Mem. S.A.It. Vol. 85, 96

Memorie della

A Bayesian reanalaysis of the quasar dataset

E. Cameron and A.N. Pettitt

Queensland University of Technology (QUT), School of Mathematical Sciences, GPO Box 2434, Brisbane 4001, QLD, Australia, e-mail: dr.ewan.cameron@gmail.com

Abstract. We investigate recent claims of spatial variation in the fine structure constant on cosmic distance scales based on estimates of its extra-galactic–to–on-Earth ratio recovered from "many multiplet" fitting of quasar absorption spectra. To overcome the limitations of previous analyses requiring the assumption of a strictly unbiased and Normal distribution for the "unexplained errors" of this quasar dataset we employ a Bayesian model selection strategy with prior-sensitivity analysis. A particular strength of the hypothesis testing methodology advocated herein is that it can handle both parametric and semi-parametric models self-consistently through a combination of recursive marginal likelihood estimation and importance sample reweighting. We conclude from the presently-available data that the observed trends are more likely to arise from biases of opposing sign in the two telescopes used to undertake these measurements than from a genuine large-scale trend in this fundamental "constant".

Key words. Cosmology: cosmological parameters – methods: data analysis – methods: statistical.

1. Introduction

The fine structure constant, α, plays a crucial role in setting the scale for all electromagnetic interactions in quantum electrodynamics; and thereby the scale of the fine structure splitting in atomic spectral lines by which it is named. Although laboratory experiments based on the comparison of precise atomic clocks (Forier et al. 2007; Rosenband et al. 2008) and the nuclear modelling of samarium isotopes from the Oklo natural fission reactor (Shlyakhter 1976; Damour and Dyson 1996) have failed to recover evidence for spatial and/or temporal variation of α locally (i.e., on the scale of the Earth and its Solar orbit), the possibility of variation on cosmic scales is not so well-constrained. In particular, although pre-Millennial analyses of quasar emission and absorption spectra (Savedoff 1956; Bahcall, Sargent and Schmidt 1967; Ivanchik, Potekhin and Varshalovich 1999) were ultimately able to establish a bound of less than one part in ten-thousand on its fractional variation at cosmological redshifts (~1-3) with respect to its laboratory benchmark, $\Delta\alpha/\alpha \equiv (\alpha - \alpha_0)/\alpha_0$, more recent studies by Webb et al. (2011) and King et al. (2012) have claimed "4σ" evidence for spatial variation at an order of magnitude below this level using the "many multiplet" analysis technique.

The basis for the "many multiplet" technique (Dzuba, Flambaum and Webb 1999) is the model-based comparison of relative frequency shifts between multiple ionic spec-

Send offprint requests to: E. Cameron

tral lines; with the positions of α-insensitive species used as reference points for the shifting of their (theoretically) α-sensitive counterparts. The accuracy of such "many muliplet" $\Delta\alpha/\alpha$ estimates though depends on both the reliability of the absorption profile model adopted (e.g. whether thermal or turbulent broadening should be assumed) and the accuracy of the instrumental calibration; the magnitude of the former being more easily gauged (via maximum likelihood statistical methods) than that of the latter (which is effectively unknown for any single observation). In the Webb et al. team's own analysis of their quasar dataset, consisting of $\Delta\alpha/\alpha$ estimates from 295 intervening absorbers on 131 quasar sightlines observed with the VLT and Keck telescopes, it was revealed that "unexplained errors", including the contribution of the latter calibration uncertainties, were (at least for a large subset of these absorbers) of similiar magnitude to their "explained errors". Hence, the statistical modelling of both error components is necessary for robust inference and hypothesis testing with this dataset.

In the Webb et al. (2011) and King et al. (2012) studies the "unexplained errors" of each instrumental subgroup in the quasar dataset were assumed to originate from a hidden error term of strictly unbiased (zero mean), Normal form; and a trimmed least squares procedure was then used to estimate its standard deviation. However, with the equator of the Webb et al. team's alleged dipole signal running surprisingly close to that of the Earth—and thereby providing a near-perfect subdivision of the quasar sample into its VLT and Keck subgroups—a skeptical interpretation of their result would be that biases of opposing sign in the error terms belonging to each might be contributing the apparent dipole effect instead. Hence, we have sought to distinguish between these competing hypotheses through Bayesian model selection. The full mathematical treatment of this work, as presented at the *Varying Fundamental Constants and Dynamical Dark Energy Conference* (Sesto, Italy, 2013), is available in a pair of research papers which will be referred to as Paper I (Cameron & Pettitt 2013a) and Paper II (Cameron & Pettitt 2013b) throughout. Here we give a more qualitative description of our model selection methodology, which we hope might serve as a useful summary and introduction to the topic.

2. Bayesian model selection methodology

The aim of Bayesian model selection is to identify the most likely hypothesis to explain the observed data using a quantitative evaluation metric comparing their prior predictive accuracies; in effect, favouring model simplicity over model complexity. The key quantity for Bayesian model selection is the posterior Bayes factor for pairwise comparision of hypotheses, defined as the ratio of their marginal likelihoods, Z_1/Z_0. The marginal likelihood in turn being the likelihood, $L(y|\theta)$, of the observed data under the given hypothesis averaged over one's prior probability for its governing parameters. For problems where the prior admits an 'ordinary' probability density with respect to Lebesgue measure, $\pi(\theta)d\theta$, we have simply, $Z = \int_\Omega L(y|\theta)\pi(\theta)d\theta$ where the range of integration is the entire parameter space of the prior; and more generally (as will be the case for our semi-parametric error model) we have an integration with respect to the prior measure, $Z = \int_\Omega L(y|\theta)\{dP_{\rm prior}(\theta)\}$. The product of this Bayes factor with our prior probability ratio for the two hypotheses, Π_1/Π_0, gives their posterior odds ratio, $\Pi_1/\Pi_0 \times Z_1/Z_0$.

Applications of Bayesian model selection methods are becoming increasingly popular in astronomy and cosmology (e.g. Trotta 2007; Feroz & Hobson 2013) and there exist various astronomically-themed introductions to this technique (e.g. Trotta 2008); although one cannot over-state the value of the many classic statistical references as a guide for the newcomer (esp. Kass & Raftery 1995 and Wasserman 2000). The two greatest practical challenges for robust Bayesian model selection are (i) the accurate computation of the required marginal likelihoods (or, at least, their ratios), and (ii) quantification of the inherent sensitivity of the posterior Bayes factor to the choice of parameter priors on each hypothesis. While much thought is often given to the first problem in astronomical model selection studies, the second is not generally so well recognised. Below we outline a two-stage Monte Carlo integration strategy for addressing both challenges in an easily applied, computationally efficient and self-consistent manner.

2.1. Recursive marginal likelihood estimation

With the marginal likelihood integral not typically admitting a simple analytical solution it becomes necessary to seek a numerical approximation for use in computing the Bayes factor. And, with the efficiency of deterministic (grid-based) approximation strategies, such as numerical quadrature, limited in high-dimensional parameter spaces ($N \gtrsim 3$), the standard approach is to use Monte Carlo-based integration. Popular versions of the Monte Carlo approach to marginal likelihood estimation in astronomy include nested sampling (cf. Mukherjee et al. 2006; Feroz & Hobson 2008; Brewer et al. 2009) and the Savage-Dickey density ratio (Trotta 2007)—with the former particularly suited to models with (transformably) uniform priors and the latter restricted to embedded models with shared parameters. The technique we favour for our work is that of the recursive approach (see Cameron & Pettitt 2013c for a review) as represented by biased sampling (Vardi 1985) and reverse logistic regression (Geyer 1994). The advantages of this particular method being that the marginal likelihood can be computed directly from the likelihood draws under tempered Markov Chain Monte Carlo (MCMC) sampling of the posterior (regardless of prior structure; including for stochastic process priors), and that these draws may then be efficiently re-used for prior-sensitivity analysis via importance sample reweighting.

The particular version of recursive marginal likelihood estimation that we use here is as follows. First, we explore via MCMC the tempered posterior, $\propto \pi(\theta)L(y|\theta)^{\beta_j}d\theta$ (or $\propto L(y|\theta)^{\beta_j}\{dP_{\text{prior}}(\theta)\}$ more generally), saving the drawn θ_i and $L_i = L(y|\theta_i)$ at each point, for a series of (typically about twenty) temperatures, β_j, spanning the prior ($\beta_1 = 0$) and posterior ($\beta_m = 1$). Applying the globally-convergent, iterative update scheme of Vardi (1985) we recover estimates for the desired marginal likelihood, $Z = Z_m$, and the corresponding normalising constants of the tempered bridging distributions, Z_j ($1 < j < m$):

$$\hat{Z}_j = \sum_{i=1}^{n} \left(L_i^{\beta_j} / [\sum_{k=1}^{m} n_k L_i^{\beta_k} / \hat{Z}_k] \right). \tag{1}$$

While only $\hat{Z}$ is specifically required for Bayes factor computation under our default priors, the remaining $\hat{Z}_j$ are not unimportant since they facilitate our importance sample reweighting scheme for prior-sensitivity analysis.

2.2. Importance sample reweighting for prior-sensitivity analysis

With the Bayes factor directly dependent on the choice of parameter priors on each model an essential stage of any Bayesian model selection analysis is to quantify the inherent degree of prior-sensitivity (cf. Kass & Raftery 1995). One way to do this is to recompute Bayes factors under a range of alternative priors, typically set by moderate rescalings of the hyperparameters controlling one's prior forms. At face value this seems inevitably like a computationally expensive exercise since it calls for repeated marginal likelihood estimation. However, with our tempered posterior draws providing a thorough sampling of the parameter space at and more broadly around the location of the posterior mode(s)—and the posterior being typically less sensitive to prior change than the marginal likelihood itself—it makes sense to reuse these in some way.

In order to do this we utilise the estimated normalization constants, $\hat{Z}_j$, computed as above to define a pseudo-mixture density, $f(\theta)d\theta = \left(\sum_{j=1}^{m} L(y|\theta)^{\beta_j}/\hat{Z}_j\right)\pi(\theta)d\theta$ (or pseudo-mixture measure, more generally, $\{dF(\theta)\} = \left(\sum_{j=1}^{m} L(y|\theta)^{\beta_j}/\hat{Z}_j\right)\{dP_{\text{prior}}(\theta)\}$), from which we imagine our tempered likelihood draws to have originated. Importance sample reweighting of these draws according to the ratio of densities, $\pi_{\text{alt}}(\theta)/f(\theta)$, (or, more generally, the

Radon-Nikodym derivative; $\frac{dP_{\text{prior, alt}}}{df}(\theta)$) between some alternative prior and our nominal prior gives an estimate of the marginal likelihood under the former:

$$\hat{Z}_{\text{alt}} = 1/n \sum_{i=1}^{n} L_i \pi_{\text{alt}}(\theta_i)/f(\theta_i) \quad (2)$$

$$[\text{or } \hat{Z}_{\text{alt}} = 1/n \sum_{i=1}^{n} L_i \frac{dP_{\text{prior, alt}}}{df}(\theta_i)]. \quad (3)$$

Evidently this estimator requires no new likelihood function calls, often the most expensive part of a modern Bayesian computer code.

3. Reanalysis of the quasar dataset

As stressed already in the Introduction, the novelty of our Bayesian model selection reanalysis of the quasar dataset is that we are able to allow for the possibility of biased and/or non-Normal "unexplained error" terms in each instrumental subgroup; thereby relaxing the contested assumptions of the earlier studies claiming a 4σ dipole significance. In Paper I we chose (for simplicity) to restrict our analysis to the regime of parametric modelling, trialling only specific functional forms for the "unexplained error" distribution; but then in Paper II we extended our investigation to the semi-parametric modelling regime using an "infinitely-flexible" mixture of Dirichlet processes prior to represent the hidden systematic error on each $\Delta\alpha/\alpha$ datapoint.

3.1. Parametric error models

Four basic forms were examined here for the "unexplained error" term, such that when convolved with the well-motivated Normal form (cf. King et al. 2010) for the "explained error" term the resulting total errors were thus effectively: (i) strictly Normal; (ii) Voigt (or Cauchy); (iii) skew-Normal; and (iv) "bimodal". In each case we also specifically identified and trialled separately a "biased" and "unbiased" version via suitable hyperparameter restrictions (e.g. for the strict Normal, that its mean be also strictly zero for the "unbiased" case, and free for the "biased" case).

3.2. Semi-parametric error model

The Dirichlet process, being an infinite-dimensional extension of the familiar Dirichlet distribution, constitutes a stochastic process on the space of (atomic) probability distributions, and is thus a commonly-used tool for non-parametric modelling in applied statistics (see Ferguson 1973; and Ghosh 2010 for a brief introduction). When combined with a parametric density kernel, as represented here by the Normal distribution of the "explained error" term, the result is a semi-parametric model most often encountered in clinical meta-analysis studies. Prior specification for the Dirichlet process is via the choice of a reference density and concentration index; and in the mixture of Dirichlet processes setting one may also specify hyperpriors on both the controlling parameters of this reference density and the concentration index itself; which we elected to do as described in Paper II.

3.3. Marginal likelihoods and posterior Bayes factors

For the purposes of our model selection analysis we identified three basic hypotheses for the cosmic variation (or lack thereof) in α. Namely, (i) the strict null ($\Delta\alpha/\alpha = 0$ everywhere); (ii) the monopole null ($\Delta\alpha/\alpha = m$; i.e. a fixed quasar-to-Earth offset); and (iii) the monopole plus $r(z)$-dipole model (a direction and "lookback distance" dependent cosmic α variation). After specifying well-motivated priors for the controlling parameters of these hypotheses we proceeded to estimate marginal likelihoods for each, coupled in turn to each of our candidate error terms. Within the restricted class of "unbiased" error models (parametric only) our marginal likelihood comparisons indeed strongly favoured the dipole hypothesis over the monopole and null alternatives at a Bayes factor of ~300, in broad agreement with the Webb et al. team's original conclusions. *However*, for all "unbiased" error models this ranking of hypotheses was reversed, with the strict null mildly favoured over the monopole and dipole at a Bayes factor of up to ~12.5. And, more importantly, the marginal likelihoods for the null hypotheses under at least two of our "biased" parametric models (the Normal and skew-Normal error terms) was actually greater than that of the dipole hypothesis under any of the "unbiased" models; again at a moderate Bayes factor of $\sim$10.

Although the marginal likelihoods recovered under our semi-parametric error model were all significantly lower than those for our preferred parametric models (owing to the inherently broader range of prior predictives for this maximally flexible error form) the Bayes factor ordering of null over monopole and dipole was nevertheless preserved. In this sense our semi-parametric model may be seen at this stage as primarily a robustness check of our parametric analysis; but one may well anticipate this error model eventually becoming preferred with the addition of further data straining the plausibility of a perfectly Normal or skew-Normal (biased) distribution of "unexplained errors".

To explore the prior-sensitivity of our Bayes factors under the best of our "biased" parametric error models we applied the above-mentioned importance sample reweighting technique to recompute the marginal likelihoods under multiple rescalings of our key hyperparameters (Paper I). We were thereby able to demonstrate the preservation of hypothesis rankings over a wide range of alternative prior choices, with a restriction of the prior bias strength to less than one part in a million ultimately deemed necessary to overturn our support for the null. In the case of the semi-parametric error model we examined the effect of changing the distributional form of our Dirichlet process centering density from Normal to Student's t of varying degree. Once again the robustness of our hypothesis rankings was thereby confirmed; however, it is interesting to note that we were actually able to increase substantially the marginal likelihood of our null hypothesis (and conversely to decrease the marginal likelihoods of our monopole and dipole hypotheses) by adopting a progressively "fatter-tailed" reference (with a peak in the former at a $d.f.$ of 4). Hence, we stress that although the semi-parametric model selection approach does allow for the relaxation of some important modelling assumptions it does not entirely free the practioner from concerns of prior-sensitivity.

4. Conclusions

We conclude from the presently-available data that the observed trends are more likely to arise from biases of opposing sign in the two telescopes used to undertake these measurements than from a genuine large-scale trend in this fundamental "constant". Nevertheless, to strengthen these conclusions (beyond a Bayes factor of $\sim$10 in favour of the null under our skeptical interpretation) requires either more $\Delta\alpha/\alpha$ estimates from each telescope (particularly along equatorial sightlines, as well as along the alleged dipole's axes) or further calibration studies to constrain the instrumental systematics.

Acknowledgements. [1] E.C. and A.N.P. are grateful for financial support from the Australian Research Council.
[2] E.C. would like to thank the organisers of the *Varying Fundamental Constants and Dynamical Dark Energy Conference* and staff at the Sexten Center for Astrophysics, Italy, for their many efforts, which led to a productive and stimulating meeting.

References

Bahcall J.N., Sargent W.L.W. and Schmidt M., 1967, ApJ, 149, L11
Brewer B.J., Pártay L.B. and Csányi G. 2009, arXiv:0912.2380
Cameron E. and Pettitt A.N. 2013a, arXiv:1207.6223
Cameron E. and Pettitt A.N. 2013b, arXiv:1309.2737
Cameron E. and Pettitt A.N. 2013c, arXiv:1301.6450
Damour T. and Dyson F. 1996, Nuclear Phys. B, 480, 37
Dzuba V.A., Flambaum V.V.and Webb J.K. 1999, Phys. Rev. A, 59, 230
Feroz F. & Hobson M. 2008, MNRAS, 384, 449
Feroz F. & Hobson M. 2013, arXiv:1307.6984

Fortier T.M, et al. 2007, Phys. Rev. Lett. 98, 070801

Geyer C.J. 1994, Estimating normalizing constants and reweighting mixtures in MCMC. Technical Report 568, School of Statistics, Univ. Minnesota.

Ivanchik A.V., Potekhin A.Y. and Varshalovich D.A. 1999, A&A, 343, 439

Kass R.E. and Raftery A.E. 1995, JASA, 90, 773

King J.A., et al. 2012, MNRAS, 422, 3370

Mukherjee P., Parkinson D. and Liddle A.R. 2006, ApJ, 638, L51

Rosenband T., et al. 2008, Science, 319, 1808

Savedoff J.D. 1956, Nature, 178, 688

Shlyakhter A.I. 1976, Nature, 264, 340

Trotta R. 2007, MNRAS, 378, 72

Trotta R. 2008, Contemporary Physics, 49, 71

Vardi Y. 1985, Ann. Stat., 13, 178

Wasserman L. 2000, J. of Math. Psych., 44, 92

Webb J.K., et al. 2011, Phys. Rev. Lett., 107, 191101