NCTS

The Third NCTS School on FeynRules-Madgraph for LHC Physics

Bayesian Applications: Global Fitting in Dark Matter Particle Model

Yue-Lin Sming Tsai

Kavli IPMU

University of Tokyo

2014, June

Bayesian and Madgraph

- Bayesian+MCMC can be used for the calculation of cross sections by integration over multi-dimensional phase space, e.g. Bayesian Neural Networks.

- **Bayesian theory is a good tool to marginalize the nuisance parameters and estimate background and signal strength.**

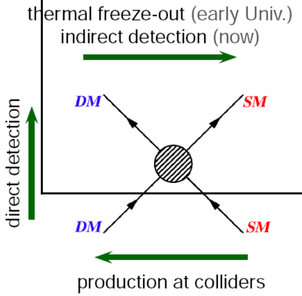
Bayesian and Madgraph

- Bayesian+MCMC can be used for the calculation od cross sections by integration over multi-dimensional phase space, e.g. Bayesian Neural Networks.

- **Bayesian theory is a good tool to marginalize the nuisance parameters and estimate background and signal strength.**

Too many free parameters in DM search!

	Particle physics	Astrophysics	Other nuisance parameters
Indirect detection	1. cross-section/decay-time 2. DM mass 3. energy spectrum	1. DM halo profile 2. velocity distributions 3. propagation models	1. background parameters 2. Standard Model parameters
Direct detection	1. cross-section 2. DM mass	1. DM local density 2. velocity distributions 3. propagation models	1. Hadronic parameters 2. Standard Model parameters
Colliders	1. couplings 2. masses 3. particle contents (model dependent)		1. Standard Model parameters



The modern high energy experiments always report the result in energy dependent event numbers. How do we interpret them?

Contents

- Recap Chi-squared method.
- What is probability?
- Bayesian statistics:
    1. Likelihood
    2. Prior
    3. Posterior
    4. Evidence
- How to perform a Bayesian global scan, a beginners guide.

Contents

- Recap Chi-squared method.
- What is probability?
- Bayesian statistics:
  1. Likelihood
  2. Prior
  3. Posterior
  4. Evidence
- How to perform a Bayesian global scan, a beginners guide.

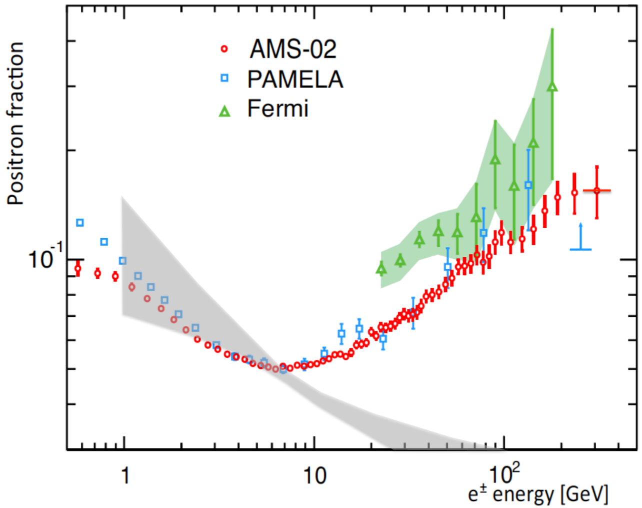
Lecture I

Lecture II

Recap chi-squared method

1. Theoretical prediction per energy bin, P(M,E), function of model parameters.
2. N data points from ONE experiment, with central value D(E) and errors sigma(E).
3. The total chi-squared for this experiment is:

$$\chi^2 = \sum_{i=1}^{N} \frac{[p(M, E_i) - D(E_i)]^2}{\sigma^2(E_i)}$$



Sometimes, if one wants to compute the chi-squared by combining several data sets from different experiment, the weight of each experiment has to be properly considered.

$$\chi^2_{tot} = \sum_i w_i \times \chi^2_i$$

where i runs over the experiments, e.g., AMS02, Fermi, and PAMELA.

Recap chi-squared method

Weighted combination:

For Poisson distribution, with detected events, N(obs)>>1, we define two experiments' chi^2 as

$$\chi^2(D_1) = (\frac{N_1 - N_1^{obs}}{\sigma_1})^2; \ \chi^2(D_2) = (\frac{N_2 - N_2^{obs}}{\sigma_2})^2$$

$$\chi^2(D_1 + D_2) = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \times \chi^2(D_1) + \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \times \chi^2(D_2)$$

$$X_{tot}^2 = \sum_i w_i \times X_i^2 \qquad \sigma = \sqrt{N^{obs}}; \ W_{1,2} = \frac{N_{1,2}^{obs}}{N_1^{obs} + N_2^{obs}}$$

Here, we ingnore systematic errors of two experiments.

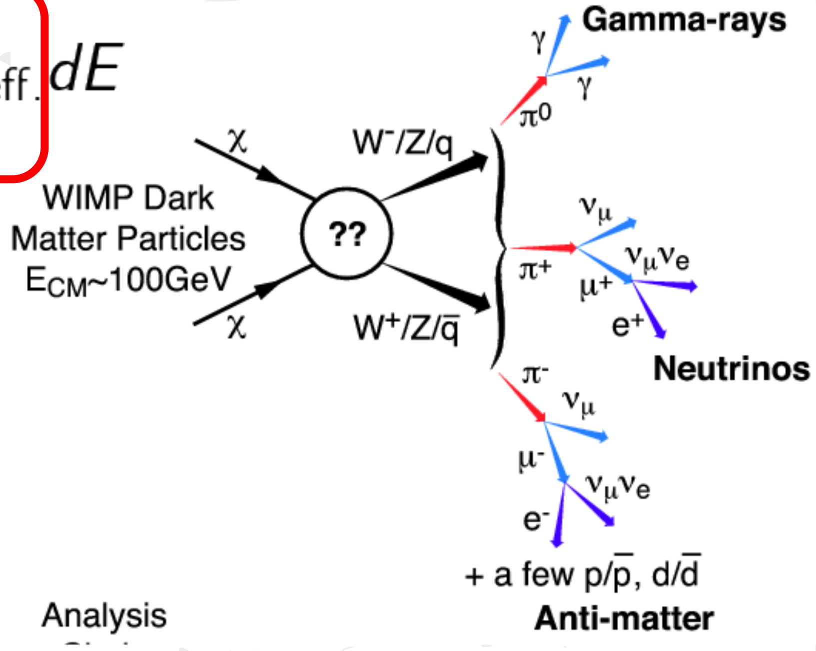
Recap chi-squared method

Statistical & Systematic errors:

$$N_{\mathrm{obs}} = N_{\mathrm{bkg.}} + T \int \sigma_{\mathrm{ann.}} \times \frac{dN}{dE} \times A_{\mathrm{eff.}} \, dE$$

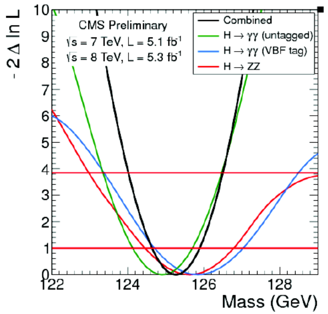
Systematic errors

statistical errors : $\sqrt{N_{\mathrm{obs}}}$



What is probabilty?

Characterization of the excess: mass



July 4th 2012 The Status of the Higgs Search J. Incandela for the CMS COLLABORATION

To reduce model dependence, allow for free cross sections in three channels and fit for the common mass:

$m_X = 125.3 \pm 0.6$ GeV

$$m_h^{Exp.} = 126.0 \pm \sigma$$

$$126.0 - n\sigma \leq m_h^{theor.} \leq 126.0 + n\sigma$$

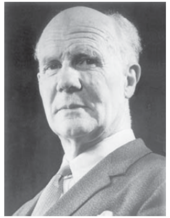
Terminologically, one can say the prediction with n sigma far away from the central value. However, can we discribe this in terms of probability?

96

What is probabilty?

Two schools in statistics: *frequentists* and *Bayesians*.

Thomas Bayes



Jerzy Neyman, Egon Pearson and Ronald Fischer

We have two choices:

1. **P(data|Model):** $P(data|model) \sim e^{-\frac{\chi^2}{2}}$
   Probability (data, given parameter), the probability constructed "in the data", Frequentist approach.
2. **P(Model|data):**
   Probability (parameter, given data), the probability constructed "in the model", Bayesian approach.

Model = male or female
Data = pregnant or not pregnant

P(female|pregnant) .ne. P(pregnant|female)
>>0.3 ~0.3

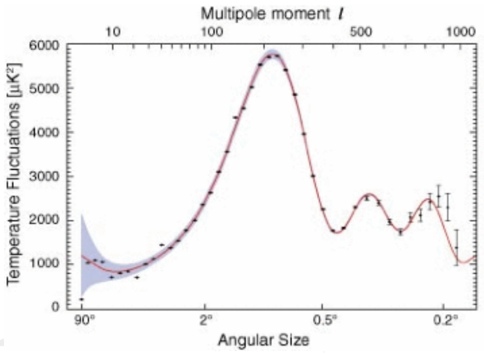
$$P(model|data) \times P(data) = P(data|model) \times P(model)$$

There is no single, “right” statistics...

- Bayesians: "probability" = degree of believability. Unknown quantities are treated probabilistically and the state of the world can always be updated. By Bayes (18th century).
- Frequentists: "probability" = long-run fraction having this characteristic. Sampling is infinite. (19th century)
- Likelihoodists: Single sample inference based on maximizing the likelihood function. By Fisher et al. (20th century).

Bayesian statistics is very popular in many branches of science (astronomy, cosmology, etc.).

For example, The Wilkinson Microwave Anisotropy Probe (WMAP) analysis of cosmic microwave background (CMB) spectrum:



Bayesian Statistics…

$$P(model|data) \times P(data) = P(data|model) \times P(model)$$

Bayes theorem:

$$\text{Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}}.$$

- **Likelihood**: the probability of obtaining data if hypothesis is true.
- **Prior**: what we know about hypothesis **BEFORE** seeing the data.
- **Evidence**: normalization constant, crucial for model comparison.
- **Posterior**: the probability about hypothesis **AFTER** seeing the data.

Profile likelihood method

Mixed Frequentist - Bayesian

The main disadvantage of the frequentist method for current experiments (such as XENON100, LHC, IceCube, etc...) is not able to repeat exactly the same setting, for example a fixed background !!

Bayesian for nuisance parameters and approximate Frequentist to reduce dimensions to just physics parameters

So,
some frequentist's point of view: PL method is minimal Bayesian.
some Bayesian's point of view: PL method is frequentist.

$$\mathcal{L}(\psi_{i=1,\ldots,r}) = \max_{m \in \mathbb{R}^{n-r}} \mathcal{L}(m)$$

Profile Likelihood vs Marginal Posterior

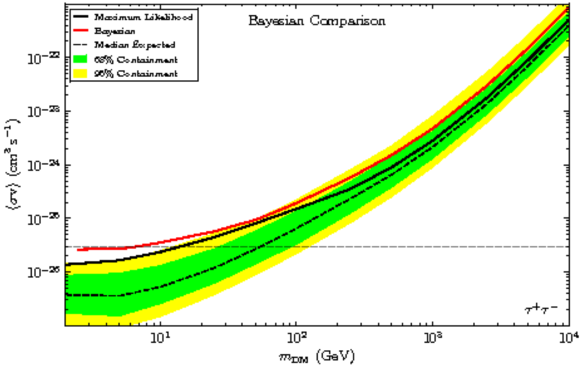
	Needs Prior	Coverage	Inference	update beliefs	Combine data	Unphysical region	Systematic/ theoretical errors
Profile Likelihood	No	very important	No	No	Easy	1/0 hard cut likelihood	profiled out over the nuisance parameters
Marginal Posterior	Yes	Unimporta nt	Yes	Yes	Hard	Excluded by prior	intergrated over prior

We will see later it is numerically easy to perform a scan by using Marginal Posterior.

Table for statistical methods in DM searching experiments!

	Fermi LAT	LHC	IceCube	XENON100/LUX	PLANCK/WMAP
Marginal Posterior	Yes	Yes	NO	NO	Yes
Profile Likelihood	Yes	Yes	Yes	Yes	NO

- More and more experiments adopt the Bayesian (Marginal posterior) approach.
- Marginal posterior method gives a conservative result.



Bayesian statistics: Part I- Likelihood

Likelihood function

- The probability of obtaining data if hypothesis is true.
- In this talk, we will show 3 standard distributions of likelihood function but there can be several distributions of likelihood, as long as the distribution is "*the probability of a given sample being randomly drawn regarded as a function of the parameters of the population*".

Gaussian Likelihood

Take a single observable $\xi(m)$ that has been measured

(e.g., M_W)

- c – central value, σ – standard exptal error
- define

$$\chi^2 = \frac{[\xi(m)-c]^2}{\sigma^2}$$

- assuming Gaussian distribution ($d \to (c,\sigma)$):

$$\mathcal{L} = p(\sigma, c|\xi(m)) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\chi^2}{2}\right]$$

- when include theoretical error estimate τ (assumed Gaussian):

$$\sigma \to s = \sqrt{\sigma^2 + \tau^2}$$

TH error "smears out" the EXPTAL range

- for several uncorrelated observables (assumed Gaussian):

$$\mathcal{L} = \exp\left[-\sum_i \frac{\chi_i^2}{2}\right]$$

Poisson Likelihood

Poisson distribution to characterize counting

$$\mathcal{L} = \prod_i \frac{e^{-(s_i+b_i)} (s_i + b_i)^{o_i}}{o_i!}$$

o_i : observed events in LHC.

b_i : expected SM background events.

s_i : $s_i = \epsilon_i \times \sigma \times \int L$.

ϵ_i : $N_i(\alpha_T > 0.55)/N_{\text{total}}$

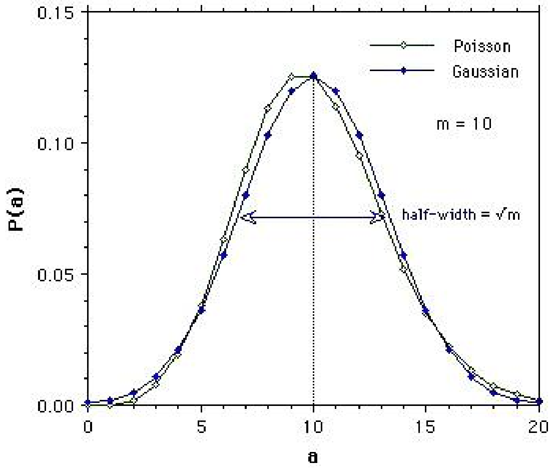
$i = 1, 2, 3..., 8.$

Poisson Likelihood:

the signal over background ratio

$$P(a) = e^{-m}\left[\frac{m^a}{a!}\right]$$

The standard deviation of Poisson distribution is: √m



$$\frac{e^{-(s_i+b')}(s_i+b')^{o_i}}{o_i!}$$

For a signal search the significance is:

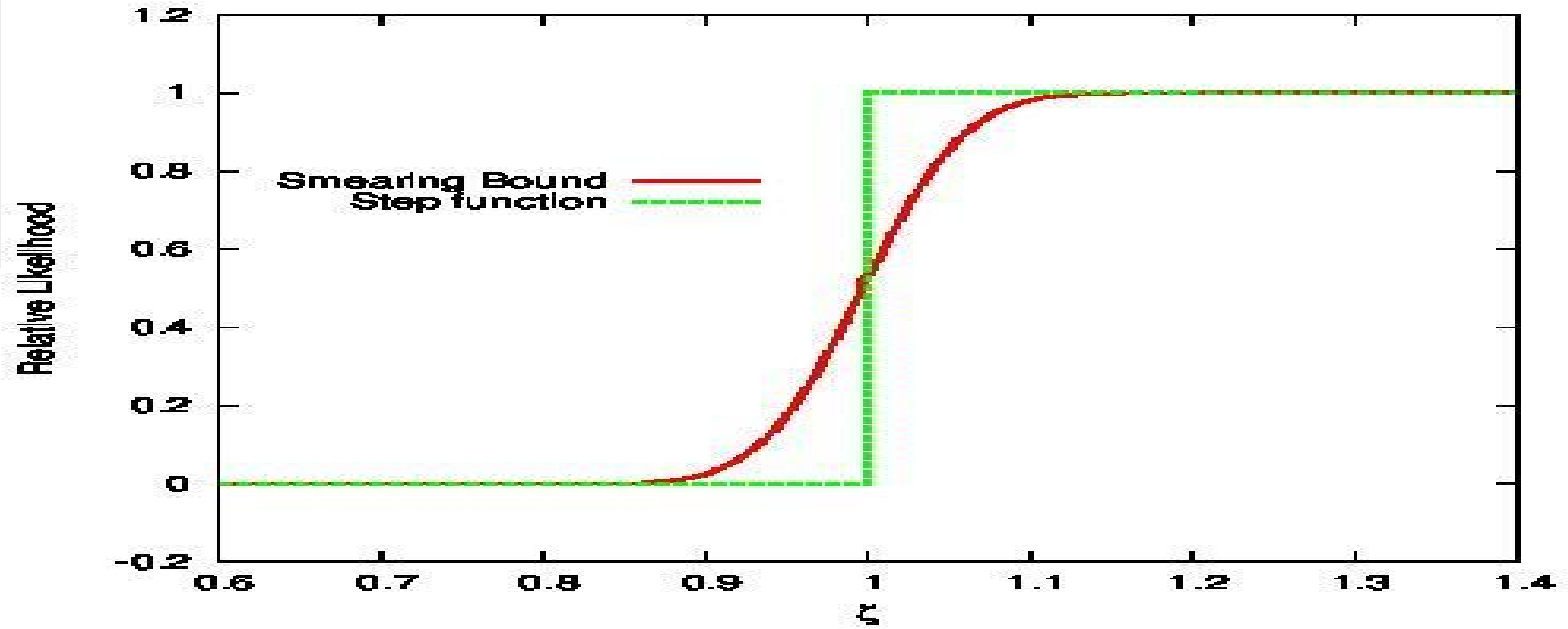
$$\text{significance} = \frac{s}{\sqrt{s+b}}$$

For a null signal (s=0) search the significance is:

$$\text{significance} = \frac{s}{\sqrt{b}}$$

Here significance means how many sigma the signal excesses/reaches to the background or signal+background?

Trick: Likelihood from limits



- Use error function to smear the bound!
- Can add theory error as well.

Test Statistic or Chi Square

$$TS = -2\ln\mathcal{L}$$

$$\text{For } \mathcal{L} \propto \exp(-\frac{\chi^2}{2}), \text{TS} = \chi^2.$$

Test DM signals

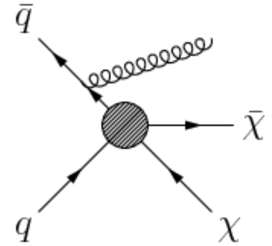
$$\delta\chi^2_s = \chi^2(s+b) - \chi^2_{\rm min}(b);$$

$$\delta\chi^2_{s+b} = \chi^2(s+b) - \chi^2_{\rm min}(s+b);$$

The first method is more conservative and
delta chi-squared can be negative!
For combined analysis, we are always using the
second method.

An example: Mono-jet Likelihood

E_T^{miss} (GeV) →	> 250	> 300	> 350	> 400	> 450	> 500	> 550
$Z(\nu\nu)$+jets	30600 ± 1493	12119 ± 640	5286 ± 323	2569 ± 188	1394 ± 127	671 ± 81	370 ± 58
W+jets	17625 ± 681	6042 ± 236	2457 ± 102	1044 ± 51	516 ± 31	269 ± 20	128 ± 13
$t\bar{t}$	470 ± 235	175 ± 87.5	72 ± 36	32 ± 16	13 ± 6.5	6 ± 3.0	3 ± 1.5
$Z(\ell\ell)$+jets	127 ± 63.5	43 ± 21.5	18 ± 9.0	8 ± 4.0	4 ± 2.0	2 ± 1.0	1 ± 0.5
Single t	156 ± 78.0	52 ± 26.0	20 ± 10.0	7 ± 3.5	2 ± 1.0	1 ± 0.5	0 ± 0
QCD Multijets	177 ±88.5	76 ±38.0	23 ±11.5	3 ±1.5	2 ±1.0	1 ± 0.5	0 ± 0
Total SM	49154 ± 1663	18506 ± 690	7875 ± 341	3663 ± 196	1931 ± 131	949 ± 83	501 ± 59
Data	50419	19108	8056	3677	1772	894	508
Exp. upper limit	3580	1500	773	424	229	165	125
Obs. upper limit	4695	2035	882	434	157	135	131



$$L(o_i|b_i + s_i)) = \max_{b'}\{\frac{e^{s_i+b'}(s_i+b')^{o_i}}{o_i!}\exp[-\frac{(b'-b_i)^2}{2\delta b_i}]\}$$

o_i : observed events in LHC.

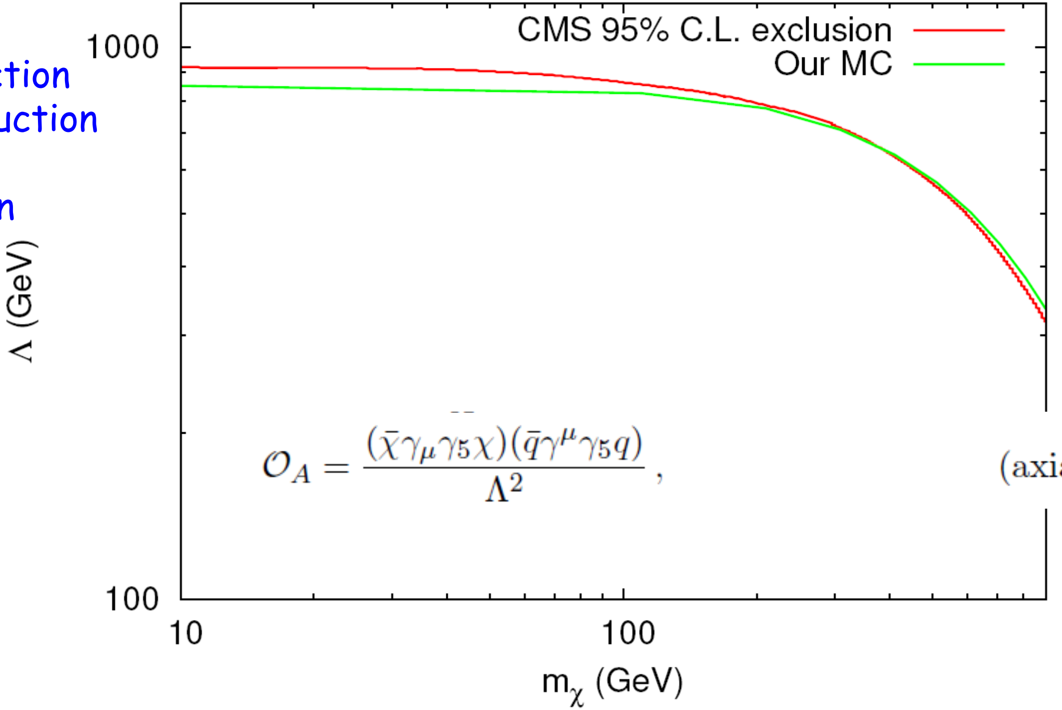
b_i : expected SM background events.

s_i : $s_i = \epsilon_i \times \sigma \times \int L$.

An example: Mono-jet Likelihood

1. Feynrules -> vertex
2. Madgraph-> cross section
3. MadEvent-> reconstruction
4. Cut applied
5. likelihood computation
6. TS=2.71

$$TS = -2\ln\mathcal{L}$$



$$\mathcal{O}_A = \frac{(\bar{\chi}\gamma_\mu\gamma_5\chi)(\bar{q}\gamma^\mu\gamma_5 q)}{\Lambda^2}\,, \qquad \text{(axial vector, } s\text{-channel)}$$

CL, p-value, and Chi-square

File Edit View Search Terminal Tabs Help

Terminal × Terminal ×

sigma	1.00	1.28	1.64	1.96	2.00	2.58	3.00	3.29	4.00
conf_int	68.27%	80.00%	90.00%	95.00%	95.45%	99.00%	99.73%	99.90%	99.99%
p-value	0.31731	0.20000	0.10000	0.05000	0.04550	0.01000	0.00270	0.00100	0.00006
chi2(k=1)	1.00	1.64	2.71	3.84	4.00	6.63	9.00	10.83	16.00
chi2(k=2)	2.30	3.22	4.61	5.99	6.18	9.21	11.83	13.82	19.33
chi2(k=3)	3.53	4.64	6.25	7.81	8.02	11.34	14.16	16.27	22.06
chi2(k=4)	4.72	5.99	7.78	9.49	9.72	13.28	16.25	18.47	24.50
chi2(k=5)	5.89	7.29	9.24	11.07	11.31	15.09	18.21	20.52	26.77
chi2(k=6)	7.04	8.56	10.64	12.59	12.85	16.81	20.06	22.46	28.91
chi2(k=7)	8.18	9.80	12.02	14.07	14.34	18.48	21.85	24.32	30.96
chi2(k=8)	9.30	11.03	13.36	15.51	15.79	20.09	23.57	26.12	32.93
chi2(k=9)	10.42	12.24	14.68	16.92	17.21	21.67	25.26	27.88	34.85
chi2(k=10)	11.54	13.44	15.99	18.31	18.61	23.21	26.90	29.59	36.72
chi2(k=11)	12.64	14.63	17.28	19.68	19.99	24.72	28.51	31.26	38.54
chi2(k=12)	13.74	15.81	18.55	21.03	21.35	26.22	30.10	32.91	40.33
chi2(k=13)	14.84	16.98	19.81	22.36	22.69	27.69	31.66	34.53	42.09
chi2(k=14)	15.94	18.15	21.06	23.68	24.03	29.14	33.20	36.12	43.82
chi2(k=15)	17.03	19.31	22.31	25.00	25.34	30.58	34.71	37.70	45.52
chi2(k=16)	18.11	20.47	23.54	26.30	26.65	32.00	36.22	39.25	47.20
chi2(k=17)	19.20	21.61	24.77	27.59	27.95	33.41	37.70	40.79	48.86
chi2(k=18)	20.28	22.76	25.99	28.87	29.24	34.81	39.17	42.31	50.50
chi2(k=19)	21.36	23.90	27.20	30.14	30.52	36.19	40.63	43.82	52.13
chi2(k=20)	22.44	25.04	28.41	31.41	31.80	37.57	42.08	45.31	53.73

```
smingtsai@smingtsai-ThinkPad-T520 ~/test $
```

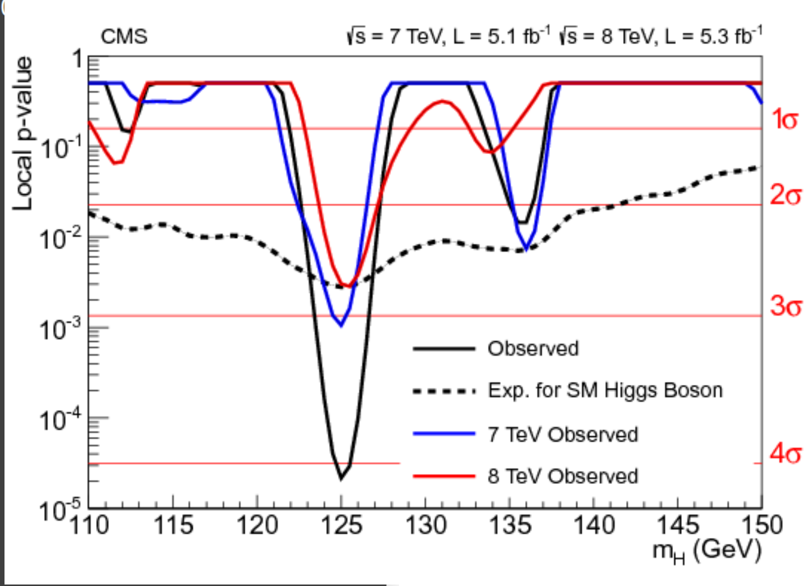
$$CL = \frac{\int_0^y \mathcal{L}(x)dx}{\int_0^\infty \mathcal{L}(x)dx} = 1 - \mathcal{P}$$

Only vaild for Gaussian-like likelihood!

CL, p-value, and Chi-square

sigma	1.00	1.28	1.64	1.96	2.00	2.58	3.00	3.29	4.00
conf_int	68.27%	80.00%	90.00%	95.00%	95.45%	99.00%	99.73%	99.90%	99.99%
p-value	0.31731	0.20000	0.10000	0.05000	0.04550	0.01000	0.00270	0.00100	0.00006
chi2(k=1)	1.00	1.64	2.71	3.84	4.00	6.63	9.		
chi2(k=2)	2.30	3.22	4.61	5.99	6.18	9.21	11		
chi2(k=3)	3.53	4.64	6.25	7.81	8.02	11.34	14		
chi2(k=4)	4.72	5.99	7.78	9.49	9.72	13.28	16		
chi2(k=5)	5.89	7.29	9.24	11.07	11.31	15.09	18		
chi2(k=6)	7.04	8.56	10.64	12.59	12.85	16.81	20		
chi2(k=7)	8.18	9.80	12.02	14.07	14.34	18.48	21		
chi2(k=8)	9.30	11.03	13.36	15.51	15.79	20.09	23		
chi2(k=9)	10.42	12.24	14.68	16.92	17.21	21.67	25		
chi2(k=10)	11.54	13.44	15.99	18.31	18.61	23.21	26		
chi2(k=11)	12.64	14.63	17.28	19.68	19.99	24.72	28		
chi2(k=12)	13.74	15.81	18.55	21.03	21.35	26.22	30		
chi2(k=13)	14.84	16.98	19.81	22.36	22.69	27.69	31		
chi2(k=14)	15.94	18.15	21.06	23.68	24.03	29.14	33		
chi2(k=15)	17.03	19.31	22.31	25.00	25.34	30.58	34		
chi2(k=16)	18.11	20.47	23.54	26.30	26.65	32.00	36		
chi2(k=17)	19.20	21.61	24.77	27.59	27.95	33.41	37		
chi2(k=18)	20.28	22.76	25.99	28.87	29.24	34.81	39		
chi2(k=19)	21.36	23.90	27.20	30.14	30.52	36.19	40		
chi2(k=20)	22.44	25.04	28.41	31.41	31.80	37.57	42		

```
smingtsai@smingtsai-ThinkPad-T520 ~/test $
```



evidence (3 sigma) but discovery (5 sigma)!

P-value in Global fitting

$$\mathcal{L}_{D5} = \frac{g_S}{\Lambda}\bar{\chi}\chi H^\dagger H \qquad \mathcal{L}_{D6}^{Higgs} = \frac{g_D}{\Lambda^2}(\bar{\chi}\gamma_\mu\gamma_5\chi)(H^\dagger iD^\mu H)$$

$$\mathcal{L}_{D6}^{Lepton} = \sum_{i=1}^{3}\frac{1}{\Lambda^2}(\bar{\chi}\gamma_\mu\gamma_5\chi)(g_{LL}\bar{L}^i\gamma^\mu L^i + g_{RE}\bar{E}_R^i\gamma^\mu E_R^i)$$

$$\mathcal{L}_{D6}^{Quark} = \sum_{i=1}^{3}\frac{1}{\Lambda^2}(\bar{\chi}\gamma_\mu\gamma_5\chi)(g_{LQ}\bar{Q}^i\gamma^\mu Q^i + g_{Ru}\bar{U}^i\gamma^\mu U^i + g_{Rd}\bar{D}^i\gamma^\mu D^i)$$

Constraints	PLANCK (relic)	LUX (SI)	X100 (SD)	gamma-ray	Mono-jet	Mono-photon	inv. Z	inv. H
constrained couplings	ALL	gS	Quark, gD	ALL	Quark, gD	Lepton, gD	gD	gS

Problem: we do not know how many d.o.f. precisely are.

P-value in Global fitting

We can always determine a p-value for any experiment by repeating the psudo-experiments. Howevr, we cannot decide a p-value for a combined analysis from several independent DM experiments result.

Contents

- Recap Chi-squared method.
- What is probability?
- Bayesian statistics:
  - ① Likelihood
  - ② Prior
  - ③ Posterior
  - ④ Evidence
- How to perform a Bayesian global scan, a beginners guide.

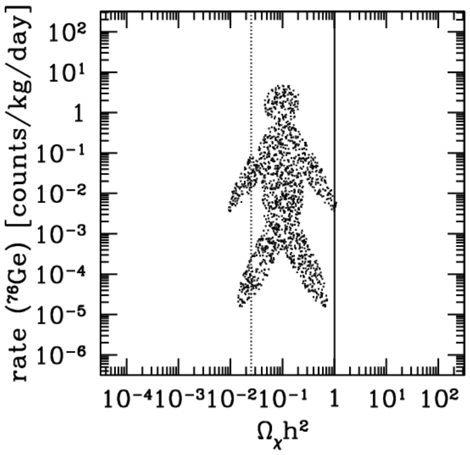
Lecture II

Bayesian statistics: Part II- Prior

Prior: what we know about hypothesis **BEFORE** seeing the data, as "state of knowledge".

Prior choice

"The definition of "interesting" is different for different investigators, and the way points are generated always involves a prior in parameter space (even grid methods can be said to have a prior, namely a series of Dirac delta functions at each grid point). One could go to the extreme of producing any kind of results by choosing appropriate priors."



Boudjema et al., arXiv:1003.4748

Prior choice

- There is no "right" choice of prior.

- There are wrong/dishonest choices, e.g., a delta-function for a parameter that you know nothing about.

- Bayesian statistics is a calculus of beliefs. It cannot tell you what your prior beliefs should be.

Question:

Given a interesting range, how do we map a random seed, $0<f<1$, generated by computer, to a physical parameter?

For example, we are interesting on scaning top mass.

mt.py (~/Dropbox/TaipeiMadgraph) - VIM

File Edit View Search Terminal Help

```python
#!/usr/bin/env python

from math import *
import random

def FlatPrior(vmin,vmax,seed):
   ans=(vmax-vmin)*seed + vmin
   return ans

def LogPrior(vmin,vmax,seed):
   ans=(log10(vmax)-log10(vmin))*seed + log10(vmin)
   ans=10.0**ans
   return ans


def GauPrior(cv,width,seed1,seed2):
   ans=cv+width*sqrt(-2.0*log(seed1))*cos(2.0*pi*seed2)
   return ans



vmin=10.0
vmax=1.0e3
cv=173.5
width=1.0

seed=random.random()
seed1=random.random()
seed2=random.random()

print FlatPrior(vmin,vmax,seed),LogPrior(vmin,vmax,seed),GauPrior(cv,width,seed1,seed2)
~
```

r : uniform random number in interval $[0, 1]$

$f(x)$: given probability density function

$$f: x \in [-\infty, \infty] \rightarrow f(x) \in [0, \infty] \qquad \int_{-\infty}^{\infty} f(x)\, dx = 1$$

We define the following two monotone nondecreasing functions $F(x), G(x)$

$$F(x) \equiv \int_{-\infty}^{x} f(x) d\,x \qquad F(-\infty) = 0, F(\infty) = 1$$

$$G(x) \equiv F^{-1}(x) \qquad G(0) = -\infty, G(0) = \infty$$

Using this function G, map of uniform random numbers

$$y = G(r)$$

yields density function $f(x)$.

$\text{Flat prior} : 10 \leq \frac{m_t}{\text{GeV}} \leq 10^3$

$\text{Log prior} : 1 \leq \log\left[\frac{m_t}{\text{GeV}}\right] \leq 3$

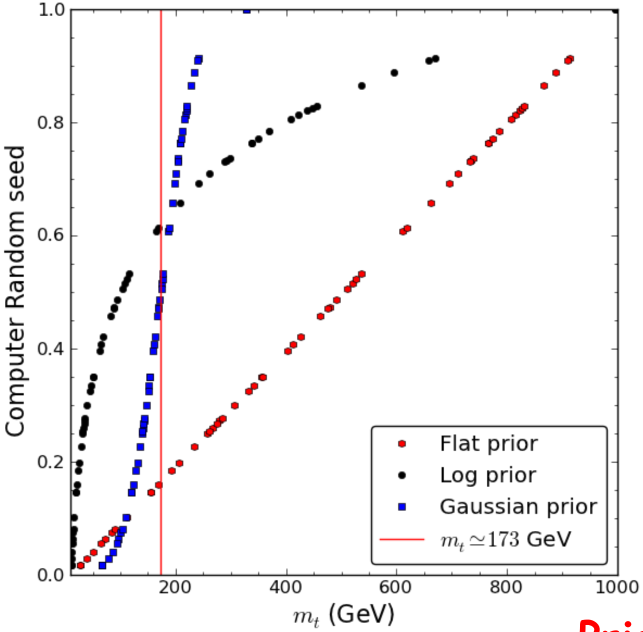
Gaussian prior : Gaussian Distribution

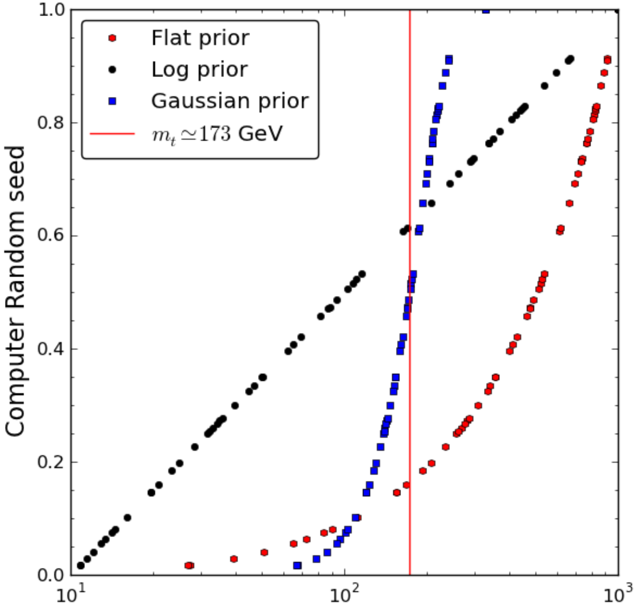
Prior: what we know about hypothesis BEFORE seeing the data.

$$\text{Flat prior}: 10 \leq \frac{m_t}{\text{GeV}} \leq 10^3$$

$$\text{Log prior}: 1 \leq \log\left[\frac{m_t}{\text{GeV}}\right] \leq 3$$

$$\text{Gaussian prior}: \text{Gaussian Distribution}$$





Prior dependence is two-fold:

- **Prior range.**
- **Prior distribution.**

PERSONAL PROBABILITIES

This is a story I originally heard from Nobel Prize winner Frank Wilczek in a slightly different context, but it illustrates the way that for Bayesians the assessment of probability can differ from person to person.

A shy postdoc is attending a workshop on the topic of 'Extra Dimensions'. Each evening, after an intensive day's work, he goes to the local bar, sits next to an empty chair and orders two glasses of wine, one for himself and the other for the empty chair. By the third evening, the barman's curiosity cannot be controlled and he asks the postdoc why he always orders the extra glass of wine. 'I work on the theory of extra dimensions', explains the postdoc, 'and it is possible that there are beautiful girls out there in 12 dimensions, and maybe by quantum mechanical tunneling they might appear in our 3-dimensional world, and perhaps one of them might materialise on this empty chair, and I would be the first person talking to her, and then she might go out with me'. 'Yes', says the barman, 'but there are three very attractive real girls sitting over there on the other side of this bar. Why don't you go and ask them if they would go out with you?' 'There's no point', replies the postdoc, 'that would be very unlikely.'

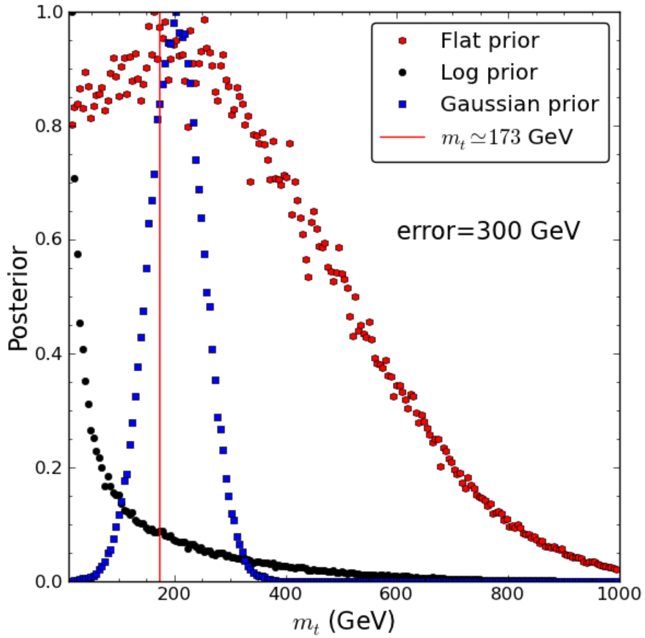
Taken from: arXiv: 1301.1273v1
Louis Lyons

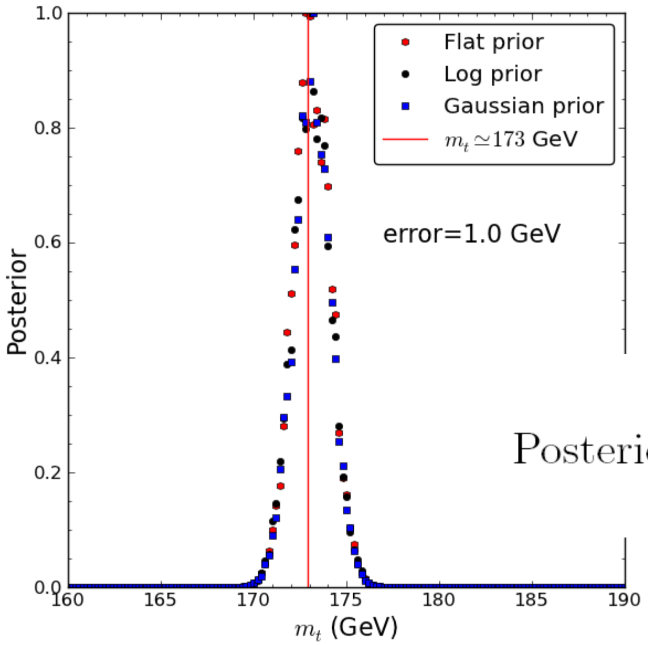
Bayesian statistics: Part III- Posterior

- Posterior: the probability about hypothesis **AFTER** seeing the data.
- The probability constructed "in the model", i.e., we think about only the data we have, not pseudo-data from imaginary experiments!

- If the Likelihood is well-peaked, the posterior follows the Likelihood.
- Otherwise, it follows the prior.

$$\mathcal{L} = \exp(-\left[\frac{m_b - 173}{2 \times \text{error}}\right]^2)$$

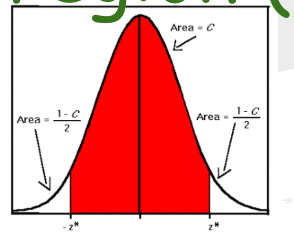




$$\text{Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}}.$$

Confidence limits (Profile Likelihood) and Credible region (Marginal Posterior)

o: observed events
b: background expected events
s: signal expected events



likelihood

Profile Likelihood

$$1 - CL = \frac{\int_{s'}^{\infty} \max[\mathcal{L}(o|s+b')]ds}{\int_{0}^{\infty} \max[\mathcal{L}(o|s+b')]ds}$$

Marginal Posterior

$$1 - CR = \frac{\int_{s'}^{\infty} \mathrm{d}s \int_0^{\infty} \mathcal{P}(s+b|o)\mathrm{d}b}{\int_{0}^{\infty} \mathrm{d}s \int_0^{\infty} \mathcal{P}(s+b|o)\mathrm{d}b}$$

Posterior

Numerical binning

Y-axis

0.0041, 0.25, 0.095	0.86, 0.56, 0.097	0.62, 0.79, 0.21
0.47, 0.86, 0.51	0.54, 0.041, 0.052	0.71, 0.48, 0.54
0.61, 0.27, 0.13	0.91, 0.51, 0.99	0.82, 0.63, 0.27

X-axis

Marginalizing

0.35	1.52	1.62
1.84	0.63	1.73
1.01	2.41	1.72

$$1 - CR = \frac{\int_{s'}^{\infty} \mathrm{d}s \int_0^{\infty} \mathcal{P}(s+b|o)\mathrm{d}b}{\int_0^{\infty} \mathrm{d}s \int_0^{\infty} \mathcal{P}(s+b|o)\mathrm{d}b}$$

normalization factor=12.83

Sort and find 68% and 95% C.L.

>95%	<95%	<68%
<68%	<95%	<68%
<95%	<68%	<68%

Profiling

0.25	0.86	0.79
0.86	0.54	0.71
0.61	0.99	0.82

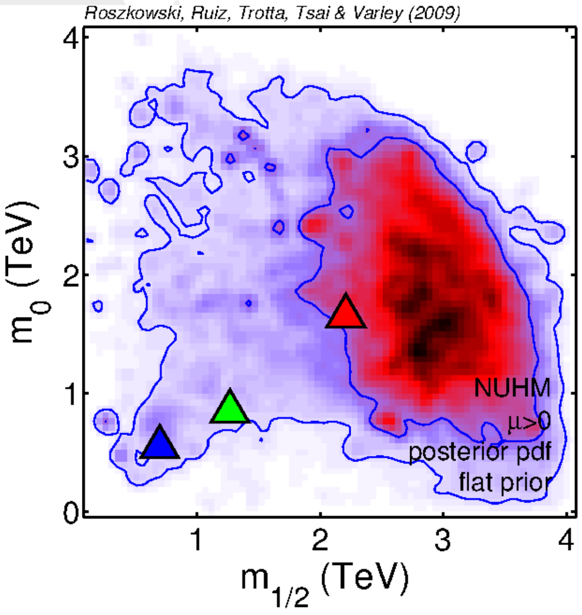
$$1 - CL = \frac{\int_{s'}^{\infty} \max[\mathcal{L}(o|s+b')]ds}{\int_0^{\infty} \max[\mathcal{L}(o|s+b')]ds}$$

Sort and find 68% and 95% C.L.

>95%	<68%	<68%
<68%	<95%	<95%
<95%	<68%	<68%

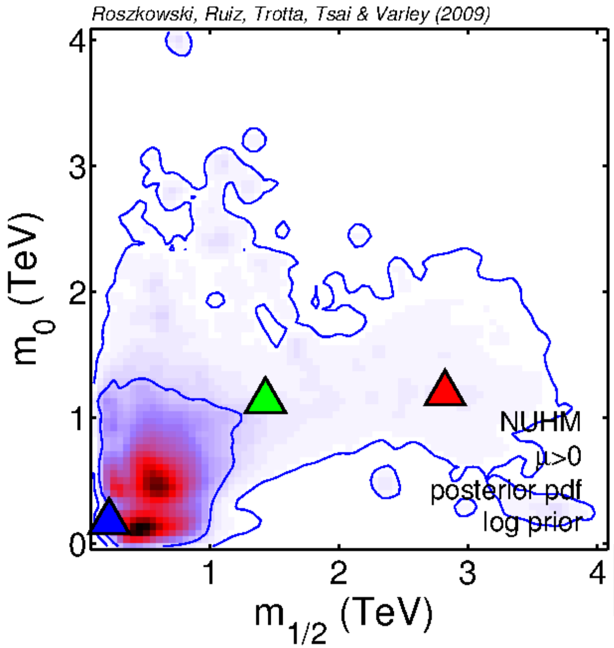
normalization factor=6.43

prior dependence



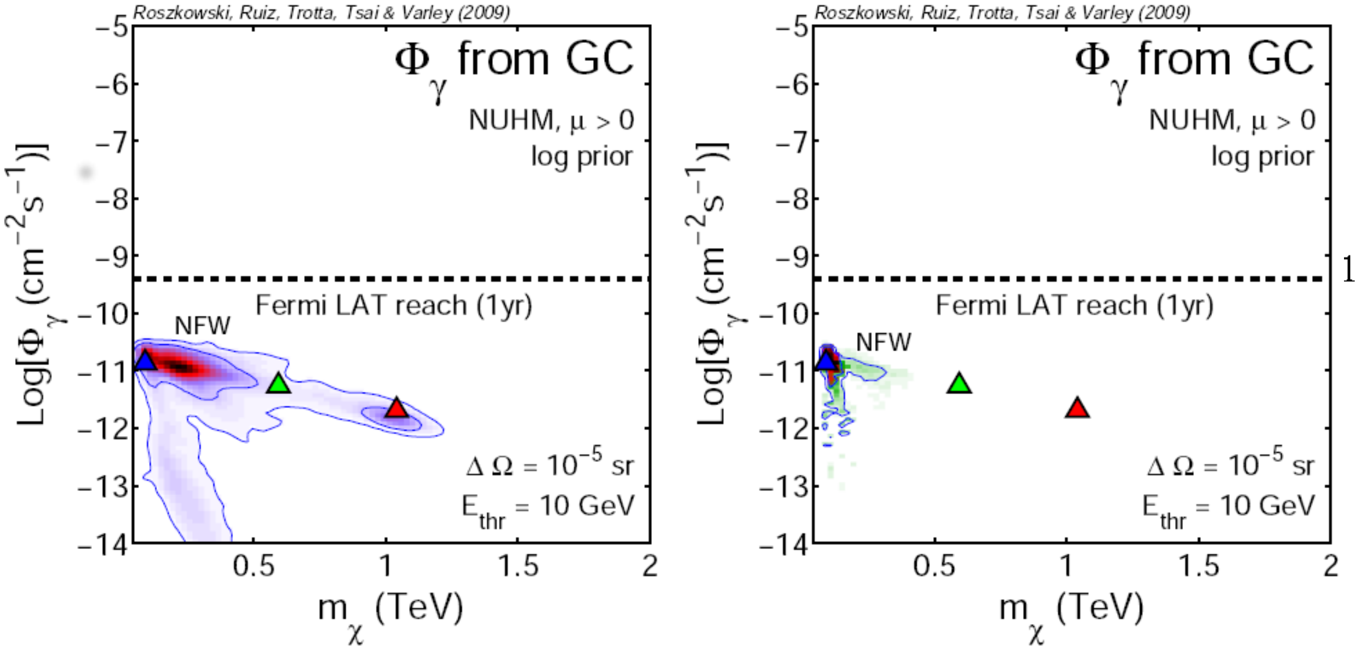
Flat prior

$$1 - CR = \frac{\int_{s'}^{\infty} \mathrm{d}s \int_0^{\infty} \mathcal{P}(s+b|o)\mathrm{d}b}{\int_0^{\infty} \mathrm{d}s \int_0^{\infty} \mathcal{P}(s+b|o)\mathrm{d}b}$$



Log prior

Volume effect



$$1 - CR = \frac{\int_{s'}^{\infty} \mathrm{d}s \int_0^{\infty} \mathcal{P}(s+b|o)\mathrm{d}b}{\int_0^{\infty} \mathrm{d}s \int_0^{\infty} \mathcal{P}(s+b|o)\mathrm{d}b}$$

The Volume effect can be very strong if the likelihood function is not very stronge.

Bayesian statistics:
Part IV- Evidence

Evidence: normalization constant, crucial for model comparison.

$$\mathcal{Z} = \int \mathcal{L}(D|\theta)\pi(\theta)d^n\theta$$

- Evidence is probability of data given model. One model in one scan only has one value of evidence.
- It contains information from both likelihood and prior.
- If evidence is small, model is fine-tuned, namely, prior agrees with data only in small part of parameter space.
- The ratio of two evidences reveals which model is better.

Model comparsion

Bayes factor and p-value

Given two competing models, $\mathcal{M}_0$ and $\mathcal{M}_1$, the Bayes factor B_{01} is the ratio of the models' evidences

$$B_{01} \equiv \frac{p(d|\mathcal{M}_0)}{p(d|\mathcal{M}_1)}, \quad (3)$$

where large values of B_{01} denote a preference for $\mathcal{M}_0$, and small values of B_{01} denote a preference for $\mathcal{M}_1$. The "Jeffreys' scale" (Table I) gives an empirical prescription for translating the values of B_{01} into strengths of belief.

Given two or more models, specified in terms of their parameterisation *and* priors on the parame-

$\lvert\ln B_{01}\rvert$	Odds	Strength of evidence
< 1.0	$\lesssim 3:1$	Inconclusive
1.0	$\sim 3:1$	Weak evidence
2.5	$\sim 12:1$	Moderate evidence
5.0	$\sim 150:1$	Strong evidence

Taken from 0811.2415v1

Table 2. **Relation between Fixed Sample Size *P* Values and Minimum Bayes Factors and the Effect of Such Evidence on the Probability of the Null Hypothesis**

P Value (Z Score)	Minimum Bayes Factor	Decrease in Probability of the Null Hypothesis, %		Strength of Evidence
		From	To No Less Than	
0.10 (1.64)	**0.26** (1/3.8)	75 50 17	44 21 5	Weak
0.05 (1.96)	**0.15** (1/6.8)	75 50 26	31 13 5	Moderate
0.03 (2.17)	**0.095** (1/11)	75 50 33	22 9 5	Moderate
0.01 (2.58)	**0.036** (1/28)	75 50 60	10 3.5 5	Moderate to strong
0.001 (3.28)	**0.005** (1/216)	75 50 92	1 0.5 5	Strong to very strong

Goodness of fit

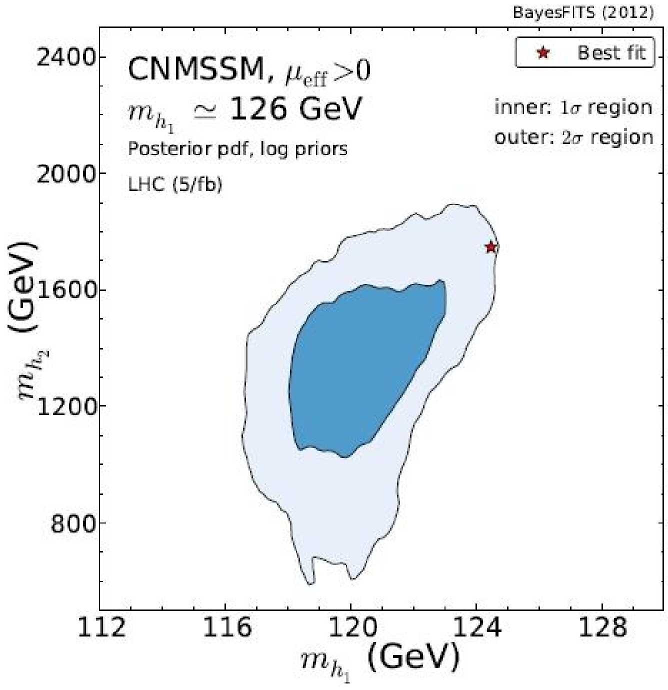
Connection between the Bayesian approach and the fine-tuning measure

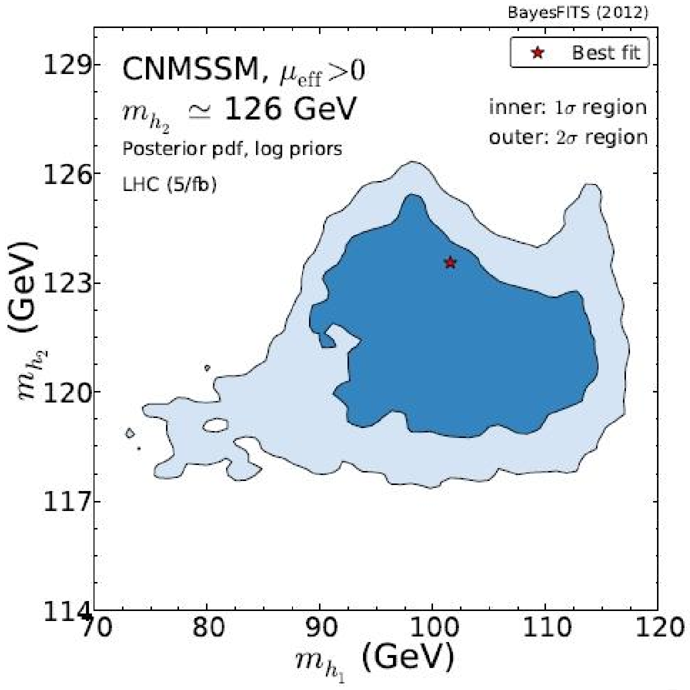
"A frequentist analysis is not sensitive to the fine-tuning. Fine-tuning has to do with statistical weight and a frequentist analysis is based entirely on likelihood, i.e. the ability to reproduce the experiment, and thus cannot see the fine-tuning. "

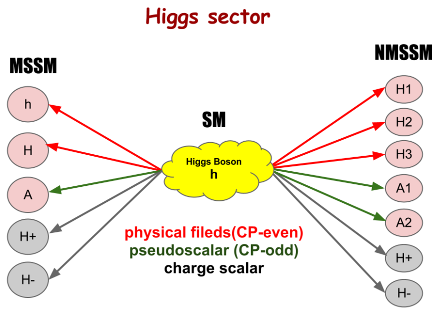
"It may happen that a point (or a region) in the parameter space can present an optimal likelihood, but only after an extreme tuning of the unplotted parameters, involving cancellations. Usually that point is considered very implausible or disfavored since, a priori, cancellations are not likely unless there exists some known theoretical reason for them. However, as long as the point is capable to reproduce the experimental data, the fine-tuning considerations do not affect its privileged condition in a frequentist analysis. This fact can favor points in the frequentist approach, e.g. in the low-energy regions, which are suppressed in the Bayesian one."

Taken from Ref. 1212.4821 (Cabrera, Casas, and Ruiz de Austri)

Fine-tuning: Higgs masses





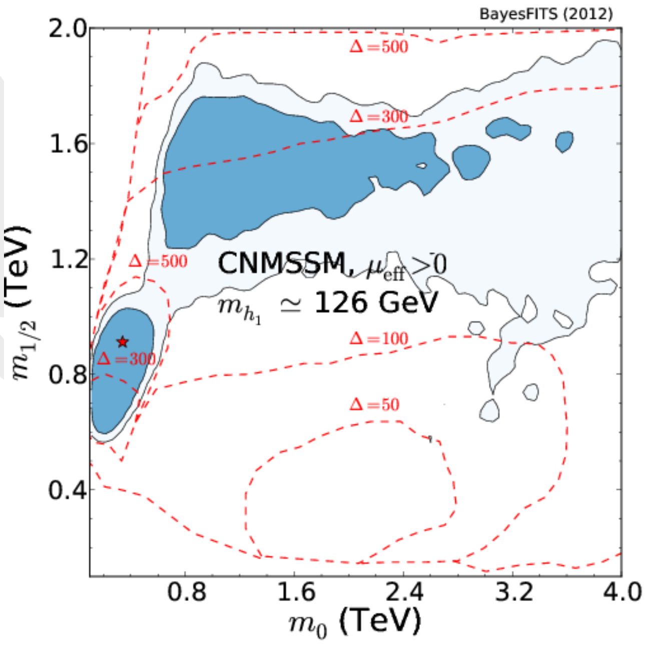


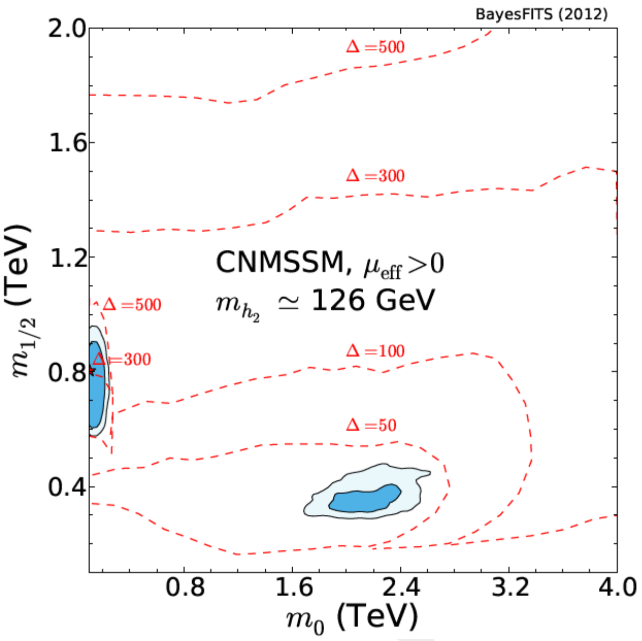
Although best-fit point is not the language of Bayesian, we can still see fine-tunning from the relationship between the location of best-fit and poterior.

The "Barbieri-Giudice" measure

$$\Delta_{BG}(p_i) = \left| \frac{\partial \ln m_Z^2}{\partial \ln p_i^2} \right|, \quad \Delta_{BG} = \max\{\Delta_{BG}(p_i)\}.$$

$$\frac{m_Z^2}{2} = \frac{(m_{H_d}^2 + \delta m_{H_d}^2) - (m_{H_u}^2 + \delta m_{H_u}^2)\tan^2\beta}{\tan^2\beta - 1} - \mu^2.$$



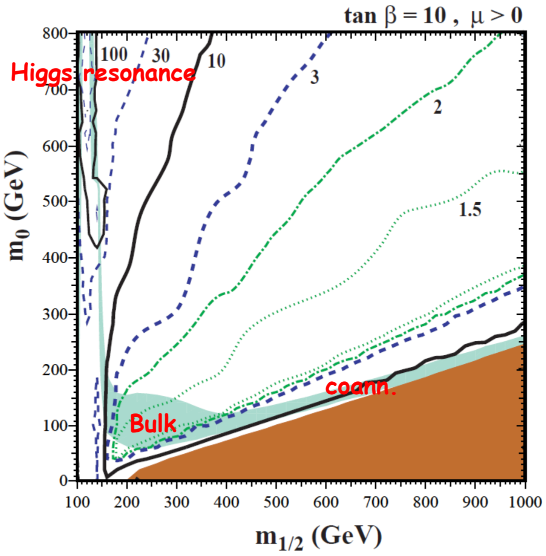


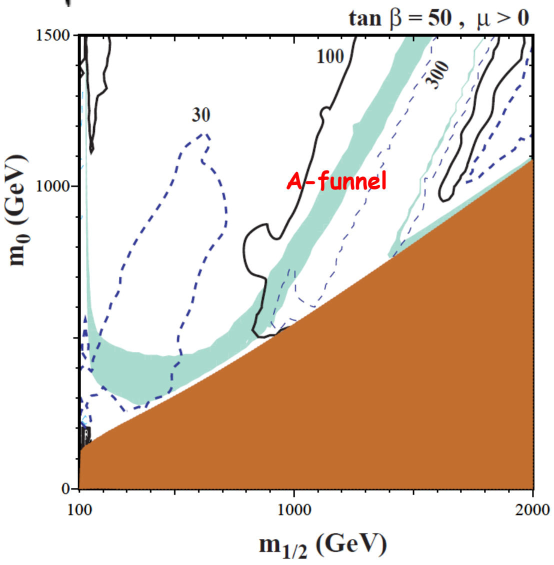
We are approaching the Higher fine-tuning region.

Fine-tuning: relic density

$$\Delta_{P_i}^{\Omega_\chi} = \sqrt{\sum_{P_i} \left| \frac{\partial \ln \Omega_\chi}{\partial \ln P_i} \right|^2}$$

arXiv:hep-ph/0105004
J. Ellis and K. Olive





- m_\chi ~ m_h/2 (Higgs resonance)
- m_\chi ~ m_A/2 (A-funnel)
- m_{stop,stau}-m_\chi ~ 10 GeV, stop (stau) coannihilation.
- All the Sfermion mass light (Bulk)

We can see that fine-tuning due to the mechianism.

Fine-tuning and Bayesian evidence

- Fine-tuning measure where the likelihood changes by varying input parameters in a small region.
- Bayesian evidence tells us the how much prior and likelihood agree each other.
- The fine-tuning measure in Bayesian statistics is the evidence!
- Whether people believe fine-tuning in the nature or not, this is issue of belief. One can introduce them in Bayesian prior so that the evidence can be improved.
- Should we take relic density and higgs mass distributions into our prior?

How to perform a Bayesian global scan, a beginners guide.

Useful slides at PHYSTAT
http://indico.cern.ch/event/107747/other-view?view=standard

Bayesian Tools

Public Codes

- SuperBayeS
- ROOTStat
- MultiNest
- CosmoMC
- SuperPy&SuperPlot
- BAT
- pippi
- …

Non-Public Codes/groups

- **BayesFITS**
- S. Akula, and P. Nath
- B. Allanach
- S. AbdusSalam
- C. Arina
- R. Catena
- J. Edsjo, and P. Gondolo
- …

There are many, many more to name here...

Useful slides at PHYSTAT
http://indico.cern.ch/event/107747/other-view?view=standard

Bayesian Tools

Public Codes

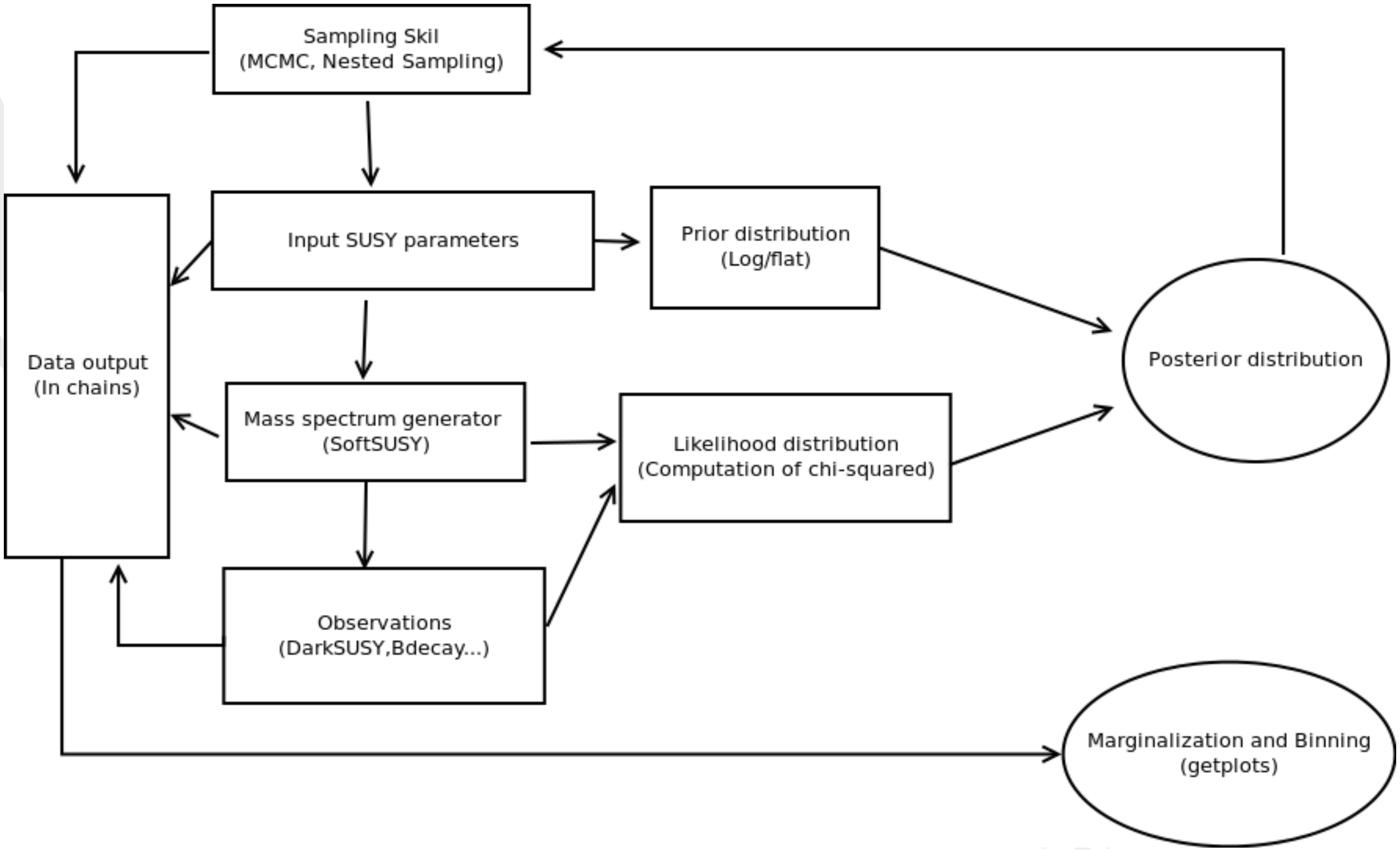
- SuperBayeS
- ROOTStat
- MultiNest
- CosmoMC
- SuperPy&SuperPlot
- BAT
- pippi
- …

Non-Public Codes/groups

- **BayesFITS**
- S. Akula, and P. Nath
- B. Allanach
- S. AbdusSalam
- C. Arina
- R. Catena
- J. Edsjo, and P. Gondolo
- …

There are many, many more to name here...

How does a scan run?

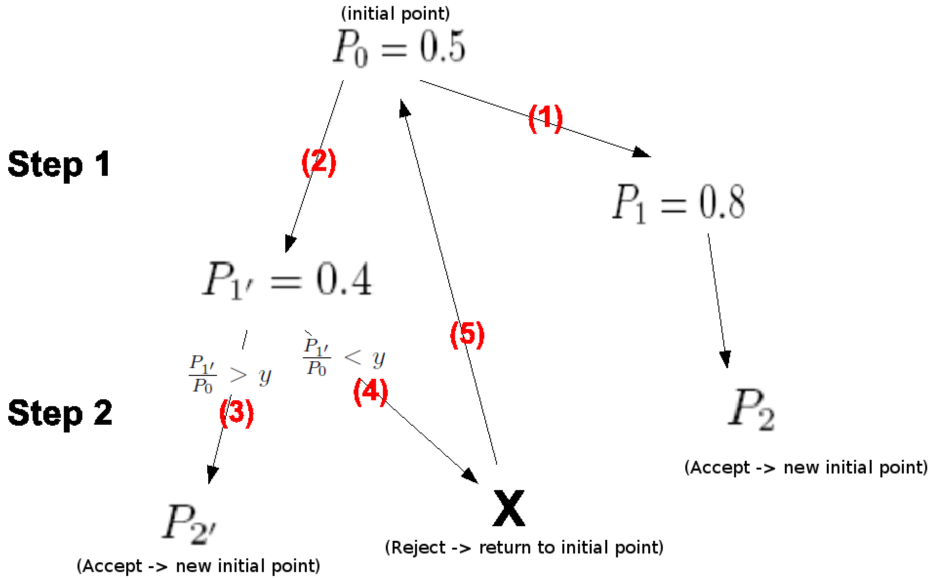


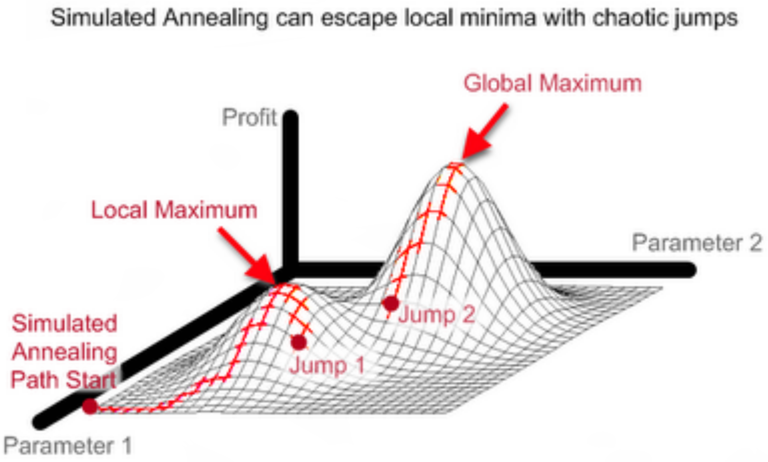
Grid Scan v.s. MultiNest

	Efficiency	nuisance parameters	Likelihood	Marginalisation	Global maxmum
Grid Scan + hard cut	very poor if D>3.	increase parameters	No need	No	Not clear
Bayesian+ MultiNest	much better	Simply include them	Need	Esay to use	prior dependency

VS

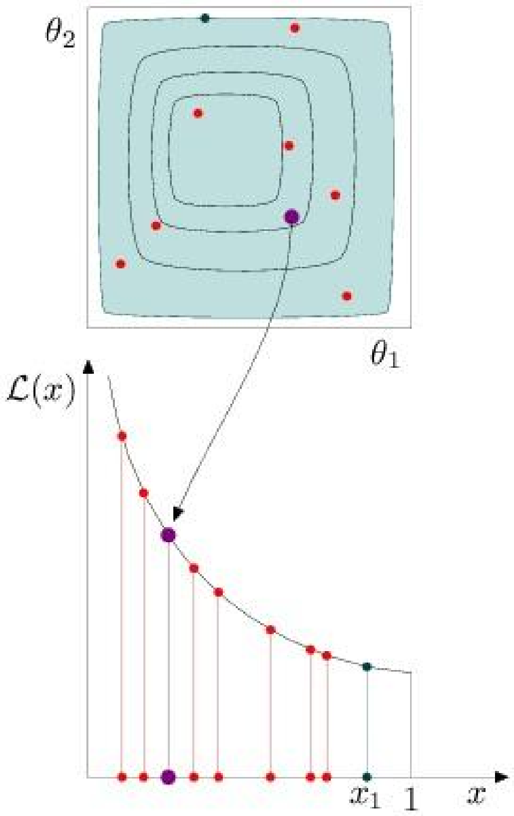
Markov chain Monte Carlo: Metropolis-Hasting Sampling





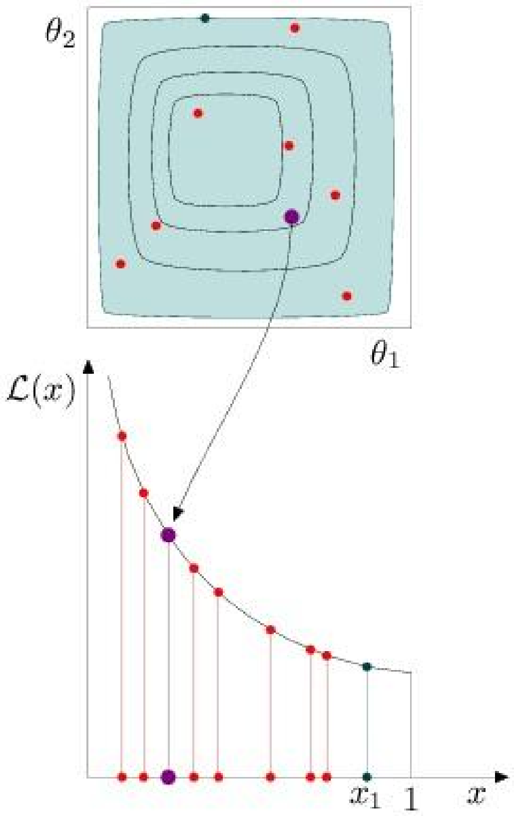
It requires "Burn in" for local Maximum.

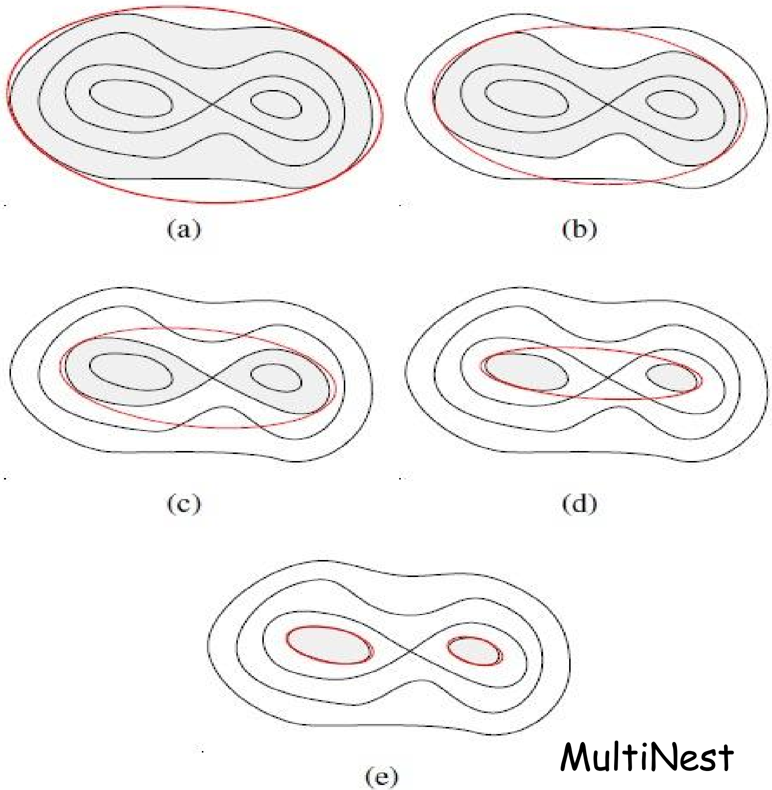
Sampling skills: Nested-Sampling



1. Randomly sample in the parameter space.
   a) Get rid of points with TS>500.
   b) Collect total number of living points equal to "nlive".
2. Sort all the likelihoods.
3. Get rid of the lowest likelihood.
4. Project the rest living points to parameter space
5. Find out the occupied region by living points.
6. Generate a new point within the occupied region, including a enlargement factor.
7. Accept the point with likelihood great than the lowest living points' likelihood, otherwise reject this point.
8. Return to step of sorting (2.) and do it again until the stop criteria satisfied.

The "MultiNest" algorithm

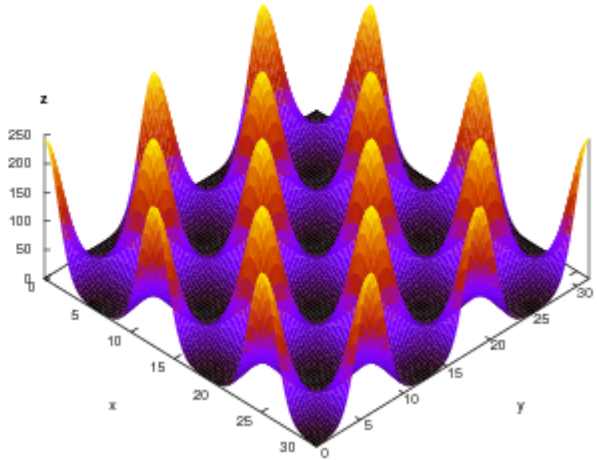


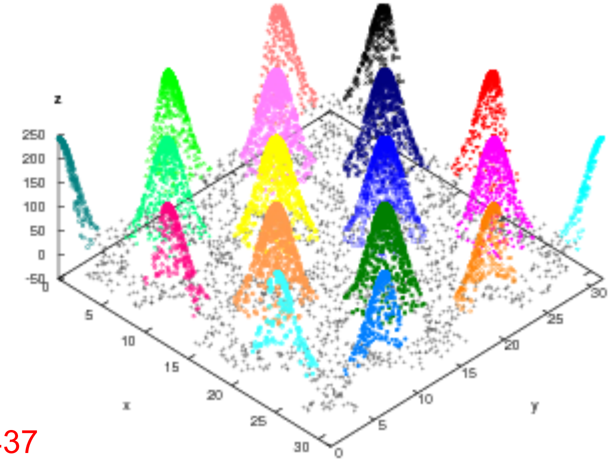


The "MultiNest" algorithm:
stop criteria

$$\delta \mathcal{Z} = \mathcal{L}_{max} \times \text{prior volume}$$

One can always set the above converge criteria less than some certain number.



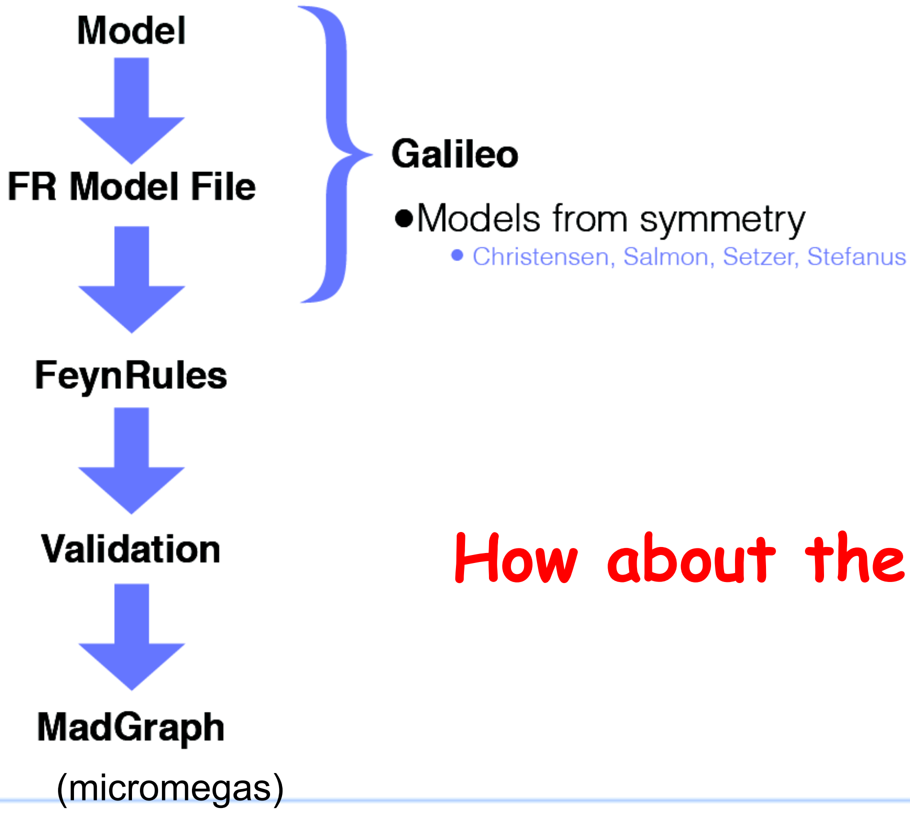


Taken from arXiv:0809.3437
F. Feroz, M.P. Hobson, M. Bridges

Challenges of Bayesian approach in particle physics

1. Pole/fine-tuning regions (if this is interesting in physics).
2. Prior dependency (weak likelihoods).
3. Reusable datasets?
4. Referees confused with frequentist's approach.

Advertisement

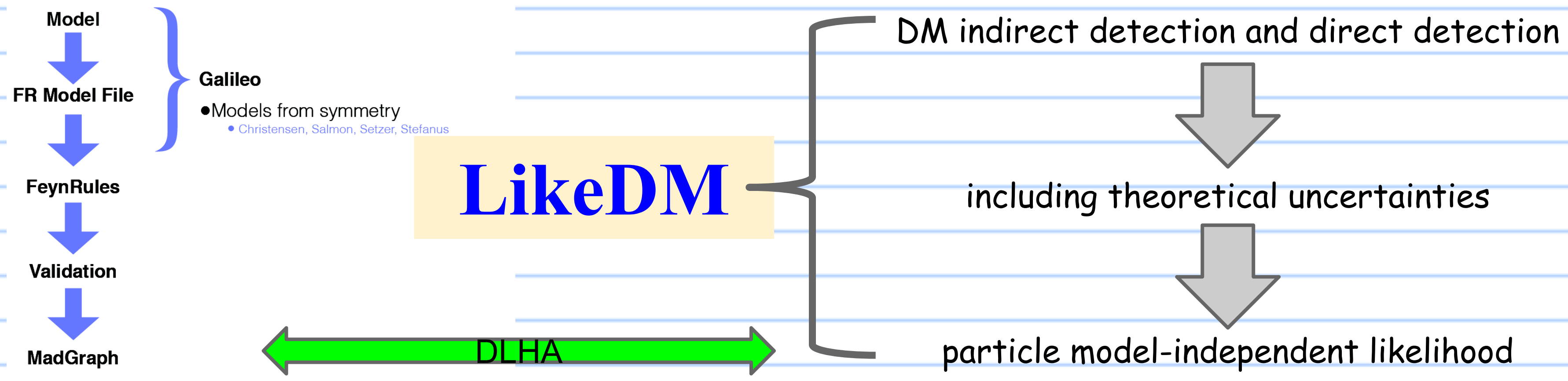


How about the experimental constraints?

Slide taken from MadGraph School 2013 Taipei, Neil Christensen

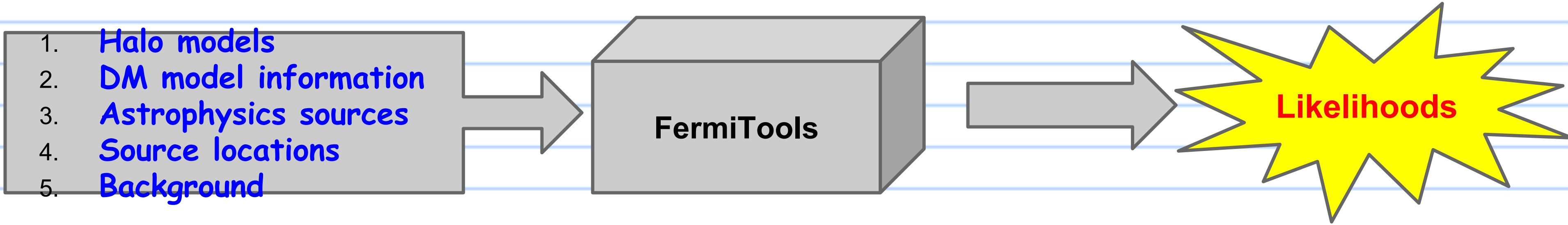
LikeDM code

in Collaboration with Q. Yuan and X. Huang



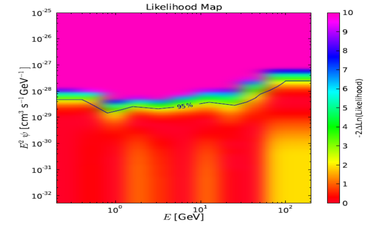
- We can more confidently and efficiently check every dark matter model.
- Can be extend to cosmology constraints.
- Similar to “DMFIT“/”HiggsBound” but starting from data level

Fitting DM gamma rays by using FermiTools



Too much CPU time consuming to do particle model fitting

Fitting DM gamma rays by using LikeDM



1. Halo models
2. Astrophysics sources
3. Source locations
4. Background

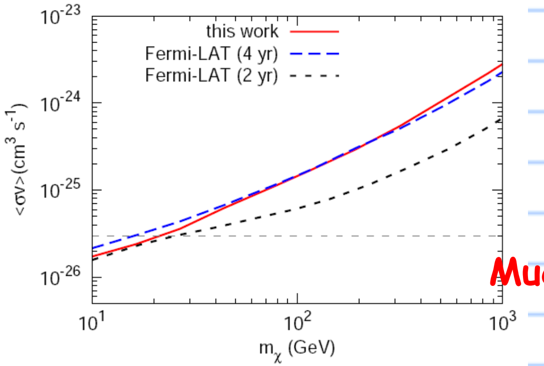
FermiTools

Energy-Residual likelihood map

Y-L Sming Tsai, Qiang Yuan, Xiaoyuan Huang
Published in JCAP 1303 (2013) 018
e-Print: arXiv:1212.3990

DM particle model information

Likelihoods



Much fast to do particle model fitting