Various statistical issues

14.02.2013

Statistics miniworkshop, CERN

A. Read (U. Oslo)

UiO University of Oslo

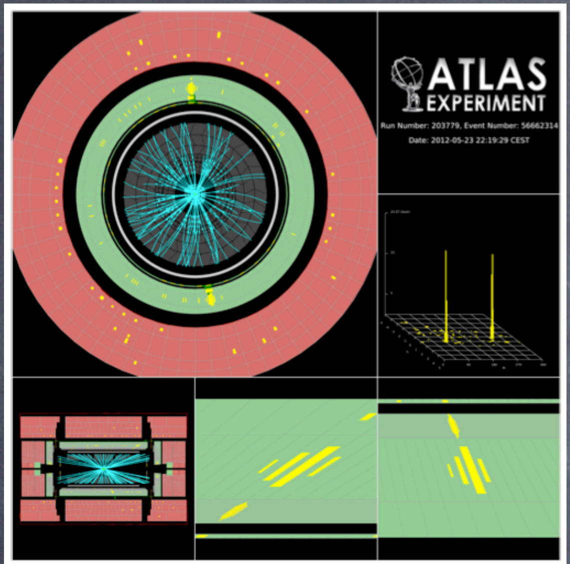
The Research Council of Norway

Issues

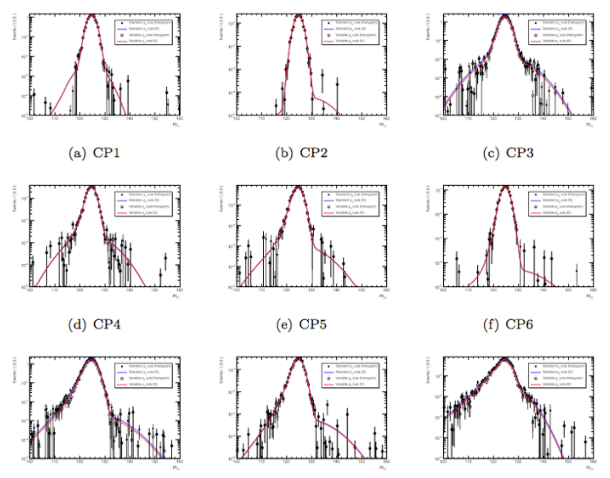
- Background parameterization
- Look elsewhere, empirical study
- Uncapping revisited
- Energy scale systematics revisited
- Importance sampling

Parameterized background model
e.g. ATLAS H->$\gamma\ \gamma$ search

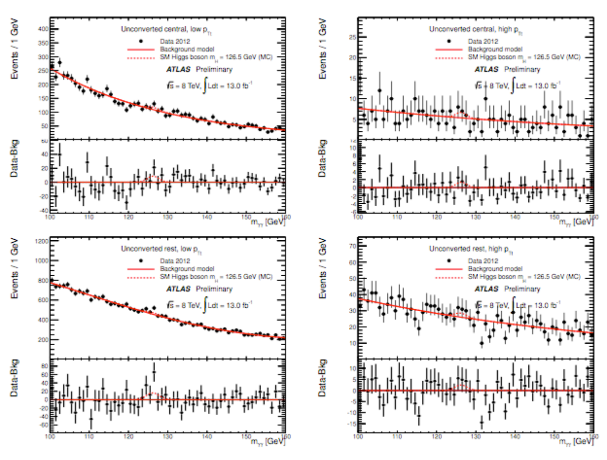
9 categories of unbinned likelihood



Name	Criteria		
CP1	unconverted	central	low p_{Tt}
CP2	unconverted	central	high p_{Tt}
CP3	unconverted	non-central	low p_{Tt}
CP4	unconverted	non-central	high p_{Tt}
CP5	converted	central	low p_{Tt}
CP6	converted	central	high p_{Tt}
CP7	converted	non-central	low p_{Tt}
CP8	converted	non-central	high p_{Tt}
CP9	converted	transition	



Parameterized signal model from fits to MC



4/9 categories

Background model: <u>selected</u> functions with unconstrained nuisance parameters

Various terms in L

$$\mathcal{L}_c(\mu, \boldsymbol{\theta}_c) = e^{-N_c} \prod_{n=1}^{N_c} \mathcal{L}_{c,n}(m_{\gamma\gamma}(n); \mu, \boldsymbol{\theta}_c)$$

L per event in a category

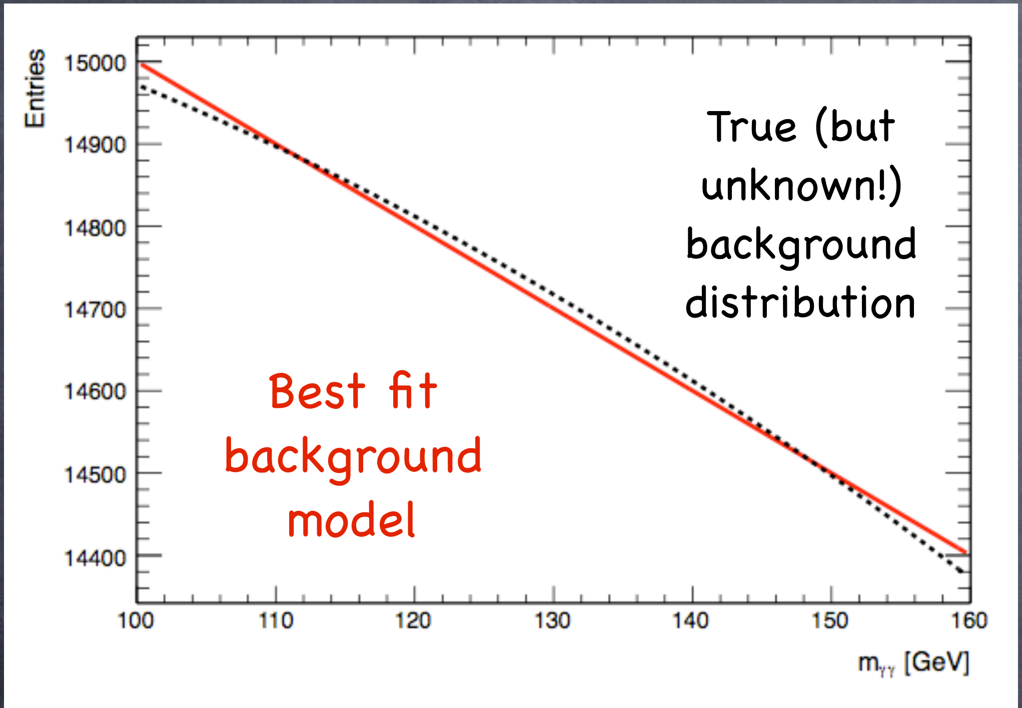
$$\mathcal{L}_{c,n}(m_{\gamma\gamma}(n); \mu, \boldsymbol{\theta}_c) = N_{s,c}(\mu, \boldsymbol{\theta}_c^{norm}) f_{s,c}(m_{\gamma\gamma}; \boldsymbol{\theta}_c^{shape}) + N_{bkg,c} f_{bkg,c}(m_{\gamma\gamma}; \boldsymbol{\theta}_c^{bkg}) \, ,$$

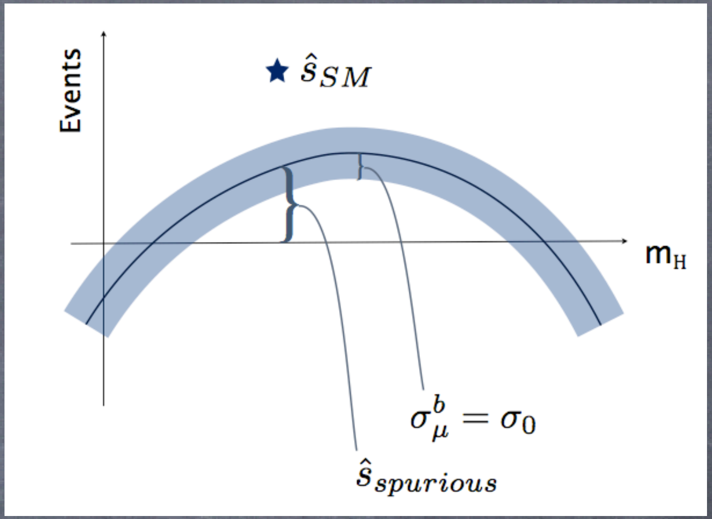
Mass distribution

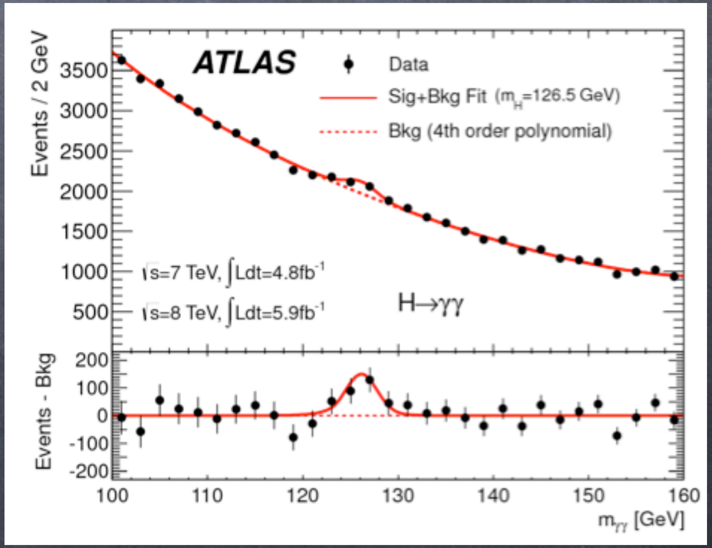
$$\begin{aligned} N_{s,c}(\mu, \boldsymbol{\theta}_c^{norm}) = \; & \mu \, [N_c^{ggH,SM}(\boldsymbol{\theta}_c^{ggH}) + N_c^{VBF,SM}(\boldsymbol{\theta}_c^{VBF}) \\ & + N_c^{WH,SM}(\boldsymbol{\theta}_c^{WH}) + N_c^{ZH,SM}(\boldsymbol{\theta}_c^{ZH}) + N_c^{ttH,SM}(\boldsymbol{\theta}_c^{ttH})] \\ & \cdot K_{BR}(\theta_{BR}) \, K_{lumi}(\theta_{lumi}) \, K_{eff}(\theta_{eff}) \, K_{isol}(\theta_{isol}) \\ & K_{pile-up}(\theta_{pile-up}) K_{EScale}(\theta_{EScale}) \\ & K_{pile-up,c}(\theta_{pile-up,c}) K_{mat,c}(\theta_{mat}) \\ & + \sigma_{spurious,c} \, \theta_{spurious,c} \, . \end{aligned} \tag{8.12}$$

Signal normalization

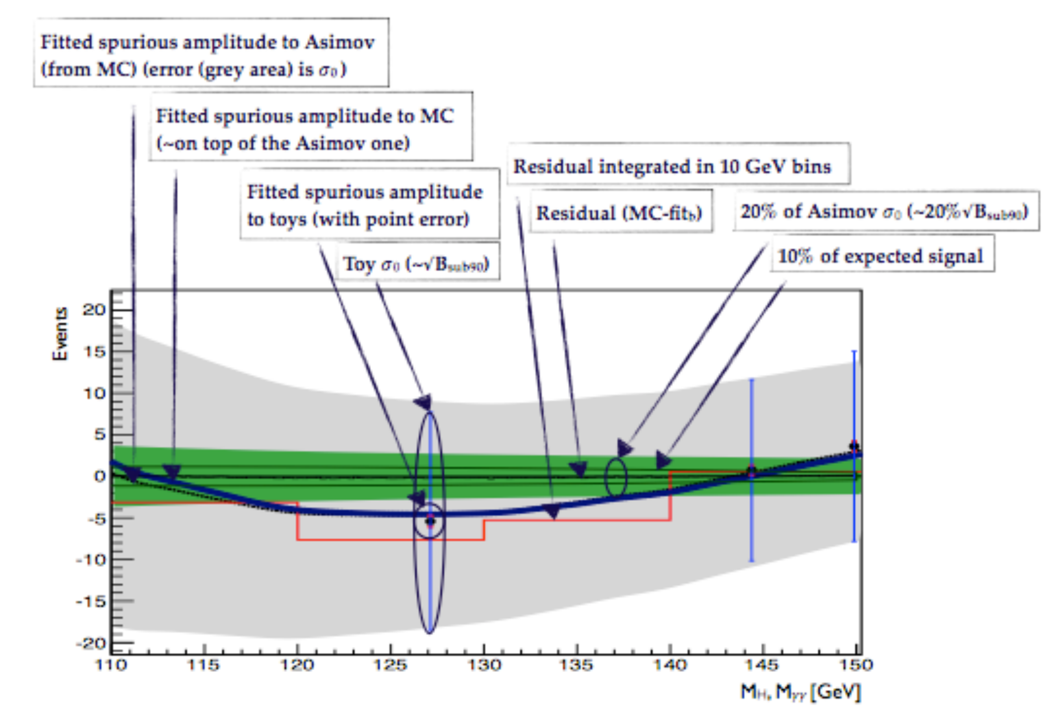
Distinguish signal from spurious signal







Model tests (on MC)



- 9 categories
- No CPU time for full simulation
- 3 MC generators, don't expect them to perfectly reproduce the background data
- Select parameterizations which can incorporate shape uncertainty in unconstrained nuisance parameters without producing false signals

BG model selection

Category	Function	Max $\|S_{SP}\|$ (m_H [GeV])	$\%\sqrt{S}\,(N_S)$	$\%\sigma_0\,(\sigma_0)$	σ^{N_S}	$\sigma^{S_{SP}}$	Pass	Pass$_{all}$
CP1	Exp	-4.7 (126)	-45 (11)	-35 (14)	0.78	-0.35		
CP1	Epoly2	2.1 (117)	18 (12)	13 (16)	0.70	0.13	✓	✓
CP2	Exp	-0.23 (110)	-15 (1.5)	-6.4 (3.5)	0.43	-0.064	✓	✓
CP3	Exp	12 (117)	50 (23)	35 (33)	0.71	0.35		
CP3	Epoly2	9.2 (112)	41 (23)	26 (36)	0.64	0.26		
CP3	Epoly3	3.4 (111)	15 (22)	8.8 (38)	0.59	0.088	✓	✓
CP3	Bern3	5.8 (111)	26 (22)	16 (36)	0.62	0.16		
CP3	Bern4	2.8 (111)	13 (22)	7.1 (40)	0.56	0.071	✓	✓
CP4	Exp	0.5 (132)	19 (2.6)	7.2 (6.9)	0.38	0.072	✓	✓
CP5	Exp	-4.4 (126)	-64 (6.8)	-34 (13)	0.54	-0.34		
CP5	Epoly2	1.6 (117)	22 (7.4)	10 (16)	0.47	0.10	✓	✓
CP6	Exp	-0.27 (110)	-27 (0.98)	-8.0 (3.4)	0.29	-0.080	✓	✓
CP7	Exp	6.5 (122)	29 (22)	18 (37)	0.60	0.17		
CP7	Epoly2	5.8 (122)	26 (22)	14 (40)	0.56	0.14		
CP7	Epoly3	-6.3 (110)	-29 (22)	-13 (48)	0.46	-0.13	✓	✓
CP7	Bern3	-6.3 (110)	-29 (22)	-14 (46)	0.47	-0.14	✓	✓
CP7	Bern4	-4.5 (110)	-20 (22)	-8.8 (50)	0.43	-0.088	✓	✓
CP8	Exp	0.45 (134)	18 (2.5)	5.7 (7.9)	0.32	0.057	✓	✓
CP9	Exp	-16 (130)	-179 (9.1)	-59 (28)	0.33	-0.59		
CP9	Epoly2	-3.2 (110)	-33 (9.9)	-8.3 (39)	0.26	-0.083	✓	✓

- the exponential function

$$N e^{-\beta m_{\gamma\gamma}}\ , \quad (8.25)$$

where N and β were the fitted parameters – the normalization and slope of the exponential, respectively;

- the exponential polynomial of order n (orders 2 and 3 were used)

$$e^{\sum_{i=0}^{n} \beta_i m_{\gamma\gamma}^{i}}\ , \quad (8.26)$$

where β_i were the fitted parameters. Note that the latter i is not an index, but the power $m_{\gamma\gamma}$ is raised to. The normalization, N, is described by the first term, e^{β_0};

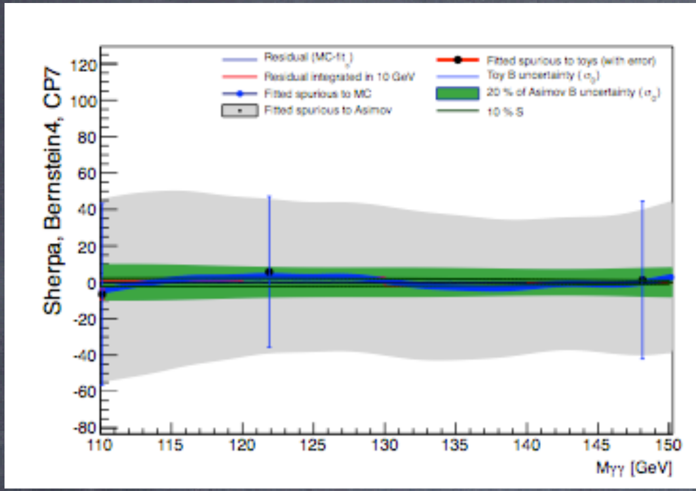
- the Bernstein polynomial of order n (orders 3 – 7 were used)

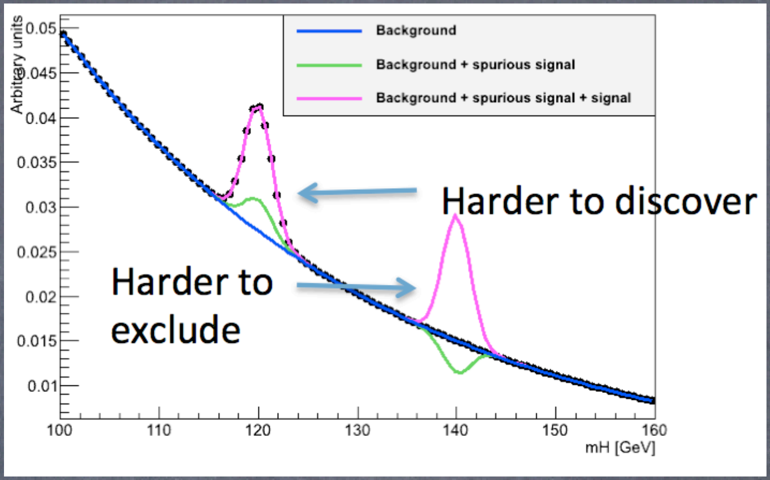
$$b_n(t) = \sum_{i=0}^{n} \beta_i \binom{n}{i} t^i (1-t)^{n-i}\ , \quad (8.27)$$

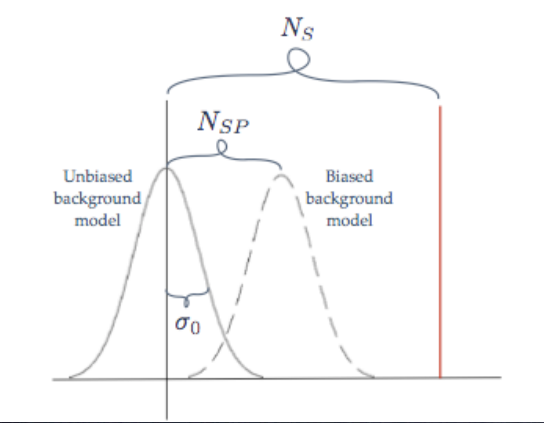
where $t = \frac{m_{\gamma\gamma}[GeV]-100}{60}$, and where β_i were fitted parameters.

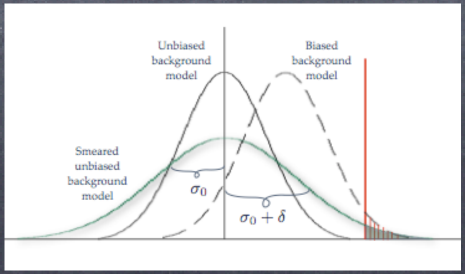
Maximum spurious signal amplitude

Residual (unknown!) bias: Spurious signal term in likelihood









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

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

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

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

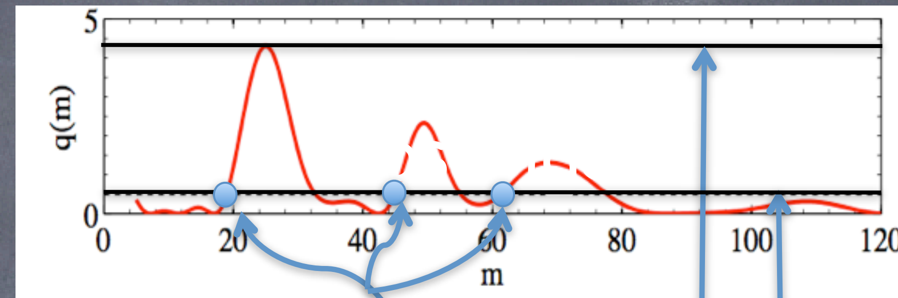
- Wanted to verify conclusions of Gross&Vitells Look elsewhere paper with higher-stats MC.
- (Illustrates fits, asymptotics, limits of asymptotics)
- Hypothetical signal is gaussian with fixed width of 0.05
- Background is mean of 200 events uniformly distributed between 0 and 1

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 5.2 σ 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.



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

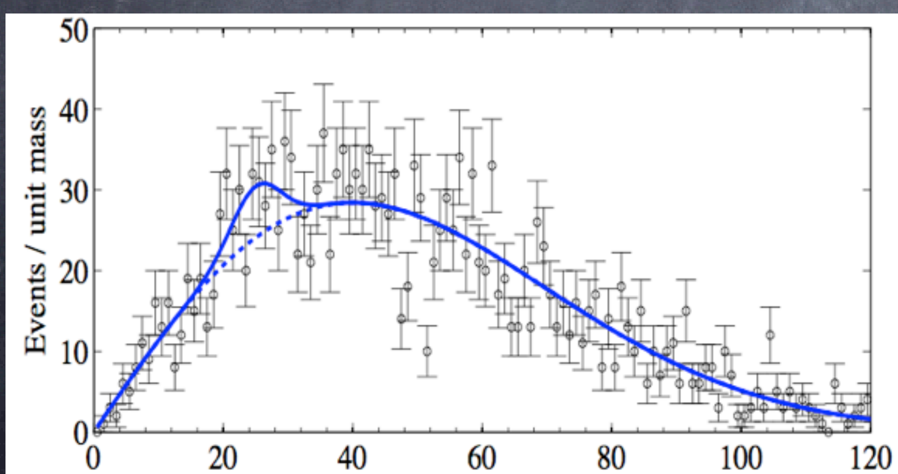
$$p_0^{\mathrm{global}} = P(q(\hat{m}) > Z^2)$$

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

$$q \leftrightarrow q_o$$

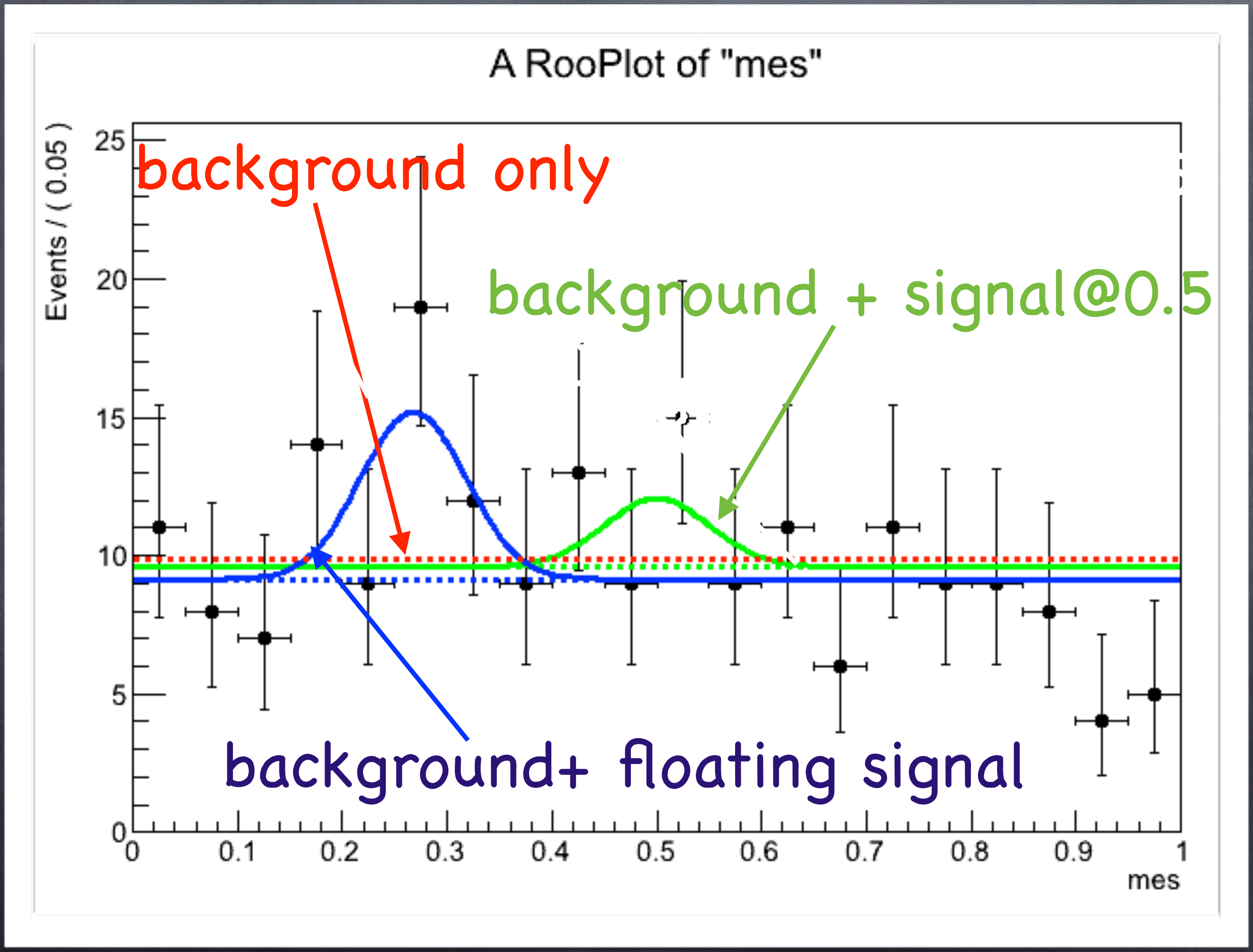
$$TF = \frac{P(q(\hat{m}) > Z^2)}{P(q(m) > Z^2)} \simeq 1 + \mathcal{N}\frac{P(\chi^2_2 > Z^2)}{P(\chi^2_1 > Z^2)}$$

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

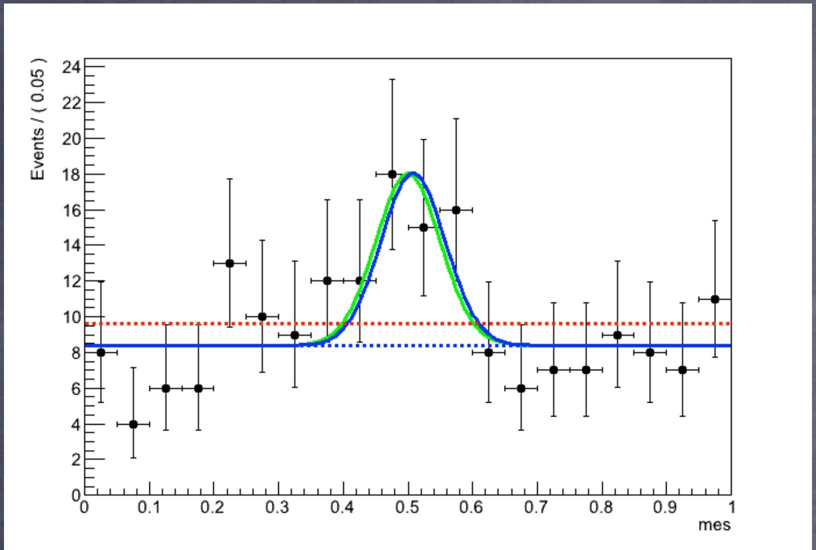


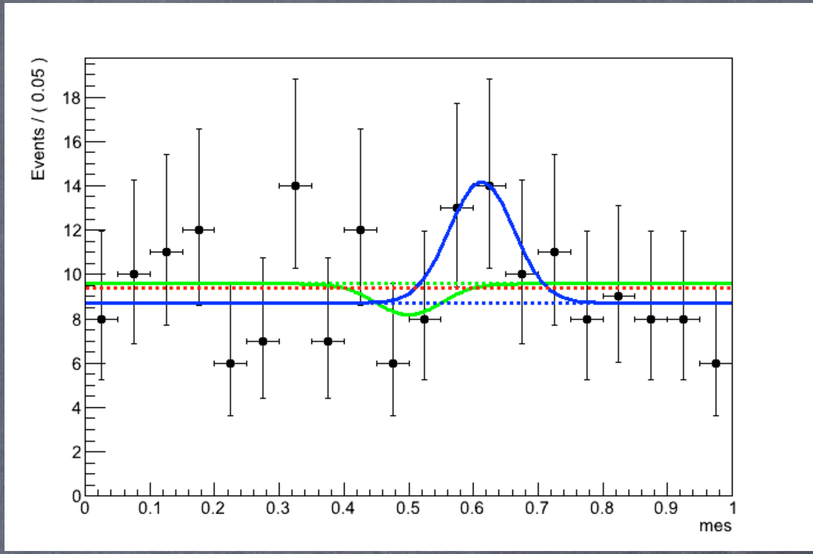
Want to verify these with high-statistics at high significance

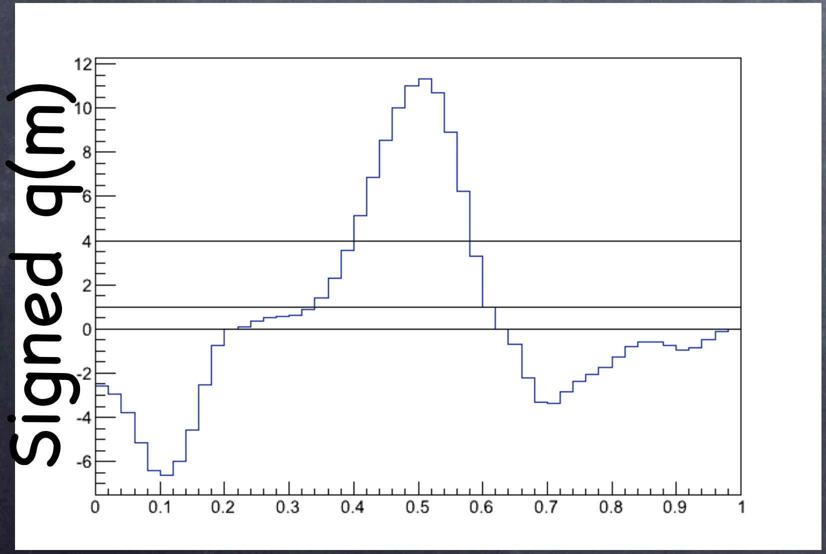
Fits to background toy

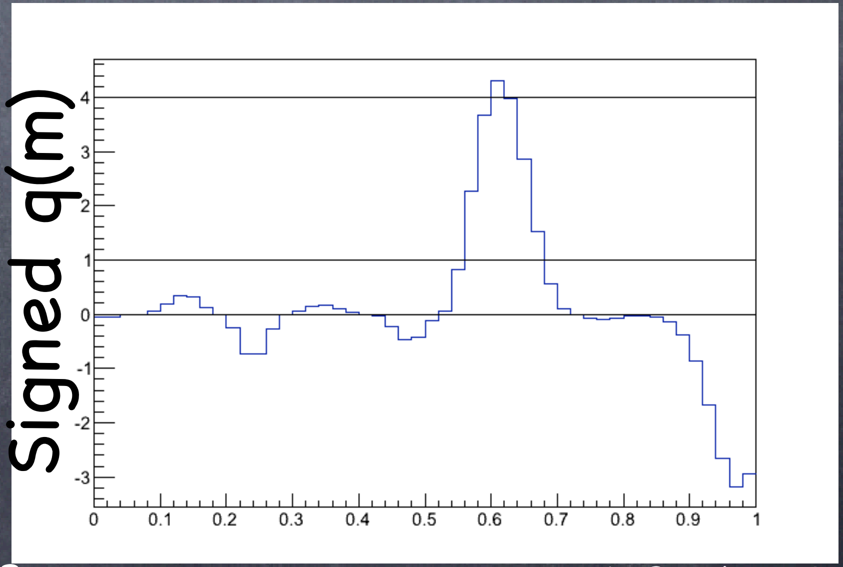


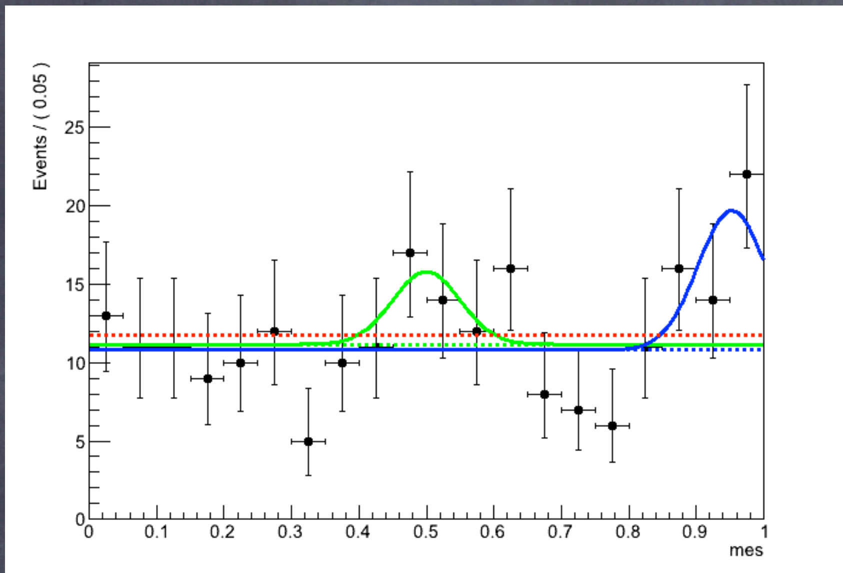
Six pseudo-experiments

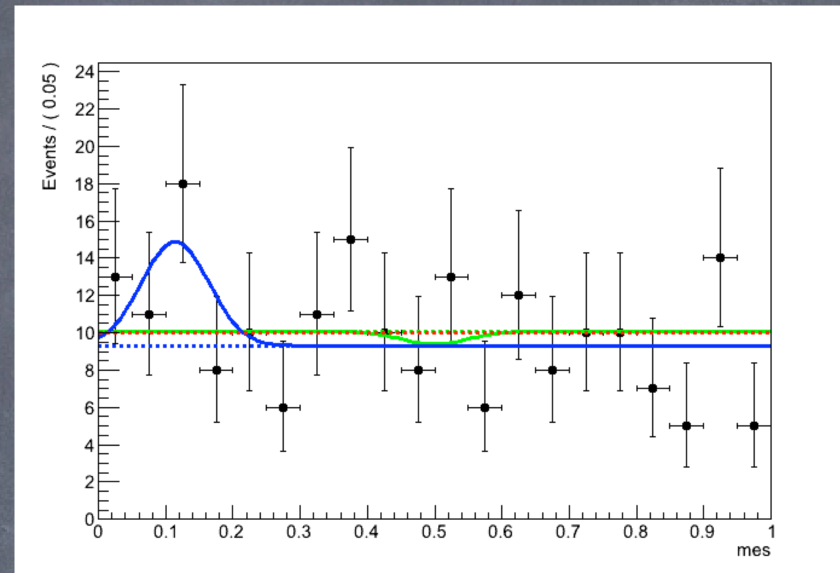


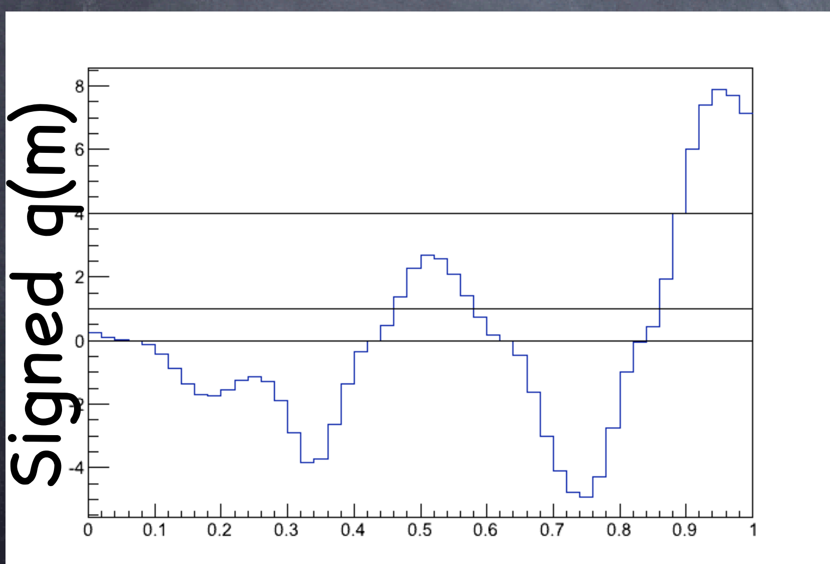


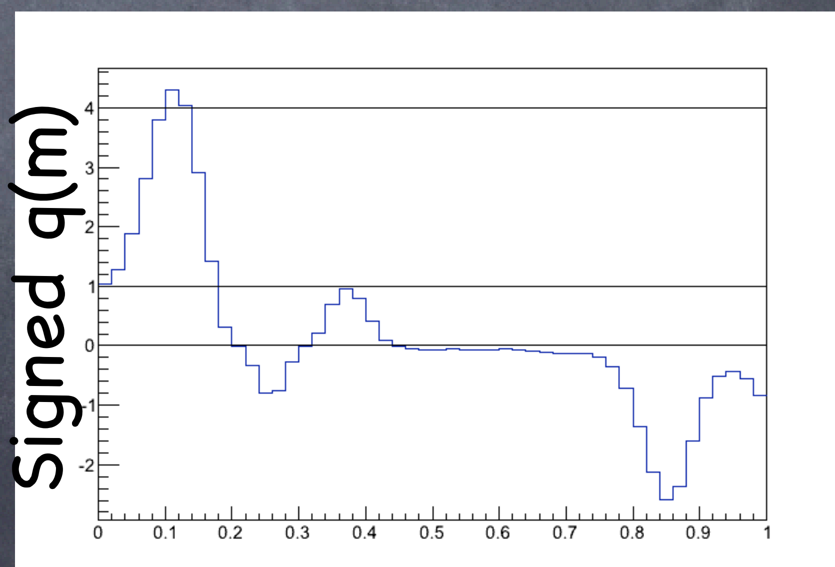


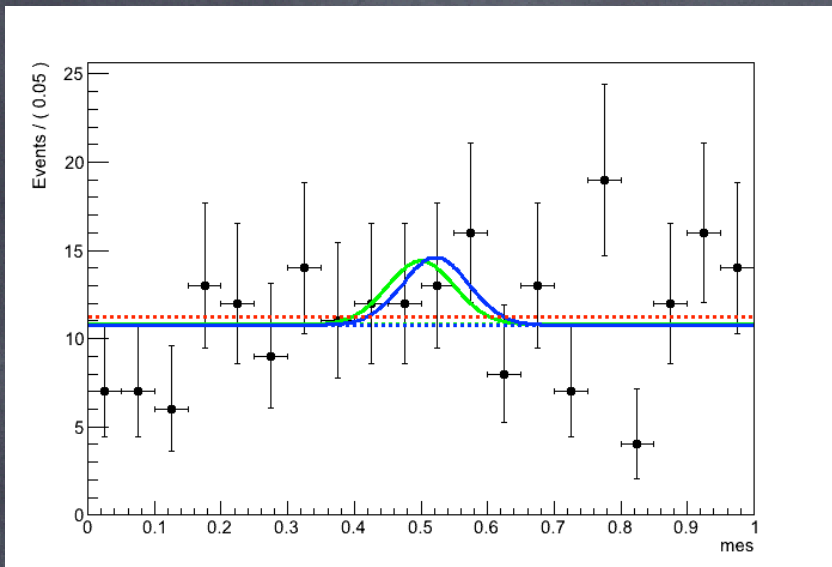


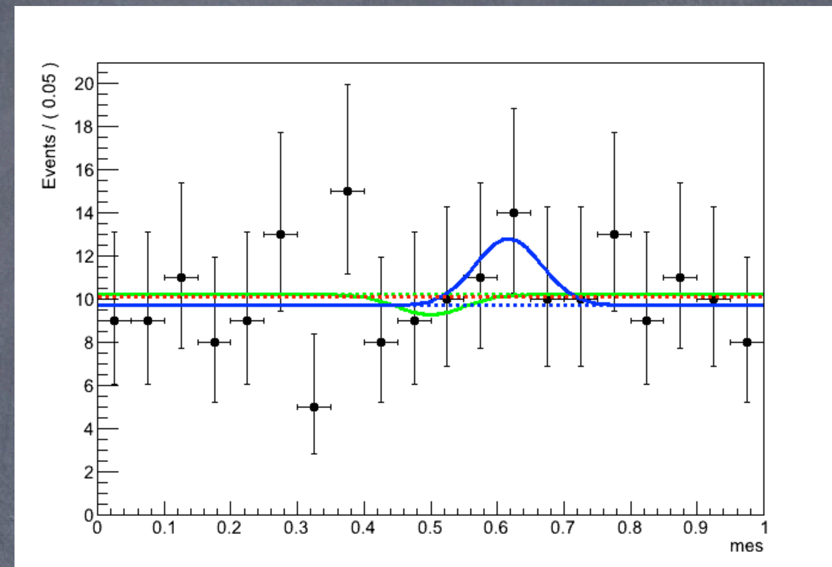

Events / (0.05)
0
5
10
15
20
25
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
mes

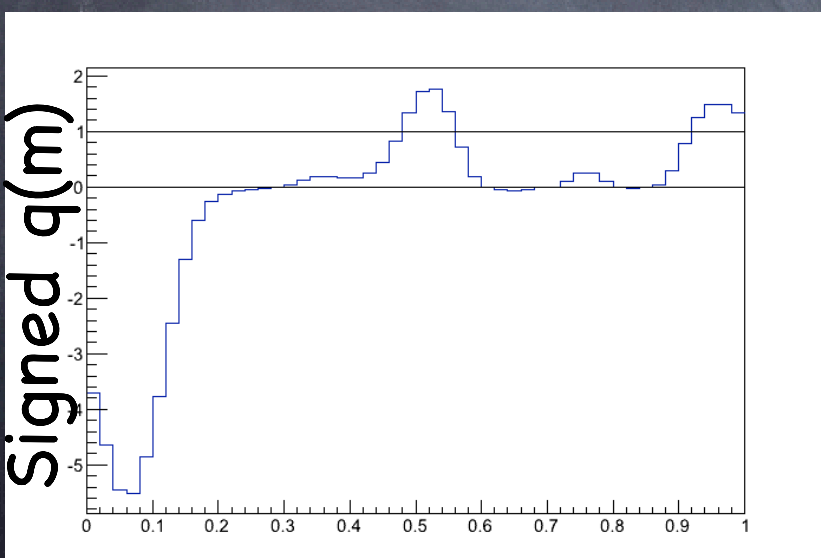

Events / (0.05)
0 2 4 6 8 10 12 14 16 18 20 22 24
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
mes

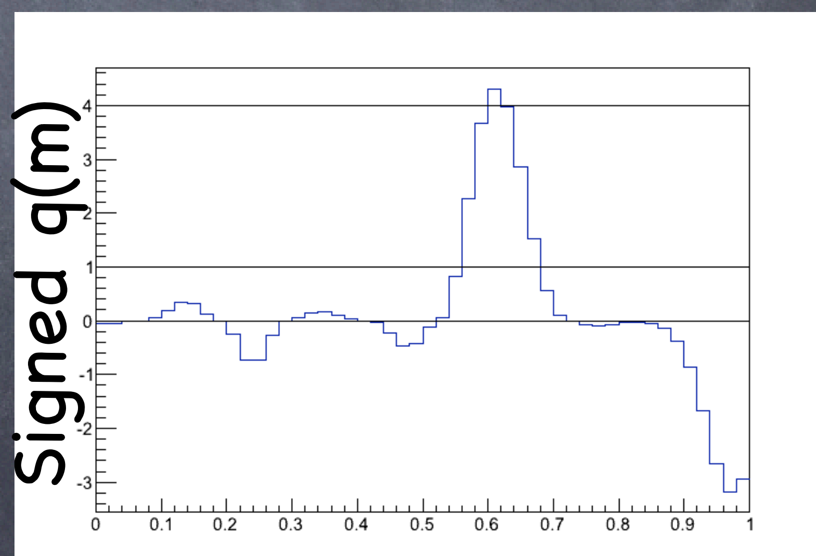

Signed q(m)
-4 -2 0 2 4 6 8
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1


Signed q(m)
-2 -1 0 1 2 3 4
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

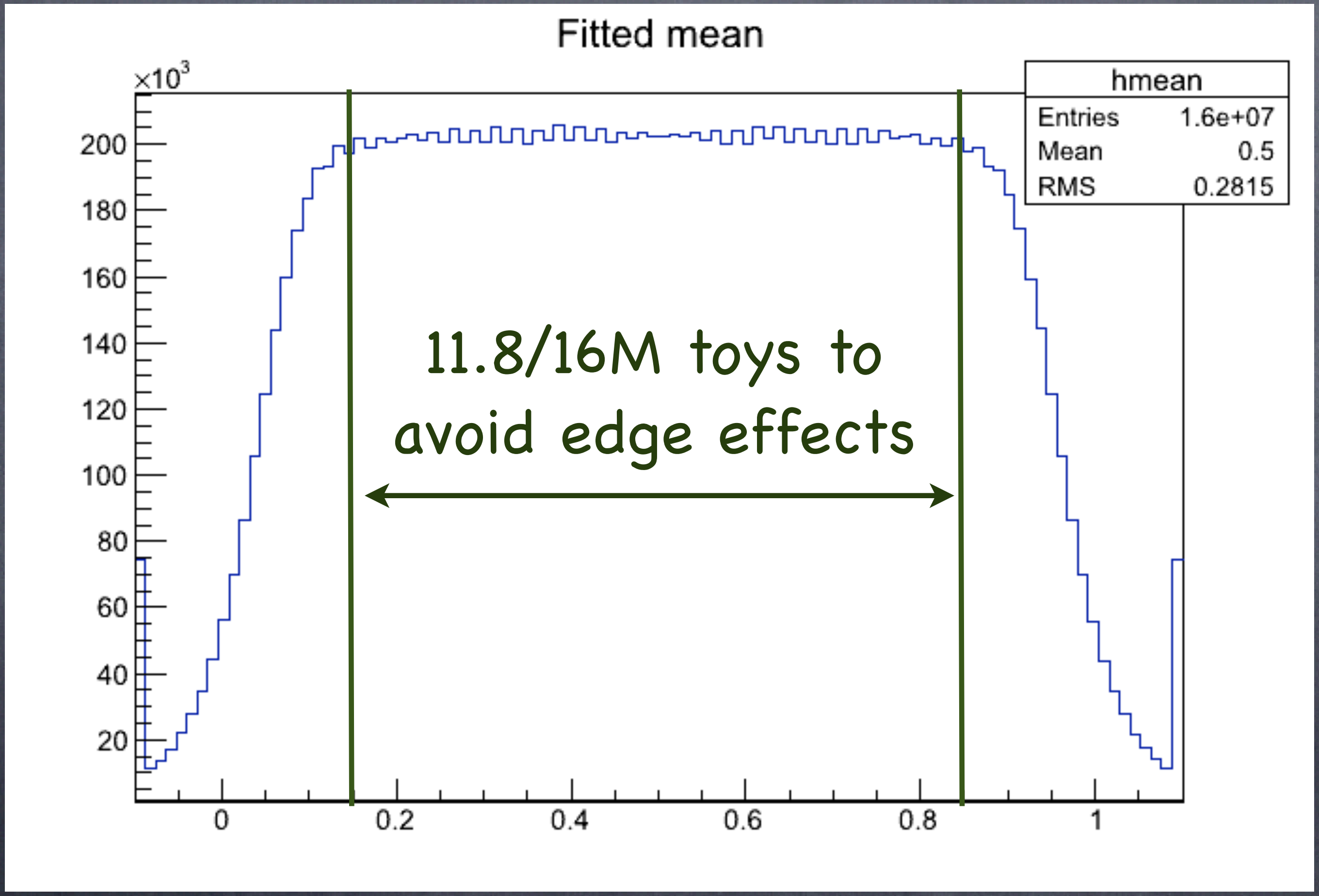

Events / (0.05)
mes


Events / (0.05)
mes

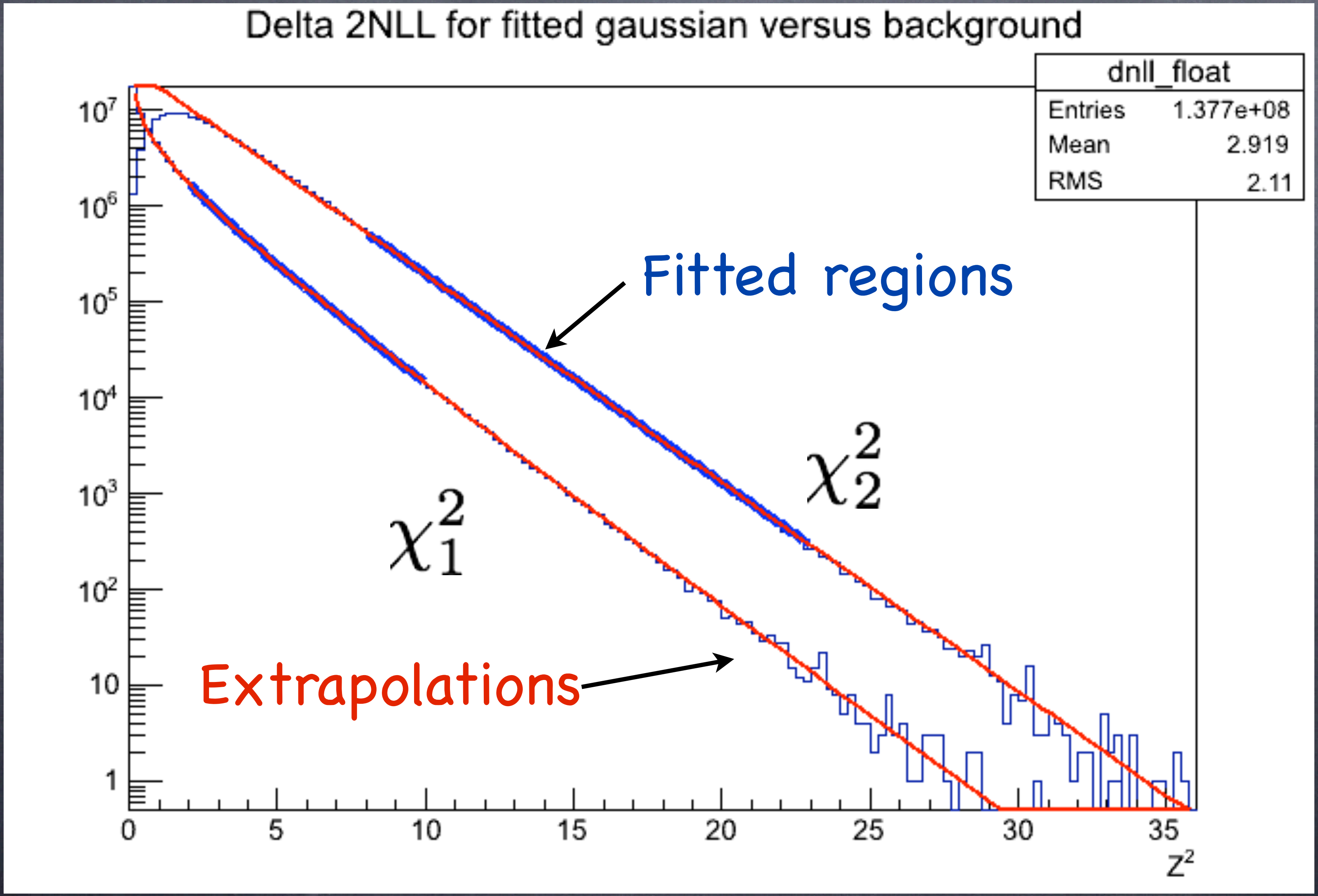

Signed q(m)


Signed q(m)

Extend to 4 σ - need Mfits!



138 Mfits away from edges



Study of Trials Factor

	2 σ	3 σ	4 σ	5 σ
$P(\hat{q}(m) > Z^2)$	2.2 10^{-2}	1.3 10^{-3}	3.2 10^{-5}	-
$P(\hat{q}(\hat{m}) > Z^2)$	2.8 10^{-1}	2.3 10^{-2}	6.9 10^{-4}	-
TF	12.9	17.9	21.5	-
$P(\chi^2_1 > Z^2)/2$	2.3 10^{-2}	1.3 10^{-3}	3.2 10^{-5}	2.9 10^{-7}
$P(\chi^2_2 > Z^2)/2$	6.8 10^{-2}	5.6 10^{-3}	1.7 10^{-4}	1.9 10^{-6}
$\mathcal{N}$	3.84	3.92	3.90	-
$TF \simeq 1 + \sqrt{\frac{\pi}{2}}\mathcal{N}Z$	10.7	15.6	20.5	25.4

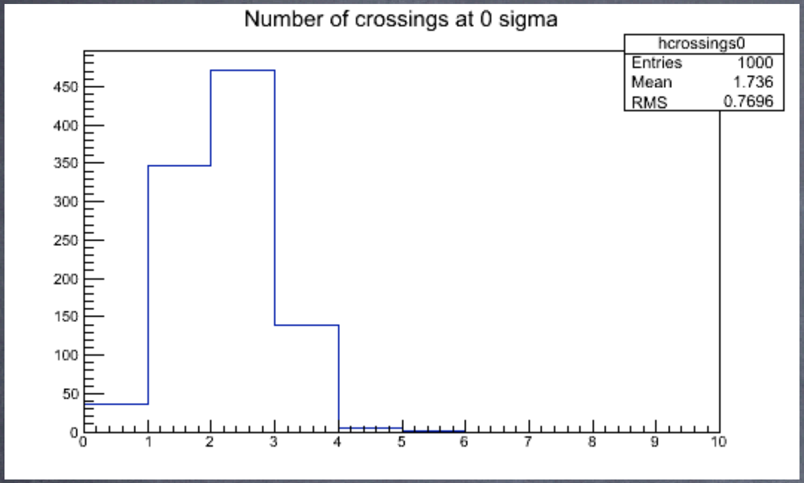
$$TF = \frac{P(q(\hat{m}) > Z^2)}{P(q(m) > Z^2)} \simeq 1 + \mathcal{N}\frac{P(\chi^2_2 > Z^2)}{P(\chi^2_1 > Z^2)}$$

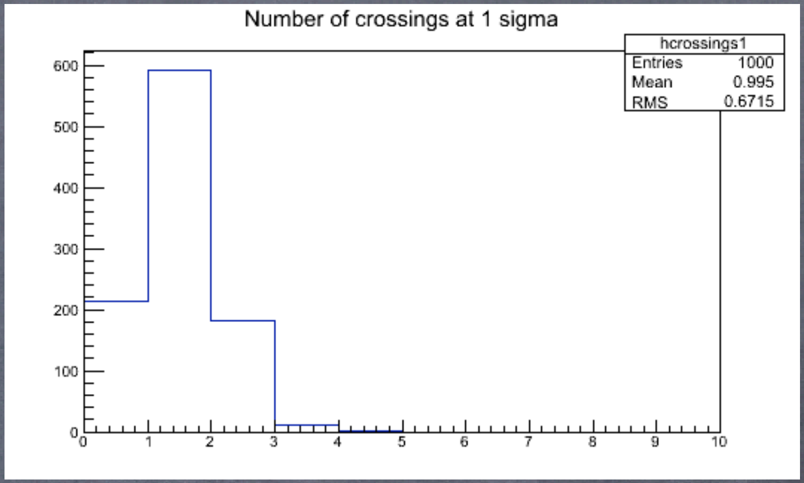
TF for 3 σ

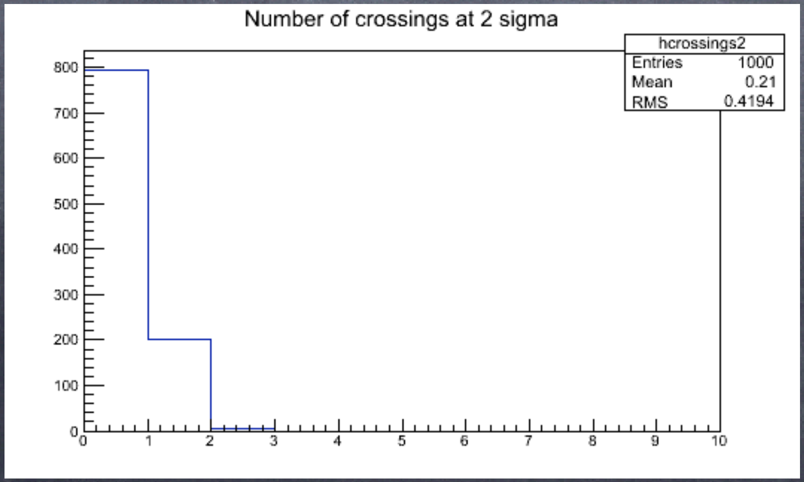
$$TF \simeq 1 + \sqrt{\frac{\pi}{2}} \mathcal{N} Z$$

- With N=3.89, TF=15.6
- Simple estimate Δm/ σ_m=14

Counting crossings





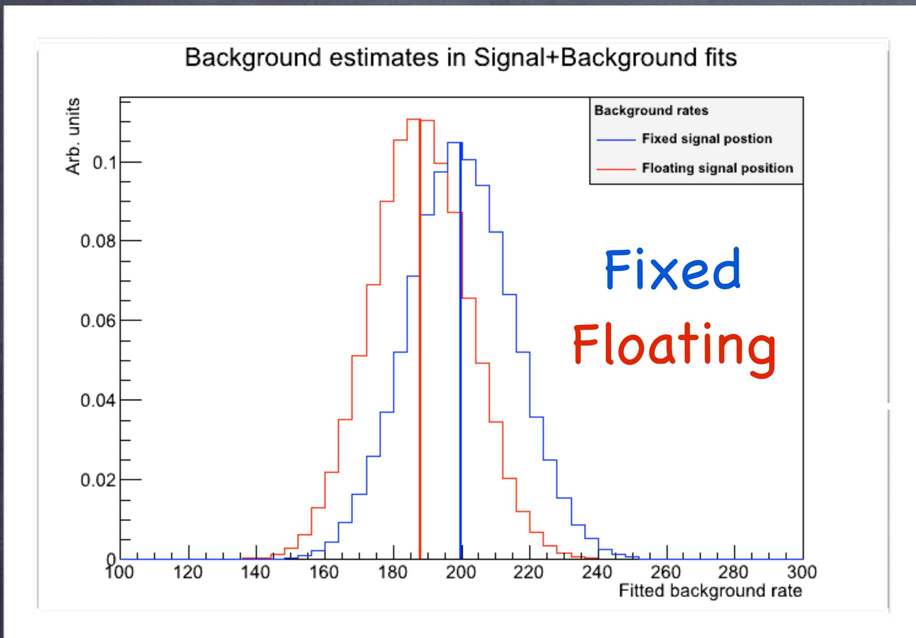


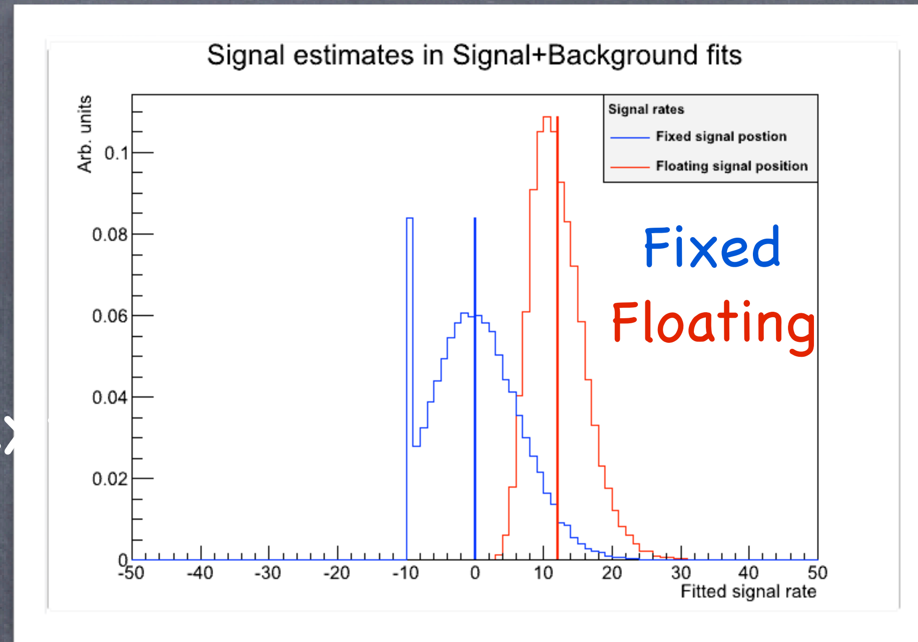
Study of crossings (see p. 5-7)

"Predicted" from: $$P(q(\hat{\theta}) > c) \leqslant P(\chi_s^2 > c) + \langle N(c_0)\rangle (\frac{c}{c_0})^{(s-1)/2} e^{-(c-c_0)/2} \quad (3)$$

	Average measured crossings <N>			
Crossing threshold	Range 0.05-0.95	Range 0.10-0.90	Range 0.15-0.85	Predicted
0 σ	2.19	1.99	1.74	1.95
1 σ	1.23	1.12	0.99	1.18
2 σ	0.29	0.26	0.23	0.26

Alternate view of Z^2 bias





- Background biased down
- Signal biased up

Uncapping revisted

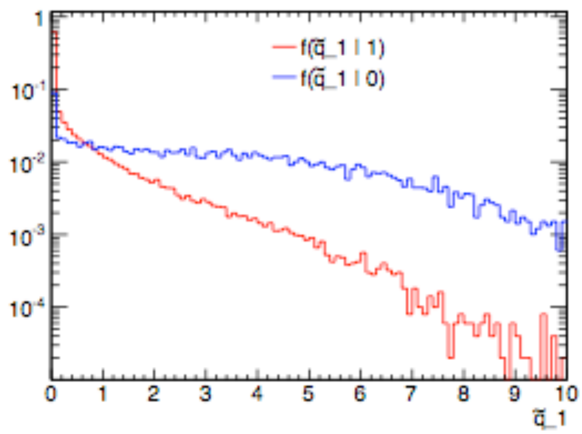
Motivation:

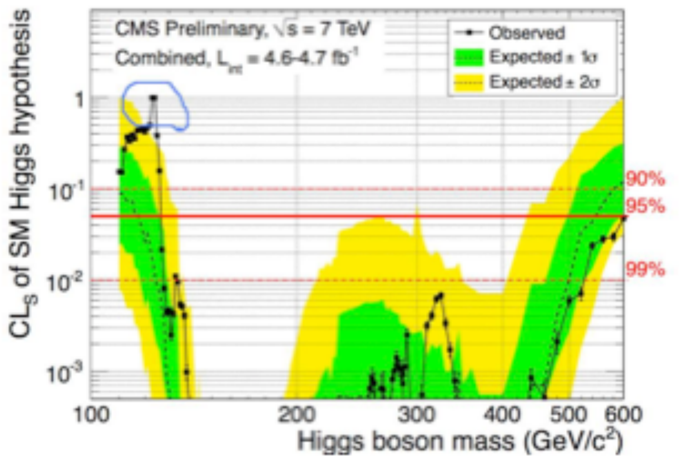
(1) Discuss deficits of background with p_0, not p_μ

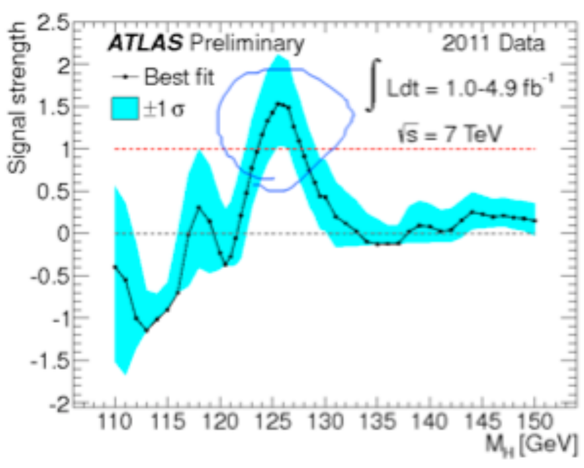
(2) Discuss excess of signal with

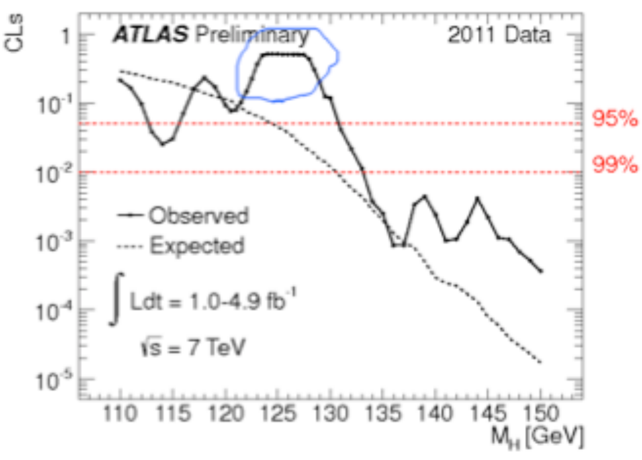
$$CL_S \simeq p_\mu (when\ p_0 \simeq 1 - p_b \rightarrow \sim 0)$$

All but $\hat{\mu}$ capped

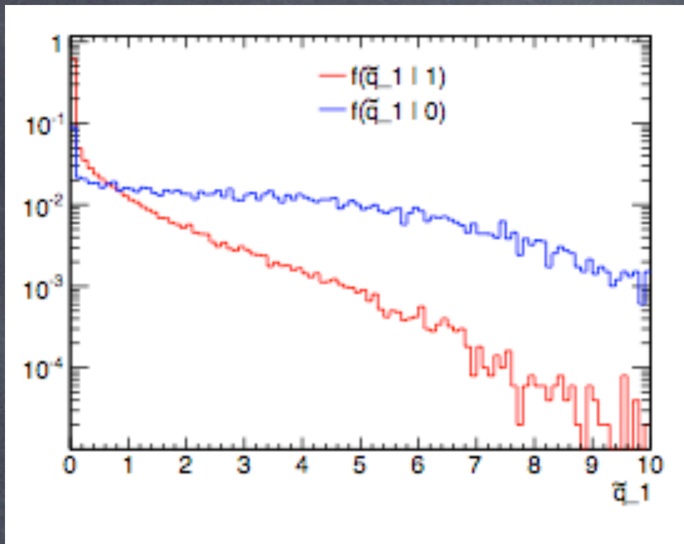


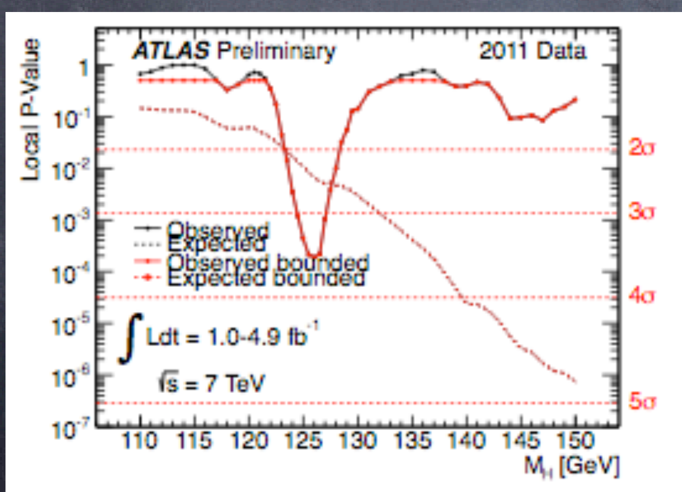


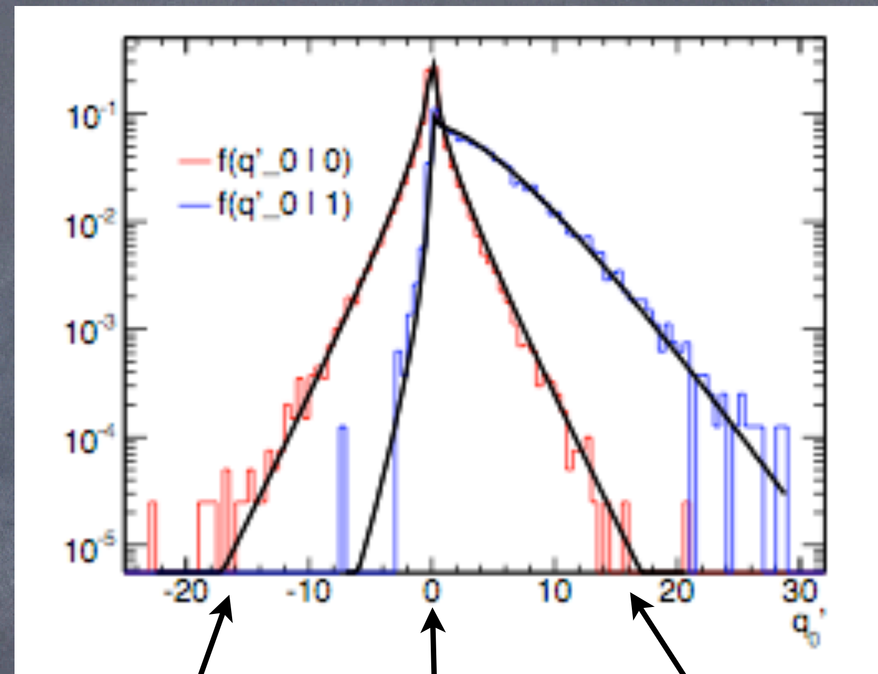




Uncapping p_0, i.e. $1-p_b$

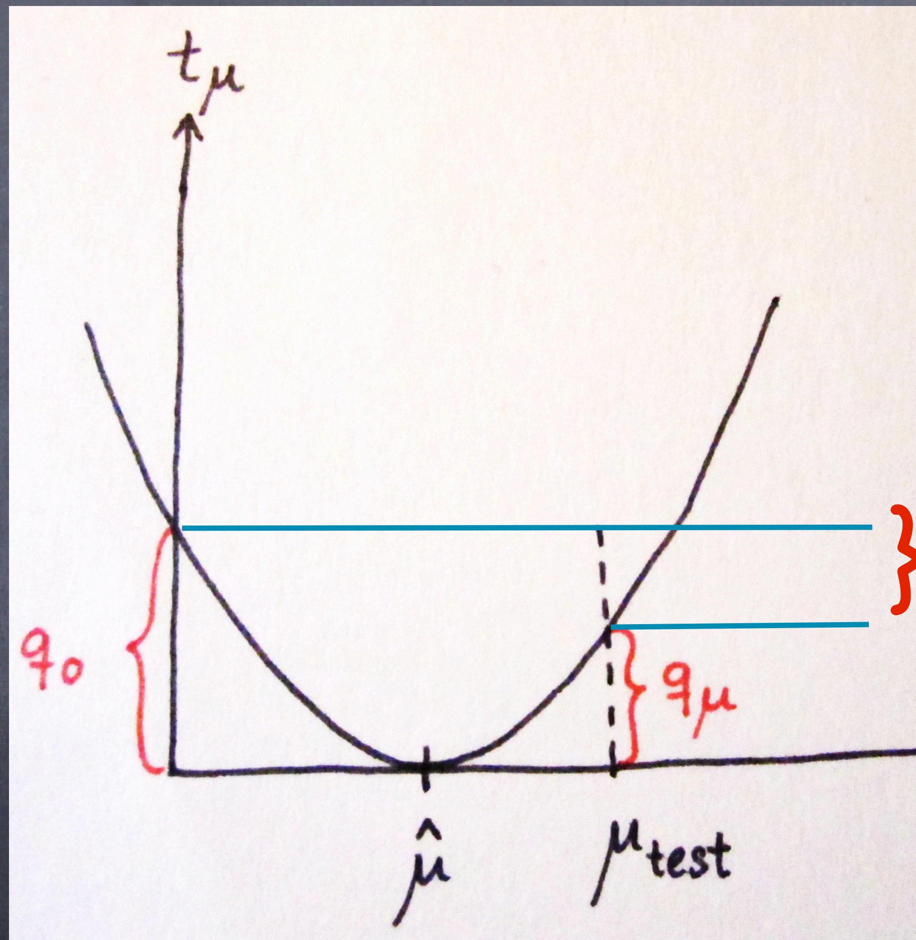






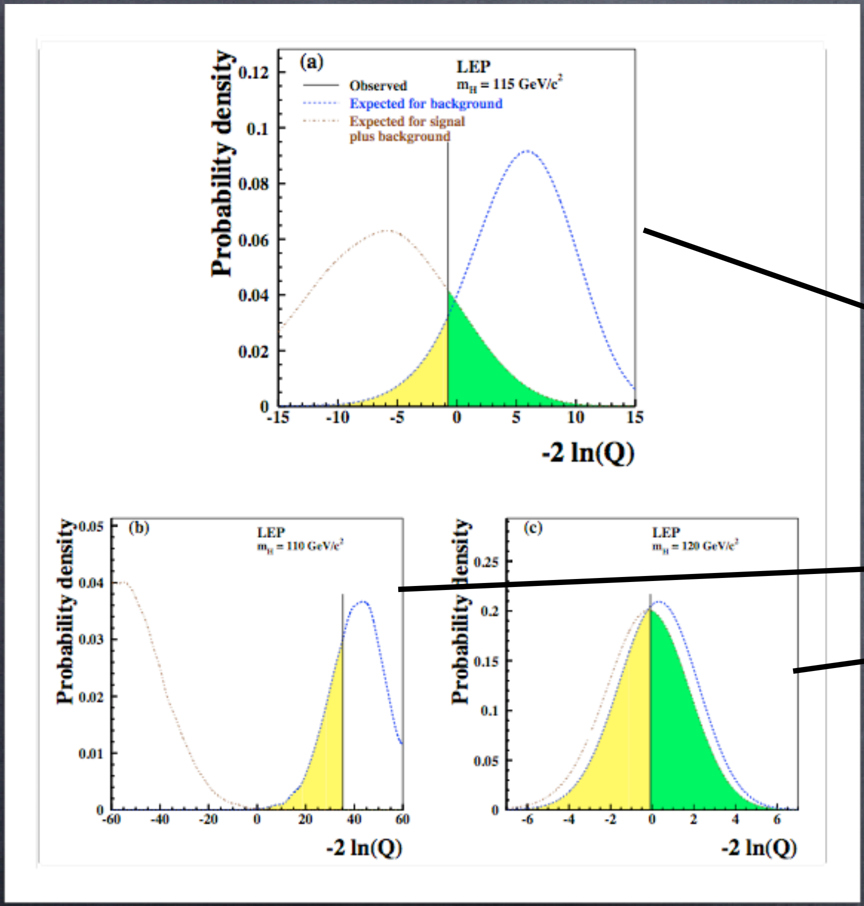
Large, median (0.5), small p^0

Q_{LEP} (Q_{TeV} w/o nuisances)

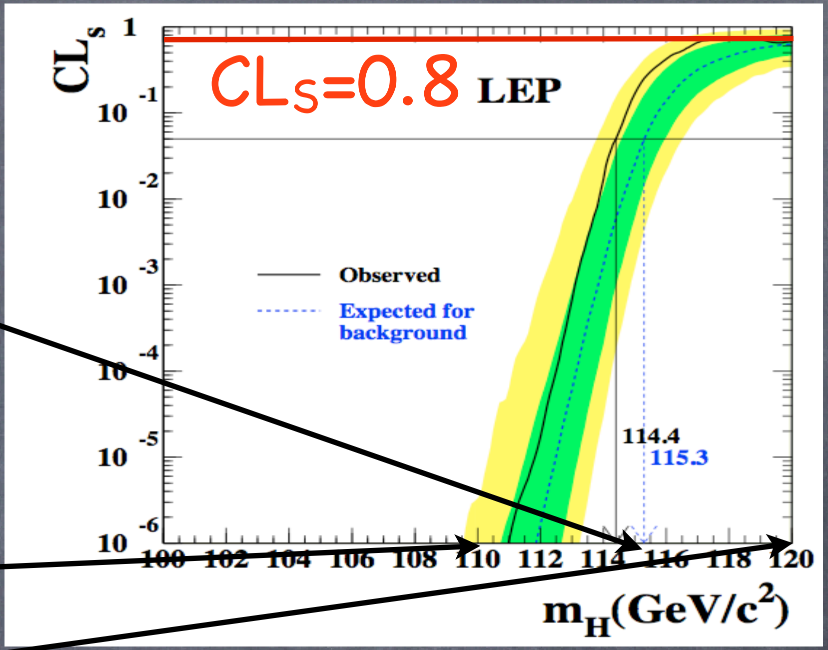


$$q_{LEP/TeV} = q_\mu - q_0$$

LEP CLs

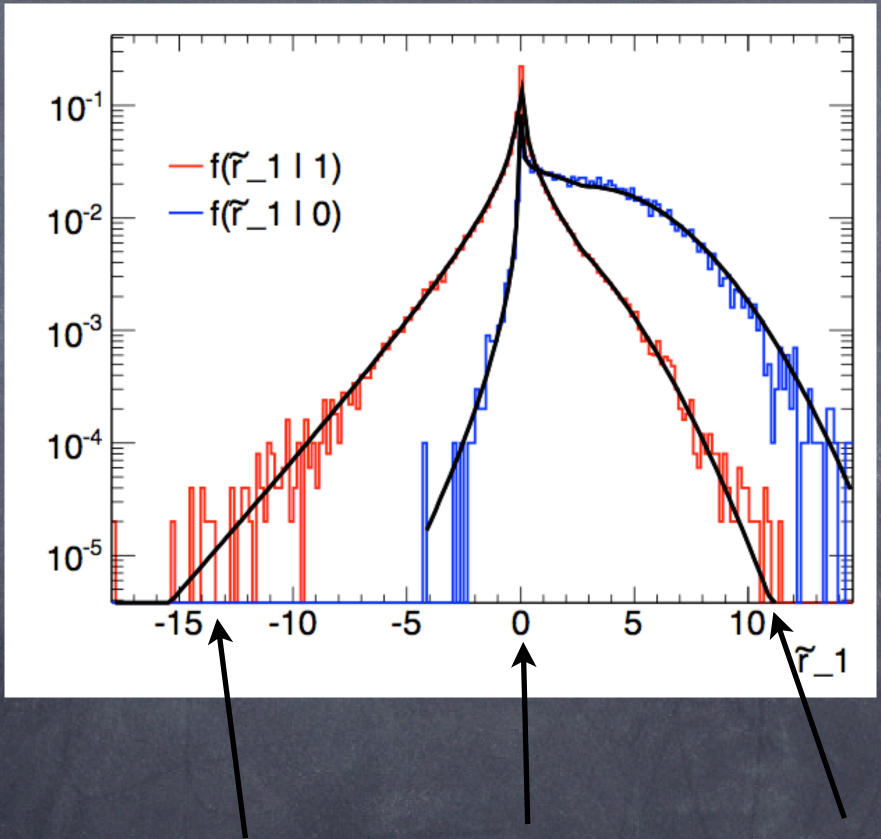


(a)
LEP
m_H = 115 GeV/c²
Observed
Expected for background
Expected for signal
plus background
Probability density
-2 ln(Q)
(b)
LEP
m_H = 110 GeV/c²
Probability density
-2 ln(Q)
(c)
LEP
m_H = 120 GeV/c²
Probability density
-2 ln(Q)



CLs=0.8
LEP
CLs
Observed
Expected for
background
114.4
115.3
m_H(GeV/c²)

Uncapping p_μ (almost CL_s)



Large, median (0.5), small p_μ

Energy scale systematics at high significance

Energy scale systematic uncertainties

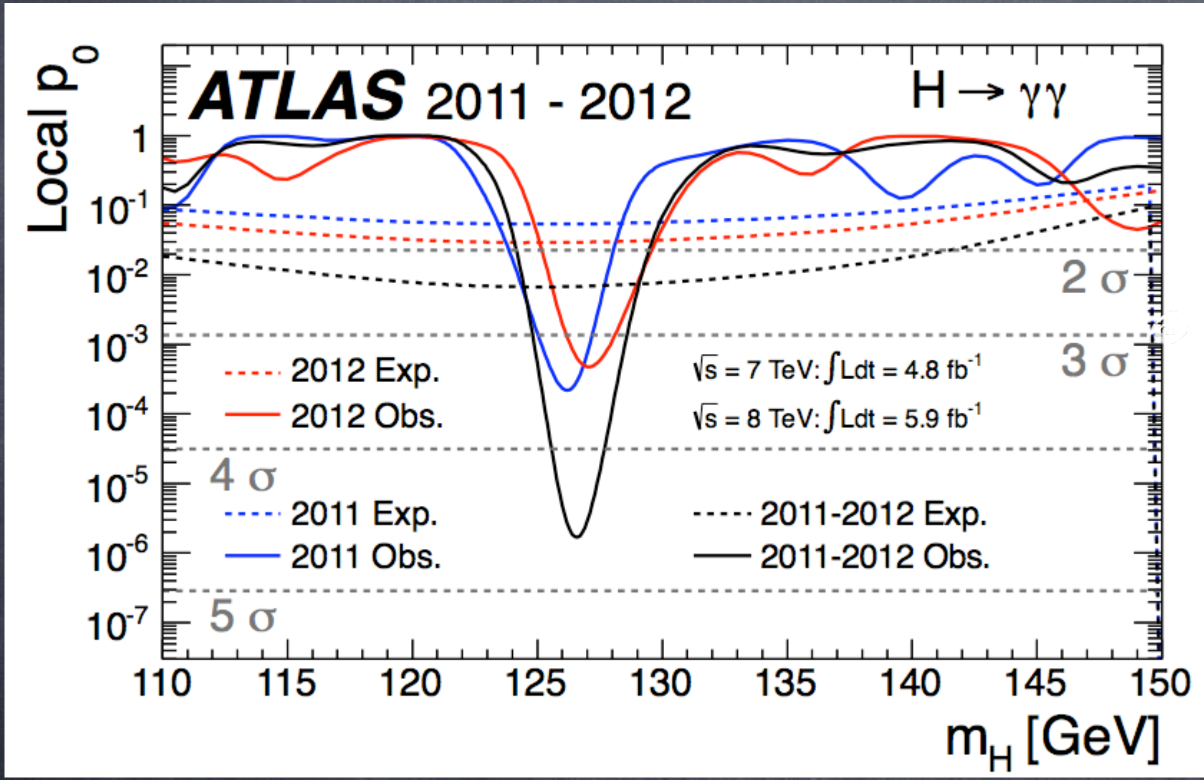
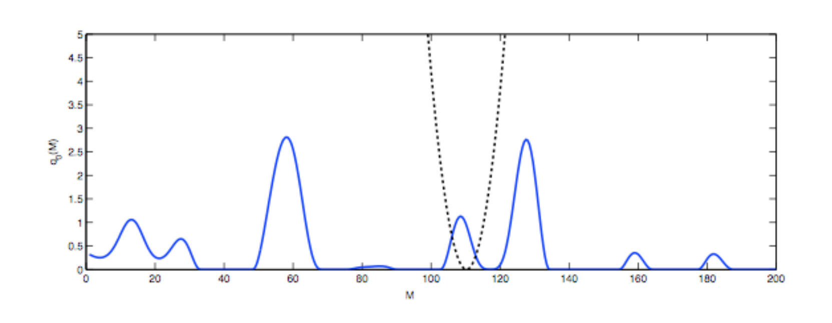


Illustration: Imagine we had aligned the red and blue before combining...

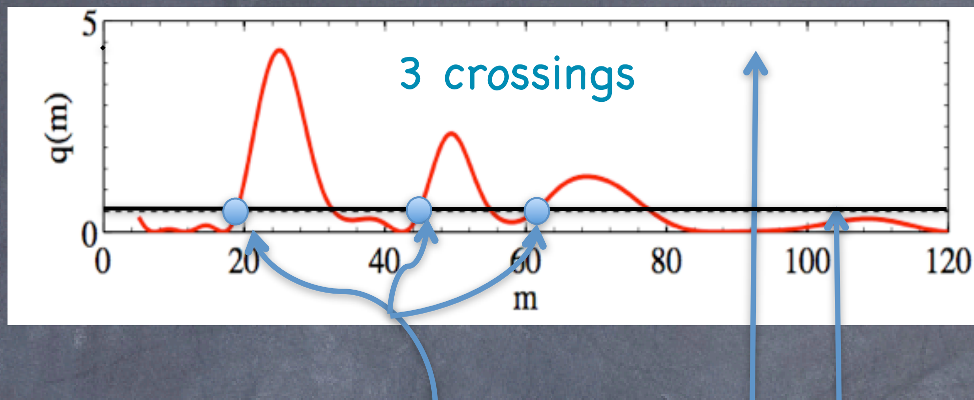
1 uncertain mass scale



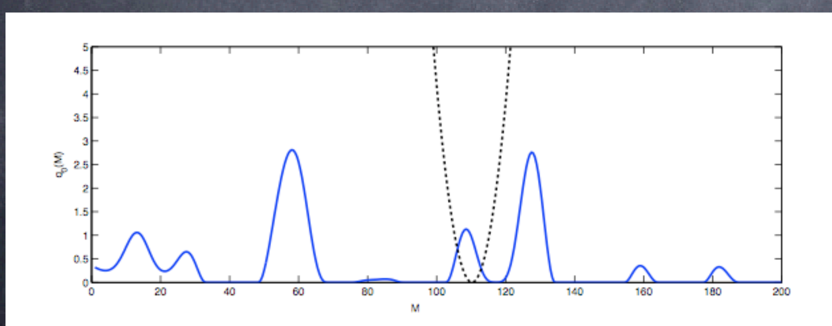
$$\mathbb{E}[N'_u] \leq \frac{1}{2}\mathbb{P}(\chi^2 > u) + \mathcal{N}_1 e^{-u/2} \frac{\sqrt{2\pi}\sigma_M}{|\Delta|}$$

Leadbetter (1965), Vitells (2012)

1 uncertain mass scale



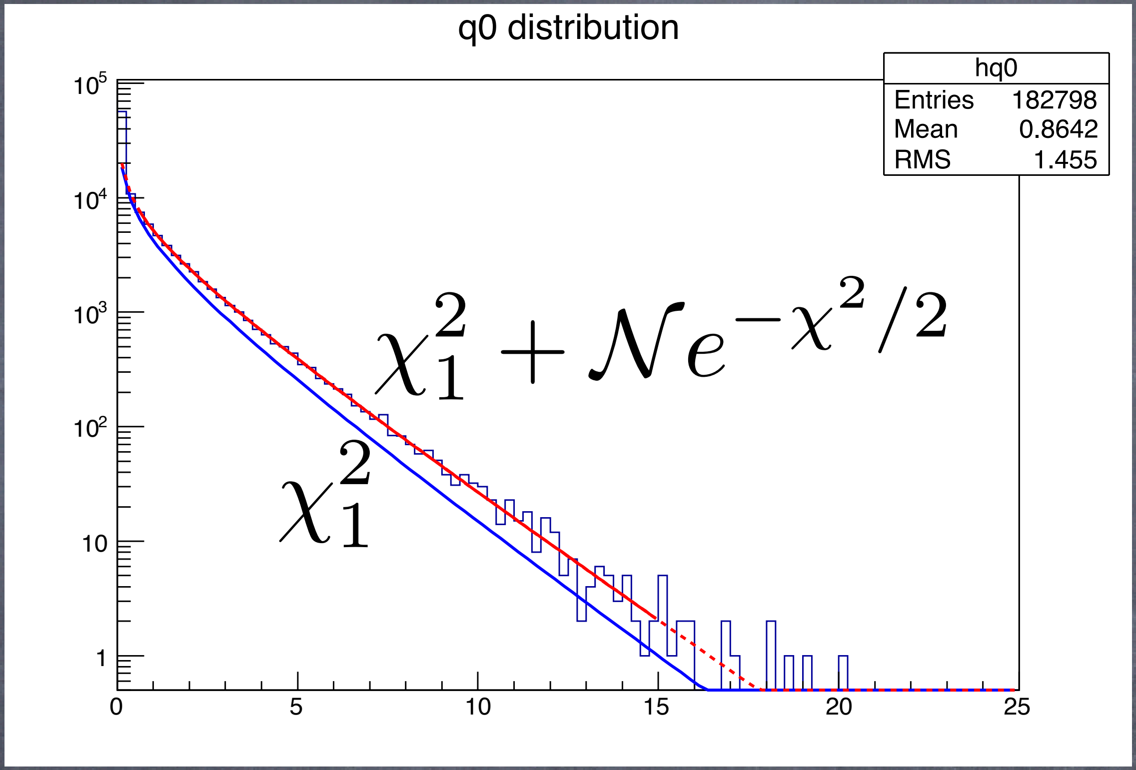
$$p_0^{\text{global}} \simeq p_0^{\text{local}} + < N(q_{\text{ref}}) > e^{-(q-q_{\text{ref}})/2}$$



$$\mathbb{E}[N'_u] \leq \frac{1}{2}\mathbb{P}(\chi^2 > u) + \mathcal{N}_1 e^{-u/2} \frac{\sqrt{2\pi}\sigma_M}{|\Delta|}$$

Leadbetter (1965), Vitells (2012)

Some random result



- Don't need $O(10^8)$ fits to MC toys to estimate tiny effect!

Importance Sampling

Sven Kreiss

Kyle Cranmer, Alex Read

Based on ideas from Alex Read and Michael Woodroofe

- ➡ Michael Woodroofe's talk at Banff 2010: http://people.stat.sfu.ca/~lockhart/richard/banff2010/woodroofe.pdf where he covered importance sampling and a method to create a suitable importance density.

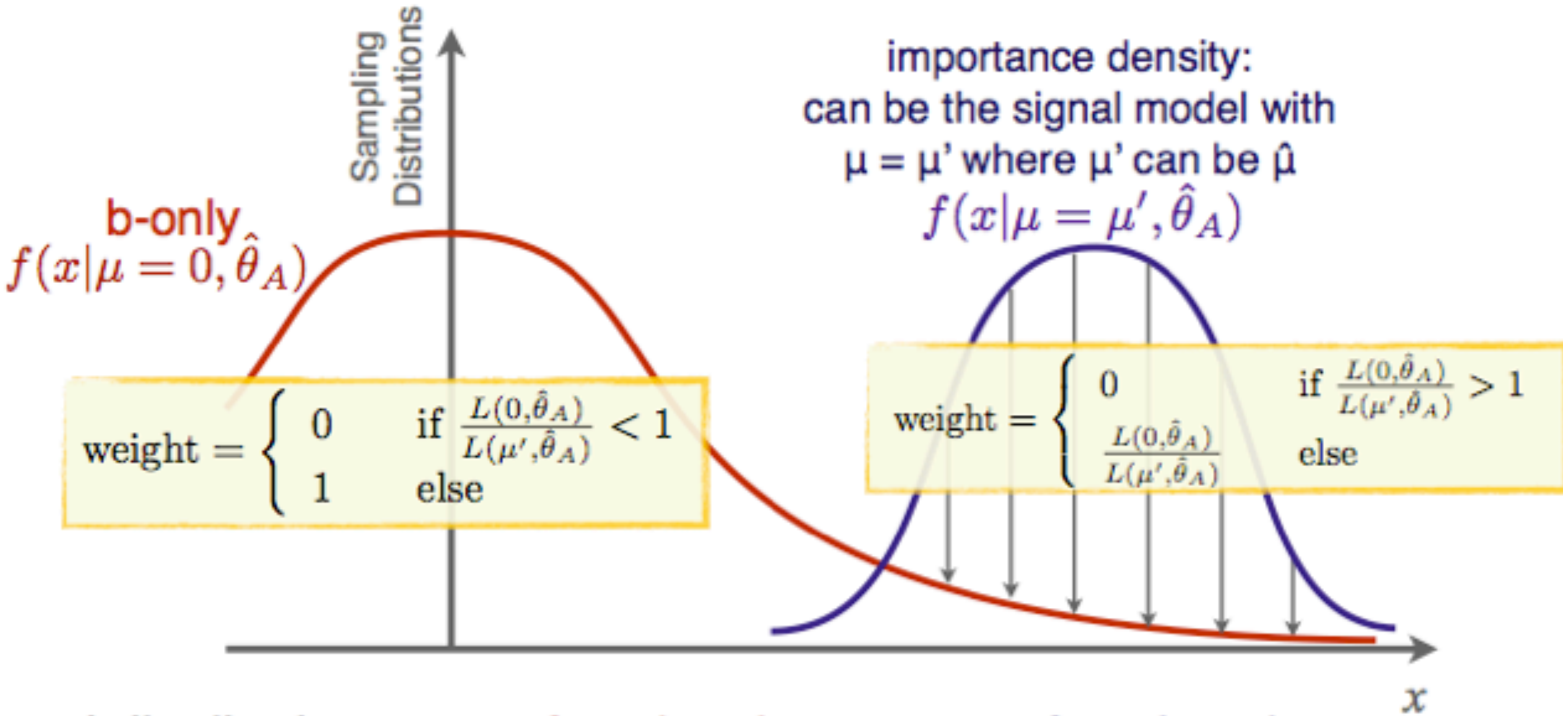
Technical: The current implementation uses a modified RooStats::ToyMCSampler and a custom script.

Idea: use a different density (importance density) to generate toys, but re-weight the result according to the ratio of their Likelihoods.

- ➡ Can populate a small tail with few toys.

P.S. Implemented in alrmc program, used by DELPHI and LEP HWG

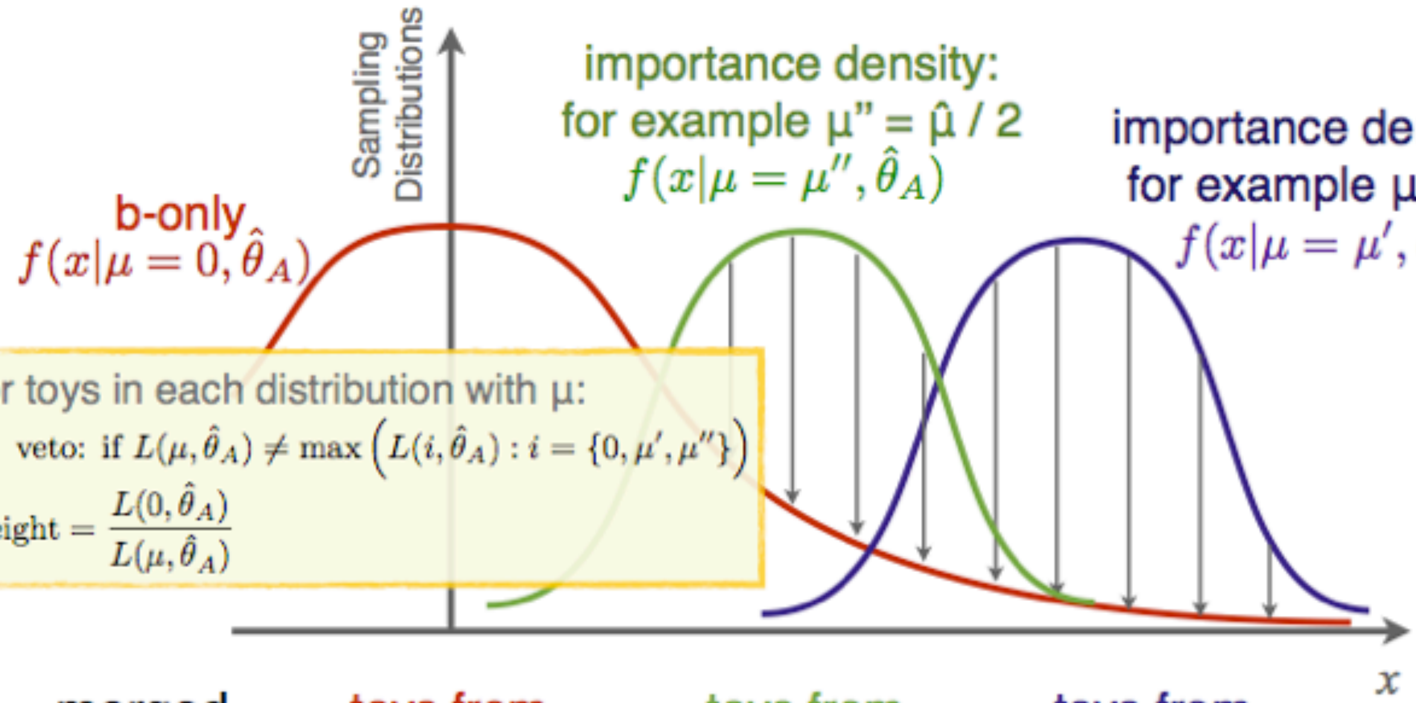
Importance Sampling with One Importance Density



merged distribution = toys from b-only + toys from imp dens

A point is only used in the merged sampling distribution when its Likelihood is the largest of the tested densities.

Importance Sampling with Multiple Importance Densities



For toys in each distribution with μ:

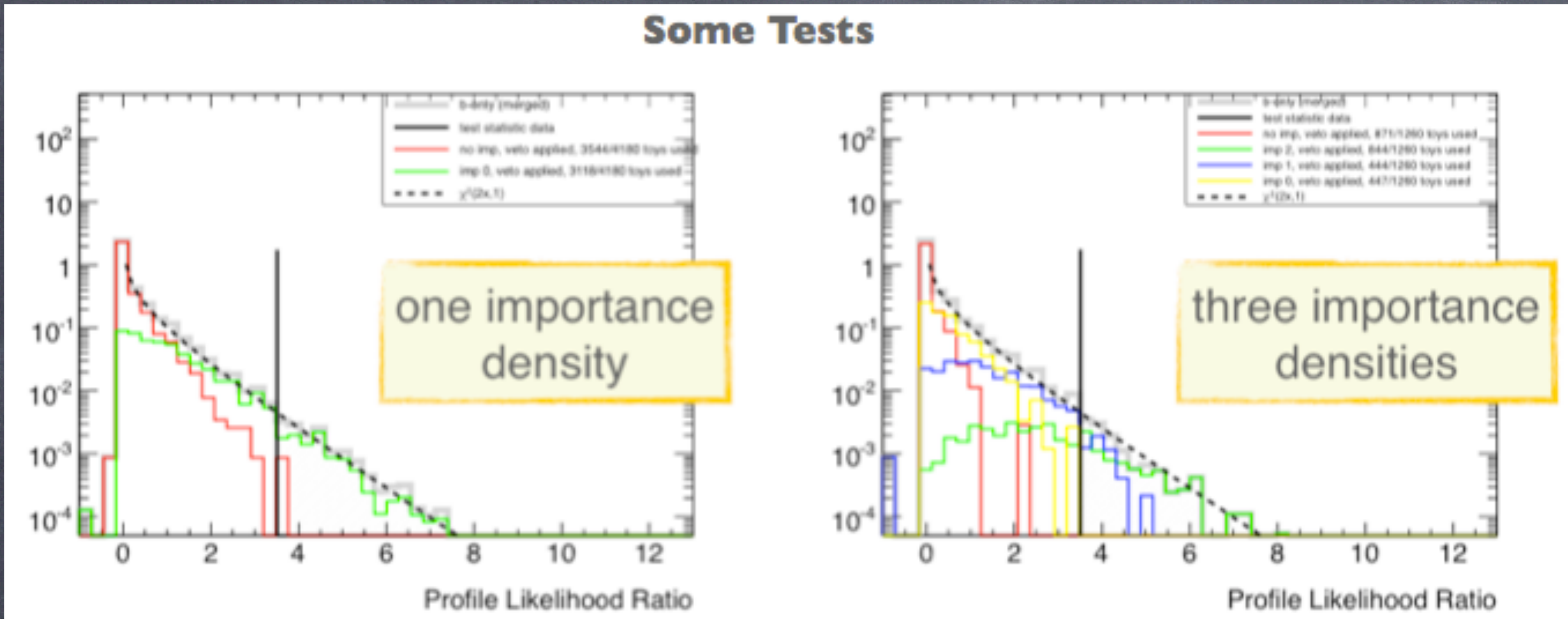
$$\text{veto: if } L(\mu, \hat{\theta}_A) \neq \max\left(L(i, \hat{\theta}_A) : i = \{0, \mu', \mu''\}\right)$$

$$\text{weight} = \frac{L(0, \hat{\theta}_A)}{L(\mu, \hat{\theta}_A)}$$

merged distribution = toys from b-only + toys from imp dens 1 + toys from imp dens 2

Adaptive number of importance densities:

➡ get error at μ̂ and use this to estimate the number of necessary importance densities to have the tails overlap at some target maximum number of standard deviations



Work in progress...
But much less urgent than we feared at some point!

Summary

- H->$\gamma\ \gamma$ background modeled, residual spurious signal accounted for
- Study of 180 Mtoys entirely consistent with GV-LEE paper
- Uncapping reveals information inside the other ~half of p-value results
- We have a "carbon-light" method to deal with ESS at high significance
- Importance sampling promising but needs careful validation, perhaps some optimization