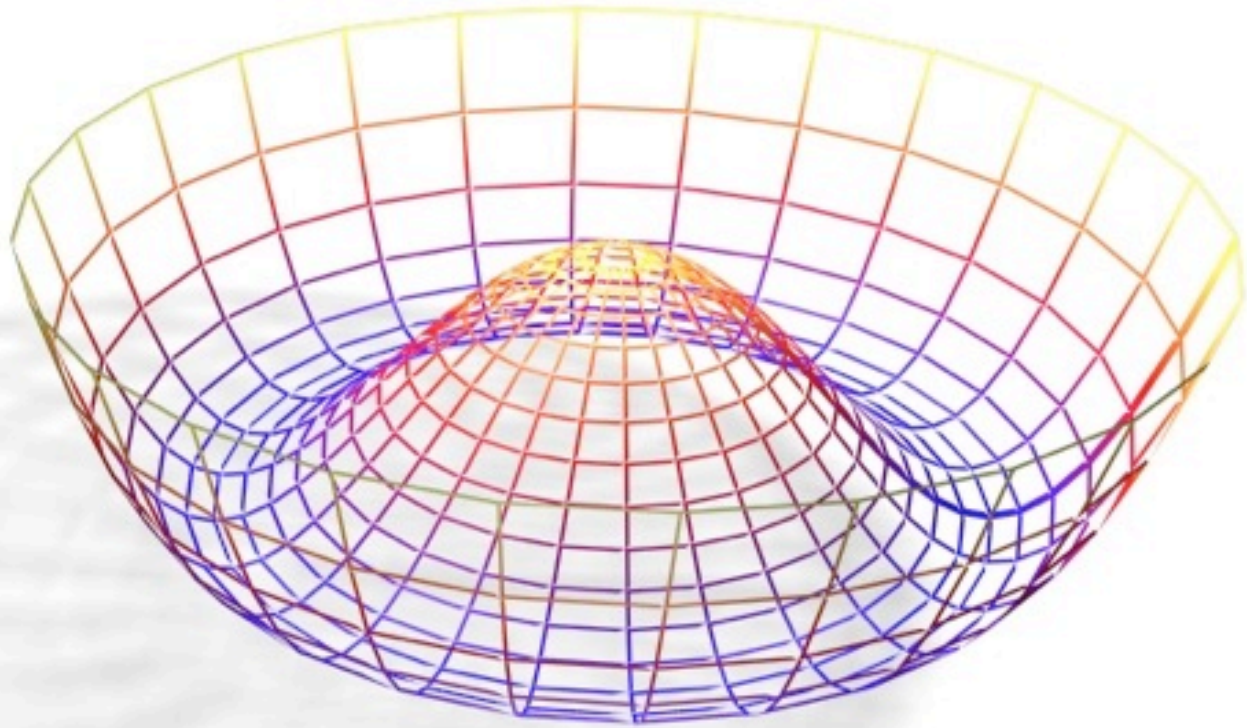
Center for
Cosmology and
Particle Physics

The Theory/Experiment Interface:
Publishing the Likelihood Function with the RooFit/RooStats Workspace



Kyle Cranmer,
New York University

Some Personal History

Some Personal History

Archbishop of Canterbury Thomas Cranmer (born: 1489, executed: 1556) author of the “Book of Common Prayer”

Some Personal History

Archbishop of Canterbury Thomas Cranmer (born: 1489, executed: 1556) author of the “Book of Common Prayer”

Two centuries later (when this Book had become an official prayer book of the Church of England) Thomas Bayes was a non-conformist minister (Presbyterian) who refused to use Cranmer’s book

"Bayesians address the question everyone is interested in, by using assumptions no-one believes"

"Frequentists use impeccable logic to deal with an issue of no interest to anyone"

-L. Lyons

A few points

Objective part of Bayesian inference is encoded in Likelihood

$$P(\text{theory}|\text{data}) = \frac{L(\text{data}|\text{theory})\pi(\text{theory})}{P(\text{data})}$$

- improvements in Likelihood is not Bayesian vs. Frequentist

Prior may be based on data $\pi(\text{theory}) \propto L'(\text{data'}|\text{theory})\eta(\text{theory})$

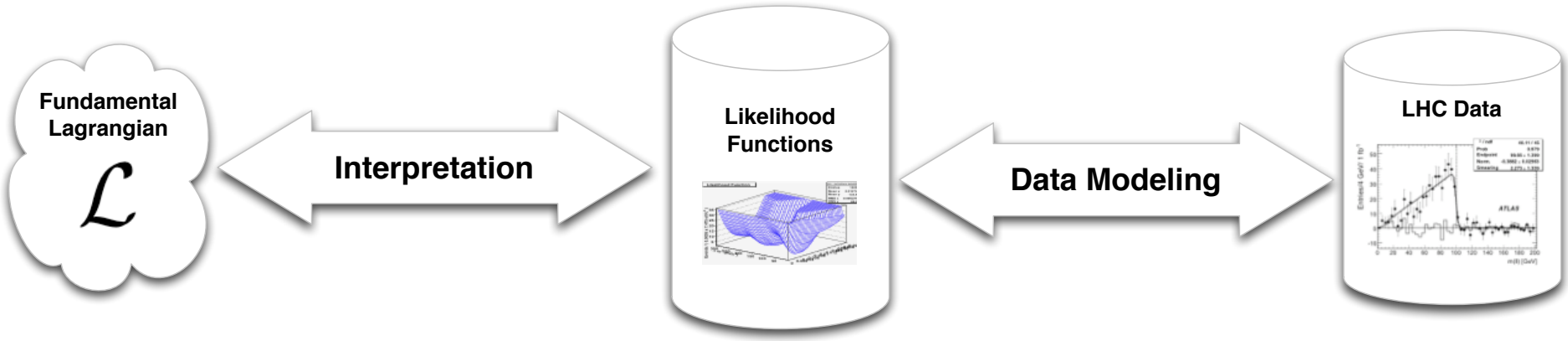
- but it also depends on the initial prior $\eta(\text{theory})$

In the same way that the "Bayesian calculus" allows for propagation of belief, the measurements can be combined with the likelihood function

$$L_{\text{tot}}(\text{data'}|\text{theory}) = L(\text{data'}|\text{theory})L'(\text{data'}|\text{theory})$$

Ideal scenario

The ideal scenario for the interface between the data and the inference to the fundamental lagrangian parameters is through a likelihood function that accurately incorporates all the experimental systematics and retains as much power in the data as possible



Is this feasible?

- It is the basic model on which Zfitter, GFitter, SFitter, Fittino, MasterCode, Kismet, SuperBayes, etc. are based
- unfortunately, likelihood functions are usually simplistic and based on a few 1-d measurements

Current scenario

Taken from the GFitter paper

[23]This procedure only uses the M_H value under consideration, where Higgs-mass hypothesis and measurement are compared. It thus neglects that in the SM a given signal hypothesis entails background hypotheses for all M_H values other than the one considered. An analysis accounting for this should provide a statistical comparison of a given hypothesis with all available measurements. This however would require to know the correlations among all the measurement points (or better: the full experimental likelihood as a function of the Higgs-mass hypothesis), which are not provided by the experiments to date. The difference to the hypothesis-only test employed here is expected to be small at present, but may become important once an experimental Higgs signal appears, which however has insufficient significance yet

Origins I: The First "Statistics in HEP" conference

WORKSHOP ON CONFIDENCE LIMITS

CERN, Geneva, Switzerland
17–18 January 2000 CERN 2000–005

Massimo Corradi

Does everybody agree on this statement, to publish likelihoods?

Louis Lyons

Any disagreement ? Carried unanimously. That's actually quite an achievement for this Workshop.

…[Fred James wants to be able to calculate coverage, Don Groom wants to able to calculate goodness of fit]…

Cousins

I thought the point of unanimity was that publishing the likelihood function was a *necessary* condition, not a sufficient condition.

But a practical problem remained: *How* to communicate multi-D likelihood?

http://indico.cern.ch/conferenceDisplay.py?confId=100458

Outline

Information:

- What is the RooStats Project?
- What the workspace can do for SUSY/BSM Fits
- Real-life examples from the LHC

Example Use cases

- A critical look at the weak points in our current chain

Moving forward:

- Hard problems that can be solved with planning
- Making a clear request to the experiments (discussion)
- Preparing toy benchmark examples (discussion)

RooStats: Project info

Started in 2005, when René Brun asked me to help organize statistical tools in ROOT

- Main goals are to provide a common framework for various statistical techniques (Frequentist, Bayesian, Likelihood based,...)

We want tools to work with probability models of arbitrary complexity (which implies interfaces, etc.)

- Decided to base tools on RooFit's data modeling language and core interfaces

Initially an ATLAS/CMS project, but other experiments are interested (LHCb, Fermi, ...)

- core developers
  - K. Cranmer (ATLAS), Lorenzo Moneta (ROOT), Gregory Schott (CMS), Wouter Verkerke (RooFit)
- open project, you are welcome to contribute
  - ~10 others contributing now, growing fast

Included since ROOT v5.22 (we are now on 5.27)

https://twiki.cern.ch/twiki/bin/view/RooStats/WebHome

RooStats has been a topic of conversation in every combined ATLAS/CMS statistics forum meeting

- In July, we showed the first toy ATLAS/CMS Higgs combination using the tools of RooFit/RooStats.
  - see agenda:

http://indico.cern.ch/conferenceDisplay.py?confId=100458

Major Goals and Status

Goal: Standardize interface for major statistical procedures so that they can work on an arbitrary RooFit model & dataset and handle many parameters of interest and nuisance parameters (systematics).

- **Status**: Done
  - **ConfIntervalCalculator** & **HypoTestCalculator** interface for tools
  - they return **ConfidenceInterval** and **HypoTestResult**

Goal: Implement most accepted techniques from Frequentist, Bayesian, and Likelihood-based approaches

- **Status**: Done / Ongoing
  - **ProfileLikelihoodCalculator**: (Likelihood) the method of MINUIT/MINOS
  - **FeldmanCousins**: (Frequentist) a generalization of F-C that can incorporate systematics
  - **MCMCCalculator**: (Bayesian) uses Metropolis-Hastings algorithm (native or BAT)
  - **HybridCalculator**: (Bayesian/Frequentist Hybrid) used at LEP and Tevatron

Goal: Provide utilities to perform combined measurements

- **Status:** Partially done / Ongoing
  - **RooWorkspace** allows one to save arbitrary RooFit model (even with custom code) into a .root file. PDFs and DataSets have been extended to facilitate combinations.
  - Same technology can aid in digital publishing ← today's focus

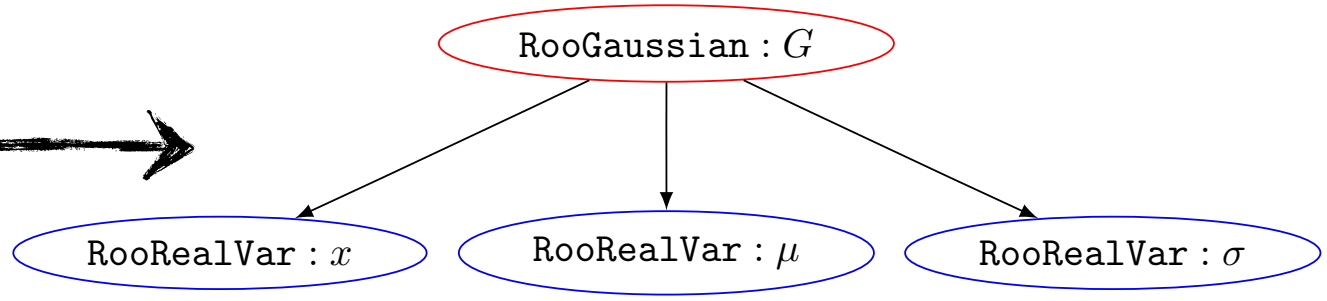
What goes in a Workspace

The workspace stores the full probability model and any data necessary to evaluate the likelihood function

- it is the code necessary to evaluate the likelihood function at an arbitrary point in the parameter space. It is not a big table of likelihood values!
- we are using the same ROOT technology that the LHC experiments are using to save their data
  - well supported, and supports "schema evolution" / backwards compatibility
- the probability model also allows you to generate toy data for any given parameter point
  - necessary for frequentist methods, goodness of fit, coverage
- PDFs and functions can be extended by the user (source stored in workspace)

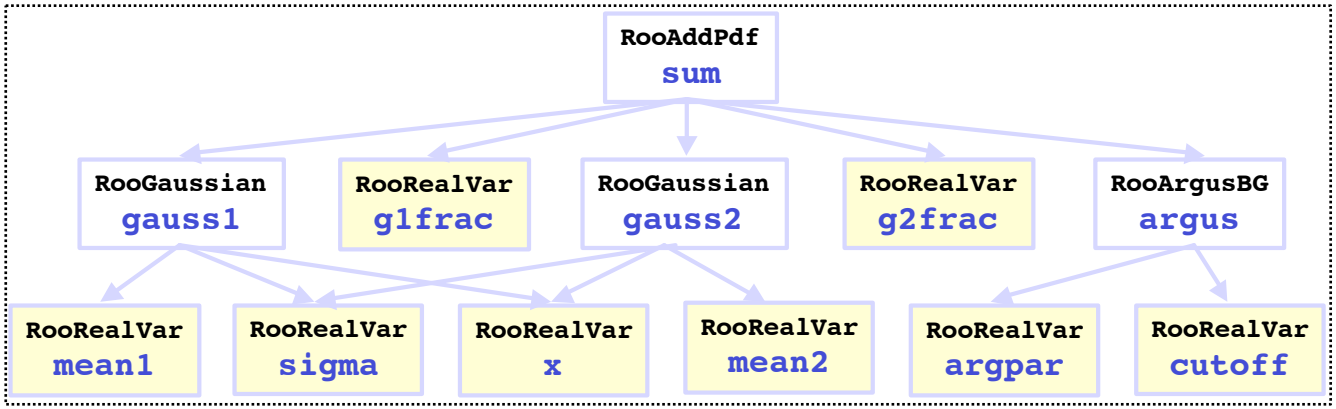
I will show some visualization of real-life LHC probability models. Let's start with a simple example:

$$G(x|\mu,\sigma) \longrightarrow$$

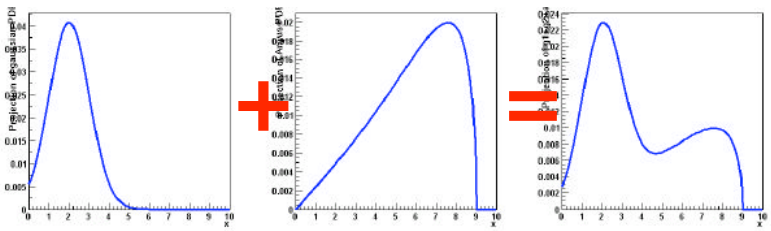


RooFit: A data modeling toolkit

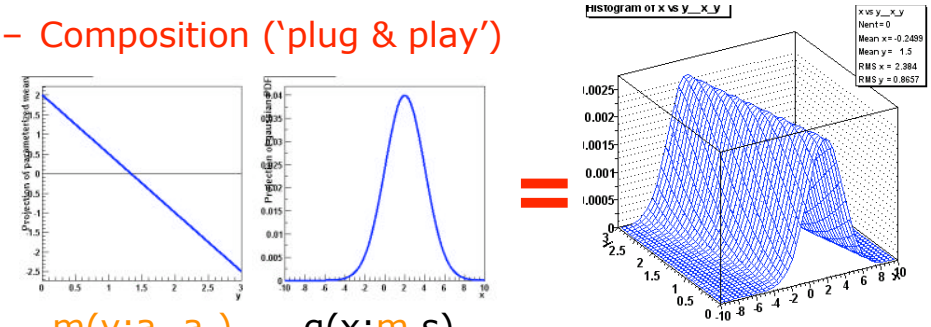
A major tool at BaBar. Fit complicated models with >100 parameters!



– Addition



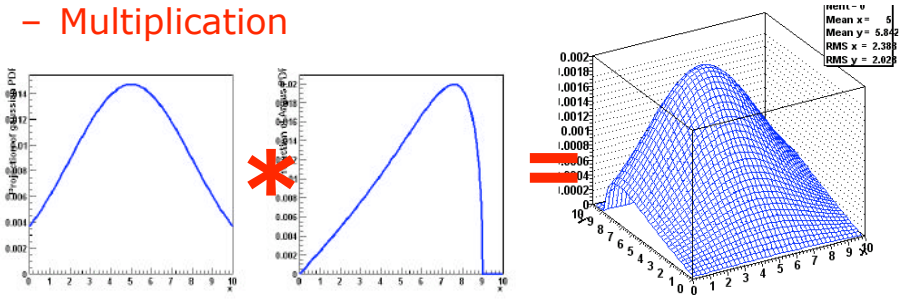
– Composition ('plug & play')



$m(y;a_0,a_1)$ $g(x;m,s)$ $g(x,y;a0,a1,s)$

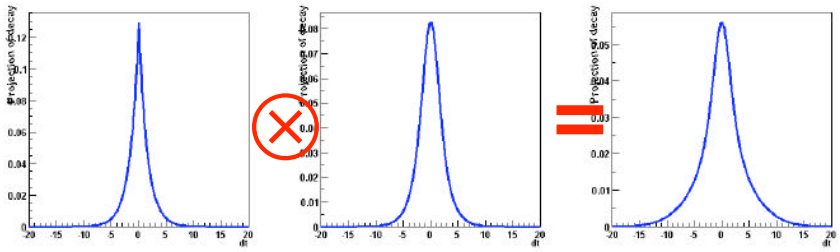
Possible in *any* PDF
No explicit support in PDF code needed

– Multiplication



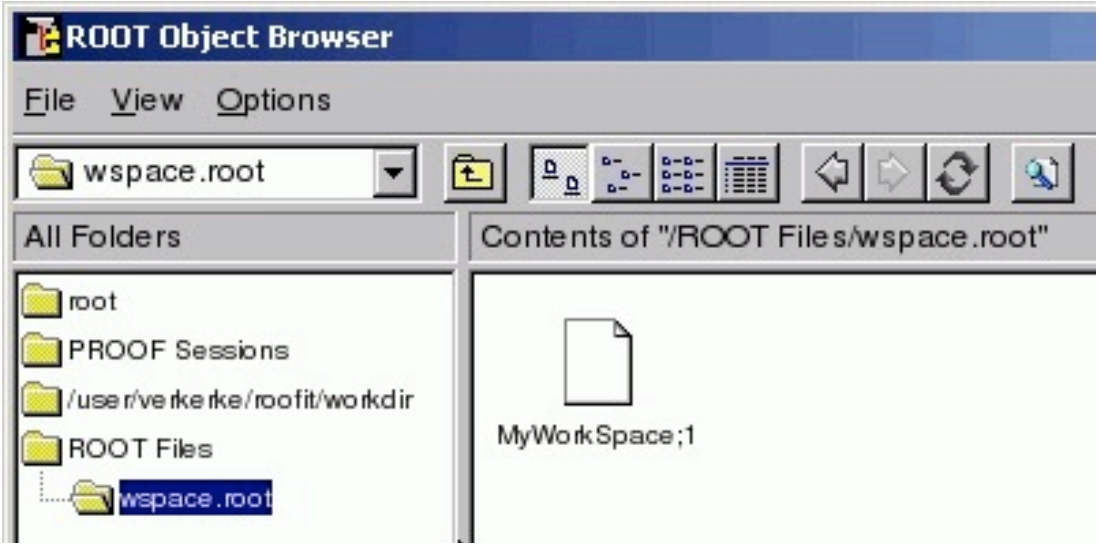
Wouter Verkerke,

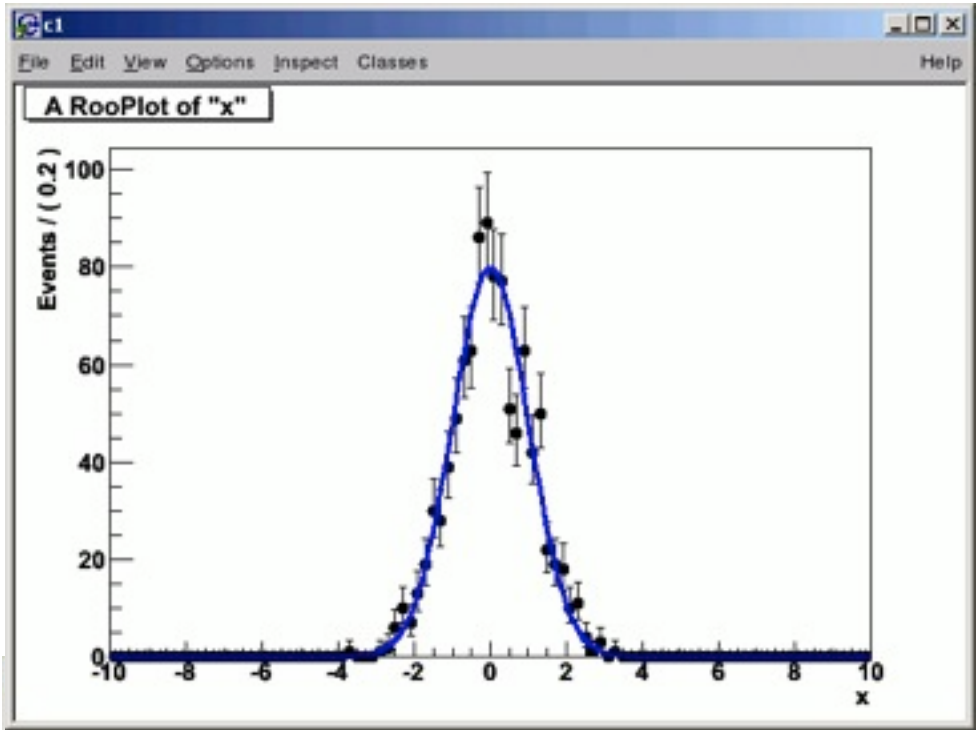
– Convolution

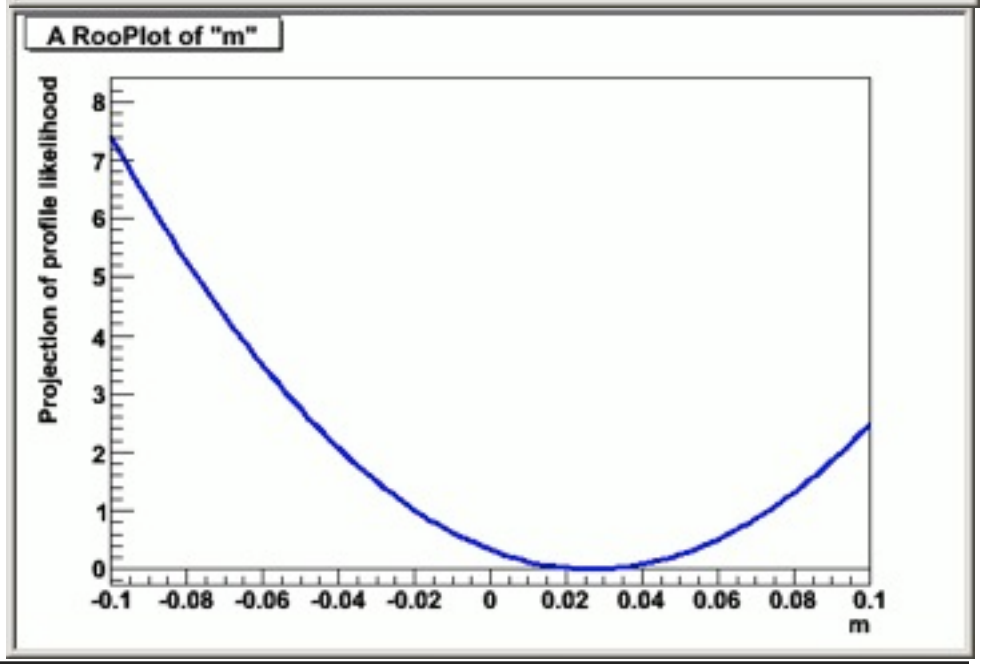


Wouter Verkerke, UCSB

The RooFit/RooStats workspace







RooStat's Workspace can save in a file the full likelihood model and the minimal data necessary to reproduce likelihood function.

The technology is generic, we decide how to parametrize the model.

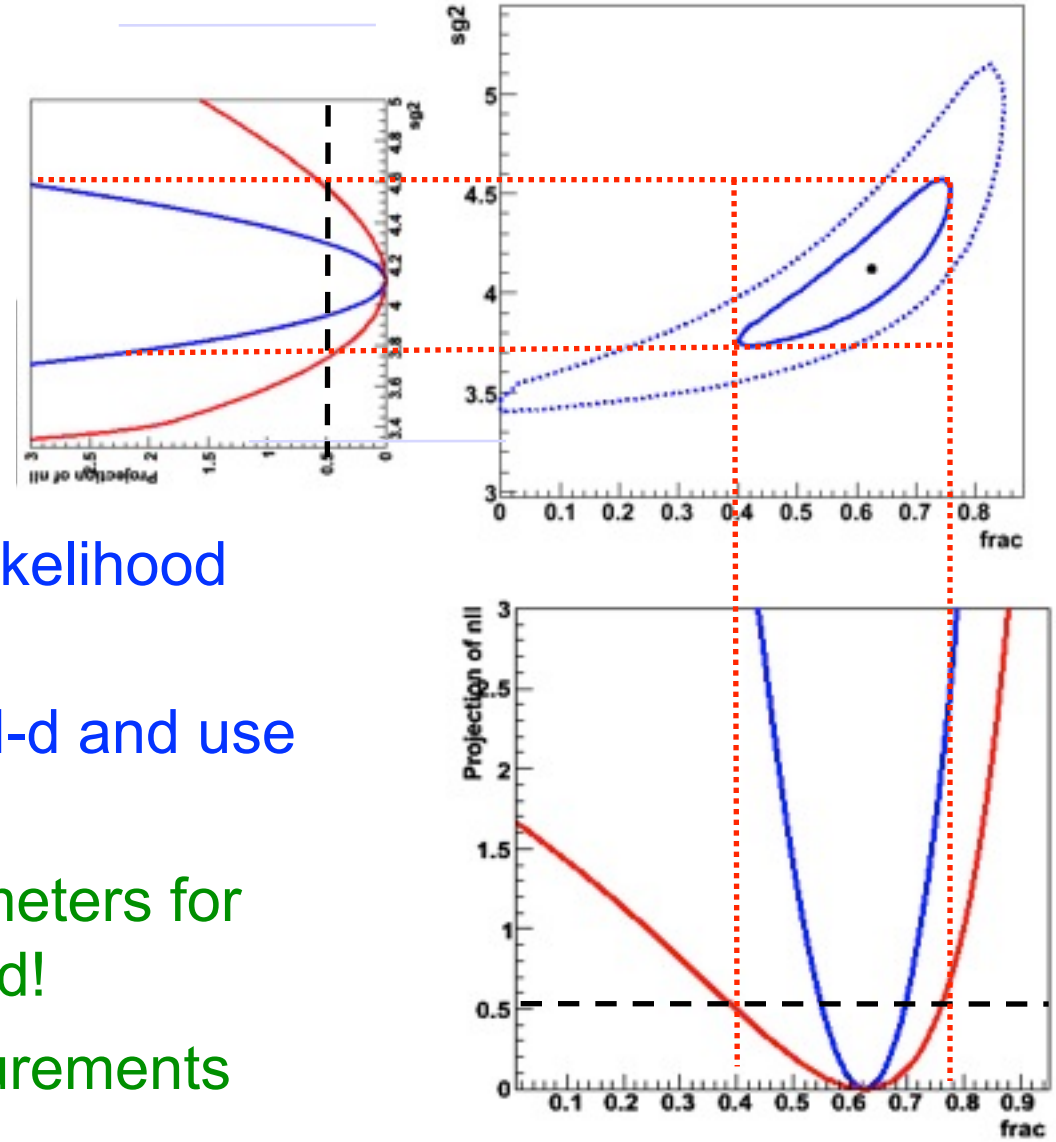
Being used by ATLAS/CMS for very complicated models

Need this for combinations, exciting potential for publishing results.

Extracting Contours from these results

The workspace can represent arbitrary models with many parameters of interest and many nuisance parameters

This contour is NOT an ellipse!

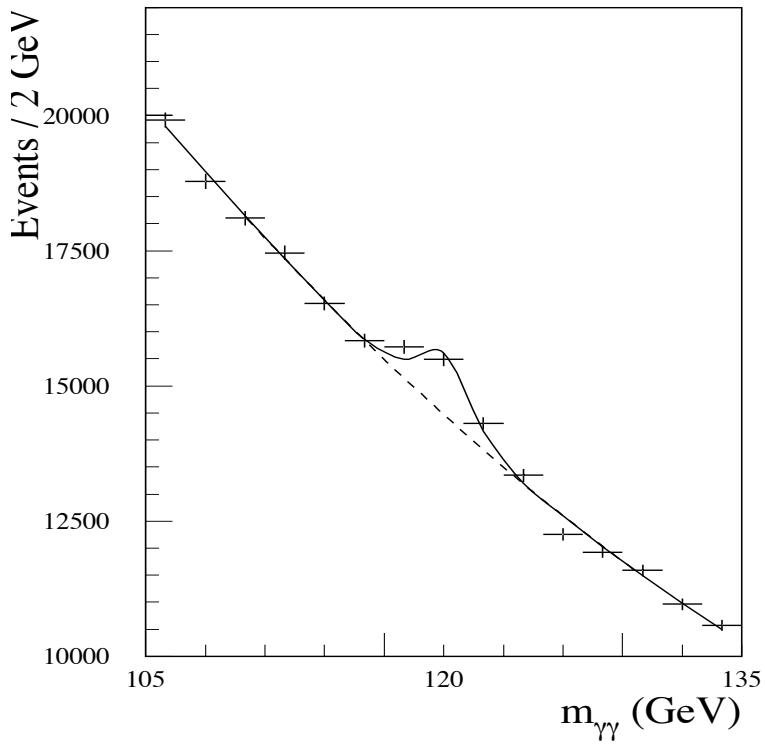


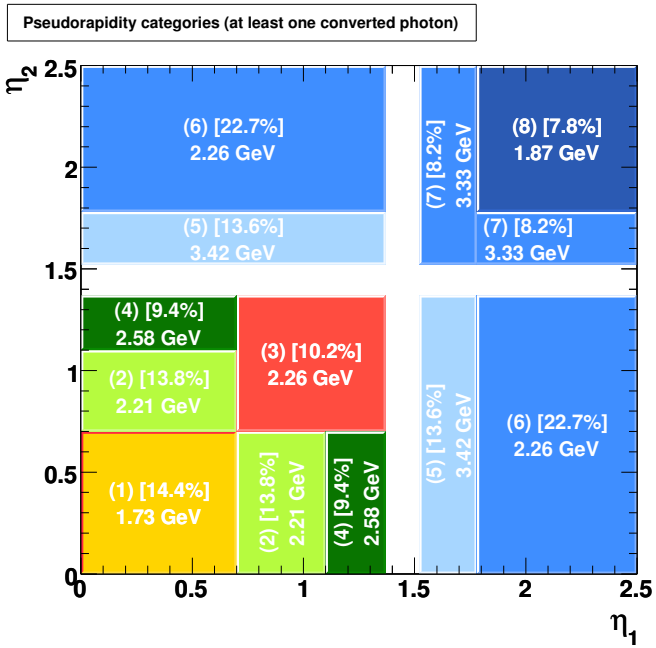
- One can plot 2-d contours, 1-d likelihood functions.
- One can evaluate likelihood in N-d and use to evaluate a theoretical model
- If the model has nuisance parameters for systematics, they will be included!
- Easy to combine multiple measurements

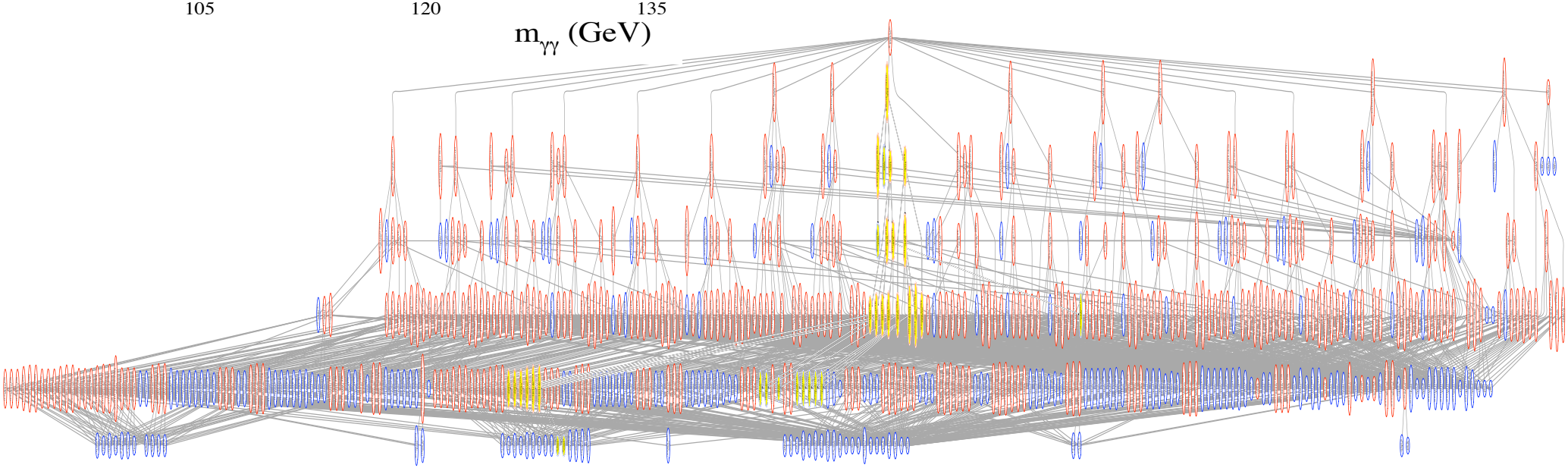
Taken from Wouter Verkerke, NIKHEF

Examples of Real-Life LHC Models

ATLAS H->γγ


Events / 2 GeV
20000
17500
15000
12500
10000
105
120
135
$m_{\gamma\gamma}$ (GeV)


Pseudorapidity categories (at least one converted photon)
η_2
η_1
(6) [22.7%] 2.26 GeV
(5) [13.6%] 3.42 GeV
(7) [8.2%] 3.33 GeV
(8) [7.8%] 1.87 GeV
(7) [8.2%] 3.33 GeV
(4) [9.4%] 2.58 GeV
(2) [13.8%] 2.21 GeV
(3) [10.2%] 2.26 GeV
(1) [14.4%] 1.73 GeV
(2) [13.8%] 2.21 GeV
(4) [9.4%] 2.58 GeV
(5) [13.6%] 3.42 GeV
(6) [22.7%] 2.26 GeV



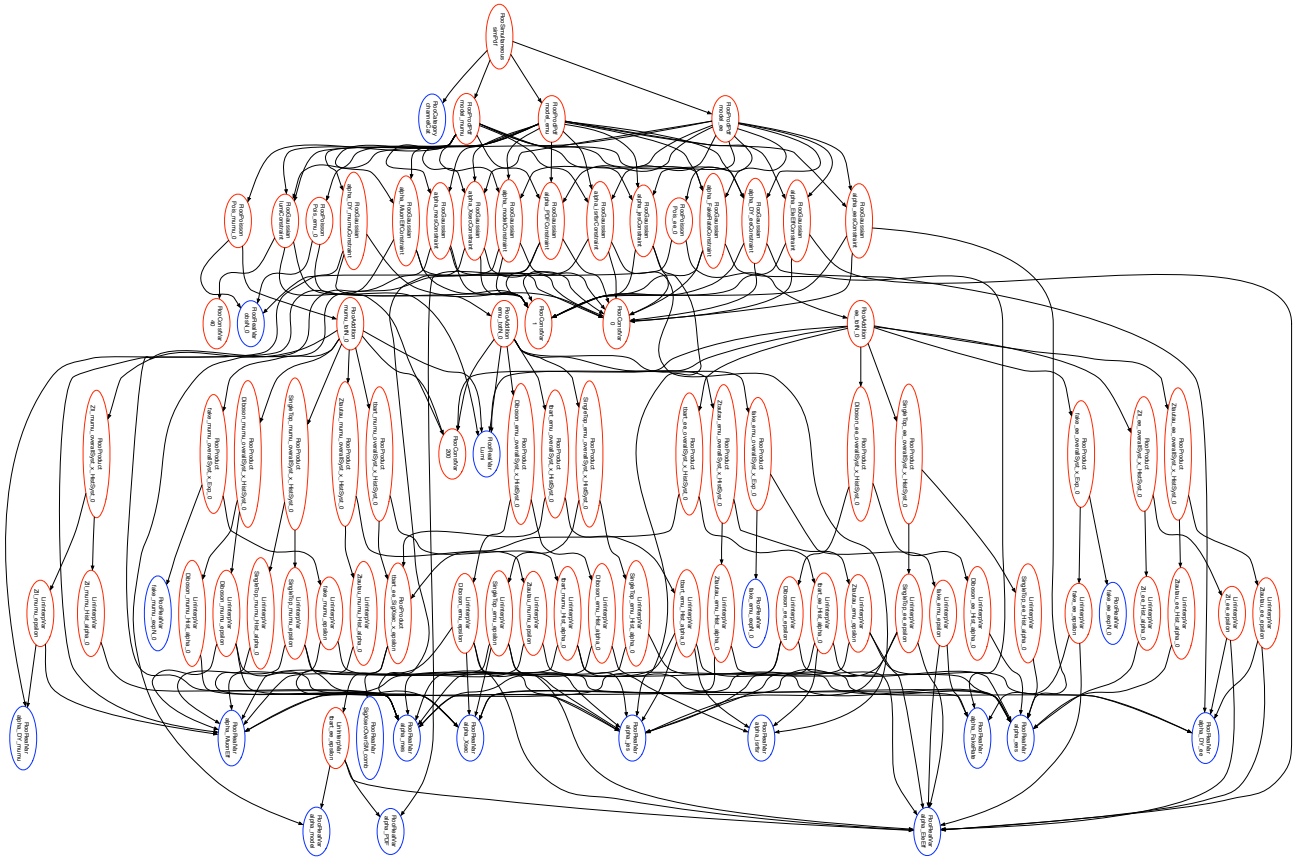
3-channel top combination

The graph below represents this PDF

$$L(\sigma_{sig}, \mathscr{L}, \alpha_j) = \prod_{l \in \{ee, \mu\mu, e\mu\}} \left\{ \prod_{i \in bins} \left[Pois(N_i^{obs} | N_{i,tot}^{exp}) \, Gaus(\tilde{\mathscr{L}} | \mathscr{L}, \sigma_{\mathscr{L}}) \prod_{j \in syst} Gaus(0 | \alpha_j, 1) \right] \right\}$$

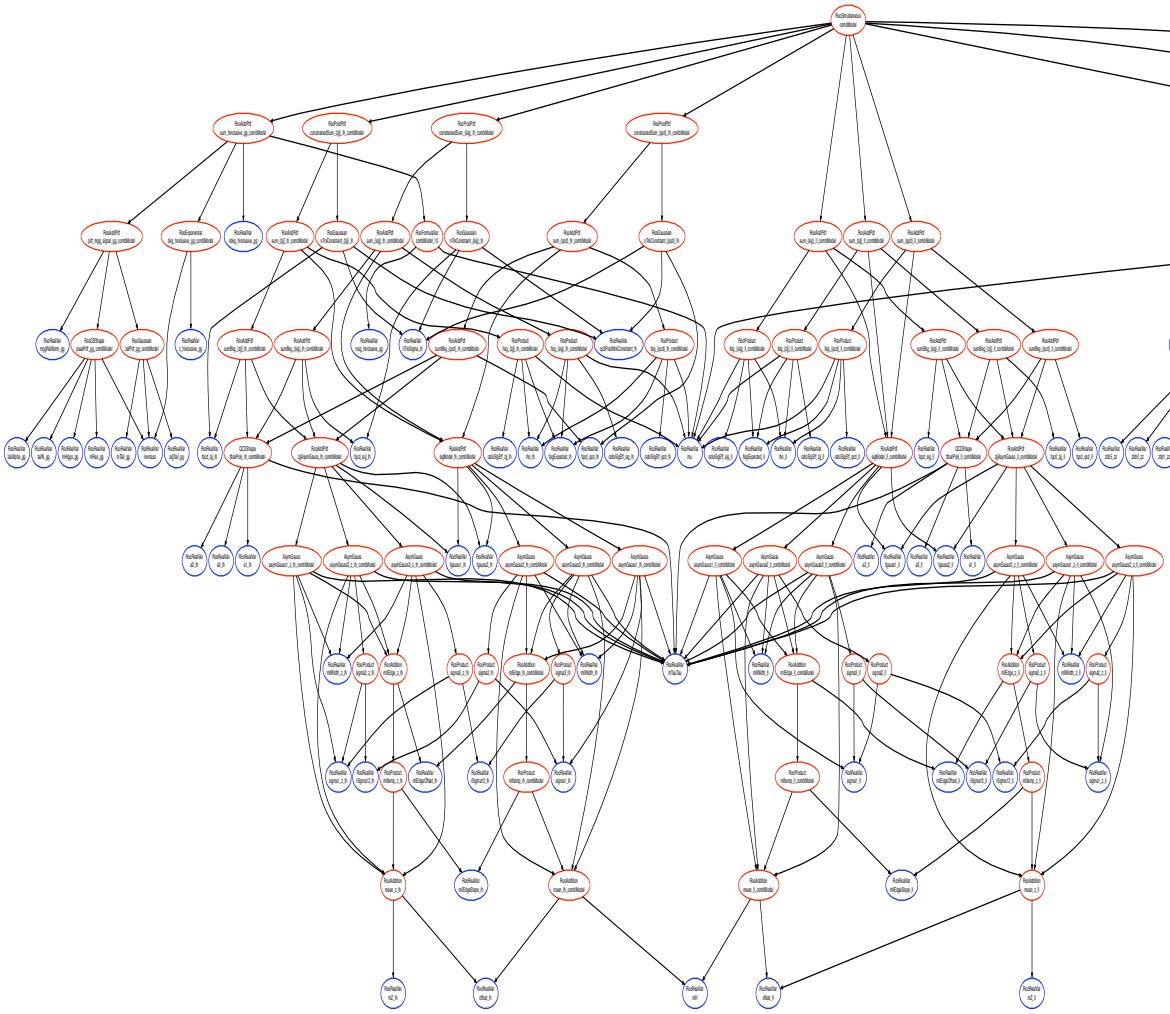
- where there are several relations between the expected means in the different channels

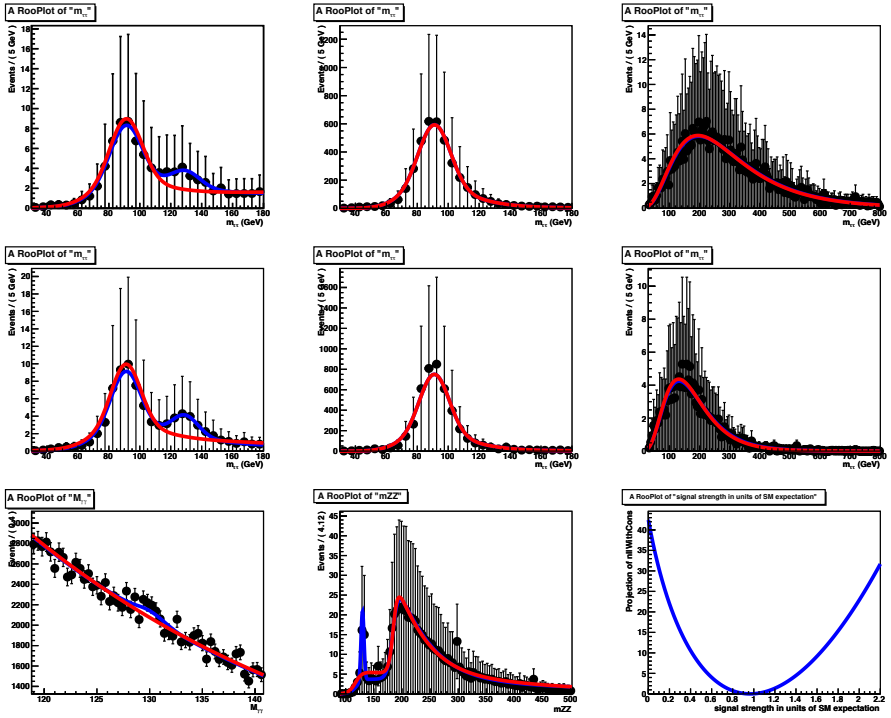
3 observations from data
13 control samples
1 parameter of interest
13 nuisance parameters



4-channel ATLAS Higgs combination

Center for Cosmology and Particle Physics

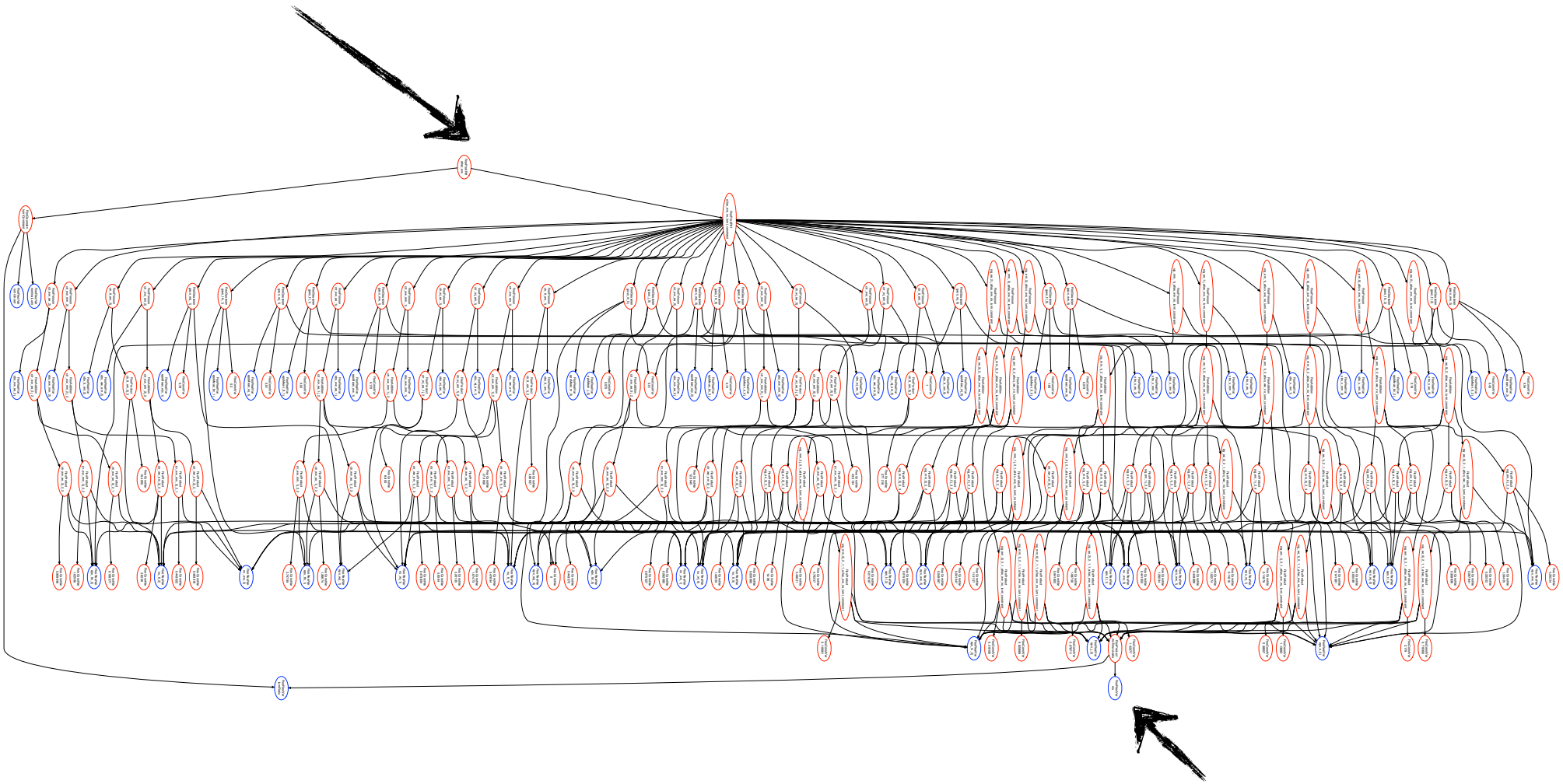




9 observations of continuous variables
1 parameter of interest
27 nuisance parameters

9 Channel ATLAS H->WW combination

top level model



25 measurements from data

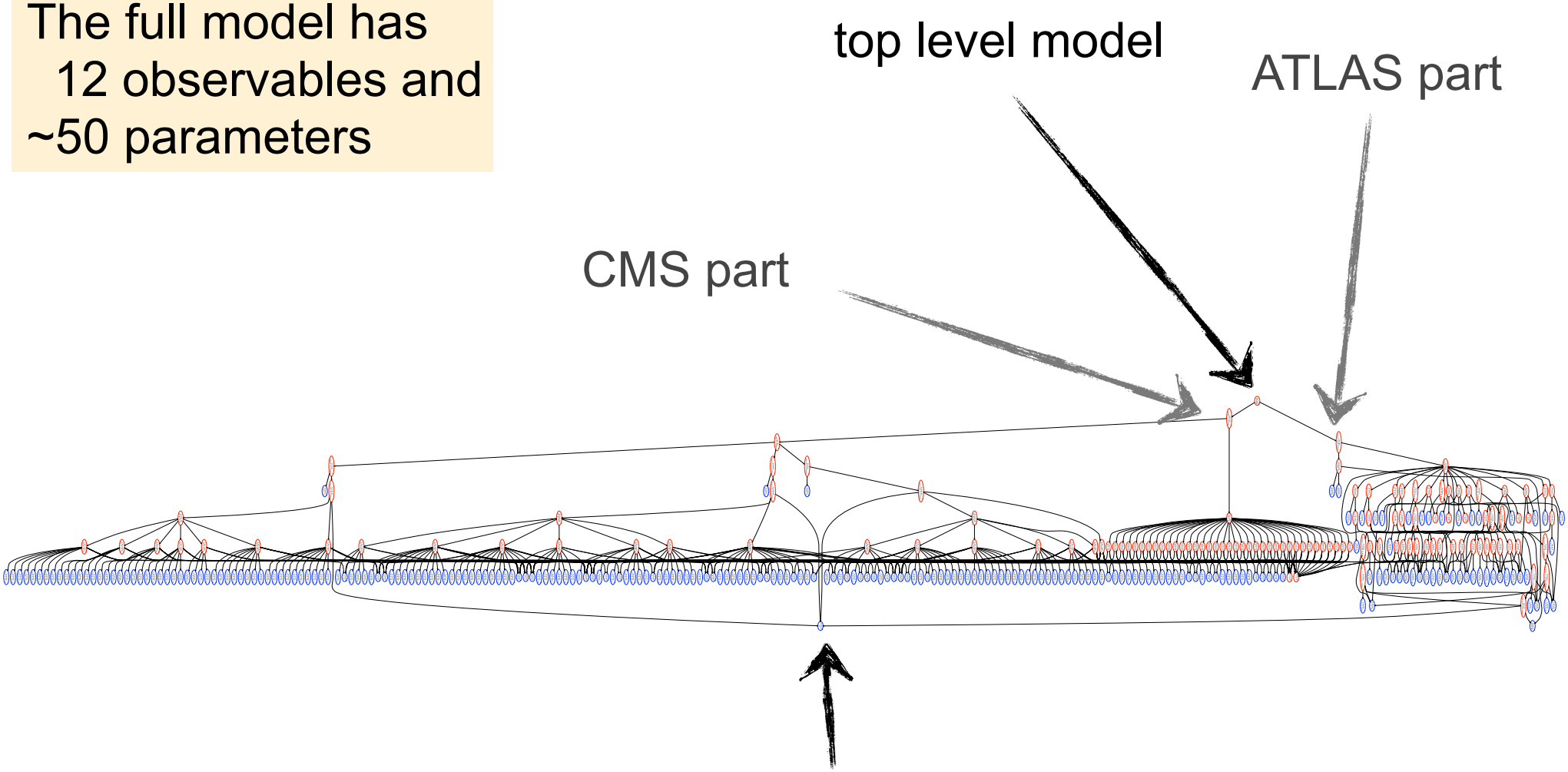
1 parameter of interest and 24 nuisance parameters

parameter of interest

$$\mu = \frac{\sigma BR}{\sigma_{SM} BR_{SM}}$$

Visualization of the ATLAS+CMS Workspace

The full model has
12 observables and
~50 parameters



parameter of interest

$$\mu = \frac{\sigma BR}{\sigma_{SM} BR_{SM}}$$

How to use the workspace

Using the likelihood function stored in a workspace does not mean:

- becoming an expert in RooFit/RooStats
- ever using a ROOT prompt.

To get started, I have a simple example:

- a C++ program with a main()
- with a real Makefile that links to the RooFit and RooStats libraries in ROOT

```
int main(int argc, char** argv) {

  evaluateLikelihoodFunction();

  return 0;
}
```

The program:

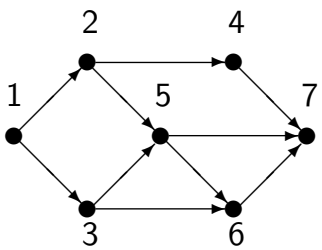
- opens the ROOT file and gets the workspace
- gets the log-likelihood function
- evaluates it at some random parameter points

```
void evaluateLikelihoodFunction(){

  // open root file with workspace
  // here you have to know name of workspace
  // you can get that by opening the file in root
  // and typing ".ls"
  // here we know it's called "w"
  TFile f("AnExampleWorkspace.root");
  RooWorkspace* w = (RooWorkspace*) f.Get("w");

  // now we need to get stuff out of the workspace
  // we can see what's inside with the Print() method
  w->Print();

  // the top-level PDF is called "model" in this case
  RooAbsPdf* model = w->pdf("model");
  // the only dataset is called "modelData"
  RooAbsData* data = w->data("modelData");

  // from the PDF and the data we can create the likelihood function
  RooNLLVar* nll = (RooNLLVar*) model->createNLL(*data);

  // now we need to know the parameters it depends on
  cout << " this likelihood depends on the following parameters"<<endl;
  RooArgSet* parameters = nll->getParameters(*data);
  parameters->Print();
  // get a pointer to the parameters
  RooRealVar* m1 = w->var("m1");
  RooRealVar* m2 = w->var("m2");

  cout << "\n\n--------------------------" << endl;
  cout << "Here are some random evaluations " << endl;

  for(int i=0; i<100; ++i){
    m1->randomize();
    m2->randomize();
    cout << "at (m1,m2) = ("<<m1->getVal()
         << ", " << m2->getVal()
         << ") -log likleihood = "
         <<  nll->getVal() << endl;
  }

}
```

Graphical Models

Given all these graphs, it's not surprising that one might think there's an application for Graphical Models

- graphs are different, but let's discuss connection

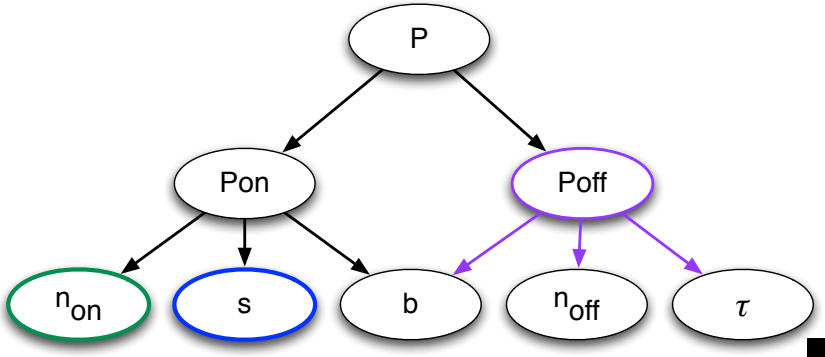


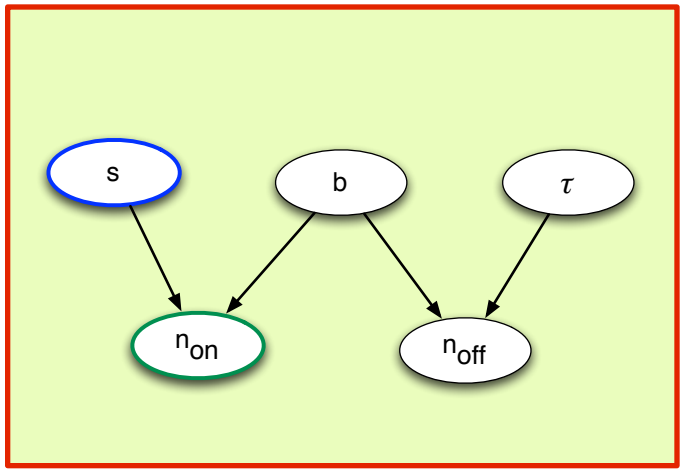
Directed Markov means

$$
\begin{aligned}
f(x) &= f(x_1)f(x_2\,|\,x_1)f(x_3\,|\,x_1)f(x_4\,|\,x_2) \\
&\times f(x_5\,|\,x_2,x_3)f(x_6\,|\,x_3,x_5)f(x_7\,|\,x_4,x_5,x_6).
\end{aligned}
$$

Theory exists for deriving all conditional independencies and exploiting local structure in graph for gross computational simplifications in complex models. Has been successfully exploited in AI, machine learning, and Bayesian statistics.

Graph of on/off model





A Critical Look at What We Are Doing Now

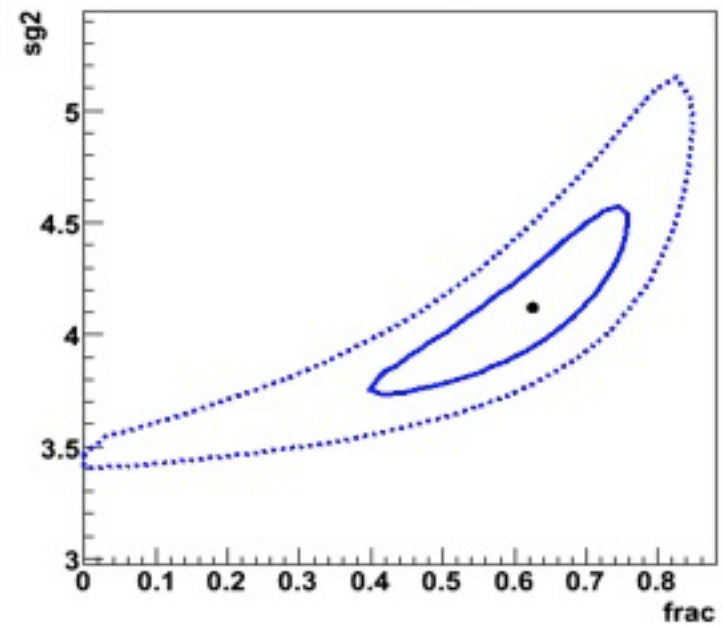
Correlated systematics

Clearly, several systematic effects will be correlated between the different measurements, and this must be taken into account

- That means the likelihood function needs to be a function of nuisance parameters.

Table 4: Endpoint positions for SU3 and SU4, in GeV. The first error is statistical, the second and third are the systematic and the jet energy scale uncertainty, respectively. The theoretical values are also given for ease of comparison to the left of the fitted values. The integrated luminosity assumed is 1 fb^{-1} for SU3 and 0.5 fb^{-1} for SU4.

Endpoint	SU3 truth	SU3 measured	SU4 truth	SU4 measured
$m_{\ell\ell q}^{\mathrm{edge}}$	501	$517 \pm 30 \pm 10 \pm 13$	340	$343 \pm 12 \pm 3 \pm 9$
$m_{\ell\ell q}^{\mathrm{thr}}$	249	$265 \pm 17 \pm 15 \pm 7$	168	$161 \pm 36 \pm 20 \pm 4$
$m_{lq(\mathrm{low})}^{\mathrm{max}}$	325	$333 \pm 6 \pm 6 \pm 8$	240	$201 \pm 9 \pm 3 \pm 5$
$m_{lq(\mathrm{high})}^{\mathrm{max}}$	418	$445 \pm 11 \pm 11 \pm 11$	340	$320 \pm 8 \pm 3 \pm 8$



Current approach attempts to

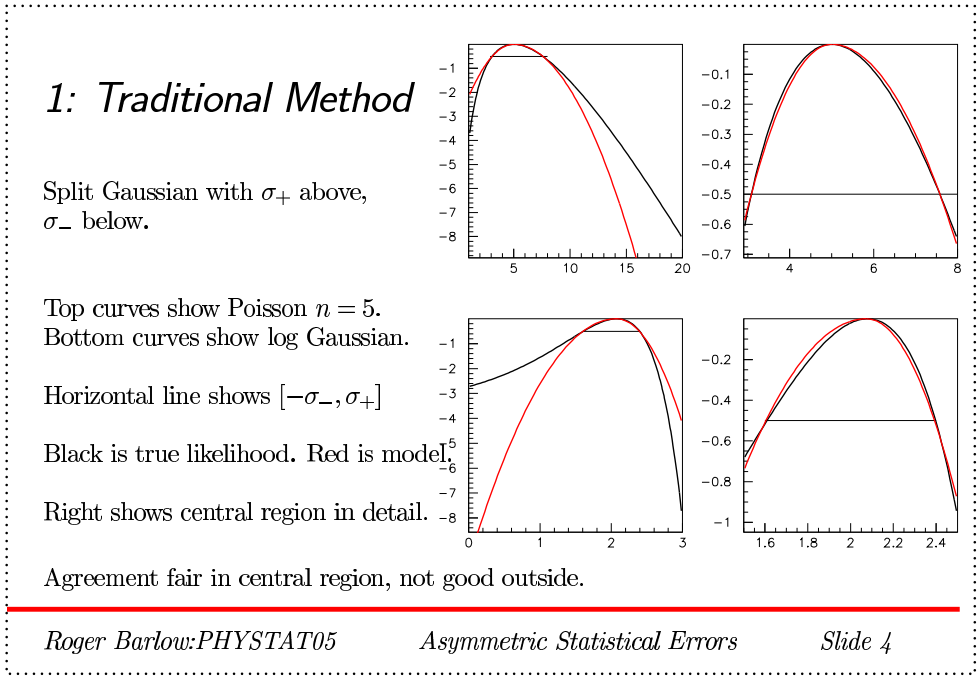
- Not clear from a table like this if errors are anti-correlated
- The relationship can be non-trivial, not able to be represented by a simple covariance matrix

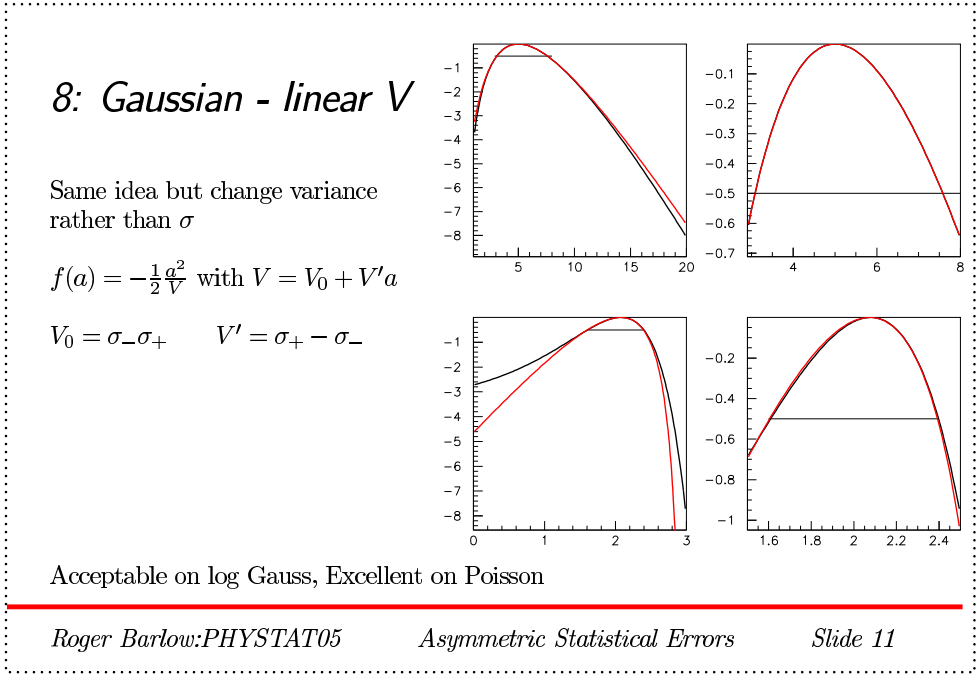
Wouldn't it be nice to be able to inflate the uncertainties and see what happens? Especially, when results don't agree.

Asymmetric Errors

I see that several groups interpret asymmetric errors reported by the experiments into a likelihood function:

- Of course, this is an ill-posed problem
- Rober Barlow considered 8 ways of doing this for cases in HEP, and the results can vary significantly





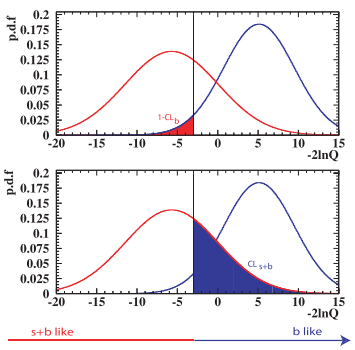
http://www.physics.ox.ac.uk/phystat05/Talks/slides.pdf

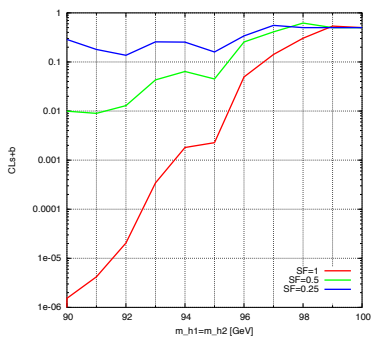
Likelihoods from CL_{s+b}

Because the experiments do not publish likelihoods, groups are often forced to try to infer the likelihood from other information, like CL_{s+b}

- The basis of this is that the CL_{s+b} was based on a LLR test statistic, which is assumed to be distributed as a chi-square distribution but let's look more closely
  - what is the "LLR" exactly?
  - Is it distributed as a chi-square distribution?

From Exclusions to Fits



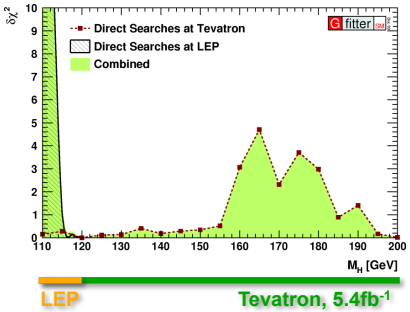


- Transfer (single-sided) CL_{s+b} into (double-sided) χ^2:
  $\chi^2_H = 2\,\mathrm{InvErf}^2(1 - 2CL_{s+b})$
- In the future (see below): Using HiggsBounds we can have χ^2 for each model with (pseudo)scalar particles
- Same for SUSY would be nice, but even more complicated . . .

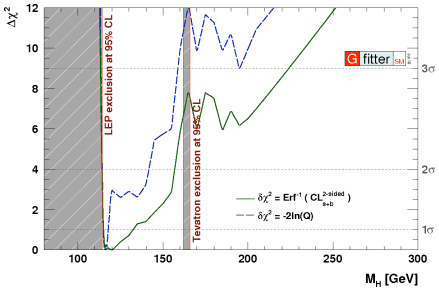
New: e.g. CL_{s+b} for MSSM/2HDM/. . .

Statistical interpretation Higgs limits

- Stat. interpretation in global fit: 2-sided CL_{S+B}
  - Experiments measure test statistics: $LLR = -2lnQ$, where $Q=L_{S+B}/L_B$
  - LLR is transformed by experiments into one-sided CL_{S+B} using toy-MC experiments
    - Sensitive to too few Higgs-like events
  - We transform 1-sided CL_{S+B} into 2-sided CL^{2s}_{S+B}
    - We measure deviations from the SM.
    - Eg. also interested in too many Higgs-like events.
  - χ^2 contribution calculated via inverse error function, assuming symmetric pdf : $\delta\chi^2 = Erf^{-1}(1-CL^{2s}_{S+B})$



- Alternative treatments (thanks to fruitful discussion with Tevatron people):
  - Use one-sided CL_{S+B}: however, different interpretation – want SM Higgs (not any Higgs)
  - Directly use $\delta\chi^2 \approx$ LLR: Bayesian interpretation, lacks pseudo-MC information



Three common "LLR" test statistics

We express cross-section as $\mu = \sigma / \sigma_{SM}$ for convenience.

Effect of systematics is parametrized by one or more "nuisance parameters" denoted ν .

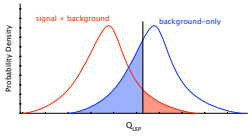
- best fit point is: $\hat{\mu}, \hat{\nu}$
- best fit of nuisance parameters with μ fixed is $\hat{\hat{\nu}}$ (aka "profiled")

In principle, s+b and b-only models can have different parametrizations

Three common test statistics used in the field are:

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

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

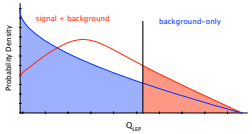


- ratio of profiled likelihoods (used commonly at Tevatron)

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

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

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

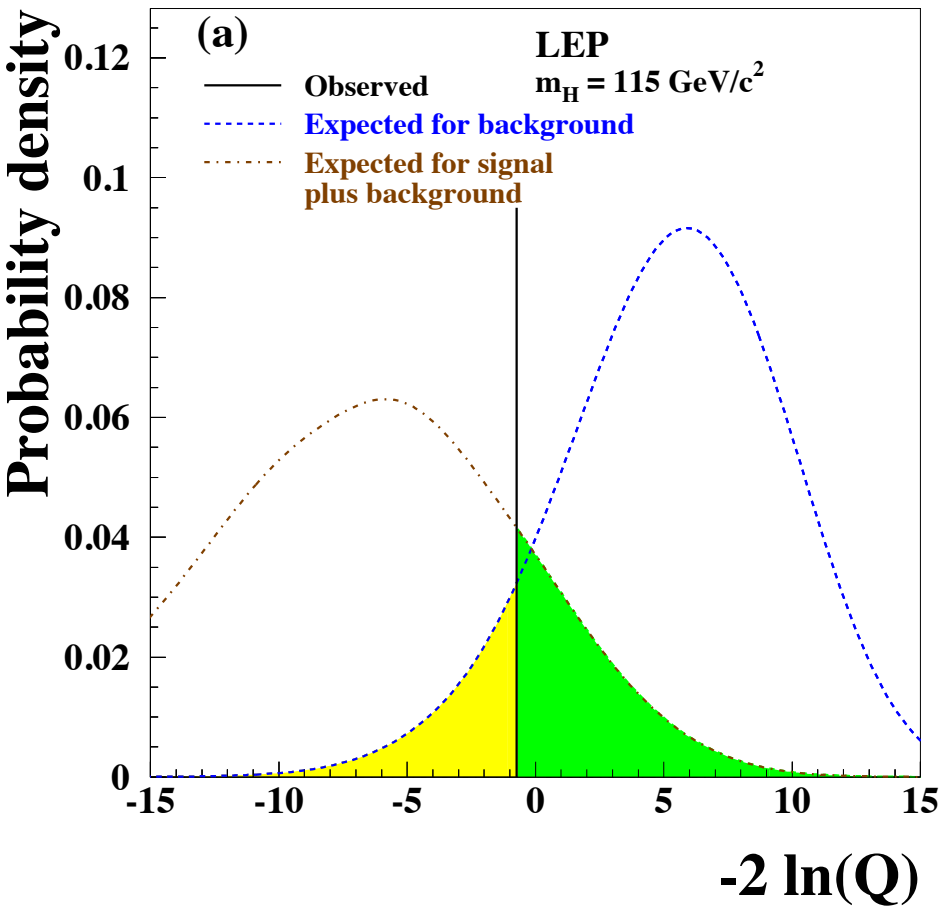


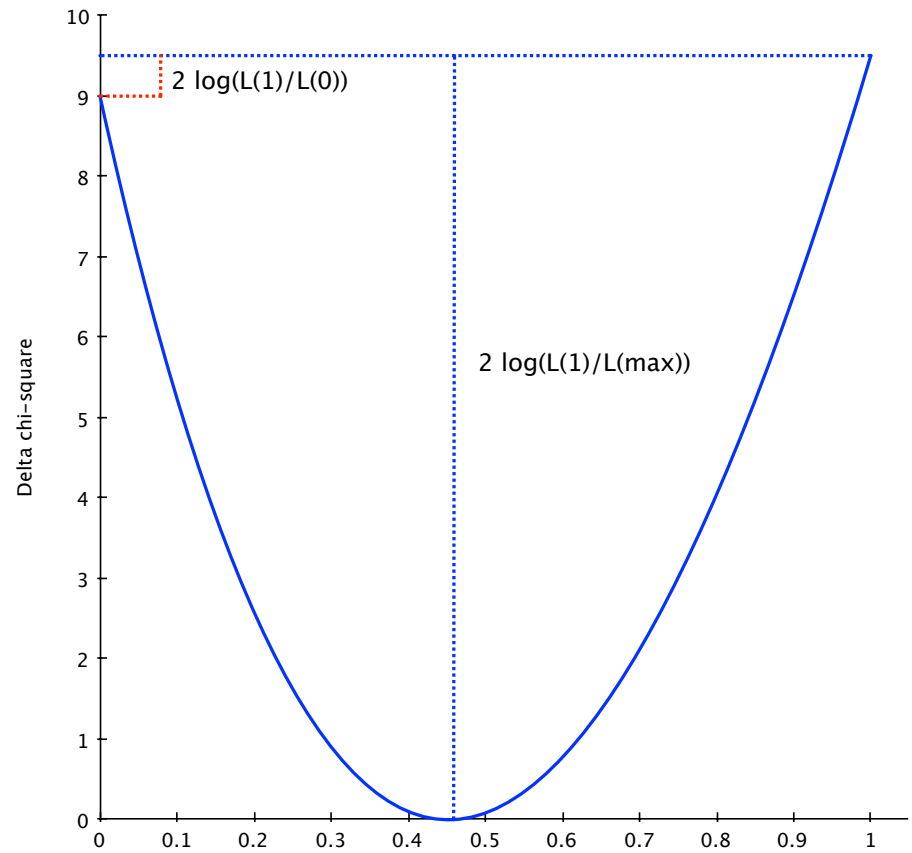
The LEP Test Statistic

The "LLR" used by LEP was $Q_{LEP} = L_{s+b}(\mu = 1)/L_b(\mu = 0)$

- This is definitely not distributed as a chi-square,

To get a chi-square distribution, the denominator needs to be the best-fit point and the variable is non-negative





$\mu = \sigma/\sigma_{SM}$

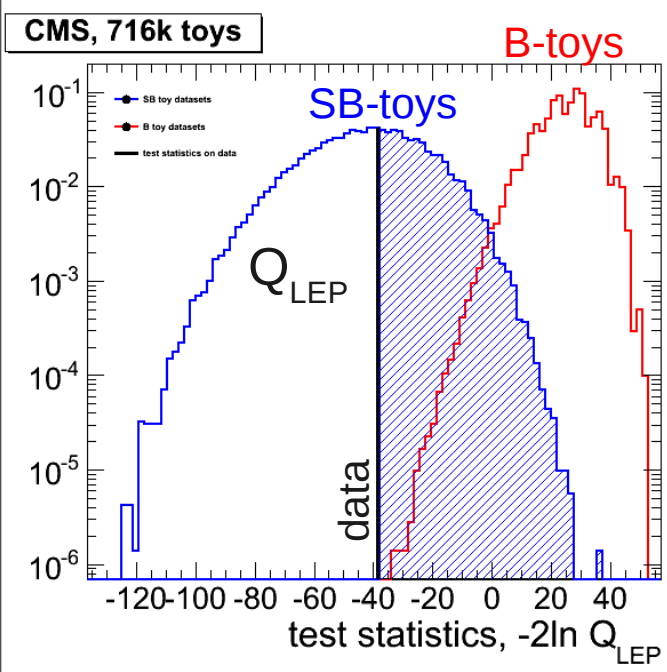
None are chi-square

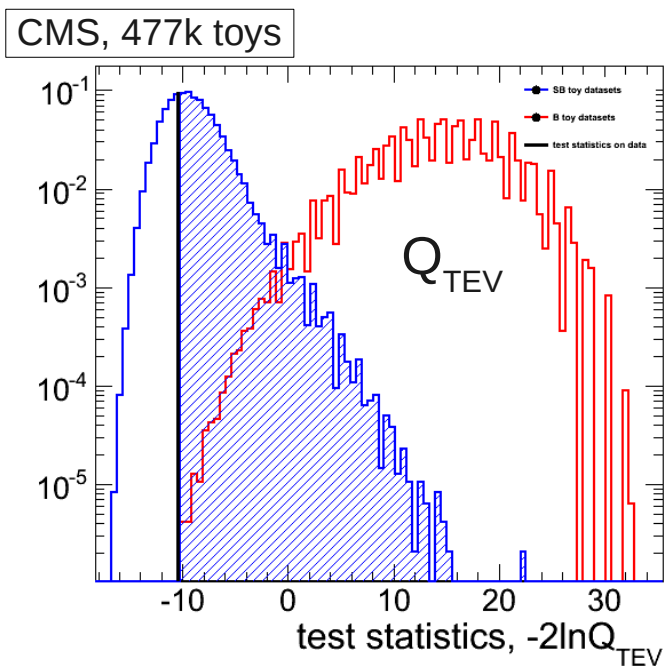
Not only are there three types of "LLR" test statistics around, but

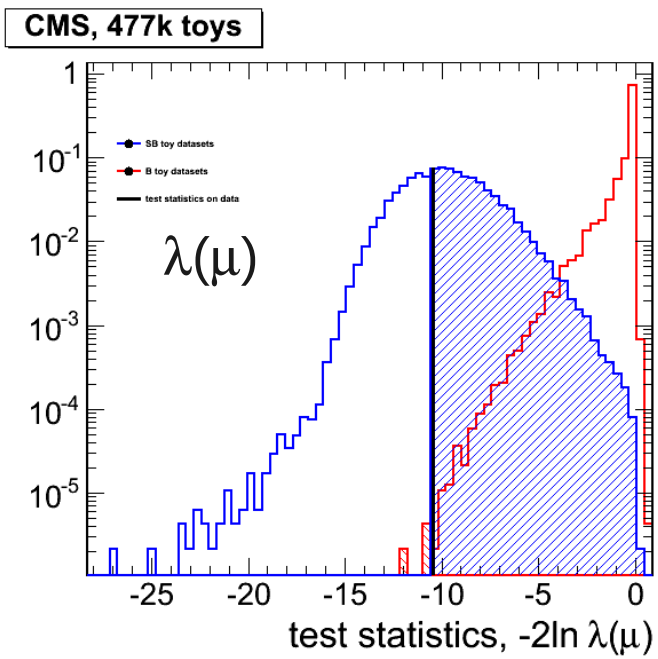
- they have very different distributions
- the only one that is expected to be distributed as chi-square is the profile $\lambda(\mu)$
  - and even it is typically distorted because fits don't allow a negative signal cross-section.
  - So you should expect the CLs+b from Toy MC to be different by ~x2 from the assumed distribution.

Forced to use approximations because we don't have the true likelihood functions.

- Makes no sense to ignore these problems and focus the precision of the underlying machinery of the lagrangian fitters







Do the Intervals Cover?

Michael Bridges, KC, Farhan Feroz, Mike Hobson, Roberto Ruiz de Austri, Roberto Trotta

See also next talk by Yashar Akrami

Coverage

Coverage is the probability that interval contains (covers) the true value

- Property of the method used to produce confidence/credible interval
- For any given data, the interval either covers or it doesn't
- Requires repeating the procedure on pseudo-data several times

Hard-core frequentist intervals "cover by construction"

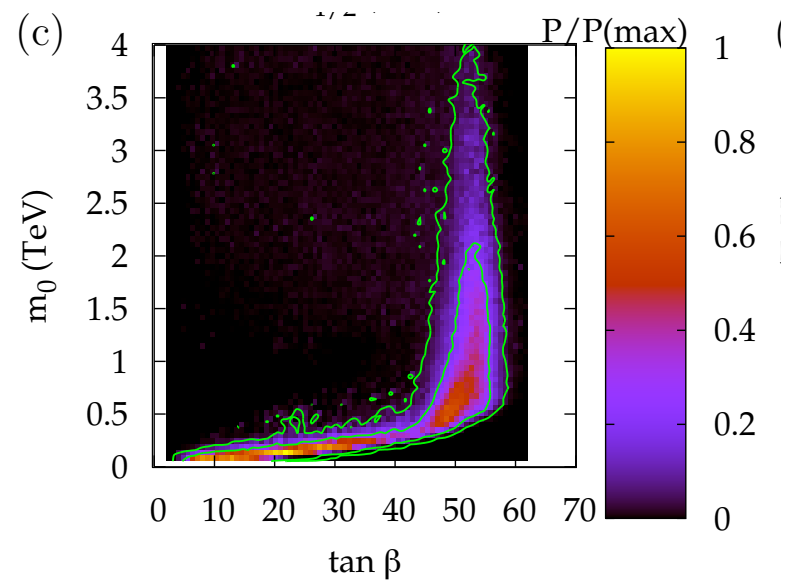
- Neyman Construction, aka "Feldman-Cousins" in HEP
- Likelihood-based intervals **are not** guaranteed to cover
- Bayesian methods **are not** guaranteed to cover (not their goal)

I think of coverage as a **calibration of our statistical apparatus**

- How often do the intervals cover the true value?
- Is it really 95%?

What should be the view today; Objective Bayesian analysis is the best frequentist tool around.

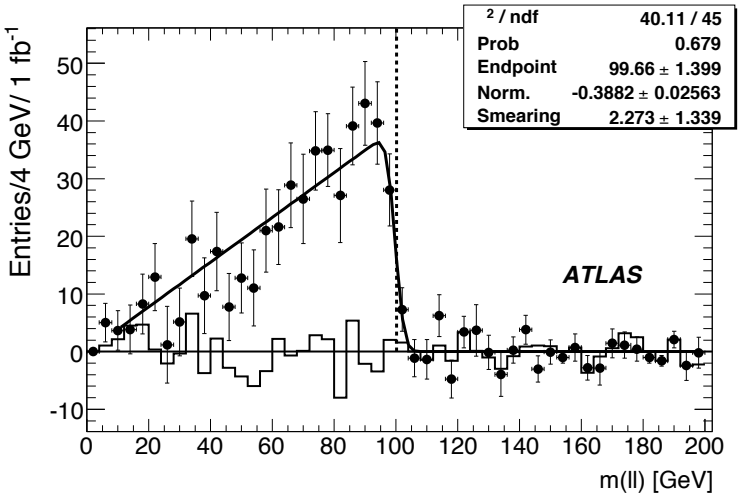
-Jim Berger



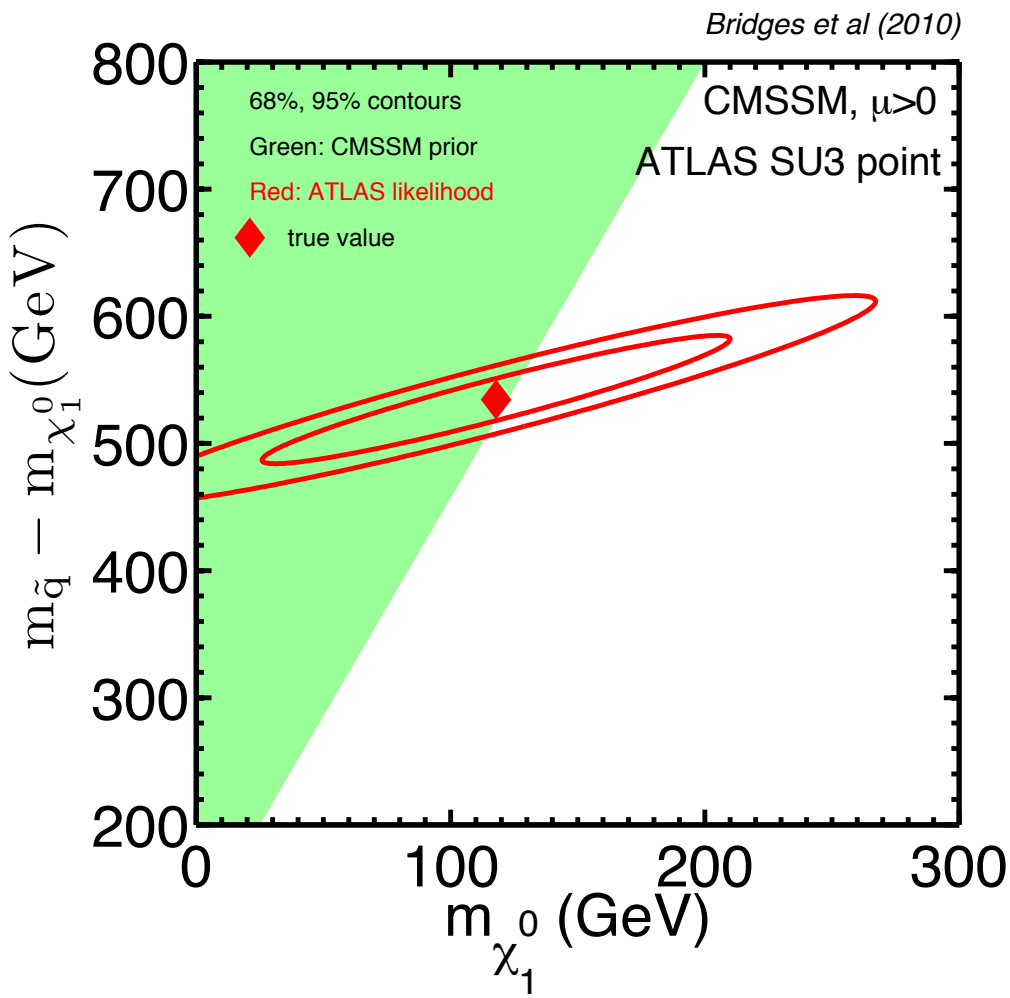
Do our current intervals cover?

To study this we considered a simplified model based on the ATLAS analysis of the "SU3" benchmark point.

- **Model**: A multivariate Gaussian likelihood function based on the published ATLAS covariance matrix SM. Is likelihood really a multi-variate Gaussian?



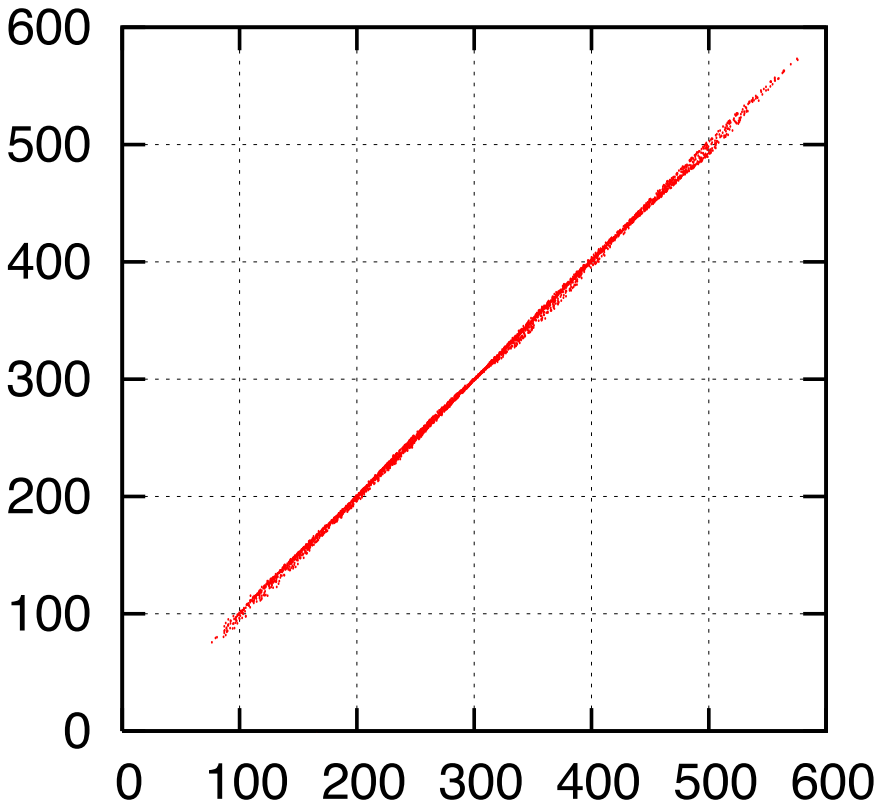
Observable	SU3 m_{meas} [GeV]
$m_{\tilde{\chi}_1^0}$	$88 \pm 60 \mp 2$
$m_{\tilde{\chi}_2^0}$	$189 \pm 60 \mp 2$
$m_{\tilde{q}}$	$614 \pm 91 \pm 11$
$m_{\tilde{\ell}}$	$122 \pm 61 \mp 2$
Observable	**SU3 Δm_{meas} [GeV]**
$m_{\tilde{\chi}_2^0} - m_{\tilde{\chi}_1^0}$	$100.6 \pm 1.9 \mp 0.0$
$m_{\tilde{q}} - m_{\tilde{\chi}_1^0}$	$526 \pm 34 \pm 13$
$m_{\tilde{\ell}} - m_{\tilde{\chi}_1^0}$	$34.2 \pm 3.8 \mp 0.1$



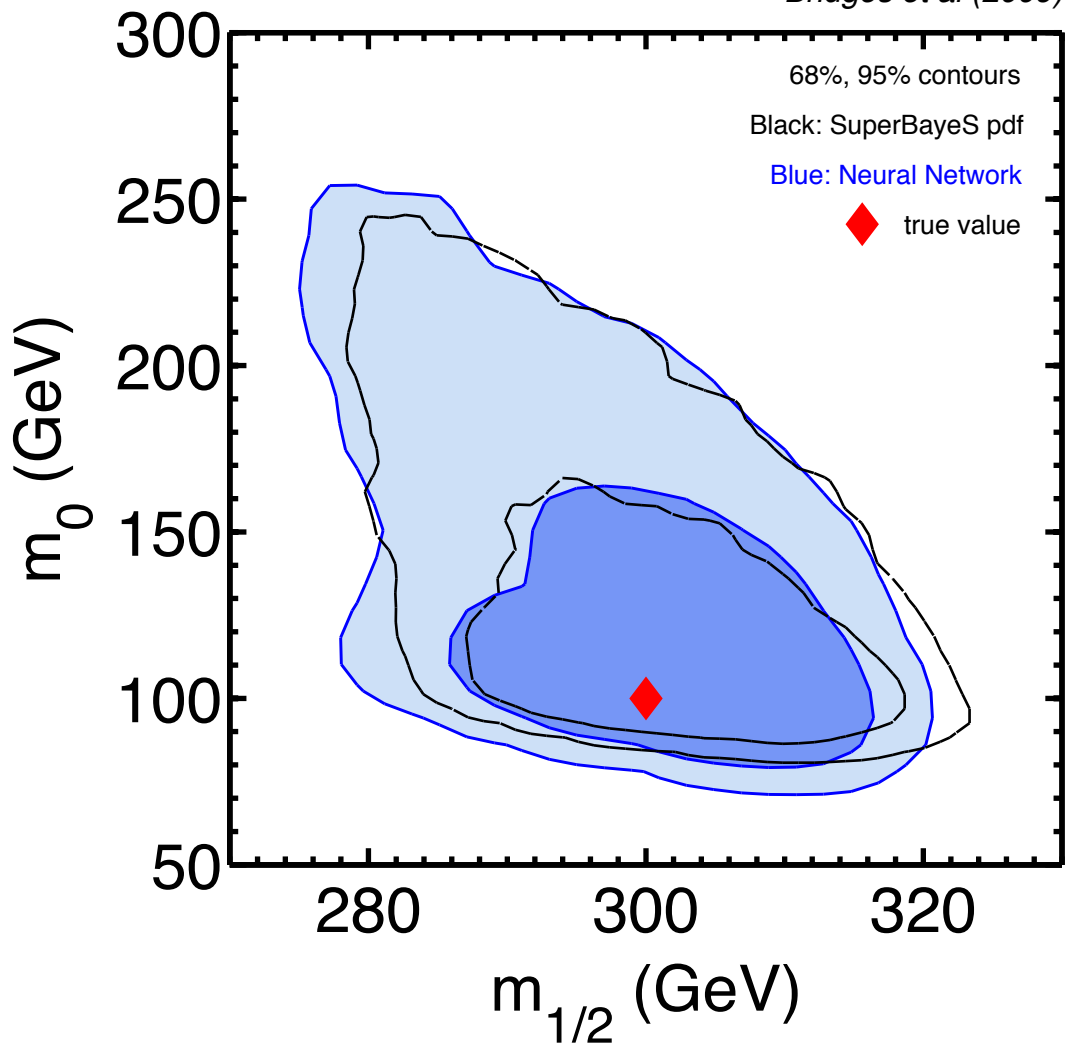
Neural Nets for CMSSM

Coverage studies are computationally intensive: 10,000 scans!

- use NN's to learn mapping from CMSSM <-> spectrum
  - speeds up scans dramatically: $O(10^5)$



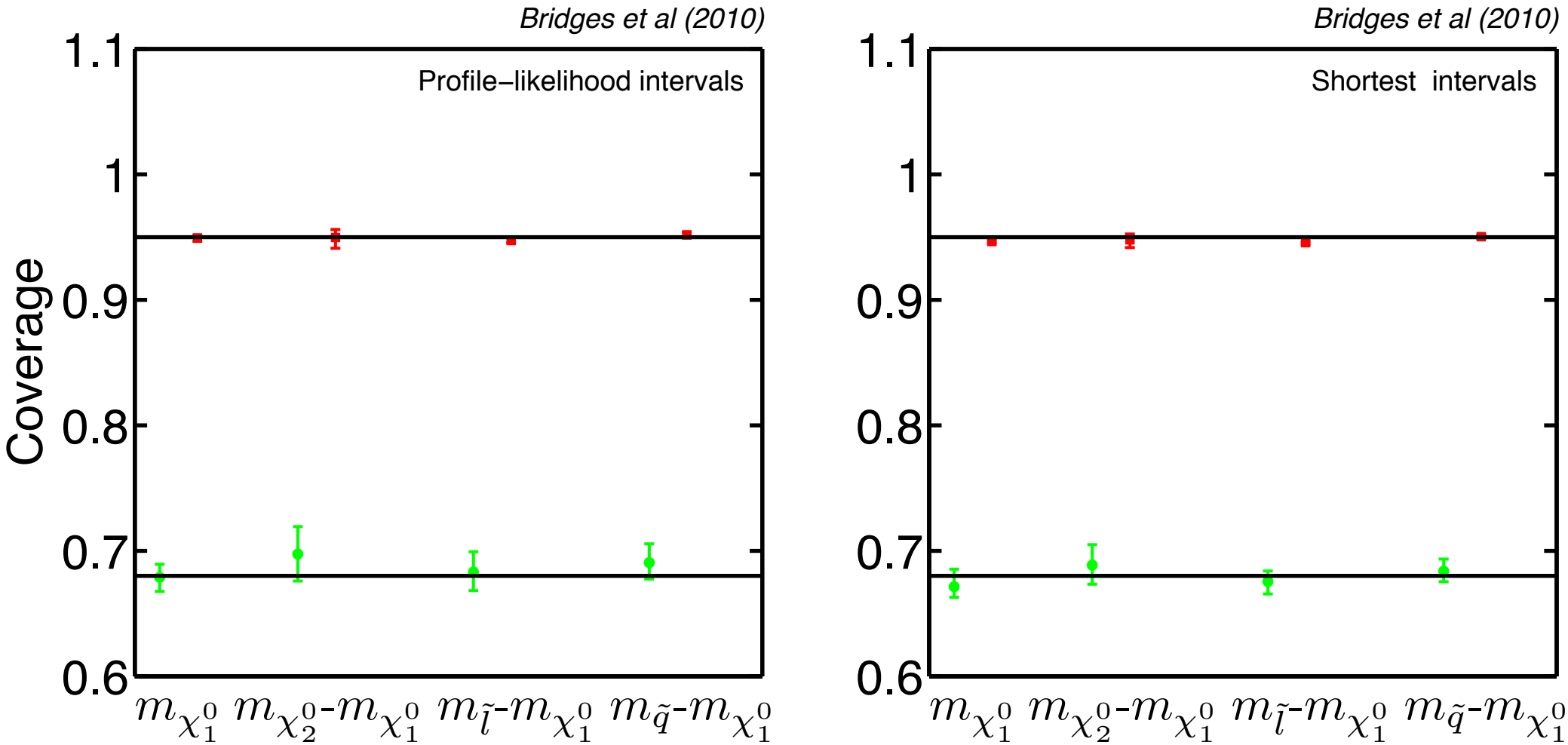
Bridges et al (2009)



Initial Coverage Result

Start by checking coverage of the weak-scale model:

- no pull-back to CMMS, parameters are mean of multivariate Gaussian

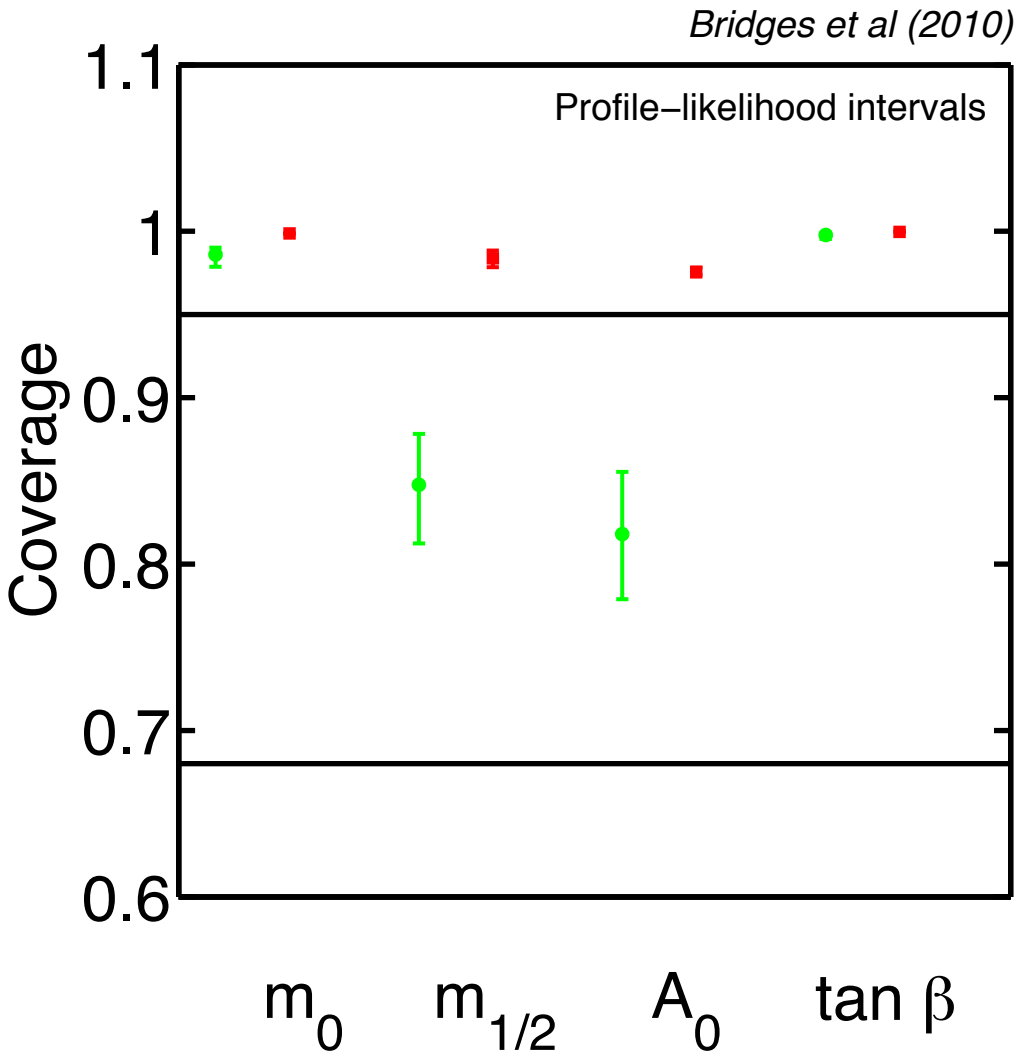


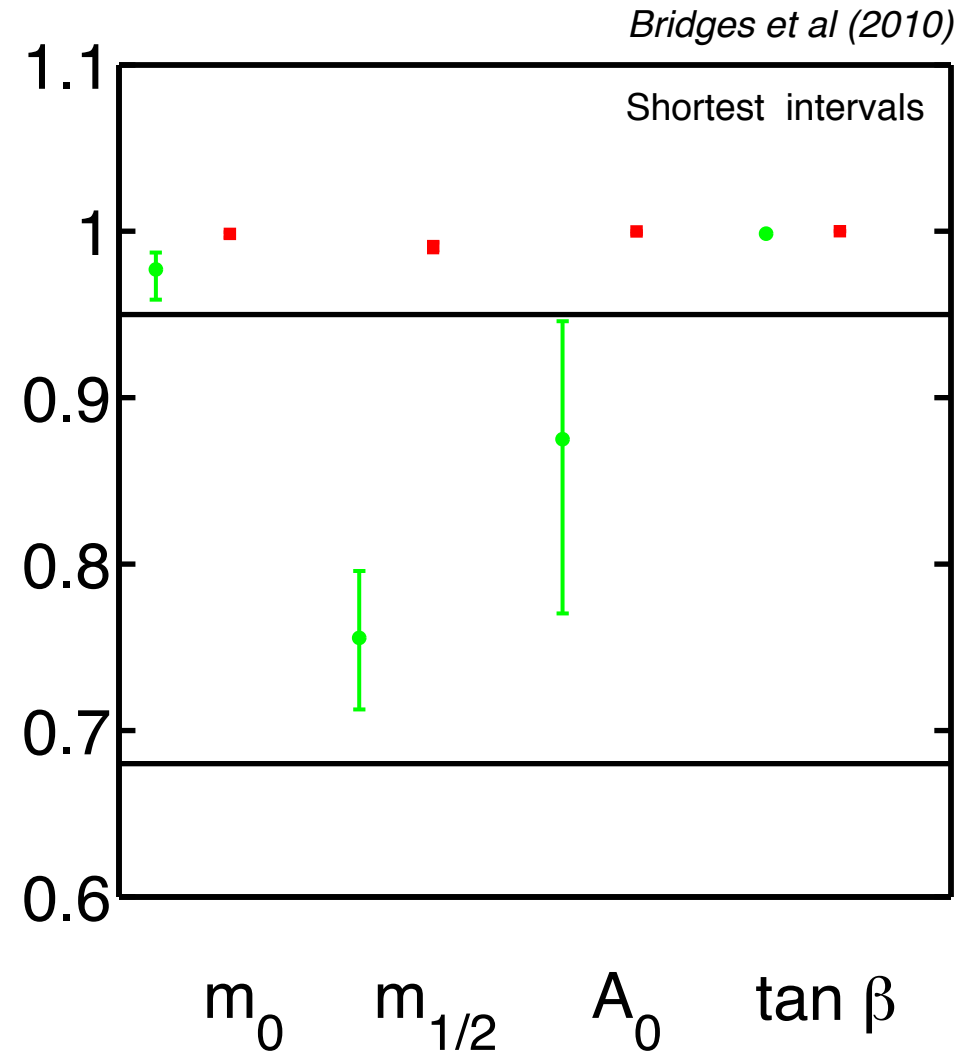
This “has to work”. If it didn’t, would be an algorithmic problem.

Initial Coverage Result

When we pull back to the CMSSM params., we see significant over-coverage

- consistently with profiling, MCMC, Multinest, etc.



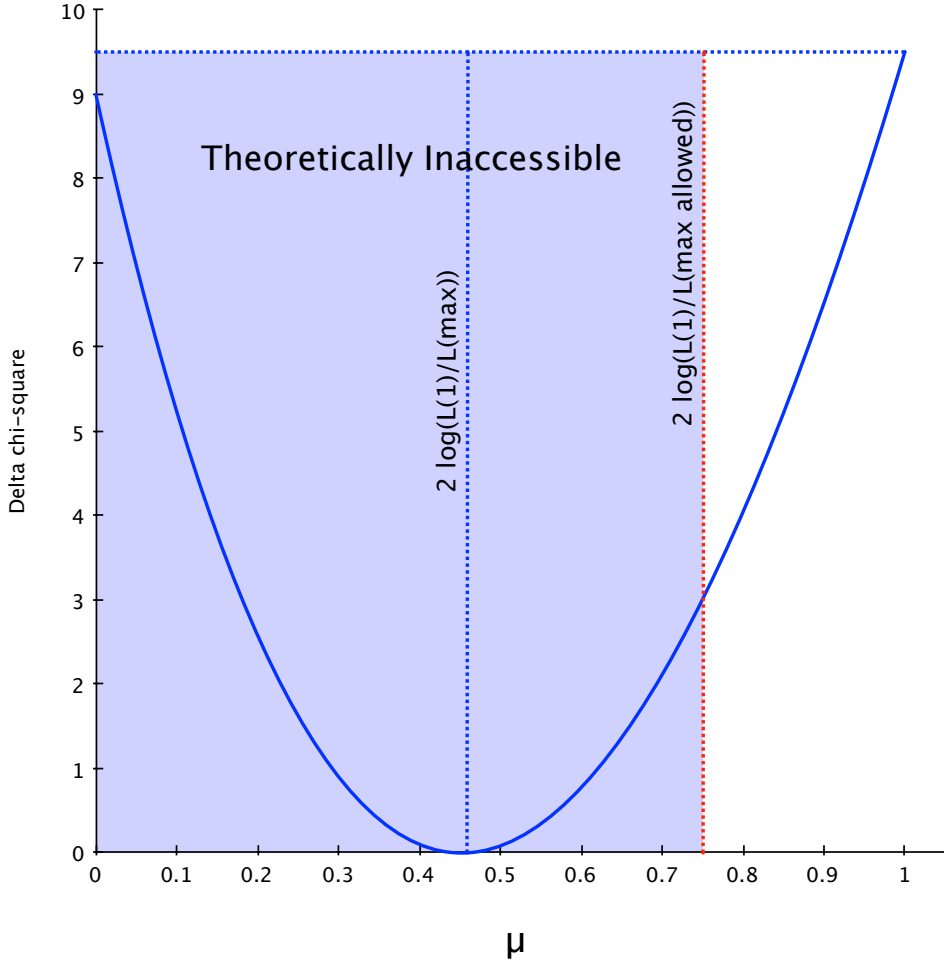


What is the source of the over-coverage?

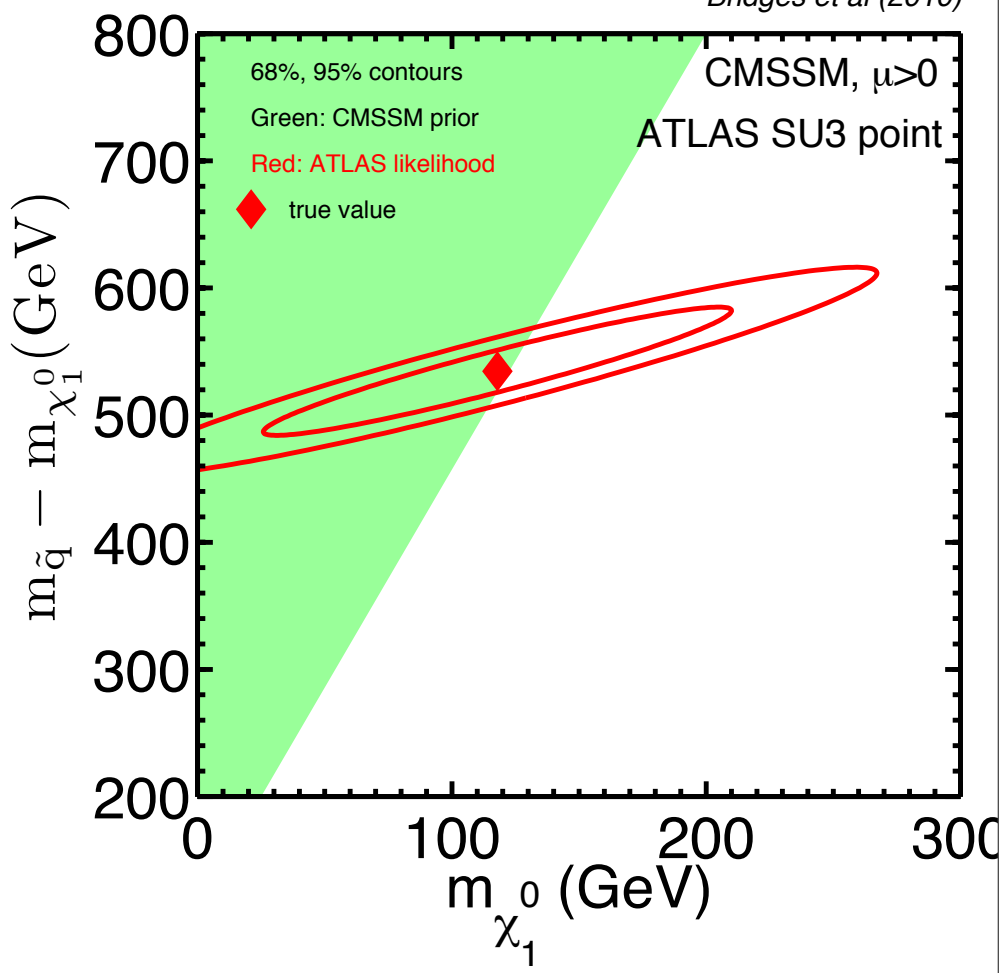
Effect of boundaries

The requirements that a CMSSM point is physical (LSP, EWSB, Tachions) introduce boundaries in the parameter space.

- These boundaries mean convergence to a χ^2 distribution (Wilks) is slow
- leads to a higher cut-off on -2 ln L ==> larger interval ==> over-coverage



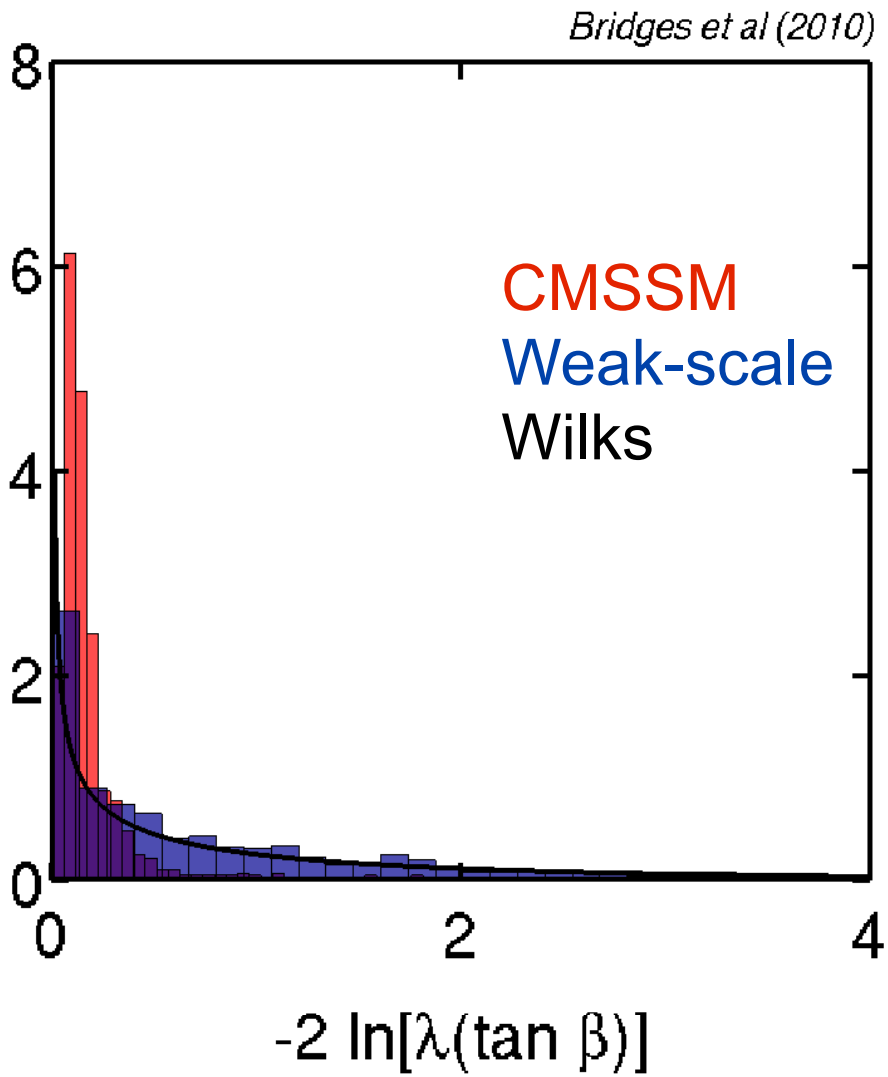
Bridges et al (2010)

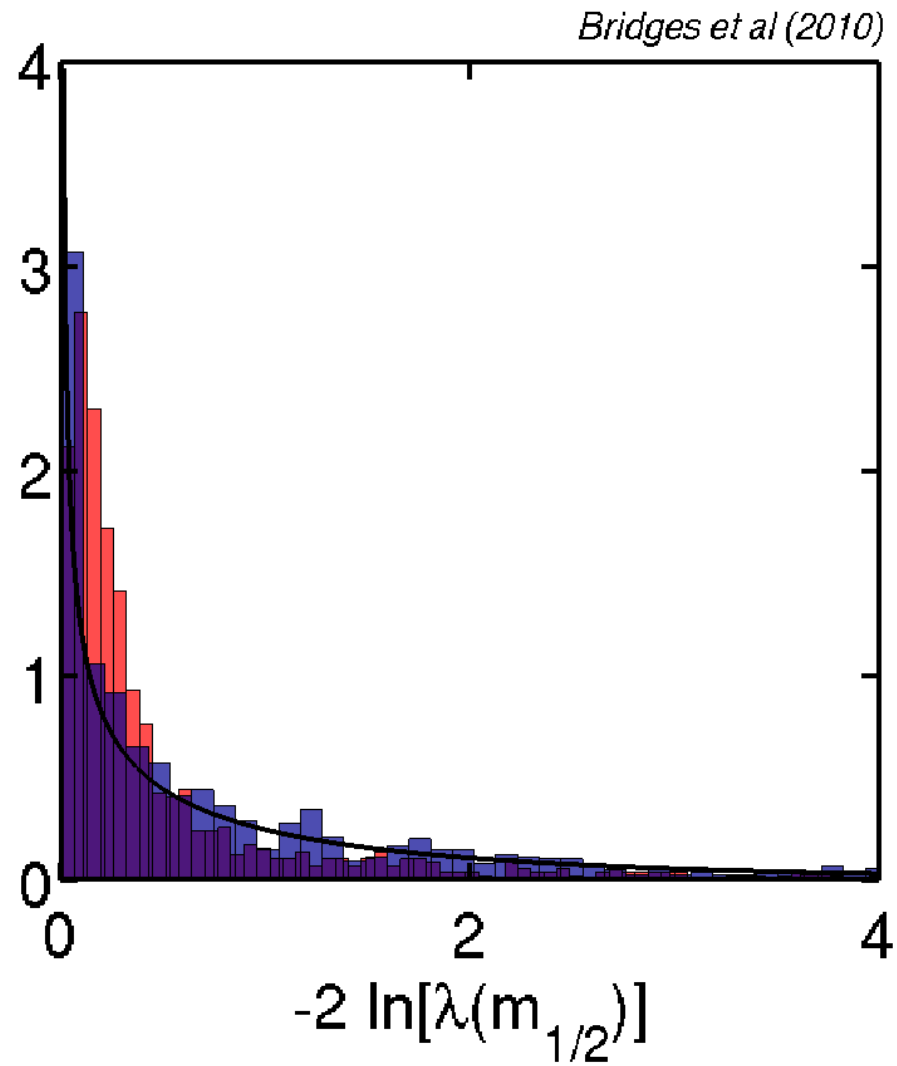


Checking "Validity" of Wilks

To check that this was the effect, we plot $-2\ln\lambda$ evaluated at true point

- Confirms expectation: distribution is **not** χ^2 for CMSSM (is for weak-scale)



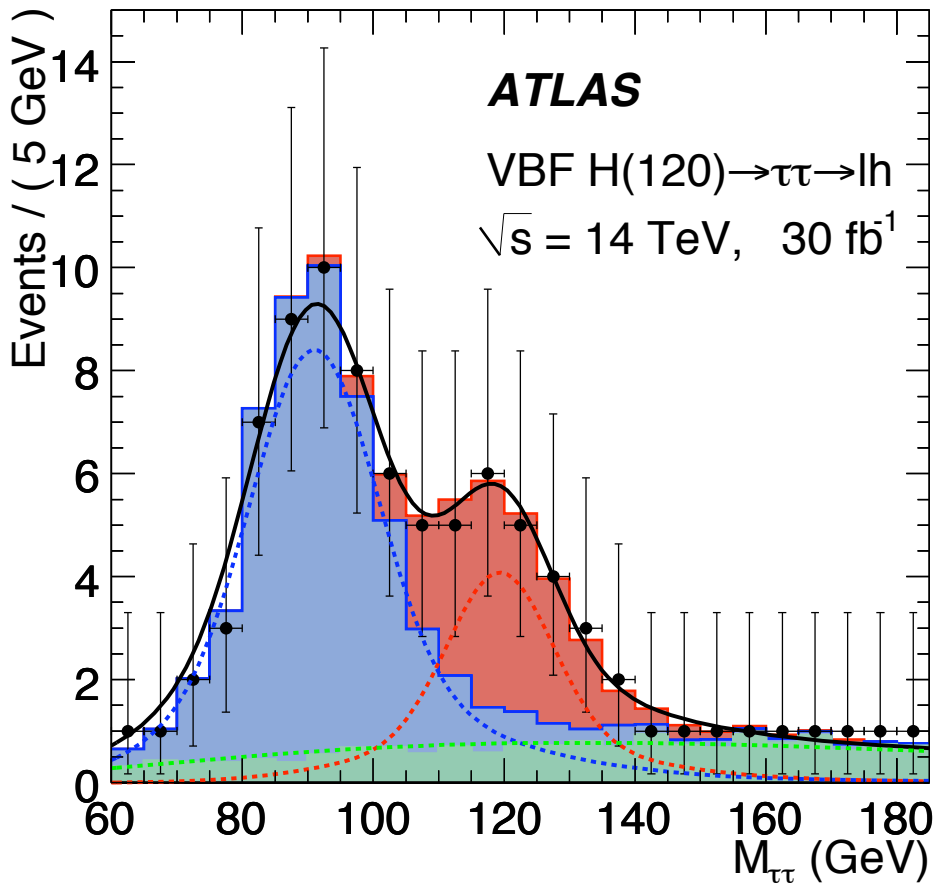


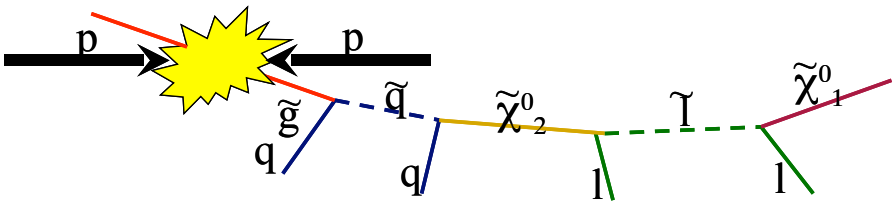
Moving Forward

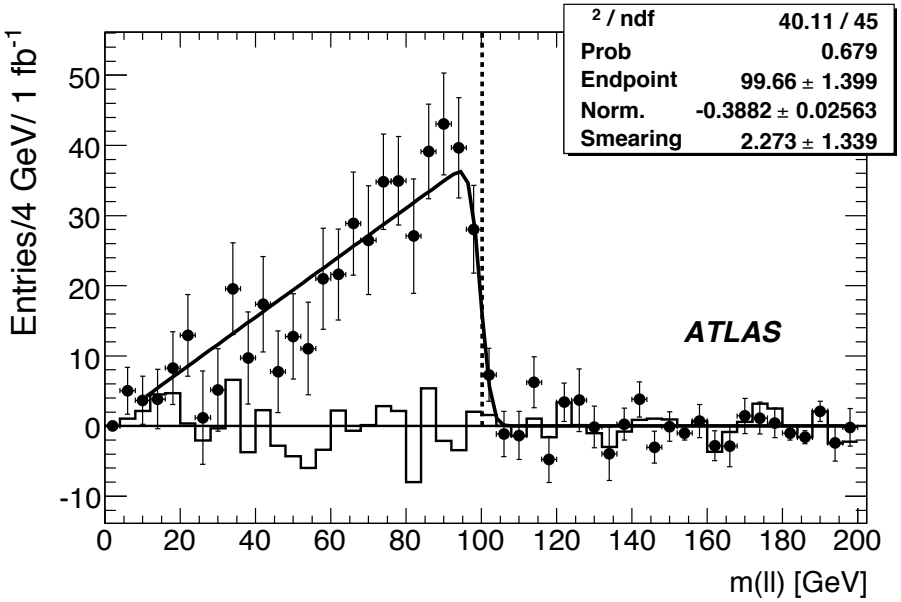
How will we model the data

The technology is general, but it is still up to the experiments to decide how they will model the data.

- Thoughtful parametrization requires planning and clear requests from groups like this



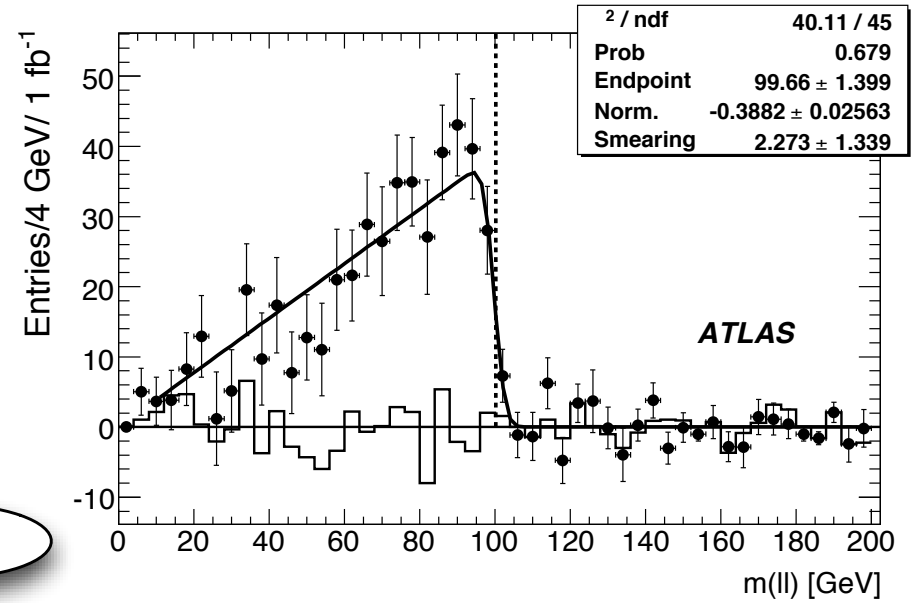




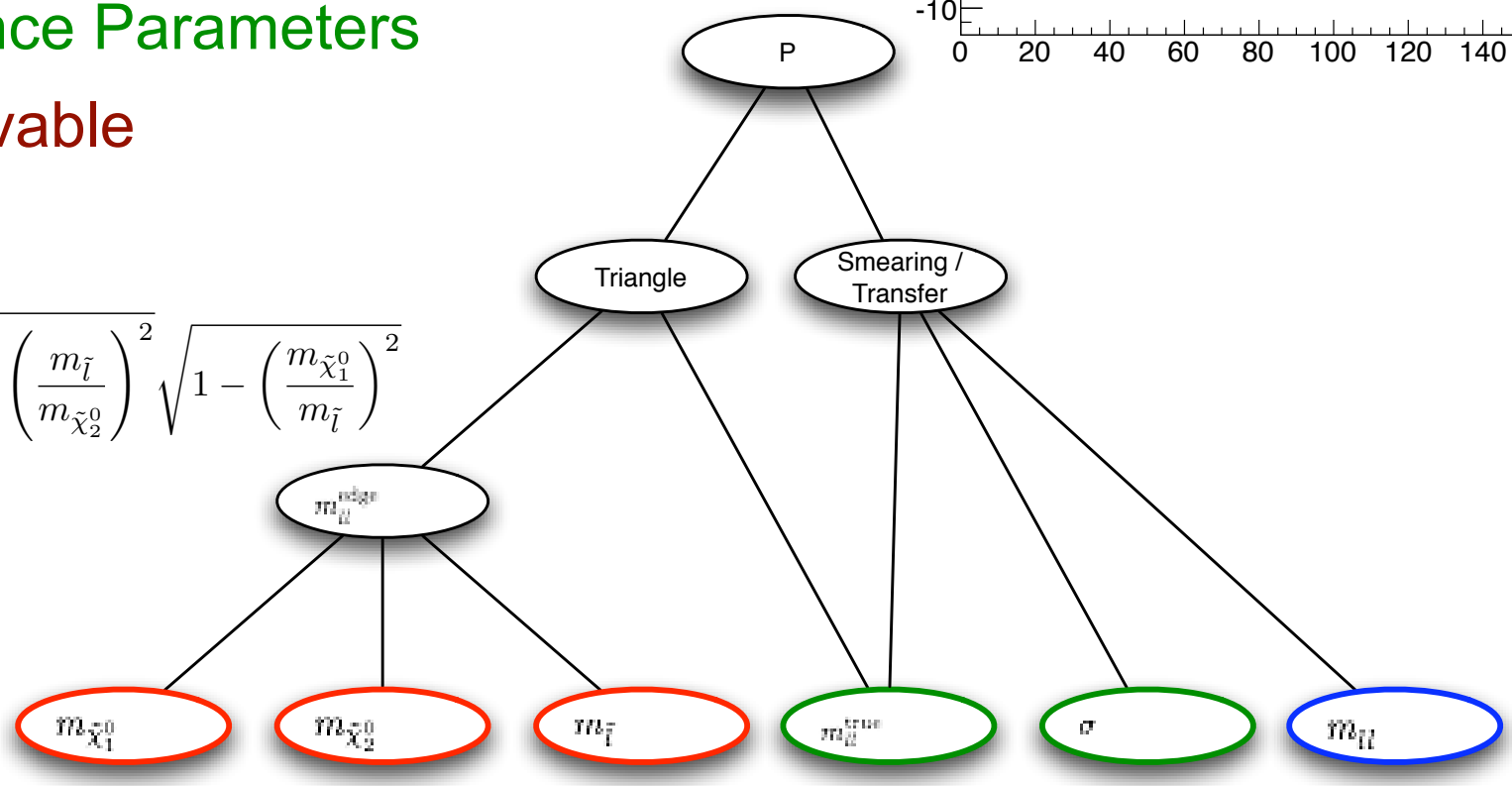
A graphical representation

Here is a graphical representation of this model that outlines its structural relationships

- Functions
- Parameters of Interest
- Nuisance Parameters
- Observable



$$m_{ll}^{\text{edge}} = m_{\tilde{\chi}_2^0}\sqrt{1-\left(\frac{m_{\tilde{l}}}{m_{\tilde{\chi}_2^0}}\right)^2}\sqrt{1-\left(\frac{m_{\tilde{\chi}_1^0}}{m_{\tilde{l}}}\right)^2}$$



$$P(m_{ll}|m_{\tilde{\chi}_1^0}, m_{\tilde{\chi}_2^0}, m_{\tilde{l}}, \sigma) = \text{Triangle}(m_{ll}^{true}, m_{ll}^{\text{edge}}(m_{\tilde{\chi}_1^0}, m_{\tilde{\chi}_2^0}, m)) \oplus \text{Smearing}(m_{ll}^{true}, m_{ll})$$

Matrix Element Method as a Parametrization

$$m_{ll}^{\text{edge}} = m_{\tilde{\chi}_2^0}\sqrt{1-\left(\frac{m_{\tilde{l}}}{m_{\tilde{\chi}_2^0}}\right)^2}\sqrt{1-\left(\frac{m_{\tilde{\chi}_1^0}}{m_{\tilde{l}}}\right)^2}$$

Matrix-element likelihood:
Calculate probability directly

P(event z | SM) = P (z | process A) + P(z | process B) + ...

where

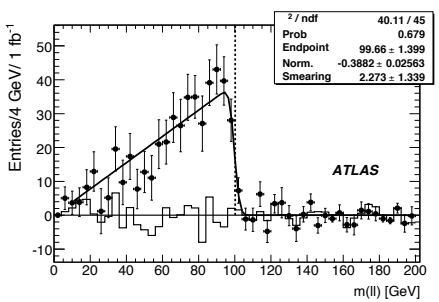
$$P(z \mid A) = \int dy \; |\mathcal{M}_A|^2 \, f_p f_p \, f_{TF}(y,z) = d\sigma_A/dz$$

Parton(y) to detector(z) transfer function (TF) describes parton-shower and detector response in parametrized form *(Issue 2)*

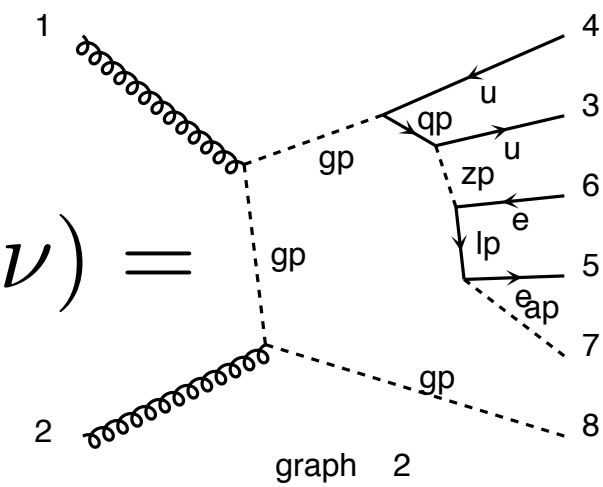
Matrix-element*PDFs for process A *(Issue 2)*

Integration over parton-level quantities

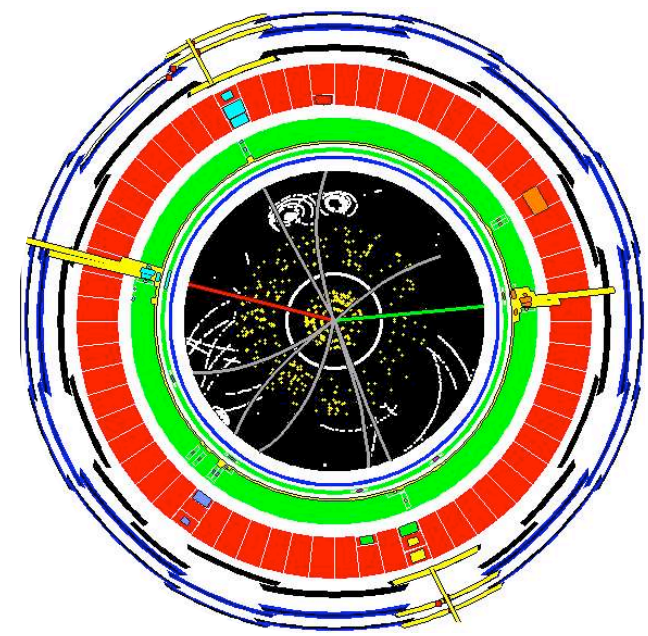
$P($



$\mu, \nu) =$



$\oplus$



Matrix Element Method as a Parametrization

$$m_{ll}^{\text{edge}} = m_{\tilde{\chi}_2^0}\sqrt{1-\left(\frac{m_{\tilde{l}}}{m_{\tilde{\chi}_2^0}}\right)^2}\sqrt{1-\left(\frac{m_{\tilde{\chi}_1^0}}{m_{\tilde{l}}}\right)^2}$$

Matrix-element likelihood:
Calculate probability directly

P(event z | SM) = P (z | process A) + P(z | process B) + ...

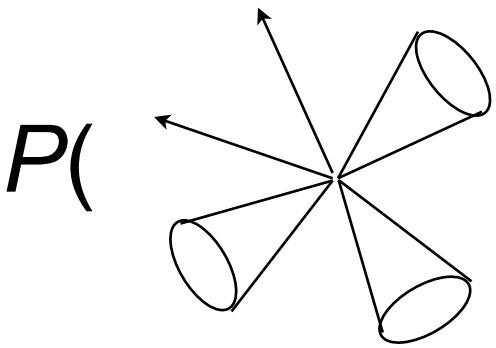
where

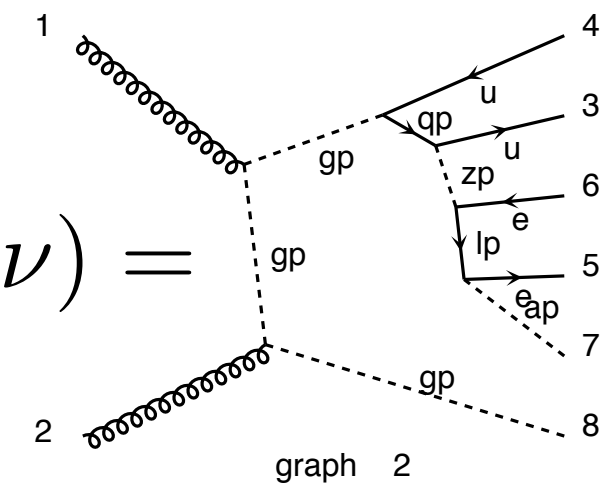
$$P(z \mid A) = \int dy \; |\mathcal{M}_A|^2 \, f_p f_p \, f_{TF}(y,z) = d\sigma_A/dz$$

Parton(y) to detector(z) transfer function (TF) describes parton-shower and detector response in parametrized form *(Issue 2)*

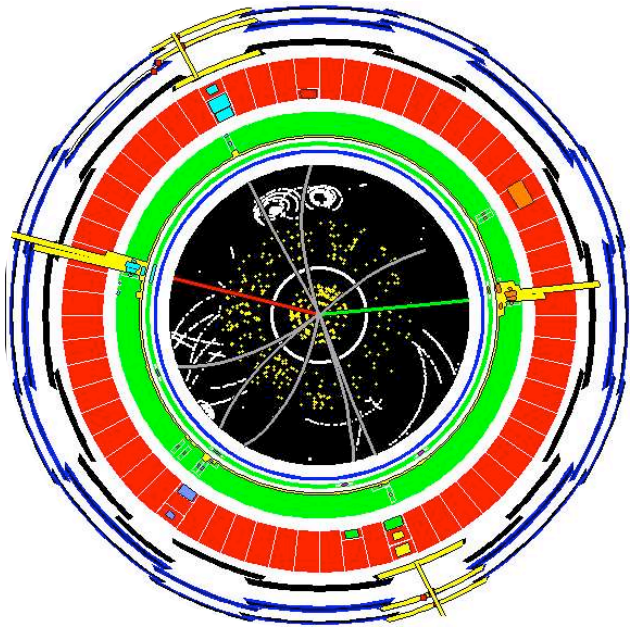
Matrix-element*PDFs for process A *(Issue 2)*

Integration over parton-level quantities





⊕



Matrix Element Method as a Parametrization

With the same di-lepton mass distribution, we can either:

- relate edge according to:

$$m_{ll}^{\text{edge}} = m_{\tilde{\chi}_2^0}\sqrt{1-\left(\frac{m_{\tilde{l}}}{m_{\tilde{\chi}_2^0}}\right)^2}\sqrt{1-\left(\frac{m_{\tilde{\chi}_1^0}}{m_{\tilde{l}}}\right)^2}$$

- incorporate matrix element techniques
  - naturally, could include more kinematic info -> more power.

Matrix-element likelihood:
Calculate probability directly

P(event z | SM) = P (z | process A) + P(z | process B) + ...

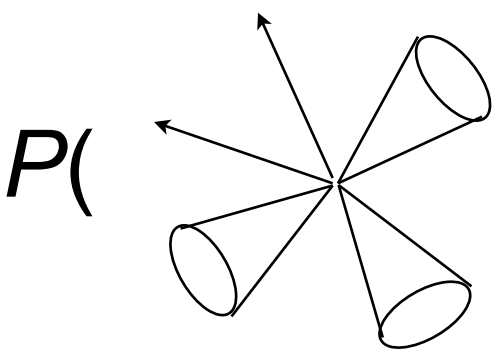
where

$$P(\ z\ |\ A\ ) = \int dy\ |\mathcal{M}_A|^2\ f_p f_p\ f_{TF}(y,z) = d\sigma_A/dz$$

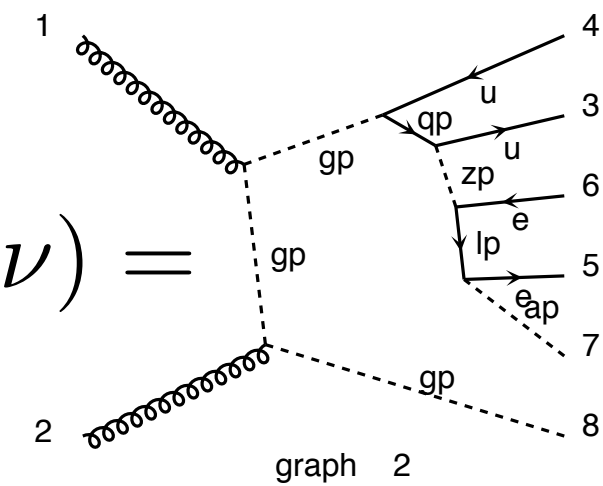
Parton(y) to detector(z) transfer function (TF) describes parton-shower and detector response in parametrized form *(Issue 2)*

Matrix-element*PDFs for process A *(Issue 2)*

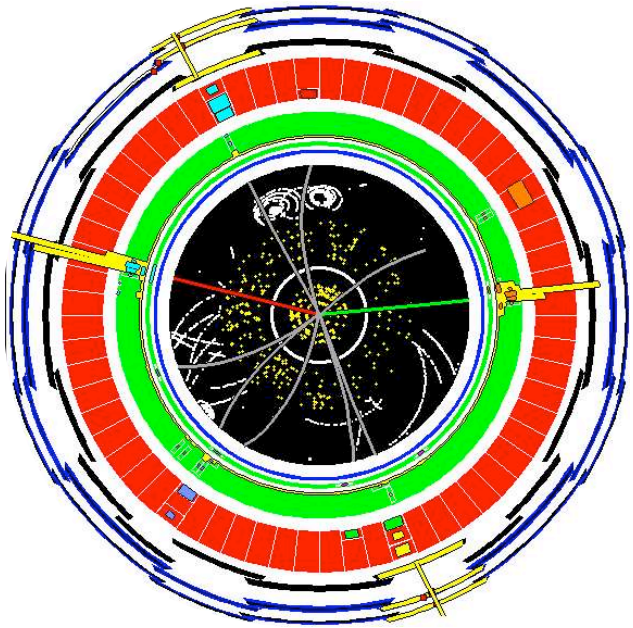
Integration over parton-level quantities



$P(\ \ \mu, \nu) =$



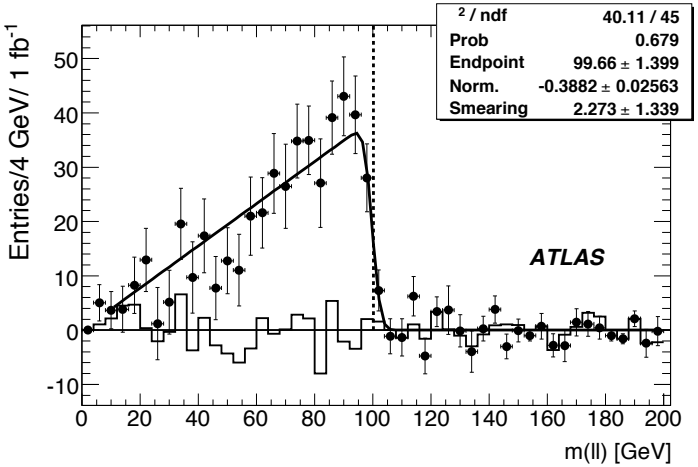
$\oplus$



Moving Forward

It seems the best way to move forward is to prepare some workspaces corresponding to benchmark tests.

- This will be very helpful for comparison of different fitters
- a very simple multivariate Gaussian for degugging
- eg. a SUSY ll_{max}, ql_{low}, ql_{high}, qll_{thresh}, qll_{edge} example
  - start with only the measured edges
  - could extend to the full shape
- A Higgs example:
  - h/H -> tau tau with both taus visible?
  - the toy ATLAS+CMS combination?



Asimov, Fisher, Wilks, Wald, Cramér, and Rao

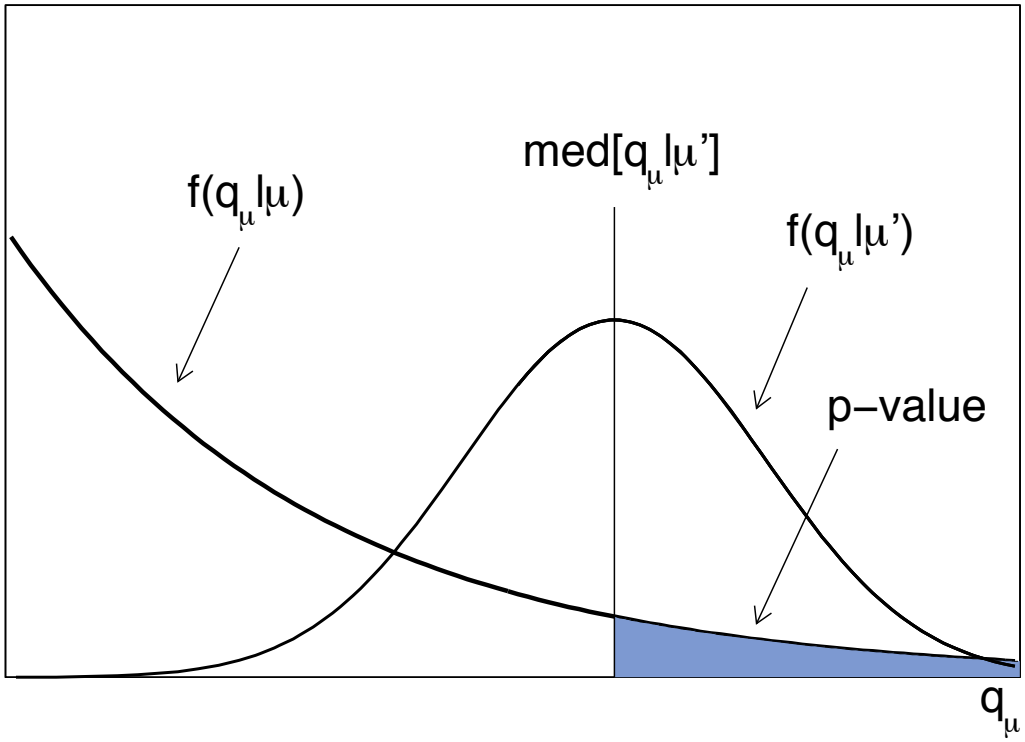
Glen Cowan, KC, Eilam Gross, Ofer Vitells: [arXiv:1007.1727]

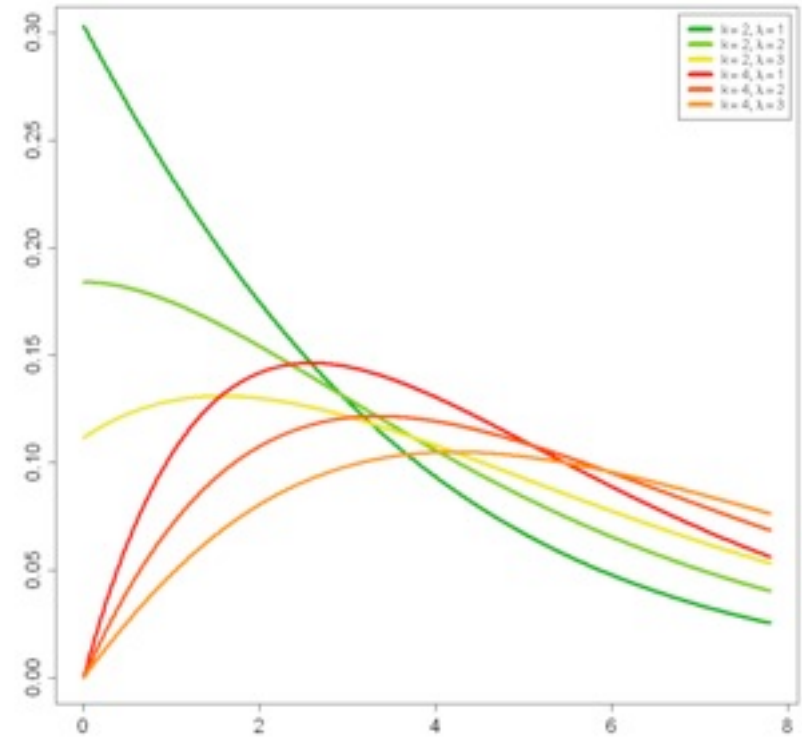
Ben Allanach, KC

From Wilks to Wald

Wilks only tells you the asymptotic distribution for the true point!

- for expected contours, one needs to know what distribution looks like for other points
- Walds theorem: non-central chi-square with noncentrality parameter Λ





$$\Lambda_r = \sum_{i,j=1}^{r} (\theta_i - \theta_i')\, \tilde{V}_{ij}^{-1}\, (\theta_j - \theta_j')$$

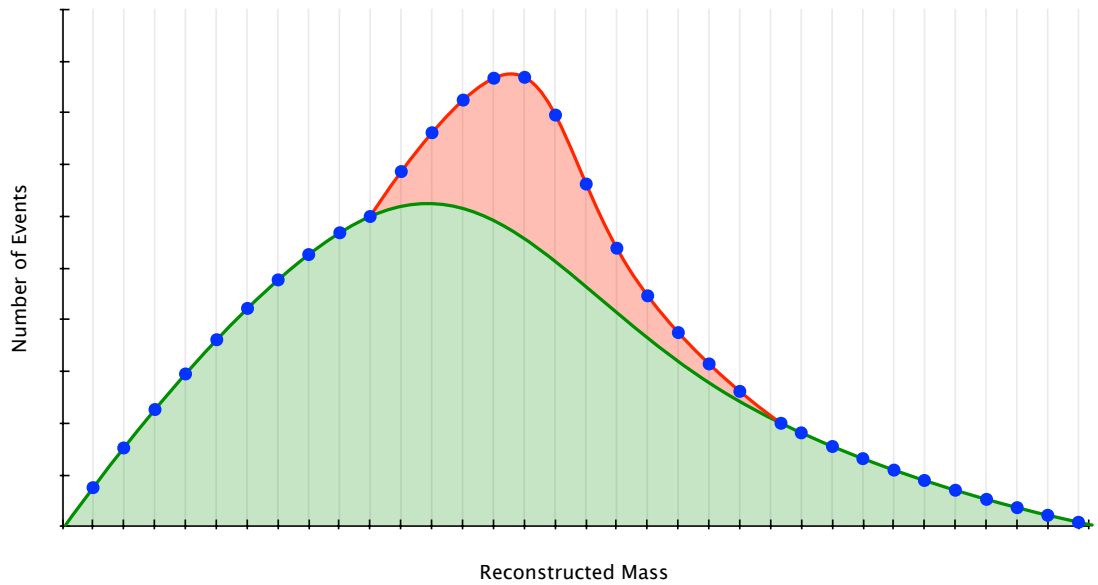
$$g_{ij}(\boldsymbol{\theta}) = V_{ij}^{-1}(\boldsymbol{\theta}) = -E\left[\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right] = \int dx L(x|\boldsymbol{\theta}) \left[\frac{\partial \ln L}{\partial\theta_i}\right]\left[\frac{\partial \ln L}{\partial\theta_j}\right]$$

The Asimov dataset

The name of the "Asimov" data set is inspired by the short story *Franchise*, by Isaac Asimov.

Glen Cowan, KC, Eilam Gross, Ofer Vitells
http://arxiv.org/abs/1007.1727

"Multivac picked you as the most representative this year. Not the smartest, or the strongest, or the luckiest, but just the most representative. Now we don't question Multivac, do we?"



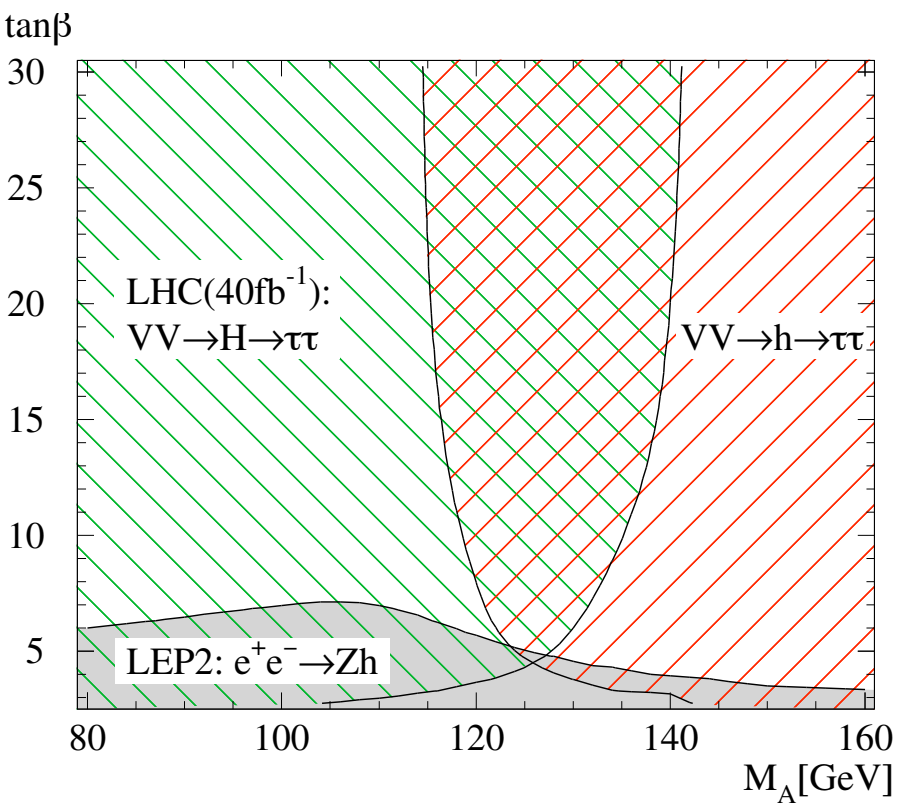
Coincidentally, the story takes place in 2008, when we started to formalize the properties of our "Asimov" Dataset

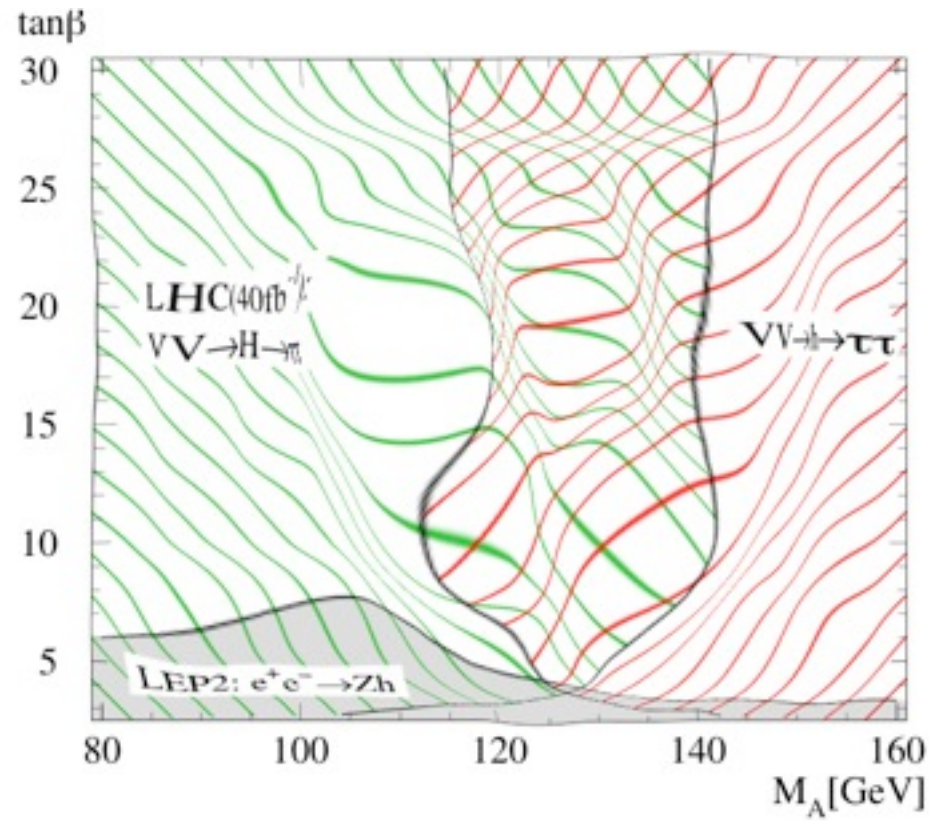
The Fisher information metric

Our theories are parametrized in some form convenient for our underlying quantum field theories. But this parametrization is somewhat arbitrary, and

- phenomenology nearly constant in large regions and changes quickly in others.
- a metric (not prior) on the parameters which is inherited from observables
- invariant to reparametrizing observables, covariant to reparametrizing theory

$$g_{ij}(\boldsymbol{\theta}) = V_{ij}^{-1}(\boldsymbol{\theta}) = -E\left[\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right] = \int dx L(x|\boldsymbol{\theta}) \left[\frac{\partial \ln L}{\partial\theta_i}\right]\left[\frac{\partial \ln L}{\partial\theta_j}\right]$$



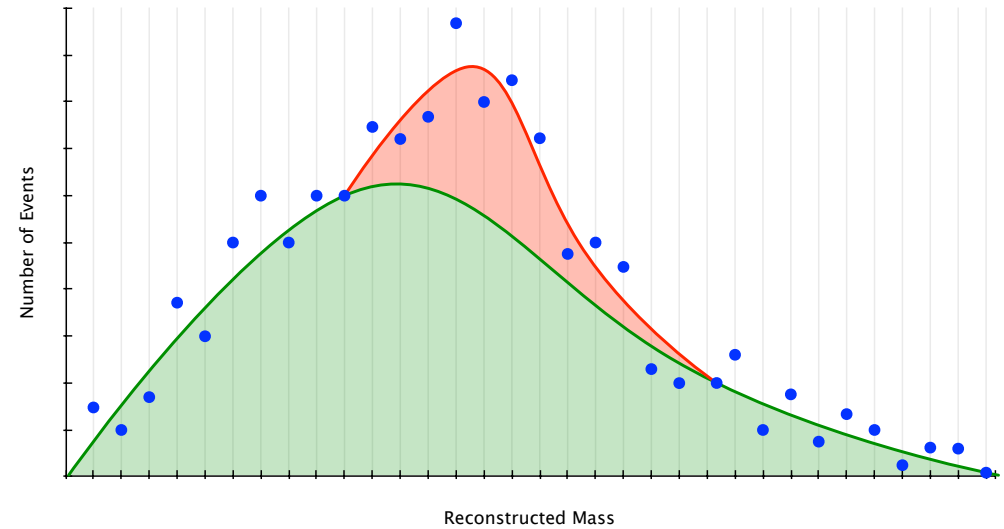


Spinoffs from the Asimov idea

Calculating the Fisher info. matrix requires an expectation over possible data.

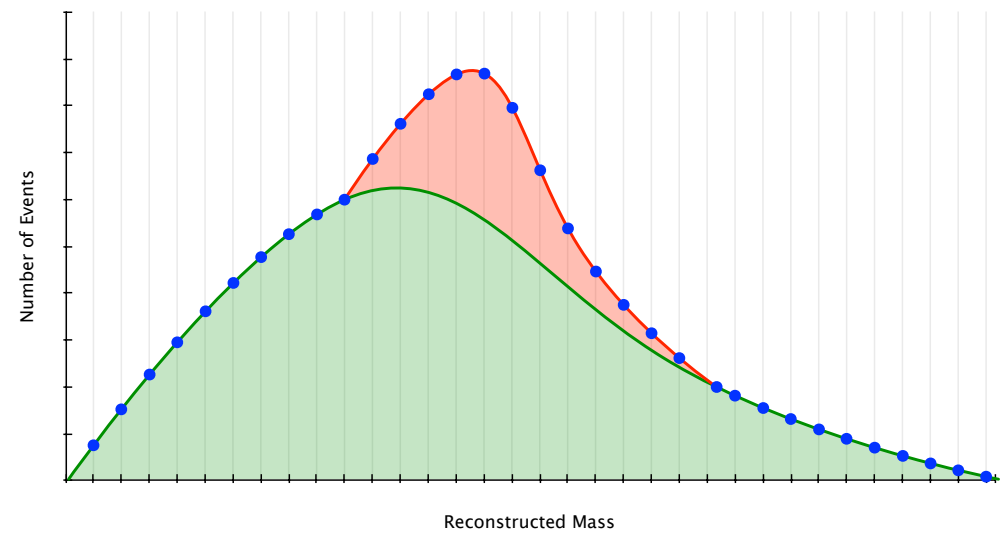
$$g_{ij}(\theta) = \mathrm{E}\left[\left(\frac{\partial}{\partial\theta_i}\ln L(\theta)\right)\left(\frac{\partial}{\partial\theta_j}\ln L(\theta)\right)\middle|\,\theta\right].$$

In many problems, this is too computationally expensive to be useful.



We found that the curvature of the likelihood function on the Asimov data gives a very good estimate of g_{ij}

$$g_{ij}(\theta) \approx \left(\frac{\partial}{\partial\theta_i}\ln L_A(\theta)\right)\left(\frac{\partial}{\partial\theta_j}\ln L_A(\theta)\right)$$



In Banff, two statisticians Earl Lawrence and Richard Lockhart helped us see that this curvature of this single Asimov dataset can be seen as a numerical integration for calculating the expectation of the curvature.

This also provides a convenient algorithm determining for Jeffreys's prior numerically, but I know their are issues with numerics and improper priors.

Proof: Curvature of Asimov Likelihood = Fisher Information

In Banff, two statisticians Earl Lawrence and Richard Lockhart helped us see that this curvature of this single Asimov dataset can be seen as a numerical integration for calculating the expectation of the curvature.

$$L(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \mathrm{Pois}(n_j | \mu s_j + b_j) \prod_{k=1}^{M} \mathrm{Pois}(m_k | u_k)$$

$$\frac{\partial \ln L}{\partial \theta_j} = \sum_{i=1}^{N} \left(\frac{n_i}{\nu_i} - 1 \right) \frac{\partial \nu_i}{\partial \theta_j} + \sum_{i=1}^{M} \left(\frac{m_i}{u_i} - 1 \right) \frac{\partial u_i}{\partial \theta_j} = 0$$

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_j \partial \theta_k} &= \sum_{i=1}^{N} \left[\left(\frac{n_i}{\nu_i} - 1 \right) \frac{\partial^2 \nu_i}{\partial \theta_j \partial \theta_k} - \frac{\partial \nu_i}{\partial \theta_j} \frac{\partial \nu_i}{\partial \theta_k} \frac{n_i}{\nu_i^2} \right] \\ &+ \sum_{i=1}^{M} \left[\left(\frac{m_i}{u_i} - 1 \right) \frac{\partial^2 u_i}{\partial \theta_j \partial \theta_k} - \frac{\partial u_i}{\partial \theta_j} \frac{\partial u_i}{\partial \theta_k} \frac{m_i}{u_i^2} \right] \end{aligned}$$

$$V_{jk}^{-1} = -E\left[\frac{\partial^2 \ln L}{\partial \theta_j \partial \theta_k} \right] = -\frac{\partial^2 \ln L_{\mathrm{A}}}{\partial \theta_j \partial \theta_k} = \sum_{i=1}^{N} \frac{\partial \nu_i}{\partial \theta_j} \frac{\partial \nu_i}{\partial \theta_k} \frac{1}{\nu_i} + \sum_{i=1}^{M} \frac{\partial u_i}{\partial \theta_j} \frac{\partial u_i}{\partial \theta_k} \frac{1}{u_i}$$

Conclusions

The RooStats project has reached a certain level of maturity and is rapidly being adopted by the LHC experiments

- The toy ATLAS/CMS Higgs combination was a milestone for the project

The workspace technology that is so important for combinations (of different channels within an experiment or between experiments) also provides enormous opportunity for communicating experimental results to the fundamental lagrangian fitters.

- Given the effort that is going into making the RGE's more precise and the fitting techniques, we should make sure the inputs (likelihood functions) are sensible

Even if we get the likelihood right, we must remember that the intervals may not cover

- fundamental arguments aside, coverage is a useful and standard **calibration**

In order to move forward, we should

- agree on some useful benchmark examples and prepare the workspace for them so different tools can start working on their interfaces to the workspace
- realize that a thoughtful parametrization of the model requires planning and clear requests to the experimental community