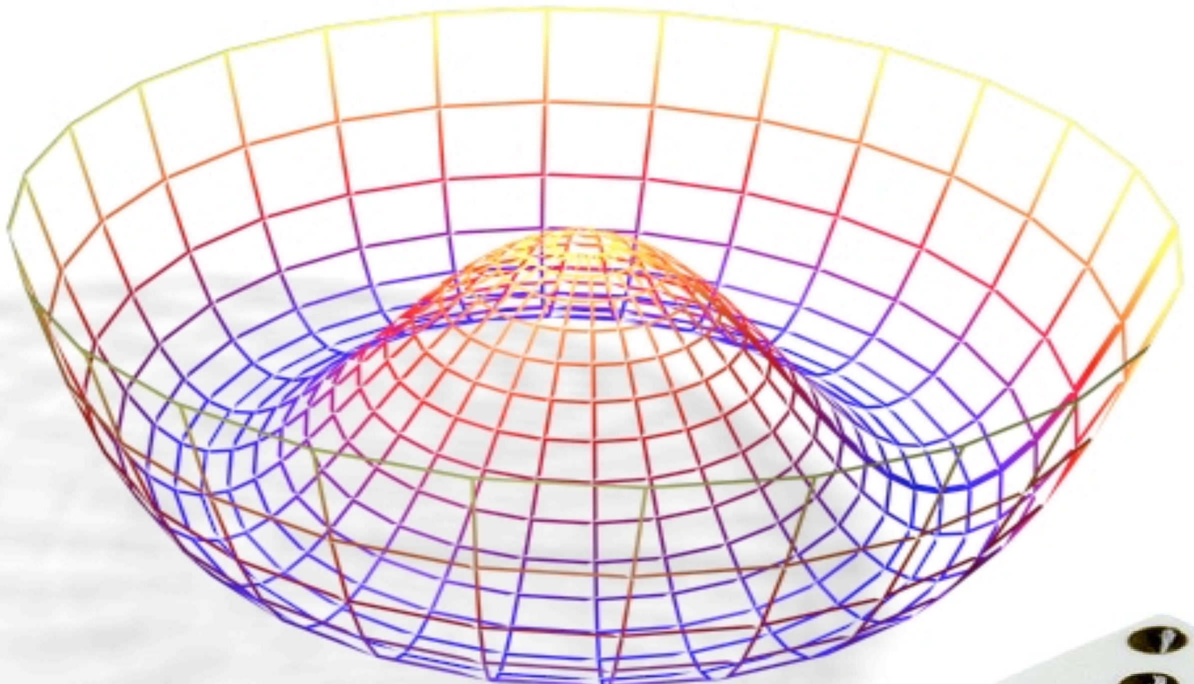
CENTER FOR
COSMOLOGY AND
PARTICLE PHYSICS

Practical Statistics for Particle Physics



Kyle Cranmer,
New York University

Introduction

Statistics plays a vital role in science, it is the way that we:

- quantify our knowledge and uncertainty
- communicate results of experiments

Big questions:

- how do we make discoveries, measure or exclude theoretical parameters, ...
- how do we get the most out of our data
- how do we incorporate uncertainties
- how do we make decisions

Statistics is a very big field, and it is not possible to cover everything in 4 hours. In these talks I will try to:

- **explain** some fundamental ideas & prove a few things
- **enrich** what you already know
- **expose** you to some new ideas

I will try to go slowly, because if you are not following the logic, then it is not very interesting.

- Please feel free to ask questions and interrupt at any time

Further Reading

By physicists, for physicists

G. Cowan, Statistical Data Analysis, Clarendon Press, Oxford, 1998.

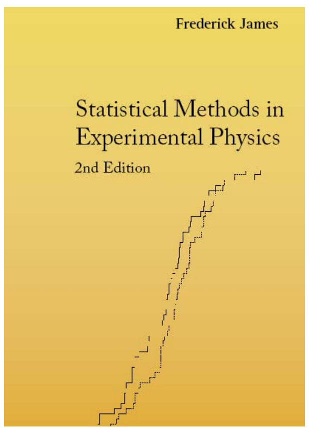
R.J.Barlow, A Guide to the Use of Statistical Methods in the Physical Sciences, John Wiley, 1989;

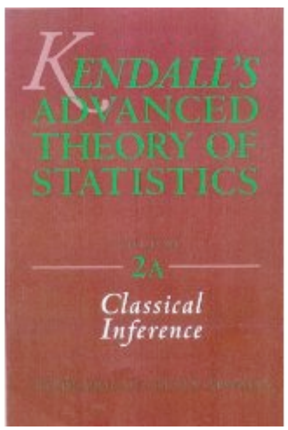
F. James, Statistical Methods in Experimental Physics, 2nd ed., World Scientific, 2006;

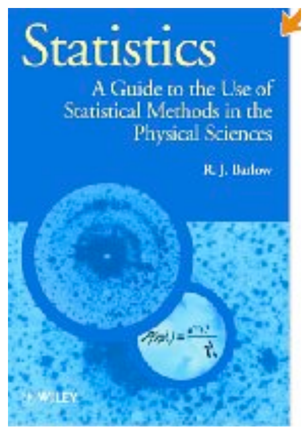
- W.T. Eadie et al., North-Holland, 1971 (1st ed., hard to find);

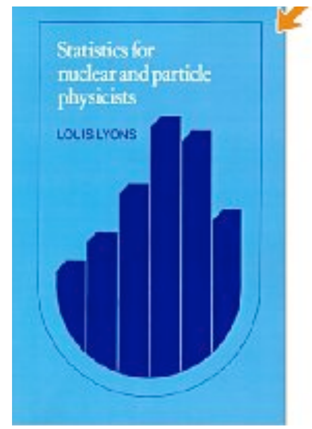
S.Brandt, Statistical and Computational Methods in Data Analysis, Springer, New York, 1998.

L.Lyons, Statistics for Nuclear and Particle Physics, CUP, 1986.









My favorite statistics book by a statistician:

Stuart, Ord, Arnold. "Kendall's Advanced Theory of Statistics" Vol. 2A *Classical Inference & the Linear Model.*

Other lectures

Center for Cosmology and Particle Physics

Fred James's lectures

http://preprints.cern.ch/cgi-bin/setlink?base=AT&categ=Academic_Training&id=AT00000799

http://www.desy.de/~acatrain/

Glen Cowan's lectures

http://www.pp.rhul.ac.uk/~cowan/stat_cern.html

Louis Lyons

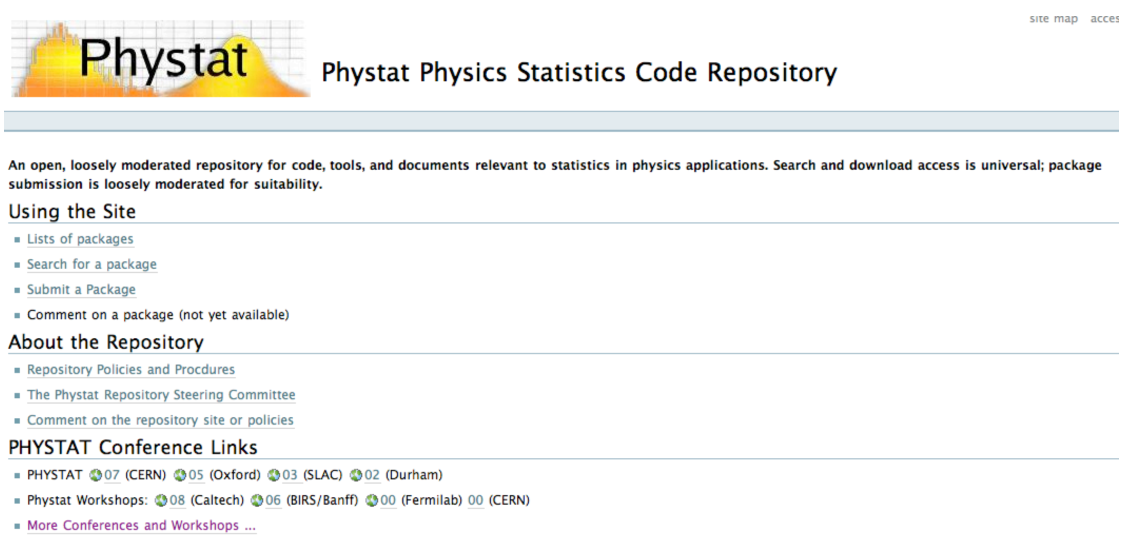
http://indico.cern.ch/conferenceDisplay.py?confId=a063350

Bob Cousins gave a CMS lecture, may give it more publicly

Gary Feldman "Journeys of an Accidental Statistician"

http://www.hepl.harvard.edu/~feldman/Journeys.pdf

The PhyStat conference series at PhyStat.org:


site map acces
Phystat
Phystat Physics Statistics Code Repository

An open, loosely moderated repository for code, tools, and documents relevant to statistics in physics applications. Search and download access is universal; package submission is loosely moderated for suitability.

Using the Site

- Lists of packages
- Search for a package
- Submit a Package
- Comment on a package (not yet available)

About the Repository

- Repository Policies and Procdures
- The Phystat Repository Steering Committee
- Comment on the repository site or policies

PHYSTAT Conference Links

- PHYSTAT 07 (CERN) 05 (Oxford) 03 (SLAC) 02 (Durham)
- Phystat Workshops: 08 (Caltech) 06 (BIRS/Banff) 00 (Fermilab) 00 (CERN)
- More Conferences and Workshops ...

Practical Statistics for the LHC

Kyle Cranmer
Center for Cosmology and Particle Physics, Physics Department, New York University, USA

Abstract
This document is a pedagogical introduction to statistics for particle physics. Emphasis is placed on the terminology, concepts, and methods being used at the Large Hadron Collider. The document addresses both the statistical tests applied to a model of the data and the modeling itself . I expect to release updated versions of this document in the future.

Contents

Lecture 1

Outline

Lecture 1: Building a probability model

- preliminaries, the marked Poisson process
- incorporating systematics via nuisance parameters
- constraint terms
- examples of different "narratives" / search strategies

Lecture 2: Hypothesis testing

- simple models, Neyman-Pearson lemma, and likelihood ratio
- composite models and the profile likelihood ratio
- review of ingredients for a hypothesis test

Lecture 3: Limits & Confidence Intervals

- the meaning of confidence intervals as inverted hypothesis tests
- asymptotic properties of likelihood ratios
- Bayesian approach

Preliminaries

Probability Density Functions

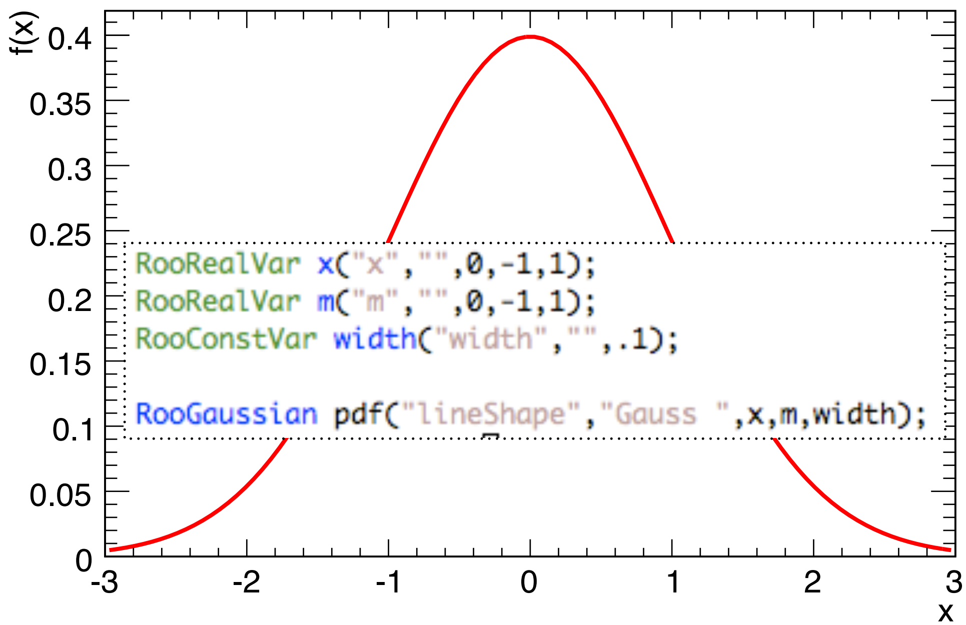
When dealing with continuous random variables, need to introduce the notion of a **Probability Density Function** (PDF... not parton distribution function)

$$P(x \in [x, x+dx]) = f(x)dx$$

Note, $f(x)$ is NOT a probability

PDFs are always normalized

$$\int_{-\infty}^{\infty} f(x)dx = 1$$



The Likelihood Function

A Poisson distribution describes a discrete event count n for a real-valued mean μ.

$$Pois(n|\mu) = \mu^n \frac{e^{-\mu}}{n!}$$

The likelihood of μ given n is the same equation evaluated as a function of μ

- Now it's a continuous function
- But it is not a pdf!

$$L(\mu) = Pois(n|\mu)$$

Common to plot the -2 ln *L*

- helps avoid thinking of it as a PDF
- connection to χ^2 distribution

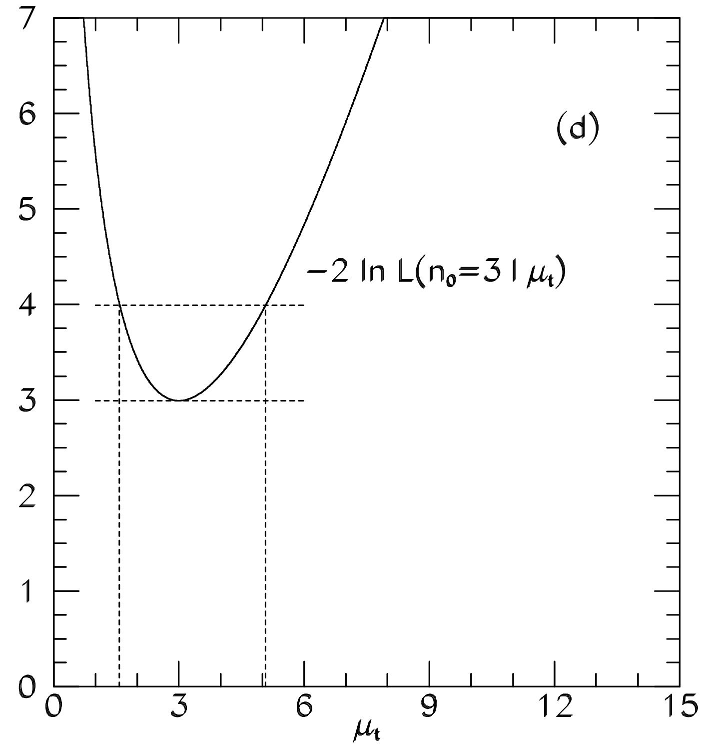


Figure from R. Cousins, Am. J. Phys. 63 398 (1995)

Change of variable x, change of parameter θ

- For pdf $p(x|\theta)$ and change of variable from x to y(x):

  $p(y(x)|\theta) = p(x|\theta) \,/\, |dy/dx|.$

  Jacobian modifies probability *density*, guaranties that

  $P(\, y(x_1) < y < y(x_2)\,) \;=\; P(x_1 < x < x_2\,)$, i.e., that

  Probabilities are invariant under change of variable x.

  – Mode of probability *density* is *not* invariant (so, e.g., criterion of maximum probability density is ill-defined).

  – Likelihood *ratio* is invariant under change of variable x. (Jacobian in denominator cancels that in numerator).

- For likelihood $\mathcal{L}(\theta)$ and reparametrization from θ to u(θ):

  $\mathcal{L}(\theta) \;=\; \mathcal{L}(u(\theta))$ (!).

  – Likelihood $\mathcal{L}(\theta)$ is invariant under reparametrization of parameter θ (reinforcing fact that $\mathcal{L}$ is *not* a pdf in θ).

Bob Cousins, CMS, 2008

Change of variables

CENTER FOR
COSMOLOGY AND
PARTICLE PHYSICS

If *f(x)* is the pdf for *x* and y(x) is a change of variables, then the pdf *g(y)* must satisfy

$$\int_{x_a}^{x_b} f(x)dx = \int_{y(x_a)}^{y(x_b)} g(y)dy$$

We can rewrite the integral on the right

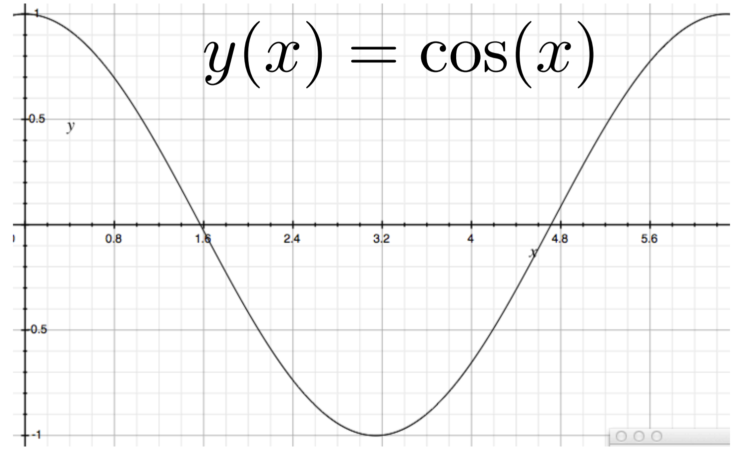
$$\int_{y(x_a)}^{y(x_b)} g(y)dy = \int_{x_a}^{x_b} g(y(x)) \left|\frac{dy}{dx}\right| dx$$

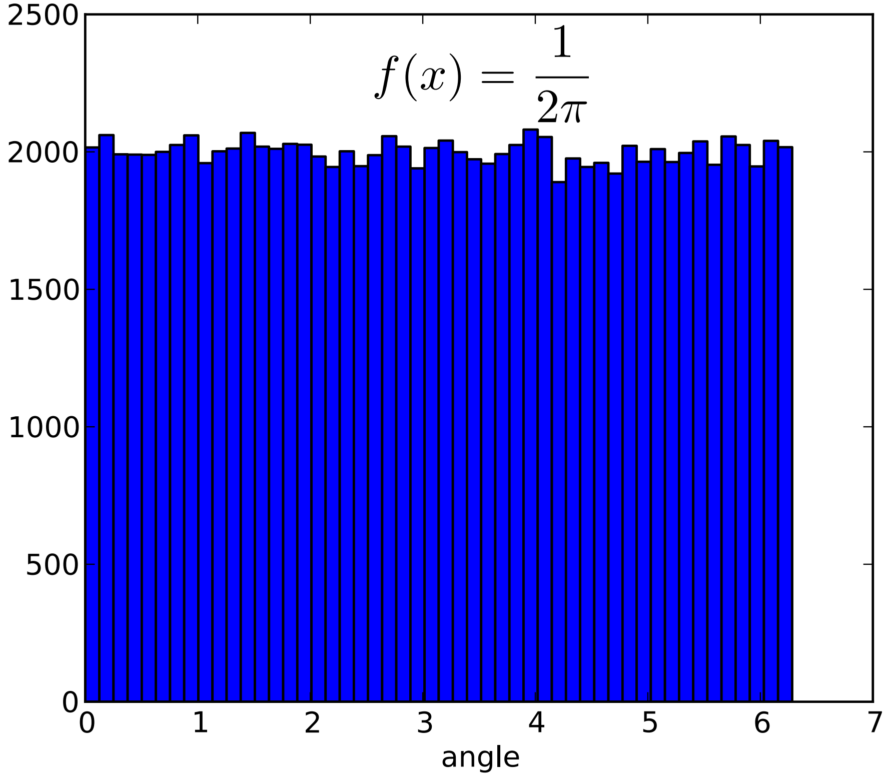
therefore, the two pdfs are related by a Jacobian factor

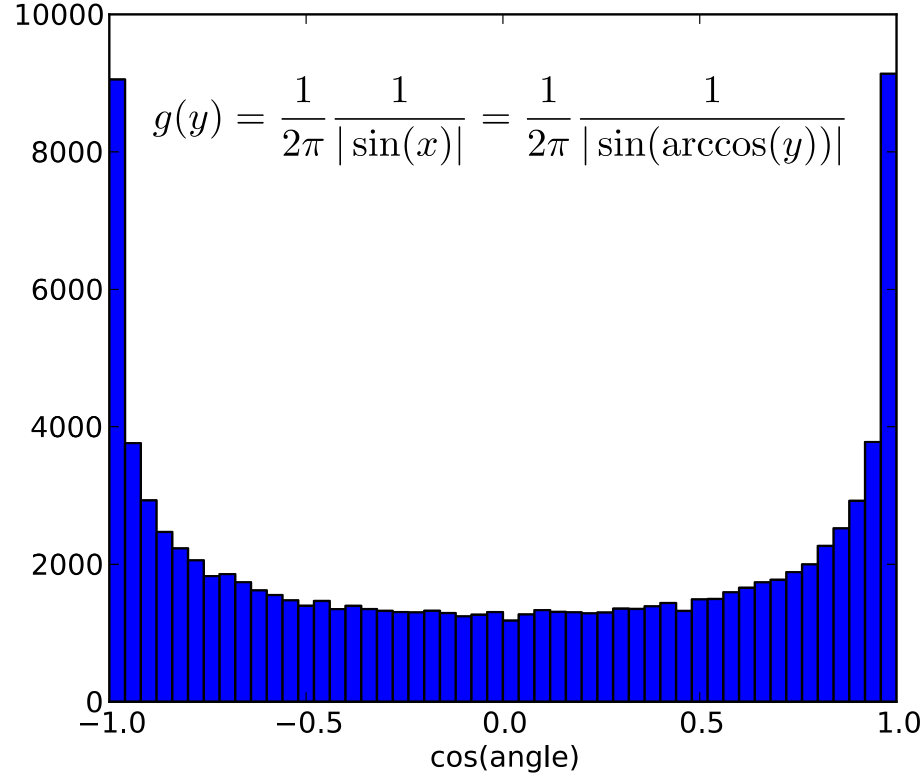
$$f(x) = g(y) \left|\frac{dy}{dx}\right|$$

An example

$$f(x) = g(y)\left|\frac{dy}{dx}\right|$$







Probability Integral Transform

"…seems likely to be one of the most fruitful conceptions introduced into statistical theory in the last few years"
– Egon Pearson (1938)

Given continuous x $\in$ (a,b), and its pdf p(x), let

$$y(x) = \int_a^x p(x')\,dx' .$$

Then y $\in$ (0,1) and p(y) = 1 (uniform) for all y. (!)

So there always exists a metric in which the pdf is uniform.

***Many* issues become more clear (or trivial) after this transformation*. (If x is discrete, some complications.)**

The specification of a Bayesian prior pdf p(μ) for parameter μ is equivalent to the choice of the metric f(μ) in which the pdf is uniform. This is a *deep* issue, not always recognized as such by users of flat prior pdf's in HEP!

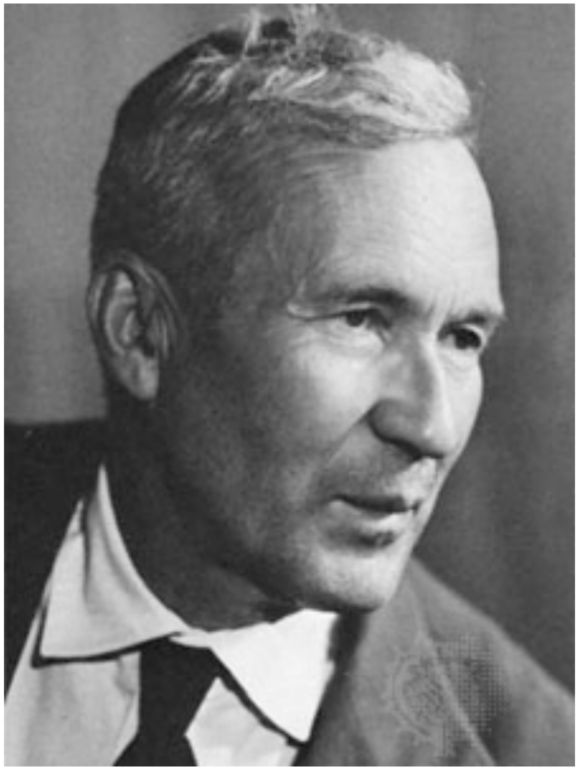
***And the inverse transformation provides for efficient M.C. generation of p(x) starting from RAN().**

Bob Cousins, CMS, 2008

Axioms of Probability

These Axioms are a mathematical starting point for probability and statistics

1. probability for every element, E, is non-negative $P(E) \geq 0 \qquad \forall E \subseteq \mathcal{F} = 2^{\Omega}$
2. probability for the entire space of possibilities is 1 $P(\Omega) = 1.$
3. if elements E_i are disjoint, probability is additive $P(E_1 \cup E_2 \cup \cdots) = \sum_i P(E_i).$



Kolmogorov axioms (1933)

Consequences:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(\Omega \setminus E) = 1 - P(E)$$

Different definitions of Probability

Frequentist

- defined as limit of long term frequency
- probability of rolling a 3 := limit of (# rolls with 3 / # trials)
  - you don't need an infinite sample for definition to be useful
  - sometimes ensemble doesn't exist
    - eg. P(Higgs mass = 120 GeV), P(it will snow tomorrow)
- Intuitive if you are familiar with Monte Carlo methods
- compatible with orthodox interpretation of probability in Quantum Mechanics. Probability to measure spin projected on x-axis if spin of beam is polarized along +z

$$|\langle \rightarrow \mid \uparrow \rangle|^2 = \frac{1}{2}$$

Subjective Bayesian

- Probability is a degree of belief (personal, subjective)
  - can be made quantitative based on betting odds
  - most people's subjective probabilities are not ***coherent*** and do not obey laws of probability

http://plato.stanford.edu/archives/sum2003/entries/probability-interpret/#3.1

Bayes’ Theorem

CENTER FOR
COSMOLOGY AND
PARTICLE PHYSICS

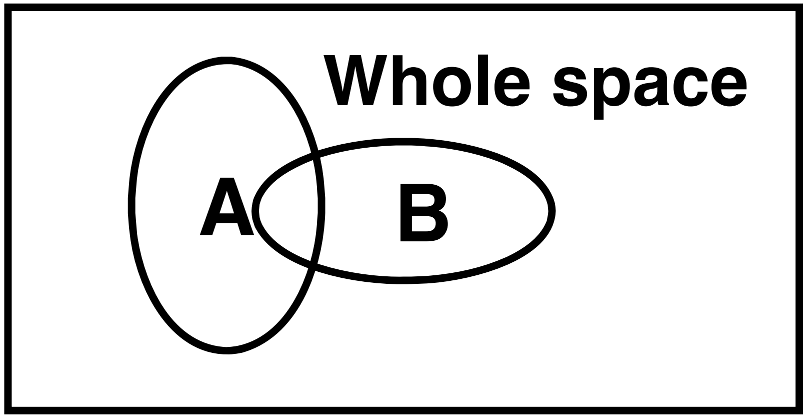
Bayes’ theorem relates the conditional and margi
events A & B

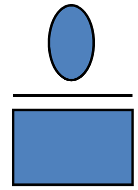
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

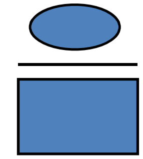
- P(*A*) is the prior probability or marginal probability of *A*. It is "prior" in the sense that it does not take into account any information about *B*.
- P(*A*|*B*) is the conditional probability of *A*, given *B*. It is also called the posterior probability because it is derived from or depends upon the specified value of *B*.
- P(*B* |*A*) is the conditional probability of *B* given *A*.
- P(*B*) is the prior or marginal probability of *B*, and acts as a normalizing constant.

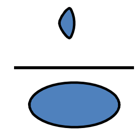
... in pictures (from Bob Cousins)

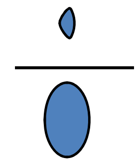
P, Conditional P, and Derivation of Bayes' Theorem in Pictures

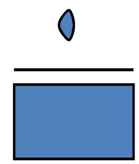


P(A) = 

P(B) = 

P(A|B) = 

P(B|A) = 

$P(A \cap B) =$ 

Don't forget about "Whole space" Ω. I will drop it from the notation typically, but occasionally it is important.

Bob Cousins, CMS, 2008

$\Rightarrow$ **P(B|A) = P(A|B) $\times$ P(B) / P(A)**

Louis's Example

$$P\,(\text{Data;Theory}) \neq P\,(\text{Theory;Data})$$

Theory = male or female

Data = pregnant or not pregnant

P (pregnant ; female) ~ 3%

but

P (female ; pregnant) >>>3%

Bayes’ Theorem

Bayes’ theorem relates the conditional and margi
events A & B

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\mathcal{N}} \propto L(\theta)\pi(\theta)$$

- P(*A*) is the prior probability or marginal probability of *A*. It is "prior" in the sense that it does not take into account any information about *B*.
- P(*A*|*B*) is the conditional probability of *A*, given *B*. It is also called the posterior probability because it is derived from or depends upon the specified value of *B*.
- P(*B* |*A*) is the conditional probability of *B* given *A*.
- P(*B*) is the prior or marginal probability of *B*, and acts as a normalizing constant.

Modeling:
The Scientific Narrative

Building a model of the data

Center for Cosmology and Particle Physics

Before one can discuss statistical tests, one must have a “**model**” for the data.

- by “model”, I mean the full structure of P(data | parameters)
  - holding parameters fixed gives a PDF for data
  - ability to evaluate generate pseudo-data (Toy Monte Carlo)
  - holding data fixed gives a **likelihood function** for parameters
    - note, likelihood function is not as general as the full model because it doesn’t allow you to generate pseudo-data

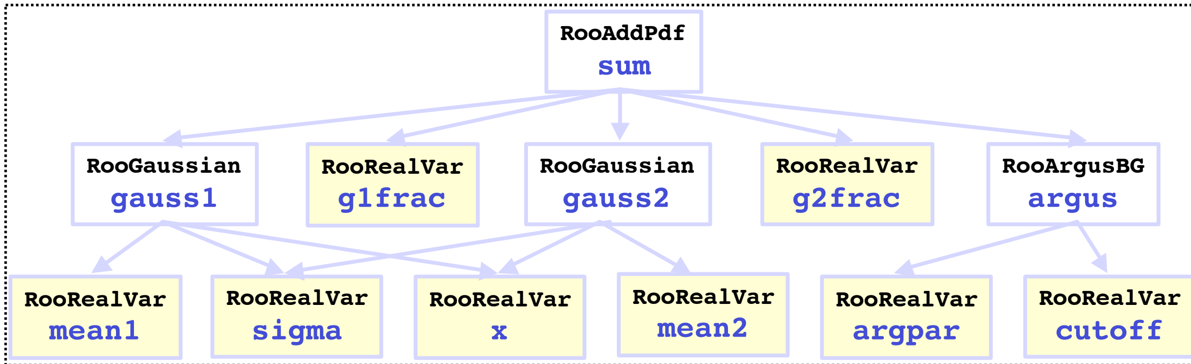
Both Bayesian and Frequentist methods start with the model

- it’s the objective part that everyone can agree on
- it’s the place where our physics knowledge, understanding, and intuiting comes in
- building a better model is the best way to improve your statistical procedure

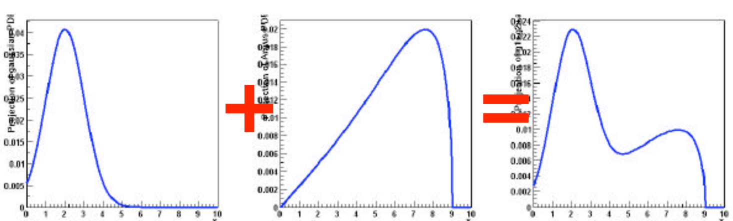
RooFit: A data modeling toolkit

Center for Cosmology and Particle Physics

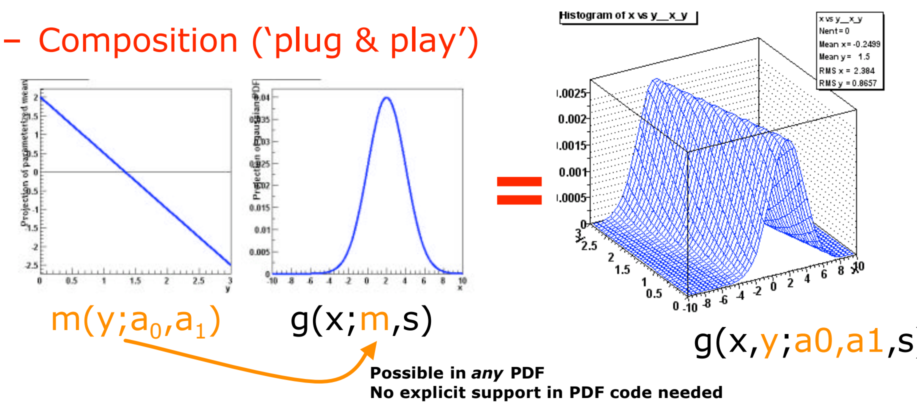
RooFit is a major tool developed at BaBar for data modeling.
RooStats provides higher-level statistical tools based on these PDFs.



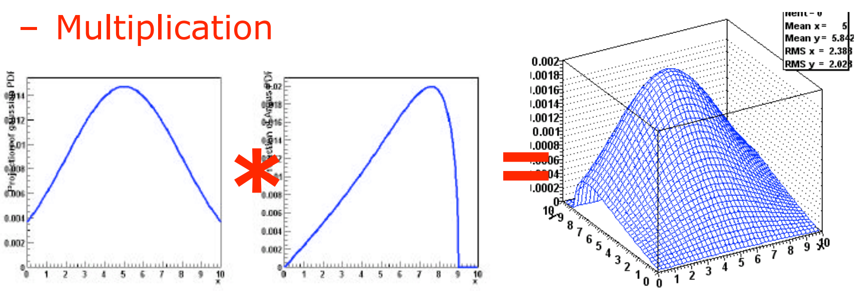
– Addition



– Composition ('plug & play')

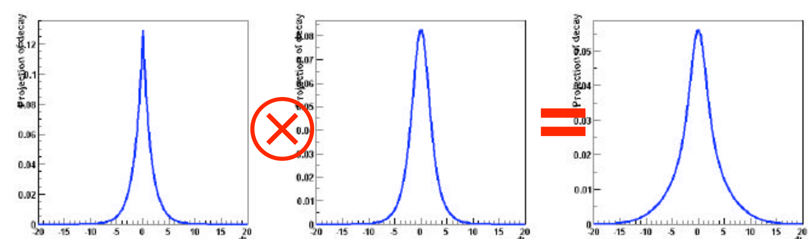


– Multiplication



Wouter Verkerke,

– Convolution



Wouter Verkerke, UCSB

The Scientific Narrative

The model can be seen as a quantitative summary of the analysis

- If you were asked to justify your modeling, you would tell a **story** about why you know what you know
  - based on previous results and studies performed along the way
- the quality of the result is largely tied to how convincing this story is and how tightly it is connected to model

I will describe a few "narrative styles"

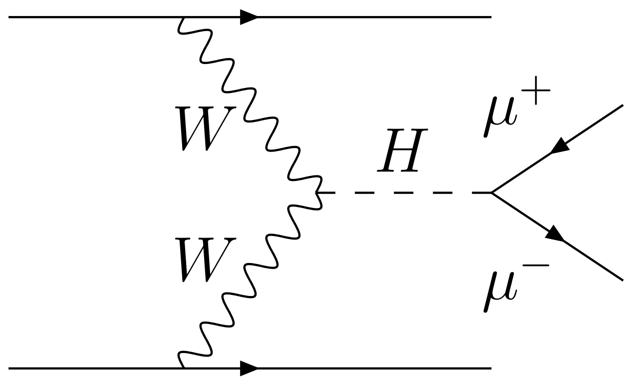
- The "Monte Carlo Simulation" narrative
- The "Data Driven" narrative
- The "Effective Modeling" narrative
- The "Parametrized Response" narrative

Real-life analyses often use a mixture of these

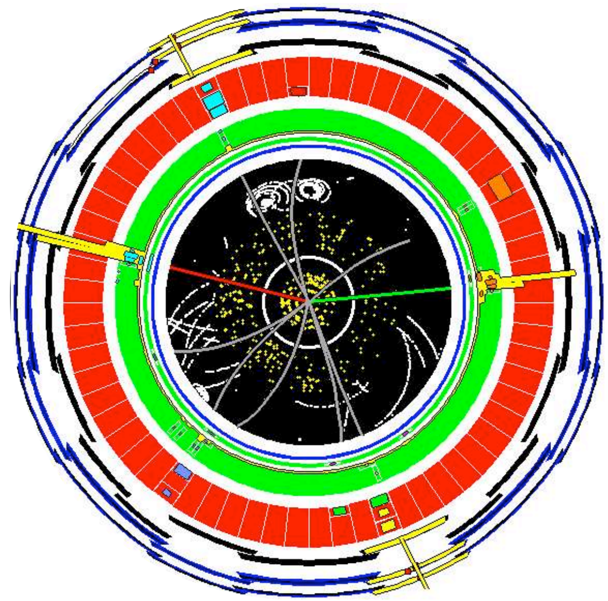
The Monte Carlo Simulation narrative

Let's start with "the Monte Carlo simulation narrative", which is probably the most familiar

$P =$



$\oplus$

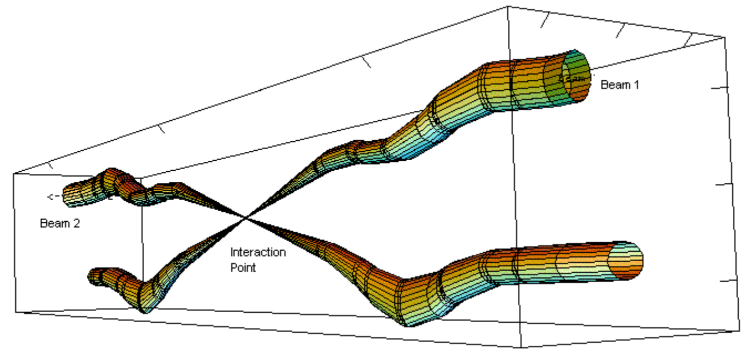


1) The language of the Standard Model is Quantum Field Theory
Phase space Ω defines initial measure, sampled via Monte Carlo

$$P = \frac{|\langle f|i\rangle|^2}{\langle f|f\rangle\langle i|i\rangle}$$

$$P \to L\sigma$$

$$d\sigma \to |\mathcal{M}|^2 d\Omega$$



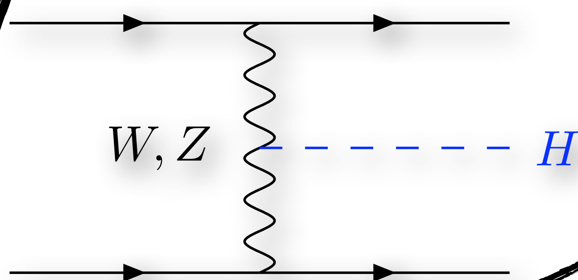
Relative beam sizes around IP1 (Atlas) in collision

$$\mathcal{L}_{SM} = \underbrace{\frac{1}{4}\mathbf{W}_{\mu\nu}\cdot\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a}_{\text{kinetic energies and self-interactions of the gauge bosons}}$$

$$+ \underbrace{\bar{L}\gamma^\mu(i\partial_\mu - \frac{1}{2}g\tau\cdot\mathbf{W}_\mu - \frac{1}{2}g'YB_\mu)L + \bar{R}\gamma^\mu(i\partial_\mu - \frac{1}{2}g'YB_\mu)R}_{\text{kinetic energies and electroweak interactions of fermions}}$$

$$+ \underbrace{\frac{1}{2}\left|(i\partial_\mu - \frac{1}{2}g\tau\cdot\mathbf{W}_\mu - \frac{1}{2}g'YB_\mu)\,\phi\right|^2 - V(\phi)}_{W^\pm, Z, \gamma, \text{and Higgs masses and couplings}}$$

$$+ \underbrace{g''(\bar{q}\gamma^\mu T_a q)\,G^a_\mu}_{\text{interactions between quarks and gluons}} + \underbrace{(G_1\bar{L}\phi R + G_2\bar{R}\phi_c L + h.c.)}_{\text{fermion masses and couplings to Higgs}}$$

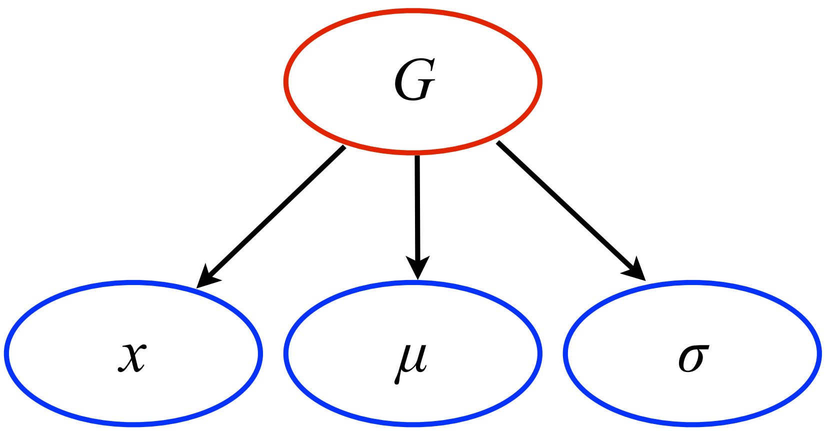


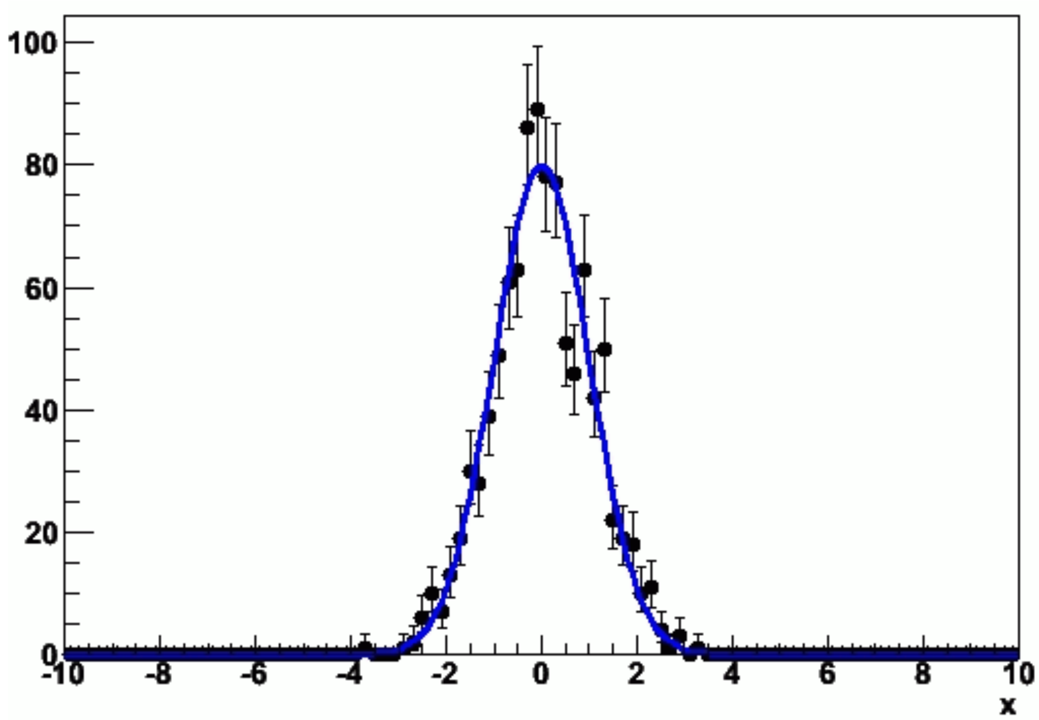
A General Purpose Statistical Model

Visualizing probability models

I will represent PDFs graphically as below (directed acyclic graph)

- eg. a Gaussian $G(x|\mu,\sigma)$ is parametrized by (μ,σ)
- every node is a real-valued function of the nodes below





Marked Poisson Process

Channel: a subset of the data defined by some selection requirements.

- eg. all events with 4 electrons with energy > 10 GeV
- n: number of events observed in the channel
- ν: number of events expected in the channel

Discriminating variable: a property of those events that can be measured and which helps discriminate the signal from background

- eg. the invariant mass of two particles
- $f(x)$: the p.d.f. of the discriminating variable x

$$\mathcal{D} = \{x_1, \ldots, x_n\}$$

Marked Poisson Process:

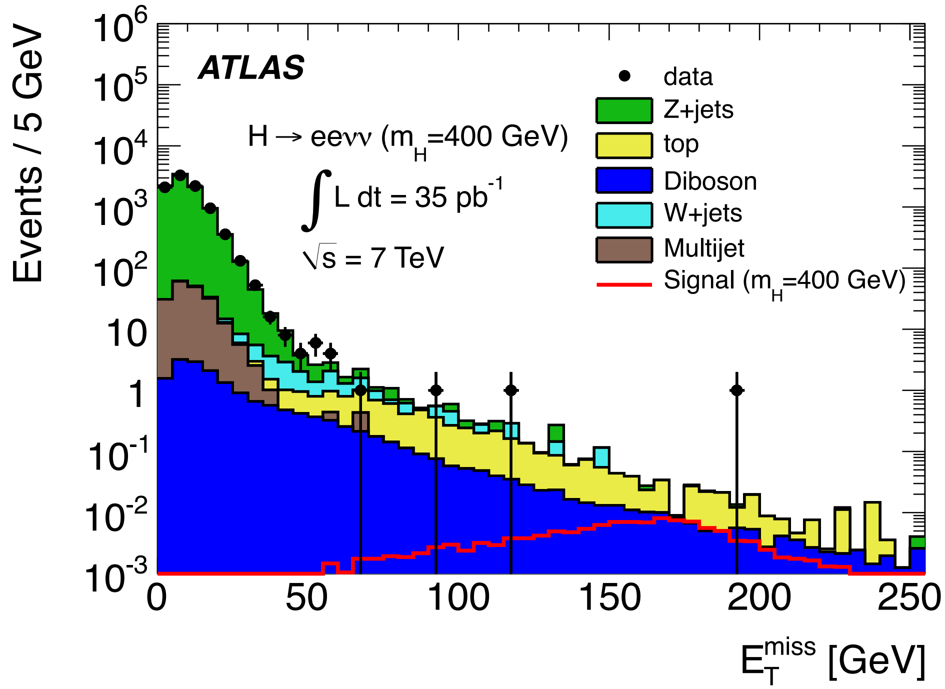
$$\mathbf{f}(\mathcal{D}|\nu) = \mathrm{Pois}(n|\nu) \prod_{e=1}^{n} f(x_e)$$

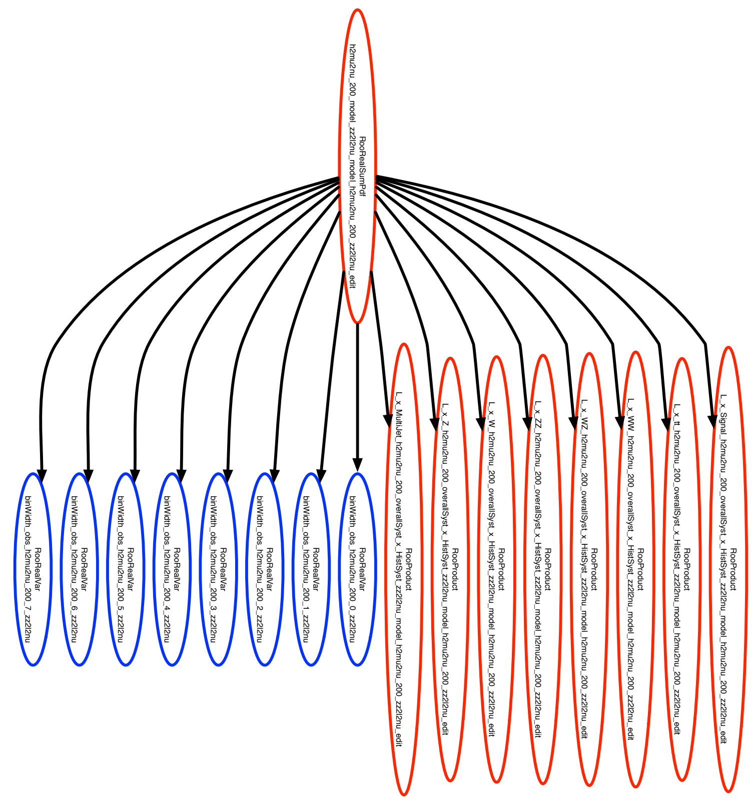
Mixture model

Sample: a sample of simulated events corresponding to particular type interaction that populates the channel.

- statisticians call this a mixture model

$$f(x) = \frac{1}{\nu_{\rm tot}} \sum_{s \in {\rm samples}} \nu_s f_s(x) \; , \qquad \nu_{\rm tot} = \sum_{s \in {\rm samples}} \nu_s$$





Parametrizing the model $\boldsymbol{\alpha} = (\mu, \boldsymbol{\theta})$

Parameters of interest (*μ*): parameters of the theory that modify the rates and shapes of the distributions, eg.

- the mass of a hypothesized particle
- the “signal strength” μ=0 no signal, μ=1 predicted signal rate

Nuisance parameters (*θ* or α_p**):** associated to uncertainty in:

- response of the detector (calibration)
- phenomenological model of interaction in non-perturbative regime

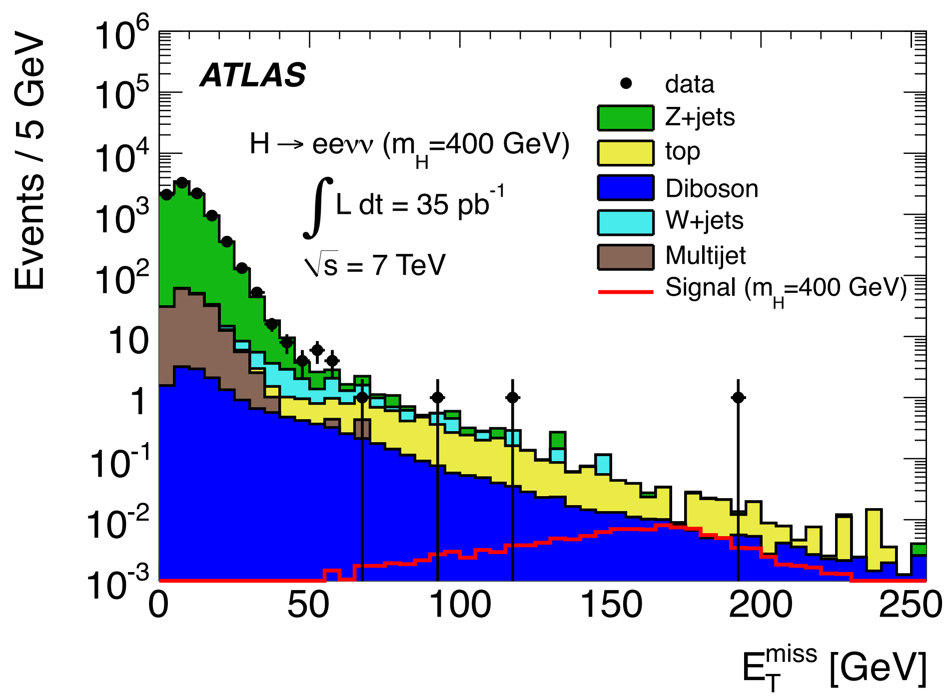
Lead to a parametrized model: $\nu \to \nu(\boldsymbol{\alpha}), f(x) \to f(x|\boldsymbol{\alpha})$

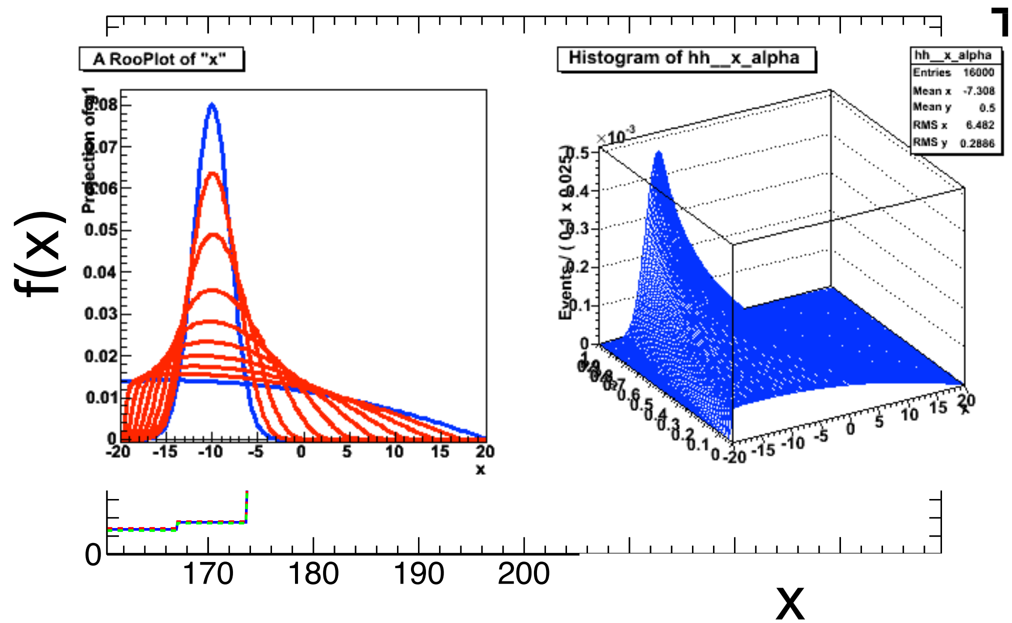
$$\mathbf{f}(\mathcal{D}|\boldsymbol{\alpha}) = \mathrm{Pois}(n|\nu(\boldsymbol{\alpha})) \prod_{e=1}^{n} f(x_e|\boldsymbol{\alpha})$$

Incorporating Systematic Effects

Tabulate effect of individual variations of sources of systematic uncertainty

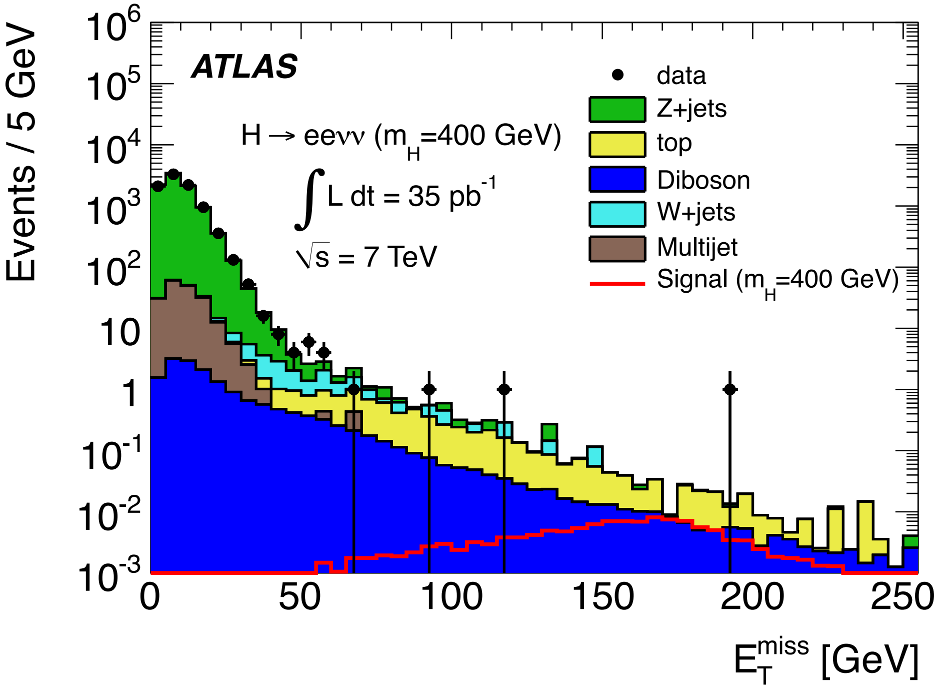
- typically one at a time evaluated at nominal and “± 1 σ”
- use some form of interpolation to parametrize p^{th} variation in terms of **nuisance parameter** α_p

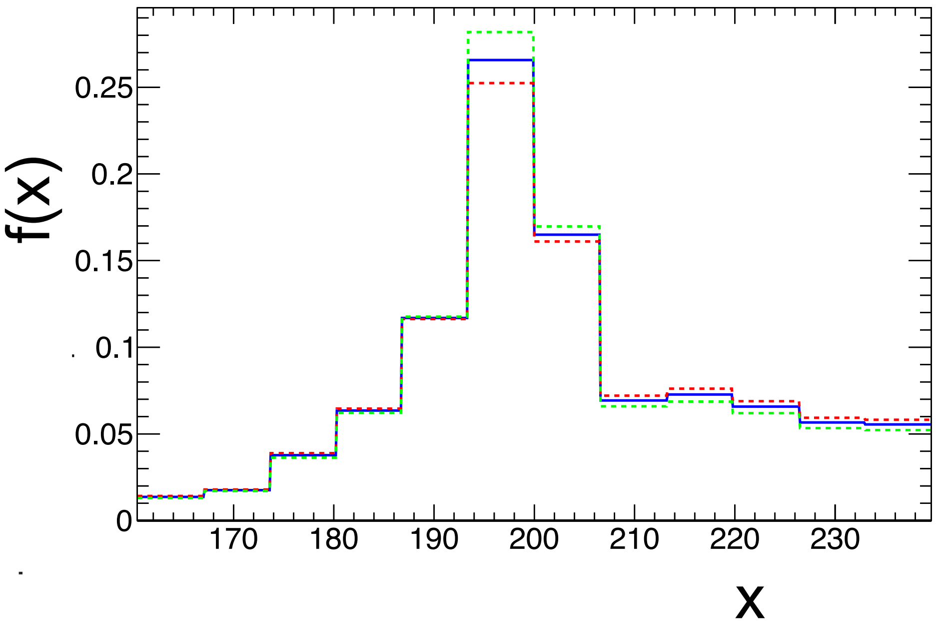


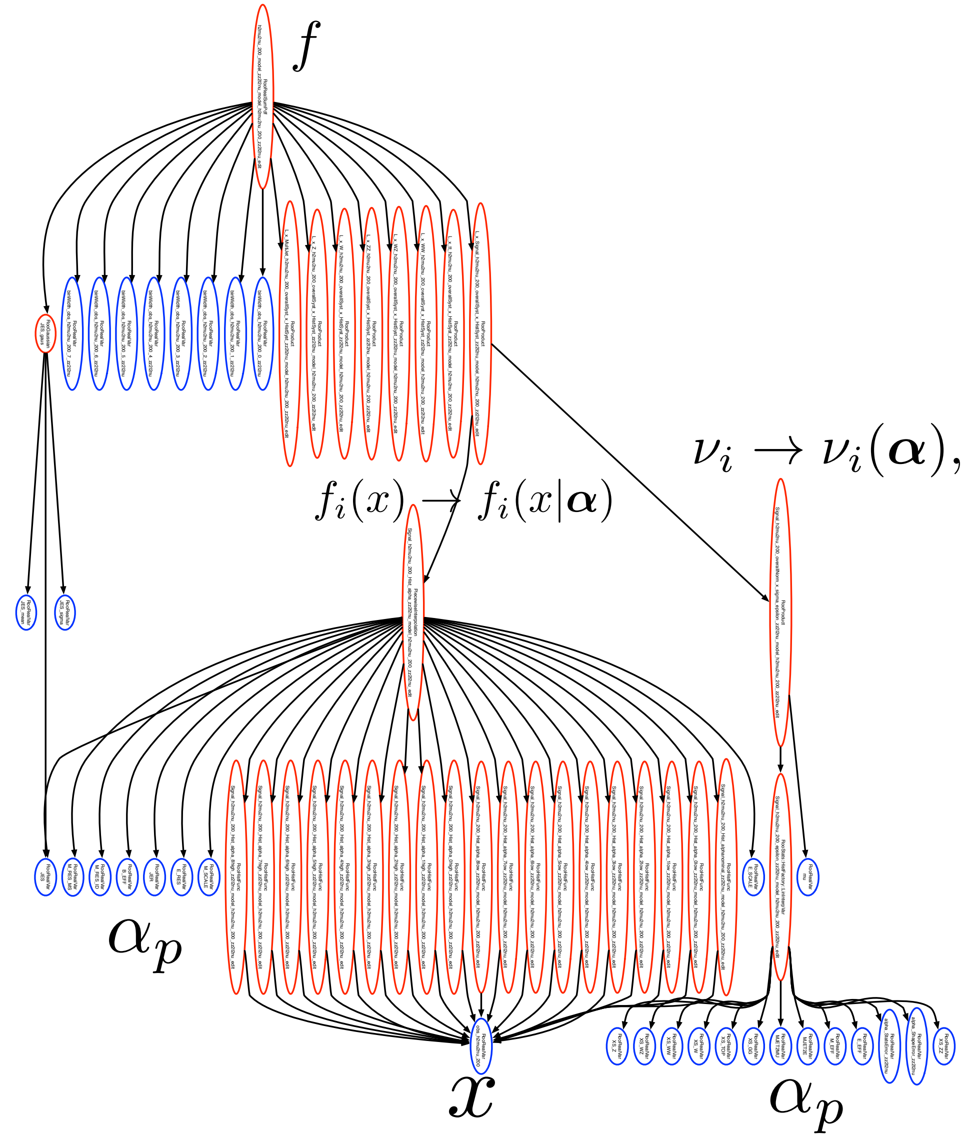


$$\mathbf{f}(\mathcal{D}|\boldsymbol{\alpha}) = \mathrm{Pois}(n|\nu(\boldsymbol{\alpha})) \prod_{e=1}^{n} f(x_e|\boldsymbol{\alpha})$$

Visualizing the model for one channel

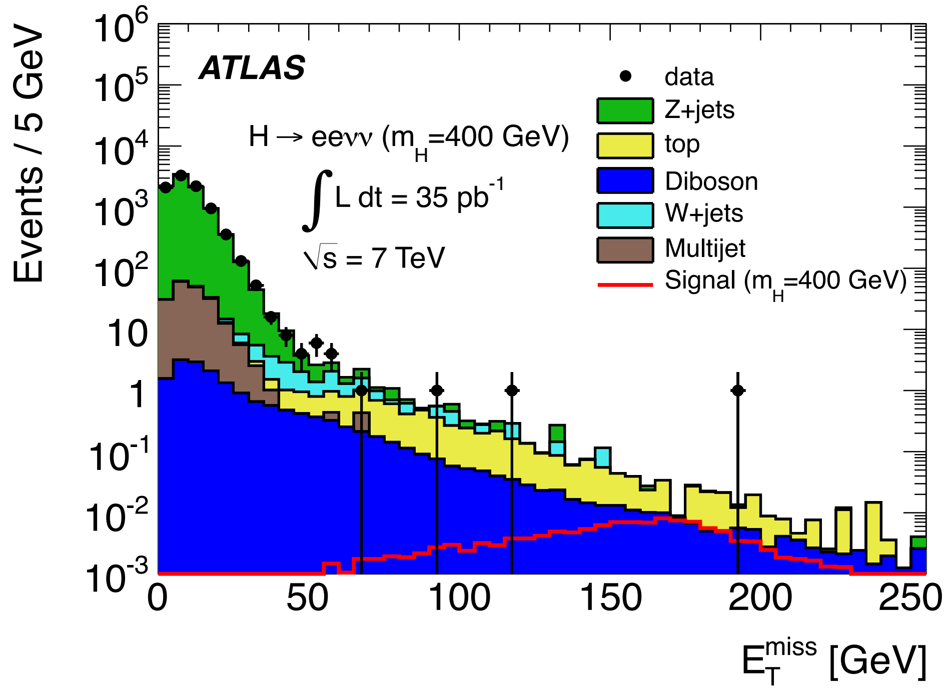


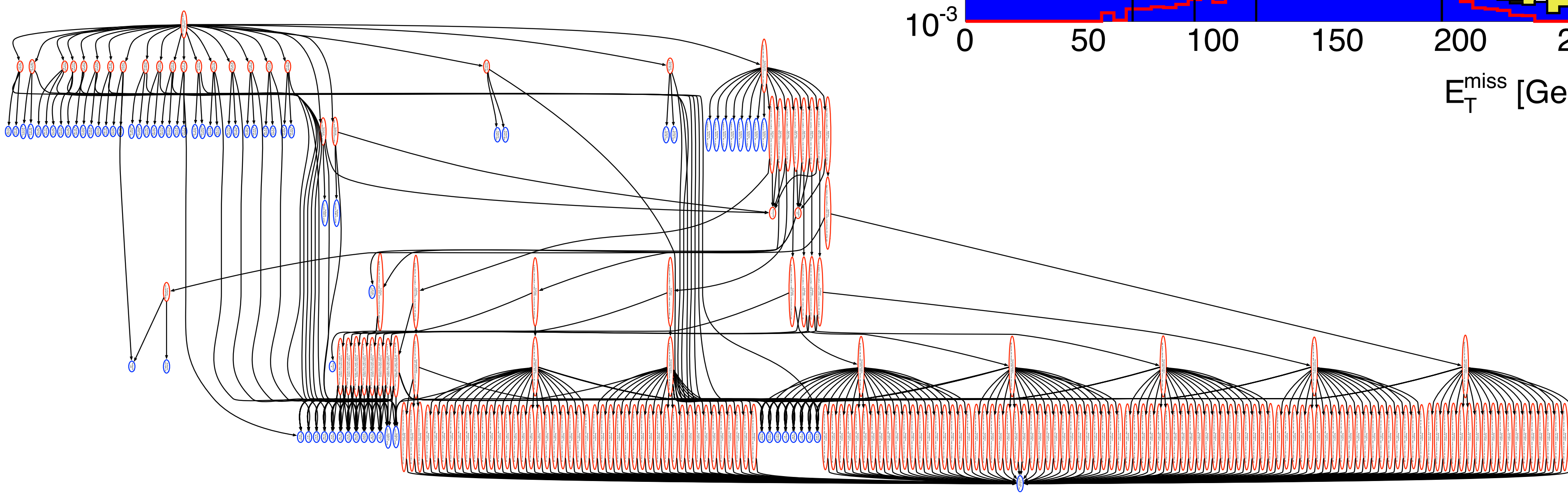




Visualizing the model for one channel

After parametrizing each component of the mixture model, the pdf for a single channel might look like this





Simultaneous multi-channel model

Simultaneous Multi-Channel Model: Several disjoint regions of the data are modeled simultaneously. Identification of common parameters across many channels requires coordination between groups such that meaning of the parameters are really the same.

$$\mathbf{f}_{\rm sim}(\mathcal{D}_{\rm sim}|\boldsymbol{\alpha}) = \prod_{c\in {\rm channels}} \left[{\rm Pois}(n_c|\nu_c(\boldsymbol{\alpha})) \prod_{e=1}^{n_c} f_c(x_{ce}|\boldsymbol{\alpha}) \right]$$

where $\mathcal{D}_{\rm sim} = \{\mathcal{D}_1, \ldots, \mathcal{D}_{c_{\rm max}}\}$

Control Regions: Some channels are not populated by signal processes, but are used to constrain the nuisance parameters

- attempt to describe systematics in a statistical language
- Prototypical Example: “on/off” problem with unknown ν_b

$$\mathbf{f}(n,m|\mu,\nu_b) = \underbrace{{\rm Pois}(n|\mu+\nu_b)}_{\text{signal region}} \cdot \underbrace{{\rm Pois}(m|\tau\nu_b)}_{\text{control region}}$$

Constraint terms

Often detailed statistical model for auxiliary measurements that measure certain nuisance parameters are not available.

- one typically has MLE for α_p, denoted a_p and standard error

Constraint Terms: are idealized pdfs for the MLE.

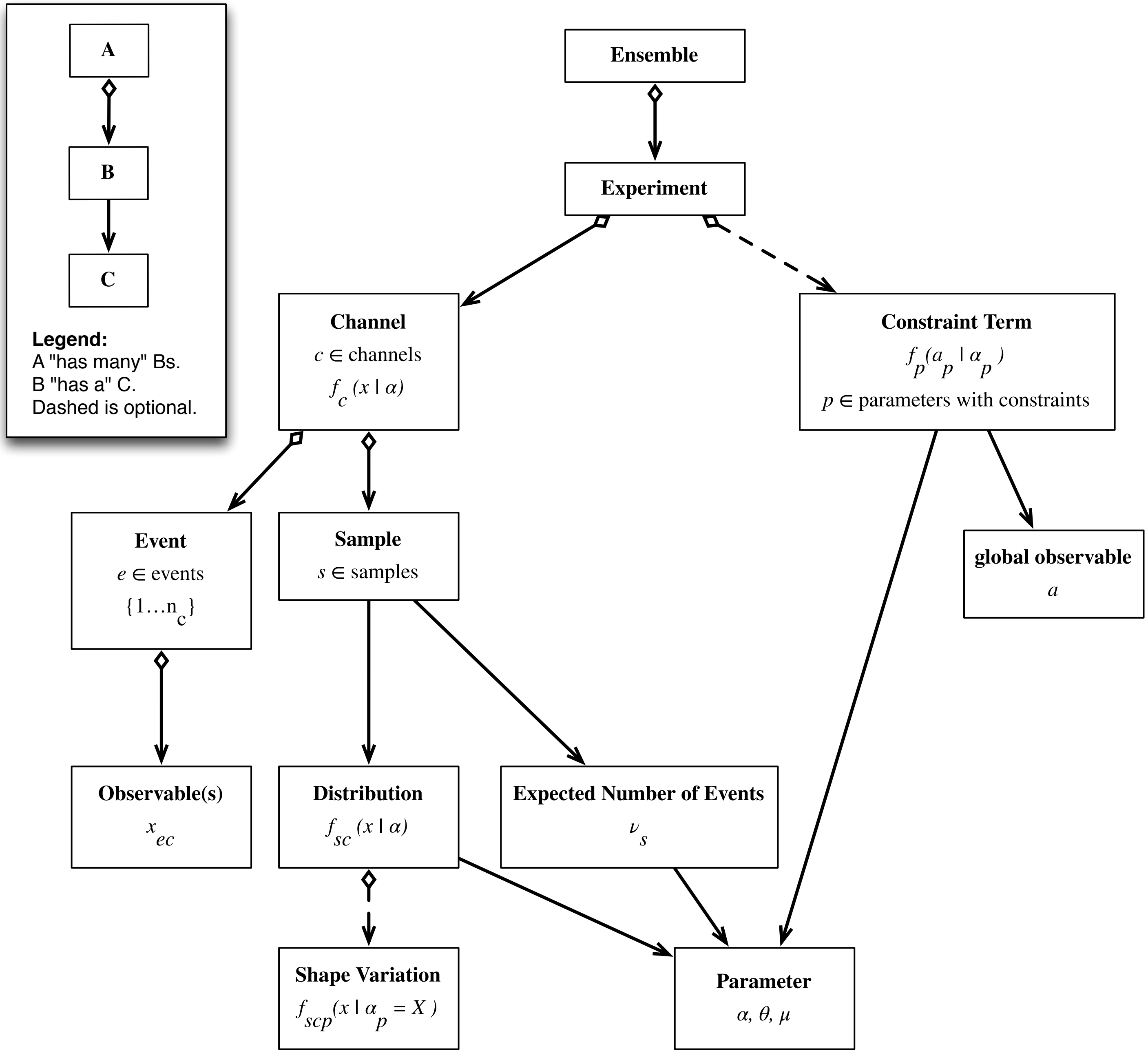
$$f_p(a_p|\alpha_p) \quad \text{for} \quad p \in \mathbb{S}$$

- common choices are Gaussian, Poisson, and log-normal
- New: careful to write constraint term a frequentist way
- Previously: $\pi(\alpha_p|a_p) = f_p(a_p|\alpha_p)\eta(\alpha_p)$ with uniform η

Simultaneous Multi-Channel Model with constraints:

$$\mathbf{f}_{\text{tot}}(\mathcal{D}_{\text{sim}}, \mathcal{G}|\boldsymbol{\alpha}) = \prod_{c \in \text{channels}} \left[\text{Pois}(n_c|\nu_c(\boldsymbol{\alpha})) \prod_{e=1}^{n_c} f_c(x_{ce}|\boldsymbol{\alpha}) \right] \cdot \prod_{p \in \mathbb{S}} f_p(a_p|\alpha_p)$$

where $\mathcal{D}_{\text{sim}} = \{\mathcal{D}_1, \ldots, \mathcal{D}_{c_{\max}}\}, \quad \mathcal{G} = \{a_p\} \quad \text{for} \quad p \in \mathbb{S}$


A
B
C
Legend:
A "has many" Bs.
B "has a" C.
Dashed is optional.
Ensemble
Experiment
Channel
$c \in$ channels
$f_c(x \mid \alpha)$
Constraint Term
$f_p(a_p \mid \alpha_p)$
$p \in$ parameters with constraints
Event
$e \in$ events
$\{1 \ldots n_c\}$
Sample
$s \in$ samples
global observable
a
Observable(s)
x_{ec}
Distribution
$f_{sc}(x \mid \alpha)$
Expected Number of Events
ν_s
Shape Variation
$f_{scp}(x \mid \alpha_p = X)$
Parameter
α, θ, μ

Combined ATLAS Higgs Search

State of the art: At the time of the discovery, the combined Higgs search included 100 disjoint channels and >500 nuisance parameters

- Models for individual channels come from about 11 sub-groups performing dedicated searches for specific Higgs decay modes
- In addition low-level performance groups provide tools for evaluating systematic effects and corresponding constraint terms

Higgs Decay	Subsequent Decay	Additional Sub-Channels	m_H Range	L [fb^{-1}]
$H \to \gamma\gamma$	–	9 sub-channels ($p_{\mathrm{T}_t} \otimes \eta_\gamma \otimes$ conversion)	110-150	4.9
$H \to ZZ$	$\ell\ell\ell'\ell'$	$\{4e, 2e2\mu, 2\mu2e, 4\mu\}$	110-600	4.8
	$\ell\ell\nu\nu$	$\{ee, \mu\mu\} \otimes$ {low pile-up, high pile-up}	200-280-600	4.7
	$\ell\ell qq$	{b-tagged, untagged}	200-300-600	4.7
$H \to WW$	$\ell\nu\ell\nu$	$\{ee, e\mu, \mu\mu\} \otimes$ {0-jet, 1-jet, VBF}	110-300-600	4.7
	$\ell\nu q\overline{q'}$	$\{e, \mu\} \otimes$ {0-jet, 1-jet}	300-600	4.7
$H \to \tau^+\tau^-$	$\ell\ell 4\nu$	$\{e\mu\} \otimes$ {0-jet} $\oplus$ {1-jet, VBF, VH}	110-150	4.7
	$\ell\tau_{\mathrm{had}}3\nu$	$\{e, \mu\} \otimes$ {0-jet} $\otimes$ $\{E_T^{\mathrm{miss}} \gtrsim 20$ GeV$\}$ $\oplus$ $\{e, \mu\} \otimes$ {1-jet, VBF}	110-150	4.7
	$\tau_{\mathrm{had}}\tau_{\mathrm{had}}2\nu$	{1-jet}	110-150	4.7
$VH \to b\bar{b}$	$Z \to \nu\bar{\nu}$	$E_T^{\mathrm{miss}} \in \{120-160, 160-200, \geq 200$ GeV$\}$	110-130	4.6
	$W \to \ell\nu$	$p_T^W \in \{< 50, 50-100, 100-200, \geq 200$ GeV$\}$	110-130	4.7
	$Z \to \ell\ell$	$p_T^Z \in \{< 50, 50-100, 100-200, \geq 200$ GeV$\}$	110-130	4.7

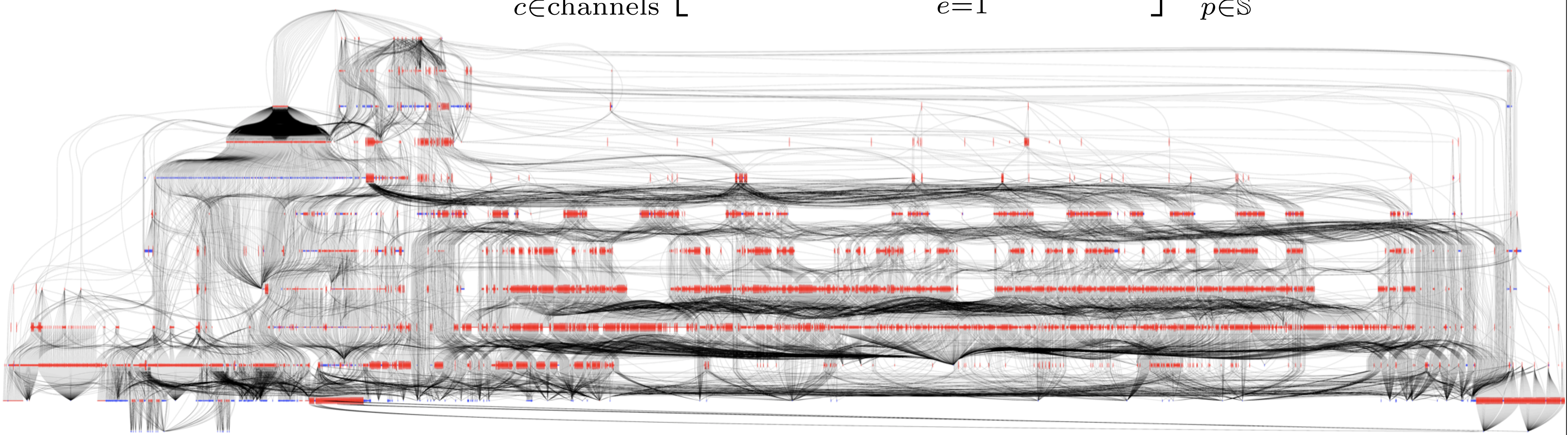
Visualizing the combined model

State of the art: At the time of the discovery, the combined Higgs search included 100 disjoint channels and >500 nuisance parameters

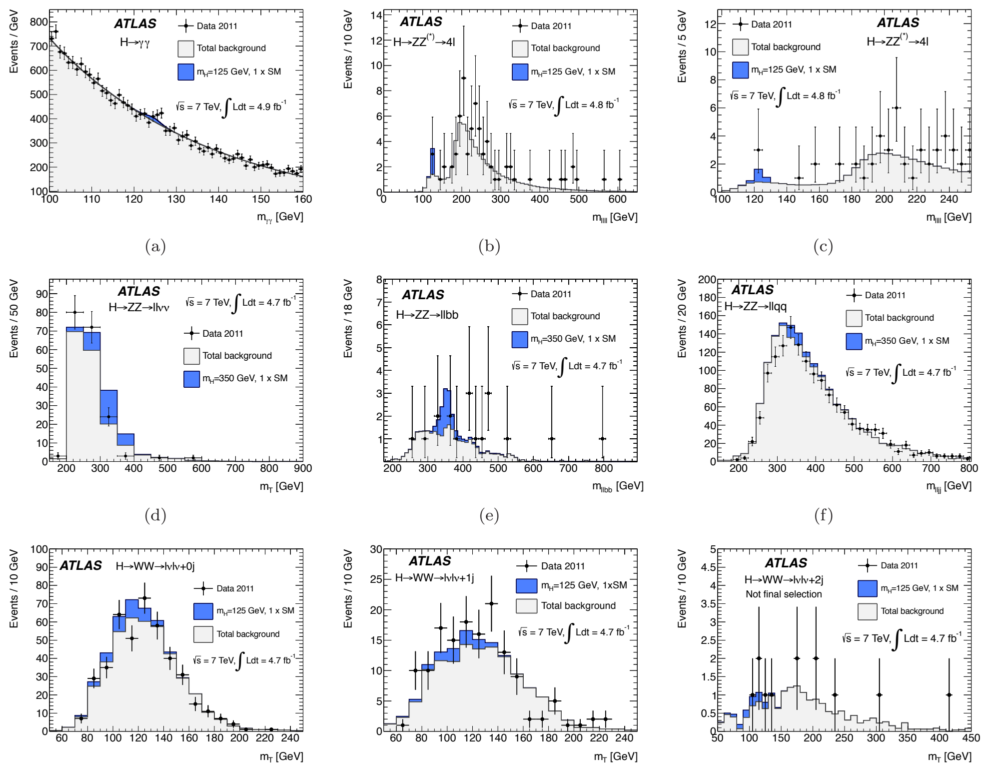
RooFit / RooStats: is the modeling language (C++) which provides technologies for collaborative modeling

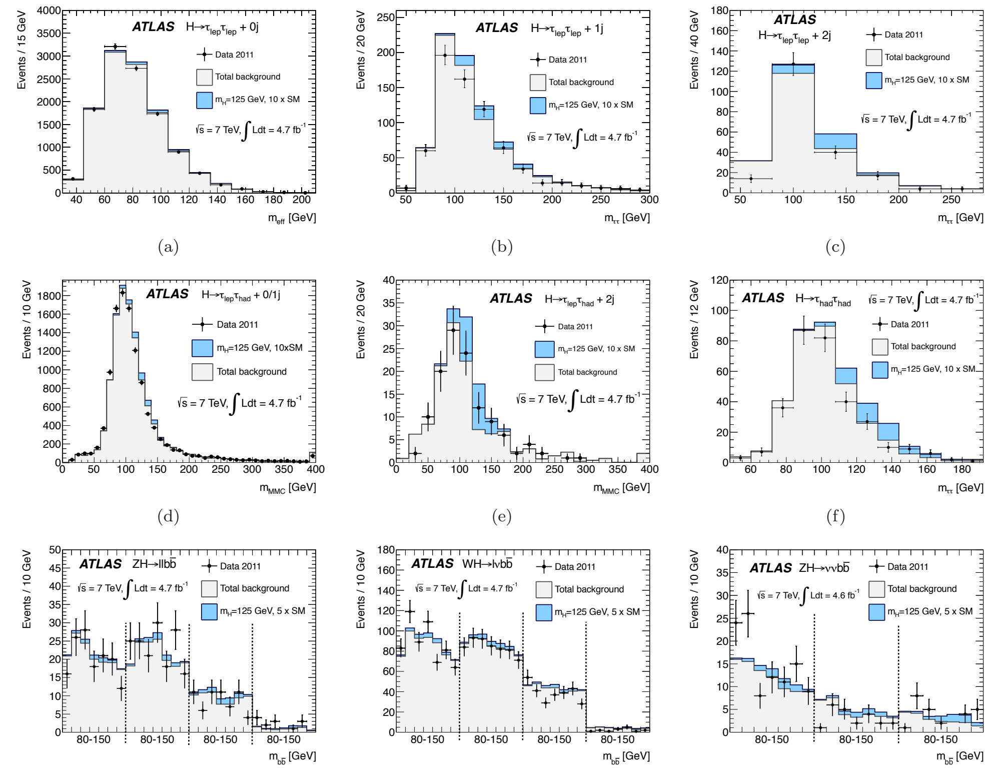
- provides technology to publish likelihood functions digitally
- and more, it's the full model so we can also generate pseudo-data

$$\mathbf{f}_{\rm tot}(\mathcal{D}_{\rm sim}, \mathcal{G}|\boldsymbol{\alpha}) = \prod_{c \in \rm channels} \left[\mathrm{Pois}(n_c|\nu_c(\boldsymbol{\alpha})) \prod_{e=1}^{n_c} f_c(x_{ce}|\boldsymbol{\alpha}) \right] \cdot \prod_{p \in \mathbb{S}} f_p(a_p|\alpha_p)$$

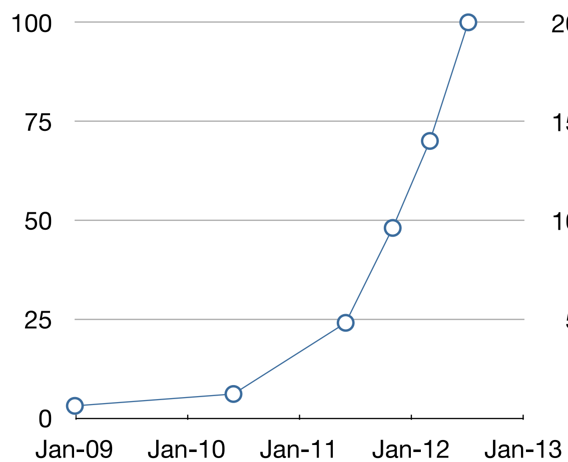


Evolution of Model Complexity

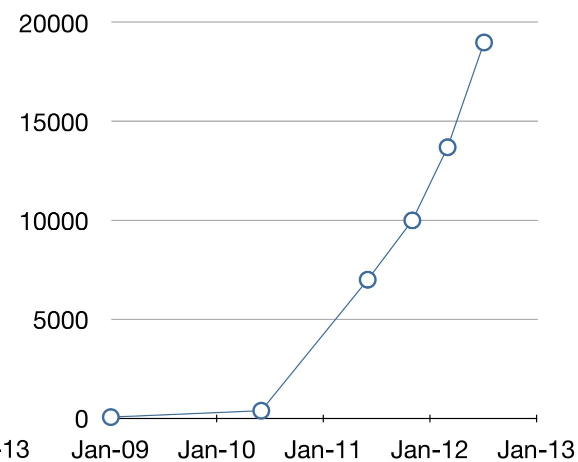




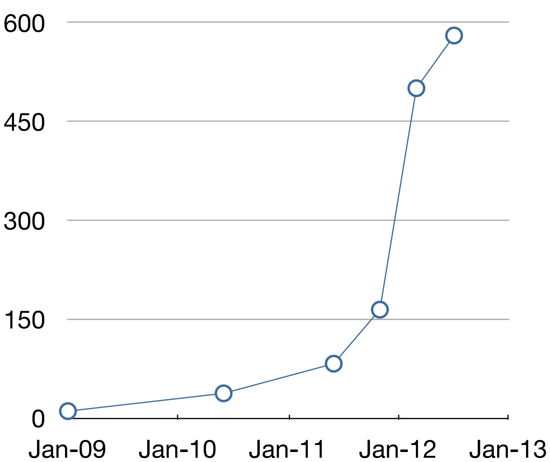
Number of Datasets Combined



Number of Model Components



Number of Parameters in Likelihood



Constraint Terms
Auxiliary Measurements
and Priors on Nuisance Parameters

What do we mean by uncertainty?

Let's consider 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
  - in our main measurement we observe n_{on} with $s+b$ expected

$$\mathrm{Pois}(n_{\mathrm{on}}|s+b)$$

- and the background has some uncertainty
  - but what is "background uncertainty"? Where did it come from?
  - maybe we would say background is known to 10% or that it has some pdf $\pi(b)$
    - then we often do a **smearing** of the background:

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

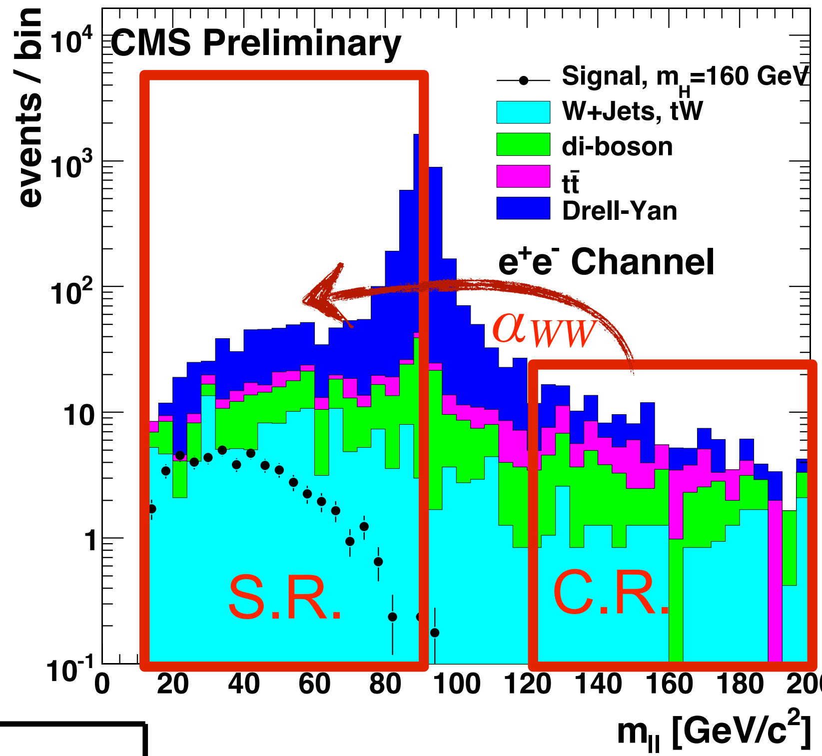
  - Where does $\pi(b)$ come from?
    - did you realize that this is a Bayesian procedure that depends on some prior assumption about what b is?

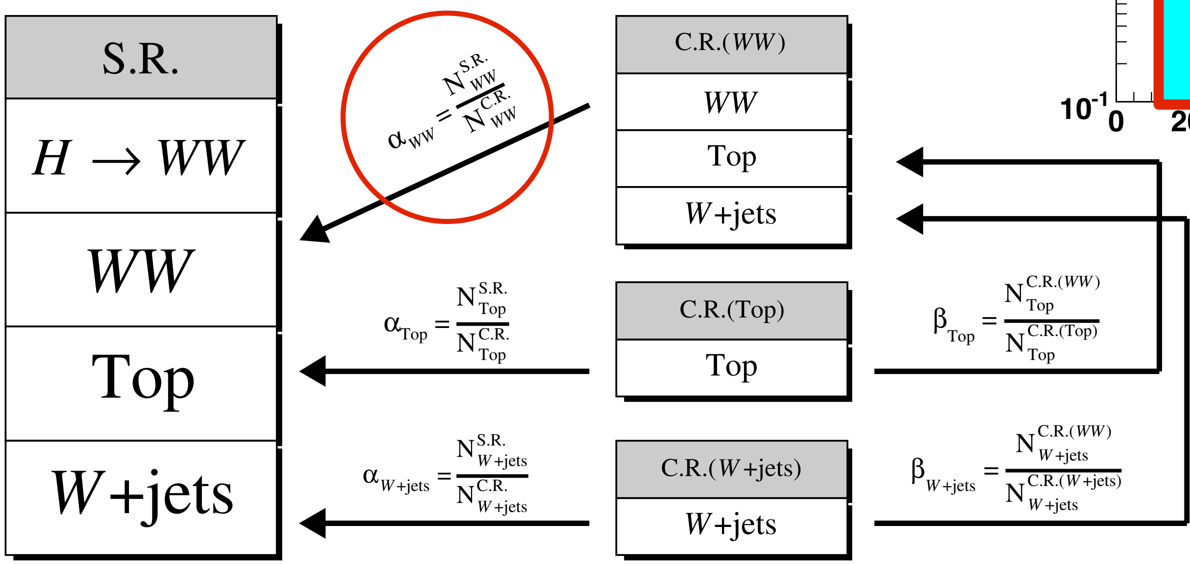
The Data-driven narrative

Center for Cosmology and Particle Physics

Regions in the data with negligible signal expected are used as control samples

- simulated events are used to estimate extrapolation coefficients
- extrapolation coefficients may have theoretical and experimental uncertainties





Notation for next slides:

\# in S.R. $\to n_{\text{on}}$

\# in C.R. $\to n_{\text{off}}$

$\alpha_{WW} \to \tau$

Figure 10: Flow chart describing the four data samples used in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis. S.R and C.R. stand for signal and control regions, respectively.

The "on/off" problem

Center for Cosmology and Particle Physics

Now let's say that the background was estimated from some control region or sideband measurement.

- We can treat these two measurements simultaneously:
  - main measurement: observe n_{on} with $s+b$ expected
  - sideband measurement: observe n_{off} with τb expected

$$\underbrace{P(n_{\rm on}, n_{\rm off}|s,b)}_{\text{joint model}} = \underbrace{\mathrm{Pois}(n_{\rm on}|s+b)}_{\text{main measurement}}\ \underbrace{\mathrm{Pois}(n_{\rm off}|\tau b)}_{\text{sideband}}$$

  - In this approach "background uncertainty" is a statistical error
  - justification and accounting of background uncertainty is much more clear

How does this relate to the smearing approach?

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

- while $\pi(b)$ is based on data, it still depends on some original prior $\eta(b)$

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

Common Constraints Terms

Center for Cosmology and Particle Physics

Many uncertainties have no clear statistical description or it is impractical to provide

Traditionally, we use Gaussians, but for large uncertainties it is clearly a bad choice

- quickly falling tail, bad behavior near physical boundary, optimistic p-values, ...

For systematics constrained from control samples and dominated by statistical uncertainty, a Gamma distribution is a more natural choice [PDF is Poisson for the control sample]

- longer tail, good behavior near boundary, natural choice if auxiliary is based on counting

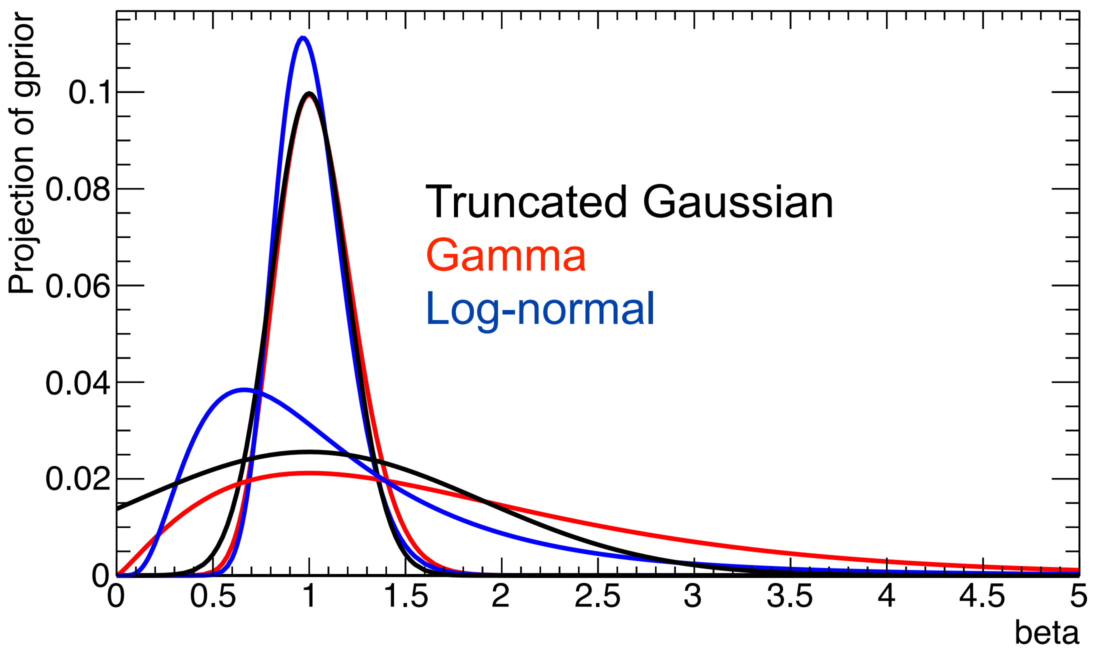
For "factor of 2" notions of uncertainty log-normal is a good choice

- can have a very long tail for large uncertainties

None of them are as good as an actual model for the auxiliary measurement, if available

To consistently switch between frequentist, Bayesian, and hybrid procedures, need to be clear about prior vs. likelihood function

PDF(y| β)	Prior(β)	Posterior(β|y)
Gaussian	uniform	Gaussian
Poisson	uniform	Gamma
Log-normal	1/β	Log-Normal



Classification of Systematic Uncertainties

Taken from Pekka Sinervo's PhyStat 2003 contribution

Type I - "The Good"

- can be constrained by other sideband/auxiliary/ancillary measurements and can be treated as statistical uncertainties
  - scale with luminosity

Type II - "The Bad"

- arise from model assumptions in the measurement or from poorly understood features in data or analysis technique
  - don't necessarily scale with luminosity
  - eg: "shape" systematics

Type III - "The Ugly"

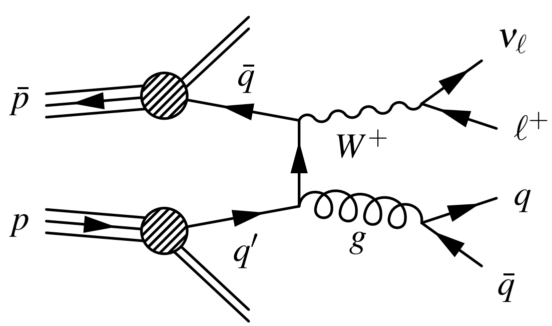
- arise from uncertainties in underlying theoretical paradigm used to make inference using the data
  - a somewhat philosophical issue

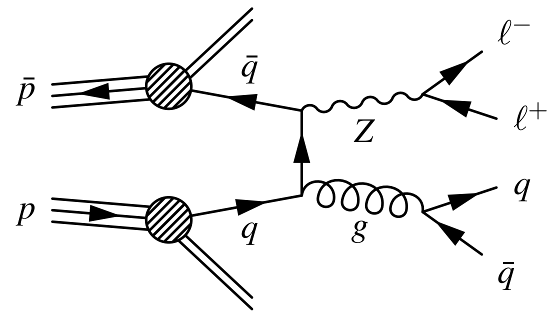
Modeling:
The Scientific Narrative
(continued)

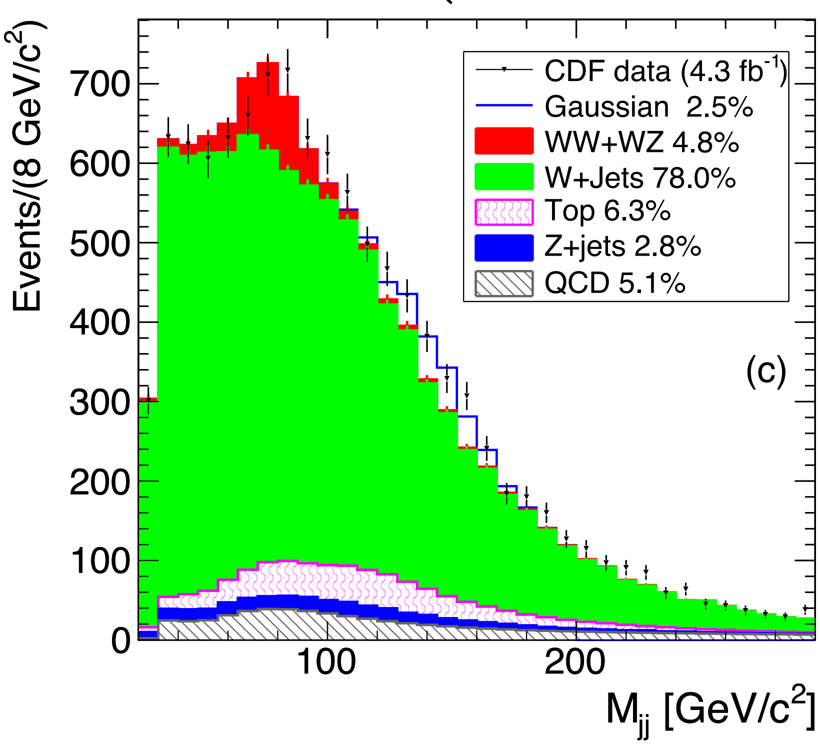
Choice: Data driven vs. Simulation

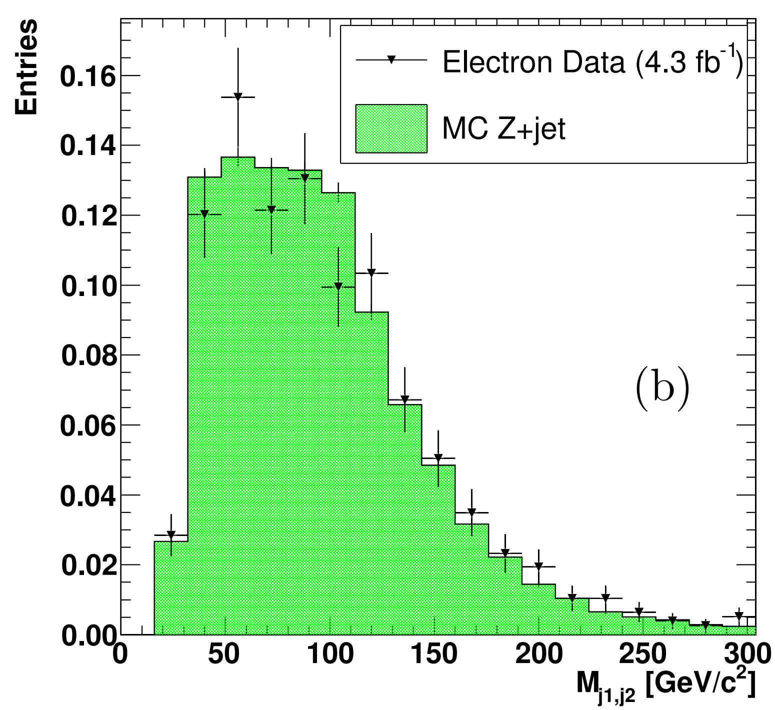
CENTER FOR COSMOLOGY AND PARTICLE PHYSICS

In the case of the CDF bump, the Z+jets control sample provides a data-driven estimate, but limited statistics. Using the simulation narrative over the data-driven is a **choice**. If you trust that narrative, it's a good choice.







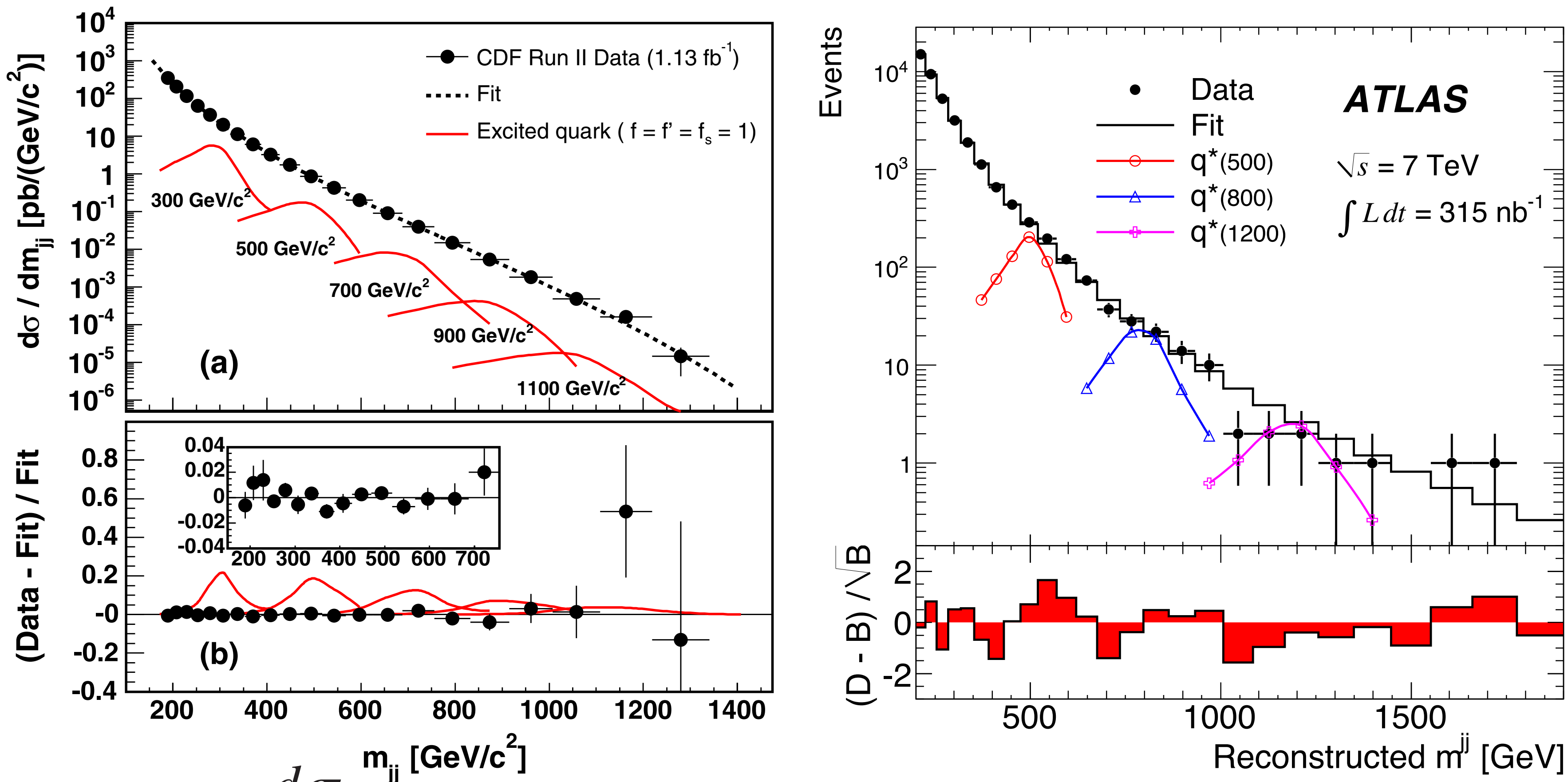


The Effective Model Narrative

It is common to describe a distribution with some parametric function

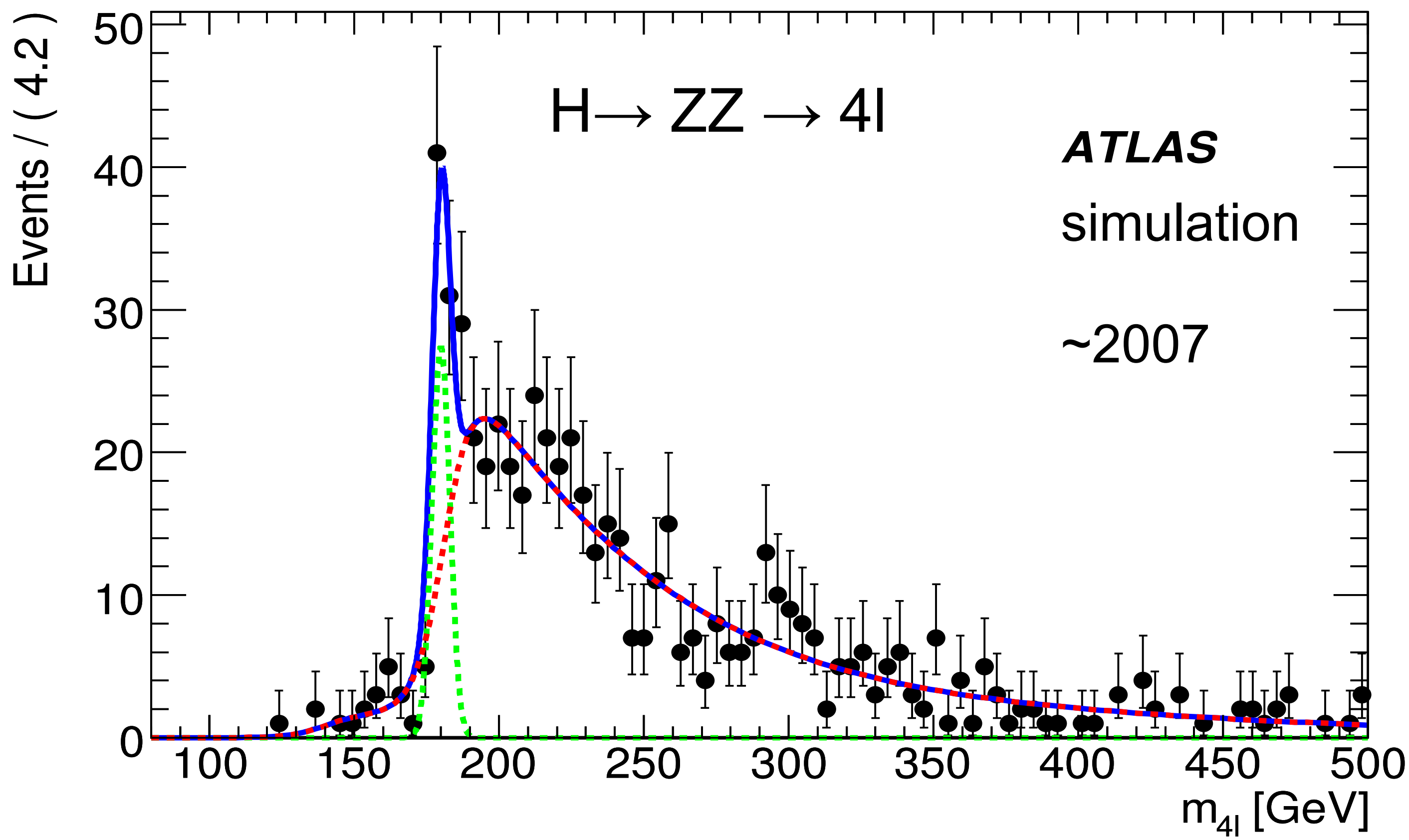
- "fit background to a polynomial", exponential, ...
- While this is convenient and the fit may be good, the narrative is weak

PHYSICAL REVIEW D **79,** 112002 (2009)



$$\frac{d\sigma}{dm_{jj}} = p_0(1-x)^{p_1}/x^{p_2+p_3\cdot\ln(x)}, \qquad x = m_{jj}/\sqrt{s},$$

The Effective Model Narrative

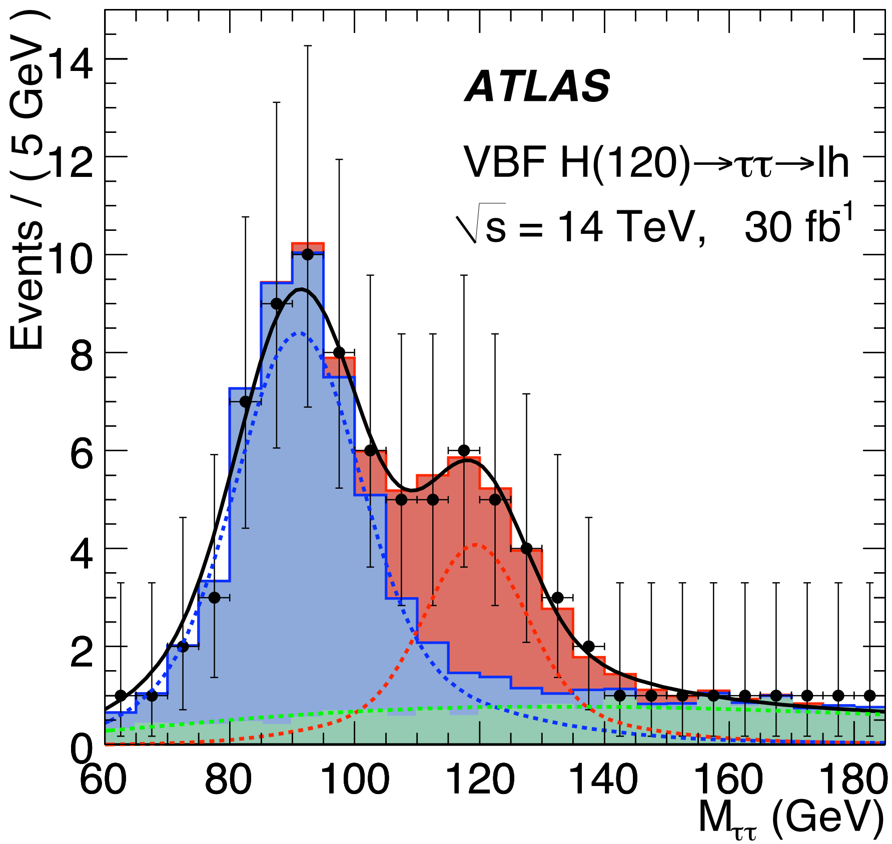


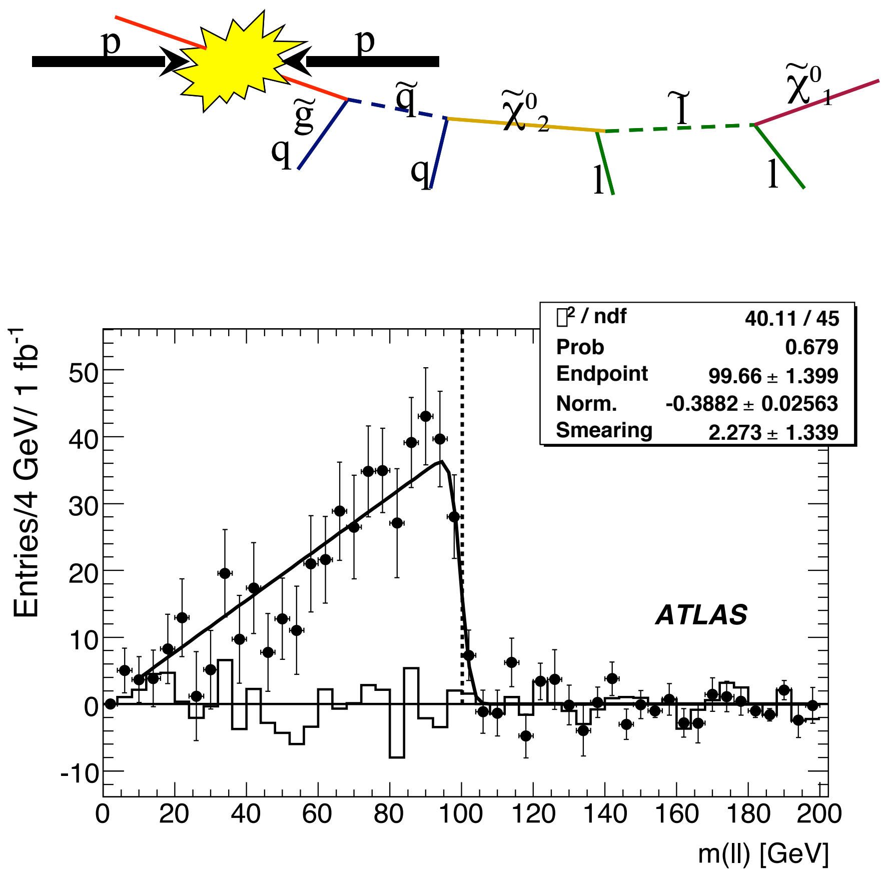
$$f(m_{ZZ}) = \frac{p0}{(1+e^{\frac{p6-m_{ZZ}}{p7}})(1+e^{\frac{m_{ZZ}-p8}{p9}})} + \frac{p1}{(1+e^{\frac{p2-m_{ZZ}}{p3}})(1+e^{\frac{p4-m_{ZZ}}{p5}})}$$

The Effective Model Narrative

Sometimes the effective model comes from a convincing narrative

- convolution of detector resolution with known distribution
  - Ex: MissingET resolution propagated through $M_{\tau\tau}$ in collinear approximation
  - Ex: lepton resolution convoluted with triangular M_{ll} distribution



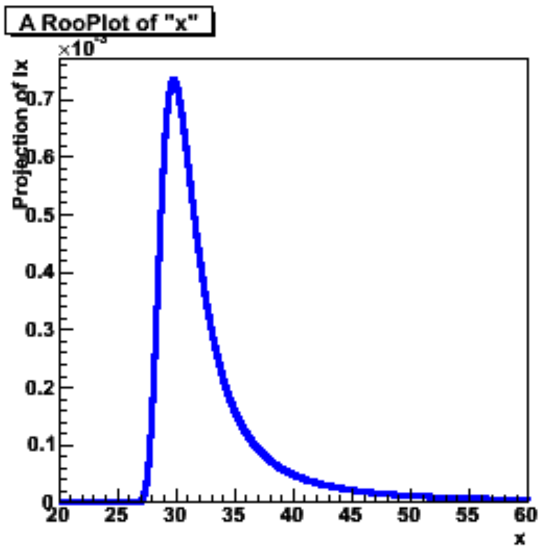


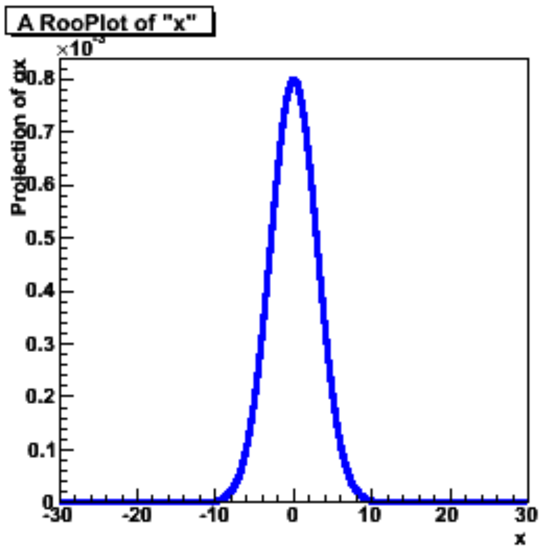
Tools for building effective models

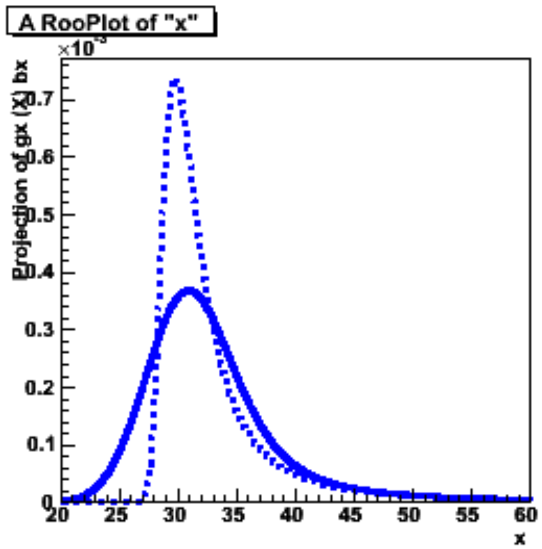
Center for
Cosmology and
Particle Physics

- RooFit's convolution PDFs can aid in building more effective models with a more convincing narrative

```
// Construct landau (x) gauss (10000 samplings 2nd order interpolation)
t.setBins(10000,"cache") ;
RooFFTConvPdf lxg("lxg","landau (X) gauss",t,landau,gauss,2) ;
```







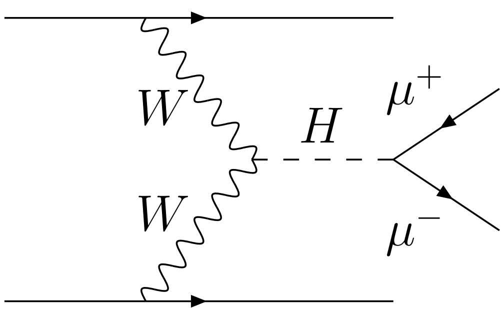
Wouter Verkerke, NIKHEF

The parametrized response narrative

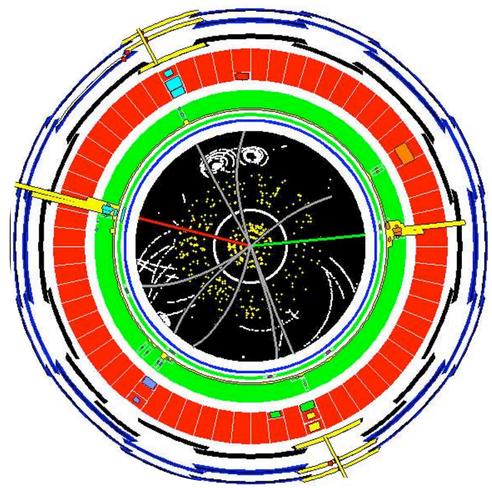
The Matrix-Element technique is conceptually similar to the simulation narrative, but the detector response is parametrized.

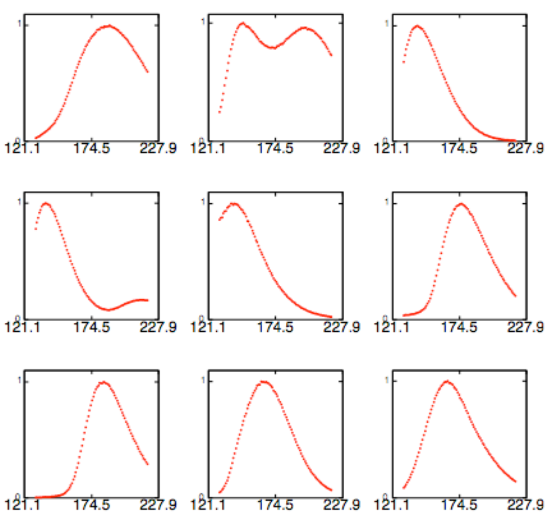
- Doesn't require building parametrized PDF by interpolating between non-parametric templates.

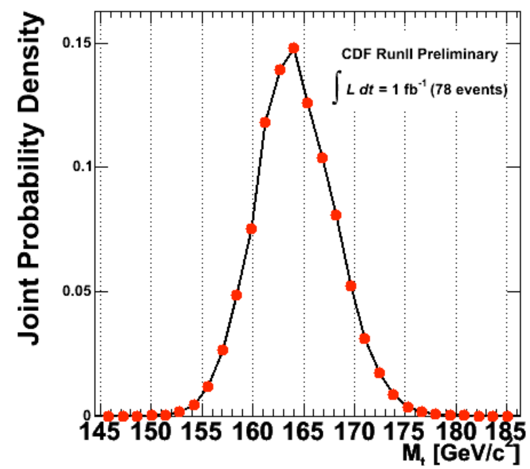
$L(x|H_0) =$



$\oplus$







Homework / Discussion Problem

Imagine a dataset with with observable *m* with range [0,10]

- ~100 events uniformly distributed in *m* $f_b(m) = 1/10$
- ~30 events Gaussian distributed with *μ= 5, σ=1* $f_s(m) = G(m|\mu,\sigma)$

Recall our "marked Poisson", where *s,b* are model parameters

$$P(\mathbf{m}|s) = \mathrm{Pois}(n|s+b) \prod_j^n \frac{s f_s(m_j) + b f_b(m_j)}{s+b}$$

- What will be the "best fit", maximum likelihood estimates of s,b?
- What is the role of the Poisson term here?

Now imagine that we have a theory that predicts s/b=0.5. How do things change?

Now imagine that we have a theory where *μ* is a free parameter, but the signal rate is given by *s(μ) = 10*μ*. How do things change?