

# Lecture 11

Higgs Boson Discovery  
Towards Deep Learning

# Hypothesis Test

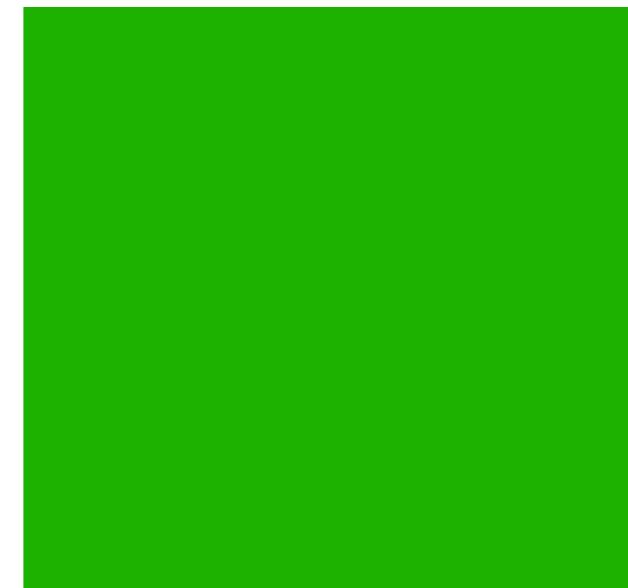
- Null Hypothesis :  
•  $p_a(x) = p_b(x)$   

Prior	Posterior
$p_a(x)$	$p_b(x)$
- Alternative Hypothesis  
•  $p_a(x) \neq p_b(x)$
- Answers these questions:
  - **Is my data consistent?**
  - **Can I detect an effect?**
  - **Did I just make a discovery?**

# T-test



Sample A



Sample B



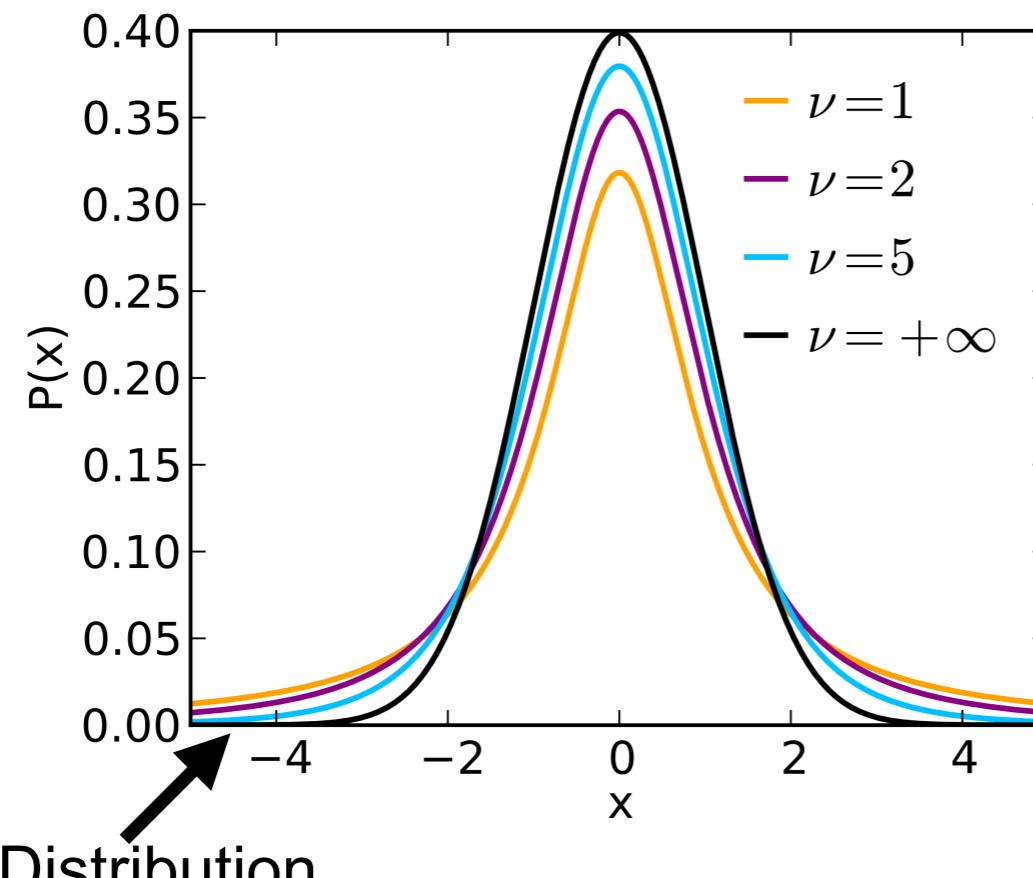
Sample C

Are Sample means A, B, and C consistent with each other?

# Frequentist :t-test

- Often called the “students” t-test (<https://www.youtube.com/watch?v=U9Wr7VEPGXA>)
- Invented by brewmaster to test flavor variations in Guinness

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad \nu = N - 1$$

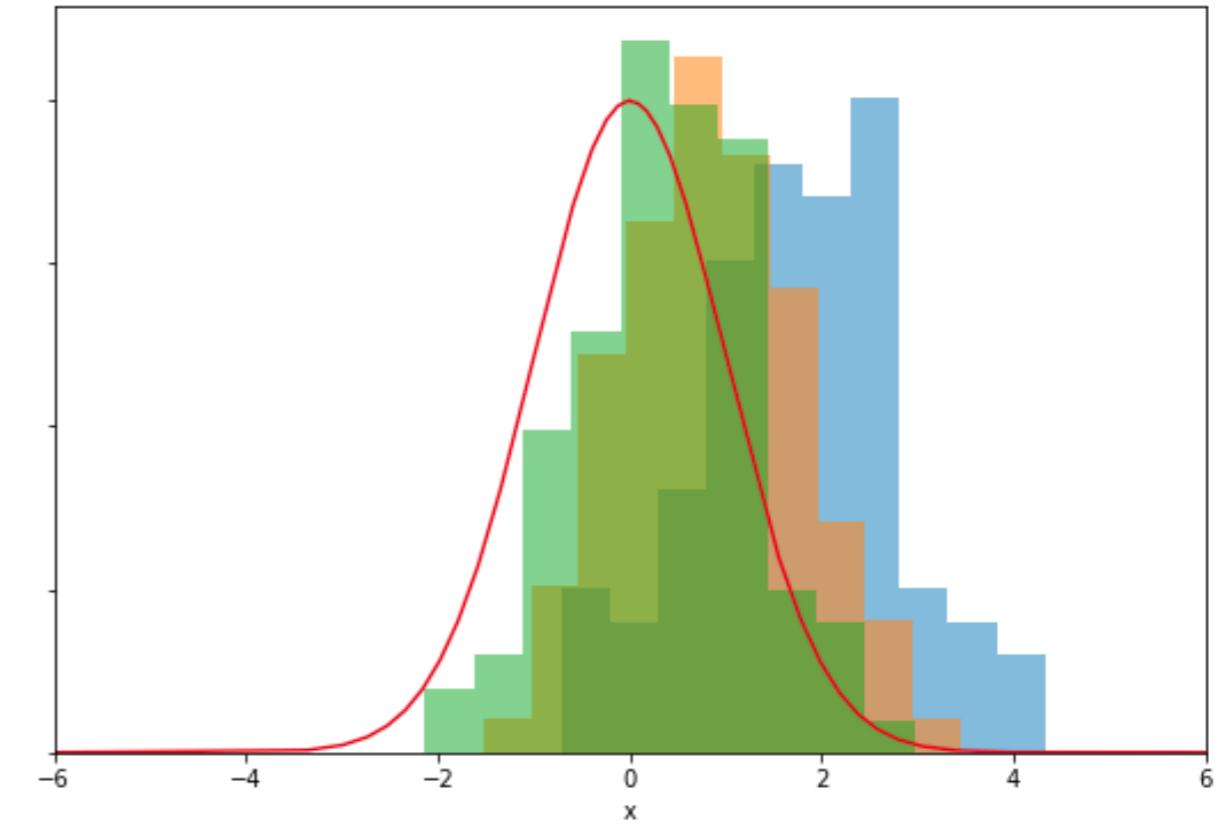
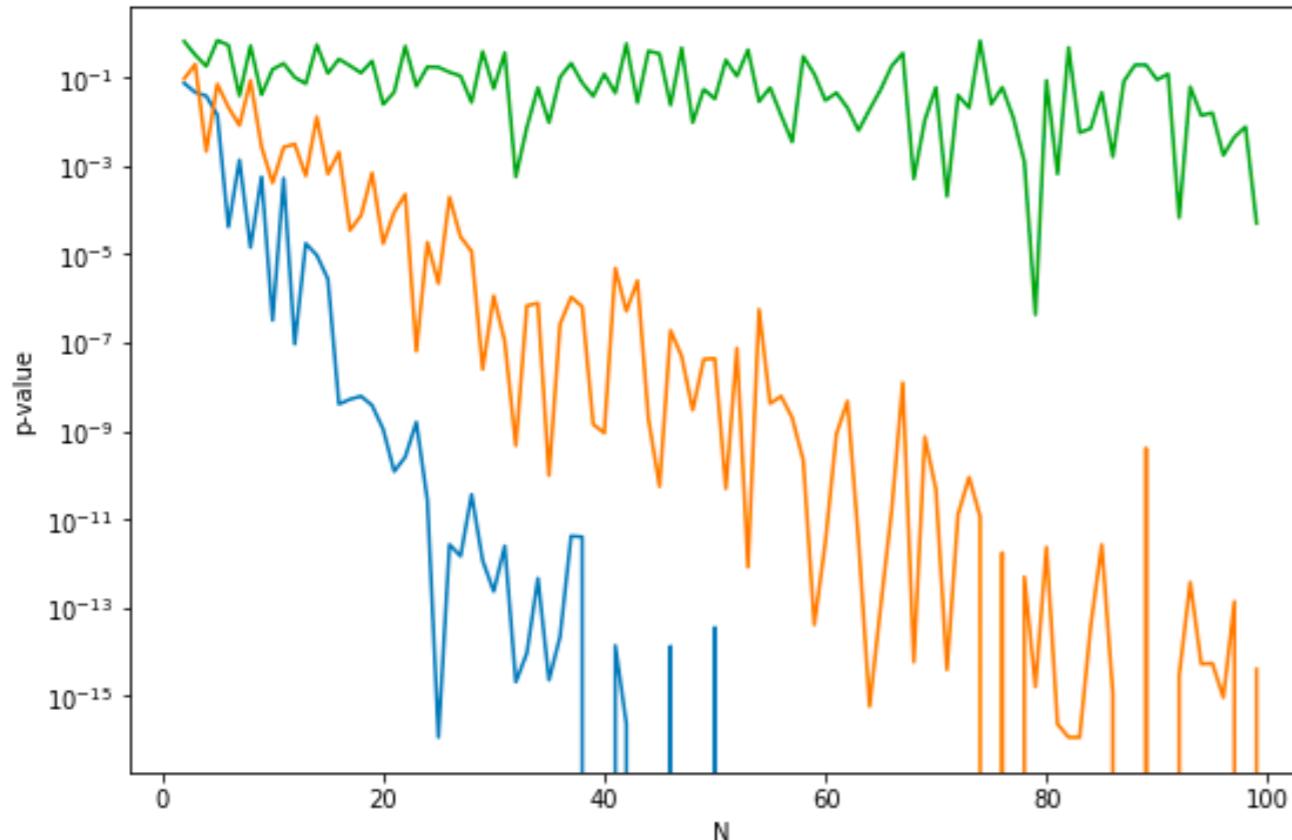


# Frequentist: t-test

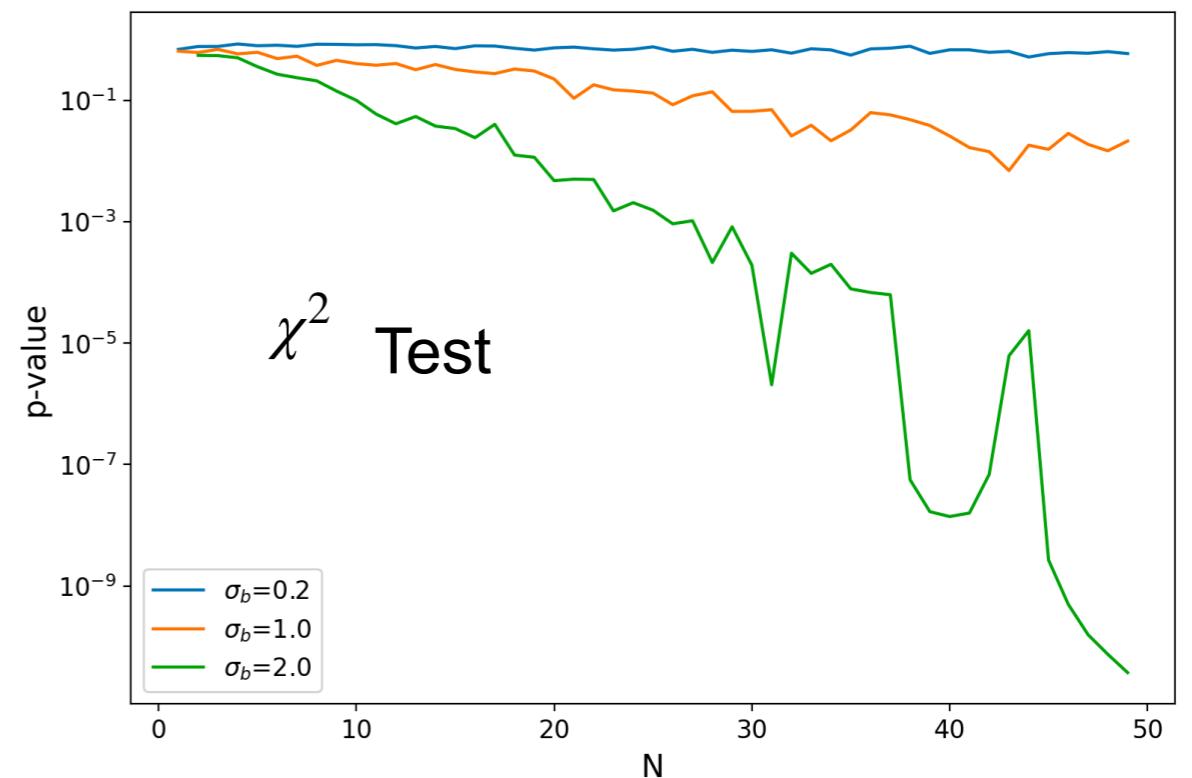
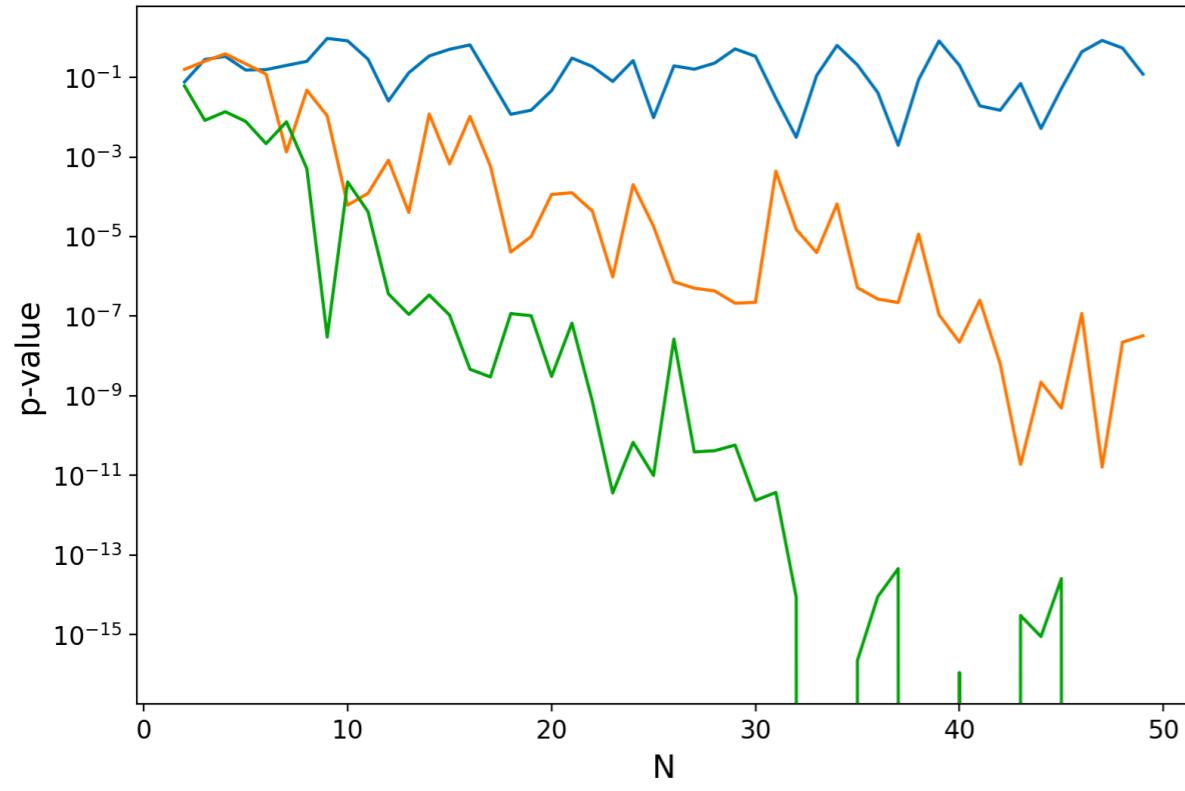
- Perform a t-test:
  - If samples  $X_1, X_2, \dots, X_N$  from the same underlying distribution
  - Underlying distribution is Gaussian with mean  $\mu$  and variance  $\sigma^2$
  - If the above is true then we have a measured mean+variance:
    - given by  $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$  and  $S^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$
    - And the distribution  $\frac{\bar{X} - \mu}{S/\sqrt{N}}$  is given by a t-distribution

# t-test Consistency

- Variation as a number of events



# t-test Consistency<sup>7</sup>



# KS-test

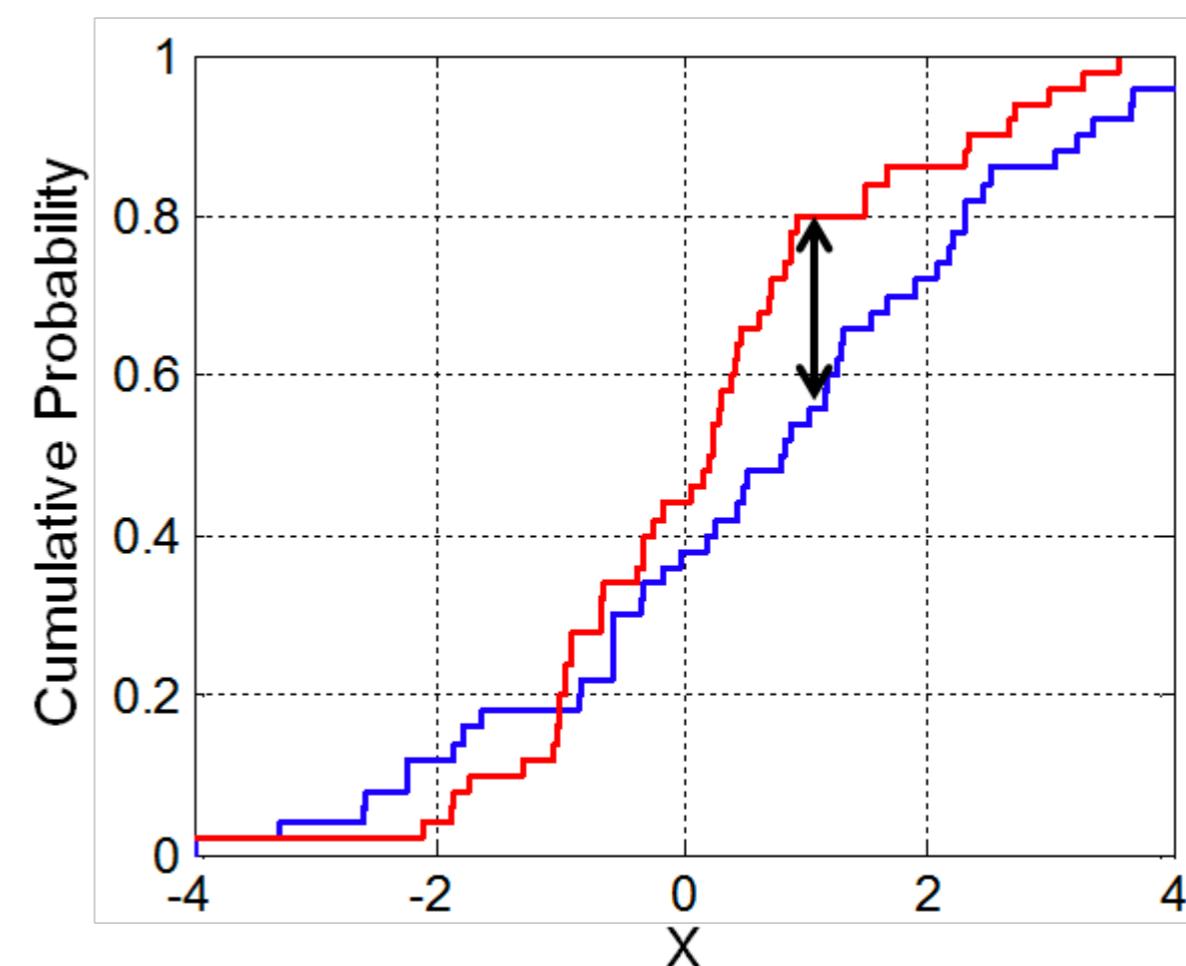
- Komolgorov-Smirnov Test
  - Estimate the maximum difference between observed and predicted cumulative distribution functions and compare with expectations.
  - An unbinned way to check histograms are equal

$$\hat{F}_n(t) = \frac{\text{number of elements in the sample } \leq t}{n} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq t},$$

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|,$$

↑  
*x*

Supremum Function (ie max difference)



# Likelihood Ratio Test

- Perhaps the most important test is the likelihood ratio test
- This is what we will focus for the rest of these slides
  - Likelihood ratio is the standard in hypothesis testing
  - This approach is very similar to a chi<sup>2</sup> test

$\mathcal{L}(\theta_i) \forall \theta_i \in \Theta_0$  where  $\Theta_0$  refers to null – hypothesis

$\mathcal{L}(\theta_i) \forall \theta_i \in \Theta$  where  $\Theta$  refers to alternative – hypothesis

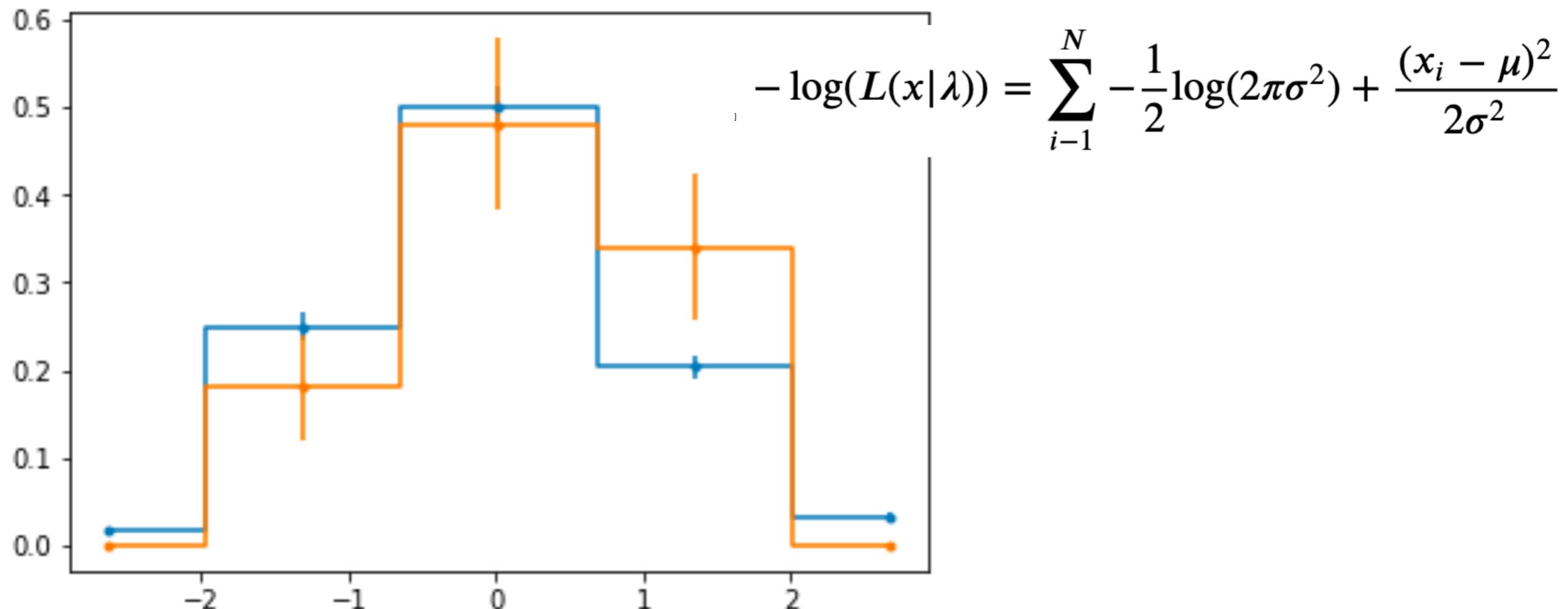
$\sup_{\theta \in \Theta_0} \mathcal{L}(\theta)$  is the maximum likelihood over all  $\theta$

$$\lambda_{LR} = -2 \ln \left[ \frac{\sup_{\theta \in \Theta_0} \mathcal{L}(\theta)}{\sup_{\theta \in \Theta} \mathcal{L}(\theta)} \right]$$

Neyman-Pearson Lemma states the likelihood ratio converges to the optimal test statistic

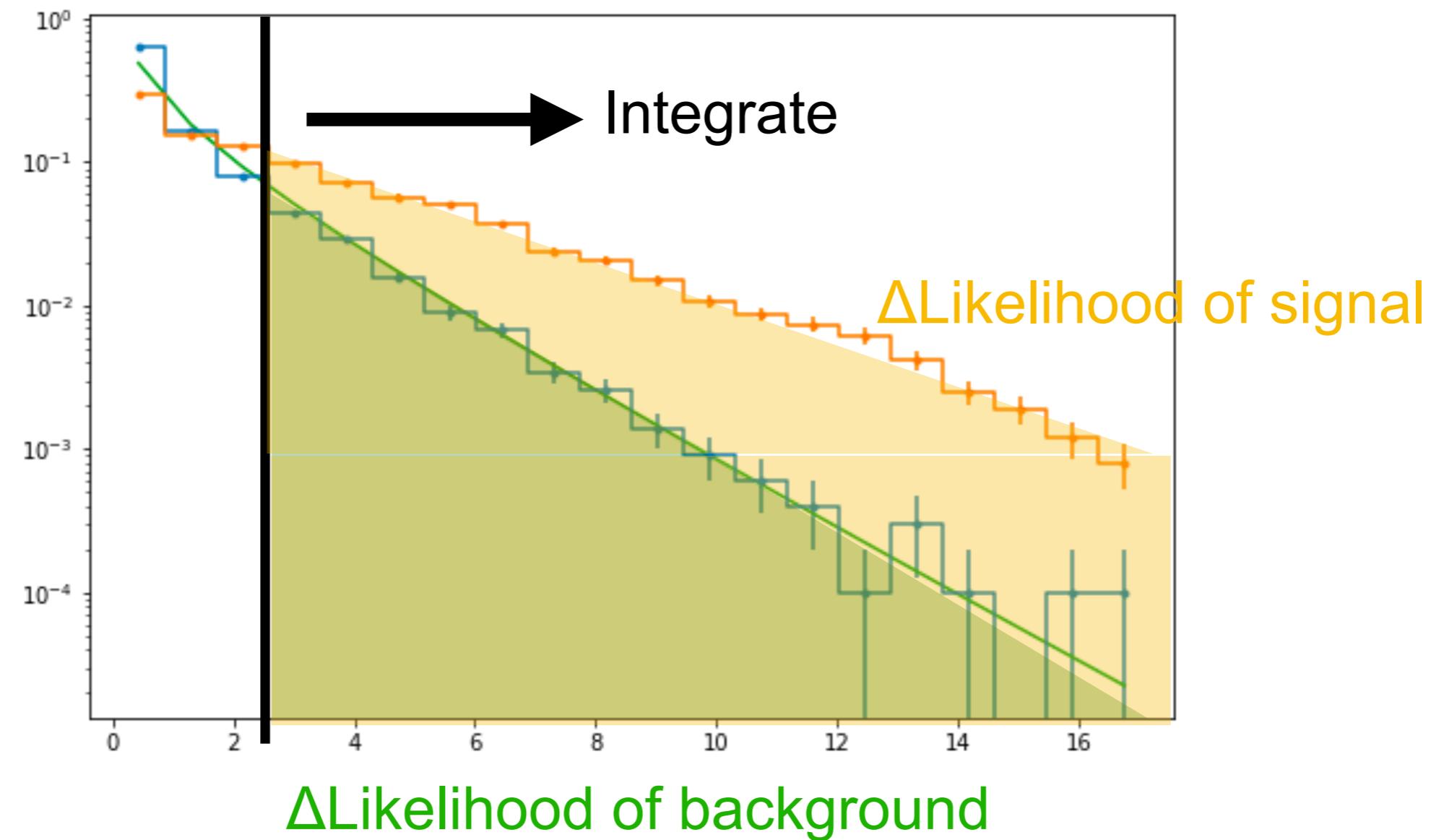
# Likelihood Test: Concept

- We would like to see if these two histograms are consistent
  - Hypothesis: Orange is a gaussian with mean 0 (blue)
  - True: Orange is sampled from a gaussian with mean 0.2



# Likelihood Test: Gaussian

- We can see how these on average a deviation



# Blind Analysis



By Judging unbiased  
the analysis  
Without knowledge of  
the result  
you remain unbaised

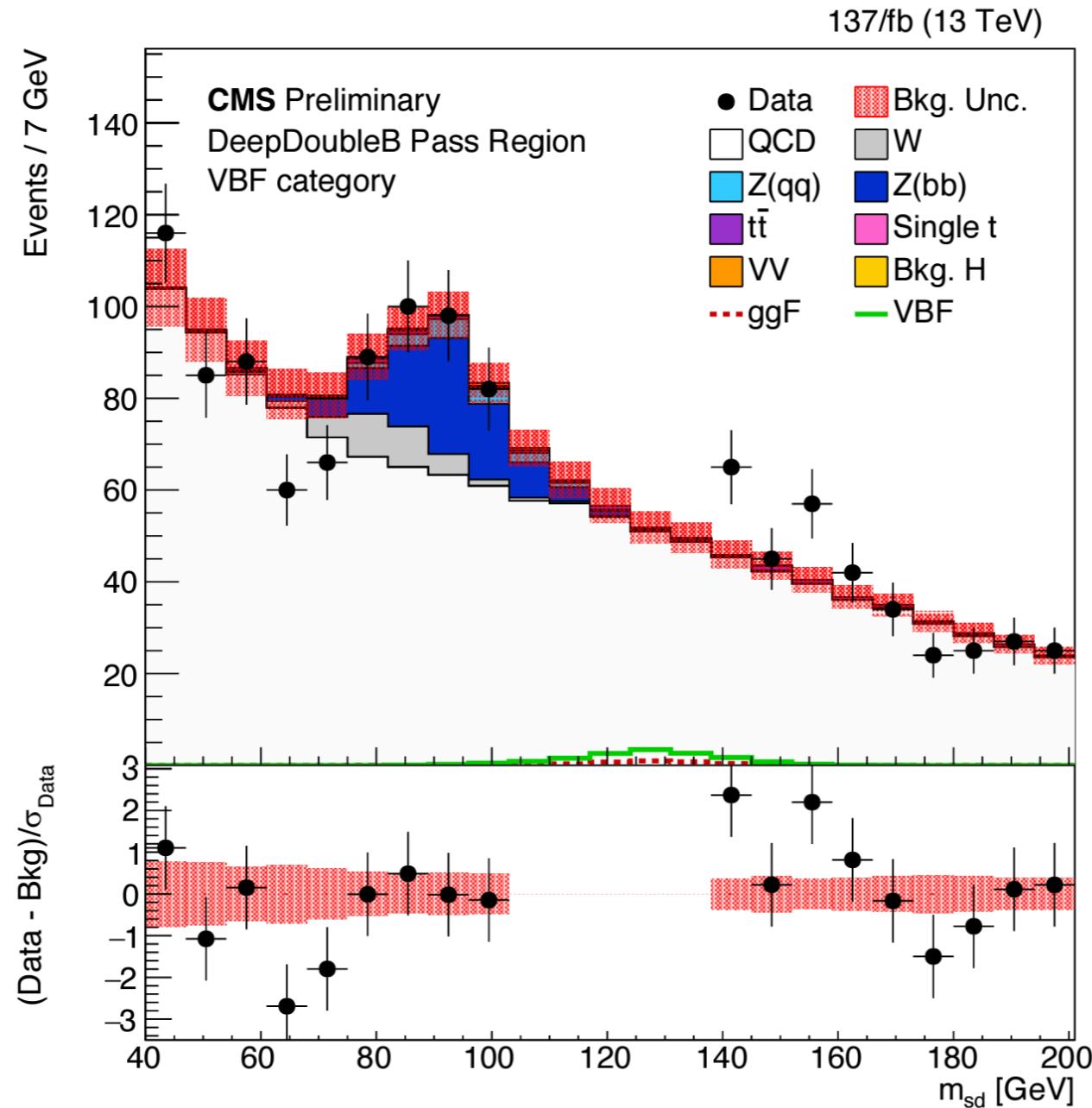
- Strategy :
  - Don't look at the signal region with the full selection
    - ▶ In this case don't make the Mass Plot
    - ▶ You can make plots of other non-sensitive variables
  - Make a simulated sample to tune everything

# Unblinding

- Once you are ready to search **you look in the box (unblind)**
  - ▶ Done in multiple steps usually
    - Compute the chi2/Goodness of fit of the sample
    - Look at best fit parameters on the fit (excluding signal)
    - Look at the actual Mass plot
  - Once you have finally looked at everything you get significance
- <https://www.youtube.com/watch?v=KezwARhBlc>

# Unblinding

- Once you
  - Done
  - Cc
  - L<sub>0</sub>
  - L<sub>0</sub>
  - Once you have finally looked at everything you get significance
- <https://www.youtube.com/watch?v=KezwARhBlc>

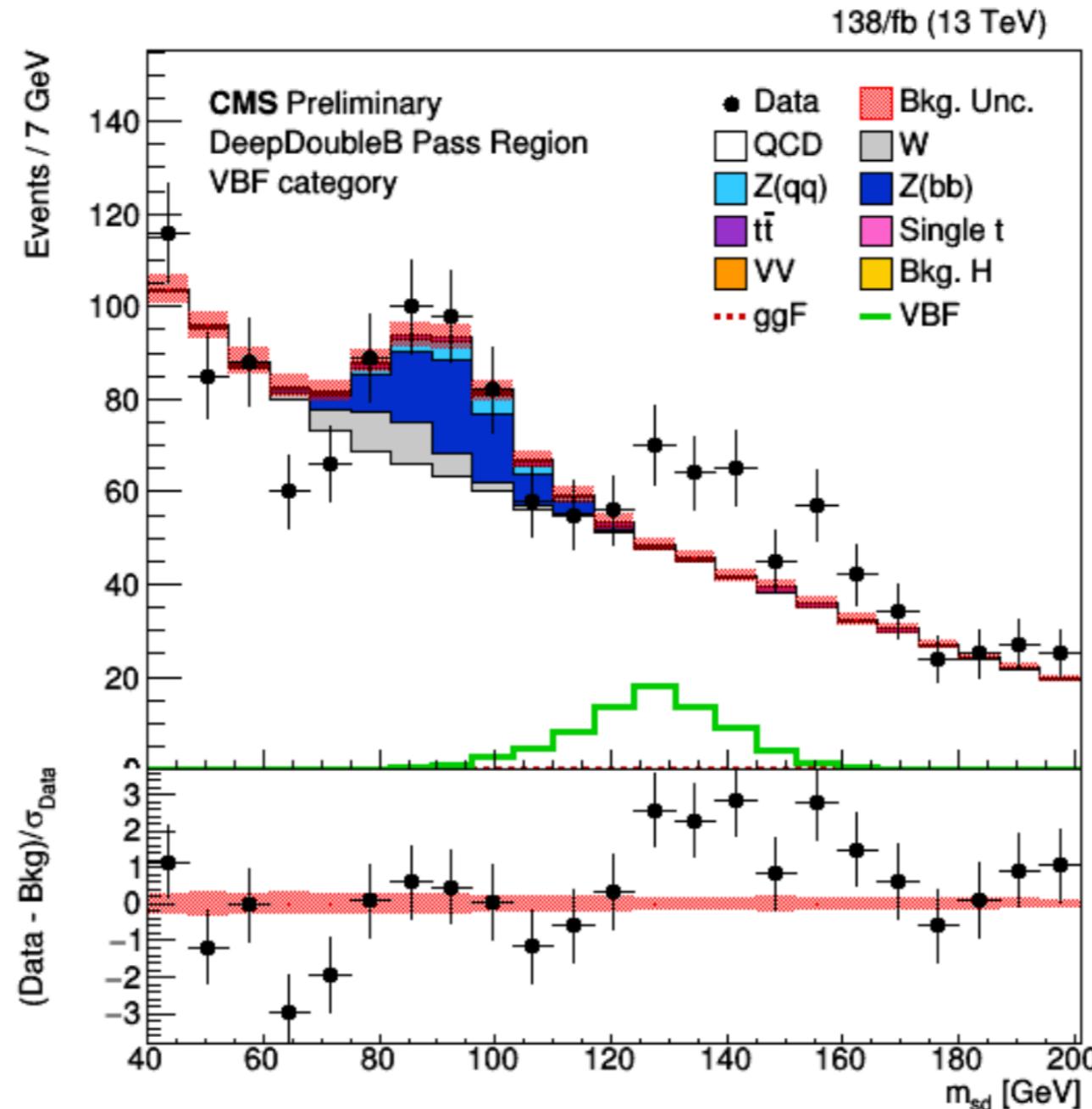


box (unblind)

the sample  
excluding signal)

# Unblinding

- Once you
  - Don't open the box (unblind) the sample (excluding signal)
  - C
  - L
  - L
- Once you have finally looked at everything you get significance
- <https://www.youtube.com/watch?v=KezwARhBlc>



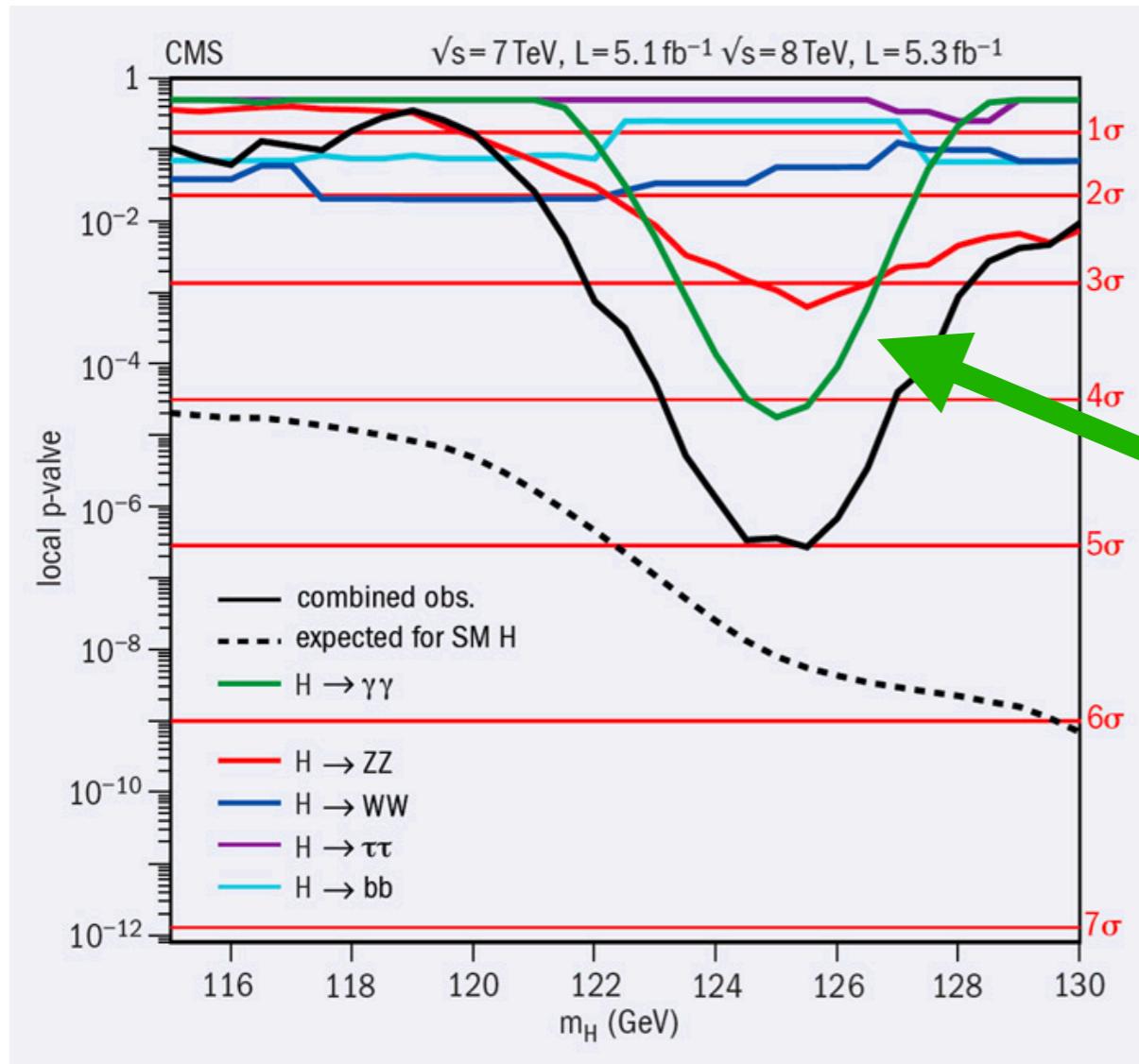
# Higgs Boson

- We discovered the Higgs boson not so long ago



# In this lecture we are going to make the discovery plot

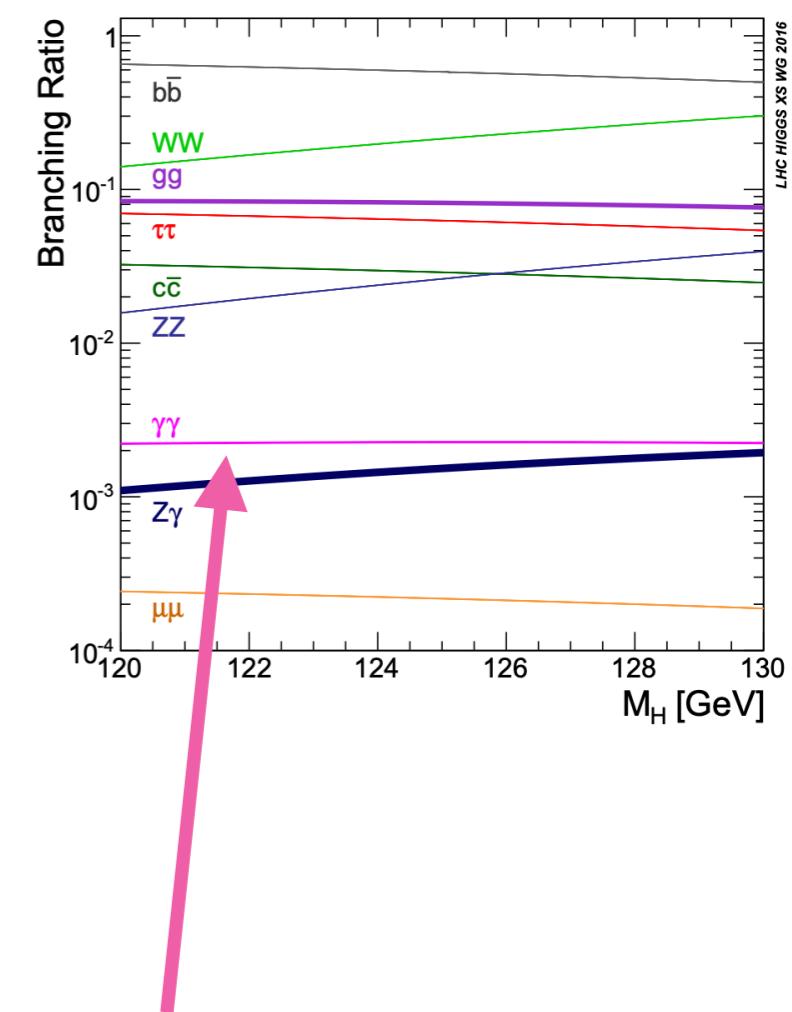
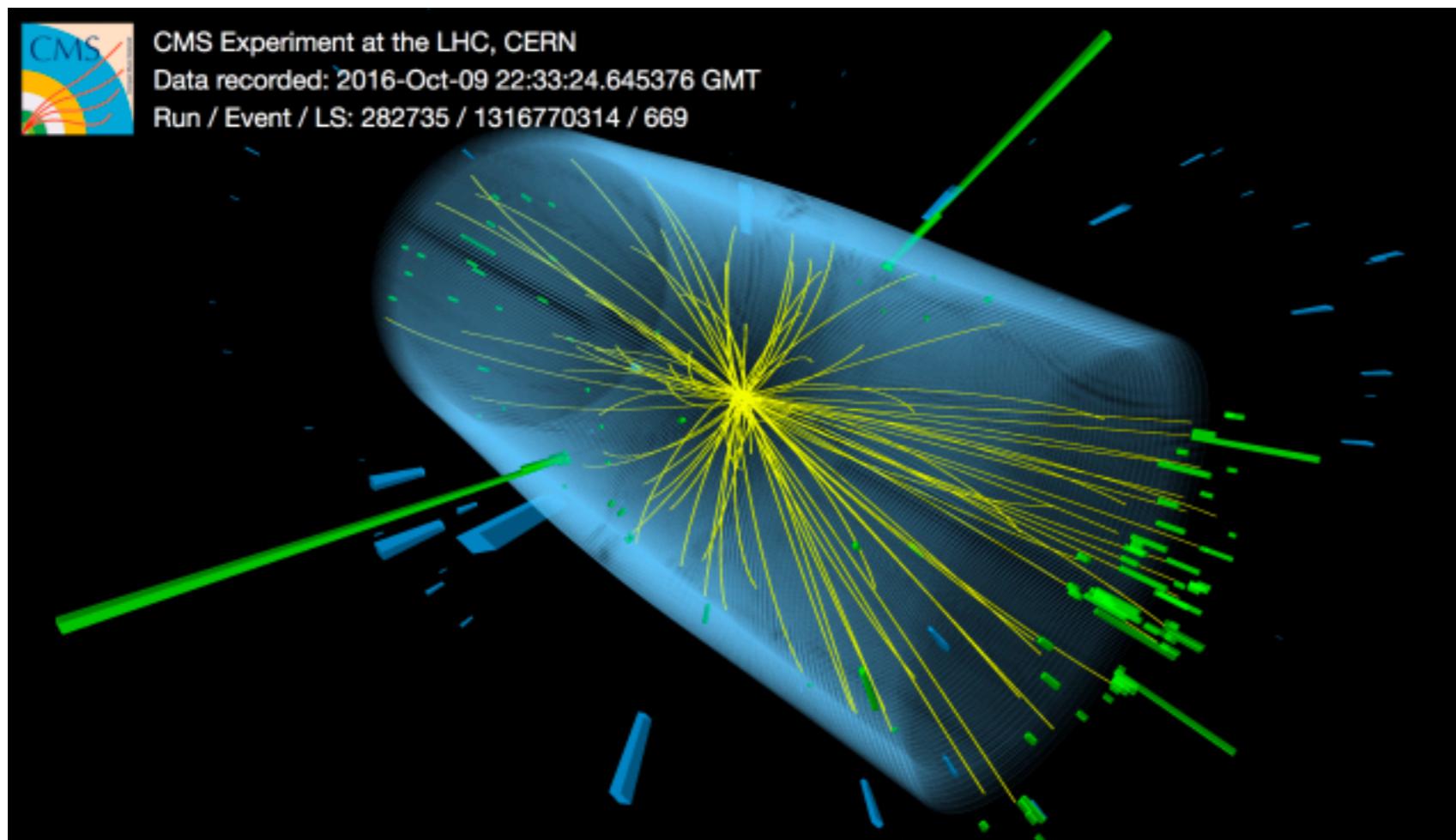
- To discovery it we made a plot like this:



We are going to focus on his object

# Higgs to two photons

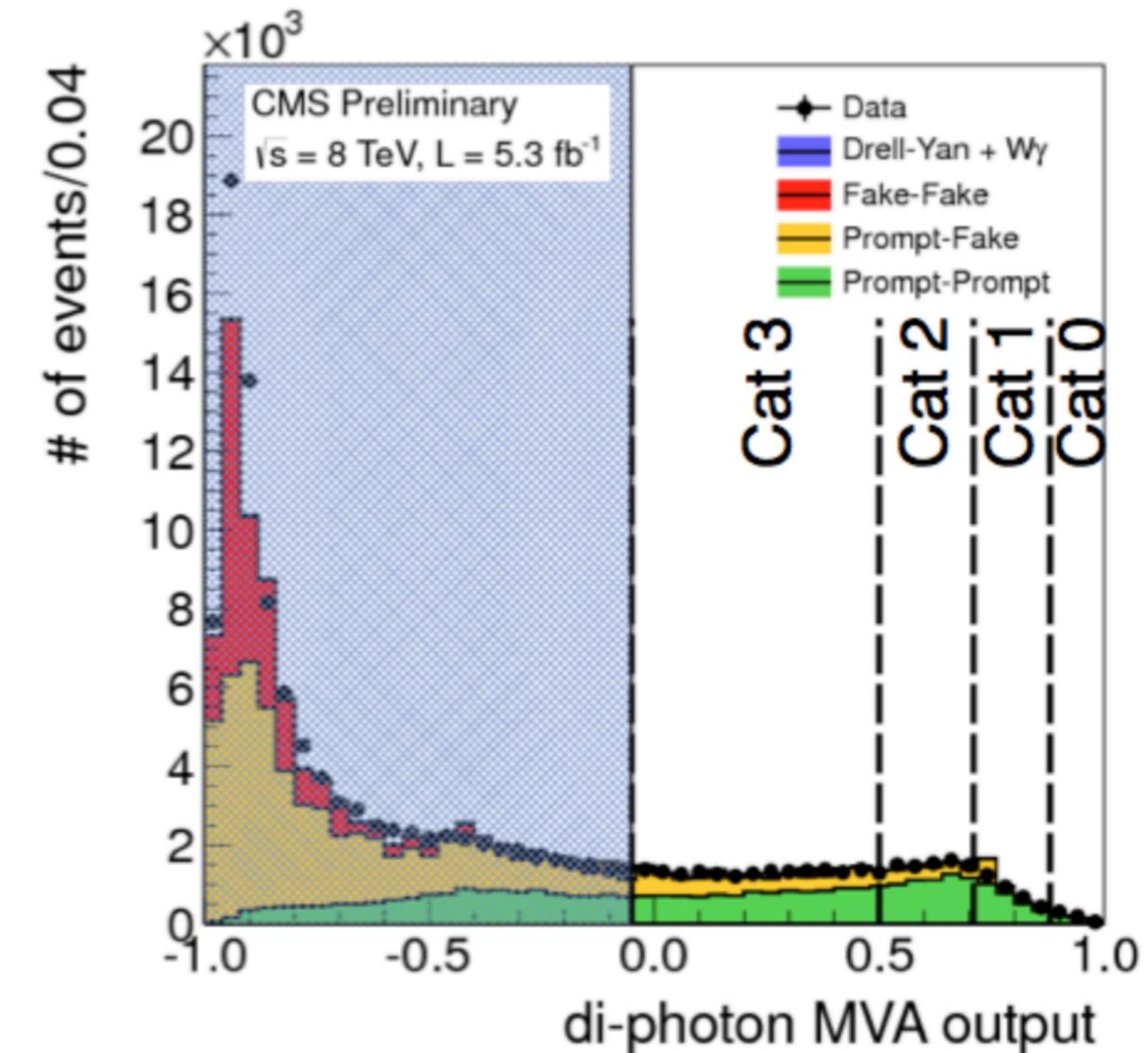
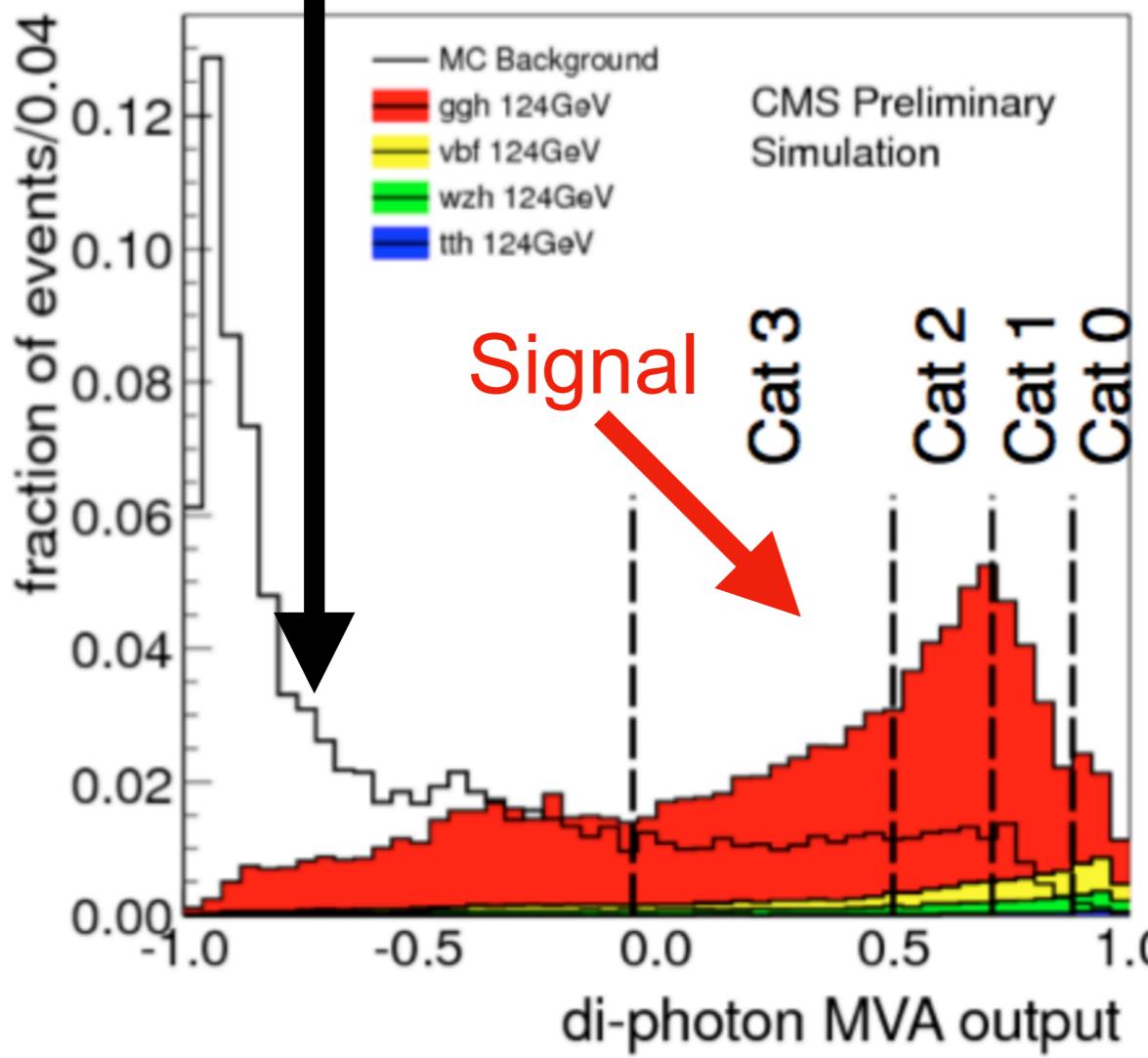
- Higgs to boson decays to nearly every other particle



Higgs bosons to two photons occur about 0.3% of the time

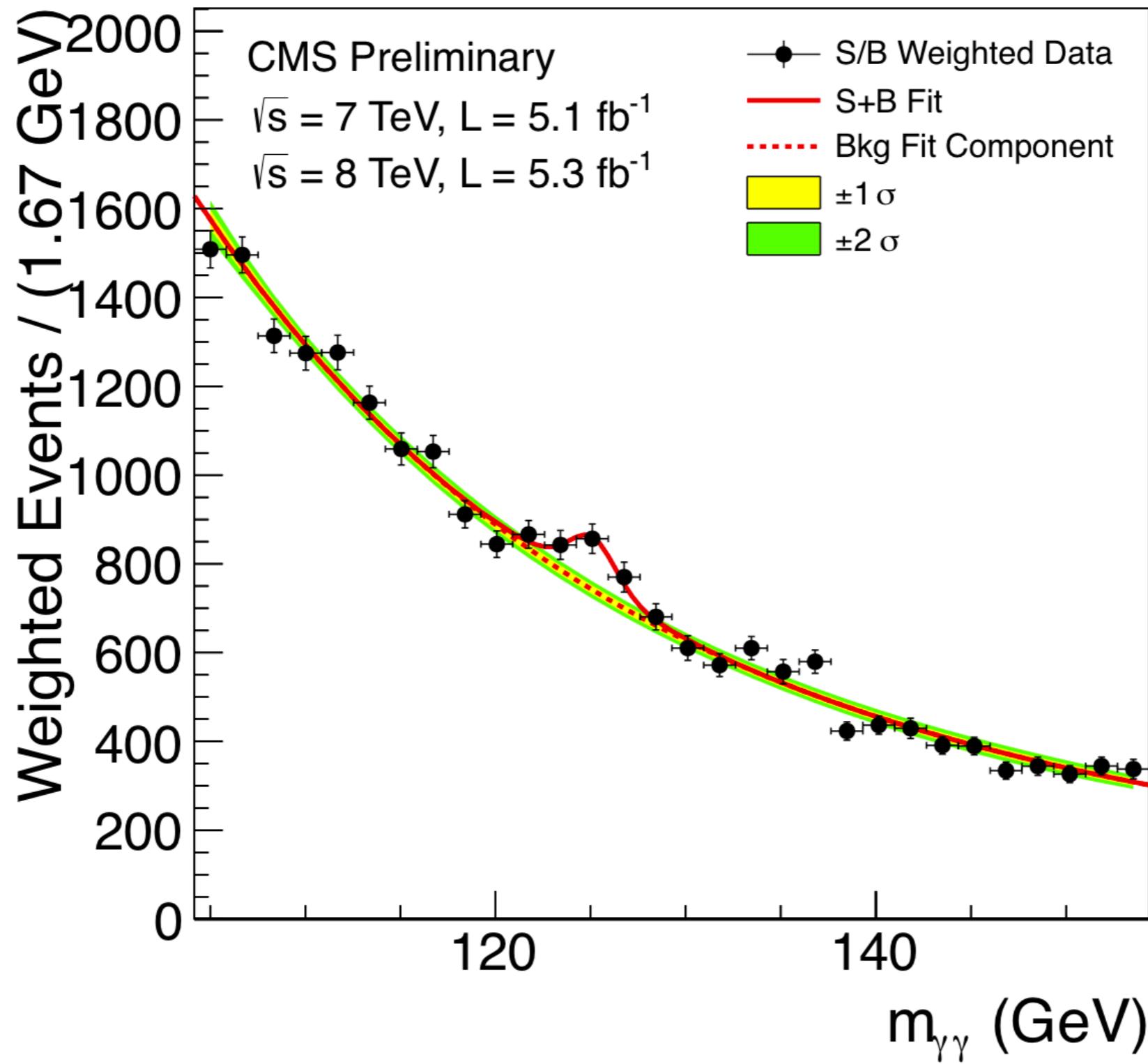
Background

# Di Photon selection



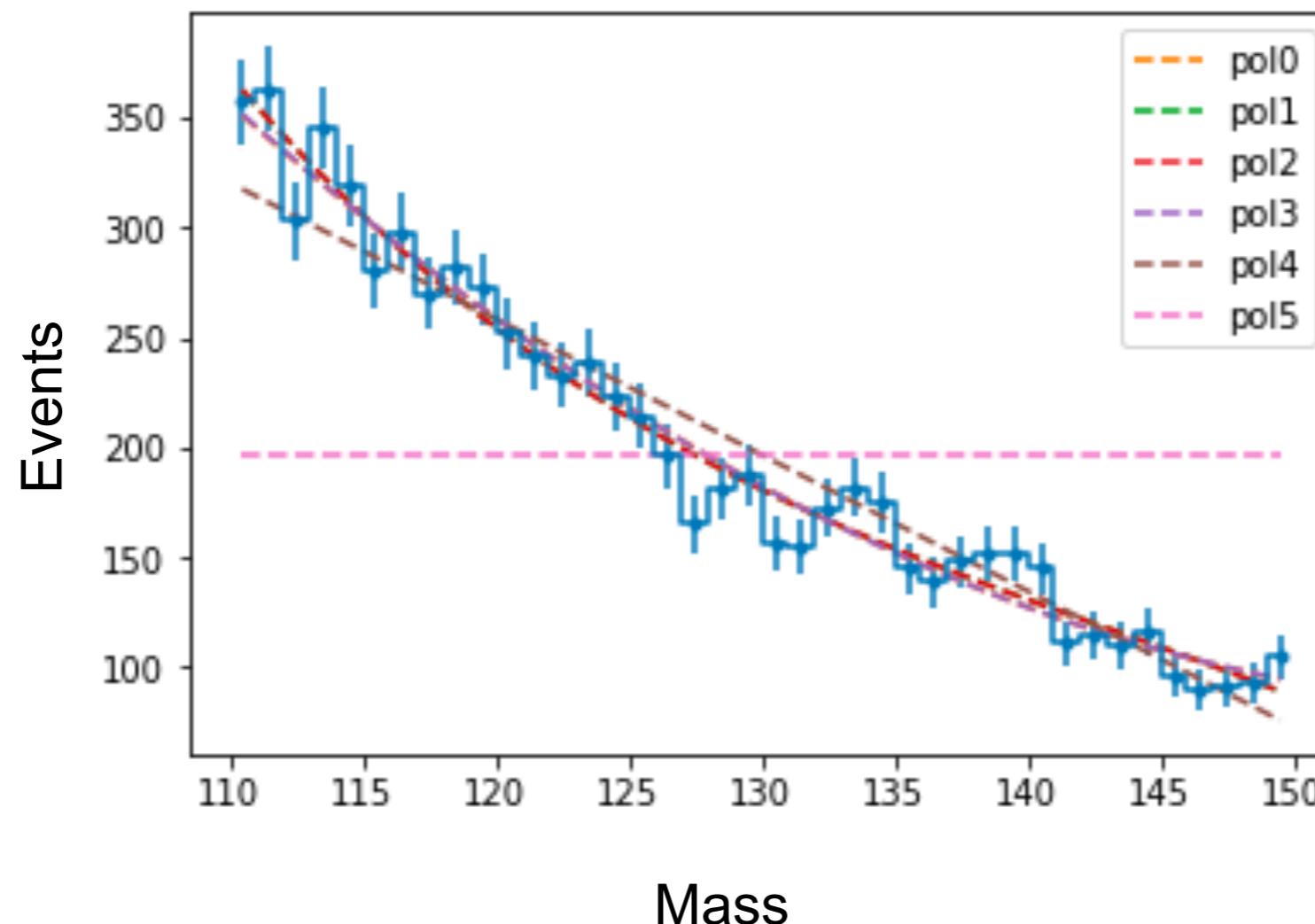
- To select di-photons we build a neural network
  - This takes into account many different Higgs properties

# Higgs Discovery Plot



# Building a Model

- The frequentist will look at this data and guess a model

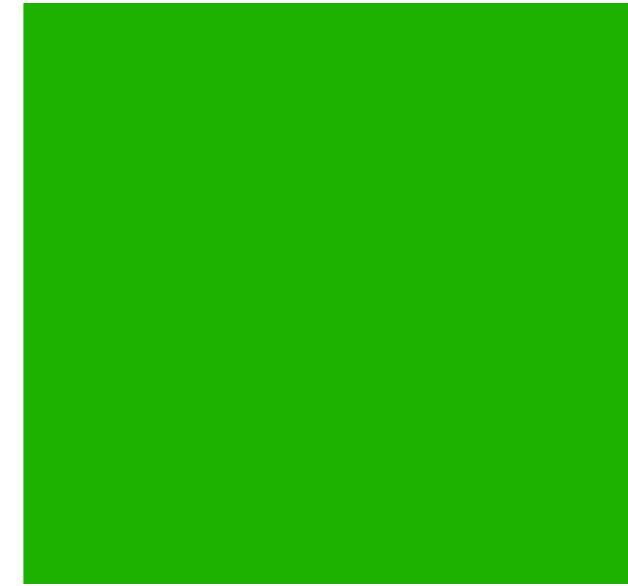


Often when we fit, we throw a barrage of functions at it  
As a rule of thumb we do the “chi-by-eye” (If chi<sup>2</sup> is good, we are ok)  
A more robust is an f-test (see notes)

# T-test



Sample A



Sample B



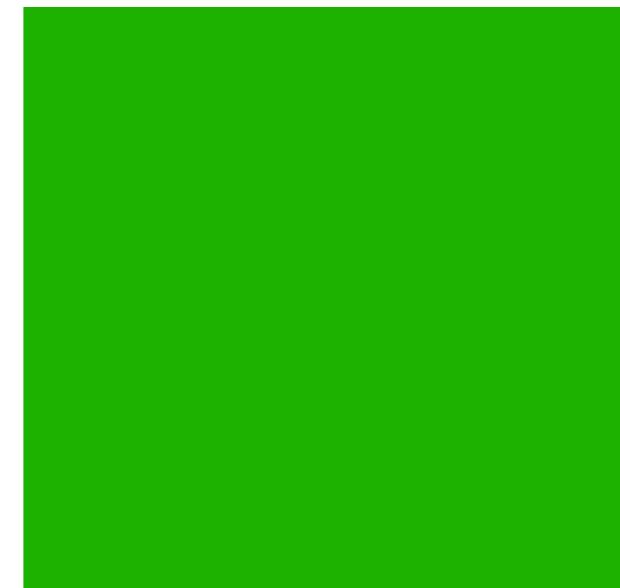
Sample C

Are Sample means A, B, and C consistent with each other?

# F-test



Sample A



Sample B



Sample C

Are the Variances of A, B, and C consistent with each other?

# F-Statistic

- F distribution leads to a hypothesis about groupings

$$\bullet \quad F = \frac{\text{explained Variance}}{\text{Unexplained Variance}} = \frac{\text{Between Group Variability}}{\text{Single Group Variability}}$$

- For fitting functions as a polynomial we can write this as

$$\bullet \quad F = \frac{\frac{1}{\Delta_{DOF}(f, g)} \left( \sum_i (y_i - f(x_i))^2 - \sum_i (y_i - g(x_i))^2 \right)}{\left( \frac{1}{n - p_f} \sum_i (y_i - f(x_i))^2 \right)}$$

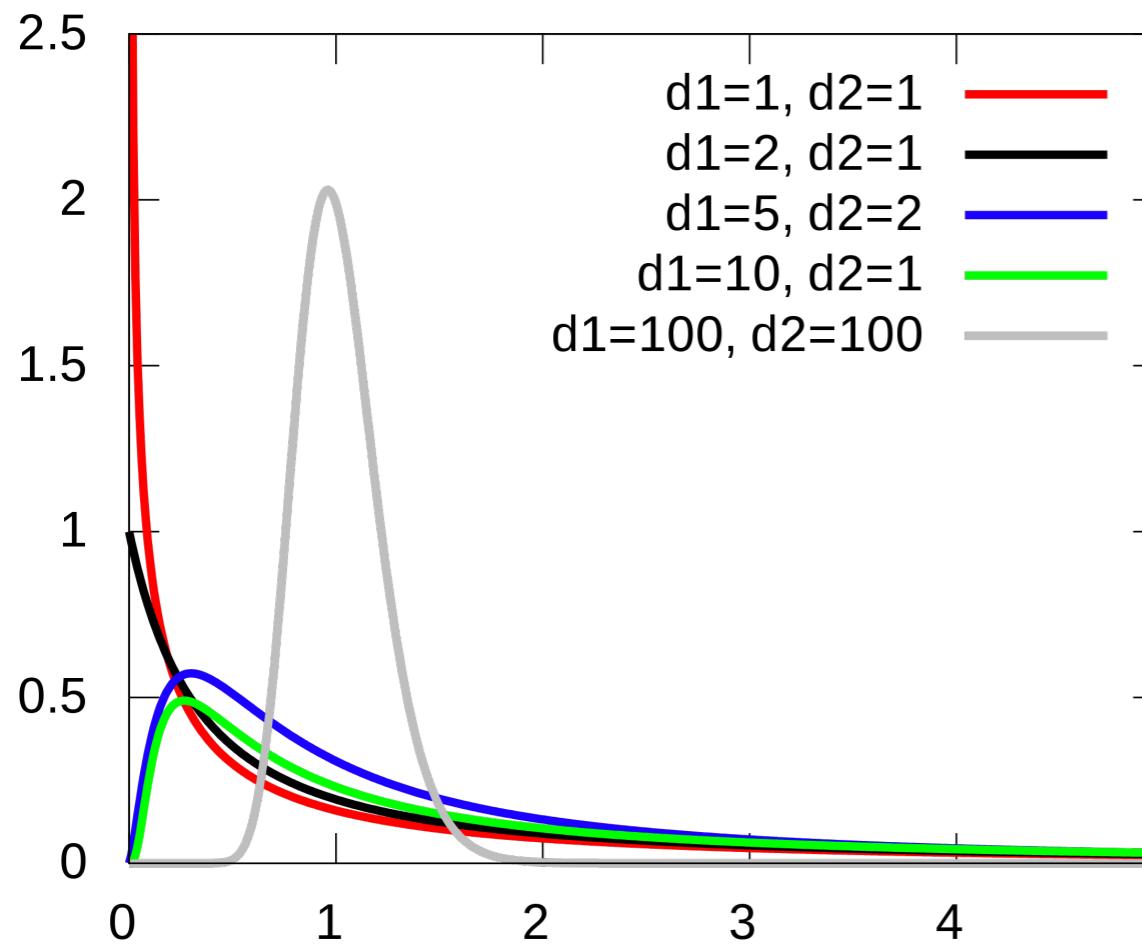
# F-Statistic

- F distribution leads to a hypothesis about groupings
- $$F = \frac{\text{explained Variance}}{\text{Unexplained Variance}} = \frac{\text{Between Group Variability}}{\text{Single Group Variability}}$$
- For fitting functions as a polynomial we can write this as
 
$$F = \frac{\frac{1}{\Delta_{DOF}(f, g)} \left( \sum_i (y_i - g(x_i))^2 - \sum_i (y_i - f(x_i))^2 \right)}{\left( \frac{1}{n - p_f} \sum_i (y_i - f(x_i))^2 \right)}$$

# F-Distribution

- Approaches gaussian about 1 for large N

$$\begin{aligned}
 f(x; d_1, d_2) &= \frac{\sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1 + d_2}}}}{x B\left(\frac{d_1}{2}, \frac{d_2}{2}\right)} \\
 &= \frac{1}{B\left(\frac{d_1}{2}, \frac{d_2}{2}\right)} \left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}} x^{\frac{d_1}{2}-1} \left(1 + \frac{d_1}{d_2} x\right)^{-\frac{d_1+d_2}{2}}
 \end{aligned}$$



The F is for Fisher  
He's the dapper gent above

# Higher Order Polynomial

- $$\frac{\frac{RSS_1 - RSS_2}{p_2 - p_1}}{\frac{RSS_2}{n - p_2}} \approx F_{p_2 - p_1, n - p_2}$$
 generically
- $$\frac{\frac{\mathcal{L}_1 - \mathcal{L}_2}{p_2 - p_1}}{\frac{\mathcal{L}_2}{n - p_2}} \approx F_{p_2 - p_1, n - p_2}$$



Test of higher polynomial order with F-test form is called the chow test

# Higher Order Polynomial

- $$\frac{\frac{RSS_1 - RSS_2}{p_2 - p_1}}{\frac{RSS_2}{n - p_2}} \approx F_{p_2 - p_1, n - p_2}$$
 generically
- $$\frac{\frac{\mathcal{L}_1 - \mathcal{L}_2}{p_2 - p_1}}{\frac{\mathcal{L}_2}{n - p_2}} \approx F_{p_2 - p_1, n - p_2}$$



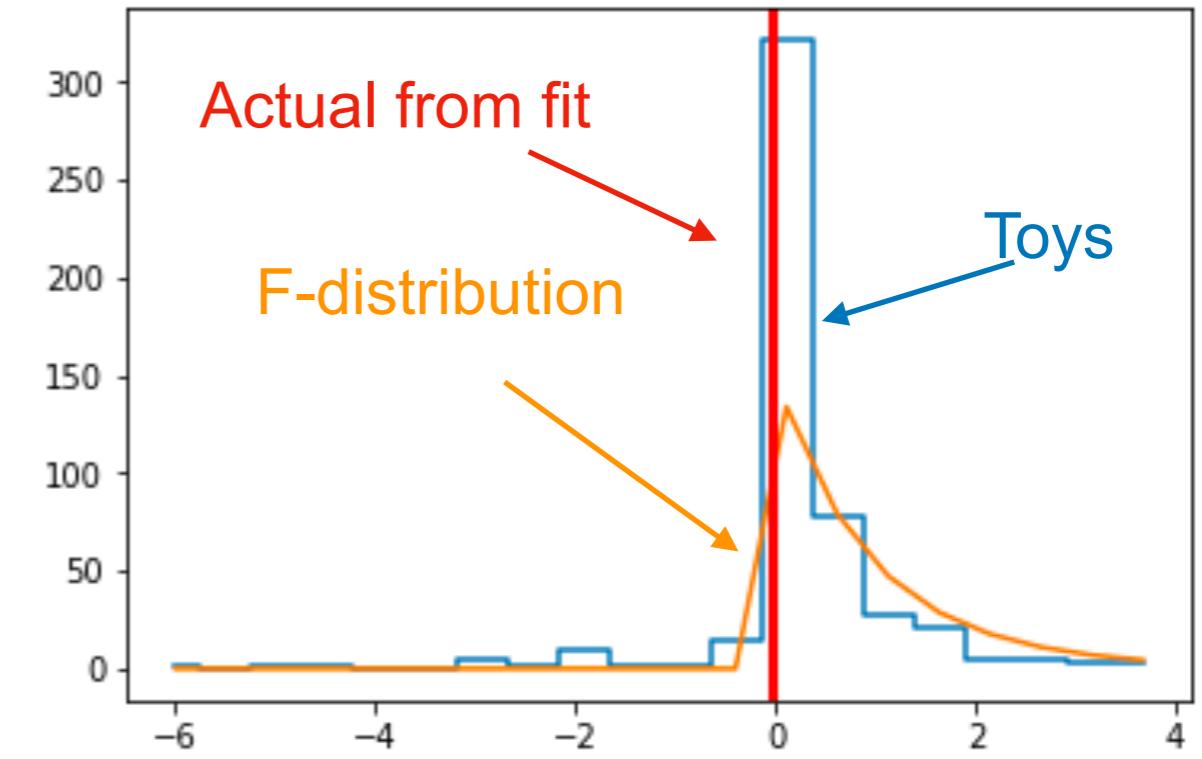
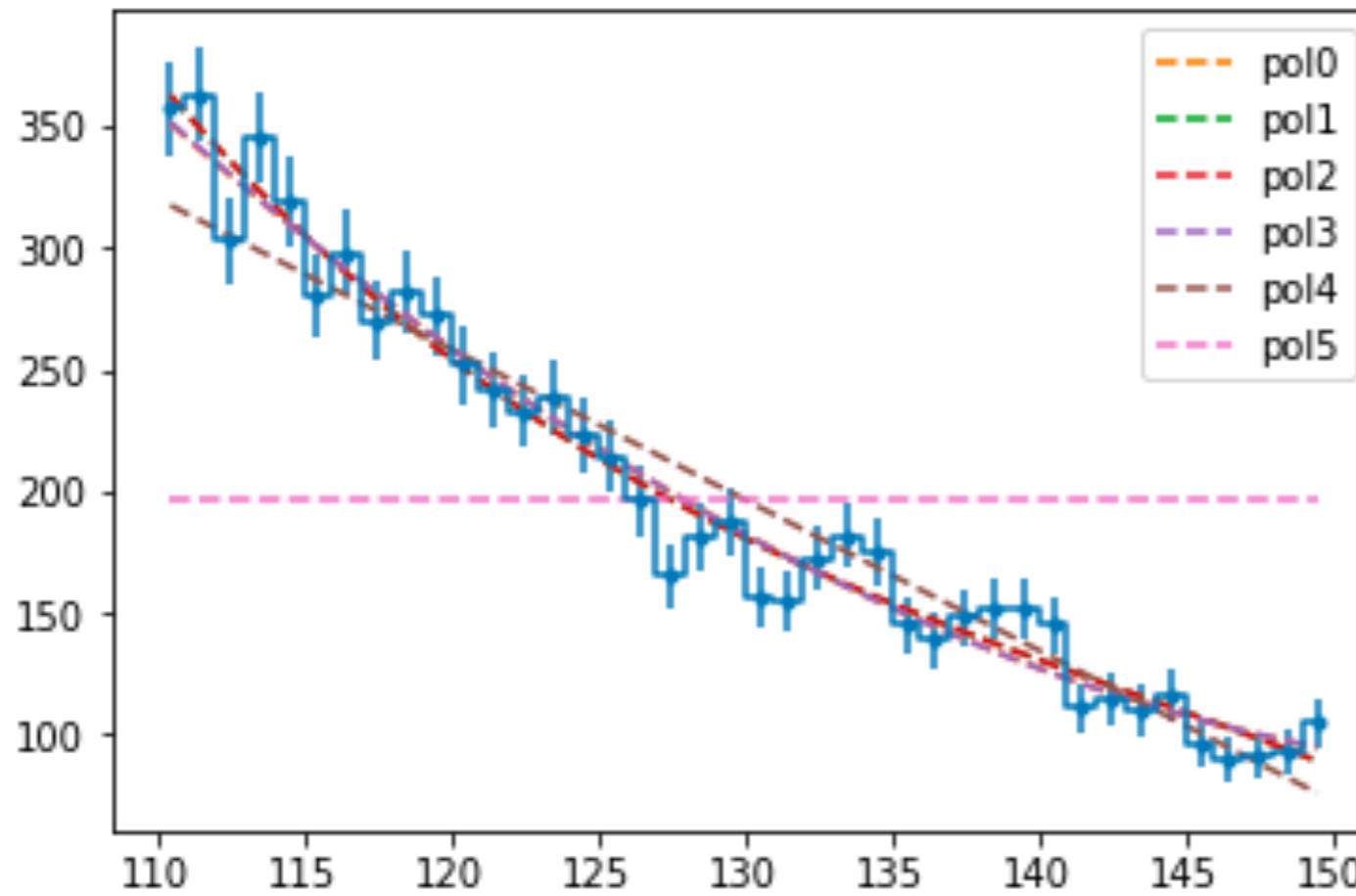
# Frequentist F-test

- $n$ -groups of **fits** each with separate fitted likelihoods
- Define :  $\bar{X}_j = \frac{1}{m} \sum_{i=1}^m X_i$  and  $S_j^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X}_j)^2$
- Define:  $MS_B = \frac{m}{n-1} \sum_{i=1}^n (\bar{X}_j - \bar{X})^2$  and  $MS_R = \sigma^2$   
Difference in likelihoods
- If  $\mu_i = \mu$  or in other words are from the same distribution the
- $\frac{MS_B}{MS_R} \approx 1 = F_{n-1, m(n-1)}$  where  $F_{n-1, m(n-1)}(x)$  is an F distrib

# F-test Example

- Here is an example from previous fit
- Our Actual  $\Delta$  (x-axis) is consist with a high p-value

Toys: Randomly sample 3rd order dist and fit with 4th order

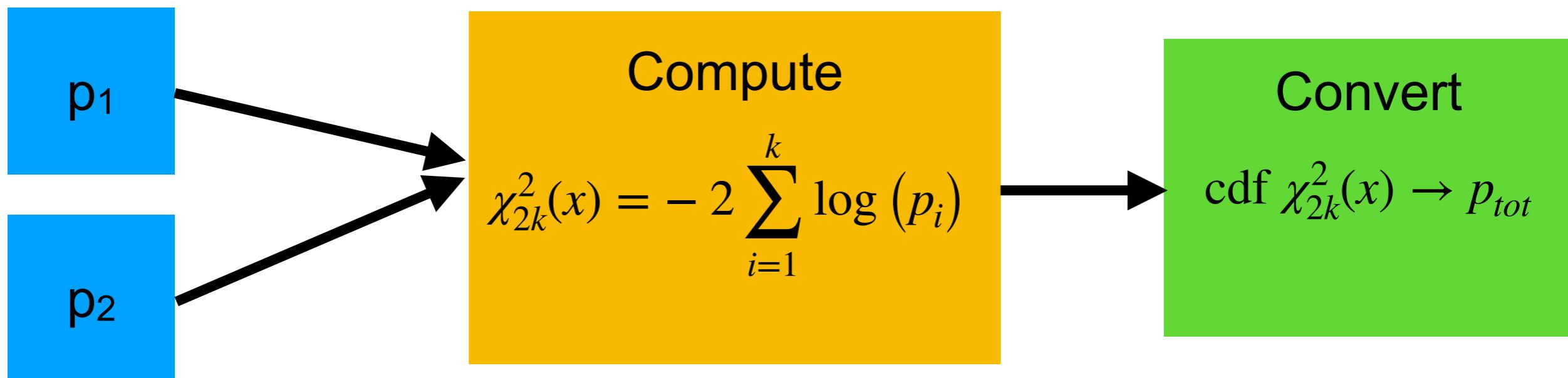


$$\frac{\mathcal{L}_1 - \mathcal{L}_2}{\frac{p_2 - p_1}{\mathcal{L}_2}} \quad \text{For a 4th order to a 3rd order}$$

# Combining Categories

- Lets say we have  $k$  measurements with probability  $p_i$ 
  - Each measurement is independent of each other
- We can combine categories using the following formula

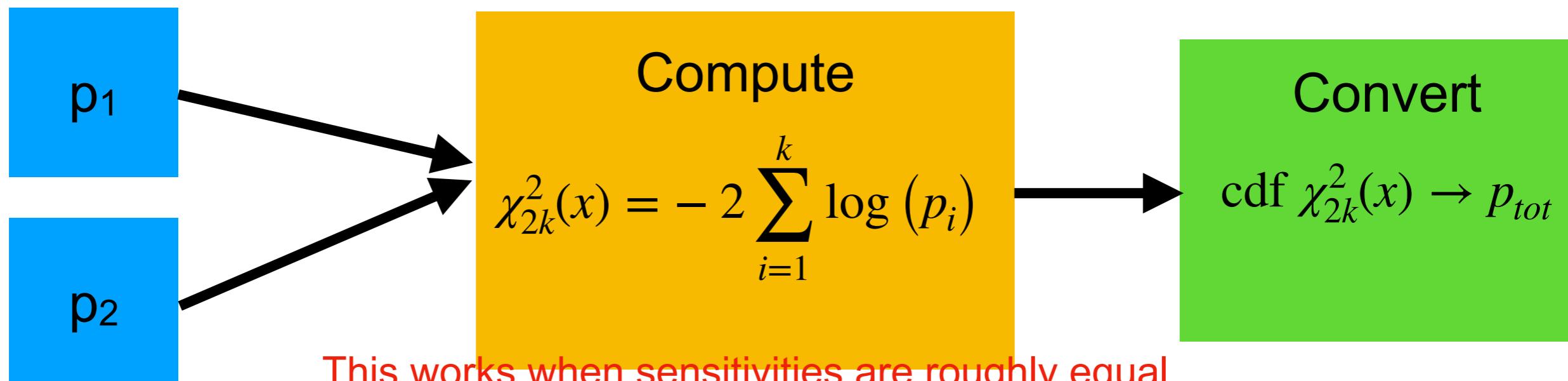
- $\chi^2_{2k}(x) = - 2 \sum_{i=1}^k \log(p_i)$
- Where left is a chi2 distribution with  $2k$  degrees of freedom



# Combining Categories

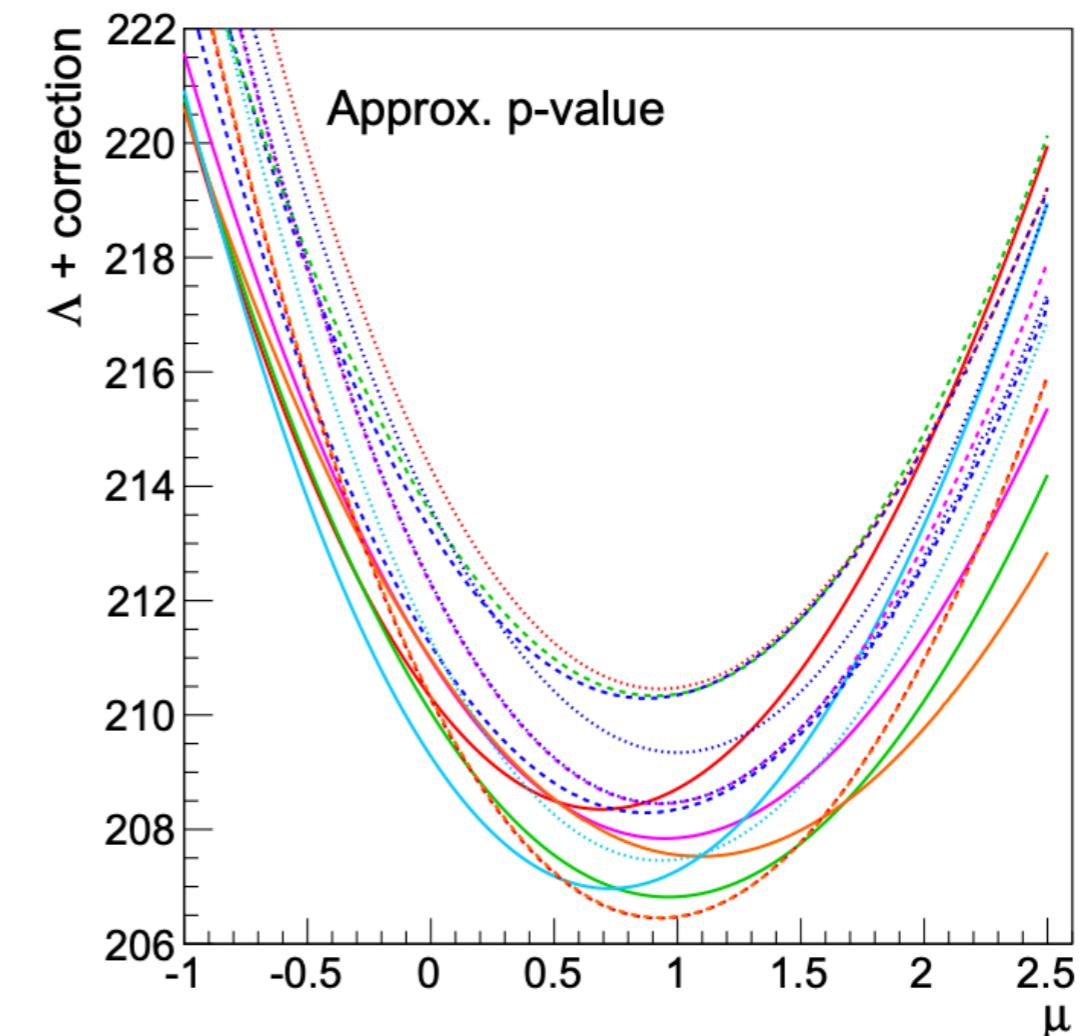
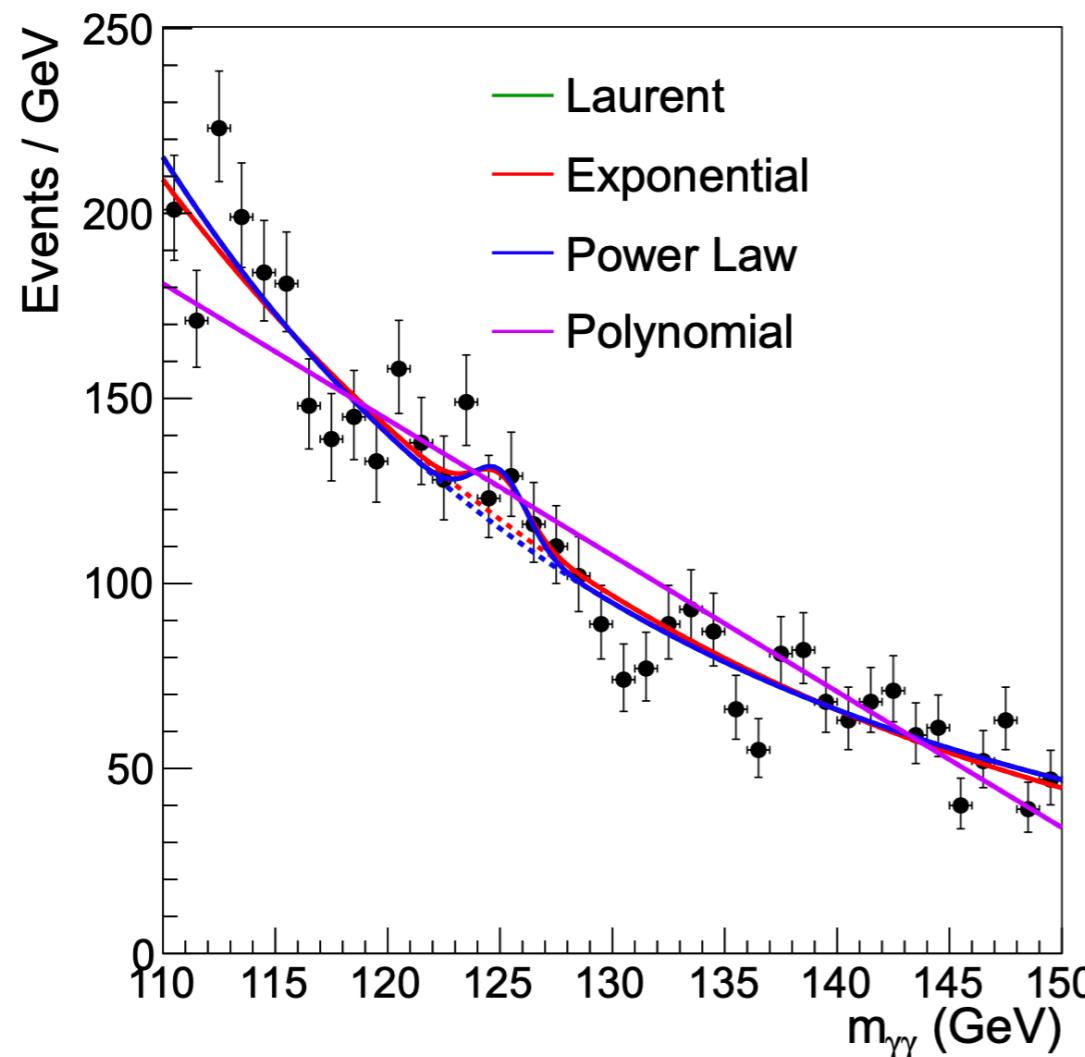
- Lets say we have  $k$  measurements with probability  $p_i$ 
  - Each measurement is independent of each other
- We can combine categories using the following formula

- $\chi^2_{2k}(x) = - 2 \sum_{i=1}^k \log(p_i)$
- Where left is a chi2 distribution with  $2k$  degrees of freedom



# Building a Model

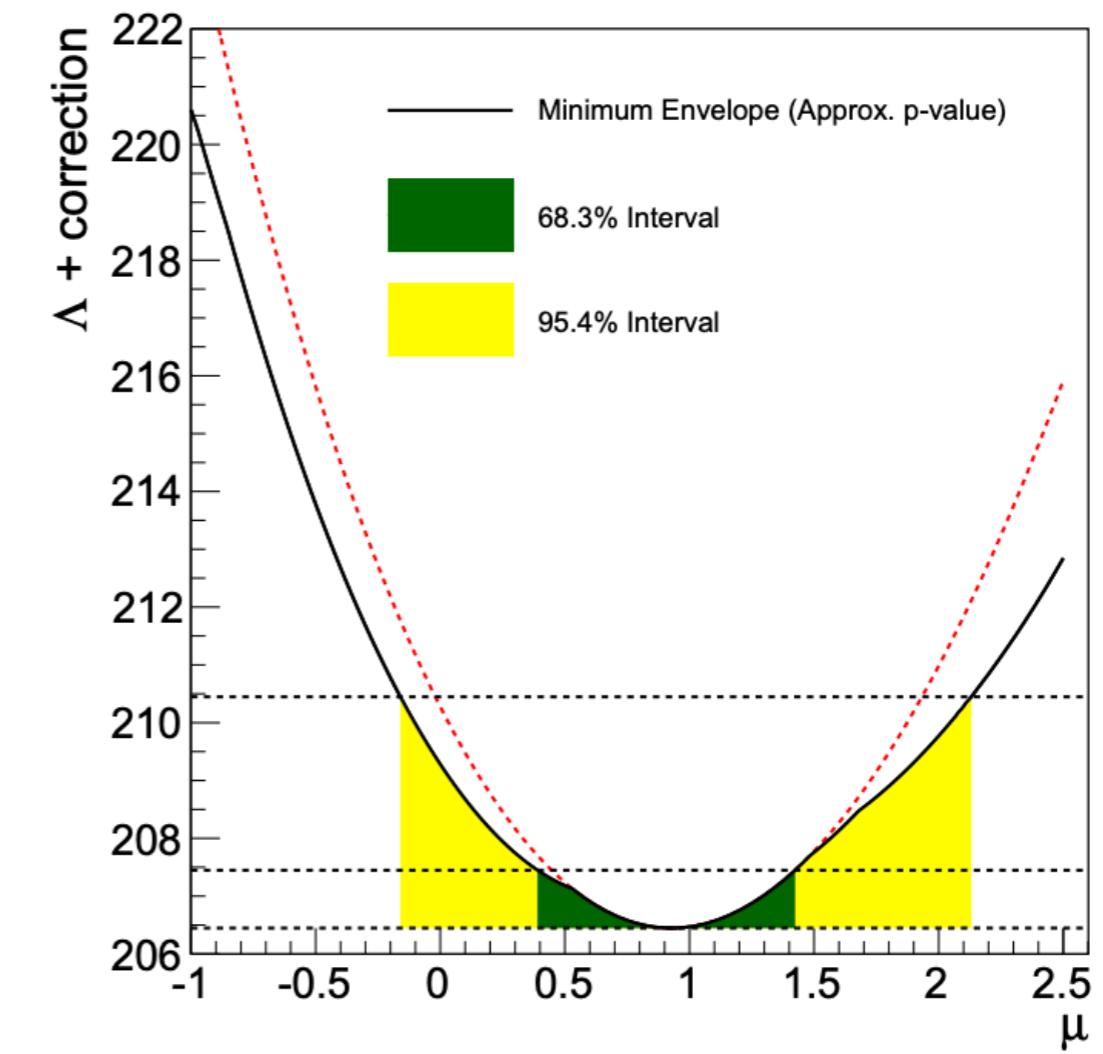
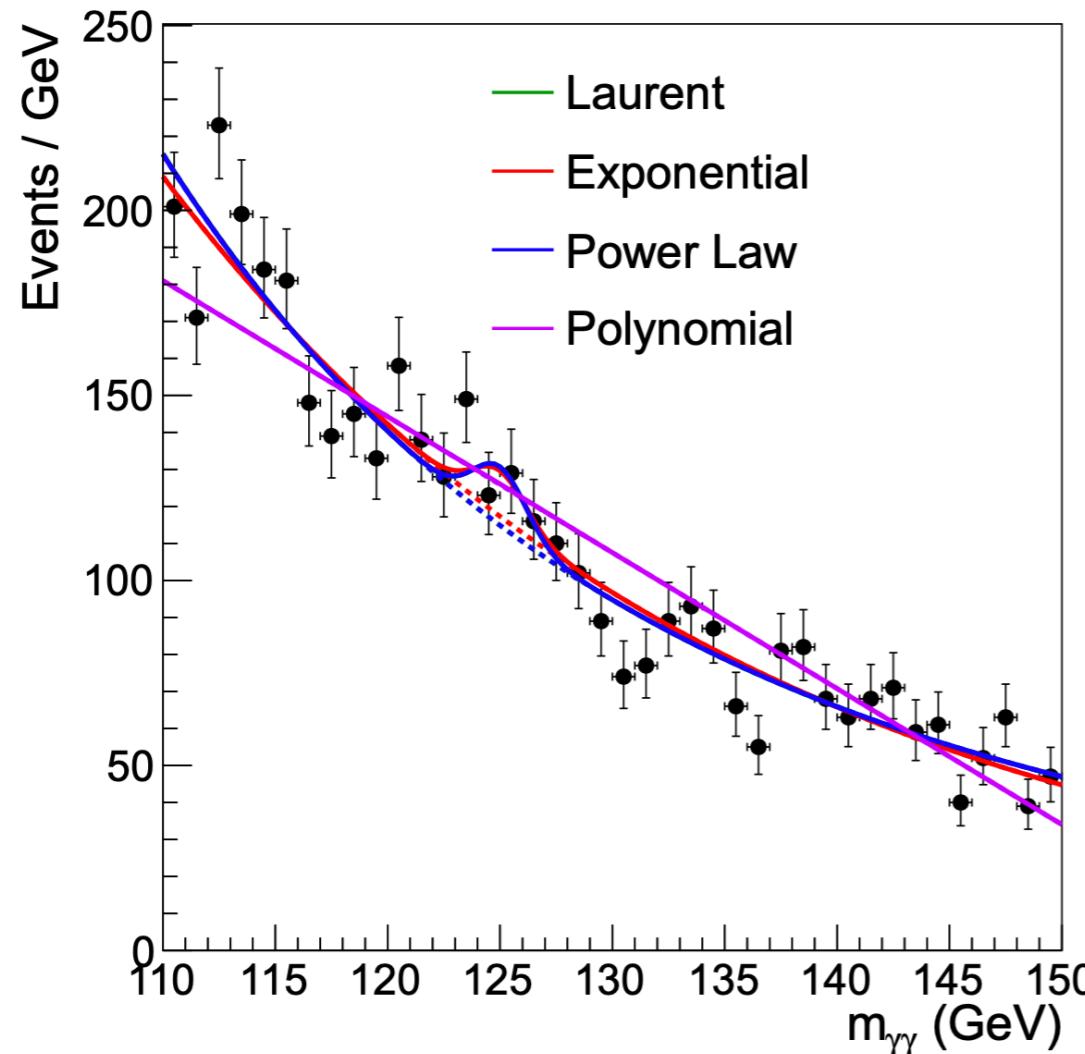
- Throwing a barrage of functions at the problem



We can try a whole library of functions  
 The likelihood we get translates to our fit

# Building a Model

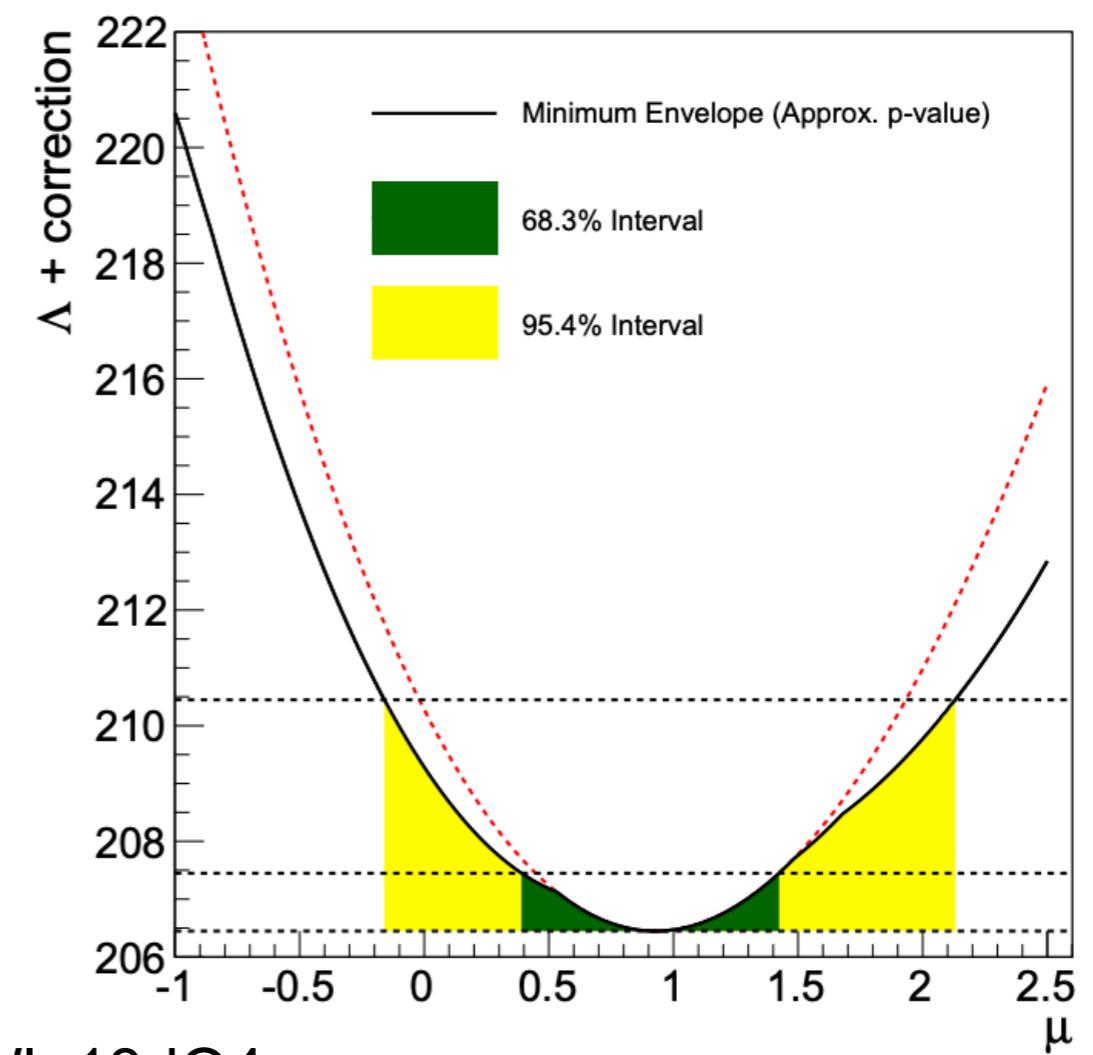
- Throwing a barrage of functions at the problem



We can try a whole library of functions  
The likelihood we get translates to our fit

# Combined Likelihood

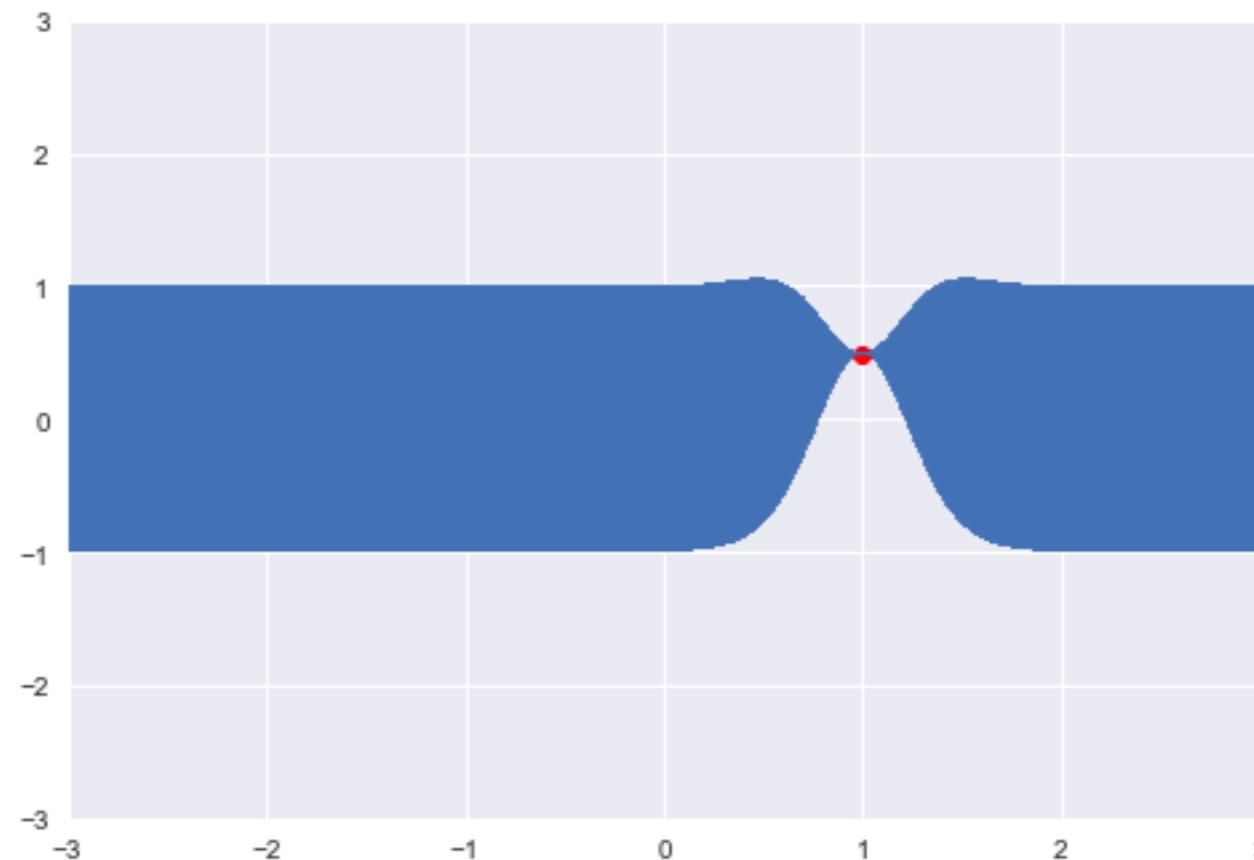
- Throwing a barrage of functions at the problem



<https://www.youtube.com/watch?v=3cHWIp13dQ4>

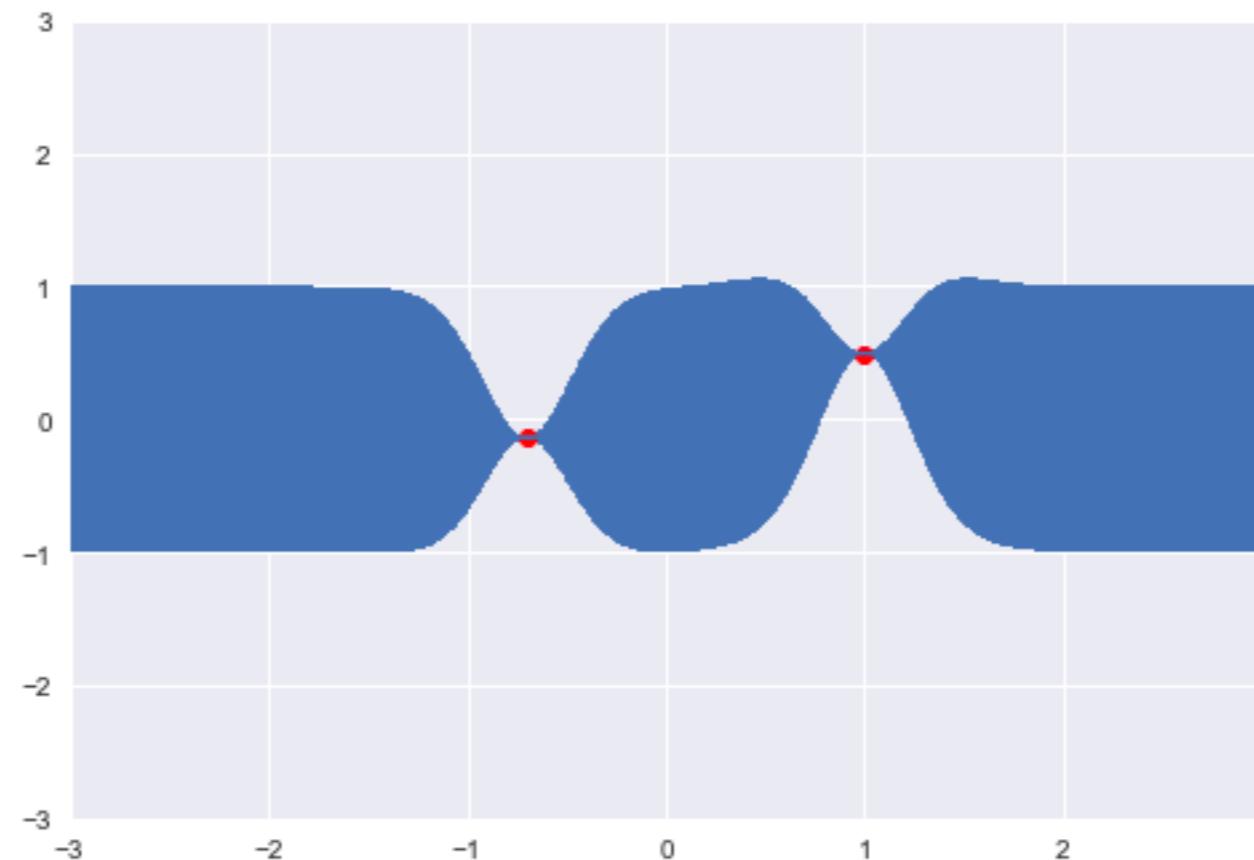
We can try a whole library of functions  
The likelihood we get translates to our fit

# Splines+GaussianProcesses



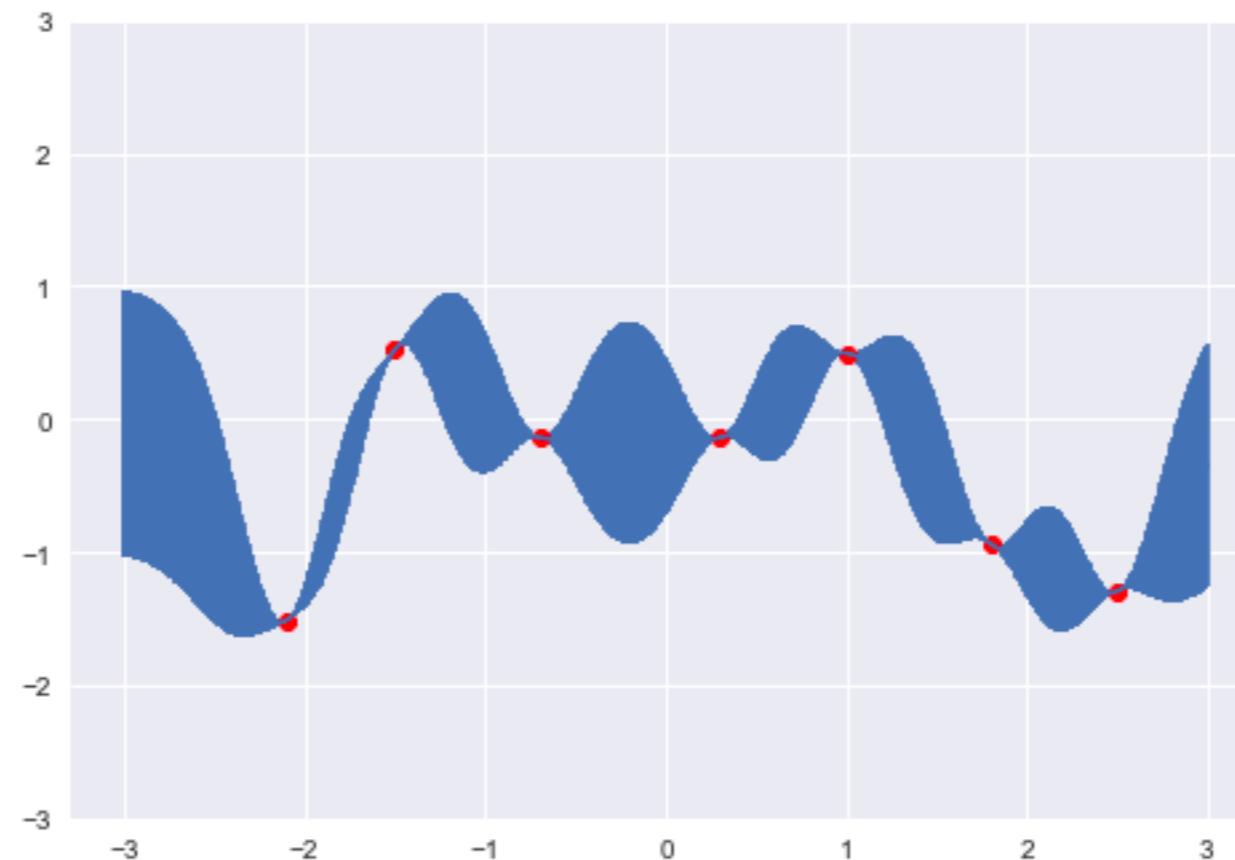
- Gaussian processes allow us to build function choice from the data

# Splines+GaussianProcesses



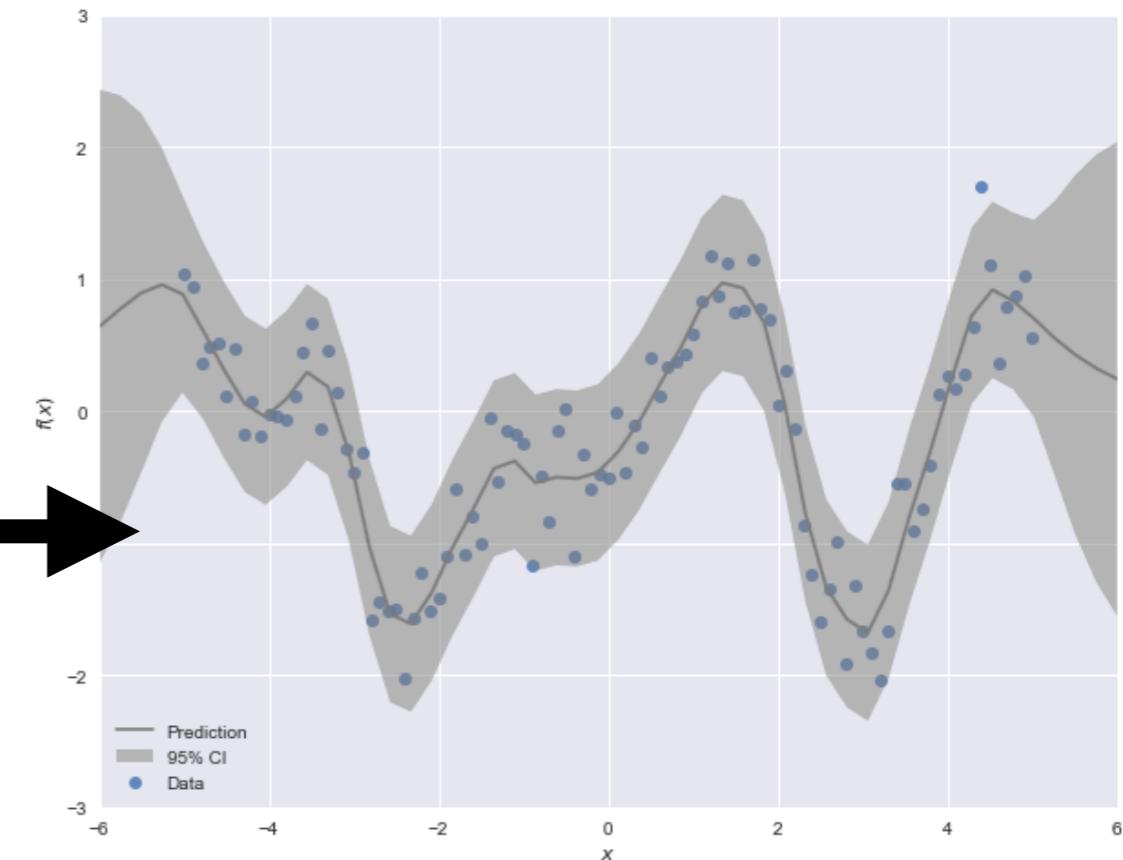
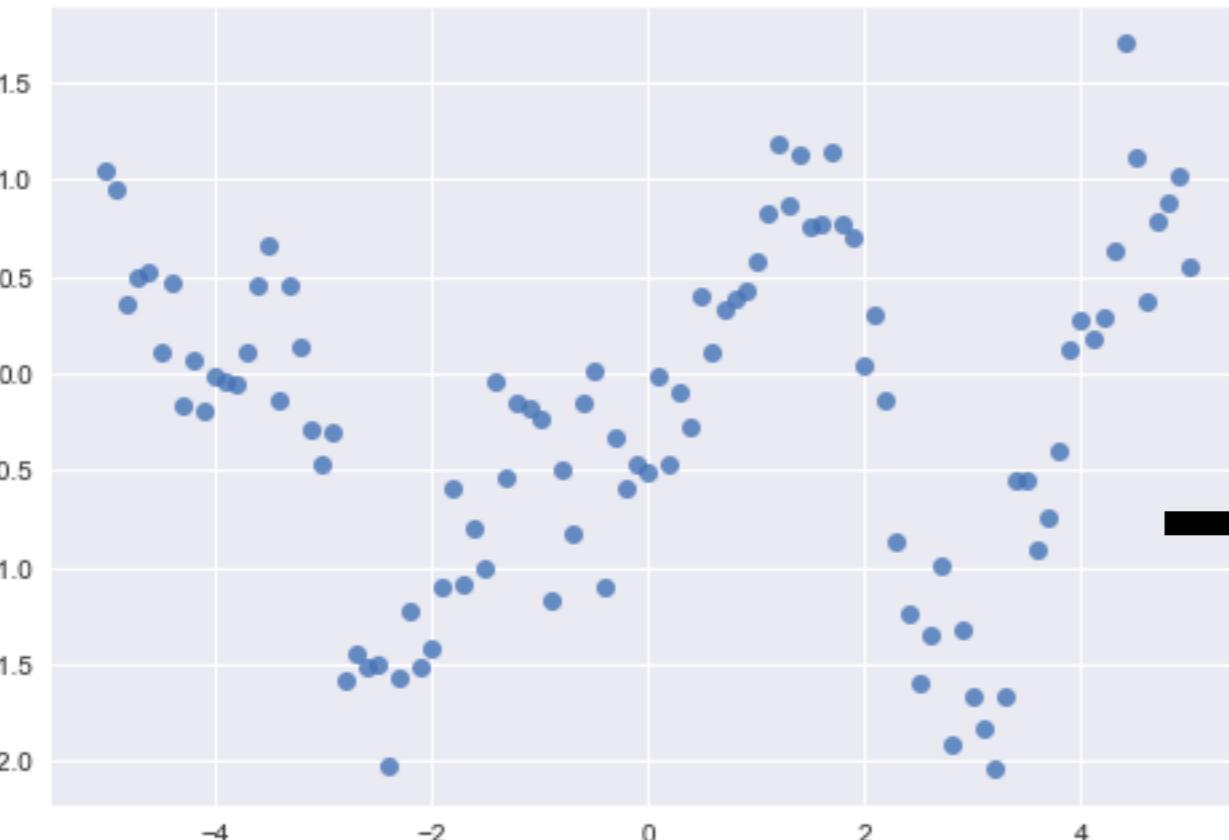
- Gaussian processes allow us to build function choice from the data

# Splines+GaussianProcesses



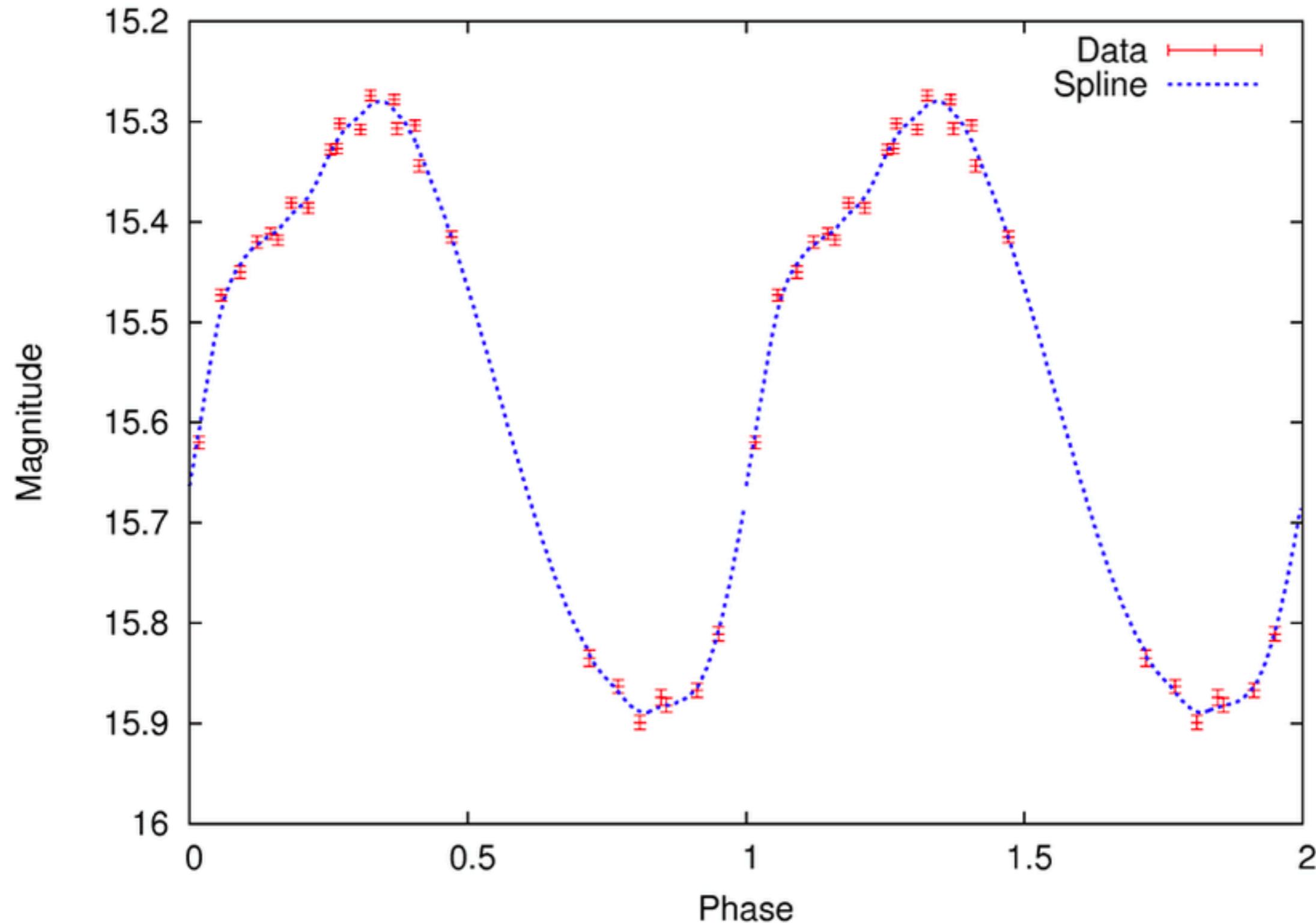
- Gaussian processes allow us to build function choice from the data

# Splines+GaussianProcesses

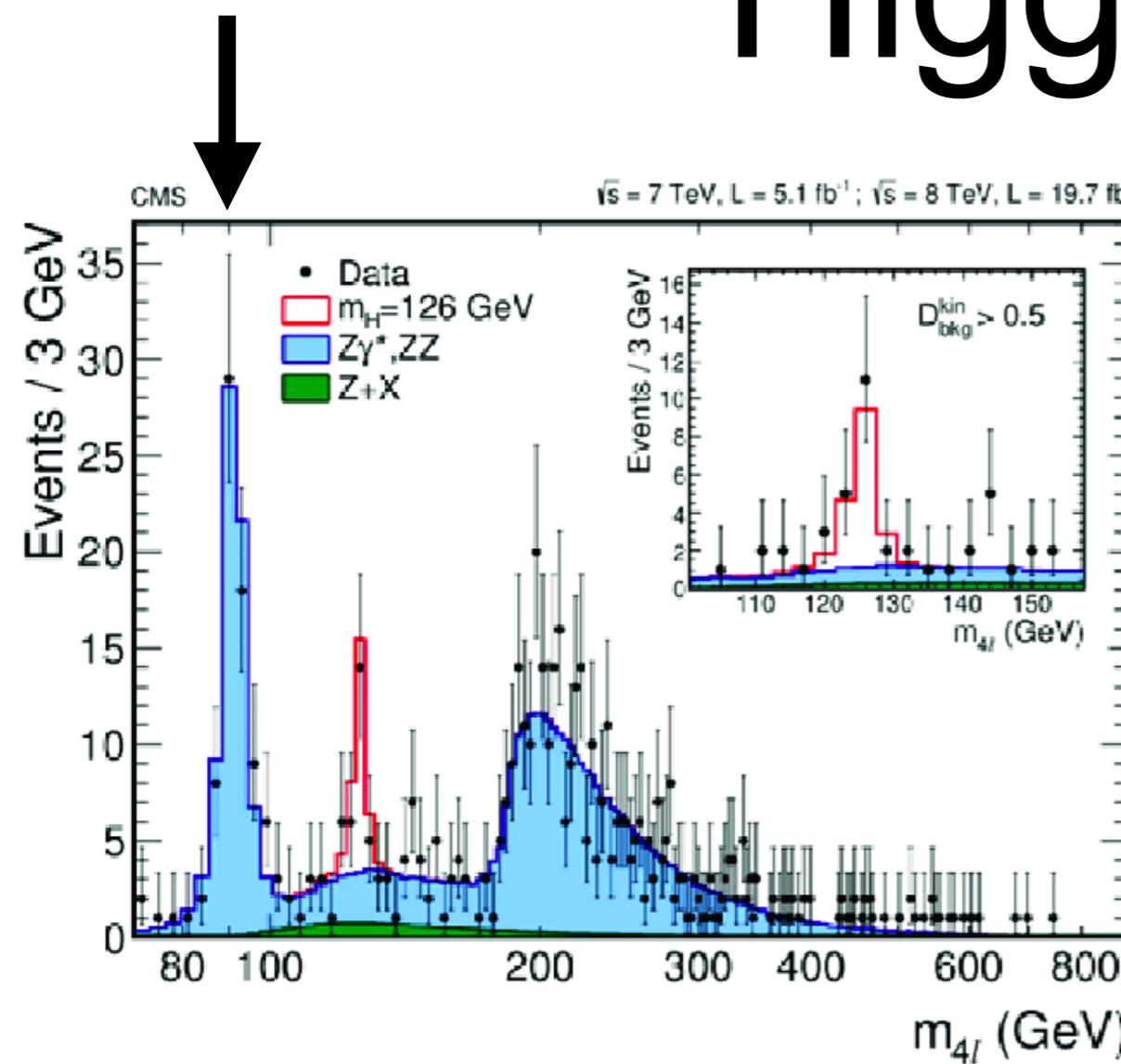


- To go from points to spline automatically

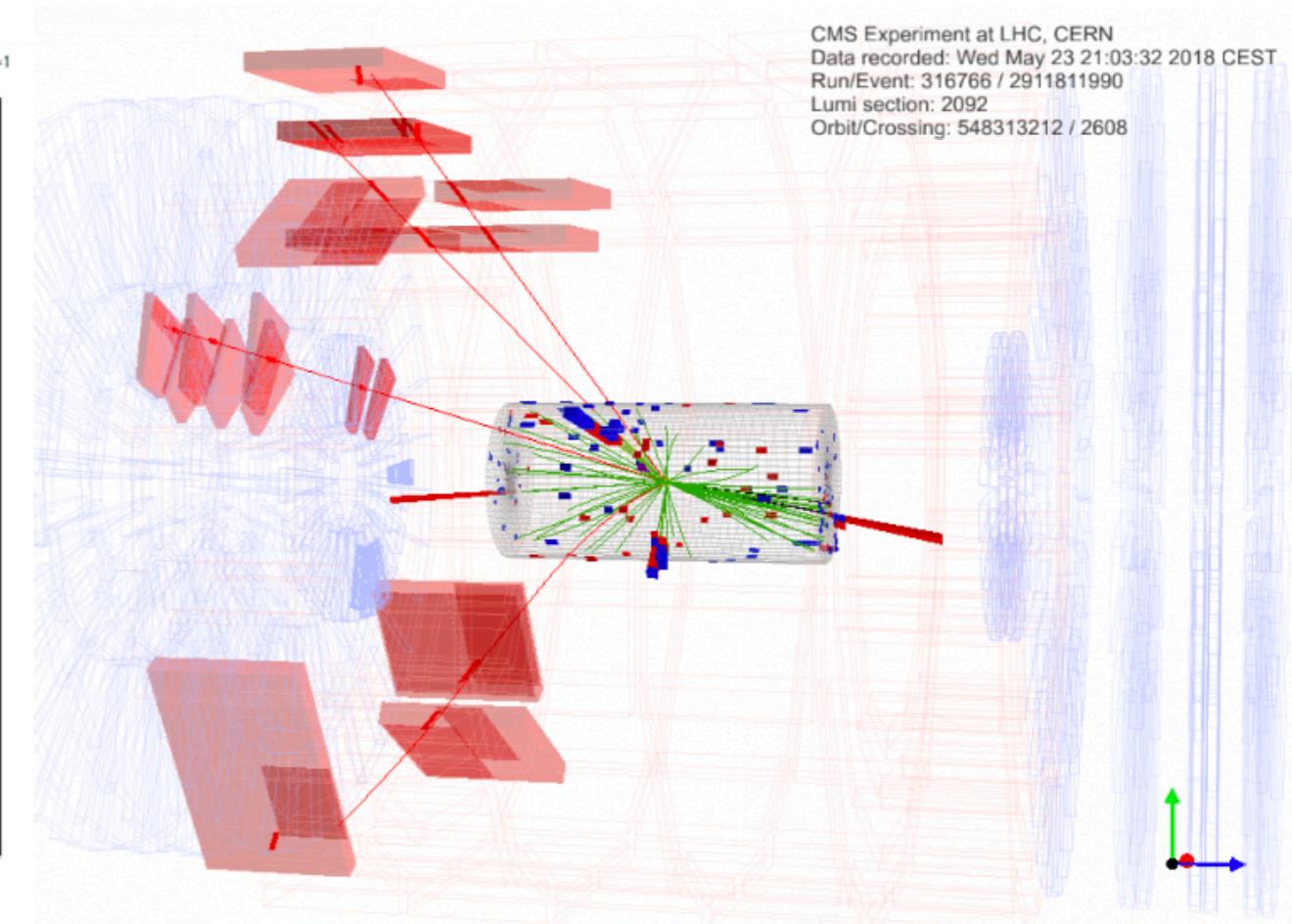
# Splines+GaussianProcesses



Z-boson Peak



# Higgs to 4 Leptons



- Higgs to 4 leptons aims at taking the mass of 4 leptons
  - A way to test the 4-leptons is the Z boson peak



# Backup

# Remind me at some point



**Explain the Chow Test**

# Higher Order Polynomial

- We can evaluate this through an F-test

- Recall  $\frac{MS_B}{MS_R} \approx 1 = F_{n-1, m(n-1)}$

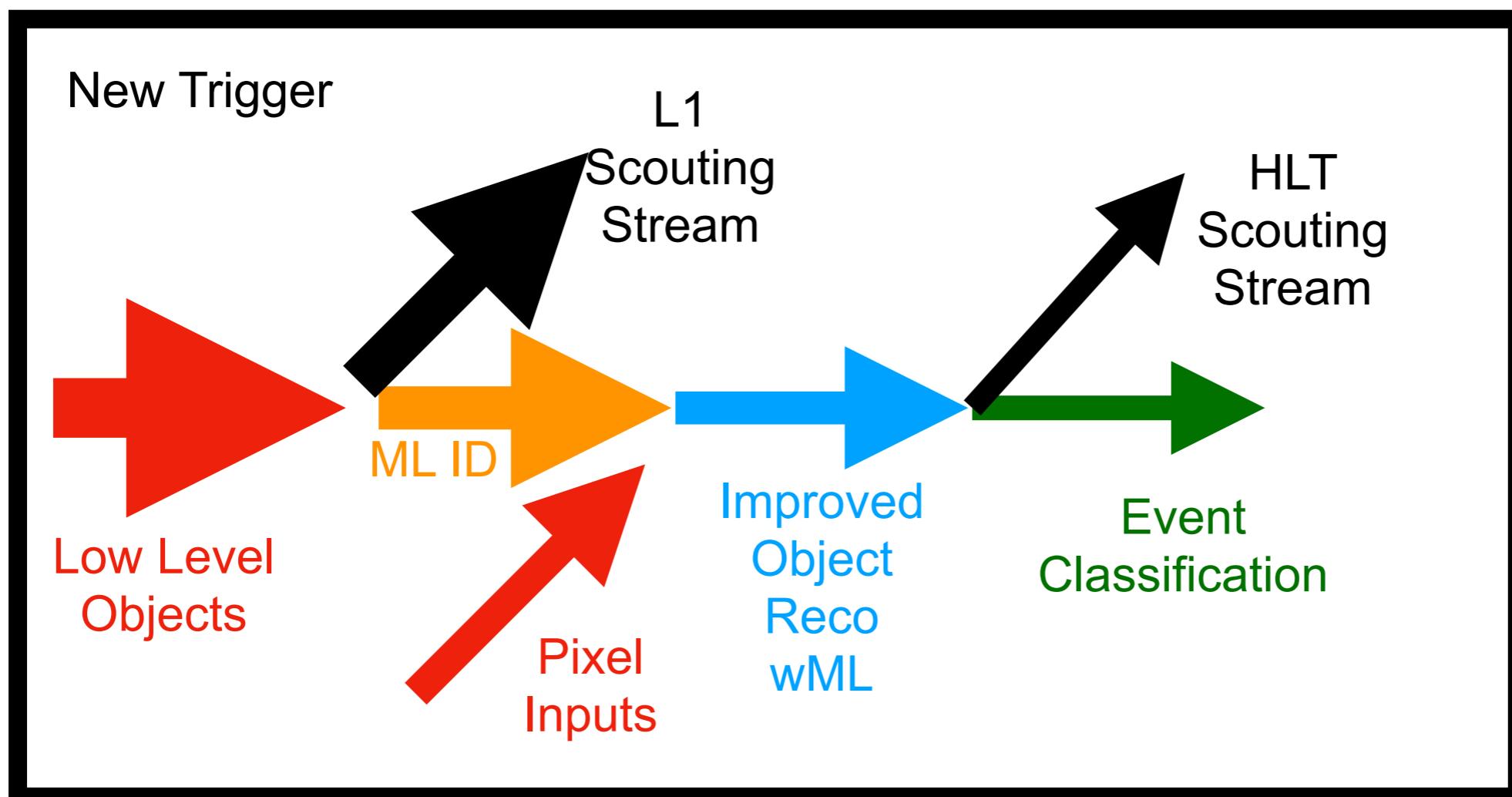


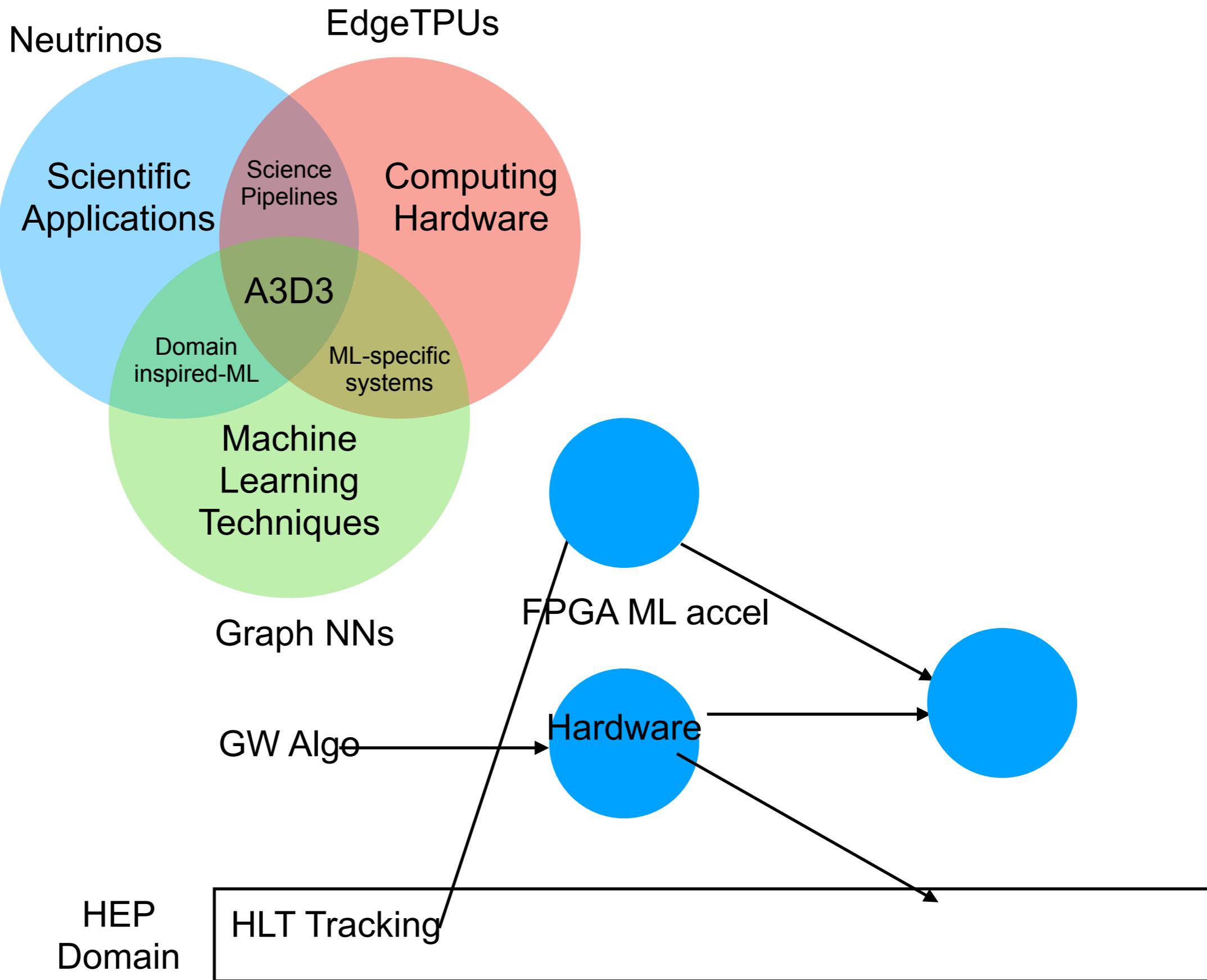
Test of higher polynomial order with F-test form is called the chow test

- Here we have  $\frac{MS_B}{MS_R} \approx 1 = F_{p_2-p_1, n-p_2}$

$$F = \frac{\left( \frac{RSS_1 - RSS_2}{p_2 - p_1} \right)}{\left( \frac{RSS_2}{n-p_2} \right)},$$

# Title Text





# Elastic Scatter

