87944 - STATISTICAL DATA ANALYSIS FOR NUCLEAR AND SUBNUCLEAR PHYSICS

Module 3 : Laboratory of Stat. Data Analysis for Nucl. and SubNucl. Physics
teacher: G. Sirri

Hands-on 1 : RooFit

To be submitted in one week

DOCUMENTATION:

- slides shown during the lecture, available on VIRTUALE:
- RooFit website: https://root.cern/manual/roofit/
- RooFit Manual (PDF A4 format)
- RooFit Quick Start Guide (PDF A4 format)
- RooFit tutorials: https://root.cern/doc/master/group__tutorial__roofit.html

Download

roofit_empty.cpp	an empty macro
roofit_empty.ipynb	an empty notebook
B0sInvariantMass.root	invariant mass spectrum for a B0s meson (from LHCb)

[0] WARM UP

Download and run the tutorial **rf101**, **rf201**, and **rf202**
https://root.cern/doc/master/group__tutorial__roofit.html

[1] Hands-on: Basic functionality: fitting, plotting, toy data generation on one-dimensional PDFs. (BASIC)

CHOOSE one exercise among the following two

Exercise 2020.1 – fit a model to unbinned dataset

Edit the macro **roofit_empty** and, following the comments inside, create a Crystal Ball p.d.f. (instead of the Gaussian) with mean = 0, sigma = 1, alpha = 1.5, n = 1.5. Change the sigma to 0.3. Visualize the p.d.f. . Generate an **unbinned** dataset of 10000 events. Make a Fit with Maximum Likelihood. Visualize the results.

Tips:

- *Use information from the slides shown during the lecture or from RooFit Manual at par 2*
- *Refer to the tutorial rf101_basics.cxx*
- *Find the definition of RooCBshape in the ROOT Class reference*

(Submit the macro and the image of the canvas)

Exercise 2020.2 – fit a model to binned dataset

Edit the macro **roofit_empty** and, following the comments inside, create an Exponential p.d.f. to with rate = -1/tau where tau = 3 is the mean life. Visualize the p.d.f. . Generate a **binned** dataset of 1000 events **(bin width = 0.5)**. Make a Fit with Maximum Likelihood. Visualize the results.

Tips:

- *Define the mean life as a RooRealVar and express the exponential rate using RooFormulaVar*
- *The binning of the returned RooDataHist is controlled by the default binning associated with the observables generated. To set the number of bins in x to 200, do e.g. x.setBins(200) prior to the call to generateBinned()*

(Submit the macro and the image of the canvas)

[2] Hands-on: Basic functionality: fit, plot, toy data generation on one-dimensional PDFs. (ADVANCED)

CHOOSE one exercise among the following two

Exercise 12.2 – import a binned dataset, create a model fit the model (LHCb)

One of the main objectives of the LHCb experiment is the study of the CP violation through the decay of different particles, like, for example, the b-flavored mesons.

Due to the short half-life, these particles can be observed by reconstructing their decay products and analyzing the so-called "invariant mass spectrum".

The mass of the mesons is estimated by fitting a Breit-Wigner distribution.

In "B0sInvariantMass.root" an example of invariant mass spectrum for a B0s meson is stored.

Create a macro to open the "B0sInvariantMass.root" and import the corresponding binned dataset. Create a Breit-Wigner model. Fit the model to the binned dataset. Create a Gaussian function and fit to the data. Finally plot the data, and the BW and Gaussian distribution to the same canvas.

Compare the fitted value with the particle mass reported in the Particle Data Group.

Tips: You can see how to import data here: https://root.cern.ch/doc/master/rf102__dataimport_8C.html

(submit the macro and the image of the canvas)

Exercise 16.1 - The Central Limit Theorem - Sum of random variables

Using RooFit, define N=8 independent random variables $x_1, x_2, x_3, \ldots$ uniformly distributed between zero and one. Then, define a new variable as the sum of the N random variables $xsum = \sum x_i$

Generate an unbinned dataset of 10000 events of variables $x_1, x_2, x_3, \ldots$.

You may be interested in the estimation of the expected mean and standard deviation of $xsum$ (the sum of N measurements with a uniform distribution). It's up to you to calculate it on the paper (it's not expected you submit the solution).

> use the formulas $\sigma = \sqrt{V(x)} = \sqrt{\overline{x^2} - \bar{x}^2}$, $\bar{x} = \int x\, F(x)dx$ and $\overline{x^2} = \int x^2 F(x)dx$
> where $F(x)$ is the distribution you are averaging over (For this case where $F(x)$ is a uniform distribution in range [0,1])

The Central Limit Theorem predicts that the sum of N measurements has a Gaussian distribution in the limit of N→∞ independent of the distribution of each individual measurement.

Then, create a Gaussian p.d.f. for $xsum$ with mean and sigma as calculated before.
Make a Fit with Maximum Likelihood. Visualize the results.

Tips:

- the product of random variable is distributed according a p.d.f. which is the product of the single p.d.f.s (use `RooProdPdf`)

- the variable "sum of random variable" can be defined in RooFit by adding a formula to the generated dataset, as shown in this example for 2 variables:

```
// Construct formula to calculate the sum events
RooFormulaVar fsum{"xsum", "var1+var2 ", RooArgList{var1, var2} ;

// Add column with variable xsum to previously generated dataset
auto xsum = (RooRealVar*)data->addColumn(fsum);
```

- define a Gaussian model for xsum (sum of variables). With this respect, xsum behaves exactly as any other RooRealVar, even defined using RooFormulaVar.

- Range and binning are property of a RooRealVar. Given that xsum is defined using a RooFormulaVar and not with the RooRealVar class constructor, you have to explicitly specify a range and a binning to obtain a frame from xsum

```
auto plot = xsum->frame(Bins(40), Title("Sum of Random variables"), Range(0., 6.));
```

- Don't' forget to adjust the range...

(Submit the macro and the image of the canvas)

[3] Hands-on: composite pdf with signal and background component setting up an extended maximum likelihood fit , one-dimensional numeric convolution (BASIC)

Exercise 13.1 - Composite Model: Signal + Background (Non-extended)

Edit the macro **roofit_empty**, and following the outline, build a Gaussian p.d.f. with mean 0 and sigma 3. Called this Gaussian p.d.f. "sig" and add an exponential background component "bkg" to the model, expressed as a function of the parameter tau, exp(-x/tau). Set the initial value of tau to 10.

Define a parameter "fsig" to represent the signal-to-background ratio. Build a composite model in the following form:

$model(x)=fsig\times sig(x)+(1-fsig)\times bkg(x)$

Visualize the p.d.f. . Generate a **binned** dataset of 1000 events (bin width = 0.5).

Make a Fit with Maximum Likelihood. Visualize the results.

Tips:
- Use RooFormulaVar to express -1./tau.
- Use RooAddPdf for the composite model
- Use the RooAbsPdf method generateBinned(…) to generate a binned data set.

(Submit the macro and the image of the canvas)

Exercise 13.2 - Making an extended ML fit

Rewrite the Exercise 13.1 so that it construct a pdf suitable for extended ML fitting.
Multiply the signal pdf by Nsig (200 events, range 0,10000) and the background pdf by Nbkg (800 events, range 0,10000)

(Submit the macro and the image of the canvas)

Hands-on activities continue on the next page…

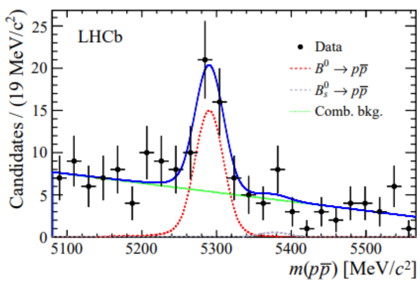
[4] Hands-on: composite pdf with signal and background component setting up an extended maximum likelihood fit , one-dimensional numeric convolution (ADVANCED)

Choose **one** exercise between:
- first observation of the rare purely baryonic decay B^0→p p̄ by LHCb
- ARIADNE Liquid Argon Time Projection Chamber

First observation of the rare purely baryonic decay $B^0 \to p\,\bar{p}$ by LHCb composite model

Inspired by Figure 1 of: "First observation of the rare purely baryonic decay B^0 $\to p\,\bar{p}$" arXiv:1709.01156v2 [hep-ex] 6 Dec 2017; https://arxiv.org/abs/1709.01156

See also https://cerncourier.com/a/the-rarest-b0-decay-ever-observed/



In 2017 the LHCb collaboration has observed the rare baryonic decay $B^0 \to p\,\bar{p}$. The branching fraction was measured at the level of about 1.3 per 100 million decays, which makes this decay mode the rarest decay of a B0 meson ever observed. It is also the rarest observed hadronic decay of all beauty mesons.

The search for the rare decays $B^0 \to p\,\bar{p}$ and $B_s^0 \to p\,\bar{p}$ had previously been performed by LHCb with the full 3 fb–1 data sample collected during the first run of the LHC. An excess of $B^0 \to p\,\bar{p}$ candidates with respect to the background-only hypothesis is observed with a statistical significance of 5.3 standard deviations. The hint of a $B_s^0 \to p\,\bar{p}$ signal reported in 2013 is, however, not confirmed, and an upper limit for the corresponding branching fraction has been set. The measured $B^0 \to p\,\bar{p}$ and $B_s^0 \to p\,\bar{p}$ branching fractions are compatible with the latest theoretical calculations

The exercise aims to reproduce the Invariant Mass Distribution Figure 1

Download
rarest_b0_decay.dat dataset collected by a B-meson experiment

Load the **unbinned** dataset from the file rarest_b0_decay.dat
*Tip: RooDataSet data = *RooDataSet::read("rarest_b0_decay.dat", x, "v");*

Using RooFit, define an extended composite model for invariant mass.
The model components are:
i) a background (you're free to choose the model: flat, polynomial, exponential, …);
ii) a Gaussian peak around the B^0 mass,
iii) a Gaussian peak around the B_s^0 mass

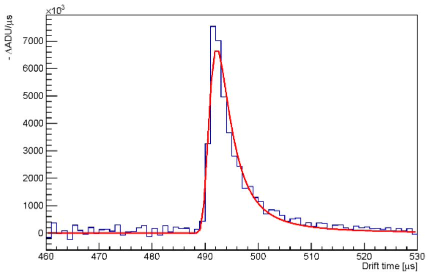
Fit the model to the data using a maximum likelihood fit. Plot data and model. Superimpose each single component with different color
Tip: use the named functions RooFit::Components(…) and RooFit::LineColor(…).

Make histogram of residual and pull distributions
Have a look to https://root.cern.ch/doc/master/rf109__chi2residpull_8C.html

Note: methods residHist(…) and pull(..) by default compute the residuals (pulls) of the latest-plotted histogram with respect to the latest-plotted curve.
-> Construct a histogram with the residuals of the data w.r.t. the curve
-> Construct a histogram with the pulls of the data w.r.t the curve
-> Create a new frame to draw the residual distribution and add the distribution to the frame
-> Create a new frame to draw the pull distribution and add the distribution to the frame

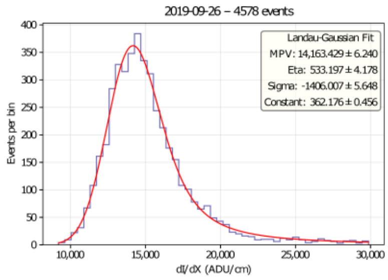
(submit source code, plots)

ARIADNE Liquid Argon Time Projection Chamber
composite model, Convolution



Inspired by Figure 6 (right) "The negative gradient of the summed ToT distribution." and by Figure 12 "The distribution of the lifetime-corrected ToT summation for a population of through-going muon tracks." of "Optical Readout of the ARIADNE LArTPC Using a Timepix3-Based Camera" Instruments 2020, 4(4), 35; https://doi.org/10.3390/instruments4040035

[*https://www.mdpi.com/2410-390X/4/4/35#*](https://www.mdpi.com/2410-390X/4/4/35#)



Future Liquid Argon Time Projection Chambers in the neutrino sector will be able to reach the kiloton-scale—for example, four 17,000 ton LArTPCs have been proposed for use on the DUNE project. Given the high construction and operating costs, as well as the sheer complexity of such large detectors, early and innovative R&D therefore has the potential for a large return on investment over an experiment's lifetime. The ARIADNE (ARgon ImAging DetectioN chambEr) Experiment is based around a 1-ton dual-phase LArTPC. Using this detector, the project aims to demonstrate the feasibility of optical readout of LArTPCs on a large scale and develop an ongoing program for the characterisation and maturation of such technology.

Part 1: COMPOSITE MODEL

This exercise aims to reproduce the Figure 6 representing negative gradient of the summed Time Over Threshold distribution as a function of the drift time.

- Define the drift time as the observable.
  Take the range, binning and units from the final plot.

- Download
  ariadne_g006_plus_400.dat binned dataset taken from figure 6
  (with an additional pedestal of 400 counts)

- As described in the paper data are distributed according to a Landau peak
  Make a composite UNIFORM + LANDAU model. Make a non-extended model but let the fraction of signal having a range larger than usual (say -0.5 < fs < 1.5) to let background underfluctuate.

- Read the **binned** dataset and fill a RooDataHist:

```
// Example code for c++
// (x is the RooRealVar observable):
    RooDataHist data{"data", "data", x};
    ifstream file("filename.txt");
    double val, weight;
    while (!file.eof()) {
          file >> val >> weight;
          x.setVal(val);
          data.set(x weight);   // (*)
    }
```

> *Hint:*
> *Range and binning of the observable shall be defined in such a way to match the histogram of the input file*
>
> *Use: RooRealVar::setBins(…)*
>
> *<u>(*) Don't import underflow (overflow) bin</u>*

- Fit the model to the data.
  Save the fit results to an object RooFitResults.
  Plot the data and the fitted model.
  Plot also the signal component in red and the bkg component in black.
  Write the best fit parameters on the plot too. (hint: paramOn(…) method of RooAbsPdf).

(submit source code, plots and a text file with the fit results)

PART 2 CONVOLUTION

This exercise aims to reproduce the Figure 12 representing the distribution of the lifetime-corrected Time over Threshold (ToT) summation for a population of through-going muon tracks.

- Define the time as the observable.
  Take the range, binning and units from the final plot.

- Download
  ariadne_g012.dat binned dataset taken from figure 12

- As described in the paper a Landau-Gaussian convolution function has been fitted to the entire distribution.
  Make a model where a LANDAU (the physics) is convolved with the GAUSSIAN (the detector resolution)

- Read the **binned** dataset and fill a RooDataHist:

```
// Example code for c++
// (x is the RooRealVar observable):
     RooDataHist data{"data", "data", x};
     ifstream file("filename.txt");
     double val, weight;
     while (!file.eof()) {
          file >> val >> weight;
          x.setVal(val);
          data.set(x weight);  // (*)
     }
```

> *Hint:*
> *Range and binning of the observable shall be defined in such a way to match the histogram of the input file*
>
> *Use: RooRealVar::setBins(…)*
>
> <u>*(*) Don't import underflow (overflow) bin*</u>

- Fit the model to the data.
  Save the fit results to an object RooFitResults.
  Plot the data and the fitted model.
  Write the best fit parameters on the plot too. (hint: paramOn(…) method of RooAbsPdf).

(submit source code, plots and a text file with the fit results)