Recent advances in cosmological Bayesian model comparison

Roberto Trotta
[DRAFT VERSION, May 23rd 2011]

Abstract I review the framework of Bayesian model comparison as applied to cosmological model building. I then discuss some recent developments in the evaluation of the Bayesian evidence, the central quantity for Bayesian model comparison, and present applications to inflationary model building and to constraining the curvature and minimum size of the Universe. I conclude by discussing what I think are some of the open challenges in the field.

1 Introduction

Many problems in cosmology and astrophysics are about deciding whether the available data require the inclusion of a new parameter in a baseline model. Examples of such problems include identifying astronomical sources in an image; deciding whether the Universe is flat or not, or whether the dark energy equation of state parameter changes with time; detecting an exo-planet orbiting a distant star; identifying a line in a spectrum, and many others.

The classical approach to this kind of questions takes the form of hypothesis testing: a null hypothesis is set up (where the effect one is looking for is supposed absent) and a test is performed to reject it, at a certain significance level. This involves comparing the observed value of a test statistics (typically, the χ^2) with the value it would assume *if the null hypothesis were true.* The shortcomings of this methodology are that (i) it does not return a probability for the hypothesis (contrary to a common misunderstanding among astrophysicists) and (ii) it cannot confirm a hypothesis, merely fail to reject it (see [1, 2, 3] for a more detailed discussion).

Some of those problems are resolved if one takes a Bayesian outlook, and adopts the framework of Bayesian model comparison. When there are several competing theoretical models, Bayesian model comparison provides a formal way of evaluating

Imperial College London, Blackett Laboratory, Prince Consort Road, London, SW7 2AZ, UK. e-mail: r.trotta@imperial.ac.uk

their relative probabilities in light of the data and any prior information available. The “best” model is then the one which strikes an optimum balance between quality of fit and predictivity. In fact, it is obvious that a model with more free parameters will always fit the data better (or at least as good as) a model with less parameters. However, more free parameters also mean a more “complex” model (a precise definition of “model complexity” can be found in [4]). Such an added complexity ought to be avoided whenever a simpler model provides an adequate description of the observations. This guiding principle of simplicity and economy of an explanation is known as *Occam’s razor* — the simplest theory compatible with the available evidence ought to be preferred.

An important feature is that an alternative model must be specified against which the comparison is made. In contrast with frequentist goodness–of–fit tests, Bayesian model comparison maintains that it is pointless to reject a theory unless an alternative explanation is available that fits the observed facts better. In other words, unless the observations are totally impossible within a model, finding that the data are improbable given a theory does not say anything about the probability of the theory itself *unless we can compare it with an alternative*. A consequence of this is that the probability of a theory that makes a correct prediction can increase if the prediction is confirmed by observations, provided competitor theories do not make the same prediction.

2 Bayesian model comparison

2.1 Shaving theories with Occam’s razor

Bayesian inference is often the statistical framework of choice in cosmology (see e.g. [5, 3]), and, increasingly so, in astroparticle physics. The posterior pdf $p(\Theta|d,\mathscr{M})$ for the n-dimensional parameters vector Θ of a model $\mathscr{M}$ is given by

$$p(\Theta|d,\mathscr{M}) = \frac{p(d|\Theta,\mathscr{M})p(\Theta|\mathscr{M})}{p(d|\mathscr{M})}\,. \tag{1}$$

Here, $p(\Theta|\mathscr{M})$ is the prior, $p(d|\Theta,\mathscr{M})$ the likelihood and $p(d|\mathscr{M})$ the model likelihood, or marginal likelihood (usually called “Bayesian evidence” by physicists), the central quantity for Bayesian model comparison.

In the context of model comparison it is appropriate to think of a model as a specification of a set of parameters Θ *and* of their prior distribution, $p(\Theta|\mathscr{M})$. It is the number of free parameters *and* their prior range that control the strength of the Occam’s razor effect in Bayesian model comparison: models that have many parameters that can take on a wide range of values but that are not needed in the light of the data are penalized for their unwarranted complexity. Therefore, *the prior choice ought to reflect the available parameter space under the model $\mathscr{M}$, independently of experimental constraints we might already be aware of*. This is because we are trying to assess the economy (or simplicity) of the model itself, and hence the prior should be based on theoretical or physical constraints on the model under consideration. Often these will take the form of a range of values that are deemed "intuitively" plausible, or "natural". Thus the prior specification is inherent in the model comparison approach.

2.2 The Bayesian evidence

The evaluation of a model's performance in the light of the data is based on the *Bayesian evidence*, the normalization integral on the right–hand–side of Bayes' theorem, Eq. (1):

$$p(d|\mathscr{M}) \equiv \int p(d|\Theta,\mathscr{M})p(\Theta|\mathscr{M})\mathrm{d}^n\Theta. \tag{2}$$

Thus the Bayesian evidence is the average of the likelihood under the prior for a specific model choice. From the evidence, the model posterior probability given the data is obtained by using Bayes' Theorem to invert the order of conditioning:

$$p(\mathscr{M}|d) \propto p(\mathscr{M})p(d|\mathscr{M}), \tag{3}$$

where $p(\mathscr{M})$ is the prior probability assigned to the model itself. Usually this is taken to be non–committal and equal to $1/N_m$ if one considers N_m different models. When comparing two models, $\mathscr{M}_0$ versus $\mathscr{M}_1$, one is interested in the ratio of the posterior probabilities, or *posterior odds*, given by

$$\frac{p(\mathscr{M}_0|d)}{p(\mathscr{M}_1|d)} = B_{01}\frac{p(\mathscr{M}_0)}{p(\mathscr{M}_1)} \tag{4}$$

and the *Bayes factor* B_{01} is the ratio of the models' evidences:

$$B_{01} \equiv \frac{p(d|\mathscr{M}_0)}{p(d|\mathscr{M}_1)} \quad \text{(Bayes factor)}. \tag{5}$$

A value $B_{01} > (<)\ 1$ represents an increase (decrease) of the support in favour of model 0 versus model 1 given the observed data. From Eq. (4) it follows that the Bayes factor gives the factor by which the relative odds between the two models have changed after the arrival of the data, regardless of what we thought of the relative plausibility of the models before the data, given by the ratio of the prior models' probabilities.

Bayes factors are usually interpreted against the Jeffreys' scale [6] for the strength of evidence, given in Table 1. This is an empirically calibrated scale, with thresholds at values of the odds of about 3 : 1, 12 : 1 and 150 : 1, representing weak, moderate and strong evidence, respectively.

Bayesian model comparison *does not* replace the parameter inference step (which is performed within each of the models separately). Instead, model comparison *ex-*

$\lvert \ln B_{01} \rvert$	Odds	Probability	Strength of evidence
< 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

Table 1 Empirical scale for evaluating the strength of evidence when comparing two models, $\mathscr{M}_0$ versus $\mathscr{M}_1$ (so–called "Jeffreys' scale"). Threshold values are empirically set, and they occur for values of the logarithm of the Bayes factor of $\lvert \ln B_{01} \rvert = 1.0$, 2.5 and 5.0. The right–most column gives our convention for denoting the different levels of evidence above these thresholds. The probability column refers to the posterior probability of the favoured model, assuming non–committal priors on the two competing models, i.e. $p(\mathscr{M}_0) = p(\mathscr{M}_1) = 1/2$ and that the two models exhaust the model space, $p(\mathscr{M}_0|d) + p(\mathscr{M}_1|d) = 1$.

tends the assessment of hypotheses in the light of the available data to the space of theoretical models, as evident from Eq. (4).

3 Recent developments

3.1 Numerical evaluation of the evidence

The computation of the Bayesian evidence, Eq. (2), is in general a numerically challenging task, as it involves a multi–dimensional integration over the whole of parameter space. Fortunately, several methods are now available, each with its own strengths and domains of applicability. Some of them have been developed by astronomers/cosmologists and are rapidly finding applications in other domains.

1. The numerical method of choice until recently has been thermodynamic integration, whose computational cost can however be fairly large. In typical cosmological applications [7, 8, 9], thermodynamic integration can require up to $\sim 10^7$ likelihood evaluations, two orders of magnitude more than MCMC–based parameter estimation. Recently, population Monte Carlo algorithms have been used succesfully to compute the evidence [10].
2. Skilling [11, 12] has put forward an elegant algorithm called "nested sampling", which has been implemented in the cosmological context by [13, 14, 15, 16, 17] (for a theoretical discussion of the algorithmic properties, see [18]). It calculates the evidence by transforming the multi-dimensional evidence integral of Eq. (2) into a one–dimensional integral that is easy to evaluate numerically. This is accomplished by defining the prior volume X as $\mathrm{d}X = p(\Theta)\mathrm{d}^n\Theta$, so that

$$X(\lambda) = \int_{\mathscr{L}(\Theta)>\lambda} p(\Theta)\mathrm{d}^n\Theta, \tag{6}$$

where $\mathscr{L}(\Theta) \equiv p(d|\Theta)$ is the likelihood function and the integral extends over the region(s) of parameter space contained within the iso-likelihood contour $\mathscr{L}(\Theta) = \lambda$ (in this section we drop the explicit conditioning on model $\mathscr{M}$, as this is understood). Assuming that $\mathscr{L}(X)$, i.e. the inverse of (6), is a monotonically decreasing function of X (which is trivially satisfied for most posteriors), the evidence integral (2) can then be written as

$$\mathscr{Z} \equiv p(d) = \int_0^1 \mathscr{L}(X)\mathrm{d}X, \tag{7}$$

Thus, if one can evaluate the likelihoods $\mathscr{L}_j = \mathscr{L}(X_j)$, where X_j is a sequence of decreasing values,

$$0 < X_M < \cdots < X_2 < X_1 < X_0 = 1, \tag{8}$$

as shown schematically in Fig. 1, the evidence can be approximated numerically using standard quadrature methods as a weighted sum

$$\mathscr{Z} = \sum_{i=1}^{M} \mathscr{L}_i w_i. \tag{9}$$

If one uses a simple trapezium rule, the weights are given by $w_i = \frac{1}{2}(X_{i-1} - X_{i+1})$. An example of a posterior in two dimensions and its associated function $\mathscr{L}(X)$ is shown in Fig. 1.

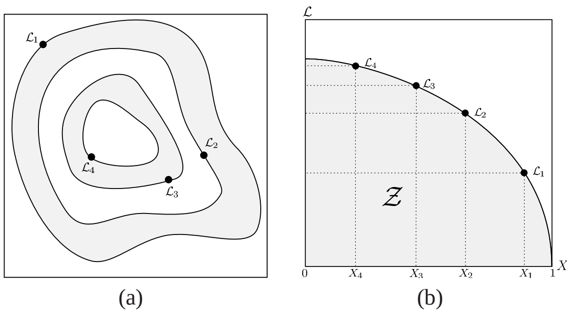

Fig. 1 Cartoon illustrating (a) the likelihood of a two-dimensional problem; and (b) the transformed $\mathscr{L}(X)$ function where the prior volumes X_i are associated with each likelihood $\mathscr{L}_i$. From [17].

This technique allows to reduce the computational burden to about $\sim 10^5$ likelihood evaluations. Recently, the development of what is called “multi–modal nested sampling” has allowed to increase significantly the efficiency of the method [16, 17], reducing the number of likelihood evaluations by another order of magnitude.

3. Useful approximations to the Bayes factor, Eq. (5), are available for situations in which the models being compared are *nested* into each other, i.e. the more complex model ($\mathscr{M}_1$) reduces to the original model ($\mathscr{M}_0$) for specific values of the new parameters. This is a fairly common scenario in cosmology, where one wishes to evaluate whether the inclusion of the new parameters is supported by the data. For example, we might want to assess whether we need isocurvature contributions to the initial conditions for cosmological perturbations, or whether a curvature term in Einstein's equation is needed, or whether a non–scale invariant distribution of the primordial fluctuation is preferred. Writing for the extended model parameters $\Theta = (\alpha, \beta)$, where the simpler model $\mathscr{M}_0$ is obtained by setting $\beta = 0$, and assuming further that the prior is separable (which is usually the case in cosmology), i.e. that

$$p(\alpha, \beta|\mathscr{M}_1) = p(\beta|\mathscr{M}_1)p(\alpha|\mathscr{M}_0), \tag{10}$$

the Bayes factor can be written in all generality as

$$B_{01} = \left. \frac{p(\beta|d, \mathscr{M}_1)}{p(\beta|\mathscr{M}_1)} \right|_{\beta=0}. \tag{11}$$

This expression is known as the Savage–Dickey density ratio (SDDR, see [19, 20]). The numerator is simply the marginal posterior under the more complex model evaluated at the simpler model's parameter value, while the denominator is the prior density of the more complex model evaluated at the same point. This technique is particularly useful when testing for one extra parameter at the time, because then the marginal posterior $p(\beta|d, \mathscr{M}_1)$ is a 1–dimensional function and normalizing it to unity probability content only requires a 1–dimensional integral, which is simple to do using for example the trapezoidal rule.

4. An instructive approximation to the Bayesian evidence can be obtained when the likelihood function is unimodal and approximately Gaussian in the parameters [21]. Expanding the likelihood around its peak to second order one obtains the Laplace approximation

$$p(d|\Theta, \mathscr{M}) \approx \mathscr{L}_{\max} \exp\left[-\frac{1}{2}(\Theta - \Theta_{\mathrm{ML}})^t L(\Theta - \Theta_{\mathrm{ML}})\right], \tag{12}$$

where Θ_{ML} is the maximum–likelihood point, $\mathscr{L}_{\max}$ the maximum likelihood value and L the likelihood Fisher matrix (which is the inverse of the covariance matrix for the parameters). Assuming as a prior a multinormal Gaussian distribution with zero mean and Fisher information matrix P one obtains for the evidence, Eq. (2)

$$p(d|\mathscr{M}) = \mathscr{L}_{\max} \frac{|F|^{-1/2}}{|P|^{-1/2}} \exp\left[-\frac{1}{2}({\Theta_{\mathrm{ML}}}^t L \Theta_{\mathrm{ML}} - \overline{\Theta}^t F \overline{\Theta})\right], \tag{13}$$

where the posterior Fisher matrix is $F = L + P$ and the posterior mean is given by $\overline{\Theta} = F^{-1} L \Theta_{\rm ML}$.

From Eq. (13) we can deduce a few qualitatively relevant properties of the evidence. First, the quality of fit of the model is expressed by $\mathscr{L}_{\max}$, the best–fit likelihood. Thus a model which fits the data better will be favoured by this term. The term involving the determinants of P and F is a volume factor, encoding the Occam's razor effect. As $|P| \leq |F|$, it penalizes models with a large volume of wasted parameter space, i.e. those for which the parameter space volume $|F|^{-1/2}$ which survives after arrival of the data is much smaller than the initially available parameter space under the model prior, $|P|^{-1/2}$. Finally, the exponential term suppresses the likelihood of models for which the parameters values which maximise the likelihood, $\Theta_{\rm ML}$, differ appreciably from the expectation value under the posterior, $\overline{\Theta}$. Therefore when we consider a model with an increased number of parameters we see that *its evidence will be larger only if the quality–of–fit increases enough to offset the penalizing effect of the Occam's factor*.

On the other hand, it is important to notice that the Bayesian evidence does *not* penalize models with parameters that are unconstrained by the data. It is easy to see that unmeasured parameters (i.e., parameters whose posterior is equal to the prior) do not contribute to the evidence integral, and hence model comparison does not act against them, awaiting better data.

3.2 Principled application of model selection

I'd like to discuss the inflationary model comparison carried out in Ref. [22] as an example of the application of the above formalism to the problem of deciding which theoretical model is the best description of the available observations. Although the technical details are fairly involved, the underlying idea can be sketched as follows.

The term "inflation" describes a period of exponential expansion of the Universe in the very first instants of its life, some 10^{-32} seconds after the Big Bang, during which the size of the Universe increased by at least 25 orders of magnitude. This huge and extremely fast expansion is required to explain the observed isotropy of the cosmic microwave background on large scales. It is believed that inflation was powered by one or more "scalar fields". The behaviour of the scalar field during inflation is determined by the shape of its potential, which is a real-valued function $V(\phi)$ (where ϕ denotes the value of the scalar field). The detailed shape of $V(\phi)$ controls the duration of inflation, but also the spatial distribution of inhomogeneities (perturbations) in the distribution of matter and radiation which emerge from inflation. It is from those perturbations that galaxies and cluster form out of gravitational collapse. Hence the shape of the scalar field can be constrained by observations of the large scale structures of the Universe and of the CMB anistropies.

Theories of physics beyond the Standard Model motivate certain functional forms of $V(\phi)$, which however typically have a number of free parameters, Ψ. The

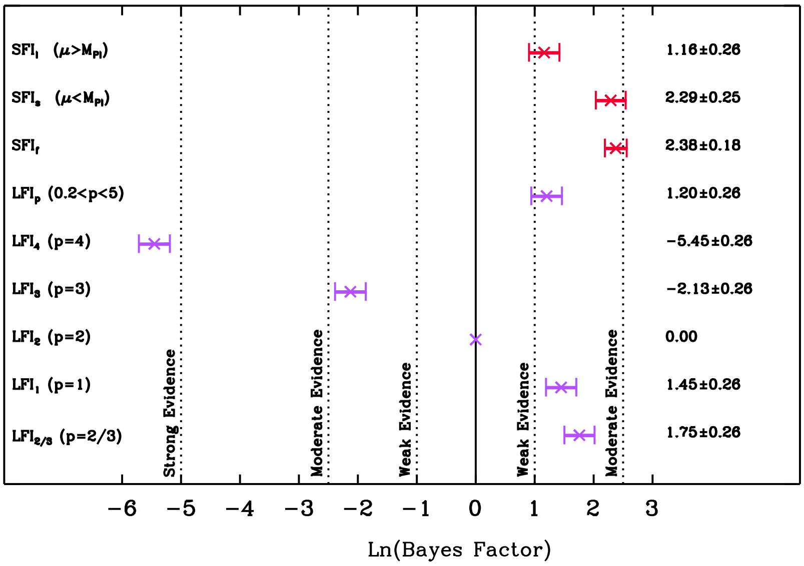

Fig. 2 Results of Bayesian model comparison between 9 inflationary models (vertical axis), subdivided in two categories (SFI models and LFI models), from Ref. [22]. Errorbars reflect the 68% uncertainty on the value of the Bayes factor from the numerical evaluation.

fundamental model selection question is to use cosmological observations to discriminate between alternative models for $V(\phi)$ (and hence alternative fundamental theories). The major obstacle to this programme is that very little if anything at all is known *a priori* about the free parameters Ψ describing the inflationary potential. What is worse, such parameters can assume values across several orders of magnitude, according to the theory. Hence the Occam's razor effect of Bayesian model comparison can vary in a very significant way depending on the prior choices for Ψ. Furthermore, a non-linear reparameterization of the problem (which leaves the physics invariant) does in general change the Occam's razor factor, and hence the model comparison result.

In Ref. [22] a first attempt was made to tackle inflationary model selection from a principled point of view. The main result of the analysis is shown in Fig. 2, which presents the Bayes factors between models (suitably normalized w.r.t. a reference model, here the so-called LFI_2 model). Two classes of models for $V(\phi)$ have been considered, namely so-called Small Field Inflation (SFI) models and Large Field Inflation (LFI) models. The two classes of model differ in the parameterized form of $V(\phi)$, and have different sets of parameters, differing in dimensionality, as well. Within each class of models, sub-classes are defined (denoted by subscripts in Fig. 2) based on theoretical considerations, e.g. by fixing some of the parameters to certain values. The priors on the models' parameters have been chosen based on theoretical considerations of possible values achievable under each class of models. Typical priors are uniform on the log of the parameter (to reflect indifference w.r.t. the characteristic scale of the quantity), within a range chosen as a reflection of physical model building. The models' priors are chosen in such a way to lead to non-committal priors for the two classes as a whole, i.e. $p(\mathrm{SFI}) = p(\mathrm{LFI}) = 1/2$.

Fig. 2 shows that some models in the LFI class are fairly strongly disfavoured by the data (e.g., LFI_3 and LFI_4), while the model comparison is inconclusive in most other cases. One finds that the posterior probability for the SFI model class evaluates to $p(\mathrm{SFI}|d) \approx 0.77$. Therefore, the probability of the SFI class has increased from 50% in the prior to about 77% in the posterior, signalling a weak preference for this type of models in the light of the data.

3.3 Bayesian model averaging

Bayesian model averaging represents the third level of Bayesian inference – incorporating model uncertainty (level 2) into parameter inferences (level 1). The idea is to average the posterior distribution for the parameters of interest over the space of available models, with a weight given by the models' posterior probability:

$$p(\Theta|d) = \sum_{\mathscr{M}} p(\Theta|d,\mathscr{M})p(\mathscr{M}|d)\,. \tag{14}$$

Of course, the above *caveats* about the choice of prior for model selection apply equally to model averaging. An interesting consequence of Bayesian model averaging is that in certain cases model averaged parameter constraints can be tighter than non-model averaging ones, a consequence of the concentration of posterior probability onto simpler models due to the Occam's razor effect. We illustrate this with the example of model averaged constraints on the curvature parameter, a problem recently investigated in Ref. [23] (for applications of Bayesian model averaging to the dark energy equation of state, see [24]; to the scalar spectral index, see [25] and to weak lensing and Sunyaev-Zel'dovich effect data, see [26]).

In the Friedmann-Robertson-Walker (FRW) Universe there are only three discrete possibilities for the underlying geometry, namely flat, open or closed. The amount of curvature is usually characterized by the curvature parameter Ω_κ: if $\Omega_\kappa < 0$ the geometry of spatial sections is spherical (i.e., the Universe is closed) and the Universe has a finite size. If instead $\Omega_\kappa > 0$ the geometry is hyperbolic (i.e., the Universe is open), while for $\Omega_\kappa = 0$ spatial sections are flat. In both the two latter cases, the spatial extent of the Universe is infinite. Limits on the value of Ω_κ can be derived in a geometrical way by observing the angular size subtended by cosmological features of known physical length, such as the acoustic peaks in the cosmic microwave background (CMB) and the corresponding baryonic acoustic oscillations (BAO) in the distribution of large scale structures. Furthermore, type Ia supernovae (SNIa) can be used as standard candles to determine the luminosity distance as a function of redshift. A combination of these three probes has been succesfully used to set very tight limits to the curvature parameter, which is now constrained at better than the $\sim 10^{-3}$ level. For example, [27] find $\Omega_\kappa = -0.0057^{+0.0066}_{-0.0068}$ at 68 % CL, employing a combination of WMAP7, BAO [28] and SNIa data [29]. Impressive as such limits are, they *assume* the Universe to be curved, and carry out parameter inference on the quantity describing curvature. A different methodological perspective is required to go beyond that assumption: model-averaged limits on the curvature of the Universe, fully accounting for the uncertainty in selecting the correct model for the FRW Universe. Given current data, flat models are preferred by Bayesian model selection from an Occam's razor perspective, and therefore most of the probability mass becomes concentrated in models with vanishing spatial curvature. However, this "concentration of probability" effect remains quite strongly dependent on the prior chosen on the curvature parameter (which controls the strength of the Occam's razor). A choice of prior based on requiring consistency with basic observational properties of the Universe (such as the age of the oldest objects, the so-called "Astronomer's prior") leads to a posterior probability for a flat Universe of 98.6%, while a prior based on inflationary consideration (the "curvature scale prior") leads to a much reduced probability of only about 46%. As in any good Bayesian analysis, examining the effect of a reasonable change of prior remains paramount.

The model averaged constraints on Ω_κ for those two choices of priors are depicted in Fig. 3. Even the most conservative prior choice gives model-averaged constraints on curvature that are a factor of ~ 2 better than non model-averaged intervals. A more aggressive prior choice (the Astronomer's prior) leads to an improvement in the constraints on Ω_κ by a factor ~ 100, giving $|\Omega_\kappa| \leq 2 \times 10^{-4}$ at 99%. The same formalism can be used to derive model averaged constraints on the size of the Universe, which is robustly constrained to encompass $N_U \gtrsim 251$ Hubble spheres, an improvement of a factor ~ 40 on previous constraints. Finally, the radius of curvature of spatial section is found to be $R_c \gtrsim 42$ Gpc.

4 Open challenges and conclusions

I conclude by listing what I think are some of the open questions and outstanding challenges in the application of Bayesian model selection to cosmological model building.

- **Is Bayesian model selection always applicable?** The Bayesian model comparison approach as applied to cosmological and particle physics problems has been strongly criticized by some authors. E.g., George Efstathiou [30] and Bob Cousins [31] pointed out (in different contexts) that often insufficient attention is given to the selection of models and of priors, and that this might lead to posterior model probabilities which are largely a function of one's unjustified assumptions. This draws attention to the difficult question of how to choose priors on phenomenological parameters, for which theoretical reasoning offers poor or no guidance (as in the inflationary model comparison example above).

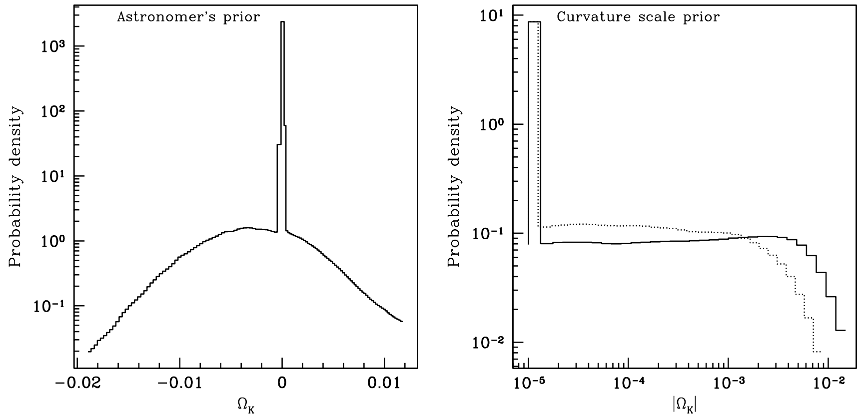

Fig. 3 Model-averaged marginal posterior probability distribution for the curvature parameter, assuming the Astronomers' prior (left panel) and the Curvature scale prior (right panel) for the curvature parameter. In the right panel, the solid line applies to closed Universes ($\Omega_\kappa < 0$), while the dotted line to open Universes ($\Omega_\kappa > 0$). The peaks represent the Dirac delta function encompassing the probability mass associated with flat models, a concentration of probability effect coming from Occam's razor. From [23].

- **How do we deal with Lindley's paradox?** It is simple to construct examples of situations where Bayesian model comparison and classical hypothesis testing disagree (Lindley's paradox [32]). This is not surprising, as frequentist hypothesis testing and Bayesian model selection really ask different questions of the data [2]. As Louis Lyons aptly put it: "Bayesians address the question everyone is interested in by using assumptions no–one believes, while frequentists use impeccable logic to deal with an issue of no interest to anyone" [33]. However, such a disagreement is likely to occur in situations where the signal is weak, which are precisely the kind of "frontier science" cases which are the most interesting ones (e.g., discovery claims). Is there a way to evaluate e.g. the loss function from making the "wrong" decision about rejecting/accepting a model?
- **How do we assess the completness of the set of known models?** Bayesian model selection always returns a best model among the ones being compared, even though that model might be a poor explanation for the available data. Is there a principled way of constructing an *absolute* scale for model performance in a Bayesian context? Recently, the notion of Bayesian doubt, introduced in [34], has been used to extend the power of Bayesian model selection to the space of unknown models in order to test our paradigm of a ΛCDM cosmological model. It would be useful to have feedback from the statistics community about the validity of such an approach, and whether similar tools have already been developed in other contexts.
- **Is there such a thing as a "correct" prior?** In fundamental physics, models and parameters (and their priors) are supposed to represent (albeit in an ideal-

ized way) the real world, i.e., they are not simply useful representation of the data (as they are in other statistical problems, e.g. as applied to social sciences). In this sense, one could imagine that there exist a "correct" prior for e.g. the parameters Θ of our cosmological model, which could in principle be derived from fundamental theories such as string theory (e.g., the distribution of values of cosmological parameters across the landscape of string theory). This raises interesting statistical questions about the relationship between physics, reality and probability.

Acknowledgements I would like to thank the organizers of the GREAT Workshop on Astrostatistics and Data Mining in Astronomical Databases conference for inviting me to talk to this very stimulating audience. Many thanks to several colleagues with whom I have had a number of stimulating discussions over the years: J. Berger., R. Cousin, K. Cranmer, A. Jaffe, M. Kunz, O. Lahav, A. Liddle, T. Loredo, L. Lyons and D. van Dyk.

References

1. T. J. Loredo, From Laplace to Supernova SN 1987A: Bayesian Inference in Astrophysics, in T. Fougere (Editor) *Maximum-Entropy and Bayesian Methods*, Available from: http://bayes.wustl.edu/gregory/articles.pdf (accessed Jan 15 2008) (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990), pp. 81–142.
2. T. Sellke, M. Bayarri and J. O. Berger, American Statistician **55** 62–71 (2001).
3. R. Trotta, Contemp. Phys. **49**, 71 (2008)
4. M. Kunz, R. Trotta and D. Parkinson, Phys. Rev. **D74** 023503 (2006).
5. *Bayesian Methods in Cosmology*, M. Hobson *et al.*, (Eds), Cambridge University Press (2010).
6. H. Jeffreys, *Theory of probability*, 3rd edn , Oxford Classics series (reprinted 1998) (Oxford University Press, Oxford, UK, 1961).
7. A. Slosar *et al.*, Mon. Not. Roy. Astron. Soc. **341** L29 (2003).
8. M. Beltran, J. Garcia-Bellido, J. Lesgourgues, A. R. Liddle, *et al.*, Phys. Rev. **D71** 063532 (2005).
9. M. Bridges, A. N. Lasenby and M. P. Hobson, Mon. Not. Roy. Astron. Soc. **369** 1123–1130 (2006).
10. M. Kilbinger *et al.*, preprint: arXiv:0912.1614.
11. J. Skilling, Nested sampling, in R. Fischer, R. Preuss and U. von Toussaint (Eds) *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, 735 (Amer. Inst. Phys. conf. proc. 2004), pp. 395–405.
12. J. Skilling, Bayesian Analysis **1** 833–861 (2006).
13. B. A. Bassett, P. S. Corasaniti and M. Kunz, Astrophys. J. **617** L1–L4 (2004).
14. P. Mukherjee, D. Parkinson and A. R. Liddle, Astrophys. J. **638** L51–L54 (2006).
15. R. Shaw, M. Bridges and M. P. Hobson, Mon. Not. Roy. Astron. Soc. **378** 1365–1370 (2007).
16. F. Feroz and M. P. Hobson, Mon. Not. Roy. Astron. Soc., 384, 2, 449-463 (2008).
17. F. Feroz, M. P. Hobson, M. Bridges, [arXiv:0809.3437 [astro-ph]].
18. N. Chopin and C. P. Robert, Contemplating Evidence: properties, extensions of, and alternatives to Nested Sampling.. (Available as preprint from: http://www.crest.fr/pageperso/Nicolas.Chopin/Nicolas.Chopin.htm. Accessed Jan 2008.).
19. I. Verdinelli and L. Wasserman, J. Amer. Stat. Assoc. **90** 614–618 (1995).
20. R. Trotta, Mon. Not. Roy. Astron. Soc. **378** 72–82 (2007).
21. A. F. Heavens, T. D. Kitching and L. Verde, Mon. Not. Roy. Astron. Soc. **380** 1029–1035 (2007).

22. J. Martin, C. Ringeval, R. Trotta, Phys. Rev. D in press (2011), arXiv:1009.4157.
23. M. Vardanyan, R. Trotta, J. Silk, Mon. Not. Roy. Astron. Soc. in print, preprint: arXiv:1101.5476.
24. A. R. Liddle, P. Mukherjee, D. Parkinson, Y. Wang, Phys. Rev. **D74**, 123506 (2006).
25. D. Parkinson, A. R. Liddle, Phys. Rev. **D82**, 103533 (2010).
26. Marshall P. J., Hobson M. P., Slosar A., 2003, Mon. Not. Roy. Astron. Soc., 346, 489
27. Komatsu E., et al., Astrophys. J. Supp., 192, 18 (2011).
28. Reid B. A., et al., 2010, Mon. Not. Roy. Astron. Soc., 401, 2148
29. Hicken M., et al., 2009, Astrophys. J., 700, 1097
30. G. Efstathiou, preprint: arXiv:0802.3185.
31. R. D. Cousins, Phys. Rev. Lett. **101**, 029101 (2008).
32. D. Lindley, Biometrika **44** 187–192 (1957).
33. L. Lyons, A particle physicist's perspective on astrostatistics, in *Proceedings of the Statistical Challenges in Modern Astronomy IV Conference*, 371, Pennsylvania State University, Pennsylvania, USA, 12-15 June 2006 (Astronomical Society of the Pacific, San Francisco, 2007), pp. 361–372.
34. M. C. March, G. D. Starkman, R. Trotta, P. M. Vaudrevange, Mon. Not. Roy. Astron. Soc. **410**, 2488-2496 (2011).
35. M. Tegmark, JCAP **0504**, 001 (2005)
36. R. Trotta, F. Feroz, M. P. Hobson, L. Roszkowski and R. Ruiz de Austri, JHEP **0812**, 024 (2008)
37. M. Vardanyan, R. Trotta and J. Silk, Mon. Not. Roy. Astron. Soc. **397**, 431 (2009)