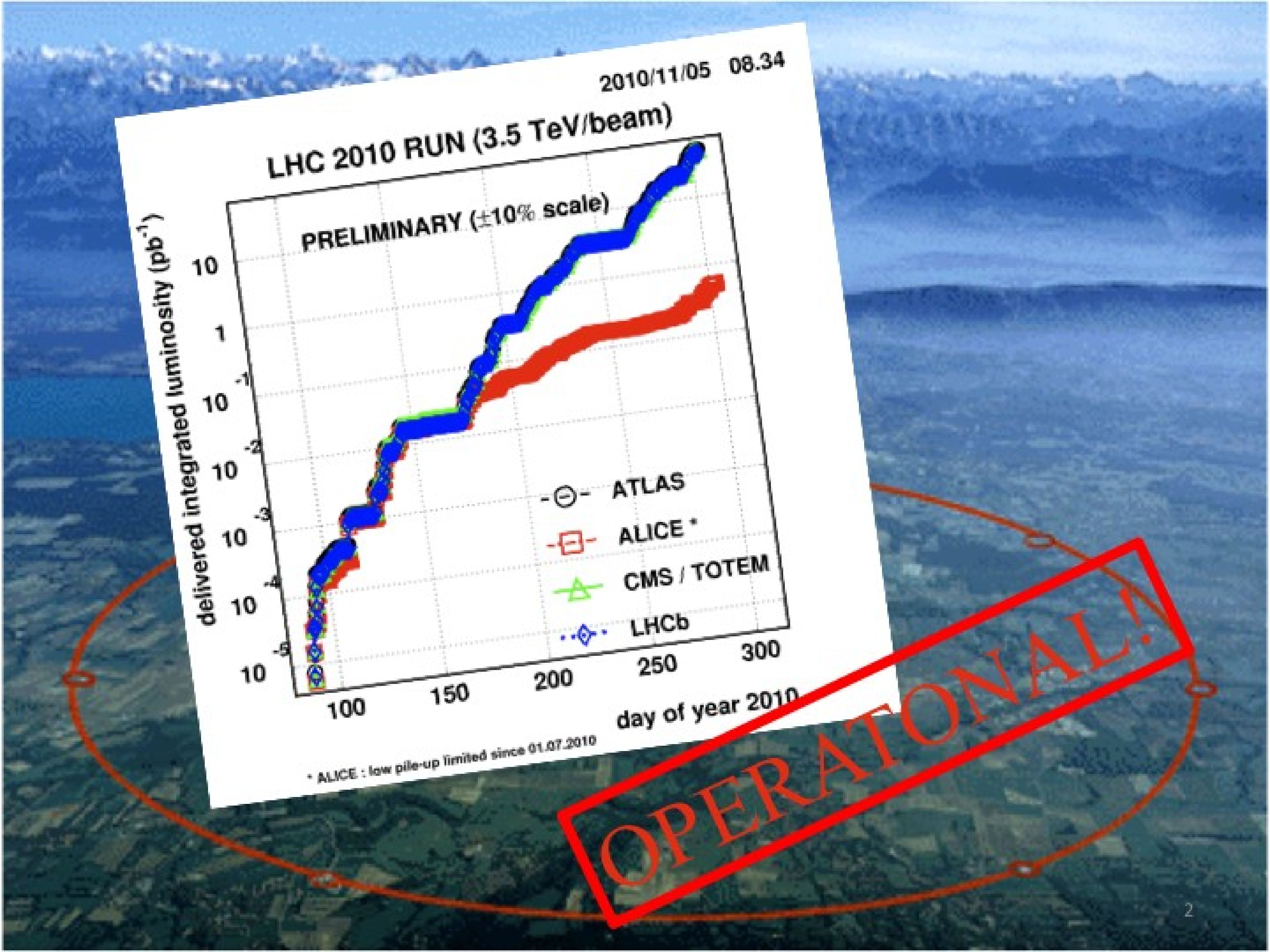

2010/11/05 08.34
LHC 2010 RUN (3.5 TeV/beam)
PRELIMINARY (±10% scale)
delivered integrated luminosity (pb^{-1})
10
1
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
100
150
200
250
300
day of year 2010
ATLAS
ALICE *
CMS / TOTEM
LHCb
* ALICE : low pile-up limited since 01.07.2010
OPERATONAL!


data's the best companion!
faithful, o, but stubborn.
need some kind of natural push
to get it going on and on..

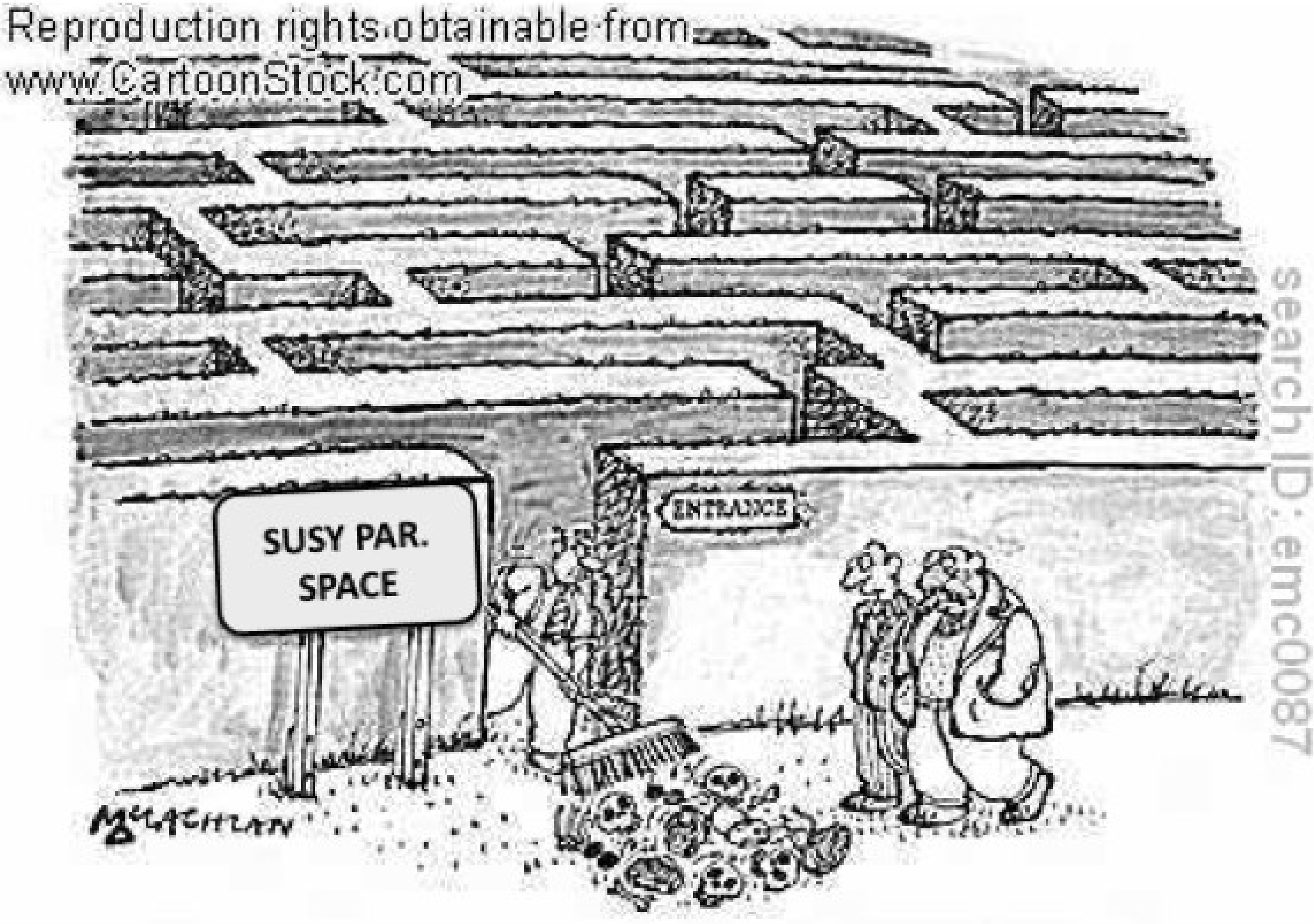

introducing "reference priors" to HEP

Searching for New Physics with Reference Priors

H. Prosper[1], M. Pierini[2], S. Sekmen[1], M. Spiropulu[2,3]
[1]FSU, [2]CERN, [3]Caltech

Global BSM fits and LHC data workshop, CERN, 10-11 Feb 2011

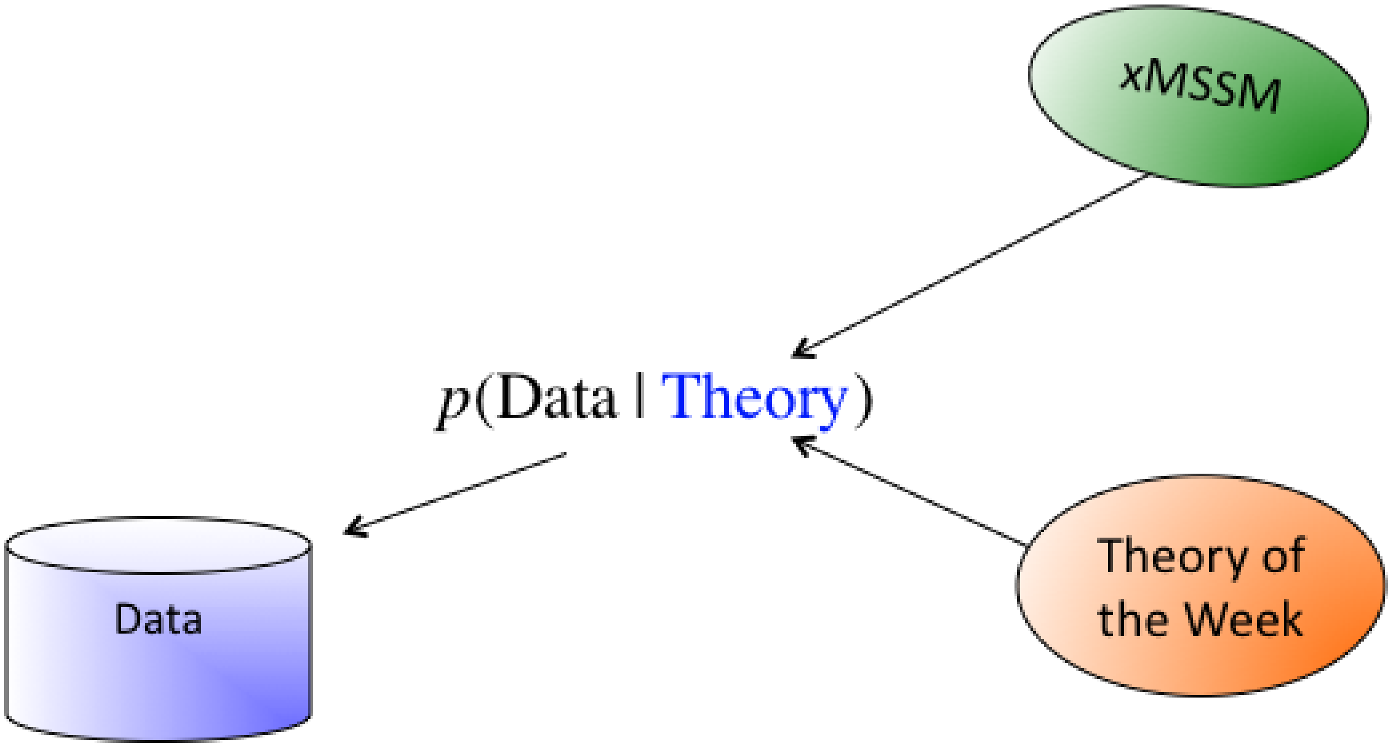
Scope of the problem:



'They're just finalizing the spring cleaning before the next collider season begins'

Expression of the problem

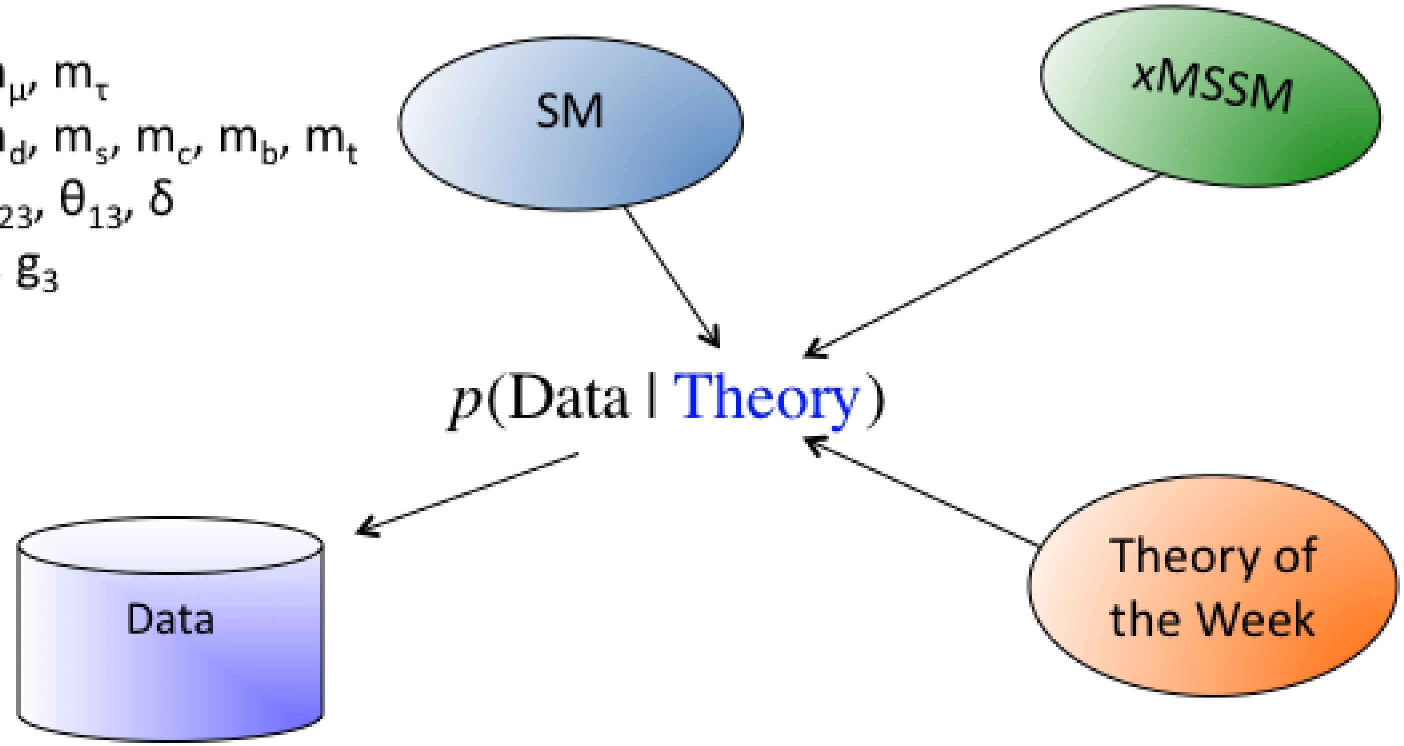
Basic problem: All interesting theories are multi-parameter models.

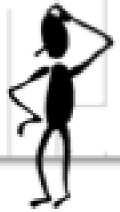


Expression of the problem

Basic problem: All interesting theories are multi-parameter models.

m_e, m_μ, m_τ
$m_u, m_d, m_s, m_c, m_b, m_t$
$\theta_{12}, \theta_{23}, \theta_{13}, \delta$
g_1, g_2, g_3
θ_{QCD}
μ, λ

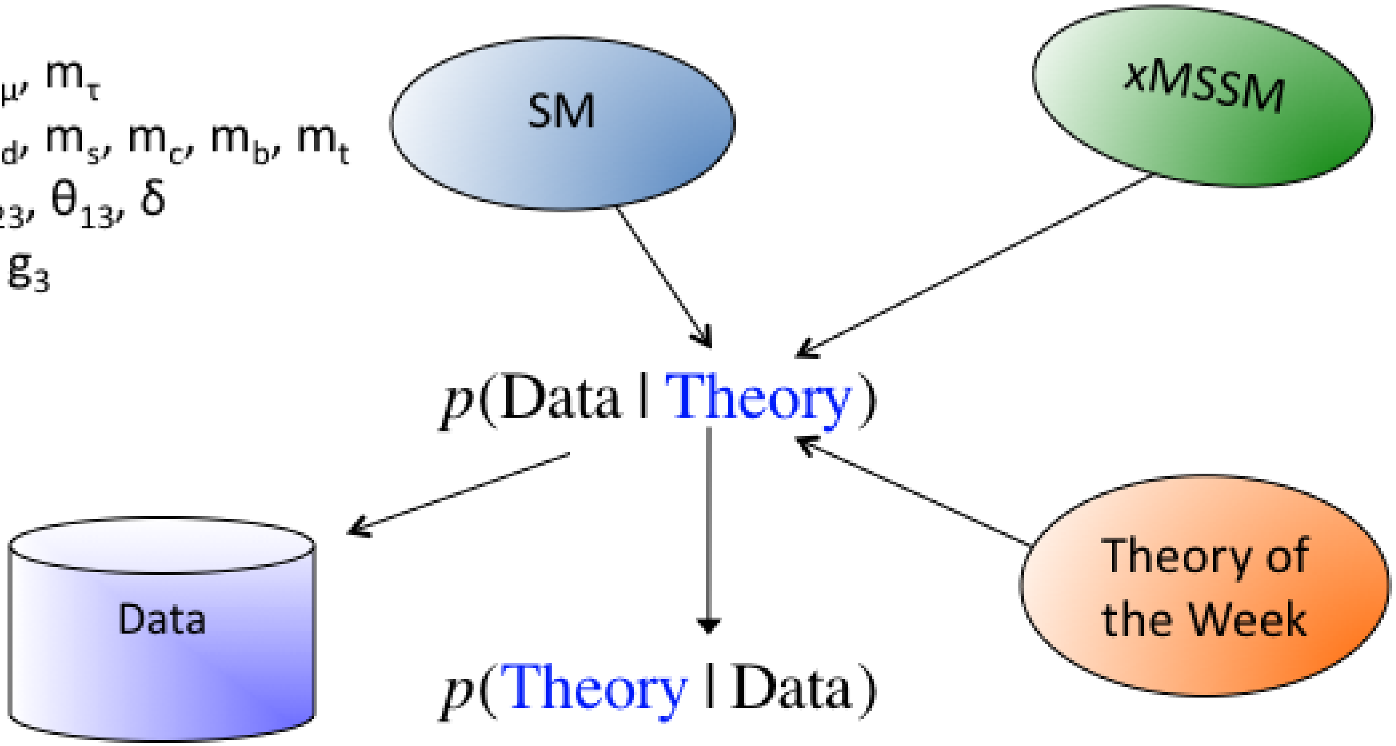




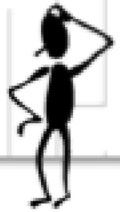
Expression of the problem

Basic problem: All interesting theories are multi-parameter models.

m_e, m_μ, m_τ
m_u, m_d, m_s, m_c, m_b, m_t
θ_{12}, θ_{23}, θ_{13}, δ
g_1, g_2, g_3
θ_{QCD}
μ, λ



Basic questions: Which theories and which parameter sub-spaces are preferred given the data?



The Bayesian reasoning

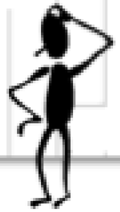
Given a model with parameters θ, and data x, Bayes' theorem is

$$p(\theta \mid x) \sim p(x \mid \theta)\, p(\theta)$$

Posterior probability of θ given x — Likelihood (contribution from data) — Prior knowledge on the model

Appealing features:

- Has strong theoretical foundations, is very general and conceptually straightforward
- Systematic learning from data through a recursive algorithm: posterior at a given stage becomes prior for the next.
- Coherent way to incorporate uncertainties regardless of their origin
- Given just the posterior, one can extract details such as point estimates, credible regions, etc.
- Can rank models according to their concordance with observation.



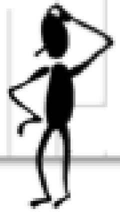
The question of priors - I

It is not possible to make progress without making some assumptions about the nature of the physics question:

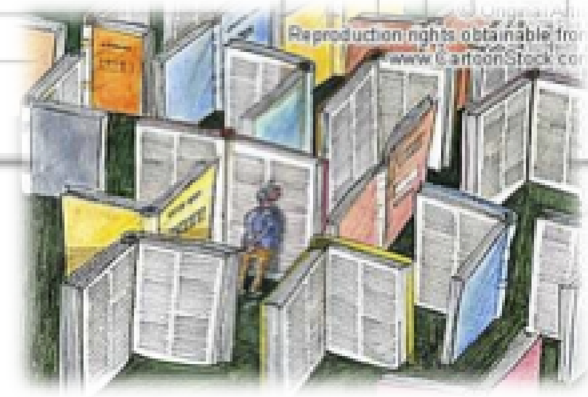
- ✓ To model backgrounds in a data-driven way, we assume that signal << background in the background region
- ✓ The LHC was designed assuming that BSM physics will be revealed at the TeV scale and will have high p_T signatures!

We use priors to incorporate our knowledge about a given model.

The question of "what prior to choose" arises when we lack intuition about the parameter space of the model. Defining suitable priors is a critical task!



The question of priors - II



Different priors will lead to different results.

The discrepancy among results obtained using different priors has been viewed, by some, as problematic. But this is a conceptual advantage that provides a way to assess whether the data are sufficient to make firm conclusions.

Current BSM studies generally adopt flat (or log) priors on the parameters. However:

- Suppose we make the transformation $\theta \to 1/\alpha$. The new prior becomes $\sim 1/\alpha^2$. Why choose the prior to be flat in θ rather than in α?
- Flat priors can be successfully used for single parameter models, but they can easily lead to pathological results in multi-parameter cases.

Therefore we need a formal way to construct priors.

Reference priors - I

In 1979, J. Bernardo introduced a formal rule to construct what he called reference priors. By construction, a reference prior contributes as little information as possible relative to the data.

A reference prior $\pi(\theta)$ maximizes the difference

$$D[\pi, p] \equiv \int p(\theta|x)\ \ln \frac{p(\theta|x)}{\pi(\theta)}\, d\theta$$

between the prior $\pi(\theta)$ and the posterior $p(\theta|x)$. D is called the Kullback-Leibler divergence. It is a measure of the information gained from the experiment.

But maximizing D is not quite right because it would yield a prior that depends on the observations n!

Reference priors - II

Reference analysis averages over all possible observations from K repetitions of the experiment:

$$I_K[\pi] \equiv \sum_{x_1=0}^{\infty} \cdots \sum_{x_K=0}^{\infty} m(x_{(K)})\, D[\pi, p(\theta|x_{(K)})],$$

in the limit K -> ∞, where

$$m(x_{(K)}) \;=\; \int p(x_{(K)}|\theta)\, \pi(\theta)\, d\theta,$$

$$\text{with } p(x_{(K)}|\theta) \;=\; \prod_{i=1}^{K} p(x_i|\theta),$$

is the marginal density for K experiments.

The reference prior is the $\pi(\theta)$ that maximizes $I_K[\pi]$, in the limit K -> ∞.

Reference priors - III

For the cases where the posterior densities are asymptotically normal, that is, become Gaussian as more data are included, the reference prior coincides with Jeffreys' prior:

$$\pi(\theta) = \sqrt{\mathbb{E}\left[-\frac{d^2 \ln p(x|\theta)}{d\theta^2}\right]}$$

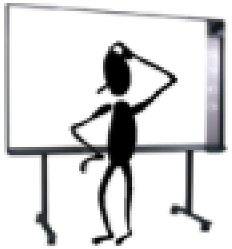
Therefore, constructing reference priors for single parameter scenarios is straightforward.

Direct generalizations to multi-parameter scenarios exist, but they are computationally demanding. Here we will propose a different way to approach the problem that is computationally tractable.

Usage of reference priors

Using the reference prior formalism, we can

- ✓ Rank new physics models according to their compatibility with observations independent of their dimensionality
- ✓ Estimate parameters of the new physics models
- ✓ Design an optimal analysis for a given model and given integrated luminosity
- ✓ …

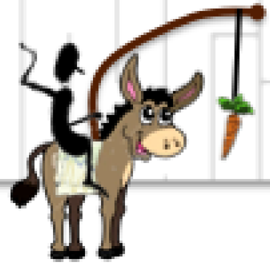


The plan

The Idea: Construct a proper posterior density for a simple experiment, starting with a reference prior, and map the posterior density into the parameter space of the model under investigation.

- We use the example of a single count experiment for which the signal and background model is well understood, and construct a reference prior $\pi(s)$ for the signal count s.
- Using $\pi(s)$, we obtain the posterior density $p(s|N)$, where N is the observed event count (background + signal).
- We use a "look-alike principle" to map the posterior density $p(s|N)$ to a prior $\pi(\theta)$ on the model parameter space.
- The prior $\pi(\theta)$ can now be used to continue the inference chain, recursively incorporating additional measurements x to get to the posterior $p(\theta|x)$.


so, let's get going!



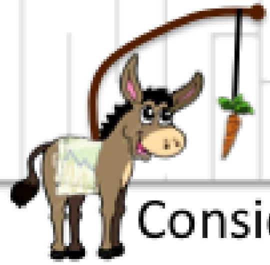
Simple mSUGRA example

- We illustrate our approach by investigating the mSUGRA scenario with
  - free parameters: $150 < m_0 < 600$ and $0 < m_{1/2} < 1500$
  - fixed parameters: $A_0 = 0$, $\tan\beta = 10$ and $\mu > 0$
- We use the CMS SUSY benchmark point LM1 with

$$m_0 = 60,\ \ m_{1/2} = 250,\ \ A_0 = 0,\ \ \tan\beta = 10,\ \ \mu > 0$$

as the “true state of nature”, which will provide the observed count N.

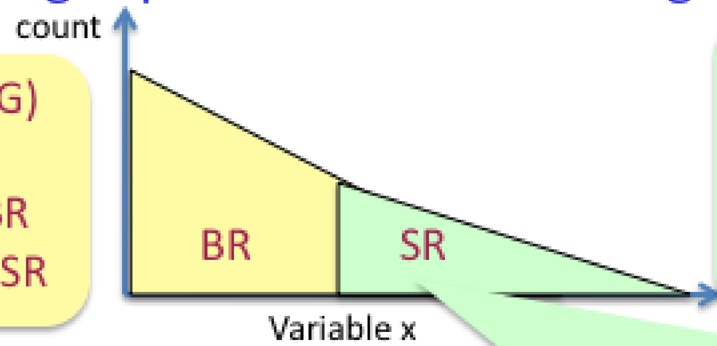
- For LM1 and for each point in a grid in the m_0-$m_{1/2}$ space, we generate 1000 7 TeV LHC events (PYTHIA) and simulate those with an approximate CMS detector response (modified PGS)
- We implement a multijets + missing ET selection and obtain the event yields for the LM1 and for the grid points. For background, we get the numbers from an existing CMS analysis.
- We quote results for 1pb^{-1}, 100pb^{-1} and 500pb^{-1}.



The single count model: Construction - I

Consider a counting experiment where the signal is due to new physics:

BR: The BG region (sig << BG)
Y: Observed count in BR
μb: Expected BG/count in BR
b: exp BG in BR / exp BG in SR



SR: The signal region
N: Observed count in SR
s: Expected signal in SR
μ: Expected BG in SR
n = s+μ: Expected count in SR

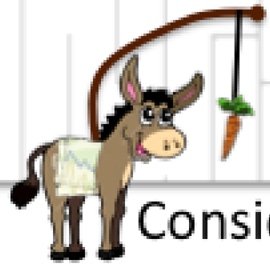
In SR, likelihood for observing N events is given by the Poisson distribution

$$p(N|\mu, s) = \frac{(\mu+s)^N}{N!}e^{-(\mu+s)}$$

To get the posterior

$$p(s|N) = p(N|s)\pi(s) = \int p(N|\mu, s)\pi(\mu, s)d\mu$$

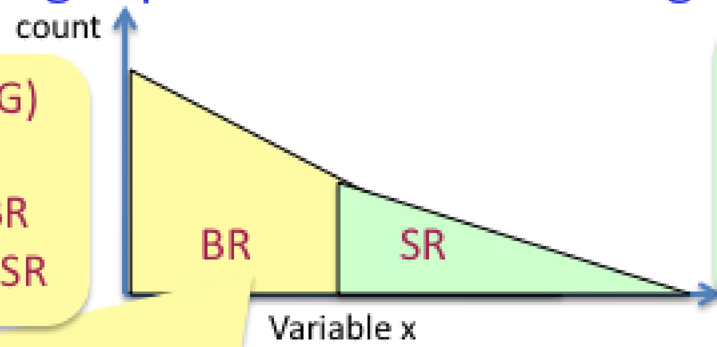
we need the prior $\pi(\mu, s)$ which we factorize as: $\pi(\mu, s) = \pi(\mu|s)\,\pi(s)$
We further assume that $\pi(\mu|s) = \pi(\mu)$, the prior on μ is independent on s.



The single count model: Construction - II

Consider a counting experiment where the signal is due to new physics:

BR: The BG region (sig << BG)
Y: Observed count in BR
μb: Expected BG/count in BR
b: exp BG in BR / exp BG in SR



SR: The signal region
N: Observed count in SR
s: Expected signal in SR
μ: Expected BG in SR
$n = s+\mu$: Expected count in SR

In BR, likelihood for observing Y events is given by the Poisson distribution

$$p(Y|b,\mu) = p(Y|\mu) = \frac{(b\mu)^Y}{Y!} e^{-(b\mu)}$$

b is a known constant

To get the posterior $p(\mu|Y) = \pi(\mu) = p(Y|\mu)\pi_0(\mu)$
we need the prior $\pi_0(\mu)$, "the initial prior". We get this by calculating the reference prior (Jeffrey's prior) using the likelihood $p(Y|\mu)$.

This gives $\pi_0(\mu) \sim 1/\sqrt{\mu}$. From $\pi_0(\mu)$ and $p(Y|\mu)$ we obtain

$$p(\mu|Y) = \pi(\mu) = \frac{b(b\mu)^{Y-1/2}}{\Gamma(Y+1/2)} e^{-b\mu}$$

The single count model: Likelihood

We marginalize p(N | μ, s) over μ to get the likelihood:

$$
\begin{aligned}
p(N \mid s) &= \int p(N \mid \mu, s)\, \pi(\mu)\, d\mu, \\
&= \int \frac{(\mu + s)^N}{N!}\, e^{-\mu - s}\, \frac{b(b\mu)^{y-1/2}}{\Gamma(y + 1/2)}\, e^{-b\mu}\, d\mu, \\
&= e^{-s} \left[\frac{b}{b+1}\right]^{y+\frac{1}{2}} \sum_{k=0}^{N} v_{Nk}\, \frac{s^k}{k!},
\end{aligned}
$$

$$
\text{where} \quad v_{Nk} \equiv \frac{\Gamma(y + \frac{1}{2} + N - k)}{\Gamma(y + \frac{1}{2})\,(N-k)!} \left[\frac{1}{b+1}\right]^{N-k}.
$$

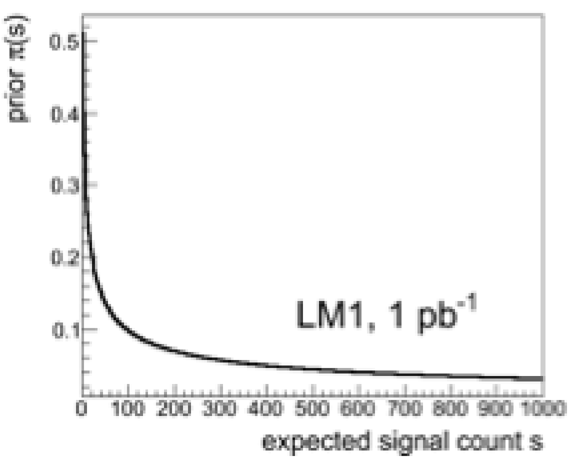
Having reduced the likelihood to a single parameter, we can use the 1-parameter algorithm to construct the reference prior π(s) (Jeffreys' prior) for this likelihood.

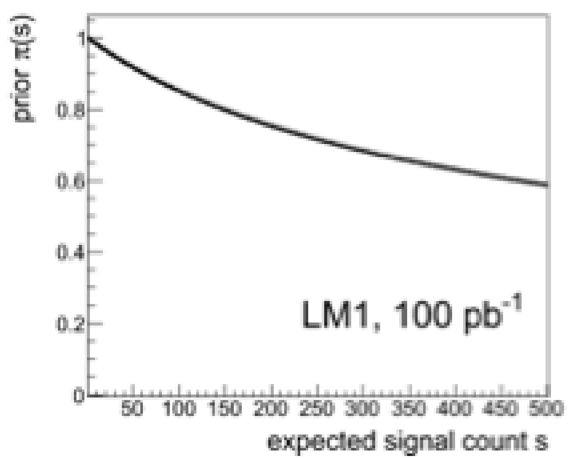
The single count model: The prior

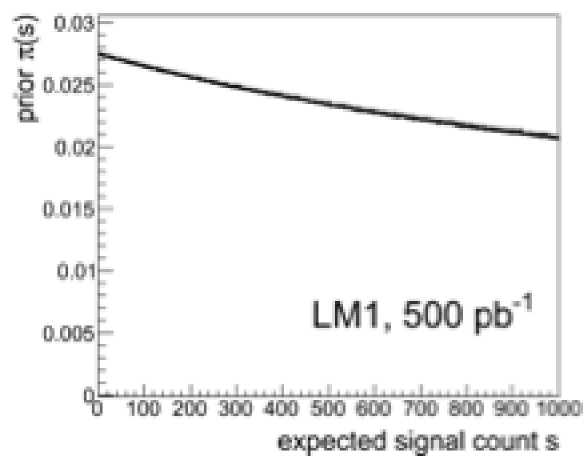
Reference prior on s calculated from likelihood p(N|s) :

$$\pi(s) \propto \sqrt{e^{-s} \sum_{n=0}^{\infty} \frac{[T_n^0 - T_n^1/s]^2}{T_n^0}},$$

$$\text{where} \quad T_n^m(s) \equiv \sum_{k=0}^{n} k^m \, v_{nk} \, \frac{s^k}{k!} \quad \text{for} \quad m = 0, 1.$$

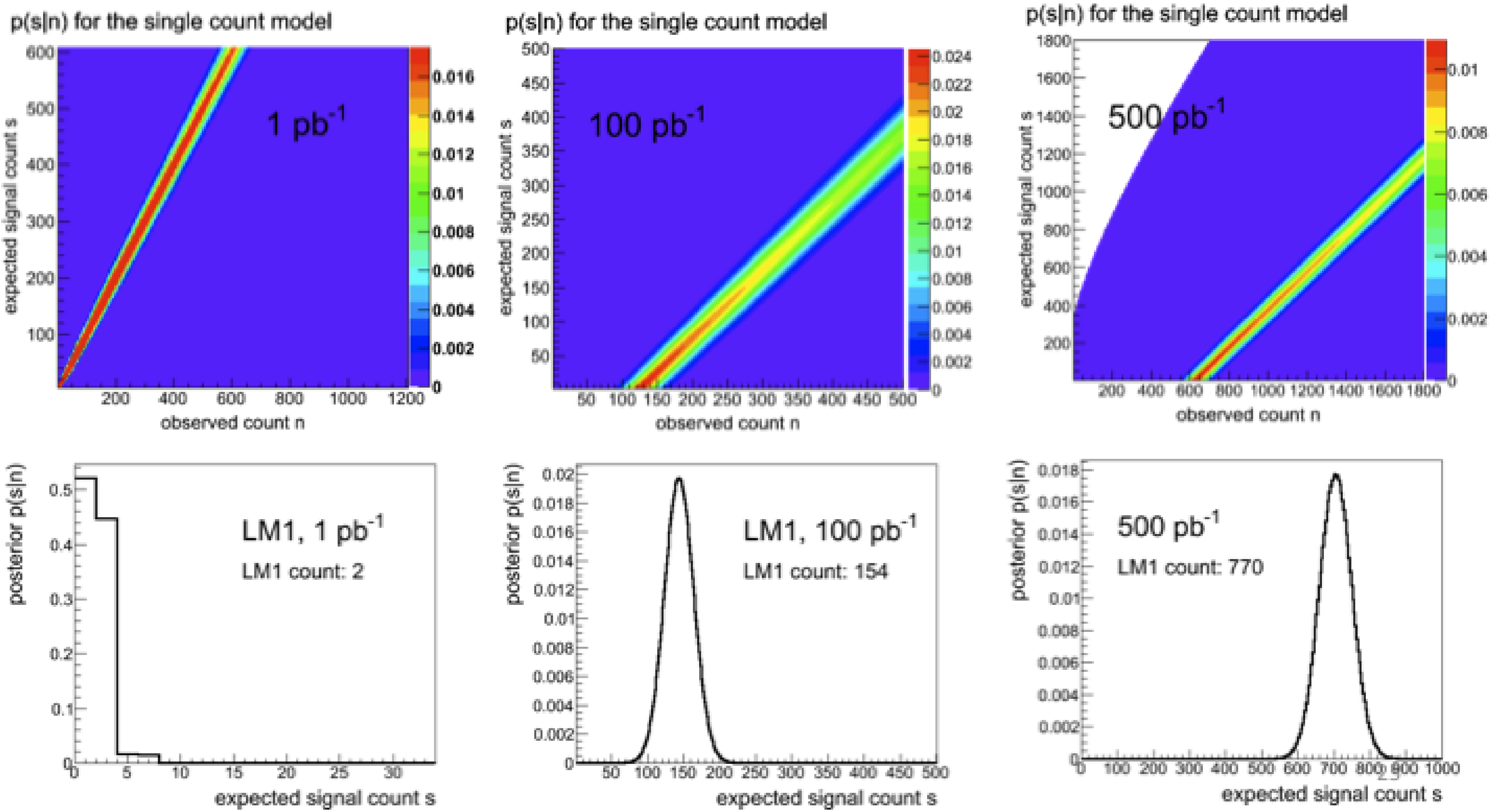


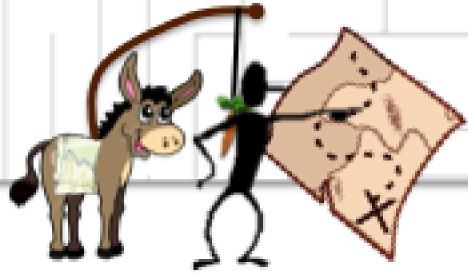




The single count model: The posterior

$$p(s|n) = p(n|s)\,\pi(s) / \int_0^\infty p(n|s)\,\pi(s)\,ds.$$





Mapping to multi-dimensional SUSY space - I

p(s|N) is a proper density based on a reference prior, and hence is invariant under one-to-one transformations of s.

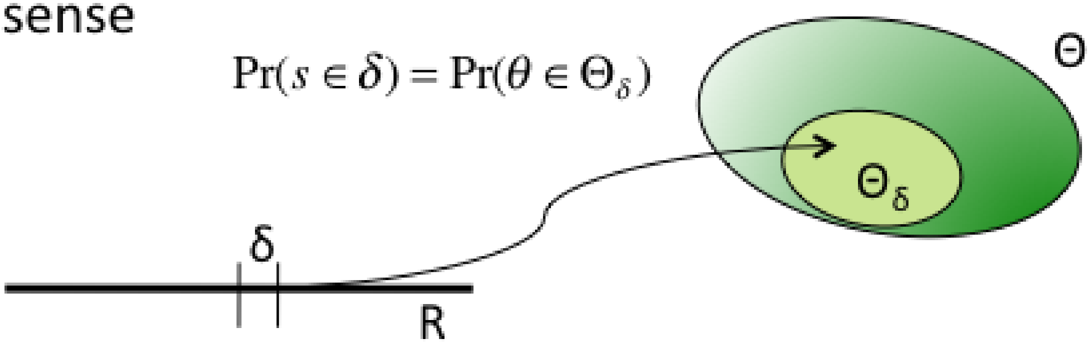
Model parameters θ are related to signal count s as s = f(θ).

We would like to find the reference prior π(θ) induced on the model parameter space by p(s|N).

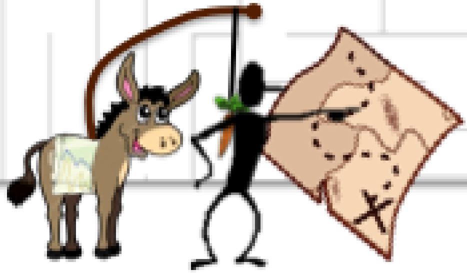
To find π(θ), we make use of a generic probability statement in two parts:

1st part - Mapping to regions: p(s|N) and π(θ) should be consistent in the following sense

$$\Pr(s \in \delta) = \Pr(\theta \in \Theta_\delta)$$



$$p(s|N) = \int_\Theta \delta(s - f(\theta))\,\pi(\theta)\,d\theta, = \int_{\mathbb{S}} \frac{\pi(\theta)}{|\nabla f|}\,d\sigma(\theta)$$



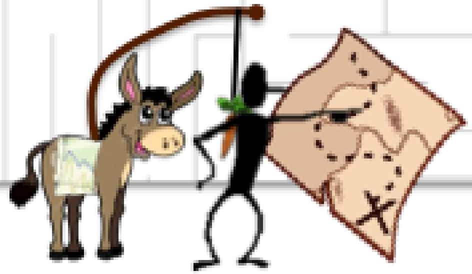
Mapping to multi-dimensional SUSY space - II

2nd part – Mapping to points: The expected signal s is the same for all points in Θ_δ. Therefore, in that sense, the points in Θ_δ are indistinguishable.

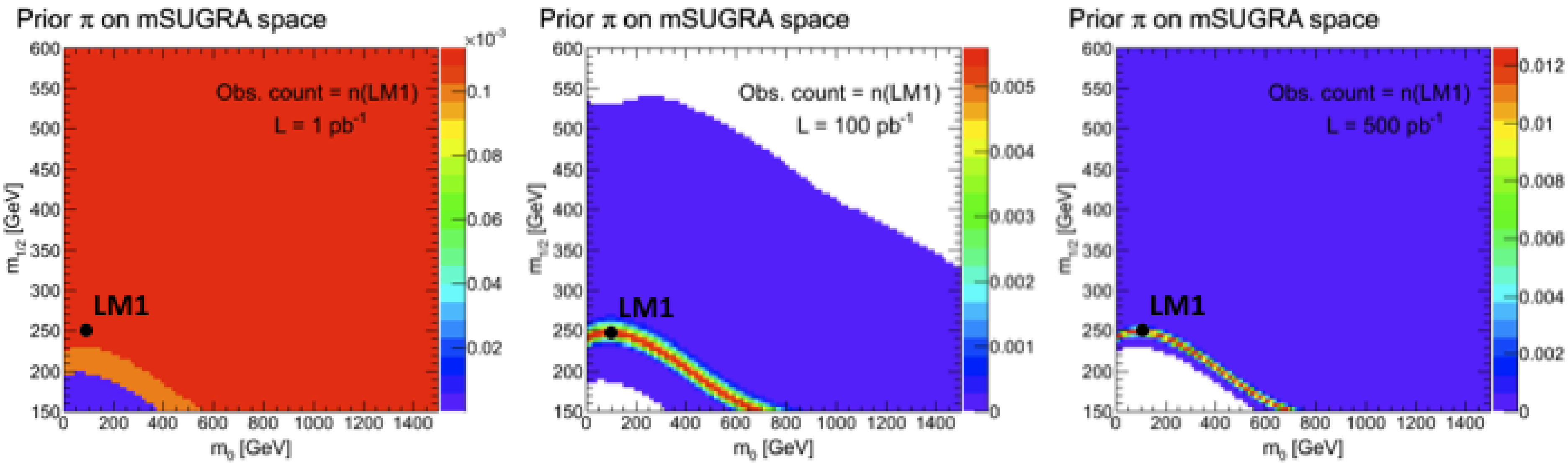
We propose, therefore, assigning the same probability density to every point in Θ_δ.

$$\pi(\theta) = p(s|N) / \underbrace{\int_{\mathbb{S}} \frac{d\sigma(\theta)}{|\nabla f|}}_{\text{surface term}} \qquad s - f(\theta) = 0.$$

For simplicity in this study the surface term was neglected, because we expect it to be a much gentler function compared to p(s|N).



Reference prior π(θ) on the mSUGRA space



The new Bayesian procedure is consistent in that the posterior/prior converge to the correct subspace of the parameter space.

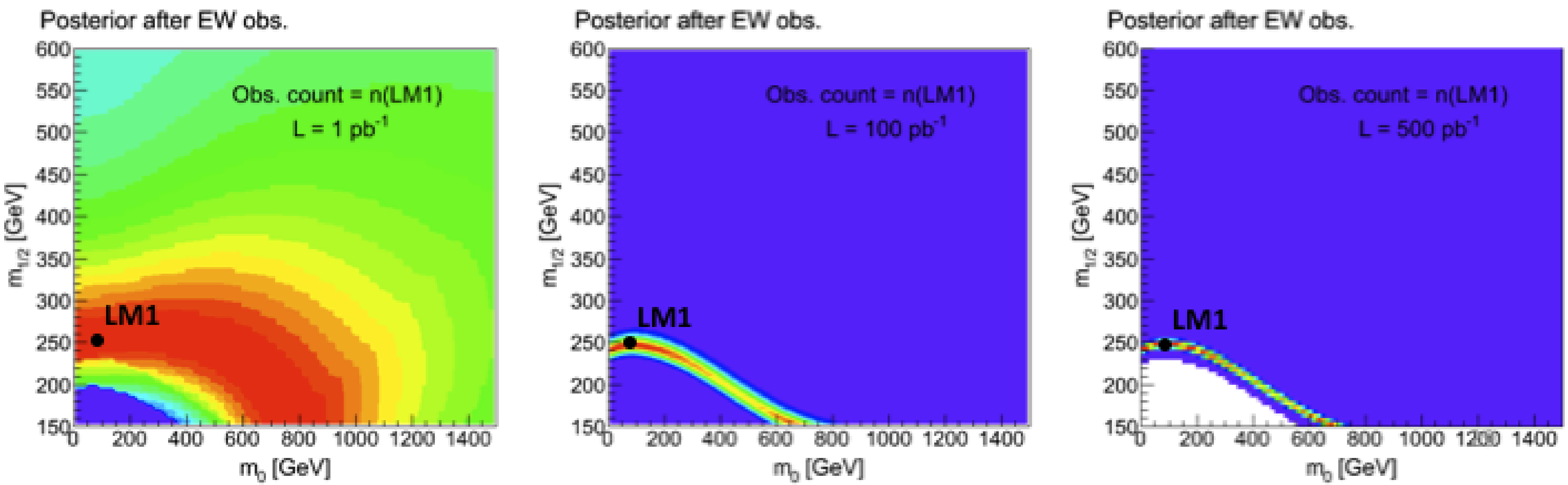
Adding the EW/flavor observables

We continue the inference chain by incorporating the likelihood

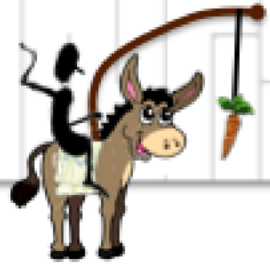
$$\mathcal{L}(\vec{\alpha}|m_0, m_{1/2}) \propto \prod_i e^{-\frac{(\alpha_i(m_0,m_{1/2})-m_i)^2}{2\sigma_i^2}}$$

for a set of EW/flavor observables I, that are BR(b -> sγ), R(BR(b -> τν)), BR(b -> Dτν), BR(b -> Dτν)/BR(b -> eτν), R_{l23}, BR(D_s -> τν), BR(D_s -> μν) and Δρ.

Since the nature is LM1, we used the LM1 values for the observables along with the measured uncertainties.




First attempt with higher dimensionality



5D SUGRA example

We use two simple extensions of mSUGRA, each with 5 free parameters:

- Model 1: Non-universal m0(1,2), with parameterization:

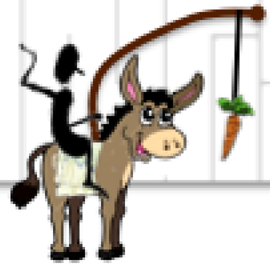
  $m_0 = m_0(3) = m_{Hu,d}$, $m_0(1,2)$, $m_{1/2}$, A_0, $\tan\beta$, $\mu > 0$

Model 2: Non-universal M3, with parameterization:

  m_0, $m_{1/2}$, M3, A_0, $\tan\beta$, $\mu > 0$

The "true state of nature (TSN)" is chosen from Model 1, and is defined as

  $m_0 = 1000$, $m_0(1,2) = 60$, $A_0 = 0$, $\tan\beta = 10$



5D SUGRA example - method

- The expected signal s is given by

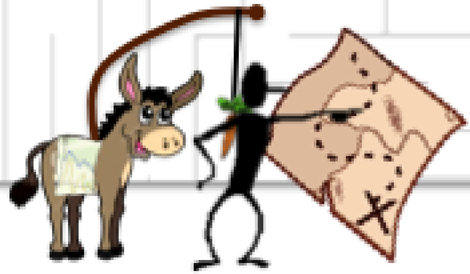
  s = cross section × efficiency × integrated luminosity.

- The "observed" count N is obtained as the sum of the expected signal for the true state of nature and the expected background obtained from the CMS analysis.

- We generate a sample of points θ from the posterior $p(s|N)$, given that $s = f(\theta)$, using a Markov Chain Monte Carlo method.

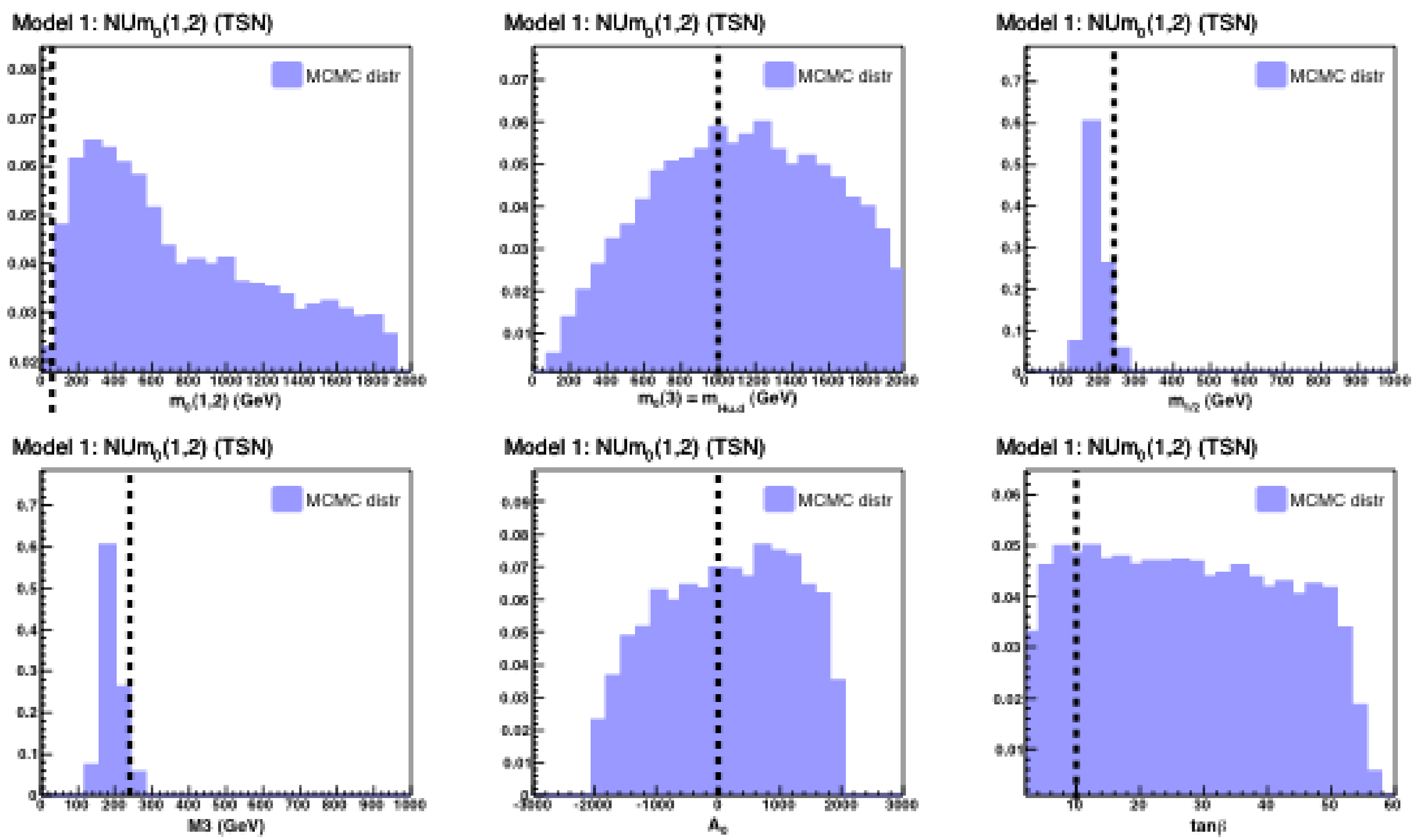
- In general, the efficiency is a function of the parameter space of the model. It is obtained by performing the analysis on a given point θ. This requires simulating a sufficient number of events for θ.

- Since computing this efficiency with official tools is time consuming, we have explored the possibility of using a constant efficiency and reweighting the points afterwards with $p(s|N)$ / $p(s_{const})|N)$.

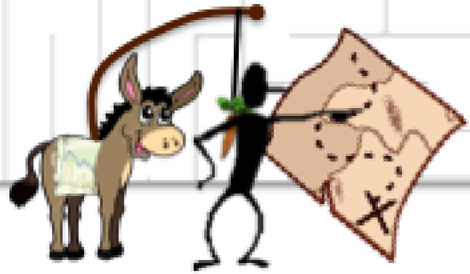


Mapping to the 5D SUGRA space - I

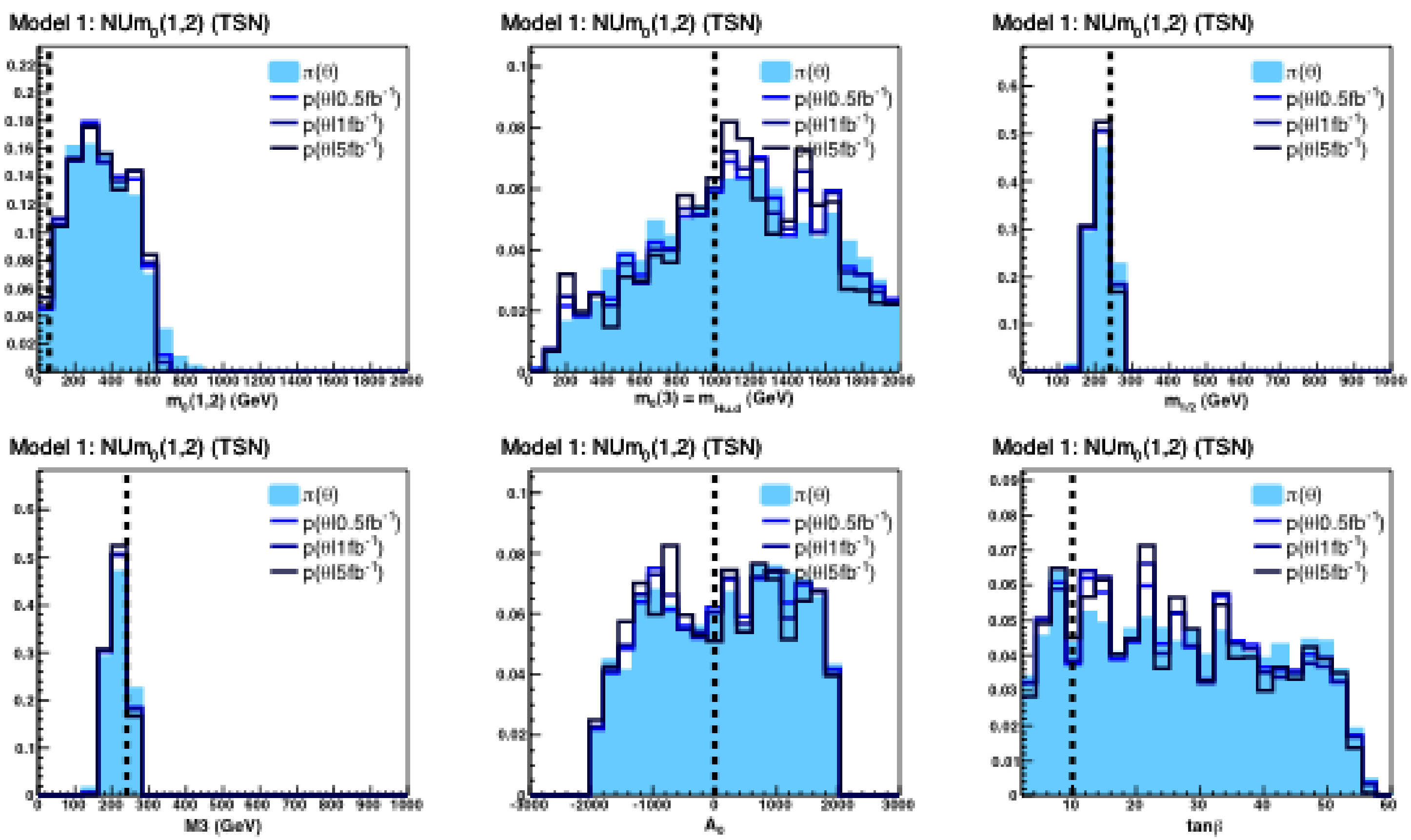
Plots show the distributions of the points sampled by the MCMC, before the corrective weights are applied.



Mapping to the 5D SUGRA space - II

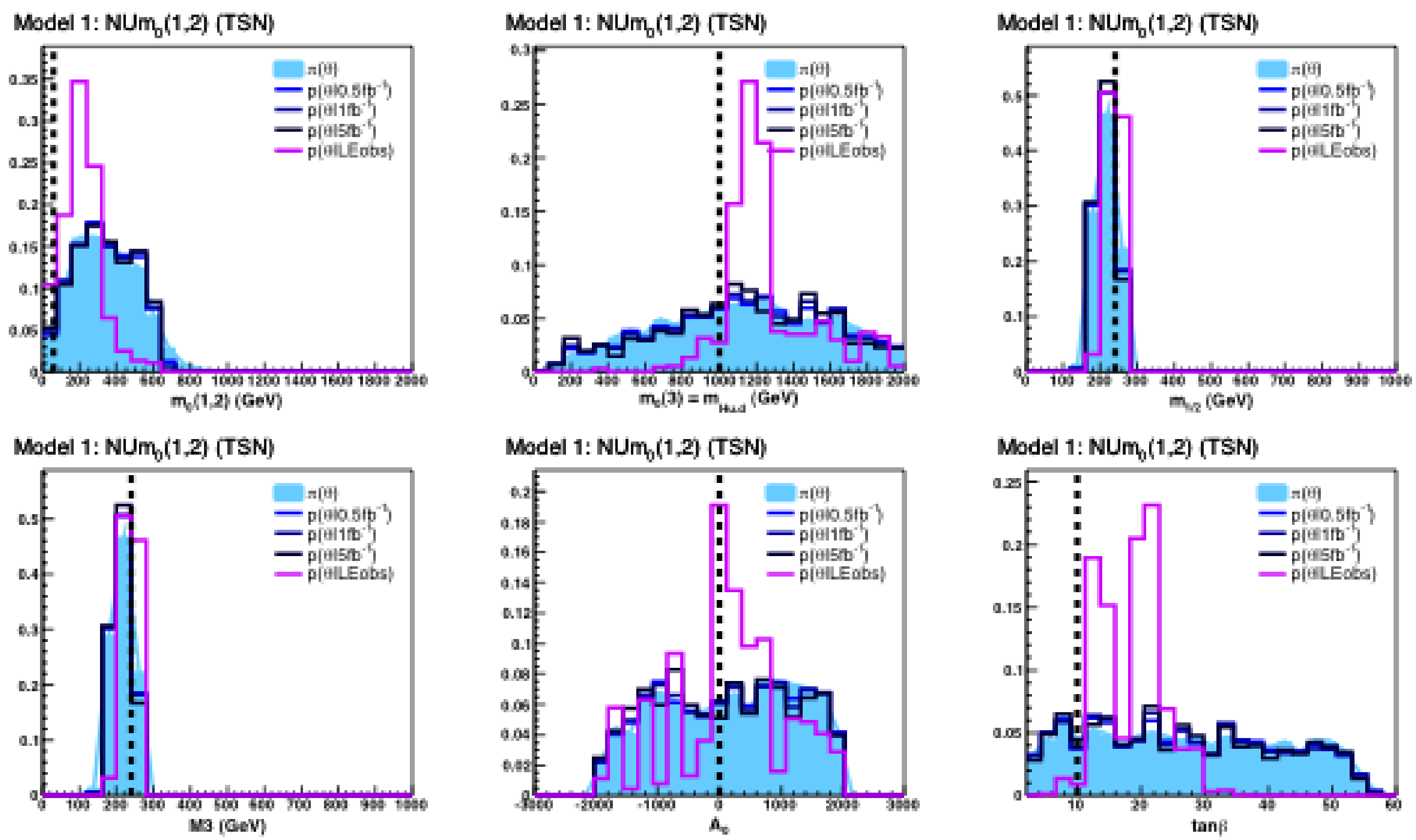


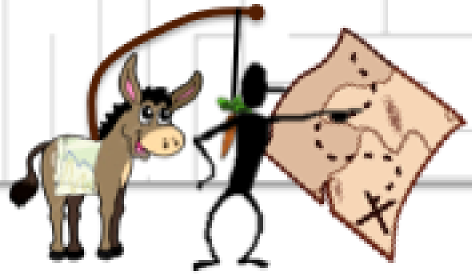
Distribution after weighting, which gives the prior $\pi(\theta)$, which is calculated using 100pb^{-1} data. Then we recursively add information from more data through multiplying with the likelihood of data given the signal count.



Adding the EW/flavor observables

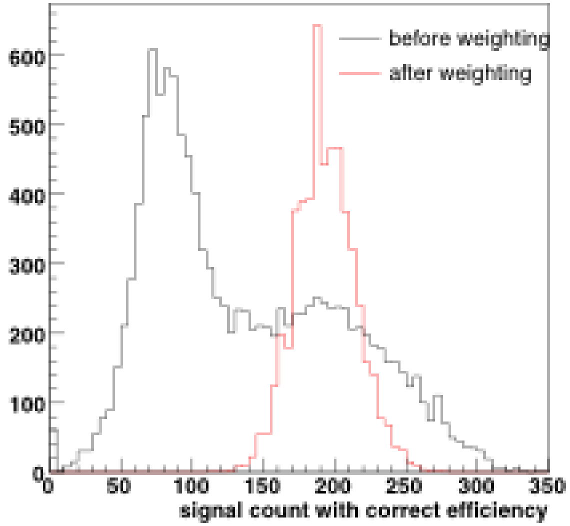
Distributions after adding the input from EW/flavor observables obtained through multiplying by the likelihood of EW/flavor energy data.

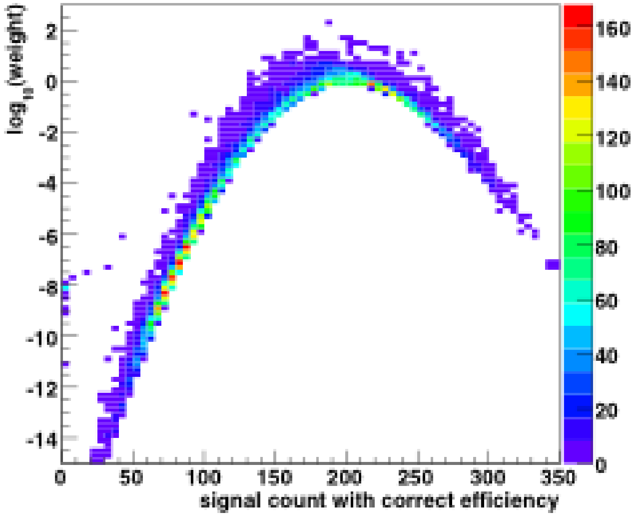




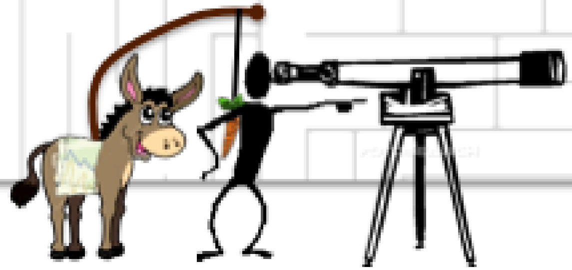
Diagnosis

The relevant subspace is not being sampled efficiently because of the use of a constant efficiency.





To do the MCMC more efficiently, we need to sample from the p(N|s) calculated using the correct efficiency.
This means we have to calculate the efficiencies during the MCMC.
To be able to do this, we need to have **VERY FAST AND ACCURATE SIMULATION TOOLS!**



Summary and outlook

- We proposed a way to construct multi-dimensional priors from the posterior density for a simple experiment. The key idea is to start with a reference prior, and map the posterior density into the parameter space of the model under investigation.
- It is necessary to use the correct efficiencies to ensure efficient sampling of the parameter space. This requires the use of fast and accurate event simulators.
- The single count model we used for building the reference prior can be replaced by any for which the signal and background modeling is well-understood.
- Reference analysis provides a procedure for ranking models (i.e., hypothesis testing), parameter estimation, etc.
- We need to find observables that will break the degeneracy in the look-alike regions.