Bump search & limit calculation

Igal Jaeglé

Thomas Jefferson National Accelerator Facility

1st Workshop on New Light Physics and Photon-beam Experiments

Tuesday March 09, 2021

Outline of this talk

- Introduction
- Prerequisite to a bump search
- Count experiment
- Shape experiment
- Limit calculation
- Conclusion

Introduction

This is not a statistical course but a schematic on how to determine an expected sensitivity or limit

Good tutorials and materials to read and watch:

- http://www.hep.fsu.edu/~harry/teaching.html, H.B. Prosper
- ROOFIT examples, https://root.cern/doc/master/group__tutorial__roofit.html
  - Google: INFNStatRooStats schools
- BAT examples, https://bat.mpp.mpg.de/?page=tutorials&version=0.9
- https://arxiv.org/abs/1705.03578, M. Williams

git clone https://github.com/igjaegle/workshop

Blind analysis

Applied when looking for new physics or particles, It means you are not allowed to look at the signal box i.e.:

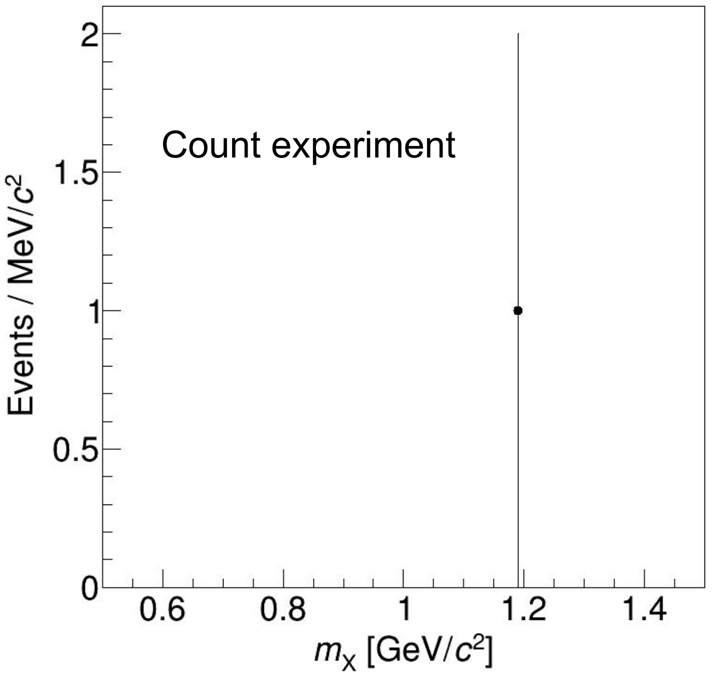
- Selection criteria must be established before opening signal box and cannot be changed once the signal box is opened
- In principle, bump search methodology should also be determined before opening the signal box

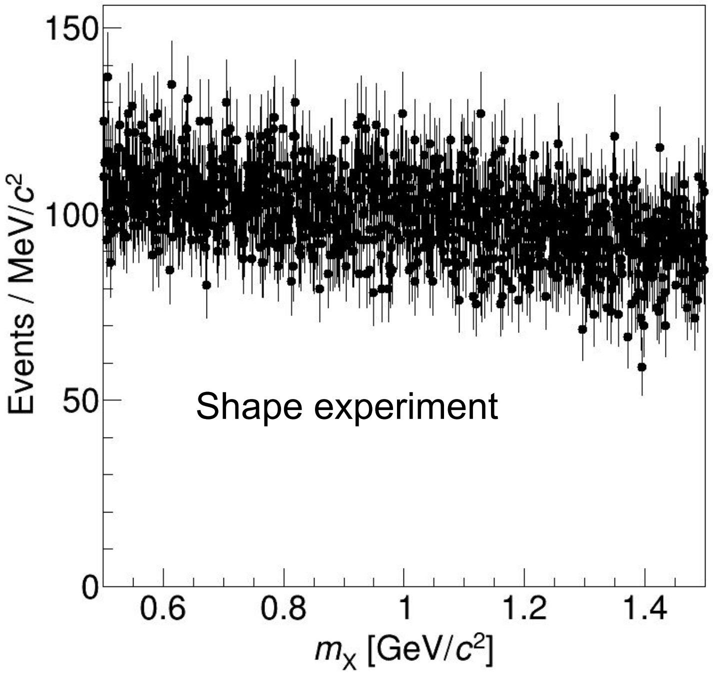
So how to determine selection criteria and bump search methodology:

- Signal MC simulation
- MC simulation sample of expected background or
- Look at very small fraction of data (< 5%), it implies this sample cannot be used in the end
- MVA
- Figure-Of-Merit:
    - Basic: S / \sqrt{B} or
    - More complicated

Count experiment or shape experiment

Let says that you did determine your selection criteria and apply it to your MC simulation or x fraction of data sample. If



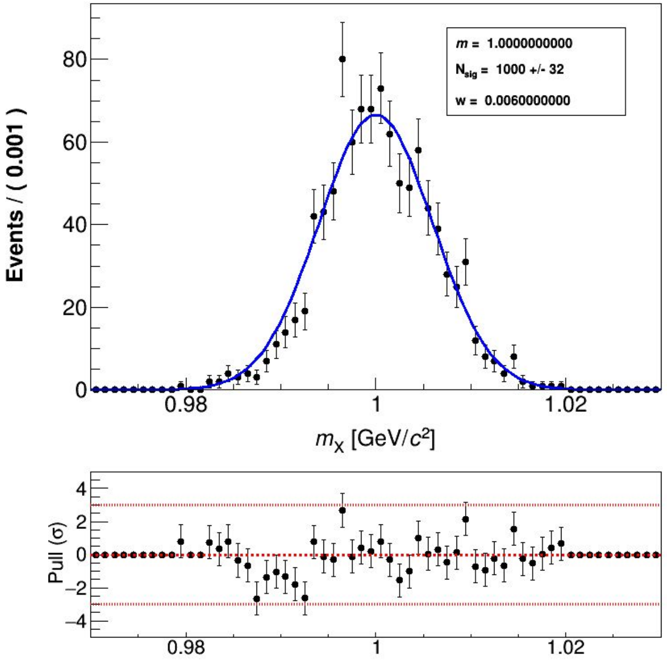


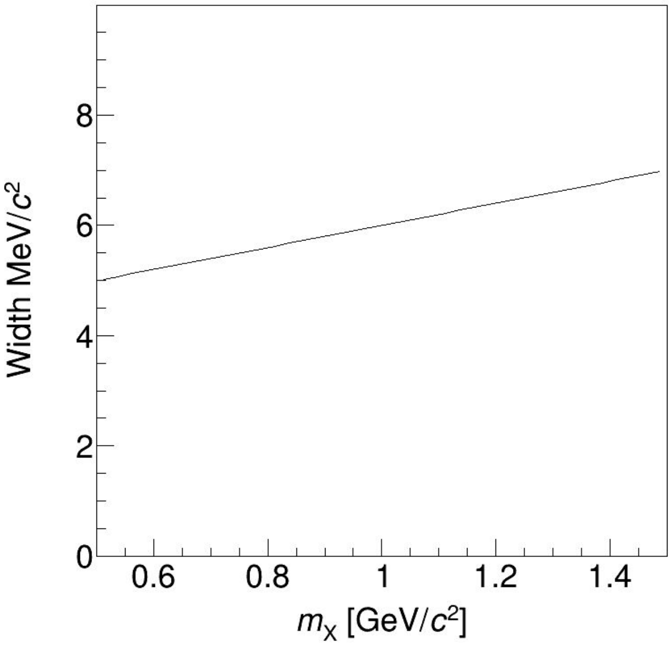
sig_pdf.c

Determination of the signal PDF

Your signal can be a narrow Gaussian-like bump or just be a wide potato shape

If your signal is a Gaussian-like, you can simulate a few mass and parametrize the width vs. mass

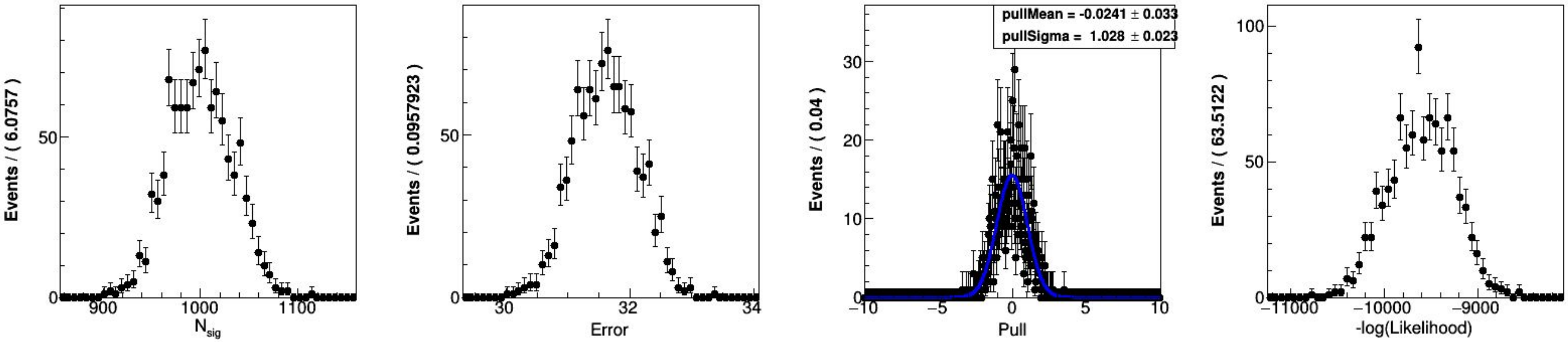




sig_pdf.c

Find sources and estimate systematic error

For example as you only simulate few mass, one need determined how is your parametrization, one way to do so is to use RooMCStudy



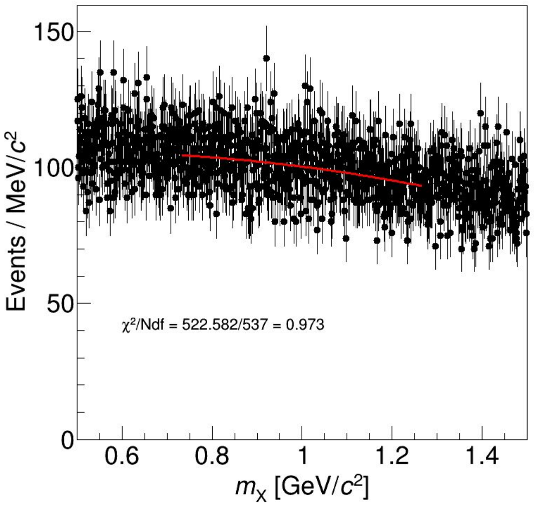
sig_pdf.c

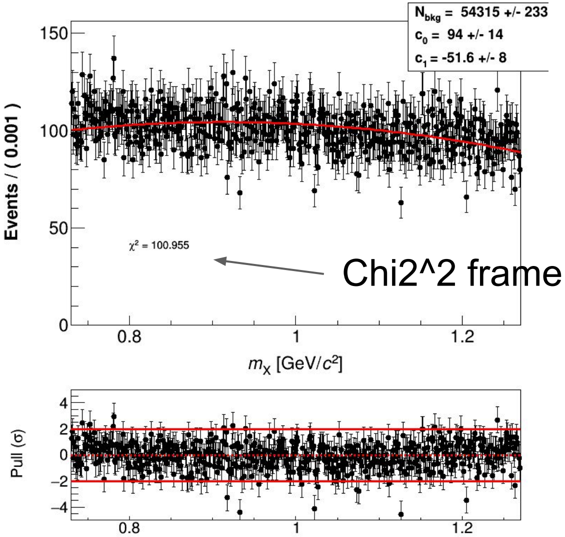
Shape experiment, determination of the background PDF

Either you know the background behavior (i.e. you know which function it is following) or do not know

What to do when you have to guess the background PDF?

- Test Polynomial of xth order and vary the fit range so that Chi^2/ndf -> 1





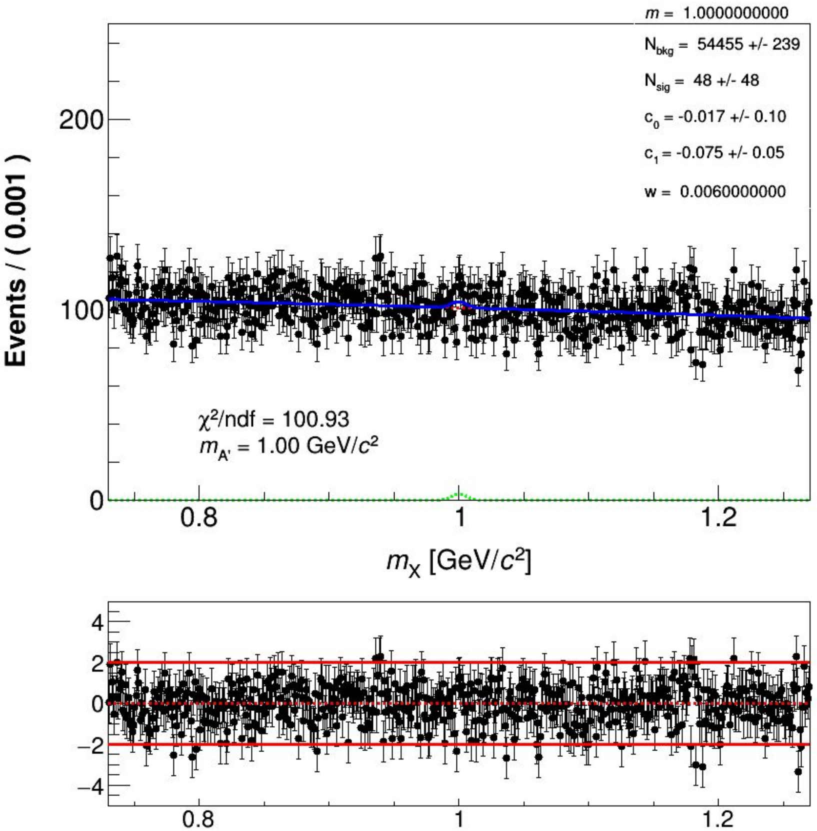
Chi2^2 frame

make_bkg.c

Jefferson Lab

Shape experiment: bkg + sig

binned maximum likelihood fit



limit.c

Shape experiment

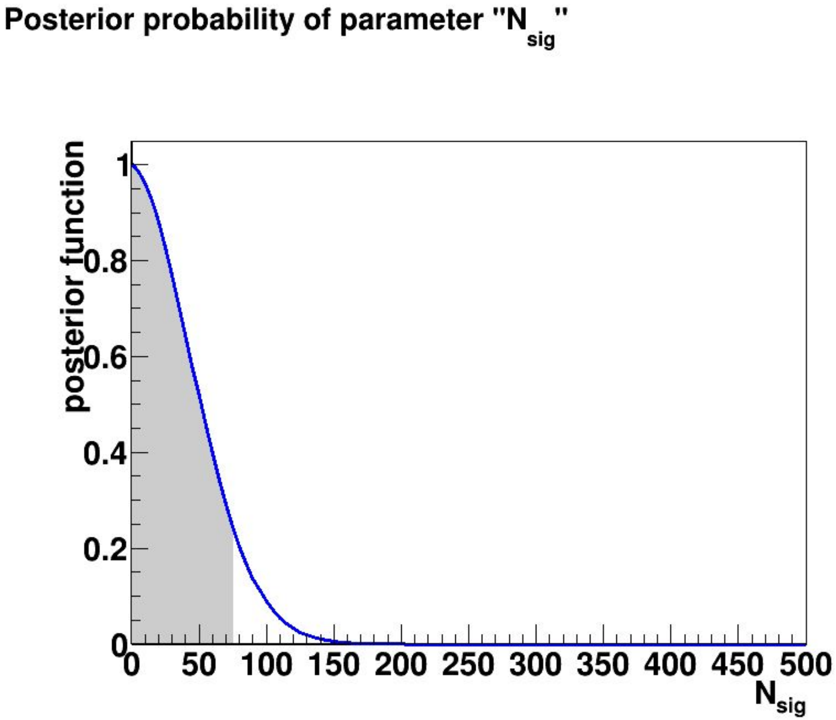
```
w->factory("SUM::pdf(s[0,15]*Uniform(x[0,1]),b[1,0,2]*Uniform(x))");
w->factory("Gaussian::prior_b(b,1,1)");
w->factory("Uniform::priorPOI(s)");
```

SUM -> RooAddPdf
s[] -> variable RooRealVar
Uniform -> Uniform pdf on x (obs), (x is also created
b -> RooRealVar from 0-2 init to 1
b is gaussian constrained to one with sigma 1 (you probably want a PROD pdf at some point)

Factories are just an interface to workspaces,
you can also import stuff if that is simplier to understand.

for instance:
your signal can be a gaussian with mean m,sigma and uncertainty dm
RooGaussian sig(x, RooFormulaVar(m+dm), sigma)

Upper limit on yield extracted



A Bayesian method is used to estimate the 90% upper limit on yield extracted

Yield range start at zero (done after the fit in order not to affect likelihood function)
Likelihood function integrated from zero up to 90% of the curve intersection point

UL_Bayes = 75.4462

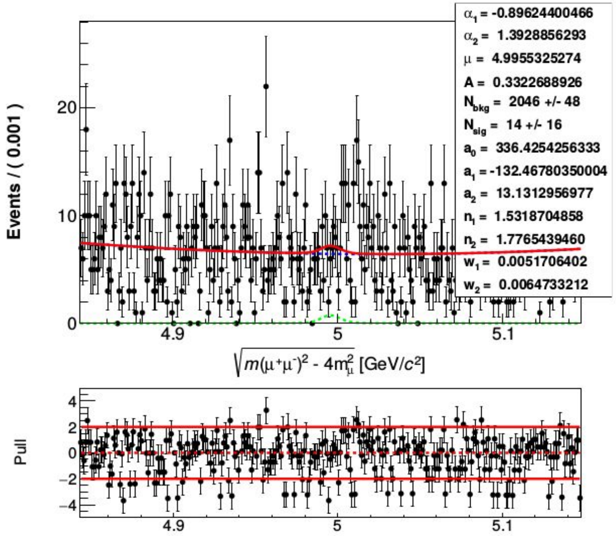
Jefferson Lab

Bias on yield extracted

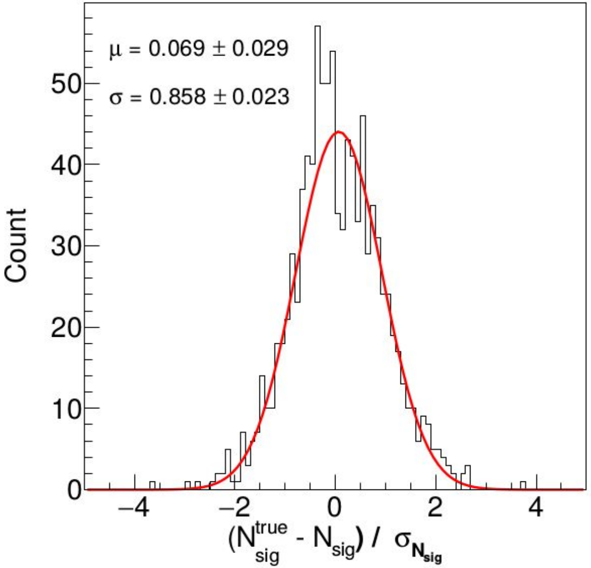
To check if maximum likelihood fit is bias

- Randomly sample Poisson(S+B) events from the combined dataset (with repetitions)
- Fit this sample with floating background and fix signal parameters
- Compare the yield extracted with S, S = 90% UL on yield

- Fit example



- Corresponding comparison for 1000 fit experiments



if (mu < 1.0) mu = 1 - mu
if (mu >= 1.0) mu = mu - 1

UL = N_{UL} + mu x sigma_{N_sig,observed} x sigma