Data Analysis & Machine Learning

Checkpoint

Peter Clarke

THE UNIVERSITY of EDINBURGH

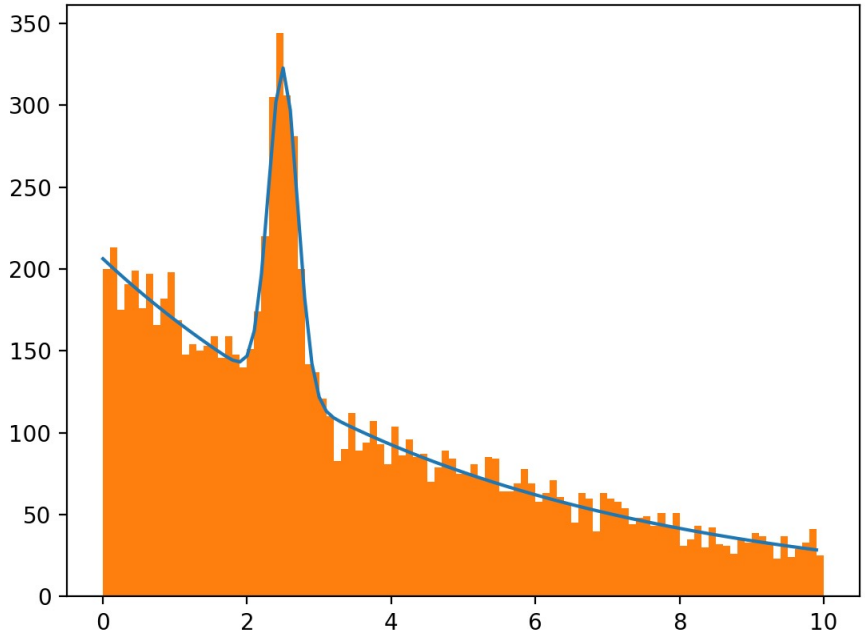
Checkpoint 7:

Systematic Errors

❑ In last CP you were given this dataset to fit to

It contains 10,000 measurements from a more complicated distribution

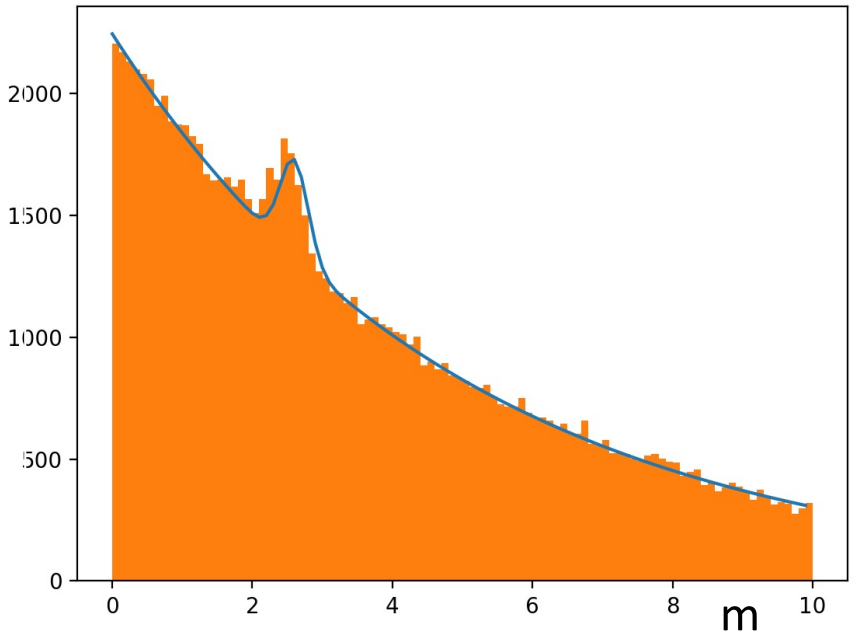
- A fraction F=0.9 of events from an exponential with a lifetime of 5
- A fraction (1-F) =0.1 of a Gaussian with mean at 2.5 and width of 0.2
- There are only events in the range 0-10



- In this CP you are going to fit to a similar dataset, but with far fewer signal events. This looks more like the Higgs signal
- Date in file:
  datafile-expresonance.txt

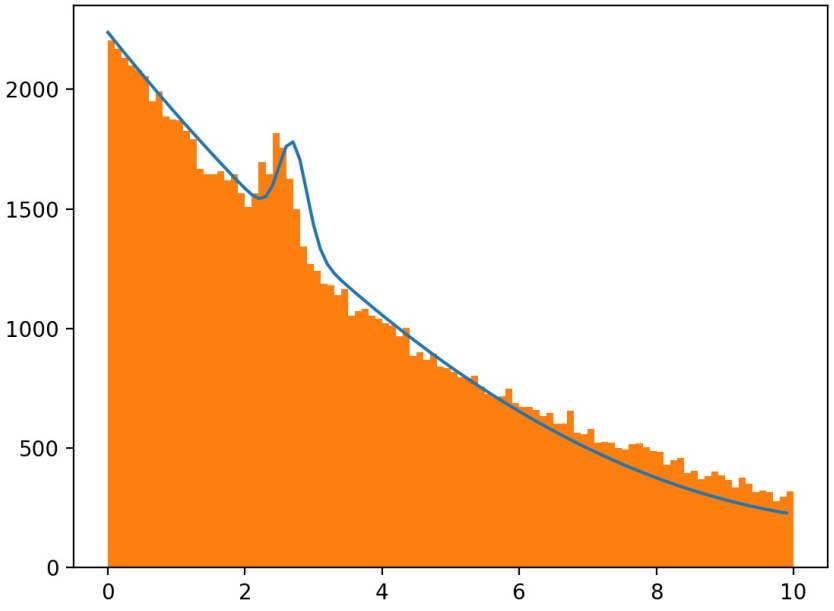
- It contains 10,000 measurements from a mass distribution, where m = mass

- It is described by:
  - A fraction F=0.98 of events from a background exponential ~ exp(-m/A) where A = 5
  - A fraction (1-F) =0.02 of a signal Gaussian
  - Gaussian mean M = 2.5 and width of 0.2
  - There are only events in the range 0-10



- CP Part 1:
  - i. Fit to these data with the above model to determine the best fit parameters for F, A and M
  - ii. Determine the statistical errors on each fit parameter
    - If using optimise you will have to write a bit of code to find the NLL+0.5 points
    - If using Minuit then errors come free in the fit - but remember to set errordef=0.5
  - iii. Present the best fit values and errors on all parameters in clear way, i.e. with descriptive text and numerical results presented as: M = Value $\pm$ Estat
  - iv. Make sure the number of digits quoted is sensible – quote errors to 2 significant digits

- CP Part 2:
  - i. Determine a systematic error on the mean of the Gaussian (M) by the "shift" method
  - ii. Fit to the same data using the same Gaussian signal model but now using **a 2nd order polynomial background model**
  - iii. Present the results clearly as Value $\pm$ Estat $\pm$ Esyst
  - iv. Present the results as Value $\pm$ Etot
  - v. Comment on the importance of the systematic error compared to the statistical error for the measurement of M



Suitable start values for the polynomial

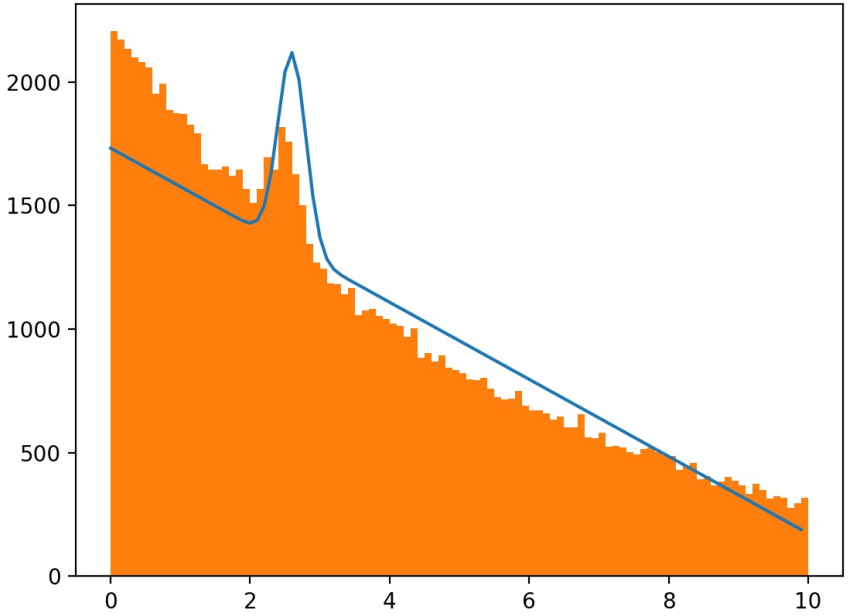
a + b*x + c*x**2

Are (I think):

b/a = -0.016
c/a = 0.007

Fix a=1
It should be irrelevant if your PDF normalizes itself properly

- CP Part 3:
  - i. Determine a systematic error on the mean of the Gaussian (M) by the "shift" method
  - ii. Fit to the same data using the same Gaussian signal model but now using **a straight line background model** (bad model - this is harder)
  - iii. Present the result clearly as M $\pm$ Estat $\pm$ Esyst
  - i. Present the results as M $\pm$ Etot
  - ii. Comment on the importance of the systematic error compared to the statistical error for the measurement of M



Suitable start values for the line

a + b*x

are:

b/a = -0.09

Fix a=1
It should be irrelevant if your PDF normalizes itself properly

Results should always be presented clearly with enough explanatory text and plots so that the marker can understand what you have done and what you are presenting

- E.g. a code+short pdf document with description
- E.g. adding text cells to your jupyter notebook

Code should have enough comments to explain what each function is

Marks are

Part 1: 40%
Part 2: 30%
Part 3: 30%