Statistics

A pragmatic way in particle physics

Lecture III

Xiaohu SUN, IHEP, 11-03-2014

What is this lecture about?

- A fresh reminder of why we need statistics in particle physcis
  - It helps to quantify the agreement/disagreement between data and our expectation (model), especially when the uncertianties are introduced
  - It helps to make decision “find or not a new particle?”
  - It helps to communicate physics results among experimentists as well as theorists

- Previously, I introduced the basics of probability: the definition, E[x], V[x], error propagation and some important distributions in Lecture I, parameter estimation and test statistic in Lecture II
- This time, I discuss
  - Theo model → stat model → likelihood → likelihood ratio
  - Profile likelihood ratio fit
  - Likehood ratio test

PDF – statistical model

- In high energy physics, one usually considers the signal and background pdfs as well as the total number of evens in constructing our statistical model, simply making the product

$$\mathcal{P}(\{x_1 \ldots x_n\}|\mu) = \mathrm{Pois}(n|\mu S + B) \left[\prod_{e=1}^{n} \frac{\mu S f_{\mathrm{S}}(x_e) + B f_{\mathrm{B}}(x_e)}{\mu S + B} \right]$$

Marked Poisson model

The total Nb

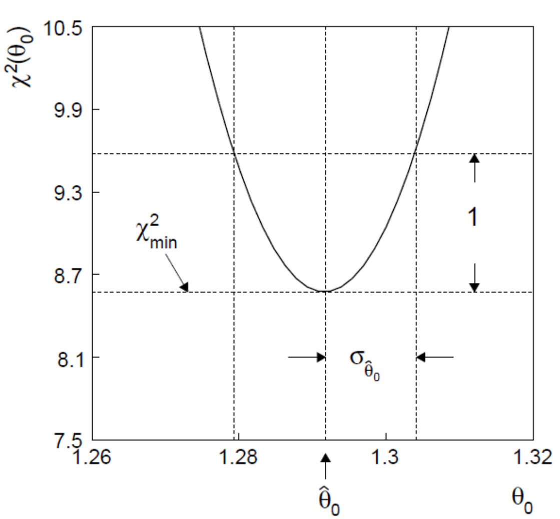
$$\prod_{i=1}^{n} f(y_i; \boldsymbol{\theta})$$ where,
pdf = $r_S$pdf(S)+$r_B$pdf(B)

- This statistical model is a function of observables (inv mass, BDT etc.) with our POI embedded in the theoretical model
  - (μ*Sf(S)+Bf(B))/(S+B)
- Equivalently, one can also interprete the stat model as a product of Poissonian terms for each bin (like a binned likelihood)

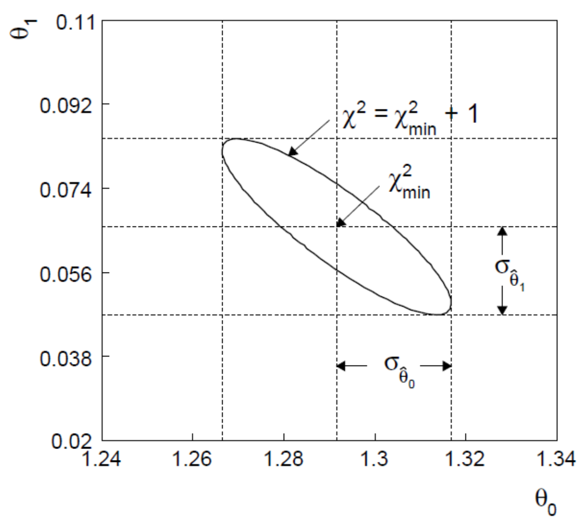
$$\mathcal{P}(n_b|\mu) = \mathcal{N}_{\mathrm{comb}} \prod_{b \in \mathrm{bins}} \mathrm{Pois}(n_b|\mu\nu_b^{\mathrm{sig}} + \nu_b^{\mathrm{bkg}})$$

Systematic uncertainties → nuisance

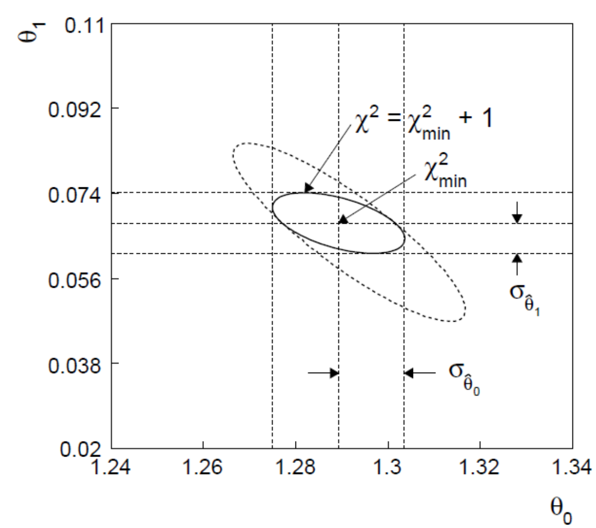
- From the Frequentist's point of view, one introduces nuisance parameter to include the effects from systematic uncertainties
  - from the detector: jet enery scale, muon momentum scale ...
  - from the data-driven backgrouds: rescaled QCD backgrounds
  - from the theoretical uncertainties: ttbar cross section ...
- As I showed in Lecture II, the introduction of nuisance parameter no matter with a prior or without one, enlarges the error of POI



no nui



one free nui



one constrined nui

Nuisance → decoration on stat model

- To introduce the preknowledge on nuisance parameters, one use Gaussian contraints or others in the stat model additionally

$$\mathcal{P}(n_c, x_e, a_p \mid \phi_p, \alpha_p, \gamma_b) = \prod_{c\in\text{channels}} \left[\text{Pois}(n_c|\nu_c) \prod_{e=1}^{n_c} f_c(x_e|\boldsymbol{\alpha})\right] \cdot G(L_0|\lambda, \Delta_L) \cdot \prod_{p\in\mathbb{S}+\boldsymbol{\Gamma}} f_p(a_p|\alpha_p)$$

from different analysis channels

usually, a guassian constraint for each nuisance parameter

- Similarly, one can regard it as a product of many Poissonian terms

$$\mathcal{P}(n_{cb}, a_p \mid \phi_p, \alpha_p, \gamma_b) = \prod_{c\in\text{channels}} \prod_{b\in\text{bins}} \text{Pois}(n_{cb}|\nu_{cb}) \cdot G(L_0|\lambda, \Delta_L) \cdot \prod_{p\in\mathbb{S}+\boldsymbol{\Gamma}} f_p(a_p|\alpha_p)$$

Likelihood as usual

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i; \boldsymbol{\theta})$$

Systematic uncertainties [profile]

- The problem is that one is definitely NOT interested in the values of any nuisance parameter, since one is ONLY interested by POI
- With all the nui, are we are going to do the fit and get a multidimensional LL or LS? Better yet, how do we visualize it?
  - NO, we don't care other dimensions but the dimension on POI
  - Profile them
- In frequentist statistics, one can construct a profile likelihood ratio (PLR) to estimate only the POI while profiling all the not-interesting nuisance parameter

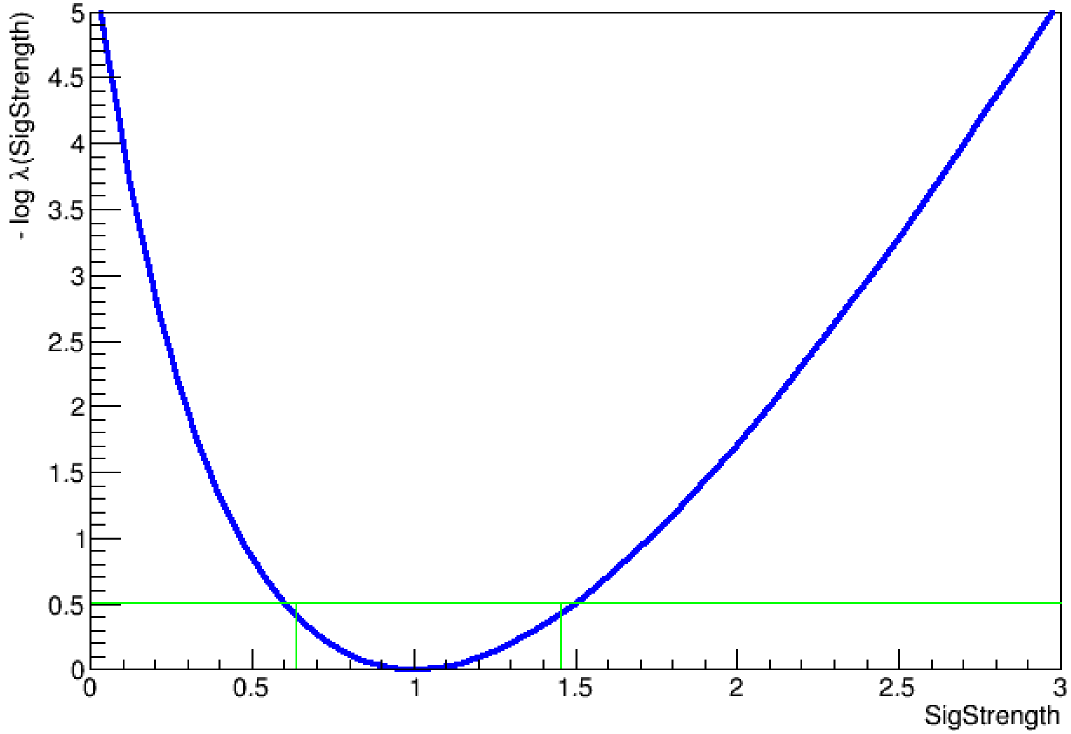
$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} \qquad 0 \leq \lambda \leq 1$$

- Single hat θ and μ in the denominator are obtained in the global unconditional maximized likelihood
- Double hat θ in the numerator is the value of θ that maximizes L for a cerntain μ

Profile likelihood ratio

- In profile likelihood ratio fit, one can get the one-dimensional likelihood curve for only POI

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



* -2Logλ is the profile χ^2 function
* Be ware of that, one has to check the behaviours of all nuisance parameter in the profilling

PLR → test statistic

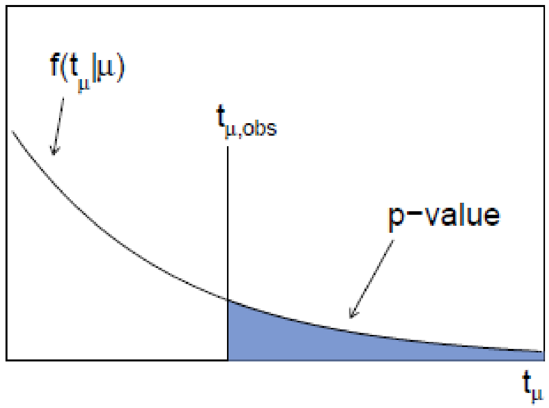
- To build a test statstistic, which can help to make decision to reject H_0 or H_1, one needs to find a quantity that discriminate S+B hypothesis and B-only hypothesis the most

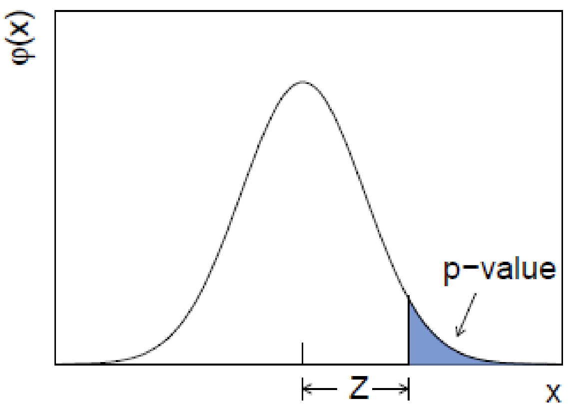
$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} \qquad 0 \le \lambda \le 1$$

More conveniently $$t_\mu = -2\ln\lambda(\mu)$$ [inf , **0**]

Similarly, the p-value $$p_\mu = \int_{t_{\mu,\mathrm{obs}}}^{\infty} f(t_\mu|\mu)\, dt_\mu$$

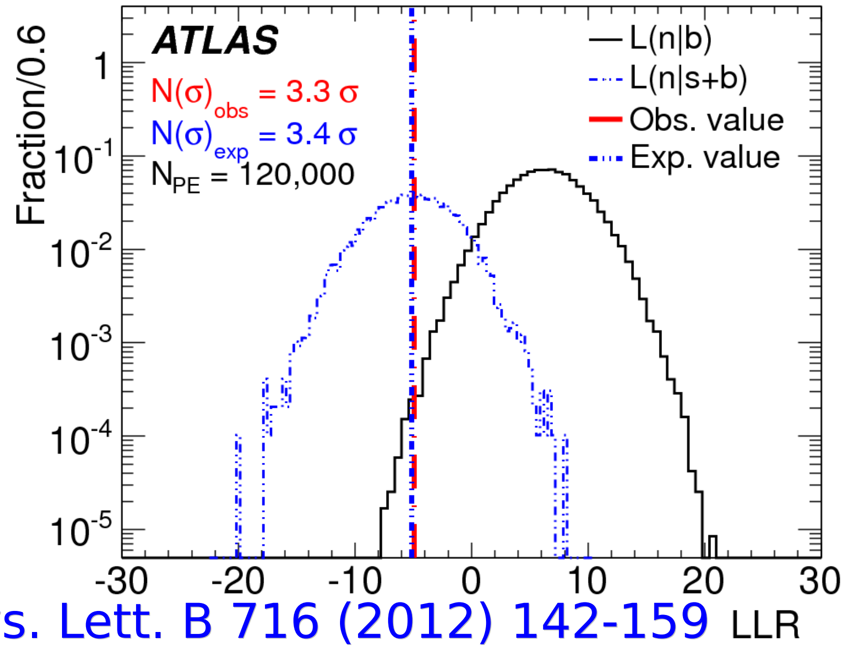
(a)





Toy Monte Carlo method

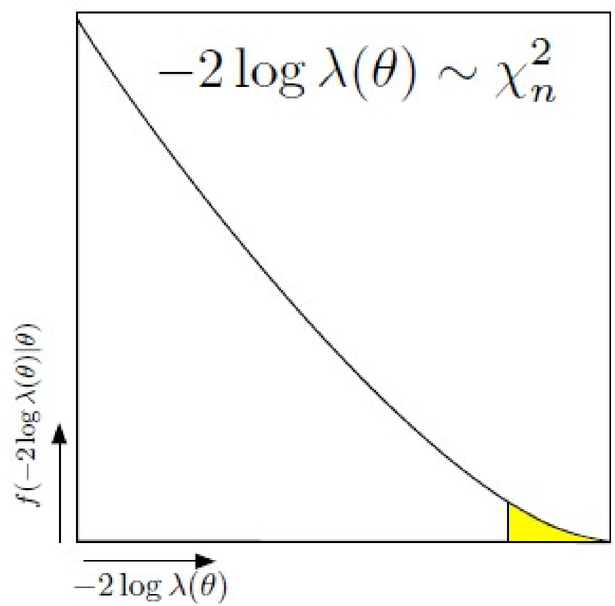
- To produce the distribution of the test statistic, traditionally one generates toy experiements
  - Randomize the POI (μ) measurements and all nuisance measurements (NOT parameters) around presumed values
    - μ=0 for B-only, 1 for S+B, nui measurements = 0 usually
  - Fit our statistical model to the toy to obtain the LR → -2lnλ
- The more toys experiement one throws, the more time will be cost

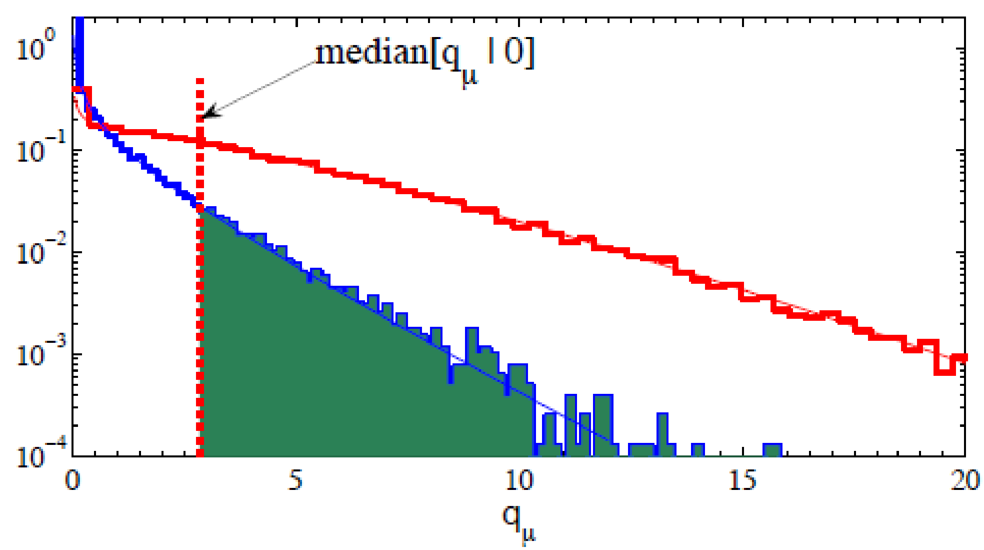


Phys. Lett. B 716 (2012) 142-159

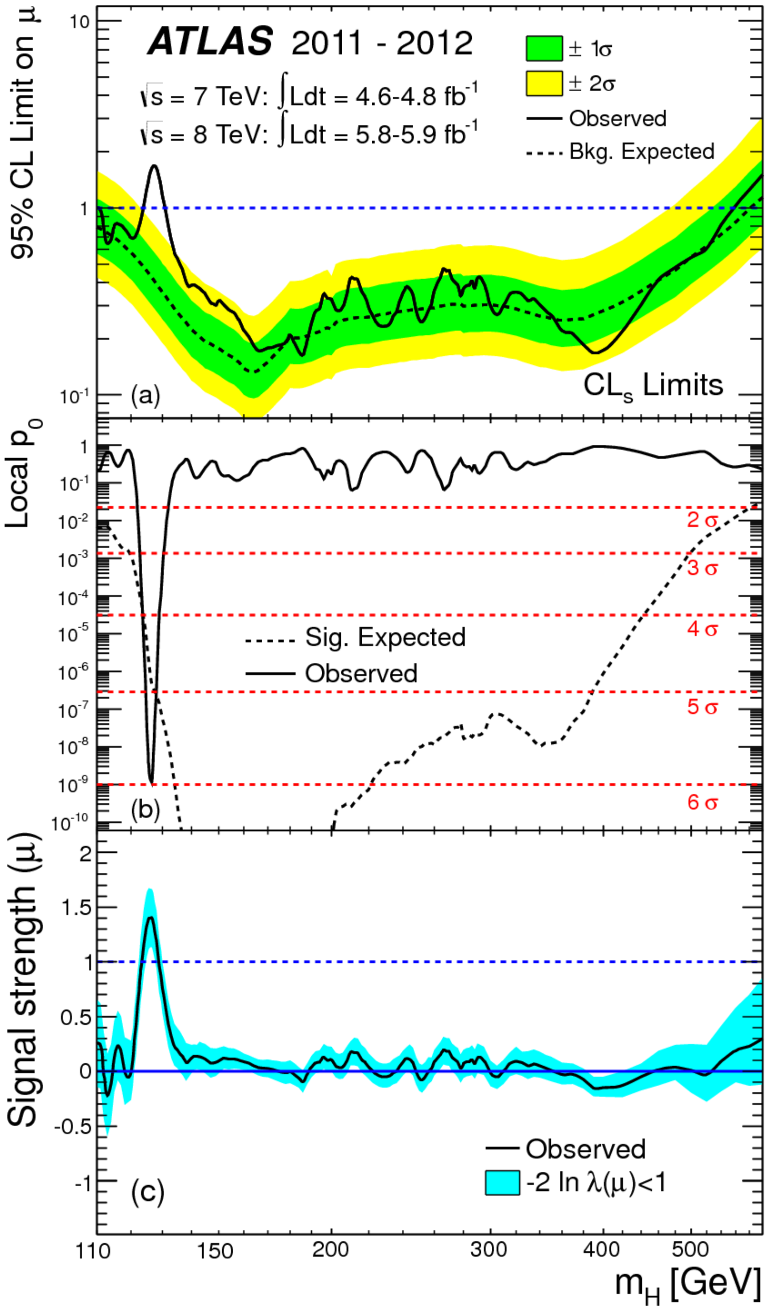
Asymptotic property

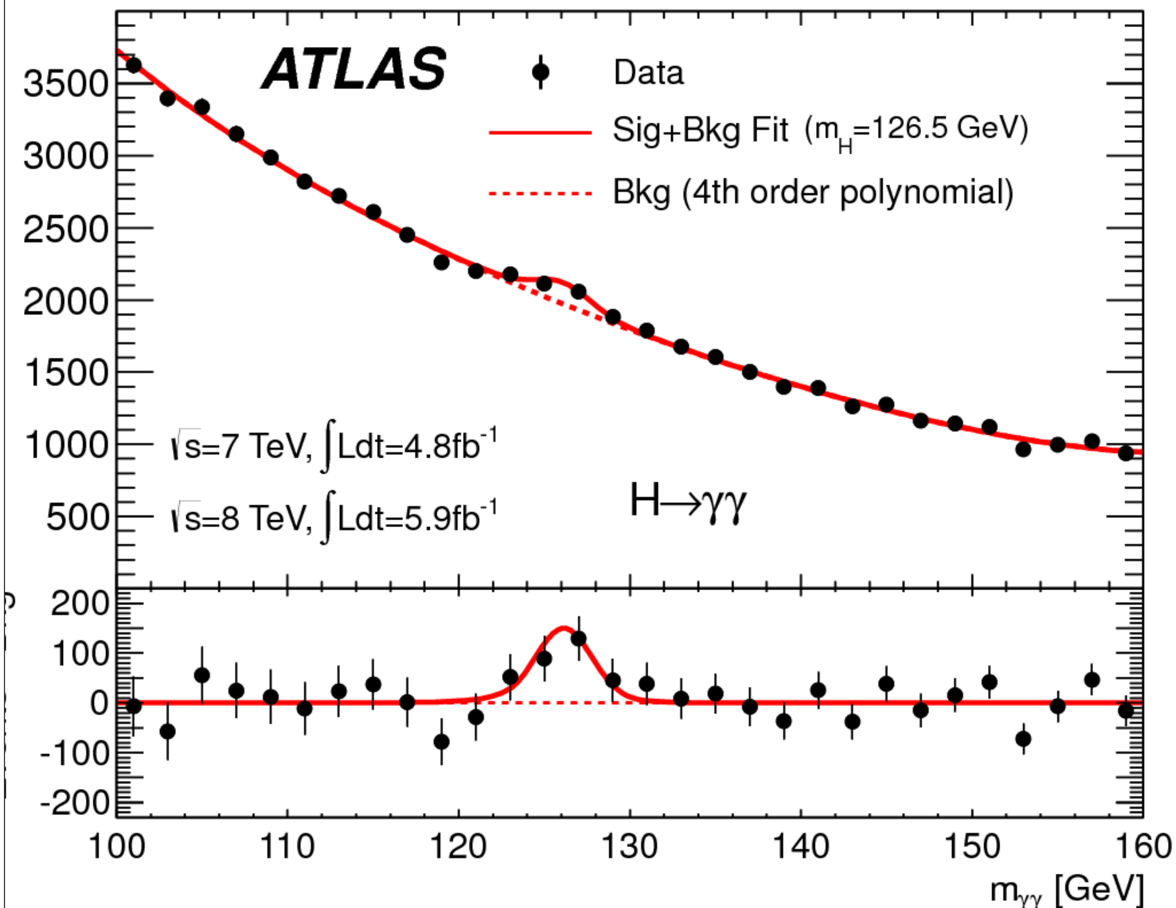
- Wilks's theorem states that under certain conditions the distribution of $-2\ln\lambda(\mu=\mu_0)$ given that the true value of μ is μ_0 converges to a chi-square distribution
  - When there is sufficient data that the log-likelihood function is parabolic
- Thus, we can calculate the p-value for the background-only hypothesis without having to generate Toy Monte Carlo!





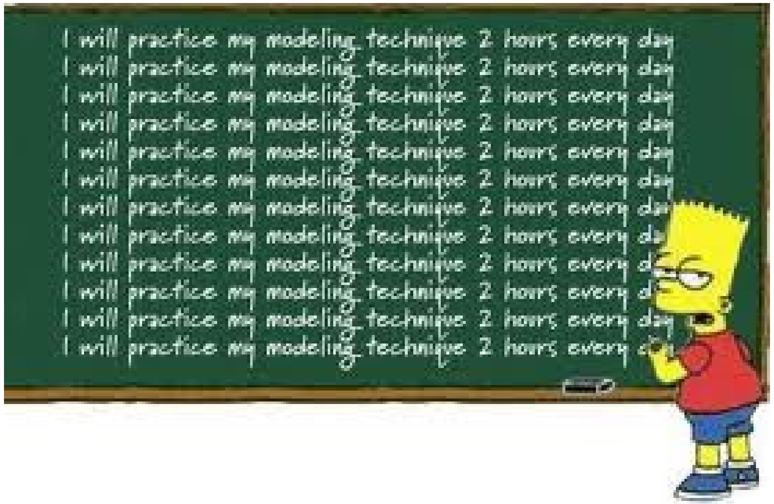
Discoveries, upper limits





Summary

- In this lecture, some pratical issues are introduced
  - The construction of statistical model
  - The construction of profile likelihood ratio
    - for fit as well as for test statistic
  - The toy experiment vs asymptotic distributions
- In all three lectures, we cover most useful aspects on statstics used in ATLAS collaboration
- Hope you enjoyed
- Now it is time to practice in your real analysis!



It is time
to find the POI