Bayesian Inference and Particle Physics

Matthew Dolan

DAMTP
University of Cambridge

"That Scanning Conference", Stockholm

Bayesian Inference

- Consider a hypothesis H, defined by parameters $\mathbf{\Theta}$ which describe some data $\mathbf{D}$.
- Bayes theorem:

$$\Pr(\mathbf{\Theta}|\mathbf{D}, H) = \frac{\Pr(\mathbf{D}|\,\mathbf{\Theta}, H)\Pr(\mathbf{\Theta}|H)}{\Pr(\mathbf{D}|H)}, \tag{1}$$

- $\Pr(\mathbf{\Theta}|\mathbf{D}, H)$ is the posterior probability distribution of the parameters.
- $\Pr(\mathbf{D}|\,\mathbf{\Theta}, H)$ is the likelihood.

Priors and Evidence

$$\Pr(\boldsymbol{\Theta}|\mathbf{D}, H) = \frac{\Pr(\mathbf{D}|\,\boldsymbol{\Theta}, H)\,\Pr(\boldsymbol{\Theta}|H)}{\Pr(\mathbf{D}|H)},$$

- $\Pr(\boldsymbol{\Theta}|H) \equiv \pi(\boldsymbol{\Theta})$ is the prior distribution.
- The priors represent assumptions and knowledge about the problem and parameter space before the appearance of data.
- $\Pr(\mathbf{D}|H) \equiv \mathcal{Z}$ is the Bayesian evidence.

Priors and Evidence

$$\Pr(\mathbf{\Theta}|\mathbf{D}, H) = \frac{\Pr(\mathbf{D}|\,\mathbf{\Theta}, H)\,\Pr(\mathbf{\Theta}|H)}{\Pr(\mathbf{D}|H)},$$

- $\Pr(\mathbf{\Theta}|H) \equiv \pi(\mathbf{\Theta})$ is the prior distribution.
- The priors represent assumptions and knowledge about the problem and parameter space before the appearance of data.
- $\Pr(\mathbf{D}|H) \equiv \mathcal{Z}$ is the Bayesian evidence.

Parameter Estimation

- Bayesian evidence unnecessary for parameter estimation.

$$\frac{\Pr(\boldsymbol{\Theta_1}|\mathbf{D})}{\Pr(\boldsymbol{\Theta_2}|\mathbf{D})} = \frac{\Pr(\mathbf{D}|\,\boldsymbol{\Theta_1})\,\Pr(\boldsymbol{\Theta_1})}{\Pr(\mathbf{D}|\,\boldsymbol{\Theta_2})\,\Pr(\boldsymbol{\Theta_2})}$$

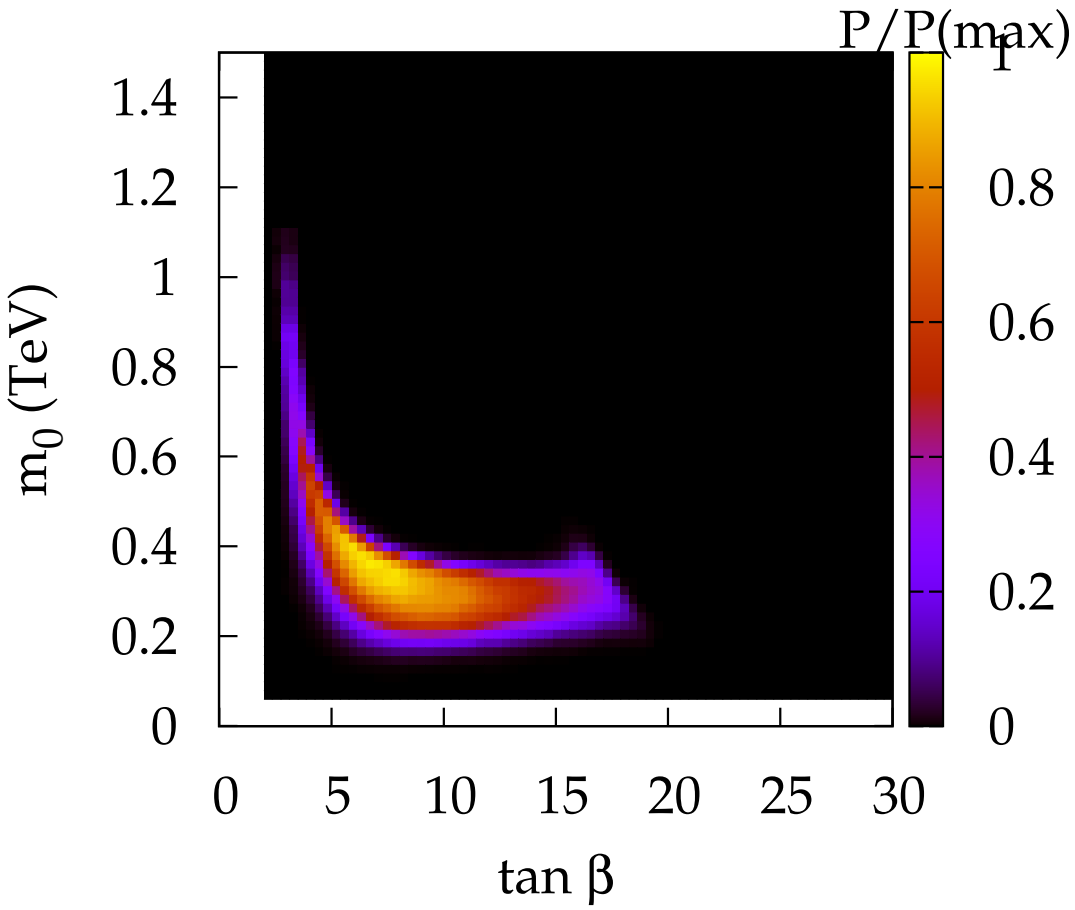
- Need to specify priors and likelihood (data) to calculate ratio of posteriors.
- As data quality increases, the posterior will become dominated by the likelihood and independent of the priors

Prior Independence in 2D Model Fits [1]

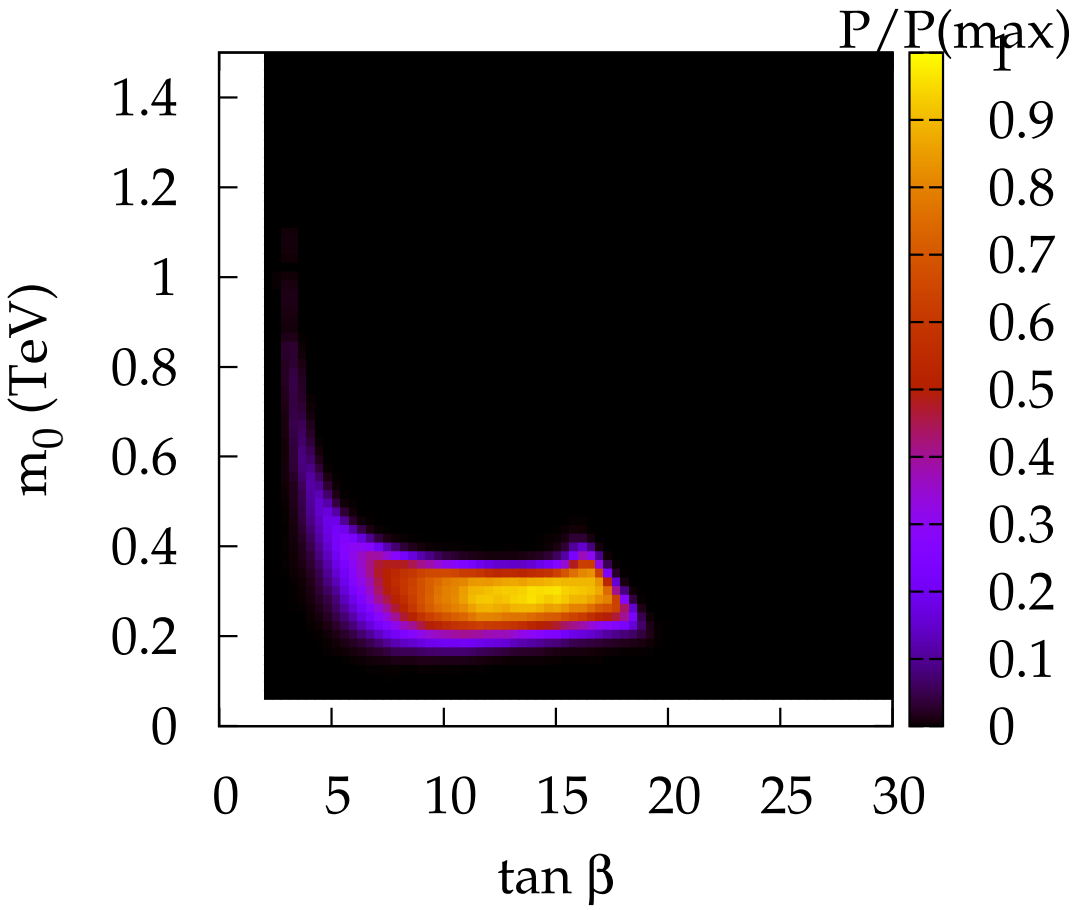
- Large Volume Scenario: IIB String theory model from moduli stabilisation.
- Soft breaking terms: m_0 and $\tan\beta$.
- Fit to $\Omega_{DM}h^2$, plus usual suite of SM observables.
- Posterior relatively independent of priors.
- Agrees with profile likelihood.

[1] Allanach, Dolan and Weber, hep-ph/0806.1184

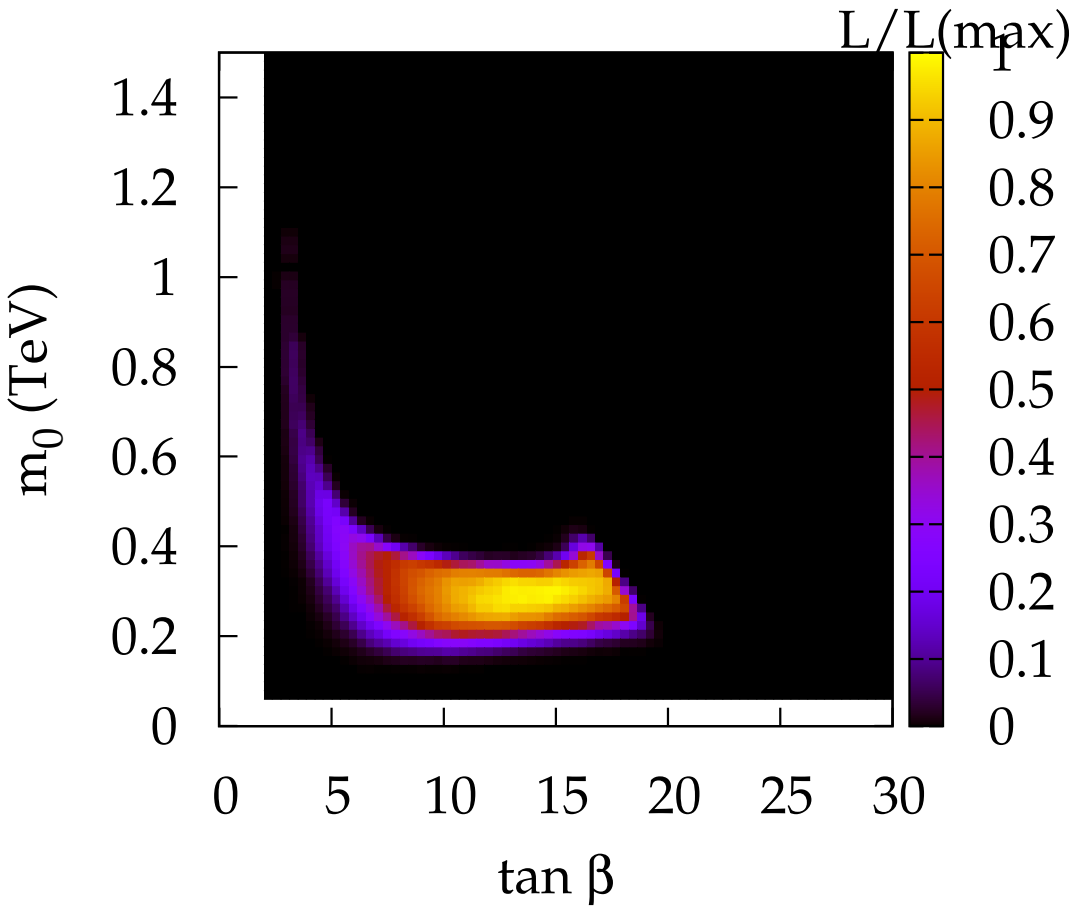
Large Volume Scenario (Flat in B_μ)



Large Volume Scenario (Flat in tan β)



Large Volume Scenario: Profile Likelihood



Bayesian Evidence

$$\Pr(\mathbf{\Theta}|\mathbf{D}, H) = \frac{\Pr(\mathbf{D}|\,\mathbf{\Theta}, H)\Pr(\mathbf{\Theta}|H)}{\Pr(\mathbf{D}|H)},$$

- The Bayesian Evidence $\Pr(\mathbf{D}|H)$ is

$$\mathcal{Z} = \int \mathcal{L}(\mathbf{\Theta})\pi(\mathbf{\Theta})d^N\mathbf{\Theta}, \tag{2}$$

- To select between models compare their posteriors

$$\frac{\Pr(H_1|\mathbf{D})}{\Pr(H_0|\mathbf{D})} = \frac{\Pr(\mathbf{D}|H_1)\Pr(H_1)}{\Pr(\mathbf{D}|H_0)\Pr(H_0)} = \frac{\mathcal{Z}_1}{\mathcal{Z}_0}\frac{\Pr(H_1)}{\Pr(H_0)}, \tag{3}$$

Bayesian Evidence

$$\Pr(\mathbf{\Theta}|\mathbf{D}, H) = \frac{\Pr(\mathbf{D}|\,\mathbf{\Theta}, H)\,\Pr(\mathbf{\Theta}|H)}{\Pr(\mathbf{D}|H)},$$

- The Bayesian Evidence $\Pr(\mathbf{D}|H)$ is

$$\mathcal{Z} = \int \mathcal{L}(\mathbf{\Theta})\pi(\mathbf{\Theta})d^N\mathbf{\Theta}, \tag{2}$$

- To select between models compare their posteriors

$$\frac{\Pr(H_1|\mathbf{D})}{\Pr(H_0|\mathbf{D})} = \frac{\Pr(\mathbf{D}|H_1)\,\Pr(H_1)}{\Pr(\mathbf{D}|H_0)\,\Pr(H_0)} = \frac{\mathcal{Z}_1}{\mathcal{Z}_0}\frac{\Pr(H_1)}{\Pr(H_0)}, \tag{3}$$

Bayesian Evidence

- Jeffreys' scale for evaluating the strength of the log evidence:

$\lvert \log \Delta \mathcal{Z} \rvert$	Odds	Probability	Remark
< 1.0	$\lesssim 3:1$	< 0.750	Inconclusive
1.0	$\sim 3:1$	0.750	Weak Evidence
2.5	$\sim 12:1$	0.923	Moderate Evidence
5.0	$\sim 150:1$	0.993	Strong Evidence

The MSSM and Breaking Supersymmetry[2]

- We like SUSY because it solves the hierarchy problem, dark matter candidate etc.
- Experimental fact: SUSY is broken at low energies.

[2] AbdusSalam, Allanach, Dolan, Feroz, Hobson; 0906.0957

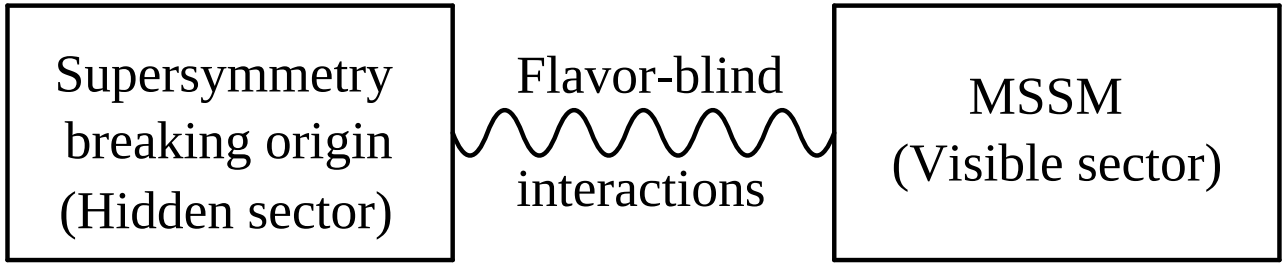
The MSSM and Breaking Supersymmetry[2]

- We like SUSY because it solves the hierarchy problem, dark matter candidate etc.
- Experimental fact: SUSY is broken at low energies.

[2] AbdusSalam, Allanach, Dolan, Feroz, Hobson; 0906.0957

The MSSM and Breaking Supersymmetry

- We like SUSY because it solves the hierarchy problem, dark matter candidate etc.
- Experimental fact: SUSY is broken at low energies.
- Mediation: SUSY broken in hidden sector, communicated via messenger sector to visible sector.



Mediation Mechanisms

Gravity Mediation

SUSY breaking via non-renormalisable terms in SUGRA Lagrangian: CMSSM (4 d.o.f.).

Anomaly Mediation

SUSY broken via superconformal anomaly: mAMSB (3).

Gauge Mediation

SUSY breaking communicated via messenger gauge multiplets: mGMSB (3)

Moduli Mediation

String theoretic Calabi-Yau moduli fields mediate SUSY breaking: LVS (2)

Aim

- Calculate the evidence for these 4 different avatars of the MSSM.
- Use MultiNest, available as part of SuperBayes package.
- Need to specify priors and likelihood calculation.

$$\mathcal{Z} = \int \mathcal{L}(\mathbf{\Theta})\pi(\mathbf{\Theta})d^N\mathbf{\Theta},$$

Priors

- Linear priors (in $\tan\beta$): $\pi(\mathbf{\Theta}_1) = \pi(\mathbf{\Theta}_2)$.
- Log priors: flat in the logarithm of the parameter.
- Natural priors[3] are flat in B and μ. Related via Jacobian factor to flat priors:

$$J = \frac{M_Z}{2}\left|\frac{B}{\mu\tan\beta}\frac{\tan^2\beta - 1}{\tan^2\beta + 1}\right| \tag{4}$$

[3] Allanach *et al*, hep-ph/0705.0487

Priors

- Linear priors (in $\tan\beta$): $\pi(\mathbf{\Theta}_1) = \pi(\mathbf{\Theta}_2)$.
- Log priors: flat in the logarithm of the parameter.
- Natural priors[3] are flat in B and μ. Related via Jacobian factor to flat priors:

$$J = \frac{M_Z}{2}\left|\frac{B}{\mu\tan\beta}\frac{\tan^2\beta - 1}{\tan^2\beta + 1}\right| \tag{4}$$

[3] Allanach *et al*, hep-ph/0705.0487

Priors

- Linear priors (in $\tan\beta$): $\pi(\mathbf{\Theta}_1) = \pi(\mathbf{\Theta}_2)$.
- Log priors: flat in the logarithm of the parameter.
- Natural priors[3] are flat in B and μ. Related via Jacobian factor to flat priors:

$$J = \frac{M_Z}{2}\left|\frac{B}{\mu\tan\beta}\frac{\tan^2\beta - 1}{\tan^2\beta + 1}\right| \tag{4}$$

[3] Allanach *et al*, hep-ph/0705.0487

The Likelihood

- Direct search constraints from LEP/Tevatron.
- Apply Gaussian constraints to experimental observables.

$$\log \mathcal{L}_i = -\frac{\chi_i^2}{2} - \frac{1}{2}\log(2\pi) - \log(\sigma_i)$$

- Except DM relic density $\Omega_{DM}h^2$, $BR(B_s \to \mu^+\mu^-)$ and the Higgs mass m_h

The Likelihood

- Direct search constraints from LEP/Tevatron.
- Apply Gaussian constraints to experimental observables.

$$\log \mathcal{L}_i = -\frac{\chi_i^2}{2} - \frac{1}{2}\log(2\pi) - \log(\sigma_i)$$

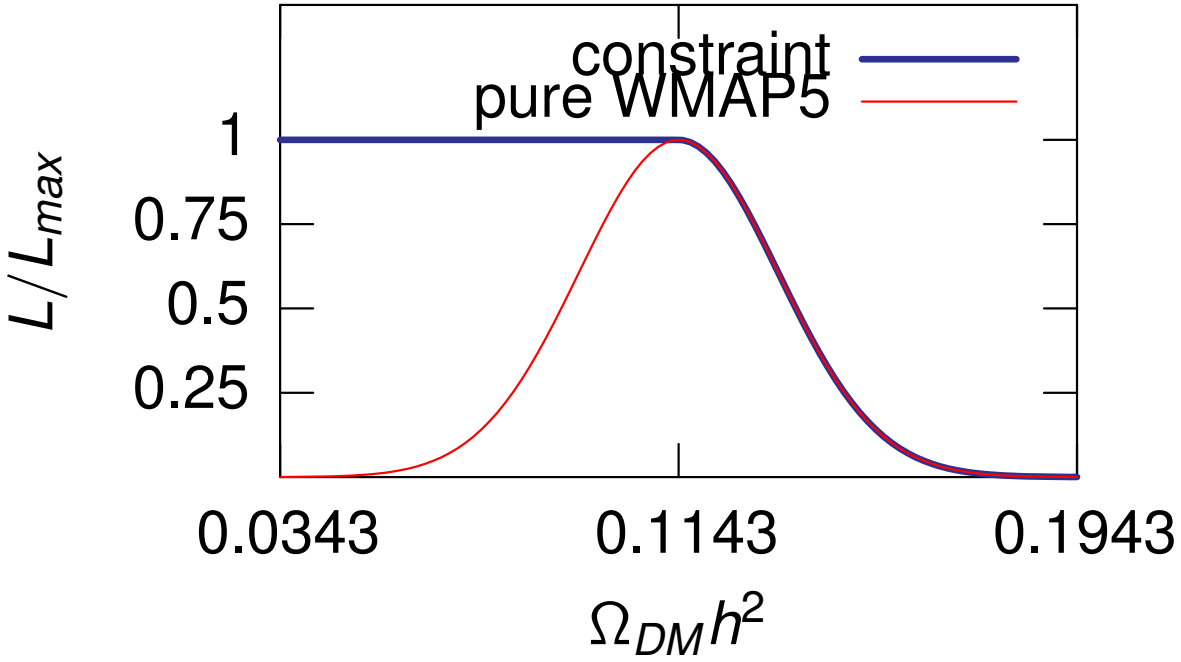
- Except DM relic density $\Omega_{DM}h^2$, $BR(B_s \to \mu^+\mu^-)$ and the Higgs mass m_h

Observables

Observable	Observable
m_W	$BR(B \to X_s\gamma)$
$\sin^2\theta^l_{eff}$	$BR(B_s \to \mu^+\mu^-)$
$(g-2)_\mu$	$BR(B \to D\tau\nu)$
$\Omega_{\rm DM}h^2$	$R_{\Delta M_s}$
m_h	$R_{B\tau\nu}$
Γ_Z^{tot}	Δ_{0-}
R_l^0	R_{l23}
$A^0_{LR}(SLD)$	m_t
R_b^0	m_b
R_c^0	m_Z
$A_{fb}^{0,b}$	$\alpha_s^{\bar{MS}}(M_Z)$
$A_{fb}^{0,c}$	$1/\alpha^{\bar{MS}}$
$\mathcal{A}_b$	$\mathcal{A}_c$

Dark Matter Constraints

- Symmetric DM constraint: Gaussian centered on WMAP5 central value.
- Asymmetric DM constraint: Half Gaussian centred on WMAP5 central value
- No DM constraint.



Preference for $\mu > 0$.

- Table shows odds of likelihood of $\mu > 0$ vs. $\mu < 0$.

	symmetric $\mathcal{L}_{\mathrm{DM}}$		
Model/Prior	flat	log	natural
mSUGRA	3	11	1.5
mAMSB	4	12	1.5
LVS	25	22	13.5

- Preference for $\mu > 0$ weak under natural priors, and moderate under log priors.
- Strongest and most consistent for Large Volume Scenario - the most constrained model.

Model Selection

	symmetric $\mathcal{L}_{\mathrm{DM}}$		
Model/Prior	flat	log	natural
mSUGRA	3000	3000	30000
mAMSB	1.5	2	1
LVS	6000	7300	130,000

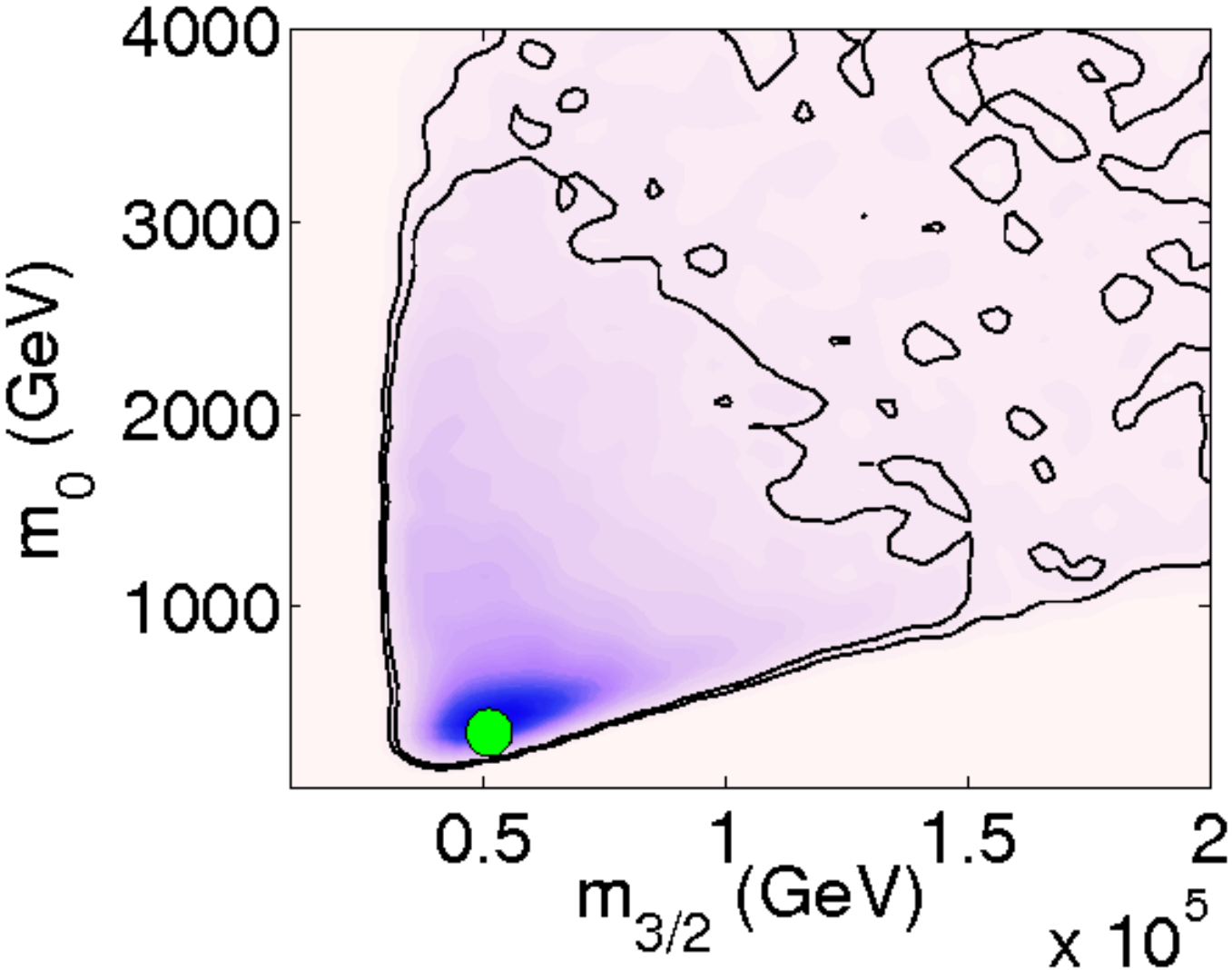
- Normalised to natural priors mAMSB.
- AMSB strongly disfavoured since DM relic density very low due to degenerate wino triplet.

Model Selection

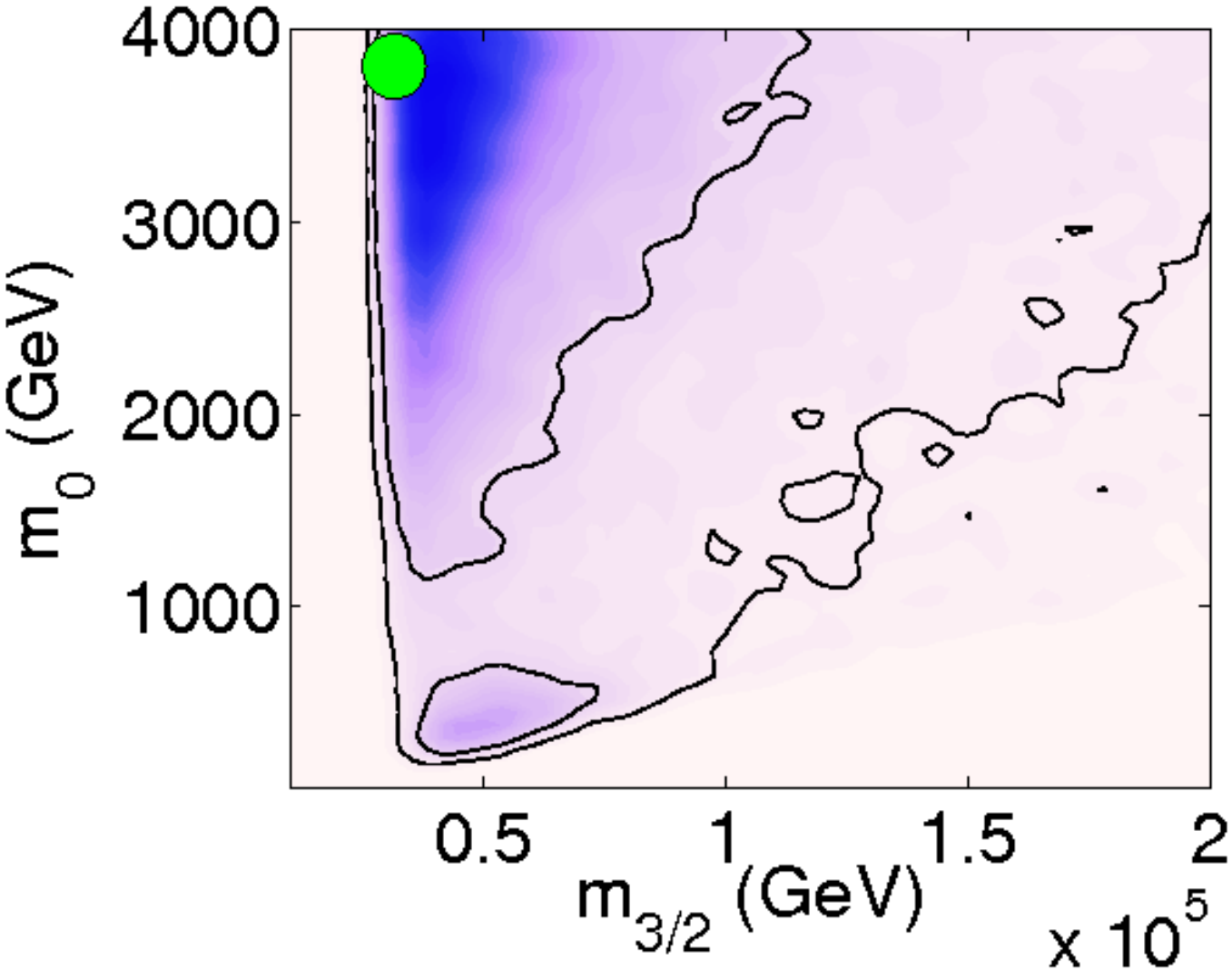
- Asymmetric $\mathcal{L}_{\mathrm{DM}}$ normalised to flat priors mSUGRA.

	symmetric $\mathcal{L}_{\mathrm{DM}}$		
Model/Prior	flat	log	natural
mSUGRA	3000	3000	30000
mAMSB	1.5	2	1
LVS	6000	7300	130,000
	asymmetric $\mathcal{L}_{\mathrm{DM}}$		
mSUGRA	1	3	4
mAMSB	160	400	150
LVS	20	20	22

Prior Dependence: AMSB (Linear Priors)



Prior Dependence: AMSB (Natural Priors)



Model Selection without WMAP

- Log evidence without WMAP bound.
- Normalised to mGMSB with natural priors.
- Fits are dark matter dominated

Model/Prior	flat	log	natural
mSUGRA	2.5	3	3
mAMSB	3	3	18
mGMSB	4	4.5	1
LVS	3	2	4.5

Information Content

- Constraining power of an observable measured using Kullback-Leibler divergence

$$D_{\mathrm{KL}}(p,\pi) = \int p(\mathbf{\Theta}|\mathbf{D},H)\log\frac{p(\mathbf{\Theta}|\mathbf{D},H)}{\pi(\mathbf{\Theta})}d\mathbf{\Theta}. \quad (5)$$

- Quantifies the information gained in going from the prior to the posterior.
- Dark Matter constraint dominates, contributing between around 60% for natural priors, up to 80% for flat priors (in CMSSM).
- Next most important is electroweak observables.
- B-Physics contributions almost entirely given by $BR(B \to s\gamma)$.

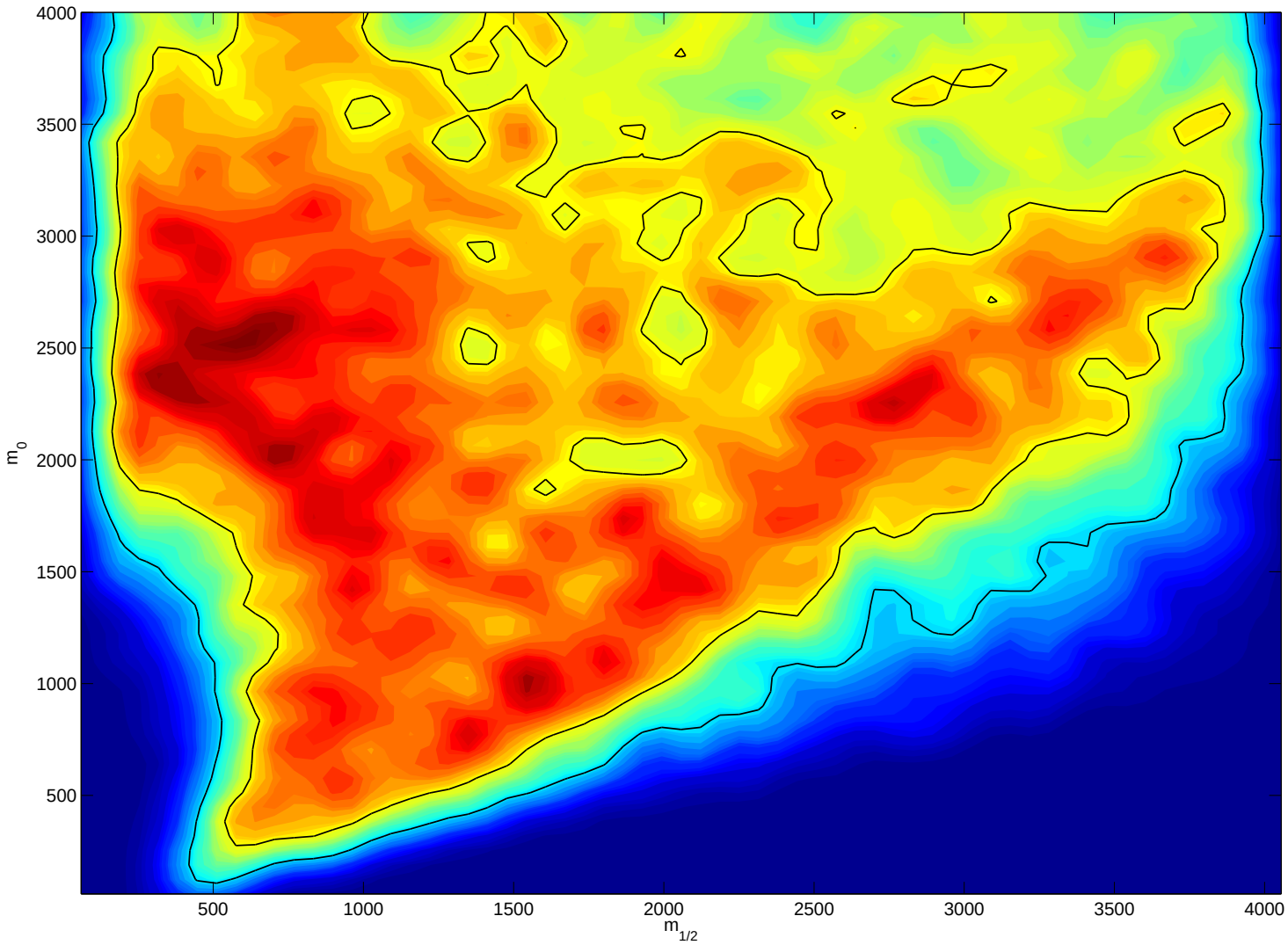
Information Content

- Constraining power of an observable measured using Kullback-Leibler divergence

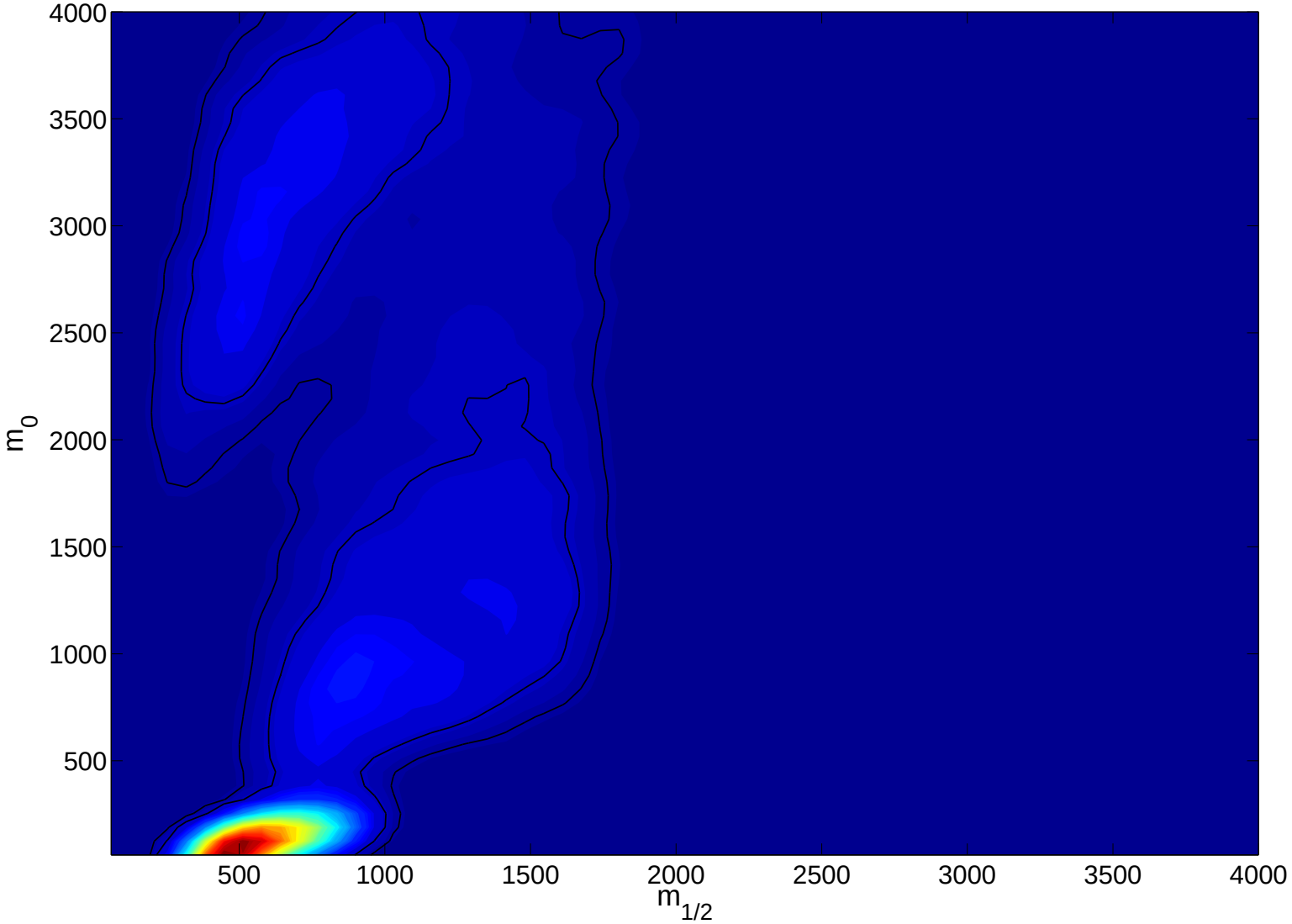
$$D_{\mathrm{KL}}(p,\pi) = \int p(\mathbf{\Theta}|\mathbf{D},H)\log\frac{p(\mathbf{\Theta}|\mathbf{D},H)}{\pi(\mathbf{\Theta})}d\mathbf{\Theta}. \quad (5)$$

- Quantifies the information gained in going from the prior to the posterior.
- Dark Matter constraint dominates, contributing between around 60% for natural priors, up to 80% for flat priors (in CMSSM).
- Next most important is electroweak observables.
- B-Physics contributions almost entirely given by $BR(B \to s\gamma)$.

mSUGRA without WMAP



mSUGRA with symmetric WMAP



Conclusions

- Prior Dependence for models with more than 2D.
- Preference for $\mu > 0$ moderate at best.
- Symmetric $\mathcal{L}_{DM}$: Large Volume.
- Asymmetric $\mathcal{L}_{DM}$: AMSB.
- No DM constraint: Inconclusive.

Bonus slides!

Best Fit Points (Asymmetric $\Omega_{DM}h^2$)

- AMSB: $m_0 = 312$ GeV, $m_{3/2} = 45$ TeV, $\tan\beta = 15.9$.
- GMSB: $\log(M_{mess}) = 7.1$, $\Lambda = 19.5$ TeV, $N_{mess} = 7$, $\tan\beta = 15.9$.
- LVS: $m_0 = \frac{m_{1/2}}{\sqrt{3}} = \frac{-A_0}{\sqrt{3}} = 189.2$ GeV, $\tan\beta = 11.6$.
- mSUGRA: $m_0 = 3338$ GeV, $m_{1/2} = 382$ Gev, $A_0 = 634$ GeV, $\tan\beta = 8.6$.
- MultiNest is optimised to efficiently calculate Bayesian evidence not best-fit points.