Probability and Statistics

for experimental physicists : part III

José Ocariz
Université Paris Diderot and IN2P3

Outline

Lecture 1 (Monday)

Basic concepts in Probability and Statistics

Lecture 2 (Tuesday)

Maximum Likelihood theorem
Multivariate techniques

Lecture 3 (today)

An analysis example from *BaBar*
Hypothesis testing, limit settings

Disclaimer

Most, if not all of you, are already familiar with many of these topics… for consistency, the scope spans from the very general concepts towards more advanced developments…

A Complete Analysis Example from BaBar

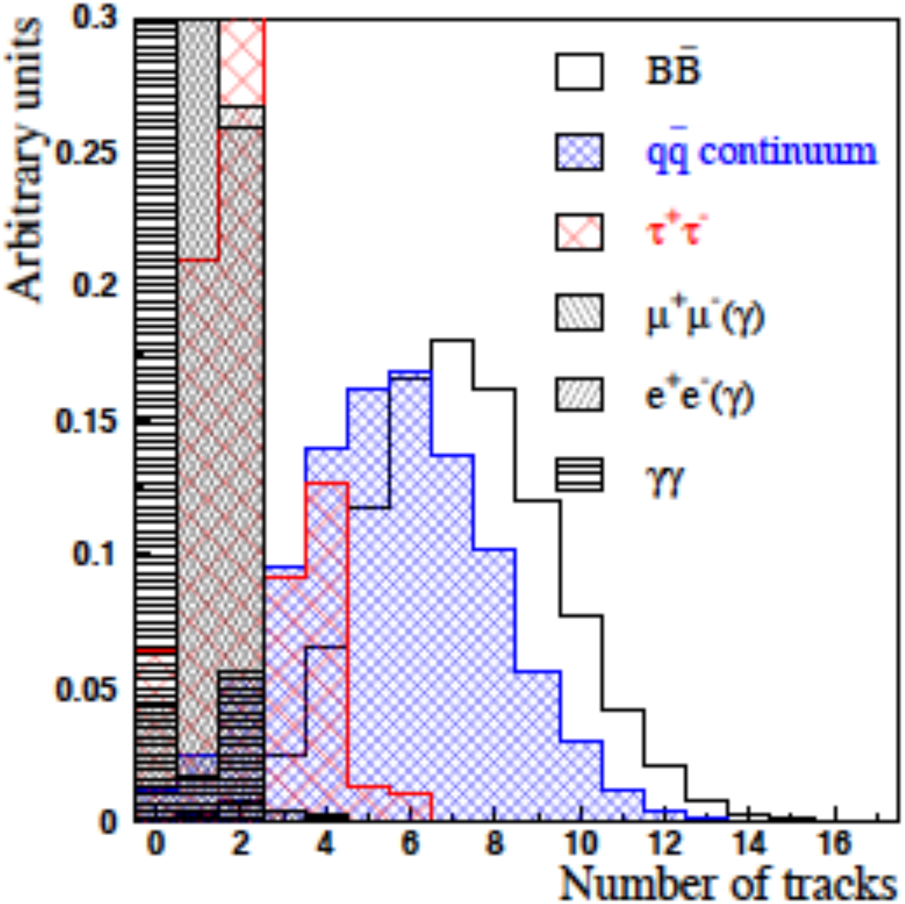
Amplitude analysis of neutral B mesons decaying into 3 particles : $B^0 / \overline{B^0} \rightarrow K_s^0 \pi^+ \pi^-$

- **requires large levels of background suppression**
- **requires B-meson "flavor" tagging**
- **requires ML discrimination of signal against irreducible background**
- **requires composition of signal amplitude and interference patters**

A Complete Analysis Example from BaBar

Amplitude analysis of neutral B mesons decaying into 3 particles : $B^0 / \overline{B^0} \rightarrow K_s^0 \pi^+ \pi^-$

- **<u>requires large levels of background suppression</u>**
- **requires B-meson "flavor" tagging**
- **requires ML discrimination of signal against irreducible background**
- **requires composition of signal amplitude and interference patters**



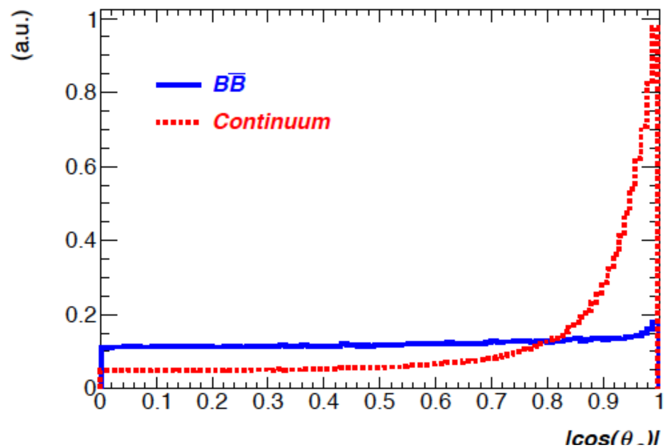
Cut-based selection, 1st stage :

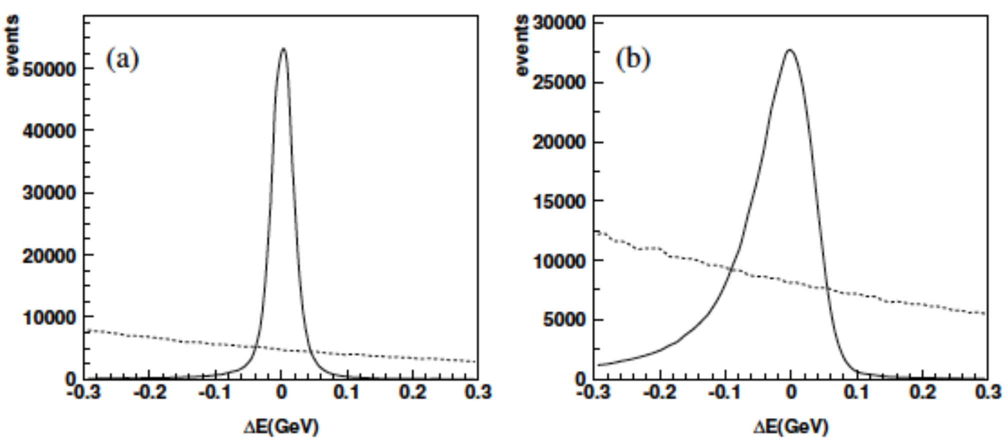
filter on simple, robust variables

A Complete Analysis Example from BaBar

Amplitude analysis of neutral B mesons decaying into 3 particles : $B^0 / \overline{B^0} \to K_s^0 \pi^+ \pi^-$

- **<u>requires large levels of background suppression</u>**
- **requires B-meson "flavor" tagging**
- **requires ML discrimination of signal against irreducible background**
- **requires composition of signal amplitude and interference patters**





Cut-based selection, 2nd stage :

kinematical variables

"event-shape" variables

Define physics sample

Amplitude analysis of neutral B mesons decaying into 3 particles : $B^0 / \overline{B^0} \rightarrow K_s^0 \pi^+ \pi^-$

- requires large levels of background suppression
- <u>requires B-meson "flavor" tagging</u>
- requires ML discrimination of signal against irreducible background
- requires composition of signal amplitude and interference patters

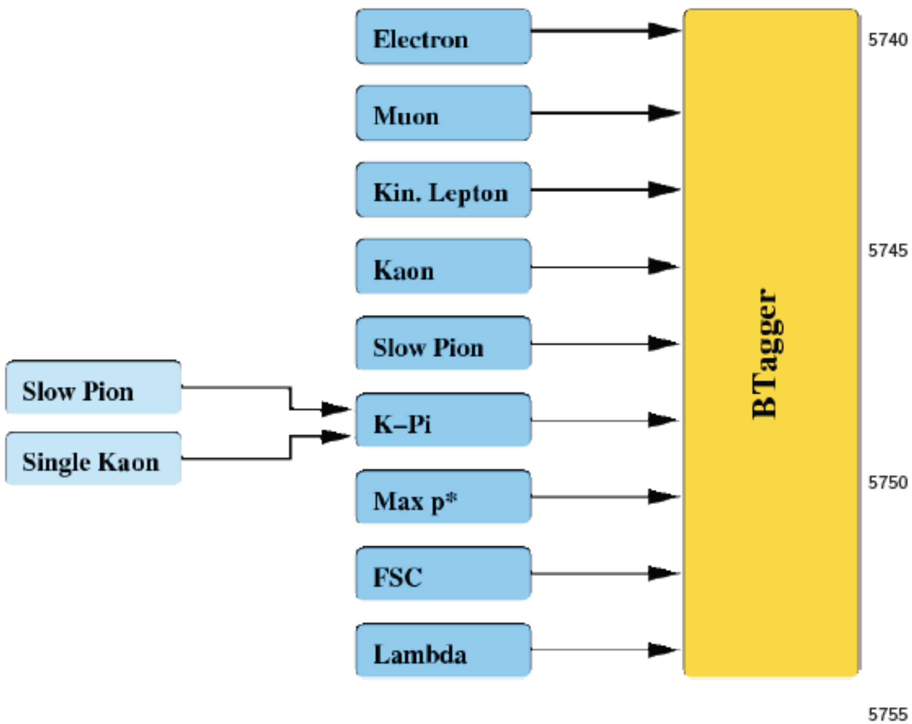


Fig. 8.5.1. Schematic overview of the BABAR tagging algorithm. Each box corresponds to a separate neural network.

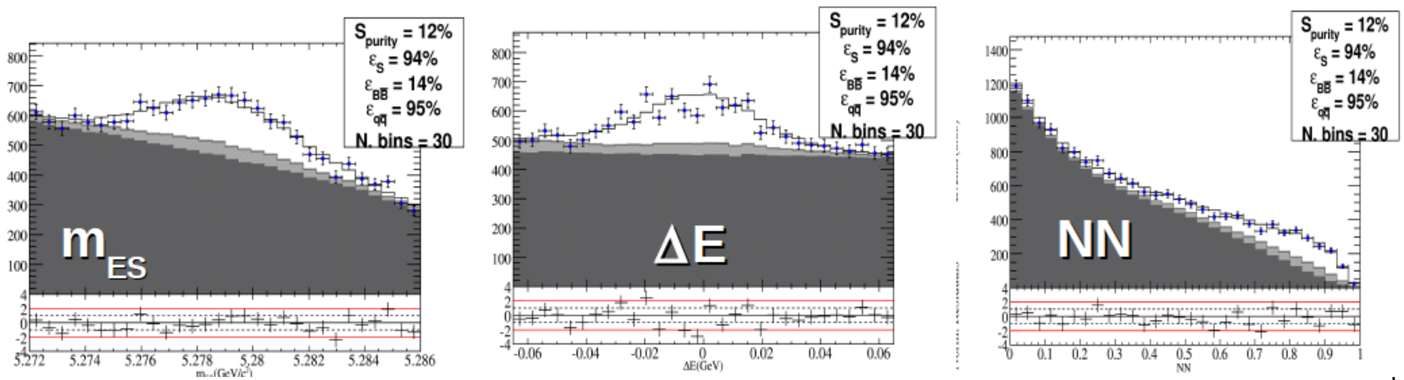
B-flavor tagging :

exploit available information

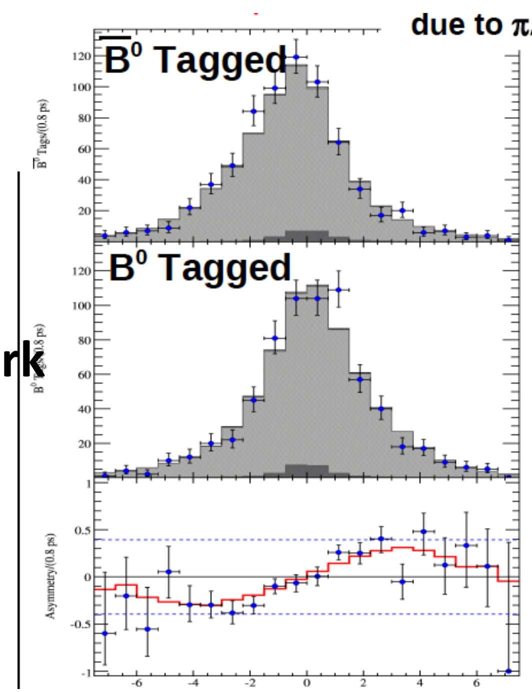
combine into a neural network

Amplitude analysis of neutral B mesons decaying into 3 particles : $B^0 / \overline{B^0} \to K_s^0 \pi^+ \pi^-$

- **requires large levels of background suppression**
- **requires B-meson "flavor" tagging**
- **requires ML discrimination of signal against irreducible background**
- **requires composition of signal amplitude and interference patters**



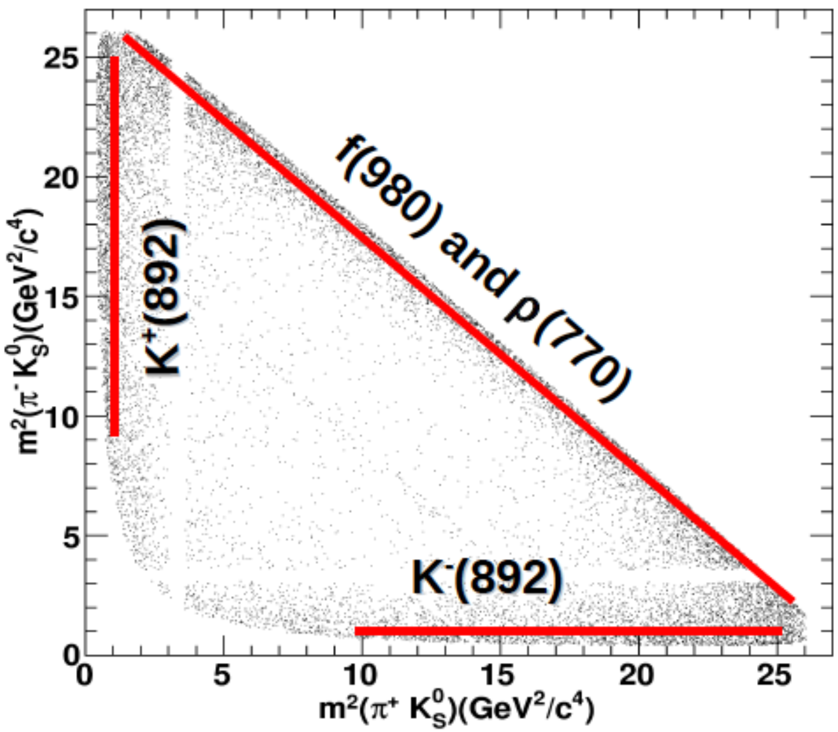
Information implemented in the Likelihood :
kinematical variables
various event-shape variables, combined into a neural network
decay time information
…
signal composition in terms of amplitudes and phases
…
and perform a blind analysis !

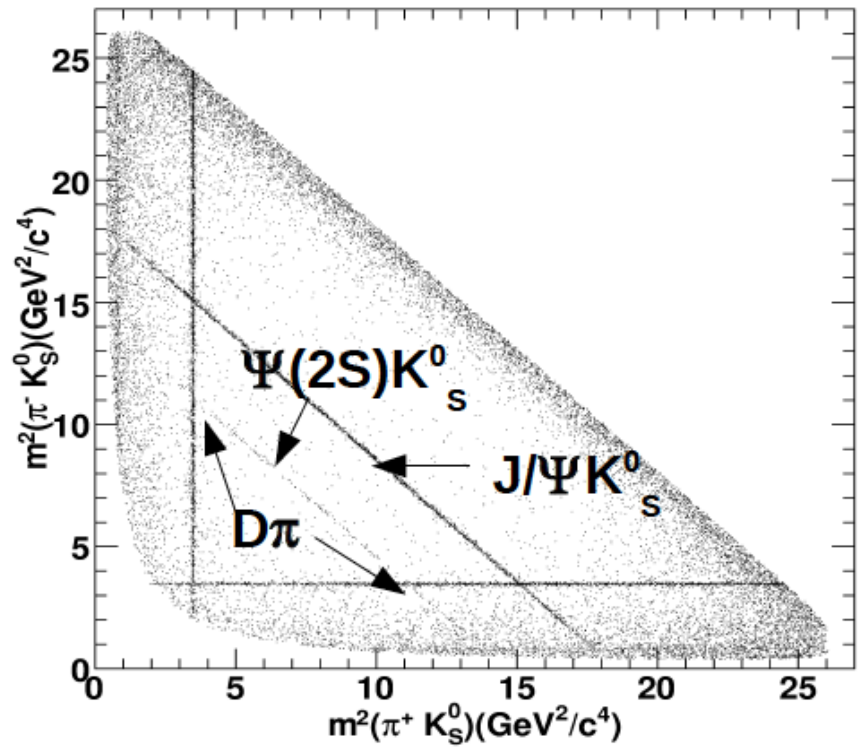


A Complete Analysys Example from BaBar

Amplitude analysis of neutral B mesons decaying into 3 particles : $B^0 / \overline{B^0} \rightarrow K_s^0 \pi^+ \pi^-$

- **requires large levels of background suppression**
- **requires B-meson “flavor” tagging**
- **requires ML discrimination of signal against irreducible background**
- **<u>requires composition of signal amplitude and interference patters</u>**

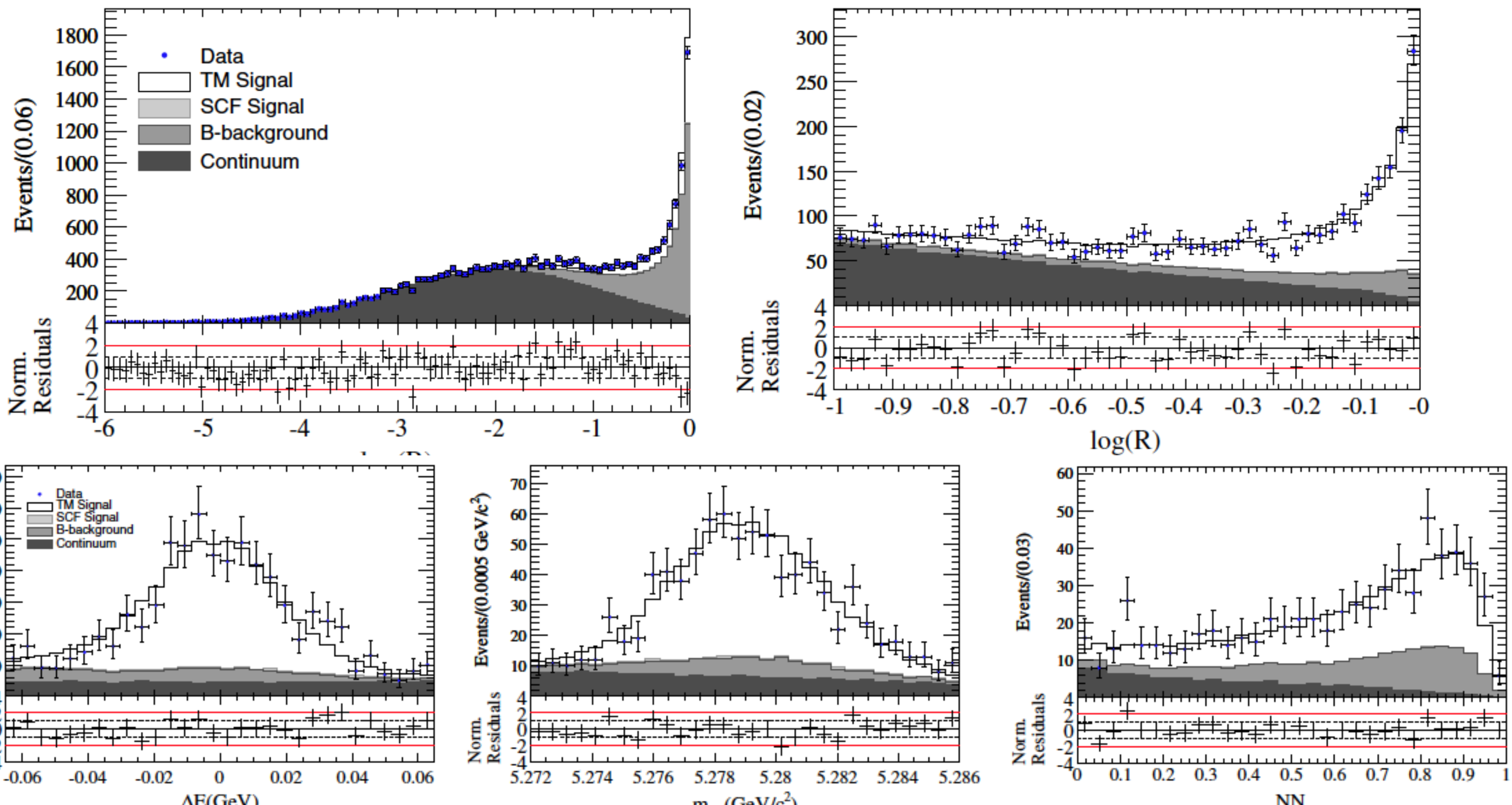




A Complete Analysys Example from BaBar

Amplitude analysis of neutral B mesons decaying into 3 particles : $B^0 / \overline{B^0} \rightarrow K_s^0 \pi^+ \pi^-$

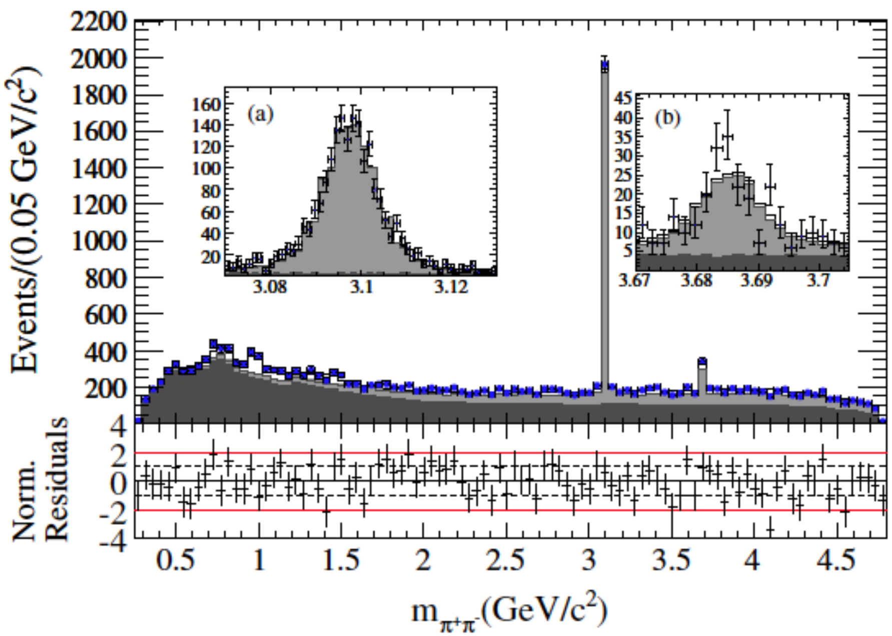
- **Blind analysis : converge on all analysis criteria using control samples**
- **Goodness-of-fit studies : Likelihood ratio projections**

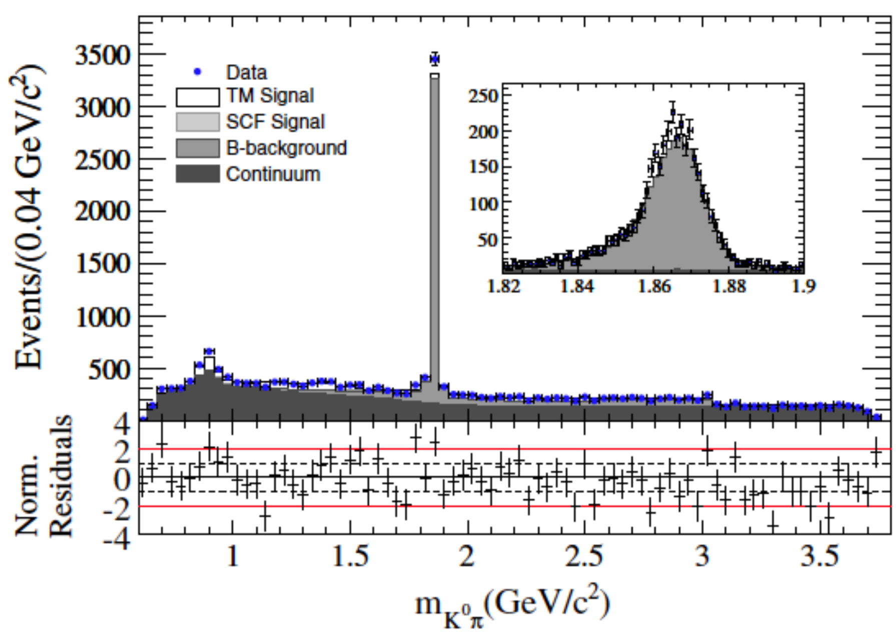


A Complete Analysys Example from BaBar

Amplitude analysis of neutral B mesons decaying into 3 particles : $B^0 / \overline{B^0} \rightarrow K_s^0 \pi^+ \pi^-$

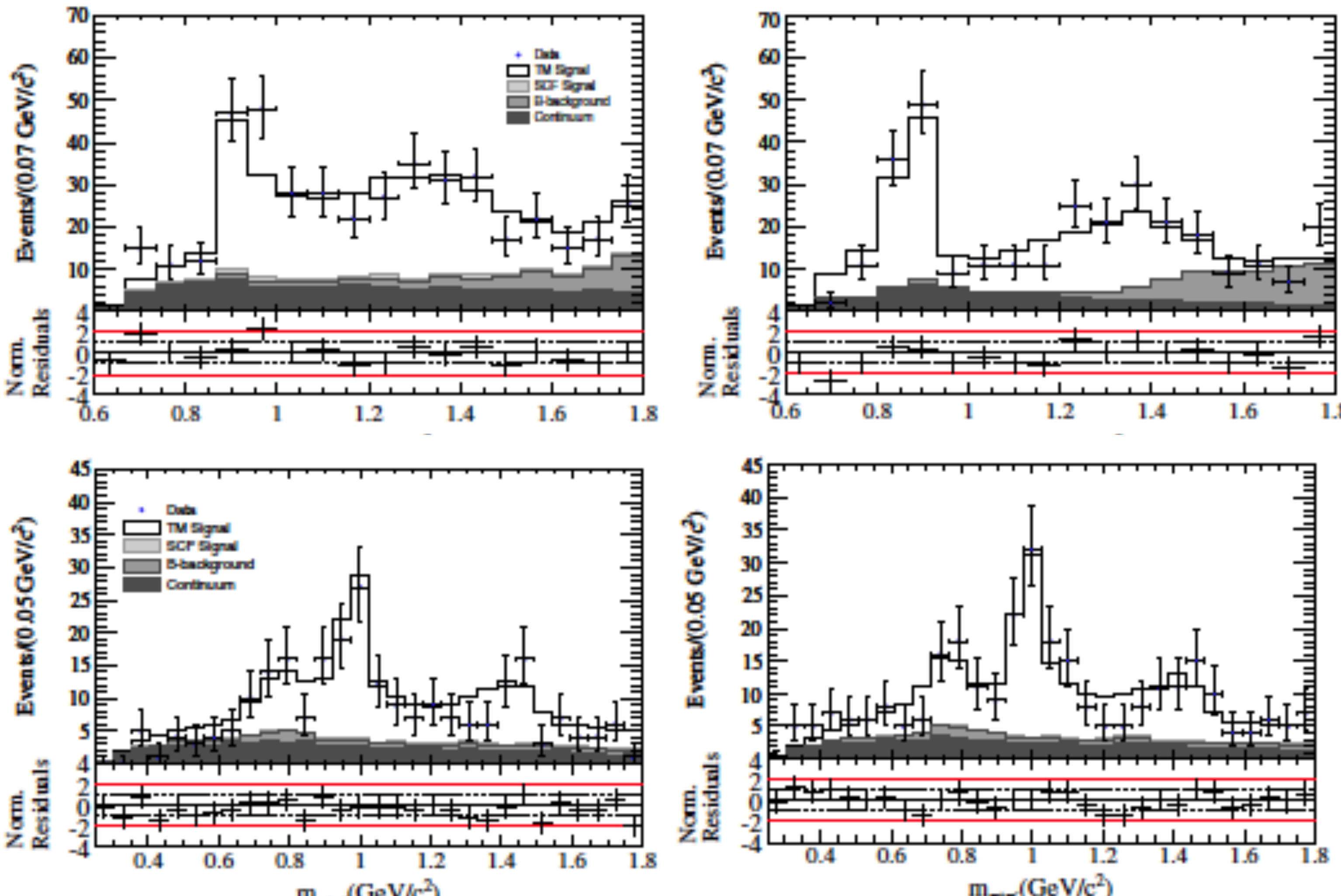
- **Goodness-of-fit studies : Likelihood ratio projections : zoom on control samples**





Amplitude analysis of neutral B mesons decaying into 3 particles : $B^0 / \overline{B^0} \rightarrow K_s^0 \pi^+ \pi^-$

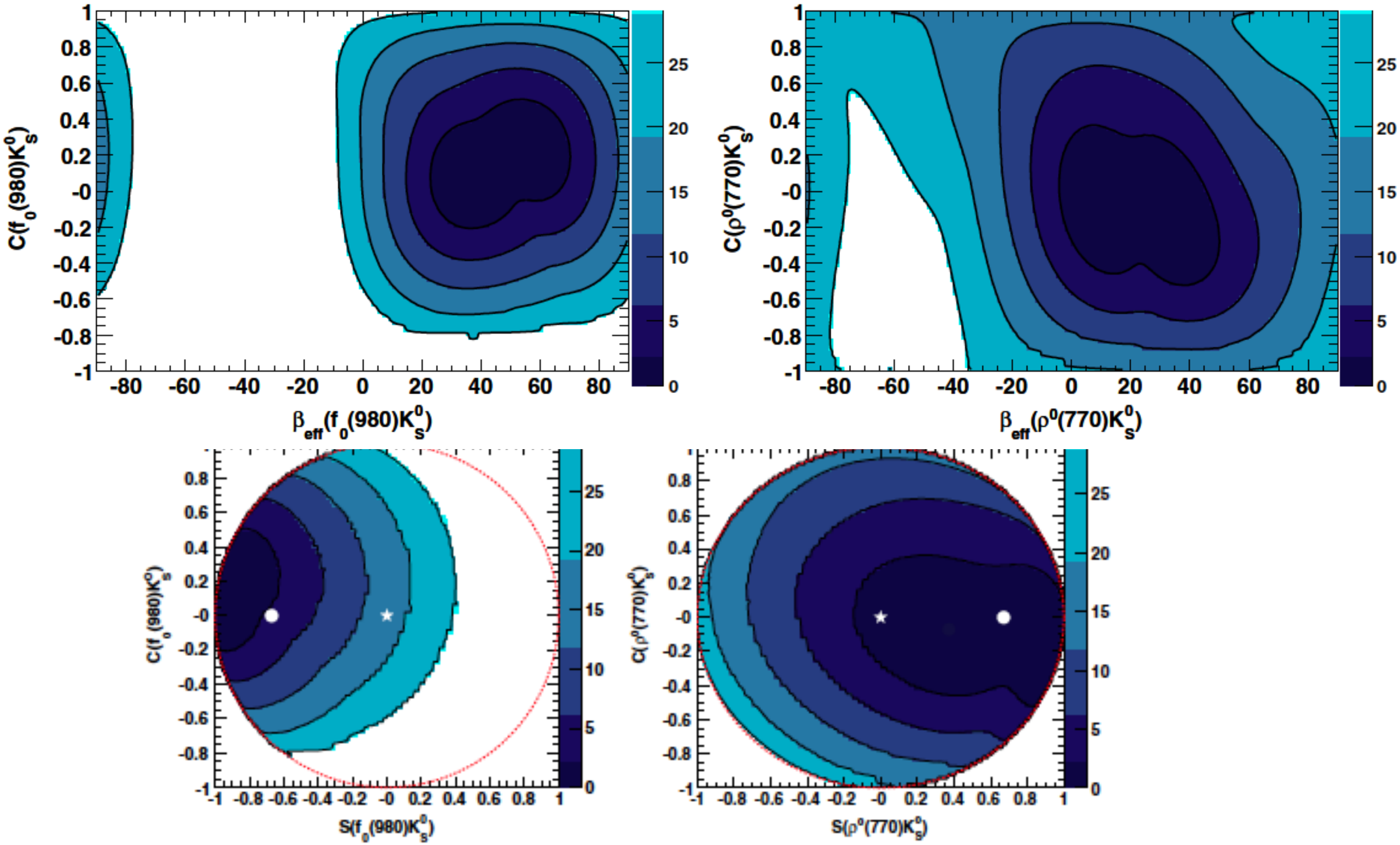
- **Goodness-of-fit studies : Likelihood ratio projections on signal-enriched areas**



A Complete Analysys Example from BaBar

Amplitude analysis of neutral B mesons decaying into 3 particles : $B^0 / \overline{B^0} \rightarrow K_s^0 \pi^+ \pi^-$

- **Results : Likelihood contours on physical observables**



Hypothesis-Testing (I)

Hypothesis (in physicist-oriented wording) : a model attempting at describing data
Some standard definitions :

- H_0 **: the "null" hypothesis (e.g. background-only model)**
- H_1 **: the "alternative" hypothesis (e.g. signal-plus-background model)**
- **Hypothesis-testing :**
  - **build a statistics** μ **, define a confidence interval** W**, measure** $\hat{\mu}$
  - **if inside (outside)** W**, accept (reject)** H_0
- **Type-I error : false positive, reject** H_0 **despite being true (inefficiency)**
  - α **is the rate of Type-I error**
  - α **is also called "size"**

$$1-\alpha = \int_W d\mu \mathcal{L}(\mu \mid H_0)$$

- **Type-II error : false negative, accept** H_0 **despite being false (fake-rate)**
  - β **is the rate of Type-II error**
  - $1\text{-}\beta$ **is also called "power"**

$$1-\beta = \int_W d\mu \mathcal{L}(\mu \mid H_1)$$

- **A good statistics : both** α **and** β **small (usually conflicting)**
- **Neyman-Pearson : at a fixed size, the *Likelihood Ratio*** λ **is optimal**

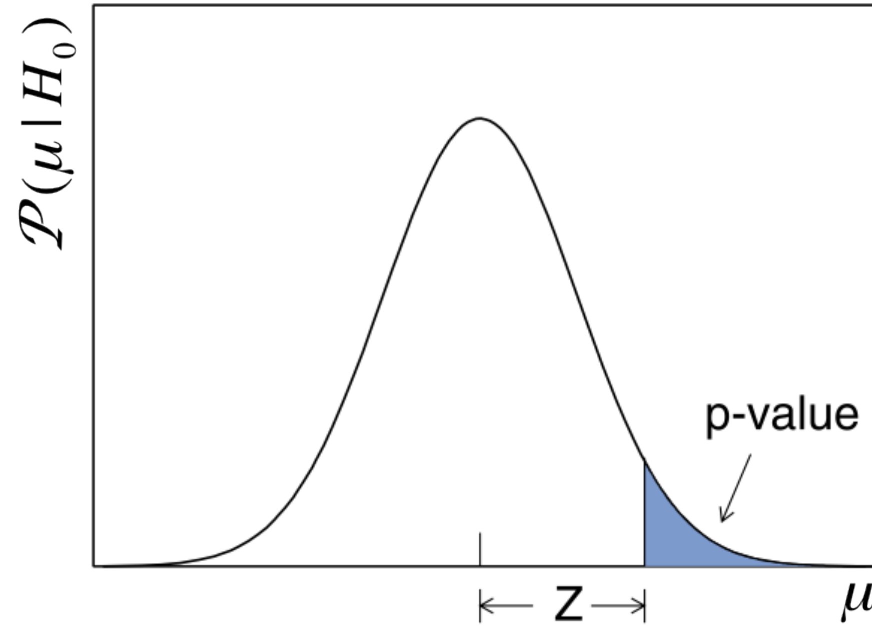
$$\lambda(\mu) = \frac{\mathcal{L}(\mu \mid H_0)}{\mathcal{L}(\mu \mid H_1)} > \kappa(\mu)$$

beware of underlying assumptions …

Hypothesis-Testing : p-values

Statistical significance :
probability within H0 of μ being larger than the observed $\hat{\mu}$ value

$$\text{p-value} = \int_{\hat{\mu}}^{\infty} d\mu \mathcal{P}(\mu \mid H_0)$$



Popular preference :
quote the p-value in terms of "sigmas",

$$\text{p-value} = \int_{n\sigma}^{\infty} dx \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = 1 - \frac{1}{2} erf\left(\frac{n}{\sqrt{2}}\right)$$

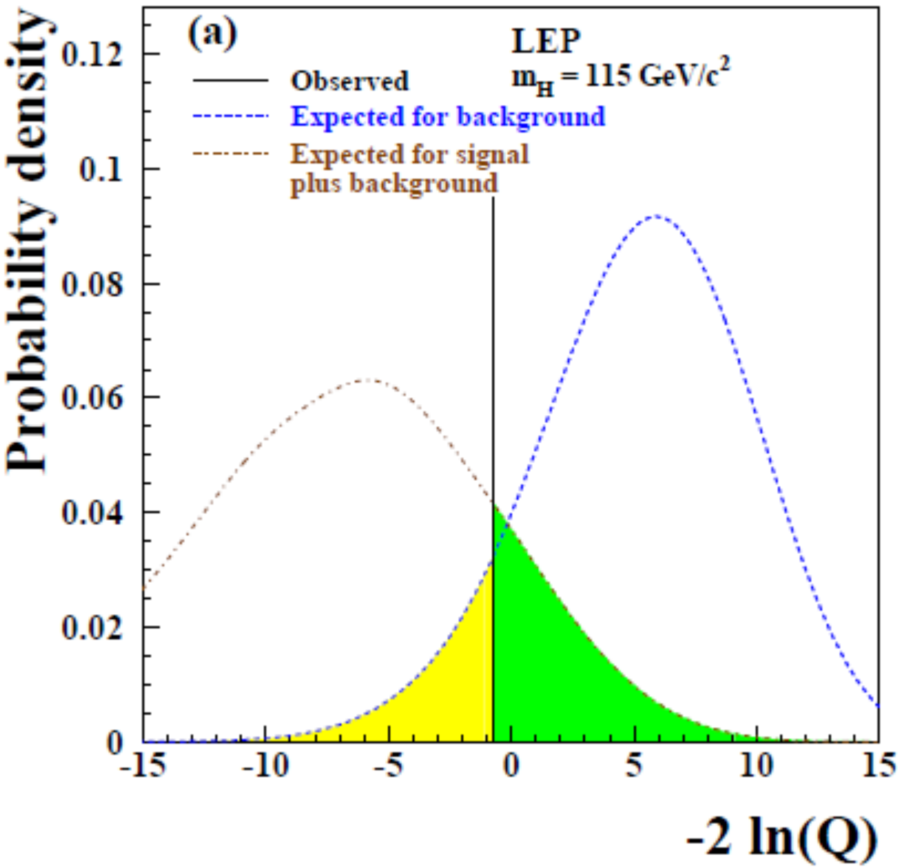
Usual benchmarks are asymmetric for exclusion and discovery :

- **require** $p<0.05$ **(i.e.** 1.64σ**, or 95% C.L.) applied to** $H_0=sig+bkg$
  - **excluding a signal hypothesis**
- **require** $p<1.35\times10^{-3}$ **(i.e.** 3σ**) applied to** $H_0=bkg\text{-}only$
  - **"evidence" benchmark**
- **require** $p<2.87\times10^{-7}$ **(i.e.** 5σ**) applied to** $H_0=bkg\text{-}only$
  - **"observation" benchmark**

Test Statistics : LEP as an Example

$$Q = \frac{\mathcal{L}(H_1)}{\mathcal{L}(H_0)}, \text{ with } \begin{cases} \mathcal{L}(H_1) = \prod_{i=1}^{N_{ch}} \mathcal{P}_{Poisson}(n_i, s_i + b_i) \prod_{j=1}^{n_i} \frac{s_j \mathcal{S}(\vec{x}_j) + b_j \mathcal{B}(\vec{x}_j)}{s_j + b_j} \\ \mathcal{L}(H_0) = \prod_{i=1}^{N_{ch}} \mathcal{P}_{Poisson}(n_i, b_i) \prod_{j=1}^{n_i} \mathcal{B}(\vec{x}_j) \end{cases}$$

$$\lambda = -2\ln(Q)$$

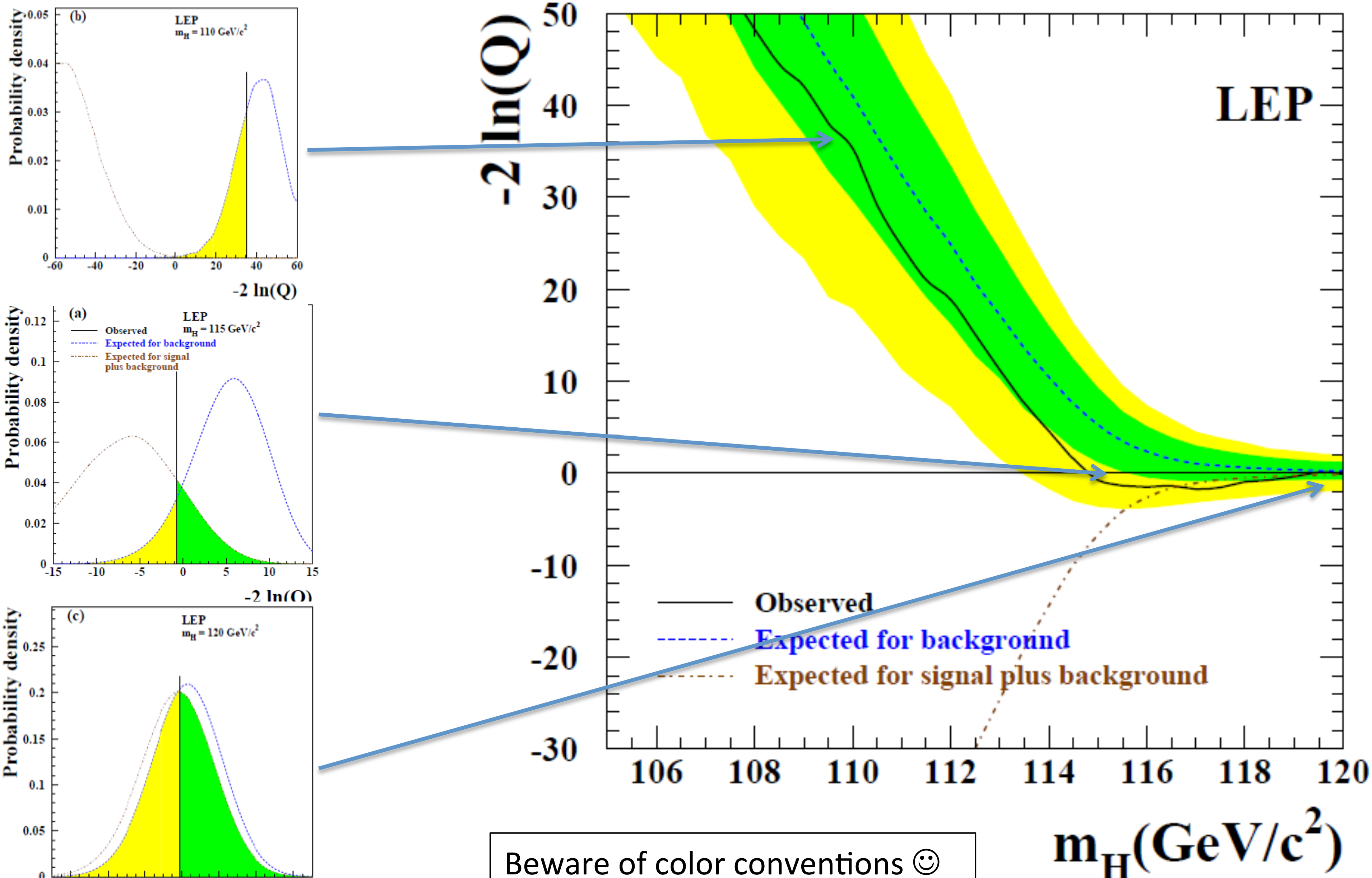


Standard definitions :

CL(s+b) **: p-value under** *H0 = sig+bkg*
probability of having sig+bkg yielding *-2lnQ*
larger than the observed one.
In the plot : *green = CL(s+b)*

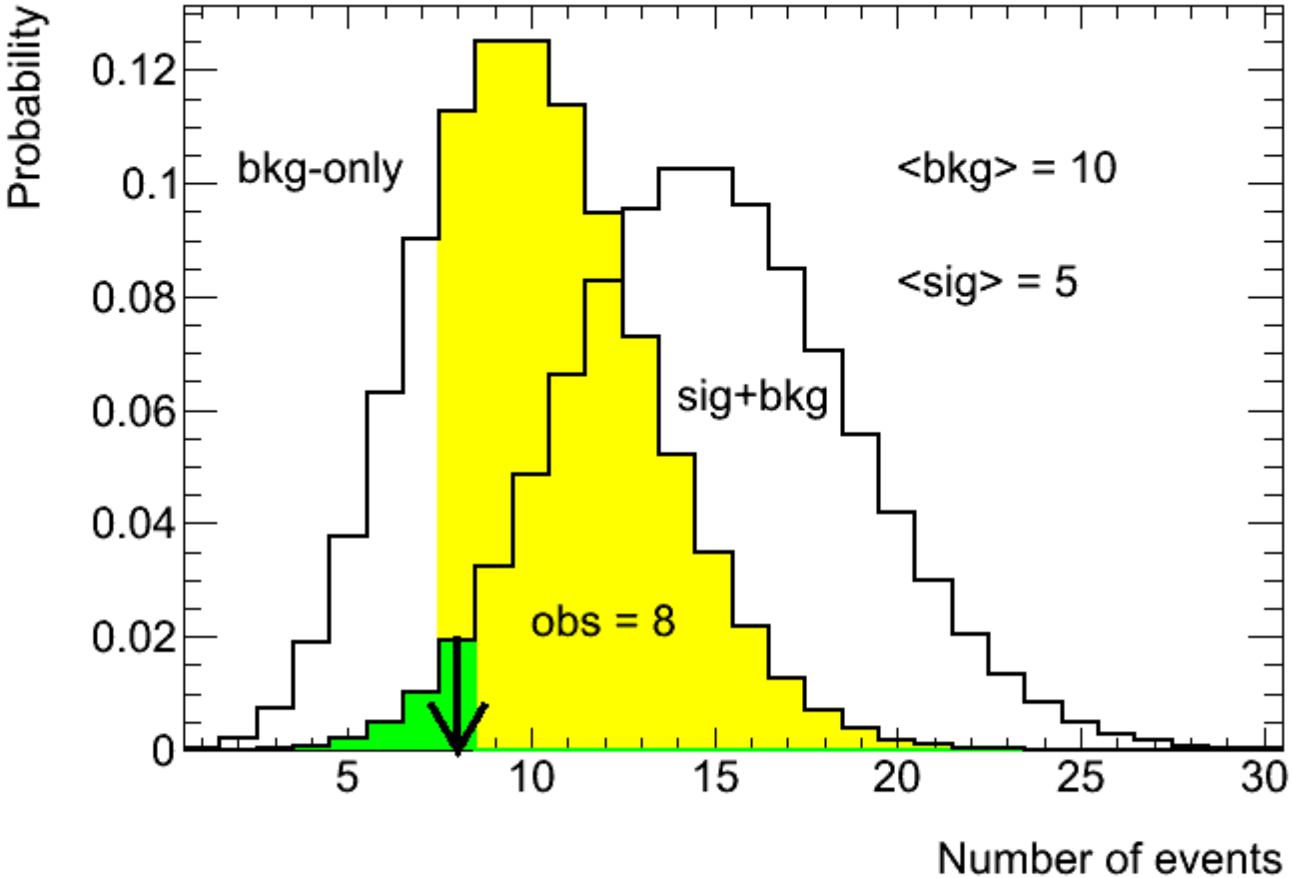
CL(b) **: p-value under** *H0 = bkg-only*
probability of having bkg yielding *-2lnQ*
larger than the observed one.
In the plot : *yellow = 1- CL(b)*

Test Statistics : LEP as an Example



Beware of color conventions ☺

Test Statistics : another Example



CL(s+b) = 3.7 %

1-CL(b) =33 %

CL(s)= 11 %

(check ☺)

Clearly, both signal and background had a (not-so-large) negative fluctuation … to avoid a 95% exclusion, define

$$CL(s) = \frac{CL(s+b)}{1-CL(b)}$$

(beware : NOT a p-value!)

Profile Likelihood Revisited

Likelihood Ratio defined previously

- **allows to extract POI's μ out of variables in sample $\{x\}$**
- **did not include nuisances**

Profile Likelihood approach : Introduce a control sample $\{y\}$ to constrain nuisances :

$$\mathcal{L}(\vec{x},\vec{y}\,|\,\mu,\theta) = \mathcal{L}(\vec{x}\,|\,\mu,\theta)\ \mathcal{L}(\vec{y}\,|\,\theta)$$

(n.b.: variables x and y , nor samples, need to be disjoint, i.e. sidebands…)

PL Test statistic:

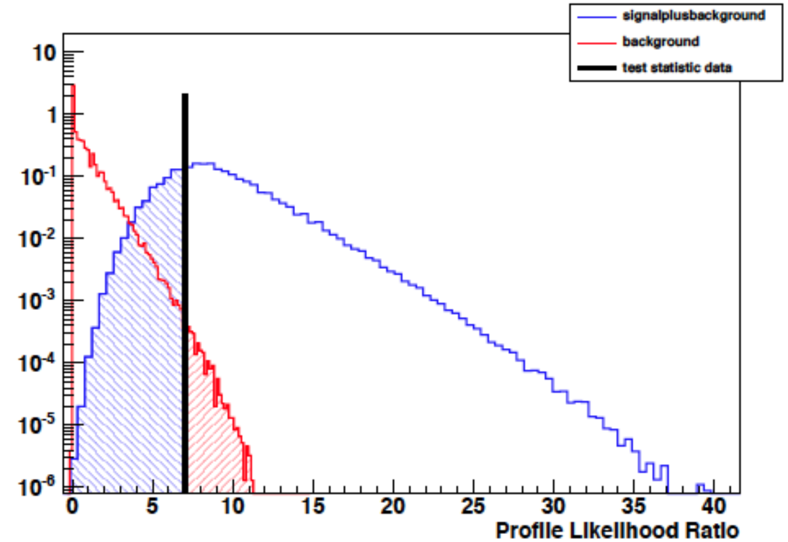
$$q_\mu(\mu) = -2\ln\lambda(\mu) = -2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu},\hat{\theta})} \quad \Rightarrow \begin{array}{l} \text{fix } \mu \text{ , fit } \theta \\ \text{fit } \mu \text{ , fit } \theta \end{array}$$

tends to a χ2 distribution with 1 ndof

p-value :

$$p(\mu^{OBS}) = \int_{q_\mu^{OBS}}^{\infty} dq_\mu \mathcal{P}(q_\mu\,|\,\mu)$$

For complex likelihoods, distributions of qm via MonteCarlo integration: CPU-consuming…



A Quasi-Academic Illustration : Testing Exclusion

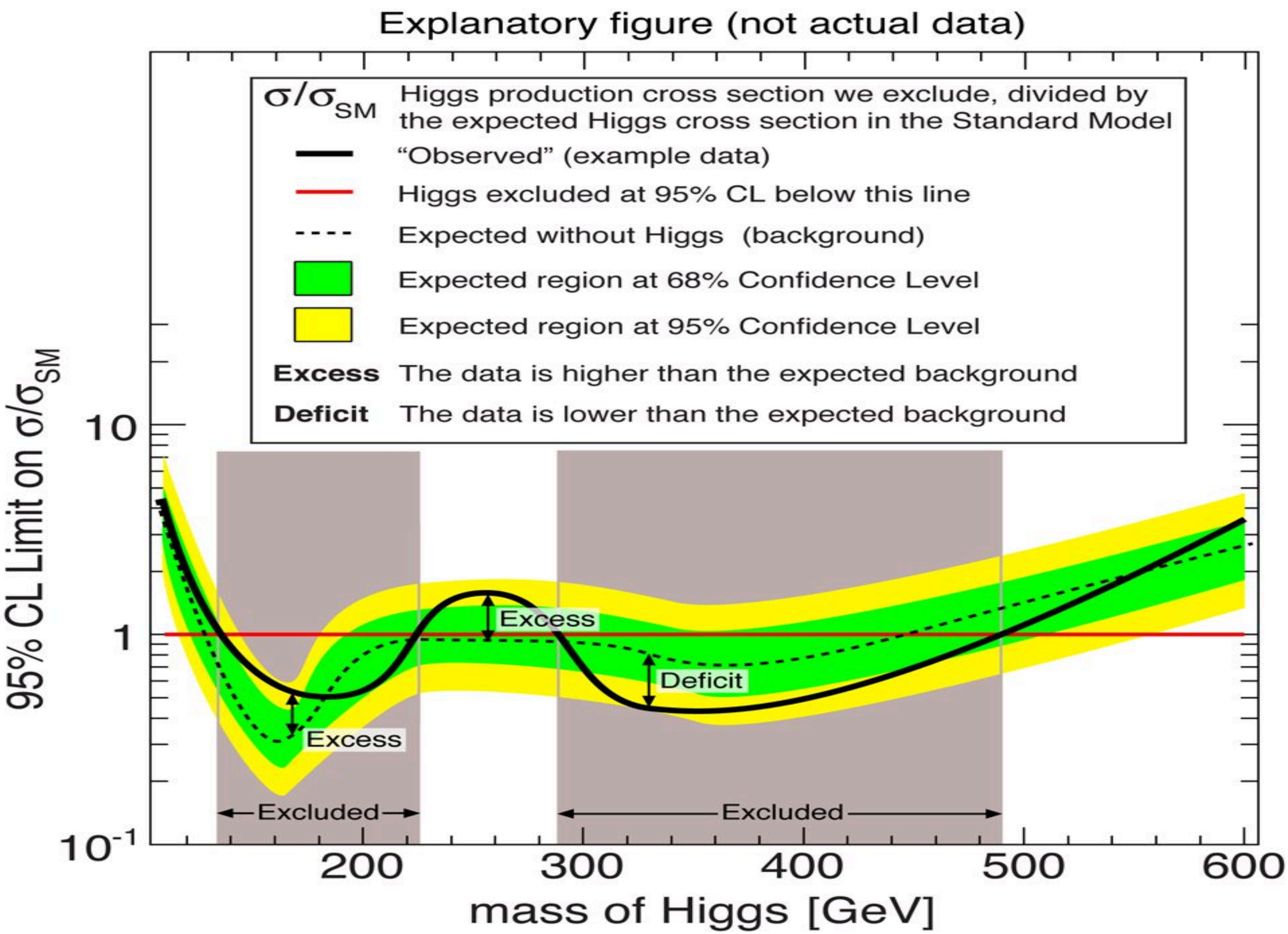
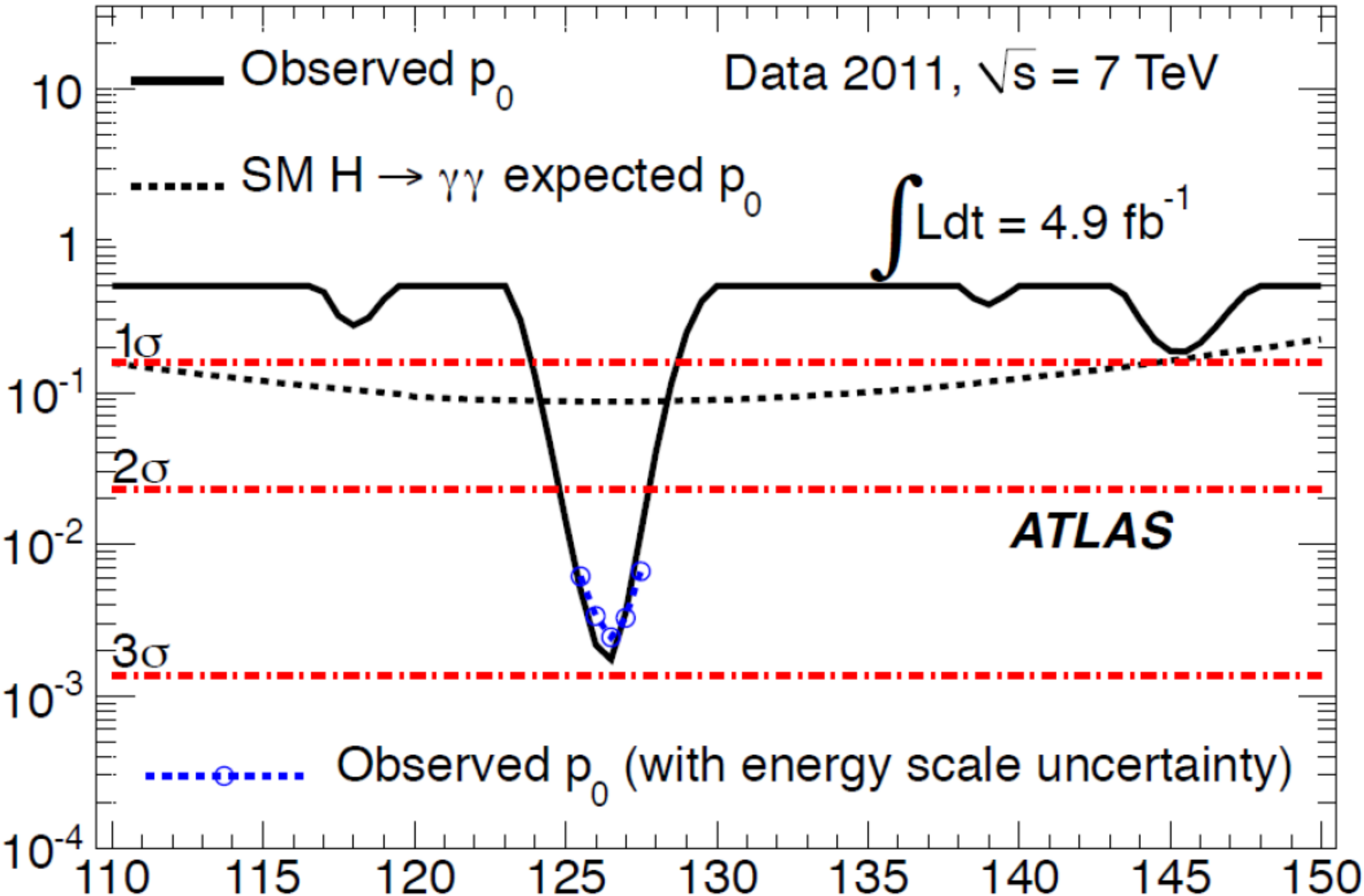


Figure A

Less Academic Illustration : Testing Observation



ATLAS+CMS : Profile-Likelihood and CL(S)

In practice, ATLAS and CMS agreed on using a modified version :

$$\tilde{q}_\mu(\mu) = -2\ln\lambda(\mu) = -2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu},\hat{\theta})} \; , \; 0 \le \hat{\mu} \le \mu$$

In fact it's slightly different, but this one is sufficient for our discussion here…

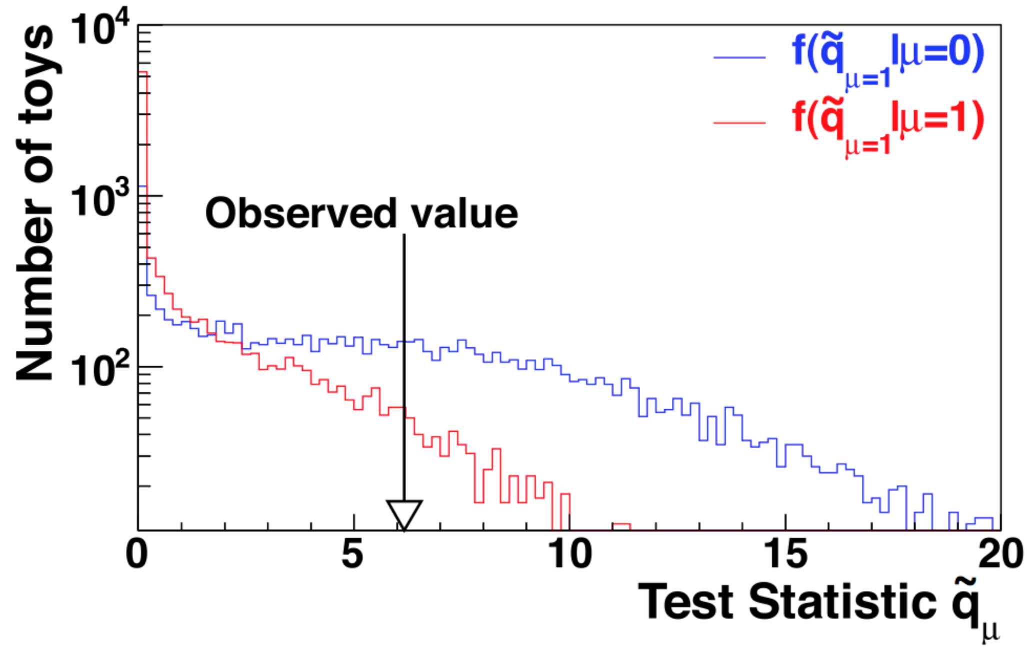
Constraint on $0 \le \hat{\mu} \le \mu$: upwards fluctuations (i.e. $\hat{\mu} > \mu$) should not be interpreted as a disfavor of signal hypothesis

Also, ATLAS and CMS agreed on using

$$CL_S(\mu) = \frac{p_\mu}{1-p_b}$$

Note : the modified test statistic follows roughly a distribution $\frac{1}{2}\delta(\tilde{q}_\mu)+\frac{1}{2}\chi^2$

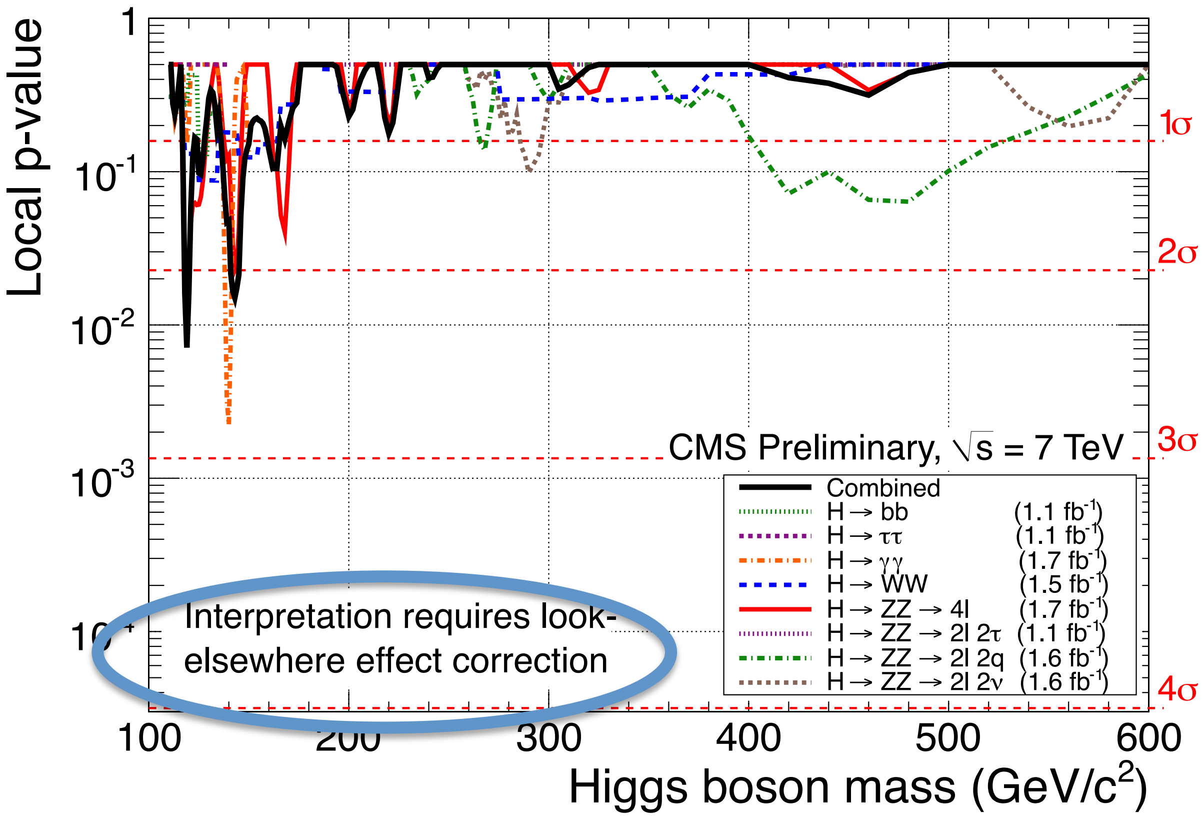
Useful asymptotic property, allows for CPU-saving approximations (i.e. the "Asimov" dataset)



THE "LOOK-ELSEWHERE-EFFECT"

Trials Factor : an apparently statistically significant hypothesis-testing conclusion

- **may actually arise due to fluctuations**
- **should likely arise in cases of large populations or large parameter spaces**
  - **i.e. in a ~1K sampling, an event showing a ~3σ deviation is no surprise!**
- **To be specific : take one HEP-motivated scenario : the Higgs search**



The "Look-Elsewhere-Effect"

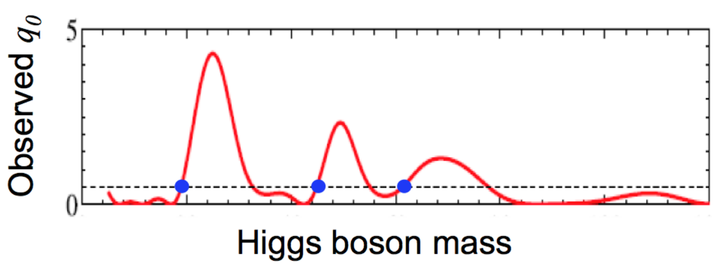
Scenario : apparently statistically significant hypothesis-testing conclusion

- may actually arise due to fluctuations
- should likely arise in cases of large populations or large parameter spaces
  - i.e. in a ~1K sampling, an event showing a ~3σ deviation is no surprise!
- To be specific : take one HEP-motivated scenario : the Higgs search

The effect can be evaluated via MonteCarlo

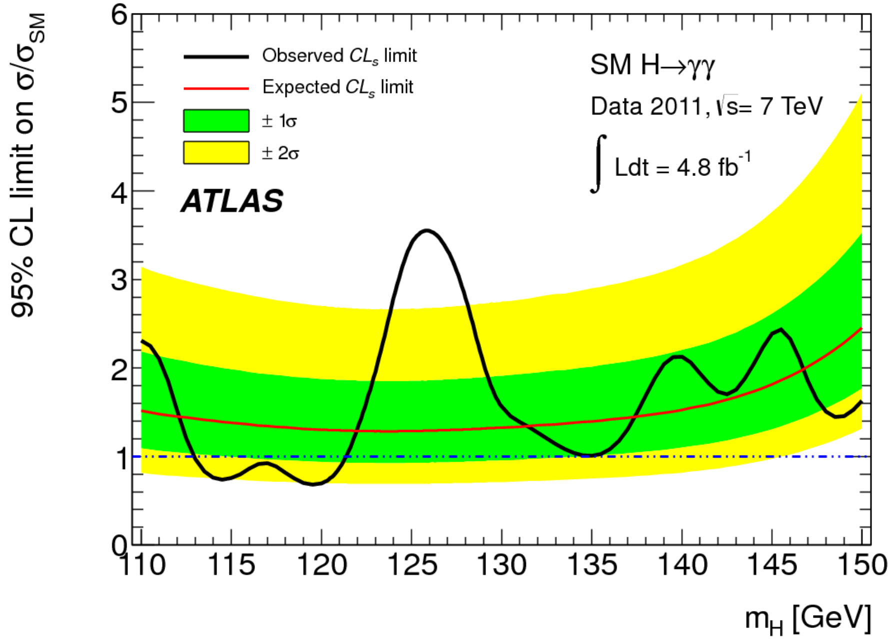
- generate simulated experiment outcomes out of a background-only PDF
- find the largest 'local' significance over the entire search range,
- extract the distribution of "local" significances

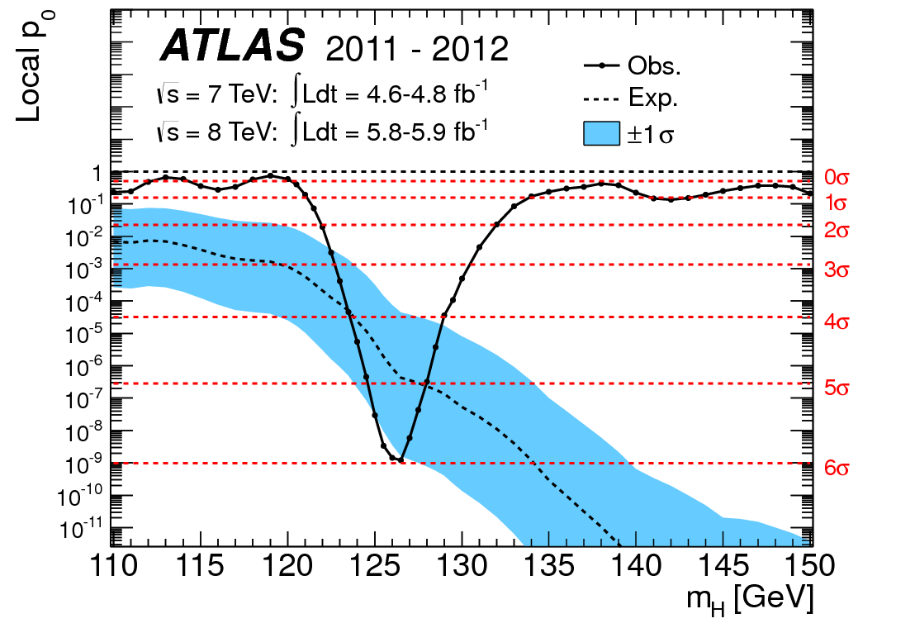
Can be approximated via asymptotic formula, and extrapolated from lower-statistics toy MonteCarlo

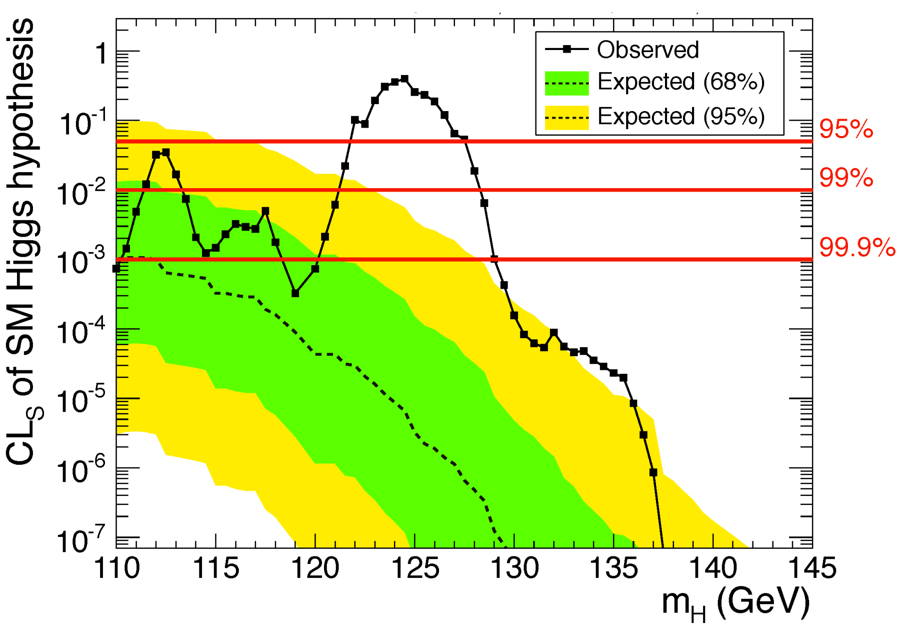


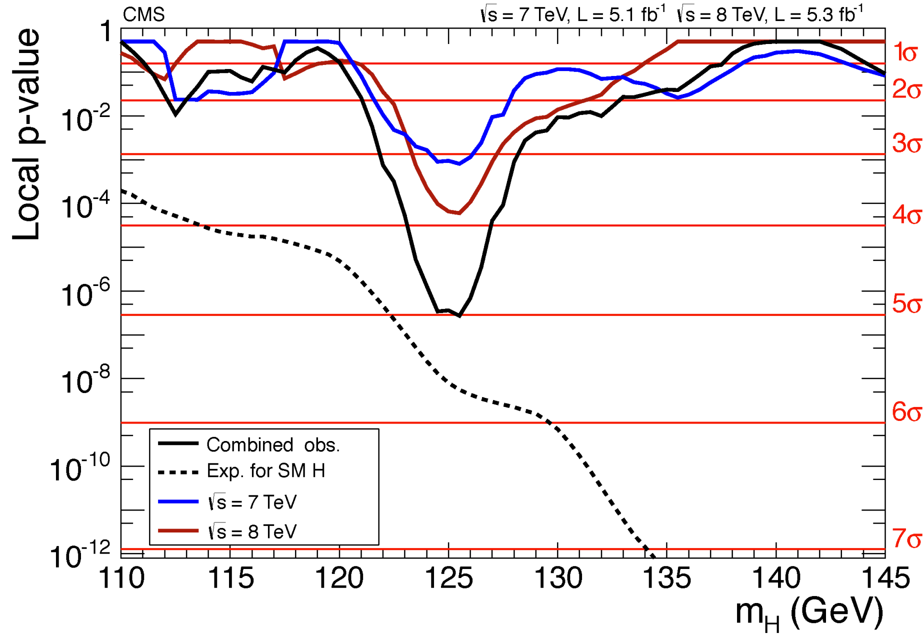
$$p_b^{global} = P(q_0(\hat{m}_H) > u) \leq \langle N_u \rangle + \frac{1}{2} P_{\chi^2_1}(u)$$

ATLAS+CMS : Profile-Likelihood and CL(s)

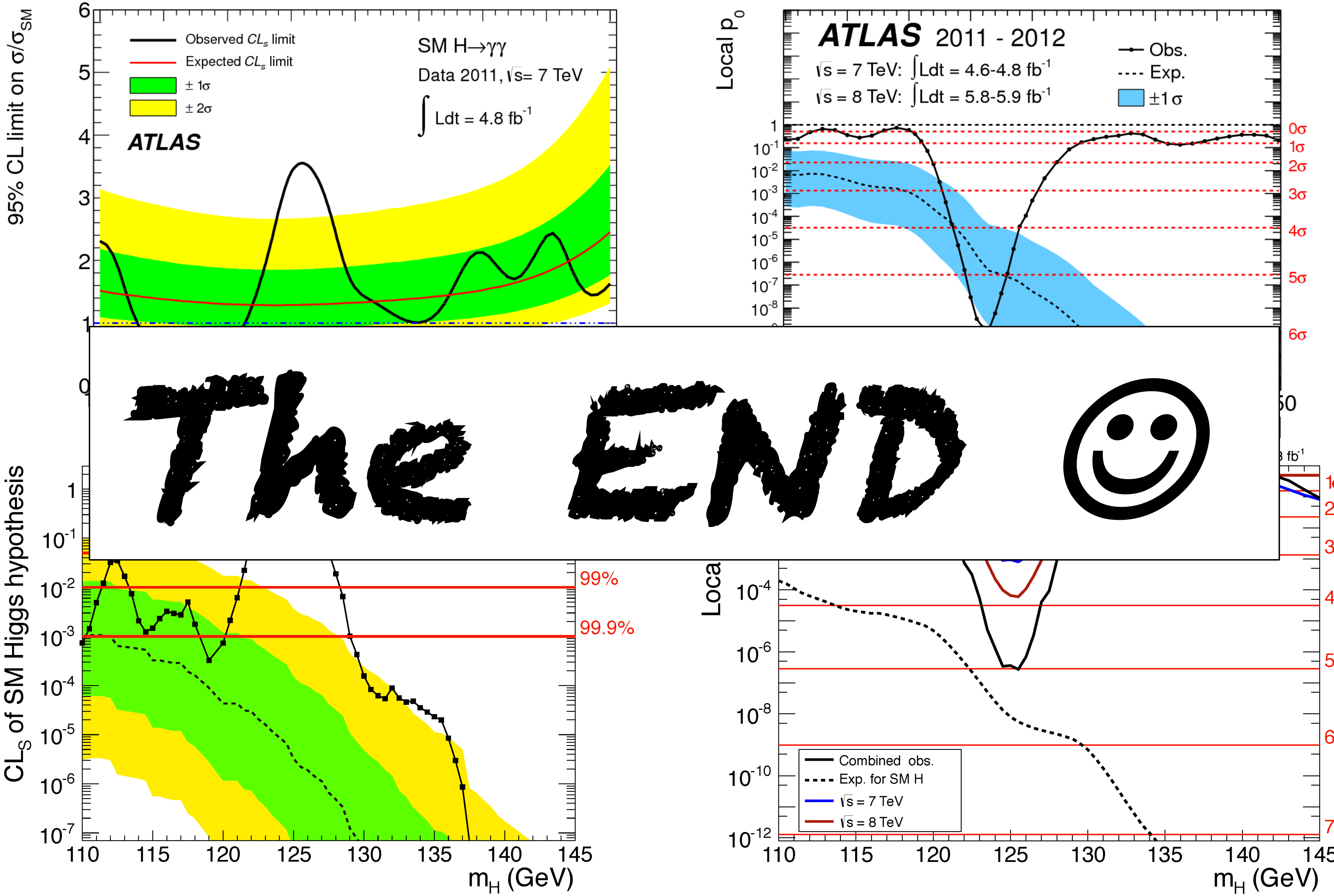








ATLAS+CMS : Profile-Likelihood and CL(S)



The END ☺