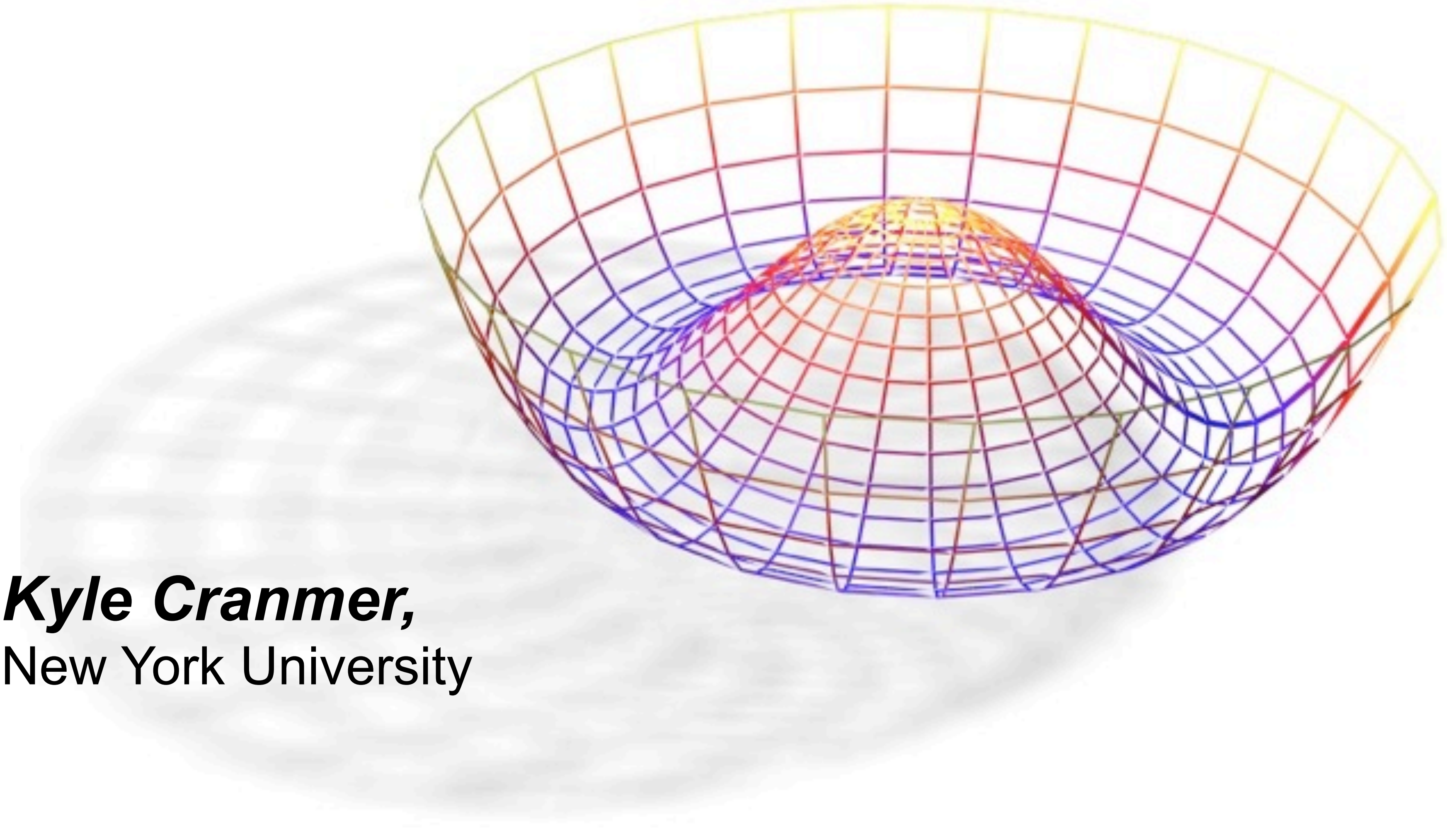
Center for
Cosmology and
Particle Physics

Some statistical / particle physics ideas and questions



Kyle Cranmer,
New York University

Topics:

- Follow up to the on/off problem
  - Hybrid (prior predictive) method for general problems
- Show and tell of some complicated particle physics models
- Graphical models?
- Fisher Information Matrix/Metric and the “Asimov” data set
- Dealing with the look-elsewhere effect via conditioning?

The "on/off" problem

Center for Cosmology and Particle Physics

This is a simplified problem that has been studied quite a bit to gain some insight into our more realistic and difficult problems

- number counting with background uncertainty
  - main measurement: observe n_{on} with $s+b$ expected
  - auxiliary measurement: observe n_{off} with τb expected

$$P(n_{\rm on}, n_{\rm off}|s, b) = {\rm Pois}(n_{\rm on}|s+b)\,{\rm Pois}(n_{\rm off}|\tau b).$$

- Note: $n_{\rm off}$ is used to constrain background uncertainty
  - In this approach "background uncertainty" is a statistical error

We learned that exact frequentist solution (construction) is formally identical to prior predictive treatment with flat prior

- eg. choose $\pi(b)$ as posterior from a flat prior and n_{off} term

$$P(n_{\rm on}|s) = \int db\,{\rm Pois}(n_{\rm on}|s+b)\,\pi(b),$$

Bayesian-Frequentist Hybrid Solutions

Goal of Bayesian-frequentist hybrid solutions is to provide a frequentist treatment of the main measurement, while eliminating nuisance parameters (deal with systematics) with an intuitive Bayesian technique.

$$P(n_{\rm on}|s) = \int db \, {\rm Pois}(n_{\rm on}|s+b) \, \pi(b),$$

Recommendations:

- clearly state prior $\eta(b)$; identify control samples or other auxiliary measurements, then base prior on

$$\pi(b) = P(b|n_{\rm off}) = \frac{P(n_{\rm off}|b)\eta(b)}{\int db P(n_{\rm off}|b)\eta(b)}.$$

Generalizing the Hybrid approach

In RooStats we are providing several techniques given a common specification of the problem that relies on:

- the joint model $P(x, y|s, b, \tau)$
- a Bayesian prior $\eta(s, b)$
- and some data (x_0, y_0)

The question is "how do we generalize the Hybrid (prior predictive) approach" given this information

Generalizing the Hybrid approach

Center for Cosmology and Particle Physics

Start with “on/off” example

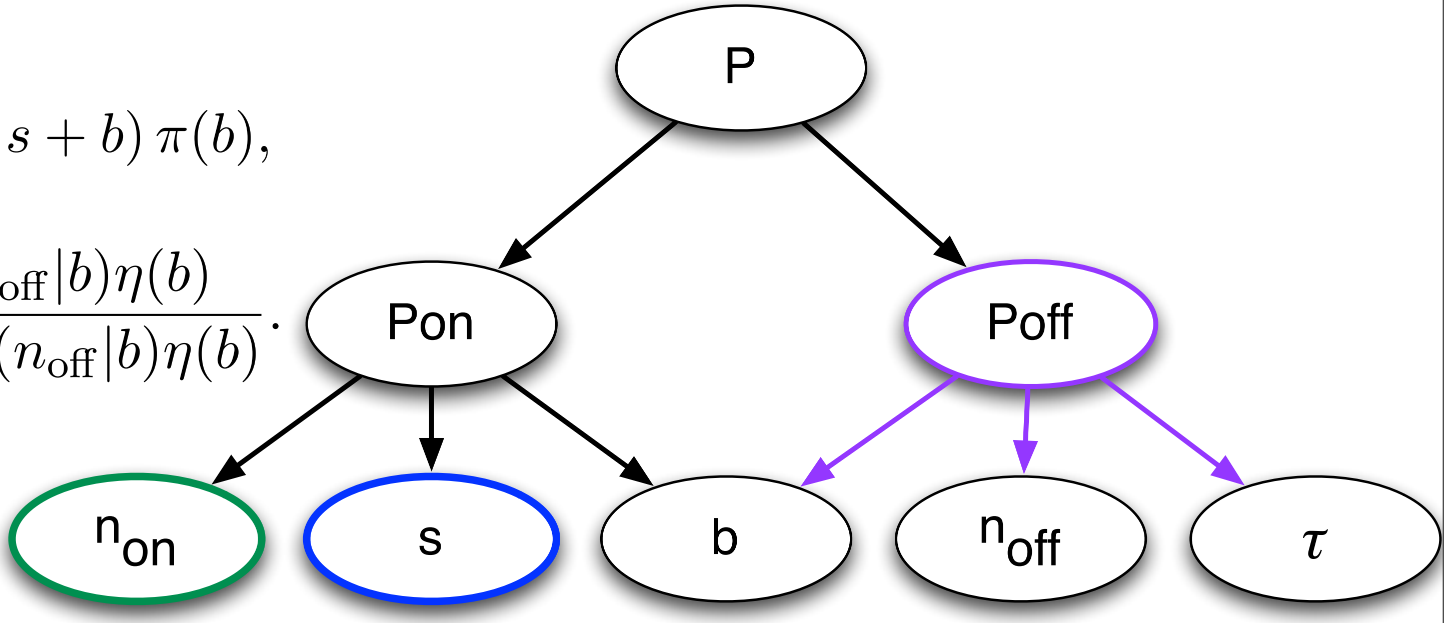
- the joint model $P(x, y|s, b, \tau)$ $P(n_{\rm on}, n_{\rm off}|s, b) = {\rm Pois}(n_{\rm on}|s + b)\,{\rm Pois}(n_{\rm off}|\tau b).$
- a Bayesian prior $\eta(s, b)$

How do we identify the “off” part of the model

- was an average model for n_{on}, so use largest factor independent of n_{on}, or
- think find largest parameter independent of parameter of interest

$$P(n_{\rm on}|s) = \int db\,{\rm Pois}(n_{\rm on}|s + b)\,\pi(b),$$

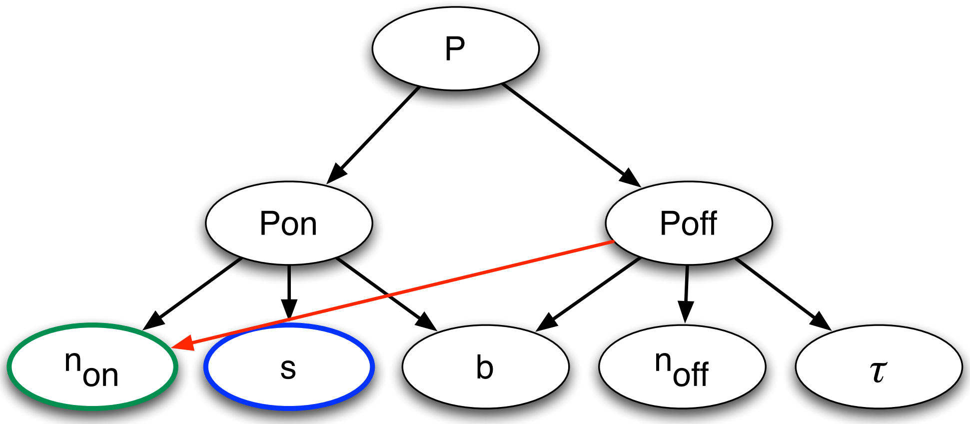
$$\pi(b) = P(b|n_{\rm off}) = \frac{P(n_{\rm off}|b)\eta(b)}{\int db P(n_{\rm off}|b)\eta(b)}.$$

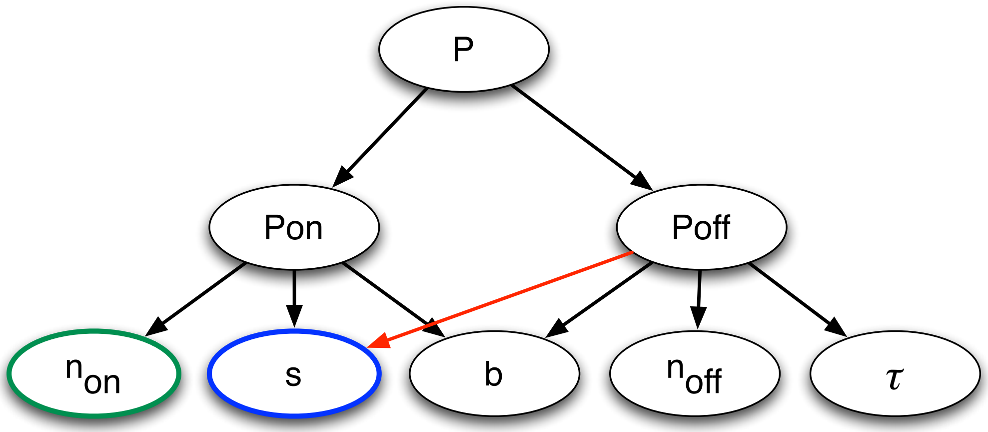


Generalizing the Hybrid approach

The two approaches are not equivalent for joint models like this:

- was an average model for n_{on}, so use largest factor independent of n_{on}, or
- think find largest parameter independent of parameter of interest

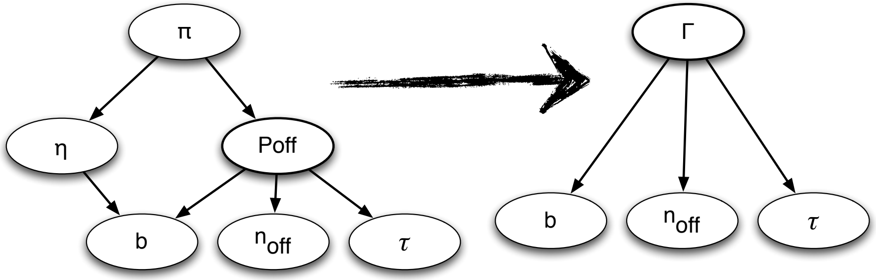




And then there is the question of the prior... what if $\eta(s,b)$ doesn't factorize

- marginal over *s* will have some residual prior dependence on *s*

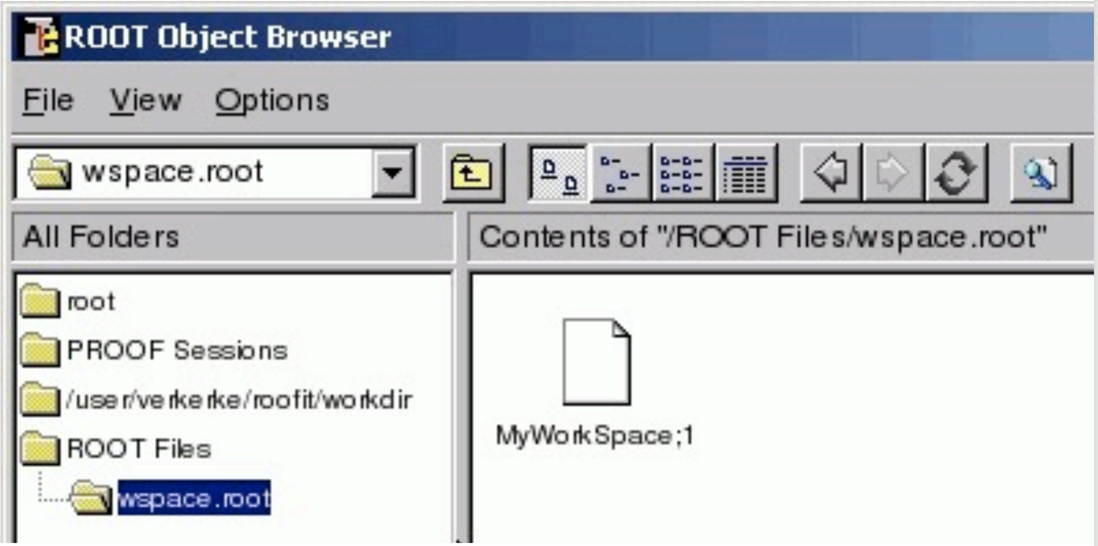
For numerical reasons, we would like to have a table of replacements:

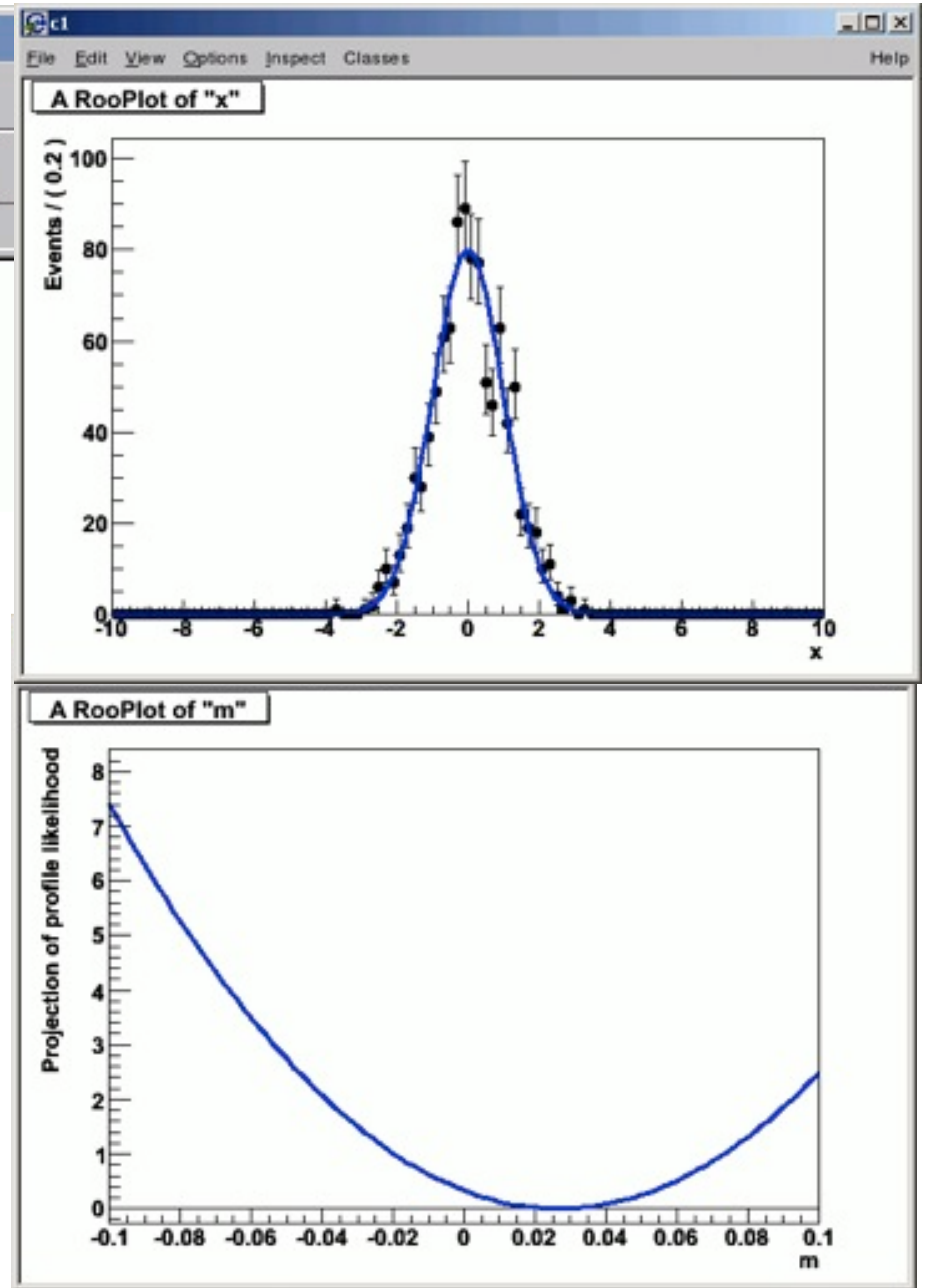


PDF	Prior	Posterior
Gaussian	uniform	Gaussian
Poisson	uniform	Gamma
Log-normal	reference	Log-Normal

The RooFit/RooStats workspace

CENTER FOR COSMOLOGY AND PARTICLE PHYSICS



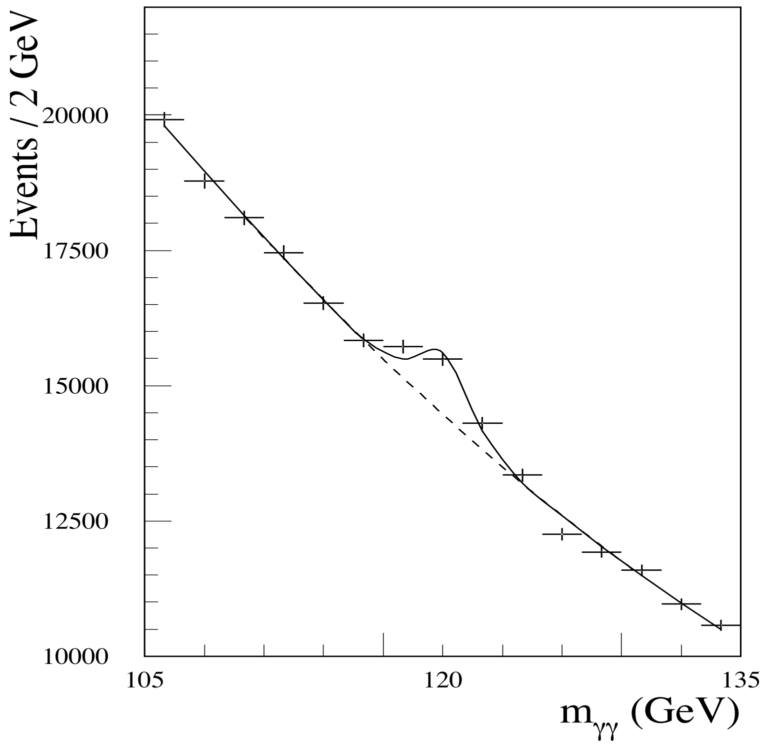


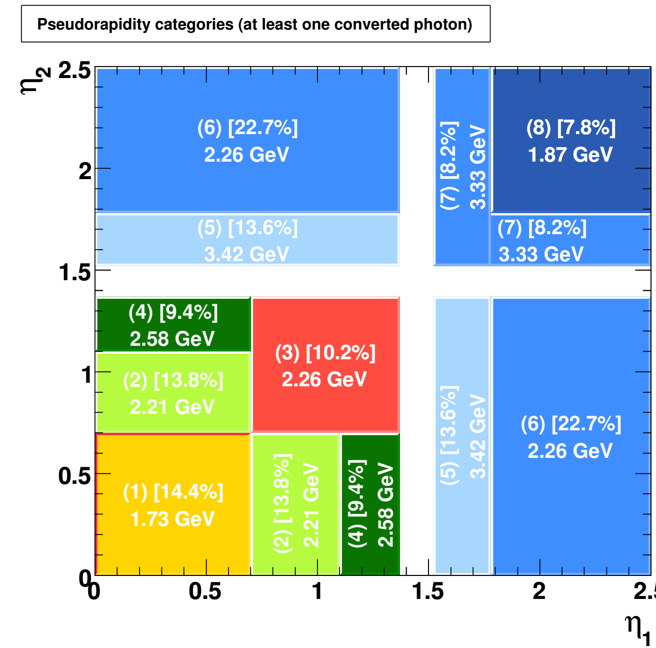
RooFit's Workspace now provides the ability to save in a ROOT file the full probability model, any priors you might need, and the minimal data necessary to reproduce likelihood function.

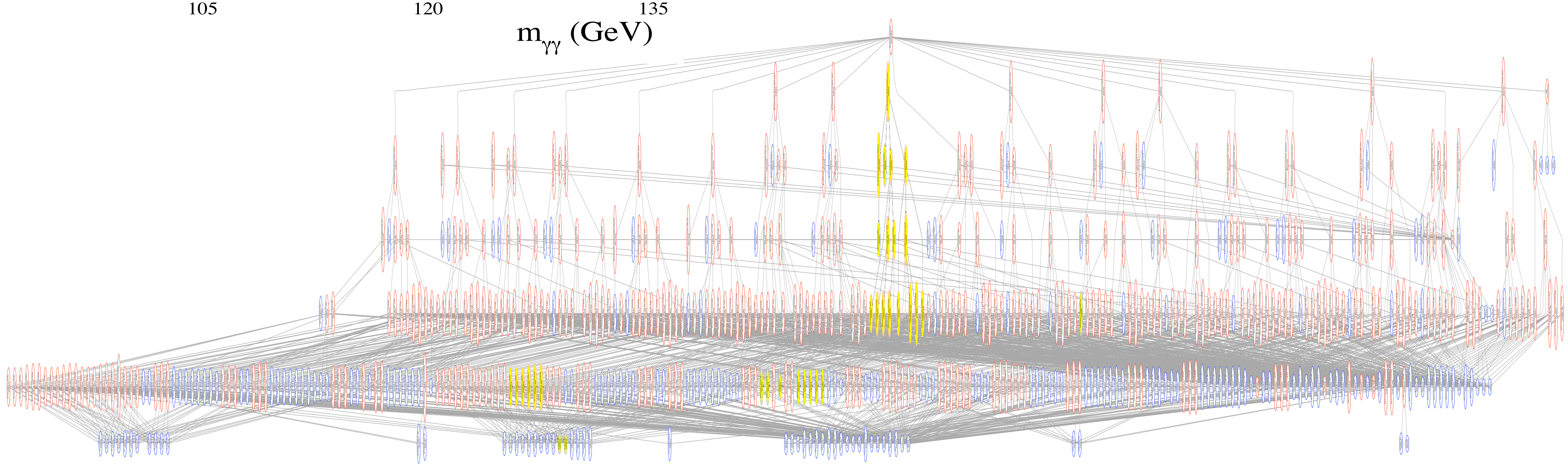
Need this for combinations, exciting potential for publishing results.

ATLAS H->γγ

Center for Cosmology and Particle Physics







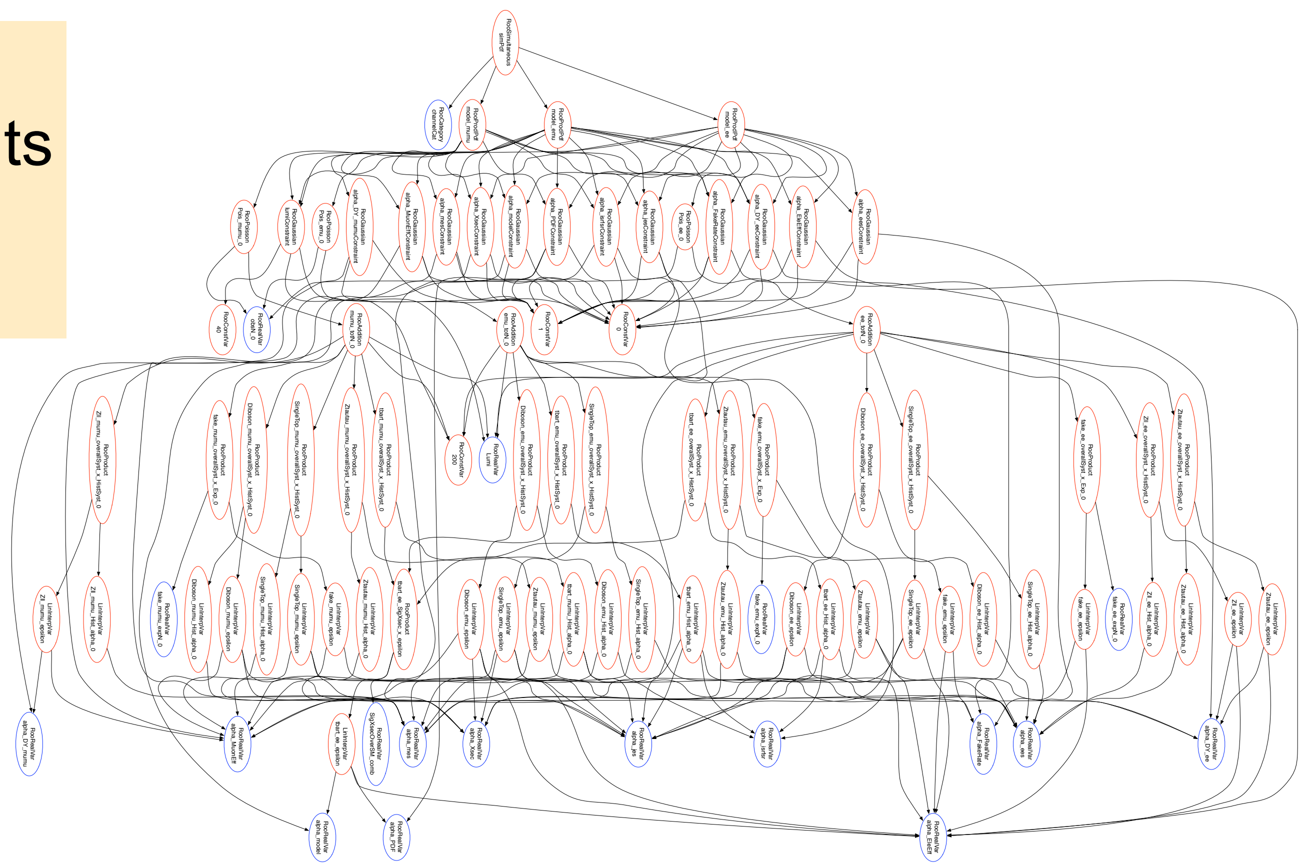
3-channel top combination

The graph below represents this PDF

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

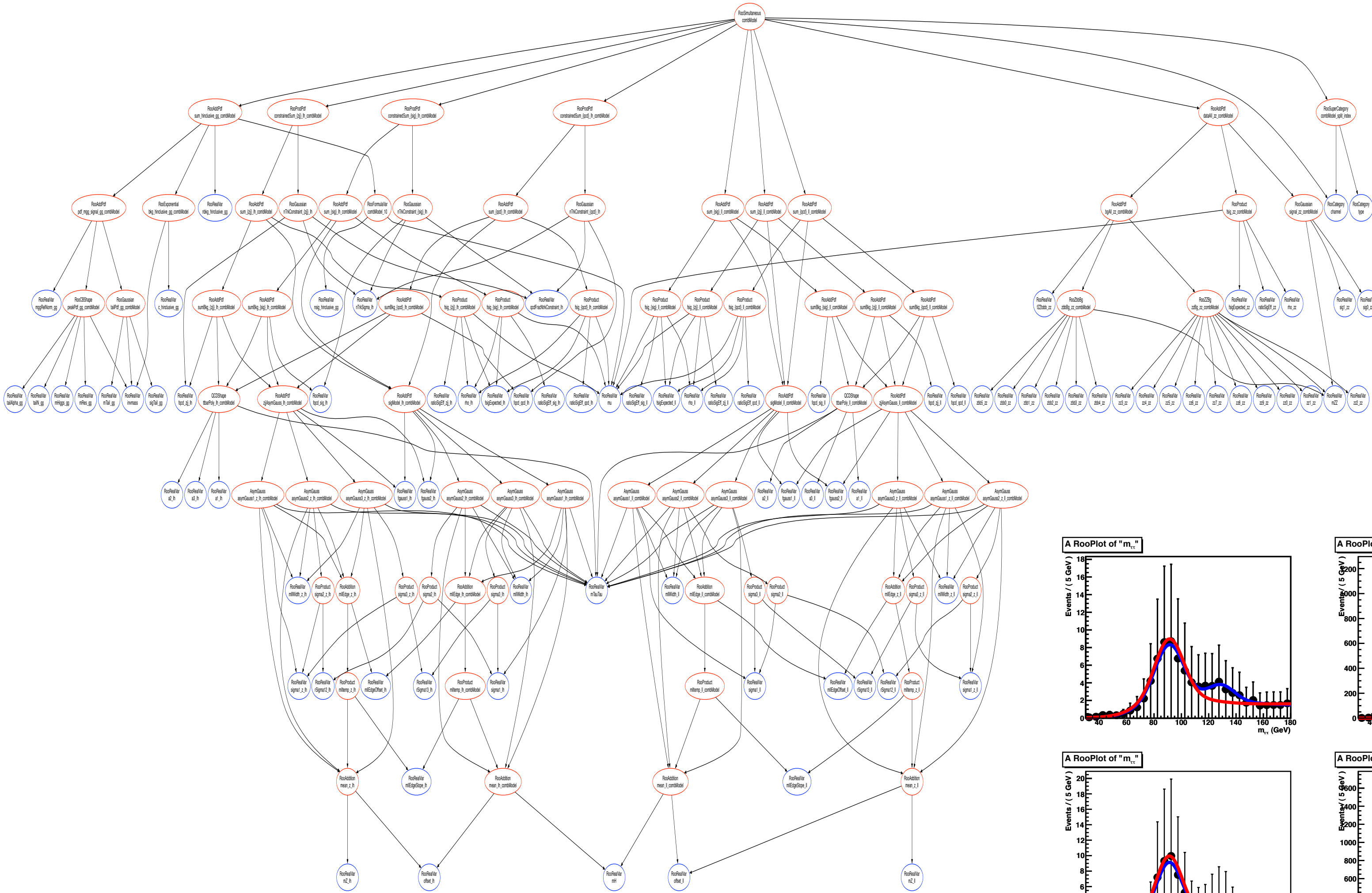
- where there are several relations between the expected means in the different channels

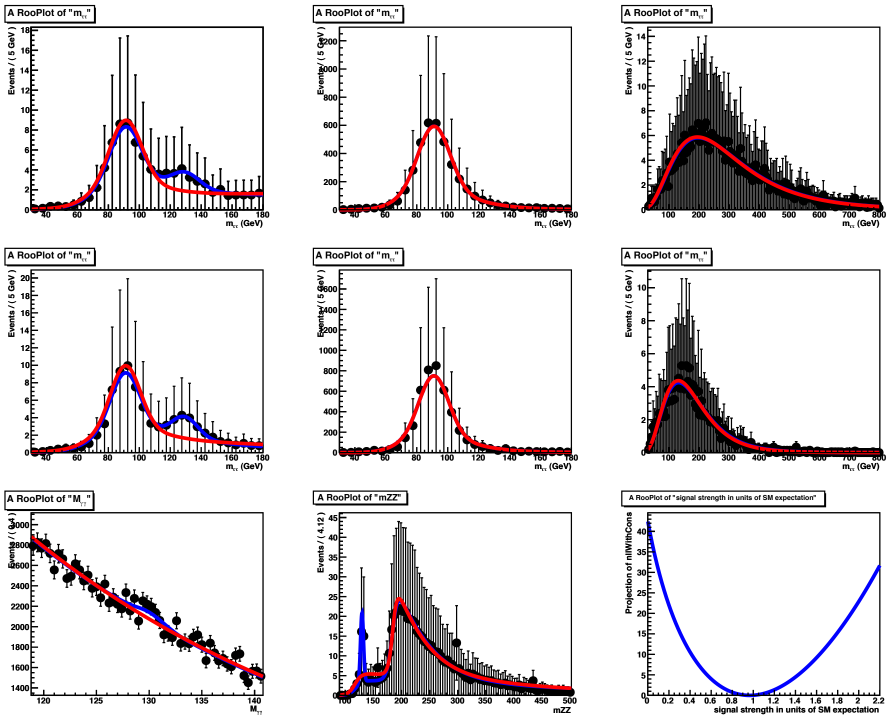
3 observations from data
13 auxiliary measurements
1 parameter of interest
13 nuisance parameters



4-channel ATLAS Higgs combination

Center for
Cosmology and
Particle Physics

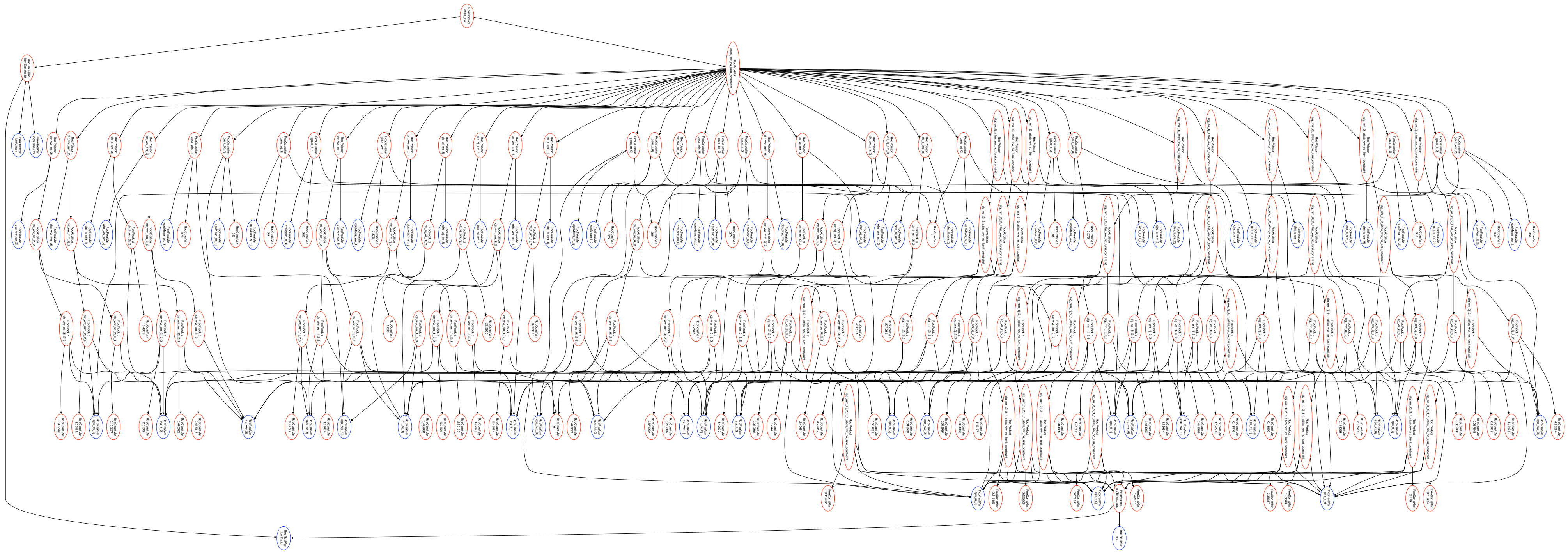




9 observations of continuous marks
1 parameter of interest
27 nuisance parameters

9 Channel ATLAS H->WW combination

top level model



parameter of interest

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

25 measurements from data

1 parameter of interest and 24 nuisance parameters

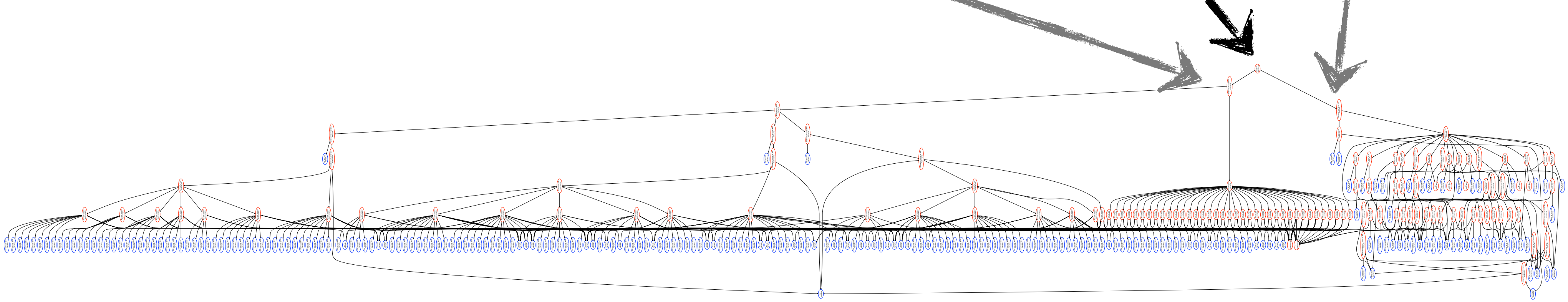
Visualization of the ATLAS+CMS Workspace

The full model has
12 observables and
~50 parameters

top level model

ATLAS part

CMS part



parameter of interest

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

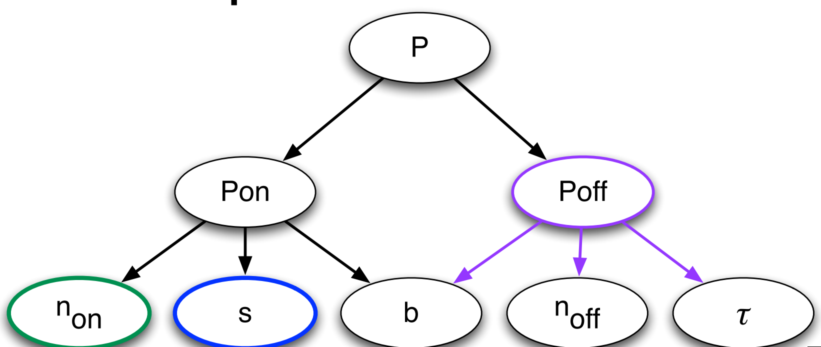
Graphical Models

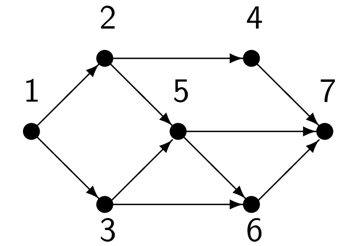
CENTER FOR COSMOLOGY AND PARTICLE PHYSICS

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

Graph of on/off model





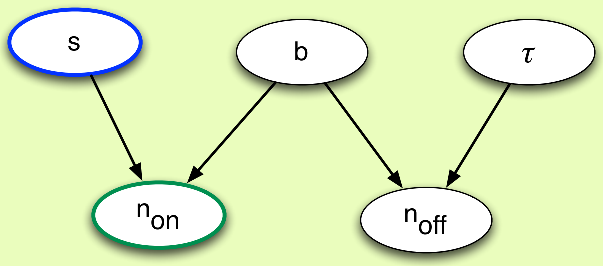
Directed Markov means

$$f(x) = f(x_1)f(x_2 \mid x_1)f(x_3 \mid x_1)f(x_4 \mid x_2) \times f(x_5 \mid x_2, x_3)f(x_6 \mid x_3, x_5)f(x_7 \mid x_4, x_5, x_6).$$

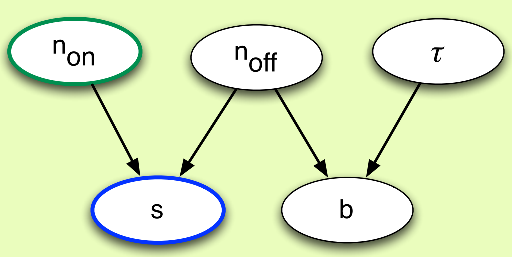
Theory exists for deriving all conditional independencies and exploiting local structure in graph for gross computational simplifications in complex models. Has been successfully exploited in AI, machine learning, and Bayesian statistics.

Steffen Lauritzen University of Oxford | Statistical respondent

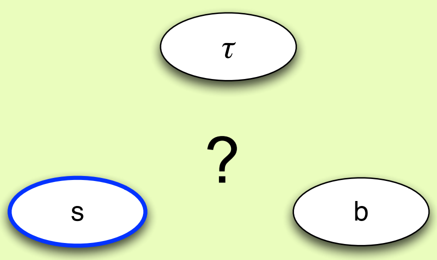
P(data | parameters)



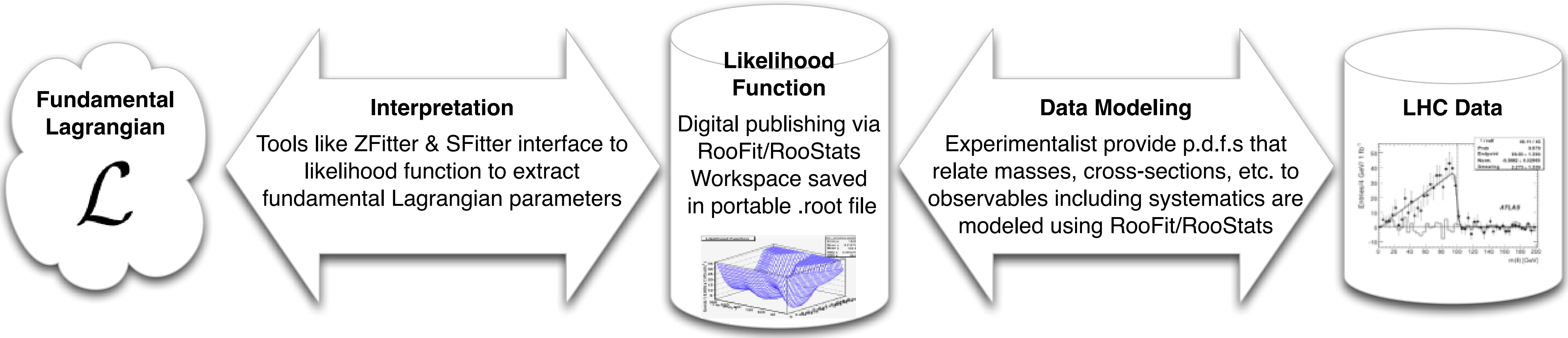
P(parameters|data)



P(parameters)



Raw Measurements to Interpretation



From the raw LHC data, the experiments estiamate (“measure”) several interesting quantities (like masses of particles)

- ideally the likelihood function for those quantities is provided

These quantities are not the parameters of the fundamental theory, but the are usually functions of the fundamental parameters

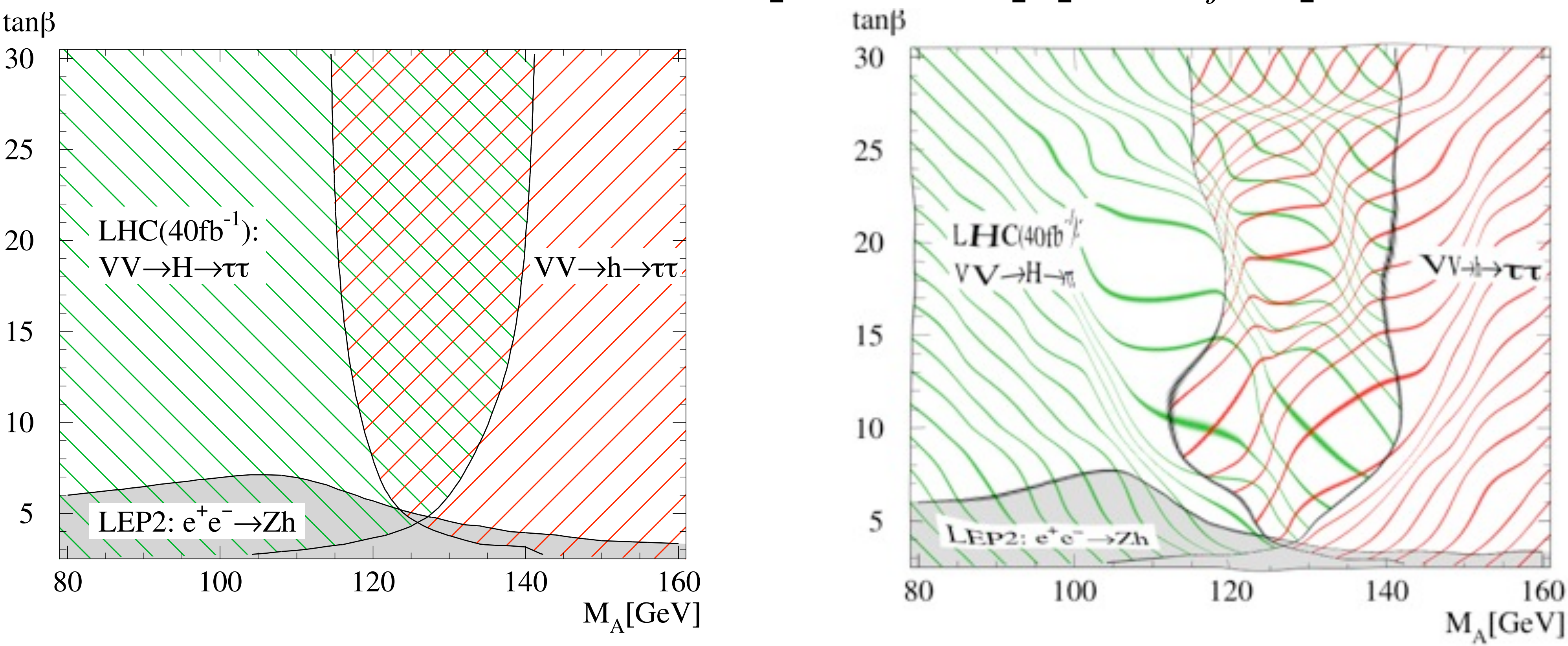
- an entire industry has emerged that interprets these observations in terms of a specific theory (see Roberto Trotta’s talk for an example)

Some applications of 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.
- It would be useful to efficiently sample this space efficiently
- eg... uniform in fisher information metric

$$g_{ij}(\alpha) = \int dx\, f_\alpha(x) \left[\frac{\partial \log f_\alpha(x)}{\partial \alpha_i}\right] \left[\frac{\partial \log f_\alpha(x)}{\partial \alpha_j}\right]$$



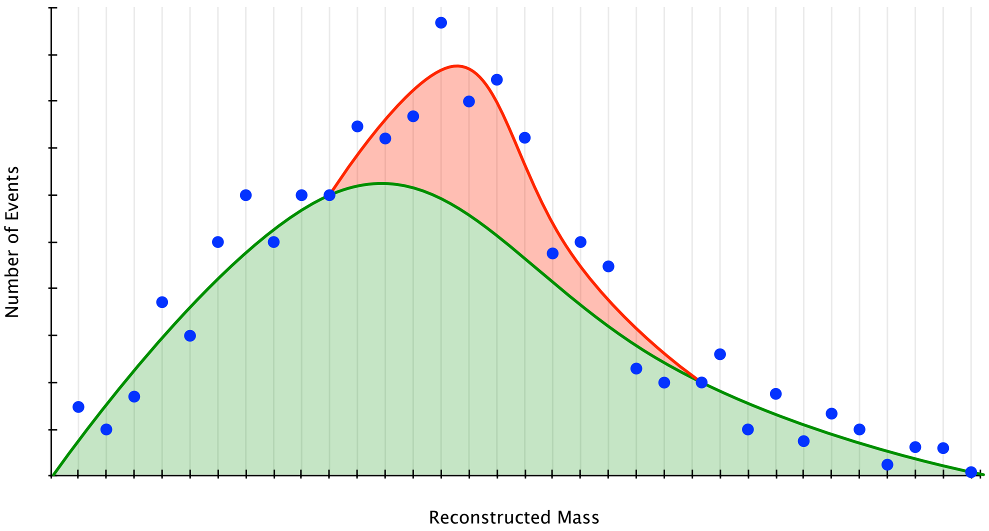
Spinoffs from the Asimov idea

Center for Cosmology and Particle Physics

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

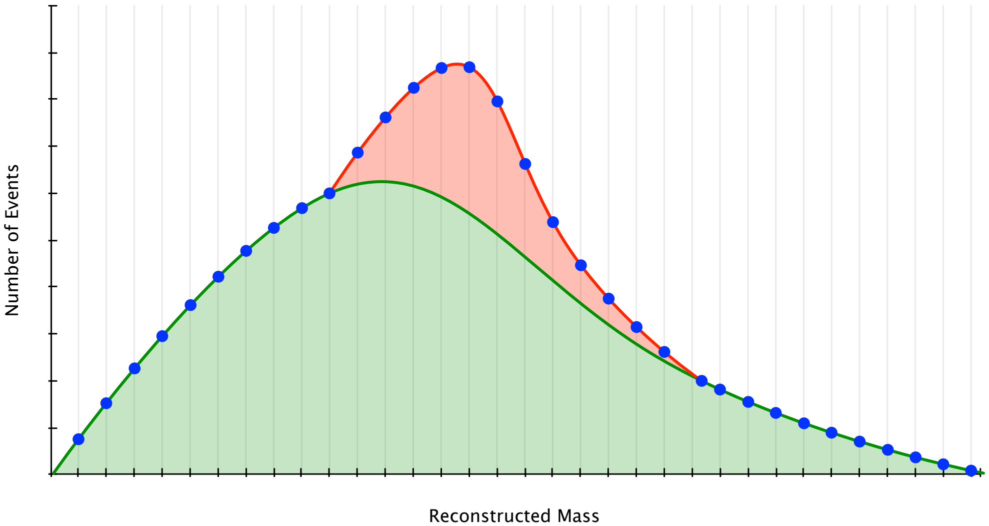
$$g_{ij}(\theta) = \mathrm{E}\left[\left(\frac{\partial}{\partial \theta_i} \ln L(\theta) \right) \left(\frac{\partial}{\partial \theta_j} \ln L(\theta) \right) \middle| \theta \right].$$

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}

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



Last night, Earl L. and Richard L. 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.

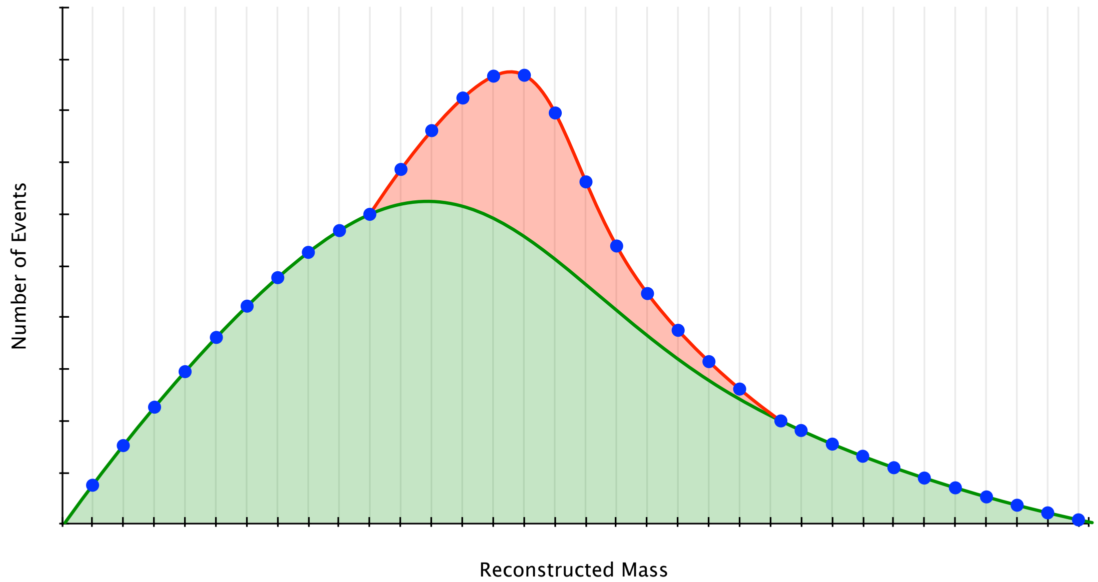
This also provides a convenient algorithm determining for Jeffreys's prior numerically, but I know their are issues with numerics and improper priors.

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

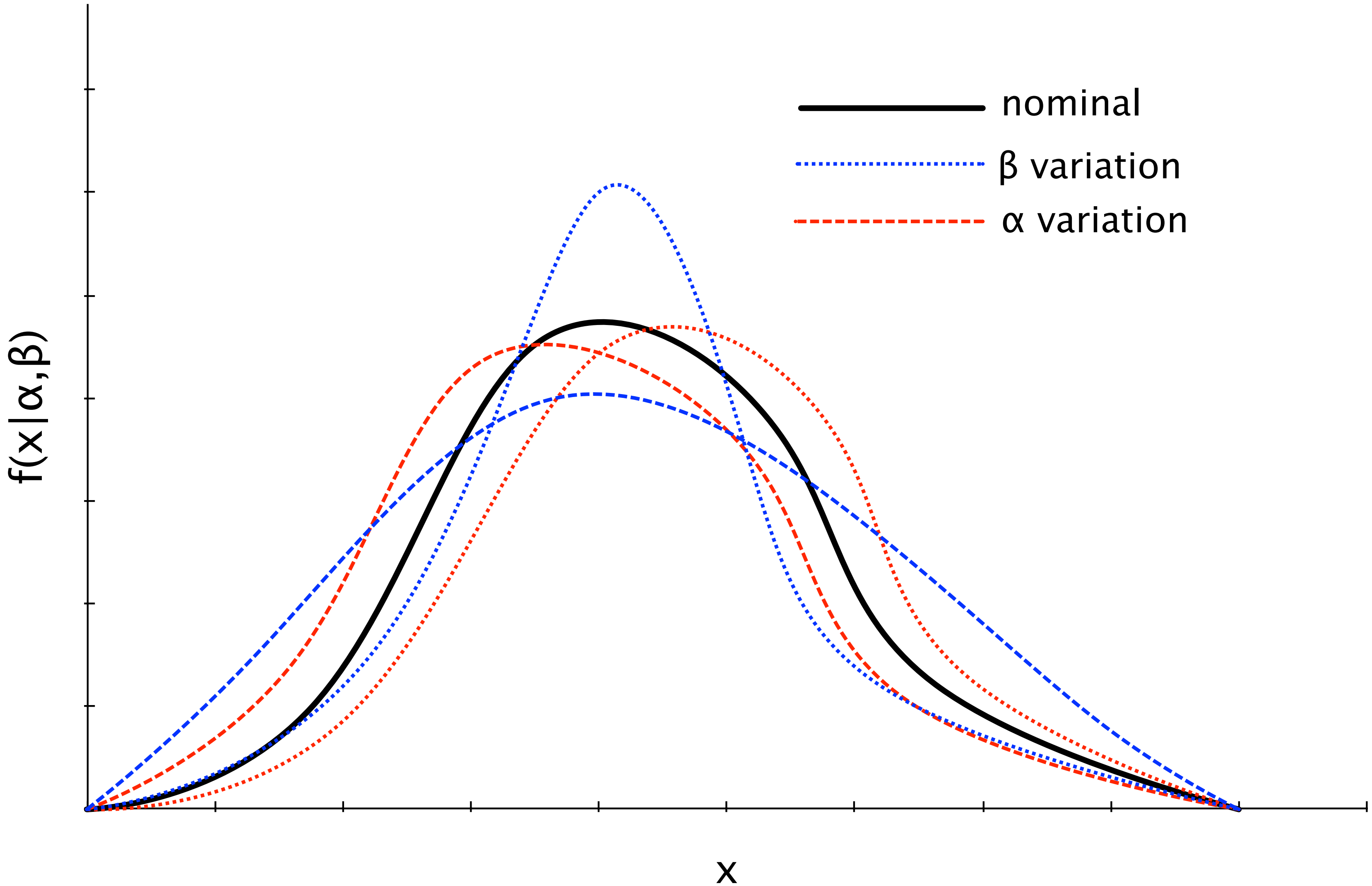
"Multivac picked you as the most representative this year. Not the smartest, or the strongest, or the luckiest, but just the most representative. Now we don't question Multivac, do we?"



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

Implicit vs. Explicit systematics

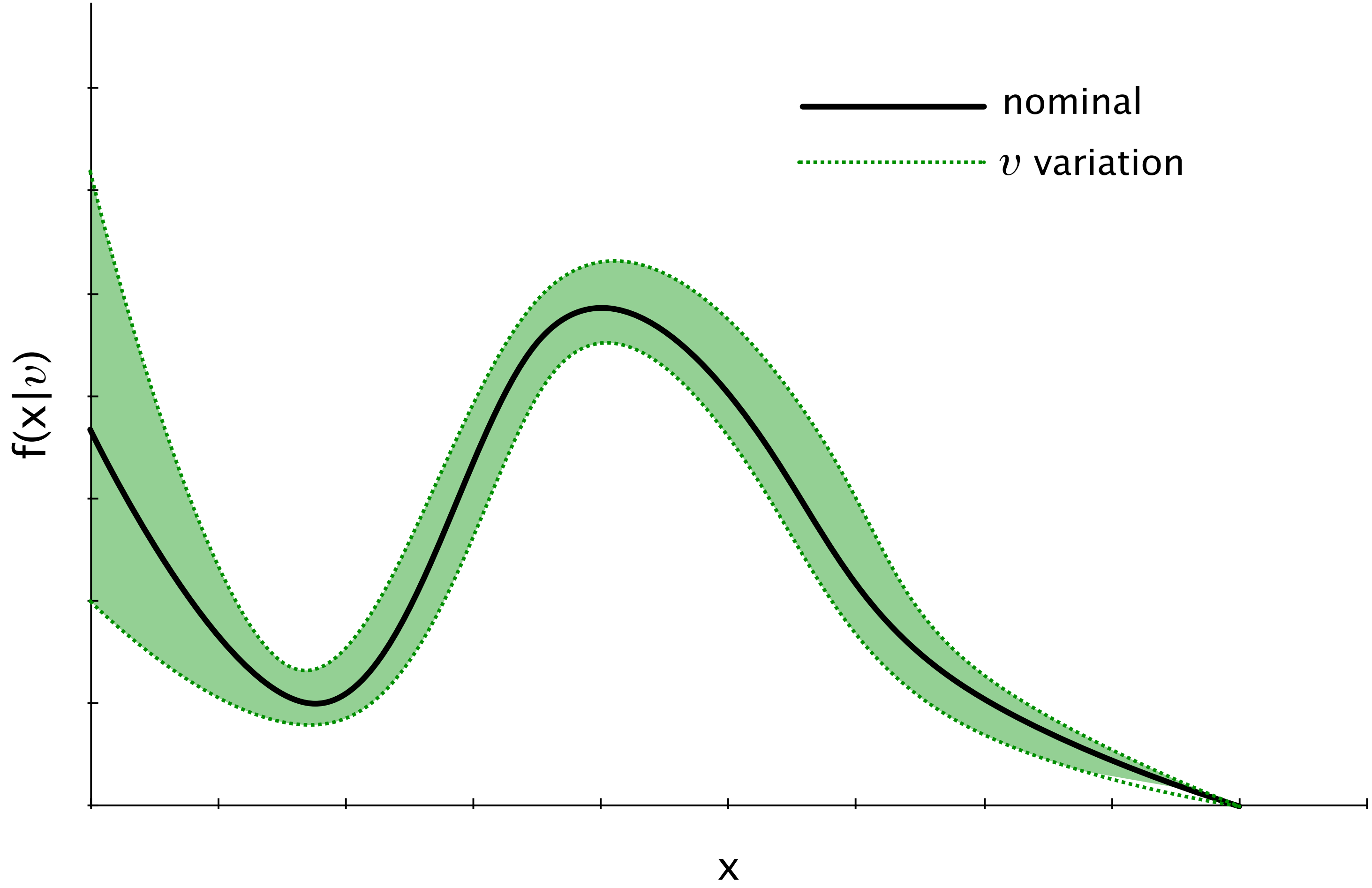
In some cases, effect of systematics is explicitly parametrized with nuisance parameters.



Implicit vs. Explicit systematics

In other cases, one simply has a flexible model parametrized by υ, which is flexible enough to incorporate the systematic effects

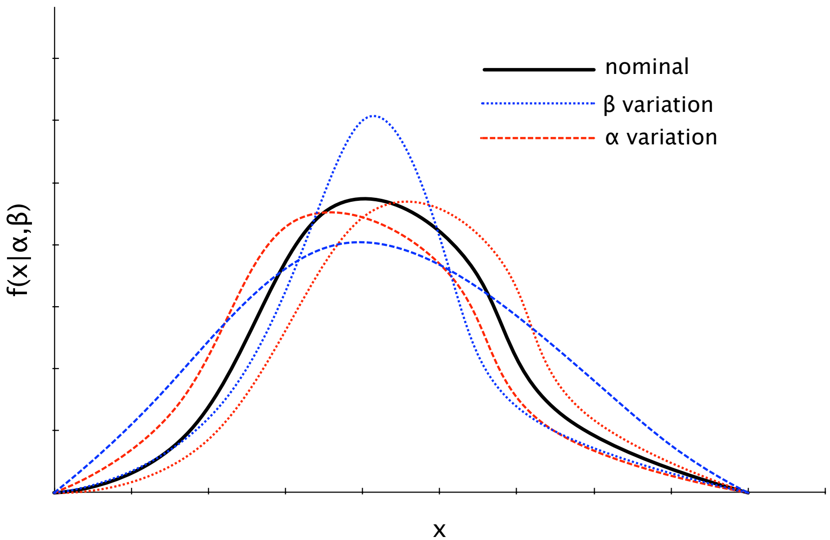
- so dependence on α,β (previously identified with specific systematic effects) is implicit

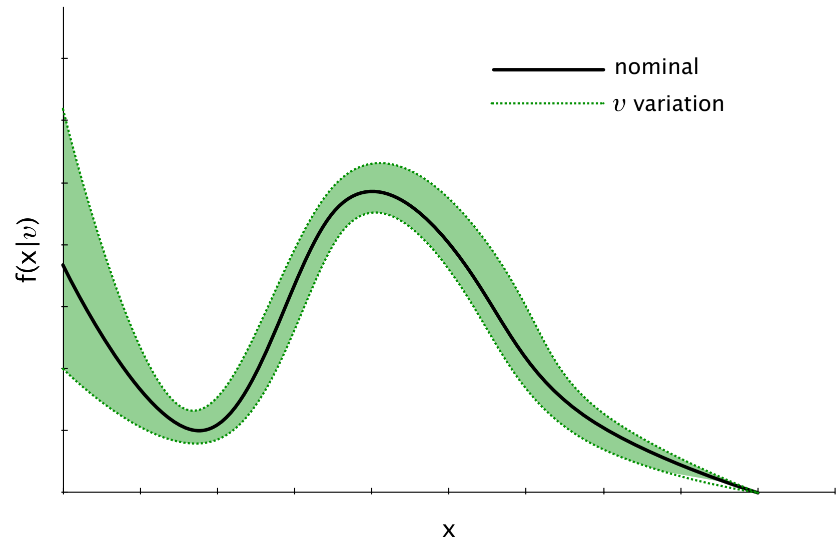


Implicit vs. Explicit systematics

A problem arises when one wants to combine these two measurements knowing that the systematic effects of α,β are correlated between the two measurements

- but there is no explicit handle on α,β in the implicit model





- in some cases this may reduce to reparametrization $\nu(\alpha, \beta)$
- in some cases effect of several systematic effects may produce a degenerate deformation of the shape, so it's not clear the dimensionality of the parametrization is even the same

Basic idea is to add a term $P(\alpha, \beta, \nu)$ that summarizes the correlation between the parameters...

- what is the procedure for determining this term, especially if we want to maintain a frequentist interpretation

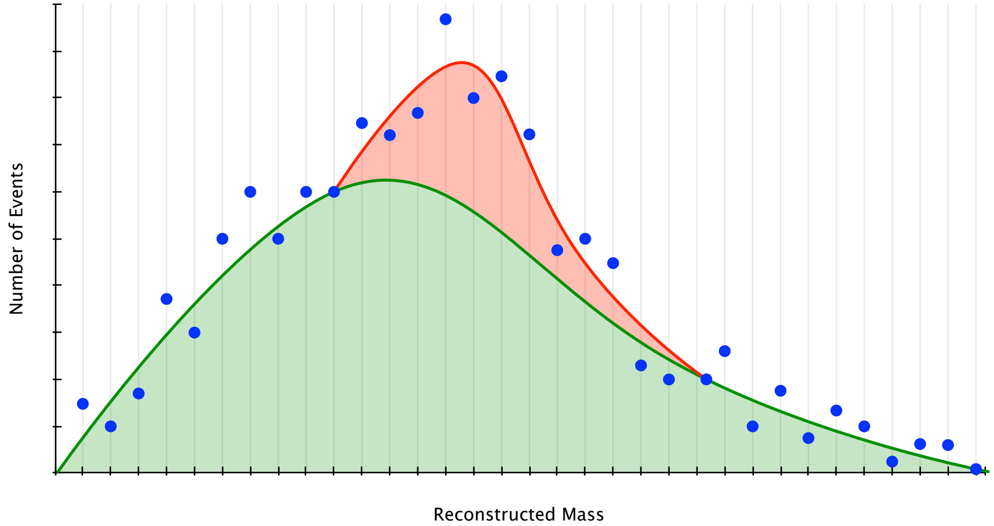
Conditioning & Look elsewhere

Look-elsewhere effect:

- location of peak is meaningless in null model
- several possibilities for background to fluctuate
  - typical approach is to understand and/or correct for "trials" factor (Bonferroni, talks by Eilam and Ofer$_2$, ...)

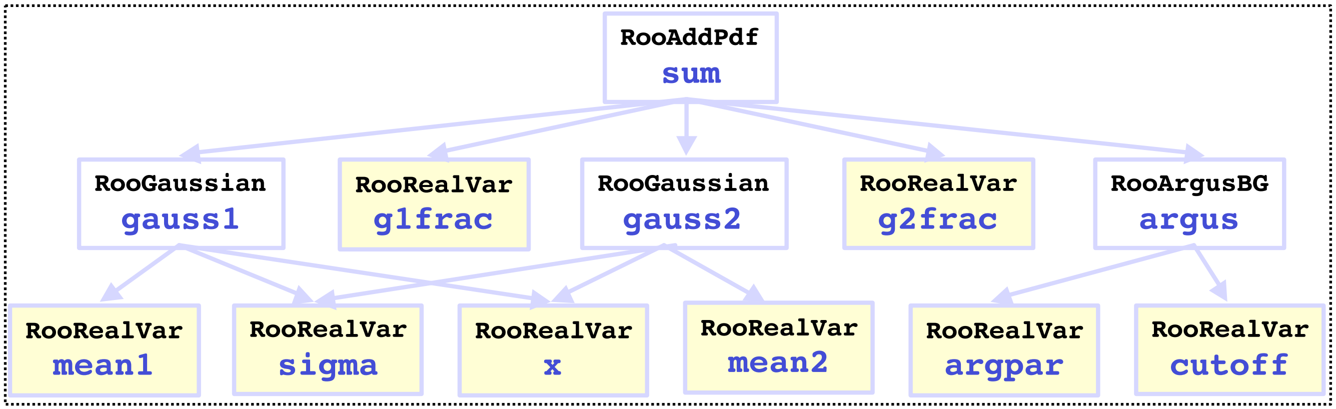
Is there an alternative approach based on conditioning:

- eg., what is p-value for a peak this large in the background for the ensemble where the biggest peak is located at this point.

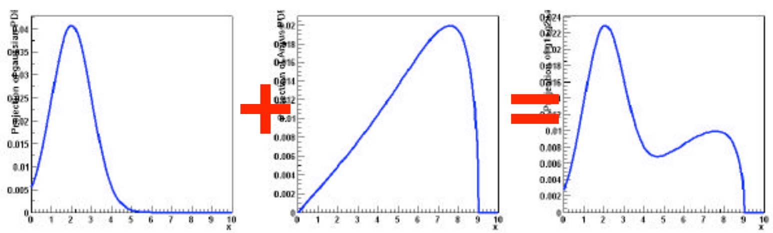


Extras

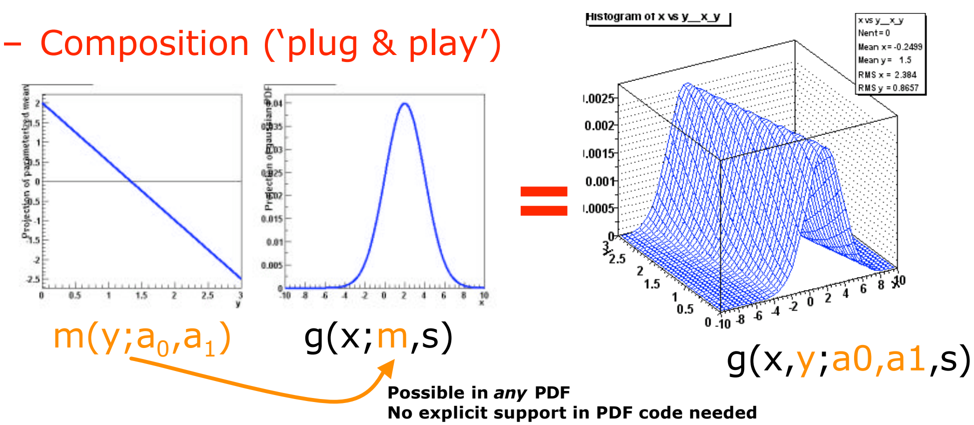
A major tool at BaBar. Fit complicated models with >100 parameters!



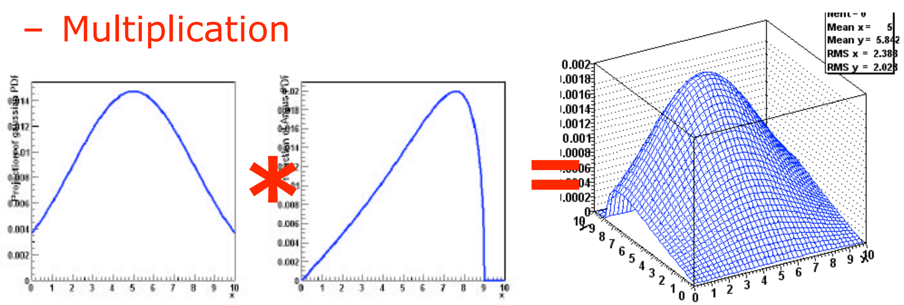
– Addition



– Composition ('plug & play')

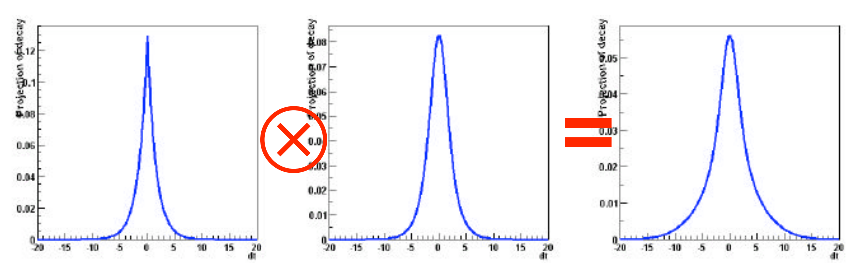


– Multiplication



Wouter Verkerke,

– Convolution

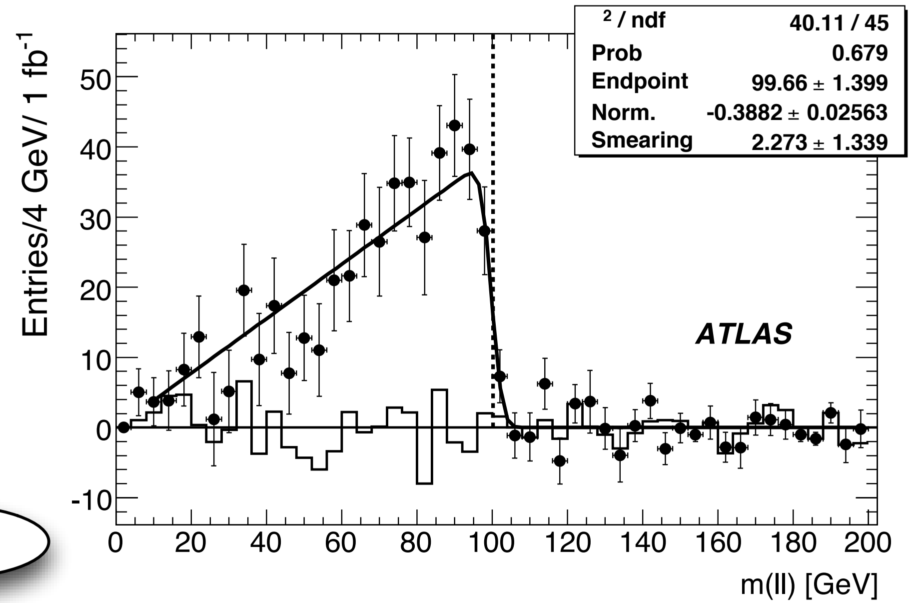


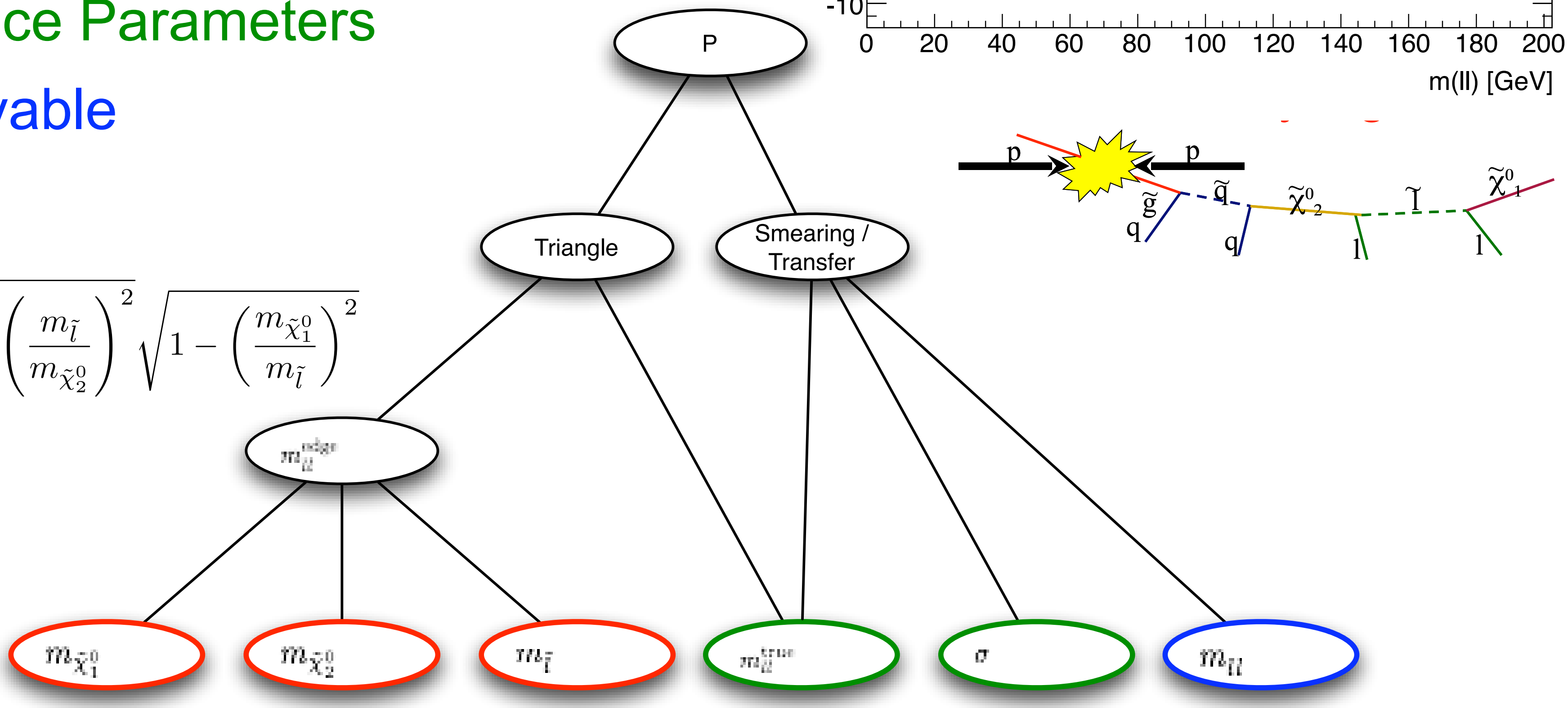
Wouter Verkerke, UCSB

Supersymmetric Mass Measurements

Here is a graphical representation of a measurement used for supersymmetric parameter estimation

- 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

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

- relate edge according to:

$$m_{ll}^{\rm edge} = m_{\tilde\chi_2^0}\sqrt{1-\left(\frac{m_{\tilde l}}{m_{\tilde\chi_2^0}}\right)^2}\sqrt{1-\left(\frac{m_{\tilde\chi_1^0}}{m_{\tilde l}}\right)^2}$$

- incorporate matrix element techniques

Matrix-element likelihood:
Calculate probability directly

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

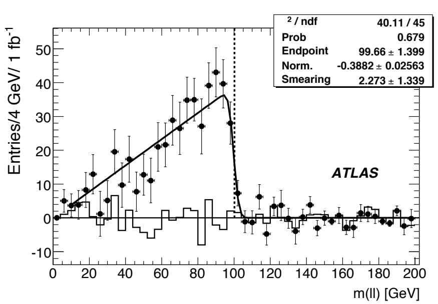
where

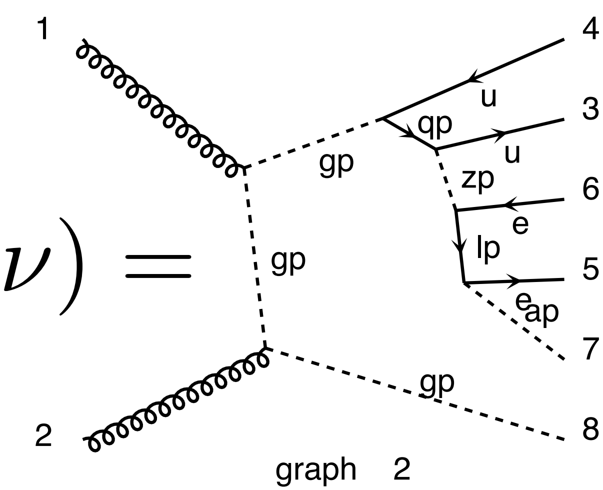
$$P(z \mid A) = \int dy \; |\mathcal{M}_A|^2 f_p f_{\bar p} \; f_{TF}(y,z) = d\sigma_A/dz$$

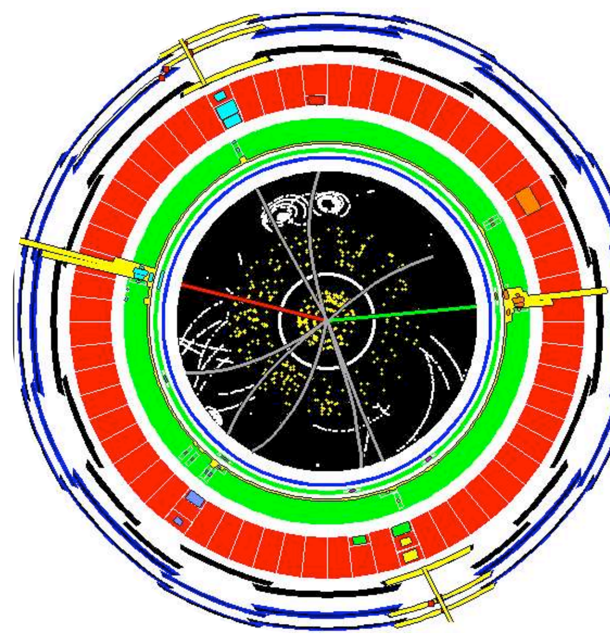
Parton(y) to detector(z) transfer function (TF) describes parton-shower and detector response in parametrized form *(Issue 2)*

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

Integration over parton-level quantities

$P($ 

$|\mu,\nu) =$ 

$\oplus$ 

Matrix Element Method as a Parametrization

Center for Cosmology and Particle Physics

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) + ...

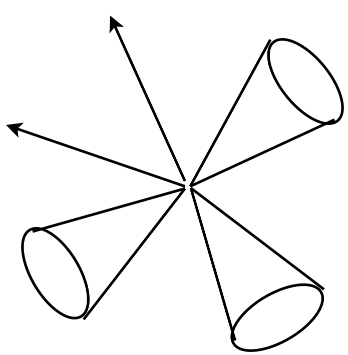
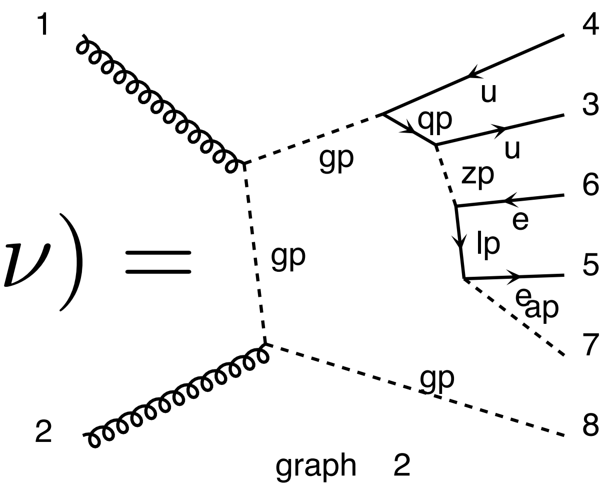
where

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

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

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

$\oplus$