Mem. S.A.It. Vol. 80, 864

Memorie della

Markov Chain Monte Carlo methods applied to measuring the fine structure constant from quasar spectroscopy

J.A. King[1], D.J. Mortlock[2], J.K. Webb[1], and M.T. Murphy[3]

[1] Department of Astronomy, University of New South Wales, Sydney, New South Wales, 2052, Australia. e-mail: jking.phys@gmail.com
[2] Astrophysics Group, Blackett Laboratory, Price Consort Road, Imperial College London, London SW7 2AZ, U.K.
[3] Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Victoria, 3122, Australia.

Abstract. Recent attempts to constrain cosmological variation in the fine structure constant, α, using quasar absorption lines have yielded two statistical samples which initially appear to be inconsistent. One of these samples was subsequently demonstrated to not pass consistency tests; it appears that the optimisation algorithm used to fit the model to the spectra failed. Nevertheless, the results of the other hinge on the robustness of the spectral fitting program VPFIT, which has been tested through simulation but not through direct exploration of the likelihood function. We present the application of Markov Chain Monte Carlo (MCMC) methods to this problem, and demonstrate that VPFIT produces similar values and uncertainties for $\Delta\alpha/\alpha$, the fractional change in the fine structure constant, as our MCMC algorithm, and thus that VPFIT is reliable.

Key words. Atomic processes – Methods: numerical – Methods: statistical – Quasars: absorption lines – Quasars: individual: LBQS 2206−1958 – Quasars: individual: LBQS 0013−0029 – Quasars: individual: Q 0551−366 – Cosmology: observations

1. Introduction

Recent years have seen sustained interested in attempting to determine whether any of the fundamental constants of nature vary. Efforts have been focused in particular on the fine structure constant, α, and the proton-to-electron mass ratio, μ, although other dimensionless constants have also been considered. Quasars provide a method of probing the value of α in the early universe by illuminating gas clouds along the line of sight to Earth. In particular, certain absorption lines in these gas clouds are sensitive to changes in one or more of the fundamental constants of nature. For an atom/ion, the relativistic corrections to the energy levels of an electron are proportional to α^2, although the magnitude of the change depends on the transition under consideration. By comparing the wavelengths of absorption lines with differing sensitivities to a change in α, we are able to place constraints on a change in α between the early universe and today. That is, we are able to measure $\Delta\alpha/\alpha = (\alpha_z - \alpha_0)/\alpha_0$, where α_z is the value of α at redshift z and α_0 is the laboratory value.

The tentative detection of a variation in α reported by Webb et al. (1999) has further increased interest in this field. Subsequent efforts refined this work by increasing the number of absorption systems considered. This yielded a constraint of $\Delta\alpha/\alpha = (-0.57 \pm 0.11) \times 10^{-5}$ (Murphy et al., 2004) from 143 absorption systems. However all of these observations are from the Keck/HIRES (High Resolution Echelle Spectrometer) instrument, and so it remains important to confirm such unexpected results with independent equipment.

Chand et al. (2004) reported $\Delta\alpha/\alpha = (-0.06 \pm 0.06) \times 10^{-5}$ from 23 measurements using the Ultraviolet and Visual Echelle Spectrograph (UVES), on the Very Large Telescope (VLT). These two measurements of $\Delta\alpha/\alpha$ are clearly discrepant. A later analysis (Murphy et al., 2008) found that the analysis of Chand et al. cannot be correct as it states a statistical precision in excess of the maximum theoretical precision allowed by the data. Similarly, Murphy et al. analysed the χ^2 vs $\Delta\alpha/\alpha$ curves produced by the Chand et al. optimisation algorithm, and concluded that the shape of the curves demonstrate a failure of the algorithm to find the maximum likelihood value of $\Delta\alpha/\alpha$ implied by the model fits and data, and thus that the estimate of $\Delta\alpha/\alpha$ given by Chand et al. is unreliable.

Although it would appear that the Murphy et al. results are robust, it is worth directly investigating the optimisation algorithm used in order to confirm that it is reliable. Furthermore, each measurement of $\Delta\alpha/\alpha$ will be subject to systematic errors, but some systematic errors should have an expectation value of zero, and thus averaging over many absorption systems will in principle eliminate such errors. It has become commonplace to quote values of $\Delta\alpha/\alpha$ for individual systems; in these cases, one has no method of determining the size of certain systematic errors through comparison with other systems, and so one should be particularly careful that the statistical errors are stated correctly. The MCMC method we describe herein allows one to confirm the validity of the statistical errors produced by a different method of analysis.

2. Motivation for MCMC

Ideally one would like an independent method of demonstrating whether or not a purported value of $\Delta\alpha/\alpha$ and the associated statistical uncertainty given by an optimisation algorithm are reliable.

Optimisation algorithms seek to minimise $\chi^2 = \sum_i [I(\mathbf{x})_i - d_i]^2/\sigma_i^2$, where d_i are the spectroscopic data points, $I(\mathbf{x}_i)$ is the prediction of the model at each data point, and σ_i is the standard error associated with each spectroscopic data point. However, $\chi^2 = -2\ln(L(\mathbf{x}))$ up to an additive constant which can be neglected, as we only ever consider differences of χ^2 for finding model parameters. We thus use the definition of the likelihood function $L(\mathbf{x}) \equiv \exp[-\chi^2(\mathbf{x})/2]$.

One option is to use an alternate optimisation algorithm. Although optimisation algorithms are in principle simple, numerical issues can cause them to fail inconsistently; this may be the case for the algorithm utilised by Chand et al. (2004). In particular, the optimisation algorithms employed by Chand et al. (2004), Murphy et al. (2008) and Murphy et al. (2004) are of the Newton type, which require all first and second partial derivatives of χ^2 with respect to to the parameters to be known. The Voigt function used to model the absorption lines is not analytic, and nor are its derivatives. As such, partial derivatives must be approximated by finite difference methods. Inappropriate choices of the step size for the finite differencing scheme can either produce poor approximations to the derivatives (step size too large) or be rendered useless by roundoff error (step size too small), leading to poor performance of the optimisation algorithm. There are number of other numerical issues which may cause failure of the optimisation algorithm, but we do not consider these here.

On account of these numerical issues, one would desire to explore the parameter space itself to directly determine the confidence limits on $\Delta\alpha/\alpha$.

3. Description of the MCMC method

Traditional Monte Carlo methods suffer from the "curse of dimensionality". That is, their performance degrades exponentially with increasing dimensionality. The Markov Chain Monte Carlo (MCMC) method degrades only polynomially with increased dimensionality, at the expense of introducing correlation between samples. Additionally, MCMC methods must be tuned to the probability distribution under consideration so as to explore the parameter space efficiently.

We implement a variant of the Metropolis algorithm (Metropolis et al., 1953) to explore our parameter space. The Metropolis algorithm proposes a new position in the parameter space, $\mathbf{x}'$, based on the current position, $\mathbf{x}$, according to some proposal function, $T(\mathbf{x}, \mathbf{x}')$. The only requirement imposed is that $T(\mathbf{x}, \mathbf{x}') = T(\mathbf{x}', \mathbf{x})$ (i.e the proposal distribution is symmetric).

Although in principle there are large numbers of possible proposal funcitons, T, in practice the most common choice is a multidimensional Gaussian centred on the current point, such that $\mathbf{x}' = \mathbf{x} + gN(0, \mathbf{\Sigma})$ where $\mathbf{\Sigma}$ is the covariance matrix obtained from the optimisation algorithm at the purported best fit solution, and g is a scalar tuning factor. The choice of T influences only the efficiency of the algorithm, not the formal correctness of the solution. The initial $\mathbf{\Sigma}$ may or may not be a good approximation to the true convariance matrix. The use of $\mathbf{\Sigma}$ ensures that the distribution of proposed parameters is approximately the same as the underlying distribution; the closer $\mathbf{\Sigma}$ is to the true covariance matrix, the faster the MCMC algorithm will be.

The tuning factor g effectively controls the size of steps taken. If g is too large, most trial steps will land in regions of low likelihood, and therefore most steps will be rejected (the chain will not move). On the other hand, if g is too small, the acceptance rate will be $\approx 100\%$, but the parameter space will be explored too slowly. If both the target and proposal distributions are Gaussian then the ideal acceptance rate is $\approx 44\%$ (Gelman et al., 1995).

The algorithm generates a sequence of points, $\{\mathbf{x}^t\}$, according to a two step prescription. First, from the current point, $\mathbf{x}$, propose a new point, $\mathbf{x}'$, via $T(\mathbf{x}, \mathbf{x}')$. Then calculate the ratio $q = L(\mathbf{x}')/L(\mathbf{x})$. Secondly, with probability $\min(q, 1)$ move to the new point i.e. set $\mathbf{x}^{t+1} = \mathbf{x}'$. Otherwise, retain the current point i.e. $\mathbf{x}^{t+1} = \mathbf{x}^t$. In this fashion, proposed moves to a point which is more likely than the existing point are always accepted, whereas proposed moves to a point which is less likely than the existing point are sometimes accepted, depending on the ratio of likelihoods. For a sufficiently large numbers of iterations, and with proper tuning of the algorithm, the distribution of $\{\mathbf{x}^t\}$ will sample from the underlying probability distribution, up to a normalisation constant. In particular, $\{\mathbf{x}^t_{\Delta\alpha/\alpha}\}$ will sample from the probability distribution of $\Delta\alpha/\alpha$, from which we can obtain a best estimate and confidence limits.

To minimise running time, for each model fit we run our MCMC algorithm several times (usually five to ten, depending on the complexity of the situation, with several hundred thousand iterations per stage) and re-estimate $\mathbf{\Sigma}$ at each stage from the chain. Prior to starting each MCMC run, we execute small runs (typically 250 iterations) in which we tune g to be such that the acceptance rate is between 30% and 50% (the outputs from these small runs do not count towards each stage). Thus, even if the initial covariance matrix does not allow a good exploration of the parameter space, by re-estimating $\mathbf{\Sigma}$ and retuning g several times we can drastically increase the chance that the final MCMC run will produce a good approximation of the underlying probability distribution. We determine whether the final MCMC run is useful by examining the chain for autocorrelation. If the autocorrelation length in the chain is much smaller than the chain length, then we deem the final run acceptable.

We do not require the usual "burn-in" period, where one discards a certain number of samples from the start of the chain, because we believe our parameters already start at the likelihood maximum. We can determine whether this assumption is robust by examining the chain – parameters should stay near their starting values, on average, if our initial parameter estimates were good.

We implement the Multiple Try Metropolis algorithm (Liu et al., 2000), which expands the Metropolis algorithm to allow multiple attempts at each step. If the initial proposal distribution is poorly tuned, this variant of the Metropolis algorithm tends to be much more robust for larger number of dimensions.

Similarly, we do not use a Gaussian proposal distribution, but start with a radial distribution which has $P(r) \propto (2/3)r^2 \exp(-r^2/2) + (1/3)\exp(-r)$. This mixture of an exponential distribution and the radial component of a 2D Gaussian allows the algorithm to occasionally take large steps, whilst otherwise taking steps clustered about some value; this speeds exploration of the parameter space where $\mathbf{\Sigma}$ is initially poorly tuned. For our proposal distribution, T, we generate our parameters from a spherically symmetric distribution with radial probability density $P(r)$ and then left multiply by $\mathbf{L}$ (where $\mathbf{LL^T} = \mathbf{\Sigma}$) so that the proposal distribution has the correct covariance structure.

MCMC can be used to directly estimate posterior probabilities in the Bayesian framework with the appropriate choice of prior distribution. The likelihood ratio then becomes $L(\mathbf{x}) \rightarrow L(\mathbf{x})\pi(\mathbf{x})$, where $\pi(\mathbf{x})$ is the Bayesian prior for a particular set of parameters. We utilise improper flat priors for the column densities and redshifts of each component. Similarly, we utilise a flat prior on $\Delta\alpha/\alpha$.

We utilise a flat prior for the logarithm of the Doppler parameters, rather than the Doppler parameters, to suppress movements to small b. Otherwise, the algorithm tends to propose many jumps to $b < 0$ for narrow lines, which must be rejected – this substantially reduces the efficiency of the algorithm. This is somewhat reasonable also on physical grounds, as we do not expect large numbers of gas clouds described by arbitrarily small temperatures, and certainly our fitting procedure would reject a profile decomposition of this nature. Using a flat prior for the Doppler parameters results in unacceptably large running times, and so we use the logarithmic prior as an easy and practical option.

We have modified the spectral profile fitting program VPFIT[1] to incorporate our MCMC algorithm. The outputs of the optimisation algorithm are fed directly into the initialisation for the MCMC code. Our MCMC algorithm uses the same code as VPFIT to generate the Voigt profiles and calculate χ^2, and thus our algorithm does not eliminate the possibilty of a code bug entirely. However, with this caveat, our algorithm can determine whether the optimisation code used by VPFIT does or does not converge to the desired solution and produce appropriate uncertainty estimates.

4. Results

We have applied our MCMC algorithm as described above to the three quasar absorption systems described below. The numerical values produced by the optimisation algorithm ("VPFIT") and the MCMC code ("MCMC") are given in table 1. In all cases we find good agreement between the VPFIT result and that produced by our MCMC code, although the statistical uncertainties produced by our MCMC code are mildly smaller than that produced by VPFIT, indicating that VPFIT may be conservative.

Our fits all pass appropriate robustness tests (in particular, $\chi^2_\nu \approx 1$ where ν is the number of degrees of freedom for the fit). All of our final chains mix well, although with the second and third objects considered here the initial chains do not – re-estimation of the covariance matrix multiple times is necessary to achieve a well mixed chain. All redshifts here refer to the redshift of the absorption system.

4.1. LBQS 2206–1958 $z = 1.018$

This absorption system appears to be well fitted by a single Voigt profile. We use the Mg II $\lambda\lambda 2796, 2803$Å transitions, which are relatively insensitive to α variation, and the Fe II $\lambda\lambda\lambda\lambda 2382, 2600, 2344, 2587$Å transitions, which are strongly sensitive.

The parameters are approximately jointly normally distributed. We expect this for single

[1] See http://www.ast.cam.ac.uk/~rfc/vpfit.html

Table 1. Comparison of purported values of $\Delta\alpha/\alpha$ calculated by VPFIT, and the results of the MCMC algorithm. Quoted uncertainties are 1σ.

Object	Redshift	$\Delta\alpha/\alpha$ – VPFIT	$\Delta\alpha/\alpha$ – MCMC
LBQS 2206−1958	1.018	$(-0.51 \pm 1.07) \times 10^{-5}$	$(-0.51 \pm 0.88) \times 10^{-5}$
LBQS 0013−0029	2.029	$(-0.86 \pm 0.94) \times 10^{-5}$	$(-0.83 \pm 0.77) \times 10^{-5}$
Q 0551−366	1.748	$(-0.80 \pm 1.08) \times 10^{-5}$	$(-0.89 \pm 0.84) \times 10^{-5}$

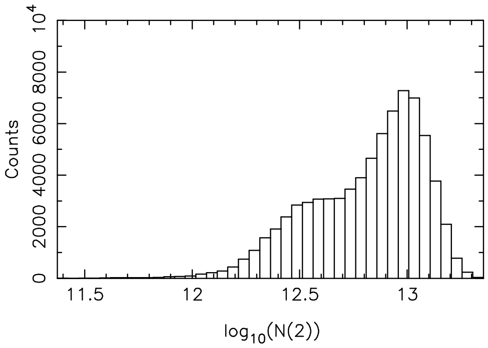

Fig. 1. Histogram of the chain for $\log_{10}(N(2)/\mathrm{cm}^2)$, where $N(2)$ is the column density of the central component to the fit for the $z = 1.748$ absorption system toward Q 0551−366.

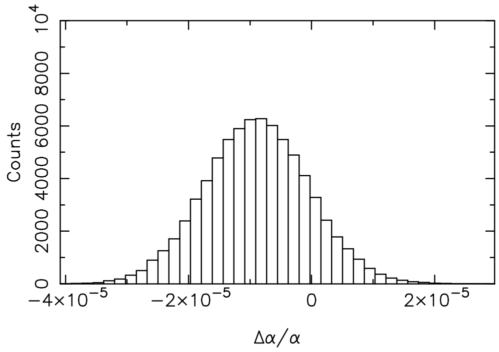

Fig. 2. Histogram of the chain for $\Delta\alpha/\alpha$ for the $z = 1.748$ absorption system toward Q 0551−366.

component fits – the Voigt profile decomposition is effectively unique with one component.

4.2. LBQS 0013–0029 $z = 2.029$

This system appears with two obvious absorption features. We find that the bluer feature is better fitted by two components than it is by one on the basis of a statistically significant reduction in χ^2 when using two components. That is, we fit three components to this absorption profile. We use a wide variety of transitions, namely: Si II λ1526Å, Al III $\lambda\lambda$1854, 1862Å, Fe II $\lambda\lambda\lambda\lambda\lambda$2382, 2600, 2344, 2587, 1608Å and Mg I λ2852Å.

The chain values of $\Delta\alpha/\alpha$ are approximately Gaussian-distributed.

4.3. Q 0551–366 $z = 1.748$

This absorption feature appears as one weak feature next to one relatively strong feature, with some overlap. We find the bluer feature to be well modelled by one component, however the higher wavelength feature appears to require two closely spaced components to achieve a statistically acceptable fit. Hence, we use three components in total to model the observed profile. The transitions we use are: Si II λ1526Å, Mg I λ2852Å and Fe II $\lambda\lambda\lambda\lambda\lambda\lambda$ 2382, 2600, 2344, 2587, 1608, 2374Å.

We note that the parameters corresponding to the two highest wavelength components are manifestly not normally distributed (see Fig. 1 for an example). We confirm, by inspection of the chain, that this effect is not due to permutations of corresponding parameters, which would leave χ^2 unchanged.

Despite the gross departures from Gaussianity for the column density and Doppler parameters corresponding to the two reddest components, the histogram of $\Delta\alpha/\alpha$

remains approximately Gaussian (see Fig. 2). The Voigt profile decomposition is not unique, and so we can only try to find the model which best describes the observations statistically. However, for a given model we would naively expect that there is a unique value of $\Delta\alpha/\alpha$ which minimises χ^2, and additionally that $\Delta\alpha/\alpha$ should be approximately Gaussian. We find both of these statements to be true here.

It is reassuring that we find concordance between the VPFIT and MCMC results for $\Delta\alpha/\alpha$ given the significant non-Gaussianity in some parameters. For non-Gaussian parameters, the parameter estimates produced by VPFIT will be the correct maximum likelihood estimates, however the confidence intervals will be biased. For our present purposes, we are only interested in the confidence limits on $\Delta\alpha/\alpha$, and here we find an acceptable level of agreement.

4.4. Combination of results

If we assume that a single value of α underlies these, we can combine the three VPFIT results above using a weighted mean to estimate $\Delta\alpha/\alpha = (-0.74 \pm 0.59) \times 10^{-5}$, which is statistically consistent with no change in α. We use the VPFIT results as a more conservative estimate.

5. Discussion & conclusion

Given suitable knowledge of the observed distribution of column densities and Doppler parameters, we could implement these distributions as priors to our model. However, our statistical constraints are generally sufficiently good that this should not significantly alter our parameter estimates. In any event, we are primarily interested in $\Delta\alpha/\alpha$, for which a flat prior is reasonable on ignorance grounds.

We would like to apply our algorithm to verify the results of King et al. (2008) in relation to $\Delta\mu/\mu$; however, those fits involve greater than a thousand parameters each. In this context, our algorithm is hopelessly inadequate. Our running times for the objects described herein varied from a few hours to a few days, and have only a few tens of parameters. Exploration of more complicated cases must wait for advances in computing power.

Our results demonstrate that VPFIT produces reliable parameters estimates and uncertainties for relatively simple situations. Experience with VPFIT suggests that there does not appear to be any indication of failure with moderately complicated circumstances, and so we would argue that the optimisation algorithm used by VPFIT is robust. The implication of is this it that the results of Murphy et al. (2004) are unlikely to be explained by some failure of the optimisation algorithm used by VPFIT. Thus the detection of a change in α must either be real, or due to some other unknown issue.

Acknowledgements. Presentation of this work at the 2009 IAU XXVII General Assembly JD9 conference was supported through financial assistance from the UNSW PRSS scheme.

References

Chand, H., Srianand, R., Petitjean, P., & Aracil, B. 2004, Astron. Astrophys., 417, 853

Gelman, A., Roberts, G. O., & Gilks, W. R. 1995, Bayesian Statistics, Vol. 5, Efficient Metropolis jumping rules, ed. J. Bernardo, J. Berger, A. Dawid, & A. Smith (Oxford University Press, Oxford)

King, J. A., Webb, J. K., Murphy, M. T., & Carswell, R. F. 2008, Phys. Rev. Lett., 101, 251304

Liu, J. S., Liang, F., & Wong, W. H. 2000, J. Amer. Statistical. Assoc., 95, 121

Metropolis, N., Rosenbluth, A. W., Teller, A. H., & Teller, E. 1953, J. Chem. Phys., 21, 1087

Murphy, M. T., Flambaum, V. V., Webb, J. K., et al. 2004, in Lecture Notes in Physics, Berlin Springer Verlag, Vol. 648, Astrophysics, Clocks and Fundamental Constants, ed. S. G. Karshenboim & E. Peik, 131–150

Murphy, M. T., Webb, J. K., & Flambaum, V. V. 2008, Mon. Not. Roy. Astron. Soc., 384, 1053

Webb, J. K., Flambaum, V. V., Churchill, C. W., Drinkwater, M. J., & Barrow, J. D. 1999, Phys. Rev. Lett., 82, 884