NEW YORK UNIVERSITY

RooFit/RooStats Tutorial

Statistics School
Hamburg
April 2-5, 2012

Sven Kreiss (NYU)
Lorenzo Moneta (CERN)

A (NOT SO) SIMPLE MODEL

- ModelConfig: GlobalObservables, Snapshots
- StandardProfileLikelihoodDemo.C

Nomenclature

Parameters: POI vs Nuisance Parameters

- ➡ POI are never profiled or marginalized. Each POI has one value. You can collect results for different values of the POI and make a plot.

Frequentist: profiling nuisance parameters to the observed data

Hybrid: frequentist test, but marginalizing nuisance parameters
(currently **not** used in CMS nor ATLAS; was used at LEP)

Bayesian: all parameters (POI and Nuisance Parameters) need a prior. The prior on the signal strength parameter is a delicate issue (see http://root.cern.ch/root/html/tutorials/roostats/JeffreysPriorDemo.C.html).

CLs: Frequentist method. Intervals for Gaussians and Poissons are numerically equal to Bayesian intervals with flat Priors. This is an “attractive” feature and makes it possible to cross check a Frequentist result with a Bayesian method, but the intervals mean different things.

The on/off problem (arxiv:0702156v4, HybridInstructional.C and references, ATLAS StatForum recom.)

Often used for counting experiment:
$$P(n_{\mathrm{on}}|s) = \int db\, \mathrm{Pois}(n_{\mathrm{on}}|s+b)\, \pi(b)$$

$\pi(b)$ is a Bayesian prior and cannot be used in pure Frequentist methods. Choice of prior is not easy in most cases.

Alternatively, one can introduce a sideband measurement (which is also how knowledge about b was obtained in reality):
$$\underbrace{P(n_{\mathrm{on}}, n_{\mathrm{off}}|s,b)}_{\text{joint model}} = \underbrace{\mathrm{Pois}(n_{\mathrm{on}}|s+b)}_{\text{main measurement}}\, \underbrace{\mathrm{Pois}(n_{\mathrm{off}}|\tau b)}_{\text{sideband}}$$

➡ Using the Likelihood of a previous measurement propagates all errors properly: do whenever that information is available!
(also see current discussions of “publishing the Likelihood” in ATLAS and CMS)

Using this approach, the knowledge about b is defined in a consistent way.

From the paper by Cousins, Linneman,Tucker:

This HEP prototype problem has an exact analog in gamma ray astronomy (GRA), upon which we base our notational subscripts “on” and “off”. The observation of n_{on} photons when a telescope is pointing at a potential source (“on-source”) includes both background and the source, while the observation of n_{off} photons with the telescope pointing at a source-free direction nearby (“off-source”) is the subsidiary measurement. In both the HEP and GRA examples, we let the parameter τ denote the ratio of the expected means of

Exercise

Build the Likelihood and print the tree.

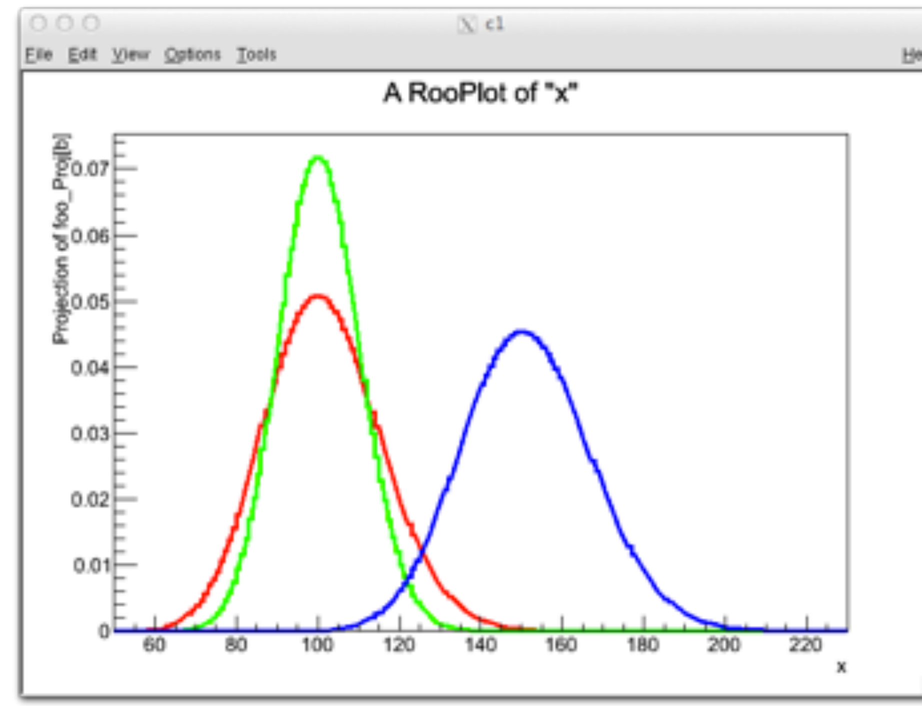
```
w->pdf("model")->Print("t");

0x7fbd338e7800 RooProdPdf::model = 2.59541e-08 [Auto,Dirty]
  0x7fbd338d8a00/V- RooPoisson::px = 6.51116e-07 [Auto,Dirty]
    0x7fbd338b7c00/V- RooRealVar::x = 150
    0x7fbd338ab600/V- RooAddition::splusb = 100 [Auto,Clean]
      0x7fbd338b6e00/V- RooRealVar::s = 0
      0x7fbd338b7400/V- RooRealVar::b = 100
  0x7fbd338d6000/V- RooPoisson::py = 0.039861 [Auto,Dirty]
    0x7fbd338d9600/V- RooRealVar::y = 100
    0x7fbd338d6600/V- RooProduct::taub = 100 [Auto,Clean]
      0x7fbd338b9c00/V- RooRealVar::tau = 1
      0x7fbd338b7400/V- RooRealVar::b = 100
```

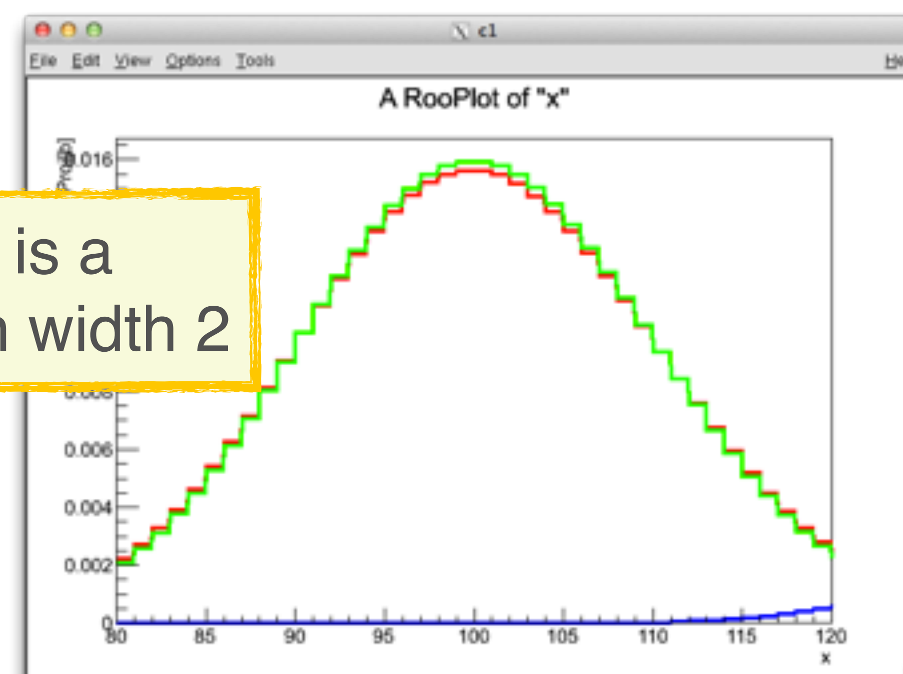
Exercise

Do Bayesian integration using RooFit over b with a uniform prior.
(w->factory("PROJ::averagedModel(PROD::foo(pxlb,pylb,prior_b),b)");)

- ➡ red is b-only averaged model
- ➡ green is b known exactly
- ➡ blue is s+b averaged model



Also do: Gaussian (σ=20) and Lognormal (k=1.2) terms for the sideband.



Sideband is a Gaussian with width 2

Global Observables

So far, the observable from a previous measurement (sideband measurement) was a constant and that is fine for plotting the Profile Likelihood. When generating pseudo experiments, one should take into account that this is not the true value. If the experiment is repeated 1000 times in reality, the observed number of events would fluctuate according to the sideband's Likelihood function. The same has to be done for the pseudo experiments. Those observables need to be "labeled" as Global Observables so that RooFit/RooStats can generate them properly.

Global Observables are not Observables

Observables are generated for every event, Global Observables are generated once per toy. This is easy to understand with an example. The JES input as given by the performance group is a Global Observable. In every toy experiment, it has exactly one value and does not change from one event to the next. On the other hand, an invariant mass which is an Observable (not global) is different for every event.

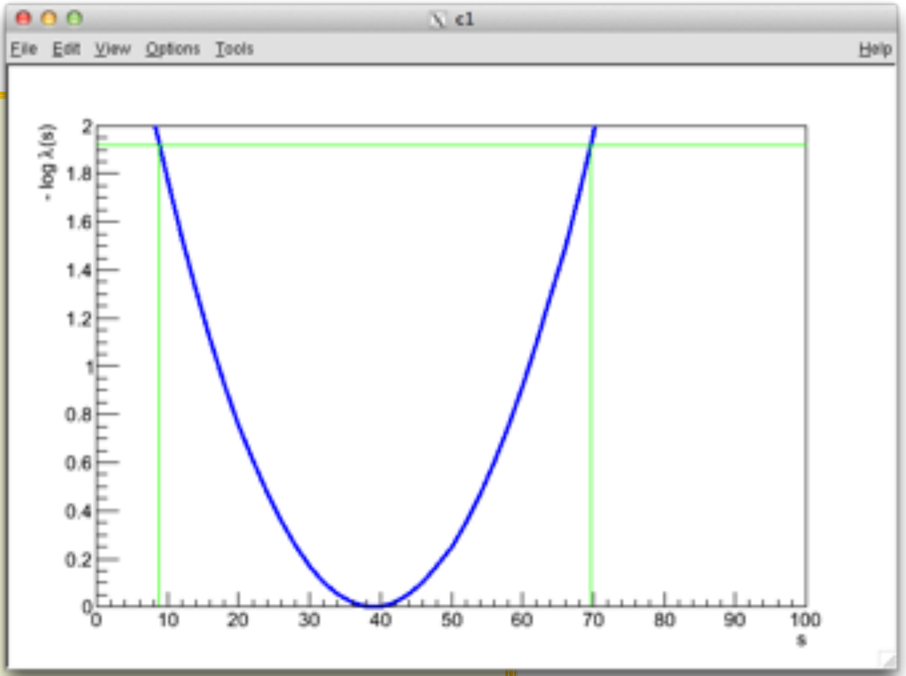
Snapshots

A *RooArgSet* usually just specifies which variables belong to a set, but not what value they have. To fix the values, this is what snapshots are for.

For a S+B model, the “signal strength modifier” μ might be 1 and for the B only model it might be 0. The ModelConfigs for these two cases contain snapshots with μ=1 and μ=0 but are otherwise identical.

Exercise: on/off Problem continued

Generate one fake observed event and call it “obsData”. Make a ModelConfig for the on/off problem. Print it and write it to a root file.




```
DataStore modelData (Generated From model)
  Contains 1 entries
  Observables:
    1)  x = 139  L(0 - 500)  "x"

=== Using the following for onOffProblem ===
Observables:             RooArgSet:: = (x)
Parameters of Interest:  RooArgSet:: = (s)
Nuisance Parameters:     RooArgSet:: = (b)
Global Observables:      RooArgSet:: = (y)
PDF:                     RooProdPdf::model[ px * py ] = 0.000915446
Snapshot:
  1) 0x7f8194229a00 RooRealVar:: s = 40  L(0 - 100)  "s"
```

Run: `root -l '$ROOTSYS/tutorials/roostats/StandardProfileLikelihoodDemo.C("onOffProblem.root","w","onOffProblem","obsData")'`

to produce the plot.

95% interval on s is : [8.78761, 69.6256]

Optional Exercise

Re-do the previous exercise with Gaussian sidebands with a few different widths.

TOOLS

- **Hypothesis Tests: Test Statistics, ToyMCSampler, Detailed Output, Multiple Test Statistics**

Hypothesis Testing

Kendall 2A:

“In general, any hypothesis concerning the generating mechanism for observable random variables is a statistical hypothesis.”

“... no hypothesis can be tested in isolation; there must be at least two competing hypothesis (and we usually consider exactly two) even if one asserts proposition A and the other asserts ‘not-A’.”

Null Hypothesis: the hypothesis that is tested
Alternative Hypothesis: another hypothesis that defines the choice of the critical region

Test Statistic: Maps high dimensional data (points in "observable"-space) to a real number. (source?), Fred James: "Any function of the data is called a statistic."

➡ a complicated shape that defines the boundary between acceptance and critical region gets mapped to a point on a line

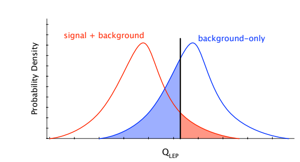
At the LHC, the Profile-Likelihood-Test-Statistic is used.

➡ takes nuisance parameters into account

RooStats has the three common test statistics used in the field (and more)

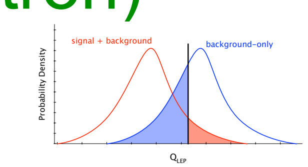
- simple likelihood ratio (used at LEP, nuisance parameters fixed)

$$Q_{LEP} = L_{s+b}(\mu = 1)/L_b(\mu = 0)$$



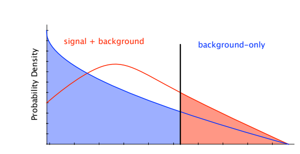
- ratio of profiled likelihoods (used commonly at Tevatron)

$$Q_{TEV} = L_{s+b}(\mu = 1, \hat{\hat{\nu}})/L_b(\mu = 0, \hat{\hat{\nu}}')$$



- profile likelihood ratio (related to Wilks's theorem)

$$\lambda(\mu) = L_{s+b}(\mu, \hat{\hat{\nu}})/L_{s+b}(\hat{\mu}, \hat{\nu})$$



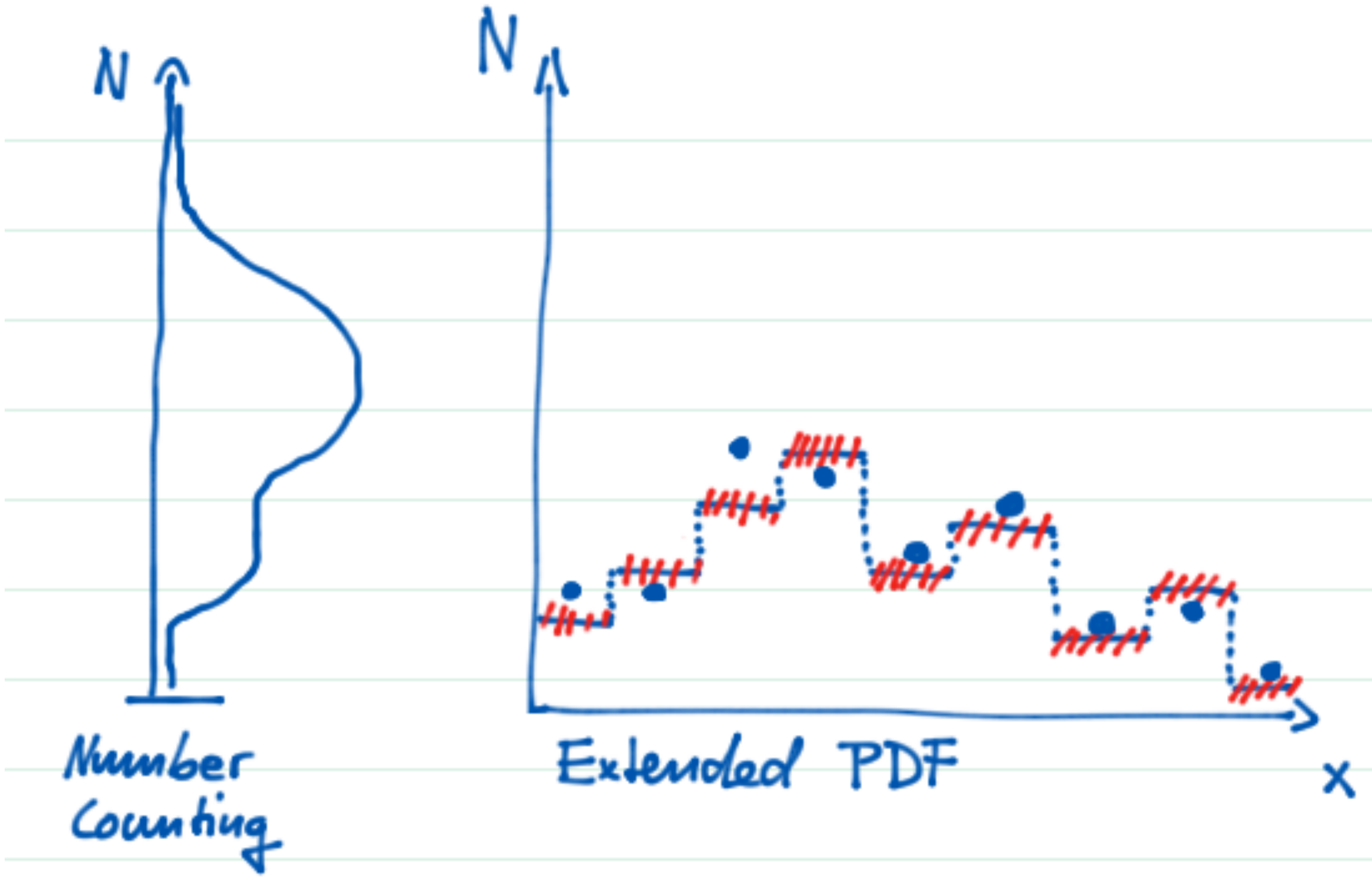
Generating Toys

Toys are generated to get a distribution of possible outcomes of the test statistic values. This distribution defines what is called: “the probability of obtaining data that is at least as discrepant as the observed data”.

ToyMCSampler is at the center of all Tools that use Toys

FrequentistCalculator and HybridCalculator use the ToyMCSampler to generate sampling distributions. The HypoTestInverter can also use these calculators to set limits using toys.

Do not use the ToyMCSampler directly. To get a sampling distribution with profiled nuisance parameters of the Null hypothesis, use the FrequentistCalculator and set the number of toys for the Alternative hypothesis to zero.



FrequentistCalculator

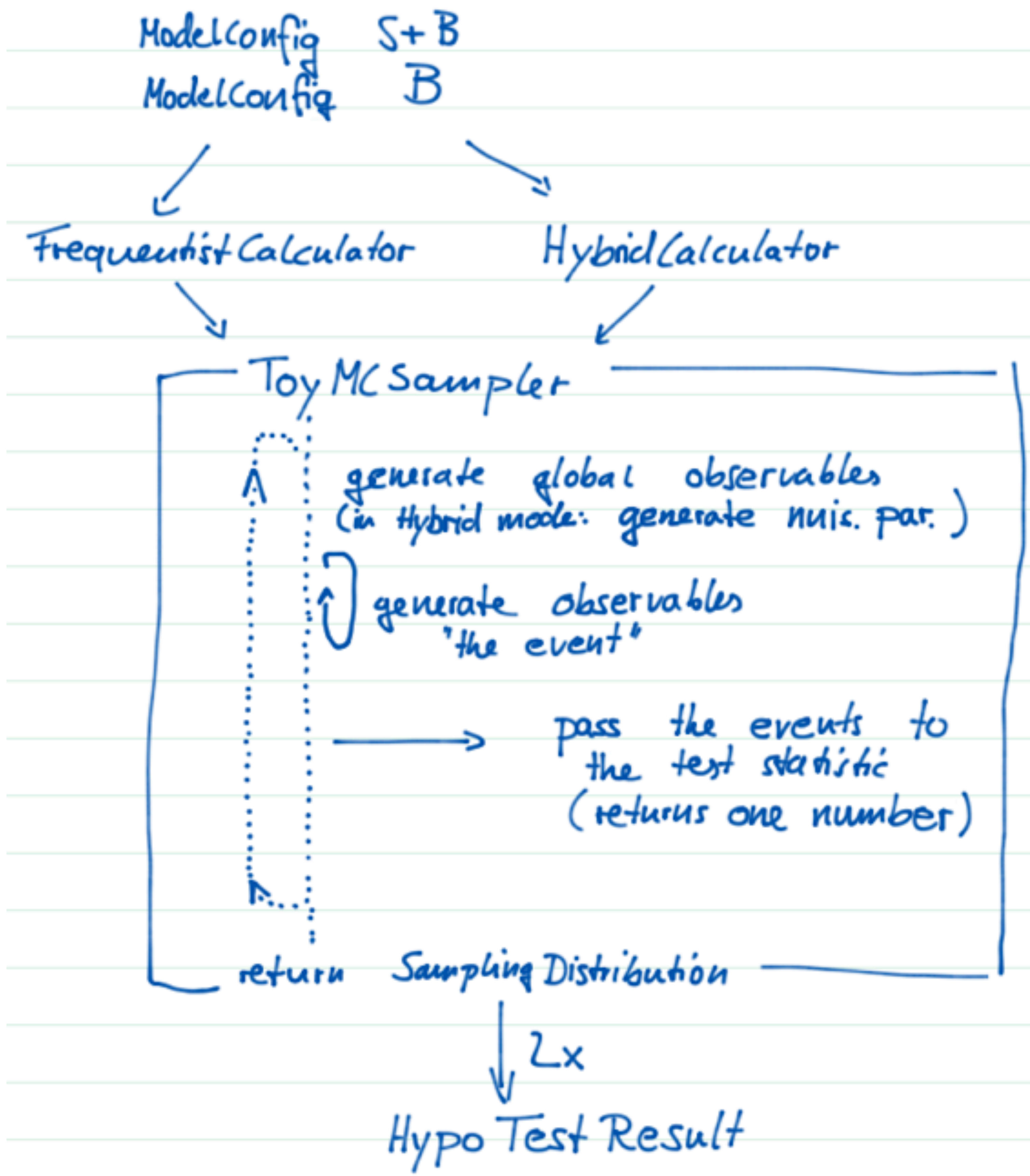
Profiling nuisance Parameters (conditional ensemble):

Nuisance parameters have to be fixed before doing a Frequentist Hypothesis test. Essentially only two possible choices: set them to their nominal values (unconditional ensemble) or do a fit (aka “profile”) to the observed data (conditional ensemble).

➡ pros and cons exist for both choices, but **asymptotic formulae correspond to the conditional ensemble**

Alternatively, it is also possible to integrate instead of profile the nuisance parameters. The integration is only defined when there is a Prior which is an inherit Bayesian object. Because a Bayesian integration is used inside a Frequentist calculation, this is called the Hybrid method and is implemented in the HybridCalculator which has an almost identical interface to the FrequentistCalculator.

Hypotheses Testing in RooStats



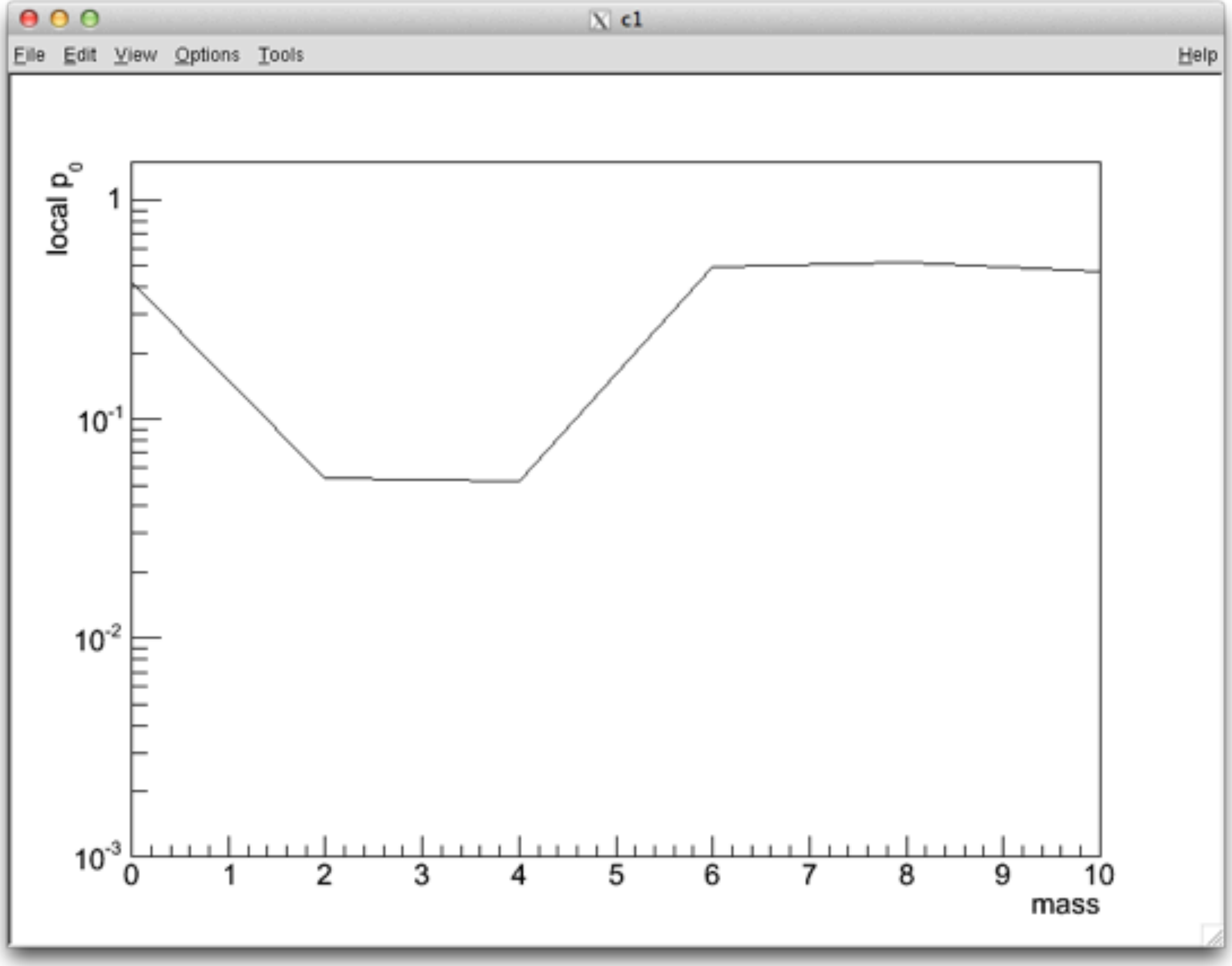
Exercise

Write a macro that reads in a Workspace, ModelConfig and Data and then runs FrequentistCalculator with MaxLikelihoodEstimateTestStat ($\hat{\mu}$), BinCountTestStat, SimpleLikelihoodRatioTestStat, RatioOfProfiledLikelihoodsTestStat and ProfiledLikelihoodTestStat. Each result should be plotted and collected into a PDF file.

Optional: Look at $ROOTSYS/tutorials/roostats/HybridInstructional.C where the on/off Problem is investigated with the HybridCalculator.

Exercise

Now that you have a way to get a p_0-value with toys, put that function into a loop over another parameter and draw a graph of this parameter. If you use this with the previous "falling exponential + Gauss" model and call the mean of the Gaussian m_H and use it as your "other parameter" then you just created a discovery plot with toys for the $H \to \gamma\gamma$ channel.

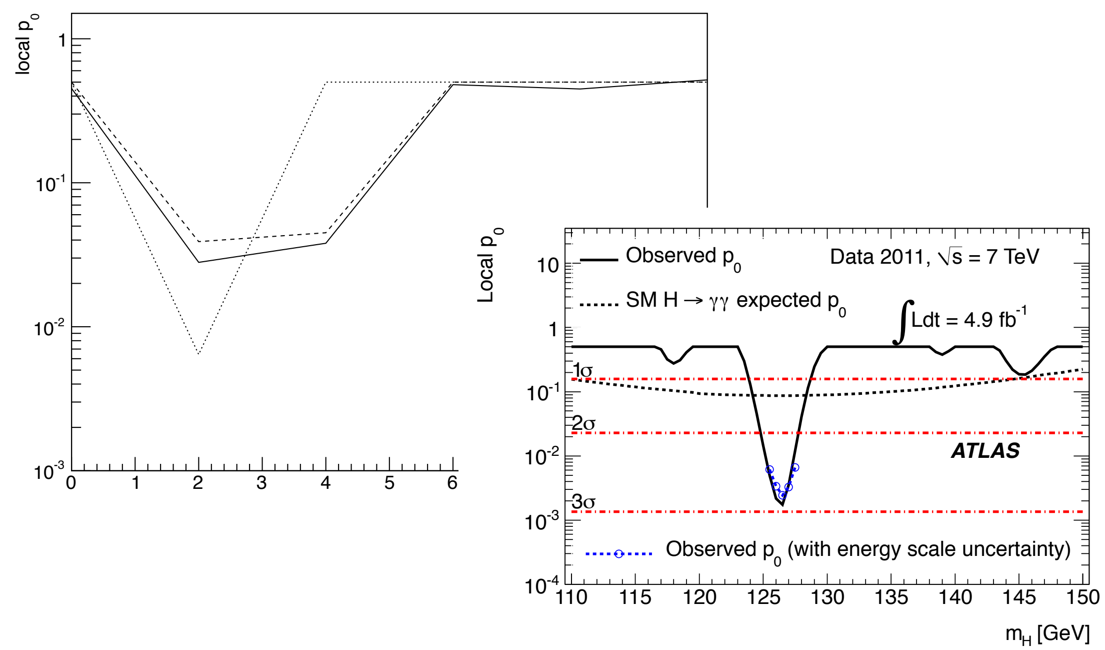


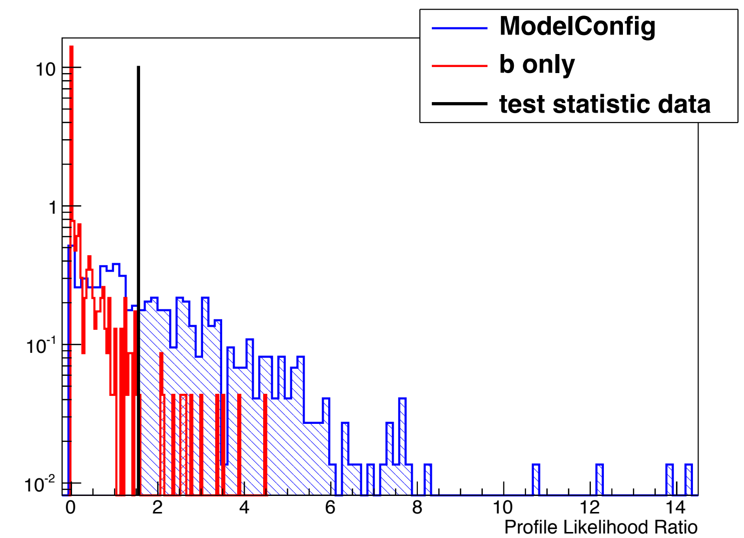
Asymptotic Calculations

Always good to know: the significance $Z = \sqrt{q_0}$ for the profile Likelihood for discovery.

➡ **Asymptotically, the significance is the square-root of the observed profile Likelihood value under the null hypothesis.**

RooStats::ProfileLikelihoodTestStat returns $q_0/2$.





Exercise: go back and add this asymptotic result to the previous exercise.

Advanced I: Generate expected data and add the expected p_0 to the plot.
Advanced II: Add energy scale systematic uncertainty and observe the breaking of the Asymptotics.