A NON-PARAMETRIC ANALYSIS OF THE CMB POWER SPECTRUM

Christopher J. Miller and Robert C. Nichol

Department of Physics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213

Christopher Genovese and Larry Wasserman

Department of Statistics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213

Draft version October 25, 2007

ABSTRACT

We examine Cosmic Microwave Background (CMB) temperature power spectra from the BOOMERANG, MAXIMA, and DASI experiments. We non-parametrically estimate the true power spectrum with no model assumptions. This is a significant departure from previous research which used either cosmological models or some other parameterized form (e.g. parabolic fits). Our non-parametric estimate is practically indistinguishable from the best fit cosmological model, thus lending independent support to the underlying physics that governs these models. We also generate a confidence set for the non-parametric fit and extract confidence intervals for the numbers, locations, and heights of peaks and the successive peak-to-peak height ratios. At the 95%, 68%, and 40% confidence levels, we find functions that fit the data with one, two, and three peaks respectively ($0 \leq \ell \leq 1100$). Therefore, the current data prefer two peaks at the 1σ level. However, we also rule out a constant temperature function at the $> 8\sigma$ level. If we assume that there are three peaks in the data, we find their locations to be within ℓ_1 = (118,300), ℓ_2 = (377,650), and ℓ_3 = (597,900). We find the ratio of the first peak-height to the second $(\frac{\Delta T_1}{\Delta T_2})^2 = (1.06, 4.27)$ and the second to the third $(\frac{\Delta T_2}{\Delta T_3})^2 = (0.41, 2.5)$. All measurements are for 95% confidence. If the standard errors on the temperature measurements were reduced to a third of what they are currently, as we expect to be achieved by the MAP and Planck CMB experiments, we could eliminate two-peak models at the 95% confidence limit. The non-parametric methodology discussed in this paper has many astrophysical applications.

1. INTRODUCTION

There has been growing evidence for the existence of peaks and valleys in the temperature power spectrum of the CMB. From a theoretical standpoint, such features are a direct result of the physics in the primordial photon-electron plasma, predicted by gravitational instability models of structure formation (Peebles & Yu 1970; Hu and Sugiyama 1995). These features are important for constraining the cosmology of our Universe. For instance, in many models, the ratio of the height of the first peak to the second peak is dependent on the spectral tilt, n_s and the baryon fraction, $\Omega_{baryons}/\Omega_{matter}$. The ratio of the third peak to the second peak is dependent on $\Omega_{matter}h^2$ and n_s (see Hu et al. 2001 for further discussion).

Most often in the literature, the CMB power spectra are fit to a suite of cosmological models (Tegmark et al., 1999,2000; Jaffe et al. 2001). These physical models are well-motivated and sophisticated, but they contain many free parameters (*e.g.*, eleven in the work of Wang, Tegmark, and Zaldarriaga 2001– WTZ), some of which are unknown (ionization depth, contribution from gravity waves) or degenerate (e.g. see Efstathiou 2001). Typically, some sort of likelihood analysis is performed to determine which cosmological model best fits the data.

There is however, another approach: place constraints on the features of the power spectrum and use these features to determine the cosmological parameters. The assumptions here are that the peaks and valleys are best described by the broad range of cosmological models (as in Hu et al.) or by parabolas or some other chosen function (as in Knox & Page 2000 and de Bernardis et al. 2001). A potential problem in all of these approaches is that it is difficult to get valid statistical confidence intervals (see Section 2). There is also the concern that the fitted features may be artifacts from the multitude of assumptions.

In this paper, we take what may be considered a more conservative approach: we make no assumptions whatsoever about the true underlying function. Our new statistical technique is non-parametric and allows for valid confidence intervals to be measured for peak characteristics. One theme of our work is that confidence intervals for any quantity of interest can be extracted from a confidence set for the unknown spectrum. These techniques for fitting and inference are applicable to a wide variety of astrophysical data-analysis problems.

2. OVERVIEW OF NONPARAMETRIC ANALYSIS

In general, non-parametric statistical methods estimate functions without imposing a finite-dimensional parametric form. The resulting estimates are obtained by carefully smoothing the data to balance bias and variance. See Hastie and Tibshirani (1990) for details and examples.

The CMB data, after suitable preprocessing, take the form $(X_1, Y_1), \ldots, (X_n, Y_n)$ where X_i is the multipole moment (usually denoted ℓ) ordered according to increasing X, and Y_i is the estimated power spectrum at X_i (usually denoted C_ℓ with some constants). Let $f(X_i)$ be the true power spectrum at X_i. Then,

$$Y_i = f(X_i) + \epsilon_i \qquad (1)$$

where ϵ_i is the error in estimating $f(X_i)$. We require that

the ϵ_i are uncorrelated, zero-mean Gaussians. Therefore, we take the uncorrelated statistical errors as given by the experiments. There are additional, correlated noise terms in all of the experiments due to calibration, beam width and pointing uncertainties, which were not used in our analysis. The magnitude of these correlated errors is typically 5% - 20%, although the Boomerang effective beam uncertainties can be higher for $\ell > 600$. The method we use in this work can be generalized to include correlated errors (Genovese et al., in preparation). We assume that f^2 is integrable; otherwise, we make no further assumptions. Our non-parametric technique yields a vector $\widehat{\mathbf{f}}$ that estimates the vector $\mathbf{f} = (f(X_1), \ldots, f(X_n))$ of the true spectrum's values at the X_is (see Figure 1 top).

After we perform the fit, we need to quantify the uncertainty to make any inferences. We begin by constructing a set of vectors C_n from the data that traps the true power spectrum, $\mathbf{f}$, with probability $1 - \alpha$, where $1 - \alpha$ is a pre-specified confidence level. With C_n we can derive a confidence interval for any quantity of interest. Consider, for example, the number of peaks. For any vector $\mathbf{f} \in C_n$, define $P(\mathbf{f})$ to be the number of peaks in $\mathbf{f}$. Then, because C_n contains the true spectrum with confidence $1 - \alpha$, the range on the measured peaks is $(\min_{\mathbf{f} \in C_n} P(\mathbf{f}), \max_{\mathbf{f} \in C_n} P(\mathbf{f}))$. We can find confidence intervals for the locations and heights of peaks in a similar way. There are important advantages for these confidence measurements over "standard" χ^2 or even Bayesian techniques, which we discuss further in the next section.

3. METHODOLOGY

Refer to the model in equation (1). Let the functions $\phi_1, \phi_2, \ldots$ be an orthonormal basis over the range of the observed X_is. The choice of basis is somewhat important. For instance, if we were fitting to a galaxy spectrum, with highly peaked emission lines on top of a smooth, broad continuum, a wavelet basis would allow for the simultaneous fitting of broad and narrow features. On the other hand, a cosine basis would require large amplitude high frequency terms to match the emission lines. This might cause the continuum fit to be wiggly. In our work, we use a discrete cosine basis, $\phi_1(x) \equiv 1$, $\phi_2(x) = \sqrt{2}cos(\pi x)$, $\phi_3(x) = \sqrt{2}cos(2\pi x), \ldots$, since it has well determined properties for confidence limits. For simplicity in the derivation, we assume that the variance (second moment about the mean) of each ϵ_i is the same value σ^2. This is not necessary in practice, nor was it assumed in the full analysis.

Any square integrable function f can be expanded as $f(x) = \sum_{j=1}^{\infty} \beta_j \phi_j(x)$. Estimating f amounts to estimating the β_j's. If we have chosen a good basis for representing f, the higher-order terms in this series will tend to decay rapidly. Hence, we can approximate the infinite sum with finite sum $f(x) \approx \sum_{j=1}^{N} \beta_j \phi_j(x)$. So, to estimate the underlying function, we need to find the βs.

Let $Z_j = N^{-1/2} \sum_{i=1}^{N} Y_i \phi_j(X_i)$, for $j = 1, \ldots, N$. Based on the theory in Beran (2000), we take $N = n$. This choice ensures that the estimate of f is optimal and that the resulting confidence intervals remain valid. It can be shown that each statistic Z_j has approximately a Gaussian distribution with mean $\theta_j = \sqrt{n}\beta_j$ and variance σ^2. This re-parameterization means that we need to estimate $\theta = (\theta_1, \ldots, \theta_n)$. Given an estimate $\widehat{\theta}$ of θ, we can estimate $f(X_i)$ via $\widehat{f}(X_i) = (1/\sqrt{n}) \sum_{j=1}^{n} \widehat{\theta}_j \phi_j(X_i)$.

We could use Z_j as an estimate of θ_j, but this yields a poor estimate of f because it is too variable. A smoother estimate can be obtained by damping the higher frequency terms in the expansion. In statistics, this is called a "shrinkage estimator". We consider shrinkage estimators of the form $\widehat{\theta} = (\gamma_1 Z_1, \gamma_2 Z_2, \cdots, \gamma_n Z_n)$, where $1 \geq \gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_n \geq 0$ are called the shrinkage coefficients. The smaller γ_j, the smaller the contribution of ϕ_j in the estimating expansion for f. With the cosine basis, for example, such shrinkage estimators damp down the contribution from high-frequency terms.

Every choice of $\gamma = (\gamma_1, \ldots, \gamma_n)$ gives an estimate $\widehat{\theta}^\gamma$ which then yields an estimate $\widehat{f}^\gamma$ of the function f. We would like to choose γ to minimize the mean squared error, $\text{MSE}(\gamma) = \langle \int (f(x) - \widehat{f}^\gamma(x))^2 dx \rangle$. Unfortunately $\text{MSE}(\gamma)$ is unknown, because it depends on the true f, but it can be estimated by Stein's Unbiased Risk Estimator (Stein 1981):

$$\widehat{\text{MSE}}(\gamma) = \sum_j \left[\sigma^2 \gamma_j^2 + \left(Z_j^2 - \sigma^2 \right) (1 - \gamma_j)^2 \right]. \quad (2)$$

We use the Pooled Adjacent Values (PAV) algorithm (Robertson, Wright, and Dykstra, 1988) to minimize $\widehat{\text{MSE}}(\gamma)$ as a function of γ while maintaining the ordering constraint on γ. The minimizer is denoted $\widehat{\gamma}$ and the final estimate is therefore $\widehat{\theta} = (\widehat{\gamma}_1 Z_1, \widehat{\gamma}_2 Z_2, \cdots, \widehat{\gamma}_n Z_n)$.

Next we construct a confidence set, C_n for the vector of function values at the observed data, $\mathbf{f}_n = (f(X_1), f(X_2), \cdots, f(X_n))$. Throughout the paper we have said that the confidence sets are "valid." Formally, what this means is that for any $c > 0$,

$$\lim {}^1{}_{n \to \infty} \sup_{||\mathbf{f}_n|| \leq c} |\text{Pr}(\mathbf{f}_n \in C_n) - (1 - \alpha)| \to 0 \quad (3)$$

as $n \to \infty$, where the operation $||a||$ denotes $\sqrt{n^{-1} \sum_i a_i^2}$. This means that for large n, the confidence set traps the values of the true function with probability very close to $1 - \alpha$. The confidence set is an ellipse of the form

$$C_n = \left\{ \theta : \; n^{-1} \sum_j (\theta_j - \widehat{\theta}_j)^2 \leq \widehat{MSE}(\widehat{\gamma}) + n^{-1/2} \tau z_\alpha \right\} \quad (4)$$

where z_α is the number that has probability α to the right under a standard Gaussian. For instance, if we choose 95% confidence ($\alpha = 0.05$), then $z_\alpha = 1.645$, while for 67% confidence ($\alpha = 0.33$), the confidence "radius" is smaller with $z_\alpha = 0.44$. Here, τ is defined similar to Beran (2000):

$$\tau^2 = 4 \sigma^2 \sum_j (Z_j^2 - \sigma^2)(1 - \widehat{\gamma}_j)^2 + 2 \sigma^4 \sum_j (2\widehat{\gamma}_j - 1)^2. \quad (5)$$

---

[1]Strictly speaking, we should use a $\lim \sup$, which refers to the largest difference between the true coverage and the claimed coverage that can be obtained as the sample size gets large

We can in turn write the confidence set for $\mathbf{f}$ as

$$\mathcal{D}_n = \left\{ \mathbf{f} : \ \mathbf{f} = n^{-1/2} \sum_{j=1}^{n} \theta_j \phi_j \ \text{and} \ \theta \in \mathcal{C}_n \right\}. \quad (6)$$

Here, the true power spectrum $\mathbf{f} = (f(X_1), ..., f(X_n))$ is estimated at each of the original data points. Prior information about $\mathbf{f}$ of the form $\mathbf{f} \in \mathcal{P}_n$ allows us to replace $\mathcal{D}_n$ with $\mathcal{D}_n \cap \mathcal{P}_n$ while maintaining a $1 - \alpha$ confidence level. In particular, we take $\mathcal{P}_n$ to be the set of vectors $\mathbf{f}$ corresponding to spectra with zero to three peaks for $\ell \leq 1100$. From the confidence set $\mathcal{D}_n \cap \mathcal{P}_n$, we can derive confidence sets for any interesting feature of $\mathbf{f}$.

A key point is that the resulting intervals on any measured quantity are *simultaneously valid*, meaning that all of the intervals contain the corresponding true quantity with probability $1 - \alpha$. In contrast, deriving $1 - \alpha$ confidence intervals from a collection of individual chi-squares does not obtain $1-\alpha$ simultaneous coverage, but often substantially lower coverage. For example, a common technique to determine the 95% confidence range for a specific parameter is to "marginalize" over the other parameters (see Tegmark et al. 1999,2000). Such a technique will provide full coverage for that one parameter. However, when parameter ranges are combined, the confidence is lower than 95%. Bayesian intervals derived from a posterior distribution suffer from a similar problem in that the long-run frequency that the interval contains the true quantity may be much less than $1 - \alpha$.

4. RESULTS

Figure 1 shows the combined data from BOOMERANG, MAXIMA, and DASI (Halverson et al. 2001; Lee et al. 2001; and Netterfield et al. 2001). The bottom panel compares our non-parametric fit to the the fit of WTZ who use more experiments than the three examined here. Recall, our fit requires no assumptions about the data or the underlying cosmology. The WTZ fit, on the other hand, requires an 11 dimensional parameter space, with numerous prior assumptions placed on those parameters (for example, the Hubble constant is constrained to 72 ± 8 km s^{-1}Mpc^{-1}). The agreement between the two fits is very good, considering the difference in methodology.

The power of this technique lies in the ability to make quantitative statements about the true function with some specified confidence. First, we checked the 95% confidence set and find that every function within this set has at least one peak. Specifically, we set $\alpha = 0.05$ and determined the "radius" of this confidence ellipse (ie. the right side of the inequality in Eq 4). We then searched all possible functions with zero to three peaks over the specified range in ℓ to see if the condition in Eq 4 was met. Our definition of one peak requires the sorted data (according to increasing ℓ) to have a section with increasing temperature followed by a section with decreasing temperature. For $\alpha = 0.05$, no zero-peaked functions met this condition. Our set of zero-peaked functions includes those with constant temperature as well as those with either increasing temperature or decreasing temperature, but not both. We rule out functional forms with zero peaks at the level 95% confidence level. We also compared specifically against the best fit constant function for the power spectrum. We rule out a flat temperature spectrum (based on the weighted average of the data) at $> 8\sigma$ confidence. We perform the same analysis at the 68% confidence level (e.g. $\alpha = 0.32$ in Eq 4). The 68% confidence set rules out all single-peak functions. So at the one sigma level, we have found at least two peaks in the data. Only in the 40% ($\alpha = 0.6$) confidence set can we rule out two-peak functions. From our confidence sets, the data supports two peaks out of three peaks in the CMB power spectrum (for $0 \leq \ell \leq 1100$).

We calculate ranges for the peak heights and locations in Table 1. Figure 2 shows the non-parametric fit with confidence intervals for peak heights and locations at 95% confidence. Finally, we computed confidence intervals for the ratios of successive peaks under a three-peak model. The 95% confidence interval for the ratio of the height of the first peak to the height of the second peak is $(1.06, 4.27)$. The 95% confidence interval for the ratio of the height of the second peak to the height of the third peak is $(0.41, 2.5)$. This rules out equal heights for the first two peaks at the 95% level. These results are consistent with Hu et al. (2001) who find much stronger constraints on the height-height ratios by fitting to cosmological models.

5. DISCUSSION AND CONCLUSIONS

We present an application of a new and powerful non-parametric technique to CMB temperature data. Past approaches were based on complicated cosmological models or parameterized forms. There is superb visual agreement between the non-parametric fit and the best fitting cosmological model. Quantitatively, we provide constraints on the peak locations, heights, and height ratios of the power spectrum. These constraints can be used to place corresponding limits on the cosmological parameters that they describe. For instance, Hu et al. (2001) derive relationships which could in principle, be used for this purpose.

At the 2σ confidence level, we find at least one peak in the current CMB power spectra data, while at the 1σ level, we find two or more peaks. Only for a very low confidence, 40%, can we rule out two peak functions. Therefore, the data do not yet show the three expected peaks for $\ell \leq 1100$ (in the three CMB datasets examined here). There are two explanations for this: the model is right, but there is insufficient precision in the current data, or the model is wrong. If in fact the errors on the current measurements are simply too large, then these standard errors would have to be reduced to one-third of their current values to rule out a two-peak spectrum at the 95% confidence level. This suggests a range for the maximum required errors for future CMB experiments (via MAP and Planck) to "discover" three peaks in the CMB spectrum.

We point out that the lack of assumptions used to arrive at our best fit is conservative. On the other hand, results from fitting assumed cosmological models are optimistic, since those models all have a multi-peaked spectra (e.g. Hu et al. 2001). While the physical underpinnings for cosmological models are well founded, the last 50 years (or even five) have seen radical changes in those models which best fit the data. Therefore, a method to describe the CMB that is "cosmology free" has scientific value. Finally, we note that the methods described here can be applied to the many astrophysical problems that are not well suited for standard parametric techniques.

During the refereeing process of our paper, two related papers came to our attention (Durrer, Novosyadlyj & Apunevych 2001; Douspis & Ferreira 2001). These papers perform a model–independent measurement of the CMB power spectrum but they are not non-parametric estimates of the CMB acoustic peaks, as discussed herein, since they use phenomenological models to describe the underlying power spectrum. It is interesting to note however, that all three analyses find low statistical significances for the detection of the second and third peaks. We await higher precision measurements of the CMB power spectrum to secure the detection, location, shape and amplitude of these peaks.

REFERENCES

de Bernardis, P., et al. 2001, ApJsubmitted, astro-ph/0105296
Beran, R. 2000, J. Amer. Stat. Assoc. 95, 155
Douspis, M. and Ferreira, P. G., 2001, see astro–ph/0111400
Durrer, R., Novosyadlyj, B., Apunevych, S., 2001, ApJ, submitted, see astro-ph/0111594
Efstathiou, G. 2001, submitted to MNRAS, astro-ph/0109151
Halverson, N.W., et al. 2001, ApJsubmitted, astro-ph/0104489
Hastie, T. J. and Tibshirani, R. J. 1990, *Generalized Additive Models*, Chapman and Hall, London
Hu, W. & Sugiyama, N. 1996, ApJ, 471, 542
Hu, W., Fukugita, M., Zaldarriaga, M., & Tegmark, M. 2001, ApJ, 549, 669

Jaffe. A., et al. 2001, astro-ph/0007333
Knox, L. and Page, L. 2000, Phys. Rev. L., 85, 1366
Lee et al. 2001, astro-ph/0104459
Netterfield et al. 2001, ApJsubmitted, astro-ph/0104460
Peebles, P. J. E. & Yu, J. T. 1970, ApJ, 162, 815
Robertson, T., Wright, F.T., and Dykstra, R.L. 1988, *Order Restricted Statistical Inference*, Wiley, New York
Stein, C. 1981, Ann. of Statist., 9, 1135
Tegmark, M. 1999, ApJ, 514, L69
Tegmark, M. & Zaldarriaga, M. 2000, ApJ, 544, 30
Wang, X., Tegmark, M. and Zaldarriaga, M. 2001, Phys. Rev. D. submitted, astro-ph/0105091 – WTZ

TABLE 1
THREE PEAK CONFIDENCE INTERVALS

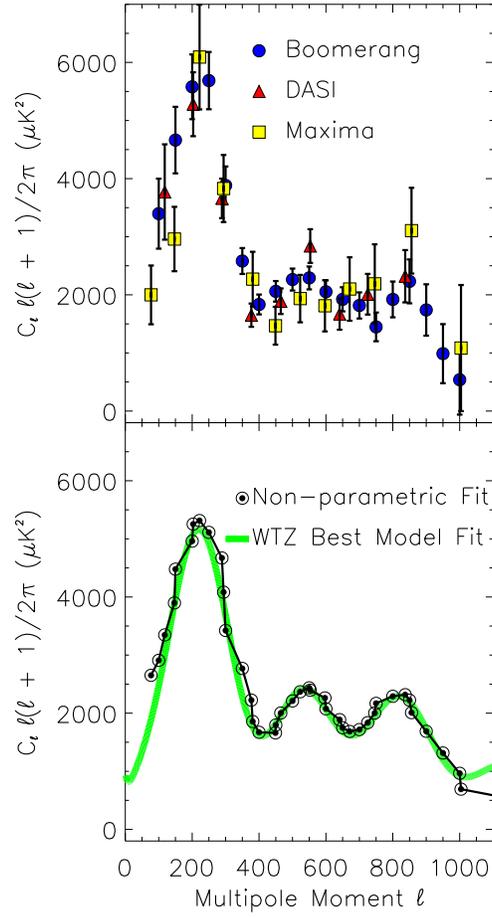| Peak	Location	Height
1	(118,300)	(4361,8055)
2	(377,650)	(1829,4798)
3	(597,900)	(1829,4688)

FIG. 1.— The **top** panel shows the raw CMB data from the Boomerang, MAXIMA, and DASI experiments. The **middle** panel is the raw CMB data over–plotted with the best fit using the non-parametric technique. The **bottom** panel shows the non-parametric fit against the best cosmological model fit from Wang, Tegmark, and Zaldarriaga (2001).
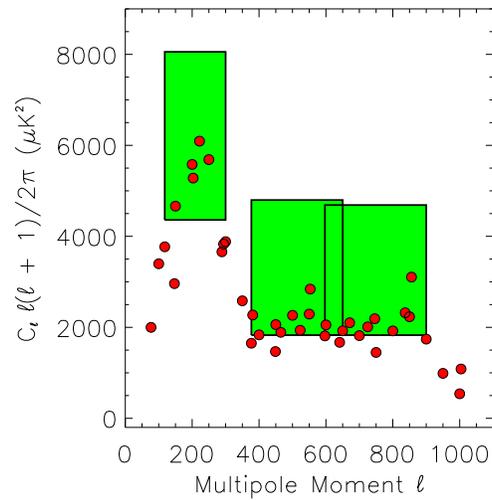

FIG. 2.— The ranges on the locations and peak heights of the non-parametric three peak (and two dip) fit function (95% confidence).