
STATUS:
OPERATIONAL


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

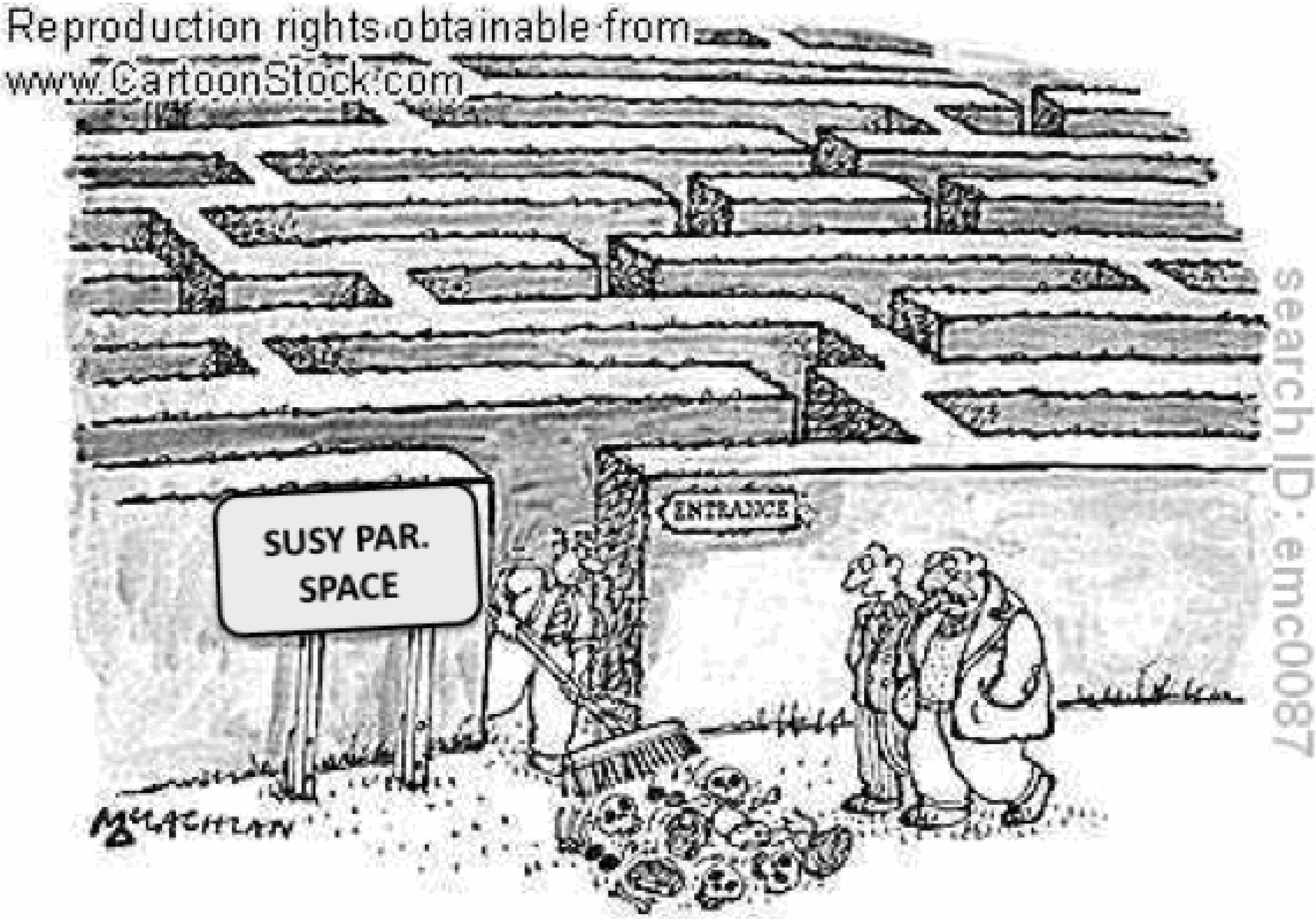

introducing "reference priors" to HEP

Exploring the SUSY parameter space with a new Bayesian approach

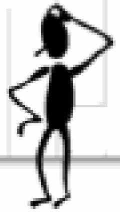
H. Prosper[1], M. Pierini[2], S. Sekmen[1], M. Spiropulu[2,3] : [1]FSU, [2]CERN, [3]Caltech

SUSY10 Bonn, 23-28 August 2010

Scope of the problem:



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



The Bayesian reasoning

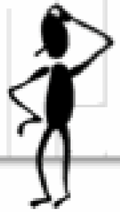
Given a model with parameters θ, and data x, Bayes' theorem is

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

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

Appealing features:

- 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 issue with priors

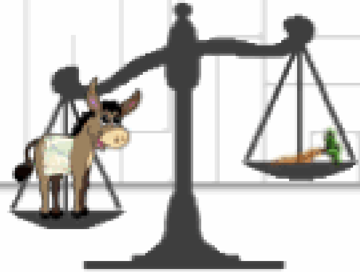
Defining suitable priors is a critical task!

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 SUSY studies generally adopt flat priors (that assign the same probability to every point) 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 lead to pathological results.

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 x!

Reference priors - II

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

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

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 π that maximizes $I_K[\pi]$, in the limit K -> ∞.

Reference priors - III

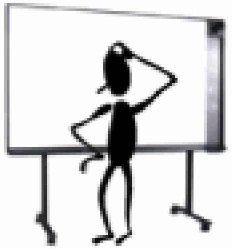
For the cases where the posterior densities are asymptotically normal, that is, become Gaussian as more data are included, the reference prior coincides with Jeffreys' prior:

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

Likelihood

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

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

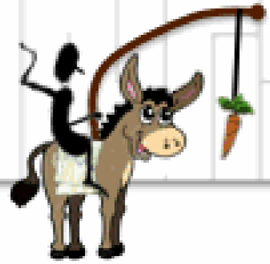


The plan

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

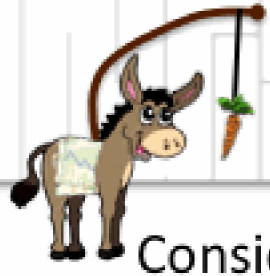
- We illustrate our approach by investigating the mSUGRA scenario with
  - free parameters: $150 < m_0 < 600$ and $0 < m_{1/2} < 1500$
  - fixed parameters: $A_0 = 0$, $\tan\beta = 10$ and $\mu > 0$
- We use the CMS SUSY benchmark point LM1 with

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

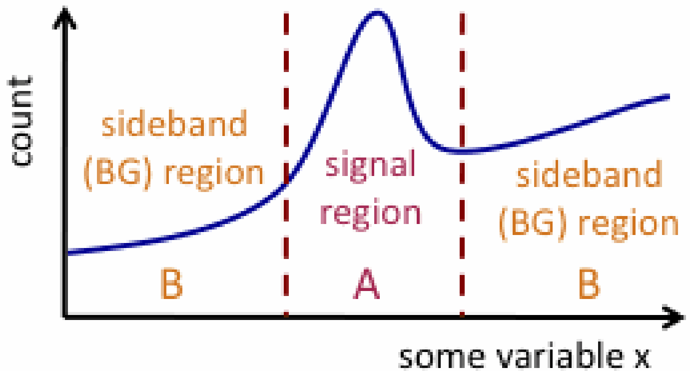
as the "true state of nature", which will provide the observed count n.

- For LM1 and for each point in a grid in the m_0-$m_{1/2}$ space, we generate 1000 7 TeV LHC events (using PYTHIA and 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



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



s: Expected signal in A

μ: Expected background in A

n: Observed count in A

bμ: Expected background in B

b: Expected BG for B / expected BG for A

y: Observed count in B

The likelihood for n events is given by a Poisson distribution

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

and we factorize the associated prior $\pi(s, \mu)$, as $\pi(s, \mu) = \pi(\mu|s)\, \pi(s)$.

We assume that $\pi(\mu|s) = \pi(\mu)$ (BG is independent of the signal), and model $\pi(\mu)$ as

$$\pi(\mu) \; = \; \frac{b(b\mu)^{y-1/2}}{\Gamma(y+1/2)}\, e^{-b\mu},$$

The single count model: Likelihood

We marginalize the likelihood p(n|s,μ) (integrate it) over μ:

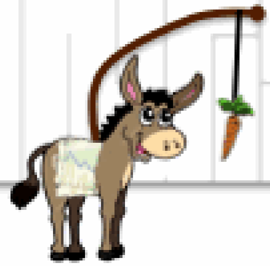
$$p(n\,|\,s) = \int p(n\,|\,s,\mu)\,\pi(\mu)\,d\mu,$$

$$= \int \frac{(s+\mu)^n}{n!}\,e^{-s-\mu}\,\frac{b(b\mu)^{y-1/2}}{\Gamma(y+1/2)}\,e^{-b\mu}\,d\mu,$$

$$= e^{-s}\left[\frac{b}{b+1}\right]^{y+\frac{1}{2}} \sum_{k=0}^{n} v_{nk}\,\frac{s^k}{k!},$$

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

Likelihood is reduced to a single parameter s.

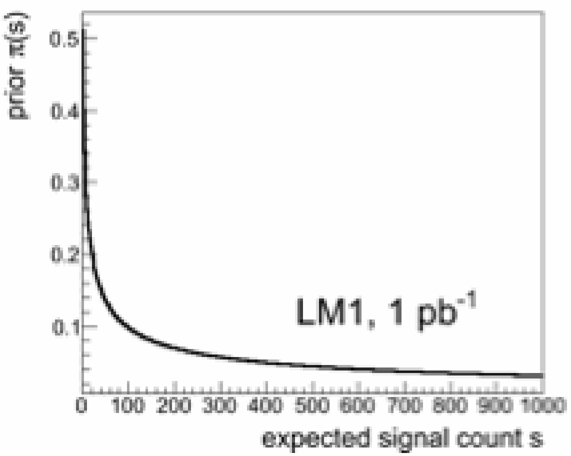


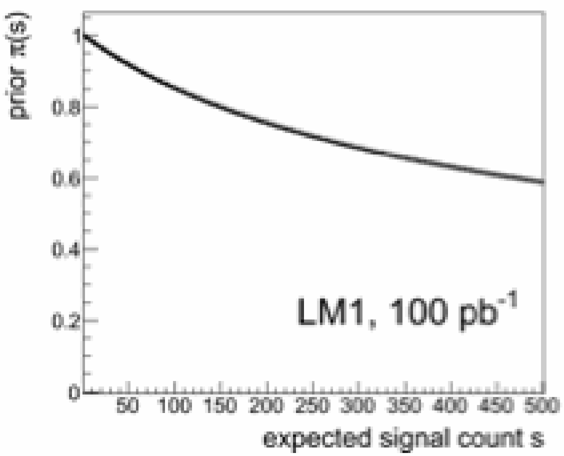
The single count model: The prior

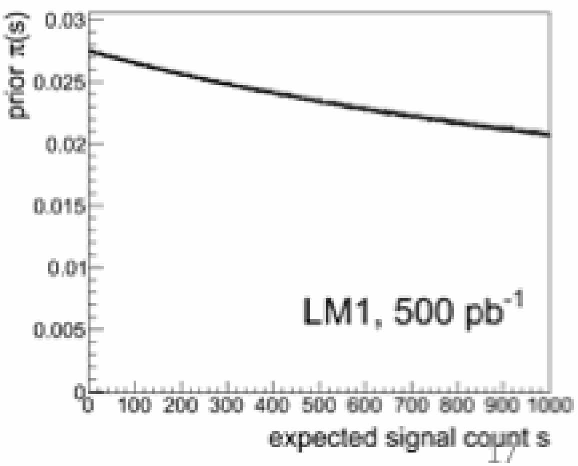
Now we can use the 1-parameter reference algorithm to construct the reference prior π(s) (Jeffreys' prior) for the likelihood p(n|s):

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

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

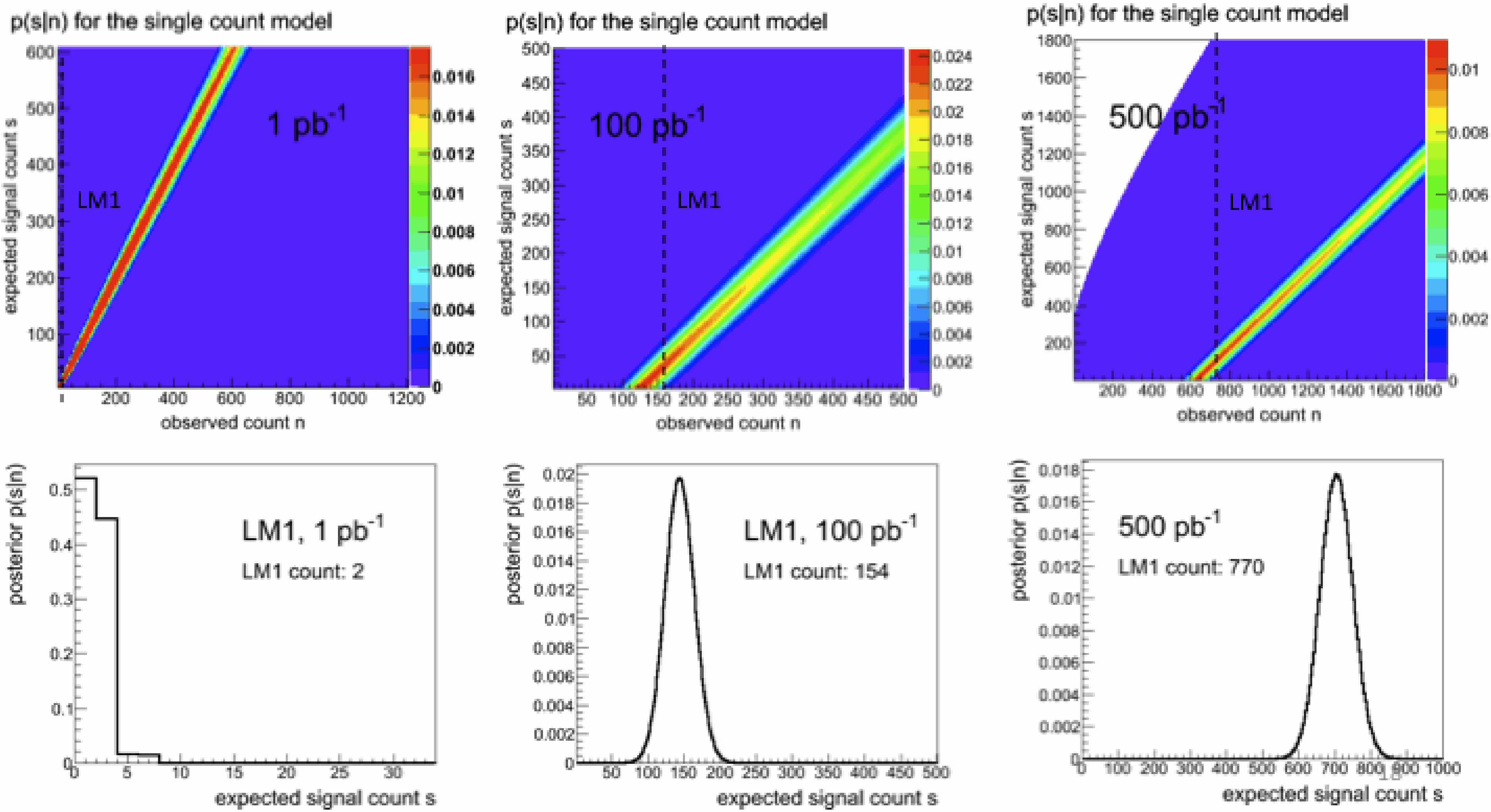


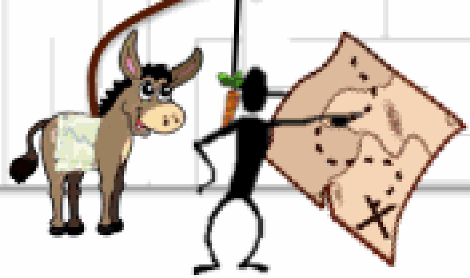




The single count model: The posterior

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



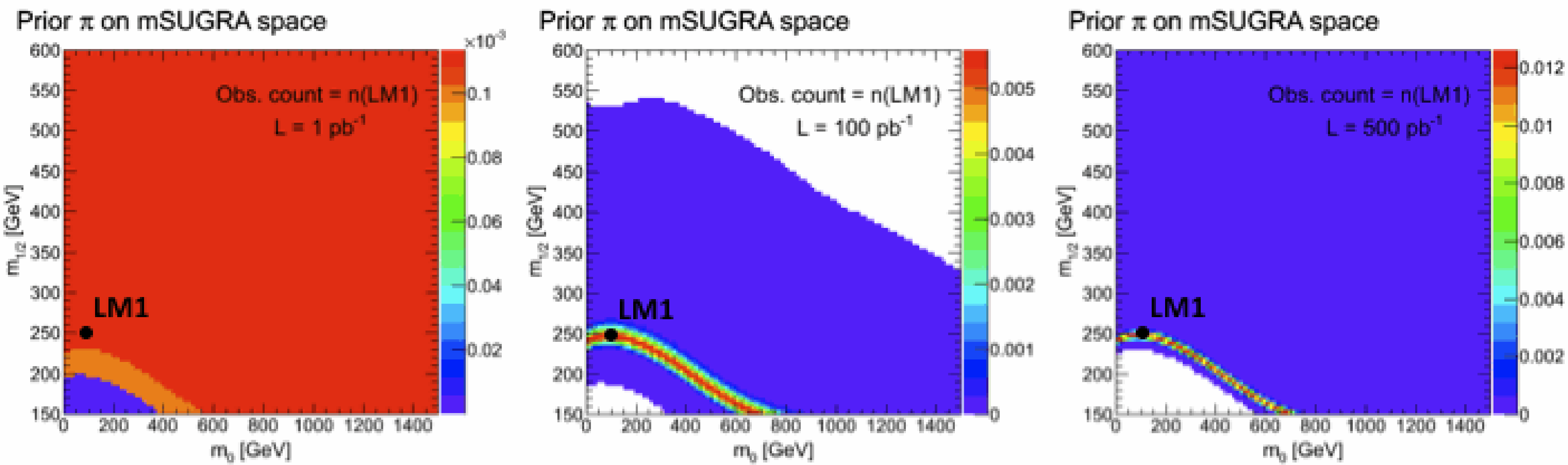


Mapping to the SUSY space

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

Given $s = f(\theta = m_0, m_{1/2})$, intervals $\delta s \in R$ map into regions $\Theta_{\delta s} \in \Theta$ which will have the same probability content. We then impose the look-alike condition:

"every point within $\Theta_{\delta s}$ be assigned the same prior density."



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

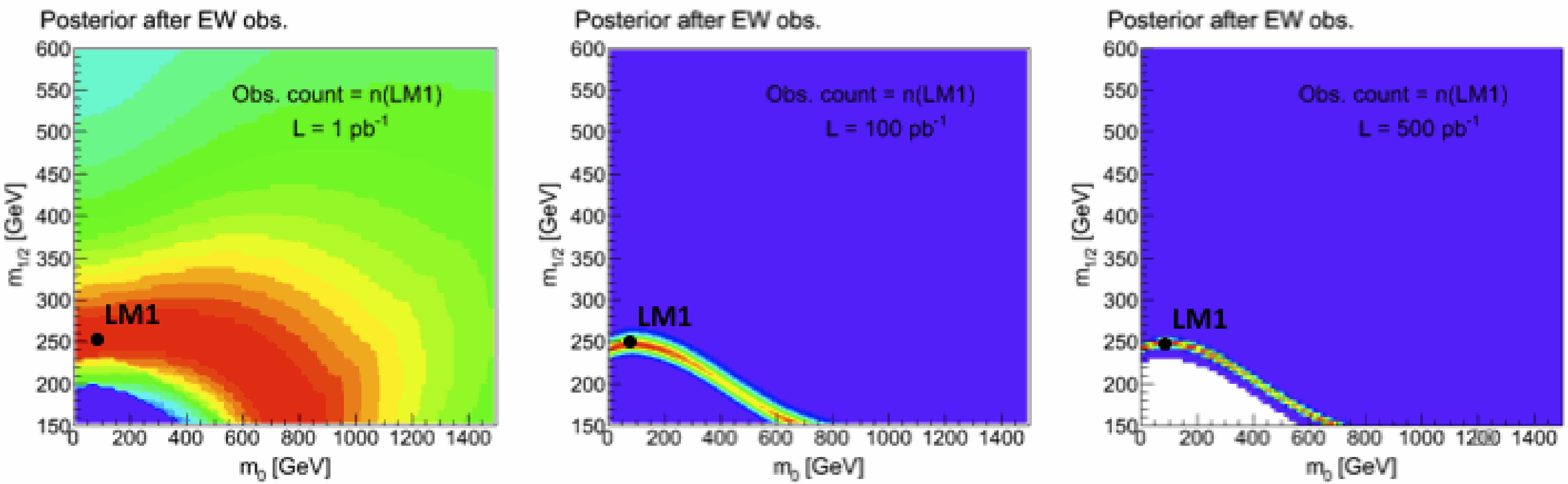
Add EW/flavor observables

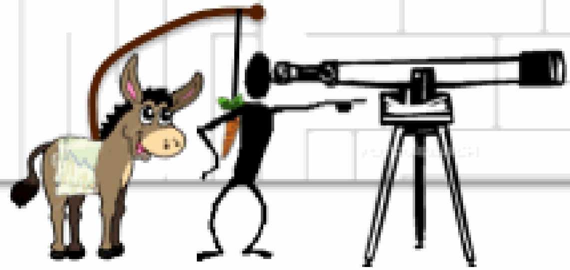
We continue the inference chain by incorporating the likelihood

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

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

Since the state of nature is LM1, we use the LM1 values for the observables along with the measured uncertainties from current experiments.





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.
- 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.
- Work is in progress to use Tevatron results to construct priors suitable for analyses at the LHC.
- We need to find observables that will break the degeneracy in the look-alike regions.