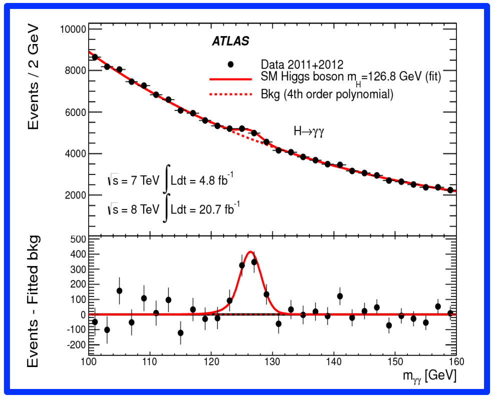
Estimation of statistical significance of a new signal within the GooFit framework on GPUs

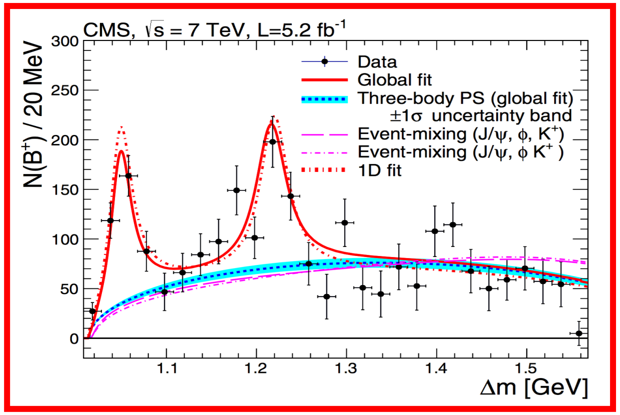
Adriano Di Florio, **Alexis Pompili**

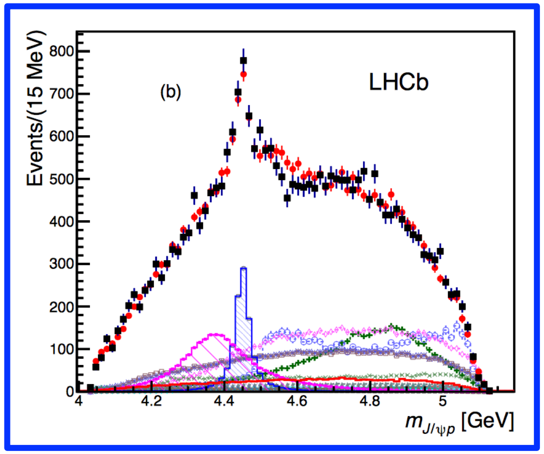
Università degli Studi di Bari
and
INFN, Sezione di Bari

Introduction

- In particle physics we often have to deal with **"signals"** that highlight a **discrepancy** with what the theory (**SM**) predicts. These signals can be **already known** or **completely new**. In any case when a **signal** is observed, we need to asses the **statistical significance, local or global.**







- In literature many papers deals with the problem of ***hypothesis testing*** and ***significance estimation*** looking, also, for analytical solutions to the problem.

Trial factors for the look elsewhere effect in high energy physics

Eilam Gross, Ofer Vitells[a]

OPEN STATISTICAL ISSUES IN PARTICLE PHYSICS[1]

BY LOUIS LYONS

Oxford University

Hypothesis testing when a nuisance parameter is present only under the alternative

BY R. B. DAVIES

Applied Mathematics Division, Department of Scientific and Industrial Research, Wellington, New Zealand

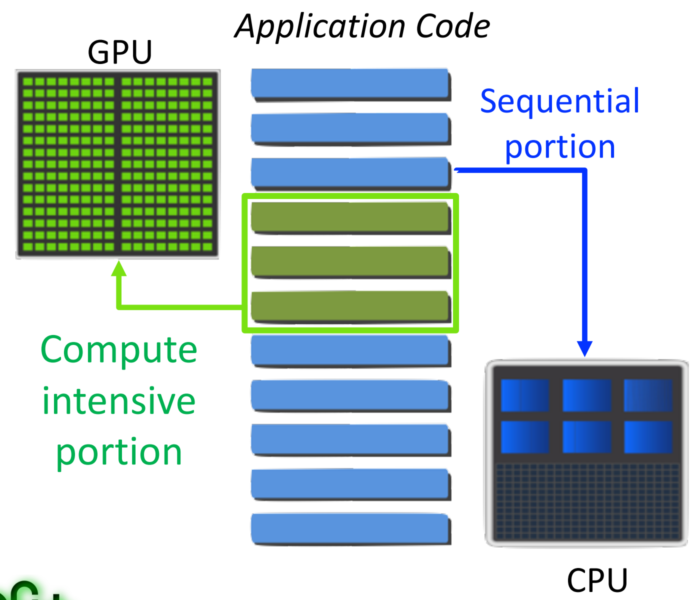
THE LARGE-SAMPLE DISTRIBUTION OF THE LIKELIHOOD RATIO FOR TESTING COMPOSITE HYPOTHESES[1]

BY S. S. WILKS

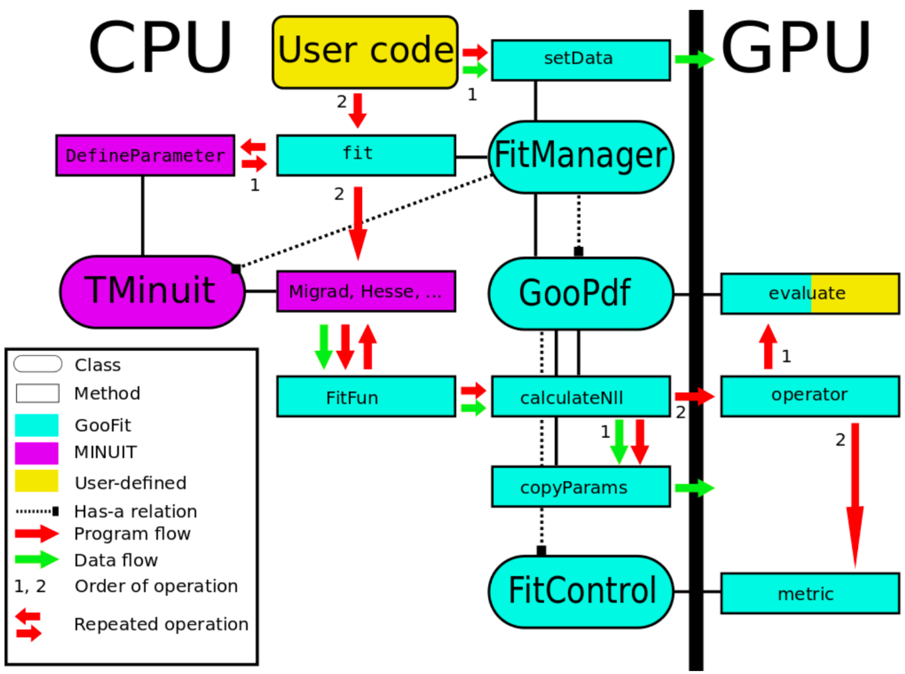
- But **sometimes** the regularity conditions of these results are not met in the typical particle physics context and, in order to estimate the statistical significance of a signal we should rely on MC Toys / pseudo experiments simulations. This kind of approach can obviously **very time consuming**! Here we show how the availability of **new tools** running on **new heterogeneous computing oriented servers** can ease the task.

GPU Computing in HEP analysis: the GooFit framework

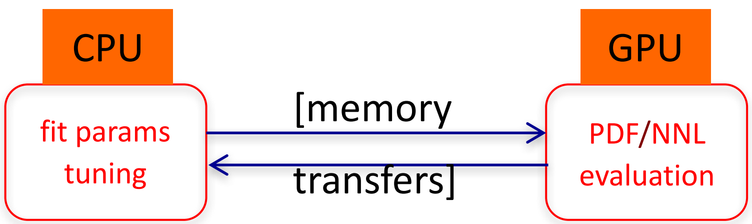
- Hetherogeneous GPU-acccelerated computing is the use of a Graphics Processing Unit to accelerate scientific applications (among other apps).



We explored the capabilities of GPU computing in the context of the 'end-user HEP analyses' by using *GooFit*.



GooFit is a data analysis tool for HEP, that interfaces ROOT/RooFit to CUDA parallel computing platform on *nVidia* GPU. It also supports OpenMP.



From the user's perspective? Applications simply run significantly faster! How much faster ? It depends - of course - on the application… We tested it firstly with the estimation of the local significance of a known signal.

Since v2.0 **Goofit** is completely integrated in python through **PyBindings** and it can run within jupyter notebooks that makes its use even easier.

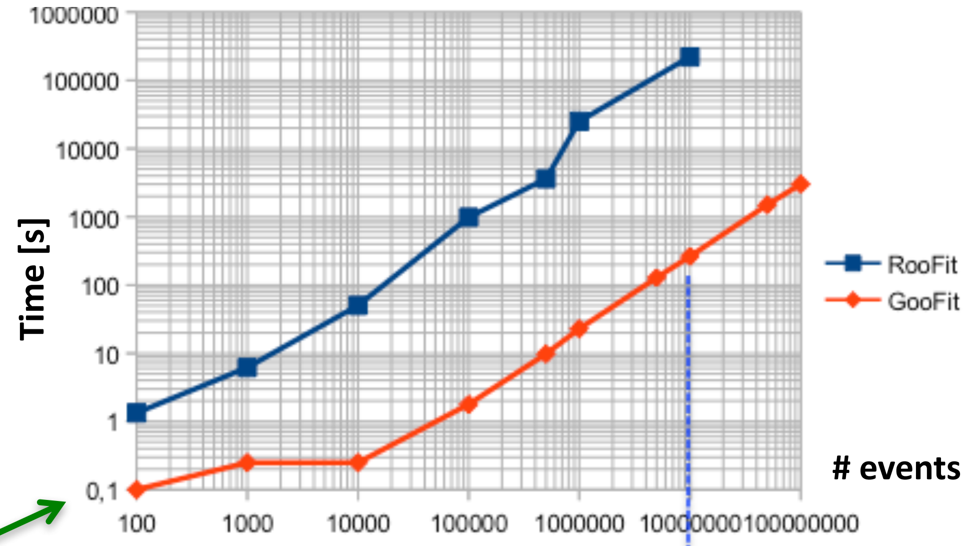
- **Parameter estimation is a crucial part of many physics analyses.**

PDF evaluation on large datasets is usually the bottleneck in the MINUIT algorithm.

***GooFit* acts as an interface between the MINUIT minimization algorithm and a parallel processor which allows a Probability Density Function to be evaluated in parallel.**

- **A preliminary test was done with an Unbinned ML fit either by using a single CPU and by using an additional GPU** (an nVIDIA Tesla C2070 hosted @ Bari T2).

  Events according to a Voigtian model (convolution is CPU-intensive) are generated & fitted. The time needed (the negligible generation time is not included) **is studied as a function of the #events:**



For 10M events: *RooFit* needs 61h+23m & *GooFit* takes 4m+39s : speed-up ~ 750

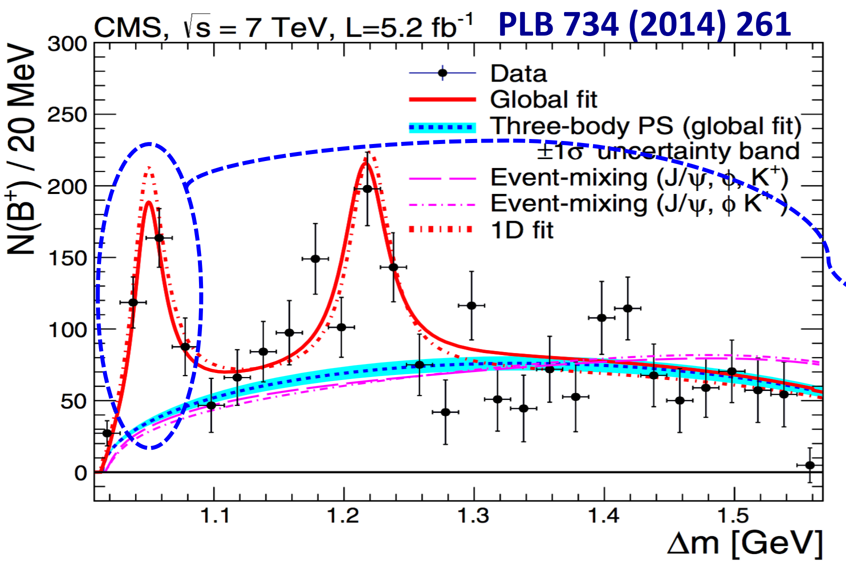
For 1M fitted events with *RooFit* … you need to wait overnight,

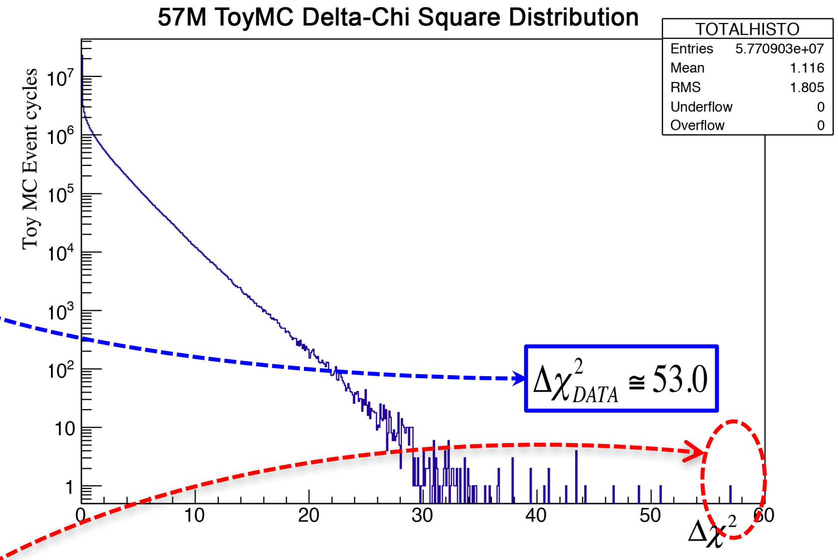
For 10M fitted events with *GooFit* … you need to take an espresso!

A first use case: local significance estimation

A first use case: local significance estiamtion

An high-statistics pseudo-experiments (toys) technique has been implemented in the `GooFit` framework in order to estimate a *p-value* and thus the (local or global) statistical significance of a signal reconstructed from data. The p-value is the probability that background fluctuations would - alone - give rise to a signal as much significant as that seen in the data.





- MC toys production was stopped once a **single fluctuation** with $\Delta\chi^2 > \Delta\chi^2_{DATA}$ was found. **Then the p-value estimation is straightforward**:

$$P = \int_{\Delta\chi^2_{obs}}^{\infty} f(\Delta\chi^2)d(\Delta\chi^2) \simeq (57.7 \cdot 10^6)^{-1} \simeq 1.73 \cdot 10^{-8}$$

Equivalent (gaussian) statistical significance:

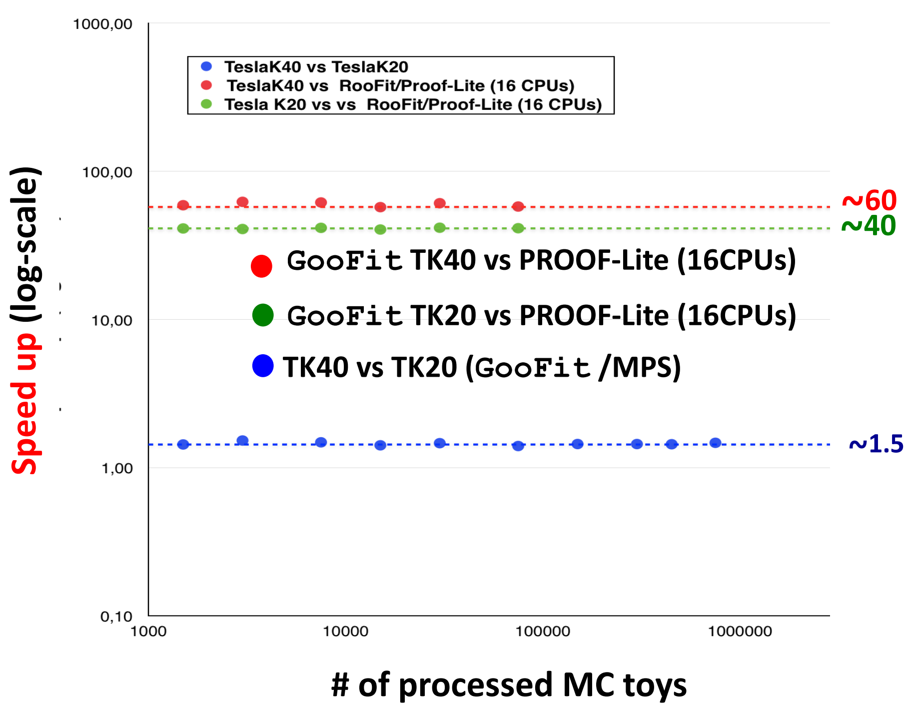
$$Z\sigma = \Phi^{-1}(1-P)\sigma \cong 5.52\sigma$$

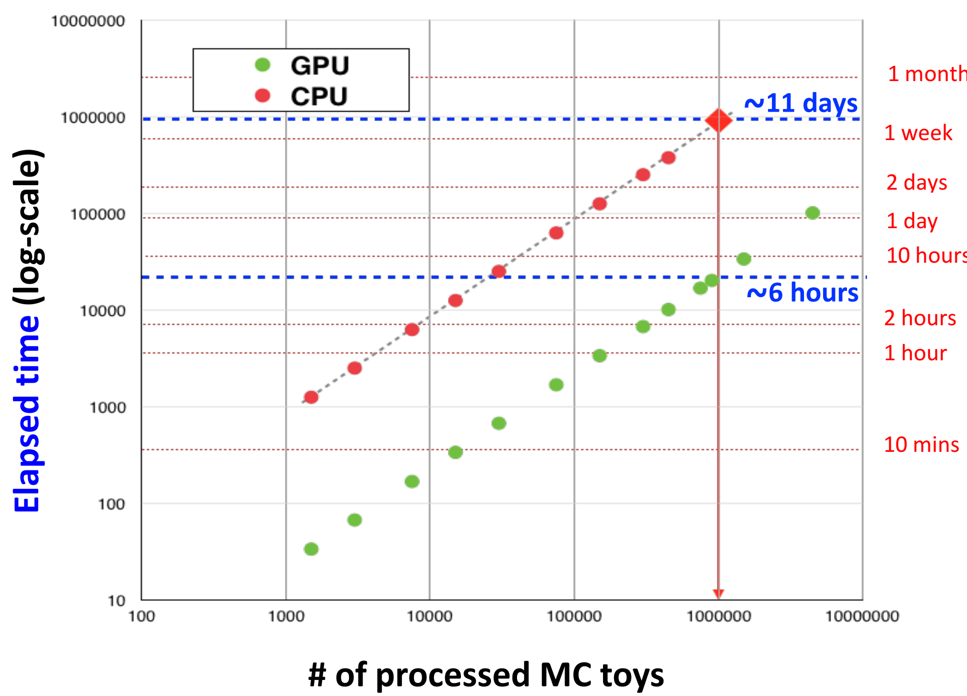
Compatible with the lower limit of 5σ for the statistical significance quoted in the CMS paper PLB 734 (2014) 261 on the basis of 50.5 millions of MC toys (by *RooFit)*.

- The optimized *GooFit* applications running, by means of the MPS, on GPUs, hosted by the servers used in the presented test, has provided a striking speed-up performance with respect to the *RooFit* application parallelized on multiple CPUs by means of *PROOF-Lite*.

- A first performances' comparison is carried out on both the servers hosting both type of GPUs (TK20 & TK40) as a function of the # of pseudo-experiments produced keeping constant the number of workers/processes.

- A second comparison is done from the point of view of the end-user/analyst having at disposal **72 CPUs and 3 GPUs (1 TK40 & 2 TK20) on 2 servers**





Exploring the applicability limits of Wilks theorem

- By means of ***GooFit*** , given the speed ups shown, it has also been feasible to explore the (asymptotic) behaviour of a likelihood ratio test statistic!

> The Wilks[*] theorem is often used to estimate the p-value associated to a new/unexpected signal. But when null hypothesis is background-only and the alternative is background+signal, often the theorem regularity conditions (see backup) are not all satisfied, and MC toys are mandatory !

- Consider the test statistic $t_\mu = -2\ln\lambda(\mu)$ [μ: *strength parameter*] as the basis of the statistical test. This could be a test for purposes of establishing the existence of a signal process (no constrain on μ)

The test statistic approaches a chi-square distribution for 1 d.o.f.

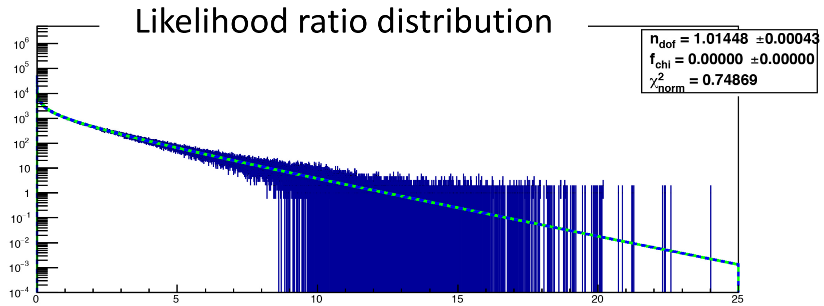
$$f(t_\mu|\mu) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{t_\mu}}e^{-t_\mu/2}$$

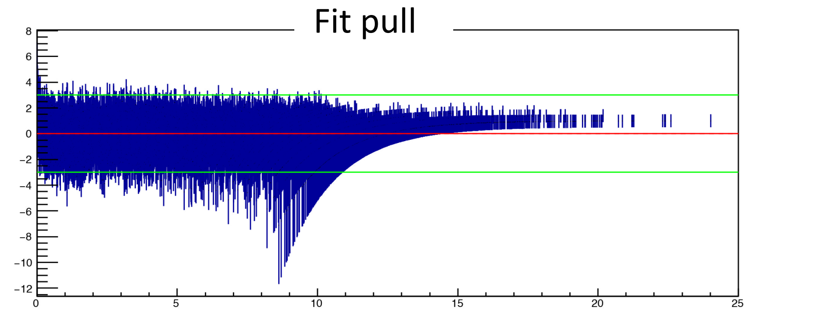
- Let us fix the m & Γ parameters, (to the CMS estimates from the fit to data) while leaving μ free in our ML fits (μ is not properly a signal yield).

By fitting our likelihood ratio distrib. we indeed get

$$\text{d.o.f.} \approx 1.014 \pm 0.001$$

$$\chi^2_{norm} = 1.009 \quad P(fit) = 0.118$$





$-2\ln\lambda$

[*] S.S.Wilks, *Ann.Math.Stat.* 9 (1938) 60-62

Special case : asymptotic formula by Cowan et al. [*] holds

- Consider the special case of the test statistic t_μ with the purpose to test $\mu = 0$ in a class of model where we assume $\mu \geq 0$. Rejecting $\mu = 0$ (the null hypothesis) leads to the discovery of a new signal.

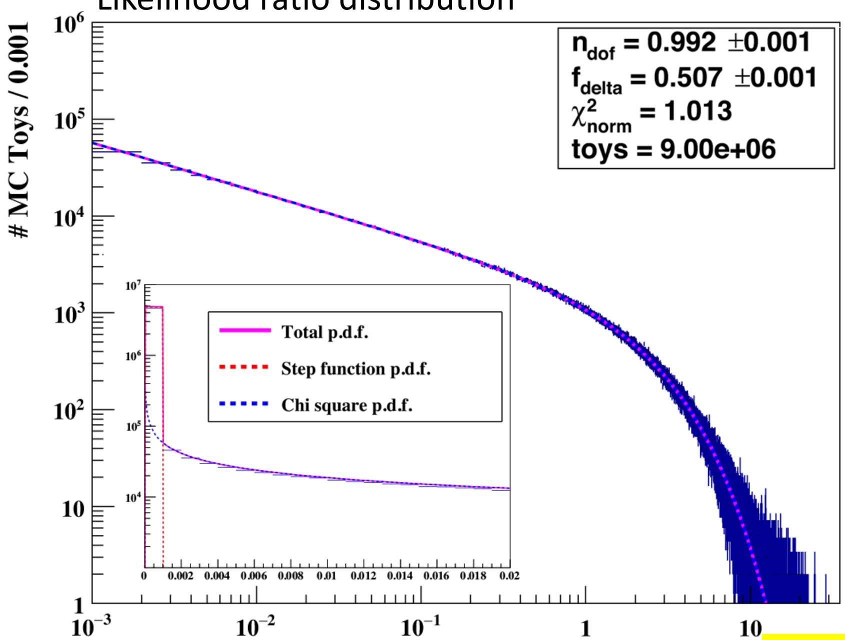
In this case following Cowan *et al*. the test statistic is :
$$q_0 = \begin{cases} -2\ln\lambda(0) \\ 0 \end{cases} \text{with} \begin{cases} \hat{\mu} \geq 0 \\ \hat{\mu} < 0 \end{cases}$$

Cowan *et al*. derive analitically that the PDF of q_0 is an equal mixture of a delta function at 0 & a chi-square distribution for 1 d.o.f. :
$$g(q_0|\mu=0) = \frac{1}{2}\boxed{\delta(q_0)} + \frac{1}{2}\boxed{\left[\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}e^{-q_0/2}\right]}$$

- Let us fix the m & Γ parameters (to the CMS estimates from fit to data) while constraining $\mu \geq 0$ in our ML fits (μ represents a signal yield here).

By fitting our likelihood ratio distrib. we indeed get :

$$\text{d.o.f.} \approx 0.992 \pm 0.001$$
$$\text{weight } C_{\chi^2} \approx 0.507 \pm 0.01$$

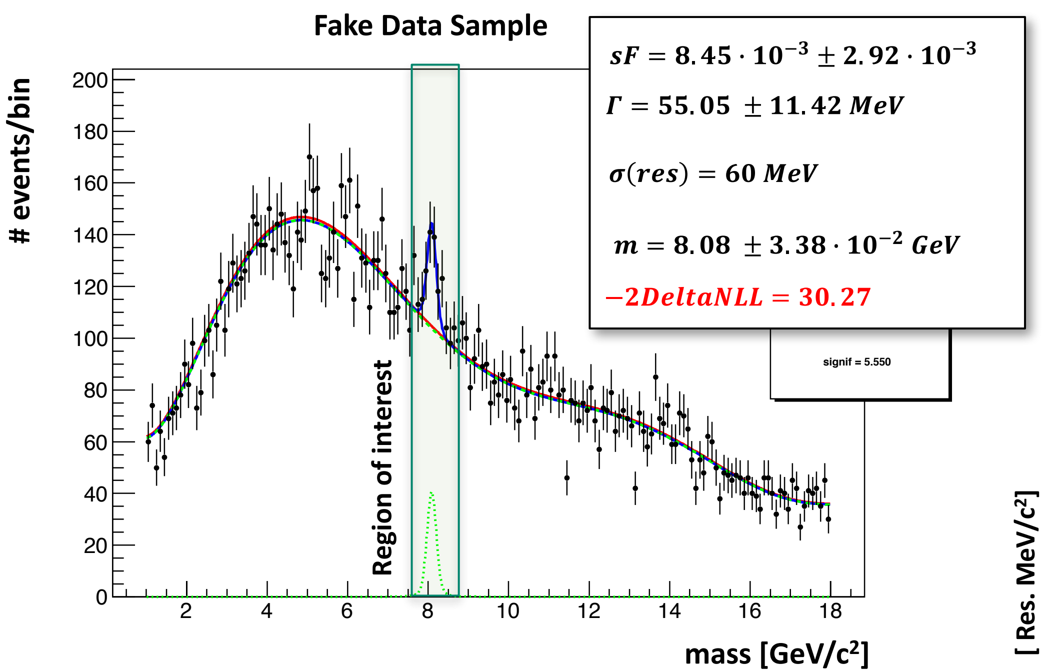


[*] Cowan *et al.*, EPJ C71 (2011) 1554

Global significance estimation for a new signal

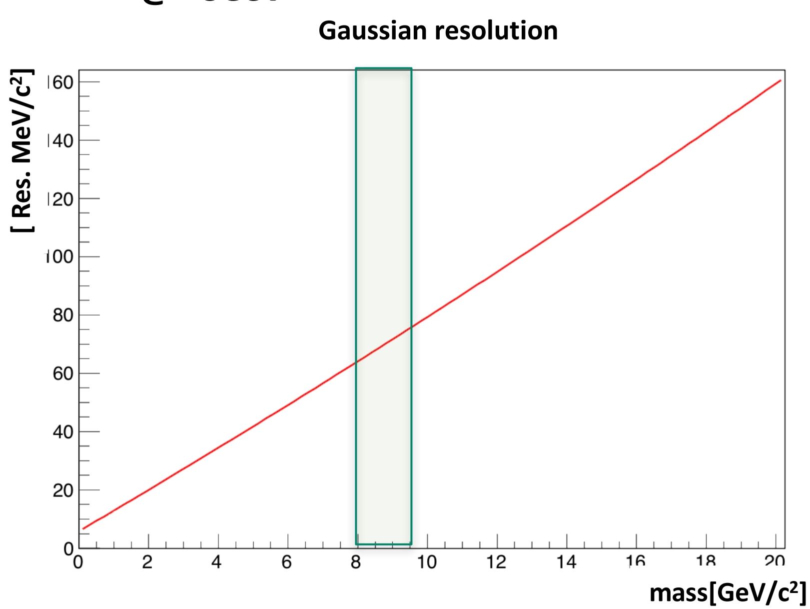
Global significance estimation for a new signal

When dealing with an unexpected new signal, a *global statistical significance* must be estimated and the Look-Elsewhere-Effect (LEE) must be taken into account. This implies to consider – within the same background-only fluctuation and everywhere in the relevant mass spectrum – any peaking behavior with respect to the expected background model and then a scanning technique must be implemented.



In order to test the effects of the LEE we generated a **pseudo-data inv. mass distribution** of 15K candidates in a generic region of interest (1-18GeV)

- ***Background* model :** 7^{th} order polynomial on
- ***Signal model:*** convolution of a B.W. and a Gaussian (resolution) p.d.f.s, **artificially added @ ~8GeV**



From the approximation:

$$Z \simeq \sqrt{-2[ln(L_{H_1}) - ln(L_{H_0})]}$$

$$\downarrow$$

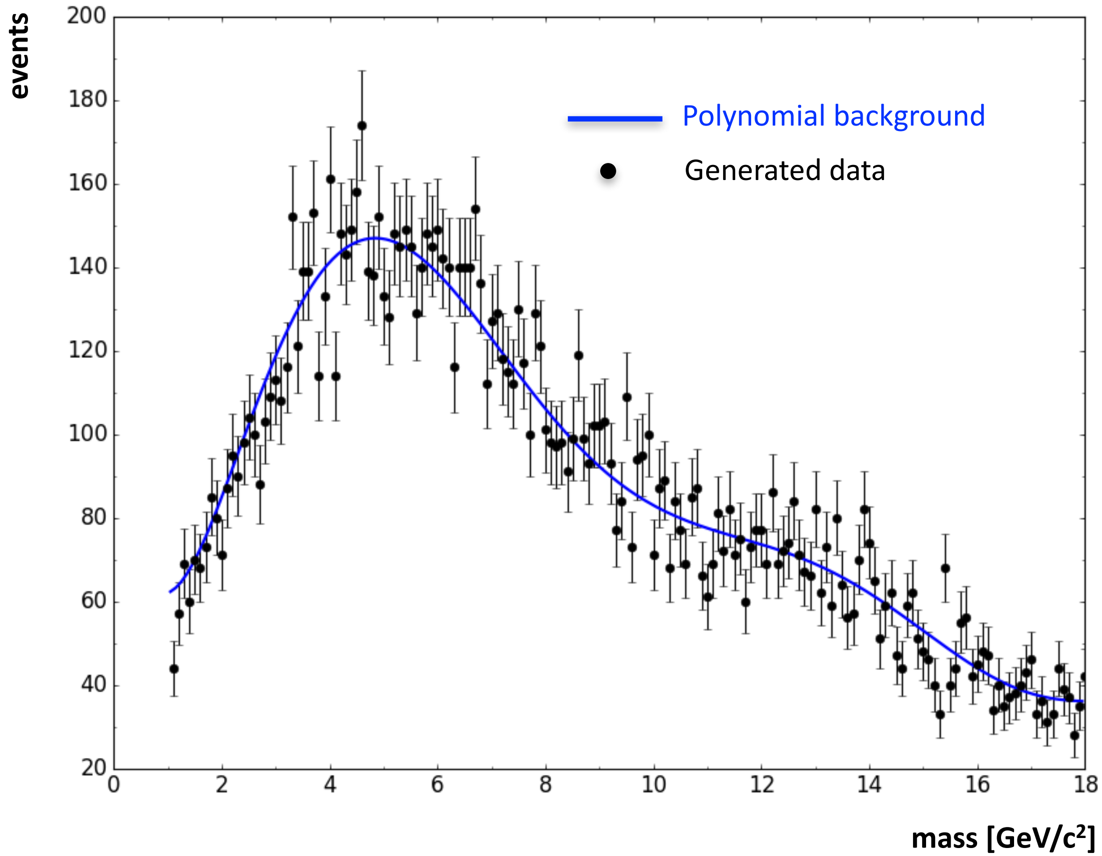
Local signficance ~ 5.50

Scanning technique: clustering approach

The **scanning technique** has been configured on the basis of a clustering approach and has been designed in advance with the aim to satisfy two concurrent requirements:

A) Do not miss any interesting fluctuation

B) Do not select too many small fluctuations



The procedure:

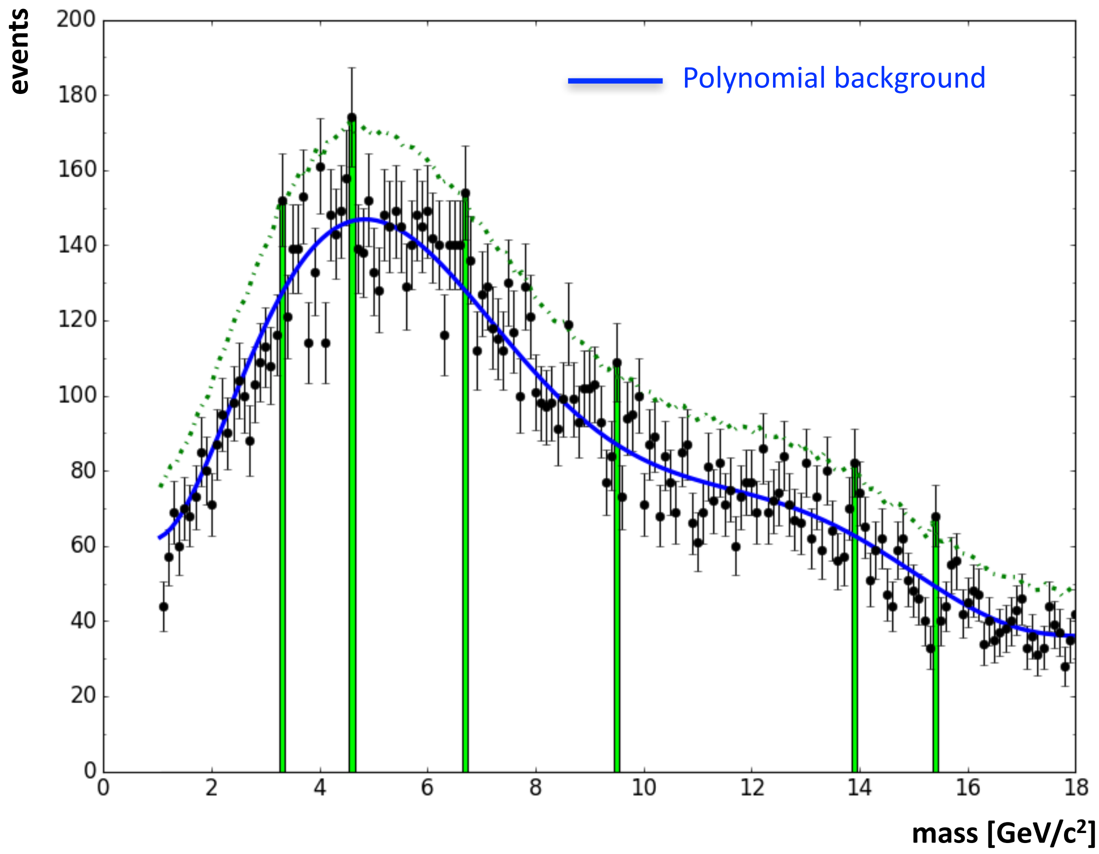
1. For **each MC Toy iteration** a distribution based on the **background p.d.f.** model is generated.

2. The ***H0 Null Hypothesis*** fit is performed with the background function only.

The **scanning technique** has been configured on the basis of a clustering approach and has been designed in advance with the aim to satisfy two concurrent requirements:

A) Do not miss any interesting fluctuation

B) Do not select too many small fluctuations



The procedure:

1. For **each MC Toy iteration** a distribution based on the **background p.d.f.** model is generated.

2. The ***H0 Null Hypothesis*** fit is performed with the background function only.

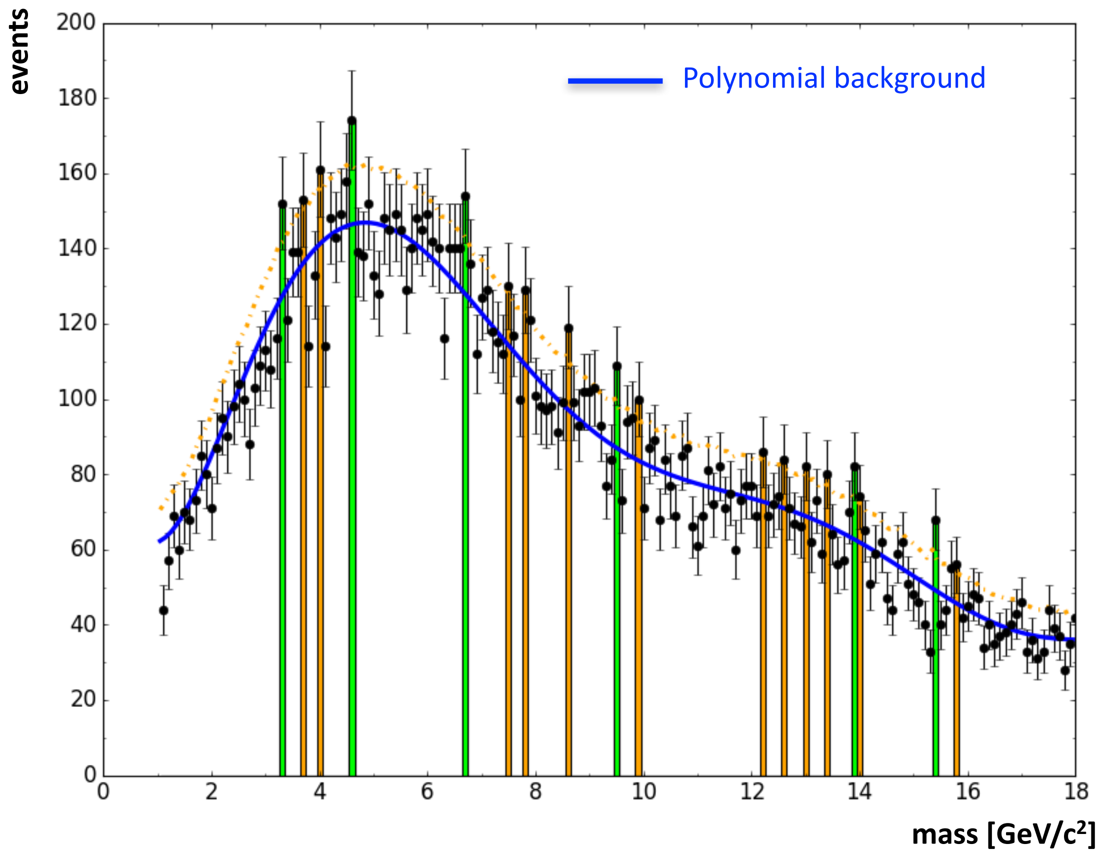
3. A first scan is performed to search for a **main seed** defined as a bin whose content fluctuates more than $x\sigma$ strictly above the value of the background function.

Scanning technique: clustering approach

The **scanning technique** has been configured on the basis of a clustering approach and has been designed in advance with the aim to satisfy two concurrent requirements:

A) Do not miss any interesting fluctuation

B) Do not select too many small fluctuations



The procedure:

1. For **each MC Toy iteration** a distribution based on the **background p.d.f.** model is generated.

2. The ***H0 Null Hypothesis*** fit is performed with the background function only.

3. A first scan is performed to search for a **main seed** defined as a bin whose content fluctuates more than $x\sigma$ strictly above the value of the background function.

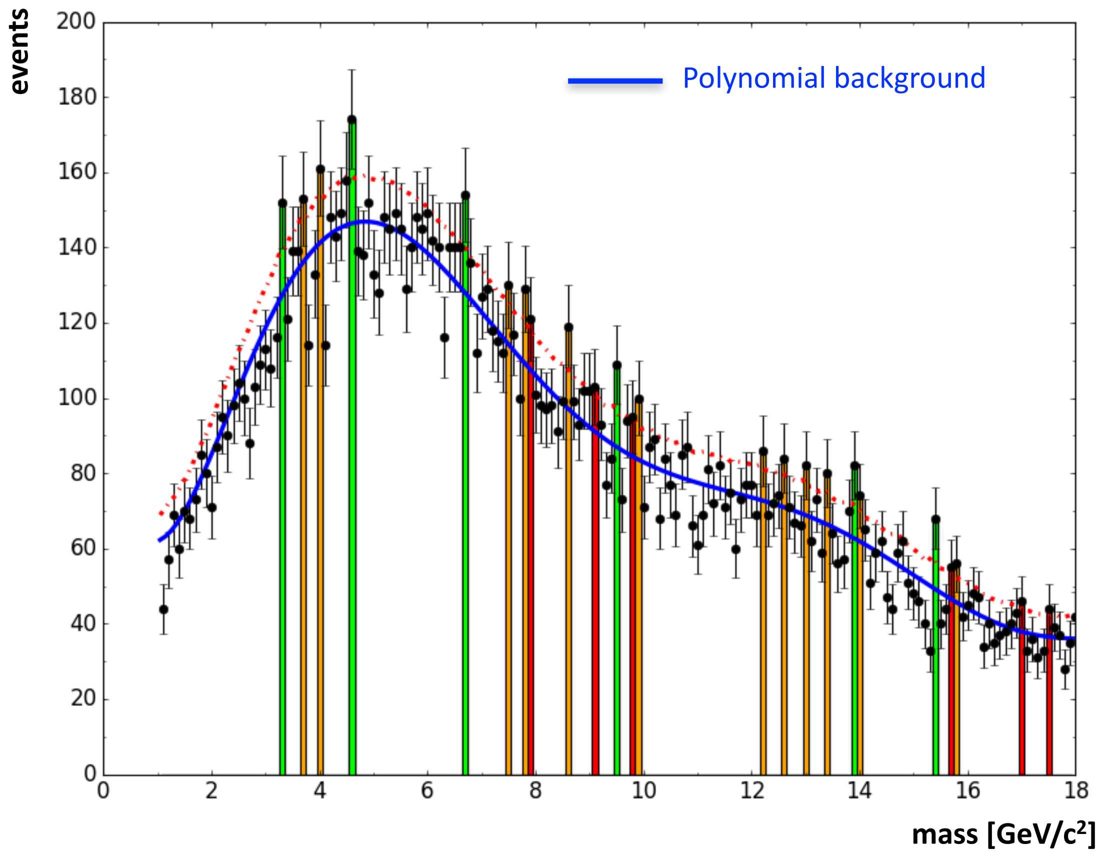
4. A second scan is performed to search for a **light seeds** defined as a bin whose content fluctuates more than $y\sigma$ $(y<x)$ strictly above the value of the background function.

Scanning technique: clustering approach

The **scanning technique** has been configured on the basis of a clustering approach and has been designed in advance with the aim to satisfy two concurrent requirements:

A) Do not miss any interesting fluctuation

B) Do not select too many small fluctuations



The procedure:

5. A final scan is performed to search for a **side seeds** defined as a bin whose content fluctuates more than $z\sigma$ **(z<y<x)** strictly above the value of the background function.

6. The final step consists of cleaning up the seeds.
   - **All** the **main** (x) **seeds** are reained.
   - The **light** (y) **seeds** are kept only if **at least one** of the **side** bins **is a seed** (of any kind).
   - The **side** (z) **seeds** are kept only if **at least one** of the **side** bins is a **main** or **light seed**.

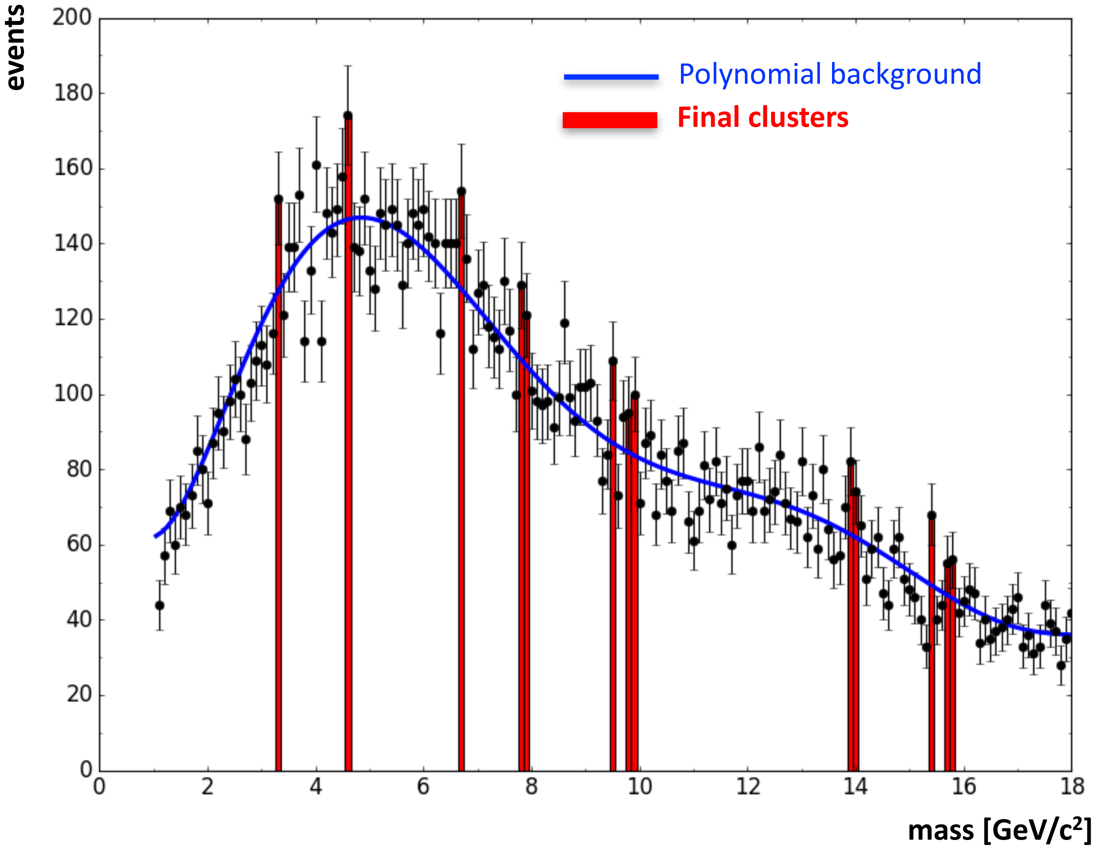
7. **The clusters are then formed**

Scanning technique: clustering approach

The **scanning technique** has been configured on the basis of a clustering approach and has been designed in advance with the aim to satisfy two concurrent requirements:

A) Do not miss any interesting fluctuation

B) Do not select too many small fluctuations



The procedure:

5. A final scan is performed to search for a **side seeds** defined as a bin whose content fluctuates more than $z\sigma$ **(z<y<x)** strictly above the value of the background function.

6. The final step consists of cleaning up the seeds.
   - **All** the **main** (x) **seeds** are reained.
   - The **light** (y) **seeds** are kept only if **at least one** of the **side** bins **is a seed** (of any kind).
   - The **side** (z) **seeds** are kept only if **at least one** of the **side** bins is a **main** or **light seed**.

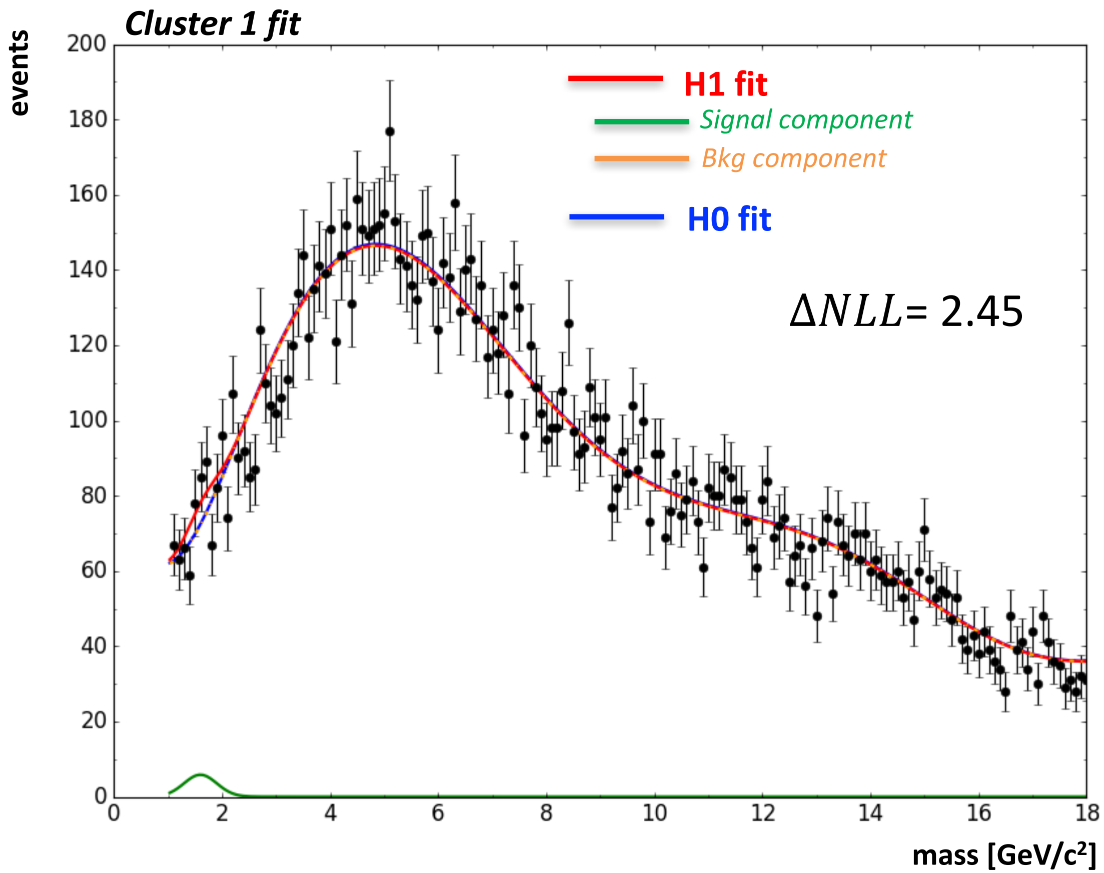
7. The clusters are then formed

Scanning technique: clustering approach

The **scanning technique** has been configured on the basis of a clustering approach and has been designed in advance with the aim to satisfy two concurrent requirements:

A) Do not miss any interesting fluctuation

B) Do not select too many small fluctuations



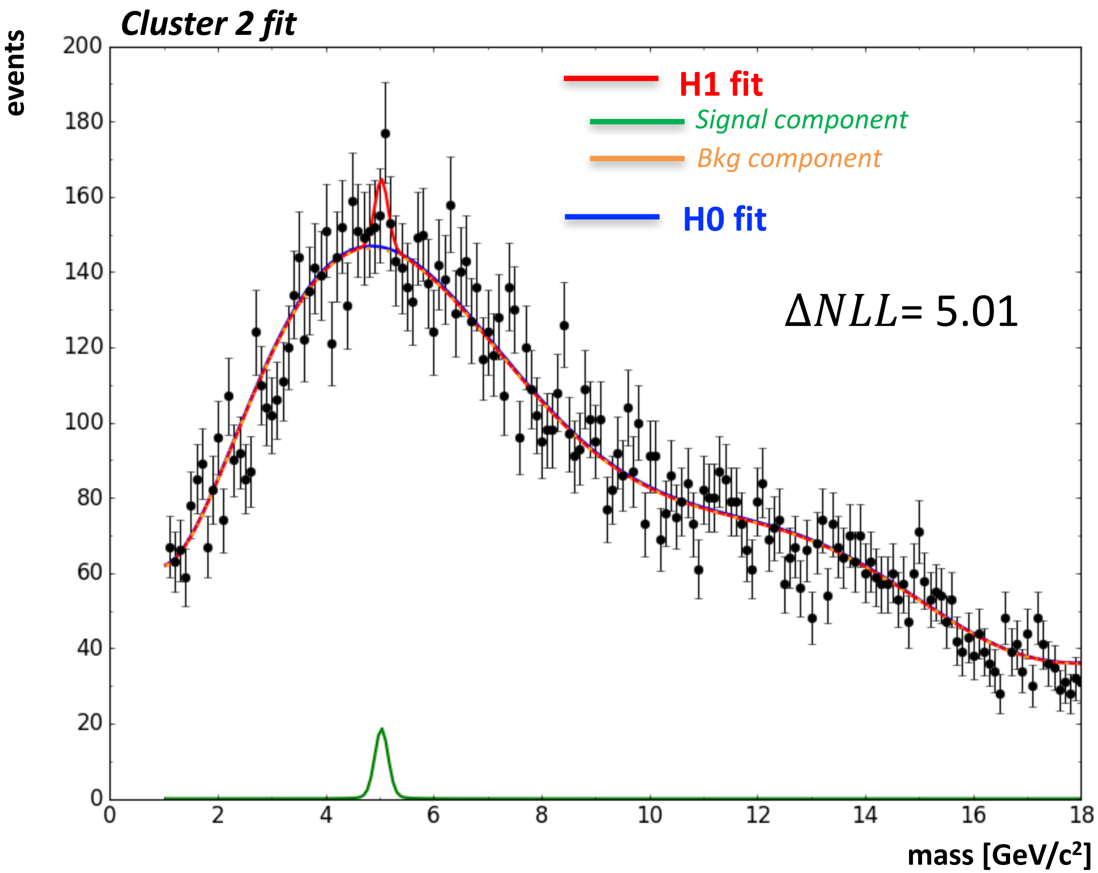
The procedure:

8. For **each cluster,** the **Alternative Hypothesis H1** fits are performed with the *polynomial H0-function* + a *Convolution* of a B.W. (signal) and a Gaussian (resolution) for the peak. For each seed a set of fits is performed **changing the parameters'** (*m* , *Γ* , ***σ***) range and starting values:
   - mass **m values** are changed scanning the **whole cluster**;
   - width *Γ* values are changed from 1 MeV to the whole cluster width [anyway always limited to 0.3 GeV] ;
   - resolution ***σ*** values is varied as a **function of the resonance** mass;

The **scanning technique** has been configured on the basis of a clustering approach and has been designed in advance with the aim to satisfy two concurrent requirements:

A) Do not miss any interesting fluctuation

B) Do not select too many small fluctuations



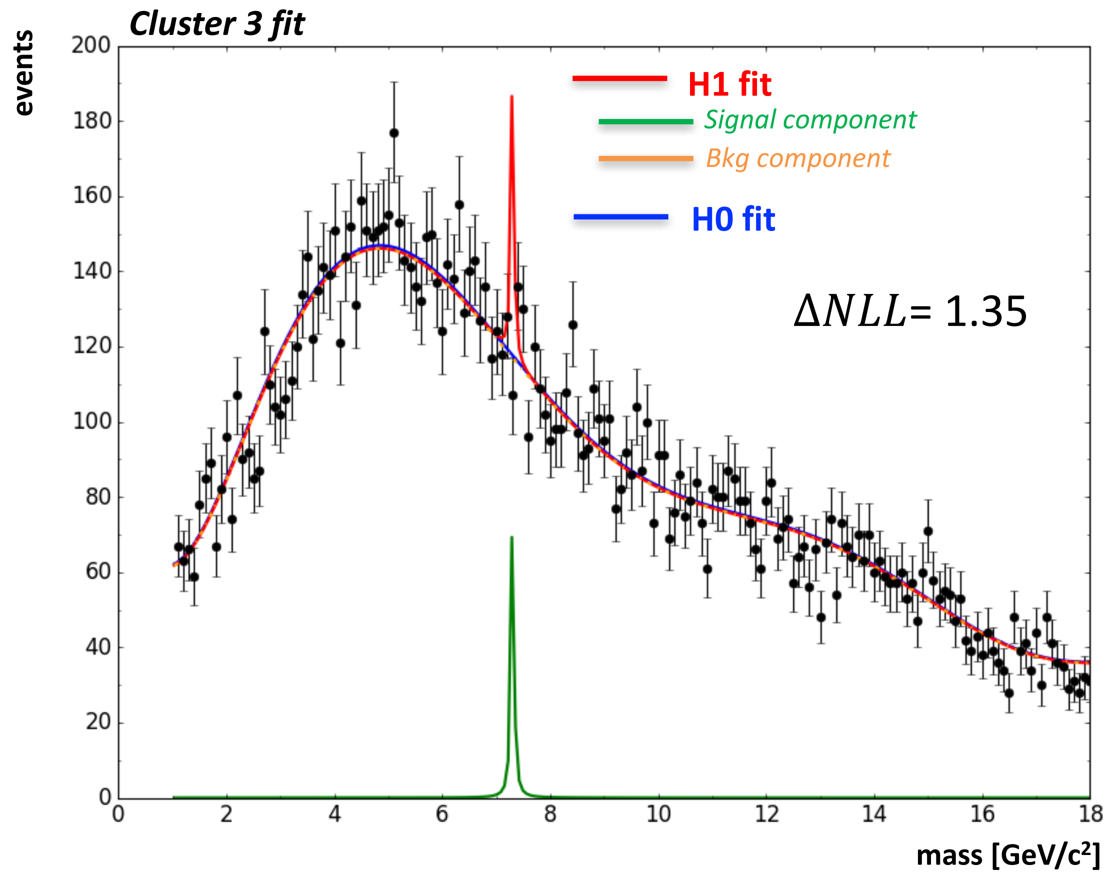
The procedure:

8. For **each cluster,** the **Alternative Hypothesis H1** fits are performed with the *polynomial H0-function* + a *Convolution* of a B.W. (signal) and a Gaussian (resolution) for the peak. For each seed a set of fits is performed **changing the parameters'** (m , Γ , σ) range and starting values:
   - mass **m values** are changed scanning the **whole cluster**;
   - width Γ values are changed from 1 MeV to the whole cluster width [anyway always limited to 0.3 GeV] ;
   - resolution σ values is varied as a **function of the resonance** mass;

The **scanning technique** has been configured on the basis of a clustering approach and has been designed in advance with the aim to satisfy two concurrent requirements:

A) Do not miss any interesting fluctuation

B) Do not select too many small fluctuations



The procedure:

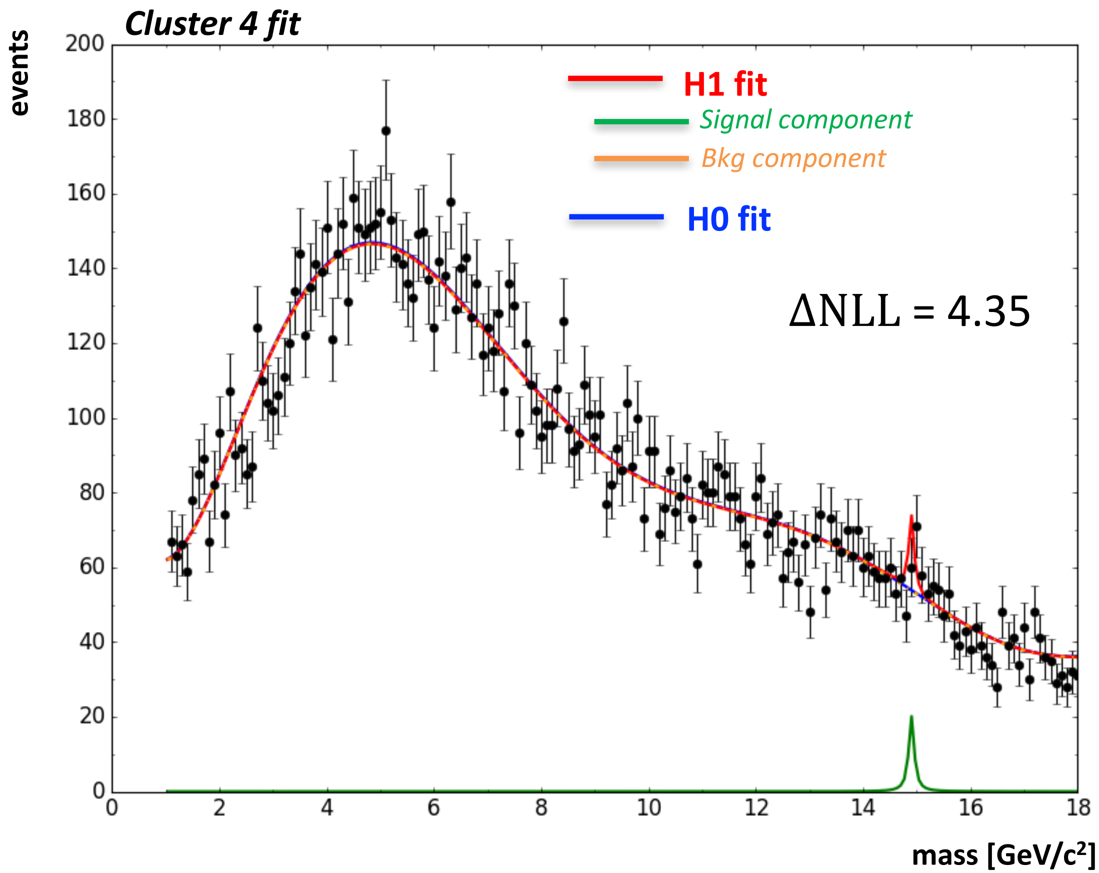
8. For **each cluster,** the **Alternative Hypothesis H1** fits are performed with the *polynomial H0-function* + a *Convolution* of a B.W. (signal) and a Gaussian (resolution) for the peak. For each seed a set of fits is performed **changing the parameters'** (m , Γ , $\boldsymbol{\sigma}$) range and starting values:
   - mass **m values** are changed scanning the **whole cluster**;
   - width Γ values are changed from 1 MeV to the whole cluster width [anyway always limited to 0.3 GeV] ;
   - resolution $\boldsymbol{\sigma}$ values is varied as a **function of the resonance** mass;

Scanning technique: clustering approach

The **scanning technique** has been configured on the basis of a clustering approach and has been designed in advance with the aim to satisfy two concurrent requirements:

A) Do not miss any interesting fluctuation

B) Do not select too many small fluctuations



The procedure:

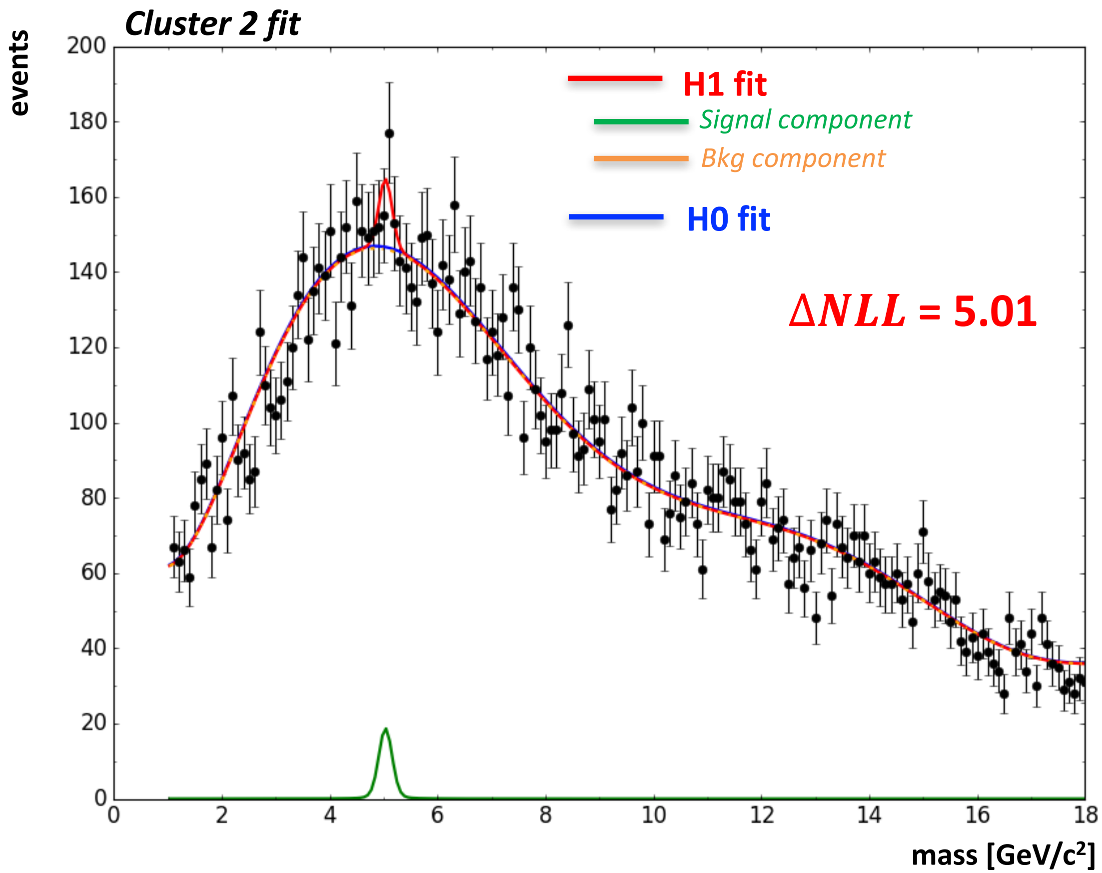
8. For **each cluster,** the **Alternative Hypothesis H1** fits are performed with the *polynomial H0-function* + a *Convolution* of a B.W. (signal) and a Gaussian (resolution) for the peak. For each seed a set of fits is performed **changing the parameters'** (*m* , *Γ* , ***σ***) range and starting values:
   - mass **m values** are changed scanning the **whole cluster**;
   - width *Γ* values are changed from 1 MeV to the whole cluster width [anyway always limited to 0.3 GeV] ;
   - resolution ***σ*** values is varied as a **function of the resonance** mass;

Scanning technique: clustering approach

The **scanning technique** has been configured on the basis of a clustering approach and has been designed in advance with the aim to satisfy two concurrent requirements:

A) Do not miss any interesting fluctuation

B) Do not select too many small fluctuations



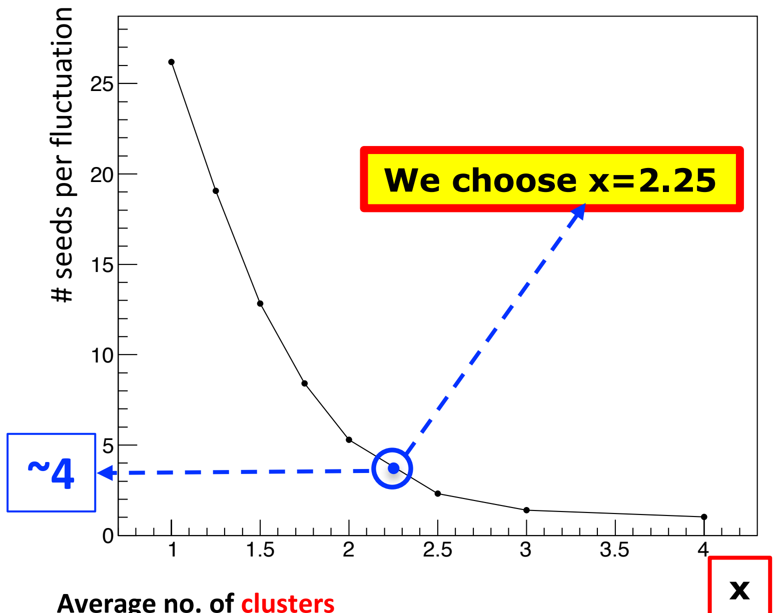
The procedure:

8. For **each cluster,** the **Alternative Hypothesis H1** fits are performed with the *polynomial H0-function* + a *Convolution* of a B.W. (signal) and a Gaussian (resolution) for the peak. For each seed a set of fits is performed **changing the parameters'** (m , Γ , σ) range and starting values:
   - mass **m values** are changed scanning the **whole cluster**;
   - width Γ values are changed from 1 MeV to the whole cluster width [anyway always limited to 0.3 GeV] ;
   - resolution σ values is varied as a **function of the resonance** mass;
8. The best ΔNLL is registered to build the test statistic distribution

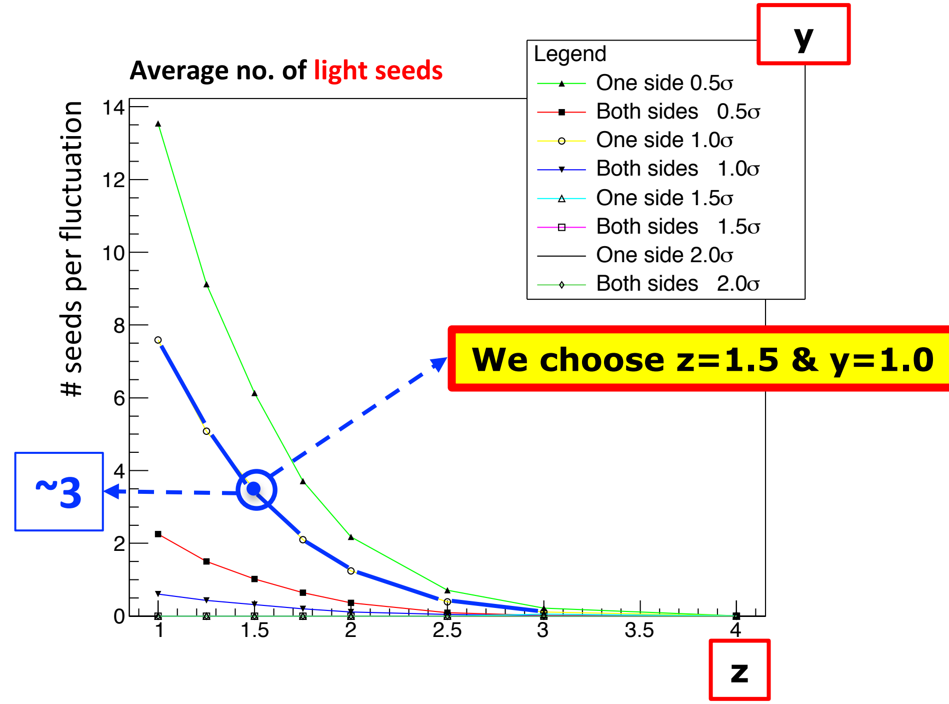
Scanning technique: cuts tuning

Once defined the scanning technique, the next step is to tune the procedure parameters **x (main seed threshold), y (light seed threshold)** and **z (sided seed threshold)** in order to fullfill the requirements [A,B]. A set of **1M** toys were produced to count the mean value of the distribution **of the number of main and light seeds** per single fluctuation.

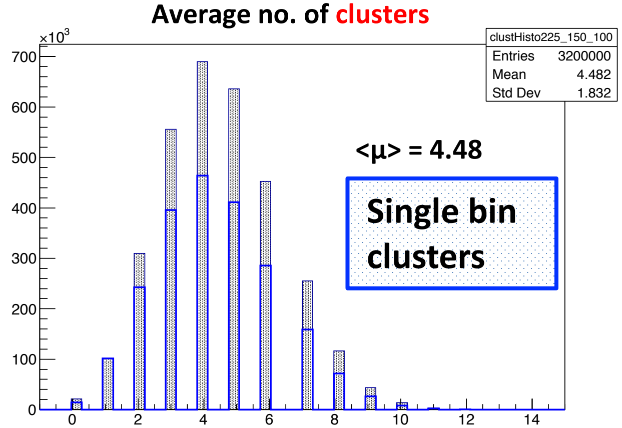
Average no. of main seeds



Average no. of light seeds



Average no. of clusters

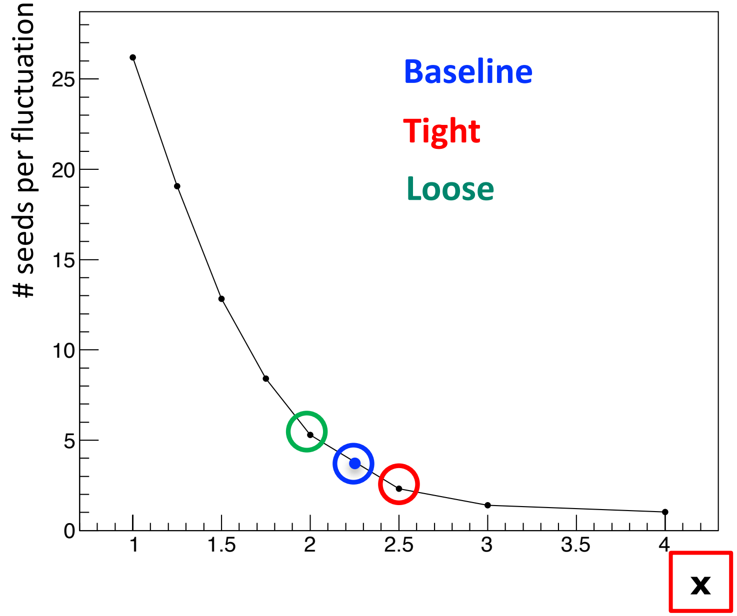


These cuts assure us to build **~4-5 clusters** in average for each Toy MC distribution and that at least 1 cluster is found in more than 99% of fluctuations in order to perform at least 1 *Alternative Hypothesis* fit.

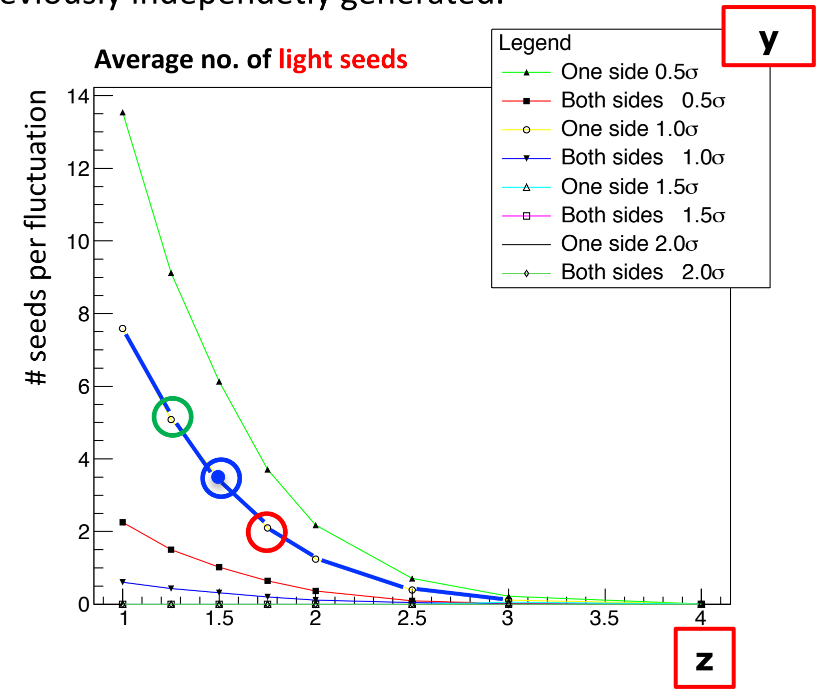
Scanning technique: systematic uncertainty

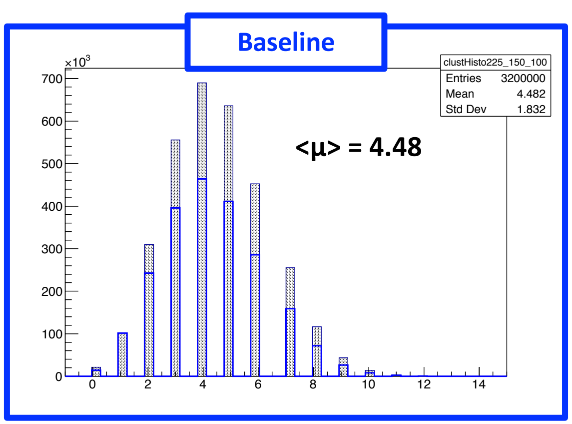
In order to study the possible **systematic uncertainties** of this method to the estimation of a global significance we have ***selected*** also two other combinations of (x,y,z). One ***looser*** than the selecte one and one ***tighter***. In addition, to avoid any possible influence of statistical fluctuations, we have run the MC Toys fitting procedure **three times** for the three different cuts on ***the same* set of MC toys fluctuations**, that have been previously independetly generated.

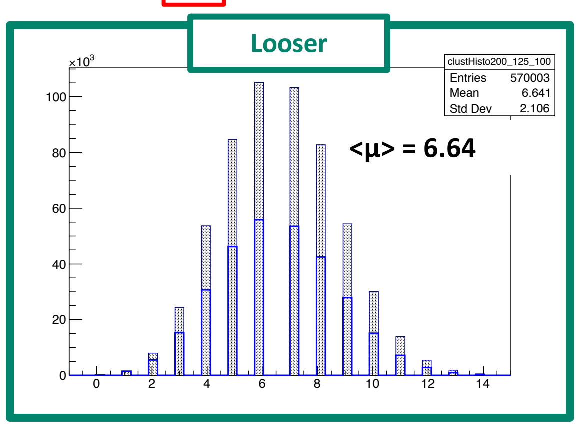
Average no. of main seeds

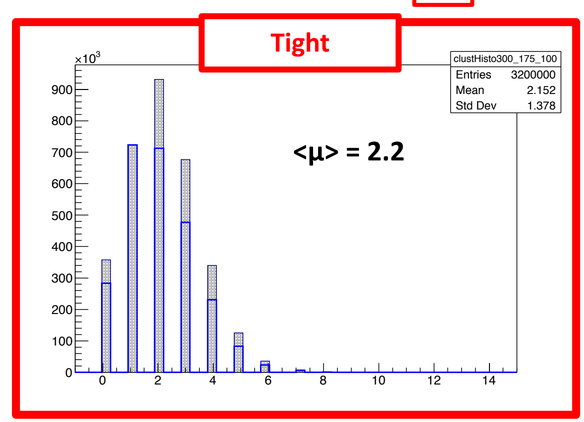


Average no. of light seeds



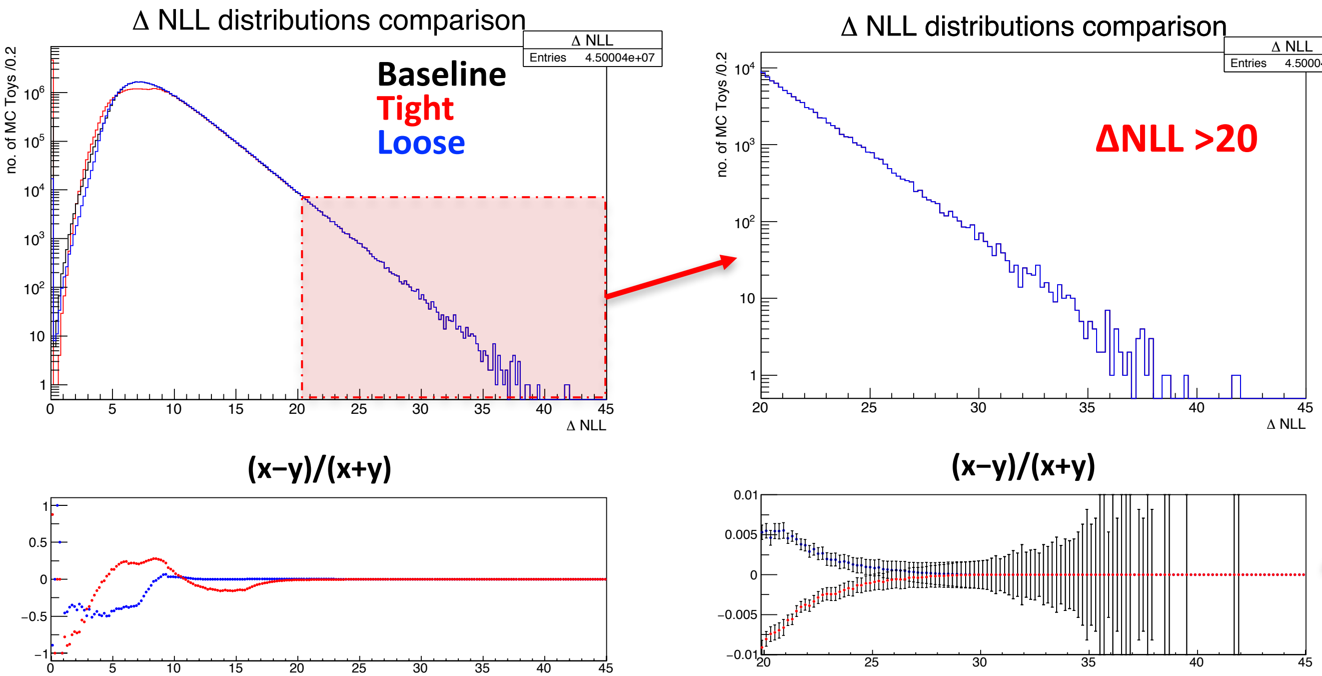






Scanning technique: comparison results

- The resulting distributions from 45M common MC Toys fluctuations are shown superimposed and compared. By focusing on the **region of interest** for the estimation of the statistical significance, **i.e. the tail of the ΔNLL distribution (ΔNLL >20)**, it is evident that there is no relevant difference among the three configurations.



This can furtherly be appreciated by inspecting the normalized deviations **(x−y)/(x+y)** of the other two distributions with respect to the baseline distribution

- Also we can examine the estimated global significances for the **p-values** corresponding to **different values of local significances**

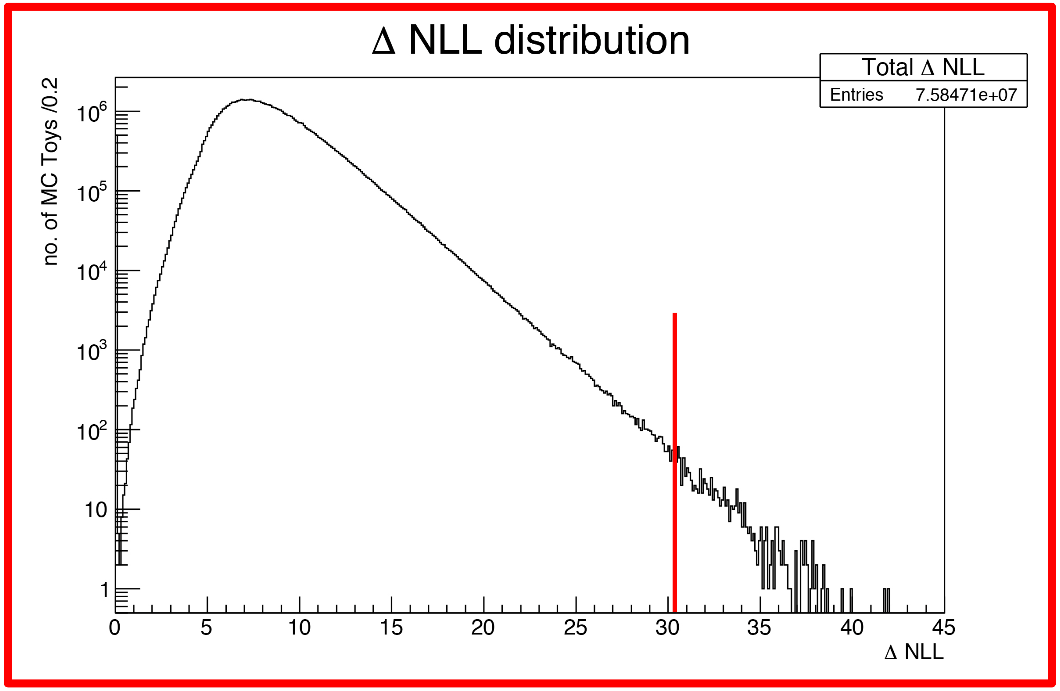
Clustering configs.	$< fit_{H1} >$	f_{nofit}
Tight (3.00, 1.75, 1.00)	2.2	~10%
Baseline (2.25,1.50, 1.00)	4.5	~1%
Loose (2.00, 1.25, 1.00)	6.6	0.1%

Local Significance	**4.0σ**	**4.5σ**	**5.0σ**	**5.5σ**	**6.0σ**
Tight (3.00, 1.75, 1.00)	2.21	2.91	3.58	4.23	5.19
Baseline (2.25,1.50, 1.00)	2.20	2.91	3.58	4.23	5.19
Loose (2.00, 1.25, 1.00)	2.19	2.92	3.58	4.23	5.19

It can be concluded that the systematic uncertainty on the p-values associated to the method is negligible.

Final test statistic distribution

- The **baseline configuration** has been run on about **76M** pseudo experiments and the **ΔNLL** distribution is shown with the superimposed **red line** indicating the ΔNLL data value for the **original pseudo-data.**



- The **global p-value** is then estimated by

$$p = \int_{\Delta NLL_{data}}^{\infty} f(\Delta NLL) d(\Delta NLL) \simeq \frac{9.820 \cdot 10^2}{7.584 \cdot 10^7} \simeq 1.295 \cdot 10^{-5}$$

Which corresponds to a **global statistical significance** of

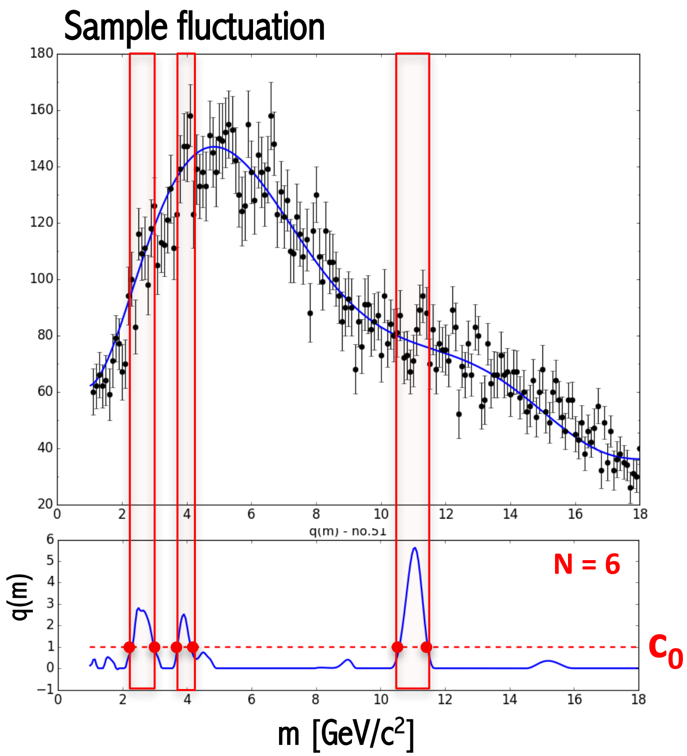
$$\boxed{Z\sigma \ = \ \Phi^{-1}(1-p)\sigma \ \simeq \ 4.22\sigma}$$

Comparison with asymptotic limit by Gross & Vittels

Comparison with asymptotic limit by Gross & Vittels

- In their 2010 paper **[*]**, *E. Gross and O. Vittels*, proposed **(among other results)** a method to estimate an **upper limit** for the **global p-value** when the ***signal hypothesis (H1)*** depends on **one or more** [nuisance] parameters ($\vec{\boldsymbol{\theta}}$) that don't exist under the ***null hypothesis (H0)***. In our case $\vec{\boldsymbol{\theta}} = (\boldsymbol{m}; \boldsymbol{\Gamma})$ and we denote as $q(\vec{\theta})$ the ΔNLL test statistics.
We are interested in the maximum of $q(\vec{\theta})$ over θ, $\boldsymbol{q(\hat{\theta}) = \max_{\vec{\theta}} q(\vec{\theta})}$.

- The **G-V method** relies on the estimation of the **average number of *upcrossings*** $< N(c) >$ of $q(\vec{\theta})$, spanning along the $\vec{\theta}$ parameter space, w.r.t. to a desired threshold **c** for the test statistics (in our case the ΔNLL_{data} **):**



$$P\left(q(\hat{\theta}) > c\right) \leq \underbrace{P\left(\chi_s^2 > c\right)}_{\text{Wilks' local significance}} + \underbrace{\left\langle N(c) \right\rangle}_{\text{average number of upcrossings}}$$

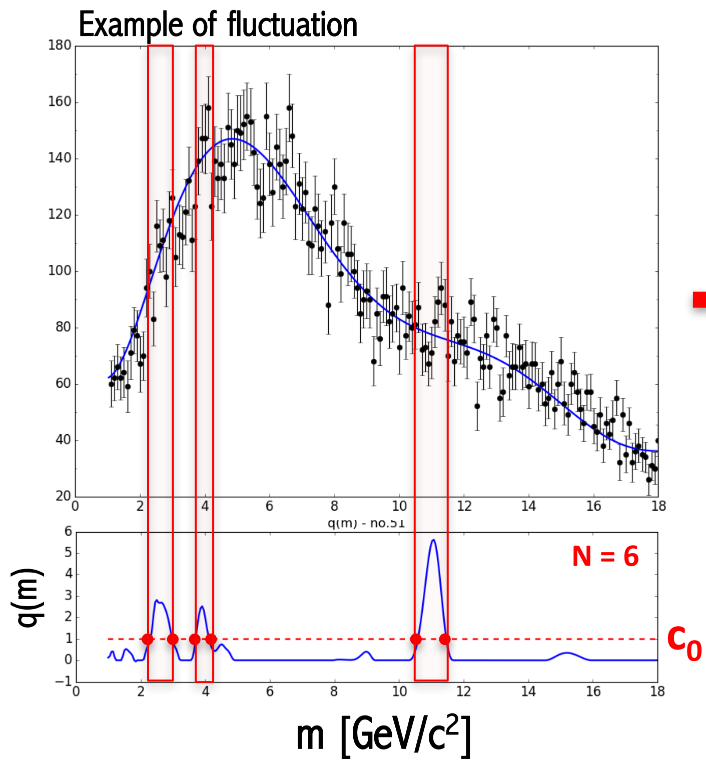
The $\boldsymbol{N(c)}$ function depends specifically on the details of the statistical model and can be difficult to calculate it analytically. In the paper, it is instead proposed to estimate **the number of upcrossings** $< N(c_0) >$ w.r.t. a **reference level *C_0=S-1*** with **S number of nuisance parameters** in a small set of background only MC toys:

$$P\left(q(\hat{\theta}) > c\right) \leq P\left(\chi_s^2 > c\right) + \left\langle N(c_0) \right\rangle \left(\frac{c}{c_0}\right)^{(s-1)/2} e^{-(c-c_0)/2} \qquad [1]$$

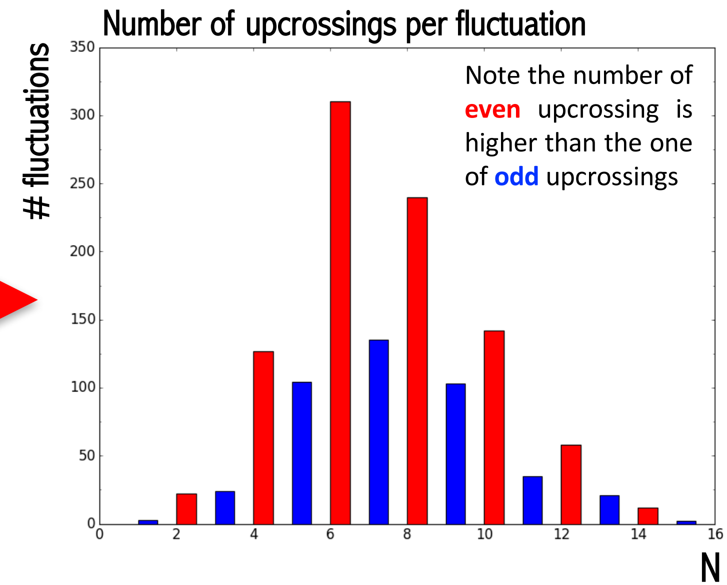
In our case the **reference level *C_0=S-1=1*** with **S=2, number of nuisance parameters**

[*] Eur. Phys. J. C (2010) 70: 525–530

- We set up a procedure [within **GooFit** framework] to estimate $< N(c_0) >$ for **our pseudo-data configuration.** **10k** toys are produced and for each toy a **complete scan** (in **1000** steps) of the mass spectrum is performed.



10k Toys



- The procedure took **~3days** on a single GPU, the time equivalent of **~4-5M MC toys** produced.

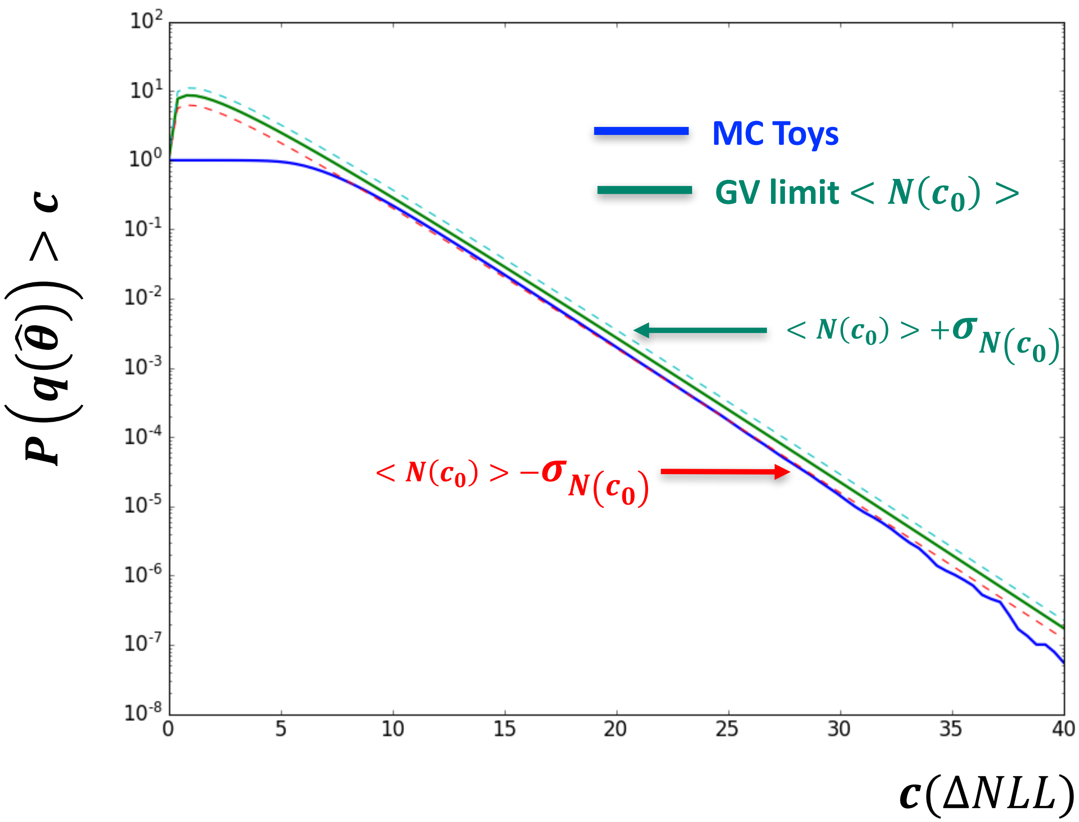
- From the **distribution** : $\boxed{< N(c_0) > = 7.3}$ $\sigma_{N(c_0)} = 2.4$ $\boxed{c_0\text{=s-1=1}}$

and the upper limit can be evaluated from:

$$P\left(q(\hat{\theta}) > c\right) \leq P\left(\chi_s^2 > c\right) + \boxed{\langle N(c_0) \rangle}\left(\frac{c}{\boxed{c_0}}\right)^{(s-1)/2} e^{-(c-\boxed{c_0})/2}$$

- Thus we can compare the $P\left(q(\hat{\theta})\right)$ computed from the ΔNLL distribution obtained with MC Toys (in the **baseline** configuration) with the upper limit just **estimated** with the **G-V method**.

  In the case of the MC Toys, $P\left(q(\hat{\theta})\right)(c)$ is calculated as the integral $P(q(\hat{\theta}))(c) = \int_c^{\infty} f(\Delta NLL)d(\Delta NLL)$



- As shown in the plot and in the table the **G-V** upper limit is **conservative** w.r.t the MC toys and, for a given ΔNLL value, always **underestimate** the global statistical significance:

Local Sig.	**4.0σ**	**4.5σ**	**5.0σ**	**5.5σ**	**6.0σ**
GV method	2.09	2.82	3.48	4.10	4.71
MC Toys	2.20	2.91	3.58	4.22	4.87

The limit is perfectly compatible with our results with the MC toys procedure

Summary

- With the advent of GPU computing the **pseudo experiment** approach is **feasible** and within the GooFit framework we built a tool to estimate the **global** (**local**) **p-value** of a signal within few days : **~1.5M** (**5M**) **toys** per day can be produced with a single GPU (TeslaK40) equipped machine [for **Z>5** **~3.5M toys** are needed]

- Also, thanks to the striking speed-ups, it was possible to **explore the validity of asimptotic** results **commonly used in HEP** (*when the regulaity conditions are met*):
  - **Cowan & Wilks'** : local significance
  - **Gross & Vittels method**: global significance.

- If you are interested to start **learning** & **working** with *GooFit*, it source code lives in a GitHub repository (https://github.com/GooFit) and its applications go **way further** than statistical significance estimation (for us in Bari it has become a "common" fitting tool particularly usefull when dealing with **multidimensional unbinned likelihood** fit at **high statistics**)

We are grateful for valuable support to all the people involved in the maitainance of the High Performance Cluster hosted by the ReCas Data Center, specifically to its manager Giacinto Donvito.

THANK YOU

"I am putting myself to the fullest possible use, which is all I think that any conscious entity can ever hope to do"

HAL9000

BACKUP

Wilks' theorem

[*] S.S.Wilks, *Ann.Math.Stat.* **9** (1938) 60-62

- The Wilks[*] theorem is often used to estimate the p-value associated to a new/unexpected signal :

  Given two hypotheses:
  - Null hypotheses H_0 with ν_0 d.o.f.
  - Alternative hypotheses H_1 with ν_1 d.o.f.

  … any test statistic t, defined as a likelihood ratio $-2\ln\lambda = -2\ln\left(\frac{L_{H_0}}{L_{H_1}}\right)$

  [or similarly (in the asymptotic limit) as a $\Delta\chi^2 = \chi^2_{H_0} - \chi^2_{H_1}$],

  approaches a χ^2 distribution with $\nu = \nu_1 - \nu_0$ d.o.f., provided that these regularity conditions hold :
  - H_0 and H_1 are nested (H_1 "includes" H_0)
  - while $H_1 \rightarrow H_0$ the H_1 parameters are well behaving (defined and not approaching some limit)
  - asymptotic limit (of a large data sample)

- Once this theorem holds, the p-value associated to the signal is given by : $P = \int_{t_{obs}}^{\infty} \chi^2_{\nu_1-\nu_0}(t)\,dt$

  The use of pseudo-experiments to estimate the p-value is not needed

  (but still suggested)

- When null hypothesis is background-only and the alternative is background+signal, often the above regularity conditions are not all satisfied, and MC toys are mandatory !