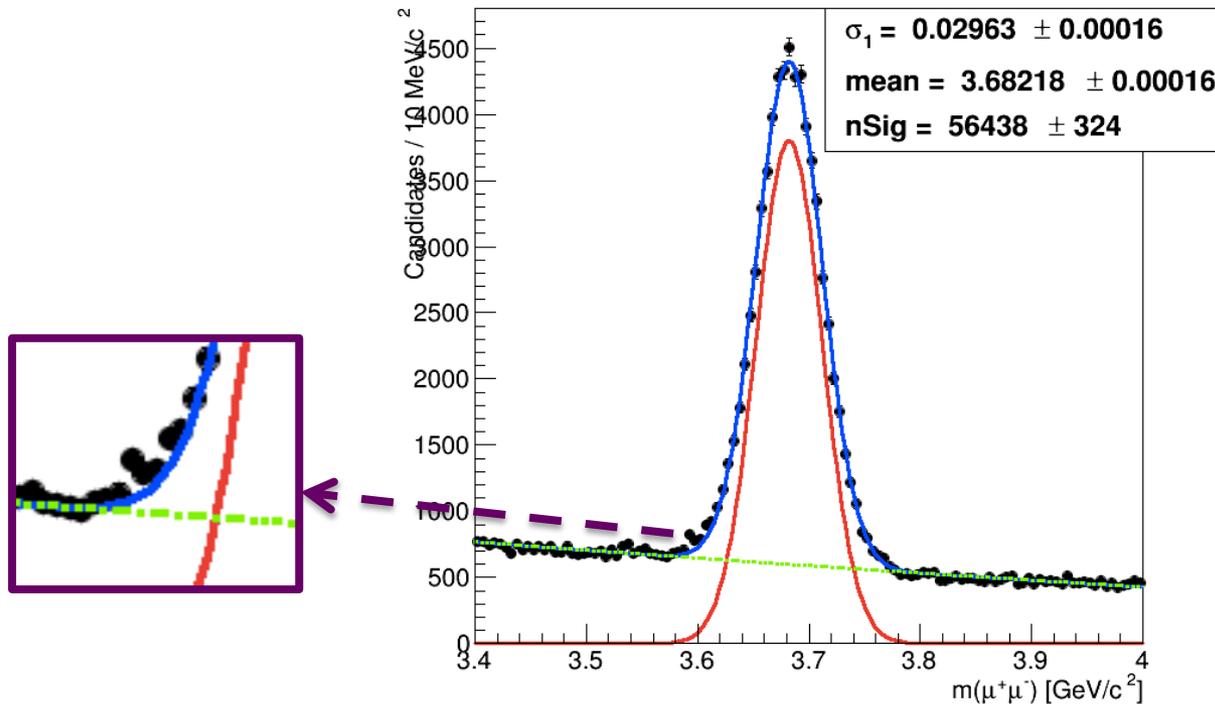


Exercise/Lesson #4b

Scientific Data Analysis Lab course

Alexis Pompili - UniBA

Nella precedente esercitazione abbiamo interpolato la distribuzione di massa invariante $m(\mu^+\mu^-)$ contenuta nel file *psiprime_bin9_histo.root* :



Si noti che **la coda a valori bassi di massa inv. per il picco di segnale non e' ben descritta dalla gaussiana.**

In effetti e' preferibile – invece di una gaussiana - usare una singola **funzione Crystal Ball**, la quale integra una funzione gaussiana rappresentante la risoluzione sperimentale con una funzione potenza rappresentante la **coda radiativa** (dovuta a **bremstrahlung interna**, un processo di QED con un muone che “emette” radiazione di stato finale).

Crystal ball function

The Crystal Ball function, named after the Crystal Ball Collaboration (hence the capitalized initial letters), is a probability density function commonly used to model various lossy processes in high-energy physics. It consists of a Gaussian core portion and a power-law low-end tail, below a certain threshold. The function itself and its first derivative are both continuous.

The Crystal Ball function is given by:

$$f(x; \alpha, n, \bar{x}, \sigma) = N \cdot \begin{cases} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right), & \text{for } \frac{x-\bar{x}}{\sigma} > -\alpha \\ A \cdot \left(B - \frac{x-\bar{x}}{\sigma}\right)^{-n}, & \text{for } \frac{x-\bar{x}}{\sigma} \leq -\alpha \end{cases}$$

where

$$A = \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right)$$

$$B = \frac{n}{|\alpha|} - |\alpha|$$

N is a normalization factor and α , n , \bar{x} and σ are parameters which are fitted with the data.

Per avere un'idea dell'effetto dei due parametri descrittivi la coda si osservi la seguente figura (tratta da *CMS AN-14-003*): 

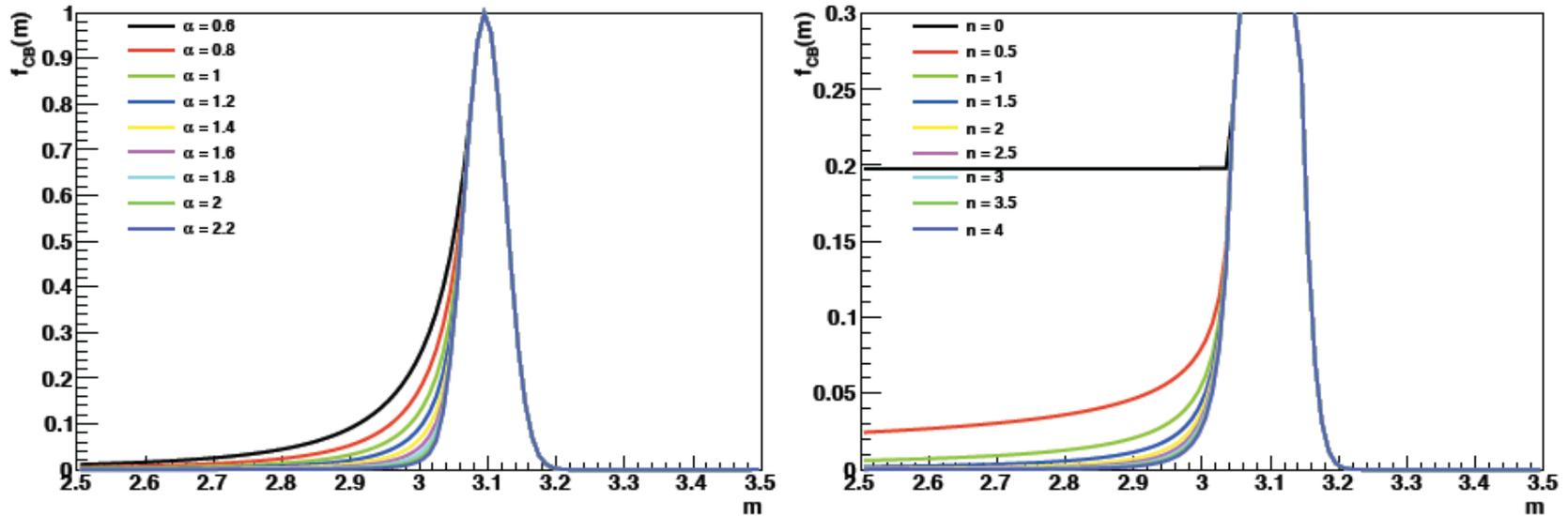


Figure 16: Shapes of the CB function for several different (n_{CB}, α_{CB}) values, fixing $n_{CB} = 2.5$ (left) or $\alpha_{CB} = 1.8$ (right).

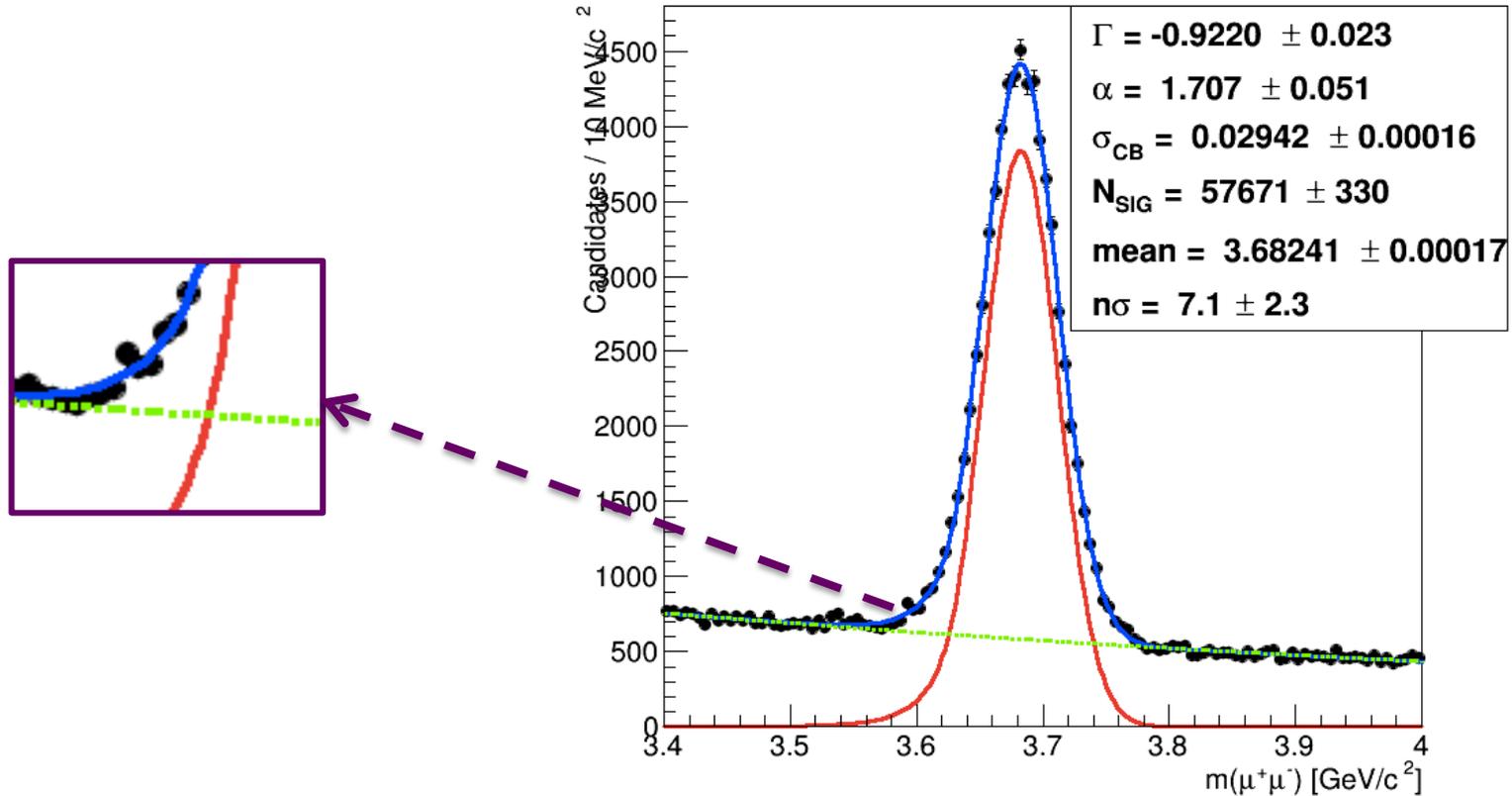
Inoltre, empiricamente, il fondo combinatorio della massa inv. dei due muoni risulta essere descritto da una funzione esponenziale, quindi con un parametro in meno rispetto alla polinomiale di ord.1!

Si puo' quindi provare anche a passare dall'esponenziale alla retta.

Si tenga conto che in *RooFit* si tratta di costruire modelli di segnale e fondo con funzioni del tipo *RooCBSShape* e *RooExponential* al posto di funzioni del tipo *RooGaussian* e *RooChebyshev* rispettivamente.

L'esercizio consiste quindi nel:

- 1) ripetere l'interpolazione **usando, come modello per il segnale, una funzione Crystal Ball e, come modello per il fondo, la funzione esponenziale, ottenendo:**

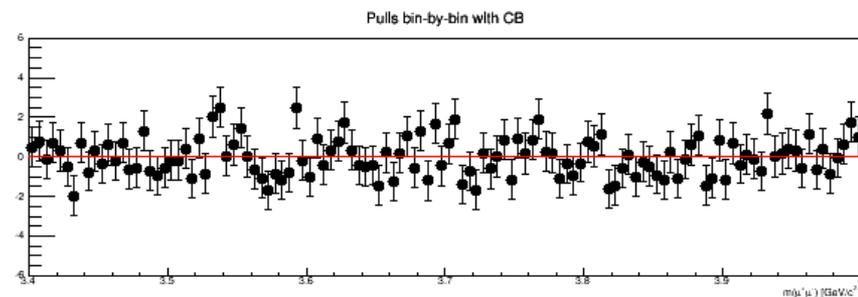
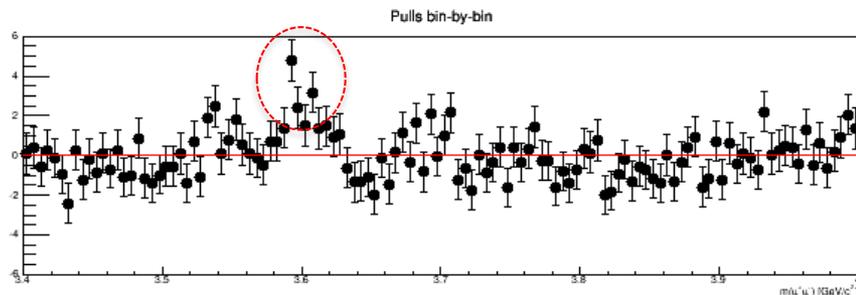
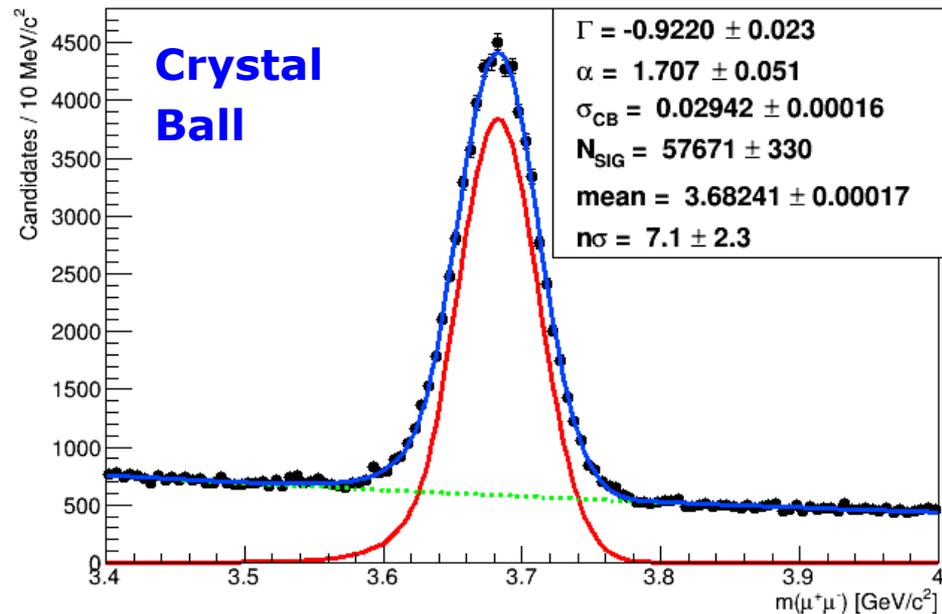
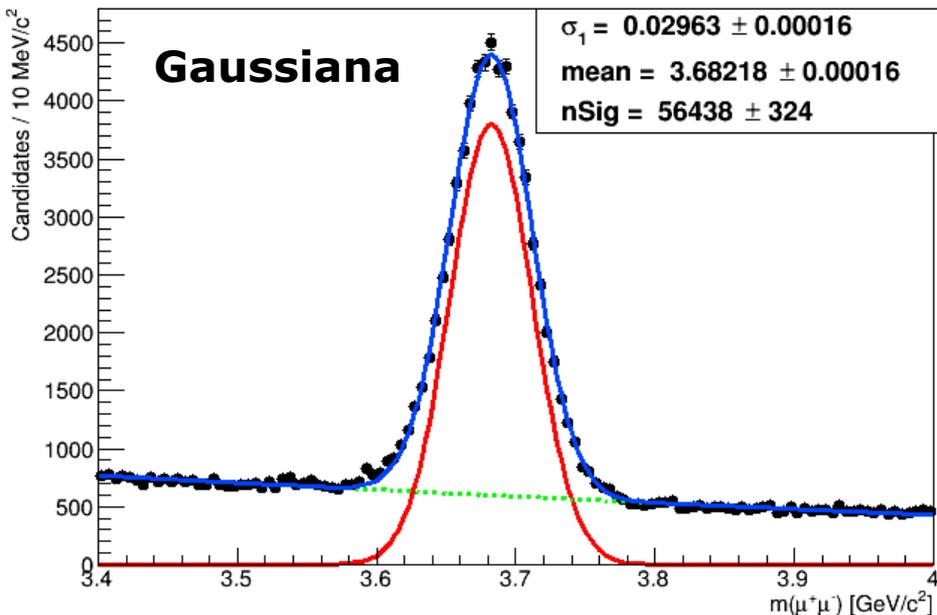


- 2) rifare l'interpolazione usando la polinomiale di ord.1 per il fondo, mentre si mantiene una Crystal Ball per il segnale.
Discutere come varia la stima dei candidati ψ di segnale nei 3 fit.

This is the comparison between the two fits; **the pulls' method allows to monitor the *goodness-of-fit*** and how it changes between the two models used in the fit.

ψ' for y in $[-0.8, -0.6]$

ψ' for y in $[-0.8, -0.6]$



In the right fit a one-sided CB replaces the Gaussian and a simple exponential is used for the bkgd

I plot alla slide precedente si ottengono aggiungendo il codice relativo alle *pull* (vedi definizione alla slide successiva):

```
////////// goodness of fits with pulls for each bin :  
//  
RooPlot *framePull = xVar.frame();  
framePull->SetTitle("Pulls bin-by-bin");  
framePull->addObject( (TObject*)xframe->pullHist(), "p" );  
framePull->SetMinimum(-6);  
framePull->SetMaximum(6);  
//  
myC->Divide(0,2);  
myC->cd(2);  
gPad->SetPad(0.,0.,1.,0.3);  
framePull->Draw();  
TLine *line = new TLine(3.4,0.,4.,0.);  
line->SetLineColor(2);  
line->Draw("same");  
myC->cd(1);  
gPad->SetPad(0.,0.3,1.,1.);  
xframe->Draw();  
//
```

Proprieta' delle pull:

- 1) dai plot alla slide precedente si evince come **l'errore sulla pull bin-by-bin sia unitario** (il motivo e' spiegato alla slide successiva);**
- 2) La proiezione sull'asse delle ordinate delle pull bin-by-bin deve fornire una distribuzione complessiva attesa essere una gaussiana standard (media 0 e varianza 1).**

Proprieta' delle pull:

Lo **scarto normalizzato** e' simile alla radice di un chi-quadrato corredato di segno (motivo per il quale tecnicamente e' uno "**pseudo chi-quadrato**"). Lo denoto con $\pm\sqrt{\chi^2}$

L'istogramma degli scarti normalizzati desidero che abbia lo stesso # di bin dell'istogramma della distribuzione di massa invariante.

E' necessario poi rappresentare lo scarto corredato dalla propria barra di errore!

$$\pm\sqrt{\chi^2}(i) = \frac{x_S^i - x_T^i}{\sigma_i} \equiv \frac{N_i - F_i}{\sqrt{N_i}} \quad \dots \text{essendo : } \begin{cases} N_i = \# \text{ candidati nel bin } i\text{-esimo dell'istogramma} \\ F_i = \# \text{ candidati nel bin } i\text{-esimo atteso} \\ \text{(assumendo corretto il modello di fit)} \end{cases}$$

Diagram illustrating the components of the normalized residual:

- x_S^i : Valore sperimentale
- x_T^i : Valore atteso (teorico)
- σ_i : Incertezza associata al valore sperimentale

L'incertezza (errore) sullo scarto normalizzato (per ogni bin) si calcola applicando l'usuale **legge di proporzazione degli errori casuali** (tralascio l'indice per alleggerire la notazione):

$$\sigma_{\pm\sqrt{\chi^2}}^2 = \left(\frac{d}{dN} \left(\frac{N-F}{\sqrt{N}} \right) \right)^2 \cdot (\sqrt{N})^2 = \left(\frac{\sqrt{N} - \frac{1}{2\sqrt{N}}(N-F)}{N} \right)^2 \cdot N = \left(\frac{N - \frac{1}{2}(N-F)}{N\sqrt{N}} \right)^2 \cdot N = \left(\frac{1}{2} \left(\frac{N+F}{N} \right) \right)^2$$

In conclusione: $\sigma_{\pm\sqrt{\chi^2}} = \frac{1}{2} \frac{N+F}{N}$ e, ad **alta statistica** (N grande) si ha: $N \approx F \Rightarrow \sigma_{\pm\sqrt{\chi^2}} \approx 1$