Statistics for High Energy Physics
1.Profile Likelihood, Asimov, CLs
2. Exclusion,Discovery
3.Look Elsewhere Effect

Eilam Gross,
Weizmann Institute of Science

Basically ITS ALL ABOUT NUMBERS

$$n_{expected} = L \cdot \sigma \cdot eff$$

$$[L] = events/cm^2$$

$$[\sigma] = cm^2$$

So Far

A counting experiment

- The Higgs hypothesis is that of signal $s(m_H)$

  $$s(m_H) = L\sigma_{SM} \cdot \epsilon$$

  For simplicity unless otherwise noted $s(m_H) = L\sigma_{SM}$

- In a counting experiment $n = \mu s(m_H) + b$

  $$\mu = \frac{L\sigma_{obs}(m_H)}{L\sigma_{SM}(m_H)} = \frac{\sigma_{obs}(m_H)}{\sigma_{SM}(m_H)}$$

- μ is the strength of the signal (with respect to the expected Standard Model one)

- The hypotheses are therefore denoted by H_μ

- H_1 is the SM with a Higgs, H_0 is the background only model

A Tale of Two Hypotheses

NULL

H_0 - SM w/o Higgs

ALTERNATE

H_1 - SM with Higgs

Rejecting H_0 in favour of $H_1(m_H)$

→Discovery of a Higgs with a mass m_H

We quantify rejection by p-value (later)

Swapping Hypotheses→exclusion

NULL

H_0- SM w/o Higgs

ALTERNATE

H_1- SM with Higgs

- Reject H_1 in favor of H_0

Excluding $H_1(m_H)$→Excluding the Higgs with a mass m_H

We quantify rejection by p-value (later)

Likelihood

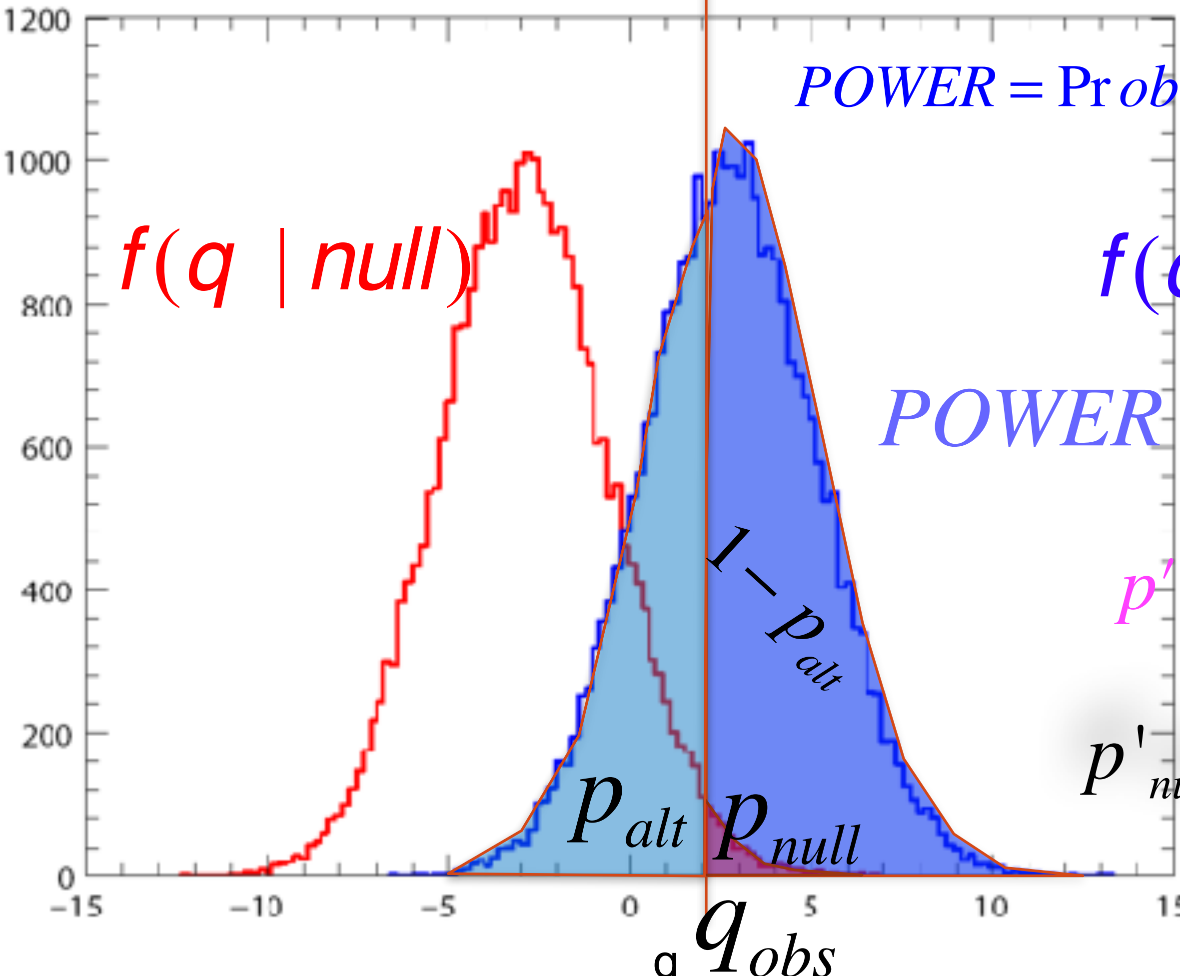
- Likelihood is the compatibility of the Hypothesis with a given data set.
  But it depends on the data

$$L(H) = P(x \mid H)$$

Test statistic and p-value

CLs

$$If\ \ p \le \alpha\ \ reject\ null$$



$$POWER = \Pr ob(rej\ H_{null} \mid H_{alt})$$

$$POWER = 1 - p_{alt}$$

$$p' = \frac{p_{null}}{Power}$$

$$p'_{null} = \frac{p_{null}}{1 - p_{alt}}$$

$$p' = CL_s$$

Null like → alt like

Test Spin 0 parity

$$H_0 = 0^+$$

$$H_1 = 0^-$$

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

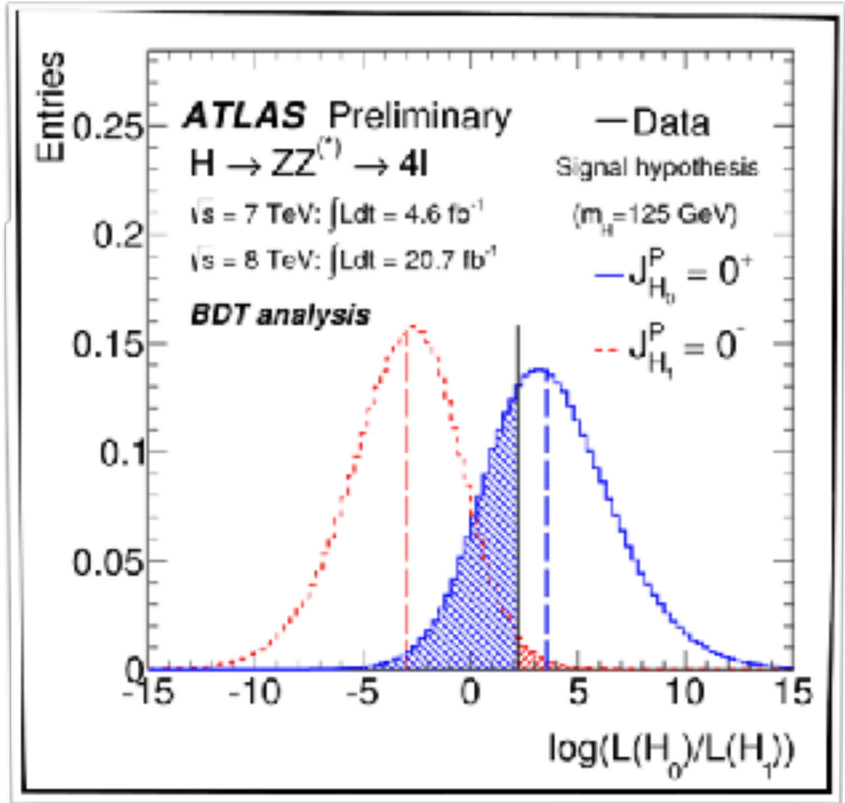
$$p_{H_1}(\exp \mid H_0) = 0.37\%,$$

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

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

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

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

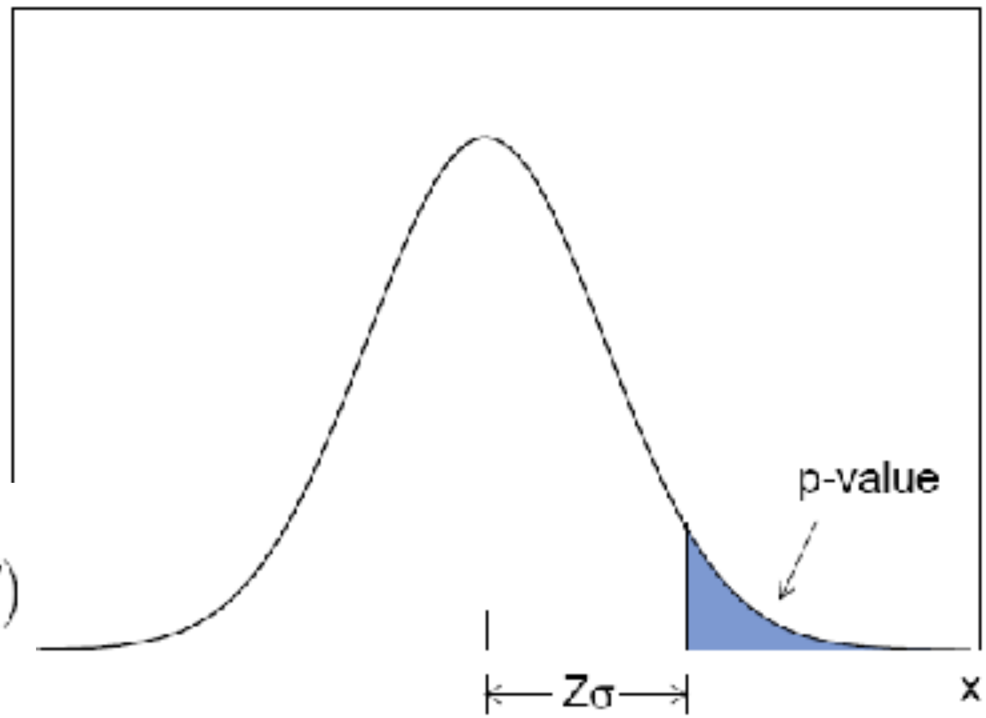


H_1 like ← H_0 like

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

From p-values to Gaussian significance

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



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

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

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

Beware of 1 vs 2-sided definitions!

p-value – testing the null hypothesis

When testing the b hypotheis (null=b), it is custom to set

$\alpha = 2.9\ 10^{-7}$

→ if $p_b < 2.9\ 10^{-7}$ the b hypothesis is rejected

→ Discovery

When testing the s+b hypothesis (null=s+b), set α =5%

if p_{s+b}<5% the signal hypothesis is rejected at the 95% Confidence Level (CL)

→ Exclusion

Nuisance Parameters

or Systematics

Nuisance Parameters (Systematics)

- There are two kinds of parameters:
  - Parameters of interest (signal strength… cross section… μ)
  - Nuisance parameters (background (b), signal efficiency, resolution, energy scale,…)

- The nuisance parameters carry systematic uncertainties

Pulls and Ranking of NPs

The pull of θ_i is given by $\dfrac{\hat{\theta}_i - \theta_{0,i}}{\sigma_0}$

without constraint $\sigma\left(\dfrac{\hat{\theta}_i - \theta_{0,i}}{\sigma_0}\right) = 1 \quad \left\langle \dfrac{\hat{\theta}_i - \theta_{0,i}}{\sigma_0} \right\rangle = 0$

It's a good habit to look at the pulls of the NPs and make sure that Nothing irregular is seen

In particular one would like to guarantee that the fits do not over constrain a NP in a non sensible way

Implementation of Nuisance Parameters

- Implement by marginalizing (Bayesian) or profiling
  (Frequentist)

- Marginalization (Integrating))
  - Integrate the Likelihood, L, over possible values of nuisance parameters (weighted by their prior belief functions -- Gaussian,gamma, others...)
  - $$L(\mu) = \int L(\mu,\theta)\pi(\theta)d\theta$$

The Hybrid Cousins-Highland Marginalization

Cousins & Highland

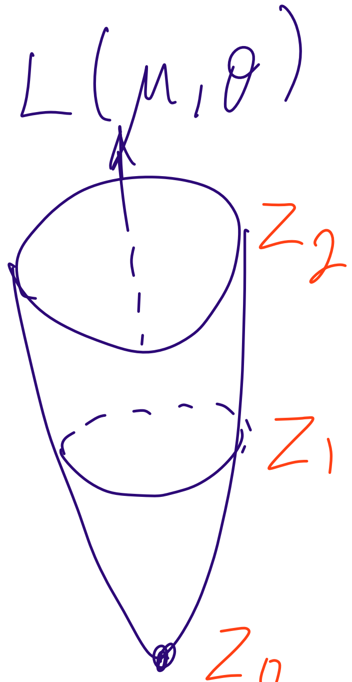
$$q = \frac{L(s+b(\theta))}{L(b(\theta))} \Rightarrow \frac{\int L\big(s+b(\theta)\big)\pi(\theta)d\theta}{\int L\big(b(\theta)\big)\pi(\theta)d\theta}$$

Profiling the NPs

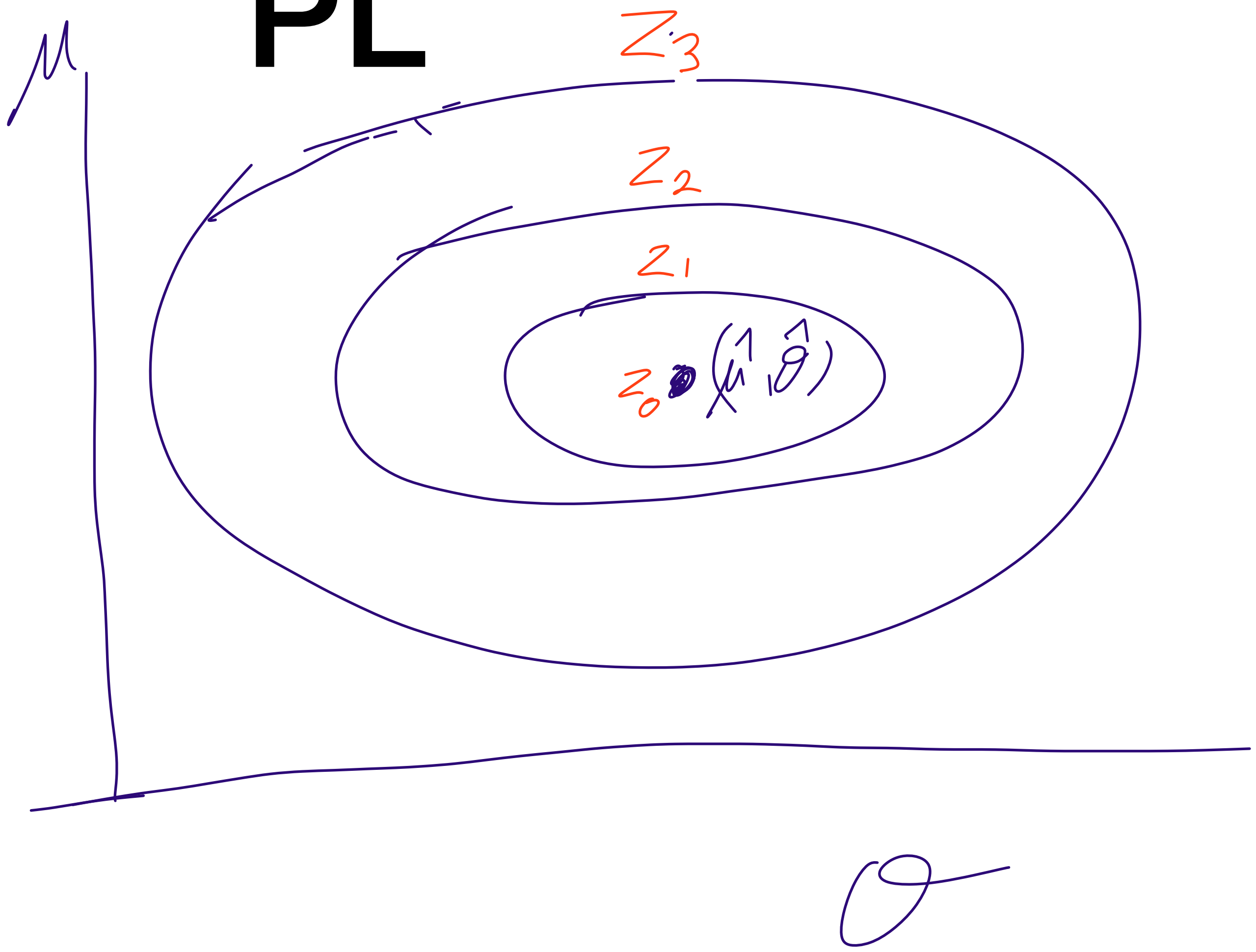
$$q = \frac{L(s+b(\theta))}{L(b(\theta))} \Rightarrow \frac{L(s+b(\hat{\hat{\theta}}_s))}{L(b(\hat{\hat{\theta}}_b))}$$

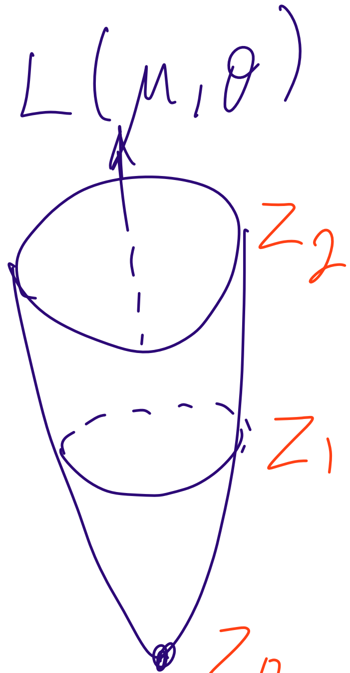
$\hat{\hat{\theta}}_s$ *is the MLE of* θ *fixing s*

PL



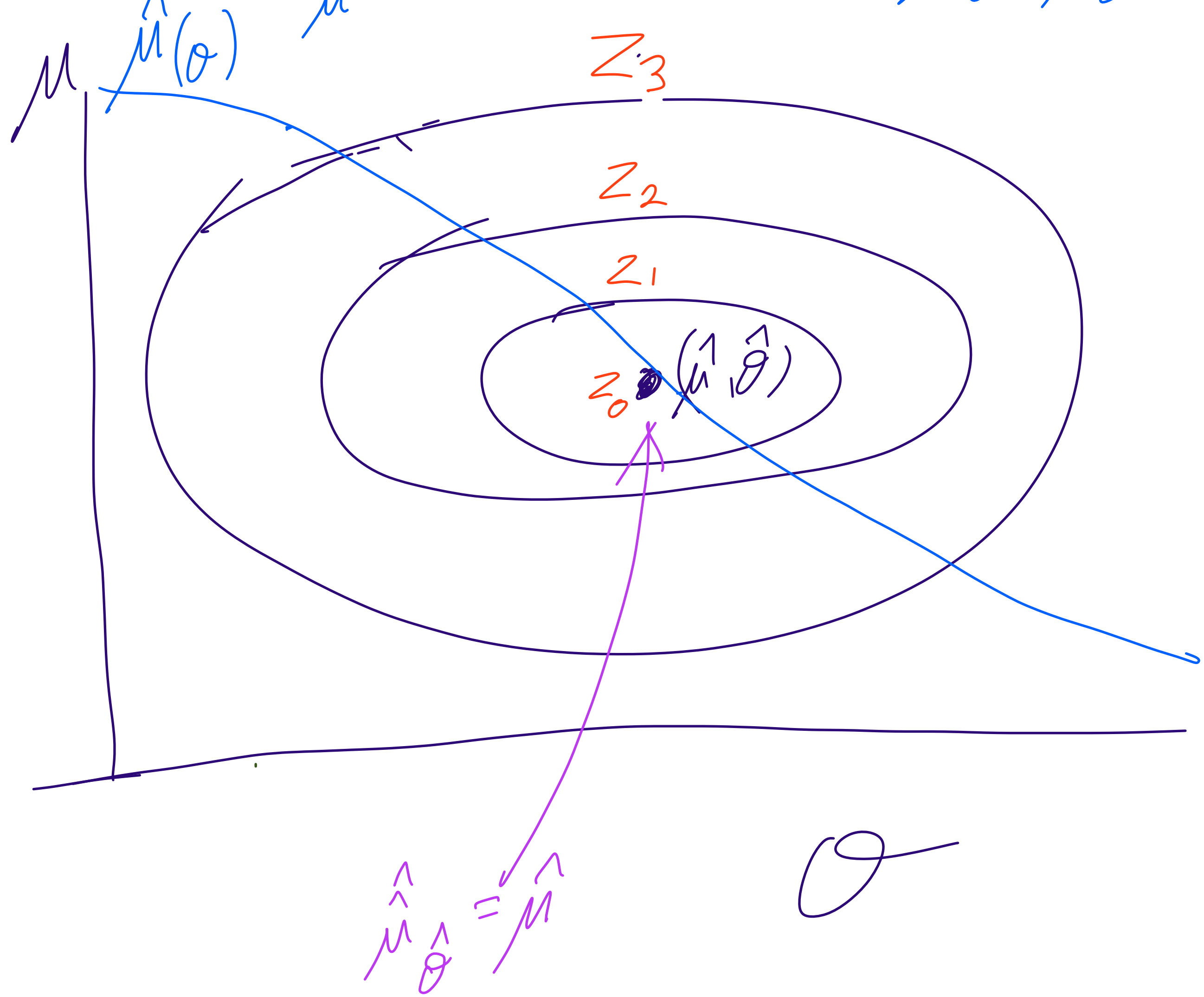
$$\min_{\mu,\theta} L = L(\hat{\mu}, \hat{\theta})$$



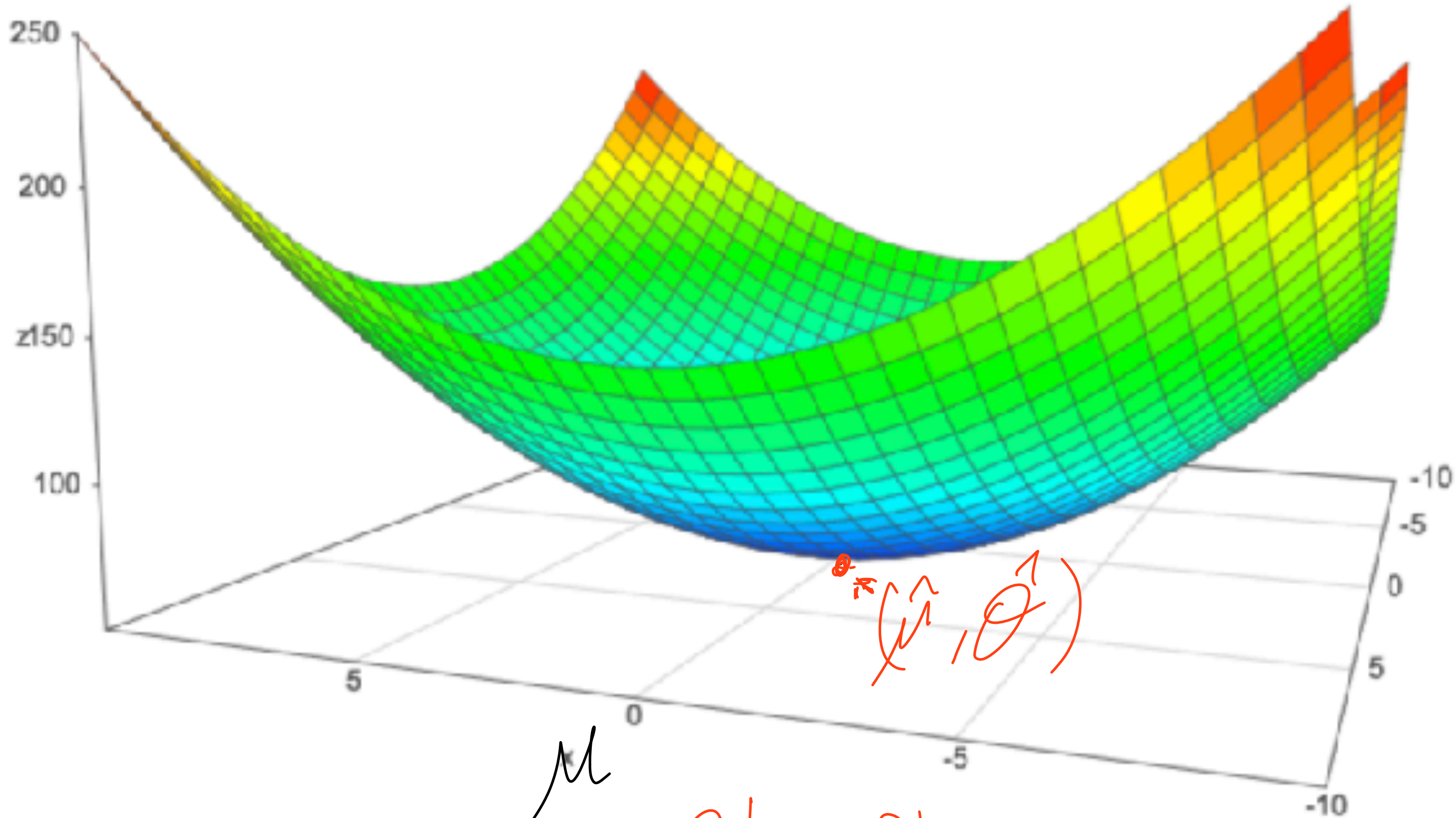


$$\min_{\mu,\theta} L = L(\hat{\mu}, \hat{\theta})$$

$$\min_{\mu} L = L(\hat{\hat{\mu}}, \theta) \qquad \hat{\hat{\mu}} = \hat{\mu}_\theta = \hat{\hat{\mu}}_\theta$$

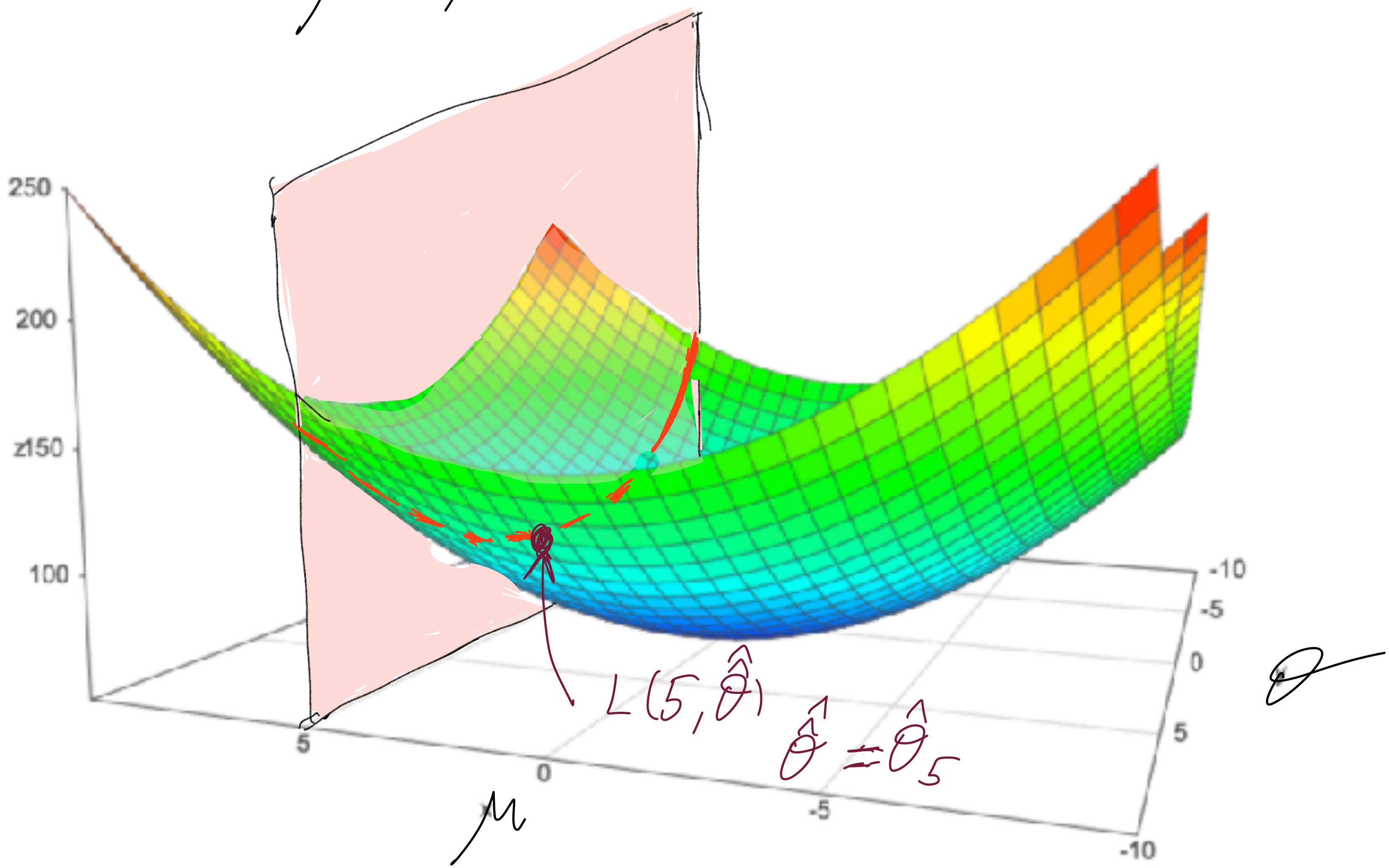


$$L = L(\mu, \theta)$$



$$\frac{\partial L}{\partial \mu} = \frac{\partial L}{\partial \theta} = 0 \Rightarrow \min L = L(\hat{\mu}, \hat{\theta})$$

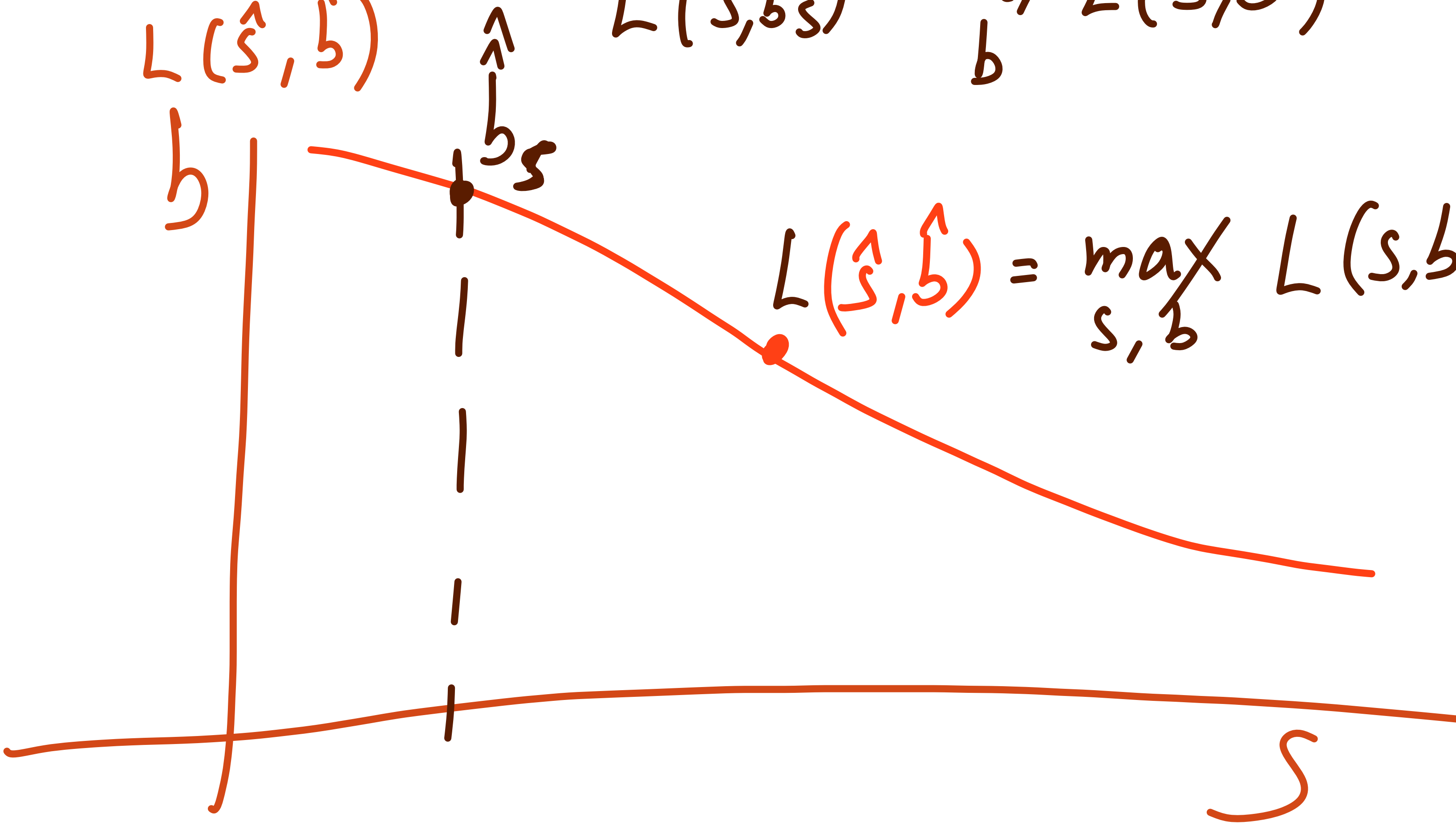
$L = L(\mu, \theta)$ fix $\mu = 5$



Profile Likelihood

$$q = \frac{L(s, \hat{\hat{b}})}{L(\hat{s}, \hat{b})}$$

$$L(s, \hat{\hat{b}}_s) = \max_{b} L(s, b)$$



$$L(\hat{s}, \hat{b}) = \max_{s,b} L(s, b)$$

Wilks theorem in the presence of NPs

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

$$\lambda = \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})}$$

$$q(\mu, \theta) = -2ln\lambda(\mu) \sim \chi^2$$

$$q(\mu_i, \theta) = -2ln\lambda(\mu_i) \sim \chi^2_n \quad Wilks\ Theorem$$

Profile Likelihood with Nuisance Parameters

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

$$q_\mu = -2\ln\frac{\max_b L(\mu s + b)}{\max_{\mu,b} L(\mu s + b)}$$

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

$\hat{\mu}$ MLE of μ

$\hat{b}$ MLE of b

$\hat{\hat{b}}_\mu$ MLE of b fixing μ

$\hat{\hat{\theta}}_\mu$ MLE of θ fixing μ

This Lecture's Questions

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

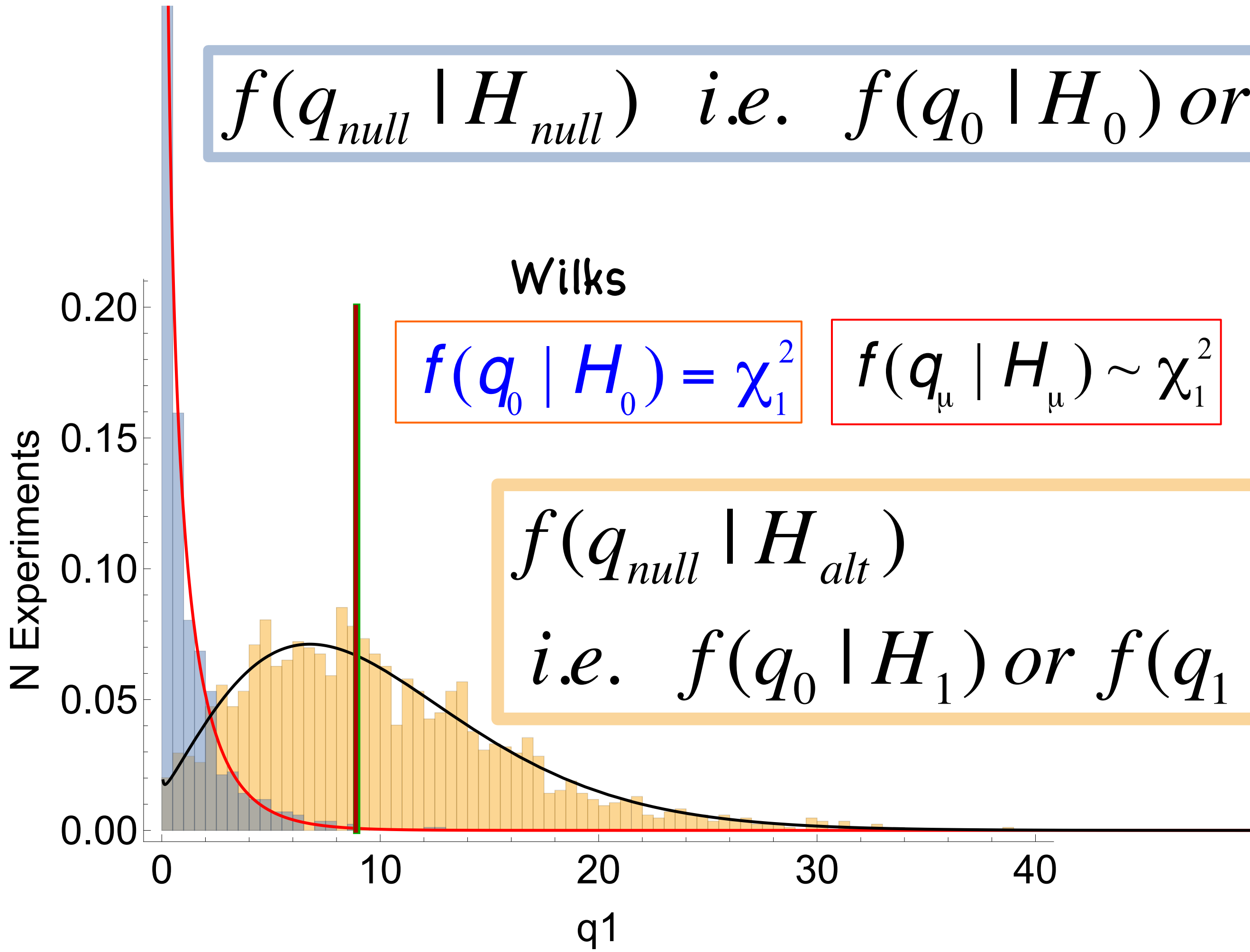
Wilks

$$f(q_0 \mid H_0) = \chi^2_1$$

$$f(q_\mu \mid H_\mu) \sim \chi^2_1$$

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

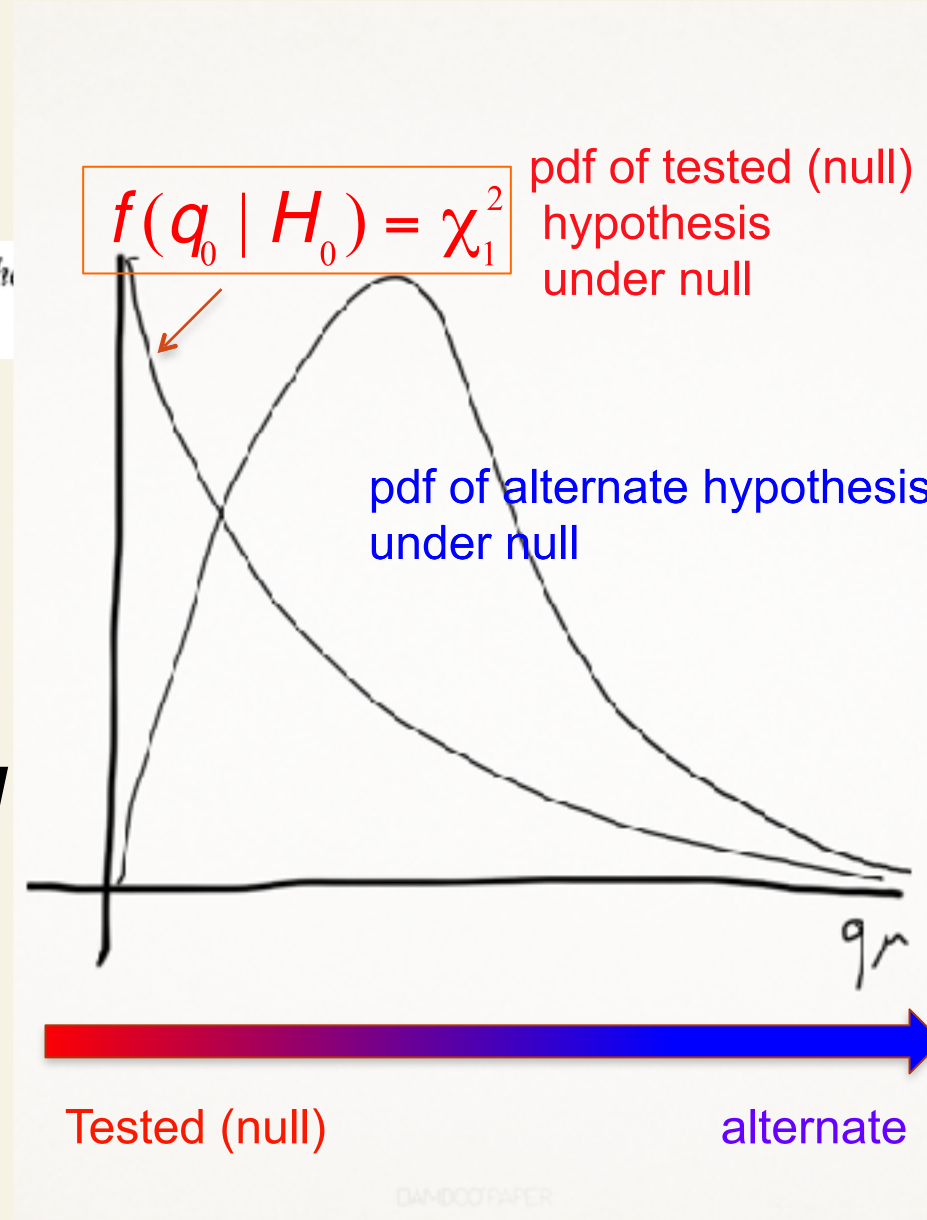
$$i.e. \quad f(q_0 \mid H_1) \; or \; f(q_1 \mid H_0)$$



Wilks Theorem

S.S. Wilks, *The large-sample distribution of th*
Ann. Math. Statist. 9 (1938) 60-2.

- *Wilks' theorem says that* the pdf of the statistic under the null hypothesis approaches a chi-square PDF for one degree of freedom



$$f(q_0 \mid H_0) = \chi_1^2$$

$$f(q_\mu \mid H_\mu) \sim \chi_1^2$$

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

Discovery vs Exclusion

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

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

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

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

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

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* $H_{\mu'}$ $(\mu'\neq\mu)$

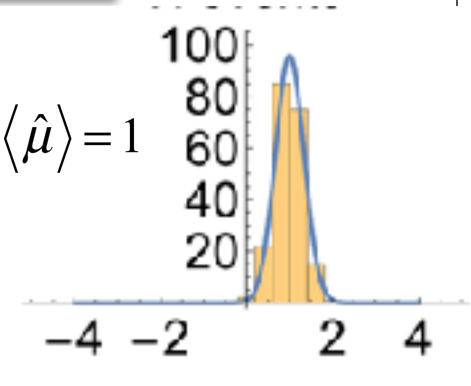
A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*, Transactions of the American Mathematical Society, Vol. **54**, No. 3 (Nov., 1943), pp. 426-482.

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

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

N is the sample size

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



(*Use the Asimov Dataset to estimate* σ)

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

with non-centrality parameter $\Lambda=\dfrac{(\mu-\mu')^2}{\sigma^2}$ 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

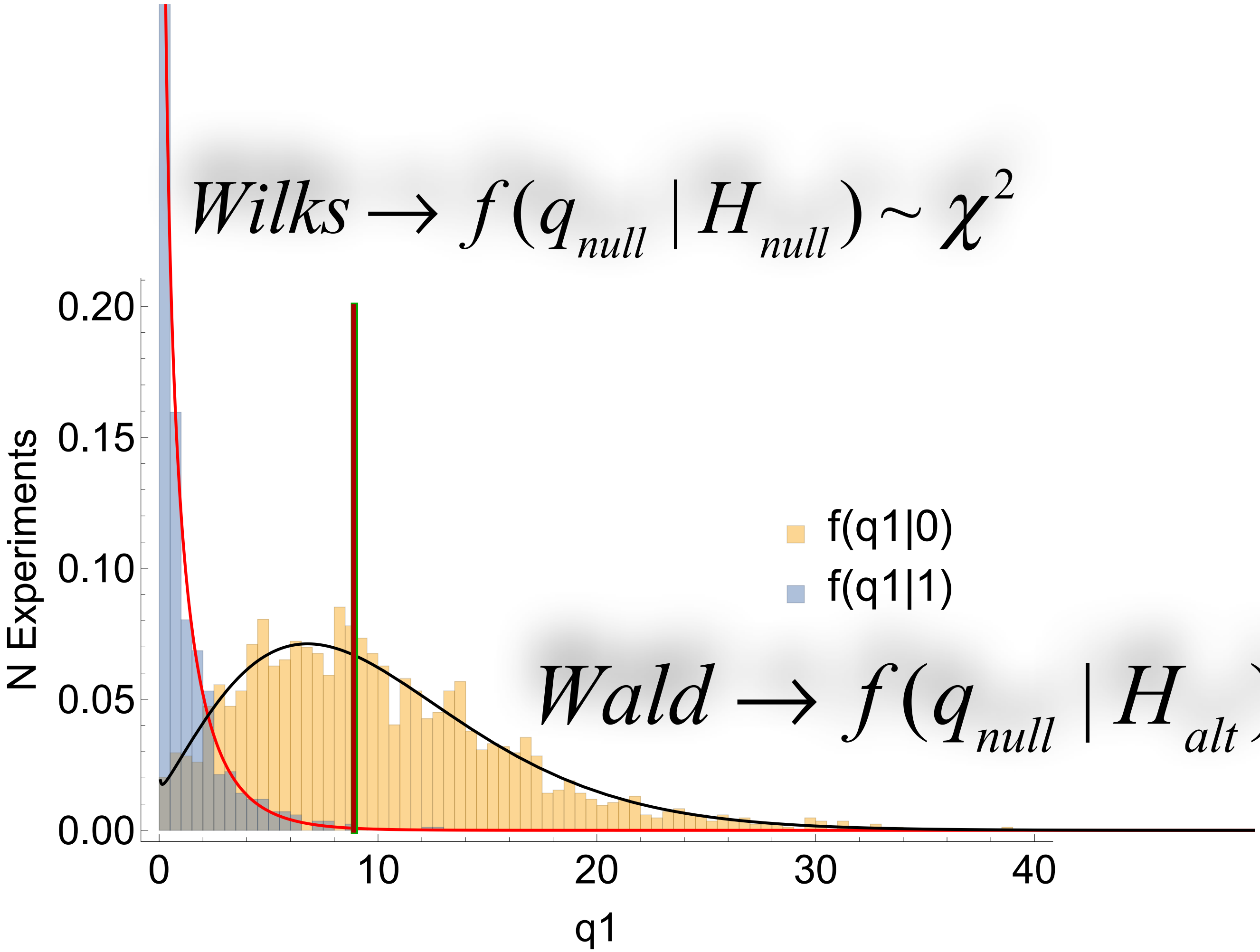
$$f(t_\mu;\Lambda) = \frac{1}{2\sqrt{t_\mu}}\frac{1}{\sqrt{2\pi}}\left[\exp\left(-\frac{1}{2}\left(\sqrt{t_\mu}+\sqrt{\Lambda}\right)^2\right)+\exp\left(-\frac{1}{2}\left(\sqrt{t_\mu}-\sqrt{\Lambda}\right)^2\right)\right]$$

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

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

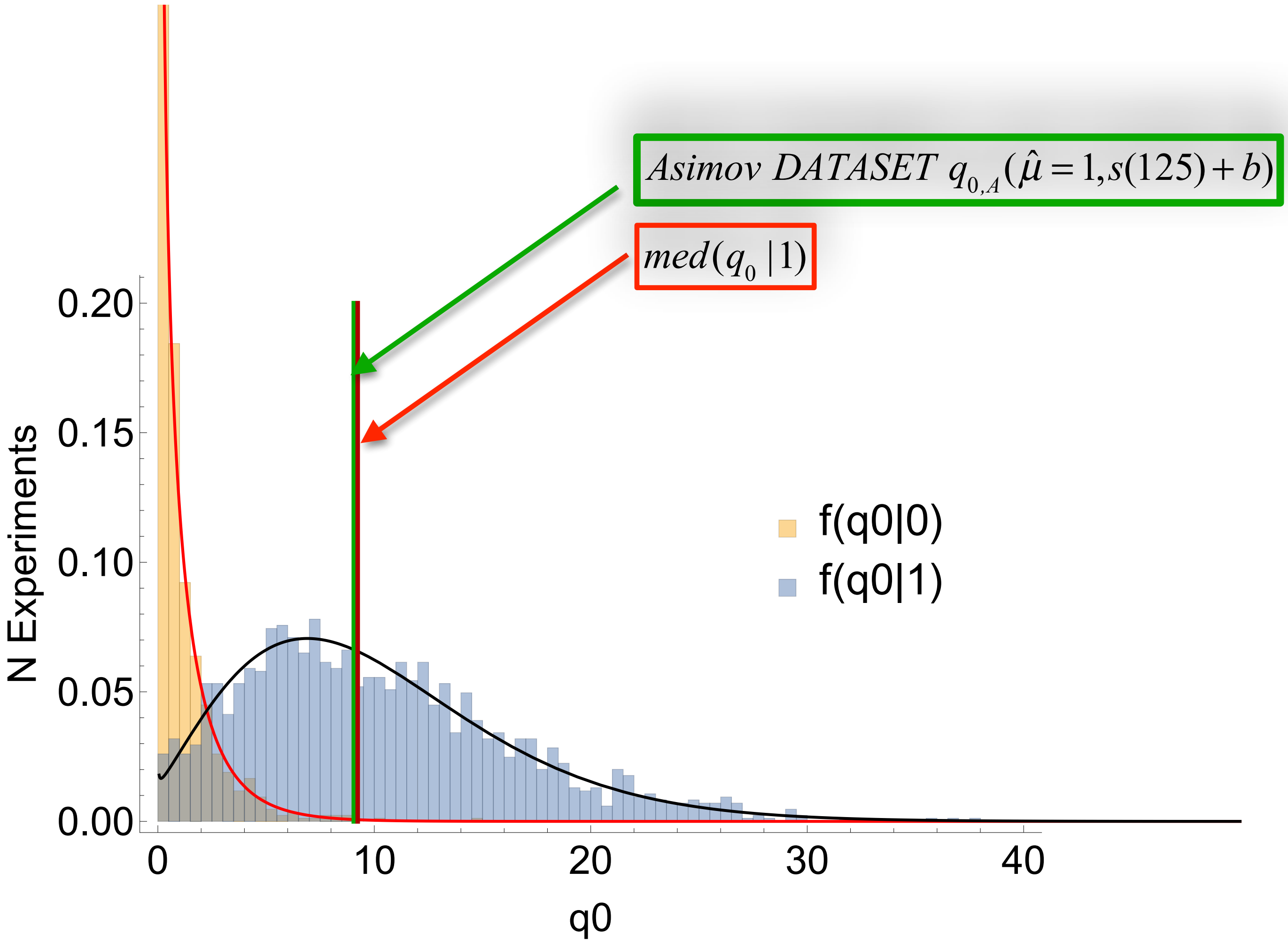
Asymptotics

$$Wilks \rightarrow f(q_{null} \mid H_{null}) \sim \chi^2$$



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

The Magic of Asimov



q_μ for exclusion

CCGV

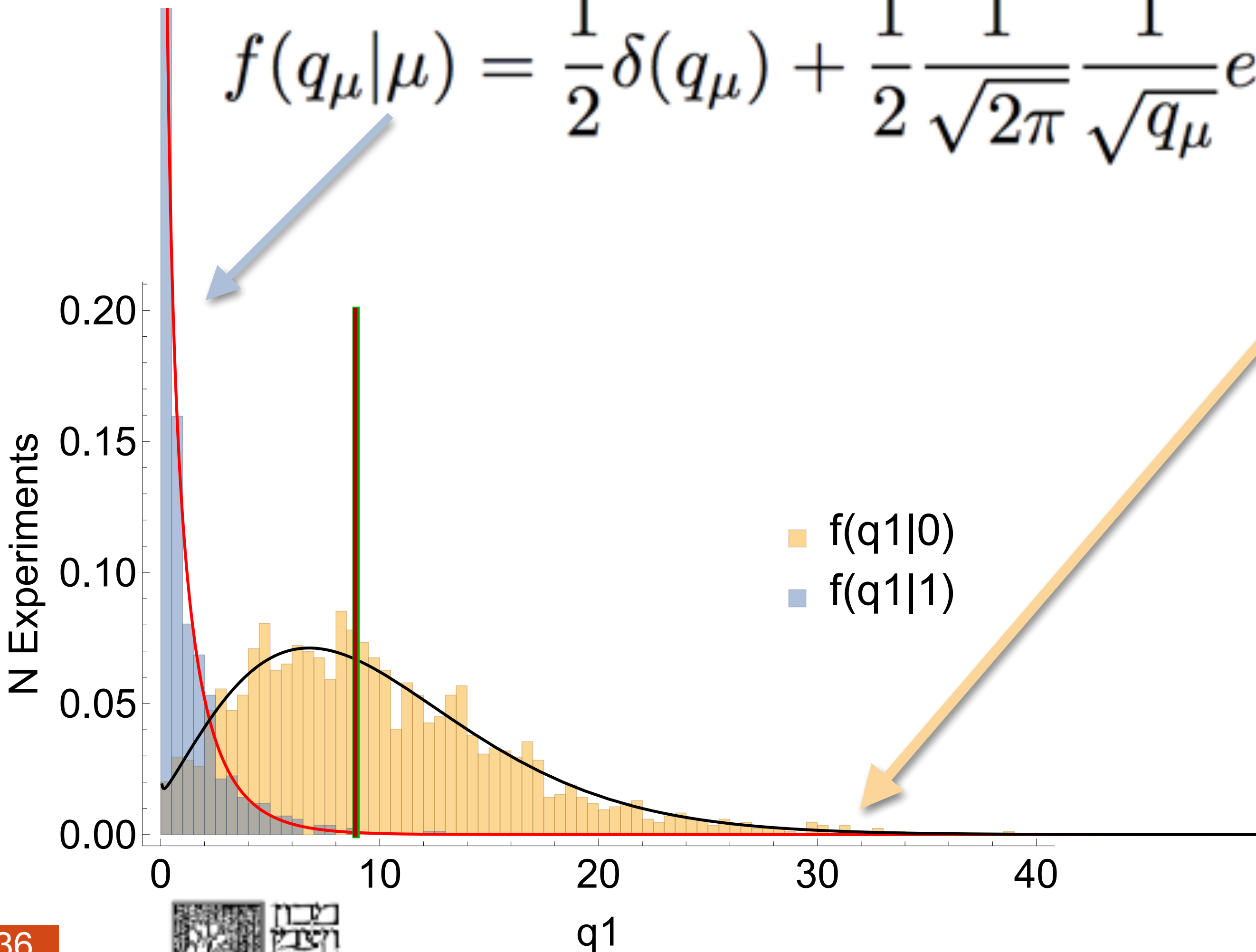
$$q_\mu = \begin{cases} -2\log\lambda(\mu) & \hat{\mu} \le \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

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

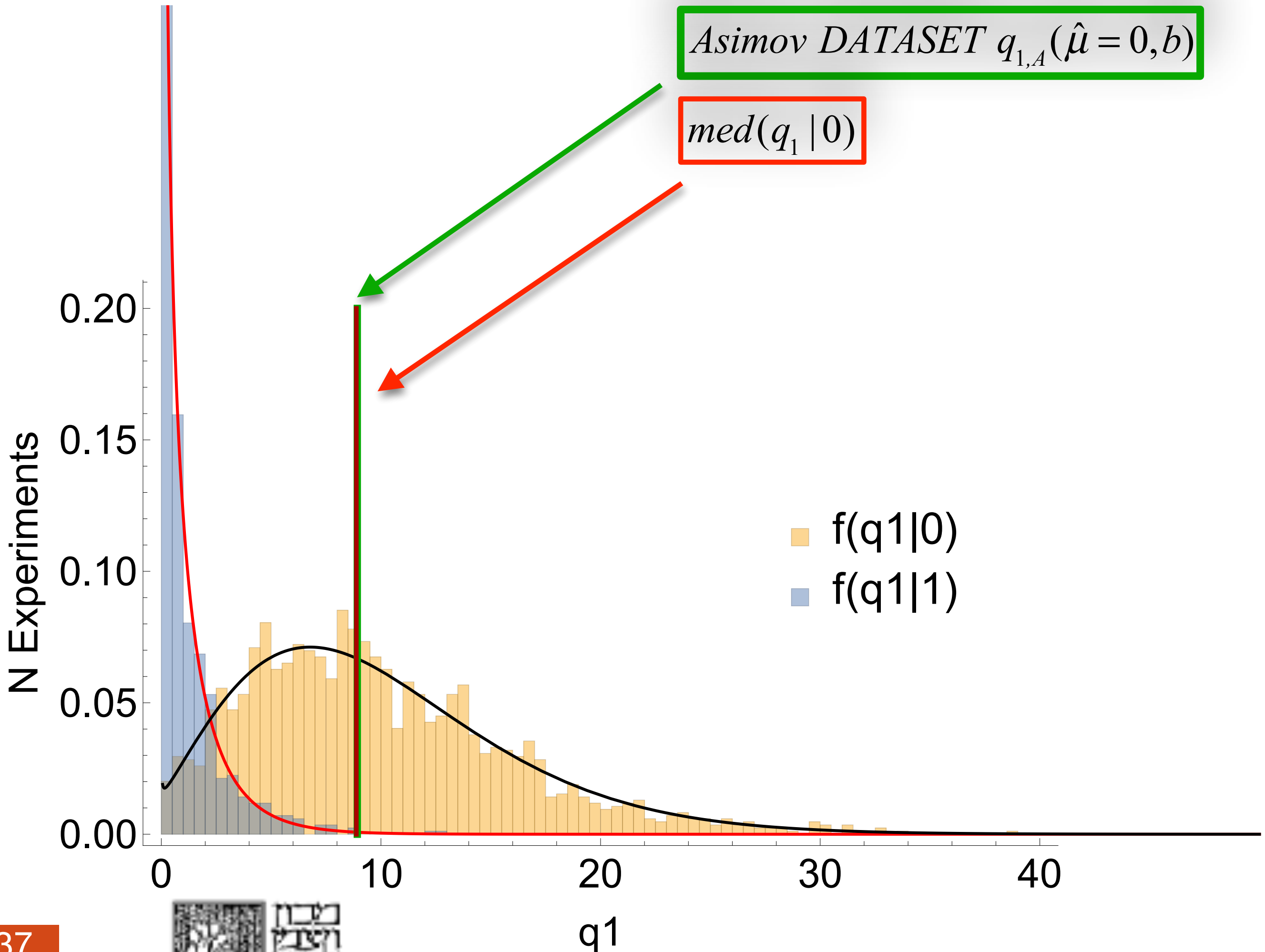
PDF of (q1|1) and (q1|1)

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

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

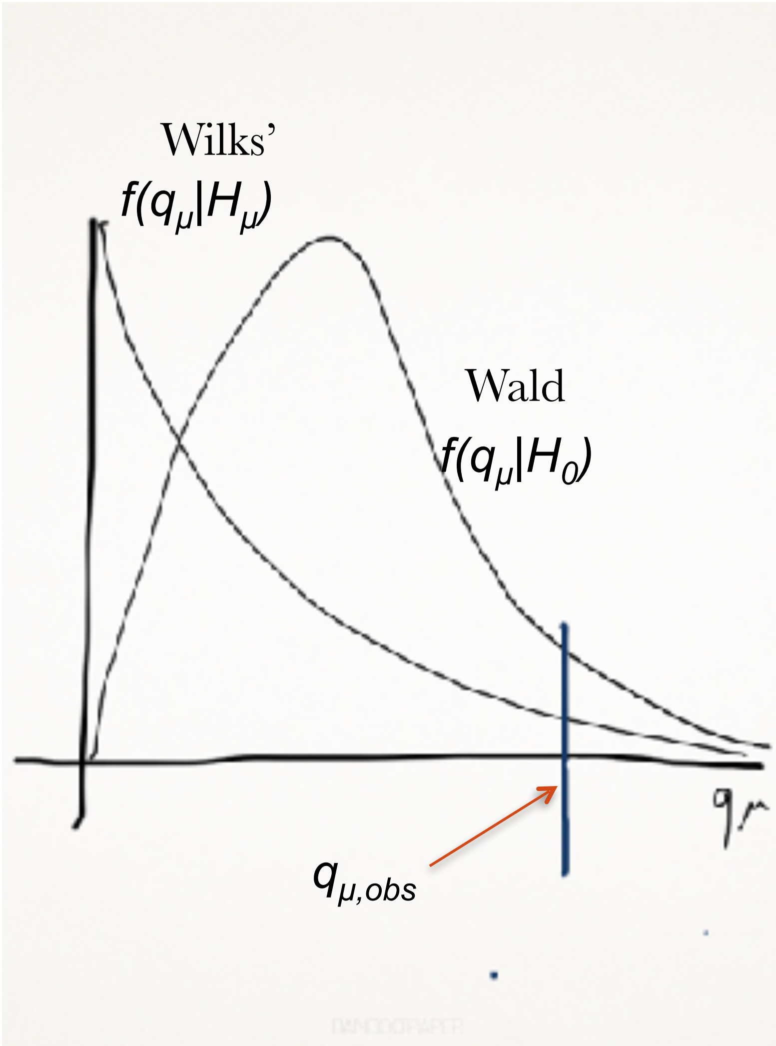


PDF of (q1|1) and (q1|1)



Exclusion at 95% CL

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

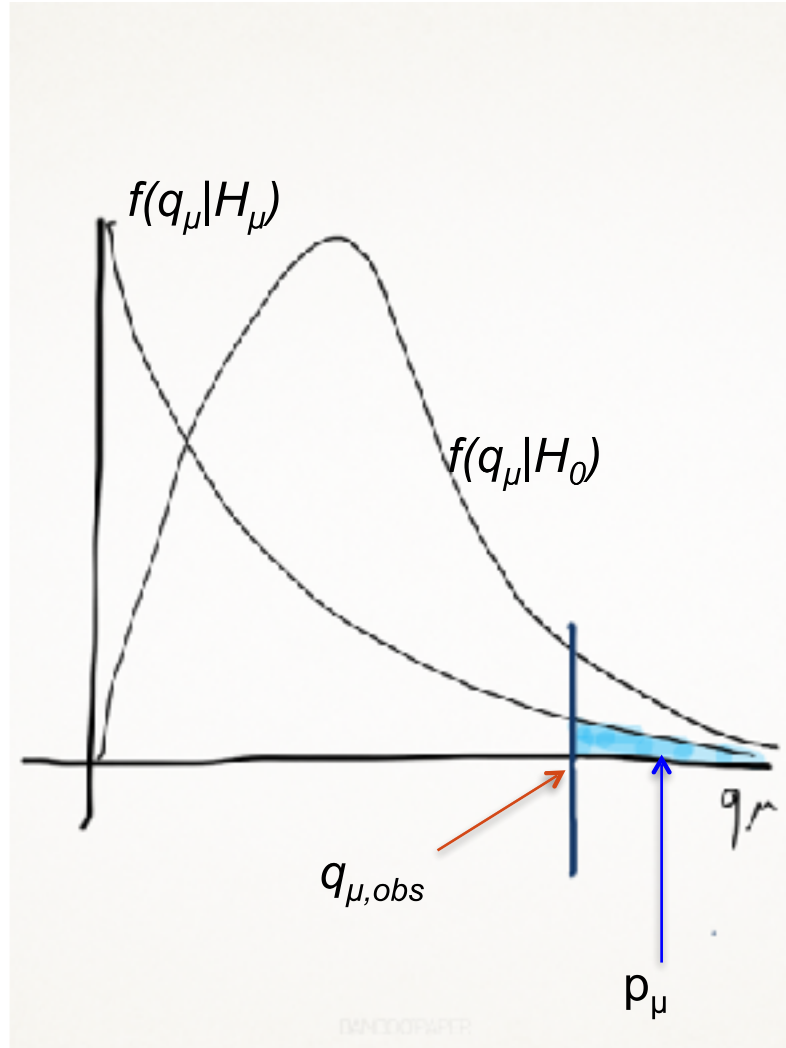


Exclusion at the 95% CL

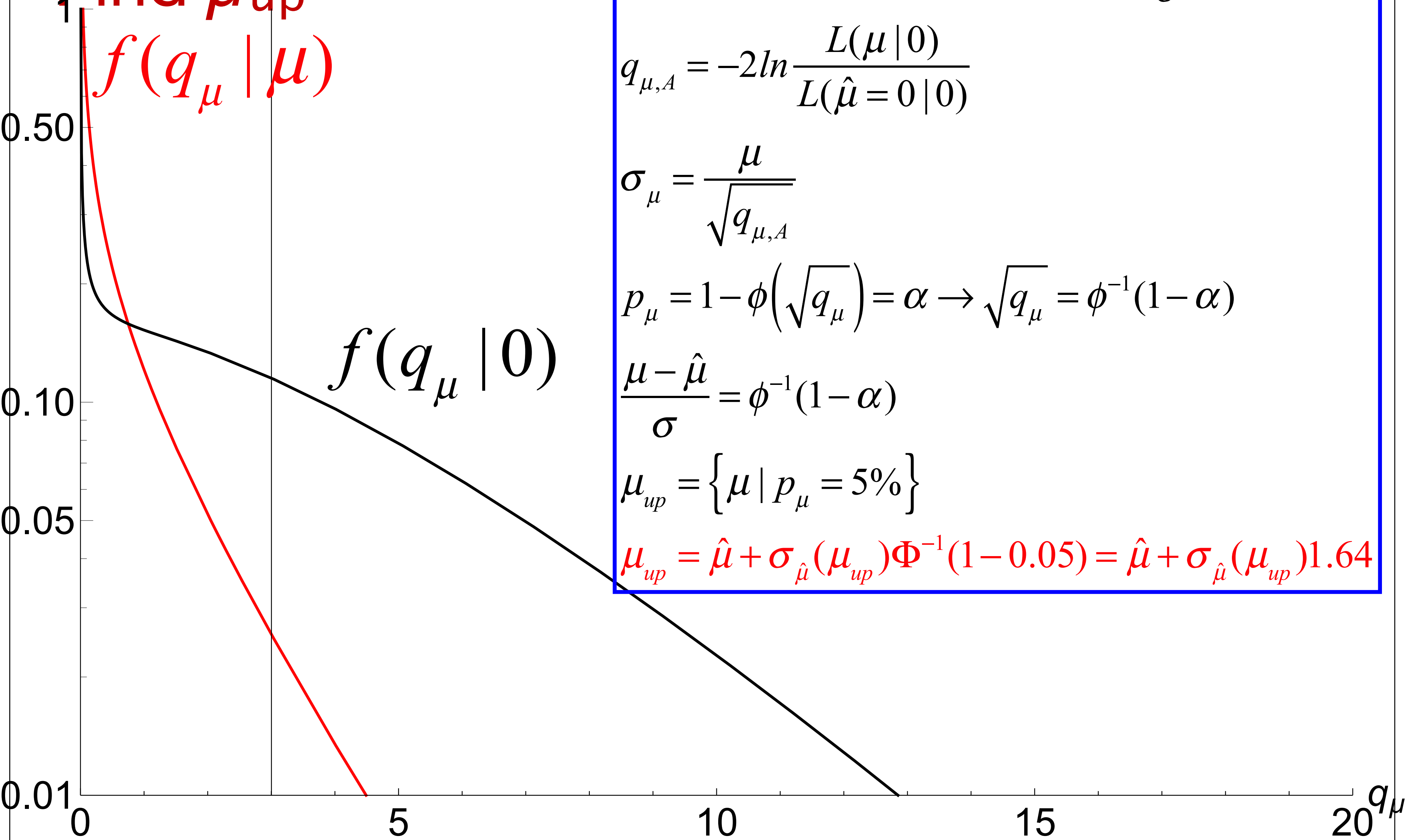
- Find the p-value of the signal hypothesis H_μ

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

- 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}



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

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

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

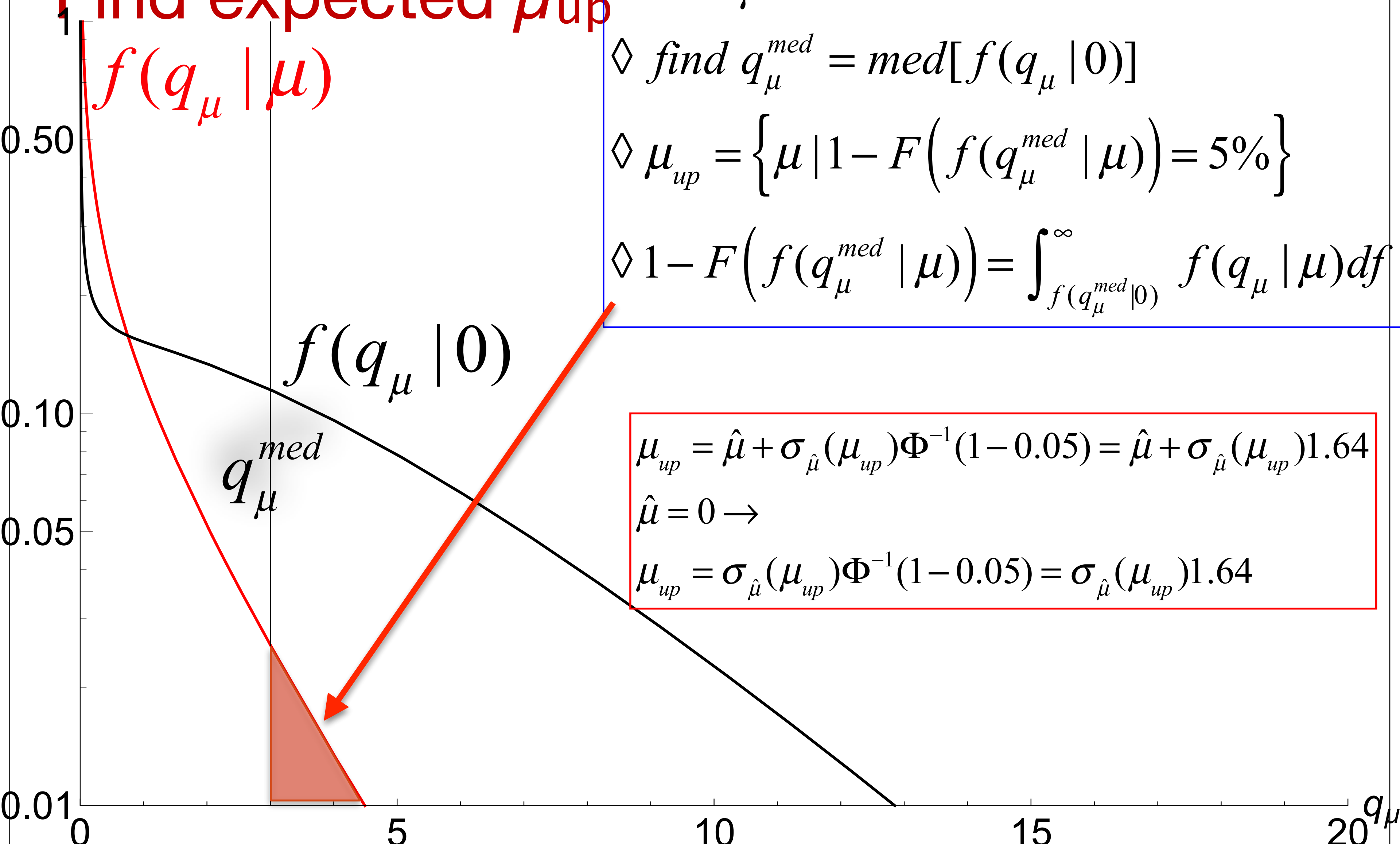
$$p_\mu = 1 - \phi\left(\sqrt{q_\mu}\right) = \alpha \rightarrow \sqrt{q_\mu} = \phi^{-1}(1-\alpha)$$

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

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

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

Find expected μ_{up}



- $\Diamond$ *scan* μ
- $\Diamond$ *find* $q_\mu^{med} = med[f(q_\mu|0)]$
- $\Diamond$ $\mu_{up} = \left\{\mu \,|\, 1 - F\left(f(q_\mu^{med}|\mu)\right) = 5\%\right\}$
- $\Diamond$ $1 - F\left(f(q_\mu^{med}|\mu)\right) = \int_{f(q_\mu^{med}|0)}^{\infty} f(q_\mu|\mu)df$

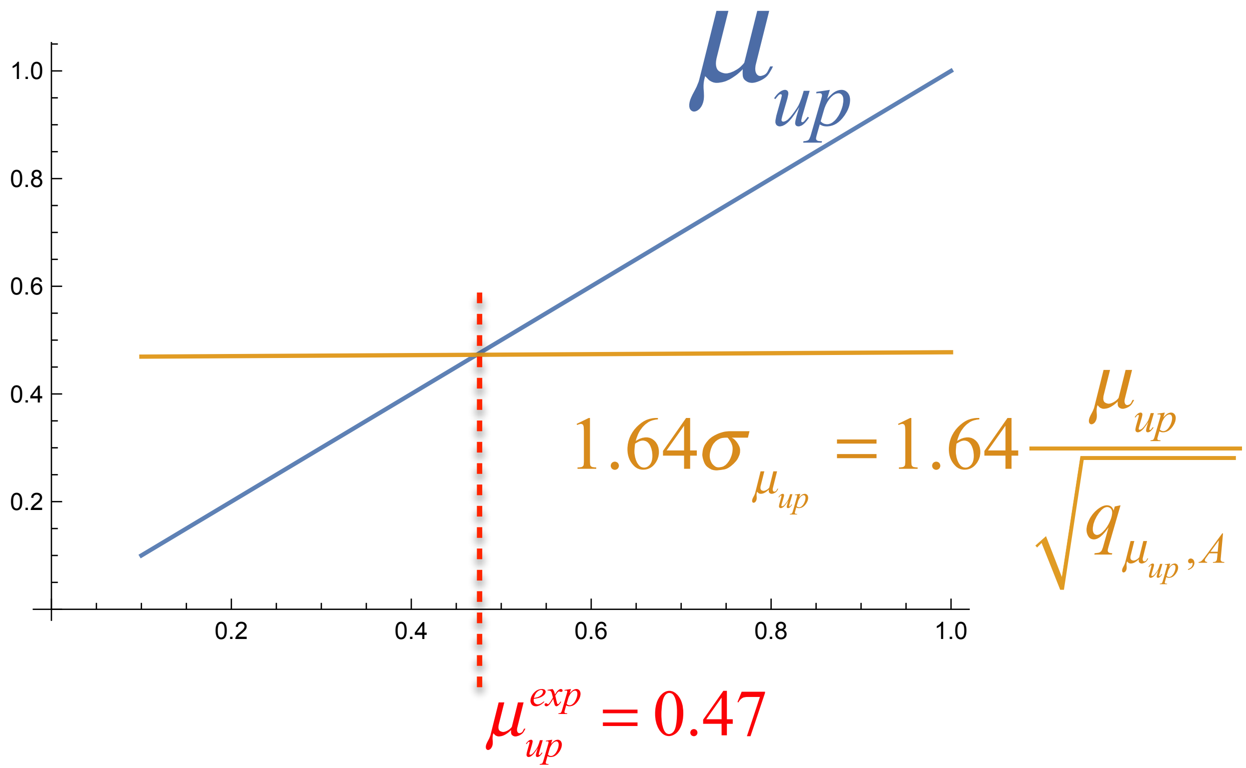
$$\mu_{up} = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up})\Phi^{-1}(1-0.05) = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up})1.64$$
$$\hat{\mu} = 0 \rightarrow$$
$$\mu_{up} = \sigma_{\hat{\mu}}(\mu_{up})\Phi^{-1}(1-0.05) = \sigma_{\hat{\mu}}(\mu_{up})1.64$$

Find expected μ_{up}

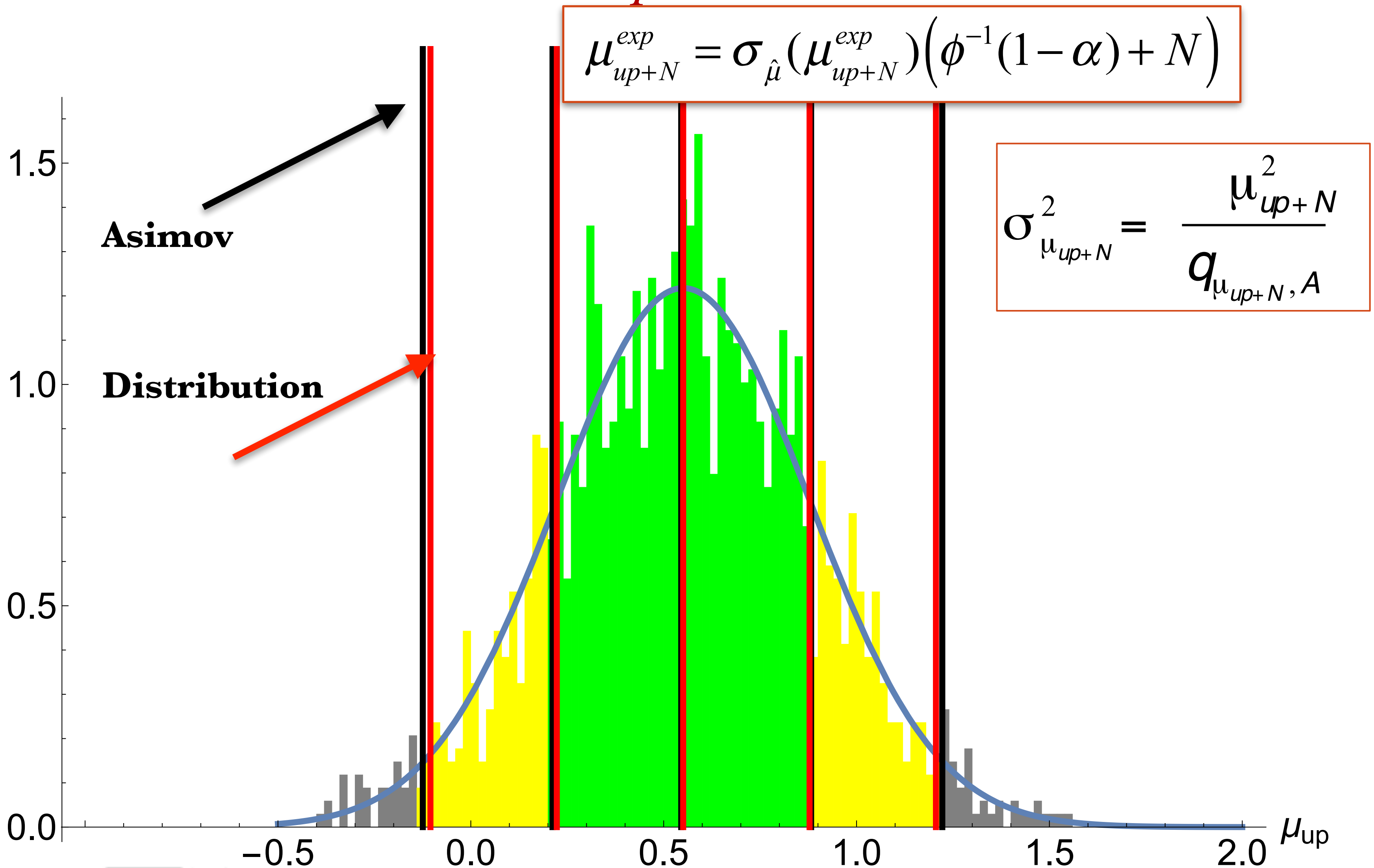
$$\mu_{up} = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up})\Phi^{-1}(1-0.05) = \hat{\mu} + \sigma_{\hat{\mu}}(\mu_{up})1.64$$

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

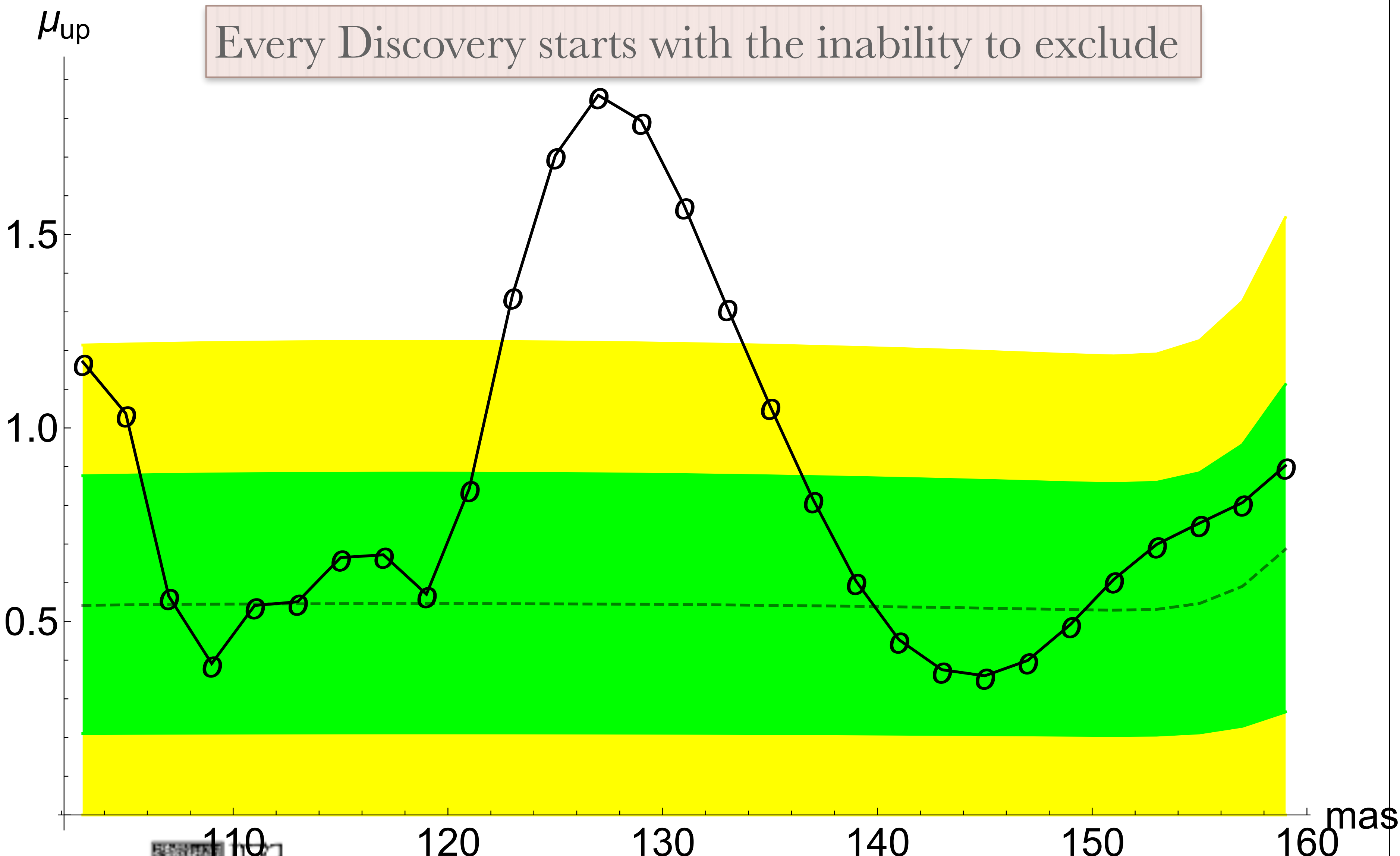
$$\mu_{up}^{exp} = \sigma_{\hat{\mu}}(\mu_{up})\Phi^{-1}(1-0.05) = \sigma_{\hat{\mu}}(\mu_{up}^{exp})1.64$$



Expected μ_{up} Bands at m=125



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

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

q_0 for discovery

CCGV

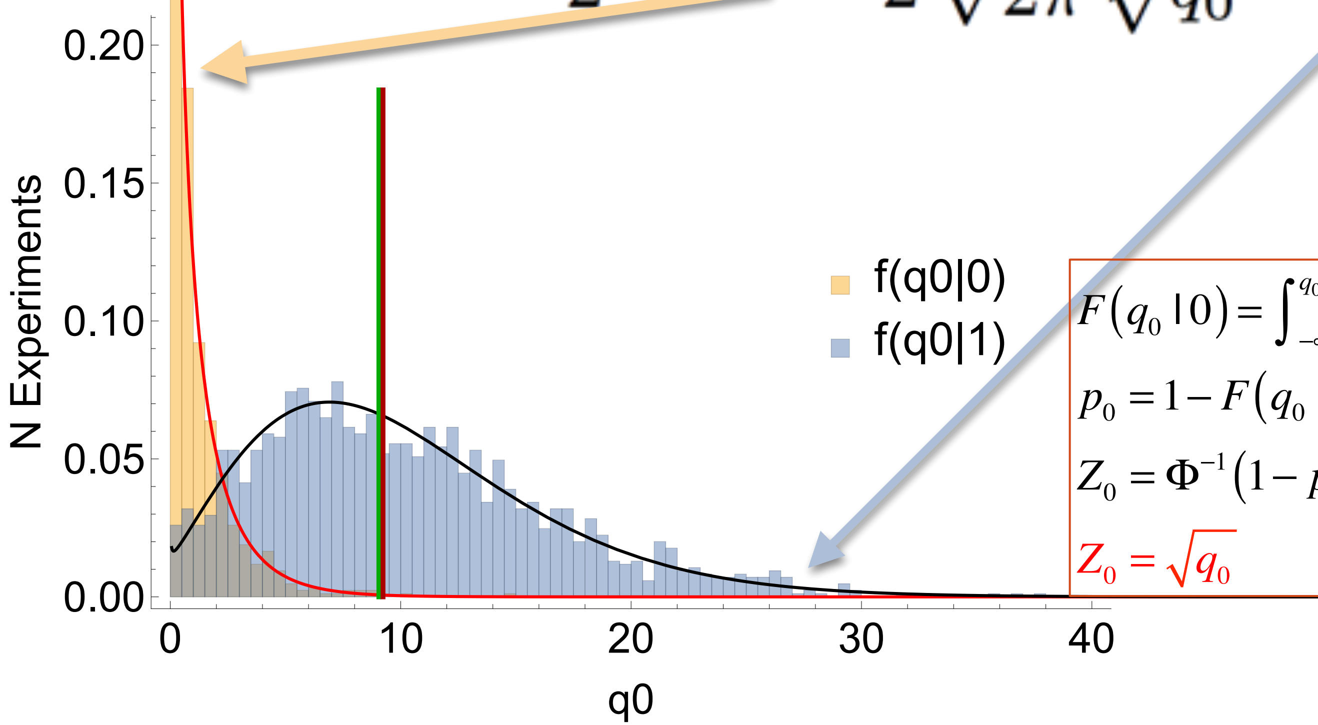
$$q_0 \equiv \tilde{t}_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}$$

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

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

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

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



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

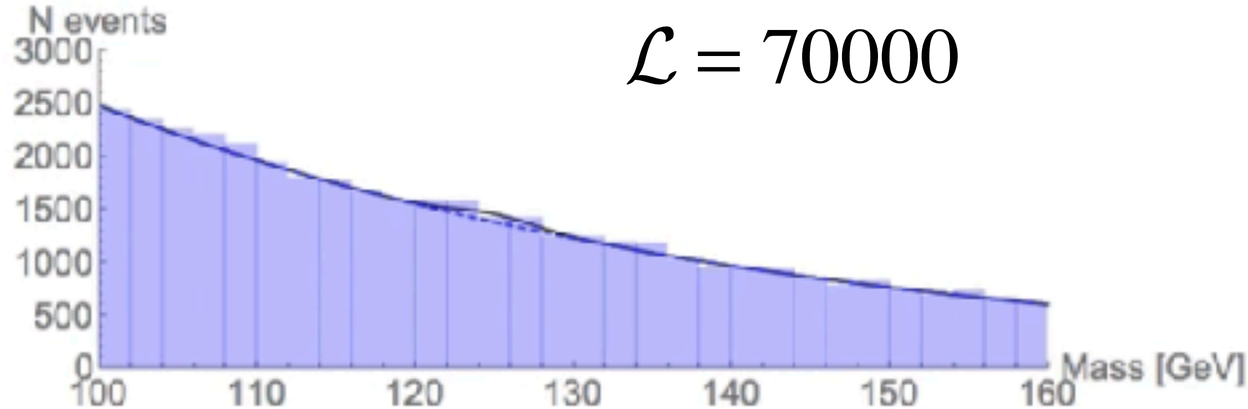
$$p_0 = 1 - F(q_0|0)$$

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

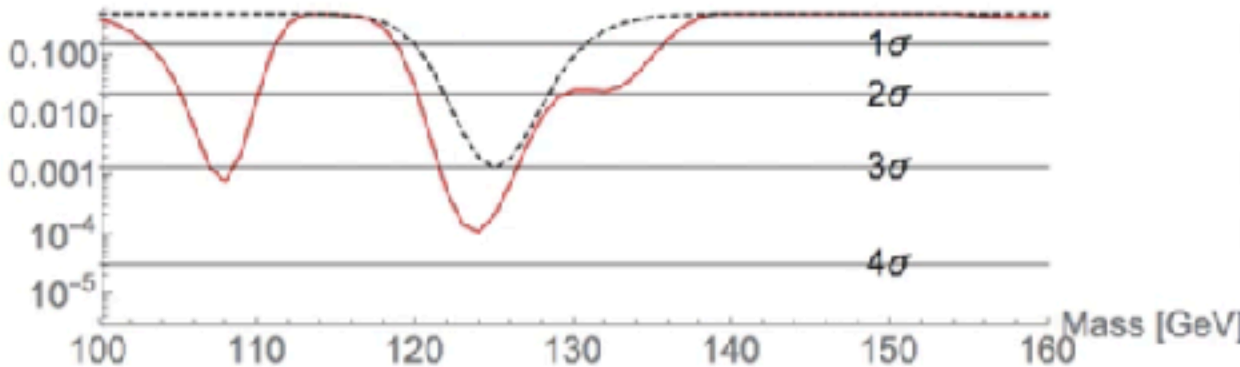
$$Z_0 = \sqrt{q_0}$$

p- value

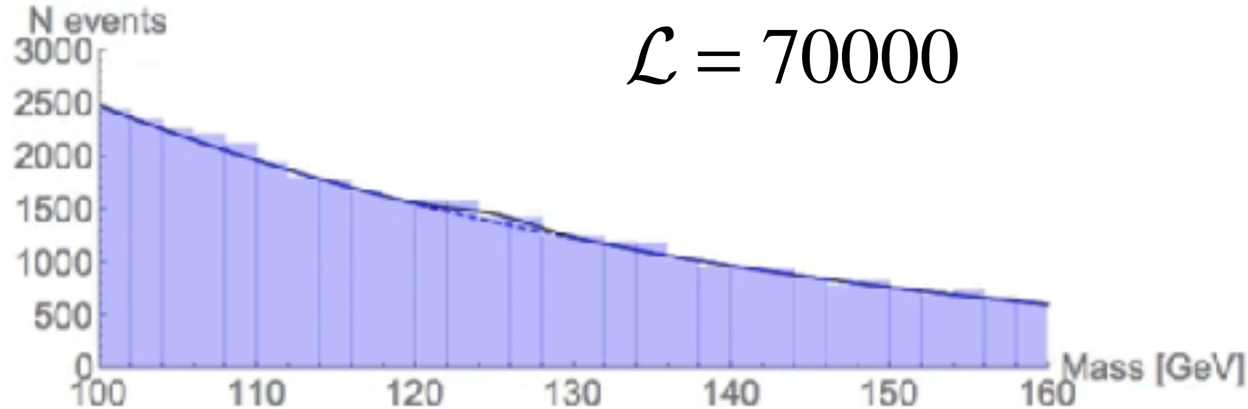
$$\mathcal{L} = 70000$$



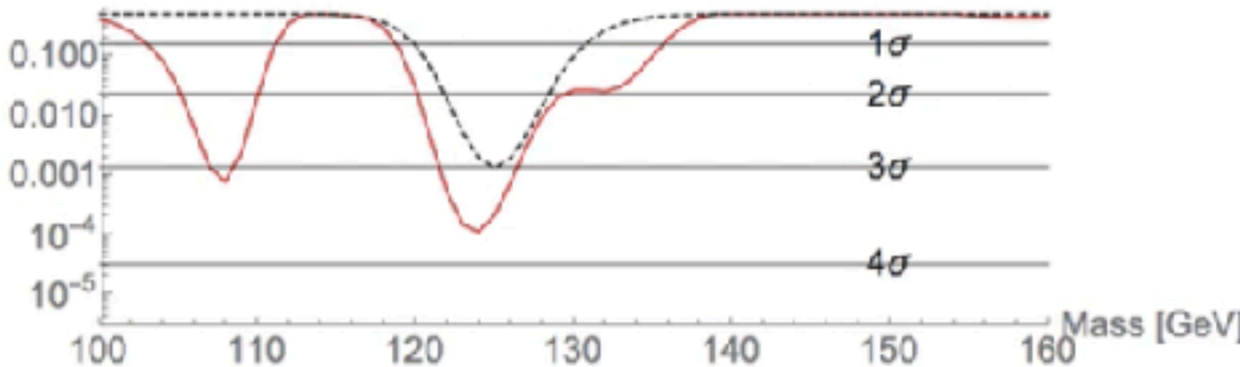
$$p = prob(q_0 \geq q_{0,obs})$$



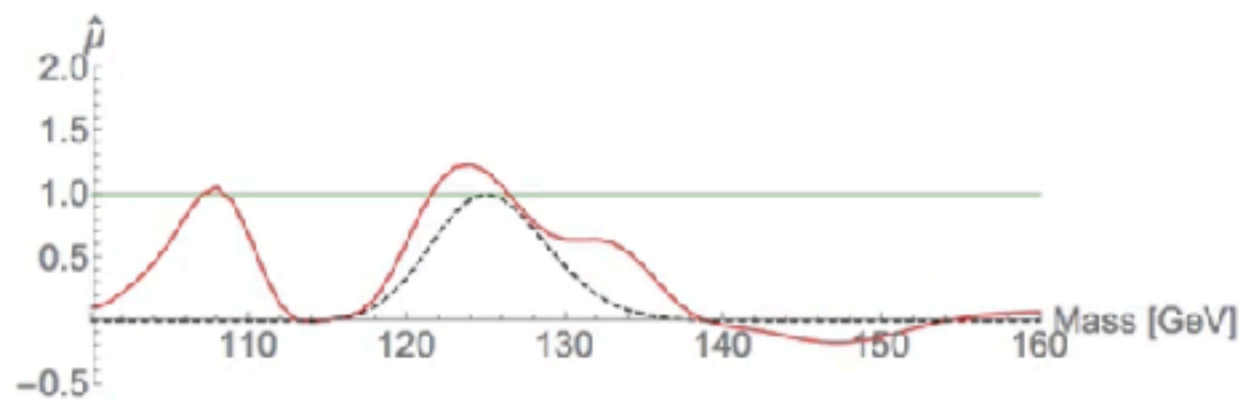
p- value



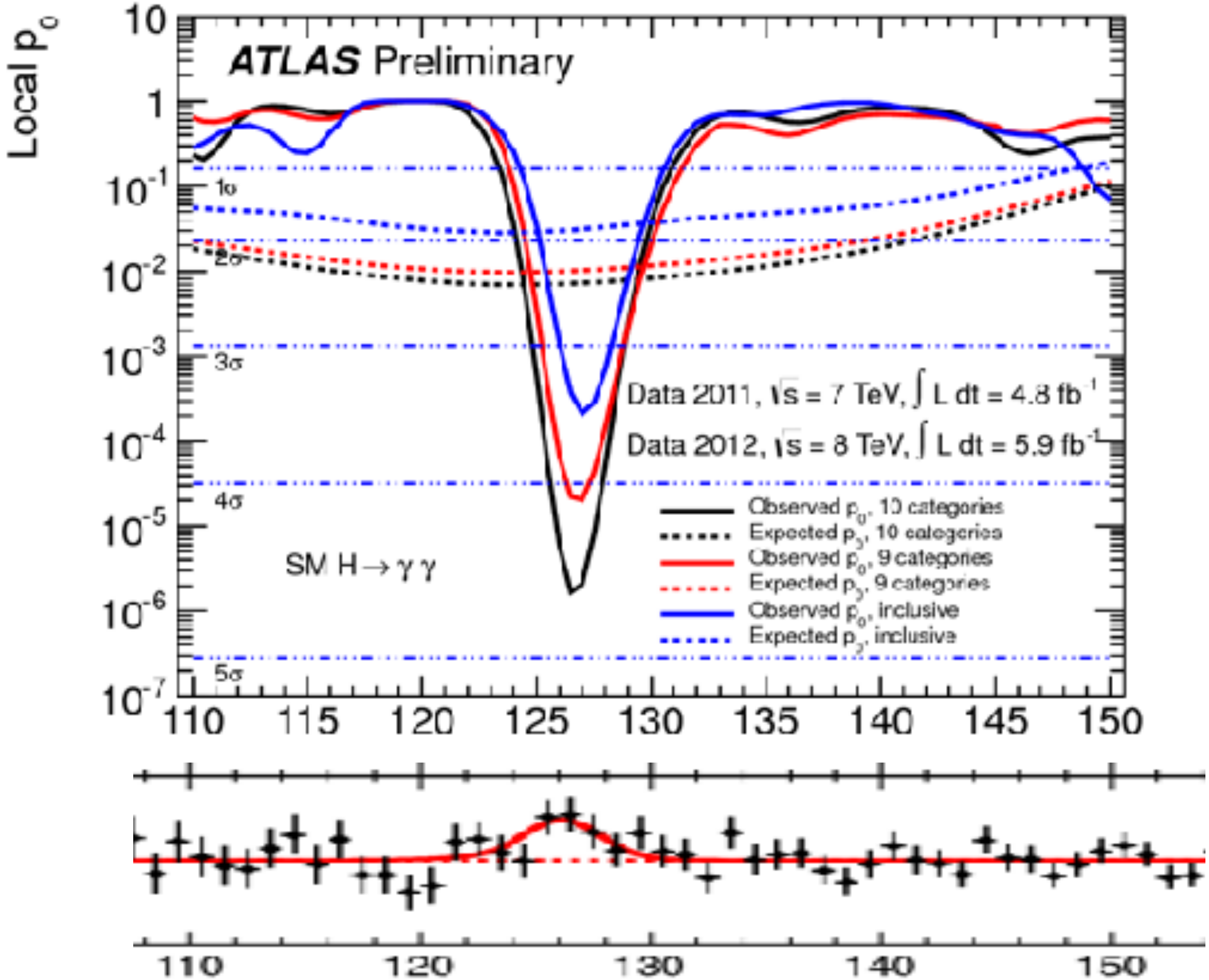
$$p = prob(q_0 \geq q_{0,obs})$$



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



H→γγ


ATLAS Preliminary
Local p_0
Data 2011, $\sqrt{s}$ = 7 TeV, ∫ L dt = 4.8 fb^{-1}
Data 2012, $\sqrt{s}$ = 8 TeV, ∫ L dt = 5.9 fb^{-1}
SM H→γγ
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
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

More Magic

The New $s/\sqrt{b}$

The new $s/\sqrt{b}$

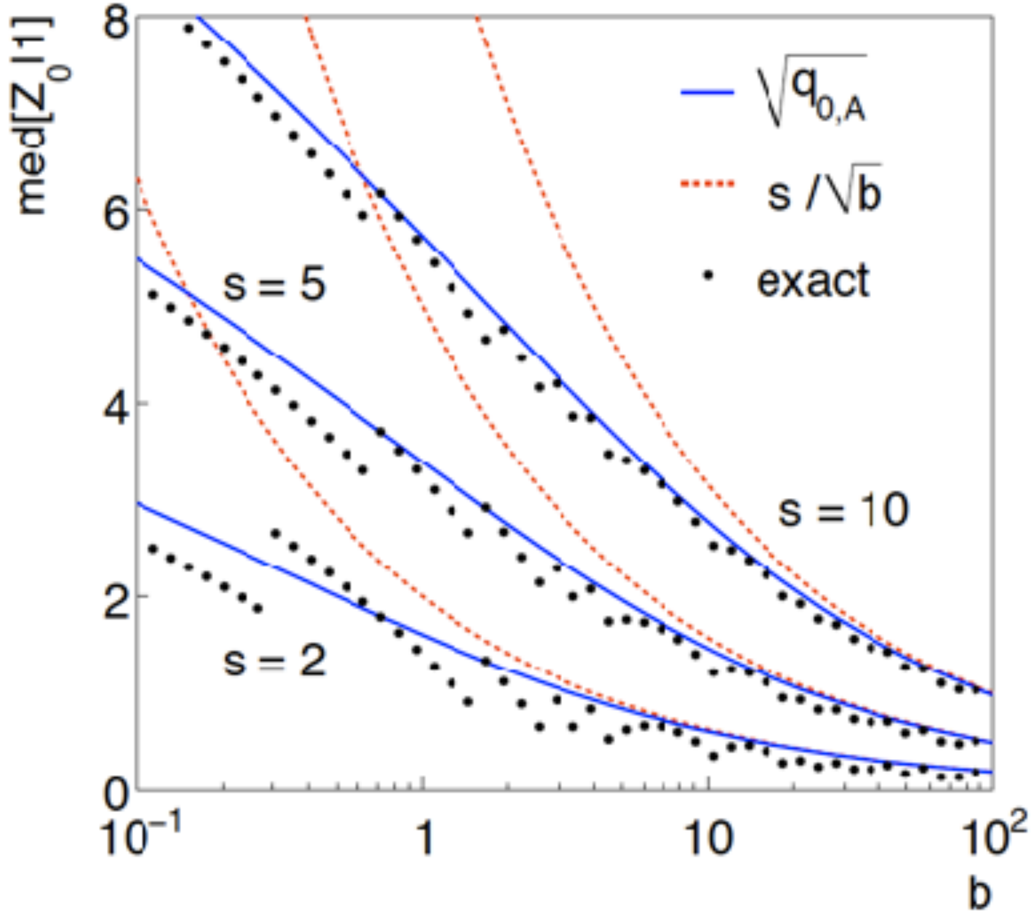
$$Z_A = \sqrt{q_{0,A}}$$

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

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

The New $s/\sqrt{b}$

$s/\sqrt{b}$?

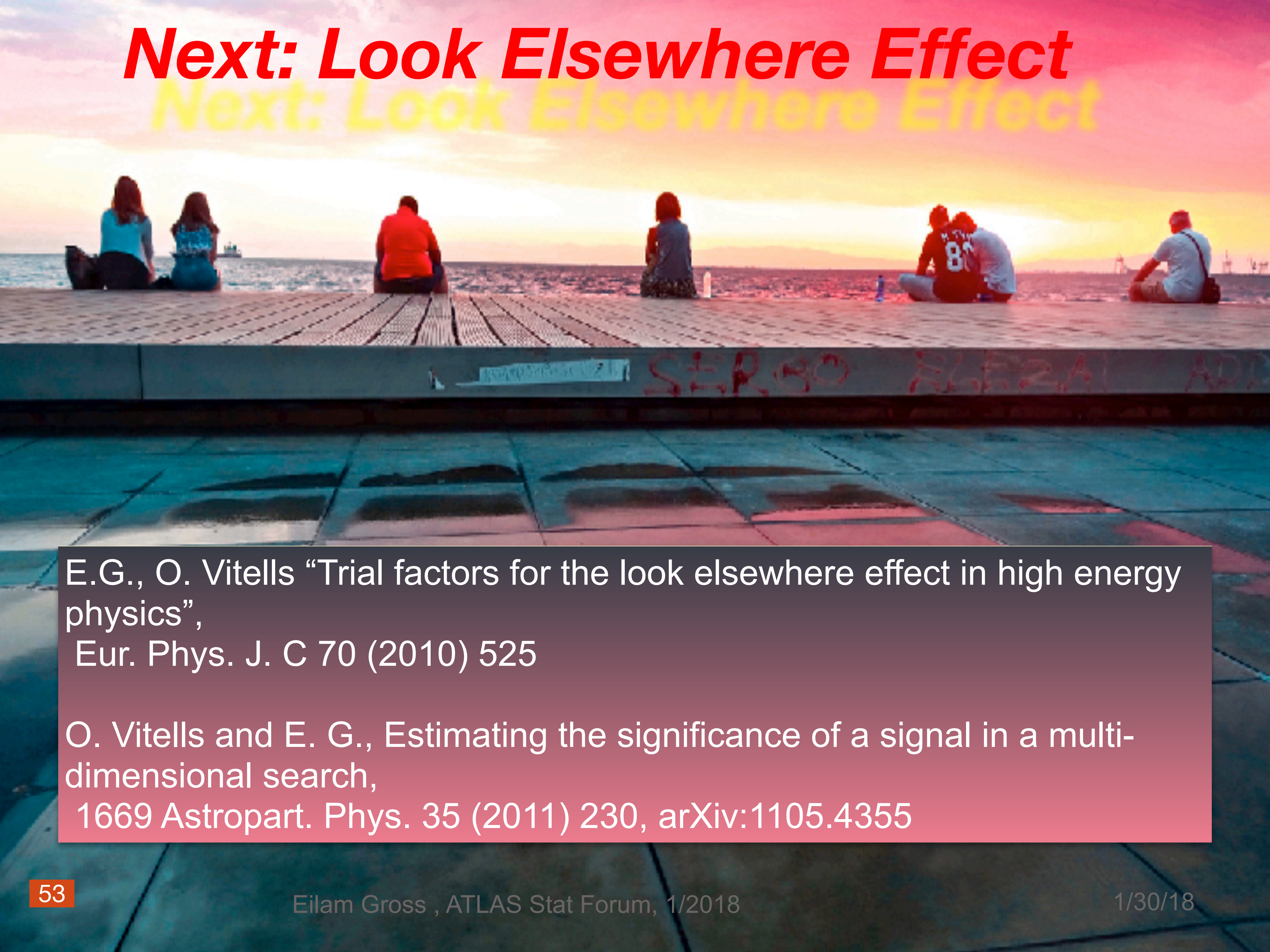


The new $s/\sqrt{b}$

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

Next: Look Elsewhere Effect

Next: Look Elsewhere Effect



E.G., O. Vitells “Trial factors for the look elsewhere effect in high energy physics”,
Eur. Phys. J. C 70 (2010) 525

O. Vitells and E. G., Estimating the significance of a signal in a multi-dimensional search,
1669 Astropart. Phys. 35 (2011) 230, arXiv:1105.4355