Look elsewhere effect

Ofer Vitells

Statistics miniworkshop at CERN , February 2013

LEE Topics

- Introduction
  - Definition of gaussian & gaussian-related fields
- Z-dependence of trial factor
  - Variance of m-hat
  - Bayesian comparison
- Different possibilities for critical region
  - Constatnt LR (“Tevatron” test statistic) curves
  - Leadbetter formula
- Location (“energy-scale”) uncertainties
  - Single channel
  - Combination
- Approximation/estimation problems
  - Sliding window effect on upcrossings counting
  - Uncertainty on observed number of upcrossings (poisson?)
  - When asymptotic formulae break down in practice

- Gaussian & Gaussian related fields

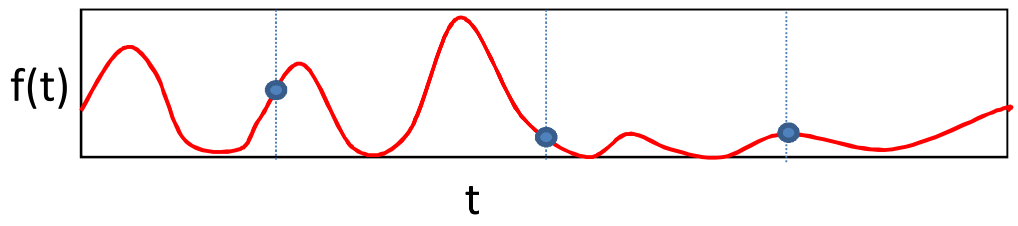
  - The joint distribution of any collection $\{f(t_1), f(t_2), \ldots, f(t_n)\}$ is multivariate Gaussian
  - *Gaussian related* fields are functions of Gaussian fields, e.g.

$$\chi_k^2(t) = \sum_{i=1}^{k} f_i^2(t)$$

(chi-squared field)

Wilks :

$$q(\theta) = -2\log\frac{\mathcal{L}(\mu=0)}{\mathcal{L}(\hat{\mu},\theta)} \sim \chi^2$$



- Z-dependance

$$\text{p-value}=\mathrm{P}(\max_{\theta}[q_0(\theta)]\geq u)\ \approx \frac{1}{2}P(\chi_1^2>u)+\mathcal{N}_1 e^{-u/2}=p_{local}+\mathcal{N}_1 e^{-Z^2/2}$$

Trial-factor $\approx \mathcal{N}_1 Z$

- Variance of $\hat{m}$:

Example : $n_i \sim Poiss(\mu s_i(m)+b_i)$

In the large sample limit $$Var[\hat{m}]\simeq(-E[\frac{\partial^2 \log L}{\partial m^2}])^{-1}$$

$$-E[\frac{\partial^2 \log L}{\partial m^2}]=\mu^2\sum_i \frac{1}{(\mu s_i(m)+b_i)^2}\left(\frac{\partial s_i(m)}{\partial m}\right)^2$$

$$\Rightarrow \quad Var[\hat{m}]\propto\frac{1}{\mu^2},\quad \sigma_{\hat{m}}\propto\frac{1}{\mu}\propto\frac{1}{Z}\qquad (b_i\gg\mu s_i)$$

$$TF\propto\frac{range}{\sigma_{\hat{m}}(Z)}$$

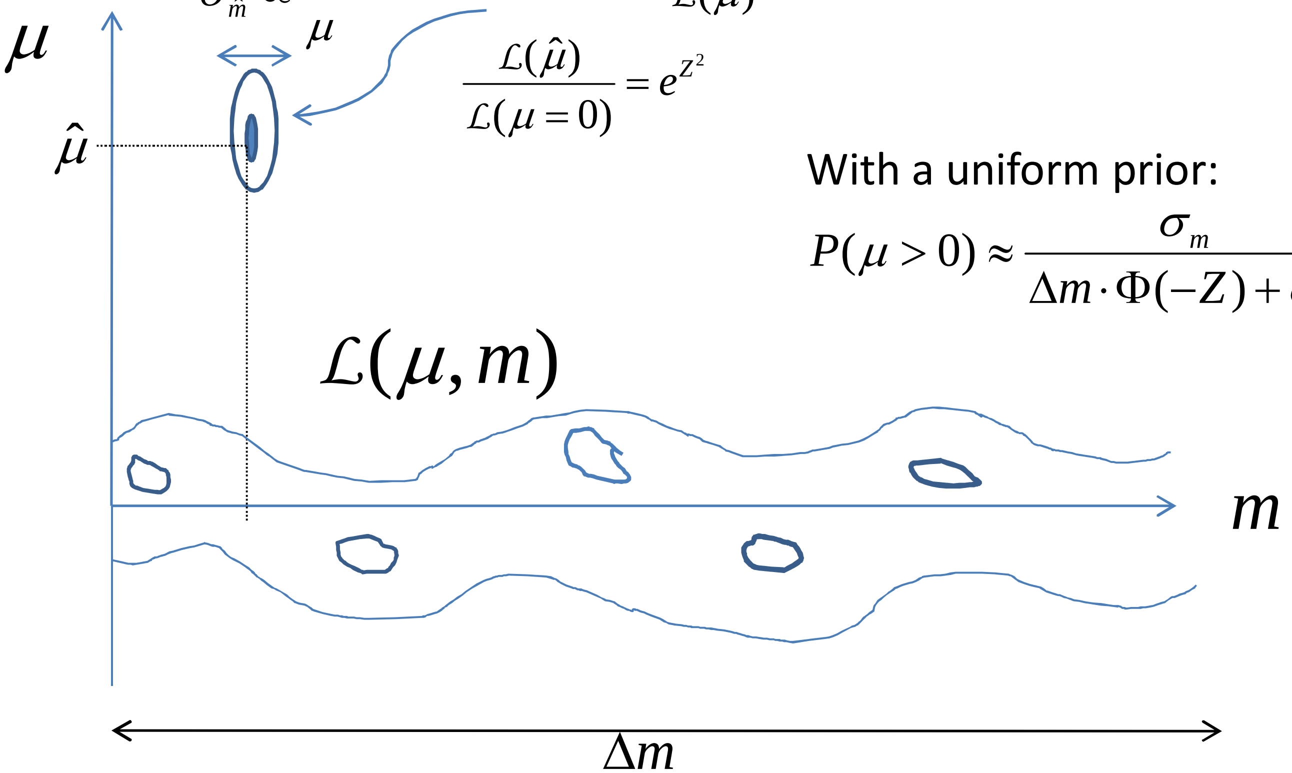
Bayesian estimate

There is less posterior probability in the peak as it narrows (~1/Z)

$$\sigma_{\hat{m}} \propto \frac{1}{\mu}$$

$$q = -2\log\frac{\mathcal{L}(\mu=0)}{\mathcal{L}(\hat{\mu})} = Z^2$$

$$\frac{\mathcal{L}(\hat{\mu})}{\mathcal{L}(\mu=0)} = e^{Z^2}$$



With a uniform prior:

$$P(\mu>0) \approx \frac{\sigma_m}{\Delta m \cdot \Phi(-Z) + \sigma_m} \approx 1 - \frac{\Delta m}{\sigma_m(Z)}\Phi(-Z)$$

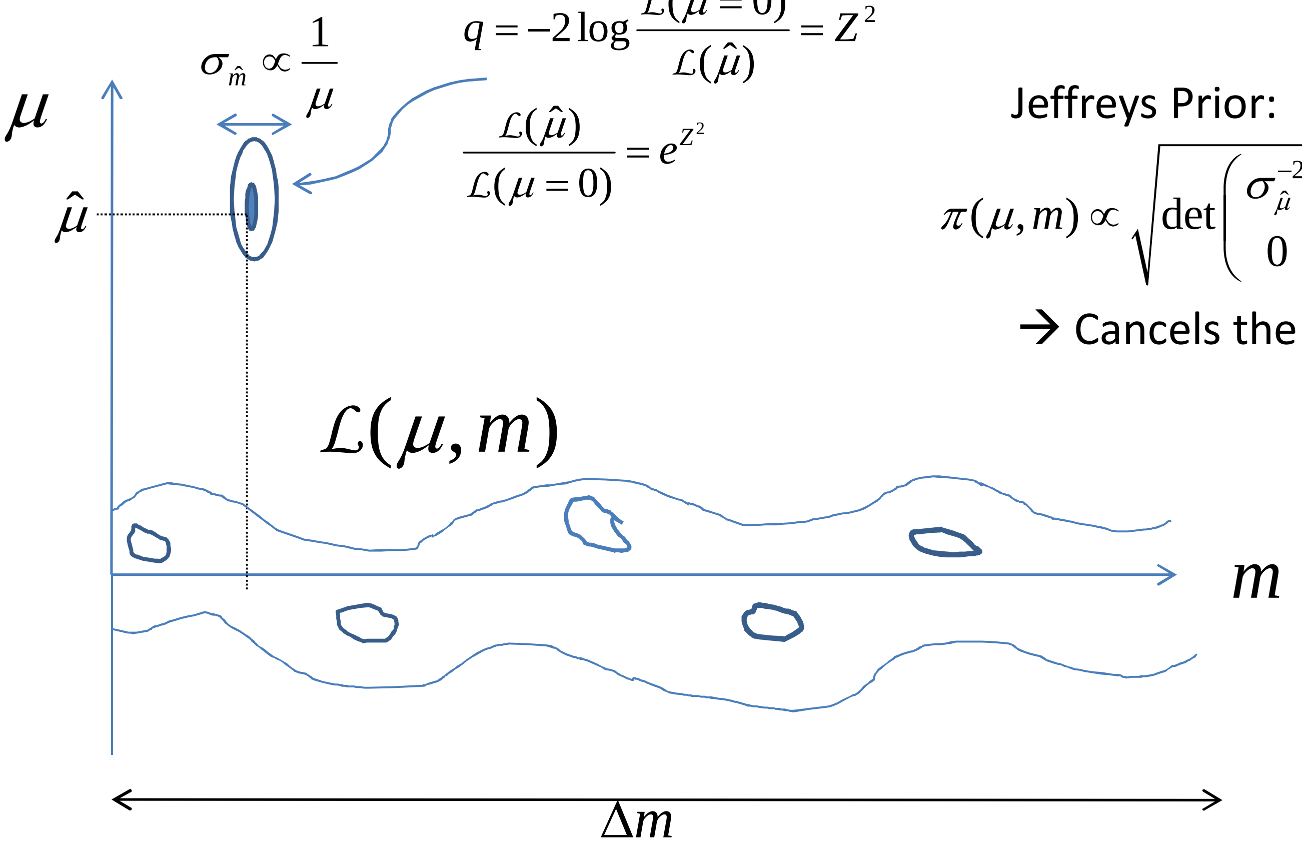
“Trial factor”

Bayesian estimate

There is less posterior probability in the peak as it narrows (~1/Z)

$$q = -2\log\frac{\mathcal{L}(\mu=0)}{\mathcal{L}(\hat{\mu})} = Z^2$$

$$\frac{\mathcal{L}(\hat{\mu})}{\mathcal{L}(\mu=0)} = e^{Z^2}$$



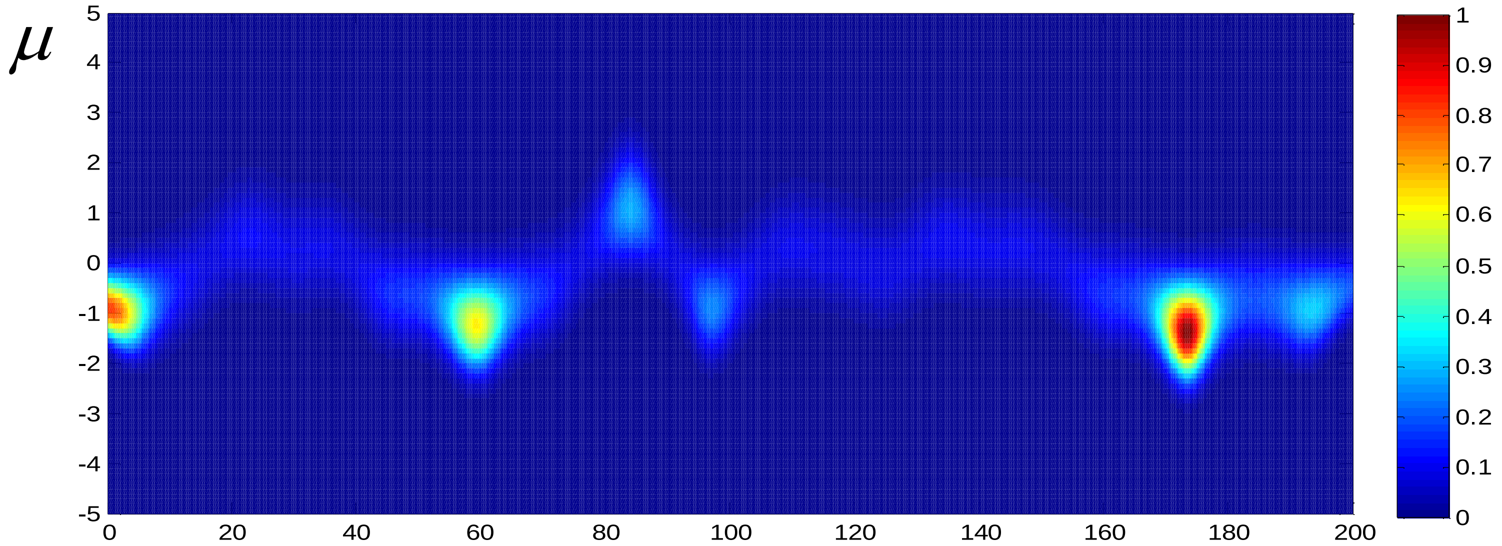
Jeffreys Prior:

$$\pi(\mu, m) \propto \sqrt{\det\begin{pmatrix} \sigma_{\hat{\mu}}^{-2} & 0 \\ 0 & \mu^2\sigma_{\hat{m}}^{-2}(1) \end{pmatrix}} \propto \mu$$

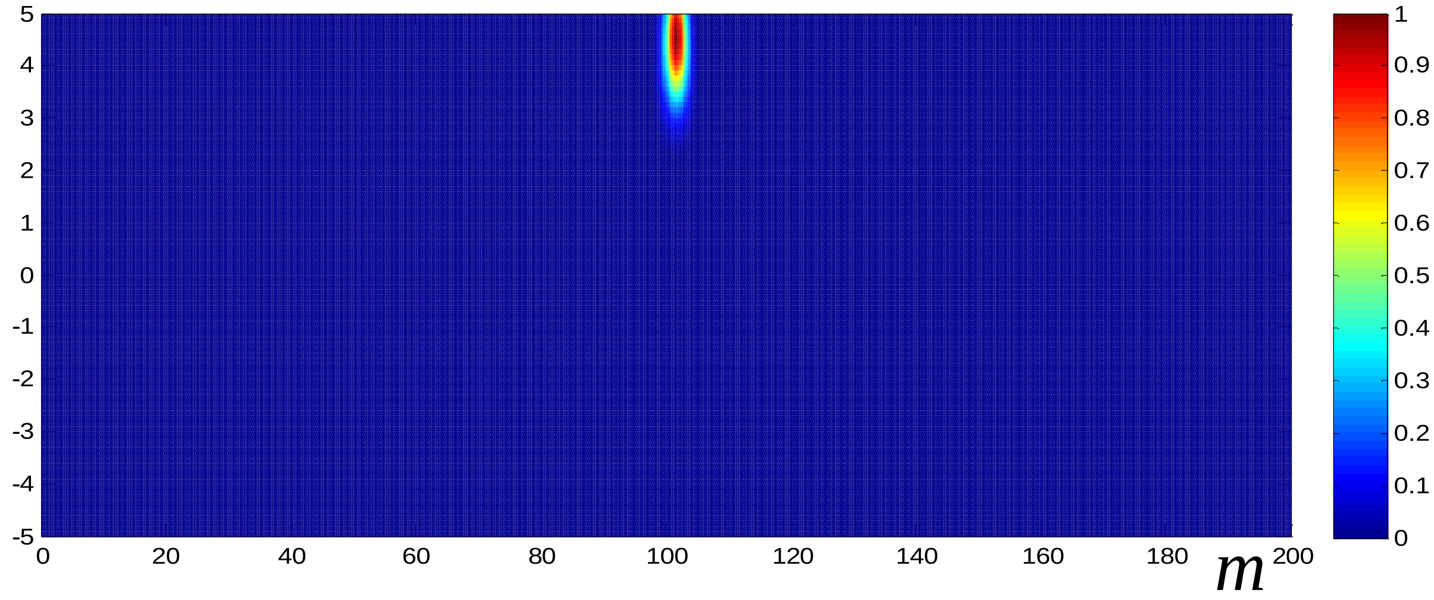
→ Cancels the Z dependence

Example normalized likelihoods

Background

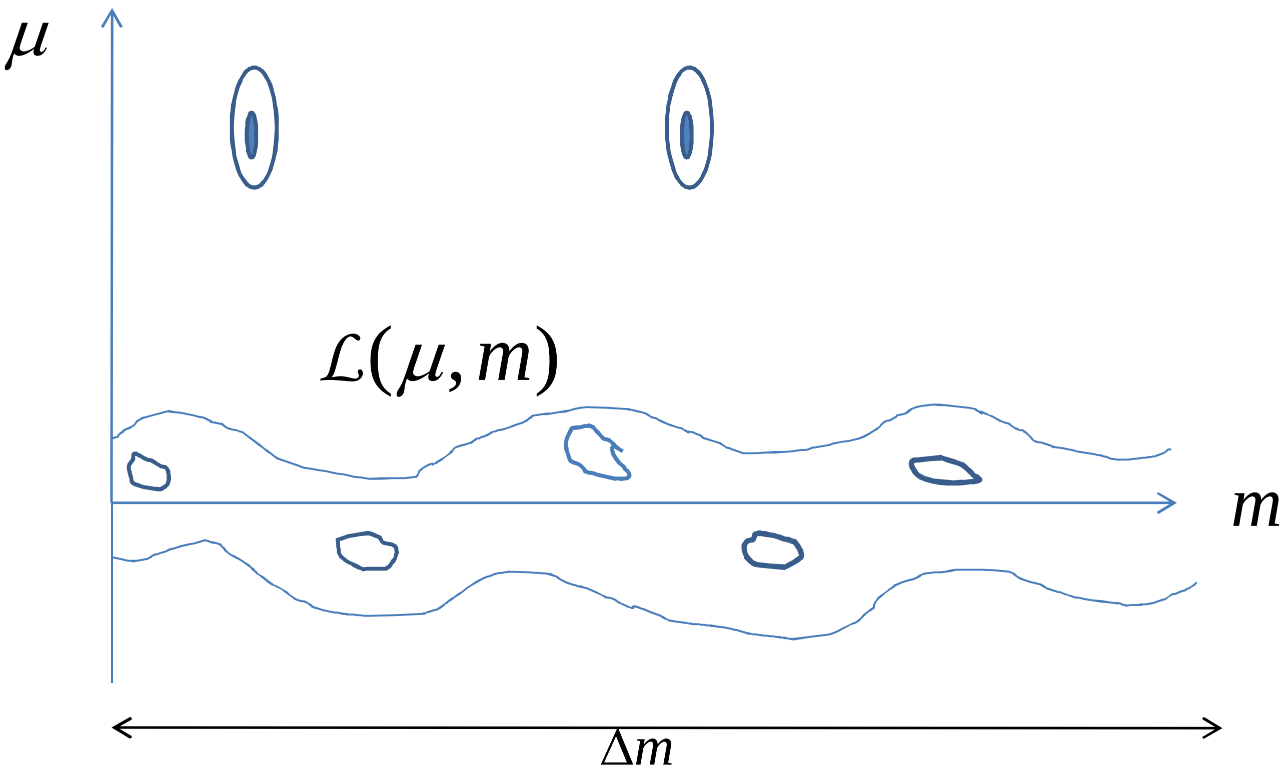


Background +
signal



Bayesian estimate

Integrate m ?



what exactly is meant by 'more impressive than observed in data' ? q(PL) vs. q_{TEV}

$$q = -2\log\frac{L(0)}{L(\hat{\mu})} = \left(\frac{\hat{\mu}}{\sigma}\right)^2$$

observed significance : $Z^{obs} = \sqrt{q} = \dfrac{\hat{\mu}}{\sigma}$

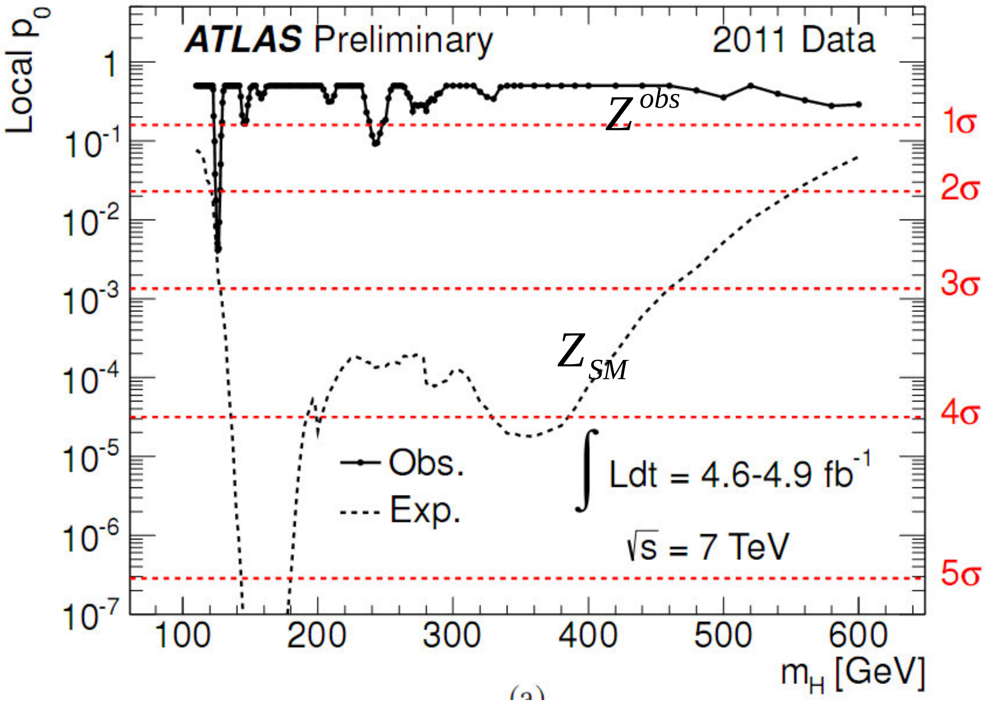
$$q_{TEV} = -2\log\frac{L(0)}{L(\mu_{SM})} = \left(\frac{\hat{\mu}}{\sigma}\right)^2 - \left(\frac{\hat{\mu}-\mu_{SM}}{\sigma}\right)^2 = 2\left(\frac{\hat{\mu}}{\sigma}\right)\left(\frac{\mu_{SM}}{\sigma}\right) - \left(\frac{\mu_{SM}}{\sigma}\right)^2$$

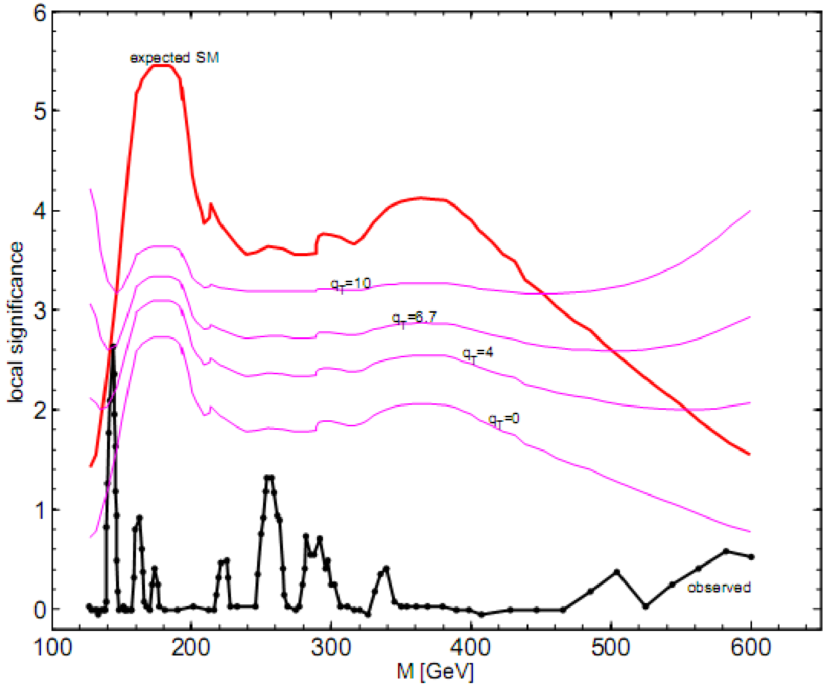
SM Expected significance (sensitivity): $Z_{SM} = \left(\dfrac{\mu_{SM}}{\sigma}\right)$ $\Rightarrow$ $\boxed{q_{TEV} = 2Z^{obs}Z_{SM} - Z_{SM}^2}$

At a given mass point the two tests are equivalent (1-to-1 functions of $\hat{\mu}$),

But give different answers to what is the “best fit” mass - $\max_m[q(m)]$ or $\max_m[q_{TEV}(m)]$

* note q_{TEV} = 0 if Z_{SM}=0. max[q_{TEV}] is generally not at the point of largest local significance.

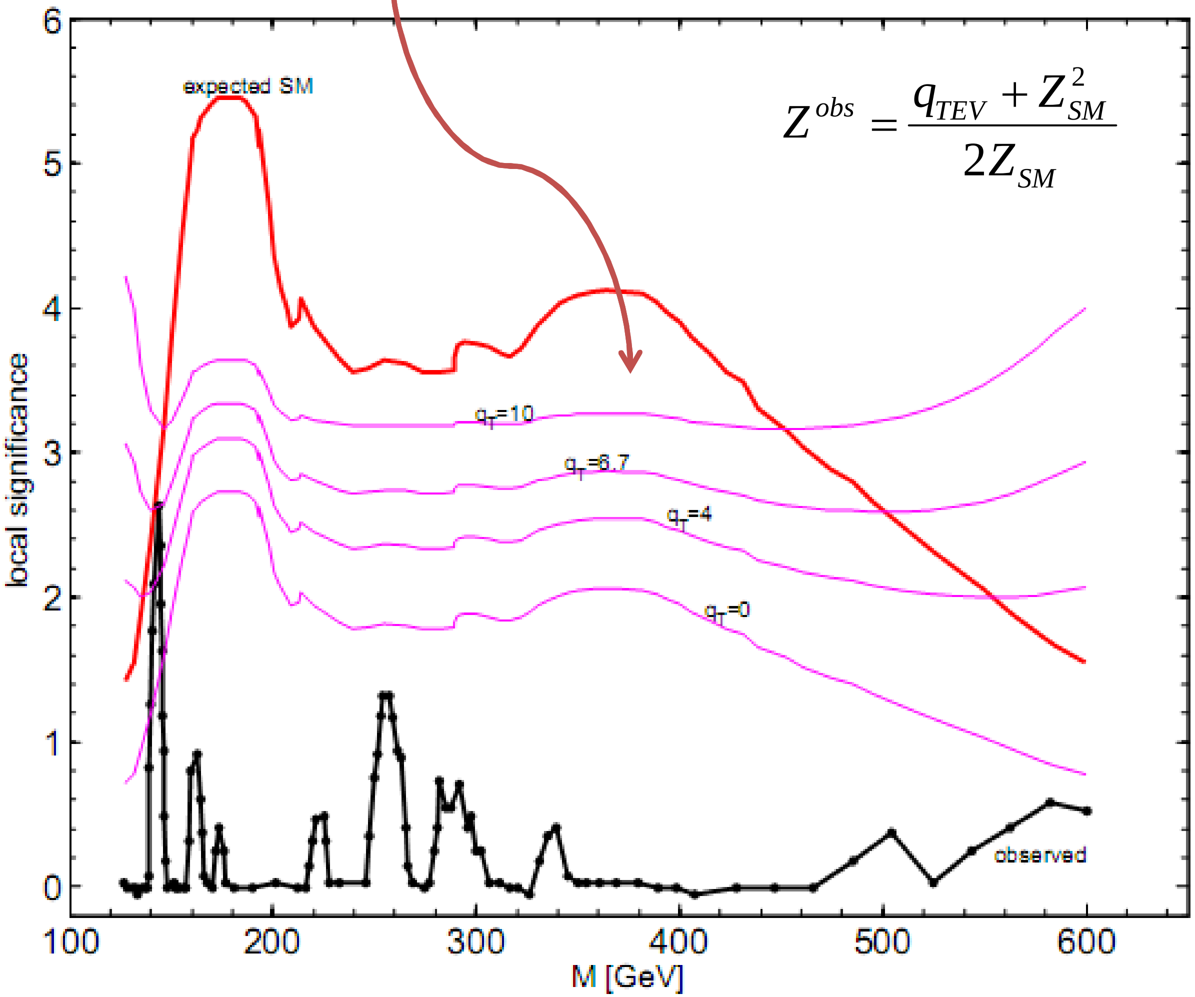




$$\boxed{q_{TEV} = 2Z^{obs} Z_{SM} - Z_{SM}^2}$$

$$Z^{obs} = \frac{q_{TEV} + Z_{SM}^2}{2Z_{SM}}$$

Curves of constant q_{TEV}

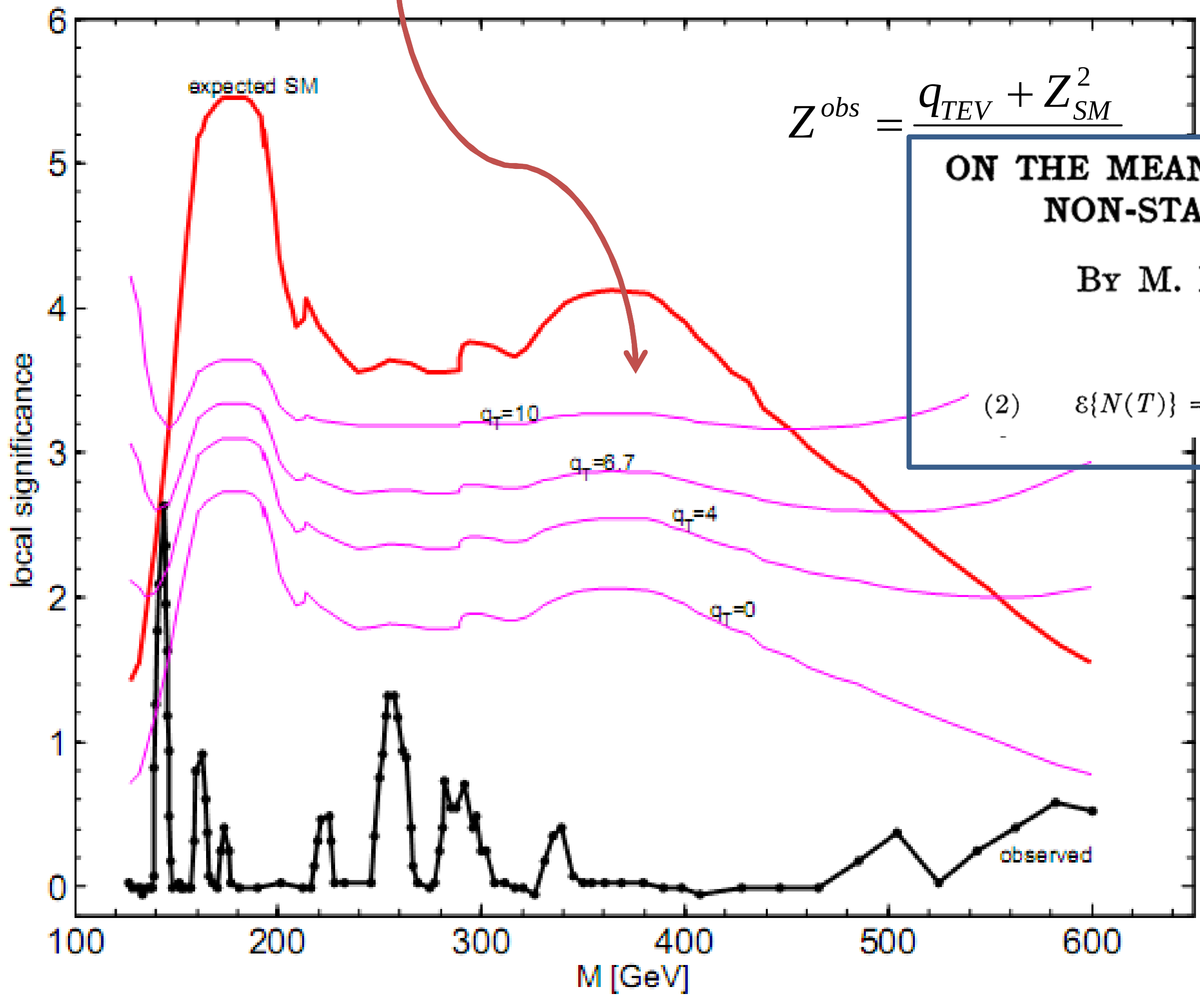


p-value =
Prob(max[q_{TEV}] > c)

A similar signal at 160 GeV would give much smaller global significance (because less consistent with the SM) - same as local 1σ @ 600 GeV

Curves of constant q_{TEV}

p-value = Prob(max[q_{TEV}] > c)



$$Z^{obs} = \frac{q_{TEV} + Z_{SM}^2}{\quad}$$

ON THE MEAN NUMBER OF CURVE CROSSINGS BY NON-STATIONARY NORMAL PROCESSES[1]

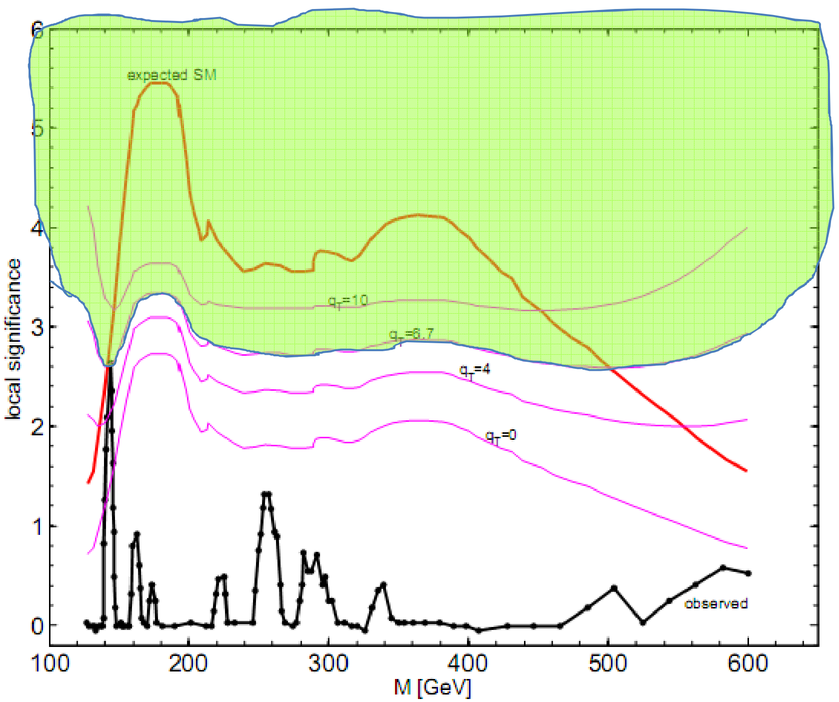
By M. R. Leadbetter and J. D. Cryer

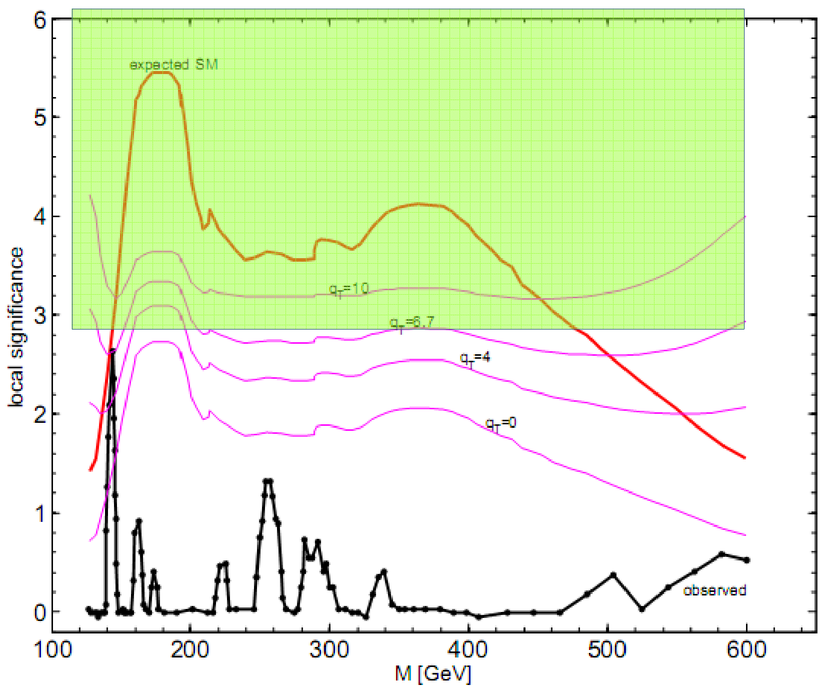
Research Triangle Institute

(2) $\quad \mathcal{E}\{N(T)\} = \int_0^T \gamma\sigma^{-1}(1 - \rho^2)^{\frac{1}{2}}\phi(m/\sigma)\{2\phi(\eta) + \eta(2\Phi(\eta) - 1)\}\, dt$

A similar signal at 160 GeV would give much smaller global significance (because less consistent with the SM) - same as local 1σ @ 600 GeV

Where to put the critical region





Energy-scale uncertainties

$$\mathcal{L}(\mu, M_0, \delta) = \mathcal{L}_0(\mu, M_0(1+\delta))G(\delta^*|\delta, \sigma_{ES}).$$

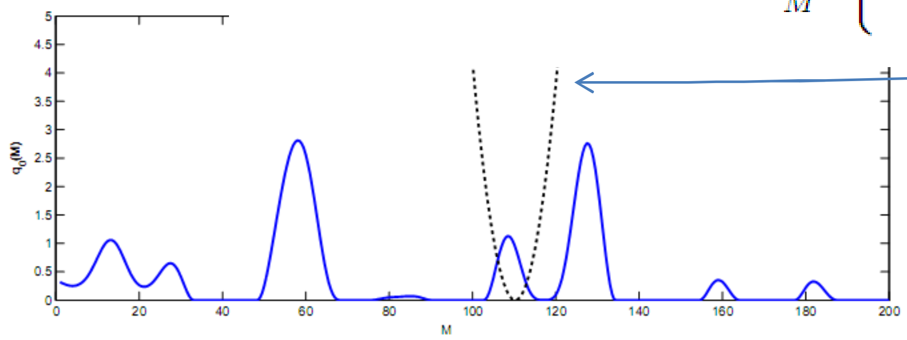
Energy-scale nuisance parameter

Likelihood at a **fixed** mass M_0

Defining $M \equiv M_0(1+\delta)$, this can be written as

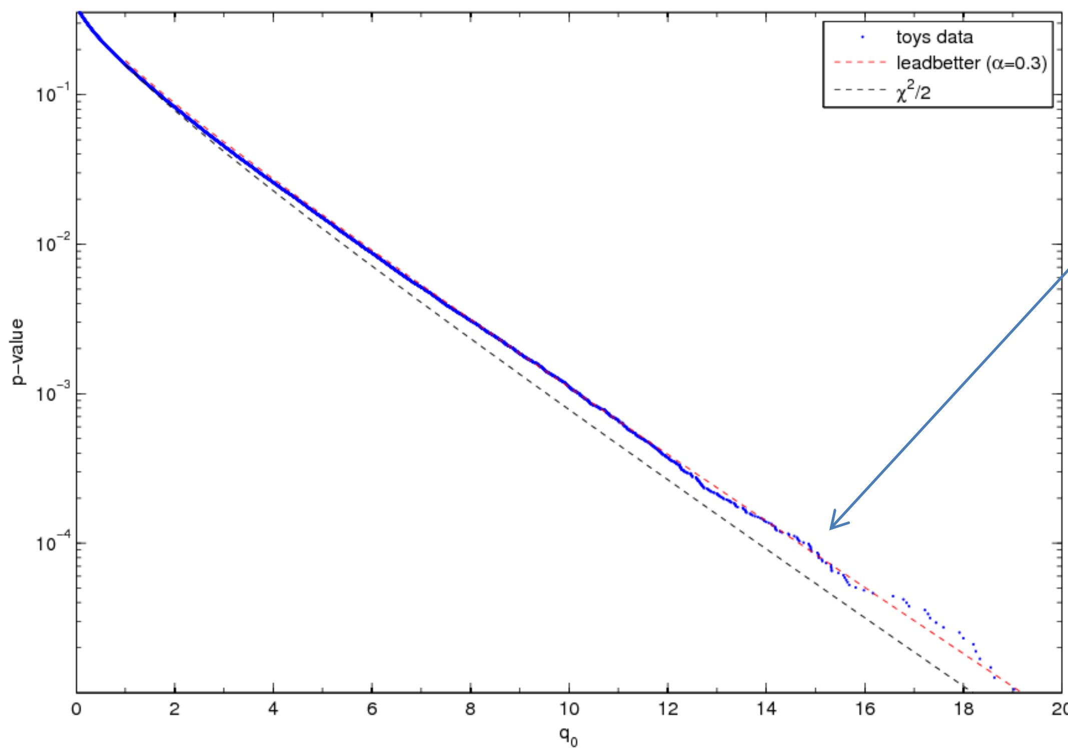
$$q(M_0) = \max_M \left\{ 2\log\frac{\mathcal{L}_0(\hat{\mu}, M)}{\mathcal{L}_0(\mu=0)} - \left(\frac{M/M_0 - 1 - \delta^*}{\sigma_{ES}}\right)^2 \right\}$$

$$= \max_M \left\{ q_0(M) - \left(\frac{M/M_0 - 1 - \delta^*}{\sigma_{ES}}\right)^2 \right\}$$



"local" LEE (Leadbetter)

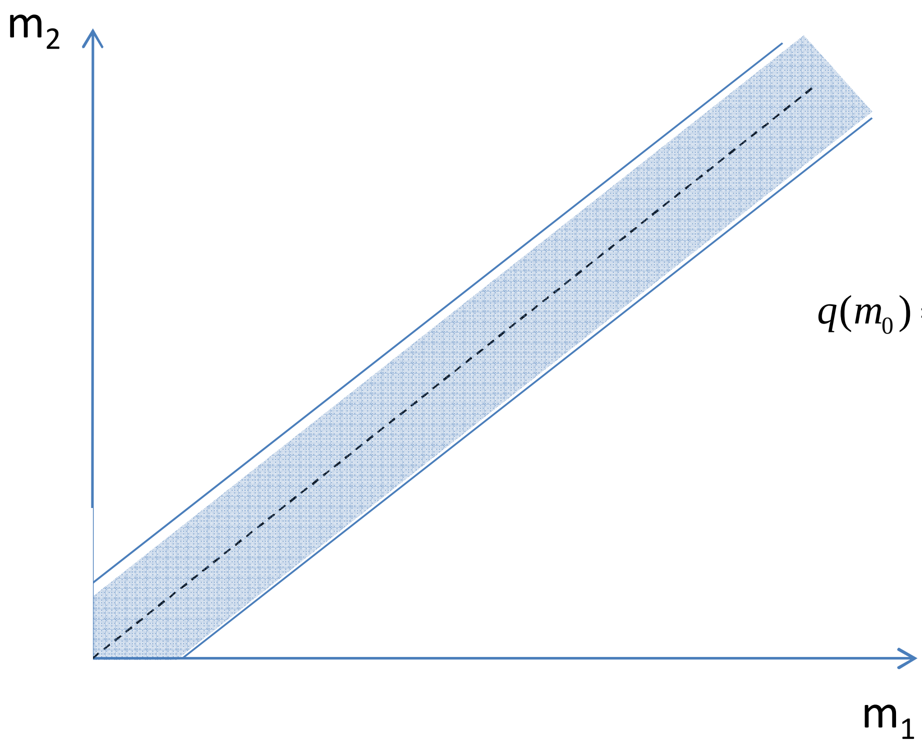
ATLAS combined Higgs workspace toy sampling at 126.5 GeV with ES uncertainty



Leadbetter formula with a parabolic curve (gaussian constraint)

Similar to a LEE in the range defined by σ_{ES}

combination

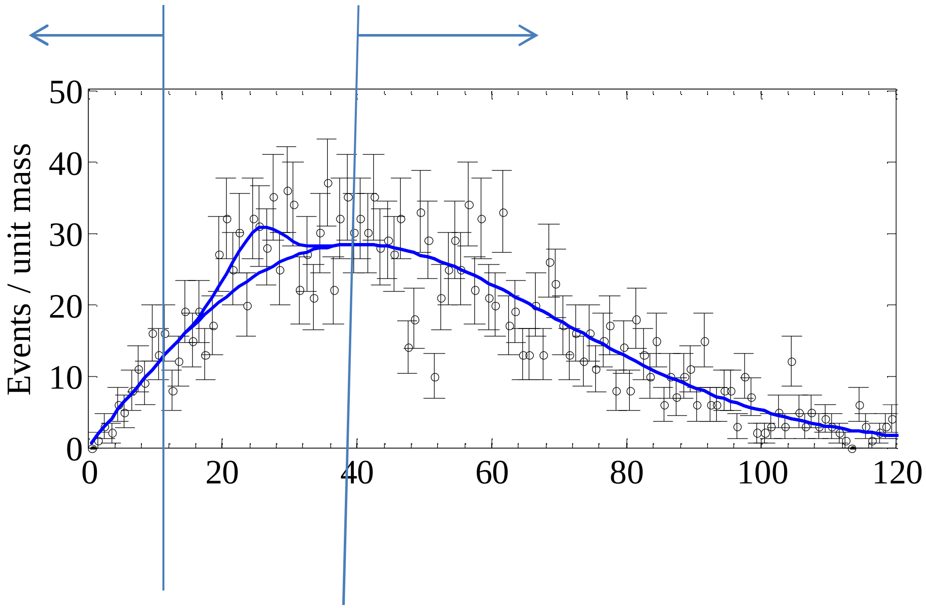


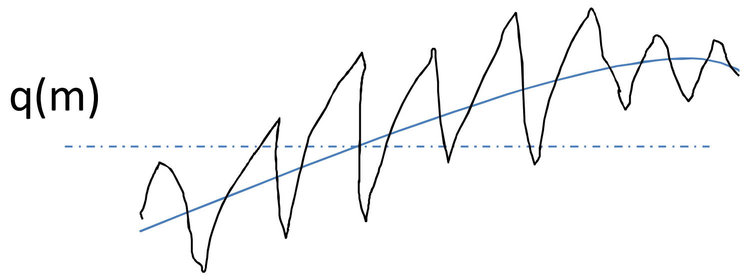
2D χ^2 field

$$q(m_0) = \max_{m_1, m_2} \left\{ q(m_1, m_2) - \left(\frac{m_1 - m_0}{\sigma_1} \right)^2 - \left(\frac{m_2 - m_0}{\sigma_2} \right)^2 \right\}$$

$p \propto Z^{d-1} e^{-Z^2/2}$ (trial factor $\sim Z^d$)

- Sliding windows (mass dependant cuts)==>discontinuity in q(m) due to events getting in/out





- Uncertainty on observed number of upcrossings
  - Usually assumed Poisson $p_{global} = p_{local} + (\mathcal{N}_1 \pm \Delta\mathcal{N}_1)e^{-Z^2/2}$
  - Effect on significance is logarithmic

- When & how asymptotic formulae break down in practice
  - ?

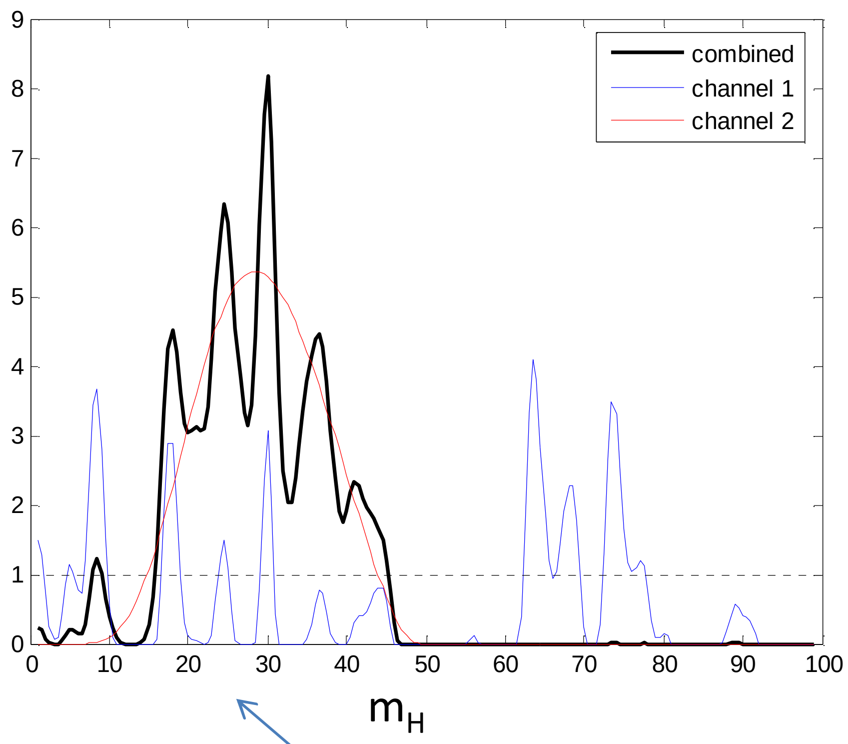
Extra slides

Example of combination of channels with different mas resolutions

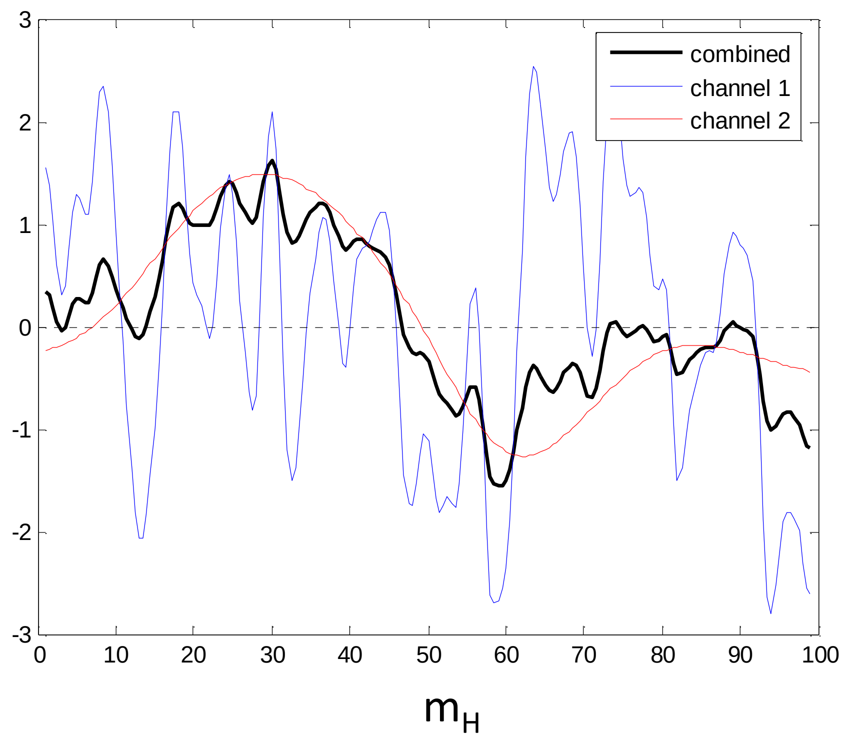
- Toy combination of two channels:
  (both gaussian signal + flat bkg)
  - channel 1: σ_m=1 GeV
  - channel 2: σ_m=10 GeV

Combination example

$q_0(m_H)$



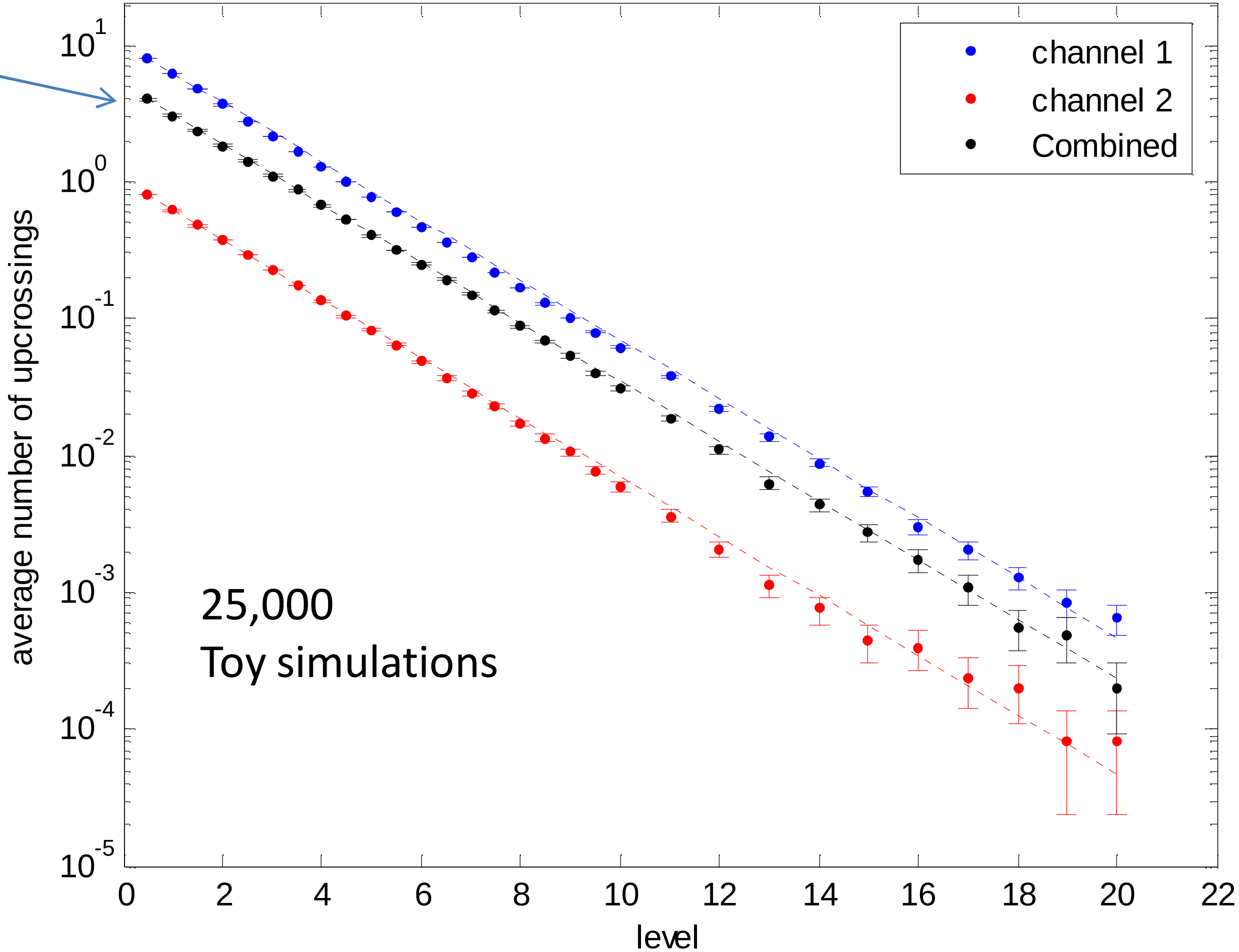
$\hat{\mu}(m_H)$



Note the effect of the wide bump on
the number of upcrossings at 1

Average number of upcrossings

Note that the average number of upcrossins in the combination is always smaller than in channel 1 alone



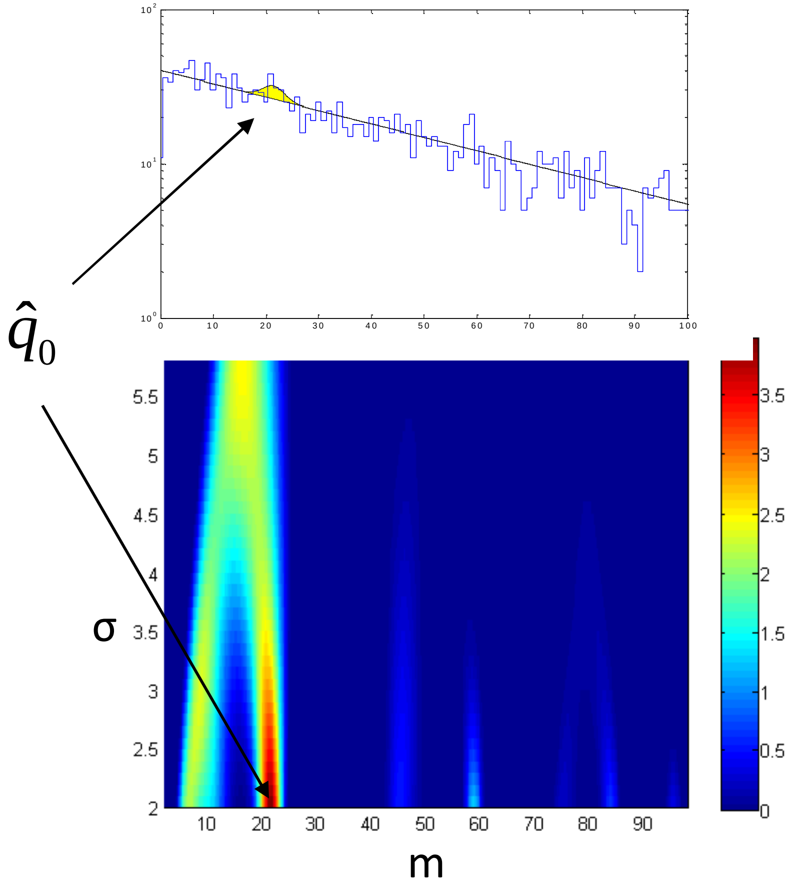
$$\langle N(c) \rangle = N_0 e^{-c/2}$$

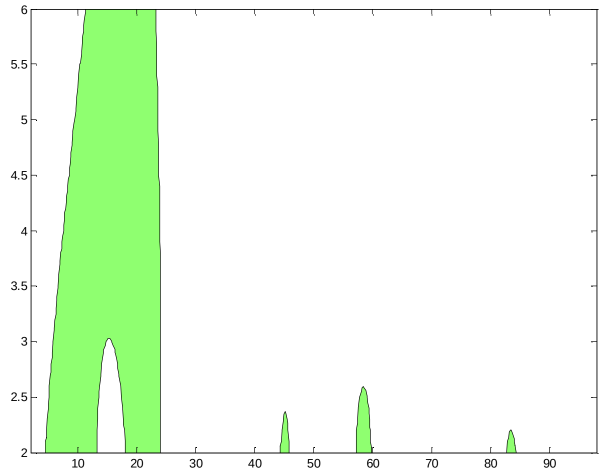
2-D exapmle #2: resonance search with unknown width

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

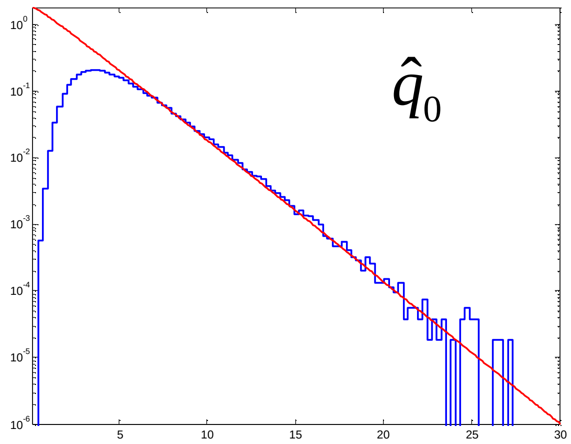
$$\mathcal{L} = \prod_i \frac{N_s f_s(x_i) + N_b f_s(x_i)}{N_s + N_b} \times Poiss(N \mid N_s + N_b)$$

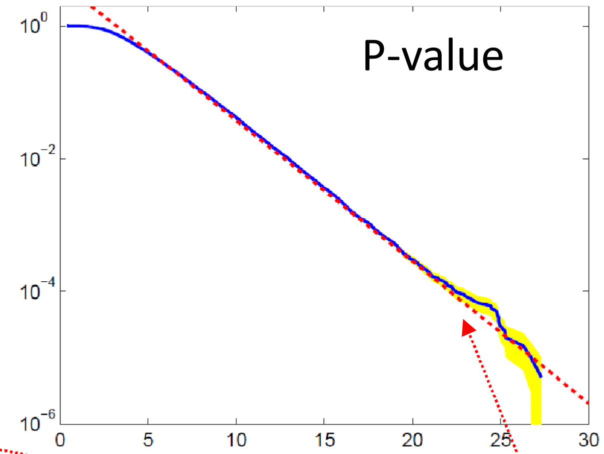
$$f_s(x;m,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \qquad f_b(x) = ce^{-cx}$$





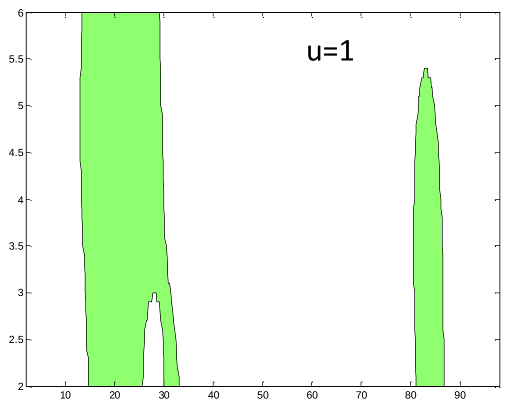
2-D exapmle #2: resonance search with unknown width



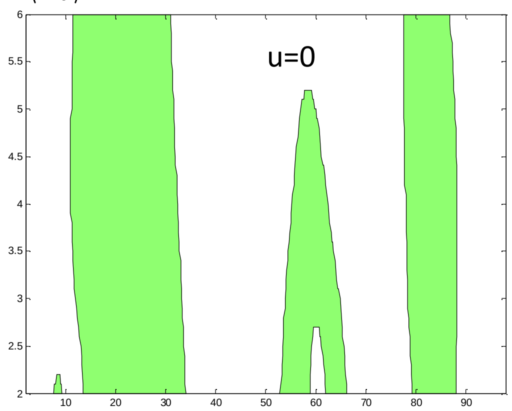


Excellent approximation above the ~2σ level

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



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

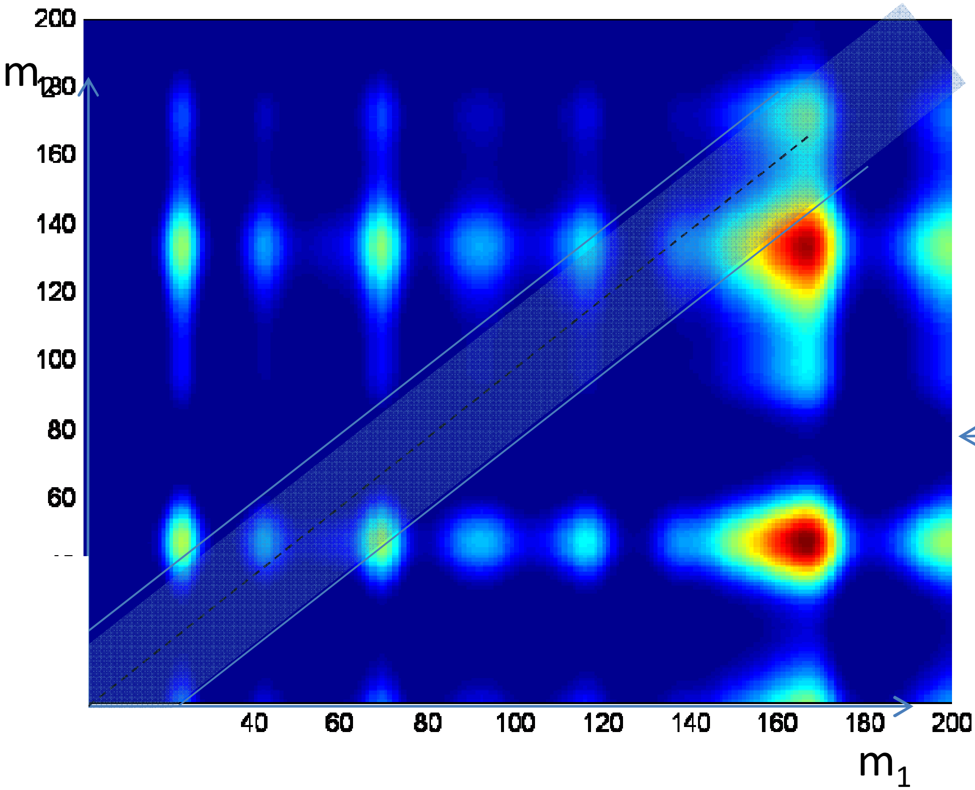


$$E[\varphi(A_u)] = \frac{1}{2}\mathrm{P}(\chi^2 > u) + (\mathcal{N}_1 + \mathcal{N}_2\sqrt{u})e^{-u/2}$$

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

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

(2nd term is dominant for $Z \gg \mathcal{N}_1 / \mathcal{N}_2 = 5.7$)



$$q(m_0) = \max_{m_1, m_2} \left\{ q(m_1, m_2) - \left(\frac{m_1 - m_0}{\sigma_1} \right)^2 - \left(\frac{m_2 - m_0}{\sigma_2} \right)^2 \right\}$$

Asymptotic formulae

Wald approximation for profile likelihood ratio

To find *p*-values, we need: $f(q_0|0), \;\; f(q_\mu|\mu)$

For median significance under alternative, need: $f(q_\mu|\mu')$

Use approximation due to Wald (1943)

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

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

sample size

$$\text{i.e., } E[\hat{\mu}] = \mu'$$

σ from covariance matrix V, use, e.g.,

$$V^{-1} = -E\left[\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right]$$

Noncentral chi-square for $-2\ln\lambda(\mu)$

If we can neglect the $O(1/\sqrt{N})$ term, $-2\ln\lambda(\mu)$ follows a noncentral chi-square distribution for one degree of freedom with noncentrality parameter

$$\Lambda = \frac{(\mu - \mu')^2}{\sigma^2}$$

As a special case, if $\mu' = \mu$ then $\Lambda = 0$ and $-2\ln\lambda(\mu)$ follows a chi-square distribution for one degree of freedom (Wilks).

To have the distribution well defined for $N \to \infty$, take $\mu' = \mu + \delta / \sqrt{N}$, since $\sigma \sim 1/\sqrt{N}$

e.g. If $b \to \infty$ take $s \propto \sqrt{b} \to \infty$, such that $s/\sqrt{b} \propto$ constant ($s/b \propto 1/\sqrt{b} \to 0$)

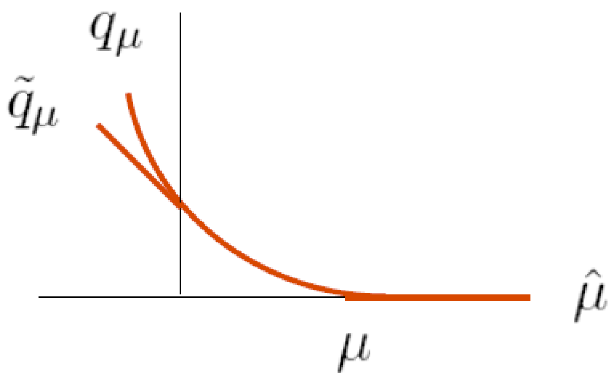
In this limit σ is independent of μ up to $O(1/\sqrt{N})$

e.g. $\dfrac{s}{\sqrt{b}} = \dfrac{s}{\sqrt{s+b}} + O(s/b)$

Relation between test statistics and $\hat{\mu}$

Assuming the Wald approximation for $-2\ln\lambda(\mu)$, q_μ and $\tilde{q}_\mu$ both have monotonic relation with μ.

$$q_\mu = \begin{cases} \frac{(\mu-\hat{\mu})^2}{\sigma^2} & \hat{\mu} < \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$



$$\tilde{q}_\mu = \begin{cases} \frac{\mu^2}{\sigma^2} - \frac{2\mu\hat{\mu}}{\sigma^2} & \hat{\mu} < 0 \\ \frac{(\mu-\hat{\mu})^2}{\sigma^2} & 0 \le \hat{\mu} \le \mu \\ 0 & \hat{\mu} > \mu \ , \end{cases}$$

And therefore quantiles of q_μ, $\tilde{q}_\mu$ can be obtained directly from those of $\hat{\mu}$ (which is Gaussian).

The Asimov data set

To estimate median value of $-2\ln\lambda(\mu)$, consider special data set where all statistical fluctuations suppressed and n_i, m_i are replaced by their expectation values (the "Asimov" data set):

$$n_i = \mu' s_i + b_i \qquad m_i = u_i$$

$$\rightarrow \quad \hat{\mu} = \mu' \qquad \hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$$

$$\lambda_{\rm A}(\mu) = \frac{L_{\rm A}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L_{\rm A}(\hat{\mu}, \hat{\boldsymbol{\theta}})} = \frac{L_{\rm A}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L_A(\mu', \boldsymbol{\theta})}$$

$$-2\ln\lambda_{\rm A}(\mu) = \frac{(\mu - \mu')^2}{\sigma^2} = \Lambda$$

Asimov value of $-2\ln\lambda(\mu)$ gives non-centrality param. Λ, or equivalently, σ.

Assuming Wald approximation, the relation between q_0 and $\hat{\mu}$ is Monotonic, therefore quantiles of $\hat{\mu}$ map one-to-one onto those of q_0, e.g.,

$$\mathrm{med}[q_0] = q_0(\mathrm{med}[\hat{\mu}]) = q_0(\mu') = \frac{\mu'^2}{\sigma^2} = -2\ln\lambda_{\rm A}(0)$$