LHCb data analysis hands-on

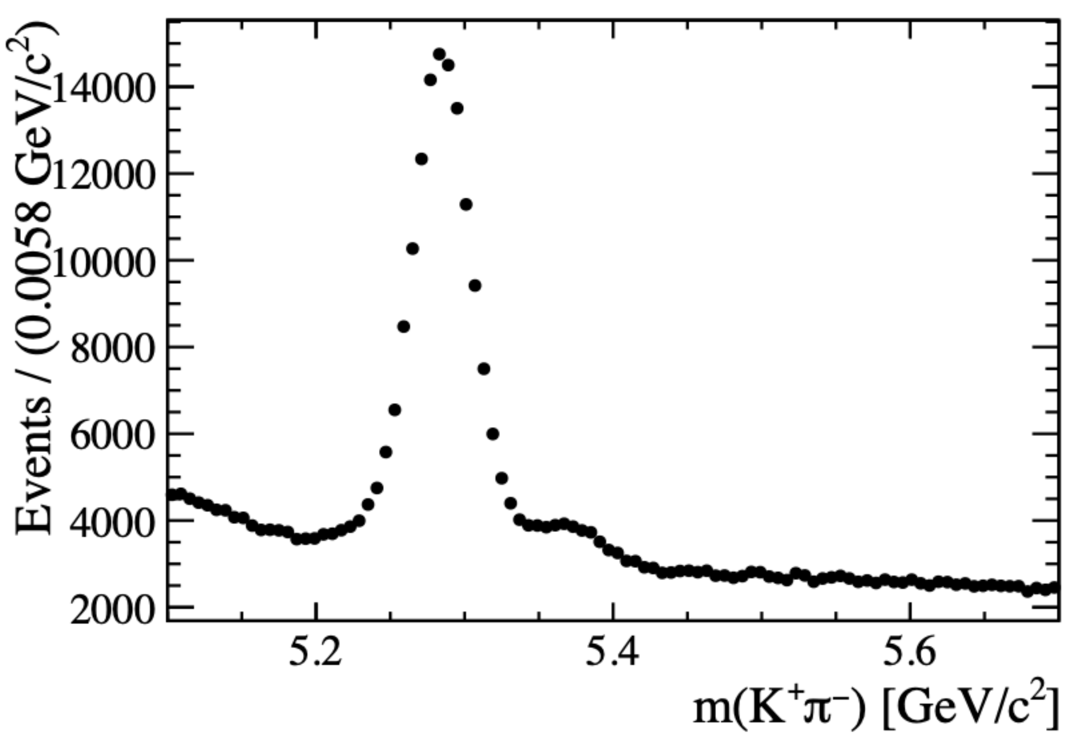
IDPASC 2025 - Orsay, France

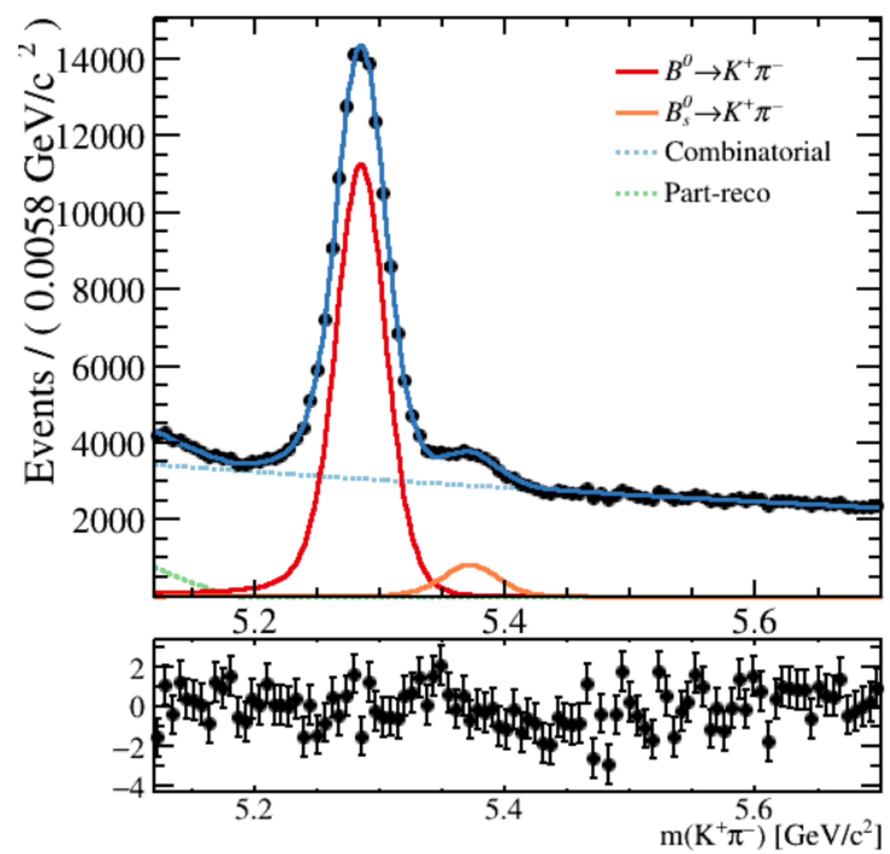
Christina Agapopoulou (IJCLab/CNRS)

On today’s menu: Fits

Why we fit

- We want to say something about our data… this can be:
  - Extract a parameter that tells us something about our theory (SM)
  - Discover a new particle
  - Precisely test the Standard Model
- How do we do it?
  - Find a discriminating variable that gives access to the information that we want : typically the inv. mass, but can also be angles, BDT scores etc
  - Create a model that describes the data. For this we need:
    - A model for our signal
    - Models for our backgrounds
  - Extract parameters of interest and their uncertainties





Finding the model

Ideally we can guess the model from the underlying physics. For example:

- Particles : resonance + some smearing (loss from Bremsstrahlung, detector resolutions etc): Gaussian core + power-law tails -> Crystal-ball (can be one or two-sided)
- Energy loss of charged particle transversing some material: Landau function
- But there's not always a 1-1 correspondence of our process to an analytic function
  - The reality is that some trial-and-error may be required
  - Empiric functions and kernel estimators are very useful tools
  - Alternative models can be considered as cross-checks / syst. Uncertainties

estimators…

- You already covered estimators in your statistics lecture this morning

- Some very nice ones are: ML and χ^2 , but which one should we be using?

Maximum Likelihood or χ^2 – What should you use?

- χ^2 fit is fastest, easiest
  - Works fine at high statistics
  - Gives absolute goodness-of-fit indication
  - Make (incorrect) Gaussian error assumption on low statistics bins
  - Has bias proportional to 1/N
  - Misses information with feature size < bin size

- Full Maximum Likelihood estimators most robust
  - No Gaussian assumption made at low statistics
  - No information lost due to binning
  - Gives best error of all methods (especially at low statistics)
  - No intrinsic goodness-of-fit measure, i.e. no way to tell if ‘best’ is actually ‘pretty bad’
  - Has bias proportional to 1/N
  - Can be computationally expensive for large N

- Binned Maximum Likelihood in between

$$-\ln L(p)_{\text{binned}} = \sum_{\text{bins}} n_{\text{bin}} \ln F(\vec{x}_{\text{bin-center}}; \vec{p})$$

  - Much faster than full Maximum Likihood
  - Correct Poisson treatment of low statistics bins
  - Misses information with feature size < bin size
  - Has bias proportional to 1/N

Wouter Verkerke, UCSB

RooFit basics

RooFit core design philosophy

- Mathematical objects are represented as C++ objects

Mathematical concept		RooFit class
variable	x	`RooRealVar`
function	$f(x)$	`RooAbsReal`
PDF	$f(x)$	`RooAbsPdf`
space point	$\vec{x}$	`RooArgSet`
integral	$\int_{x_{min}}^{x_{max}} f(x)dx$	`RooRealIntegral`
list of space points		`RooAbsData`

Wouter Verkerke, NIKHEF

From https://root.cern/download/roofit-strasbourg-v10.pdf

Minimisation

- Let's say we've chosen our estimator - we then need to find the best (minimum) value

(Default estimator is ML in RooFit)

Minuit

Watch for correlated parameters, find ways to avoid them

E.g.: instead of using two correlated parameters, take the 1st one and the ratio of the two.

Fix some of them (iteratively) or constrain allowed ranges / starting values

A brief description of MINUIT functionality

https://web2.ba.infn.it/~pompili/teaching/data_analysis_lab/Verkerke-RooFit-part2.pdf

- MIGRAD
  - Find function minimum. Calculates function gradient, follow to (local) minimum, recalculate gradient, iterate until minimum found
    - To see what MIGRAD does, it is very instructive to do RooMinuit::setVerbose(1). It will print a line for each step through parameter space
  - Number of function calls required depends greatly on number of floating parameters, distance from function minimum and shape of function

Beware of local minima: starting values might matter

- HESSE
  - Calculation of error matrix from 2nd derivatives at minimum
  - Gives symmetric error. Valid in assumption that likelihood is (locally parabolic)

$$\hat{\sigma}(p)^2 = \hat{V}(p) = \left(\frac{d^2 \ln L}{d^2 p}\right)^{-1}$$

  - Requires roughly N^2 likelihood evaluations (with N = number of floating parameters)

Good approximation for large number of events

Wouter Verkerke. NIKHEF

- MINOS
  - Calculate errors by explicit finding points (or contour for >1D) where Δ-log(L)=0.5
  - Reported errors can be asymmetric
  - Can be very expensive in with large number of floating parameters

Useful in low-stats case

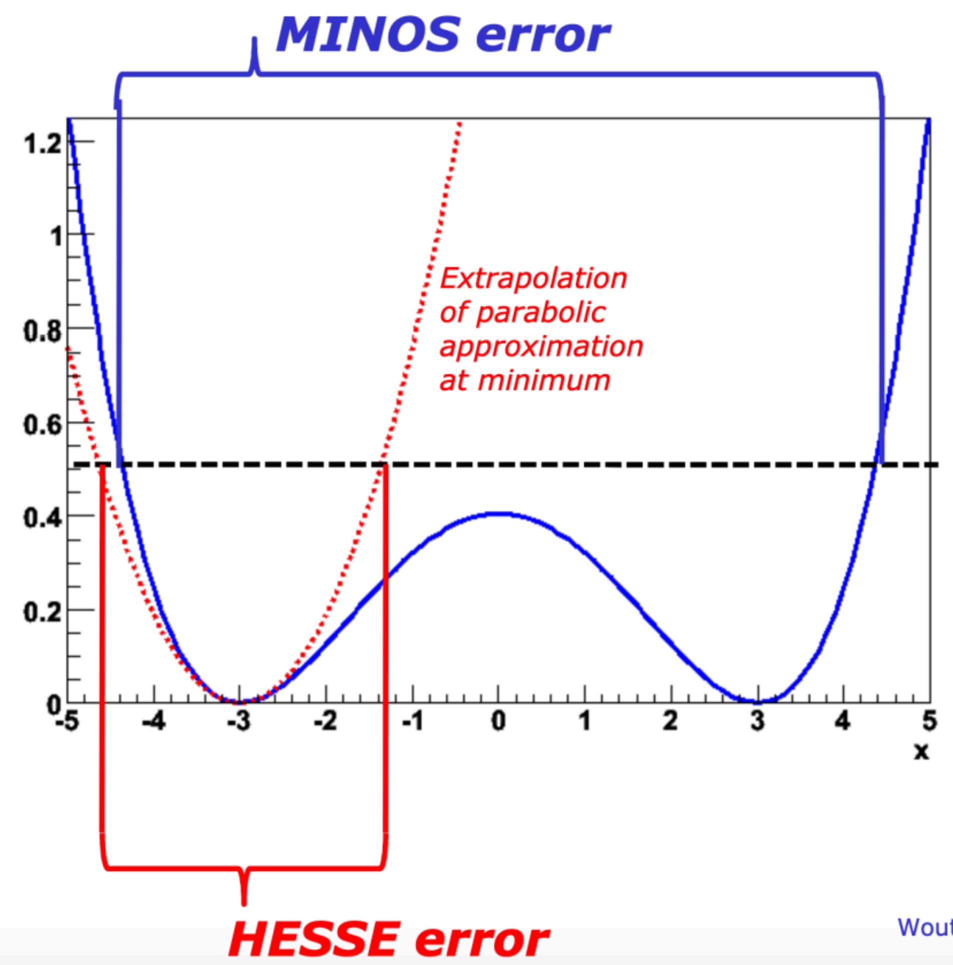
Minimisation

Minuit

https://web2.ba.infn.it/~pompili/teaching/data_analysis_lab/Verkerke-RooFit-part2.pdf

Illustration of difference between HESSE and MINOS errors

- 'Pathological' example likelihood with multiple minima and non-parabolic behavior



Wouter Verkerke, NIKHEF

Enough talking, let’s get coding!