Recent Advances in Model Comparison Techniques

Anže Slosar

Trieste, 31 May 2005

1 – Introduction

- Done in collaboration with Andrew Liddle (Sussex), Juan García-Bellido and María Beltrán (Madrid)
- Work in progress
- Bayesian parameter estimation is a well established procedure with MCMC as a standard method
- Parameter estimation will “estimate” parameters, regardless of whether the model is appropriate or not.
- Bayesian model comparison allow one to compare models. It picks the most suitable model based on
  - ability to fit the data
  - complexity

2 – Introduction

- Basic quantity that describes the relative probability of a model is *Evidence*:
  - it is a Bayesian equivalent of χ^2
  - Occam's razor built in
  - integral of likelihood over the prior:

$$E = \int L(\mathbf{x})\pi(\mathbf{x})\mathrm{d}^{\mathrm{n}}\mathbf{x} \tag{1}$$

3 – Example

Isocurvature models (astro-ph/0501477):

Parameter	Prior Range	Model
$\omega_{\rm b}$	(0.018,0.032)	AD-HZ,AD-$n_{\rm s}$,ISO
$\omega_{\rm dm}$	(0.04,0.16)	AD-HZ,AD-$n_{\rm s}$,ISO
θ	(0.98,1.10)	AD-HZ,AD-$n_{\rm s}$,ISO
τ	(0,0.5)	AD-HZ,AD-$n_{\rm s}$,ISO
$\ln[10^{10}\mathcal{R}_{\rm rad}]$	(2.6,4.2)	AD-HZ,AD-$n_{\rm s}$,ISO
$n_{\rm s}$	(0.8,1.2)	AD-$n_{\rm s}$,ISO
$n_{\rm iso}$	(0,3)	ISO
$\delta_{\rm cor}$	($-$0.14,0.4)	ISO
$\sqrt{\alpha}$	($-$1,1)	ISO
β	($-$1,1)	ISO

Model	ln(Evidence)
AD-HZ	0.0 ± 0.1
AD-$n_{\rm s}$	0.0 ± 0.1
CDI	-1.0 ± 0.2
NID	-1.0 ± 0.2
NIV	-1.0 ± 0.3

4 – Thermodynamic integration

- The iso-curvature paper was a proof of concept
- The Evidence was calculated by the thermodynamic integration:

$$E = \int_0^1 \mathrm{d}\lambda \, \langle \log \mathrm{L} \rangle_{\mathrm{L}^\lambda \pi} \tag{2}$$

- Computationally extremely inefficient: 10^6 samples required for reasonable accuracy
- Method "learns" about the extend of the prior by seeing how samples behave at low lambda: clearly inefficient.
- Naive approaches don't work:
  - random sampling never hits high L region
  - average of burned-in samples L is evidence if prior=posterior (bad!)

5 – Thermodynamic integration

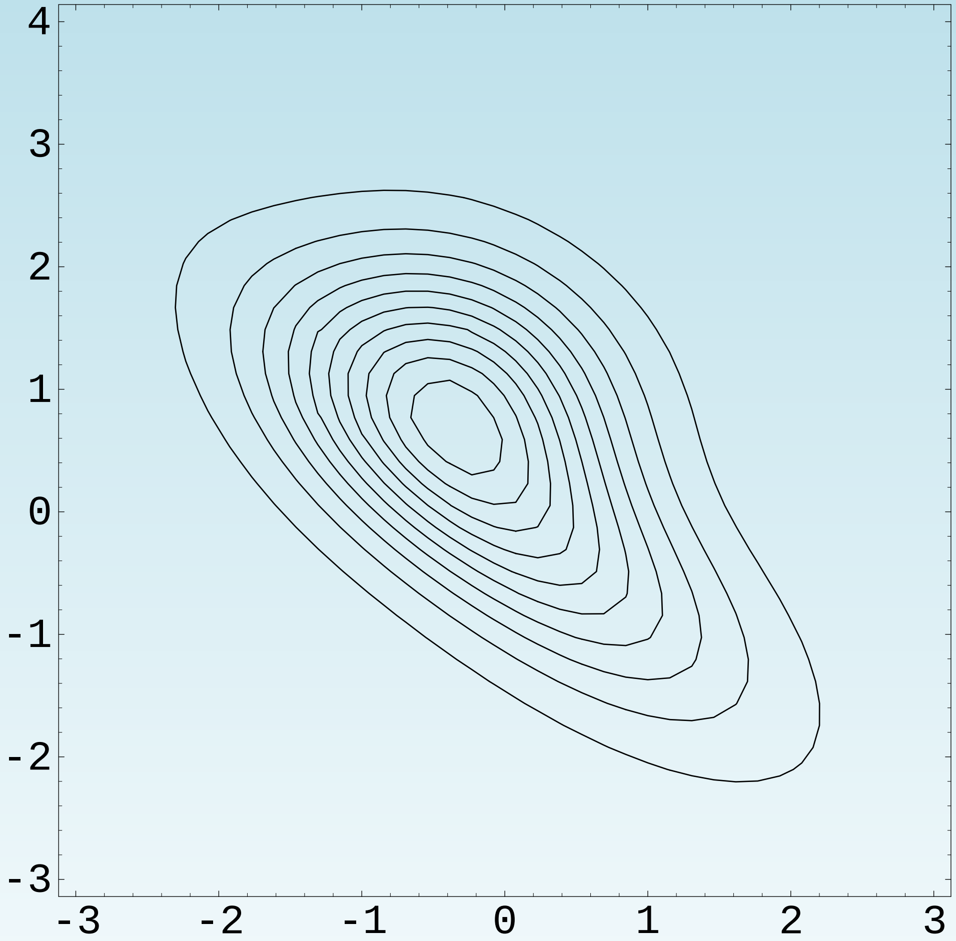
We attempted several methods to get evidence, or an approximation of evidence from burned-in samples alone:

- Would put evidence calculation into mainstream allowing people to reuse parameter estimation chains
- Must put in prior width by hand (good!)
- Gaussian approximation + expansion
- Savage - Dickey approach

6 – Gaussian + perturbative expansion

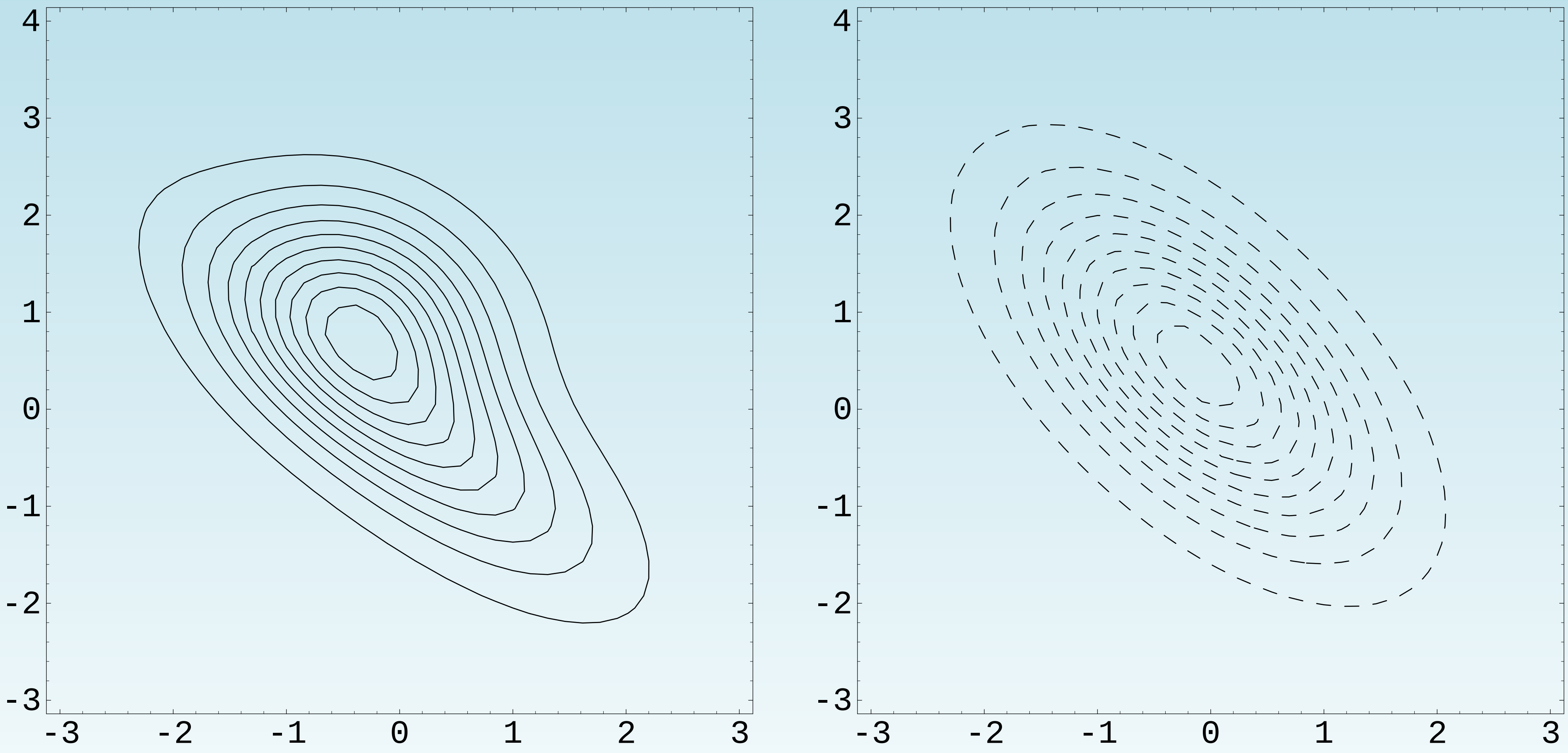
- Approximate posterior by a Gaussian and calculate evidence.
- Allows one to add Skewnes, Kurtosis and higher moments to improve fit.

7 – Gaussian - example

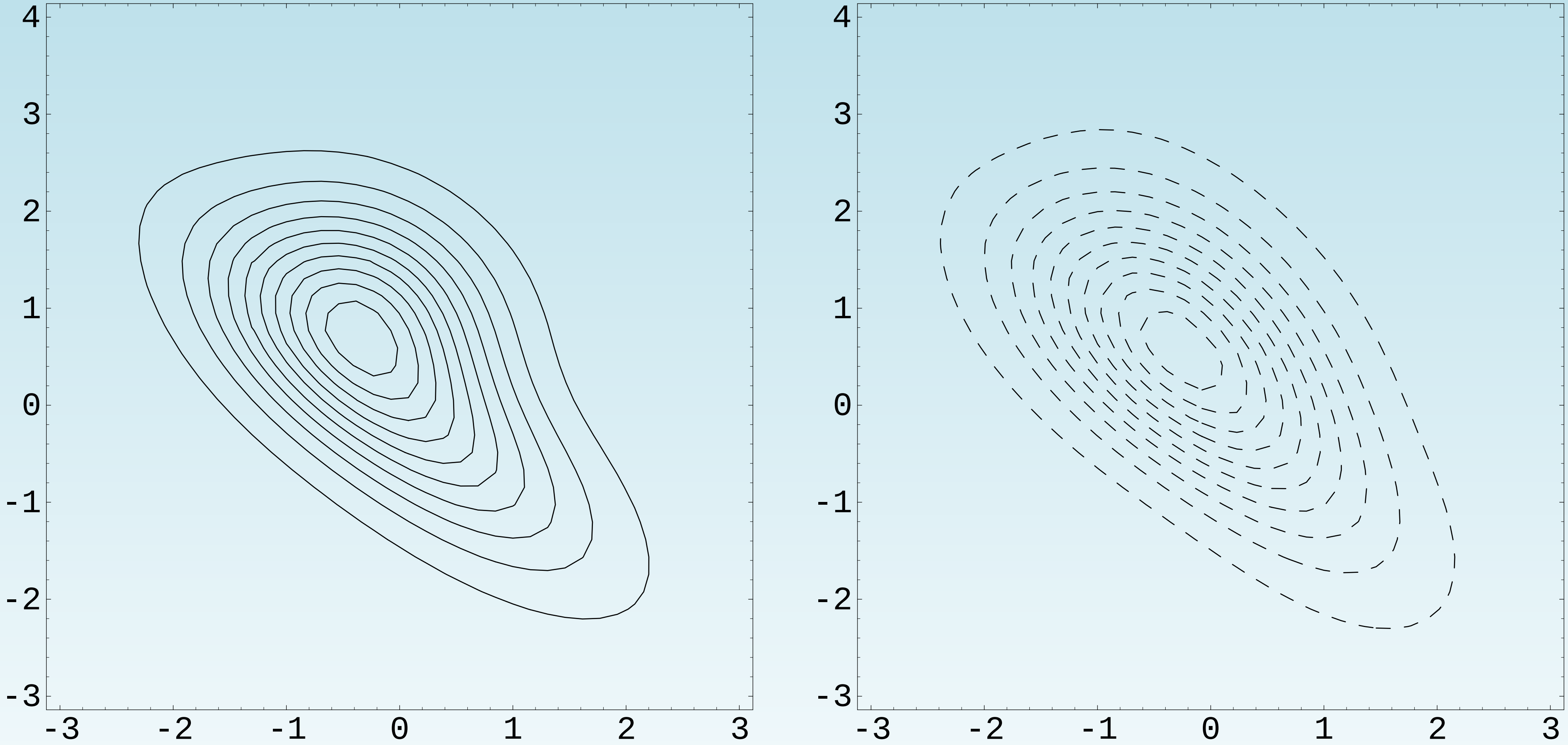


- A non-gaussian function made by summing two Gaussians
- Approximate by Gaussian with the same mean / covariance
- Add skewness and kurtosis corrections

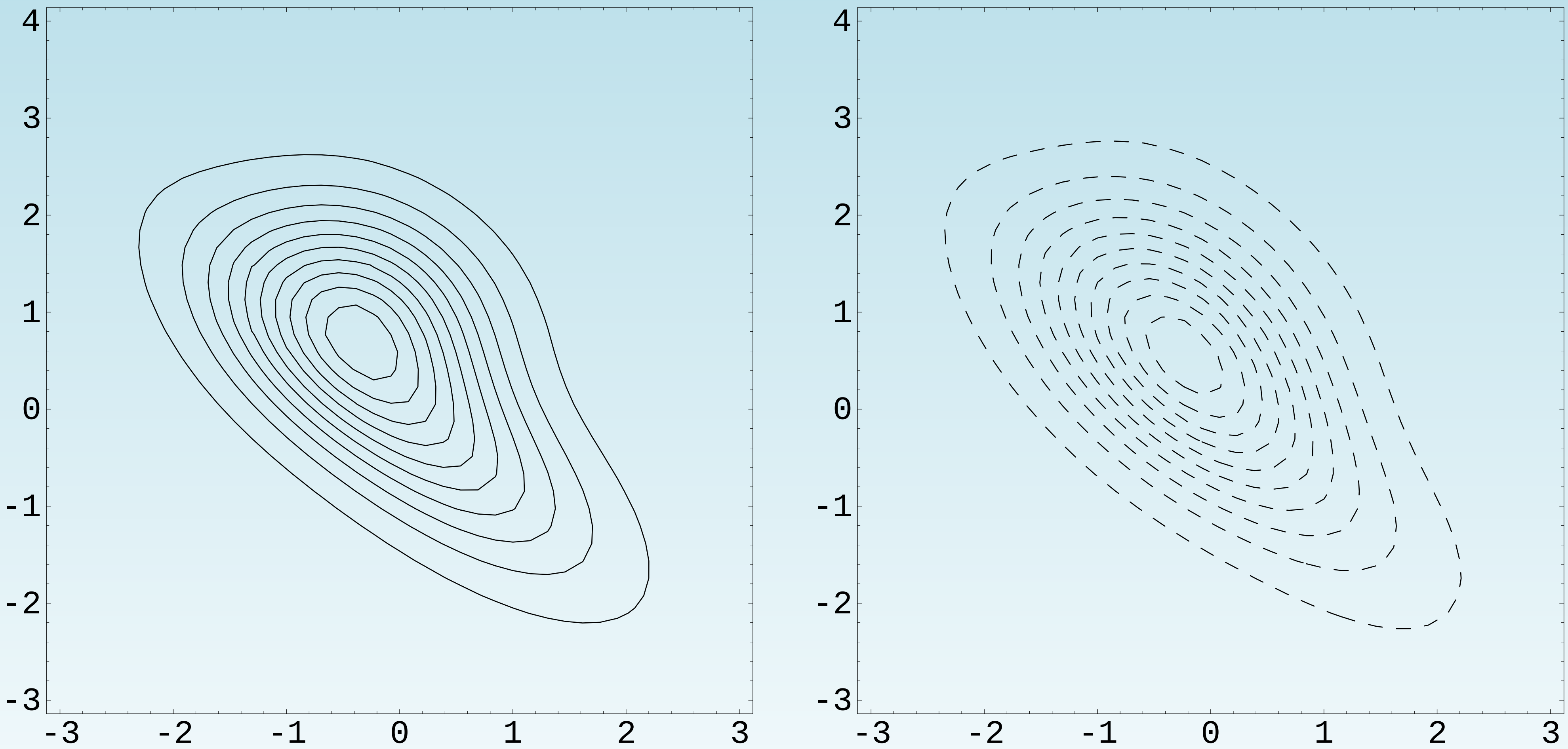
8 – Gaussian - example



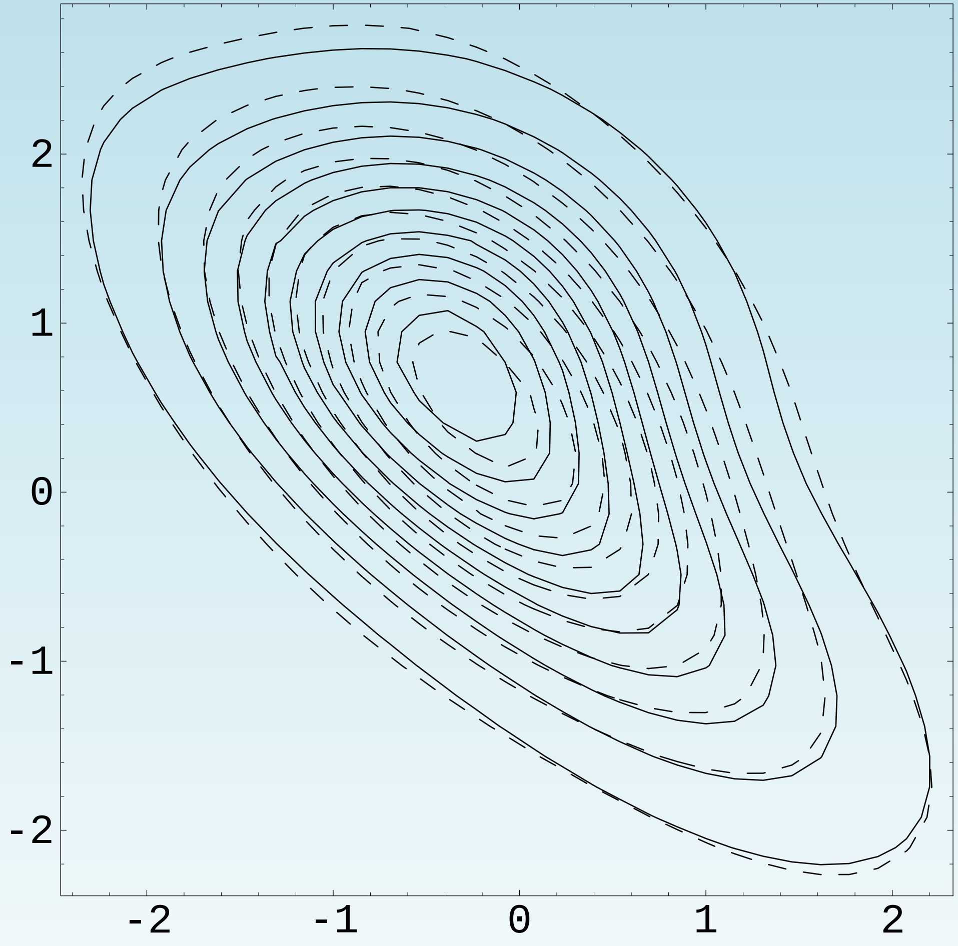
9 – Gaussian - example



10 – Gaussian - example

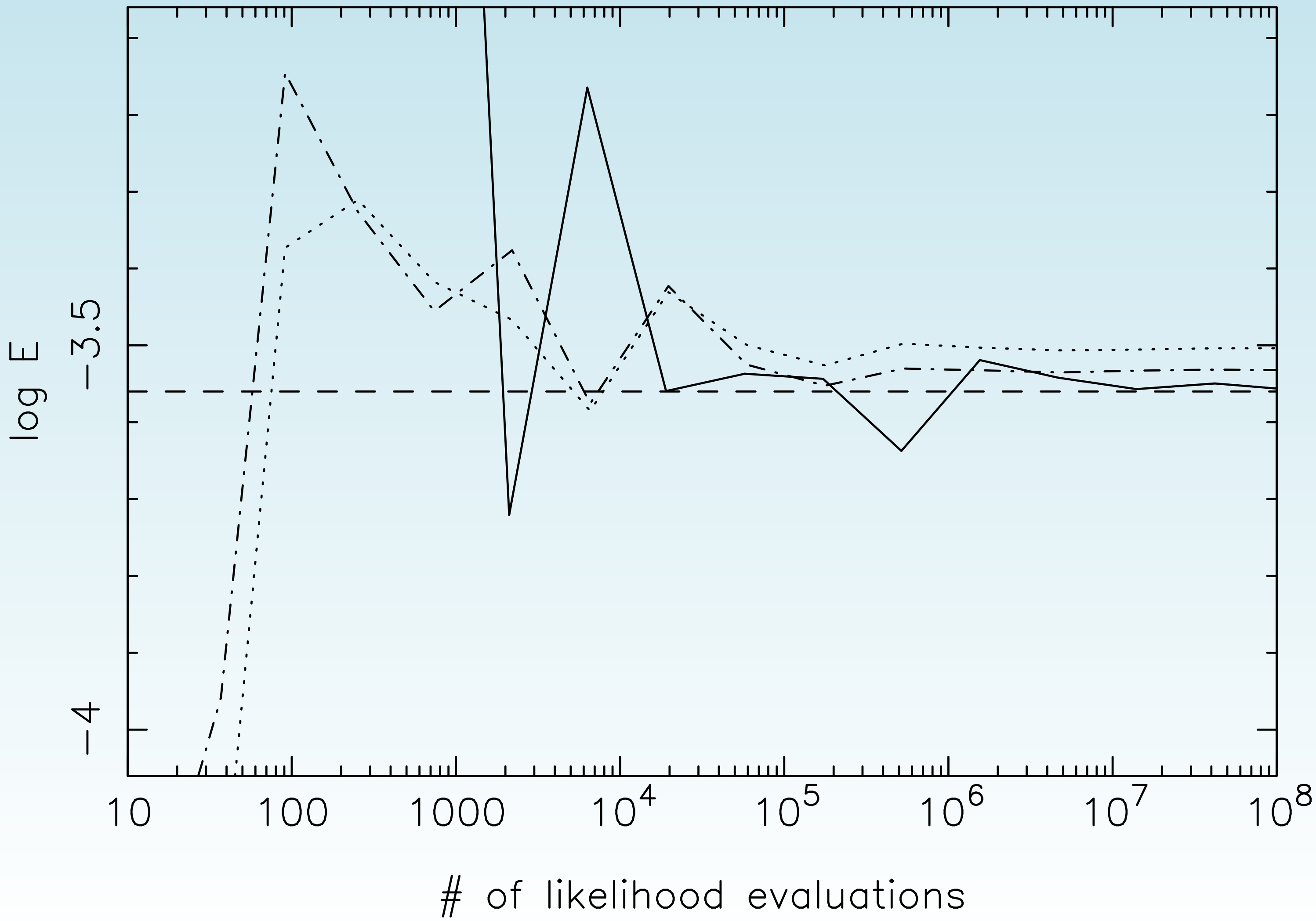


11 – Gaussian - example



- Works extremely well
- Evidence the same to better that 1%
- Leading error comes from overall normalisation of the fitting function

12 – Speed of convergence



13 – In practice

- In principle it should be easy:
  - Calculate means, Covariances and higher order matrices if necessary
  - Plug them into equation and voila
- Turns to be fairly difficult thing to do accurately
- Expanding around mean gives different results that expanding around most likely point
- Inherently prone to systematics if prior small
- Problem with overall normalisation
- Gives imaginary evidence occasionally (yuck!)

14 – Preliminary results

	6-param flat basic model	NID isocurvature model (10 param)
Thermodynamic integration	-855.1	-856.1
Gaussian expansion around mean	-855.5	-853.6
+Skew+Kurtosis expansion around mean	-854.5	-855.1
+fitting likelihood normalisation	-856.5	-857.5
Gaussian expansion around MaxLike	-854.2	-850.6
+Skew+Kurtosis expansion around MaxLike	-855.6	x
+fitting likelihood normalisation	-856.6	x

Works fine for basic model but not for isocurvature models

15 – Direct fitting of the likelihood function

- Do not infer the covariance matrix, etc. from moments, but rather fit directly.
- Much more stable and efficient (compared 1d Gaussian)
- No ambiguities wrt to where from to expand, normalise, etc.
- We minimise $\sum(\log L_{\mathrm{theory}} - \log L_{sample})^2$
- Becomes difficult problem again:
  - 126 numbers to fit for a 6d problem
  - 626 numbers to fit for a 10d problem
  - need a supercomputer, but much faster than taking 10^6 samples
- Can actually estimate a typical error of the fitting function

16 – Gaussian Fitting Results

	6-param flat basic model	NID isocurvature model (10 param)
Thermodynamic integration	-855.1	-856.1
Gaussian expansion	-854.2	-847.5
rms error in $\log L$	0.81	5.4
+Skew+Kurtosis	-855.6	
rms error in $\log L$	0.6	

Again, works fine for basic model but does not seem to converge to the correct minimum for NID.

17 – Savage - Dickey Method

- Recently advocated by R. Trotta (astro-ph/0504022).
- Works for nested models (for example flat models are nested in more general variable Ω_k models).
- No assumption about the shape of likelihood
- Essentially:

$$E \propto \frac{\text{\# of samples in a model}}{\text{prior volume}} \tag{3}$$

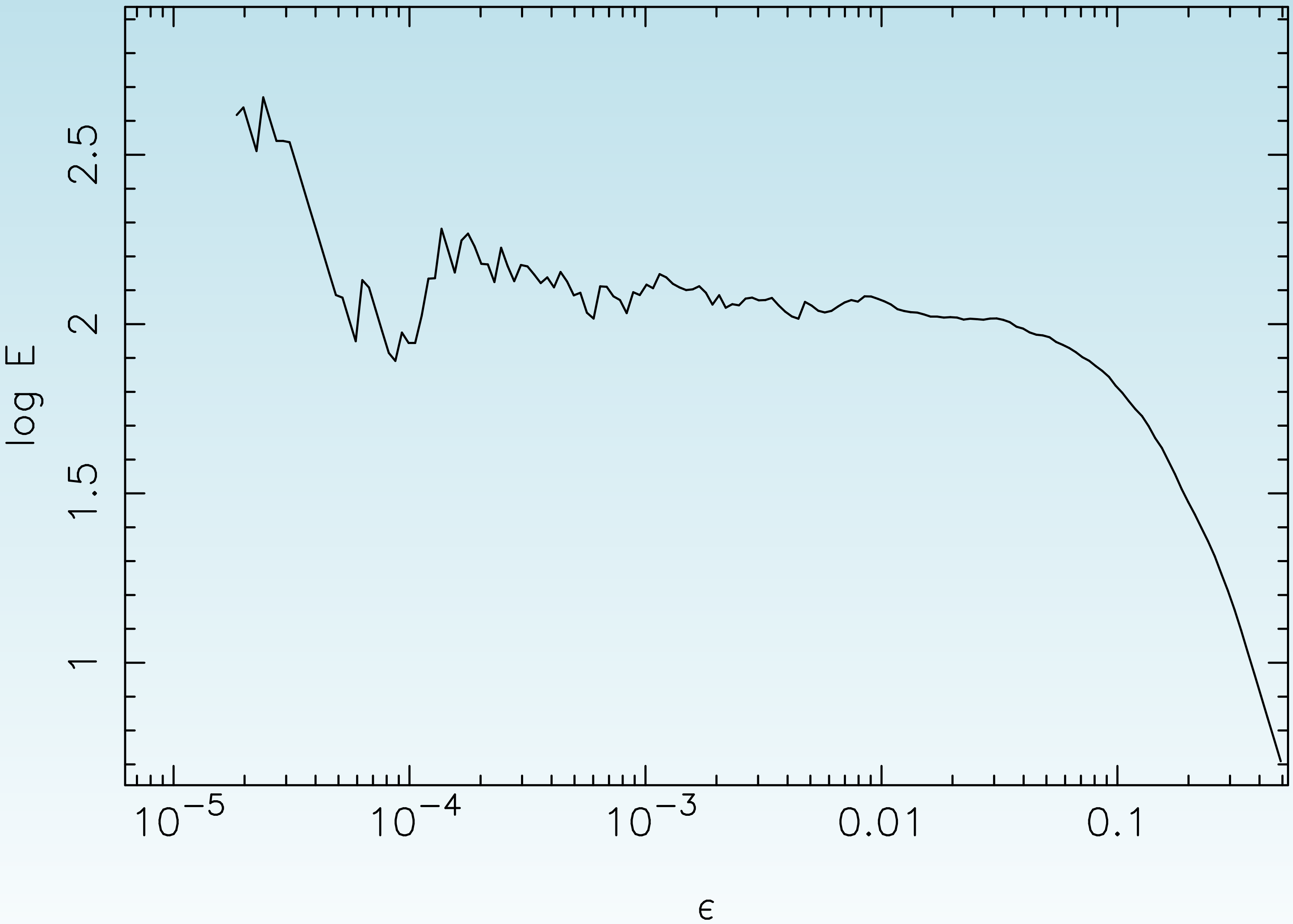
18 – S-D in practice

NID vs basic model:

- NID has 4 more parameters: α, β, $n_{\rm iso}$, $\delta_{\rm cross}$.
- $\alpha = 0$ corresponds to purely adiabatic modes: β, $n_{\rm iso}$, $\delta_{\rm cross}$ unconstrained as they do not affect likelihood.
- Adding an unconstrained parameter to a model doesn't change its evidence.
- $\alpha = 0$ is thus a nested adiabatic model

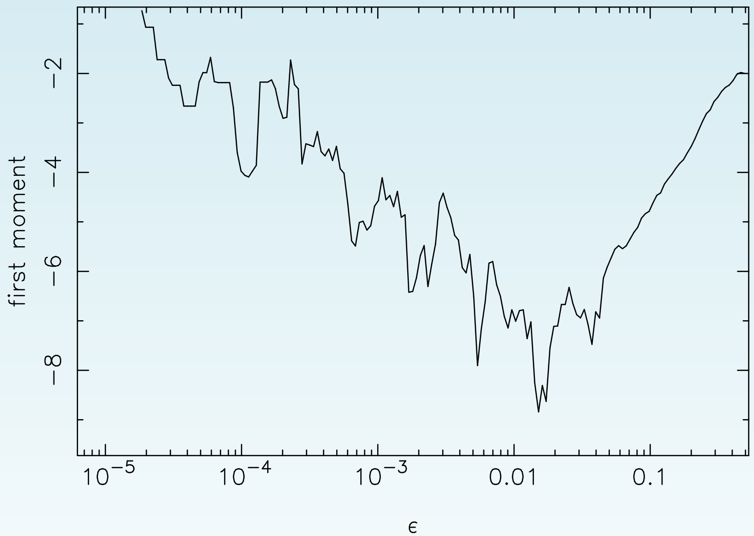
19 – S-D in practice

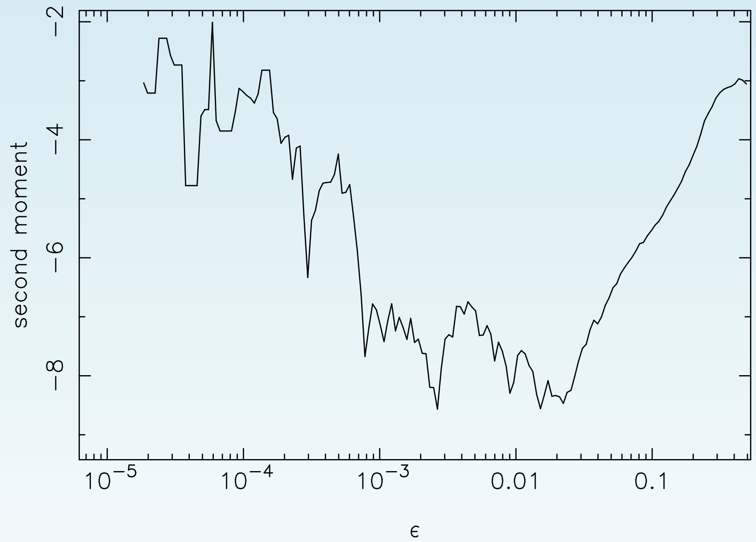
Real sampling never hits $\alpha = 0$. So we approximate the adiabatic model by $\alpha < \epsilon$.



20 – How to choose ϵ

- Want to hit the plateau at small ϵ
- $\alpha < \epsilon$ must be a good approximation to $\alpha = 0$
- Compare first and second moments of distributions from $\alpha < \epsilon/2$ and $\epsilon/2 < \alpha < \epsilon$.





21 – S-D results

- Dominating source of uncertainty is sampling error - calculate error from scatter from different chains
- $\Delta E_{\mathrm{nid-ad}} = -1.9 \pm 0.15$
- Inconsistent with thermodynamic results of -1.0 ± 0.2 (but issues wrt to Gaussian errors)
- *Very* fast!

22 – Conclusions

- Can do model comparison, but computationally extremely expensive
- At the moment no-one is doing it regularly, but given the right tools, people would use it (?!)
- We are developing these methods: not quite there, but close.