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

More Magic
(Asimov Significance)

The New $s/\sqrt{b}$

The new $s/\sqrt{b}$

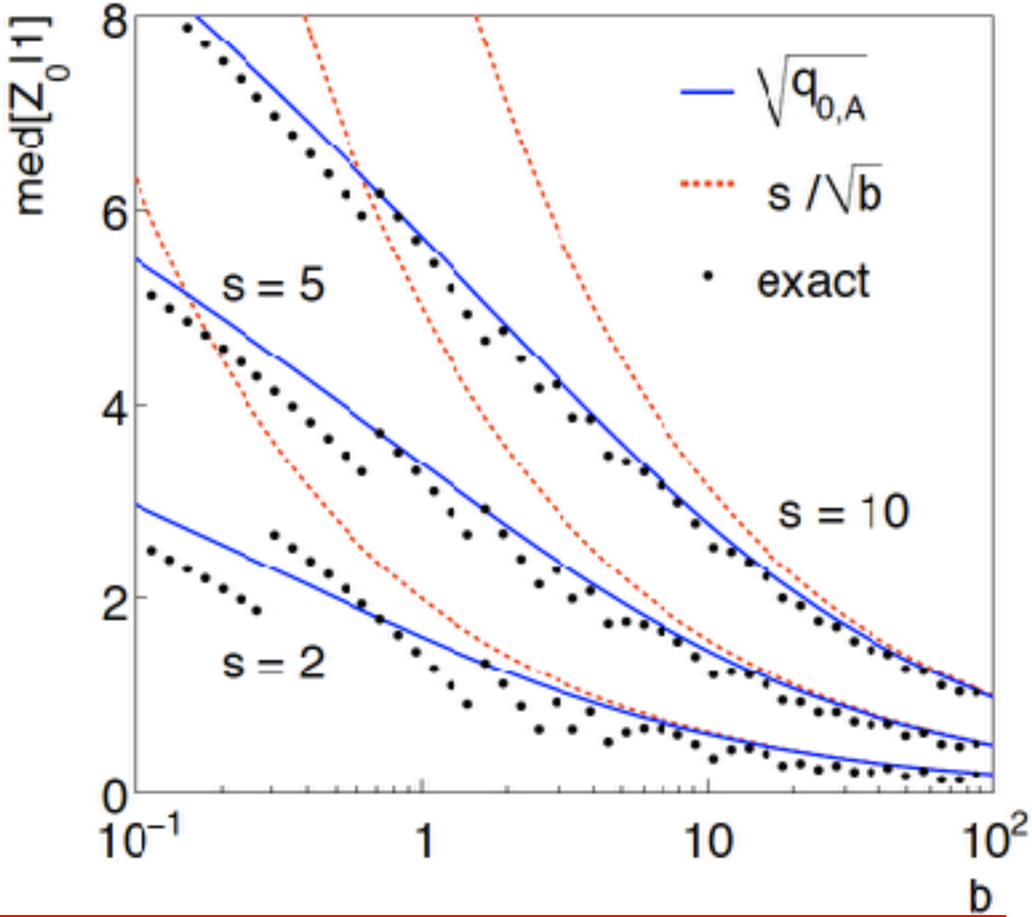
$$Z_A = \sqrt{q_{0,A}}$$

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

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

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)

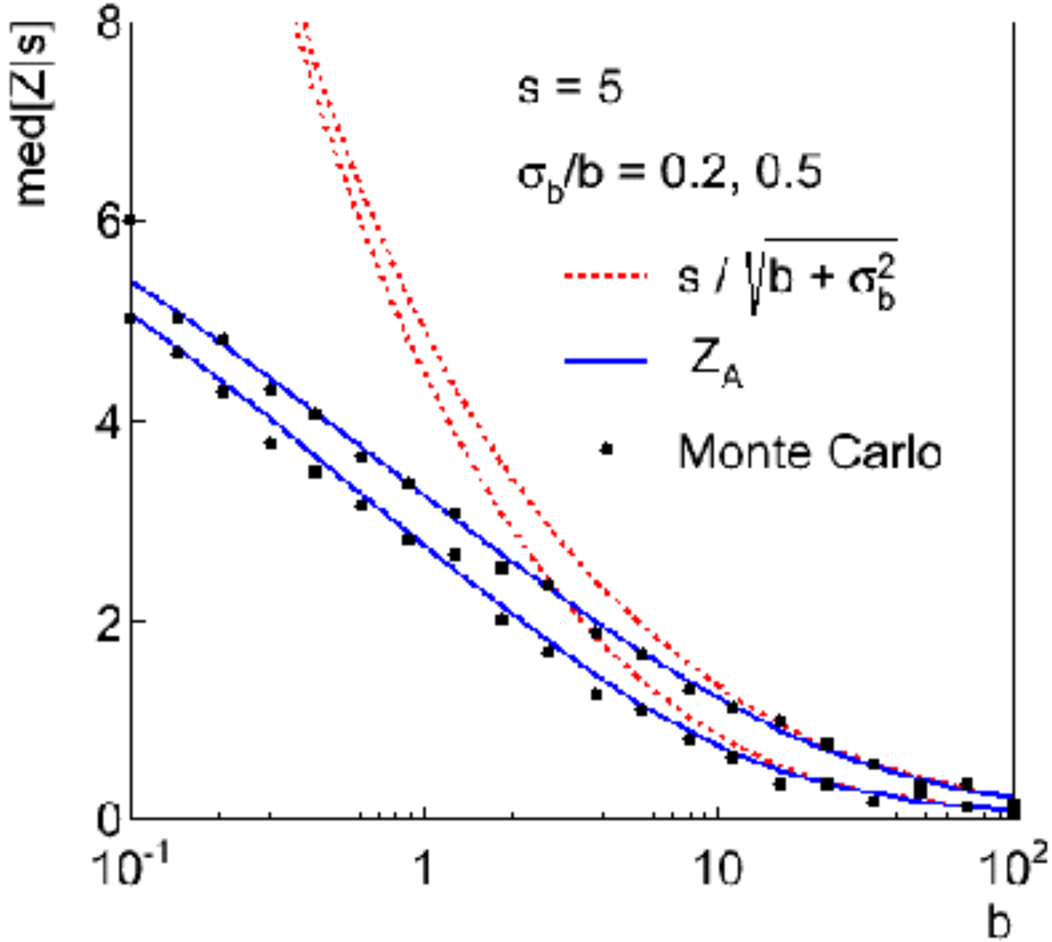
$$Z_{\mathrm{A}} = \left[2 \left((s+b) \ln \left[\frac{(s+b)(b+\sigma_b^2)}{b^2 + (s+b)\sigma_b^2} \right] - \frac{b^2}{\sigma_b^2} \ln \left[1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)} \right] \right) \right]^{1/2}$$

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



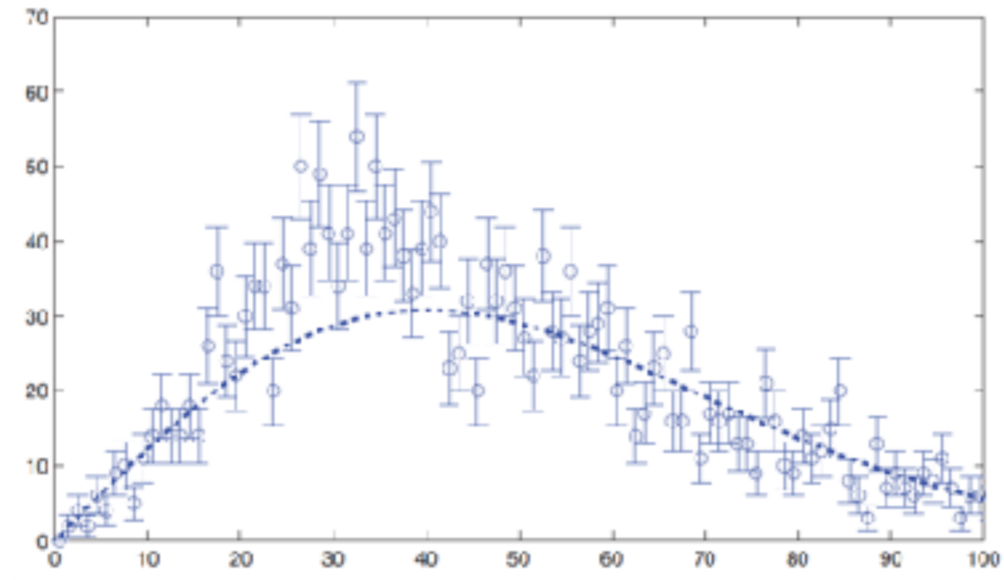
Look Elsewhere Effect

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

Look Elsewhere Effect

- Is there a signal here?

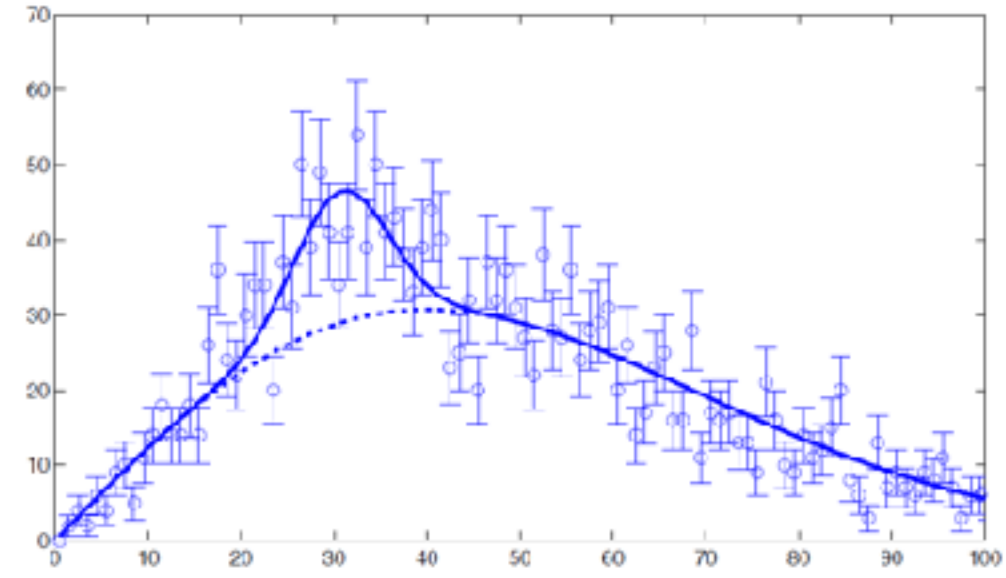


Look Elsewhere Effect

- Looks like a signal at m=30
- What is its significance?

Test the BG hypothesis
At m=30

$$q_0(\theta) = \begin{cases} -2\log \dfrac{L(\mu=0)}{L(\hat{\mu},\theta)} \\ 0 \end{cases}$$



$$q_{fix,obs} = -2\ln \frac{L(b)}{L(\hat{\mu}s(m=30)+b)}$$

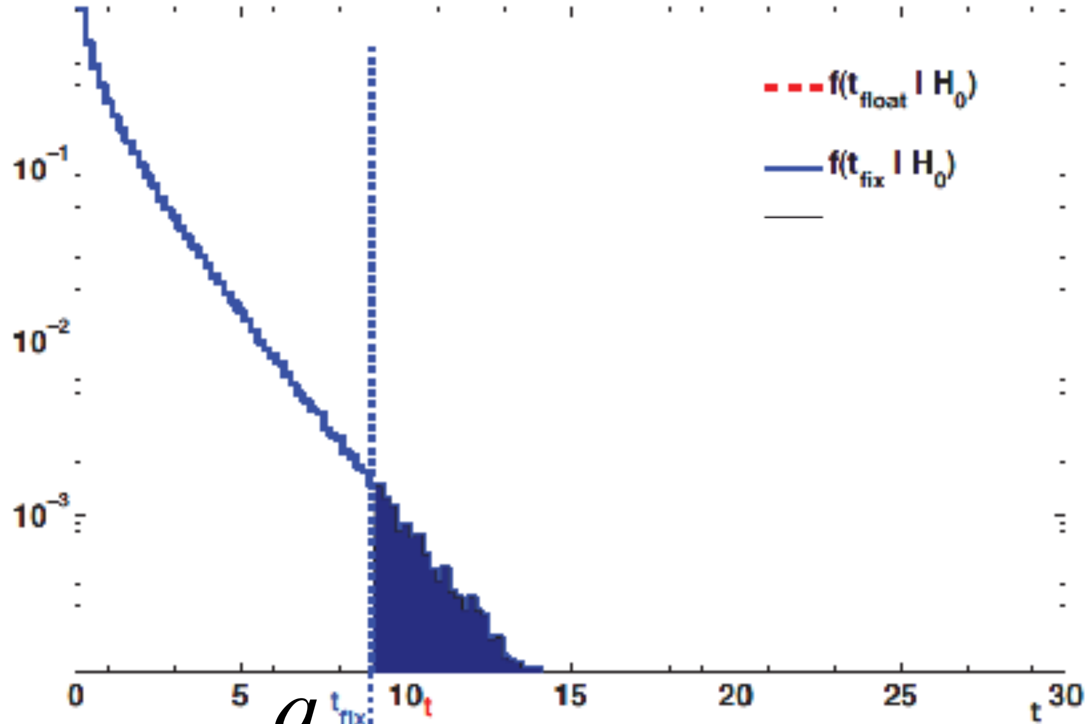
$$Z = \sqrt{q_{0,fix,obs}}$$

Look Elsewhere Effect

$$q_{0,fix} = -2ln\frac{L(\mu=0)}{L(\hat{\mu}s(30)+b)}$$

$$f(q_{0,fix} | H_0) \sim \chi^2$$

$$p_{fix} = \int_{q_{fix,obs}}^{\infty} f(q_0 | H_0)dq_0$$



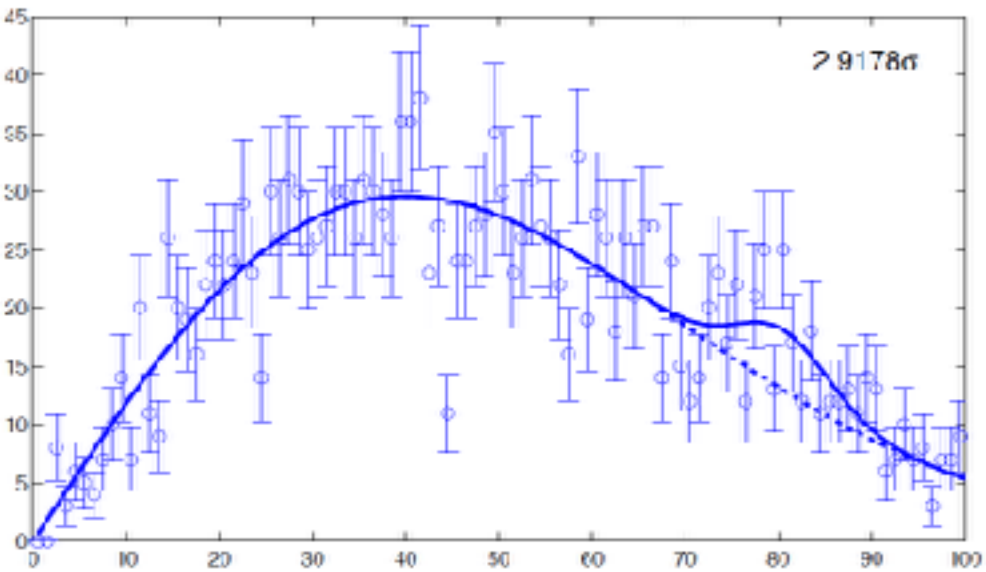
$q_{0,fix,obs}$

p_{fix} *answers the question:*

What is the probability to have a fluctuation as or bigger than the observed one?

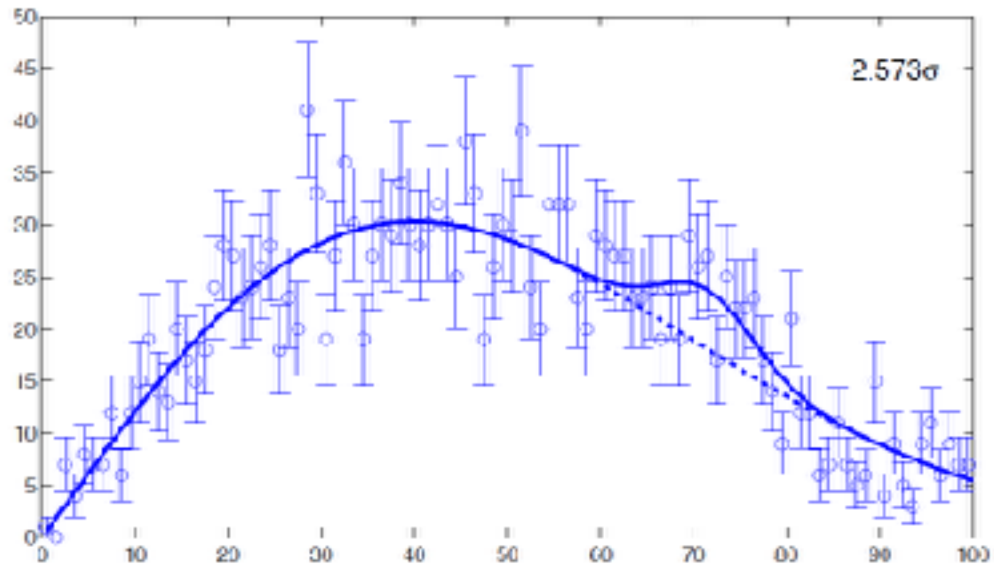
Look Elsewhere Effect

- Would you ignore this signal, had you seen it?



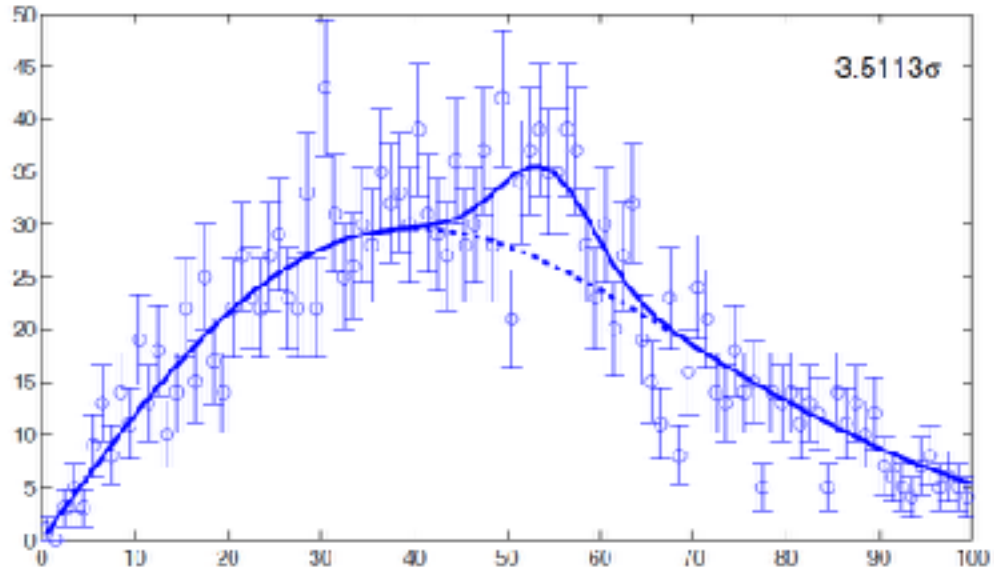
Look Elsewhere Effect

- Or this?



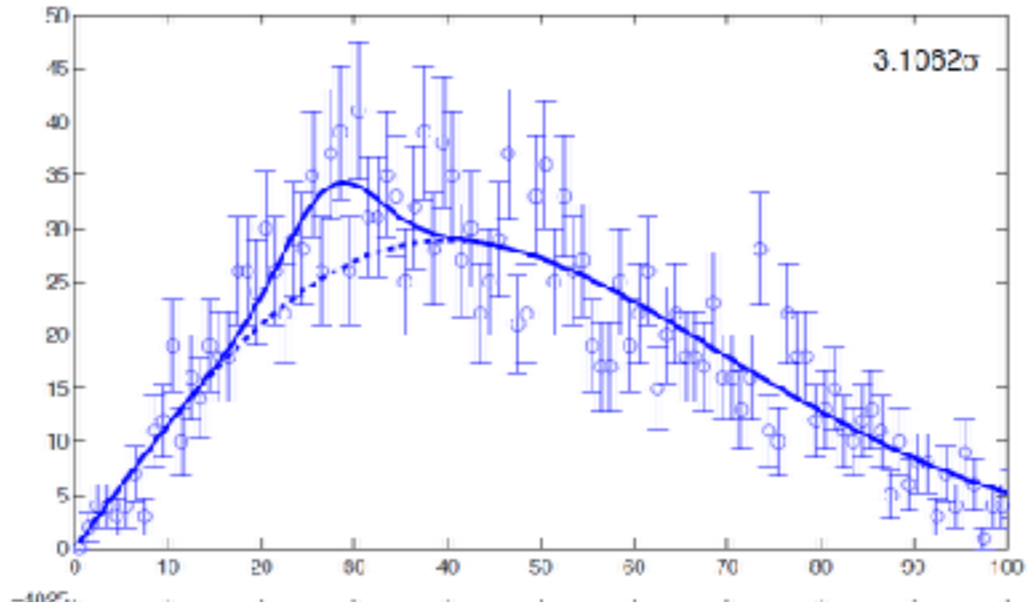
Look Elsewhere Effect

- Or this?



Look Elsewhere Effect

- Or this?
- Obviously NOT!
- ALL THESE "SIGNALS" ARE BG FLUCTUATIONS



The right question :

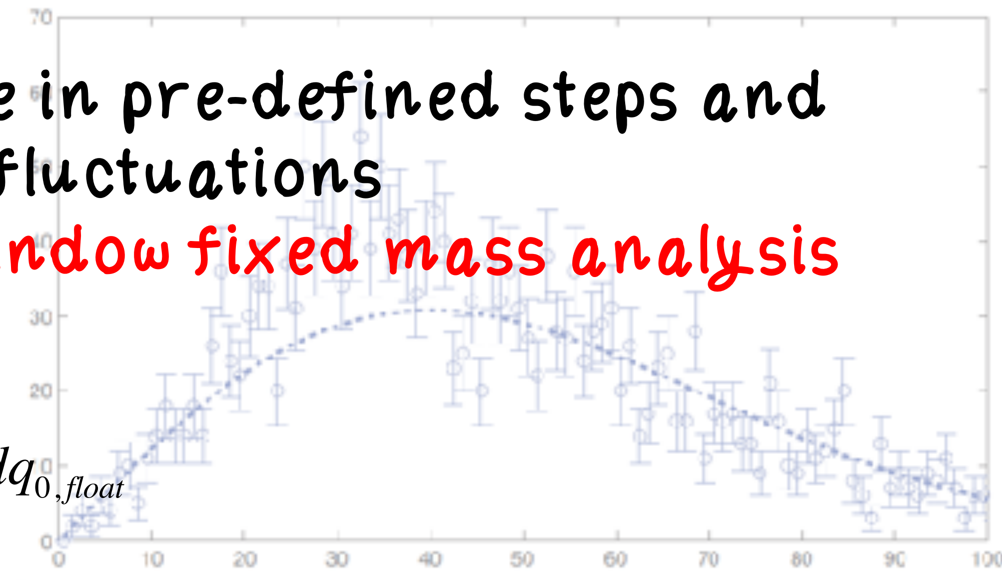
What is the probability to have a fluctuation as or bigger than the observed one ***ANYWHERE in the mass search range****?*

Look Elsewhere Effect

- Having no idea where the signal might be there are two equivalent options
- OPTION I:

scan the mass range in pre-defined steps and test any disturbing fluctuations

Perform a sliding window fixed mass analysis



$$q_{0,float} = \max_{m}\left(q_0(m)\right)$$

$$p_{float} = \int_{q_{float,obs}}^{\infty} f(q_{0,float} \mid H_0)\, dq_{0,float}$$

- OPTION II:

Perform a floating mass analysis

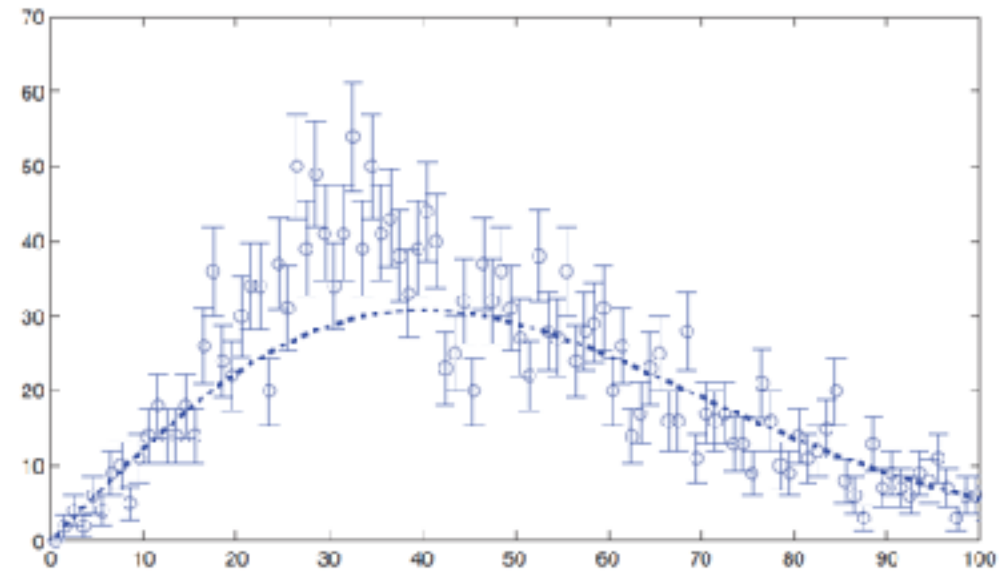
$$q_{0,float} = q_0(\hat{m}) = -2\ln\frac{L(b)}{L(\hat{\mu}s(\hat{m})+b)}$$

$$p_{float} = \int_{q_{float,obs}}^{\infty} f(q_{0,float} \mid H_0)\, dq_{0,float}$$

Sliding Window

- Scan and perform a fixed mass analysis at each point

$$q_0 = -2\ln\frac{L(\mu=0)}{L(\hat{\mu}s(m)+b)}$$



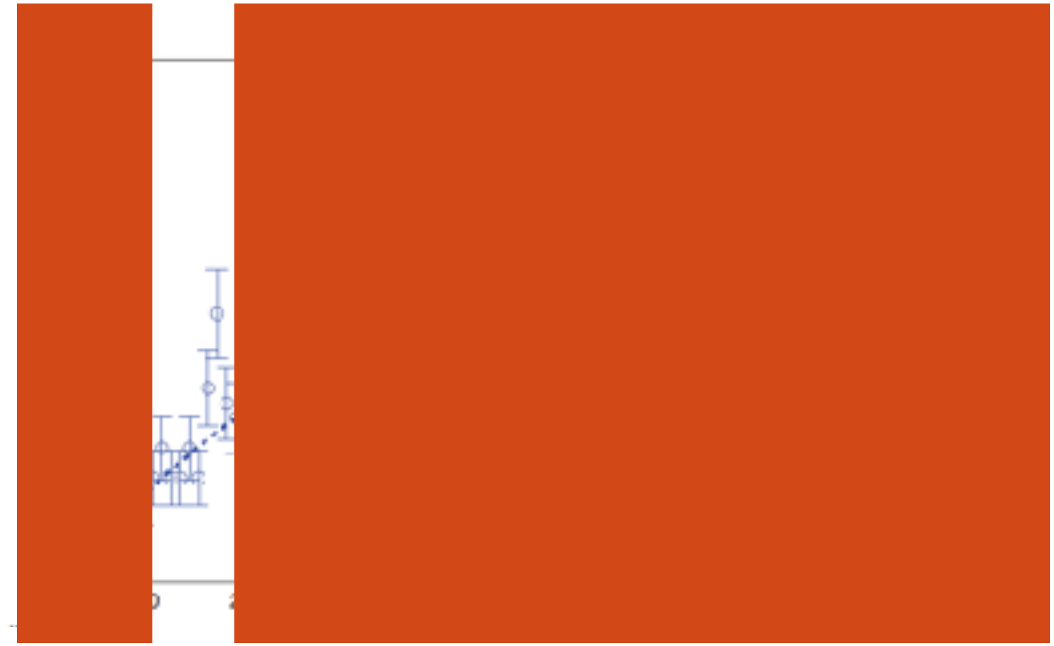
- The scan resolution must be less than the signal mass resolution

Sliding Window

$$q_0 = -2ln\frac{L(\mu=0)}{L(\hat{\mu}s(m)+b)}$$

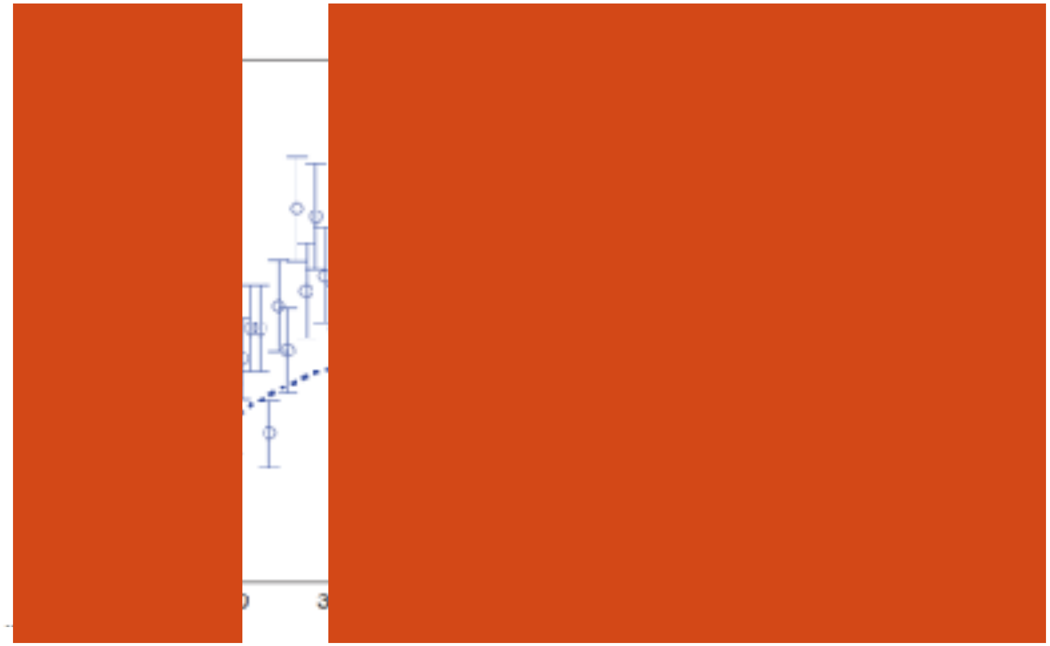
Sliding Window

$$q_0 = -2ln\frac{L(\mu=0)}{L(\hat{\mu}s(m)+b)}$$



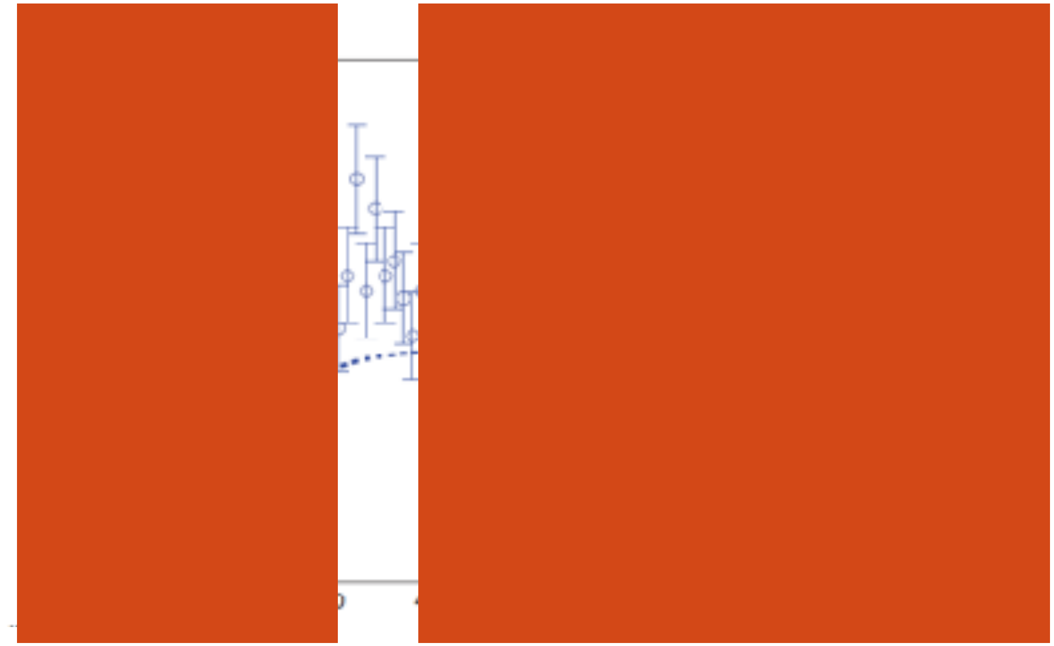
Sliding Window

$$q_0 = -2ln\frac{L(\mu=0)}{L(\hat{\mu}s(m)+b)}$$



Sliding Window

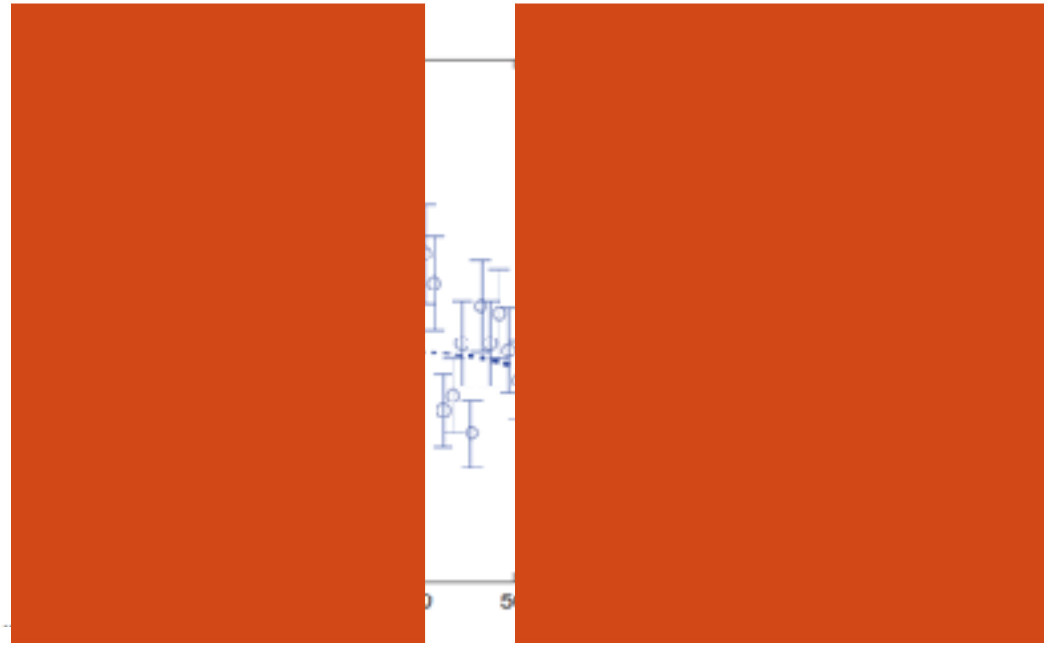
$$q_0 = -2ln\frac{L(\mu=0)}{L(\hat{\mu}s(m)+b)}$$



1/30/18

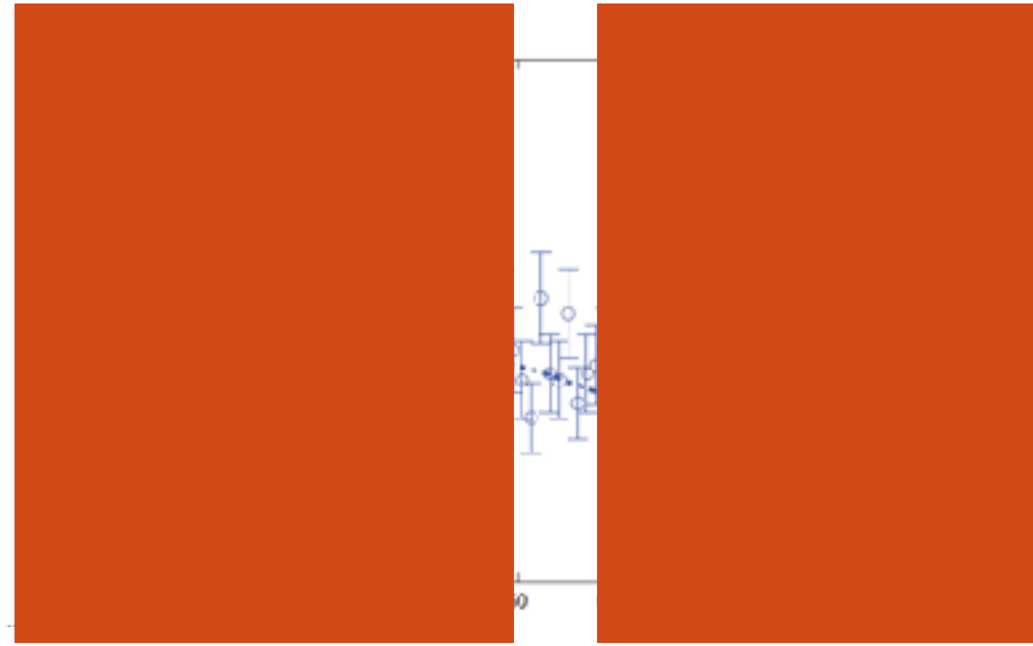
Sliding Window

$$q_0 = -2ln\frac{L(\mu = 0)}{L(\hat{\mu}s(m) + b)}$$



Sliding Window

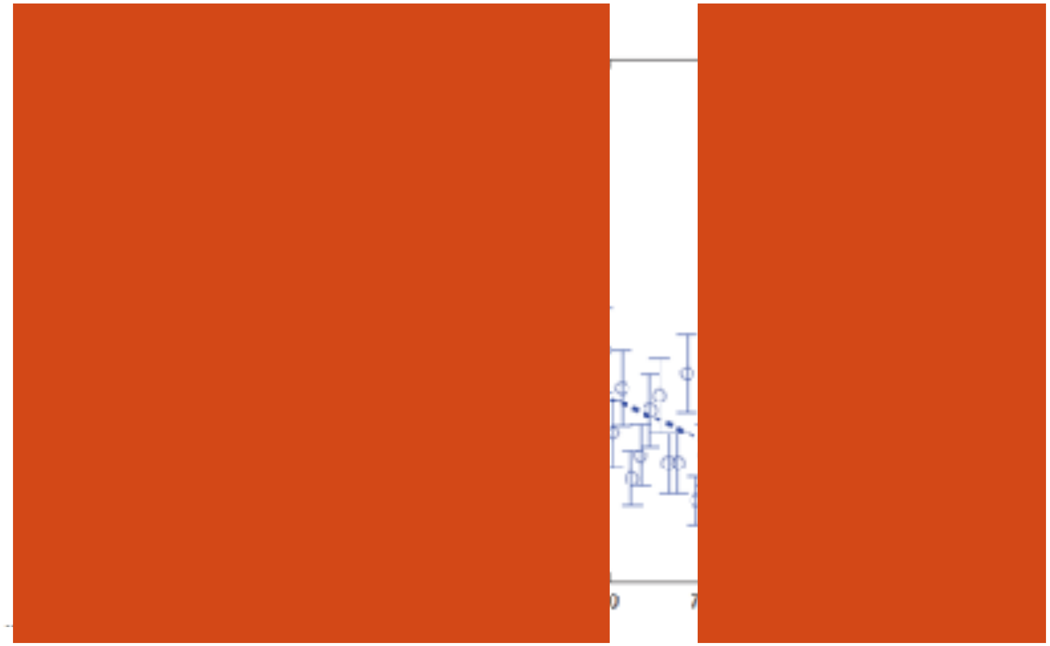
$$q_0 = -2ln\frac{L(\mu = 0)}{L(\hat{\mu}s(m) + b)}$$



1/30/18

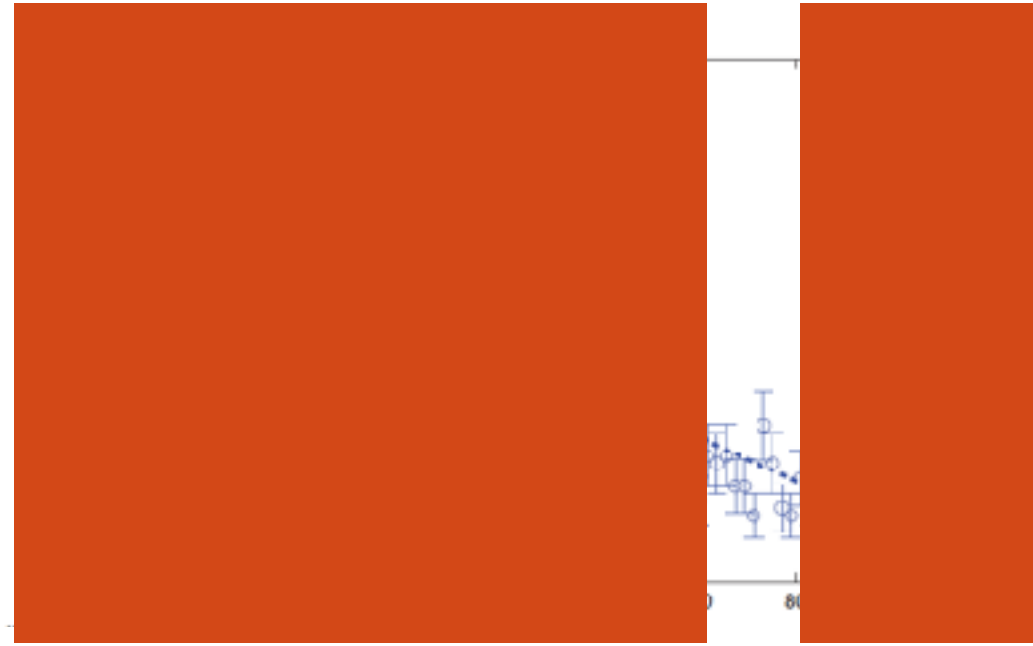
Sliding Window

$$q_0 = -2ln\frac{L(\mu=0)}{L(\hat{\mu}s(m)+b)}$$



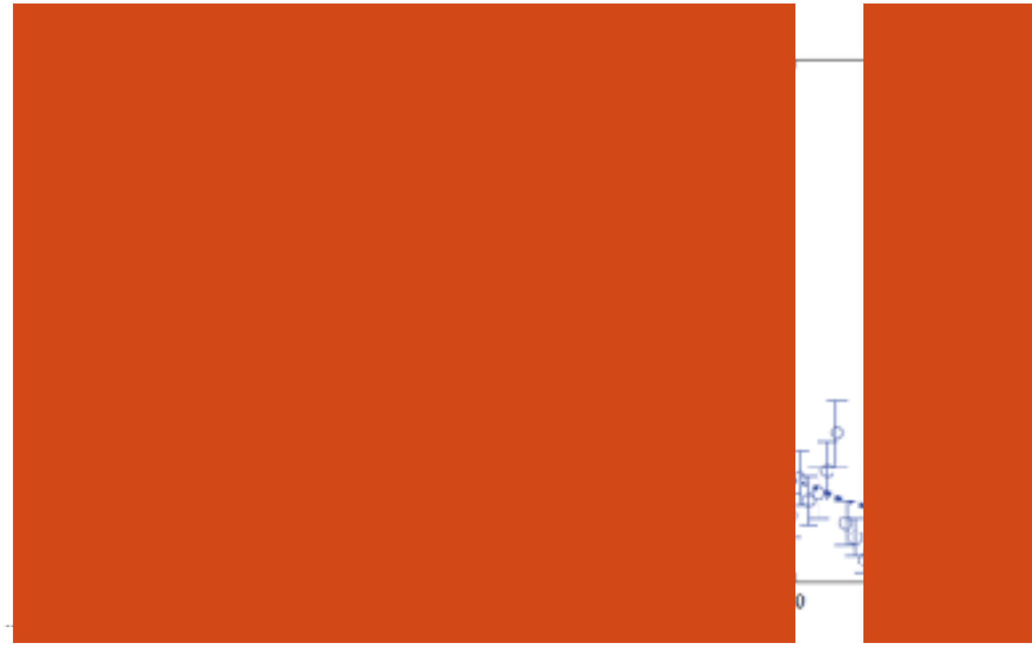
Sliding Window

$$q_0 = -2ln\frac{L(\mu=0)}{L(\hat{\mu}s(m)+b)}$$



Sliding Window

$$q_0 = -2ln\frac{L(\mu=0)}{L(\hat{\mu}s(m)+b)}$$



Sliding Window

- Assuming the signal can be only at one place
- pick the one with the MAXIMUM SIGNIFICANCE

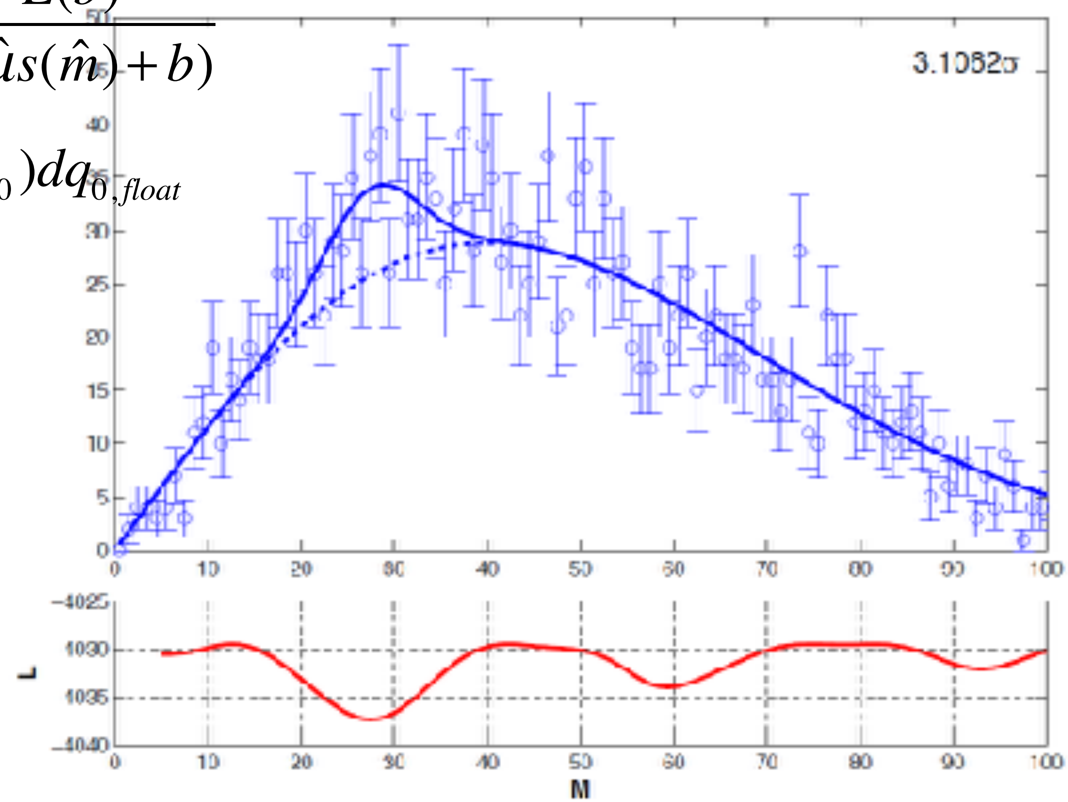
$$q_{0,float} = \max_{m}\left(q_0(m)\right)$$

Look Elsewhere Effect: Floating Mass

- **OPTION II**

$$q_{0,float} = q_0(\hat{m}) = -2\ln\frac{L(b)}{L(\hat{\mu}s(\hat{m})+b)}$$

$$p_{float} = \int_{q_{float,obs}}^{\infty} f(q_{0,float} \mid H_0)\, dq_{0,float}$$

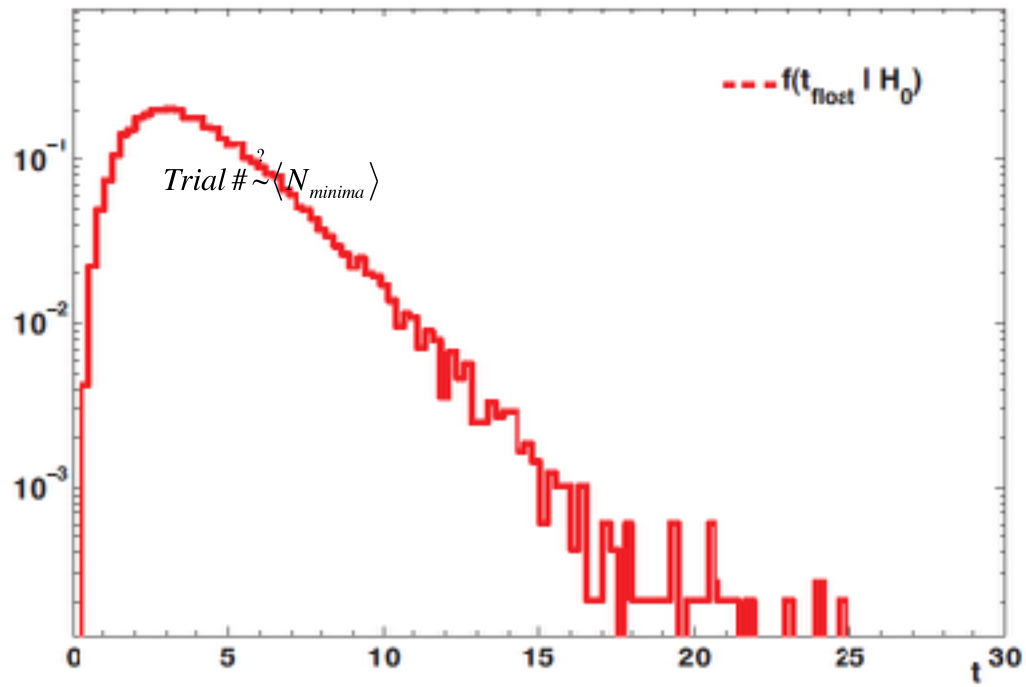


Look Elsewhere Effect

- The distribution $f(q_{float}|H_0)$ does not follow a chi-squared with 2dof because the mass parameter is not defined under the null hypothesis

$$\exists m_{fix} \quad q_0(\hat{m}) \geq q_0(m_{fix})$$

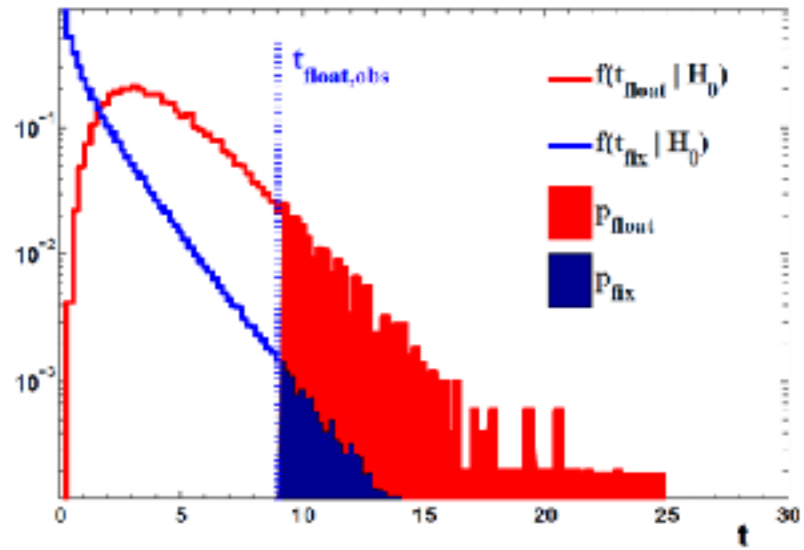
The χ_1^2 distribution is pushed to the right



trial#

- Assume a maximal local fluctuation at mass $\hat{m} = 30$
- The observed q_0 is given by

$$q_{0,obs} = -2ln\frac{L(\mu=0)}{L(\hat{\mu}s(m)+b)}$$



$$p_{fix} = \int_{q_{0,obs}}^{\infty} f(q_{0,fix} \mid H_0) dq_{0,fix}$$

$$p_{float} = \int_{q_{0,obs}}^{\infty} f(q_{0,float} \mid H_0) dq_{0,float}$$

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

Can we calculate analytically the floating mass p-value

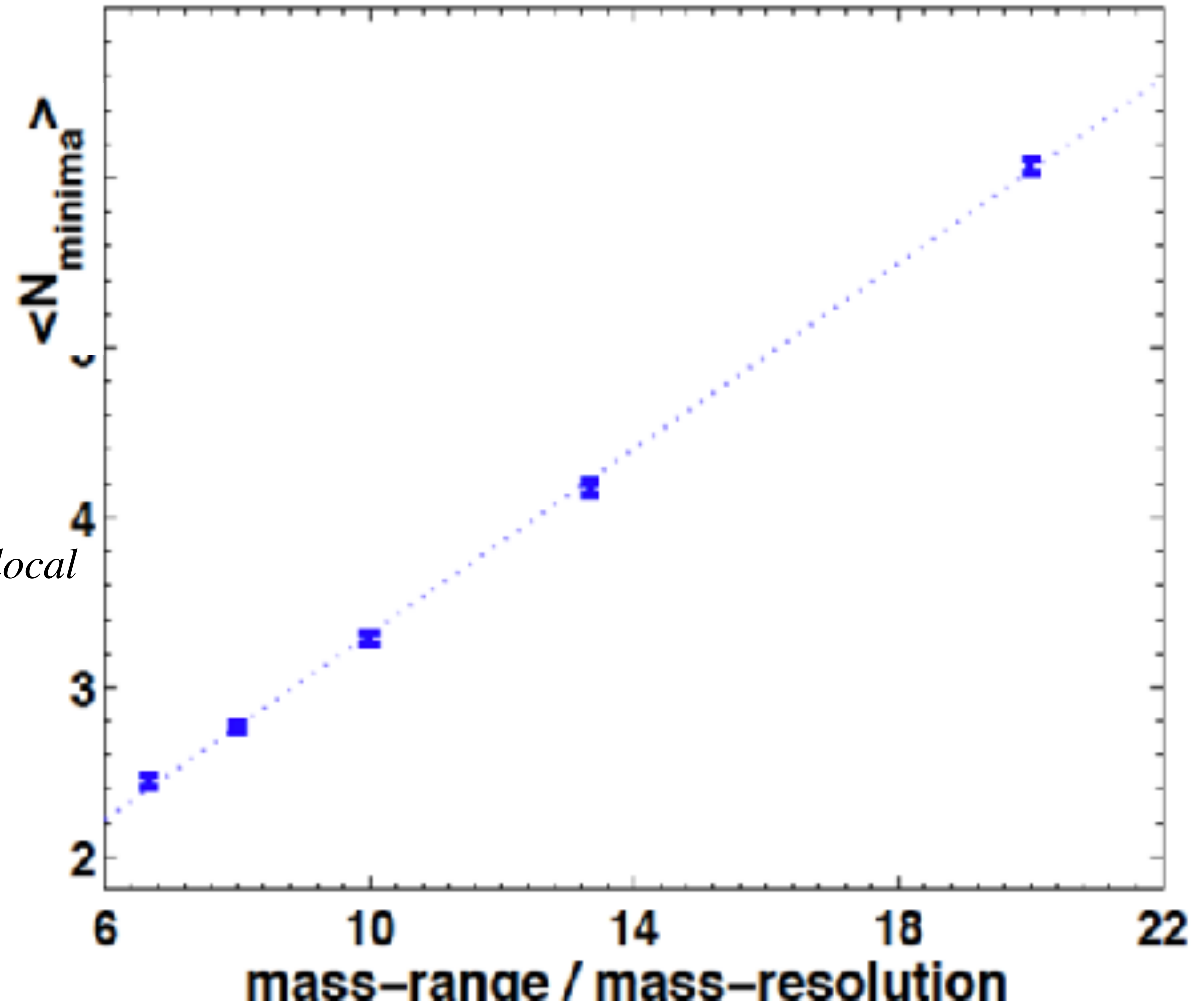
(Wrong) Thumb Rule

$$\langle N_{minima} \rangle \sim \frac{Mass\ Range}{Mass\ Resolution}$$

$$Trial\,\# \sim \langle N_{minima} \rangle$$

$$Trial\,\# \overset{?}{=} \langle N_{minima} \rangle\, p_{local}$$

The answer is NO



The right question:

What is the probability to have a fluctuation as or bigger than the observed one ***ANYWHERE*** *in the mass search range?*

Let θ *be a nuisance parameter undefined under the null hypothesis.*

Define $q(\hat{\theta}) = max_{\theta}(q(\theta))$

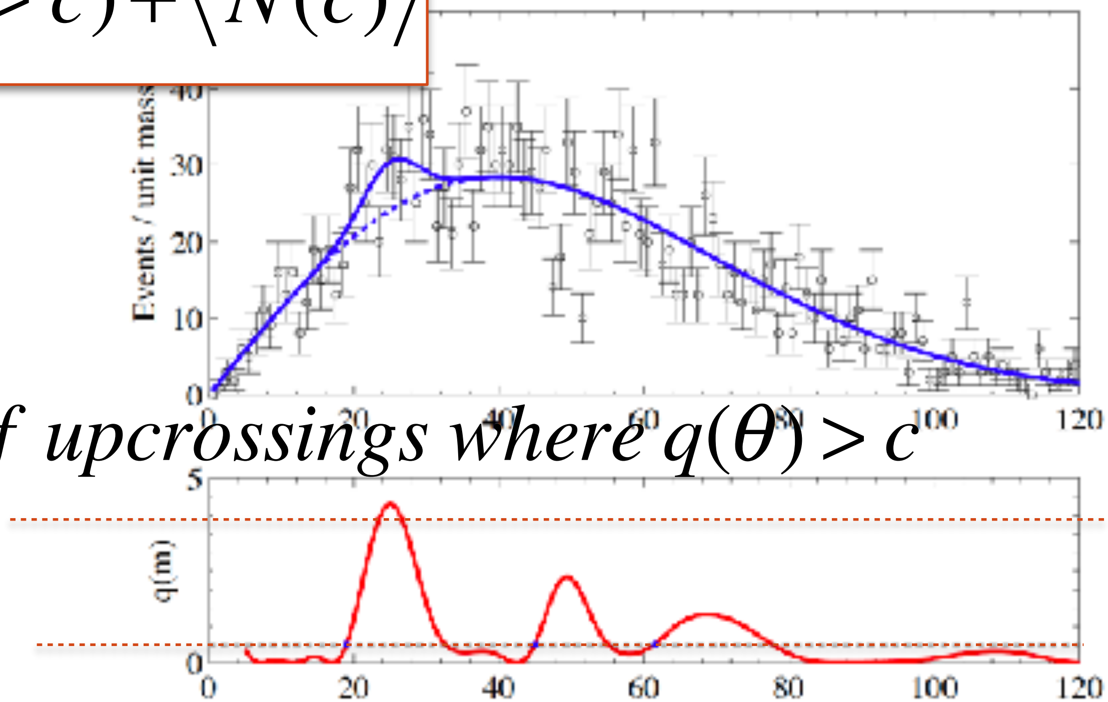
Davies (1987) finds, for $c >> 1$

$$P\left(q(\hat{\theta}) > c\right) \sim P(\chi_1^2 > c) + \langle N(c) \rangle$$

$\langle N(c) \rangle =$ *Number of upcrossings* $q(\theta) > c$

Davies Formula

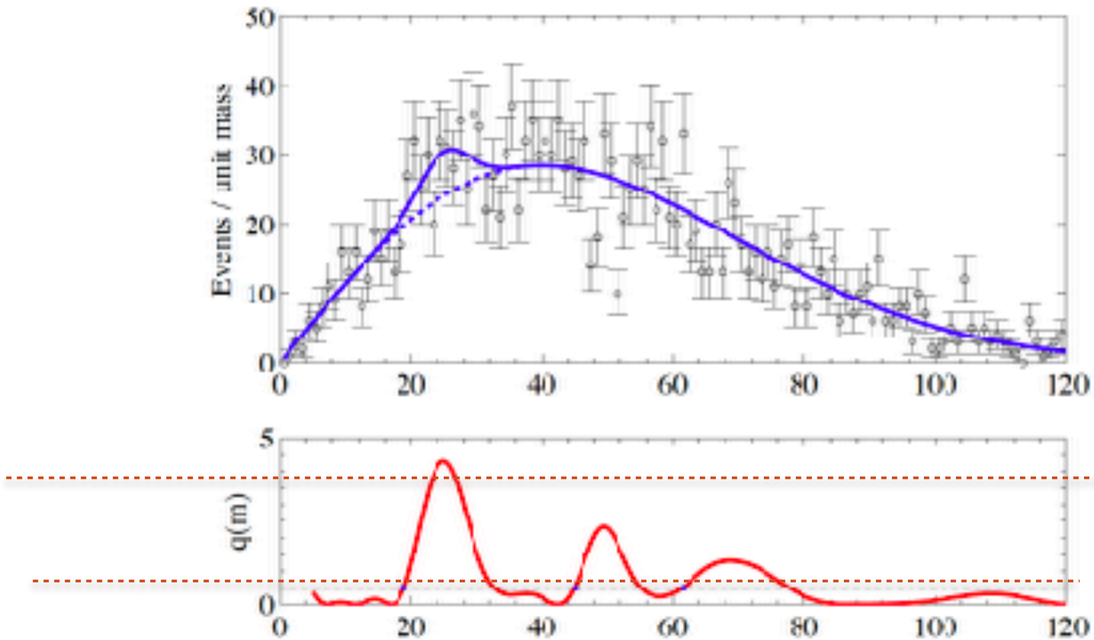
$$P\left(q(\hat{\theta})>c\right)\sim P(\chi_1^2>c)+\langle N(c)\rangle$$



$$\langle N(c)\rangle = Number\ of\ upcrossings\ where\ q(\theta)>c$$

$$or\ c>>1\rightarrow\langle N(c)\rangle<<1$$

Making Davies Formula Accessible



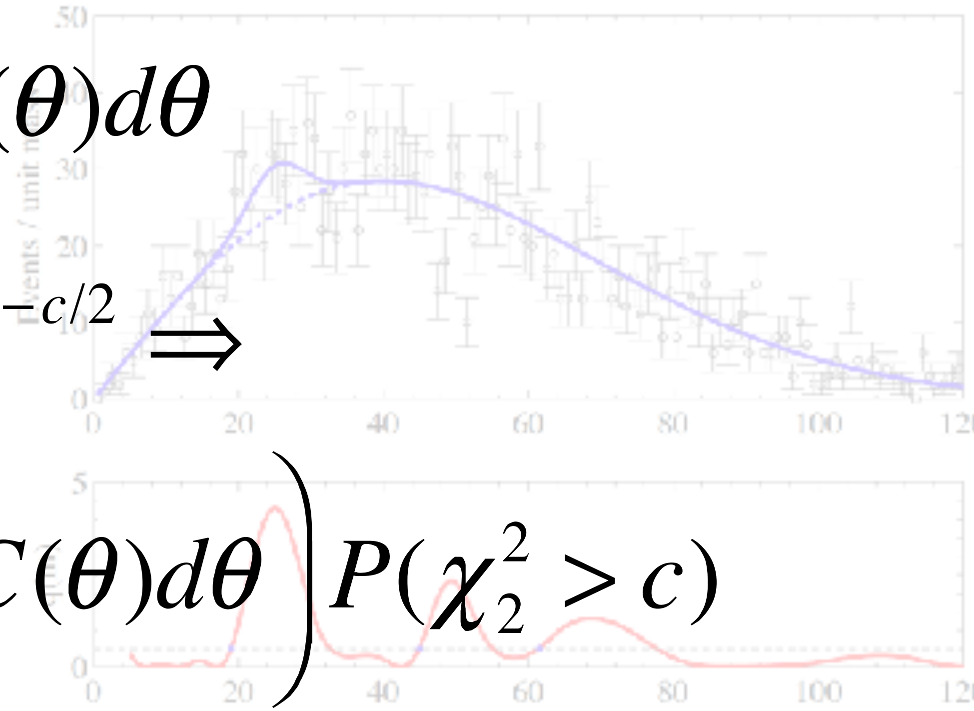
$$\langle N(c)\rangle \ll 1$$

$$\langle N(c)\rangle \sim e^{-c/2}$$

$$P(q(\hat{\theta})>c) \sim P(\chi_1^2>c)+\langle N(c_0)\rangle \frac{\langle N(c)\rangle}{\langle N(c_0)\rangle}$$

$$P(q(\hat{\theta})>c) \sim P(\chi_1^2>c)+\langle N(c_0)\rangle e^{-(c-c_0)/2}$$

Davies Formula



$$\langle N(c)\rangle \approx \frac{e^{-c/2}}{\sqrt{2\pi}}\int_\theta C(\theta)d\theta$$

$$P(\chi_2^2 > c)\xrightarrow[c\to\infty]{} e^{-c/2} \Rightarrow$$

$$\langle N(c)\rangle \approx \left(\frac{1}{\sqrt{2\pi}}\int_\theta C(\theta)d\theta\right)P(\chi_2^2 > c)$$

$$\langle N(c)\rangle = \mathcal{N}P(\chi_2^2 > c)$$

$$P(q(\hat{\theta}) > c) \sim P(\chi_1^2 > c) + \mathcal{N}P(\chi_2^2 > c)$$

Trial

$$P(\chi_1^2 > c) \xrightarrow{c>>1} \sqrt{\frac{2}{c}} \frac{e^{-c/2}}{\Gamma\left(\frac{1}{2}\right)}$$

$$P(\chi_2^2 > c) \xrightarrow{c>>1} e^{-c/2}$$

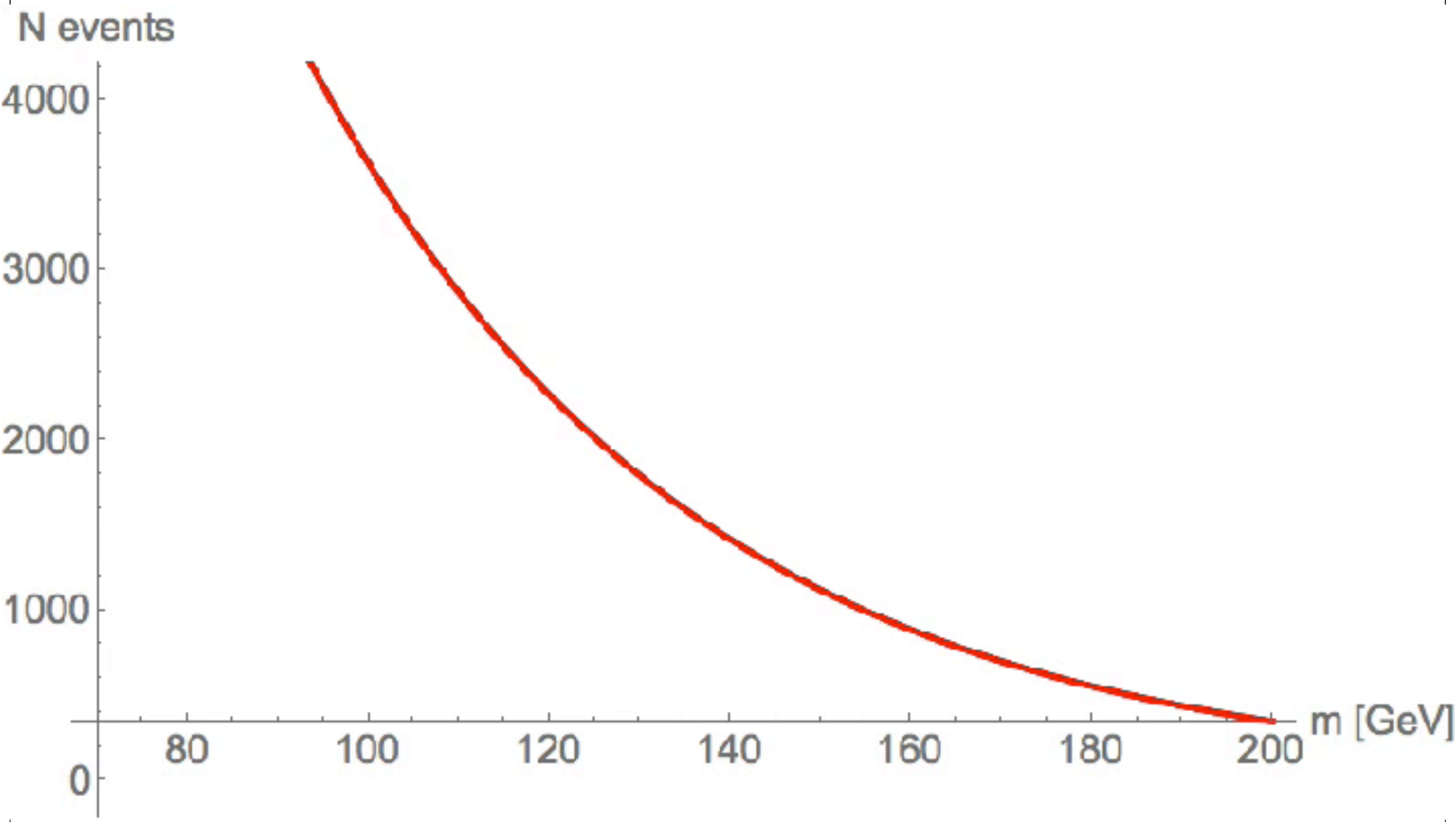
$$trial\,\# = \frac{P(q(\hat{\theta}) > c)}{P(q(\theta) > c)} \approx$$

$$\approx 1 + \mathcal{N} \frac{P(\chi_2^2 > c)}{P(\chi_1^2 > c)} \Rightarrow$$

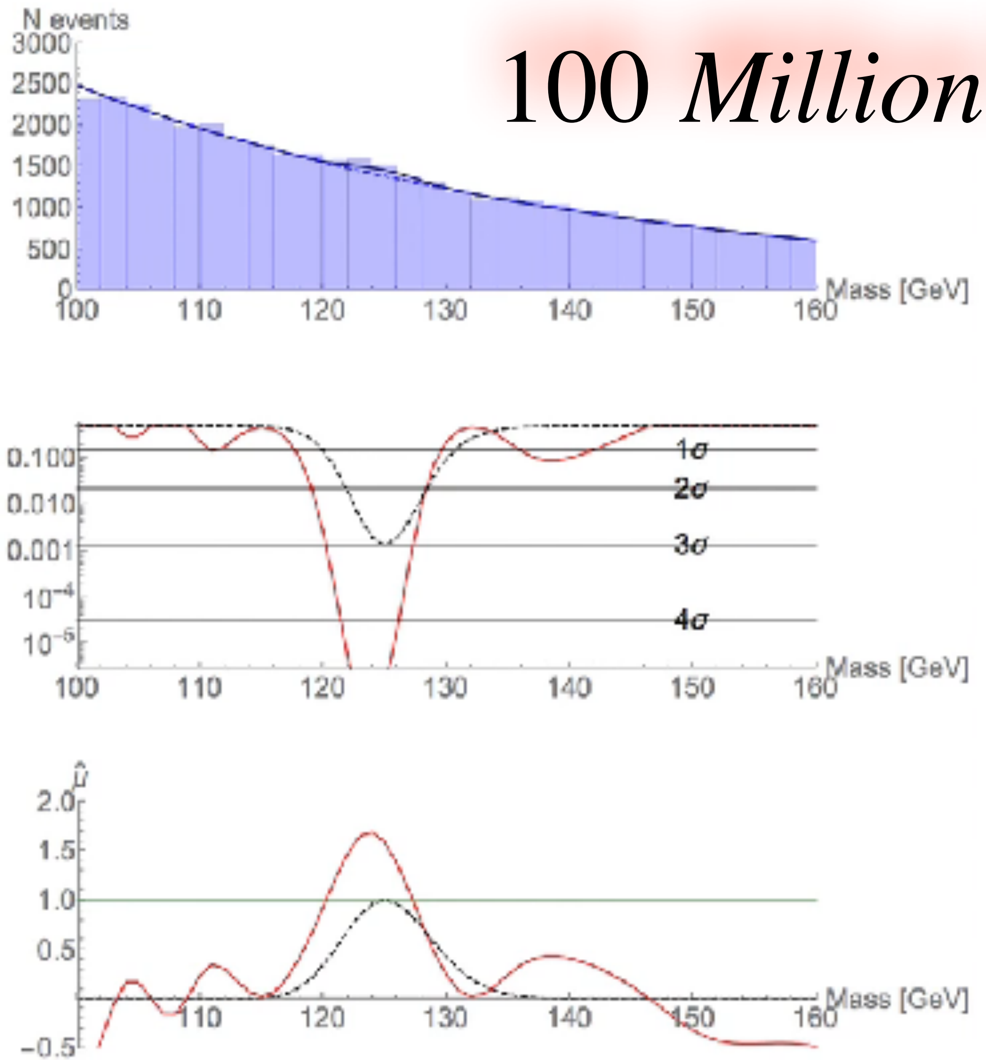
$$trial\,\# \approx 1 + \mathcal{N} \sqrt{\frac{c}{2}} \Gamma(1/2) \Rightarrow$$

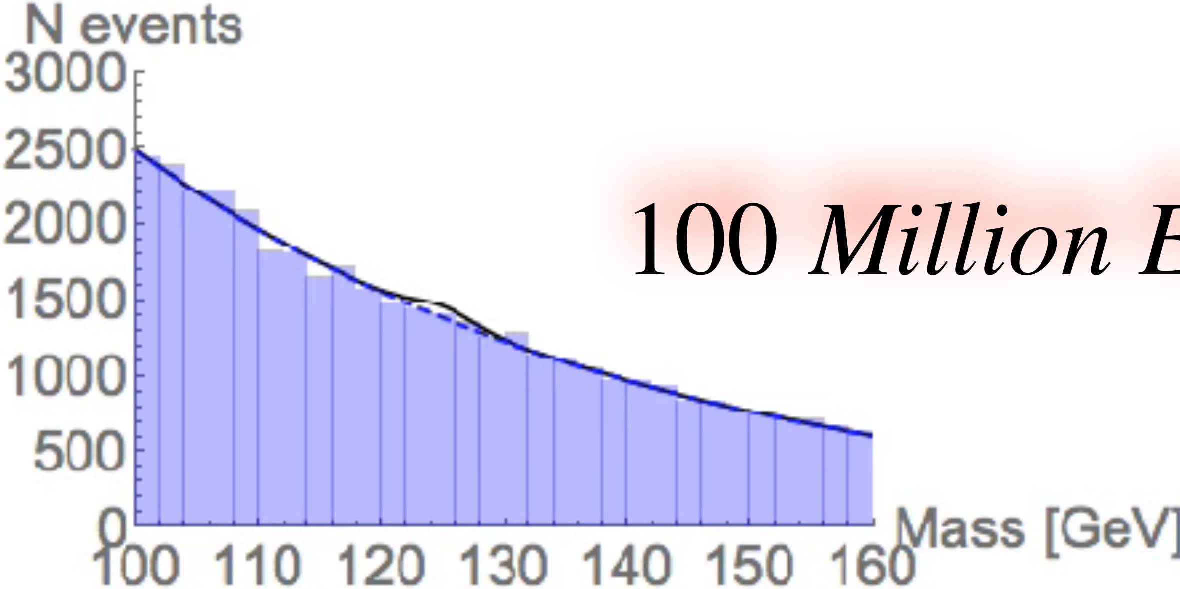
$$trial\,\# \approx 1 + \sqrt{\frac{\pi}{2}} \mathcal{N} Z_{fix}$$

What is Going On?


N events
4000
3000
2000
1000
0
80
100
120
140
160
180
200
m [GeV]

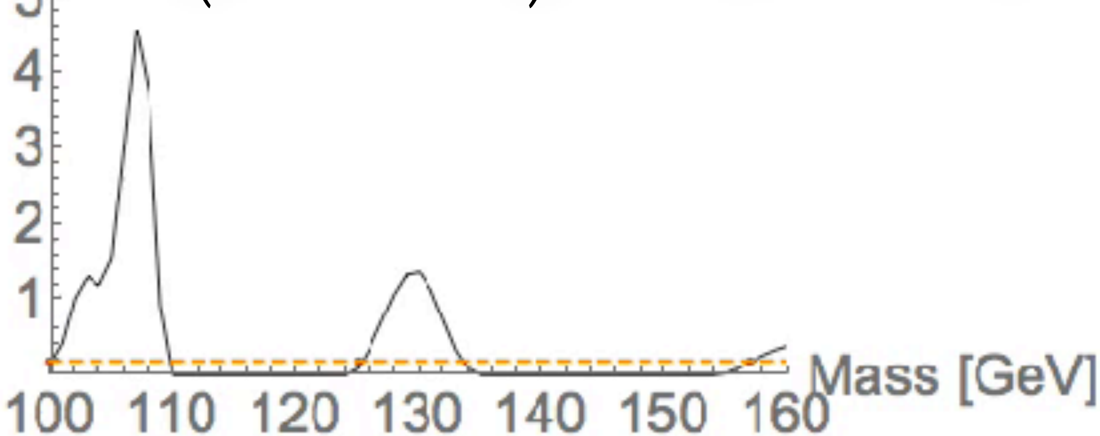
100 *Million Experiments*

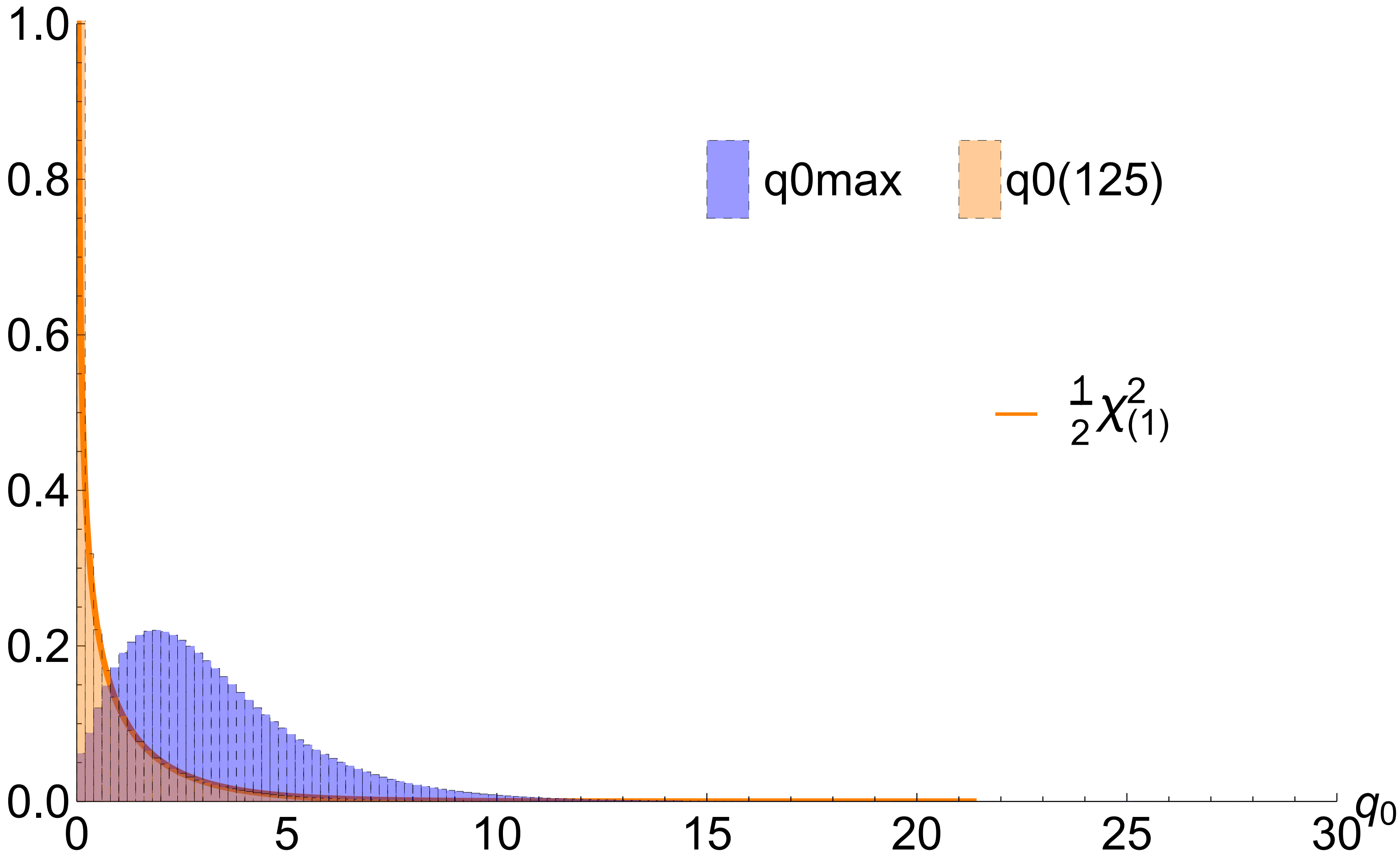


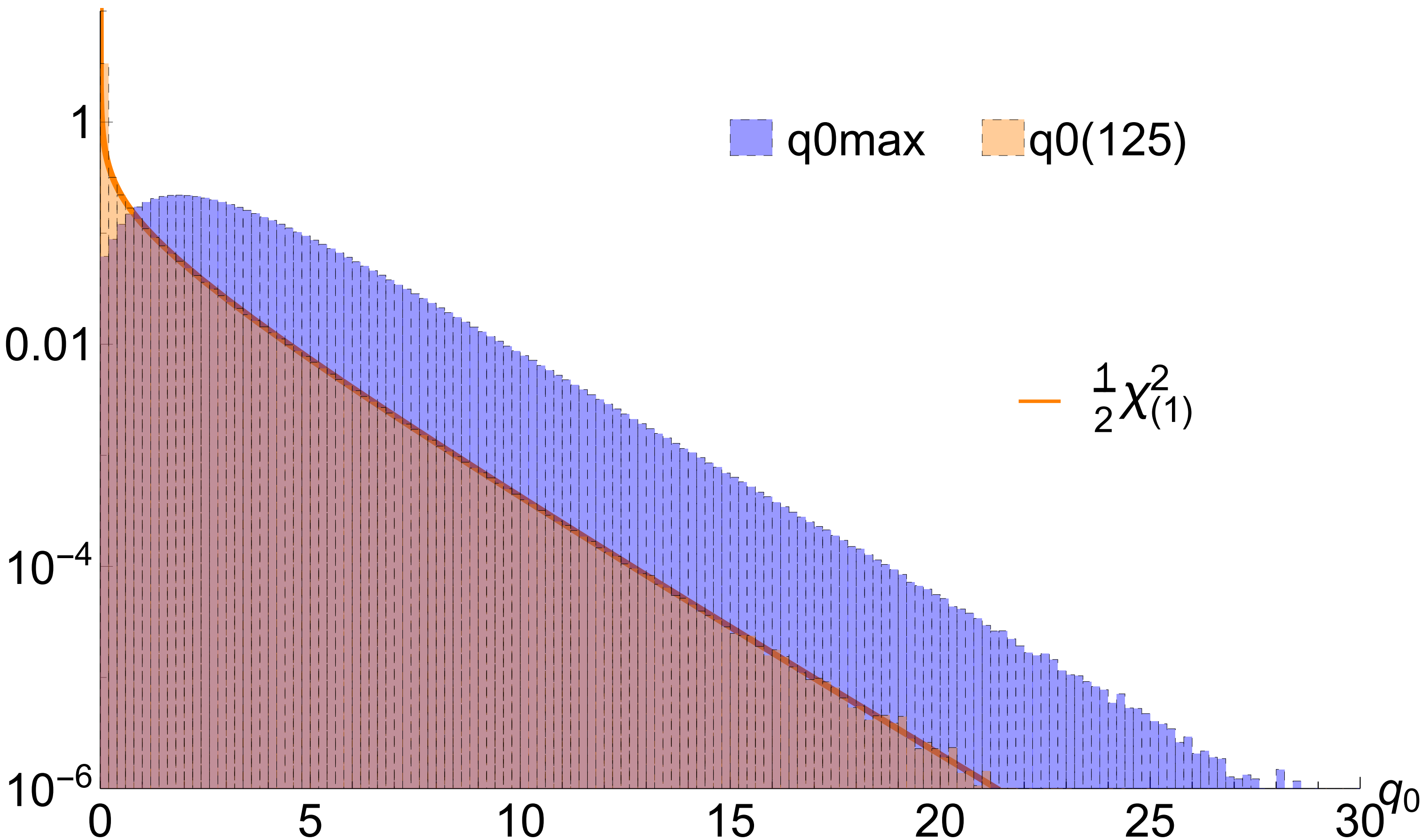


100 *Million Experiments*

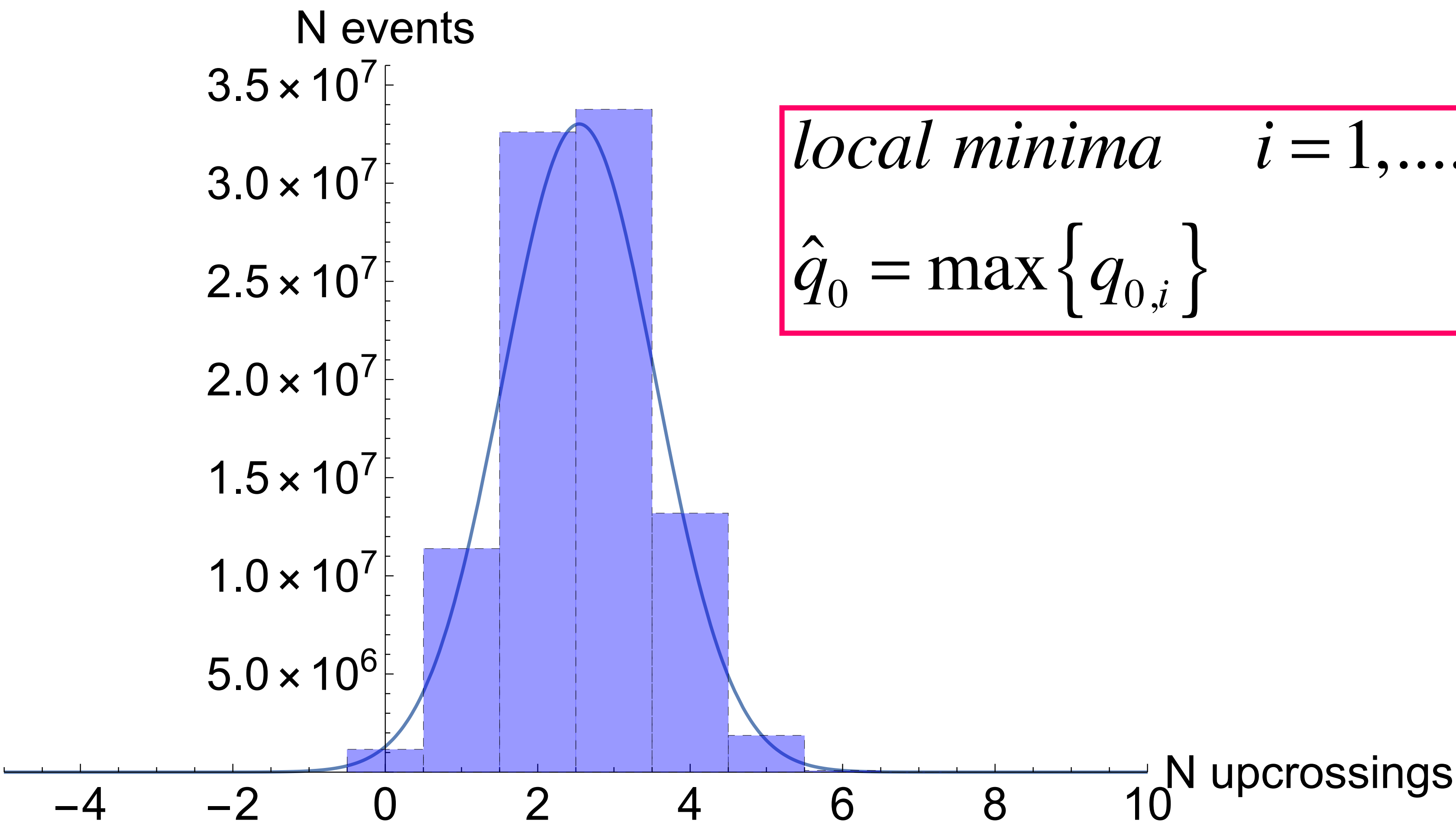
$$P\left(q(\hat{\theta}) > c\right) \sim P\left(\chi_1^2 > c\right) + \langle N(c_0)\rangle e^{-(c-c_0)/2}$$




q0max
q0(125)
$\frac{1}{2}\chi^2_{(1)}$
q_0
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30



1
0.01
10^{-4}
10^{-6}
0
5
10
15
20
25
30
q_0
q0max
q0(125)
$\frac{1}{2}\chi^2_{(1)}$

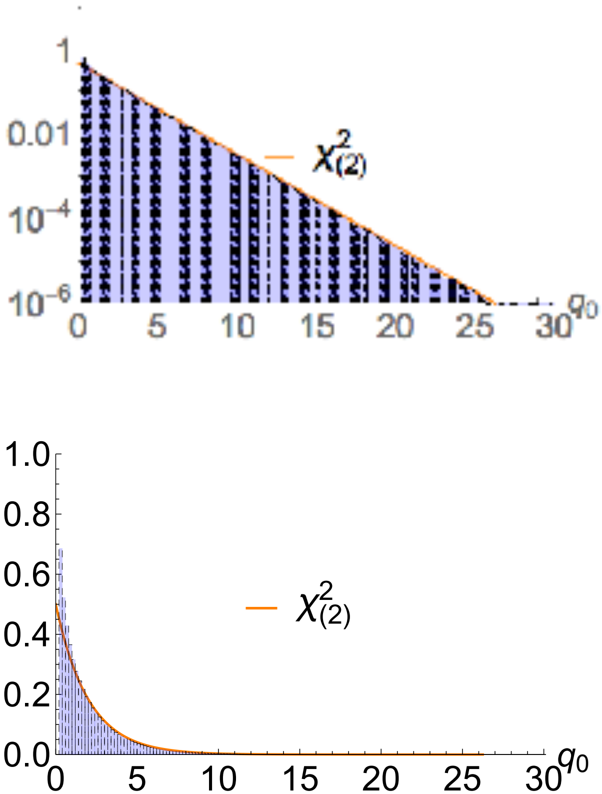


local minima $\quad i = 1, \ldots\ldots, n$

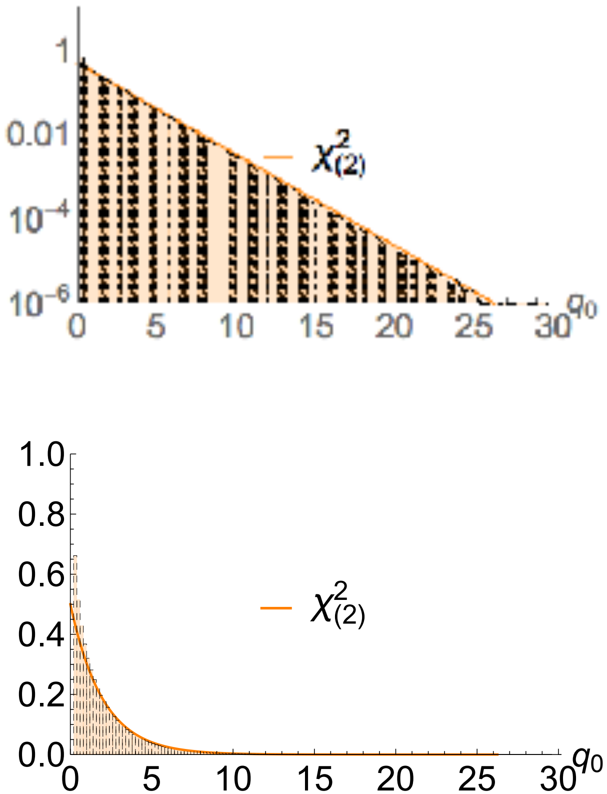
$$\hat{q}_0 = \max\left\{ q_{0,i} \right\}$$

$$\forall i \quad q_{0,i} \sim \chi_2^2$$

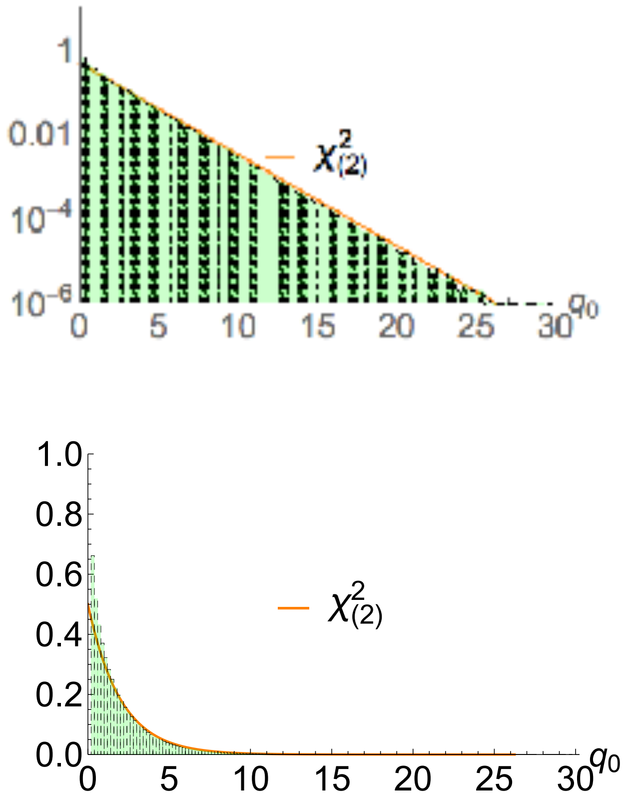
$q_{0,1}$

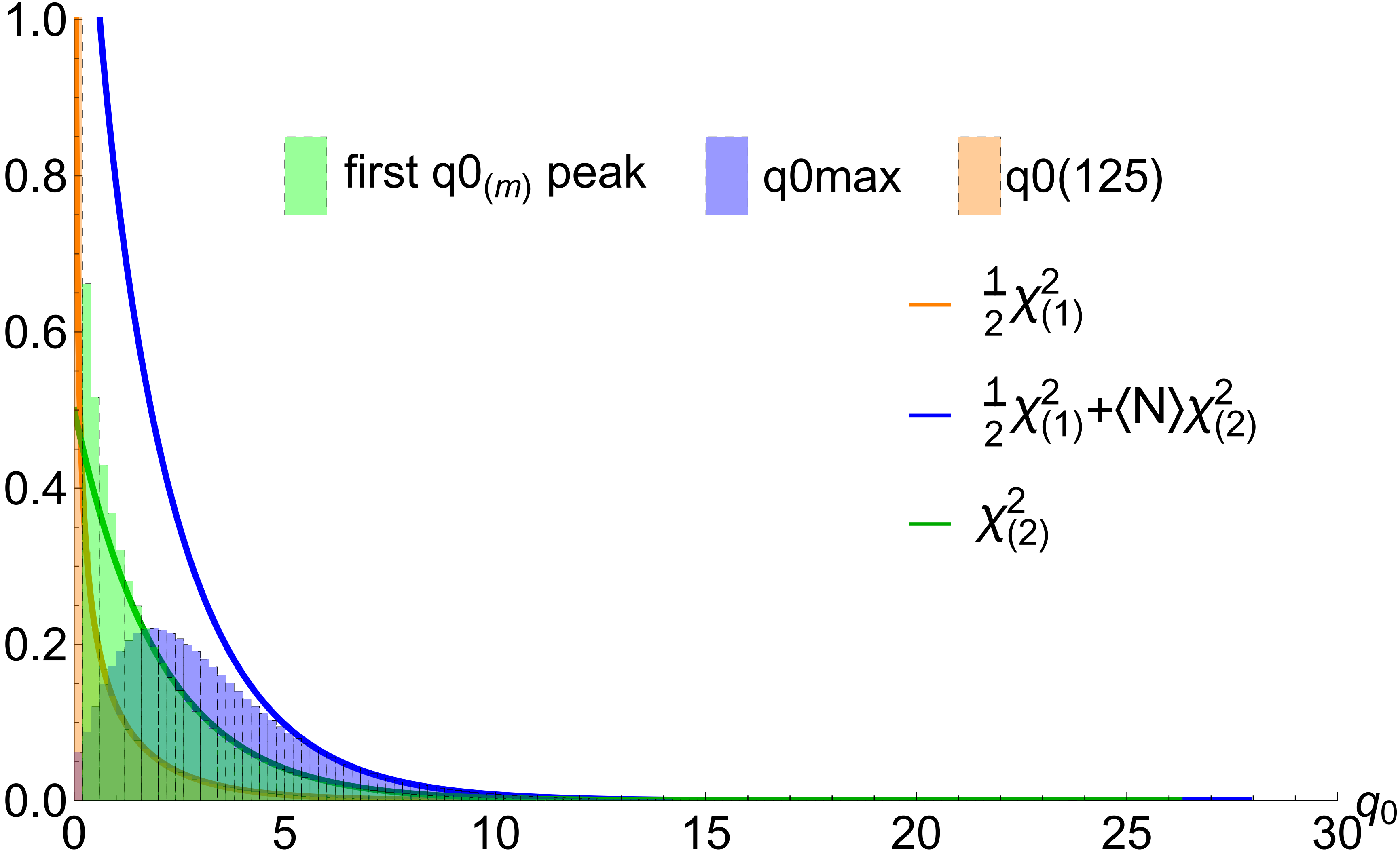


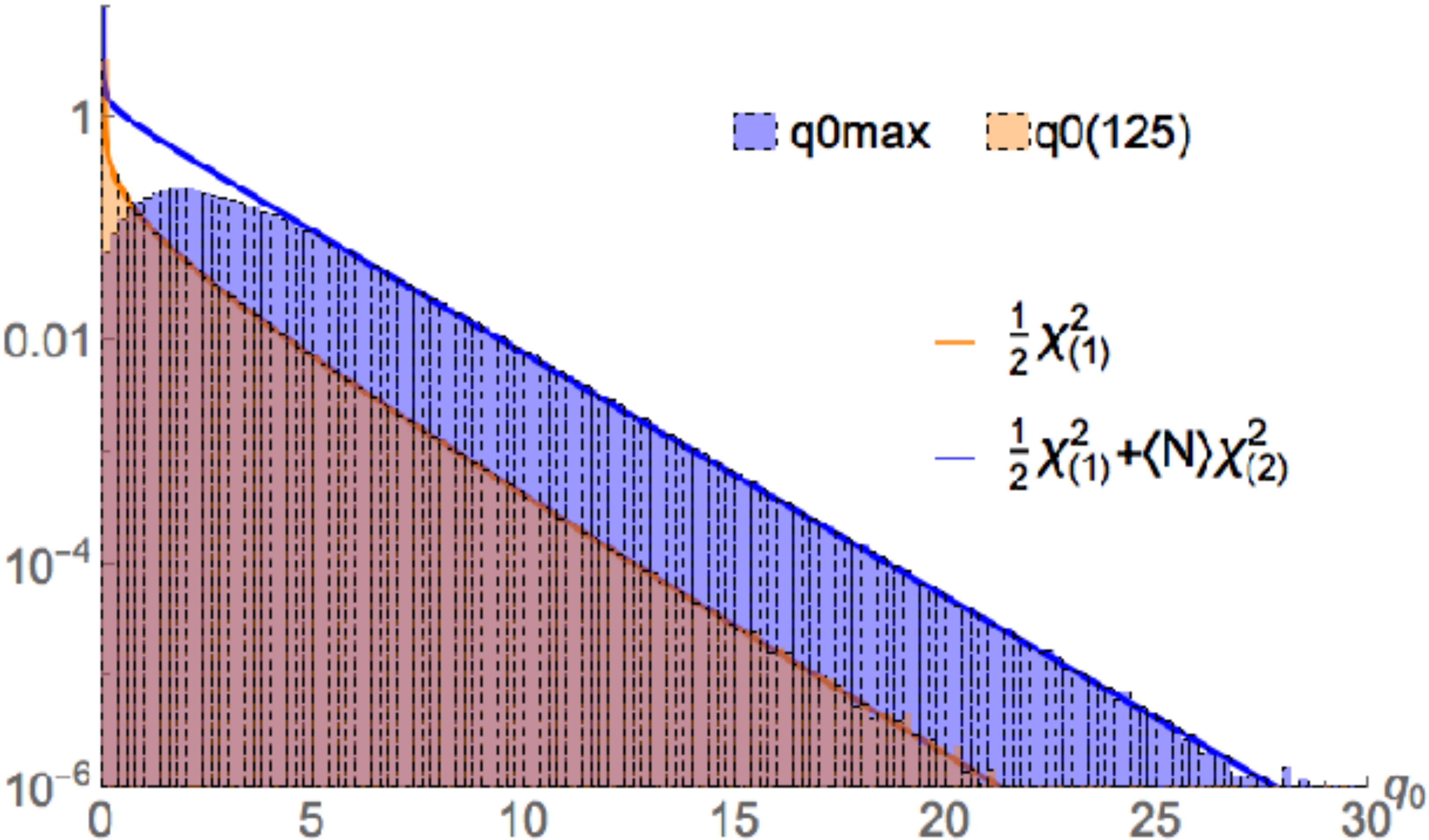
$q_{0,2}$



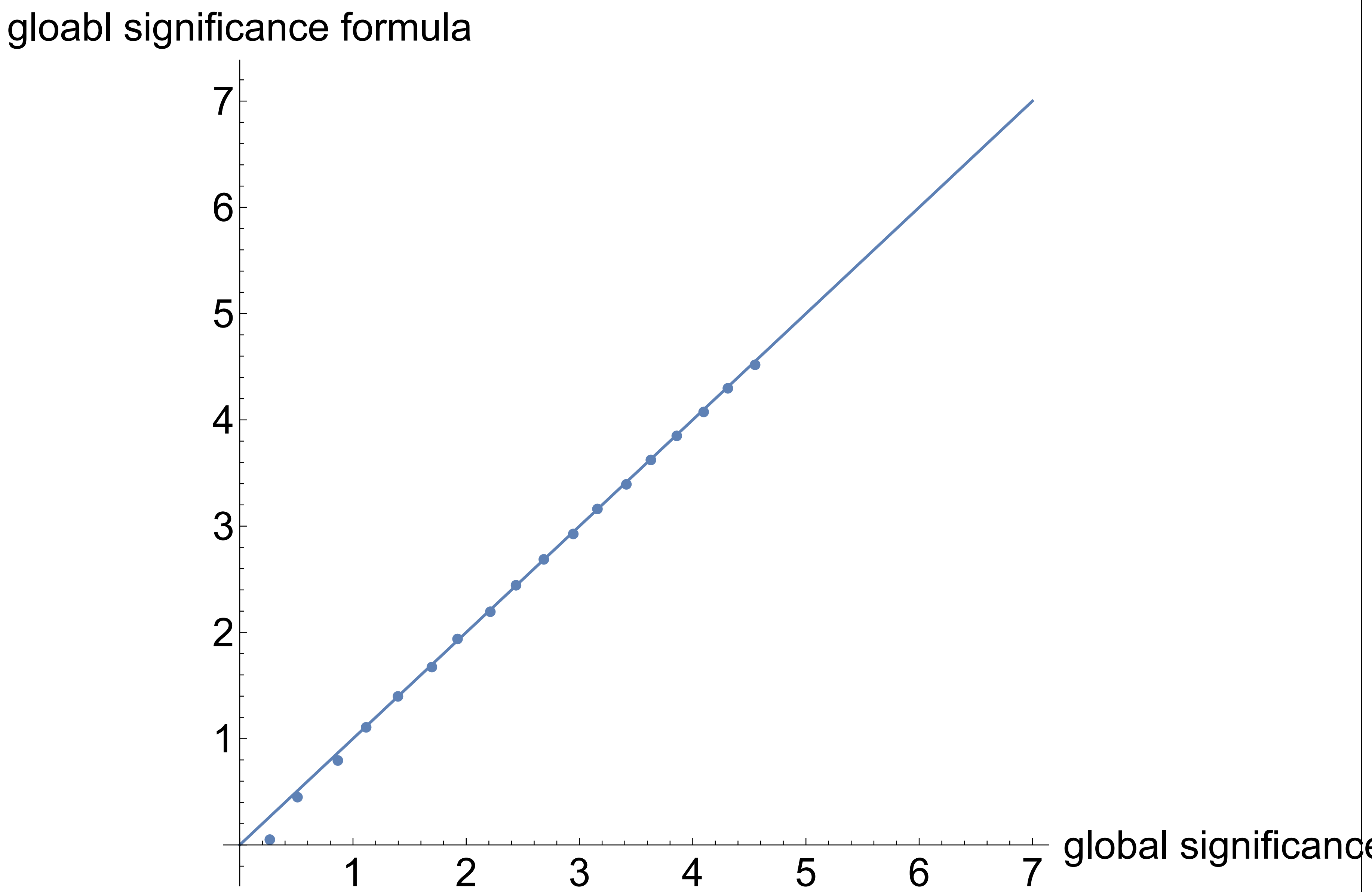
$q_{0,3}$



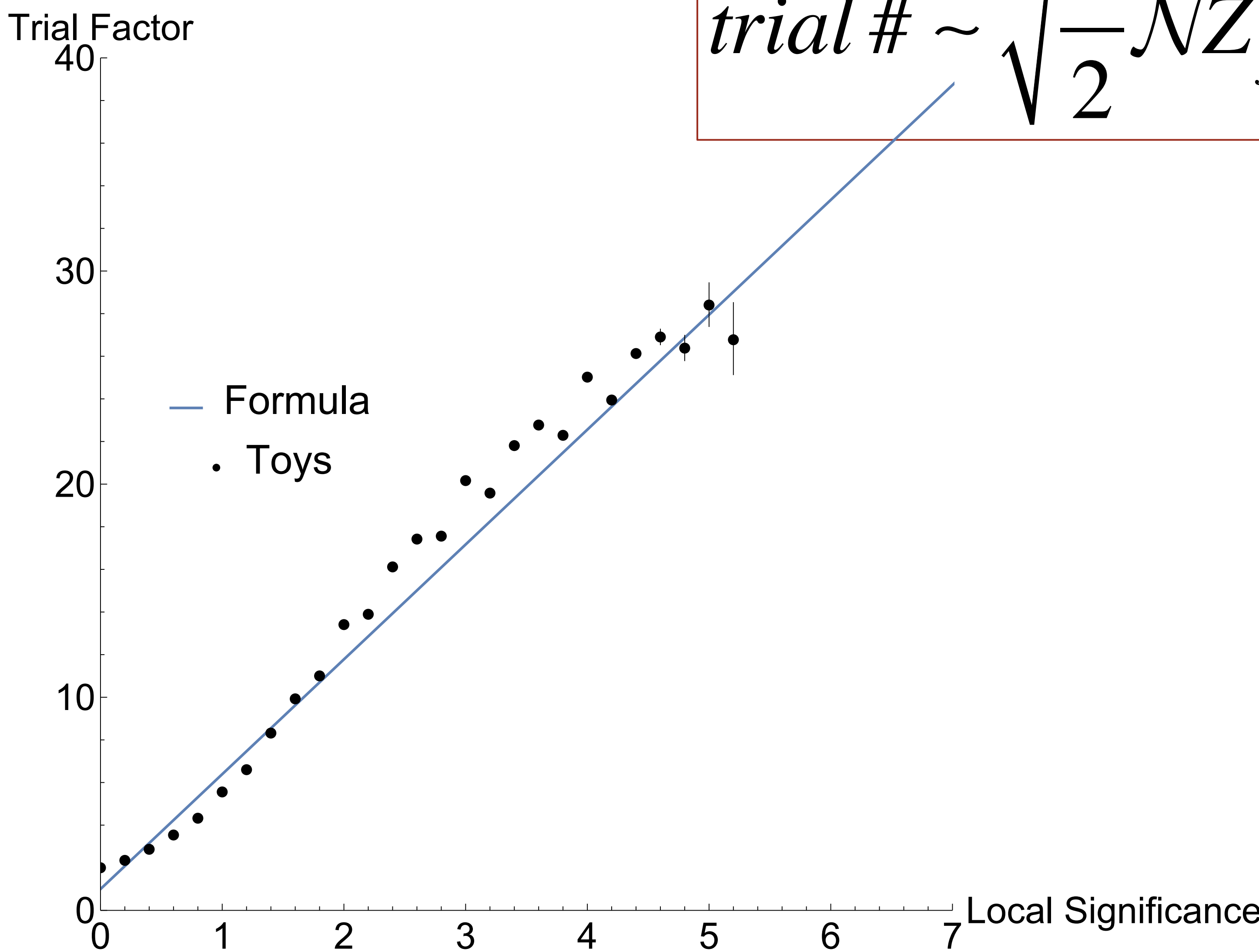

1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
q_0
first $q0_{(m)}$ peak
q0max
q0(125)
$\frac{1}{2}\chi^2_{(1)}$
$\frac{1}{2}\chi^2_{(1)}+\langle N\rangle\chi^2_{(2)}$
$\chi^2_{(2)}$



q0max
q0(125)
$\frac{1}{2}\chi^2_{(1)}$
$\frac{1}{2}\chi^2_{(1)}+\langle N\rangle\chi^2_{(2)}$
1
0.01
10^{-4}
10^{-6}
0
5
10
15
20
25
30
q_0


gloabl significance formula
7
6
5
4
3
2
1
1
2
3
4
5
6
7
global significance

$$trial\,\# \sim \sqrt{\frac{\pi}{2}}\mathcal{N}Z_{fix}$$



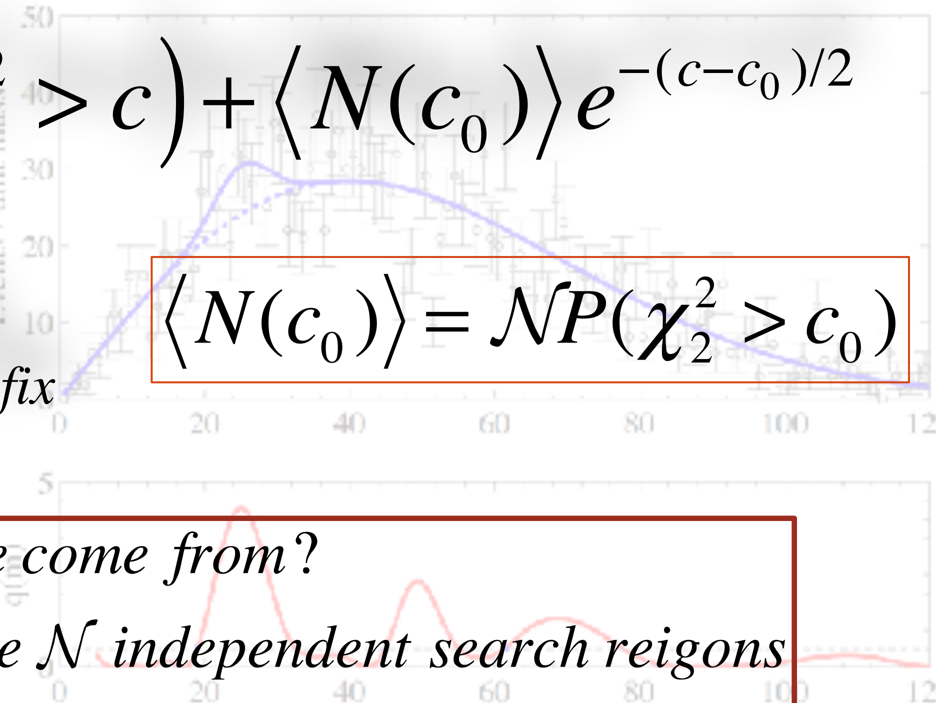
Why Trial#~Z_{fix}?

Solution of the LEE problem

$$P\left(q(\hat{\theta})>c\right)\sim P\left(\chi_1^2>c\right)+\left\langle N(c_0)\right\rangle e^{-(c-c_0)/2}$$

$$trial\,\#\sim 1+\sqrt{\frac{\pi}{2}}\mathcal{N}Z_{fix}$$

$$\left\langle N(c_0)\right\rangle=\mathcal{N}P(\chi_2^2>c_0)$$



Where does the Z dependence come from?

View the results as if there are $\mathcal{N}$ *independent search reigons*

In each one there is a χ_2^2 *distribution of* $q_0(\mu,m)$

The mass is a dof even though it is undefined under the null

$$\sigma_{\hat{m}}\sim\frac{1}{Z}\Rightarrow\frac{\Delta m}{\sigma_{\hat{m}}}\sim Z$$

Why trial#~Z

$$Var(m)=\left[-E\left(\frac{\partial^2\log\mathcal{L}}{\partial m^2}\right)\right]^{-1}$$

$$n\sim Poiss\left(\mu s(m)+b\right)\approx e^{-(\mu s(m)+b)}\left(\mu s(m)+b\right)^n$$

$$\log\mathcal{L}=-\mu s(m)-b+n\log\left(\mu s(m)+b\right)$$

$$\frac{\partial\log\mathcal{L}}{\partial m}=-\mu\frac{\partial s(m)}{\partial m}+n\frac{\mu}{\mu s(m)+b}\frac{\partial s(m)}{\partial m}$$

$$\frac{\partial^2\log\mathcal{L}}{\partial m^2}=-\mu\frac{\partial^2 s(m)}{\partial m^2}+n\frac{\mu}{\mu s(m)+b}\frac{\partial^2 s(m)}{\partial m^2}-n\frac{\mu^2}{\left(\mu s(m)+b\right)^2}\left(\frac{\partial s(m)}{\partial m}\right)^2$$

$$E\left[n\right]=\mu s(m)+b$$

$$E\left[\frac{\partial^2\log\mathcal{L}}{\partial m^2}\right]=-\frac{\mu^2}{\mu s(m)+b}\left(\frac{\partial s(m)}{\partial m}\right)^2$$

$$Var\left[m\right]\sim\frac{1}{\mu}\sim\frac{1}{Z}\Rightarrow\sigma_{\hat{m}}\sim Z\Rightarrow trial\#\sim\frac{range}{\sigma_{\hat{m}}}\sim Z$$

A real life example

$$\mathrm{P}(q_0 > u) \le E[N_u] + \mathrm{P}(q_0(0) > u)$$

$$E[N_u] = \mathcal{N}_1 e^{-u/2}$$

$$\mathcal{N}_1 \cong \langle N_{u_0} \rangle e^{u_0/2}$$

$$\mathrm{P}(q_0 > u) = \mathcal{N}_1 e^{-u/2} + \frac{1}{2} P(\chi_1^2 > u)$$

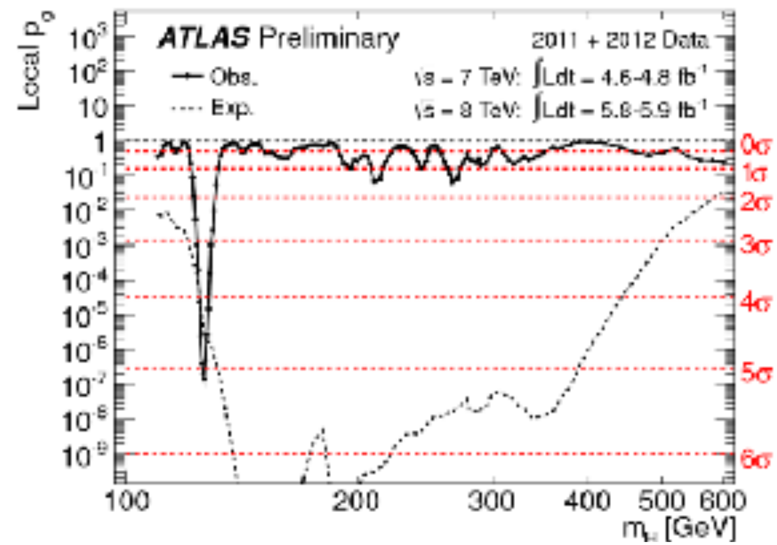
$$p_{global} = \mathcal{N}_1 e^{-u/2} + p_{local}$$

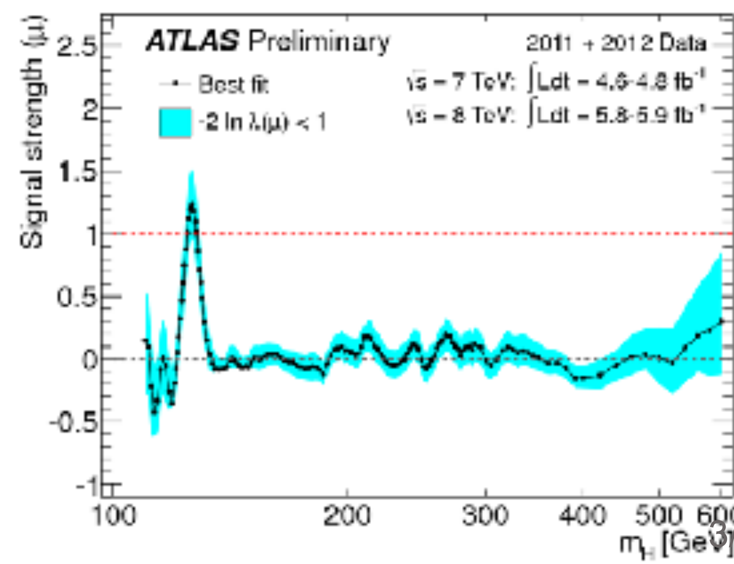
$$p_{global} = \langle N_{u_0} \rangle e^{\frac{u_0 - u}{2}} + p_{local}$$

$$N_{u_0=0} = 9 \pm 3$$

$$p_{global} = 9 \cdot \mathrm{e}^{-25/2} + O(10^{-7}) = 3.3 \cdot 10^{-5}$$

$5\sigma \rightarrow 4\sigma$ trial#~100





Example: The 750 GeV Resonance

Spin 0 2015

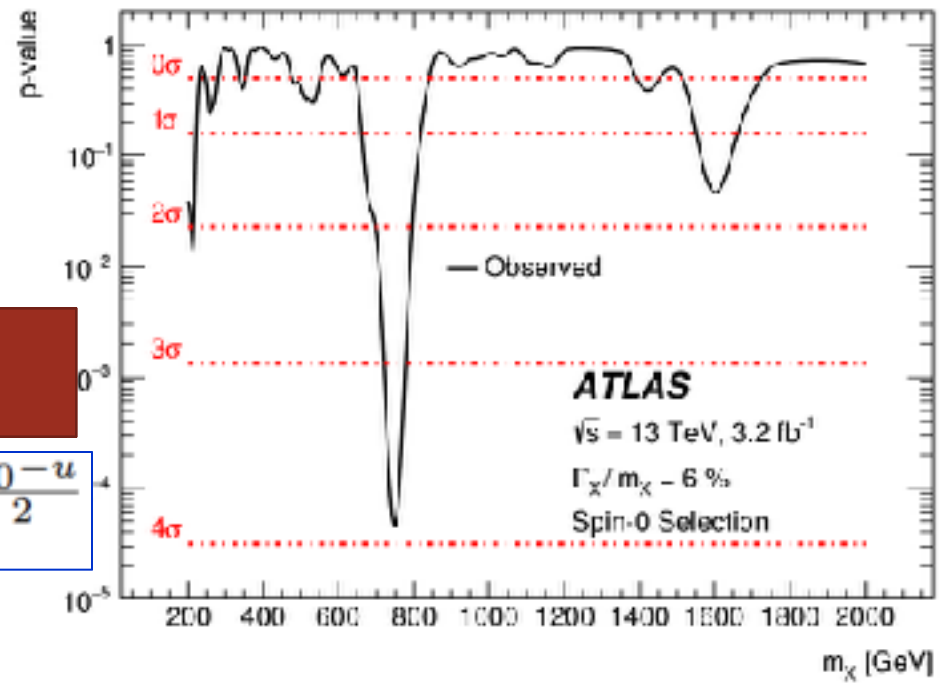
Largest significance

$m_x \sim 750\text{GeV}, \Gamma_x \sim 45\text{GeV}(6$

Local Z = 3.9σ

Any peak with Z>3.8σ
with m=500-2000 will draw our attention

$$P_{global}(u) \approx p_{local}(u) + E(n_{u_0}) e^{\frac{u_0-u}{2}}$$



$$p_{local} = 5 \cdot 10^{-5}$$
$$u_0 = 0$$
$$n_{u_0} = 7 \pm 2.6$$
$$u = Z^2 = 3.9^2 = 15.2$$
$$p_{global} = 5 \cdot 10^{-5} + (7 \pm 2.6)e^{-15.2/2} = (2.2 - 4.8)10^{-3}$$
$$Z_{global} \sim 2.7 \pm 0.1\sigma$$

The LEE is even stronger when you consider another dimension (the width range (0-10%m) should also be taken into account)

The 2D LEE

Define the Problem

- Let $n = \mu s(m,\Gamma) + b$
- m, Γ are nuisance parameters undefined under the null hypothesis $\mu = 0$
- What is the pdf of

$$\hat{q}_0 \equiv q_0(\hat{m},\hat{\Gamma}) = -2\ln\frac{L(\mu=0)}{L(\hat{\mu},\hat{m},\hat{\Gamma})} = \max_{m,\Gamma} q_0(m,\Gamma)$$

under the null hypothesis

Define the Problem

- To generalize the problem, let Θ be the nuisance parameter, undefined under the null hypothesis, and let us try to find out the pdf of

$$\hat{q}_0 \equiv q_0(\hat{\theta}) = -2\ln\frac{L(\mu=0)}{L(\hat{\mu},\hat{\theta})} = \max_{\theta} q_0(\theta)$$

for which we want to calculate

$$p-value = P\left(\max_{\theta}\left[q_0(\theta)\right] \geq u\right),\ u = Z^2$$

Chi Squared Random Field

- For fixed θ $\quad q_0(\theta) = -2\ln \frac{L(\mu=0)}{L(\hat{\mu},\theta)} \sim \chi_1^2$

- $q_0(\theta)$ is a chi squared random field over the space of θ
(a random variable indexed by a continuous parameter(s))

- We are interested in

$$\hat{q}_0 \equiv q_0(\hat{\theta}) = -2\ln \frac{L(\mu=0)}{L(\hat{\mu},\hat{\theta})} = \max_\theta q_0(\theta)$$

for which we want to calculate

$$p-value = P\left(\max_\theta \left[q_0(\theta) \right] \geq u \right), \; u = Z^2$$

Chi Squared Random Field

- We are only interested in positive signals (downward fluctuations of the background are not considered as an evidence against the background)

$$q_0(\theta) = \begin{cases} -2\log \dfrac{L(\mu = 0)}{L(\hat{\mu},\theta)} \\ 0 \end{cases}$$

$$q_0(\theta) \sim \frac{1}{2}\chi_1^2$$

[H. Chernoff, Ann. Math. Stat. 25, 573578 (1954)]

Chi Squared Random Field

- We are only interested in positive signals
(downward fluctuations of the background are not considered as an evidence against the background)

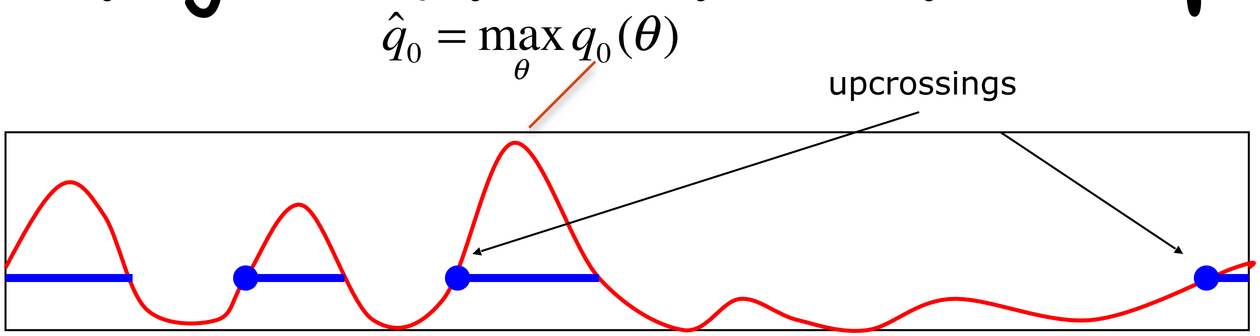
$$q_0(\theta) = \begin{cases} -2\log\frac{L(\mu=0)}{L(\hat{\mu},\theta)} \\ 0 \end{cases} \qquad q_0(\theta) \sim \frac{1}{2}\chi_1^2$$

[H. Chernoff, Ann. Math. Stat. 25, 573578 (1954)]

- $q_0(\theta) = \left(\frac{\hat{\mu}(\theta)}{\sigma}\right)^2$ $\hat{\mu}(\theta)$ is a Gaussuan Random Field over θ

1-D Random Fields

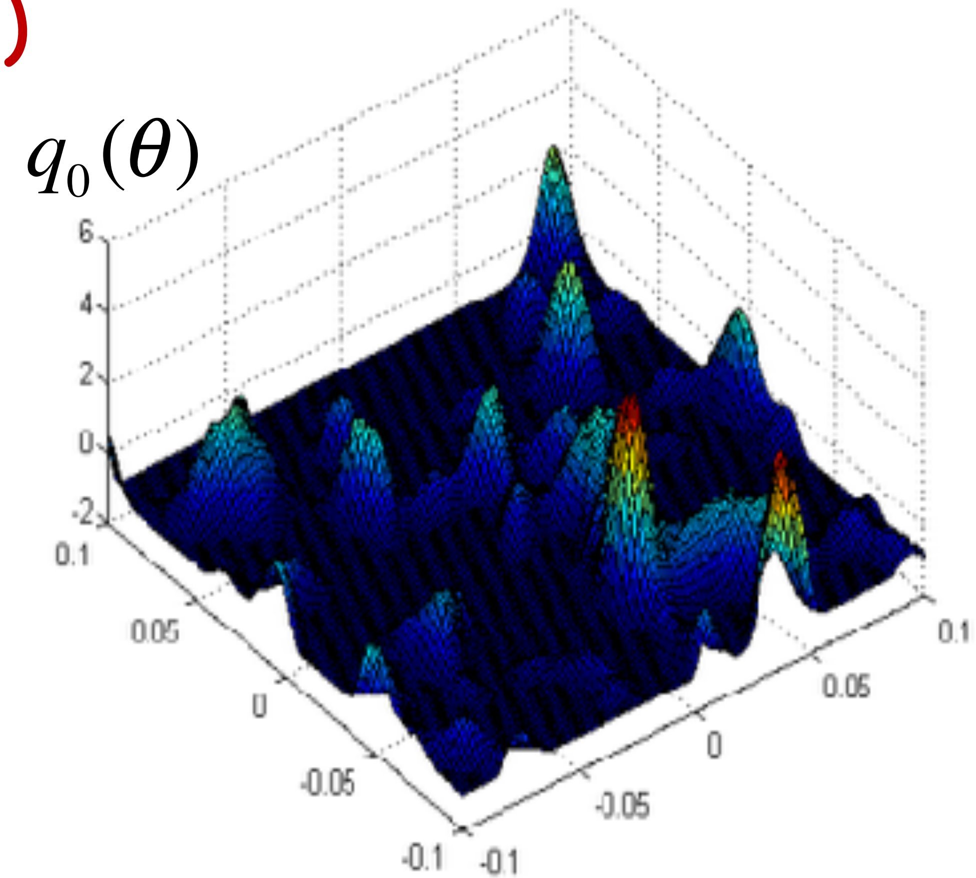
- In 1-D points where the field becomes larger than u are called upcrossings.

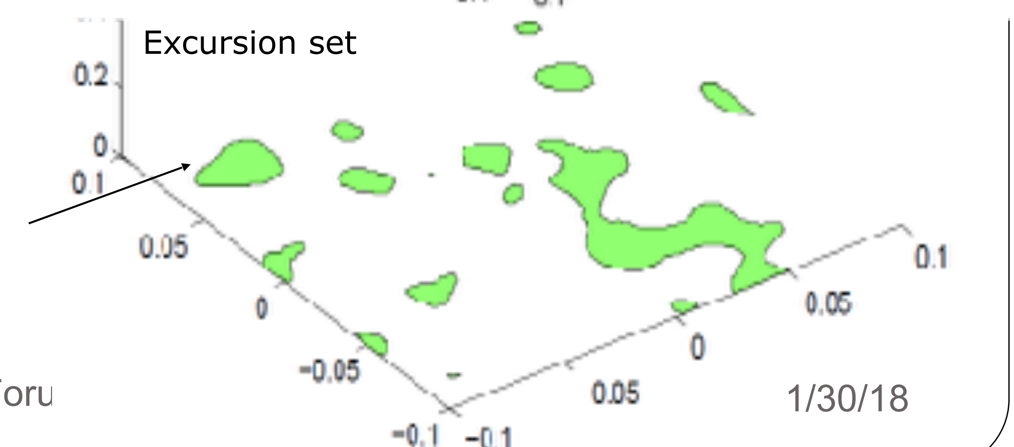


- The probability that the global maximum is above the level u is called exceedance probability. (p-value of $q_0(\hat{\theta})$) $p = P\left(\max_{\theta}\left[q_0(\theta)\right] \geq u\right), \; u = Z^2$

Random fields (>1 D)

- The set of points where the field has values larger then some number u is called the *excursion set A_u above the level u.*





Random fields

- Fortunately, quite a lot of statistical literature on the properties of random fields in D- dimensions
- Applications in Cosmology, Brain mapping, Oceanography …

[3] R.J. Adler and A.M. Hasofer, *Level Crossings for Random Fields*, Ann. Probab. 4, Number 1 (1976), 1-12.

[4] R.J. Adler, *The Geometry of Random Fields*, New York (1981), Wiley, ISBN: 0471278440.

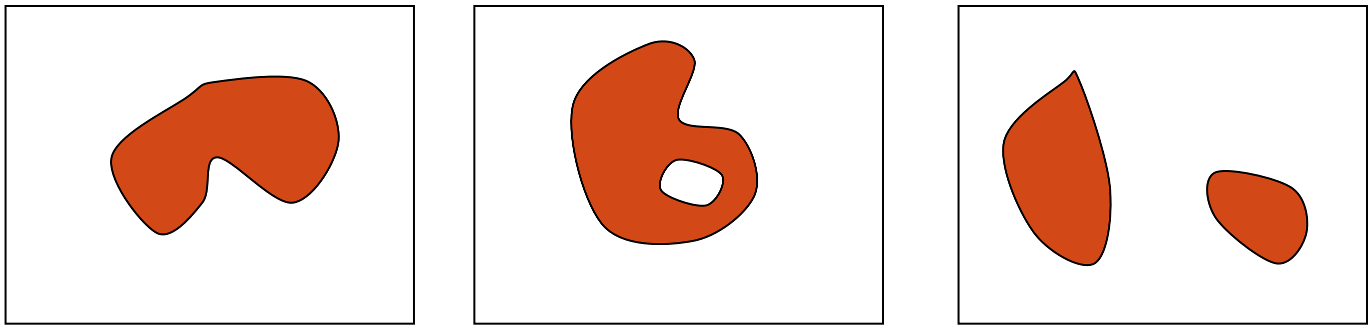
[5] K.J. Worsley, S. Marrett, P. Neelin, A.C. Vandal, K.J. Friston and A.C. Evans, *A Unified Statistical Approach for Determining Significant Signals in Location and Scale Space Images of Cerebral Activation*, Human Brain Mapping 4 (1996) 58-73.

[6] R.J. Adler and J.E. Taylor, *Random Fields and Geometry* , Springer Monographs in Mathematics (2007). ISBN: 978-0-387-48112-8.

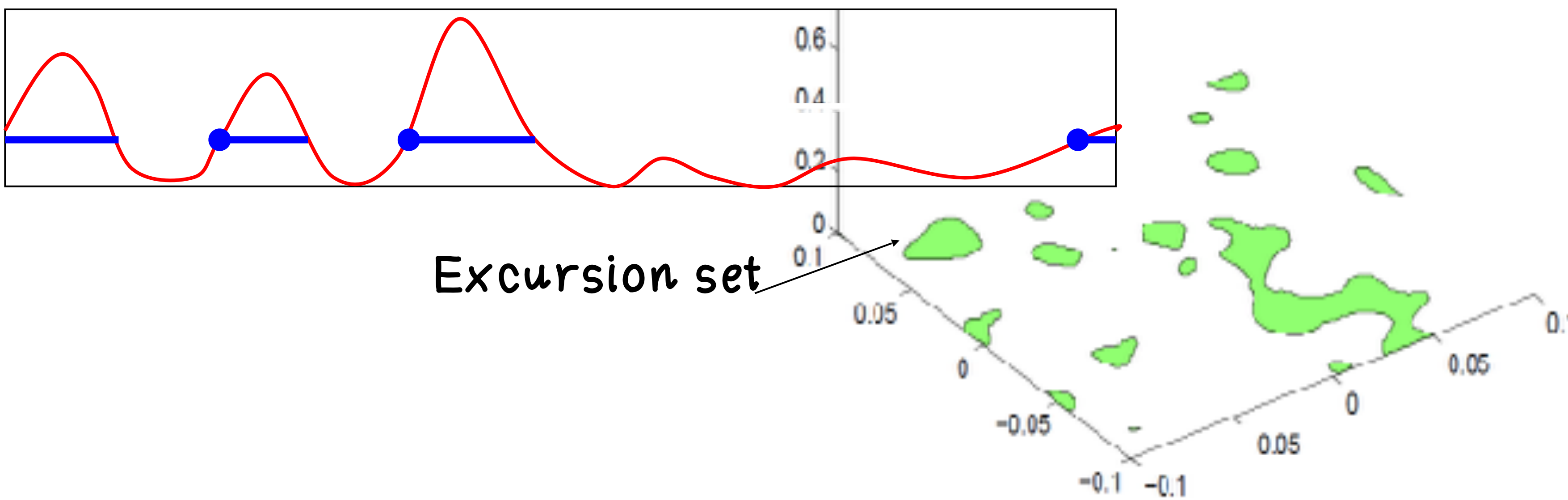
[9] J. Taylor, A. Takemura and R.J. Adler, *Validity of the expected Euler characteristic heuristic*, Ann. Probab. 33 (2005) 1362-1396.

Euler characteristic

- Number of disconnected components minus number of 'holes'



$\varphi=1$ $\varphi=0$ $\varphi=2$



The n-dumensional case

[R.J. Adler and J.E. Taylor, *Random Fields and Geometry* (2007), Springer Monographs in Mathematics]

- The upcrossings formula is a special case of a more general result which gives the expectation of the Euler characteristic of the excursion set of a random field over a general n-dimensional manifold

$$E\left[\vartheta(A_u)\right] = \sum_{d=0}^{D} \mathcal{N}_d \rho_d^s(u)$$

A_u is the excursion set of the field above a level u

(set of points where $q_0(\theta)>u$)

$\varphi(A_u)$ is it's Euler characteristic

ρ_d are 'universal' functions

(depend only on the level u and s, number of poi)

Adleret. al. Formula

$$E\left[\vartheta(A_u)\right]=\sum_{d=0}^{D}\mathcal{N}_d\,\rho_d^s(u)$$

n is the Dimension
(number of Nuisance parameters
undefined under the null hypothesis)

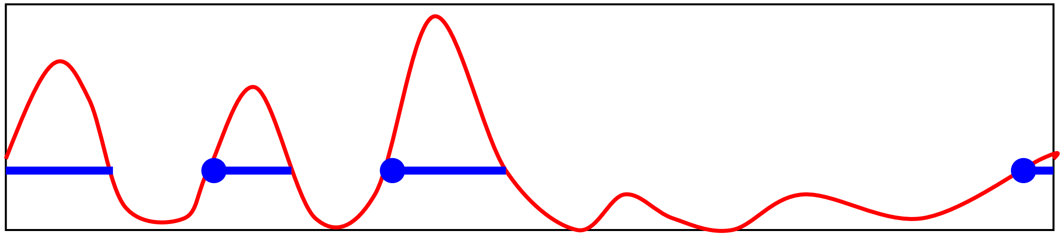
- For a Chi Squared field with
  s parameters of interest and dimension D

$D=1, s\ poi$	$D=2, s\ poi$	$D=2, s=1$
$\rho_0(u)=P(\chi_1^2>u)$	$\rho_0(u)=P(\chi_2^2>u)$	$\rho_0(u)=P(\chi_2^2>u)$
$\rho_1(u)=u^{(s-1)/2}e^{-u/2}$	$\rho_1(u)=u^{(s-1)/2}e^{-u/2}$	$\rho_1(u)=e^{-u/2}$
$-$	$\rho_2(u)=u^{(s-2)/2}\left(u-(s-1)\right)e^{-u/2}$	$\rho_2(u)=\sqrt{u}\,e^{-u/2}$

$$1D:\ E\left[\vartheta(A_u)\right]=\frac{1}{2}P(\chi_1^2>u)+\mathcal{N}_1 e^{-u/2}$$

$$2D:E\left[\vartheta(A_u)\right]=\frac{1}{2}P(\chi_2^2>u)+\left(\mathcal{N}_1+\mathcal{N}_2\sqrt{u}\right)e^{-u/2}$$

1D Euler characteristic



$$E[\varphi(A_u)] = \sum_{d=0}^{n} N_d \rho_d(u)$$

In 1 dimension:

$$\varphi(A_u) = N_u + 1_{[q_0(0)>u]}$$

$$E[\varphi(A_u)] = E[N_u] + P(q_0(0) > u)$$

$$= N_0 P(\chi_1^2 > u) + N_1 e^{-u/2}$$

$$N_0 = \varphi(manifold) = 1$$

$$E[\varphi(A_u)] = P(\chi_1^2 > u) + N_1 e^{-u/2}$$

This is Davies Formula

The general case

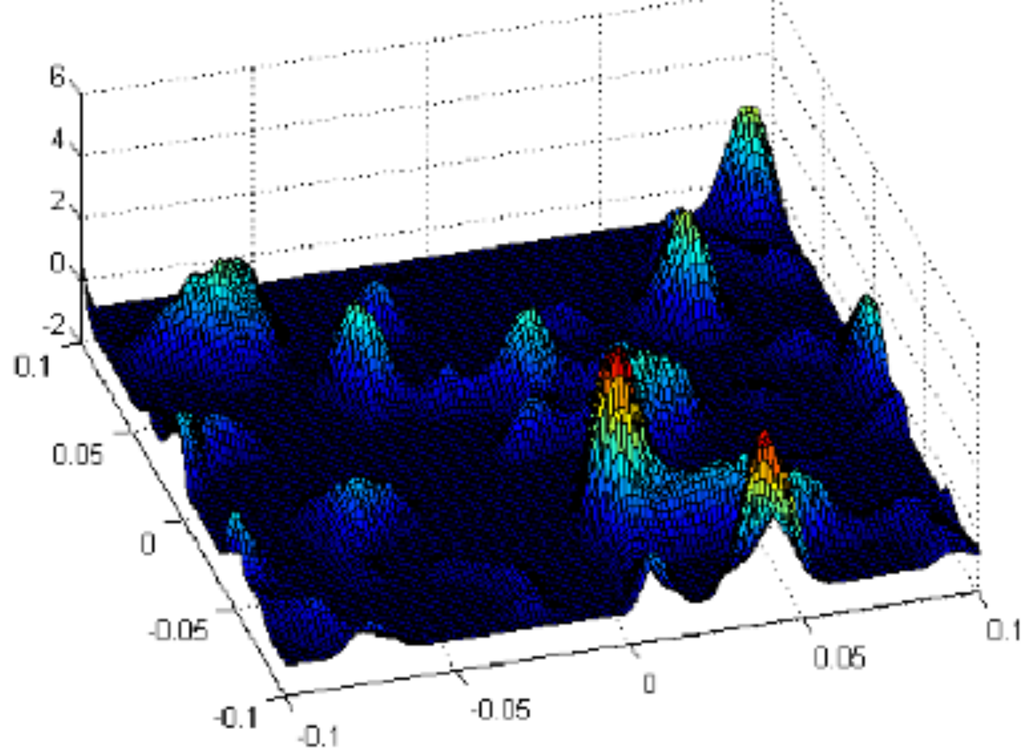
$$N_0 = \varphi(manifold)$$

$$\rho_0(u) = P(\chi_s^2 > u)$$

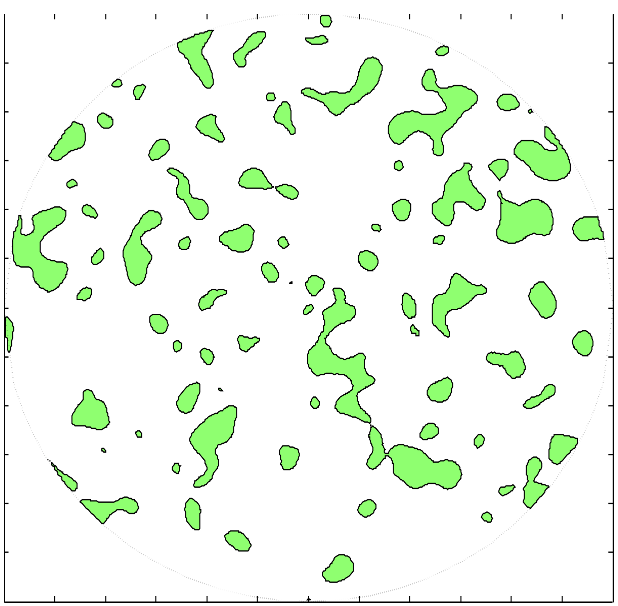
In general for high-level excursions

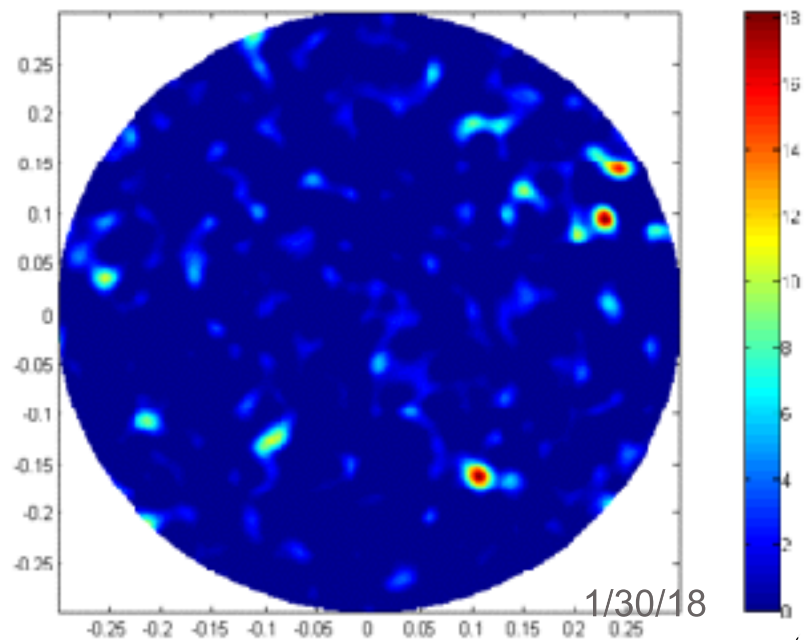
$$E[\varphi(A_u)] \xrightarrow[u \gg 1]{} P\left(\max_\theta [q_0(\theta)] \geq u\right)$$

[J. Taylor, A. Takemura, and R. J. Adler, Ann. Probab. 33, 4 (2005)]

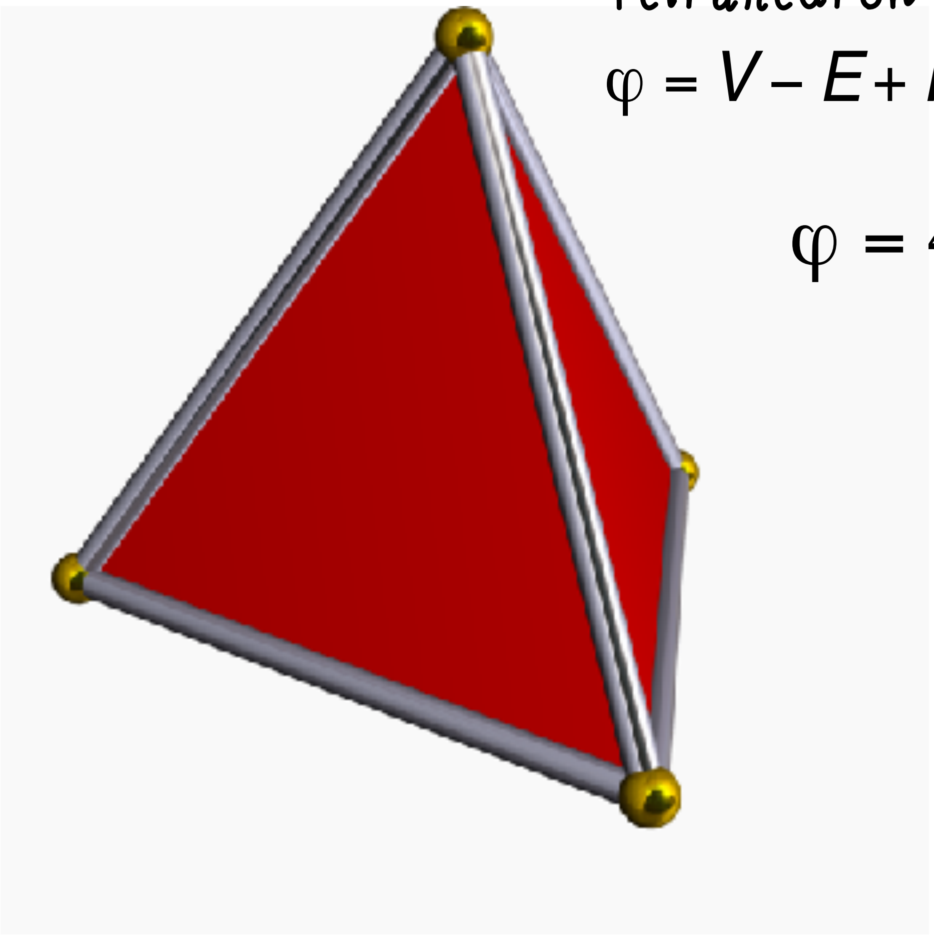


Excursion set
(u=1)





Calculation of the Euler characteristic

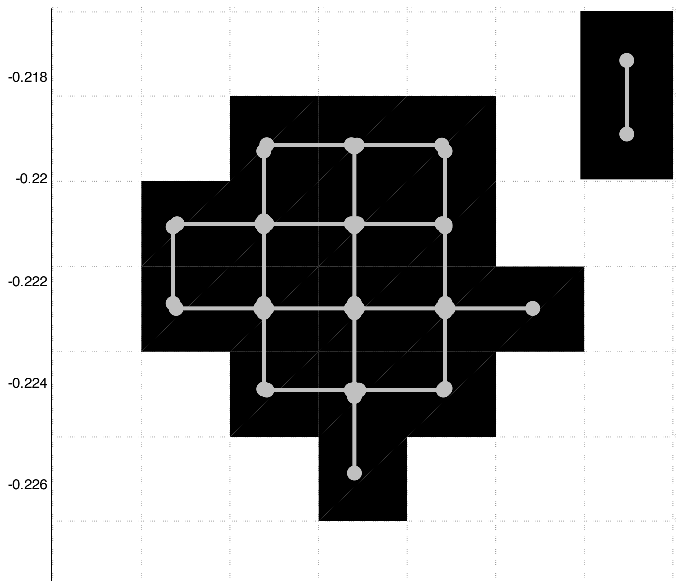


Tetrahedron

$$\varphi = V - E + F = \#vertices - \#edges + \#faces$$

$$\varphi = 4 - 6 + 4 = 2$$

Calculation of the Euler characteristic



- Usually we have $q(\theta)$ calculated on a grid of points
- Calculation of the E.C. is straightforward:
- φ = #vertices - #edges + #faces
- Generalizes to higher dimensions

$$\varphi = 18(\text{points}) - 23(\text{edges}) + 7(\text{faces}) = 2$$

2-d example: search for neutrino sources (IceCube)

For a chi[2] field in 2 dimensions:

$$E\left[\vartheta(A_u)\right] = \frac{1}{2}P(\chi^2_2 > u) + \left(\mathcal{N}_1 + \mathcal{N}_2\sqrt{u}\right)e^{-u/2}$$

Estimate E[φ] at two levels, e.g. 0 and 1, and solve for N_1 and N_2

From 20 bkg. Simulations:

$\langle\varphi_0\rangle = 33.5 \pm 2$

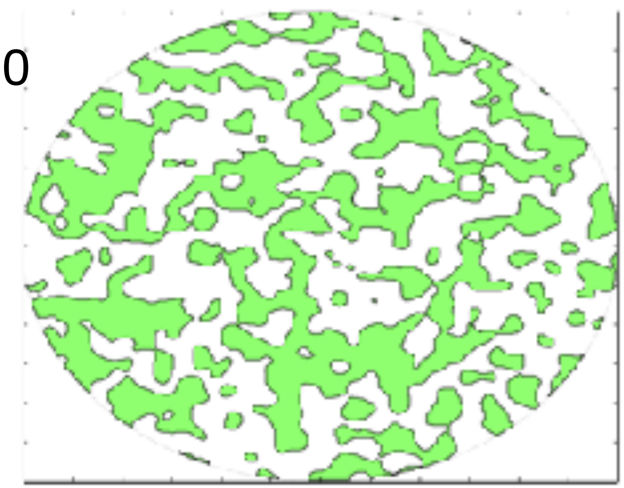
$\langle\varphi_1\rangle = 94.6 \pm 1.3$

⇩

$N_1 = 33 \pm 2$

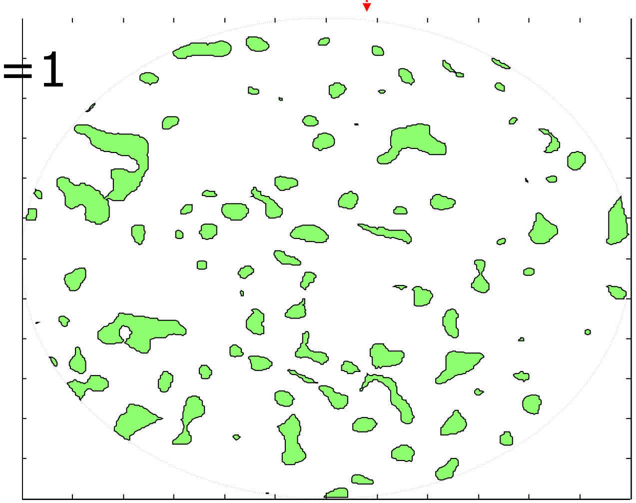
$N_2 = 123 \pm 3$

u=0



φ=35

u=1



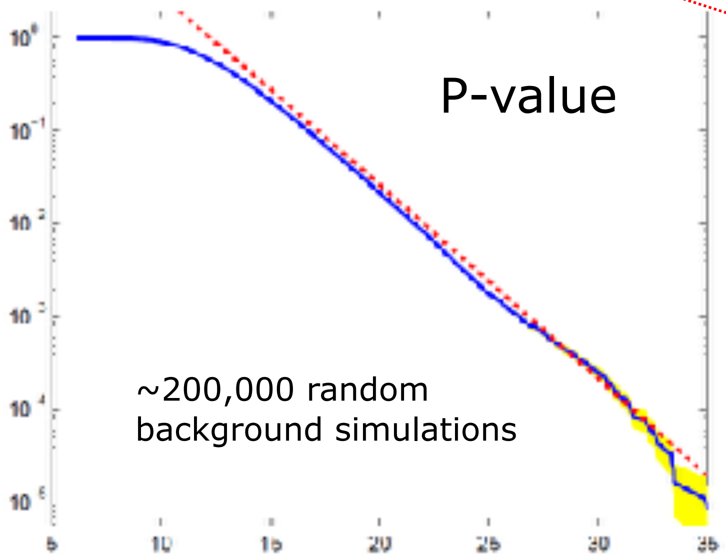
φ=95

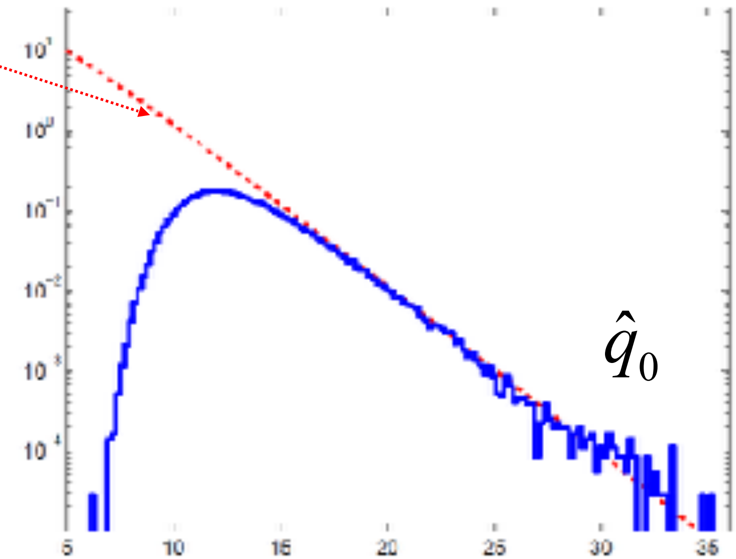
2-d example: search for neutrino sources (IceCube)

$$E\left[\vartheta(A_u)\right] = \frac{1}{2}P(\chi_2^2 > u) + \left(\mathcal{N}_1 + \mathcal{N}_2\sqrt{u}\right)e^{-u/2}$$

$\mathcal{N}_1 = 33 \pm 2$

$\mathcal{N}_2 = 123 \pm 3$





e.g.: P(max q_0>30) = $(2.5 \pm 0.4)\text{x}10^{-4}$ (estimated)

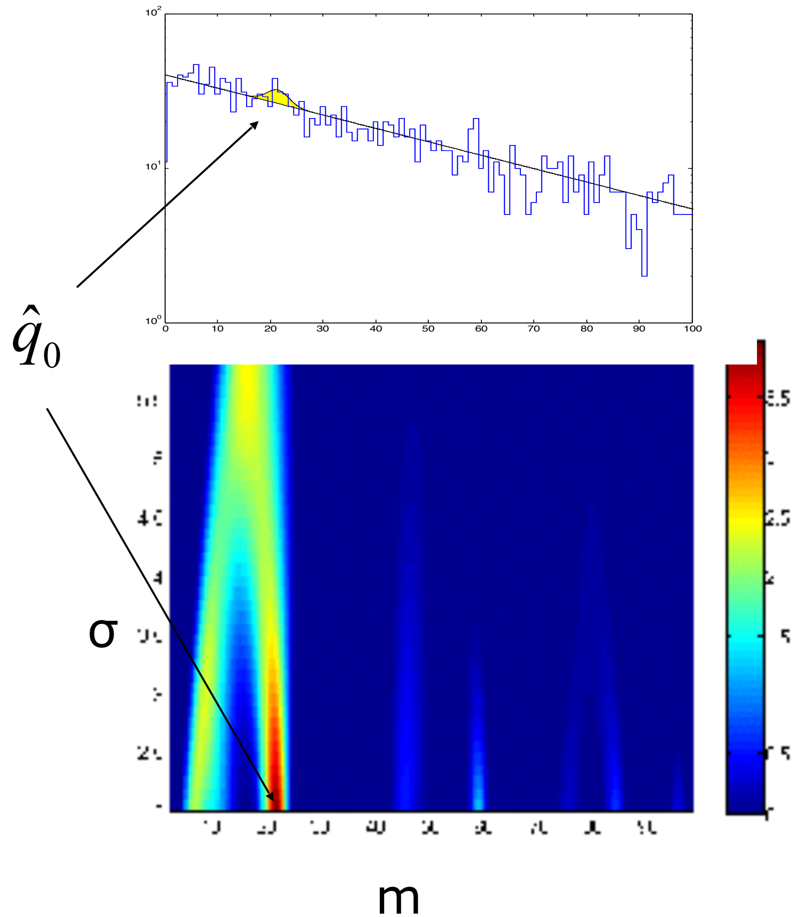
E.C. Formula : $(2.28 \pm 0.06)\text{x}10^{-4}$

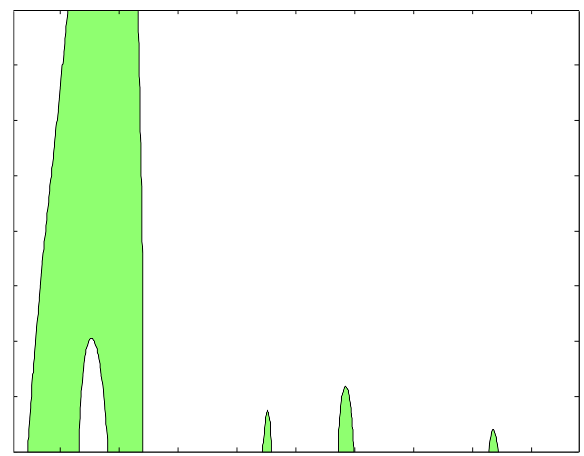
2-D exapmle #2: resonance search with unknown width

- Gaussian signal on exponential background
- Toy model : 0<m<100 ,   2<σ<6
- Unbinned likelihood:

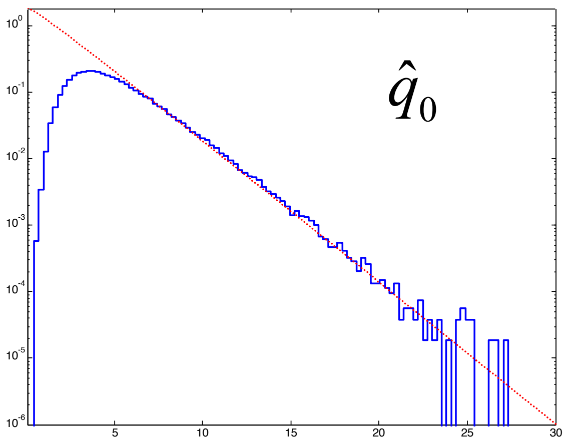
$$L = \prod_i \frac{N_s f_s(x_i) + N_b f_b(x_i)}{N_s + N_b} \times Poiss(N \mid N_s + N_b)$$

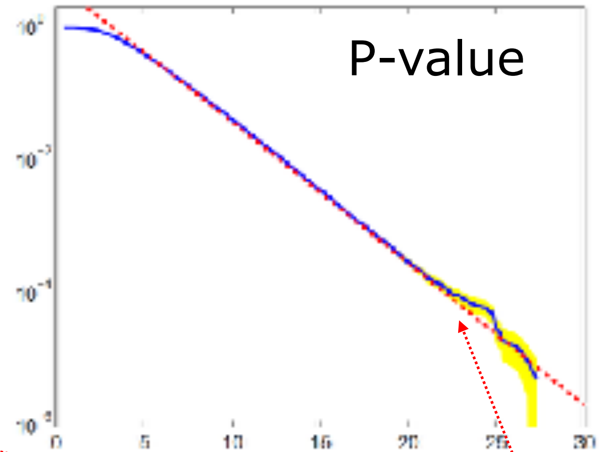
$$f_s(x;m,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \qquad f_b(x) = ce^{-cx}$$





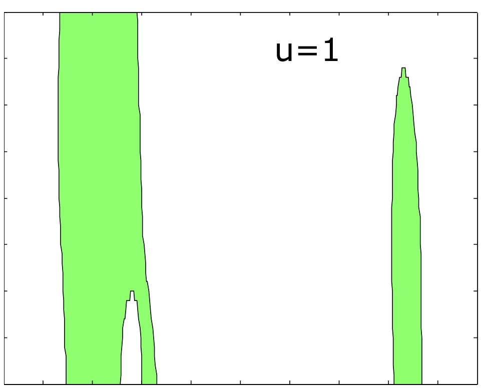
2-D exapmle #2: resonance search with unknown width



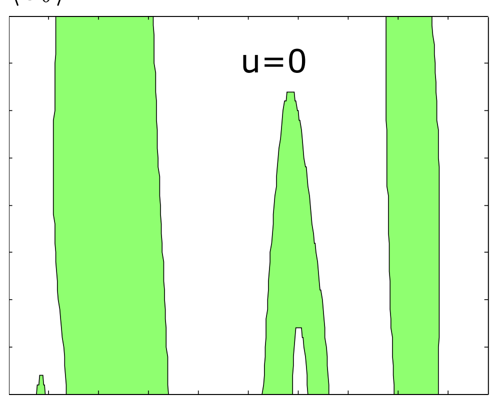


Excellent approximation above the ~2σ level

$\langle \varphi_1 \rangle = 3 \pm 0.16$



$\langle \varphi_0 \rangle = 4.5 \pm 0.2$



$$E[\vartheta(A_u)] = \frac{1}{2}P(\chi_2^2 > u) + \left(\mathcal{N}_1 + \mathcal{N}_2\sqrt{u}\right)e^{-u/2}$$

$$\mathcal{N}_1 = 4 \pm 0.2$$

$$\mathcal{N}_2 = 0.7 \pm 0.3$$

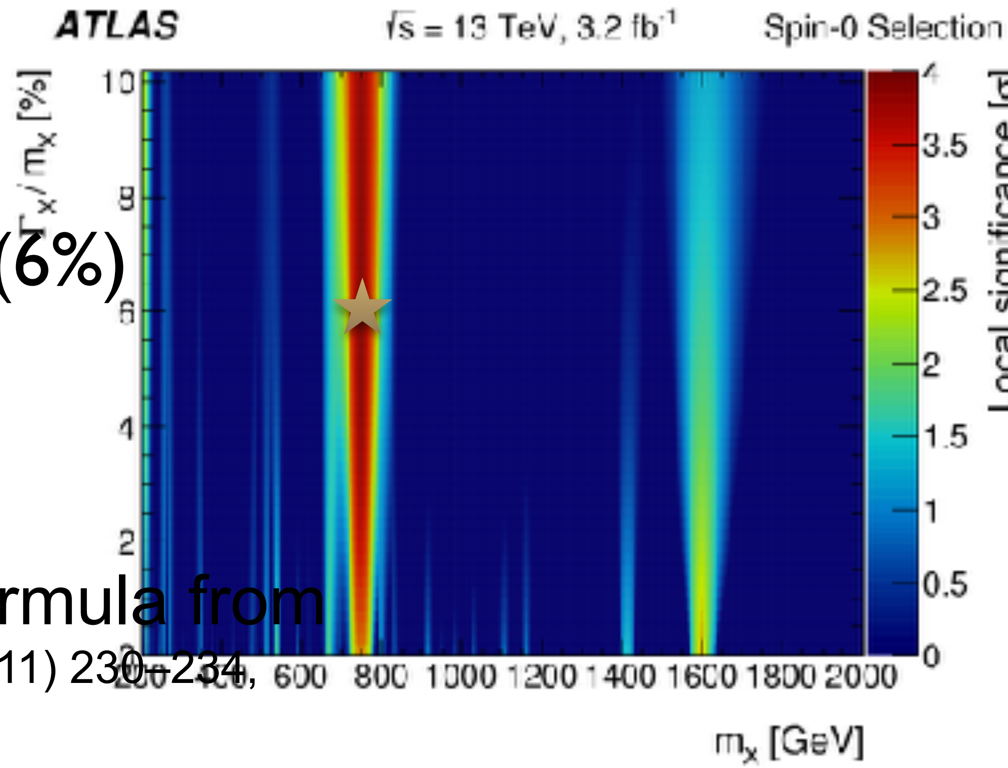
2015

2D Scan
Largest significance

$m_x \sim 750\text{GeV}, \Gamma_x \sim 45\text{GeV}(6\%)$

Local Z = **3.9σ**

m=200-2000 GeV
Γ_X/m_X=0-10%



Use toys or asymptotic formula from
O. Vitells et. al. Astropart. Phys. 35 (2011) 230-234,
arXiv:1105.4355

$$Z_{local} = 3.9\sigma$$

$$Z_{global} = 2.1\sigma$$

2.1σ is not something to write home about

Summary

$$p_{global}(s=1,D=1) \approx E[\vartheta(A_u)] = \frac{1}{2}P(\chi_1^2 > u) + \mathcal{N}_1 e^{-u/2}$$

$$p_{global}(s=1,D=2) \approx E[\vartheta(A_u)] = \frac{1}{2}P(\chi_2^2 > u) + \left(\mathcal{N}_1 + \mathcal{N}_2\sqrt{u}\right)e^{-u/2}$$

- The procedure for estimating the p-value is simple and reliable.
- The Euler characteristic formula provides a practical way of estimating the look-elsewhere effect.
- It is easily expandable to s p.o.i and D NPs (undefined under the null hypothesis)

End of Lectures
Thank You

Eilam Gross