THE ASTROPHYSICAL JOURNAL, 449:404–412, 1995 August 20

A SIMPLE TEST FOR NON-GAUSSIANITY IN COSMIC MICROWAVE BACKGROUND RADIATION MEASUREMENTS

PAUL GRAHAM AND NEIL TUROK
Joseph Henry Laboratories, Princeton University, Princeton, NJ 08544
AND
P. M. LUBIN AND J. A. SCHUSTER
Physics Department, University of California at Santa Barbara, Santa Barbara, CA 93106
Received 1993 July 1; accepted 1995 March 1

ABSTRACT

We propose a set of statistics for detecting non-Gaussianity in one-dimensional cosmic microwave background radiation (CMBR) anisotropy data sets. These statistics are both simple and, according to calculations over a space of linear combinations of three-point functions, nearly optimal at detecting certain types of non-Gaussian features. We use this statistic to analyze the anisotropy detected by the UCSB SP91 experiment. According to this statistic the mean of the four frequency channels is significantly non-Gaussian. If this signal represents primordial CMBR fluctuations, it would be highly unlikely in a Gaussian theory with a small coherence angle, such as "standard" ($n = 1$, $\Omega_b = 0.05$, $h = 0.5$, $\Lambda = 0$) inflation. We cannot tell whether the observed non-Gaussian signal is cosmological in origin, but if we assume it due instead to foreground emission, and remove the points responsible for the non-Gaussian behavior, the rms of the remaining fluctuations is improbably low for the "standard" inflation theory. Further data are clearly needed before any definitive conclusions may be drawn. We also generalize the ideas behind this statistic to non-Gaussian features that might be detected in other experimental schemes.

Subject headings: cosmic microwave background — cosmology: theory — methods: statistical

1. INTRODUCTION

Many experiments, current and proposed, are dedicated to measuring fluctuations in the cosmic microwave background radiation (CMBR). These measurements promise a strong experimental test of theories of structure formation in the early universe, as each theory predicts a distinct magnitude and form for CMBR fluctuations. In inflationary models, the structure generation mechanism is linear, resulting in a Gaussian pattern of fluctuations completely characterized by its power spectrum (see, e.g., Efstathiou 1990). By contrast, in theories based on symmetry breaking and field ordering (e.g., cosmic strings and textures) nonlinear dynamics lead to a non-Gaussian anisotropy pattern, due in part to horizon-sized topological defects at the epoch of last scattering (Kaiser & Stebbins 1984; Turok & Spergel 1990; Bennett & Rhie 1993; Pen, Spergel, & Turok 1993; Coulson et al. 1994).

CMBR measurements have not yet definitively discriminated among different structure formation theories. This is in part because the measurements are still far from perfect. *COBE* has a low signal-to-noise ratio and large angular smoothing scale (Smoot et al. 1992; Ganga et al. 1993), while other experiments are less noisy but cover only a small region of the sky (e.g., Gaier et al. 1992; Schuster et al. 1993; Cheng et al. 1993; Dragovan et al. 1993; Meinhold et al. 1993; Gundersen et al. 1993). While degree-scale fluctuations are now seen (Schuster et al. 1993; Cheng et al. 1993; Dragovan et al. 1993; Gundersen et al. 1993), the nature and origin of these structures needs further determination.

Even when fairly good CMBR anisotropy power spectrum measurements are available, it is difficult to rule out theories, because most theories include parameters (n, h, Ω, Ω_B, Λ, tensor/scalar ...) which can be adjusted to modify the power spectrum. Such adjustments do not, however, alter the more fundamental Gaussian or non-Gaussian character of the theories, which may be a more powerful discriminator. This paper is aimed at finding statistics which focus on this basic question. Other recent papers which propose statistical tests for non-Gaussianity are Falk, Rangarajan, & Srednicki (1992), Luo & Schramm (1993), Moessner, Perivaropoulos, & Brandenberger (1993), and Srednicki (1993).

We are attempting to extract information from small data sets obtained in very difficult experiments. This is of course quite hazardous: it is unlikely that the idealized assumptions we shall make about the experimental errors are correct. Any non-Gaussianity we detect in the data may well be due to foreground sources or systematic instrumental effects, rather than non-Gaussian cosmology. Nevertheless it is an interesting exercise to see how much may be learned, in principle, from experiments of the type currently being undertaken. This effort may also serve as a guide to what kind of experiment would be most informative in the future. At the very least, we can make rigorous a process which is often performed by eye: the identification of data points which are inconsistent with Gaussian theories and must be thrown out as contaminated if these theories are to be believed.

The various non-Gaussian field ordering theories predict different characteristic forms on the microwave sky—for example, linelike discontinuities for strings, hot and cold spots for textures. However, they share a common feature: they predict regions of sharp gradient (compared to gradients found in a Gaussian theory of the same power spectrum), separated by a characteristic scale of order the horizon at last scattering, from one to a few degrees depending on the reionization history of the universe (Pen et al. 1993; Coulson et al. 1993). In

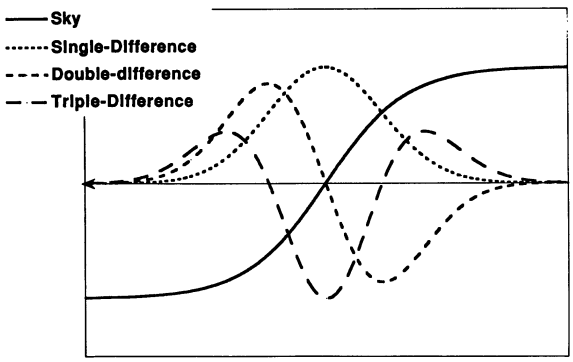

FIG. 1.—The characteristic "signatures" which would result from a step function or region of very sharp gradient in the CMBR, for single- through triple-difference experiments (that is, experiments whose results approximate the first through the third derivatives of the CMBR intensity.) The purpose of this paper is to develop methods of detecting these signatures against a background of instrument and other noise.

this paper we shall discuss statistics which are sensitive to degree-scale large-gradient regions, and use them to discriminate between Gaussian and non-Gaussian theories.

We shall follow tradition and use as our canonical Gaussian theory the "standard" inflationary model, with parameters $n = 1$, $h = 0.5$, $\Omega = 1$, $\Omega_B = 0.05$, $\Lambda = 0$, and negligible tensor mode contribution. This theory has a coherence angle of order 15′, substantially smaller than the scales degree scale experiments probe. The results we get are similar to those obtained assuming uncorrelated Gaussian noise at each point differenced on the sky. So while we shall find rather strong evidence against such theories from the UCSB SP91 data, the constraints are weaker for Gaussian theories with a large coherence angle, as we will demonstrate using tensor-mode inflation.

As a concrete example of a degree-scale CMBR measurement, we shall analyze the UCSB SP91 experiment (Gaier et al. 1992; Schuster et al. 1993). This is a "single difference" experiment; to generate a single data point, the beam moves in a sinusoidal pattern, with the antenna temperature integrated antisymmetrically. The result approximates the first spatial derivative of the fluctuations. A set of results consists of nine to 15 data points (temperature differences), on an arc of constant declination on the sky with 2°.1 separation between points. To correct for atmospheric and other drifts, a best-fit line is removed. The Schuster et al. (1993) data set currently reports a lower error per pixel than any other CMBR anisotropy measurement (error in $\Delta T/T \sim 5 \times 10^{-6}$ for the four-channel average) and shows a significant detection at a level $\Delta T/T \sim 1 \times 10^{-5}$. The Gaier et al. (1982) data four-channel average also shows a similar level of fluctuations. Other ground-based experiments share many of these features, although some integrate their intensities in such a way as to approximate the second or third derivative of the fluctuations rather than the first and are thus called double- or triple-difference experiments. Depending on how many spatial derivatives an experiment takes, a gradient region will leave certain "signature" forms on the data, as shown in Figure 1. Note that although the signatures for high-derivative experiments involve several points, they will appear even if the gradient region is infinitely sharp and narrow (because for these experiments, the instrument's response function is broader than the spacing between data points.)

2. CHOOSING STATISTICS

All current theories of the origin of structure produce fluctuations in the form of a stationary random process. Any such process may be completely characterized by the set of all n-point correlation functions $C_{0r_1r_2\cdots r_{n-1}}$. On a one-dimensional data set $\{x_i : 1 \le i \le N\}$ these may be estimated as

$$C_{0r_1r_2\cdots r_{n-1}} \equiv \frac{1}{(N - r_{n-1})} \sum_{i=1}^{N-r_{n-1}} (x_i)(x_{i+r_1}) \cdots (x_{i+r_{n-1}}) ,$$

where, by convention,

$$0 \le r_1 \le r_2 \le \cdots \le r_{n-1} \le N - 1 .$$

We shall assume that the data set of interest has zero mean, as in UCSB SP91, where a best-fit line is subtracted as explained above. We shall adopt the convention of normalizing the data set to unit variance $[(1/N)\sum_{i=1}^{N} x_i^2 = 1]$ in order to concentrate on the *shape* and not the amplitude of the signal.

The one-point function C_0 is identically zero, and the two-point function C_{00} is identically one (because of our unit-variance convention), an so the first nontrivial correlation is C_{0i}, the two-point correlation function at scale i. This does contain nontrivial information about the fluctuations—it is the Fourier transform of the power spectrum—but it is no help in distinguishing Gaussian from non-Gaussian data.

The three-point function C_{0ij} is a more promising test for non-Gaussianity. For Gaussian noise, $\langle C_{0ij} \rangle = 0$ for all i and j (although for finite data sets there will be random fluctuations about the expected value of zero.) For non-Gaussian skies we expect nonzero three-point functions. For example, consider the three-point function C_{000}, more commonly known as skewness. A data set containing a positive "bump" has several outlying high points, and thus positive skewness. A downward-pointing bump will lead to negative skewness. Either way, the high absolute value of C_{000} could be used to distinguish a data set drawn from a Gaussian model from a region containing a non-Gaussian bump (which, from Fig. 1, is a likely signature of non-Gaussianity in a single-difference experiment.)

As we shall show, however, skewness is not a very powerful statistic for reliably detecting non-Gaussianity in a noisy experiment. We can improve on its performance in two ways.

First, if the CMBR is non-Gaussian, a "bump" marking a gradient region may span two or more adjacent points. Even if the region of steep gradient on the sky were infinitely sharp, it would register in at least two data points because of the instrument's response function. Skewness fails to take advantage of these correlations among closely neighboring points (obviously, since skewness is invariant under spatial scrambling of the data points.) We can remedy this shortcoming by combining several adjacent points; for example, to look for bumps of width $\sim q$, define

$$S_q \equiv \frac{1}{(N - q + 1)} \sum_{i=1}^{N-q+1} \left(\frac{x_i + x_{i+1} + \cdots + x_{i+q-1}}{q} \right)^3 . \quad (1)$$

This statistic responds much more sharply to several adjacent high points than to the same number of high points scattered randomly over the data set, so it better distinguishes actual non-Gaussian bumps from noise. The absolute value of S_q will be near zero if no bump exists and strongly nonzero if there is a single bump.

Of course, S_q is not equal to a simple three-point function, but except for the treatment of points near the edges of a data set, it *is* equivalent to a linear combination of three-point functions. We will expand on this point later, when we show that S_3 is nearly optimal, among all linear combinations of a certain set of three-point functions, at detecting bumps of width near three.

A second way to improve the performance of almost *any* statistic which detects bumps in a data set is to apply it not to the entire data set but to shorter subsets or "windows" of length $L \leq N$. The final statistic $S_{q;L}$ is defined as the absolute value of the most extreme (positive or negative) S_q found in any of the $N-L+1$ possible window positions. Use of these "sliding windows" improves the statistic's performance for several reasons. Most importantly, it prevents a positive gradient region in one part of the data from cancelling a negative gradient in another region (by isotropy, both signs are equally likely to occur.) The procedure also reduces the effects of noise on the statistic's probability distribution by concentrating on only a few points around each gradient signature.

If the data set contains a bump with some number p of adjacent "strong" (highly positive or highly negative) points, we expect the best results from $S_{q;L}$ when $L \approx p + 2(q-1)$: This allows the window to contain every group of q adjacent points which includes at least one "strong" point, and no groups of q points with no "strong" point. S_q is most sensitive to bumps with about q "strong" points. We generally choose q to be slighly larger than the expected value of p, say $q = p + 1$, so

$$L = 3(q-1)\,. \tag{2}$$

We will show that both in Monte Carlo runs and on actual experimental data, statistics perform much better on sliding windows of about this scale than on entire data sets. For long data sets, one might better consider the probability distribution of S_q over window positions, rather than its maximal value, but we do not develop this possibility in this paper.

Our favored choice of statistic will be $S_{3;6}$, as $q = 3$ will be sensitive to bumps only slightly wider than the experimental response function, and can thus detect gradient regions whose width (relative to the two-degree scale set by the experiment) is fairly small but nonnegligible. Equation (2) then sets the window length of $L = 6$.

3. MONTE CARLO RESULTS

To test the relative power of different statistics, we devised a Monte Carlo technique based loosely on the UCSB SP91 experiment. A large number of N-point trial data sets $\{x_i: 1 \leq i \leq N\}$ were generated. Typically, $N = 13$ to match the UCSB SP91 experiment. Half these data sets were generated from a "null" Gaussian model and half from a "bumpy" non-Gaussian model.

We did not want to design statistics around specific cosmological theories, so we used rather vague, generic models at this stage (but in § 5, we *will* analyze the performance of our chosen statistic on realistic models). For the Gaussian "null" sets we used pure white noise (N independent, random points drawn from a Gaussian distribution.) For the non-Gaussian "bumpy" model, a single bump, centered at some random location n_0 within the data set, was laid down:

$$x_n = e^{-\alpha(n-n_0)^2}\,,$$

where α determines the width of the bump; we generally chose $\alpha = 0.5$, which gives the bumps a full-width half-maximum (FWHM) of $(8 \ln 2)^{1/2} \approx 2.4$ pixels. Incidentally, n_0 is not necessarily an integer; the center of the bump can lie between pixels. Independent Gaussian noise was then added to each point to simulate instrument noise. In both the null and "bumpy" models, each data set $\{x_i\}$ was normalized to zero mean and unit variance.

For each statistic S which we want to investigate, we can calculate probability distributions of S in the "null" and "bumpy" models. If S is a powerful detector of non-Gaussianity, there should be little or no overlap between the two distributions. Figure 2 shows these distributions. Figure 2 shows these distributions for our favorite statistic, $S_{3;6}$, using a signal-to-noise ratio of 1.25 (about the same level as seen in the UCSB SP91 data set). There is indeed very little overlap: the value of $S_{3;6}$ calculated on a "bumpy" data set typically exceeds the values of all but a small fraction of the null sets. This "small fraction" varies from one "bumpy" set to another, but its average value is 1.2% (from now on, we'll refer to this as "a mean significance of 1.2%).

For comparison, Figure 3 shows the distributions of absolute values of skewness (calculated in sliding windows of width 6) of data sets from the null and "bumpy" models. The overlap is tremendous; at this noise level, skewness could never reliably distinguish the two models. The idea of using sliding windows also pays off; if we abandon them and calculate S_3 on the entire data set at once, the mean significance rises from 1.2% to 3.4%. Evidently, S_3 is a much more powerful detector of non-Gaussian bumps than is skewness, and $S_{3;6}$ (that is, S_3 with sliding windows) is significantly better than S_3 alone.

To test our assertion that each statistic S_q is most sensitive to bumps of width of about q, we performed Monte Carlo runs like those described above for a variety of "bumpy" models with bumps of different widths. For each model, we measured the average significance obtained using $S_{q;3(q-1)}$ for various (integer) values of q. As expected, each statistic $S_{q;3(q-1)}$ reached its maximum power (lowest mean significance) for bumps of FWHM near q.

4. OPTIMAL STATISTICS

We have justified the statistic $S_{q;3(q-1)}$ by an incremental process, starting with skewness, the simplest detector of non-

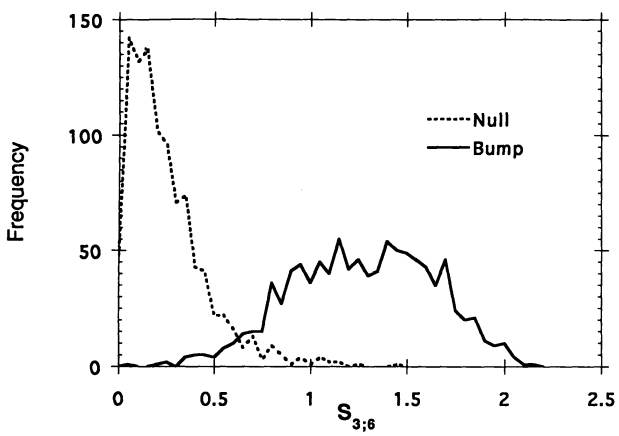

FIG. 2.—Probability distributions of the statistic $S_{3;6}$ (that is, S_3 as defined in eq. [1], calculated in "sliding windows" of length six) for the "bumpy" and null models described in the section on Monte Carlo results. The two distributions show very little overlap, so $S_{3;6}$ appears to be powerful at detecting certain types of non-Gaussianity.

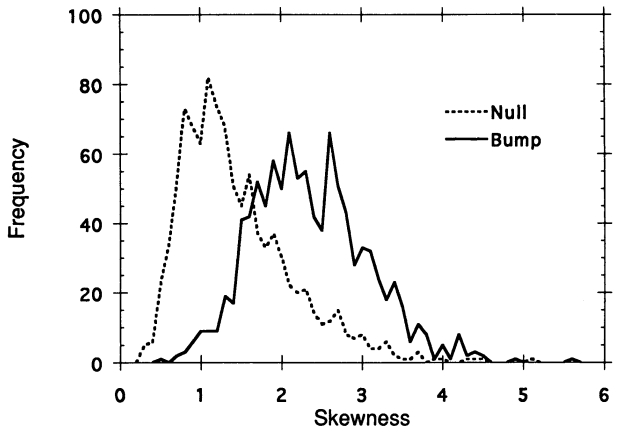

FIG. 3.—Probability distributions of skewness, calculated in "sliding windows" of length six, for the same two models used in Fig. 2. There is considerable overlap; skewness is much less powerful than $S_{3;6}$ at distinguishing between data sets drawn from these two models.

Gaussianity, and modifying it to counter its obvious shortcomings. Our Monte Carlo results showed that the resulting statistic is a much better detector of non-Gaussian "bumps" than is skewness, but we would like to go further and show that it is optimal or near-optimal for this job, at least over certain classes of related statistics.

The procedure of calculating $S_{q;3(q-1)}$ can be separated into three steps. First we convolve the data set with a square tooth of width q (a function equal to one at q adjacent points, and zero elsewhere). Next we take the third power of the convolved data points. Finally we add the results for each connected subset or "window" of length $3(q-1)$ within the data set, and take the most extreme value of S_q as our final statistic $S_{q;3(q-1)}$. For each of these three steps, we can investigate whether a modification of the procedure would produce stronger results.

4.1. *Optimal Choice of Convolution Function*

Our choice to convolve the data with a square tooth function amounts to a filtered deconvolution about the gradient signature we are searching for (in this case a bump), with the high-frequency components suppressed. It is *not* better to use the rigorous deconvolution function of the signature we are seeking; this method is notoriously vulnerable to high-frequency noise. But it *is* worthwhile to see whether convolving the data with some other function, rather than the arbitrarily chosen square tooth, would produce a better statistic.

As mentioned before, S_q is nearly equivalent to a linear combination of several three-point functions. Except for its treatment of points near the window edges, S_3 is proportional to

$$C_{000} + 2(C_{001} + C_{011}) + 2C_{012} + (C_{002} + C_{022}) . \quad (3)$$

To the extent that this approximation holds, the search for the optimal convolution function is equivalent to a search for the optimal linear combination of three-point functions. We define the generalized three-point function S_{G3} by

$$S_{G3} \equiv C_{000} + t(C_{001} + C_{011}) + u(C_{012}) + v(C_{002} + C_{022}) . \quad (4)$$

C_{001} and C_{011} share the same coefficient for reasons of symmetry, as do C_{002} and C_{022}. There is no coefficient before C_{000} because an overall multiplicative constant does not affect a statistic's ability to distinguish between distributions of different *shapes*.

S_{G3} includes terms in all six of the three-point functions which involve no more than three adjacent points at a time. Wider ranging three-point functions, such as C_{013}, are not included because we are attempting to generalize S_3, which searches most powerfully for bumps spanning two or three points.

To estimate the optimal coefficients t, u, and v, the procedure is as follows. We first adopt two simple analytic models of null (Gaussian) and "bumpy" (non-Gaussian) distributions. We then calculate the mean of S_{G3} for the bumpy model, and its mean and variance for the null model. We define the "bump-resolving power" R as the distance, measured in standard deviations of the null model, between the means of the null and bumpy models:

$$R \equiv \frac{\langle S_{G3}\rangle_{\rm bump} - \langle S_{G3}\rangle_{\rm null}}{\sqrt{\langle S_{G3}^2\rangle_{\rm null} - \langle S_{C3}\rangle_{\rm null}^2}} . \quad (5)$$

Finally we maximize R as a function of t, u, and v. (The definition of R is chosen to be both easily calculable and indicative of the statistic's ability to distinguish the two models.)

To simplify the calculation, we work in an infinitely long window of length $N \to \infty$, instead of the window of length $L = 6$ which we shall use in practice. This underscores the fact that choosing a statistic (like S_q) and choosing a window size are two separate ideas; the idea of sliding windows is not specific to S_q but improves the performance of almost any statistic.

4.1.1. *Null Model*

The null model consists of N points drawn from a Gaussian distribution of zero mean and unit variance, with one important modification: each data set $\{x_i\}$ is set explicitly to zero mean (where normally the means would fluctuate slightly about zero). In other words, if $\{y_i\}$ are a set of independent points drawn from a Gaussian distribution,

$$x_i = y_i - \frac{1}{N}\sum_{j=1}^{N} y_j .$$

This sounds like a trivial change, especially for large data sets whose means would ordinarily fluctuate very little, but the explicit normalization makes a surprising difference in the calculation even as $N \to \infty$. We do *not* explicitly normalize each set to unit variance, because it can be shown to make no difference in the $N \to \infty$ limit.

In order to calculate R from equation (5), we need to know $\langle S_{G3}\rangle$ and $\langle S_{G3}^2\rangle$ for this model. $\langle S_{G3}\rangle$ is clearly zero, and using other symmetry properties,

$$\langle S_{G3}^2\rangle = \langle C_{000}^2\rangle + 2t^2\langle C_{001}^2\rangle + u^2\langle C_{012}^2\rangle + 2v^2\langle C_{002}^2\rangle .$$

So we need to calculate expectations such as

$$\langle C_{000}^2\rangle = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\langle x_i^3 x_j^3\rangle$$

for a Gaussian distribution with unit variance.

Expectations such as $\langle x_i^3 x_j^3\rangle$ can be calculated using Wick's theorem, starting with the fundamental two-point expectation $\langle x_i x_j\rangle = \delta_{ij} - N^{-1}$. The N^{-1} term is due to the normalization

of the $\{x_i\}$ to zero mean. The results are

$$\langle C_{000}^2\rangle = 6N^{-1}\,,\quad \langle C_{001}^2\rangle = 2N^{-1}\,,$$
$$\langle C_{002}^2\rangle = 2N^{-1}\,,\quad \langle C_{012}^2\rangle = N^{-1}\,.$$

All other terms in the expression for $\langle S_{G3}^2\rangle$ are zero, so

$$\langle S_{G3}^2\rangle_{\rm null} \to (6 + 4t^2 + u^2 + 4v^2)N^{-1} \quad \text{as } N \to \infty\,. \tag{6}$$

4.1.2. *Bumpy Model*

To represent non-Gaussian, "bumpy" data sets we take

$$x_n = ay_n + bm_n\,,$$

where the "noise" y_n is drawn from a zero mean, unit variance normal distribution, and m_n is the underlying "bumpy" model:

$$m_n = g(e^{-\alpha(n-n_0)^2} - c),\quad \text{where } c = \frac{1}{N}\sqrt{\frac{\pi}{\alpha}},\quad g = N^{1/2}\left(\frac{2\alpha}{\pi}\right)^{1/4}.$$

The bump center n_0 is chosen randomly and is not necessarily an integer. The constants g and c are chosen so that, as $N \to \infty$, m_n will also have zero mean and unit variance (when averaged over all n_0). The purpose of a and b is to set the signal-to-noise ratio: S/N $= b/a$, and $a^2 + b^2 = 1$. It is straightforward to check that all terms in the noise b have zero expectation in S_{G3} (this would not be true for moments higher than the third.) Averaging over n_0 converts all sums to integrals, which in the limit $N \to \infty$ yields the results:

$$\langle C_{000}\rangle = b^3N^{1/2}\left(\frac{8\alpha}{9\pi}\right)^{1/4},\quad \langle C_{001}\rangle = \langle C_{000}\rangle e^{-(2/3)\alpha}\,,$$
$$\langle C_{002}\rangle = \langle C_{000}\rangle e^{-(8/3)\alpha}\,,\quad \langle C_{001}\rangle = \langle C_{000}\rangle e^{-2\alpha}\,.$$

For $\alpha = 0.5$ (corresponding to a bump with FWHM of about 2.4 data points, about the scale we hope to detect with S_{G3}),

$$\langle S_{G3}\rangle_{\rm bump} = b^3N^{1/2}(0.6133 + 0.8789t + 0.2256u + 0.3233v)\,. \tag{7}$$

Now we use equation (5) to estimate the statistic's power to distinguish Gaussian from non-Gaussian models:

$$R \equiv \frac{\langle S_{G3}\rangle_{\rm bump}}{\sqrt{\langle S_{G3}^2\rangle_{\rm null}}} = b^3N\,\frac{0.6133 + 0.8789t + 0.2256u + 0.3233v}{\sqrt{6 + 4t^2 + u^2 + 4v^2}}\,.$$

The optimal values of t, u, and v are those which maximize R; a numerical search for these yields the optimized three-point statistic:

$$S_{G3} = C_{000} + 2.15(C_{001} + C_{011}) + 2.21C_{012} + 0.79(C_{002} + C_{022})\,. \tag{8}$$

As we hoped, this result is similar to equation (3), confirming that our original combination of three-point functions (or equivalently, our choice to convolve the data set with a square tooth) is among the most powerful methods. Since this calculation was approximate, we checked it with more precise Monte Carlo runs, which confirmed that no other choice of convolution function gives dramatically better results. We have settled on the choice of a square tooth as the best combination of simplicity and power.

4.2. *Optimal Choice of Power*

After convolving the data set with a square tooth of width q, S_q requires us to sum the third powers of the convolved data points. We should investigate whether taking some power other than the third would give better results. The higher the power, the more emphasis is given to the most extreme points in the data set (after convolution). Emphasizing extreme points has the advantage of reducing the effects of noise, since it prevents several small bumps, caused by noise, from matching the effect of a large bump in the signal. The drawback is that for a real signal, the highest points will have neighbors which are also higher than average, since neither a physical gradient region on the sky nor the instrument's response function have perfectly sharp edges. High points due to extreme values of instrument noise will not in general have unusual neighbors. So focusing too heavily on extreme points throws away information which would help distinguish a physical signal from noise. The optimal choice of power is that which balances these two competing effects.

The answer is not obvious and is clearly model-dependent, so we turn again to Monte Carlo results. Consider a class of statistics

$$S_{3;6}^{(p)} \equiv \frac{1}{(N-2)}\sum_{i=1}^{N-2}\left(\frac{x_i + x_{i+1} + x_{i+2}}{3}\right)^p. \tag{9}$$

These statistics are calculated in sliding windows of width 6; they differ from $S_{3;6}$ only in the use of the pth power rather than the third. We investigated their ability to distinguish two different "bumpy" models from white noise. One model, described in the earlier section on Monte Carlo results, used bumps of Gaussian profile with FWHM of 2.4 data points. The other model used bumps of a square tooth profile with three adjacent, equally high points randomly placed in the data set. Gaussian white noise was added to both models at a signal-to-noise ratio of 1.25. For the Gaussian-profile model, we expect the resolving strength to peak at some finite power p, while the square bumps should be best resolved at very high p since the argument in favor of lower powers applies only to bumps in which points near the bump have nonzero expectations.

The results are shown in Figure 4, which plots mean significance of detection versus p. For detection of the Gaussian-profile bumps, the optimal power was $p = 3$. For the square bumps, higher powers are always better, as expected. Again, neither gradient regions on the CMBR nor instrument response functions are expected to have sharp cutoffs, so we view the Gaussian profiles as more physically realistic than the square ones, and continue to use $p = 3$. However, the results show that $p = 3$ is only slightly better than other nearby choices, so we will not hesitate to use $p = 4$ later in the paper when we generalize the statistic to search for other types of gradient signatures (because even powers will be more convenient than odd ones).

Also, we should note that $S_{3;6}^{(p)}$ becomes very simple as $p \to \infty$. Our statistic is then equivalent (in its relative ranking of different data sets, which is the only thing that matters) to simply convolving the data with a square tooth of width three, and then choosing the most extreme point. The "sliding window" becomes irrelevant in this limit. Readers who feel that this extra simplicity is worth sacrificing some power may prefer to use this straightforward procedure.

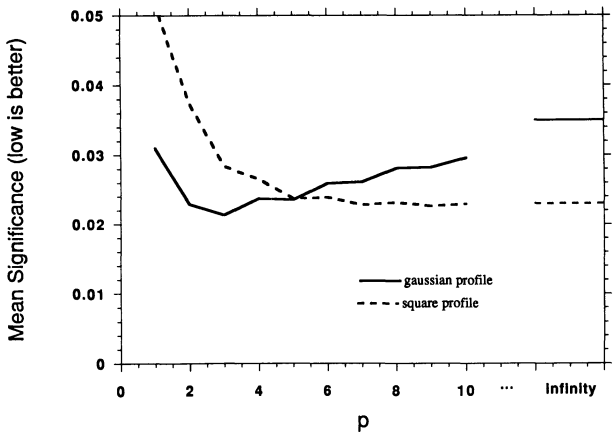

FIG. 4.—The average significance levels achieved by the statistics $S^{(p)}_{3;6}$ defined in eq. (9) as a function of the power p, when used to discriminate Gaussian noise from two different non-gaussian models. One model employs bumps of a Gaussian profile, the other bumps of a square tooth profile. We consider the former more physically reasonable, and consequently adopt the power $p = 3$ in our subsequent analysis.

4.3. *Optimizing Window Size*

The final choice that we have made is to use windows of length $L = 3(q - 1)$ (eq. [2]). There is not much of interest to say about this choice. We gave a rough justification in § 2, and Monte Carlo runs confirm that it is the best or nearly the best length for a wide range of q (when searching for bumps of FWHM near q).

5. THEORETICAL SIMULATIONS

We have developed the statistics S_q by focusing on a generic feature (regions of large gradient) which we hope to see in many non-Gaussian theories. We should now test their ability to distinguish simulated experimental results in specific topological theories from those of Gaussian theories. We have available several sets of simulated 30° square skymaps from various topological theories[1] (Coulson et al. 1994). We also generated 5000 such skymaps for "standard" inflation-plus-CDM (it is much easier to create maps from the power spectrum of a Gaussian theory than from the dynamical simulations needed for a non-Gaussian one).

It is straightforward to apply the two-dimensional SP91 response function to the skymaps and obtain simulated experimental results for each theory and calculate the value of $S_{3;6}$ for each data set. The resulting probability distributions are shown in Figure 5.

Clearly, textures give a radically different distribution than inflation; in an experiment like SP91 it would be fairly common to see texture skies give values of $S_{3;6}$ high enough to rule out inflation with great confidence. Of course, textures also commonly give values of $S_{3;6}$ which do *not* rule out inflation; not every data set can contain a texture. Nevertheless, degree-scale experiments, if they happen to cross a texture boundary, are a good tool for detecting the non-Gaussianity of the texture theory.

For monopoles and strings, the results are not nearly as promising. The non-Gaussian features of these theories tend to subtend much smaller scales than do unwinding textures, and apparently need higher resolution experiments (or perhaps a better statistical treatment) to be readily detected.

Incidentally, the distributions of $S_{3;6}$ and all other interesting statistics we used were very similar for the simulated experimental data from standard inflationary skies and for the simple white-noise data sets we used in § 3. The statistics are apparently not greatly affected either by correlations arising from the inflationary power spectrum (no great surprise, since the two-degree scale of this experiment is on the low-frequency side of these theories' power spectra, and taking a spatial derivative further shifts the spectrum toward high spatial frequencies) or by correlations arising from the instrument's response function (again no surprise; the signature that we are looking for is symmetric and thus orthogonal to the antisymmetric response function. Were we searching for point sources rather than discontinuities, response-function-induced correlations would be a more powerful confounding factor.)

We expect to have more trouble distinguishing Gaussian from non-Gaussian theories when the Gaussian theory has a large coherence length (for example, an inflationary theory with a significant tensor-mode contribution; see Crittenden et al. 1993.) To test this is an extreme case, we generated 5000 maps from the power spectrum of a theory driven *entirely* by tensor-modes resulting in a very large coherence scale. As expected, the distribution of $S_{3;6}$ for SP91 simulations made from these maps suffered a great deal more overlap with the texture distributions; it would be difficult but not impossible to distinguish the two theories in practice.

6. EXPERIMENTAL RESULTS

We focus on a run of the UCSB SP91 experiment which observed 13 points in four frequency channels (Gaier et al. 1992). Cosmological fluctuations should be frequency-independent, so we can average the four channels to better distinguish cosmological fluctuations from instrument noise and, possibly, from astrophysical and atmospheric effects. The mean of the four channels is shown in Figure 6.

We applied the statistic $S_{3;6}$ to data from each of the four channels and to their mean. We compared the results to our simulated data sets for inflationary skies described in § 5 (after

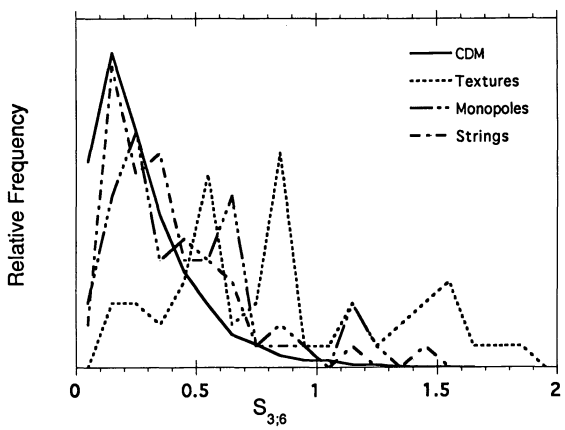

FIG. 5.—Probability distributions of $S_{3;6}$ for simulations of the SP91 experiment on skymaps drawn from four cosmological theories: "standard" inflation, textures, monopoles, and nontopological textures. Distributions such as these can be constructed for any experiment and any theory, to determine which theories are consistent or inconsistent with the value of $S_{3;6}$ given by an experimental data set.

[1] These calculations assumed a fully reionized universe: calculations for the case of standard recombination are still in progress.

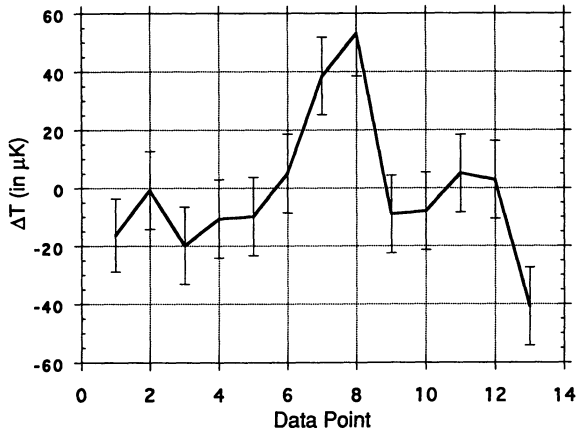

FIG. 6.—A set of results from the UCSB SP91 experiment. This shows the temperature offset δT, averaged over all frequency channels (25 GHz–35 GHz), for each of 13 points separated by 2°1. UCSB SP91 is a single-difference experiment, so these values actually represent (roughly) the difference between the CMBR intensities at two different points.

adding Gaussian noise at the estimated experimental level, and removing a best-fit line, as was done to the actual UCSB SP91 data.) The average of all four channels scores $S_{3;6} = 1.25$, or 0.26% significance relative to "standard" inflation (that is, only 13 of 5000 inflationary simulations give higher values of $S_{3;6}$. Note that this value *is* consistent with textures; as shown in Fig. 5, $S_{3;6}$ exceeds 1.25 about 25% of the time in texture simulations.)

We also compared the mean of the four channels to our ensemble of simulated results from entirely tensor-mode-driven inflationary skies; as mentioned in § 5, this theory has an extremely long coherence length and should be more difficult to disprove using $S_{3;6}$. The mean of the four SP91 channels achieves 2.8% significance relative to tensor-mode inflation; not entirely insignificant but not as strong as the result for "standard" inflation.

In addition to quoting these classical confidence intervals, we can do a Bayesian analysis, in which the change in the relative odds of theory A to theory B following a measurement of a continuous observable X is given by the "Bayes factor"

$$\frac{P(X \mid A)}{P(X \mid B)}, \tag{10}$$

where the probability of observing X in the interval dX is $P(X \mid A)dX$ according to theory A and similarly for theory B. Applying this method to the simulated distributions of $S_{3;6}$ gives a Bayes factor of 17 for textures relative to standard inflation (that is, readers who believe unquestioningly that the SP91 data are a valid map of the CMBR, marred only by white noise, and that our analysis of it is correct, would believe that these data change the relative odds of textures and standard inflation by a factor of 17, in favor of textures.) The Bayes factor for textures relative to 100% tensor-mode inflation is only 3.0, again in favor of textures.

Channels 1 and 2, taken individually, are roughly as non-Gaussian as the mean of the four channels ($S_{3;6} = 1.26$ and 1.20, respectively). Channels 3 and 4 are individually unremarkable; they are consistent with every theory we tested.

It is worth asking if we could have achieved this (tentative and qualified) rejection of "standard" inflation with a simpler statistic than $S_{3;6}$. If we had used ordinary skewness, we would conclude the mean of the four channels to be non-Gaussian at only 11% significance. Adding sliding window of length 6 would improve the significance to 5%, still not nearly as good as the 0.26% significance achieved with $S_{3;6}$.

We chose S_3 (rather than some other S_q) as the preferred statistic for analyzing data sets because we expect the perceived size of the gradient regions to be not much larger than the lower limit set by the instrument response function (two or three pixels). For comparison, Figure 7 shows the significance levels at which all four channels, as well as their mean, can be shown non-Gaussian by the various $\{S_q\}$, with q ranging from 1 to 13. $S_{3;6}$ provides the strongest overall results, although $S_{2;3}$ and $S_{4;9}$ both outperform it on individual channels.

These results show that the UCSB SP91 data are significantly non-Gaussian by these criteria, at least if we favor theories with reasonably small coherence scales. But non-Gaussian data do not necessarily imply non-Gaussian cosmology. The data sets contain a visible spike spanning about 2 pixels (clearly visible in Fig. 6.) This could be the signature of a sharp gradient generated by a non-Gaussian cosmological model, but there are several other possibilities. These include galactic foreground sources, extragalactic but noncosmological sources (unlikely; such sources could not easily match the spatial structure of the data), or a systematic instrumental effect such as sidelobe pickup. With UCSB SP91's limited range of frequency (25–35 GHz) there is not enough spectral information to reliably distinguish astrophysics from cosmology (i.e., by fitting to the spectra of synchrotron or bremsstrahlung radiation). However, if a Gaussian theory with small correlations on 2° scales, like standard inflation, is to be believed, we must assume that the signal in points number 7 and 8 is noncosmological, and remove those points from the data set. The fluctuations of the remaining eleven points may then be used to impose constraints on the cosmological fluctuations. We removed a best-fit line from the remaining points, because the line already removed from the scan must be assumed invalid if two points were contaminated. The remaining points have a χ^2 of 9.9, quite reasonable for 9 degrees of freedom. We then compared the rms to those of the simulated standard inflation data sets (with points 7 and 8 likewise removed, and a new best-fit line subtracted from the remaining points). After this procedure,

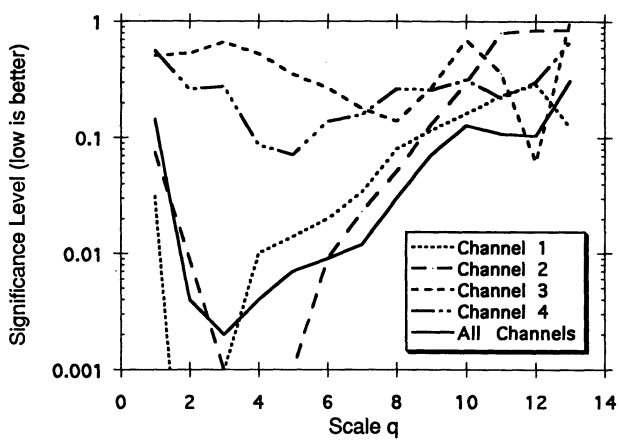

FIG. 7.—Significance of non-Gaussianity vs. scale q for UCSB SP91 data. The significance levels at which each of UCSB SP91's four channels, as well as the mean of these channels, can be distinguished from "standard" inflation by each of the statistics $S_{q;3(q-1)}$ for $1 \le q \le 13$. $S_{3;6}$ shows the best overall performance.

only six of the 5000 simulated data sets had an rms as low as that of the mean of UCSB SP91's four channels. We performed the same procedure, removing points 6, 7, 8, and 9 (since points near the bump may also be suspect) and found the UCSB SP91 data were quieter than all but 38 of the 5000 simulations.

If the UCSB SP91 data are indicative of the CMBR sky, the standard inflation theory is caught on the horns of a dilemma. If points 7 and 8 of the UCSB SP91 data are of cosmological origin, the shape of the data appears non-Gaussian and thus inconsistent with the theory. If the two points are contaminated, the measured rms is too low at high confidence. Other Gaussian models can be tested in a similar way. Before claiming that we have rejected any cosmological model, we must wait to see if these methods demonstrate non-Gaussianity in other experiments. Since these data are very limited in spatial extent, sampling less than 0.1% of the sky, one should be cautious about drawing conclusions.

7. OTHER EXPERIMENTS

$S_{q;3(q-1)}$ can be applied without modification to any single-difference experiment. We recommend $q = 3$ unless the experiment is oversampled or has a very short distance between data points (well under the $\sim$degree scale of the horizon at last scattering.) In such cases one should try several larger values of q to search for high-gradient regions typical of field ordering theories.

For double- and triple-difference experiments, the characteristic "bump" marking gradient regions will be replaced by more complicated signatures representing higher derivatives of a sharp gradient (as seen in Fig. 1). The ideas developed here, with some modification, should apply to these experiments as well. Recall that the procedure of S_q involves convolving the data set with a "square tooth" of width q, then adding the third powers of the convolved data points within each sliding window (see the section on optimal statistics for a discussion of these steps.) For more complicated signatures, we need to modify both the convolution function and the choice of the third power.

The square tooth was a natural choice for the convolution function because it approximates the "bump" signature form which we are looking for. We will continue to convolve with a function or width q that approximates the signature form being sought. Unfortunately, extra complications arise when the gradient-signature being sought crosses zero (as it does for all but single-difference experiments.) The statistics $\{S_q\}$ designed to search for bumps with width of about q data points, were fairly powerful for a wide range of other widths as well. But when searching for a signature which changes sign, a slight mismatch of widths can leave the statistic searching for a form which is *orthogonal* to that actually present, thus canceling the result. The more zero crossings the signature contains, the more critical it becomes to use an accurate width. This requires knowledge not only of the instrument response function but also of the expected width of the gradient region on the sky. The search for non-Gaussianity becomes uncomfortably theory-specific.

Even if the width is chosen perfectly, the convolution of a signature function with an approximation of itself will yield several adjacent points which are strongly nonzero but *alternate in sign* as rapidly as the signature itself does. If we added the third (or any odd) power of these convolved points, they would cancel one another. A simple solution is to raise the convolved data to the fourth power rather than the third, suffering a slight decrease in resolving power (according to our Monte Carlo simulations) but gaining robustness. For example, in a third-derivative experiment, such as Dragovan et al.'s Python (Dragovan et al. 1994), a sharp gradient might be best resolved by a statistic of the form

$$S \sim \sum_j (x_j - 2x_{j+1} + x_{j+2})^4 .$$

The coefficients of x_j, x_{j+1}, and x_{j+2} are fairly obvious guesses, matched to the expected form of the data (Fig. 1). To detect wider gradient regions or other signature forms, simply use an appropriate approximation to the shapes shown in Figure 1; for example, a bump of width four would respond to

$$S \sim \sum_j (x_j - x_{j+1} - x_{j+2} + x_j)^4$$

while in a double-difference experiment, gradient-signature regions with width of about three data points would respond well to

$$S \sim \sum_j (x_j - x_{j+2})^4 .$$

There is no need to modify the "sliding window" scheme as we look for more intricate signature forms; windows of length $3(q - 1)$ still work quite well.

We have performed Monte Carlo simulations which confirm that statistics such as these are much more powerful than simple skewness of kurtosis at detecting the signatures of sharp gradient regions in double- and triple-difference experiments. However, they are not as powerful or as robust as the methods we have developed for single-difference experiments, so there is another option that may be worth considering. Double- and triple-difference results are often constructed in stages, starting with single-difference data and combining adjacent points. It may be best to look for regions of sharp gradient in the original, single-difference data, where they will appear as simple bumps and can be detected by the comparatively robust statistics $\{S_q\}$. The disadvantage to this approach is that single-difference results may be more vulnerable to systematic errors, possibly to the point of being useless. We cannot predict in general which approach will work best for which experiments.

8. CONCLUSIONS

We have proposed a class of statistics which should be quite powerful in detecting a wide range of non-Gaussian features in one-dimensional data sets. Their greatest potential vulnerability is that Gaussian data with significant correlations on the scale of the spacing between data sets may be hard to distinguish from non-Gaussian forms. This could occur in cosmological models with unusually strong power spectra at large angular scales (Crittenden et al. 1993), in experiments with a short distance between data points, or in cases where correlations introduced by the instrument's response function are similar in form to the non-Gaussian "signature" being sought. Any of these factors may make conclusions harder to draw, but they should not lead to false rejections of a Gaussian theory, as long as the null data sets accurately model the Gaussian theory and the instrument's properties.

Our analysis indicates that the results of the UCSB SP91 experiment are in apparent conflict with "standard" inflation. If, as the theory predicts, the CMBR fluctuations are Gaussian, the non-Gaussianity of the data must be due to foreground

contamination. If we discard the contaminated points (those responsible for the non-Gaussian shape) we find a level of fluctuations significantly smaller than the theory predicts. One might worry about possible "conspiracies" here—a region of foreground contamination is more likely to be flagged if it coincides with a strong fluctuation in the CMBR. This makes regions with strong CMBR fluctuations more likely to be discarded and creates a possible bias toward low amplitudes in the remaining points. However, this bias is negligible in theories with short coherence lengths (like standard inflation), because the remaining points are very weakly correlated with the removed points and thus still provide a fair sample.

G. Efstathiou generously provided us with a collection of standard inflation-plus-CDM simulated data sets for the UCSB SP91 experiment, which we used until we could generate our own simulations using power spectra generously provided by R. Crittenden and P. Steinhardt. We thank them, and also M. Dragovan, T. Gaier, J. Gundersen, P. Meinhold, L. Page, J. Peebles, J. Ruhl, D. Spegel, S. Staggs, and D. Wilkinson for helpful conversations. The work of N. T. was supported by NSF contract PHY90-21984, and the David and Lucile Packard Foundation. The work of P. L. and J. S. was supported by NASA contract NAGW-1062, the NSF CfPA and the NSF-DPP.

REFERENCES

Bennett, D. P., & Rhie, S. H. 1993, ApJ, 406, L7
Cheng, E. S., et al. 1993, preprint
Coulson, D., Ferreira, P., Graham, P., & Turok, N. 1994, Nature,
Crittenden, R., et al. 1993, Phys. Rev. Lett., 71, 324
Dragovan, M., Ruhl, J. E., Novak, G., Platt, S. R., Crone, B., Pernic, R., & Peterson, J. 1994, ApJ, 427, L67
Efstathiou, G. 1990, in Physics of the Early Universe, ed. A. Heavens, J. Peacock, & A. Davies (Edinburgh: SUSSP), 361
Falk, T., Rangarajan, R., & Srednicki, M. 1992, ApJ, 403, L1
Gaier, T., et al. 1992, ApJ, 398, L1
Ganga, K., et al. 1993, ApJ, 410, L57
Gundersen, J. O., et al. 1993, ApJ, 413, L1
Kaiser, N., & Stebbins, A. 1984, Nature, 310, 391
Luo, X., & Schramm, D. 1993, Phys. Rev. Lett., 71, 1124
Meinhold, P., et al. 1993, ApJ, 398, L1
Moessner, R., Perivaropoulos, L., & Brandenberger, R. 1993, preprint
Pen, U., Spergel, D., & Turok, N. 1994, Phys. Rev. D, 49, 692
Schuster, J., et al. 1993, ApJ, 412, L47
Smoot, G. F., et al. 1992, ApJ, 396, L1
Srednicki, M. 1993, ApJ, 416, L1
Turok, N., & Spergel, D. 1990, Phys. Rev. Lett., 64, 2736