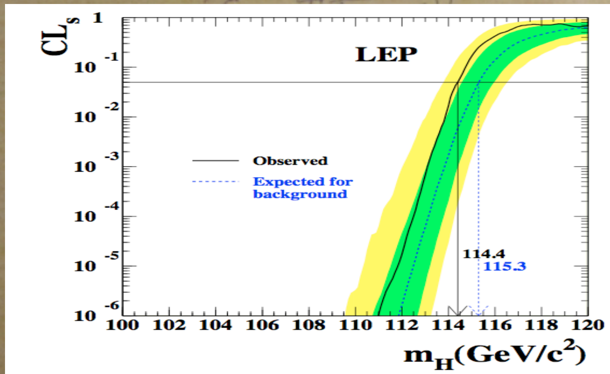
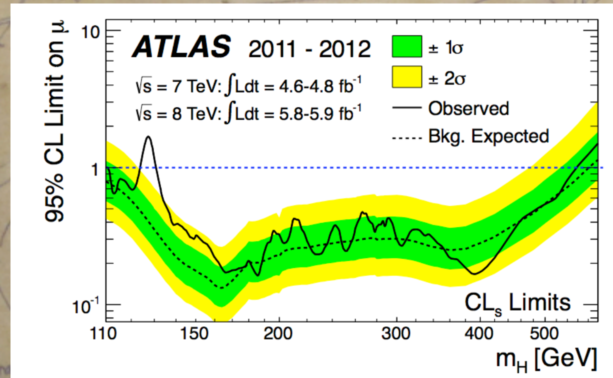
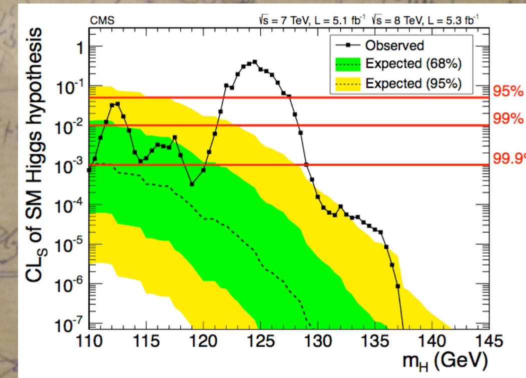






Limit calculations in HEP

20.08.2012

Terascale Alliance School and Workshop

A. Read (U. Oslo)

UiO : University of Oslo

The Research Council of Norway

Topics

- Challenge of limit-setting
- A brief history
- Likelihood ratio -> profile likelihood ratio
- Connection to discovery, measurement
- Systematics/Nuisance parameters
- Look elsewhere
- Weighted events
- Energy scale systematics

A counting problem

- Expected background b
- Want to test signal hypothesis s
- Observe n events $\mathcal{L}(n|s+b) = \frac{\exp^{-(s+b)}(s+b)^n}{n!}$
  - If n=0 we **know** we saw no signal - and can calculate how likely that is given s: Upper limit of 3 events comes from Poisson $1-e^{-s}=95\%$
  - If $n \geq 1$ what do we know?
    - $n_s \leq n$ and $n_b \leq n$ and $n_{s+b} \leq n$
  - If n << b, what do we conclude about s-hypothesis?

This result quantifies the difficulty understood intuitively by past physicists, and connects it to a body of statistics literature going back 50+ years!

Our simply stated problem is in one of the thorniest corners of the statistics literature: what to do when one *knows* post-data that the pre-data coverage probability is inapplicable to the "recognizable subset" containing the observed *x*.

The BIG LESSON: if all your discussions/arguments consider only N-P coverage and power, you can be missing important considerations about post-data inference.

Bob Cousins

Bob Cousins. Bayes...and the LHC, 12 Sep 2011

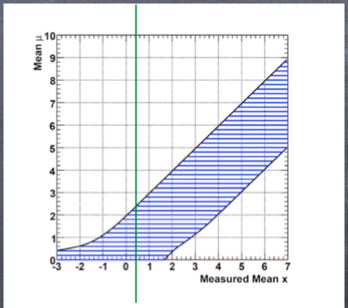
Brief (!) history of limits

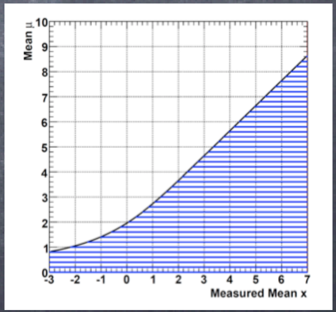
- O. Helene (1983) - Bayesian limit with flat prior on signal
- G. Zech (1988) - frequentist interpretation of Helene
- A. Read (1997) - rederived Zech from likelihood ratio and "background conditioning"; $CL_S \approx$ "confidence in the signal-only hypothesis"
- Feldman and Cousins (1998) - frequentist confidence intervals - "coverage is king" (but tests signal +background hypothesis)
- Birnbaum (1961!!) - support for CL_S in the professional statistics literature - rediscovered by O. Vitells

$$CL = \frac{\int_s^\infty \mathcal{L}(s',b)ds'}{\int_0^\infty \mathcal{L}(s',b)ds'}.$$

$$CL = 1 - \frac{\sum_{n=0}^{n_{obs}} \frac{e^{-(b+s)}(b+s)^n}{n!}}{\sum_{n=0}^{n_{obs}} \frac{e^{-b}b^n}{n!}}.$$

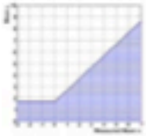
$$CL_s \equiv CL_{s+b}/CL_b.$$



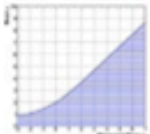


"A concept of statistical evidence is not plausible unless it finds 'strong evidence for H2 as against H1' with small probability (alpha) When H1 is true, and with much larger probability (1 -beta) when H2 is true."

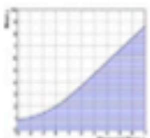
Five methods used for bounded Gaussian mean problem



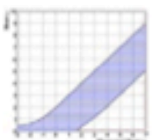
1) 1960's and beyond:
   $UL = \max(x, 0) + 1.64\sigma$



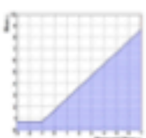
2) 1979 "PDG" (real 1986 PDG) and beyond:
   Bayesian with uniform prior



3) 1997: Alex Read et al. (LEP)
   CL_S



4) 1997: Feldman and Cousins (NOMAD)
   Unified Approach



5) 2010: Power Constrained Limits;
   Cowan, Cranmer, Gross, Vitells (ATLAS):
   $UL = \max(0, \max(x, x_{PCL}) + 1.64\sigma)$

Bob Cousins. Bayes...and the LHC, 12 Sep 2011 40

Why likelihood ratio?

- If there is no signal, we want to exclude signal hypothesis as strongly as possible
- If there is a signal, we want to exclude background hypothesis as strongly as possible
  - The alternative is signal hypothesis
  - We can't ignore that there may be alternative signals -> Continuous signal test is measurement

Selecting events

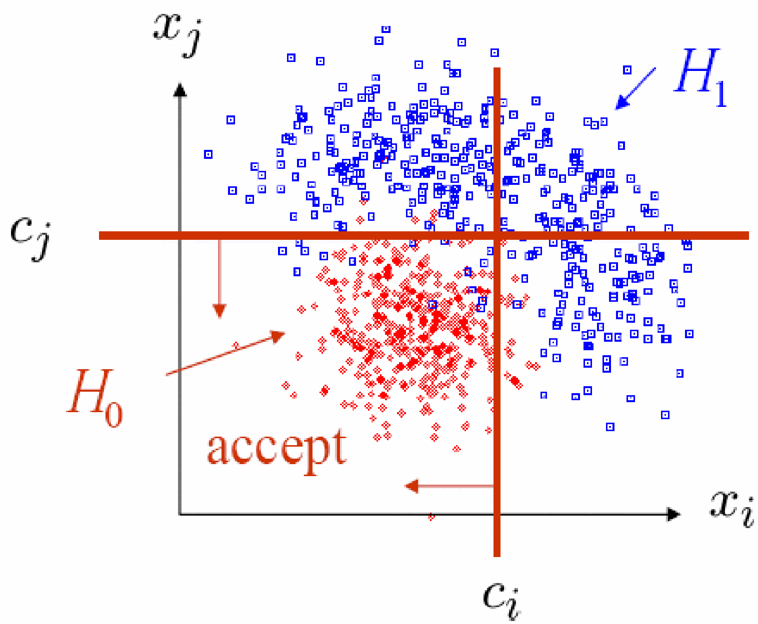
Suppose we have a data sample with two kinds of events, corresponding to hypotheses H_0 and H_1 and we want to select those of type H_0.

Each event is a point in $\vec{x}$ space. What decision boundary should we use to accept/reject events as belonging to event type H_0?

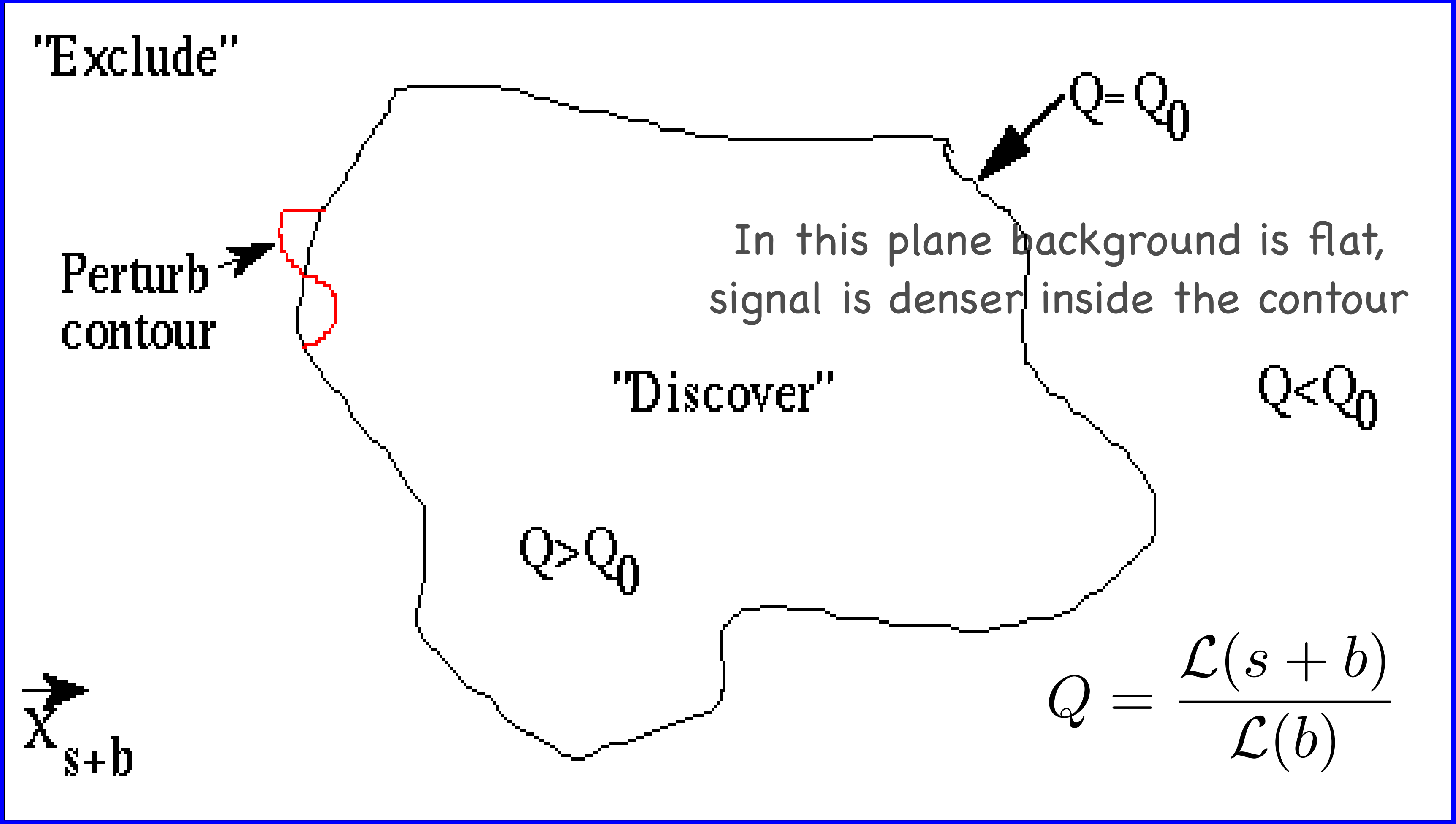
Probably start with cuts:

$$x_i < c_i$$

$$x_j < c_j$$



Neyman-Pearson lemma



$$Q = \frac{\mathcal{L}(s+b)}{\mathcal{L}(b)}$$

LR and Multivariate Analysis (MVA)

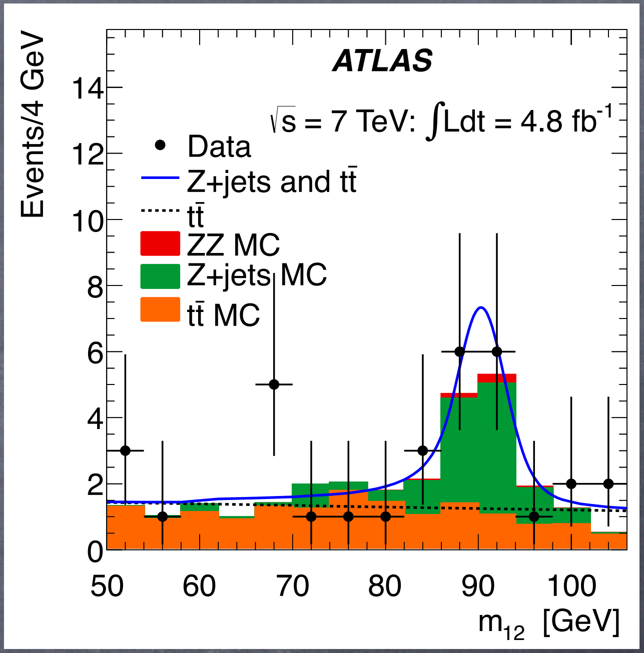
- "Multi-variate analysis attempts to find the LR contours in a complicated hyperspace" - Paraphrase of Harrison Prosper at PHYSTAT in Durham, 2002

Important aside: Nuisance parameters

- background, uncertainty, uncertainties important words in ATLAS "Higgs discovery paper"

Nuisance parameters

- Parameters fitted directly to the data but no real interest
  - E.g. parametric background
- Parameters from external measurements that incorporate uncertainty
  - E.g. normalization: luminosity, signal theory



Nuisance parameters

- Broaden the likelihood profile

$$\chi^2 = \frac{(n-(\mu+\delta))^2}{\sigma^2} + \frac{\delta^2}{\sigma_s^2}$$

$$\frac{\partial \chi^2}{\partial \delta} = 0$$

$$\frac{\partial \chi^2}{\partial \mu} = 0$$

$$\hat{\mu} = n, \delta = 0$$

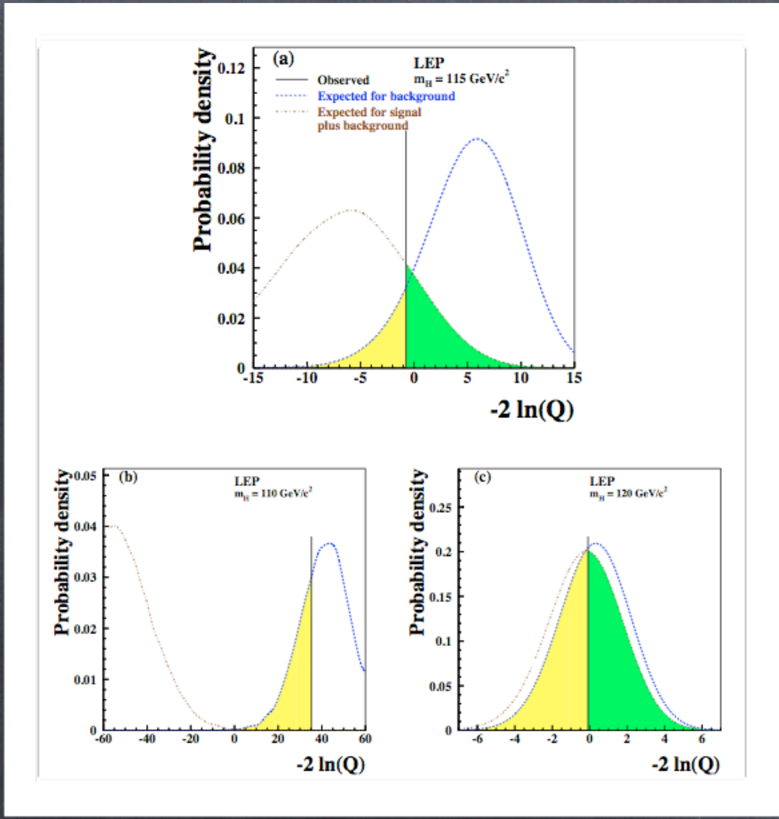
$$\frac{1}{\sigma_\mu^2} = \frac{1}{2}\frac{\partial^2 \chi^2}{\partial \mu^2}$$

$$\sigma_\mu = \sqrt{\sigma^2 + \sigma_s^2}$$

LLR test statistics

	Test statistic	Test statistic	Nuisance parameters	Pseudo-experiments
LEP	$-2\ln\frac{L(\mu,\tilde{\theta})}{L(0,\tilde{\theta})}$	Simple LR	Fixed by MC	Nuisance parameters randomized about MC
Tevatron	$-2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(0,\hat{\theta})}$	Ratio of profiled likelihoods	Extracted from priors	Nuisance parameters randomized from priors
LHC	$-2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}$	Profile likelihood ratio	Profiled (fit to data)	New nuisance parameters fitted for each pseudo-exp.

LHC sampling of test statistic is frequentist, LEP and Tevatron Bayes-frequentist hybrid. CL_s can be used together with any of these – must be specified! No longer sufficient to write e.g. "the CL_s method was used".



$$Q = \frac{\prod_{i=1}^{N_{chan}} \frac{e^{-(s_i+b_i)}(s_i+b_i)^{n_i}}{n_i!} \prod_{j=1}^{n_i} \frac{s_i S_i(x_{ij})+b_i B_i(x_{ij})}{s_i+b_i}}{\prod_{i=1}^{N_{nchan}} \frac{e^{-b_i} b_i^{n_i}}{n_i!} \prod_{j=1}^{n_i} B_i(x_{ij})}$$

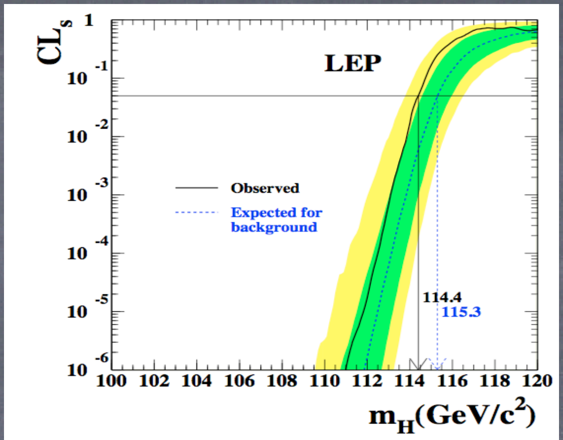
- Why do we now use λ(μ) and not Q_{LEP}(μ)?
    - LEP->Tevatron: Nuisance parameters, data-driven, fits
    - Tevatron->LHC: Asymptotics, not mix Bayes and frequentist

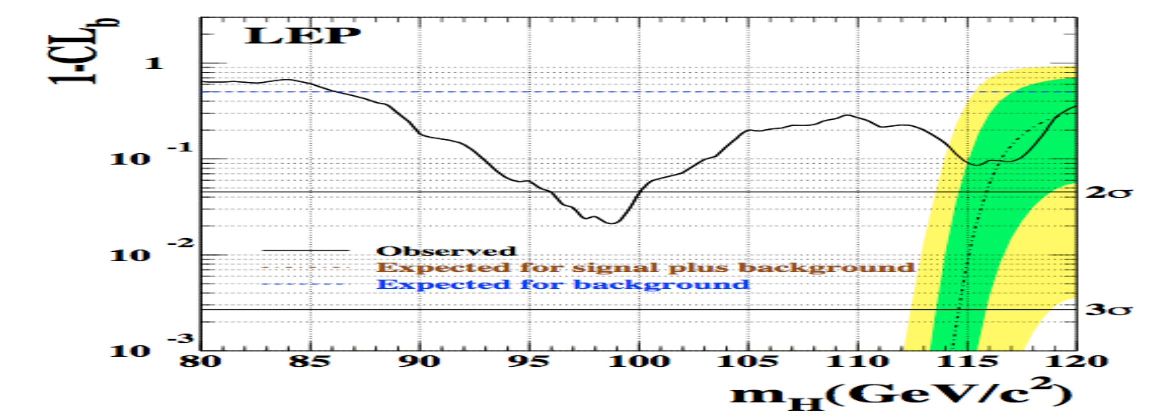
$$L(\mu, b) = \frac{(\mu s + b)^n}{n!} e^{-(\mu s + b)} \frac{(\tau b)^m}{m!} e^{-\tau b}$$

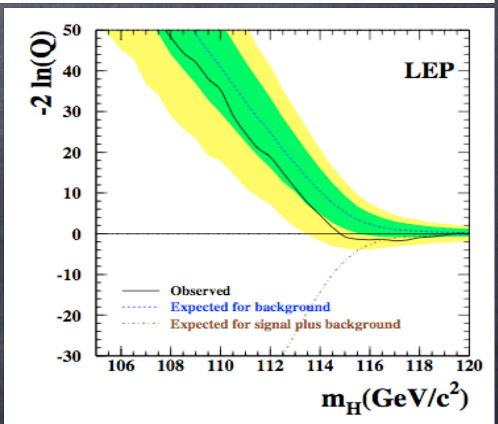
$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})}$$

More on QLEP/Tev

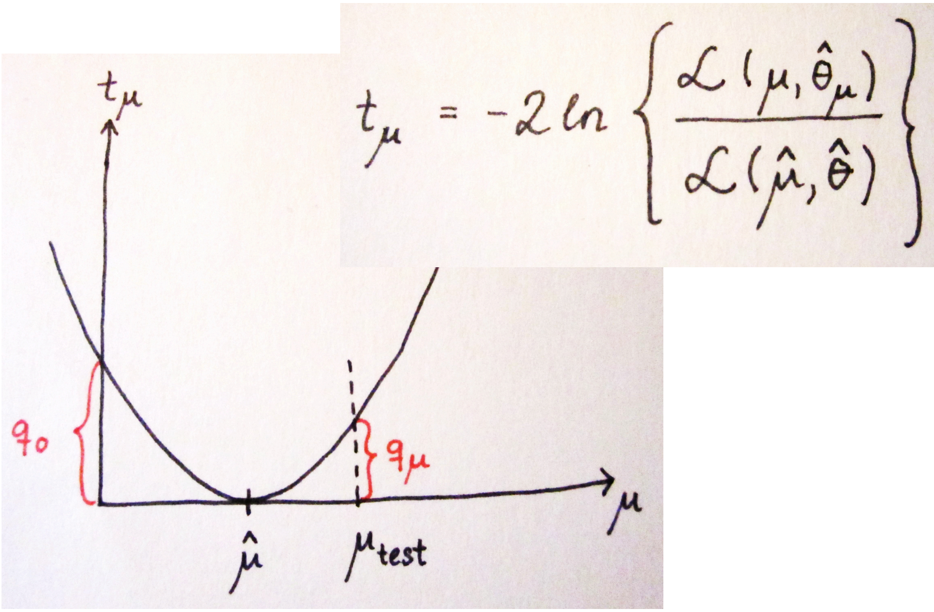
- Advantage:
  - Exclusion, discovery, -2lnQ monotonic transformations of each other
- Disadvantages:
  - Signal uncertainties affect discovery
  - Asymptotics much more complicated





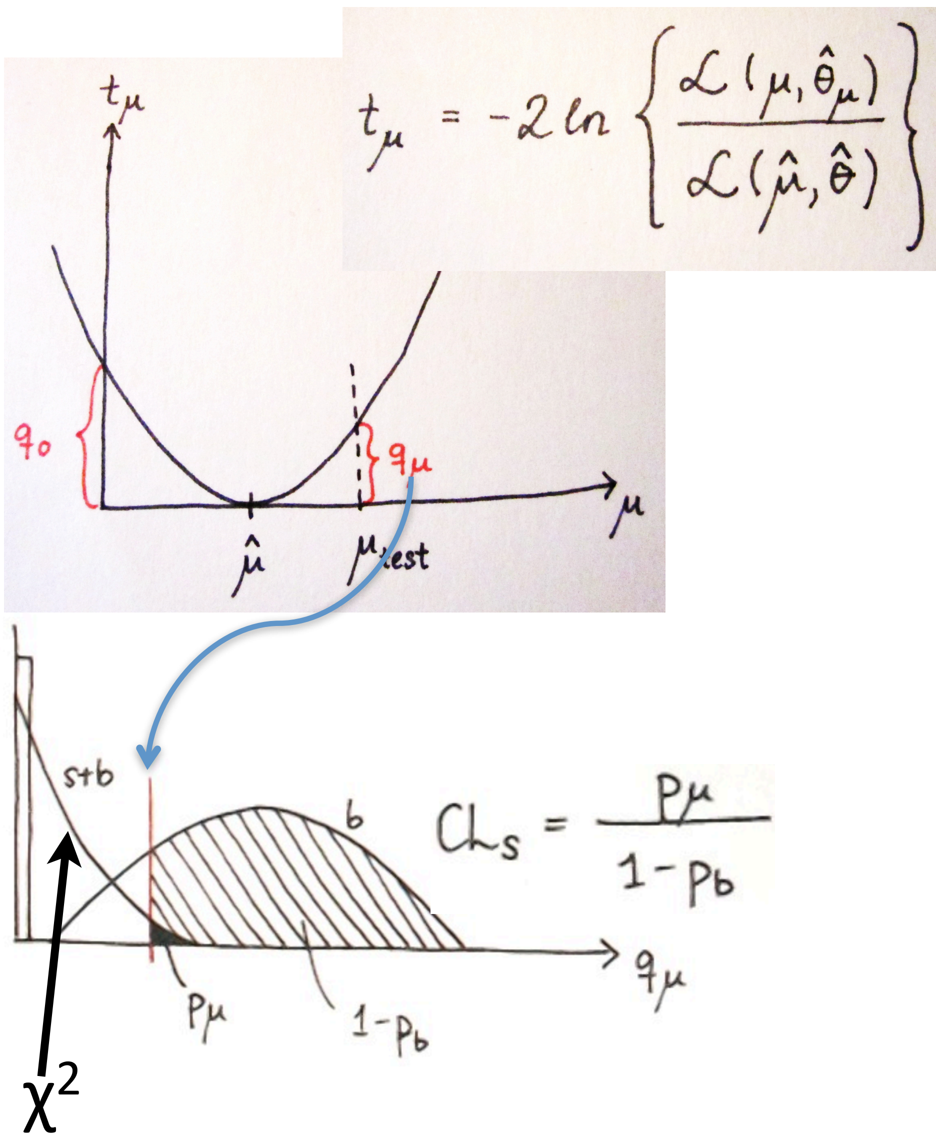


Profile likelihood ratio: CL_s and μ_{95}^{up}

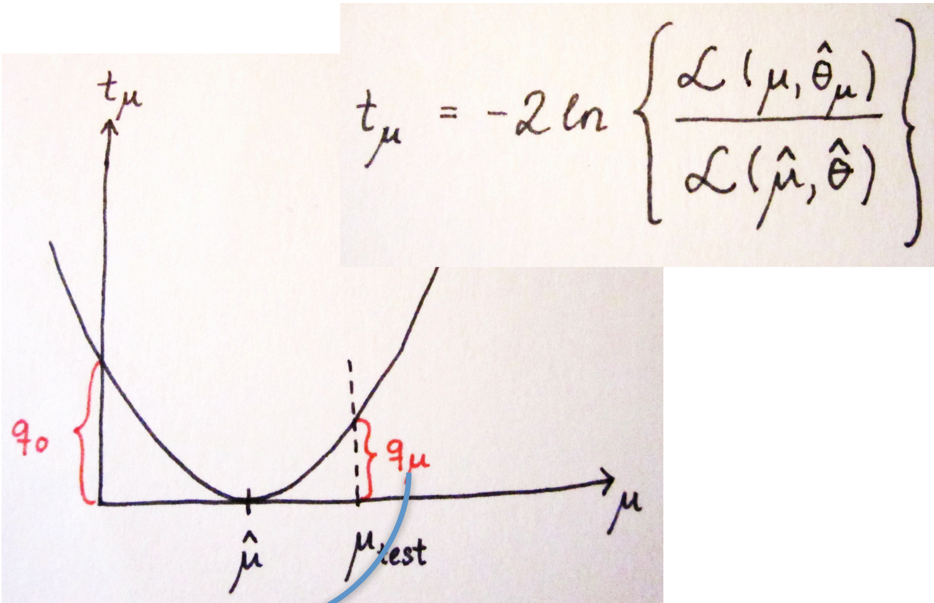


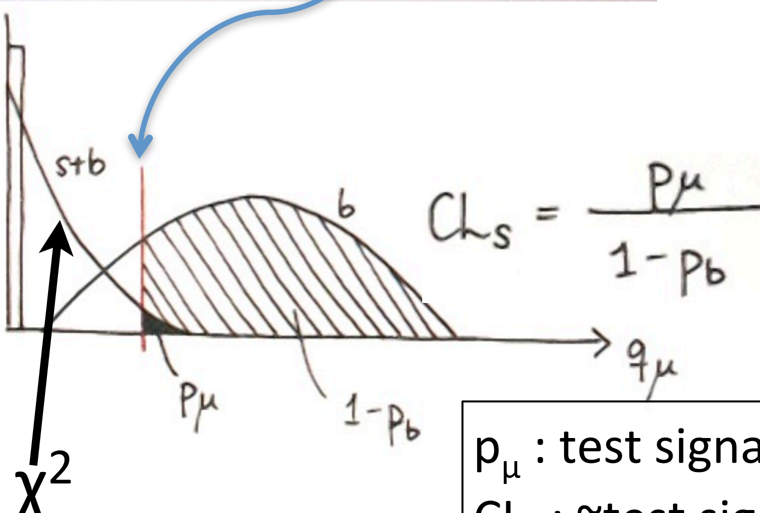
χ^2

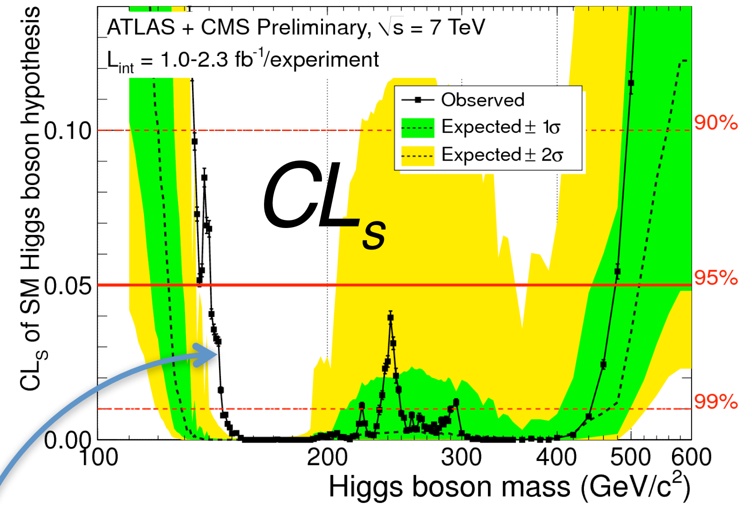
Profile likelihood ratio: CL_s and μ_{95}^{up}



Profile likelihood ratio: CL_s and μ_{95}^{up}



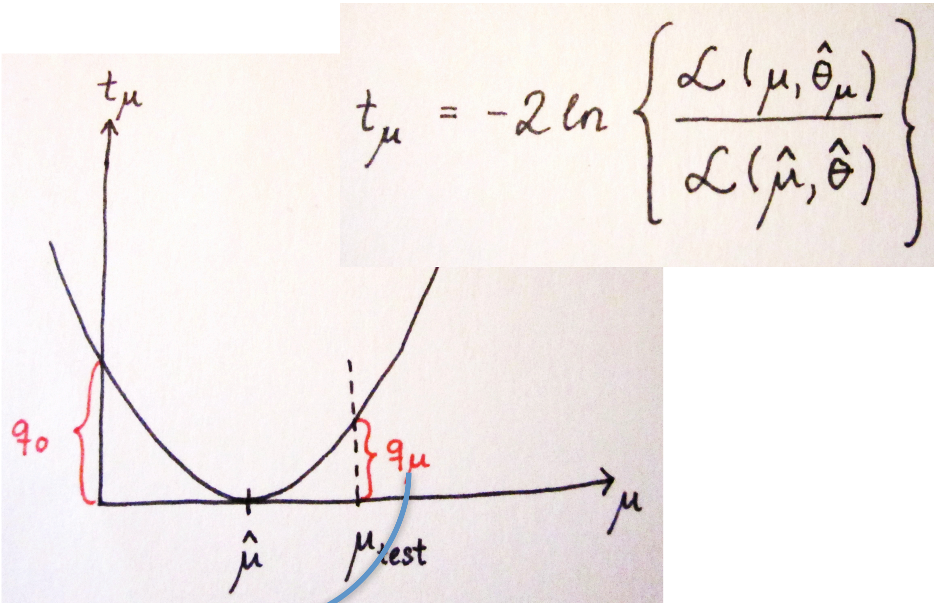


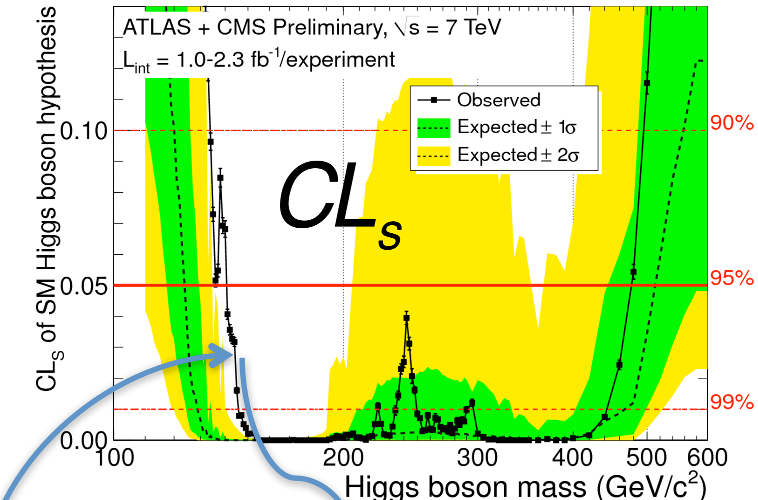


p_μ : test signal+background

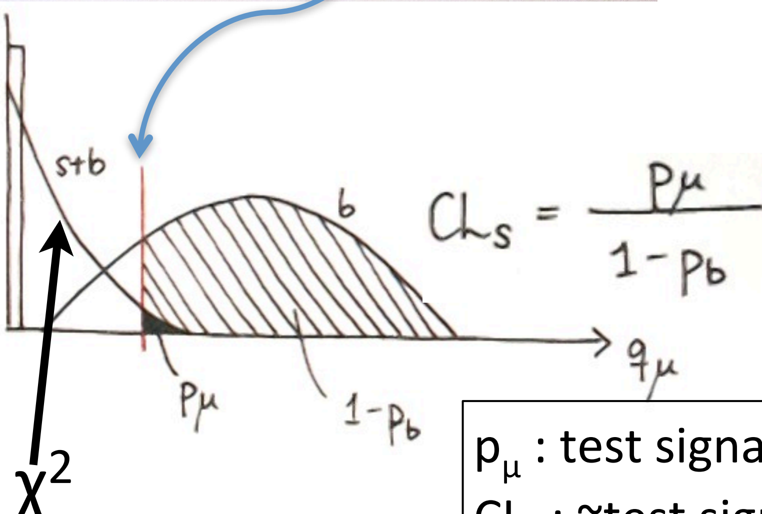
CL_s : ~test signal

Profile likelihood ratio: CL_s and μ_{95}^{up}





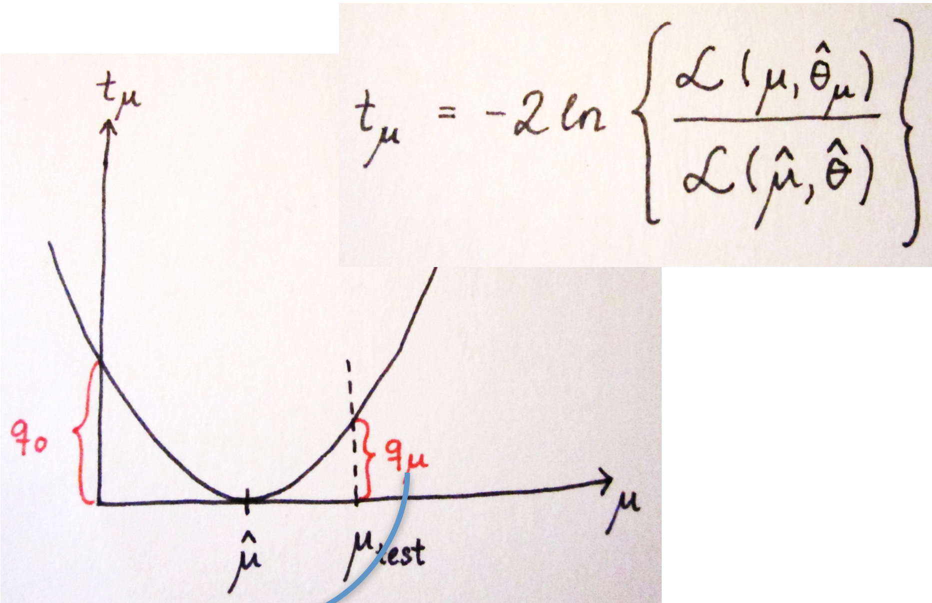
$$\mu_{95}^{up} = \mu(CL_s = 0.05)$$

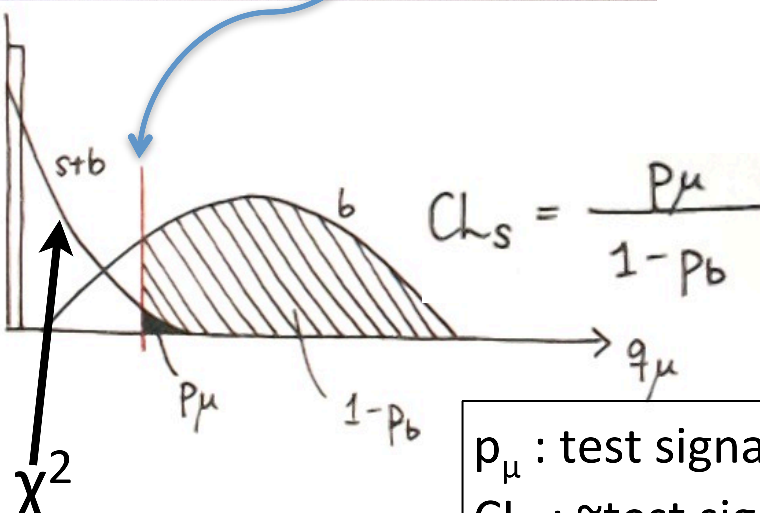


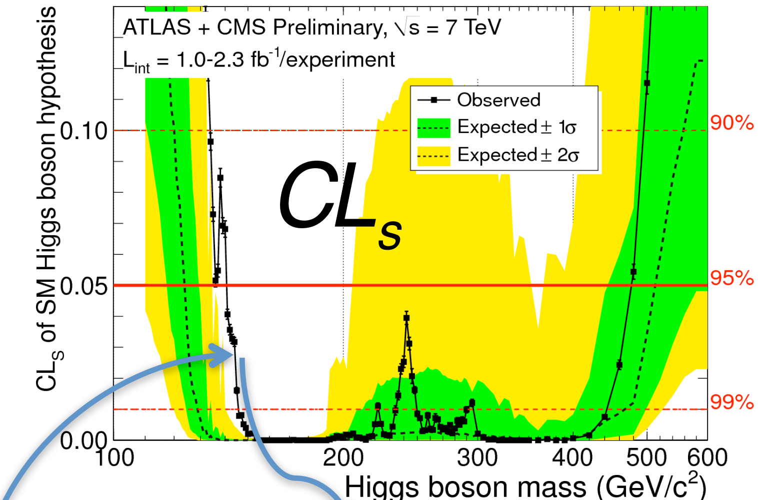
p_μ : test signal+background

CL_s : ~test signal

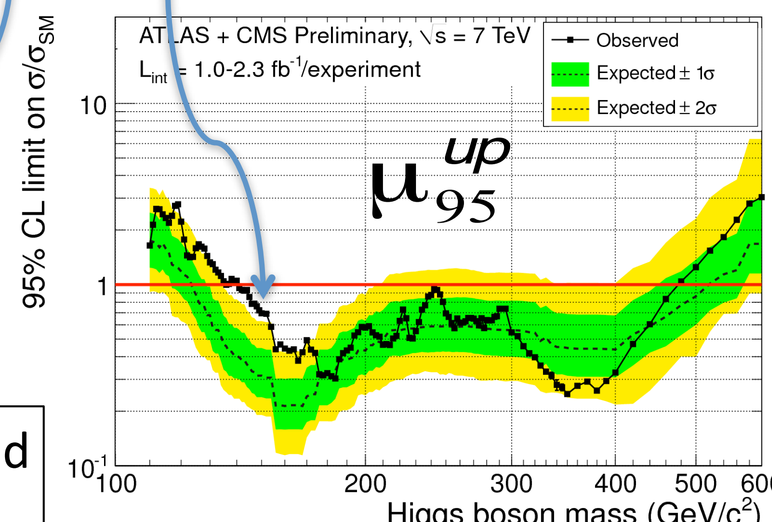
Profile likelihood ratio: CL_s and μ_{95}^{up}







$$\mu_{95}^{up} = \mu(CL_s = 0.05)$$



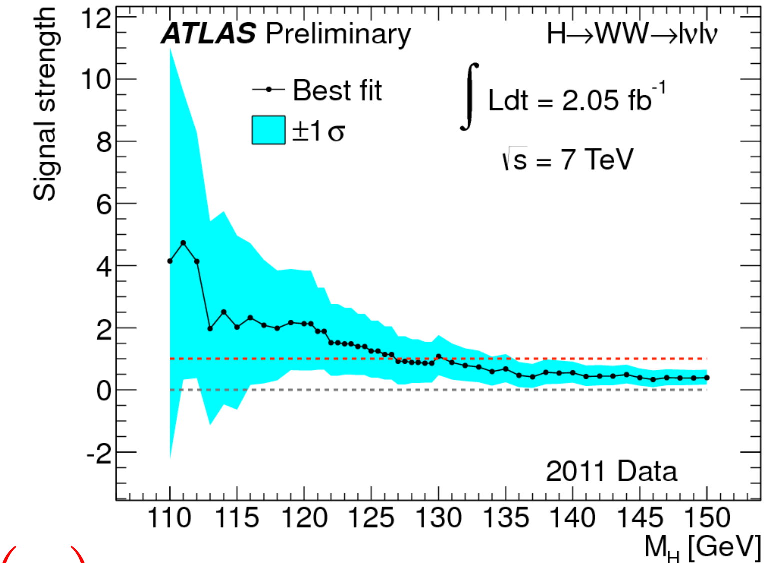
p_μ : test signal+background

CL_s : ~test signal

Profile likelihood ratio: p_0 and $\hat{\mu}$

LHCHCG Combination Procedures

$$= -2 \ln \left\{ \frac{\mathcal{L}(\mu, \hat{\hat{\theta}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \right\}$$

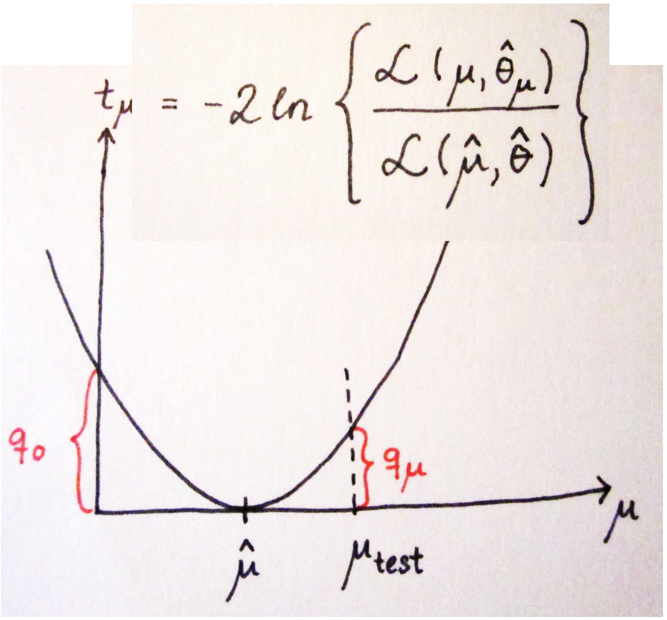


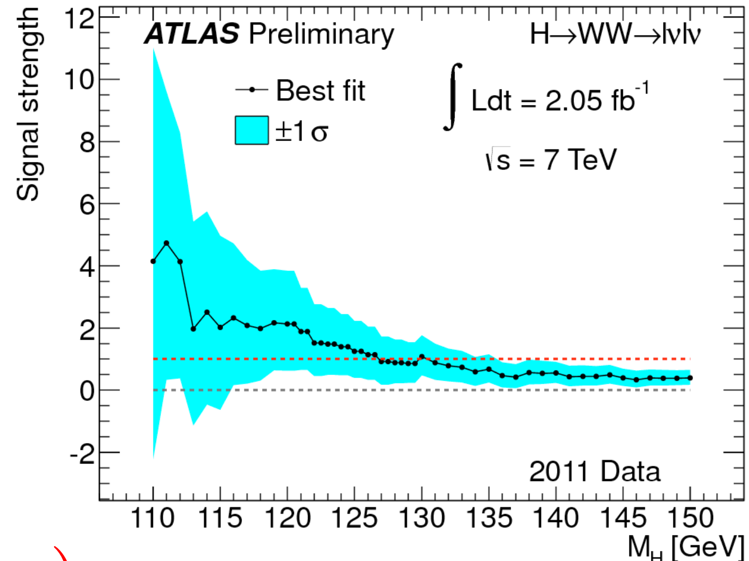
P.S. $q_{LEP}(\mu) = q_\mu - q_0$

χ^2

Profile likelihood ratio: p_0 and $\hat{\mu}$

[LHCHCG Combination Procedures]



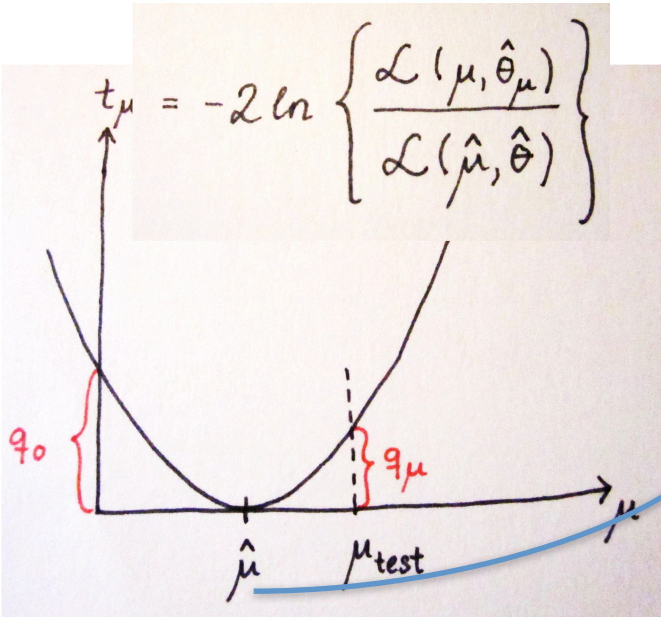


P.S. $q_{LEP}(\mu) = q_\mu - q_0$

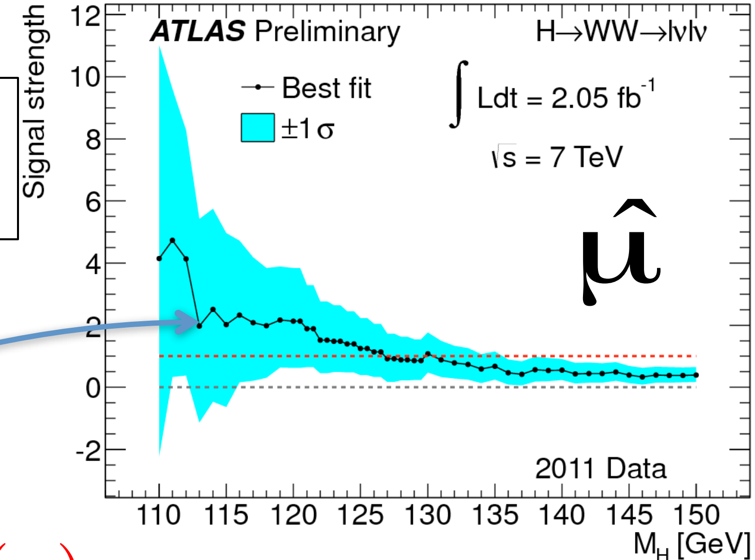
χ^2

Profile likelihood ratio: p_0 and $\hat{\mu}$

LHCHCG Combination Procedures



$\hat{\mu}$ to estimate signal strength



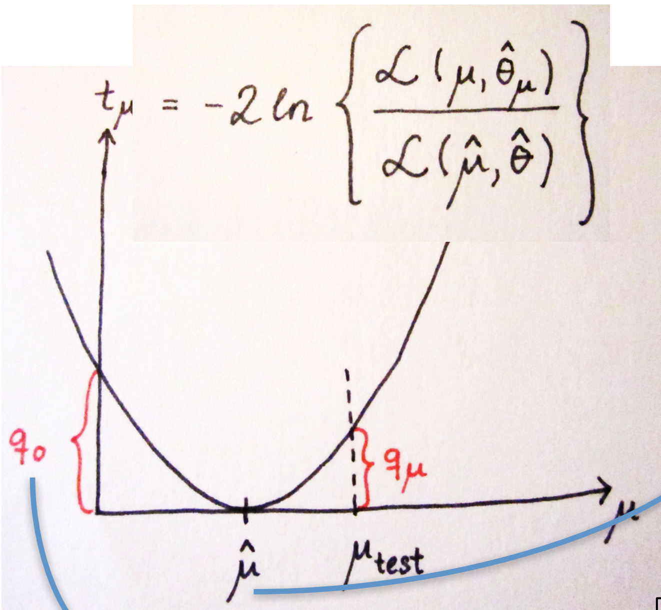
$\hat{\mu}$

P.S. $q_{LEP}(\mu) = q_\mu - q_0$

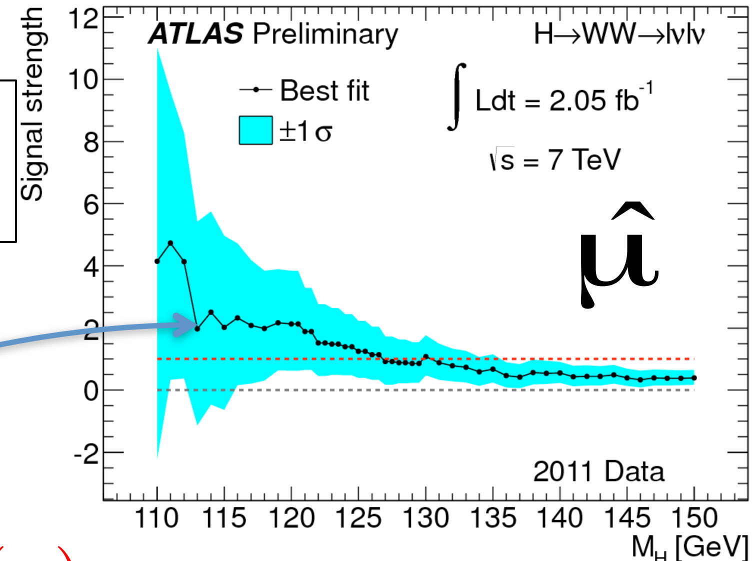
χ^2

Profile likelihood ratio: p_0 and $\hat{\mu}$

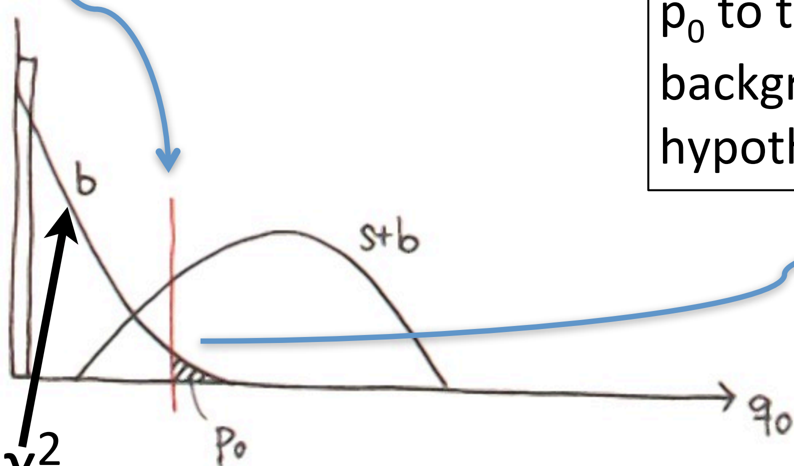
LHCHCG Combination Procedures



$\hat{\mu}$ to estimate signal strength

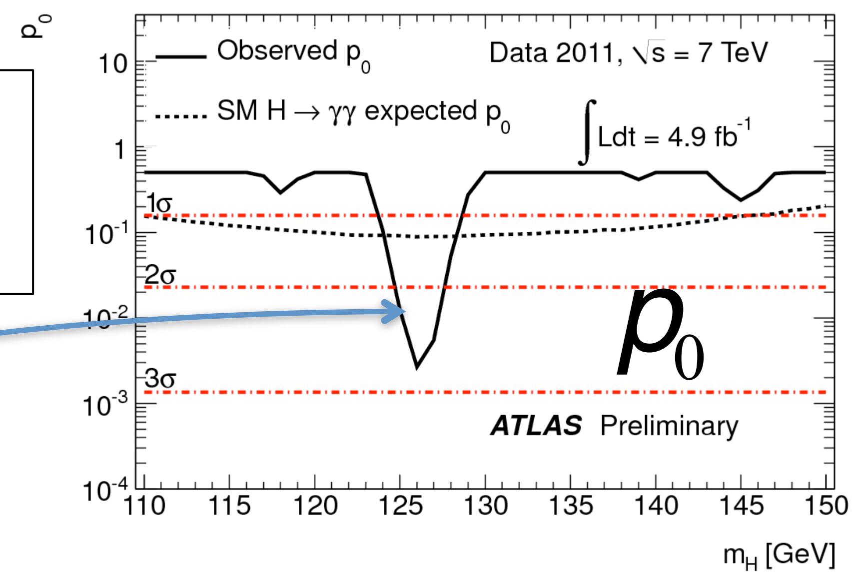


P.S. $q_{LEP}(\mu) = q_\mu - q_0$



χ^2

p_0 to test background hypothesis



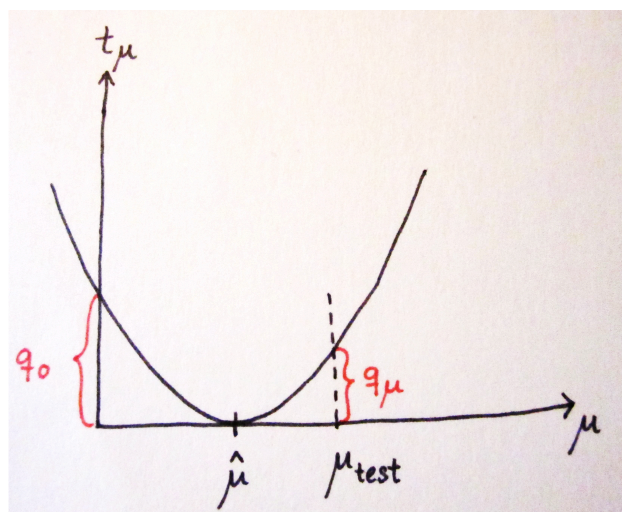
Combined Results

$$L(m_H, \mu, \vec{\vartheta}) = \prod_i L_i(m_H, \mu, \vec{\vartheta}_i)$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i/\sigma_i^2}{1/\sigma^2}$$

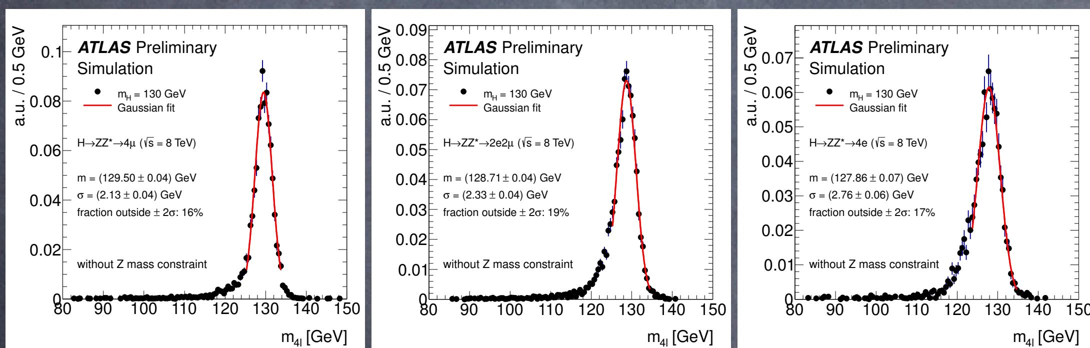
$$\frac{1}{\sigma^2} = \frac{1}{\sum_{i=1}^{n} 1/\sigma_i^2}$$

$$t_\mu = -2\ln\left\{\frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})}\right\}$$

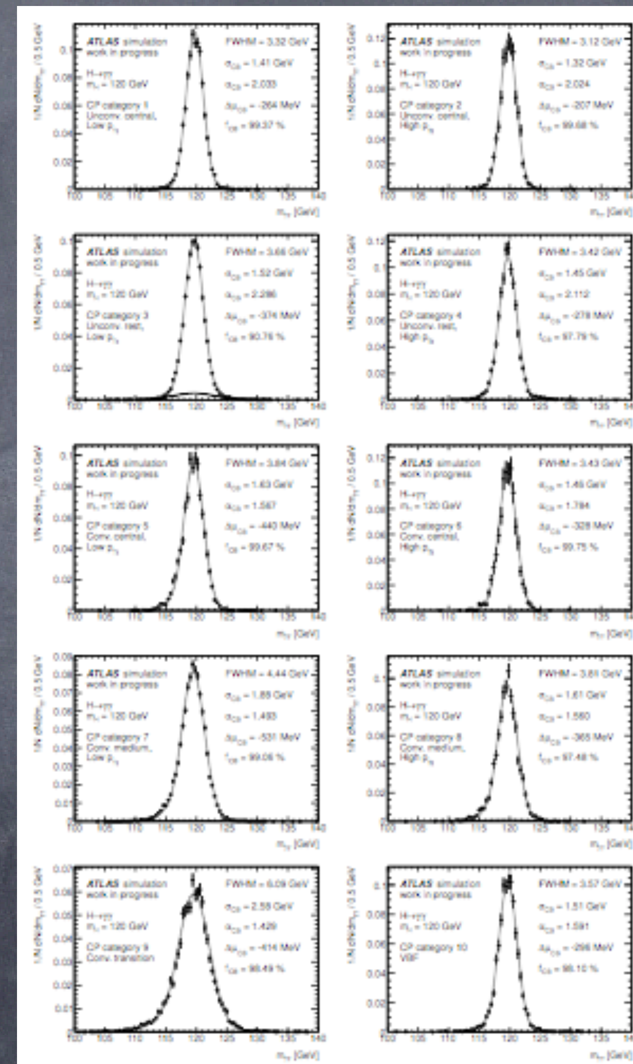


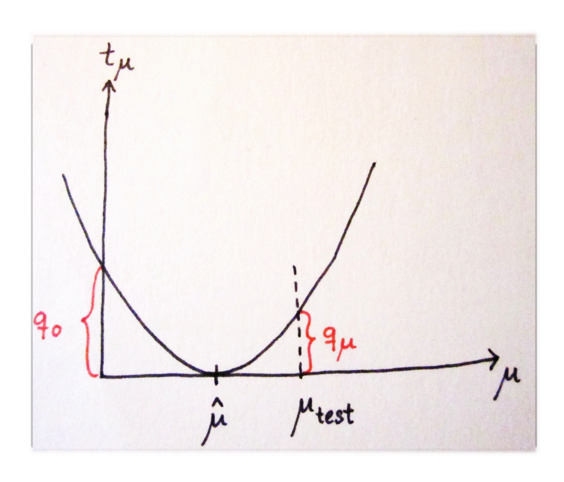
Distributions (pdfs) in LR

- Mass distributions, one for each channel (resolution not the same.
- Ultimate sensitivity with per-event error function

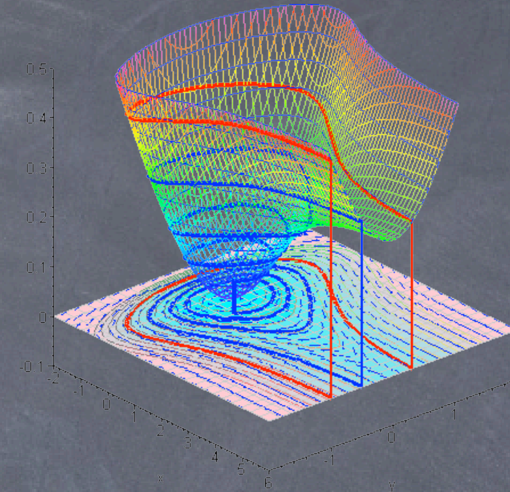


$$Q = \frac{\prod_{i=1}^{N_{chan}} \frac{e^{-(s_i+b_i)}(s_i+b_i)^{n_i}}{n_i!}}{\prod_{i=1}^{N_{nchan}} \frac{e^{-b_i} b_i^{n_i}}{n_i!}} \frac{\prod_{j=1}^{n_i} \frac{s_i S_i(x_{ij}) + b_i B_i(x_{ij})}{s_i+b_i}}{\prod_{j=1}^{n_i} B_i(x_{ij})}$$





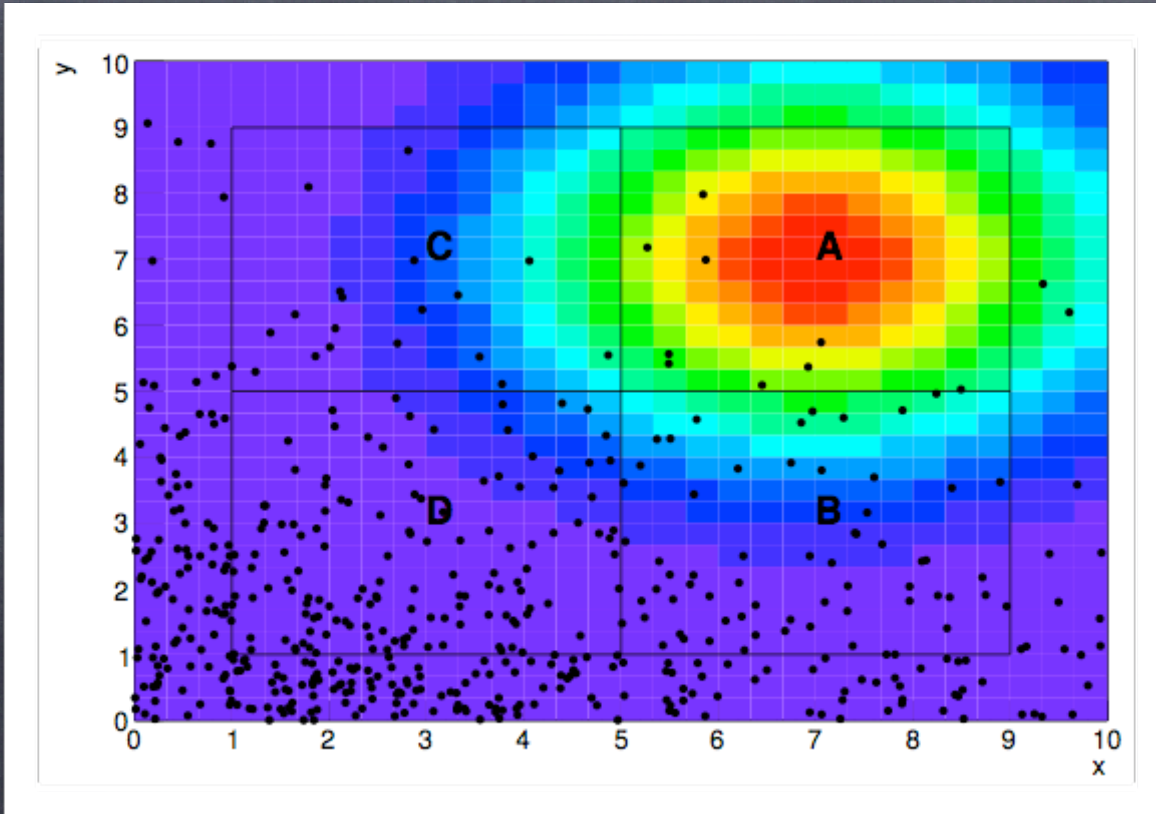
Other Hypothesis Tests



Background (scan m_H)	$\lambda(\mu=0, m_H) = \frac{L(\mu=0, m_H, \hat{\theta})}{L(\hat{\mu}, m_H, \hat{\hat{\theta}})}$
Signal (scan m_H)	$\lambda(\mu, m_H) = \frac{L(\mu, m_H, \hat{\theta})}{L(\hat{\mu}, m_H, \hat{\hat{\theta}})}$
Mass consistency	$\lambda(m_H) = \frac{L(m_H, \hat{\hat{\mu}}_1, \hat{\hat{\mu}}_2, \hat{\hat{\theta}})}{L(\hat{m_1}_H, \hat{m_2}_H, \hat{\mu}_1, \hat{\mu}_2, \hat{\theta})}$
Mass	$\lambda(m_H) = \frac{L(m_H, \hat{\hat{\mu}}_1, \hat{\hat{\mu}}_2, \hat{\hat{\theta}})}{L(\hat{m}_H, \hat{\mu}_1, \hat{\mu}_2, \hat{\theta})}$
Signal and mass	$\lambda(\mu, m_H) = \frac{L(\mu, m_H, \hat{\hat{\theta}}_\mu)}{L(\hat{\mu}, \hat{m}_H, \hat{\theta}_\mu)}$

ABCD methods

- Data-driven estimation of background in signal region
- Assumes two uncorrelated distributions of background
- For measurements linear propagation of uncertainties works fine
- For searches, small numbers, this may not be such a good approximation
- "Let's write down the likelihood function" - Glen Cowan



A: DDB

B: $DDB\tau_B$

C: $DDB\tau_C$

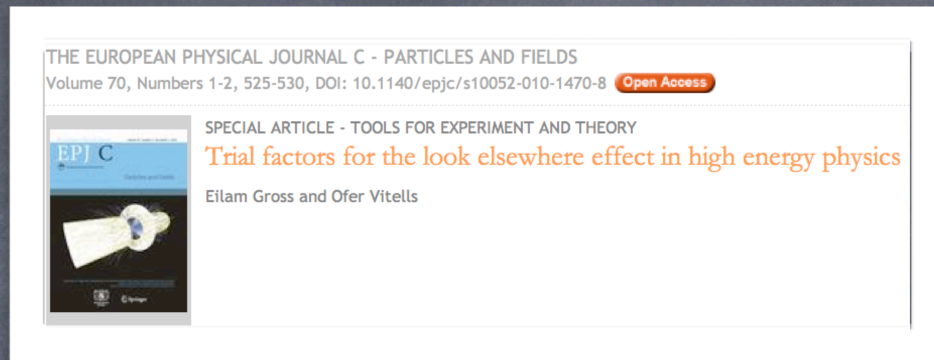
D: $DDB\tau_B\tau_C$

$$L(n_A, n_B, n_C, n_D|\mu, \theta) = \Pi_{i=A,B,C,D} \frac{e^{-\mu_i}\mu_i^{n_i}}{n_i!} f(\theta)$$

$$\begin{aligned} \mu_A &= \mu + A_{bkgMC} + DDB \\ \mu_B &= \epsilon_B \mu + B_{bkgMC} + DDB\tau_B \\ \mu_C &= \epsilon_C \mu + C_{bkgMC} + DDB\tau_C \\ \mu_D &= \epsilon_D \mu + D_{bkgMC} + DDB\tau_B\tau_C \end{aligned}$$

- 4 measurements
- μ and 3 N.P.'s
  - Uncertainty on μ is the challenge

Toy study of LEE


THE EUROPEAN PHYSICAL JOURNAL C - PARTICLES AND FIELDS
Volume 70, Numbers 1-2, 525-530, DOI: 10.1140/epjc/s10052-010-1470-8 Open Access

SPECIAL ARTICLE - TOOLS FOR EXPERIMENT AND THEORY

Trial factors for the look elsewhere effect in high energy physics

Eilam Gross and Ofer Vitells

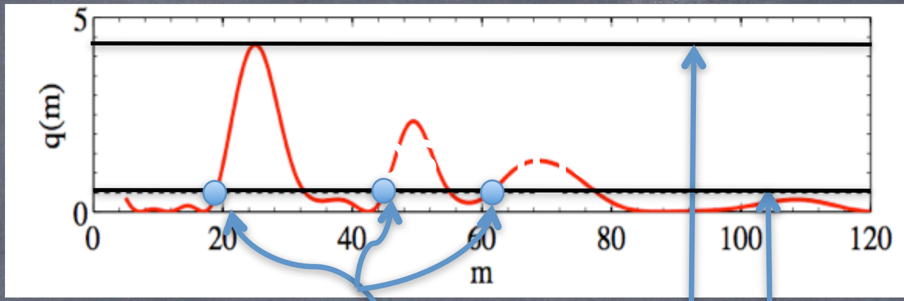
- Wanted to verify conclusions of Gross&Vitells Look elsewhere paper with higher-stats MC.
- Illustrates fits, asymptotics, limits of asymptotics
- Hypothetical signal is gaussian with fixed width of 0.05
- Background is mean of 200 events uniformly distributed between 0 and 1

Look-elsewhere effect (LEE)

Ex: 10^7 searches with 10^{-7} background

- Expect on the average 1 event with local p-value of 10^{-7}, but this is NOT a 5.2σ discovery!
- Probability to make a false discovery is $P(n \geq 1 | b = 1) = 1 - e^{-1}(-1)^0/1! = 63\%$
- Trials factor p_0^{global}/p_0^{local} from LEE is 0.63×10^7

Gross&Vitels: LEE in LLR-based search.



$$p_0^{\rm global} \simeq p_0^{\rm local} + < N(q_{\rm ref}) > e^{-(q - q_{\rm ref})/2}$$

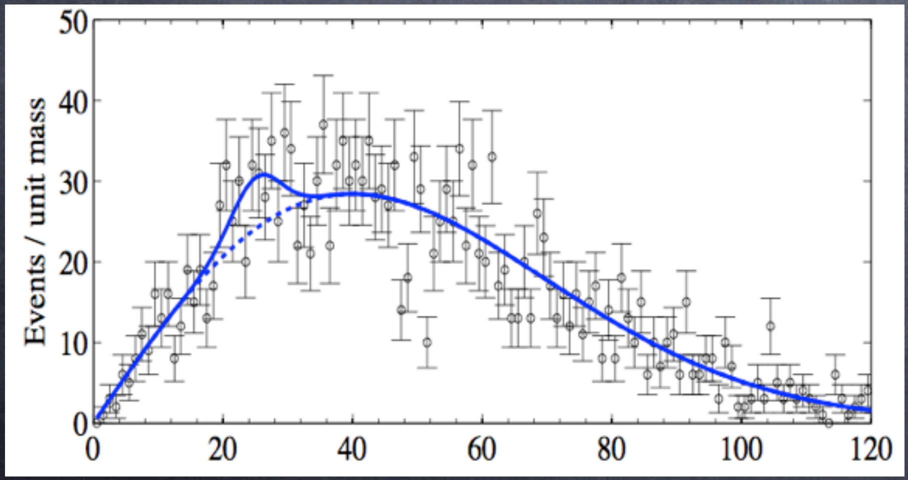
$$p_0^{\rm global} = P(q(\hat{m}) > Z^2)$$

$$p_0^{\rm local} = P(q(m) > Z^2)$$

$$q \leftrightarrow q_o$$

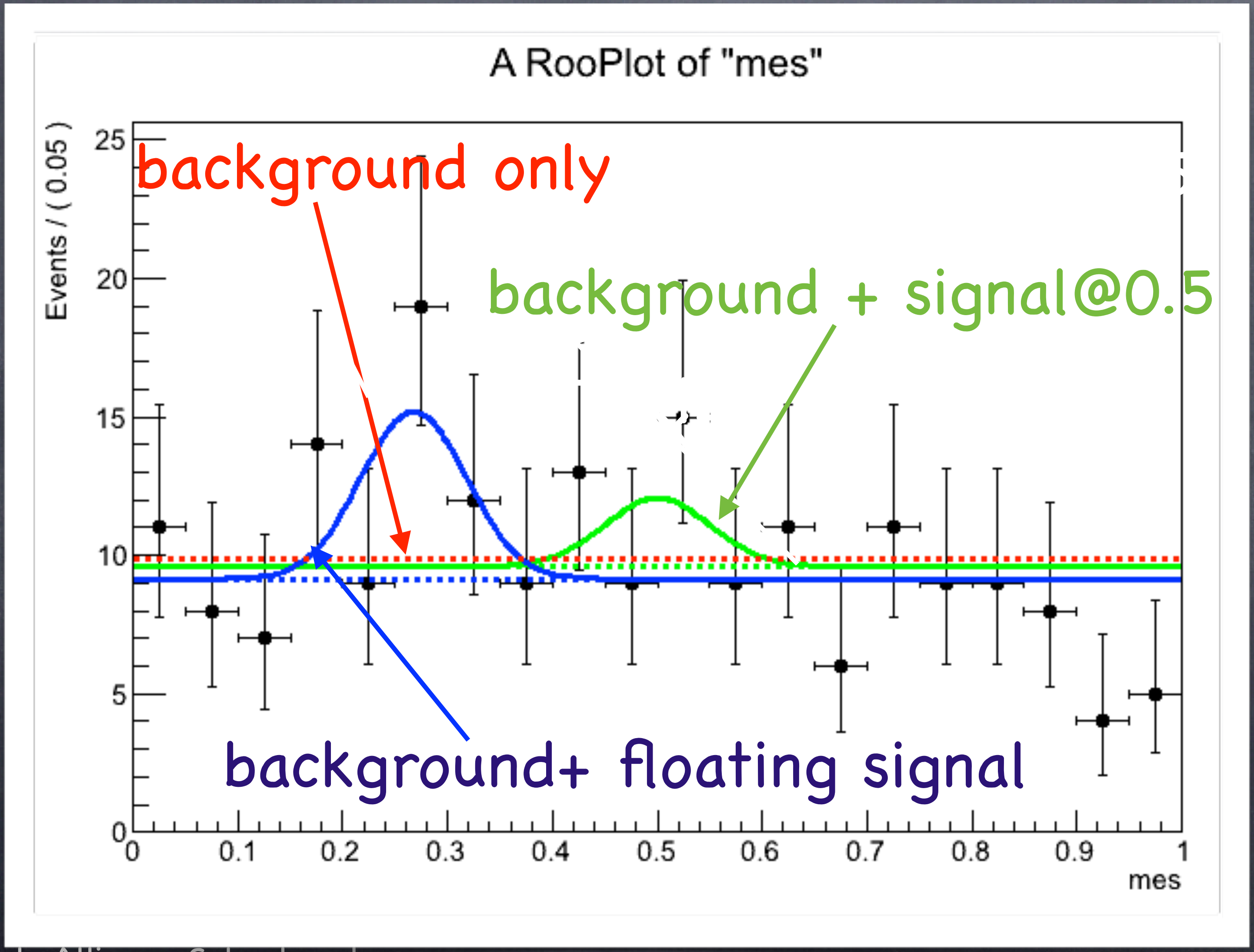
$$TF = \frac{P(q(\hat{m}) > Z^2)}{P(q(m) > Z^2)} \simeq 1 + \mathcal{N} \frac{P(\chi_2^2 > Z^2)}{P(\chi_1^2 > Z^2)}$$

$$TF \simeq 1 + \sqrt{\frac{\pi}{2}} \mathcal{N} Z$$

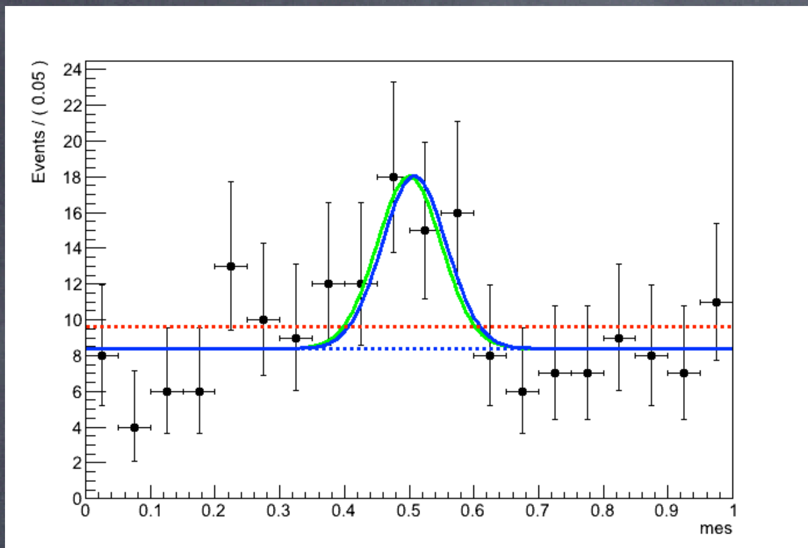


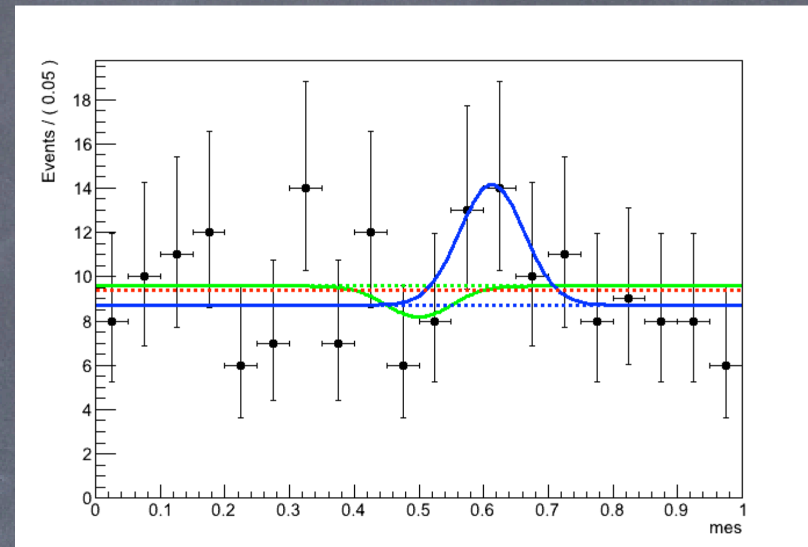
Want to verify these with high-statistics at high significance

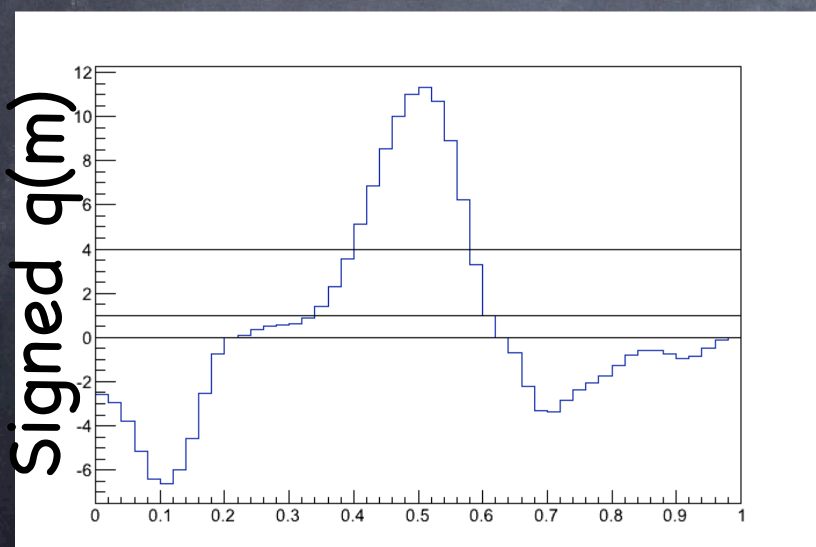
Fits to background toy

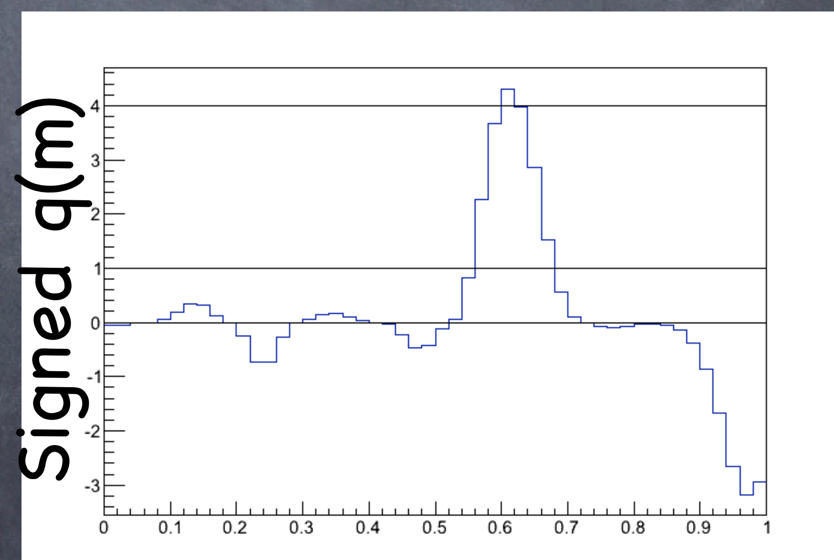


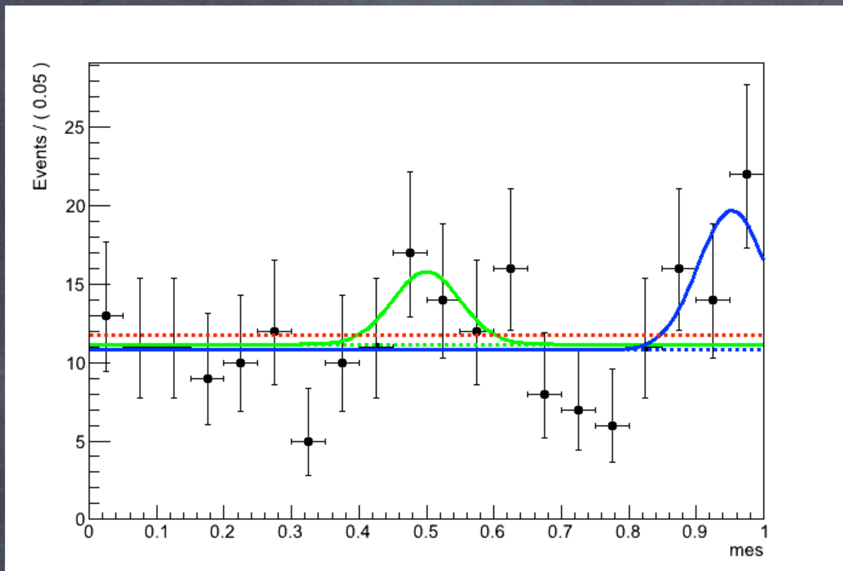
Six pseudo-experiments

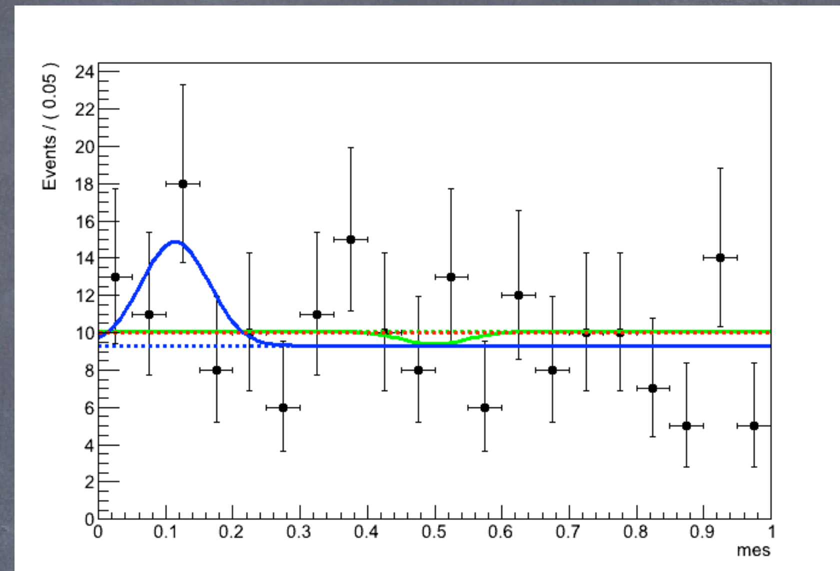


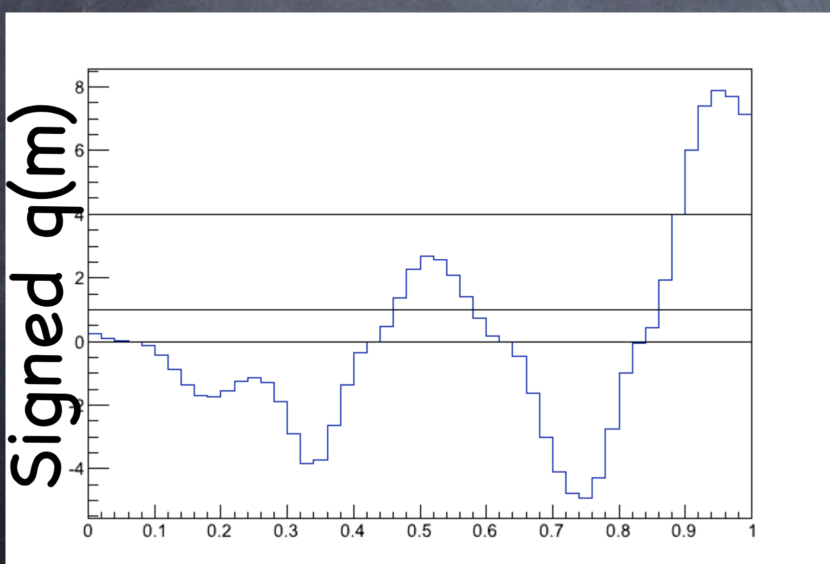


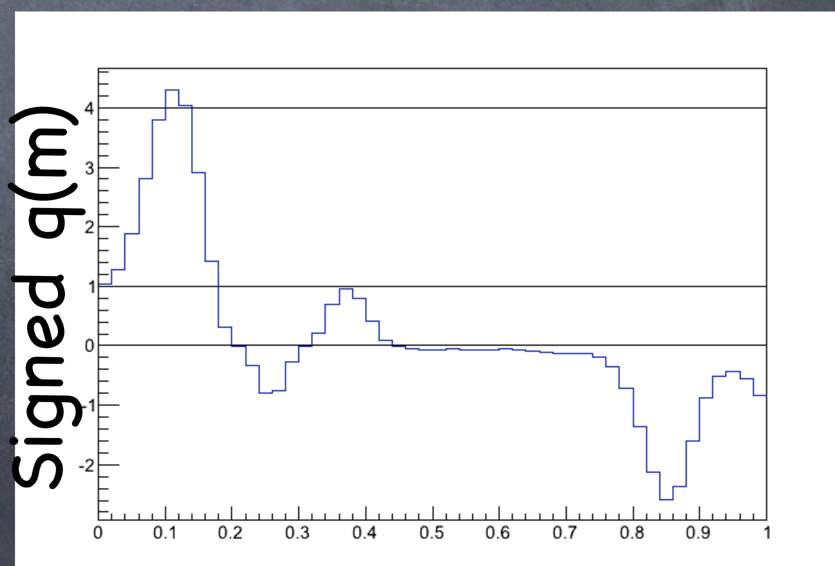


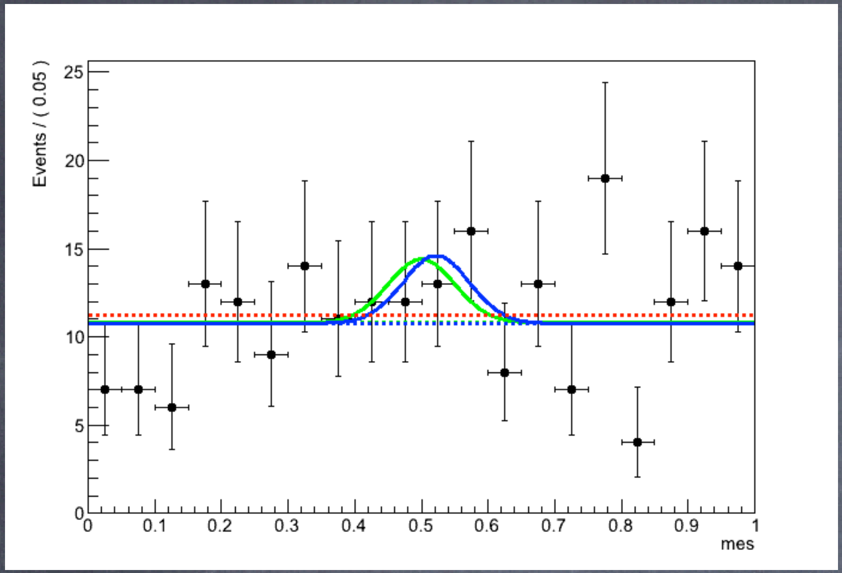



Events / (0.05)
mes

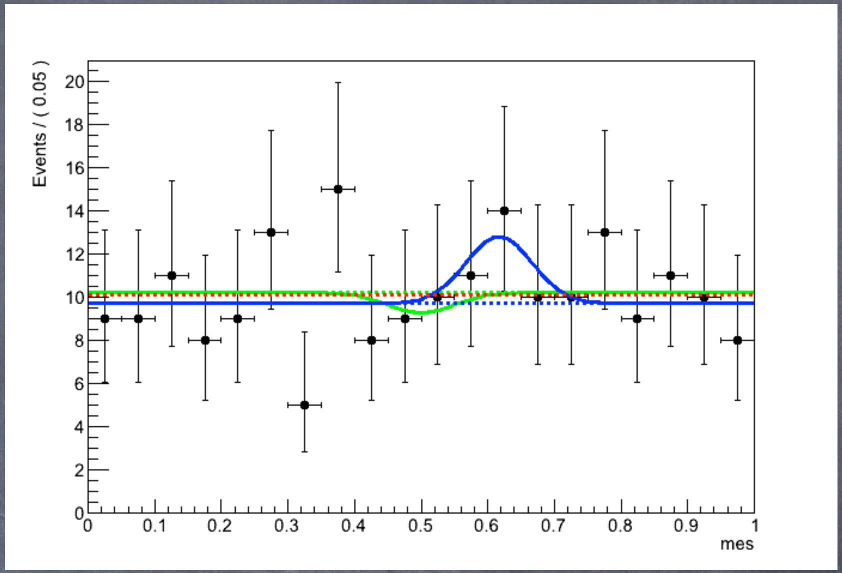

Events / (0.05)
mes


Signed q(m)

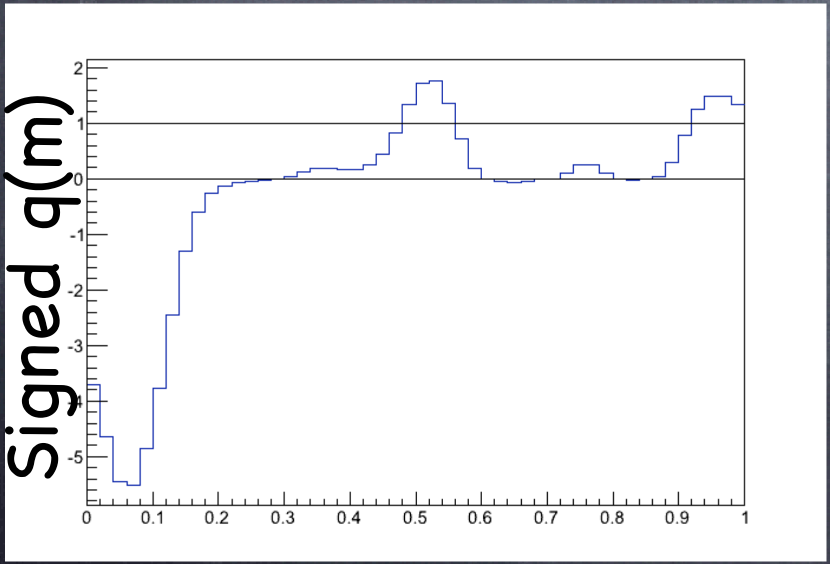

Signed q(m)



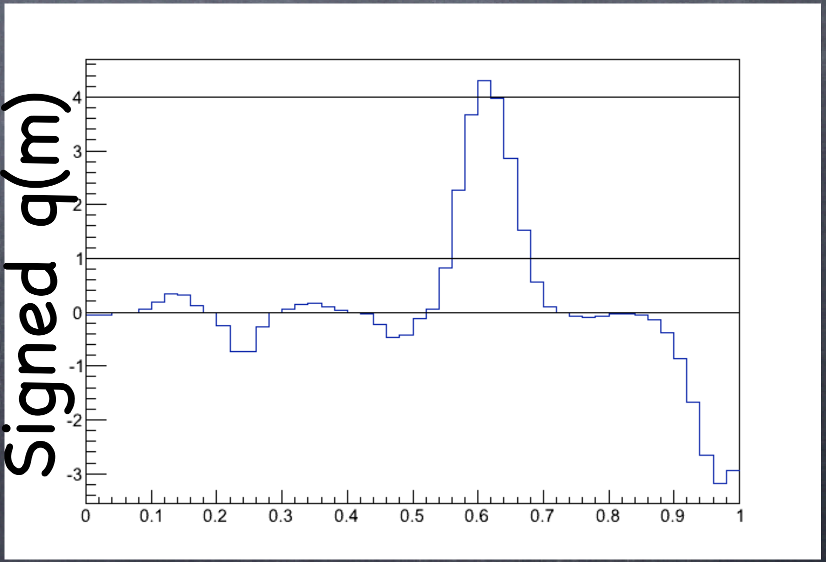
Events / (0.05)
mes



Events / (0.05)
mes

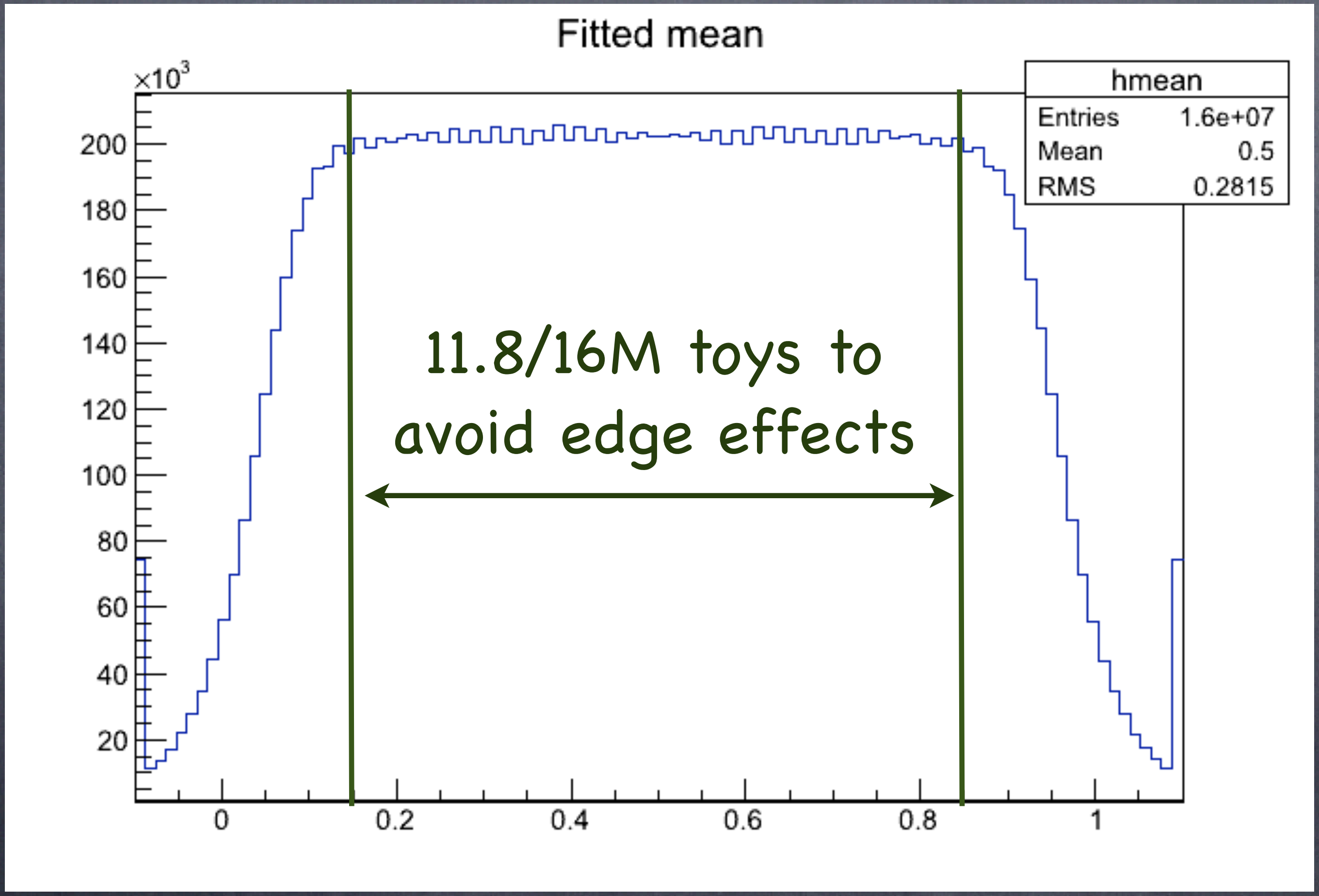


Signed q(m)

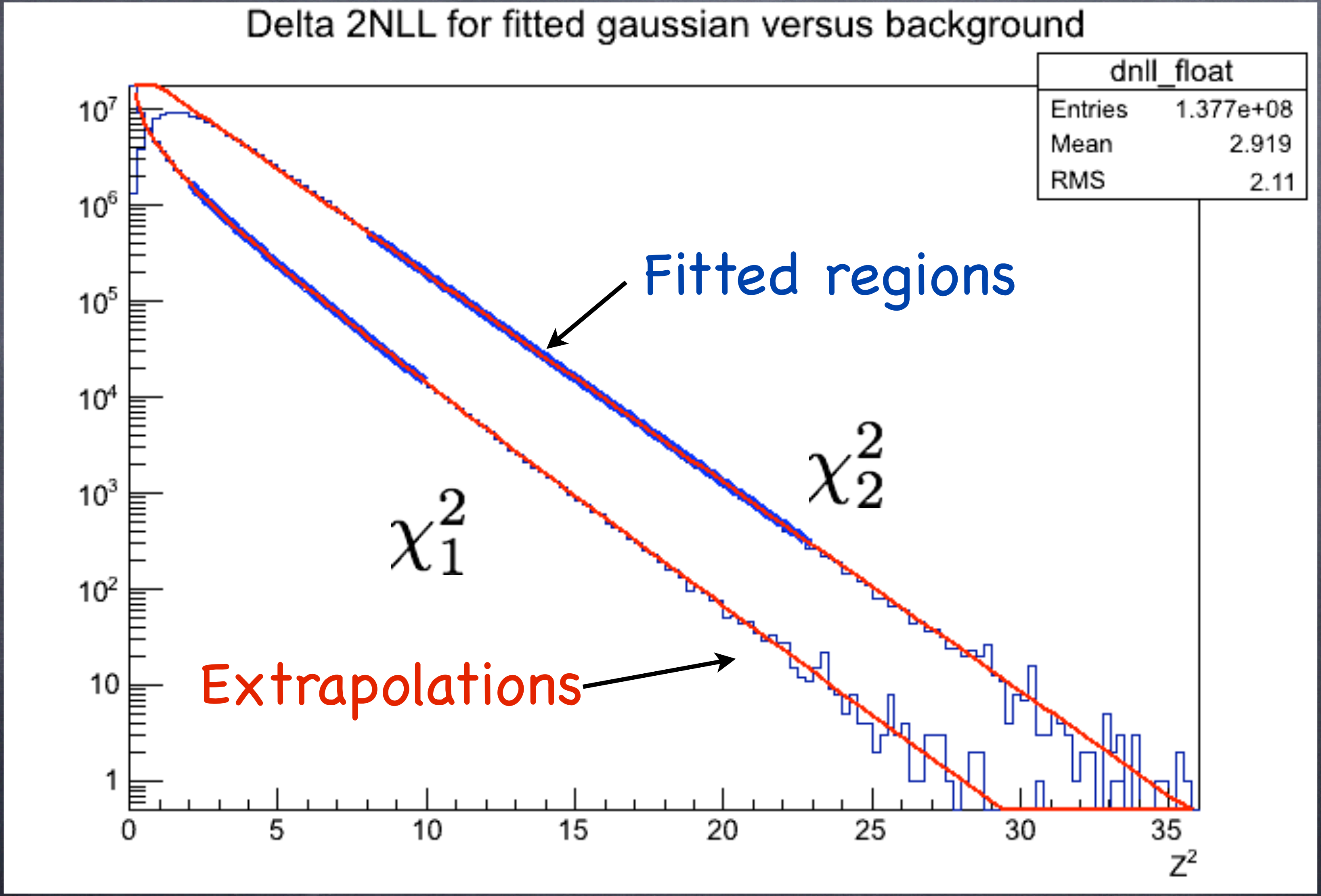


Signed q(m)

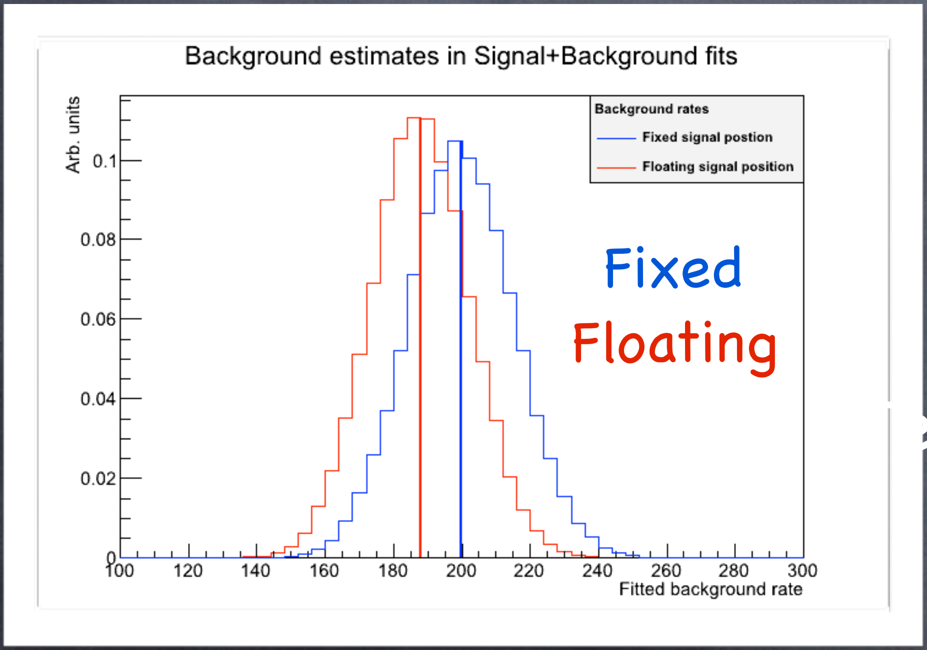
Extend to 4σ - need Mfits!

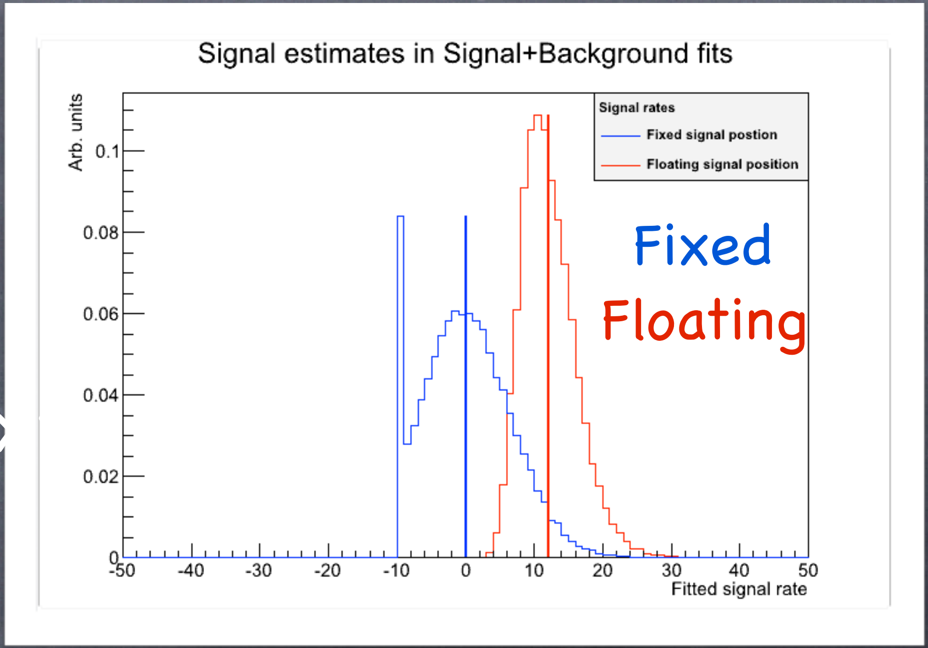


11.8 Mfits away from edges



Alternate view of Z^2 bias





- Background biased down
- Signal biased up

Summary

- "CL_S method" @ LHC rebaptized "criterion" or "procedure" to distinguish Q_{LEP} from q_μ, q_0
  - LR -> Profile LR (fits, marginalization of nuisance parameters)
  - More or less the same fits and profile likelihood ratios can be used for both exclusion/discovery and measurement
- Key to statistical analysis of searches and measurements: "Let's write down the likelihood function for your observations."
- Look-elsewhere an example of limitation of asymptotic expressions (but we know how to treat it)