Bayesian Methods for Cosmological Parameter Estimation from Cosmic Microwave Background Measurements

Nelson Christensen[1], **Renate Meyer**[2], **Lloyd Knox**[3], **Ben Luey**[1]
[*]
[1]Physics and Astronomy, Carleton College, Northfield, MN 55057,USA
[2]Department of Statistics, The University of Auckland, Auckland, New Zealand
[3]Department of Physics, University of California - Davis, Davis, CA 95616, USA

Abstract. We present a strategy for a statistically rigorous Bayesian approach to the problem of determining cosmological parameters from the results of observations of anisotropies in the cosmic microwave background. Our strategy relies on Markov chain Monte Carlo methods, specifically the Metropolis-Hastings algorithm, to perform the necessary high–dimensional integrals. We describe the Metropolis–Hastings algorithm in detail and discuss the results of our test on simulated data.

PACS numbers: 02.70.Uu, 02.60.Pn, 98.70.Vc, 98.80.Es

1. Introduction

Recent determinations of the angular power spectrum of the anisotropy of the cosmic microwave background (CMB) [1, 2, 3, 4, 5, 6, 7] have created much excitement. These data may be used to estimate cosmological parameters. The maximum in the angular power spectrum around the multipole value $l \approx 200$ ('the first peak') is consistent with inflation and vanishing mean spatial curvature [8, 9, 10]. The lack of curvature adds compelling evidence to the case for the existence of some smoothly–distributed negative–pressure component which is possibly a cosmological constant (e.g., [11]). The constraints on the amplitude of the predicted second peak lead to lower limits on the baryon density [8, 12, 13] which are in conflict with the standard results from big bang nucleosynthesis (e.g., [14]); new data are eagerly awaited to provide a resolution. Fortunately, future satellite missions will provide angular power spectrum determinations with greatly improved precision [15, 16].

The extraction of cosmological parameters from the CMB anisotropy data requires a long and computer–intensive analysis chain. Here we are interested in the last step in the chain which is a derivation of cosmological parameters from the inferred angular power spectrum, C_l.

This last step is computer–intensive as well and promises to be more so as the number of parameters in ones model increases. Current analysis exercises have included up to ten parameters[8, 12, 13]; the scalar quadrupole and gravity wave perturbation normalizations (A_s and A_t), the scalar and tensor power–law indices for primordial perturbations (n_s and n_t), the reionization optical depth τ, the spatial curvature Ω_k, the energy densities for baryonic matter (ω_b), cold dark matter (ω_{cdm}), neutrinos (Ω_ν) and the vacuum (Ω_Λ). Other logical cosmological parameters could include the number of neutrino families and their masses. Potentially infinite degrees of freedom lie in the primordial spectrum of perturbations.

The estimation of parameters can be thought of as an exercise in Bayesian inference. One starts with the likelihood function, namely the conditional probability distribution function (PDF) of the observation $\mathbf{z}$ *given* the unknown parameters, $p(\mathbf{z}|\boldsymbol{\theta})$. For studies pertaining to the CMB, the likelihood used is an approximation to the probability of the power spectrum given the data, e.g. [17]. If the likelihood function for a certain parameter vector $\boldsymbol{\theta}_1$ is higher than for a parameter vector $\boldsymbol{\theta}_2$, the observations are more likely to have occurred under $\boldsymbol{\theta}_1$ than under $\boldsymbol{\theta}_2$. Thus, the observations give higher plausibility to $\boldsymbol{\theta}_1$ and finding the parameter value that maximizes the likelihood seems to have some compelling logic. However, the only consistent way to quantify one's uncertainty about an *unknown* parameter $\boldsymbol{\theta}$ is by specifying a probability distribution or equivalently a PDF for the parameter. After observing the data, this distribution should be updated to the PDF of the parameter *given* the data by incorporating the gained information. We are not interested in the PDF of the data given the *unknown* $\boldsymbol{\theta}$ but in the PDF of $\boldsymbol{\theta}$ given the *known* data; we want the "inverse probability", not $p(\mathbf{z}|\boldsymbol{\theta})$ but $p(\boldsymbol{\theta}|\mathbf{z})$. The only coherent approach to update a *prior* probability distribution with experimental information consists of calculating the *posterior* PDF via Bayes' theorem:

$$p(\boldsymbol{\theta}|\mathbf{z}) = \frac{p(\boldsymbol{\theta})p(\mathbf{z}|\boldsymbol{\theta})}{m(\mathbf{z})} \tag{1}$$

where $m(\mathbf{z}) = \int p(\mathbf{z}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}$ is the marginal PDF of $\mathbf{z}$. The denominator can be regarded as a normalization constant since it is independent of $\boldsymbol{\theta}$. The posterior PDF thus combines prior with likelihood information. A Bayesian point estimate of the unknown parameter would be the posterior mean, median, or mode.

The usual difficulties of Bayesian inference apply here: namely, the challenge of high–dimensional integration. One needs to be able to perform high-dimensional integrations when calculating the normalization constant $m(\mathbf{z})$, for instance, or when calculating the marginal PDF of a single component of the parameter vector $\boldsymbol{\theta}$ by integrating out all the other components, or when calculating posterior means. One of the main objectives of this paper is to demonstrate the potential of what statisticians refer to as simulation–based integration techniques for Bayesian posterior computation of CMB model parameters. These are called simulation–based, not because there is any simulation of the data, but because the posterior distribution is simulated. The particular simulation–based method we use here generates a set of parameter vectors whose distribution simulates the posterior distribution.

Analyses of cosmological parameter constraints from the most recent angular power spectrum determinations start with the calculation of the likelihood on an n_p–

dimensional grid, where n_p is the number of parameters [8, 12, 13]. The next step is then the reduction of this large amount of data to a series of one– or two–dimensional probability distributions functions which can easily be plotted. This is achieved by marginalizing over the other variables; i.e., integrating over them. Note that some times (and in all applications for at least some of the variables) this marginalization is approximated by a *maximization* over the other variables.

The chief drawback to the grid–based approach is the exponential increase in computing times and storage requirements with increasing number of parameters. Even with $n_p = 10$ (plus a handful of instrumental parameters) overly coarse grids and approximate treatment of some of the marginalizations are necessary to make the problem tractable. Including more parameters, for example ones that describe instrumental effects, foreground contaminants, or departures of the primordial power spectrum from a strict power–law, will make the grid–based approaches even more difficult if not impossible.

In this paper we implement the Bayesian approach to estimation of cosmological parameters using computer–intensive Markov chain Monte Carlo (MCMC) methods. Impediments to multidimensional integrations have been overcome by the progress made within the last decade in Bayesian computational technology via MCMC methods [18]. Since its initial application in digital signal analysis [19] MCMC methods have revolutionized many areas of applied statistics and we expect there to be an impact on cosmological parameter estimation from CMB measurements as well. A distinct advantage of the MCMC approach is that computational time does not grow exponentially with parameter number, as it does for other methods [18]. MCMC techniques have been applied in numerous areas, from science to economics [20]. Applications of state–space modeling in finance, e.g. stochastic volatility models applied to time series of daily exchange rates or returns of stock exchange indices, easily have 1000–5000 parameters [21]. Specially tailored MCMC algorithms can markedly improve the calculational speed [22, 23]. Hence, the MCMC approach to cosmological parameter estimation may provide the best strategy when testing complex models with numerous parameters.

We test our MCMC parameter estimation routine using simulated data. The likelihood is produced via a prototype fast calculator[24]. The toy model depends on four parameters. We demonstrate the validity of the technique with this example. The goal of our research is to apply MCMC methods to real CMB data, and we are currently optimizing our routines for this multiparameter (˜10) analysis. We are optimistic since experience has generally shown scaling of computing times with the number of parameters to be slower than exponential [18].

In section 2 we review methods that have been used to estimate cosmological parameters from CMB measurements, plus techniques for calculating likelihoods. In section 3 we describe the Bayesian approach to statistical inference and its implementation via Markov chain Monte Carlo methods. In section 4 we describe a method for applying MCMC methods to cosmological parameter estimation with CMB data. We test our method with a four parameter example, and our results are presented in Section 5. Section 6 presents our conclusions.

2. Current Statistical Methods with CMB Data

2.1. Parameter Estimation

Parameter estimation from CMB data is usually performed with a multi–step process, the last two steps of which are determining the power–spectrum from a map (or otherwise–pixelized data) [25, 26, 27], and then determining the cosmological parameters from the power spectrum. To date it has been impractical to go straight from the map to the parameters because of the computational expense of evaluating the probability of the map given the parameters thousands of times. Fortunately, the structure of the probability of the map given the power spectrum (the likelihood of the power spectrum) is much simpler than that of the map given the parameters. The simple structure allows for the mode to be found in a small number of iterations of a Levenberg–Marquardt–type search algorithm [27].

This "radical" compression of the information in the map to information in the power spectrum is possible because we assume that the signal in the maps is Gaussian and statistically isotropic. However, the uncertainties in the resulting power spectrum are not normally–distributed and the above search procedures do not allow one to completely characterize the distribution. Fortunately there are analytic approximations to the complete distribution whose parameters one can calculate with minor adaptations of the power–spectrum mode search algorithms [17, 28].

The most recent attempts at determining cosmological parameters ([8, 12, 13]) are in fact attempts to determine the posterior probability distribution of the cosmological parameters under various prior assumptions. The first step is to evaluate the likelihood on a grid of cosmological parameters. To have constraints of reduced–dimensionality (one or two) suitable for plotting, one marginalizes over the other parameters. Sometimes marginalization is approximated by simply maximizing over the remaining parameters. This approximation is exact if the marginalized parameters are normally distributed.

Direct grid–based evaluations of the likelihood have computation time and storage requirements which rise exponentially with the number of parameters. Such approaches will be difficult to implement for models of greater complexity than have been studied so far. Further, the maximization approximation to marginalization can sometimes lead to spurious results. Numerical maximization techniques, such as the Levenberg–Marquardt method, are only guaranteed to find a local maximum. Once they have reached a local maximum they might get stuck in their search and not reach the global maximum. For this reason statisticians have applied *simulated annealing*, which is a technique for global optimization. Simulated annealing has been attempted for cosmological parameter estimation [29, 30]. Simulated annealing is related to MCMC as its core component is the Metropolis Hastings algorithm. In this method the parameter space is searched in a random way. A new parameter space point is reached with a probability that depends on the likelihood and an *effective temperature* term. In the limit where the temperature approaches zero the *thermodynamics* of this parameter space search finds the system approaching the maximum of the likelihood. Although these methods are applicable to high–dimensional problems, they can be very inefficient in certain situations. Furthermore, there is no guarantee that the global maximum will be reached in a finite time. The efficiency depends very much on specifying a good cooling schedule which involves the arbitrary and skillful choice of various cooling parameters.

2.2. Calculating the Likelihood

Implicit in the problem of parameter estimation, whether from a frequentist or Bayesian perspective, is the calculation of the likelihood of the observed power spectrum (determined from a map) given some cosmological model. Thus likelihood calculation requires us to be able to calculate the angular power spectrum for a given model (with its associated cosmological parameters). This calculational task is accomplished with computer codes such as CMBfast [31] or CAMB [32]. These codes accept the cosmological parameters as input, and return the angular power spectrum of the CMB anisotropies, C_l. These software packages serve as the work–horses of current CMB parameter–determination efforts. For example, the likelihood has been calculated 30,311,820 times in order to cover a region in a ten–dimensional cosmological parameter space [13]. In other studies the likelihood was evaluated as needed within the calculation [11, 33].

3. Bayesian Posterior Computation via MCMC

Parameter estimation can be comprehensively described within the language of Bayesian inference. Application of Bayes' theorem is well–suited to astrophysical observations [34]. In Bayesian data analysis the model consists of a joint distribution over all unobserved (parameters) and observed (data) quantities. One conditions on the data to obtain the posterior distribution of the parameters. The starting point of the Bayesian approach to statistical inference is setting up a full probability model that consists of the *joint* probability distribution of all observables, denoted by $\mathbf{z} = (z_1, \ldots, z_n)$ and unobservable quantities, denoted by $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$. Using the notion of conditional probability, this joint PDF, $p(\mathbf{z}, \boldsymbol{\theta})$can be decomposed into the product of the PDF of all unobservables, $p(\boldsymbol{\theta})$, referred to as the *prior* PDF of $\boldsymbol{\theta}$, and the conditional PDF of the observables given the unobservables, $p(\mathbf{z}|\boldsymbol{\theta})$, referred to as the sampling distribution or *likelihood*, i.e.

$$p(\mathbf{z}, \boldsymbol{\theta}) = p(\boldsymbol{\theta})p(\mathbf{z}|\boldsymbol{\theta}). \tag{2}$$

The prior PDF contains all the information about the unobservables that is known from substantive knowledge and expert opinion *before* observing the data. All the information about the $\boldsymbol{\theta}$ that stems from the experiment is contained in the likelihood. In the light of the data, the Bayesian paradigm then updates the prior knowledge about $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$, to the *posterior* PDF of $\boldsymbol{\theta}$, $p(\boldsymbol{\theta}|\mathbf{z})$. This is done via an application of Bayes' theorem through conditioning on the observations

$$p(\boldsymbol{\theta}|\mathbf{z}) = \frac{p(\boldsymbol{\theta}, \mathbf{z})}{m(\mathbf{z})} \propto p(\boldsymbol{\theta})p(\mathbf{z}|\boldsymbol{\theta}) \tag{3}$$

where $m(\mathbf{z}) = \int p(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ is the marginal PDF of $\mathbf{z}$ which can be regarded as a normalizing constant as it is independent of $\boldsymbol{\theta}$. The Bayesian approach is based on the likelihood function but also quantifies the uncertainty about the parameters through a joint prior distribution that summarizes the available information about the parameters before observing the data. In the light of the observations, the information about the unknown parameters is then updated via Bayes' theorem to the posterior distribution which is proportional to the product of likelihood and prior density [35].

As already mentioned in the introduction, the main difficulty with the Bayesian approach to parameter estimation is high–dimensional integration. To calculate the normalizing constant of the joint posterior PDF, for instance, requires d–dimensional integration. Having obtained the joint posterior PDF of $\boldsymbol{\theta}$, the posterior PDF of a single parameter θ_i of interest can be obtained by integrating out all the other components, i.e.

$$p(\theta_i|\mathbf{z}) = \int \dots \int p(\boldsymbol{\theta}|\mathbf{z})d\theta_1 \dots d\theta_{i-1}d\theta_{i+1} \dots d\theta_d. \tag{4}$$

Calculation of the posterior mean of θ_i necessitates a further integration, e.g. $E[\theta_i|\mathbf{z}] = \int \theta_i p(\theta_i|\mathbf{z})d\theta_i$. This procedure is referred to as marginalization.

Only in the simplest situations can these integrals be solved analytically. The main approximate techniques are normal and Laplace approximations based on asymptotics, quadrature approximations, Monte Carlo integrations, and stochastic simulation. MCMC methods belong to the last category. For an overview see [36]. The deterministic techniques rely upon approximate normality and asymptotic results in the sense of the sample size growing to infinity. These techniques were mostly developed before the immense technological advances that enabled computer–intensive methods to be applied. The complexity of these techniques increases substantially with the dimension of the parameter space. In very broad terms, experience suggests that deterministic techniques provide good results for low–dimensional models. Similar comments are valid for non–iterative simulation techniques, since finding a suitable auxiliary distribution in rejections or importance sampling, for instance, becomes an extremely difficult task for high dimensions. As the dimension of the model increases, only iterative simulation–based integration techniques such as sampling–importance resampling (SIR) or MCMC provide adequate solutions. One major advantage of using a sampling–based approach to posterior computation is that once a sample from the posterior PDF, say $(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N)$ is available, we can employ this to estimate the posterior mean of each parameter by the sample average of the corresponding component, the marginal PDF's using kernel density estimates, and correlation between parameters using the sample correlations.

To avoid a time–consuming direct sampling of the joint posterior we propose a MCMC method [18, 37]. Instead of generating a sequence of independent samples from the joint posterior, in MCMC a Markov chain is constructed whose equilibrium distribution is just the joint posterior. Thus, after running the Markov chain for a certain *burn–in* period, one obtains (correlated) samples from the limiting distribution, provided that the Markov chain has reached convergence.

One method for constructing a Markov chain is via the Metropolis–Hastings algorithm. The MH algorithm was developed by Metropolis *et al.* [38] and generalized by Hastings [39]. It is a MCMC method which means that it generates a Markov chain whose equilibrium distribution is just the target density, here the joint posterior PDF $p(\boldsymbol{\theta}|\mathbf{z})$. This means that after sampling from this Markov chain for a sufficiently long time to allow the chain to reach equilibrium, the samples can be regarded as samples from the joint posterior PDF.

The MH algorithm shares the concept of a generating PDF with the well–known simulation technique of *rejection sampling*, where a candidate is generated from an auxiliary PDF and then accepted or rejected with some probability. However, the candidate generating PDF, $q(\boldsymbol{\theta}|\boldsymbol{\theta}_n)$ can now depend on the current state $\boldsymbol{\theta}_n$ of the Markov chain. A new candidate $\boldsymbol{\theta}'$ is accepted with a certain *acceptance probability* $\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}_n)$ also depending on the current state $\boldsymbol{\theta}_n$ given by:

$$\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}_n) = \min\left\{\frac{p(\boldsymbol{\theta}')p(\mathbf{z}|\boldsymbol{\theta}')q(\boldsymbol{\theta}_n|\boldsymbol{\theta}')}{p(\boldsymbol{\theta}_n)p(\mathbf{z}|\boldsymbol{\theta}_n)q(\boldsymbol{\theta}'|\boldsymbol{\theta}_n)}, 1\right\} \tag{5}$$

if $(p(\boldsymbol{\theta}_n)p(\mathbf{z}|\boldsymbol{\theta}_n)q(\boldsymbol{\theta}'|\boldsymbol{\theta}_n)) > 0$ and $\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}_n) = 1$ otherwise. The steps of the MH algorithm are therefore:

Step 0:	Start with an arbitrary value $\boldsymbol{\theta}_0$
Step $n+1$:	Generate $\boldsymbol{\theta}'$ from $q(\boldsymbol{\theta}|\boldsymbol{\theta}_n)$ and u from $U(0,1)$ If $u \leq \alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}_n)$ set $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}'$ (acceptance) If $u > \alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}_n)$ set $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n$ (rejection)

The MH algorithm does not require the normalization constant of the target density. Note that the outcomes from the MH algorithm can be regarded as a sample from the invariant density only after a certain *burn–in* period. Although the theory guarantees convergence for a wide variety of proposal PDFs, it does not say anything about the speed of convergence, i.e. how long the "burn–in" period should be. Convergence rates of MCMC algorithms are important topics of ongoing statistical research with little practical findings so far. There is no formula for determining the minimum length of an MCMC run beforehand, nor a method to confirm that a given chain has reached convergence. The only tests available are based on an empirical time series analysis of the sampled values and can only detect non–convergence. Thus, by performing a whole sequence of so–called convergence diagnostics with negative results, one only gains more confidence in one's hope that the chain reached equilibrium but never a guarantee. For issues concerning convergence diagnostics, the reader is referred to Cowles and Carlin [40].

Various convergence diagnostics have been developed and are implemented in CODA [41]. CODA is a menu–driven collection of functions for analyzing the output of the Markov chain. Besides trace plots and the usual tests for convergence, CODA calculates statistical summaries of the posterior distributions and kernel density estimates.

4. Applying MCMC Methods to CMB Anisotropy Data

Probability distribution functions for cosmological parameters given CMB anisotropy data can be computed in a Bayesian fashion with the MCMC serving as a means of conducting a proper marginalization over parameters. The implementation of the MCMC method is relatively straightforward. Instead of calculating the likelihood at uniform locations in the parameter space [8, 12, 13], one lets the MCMC do its intelligent walk through the space. Uniform *a priori* distributions for the parameters seem reasonable, so the MCMC could sample the parameter space defined by them. Since the likelihood function can not be written explicitly in terms of the cosmological parameters, but instead in terms of the CMB anisotropy power spectra, C_l, it is necessary to implement a Metropolis–Hastings MCMC routine.

The Markov chain can commence at a randomly selected position in parameter space $(\theta_1^{(0)}, \ldots, \theta_d^{(0)})$. With the parameter set one can then utilize a program like CMBfast [31] or CAMB [32] to generate an angular power spectrum and a likelihood can then be calculated. New values for the parameters $(\theta_1^{(1)}, \ldots, \theta_d^{(1)})$ are selected via sampling from the *a priori* distributions. However, these values are not necessarily *accepted* as new values. First their likelihoods are evaluated. Then the new values are *accepted* or rejected according to the following test; a random number, u, would be generated between 0 and 1. If $u \leq \min\left[1, (p(\theta_1^{(1)}, \ldots, \theta_d^{(1)})p(\mathbf{z}|\theta_1^{(1)}, \ldots, \theta_d^{(1)}))/(p(\theta_1^{(0)}, \ldots, \theta_d^{(0)})p(\mathbf{z}|\theta_1^{(0)}, \ldots, \theta_d^{(0)}))\right]$ where $p(\mathbf{z}|\theta_1^{(1)}, \ldots, \theta_d^{(1)})$ is the likelihood in terms of data $\mathbf{z}$ and cosmological parameters $(\theta_1, \ldots, \theta_d)$ then the new parameters is *accepted* into the chain, if not the next chain element has values equal to that of the previous state. A new set of parameters is then randomly sampled from the a priori distributions and the procedure continues.

The generated chain of parameter values forms the set from which the statistical properties would be derived. After running the Markov chain for a certain burn–in period (in order for the Markov chain to reach convergence) one obtains (correlated) samples from the limiting distribution. This process continues for a sufficiently long time (as determined by convergence diagnostics [41]).

After the *burn–in* the frequency of appearance of parameters represents the actual posterior density of the parameter. From the posterior density one can then create confidence intervals. Summary statistics are produced from the distribution, such as posterior mean and standard deviation. A cross–correlation matrix is also easily produced—which could prove to be a useful part of quantifying the multi–dimensional constraints.

The above method constitutes the simplest implementation of the Metropolis–Hastings method, that of an independence chain. We have just used an *acceptance probability* $\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta})$, defined by

$$\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}) = \min\left\{\frac{p(\boldsymbol{\theta}')p(\mathbf{z}|\boldsymbol{\theta}')q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{p(\boldsymbol{\theta})p(\mathbf{z}|\boldsymbol{\theta})q(\boldsymbol{\theta}'|\boldsymbol{\theta})}, 1\right\} \tag{6}$$

where the generating density $q(\boldsymbol{\theta}'|\boldsymbol{\theta})$ is the uniform density over the parameter space and thus, in particular, independent of the current state. By using uniform priors, the posterior PDFs in the acceptance probability calculation reduce to the likelihoods. However, the efficiency of a Metropolis–Hastings algorithm depends crucially on the form of the generating density $q(\boldsymbol{\theta}'|\boldsymbol{\theta})$. Just using a uniform distribution that does not even depend on the current state $\boldsymbol{\theta}$ is the simplest but probably least efficient way to accomplish the task. Even with a uniform distribution, the algorithm will be irreducible/aperiodic/reversible and thus the Markov chain will converge towards its stationary distribution.

A slightly better way might be to use a uniform distribution in a neighborhood of the current $\boldsymbol{\theta}$. Any prior information could be useful, such as correlations that one could use to specify a multivariate normal centered around the current $\boldsymbol{\theta}$ with a covariance matrix that takes said correlations into account. The optimization of the Metropolis–Hastings MCMC strategy will inevitably require experimentation with the generating density $q(\boldsymbol{\theta}'|\boldsymbol{\theta})$. While this may require some detailed study, the benefit will be the ability to generate posterior distributions for a large number of cosmological parameters.

An approximate formula for the Fisher matrix of the cosmological parameters which has often been used for forecasting parameter errors for CMB experiments may provide us with a useful generating density. The Fisher matrix is the expectation value of the second–derivative of the log of the likelihood and is given by:

$$F_{pp'}(\boldsymbol{\theta}) = \sum_l \frac{\partial C_l}{\partial \theta_p}\frac{\partial C_l}{\partial \theta_{p'}}\frac{1}{\sigma_l^2} \tag{7}$$

where the variance in each C_l determination can be approximated by

$$\sigma_l^2 = \frac{2}{(2l+1)f_{\rm sky}}\left(C_l + w^{-1}B_l^{-2}\right)^2 \tag{8}$$

where $f_{\rm sky}$ is the fraction of sky observed, $w^{-1} = \sigma_{\rm pix}^2\Omega_{\rm pix}$, $\sigma_{\rm pix}$ is the standard error in each map pixel, $\Omega_{\rm pix}$ is the size of the pixel and B_l is the Legendre transform of the beam profile. The generating density would then be

$$q(\boldsymbol{\theta}'|\boldsymbol{\theta}) \propto \exp\left(-\sum_{pp'}(\theta_p - \theta'_p)F_{pp'}(\boldsymbol{\theta}^*)(\theta_{p'} - \theta'_{p'})/2\right) \tag{9}$$

Note that the Fisher matrix is always evaluated at the same $\boldsymbol{\theta}^*$ to avoid having to calculate the derivatives in Eq. 7 at each iteration.

One may wish to know the posterior distribution of the parameters for a range of choices of priors. For example, in [8] the analysis was performed with 13 different priors each of which corresponded to a different choice about our knowledge of, e.g., the Hubble constant and the baryon density. For some parameter–determination methods adoption of a variety of priors is very costly since for each choice of prior the entire calculation must be redone. This is true, for example, of the Levenberg–Marquardt–type search algorithm used in [11]. In contrast, as emphasized in [13], evaluation of the likelihood on a grid allows for rapid calculation of various posteriors with different prior assumptions.

Fortunately, the sampled values from a Markov chain constructed assuming one prior can be used for other priors as well using *importance sampling* [42]. Suppose we have devised a MCMC sampler with stationary distribution $p^*(\theta|z) \propto p_1(\theta)p(z|\theta)$ with a prior p_1. Then we can estimate the expectation of an arbitrary function $g(\theta)$ of interest under the modified posterior $p(\theta|z) \propto p_2(\theta)p(z|\theta)$, i.e. using a new prior p_2, by importance reweighting the output $\theta_1, \ldots, \theta_N$ from the chain with stationary distribution p^*. Thus,

$$E_p[g(\theta)|z] \approx \sum_{i=1}^{N} \frac{w_i g(\theta_i)}{\sum_{i=1}^{N} w_i}$$

where the importance weight is $w_i = p_2(\theta_i)/p_1(\theta_i)$.

5. Test of Metropolis–Hastings Algorithm

We have successfully tested our MH method on simulated data. The ‘data’ we simulated is an angular power spectrum from $l = 100$ to $l = 800$ with normally distributed errors. The underlying model from which we made the realization is a cold dark matter model with baryonic matter density $\omega_b = 0.019$, dark matter density $\omega_d = 0.154$, $A = 100.0$ (in some units) and $n = 1.0$. The errors were realized as uncorrelated Gaussian random variables with zero mean and variance, σ_l^2 given by Eq. 8 assuming a noise–free map covering 10% of the sky.

To evaluate the likelihood of the data at any given point in the four–dimensional parameter space we evaluated:

$$-2\ln\mathcal{L} = \sum_l \left(C_l(\omega_b, \omega_d, A, n) - C_l^{\rm data}\right)^2/\sigma_l^2. \tag{10}$$

The most time–consuming step in the likelihood calculation is calculation of $C_l(\omega_b, \omega_d, A, n)$. We sped this up using a fast C_l calculator described very briefly here and in more detail elsewhere [24]. We use CMBfast to pre–compute $\Delta_l^2(k)$ on an 8 by 8 grid of values of ω_b and ω_d. The quantity $\Delta_l^2(k)$ gives the contribution from each wave number to the angular power spectrum: $C_l = (4\pi)^2 \int dk k^2 \Delta_l^2(k) P(k)$ where $P(k)$ is the primordial spectrum of density perturbations which we parametrize as a power–law: $P(k) = Ak^n$. Thus, for a given ω_b, ω_d, A and n we perform a multi–linear interpolation over the ω_b, ω_d grid of $\Delta_l^2(k)$ values and then do the integral over k.

A simple Fortran code was used for the MH routine, and it can be obtained at *http://physics.carleton.edu/Faculty/nelsonhome.html*. This routine calls the fast likelihood calculator for the likelihood values. The *a priori* distributions were all uniform, with ranges of $0.015 < \omega_b < 0.025$, $0.1 < \omega_d < 0.2.$, $0 < n < 2$ and $0 < A < 200$. The candidate generating densities were uniform and centered at the current chain parameter values and extending with ranges of 5×10^{-4} for ω_b, 5×10^{-3} for ω_d, 0.25 for A, and 0.01 for n. The results presented here were from a run of 200000 iterations, with an acceptance ratio of 41%. This calculation took ˜24 hours on a Sun Ultra 10 (440 MHz) workstation.

For our analysis we thinned the 200,000 cycle chain by accepting every 20th observation in order to avoid highly correlated values. Of the remaining 10,000 samples we used a burn–in of 1,000 which yields a final chain length of 9,000. Extensive convergence diagnostics were calculated for the four parameters using the CODA software [41]. All chains passed the Heidelberger–Welch stationarity test. The Rafferty–Lewis convergence diagnostics confirmed that the thinning and burn–in period were sufficient. Lags and autocorrelations within each chain were reasonably low for all parameters. These convergence diagnostics are summarized in [41], and references therein.

The trace plots and resulting kernel densities for the four parameters are shown in figure 1. Summary statistics including posterior mean, standard deviation, the time series standard error (the square root of the spectral density estimate divided by the sample size), and the 2.5%, 50%, and 97.5% credible regions are listed in table 1. The cross–correlation matrix is presented in table 2.

6. Discussion

The computational demands of Bayesian inference with large numbers of parameters are best met with MCMC methods. These methods have demonstrated their importance in numerous applications. As cosmological models grow in complexity it will become necessary to use techniques such as those discussed here in order to handle marginalization of parameters.

The demonstration of the Metropolis–Hastings routine with the toy model was successful. We plan to apply this technique to real CMB anisotropy power spectrum measurements. With MCMC techniques one can easily extend this method to ˜10 parameters. We are presently working on optimizing the speed of the likelihood calculator for the larger parameter number. Markov chain Monte Carlo methods provide a means of handling large parameter numbers, and maintain a rigorous approach to Bayesian inference. They may prove to be essential for certain CMB parameter–determination efforts in the near future.

7. Acknowledgements

This work was supported by the National Science Foundation, Royal Society of New Zealand Marsden Fund, the University of Auckland Research Committee, and Carleton College. LK thanks the Institut d'Astrophysique de Paris for their hospitality while some of this work was completed.

References

[*] Email addresses: nchriste@carleton.edu, meyer@stat.auckland.ac.nz, lknox@ucdavis.edu, lueyb@carleton.edu
[1] Mason B S 2001 *et al* astro-ph/0101171.
[2] Hanany S *et al Astrophys J. Lett.* 2000 **545** L5
[3] de Bernardis P *et al* 2000 *Nature* **4404** 95
[4] Miller A D, Caldwell R, Devlin M J, Dorwart W B, Herbig T, Nolta M R, Page L A, Puchalla J, Torbet E, and Tran H T 1999 *Astrophys. J. Lett.* **524** L1
[5] Mauskopf P D *et al* 2000 *Astrophys.J.* **536** L59
[6] Coble K A 1999 astro-ph/9911419
[7] Peterson J B, Griffin G S, Newcomb M G, Alvarez D L, Cantalupo C M, Morgan D, Miller K W, Ganga K, Pernic D, Thoma M 2000 *Astrophys. Phys. Lett.* **532** L83
[8] Lange A E *et al* 2001 *Phys. Rev. D* **63** 042001
[9] Balbi A *et al* 2000 *Astrophys. J. Lett.* **545** L1
[10] Knox L and Page L 2000 *Phys. Rev. Lett.* **85** 1366
[11] Dodelson S and Knox L 2000 *Phys. Rev. Lett.* **84** 3523
[12] Jaffe A *et al* 2000 astro-ph/0007333
[13] Tegmark M and Zaldarriaga M 2000 *Astrophys. J.* **544** 30
[14] Burles S, Nollett K M, Turner M S 2000, astro-ph/0008495 and *Phys. Rev. D* in press
[15] http://map.gsfc.nasa.gov/
[16] http://astro.estec.esa.nl/Planck/
[17] Bond J R, Jaffe A H, Knox L 2000 *Astrophys. J.* **533** 19
[18] Gilks W R, Richardson S and Spiegelhalter D J 1996 *Markov Chain Monte Carlo in Practice* (London: Chapman and Hall)
[19] Geman S and Geman D 1984 *IEEE Trans. Patter. Anal Mach. Intell.* **6** 721
[20] Meyer R, *Data Analysis, Classification, and Applications* 2000 ed W Gaul, O Opitz and M. Schader, (Berlin: Springer) pp 259-271
[21] Jacquier E, N.G. Polson, P.E. Rossi, *J. Bus. Econ. Stat.* **12**, 371 (1994).
[22] Meyer R and Yu J 2000 *Econometrics J.* **3** 198
[23] Kim S, Shephard N and Chib S 1998 *Rev. Econ. Studies* **65**, 361
[24] Knox L and Skordis C 2001 in preparation
[25] Tegmark M 1997 *Phys. Rev. D* **55** 5895
[26] Oh S P, Spergel D N and Hinshaw G 1999 *Astrophys. J.* **510** 551
[27] Bond J R, Jaffe A H, Knox L 1998 *Phys. Rev. D* **57** 2117
[28] Bartlett J G, Douspis M, Blanchard A and Le Dour M 2000 Astron. Astrophys. Sup. **146** 507
[29] Hannestad S 1999 *Phys. Rev D* **61** 023002
[30] Knox L 1995 *Phys. Rev D* **52** 4307
[31] Seljak U and Zaldarriaga M 1996 *Astrophys. J.* **469** 437
[32] Lewis A, Challinor A and Lasenby A 2000 *Astrophys. J* . **538** *473*
[33] Melchiorri A *et al* 2000 *Astrophys. J. Lett.* **536** L63
[34] Gregory P C and Loredo T J 1992 *Astrophys. J.* **398** 146

[35] Carlin B P and Louis T A 1996 *Bayes and Emperical Bayes Methods for Data Analysis* (London: Chapman and Hall)
[36] Evans M. and Swartz T 1995 *Statistical Science* **10**, 254
[37] Christensen N L and Meyer R 1998 *Phys. Rev. D* **58** 082001
[38] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H, and Teller E 1953 *J. Chem. Phys.* **21**, 1087
[39] Hastings W K 1970 *Biometrika* **57**, 97
[40] Cowles M K and Carlin B P 1996 *J. Am. Stat. Assoc.* **91** 883
[41] Best N G, Cowles M K and Vines S K 1995 *CODA manual version 0.30* (Cambridge: MRC Biostatistics Unit)
[42] Ripley B D 1987 *Stochastic Simulation* (New York: Wiley)

Figure and Table Captions

Figure 1. Trace and kernel density plots of the marginal posterior distributions for the parameters A (amp), ω_b (b), ω_d (d), and n (tilt). The *true* parameters are $A = 100.0$, $\omega_b = 0.019$, $\omega_d = 0.154$ and $n = 1.0$.

Table 1. The posterior mean, standard deviation, time series standard error (SE), and the 2.5%, 50% (Median) and 97.5% credible regions of the parameters A, ω_b, ω_d and n (cf. figure 1).

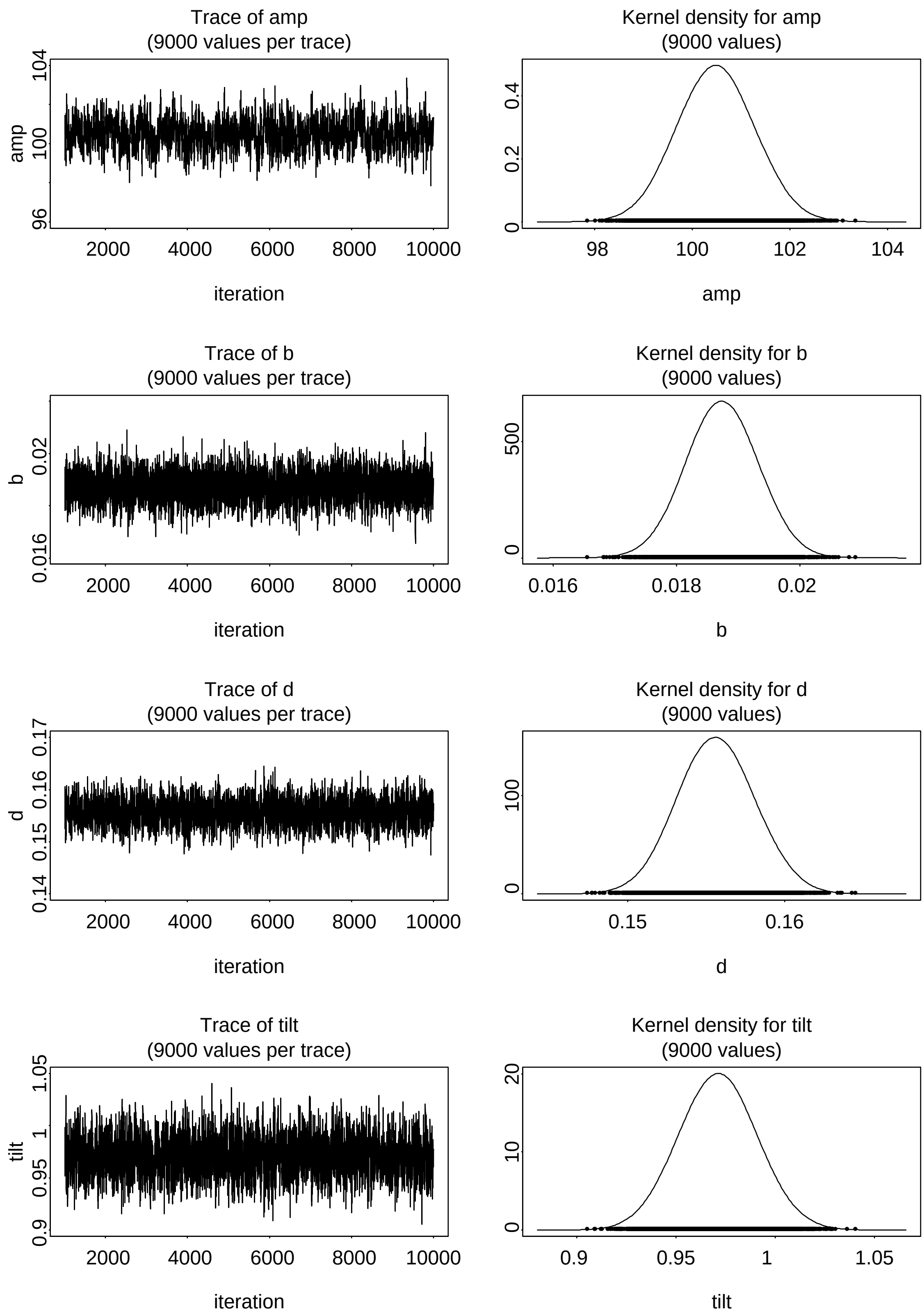
Table 2. The cross–correlation matrix of the parameters A, ω_b, ω_d and n (cf. figure 1).

TABLE 1

Parameter	Mean	SD	SE	2.5%	Median	97.5%
A	100.0	0.729	$7.69x10^{-3}$	99.1	100.0	102.0
ω_b	0.0187	$5.28x10^{-4}$	$5.56x10^{-6}$	0.0177	0.0187	0.0198
ω_d	0.156	$2.27x10^{-3}$	$2.39x10^{-5}$	0.151	0.156	0.160
n	0.971	$1.81x10^{-2}$	$1.90x10^{-4}$	0.936	0.971	1.010

TABLE 2

Parameter	A	ω_b	ω_d	n
A	1.0			
ω_b	0.220	1.0		
ω_d	0.614	0.014	1.0	
n	-0.189	0.420	-0.440	1.0



Trace of amp
(9000 values per trace)
amp
iteration
Kernel density for amp
(9000 values)
amp
Trace of b
(9000 values per trace)
b
iteration
Kernel density for b
(9000 values)
b
Trace of d
(9000 values per trace)
d
iteration
Kernel density for d
(9000 values)
d
Trace of tilt
(9000 values per trace)
tilt
iteration
Kernel density for tilt
(9000 values)
tilt