



NAMIBIA UNIVERSITY
OF SCIENCE AND TECHNOLOGY



UNAM
UNIVERSITY OF NAMIBIA



African School of Fundamental
Physics and Applications

Fifth African School of Fundamental Physics and its Applications

Search and Discovery Statistics in HEP Lecture 3




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 1-2: Basic Concepts
Histograms, PDF, Testing Hypotheses,
LR as a Test Statistics, p-value, POWER, CLs
Measurements
-  Lecture 3: Wald Theorem, Asymptotic Formalism, Asimov Data Set, Feldman-Cousins, PL & CLs
-  Lecture 4: Asimov Significance
Look Elsewhere Effect
1D LEE the non-intuitive thumb rule
(upcrossings, trial $\# \sim Z$)
2D LEE (Euler Characteristic)

Profile Likelihood & Wilks Theorem



Wilks theorem in the presence of NPs

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

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

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

Bump Hunt

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

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

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

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

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



Classification of Test Statistics

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

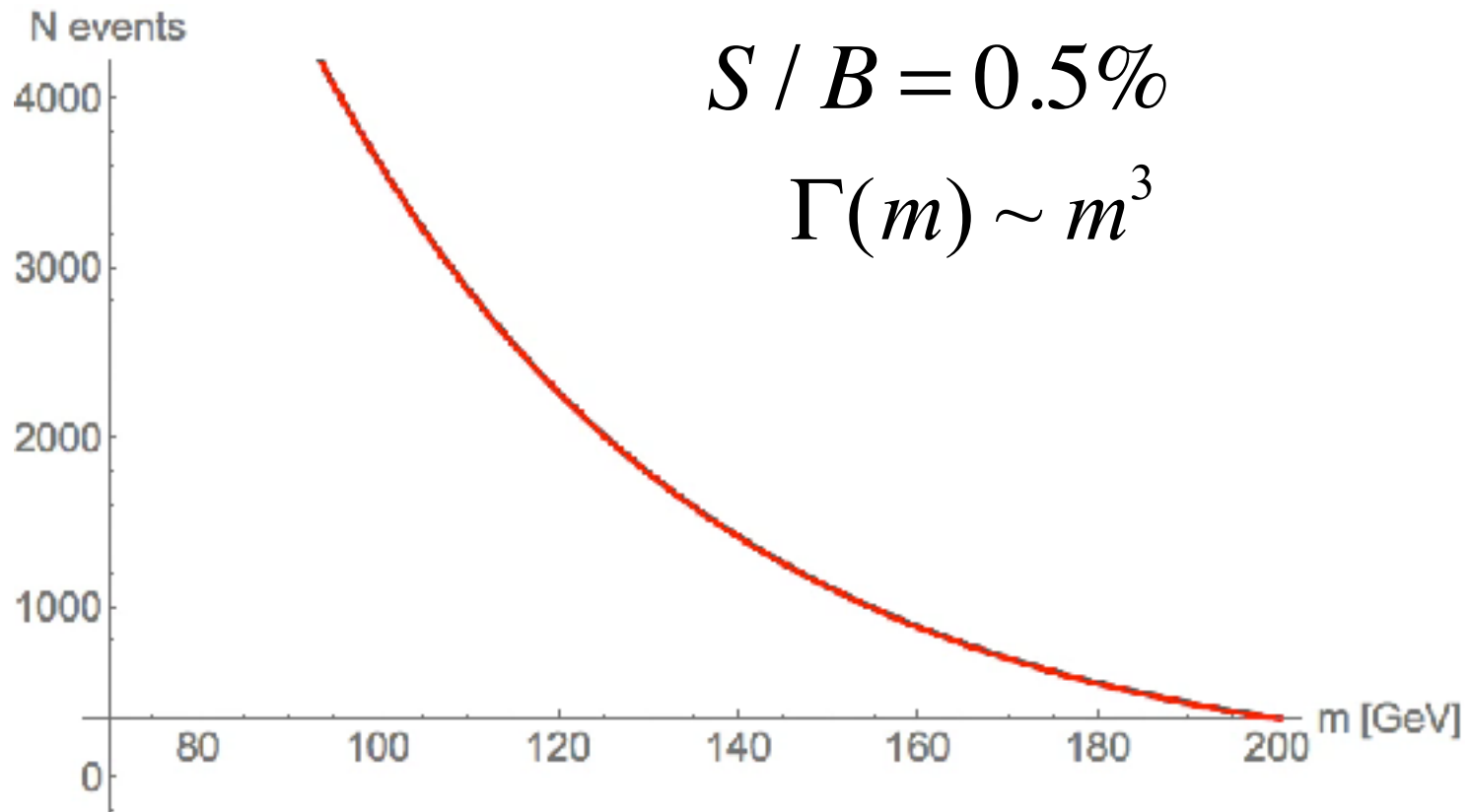


Study Case 2: Bump Hunt

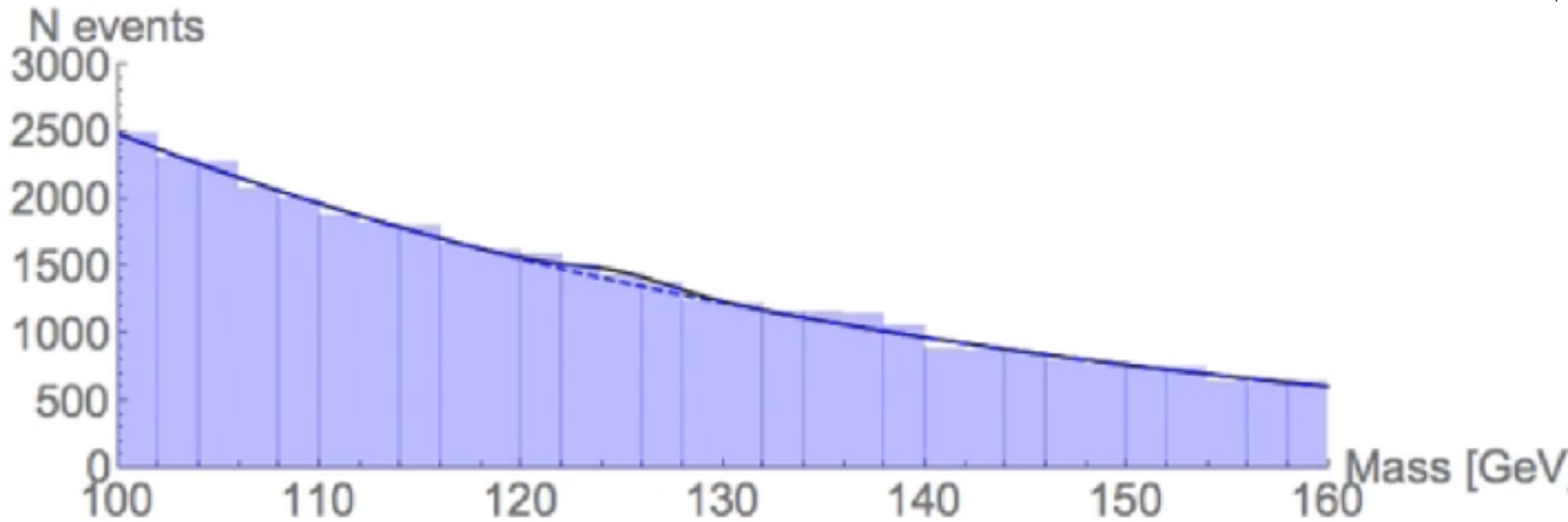


Bump Hunt

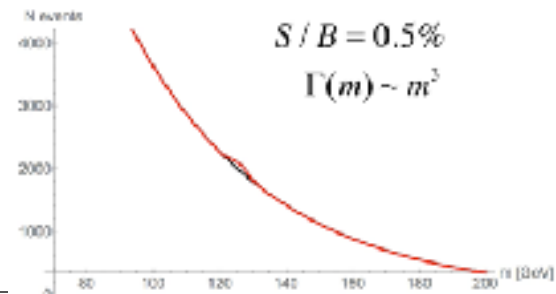
Gamma Gamma like BG and a Gaussian signal on top of it



A GammaGammaLike Signal



Luminosity is the number of events in the histogram



Asymptotic Approximation

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) 2601



Kyle
Cranmer

Glen
Cowan

Ofer
Vitells

E.G.

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

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

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



Wald Theorem

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

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

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

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

(Use the Asimov Dataset to estimate σ)

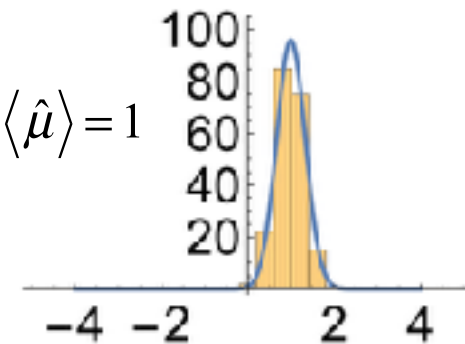
$f(t_\mu | \mu')$ follows a noncentral Chi squared distribution

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

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

N is the sample size

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



Wald Theorem

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

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

N is the sample size

$f(t_{\mu} | \mu')$ follows a noncentral Chi squared distribution

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

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

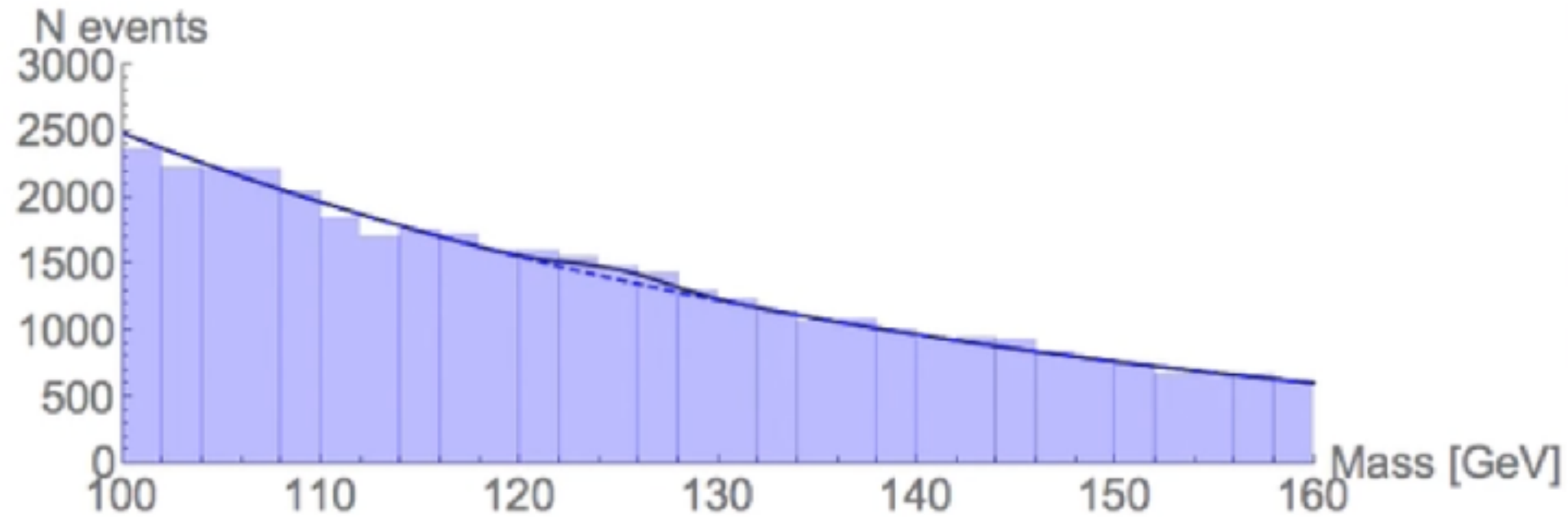
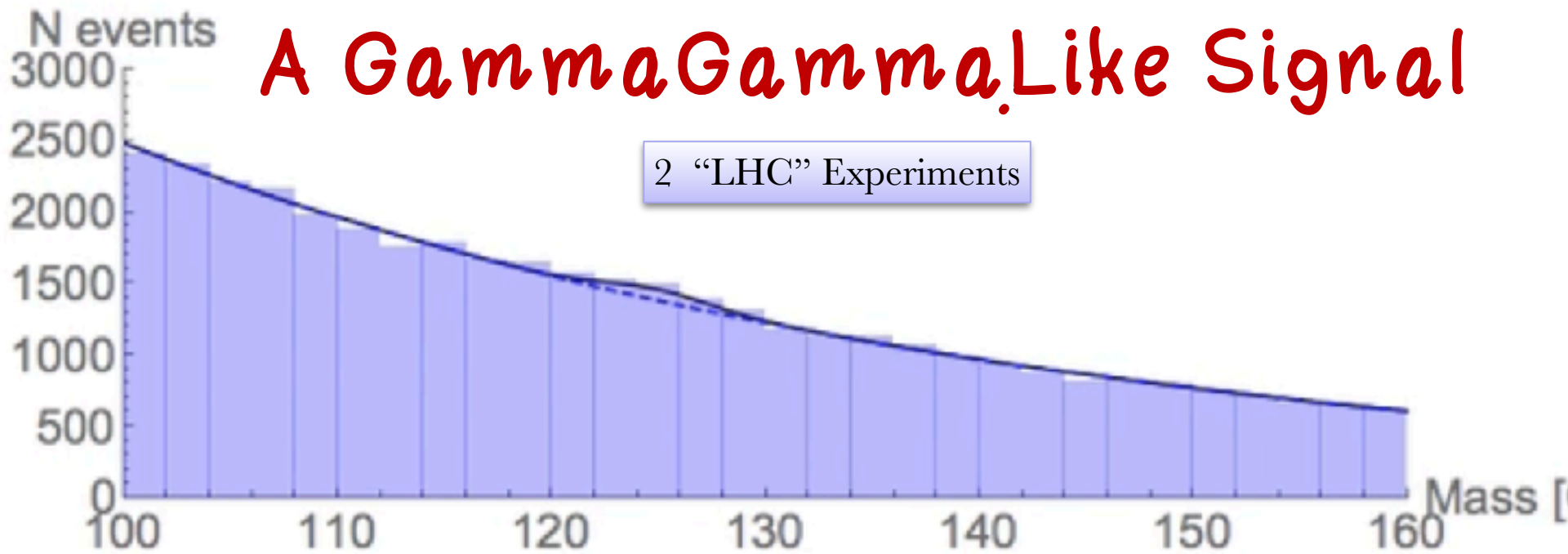
for $\mu' = \mu$ we retrieve Wilks' theorem

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



A GammaGammaLike Signal

2 "LHC" Experiments

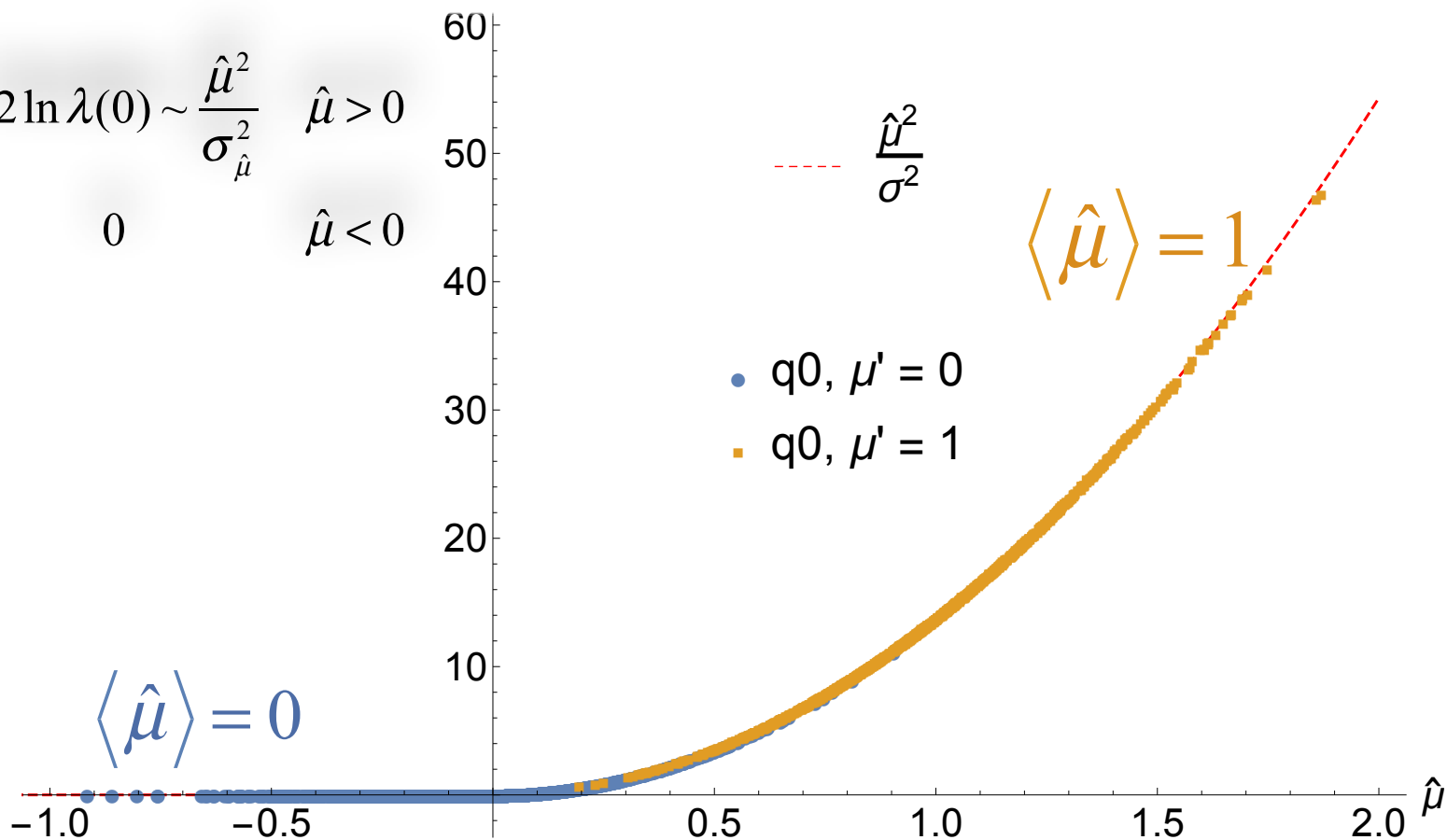


Wald Theorem Demonstration

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

$$\mathcal{L} = 60000$$

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



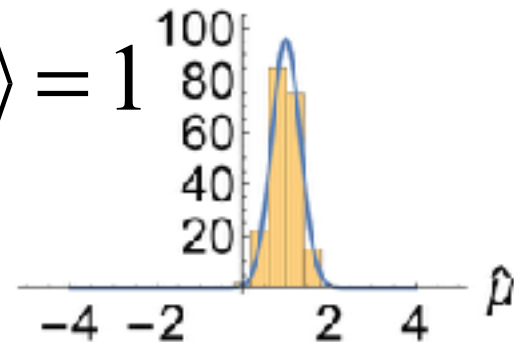
Wald Theorem Demonstration $\mathcal{L} = 60000$

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

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

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

$$\frac{(\mu - \hat{\mu})^2}{\sigma^2}$$



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

- $q_1, \mu' = 0$
- $q_1, \mu' = 1$

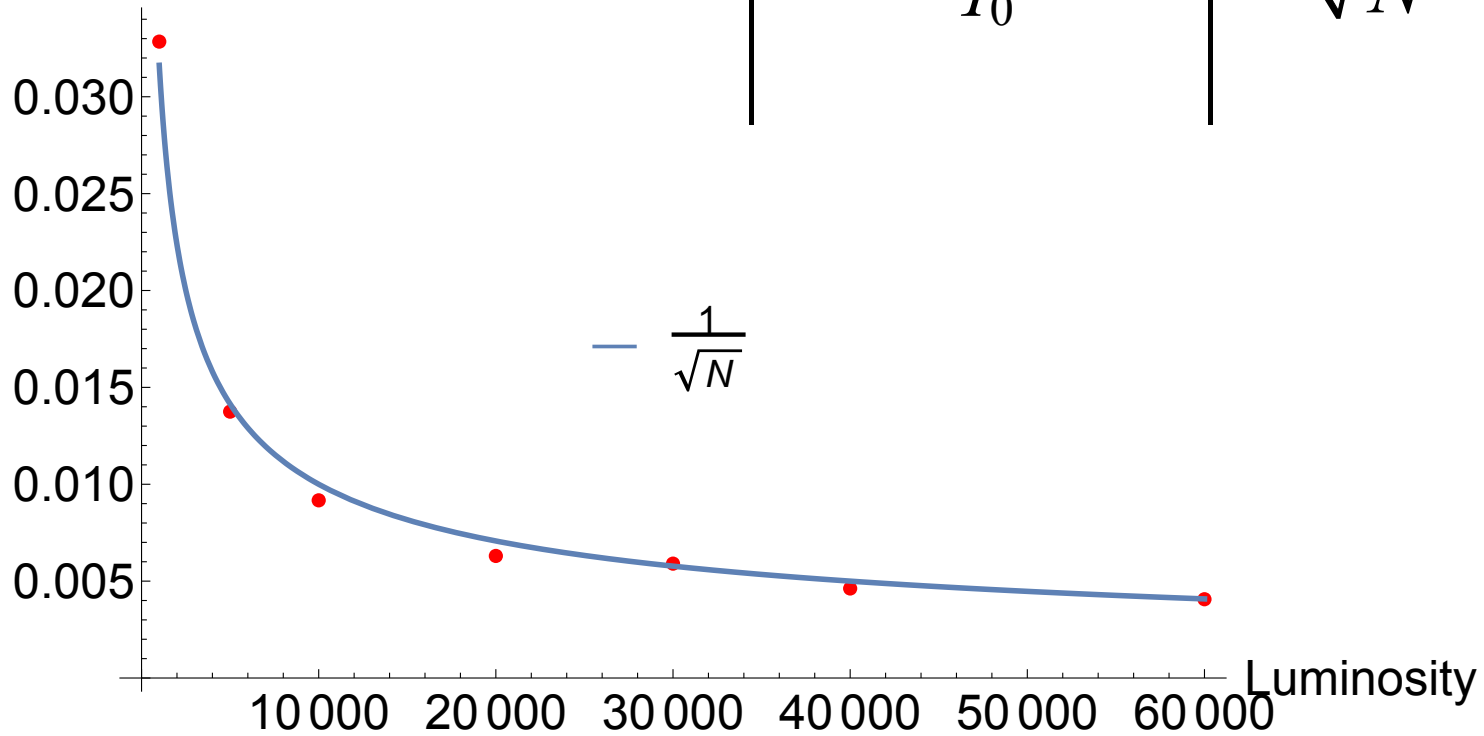
$$\langle \hat{\mu} \rangle = 1$$



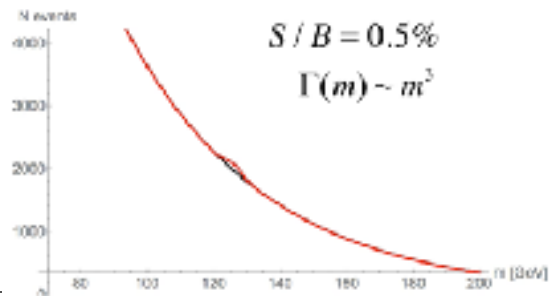
Wald Theorem

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

Relative Error



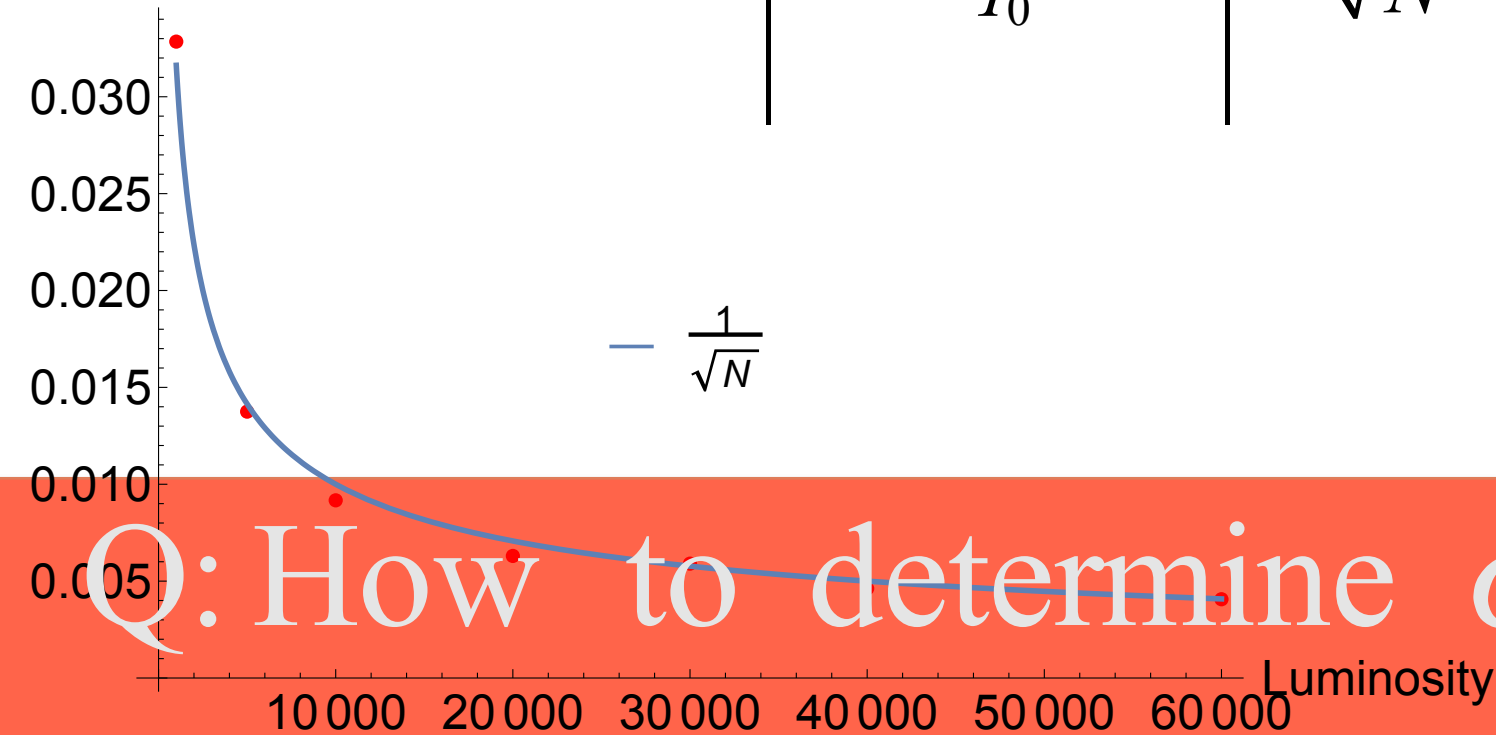
Luminosity is the number of events in the histogram



Wald Theorem

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

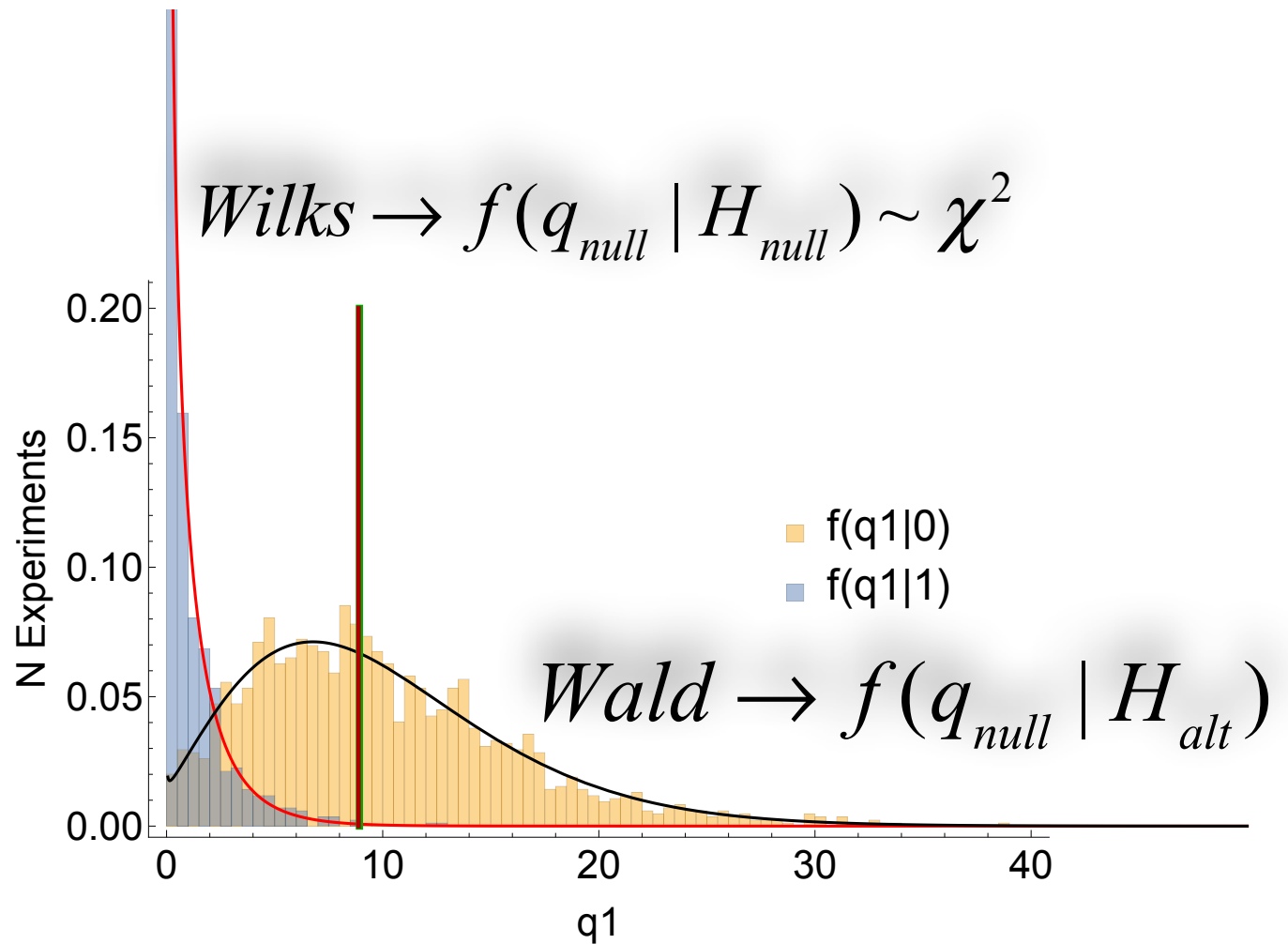
Relative Error



Luminosity is the number of events in the histogram

A: With the Asimov DATA

Asymptotics



The Feldman Cousins Unified Method



The Flip Flop Way of an Experiment

- The most intuitive way to analyze the results of an experiment would be

If the significance based on q_{obs} , is less than 3 sigma, derive an upper limit (just looking at tables), if the result is >5 sigma derive a discovery central confidence interval for the measured parameter (cross section, mass....)

- This Flip Flopping policy leads to undercoverage:
Is that really a problem for Physicists?
Some physicists say, for each experiment quote always two results, an upper limit, and a (central?) discovery confidence interval
- Many LHC analyses report both ways.

Frequentist Paradise – F&C Unified with Full Coverage

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

• The motivation:

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

• Let the test statistics be

$$q = \begin{cases} -2 \ln \frac{L(s+b)}{L(\hat{s}+b)} & \hat{s} \geq 0 \\ -2 \ln \frac{L(s+b)}{L(b)} & \hat{s} < 0 \end{cases}$$

where \hat{s} is the

physically allowed mean s that maximizes $L(\hat{s}+b)$

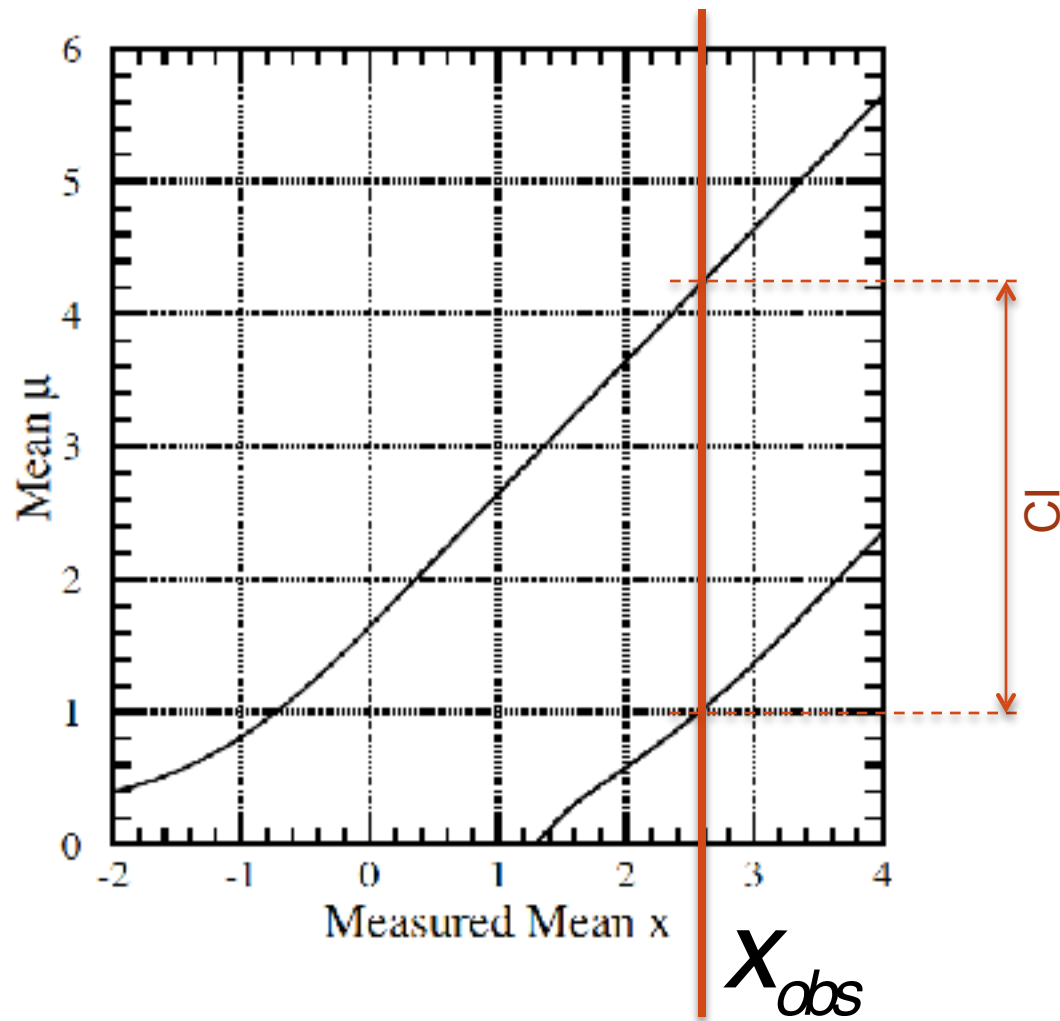
(protect a downward fluctuation of the background, $n_{\text{obs}} > b$; $\hat{s} > 0$)

- Order by taking the 68% highest q 's



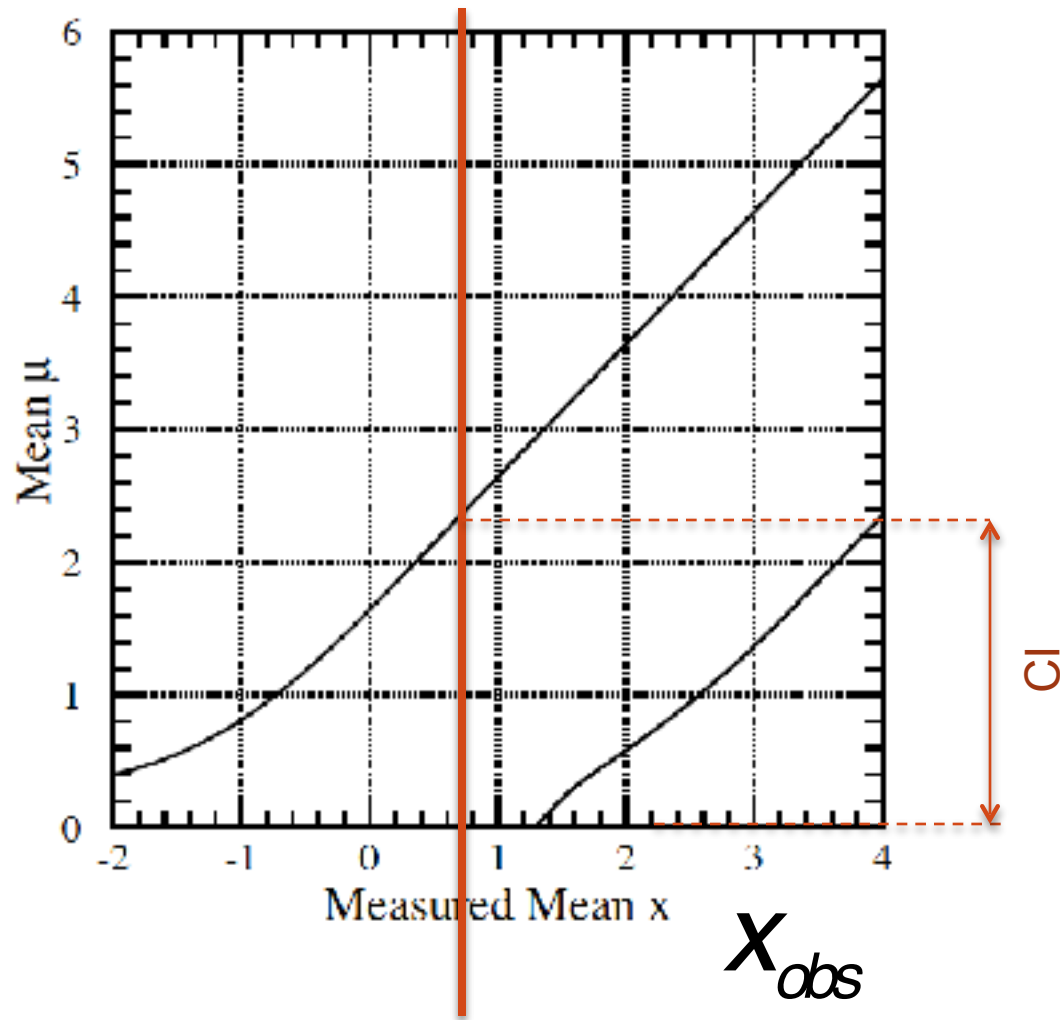
How to tell an Upper limit from a Measurement without Flip Flopping

- 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

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

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

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

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

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

CI of μ at the $(1-\alpha)$ CL = $\{\mu | p_\mu \geq \alpha\}$

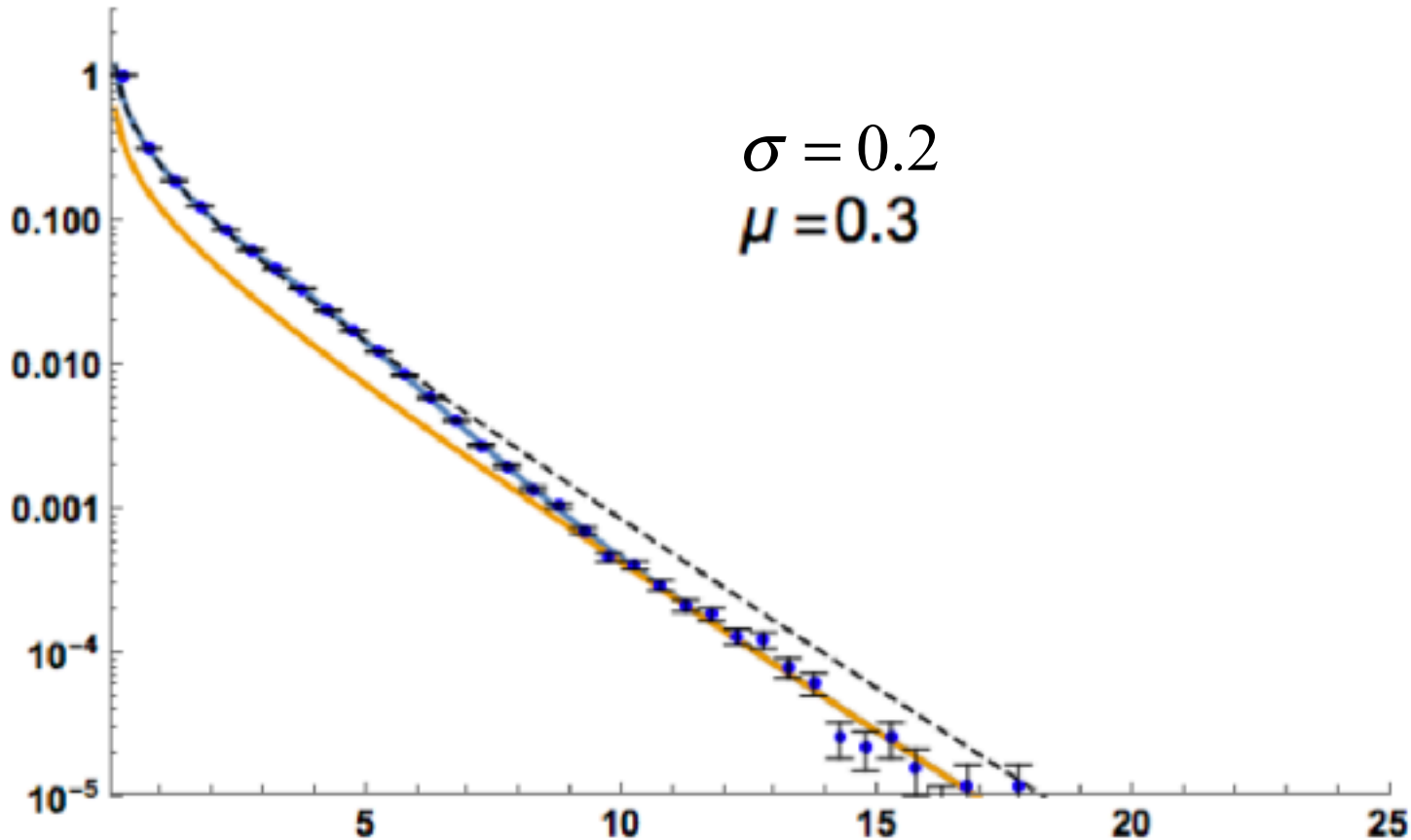
CI of μ at the 95% CL = $\{\mu | p_\mu \geq 5\%\}$



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

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

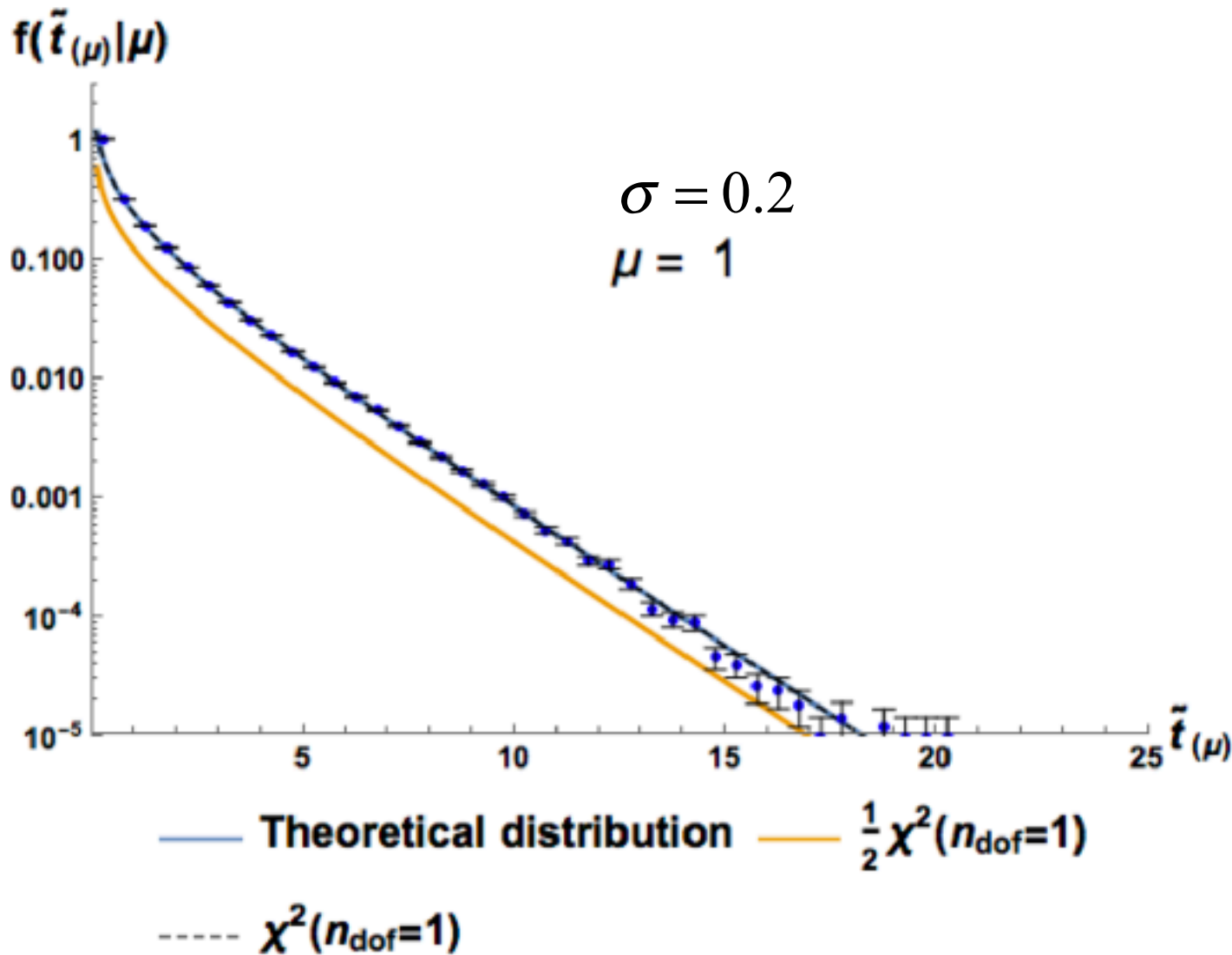
$f(\tilde{t}_{(\mu)} | \mu)$



— Theoretical distribution — $\frac{1}{2}\chi^2(n_{\text{dof}}=1)$

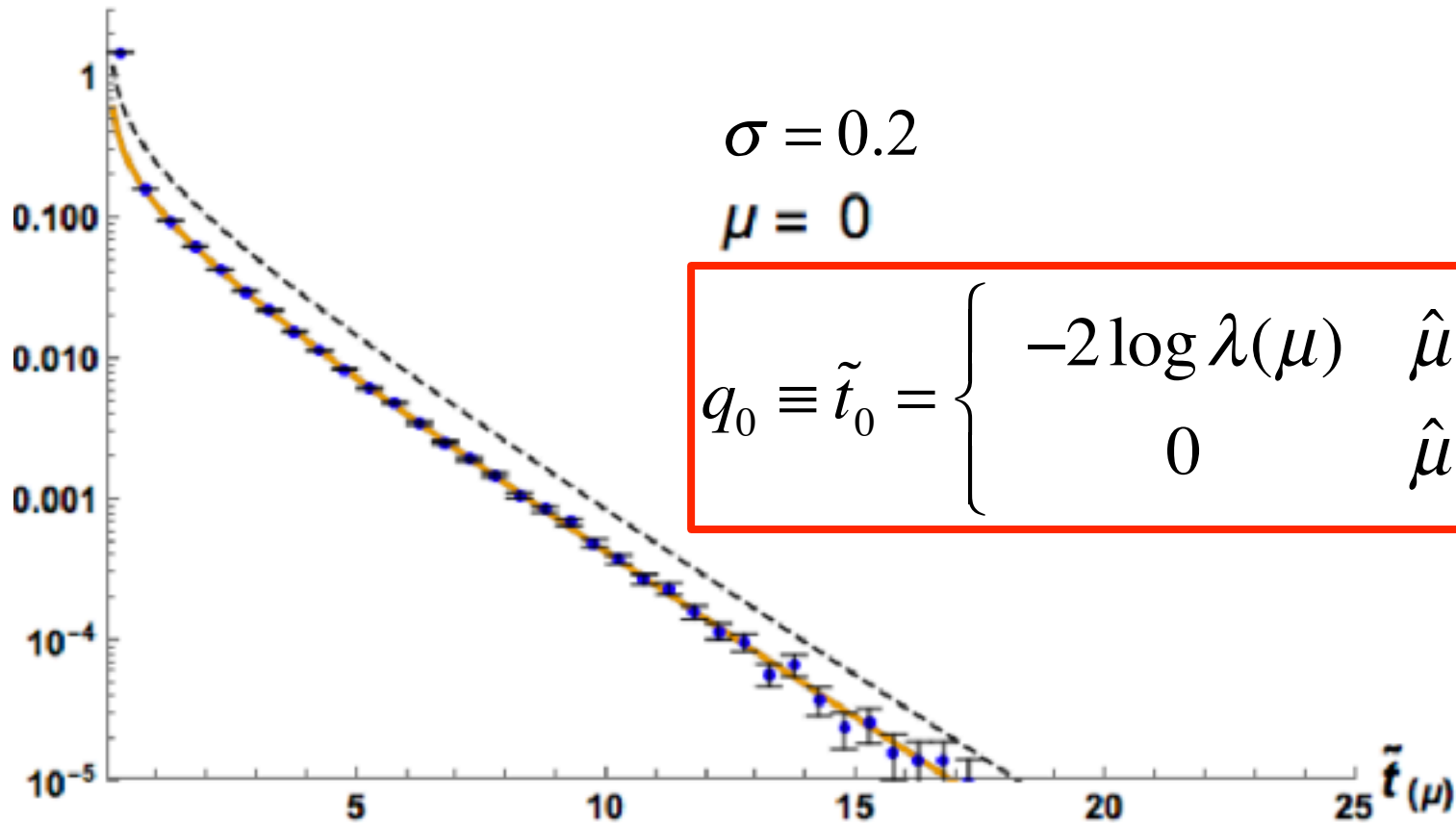
--- $\chi^2(n_{\text{dof}}=1)$

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



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

$f(\tilde{t}_\mu | \mu)$



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

— Theoretical distribution — $\frac{1}{2} \chi^2(n_{\text{dof}}=1)$

----- $\chi^2(n_{\text{dof}}=1)$

FC confidence belt

Given $\mu_{true} = 0.2$

derive the $\hat{\mu}$ interval

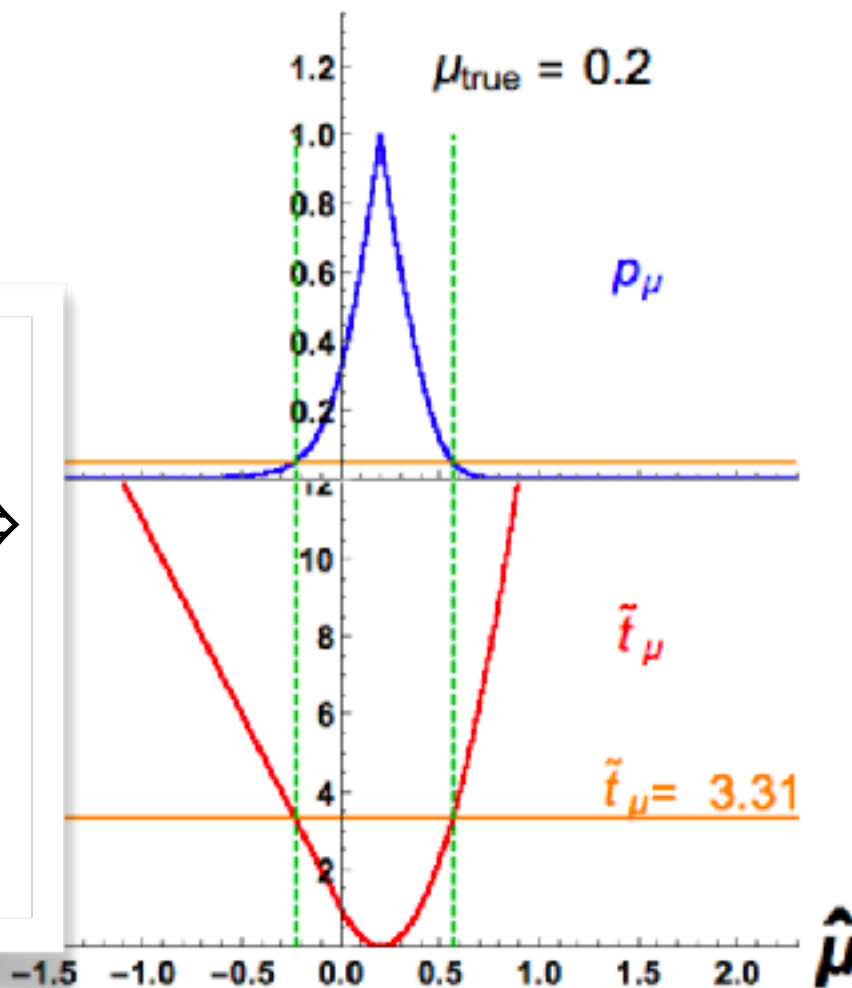
for which $p_{\mu} > 0.05$

set μ_{true} e.g. $\mu_{true} = 0.2$

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

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

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



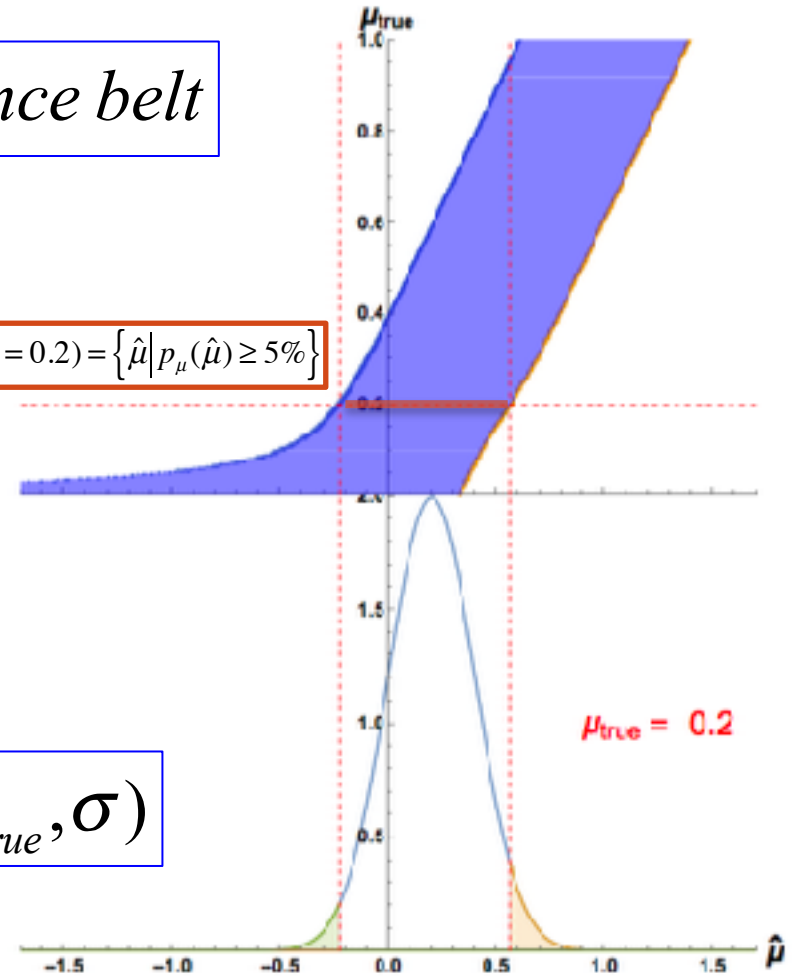
FC confidence belt

Scan μ_{true} and build the Confidence belt

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

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

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

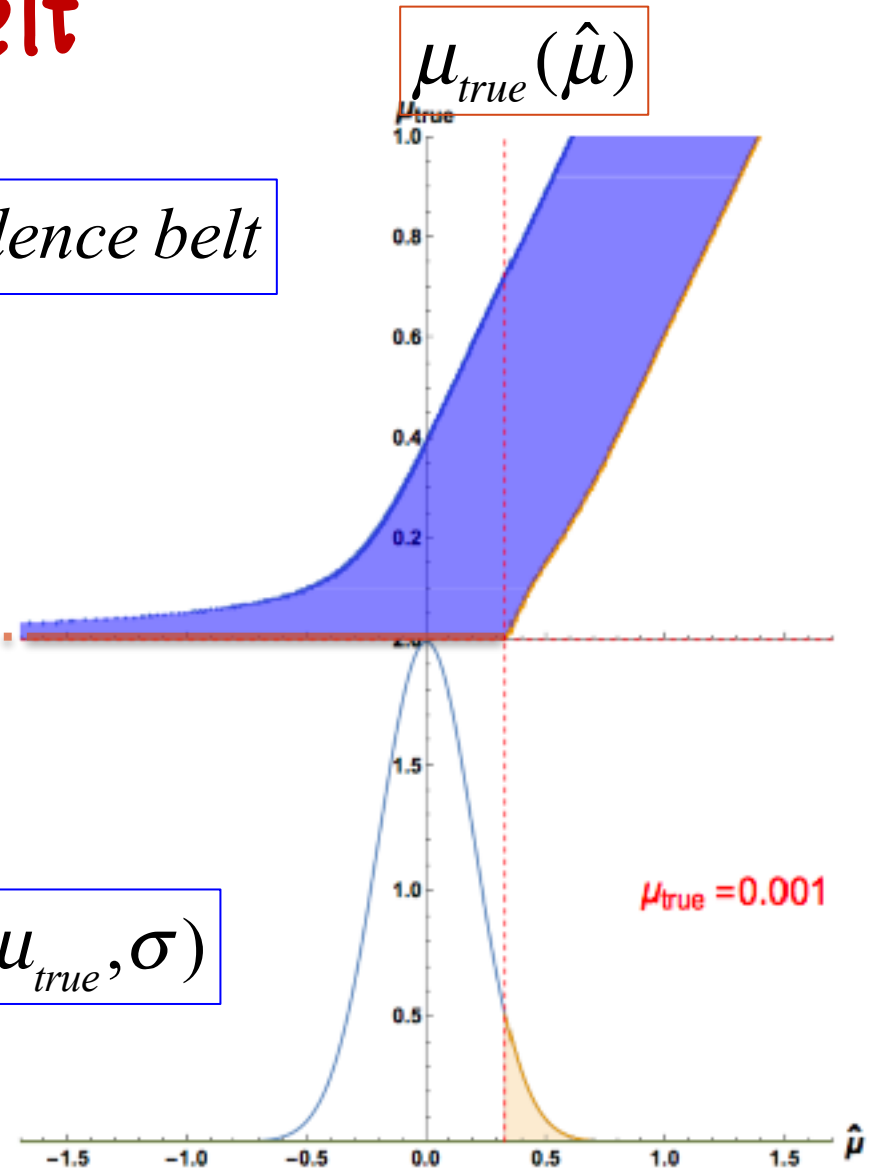


FC confidence belt

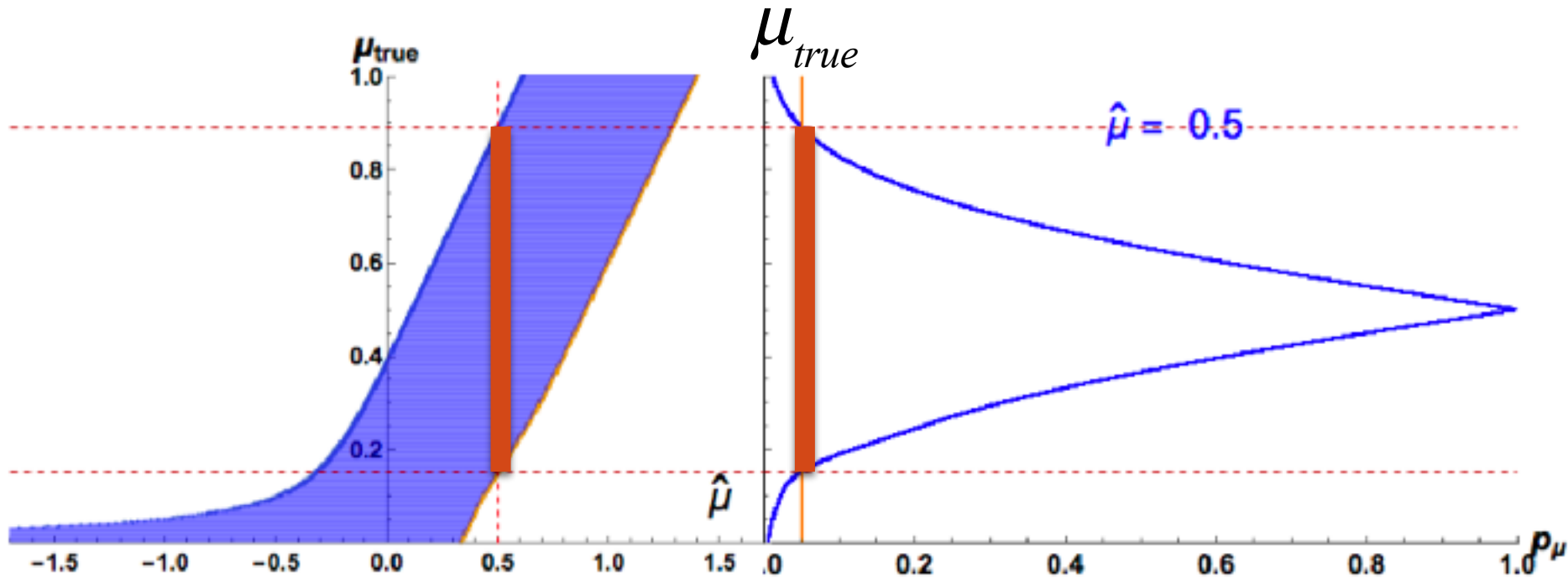
Scan μ_{true} and build the Confidence belt

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

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



FC confidence belt



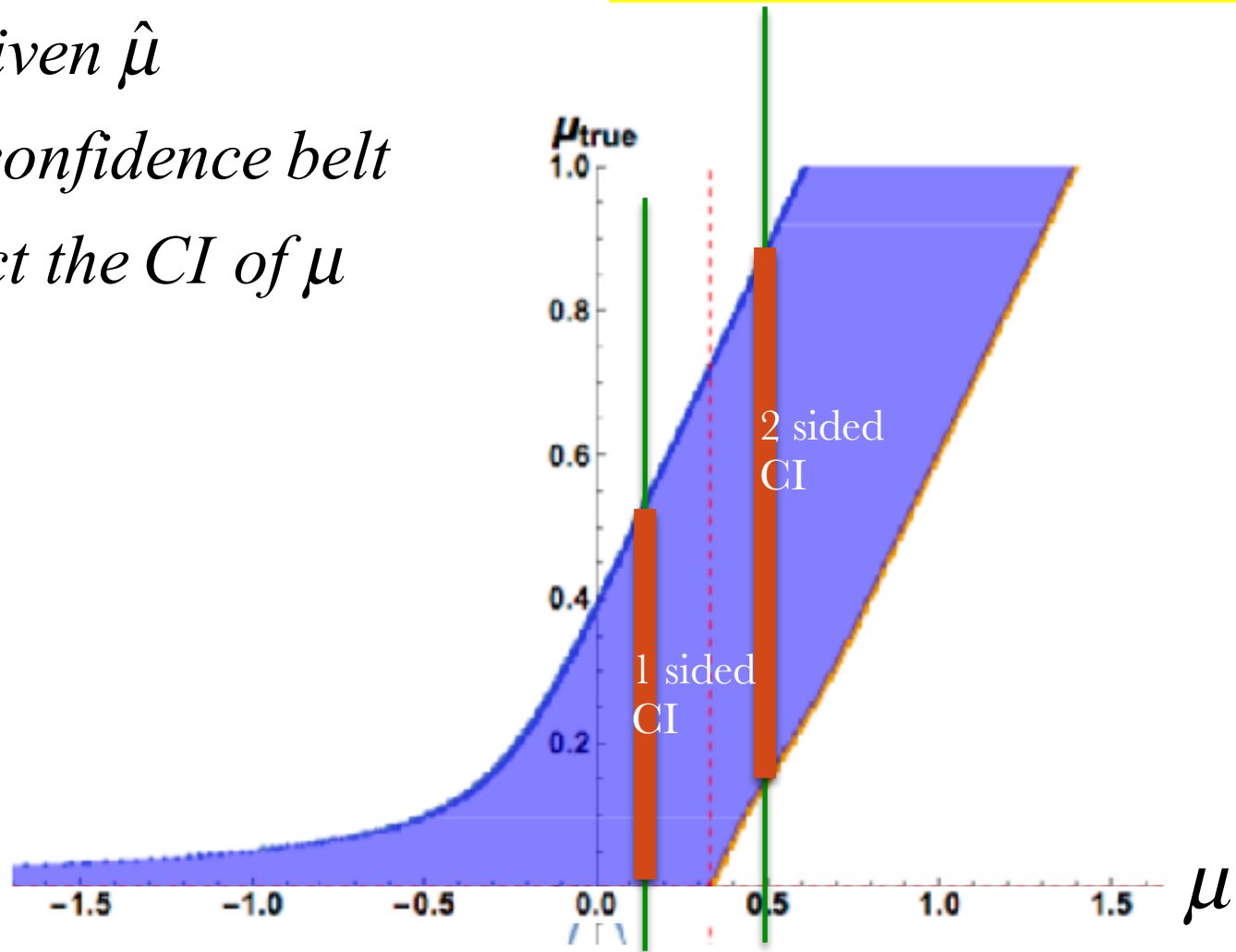
*The values of μ_{true} for which $p_\mu = 0.05$
for a given $\hat{\mu}$, are the CI of μ*



FC confidence belt

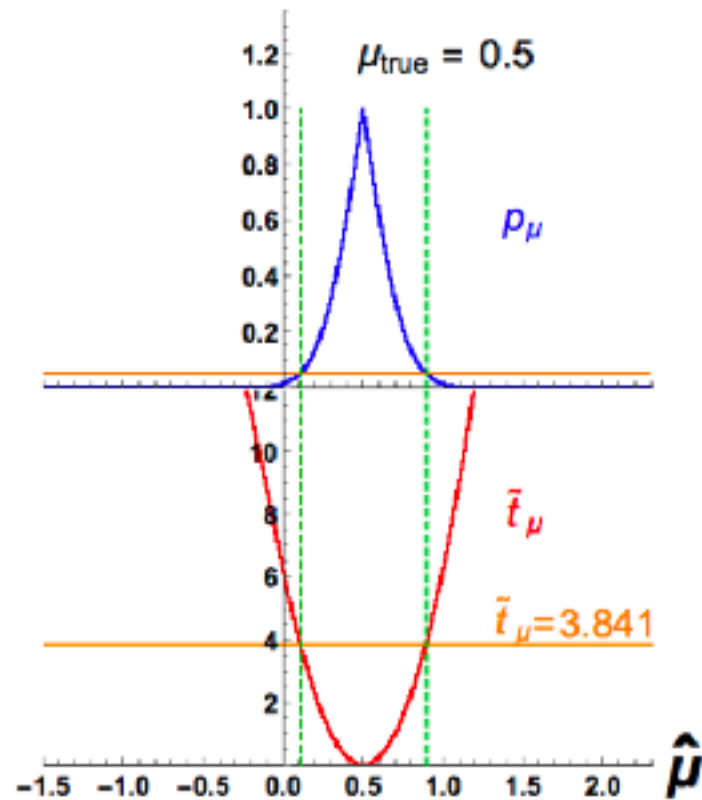
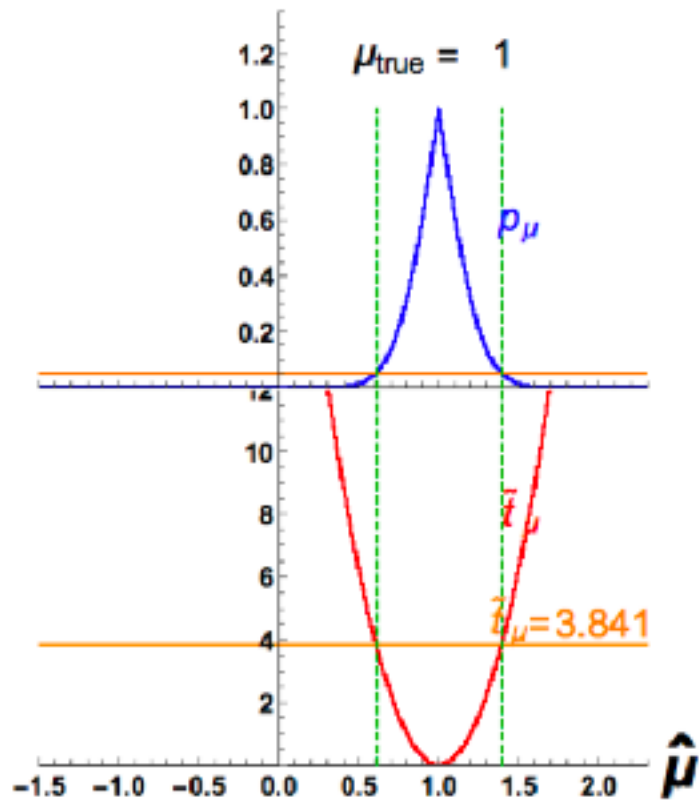
Depends on the observation
one might get 1-sided or 2-sided CI

*For a given $\hat{\mu}$
use the confidence belt
to extract the CI of μ*



FC confidence belt - A Shortcut

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



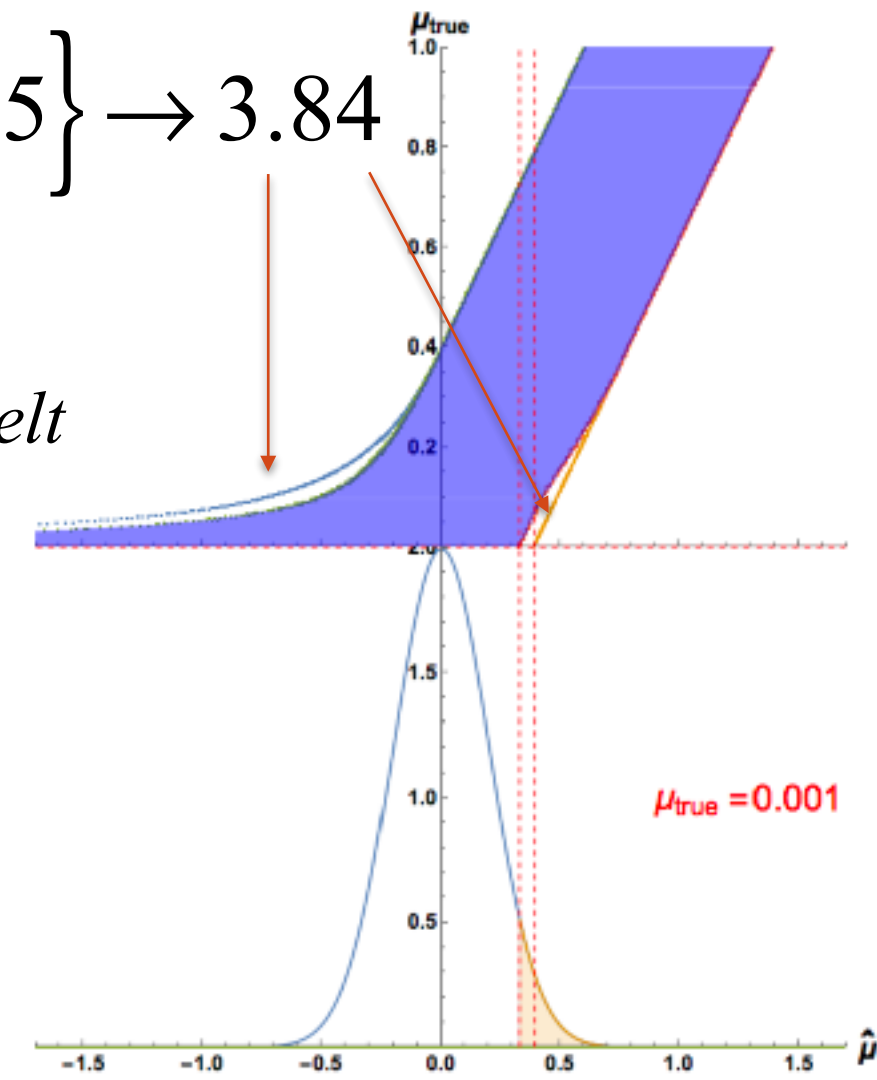
FC confidence belt - A Shortcut

$$\mu_{true} \gg 0 \Rightarrow \left\{ \tilde{t}_{\mu} \mid P_{\mu} > 0.95 \right\} \rightarrow 3.84$$

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

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

(Conservative)



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

q_0 for discovery

CCGV

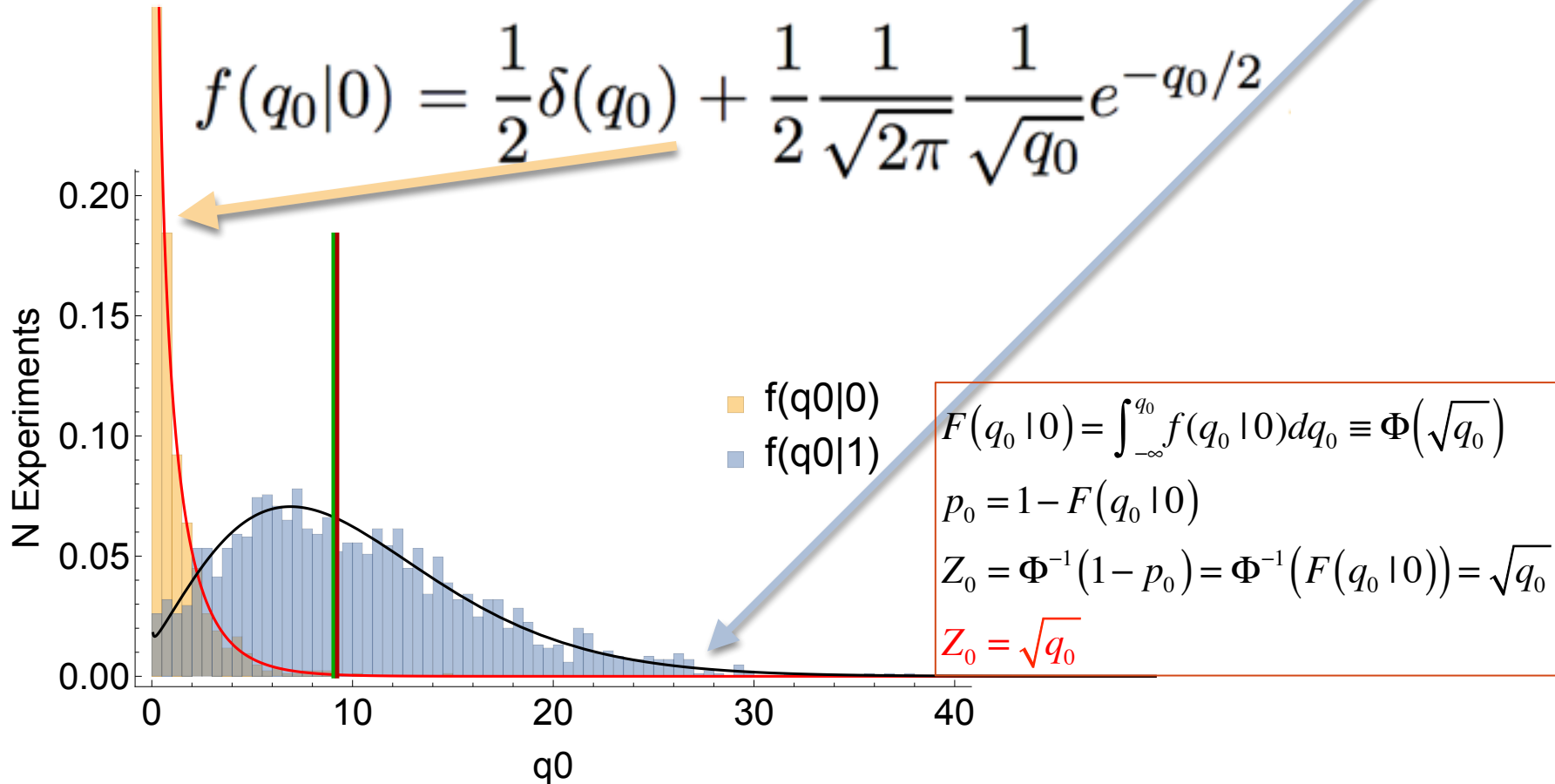
$$q_0 = \begin{cases} -2 \log \lambda(\mu) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}$$

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

PDF of $(q_0|0)$ and $(q_0|1)$

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

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

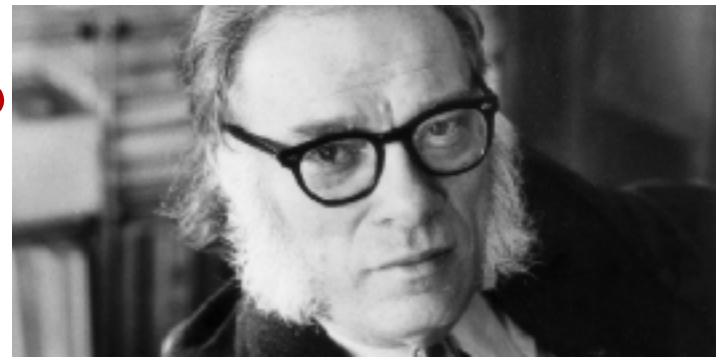
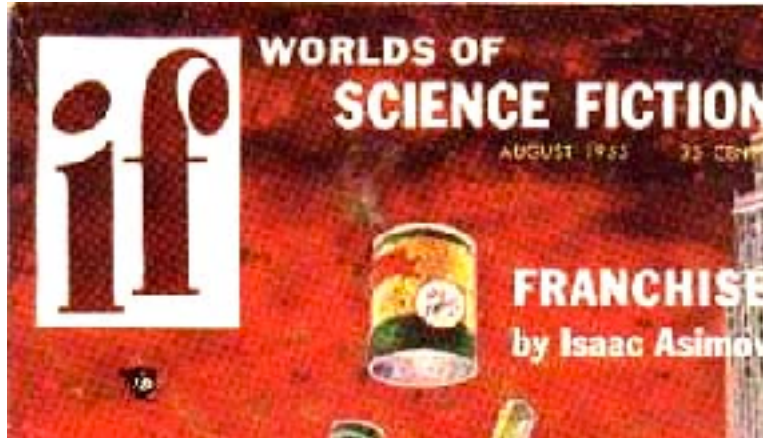


Estimating the Sensitivity of an Experiment

- 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 , $f(q_{\text{null}}|BG)$.
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

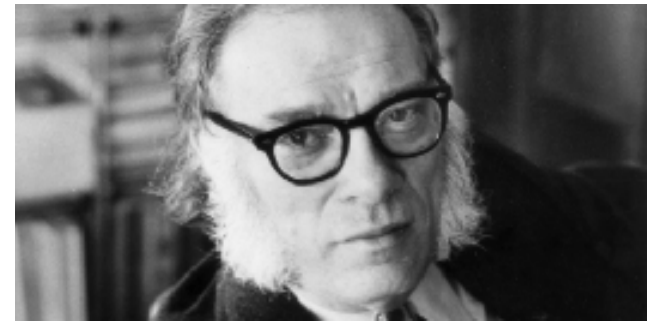


The Asimov Data S



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.

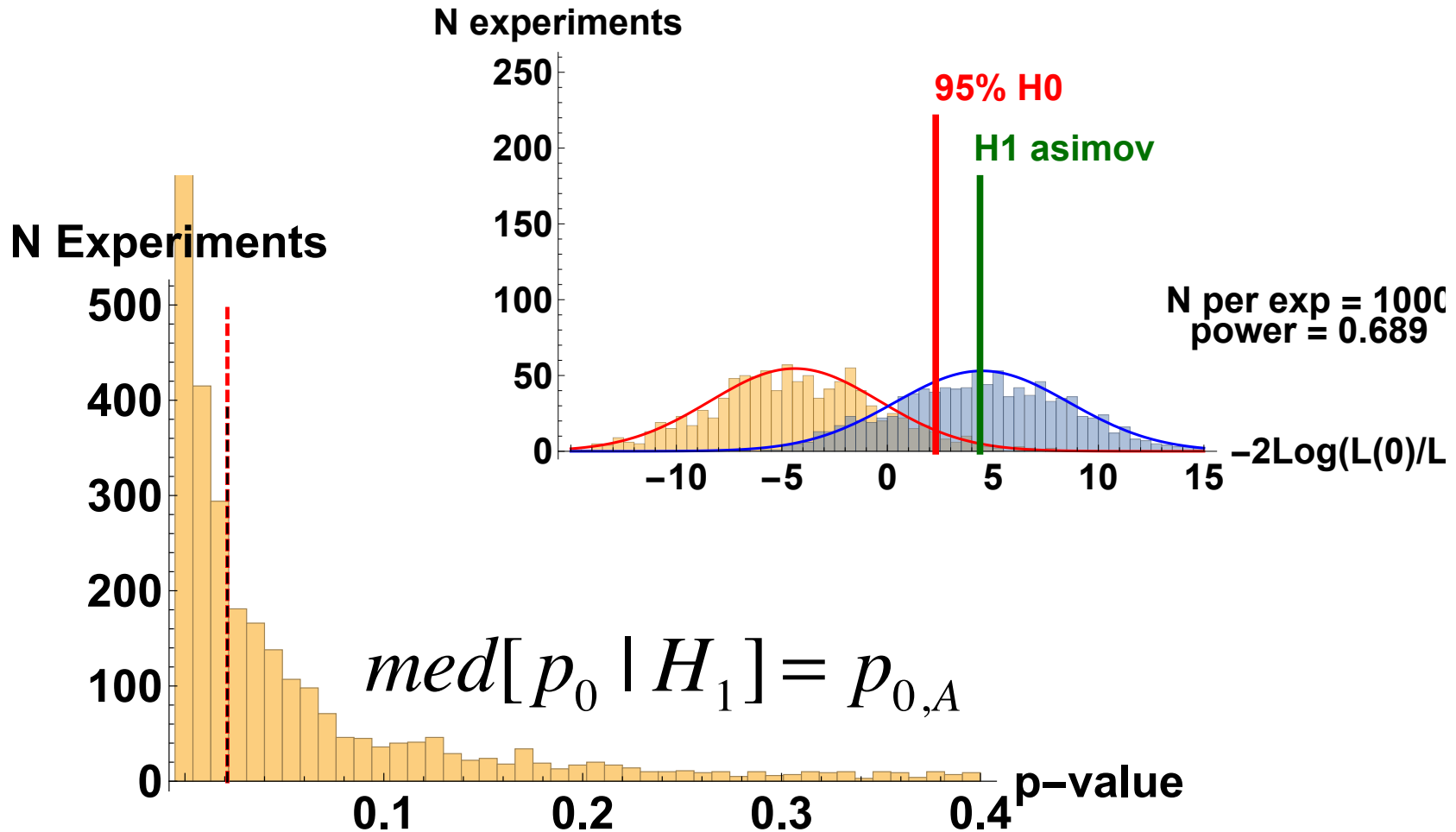
This "Asimov" data set delivers the desired median sensitivity. Hence, one is exempted from the need to perform an ensemble of experiments for each set of parameters.

- 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

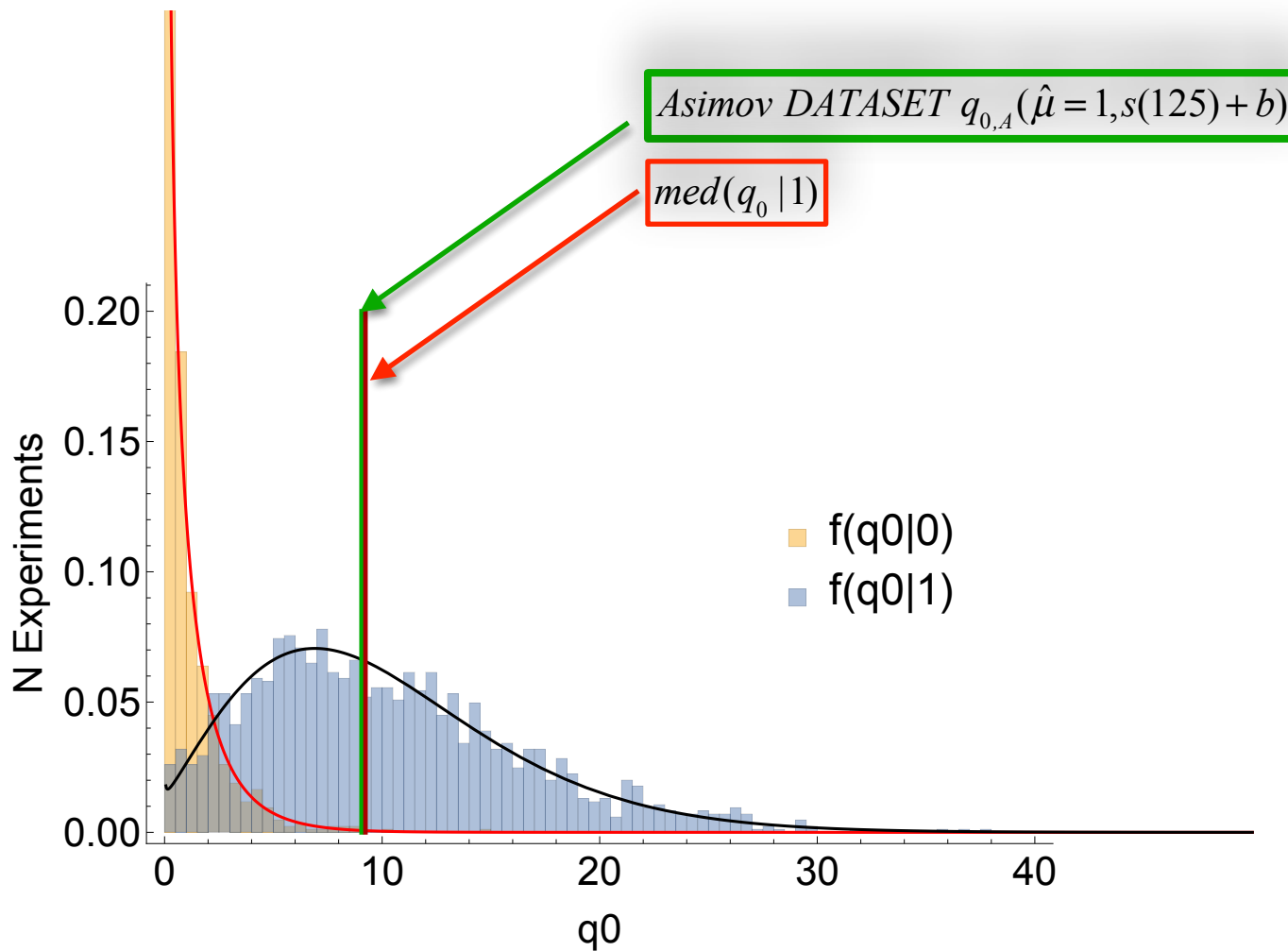
Distribution of p_0 -value under H_1



$$f(p_0 | H_1) = f([\text{prob}(q \geq q_{obs}) | H_0] | H_1)$$



The Magic of Asimov



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

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

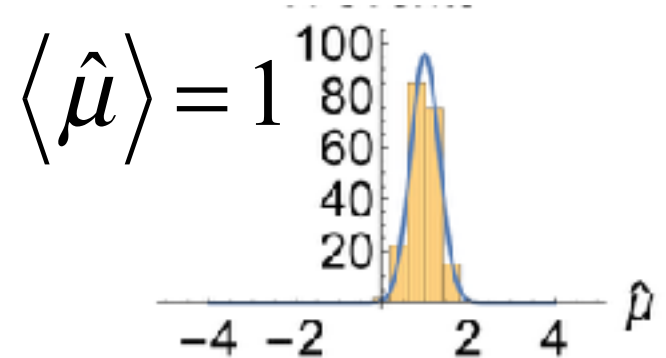
For the Asimov $\hat{\mu} = \mu_{true}$

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

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

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

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



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

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



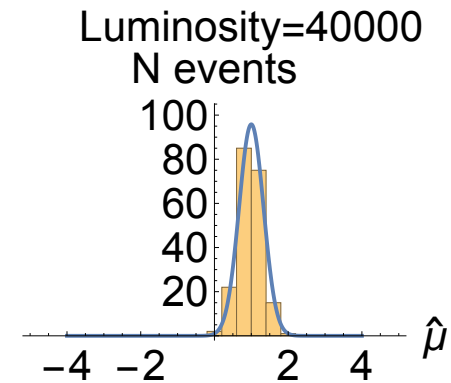
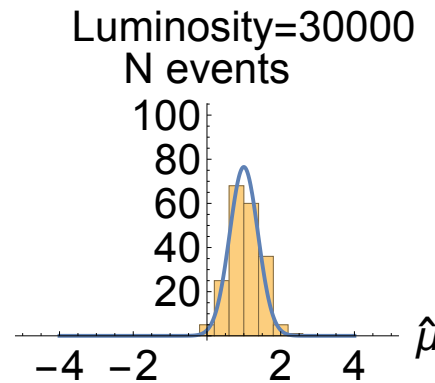
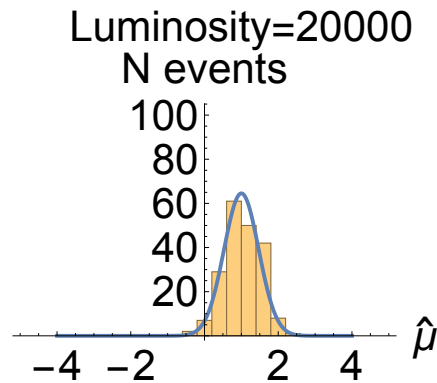
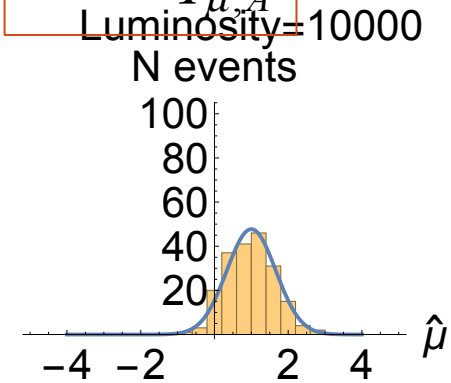
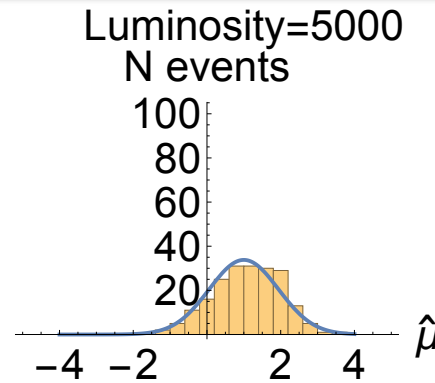
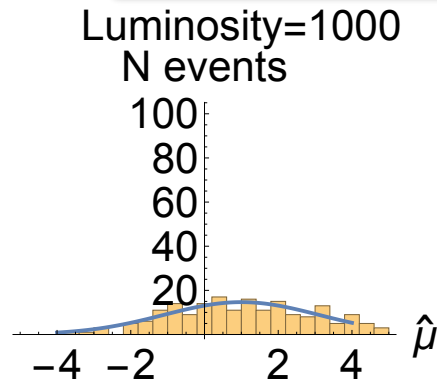
Back to Wald, what is

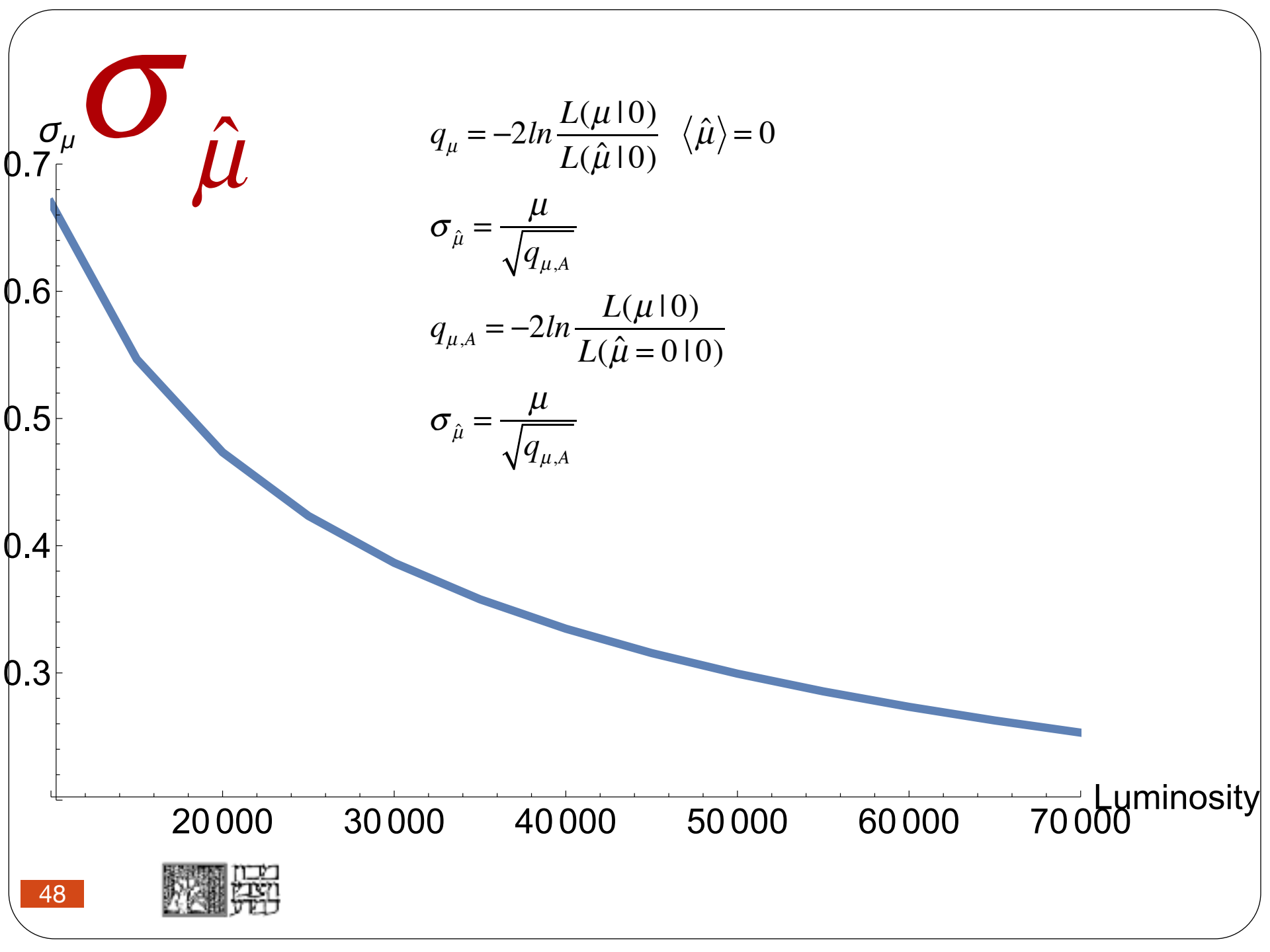
$\sigma_{\hat{\mu}}$?

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

Fit the distribution with the Asimov calculated Sigma

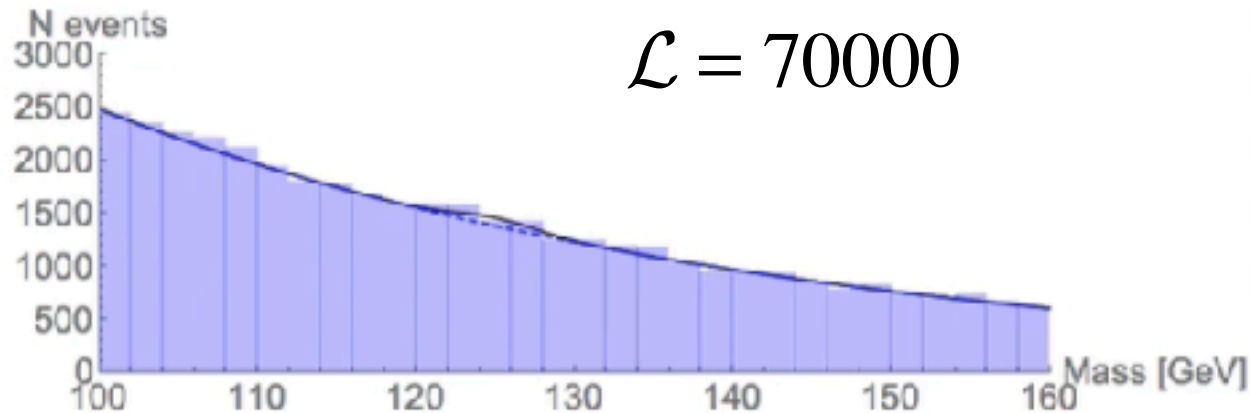
$$\sigma_{\hat{\mu}}^2 = \frac{\mu^2}{q_{\mu,A}}$$



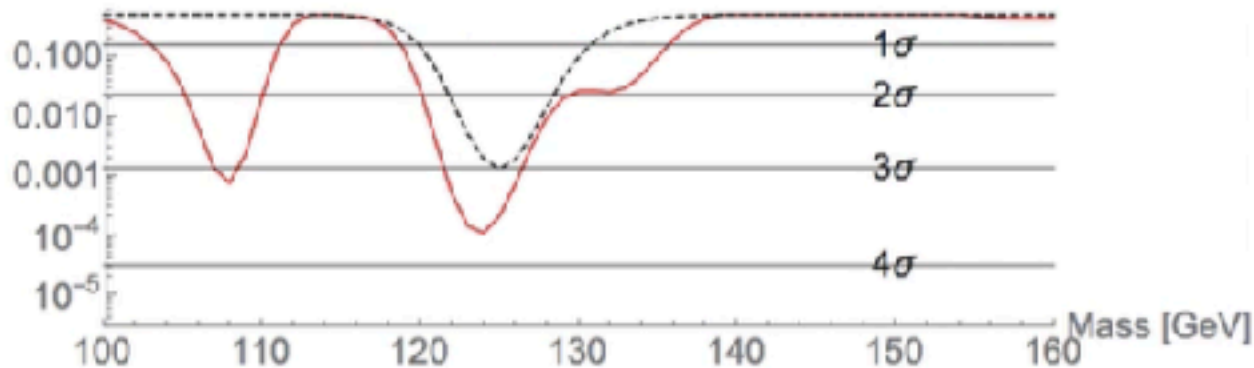


p -value

$$\mathcal{L} = 70000$$

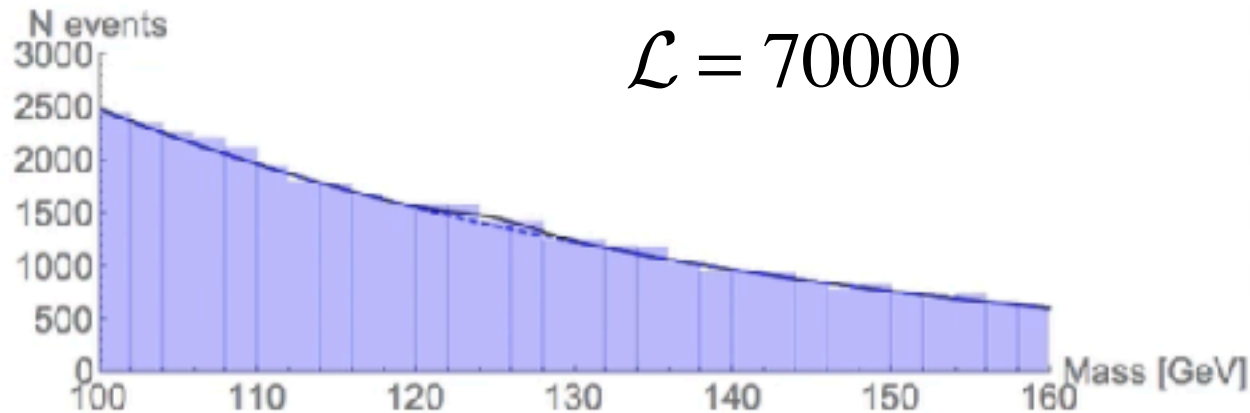


$$p = \text{prob}(q_0 \geq q_{0,obs})$$

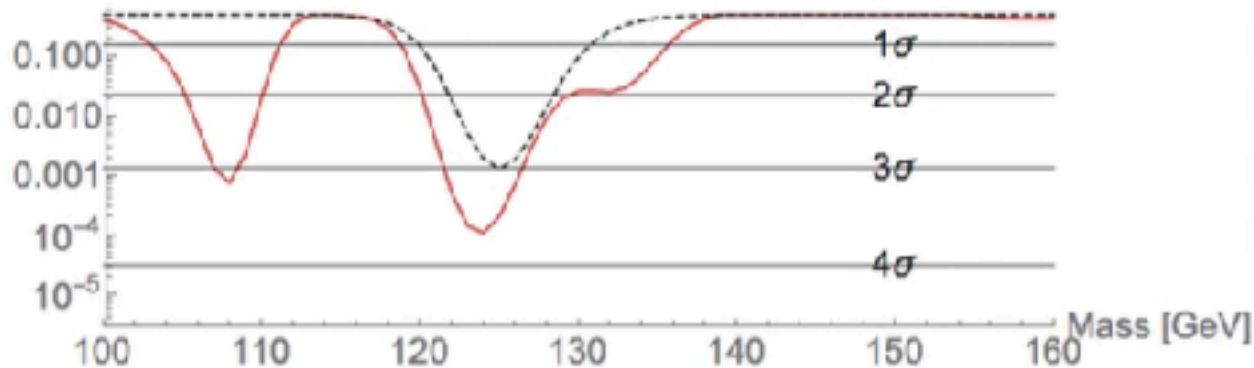


p -value

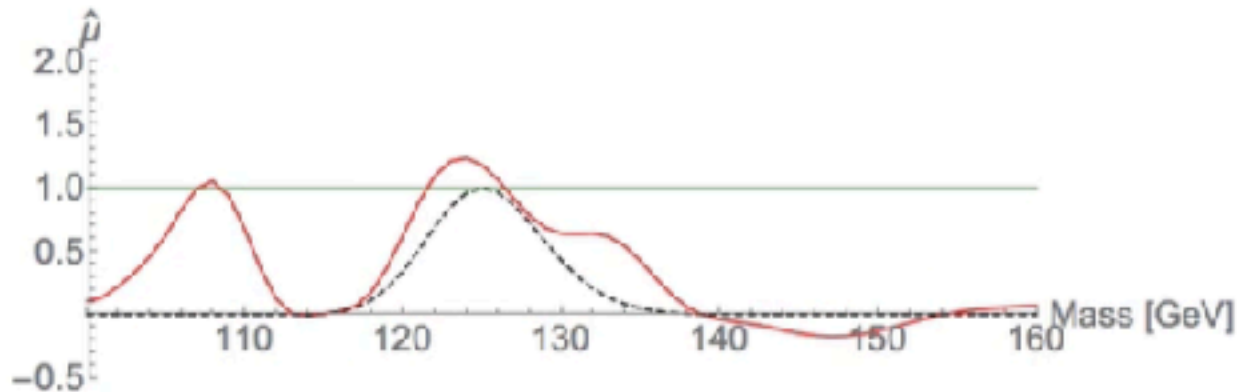
$\mathcal{L} = 70000$



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

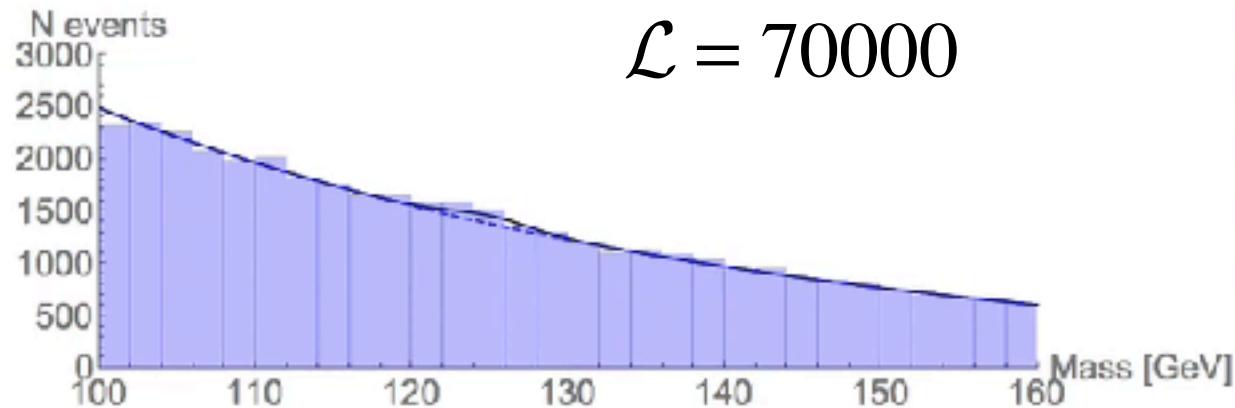


$$\hat{\mu} = \frac{\sigma_{obs}}{\sigma_{exp}}$$

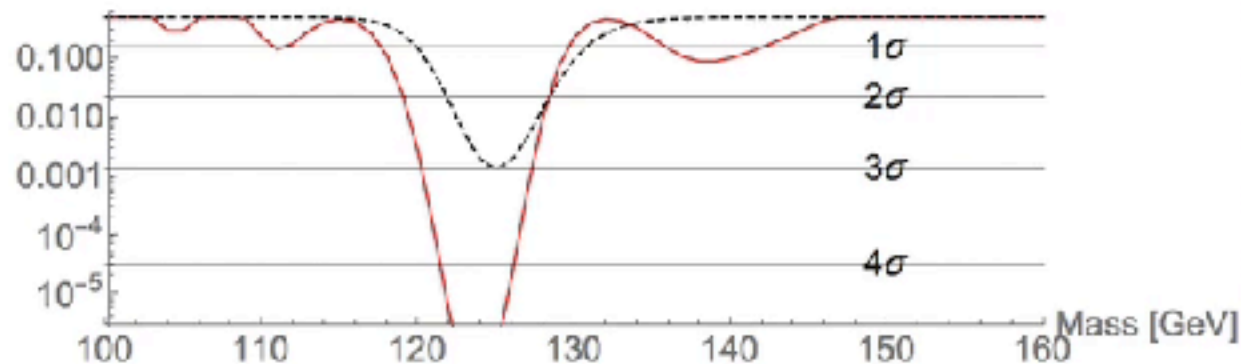


p -value

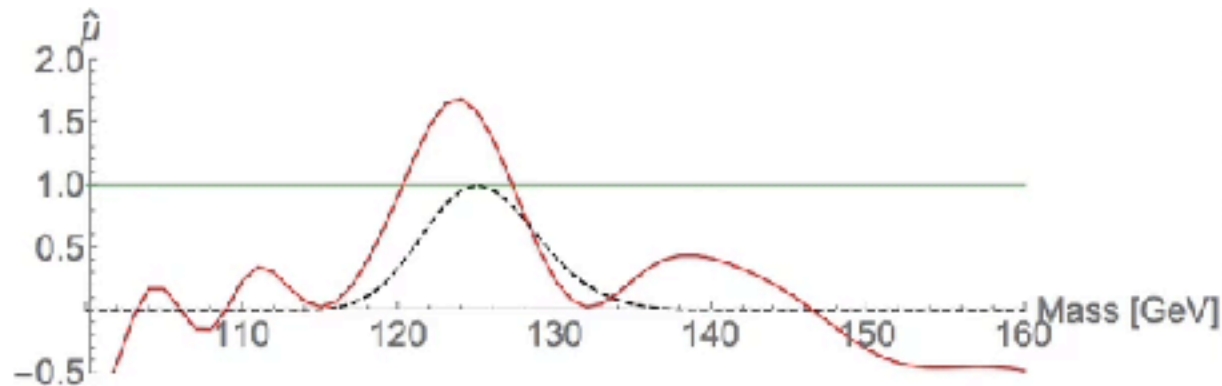
$\mathcal{L} = 70000$



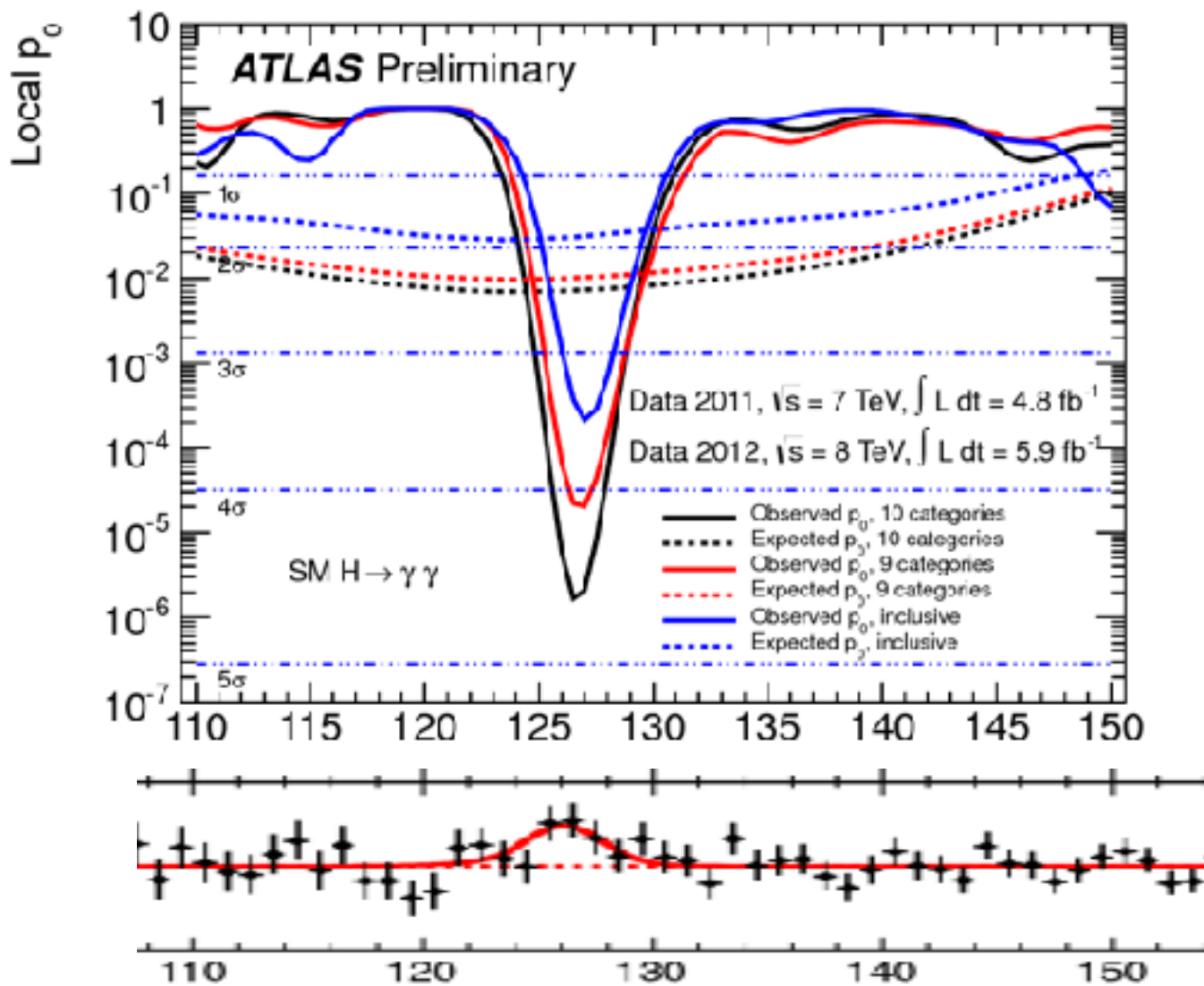
$$p = \text{prob}(q_0 \geq q_{0,obs})$$



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



$H \rightarrow \gamma\gamma$



q_μ for exclusion

CCGV

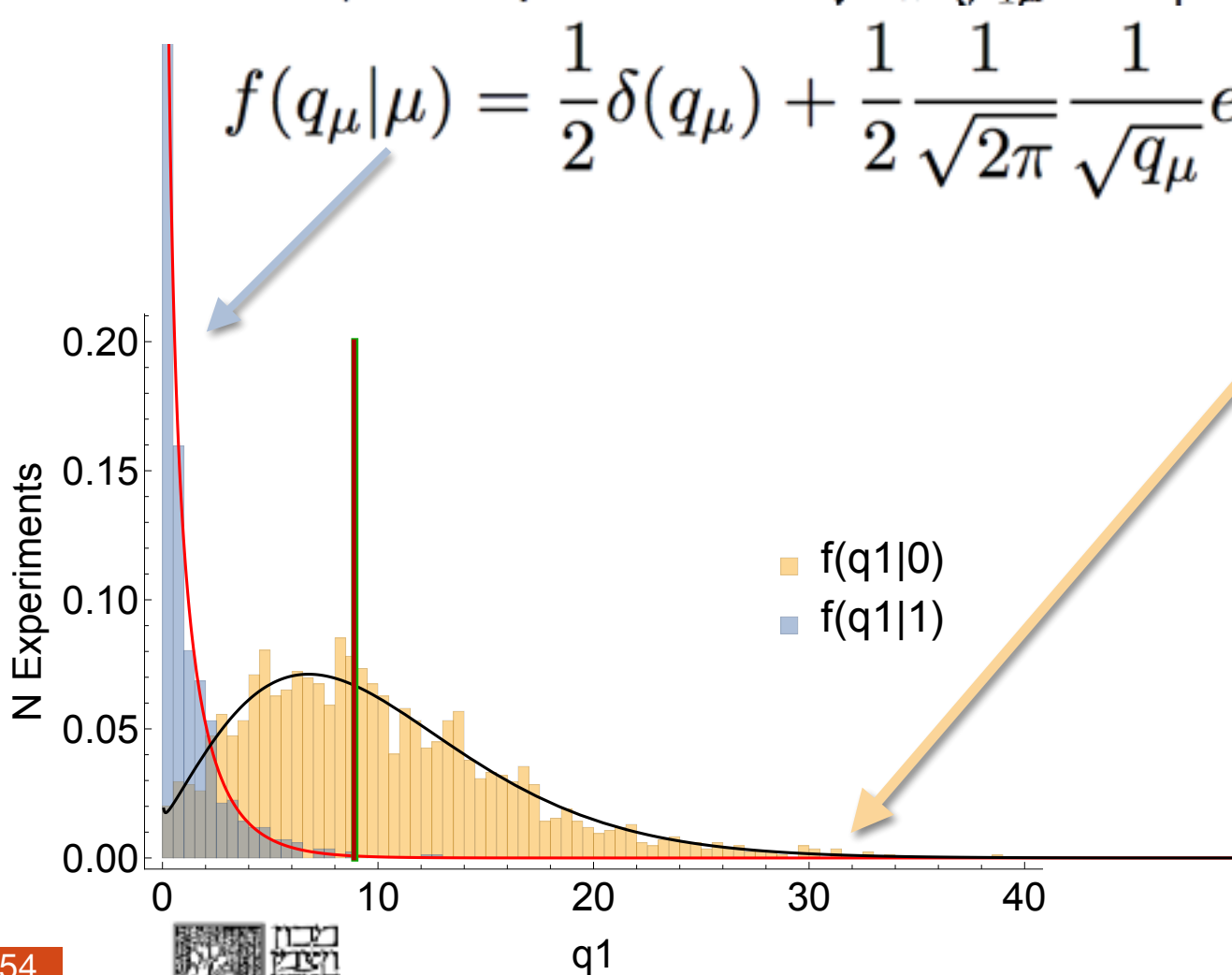
$$q_\mu = \begin{cases} -2 \log \lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

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

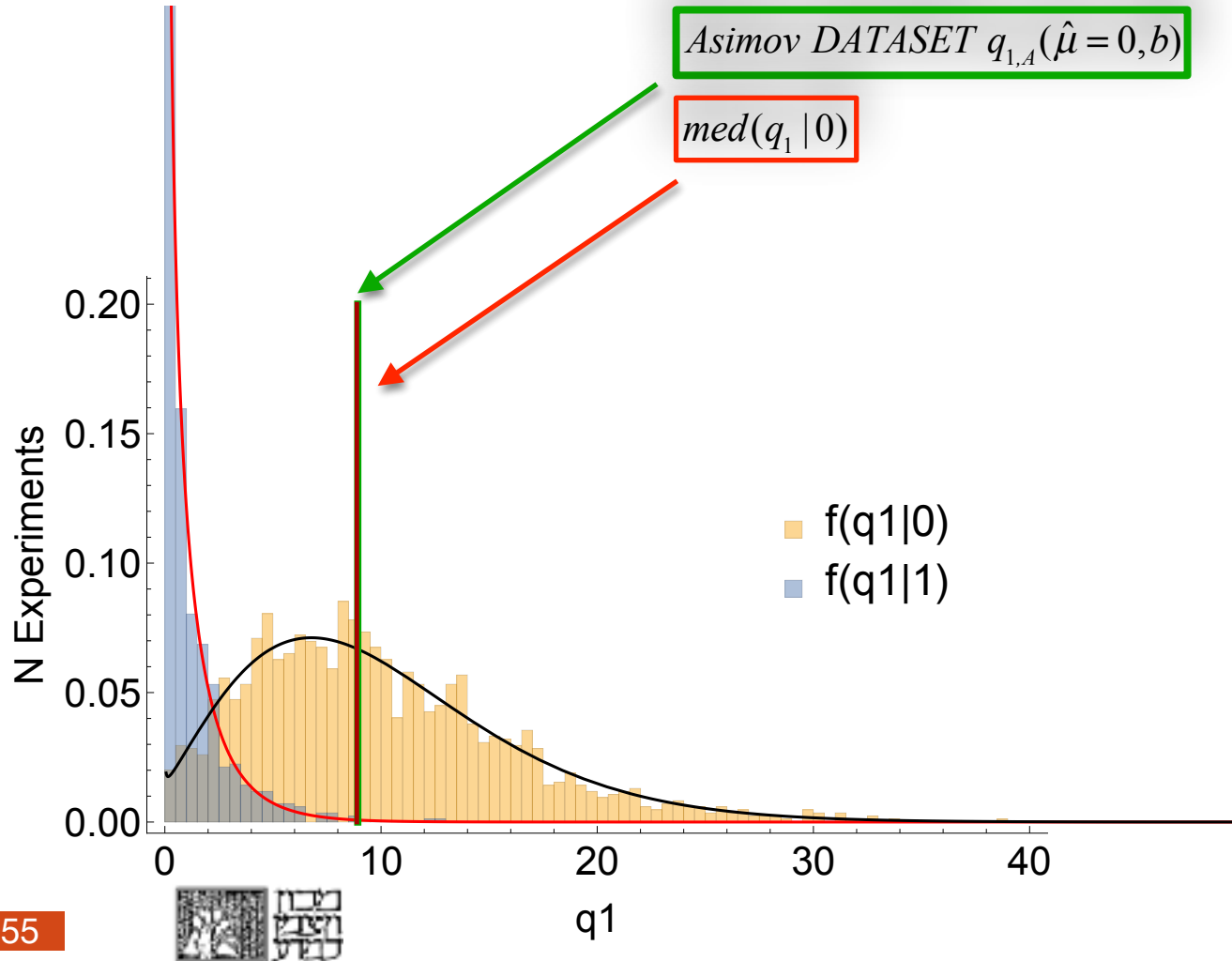
PDF of (q1|1) and (q1|1)

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

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



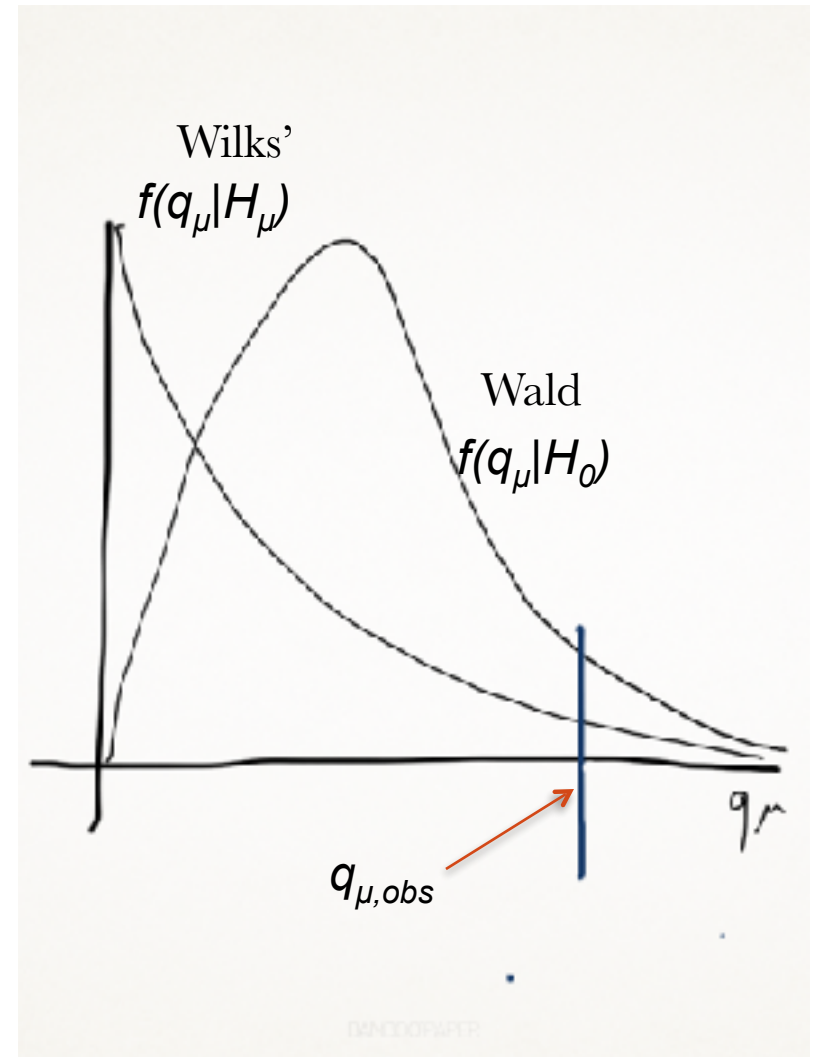
PDF of $(q_1|1)$ and $(q_1|1)$



Exclusion at 95% CL

- We test hypothesis H_μ
- We calculate the PL (profile likelihood) ratio with the one observed data

- $q_{\mu,obs}$

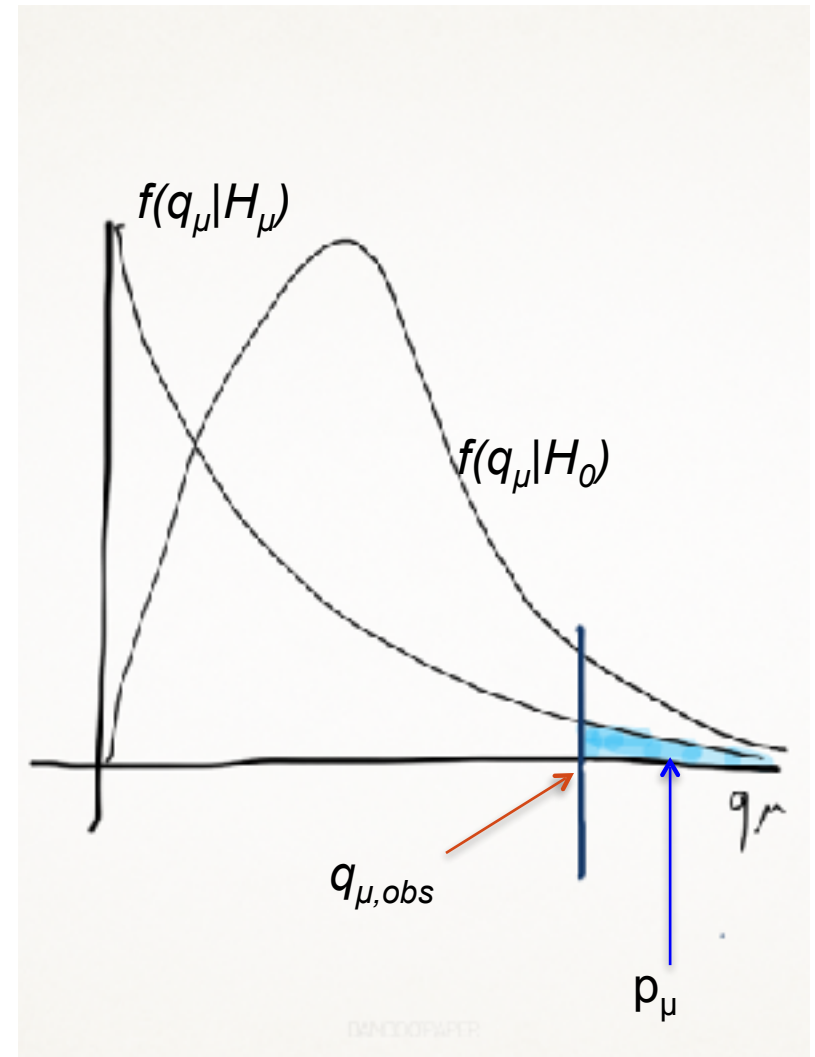


Exclusion at the 95% CL

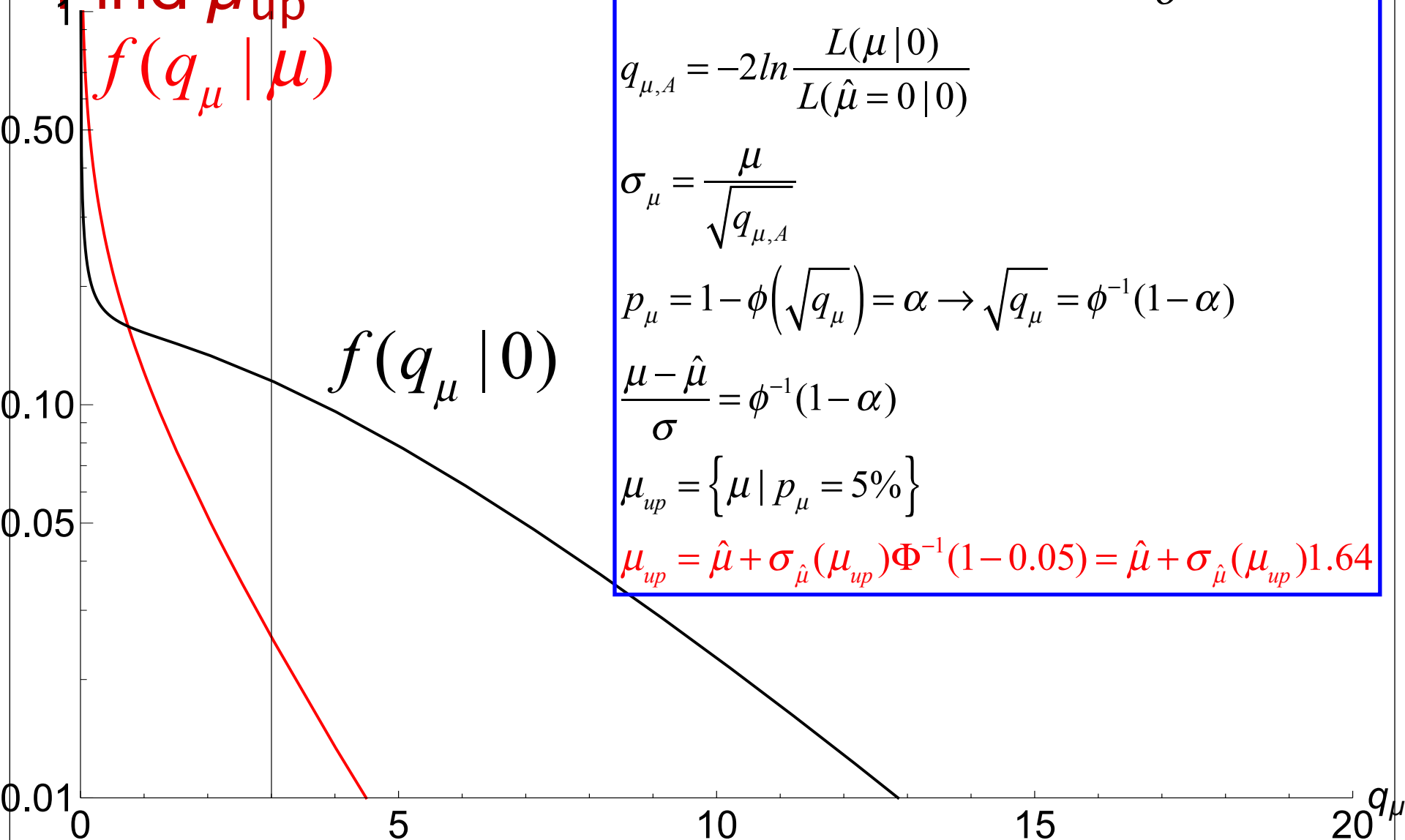
- Find the p-value of the signal hypothesis H_μ

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

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



Find μ_{up}
 $f(q_\mu | \mu)$



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

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

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

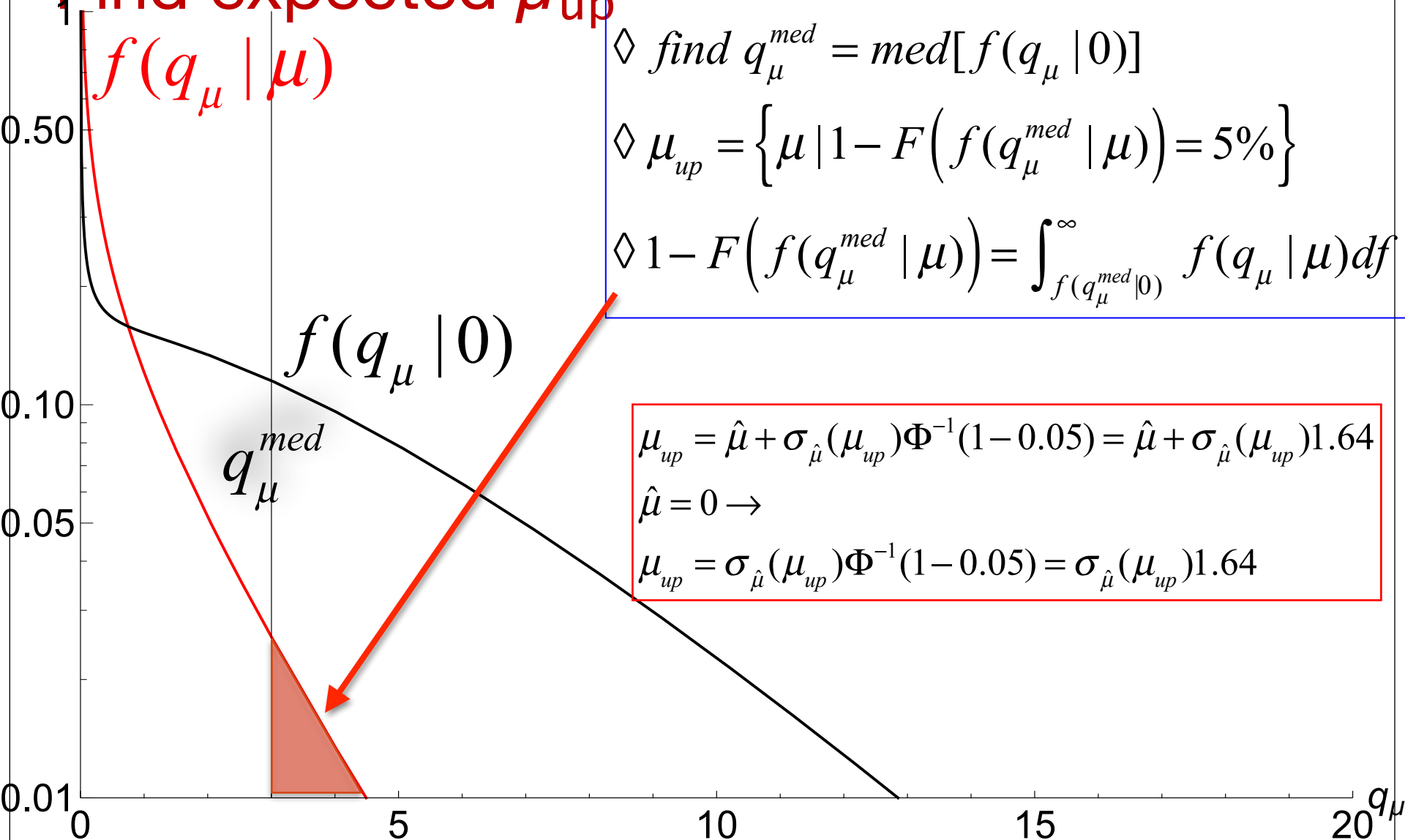
$$p_\mu = 1 - \Phi(\sqrt{q_\mu}) = \alpha \rightarrow \sqrt{q_\mu} = \Phi^{-1}(1 - \alpha)$$

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

$$\mu_{up} = \{ \mu \mid p_\mu = 5\% \}$$

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

Find expected μ_{up}



◇ scan μ

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

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

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

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

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

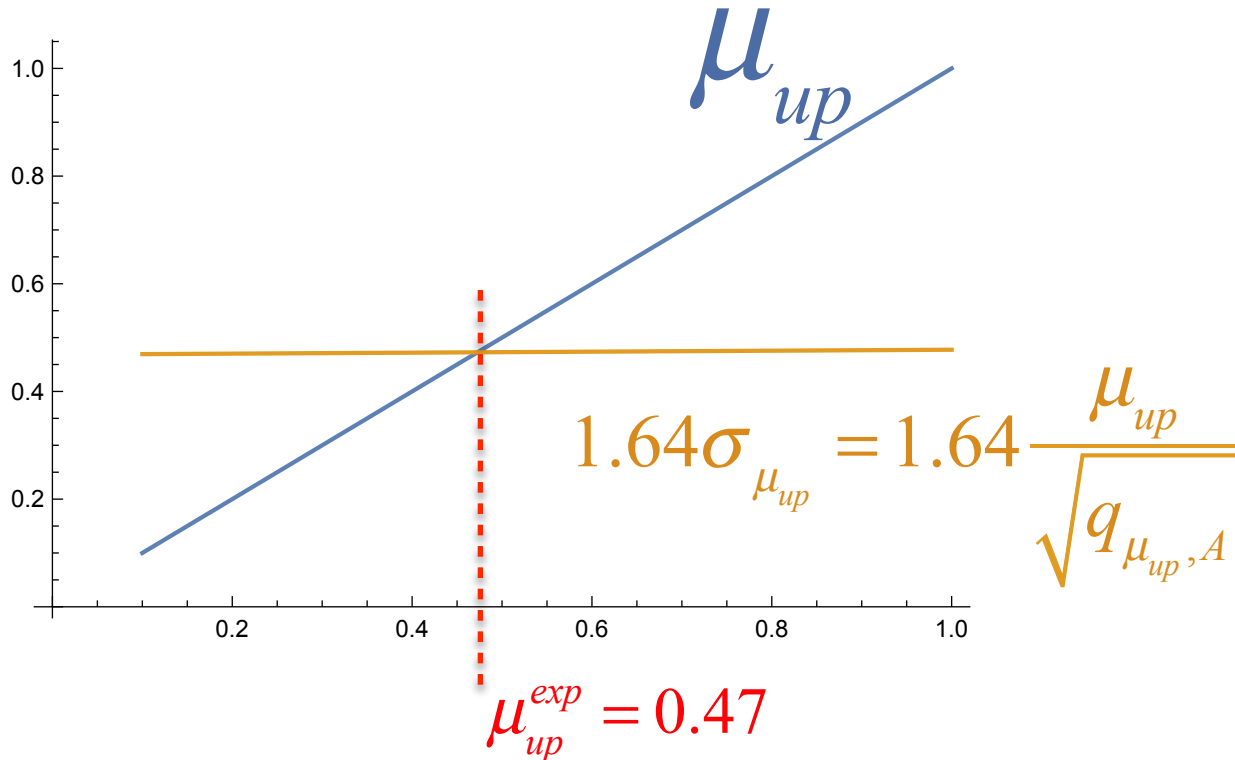
$$\mu_{up} = \sigma_{\hat{\mu}}(\mu_{up}) \Phi^{-1}(1 - 0.05) = \sigma_{\hat{\mu}}(\mu_{up}) 1.64$$

Find expected μ_{up}

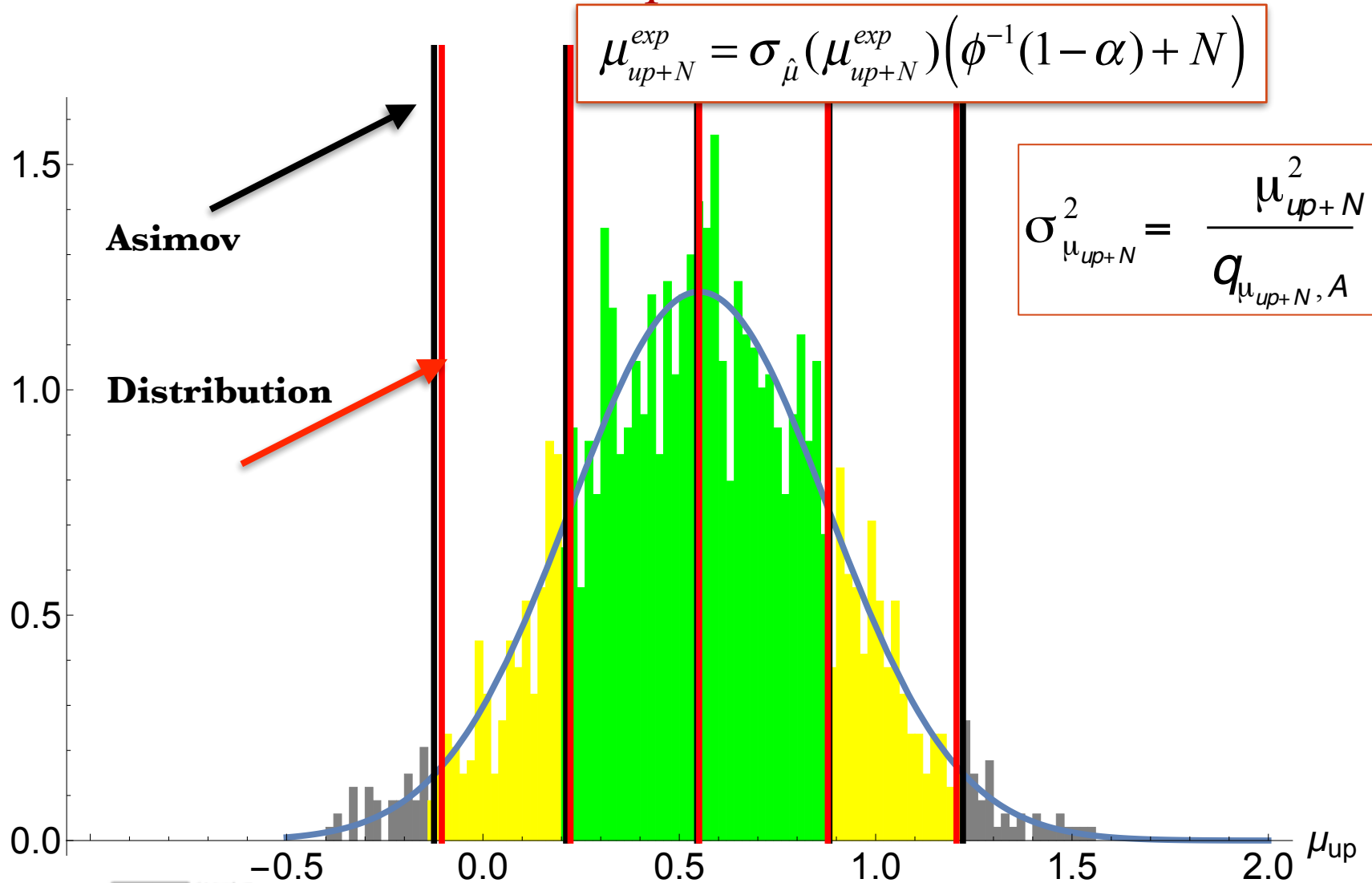
$$\mu_{up} = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up})\Phi^{-1}(1-0.05) = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up})1.64$$

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

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



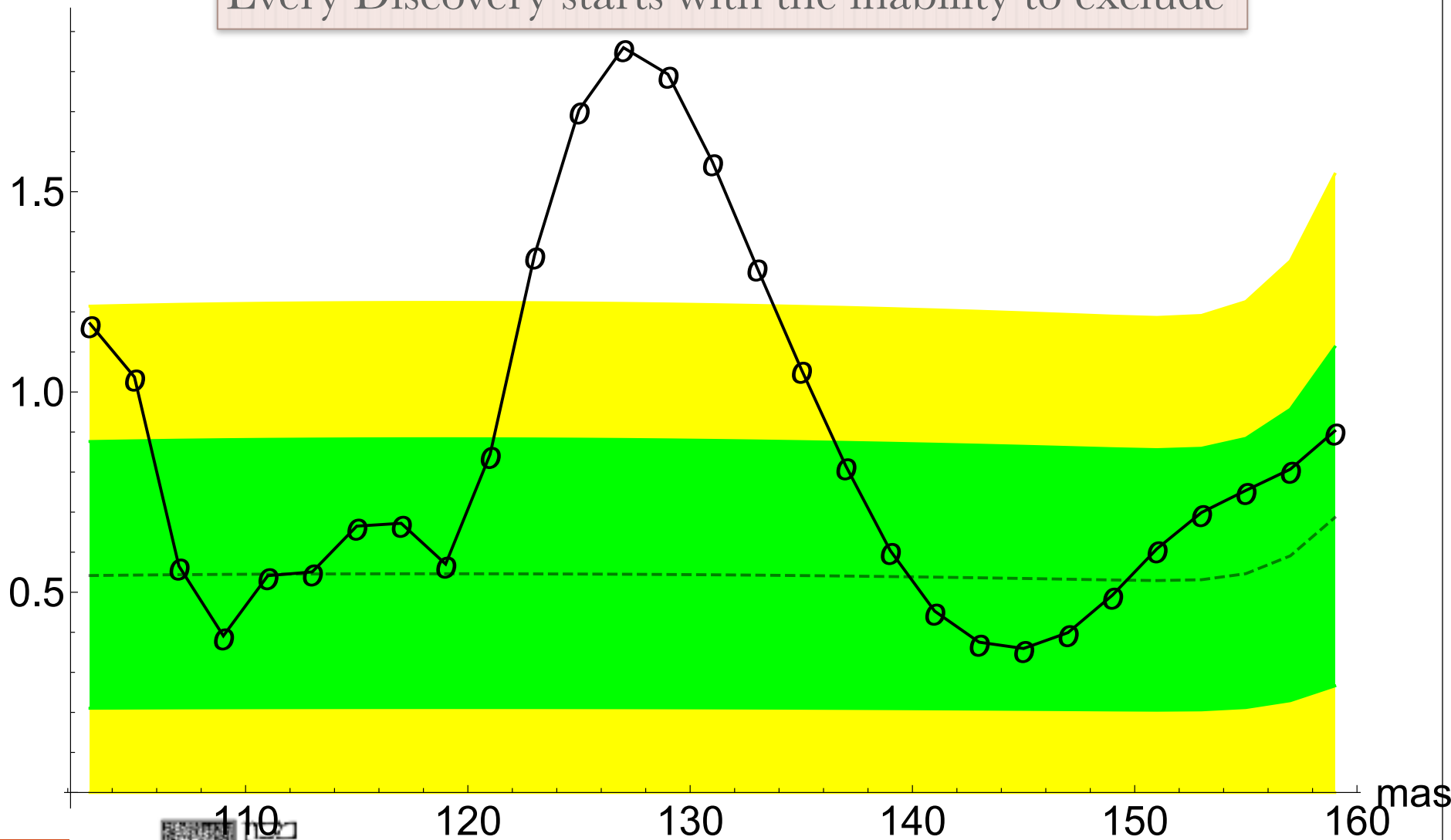
Expected μ_{up} Bands at $m=125$



Brazil Plot

μ_{up}

Every Discovery starts with the inability to exclude

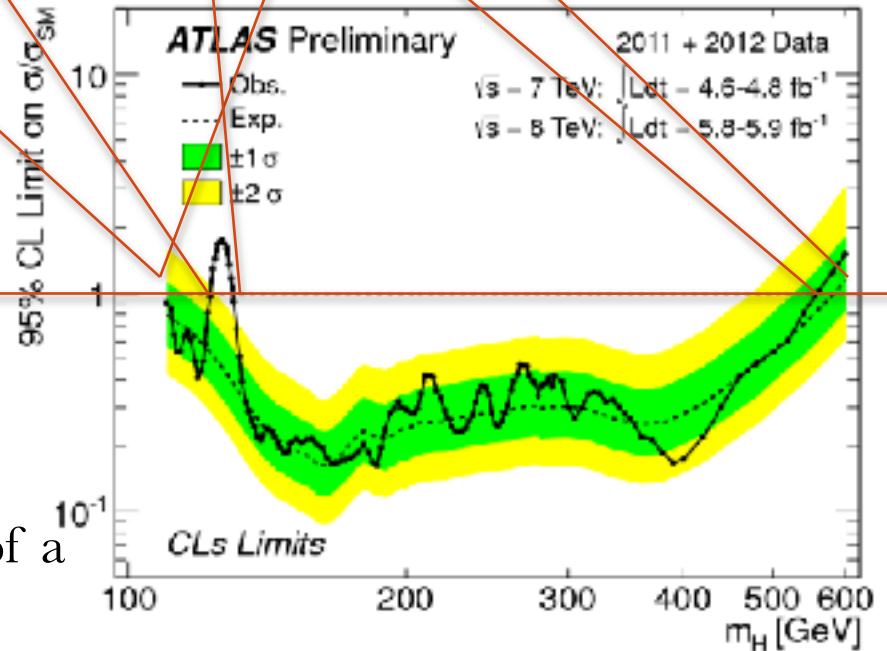


Understanding the Brazil Plot

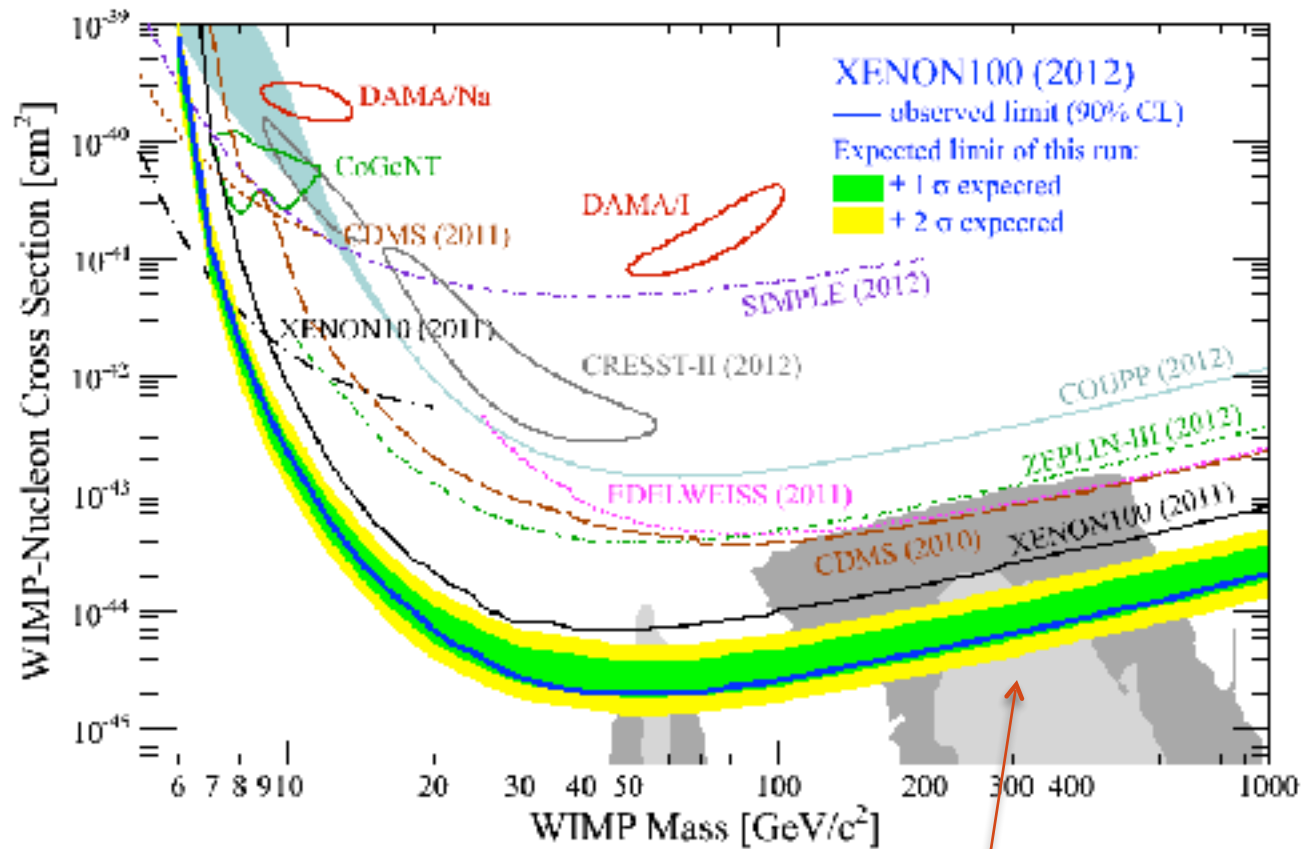
The expected 95% CL exclusion region covers the m_H range from 110 GeV to 582 GeV. The observed 95% CL exclusion regions are from 110 GeV to 122.6 GeV and 129.7 GeV to 558 GeV. The addition of

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



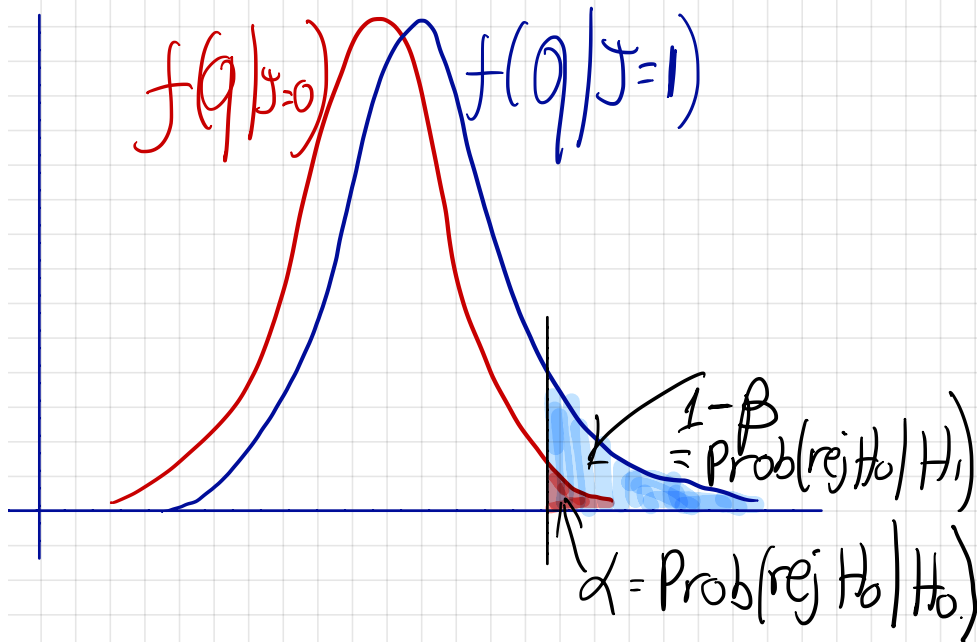
The lack of events in spite of an expected background allows us to set a better limit than the expected

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



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

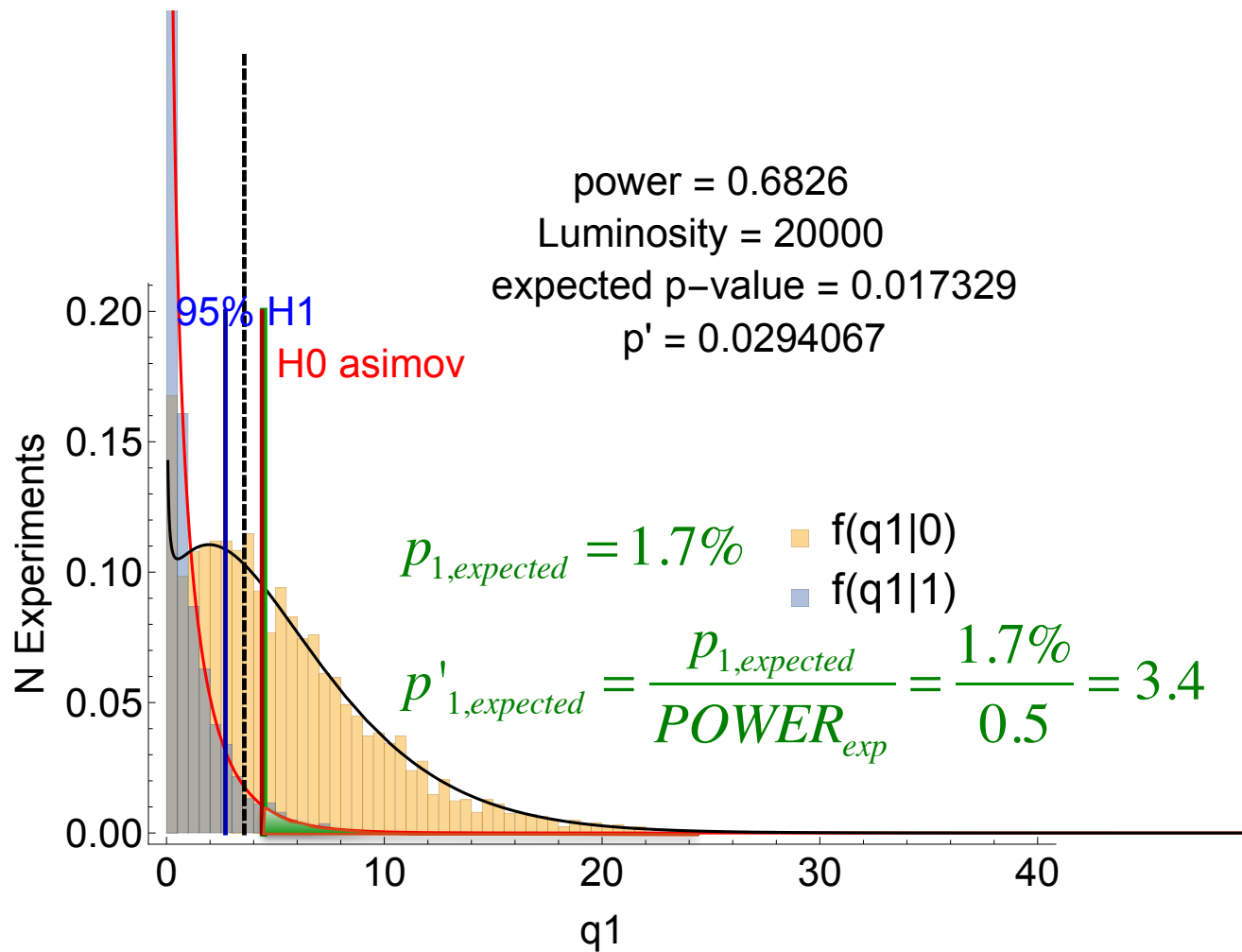
$$p' \equiv CL_s$$

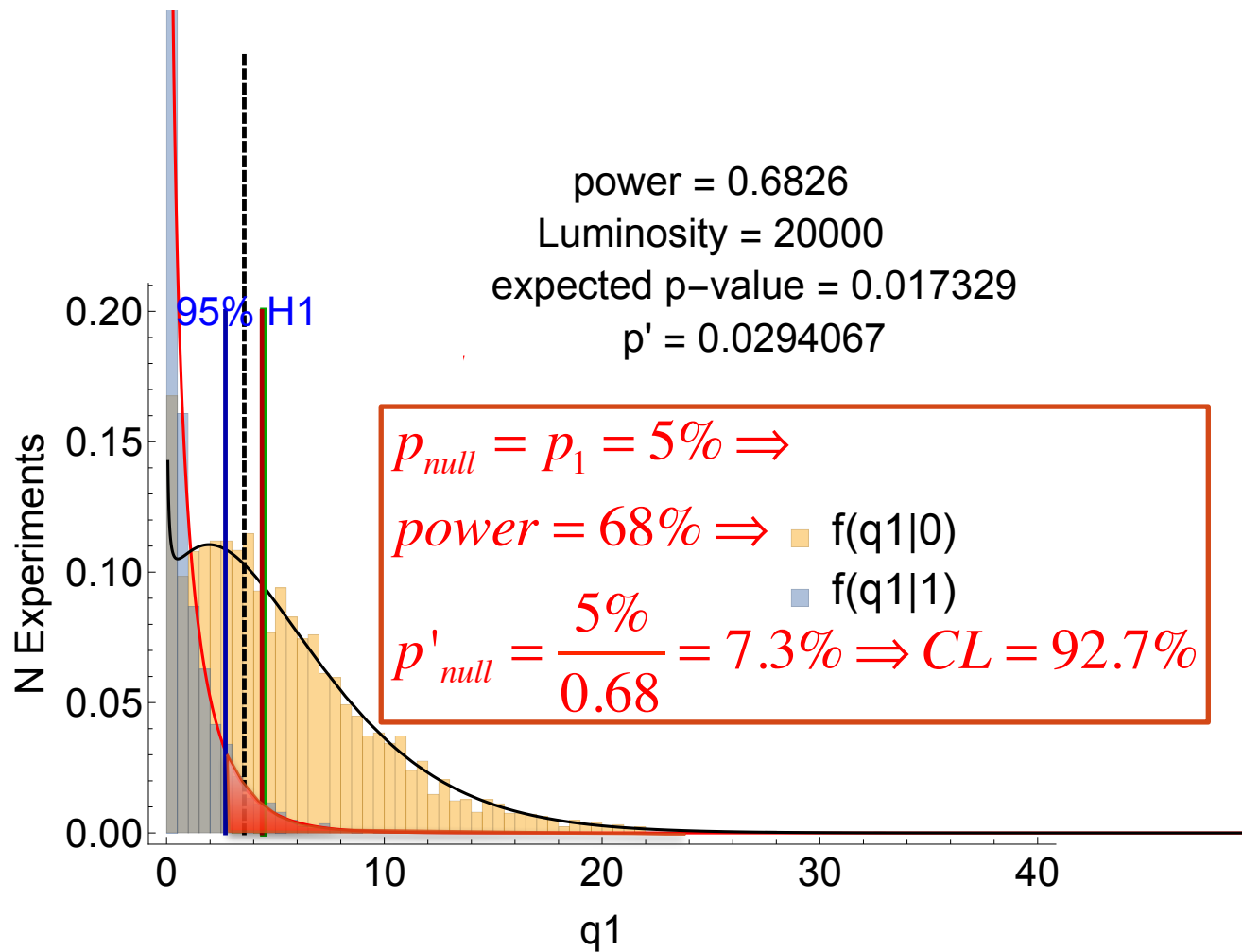
$$p'_\mu = \frac{P_\mu}{1 - p_0}$$

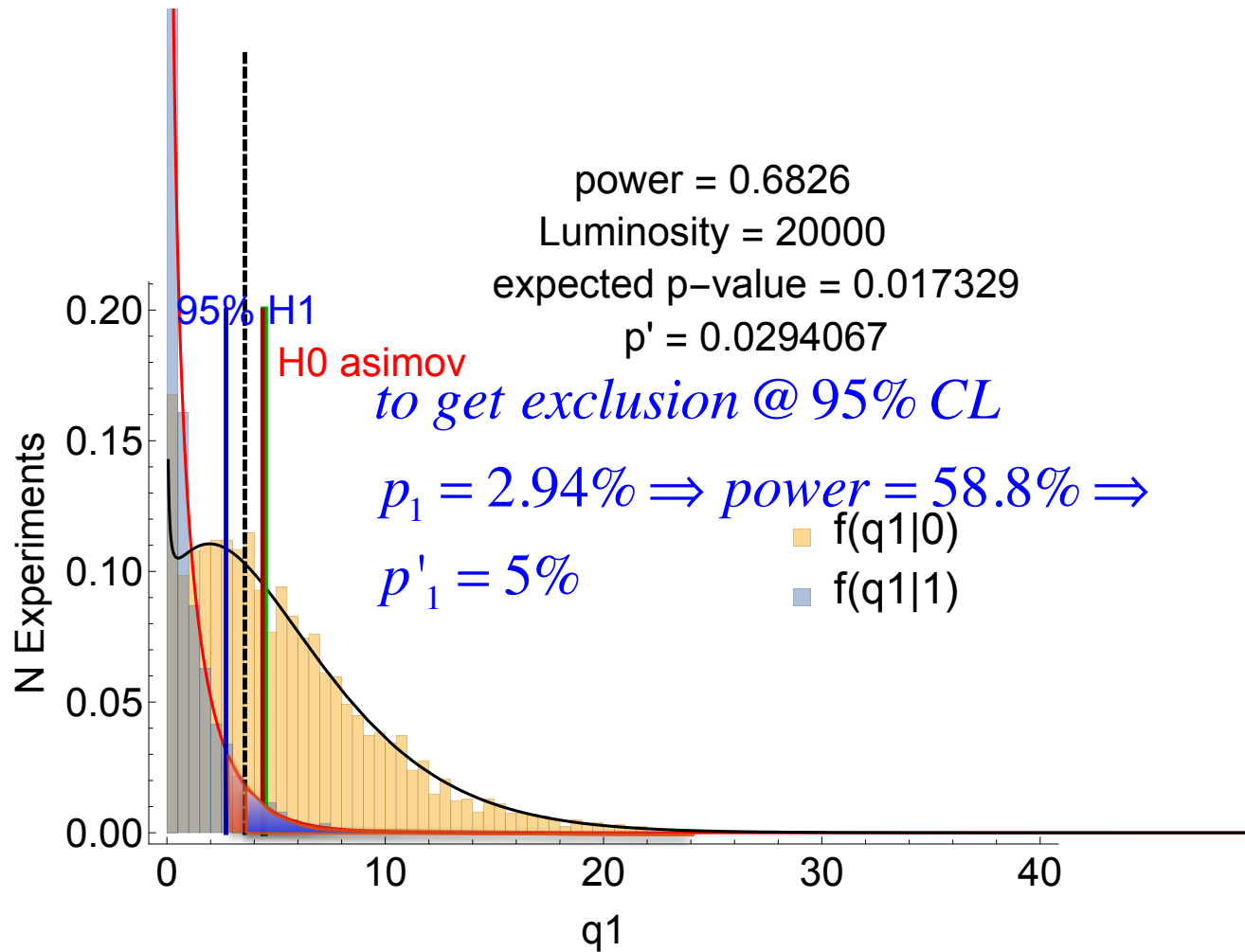
The CLs method
Was brought into
HEP By Alex Read (2002)
A.L. Read,
Presentation of search results:
The CL(s) technique,
"J. Phys. G {bf 28}, 2693 (2002).

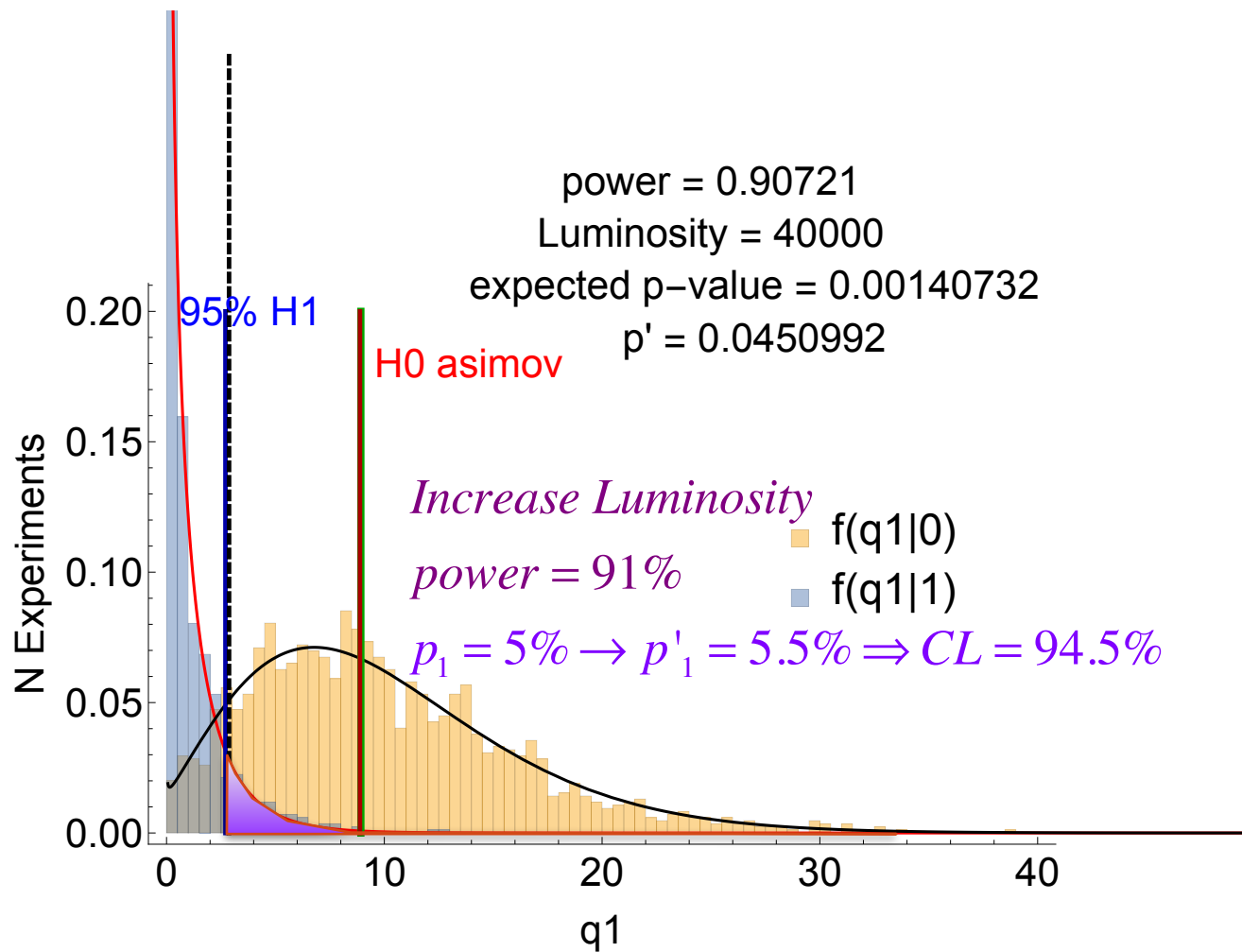
Birnbaum was re-discovered later
By O. Vitells

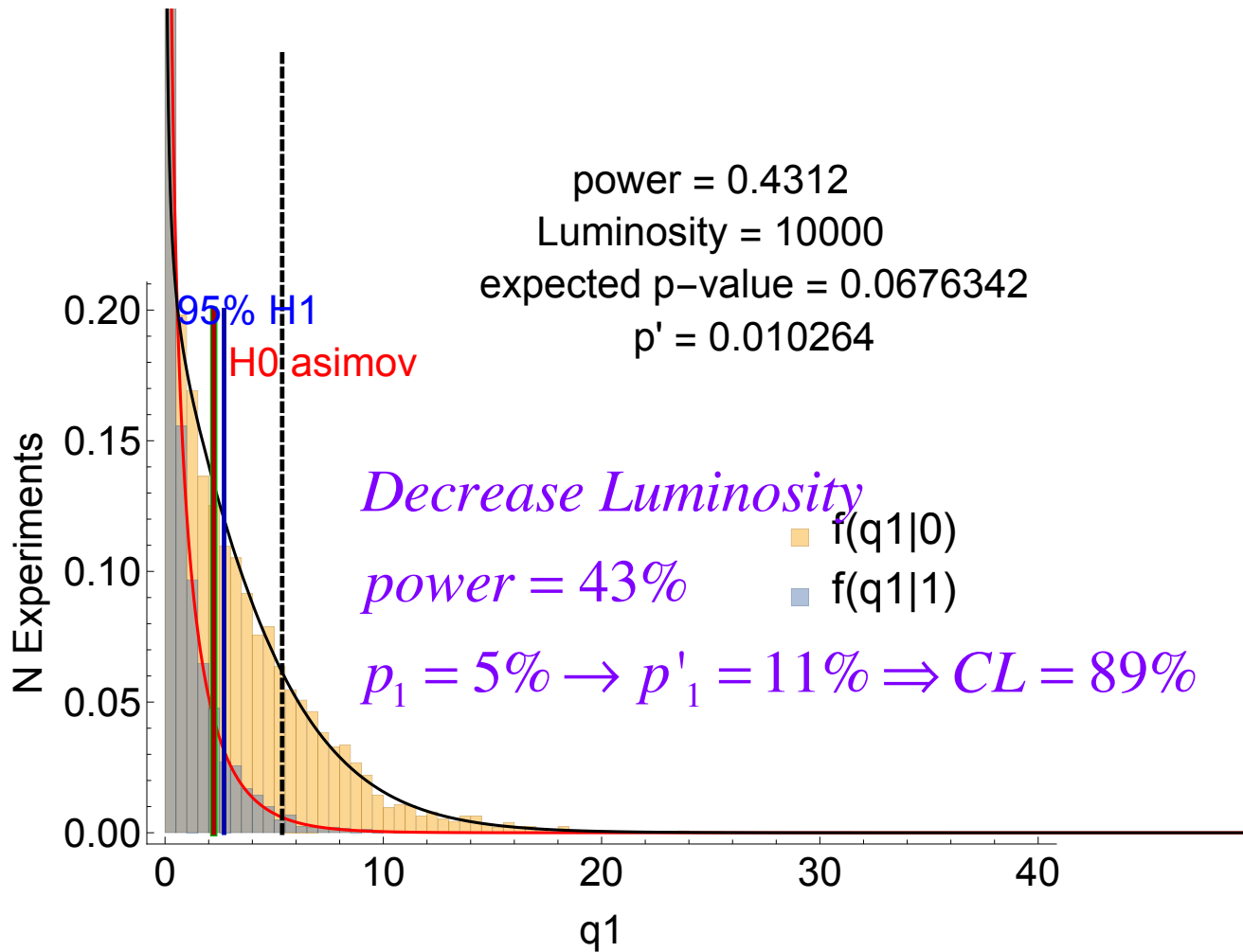


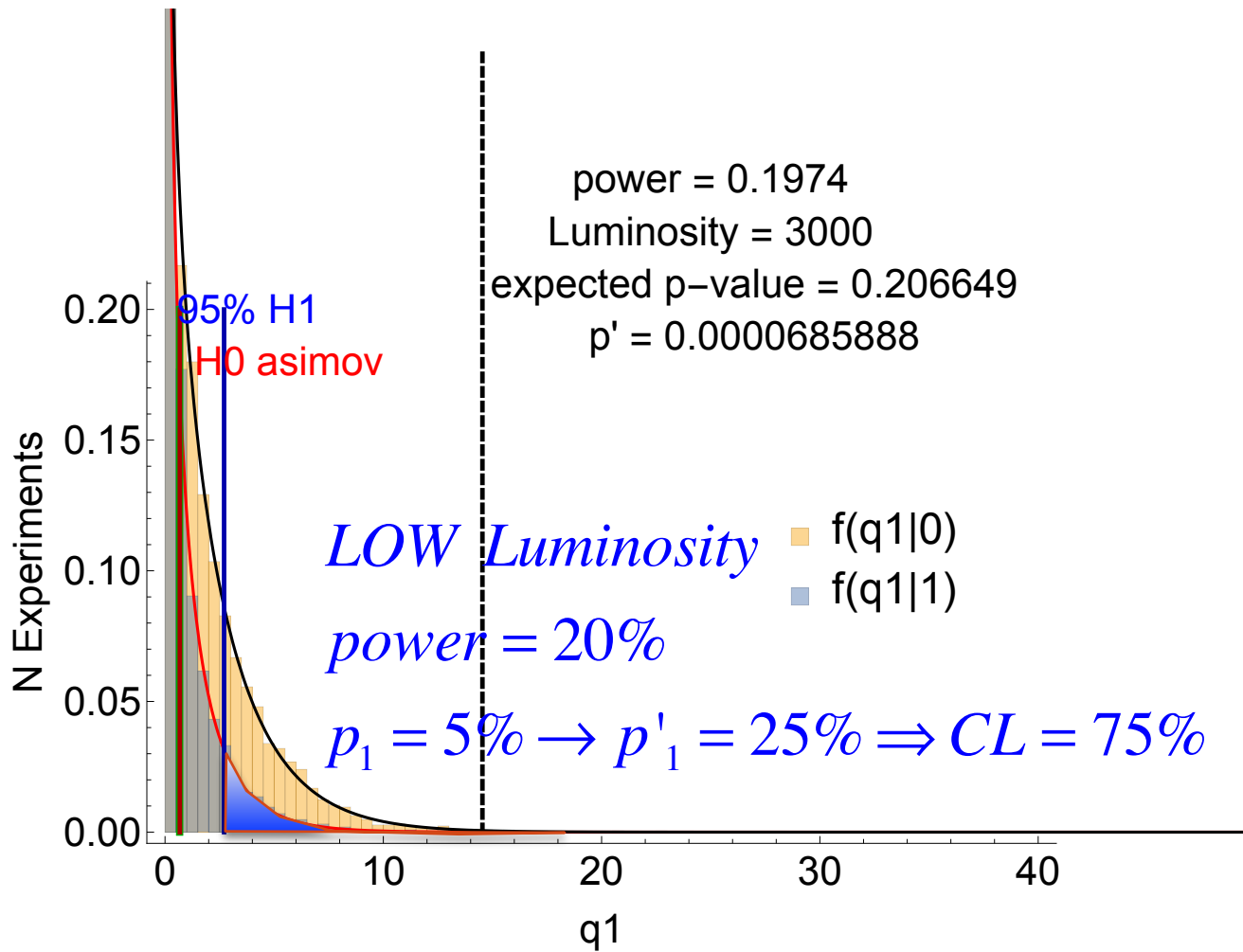


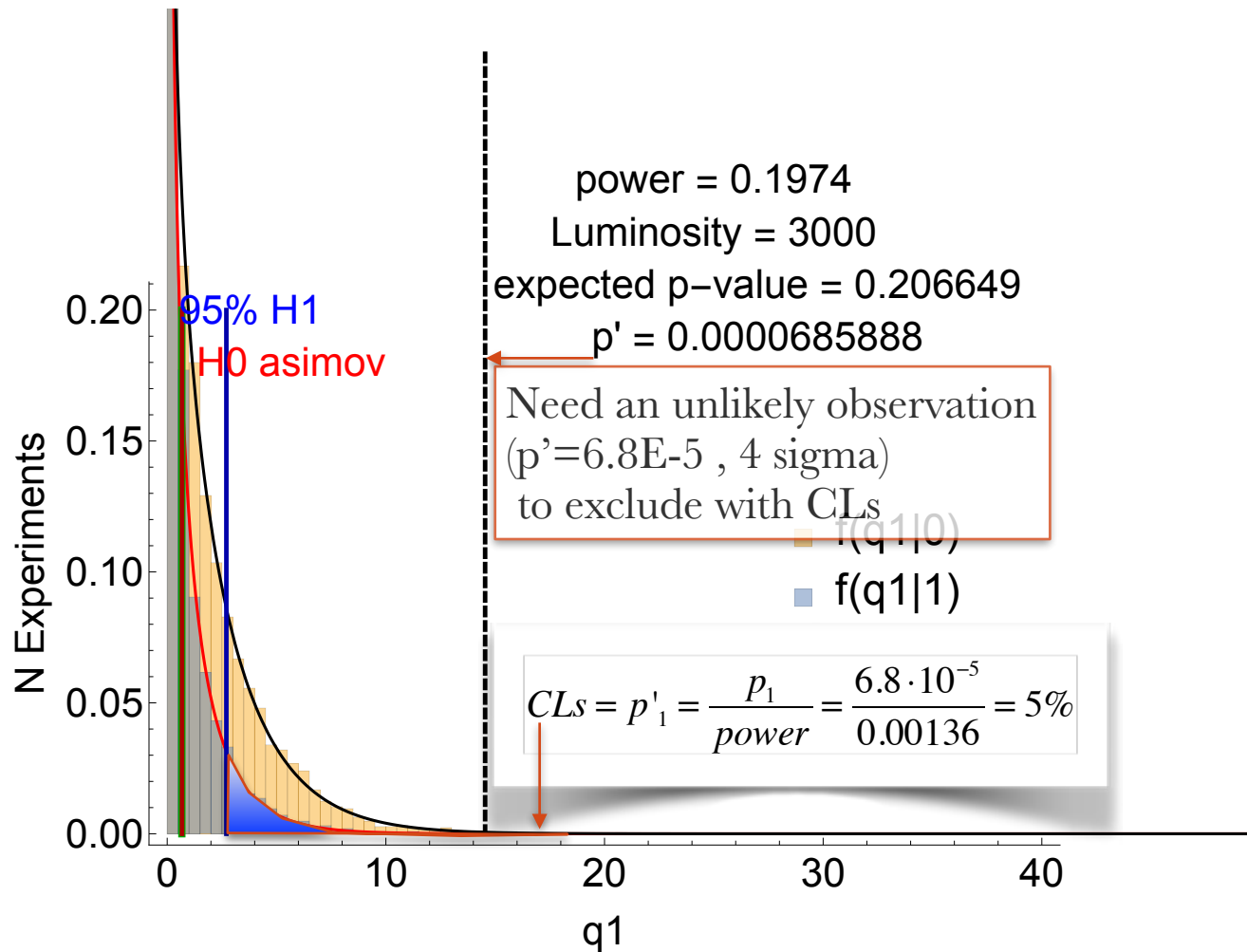












Hard to digest an observation with a power <5%



The Asymptotic and CLs

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

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

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

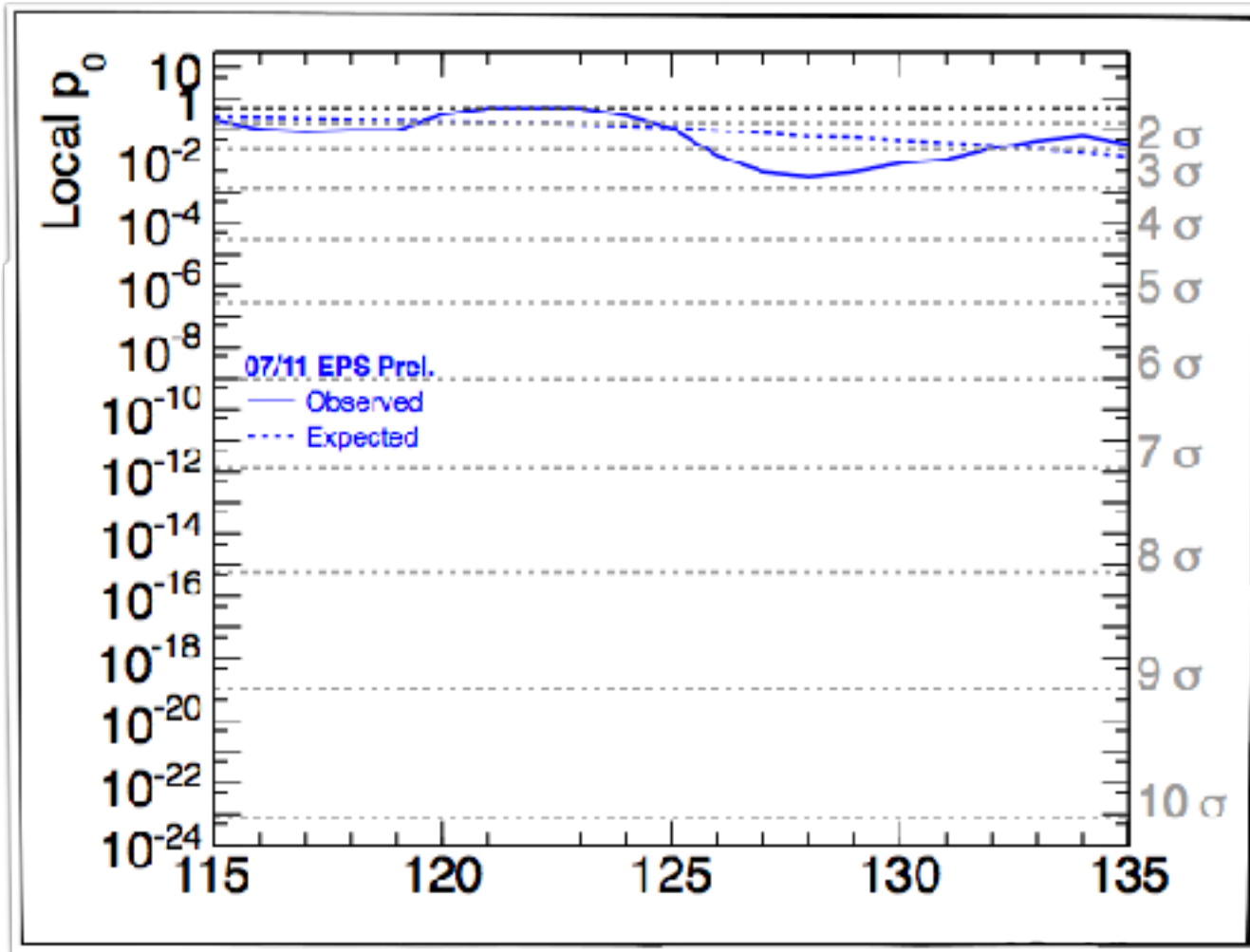
$$p'_\mu = \frac{1 - \Phi\left(\sqrt{q_{\mu,obs}}\right)}{\Phi\left(\sqrt{q_{\mu,A}} - \sqrt{q_{\mu,obs}}\right)} \Rightarrow$$

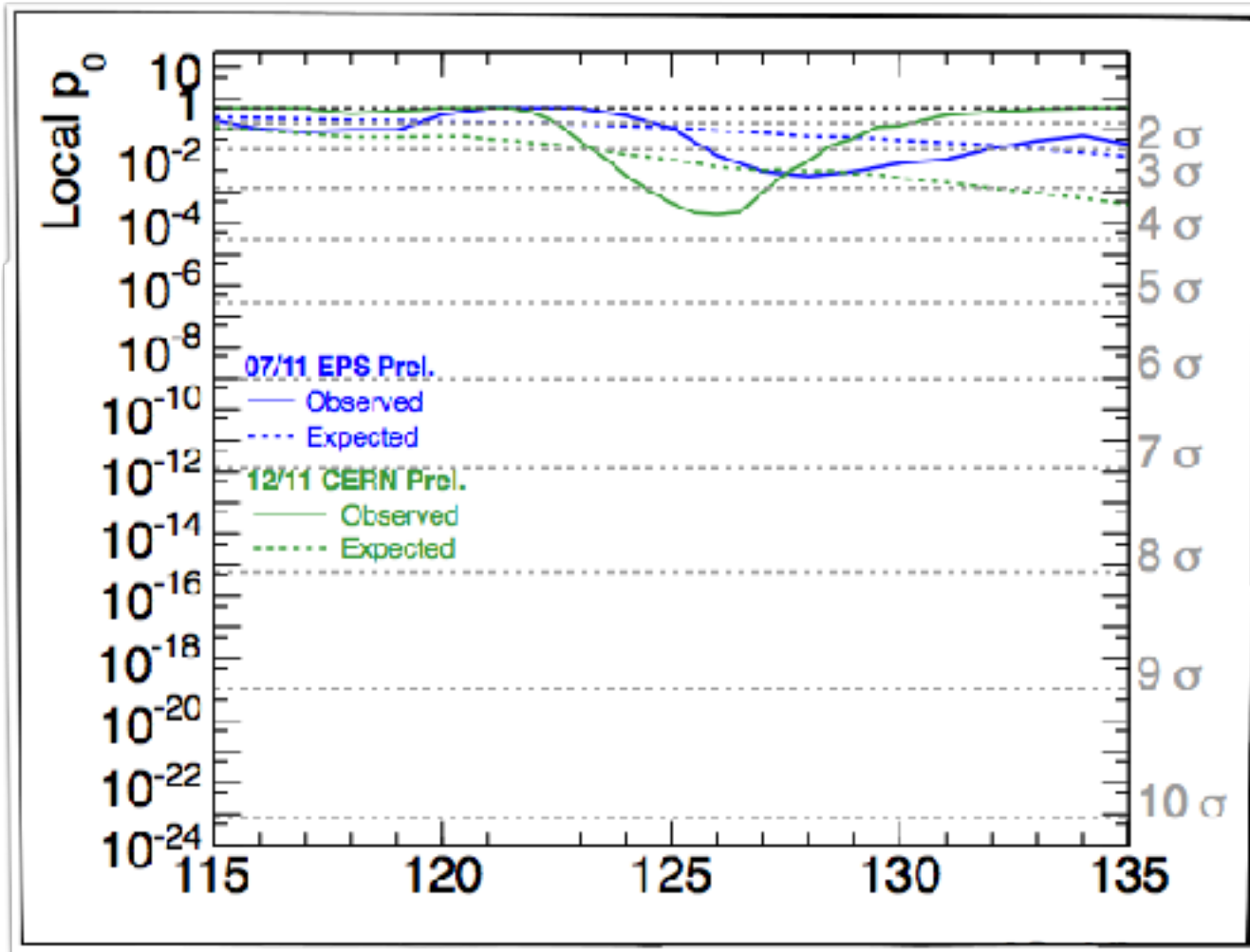
scan μ and find μ_{up}

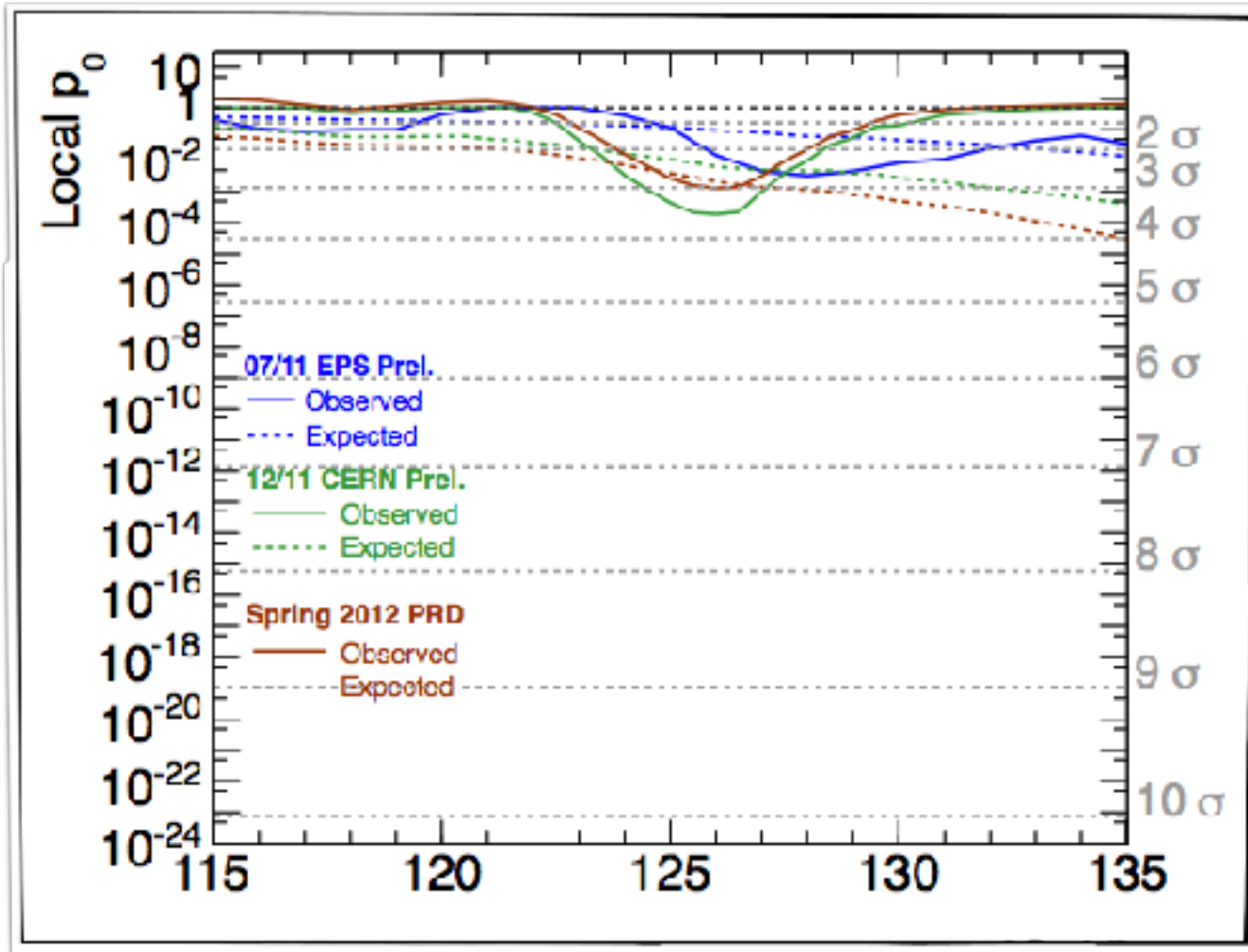
$$\mu_{up} = \left\{ \mu \left| \frac{1 - \Phi\left(\sqrt{q_{\mu,obs}}\right)}{\Phi\left(\sqrt{q_{\mu,A}} - \sqrt{q_{\mu,obs}}\right)} = 5\% \right. \right\}$$

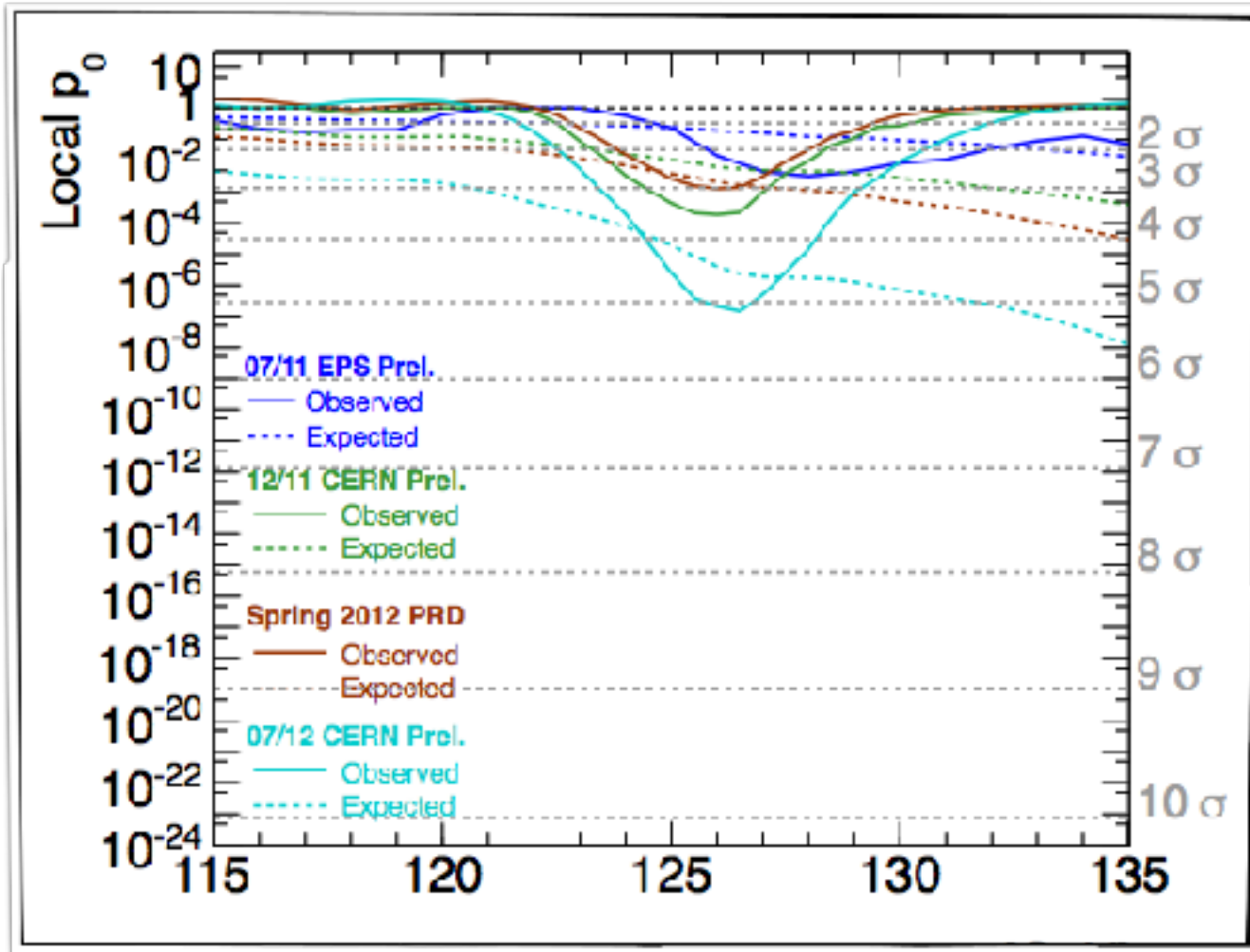
Examples (if time permits)

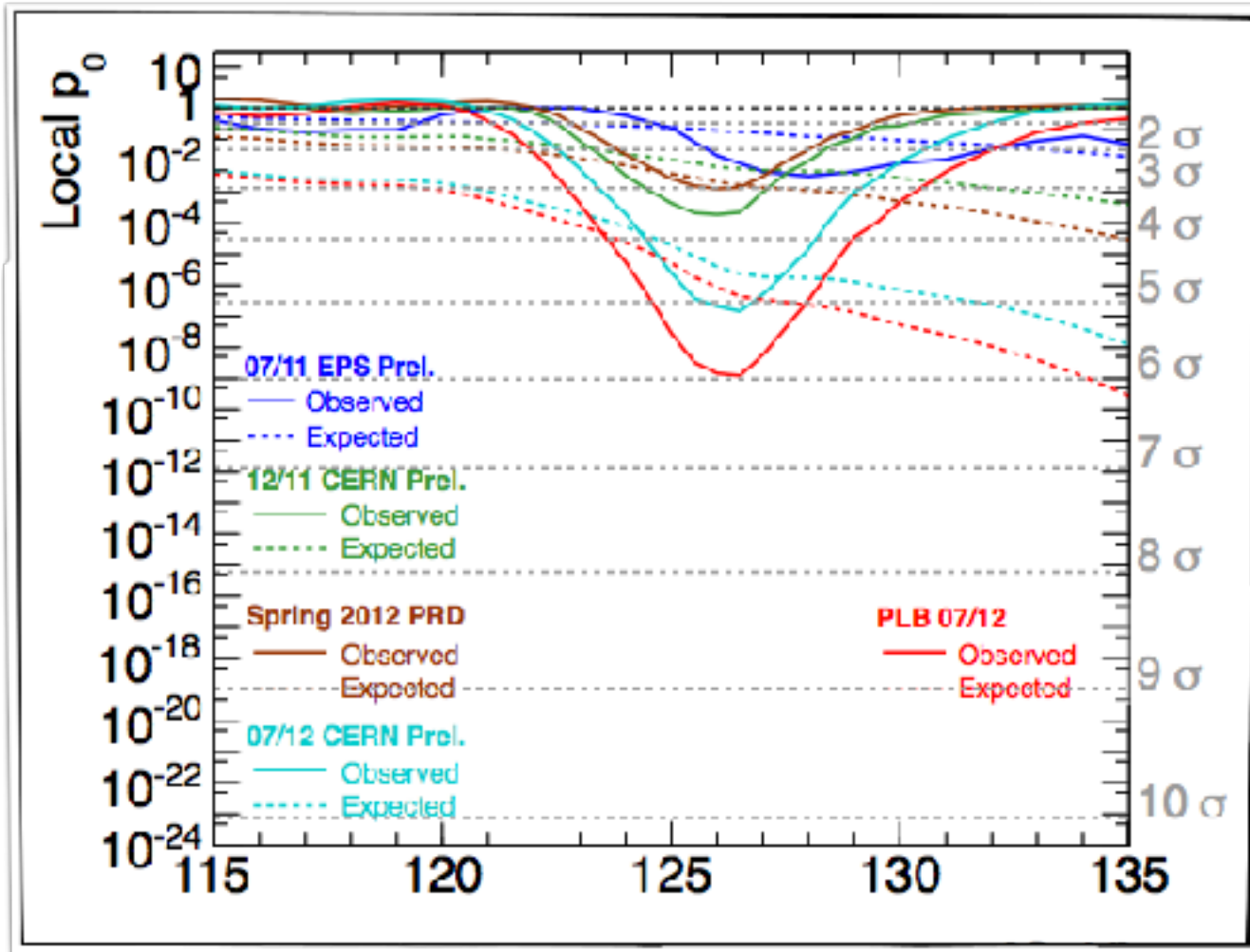


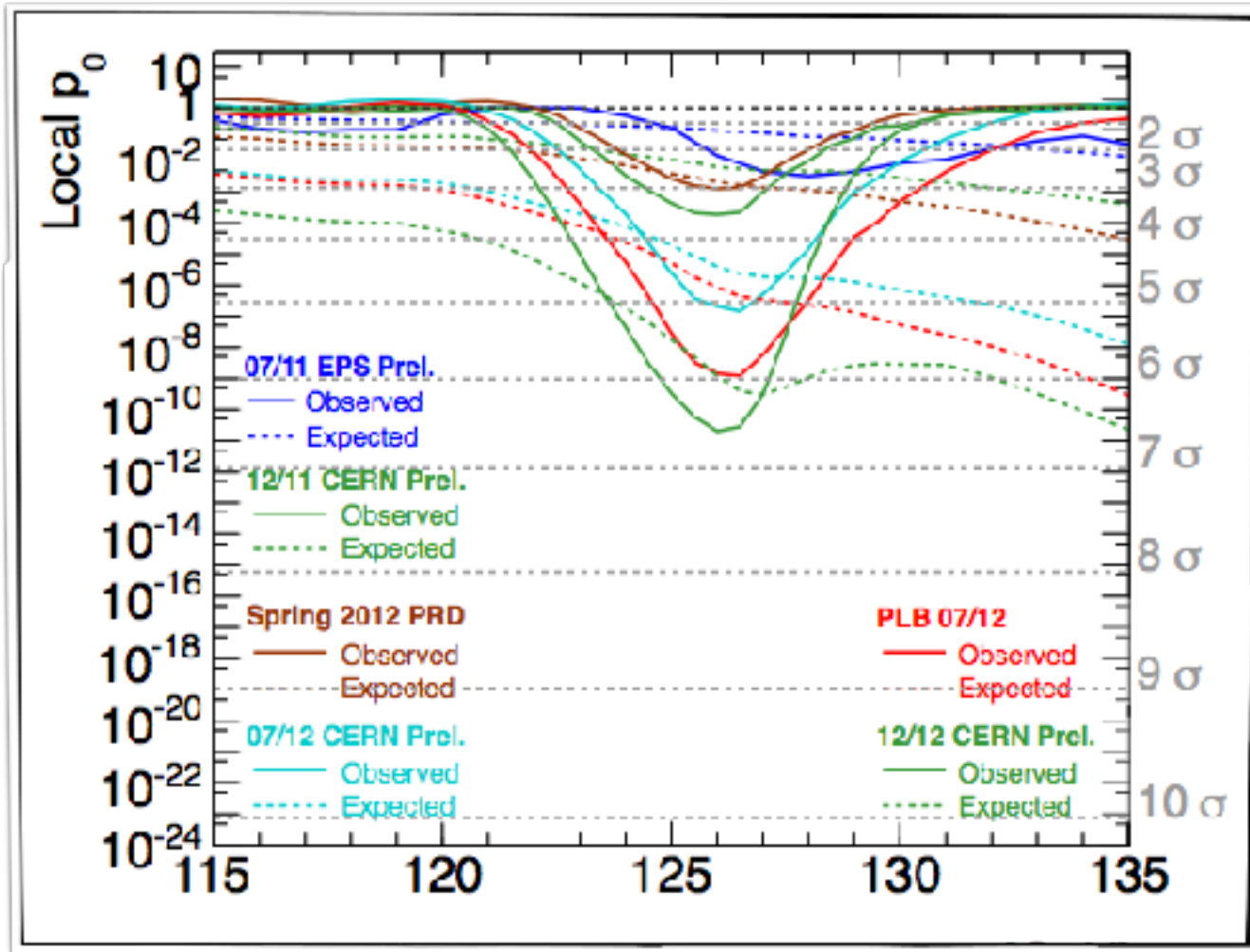


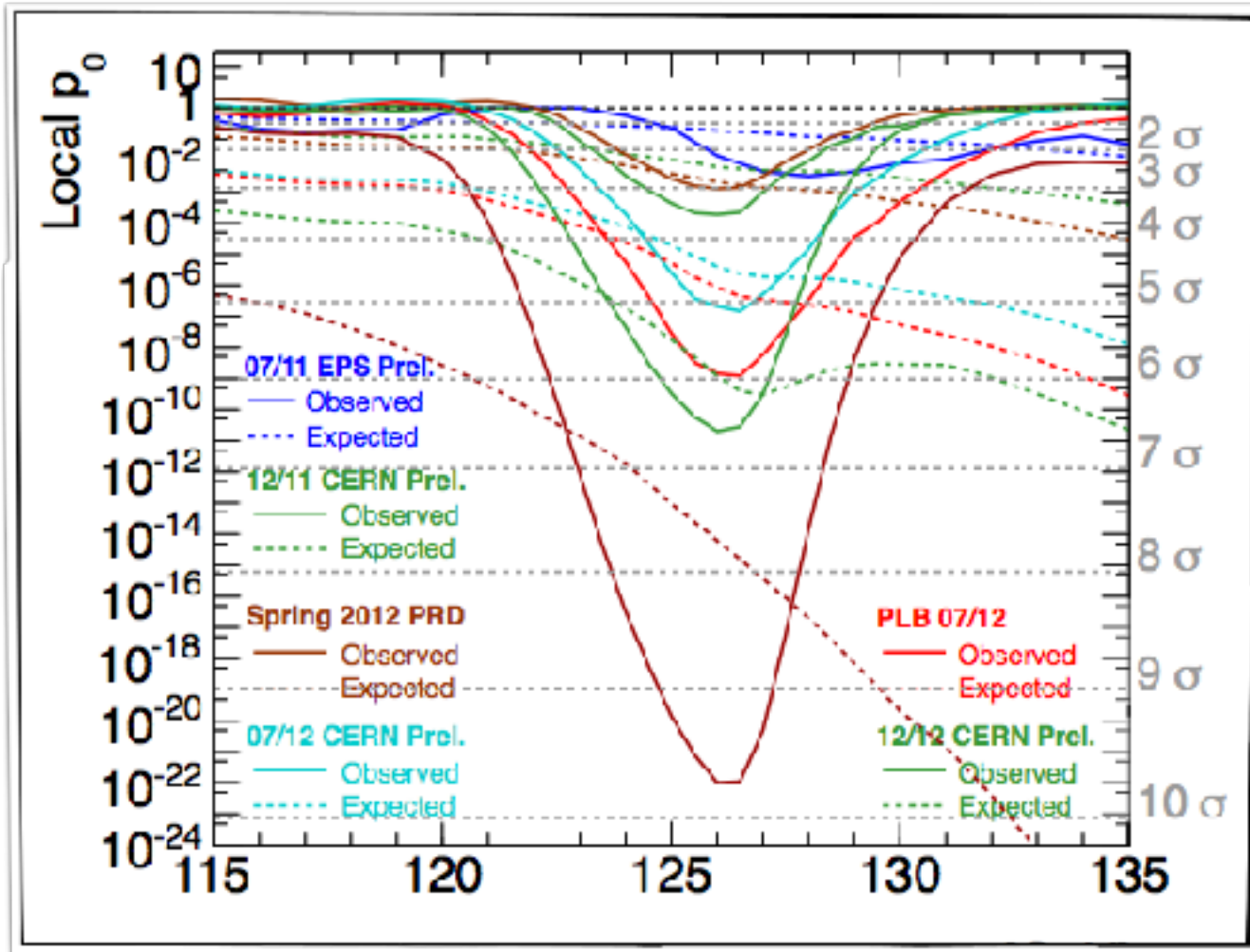




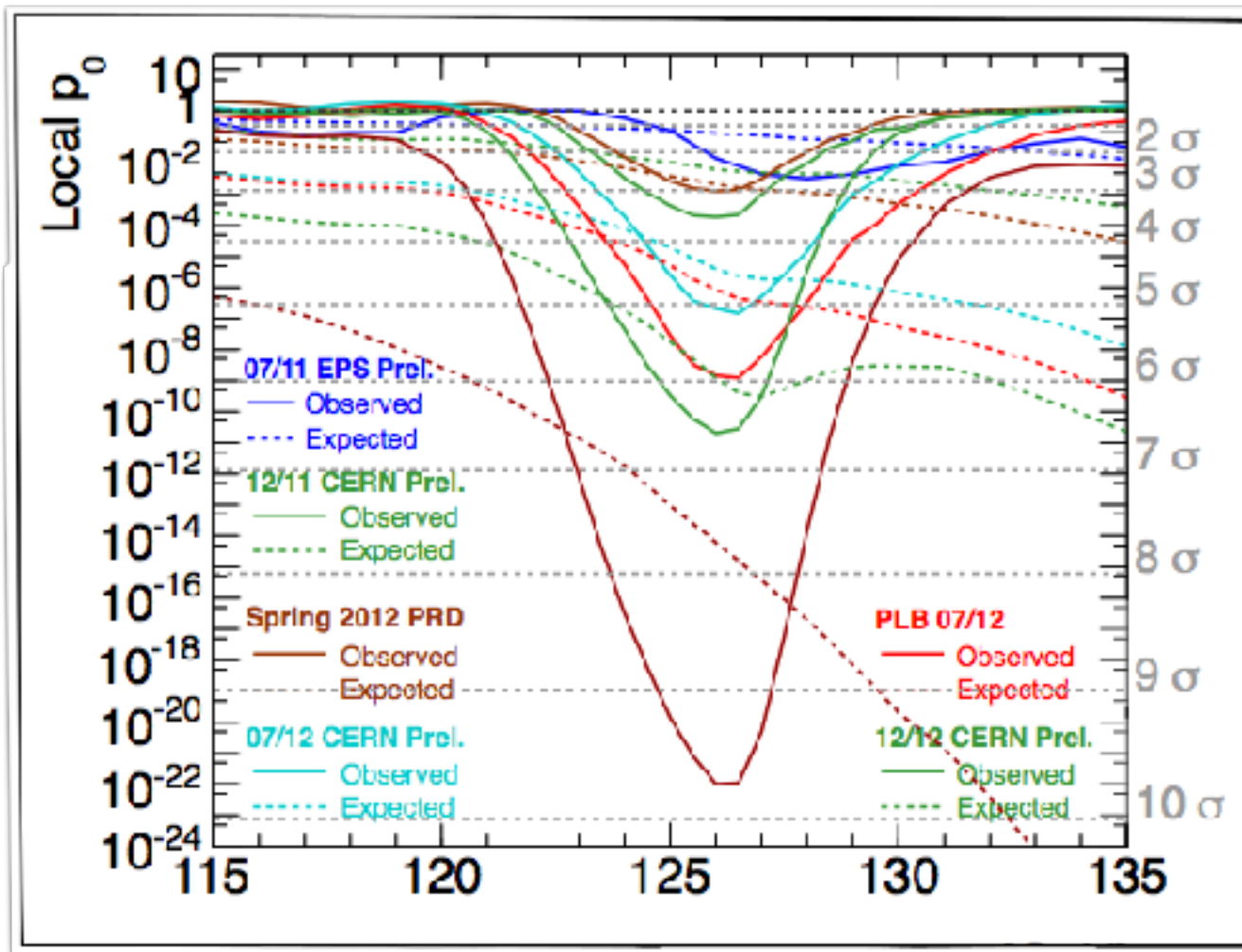








p0



p0 and the expected p0

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

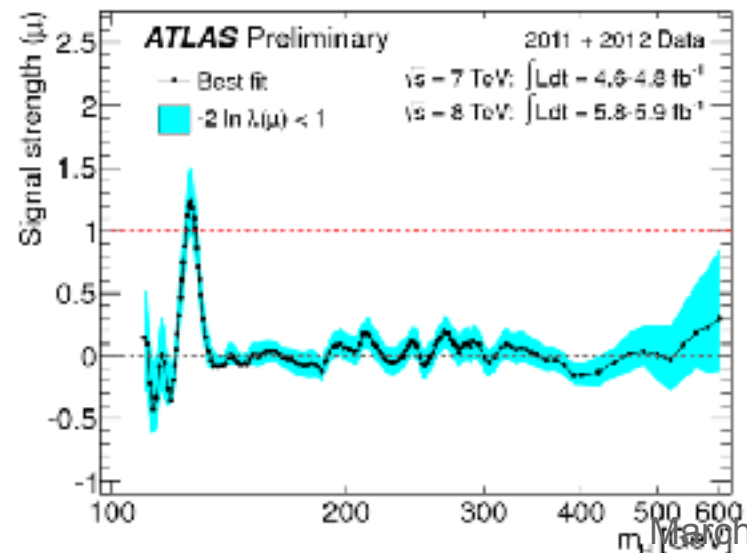
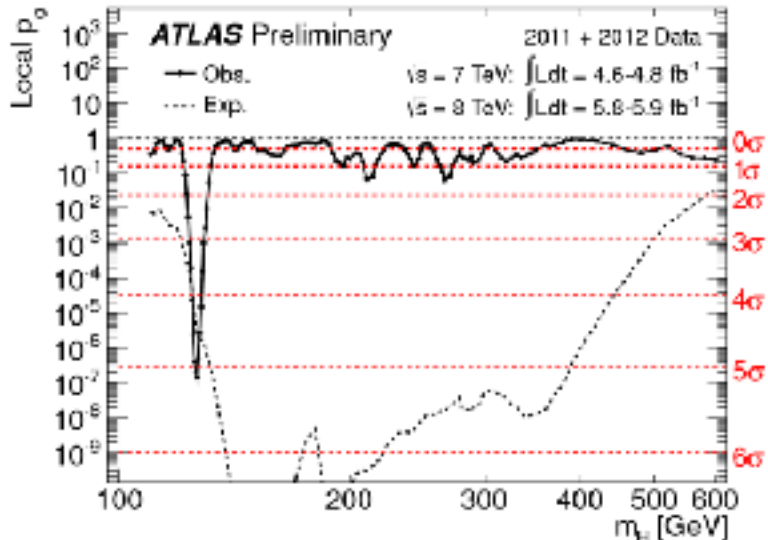
p_0 is the probability to observe a less BG like result (more signal like) than the observed one

Small p_0 leads to an observation

A tiny p_0 leads to a discovery

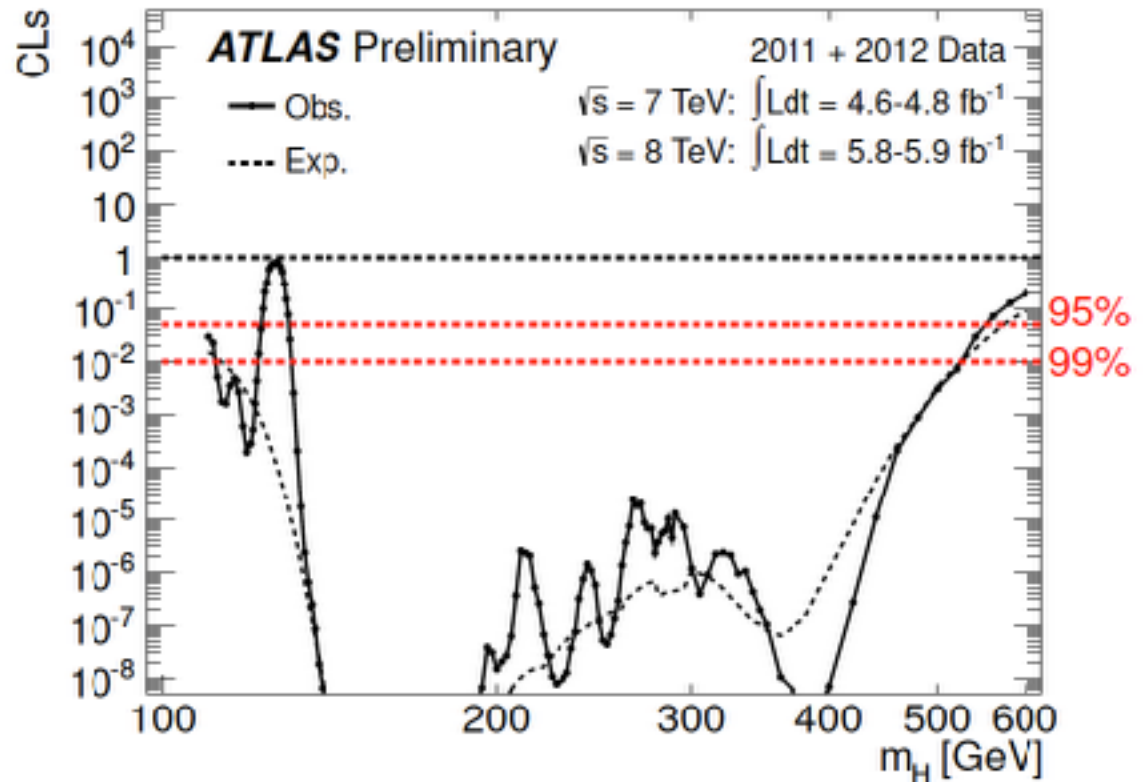
$$p = \int_Z^{p_{max}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 - \Phi(Z)$$

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



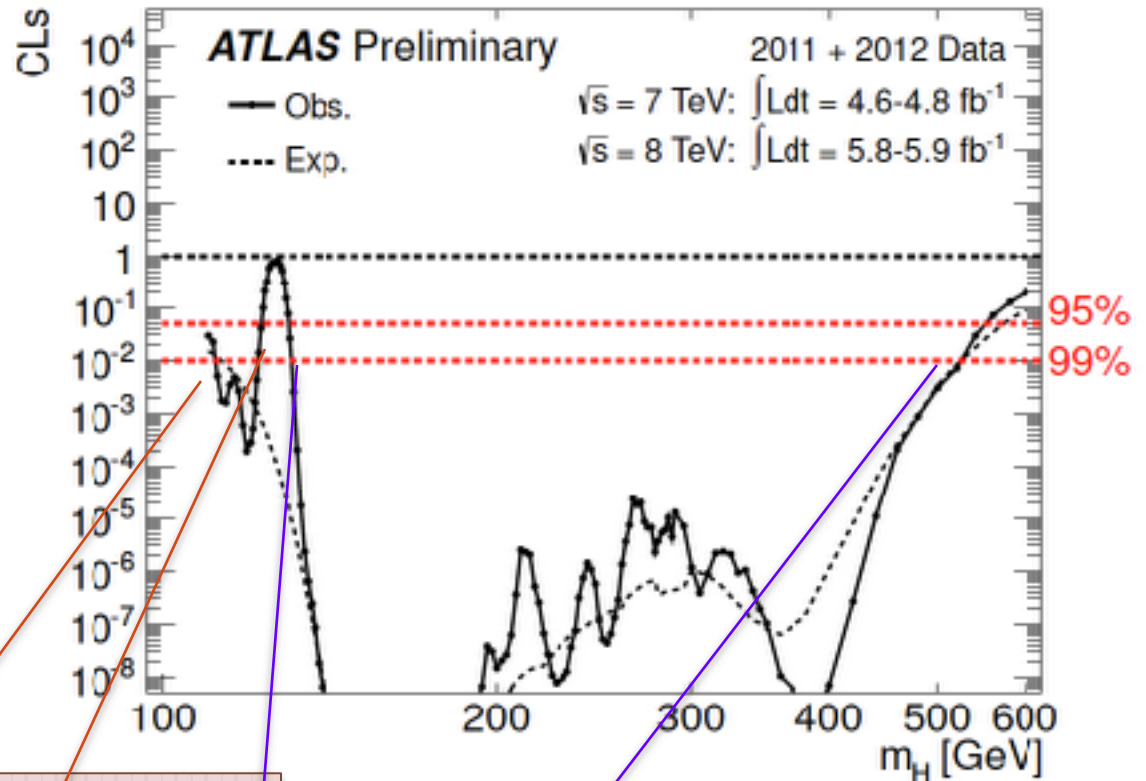
Understanding the CLs plot

- 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 $CL = 1 - CL_s = 1 - p'_s$

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.

Test Spin 0 parity

$$H_0 = 0^+$$

$$H_1 = 0^-$$

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

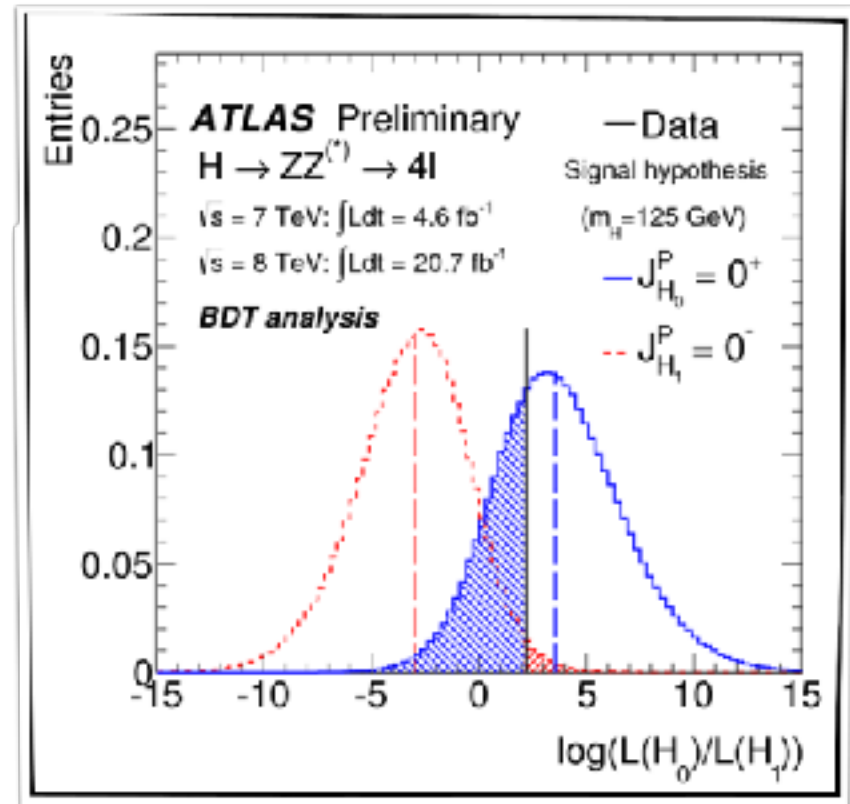
$$p_{H_1}(\text{exp} | H_0) = 0.37\%$$

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

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

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

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



Which means

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

H_1 like

H_0 like

More Magic (if time permits)



The New s/\sqrt{b}

The new s/\sqrt{b}

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

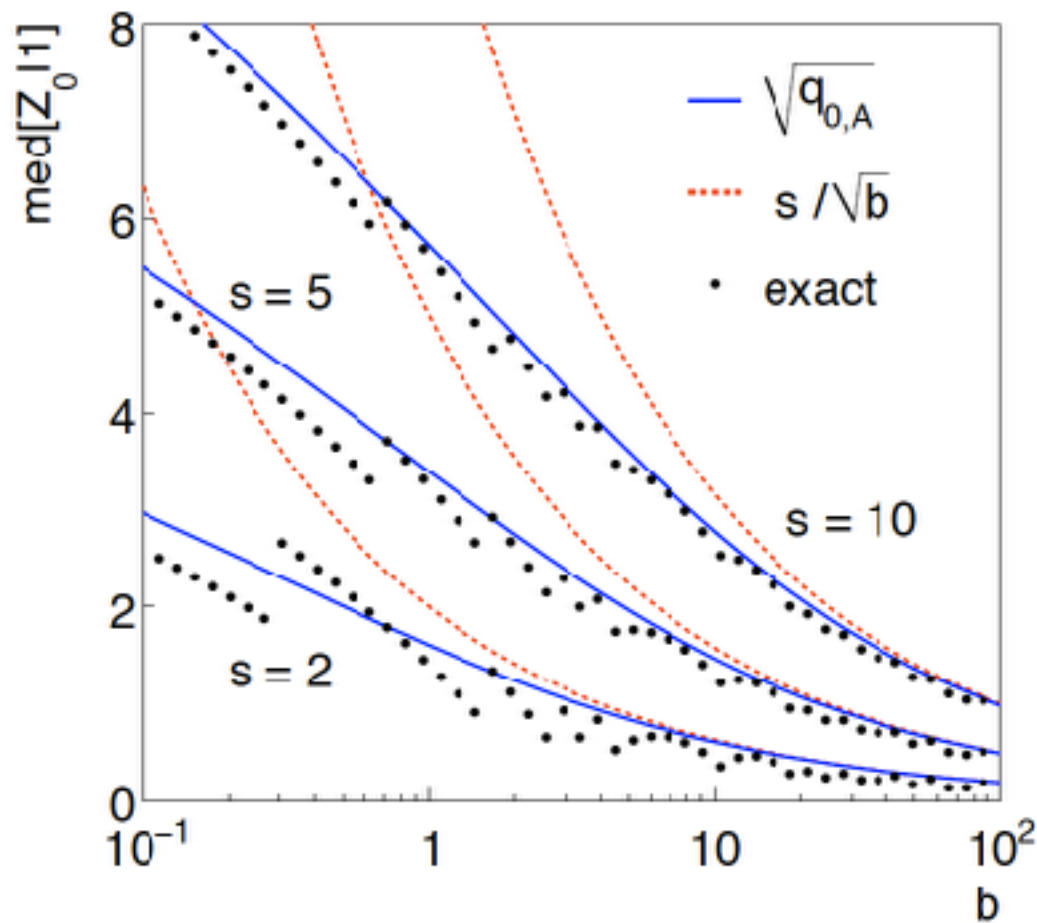
$$\text{med}[Z_0|1] = \sqrt{q_{0,A}} = \sqrt{2((s+b)\ln(1+s/b) - s)}$$

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



The New s/\sqrt{b}

s/\sqrt{b} ?



The new s/\sqrt{b}

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



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.,

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

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

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

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

But even that formula can be improved using the Asimov formalism

Significance with systematics

- We find (G. Cowan)

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

Expanding the Asimov formula in powers of s/b and σ_b^2/b gives

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

- 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

