Power Spectrum Estimation for Cosmic Microwave Background and Large Scale Structure

Suhail Dhawan
Supervisors: Dr. Filipe Abdalla and Dr. Sreekumar Balan

March 20, 2013

Abstract

In our project, we wish to apply the techniques for cosmic microwave background (CMB) power spectrum estimation to data from large scale structure (LSS) surveys. We want to look specifically at the Guided Hamiltonian Sampler (GHS), a modified Markov Chain Monte Carlo (MCMC) sampling method as a tool to estimate LSS power spectra. We wish to evaluate its efficiency with high dimensional probability distributions. In this literature review, we present the methods that have been used in the past, in both CMB and LSS cosmology, to estimate the power spectra. We look at a Bayesian framework for estimating the parameters of interest in an optimal, efficient and unbiased manner and how it has been implemented in the past to perform cosmological analysis. Since modern cosmology has entered an era of extreme precision with large datasets, we explain the motivation for using sampling methods and their implementation in the past with surveys like WMAP and SDSS .

We also present some preliminary analysis of our own, using the GHS on simulated CMB data, gauging the sampler performance with a broad range of input parameters, and give a short summary of the work to be done in the future.

Contents

1 Introduction

The discovery of anisotropy of the cosmic microwave background was one of the greatest scientific achievement of the last century. The group working with CMB maps from the COBE satellite, a dataset of ~ 4000 dimensions (Figure 1 shows the 4-year COBE power spectrum from Tegmark et al. [1996]), found temperature fluctuations of 1 part in 100000 from the mean temperature of 2.73K. This culminated in the award of the 2006 Nobel prize in physics to Professors John Mather and George Smoot. The nature of these fluctuations offers insight into conditions in the very early universe and the seeds of structure formation. A clear picture of early universe cosmology is emerging as a result of CMB probes. A review of CMB cosmology is presented in Challinor [2012], Hu and White [2004].

Galaxy redshift surveys are a powerful tool to measure a contrasting cosmic epoch (Peebles [1973]). They return a substantial amount of information about the large scale structure of the universe and its evolution through gravitational amplification from primordial perturbations. The two-point galaxy correlation function provides a test for standard inflation and different cosmological models that predict the observed structure of the universe. The large scale structure also has imprinted in it standard length scales which yield oscillatory features in the power spectrum, called Baryon Acoustic Oscillations (BAO). Since the physics of BAO is well understood, this allows for an accurate distance redshift relation and an independent probe of the accelerated expansion of the universe.

Power spectra present valuable information about the nature of the fluctuations (CMB temperature or galaxy count) at different angular scales to offer poignant insights into the history of the universe and its structure on the large scales. Estimated power spectra are also compared with their theoretical counterparts to accurately constrain cosmological parameters from the observed data (Dunkley et al. [2009]). In this literature review, we look at the methods employed by CMB and LSS research groups to estimate power spectra from survey datasets.

1.1 Motivation

Over the years, cosmology has evolved into a field where brute force methods are not feasible for data analysis. With surveys returning high resolution maps of the universe, there is a pressing need for novel techniques to be implemented so that results are optimal, efficient and computationally feasible.

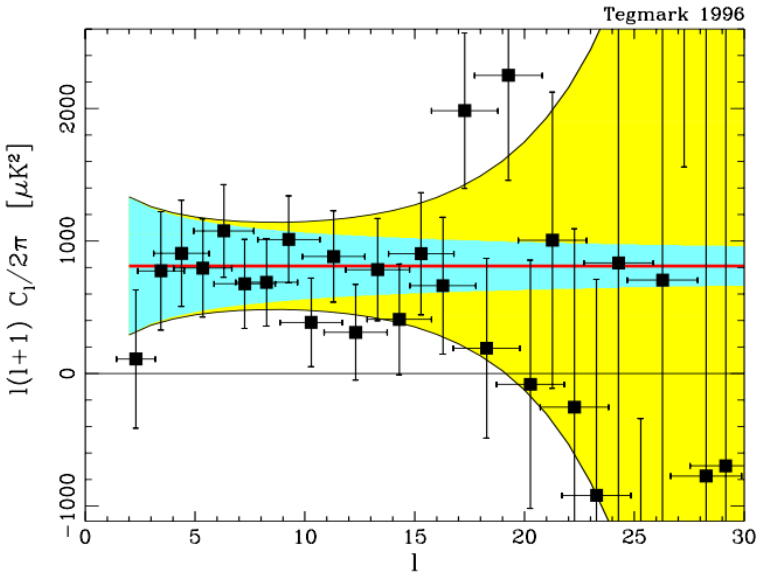

Figure 1: The 4-year COBE power spectrum Tegmark [1996]. The blue region is the 1-sigma error bar. The vertical error bars account for the error from pixel noise as well as cosmic variance. The horizontal error bars reflect the width of the window function used. The dark shaded region shows the contribution from cosmic variance. We can see that the power spectrum only extends to about l=30.

There has been an increase in usage of sampling techniques for estimating power spectrum coefficients. For our work, we wish to examine the application of GHS developed in Balan [2012] to LSS data and analyse its performance with dimensionality of the dataset. We also want compare GHS to conventional power spectrum estimation methods. It is an exciting and unique task, since this method has not been applied to LSS data before.

1.2 Overview

The estimation methods in this review include pseudo-C_l, (Peebles [1973]), Maximum Likelihood (Gorski et al. [1994]), and a plethora of sampling methods including Metropolis-Hastings (Metropolis et al. [1953], Hastings [1970]), Gibbs Sampling (Wandelt et al. [2004]), Hamiltonian Monte Carlo (Taylor et al. [2008]) and the Guided Hamiltonian Sampler (Balan [2012]), which we shall be using for our analysis.

The structure of the review is as follows. In section 2 we introduce the basics of Bayesian parameter estimation and the numerical methods that have

been used in cosmology. In section 3 we describe the nature of datasets at hand and the application of conventional power spectrum estimation methods and the numerical methods introduced in section 2 to CMB datasets. In section 4 we present a brief summary of our preliminary analysis of simulated CMB data with GHS. In section 5 we look at the application of estimators to LSS data. Section 6 is a brief conclusion. Finally, we look at the work to be done in the future for LSS power spectrum estimation in section 7.

2 Bayesian Parameter Estimation

Problems in physics mostly boil down to finding patterns in observations. These patterns are expressed as mathematical models. In order to discriminate between several different models we estimate the value of parameters contained by them. Bayes' theorem allows us make this estimate of a parameter or set of parameters given a dataset with the knowledge of the likelihood of having observed that data and the prior information we have about the parameter from the model. Bayesian inference has been used in several different areas of astronomy and astrophysics some of which are cosmic ray physics (Watson et al. [2011]), supernova cosmology (et al. [2011b], Mandel et al. [2008]), dark matter detection (Arina [2011]), gravity wave background from pulsar timing arrays (Lentati et al. [2012]) and even the search for exoplanets (Balan et al. [2008]).

In this section, we look at the fundamentals of Bayesian parameter estimation, the numerical methods for sampling the probability distribution of the parameter given the data and the tests for convergence of these numerical techniques. Trotta [2008] provides a holistic review of Bayesian parameter estimation. [1]

2.1 Basic Fundamentals

In cosmological data analysis we use a Bayesian outlook of assigning degrees of belief or 'weights' to a given proposition or value that a parameter can take. This approach was first introduced by Revd.Thomas Bayes in 1763 in his seminal article 'An essay towards solving a Problem in the Doctrine of Chances'. Bayes' theorem for estimating parameter $\mathbf{\Theta}$ given data D and model H is written as

[1] In this review we follow the notation from Balan [2012] for Bayesian Parameter Estimation and introducing the nature of CMB and LSS data

$$\Pr(\boldsymbol{\Theta}|D,H) = \frac{\Pr(D|\boldsymbol{\Theta},H)\Pr(\boldsymbol{\Theta}|H)}{\Pr(D|H)} \tag{1}$$

The theorem relates the posterior probability $\Pr(\boldsymbol{\Theta}|D,H)$, the likelihood $\Pr(D|\boldsymbol{\Theta},H)$ and the prior knowledge $\Pr(\boldsymbol{\Theta}|H)$. The Bayesian evidence $\Pr(D|H)$ is independent of the parameter $\boldsymbol{\Theta}$, we can treat is as a normalizing constant and write the above expression as

$$\Pr(\boldsymbol{\Theta}|D,H) \propto \Pr(D|\boldsymbol{\Theta},H)\Pr(\boldsymbol{\Theta}|H) \tag{2}$$

2.2 Numerical Methods

To evaluate the posterior in (2), we require numerical simulation since analytic solutions either don't exist or are not sufficiently accurate. In this subsection we look at different numerical recipes to sample the posterior. These are Metropolis-Hastings, Gibbs Sampling, Hamiltonian Sampling and GHS, Importance Sampling, Nested Sampling, Slice Sampling.

2.2.1 Metropolis-Hastings

The most common method used for sampling a posterior distribution is the Metropolis-Hastings (MH) algorithm, posited independently in Metropolis et al. [1953] and Hastings [1970].It creates a sequence of points in parameter space whose density is proportional to the posterior distribution. Given initial state x_0, a new proposed point x_1 drawn from a proposal distribution $Q(x_0, x_1)$ is accepted with probability A [2]

$$A = min[1, \frac{Pr(x_1)Q(x_1,x_0)}{Pr(x_0)Q(x_0,x_1)}] \tag{3}$$

If the proposal distribution is symmetric about x_1 and x_0, $Q(x_1, x_0) = Q(x_0, x_1)$, A reduces to the ratio of the two probabilities. The process is repeated till enough samples are obtained for estimating the desired probability distribution.

2.2.2 Gibbs Sampling

Since the acceptance in the MH algorithm depends on the ratio of the probabilities, many of the points will be rejected. Gibbs sampling removes the

[2]Some papers use a to denote acceptance criterion, we have used an upper case A to distinguish it from CMB signal realization **a** which we will introduce in section 3 and use thereafter.

need to reject points, thus, each one can be used to build the posterior. It is done by presuming prior knowledge of the conditional distributions of the parameters. A review of Gibbs sampler algorithm and its application to CMB data can be found in Groeneboom [2009].

For a model with 2 parameters θ_1 and θ_2, we wish to map $\Pr(\theta_1, \theta_2)$. We define a proposal distribution T for θ_2 as a conditional distribution

$$T(\theta_1^{i+1}, \theta_2^{i+1}|\theta_2^i, \theta_1^i) = \delta(\theta_1^{i+1} - \theta_1^i) \Pr(\theta_2^{i+1}|\theta_1^i) \tag{4}$$

Where the proposal is considered only for θ_1^{i+1}=θ_1^i, i.e. when θ_1 is fixed and θ_2 is varied. Applying the Metropolis Hastings acceptance criterion as in (3):

$$A = \frac{T(\theta_1^{i+1}, \theta_2^{i+1}|\theta_2^i, \theta_1^i)}{T(\theta_1^i, \theta_2^i|\theta_2^{i+1}, \theta_1^{i+1})} \frac{\Pr(\theta_1^{i+1}, \theta_2^{i+1})}{\Pr(\theta_1^i, \theta_2^i)} \tag{5}$$

Enforcing the delta function such that θ_1^{i+1}=θ_1^i, sampling from the conditional distribution makes it such that the terms cancel out making

$$A = 1 \tag{6}$$

Hence one can alternate from known conditional distributions, where each step is independently accepted and can be performed as many times as needed.

2.2.3 Hamiltonian Monte Carlo and GHS

This technique for sampling is based on the equations from Hamiltonian dynamics.

$$H = \sum_i^N \frac{p_i^2}{2m} + \Psi(x) \tag{7}$$

where $\Psi(x)$ is the negative log of the posterior probability distribution

$$\Psi(x) = -\log \Pr(x) \tag{8}$$

The objective in the sampling methodology is the sample from a distribution which is proportional to $\exp(-H)$ and can be written as,

$$\exp(-H) \propto \Pr(x) \prod_i^N exp(\frac{-p_i^2}{2m}) \tag{9}$$

From (x_0, p_0) as the initial starting point, the system is evolved deterministically to a new point in phase space (x_1, p_1). This process follows

$$\frac{dx}{dt} = \frac{\partial H}{\partial p_i} \tag{10}$$

$$\frac{dp}{dt} = -\frac{\partial H}{\partial x_i} \tag{11}$$

The new point is accepted with the probability

$$p_A = min(1, \exp(-\delta H)) \tag{12}$$

Where

$$\delta H = H(x_1, p_1) - H(x_0, p_0) \tag{13}$$

A leapfrog step for evolving (x_0, p_0) to (x_1, p_1) is repeated n times with a finite step size ϵ, so that the total evolution time is T=nϵ. These steps are given in [Neal, 1995]. In order to randomize the phase space trajectories we draw n and ϵ from uniform distributions U(1, N), U(0, E) respectively, where N and E are the maximum values that the number of steps and the step size can take. This allows for a random variation of the evolution time T.

For our project, we wish to use the Guided Hamiltonian Sampler. The technique is outlined in Balan [2012] and is summarized by the following steps:

1. Find the peak of the distribution
2. Calculate the Hessian of the negative log posterior at the peak
3. Perform eigen-decomposition of the Hessian and obtain the principal coordinates
4. Set step sizes as the inverse square root of the eigenvalues of the Hessian
5. Set the dimensionality scaling factor η, and sample in the principal coordinates
6. Calculate convergence statistics and stop once condition is met

2.2.4 Other methods

There are other sampling techniques for Bayesian inference like Nested Sampling, Slice Sampling and Importance Sampling that are effective variants of the traditional markov chain algorithm. In **Nested sampling**, the quantity of interest is the Bayesian evidence to select from different cosmological models. It is given as the marginalized integral of the likelihood (L) over the prior mass (X).

$$E = \int L dX \tag{14}$$

Slice Sampling works on the principle that sampling a random variable can be done by sampling a uniform distribution from the region under its density function. An introduction is provided in Neal [2003]. **Importance Sampling**, as the name suggests, does not sample from a uniform distribution, but, instead, a distribution where the points of large integral value of the function are concentrated.

$$I = \int \frac{f(x)}{g(x)} g(x) dx \tag{15}$$

where g(x) is a reasonable approximation to f(x). The integral can be evaluated by choosing random points from a distribution of g(x) and calculating the ratio at these points. Tokdar and Kass [2010] reviews Importance Sampling.

2.3 Convergence Tests

Convergence tests help us know that we are sampling from the target posterior distribution and that there is no bias in the samples obtained. In Cowles and Carlin [1996] there is a brilliant review of a collection of convergence statistics like Gelman and Rubin (Gelman and Rubin [1992]), Raftery and Lewis (Raftery and Lewis [1992]). They conclude that there are certain characteristics that all tests fail to detect whilst looking for convergence of the samples to the target posterior.

Broadly, the techniques can be split into two groups: inter-chain and intra-chain. In the first category, a number of chains are taken from different initial points in the sample space and their convergence to the target posterior is compared. In intra-chain methods, a single chain is tested for convergence to the posterior by defining a statistic and a range of values for the statistic for which the chain is said to have converged. Hanson [2001] provides a diagnostic method which is based on intra-chain convergence.

Hanson statistic for the i th dimension is:

$$H_i = \frac{\frac{1}{M}\Sigma_k (x_i^k - \bar{x}_i^k)^2}{\frac{1}{3M}\Sigma_k (x_i^k - \bar{x}_i^k)^3 \frac{\partial \Psi}{\partial x_i}|_{x_i^k}} \tag{16}$$

The chain is said to have converged when H approaches 1.The range 0.8 to 1.2 is good convergence and 0.6 and 1.4 is acceptable. Gelman and Rubin employs the inter-chain approach and compares interchain and intra-chain statistics. Hanson statistic only compares two intra-chain statistics, is easy to compute, and only requires one chain, compared to 10 for Gelman and Rubin. Since Hanson values are a bit pessimistic, other techniques like the one in Dunkley et al. (2005) can also be investigated.

It is seen that for correlated samples, the convergence takes a very long time. The introduction of masks or a high rms noise value causes the process to be less Markovian, requiring more samples for the convergence statistic to be within the desired range. We see this effect for small scales in CMB data where samples have higher autocorrelation values.

3 Methods for CMB power spectrum estimation

The creation of digitized CMB maps is the intermediate stage between getting large time-ordered data (TOD) streams and astrophysical analysis. An approach for the high resolution numerical representation of functions on a sphere and the usage of this geometry to simulating CMB maps is introduced in the HEALPIX primer (Wandelt et al. [2010]). In this section, we introduce the CMB data and methods to estimate power spectra from the CMB maps.

Both CMB and LSS data can be written as a sum of a signal **s** and noise **n**

$$\mathbf{d} = \mathbf{s} + \mathbf{n} \tag{17}$$

the temperature **t** can be mapped into the signal s through the mapping R.

$$\mathbf{d} = \mathbf{R}\mathbf{t} + \mathbf{n} \tag{18}$$

Using the spherical harmonic decomposition from the angular solution to laplace's equation in spherical polar coordinates.

$$\mathbf{t}(x_p) = \sum_{l=2}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{lm} \tag{19}$$

$$\mathbf{d} = \mathbf{RYa} + \mathbf{n} \tag{20}$$

Where Y is the orthogonal basis vector and a is the spherical harmonic coefficient.

The CMB sky is an isotropic Gaussian random field, the elements of the covariance matrix are given by

$$C_{lml'm'} = \langle a_{lm} a_{l'm'} \rangle \tag{21}$$

Where the set of coefficients Cl defines the angular power spectrum and since the sky is real

$$a_{lm} = a^*_{l,-m} \tag{22}$$

For CMB datasets, both the signal and the noise are assumed to be realizations of independent Gaussian processes

$$\langle \mathbf{dd^T} \rangle = \langle \mathbf{ss^T} \rangle + \langle \mathbf{nn^T} \rangle \tag{23}$$

Which can be written in matrix formulation as D=C+N, where C is the signal covariance matrix and N is the noise covariance matrix. d is the data vector.

With smaller datasets and lower angular scales, one can use several methods that do not involve sampling in order to obtain the power spectrum from the data. These are maximum likelihood and pseudo C_l. Efstathiou [2008] provides a review of estimators.

3.1 Maximum Likelihood

This was the first method used in the early 90's for CMB power spectrum estimation with data from the COBE satellite. Maximum likelihood (Gorski 1994, Bond et al. [1998], Oh et al. [1999]) is an optimal and unbiased method for the estimation of the power spectrum from observed TOD. It is helpful in estimating the coefficients from interferometric experiments for the measurement of CMB anisotropies (Hobson and Maisenger [2002]).

The likelihood function is defined as the probability of observing data given a set of power spectrum coefficients.

$$L(C_l) = \Pr(\mathbf{d}|C_l) \tag{24}$$

We construct the likelihood from a marginalization of the joint probability distribution of the data and the signal, conditional on the $C'_l s$.

$$L = \Pr(\mathbf{d}|C_l) = \int \Pr(\mathbf{d}, \mathbf{a}|C_l) d\mathbf{a} = \int \Pr(\mathbf{d}|\mathbf{a}, C_l) \Pr(\mathbf{a}|C_l) d\mathbf{a} \tag{25}$$

Since the data is dependent on C_l via the signal

$$L = \Pr(\mathbf{d}|C_l) = \int \Pr(\mathbf{d}|\mathbf{a}) \Pr(\mathbf{a}|C_l) d\mathbf{a} \tag{26}$$

Under assumption of gaussianity of noise and signal, the likelihood can be written completely and exactly as

$$L = \frac{exp[\frac{-(\mathbf{d}-\mathbf{a})^T N^{-1} (\mathbf{d}-\mathbf{a})}{2}]}{(2\pi)^{N/2} (det(C))^{1/2}} \tag{27}$$

Methods to perform this maximization are described in Bond et al. [1998]. Since the calculation of the likelihood requires the inversion and multiplication of the covariance matrix which is of order $O(N_{pix}^3)$, where N_{pix} is the number of pixels, this method is not feasible for megapixel datasets like Planck or WMAP. There are approximations of the covariance matrix that reduce the order to $O(N_{pix}^2)$, however, the reduction is not sufficient and it also leads to loss of information and creeping in of bias since the low $C_l s$ will be having low variance.

Borrill [1998] explains the quadratic estimator for maximize the likelihood function. In order to rapidly search for the maximum of L

$$\frac{\partial \log L}{\partial C_l} = 0 \tag{28}$$

is iteratively evaluated with the Newton Raphson method. In Tegmark [1996], the COBE power spectrum is estimated by taking a weighted linear combination of all pixels, squaring it (as is shown in the relation between $a'_{lm} s$ and $C'_l s$ and subtracting noise. The weights are plotted as a sky map, much like the temperature data itself. It is argued that if the power spectrum is parametrized by a small number of parameters an effective likelihood analysis for estimation can be performed.

3.2 Pseudo C_l

A power spectrum is the square of the Fourier transform of the density for a distribution in flat space and of the spherical harmonic transform of the density for a distribution in three dimensions. This method helps in establishing contamination of the pseudo-spectrum by irregularities in different parts of the actual spectrum. The method has been used for several CMB experiments like BOOMERANG (Jones et al. [2005]) and MAXIMA (Stompor et al. [2003]). A discussion of its application to CMB cosmology can be found in Hivon et al. [2002].

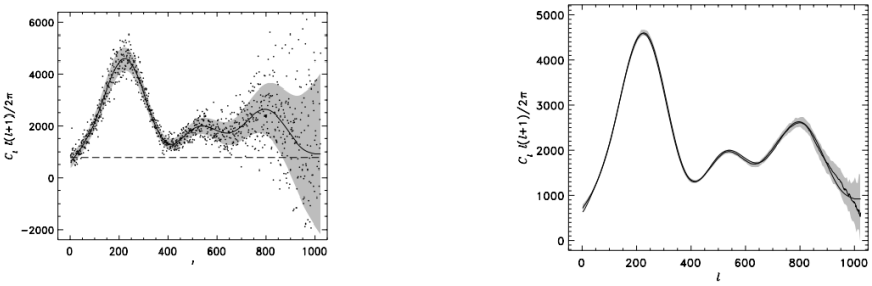

Figure 2: **Left** *: The power spectrum for a single simulated 1024 X 2048 pixel map (Oh et al.). The gray band is the 1 sigma region given by the inverse fisher matrix.* **Right** *: the smoothing spline fit to the points in the* **left** *panel. The dark line is the estimated power spectrum whereas the light line is the input power spectrum*

This method allows for the estimation of a pseudo-spectrum by using the spherical harmonic coefficients of the signal presuming noise-free, full sky coverage. It can be used for datasets of any size and is an efficient method; however it is not feasible at large scales (low multipoles). The $C_l s$ can be estimated by:

$$C_l = \frac{1}{2l+1} \sum_m |a_{lm}|^2 \tag{29}$$

Emission from galactic foregrounds prevents us from having a full sky coverage. Survey strategy is also important since number of observations for different parts of the sky may vary. There is a significant contribution from detector noise and a convolution with an observing beam. In pseudo-C_l weights w($\mathbf{x}$) are assigned to each position to encode not only the partial sky coverage, but also all other effects that cause a deviation from the ideal power spectrum.

$$\tilde{C}_l = \frac{1}{2l+1} \sum_m |a_{\tilde{lm}}|^2 \tag{30}$$

We can incorporate this weighting function into the spherical harmonic coefficients

$$\tilde{d_{lm}} = \int d(\mathbf{x}) w(\mathbf{x}) Y^*_{lm} d\mathbf{x} \tag{31}$$

$$\tilde{d_{lm}} \sim \Omega_p \sum_p d(x_p) w(x_p) Y^*_{lm}(x_p) \tag{32}$$

where Ω_p is the area of the individual pixel. Here, the integral is approximate by the sum over the pixels of the map, assuming that all pixels have the same area.

In order to correlate the pseudo-spectrum with the actual power spectrum, we use the mode-mode coupling kernel. This follows from the relation between the average of the pseudo-spectrum coefficients and the actual $C_l s$.

$$\langle \tilde{C}_l \rangle = \sum_k M_{lk} B_l^2 \langle C_k \rangle + \langle \tilde{N}_l \rangle \tag{33}$$

Where B_l is the window function describing the smoothing effect from beam convolution and finite pixel size. $\langle \tilde{N}_l \rangle$ is the average noise power spectrum. The task now is to deduce the nature of the M_{lk} coupling kernel. We can do the arithmetic in planar or spherical geometry. The former presents a more simple and familiar formulation. A detailed recipe for how M_{lk} can be written in terms of the vector in polar coordinates $\mathbf{k}$ can be seen in Appendix A of Hivon et al. [2002]

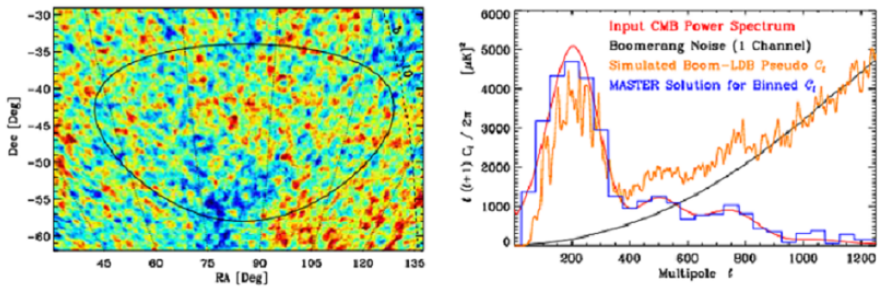

Figure 3: **Left***: a realisation of the CMB sky from a theoretical model described by the power spectrum shown (Hivon et al [2002]) by the red line on the right panel.* **Right***: Power spectrum smoothed by beam and pixel window function (red line), instrumental noise power spectrum (black), binned MASTER estimate of the full sky power spectrum (blue histogram) pseudo-Cl's from simulated Boom-LDB experiment (orange)*

3.3 Sampling Methods for CMB data

Novel algorithms based on traditional MCMC methods provide an optimal and computationally feasible route to estimate power spectrum from large datasets. We look at the application of these methods to CMB data.

3.3.1 Gibbs Sampling

The usage of Gibbs Sampling in CMB cosmology was introduced in Wandelt et al. [2004]. Their method is effective since it doesn't require any prior assumptions for the survey geometry, scanning strategy and instrument noise. The equations for sampling the conditional distributions of $a'_{lm}s$ and $C'_l s$ are given by:

$$a^{i+1} \leftarrow Pr(a^i|Cl^i, d) \tag{34}$$

$$Cl^{i+1} \leftarrow Pr(Cl^i|a^{i+1}, d) \tag{35}$$

The conditional distribution for a_{lm} is Gaussian and the one for the C_l is inverse gamma. The inverse gamma distribution is algorithmically non-trivial since the proposal distribution for the sampling needs to be carefully chosen in order to prevent a low acceptance rate or a high level of correlation. It, however, is a computationally inexpensive step. The Gaussian distribution becomes unfeasible to evaluate at million dimensions since the inversion of the covariance matrix is an expensive step.

For low signal to noise data, the direct Gibbs sampler is inefficient. Since the joint distribution is much wider than either conditional distribution, convergence is not possible. In Jewell et al. [2008], solved by using an algorithm that involves both the Gibbs sampling and the Metropolis Hastings methodologies. The method involves a simple Metropolis Hastings MC with a general proposal rule for the power spectrum, C_l, followed by a particular deterministic rescaling of the sky signal $\mathbf{s}$. The acceptance rate for this algorithm depends on the χ squared between the original and the proposed sky sample, which is close to unity for the low signal to noise regime.

Gibbs sampling the analysis of WMAP data. For the estimation of the power spectra, the pseudo C_l method from Hivon et al. [2002] is used for the low angular scales of $l \geq 32$. For the large scale multipoles, a Blackwell-Rao estimator is used on the Gibbs sampler chains. This is a direct Gibbs sampler, used for sampling the parameter space. In the evaluation of the $C'_l s$, the entire set of coefficients is found using MASTER algorithm and the l<32 are truncated.

In Larson et al. [2010], the seven-year WMAP data is used to estimate polarization and temperature power spectra as well as cosmological parameters like Ω_m and Ω_Λ. The signal to noise ratio exceeded unity for l<919. In band-power widths of $\Delta l = 10$, the SNR exceeds 1 for l =1060. The third acoustic peak is well measured by WMAP in this paper.

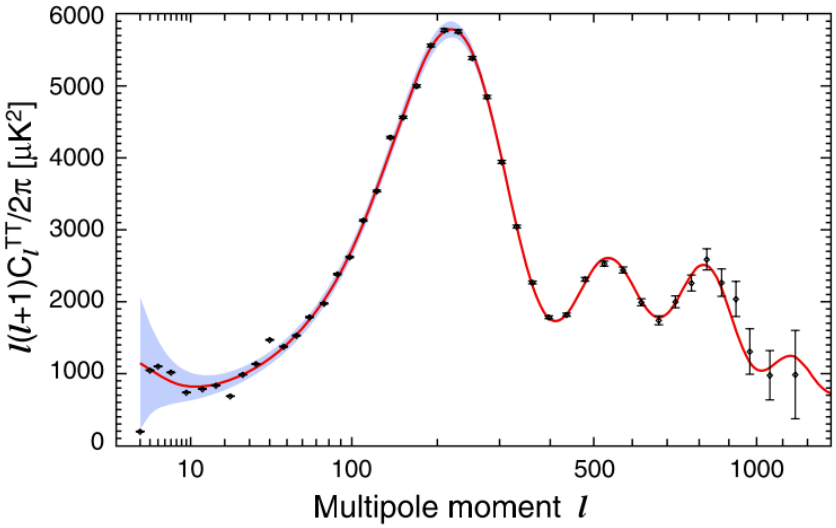

Figure 4: The WMAP seven year temperature (TT) power spectrum (Larson et al. [2010]). This was estimated by using Blackwell-Rao estimator on chains obtained from Gibbs sampling. It reaches scales as low as l=1060 (with the appropriate binning condition to have sufficiently high signal to noise), and can provide data for the third acoustic peak.

3.3.2 Hamiltonian Monte Carlo and GHS

The methods of Hamiltonian sampling and GHS have been introduced in section 2.2.3. Here we look at their application to CMB data as seen in Taylor et al. [2008] and Balan [2012]. In Taylor et al. [2008], the HMC is introduced and tested on different sized simulated data. It is found that for WMAP-scale simulations, the power spectrum can be estimated in 80 hours on a high-end desktop. The HMC is compared with the Gibbs sampler on low-resolution simulated data and is seen to perform better with less correlation between the samples at low signal to noise maps. HMC requires one order of magnitude lesser $a'_{lm}s$ than a Gibbs sampler without preconditioners and a factor of 3-4 fewer than a Gibbs sampler with carefully tuned preconditions (Eriksen et al. [2004]).

The correlation of the samples produced by the HMC depends strongly on the masses used, allowing for room to improve the technique with a better prescription for the masses. The strict condition of positivity of C_l leads to a high rejection rate and due to the susceptibility to numerical error, we sample in $\log C_l$ so that the negative values are easily rejected. This means that the uniform prior on $C'_l s$ means that there has to be an exponential prior on $\log C_l$.

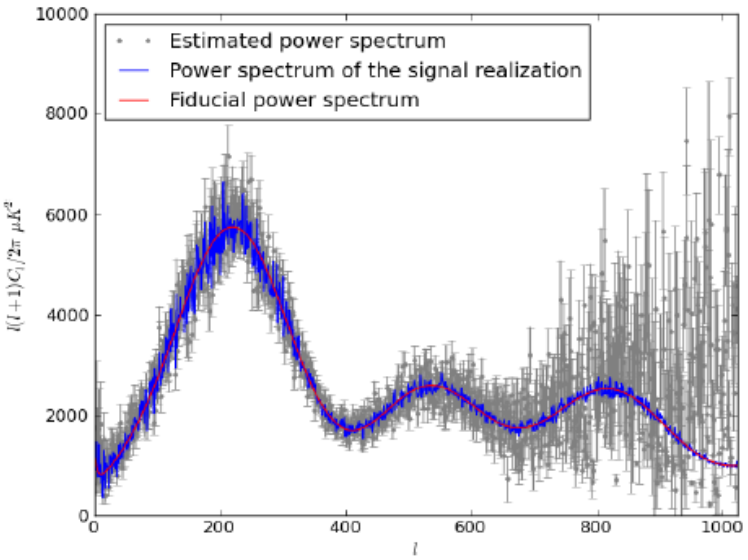

Figure 5: The figure shows the estimated power spectrum (grey) using GHS from Balan et al. [in prep]. The blue line represents the power spectrum of the signal realization and the red line is the fiducial power spectrum that is input into the sampler.

The performance of the GHS is analysed in Balan [2012] for datasets with different dimensionality of the posterior and it is seen that even for maps with 10^6 pixels, the sampler has an accpetance rate of 0.15.

4 CMB analysis with GHS

In our project, we have applied the GHS method to simulated CMB data. The maps were simulated from the theoretical power spectrum of WMAP7. A noise matrix and a beam size were input to generate the temperature data. We made different maps with varying beam sizes and noise values. We took the GHS code for estimating CMB power spectra and made a general interface to add maps for any dataset. We applied the GHS method to maps of different resolutions with different input values. We observed that the sampler converges with an acceptance rate of 0.77 in 2500 samples for $\sim$ 1000 dimensions in the posterior. For a posterior distribution of $\sim$ 4000 dimensions, there is no significant drop in the acceptance and the number of samples required for convergence is nearly the same. The sampler shows

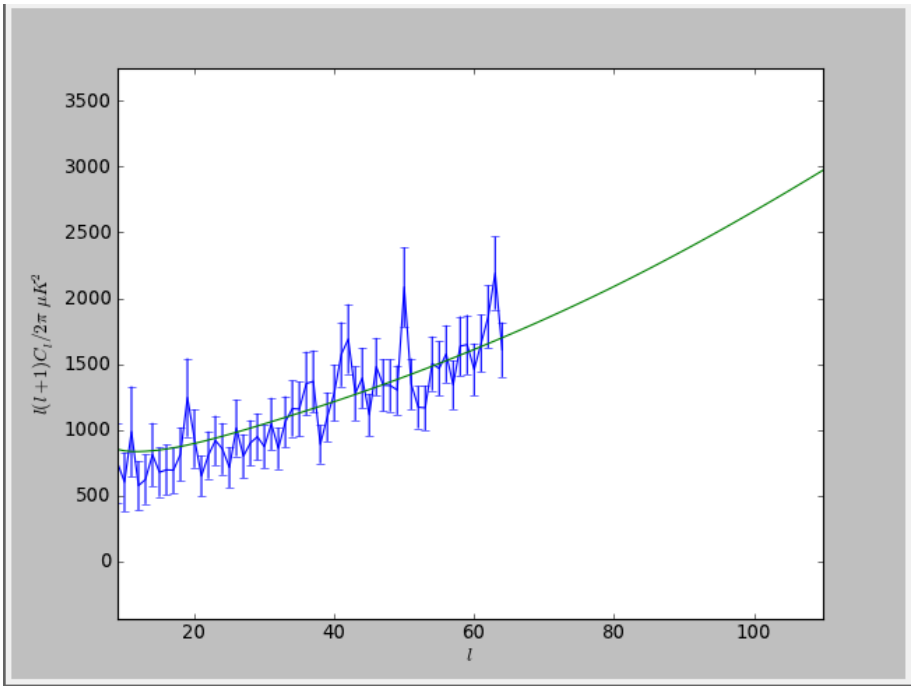

Figure 6: The estimated power spectrum (blue) with $N_{side} = 32$. The fiducial power spectrum from WMAP7 is overplotted (green). http://lambda.gsfc.nasa.gov

a drop in performance if the noise matrix input to the sampler is different from the noise value of the temperature map being input to the sampler code. The estimated power spectrum till l=64 is seen in the figure. We

Table 1: The input characteristics for estimating the power spectrum using GHS and the fiducial C_l's from WMAP7

Characteristics	Values
N_{side}	32
Beam size(radian)	0.04
rmsNoise (K)	$1 * 10^{-6}$
Input Power Spectrum	WMAP7

introduced masks to the data, to account for the data lost due to galactic foreground emission. We see that the sampler takes longer to converge, with a lower acceptance rate and a larger number of samples to converge.

5 Methods for LSS power spectrum estimation

Measurement of cosmological parameters from galaxy surveys breaks parameter degeneracies of CMB probes (Tegmark [1997]). In this section we look at the methods that have been introduced in section 3 for CMB power spectrum estimation as tools for working with LSS data. In the first part, we see the application of the pseudo-spectrum method to a set of galaxies from the Sloan Digital Sky Survey (SDSS). We also look at application of sampling methods to LSS data.

5.1 Pseudo C_l

In Thomas et al.[2010], the estimation of pseudo $C_l's$ has been applied to large scale structure data from the SDSS survey of luminous red galaxies (LRGs). Galaxy surveys offer unique insight into the establishment of a coherent cosmological model, with late-time galaxy distribution being sensitive to dark energy Riess et al. [1998]. Thomas et al. [2009] uses weak lensing data from the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) to constrain mDGP model of modified gravity which includes an extra-dimension and look at general modified gravity growth (γ) and power spectrum (Σ) parameters. Hu et al. discusses the possibility of surveys like SDSS detecting degenerate massive neutrinos m_ν >0.65 eV since eV mass neutrinos suppress power by a factor of 2.

The paper follows the spherical harmonic analysis of Peebles [1973], explicitly summing over the discrete galaxies in the incomplete sky. The formalism has been explained in section 3.2. The data is binned from z=0.45 to z=0.65 in bins of 0.05 for a total of 723,556 galaxies. Since spectroscopic follow-up of such a large number of galaxies isn't possible, these large survey use photometry from several different wavelength bands to determine the redshifts of these galaxies. The motivation for such measurements arises from the fact that the decrease in accuracy with respect to spectroscopic redshifts is outweighed by the high number of galaxies over a large area of the sky. This technique is shown to be competitive in Blake et al. (2007) and Padmanabhan et al. (2007). It forms the basis for current LSS experiments like Dark Energy Survey and future probes like LSST et al. [a] and EuCLId. Recent papers like Greisel et al. [2013] are generating spectral energy distributions for SDSS galaxies to deduce photometric redshifts from a Bayesian template fitting method.

It is observed in Thomas et al. [2010] with the Data Release 7 (DR7) of SDSS that there is an excess of power in the farthest redshift bin at the

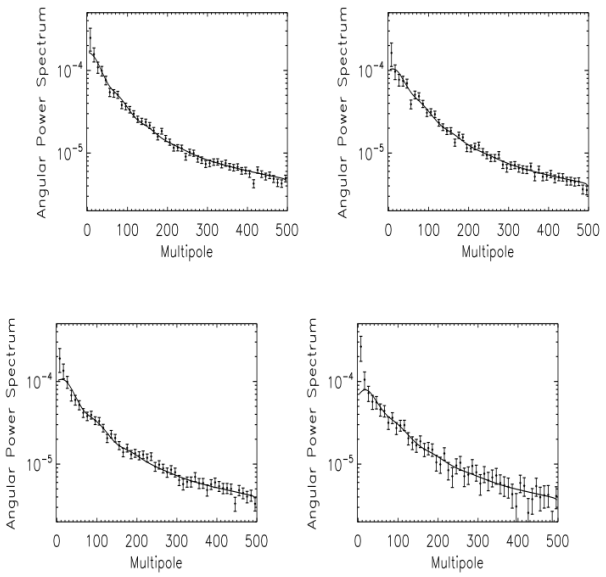

Figure 7: The power spectrum of the large scale structure of galaxies from the SDSS LRG survey as estimated in Thomas et al. [2010]. The four plots are for redshift bins dz=0.05 from z=0.45 to z=0.65. The error bars include contributions from cosmic variance and shot noise, while accounting for sky coverage. There is an excess of power in the lowest multipole at the farthest redshift.

largest angular scales. This can be seen in the figure. It is consistent with the observation in Blake et al. (2007) with SDSS DR4. In a companion paper Thomas et al. [2011] the cause of this high power is speculated. With analysis of survey data and statistics, they find that the cause isn't extinction, redshift estimation or M-star contamination. The authors posit that it could be a sign of new physics like a modified theory of gravity or exotic dark energy. It could also be a result of systematics due to the fact that highest redshift galaxies are most susceptible to error in measurement. A measurement of this anomaly with DES would be an interesting indicator of possible new cosmology.

5.2 Sampling Methods for LSS data

The application of sampling techniques to LSS data is similar to the analysis with CMB data. In this part, we look at papers on Gibbs Sampling and Hamiltonian Sampling applied to LSS data.

5.2.1 Gibbs Sampling

The technique of Gibbs sampling to sample the joint posterior of signal realization and the power spectrum is used to estimate the 3D power spectrum of the large scale structure. In Jasche et al. [2009], it is used for the first time to estimate LSS power spectrum and the dark matter density. The Bayesian framework is implemented in ARES (**A**lgorithm for**RE**construction and **S**ampling) under the assumption that the distribution is gaussian on large scales. This implementation allows for a calculation of any desired statistical summary, specially the joint uncertainty estimates.

The aim is to find $P(C_l|d)$ where C_l is the 3D power spectrum coefficient. Since this is a very difficult task, the method employs Gibbs sampling to sample the full joint posterior $Pr(a, C_l|d)$ where a is the signal realization. This joint posterior is then marginalized over the signal amplitude. The methodology followed is identical to Wandelt et al. [2004] with alternating iterations of the conditional distributions of the power spectrum and signal realization to evaluate the joint posterior.

5.2.2 Hamiltonian Sampling and GHS application

In Jasche et al. [2010], a fast HMC is used for LSS inference. They test the method with different scenarios, accounting for survey geometry and selection effects. Since the HMC is a sampled representative of the density posterior, any statistic can be estimated from the set of samples. In their inference, they use a lognormal Poissonian distribution and apply the Hamiltonian equations of motion accordingly, since the potential energy term would differ from its expression for a Gaussian distribution. The formalism is identical to the work done for CMB data analysis in Taylor et al. [2008]. Since the initial conditions of the sampler are not on the posterior surface, a burn-in phase is required till a point on the correct posterior surface is identified. As there is no theoretical criterion for demarcating the end of the burn-in phase, it is empirically tested. The figure shows the estimated power spectra.

In our approach, we wish to use GHS for LSS data upto $l \sim 200$ for power spectrum estimation. Since GHS uses a Gaussian approximation to 'guide' the sampler within the sample space and the LSS distribution is non-Gaussian on small scales due to clustering, we expect to see interesting results. In our investigation, we will look at modifying GHS with a Poisson distribution as an alternative approximation for the posterior.

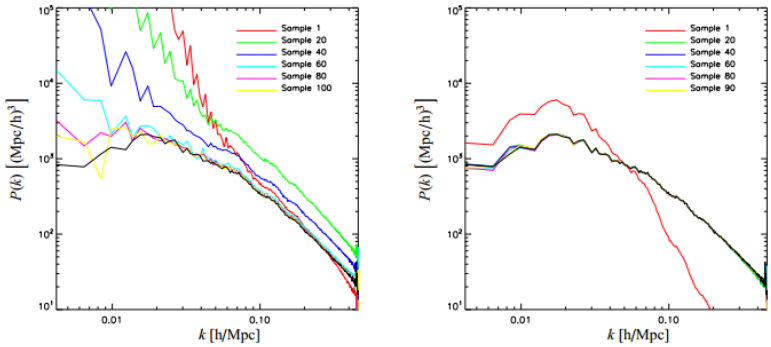

Figure 8: Successive power spectra measured from Hamiltonian samples during burn in. **Left***: The complete observational problem.* **Right***: A fiducial calculation showing the convergence of individual sample spectra towards the spectrum of the true matter field realization (black).*

6 Conclusion

In this review we have introduced the methods used in the literature for estimation of power spectra for both CMB and LSS datasets and looked at their efficacy at the scale of modern cosmological surveys. We provide a basic framework for Bayesian parameter estimation and introduce numerical methods for sampling the posterior probability distribution. The tests for convergence of these methods fall into two categories: inter-chain and intra-chain.

We looked at the application of different traditional and sampling methods to CMB datasets. We see that for smaller datasets like COBE, maximum likelihood methods were effective in estimating the power spectrum. We also looked at a pseudo-C_l method which presumes a fully-sky, noise free data and then relates the pseudo-$C_l's$ to the real $C_l's$. It has been used for power spectrum estimation for l >32 WMAP data. We then look at literature that has used sampling methods like Gibbs sampling, Hamiltonian sampling and GHS for CMB power spectrum estimation. We presented a short summary of results obtained from our own analysis of simulated CMB data and introduced methods that have been implemented for LSS power spectrum estimation and papers presenting results from SDSS data.

7 Future Work

We have applied GHS to simulated low resolution CMB maps. Results are presented in section 4. We have developed an interface to the sampler for any input dataset. In the coming months, we have the following objectives:

- Simulate large scale structure maps from theoretical galaxy power spectra
- Apply the GHS code to the simulated data using the interface setup
- Analyse the performance of the sampler for data with different input characteristics
- Compare GHS to conventional pseudo-C_l estimation

Acknowledgements

I would like to thank my supervisors Dr. Filipe Abdalla and Dr. Sreekumar Balan for their invaluable inputs to this literature review.

References

Sreekumar Thaithara Balan. Bayesian methods for astrophysical data analysis. 2012.

Julian Borrill. *arXiv:astro-ph/9712121v2*.

Anthony Challinor. CMB anisotropy science : a review. (288):1–11, 2012.

Mary Kathryn Cowles and Bradley P Carlin. Markov Chain Monte Carlo Convergence Diagnostics : A Comparative Review. 1996.

S. Dodelson. *Modern Cosmology*. Academic Press, 2003.

G Efstathiou. Myths and Truths Concerning Estimation of Power Spectra : The Case for a Hybrid Estimator. 000(February):1–27, 2008.

Abate et al. Large Synoptic Survey Telescope: Dark Energy Science Collaboration. *arXiv:1211.0310v1*, pages 1–133, a.

Balan et al. *arXiv:0805.3532*, 2008a.

Bennett et al. Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. pages 1–177, 1843.

Bennett et al. Scientific results from the Cosmic Background Explorer (COBE). *Proceedings of the National Academy of Sciences of the United States of America*, 90(11):4766–73, June 1993. ISSN 0027-8424.

Blake et al. Cosmological baryonic and matter densities from 600 , 000 SDSS Luminous Red Galaxies with photometric redshifts. 000(February), 2008b.

Bond et al. Estimating the power spectrum of the cosmic microwave background. *Physical Review D*, 57(4):2117–2137, February 1998a. ISSN 0556-2821. doi: 10.1103/PhysRevD.57.2117. URL `http://link.aps.org/doi/10.1103/PhysRevD.57.2117`.

Dunkley et al. Fast and reliable Markov chain Monte Carlo technique for cosmological parameter estimation. *Monthly Notices of the Royal Astronomical Society*, 356(3):925–936, January 2005a. ISSN 00358711.

Dunkley et al. Five-Year Wilkinson Microwave Anisotropy Probe Observations: Likelihoods and Parameters From the Wmap Data. *The Astrophysical Journal Supplement Series*, 180(2):306–329, February 2009a. ISSN 0067-0049.

Eriksen et al. Power Spectrum Estimation from HighResolution Maps by Gibbs Sampling. *The Astrophysical Journal Supplement Series*, 155(2): 227–241, December 2004. ISSN 0067-0049.

Feroz et al. MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society*, 398(4):1601–1614, October 2009b. ISSN 00358711.

Gorski et al. On Determining the Spectrum of Primordial Inhomogeneity from the COBE 1 DMR Sky Maps : II . Results of Two Year Data Analysis. pages 1–13, b.

Hivon et al. *The Astrophysical Journal, 567*, 1:2–17, 2002.

Hobson et al. *Bayesian Methods in Cosmology*. Cambridge University Press, 2010a.

Jasche et al. Bayesian power-spectrum inference for large-scale structure data. *Monthly Notices of the Royal Astronomical Society*, 406(1):60–85, July 2010b. ISSN 00358711.

Jewell et al. A Markov Chain Monte Carlo Algorithm for Analysis of Low Signal-To-Noise Cosmic Microwave Background Data. *The Astrophysical Journal*, 697(1):258–268, May 2009c. ISSN 0004-637X. doi: 10.1088/0004-637X/697/1/258.

Jones et al. A measurement of the angular power spectrum of the cmb temperature anisotropy from the 2003 flight of boomerang. 2005b.

Larson et al. Seven-Year Wilkinson Microwave Anisotropy Probe (Wmap) Observations: Power Spectra and Wmap -Derived Parameters. *The Astrophysical Journal Supplement Series*, 192(2):16, February 2011a. ISSN 0067-0049.

Lentati et al. *arXiv:1210.3578*, 2012.

Mandel et al. Type Ia Supernova Light-curve Inference: Hierarchical Bayesian Analysis In The Near-Infrared. *The Astrophysical Journal*, 704 (1):629–651, October 2009d. ISSN 0004-637X.

March et al. Improved constraints on cosmological parameters from Type Ia supernova data. *Monthly Notices of the Royal Astronomical Society*, 418(4):2308–2329, December 2011b. ISSN 00358711.

Metropolis et al. *Journal of Chemical Physics*, 21, 1087, 1953.

Mukherjee et al. A nested sampling algorithm for cosmological model selection. *arXiv:astro-ph/0508461v2*, 2006.

Padmanabhan et al. The Clustering of Luminous Red Galaxies in the Sloan Digital Sky Survey Imaging Data. (February), 2008c.

Riess et al. Observational evidence from supernovae of an accelerating Universe. 1998b.

Stompor et al. The MAXIMA Experiment : Latest Results and Consistency Tests. 2147(01):1–14, 2003.

Watson et al. *arXiv:1010.0911*, 2011c.

A. Gelman and D.B. Rubin. Inference from Iterative Simulation using multiple sequences. *Statistical Science , Vol. 7, No. 4, 457-511*, 1992.

N. et al. Greisel. *arXiv: 1303.3005v1*, 2013.

N.E. Groeneboom. *arXiv:0905.3823v1*, 2009.

Kenneth M Hanson. *Proceedings of SPIE*, 467, 456, 2001.

W. K. Hastings. *Biometrika*, 57, 97, 1970.

P.J.E. Hauser, M.G. Peebles. *ApJ, 185: 757-785, 1973.*

M.P Maisenger K. Hobson. Maximum-likelihood estimation of cosmic microwave background power spectrum from interferometer observations. *MNRAS 334, 569-588 (2002).*

Wayne Hu and Martin White. The cosmic symphony. *Scientific American*, 290(2):44–53, March 2004. ISSN 0036-8733.

Wayne Hu, Daniel J Eisenstein, and Max Tegmark. Weighing Neutrinos with Galaxy Surveys. *arXiv:astro-ph/9712057v2*, (1).

Jens Jasche and Francisco S. Kitaura. Fast Hamiltonian sampling for large-scale structure inference. *Monthly Notices of the Royal Astronomical Society*, 407(1):29–42, September 2010. ISSN 00358711.

E T Jaynes. Probability Theory : The Logic of Science. 1995.

R Neal. Slice sampling. *The Annals of Statistics*, 31(3):705–767, 2003.

Radford M Neal. An Improved Acceptance Procedure for the Hybrid Monte Carlo Algorithm arXiv : hep-lat / 9208011v2 20 Aug 1992. pages 1–16, 1992.

Radford M Neal. Suppressing Random Walks in Markov Chain. (9508): 1–22, 1995.

Siang Peng Oh, David N Spergel, and Gary Hinshaw. *The Astrophysical Journal.*

A.E. Raftery and S.M. Lewis. *Statist. Sci.*, 7(4):493–497, 1992.

D.S. Sivia. *Data Analysis: A Bayesian Tutorial.* Oxford Science, 1996.

J.S. Skilling. Bayesian Analysis. *1, 833.*

J. F. Taylor, M. a. J. Ashdown, and M. P. Hobson. Fast optimal CMB power spectrum estimation with Hamiltonian sampling. *Monthly Notices of the Royal Astronomical Society*, 389(3):1284–1292, September 2008. ISSN 00358711.

Jeremy F Taylor. The effect of foregrounds on observations of CMB polarization. (September), 2008.

M Tegmark. Angular power spectrum of the 4-year COBE data. *arXiv:astro-ph/9601077v3*, 1996.

Max Tegmark. Measuring Cosmological Parameters with Galaxy Surveys. *Physical Review Letters*, 79(20):3806–3809, November 1997. ISSN 0031-9007.

S. A. Thomas, Filipe B Abdalla, and Jochen Weller. Constraining Modified Gravity and Growth with Weak Lensing . *arXiv:0810.4863v1*, 000 (October), 2008.

Shaun A Thomas, Filipe B Abdalla, and Ofer Lahav. The Angular Power Spectra of Photometric SDSS LRGs. 000(November), 2010.

Surya T. Tokdar and Robert E. Kass. Importance sampling: a review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1):54–60, January 2010. ISSN 19395108.

Roberto Trotta. Bayes in the sky: Bayesian inference and model selection in cosmology. *arXiv:0803.4089v1*, (00):1–41, 2008.

Benjamin D Wandelt, David L Larson, and Arun Lakshminarayanan. arXiv : astro-ph / 0310080v2 5 Oct 2003. pages 1–12, 2008.

Benjamin D Wandelt, Frode K Hansen, and Anthony J Banday. The HEALPix Primer. CM(2), 2010.