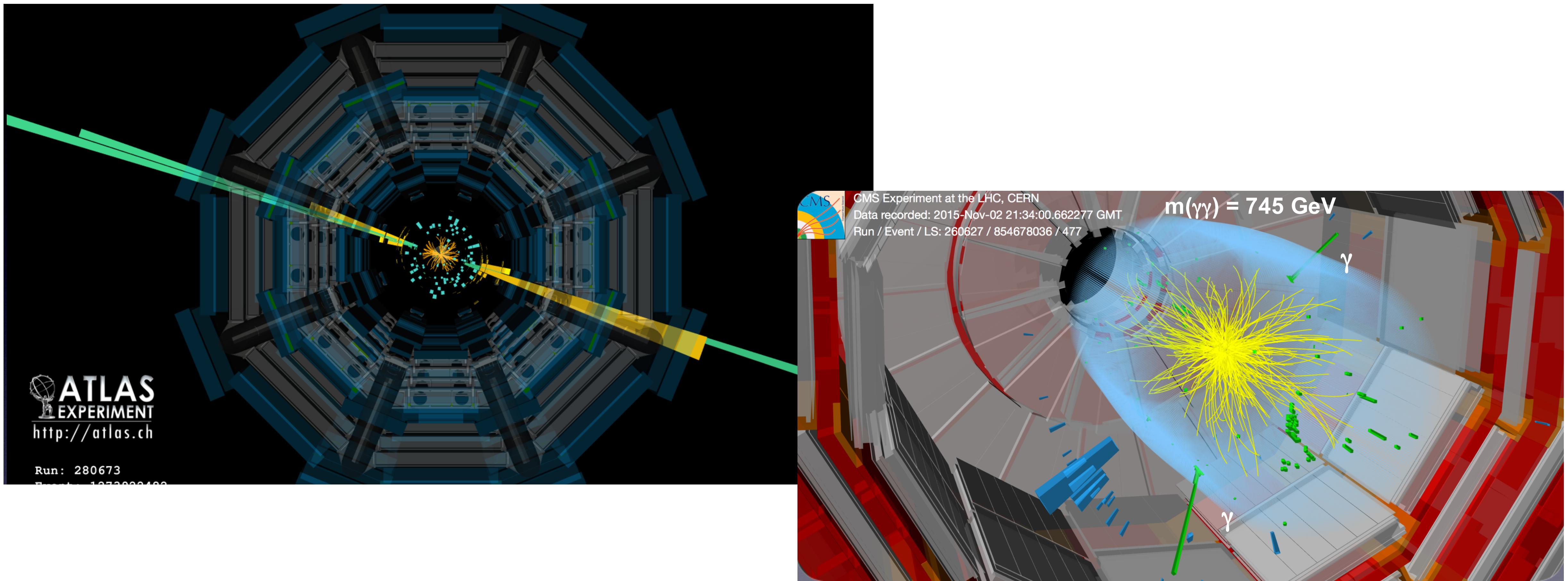

ESAME FINALE

750 GEV, NUOVA RISONANZA?

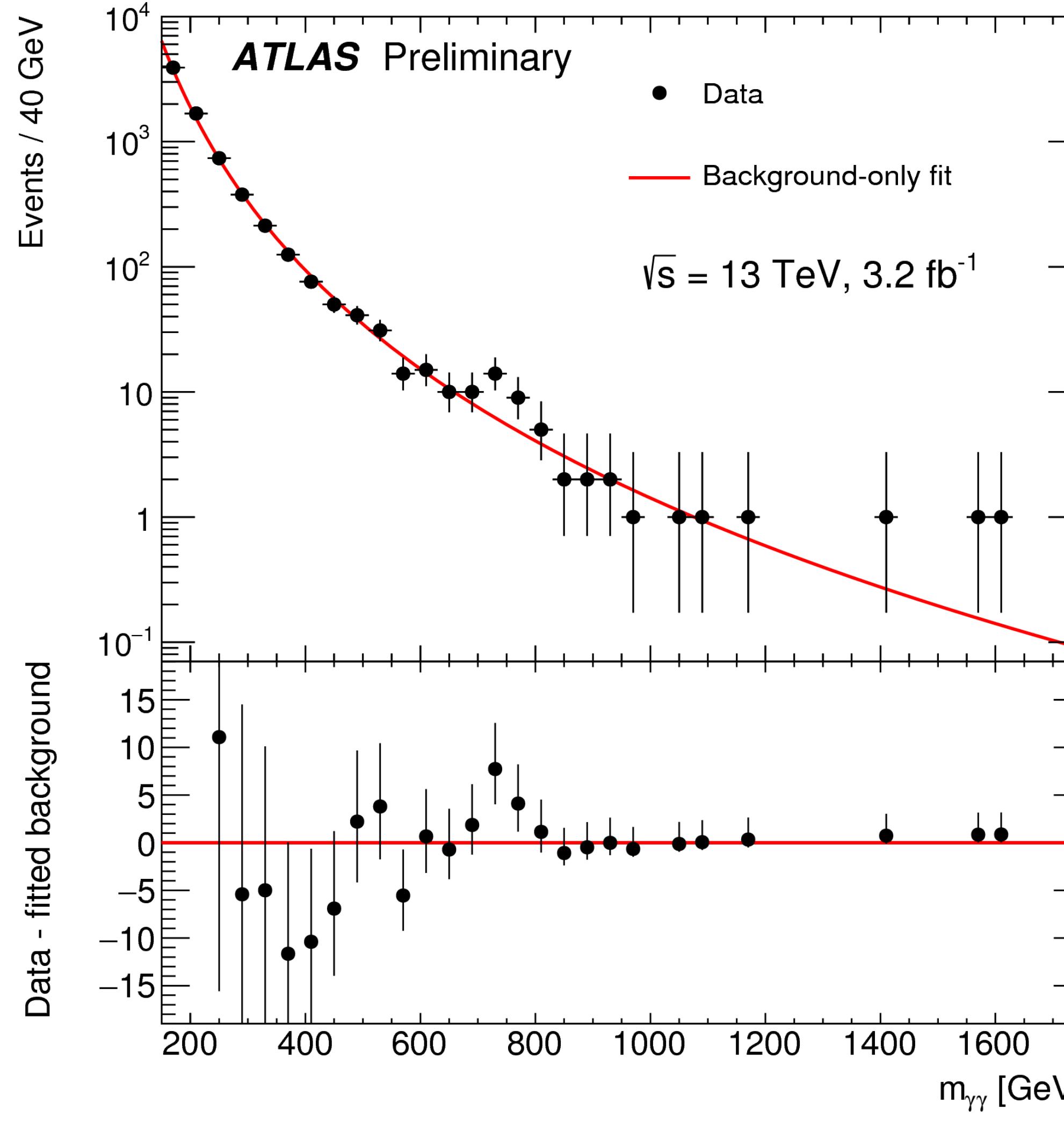
Nel 2015 ci fu molta agitazione dovuta al fatto che sia Atlas che CMS osservarono un eccesso di eventi con due fotoni attorno ad una massa di ~ 750 GeV.

Noi ci porremo nei panni di un ipotetico terzo esperimento, con l'onere di confermare/smentire l'esistenza di questa nuova particella

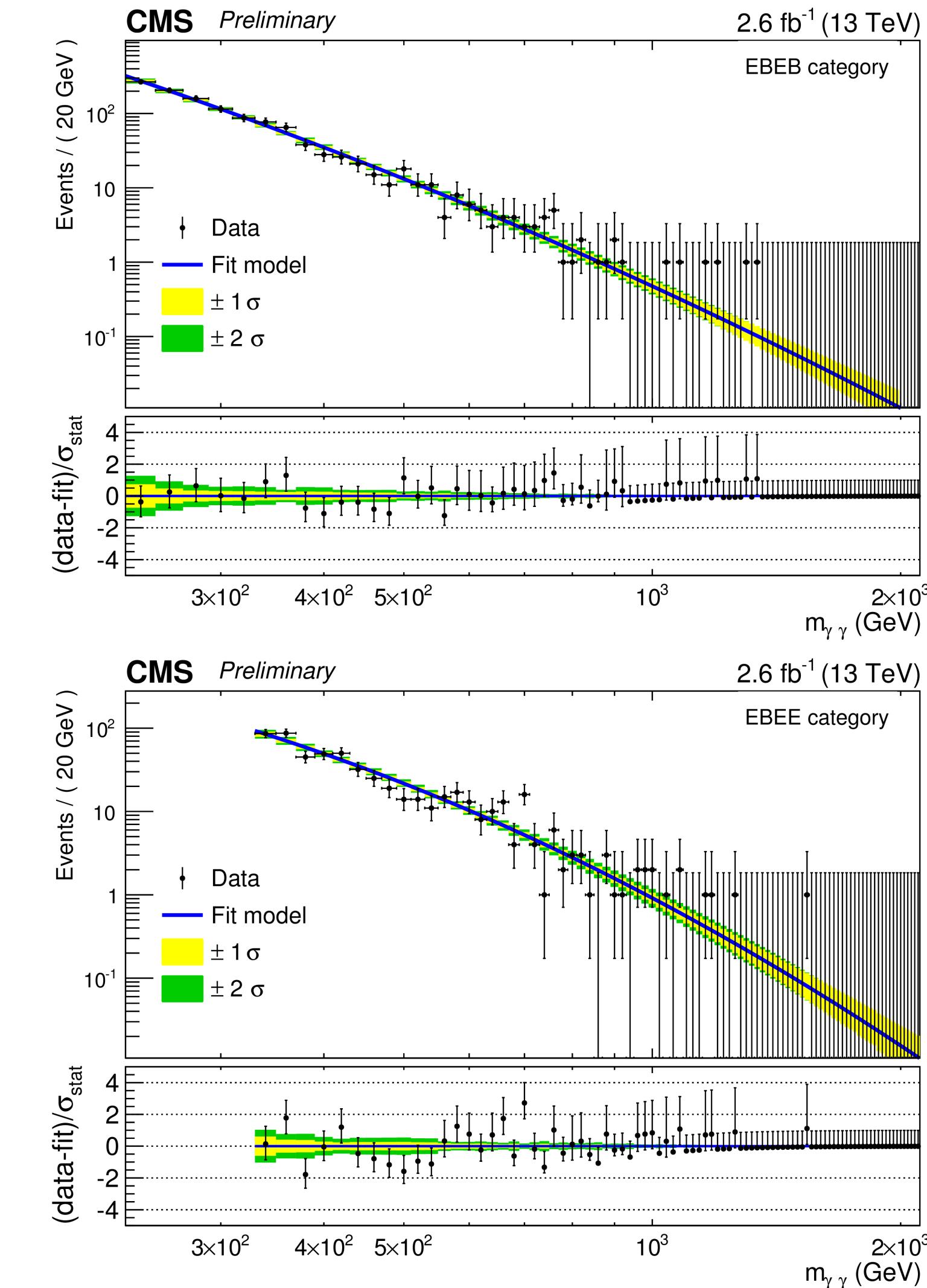


750 GEV, NUOVA RISONANZA?

<https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2015-081/>



<http://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/EXO-15-004/index.html>

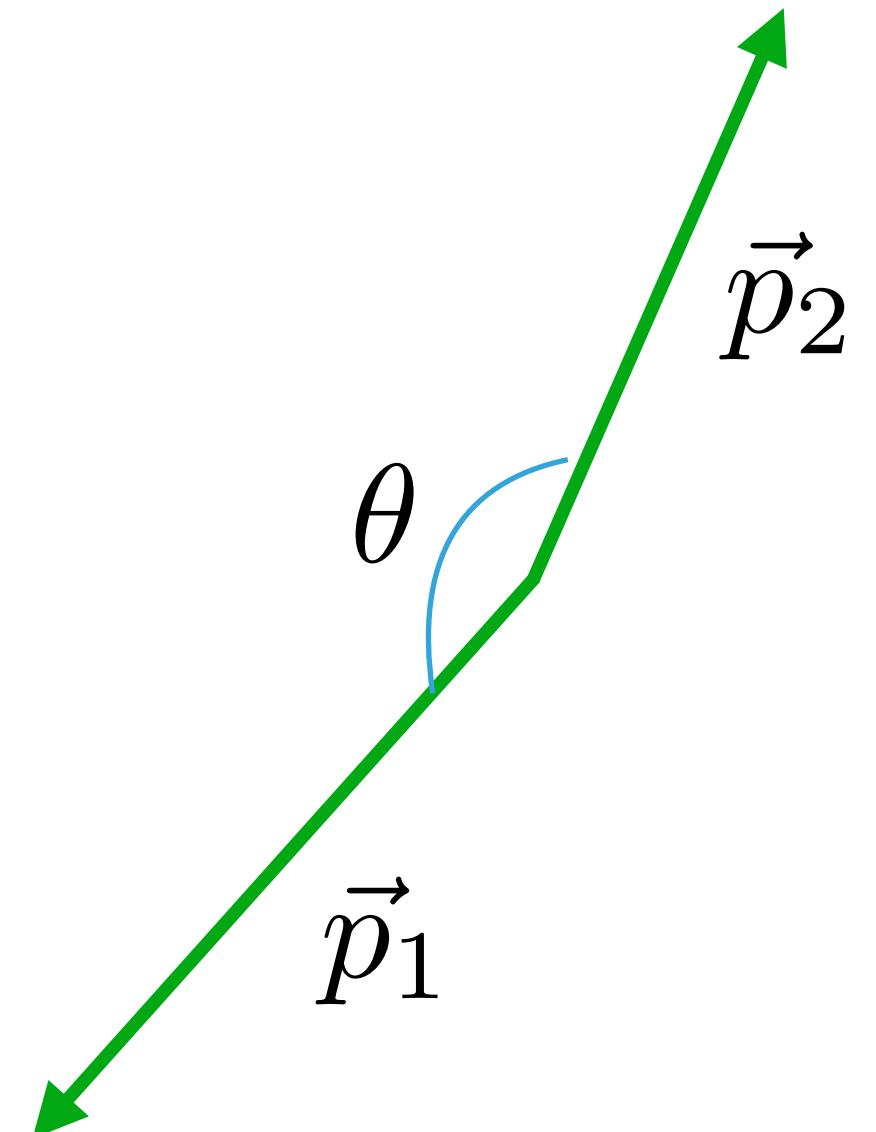
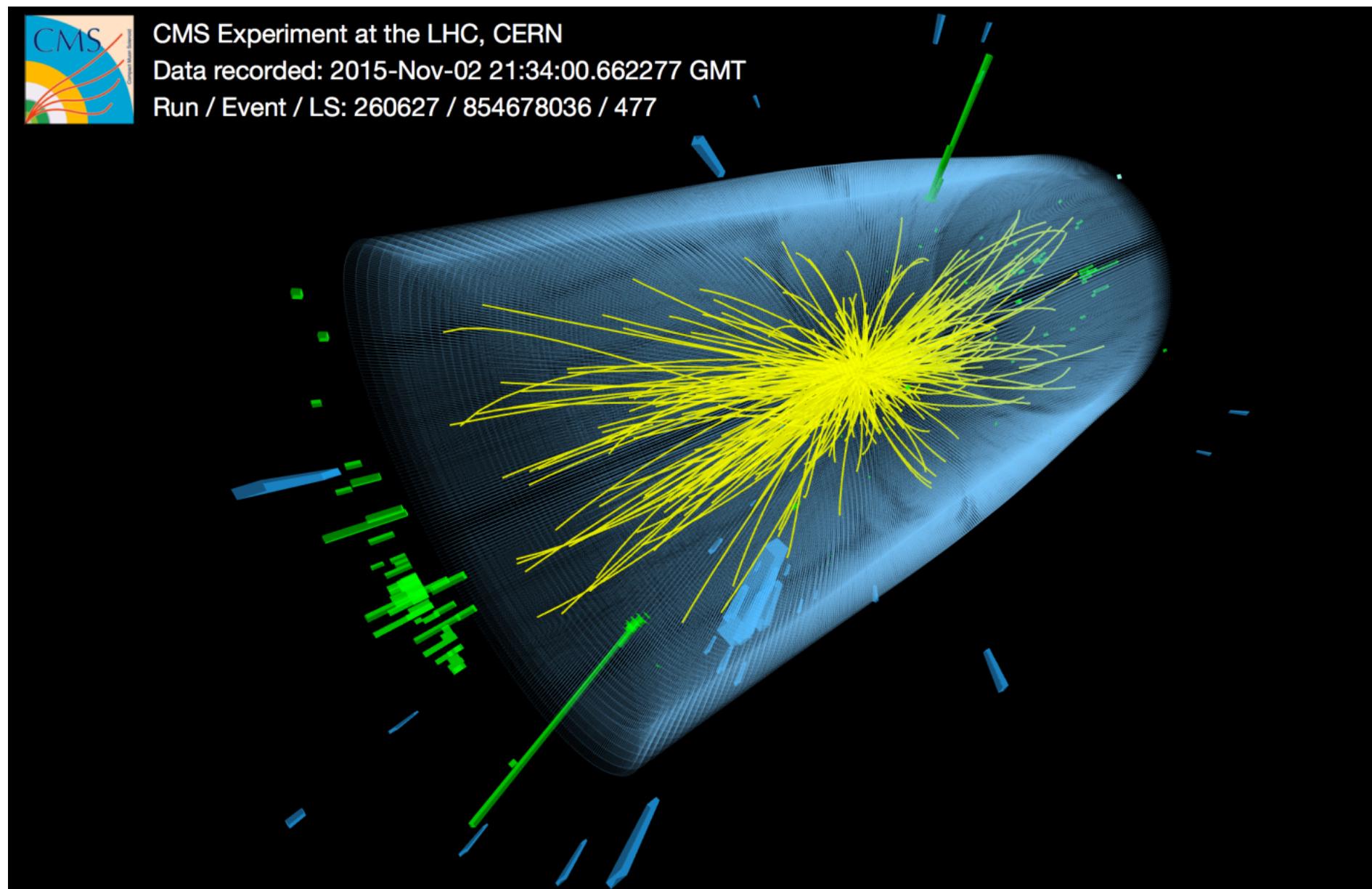


IL DATASET

Avrete a vostra disposizione due set di eventi, entrambi contenenti eventi con due fotoni.

Di entrambe i fotoni avrete a disposizione il quadri-momento, dal quale ricostruire la massa invariante

$$M = \sqrt{(p_1 + p_2)^2} = \sqrt{p_1^2 + p_2^2 + 2p_1^\mu p_2^\mu} = \sqrt{2E_1 E_2 (1 - \cos \theta)}$$



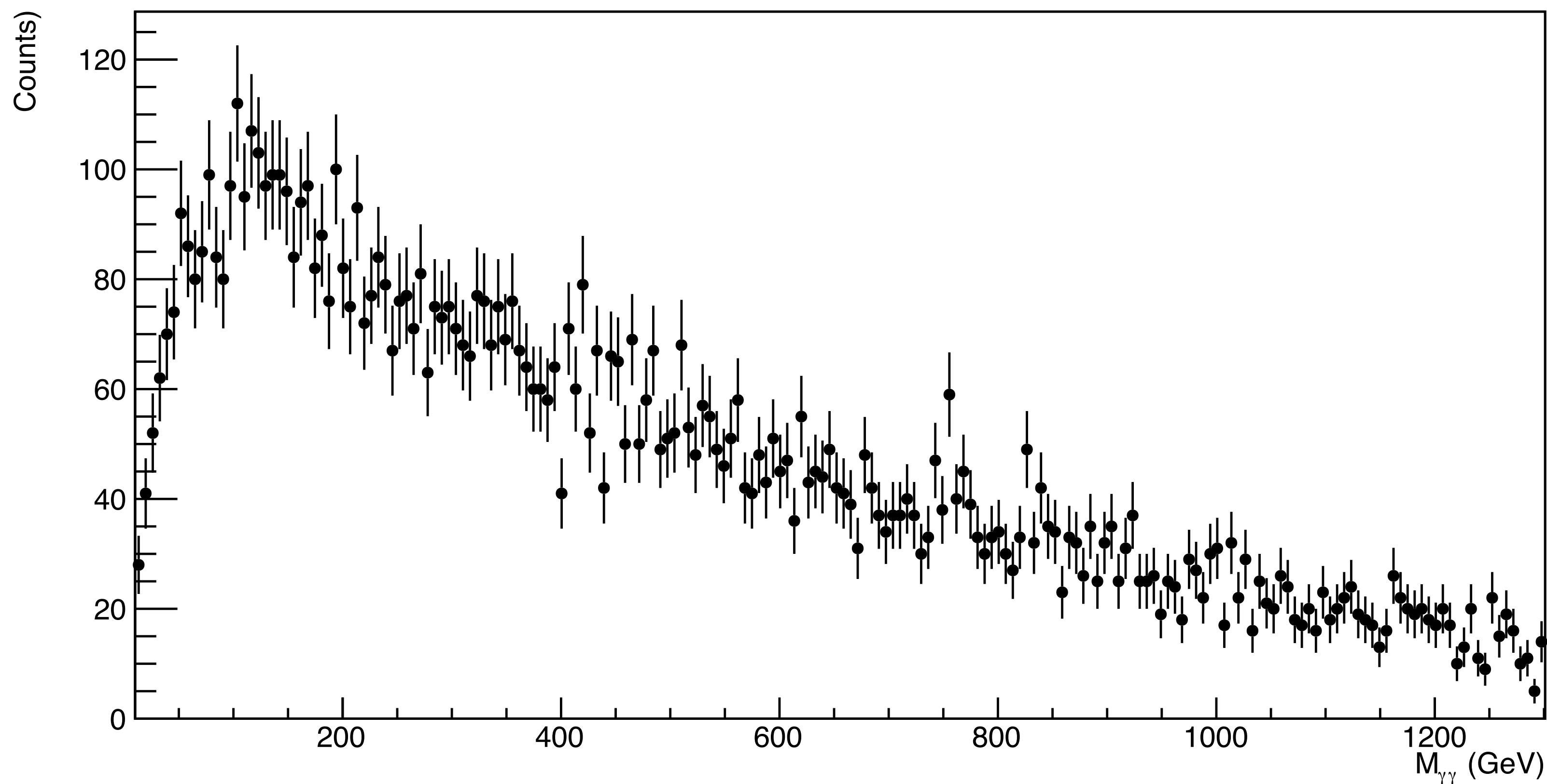
I DATASET

Avrete a vostra disposizione due set di eventi misurati, entrambi contenenti eventi con due fotoni.

Avete a disposizione anche un set di eventi con solo background proveniente dalla simulazione MC.

Di entrambe i fotoni avrete a disposizione il quadri-momento, dal quale ricostruire la massa invariante

Dataset 1: il vostro capo vi chiama nel pieno della notte e vi urge un controllo il prima possibile, iniziate a processare i dati e quando siete al 10% della statistica totale decidete di dare un'occhiata.



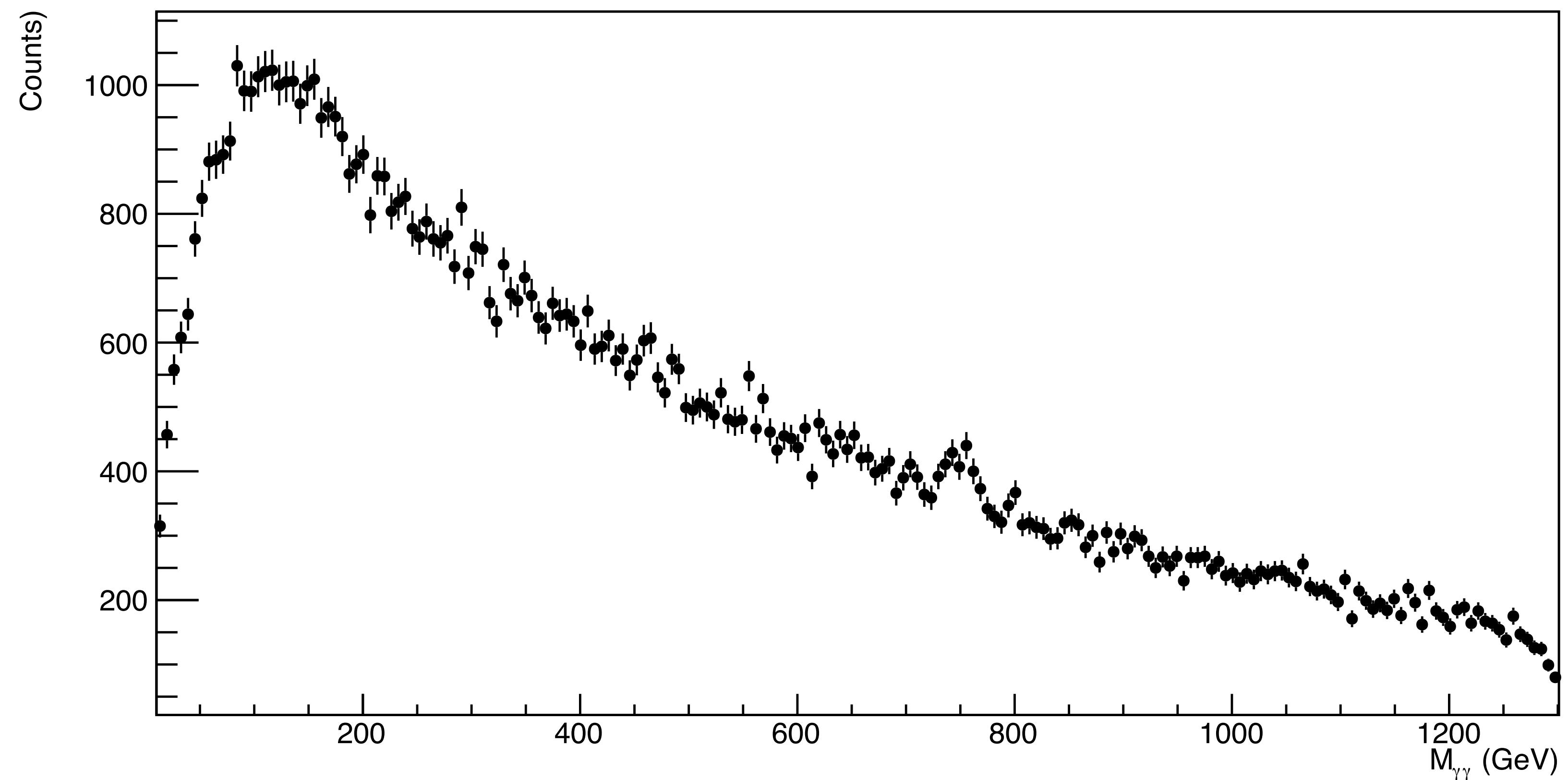
I DATASET

Avrete a vostra disposizione due set di eventi misurati, entrambi contenenti eventi con due fotoni.

Avete a disposizione anche un set di eventi con solo background proveniente dalla simulazione MC.

Di entrambe i fotoni avrete a disposizione il quadri-momento, dal quale ricostruire la massa invariante

Dataset 2: dopo qualche giorno avete ultimato il processamento di tutta la statistica e potete finalmente dare la risposta finale.



L'ESERCIZIO

Per ciascuno dei due dataset vi viene chiesto di:

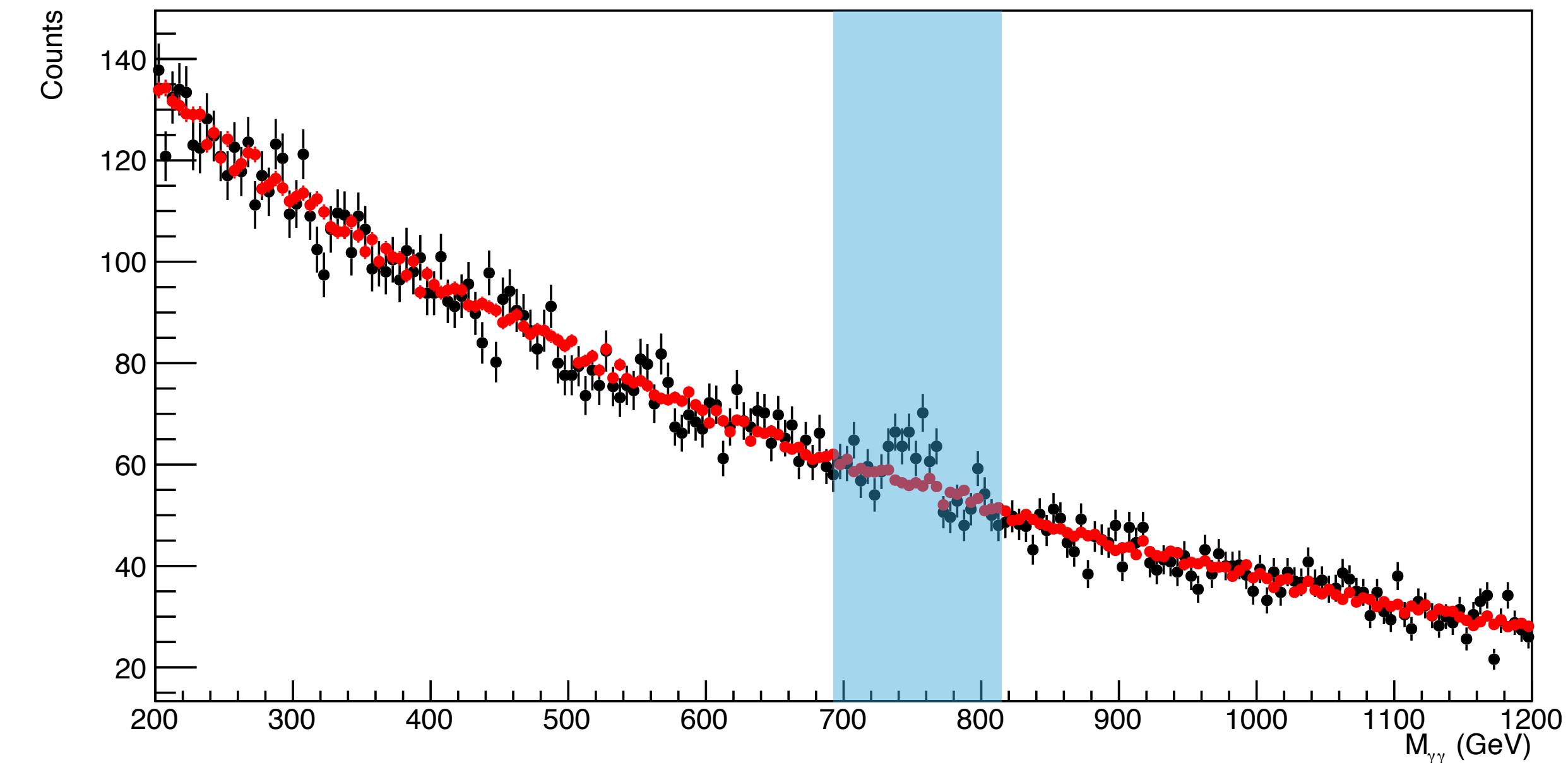
- Stimare la presenza o meno di un segnale alla massa di ~750 GeV e riportarne la significanza statistica e p-value.
 - In caso la significanza sia troppo bassa, porre un upper-limit al numero di eventi di segnale.
 - In caso la significanza sia sufficiente, stimare il numero di eventi di segnale e riportarne l'incertezza.

Potete scegliere diversi metodi per calcolare il numero di eventi di segnale:

- Cut & count
Scegliere una finestra in massa, contare il numero di eventi misurati e sottrarre la stima di quanti eventi di fondo ci sono in quella finestra.

La stima del fondo puo' avvenire:

- dal MC



L'ESERCIZIO

Per ciascuno dei due dataset vi viene chiesto di:

- Stimare la presenza o meno di un segnale alla massa di ~ 750 GeV e riportarne la significanza statistica e p-value.
 - In caso la significanza sia troppo bassa, porre un upper-limit al numero di eventi di segnale.
 - In caso la significanza sia sufficiente, stimare il numero di eventi di segnale e riportarne l'incertezza.

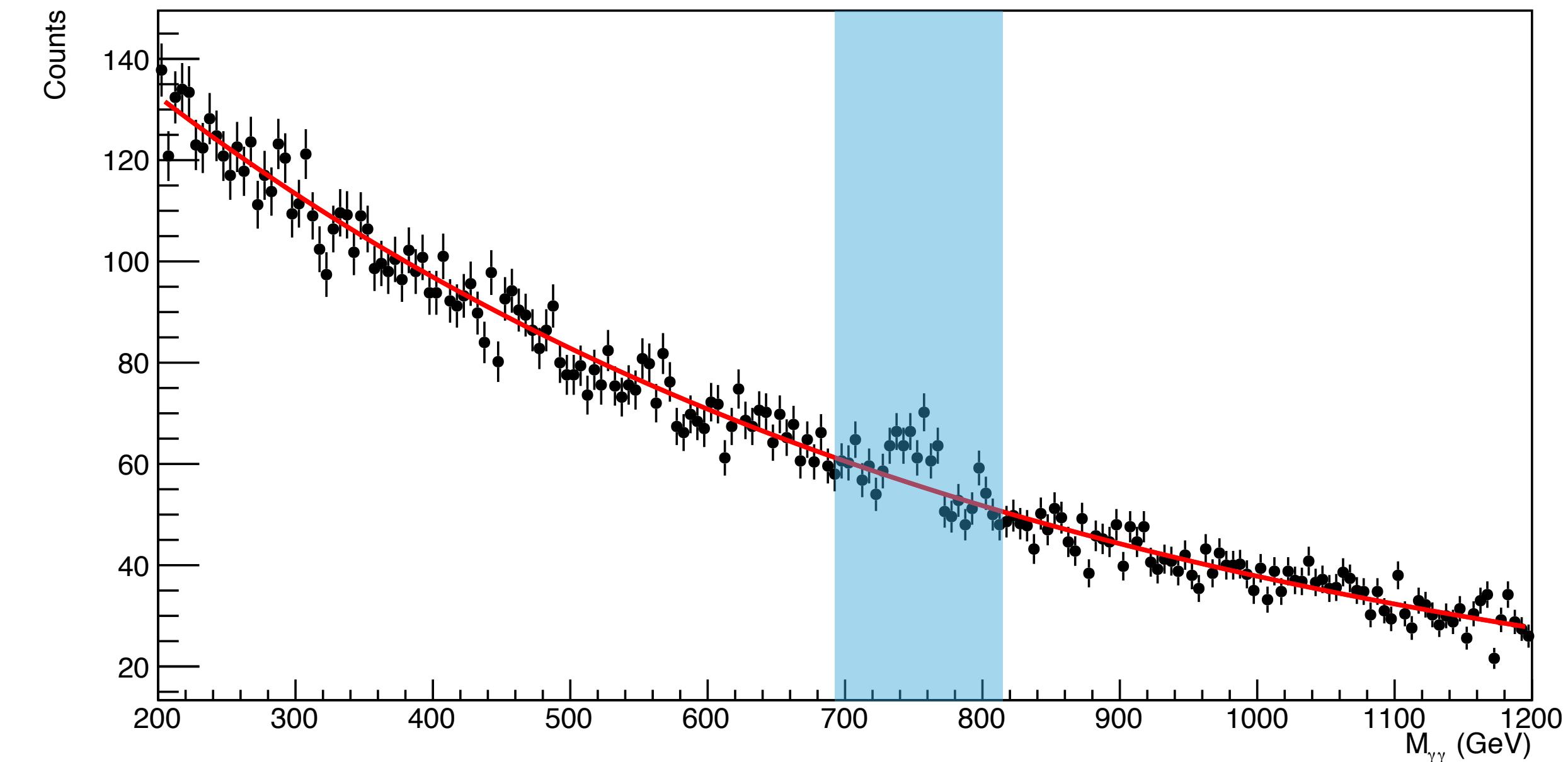
Potete scegliere diversi metodi per calcolare il numero di eventi di segnale:

- Cut & count

Scegliere una finestra in massa, contare il numero di eventi misurati e sottrarre la stima di quanti eventi di fondo ci sono in quella finestra.

La stima del fondo puo' avvenire:

- dal MC
- da un Fit analitico



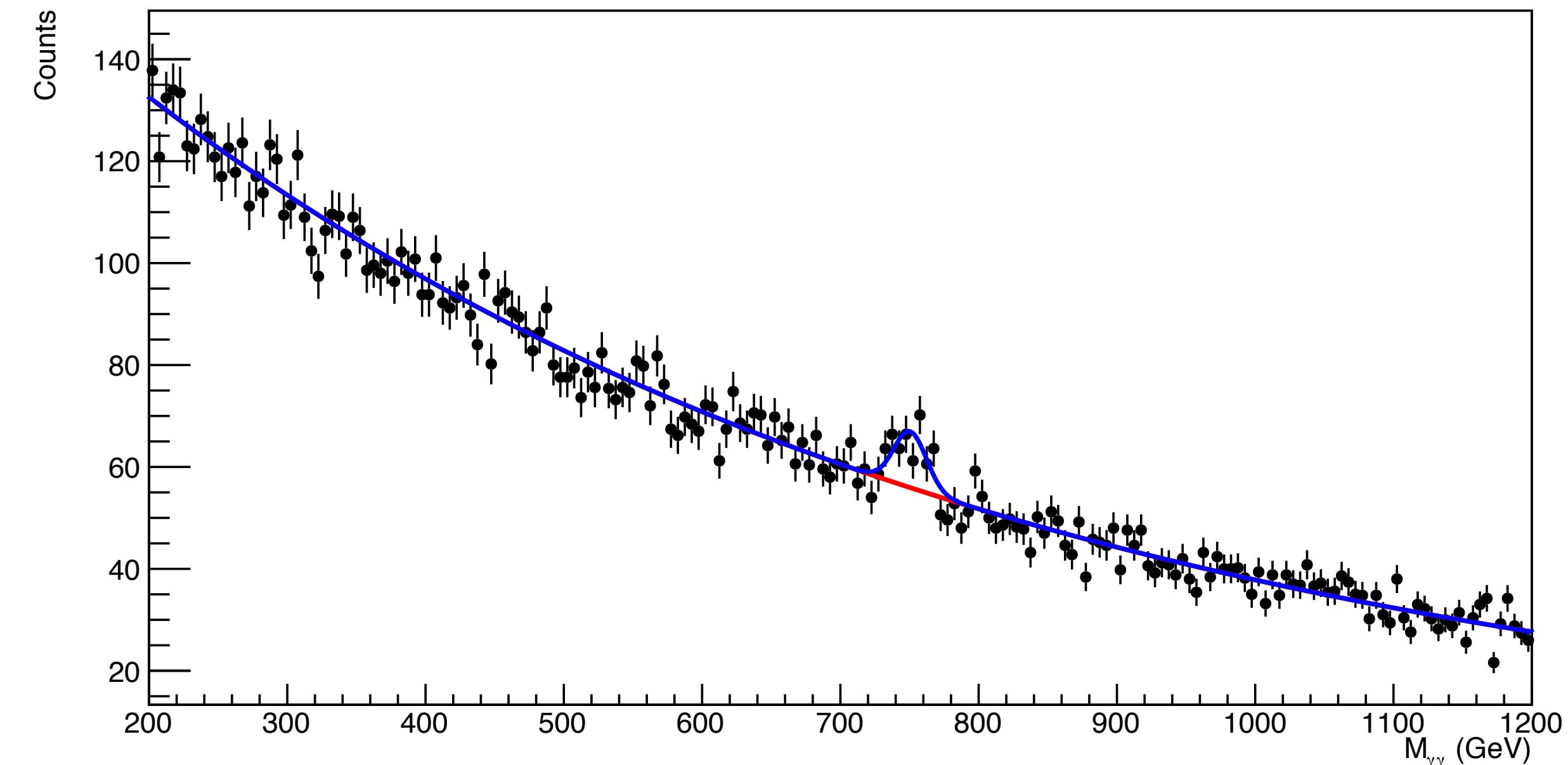
L'ESERCIZIO

Per ciascuno dei due dataset vi viene chiesto di:

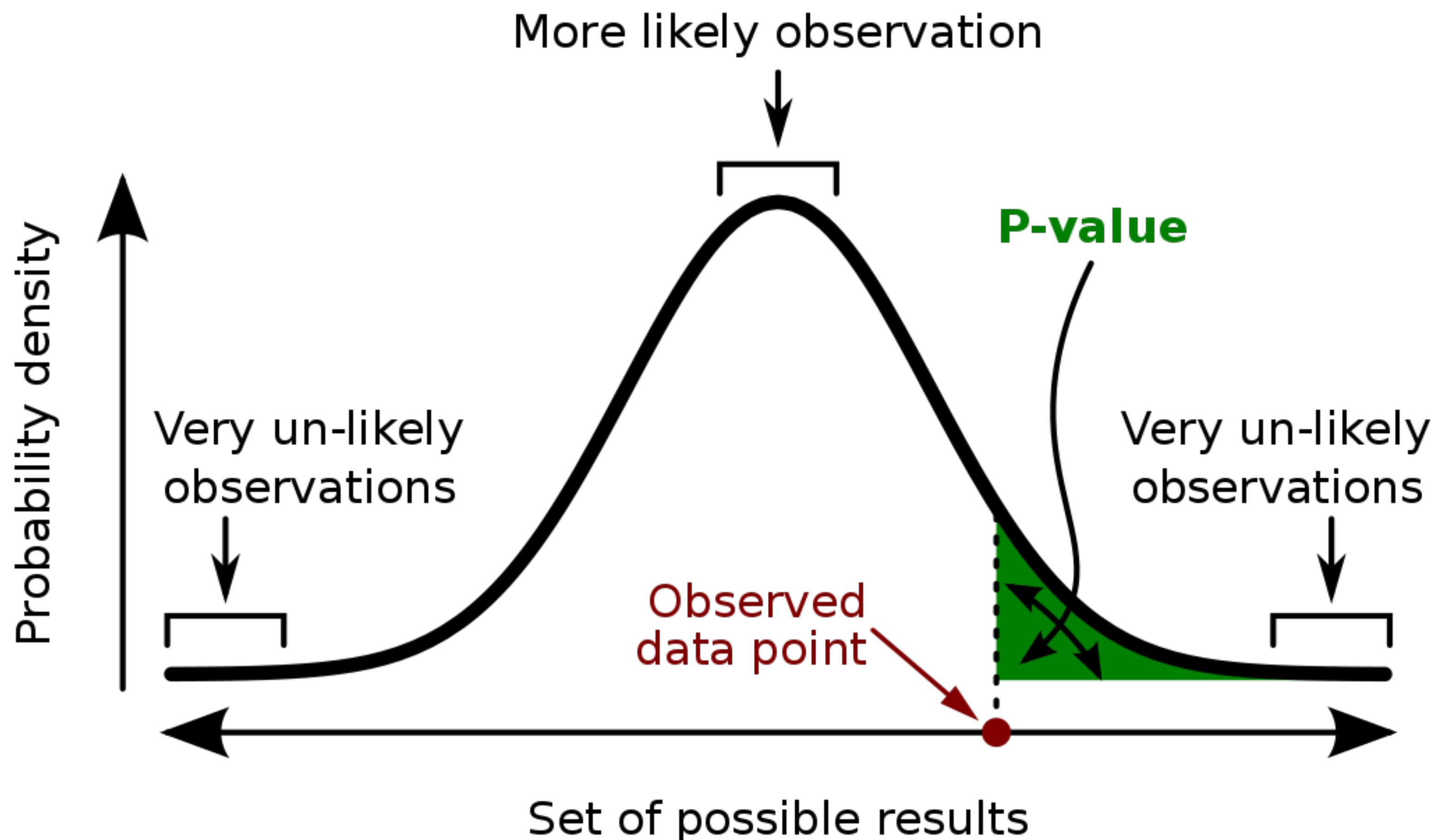
- Stimare la presenza o meno di un segnale alla massa di ~ 750 GeV e riportarne la significanza statistica e p-value.
 - In caso la significanza sia troppo bassa, porre un upper-limit al numero di eventi di segnale.
 - In caso la significanza sia sufficiente, stimare il numero di eventi di segnale e riportarne l'incertezza.

Potete scegliere diversi metodi per calcolare il numero di eventi di segnale:

- Template fit
Dovrete assumere una forma per il background (anche qui, puo' venire dal MC o da una forma funzionale analitica) e un segnale \sim gaussiano con una risoluzione in massa del $\sim 1.8\%$



P-VALUE



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

P-VALUE: TOY EXPERIMENTS

Un modo di calcolare il p-value è quello di generare N pseudo esperimenti nell'ipotesi in cui il segnale è assente (ricordate che potete utilizzare la simulazione MC per determinare le distribuzioni del background).

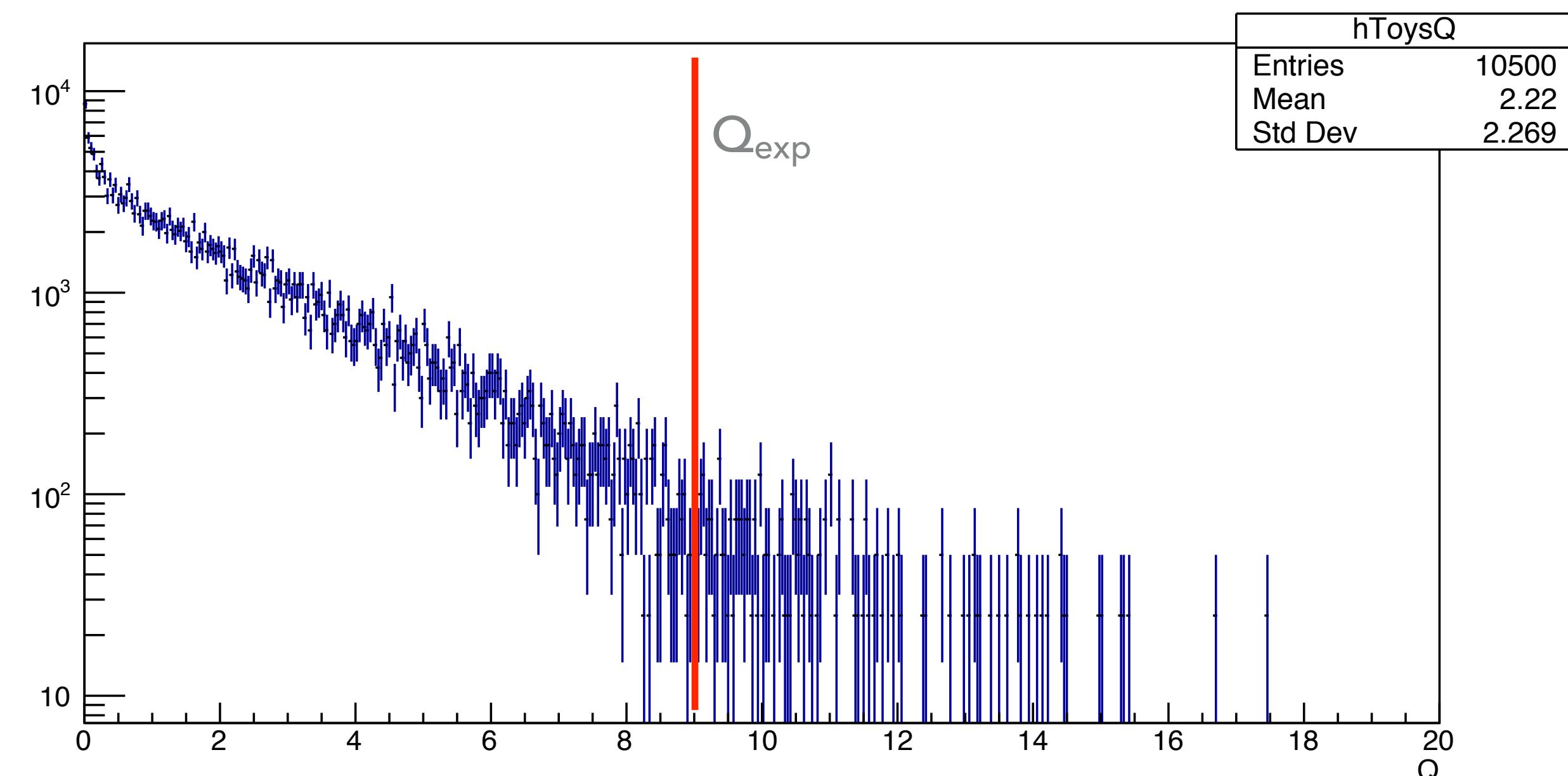
Per ogni pseudo-esperimento generato ripetete la procedura di stima del segnale che avete scelto e calcolate la test-statistics

$$Q = -2 \ln \lambda = -2 \ln \left(\frac{P(\text{obs.} | H_0)}{P(\text{obs.} | H_1)} \right)$$

Riempite un istogramma con tutti gli pseudoesperimenti generati e calcolate quanti di questi presentano un valore di $Q > Q_{\text{exp.}}$.

$$p = \frac{N(Q > Q_{\text{exp}})}{N}$$

Se convertite questo valore invertendo la cumulativa della gaussiana ottenete una significanza.

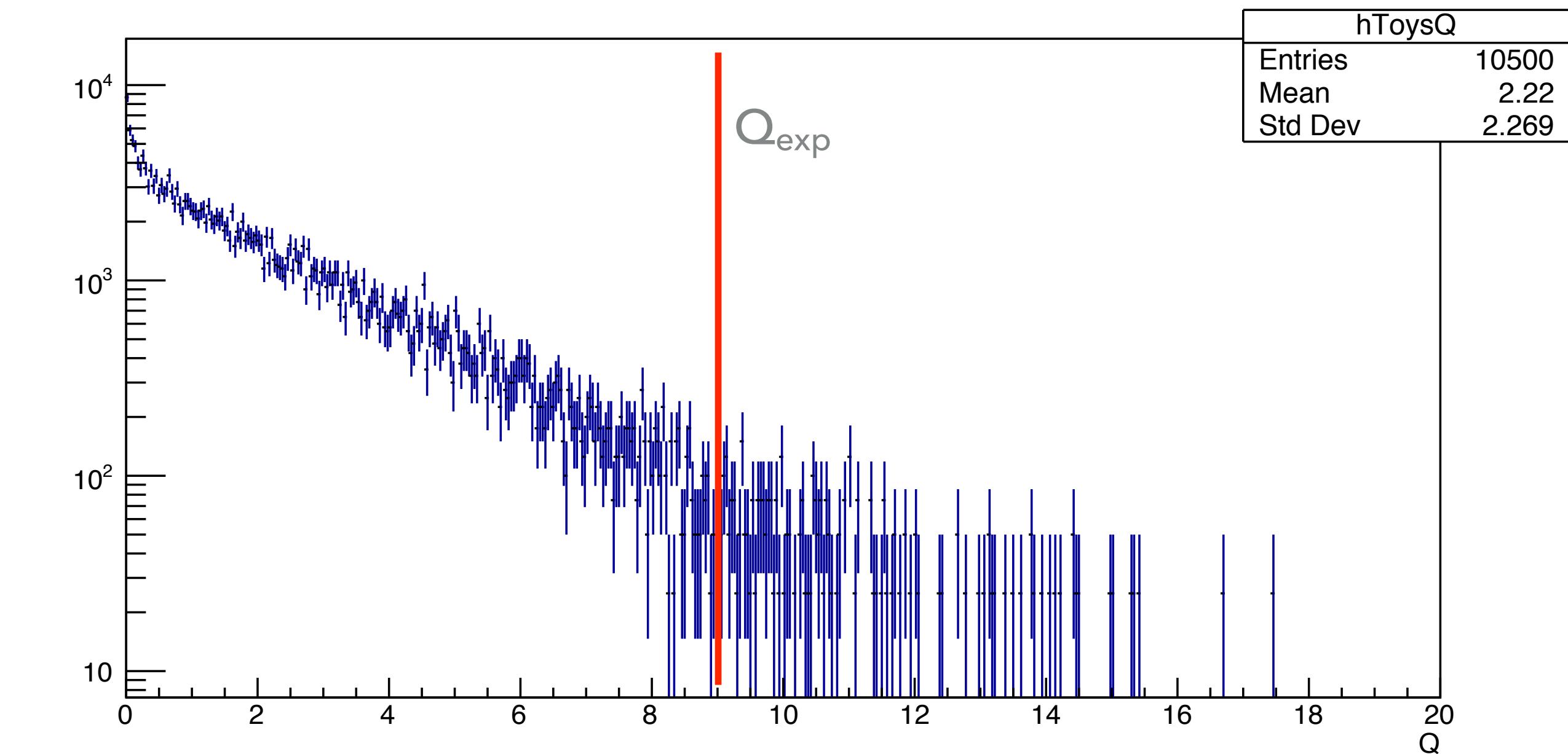
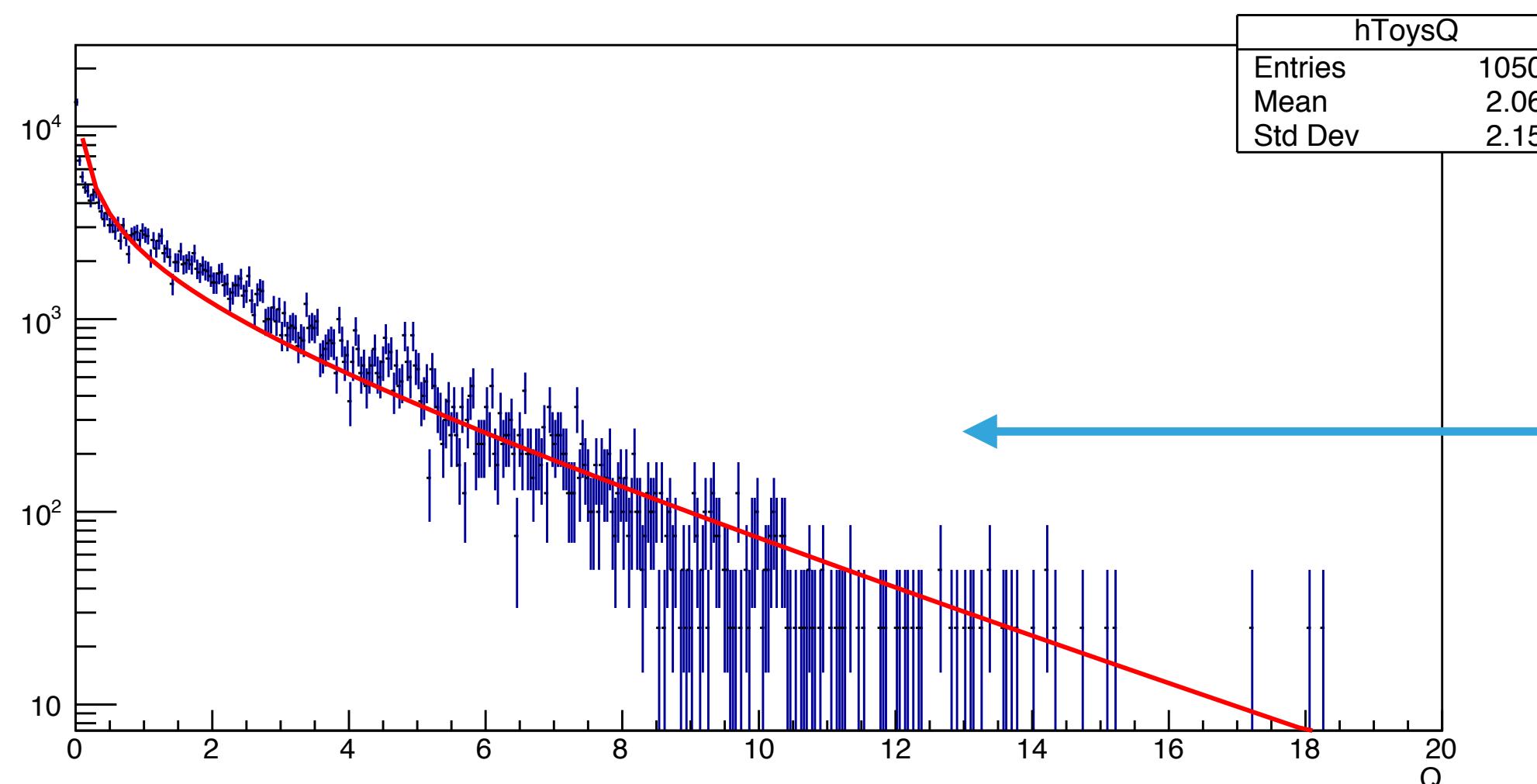


P-VALUE: TOY EXPERIMENTS

Riempite un istogramma con tutti gli pseudoesperimenti generati e calcolate quanti di questi presentano un valore di $Q > Q_{\text{exp}}$.

$$p = \frac{N(Q > Q_{\text{exp}})}{N}$$

Se convertite questo valore invertendo la cumulativa della gaussiana ottenete una significanza.



Attenzione: già per "5 sigma" il p-value e' circa ~10-5, il che richiede di generare un numero molto alto di pseudo-esperimenti. Un'alternativa è utilizzare la forma analitica

$$f(Q) = \frac{1}{\sqrt{2\pi k Q}} e^{-Q/2k}$$

Dove generalmente $k=1$ (ma controllate, puo' dipendere da come stimate il numero di eventi di segnale).

L'ESERCIZIO

Troverete i dataset e una macro di esempio per leggere i dati all'indirizzo <https://github.com/valerioformato/StatFinalExam>

```
#include <iostream>
#include <memory>
#include <random>

#include "TCanvas.h"
#include "TF1.h"
#include "TFile.h"
#include "TH1D.h"
#include "TRandom3.h"
#include "TTree.h"

void ReadData() {

    // 4-vectors are stored with the convention
    // p = (px, py, pz, E)
    float photon1[4], photon2[4];

    auto inFile = TFile::Open("data/data_high.root");
    auto tree = (TTree *)inFile->Get("Data");

    tree->SetBranchAddress("photon1", &photon1);
    tree->SetBranchAddress("photon2", &photon2);

    for (Long64_t iEv = 0; iEv < tree->GetEntries(); iEv++) {
        tree->GetEntry(iEv);

        // do stuff here...
    }
}
```

```
root -l data/data_low.root
root [0]
Attaching file data/data_low.root as _file0...
(TFile *) 0x7f8f9c114460
root [1] .ls
TFile** data/data_low.root
TFile* data/data_low.root
    KEY: TTree Data;1 Data
root [2] Data->Print()
*****
*Tree :Data : Data
*Entries : 9953 : Total = 320756 bytes File Size = 298269 *
*: : Tree compression factor = 1.07 *
*****
*Br 0 :photon1 : ph1[4]/F
*Entries : 9953 : Total Size= 160211 bytes File Size = 119298 *
*Baskets : 4 : Basket Size= 32000 bytes Compression= 1.07 *
*.....
*Br 1 :photon2 : ph2[4]/F
*Entries : 9953 : Total Size= 160211 bytes File Size = 119321 *
*Baskets : 4 : Basket Size= 32000 bytes Compression= 1.07 *
*.....
root [4] Data->Show(0)
=====> EVENT:0
ph1      = 67.1726,
          318.537, -146.11, 354.014
ph2      = 33.4587,
          -51.6474, -360.718, 360.849
root [5] Data->Show(1)
=====> EVENT:1
ph1      = 12.6363,
          -23.0406, -94.6367, 98.6812
ph2      = 213.468,
          156.496, -399.253, 475.59
```