Sample Variance of Non-Gaussian Sky Distributions

Submitted to ApJ

Xiaochun Luo

Center for Particle Astrophysics

University of California, Berkeley, CA 94720

Sample Variance of Non-Gaussian Sky Distributions

Submitted to ApJ
Recently, several groups have reported results on the measurement of anisotropies of cosmic microwave background (CMB) at degree scales (de Bernardis et al. 1994; Cheng et al. 1993; Gunderson et al. 1993; Meinhold et al. 1993; Schuster et al. 1993; Tucker et al. 1993; Wollack et al. 1993). The beam size, beam throw, the most sensitive angular scale, sky coverage and quoted \( \text{rms} \) temperature anisotropies are summarized in Table 1. The results from these experiments do not agree with each other, especially the results from the same experiment MAX at two different part of the sky, Gamma Ursa Minor (GUM) region and mu-Pegasi (MuP) region, contradict at 2\( \sigma \) level. A way to reconcile these measurements is a non-Gaussian distribution of temperature anisotropies at half-degree scales (Coulson et al. 1994). At present, there are still large uncertainties in all experiments due to foreground subtractions, therefore the need to invoke non-Gaussian temperature distributions remains to be seen. Since different experiment probes different part of the sky and the sky coverage of each experiment is small, sample variance of each experiment can be large, especially when the sky signal is non-Gaussian. The goal of this paper is to quantify the difference in the expected sample variance between non-Gaussian and Gaussian fields in order to determine if this effect could be responsible for the discrepancy between experiments. Furthermore, we also estimate the minimum sample size for each experiment in order for the sample variance to be less than 20\( \mu \text{K} \), both for Gaussian and non-Gaussian distributions.

Regarding the statistical analysis, the most relevant quantity is the number of independent measurements in a single sample. However, the sample data points in CBR experiments are correlated and therefore contained less statistical information. Thus, it is useful to determine the effective number of data point, \( N_e \), defined as the number of independent measurements of temperature anisotropies in each experiment (we will discuss how to estimate \( N_e \) for each experiment later on). Expressed in terms of \( N_e \), the sample-averaged \( \text{rms} \) temperature
anisotropy is
\[
\left( \frac{\delta T}{T} \right)_{sam}^2 = \Delta = \frac{1}{N_e} \sum_{i=1}^{N_e} \left( \frac{\Delta T_i}{T} \right)^2,
\]
where \( \Delta T_i/T \) is an independent measurement of temperature anisotropy at the angular scale of the beam size. The ensemble average of \( \Delta T_i/T \) vanishes and the ensemble average of \( \Delta \) is \( \sigma_{th}^2 \), where \( \sigma_{th} \) is the theoretical prediction of temperature anisotropy at the angular scale probed by the experiment.

Due to the stochastic nature of \( \frac{\Delta T}{T} \), \( \Delta \) will still be a random quantity which can vary between different experiments. The sample variance is the fluctuation around \( \sigma_{th}^2 \), the ensemble-average of \( \Delta \). In the case where \( \frac{\Delta T}{T} \) is Gaussian, the sample variance is given by
\[
\sigma_{sam}^2 = \langle \Delta^2 \rangle - \langle \Delta \rangle^2 = \frac{2}{N_e} \langle \Delta \rangle^2, \quad \langle \Delta \rangle = \sigma_{th}^2.
\]
Here the angle-brackets \( \langle ... \rangle \) denote an ensemble average. The sample variance will be reduced by increasing the sky coverage by an amount which will scale with the effective number of data points \( N_e \) as \( \sigma_{sam} \propto 1/\sqrt{N_e} \).

We would expect the sample variance to be larger when the temperature anisotropies are non-Gaussian distributed at degree scales. Since the functional space of non-Gaussian distributions is unlimited, one is faced with the question of choosing appropriate distributions to model the sky. As we stressed before (Luo 1994), a \( \chi_n^2 \) distribution with \( n \)-degrees of freedom is one of the simplest and most natural choices. By varying \( n \), it provides a family of distributions that range from highly non-Gaussian (small \( n \)) to nearly Gaussian (large \( n \)). Furthermore, it provides a good fit to the statistics of temperature fluctuation from global topological defects and non-topological defects in the framework of the \( O(N) \) \( \sigma \)-model where a global symmetry \( O(N) \) is broken to \( O(N-1) \) by an \( N \)-component real scalar field \( \phi = (\phi_1, \ldots, \phi_N) \) in the early universe (Turok & Spergel 1991). In this paper, we will
model the degree scale CBR sky as \( \chi^2_n \) field, with

\[
\frac{\Delta T}{T} = \sum_{j=1}^{n} (\delta_i^2 - \sigma^2),
\]

(3)

where \( \delta_i, i = 1...n \), are \( n \)-independent Gaussian variables with zero mean and variance \( \sigma^2 \).

The ensemble average of \( \Delta \) is related to \( \sigma \) through \( \langle \Delta \rangle = 2n\sigma^4 \). To calculate the sample variance of the \( r m s \) temperature anisotropy of a \( \chi^2 \) field, we will utilise some of the higher moments of a Gaussian, i.e.,

\[
\langle \delta^4 \rangle = 3\langle \delta^2 \rangle^2, \quad \langle \delta^6 \rangle = 15\langle \delta^2 \rangle^3, \quad \langle \delta^8 \rangle = 105\langle \delta^2 \rangle^4.
\]

(4)

and the following identities for Gaussian variables:

\[
\langle \sum_{i,j=1}^{n} \delta_i^2 \delta_j^2 \rangle = n(n+2)\sigma^4,
\]

(5)

\[
\langle \sum_{i,j,k=1}^{n} \delta_i^2 \delta_j^2 \delta_k^2 \rangle = n(n+2)(n+4)\sigma^6,
\]

(6)

and

\[
\langle \sum_{i,j,k,l=1}^{n} \delta_i^2 \delta_j^2 \delta_k^2 \delta_l^2 \rangle = n(n+2)(n+4)(n+6)\sigma^8.
\]

(7)

After some algebra we have

\[
\sigma^2_{sam} = \left( \frac{2}{N_e} \right) \cdot \frac{n+6}{n} \langle (\frac{\Delta T}{T})_{sam} \rangle^2
\]

(8)

This is the main result of the paper. For a \( \chi^2_n \) distribution, the sample variance is enhanced by a factor of \( (n+6)/n \) relative to a gaussian distribution. As we expected, as \( n \) becomes much larger than 6, the enhancement is negligible.

The effective number of data points, \( N_e \), depends on the detailed sampling scheme. If the experimental data is sparcely sampled, i.e. the distance between data points is much larger than the beam size, then \( N_e \) is approximately the number of experimental data points. If the data is over-sampled so that correlations between the data points are important, then \( N_e \) can
be estimated as follows. The sample variance, $\sigma_{\text{sam}}$, of a Gaussian-distributed temperature anisotropy can also be expressed as (Scott et al 1994):

$$\sigma_{\text{sam}} = \frac{2}{\Omega^2} \int d\Omega_1 d\Omega_2 C^2(\theta_1, \theta_2), \tag{9}$$

here $\Omega$ is the solid angle covered by the experiment and $C(\theta_1, \theta_2)$ is the two-point temperature correlation function. Combining the equations 9 and 2 gives an expression for $N_e$:

$$N_e = \frac{\Omega^2 C^2(0)}{\int d\Omega_1 d\Omega_2 C^2(\theta_1, \theta_2)}. \tag{10}$$

The data analysis of all experiments involves using the Gaussian auto-correlation function (GACF), where the correlation among the data points is approximated as

$$C(\theta) = C(0) \exp\left(-\frac{\theta^2}{2\theta_c^2}\right), \tag{11}$$

where $\theta$ is the angular distance between two data points and $\theta_c$ is the coherence angle. Since the two point correlation function depends only on the difference between two angular directions, after changing variable equation (10) can be reduced to

$$N_e = \Omega C^2(0) \int d\Omega C^2(\theta). \tag{12}$$

after changing variables. For a single one dimensional temperature scan, it is easy to evaluate $N_e$ since

$$\frac{1}{\Omega C^2(0)} \int d\Omega C^2(\theta) = \sqrt{\pi} \frac{\theta_m}{\theta_c}, \quad \text{for} \quad \theta_m \gg \theta_c, \tag{13}$$

where $\theta_m$ is the maximum angular difference among data points with respect to the observer at the origin. The solid angle covered by the experiment is $\Omega = \theta_m \times \theta_{FWHM}$, hence one can express $N_e$ in terms of $\Omega$ as

$$N_e = \frac{1}{\sqrt{\pi}} \left( \frac{\Omega}{\theta_{FWHM} \theta_c} \right). \tag{14}$$

Although the above formula is derived for a single one dimensional scan, it is also applicable to multiple scan experiments when the correlation among different scans is small. Estimates of $N_e$ for the different experiments are listed in Table 2.
The sample variance, $\sigma_{sam}$, is directly proportional to the theoretical prediction of the temperature anisotropy. The theoretical model that is used to calculate the numerical value of $\sigma_{sam}$ is the standard CDM ($\Omega = 1$) model, with a scale invariant primordial power spectrum. We take the Hubble constant as $h = 0.5$ and the baryonic fraction $\Omega_b = 0.06$, to be consistent with the big-bang nucleosynthesis bound (Walker et al. 1991). Theoretical values for the temperature anisotropy at degree scales for those experiments are taken from White et al. (1994).

A question of practical interest is how large a sample should be in order for the experimental result to be informative? One criteria could be that the sample variance is much smaller the the instrumental sensitivity in $\Delta T$. From equations 2 and 8, we estimate the minimum sample size $N_{\text{min}}$ that gives rise to a sample variance of $\Delta T$ be $20 \mu K$, both for Gaussian and $\chi^2$ distributions. Minimum sample size for current experiments are shown in Table 2. Sky coverages of ARGO (de Bernardis et al. 1994) and Saskatoon (Wollack et al. 1994) is already large enough so that the sample variance is subdominant in each experiment if the sky distribution is Gaussian. However, substantial sky coverage is needed if the sky distribution is highly non-Gaussian, especially for those experiments that samples around the Doppler peak.

If the CBR sky distributions are highly non-Gaussian on degree scales, the signature of non-Gaussianity can be detected in the large angular scale CBR maps (Smoot et al. 1992; Ganga et al. 1993). We will use the skewness,

$$\mu_3 = \frac{\langle(\Delta T)^3\rangle}{\langle(\Delta T)^2\rangle^{3/2}}$$

(15)

to characterize the lowest order deviation of CMB distribution from a Gaussian. The skewness $\chi_n^2$ distribution is $\mu_3 = \sqrt{8/n}$. If the sky distribution is a $\chi_n^2$ field on an angular scale $\theta_0 \sim 1^\circ$, after smoothing over with beam size $\theta_s$, the distribution is $\chi_{n'}^2$, with $n' = n(\theta_s/\theta_0)^2$. For COBE, $\theta_s = 7^\circ$ and $n' \approx 50n$ and for FIRS, $\theta_s = 3.8^\circ$ and $n' \approx 14n$. Thus for
the $n \gtrsim 4$, the skewness of temperature fluctuation on COBE and FIRS scale is at least: \[ \mu_3 \gtrsim 0.1 \quad \text{for COBE}; \quad \mu_3 \gtrsim 0.30 \quad \text{for FIRS}. \] The cosmic variance of the skewness $\mu_3$ is $\mu_3 \sim 0.18$ for large angular scale experiments like COBE or FIRS (Srednicki 1993). For COBE, the cosmic variance is larger than the non-Gaussian signals, thus it is impossible for the non-Gaussian signal to be detected. However, for FIRS, the smooth non-Gaussian field will stand above cosmic variance.

Finally, we shall comment on reionization. Early reionization is expected if the density fluctuations are non-Gaussian and objects can form at an earlier epoch. If the universe was reionized at an early epoch ($z \sim 100$), then most of the degree scale temperature anisotropies are dramatically reduced. In this case the sample variance will also be reduced dramatically. As an example, in Table 2 we will also list sample variance (for different experiment) for an re-ionized CDM model with optical depth $\tau = 1$ (Kamionkowski et al. 1994). We conclude that we cannot reconcile the high detection of MAX/GUM experiment with theoretical predictions for any $\chi^2$ distribution.

We want to thank Ben Moore for reading the manuscript. This work was supported by a grant from the NSF.
REFERENCES


Kamionkowski, M., Spergel, D.N., & Sugiyama, N. 1994, CfPA-94-01


Table 1: A brief summary of the experimental situation at degree scales.

Table 2: The sample variance $\sqrt{\sigma_{sam}^2}$ of rms temperature anisotropies in various experiments, in units of $\mu K$, for Gaussian and $\chi^2_{\text{sky}}$ sky distributions. The theoretical predictions for various experiments are given for CDM dominated universe with $\Omega = 1, \Omega_b = 0.06, h = 0.5$ and COBE normalized Harrison-Zeldovich spectrum. The numbers in bold face are for standard recombination and those in brackets are for fully reionized universe with optical depth $\tau = 1$. 
<table>
<thead>
<tr>
<th>Experiment</th>
<th>$\theta_{FWHM}$</th>
<th>$\theta_{chop}$</th>
<th>$\theta_c$</th>
<th>sample size</th>
<th>$\Delta T(\mu K)$ (2$\sigma$ range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARGO</td>
<td>52'</td>
<td>1.8°</td>
<td>30'</td>
<td>63</td>
<td>41 - 71</td>
</tr>
<tr>
<td>MAX (GUM)</td>
<td>0.5°</td>
<td>1.3°</td>
<td>25'</td>
<td>165</td>
<td>85 - 162</td>
</tr>
<tr>
<td>MAX (MuP)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>21</td>
<td>&lt; 68</td>
</tr>
<tr>
<td>MSAM (single)</td>
<td>0.5°</td>
<td>40'</td>
<td>.5°</td>
<td>14</td>
<td>16 - 60</td>
</tr>
<tr>
<td>MSAM (double)</td>
<td>-</td>
<td>-</td>
<td>.3°</td>
<td>-</td>
<td>30 - 85</td>
</tr>
<tr>
<td>Saskatoon</td>
<td>1.44°</td>
<td>2.45°</td>
<td>1.44°</td>
<td>48</td>
<td>28 - 57</td>
</tr>
<tr>
<td>SP91</td>
<td>1.5°</td>
<td>2.95°</td>
<td>1.5°</td>
<td>13</td>
<td>13 - 45</td>
</tr>
<tr>
<td>White dish</td>
<td>12'</td>
<td>23.6'</td>
<td>0.15°</td>
<td>5</td>
<td>&lt; 63</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$(\Delta T)_{th}(\mu K)$</th>
<th>$N_e$</th>
<th>$N^G_{\text{min}}$</th>
<th>$N^{N^2}_{\text{min}}$</th>
<th>Gaussian</th>
<th>$\chi^2_{n=1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARGO</td>
<td>44(27)</td>
<td>63</td>
<td>47(7)</td>
<td>289(49)</td>
<td>19(11)</td>
<td>31(18)</td>
</tr>
<tr>
<td>MAX (GUM)</td>
<td>71(38)</td>
<td>14</td>
<td>318(26)</td>
<td>2226(182)</td>
<td>44(23)</td>
<td>71(38)</td>
</tr>
<tr>
<td>MAX (MuP)</td>
<td>-</td>
<td>12</td>
<td>-</td>
<td>-</td>
<td>45(24)</td>
<td>74(39)</td>
</tr>
<tr>
<td>MSAM (single)</td>
<td>76(44)</td>
<td>8</td>
<td>417(47)</td>
<td>2919(329)</td>
<td>54(31)</td>
<td>88(50)</td>
</tr>
<tr>
<td>MSAM (double)</td>
<td>46(22)</td>
<td>13</td>
<td>56(3)</td>
<td>392(21)</td>
<td>29(14)</td>
<td>47(23)</td>
</tr>
<tr>
<td>Saskatoon</td>
<td>38(27)</td>
<td>27</td>
<td>26(7)</td>
<td>182(49)</td>
<td>20(14)</td>
<td>33(23)</td>
</tr>
<tr>
<td>SP91</td>
<td>38(27)</td>
<td>6</td>
<td>26(7)</td>
<td>182(49)</td>
<td>29(21)</td>
<td>47(33)</td>
</tr>
<tr>
<td>White dish</td>
<td>41(19)</td>
<td>5</td>
<td>35(2)</td>
<td>245(14)</td>
<td>33(15)</td>
<td>54(25)</td>
</tr>
</tbody>
</table>