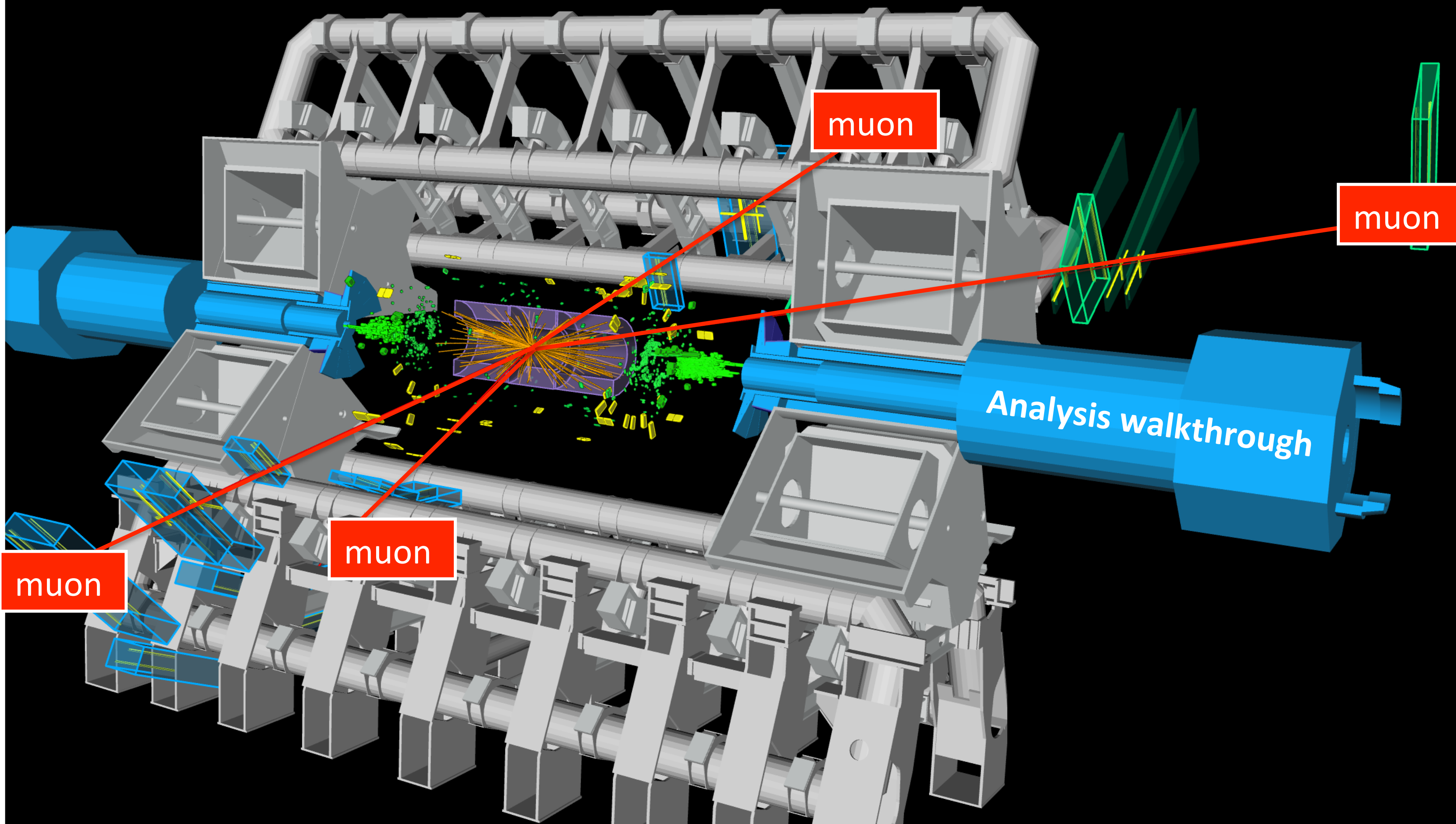
Introductory Terascale School, April 2015 (DESY)



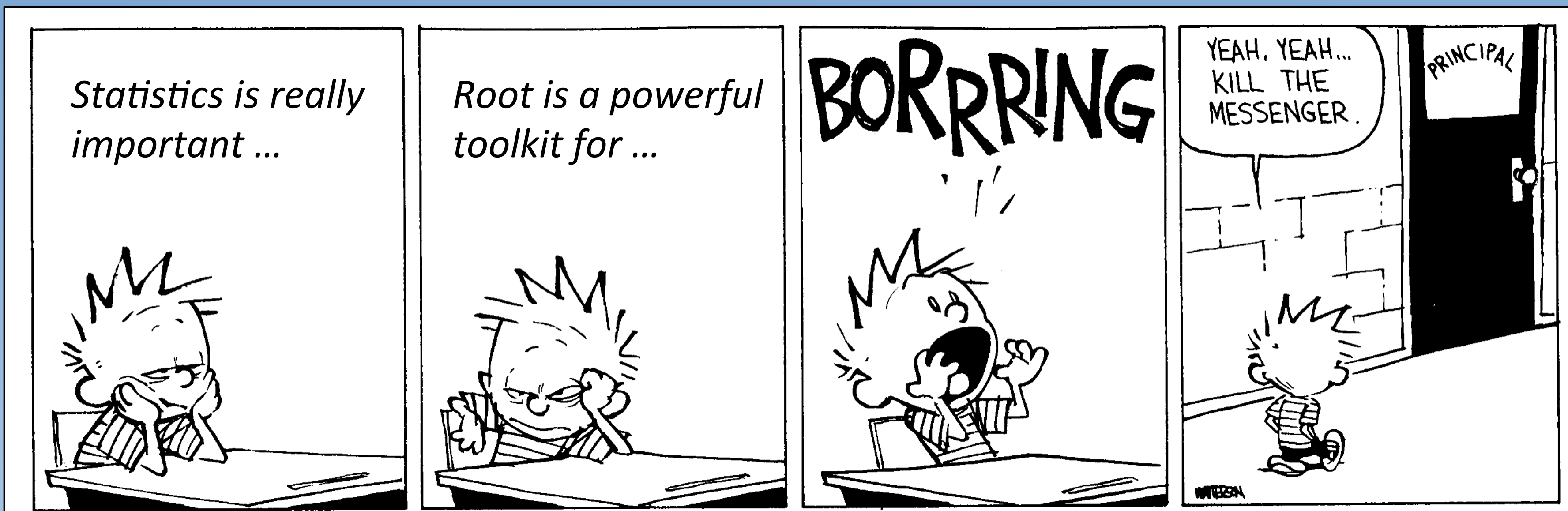
Statistics exercises

Tuesday 3.5 hours (14:00-15:30 & 16:00-18:00)

Wednesday 3.5 hours (11:00-12:30 & 14:00-16:00)

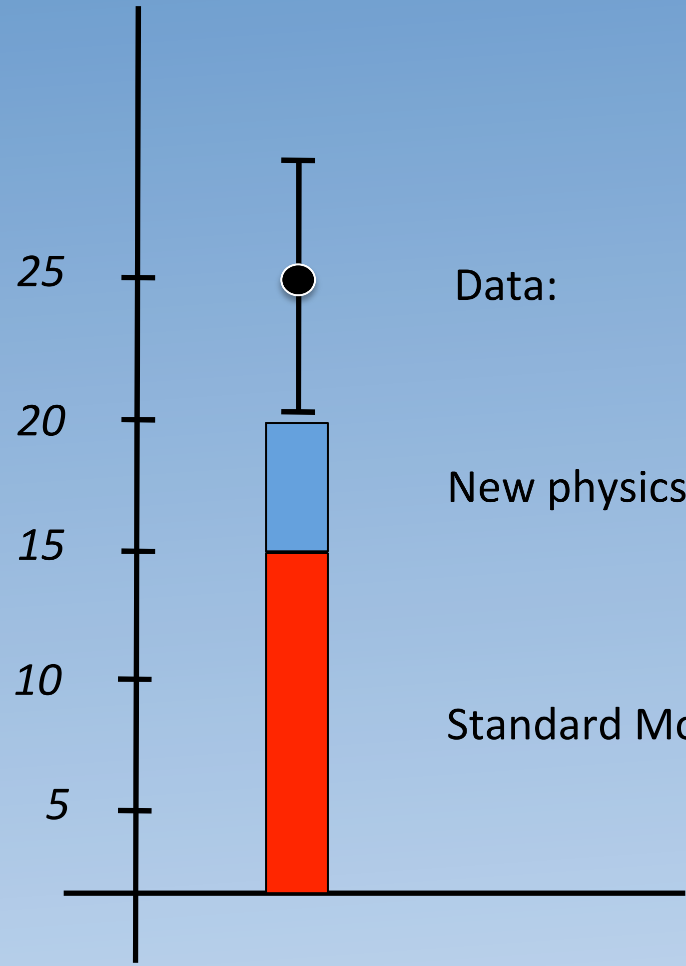
Ivo van Vulpen (UvA/Nikhef)

A short lecture on Root and statistics



Note: will be 95% of your work. Master this and you can focus on physics

Example: significance



Data: $N_{data} = 25$

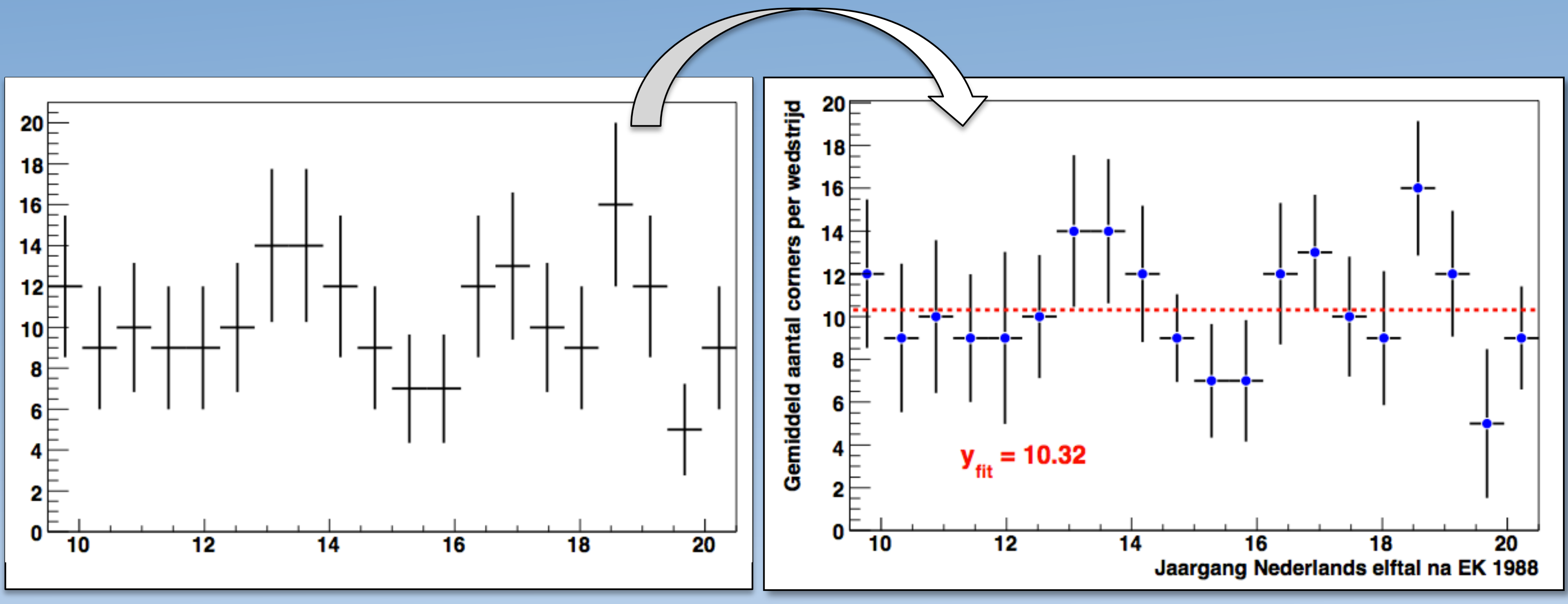
New physics: $N_{sig} = 5$

Standard Model: $N_{bgr} = 15$

What is the significance of the excess ?

Example: poisson statistics & likelihood fit

Can everybody do this ?



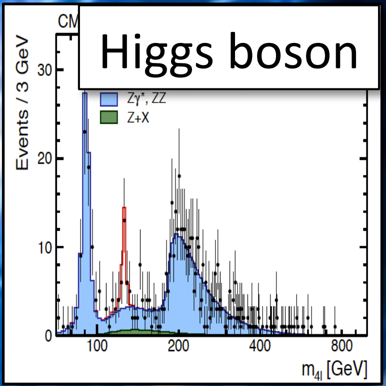
http://www.nikhef.nl/~ivov/SimpleFit/

Statistics is everywhere in science and industry

Risk analyses

Banking/consultancy

Weltmeister ?



Higgs boson

- Many mysteries, folklore, buzz-words, bluffing etc. but you **need** to master it to quantify the results of any analysis. Do **not** just follow 'what everybody else does' or your supervisor tells you.

- RooFit, Roostats, BAT, TMVA, BDT's are excellent and very powerful tools. Make sure you understand the basics so you know what you ask it to do.

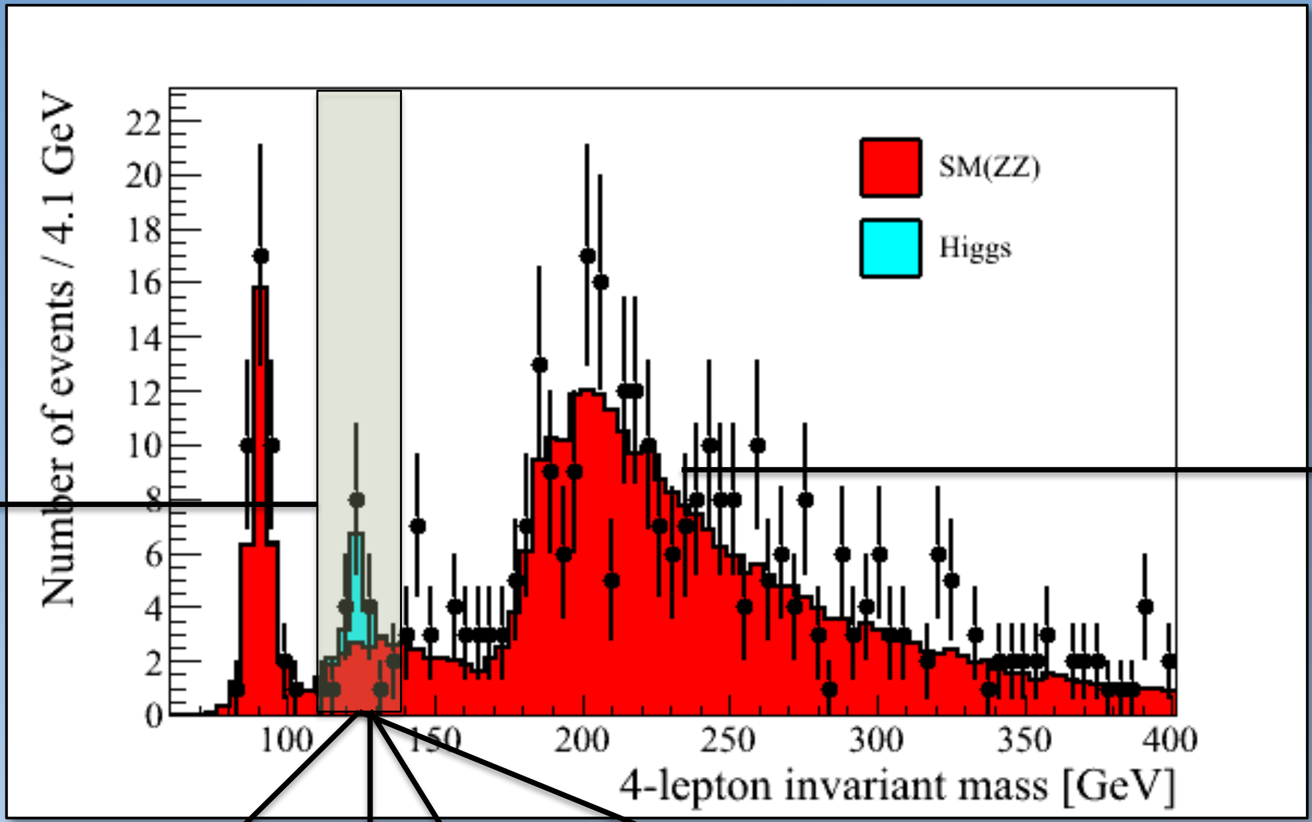
Goals of the hands-on exercises:

- manipulate histograms in root file
- understand and compute significances
- likelihood fit and parameter estimation (with error)

Extra: Convey message that you can and should do analysis
steps yourself first to understand the various concepts

Only then use power of RooFit and Roostats.

Data-set for the exercises: 4 lepton mass



Cross-section measurement

Exclusions

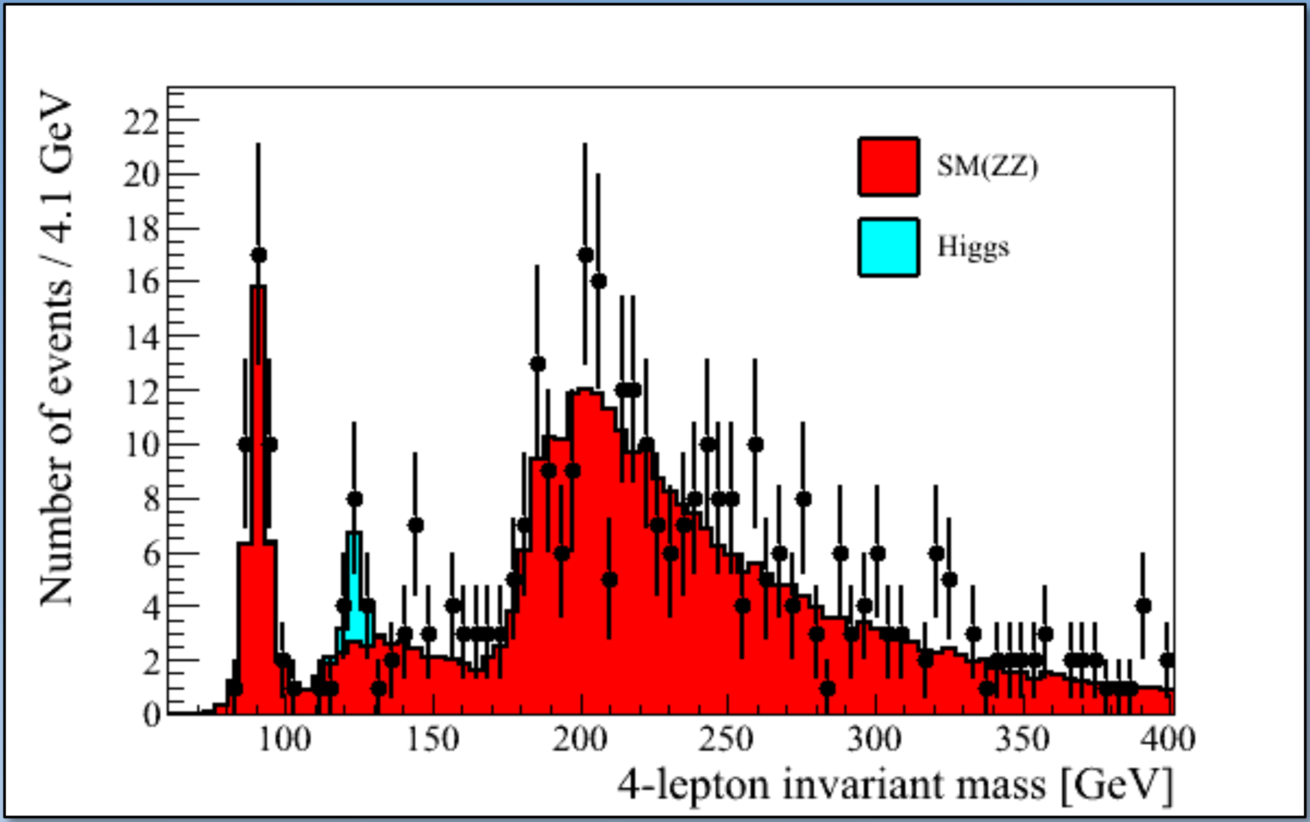
Significance optimization

Data-driven background estimate (likelihood fit using side bands)

Test statistic (Toy-MC)

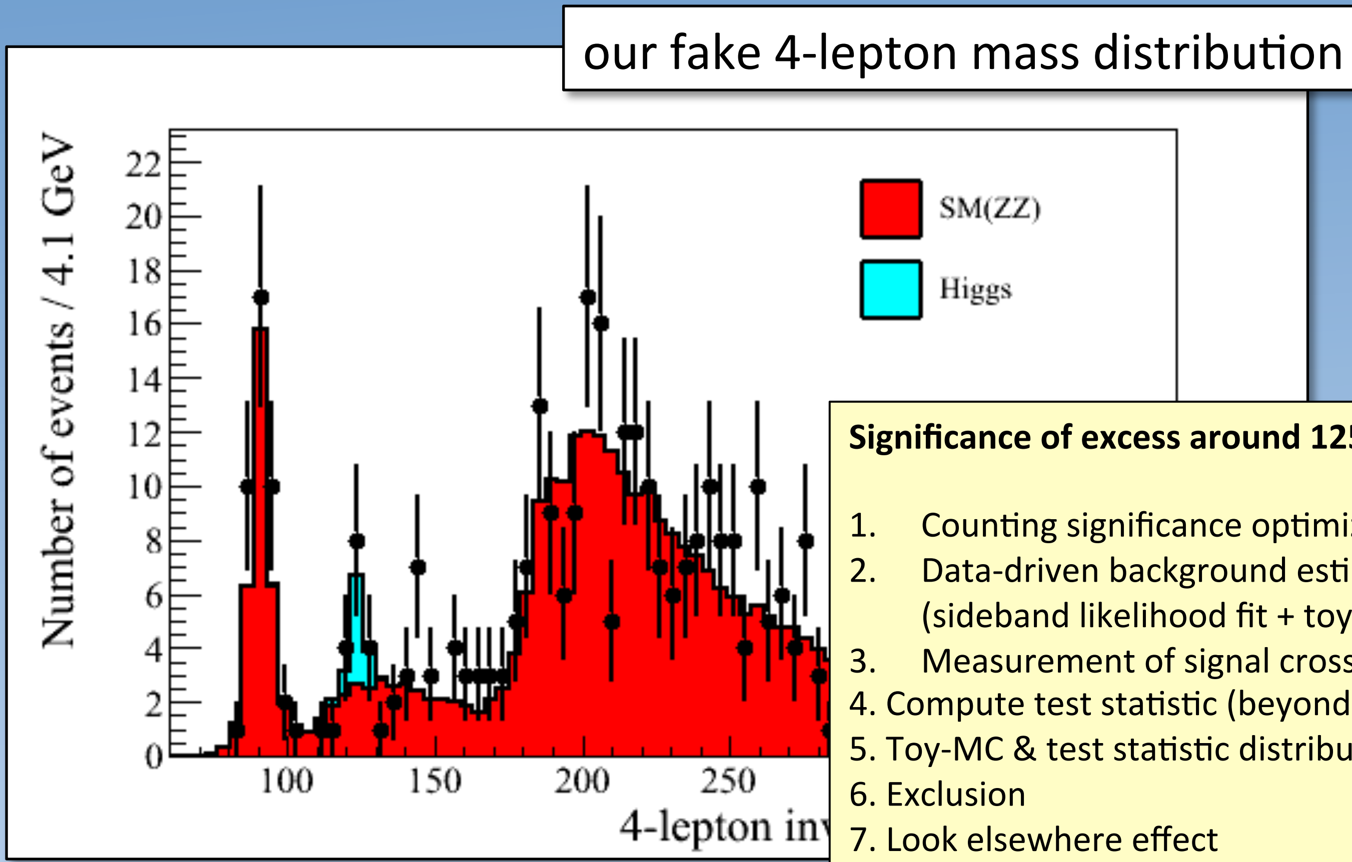
Mass measurement

Data-set for the exercises: 4 lepton mass



Note: - Original histograms have 200 MeV bins

- This is fake data



Significance of excess around 125 GeV

1. Counting significance optimization
2. Data-driven background estimate (sideband likelihood fit + toy MC Poisson)
3. Measurement of signal cross-section
4. Compute test statistic (beyond counting)
5. Toy-MC & test statistic distribution
6. Exclusion
7. Look elsewhere effect
8. Complex/correlated measurements

Full set of exercises (this talk):

http://www.nikhef.nl/~ivov/Talks/2015_04_21_DESY_HandsOnExercises.pdf

Basic material for the exercises:

1) Download tarball: **DesyExercises.tgz**

2) Unpack everything: **tar –vzxf DesyExercises.tgz**

a) Histograms_fake.root
4 histograms with the 4 lepton invariant mass (H125, H200, ZZ, data)

b) DESY_skeleton.C
Some skeleton code (different levels, as minimal as possible)

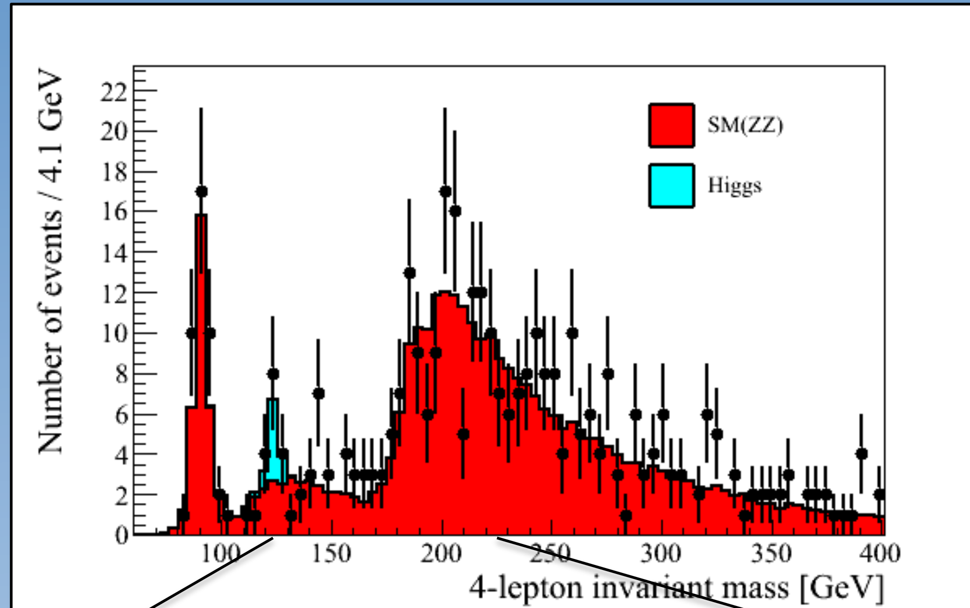
c) Rootlogon.C
Some standard Root blabla

Create the 4-lepton mass plot

```
root> .L DESY_skeleton.C++
root> MassPlot(20)
```

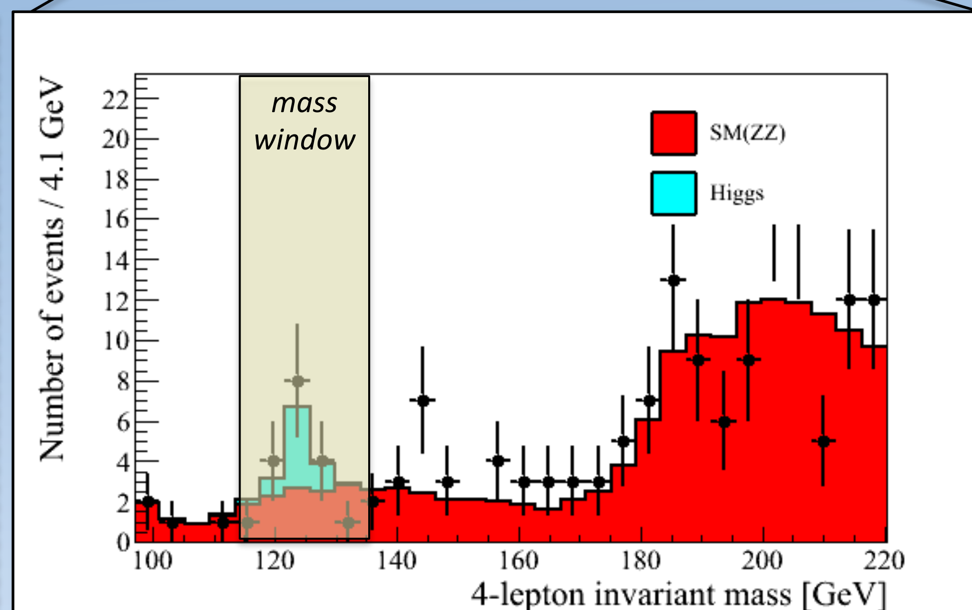
Rebin-factor

hist: h_bgr, h_sig, h_data



Summary in signal mass region
(using 200 MeV bin and 10 GeV window)

Ndata = 16
Nbgr = 6.42
Nsig = 5.96



Exercises: significance

DESY_skeleton.C

```
//======================
void MassPlot(int Irebin){
//======================
  // ------------------------------------------
  // Goal: produce SM+Higgs+data plot
  //       Note: rebinning is only for plotting
  // ------------------------------------------

  //------------------------------------
  //-- Standard stuff and prepare canvas
  //------------------------------------
  gROOT->Clear();
  gROOT->Delete();

  //-- Prepare canvas and plot histograms
  TCanvas * canvas1 = new TCanvas( "canvas1","Standard Canvas",600,400);
  canvas1->SetLeftMargin(0.125);
  canvas1->SetBottomMargin(0.125);
  canvas1->cd();

  //------------------------------------------------------------------
  //-- [1] Prepare histograms
  //--     o Get histograms from the files (signal, background and data)
  //--     o Make cumulative histograms (for signal and background)
  //------------------------------------------------------------------

  //-- Get histograms from the files (higgs, zz and data)
  TH1D *h_sig, *h_bgr, *h_data;
  h_sig  = GetMassDistribution(125);
  h_bgr  = GetMassDistribution(1);
  h_data = GetMassDistribution(2);

  //-----------------------------------
  //-- [2] Plot histograms and make gif
  //--     o rebin histograms
  //--     o prepare cumulative histogram
  //--     o make plot + opsmuk + gif
  //-----------------------------------

  //-- Rebin histograms (only for plotting)
  h_sig->Rebin(Irebin);
  h_bgr->Rebin(Irebin);
  h_data->Rebin(Irebin);

  //-- Prepare cumulative histogram for signal + background
  TH1D *h_sig_plus_bgr = (TH1D* ) h_bgr->Clone("h_sig_plus_bgr");
  h_sig_plus_bgr->Reset();
  for (int i_bin = 1; i_bin < h_bgr->GetNbinsX(); i_bin++){
       h_sig_plus_bgr->SetBinContent( i_bin, h_sig->GetBinContent(i_bin) + h_bgr->GetBinContent(i_bin));
       printf("  REBINNED HISTOGRAM:  bin %d, Ndata = %d\n",i_bin,(int)h_data->GetBinContent(i_bin));
  }

 //-- prepare histograms and plot them on canvas
 double Data_max = h_data->GetBinContent(h_data->GetMaximumBin());
 double Ymax_plot = 1.10* (Data_max + TMath::Sqrt(Data_max));
 h_sig_plus_bgr->SetFillColor(7);
 h_sig_plus_bgr->SetAxisRange(0.,Ymax_plot,"Y");
 h_sig_plus_bgr->SetAxisRange(0.,400.,"X");
 h_bgr->SetFillColor(2);
 h_sig_plus_bgr->Draw("hist");
 h_bgr->Draw("same");
 h_bgr->Draw("axis same");
 h_data->Draw("e same");

 //-- some nice axes and add legend
 AddText( 0.900, 0.035, "4-lepton invariant mass [GeV]",0.060, 0.,"right");                  // X-axis
 AddText( 0.040, 0.900, Form("Number of events / %3.1f GeV",h_bgr->GetBinWidth(1)) ,0.060,90.,"right"); // Y-axis
 TLegend *leg1 = new TLegend(0.65,0.65,0.90,0.85);
 leg1->SetBorderSize(0); leg1->SetFillColor(0);
 TLegendEntry *leg1a = leg1->AddEntry(h_bgr,          " SM(ZZ)", "f");  leg1a->SetTextSize(0.04);
 TLegendEntry *leg1b = leg1->AddEntry(h_sig_plus_bgr, " Higgs" , "f");  leg1b->SetTextSize(0.04);
 leg1->Draw();

 //-- prepare gif
 canvas1->Print(Form("./MassPlot_rebin%d.gif",Irebin));

 return;

  //==============
} // end MassPlot()
  //==============
```

→ *Define canvas*

→ *Get histograms from root-file*

→ *Rebin histograms*

Print bin content

Make cumulative histogram

→ *Histogram characteristics & plot*

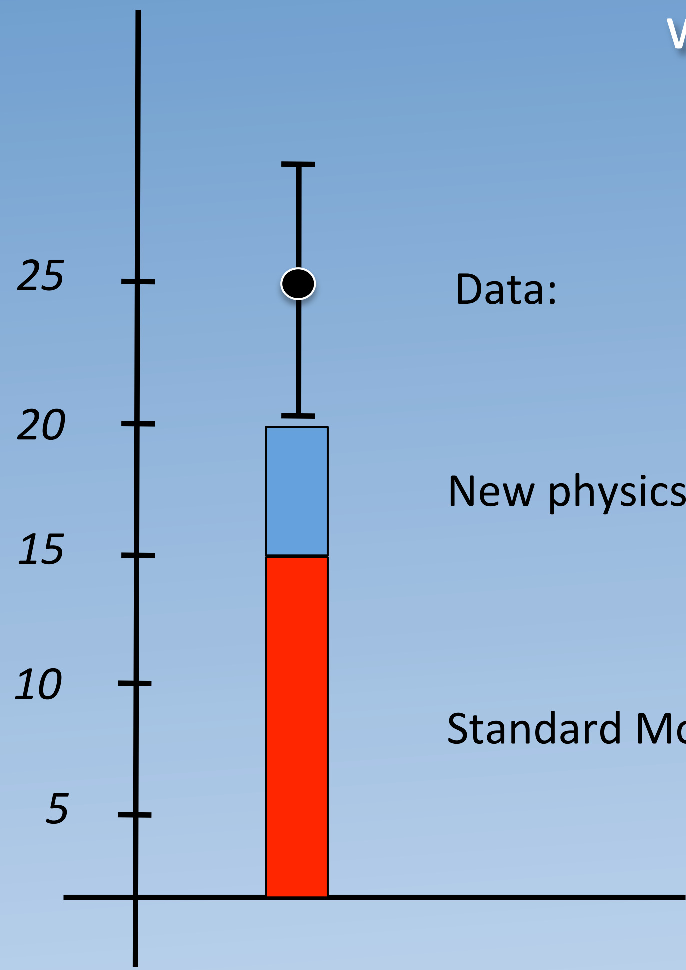
Add text

Save plot as gif in your directory

10-slide mini lecture on significance:

- discovery and exclusion -

General remark : what is the significance ?



Data: $N_{data} = 25$

New physics: $N_{sig} = 5$

Standard Model: $N_{bgr} = 15$

Significance for N events: probability to observe N events (or even more) under the background-only hypothesis

Observed significance:

$$\int_{25}^{\infty} \text{Poisson}(N \mid 15)\, dN = 0.0112$$ ← *p-value*

$$= 2.28 \text{ sigma}$$ ← *significance*

Expected significance:

$$\int_{20}^{\infty} \text{Poisson}(N \mid 15)\, dN = 0.1248$$

$$= 1.15 \text{ sigma}$$

Discovery if p-value < 2.87×10^{-7}

→ 39 events

Poisson distribution

The Poisson distribution

Binomial with n→∞, p →0 and np=λ

$$P(n \mid \lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$$

Poisson distribution

Probability to observe n events when λ are expected

$$P(0 \mid 4.0) = 0.01832$$

$$P(2 \mid 4.0) = 0.14653 \quad !$$

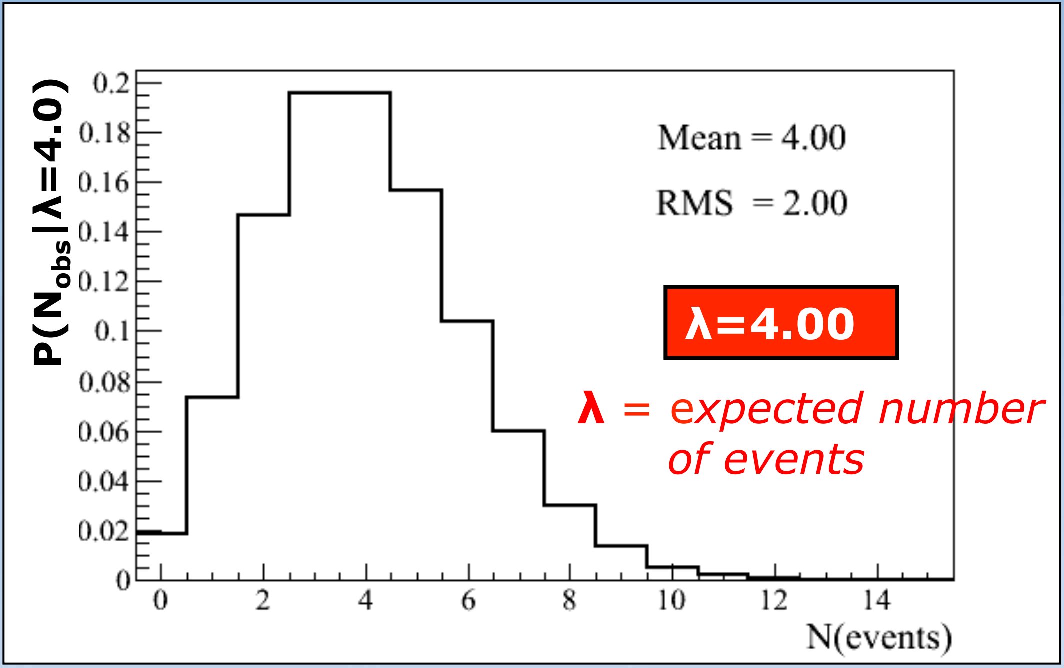
$$P(3 \mid 4.0) = 0.19537$$

$$P(4 \mid 4.0) = 0.19537$$

$$P(6 \mid 4.0) = 0.10420 \quad !$$

#observed — **varying**

λ hypothesis — **fixed**



λ = *expected number of events*

Known λ (Poisson)

Binomial with n→∞, p →0 np=λ

$$P(n \mid \lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$$

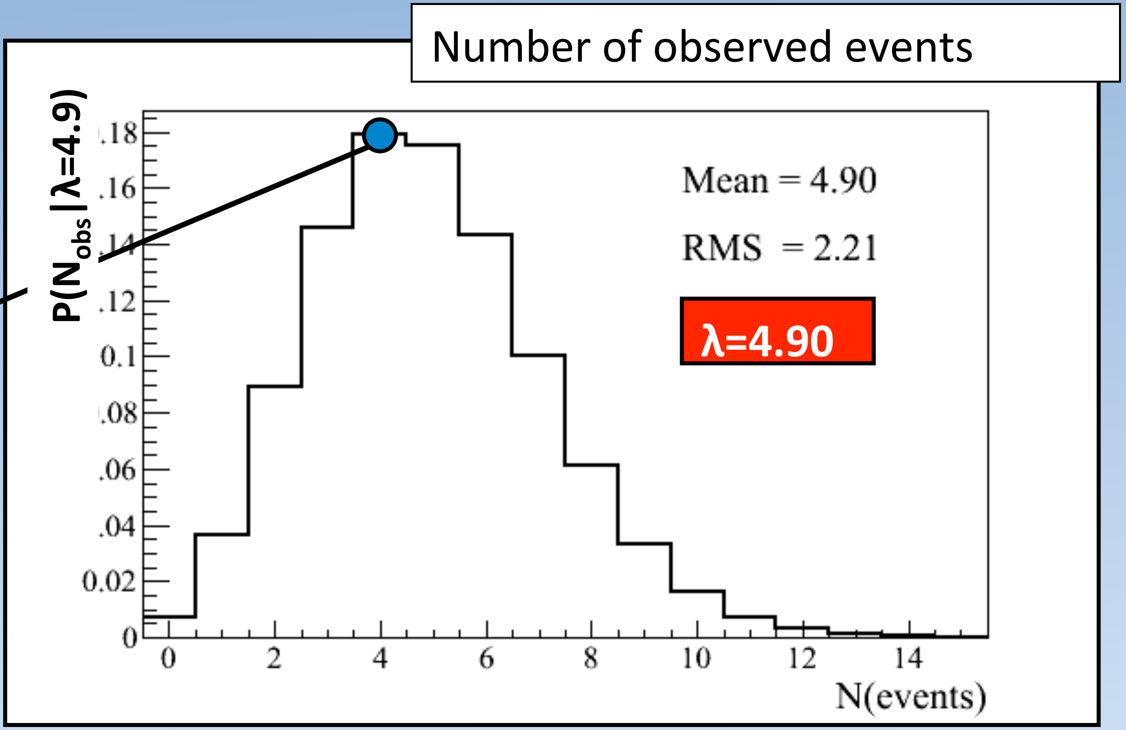
Poisson distribution

Probability to observe n events when λ are expected

$$P(0 \mid 4.9) = 0.00745$$
$$P(2 \mid 4.9) = 0.08940$$
$$P(3 \mid 4.9) = 0.14601$$
$$P(4 \mid 4.9) = 0.17887$$

#observed — varying

λ hypothesis — fixed



the famous √N

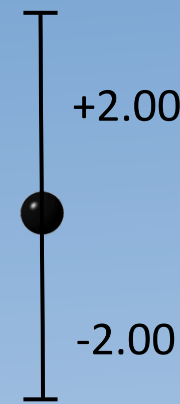
Properties Poisson distribution

(1) Mean: $\langle n \rangle = \lambda$

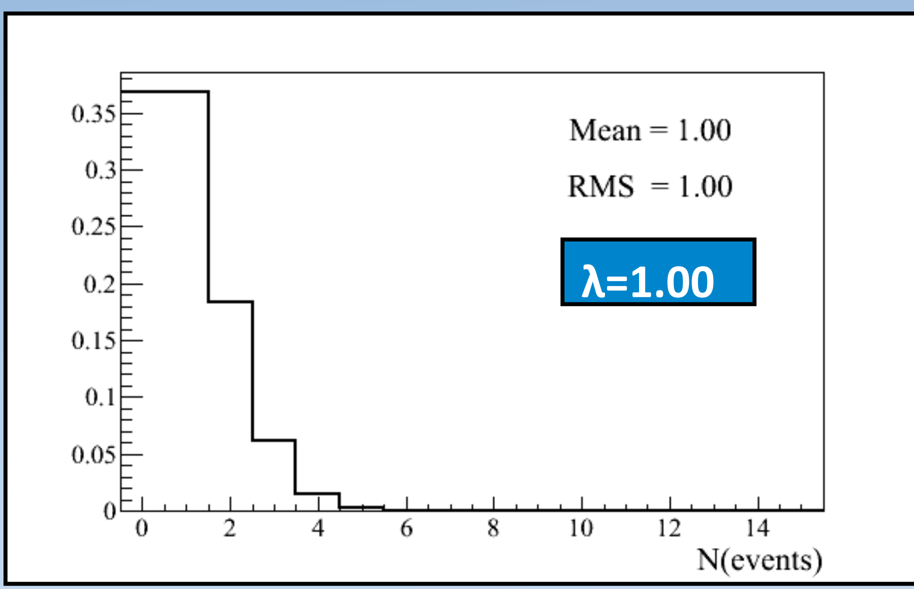
(2) Variance: $\langle (n - \langle n \rangle)^2 \rangle = \lambda$

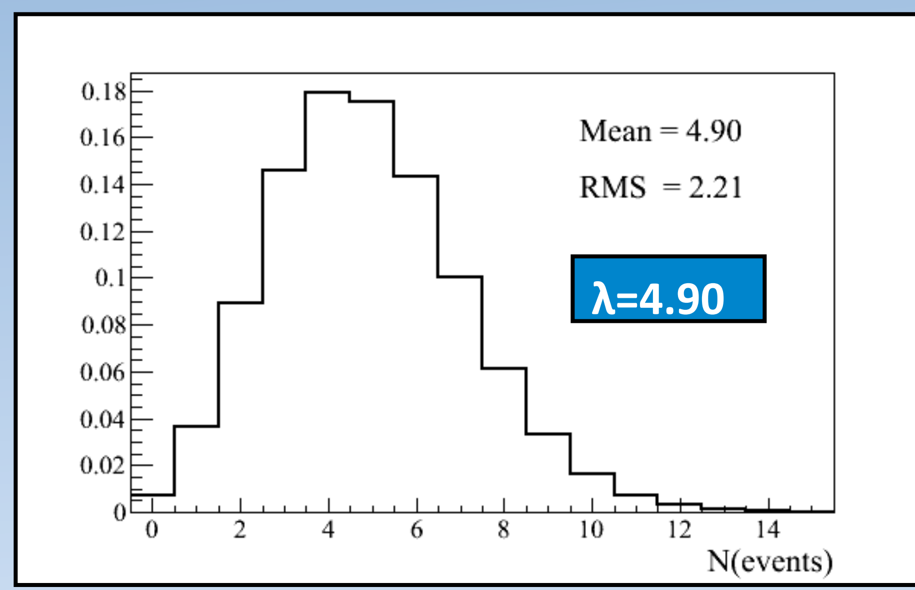
(3) Most likely: first integer ≤ λ

Usual way to represent the error on a data-point



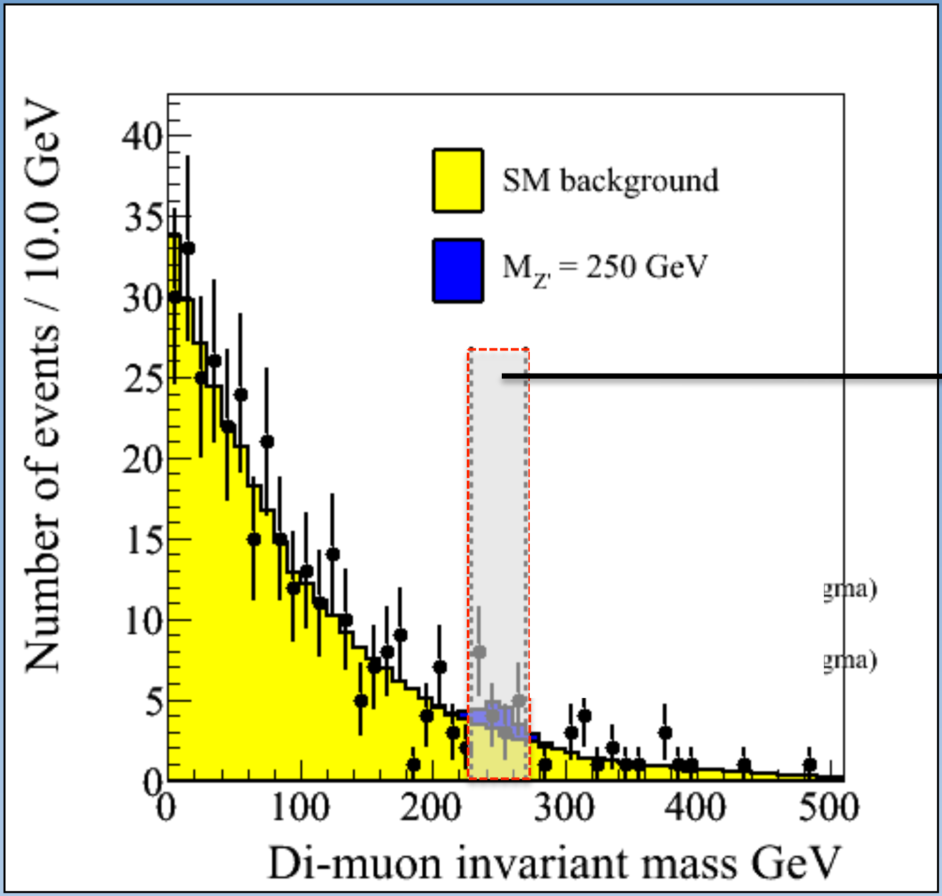
Not default in Root





Significance example

Counting events in a mass window



Standard Model

SM	10
Higgs	5
Data	12

Ok, now what ?

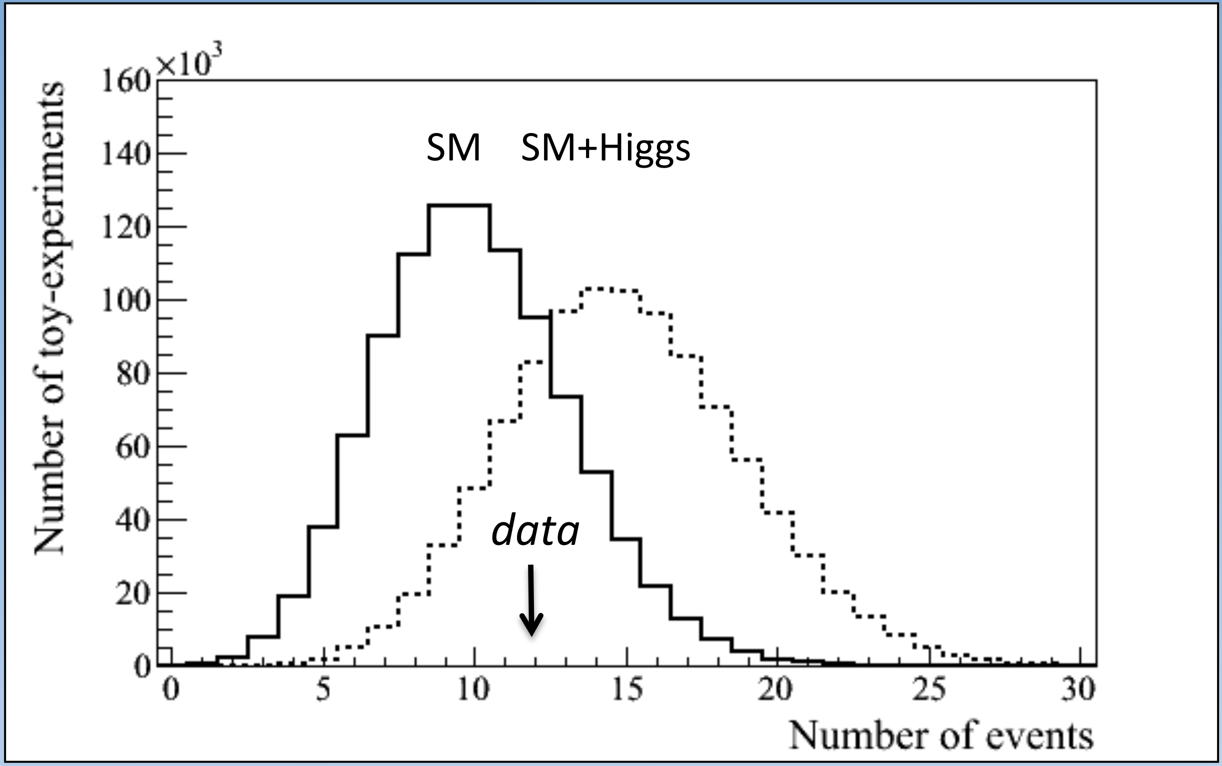
Next slides:
- *discovery*
- *exclusion*

Standard Model

SM	10
Higgs	5
Data	12

Ok, now what ?

Poisson distribution



Significance for N events: probability to observe N events (or even more) under the background-only hypothesis

Interpretation

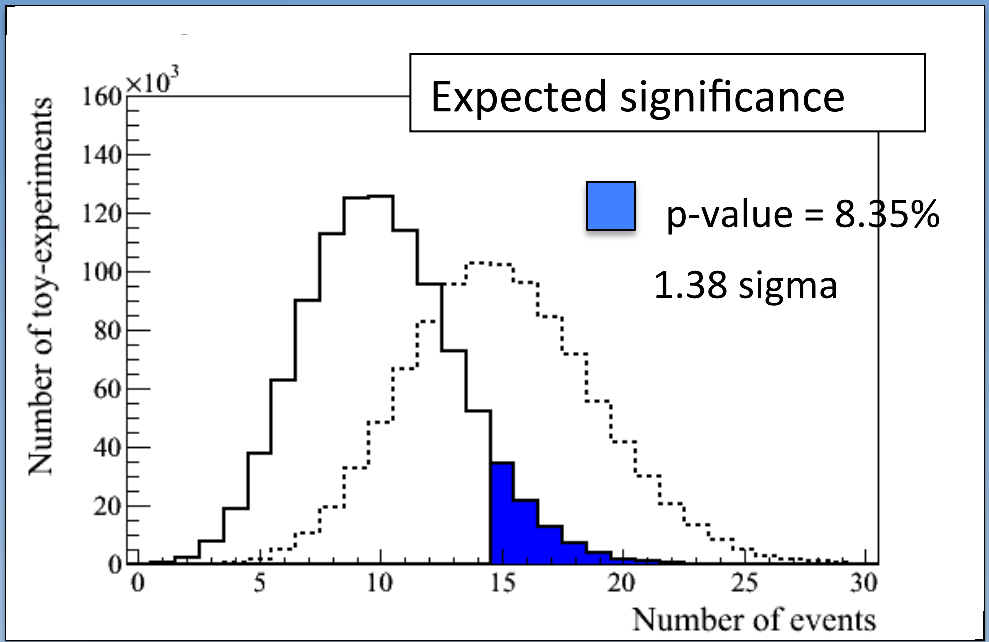
optimistic: discovery

Incompatibility with SM-hypothesis

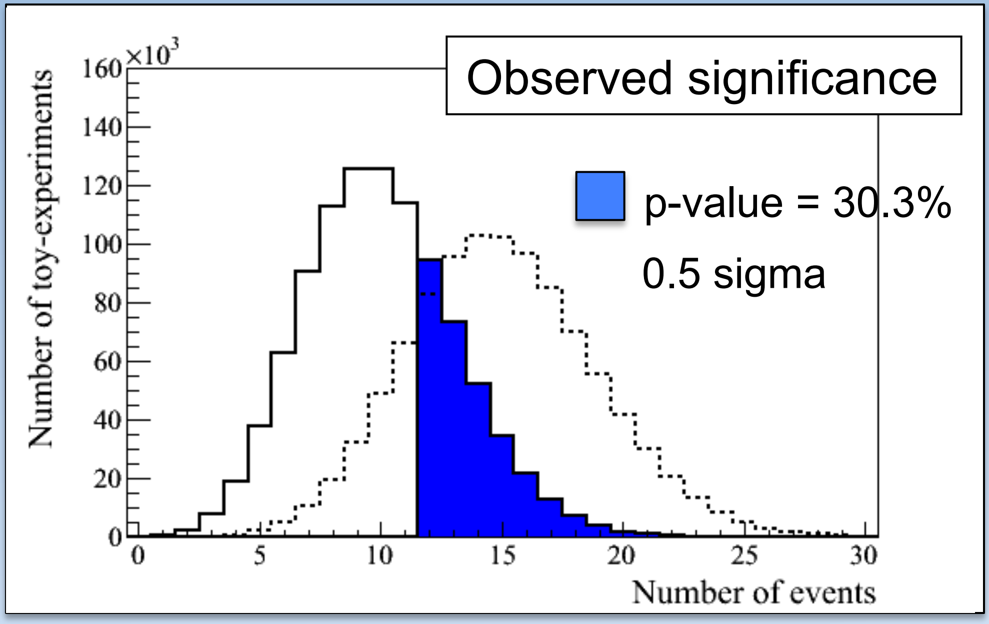
p-value: incompatibility with SM-only hypothesis

SM	10
Higgs	5
Data	12

*1) What is the **expected** significance ?*

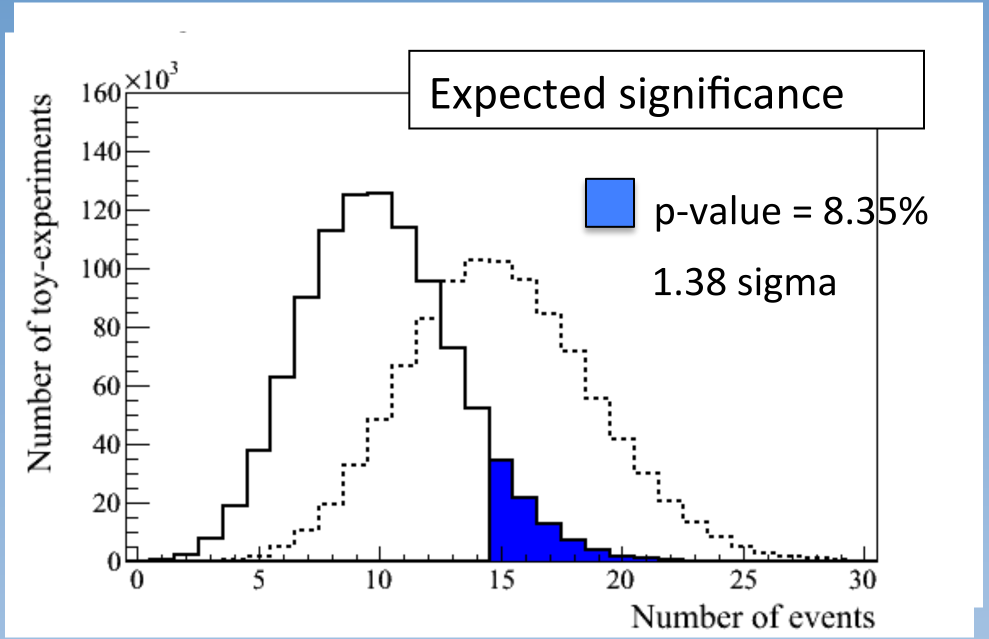


*2) What is the **observed** significance ?*



p-value: incompatibility with SM-only hypothesis

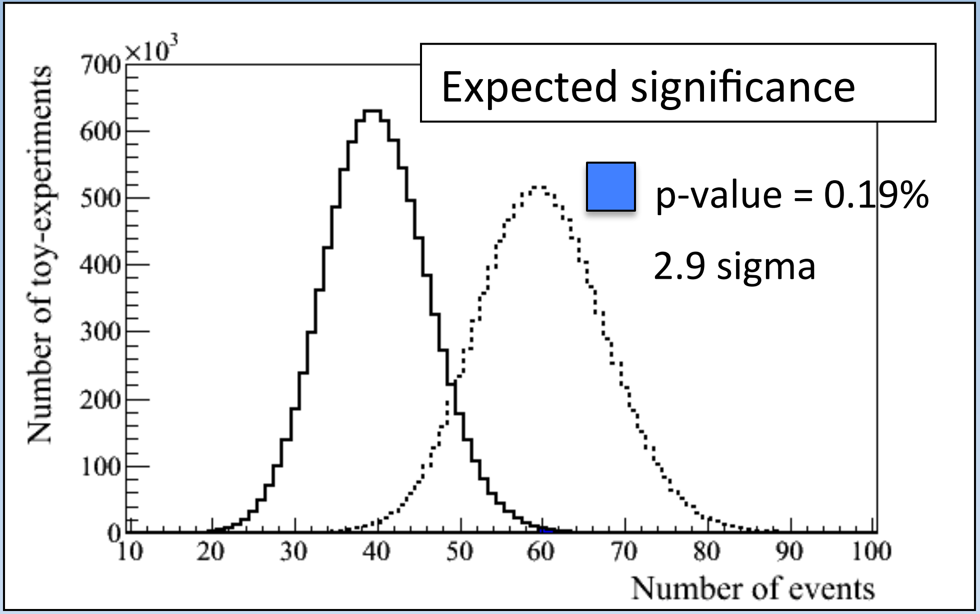
SM	10
Higgs	5



3) At what Lumi do you expect to be able to claim a discovery ?

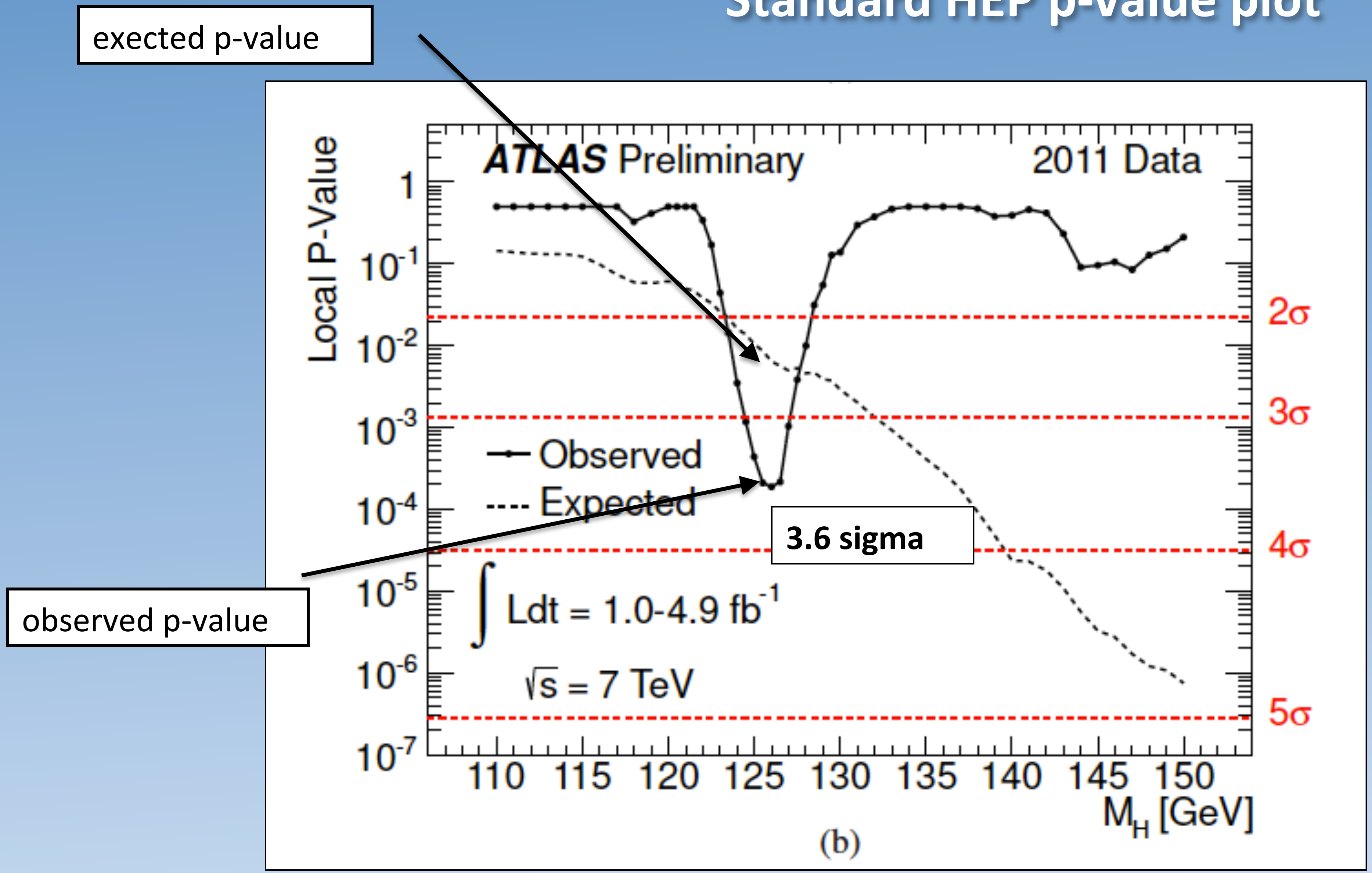
3 TIMES MORE LUMINOSITY

SM	30
Higgs	15



Discovery if p-value < 2.87×10^{-7}

Standard HEP p-value plot



Interpretation

pessimistic: exclusion

Incompatibility with New Physics-hypothesis

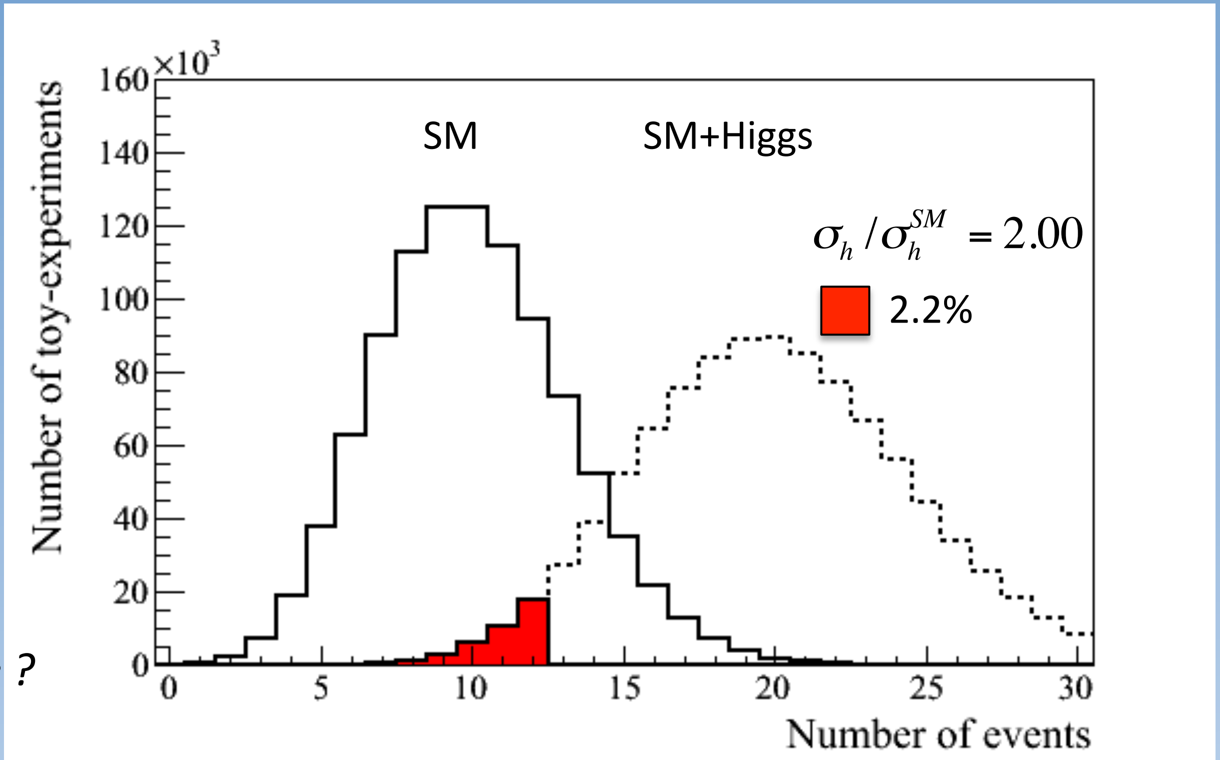
Excluding a signal: Incompatibility with s+b hypothesis

Standard Model

SM	10
Higgs	5
Data	12

Can we exclude the SM+Higgs hypothesis ?

What σ_h/σ_h^{SM} can we exclude ?



Exclusion: probability to observe N events (or even less) under the signal + background hypothesis

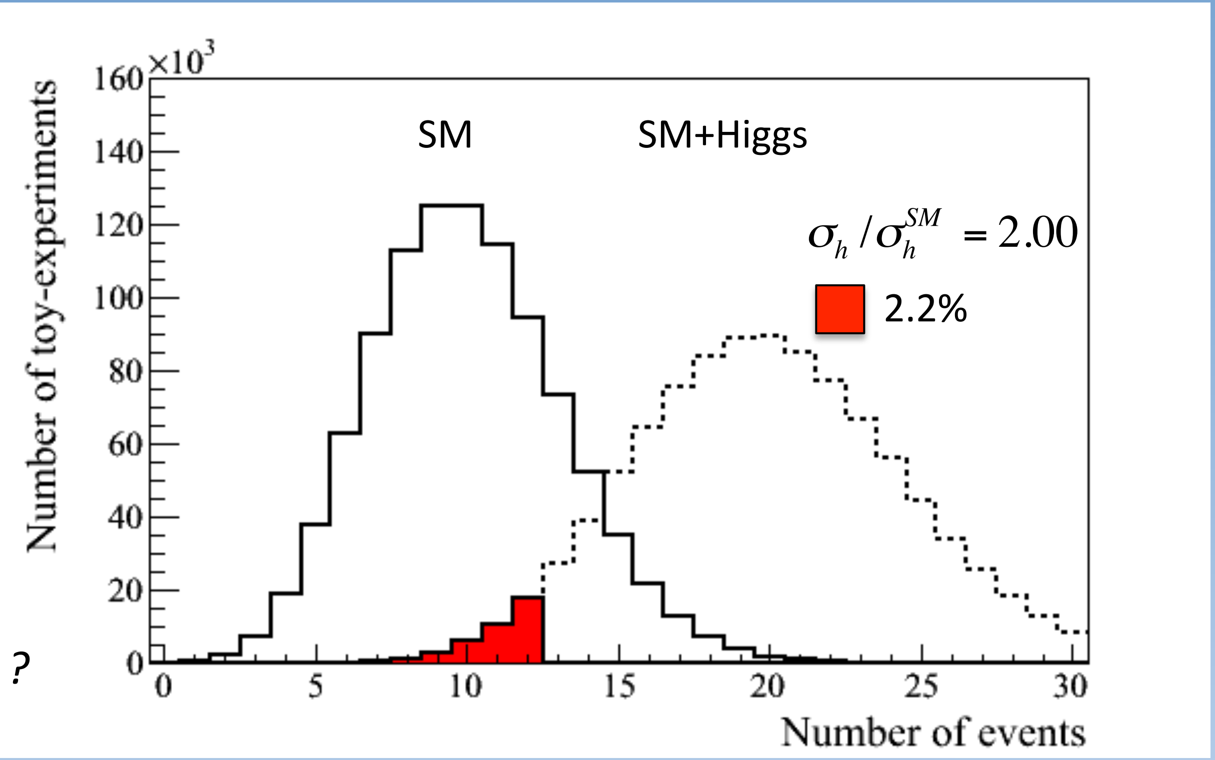
Excluding a signal: Incompatibility with s+b hypothesis

Standard Model

SM	10
Higgs	5
Data	12

Can we exclude the SM+Higgs hypothesis ?

What σ_h/σ_h^{SM} can we exclude ?



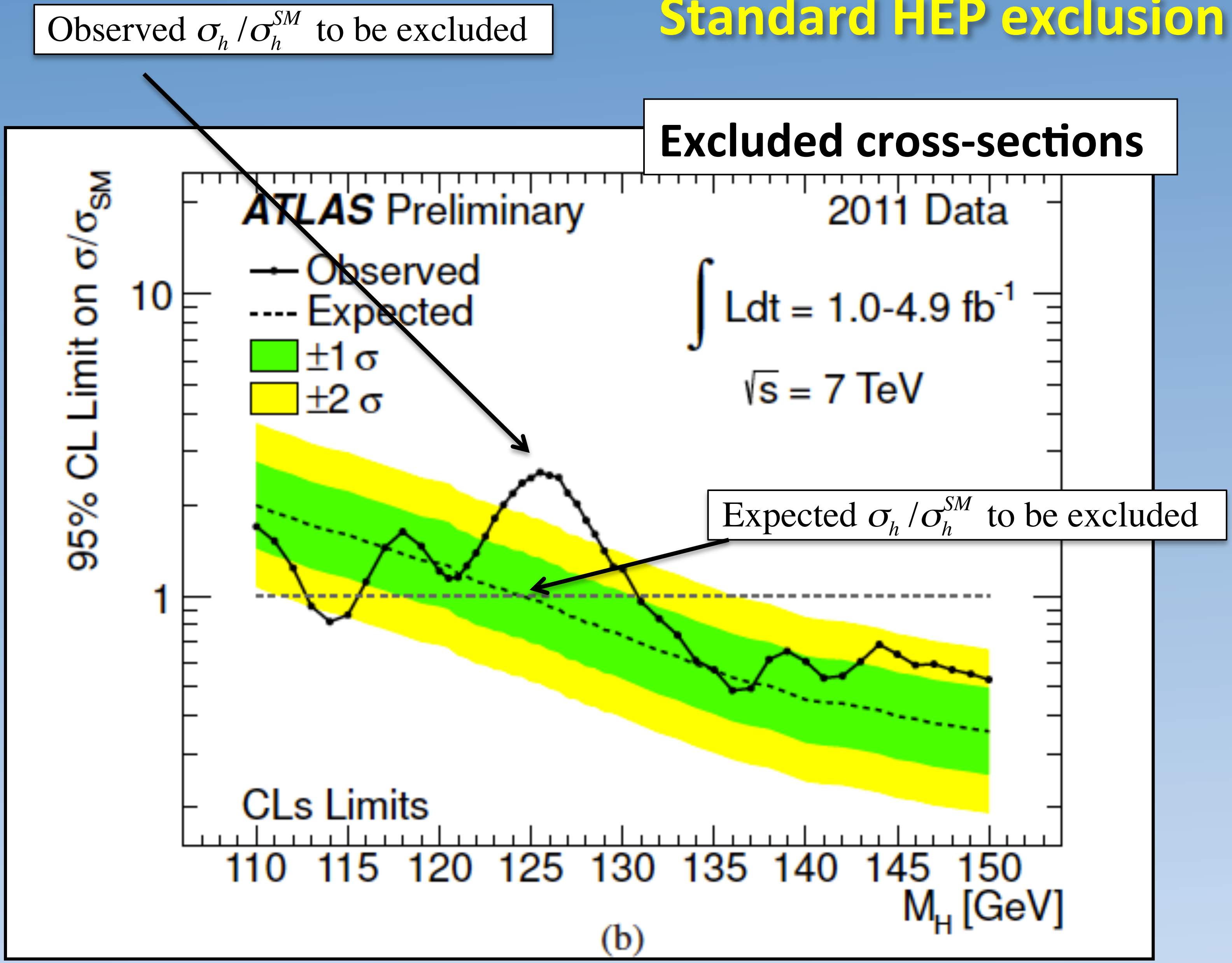
σ/σ_{SM}	SM	# data	SM+Higgs	
1.0	10	12	15.0	18.5 %
1.5	10	12	17.5	6.8%
2.0	10	12	20.0	2.2%

excluded

Expected exclusion ? Use mean SM instead of Ndata

Observed excluded cross-section, σ_h/σ_h^{SM} , = 1.64

Standard HEP exclusion plot

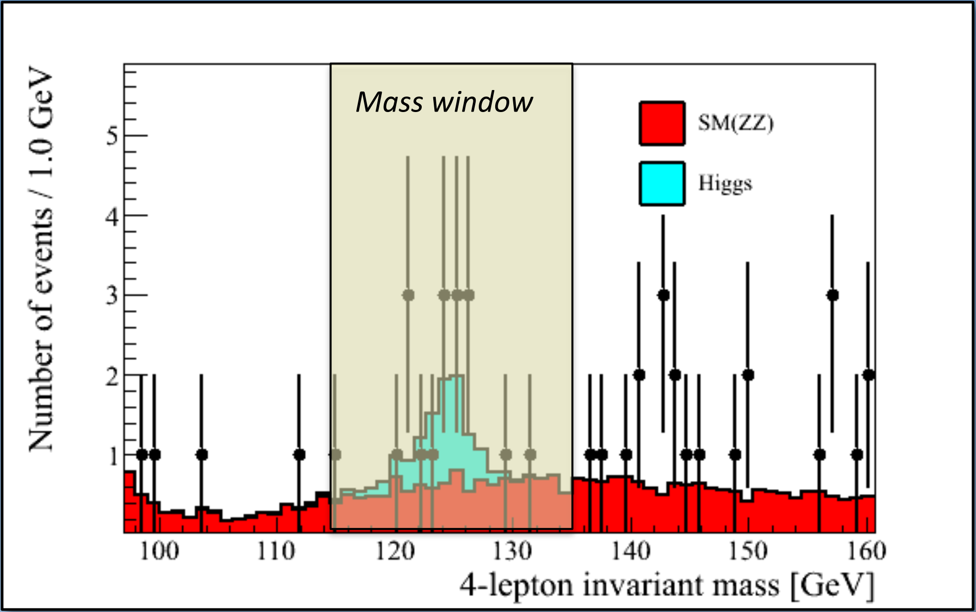



Exercises
PART 1

1) significance optimization

Exercise 1:

Optimizing the counting experiment



Code you could use:

```
IntegratePoissonFromRight()
Significance_Optimization()
```

Exercise 1: significance optimization of search window (Poisson counting)

1.1 Find the window that optimizes the expected significance

1.2 Find the window that optimizes the observed significance (and never do it again)

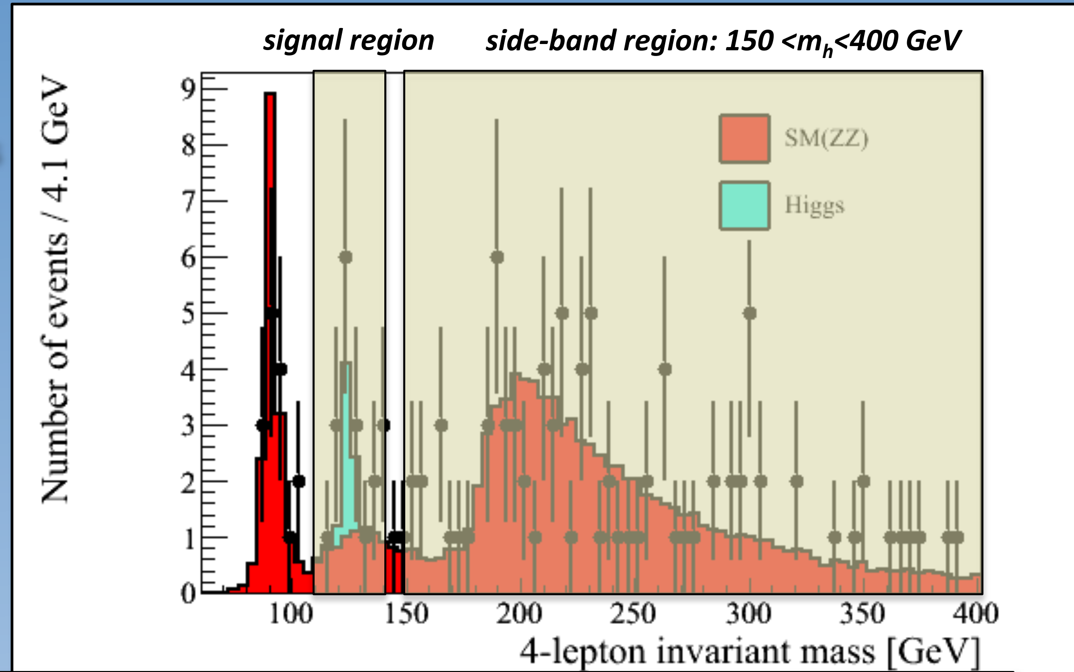
1.3 Find the window that optimizes the expected significance for 5x higher luminosity

1.4 At what luminosity do you expect to be able to make a discovery ?

2) background estimate from side-band fit

Exercise 2:

Data driven bkg estimate in 10 GeV ,mass window or optimal one from Exercise 1



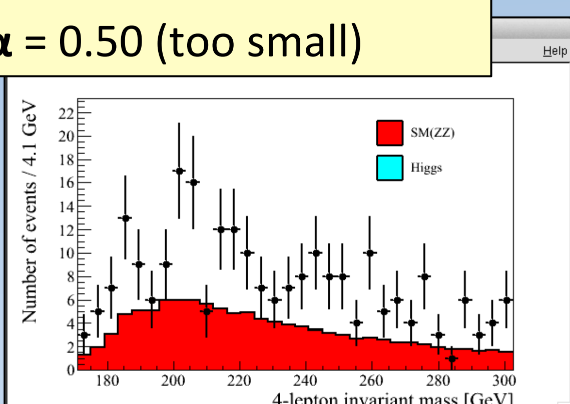
Code you could use:

```
SideBandFit()
```

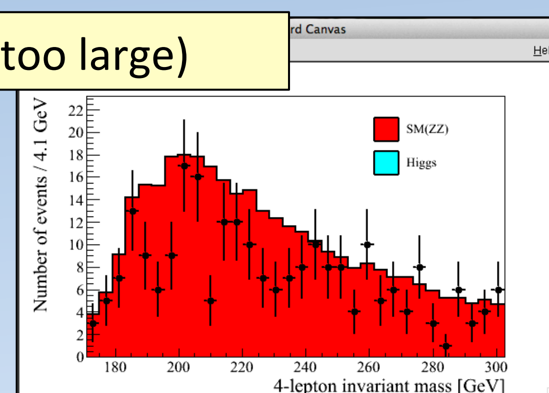
Exercise 2: background estimation from side-band fit

2.1 What is the optimal scale-factor for the background (α) ?
Do a likelihood fit to the side-band region $150 \leq m_h \leq 400$ GeV

α = 0.50 (too small)



α = 1.50 (too large)

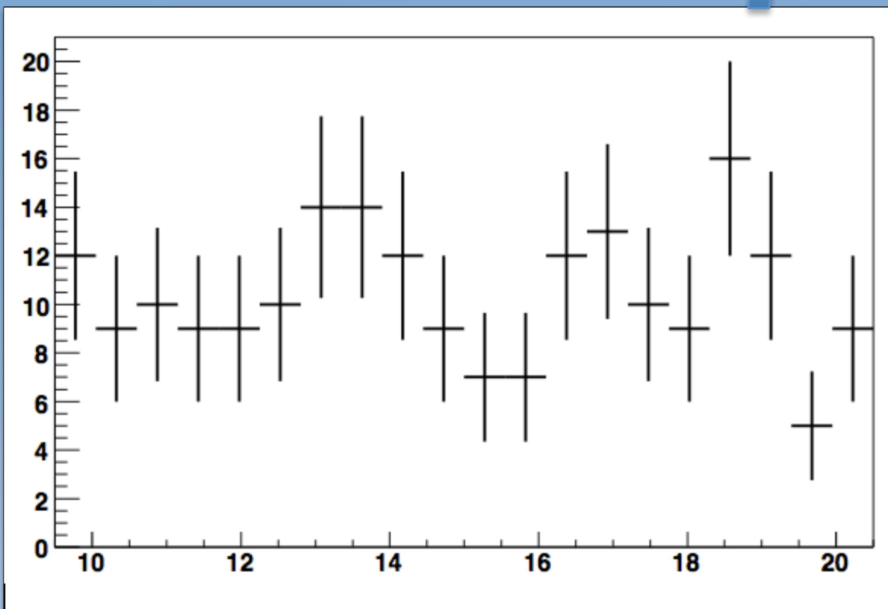


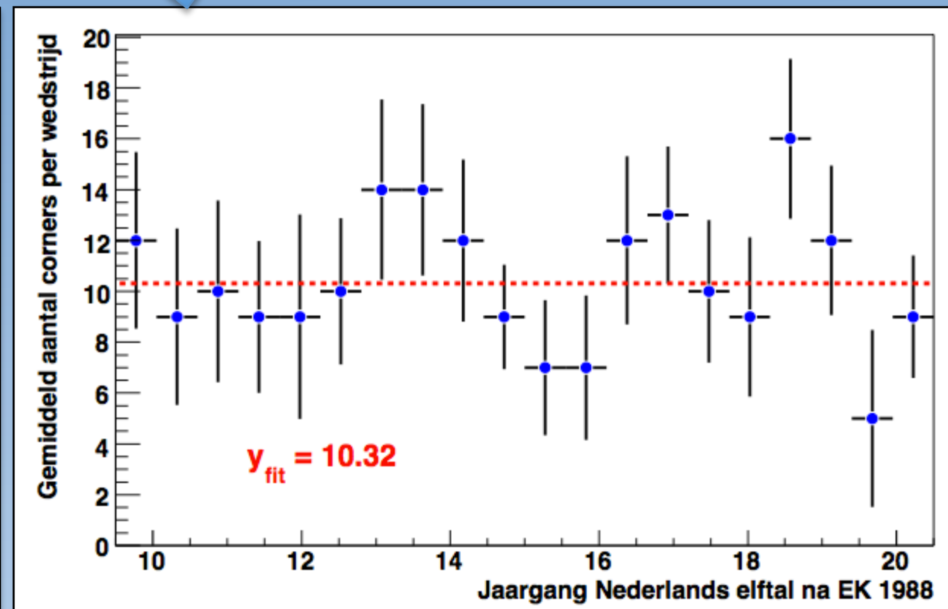
10-slide mini lecture on fitting

- Likelihood fits and uncertainties -

Simple likelihood fit

Can everybody do this ?



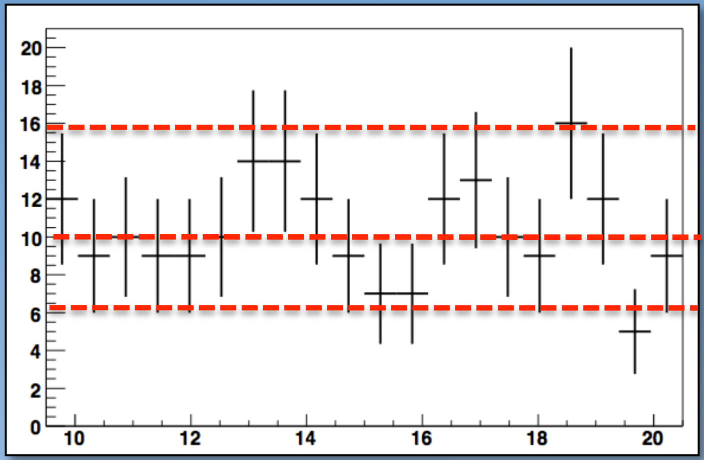


If you want to reproduce this plot, but cannot please let me know

TMath::Poisson(Nevt_bin, alpha)

http://www.nikhef.nl/~ivov/SimpleFit/

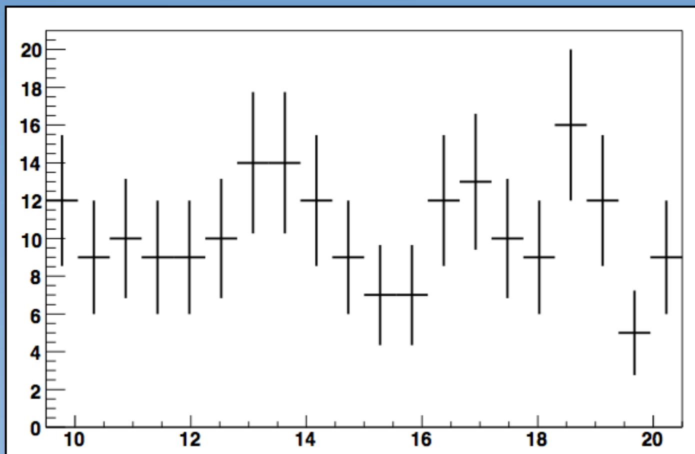
Fitting in 1 slide



You model: $f(x) = \lambda$

Try different values of λ
and for each one compute:

Fitting in 1 slide



You model: $f(x) = \lambda$

Try different values of λ and for each one compute the ***compatibility*** of the model with the data

χ^2-fit

Compatibility number :

$$\chi^2 = \sum_{bins} \frac{\left(N_{bin}^{data} - \lambda_{bin}^{expected}\right)^2}{N_{bin}^{data}}$$

Best value:
Value of λ that minimizes χ^2 (χ_{min}^2)

Errors:
Values of λ for which $\chi^2 = \chi_{min}^2 + 1$

Likelihood-fit

Compatibility number :

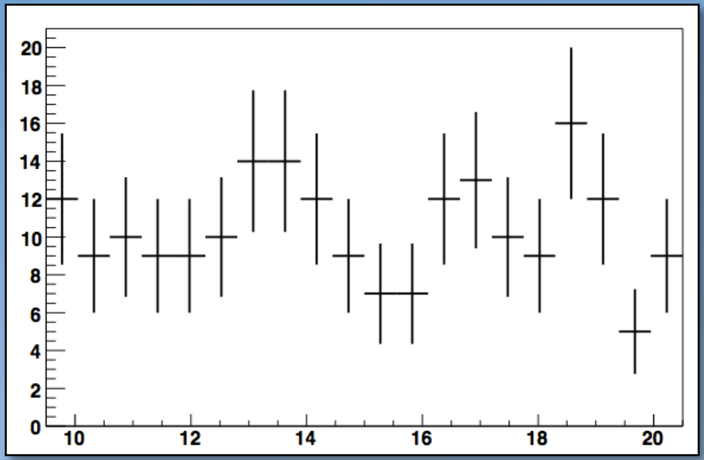
$$-2\log(L) = -2 \cdot \sum_{bins} \log(Poisson(N_{bin}^{data} \mid \lambda))$$

TMath::Poisson(Nevt_bin, λ)

Best value:
Value of λ that minimizes $-2Log(L)$ ($-2log(L)_{min}$)

Errors:
Values of λ for which $2Log(L)$ = ($-2log(L)_{min}$) +1

Fitting in 1 slide



You model: $f(x) = \lambda$

Try different values of λ and for each one compute the ***compatibility*** of the model with the data

Recipe for each value of λ:

- Set LogLik = 0
- Loop over all bins:
  - o For each bin: compute prob. to observe N_i evts when you expect λ. **Poisson distribution**
  - o take -2*Log of bin-probability
  - o Add to existing LogLik

→ Output LogLik (1 number)

Likelihood-fit

Compatibility number :

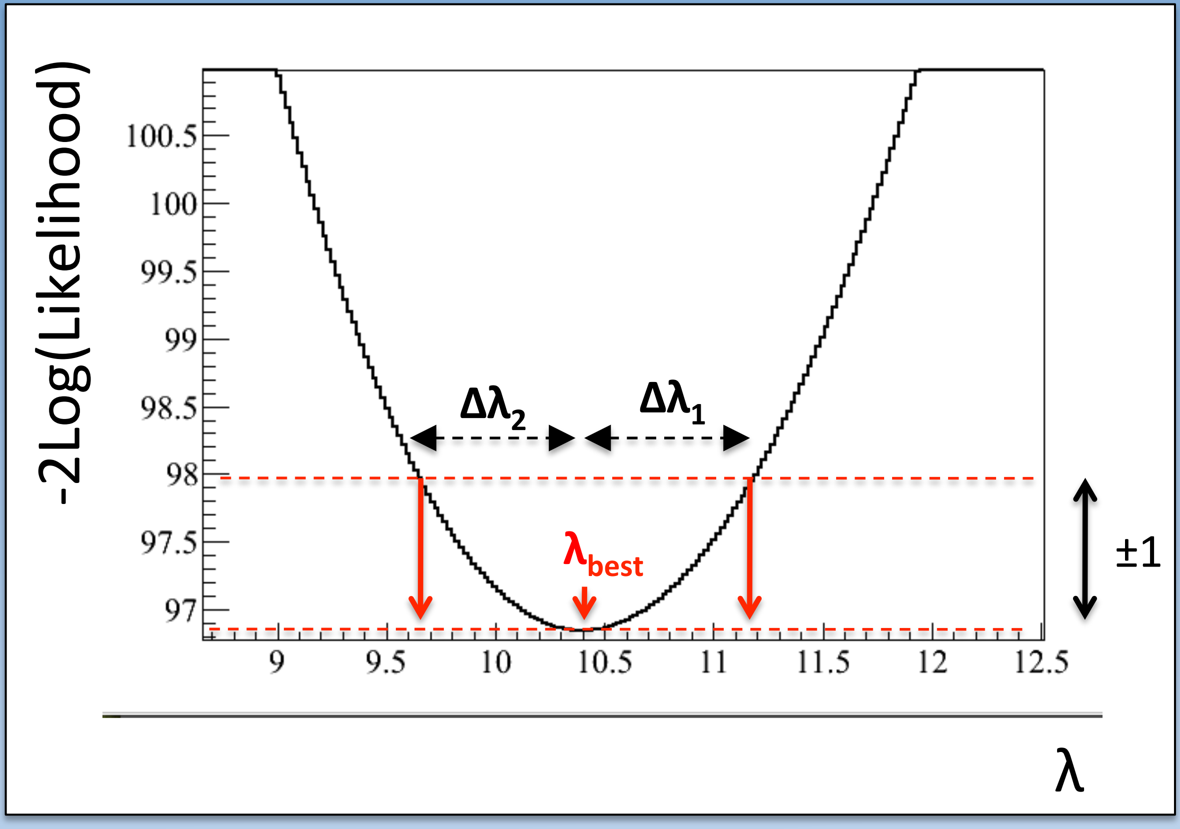
$$-2\log(L) = -2\cdot\sum_{bins}\log(Poisson(N_{bin}^{data}\mid\lambda))$$

TMath::Poisson(Nevt_bin, λ)

Best value:
Value of λ that minimizes -2Log(*L*) ($-2\log(L)_{min}$)

Errors:
Values of λ for which 2Log(*L*) = ($-2\log(L)_{min}$) +1

Result from the fit

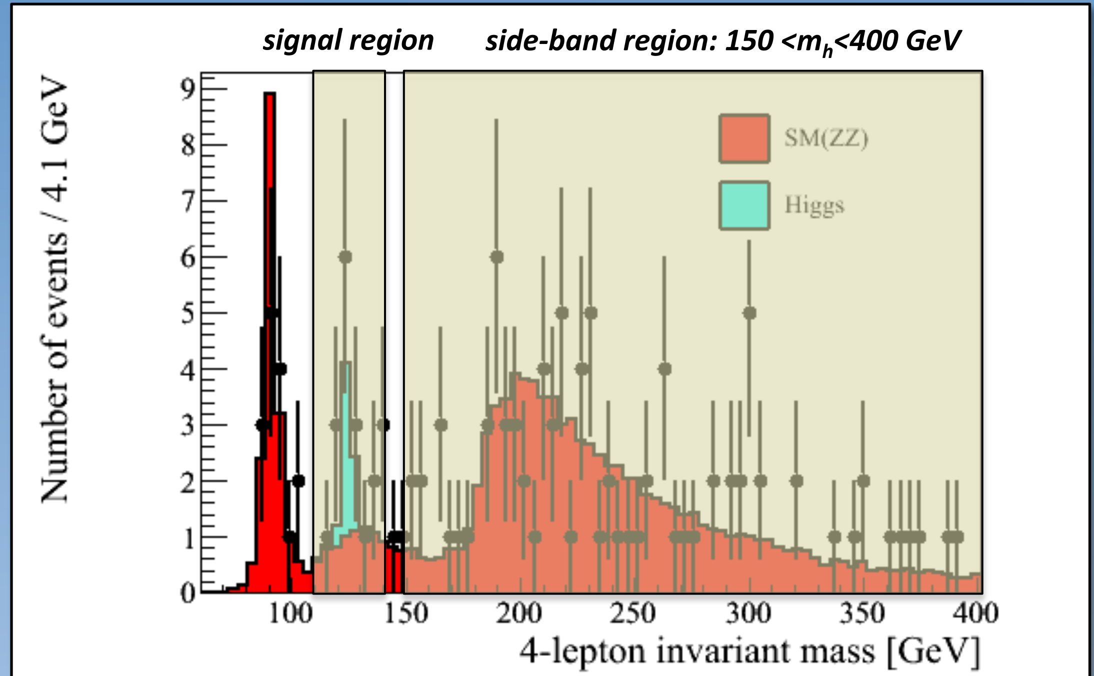


$$\text{result}: \quad \lambda = \lambda_{\text{best}}{}^{+\Delta\lambda_1}_{-\Delta\lambda_2}$$

Exercise 2:
continued

Code you could use:

```
SideBandFit()
```

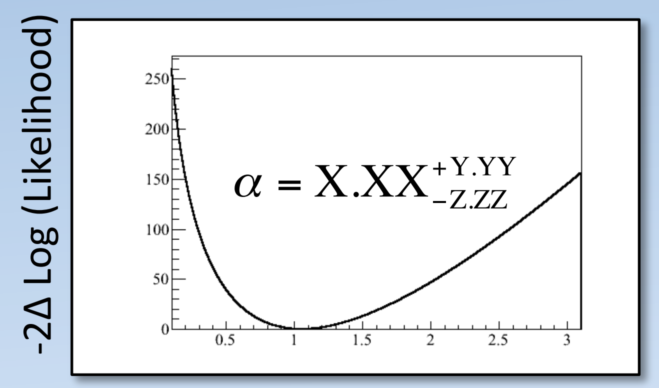


Exercise 2: significance optimization of mass/search window (use Poisson counting)

2.1 What is the optimal scale-factor for the background (α) ?
Do a likelihood fit to the side-band region $150 \leq m_h \leq 400$ GeV

Computing the likelihood:
For each 'guess' of α:

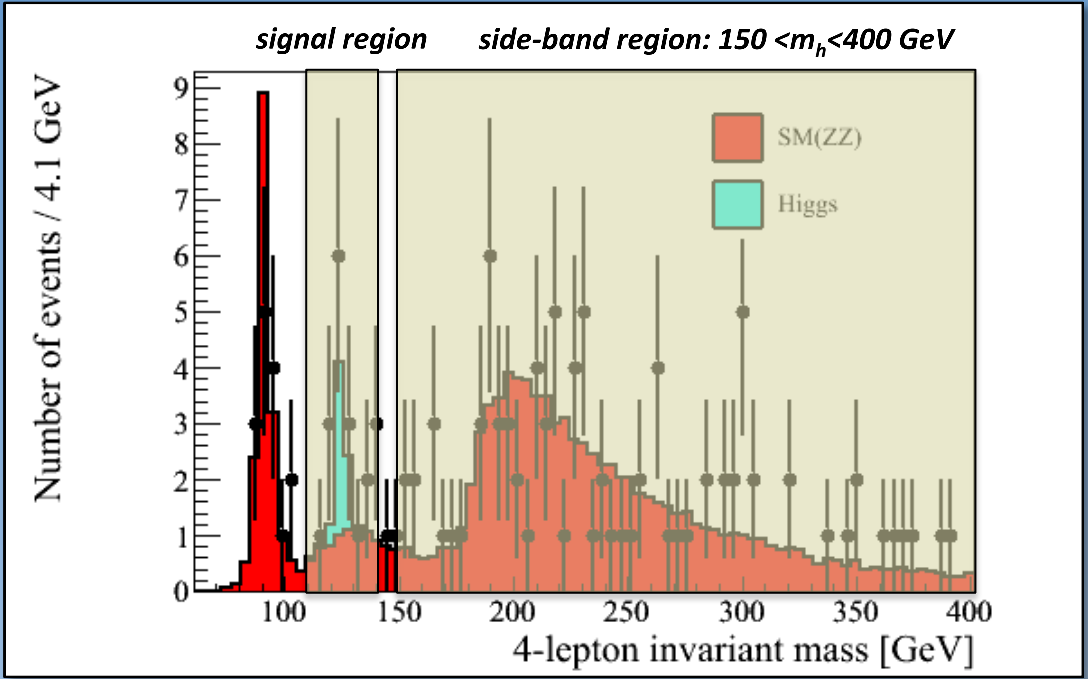
$$-2\log(L) = -2 \cdot \sum_{bins} \log(Poisson(N_{bin}^{data} \mid \alpha \cdot f_{bin}^{SM}))$$



Exercise 2:
continued

Code to use:

none



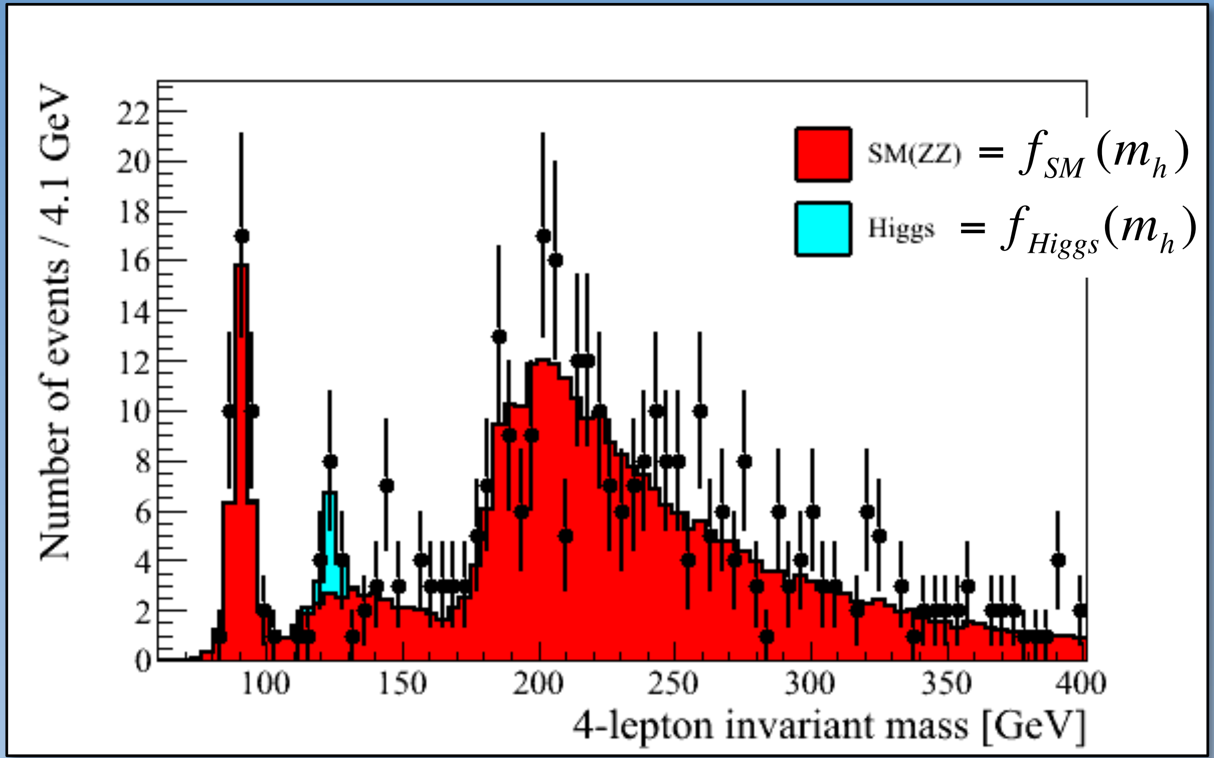
2.2 Estimate background and its uncertainty b±Δb in the mass window around 125 GeV (your optimal one from Exercise 1 or a simply a 10 GeV window)

2.3 Compute the expected and observed significance using Toy-MC

Note: Draw random # events in the mass window (for b-only and s+b)
For each toy-experiment, not just draw a Poisson number,
but also take a new central value using the (Gauss) Δb from 3.2

Compare it to the significance in exercise 1

3) signal cross-section



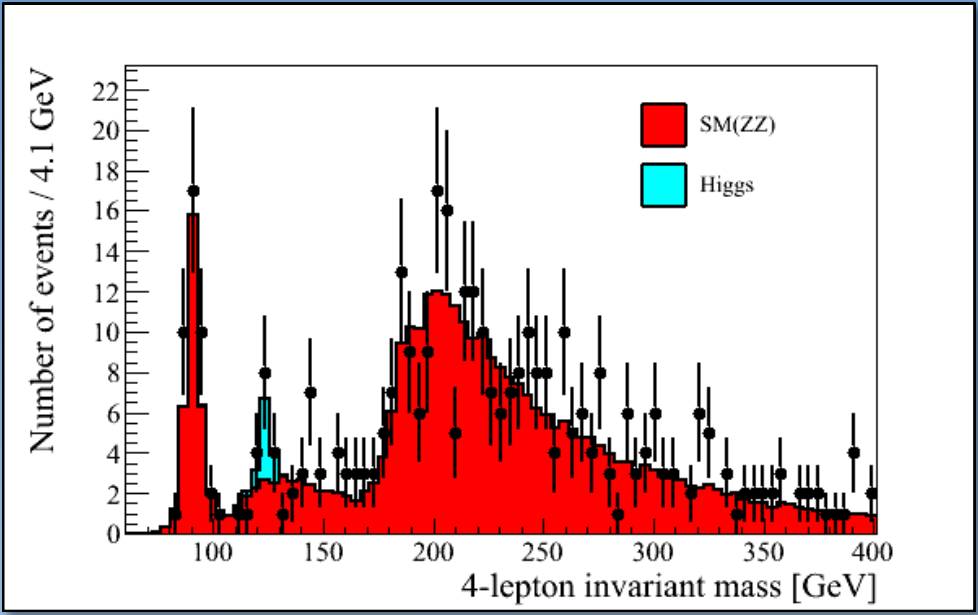
$$f(m_h) = \mu \times f_{Higgs}(m_h) + \alpha \times f_{SM}(m_h)$$

Scale factor for the Higgs

Scale factor for the SM background

Exercise 3:

Estimate of Higgs cross-section

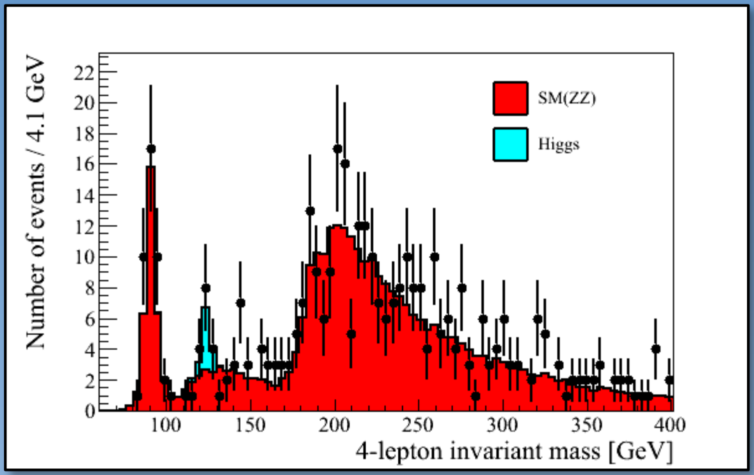


$$-2\cdot \log\left(Likelihood\right) = -2\cdot \sum_{bins} \log\left(Poisson(N_{bin}^{data} \mid \mu\cdot\ f_{bin}^{Higgs} + \alpha\cdot\ f_{bin}^{SM})\right)$$

Exercise 3: Measurement of the signal cross-section

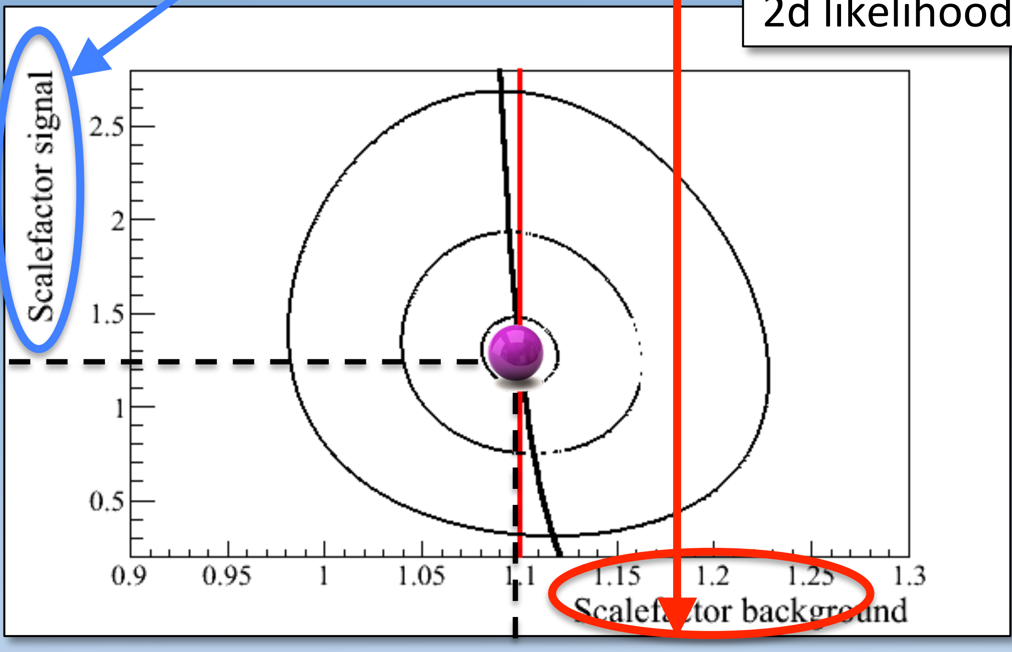
3.1 Do a fit where you fix background (to level from exercise 2) and leave the signal cross-section (μ) free. What is the best value for μ and what is its uncertainty ?

3.2 Do a fit where you leave both α and μ free. What are the optimal values ? How would you estimate the uncertainty on each of the parameters ?



$$f(m_h) = \mu \times f_{Higgs}(m_h) + \alpha \times f_{SM}(m_h)$$

2d likelihood



$$\mu_{bgr}^{best} = 1.29$$

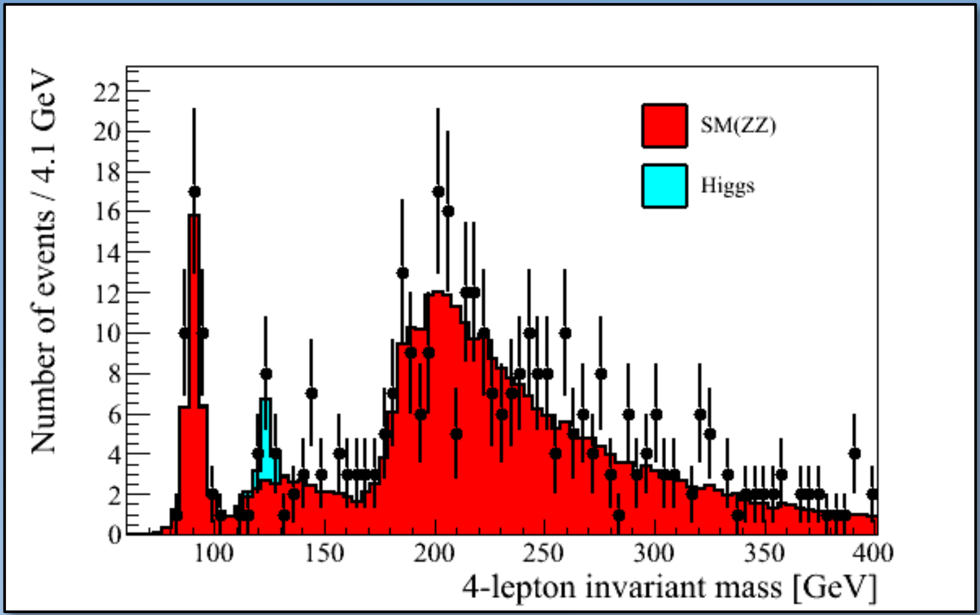
$$\alpha_{bgr}^{best} = 1.10$$


Good luck


Exercises
PART 2

4) More complex test statistics

Beyond simple counting: profile likelihood ratio test-statistic



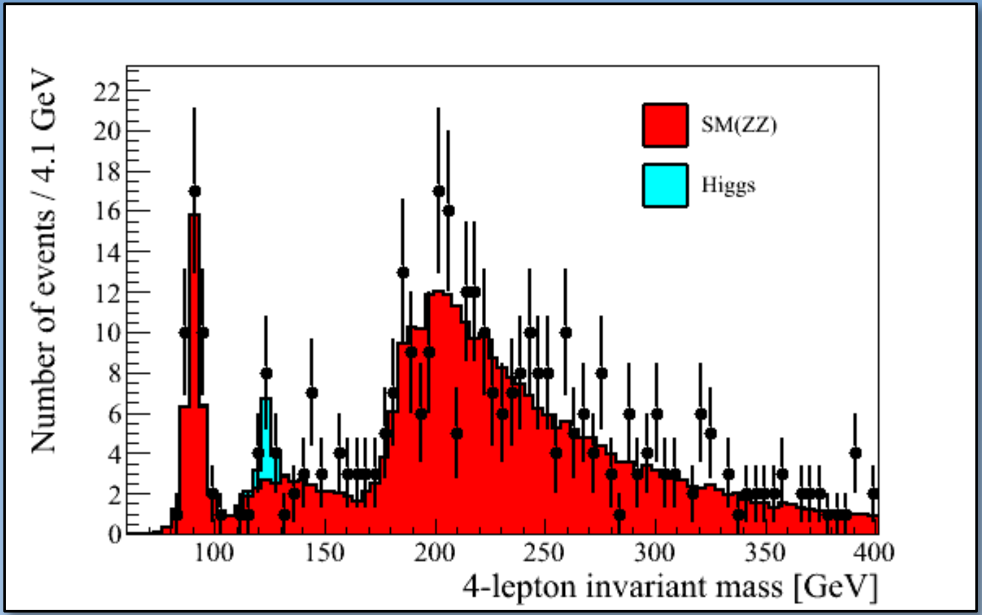
Condense data in **one** number: X

LHC experiments:

$$X(\mu) = -2\ln(Q(\mu)), \ \text{with} \quad Q(\mu) = \frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(\hat{\mu},\hat{\theta})}$$

We'll use something a bit simpler, but same idea

Beyond simple counnting: likelihood ratio test-statistic



$$-2\cdot \log\left(Likelihood\right) = -2\cdot \sum_{bins} \log\left(Poisson(N_{bin}^{data} \mid \mu\cdot f_{bin}^{Higgs} + \alpha\cdot f_{bin}^{SM})\right)$$

$$X = -2\ln(Q), \quad \text{with} \quad Q = \frac{L(\mu_s = 1)}{L(\mu_s = 0)}$$

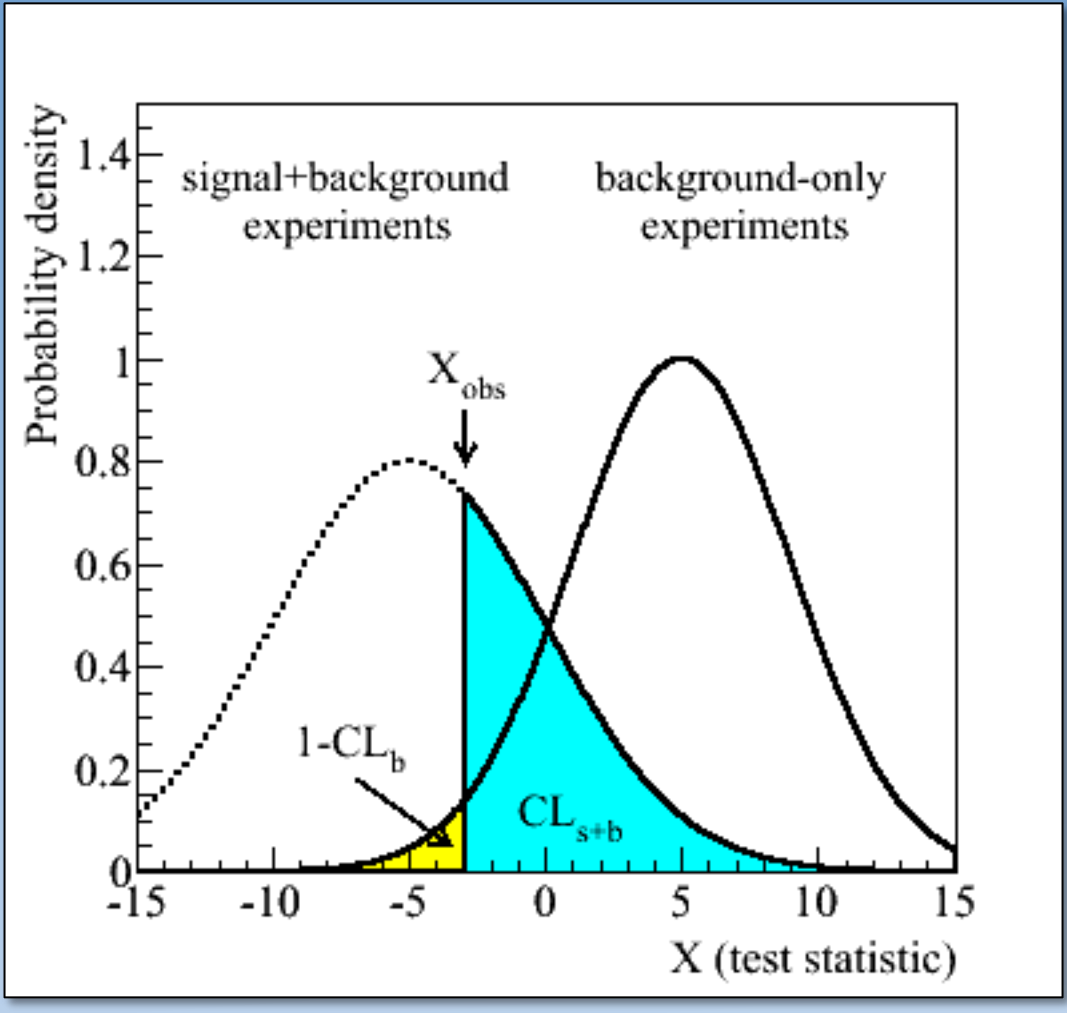
Likelihood assuming μ_s=1 (signal+background)

Hypothesis 1

Hypothesis 0

Likelihood assuming μ_s=0 (only background) signal like

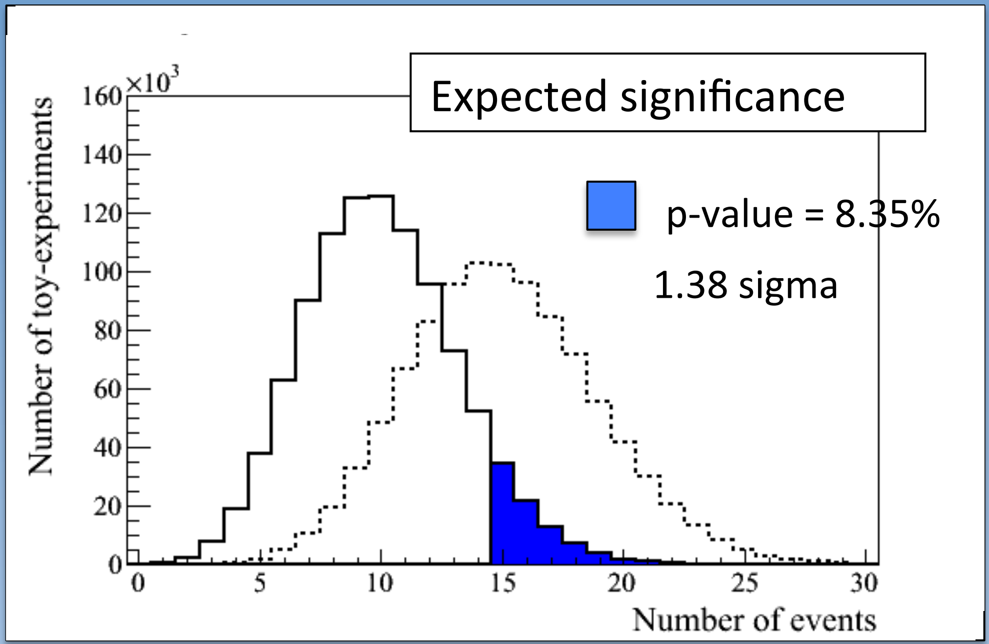
background like



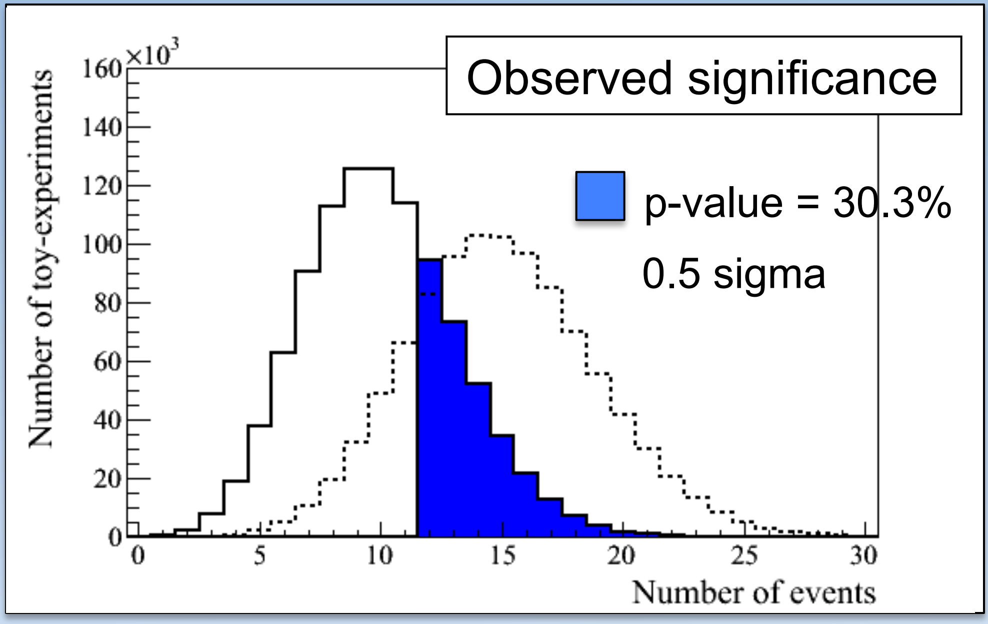
Discovery-aimed: p-value and significance

incompatibiliy with SM-only hypothesis

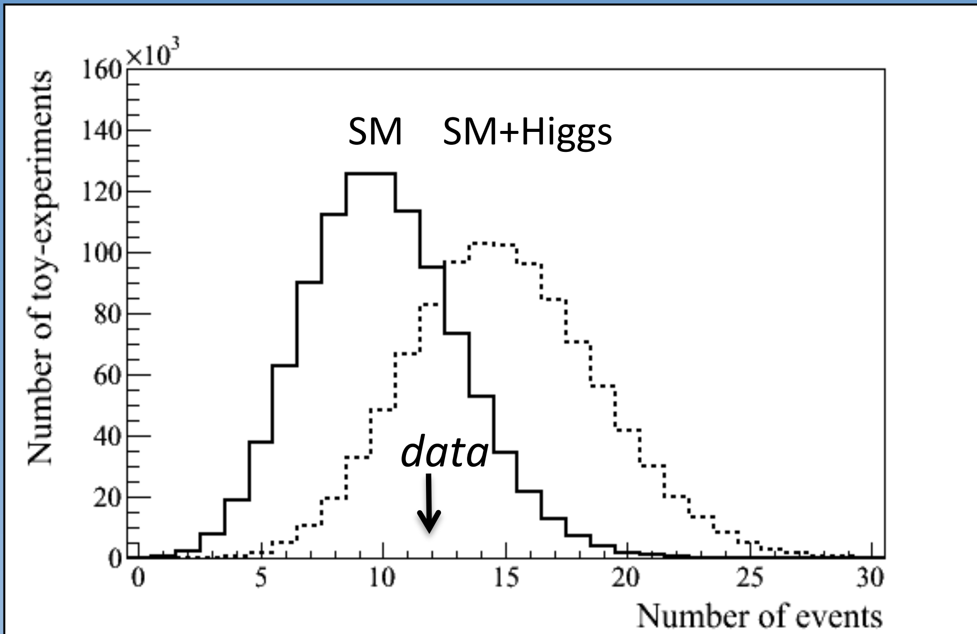
SM	**10**
Higgs	5
Data	12

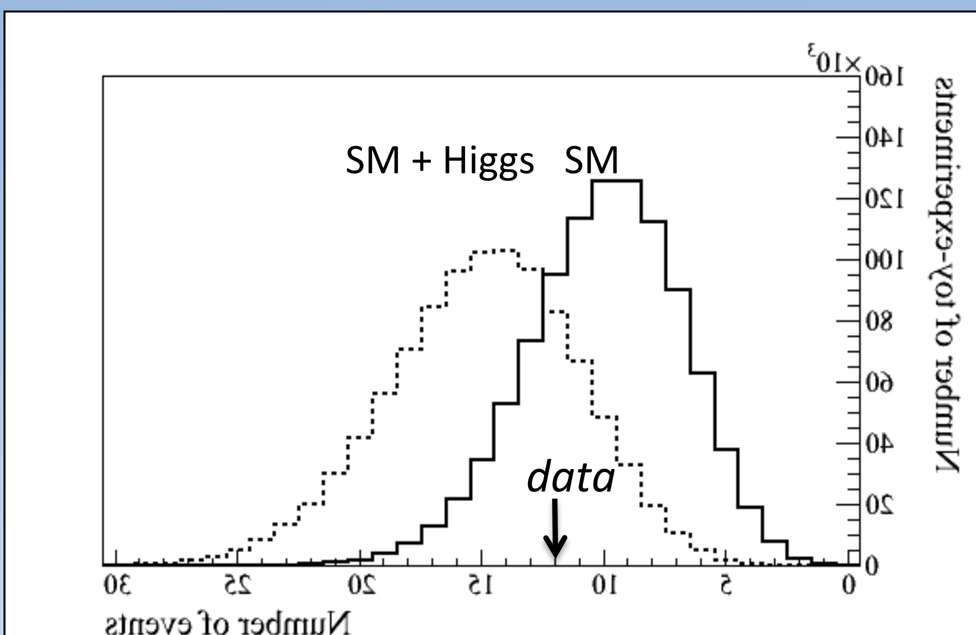


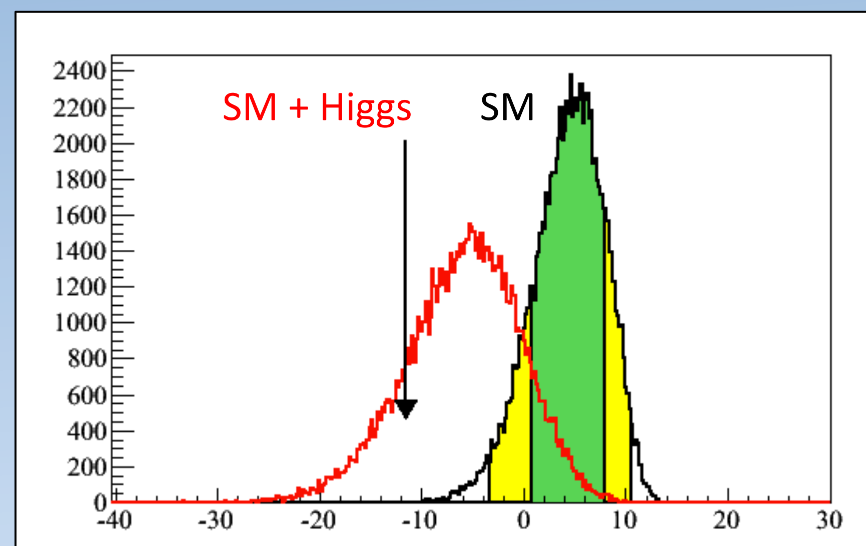
*1) What is the **expected** significance ?*



*2) What is the **observed** significance ?*

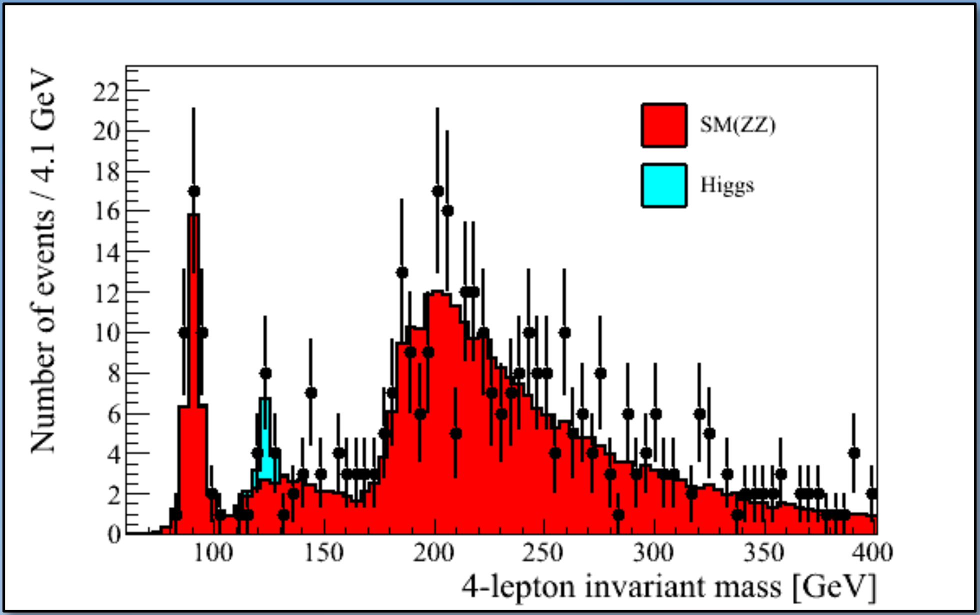

Number of toy-experiments
×10^3
Number of events
SM
SM+Higgs
data


SM + Higgs
SM
data


SM + Higgs
SM

Question: does the window not matter ?

$$X = -2\ln(Q), \quad \text{with} \quad Q = \frac{L(\mu_s = 1)}{L(\mu_s = 0)}$$



X = log(a/b) = log(A) – log(B)

What happens if you add a bin at 300 GeV ?
Will it not dilute the channel like in counting ?

In that bin Lik_{bin} = Constant = C

X = log(a/b) = [log(A) + log(C)] – [log(B) + log (C)]
= log(A) – log(B)

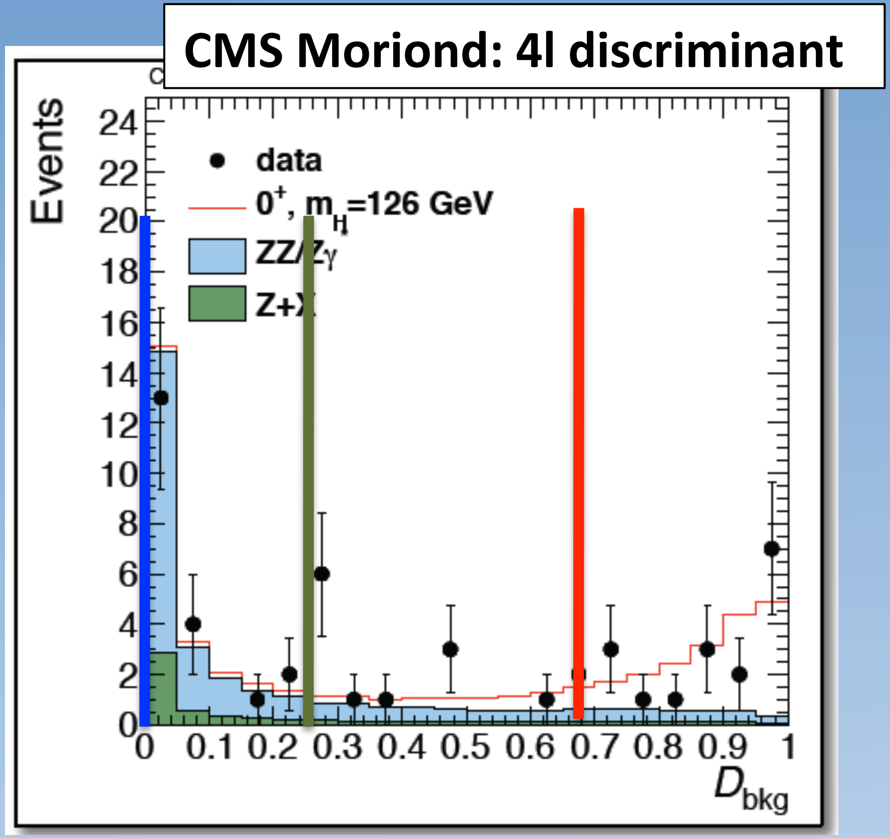
ANY discrimination info is good !

Question: what about more info than mass alone ?

1) Optimal for counting

2) Optimal for LR test stat.

3) Normal procedure



Why: because the 'information' you add below D< 0.25 is maybe difficult to verify in terms of correctness: needs signal descripption in very background-like region: systematics. Need to find optimum.

Note: they still evaluate, like you: $X = -2\ln(Q)$, with $Q = \dfrac{L(\mu_s = 1)}{L(\mu_s = 0)}$

We will use a very simple form for the test statistic

Our exercise (α=1 or from Ex.3):

$$X = -2\ln(Q), \quad \text{with} \quad Q = \frac{L(\mu_s = 1)}{L(\mu_s = 0)} =$$  / 

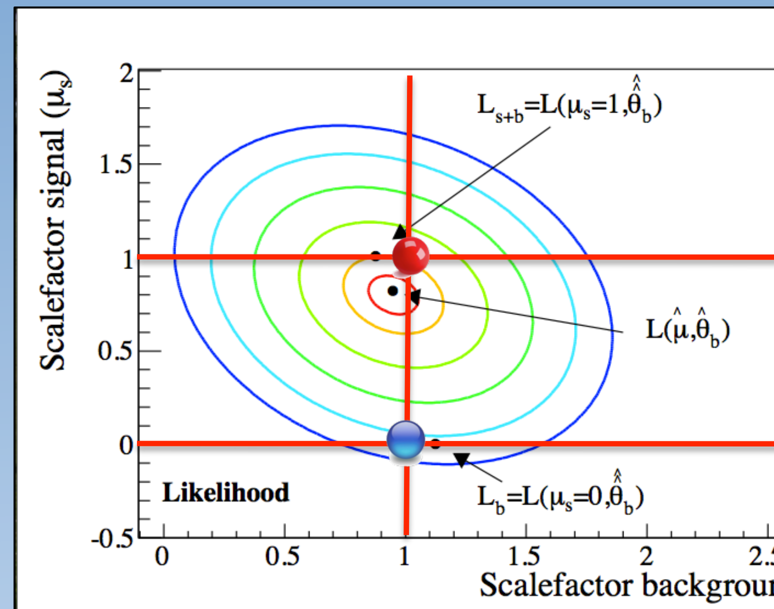
Tevatron-style:

$$X = -2\ln(Q), \quad \text{with} \quad Q = \frac{L(\mu_s = 1, \hat{\hat{\theta}}_{(\mu_s = 1)})}{L(\mu_s = 0, \hat{\hat{\theta}}_{(\mu_s = 0)})}$$

LHC experiments:

$$X(\mu) = -2\ln(Q(\mu)), \quad \text{with} \quad Q(\mu) = \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})}$$

2-dimensional fit (α and μ free)



Note:

α_{bgr} is just one of the nuissance parameters θ in a 'real' analysis

Exercise 4:

Likelihood ratio test statistic (X)

$$X = -2\ln(Q), \quad \text{with} \quad Q = \frac{L(\mu_s = 1)}{L(\mu_s = 0)}$$

→ Likelihood assuming μ_s=1 (signal+background)

→ Likelihood assuming μ_s=0 (only background)

Exercise 4: create the likelihood ration test statistic – beyond simple counting

4.1 Write a routine that computes the likelihood ratio test-statistic for a given data-set

```
double Get_TestStatistic(TH1D *h_mass_dataset, TH1D *h_template_bgr, TH1D *h_template_sig)
```

$$-2Log\left(Likelihood_{(\mu,\alpha\ =\ 1)}\right) = -2\cdot \sum_{bins} \log\left(Poisson(N_{bin}^{data} \mid \mu\cdot f_{bin}^{Higgs} + \alpha\cdot f_{bin}^{SM})\right)$$

Note: log(a/b) = log(a) – log(b)

4.2 Compute the likelihood ratio test-statistic for the ‘real’ data

bonus: Implement the conditional profile likelihood ratio, i.e. find for each of the two hypotheses (μ_s=1 and μ_s=0) the best value for the background scaling (α_{bgr})

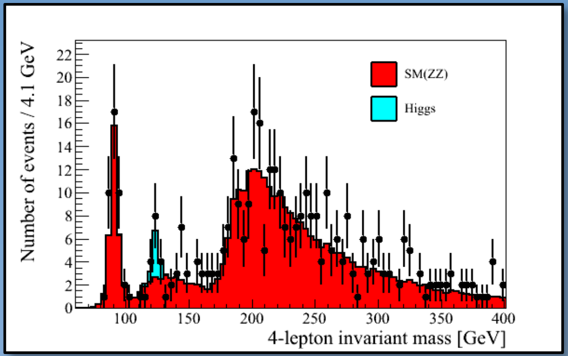
$$X = -2\ln(Q), \quad \text{with} \quad Q = \frac{L(\mu_s = 1, \hat{\hat{\theta}}_{(\mu_s = 1)})}{L(\mu_s = 0, \hat{\hat{\theta}}_{(\mu_s = 0)})}$$

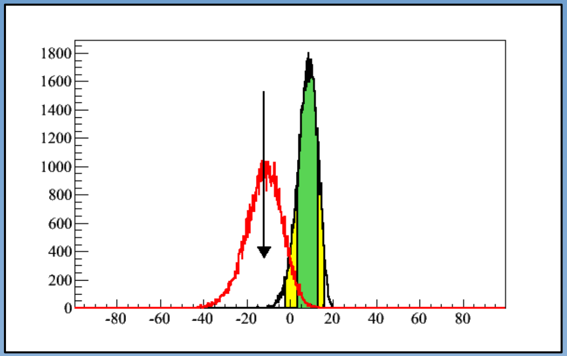
5) - Toy Monte Carlo

- distribution of test statistic for different hypotheses

Exercise 5:

- Generate toy data-sets
- Test statistic distribution





Exercise 5: create toy data-sets

5.1 Write a routine that generates a toy data-set from a MC template (b or s+b)

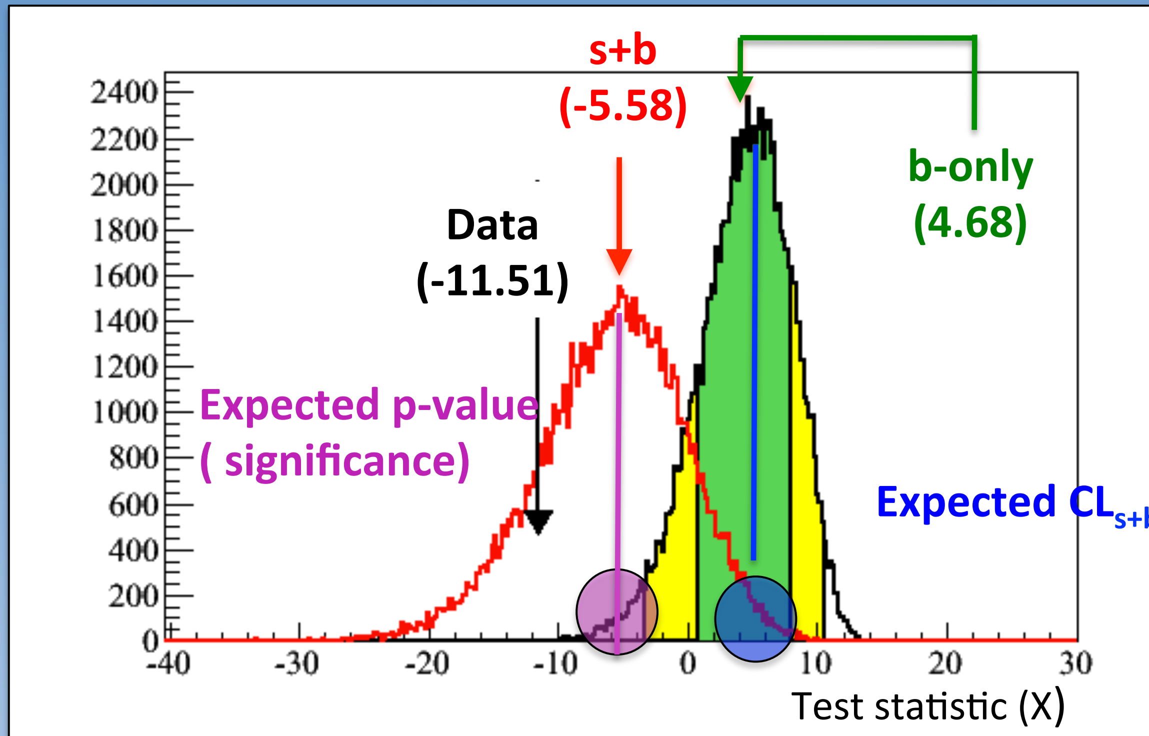
```
TH1D * GenerateToyDataSet(TH1D *h_mass_template)
```

How: Take the histogram h_mass_template and draw a Poisson random number in each bin using the bin content in h_mass_template as the central value. Return the new fake data-set.

5.2 Generate 1000 toy data-sets for *background-only* & get test statistic distribution
Generate 1000 toy data-sets for *signal+background* & get test statistic distribution

→ plot both in one plot

5.3 Add the test-statistic from the data(exercise 4.2) to the plot signal like

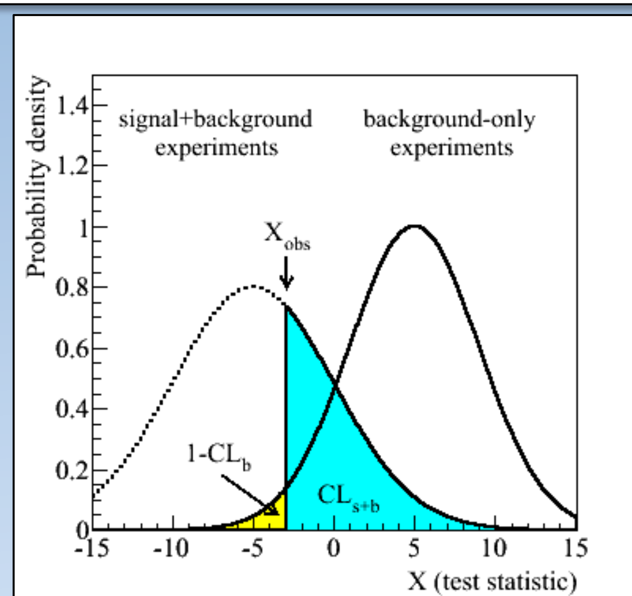


background like

Discovery: $1\text{-}Cl_b < 2.87 \times 10^{-7}$

Incompatibility with b-only hypothesis

signal like



Exclusion: $Cl_{s+b} < 0.05$

Incompatibility with s+b hypothesis

background like

6) Discovery potential

Exercise 6

Summarize separation power: conclusion

Exercise 5: compute p-value

6.1 Compute the p-value or 1-CL_b (under the background-only hypothesis):

- For the average(median) b-only experiment
- For the average(median) s+b-only experiment [expected significance]
- For the data [observed significance]

6.2 Draw conclusions:

- Can you claim a discovery ?
- Did you expect to make a discovery ?
- At what luminosity did/do you expect to be able to make a discovery ?

7) Excluding hypotheses

Exercise 6 continued

Exclude a cross-section for a given Higgs boson mass

Some shortcomings, but we'll use it anyway

$$\sigma_h(m_h) = \zeta \cdot \sigma_h^{SM}(m_h)$$

Scale factor wrt SM prediction

Exercise 6: compute CL_{s+b} and exclude Higgs masses or cross-sections

6.3 Compute the CL_{s+b}:

- For the average(median) s+b experiment
- For the average(median) b-only experiment
- For the data

6.4 Draw conclusions:

- Can you exclude the m_h=200 GeV hypothesis ? What ς can you exclude ?

- Did you expect to be able to exclude the m_h=200 GeV hypothesis ? What ς did you expect to be able to exclude ?

BACKUP

Exercise 8

Pulls

Exercise 8: Is your procedure unbiased and has correct uncertainty ?

8.1 Generate 100 fake data-sets (at $\mu=1$) and extract μ and its error → extract pull

… more later

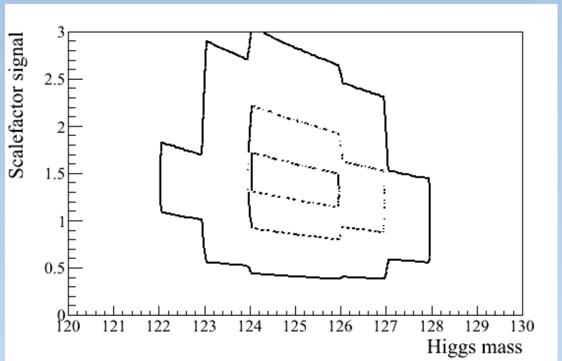
Exercise 9

Mass versus mu

Exercise 9: lmass versus pull

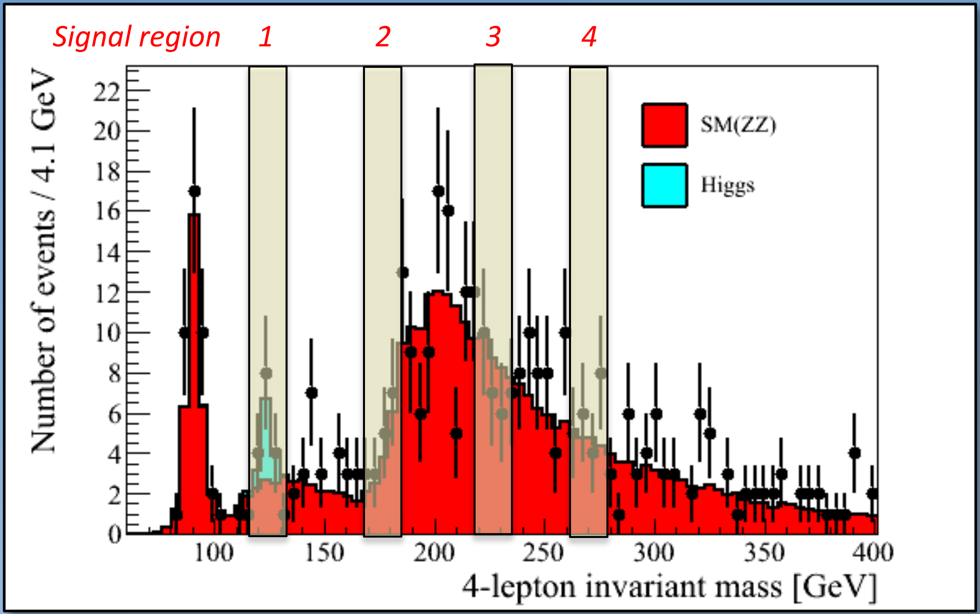
9.1 Do a 2d fit on m_h and mu

… more later



Exercise 3

Look-Elsewhere effect (Trial factor)



Exercise 3: trial factors – global versus local p-values (Look Elsewhere effect)

3.1 Simulate different event yields in 4 possible mass regions. What fraction of LHC experiments is expected to have an excess ≥ 2 sigma: at 125, 175, 225 & 275 GeV ?
Use 4 Poisson Random numbers in a 10 GeV mass window per toy-experiment

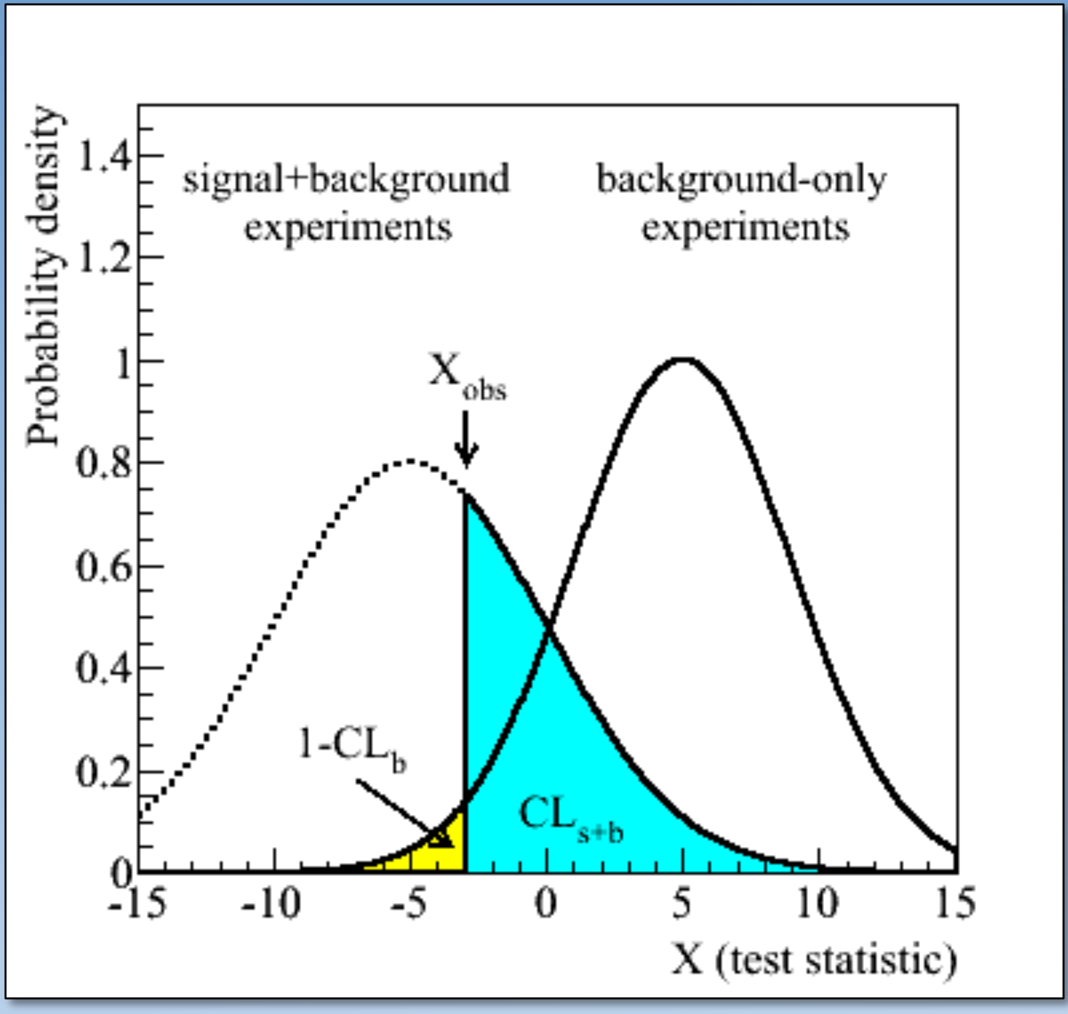
3.2 What is the fraction of toy-experiments that have a maximum excess (at least one out of the 4) above 2 sigma?

↓

To what global significance does a 2 sigma local significance correspond ?

X_{obs}: rules for discovery and exclusion

signal like



background like

Discovery: $1\text{-}Cl_b < 2.87 \times 10^{-7}$
Incompatibility with b-only hypothesis

Exclusion: $Cl_{s+b} < 0.05$
Incompatibility with s+b hypothesis