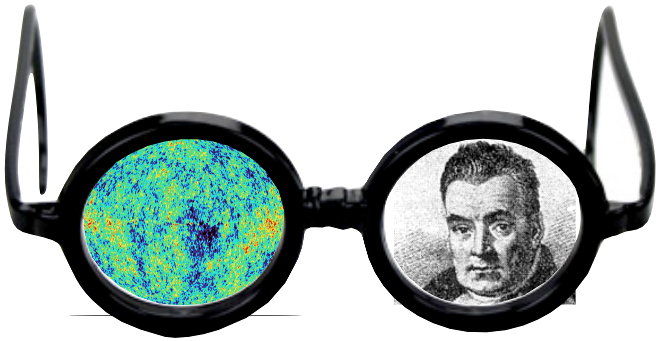
Recent advances in cosmological Bayesian model comparison



Roberto Trotta
Astrophysics, Imperial College London
www.robertotrotta.com

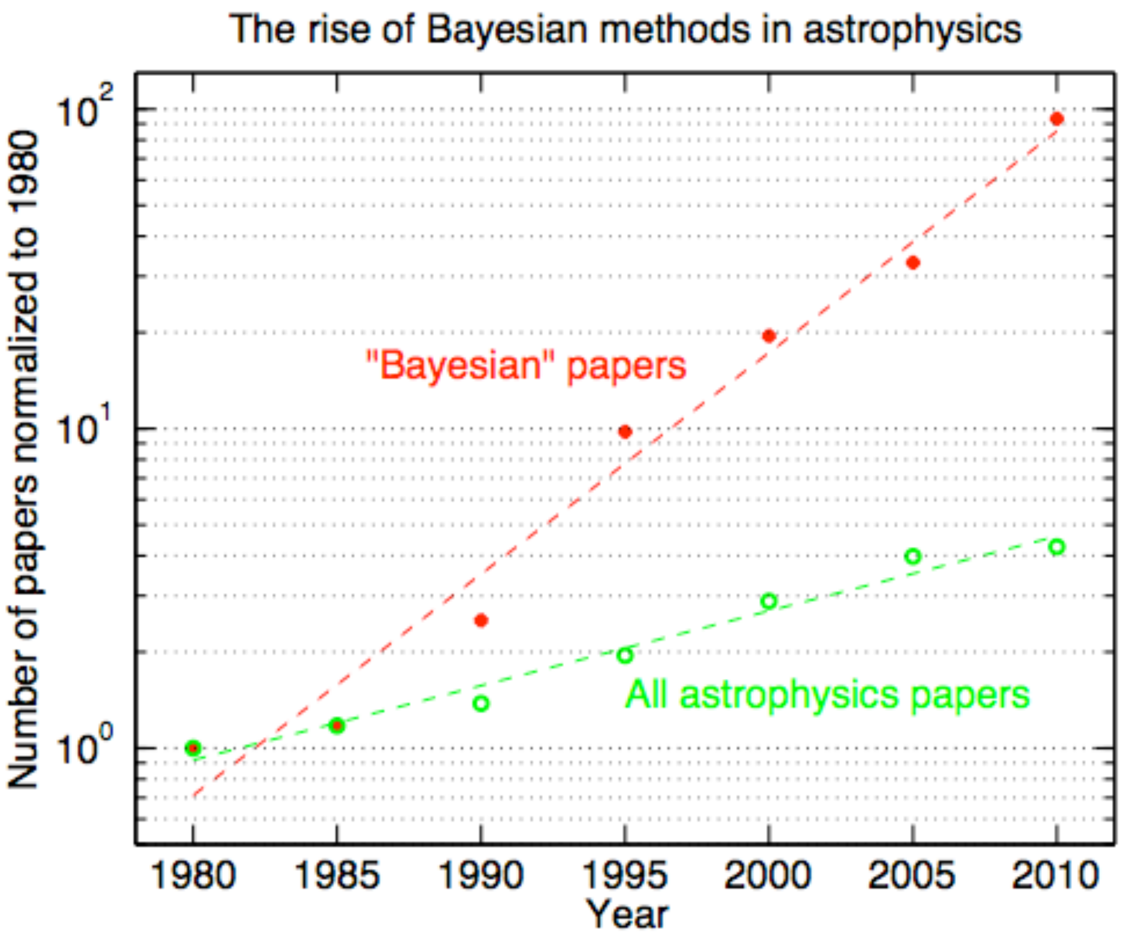
1. What is model comparison?
2. The Bayesian model comparison framework
3. Cosmological applications (curvature, inflation)

Imperial College London

Bayes in the sky

Imperial College London

Review of Bayesian methods in cosmology: Trotta (2008)



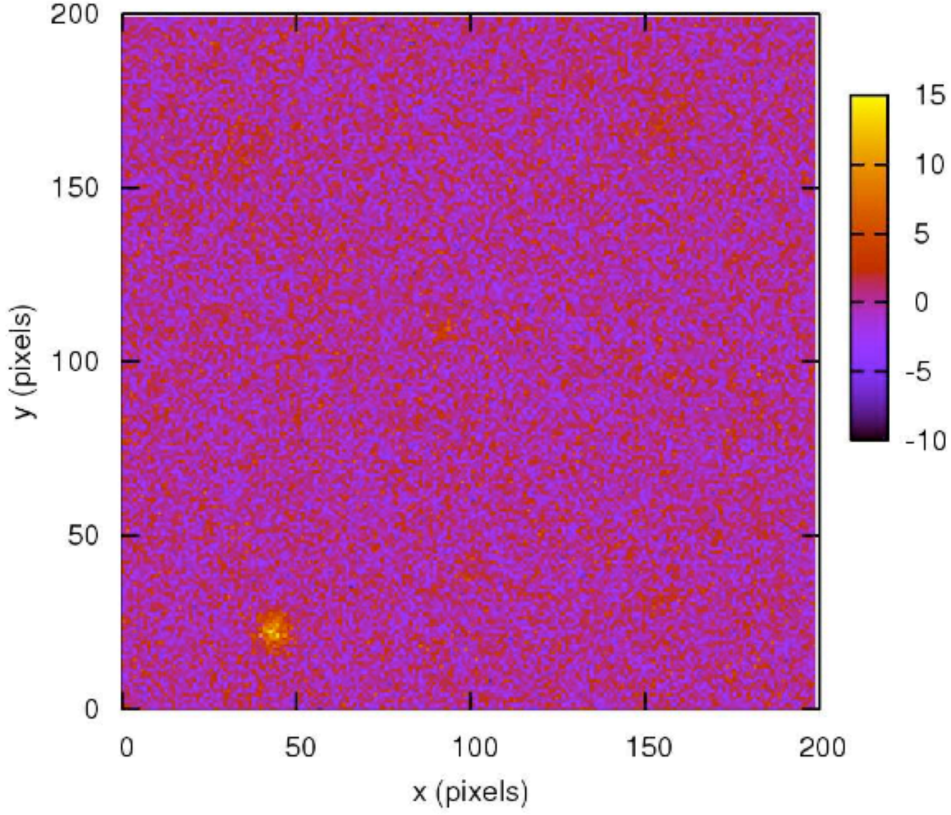
Roberto Trotta

Model comparison: how many sources?

Imperial College London

Feroz and Hobson (2007)

Signal + Noise



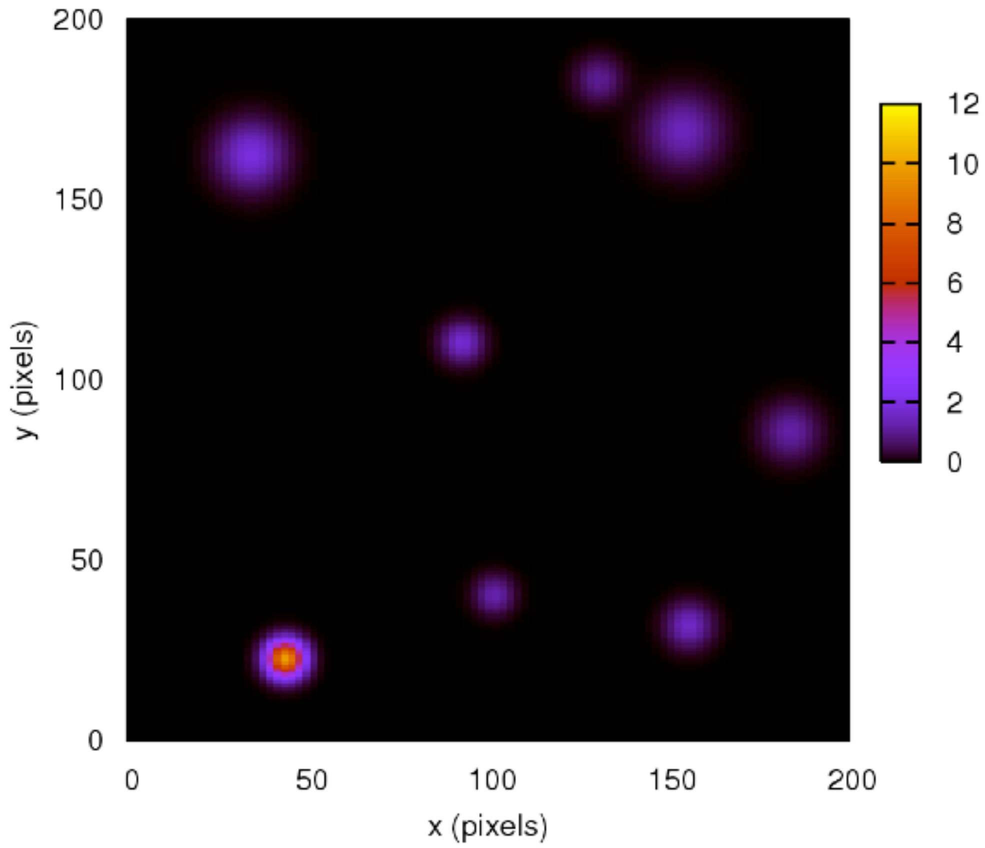
Roberto Trotta

Model comparison: how many sources?

Imperial College London

Feroz and Hobson (2007)

Signal: 8 sources

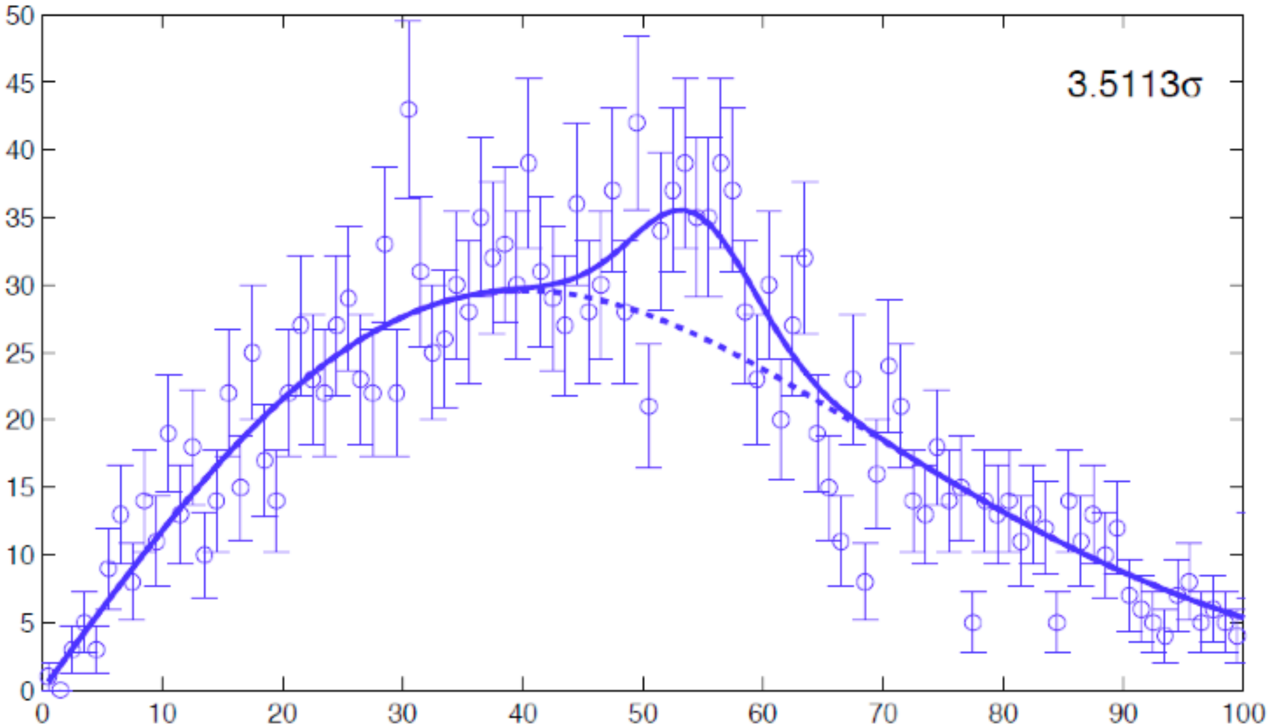


Roberto Trotta

Model comparison: evidence for new physics?

Imperial College London

"Look Elsewhere" effect - see Eilam Gross' talk



Roberto Trotta

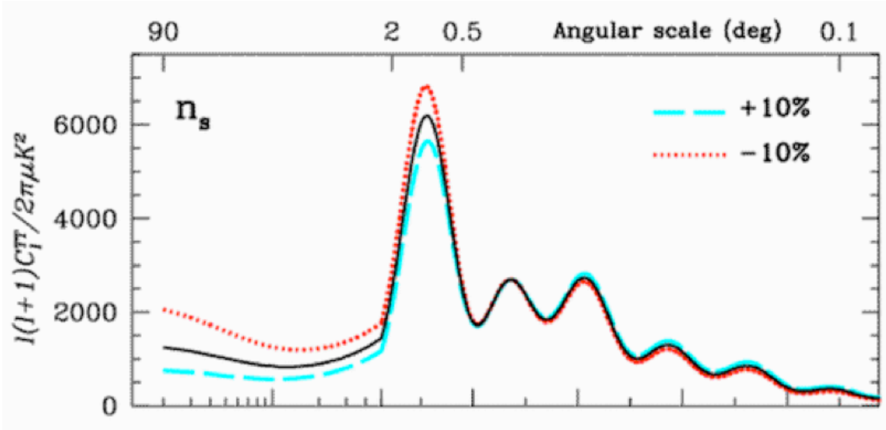
Cosmological model comparison

Imperial College London

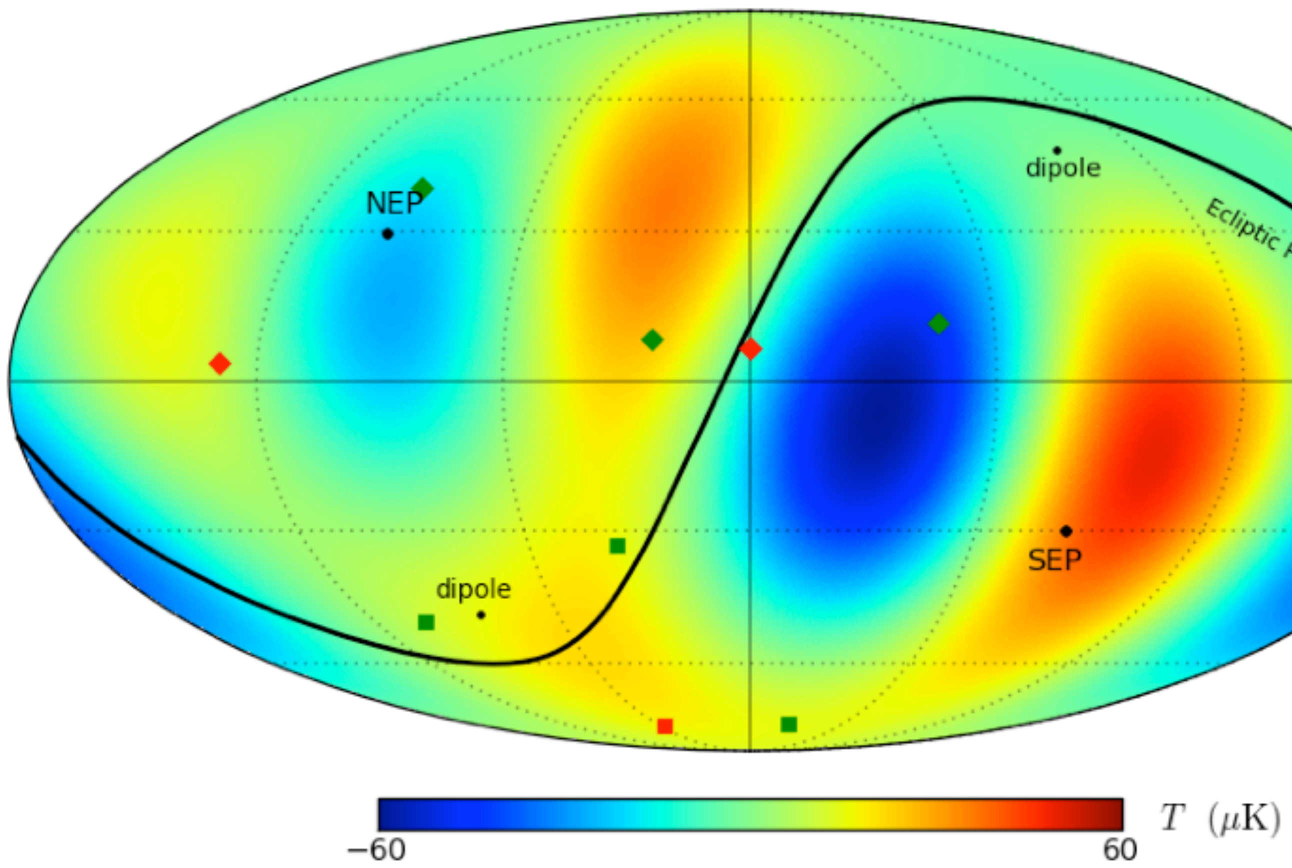
- Is the spectrum of primordial fluctuations scale-invariant (n = 1)?

- Model comparison:
  n = 1 vs n ≠ 1 (with inflation-motivated prior)

- Results:
  n ≠ 1 favoured with odds of 17:1
  (Trotta 2007)
  n ≠ 1 favoured with odds of 15:1
  (Kunz, Trotta & Parkinson 2007)
  n ≠ 1 favoured with odds of 7:1
  (Parkinson 2007 et al 2006)

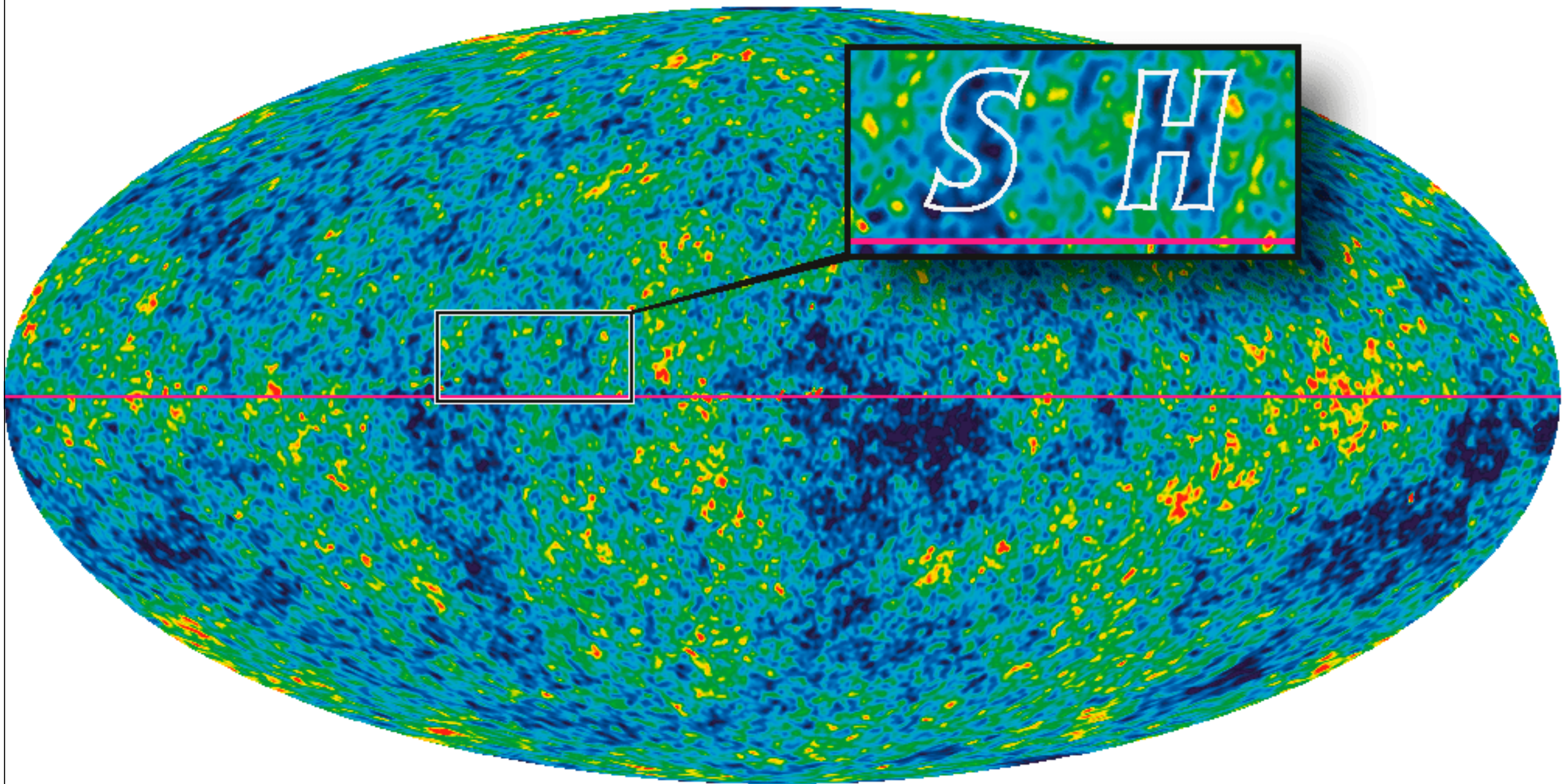


Large scale CMB anomalies



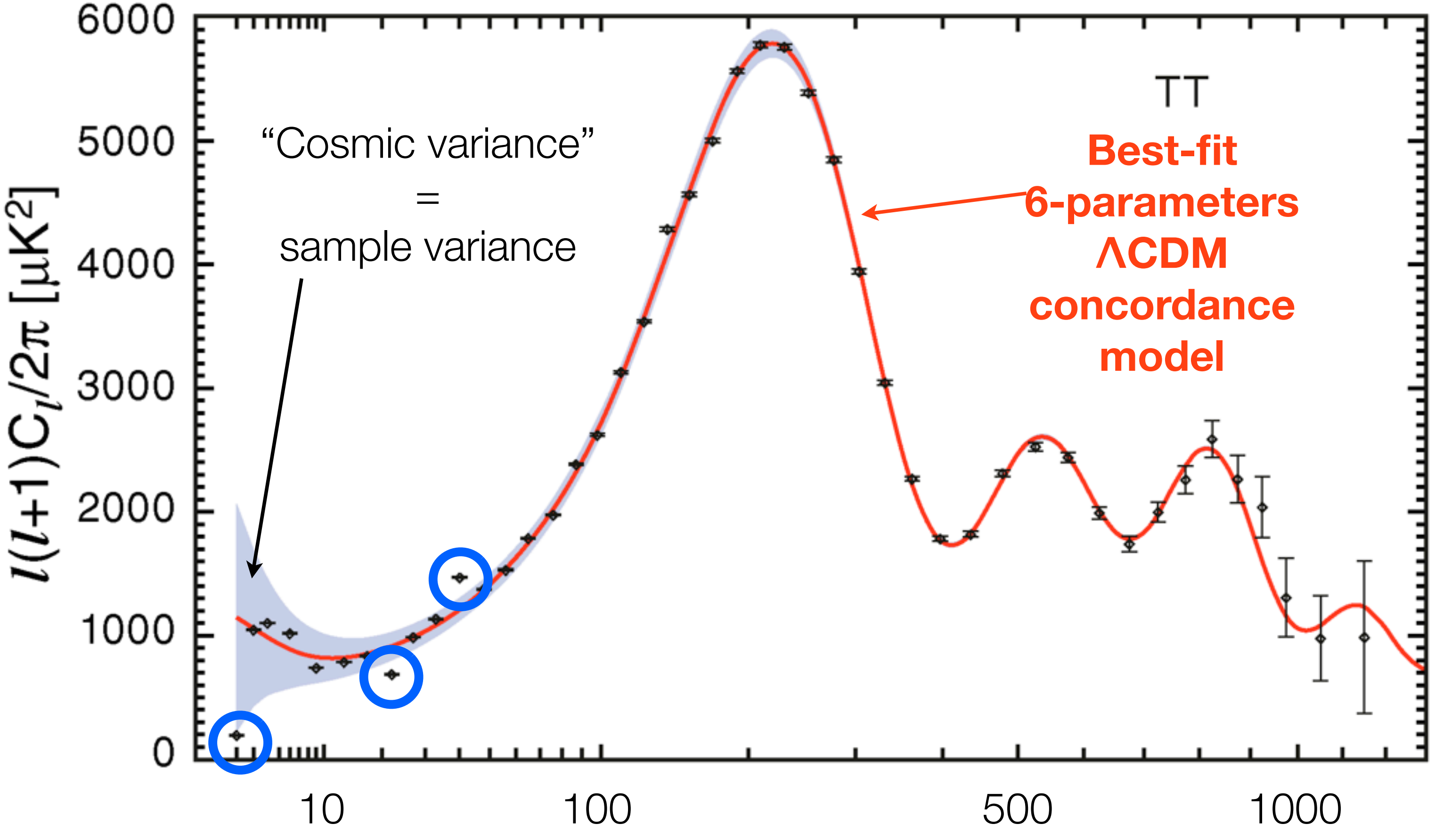
Copi et al (2010)

The “SH” initials of Stephen Hawking are shown in the ILC sky map.



WMAP team

WMAP 7-years temperature power spectrum

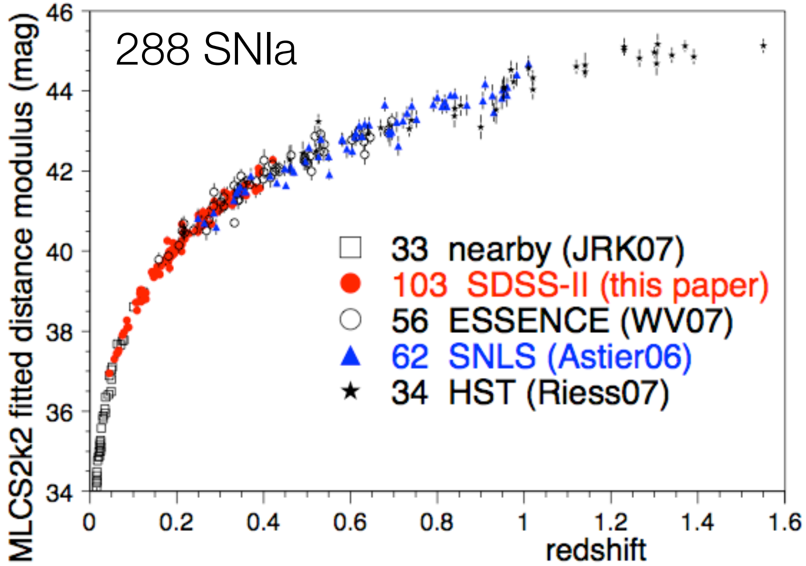


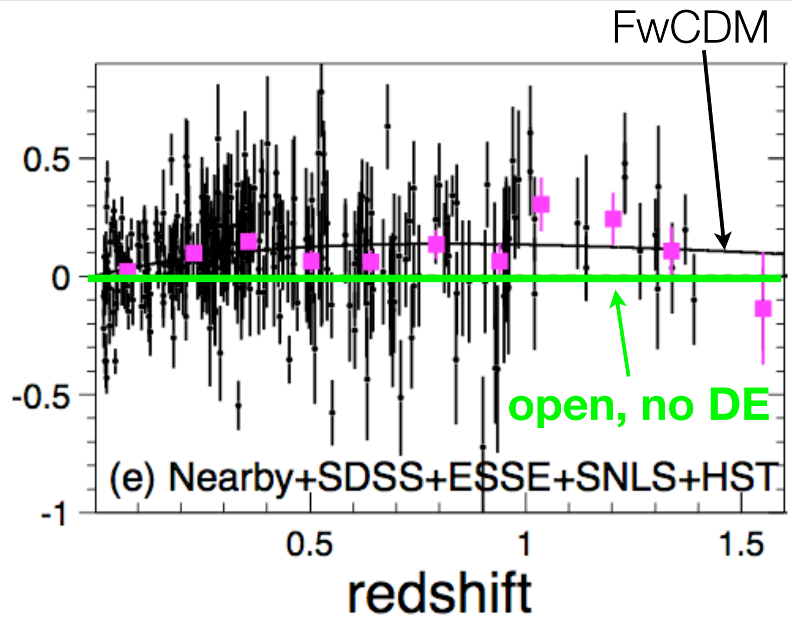
Jarosik et al (2010)

Additional data from SNIa and BAO

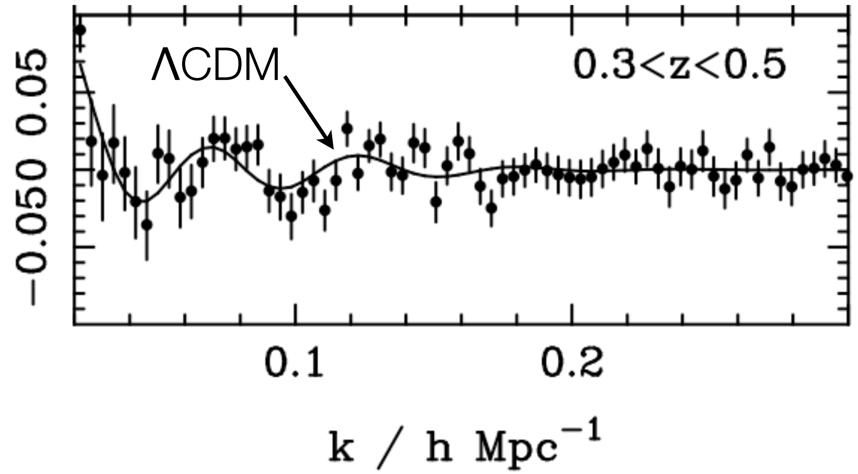
Imperial College London

Kessler et al (SDDS collaboration) (2010)





Percival et al (2010)



Baryonic wiggles in SDSS DR7

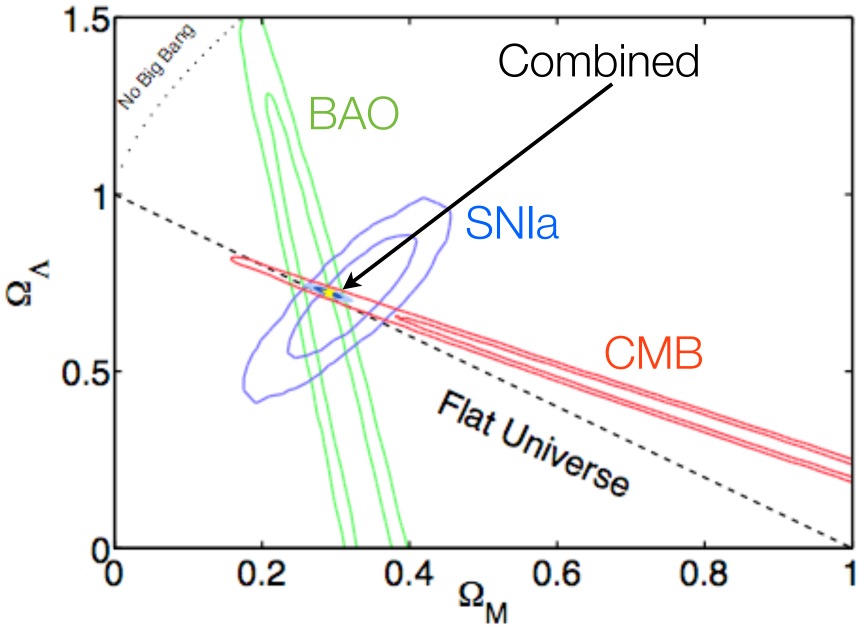
Roberto Trotta

Putting it all together... precision cosmology!

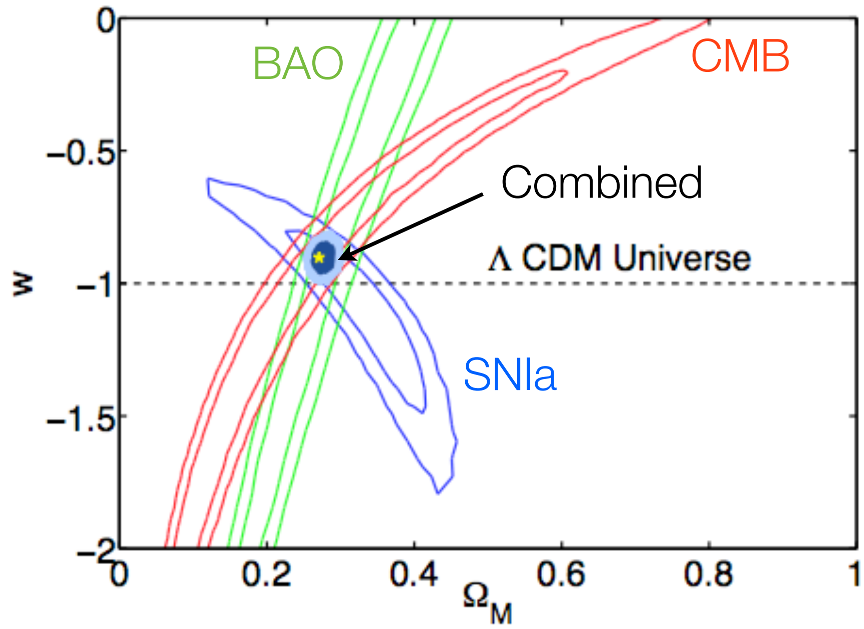
Imperial College London

- Combined cosmological constraints on matter and dark energy content:

Assuming Λ



Assuming flatness



March, RT et al (2011)

The Bayesian framework

Bayes' Theorem:
The Equation of Knowledge

Imperial College London

posterior | likelihood | prior

$$P(\theta|d, I) = \frac{P(d|\theta, I)P(\theta|I)}{P(d|I)}$$

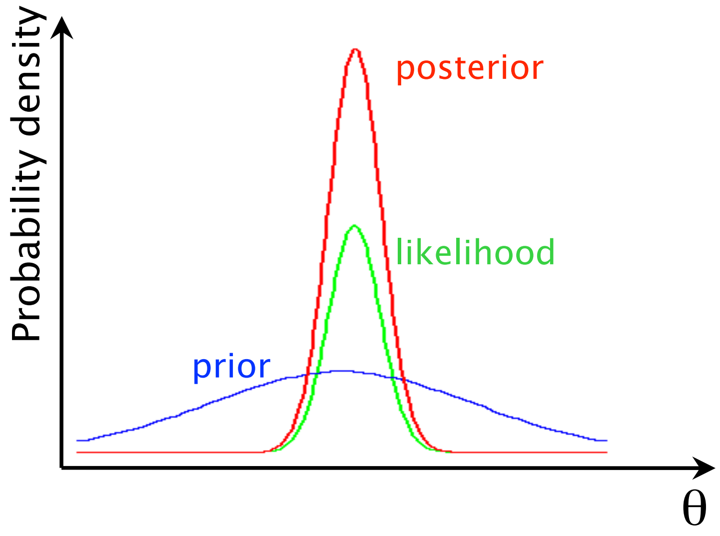
evidence

θ: parameters
d: data
I: any other external information, or the assumed model

For parameter inference it is sufficient to consider

$$P(\theta|d, I) \propto P(d|\theta, I)P(\theta|I)$$

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$



The 3 levels of inference

Imperial College London

LEVEL 1
I have selected a model M and prior P(θ|M)

Parameter inference
(assumes M is the true model)

$$P(\theta|d,M) = \frac{P(d|\theta,M)P(\theta|M)}{P(d|M)}$$

LEVEL 2
Actually, there are several possible models: M_0, M_1,...

Model comparison
What is the relative plausibility of M_0, M_1,... in light of the data?

$$\mathrm{odds} = \frac{P(M_0|d)}{P(M_1|d)}$$

LEVEL 3
None of the models is clearly the best

Model averaging
What is the inference on the parameters accounting for model uncertainty?

$$P(\theta|d) = \sum_i P(M_i|d)P(\theta|d,M_i)$$

The many uses of Bayesian model comparison

Imperial College London

ASTROPHYSICS

Exoplanets detection

Is there a line in this spectrum?

Is there a source in this image?

Cross-matching of sources

COSMOLOGY

Is the Universe flat?

Does dark energy evolve?

Are there anomalies in the CMB?

Which inflationary model is best?

Is there evidence for modified gravity?

Are the initial conditions adiabatic?

Many scientific questions are of the model comparison type

ASTROPARTICLE

Gravitational waves detection

Do cosmic rays correlate with AGNs?

Dark matter signals

Level 2: model comparison

Imperial College London

$$P(\theta|d,M) = \frac{P(d|\theta,M)P(\theta|M)}{P(d|M)}$$

Bayesian evidence or model likelihood

The evidence:

$$P(d|M) = \int_\Omega d\theta P(d|\theta,M)P(\theta|M)$$

The model's posterior:

$$P(M|d) = \frac{P(d|M)P(M)}{P(d)}$$

When comparing two models:

$$\frac{P(M_0|d)}{P(M_1|d)} = \frac{P(d|M_0)}{P(d|M_1)} \frac{P(M_0)}{P(M_1)}$$

The Bayes factor:

$$B_{01} \equiv \frac{P(d|M_0)}{P(d|M_1)}$$

Posterior odds = Bayes factor × prior odds

Roberto Trotta

Scale for the strength of evidence

Imperial College London

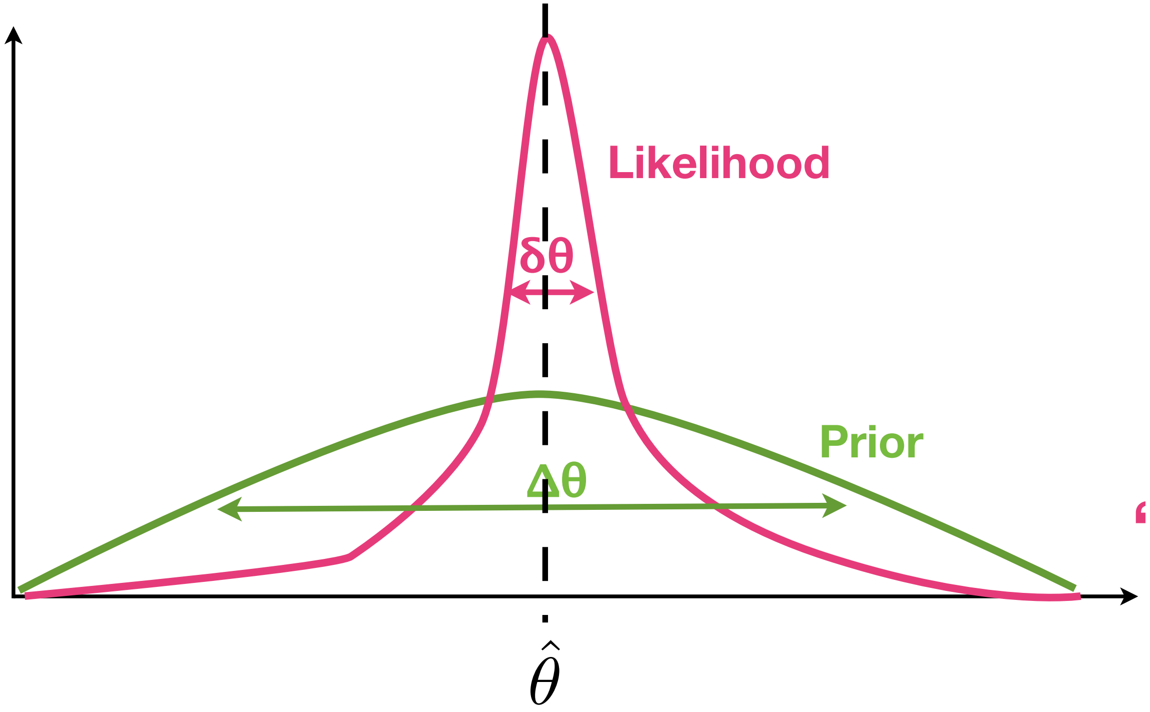
- A (slightly modified) Jeffreys' scale to assess the strength of evidence (**Notice:** this is empirically calibrated!)

$\|\ln B\|$	relative odds	favoured model's probability	Interpretation
< 1.0	$< 3:1$	< 0.750	not worth mentioning
< 2.5	$< 12:1$	0.923	weak
< 5.0	$< 150:1$	0.993	moderate
> 5.0	$> 150:1$	> 0.993	strong

An in-built Occam's razor

Imperial College London

- Bayes factor balances quality of fit vs extra model complexity.
- It rewards highly predictive models, penalizing "wasted" parameter space



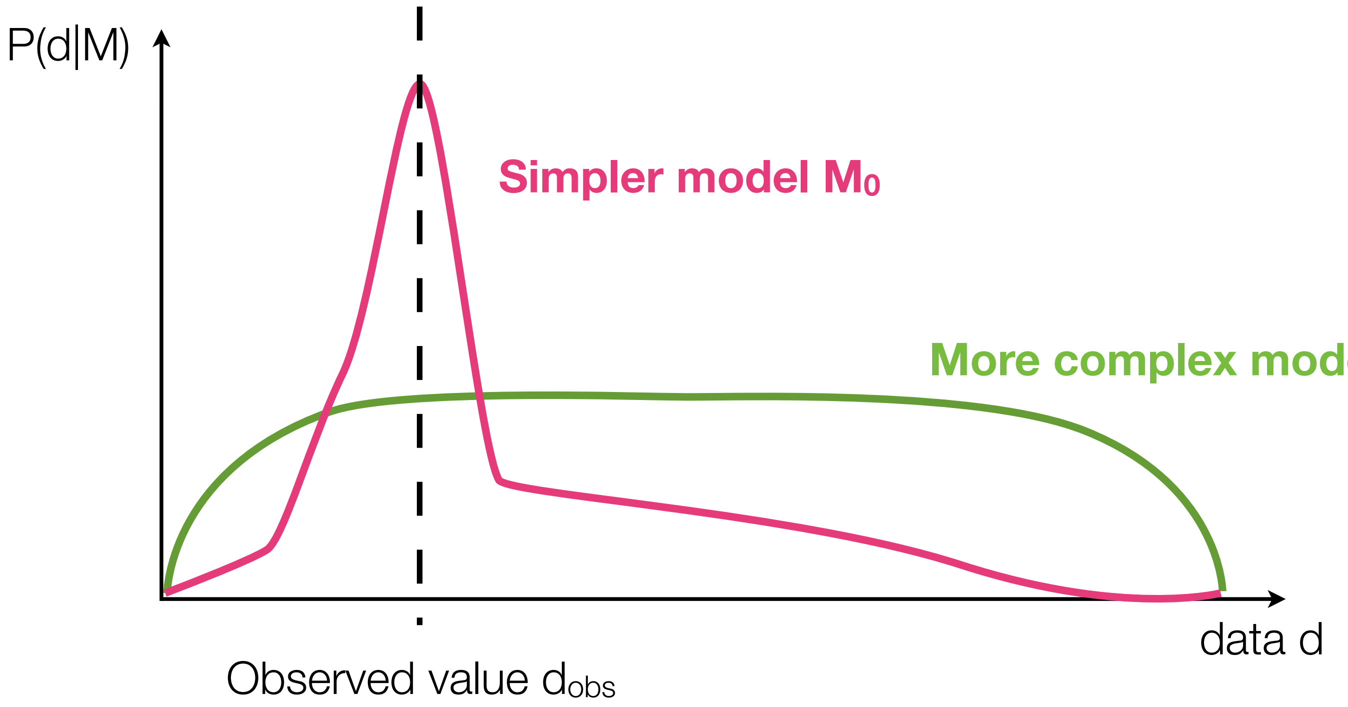
$$P(d|M) = \int d\theta L(\theta) P(\theta|M) \approx L(\hat{\theta}) \delta\theta P(\hat{\theta}) \approx \frac{\delta\theta}{\Delta\theta} L(\hat{\theta})$$

"Occam's factor" — Quality of fit

The evidence as predictive probability

Imperial College London

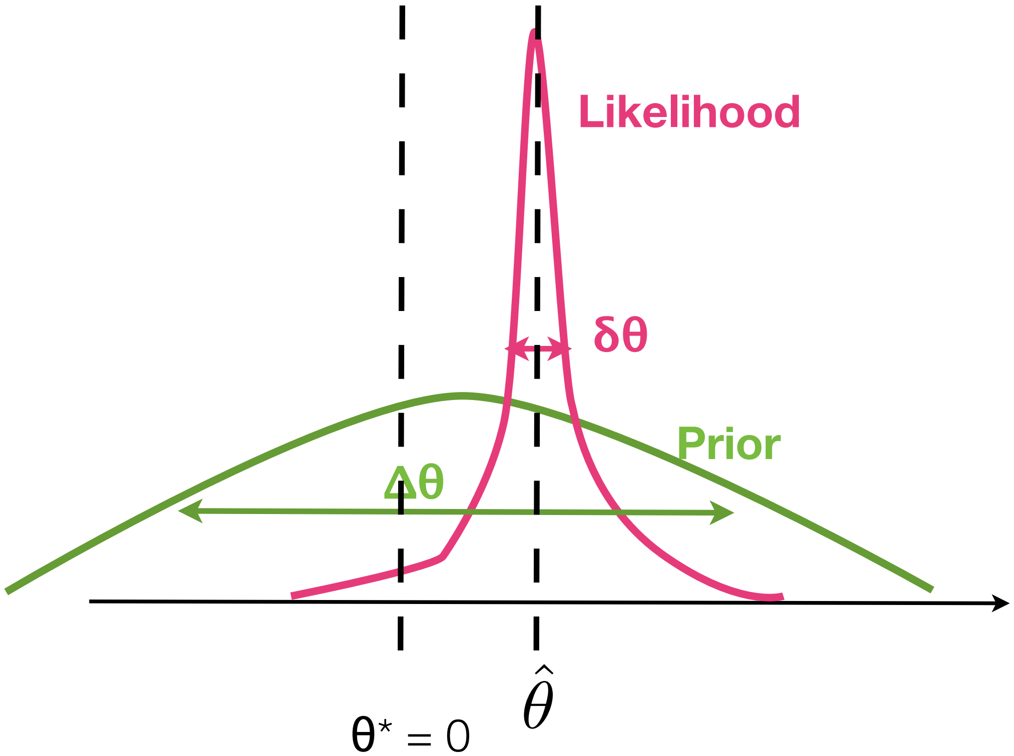
- The evidence can be understood as a function of d to give the predictive probability under the model M:



Model comparison for nested models

- This happens often in practice: we have a more complex model, M_1 with prior $P(\theta|M_1)$, which reduces to a simpler model (M_0) for a certain value of the parameter, e.g. $\theta = \theta^* = 0$ (**nested models**)

- **Is the extra complexity of M_1 warranted by the data?**



Model comparison for nested models

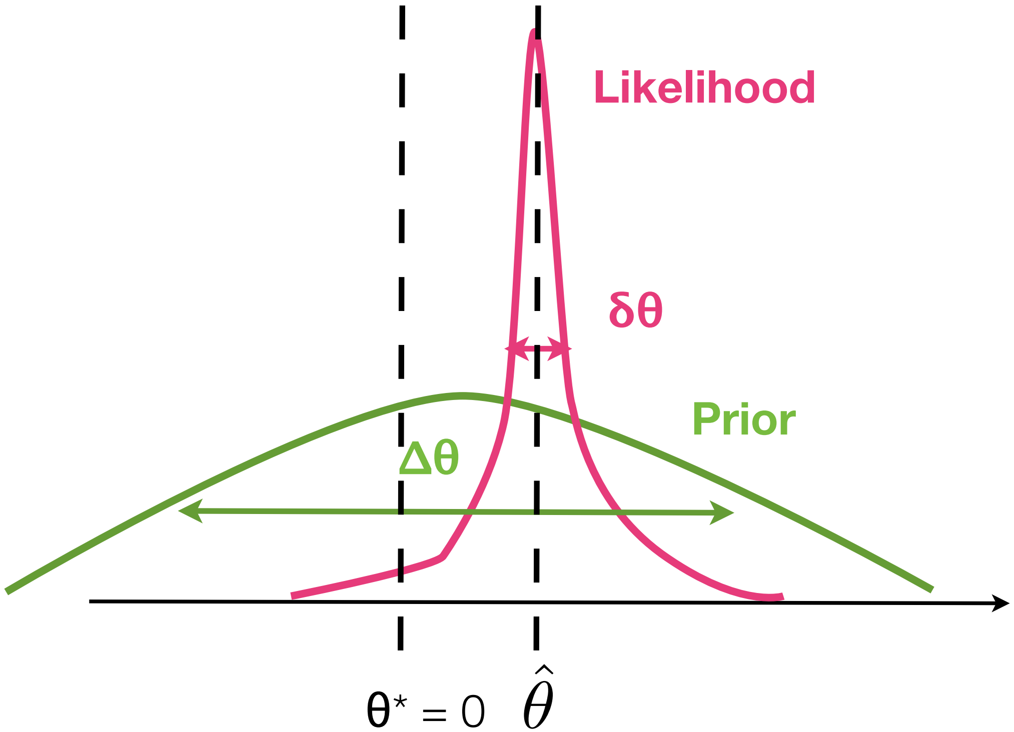
Define: $\lambda \equiv \frac{\hat{\theta}-\theta^\star}{\delta\theta}$

For "informative" data:

$$\ln B_{01} \approx \ln \frac{\Delta\theta}{\delta\theta} - \frac{\lambda^2}{2}$$

"wasted" parameter space (favours simpler model)

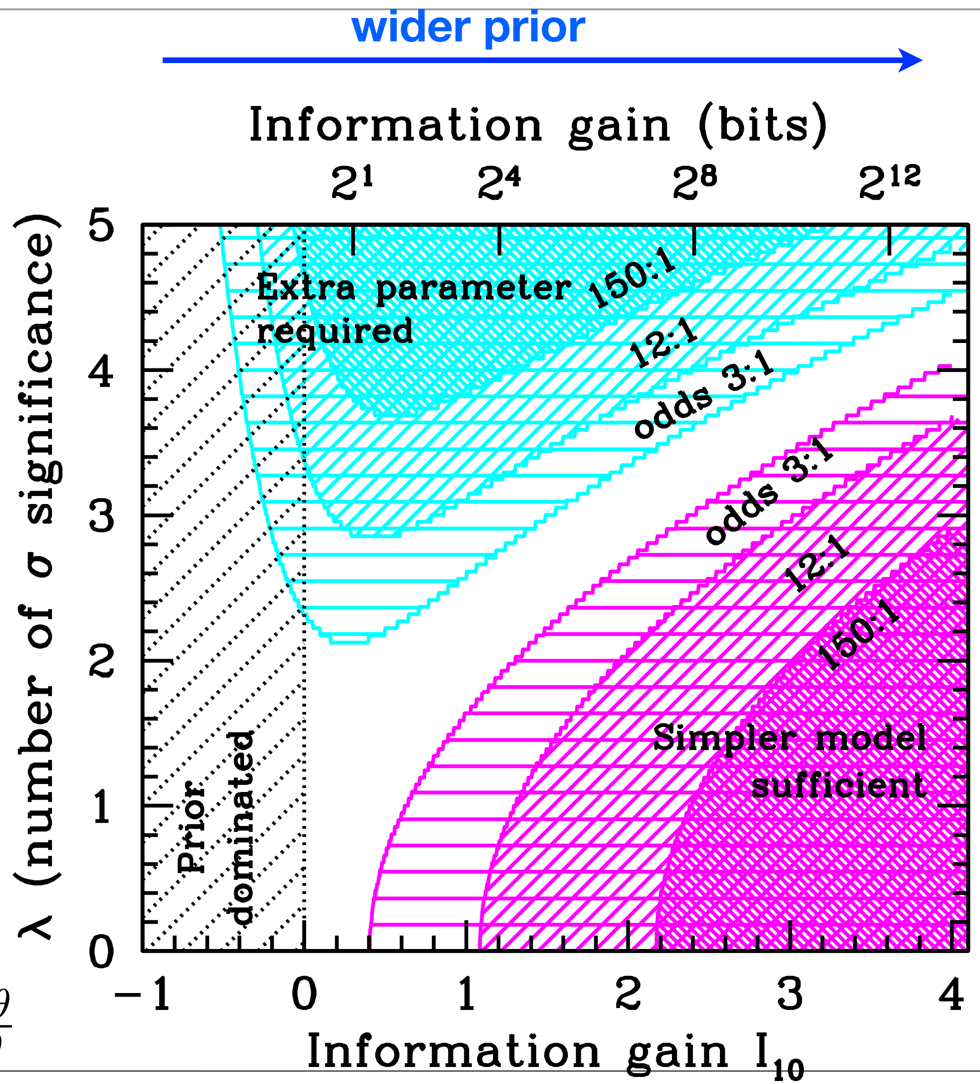
mismatch of prediction with observed data (favours more complex model)



The rough guide to model comparison

Imperial College London

Trotta (2007)



$$I_{10} \equiv \log_{10} \frac{\Delta\theta}{\delta\theta}$$

Roberto Trotta

About frequentist hypothesis testing

Imperial College London

- **Warning:** frequentist hypothesis testing **cannot be interpreted as a statement about the probability of the hypothesis!**
- **Example:** to reject the null hypothesis H_0: $\theta = 0$, draw *n* normally distributed points (with known variance σ^2). The χ^2 is distributed as a chi-square distribution with *(n-1)* degrees of freedom (dof). Pick a significance level α (or p-value, e.g. $\alpha = 0.05$). If $P(\chi^2 > \chi^2_{obs}) < \alpha$ reject the null hypothesis.
- This is a statement about the probability of observing data as extreme or more extreme than have been measured *assuming the null hypothesis is correct*.
- **It is not a statement about the probability of the null hypothesis itself and cannot be interpreted as such! (or you'll make gross mistakes)**
- *The use of p-values implies that a hypothesis that may be true can be rejected because it has not predicted observable results that have not actually occurred.* (Jeffreys, 1961)

Assessing hypotheses

- The fundamental mistake is to confuse:

P(data|hypothesis) ≠ P(hypothesis|data)

p-value, frequentist

Assumes hypothesis to be true. Rejected if data improbable under the null (so what?)

Requires Bayes' Theorem

This is typically the question we are interested in!

Highly recommended: Sellke, Bayarri & Berger, *The American Statistician*, 55, 1 (2001)

Assessing hypotheses

P(data|hypothesis) ≠ P(hypothesis|data)

Example:

Hypothesis (H): is a random person female (H=F or H=M)?

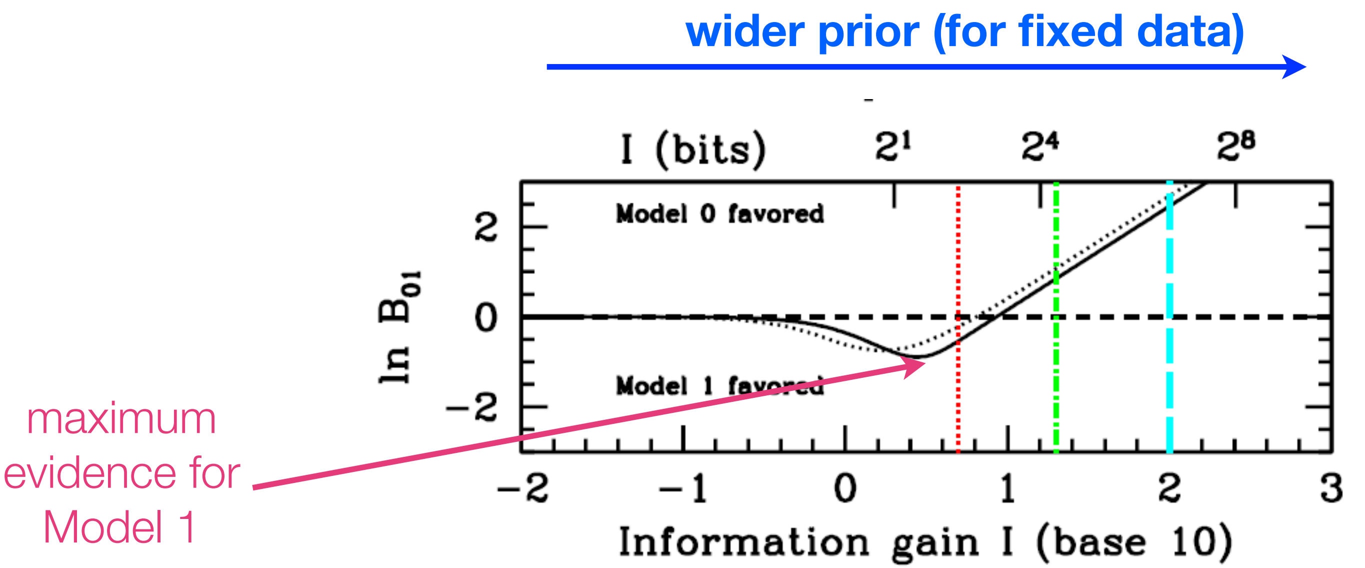
Observation (data): is the person pregnant? (D = Y)

Caution: P(D=Y|H=F) = 0.03
but
P(H=F|D=Y) >> 0.03

Prior-free evidence bounds

Imperial College London

- What if we do not know how to set the prior?
- Then our physical theory is probably not good enough! (e.g., dark energy, inflationary potentials)
- E.g.: for nested models, we can still choose a prior that will maximise the support for the more complex model:



Roberto Trotta

Maximum evidence for a detection

Imperial College London

- **The absolute upper bound:** put all prior mass for the alternative onto the observed maximum likelihood value. Then

$$B < \exp(-\chi^2/2)$$

- **More reasonable class of priors:** symmetric and unimodal around Ψ=0, then (α = significance level)

$$B < \frac{-1}{\exp(1)\alpha \ln \alpha}$$

If the upper bound is small, no other choice of prior will make the extra parameter significant.

Gordon & Trotta (2007)

How to interpret the "number of sigma's"

Imperial College London

α	sigma	Absolute bound on lnB (B)	"Reasonable" bound on lnB (B)
0.05	2.0	2.0 (7:1) weak	0.9 (3:1) undecided
0.003	3.0	4.5 (90:1) moderate	3.0 (21:1) moderate
0.0003	3.6	6.48 (650:1) strong	5.0 (150:1) strong

Numerical evaluation of the Bayesian evidence

Computing the evidence

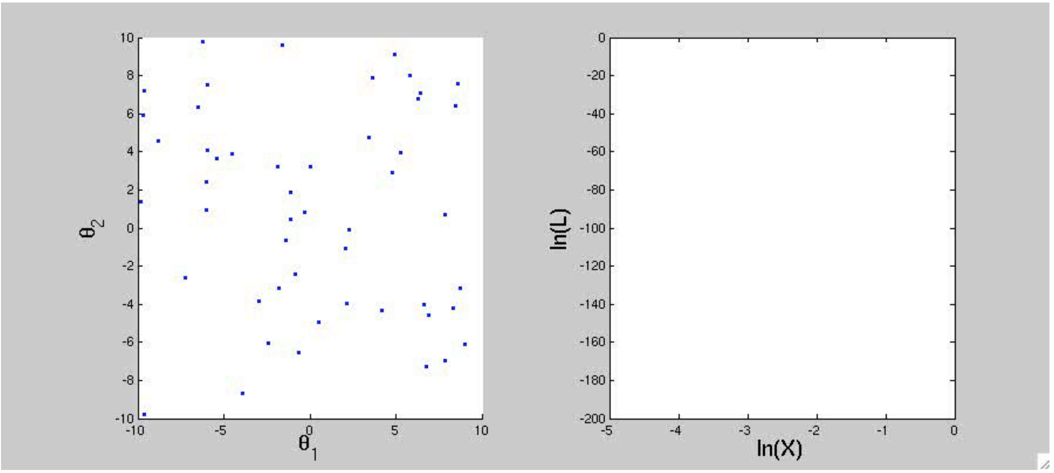
Imperial College London

Evidence: $P(d|M) = \int_\Omega d\theta P(d|\theta, M)P(\theta|M)$

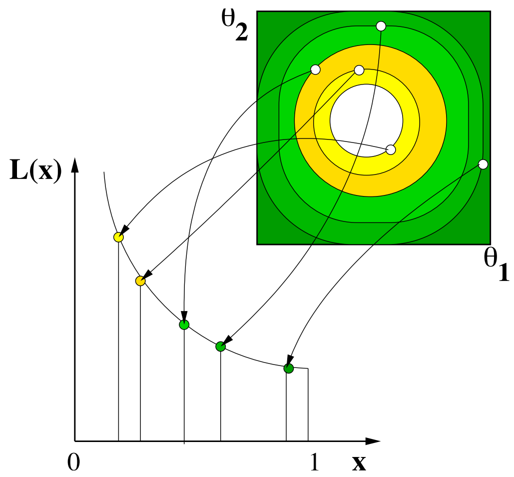
Bayes factor: $$B_{01} \equiv \frac{P(d|M_0)}{P(d|M_1)}$$

- Usually a computational demanding multi-dimensional integral!
- Several numerical/semi-analytical techniques available:
  - **Thermodynamic integration** or **Population Monte Carlo**
  - **Laplace approximation:** approximate the likelihood to second order around maximum gives Gaussian integrals (for normal prior). Can be inaccurate.
  - **Savage-Dickey density ratio:** good for nested models, gives the Bayes factor
  - **Nested sampling:** clever & efficient, can be used generally

The “Nested Sampling” algorithm



(animation courtesy of David Parkinson)



An algorithm to simplify the computation of the Bayesian evidence (Skilling, 2006):

$$X(\lambda) = \int_{\mathcal{L}(\theta)>\lambda} P(\theta)d\theta$$

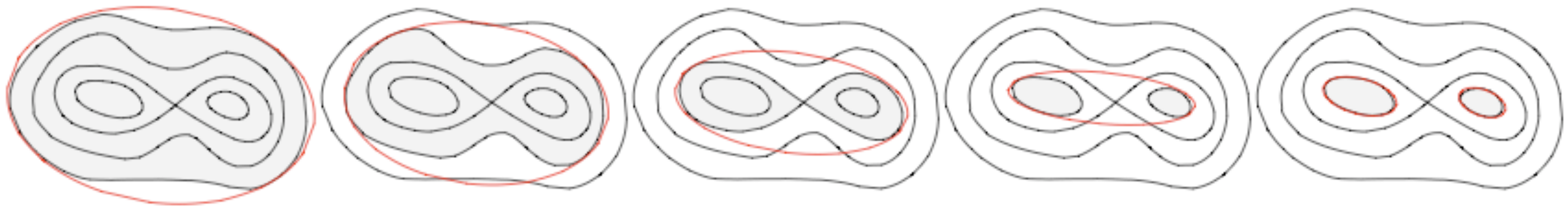
$$P(d) = \int d\theta \mathcal{L}(\theta)P(\theta) = \int_0^1 X(\lambda)d\lambda$$

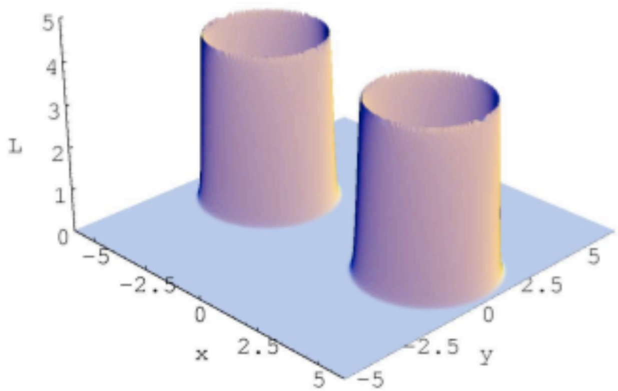
Feroz et al (2008), *arxiv: 0807.4512* Trotta et al (2008), *arxiv: 0809.3792*

The MultiNest algorithm

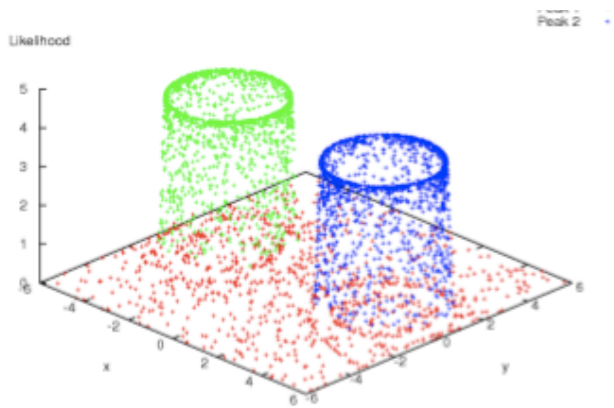
Imperial College London

- MultiNest: Also an extremely efficient sampler for multi-modal likelihoods!
  Feroz & Hobson (2007), RT et al (2008), Feroz et al (2008)





Target likelihood

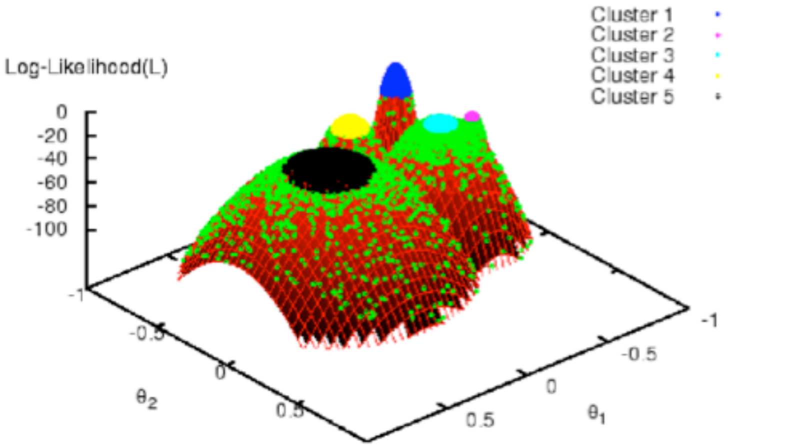


Sampled likelihood

Computation of the evidence with Multinest

Imperial College London

Feroz and Hobson (2007)



Gaussian mixture model:

True evidence: log(E) = -5.27

Multinest:

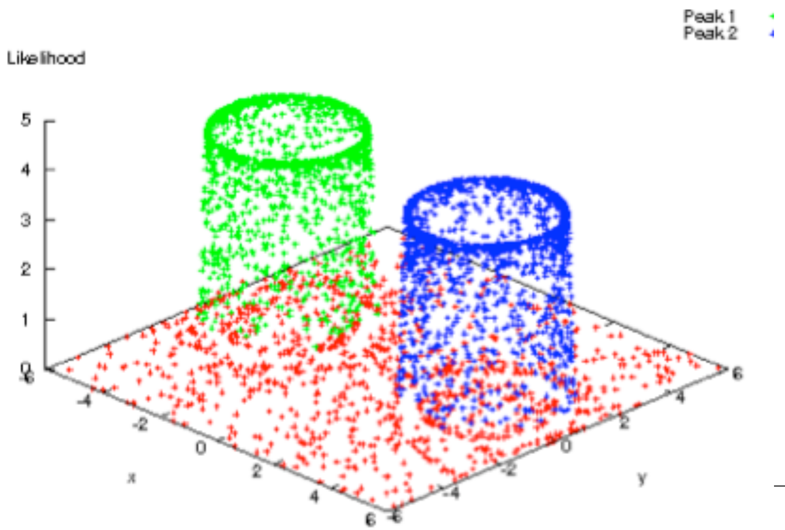
Reconstruction: log(E) = -5.33 ± 0.11

Likelihood evaluations ~ 10^4

Thermodynamic integration:

Reconstruction: log(E) = -5.24 ± 0.12

Likelihood evaluations ~ 10^6



Courtesy Mike Hobson

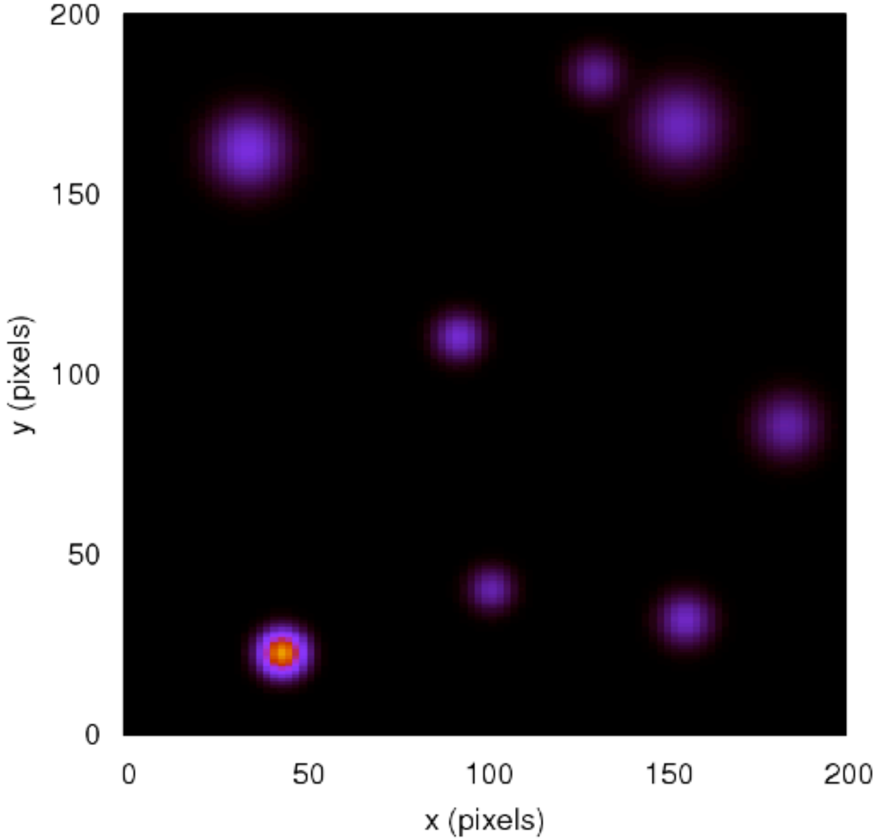
D	N_{like}	efficiency	likes per dimension
2	7000	70%	83
5	18000	51%	7
10	53000	34%	3
20	255000	15%	1.8
30	753000	8%	1.6

Roberto Trotta

MultiNest applied to object detection

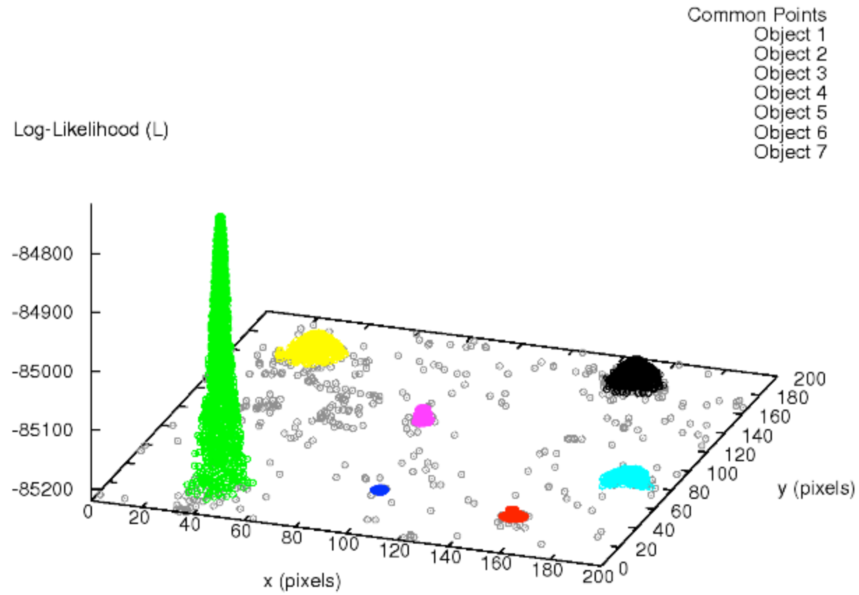
Imperial College London

Feroz and Hobson (2007)



Bayesian reconstruction

7 out of 8 objects correctly identified.
Confusion happens because 2 objects very close.



Roberto Trotta

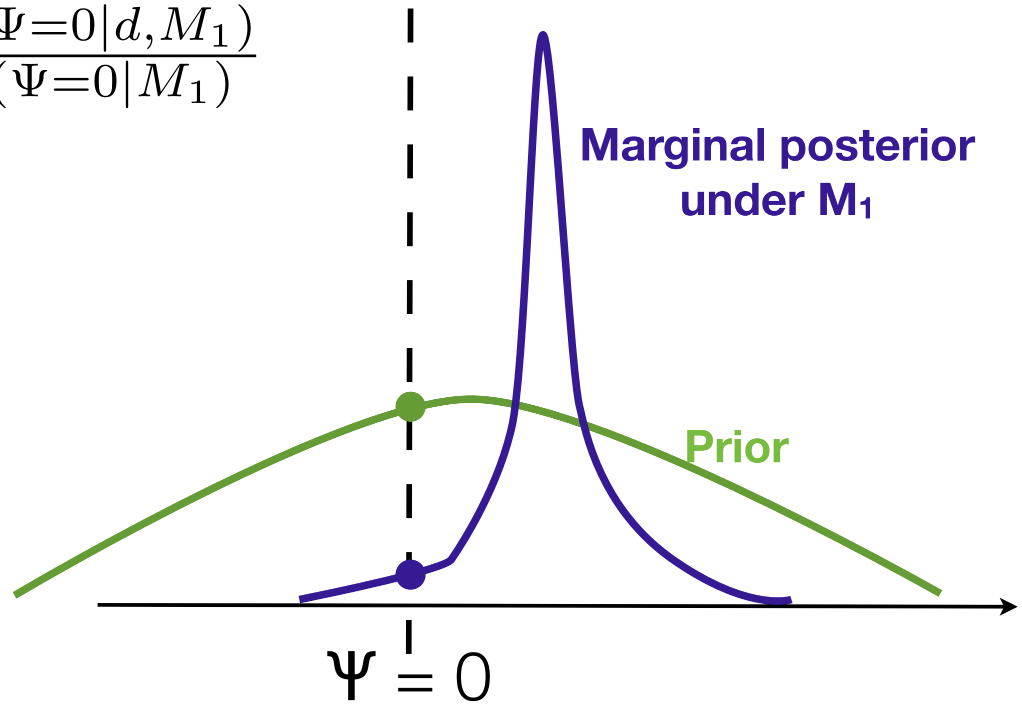
The Savage-Dickey density ratio

Imperial College London

- This methods works for nested models and gives the Bayes factor analytically.
- **Assumptions:** nested models (M_1 with parameters θ,Ψ reduces to M_0 for e.g. $\Psi=0$) and separable priors (i.e. the prior $P(\theta,\Psi|M_1)$ is uncorrelated with $P(\theta|M_0)$)
- Result:

$$B_{01} = \frac{P(\Psi=0|d,M_1)}{P(\Psi=0|M_1)}$$

- **Advantages**:
  - analytical
  - often accurate
  - clarifies the role of prior
  - does not rely on Gaussianity



Roberto Trotta

Cosmological applications

Cosmological model building: results

Imperial College London

Competing model	ΔN_{par}	ln B	Ref	Data	Outcome
Initial conditions Isocurvature modes					
CDM isocurvature	+1	−7.6	[58]	WMAP3+, LSS	Strong evidence for adiabaticity
+ arbitrary correlations	+4	−1.0	[46]	WMAP1+, LSS, SN Ia	Undecided
Neutrino entropy	+1	$[-2.5, -6.5]^p$	[60]	WMAP3+, LSS	Moderate to strong evidence for adiabaticity
+ arbitrary correlations	+4	−1.0	[46]	WMAP1+, LSS, SN Ia	Undecided
Neutrino velocity	+1	$[-2.5, -6.5]^p$	[60]	WMAP3+, LSS	Moderate to strong evidence for adiabaticity
+ arbitrary correlations	+4	−1.0	[46]	WMAP1+, LSS, SN Ia	Undecided
Primordial power spectrum					
No tilt ($n_s = 1$)	−1	+0.4	[47]	WMAP1+, LSS	Undecided
		$[-1.1, -0.6]^p$	[51]	WMAP1+, LSS	Undecided
		−0.7	[58]	WMAP1+, LSS	Undecided
		−0.9	[70]	WMAP1+	Undecided
		$[-0.7, -1.7]^{p,d}$	[186]	WMAP3+	$n_s = 1$ weakly disfavoured
		−2.0	[185]	WMAP3+, LSS	$n_s = 1$ weakly disfavoured
		−2.6	[70]	WMAP3+	$n_s = 1$ moderately disfavoured
		−2.9	[58]	WMAP3+, LSS	$n_s = 1$ moderately disfavoured
		$< -3.9^c$	[65]	WMAP3+, LSS	Moderate evidence at best against $n_s \neq 1$
Running	+1	$[-0.6, 1.0]^{p,d}$	[186]	WMAP3+, LSS	No evidence for running
		$< 0.2^c$	[166]	WMAP3+, LSS	Running not required
Running of running	+2	$< 0.4^c$	[166]	WMAP3+, LSS	Not required
Large scales cut–off	+2	$[1.3, 2.2]^{p,d}$	[186]	WMAP3+, LSS	Weak support for a cut–off
Matter–energy content					
Non–flat Universe	+1	−3.8	[70]	WMAP3+, HST	Flat Universe moderately favoured
		−3.4	[58]	WMAP3+, LSS, HST	Flat Universe moderately favoured
Coupled neutrinos	+1	−0.7	[193]	WMAP3+, LSS	No evidence for non–SM neutrinos
Dark energy sector					
$w(z) = w_{\rm eff} \neq -1$	+1	$[-1.3, -2.7]^p$	[187]	SN Ia	Weak to moderate support for Λ
		−3.0	[50]	SN Ia	Moderate support for Λ
		−1.1	[51]	WMAP1+, LSS, SN Ia	Weak support for Λ
		$[-0.2, -1]^p$	[188]	SN Ia, BAO, WMAP3	Undecided
		$[-1.6, -2.3]^d$	[189]	SN Ia, GRB	Weak support for Λ
$w(z) = w_0 + w_1 z$	+2	$[-1.5, -3.4]^p$	[187]	SN Ia	Weak to moderate support for Λ
		−6.0	[50]	SN Ia	Strong support for Λ
		−1.8	[188]	SN Ia, BAO, WMAP3	Weak support for Λ
$w(z) = w_0 + w_a(1 - a)$	+2	−1.1	[188]	SN Ia, BAO, WMAP3	Weak support for Λ
		$[-1.2, -2.6]^d$	[189]	SN Ia, GRB	Weak to moderate support for Λ
Reionization history					
No reionization ($\tau = 0$)	−1	−2.6	[70]	WMAP3+, HST	$\tau \neq 0$ moderately favoured
No reionization and no tilt	−2	−10.3	[70]	WMAP3+, HST	Strongly disfavoured

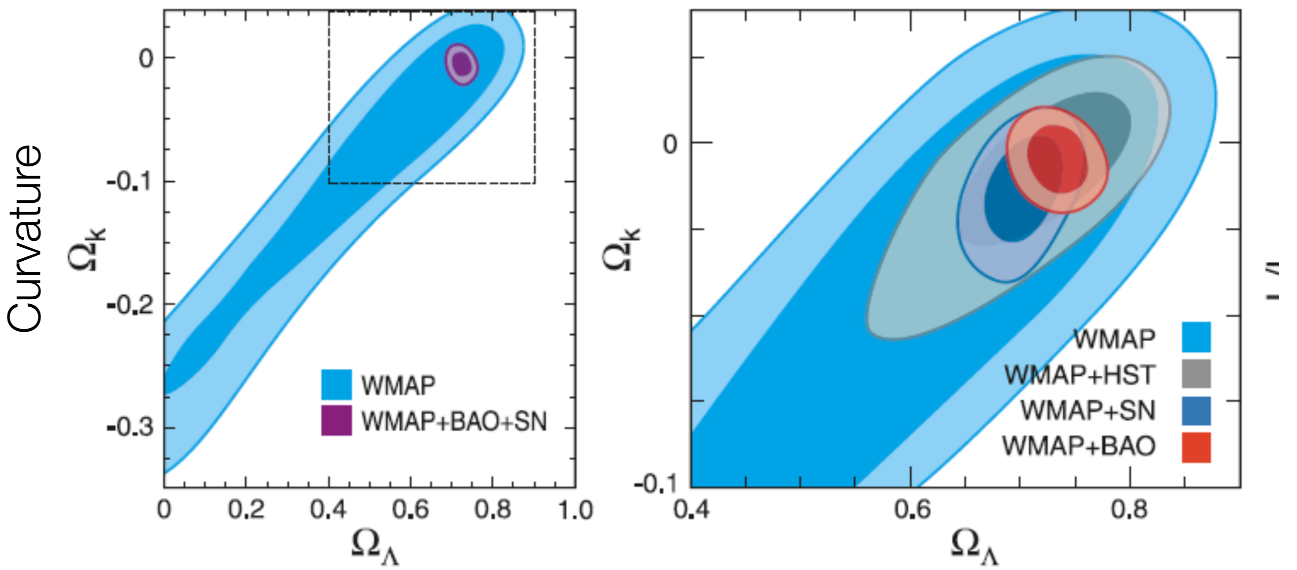
lnB < 0: ΛCDM remains the "best" model from a Bayesian perspective!

Trotta (2008)

Roberto Trotta

Level 1 inference:
Constraints on curvature

Imperial College London



Komatsu et al (WMAP Team) (2006)

Assuming flatness ($\Omega_\kappa = 0$):

Ω_Λ = **0.721 ± 0.015**
Ω_{cdm} = **0.233 ± 0.013**
Ω_b = **0.0462 ± 0.0015**

Assuming dark energy is Λ:

$0.0170 < \Omega_\kappa < 0.0068$ (95%)

Roberto Trotta

Level 2 inference: a three-way model comparison

Vardanyan, RT & Silk (2009)

Imperial College London

- For a FRW Universe, there are only 3 discrete models for the geometry:

$$ds^2 = -dt^2 + a^2 \left(\frac{dr^2}{1-\kappa r^2} + r^2 d\Omega \right)$$

$$\Omega_\kappa = -\frac{\kappa}{H_0^2 a_0^2}$$

Model 0: $\kappa = 0$	Model 1: $\kappa = +1$	Model -1: $\kappa = -1$
Flat	Closed	Open
$\Omega_\kappa = 0$	$\Omega_\kappa < 0$	$\Omega_\kappa > 0$
$P(M_0) = 1/3$	$P(M_{+1}) = 1/3$	$P(M_{-1}) = 1/3$

Roberto Trotta

Curvature priors

Imperial College London

- The Astronomer's prior:

  motivated by consistency with basic properties of the observed Universe (age of oldest objects, obviously non-empty)

$$-1 \leq \Omega_\kappa \leq 1 \text{ (flat on } \Omega_\kappa)$$

- The Curvature scale prior:

  gives the same prior probability to all orders of magnitude for the curvature radius (a_0), between 10^{-5} for the curvature parameter (size of curvature perturbation) to unity (Universe not empty)

$$-5 \leq \log|\Omega_\kappa| \leq 0 \text{ (flat on } \log\Omega_\kappa)$$

Results: current model comparison

Imperial College London

- A positive lnB favours the flat model over curved one

prior = 1/3 prior = 2/3

Data sets and models	$\ln B_{01}$	$\ln B_{0-1}$	$p(\mathcal{M}_0\|d)$	$p(N_U = \infty\|d)$	Notes
				Astronomer's prior (flat in Ω_κ)	
WMAP5+BAO ($w = -1$)	4.1	5.3	0.98	0.98	Moderate evide
WMAP5+BAO+SNIa ($w = -1$)	4.2	5.3	0.98	0.98	Moderate evide
WMAP5+BAO ($w \neq -1$)	1.0	6.1	0.74	0.74	Weak evidence
WMAP5+BAO+SNIa ($w \neq -1$)	3.9	5.3	0.98	0.98	Moderate evide
				Curvature scale prior (flat in o_κ)	
WMAP5+BAO ($w = -1$)	0.4	0.6	0.45	0.69	Inconclusive
WMAP5+BAO+SNIa ($w = -1$)	0.4	0.6	0.45	0.69	Inconclusive
WMAP5+BAO ($w \neq -1$)	−0.8	0.5	0.26	0.42	Inconclusive
WMAP5+BAO+SNIa ($w \neq -1$)	0.3	0.6	0.44	0.67	Inconclusive

posterior probability of flatness

posterior probability of an infinite Universe

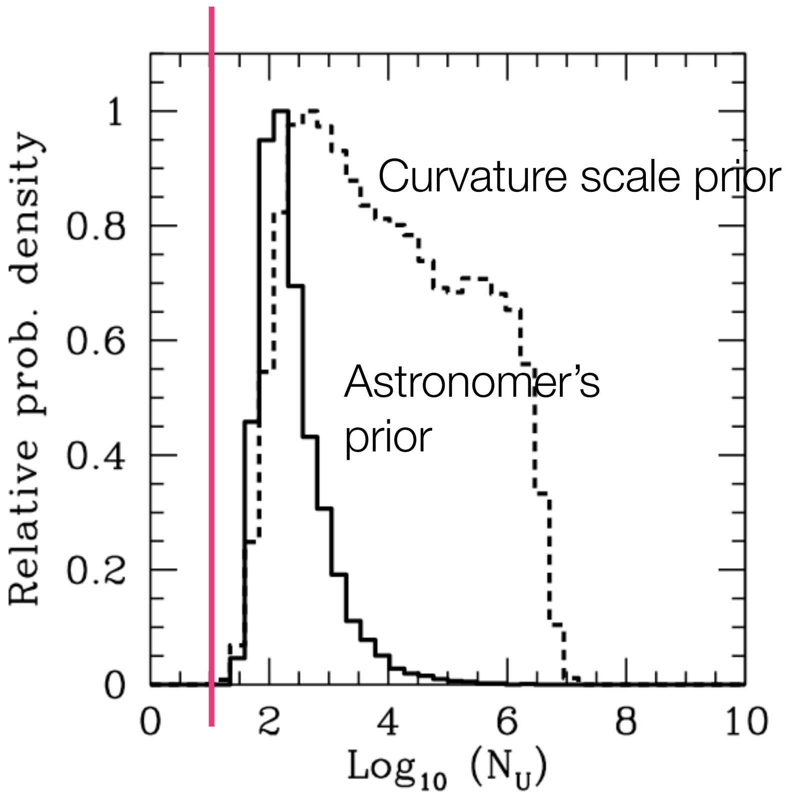
Vardanyan, RT & Silk (2009)

Roberto Trotta

The number of Hubble spheres

- For closed models, we can compute the probability distribution of the number of Hubble spheres (apparent particle horizon) contained in a spatial slice:

$N_U > 5$ a robust lower bound



Level 3 inference:
Bayesian model-averaged constraints

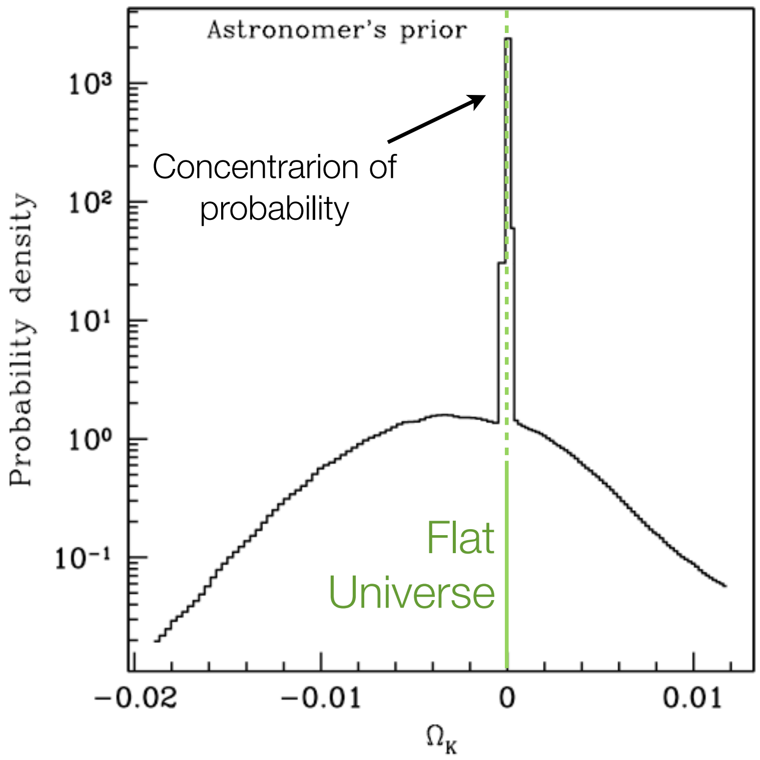
Vardanyan, RT & Silk (2011)

Imperial College London

$$P(\theta|d) = \sum_i P(M_i|d)P(\theta|d, M_i)$$

- **Aim:** model-independent constraints that account for model uncertainty
- **Model posterior:** flat models are preferred by Bayesian model selection → probability gets concentrated onto those models
- **Consequence:** constraints on the curvature, number of Hubble spheres and size of the Universe can be **stronger** after Baysian model averaging!
- **Number of Hubble spheres** $N_U > 251$ (99%)
  ~8 times stronger
  Radius of curvature > 42 Gpc (99%)
  1.5 times stronger

Vardanyan, RT & Silk (2011)



Roberto Trotta

Hunting down the best model of inflation

Inflationary models: large and small field

Imperial College London

- The simplest inflationary scenario is based on one single scalar field (adiabatic perturbations)
- Taylor expansion of the potential V(φ) of single-field models gives two classes:

$$V(\phi) = M^4 \left(\frac{\phi}{M_{\rm Pl}}\right)^p \quad \text{(large field)}$$

$$V(\phi) = M^4 \left[1 - \left(\frac{\phi}{\mu}\right)^p\right] \quad \text{(small field)}$$

Priors on inflationary potential parameters

Imperial College London

- Priors need to be chosen carefully based on physical considerations!
- Some arbitrariness involved in some choices, but mostly dictated by physical boundaries or theoretical prejudice - see Martin, Ringeval & Trotta (2011)
- Data: WMAP7. Parameters and priors (Martin et al, arxiv: 1009.4157):

$$V(\phi) = M^4 \left[1 - \left(\frac{\phi}{\mu}\right)^p\right] \qquad V(\phi) = M^4 \left(\frac{\phi}{M_{\rm Pl}}\right)^p$$

Parameter	Small field models, Eq. (4)			Large field models, Eq. (3)					
	SFI_s	SFI_l	SFI_f	LFI_p	$\mathrm{LFI}_{2/3}$	LFI_1	LFI_2	LFI_3	LFI_4
Normalization, $\ln P_*$	$[2.7\times10^{-10}, 4.0\times10^{-10}]$			$[2.7\times10^{-10}, 4.0\times10^{-10}]$					
Exponent, p	$[2.4, 10]$			$[0.2, 5]$	$2/3$	1	2	3	4
Vacuum expectation, $\log(\mu/M_{\rm Pl})$	$[-1, 0]$	$[0, 2]$	$[-1, 2]$	Not applicable					
Reheating, $\ln R$	$[-46, 15]$			$[-46, 15]$					
n number of free parameters	4	4	4	3	2	2	2	2	2

Roberto Trotta

Effective model complexity

Imperial College London

- "Number of free parameters" is a relative concept. The relevant scale is set by the prior range
- How many parameters can the data support, regardless of whether their detection is significant?
- **The Bayesian complexity** measures the effective number of parameters:

$$\mathcal{C}_b = \overline{\chi^2(\theta)} - \chi^2(\hat{\theta}) = \sum_i \frac{1}{1 + (\sigma_i/\Sigma_i)^2}$$

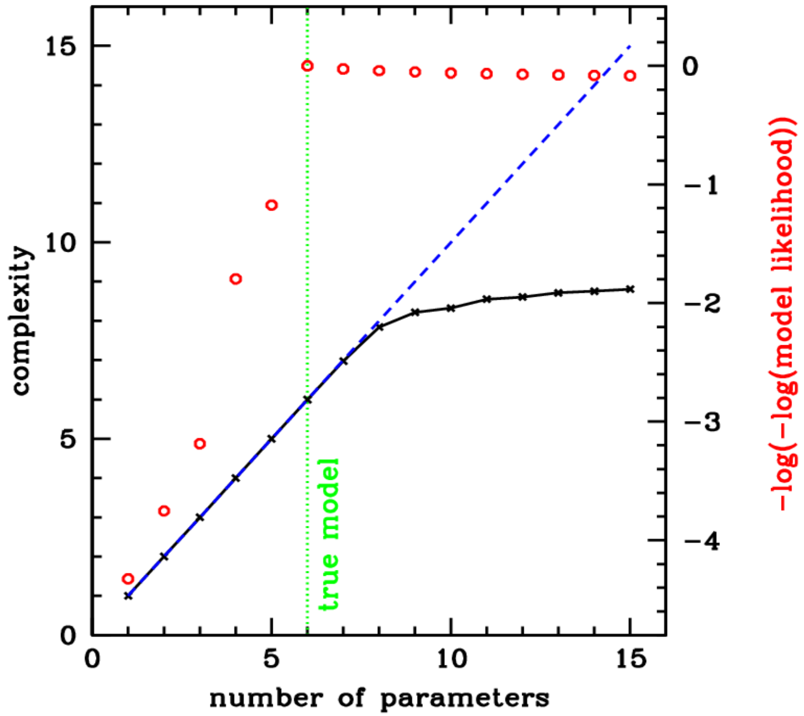
Kunz, RT & Parkinson (2006), Spiegelhalter et al (2002)

Example: polynomial fitting

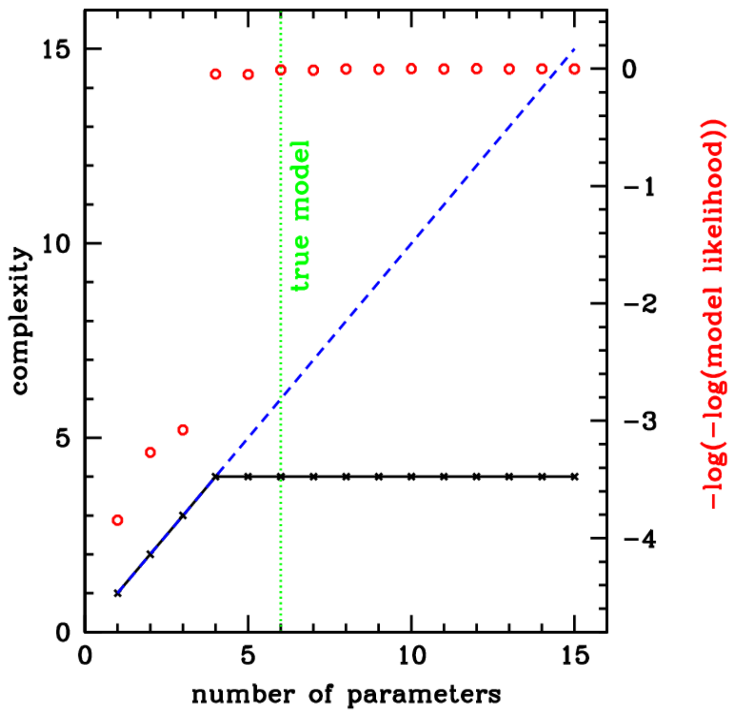
Imperial College London

- Data generated from a model with n = 6:

GOOD DATA
Max supported complexity ≈ 9



INSUFFICIENT DATA
Max supported complexity ≈ 4

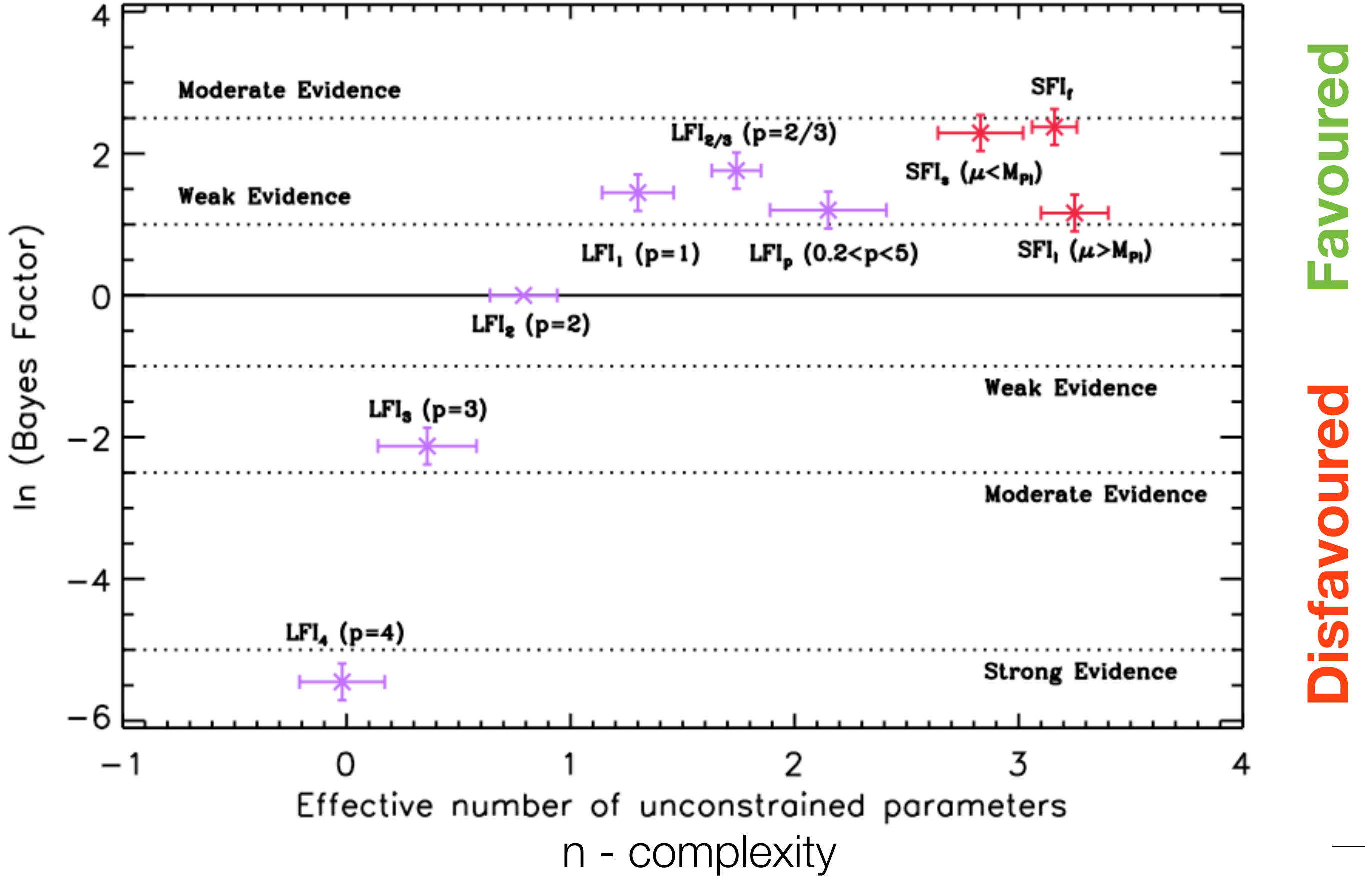


Roberto Trotta

Results: Small field models favoured

Imperial College London

The probability of small field models rises from an initial 50% to
P(small field | data) = 0.77 ± 0.03



Roberto Trotta

Conclusions

Imperial College London

- Determining the presence of new parameters is a model comparison task: this requires the Bayesian evidence
- Bayesian model comparison allows to quantify the preference between two or more competing models, automatically implementing Occam's razor.
- The prior choice for the extra parameters is critical in controlling the strength of the Occam's razor effect. As such, a sensitivity analysis is mandatory.