Working with Collie:
A Confidence Level Limit Evaluator

Jun, 02nd 2010

Angelo Santos (IFT-UNESP)

Outline

- What is Collie?
- Collie inputs and Effects from Systematics
- Example of Collie input
- Creating a Collie IO file
- Example of Log Likelihood Ratio distribution
- Confidence level calculations
- χ^2 formulas
- χ^2 fit results
- Matrix correlation between systematics.

What is Collie?

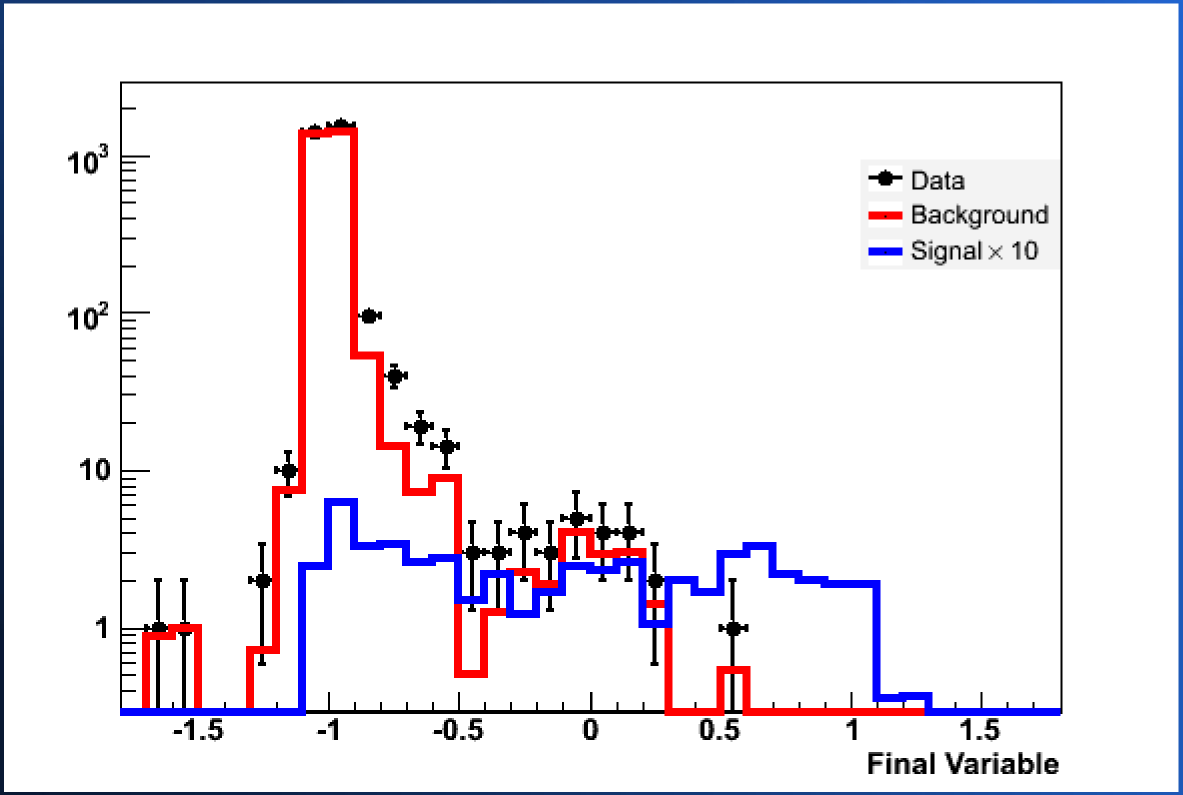
- Collie is a software suite designed to generate Confidence Levels (CL) which is used to calculate upper limits for TEST and NULL hypothesis.
    - TEST hypothesis: Signal + background hypothesis, i.e. new Physics.
    - NULL hypothesis: background only.
- This tool tries to separate TEST and NULL hypothesis in a way to get observed and expected limits with ± 1 / 2 sigmas.
- χ^2 fits are applied to compare data/MC considering:
    - Gaussian statistics
    - Poisson likelihood statistics
- Limits are calculated for the minimum χ^2 value

Collie inputs and
Effects from Systematics

- Hight Energy Physics search analysis often has a final variable chosen to be sensitive to a parameter of the search system.
- This final variable is composed by signal, background and data distributions.
- Measurements are affected by nuisance parameters.
  - Nuisance parameters are not parameters of interest, but somehow affect measurements.
  - Example of nuisance parameters: luminosity, efficiencies, background cross sections.
- Nuisance parameters have systematic uncertainties which can degrade the sensitive of the final variable.
- So Collie tries to reduce the effects from systematics.

Example of Collie input

- Final variable from TMVA analysis



Creating a Collie IO file

- int main(int argc, char* argv[ ]) {
- CollieIOFile* cfile = new CollieIOFile();

  //outputfile and channel name
- cfile->initFile("exampleCollieIOfile.root", "Test Channel");
- cfile->setInputHist(0.0,1.0,20); //Define your input histograms
- cfile->setSmooth(false); //Option to smooth histograms
- //Define backgrounds
- vector<string> bkgdNames;
- bkgdNames.push_back("Bkgd1");
- bkgdNames.push_back("Bkgd2");
- cfile-> createChannel(bkgdNames);

...creating collie IO file

// Get inputs from a file:

- TFile infile("myInputFile.root");
- TH1D* data = (TH1D*)infile.Get("data");
- TH1D* signal = (TH1D*)infile.Get("signal");
- TH1D* bkgd1 = (TH1D*)infile.Get("bkgd1");
- //Backgrounds are passed in via vector
- vector<TH1D*> vbkgd;
- vbkgd.push_back(bkgd1);
- vbkgd.push_back(bkgd2);

// Mass points are passed here

- cfile->createMassPoint(100, data, sig, 1,vbkgd,valpha);

…creating IO files

//Add systematics...either flat or by shape (ie, function of final variable)

// if by shape, must supply a histogram of the values in percent(%) fluctuations...

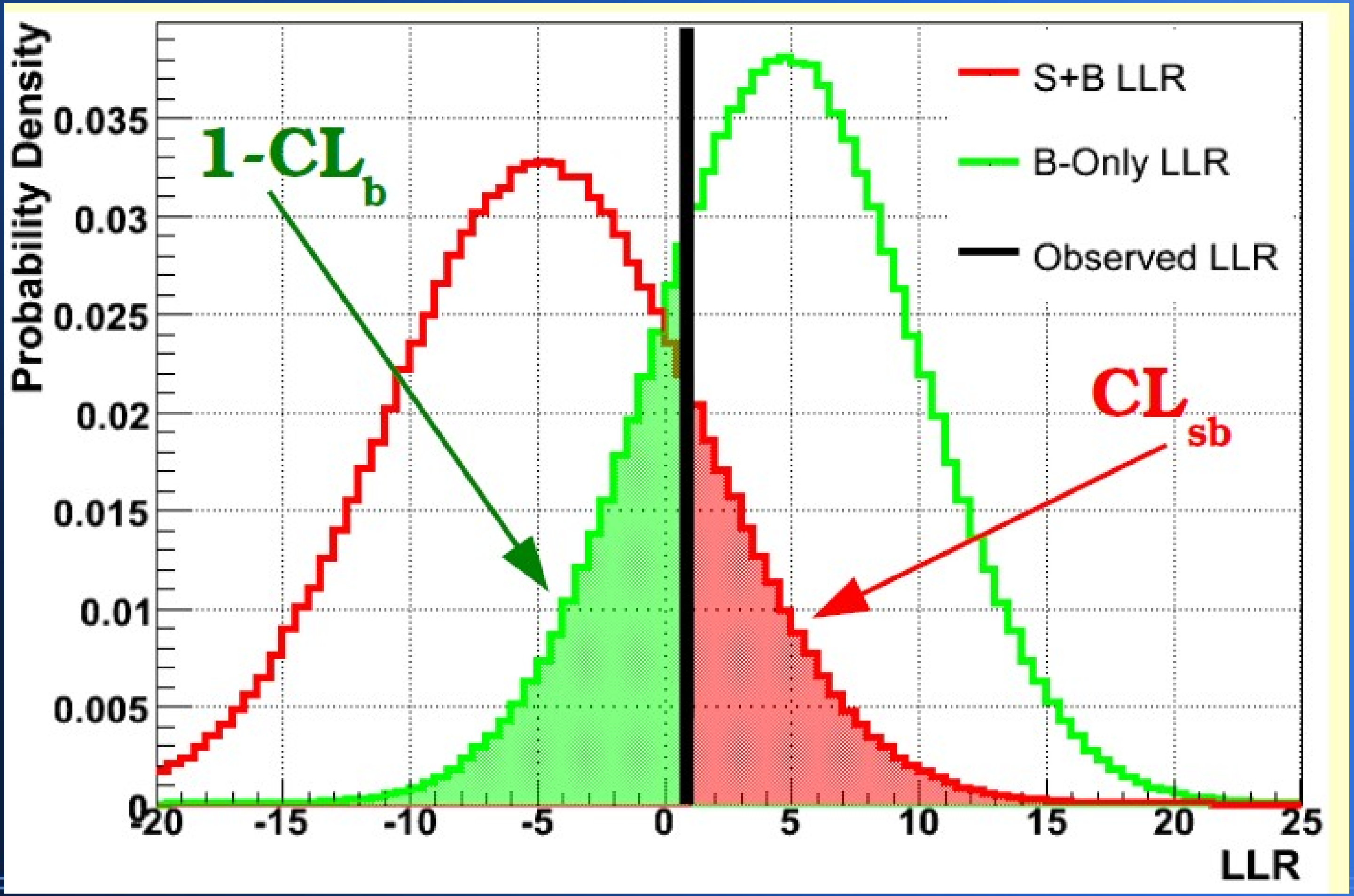
// Signal requires no index, but backgrounds must be specifically indexed (0>N bkgds)

- cfile->createFlatSigSystematic("Lumi",0.06,0.06,100);
- cfile->createShapeSigSystematic("SigShape",sigSystP,sigSystN,100);
- cfile->createFlatBkgdSystematic(0,"Lumi",0.06,0.06,100);
- cfile->createFlatBkgdSystematic(1,"Lumi",0.06,0.06,100);
- ///store channel
- cfile->storeFile();
- }

Log Likelihood Ratio distributions

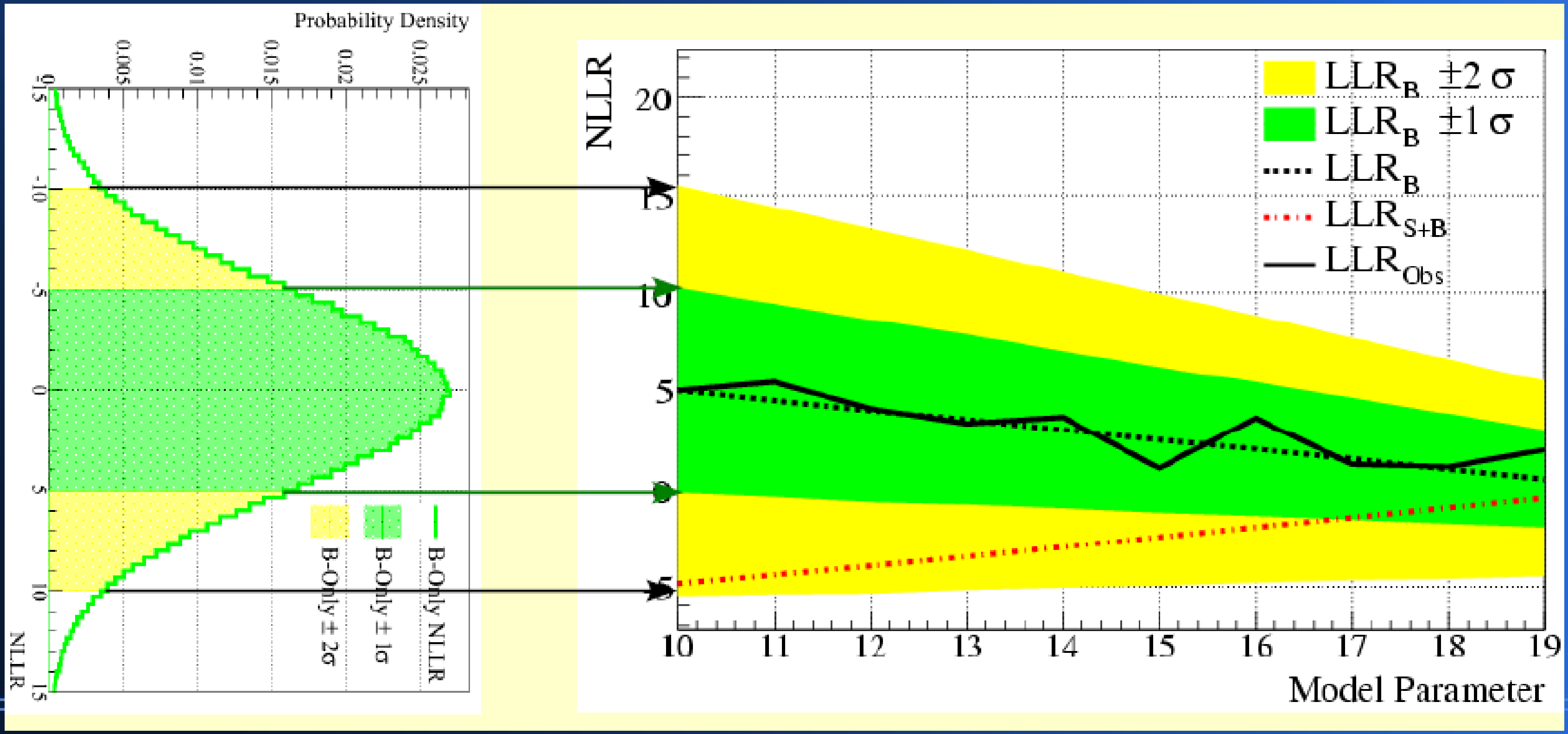
- 1 – Cls = 1 – α = confidence level

$$CL_s = \frac{CL_{s+b}}{CL_b}$$



Confidence level calculations

- Confidence level can be evaluated by 4 different kinds of tools:
  - CLfast: fast tool, but ignores systematics
  - CLsyst: standard gaussian including systematics
  - CLfit: single likelihood fit, maximized over systematics
  - CLfit2: double likelihood fit, maximized over systematics



χ^2 formulas

- Gaussian statistics

$$\chi^2 = \sum_i^{N^{\text{bins}}} \frac{(d_i - p_i)^2}{\sigma_i^2} = -2\ln\left(\frac{\mathcal{L}}{\mathcal{L}_0}\right)$$

$$\mathcal{L} = \prod_i^{N^{\text{bins}}} \left(\frac{1}{\sigma_i\sqrt{2\pi}}\right) \exp^{-\frac{(d_i - p_i)^2}{2\sigma_i^2}}$$

$$\mathcal{L}_0 = \prod_i^{N^{bins}} \left(\frac{1}{\sigma_i\sqrt{2\pi}}\right)$$

- Poisson statistics

$$\mathcal{L}^P = \prod_i^{N^{\text{bins}}} \frac{p_i^{d_i} \exp^{-p_i}}{d_i!}$$

$$\mathcal{L}_0{}^P = \prod_i^{N^{\text{bins}}} \frac{d_i^{d_i} \exp^{-d_i}}{d_i!}$$

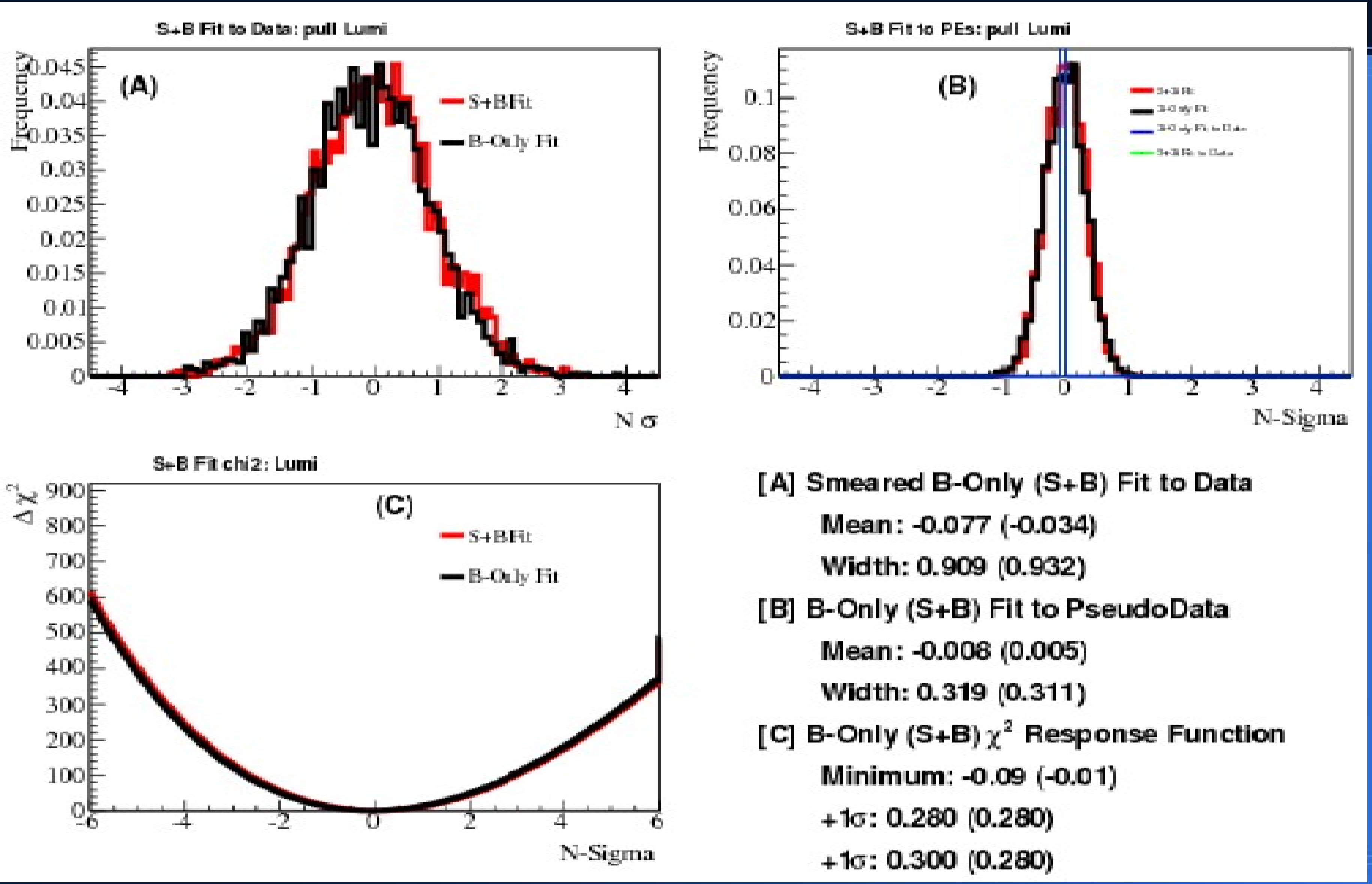
→

$$\chi^2 = -2\ln\left(\frac{\mathcal{L}^P}{\mathcal{L}_0{}^P}\right)$$

$$= -2\ln \prod_i^{N^{\text{bins}}} \left(\frac{p_i}{d_i}\right)^{d_i} \exp^{(d_i - p_i)}$$

$$= 2\sum_i^{N^{\text{bins}}} (p_i - d_i) - d_i \ln\left(\frac{p_i}{d_i}\right)$$

χ^2 fit results



Matrix correlation between systematics