Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 3 February 2014 (MN LATEX style file v2.2)

Bayesian Calibrated Significance Levels Applied to the Spectral Tilt and Hemispherical Asymmetry

Christopher Gordon and Roberto Trotta
Oxford University, Astrophysics Department, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK

ABSTRACT

Bayesian model selection provides a formal method of determining the level of support for new parameters in a model. However, if there is not a specific enough underlying physical motivation for the new parameters it can be hard to assign them meaningful priors, an essential ingredient of Bayesian model selection. Here we look at methods maximizing the prior so as to work out what is the maximum support the data could give for the new parameters. If the maximum support is not high enough then one can confidently conclude that the new parameters are unnecessary without needing to worry that some other prior may make them significant. We discuss a computationally efficient means of doing this which involves mapping p–values onto upper bounds of the Bayes factor (or odds) for the new parameters. A p–value of 0.05 (1.96σ) corresponds to odds less than or equal to 5:2 which is below the 'weak' support at best threshold. A p–value of 0.0003 (3.6σ) corresponds to odds of less than or equal to 150:1 which is the 'strong' support at best threshold. Applying this method we find that the odds on the scalar spectral index being different from one are 49:1 at best. We also find that the odds that there is primordial hemispherical asymmetry in the cosmic microwave background are 9:1 at best.

1 INTRODUCTION

When there are several competing theoretical models, Bayesian model selection provides a formal way of evaluating their relative probabilities in light of the data and any prior information available. A common scenario is where a model is being extended by adding new parameters. Then the relative probability of the model with the extra parameters can be compared with that for the original model. This provides a way of evaluating whether the new parameters are supported by the data. Often the original model is "nested" in the new model in that the new model reduces to the original model for specific values of the new parameters. The Bayesian framework automatically implements an Occam's razor effect as a penalization factor for less predictive models – the best model is then the one that strikes the best balance between goodness of fit and economy of parameters (Trotta 2007).

For nested models, the Occam's razor effect is controlled by the volume of parameter space enclosed by the prior probability distributions for the new parameters. The relative probability of the new model can be made arbitrarily small by increasing the broadness of the prior. Often this is not problematical as prior ranges for the new parameters can (and should) be motivated from the underlying theory. For example, in estimating whether the scalar spectral index (n) of the primordial perturbations is equal to one (see Sec. 4), the prior range of the index can be constrained to be $0.8 \lesssim n \lesssim 1.2$ by assuming the perturbations were generated by slow roll inflation. The sensitivity of the model selection result can also be easily investigated for other plausible, physically motivated choice of prior ranges (e.g., Trotta (2007a,b)).

However, there are cases like the asymmetry seen in the WMAP cosmic microwave background (CMB) temperature data (see Sec. 5) where there is not a specific enough model available to place meaningful limits on the prior ranges of the new parameters. This hurdle arises frequently in cases when the new parameters are a phenomenological description of a new effect, only loosely tied to the underlying physics, such as for example expansion coefficients of some series. In these cases, an alternative is to choose the prior on the new parameters in such a way as to maximise the probability of the new model, given the data. If, even under this best case scenario, the new model is not significantly more probable than the old model, then one can confidently say that the data does not support the addition of the new parameters, regardless of the prior choice for the new parameters.

2 UPPER BOUNDS ON THE BAYES FACTOR

A model (M_0) may be compared to a new model with extra parameters (M_1) using the Bayes factor (also known as the odds)

$$B = \frac{p(x|M_1)}{p(x|M_0)}, \tag{1}$$

where x is the data and the model likelihood $p(x|M_i)$ ($i = 0, 1$) is given by

$$p(x|M_i) = \int \mathrm{d}\theta_i p(x|\theta_i, M_i) p(\theta_i|M_i) \tag{2}$$

with θ_i denoting the parameters under model M_i. The Bayes factor gives the change in the relative probability of the two models brought about by the data x, i.e.

$$\frac{p(M_1|x)}{p(M_0|x)} = B\frac{P(M_1)}{P(M_0)}, \tag{3}$$

where $P(M_i)$ $(i = 0, 1)$ are the prior probabilities for the two models and $P(M_i|x)$ the posterior probabilities. The level of support is usually categorized as either 'inconclusive' ($|\ln B| < 1$), 'weak' ($1 \leqslant |\ln B| \leqslant 2.5$), 'moderate' ($2.5 \leqslant |\ln B| \leqslant 5$), or 'strong' ($|\ln B| \geqslant 5$).

We denote the new parameters by θ and they are fixed to be θ^* under the simpler model (we restrict our considerations to nested models). The Bayes factor is then, using a generalized version of the Savage–Dickey density ratio (see Trotta (2007) for details)

$$B = \frac{p(\theta^*|M_1)}{p(\theta^*|x, M_1)}, \tag{4}$$

where $p(\theta^*|x, M_1)$ is the posterior distribution under M_1, evaluated at $\theta = \theta^*$. If $p(\theta|M_1)$ is made sufficiently broad, $p(\theta^*|x, M_1)$ depends only on the likelihood. Thus, B can be made arbitrarily small by making $p(\theta|M_1)$ sufficiently broad (since the prior must be normalized to unity probability content, a broader $p(\theta|M_1)$ corresponds to a smaller value of $p(\theta^*|M_1)$). This is not problematical if the physical model underlying M_1 is specific enough to provide some well–motivated prior bounds on θ. When this is not the case, an upper bound can still be obtained on B by optimizing over all priors and choosing $p(\theta|M_1)$ to be a delta function centered at the maximum likelihood value, $\theta_{\max}$. This is the choice that maximally favours M_1, and the upper bound on the odds is then

$$\bar{B} = \frac{p(x|\theta_{\max}, M_1)}{p(x|\theta^*, M_0)}, \tag{5}$$

corresponding to the likelihood ratio between $\theta_{\max}$ and θ^*. However, such a choice of prior fails to capture that M_1 is supposed to be a more complex model than M_0. Since θ^* represents the theoretically motivated simpler hypothesis, it makes sense that the alternative hypothesis has a more spread out prior distribution for θ. Furthermore, if there is *a priori* no strong preference for either $\theta > \theta^*$ or $\theta < \theta^*$, then it may be preferable to maximize over priors that are symmetric about θ^* and unimodal (the latter requirement coming again from a principle of indifference). Berger & Sellke (1987) show that maximizing B over all such $p(\theta|M_1)$ is the same as maximizing over all $p(\theta|M_1)$ that are uniform and symmetric about θ^*. We give an explicit example of this procedure in Eq. (11).

However, this optimization may be computationally prohibitive as evaluating Eq. (4) usually requires numerical evaluation of high dimensional integrals (Mukherjee et al. 2006; Feroz & Hobson 2007). An alternative way of obtaining an upper bound on B, that does not rely on explicitly specifying a class of alternative priors for θ, is to use Bayesian calibrated p–values (Sellke et al. 2001). First, frequentist methods are used to obtain the p–value. To do this, a test statistic (t) needs to be chosen, with the general property that the larger the value the less well the data agree with M_0. A common choice is the improvement in the maximum likelihood value when the additional parameters are allowed to vary. However, if the likelihood is computationally expensive to obtain, then other measures may be used. The p–value is given by

$$\wp = p(t \geqslant t_{\rm obs}(x)|M_0), \tag{6}$$

where $t_{\rm obs}(x)$ is the value of t estimated from the data. The key property of p–values is that if M_0 is correct, and t is a continuos statistic, then the probability distribution of $\wp$ will be uniform, $p(\wp|M_0) = 1$ for $0 \leqslant \wp \leqslant 1$. The final result will not be sensitive to the precise choice of t. The only property needed for t is that it should be a continuous statistic and larger values of t should correspond to less agreement with M_0. It follows that $p(\wp|M_1)$ will be monotonically decreasing for $0 \leqslant \wp \leqslant 1$. Sellke et al. (2001) express the Bayes factor in terms of the distribution of the p–values

$$B = \frac{p(\wp|M_1)}{p(\wp|M_0)} = p(\wp|M_1)\,. \tag{7}$$

They look at a wide range of non-parameteric monotonically decreasing distributions for $p(\wp|M_1)$ and under mild regularity conditions, they find the upper bound

$$B \leqslant \bar{B} = \frac{-1}{\mathrm{e}\wp\ln\wp} \tag{8}$$

for $\wp \leqslant \mathrm{e}^{-1}$, where e is the exponential of one. Table 1 lists $\bar{B}$ for some common thresholds of $\wp$ and $\ln B$. Note how the p–value of 0.05 (a 95% confidence level result) only corresponds to an odds ratio upper bound of $\bar{B} = 2.5$ and so does not quite reach the "weak" support threshold even for an optimized prior. Also note that in order order for the "strong" support threshold to be reachable, $\sigma \geqslant 3.6$ is required.

p–value	$\bar{B}$	$\ln \bar{B}$	sigma	category
0.05	2.5	0.9	2.0	
0.04	2.9	1.0	2.1	'weak' at best
0.01	8.0	2.1	2.6	
0.006	12	2.5	2.7	'moderate' at best
0.003	21	3.0	3.0	
0.001	53	4.0	3.3	
0.0003	150	5.0	3.6	'strong' at best
6×10^{-7}	43000	11	5.0	

Table 1. Translation table (using Eq. (8)) between p–values and the upper bounds on the odds ($\bar{B}$) between the two models. The 'sigma' column is the corresponding number of standard deviations away from the mean for a normal distribution. In the 'category' column are the descriptions for the different categories of support reachable for the corresponding p–value.

In general, for large sample size and under mild regularity conditions, the p–value for the addition of one or more new parameters can be estimated by finding the maximum likelihood with the new parameters fixed ($\mathcal{L}^*_{\max}$), and when the new parameters are allowed to vary ($\mathcal{L}_{\max}$). Then the quantity

$$\Delta\chi^2_{\rm eff} \equiv -2\ln(\mathcal{L}^*_{\max}/\mathcal{L}_{\max}) \tag{9}$$

has a Chi squared distribution with the number of degrees of freedom equal to the number of new parameters (Wilks 1938). It is important to note that for this to be valid none

of the new parameters can have their fixed values on the boundary of the parameter space (see e.g. Protassov et al. (2002) for an astronomy-oriented example where this condition does not hold). The p–value can then be estimated by

$$\wp = \int_{y=\Delta\chi^2_{\rm eff}}^{\infty} \chi^2_\nu(y)\, dy = 1 - \frac{\Gamma(\nu/2, \Delta\chi^2_{\rm eff}/2)}{\Gamma(\nu/2)} \qquad (10)$$

where χ^2_ν is the Chi squared distribution with ν degrees of freedom, and ν is the number of new parameters. Eq. (10) is simply the asymptotic probability of obtaining a $\Delta\chi^2$ as large or larger then what has actually been observed, $\Delta\chi^2_{\rm eff}$, assuming the null hypothesis is true. If the above procedure cannot be applied (for instance because the new parameters lie at a boundary of the parameter space), then the p–value can still be obtained by Monte Carlo simulations.

A very different approach to estimating the Bayes factor without having to specify a prior is the Bayesian Information Criteria (BIC) (Schwarz 1978; Magueijo & Sorkin 2007; Liddle 2007). The BIC assumes a prior for the new parameters which is equivalent to a single data point (Raftery 1995). Therefore, it will in general give lower values for B. The BIC is complementary to the upper bound for B presented here in that it provides a default weak rather than default strong prior.

3 AN ILLUSTRATIVE EXAMPLE

Consider the case where under M_0, $x \sim \mathcal{N}(\mu_0, \sigma)$ for fixed μ_0 (the null hypothesis), while under the alternative M_1, $x \sim \mathcal{N}(\mu, \sigma)$ and N data samples are available (with σ known). If the prior on μ is taken to be symmetric about $\mu = \mu_0$ and unimodal, then (Berger & Sellke 1987)

$$\bar{B} = \frac{\phi(K+t) + \phi(K-t)}{2\phi(t)} \qquad (11)$$

where $t \equiv \sqrt{N}|\bar{x} - \mu_0|/\sigma$, $\phi(y) \equiv \mathrm{e}^{-y^2/2}$, and K is found by solving

$$K[\phi(K+t) + \phi(K-t)] = \int_{-(K+t)}^{K-t} \phi(y)\, \mathrm{d}y\,. \qquad (12)$$

Alternatively, the p–value is given by

$$\wp = 1 - \int_{y=-t_{\rm obs}}^{t_{\rm obs}} \phi(y)\, \mathrm{d}y\,. \qquad (13)$$

This can be converted to a upper bound on the Bayes factor using Eq. (8). The results for the two methods are virtually identical and can be read off Table 1 where t is the number of sigma.

Sellke et al. (2001) present an interesting simulation study of this model. Consider the case described above, and let us generate data from a random sequence of null hypothesis (M_0) and alternatives (M_1), with $\mu_0 = 0$, $\sigma = 1$ and $\mu \sim \mathcal{N}(0, 1)$. Suppose that the proportion of nulls and alternatives is equal. We then compute the p–value using Eq. (13) and we select all the tests that give $\wp \in [\alpha - \epsilon, \alpha + \epsilon]$, for a certain value of α and $\epsilon \ll \alpha$. Among such results, which rejected the null hypothesis at the $1-\alpha$ level, we then determine the proportion that actually came from the null, i.e. the percentage of wrongly rejected nulls. We assume that either M_1 or M_0 is true. This allows us to use

$$P(M_0|x) = \frac{1}{1+B}\,. \qquad (14)$$

p–value	sigma	fraction of true nulls	lower bound
0.05	1.96	0.51	0.29
0.01	2.58	0.20	0.11
0.001	3.29	0.024	0.018

Table 2. Proportion of wrongly rejected nulls among all results reporting a certain p–value (simulation results). This illustrates that the p–value is not equal to the fraction of wrongly rejected true nulls, which can be considerably worse. This effect does not depend on the assumption of Gaussianity nor on the sample size. The right most column gives a lower bound on the fraction of true nulls derived using Eqs. (8) and (14).

The results are shown in Table 2. We notice that among all the "significant" effects at the 2σ level about 50% are wrong, and in general when there is only a single alternative at least 29% of the 2σ level results will be wrong.

The root of this striking disagreement with a common misinterpretation of the p–value (namely, that the p–value gives the fraction of wrongly rejected nulls in the long run) is twofold. While the p–value gives the probability of obtaining data that are as extreme or more extreme than what has actually been observed *assuming the null hypothesis is true*, one is not allowed to interpret this as the probability of the null hypothesis to be true, which is actually the quantity one is interested in assessing. The latter step requires using Bayes theorem and is therefore not defined for a frequentist. Also, quantifying how rare the observed data are under the null is not meaningful unless we can compare this number with their rareness under an alternative hypothesis. Both these points are discussed in greater detail in Berger & Sellke (1987); Sellke et al. (2001); Berger (2003).

4 SCALAR SPECTRAL INDEX

Here we evaluate the upper bounds on the Bayes factor for the scalar spectral index (n) using WMAP combined with other data, comparing a Harrison–Zeldovich model ($n = 1$) to a model where n can assume other values. For this problem, there are well motivated priors. If the primordial perturbations are from slow roll inflation, then

$$n = 1 + 2\eta - 6\epsilon \qquad (15)$$

where η and ϵ are the slow roll parameters and need to be much less than one. For most models $\epsilon \ll \eta$ and so a reasonable prior bound is

$$0.8 \lesssim n \lesssim 1.2\,, \qquad (16)$$

which can be implemented by taking a Gaussian prior of the form

$$p(n_s|M_1) = \mathcal{N}(\mu = 1.0, \sigma = 0.2). \qquad (17)$$

However, if the inflation potential (V) is of the form

$$V = V_0 - \frac{1}{2}m^2\phi^2 \qquad (18)$$

data	$\Delta\chi^2_{\rm eff}$	p–value	$\ln B$	$\ln \bar{B}$	$\overline{B}$
WMAP (Spergel et al. 2007)	6	0.014	–	1.8	6
WMAPext+SDSS +2df+No SZ (Parkinson et al. 2006)	8	0.005	2.0	2.7	15
WMAPext+HST (Kunz et al. 2006)	8	0.004	2.7	2.8	16
WMAPext+HST+SDSS (Trotta 2007)	11	0.001	2.9	3.9	49

Table 3. The odds against a Harrison–Zeldovich spectrum. The p–values where estimated from $\Delta\chi^2_{\rm eff}$ using Eq. (10). The upper bounds on the Bayes factor were estimated using Eq. (8). Where $\ln B$ is available it was calculated with the prior of Eq. (17).

(where ϕ is the inflaton, and V_0 and m are constants) then inflation can occur with $\eta \sim 1$ (Linde 2001; Boubekeur & Lyth 2005) and so a larger range of n may be considered for the prior.

As there is such a broad range for the the prior on n, it is useful to evaluate what is the upper bound on the odds for a non–Harrison–Zeldovich spectrum, $n \neq 1$. In Table 3 we list a number of different studies of the variation of the spectral index for a range of data. Where the Bayes factor has been worked out it can be seen that our estimate of the upper bound is always more than the evaluated version. Also, for the case with the greatest amount of data there is quite a large discrepancy between the upper bound and the evaluated odds. This makes sense as the same prior for n was used (Eq. (17)) but now the data is more constraining and so the maximizing prior is narrower. Using the most constraining data combination (WMAPext+HST+SDSS) the upper limits on the odds against $n = 1$ is 49:1. However, the odds against Harrison–Zeldovich could be weakened by various systematic effects in data analysis choices, e.g. inclusion of gravitational lensing, beam modelling, not including Sunyaev-Zeldovich (SZ) marginalization, and point-source subtraction (Peiris & Easther 2006; Lewis 2006; Parkinson et al. 2006; Eriksen et al. 2007; Huffenberger et al. 2006; Spergel et al. 2007).

5 ASYMMETRY IN THE CMB

In the recent WMAP 3–yr release the isotropy of the CMB fluctuations was tested using a dipolar modulating function (Spergel et al. 2007)

$$\Delta T(\hat{\mathbf{n}}) = \Delta T_{\rm iso}(\hat{\mathbf{n}})(1 + A\hat{\mathbf{n}} \cdot \hat{\mathbf{d}}) \tag{19}$$

where ΔT is the CMB temperature fluctuations in direction $\hat{\mathbf{n}}$, $\Delta T_{\rm iso}$ are the underlying isotropically distributed temperature fluctuations, A is the amplitude of the isotropy breaking, and $\hat{\mathbf{d}}$ is the direction of isotropy breaking. The isotropy of the fluctuations can then be tested by evaluating whether $A = 0$. The problem with using the Bayes ratio in this case is that there is no good underlying model which produces this type of isotropy breaking. An attempt was made by Donoghue et al. (2007) to allow an initial gradient in the inflaton field but they found that the modulation dropped sharply with scale. However, the required modulation should probably extend all the way to scales associated with the harmonic $\ell = 40$ (Hansen et al. 2004). Also, Inoue & Silk (2006) postulated that Poisson distributed voids may be responsible for the asymmetry. But, a generating mechanism for the voids and a detailed likelihood analysis are presently lacking.

data	$\Delta\chi^2_{\rm eff}$	p–value	$\ln B$	$\ln \bar{B}$	$\overline{B}$
WMAP (7°) (Spergel et al. 2007)	3	0.4	–	–	–
WMAP (7°)+$C_{\rm marg}$ (Gordon 2007)	9	0.03	–	1.3	4
WMAP (3.6°)+$C_{\rm marg}$ (Eriksen et al. 2007)	11	0.01	1.8	2.16	9

Table 4. The odds for dipolar modulation, $A \neq 0$. The resolution of the data used is also indicated. The $C_{\rm marg}$ refers to marginalisation over a non-modulated monopole and dipole. $\ln \bar{B}$ was evaluated using Eq. (8).

Therefore at present there is not a concrete enough theory to place meaningful prior limits on A. However, we can still work out the upper limit on the Bayes factor. The p–values can be evaluated from Eq. (10). Although $A = 0$ is on the boundary of the parameter space, the problem can be reparameterized in Cartesian coordinates where $A = w_x^2 + w_y^2 + w_z^2$ and w_i is a linear modulation weight for spatial dimension i. Then the $w_i = 0$ point, for all i, will not be on the edge of the parameter space and so Eq. (10) can be used.

The results are shown in Table 4. Simulations had been done for the last row's p–value (Eriksen et al. 2007) and were in excellent agreement with the result from Eq. (10). Eriksen et al. (2007) did compute the Bayes factor, taking as the prior $0 \leqslant A \leqslant 0.3$ but did not give a justification for that prior except that it contained all the non-negligible likelihood. This is unproblematic for parameter estimation, but is ambiguous for working out the Bayes factor. For example if the prior range for A was extended to be $0 \leqslant A \leqslant 0.6$ then the Bayes factor would decrease by 2 but the parameter estimates would be unaffected.

6 CONCLUSIONS

Bayesian model selection provides a powerful way of evaluating whether new parameters are needed in a model. There are however cases where the prior for the new parameter can be uncertain, or physically difficult to motivate. Here we have looked at priors which maximize the Bayes factor for the new parameters. This puts the reduced model under the most strain possible and so tells the user what the best case scenario is for the new parameters. We have pointed out a common misinterpretation of the meaning of p–values, which often results in an overestimation of the true significance of rejection tests for null hypotheses.

Using Bayesian calibrated p–values we have evaluated upper bounds on the Bayes factor for the spectral index. We have found that the best the current data can do is provide moderate support (odds $\leqslant 49:1$) for $n \neq 1$. We also looked at the maximum Bayes factor for a modulation

in the WMAP CMB temperature data. We found that the current data can at best provide weak support (odds $\leqslant 9:1$) for a departure from isotropy.

The comparison between p–values and Bayes factors suggests a threshold of $\wp = 3 \times 10^{-4}$ or $\sigma = 3.6$ is needed if the odds of 150:1 ("strong" support at best) are to be obtained. It is difficult to detect systematics which are smaller than the statistical noise and so systematic effects in the data analysis typically lead to a shift of order a sigma. It follows that the "particle physics discovery threshold" of 5σ may be required in order to obtain odds of at best 150:1.

ACKNOWLEDGMENTS

We are grateful to Kate Land for useful conversations and to Uros Seljak for interesting comments. CG is supported by Beecroft Institute for Particle Astrophysics and Cosmology. RT is supported by the Royal Astronomical Society through the Sir Norman Lockyer Fellowship, and by St Anne's College, Oxford.

REFERENCES

Berger J., 2003, Statistical Science, 18, 1
Berger J. O., Sellke T., 1987, The American Statistician, 55, 62
Boubekeur L., Lyth D. H., 2005, Journal of Cosmology and Astro-Particle Physics, 7, 10, ADS, arXiv:hep-ph/0502047
Donoghue J. F., Dutta K., Ross A., 2007, ArXiv Astrophysics e-prints, astro-ph/0703455, astro-ph/0703455
Eriksen H. K., Banday A. J., Gorski K. M., Hansen F. K., Lilje P. B., 2007, ApJ, 660, L81, ADS, astro-ph/0701089
Eriksen H. K., et al., 2007, ApJ, 656, 641, ADS, arXiv:astro-ph/0606088
Feroz F., Hobson M. P., 2007, ArXiv Astrophysics e-prints, arXiv:0704.3704, arXiv:0704.3704 [astro-ph]
Gordon C., 2007, ApJ, 656, 636, ADS, arXiv:astro-ph/0607423
Hansen F. K., Banday A. J., Gorski K. M., 2004, MNRAS, 354, 641, astro-ph/0404206
Huffenberger K. M., Eriksen H. K., Hansen F. K., 2006, ApJ, 651, L81, ADS, arXiv:astro-ph/0606538
Inoue K. T., Silk J., 2006, ApJ, 648, 23, ADS, arXiv:astro-ph/0602478
Kunz M., Trotta R., Parkinson D. R., 2006, PRD, 74, 023503, ADS, arXiv:astro-ph/0602378
Lewis A., 2006, ArXiv Astrophysics e-prints, astro-ph/0603753, ADS, astro-ph/0603753
Liddle A. R., 2007, Mon. Not. Roy. Astron. Soc. Lett., 377, L74, ADS, astro-ph/0701113
Linde A., 2001, Journal of High Energy Physics, 11, 52, ADS, arXiv:hep-th/0110195
Magueijo J., Sorkin R. D., 2007, Mon. Not. Roy. Astron. Soc. Lett., 377, L39, ADS, astro-ph/0604410
Mukherjee P., Parkinson D., Liddle A. R., 2006, ApJ, 638, L51, ADS, arXiv:astro-ph/0508461
Parkinson D., Mukherjee P., Liddle A. R., 2006, Phys. Rev., D73, 123523, astro-ph/0605003
Peiris H. V., Easther R., 2006, Journal of Cosmology and Astro-Particle Physics, 10, 17, ADS, arXiv:astro-ph/0609003
Protassov R., van Dyk D. A., Connors A., Kashyap V. L., Siemiginowska A., 2002, ArXiv Astrophysics e-prints, astro-ph/0201547, astro-ph/0201547
Raftery A., 1995, Sociological Methodology, 25, 11
Schwarz G., 1978, Annals of Statistics, 6, 461
Sellke T., J. B. M., Berger J. O., 2001, Journal of the American Statistical Association, 82, 112
Spergel D. N., et al., 2007, ApJS, 170, 377, astro-ph/0603449
Trotta R., 2007, Mon. Not. R. Astron. Soc., 378, 72, ADS, astro-ph/0504022
Trotta R., 2007a, MNRAS, 378, 819, astro-ph/0703063
Trotta R., 2007b, Mon. Not. Roy. Astron. Soc. Lett., 375, L26, astro-ph/0608116
Wilks S. S., 1938, The Annals of Mathematical Statistics, 9, 60