

Comment on “Bayesian evidence: can we beat MultiNest using traditional MCMC methods”, by Rutger van Haasteren (arXiv:0911.2150)

F. Feroz^{1,*}, M.P. Hobson¹ and R. Trotta²

¹ *Astrophysics Group, Cavendish Laboratory, JJ Thomson Avenue, Cambridge CB3 0HE, UK*

² *Astrophysics Group, Imperial College, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK*

Accepted —. Received —; in original form 8 January 2010

ABSTRACT

In arXiv:0911.2150, Rutger van Haasteren seeks to criticize the nested sampling algorithm for Bayesian data analysis in general and its MULTINEST implementation in particular. He introduces a new method for evidence evaluation based on the idea of Voronoi tessellation and requiring samples from the posterior distribution obtained through MCMC based methods. He compares its accuracy and efficiency with MULTINEST, concluding that it outperforms MULTINEST in several cases. This comparison is completely unfair since the proposed method can not perform the complete Bayesian data analysis including posterior exploration and evidence evaluation on its own while MULTINEST allows one to perform Bayesian data analysis end to end. Furthermore, their criticism of nested sampling (and in turn MULTINEST) is based on a few conceptual misunderstandings of the algorithm. Here we seek to set the record straight.

Key words: methods: data analysis – methods: statistical

1 INTRODUCTION

In a recent paper (van Haasteren 2009), Rutger van Haasteren has criticized the MULTINEST algorithm for Bayesian analysis and suggested another method to calculate the Bayesian evidence with the claim that it significantly outperforms MULTINEST both in terms of accuracy and computational accuracy. Our aim in this short note is to highlight the conceptual misunderstandings and the false premise on which the comparison is based.

The outline of the paper is as follows. In Sec. 2 we give an introduction to Bayesian inference and describe the nested sampling algorithm and its implementation in MULTINEST package in Sec. 3. In Sec. 4 we give an account of the conceptual misunderstandings in van Haasteren 2009 which are the basis of the attack mounted by the author on MULTINEST. Finally, our conclusions are presented in Sec. 5.

2 BAYESIAN INFERENCE

Bayesian inference methods provide a consistent approach both to the estimation of a set of parameters Θ in a model (or hypothesis) H for the data \mathbf{D} and to the evaluation of the relative merits of different models for the data (see Trotta (2008) for a review of

Bayesian methods in cosmology and astrophysics). Bayes’ theorem states that

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

where $\Pr(\Theta|\mathbf{D}, H) \equiv P(\Theta)$ is the posterior probability distribution of the parameters, $\Pr(\mathbf{D}|\Theta, H) \equiv \mathcal{L}(\Theta)$ is the likelihood, $\Pr(\Theta|H) \equiv \pi(\Theta)$ is the prior distribution, and $\Pr(\mathbf{D}|H) \equiv \mathcal{Z}$ is the Bayesian evidence.

Bayesian evidence is the factor required to normalise the posterior over Θ :

$$\mathcal{Z} = \int \mathcal{L}(\Theta)\pi(\Theta)d^D\Theta, \quad (2)$$

where D is the dimensionality of the parameter space. Since the Bayesian evidence is independent of the parameter values Θ , it is usually ignored in parameter estimation problems and the posterior inferences are obtained by exploring the un-normalized posterior using standard MCMC sampling methods.

Bayesian parameter estimation has been used quite extensively in a variety of astronomical applications, including gravitational wave astronomy, although standard MCMC methods, such as the basic Metropolis–Hastings algorithm or the Hamiltonian sampling technique (see e.g. Mackay 2003), can experience problems in sampling efficiently from a multi-modal posterior distribution or one with large (curving) degeneracies between parameters. Moreover, MCMC methods often require careful tuning of the proposal

* E-mail: f.feroz@mrao.cam.ac.uk

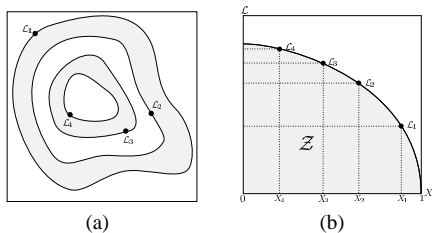


Figure 1. Cartoon illustrating (a) the posterior of a two dimensional problem; and (b) the transformed $\mathcal{L}(X)$ function where the prior volumes X_i are associated with each likelihood \mathcal{L}_i .

distribution to sample efficiently, and testing for convergence can be problematic.

In order to select between two models H_0 and H_1 one needs to compare their respective posterior probabilities given the observed data set \mathbf{D} , as follows:

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

where $\Pr(H_1)/\Pr(H_0)$ is the prior probability ratio for the two models, which can often be set to unity but occasionally requires further consideration (see, for example, Feroz et al. 2009e, 2008 for the cases where the prior probability ratio should not be set to unity) and the ratio of the evidences $\frac{\mathcal{Z}_1}{\mathcal{Z}_0}$ is called the Bayes factor between the two models. It can be seen from Eq. (3) that the Bayesian evidence plays a central role in Bayesian model selection. As the average of likelihood over the prior, the evidence automatically implements Occam’s razor: a simpler theory which agrees well enough with the empirical evidence is preferred. A more complicated theory will only have a higher evidence if it is significantly better at explaining the data than a simpler theory (e.g., Liddle (2004); Trotta (2007a)).

The evaluation of the Bayesian evidence involves the multi-dimensional integral (Eq. 2) and thus presents a challenging numerical task. Standard techniques like thermodynamic integration Ó Ruanaidh & Fitzgerald (1996) are usually extremely computationally expensive, which makes evidence evaluation typically at least an order of magnitude more costly than parameter estimation. Recently, a estimation of the evidence using population Monte Carlo techniques has been successfully implemented in the cosmological context Kilbinger et al. (2009). Some fast approximate methods have been used for evidence evaluation, such as treating the posterior as a multivariate Gaussian centred at its peak (see, for example, Hobson et al. 2002), but this approximation is clearly a poor one for highly non-Gaussian and multi-modal posteriors. A computationally cheap and accurate method is the Savage-Dickey density ratio Trotta (2007a,b); Vardanyan et al. (2009), which however can only be used to compute the Bayes factor between nested models. Various alternative information criteria for model selection are discussed in Liddle (2004); Liddle (2007), but the evidence remains the preferred method.

3 NESTED SAMPLING AND THE MULTINEST ALGORITHM

In this section we briefly review the Nested sampling algorithm and the MULTINEST implementation. Further details can be found in Skilling (2004); Feroz & Hobson (2008); Feroz et al. (2009c).

3.1 Nested Sampling

Nested sampling Skilling (2004) is a Monte Carlo method targeted at the efficient calculation of the evidence, but also produces posterior inferences as a by-product. It calculates the evidence by transforming the multi-dimensional evidence integral into a one-dimensional integral that is easy to evaluate numerically. This is accomplished by defining the prior volume X as $dX = \pi(\Theta)d^D\Theta$, so that

$$X(\lambda) = \int_{\mathcal{L}(\Theta) > \lambda} \pi(\Theta)d^D\Theta, \quad (4)$$

where the integral extends over the region(s) of parameter space contained within the iso-likelihood contour $\mathcal{L}(\Theta) = \lambda$. The evidence integral, Eq. (2), can then be written as

$$\mathcal{Z} = \int_0^1 \mathcal{L}(X)dX, \quad (5)$$

where $\mathcal{L}(X)$, the inverse of Eq. (4), is a monotonically decreasing function of X . Thus, if one can evaluate the likelihoods $\mathcal{L}_i = \mathcal{L}(X_i)$, where X_i is a sequence of decreasing values,

$$0 < X_M < \dots < X_2 < X_1 < X_0 = 1, \quad (6)$$

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

$$\mathcal{Z} = \sum_{i=1}^M \mathcal{L}_i w_i, \quad (7)$$

where the weights w_i for the simple trapezium rule 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 $\mathcal{L}(X)$ is shown in Fig. 1.

The summation in Eq. (7) is performed as follows. The iteration counter is first set to $i = 0$ and N ‘active’ (or ‘live’) samples are drawn from the full prior $\pi(\Theta)$, so the initial prior volume is $X_0 = 1$. The samples are then sorted in order of their likelihood and the smallest (with likelihood \mathcal{L}_0) is removed from the active set (hence becoming ‘inactive’) and replaced by a point drawn from the prior subject to the constraint that the point has a likelihood $\mathcal{L} > \mathcal{L}_0$. The corresponding prior volume contained within the iso-likelihood contour associated with the new live point will be a random variable given by $X_1 = t_1 X_0$, where t_1 follows the distribution $\Pr(t) = Nt^{N-1}$ (i.e., the probability distribution for the largest of N samples drawn uniformly from the interval $[0, 1]$). At each subsequent iteration i , the removal of the lowest likelihood point \mathcal{L}_i in the active set, the drawing of a replacement with $\mathcal{L} > \mathcal{L}_i$ and the reduction of the corresponding prior volume $X_i = t_i X_{i-1}$ are repeated, until the entire prior volume has been traversed. The algorithm thus travels through nested shells of likelihood as the prior volume is reduced. The mean and standard deviation of $\log t$, which dominates the geometrical exploration, are:

$$E[\log t] = -1/N, \quad \sigma[\log t] = 1/N. \quad (8)$$

Since each value of $\log t$ is independent, after i iterations the prior volume will shrink down such that $\log X_i \approx -(i \pm \sqrt{i})/N$. Thus, one takes $X_i = \exp(-i/N)$.

The nested sampling algorithm is terminated when the evidence has been computed to a pre-specified precision. The evidence that could be contributed by the remaining live points is estimated as $\Delta\mathcal{Z}_i = \mathcal{L}_{\max} X_i$, where \mathcal{L}_{\max} is the maximum-likelihood value of the remaining live points, and X_i is the remaining prior volume. The algorithm terminates when $\Delta\mathcal{Z}_i$ is less than a user-defined value (we use 0.5 in log-evidence).

Once the evidence \mathcal{Z} is found, posterior inferences can be easily generated using the final live points and the full sequence of discarded points from the nested sampling process, i.e., the points with the lowest likelihood value at each iteration i of the algorithm. Each such point is simply assigned the probability weight

$$p_i = \frac{\mathcal{L}_i w_i}{\mathcal{Z}}. \quad (9)$$

These samples can then be used to calculate inferences of posterior parameters such as means, standard deviations, covariances and so on, or to construct marginalised posterior distributions.

3.2 The MULTINEST Algorithm

The most challenging task in implementing nested sampling is to draw samples from the prior within the hard constraint $\mathcal{L} > \mathcal{L}_i$ at each iteration i . The MULTINEST algorithm Feroz & Hobson (2008); Feroz et al. (2009c) tackles this problem through an ellipsoidal rejection sampling scheme. The live point set is enclosed within a set of (possibly overlapping) ellipsoids and a new point is then drawn uniformly from the region enclosed by these ellipsoids. The ellipsoidal decomposition of the live point set is chosen to minimize the sum of volumes of the ellipsoids. The ellipsoidal decomposition is well suited to dealing with posteriors that have curving degeneracies, and allows mode identification in multi-modal posteriors. If there are subsets of the ellipsoid set that do not overlap with the remaining ellipsoids, these are identified as a distinct mode and subsequently evolved independently.

The MULTINEST algorithm has proven very useful for tackling inference problems in cosmology and particle physics (see e.g. Sekiguchi et al. 2009; Bridges et al. 2009; Vegetti et al. 2009; Sollom et al. 2009; AbdusSalam et al. 2009b,a; Feroz et al. 2009e, 2008, 2009d; Feroz et al. 2008a; Trotta et al. 2008; López-Fogliani et al. 2009; Raklev & White 2009; Trotta et al. 2009; Roszkowski et al. 2009) typically showing two orders of magnitude improvement in efficiency over conventional techniques. More recently, it has been shown to perform well as a search tool for gravitational wave data analysis (Feroz et al. 2009b,a).

4 DETAILED CRITIQUE OF VAN HAASTEREN (2009)

In van Haasteren (2009), the author proposes a method based on Voronoi tessellation to calculate the Bayesian evidence (Eq. 2) as follows:

$$\mathcal{Z} = \sum_i \mathcal{L}_i O_i, \quad (10)$$

where the index i iterates over the samples (may be obtained through an MCMC algorithm) in the parameter space and O_i is the area of the Voronoi cell occupied by the i^{th} sample. Although this approximation converges to the true evidence value, it can not be used in practice because of the high computational cost involved in the calculation of Voronoi tessellation. In order to overcome this problem, the author suggests to concentrate on a small subset of samples F_t , around the peak of the distribution, occupying volume V_t and exploits the following relationship:

$$V_t = \sum_{\Theta_i \in F_t} \frac{\alpha}{\mathcal{L}_i}, \quad (11)$$

where α is a proportionality constant to be determined. In order to calculate α , the author then suggests to make a Gaussian approximation around the peak so that the volume V_t can be calculated

as,

$$V_t = \frac{r^D \pi^{D/2}}{\Gamma(1 + \frac{D}{2})} \sqrt{|\mathbf{C}|}, \quad (12)$$

where D is the dimensionality of the problem, \mathbf{C} the covariance matrix and r is the Mahalanobis distance from the peak of the point with the lowest likelihood value inside the region F_t . Once α has been calculated using Eqs.(11) and (12), The evidence can be calculated as:

$$\mathcal{Z} = N\alpha, \quad (13)$$

where N is the total number samples inside the region F_t . The author then applies this method to a few toy problems and compares it with MULTINEST, attempting to show that it outperforms MULTINEST in terms of accuracy as well as computational efficiency.

We believe that the comparison of MULTINEST to the method proposed in van Haasteren (2009) is completely unfair, as it makes a very strong assumption about the availability of suitable posterior samples. Moreover, the criticism of nested sampling algorithm is due to author's misunderstanding about the method. We discuss these in detail now.

4.1 Nested Sampling solves the full inference problem

The method proposed in van Haasteren (2009) and briefly summarized in the previous section requires the region F_t to be sufficiently and adequately populated by MCMC samples but does not propose an MCMC algorithm to satisfy this condition. Therefore, the proposed method does not provide a complete solution for Bayesian inference problem, but it actually makes the strong assumption that such a solution has already been implemented using another technique. For many problems of interest (e.g. problem exhibiting strong, curving degeneracies and/or multi-modality) adequately sampling the parameter space is a difficult problem which often proves to be a stumbling block even for the parameter inference step. The algorithm proposed by van Haasteren (2009) assumes that this part of the problem has already been solved, which strongly limits the usefulness of the suggested method if no efficient way of gathering posterior samples is put forward.

The MULTINEST algorithm, on the other hand, provides a means to carry out both parameter exploration and the evidence evaluation in an efficient and robust way. This has been widely demonstrated by applying it very successfully to various real inference problems in astrophysics, cosmology and particle physics phenomenology (see e.g. Sekiguchi et al. 2009; Bridges et al. 2009; Vegetti et al. 2009; Sollom et al. 2009; AbdusSalam et al. 2009b,a; Feroz et al. 2009e, 2008, 2009d; Feroz et al. 2008a; Feroz et al. 2009b,a; Trotta et al. 2008; López-Fogliani et al. 2009; Raklev & White 2009).

The eggbox problem (discussed in Sec. 5.2 of van Haasteren 2009) is a particular example of what we believe is an unfair comparison. The author assumes that a suitable trick can be found enabling the MCMC algorithm to explore all the modes adequately so that his proposed method can be applied to this highly multi-modal problem, while MULTINEST not only finds all the modes without making any assumptions but also calculates the evidence accurately.

A fair comparison in our opinion would require the specification of another algorithm capable of providing not only the means to calculate the evidence but also to explore the parameter space without making too many assumptions. In our view, the compari-

son between the method of van Haasteren (2009) and MULTINEST is fundamentally unfair, as the latter has a much wider applicability and solves the full inference problem from scratch.

One of the most attractive features of MULTINEST is its general applicability for moderately dimensional but highly complex problems. It is completely application-independent in the sense that the specific problem being tackled enters only through the likelihood computation, and does not change how the live point set is updated. It makes no assumptions about the number and nature of modes. Even if a perfect MCMC sampler were available, the method proposed in van Haasteren (2009) although can be quite useful for uni-modal Gaussian problems, would nonetheless fail for highly non-Gaussian problems. The Gaussian-shell problem discussed in Sec. 6.2 of Feroz et al. (2009c) would perhaps be the most obvious example. Comparing MULTINEST with the proposed method on problems ideally suited to the proposed method is another reason why we think the comparison is completely unfair.

4.2 Misconceptions about Nested Sampling

In the introduction, the author argues that MULTINEST (and consequently nested sampling) algorithm by design, samples from the prior distribution and not the posterior distribution, and consequently suffers more with the ‘curse of dimensionality’ than the traditional MCMC based algorithms. This statement ignores the fact that nested sampling does not sample from the prior distribution blindly, it samples from the prior within the hard constraint $\mathcal{L} > \mathcal{L}_i$ at each iteration i . As discussed in Sec. 3.1, this hard-edge sampling scheme results in exponential decrease in the prior volume X_i occupied by the live points at the i^{th} iteration with the expected value of X_i given as,

$$\langle X_i \rangle = \exp(-i/N), \quad (14)$$

N being the total number of live points. This allows the algorithm to reach highly localized regions of the parameter space with large likelihood values in a reasonable number of iterations. We can not see why such a sampling scheme would suffer from the ‘curse of dimensionality’ more than the traditional MCMC schemes.

In our opinion, the information content H , given as follows,

$$H = \int \log \left(\frac{dP}{dX} \right) dX, \quad (15)$$

where P denotes the posterior, is more important than the dimensionality of the problem. Most of the contribution to the evidence value usually comes from the iterations around the maximum likelihood point, which occurs in the region with prior volume $X \approx e^{-H}$ and therefore because of the exponential shrinkage of the prior volume, one could argue that nested sampling has an inherent advantage over MCMC schemes. It should be noted that nested sampling framework itself leaves it to the user to design a scheme to sample from the iso-likelihood contour and it is certainly possible to come up with highly inefficient schemes, to sample a new point with $\mathcal{L} > \mathcal{L}_i$, suffering severely with the ‘curse of dimensionality’.

In Sec. 4.2, the author claims that nested sampling generates samples from the whole of the parameter space, rather than from the posterior distribution, and therefore it can never reach the efficiency achieved by the traditional MCMC based methods. This statement ignores two key points. First, as discussed earlier in this section, nested sampling does not sample from the whole of the prior distribution, but instead it samples from the hard-edge region inside the prior whose volume is reduced exponentially with each

iteration. Secondly, the samples generated by a nested sampling algorithm are not all equally weighted as the ones generated through an MCMC method are. As discussed in Sec. 3.1, one needs to assign a probability weight given by Eq. (9) to the point with lowest likelihood value at each iteration of the nested sampling algorithm. In Fig. 3 of van Haasteren (2009), the author shows a scatter plot of 40,000 samples obtained through a nested sampling algorithm. The author does not mention whether he has plotted simply the points with the lowest likelihood values at each iteration or the probability weights of these points have also been taken into account and therefore we are unable to comment on the accuracy of the figure.

4.3 Curse of Dimensionality

The author repeatedly attacks nested sampling algorithm in general and MULTINEST in particular, saying that it suffers severely from the ‘curse of dimensionality’. We have already discussed in the previous section that as far as the general nested sampling framework is concerned, this is not true.

MULTINEST implements the nested sampling algorithm through an ellipsoidal rejection sampling scheme to sample uniformly from the iso-likelihood contour as discussed in Sec. 3.2. It is well known that all rejection sampling schemes are highly inefficient for high dimensional problems (see e.g. Mackay 2003). MULTINEST was designed to work with problems with moderately high number of dimensions and it has proven highly successful to deal with highly multi-modal and complex problems in cosmology and particle physics phenomenology. It should be noted that nested sampling not only provides a way to evaluate the Bayesian evidence accurately but with a clever algorithm to sample from the hard constraint, it also provides a solution to deal with highly degenerate and multi-modal problems. A very good example is the use of MULTINEST in gravitational wave astronomy (see e.g. Feroz et al. 2009b,a) where the problems are inherently multi-modal and so far MULTINEST has mainly been used as a parameter exploration tool with great deal of success.

For parameter exploration, MULTINEST works with reasonable efficiency up to $\sim 100D$ beyond which the efficiency drops appreciably as expected. Accurate evidence evaluation requires adequate sampling from the whole of parameter space and even slight inaccuracies in estimating the iso-likelihood region through the ellipsoidal decomposition can result in large inaccuracies and therefore MULTINEST can calculate the evidence value accurately with reasonable efficiency for problems up to $\sim 50D$. In order to demonstrate this, we ran MULTINEST on an n -dimensional ellipsoidal Gaussian with likelihood¹,

$$\mathcal{L} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{\theta_i - 0.5}{2\sigma_i}\right), \quad (16)$$

with

$$\sigma_i = 0.001i. \quad (17)$$

We set uniform prior $\mathcal{U}(0, 1)$ for all the parameters so that the analytical log-evidence value is 0.0 regardless of the dimensionality of the problem. We used 1,000 live points with target efficiency e set to 1.0 for the 2D problem and reducing it to 0.01 for the 32D

¹ We would have liked to test MULTINEST on the same toy problem described in Sec. 5.1 of van Haasteren (2009) but we were unable to do so as the author did not describe the prior distribution he used and hence we chose a similar but not exactly the same problem.

n	N_{like}	$\log(\mathcal{Z})$	H
2	14, 184	0.04 ± 0.10	10.00
4	26, 033	-0.07 ± 0.14	19.60
8	107, 876	-0.24 ± 0.18	32.40
16	651, 345	0.11 ± 0.24	57.60
32	14, 134, 227	0.40 ± 0.31	96.10

Table 1. The log-evidence ($\log(\mathcal{Z})$) and information content (H) values obtained by the MULTINEST algorithm when applied to the problem described in Eq. (16). The analytical $\log(\mathcal{Z})$ is 0.0 regardless of dimensionality n .

problem. A value of $e = 0.3$ was suggested in Feroz et al. (2009c) for the standard CMB data analysis, but it was accompanied by the caveat that the user should check that the evidence value is consistent when e is lowered. We list the number of likelihood evaluations, recovered $\log(\mathcal{Z})$ and information content H in Table 1. These results clearly show that MULTINEST is able to correctly evaluate the evidence values even for the 32D problem with the posterior occupying only $e^{-96.10}$ of the prior volume, although the efficiency does drop appreciably with the increase in dimensionality of the problem. This drop in efficiency is mainly due to the exponential increase in the information content.

We should also mention that for a Gaussian problem, we have been able to calculate the evidence value accurately up to $\sim 1000D$ using nested sampling with a Hamiltonian sampling scheme to sample from the hard-constraint. This method is still under development and we would present the results in a forthcoming publication.

5 CONCLUSIONS

MULTINEST has proven to be a very useful and powerful tool to carry out Bayesian inference for a wide variety of problems in cosmology and particle physics phenomenology as well as for general moderately dimensional inference problems. While the method proposed in van Haasteren 2009 to calculate the Bayesian evidence is interesting for sufficiently simple problems, it does not have general applicability and relies on the availability of samples distributed according to the posterior distribution. The comparison of this method with MULTINEST is completely unfair as MULTINEST provides the means to perform full Bayesian analysis without making any assumptions about the nature of the problem nor does it rely on the availability of posterior samples, in fact it can be used to provide the posterior samples.

REFERENCES

- AbdusSalam S. S., Allanach B. C., Dolan M. J., Feroz F., Hobson M. P., 2009a, Phys.Rev.D, 80, 035017 [arXiv:0906.0957]
- AbdusSalam S. S., Allanach B. C., Quevedo F., Feroz F., Hobson M., 2009b, ArXiv e-prints [arXiv:0904.2548]
- Bridges M., Feroz F., Hobson M. P., Lasenby A. N., 2009, MNRAS, 400, 1075 [arXiv:0812.3541]
- Feroz F., Allanach B. C., Hobson M., Abdus Salam S. S., Trotta R., Weber A. M., 2008a, Journal of High Energy Physics, 10, 64 [arXiv:0807.4512]
- Feroz F., Gair J. R., Graff P., Hobson M. P., Lasenby A., 2009a, ArXiv e-prints [arXiv:0911.0288]
- Feroz F., Gair J. R., Hobson M. P., Porter E. K., 2009b, Classical and Quantum Gravity, 26, 215003 [arXiv:0904.1544]
- Feroz F., Hobson M. P., 2008, MNRAS, 384, 449 [arXiv:0704.3704]
- Feroz F., Hobson M. P., Bridges M., 2009c, MNRAS, 398, 1601 [arXiv:0809.3437]
- Feroz F., Hobson M. P., Roszkowski L., Ruiz de Austri R., Trotta R., 2009d, ArXiv e-prints [arXiv:0903.2487]
- Feroz F., Hobson M. P., Zwart J. T. L., Saunders R. D. E., Grainge K. J. B., 2009e, MNRAS, 398, 2049 [arXiv:0811.1199]
- Feroz F., Marshall P. J., Hobson M. P., 2008, ArXiv e-prints [arXiv:0810.0781]
- Hobson M. P., Bridle S. L., Lahav O., 2002, MNRAS, 335, 377 [arXiv:astro-ph/0203259]
- Kilbinger M., et al., 2009, ArXiv e-prints [arXiv:0912.1614]
- Liddle A. R., 2004, Mon. Not. Roy. Astron. Soc., 351, L49 [arXiv:astro-ph/0401198]
- Liddle A. R., 2007, Mon. Not. Roy. Astron. Soc., 377, L74 [arXiv:astro-ph/0701113]
- López-Fogliani D. E., Roszkowski L., de Austri R. R., Varley T. A., 2009, Phys.Rev.D, 80, 095013 [arXiv:0906.4911]
- Mackay D. J. C., 2003, Information Theory, Inference and Learning Algorithms. Information Theory, Inference and Learning Algorithms, by David J. C. MacKay, pp. 640. ISBN 0521642981. Cambridge, UK: Cambridge University Press, October 2003.
- Ó Ruanaidh J., Fitzgerald W., 1996, Numerical Bayesian Methods Applied to Signal Processing. Springer Verlag:New York
- Raklev A. R., White M. J., 2009, ArXiv e-prints [arXiv:0911.1986]
- Roszkowski L., Ruiz de Austri R., Trotta R., 2009, ArXiv e-prints [arXiv:0907.0594]
- Sekiguchi T., Ichikawa K., Takahashi T., Greenhill L., 2009, ArXiv e-prints [arXiv:0911.0976]
- Skilling J., 2004, in Fischer R., Preuss R., Toussaint U. V., eds, American Institute of Physics Conference Series Nested Sampling. pp 395–405
- Sollom I., Challinor A., Hobson M. P., 2009, Phys.Rev.D, 79, 123521 [arXiv:0903.5257]
- Trotta R., 2007a, Mon. Not. Roy. Astron. Soc., 378, 72 [arXiv:astro-ph/0504022]
- Trotta R., 2007b, Mon. Not. Roy. Astron. Soc. Lett., 375, L26 [arXiv:astro-ph/0608116]
- Trotta R., 2008, Contemp. Phys., 49, 71 [arXiv:0803.4089]
- Trotta R., de Austri R. R., Heros C. P. d. I., 2009, JCAP, 0908, 034 [arXiv:0906.0366]
- Trotta R., Feroz F., Hobson M., Roszkowski L., Ruiz de Austri R., 2008, Journal of High Energy Physics, 12, 24 [arXiv:0809.3792]
- van Haasteren R., 2009, ArXiv e-prints [arXiv:0911.2150]
- Vardanyan M., Trotta R., Silk J., 2009, Mon. Not. Roy. Astron. Soc., 397, 431 [arXiv:0901.3354]
- Vegetti S., Koopmans L. V. E., Bolton A., Treu T., Gavazzi R., 2009, ArXiv e-prints [arXiv:0910.0760]