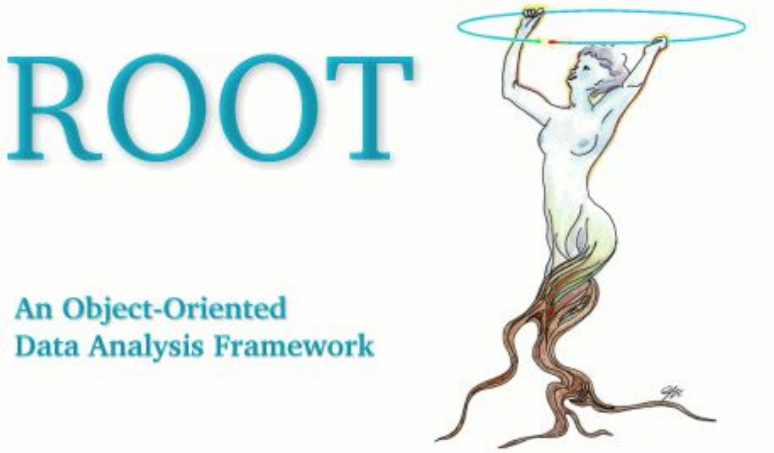

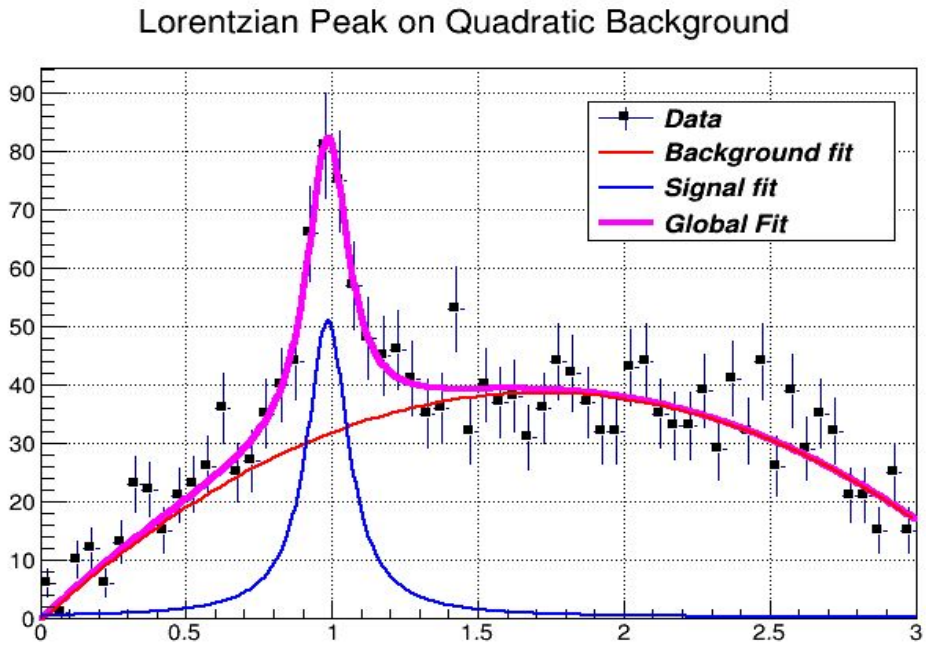

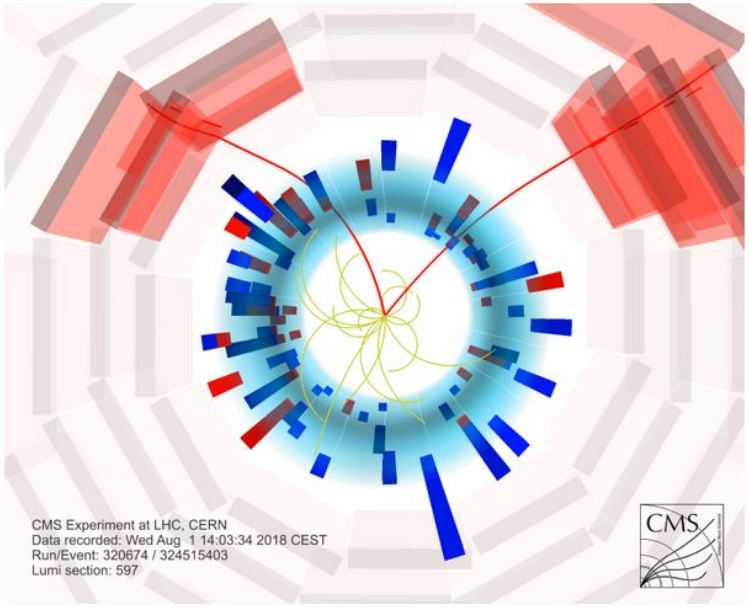

Big data en el CERN y otros contextos

Jhovanny Andres Mejia Guisao
UNIVERSIDAD DE ANTIOQUIA, COLOMBIA

Motivation

What we hope to discuss about scientific data analysis?

- Advanced graphical user interface
- Interpreter for the C++ programming language
- Persistency mechanism for C++ objects
- Used to write every year petabytes of data recorded by the Large Hadron Collider experiments

Input and plotting of data from measurements and fitting of analytical functions.

Credit

RooFit slides and example are extracted from material prepared by W. Verkerke (NIKHEF), author of RooFit

– more information and additional slides from W. Verkerke are available at

- http://indico.in2p3.fr/getFile.py/access?contribId=15&resId=0&materialId=slides&confId=750

Roofit manual

https://root.cern/download/doc/RooFit_Users_Manual_2.91-33.pdf

Purpose

The RooFit library provides a toolkit for modeling the expected distribution of events in a physics analysis. Models can be used to perform unbinned maximum likelihood fits, produce plots, and generate "toy Monte Carlo" samples for various studies. RooFit was originally developed for the BaBar collaboration, a particle physics experiment at the Stanford Linear Accelerator Center. The software is primarily designed as a particle physics data analysis tool, but its general nature and open architecture make it useful for other types of data analysis also.

Mathematical model

The core functionality of RooFit is to enable the modeling of ‘event data’ distributions, where each event is a discrete occurrence in time, and has one or more measured observables associated with it. Experiments of this nature result in datasets obeying Poisson (or binomial) statistics. The natural modeling language for such distributions are probability density functions F(x;p) that describe the probability density the distribution of observables x in terms of function in parameter p.

The defining properties of probability density functions, unit normalization with respect to all observables and positive definiteness, also provide important benefits for the design of a structured modeling language: p.d.f.s are easily added with intuitive interpretation of fraction coefficients, they allow construction of higher dimensional p.d.f.s out of lower dimensional building block with an
intuitive language to introduce and describe correlations between observables, they allow the universal implementation of toy Monte Carlo sampling techniques, and are of course an prerequisite for the use of (unbinned) maximum likelihood parameter estimation technique.

Roofit manual

https://root.cern/download/doc/RooFit_Users_Manual_2.91-33.pdf

Design

RooFit introduces a granular structure in its mapping of mathematical data models components to C++ objects: rather than aiming at a monolithic entity that describes a data model, each math symbol is presented by a separate object. A feature of this design philosophy is that all RooFit models always consist of multiple objects. For example a Gaussian probability density function consists typically of four objects, three objects representing the observable, the mean and the sigma parameters, and one object representing a Gaussian probability density function. Similarly, model building operations such as addition, multiplication, integration are represented by separate operator objects and make the modeling language scale well to models of arbitrary complexity.

Math concept	*Math symbol*	*RooFit (base)class*
Variable	x	`RooRealVar`
Function	$f(x)$	`RooAbsReal`
P.D.F.	$F(x;p)$	`RooAbsPdf`
Integral	$\int_{x_{min}}^{x_{max}} f(x)dx$	`RooRealIntegral`
Space point	$\vec{x}$	`RooArgSet`
Addition	$fF(x)+(1-f)G(x)$	`RooAddPdf`
Convolution	$f(x)\otimes g(x)$	`RooFFTConvPdf`

Table 1 - Correspondence between selected math concepts and RooFit classes

Scope

RooFit is strictly a data modeling language: It implements classes that represent variables, (probability density) functions, and operators to compose higher level functions, such as a class to construct a likelihood out of a dataset and a probability density function. All classes are instrumented to be fully functional: fitting, plotting and toy event generation works the same way for every p.d.f., regardless of its complexity

RooFit

- ➔ is a library which provides a toolkit for data analysis
- ➔ is included in ROOT framework
- ➔ is used to model expected event distributions in physics analysis
- ➔ can perform (un)binned maximum likelihood fits, produce plots and study goodness-of-fit with toy Monte Carlo samples
- ➔ was originally developed for the BaBar collaboration @ Stanford Linear Accelerator Center

To use RooFit in ROOT CINT
Load library as:

- gSystem->Load("libRooFit") ;
  using namespace RooFit ;

OR

- Load prepared macro file
  .x path-to-file