Nonparametric Inference for the Cosmic Microwave Background

Christopher R. Genovese,[1] Christopher J. Miller,[2] Robert C. Nichol,[3] Mihir Arjunwadkar,[4] and Larry Wasserman[5]

Carnegie Mellon University

The Cosmic Microwave Background (CMB), which permeates the entire Universe, is the radiation left over from just 380,000 years after the Big Bang. On very large scales, the CMB radiation field is smooth and isotropic, but the existence of structure in the Universe – stars, galaxies, clusters of galaxies, ... – suggests that the field should fluctuate on smaller scales. Recent observations, from the Cosmic Microwave Background Explorer to the Wilkinson Microwave Anisotropy Project, have strikingly confirmed this prediction.

CMB fluctuations provide clues to the Universe's structure and composition shortly after the Big Bang that are critical for testing cosmological models. For example, CMB data can be used to determine what portion of the Universe is composed of ordinary matter versus the mysterious dark matter and dark energy. To this end, cosmologists usually summarize the fluctuations by the

[1]Research supported by NSF Grants SES 9866147 and NSF-ACI-0121671.
[2]Research supported by NSF-ACI-0121671.
[3]Research supported by NSF Grant DMS-0101360.
[4]Research supported by NSF Grant DMS-0101360.
[5]Research supported by NIH Grant R01-CA54852-07 and NSF Grants DMS-98-03433, DMS-0104016, and NSF-ACI-0121671.

power spectrum, which gives the variance as a function of angular frequency. The spectrum's shape, and in particular the location and height of its peaks, relates directly to the parameters in the cosmological models. Thus, a critical statistical question is how accurately can these peaks be estimated.

We use recently developed techniques to construct a nonparametric confidence set for the unknown CMB spectrum. Our estimated spectrum, based on minimal assumptions, closely matches the model-based estimates used by cosmologists, but we can make a wide range of additional inferences. We apply these techniques to test various models and to extract confidence intervals on cosmological parameters of interest. Our analysis shows that, even without parametric assumptions, the first peak is resolved accurately with current data but that the second and third peaks are not.

Key words and phrases: Confidence sets, nonparametric regression, cosmology.

1 Introduction

The "Big Bang" model is misnamed, as one might expect when a term is coined as an insult. Cosmologist Fred Hoyle first used the name in a BBC radio interview to denigrate the theory, which opposed the then-dominant Steady State model. The name Big Bang stuck, as did its evocation of a mighty explosion in space. But the image of an *explosion* is highly misleading. What the model actually posits is that the Universe began hot, dense, and expanding.

Within the first second, roughly 13.7 billion years ago, the Universe achieved temperatures on the order of 1 trillion Kelvin (K, degrees above absolute zero; Schwarz, 2003). The density during that second was high enough to stop neutrinos, which interact so weakly with matter that they can pass unmolested through a quadrillion kilometers of lead. What ties this hot, dense beginning to the Universe we see today is expansion. A useful metaphor for the expanding universe is the surface of an inflating balloon. As the balloon inflates, space-time itself is stretched; every point moves away from every other point. Density falls as the universe expands. If you picture a wave oscillating over the surface of the balloon, the wavelength increases. Increasing the wavelength of light corresponds to reducing its temperature. The Universe thus cools as it expands.

Within the first three minutes, the Universe's temperature was over one billion K. The energy density in space was so high that atoms could not form.

Space was filled with a stew of photons, baryons (e.g., protons and neutrons), electrons, neutrinos, and other matter. As the temperature cooled below 1 billion K, light-element nuclei (deuterium, helium, some lithium) formed as well, in proportions that fit well with observations. During this period, photons (radiation) were the dominant form of energy in the Universe. Any fluctuations in density caused by gravity (which affects light and matter) were quickly smoothed out and so could not grow.

When the temperature of the primordial photons had fallen below approximately 12,000 K, photons were no longer dominating the interactions among all particles. Photons and baryons became coupled in a mathematically perfect fluid, while exotic kinds of matter began to clump under the influence of gravity. The interaction between this photon-baryon fluid and such gravitational overdensities are of critical importance and will be described below.

When the temperature reached about 3000 K, roughly 380,000 years after the Big Bang, electrons and protons could combine to form atoms. This decoupled the photon-baryon fluid, and the photons flew free through space. This period is named *recombination* and happened, in cosmic terms, very quickly. After another 200 million years, hydrogen formed after recombination had clumped enough for the first stars to form, which began the synthesis of heavy elements and the formation of galaxies that we see today.

Most of the photons released at recombination have travelled through space for billions of years without interacting with matter. The temperature

of these primordial photons has now cooled to about 2.7K, barely above absolute zero, which puts them in the microwave part of the electromagnetic spectrum. This primordial radiation field, which still pervades the Universe, is called the Cosmic Microwave Background (CMB). The CMB thus provides a snapshot of the moment of recombination, and fluctuations in the temperature across the sky contain information about the physics of the early universe.

1.1 The Cosmic Microwave Background Radiation

As we will explain in the remainder of this section, the temperature fluctuations in the CMB give a snapshot of the physics in the early Universe and provide critical tests of cosmological models. In 1992, the Cosmic Microwave Background Explorer (COBE) satellite discovered fluctuations in the blackbody temperature of the CMB (Smoot et al. 1992). These fluctuations are small: approximately one thousandth of the mean temperature over the sky. Indeed, almost thirty years of experiments since the CMB's discovery could not detect any deviation from uniformity. During the ten years following COBE, many more refined measurements were taken; notable experiments include MAXIMA, DASI, BOOMERANG (Lee et al. 2002, Halverson et al. 2002, Netterfield et al. 2002). In 2003, the Wilkinson Microwave Anisotropy Project (WMAP) considerably refined the picture, increasing spatial resolution by a factor of 33 and sensitivity by a factor of 45 over COBE (Bennett et al. 2003). In Figure 1, we compare the COBE and WMAP temperature sky

maps after removing the mean temperature $T = 2.726$ Kelvin and adjusting for the motion of our galaxy through the Universe. The fluctuations' magnitudes are just right to explain the large-scale structure in the Universe we see today. For example, if they had been much smaller, there would not be enough local concentration of mass to seed the formation of galaxies, galaxy clusters, et cetera.

Perhaps the most important summary of the temperature measurements used by cosmologists is the power spectrum, which gives the temperature variance as a function of spatial frequency. The spectrum's shape, and in particular the location and height of its peaks, relates directly to the parameters in cosmological models. (See Appendix 2 for a description of these parameters.) Thus, a critical statistical question is how accurately can these peaks be estimated. Of particular interest are the height and location of the first peak and the relative heights of the successive peaks.

Figure 2 displays an estimated spectrum commonly used by cosmologists and highlights the peaks of interest. We will give a more precise definition of the spectrum in Section 1.2, but here we want to explain how the spectrum's shape relates to the physics in the time up to recombination.

A key to understanding the physics before recombination is, as mentioned earlier, that photons and baryons became coupled into a (perfect) fluid. Mathematical techniques for studying fluid dynamics apply well in this scenario and have been investigated by many authors (see for instance Hu and Sugiyama 1995; Hu 1999, 2001, 2003; Hu and Dodelson 2002). The properties

of the fluid are determined by the relative density of photons and baryons in the fluid. Photons provided pressure, and the baryons provided inertia. As the fluid falls into a gravitational potential well around a clump of higher density, the pressure from the photons resists compression and the inertia of the baryons increases it. (Large, isolated potential wells were likely rare in the early universe; instead, there were random density fluctuations at many scales.) The result is an oscillation that produces pressure waves – sound – in the photon-baryon fluid. These *acoustic oscillations* account for much of the interesting structure in the spectrum, particularly the size and arrangement of peaks. The imprint of those waves remains in the CMB as a pattern of hot and cold spots.

To understand the peaks in the power spectrum, it is helpful to decompose the acoustic oscillations into their basic components, or modes. The first peak of the spectrum represents the fundamental tone of the oscillations, and the other peaks in the spectrum represent harmonics of this tone. The fundamental corresponds to the mode for which one compression occurs between the Big Bang and recombination. Each successive harmonic corresponds to an additional half-cycle, compressions followed by rarefication (decompression). Thus, the second peak represents modes that had time to compress and then rarefy before the photons were released from the photon-baryon fluid. The third peak represents compression-rarefication-compression, and so on.

The height of the first peak is determined by the total energy density.

Roughly, with more matter, the gravitational attraction requires more force to counteract, deepening the compression and thus increasing the amplitude of oscillation.

Now suppose we increase the density of baryons in the photon-baryon fluid. This increases the inertia of the fluid, deepening each compression phase without changing the rarefication. The oscillations become asymmetric. What this means is the odd-numbered peaks, whose modes end on a compression, are enhanced relative to the even-numbered peaks, whose modes end on a rarefication. Thus as the baryon fraction increases we should (over some range) see a differential effect on the odd and even numbered peaks.

The third peak in the spectrum provides the clearest support for the existence of "dark matter" – a substance of unknown composition that interacts at most weakly with baryons (e.g., neutrons, protons) or with photons (that's why its dark). To see why, it is illuminating to compare the oscillations in two example cases. In the "radiation-dominated era," when photons were the dominant form of interaction in the universe, density fluctuations were short-lived and unstable. A compressed region of photon-baryon fluid would rarefy as described earlier, but as it did so, the overdensity that caused the original gravitational well would disappear. Thus, in this case, at most one cycle of oscillation would occur between the Big Bang and recombination. We would see only one small peak in the temperature power spectrum corresponding to the mode (component of oscillation) that reaches maximum compression at the time of recombination. When the matter fraction is low, the peak would

be small, increasing with the baryon fraction (inertia).

In the "matter-dominated era," however, most of the energy density was in the form of dark matter. The rarefication phase of the oscillation would not eliminate the local overdensity, allowing multiple cycles of oscillation. The result is a spectrum with multiple harmonics and thus multiple peaks. The existence and contribution of dark matter is only distinguishable from that of baryons alone with three or more peaks. Moreover, the magnitude of the third peak constrains the time of transition between a radiation and matter dominated universe. In particular, a finding that the second and third peaks were comparable in magnitude would suggest that dark matter dominated before recombination, which is a fundamental prediction of Big Bang cosmology. The magnitude of the third peak is also of interest for estimating the fraction of dark matter in the Universe. Astronomers have several methods for inferring the dark matter fraction (e.g., studying the rotation of galactic disks in the recent Universe), and it is vital to determine if these estimates are comparable to those produced by the physics of the early Universe.

Finally, the pattern of CMB hot and cold spots we see on the sky corresponds to those photons just reaching us from the moment of recombination. (Recombination was relatively quick but not instantaneous, so there is some blurring of high spatial frequencies from the scatter of photons during that finite period.) The contribution to this pattern from each acoustic mode maps to a spherical mode of fluctuations on the sky. The analysis then proceeds by decomposing the observed fluctuations into spherical modes and using

the contributions of these modes to understand the acoustic oscillations. We discuss this in the next subsection.

1.2 The CMB Temperature Power Spectrum

Our focus in this paper is inference about the CMB temperature power spectrum and in particular the peaks in the spectrum. In this section, we describe the spectrum and some of the issues that arise in estimating it. Marinucci (2004, this issue) gives a more complete derivation upon which ours is based.

Let $T(\theta, \vartheta)$ denote the temperature field as a function of colatitude (zero at the zenith) $0 \leq \theta \leq \pi$ and longitude $0 \leq \vartheta < 2\pi$. Let $\overline{T}$ denote the average of the temperature field over the sphere.

Define the temperature fluctuation field by

$$Z(\theta, \vartheta) = \frac{T(\theta, \vartheta) - \overline{T}}{\overline{T}}.$$

Note that Z is a random field with mean zero and is assumed to have finite second moment. We can expand Z in terms of a orthonormal basis on the sphere. The usual choice of basis is the set of spherical harmonics $\{Y_{\ell,m}(\theta, \vartheta)\}$, for positive integers $\ell = 1, 2, \ldots$ and integers $-\ell \leq m \leq \ell$. (Here ℓ is called the multipole index, or loosely "multipole moment.") These are defined as follows:

$$Y_{\ell,m}(\theta, \vartheta) = \sqrt{\left(\frac{2\ell+1}{4\pi}\right) \frac{(\ell - |m|)!}{(\ell + |m|)!}}\, P_\ell^{|m|}(\cos\theta)\, e^{im\vartheta},$$

where the $P_{\ell,m}$ $\ell = 1, 2, \ldots$ and $m = 0, \ldots, \ell$ are the associated Legendre

functions defined by

$$P_\ell^m(x) = (-1)^m(1-x^2)^{m/2}\frac{d^m}{dx^m}P_\ell(x)$$

with Legendre polynomials

$$P_\ell(x) = \frac{1}{2^\ell \ell!}\frac{d^\ell}{dx^\ell}(x^2-1)^\ell.$$

We can now write

$$Z(\theta,\vartheta) = \sum_{\ell=1}^{\infty}\sum_{m=-\ell}^{\ell} a_{\ell,m}Y_{\ell,m}(\theta,\vartheta), \tag{1}$$

where,

$$a_{\ell,m} = \int_0^{2\pi}\int_0^{\pi} Z(\theta,\vartheta)Y_{\ell,m}(\theta,\vartheta)\sin\theta d\theta d\vartheta. \tag{2}$$

Since Z is a mean zero random field, the coefficients $a_{\ell,m}$ are random variables. They have mean 0, variance

$$C_\ell \equiv \mathsf{E}\,|a_{\ell,m}|^2,$$

and are uncorrelated. The *power spectrum* is defined to be C_ℓ as a function of ℓ.

Usually, it is assumed that Z is a Gaussian field (but see the paper by Marinucci in this issue) which implies that the $a_{\ell,m}$ have a Gaussian distribution. If we were to observe Z without measurement error, we could estimate C_ℓ by, say,

$$\widetilde{C}_\ell = \frac{1}{2\ell+1}\sum_{m=-\ell}^{\ell} a_{\ell,m}^2, \tag{3}$$

and thus for large ℓ we have $\widetilde{C}_\ell \approx C_\ell$ because we are averaging a large number of $a_{\ell,m}^2$. We call $\widetilde{C}_\ell$ the realized spectrum. Another important implication of

equation (3) is that even with perfect observations, we would not know the true power spectrum. Because our Universe is viewed as one realization of a stochastic process, $\widetilde{C}_\ell$ will in general differ from C_ℓ, especially for small ℓ. This is known as the problem of *cosmic variance*. We return to this point in Section 5.

In practice, the data are subject to various sources of measurement error, blurring, and unobserved parts of the sky. For example,the Milky Way, which is relatively bright, obscures the deep sky along a wide band. The spherical harmonics are no longer orthogonal over what is left of the sphere, which induces correlation and bias into the estimated C_ℓs. There are in addition a host of other complications in measuring Z.

Our model, in vector form, is

$$\widehat{C} = C + \epsilon, \tag{4}$$

where $\widehat{C}$ is the *observed spectrum* and the noise vector ϵ, with covariance matrix $C\aleph C^T$, incorporates the known sources of error, including measurement error. If there were no sky cut for the galaxy, $\aleph$ would be diagonal, but in practice, it incorporates the various known sources of error. In practice, the unknown C in the covariance matrix is replaced by a a pilot estimate, C^0. The choice of C^0 turns out to have surprisingly little effect on the results. We thus take the covariance matrix of ϵ in equation (4) to be known and equal to $\Sigma = C^0\aleph(C^0)^T$.

Another issue is that the observations are actually derived from a convo-

lution of the $\widetilde{C}_\ell$s with ℓ-dependent window functions; that is, the model is actually $\widehat{C} = KC + \epsilon$ for some matrix K. But as Figure 3 shows, the rows of K are very nearly delta functions. (See Knox, 1999.) And in fact, incorporating these window functions has negligible effect on our results, so we disregard them in what follows.

2 Uniform Confidence Sets For Nonparametric Regression

Taking $Y_\ell = \widehat{C}_\ell$ and $x_\ell = \ell/L_{\max}$, let $f(x_\ell) \equiv C_\ell$ denote the true power spectrum at multipole index ℓ. See Figure 4 for the Y_ℓ from the WMAP data (Hinshaw et al. 2003). We can then rewrite equation (3) in the form of a nonparametric regression problem:

$$Y_\ell = f(x_\ell) + \epsilon_\ell, \qquad \ell = L_{\min}, \ldots, L_{\max}, \tag{5}$$

where $\epsilon = (\epsilon_{L_{\min}}, \ldots, \epsilon_{L_{\max}})$ is assumed Gaussian with known covariance matrix Σ as described earlier. This is only an approximation to the model actually used, but we will not discuss the various practical complications here.

Let σ_ℓ^2 denote the diagonal elements of Σ and $n = L_{\max} - L_{\min} + 1$ be the total number of observed multipoles. Henceforth, we will use $i = \ell - L_{\min} + 1$ as an index.

Our approach is to nonparametrically estimate the regression f and find a nonparametric $1 - \alpha$ confidence ball $\mathcal{B}_n$ for f. More precisely, we want $\mathcal{B}_n$

$$\liminf_{n\to\infty} \inf_{f\in\mathcal{F}} P(f \in \mathcal{B}_n) \geq 1 - \alpha \tag{6}$$

for some large function class $\mathcal{F}$ such as a Sobolev space.

Once we have computed the confidence ball, we can construct a confidence interval for any functional $T(f)$ of interest, such as the location of the first peak. If $\mathcal{T}$ is a set of such functionals and

$$I_n(T) = \left(\min_{f \in \mathcal{B}_n} T(f),\ \max_{f \in \mathcal{B}_n} T(f) \right)$$

then we have that

$$\liminf_{n \to \infty} \inf_{f \in \mathcal{F}} P\left(T(f) \in I_n(T) \text{ for all } T \in \mathcal{T}\right) \geq 1 - \alpha. \tag{7}$$

Alternatively, we can construct the set of cosmological parameters that produces spectra within the confidence ball, which gives a joint confidence set on these parameters.

We use orthogonal series regression to estimate f and then construct a confidence ball via the Beran-Dümbgen pivot method (Beran 2000, and Beran and Dümbgen 1998), which was inspired by an idea in Stein (1981). Specifically, we expand f in the cosine basis $f = \sum_{j=0}^{\infty} \mu_j \phi_j$, where $\phi_0(x) = 1$ and $\phi_j(x) = \sqrt{2}\phi(\pi j x)$ for $j \geq 1$. If f is fairly smooth, for example if f lies in a Sobolev space, then $\sum_{j>n} \mu_j^2$ is negligible and we can write $f(x) \approx \sum_{j=0}^{n} \mu_j \phi_j(x)$. Let

$$Z_j = \frac{1}{n} \sum_{i=1}^{n} Y_i \phi_i(x_i) \tag{8}$$

for $0 \leq j < n$. Note that vector Z is approximately Normal with mean μ and variance matrix $U \Sigma U^T / \sqrt{n}$, where U is the cosine basis transformation matrix. We define the monotone shrinkage estimator by

$$\widehat{\mu}_j = \lambda_j Z_j \tag{9}$$

where $1 \geq \lambda_1 \geq \cdots \geq \lambda_n \geq 0$ are shrinkage coefficents. The estimate of f is

$$\widehat{f}(x) = \sum_{j=1}^{n} \widehat{\mu}_j \phi_j(x).$$

In this paper, we will use a special case of monotone shrinkage, called nested subset selection (NSS), in which $\lambda_j = 1$ for $j \leq J$ and $\lambda_j = 0$ for $j > J$. In this case,

$$\widehat{f}(x) = \sum_{j=1}^{J} Z_j \phi_j(x).$$

The squared error loss as a function of $\widehat{\lambda} = (\widehat{\lambda}_1, \ldots, \widehat{\lambda}_n)$ is

$$L_n(\widehat{\lambda}) = \int (\widehat{f}(x) - f(x))^2 dx \approx \sum_j (\mu_j - \widehat{\mu}_j)^2.$$

The risk is

$$R(\lambda) = \mathbb{E} \int (f(x) - \widehat{f}(x))^2 dx \approx \sum_{j=1}^{n} \lambda_j^2 \frac{\sigma_j^2}{n} + \sum_{j=1}^{n} (1 - \lambda_j)^2 \mu_j^2$$

where $\sigma_j^2 = \mathbb{V}(\epsilon_j)$. The shrinkage parameter λ is chosen to minimize the Stein's unbiased risk estimate

$$\widehat{R}(\lambda) = \sum_{j=1}^{n} \lambda_j^2 \frac{\widehat{\sigma}_j^2}{n} + \sum_{j=1}^{n} (1 - \lambda_j)^2 \left(Z_j^2 - \frac{\widehat{\sigma}_j^2}{n} \right)_+ . \tag{10}$$

Beran and Dümbgen showed that $\widehat{R}(\lambda)$ is asymptotically, uniformally close to $R(\lambda)$ in either the monotone or NSS case.

The Beran-Dümbgen method is based on the weak convergence of the "pivot process" $B_n(\widehat{\lambda}) = \sqrt{n}(L_n(\widehat{\lambda}) - \widehat{R}(\widehat{\lambda}))$ to a Normal $(0, \tau^2)$ for some $\tau^2 > 0$. (The estimator for τ^2 is given in the Appendix 3.) It follows that

$$\mathcal{D}_n \;=\; \left\{ \mu : \; \frac{L_n(\widehat{\lambda}_n) - S_n(\widehat{\lambda}_n)}{\widehat{\tau}_n / \sqrt{n}} \leq z_\alpha \right\}$$

$$= \left\{ \mu : \sum_{i=1}^{n} (\widehat{\mu}_i - \mu_i)^2 \leq \frac{\widehat{\tau}_n \, z_\alpha}{\sqrt{n}} + \widehat{R}(\widehat{\lambda}_n) \right\}$$

is an asymptotic $1-\alpha$ confidence set for the coefficients, where z_α denotes the upper α quantile of a standard Normal and where $\widehat{\mu}_i \equiv \widehat{\mu}_i(\widehat{\lambda}_n)$. Thus

$$\mathcal{B}_n = \left\{ f(x) = \sum_{j=1}^{n} \mu_j \phi_j(x) : \quad \mu \in \mathcal{D}_n \right\} \tag{11}$$

is an asymptotic $1-\alpha$ confidence set for f.

The approach to confidence sets that we use here is quite different than the more familiar confidence band approach in which one constructs bands of the form $\widehat{f}(x) \pm c\sqrt{\widehat{\mathsf{Var}}(\widehat{f}(x))}$ for some c. The advantage of bands is that by plotting them, we get a simple visual impression of the uncertainty. However, there are some drawbacks to bands. In their most naive form, the constant $c = z_{\alpha/2}$, which does not account for the multiplicity over the xs. This can be fixed by using a larger constant, although the computation of the constant is, in some cases, nontrivial. See Sun and Loader (1994). Second, the available results about coverage appear to be pointwise rather than uniform over $f \in \mathcal{F}$ although we suspect that the results can be strengthened to be asymptotically uniform. The third, and most serious problem, is that the function estimate $\widehat{f}$ is biased so the confidence interval is not centered properly, resulting in uncercoverage. Specifically, letting $s(x)$ denote the standard error of $\widehat{f}$ and $m(x) = \mathsf{E}\,\widehat{f}(x)$, we have that

$$\frac{\widehat{f}(x) - f(x)}{s(x)} = \frac{\widehat{f}(x) - m(x)}{s(x)} + \frac{m(x) - f(x)}{s(x)}.$$

The first term typically satisfies a central limit theorem. The second term

does not tend to zero since optimal smoothing causes the bias $m(x) - f(x)$ to be of the same order as $s(x)$. There have been some attempts to control this smoothing bias; see Ruppert, Wand and Carroll (2003) for a discussion.

The confidence ball approach automatically deals with the smoothing bias, at least approximately. This is because the ball takes the object $||\widehat{f}(x) - f(x)||^2$ as its starting point, rather than $\widehat{f}(x) - m(x)$ which is implicit in the band approach. The ball approach does have some bias, since $\widehat{f}$ actually estimates $f_n(x) = \sum_{j=1}^{n} \mu_j \phi_j(x)$ rather than $f(x) = \sum_{j=1}^{\infty} \mu_j \phi_j(x)$ resulting in a tail bias of $\sum_{j=n+1}^{\infty} \mu_j^2$. However, this tail bias is small relative to the smoothing bias.

3 Dealing with Heteroskedastic Errors

As Figure 5 shows, the data for the CMB power spectrum are highly heteroskedastic. The confidence set based on $\mathcal{L}^2$ loss is a ball and thus gives equal weight to deviations in all direction. Because the the CMB variances are tiny for some ℓs and huge for others, this symmetry is inappropriate. In parametric inference, confidence sets under heteroskedasticity are typically ellipses rather than balls, and we need to make a similar adjustment. We do this by constructing the confidence set under a loss function that gives more weight to points where the spectrum is measured precisely. In this section, we extend the Beran-Dümbgen method to such weighted loss functions.

We now replace the L^2 loss function with the following weighted loss:

$$L(f, \widehat{f}) = \int (f - \widehat{f})^2 w^2,$$

where we take $w^2(x) = 1/\sigma^2(x)$. We expand both the unknown function and the weight function w^2 in the orthonormal basis. Hence, we write

$$\begin{aligned} f(x) &= \sum_j \beta_j \phi_j(x) \\ w^2(x) &= \sum_j w_j \phi_j(x), \end{aligned}$$

where $\phi_0, \phi_1, \ldots$ is the cosine basis on [0,1] defined above.

The construction of $\mathcal{B}_n$ requires a new central limit theorem and a modified estimate of the asymptotic variance. We also replace the risk estimator in equation (10) by the following, which can be shown to be unbiased for the new loss function:

$$\widehat{R} = Z^T \bar{D} W \bar{D} Z + \text{trace}(DWDB) - \text{trace}(\bar{D} W \bar{D} B), \tag{12}$$

where D and $\bar{D} = I - D$ are diagonal matrices with 1's in the first J and last $n - J$ entries, $B = U \Sigma U^T$ is the covariance of Z, and $W_{jk} = \sum_\ell w_\ell \Delta_{jk\ell}$ with w_ℓ being the ℓ^{th} expansion coefficient of the function w^2 and

$$\begin{aligned} \Delta_{jk\ell} &= \int_0^1 \phi_j \phi_k \phi_\ell \\ &= \begin{cases} 1 & \text{if } \#\{j,k,\ell = 0\} = 3 \\ 0 & \text{if } \#\{j,k,\ell = 0\} = 2 \\ \delta_{jk}\delta_{0\ell} + \delta_{jl}\delta_{0k} + \delta_{kl}\delta_{0j} & \text{if } \#\{j,k,\ell = 0\} = 1 \\ \frac{1}{\sqrt{2}}(\delta_{\ell,j+k} + \delta_{\ell,|j-k|}) & \text{if } j,k,\ell > 0. \end{cases} \end{aligned}$$

The set $\mathcal{B}_n$ is defined as in equation (11) but with the new estimate of risk. The estimated variance of the pivot, $\widehat{\tau}^2$, is also different and is given in the appendix.

4 Results

We applied our method to the WMAP data to obtain a confidence set for the unknown spectrum $f(\ell/L_{\max}) \equiv C_\ell$. Figure 6 compares the center of our confidence ball with the so-called "Concordance model" (Spergel et al. 2003). The Concordance model is the maximum likelihood estimator for a likelihood of the form

$$\begin{aligned}
&\mathcal{L}_{\mathrm{Conc}}(\theta; Y_{\mathrm{WMAP}}, Y_{\mathrm{LSS}}, Y_{\mathrm{Lyman}}, Y_{\mathrm{CBI}}, Y_{\mathrm{Acbar}}) \\
&\quad = \quad \mathcal{L}_{\mathrm{WMAP}}(\theta; Y_{\mathrm{WMAP}}) \cdot \\
&\qquad \mathcal{L}_{\mathrm{LSS}}(\theta; Y_{\mathrm{LSS}}) \cdot \mathcal{L}_{\mathrm{Lyman}}(\theta; Y_{\mathrm{Lyman}}) \cdot \mathcal{L}_{\mathrm{CBI}}(\theta; Y_{\mathrm{CBI}}) \cdot \mathcal{L}_{\mathrm{Acbar}}(\theta; Y_{\mathrm{Acbar}}), \quad (13)
\end{aligned}$$

where the Ys are independent data sets from different experiments (WMAP: Bennett et al. 2003; LSS: Percival et al. 2003; Lyman: Croft et al. 2002, Gnedin and Hamilton 2002; CBI: Mason et al. 2003, Sievers et al. 2003, Pearson et al. 2003; Acbar: Kuo et al. 2001). In particular, Y_{WMAP} is the data set we are using. The parametric fit from the WMAP data alone (see Figure 10, top right) is obtained by maximizing only the first component $\mathcal{L}_{\mathrm{WMAP}}(\theta; Y_{\mathrm{WMAP}})$.

Note how well the nonparametric curve compares to the Concordance spectrum. The notable exceptions are in the very high-ℓ region around the

third peak and the low-ℓ region where the physical models curve upward sharply. We will argue that both the third peak and the rise in the spectrum at low ℓs are by-products of the model and not the data. All of the cosmological models share both features. We are not suggesting that these features are incorrect, but we believe it is useful to separate effects driven by the data from those driven by the model. See Section 5.

Once we construct the confidence ball, the next step is to use it to draw inferences. Because the ball is 900-dimensional in this case, it can seem daunting to display results. Fortunately, our construction provides simultaneous coverage over all functionals of the unknown function, pre or post hoc. We thus explore the uncertainty by creating targeted *probes* of the ball.

First, we explore the uncertainty in the location and height of the peaks. To do so, we searched through the confidence ball using local quadratic probe. Specifically, at each location ℓ_0, we defined a quadratic $q_h(\ell)$ with support on the interval $[\ell_0 - \Delta, \ell_0 + \Delta]$ for fixed $\Delta = 51$, centered at ℓ_0, and with height h. If $\widehat{f}$ is the center of our confidence ellipse, we considered perturbations of the form $f = \widehat{f} + q_h$. We varied h to find the largest and smallest values such that the resulting f is within the confidence ball and maintains three peaks over the ℓ range $[2, 900]$. This results in confidence limits on the peaks as shown in Figure 7.

One striking result is the different precisions with which the first and second peaks are resolved. This is to be expected given the large variances near the second peak. In other words, the data alone give little information about

the second peak. (The third peak is even more uncertain.) The published results in the physics literature present the second peak with much lower uncertainty. We return to this issue in Section 5.

Figure 8 shows an example of a model-directed probe. Using the CMBFAST software package (Seljak and Zaldarriaga 1996), we generated spectra in a one-dimensional family centered on the Concordance model parameters. The figure, which we call a ribbon plot, shows how the spectrum changes as the baryon fraction Ω_{b} is varied while keeping the total energy density Ω_{Total} fixed at 1. The light gray curves are in the ball; the black curves are not. The resulting interval for $\Omega_b h^2$ is [0.0169,0.0287]. To generate a valid confidence interval with such a probe we would need to search the entire 11-dimensional parameter space.

The data are much noisier for high ℓs, and we want to quantify how this propogates into local uncertainty about the function because this our ability to resolve the second and third peaks. A simple probe of the confidence set is useful for this purpose: finding how far a particular function in the confidence ball can be perturbed by localized deviations. For example, at each ℓ, we examined the one-dimensional family of spectra $f_h = \widehat{f} + h \cdot b$, where b is a boxcar of fixed width and unit height centered at ℓ. Figure 9 shows the maximum absolute height h that remains in the 95% ball relative to the height of the Concordance spectrum, for two different boxcar widths. At ℓs where this curve is greater than 1, the data arguably contain little information about the height of the curve near that location.

The confidence ball is also useful for model checking. Figure 10 shows four different spectra along with the minimum value of $1-\alpha$ for which each spectrum is in the $1-\alpha$ confidence ball. The concordance spectrum is very close to the center. But the best fitting parametric model using only the WMAP data is at best in the 73% confidence ball. Cosmologists often use 68% confidence levels, so this can be seen as a weak rejection of the best fitting model from the data. We also considered two extremal models that are in the 95% ball. These show that the data alone are consistent with eliminating the second and third peaks. While the cosmological models all predict these peaks – through the acoustic oscillations caused by dark matter – this suggests the benefits of more precise data, as from the Planck mission (Balbi et al. 2003).

5 Discussion

5.1 Findings

Our most striking finding is that the center of our nonparametric confidence ball using the WMAP data alone lies very close to the Concordance model fit over the range where the data are not noise dominated. Recall from equation (13) that the Concordance model incorporates four independent data sets, each with distinct likelihood forms. In contrast, the parametric model fit using only the WMAP data (with likelihood $\mathcal{L}_{\text{WMAP}}$) lies barely in the 73% confidence ball. Given that cosmologists often use 68% confidence intervals as their standard of evidence, this is tantamount to a rejection of the

cosmological model that underlies that parametric fit.

This raises two points. First, it is remarkable that with a fully nonparametric method we have come very close to the Concordance model based on the WMAP data alone. Second, that we obtained basically the same spectrum as the Concordance model calls into question the accuracy of the WMAP-only likelihood $\mathcal{L}_{\text{WMAP}}$.

5.2 Methods

We have presented a nonparametric method for analyzing the CMB spectrum. Our techniques have wide applicability to regression problems beyond cosmology. By starting with a confidence ball, then probing the ball using functionals, one can address a variety of questions about the unknown function while maintaining correct coverage, despite multiplicity and post-hoc selection.

The method in this paper modifies the original Beran-Dümbgen construction to account for heteroskedasticity. This modification yields a substantial reduction in the size of the confidence set. The resulting confidence set is also more useful in that it leads to tighter (looser) bounds in regions where the function is more (less) accurately measured.

One advantage of our approach, is that it allows one to separate the information in the data from the information in a model. Although we did not pursue the full calculation here, we could intersect our confidence ball with the manifold of spectra from the parametric model as a way to combine

data and model. Specifically, we could use the cosmological model to generate spectra, but then test which spectra are consistent with the data by reference to our confidence ball. This does not rely on likelihood asymptotics which, as we discuss below, are suspect in this problem. Another advantage is that by extending this analysis to a constrained noparametric model (such as a three peak model) that contains the cosmological model, we can make the same inferences without being tied to the analytic form of the model. Our approach can then be used to check the model, make inferences under the model, and compare parametric to nonparametric inferences.

We should point out that cosmologists obtain confidence intervals for parameters in their (11-dimensional) model by integrating over the nusiance parameters and producing a marginal posterior. However, the likelihood is ill-behaved, under-identified, and degenerate. Moreover, in the physics literature, there does not seem to be a clear appreciation of the fact that interval estimates obtained this way need not have correct frequentist coverage.

There are several other advantages to our approach. If a parameter is under-identified this will show up automatically as a wide confidence interval. The intervals have correct asymptotic coverage and simultaneous validity over all parameters of interest. There is no need to integrate or profile the likelihood function. Finally, the asymptotic theory for (6) is insensitive to the fact that the standard asymptotics for the likelihood approach fail.

5.3 Inferential Foundations

Interestingly, there seems to be some confusion about the validity of frequentist inference in cosmology. Since we have access to only one Universe – and thus cannot replicated it – some feel that it makes no sense to make frequentist inferences. This represents a common misunderstanding about frequentist inference in general and confidence intervals in particular. The frequency statements for confidence intervals refer to the procedure, not the target of the inference. Our method for constructing confidence balls traps the true function 95 percent of the time, even over a sequence of different, unrelated problems. There is no need to replicate the given experiment, or Universe.

Complicating matters is the fact that the coefficients $a_{\ell,m}$ of the temperature field are random and unknown. To see the importance of this point, it is useful to make a finer distinction by defining the realized spectrum $\widetilde{C}_\ell = (1/(2\ell+1))\sum_\ell |a_{\ell,m}|^2$, the "true spectrum" $C_\ell = \mathsf{E}\,(\widetilde{C}_\ell)$ and the measured spectrum $\widehat{C}_\ell$. Note that all our inferences have actually been directed at the realized spectrum. Some phyciststs find it disturbing to be making frequentist inferences about $\widetilde{C}_\ell$ since it is a realization of a random variable rather than a parameter in the usual sense. But this is no different than making inferences about a random effect in a standard random effects model.

These confusions have led to an interesting movement towards Bayesian methods in cosmology. Of course, when used properly, Bayesian methods can be very effective. Currently however, the Bayesian interval estimates in

the physics literature seem questionable, being based on unfettered use of marginalizing over high-dimensional, degenerate likelihoods using flat priors chosen mainly for convenience. Indeed, an active area of research is finding corrections for such intervals to make them have correct coverage. Moreover, the potentially poor coverage of the Bayesian interval seems not to have been widely recognized in the Physics literature.

Appendix 1. CMB Data.

The CMB is composed of photons. The temperature of these photons (2.726 Kelvin) means that the radiation will be at the microwave wavelengths. The light is collected via a dish (or reflector) and fed into either a (1) bolometer, which senses small temperature change as the photon hits the detector, or (2) a high performance transistor. In some cases (such as the aforementioned COBE experiment), the telescope is placed in orbit above the Earth. In other cases, the telescope is placed on a balloon and launched into the atmosphere. With careful attention paid to ground reflections, CMB telescopes can also be placed on the ground in regions where the atmosphere will contribute little contamination (like the South Pole). In all cases, there are a series of steps leading from the raw data collection to the final power spectrum estimation.

The raw data are collected in a time stream. For each moment in time, the telescope records a temperature difference on the sky between two widely separated points. For example, one of these locations could be a fixed source of known temperature, thus allowing the temperature at the other point

to be calculated. However, the comparison location need not be fixed (or known) and the absolute temperatures can still be solved for iteratively, using previous measurements. Throughout this process, the *pointing* of the telescope needs to be accurately known (as a function of time), as well as the *calibration* of the temperatures, and also the instrument noise.

Appendix 2. Cosmological Parameters

The physics of the Universe on large scales is well described by a small set of cosmological parameters. We describe some of the most important parameters below.

Normalized Hubble Constant h. The Hubble constant is the rate of the Universe's expansion. Specifically, $H = \frac{\dot{a}}{a}$ where a is the size of the Universe and $\dot{a}$ is the rate of change in a. "Constant" is a misnomer since this is a dynamic quantity. The "Hubble constant" refers to the value of H as measured today (H_0); this is often normalized and reported as $h = H_0/100.0$ with units km s^{-1} Mpc^{-1}.

Total Energy Density Ω_{Total}. Ω_{Total} is the energy density of the Universe divided by the critical density of the Universe: $\rho_{crit} = 3c^2H_0^2/8\pi G$ at which the Universe would be geometrically flat. Ω_{Total} can be broken down into the sum of different components, like Ω_{baryons}, $\Omega_{\text{dark matter}}$, $\Omega_{\text{neutrinos}}$.

Cosmological Constant Λ. Λ is a constant that was added by Einstein into his general relativistic field equations to produce a static Universe. The constant was later dismissed as unnecessary after the discovery by Edwin Hubble

that the Universe is not static, but expanding. However, recent studies show strong evidence for a cosmological constant term. Λ acts as a negative pressure and thus might accelerate the expansion of the Universe. We often speak of the energy density component Ω_Λ, which is then included in the sum of Ω_{Total}.

Baryon Density Ω_{b}. This is the density component of baryonic matter in the Universe (e.g. protons, neutrons, etc). The fraction of matter density that is baryonic (over the total matter density of the Universe which includes baryons and non-baryonic dark matter) is often measured to be in the range: 15%-20%.

Dark Matter Density Ω_{d}. The majority of matter in the Universe is detected indirectly through it's gravitational effects. Since it cannot be seen or measured in the laboratory, it has been dubbed "dark matter". Ω_{d} is the energy density component strictly due to dark matter.

Neutrino Fraction, f_ν. The fraction of the neutrino density over the total matter density: $f_\nu = \Omega_\nu/(\Omega_{\text{b}} + \Omega_{\text{d}})$.

Optical Depth τ. We know today that most of the hydrogen in the Universe is ionized. So at some time after recombination, the Universe was reionized. τ is the optical depth due to Thomson scattering up to a redshift of $z < z_{ionization}$: $\int_0^{t(z_{ionization})} \sigma_T n_e dt$ where σ_T is the Thomson scattering cross-section and n_e is the electron density.

Spectral Index n_s. The primordial matter density fluctuation spectrum is proportional to the scale size raised to the power n, the primordial spectral

index. On large enough scales, the the CMB temperature power spectrum's spectral index (n_s) is then close (or equal) to the primordial spectral index.

The spectrum may be approximated numerically as a function of these parameters using the CMBFAST software package (Seljak and Zaldarriaga 1996). Figure 8 shows spectra corresponding to a range of cosmological parameter settings. For example, the location and amplitude of the first peak is related to the total energy density Ω_{Total}. The baryon fraction Ω_{b} and the spectral index n_s drive the ratio of the amplitude of the first and second peaks. The ratio of the amplitudes of the second to third peaks depends on the density of matter ($\Omega_{\text{b}} + \Omega_{\text{d}} + f_\nu$), h, and n_s.

Appendix 3. Estimating τ.

Recall from Section 3 that the cosine basis is defined on [0,1] by

$$\phi_0(x) = 1, \quad \phi_j(x) = \sqrt{2}\cos(\pi j x) \quad j > 0.$$

Then, if j and k are distinct and positive,

$$\begin{aligned} \phi_j \phi_k &= 2\cos(\pi j x)\cos(\pi k x) \\ &= \cos(\pi(j+k)x) + \cos(pi(j-k)x) \\ &= \frac{1}{\sqrt{2}}\left(\phi_{j+k} + \phi_{|j-k|}\right). \end{aligned}$$

If $j > 0$

$$\phi_j^2 = 2\cos^2(\pi j x) = \cos(\pi 2 j x) + 1 = \frac{1}{\sqrt{2}}\phi_{2j} + \phi_0.$$

Hence,

$$\begin{aligned}\Delta_{jk\ell} &= \int_0^1 \phi_j\phi_k\phi_\ell \\ &= \begin{cases} 1 & \text{if } \#\{j,k,\ell=0\}=3 \\ 0 & \text{if } \#\{j,k,\ell=0\}=2 \\ \delta_{jk}\delta_{0\ell}+\delta_{jl}\delta_{0k}+\delta_{kl}\delta_{0j} & \text{if } \#\{j,k,\ell=0\}=1 \\ \frac{1}{\sqrt{2}}(\delta_{\ell,j+k}+\delta_{\ell,|j-k|}) & \text{if } j,k,\ell>0 \end{cases}\end{aligned}$$

We thus have that

$$\begin{aligned}L(f,\widehat{f}) &= \int (f-\widehat{f})^2 w^2 \\ &= \sum_{j,k,\ell}(\beta_j-\widehat{\beta}_j)(\beta_k-\widehat{\beta}_k)w_\ell\int\phi_j\phi_k\phi_\ell \\ &= \sum_{j,k}(\beta_j-\widehat{\beta}_j)(\beta_k-\widehat{\beta}_k)\sum_\ell w_\ell\Delta_{jk\ell} \\ &= (\beta-\widehat{\beta})^T W(\beta-\widehat{\beta}),\end{aligned}$$

where $W_{jk}=\sum_\ell w_\ell\Delta_{jk\ell}$.

Let $\bar{\lambda}=1-\lambda$ and let $D(x)$ denote the diagonal matrix with x along the diagonal. Write $\widehat{\beta}=D(\lambda)Z$. Assume Z has a Normal$\langle\beta,B\rangle$ distribution. Then, $\mathsf{E}\,\widehat{\beta}=D(\lambda)\beta$ and since $\mathsf{Cov}\,(\widehat{\beta}_j,\widehat{\beta}_k)=\lambda_j\lambda_k B_{jk}$, $\mathsf{Var}\,(\widehat{\beta})=D(\lambda)BD(\lambda)$. Then,

$$\begin{aligned}\mathsf{E}\,L &= \mathsf{E}\,(\widehat{\beta}-\beta)^T W(\widehat{\beta}-\beta) \\ &= \text{trace}(D(\lambda)WD(\lambda)B)+\beta^T D(\bar{\lambda})WD(\bar{\lambda})\beta.\end{aligned}$$

The latter quadratic form can be written as $\sum_{j,k}\beta_j\beta_k\bar{\lambda}_j\bar{\lambda}_k W_{jk}$. We obtain unbiased estimate $\widehat{L}$ by replacing $\beta_j\beta_k$ with $Z_jZ_k-B_{jk}$.

For convenience, let D denote $D(\lambda)$ and $\bar{D}$ denote $D(\bar{\lambda})$, The result is

$$\widehat{L} = Z^T\bar{D}W\bar{D}Z+\text{trace}(DWDB)-\text{trace}(\bar{D}W\bar{D}B).$$

It follows that,

$$\begin{aligned} L - \widehat{L} &= \beta^T W \beta - 2 Z^T D W \beta + Z^T D W D Z - \\ &\qquad Z^T (I - D) W (I - D) Z - \operatorname{trace}((W - DW - WD)B) \\ &= \beta^T W \beta - 2 Z^T D W \beta + Z^T (DW + WD - W) Z + \operatorname{trace}((DW + WD - W)B). \end{aligned}$$

Let $A = DW + WD - W$ and $C = 2DW\beta$. Then,

$$\begin{aligned} \mathsf{Var}\,(L - \widehat{L}) &= \mathsf{Var}\,(Z^T A Z - Z^T C) \\ &= \mathsf{Var}\,(Z^T A Z) + \mathsf{Var}\,(Z^T C) - 2\mathsf{Cov}\,(Z^T A Z, Z^T C) \\ &= 2 \operatorname{trace}(ABAB) + \beta^T Q \beta, \end{aligned}$$

where

$$\begin{aligned} Q/4 &= ABA + WDBDW - 2ABDW \\ &= (DW + WD - W)B(WD + DW - W) + WDBDW - 2(DW + WD - W)BDW. \end{aligned}$$

Hence, plugging in unbiased estimates of the linear and quadratic forms involving β, we get an estimate of the variance:

$$\widehat{\tau}^2 = 2 \operatorname{trace}(ABAB) + Z^T Q Z - \operatorname{trace}(QB). \tag{14}$$

Acknowledgments

The authors would like to thank the Executive Editor for his careful reading and helpful suggestions.

References

Papers below with references to `astro-ph` numbers can be found on-line in multiple formats at `http://xxx.lanl.gov/archive/astro-ph`. In addition, the following websites have useful introductory material about WMAP and the CMB:

1. The WMAP home page

   `http://map.gsfc.nasa.gov/`,

2. Wayne Hu's CMB Physics Pages

   `http://background.uchicago.edu/~whu/physics/physics.html`,

3. Ned Wright's Cosmology Tutorial

   `http://www.astro.ucla.edu/~wright/cosmolog.htm`,

4. and an introductory review of CMB studies

   `http://astron.berkeley.edu/~mwhite/rosetta/index.html`.

Balbi, A., Baccigalupi, C., Perrotta, F., Matarrese, S., & Vittorio, N. (2003). Probing Dark Energy with the Cosmic Microwave Background: Projected Constraints from the Wilkinson Microwave Anisotropy Probe and Planck *Astrophys. J.*, **588**, L5-L8.

Bennett et al. (2003). First Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results *Astrophys. J.*

Supp. **148**, 1-27.

Beran, R. (2000). REACT scatterplot smoothers: superefficiency through basis economy. *J. Amer. Statist. Assoc.* **95**, 155-171.

Beran, R. and Dümbgen, L. (1998). Modulation of estimators and confidence sets. *Ann. Statist.* **26**, 1826-1856.

Croft et al. (2002). Toward a Precise Measurement of Matter Clustering: Ly-alpha Forest Data at Redshifts 2-4, *Astrophys. J.*, **581**, 20–52

Gnedin, N. Y. and Hamilton, A. J. S., (2002). Matter power spectrum from the Lyman-alpha forest: myth or reality? *Monthly Notices of the Royal Astronomical Society*, **334**, 107–116.

Halverson, N.W and Leitch, E.M. and Pryke, C. and Kovac, J. and Carlstrom, J.E. and Holzapfel, W.L. and Dragovan, M. and Cartwright, J.K. and Mason, B.S. and Padin, S. and Pearson, T.J. and Shepherd, M.C. and Readhead, A.C.S. (2002). DASI First Results: A Measurement of the Cosmic Microwave Background Angular Power Spectrum. *Astrophy. J.l*, **568**, 38-45.

Hinshaw et al. (2003). First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: The Angular Power Spectrum *Astrophys. J. Supp.*, **148**, 135-159.

Hu, W. (1999). CMB Anisotropies: A Decadal Survey, in *Birth and Evolution of the Universe: RESCEU 1999*, astro-ph/0002520.

Hu, W. (2001). Physics of CMB Anisotropies, Warner Prize Lecture, American Astronomical Society Meeting, San Diego, Jan. 2001, http://background.uchicago.edu/∼whu/Presentations/warnerprint.pdf.

Hu, W. (2003). CMB Temperature and Polarization Anisotropy Fundamentals, *Annals of Physics*, **303**, 203, astro-ph/0210696.

Hu, W. and Dodelson, S. (2002). Cosmic Microwave Background Anisotropies, *Ann. Rev. Astron. Astrophys.*, **40**, 171, astro-ph/0110414.

Hu, W. and Sugiyama, N. (1995). Toward Understanding CMB Anisotropies and Their Implications *Phys. Rev. D*, **51**, 2199-2630.

Knox, L. (1999). Cosmic Microwave Background Bandpower Windows Revisited, *Physic Rev. D*, 60, issue 10, 103516, 5 pages.

Kuo et al. (2003). High Resolution Observations of the CMB Power Spectrum with ACBAR, astro-ph/0212289.

Lee, A. T. et. al. (2002), A High Spatial Resolution Analysis of the MAXIMA-1 Cosmic Microwave Background Anisotropy Data, *Astrophys. J.*, 561, L1-L6.

Marinucci, D. (2004). Testing For Non-Gaussianity On Cosmic Microwave Background Radiation: A Review *Statistical Science*, this issue.

Mason et al. (2002). The Anisotropy of the Microwave Background to l = 3500: Deep Field Observations with the Cosmic Background Imager, *Astrophys. J.*, **591**, 540–555.

Netterfield, C. B. et. al. (2002). A measurement by BOOMERANG of multiple peaks in the angular power spectrum of the cosmic microwave background. *Astrophys. J.*, **571**, 604-614.

Pearson et al. (2003). The Anisotropy of the Microwave Background to l = 3500: Mosaic Observations with the Cosmic Background Imager, *Astrophys. J.*, **591**, 556–574.

Percival et al. (2001). The 2dF Galaxy Redshift Survey: the power spectrum and the matter content of the Universe, *Monthly Notices of the Royal Astronomical Society*, **327**, 1297–1306.

Ruppert, D., Wand, M. and Carroll, R. (2003). *Semiparametric Regression.* Cambridge University press. Cambridge.

Schwarz, D.J. (2003). The first second of the Universe, *Annalen Phys.*, **12**, 220–270, `astro-ph/0303574`.

Seljak, U. and Zaldarriaga, M. (1996). A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies *Astrophys. J.*, **469**,

437-444.

Sievers et al. (2003). Cosmological Parameters from Cosmic Background Imager Observations and Comparisons with BOOMERANG, DASI, and MAXIMA, *Astrophys. J.*, **591**, 599–622.

Smoot, G. et al. (1992). Structure in the COBE differential microwave radiometer first-year maps. *Astrophys. J.*, **396L**, L1-L5.

Spergel, D. et al. (2003). First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters *Astrophys. J. Supp.*, **148**, 175-194.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9**, 1135-1151.

Sun, J. and Loader, C. R. (1994). Simultaneous Confidence Bands in Linear Regression and Smoothing, *Annals of Statistics*, **22**, 1328–1345.

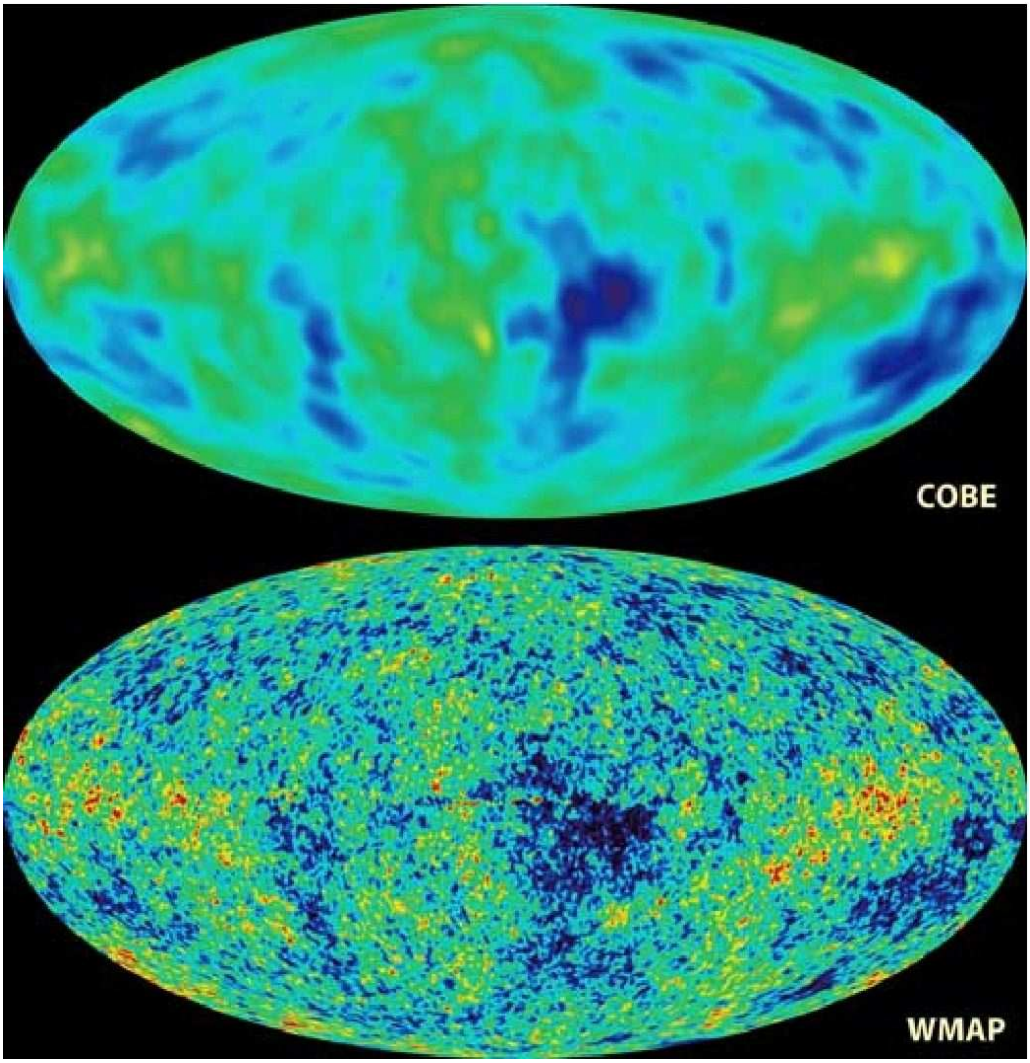

Figure 1. (top) The CMB as seen by the COBE satellite. The angular resolution of the satellite is about 10° and the various shades correspond to hot and cool spots with respect to the CMB blackbody temperature. (bottom) The CMB from the Wilkinson Microwave Anisotropy Probe. Notice the high angular resolution. Also notice that the large-scale structures are apparent in both the COBE and the WMAP data. Image courtesy of the WMAP Science Team and available at the WMAP Mission website: `http://map.gsfc.nasa.gov`.

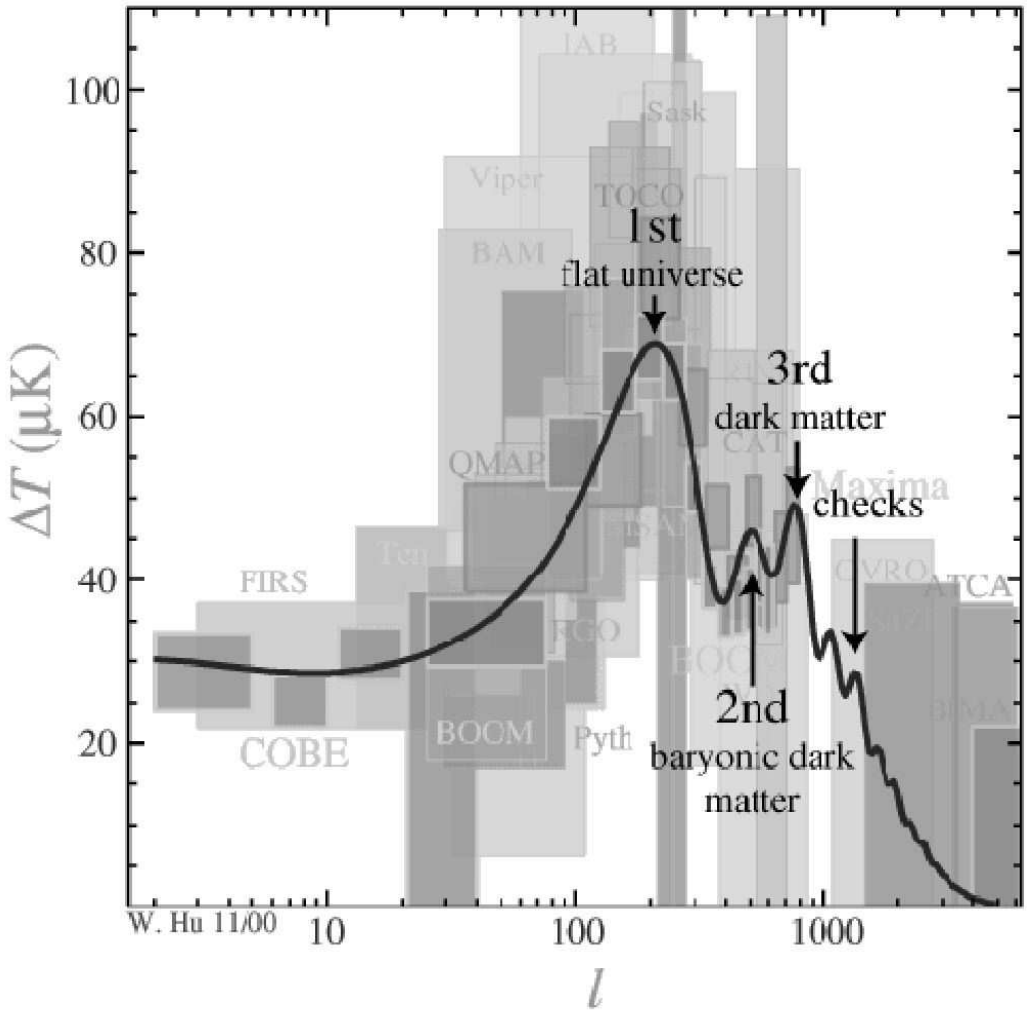

Figure 2. Estimated CMB spectrum showing the three peaks of interest. The underlayed boxes give the data ranges and uncertainties from a variety of older CMB experiments, not including WMAP. From Hu (2001).

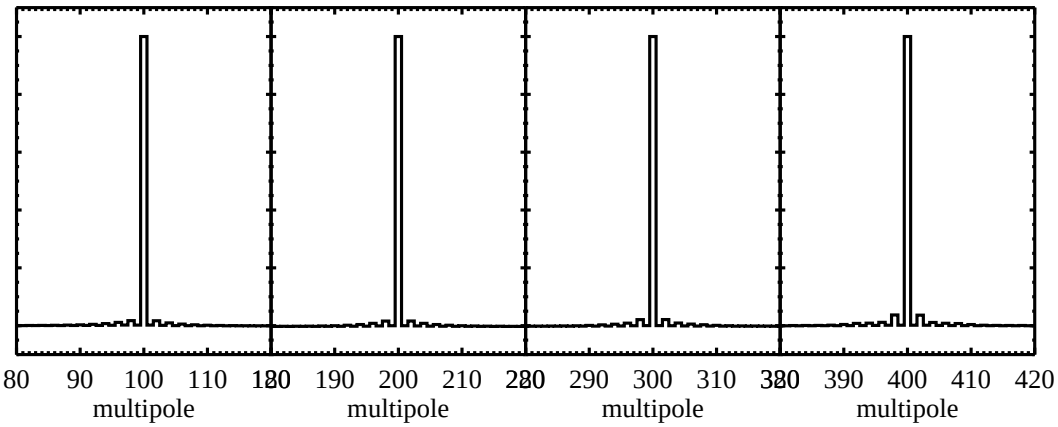

Figure 3. Bandpower windows from the matrix K centered on (left to right) $\ell = 100, 200, 300, 400$.

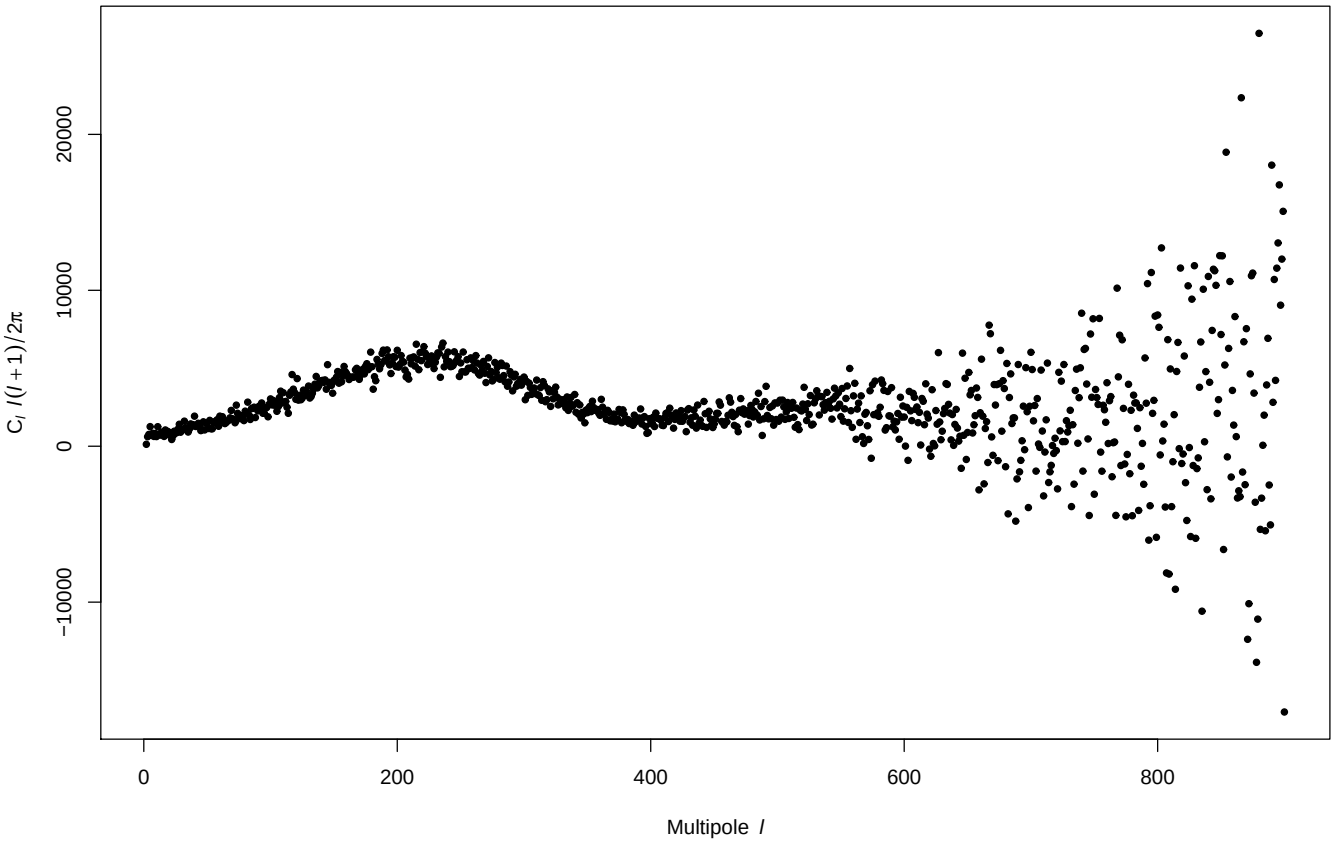

Figure 4. Y_ℓ as a function of ℓ for the WMAP data.

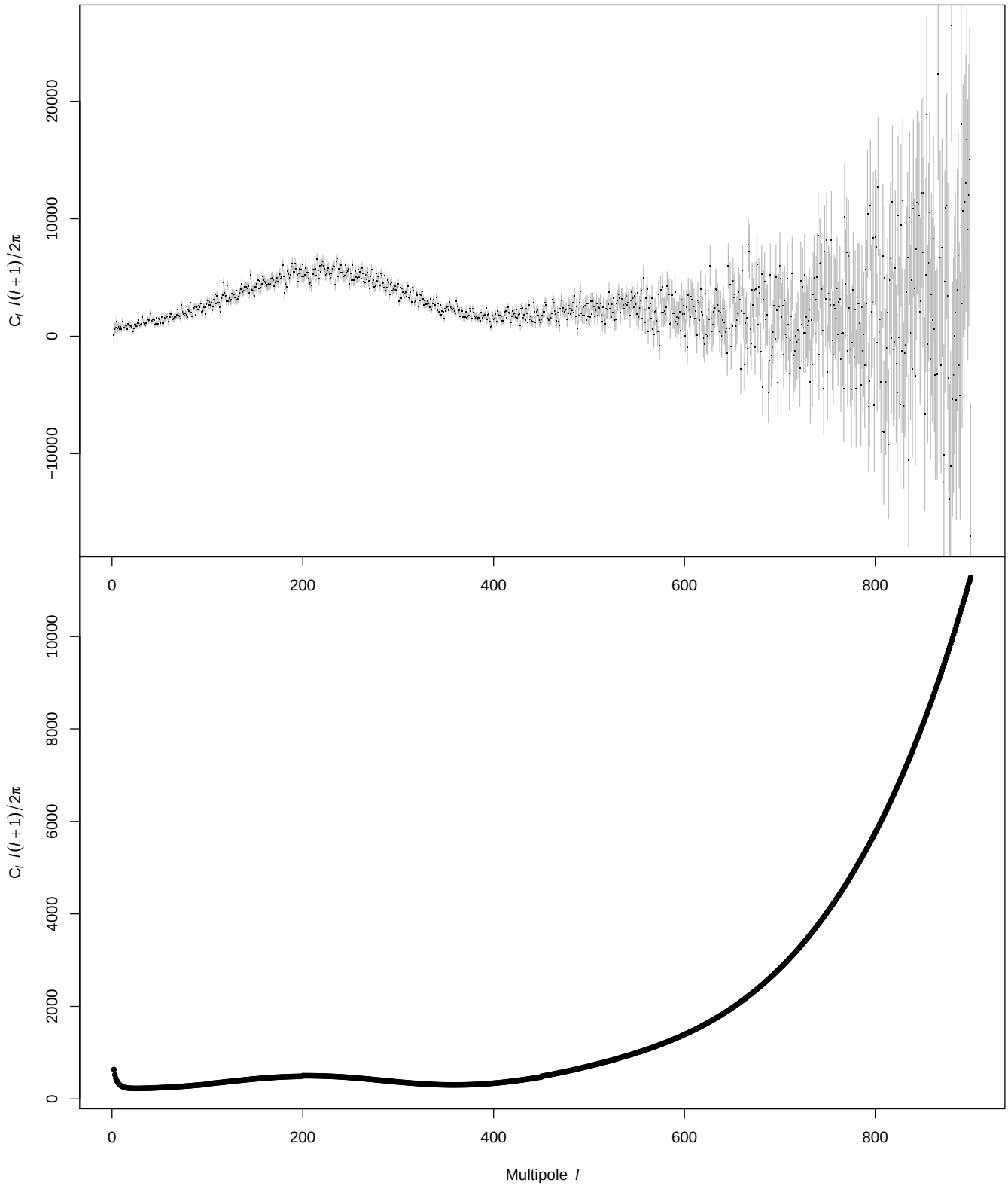

Figure 5. Noise standard deviation as error bars on data (above) and as a function of ℓ (below).

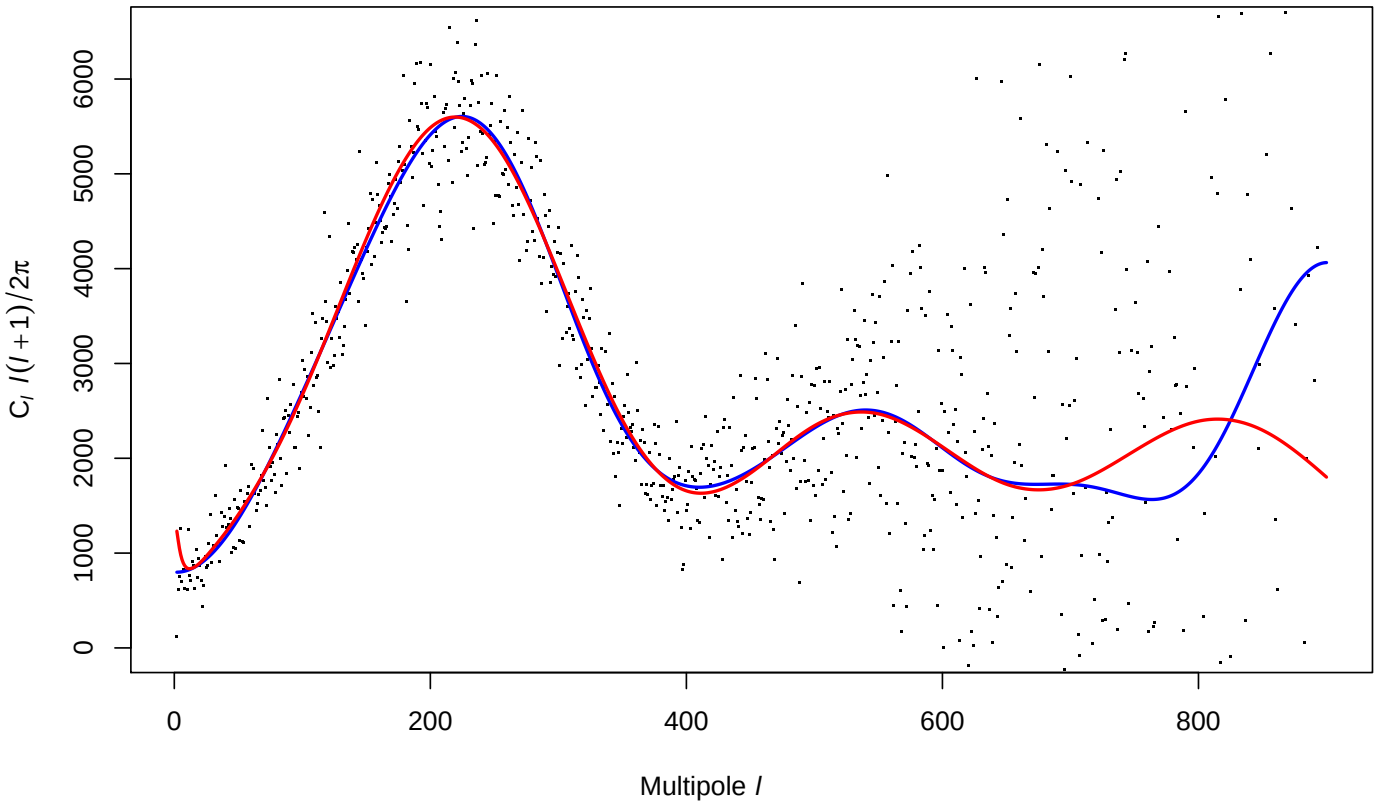

Figure 6. Center of our confidence ball (curve with sharp rise at right) and the power spectrum for the Concordance model (curve with three peaks). Note the striking agreement between the nonparametric fit and the parametric fit.

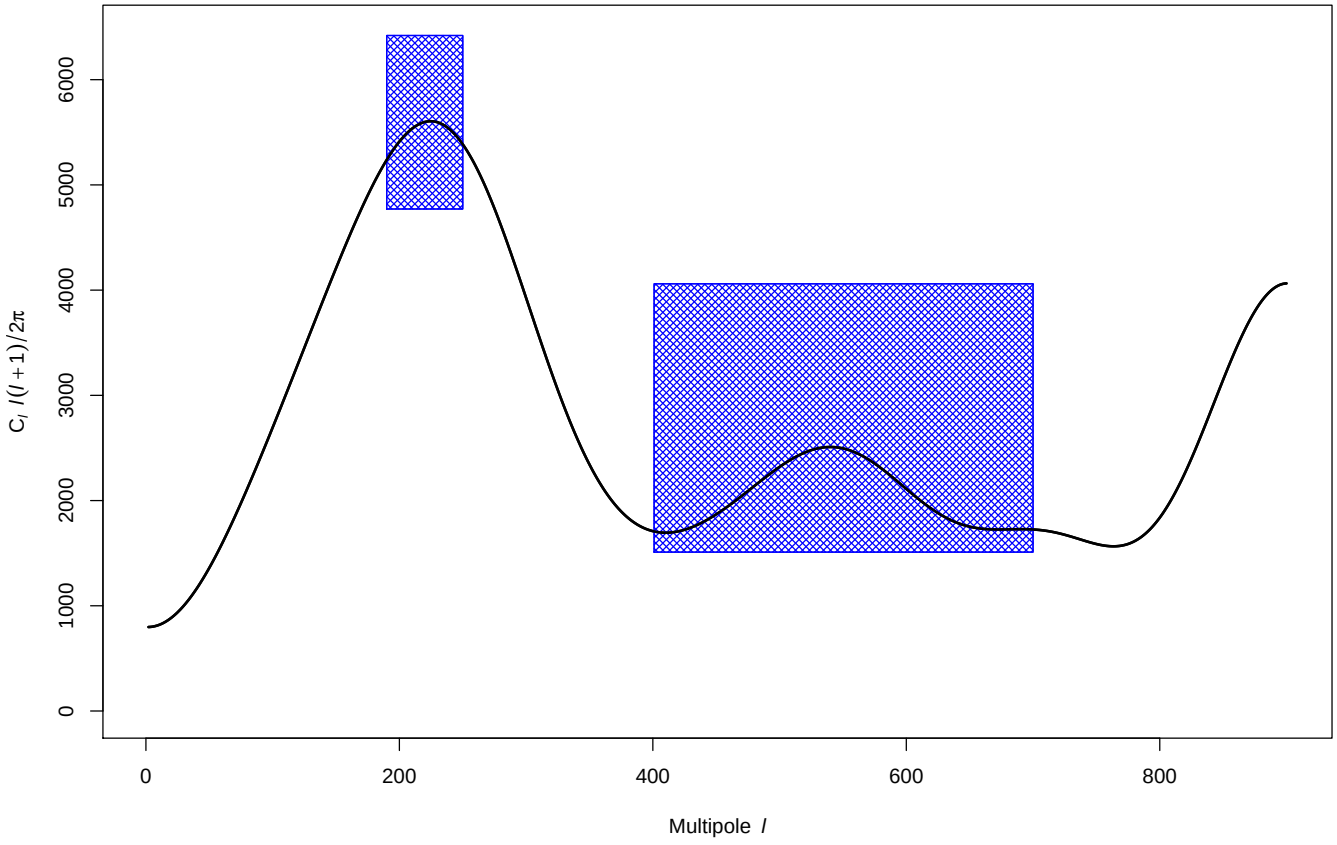

Figure 7. Center of our 95% confidence ball with superimposed 95% intervals for the heights and widths of the first two peaks.

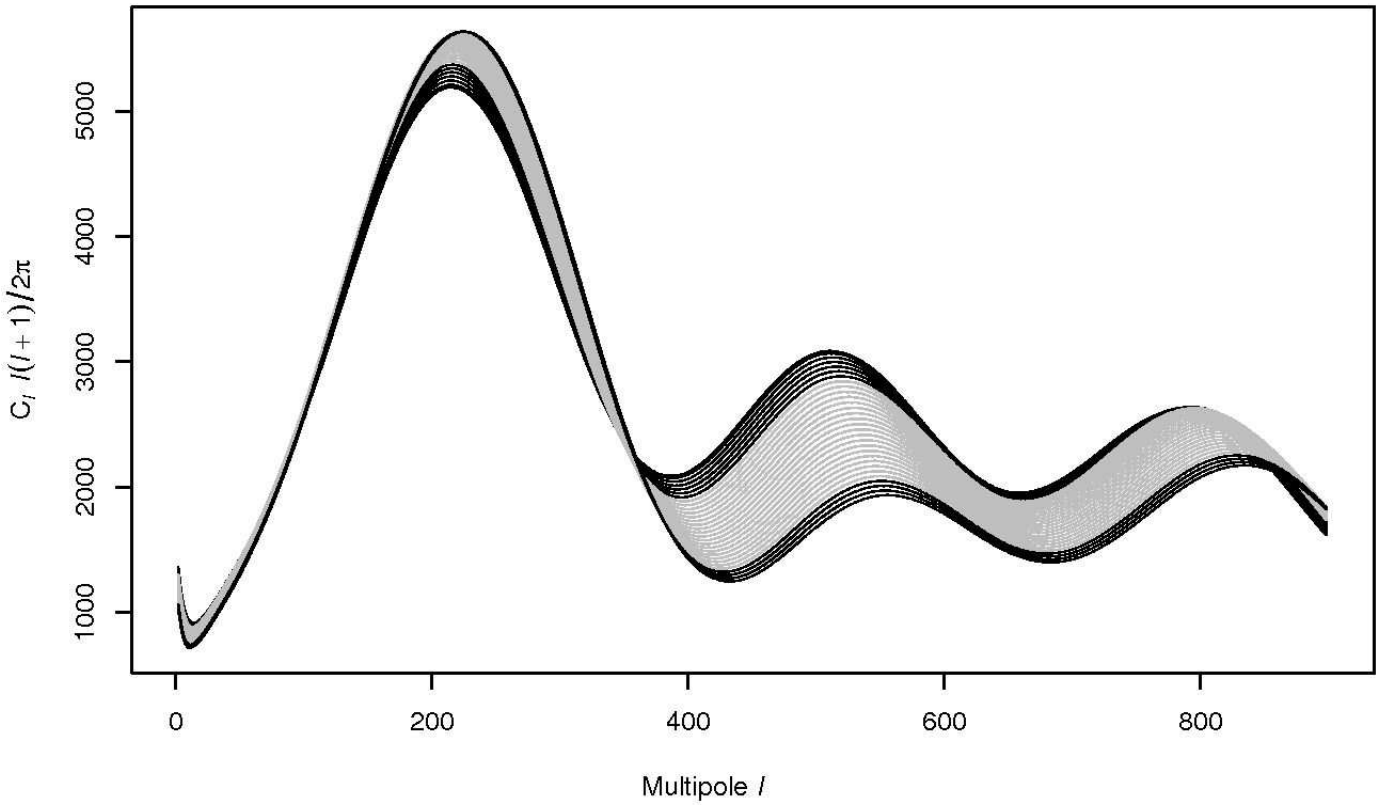

Figure 8. "Ribbon" probe of the confidence ball within the parametric model keeping $\Omega_{\rm Total}$ fixed at 1 and varying the Baryon fraction $\Omega_{\rm b}h^2$ from 0.01225 to 0.03675.

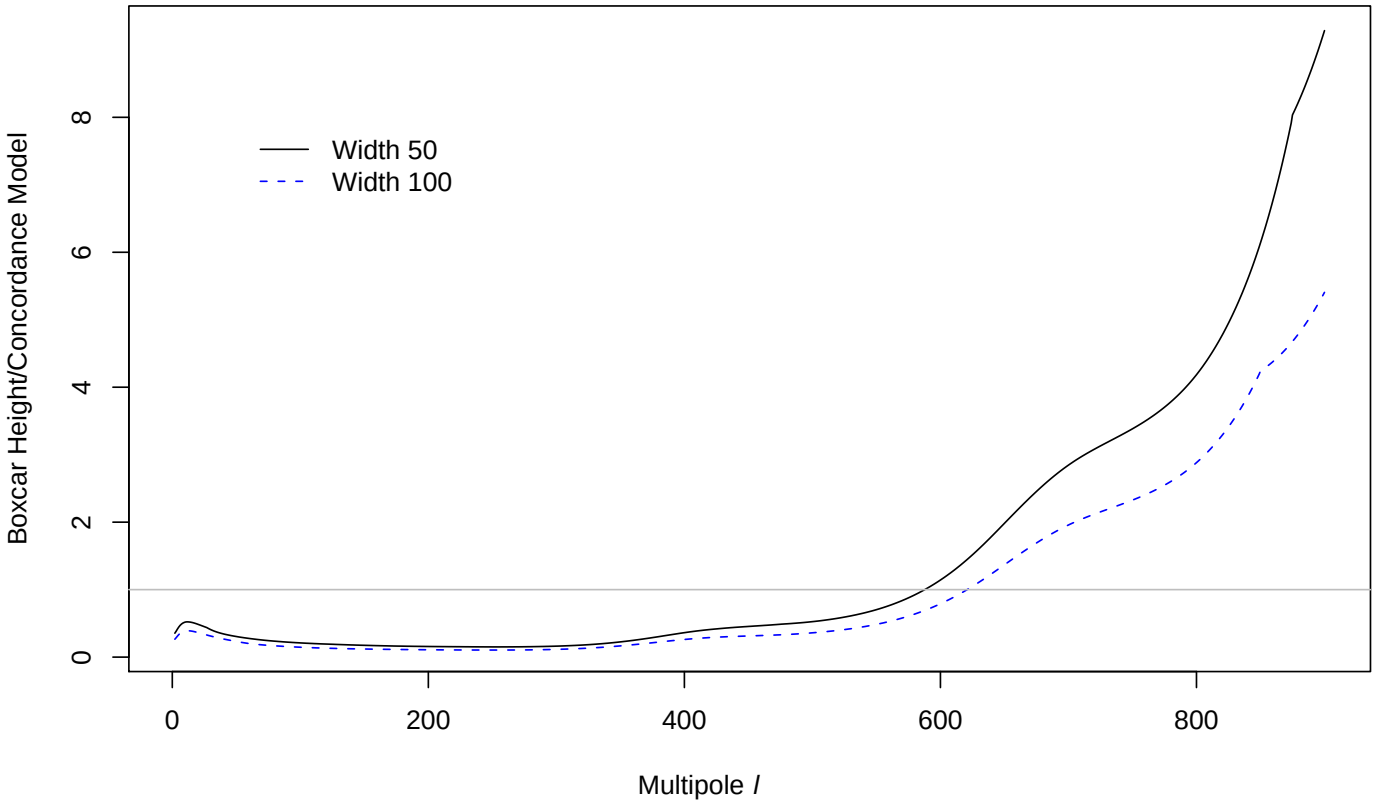

Figure 9. Height of local "box car probe" that is just in the 95% confidence ball, divided by the height of the Concordance spectrum, for two different box car widths. The horizontal line is at height 1.

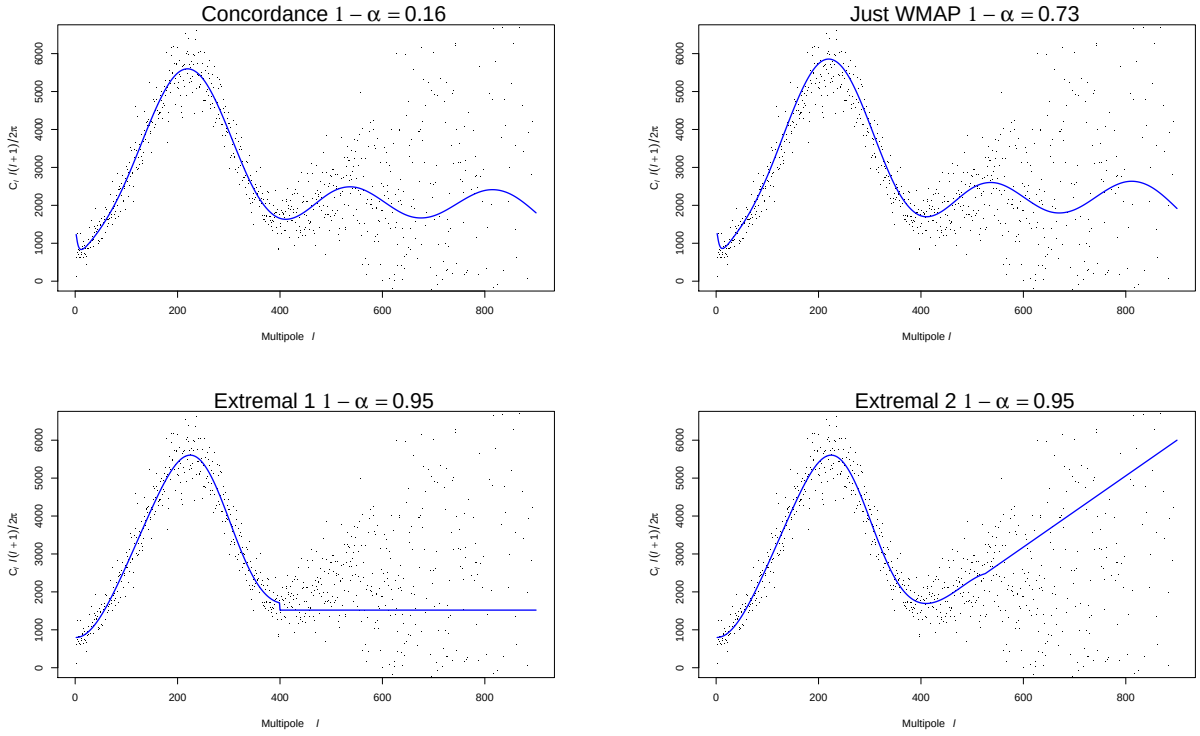

Figure 10. CMB spectra: (top left) Concordance model fit, (top right) WMAP-only model fit, and (bottom) two extremal fits.