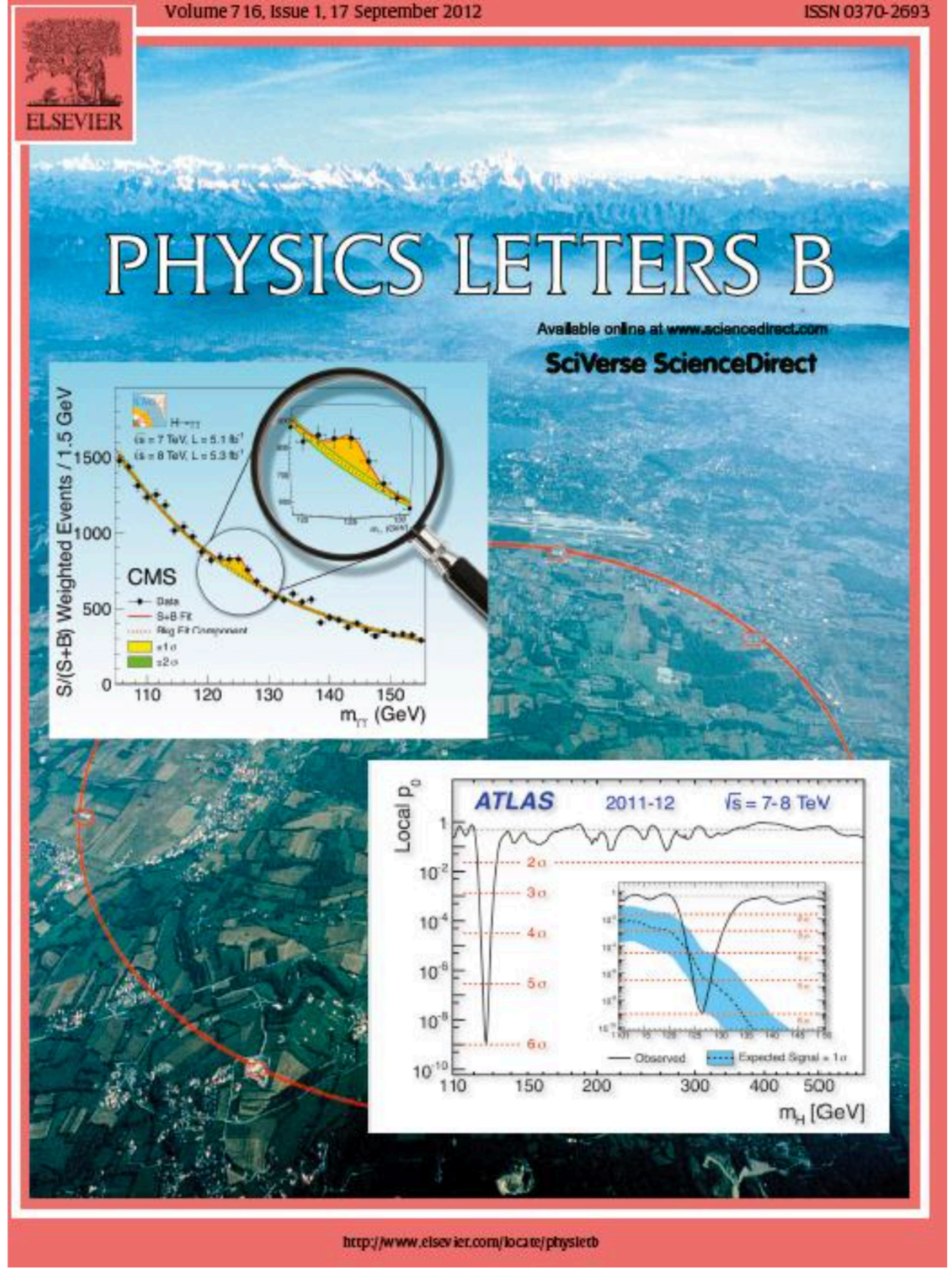
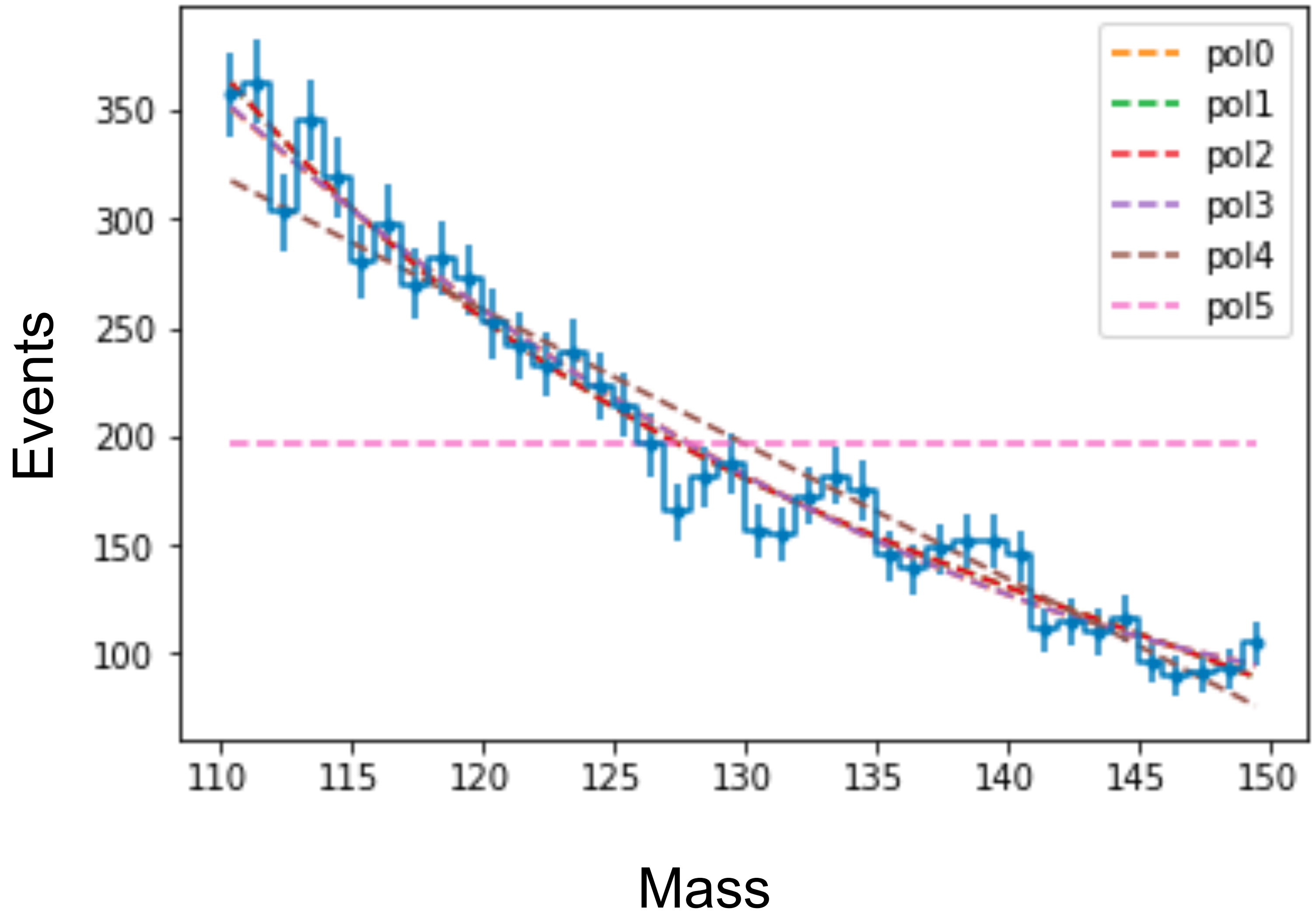


Lecture 12

Higgs Boson Discovery
Towards Deep Learning

Building a Model

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



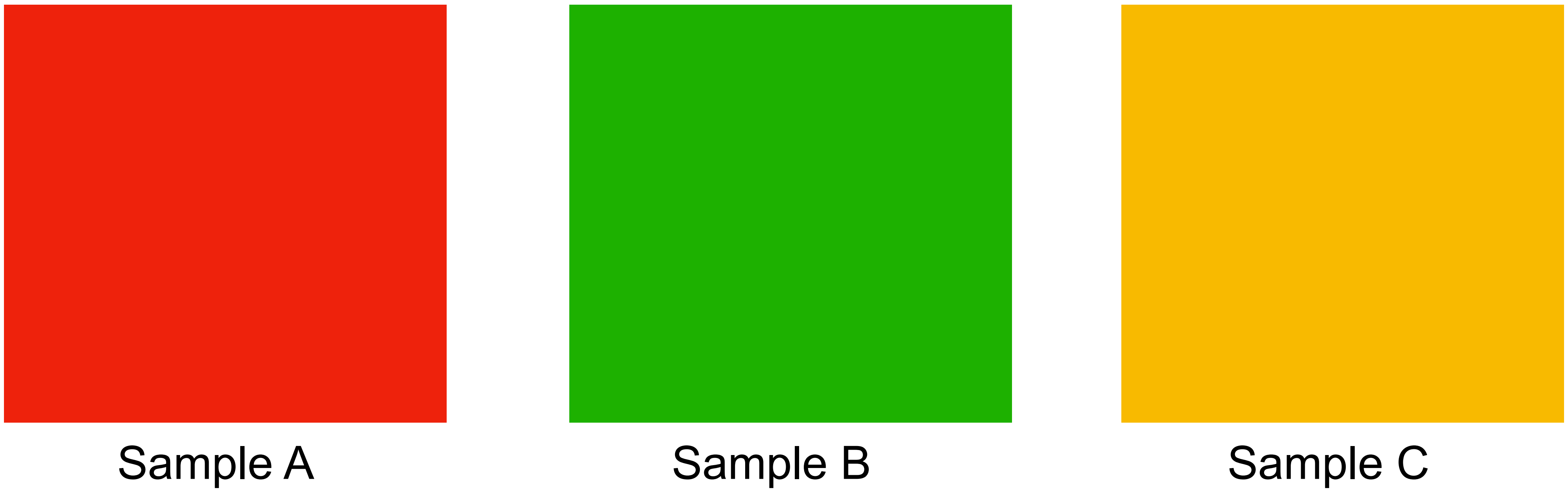
Often when we fit, we through a barrage of functions at it
As a rule of thumb we do the “chi-by-eye” (If chi2 is good, we are ok)
A more robust is an f-test (see notes)

T-test

Sample A Sample B Sample C

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

F-test



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

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

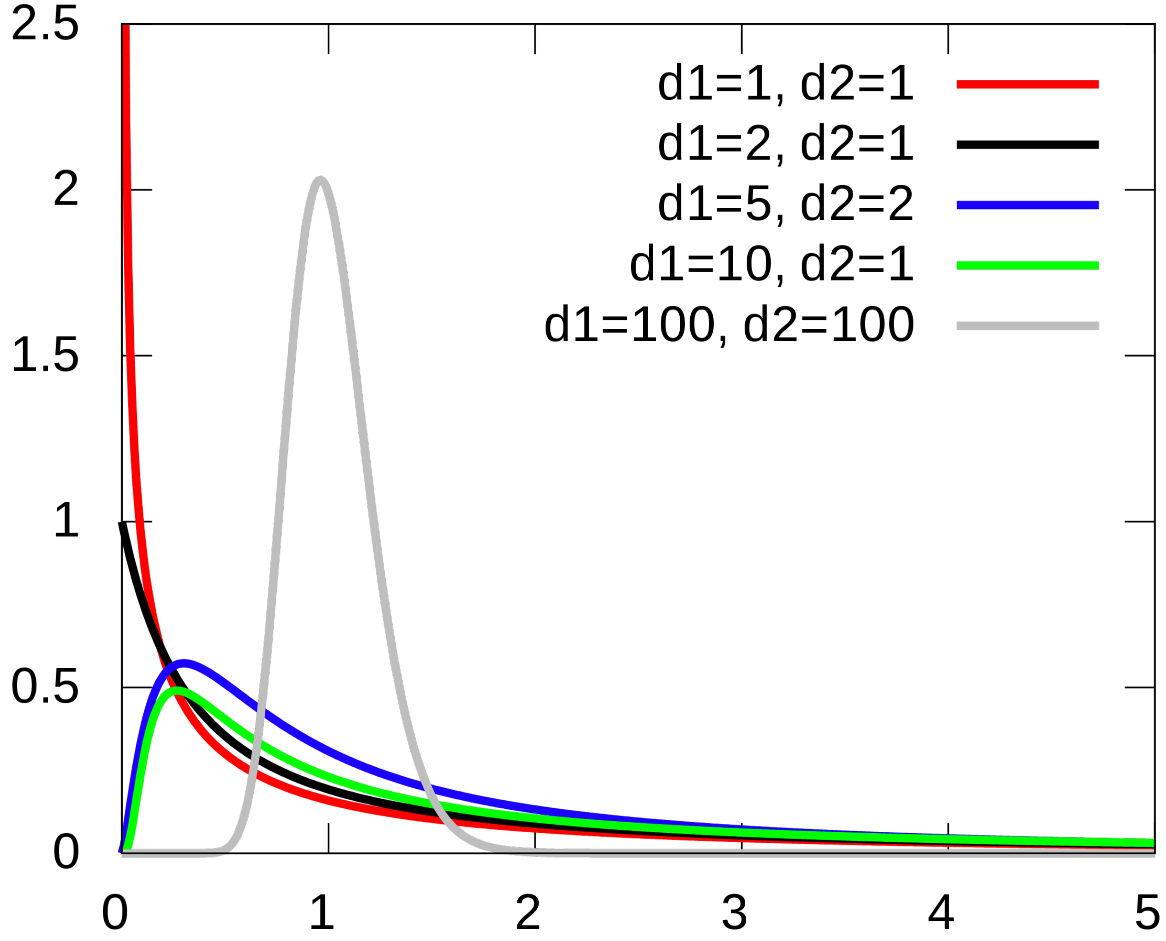
F-Statistic

- 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
- $$F = \frac{\frac{1}{\Delta_{DOF}(f,g)}\Bigg(\overbrace{\sum_i \big(y_i - g(x_i)\big)^2}^{\text{Pol1}} - \overbrace{\sum_i \big(y_i - f(x_i)\big)^2}^{\text{Pol2}}\Bigg)}{\underbrace{\Bigg(\frac{1}{n-p_f}\sum_i \big(y_i - f(x_i)\big)^2\Bigg)}_{\text{Pol2}}}$$

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

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

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

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

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

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 $\mathrm{S_j^2} = \frac{1}{\mathrm{m}-1}\sum_{\mathrm{i}=1}^{\mathrm{m}} (\mathrm{X_i} - \bar{\mathrm{X}}_\mathrm{j})^2$
- Define: $MS_B = \frac{m}{n-1}\sum_{i=1}^{n} (\bar{X}_j - \bar{X})^2$ and $\mathrm{MS_R} = \sigma^2$

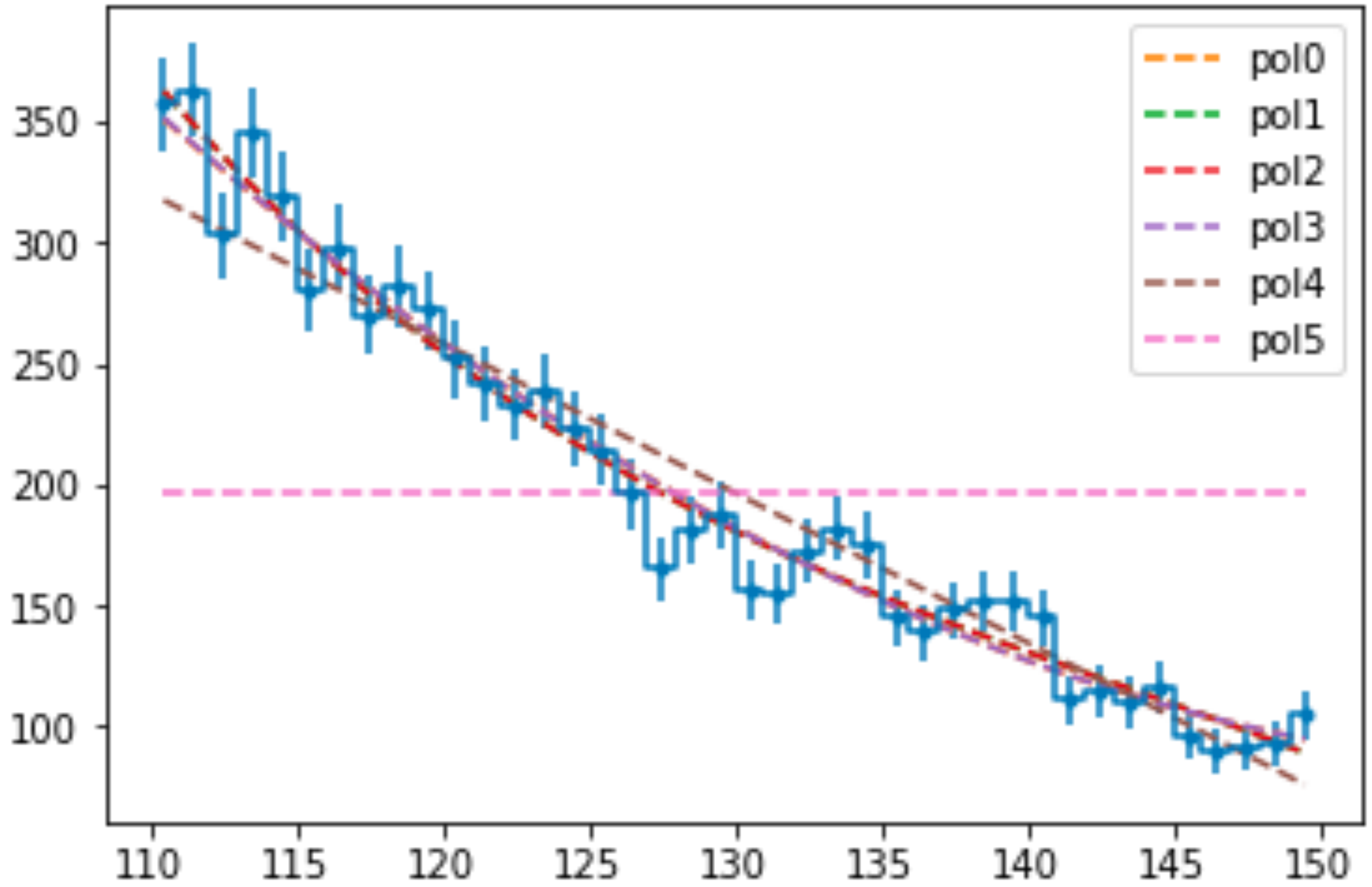
Difference in likelihoods

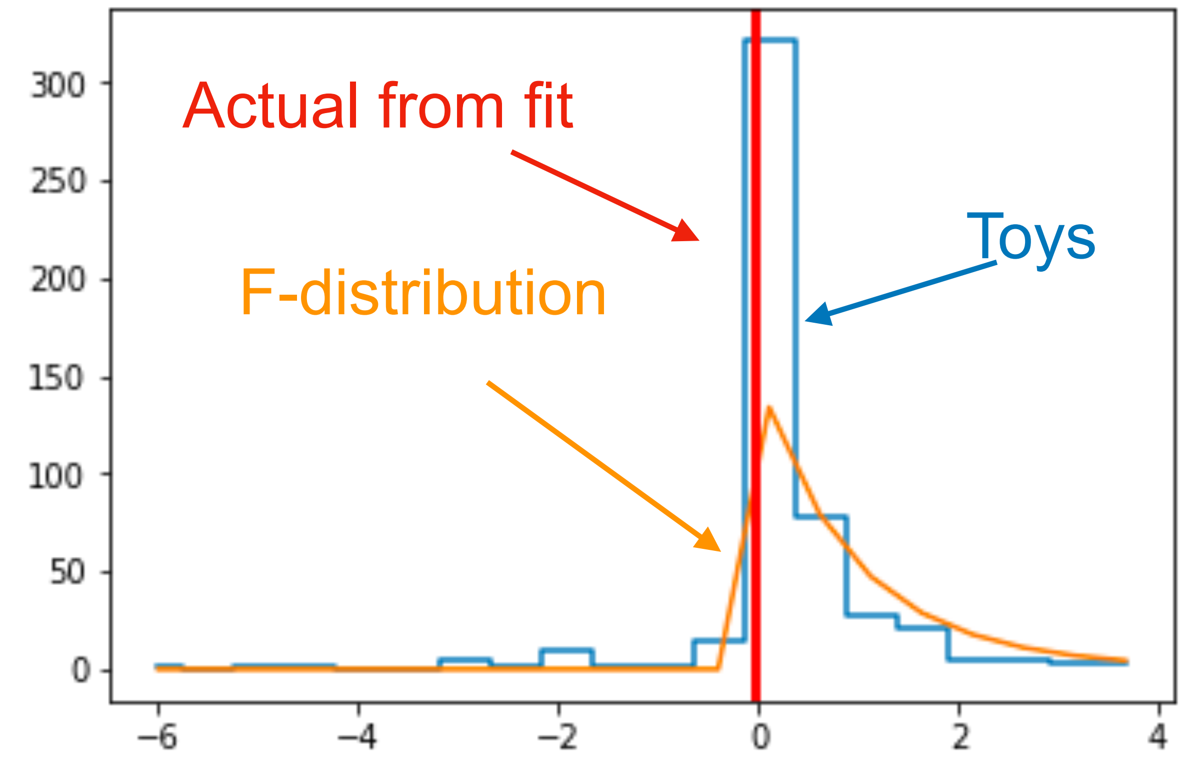
- If $\mu_i = \mu$ or in other words are from the same distribution the
- $\frac{MS_B}{MS_R} \approx 1 = F_{n-1\ m(n-1)}$ where $F_{n-1\ m(n-1)}(x)$ is an F distrib

F-test Example

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

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

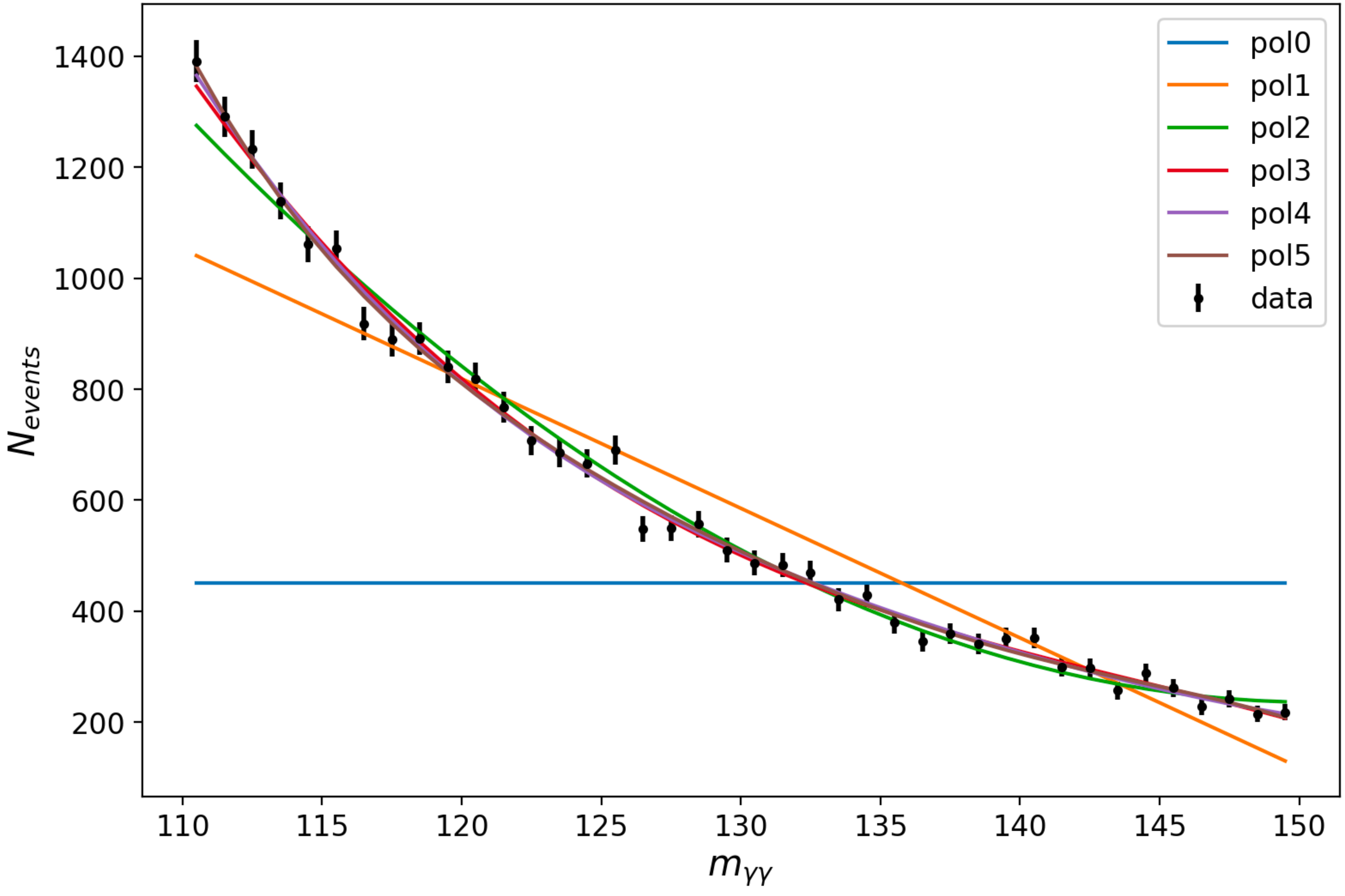




$$\frac{\frac{\mathcal{L}_1 - \mathcal{L}_2}{p_2 - p_1}}{\frac{\mathcal{L}_2}{n - p_2}}$$ For a 4th order to a 3rd order

Recap

- Sent an Army of polynomials to fit the Higgs boson



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

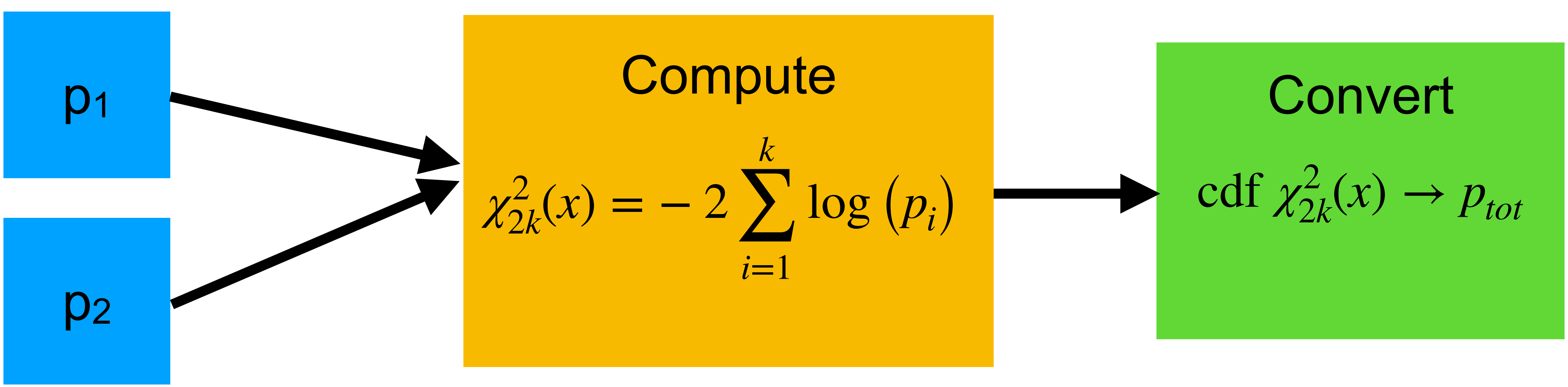
Pol1 **Pol2**

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

Pol2

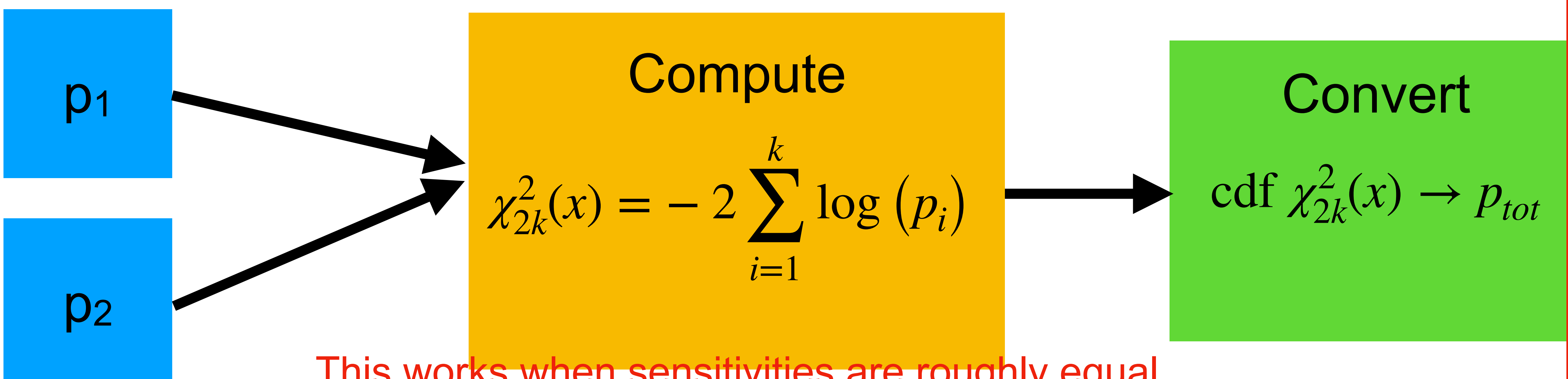
Combining Categories

- 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



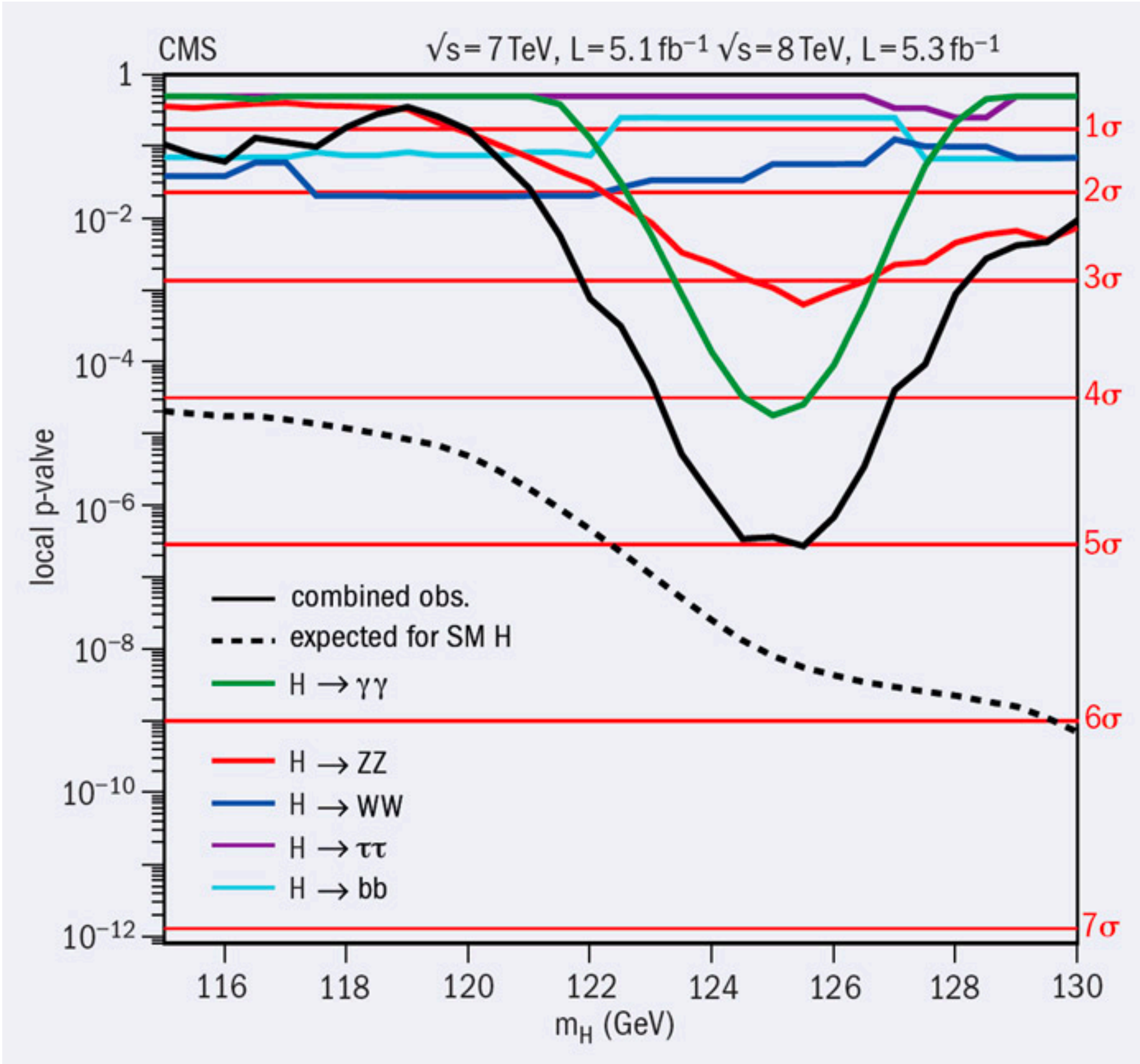
Combining Categories

- 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



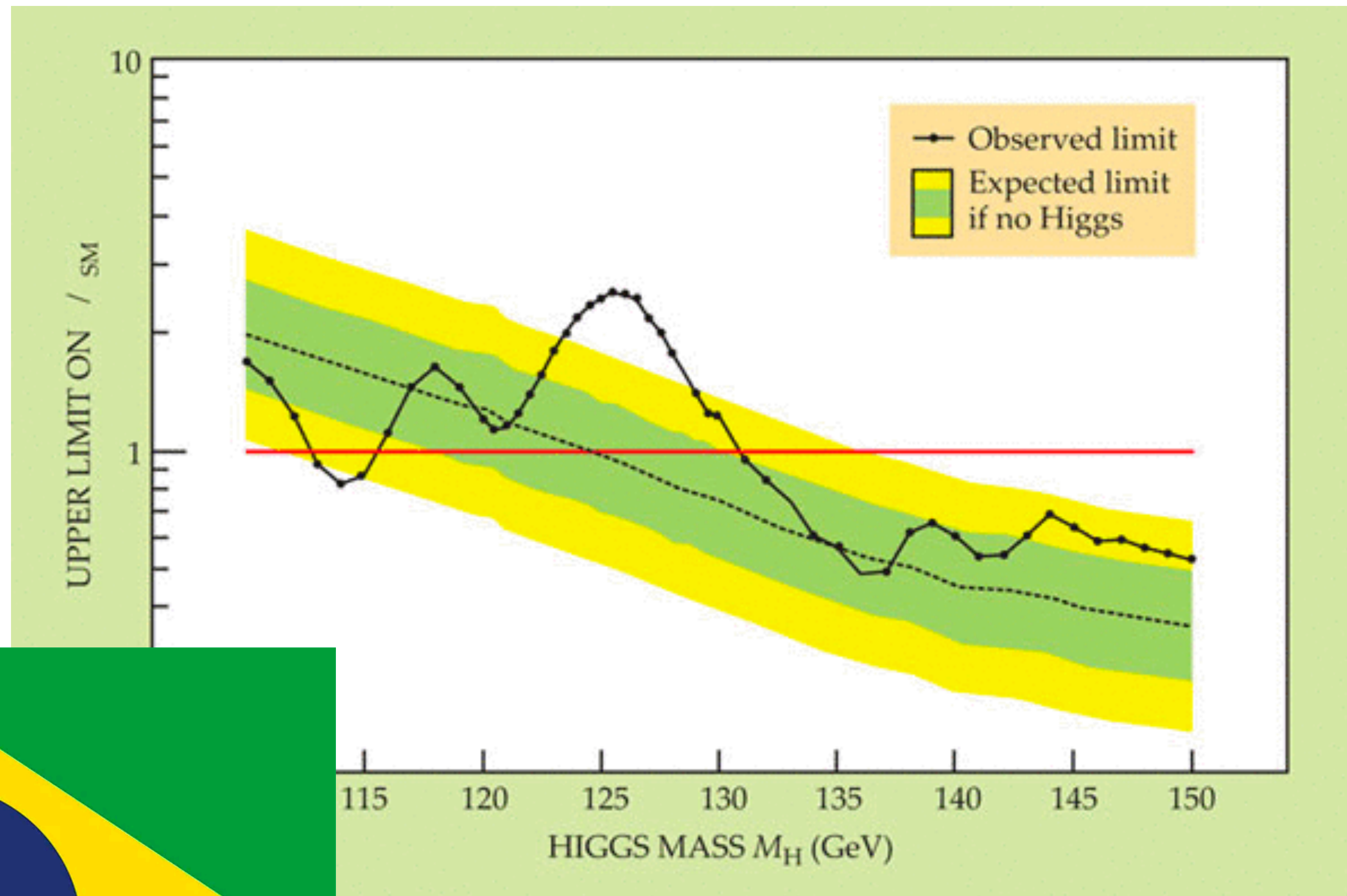
This works when sensitivities are roughly equal
More robust way is to fit all categories simulatneously

Combining Categories



Brazil Plot

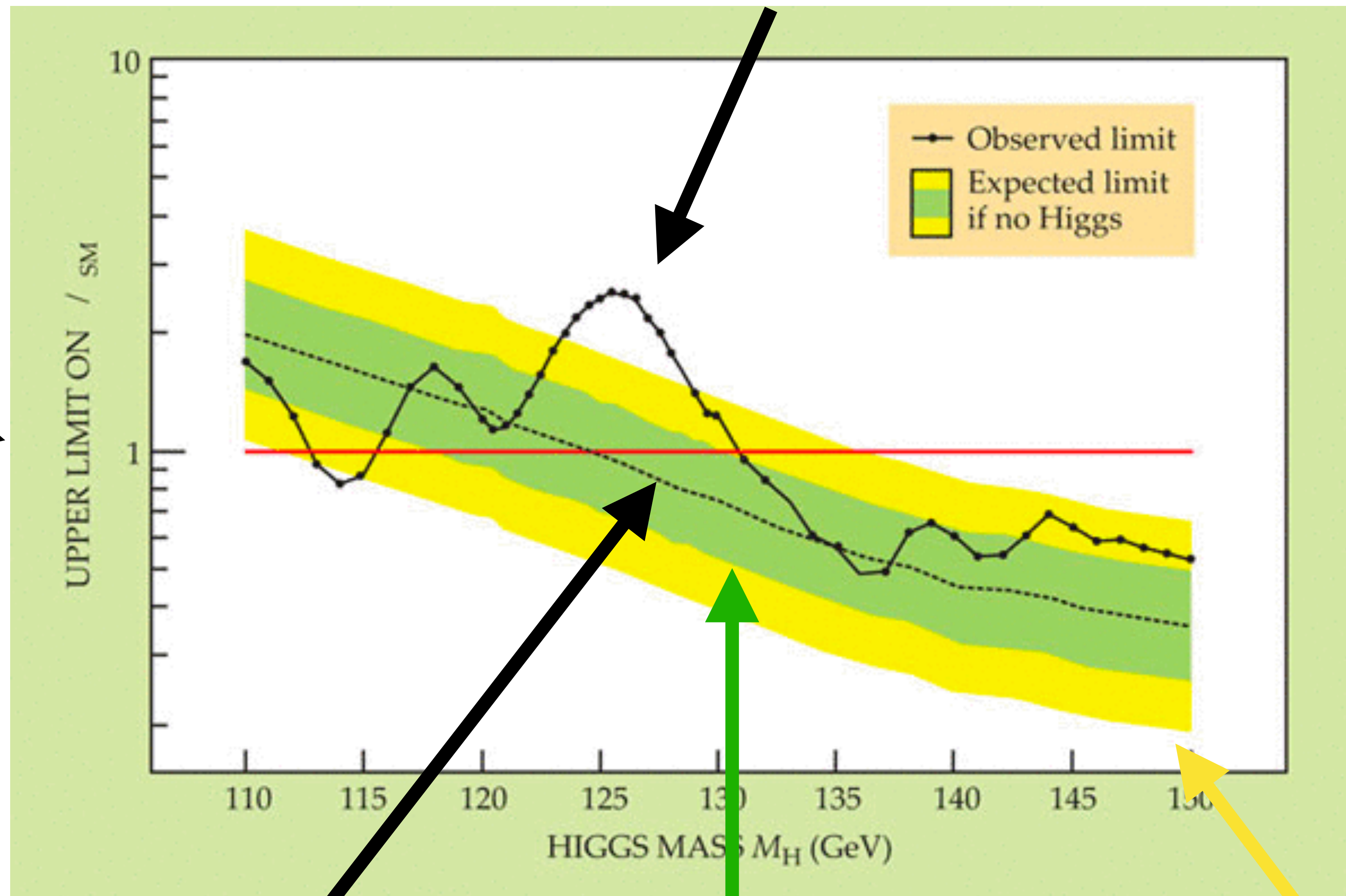
- What is this plot?



Brazil Plot

- What is this plot?

Observed Delta Log Likelihood



Signal
Strength in units
Of expected
(Defined by
prediction)

Expect signal
Strength for 95%
Exclusion

1 Standard Deviation
up and down band

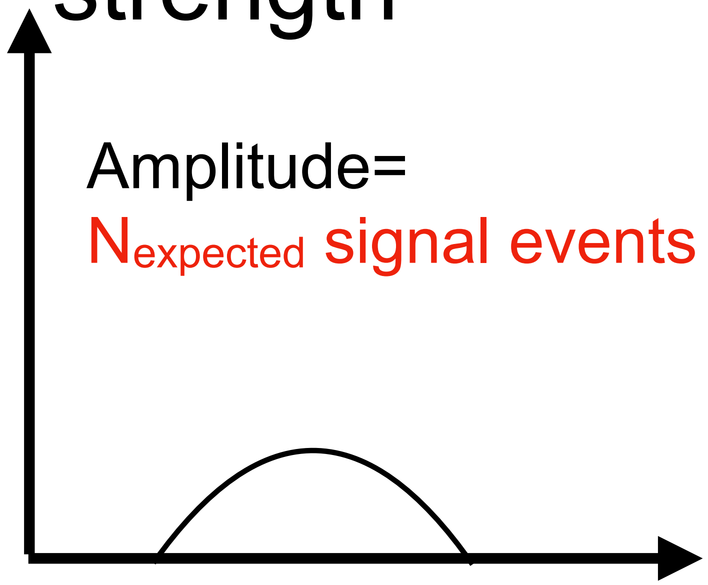
2 Standard Deviation
up and down band

The test of the limits

$$\mu = \frac{N_{\text{observed}}}{N_{\text{expected}}}$$

Alt Hypothesis

Higgs signal at a fixed mass and strength



Null Hypothesis

No signal

What is the 95% confidence level of our signal

Not being there in units of Expected signal strength

The test of the limits

Test Statistic

$$q_\mu = \log(\lambda) = 2\log\left(\frac{\mathscr{L}(\mu = N_{\text{expected}})}{\mathscr{L}(\mu = 0)}\right)$$

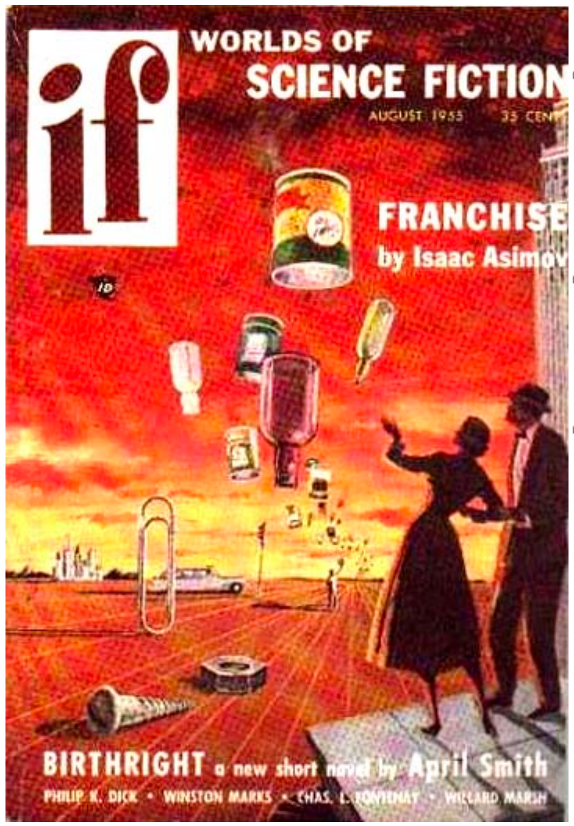
Exclusion : $p\left(q_\mu\right) > 0.95$ following Wilks' theorem

$$q_\mu > \chi^2(p_{95}) = 1.64 \rightarrow \frac{\mu}{\sigma_\mu} > 1.64$$

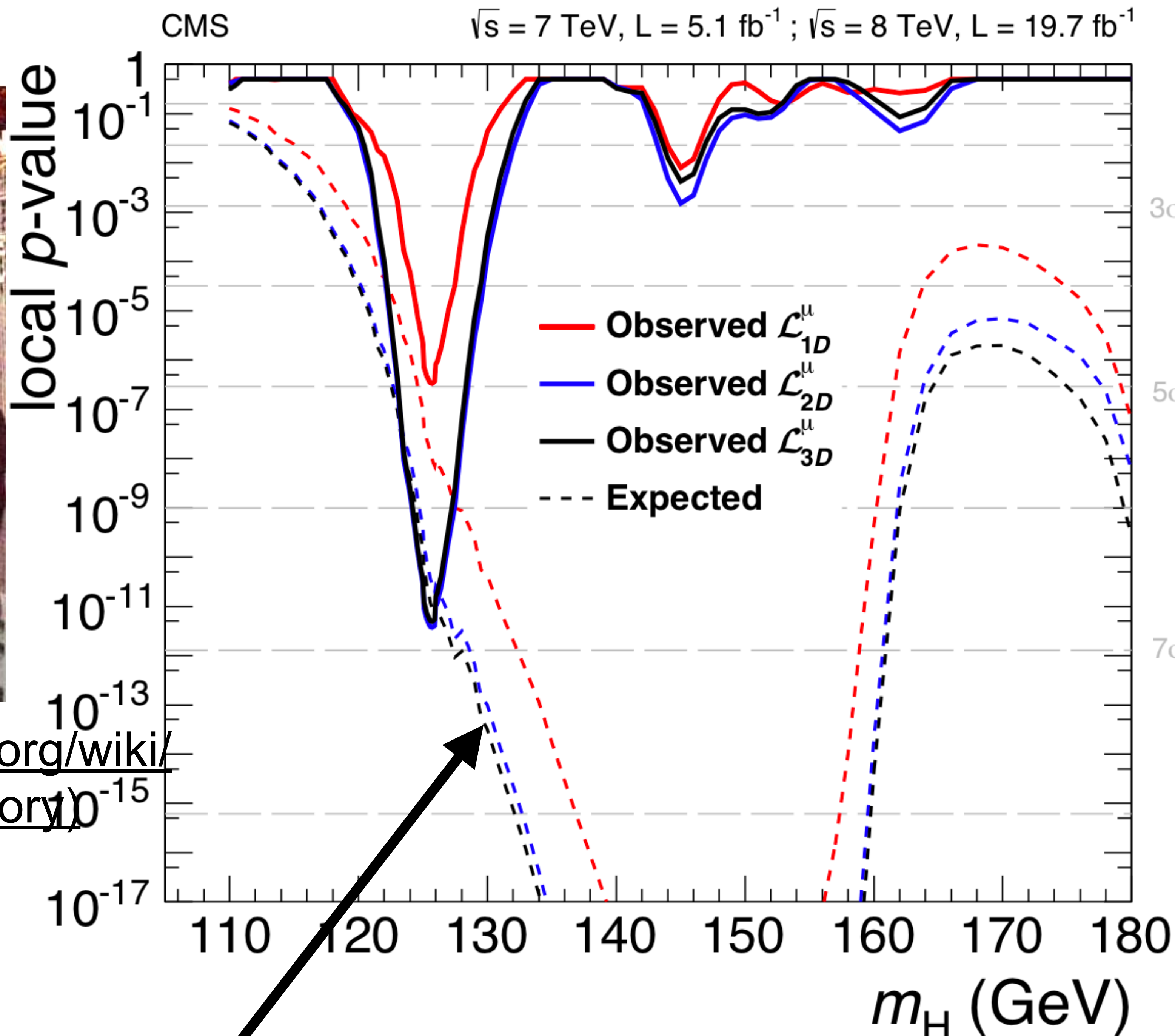
What is the 95% confidence level of our signal

Not being there in units of Expected signal strength

A p-value view of this



https://en.wikipedia.org/wiki/Franchise_(short_story)



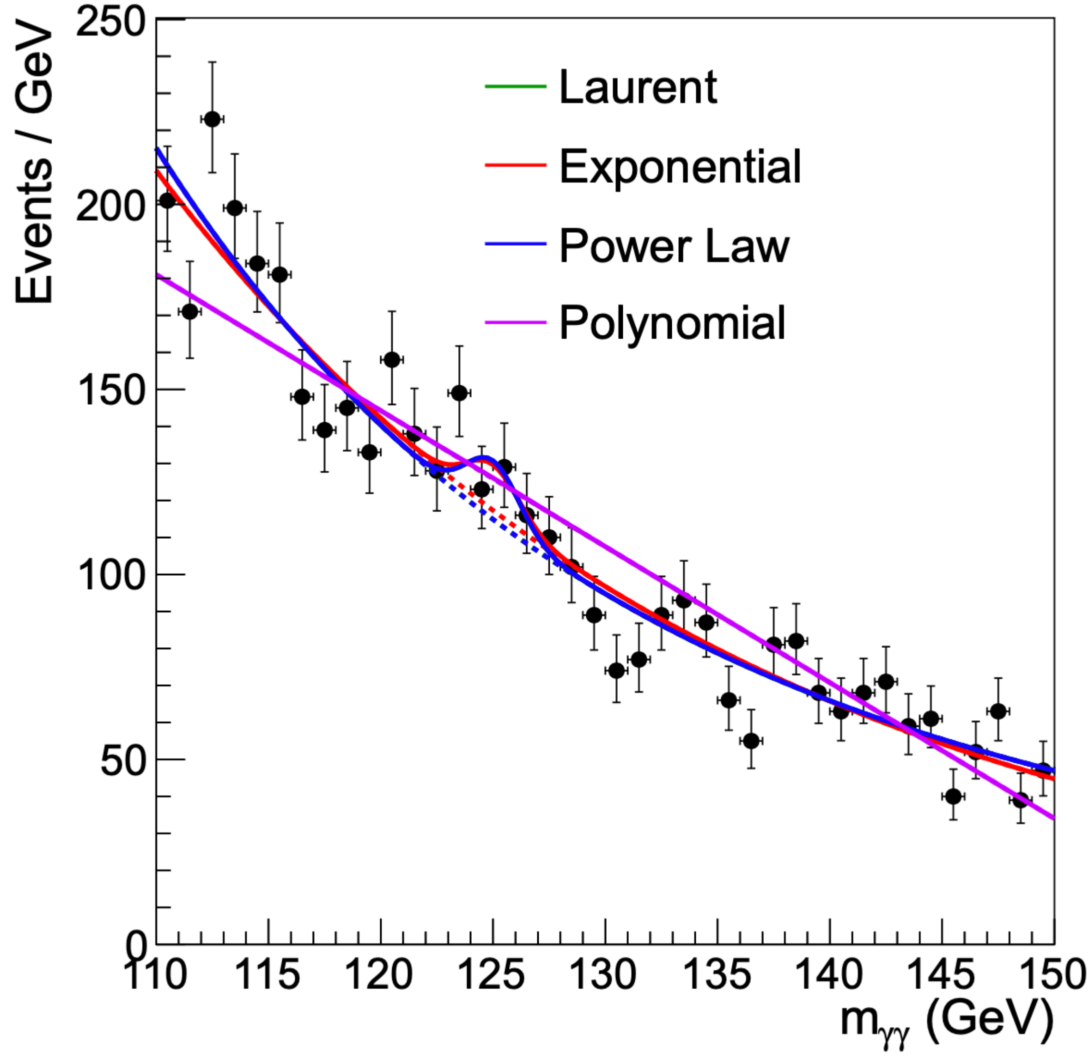
Line is based on the Asimov dataet

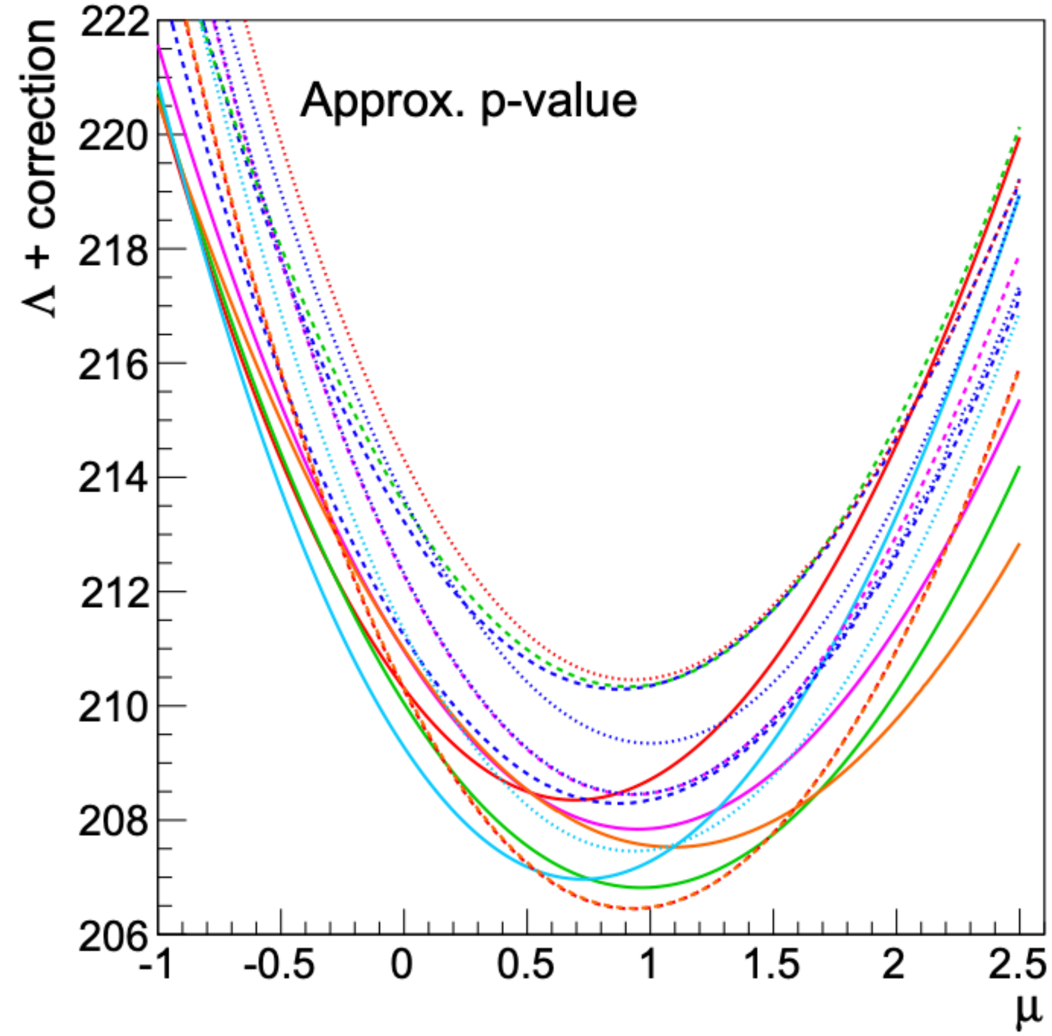
Line we get by running a limit on toy(fake) data with signal ininjected

The expected signal with magnitude by prediction is called the Asimov

Building a Model

- Throwing a barrage of functions at the problem

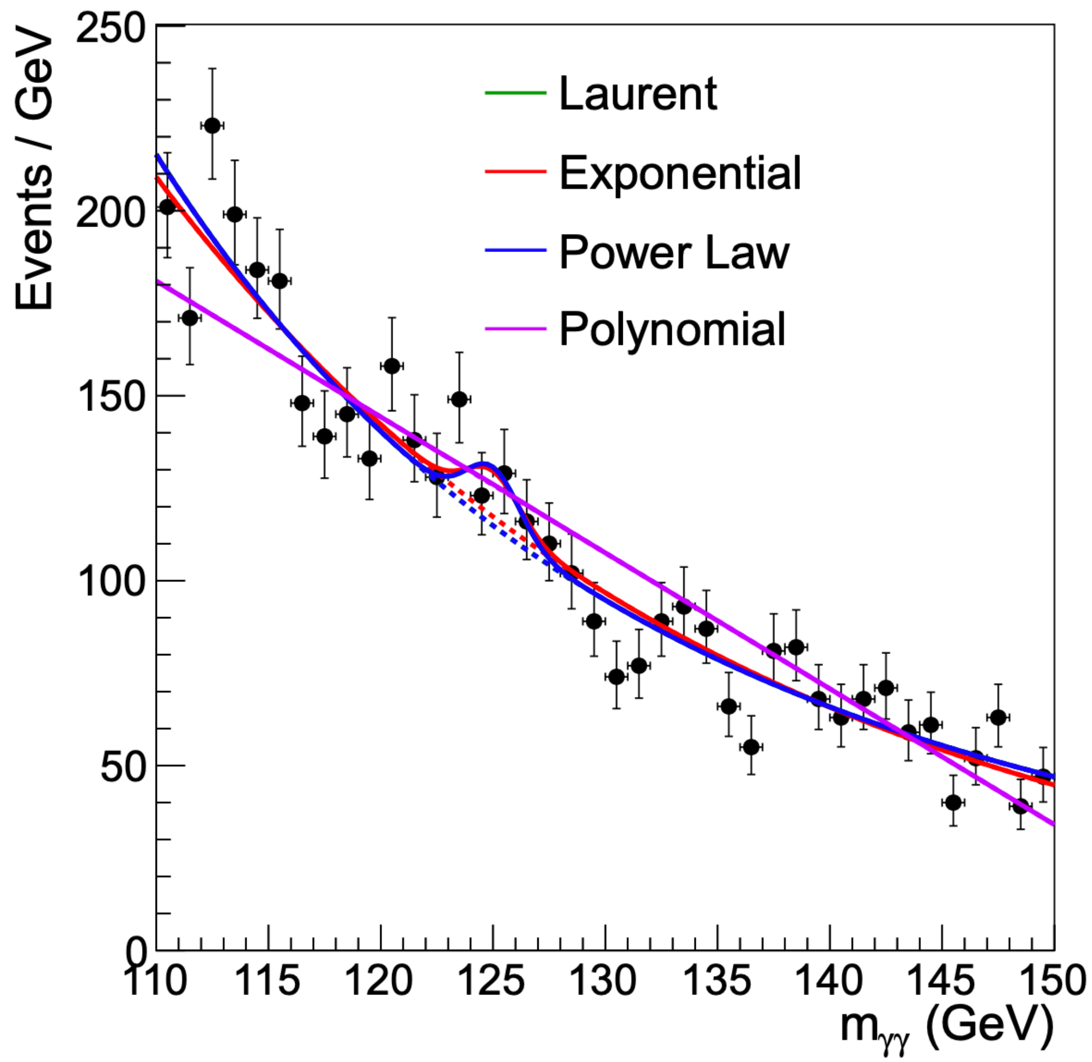


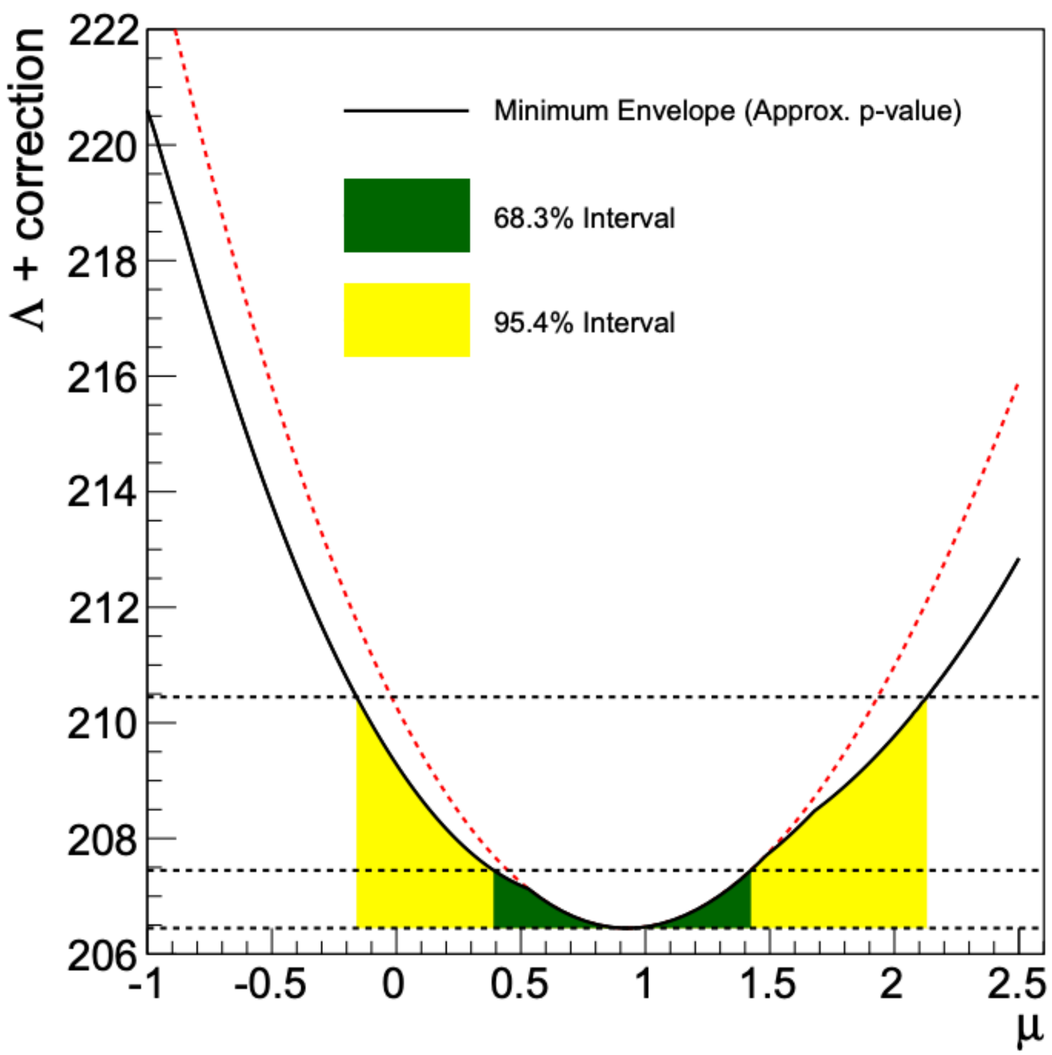


We can try a whole library of functions
The likelihood we get translates to our fit

Building a Model

- Throwing a barrage of functions at the problem

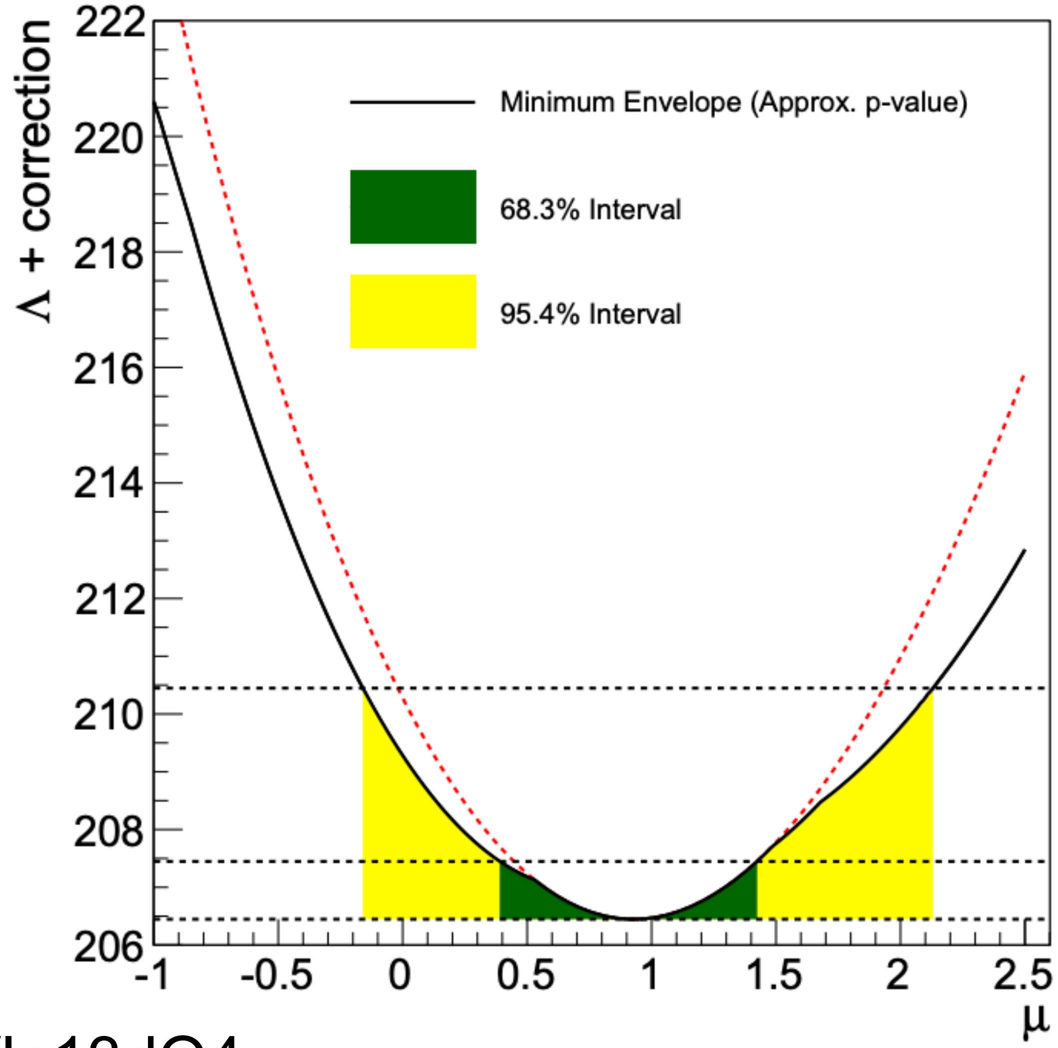




We can try a whole library of functions
The likelihood we get translates to our fit

Combined Likelihood

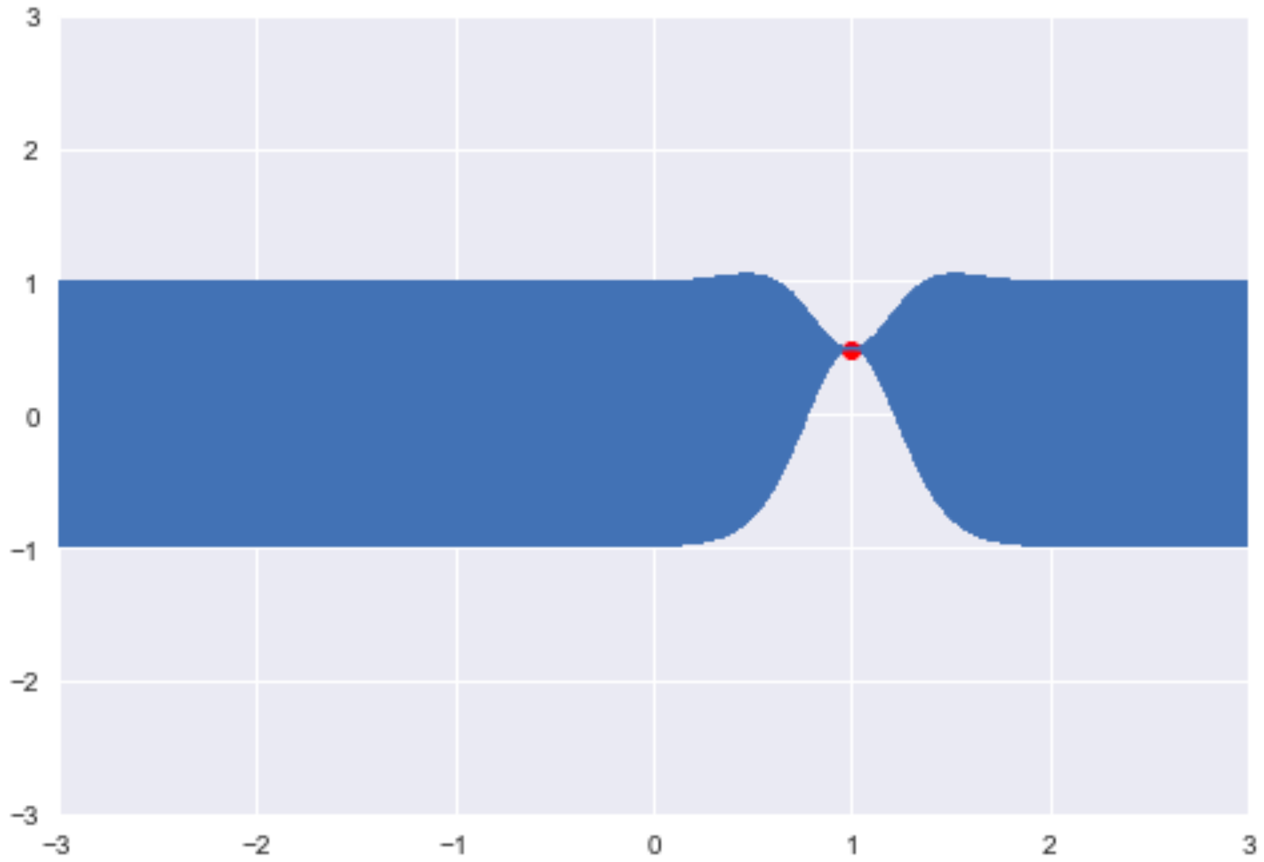
- Throwing a barrage of functions at the problem



https://www.youtube.com/watch?v=3cHWlp13dQ4

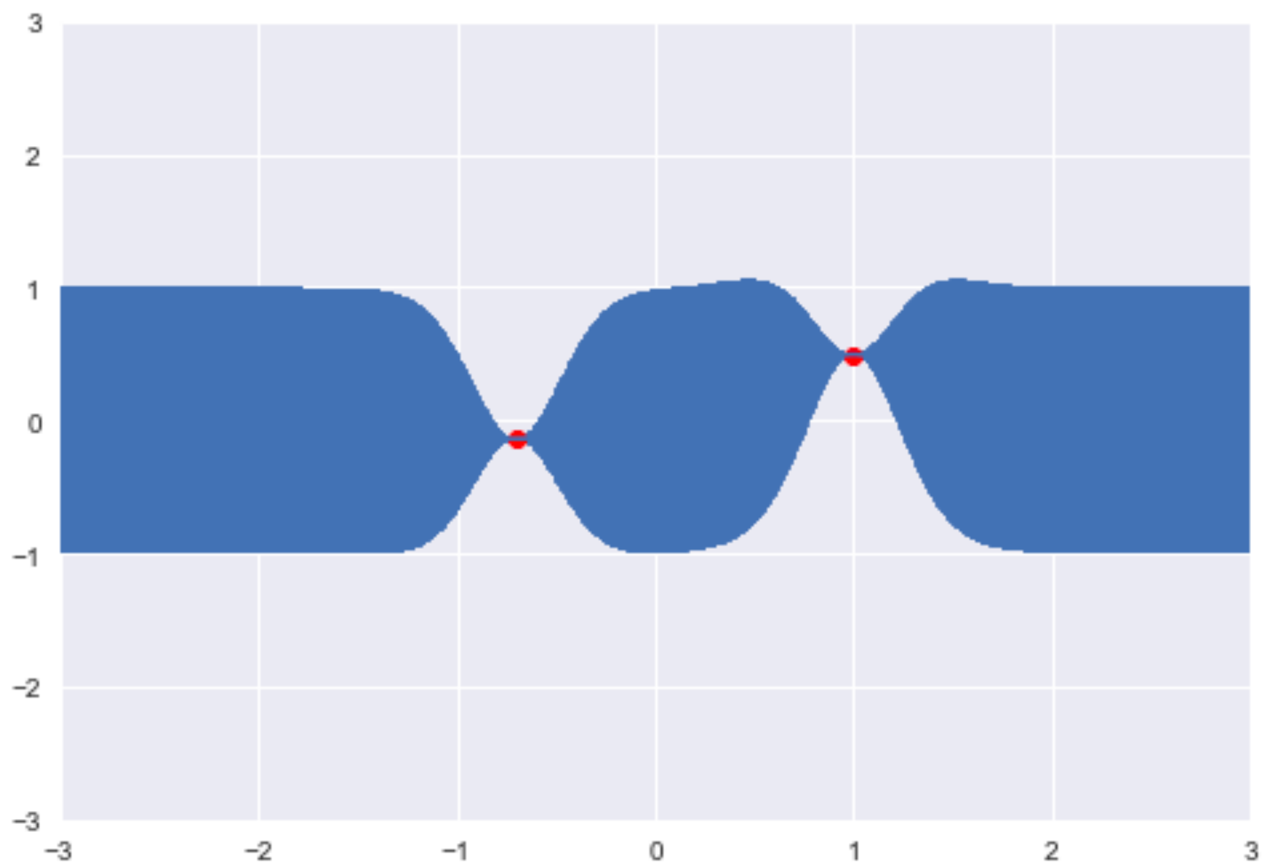
We can try a whole library of functions
The likelihood we get translates to our fit

Splines+GaussianProcesses



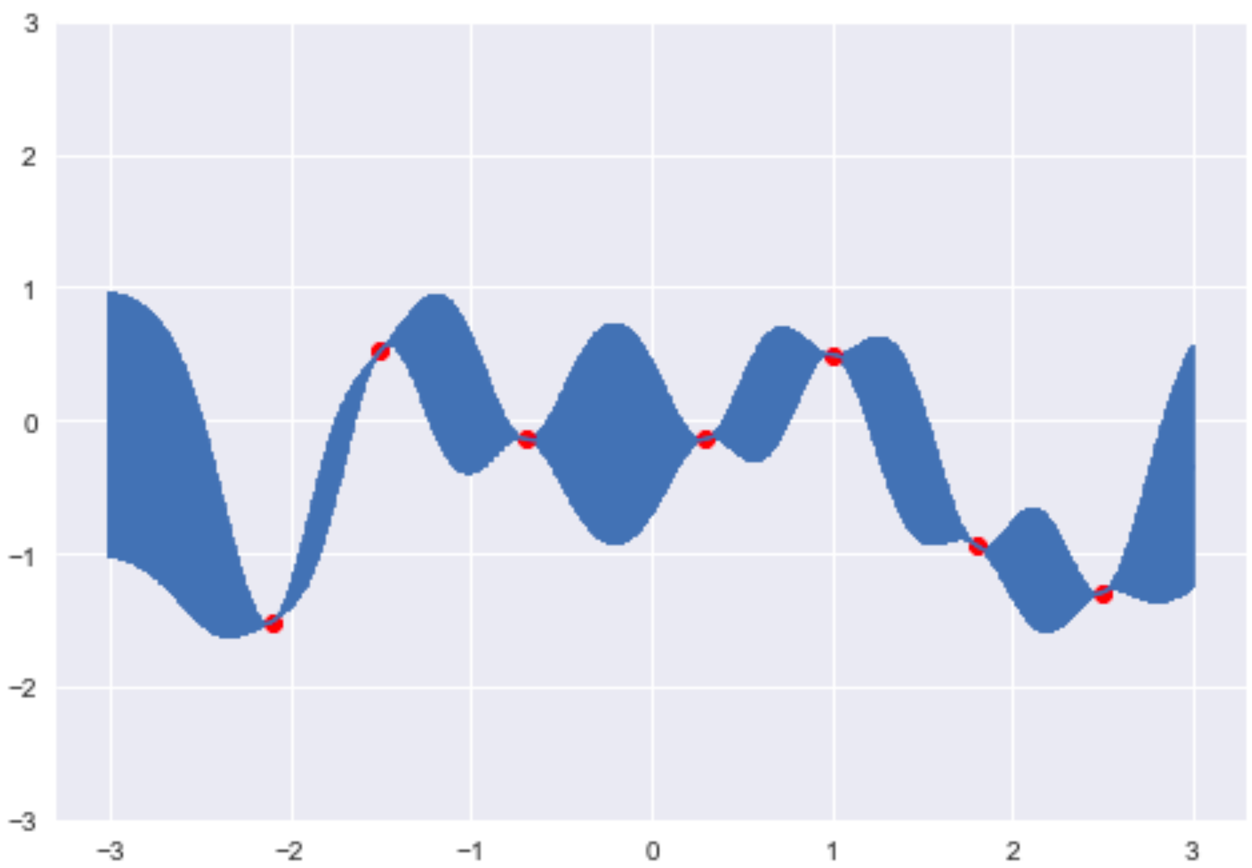
- Gaussian processes allow us to build function choice from the data

Splines+GaussianProcesses



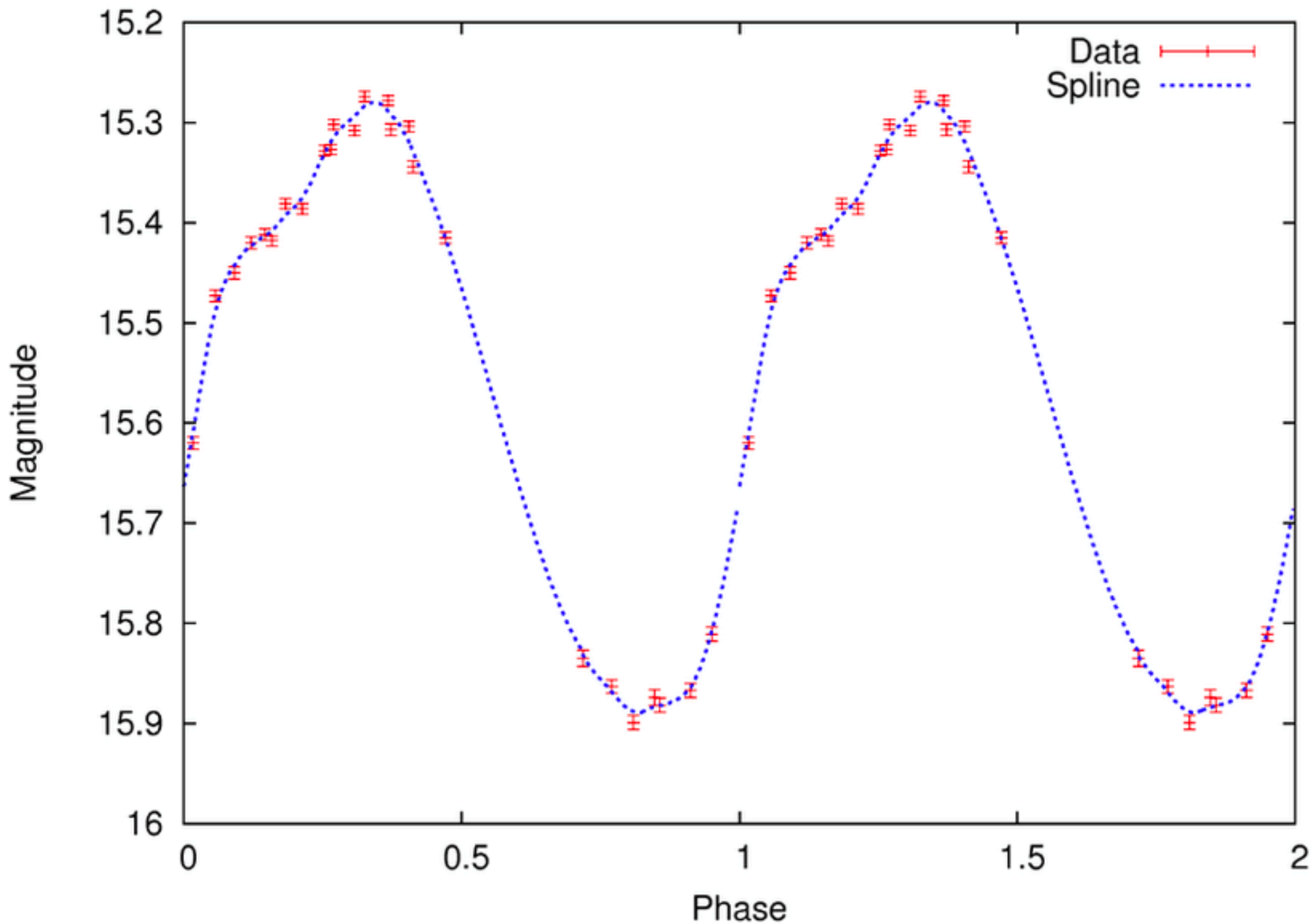
- Gaussian processes allow us to build function choice from the data

Splines+GaussianProcesses

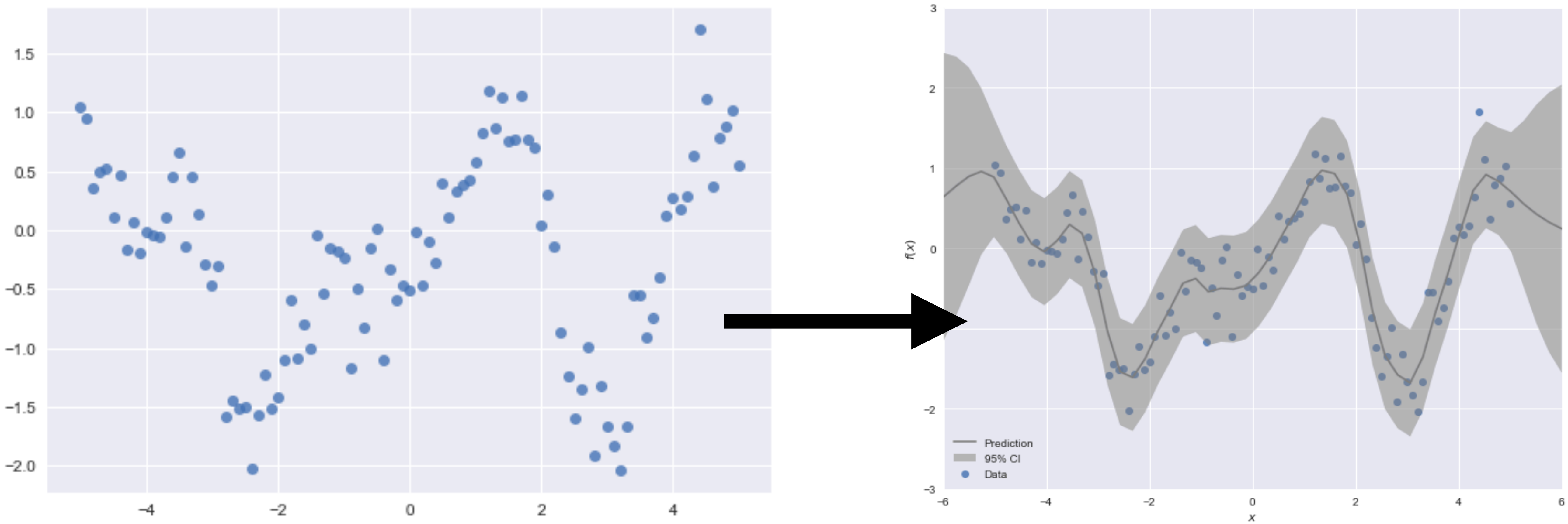


- Gaussian processes allow us to build function choice from the data

Splines+GaussianProcesses



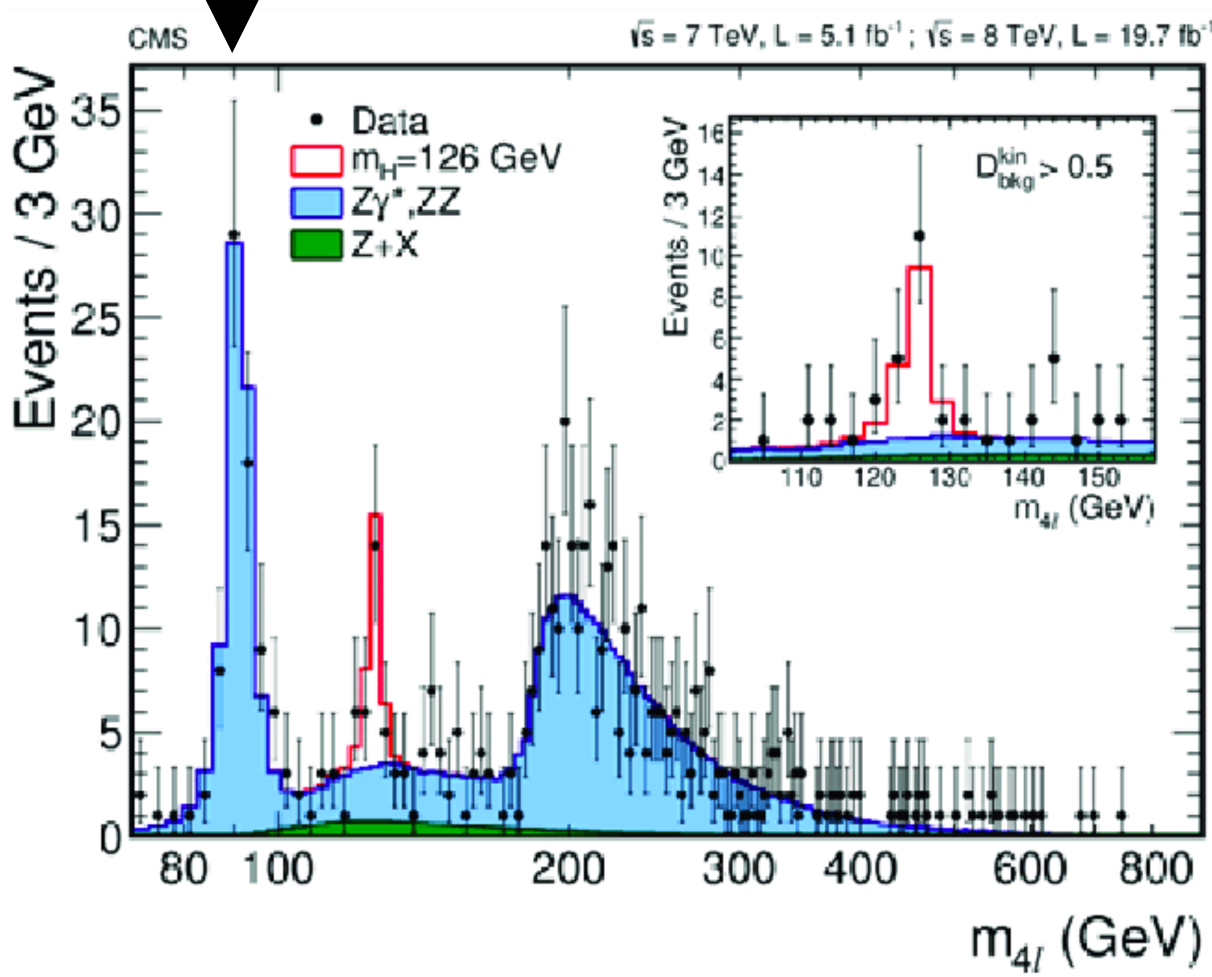
Splines+GaussianProcesses

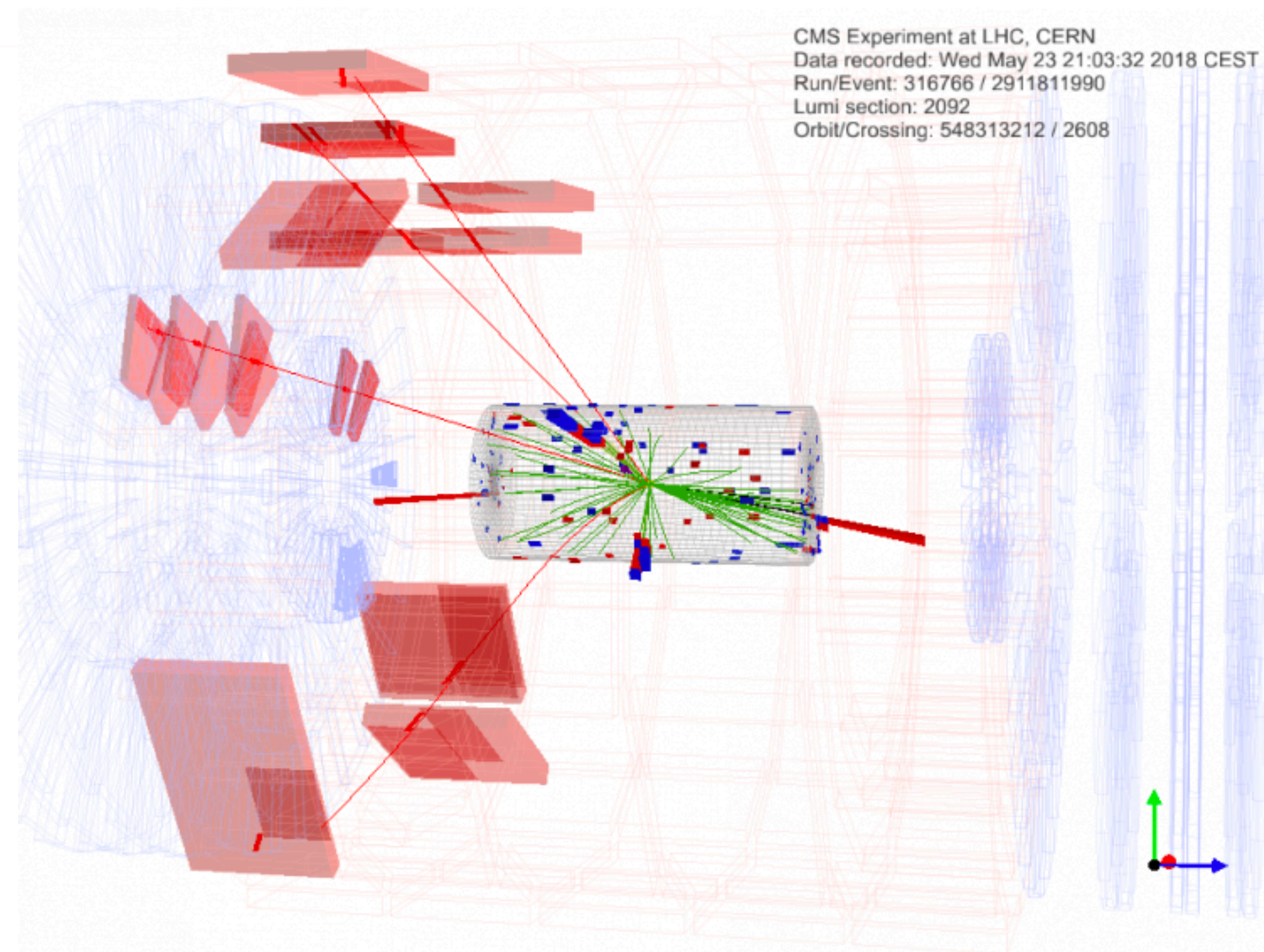


- To go from points to spline automatically

Higgs to 4 Leptons

Z-boson Peak





- 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

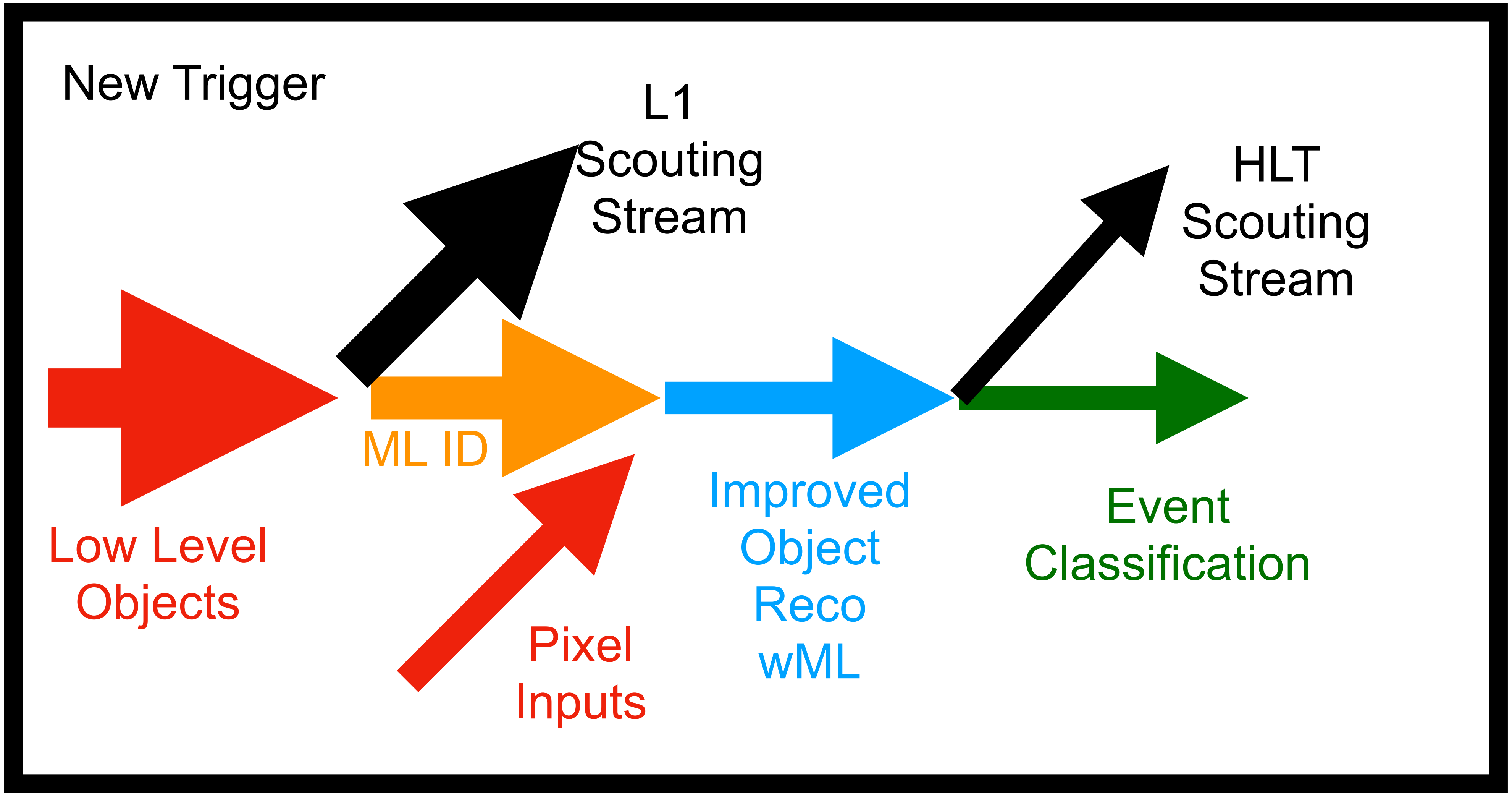
- We can evaluate this through an F-test

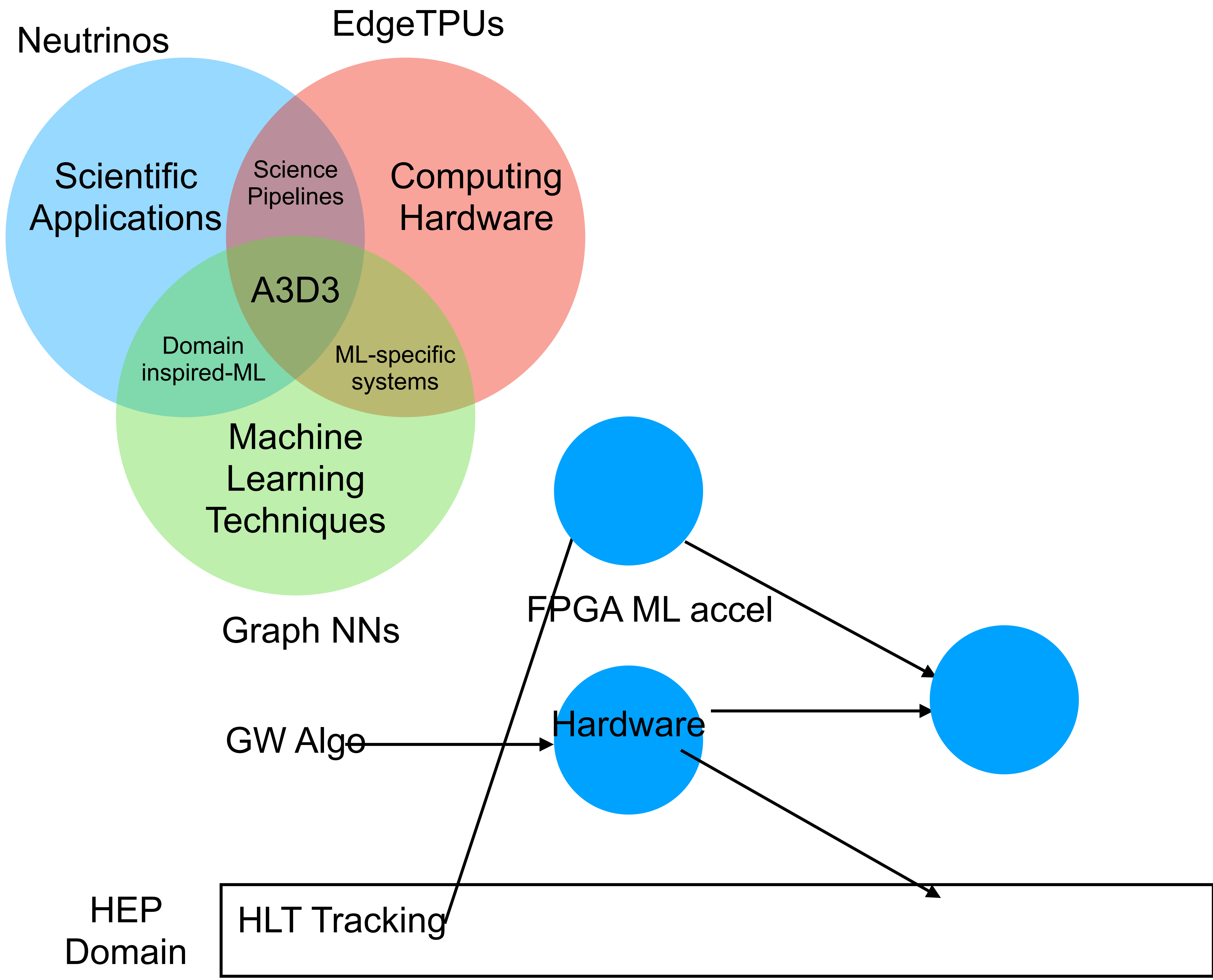
- Recall $\frac{MS_B}{MS_R} \approx 1 = F_{n-1\ m(n-1)}$

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

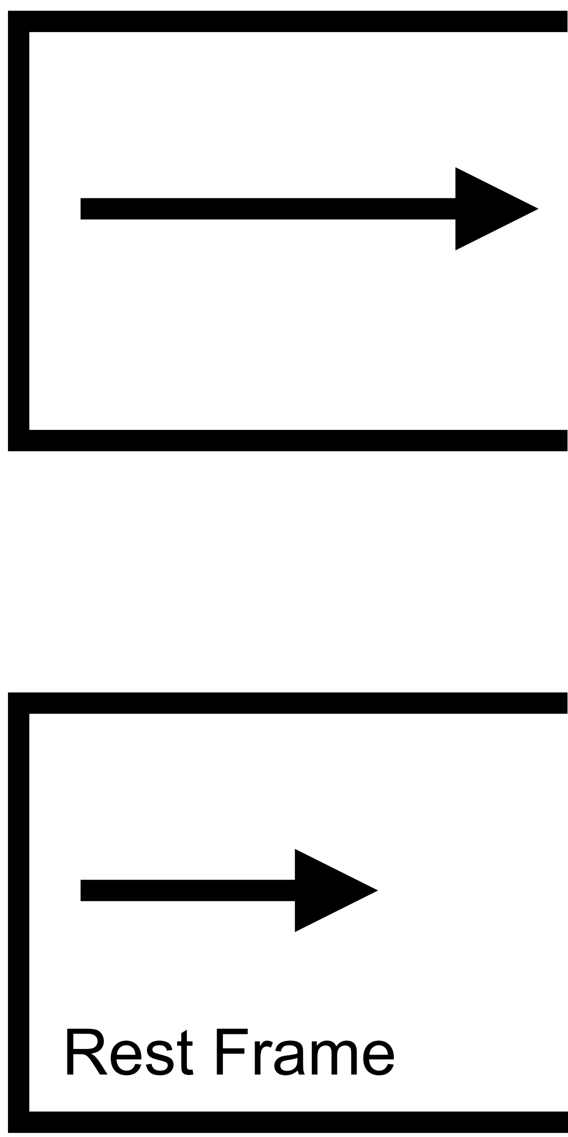
$$F = \frac{\left(\frac{\mathrm{RSS}_1 - \mathrm{RSS}_2}{p_2 - p_1}\right)}{\left(\frac{\mathrm{RSS}_2}{n - p_2}\right)},$$

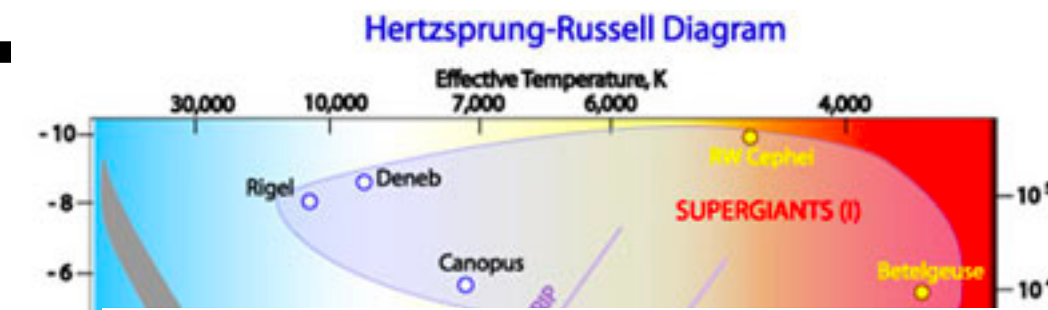
Title Text

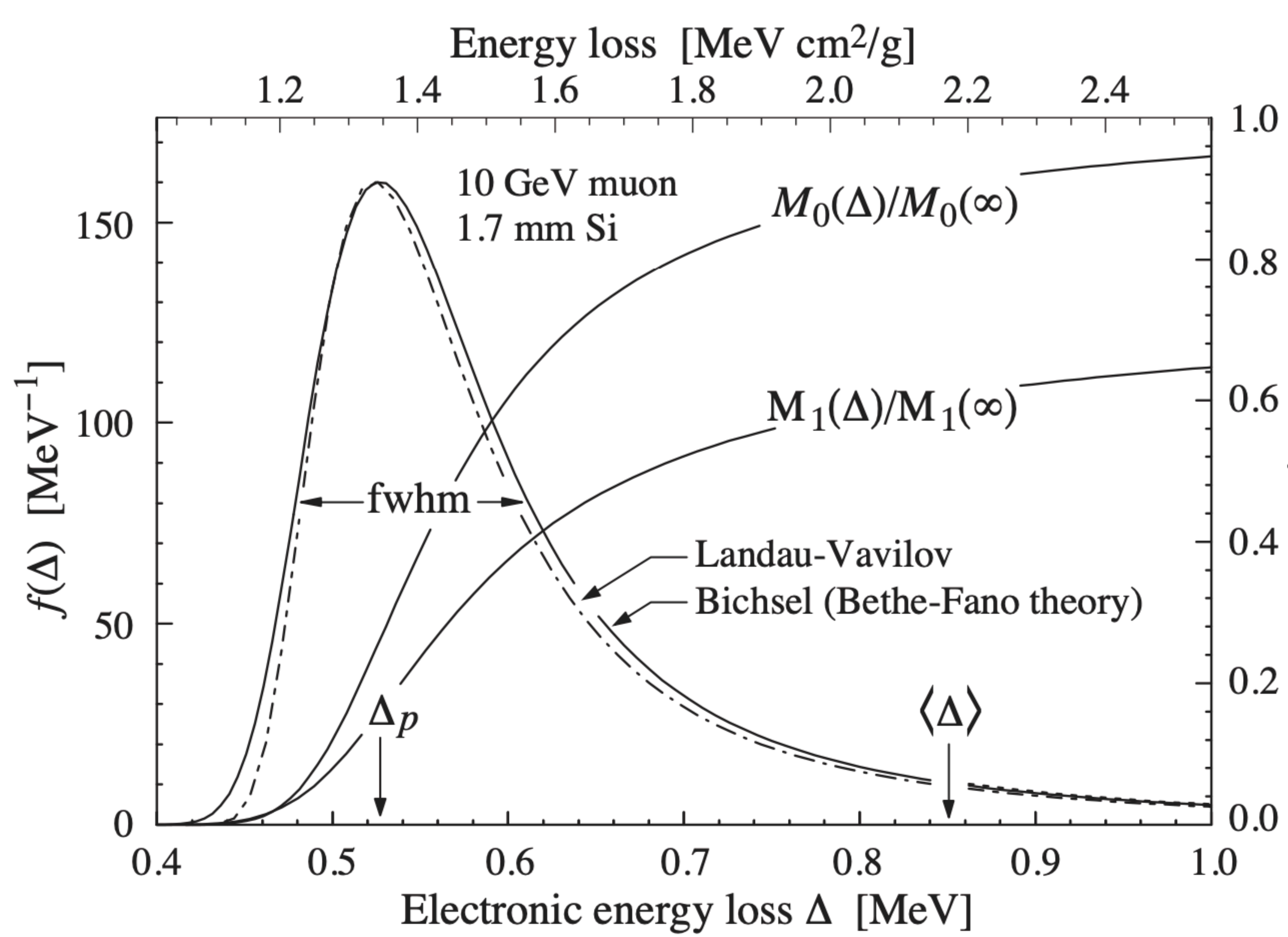

New Trigger
L1
Scouting
Stream
HLT
Scouting
Stream
ML ID
Low Level
Objects
Pixel
Inputs
Improved
Object
Reco
wML
Event
Classification

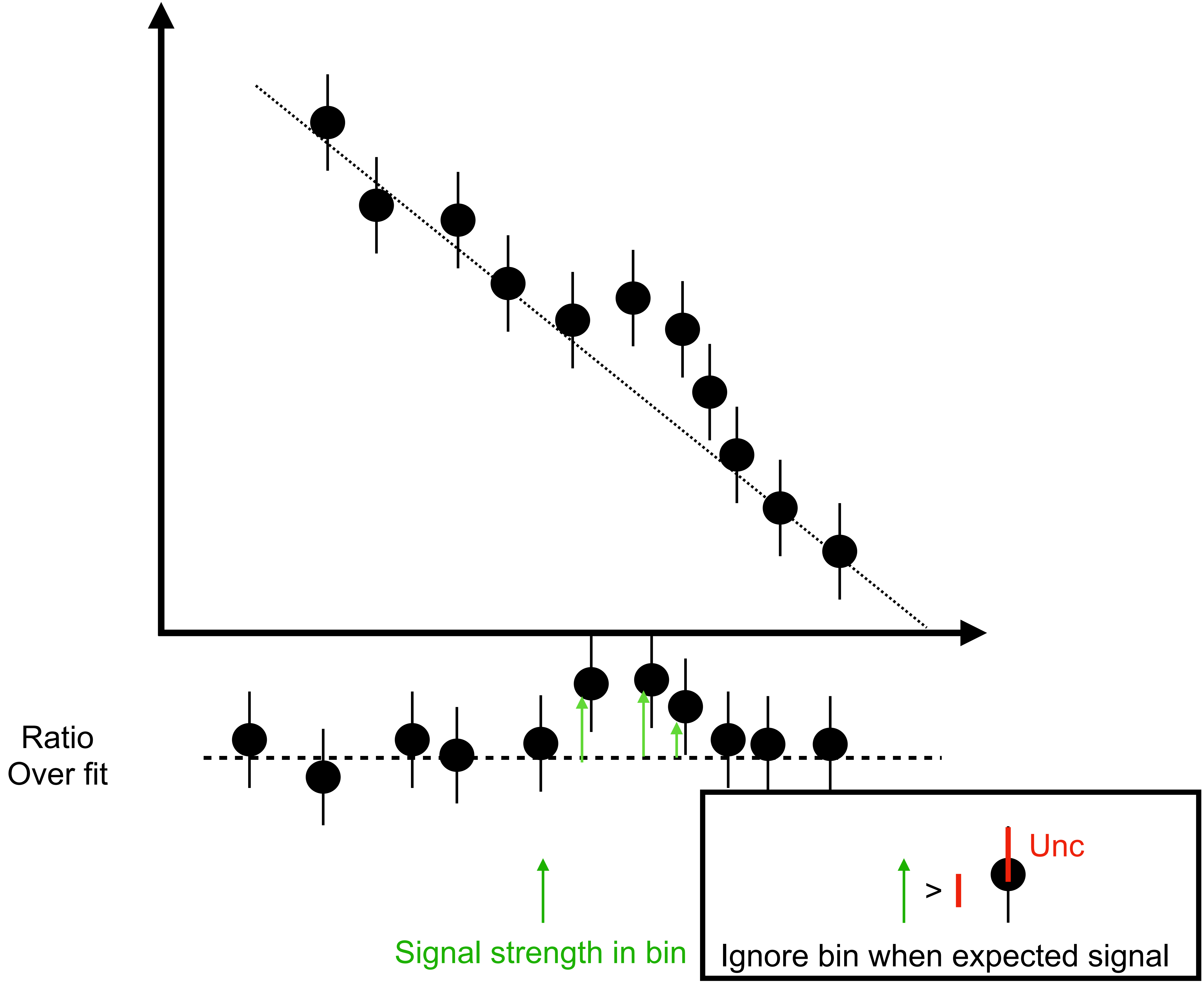

Neutrinos
EdgeTPUs
Scientific Applications
Science Pipelines
Computing Hardware
A3D3
Domain inspired-ML
ML-specific systems
Machine Learning Techniques
Graph NNs
FPGA ML accel
GW Algo
Hardware
HEP Domain
HLT Tracking

Elastic Scatter








Ratio
Over fit
Signal strength in bin
Unc
>
Ignore bin when expected signal