Cosmological Microwave Background Radiation

Likelihood Analysis - Part 1

Jayanti Prasad

Inter-University Centre for Astronomy & Astrophysics (IUCAA)
Pune, India (411007)

September 12, 2011

Plan of the talk

- CMBR Likelihood
- Gaussian Likelihood
- Optimal Quadratic Estimator
- WMAP Likelihood

CMBR Likelihood

- CMBR temperature and polarization observations can constrain cosmological parameters if the likelihood function can be computed exactly.
- Computing the likelihood function exactly in a brute force way is computationally challenging since it involves inversion of the covariance matrix i.e., $O(N^3)$ computation.

$$\mathcal{L} = \frac{1}{(2\pi)^{N/2}\sqrt{|C|}} \exp\left[-d^t C^{-1} d\right], \tag{1}$$

  where d is the data vector and C is the covariance matrix.
- Temperature and polarizations fields are correlated and partial sky coverage correlates power at different ls.
- Since the exact computation of likelihood function is challenging, approximations are generally made (Gaussian Likelihood, Gibbs sampling etc).

Gaussian Likelihood

- The likelihood function for the temperature fluctuations observed by a noiseless experiment with full-sky coverage has the form:

$$L(T|C_l) \propto \frac{1}{\sqrt{|S|}} \exp[-(TS^{-1}T)/2] \tag{2}$$

or

$$L(T|C_l) = \prod_{lm} \frac{1}{\sqrt{C_l}} \exp[-|a_{lm}|^2/(2C_l)] \tag{3}$$

where

$$T(\hat{n}) = \sum_{lm} a_{lm} Y_{lm}; \quad \text{and } S_{ij} = \sum_{l} ((2l+1)/(4\pi)) C_l P_l(\hat{n}_i . \hat{n}_j) \tag{4}$$

- Since observe only one sky, we cannot measure the power spectra directly, but instead form the rotationally invariant estimators, C_l^{XY}, for full-sky CMB maps given by

$$\hat{C}_l^{XY} = (1/(2l+1)) \sum_{m=-l}^{m=l} a_{lm}^X a_{lm}^Y \tag{5}$$

where $< \hat{C}_l^{XY} >= C_l^{XY}$ (true power spectrum). Note that the likelihood has a maximum when $C_l = \hat{C}_l$ so $\hat{C}_l$ is the MLE.

[Hamimeche & Lewis (2008)]

- Since a_{lm} are assumed to be Gaussian and statistically isotropic, they have independent distributions (for $|m| \geq 0$). The probability of a set of a_{lm} at a given l is given by:

$$
\begin{aligned}
-2 ln P(a_{lm}|C_l) &= \sum_{m=-l}^{m=l} [a_{lm}^{'} C_l^{-1} a_{lm} + ln|2\pi C_l|] \\
&= (2l+1)[Tr(\hat{C}_l C_l^{-1}) + ln|C_l|] + const. \qquad (6)
\end{aligned}
$$

- The fact that this likelihood for C_l depends only on the $\hat{C}_l^{XY}$ shows that on the full sky the CMB data can losslessly be compressed to a set of power spectrum estimators, that contain all the relevant information about the posterior distribution i.e., $\hat{C}_l$ is a sufficient statistic for the likelihood.

- If the likelihood of the theory power spectrum C_l as a function of the measured estimators $\hat{C}_l$ were Gaussian, the likelihood could be calculated straightforwardly from the measured $\hat{C}_l$.
- However the distribution is non-Gaussian because for a given temperature power spectrum C_l, the $\hat{C}_l$ is a sum of squares of Gaussian harmonic coefficients, have a (reduced) χ^2 distribution. This makes it difficult to use "χ^2 per degree of freedom" as "goodness of fit".

Maximum Likelihood Estimation

$$L = \frac{1}{(2\pi)^{N/2}\sqrt{|C|}} \exp\left[-(x-\mu)C^{-1}(x-\mu)^t\right] \tag{7}$$

or

$$2\log L = 2\mathcal{L} = ln|C| + (x-\mu)C^{-1}(x-\mu)^t = Tr[lnC + C^{-1}D] \tag{8}$$

where the covariance matrix C and data matrices D are given by

$$C = \left\langle (x-\mu)(x-\mu)^t \right\rangle \quad \text{and} \quad D = (x-\mu)(x-\mu)^t \tag{9}$$

differentiating equation (8) with respect to to θ_i

$$2\mathcal{L}_{,i} = Tr\left[C^{-1}C_{,i} - C^{-1}C_{,i}C^{-1}D + C^{-1}D_{,i}\right] \tag{10}$$

note that

$$, = \frac{\partial}{\partial\theta_i}; \quad \frac{\partial}{\partial\theta_i}(C^{-1}) = -C^{-1}\frac{\partial C}{\partial\theta_i}C^{-1} \quad \text{and} \quad \frac{\partial}{\partial\theta_i}(lnC) = C^{-1}\frac{\partial C}{\partial\theta_i} \tag{11}$$

[Bond et al. (1998); Tegmark et al. (1997)]

Fisher Information Matrix

- $$F_{ij} = \langle \mathcal{L}_{,ij} \rangle = \frac{1}{2} Tr \left(A_i A_j + C^{-1} M_{ij} \right), \tag{12}$$

  where
  $$A_i = (lnC)_{,i} \quad \text{and} \quad M_{ij} =< D_{,ij} > \tag{13}$$
  for isotropic CMBR fluctuations $\mu = 0$ and
  $$C_{ij} = \delta_{ij} \left[C_l + \frac{4\pi\sigma^2}{N} e^{\theta_b^2 l(l+1)} \right] \tag{14}$$
  where σ is the RMS pixel noise and $\theta_b = 0.425 FWHM$.
- From the above equations

$$F_{ij} = \sum_{l=2}^{l=l_{max}} \left(\frac{2l+1}{2} \right) \left[C_l + \frac{4\pi\sigma^2}{N} e^{\theta_b^2 l(l+1)} \right]^{-2} \left(\frac{\partial C_l}{\partial \theta_i} \right) \left(\frac{\partial C_l}{\partial \theta_j} \right) \tag{15}$$

Quadratic Estimator

- The best fit model i.e., parameter set $\bar{\theta}$, corresponds to a point in the parameter space at which

$$\frac{\partial \mathcal{L}}{\partial \theta}|_{\theta=\bar{\theta}} = 0 \tag{16}$$

We can use a root finding method to obtain $\bar{\theta}$ in the following way:

- We begin with some best guess value $\theta^{(0)}$ and approximate the derived to the first order

$$\frac{\partial \mathcal{L}(\bar{\theta})}{\partial \theta_i} = \frac{\partial \mathcal{L}(\theta^{(0)})}{\partial \theta_i} + \left(\bar{\theta}_j - \theta_j^{(0)}\right) \frac{\partial^2 \mathcal{L}(\theta^{(0)})}{\partial \theta_i \partial \theta_j} \tag{17}$$

setting the LHS zero

$$\left(\bar{\theta}_j - \theta_j^{(0)}\right) \approx - \left[\frac{\partial^2 \mathcal{L}(\theta^{(0)})}{\partial \theta_i \partial \theta_j}\right]^{-1} \frac{\partial \mathcal{L}(\theta^{(0)})}{\partial \theta_i} \delta\theta \tag{18}$$

for the next step we can write

$$\theta^{(1)} = \theta^{(0)} + \delta\theta \tag{19}$$

and keep iterating till we get convergence.

[Bond et al. (1998); Durrer (2008)]

- $$\frac{\partial \mathcal{L}}{\partial \theta_i} = \frac{1}{2}\left[d^t C^{-1} C_{,i} C^{-1} d - Trace\left(C^{-1} C_{,i}\right)\right] \tag{20}$$

  and

  $$F_{ij} = \left\langle \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_j} \right\rangle = \frac{1}{2} Trace\left[C^{-1} C_{,i} C^{-1} C_{,i}\right] \tag{21}$$

- $$\theta^{(1)} = \theta^{(0)} + \frac{1}{2} F_{ij}\left[d^t C^{-1} C_{,i} C^{-1} d - Trace\left(C^{-1} C_{,i}\right)\right] \tag{22}$$

- The advantage of this method is that for a given starting parameter the Fisher matrix is readily calculated from the covariance matrix alone, without having to know the data.
- The above estimated parameters $\theta^{(1)}$ depend quadratically on the data vector d so they are called a quadratic estimators.
- This method is faster than the grid based methods, however, there are problems. This method is Levenberg-Marquardt method as is implemented in GSL.

- *What if we end up in a shallow local maximum of the likelihood function and are stuck there?*
  To avoid this problem, one usually adds a small random fluctuation to the obtained $\delta\lambda$, a 'temperature', so that one can leave a shallow local maximum.
- *What if there are several local maxima, some of them quite steep and separated by deep ridges?*
  To check this, one performs not only one but many iteration chains with different starting points. One can then compare the height of the different maxima. A procedure along these lines is the Markov chain Monte Carlo method (MCMC) discussed below. It is presently the method of choice for CMB analysis. A publicly available MCMC code and more details of the method can be found in the paper by Lewis & Bridle (2002).
- *What if the maximum is somewhere at the border of parameter space?*
  The border of the parameter space is given by the prior. If the data are best fitted by parameter values lying at the border of what is allowed by the prior, this hints that either the prior is wrong or the model is incorrect. This is one of the most important drawbacks of the Fisher matrix technique. It provides relatively rapidly the best-fitting model under consideration, but it works independently of whether this model is actually a good fit to the data or not. For this an evaluation of the likelihood function at the best-fitting parameter values has to be performed. If the likelihood function is very small, this is either a sign that the model is wrong or that the real errors are much larger than those assumed.

Error Estimate

- A good first estimate for 1σ error bars are the diagonal elements of the Fisher matrix at the maximum, i.e., the best fit point.

$$\langle \delta\theta_i \delta\theta_j \rangle = F_{ij}^{-1} \tag{23}$$

- These are the true 1σ errors only if the distribution is Gaussian in the parameters θ which usually it is not.
- The Fisher matrix defines an ellipse around $\bar{\theta}$ in the parameter space via the equation $\delta\theta^t F(\bar{\theta})\delta\theta = 1$
- The principal directions of this error ellipse are parallel to the eigenvectors of F and their half length is given by the square root of the eigenvalues of F
- The total width of the error ellipse in a given direction j at $\bar{\theta}$ is $2/\sqrt{Fjj}$ and $1/\sqrt{Fjj}$ is the error of the parameter j if all other parameters are known and are equal to $\bar{\theta}$.
- In practice we do not know the other parameters any better than j so the true error in j is given by the size of the projection of the error matrix onto the j axis.

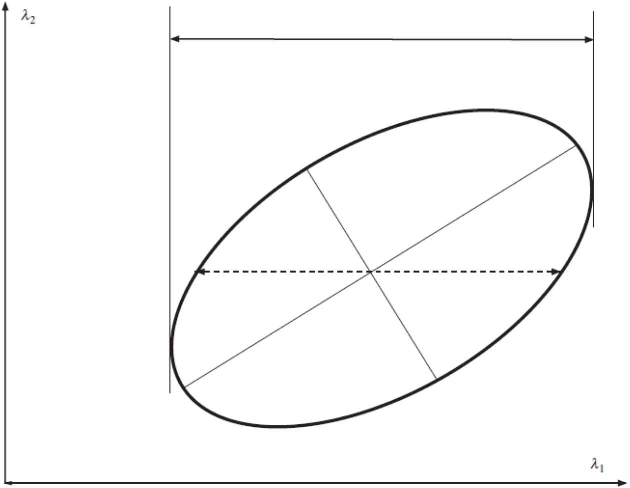

Fig. 6.8. The error ellipse is shown in a two-dimensional example. The widths $2/\sqrt{F_{11}}$ (dashed double arrow) and $2\sqrt{F_{11}^{-1}}$ (solid double arrow) are indicated.

WMAP Likelihood

- At large l likelihood is approximated as Gaussian

$$2lnL_{Gauss} \propto \sum_{ll'} (C_l - \hat{C}_l) Q_{ll'} (C_l - \hat{C}_l) \tag{24}$$

where $Q_{ll'}$, the curvature matrix, is the inverse of the power spectrum covariance matrix.

- The power spectrum covariance encodes the uncertainties in the power spectrum due to cosmic variance, detector noise, point sources, the sky cut, and systematic errors.
- Since the likelihood function for the power spectrum is slightly non-Gaussian, equation (24) is a systematically biased estimator so a log normal function is suggested

$$2lnL_{LN} \propto \sum_{ll'} (z_l - \hat{z}_l) D_{ll'} (z_l - \hat{z}_l) \tag{25}$$

where $z_l = (C_l + N_l)$ and $D_{ll'} = (C_l + N_l) Q_{ll'} (C_l + N_l)$

- It has been argued that the following estimator is more accurate then the the Gaussian or log normal only [Verde et al. (2003)]

$$L = (1/3) L_{Gauss} + (2/3) L_{LN} \tag{26}$$

WMAP seven years Likelihood

- For low-l TT ($l \leq 32$) the likelihood of a model is computed directly from the 7-year Internal Linear Combination (ILC) maps.
- For high-l ($l > 32$) TT MASTER pseudo-C_l quadratic estimator is used.
- For low-l polarization, ($l < 23$) TE, EE, & BB, pixel-space estimator is used.
- For high-l TE ($l > 23$) MASTER quadratic estimator is being used.
- Given a best-fit model one asks how well the model fits the data. Given that the likelihood function is non-Gaussian, answering the question is not as straightforward as testing the χ^2 per dof of the best-fit model. Instead one resorts to Monte Carlo simulations and compare the absolute likelihood obtained from fitting the flight data to an ensemble of simulated values.

[Larson et al. (2011)]

Summary & Discussion

- I have discussed the difficulties and approximations used to compute the CMBR likelihood function.
- My aim was to understand why and how the likelihood function is computed differently for different values of l and different type of power spectra.
- I have also discussed various estimators related to angular power spectrum and the likelihood function.
- I was not able to discuss the actual computation of the various components of the likelihood function in the WMAP likelihood code.
- It has been suggested (by Dodelson) that the Nelder Mead method or downhill simplex method (commonly known amoeba) can be be used for finding the best fit parameters.
- Another method called NEWUOA which is used for unconstrained optimization without derivatives, is also suggested (by Antony Lewis) to find the best fit parameters, optimization of the likelihood function.

References

Bond, J. R., Jaffe, A. H., & Knox, L. 1998, Phys. Rev. D , 57, 2117

Dodelson, S. 2003, Modern cosmology (San Diego, U.S.A.: Academic Press)

Durrer, R. 2008, The Cosmic Microwave Background (Cambridge University Press)

Hamimeche, S., & Lewis, A. 2008, Phys. Rev. D , 77, 103013

Larson, D., et al. 2011, Astrophys. J. Suppl. Ser. , 192, 16

Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, Astrophys. J. , 480, 22

Verde, L., et al. 2003, Astrophys. J. Suppl. Ser. , 148, 195