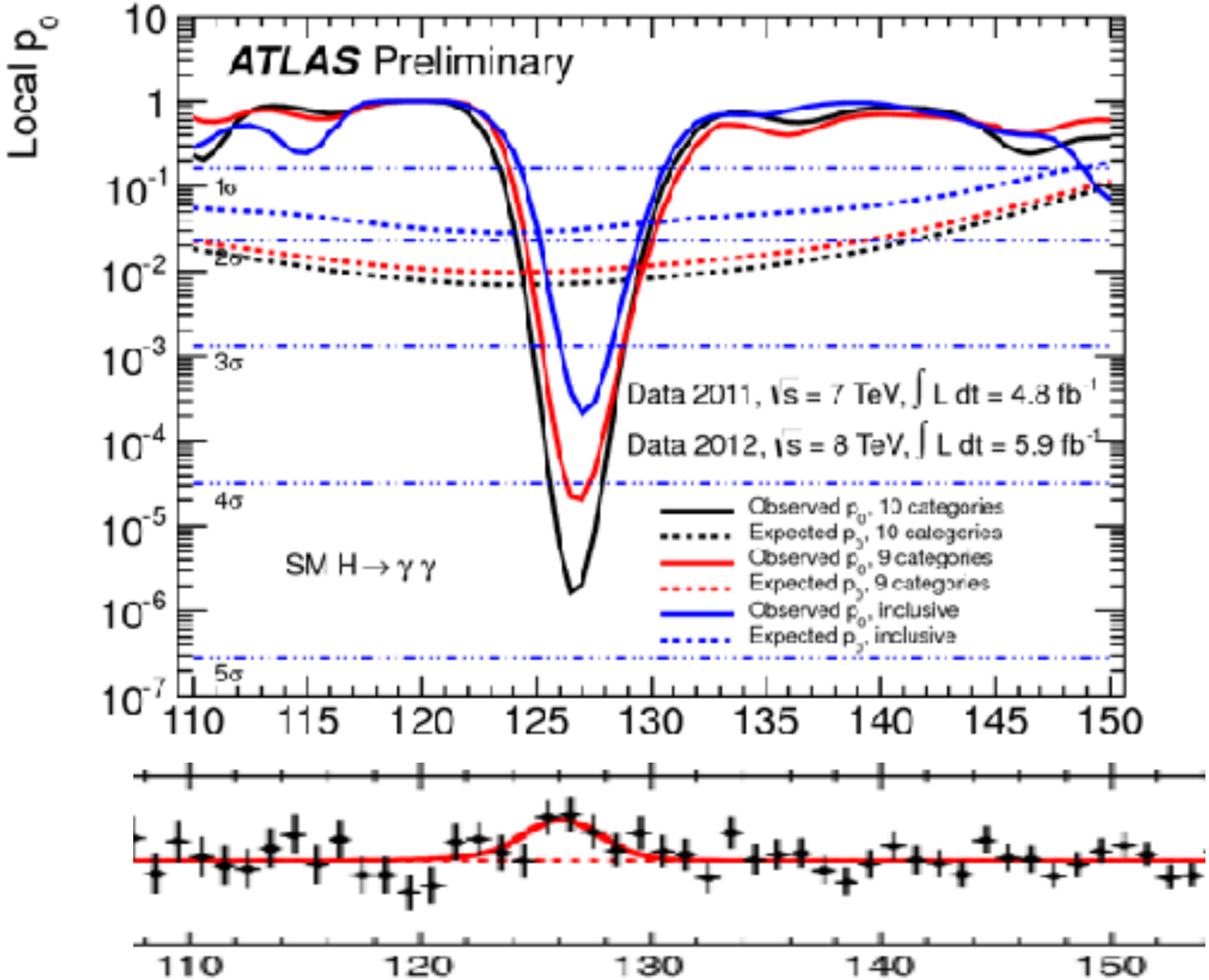
$$p = prob(q_0 \geq q_{0,obs})$$



$$\hat{\mu} = \frac{\sigma_{obs}}{\sigma_{exp}}$$



H→γγ


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

q_μ for exclusion

CCGV

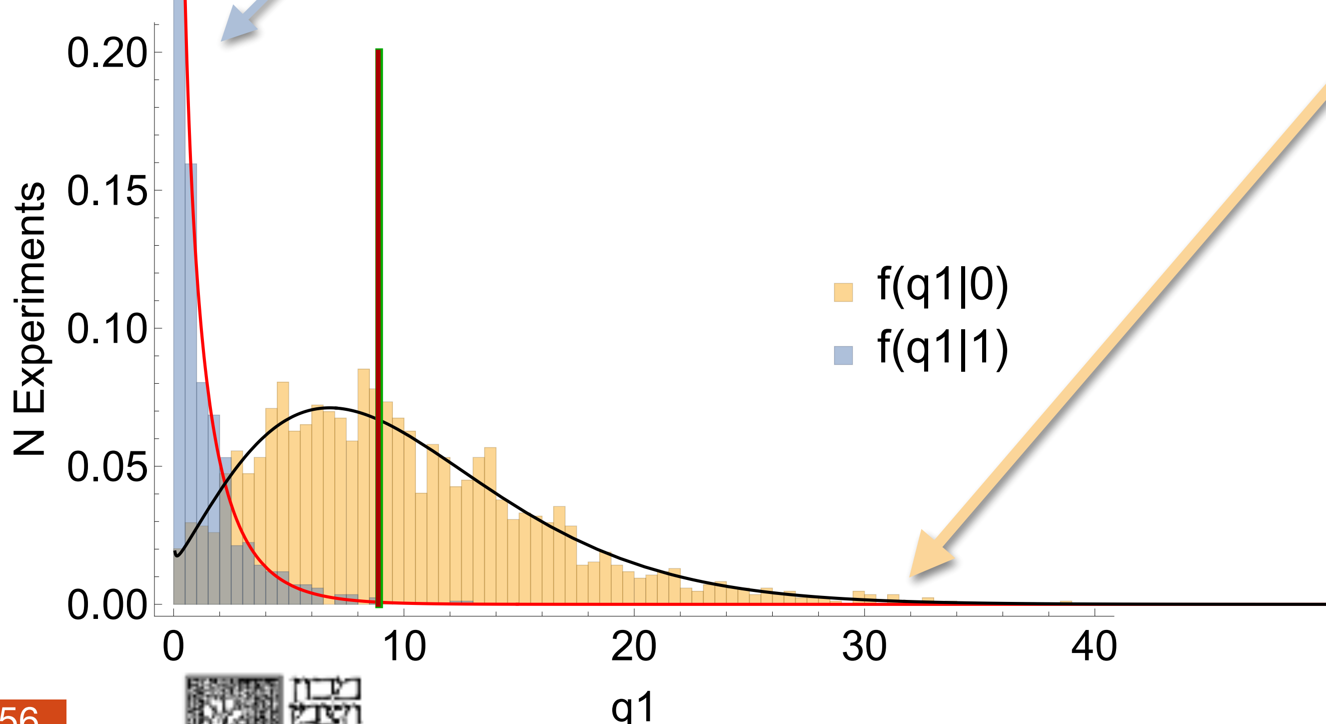
$$q_\mu = \begin{cases} -2\log\lambda(\mu) & \hat{\mu} \le \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

Upward fluctuations of the signal
do not serve as an evidence against the signal

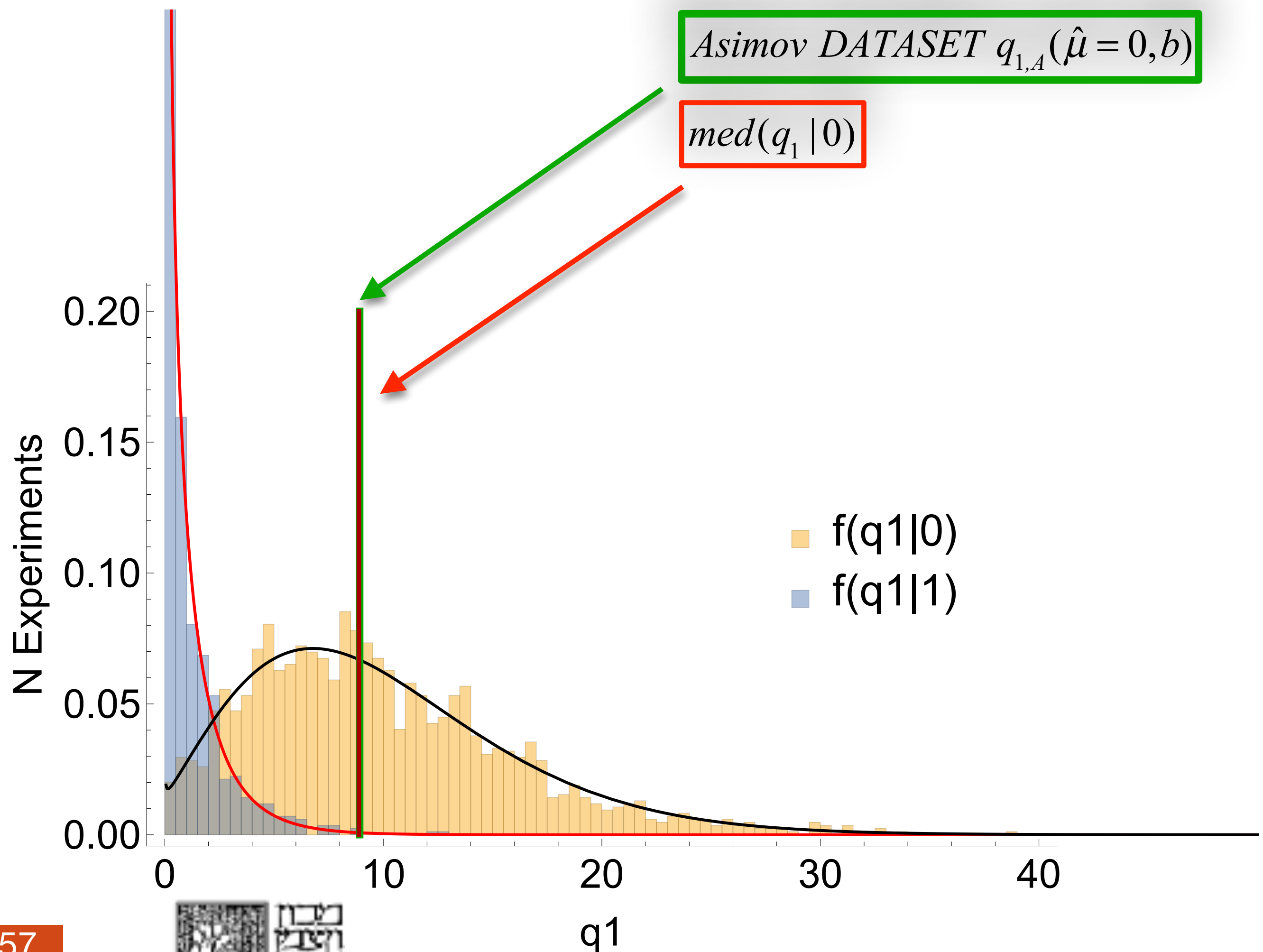
PDF of (q1|1) and (q1|1)

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

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

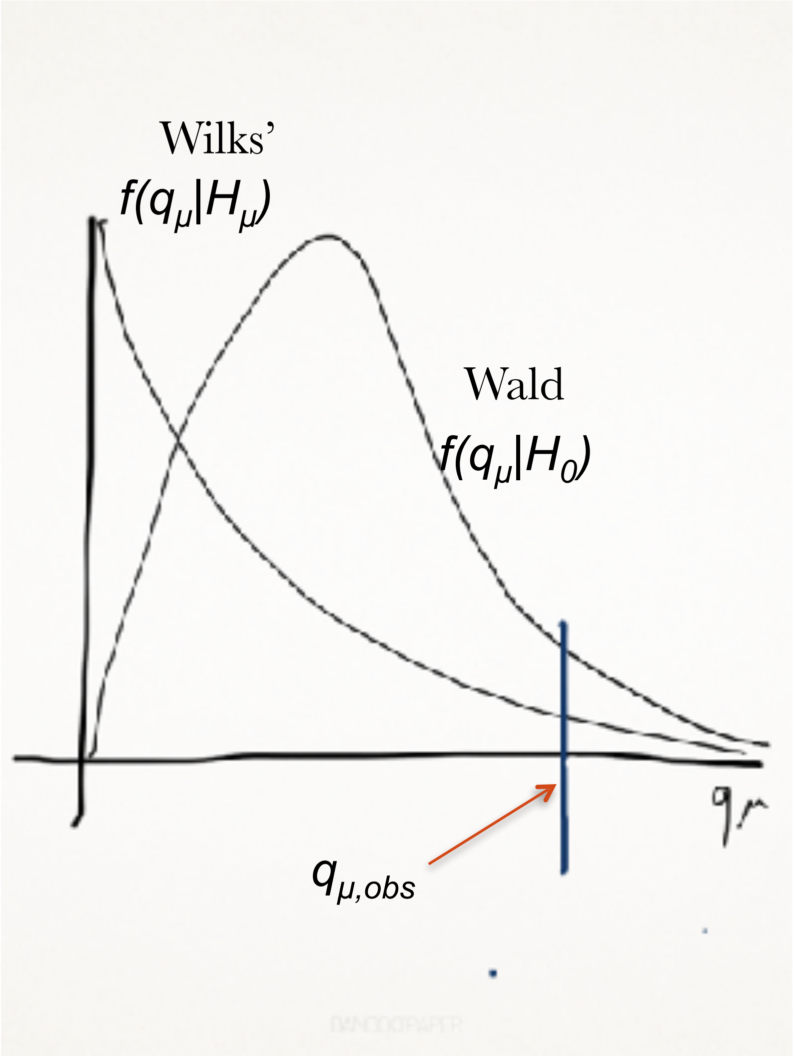


PDF of (q1|1) and (q1|1)



Exclusion at 95% CL

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

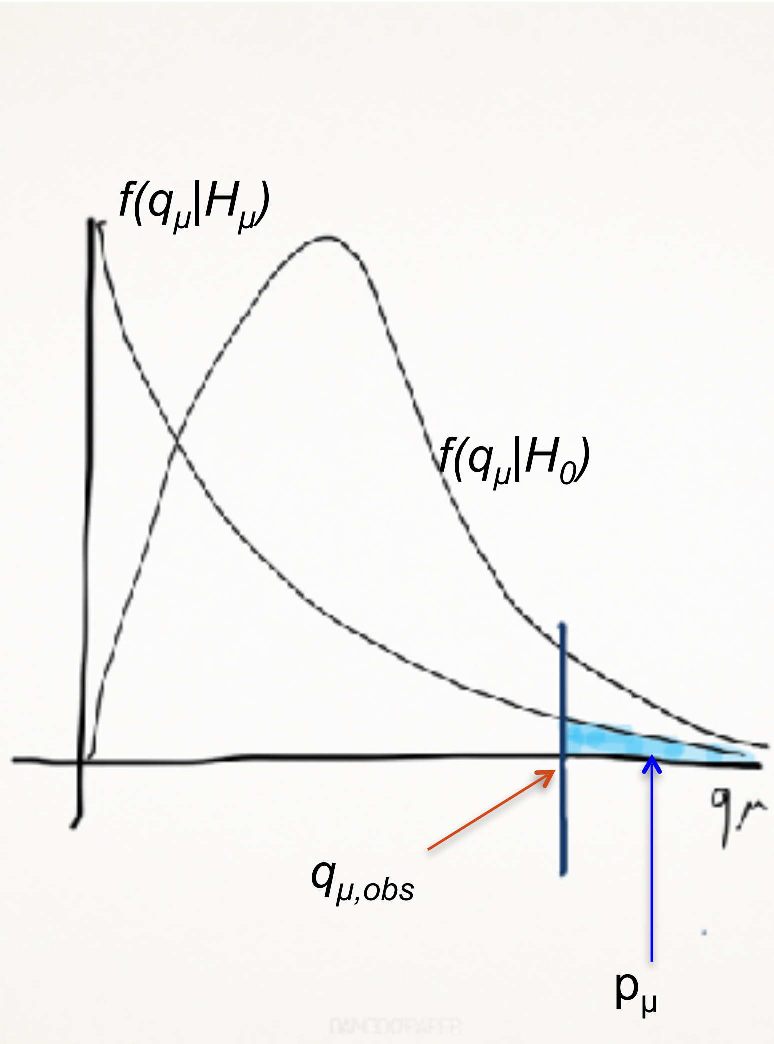


Exclusion at the 95% CL

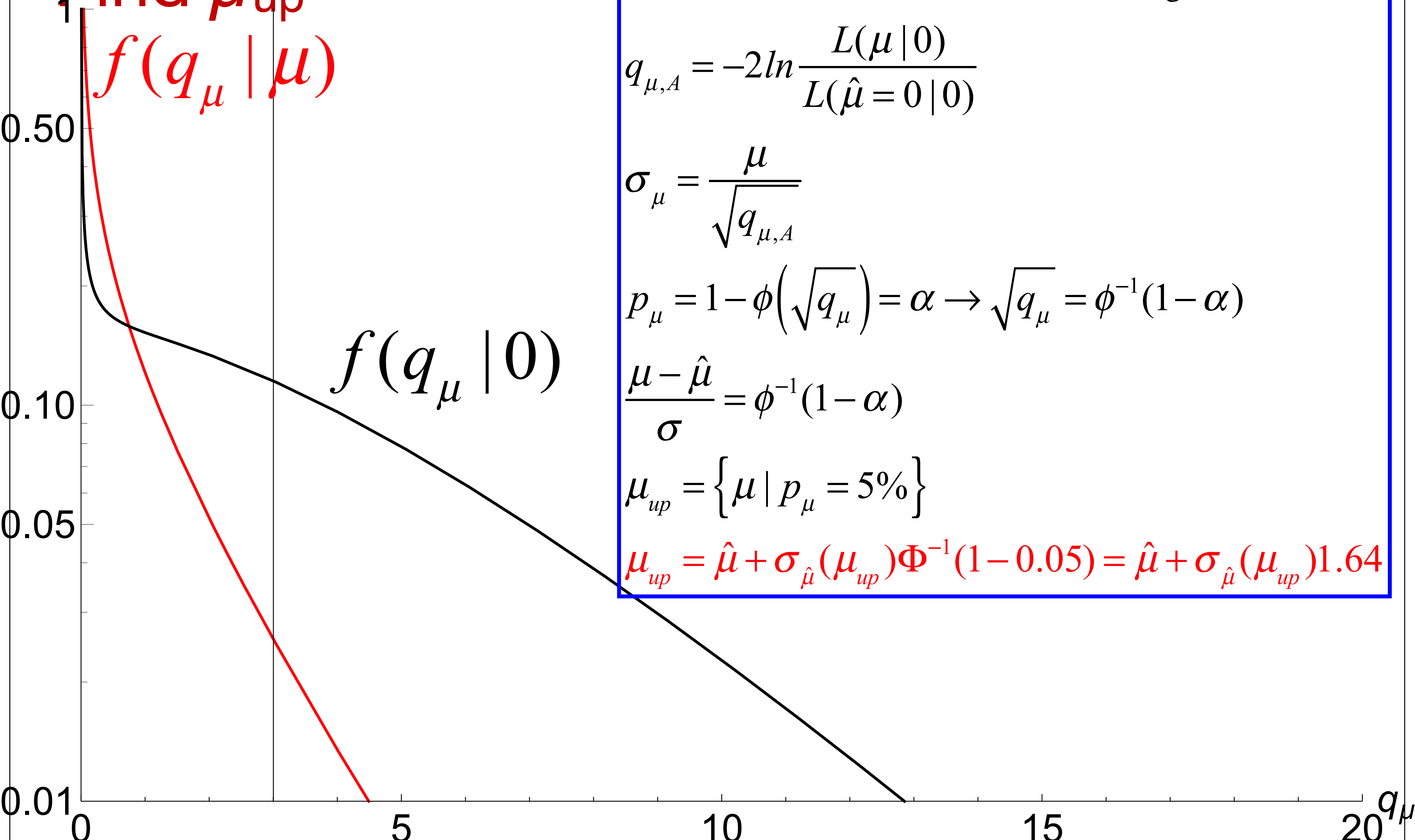
- Find the p-value of the signal hypothesis H_μ

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

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



Find μ_{up}



$$Let\ \langle \hat{\mu} \rangle = 0,\ Wald \rightarrow Z = \sqrt{q_\mu} = \frac{\mu - \hat{\mu}}{\sigma}$$

$$q_{\mu,A} = -2ln\frac{L(\mu | 0)}{L(\hat{\mu} = 0 | 0)}$$

$$\sigma_\mu = \frac{\mu}{\sqrt{q_{\mu,A}}}$$

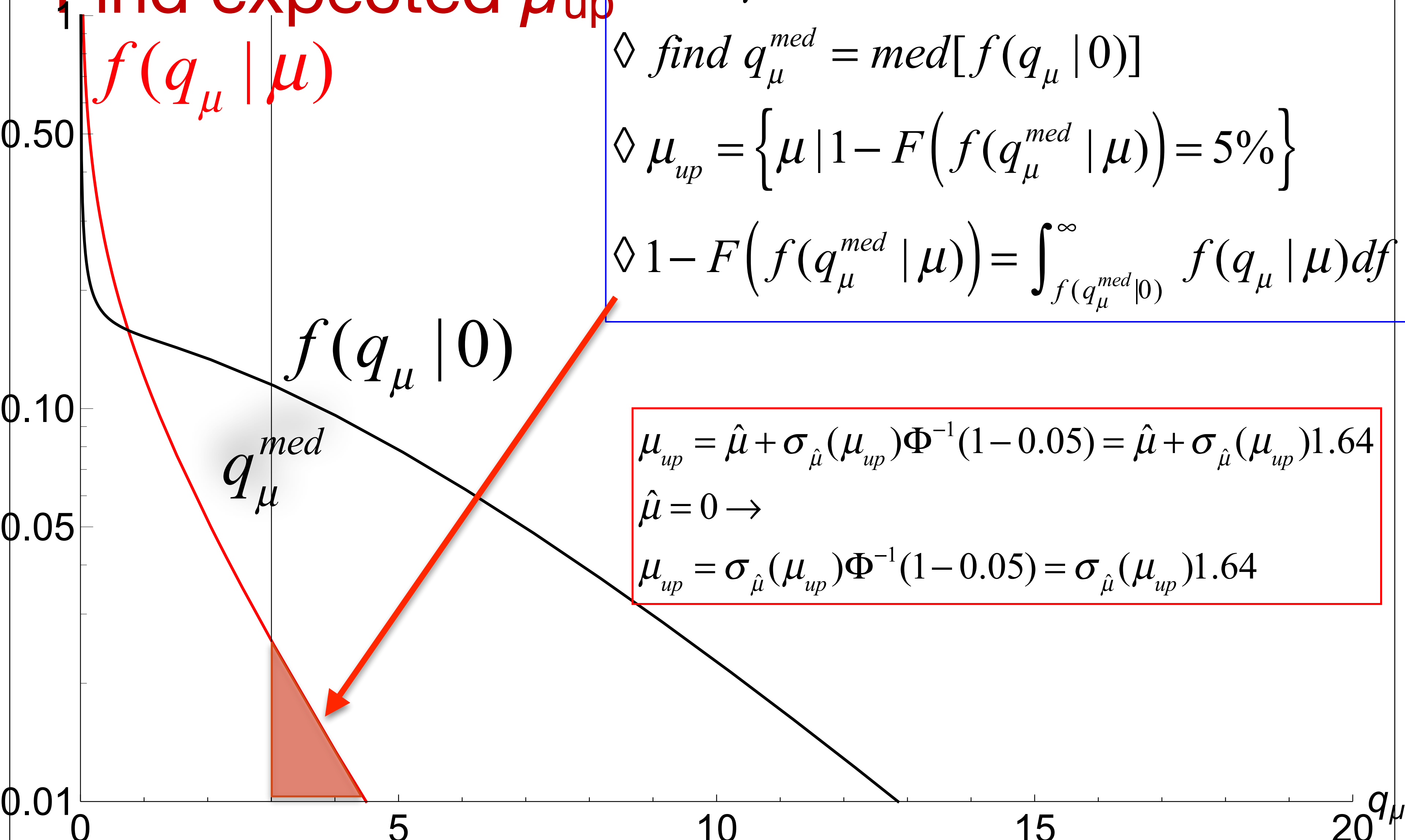
$$p_\mu = 1 - \phi\left(\sqrt{q_\mu}\right) = \alpha \rightarrow \sqrt{q_\mu} = \phi^{-1}(1-\alpha)$$

$$\frac{\mu - \hat{\mu}}{\sigma} = \phi^{-1}(1-\alpha)$$

$$\mu_{up} = \left\{ \mu \,|\, p_\mu = 5\% \right\}$$

$$\mu_{up} = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up})\Phi^{-1}(1-0.05) = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up})1.64$$

Find expected μ_{up}



$$\Diamond \ scan\ \mu$$

$$\Diamond \ find\ q_\mu^{med} = med[f(q_\mu | 0)]$$

$$\Diamond \ \mu_{up} = \left\{ \mu \,|\, 1 - F\left(f(q_\mu^{med} \,|\, \mu) \right) = 5\% \right\}$$

$$\Diamond \ 1 - F\left(f(q_\mu^{med} \,|\, \mu) \right) = \int_{f(q_\mu^{med}|0)}^{\infty} f(q_\mu \,|\, \mu) df$$

$$\mu_{up} = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up}) \Phi^{-1}(1 - 0.05) = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up}) 1.64$$

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

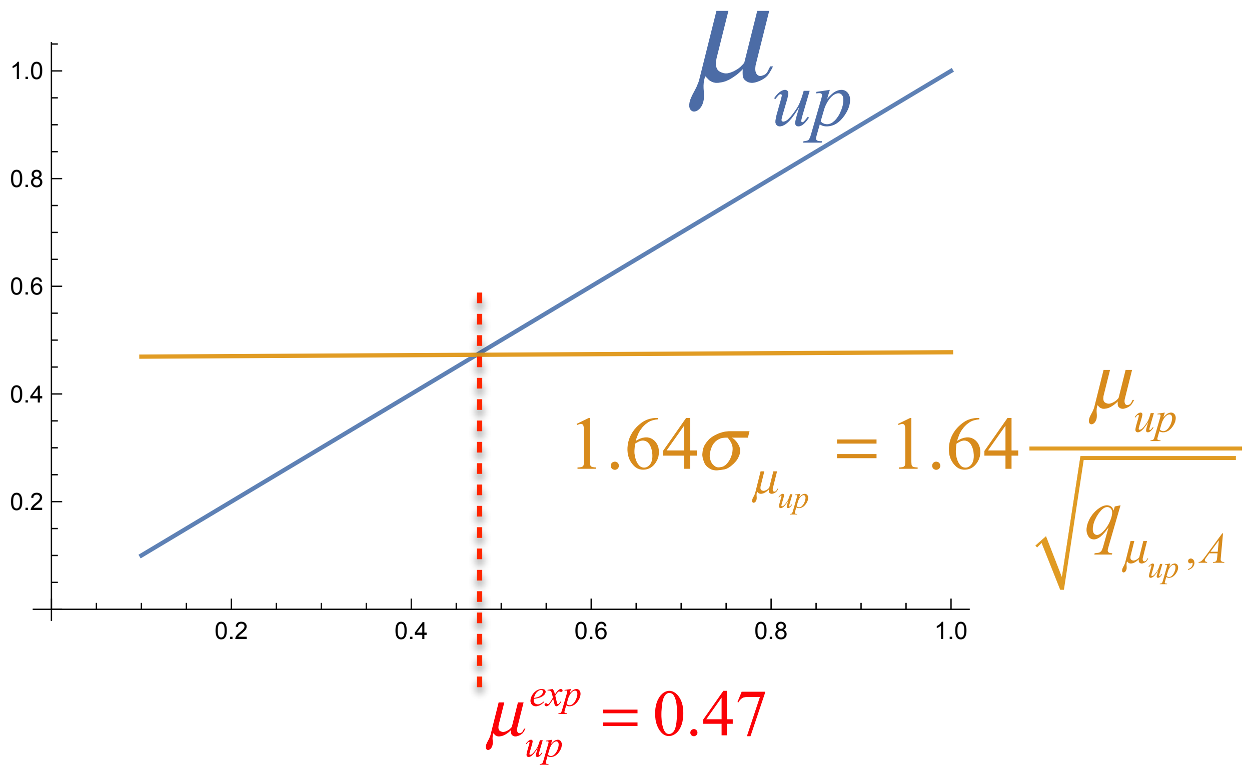
$$\mu_{up} = \sigma_{\hat{\mu}}(\mu_{up}) \Phi^{-1}(1 - 0.05) = \sigma_{\hat{\mu}}(\mu_{up}) 1.64$$

Find expected μ_{up}

$$\mu_{up} = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up})\Phi^{-1}(1-0.05) = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up})1.64$$

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

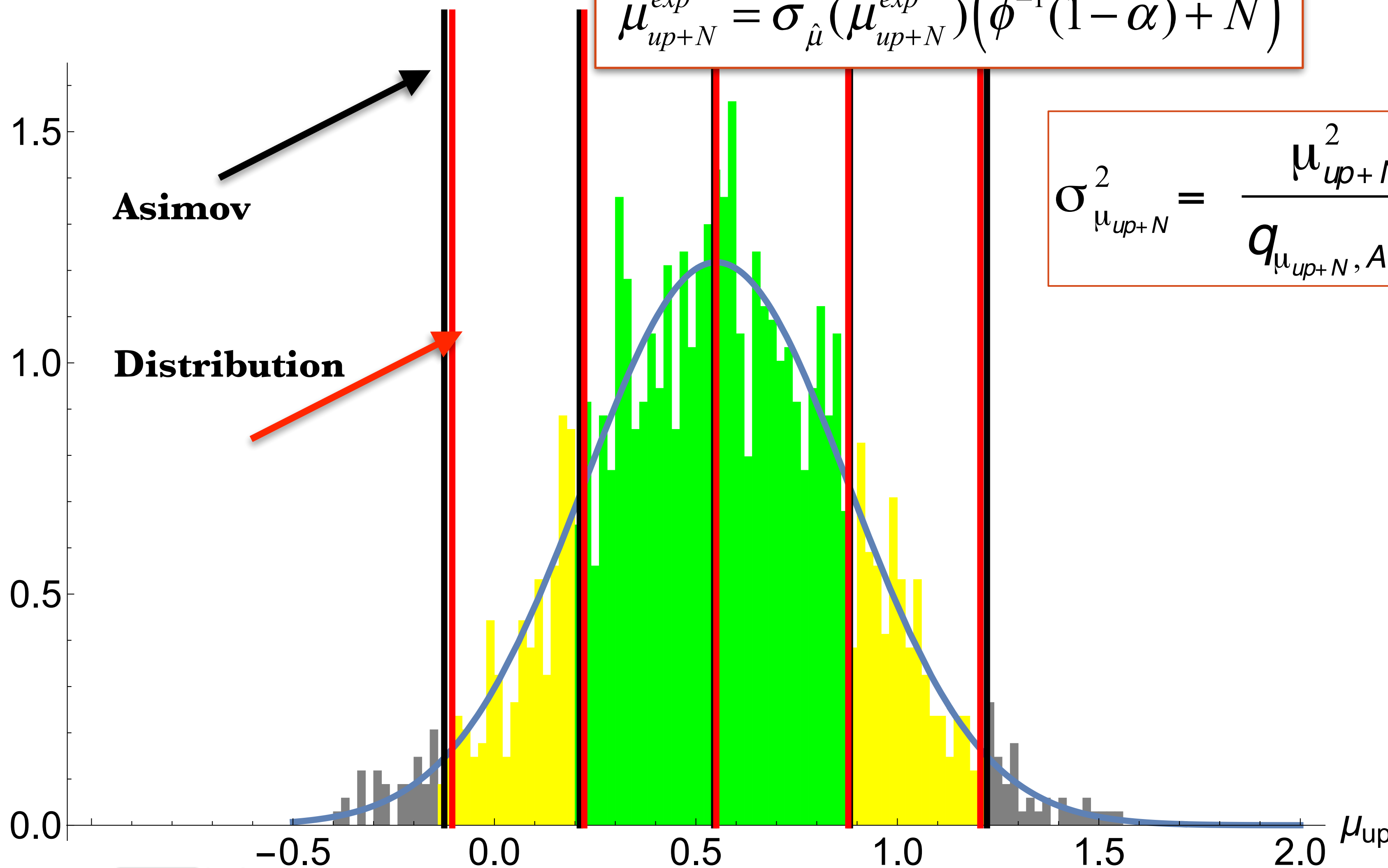
$$\mu_{up}^{exp} = \sigma_{\hat{\mu}}(\mu_{up})\Phi^{-1}(1-0.05) = \sigma_{\hat{\mu}}(\mu_{up}^{exp})1.64$$



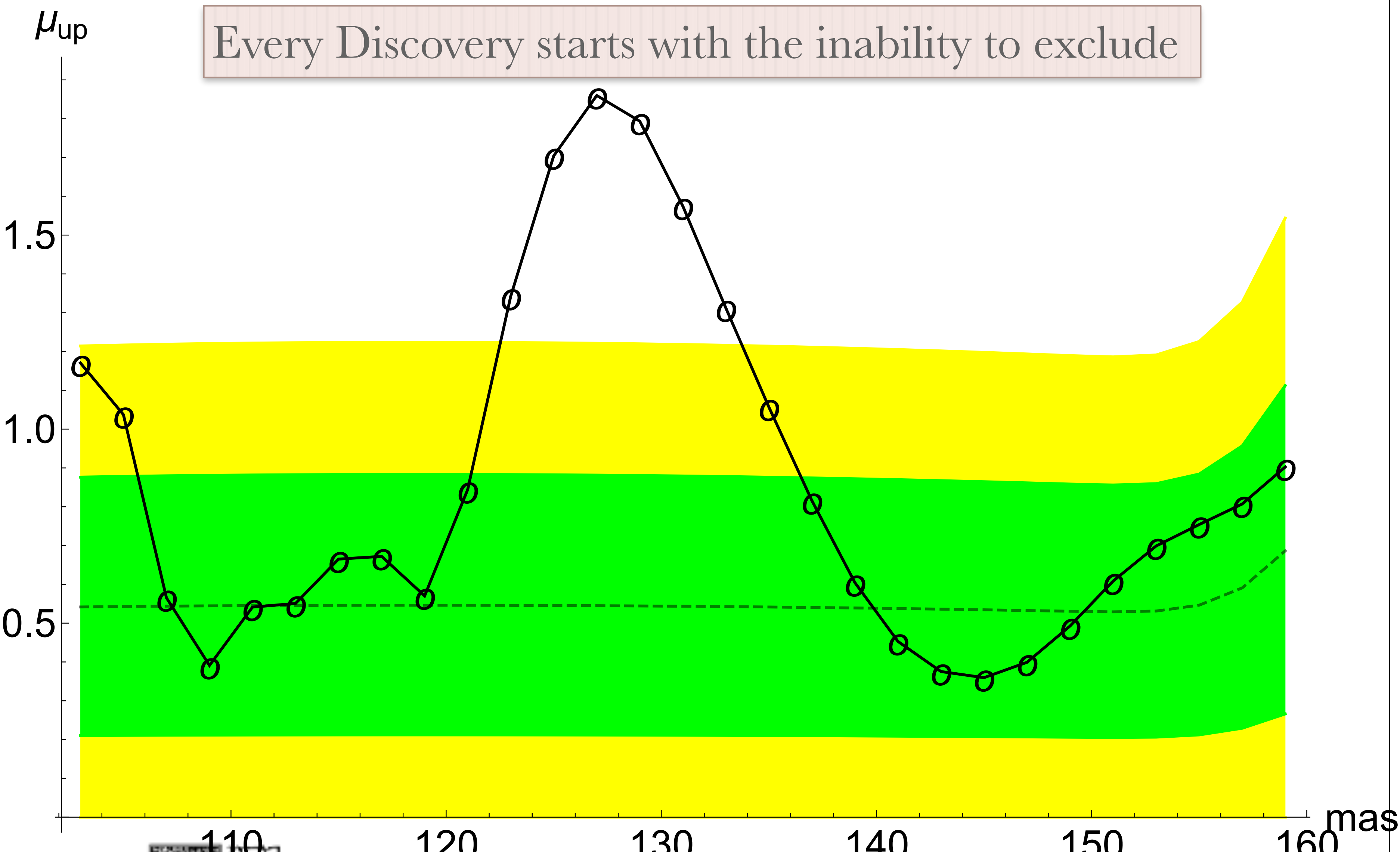
Expected μ_{up} Bands at m=125

$$\mu_{up+N}^{exp} = \sigma_{\hat{\mu}}(\mu_{up+N}^{exp})\left(\phi^{-1}(1-\alpha)+N\right)$$

$$\sigma^2_{\mu_{up+N}} = \frac{\mu^2_{up+N}}{q_{\mu_{up+N},A}}$$



Brazil Plot

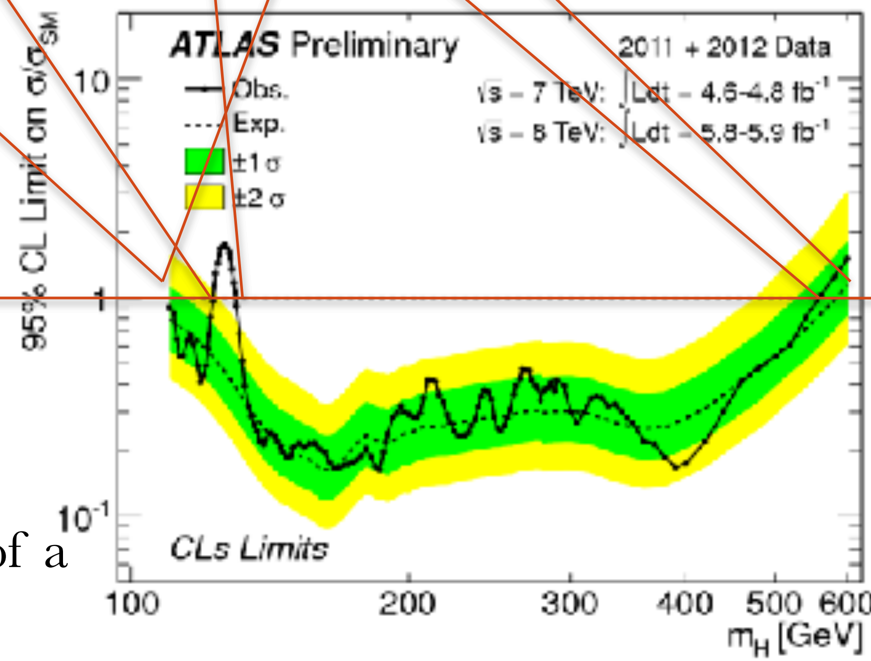


Every Discovery starts with the inability to exclude
μ_{up}
1.5
1.0
0.5
110
120
130
140
150
160
mas

Understanding the Brazil Plot

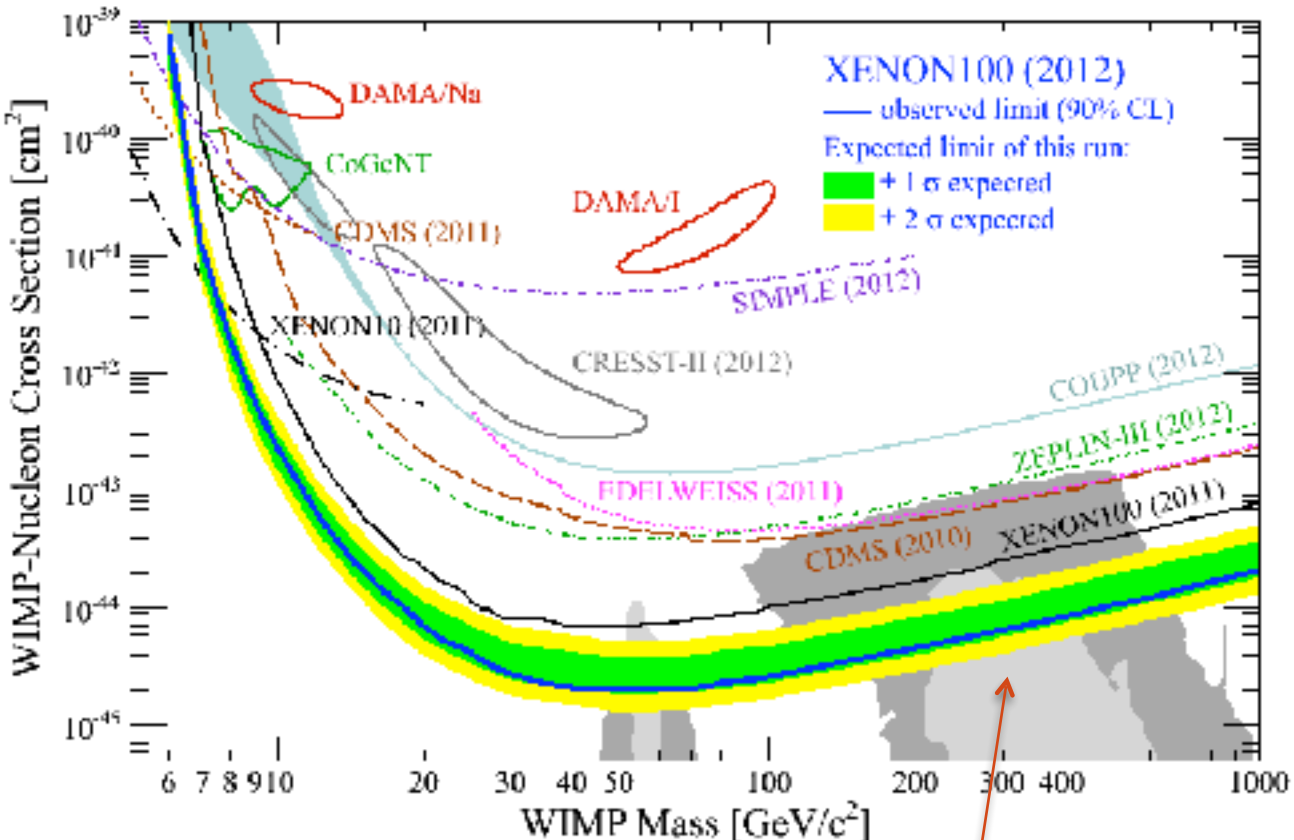
The expected 95% CL exclusion region covers the m_H range from 110 GeV to 582 GeV. The observed 95% CL exclusion regions are from 110 GeV to 122.6 GeV and 129.7 GeV to 558 GeV. The addition of

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



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

Implications in Astro-Particle Physics

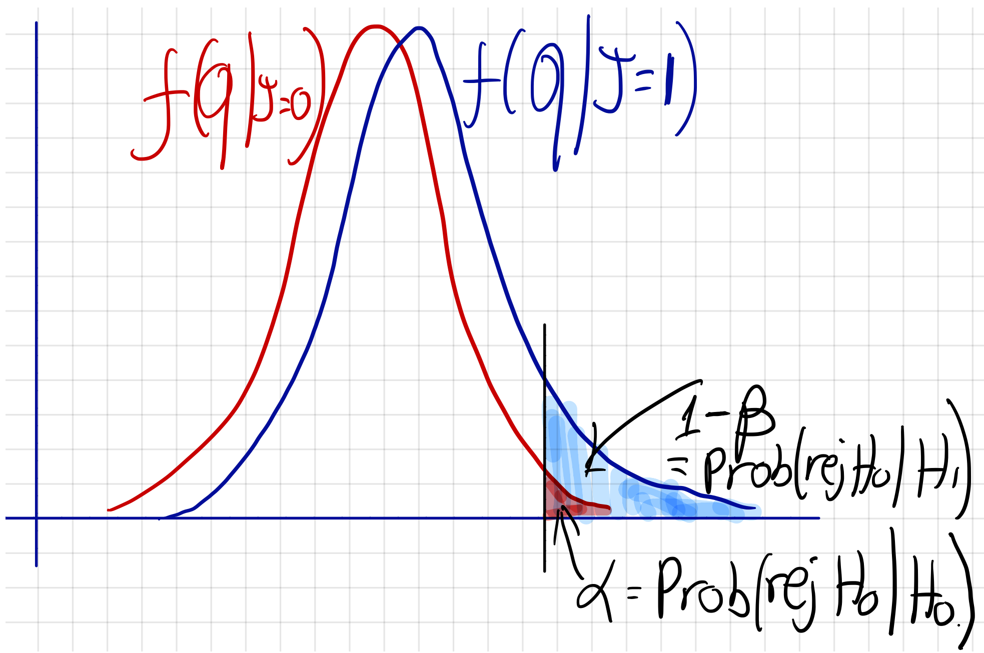


The lack of events in spite of an expected background allows us to set a better limit than the expected

Revised CLs (and Asymptotic)

CLs

Birnbaum (1962) *suggested that* $\alpha / 1 - \beta$ (*significance* / *power*) *should be used as a measure of the strength of a statistical test, rather than* α *alone*



$$p = 5\% \rightarrow p' = 5\% / 0.155 = 32\%$$

$$p' \equiv CL_S$$

$$p'_\mu = \frac{p_\mu}{1 - p_0}$$

The CLs method
Was brought into
HEP By Alex Read (2002)
A.L. Read,
Presentation of search results:
The CL(s) technique,
" J.\ Phys.\ G {\bf 28}, 2693 (2002).

Birnbaum was re-discovered later
By O. Vitells

The Asymptotic and CLs

$$p'_\mu = \frac{p_\mu}{1-p_0}$$

$$p_\mu = 1-\Phi\left(\sqrt{q_{\mu,obs}}\right)$$

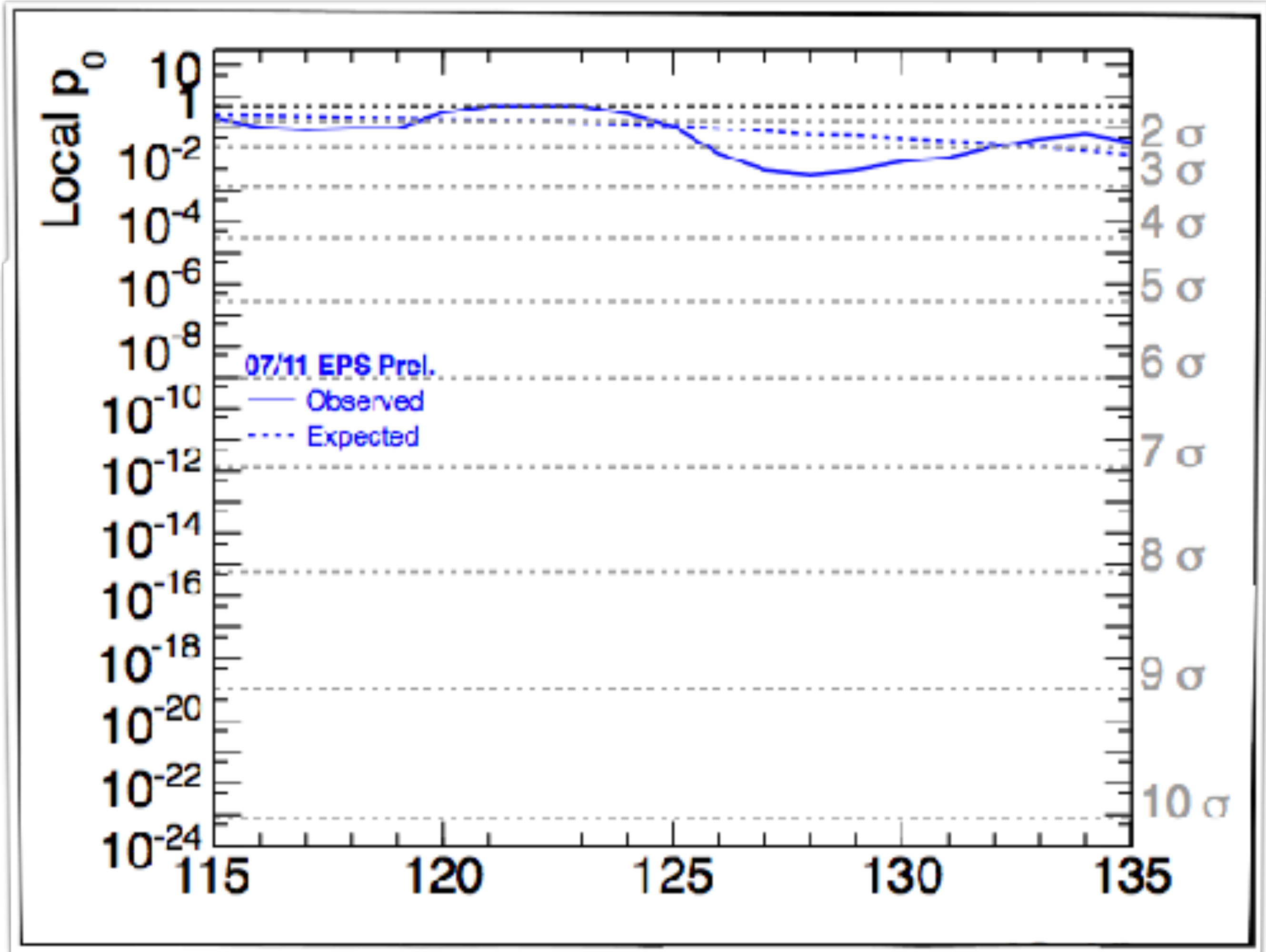
$$p_0 = \Phi\left(\sqrt{q_{\mu,obs}} - \sqrt{q_{\mu,A}}\right) \rightarrow$$

$$p'_\mu = \frac{1-\Phi\left(\sqrt{q_{\mu,obs}}\right)}{\Phi\left(\sqrt{q_{\mu,A}} - \sqrt{q_{\mu,obs}}\right)} \Rightarrow$$

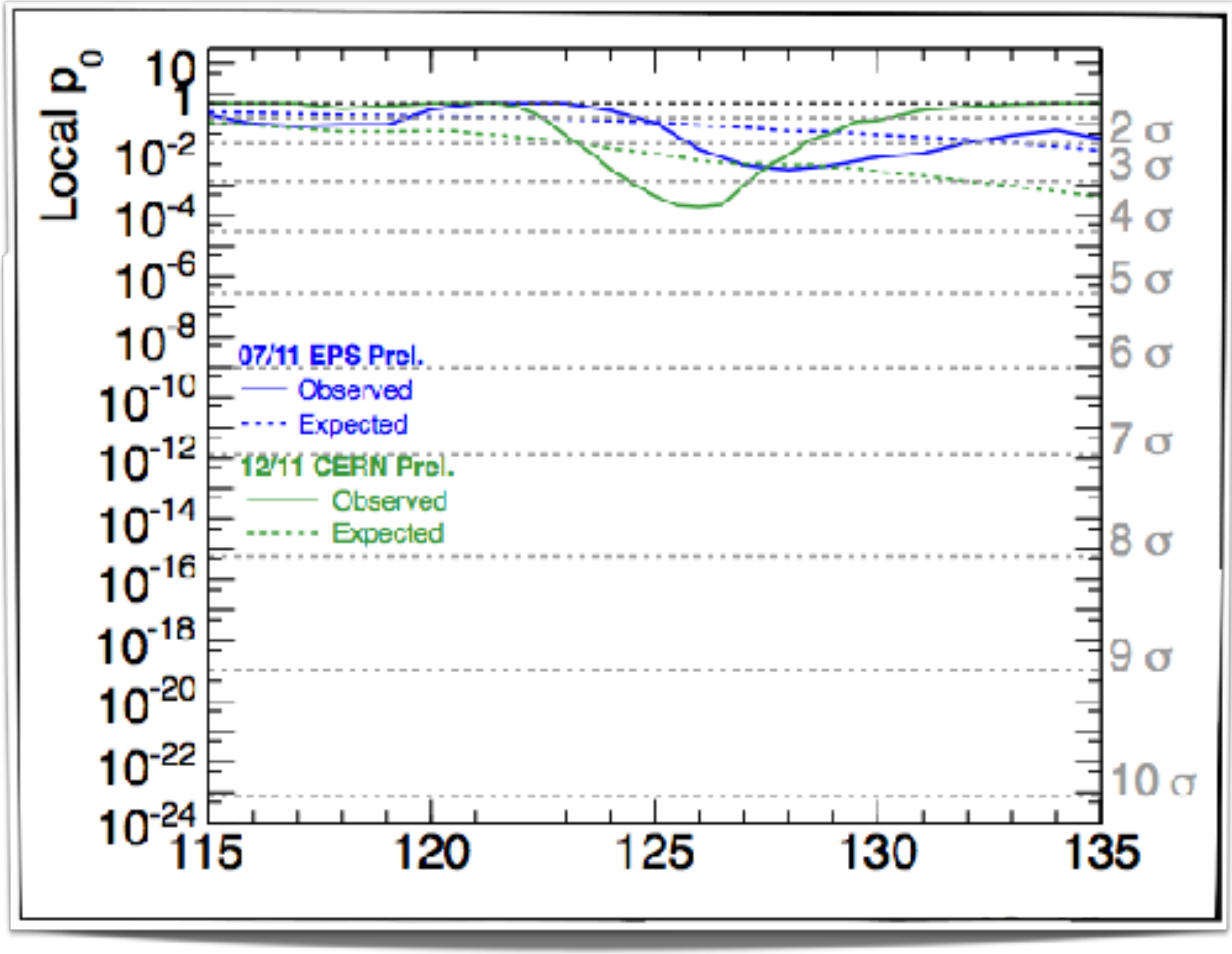
$$scan\ \mu\ and\ find\ \mu_{up}$$

$$\mu_{up} = \left\{ \mu \left| \frac{1-\Phi\left(\sqrt{q_{\mu,obs}}\right)}{\Phi\left(\sqrt{q_{\mu,A}} - \sqrt{q_{\mu,obs}}\right)} = 5\% \right. \right\}$$

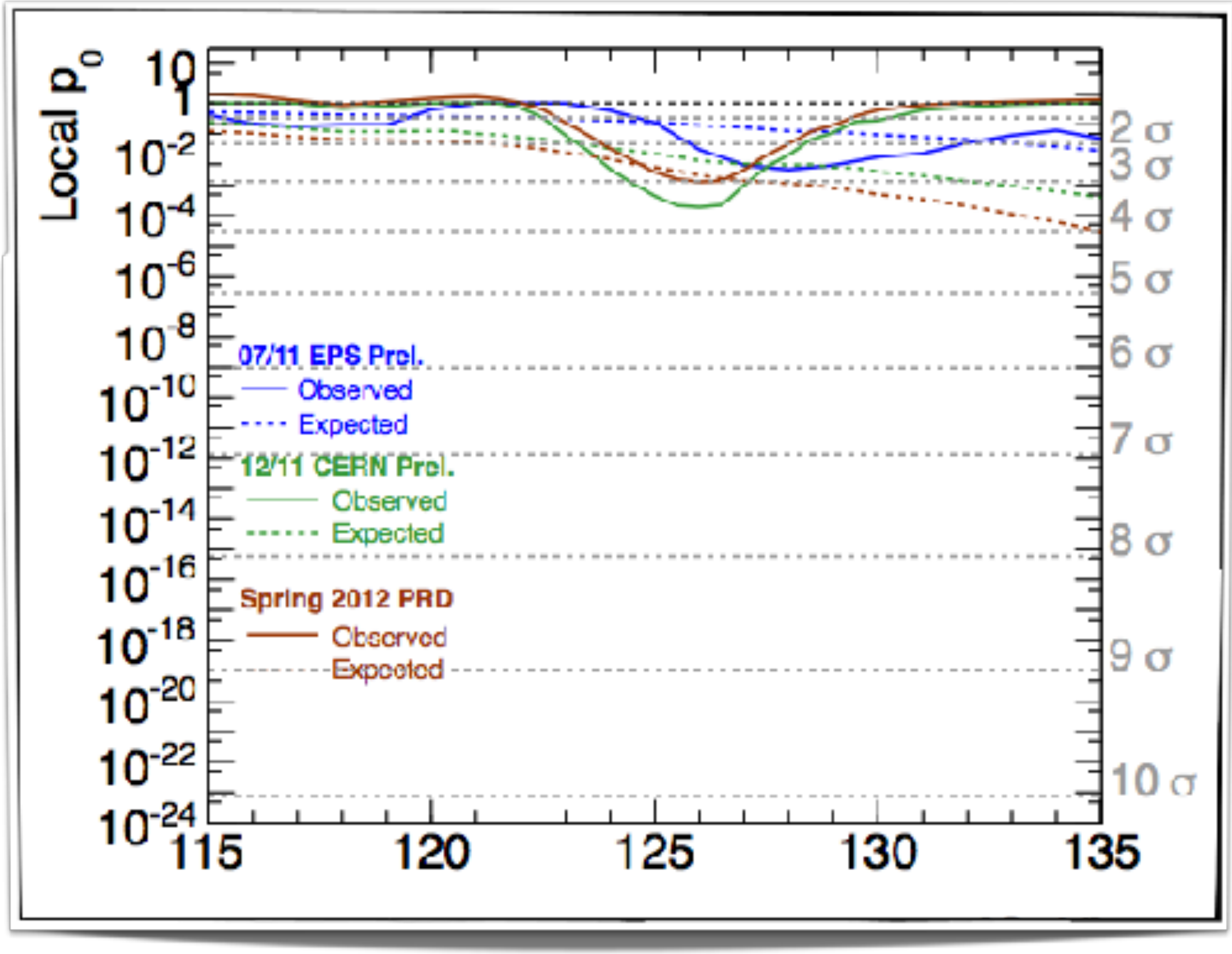
Examples (if time permits)

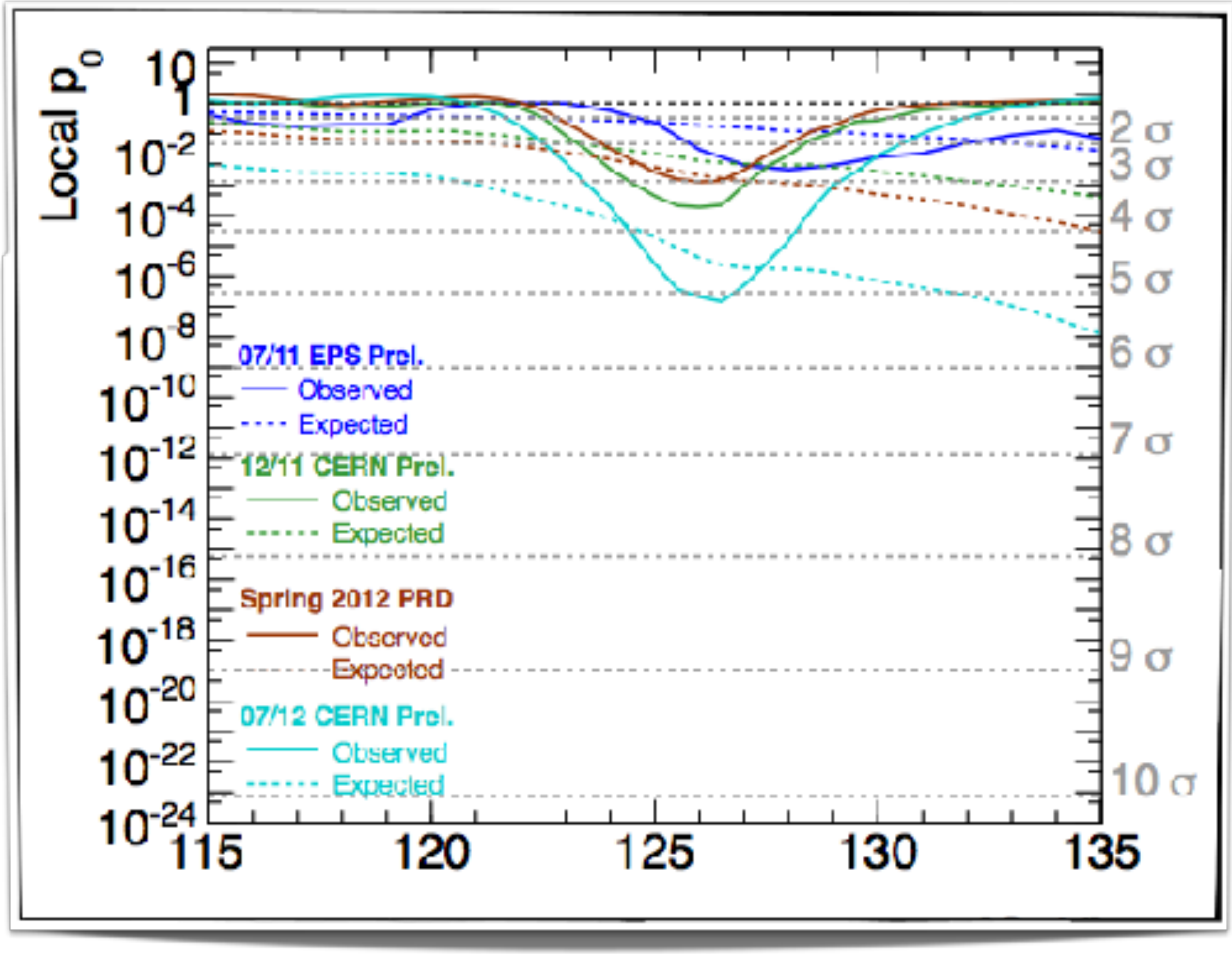


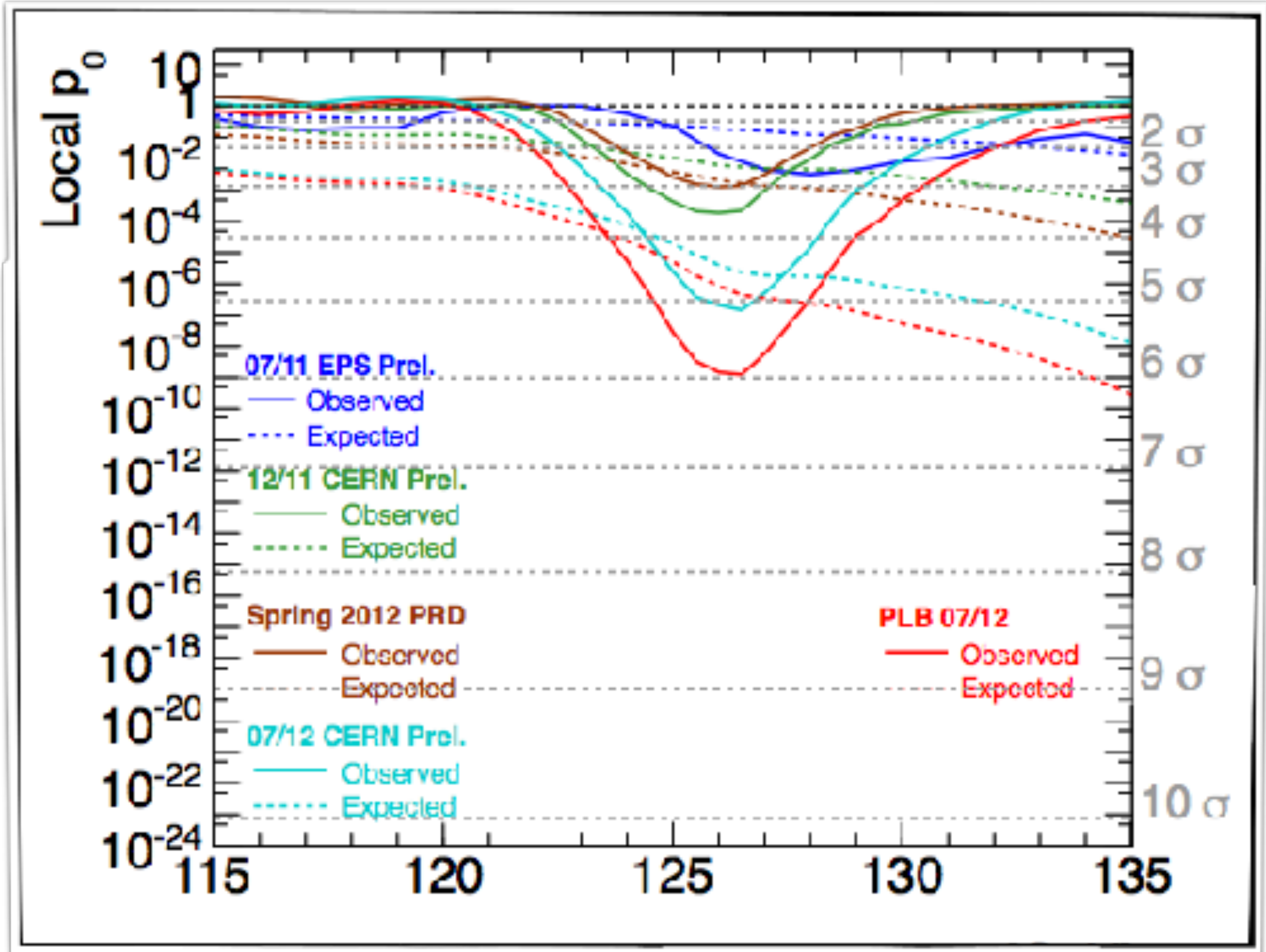
Local p_0
10
1
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
10^{-20}
10^{-22}
10^{-24}
115
120
125
130
135
2 σ
3 σ
4 σ
5 σ
6 σ
7 σ
8 σ
9 σ
10 σ
07/11 EPS Prel.
Observed
Expected



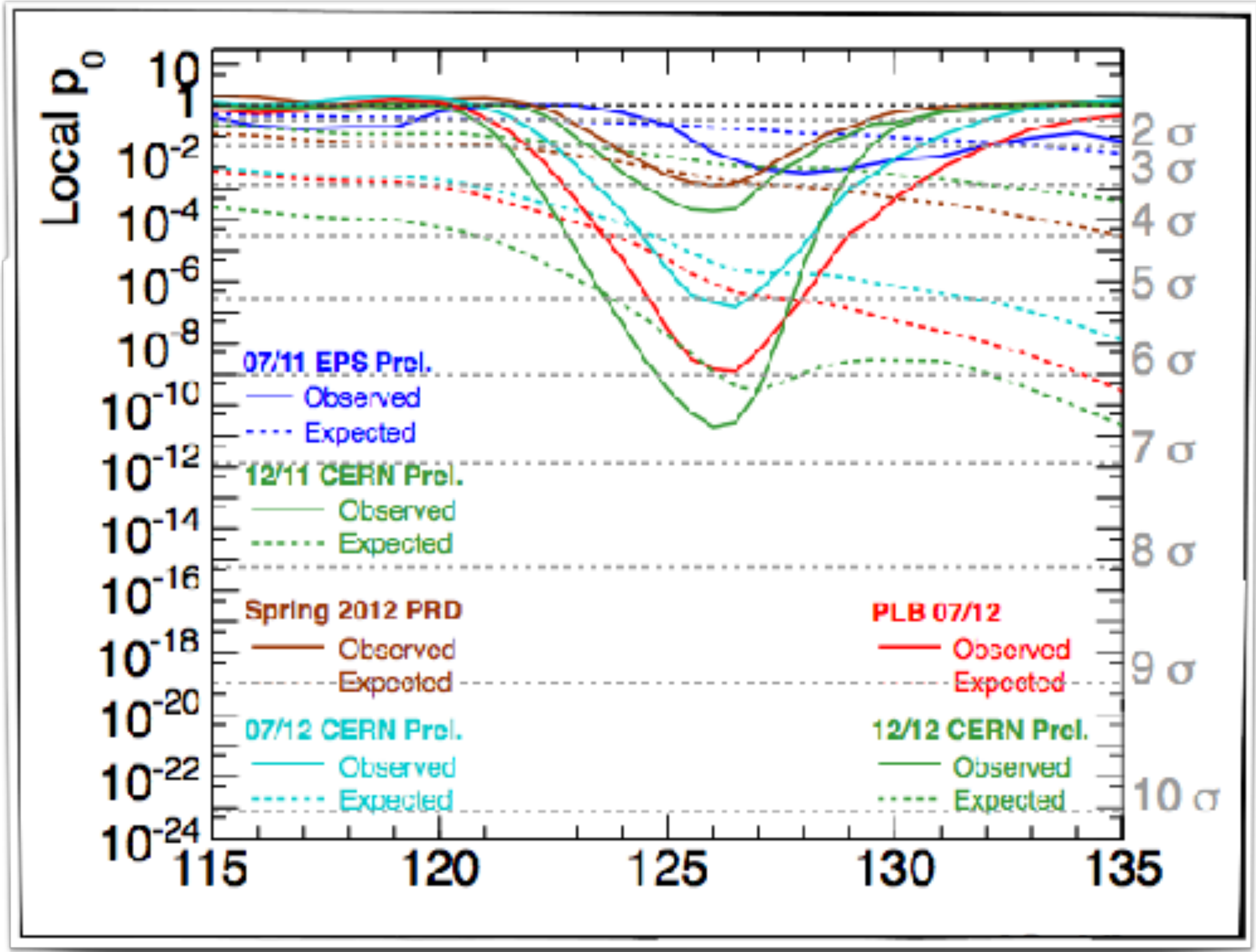
Local p_0
10
1
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
10^{-20}
10^{-22}
10^{-24}
115
120
125
130
135
2 σ
3 σ
4 σ
5 σ
6 σ
7 σ
8 σ
9 σ
10 σ
07/11 EPS Prel.
Observed
Expected
12/11 CERN Prel.
Observed
Expected


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

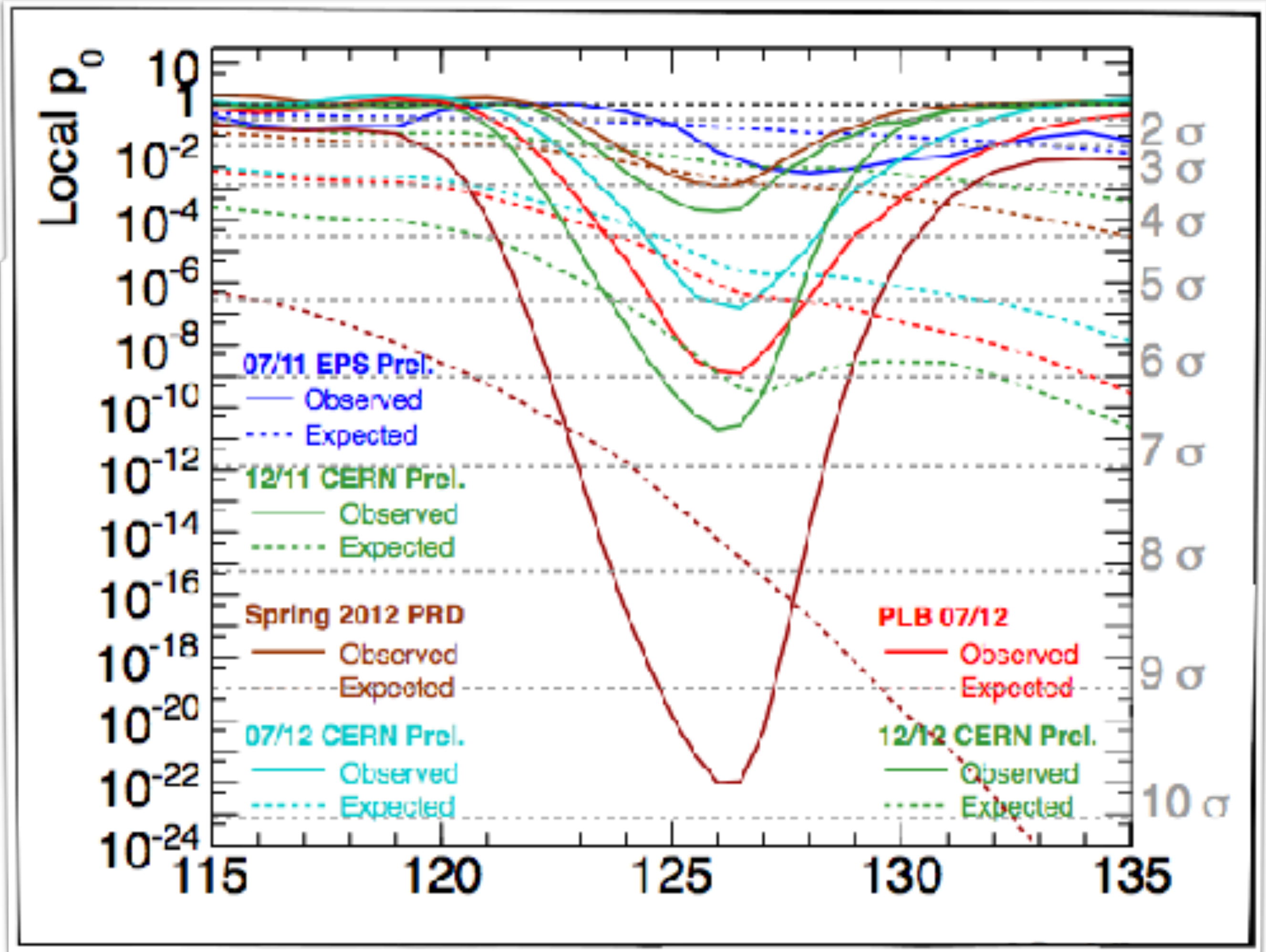

Local p_0
10
1
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
10^{-20}
10^{-22}
10^{-24}
115
120
125
130
135
2 σ
3 σ
4 σ
5 σ
6 σ
7 σ
8 σ
9 σ
10 σ
07/11 EPS Prel.
Observed
Expected
12/11 CERN Prel.
Observed
Expected
Spring 2012 PRD
Observed
Expected
07/12 CERN Prel.
Observed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



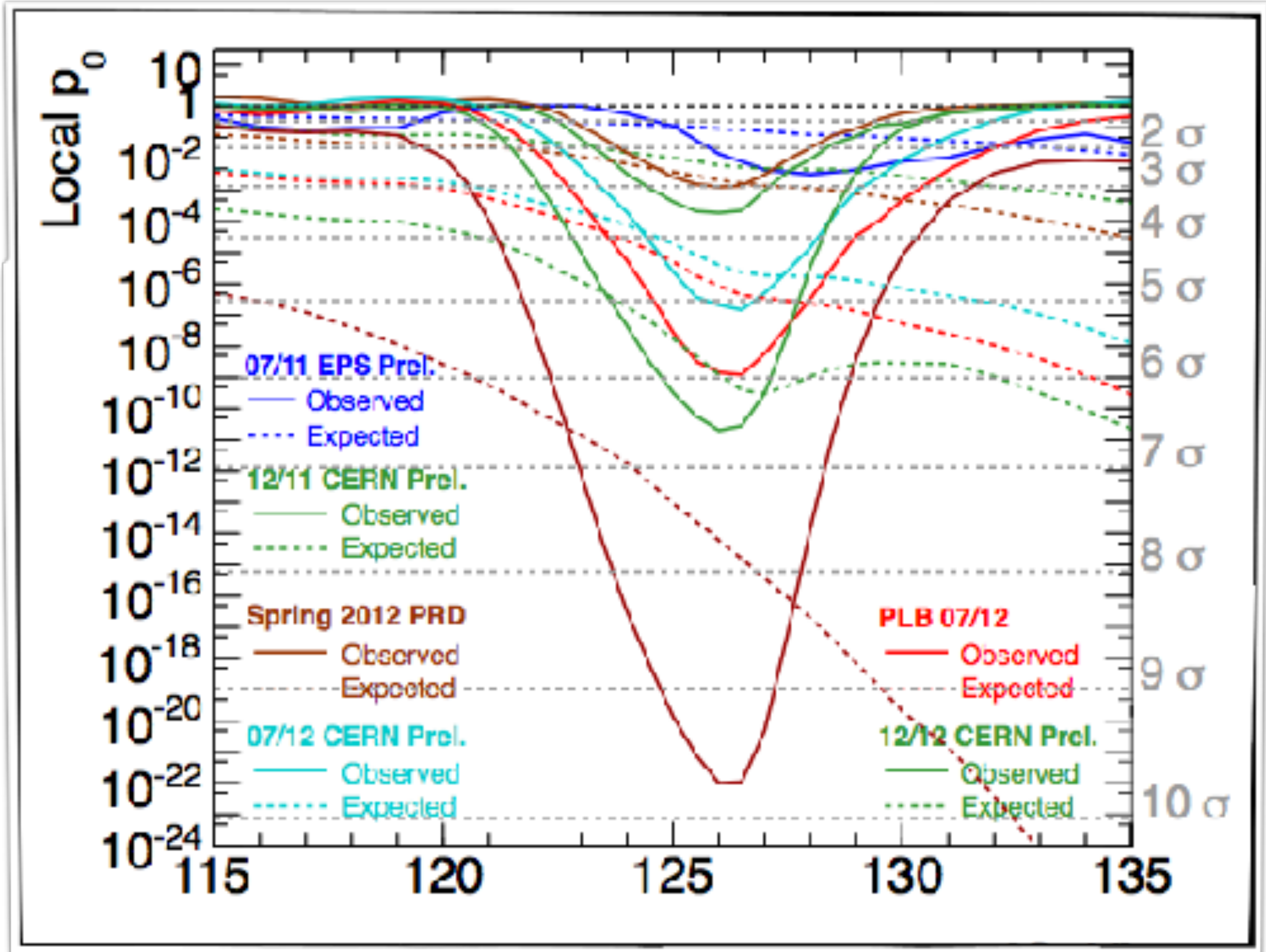
Local p_0
10
1
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
10^{-20}
10^{-22}
10^{-24}
115
120
125
130
135
2 σ
3 σ
4 σ
5 σ
6 σ
7 σ
8 σ
9 σ
10 σ
07/11 EPS Prel.
Observed
Expected
12/11 CERN Prel.
Observed
Expected
Spring 2012 PRD
Observed
Expected
07/12 CERN Prel.
Observed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



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


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

p0


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

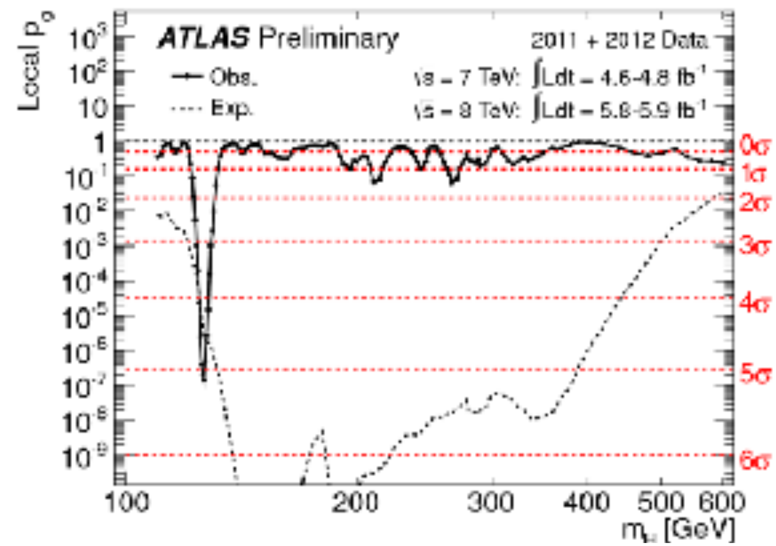
p0 and the expected p0

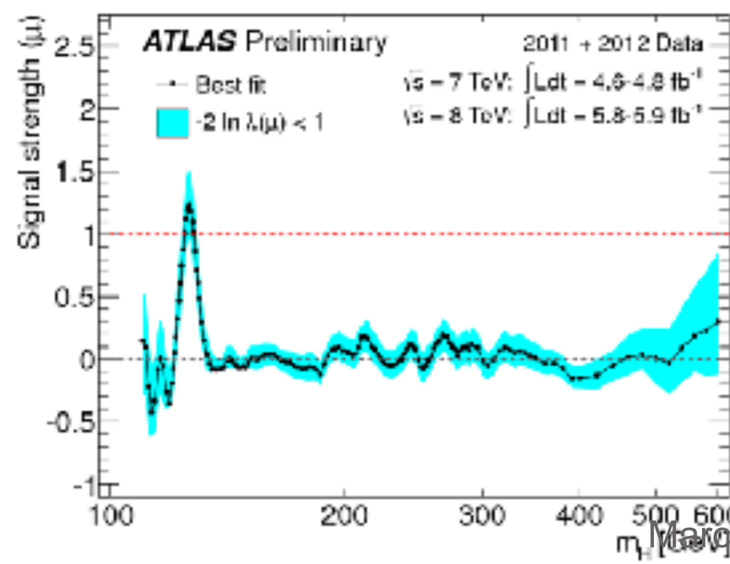
$$p_0 = \int_{q_{0,obs}}^{\infty} f(q_0|0)\, dq_0$$

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

Small p0 leads to an observation

A tiny p0 leads to a discovery



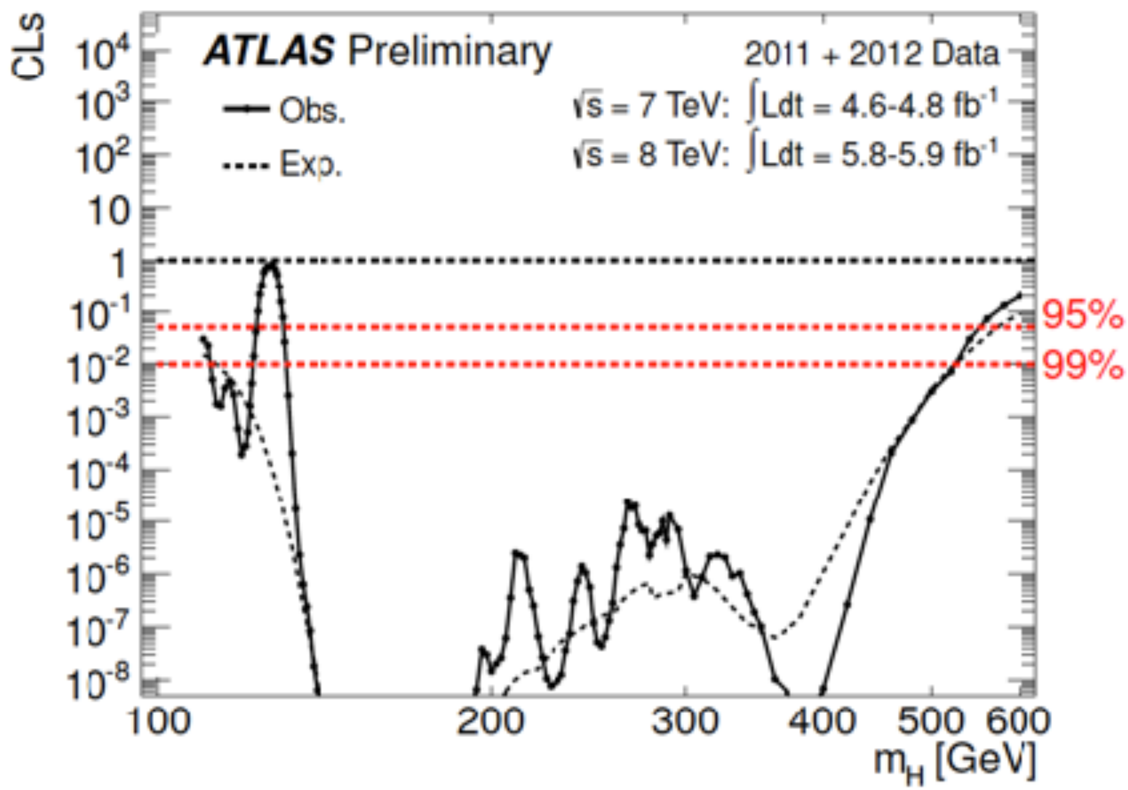


$$p = \int_Z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 1 - \Phi(Z)$$

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

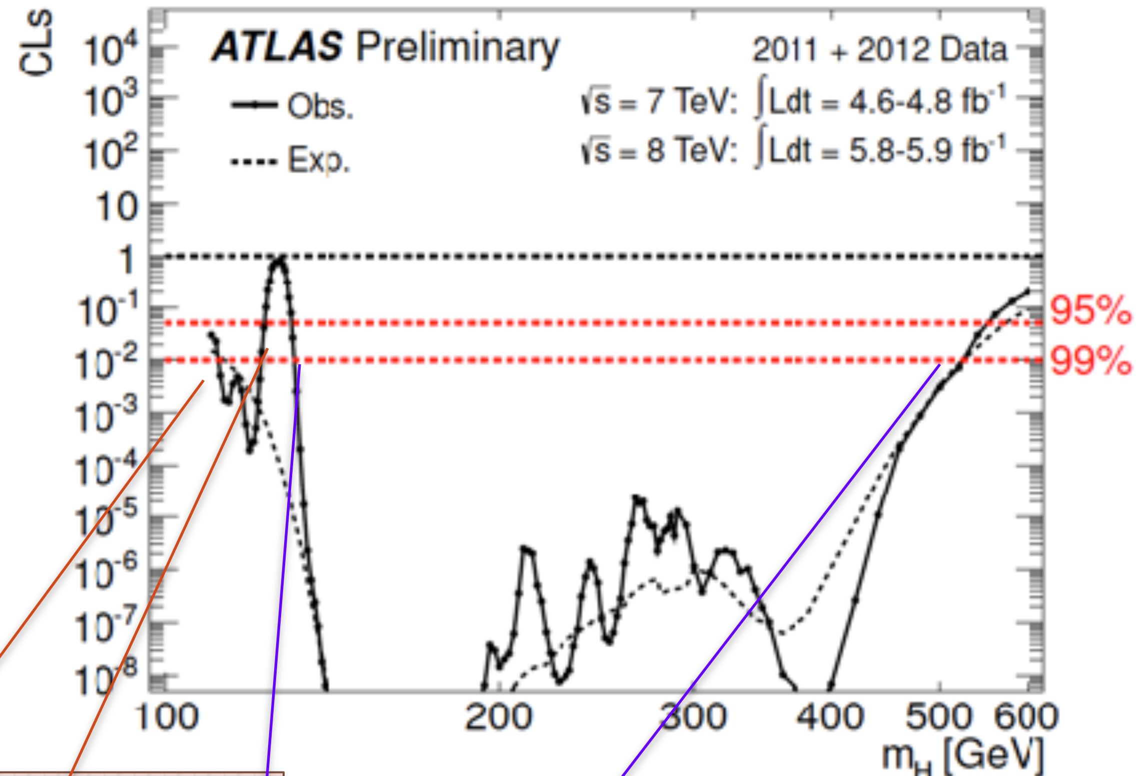
Understanding the CLs plot

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



Understanding the CLs plot

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



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

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

More Magic (if time permits)

The New $s/\sqrt{b}$

The new $s/\sqrt{b}$

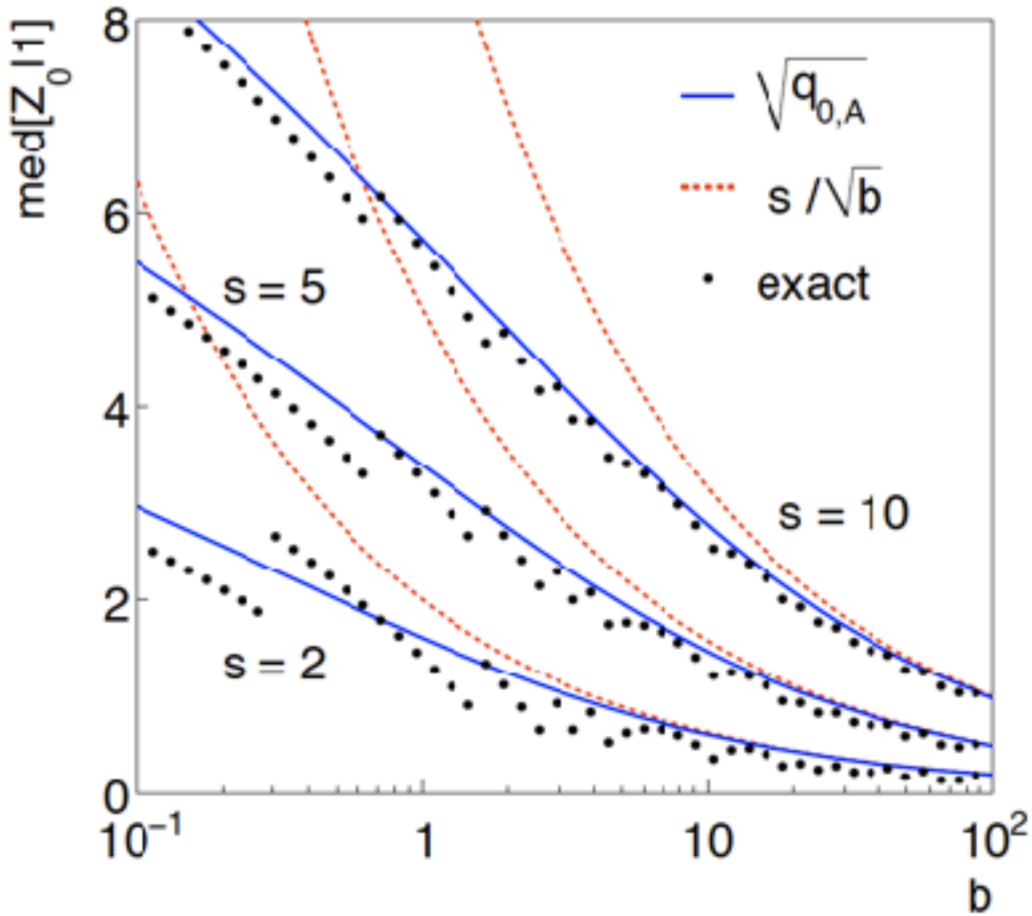
$$Z_A = \sqrt{q_{0,A}}$$

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

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

The New $s/\sqrt{b}$

$s/\sqrt{b}$?



The new $s/\sqrt{b}$

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

Taking Background Systematics into Account

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

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

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

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

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

But even that formula can be omproved using the Asimov formalism

Significance with systematics

- We find (G. Cowan)

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

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

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

- So the “intuitive” formula can be justified as a limiting case of the significance from the profile likelihood ratio test evaluated with the Asimov data set.

Significance with systematics