On the Likelihood Function of Cosmic Microwave Background Anisotropy Observations

Eric Gawiser

Center for Astrophysics and Space Sciences, University of California at San Diego,

La Jolla, CA 92037

egawiser@ucsd.edu

1. Introduction

In testing theories of the origin of Cosmic Microwave Background (CMB) anisotropy versus observations, it is customary to compute the likelihood of the data given a particular theory, $L(d|t)$. A frequentist analysis can use this likelihood (or a related misfit statistic) to see which theories are ruled out at a certain confidence level, and a Bayesian analysis can multiply by the prior distribution, $p(t)$, to determine a credible region of theories that are the most likely given the observations. This paper offers a basic illustration of the processes that must be accounted for to compute this likelihood function. We will examine the assumptions required to produce a tractable form of the likelihood and discuss their accuracy. We will conclude by discussing what approximations are possible and what the consequences are of relaxing the input assumptions.

2. Bayesian and Frequentist Approaches

Let us take a minute to set some ground rules. There is no "right" approach to statistics, although there are better and worse statistics for answering well-posed questions. Sometimes a statistic is biased but one can calculate the expected bias and correct for it. The Bayesian and Frequentist approaches ask different questions and therefore have different ways of seeking answers. (We use the words hypothesis, theory, and model interchangeably below.)

The Frequentist approach is essentially a formalization of the scientific method as taught in elementary school. Consider a range of hypotheses. For each we ask, how likely is it that the data would disagree with our hypothesis as badly as they do? A hypothesis for which the data have a 32% chance of disagreeing worse is ruled out at 68% confidence (often referred to as "1σ" based on the distance from a Gaussian's mean at which the integrated probability reaches 68%). A 5% chance of disagreeing worse rules out at 95% (2σ) and a 1% chance rules out at 99% confidence ($\approx 3\sigma$). When this is done, we are left with a restricted set of hypotheses that have not been ruled out at a given confidence level. Although there will be a "maximum-likelihood" hypothesis that the data rules out at the lowest level of confidence, the data may well be "too good" of a fit for this hypothesis, i.e. there is over a 50% chance that the data would disagree more if this hypothesis is right. The Frequentist approach does not strongly discriminate between the surviving hypotheses until new data is obtained that has the power to rule out some of them at the desired confidence level. One significant advantage of the Frequentist approach is its ability to test entire paradigms; if all of the models corresponding to a particular scientific paradigm are ruled out, then the paradigm is ruled out. The Frequentist approach allows complementary datasets to be included in the analysis as long as one can calculate the joint likelihood and then the probability that the joint likelihood would be less than observed given a particular hypothesis.

The Bayesian approach attempts to answer a question which is perhaps more satisfactory for most scientists (and certainly more likely to get the media's interest): given what we knew before, and given the information in this dataset, what is our new best guess at the true model, and what range of models are at least 68% as likely to be correct as that model (the "1σ" credible region) or at least 95% as likely to be correct (the "2σ" credible region)? Notice that we no longer worry if this best-fit model is a surprisingly good or surprisingly poor fit to the data, and as we are only considering models **within** a paradigm we have lost the ability to rule out that paradigm. The detailed calculation depends on Bayes' Theorem:

$$p(d,t) = p(d|t)p(t) = p(t|d)p(d) \; , \tag{1}$$

where the statement is clearly true for any variables d and t but in this case they represent data and theories. $p(d|t)$ is just the likelihood (probability density) of the data given the theory, $L(d|t)$, introduced above. If we know the prior probability distribution of which theory is correct, then we can invert Equation 1 to calculate the posterior probability distribution of theories,

$$p(t|d) = \frac{p(d|t)p(t)}{p(d)} \; , \tag{2}$$

where $p(d) = \int_t p(d|t)p(t)$. For comparing between theories, $p(d)$ is just a constant, so we typically set this to one and make sure that $p(t|d)$ is normalized. The real conundrum is in choosing $p(t)$ which in most paradigms is unknown. A safe approach is to try various priors that appear to span the range of "reasonableness" and to see if the results vary as a function of the prior. For sufficiently narrowly peaked likelihood functions, the choice among "reasonable" priors will be unimportant, and the Bayesian approach can be pursued despite the uncertainty in the actual prior probability distribution of the theories. Notice that the definition of "reasonableness" is not only subjective but quite important as a sufficiently different prior is guaranteed to alter the results. The Bayesian approach allows complementary datasets (d_c) to be included in the analysis either as part of the data or as a modified prior $p'(t)$ where $p'(t) = p(d_c|t)p(t)$ is still sensitive to uncertainty in the initial prior $p(t)$.

3. Assumptions, Basics, and Correlations

A Cosmic Microwave Background dataset consists of a measurement of the intensity of microwave radiation at a set of pixels on the sky. This intensity is used to estimate the temperature of the Cosmic Microwave Background radiation at the set of points on the sky corresponding to the centers of the observed pixels. The average temperature ($T_{cmb} = 2.73\ K$) is subtracted off, yielding a measurement of the temperature anisotropy.

Let the data be given by d_i where i runs from 1 to N_{pixels}. This is an estimate of the true distribution of temperatures on the sky, s_i, plus noise such that $d_i = s_i + n_i$. The first type of information needed to compute the likelihood function is the statistical distribution of the noise involved in this observation, n_i, which has mean zero at every pixel and correlations given by $< n_i n_j >= C^N_{ij}$. Idealized experiments offer a diagonal matrix for $\mathbf{C}^N$ i.e. the noise is uncorrelated between pixels. This is a good approximation for a few instruments but not for most, so we use the full matrix $\mathbf{C}^N$. Another assumption is made at this point, that the noise is Gaussianly distributed in some basis, such that all higher-order statistics of the noise distribution are determined by $\mathbf{C}^N$.

Assumption 1: The instrument noise obeys a multivariate Gaussian distribution described by $< n_i >= 0$ and the correlation matrix C^N. This seems to be relatively accurate, even for instruments with strongly correlated $1/f$ noise where the matrix is far from diagonal. Careful analysis is required to show that unsubtracted cosmic rays and other instrumental effects do not produce significant non-gaussianity./footnoteStrictly speaking, it is not sufficient to show that the distribution of noise in individual pictures is gaussian; it is possible to have a distribution of noise whose marginal distribution in each pixel is gaussian but which is not a multivariate gaussian distribution (see **?**), Problem 5). In practice, this would require a rather contrived source of noise.

Let us imagine for the moment that a theory predicts an unambiguous distribution of sky temperatures, s_i. Then the likelihood of observing the data given the theory is given by

$$L(d|t) = \frac{1}{(2\pi)^{\frac{n}{2}}}|\mathbf{C^N}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{d}^T(\mathbf{C}^N)^{-1}\mathbf{d}\right) \ , \tag{3}$$

where d^T is the transpose of d.

One can find a basis of pixel vectors, $d_\alpha = \sum_i f_{\alpha i} d_i$, under which the matrix $\mathbf{C}^N$ is diagonalized, in which case the likelihood is just the product of the Gaussian probability distributions of each of these independent variables. As a simple example, if the instrument noise were uncorrelated between pixels, $\mathbf{C}^N$ would be diagonal, and this would reduce to

$$L(d|t) = \frac{1}{(2\pi)^{\frac{n}{2}} \prod_i \sigma_i} \exp\left(-\frac{\chi^2}{2})\right) \ , \tag{4}$$

with the familiar χ^2 misfit statistic given by

$$\chi^2 = \sum_i (d_i - s_i)^2/\sigma_i^2 \ . \tag{5}$$

One can find a basis of pixel vectors under which the actual matrix $\mathbf{C}^N$ is diagonalized, in which case the likelihood is just the product of the Gaussian probability distributions of each of these independent variables.

The second type of information used to compute the likelihood function for a given theory is the set of sky temperature correlations predicted by that theory, $< s_i s_j >= \mathbf{C}^S$. Because we have subtracted off the average sky temperature, $< s_i >= 0$. Theories of structure formation predict the angular power spectrum of CMB anisotropies, C^ℓ. From this, one can determine $\mathbf{C}^S$ solely as a function of the angular separation from pixel i to pixel j and the window function of the instrument, W_ℓ, that accounts for the beam and pixel convolutions discussed above and the observing strategy used to take the data.

$$C^S_{ij} =< s_i s_j >= \frac{1}{4\pi} \sum_\ell (2\ell+1) C_\ell W_\ell P_\ell(\theta_{ij}) \ , \tag{6}$$

where P_ℓ are the legendre polynomials. For theories of cosmological structure formation in which the density perturbations are Gaussian distributed, such as those produced by quantum fluctuations during an inflationary phase in the early universe, the theory is fully described by this correlation matrix, and higher-order statistics can be predicted as a function of $\mathbf{C}^S$ alone.

Assumption 2: The sky temperatures obey a multivariate Gaussian distribution given by $< s_i >= 0$ and the correlation matrix C^S. This is true for most theories of structure formation and can be tested by comparison with the distribution of pixel temperatures, the 3-point correlation function, the bispectrum, and other statistical tests for non-Gaussianity.

Assumption 3: The combination of pixel convolution, beam convolution, and observing strategy are azimuthally symmetric and can be represented as a window function W_ℓ. This assumption is never perfect as the pixel shapes lack azimuthal symmetry but for nearly symmetric, oversampled beams this will be an unimportant effect. The effect of observing strategy and covering only part of the sky are worth investigating in more detail. Partial sky coverage leads to a coupling of various spherical harmonics over a range $\Delta\ell \propto 1/f_{sky}$. This coupling is not relevant for our calculation of the full dataset s_i but will become important when we talk about bandpower estimates below.

We now have the second set of information needed to calculate the likelihood function. Let us imagine an instrument with zero instrument noise.[1] Then $\mathbf{C}^N = 0$ and $s_i = d_i$ so we can calculate

$$L(d|t) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{C}^S|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{d}^T (\mathbf{C}^S)^{-1}\mathbf{d}\right) . \tag{7}$$

[1]Theorists consider this a reasonable concept.

4. The Likelihood Function

The general calculation of the likelihood function occurs when the theory predicts only $< s_i >= 0$ and $\mathbf{C}^S$ and the instrument has noise given by $< n_i >= 0$ and $\mathbf{C}^N$. The standard calculation of the likelihood function notes that these processes are independent and that, given the assumptions made above, both are described by only their two-point functions. This allows the correlation matrix of the data to be calculated:

$$C^D_{ij} =< d_i d_j >=< (s_i{+}n_i)(s_j{+}n_j) >=< s_i s_j > + < n_i n_j > + < s_i n_j > + < n_i s_j >= C^S_{ij}{+}C^N_{ij} \ , \tag{8}$$

since the noise has mean zero and is uncorrelated with the temperature fluctuations on the sky.

Assumption 4: The instrument noise is uncorrelated with the temperature fluctuations on the sky. This follows directly from the definition of noise as the part of a measurement that fluctuates with time such that in the limit of infinite measurements the average will approach the signal. Because the pattern of CMB anisotropies on the sky is changing incredibly slowly, this applies. Moreover, even the variance of the noise $\mathbf{C}^S$ is uncorrelated with the signal. This is quite different from the experience of optical and high-energy astrophysics where Poissonian fluctuations in the number of photons coming from an object are a primary source of noise, but it is quite reasonable for CMB anisotropy observations, where there are so many photons being detected that Poissonian fluctuations are unimportant.[2]

It is now straightforward to calculate the likelihood of the data for a given theory, as above,

$$L(d|t) = \int_{\mathbf{s}} \int_{\mathbf{n}} p(\mathbf{s}|t) p(\mathbf{n}) \delta(\mathbf{d} - \mathbf{s} - \mathbf{n})$$
$$= \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{C}^D|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{d}^T (\mathbf{C}^D)^{-1} \mathbf{d}\right) , \tag{9}$$

given the typical assumption that the expected distribution of the data is fully described by $< d_i >= 0$ and C^D.

Deduction: The data obeys a multivariate Gaussian distribution. Although this assumption appears reasonable given that both processes discussed above can be assumed to follow a multivariate Gaussian distribution, it may not be immediately clear that the convolution of one such process with another produces a net Gaussian distribution. While the convolution theorem behaves differently on the sphere, we can take advantage of its $\Re^n$ version to evaluate the likelihood mathematically before moving back to evaluating this function on the sphere. The realization of the sky from a theory is diagonal in a different basis from that in which the instrument noise is diagonal, but we will show in

---

[2]Even if Poissonian fluctuations were important, $\Delta T/T = 10^{-5}$ means that Poissonian fluctuations in T would completely dominate those in ΔT. A quick back-of-envelope says that 400 CMB photons per cm^3 times an average speed of $c/\sqrt{3}$ yields about $10^{13}/cm^2/s$ in a detector. Since each detector covers only a portion of the CMB spectrum, avoiding the Poissonian regime requires capturing larger than a cm^2 of area in a one second observation.

Appendix A that the sum of two multivariate Gaussian distributions is indeed a multivariate Gaussian distribution, i.e. that our deduction is exact given the previous assumptions.

5. Adding Beam and Calibration Uncertainties

5.1. Calibration Uncertainty

Generally, CMB anisotropy observations have a calibration uncertainty where the reported temperature fluctuations in uK are really in units of $c\,uK$ where c is the calibration factor with $<c>=1$ and an uncertainty given by σ_c.[3]

While our noise covariance matrix has been estimated from the data and is therefore already in units of $c\,uK$, our theories still predict true temperature fluctuations in uK. So, for a given c, $\mathbf{C}^N$ is fixed but the covariance matrix of observed signal is really given by $c^2\mathbf{C}^S$, i.e. $\mathbf{C}^D = c^2\mathbf{C}^S + \mathbf{C}^N$. We know how to write out the full likelihood function at this point, acknowledging that the

$$L(d|t) = \int_c \int_{\mathbf{s}} \int_{\mathbf{n}} p(\mathbf{s}|t)p(\mathbf{n})p(c)\delta(\mathbf{d} - c\mathbf{s} - \mathbf{n})$$

$$= \int dc \frac{1}{(2\pi)^{\frac{n+1}{2}}|c^2\mathbf{C}^S + \mathbf{C}^N|^{1/2}\sigma_c} \exp\{-\frac{1}{2}\left(\mathbf{d}^T(c^2\mathbf{C}^S + \mathbf{C}^N)^{-1}\mathbf{d} + \frac{(c-1)^2}{\sigma_c^2}\right)\} \,. \qquad (10)$$

---

[3]Theorists love to model this uncertainty as Gaussian even though observers will typically claim that its distribution is either uncertain or closer to a top-hat. Since c must be positive a lognormal distribution might be a better approximation, although c is usually many nominal σ away from 0 so this is not a major issue.

Even for the assumed Gaussian distribution in c, this integral is not readily evaluated analytically. It can be evaluated numerically by sampling at a discrete set of c and interpolating, in which case we have marginalized over the nuisance parameter c (see Section 6.1 below). Because c obeys a known distribution, a Frequentist analysis can still perform this integral in the context of computing $p(d|t) = \int_c p(d|t,c)p(c)$. One could argue that the most Frequentist approach is to maximize over c but this ignores the known information about the distribution of c. The misfit statistic from the exponent in the integral of Equation 10 is a more reasonable way to determine the "maximum-likelihood" value of c:

$$\chi_c^2 = \mathbf{d}^T(c^2\mathbf{C}^S + \mathbf{C}^N)^{-1}\mathbf{d} + \frac{(c-1)^2}{\sigma_c^2}\,. \tag{11}$$

Notice that the multiple dependences on c means that this statistic will not obey a χ^2 distribution. We could however simulate how likely it is to get a value of χ_c^2 larger than that observed for a given theory; see Appendix B for a fuller discussion of misfit statistics applied to non-Gaussian likelihoods.

5.2. Beam Uncertainty

The beamsize of an instrument is typically estimated from the data and described by standard deviation b^* of the best-fit Gaussian beam, with uncertainty σ_b and the corresponding window function $W_\ell = \exp[-\ell(\ell+1)b^2]$. Equation 6 illustrated how $\mathbf{C}^S$ is a function of cosmological parameters that determine C_ℓ and this beamwidth b. So we can always compute $\mathbf{C}^S(b)$ for a given theory and a given value of b. Hence our full likelihood is given by

$$L(d|t) = \int_c \int_b \int_{\mathbf{s}} \int_{\mathbf{n}} p(\mathbf{s}|t)p(\mathbf{n})p(b)p(c)\delta(\mathbf{d} - c\mathbf{s}(\mathbf{b},\mathbf{t}) - \mathbf{n})$$

$$= \int dc \int db \frac{1}{(2\pi)^{\frac{n+2}{2}} |c^2\mathbf{C}^S(b,t) + \mathbf{C}^N|^{\frac{1}{2}} \sigma_c \sigma_b} \exp\{-\frac{1}{2}\left(\mathbf{d}^T(c^2\mathbf{C}^S(b,t) + \mathbf{C}^N)^{-1}\mathbf{d} + \frac{(c-1)^2}{\sigma_c^2} + \frac{(b-b^*)^2}{\sigma_b^2}\right)\} ,$$

where we have written explicitly that the signal in true uK units is a function of both our theory t and our actual beamwidth b. Non-symmetric beamshapes have been determined for some instruments, and this adds yet another level of complexity to the likelihood function.

6. Approximations to the Likelihood Function

6.1. Maximizing versus Marginalizing over Nuisance Parameters

Let u be a nuisance parameter that we are not interested in estimating from the data, either with a known distribution like calibration or beam uncertainty or with an uncertain prior such as Hubble's constant. Then

$$p(t|d) = \int du p(d|u,t)p(u)p(t) \tag{13}$$

and we "marginalize" over the nuisance parameter u by integrating. If u obeys a known distribution, the Frequentist approach allows this integration as well, with the result that $p(d|t) = \int du p(d|u,t)p(u)$. For an uncertain distribution of u, the typical Frequentist approach is to maximize, that is find the value $u^*(t)$ which maximizes $p(d|u,t)$ for the observed data. This ignores $p(u)$ which is quite comfortable for a Frequentist but rather strange for a Bayesian analysis. Some recent Bayesian analyses (Melchiorri et al., Balbi et al., Tegmark & Zaldarriaga [check], Dodelson & Knox [check]) have made use of maximization to avoid the computationally-intensive task of integrating the likelihood over all possible values of the nuisance parameters for each prior distribution considered. Maximization corresponds to **retroactively assuming a delta-function prior** in the nuisance parameters at their optimal values for the observed data:

$$\int du p(d|u,t)\delta(u-u^*(t)) = p(d|u^*(t),t) \ . \tag{14}$$

These authors defend their assumption of a delta-function prior by arguing that in the approximation where the likelihood function is Gaussian in u,

$$\int du p(d|u^*,t) \exp\left(-\frac{(u-u^*(t))^2}{2\sigma_u^2}\right) = p(d|u^*(t),t) \ , \tag{15}$$

i.e. that maximization equals marginalization under this assumption. However, this requires that u is uncorrelated with the other theory parameters t i.e. that $\partial^2 L/\partial u \partial t = \delta_{ut}$ which is almost never correct![4] For a known distribution $p(u)$ the error induced by this approximation is calculable but not offered in these papers. For an uncertain prior $p(u)$ this approximation can only be correct for one possible choice of $p(u)$ which may not even be "reasonable."

The nuisance parameter may have a known distribution such as calibration or beamwidth uncertainty or it may not. Although the Bayesian approach assigns a prior

---

[4]The exception is the case where both the amplitude of fluctuations and the calibration uncertainty are considered nuisance parameters as they are completely degenerate and uncorrelated with all other parameters.

distribution to all parameters anyway, this prior need not be varied for parameters with known distributions. [Still simulating this for toy model.] Impressive difference for BOOMERANG of about 0.05 in best-fit Ω_{tot}, with $L(\Omega_{tot} = 1) = 0.8$ for maximizing and 0.2 for marginalizing. Their database has discrete steps of 0.05 in Ω_{tot} so you have to wonder whether to make much of such differences but people do.

6.2. Dangers of Assuming a Gaussian Shape to Set Credible Regions

We now discuss three dangers of assuming that 68% (or 95%) of area is found down to $L/L_{max} = \exp(-1/2)$ (or $\exp(-2)$) as would be the case for a one-dimensional Gaussian. The same ratios are calculable for an n-dimensional Gaussian assumption for the likelihood function of n parameters. The safer approach is to integrate the likelihood function down to the contour at which the desired percentage of the area is found! This is only relevant for calculating Bayesian credible regions, in which case it is $p(t|d) = p(d|t)p(t)$ that should be integrated, not $p(d|t)$!

1. Best-fit L_{max} flucutuates more than an integral does. My simplest toy model does not allow this effect to be simulated as the distribution is always Gaussian, but for more complex datasets this will not be the case.

2. For paradigms in which the parameters do not span the space of all possible datasets, n parameters may only use up m effective degrees of freedom in the data, altering the expected shape of the n-dimensional likelihood function away from Gaussian. My first

toy model shows that this can create both a bias and a significant variance in the actual fraction of the volume covered out to the assumed contour: a mean volume of only 65% and a standard deviation of 11%, making this a poor estimator of where 68% of the volume cuts off.

Surprisingly, this sort of maximum-likelihood at the boundary issue is precisely what Lange et al. [astro-ph/0005004] cite for the case where they switch from integrating the volume under the likelihood to checking where it drops off by e.g. $e^{-1/2}$ in determining credible regions - but our toy model shows that such boundary-maxima are a good example of where the L/L_{max} ratio fails! The real issue here is whether the boundary is physical - if the parameter really can't go past the boundary i.e. the prior probability is zero, then integrating the volume is still safest. But if the boundary is due to not considering a wide enough range of parameter space then the true maximum may not have been found and the only good solution is to expand the range of parameters considered!

Notice that the Frequentist approach is not sensitive to this problem. For n_d datapoints that obey a multivariate Gaussian distribution, $p(d|t)$ obeys a $\chi^2_{n_d}$ distribution for a given **theory**. Fitting n parameters leads to a $\chi^2_{n_d}$ distribution in the expected value of $-2\log L_{max}$ and a χ^2_m distribution in the expected value of $\log(L/L_{max})$, where $m \leq n$ and the equality only holds if the parameters span the space of all possible datasets. However, the Frequentist approach is not concerned with L_{max} but rather tests each theory versus its expected $\chi^2_{n_d}$ distribution of $p(d|t)$.

3. Varying the prior is guaranteed to vary the shape that you're integrating over! For a Gaussian likelihood only a uniform prior in all parameters will preserve this shape. In my toy model, this is a very small effect, however.

6.3. How Should I Set the Credible Region?

While the method described above has numerous pitfalls, there is no clear perfect way to determine **which** 68% of the volume under the likelihood function is what you want. A median-related statistic would integrate over a 1-dimensional likelihood and "rule out" the leftmost 16% and rightmost 16% at 1σ confidence. This method is used by Lange et al. However, in higher dimensions this is an ill-defined concept. One can start from the mode (maximum likelihood) point and move outwards symmetrically until 68% of the total volume has been encompassed. This is similar to finding the smallest continous interval in the various parameters which contains 68% of the volume. All of these methods run the risk of ignoring a distant but highly peaked secondary maximum of the likelihood function, so we advocate determining the threshold value of the likelihood for which the volume under it comprises 68% of the data, **even if this region is not simply connected**. This will most accurately tell us what range of theories are most credible. Note that a Frequentist analysis will naturally allow a non-simply-connected region of "parameter space" to survive confrontation with the data.

6.4. Data Compression into Bandpower Estimates

Intermediate theory is determined to explain the data, $t = C_B$ for a set of bandpowers B ranging from ℓ_B^{min} to ℓ_B^{max}. Data is used to determine C_B^* and $\mathbf{C}^{BB'}$, the maximum likelihood estimate of the bandpower values and their covariance matrix respectively. This covariance can either be estimated empirically as the data is fit or be calculated from the known coupling of various ℓ modes from one band into another given the sky coverage of the instrument. So the likelihood function has been compressed and we can assume that for a given set of C_B,

$$L(d|t) = \frac{1}{(2\pi)^{\frac{n_B}{2}}|\mathbf{C}^{BB'}|^{\frac{1}{2}}} \exp -\frac{1}{2} C_B^T (\mathbf{C}^{BB'})^{-1} C_B \; . \tag{16}$$

For each theory considered we can calculate what values of C_B correspond and evaluate its likelihood in this manner. Note that many different theories will predict a particular set of C_B. Given beamsize uncertainty and the increased coupling of ℓ modes near the edges of neighboring bins, it is only an approximation that all theories which predict a given C_B are equally likely. Indeed, the correlation matrix $C^{BB'}$ is really a function of the shape of the C_ℓ spectrum of a given theory in each bin. It is necessary to see how much error is caused by this approximation of bandpowers, although the computational resources required to analyze the full datasets make bandpower compression a highly attractive option for all but the most precise analyses.

[discuss $\ell\ell'$ coupling]

6.5. The Offset Lognormal Approximation

Bond, Jaffe, and Knox [astro-ph/9808264] basically assume that people will do a χ^2 analysis on the reported bandpowers. You can calculate the curvature matrix of the theoretically expected likelihood evaluated at the maximum likelihood point, which is a linear function of $C_\ell + N_\ell/B_\ell^2$, where This means that the curvature of $L(d|C_\ell)$ is changing with C_ℓ, although one could still evaluate that at C_ℓ^*. But if we switch variables to $Z_\ell = \log(C_\ell + N_\ell/B_\ell^2)$ then the curvature will be constant as a function of Z_ℓ i.e. to second order in a Taylor expansion about the maximum of the likelihood function the likelihood function is Gaussian in Z_ℓ. Then if we're going to do a χ^2 analysis Z_ℓ is the best variable to use! Of course, once you're willing to go through this much trouble, you should probably go ahead and use a better approximation to the likelihood function and not do a χ^2 analysis. A Frequentist analysis is still possible using the sort of non-Gaussian misfit statistic discussed above and in Appendix B.

7. Conclusions

Eventually I will conclude something...

Acknowledgments

I want to thank Ted Bunn and Sevil Salur for helpful conversations. I thank the Institute for Theoretical Physics at U.C. Santa Barbara for their hospitality; this research was supported in part by the National Science Foundation under Grant No. PHY94-07194.

8. Appendix A: The Sum of Two Multivariate Gaussians is a Multivariate Gaussian

The appendices are still under construction. Rough outline:

Integrals aren't fully down to negative infinity! $\Delta T > -2.73K$ That's why lognormal dist. for LSS but not an issue here at $\Delta T/T \simeq 10^{-5}$.

Proof 1: Matrix completion of the square

Proof 2: Characteristic Functions (which are just Fourier transforms) Still uses matrix completion of the square but is simpler, less fancy linear algebra needed for the final step. Convolution is just multiplication of characteristic functions.

9. Appendix B: A Misfit Statistic for Non-Gaussian Likelihood Functions