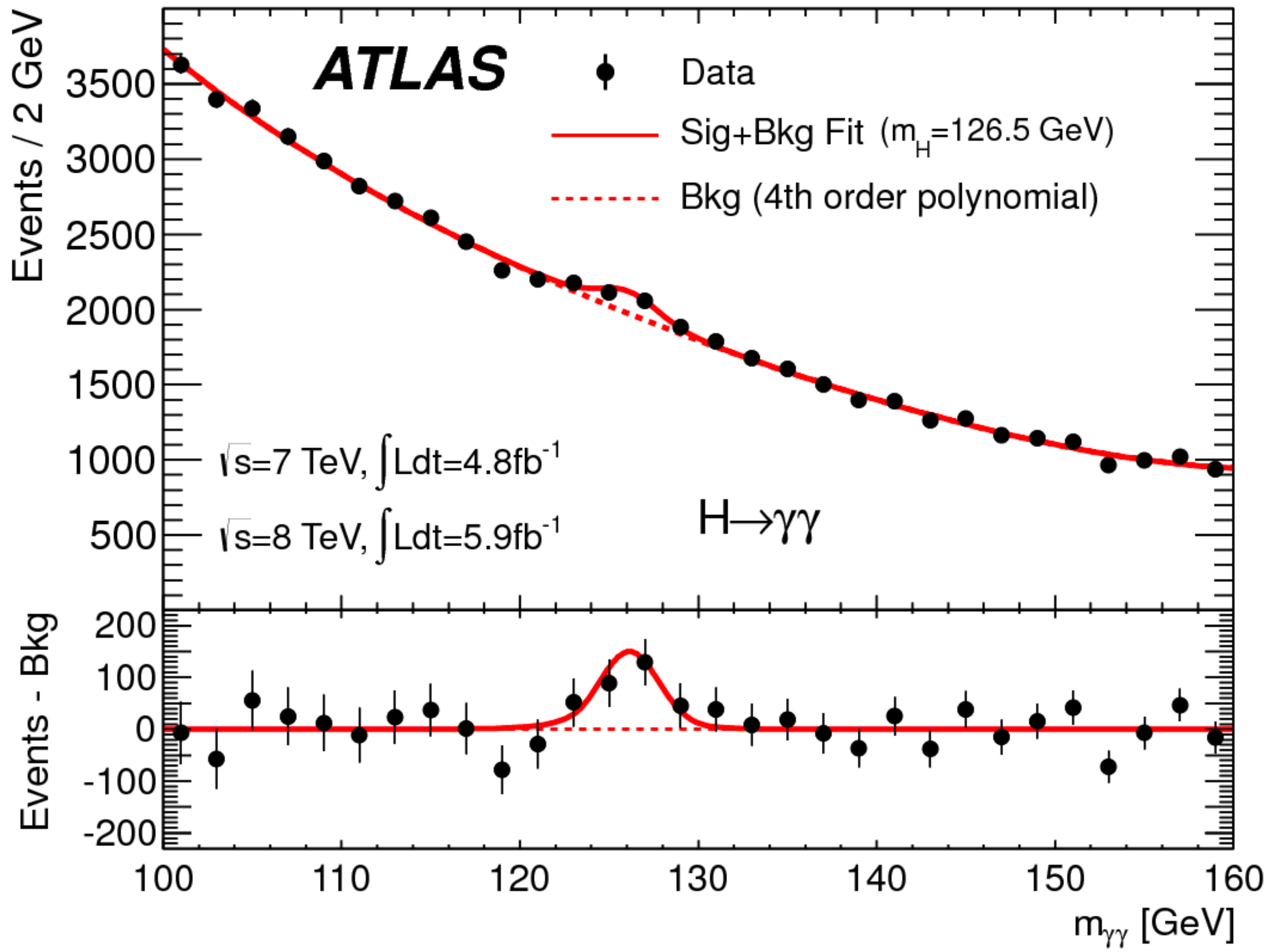


Lecture 10

Higgs Boson Discovery
Towards Deep Learning

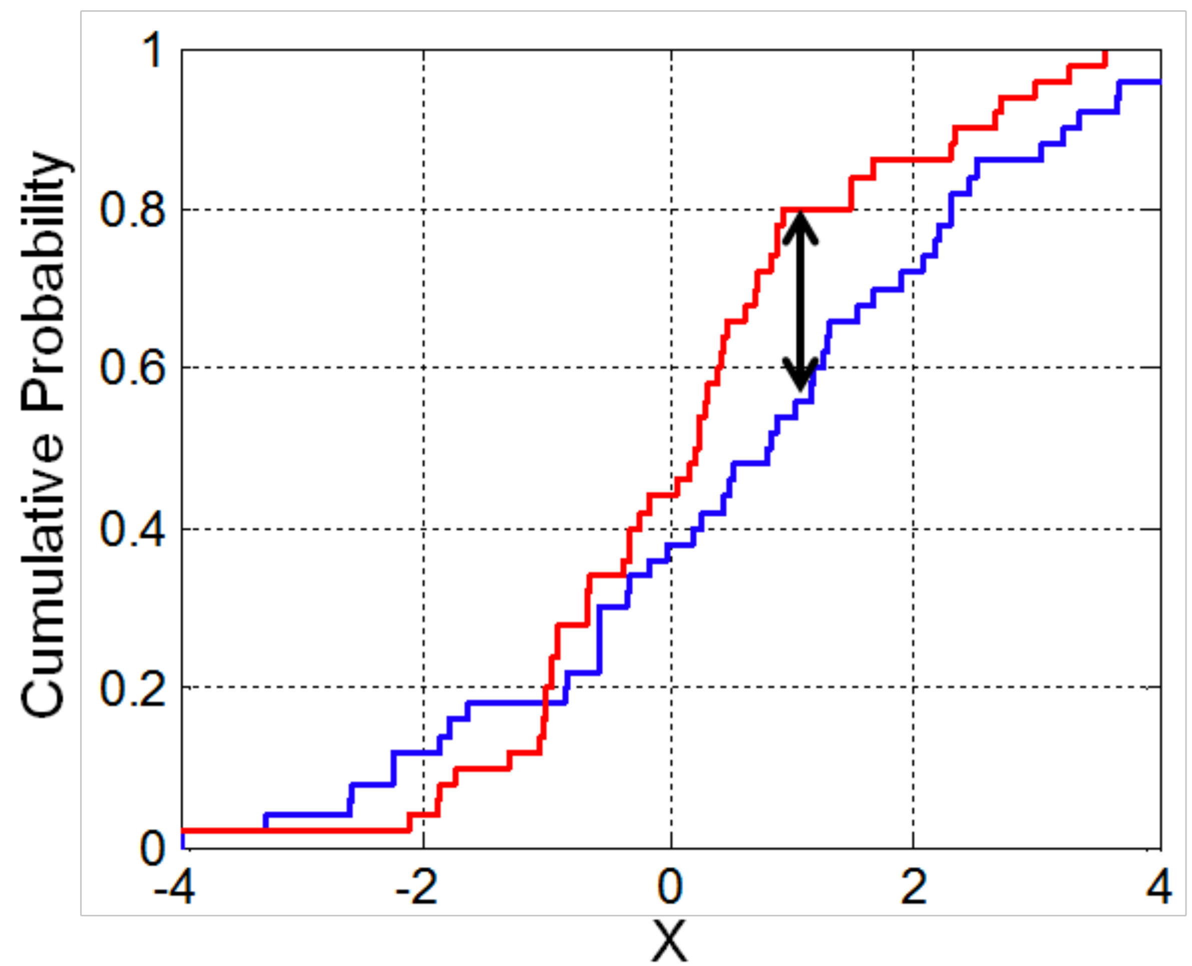
KS-test

- 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

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

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

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 a chi2 test

$\mathscr{L}(\theta_i)\forall\theta_i \in \Theta_0 \text{where } \Theta_0 \text{ refers to null} - \text{hypothesis}$

$\mathscr{L}(\theta_i)\forall\theta_i \in \Theta \text{where } \Theta \text{ refers to alternative} - \text{hypothesis}$

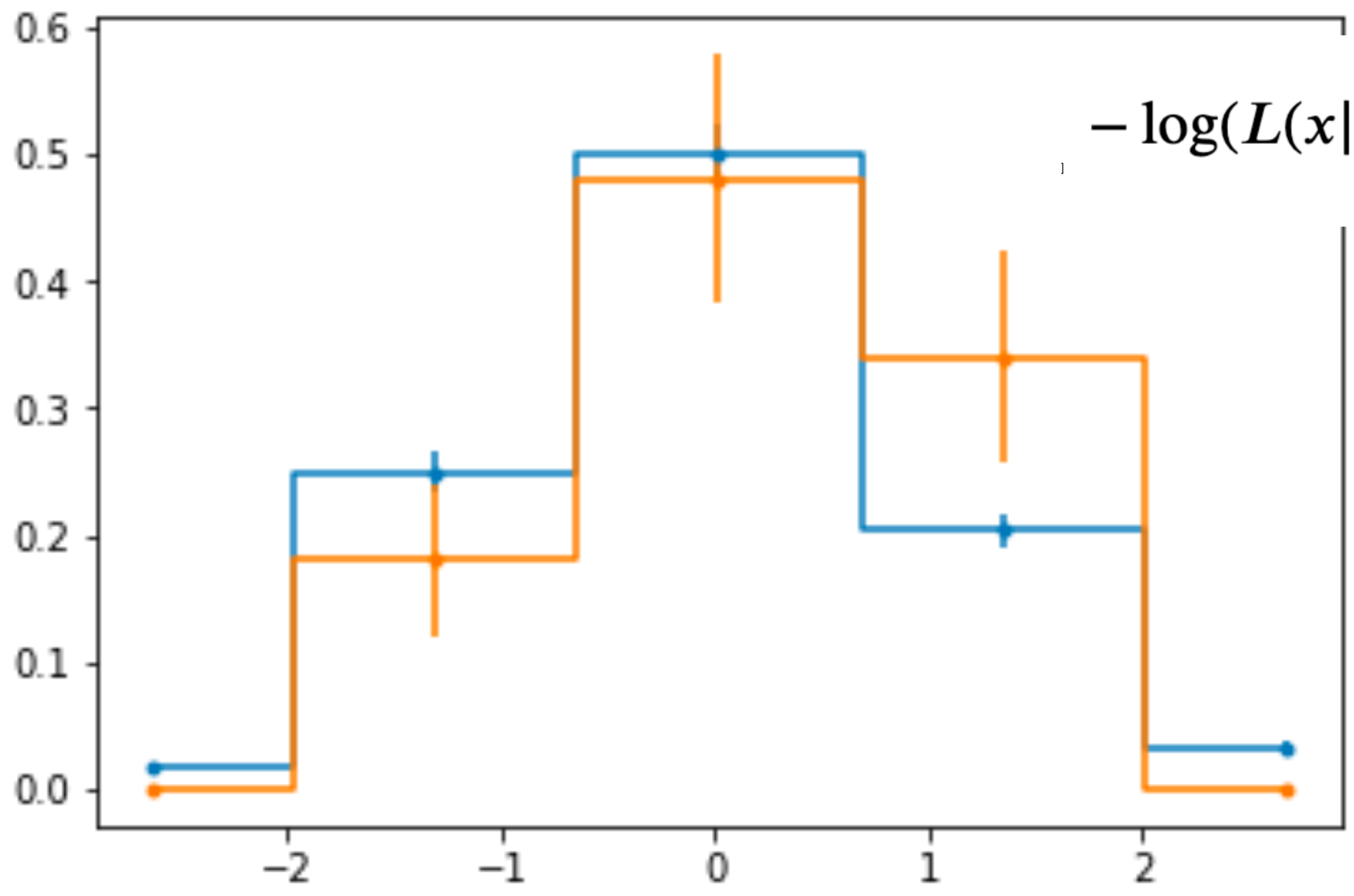
$\sup_{\theta\in\Theta_0}\mathscr{L}(\theta)$ is the maximum likelihood over all θ

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

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

Likelihood Test: Concept

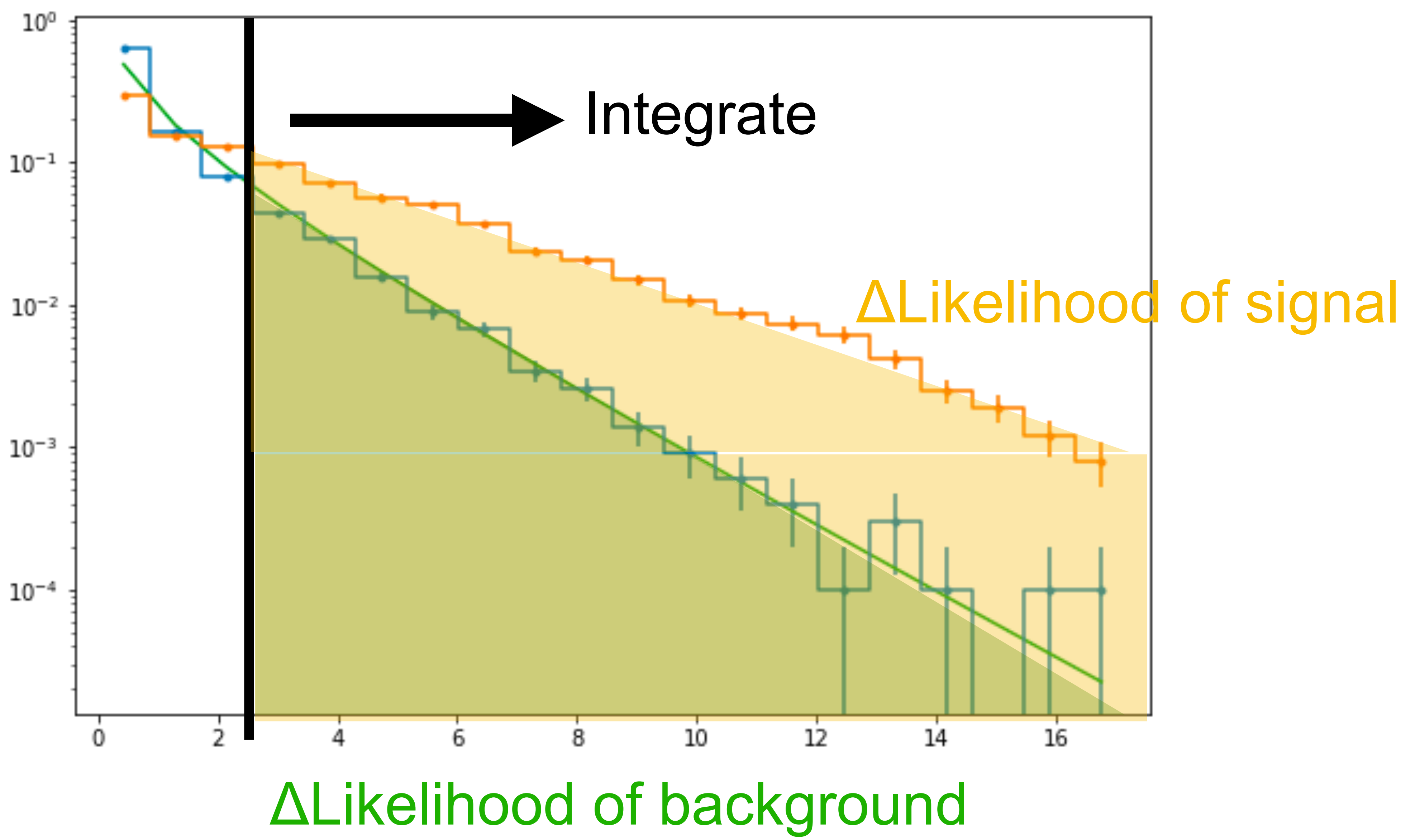
- We would like to see if these to histograms are consistent
  - Hypothesis: Orange is a guassian with mean 0 (blue)
  - True: Orange is sampled from a gaussian with mean 0.2



$$-\log(L(x|\lambda)) = \sum_{i-1}^{N} -\frac{1}{2}\log(2\pi\sigma^2) + \frac{(x_i-\mu)^2}{2\sigma^2}$$

Likelihood Test: Gaussian

- We can see how these on average a deviation



Frequentist: t-test

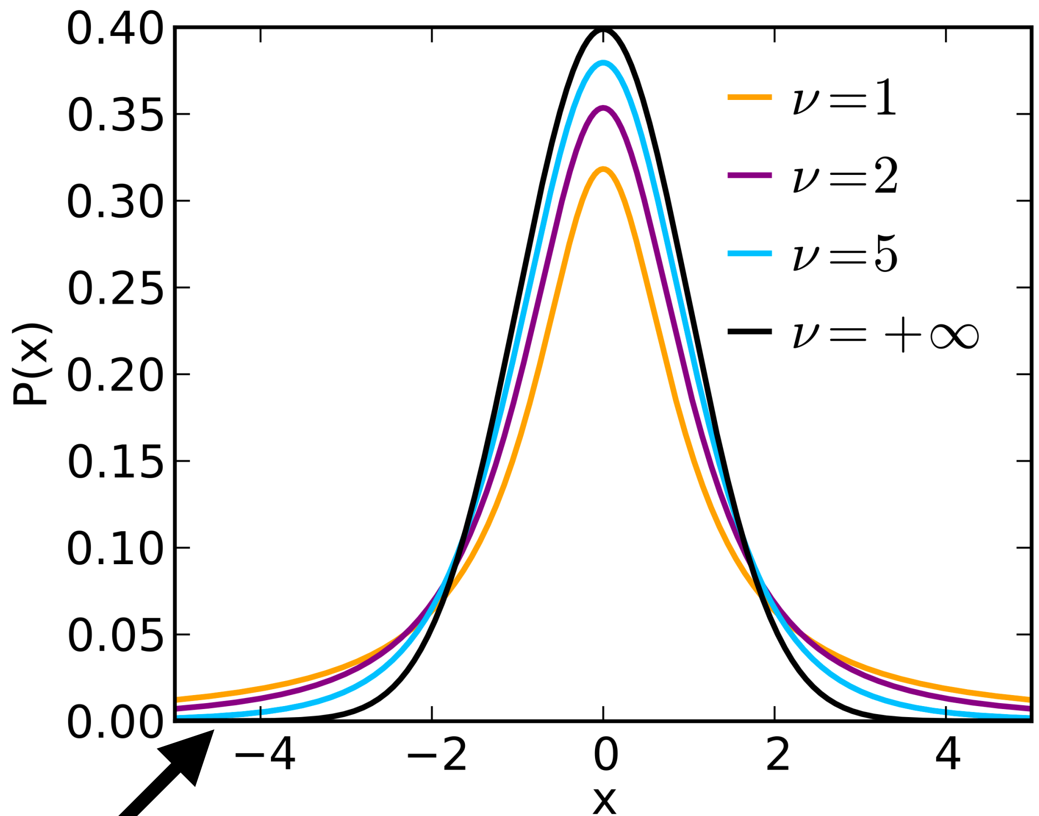
- Perform a t-test:
  - If samples $X_1, X_2, \ldots X_N$ from the same underlying distrbtion
  - Underlying distribution is Gaussian with mean μ and variance σ^2
  - If the above is true then we nave a measured mean+variance:
  - given by $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ and $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i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

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

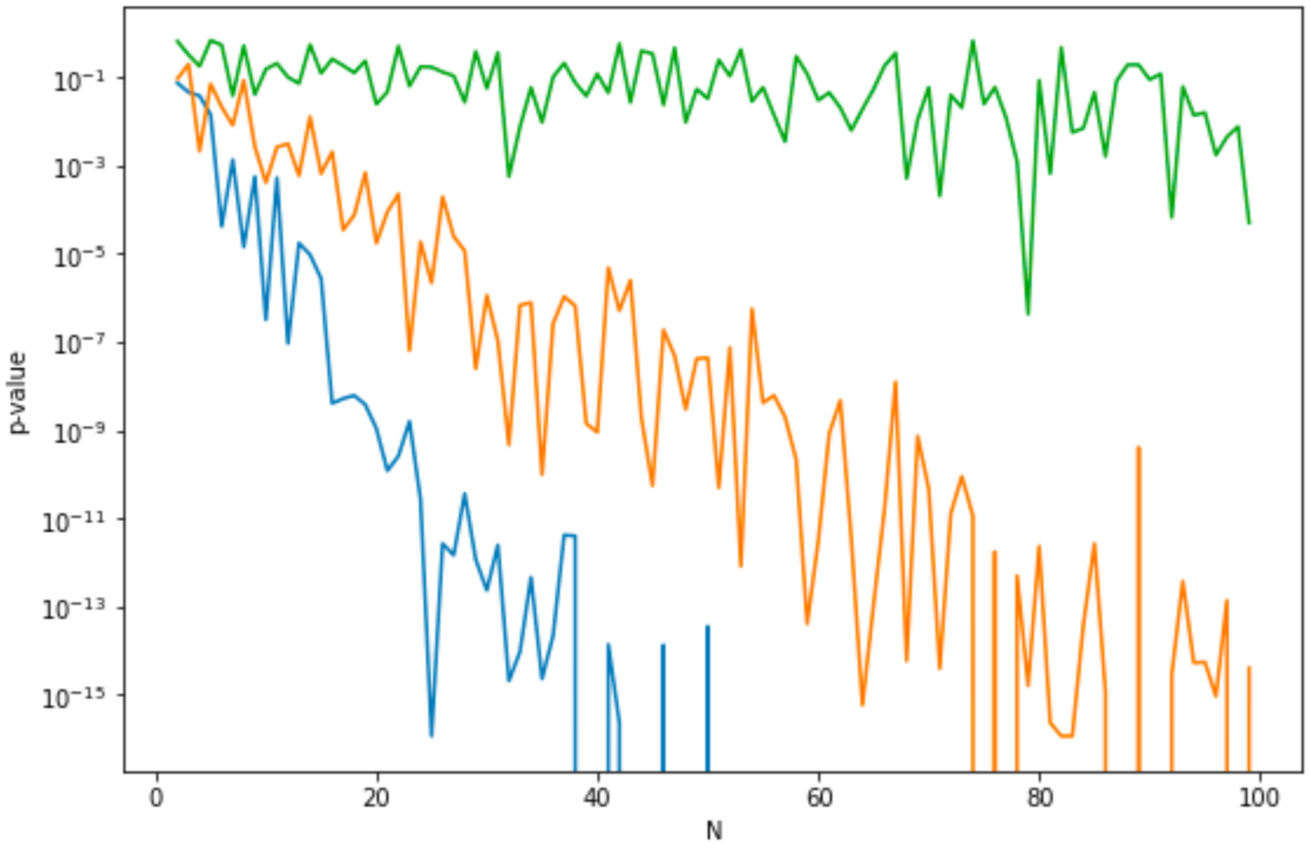
$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \qquad \nu = N - 1$$

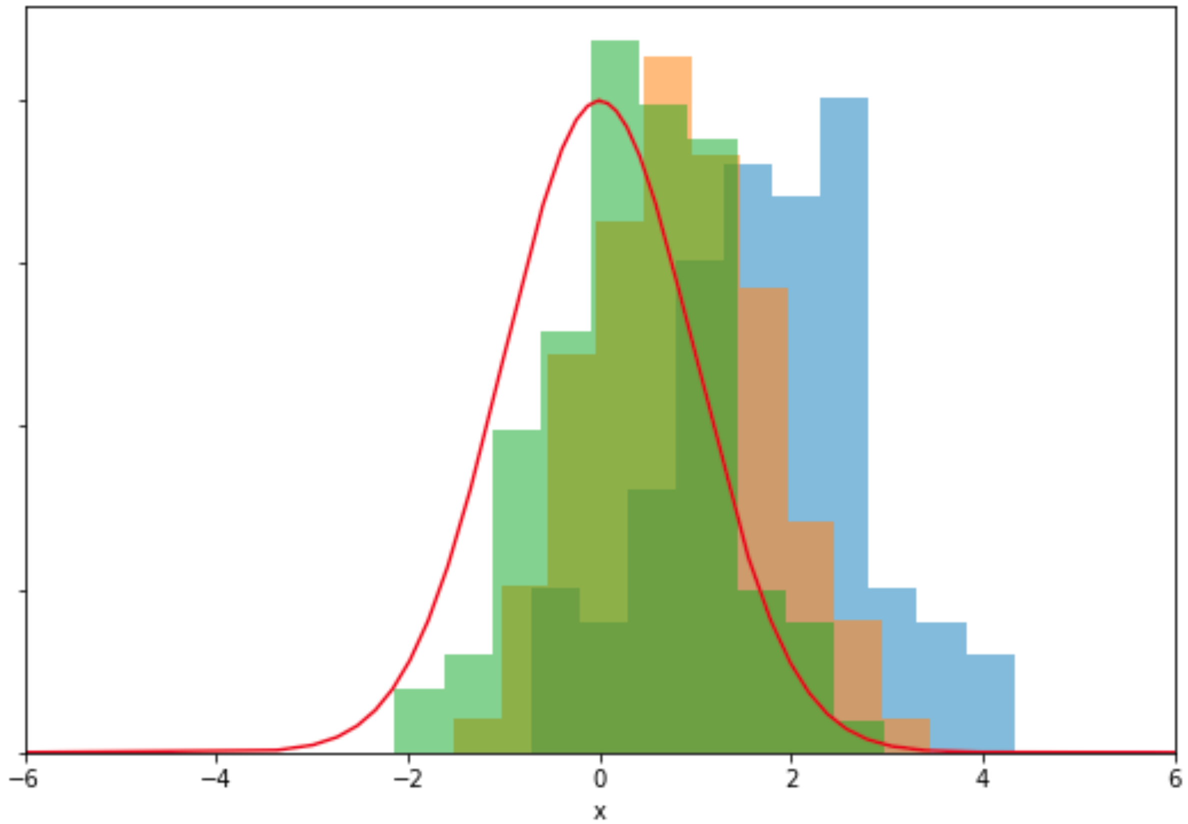


Normal Distribution

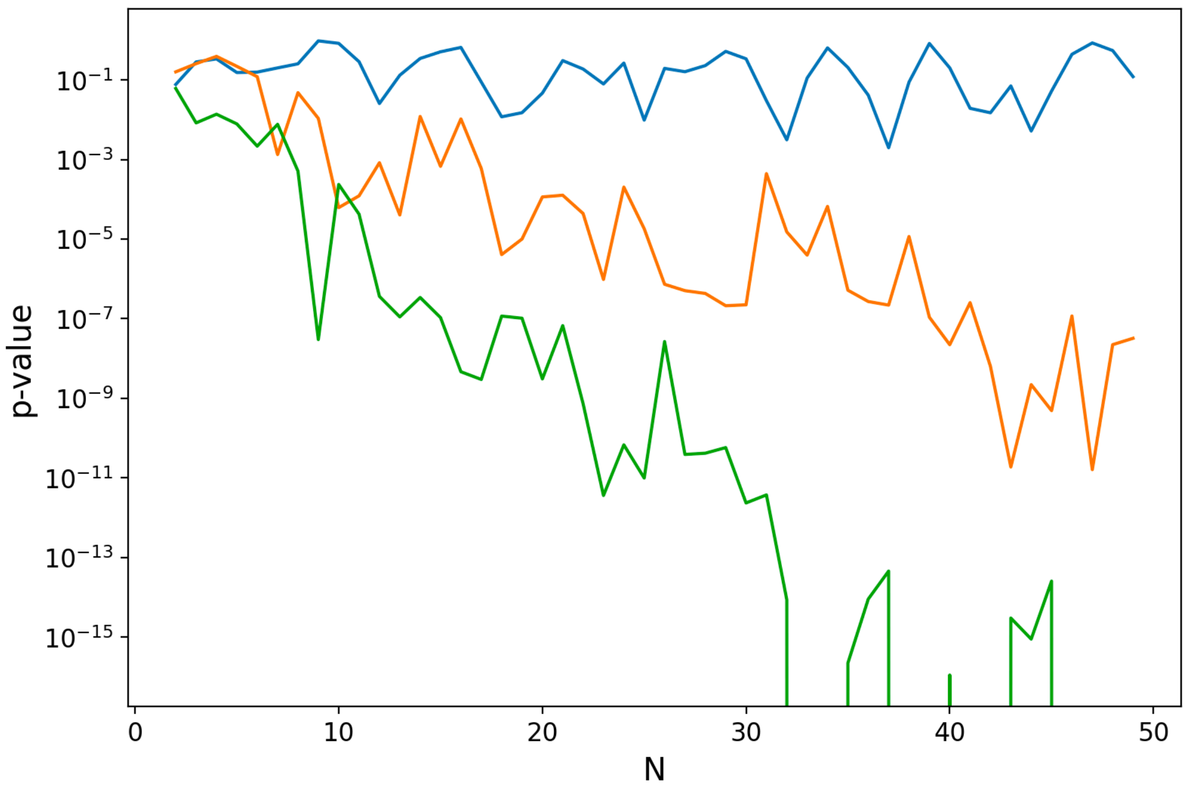
t-test Consistency

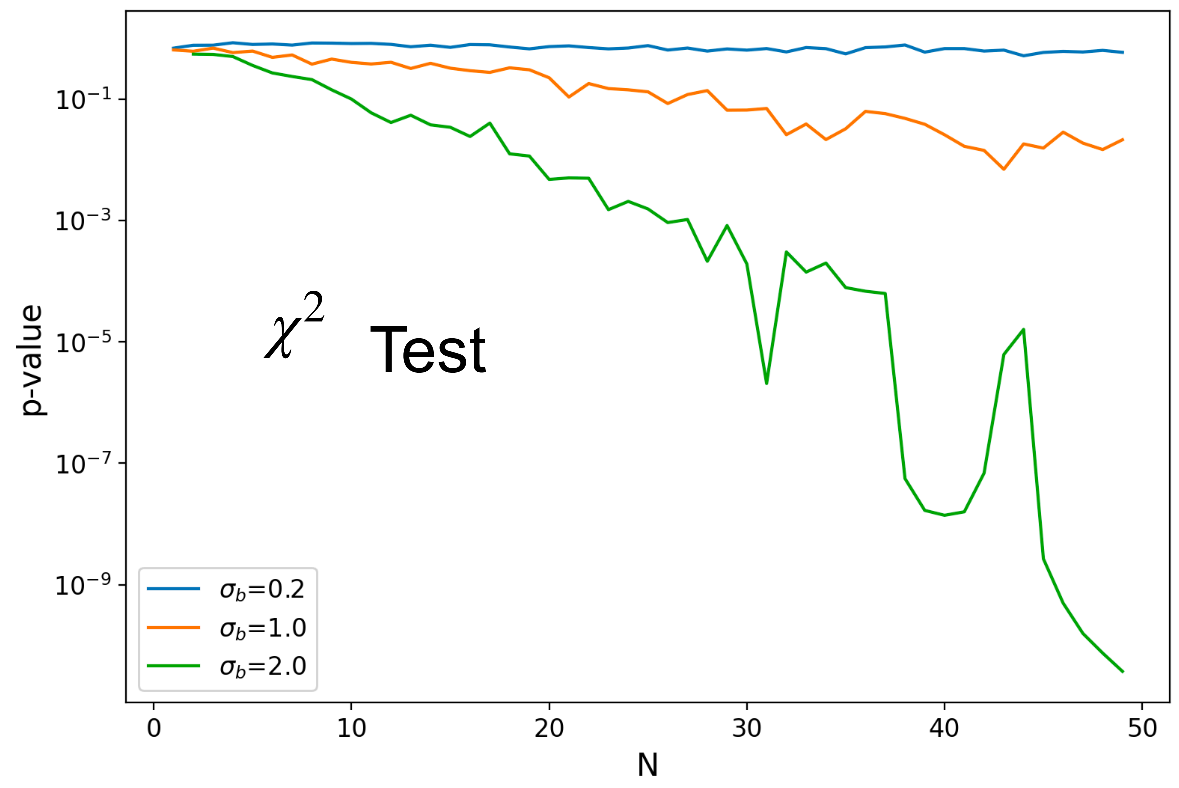
- Variation as a number of events





t-test Consistency



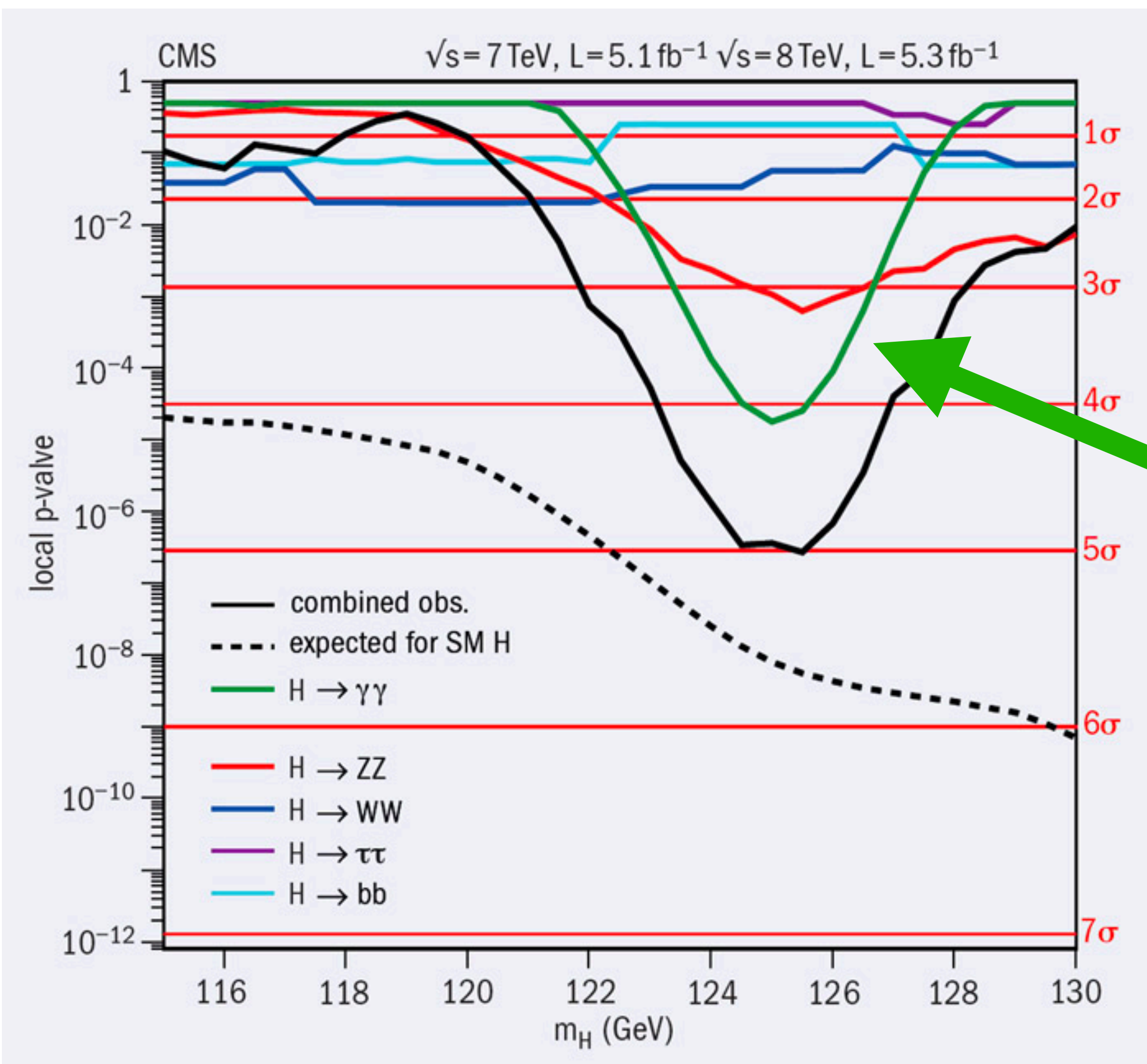


Higgs Boson

- We discovered the Higgs boson not so long ago

In this lectuere we are going to make the discovery plot

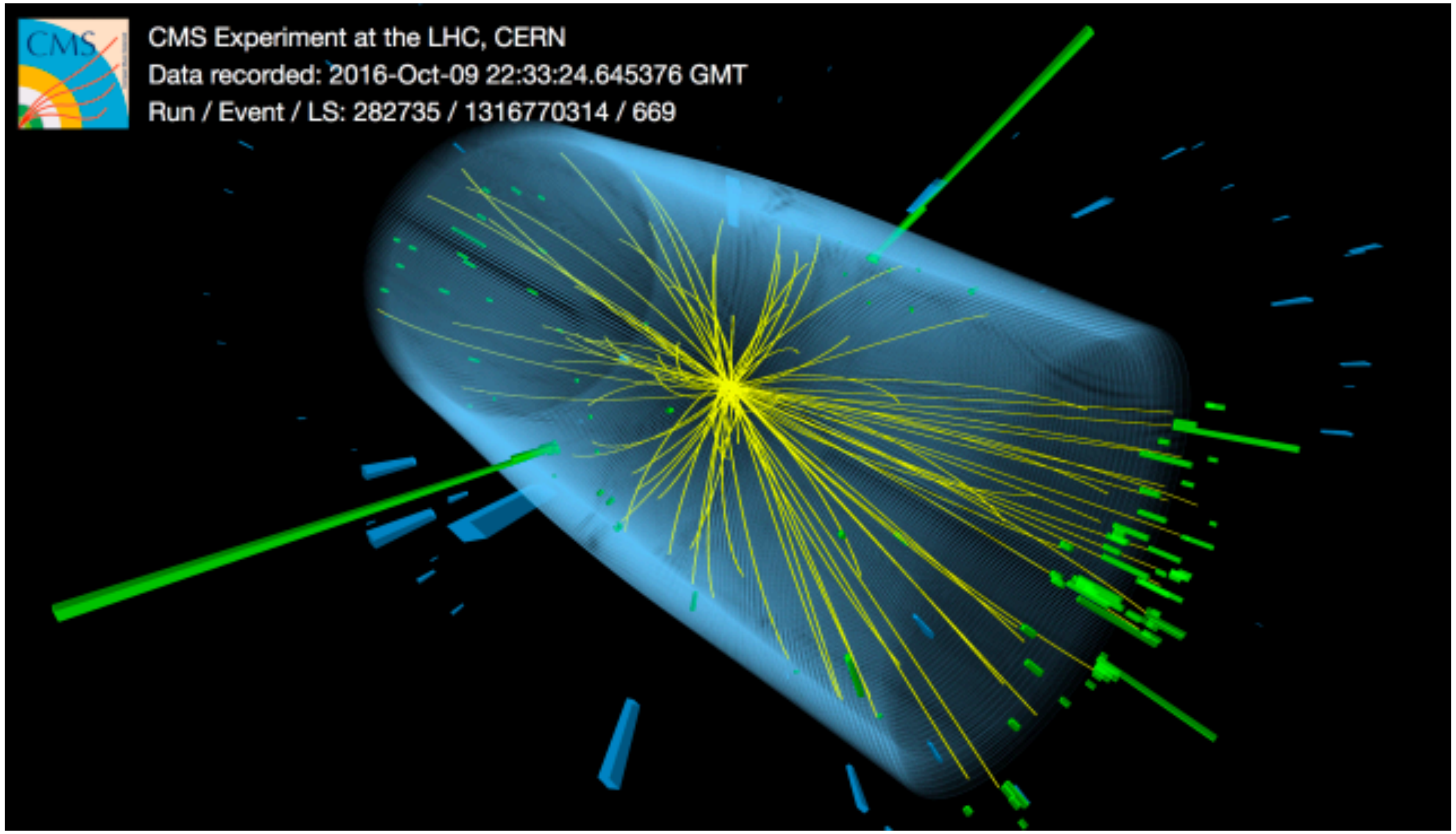
- To discovery it we made a plot like this:

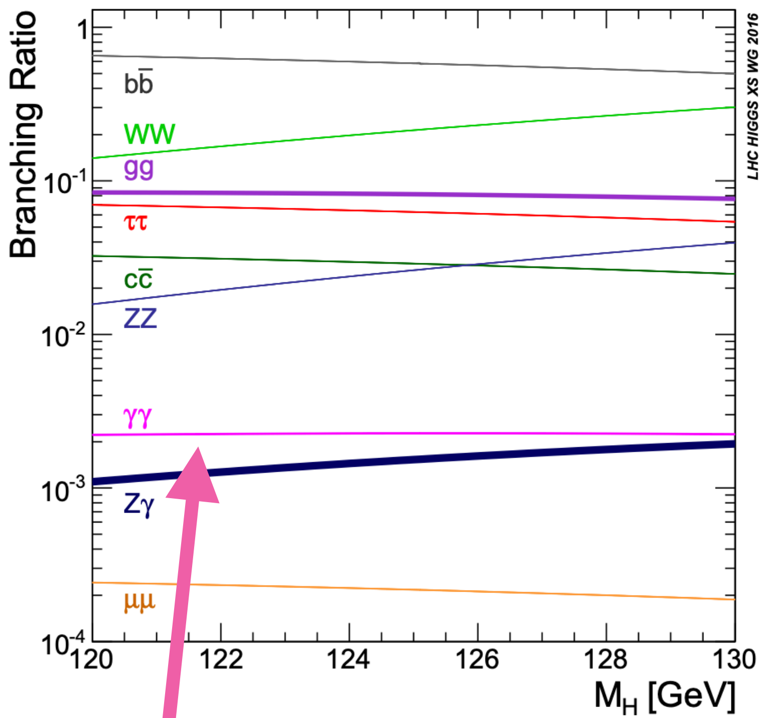


We are going to focus on his object

Higgs to two photons

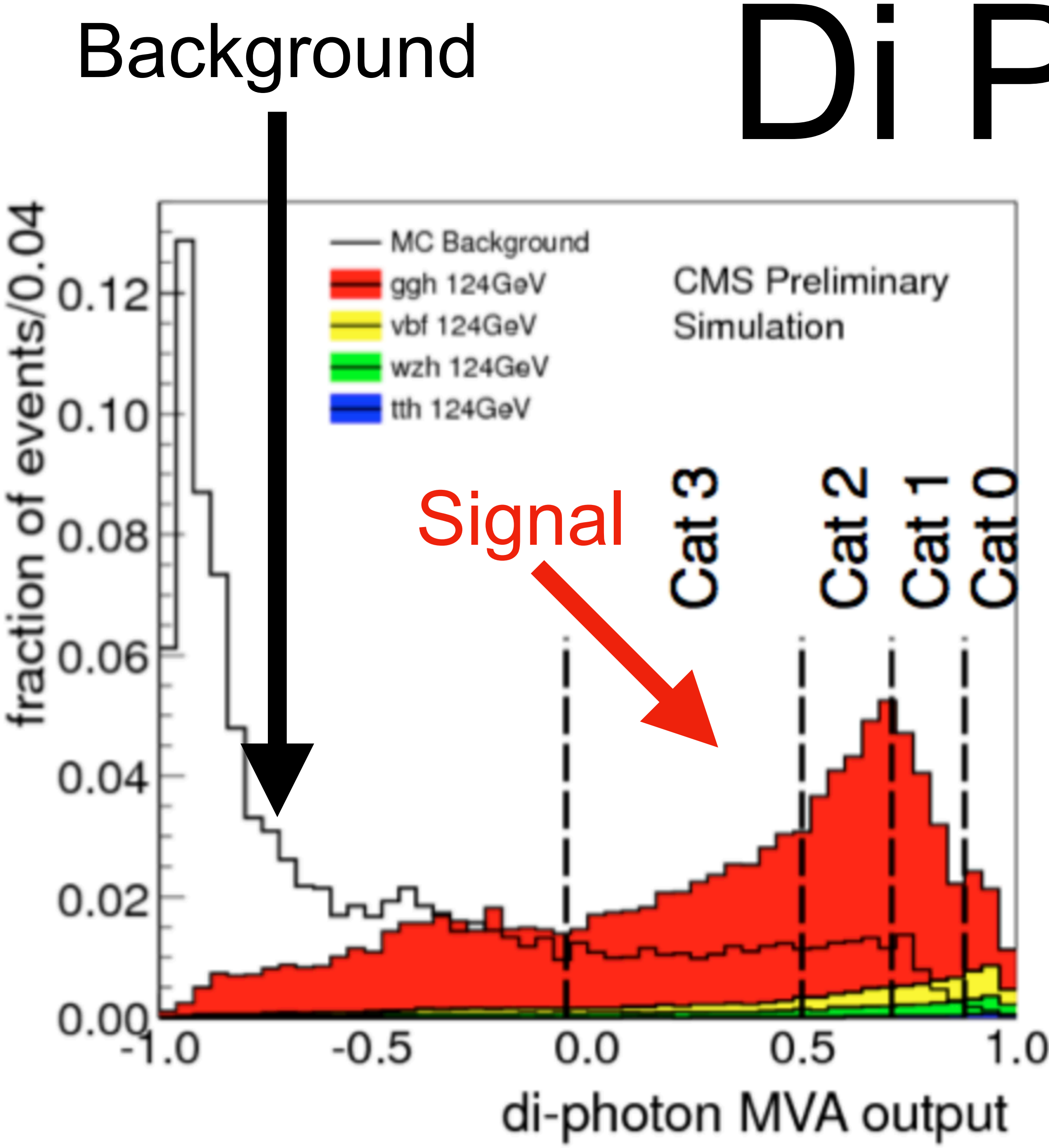
- Higgs to boson decays to nearly every other particle

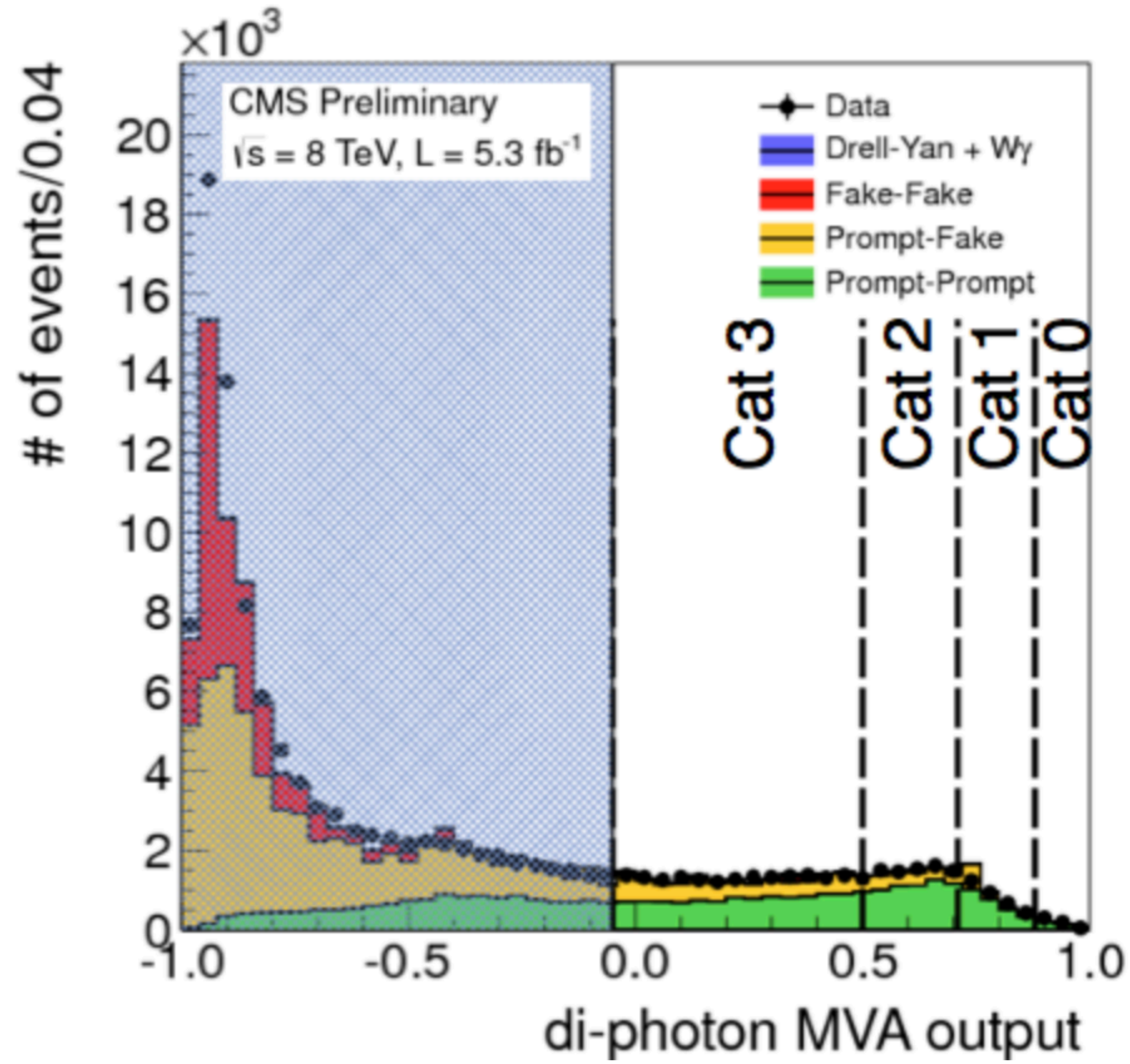




Higgs bosons to two photons occur about 0.3% of the time

Di Photon selection

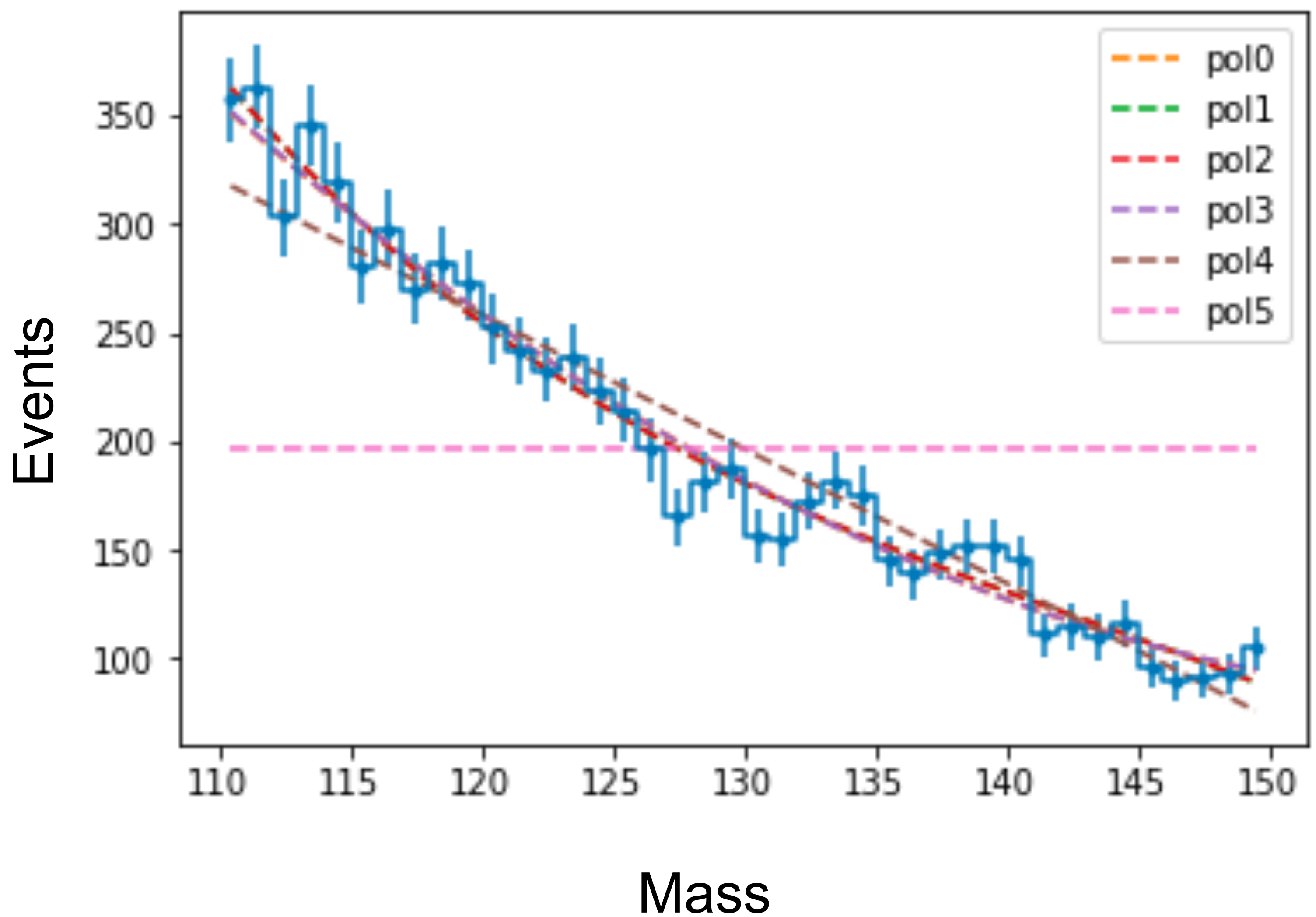




- To select di-photons we build a neural network
  - This takes into account many different Higgs properties

Building a Model

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



Often when we fit, we through a barrage of functions at it
As a rule of thumb we do the “chi-by-eye” (If chi2 is good, we are ok)
A more robust is an f-test (see notes)

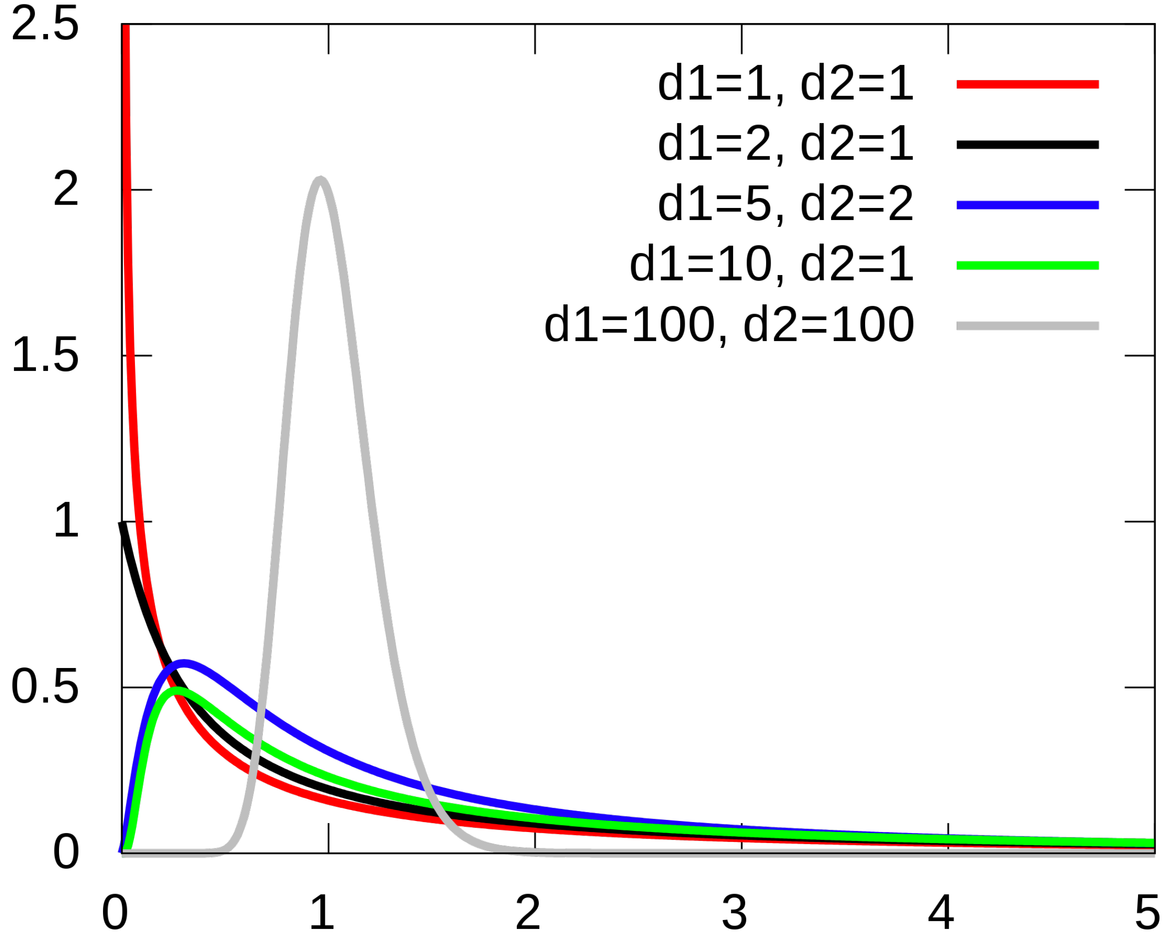
F-Statistic

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

F-Distribution

- Approaches gaussian about 1 for large N

$$f(x; d_1, d_2) = \frac{\sqrt{\frac{(d_1 x)^{d_1}\, d_2^{d_2}}{(d_1 x + d_2)^{d_1+d_2}}}}{x\,\mathrm{B}\!\left(\frac{d_1}{2}, \frac{d_2}{2}\right)}$$
$$= \frac{1}{\mathrm{B}\!\left(\frac{d_1}{2}, \frac{d_2}{2}\right)} \left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}} x^{\frac{d_1}{2}-1} \left(1 + \frac{d_1}{d_2}\,x\right)^{-\frac{d_1+d_2}{2}}$$



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

Now do Frequentist F-test

- n-groups of **fits** each with separate fitted likelihoods

Likelihood of a fit PolX

- Define : $\boxed{\bar{X}_j = \frac{1}{m}\sum_{i=1}^{m} X_i}$ and $\mathrm{S}_\mathrm{j}^2 = \frac{1}{\mathrm{m}-1}\sum_{\mathrm{i}=1}^{\mathrm{m}} (\mathrm{X}_\mathrm{i} - \bar{\mathrm{X}}_\mathrm{j})^2$

- Define: $\boxed{MS_B = \frac{m}{n-1}\sum_{i=1}^{n} (\bar{X}_j - \bar{X})^2}$ and $\mathrm{MS}_\mathrm{R} = \sigma^2$

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

Higher Order Polynomial

- $$\frac{\frac{RSS_1 - RSS_2}{p_2 - p_1}}{\frac{RSS_2}{n - p_2}} \approx F_{p_2 - p_1 \; n - p_2} \text{ generically}$$
- $$\frac{\frac{\mathscr{L}_1 - \mathscr{L}_2}{p_2 - p_1}}{\frac{\mathscr{L}_2}{n - p_2}} \approx F_{p_2 - p_1 \; n - p_2}$$

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

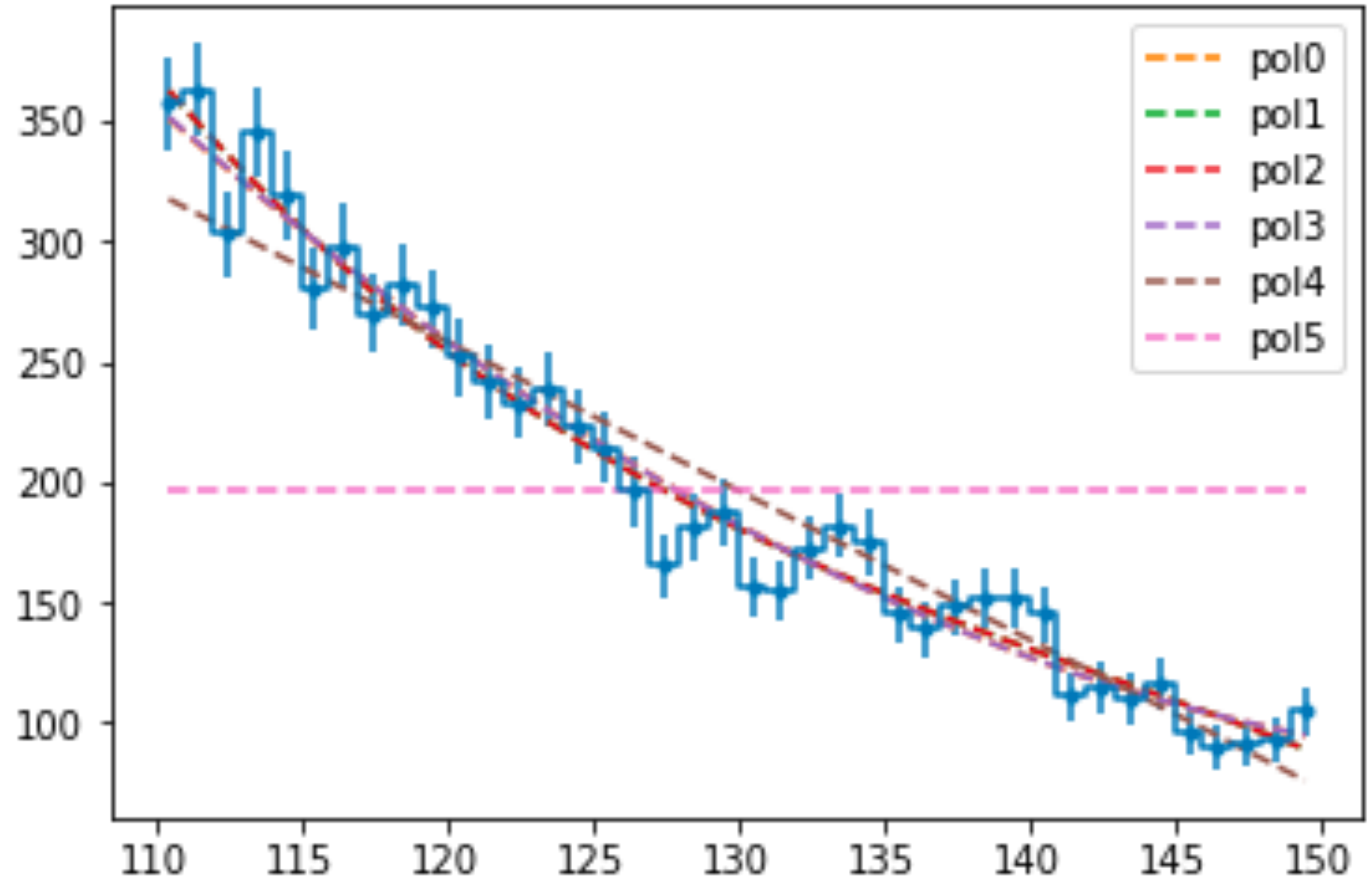
Higher Order Polynomial

- $$\frac{\frac{RSS_1 - RSS_2}{p_2 - p_1}}{\frac{RSS_2}{n - p_2}} \approx F_{p_2-p_1\ n-p_2}$$ generically

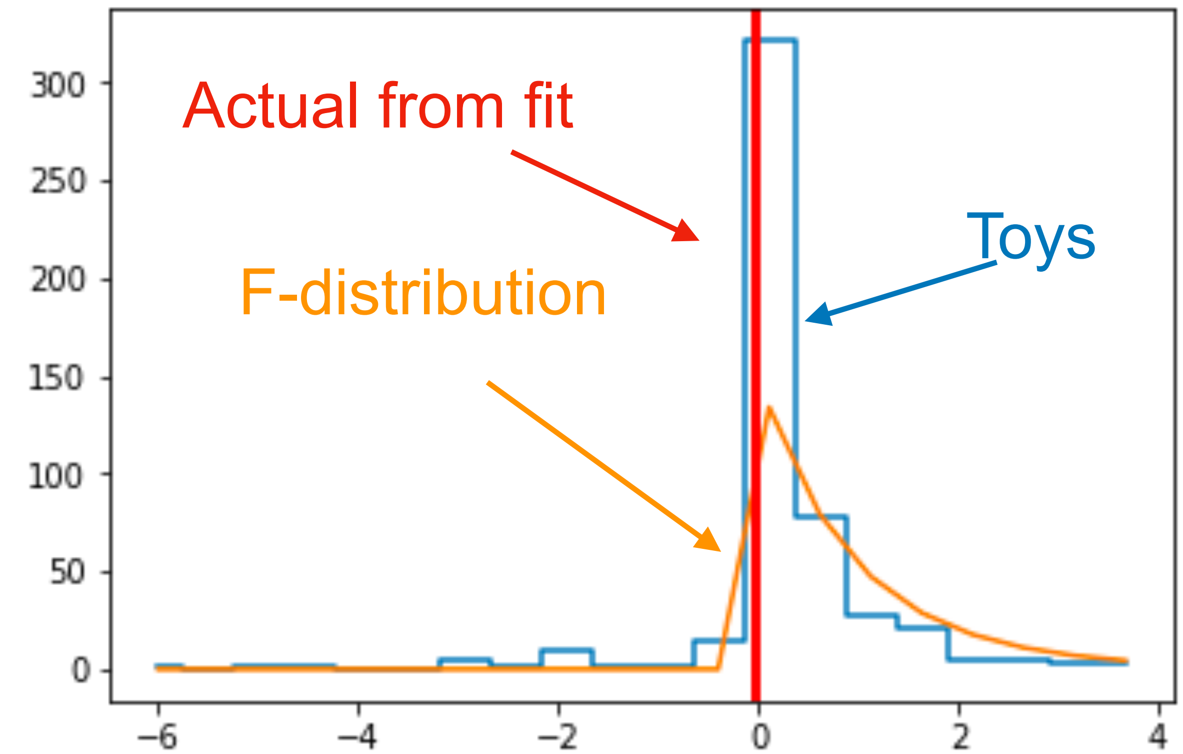
- $$\frac{\frac{\mathscr{L}_1 - \mathscr{L}_2}{p_2 - p_1}}{\frac{\mathscr{L}_2}{n - p_2}} \approx F_{p_2-p_1\ n-p_2}$$

F-test Example

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



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

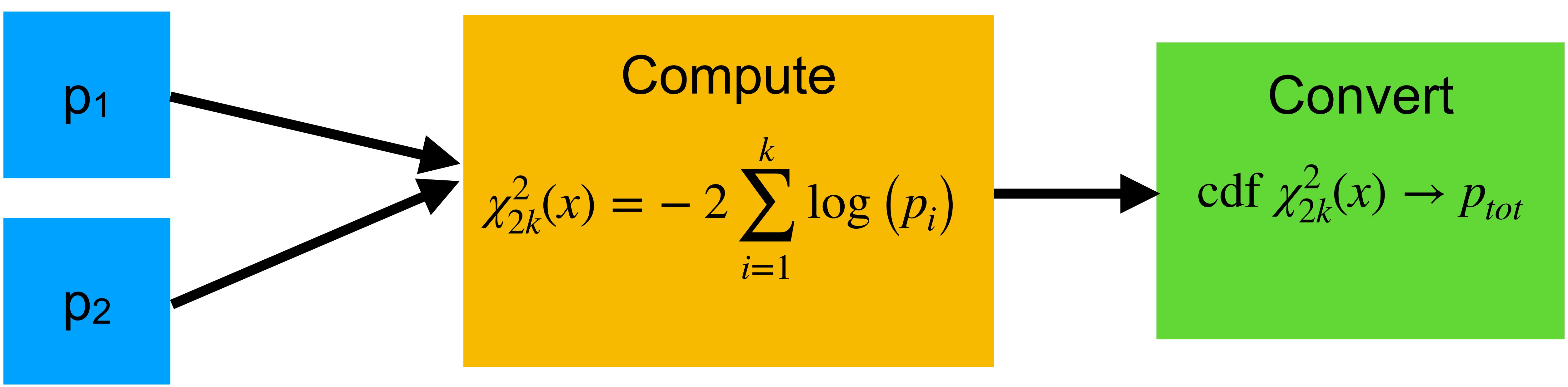


$$\frac{\frac{\mathcal{L}_1 - \mathcal{L}_2}{p_2 - p_1}}{\frac{\mathcal{L}_2}{n - p_2}}$$

For a 4th order to a 3rd order

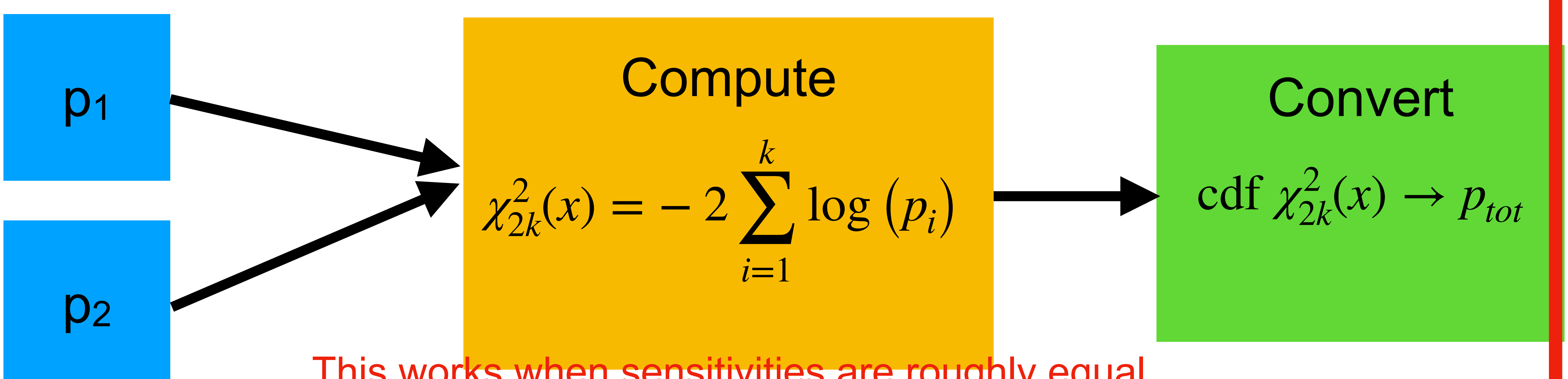
Combining Categories

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



Combining Categories

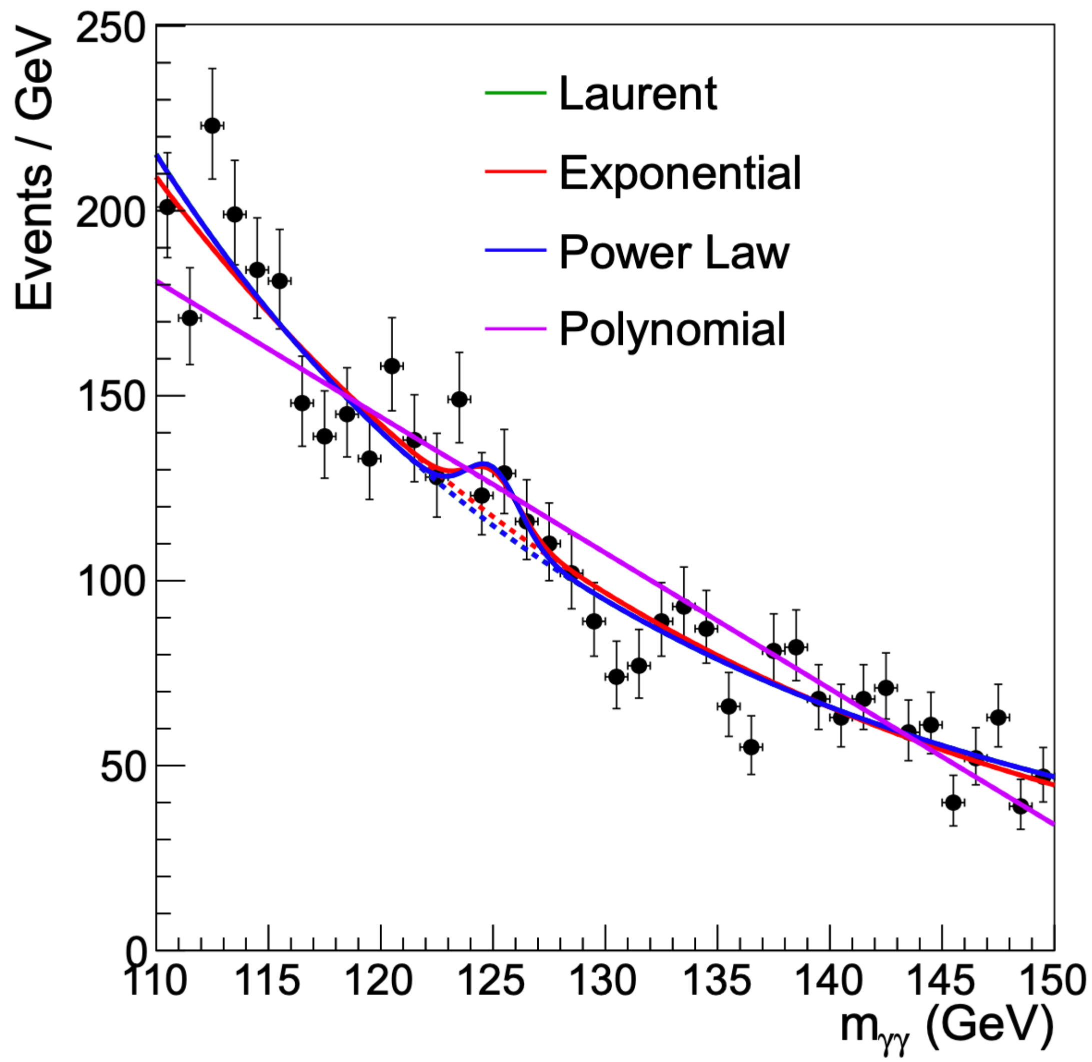
- Lets say we have k measurements with probability p_i
  - Each measurement is independent of each other
- We can combine categories using the following formula
  - $\chi^2_{2k}(x) = -2\sum_{i=1}^{k}\log\left(p_i\right)$
  - Where left is a chi2 distribution with 2k degrees of freedom

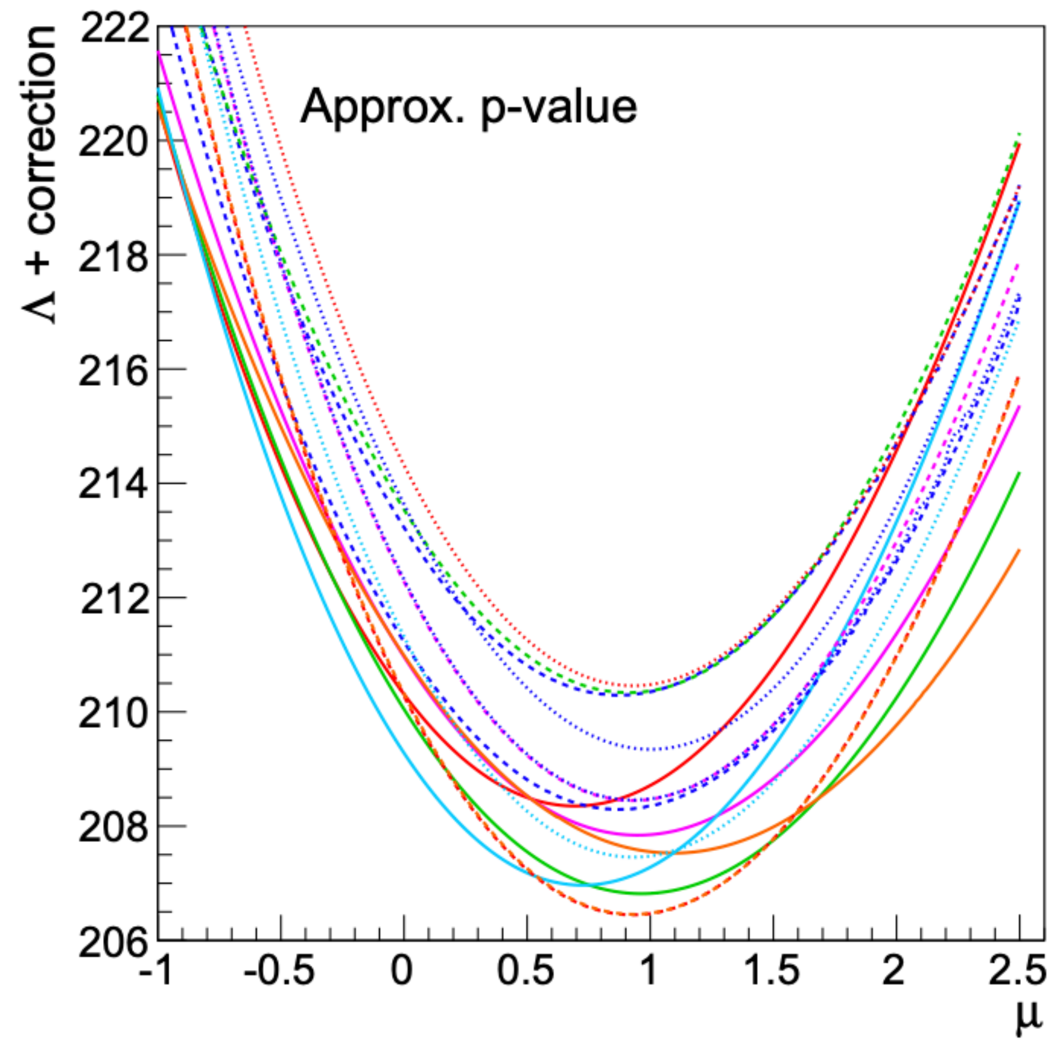


This works when sensitivities are roughly equal
More robust way is to fit all categories simulatneously

Building a Model

- Throwing a barrage of functions at the problem

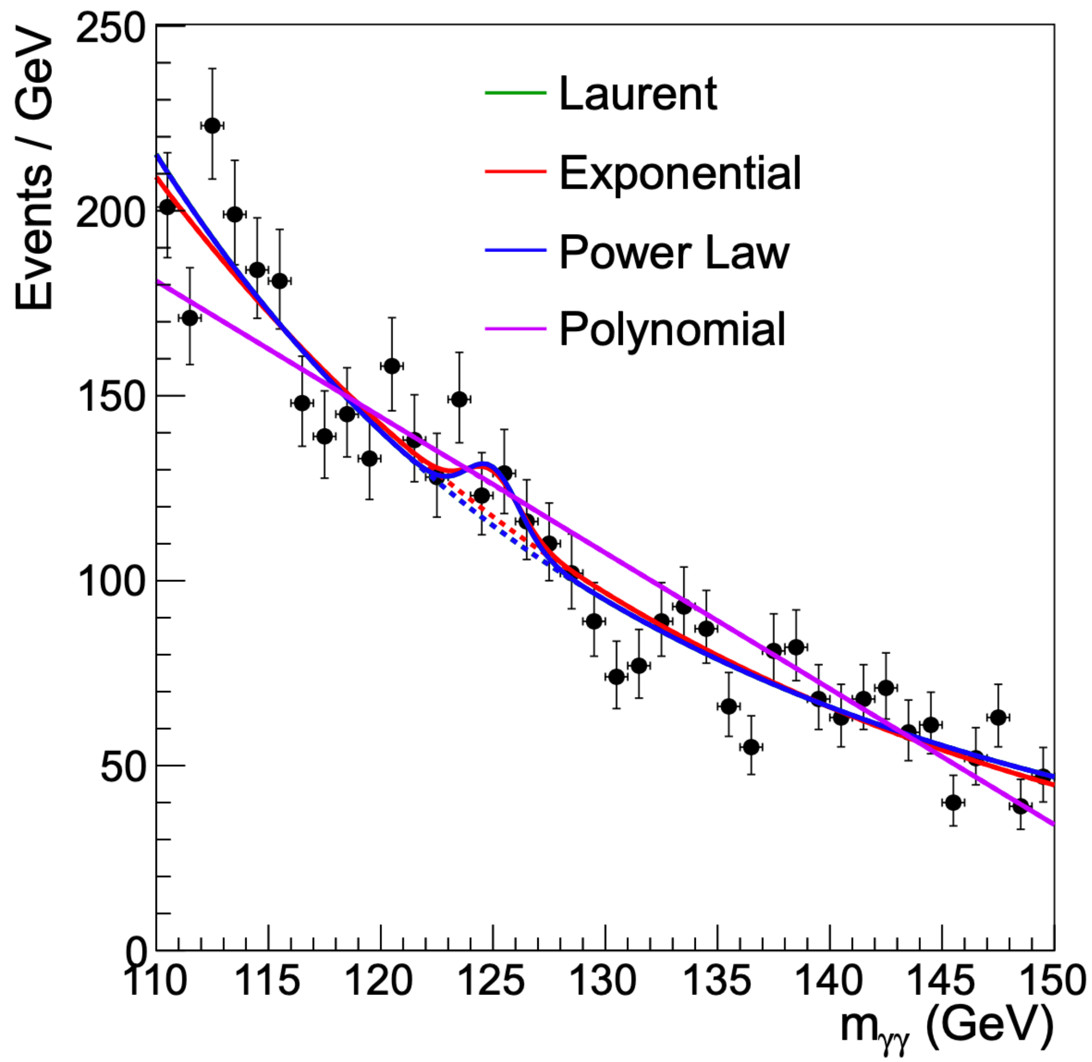


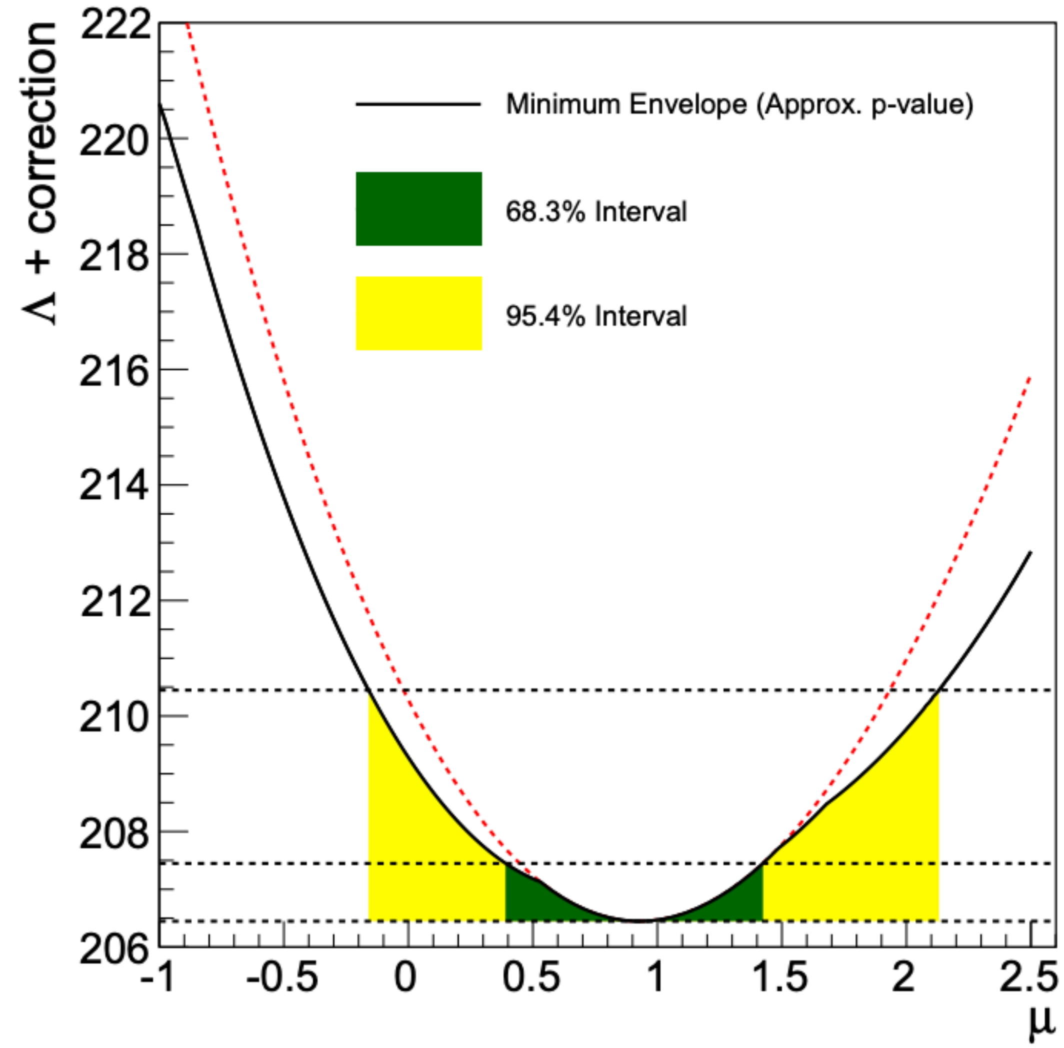


We can try a whole library of functions
The likelihood we get translates to our fit

Building a Model

- Throwing a barrage of functions at the problem

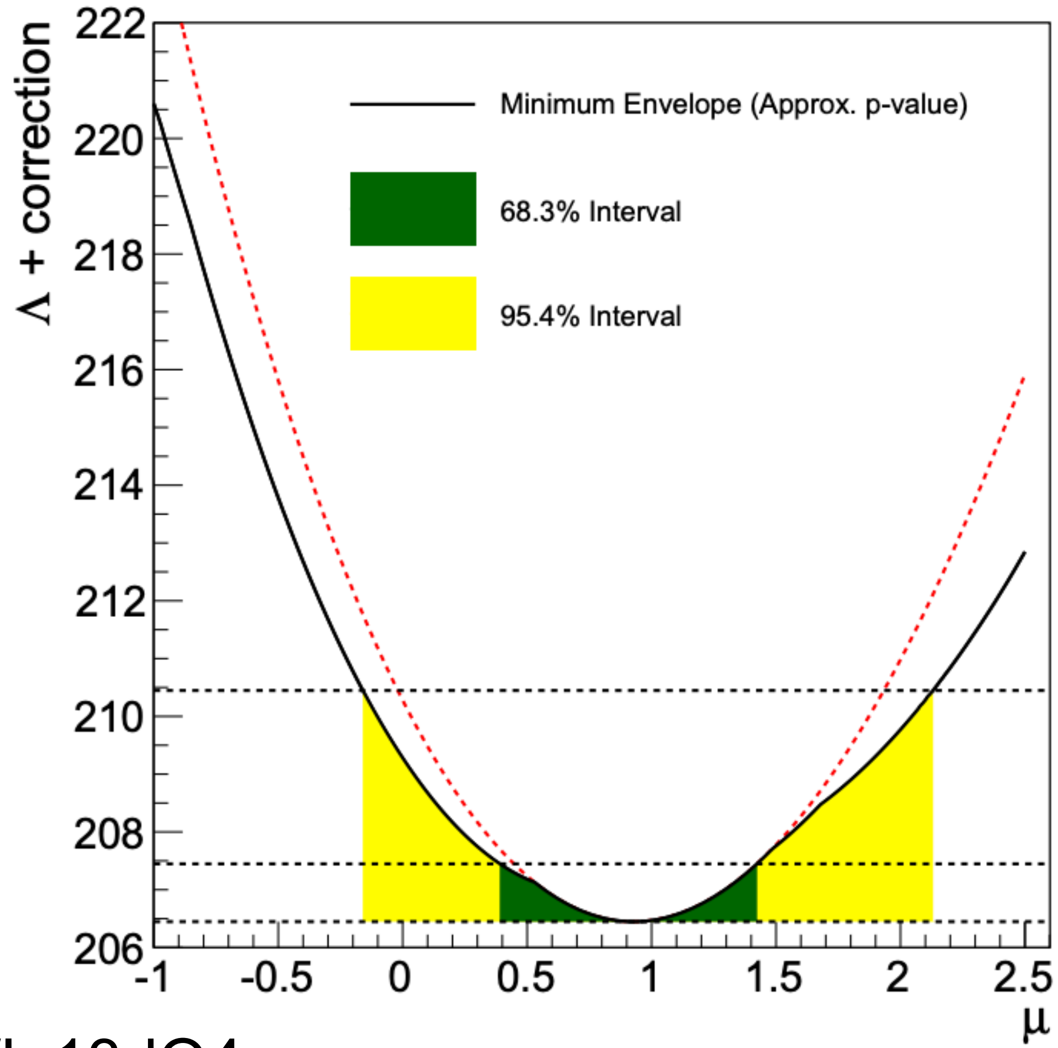




We can try a whole library of functions
The likelihood we get translates to our fit

Combined Likelihood

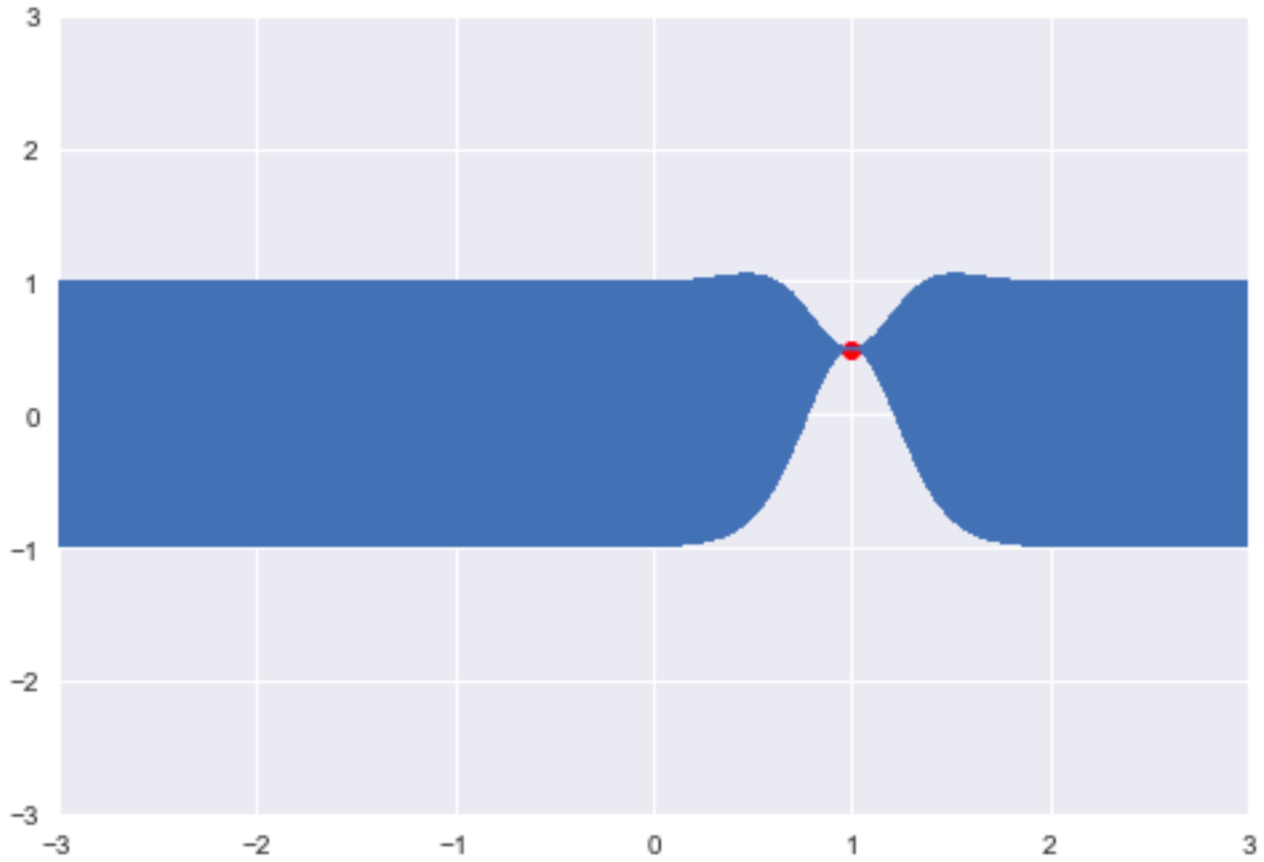
- Throwing a barrage of functions at the problem



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

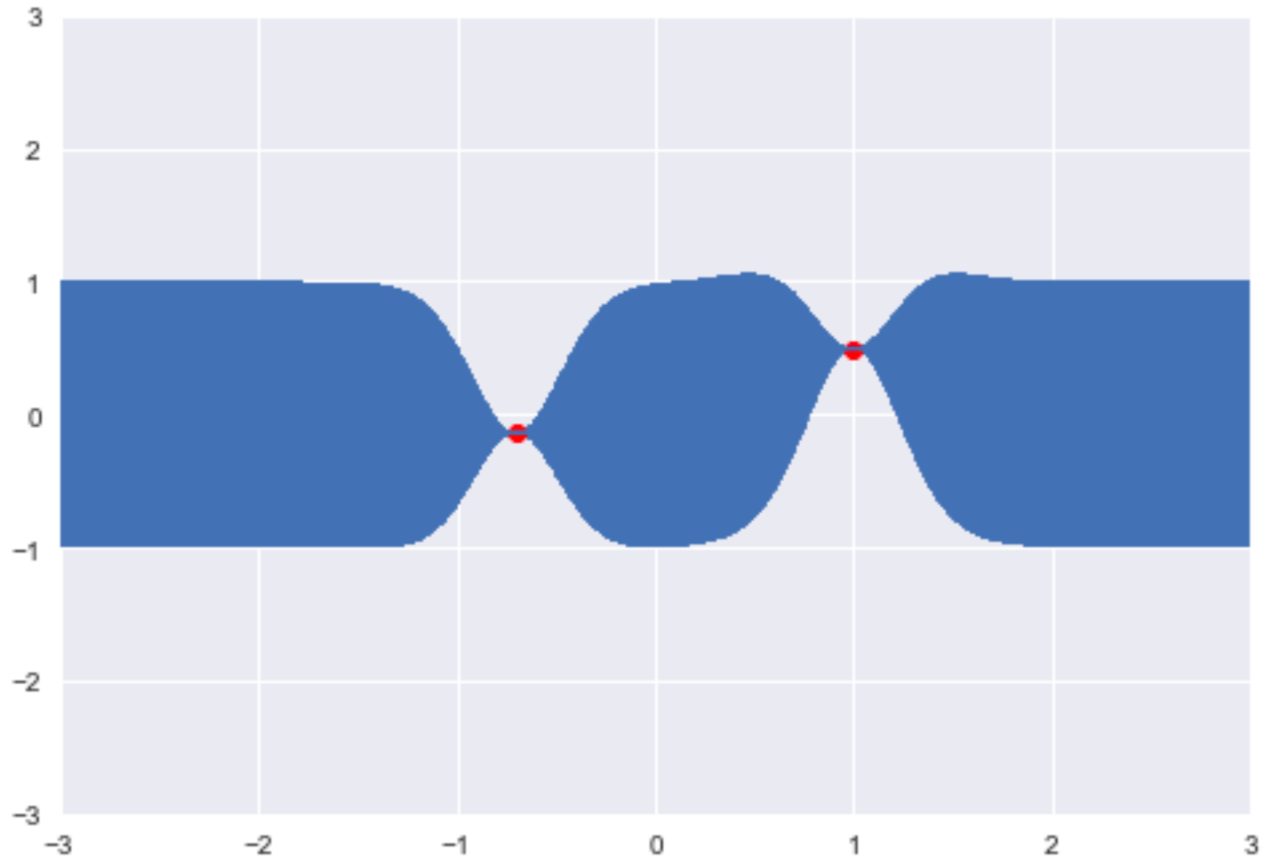
We can try a whole library of functions
The likelihood we get translates to our fit

Splines+GaussianProcesses



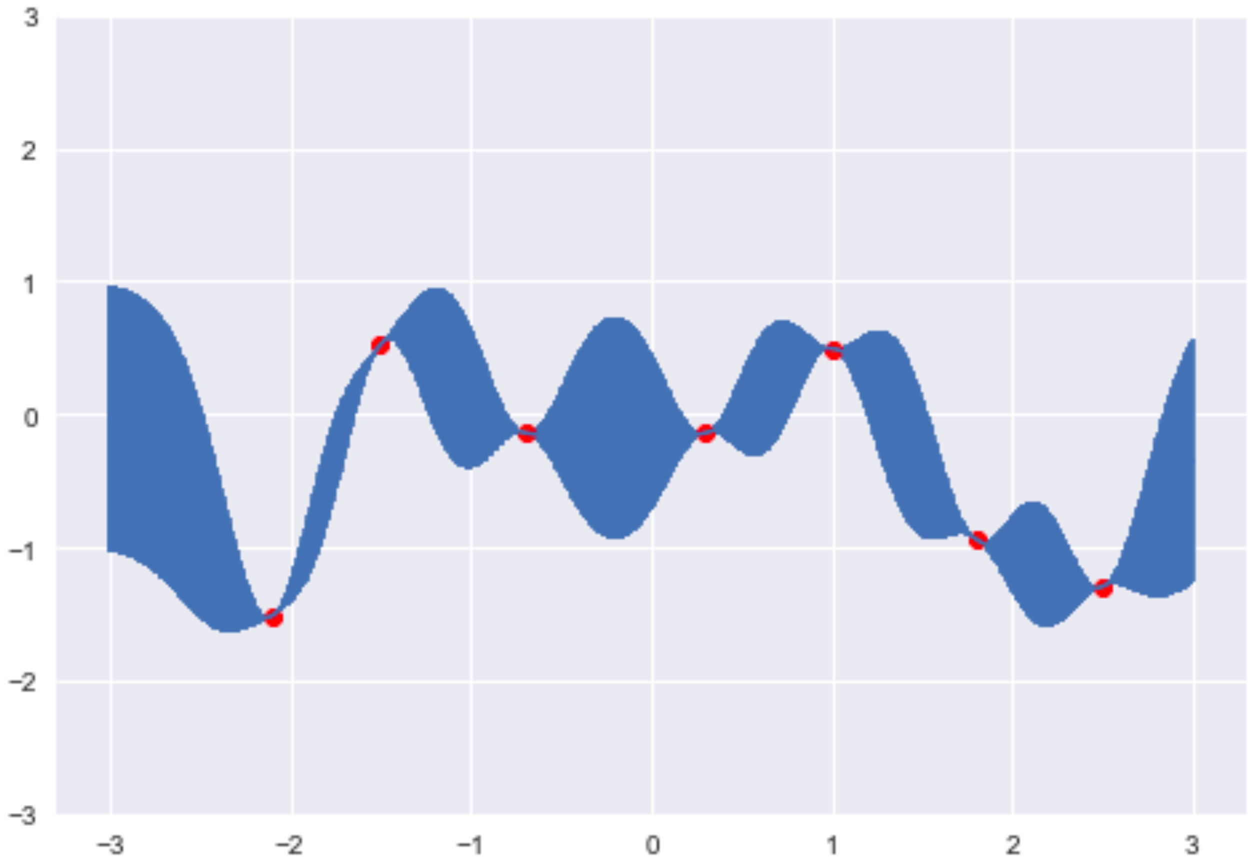
- Gaussian processes allow us to build function choice from the data

Splines+GaussianProcesses



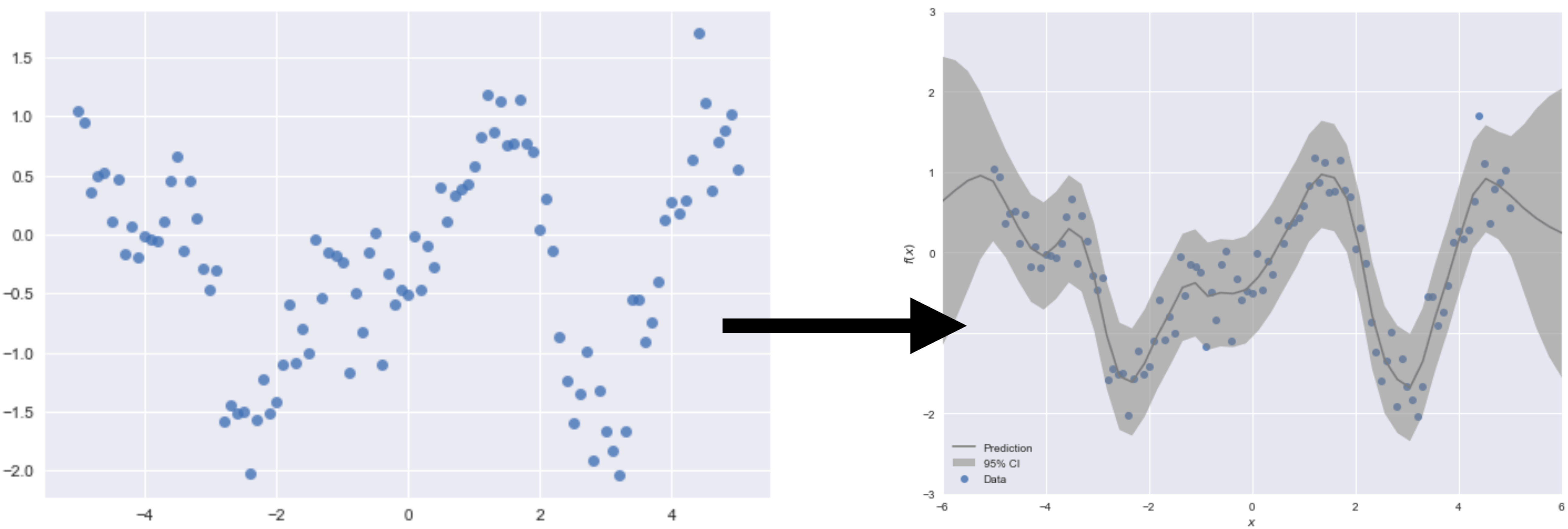
- Gaussian processes allow us to build function choice from the data

Splines+GaussianProcesses



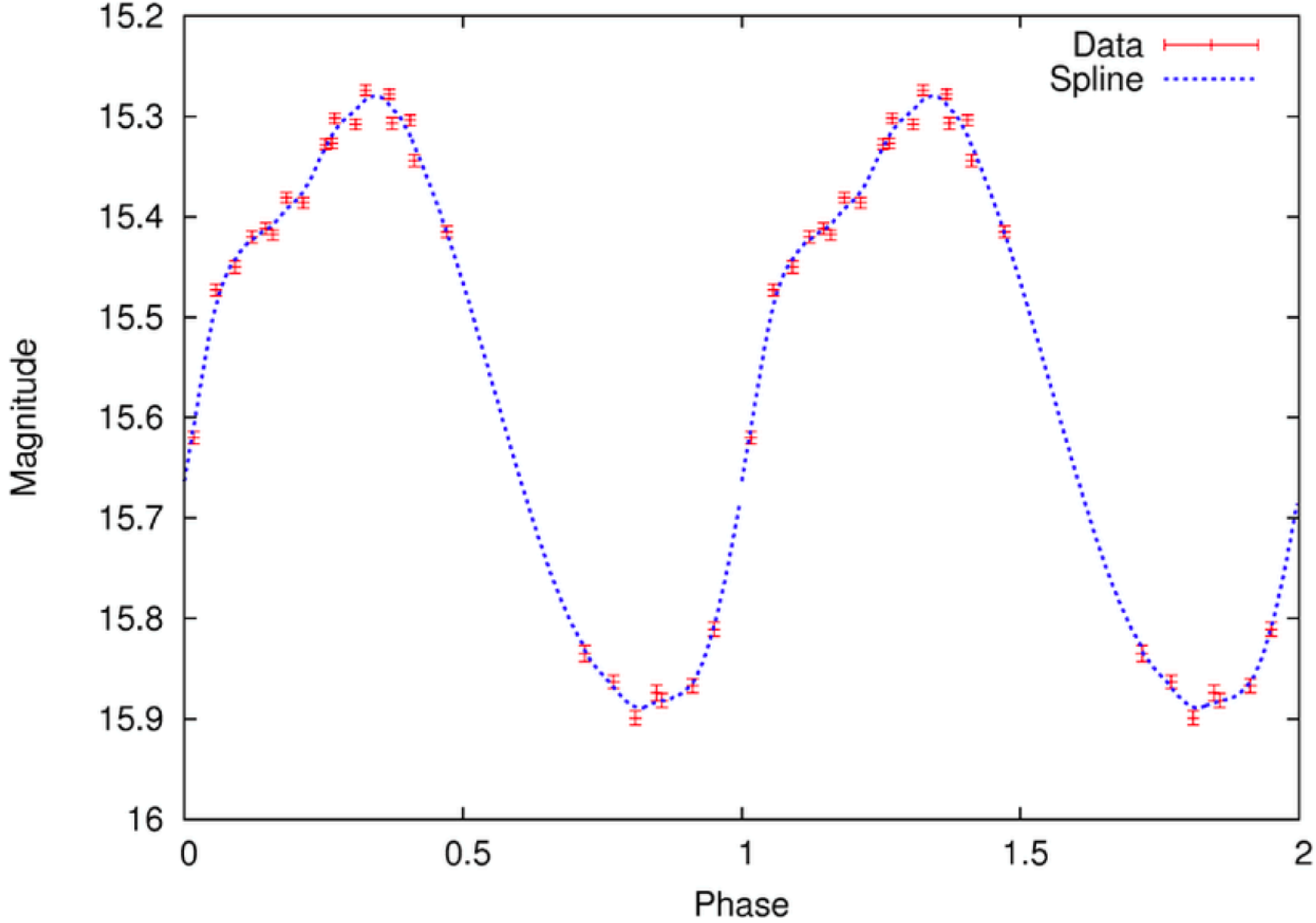
- Gaussian processes allow us to build function choice from the data

Splines+GaussianProcesses



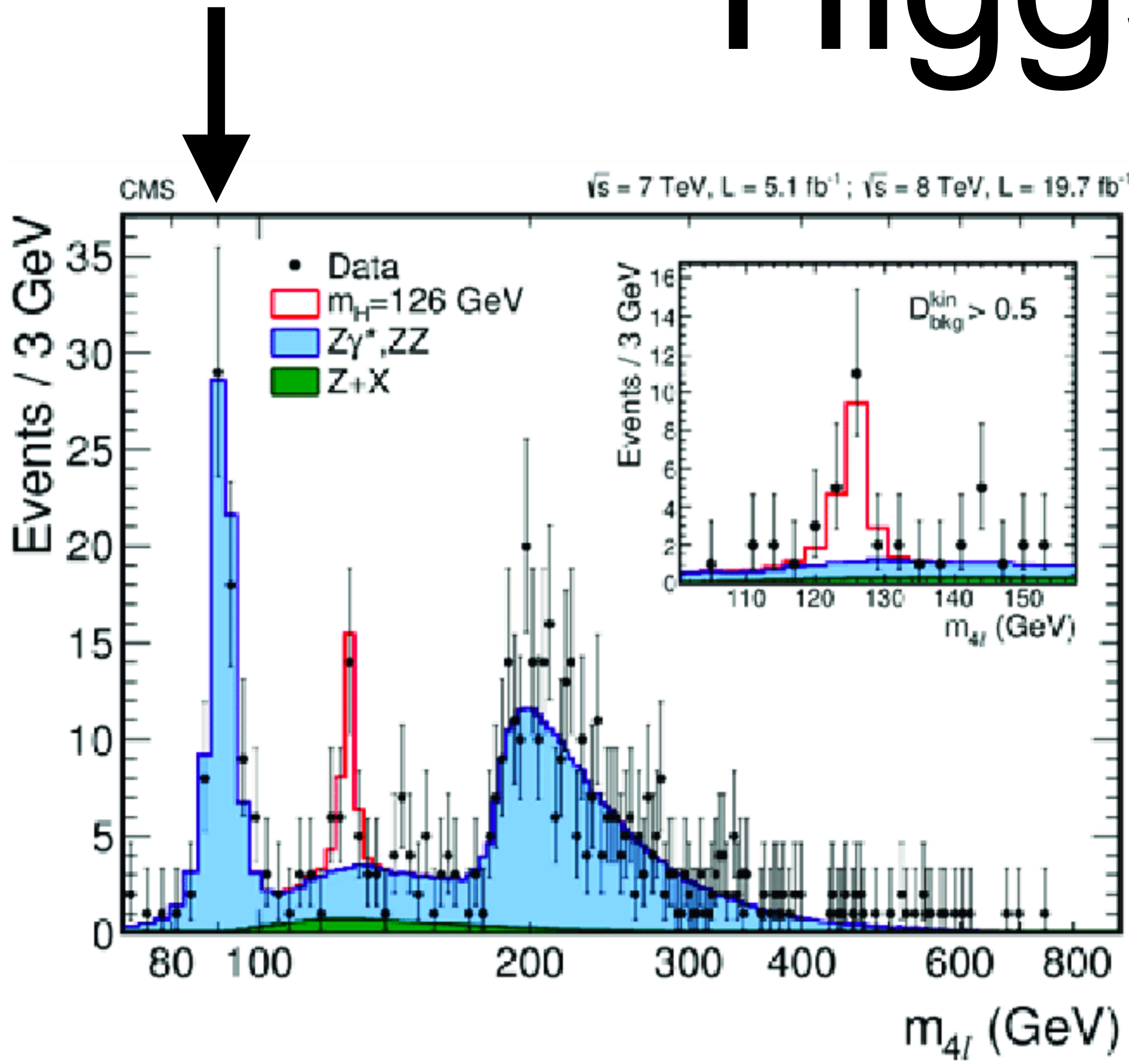
- To go from points to spline automatically

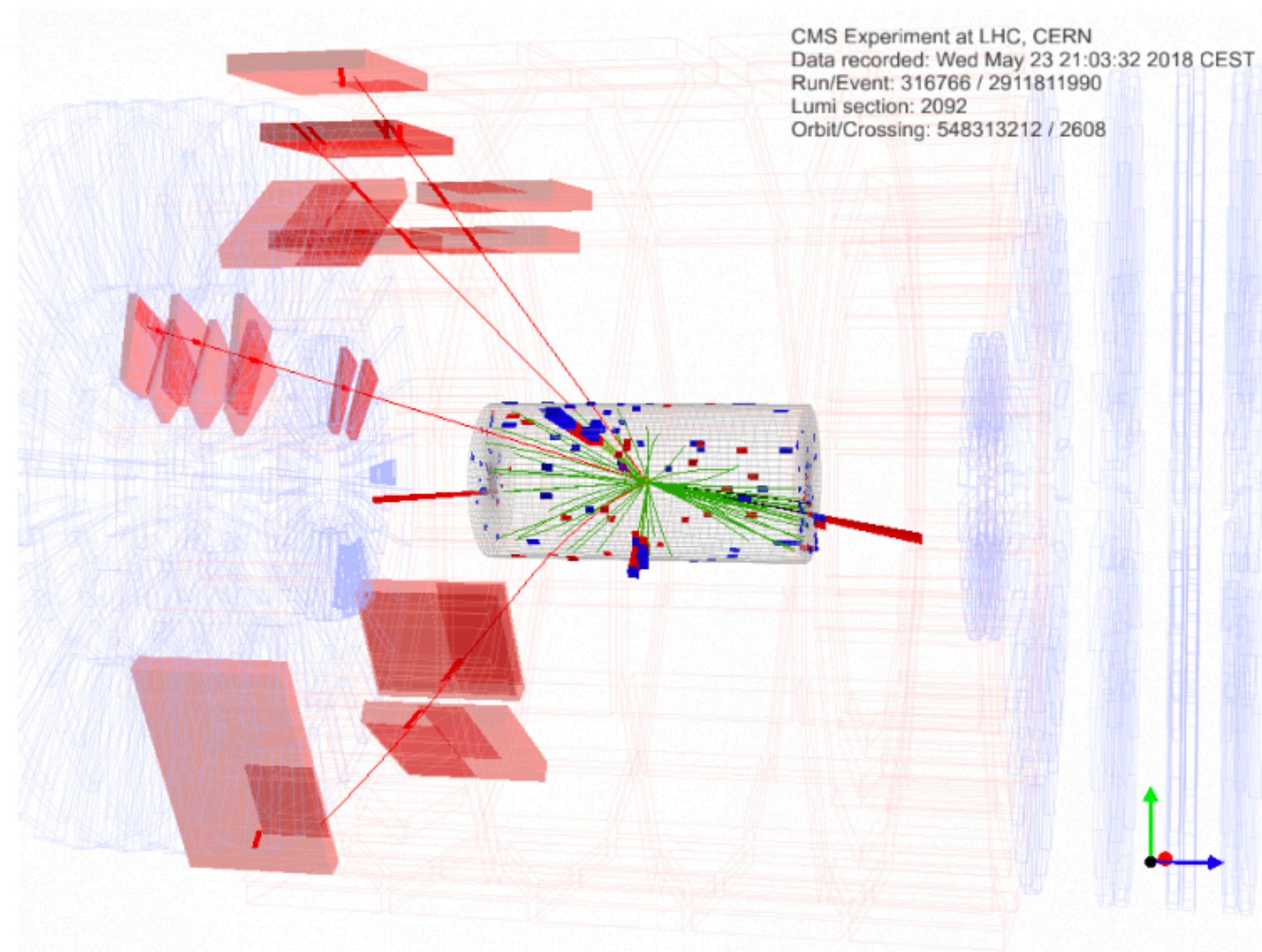
Splines+GaussianProcesses



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

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

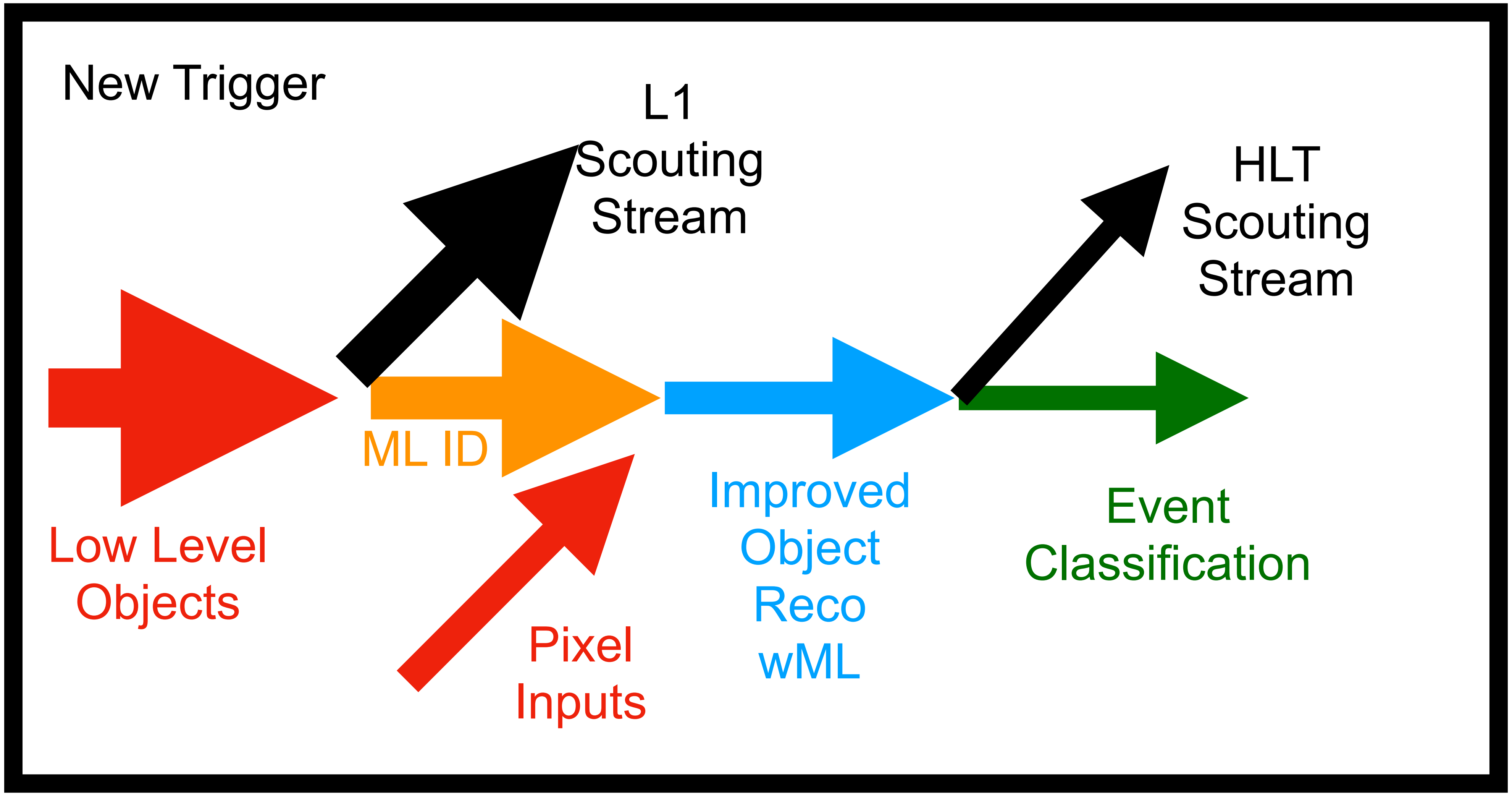
- We can evaluate this through an F-test

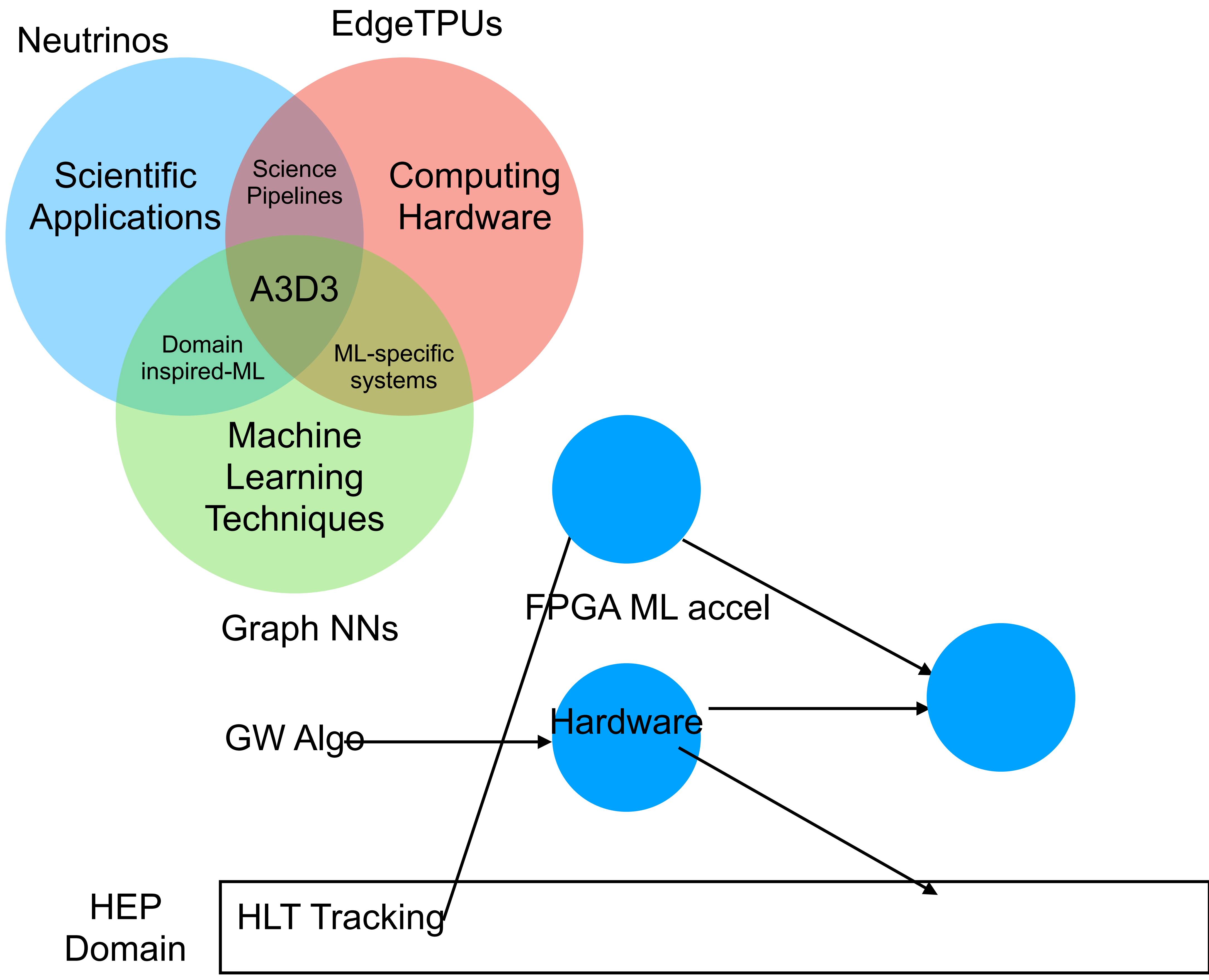
- Recall $\frac{MS_B}{MS_R} \approx 1 = F_{n-1\ m(n-1)}$

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

$$F = \frac{\left(\frac{\mathrm{RSS}_1 - \mathrm{RSS}_2}{p_2 - p_1}\right)}{\left(\frac{\mathrm{RSS}_2}{n - p_2}\right)},$$

Title Text


New Trigger
L1
Scouting
Stream
HLT
Scouting
Stream
ML ID
Low Level
Objects
Pixel
Inputs
Improved
Object
Reco
wML
Event
Classification


Neutrinos
EdgeTPUs
Scientific Applications
Science Pipelines
Computing Hardware
A3D3
Domain inspired-ML
ML-specific systems
Machine Learning Techniques
Graph NNs
FPGA ML accel
GW Algo
Hardware
HEP Domain
HLT Tracking

Elastic Scatter