ROOT and statistics tutorial

Exercise: Discover the Higgs, part 2

Attilio Andreazza

Università di Milano and INFN

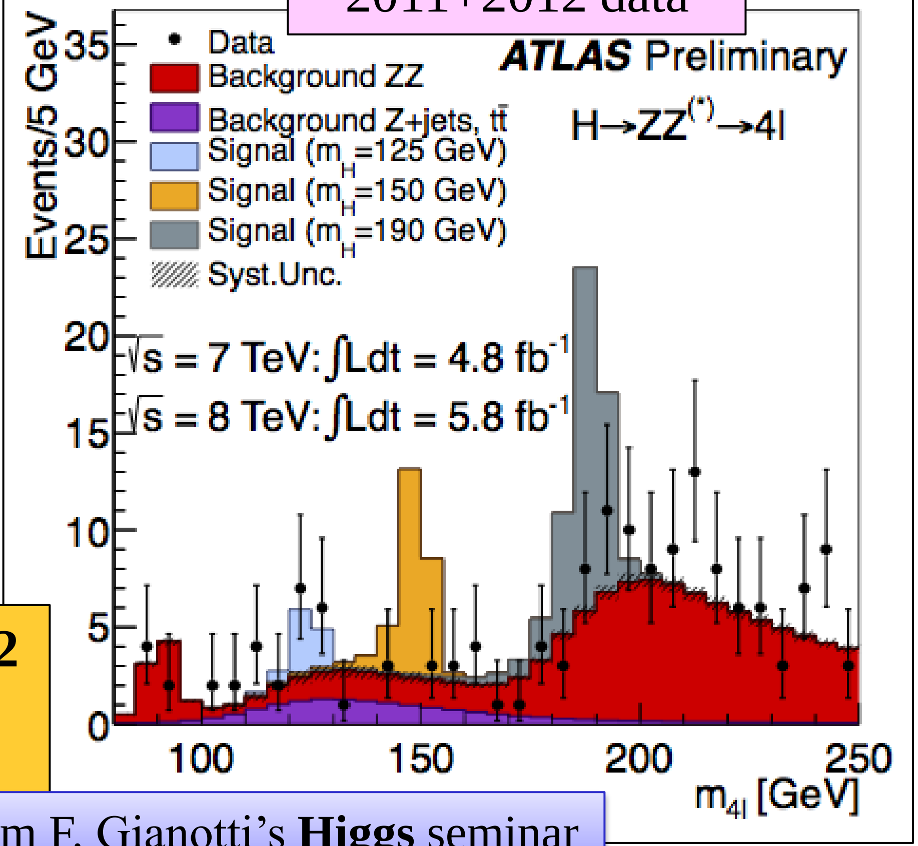
Caterina Doglioni

Université de Genève

Outline

- What will we do today:
  - **Discover the Higgs boson of course!** ...well, check ATLAS is not cheating us ☺
- What we will learn!
  - Computing confidence levels based on Poissonian statistics.
  - Compute confidence levels based if the expected $\mu_{(B,S)}$ is uncertain.
- For people running fast:
  - Expected 95% level exclusion + error bars

2011+2012 data



From F. Gianotti's **Higgs** seminar

In the region 125 ± 5 GeV

Dataset	**2011**	**2012**	**2011+2012**
Expected B only	$2_{\pm 0.3}$	$3_{\pm 0.4}$	$5.1_{\pm 0.8}$
Expected S m_H=125 GeV	$2_{\pm 0.3}$	$3_{\pm 0.5}$	$5.3_{\pm 0.8}$
Observed in the data	4	9	13

Confidence level definition

- Definition of confidence level
  - ***N.B.:*** *this is **the frequentist definition**, not the Bayesian one from the lectures, but that allows you to make all computations by yourself and to grasp the main features of the problem.*
  - A certain criterion rejects an hypothesis with C.L. α if, in case the hypothesis is true, it would be erroneously rejected by that criterion on a fraction $1-\alpha$ of the cases.
- In our exercise:
  - We observe a certain number of events $\mathbf{N_{obs}}$
  - We expect a certain number of background events: $\mathbf{N_B}$
  - **Discovery:** we reject the hypothesis our sample contains only background events if:

$$P\left(N \geq N_{\mathrm{obs}} \mid N_{\mathrm{B}}\right) < 1 - \alpha$$

  - **Exclusion:** we reject an hypothesis expecting $\mathbf{N_S}$ signal events if:

$$P\left(N \leq N_{\mathrm{obs}} \mid N_{\mathrm{B}} + N_{\mathrm{S}}\right) < 1 - \alpha$$

Confidence level computation

- At first just assume a Poisson statistics:
  - We can compute the probability summing the probabilities of all N up to N_{obs}
  - **But we want to use a Monte Carlo method!**
    Why? It will be easier to extend to the treatment of systematic uncertainties (this is what BAT or RooStats do for example).
- What to do:
  - Sample the probability distribution M times (say M=10000)
  - Count how many cases M' the value of N exceeds (or is lower than, if appopriated) N_{obs}.
  - We can reject the hypothesis if $M'/M<1-\alpha$.
  - Something like:

```
Int_t M=0;
for (Int_t i=0; i<10000; i++) {
  Int_t N = gen.Poisson(NB);
  if ( N>=Nobs ) M++;
}
Double_t CL = 1.-M/10000.
```

Computing confidence levels

- Neglecting uncertainties on $\mathbf{N_B}$
- With which CL can we reject the background-only hypothesis when using:
  - Only 2011 data
  - Only 2012 data
  - The combined set
- If one would have been observed only the expected backgroun (i.e. $\mathbf{N_{obs}}$ =2,3 and 5 events respectively for 2011, 2012 and combined dataset), with which confidence level one would have rejected the hypothesis of the Higgs presence?
- Repeat adding the uncertainty on the μ value of the Poisson distribution. Assume the uncertainties on $\mathbf{N_B}$ and $\mathbf{N_S}$ are fully correlated.
  - In such a situation $P\left(N \mid \langle N_\mathrm{B} \rangle, \sigma_B\right) = \int dN_\mathrm{B} \mathrm{Poisson}\left(N \mid N_\mathrm{B}\right) P\left(N_\mathrm{B}\right)$ and the integral can be performed by sampling $\mathbf{N_B}$ from a Gaussian distribution and afterward sampling *N*.

Computing exclusion limit

- **Observed exclusion limit:**
  - after getting $\mathbf{N_{obs}}$ events, all $\mathbf{N_S > N_{0,S}}$ are excluded at confidence level α, where is the $\mathbf{N_{0,S}}$ minimum one satisfying the relation:

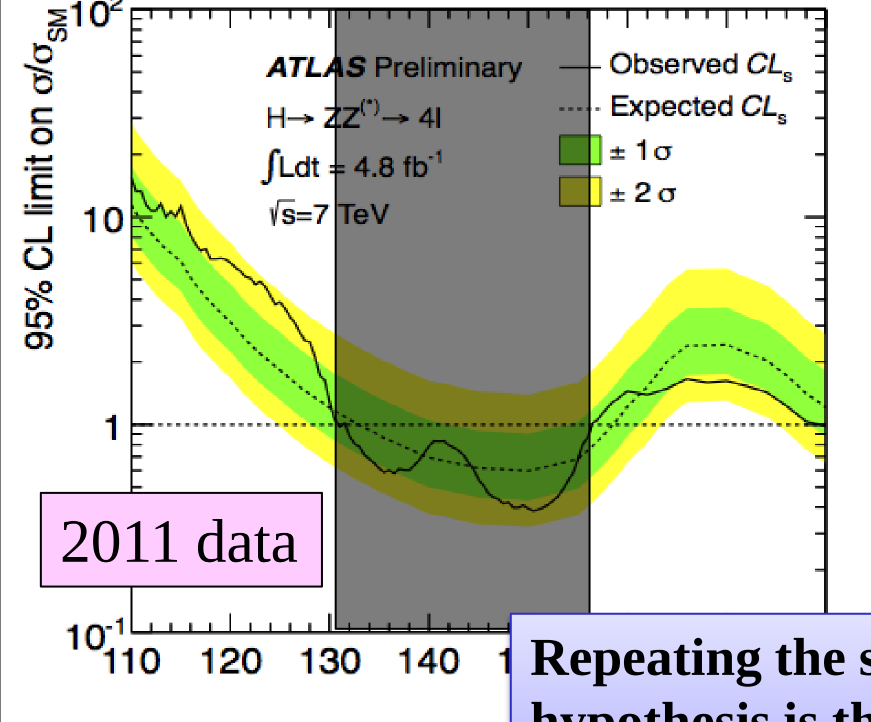
    $$P\left(N \le N_{\mathrm{obs}} \mid N_{\mathrm{B}} + N_{\mathrm{S}}\right) < 1 - \alpha$$

  - *Determining this minimum, even in this simple case is quite computationally expensive, and special tools are usually employed.*
- **Expected exclusion limit:**
  - Is the one that would obtained if $\mathbf{N_{obs}}$ would correspond to the median of

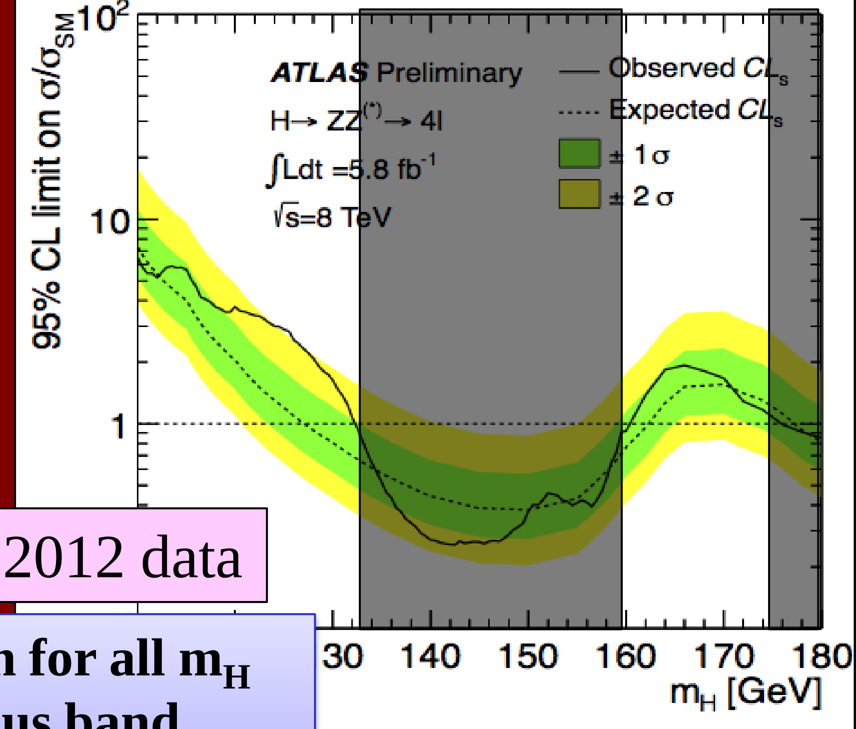
    $$P\left(N \mid \langle N_{\mathrm{B}} \rangle, \sigma_B\right)$$

  - The $\pm 1\sigma$ expected values correspond to the limit that would be obtained if $\mathbf{N_{obs}}$ would coincide with the 16% and 84% percentiles of $P\left(N \mid \langle N_{\mathrm{B}} \rangle, \sigma_B\right)$
  - The $\pm 1\sigma$ expected values correspond to the limit that would be obtained if $\mathbf{N_{obs}}$ would coincide with the 2.25% and 97.75% percentiles of $P\left(N \mid \langle N_{\mathrm{B}} \rangle, \sigma_B\right)$

– **Compute the expected limits for ATLAS and $\mathbf{m_H}$=125 GeV**



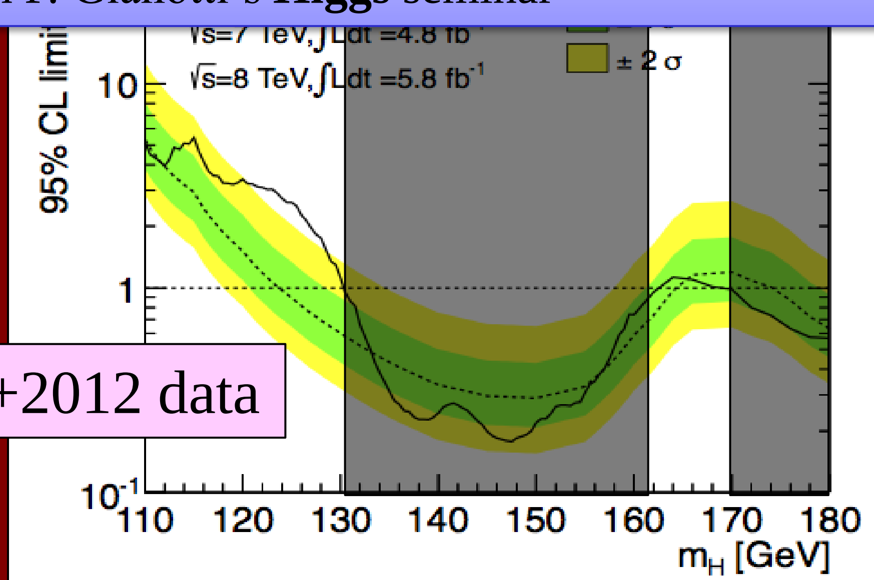
2011 data



2012 data

Repeating the same calculation for all m_H hypothesis is the way the famous band plots are produced.
From F. Gianotti's **Higgs** seminar

Excluded (95% CL):
131-162, 170-460 GeV
Expected:
124-164, 176-500 GeV



2011+2012 data