Optimal estimation of non-Gaussianity

Daniel Babich

Deparment of Physics, Harvard University, Cambridge, Massachusetts 02138, USA
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA

We systematically analyze the primordial non-Gaussianity estimator used by the Wilkinson Microwave Anisotropy Probe (WMAP) science team with the basic ideas of estimation theory in order to see if the limited cosmic microwave background (CMB) data is being optimally utilized. The WMAP estimator is based on the implicit assumption that the CMB bispectrum, the harmonic transform of the three-point correlation function, contains all of the primordial non-Gaussianity information in a CMB map. We first demonstrate that the signal-to-noise (S/N) of an estimator based on CMB three-point correlation functions is significantly larger than the S/N of any estimator based on higher-order correlation functions; justifying our choice to focus on the three-point correlation function. We then conclude that the estimator based on the three-point correlation function, which was used by WMAP, is optimal, meaning it saturates the Cramer-Rao inequality when the underlying CMB map is nearly Gaussian. We quantify this restriction by demonstrating that the suboptimal character of our estimator is proportional to the square of the fiducial non-Gaussianity, which is already constrained to be extremely small, so we can consider the WMAP estimator to be optimal in practice. Our conclusions do not depend on the form of the primordial bispectrum, only on the observationally established weak levels of primordial non-Gaussianity.

I. INTRODUCTION

The origin of the cosmological perturbations which led to the observed cosmic microwave background (CMB) temperature anisotropies and large scale structure is one of the outstanding questions in cosmology. Unfortunately there are only a limited number of independent ways we can constrain the mechanism that produced these perturbations. Windows into the production mechanism of cosmological perturbations include a spectrum of primordial gravity waves, departures from a scale-invariant curvature power spectrum, isocurvature primordial perurbations and non-Gaussianity statistics of the primordial curvature perturbations. While each of these features of the primordial perturbations will change the observed CMB anisotropies and provide insight into the production mechanism of the primordial perturbations, we will focus on the non-Gaussian characteristics of the CMB anisotropies in this paper.

The standard theory of inflation robustly predicts that the primordial curvature perturbations, and therefore the resulting CMB anisotropies, should be nearly Gaussian [1]. By Gaussian we mean that all odd n-point correlation functions exactly vanish and all even n-point correlation functions can be completely expressed as the combination of two-point correlation functions. A non-Gaussian primordial curvature perturbation field would violate one of the above criteria. The degree to which any given Gaussianity criterion is violated can be characteristized by forming a natural ratio between the particular non-Gaussian correlation function and the appropriate combi-

nation of two-point corellation functions [2]. Since we will ultimately perform calculations with a non-Gaussian probability distribution function (PDF) (from which all n-point correlation functions can be calculated), below we will provide a natural quantity that describes the degree of non-Gaussianity in terms of quantites of the PDF. Then we can define the condition nearly Gaussian to mean that this natural quantity is much less than one. Even though the standard calculation within inflation predicts extremely small amounts of non-Gaussianity, there are several viable mechanisms that can produce substantial non-Gaussian primordial curvature perturbations (see [3] and references there within). These processes not only predict different amplitudes of the total non-Gaussianity but also different functional forms [4]. Through linear gravitational and hydrodynamical evolution these curvature perturbations will produce CMB anisotropies with statistical properties that mirror the statistical properties of the primordial perturbations. Therefore it is possible to learn about primordial non-Gaussianity by studying higher-order n-point correlation functions of the CMB anisotropies.

There are more potential sources of non-Gaussianity in the CMB than just primordial non-Gaussianity. The non-linearities in the gravitational and hydrodynamical equations of motion for the baryon-photon fluid prior to recombination can produce non-Gaussianity [5–7]. Secondary anisotropies, such as, the thermal and kinetic Sunyaev-Zeldovich effects, gravitational lensing and the Ostriker-Vishniac effect can all produce non-Gaussianity in the CMB [8–10]. However we are most interested in an observation of primordial non-Gaussianity in the CMB because this would require new ideas for the production of the primordial curvature perturbations and may give a

*Electronic address: babich@physics.harvard.edu

glimpse into beyond the standard model physics. It has been demonstrated that the primordial non-Gaussian signal can be separated from non-Gaussian secondary anisotropies on scales relevant for WMAP and Planck [11].

As stated above we expect the non-Gaussianity of the CMB to be extremely small. Instead of trying to detect individual modes of a non-Gaussian correlation function[1], which would allow us to the examine the functional dependence of momentum wavevectors and better understand the mechanisms that produced the non-Gaussianity [4], it is customary to parameterize the primordial non-Gaussianity with a model. Thus we will be able to combine many bispectra modes and therefore increase the statistical significance of our conclusions. The standard model for primordial non-Gaussianity is the "local model"; the primordial curvature perturbations in real-space are expressed as

$$\Psi(\boldsymbol{x}) = \Phi(\boldsymbol{x}) + f_{NL}[\Phi(\boldsymbol{x})^2 - \langle\Phi(\boldsymbol{x})^2\rangle], \qquad (1)$$

where $\Phi(\boldsymbol{x})$ is a Gaussian field with zero mean and covariance matrix C. The "local model," even though an idealization, has a strong physical motivation; during inflation the nonlinear couplings of general relativity will produce "local model" terms in the bispectrum, the harmonic transform of the connected three-point function [1]. Standard measures of the similarity of two bispectra imply that the "local model" and the standard inflationary calculation are nearly identical [4]. Moreover, models where nonlinearities develop outside the horizon, such as the curvaton model [12] and the inhomogeneous reheating models [13,14], will produce bispectra identical to the "local model." This model will allow us to place limits on the cumulative amplitude of all bispectrum modes and a method has been developed to translate these constraints into constraints on other models of primordial non-Gaussianity [4].

The characteristic amplitude of non-Gaussianity can be found by rescaling Ψ to have unit variance. The coefficient of the rescaled quadratic term, $f_{NL}\langle\Phi(\boldsymbol{x})^2\rangle^{1/2}$, is the natural measure of the amplitude of non-Gaussianity. Below we will find that the signal-to-noise of estimators constructed out of non-Gaussian three-point correlation functions will always contain this factor (or related quantities for higher-order non-Gaussian n-point correlation functions). The best current data constrains $f_{NL}\langle\Phi(\boldsymbol{x})^2\rangle^{1/2} \lesssim 3.5 \times 10^{-3}$ (95% C.L.) [15]. The smallness of this expansion parameter not only allows us to truncate the expansion of the definition of the non-Gaussian field, Eq. (1), but it also implies that the characteristic amplitude of the bispectrum will be much larger than the characteristic amplitude of the trispectrum, the harmonic transform of the connected four-point function. By rescaling arguments analogous to above,

the characteristic amplitude of the trispectrum, either due to an additional cubic term $(f_3\Phi(\boldsymbol{x})^3)$ or from the quadratic piece $(f_{NL}\Phi(\boldsymbol{x})^2)$, would be proportional to $(f_{NL}^2 + f_3) \times \langle\Phi(\boldsymbol{x})^2\rangle$. So unless f_3 is much larger than f_{NL}, the characteristic amplitude of the bispectrum will be much larger than the characteristic amplitude of the trispectrum.

We are not simply interested in the characteristic amplitude of a higher-order correlation function, but in the cummulative signal-to-noise (S/N) of an estimator based on that higher-order correlation function. The number of relevant graphs grows quickly as the order of the correlation function increases, for example, the number of trispectrum quadrilaterals is much greater than the number of bispectrum triangles. As demonstrated in the appendix, if the primordial power spectrum of the Gaussian field is scale-invariant, then the $(S/N)^2$ of a non-Gaussianity estimator based on any non-Gaussian correlation function, regardless of order, will simply scale with the number of observed pixels (N_{pix}), not the number of graphs. Ignoring combinatorial factors we find, for an estimator constructed out of the nth-point correlation function, that

$$(S/N)_n \sim (f_{n-1} + f_{n-2}^2 + \cdots + f_2^{n-2})\langle\Phi(\boldsymbol{x})^2\rangle^{(n-2)/2}$$
$$\times \sqrt{N_{\text{pix}}}. \qquad (2)$$

Since $\langle\Phi(\boldsymbol{x})^2\rangle^{1/2} \sim 10^{-5}$ the estimator based on the three-point correlation function $(f_{NL} \equiv f_2)$ will dominate over all higher-order estimators. This implies that the three-point correlation function is the most effective means to constrain f_{NL} and that it will be significantly easier to constrain f_{NL} than any higher-order f_n. Thus we should begin to explore primordial non-Gaussianity by constraining the amplitude of the bispectrum.

These conclusions depend upon the naturalness argument that $f_{NL} \sim f_3 \sim f_n$, which implies the S/N of the various estimators will be supressed by increasing factors of $\langle\Phi(\boldsymbol{x})^2\rangle^{1/2}$, which we believe to be approximately 10^{-5}. There are several reasons to be cautious. First of all, there might be some reason due to symmetry that $f_{NL} = 0$, then the primordial bispectrum will vanish. Also in the inhomogeneous reheating models, inefficiency in the production of the gravitational potential could equally supresses all terms in expansion of the non-Gaussian field; therefore Eq. (1) would become

$$\Psi(\boldsymbol{x}) = A[\Phi(\boldsymbol{x}) + f_{NL}[\Phi(\boldsymbol{x})^2 - \langle\Phi(\boldsymbol{x})^2\rangle]], \qquad (3)$$

where $A < 1$ [16]. This mutual suppression would increase the inferred value of $\langle\Phi(\boldsymbol{x})^2\rangle^{1/2}$ and cause the suppression of the S/N of higher-order estimators to be less drastic. However within the standard slow-role single field inflationary scenario [1] our conclusions are quite robust.

The standard inflationary scenario for the generation of primordial curvature perturbations predicts that these perturbations were quantum mechanically produced. While it is impossible for us to predict the primordial curvature

---

[1]By individual mode, we mean the correlation function evaluated for a particular configuration of Fourier wavevectors.

fluctuation at any point in the universe, it is possible to predict the statistical properties of these perturbations. With these statistical properties we can create a PDF from which we can generate random realizations of the non-Gaussian curvature perturbations. Since the CMB anisotropies were generated through the linear gravitational and hydrodynamical evolution of these primordial curvature perturbations, the statistical properties of the CMB anisotropies will mirror the statistical properties of the primordial curvature perturbations. Therefore we can view the data from a given CMB experiment as random samples drawn from the appropriate CMB PDF derived from the local model PDF.

We have an explicit expression for the primordial curvature perturbations, Eq. (1), so we can write down the exact PDF for this random field

$$P(\Psi|f_{NL}, C) = \int d^N \Phi \delta^{(N)}(\Psi(\boldsymbol{x}) - \Phi(\boldsymbol{x}) - f_{NL}[\Phi(\boldsymbol{x})^2 - \langle\Phi(\boldsymbol{x})^2\rangle])P(\Phi|C), \qquad (4)$$

and we also include the assumption that Φ has a Gaussian distribution

$$P(\Phi|C) = \frac{e^{-\Phi^T C^{-1}\Phi/2}}{\sqrt{(2\pi)^N \det C}}. \qquad (5)$$

Here we have adopted standard notation for functionals and compactly written the quadratic form of functionals in vector notation. Depending on the particular situation, the quantity $\Phi^T C^{-1}\Phi$ could either be a sum over a discrete set of eigenfunctions (for example, observations of CMB anisotropies $\Phi^T C^{-1}\Phi = \sum_{lm} a_{lm}^* a_{lm}/C_l$) or an integral over a continuous set of eigenfunctions (for example, observations of primordial curvature perturbations $\Phi^T C^{-1}\Phi = \int d^3 \boldsymbol{k} \Phi^*(\boldsymbol{k})\Phi(\boldsymbol{k})/P(k)$). Assuming that the local model is correct and the PDF in Eq. (4) properly represents nature, we can ask how well the data sample will allow us to constrain the underlying parameters of the PDF (f_{NL}, C). There are standard tools from the field of statistical estimation theory that will help us rate estimation procedures and determine the smallest possible error bars we can place on a parameter with a given set of data [17–21].

The purpose of the paper is to determine what are the smallest error bar possible to place on f_{NL} with a given data set and which estimation procedure will allow us to place these constraints. The entire set of measured n-point correlation functions contains all observable information on the underlying PDF contained in the given data set. It may be possible to find a simple estimator which retains all of the available information without undergoing the time consuming process of measuring all n-point correlation functions. Currently researchers are using several different techniques to estimate the non-Gaussianity of a CMB map. An estimator based on the CMB three-point correlation function has been used to constrain the non-Gaussianity

(f_{NL}) of the results from the Wilkinson Microwave Anisotropy Probe (WMAP) [15,22], Very Small Array (VSA) [23] and MAXIMA [24] CMB experiments. In addition, Minkowski functionals [25] and the correlations of Fourier phases (see [26,27] and references within) have been used to characterize non-Gaussianity. We will demonstrate, that for weak levels of non-Gaussianity, the estimator based on the three-point function contains all possible information on f_{NL}. In particular we will show that the exact estimator used by WMAP is an optimal estimator. Therefore the procedure adopted by WMAP will, in principle, provide the best possible error bars on f_{NL}. By using this estimator none of the information potentially contained in the higher order n-point functions is lost. This conclusion is equivalent to the statement that an estimator based on the two-point correlation function contains all possible information on the power spectrum, even though higher-order even n-point correlations functions also contain information on the power spectrum. Finally we will argue that our conclusion does not depend on any characteristic of the local model, just on the observationally established weak levels of non-Gaussianity, so the amplitude of any primordial bispectrum can be optimally constrained by estimators based on the CMB three-point correlation function.

The calculations in this paper assume that the CMB map is free of both instrument noise and Galactic foreground contamination. It is straightforward to include the effects of noise in our estimator for f_{NL}. However the statistical estimation of individual bispectrum modes is very complicated for maps with inhomogeneous noise or a Galactic sky cut. The simple calculation of individual bispectrum modes by combining the appropriate coefficient of the spherical harmonic decomposition of the observed CMB map will result in suboptimal estimates for the bispectrum modes and therefore suboptimal estimates for f_{NL}. An optimal estimator for individual bispectrum modes, including the effects of inhomogeneous instrument noise, has been developed [28]. In this paper we treat the part of the problem which involves the estimation of f_{NL} from measured bispectrum modes.

In Sec. II we introduce the notion of an optimal estimator and the Cramer-Rao Inequality and show how they are related to the maximum likelihood estimator. In Sec. III we analyze a Poisson random field with the appropriate local model non-Gaussian PDF as a simple example of the essential ideas. In Sec. IV we extend our analysis to the scale-invariant distribution predicted by inflation and generalize our results to arbitrary primordial bispectrum in Sec. V. In Sec. VI we summarize our results.

II. ESTIMATION THEORY

Estimation theory is the branch of statistics developed in order to analyze the procedures used to constrain the underlying continuous parameters of a PDF. Assuming

that a given data set is drawn from a known PDF with unknown fixed parameters it allows us to determine the minimum error bars attainable on those parameters using the data. In the following subsections we develop a simple sufficient and necessary condition for a PDF to admit an estimator that saturates the famous Cramer-Rao inequality.

Then for completeness we will discuss the relationship between these concepts and the popular maximum likelihood estimator (MLE). We will only outline necessary concepts, the interested reader should consult [17–21] for more details.

A. Cramer-Rao inequality

In this subsection we discuss the Cramer-Rao Inequality, which determines the lower bound for error bars that we can place on a parameter with a given data sample. These minimum errors bars derived from the Cramer-Rao inequality are valid only for the "frequentist," and not the Bayesian, understanding of statistical estimation (see [17–21] for a discussion of these differing understandings of probability and estimation). An estimator that saturates this inequality is dubbed optimal since it weights the data in the best possible manner. Unfortunately for a general PDF no optimal estimator will exist. In order to develop some intuition regarding the information content of a data sample we will derive the Cramer-Rao inequality for a scalar parameter. The generalization to multiple parameters is conceptly straightforward, but requires linear algebra which obscures some of the essential insights.

We let $\Lambda(x)$ be an unbiased estimator of λ, the parameter we are trying to estimate from our data sample. Assuming the regularity condition

$$\int \frac{\partial \ln p(x|\lambda)}{\partial \lambda} p(x|\lambda) dx = 0, \qquad (6)$$

which holds when we can interchange the order of integration and differentiation, we can use this property to derive the following identity

$$\int (\Lambda(x) - \lambda) \frac{\partial \ln p(x|\lambda)}{\partial \lambda} p(x|\lambda) dx = 1. \qquad (7)$$

Using the Schwarz Inequality, $(A \cdot B)^2 \leq (A \cdot A)(B \cdot B)$ in the notation of linear algebra, we find

$$\left[\int (\Lambda(x) - \lambda) \frac{\partial \ln p(x|\lambda)}{\partial \lambda} p(x|\lambda) dx \right]^2$$

$$\leq \left[\int (\Lambda(x) - \lambda)^2 p(x|\lambda) dx \right]$$

$$\times \left[\int \left(\frac{\partial \ln p(x|\lambda)}{\partial \lambda} \right)^2 p(x|\lambda) dx \right]. \qquad (8)$$

Now defining the Fisher information as

$$F(\lambda) = \int \left(\frac{\partial \ln p(x|\lambda)}{\partial \lambda} \right)^2 p(x|\lambda) dx, \qquad (9)$$

and using Eq. (7) we obtain the Cramer-Rao inequality

$$Var(\Lambda(x)) \equiv \left[\int (\Lambda(x) - \lambda)^2 p(x|\lambda) dx \right] \geq \frac{1}{F(\lambda)}. \qquad (10)$$

The Cramer-Rao inequality states that no estimator of λ can produce error bars smaller than $1/F(\lambda)$.

More importantly we can identify a necessary and sufficient condition for the variance of $\Lambda(x)$ to saturate the Cramer-Rao inequality. It is clear that the Schwarz inequality in Eq. (8) will be saturated if and only if

$$\frac{\partial \ln p(x|\lambda)}{\partial \lambda} = F(\lambda)(\Lambda(x) - \lambda), \qquad (11)$$

thus the PDF must be able to be written in this form if it will admit an optimal estimator. In what follows, we will analyze the local model PDF to determine if it can be expressed in this form.

There is a multiparameter generalization of the Fisher Information defined in Eq. (9),

$$\mathcal{F}_{ij} = \int \frac{\partial \ln p(x|\lambda)}{\partial \lambda_i} \frac{\partial \ln p(x|\lambda)}{\partial \lambda_j} p(x|\lambda) dx. \qquad (12)$$

If we define C_{ij} as the covariance matrix of the set of parameters, the Cramer-Rao Inequality becomes the statement that $C - \mathcal{F}$ is a positive semidefinite matrix. This implies the more familiar statement

$$Var(\lambda_i) \equiv C_{ii} \geq \mathcal{F}_{ii}^{-1}. \qquad (13)$$

If all off diagonal terms in C_{ij} vanish then we can independently estimate all λ_i and the multiparmeter Cramer-Rao Inequality reduces the single parameter case.

B. Relationship with the maximum likelihood estimator

Previous work on the analysis of data focused on the maximum likelihood estimator (MLE). Here we will discuss the relationship between the more familiar MLE and the optimal estimator that we introduced above. The basic principle of the MLE is to consider the observed data as fixed and consequentially choose λ in order to maximize the probability of observing the fixed data. When we take this approach the PDF will be refered to as the likelihood function. We must choose λ in order to maximize the likelihood function, so equivalently we demand

$$\frac{\partial \ln p(x|\lambda)}{\partial \lambda} |_{\lambda=\lambda_{ML}} = 0, \qquad (14)$$

then λ_{ML} is the estimated value of λ. Notice that this equality is true for all data realizations and not just for ensemble averages. In general Eq. (14) will be a complicated nonlinear equation, but there are standard techniques to solve such equations. The most popular is the Newton-Raphson method which is widely used in cosmology for likelihood estimation of the CMB power spectrum from observed CMB maps [29,30].

This approach is widely used in practice because it is always possible to implement, whereas the approach described above often does not yield an estimator. In addition, the MLE is asymptotically optimal and unbiased, meaning that as the amount of data increases the MLE approaches the correct answer with error bars equal to those predicted by the Cramer-Rao inequality. Whenever an optimal estimator exists, the optimal estimator is also the MLE. This is clear from the necessary and sufficient condition for the existence of an optimal estimator, Eq. (11), which automatically satisfies the definition of the MLE, Eq. (14). However the converse does not hold and in general, for finite amounts of data, the MLE is not optimal.

III. POISSON DISTRIBUTIONS

In order to gain intuition we will first analyze a Poisson random field with the appropriate non-Gaussian PDF. This implies that each point in space is uncorrelated with one another and will be independently sampled from the non-Gaussian local model PDF. Since we measure N independent and identically distributed (IID) random variables we can simply scale the single pixel Fisher matrix by N. With these assumptions, the PDF for the single pixel primordial curvature perturbation, Eq. (4), becomes

$$P(\Psi|f_{NL}, \mu) = \int d\Phi \delta(\Psi - \Phi - f_{NL}[\Phi^2 - \mu^2])$$
$$\times \frac{e^{-\Phi^2/2\mu^2}}{\sqrt{2\pi\mu^2}}, \qquad (15)$$

where $\mu^2 = \langle\Phi^2\rangle$. Given an observed Ψ there are two possible values of the underlying Φ

$$\Phi_\pm = \frac{1}{2f_{NL}}\left[\pm\sqrt{1 + 4f_{NL}(\Psi + f_{NL}\mu^2)} - 1\right]. \qquad (16)$$

Integrating over the delta-function the PDF becomes

$$P(\Psi|f_{NL}, \mu) = \frac{1}{\sqrt{2\pi\mu^2}}\left[\frac{e^{-\Phi_+^2/2\mu^2}}{1 + 2f_{NL}\Phi_+} + \frac{e^{-\Phi_-^2/2\mu^2}}{1 + 2f_{NL}\Phi_-}\right]. \qquad (17)$$

In the weakly non-Gaussian limit, $f_{NL}\mu \ll 1$, the contribution from Φ_- is exponentially suppressed so we will ignore it in what follows; we can then write

$$\log P(\Psi|f_{NL}, \mu) = -\frac{\Phi_+^2}{2\mu^2} - \ln(1 + 2f_{NL}\Phi_+) - \ln\mu, \qquad (18)$$

which can be expanded in a power series

$$\log P(\Psi|f_{NL}, \mu) = -\frac{\Psi^2 + 2\mu^2\ln\mu^2}{2\mu^2} + \frac{f_{NL}}{\mu^2}(\Psi^3 - 3\mu^2\Psi)$$
$$- \frac{f_{NL}^2}{2\mu^2}(5\Psi^4 + 5\mu^4 - 14\mu^2\Psi^2)$$
$$+ \mathcal{O}(f_{NL}^3). \qquad (19)$$

If we rewrite $\log P$ such that the power series is expressed in terms of Ψ/μ, which is a random variable with unit variance, then we discover that we are actually expanding in the quantity $f_{NL}\mu$. Doing so we can write Eq. (19) as

$$\log P(\Psi|f_{NL}, \mu) = -I_0(\Psi/\mu) + f_{NL}\mu I_1(\Psi/\mu) - \frac{1}{2}$$
$$\times (f_{NL}\mu)^2 I_2(\Psi/\mu) + \mathcal{O}(f_{NL}^3\mu^3), \qquad (20)$$

where I_0, I_1 and I_2 are defined with respect to Eq. (19). The expectation value of the first-order piece is

$$\langle I_1 \rangle = 6f_{NL}\mu + \mathcal{O}(f_{NL}^3\mu^3), \qquad (21)$$

and the second-order piece is

$$\langle I_2 \rangle = 6 + 272f_{NL}^2\mu^2 + \mathcal{O}(f_{NL}^4\mu^4). \qquad (22)$$

Thus it is necessary to keep the second-order term, I_2, in the expansion of $\log P$. However we can ignore the non-Gaussian piece of $\langle I_2\rangle$, which is a factor of $f_{NL}^2\mu^2$ smaller than the Gaussian piece of $\langle I_2\rangle$. Our conclusions will depend on this approximation, which is equivalent to the weakly non-Gaussian approximation typically made in the literature.

If we view the PDF as a likelihood function, as described above, only I_1 contains information on f_{NL} to $\mathcal{O}(f_{NL}^4\mu^4)$. Since we only need to calculate I_1 from the observed data to specify the likelihood we can regard this quantity as an estimator for f_{NL}; we will address the need to specify I_2 below. Now we will analyze the statistical properties of this estimator.

After choosing the normalization in order to unbias the estimator we find

$$\hat{f}_{NL} = \frac{1}{6\mu^4}(\Psi^3 - 3\mu^2\Psi). \qquad (23)$$

Once we include all N independent observations, this estimator becomes

$$\hat{f}_{NL} = \frac{1}{6\mu^4 N}\sum_i^N(\Psi_i^3 - 3\mu^2\Psi_i), \qquad (24)$$

which has the variance

$$Var(\hat{f}_{NL}) = \frac{1}{6\mu^2 N} + \frac{22f_{NL}^2}{N} + \mathcal{O}(f_{NL}^4\mu^4). \qquad (25)$$

The PDF defined in Eq. (15) satisfies the regularity condition, Eq. (6), for both f_{NL} and μ so we can use the Cramer-Rao inequality to give the lower bound on the error

bars of these parameters. This lower bound will be different than the variance of $\hat{f}_{NL}$ if the PDF does not satisfy the appropriate necessary and sufficient condition for a PDF to admit an optimal estimator, Eq. (11). Simple inspection of the single pixel PDF, Eq. (15)

$$\frac{\partial \ln p(\Psi|f_{NL})}{\partial f_{NL}} = 6\mu^2 \left[\hat{f}_{NL}(\Psi) - \frac{f_{NL}}{6} I_2(\Psi/\mu) \right]$$
$$+ \mathcal{O}(f_{NL}^2 \mu^2), \qquad (26)$$

shows that the necessary and sufficient condition of Eq. (11) is strictly met only if $f_{NL} = 0$. Thus our estimator $\hat{f}_{NL}(\Psi)$ is only optimal for setting non-Gaussian limits on Gaussian maps.

The PDF does not satisfy the appropriate conditions for the existance of an optimal estimator because the function $I_2(\Psi/\mu)$ multiplies f_{NL}. When we evaluate $\langle I_2 \rangle$, we find that there is a leading order Gaussian piece and a non-Gaussian piece, suppressed by a factor of $f_{NL}^2 \mu^2$. If we replace I_2 with its expectation value, ignoring the non-Gaussian piece, we find

$$\frac{\partial \ln p(\Psi|f_{NL})}{\partial f_{NL}} = 6\mu^2 [\hat{f}_{NL}(\Psi) - f_{NL}], \qquad (27)$$

which is exactly the condition for an optimal estimator to exist. While our original conclusion, that $\Psi^3 - 3\mu^2\Psi$ is only optimal for underlying Gaussian distributions, is still true; we can clearly see that within the weakly non-Gaussian limit our estimator is optimal. If the estimator was optimal its variance would exactly equal the bound derived from the Fisher matrix. For the non-Gaussian underlying distribution (the inverse of) the Fisher matrix, Eq. (26), and the estimator variance, Eq. (25), differ by terms proportional to $f_{NL}^2 \mu^2$. This is precisely the type of term that can be ignored within the weakly non-Gaussian approximation. Therefore assuming this approximation is valid, the estimator $\Psi^3 - 3\mu^2\Psi$ is optimal even for underlying non-Gaussian distributions.

This is not a coincidence, but a direct consequence of the regularity condition, Eq. (6), and the weakly non-Gaussian approximation. Since the regularity condition must hold for all values of f_{NL}, we can require it to hold term-by-term in our expansion in f_{NL}. The first-order terms in the regularity condition are $\langle I_1 \rangle_{NG} - f_{NL} \langle I_2 \rangle_G$, where G and NG denote the Gaussian and non-Gaussian expectation values, respectively. Since we know the first-order terms must vanish, we can infer $\langle I_1 \rangle_{NG} = f_{NL} \langle I_2 \rangle_G$. Thus if we can replace I_2 with its Gaussian expectation value then $\hat{f}_{NL} = I_1 / \langle I_2 \rangle_G$ will be an unbiased optimal estimator with variance $1/\langle I_2 \rangle_G$.

Our argument depends on being able to replace I_2 with its expectation value, clearly when the sample size grows the error introduced by making this assumption vanishes. Now we will calculate conditions on the sample size for this property to hold. The expectation value will have a variance which decreases inversely with the sample size, N,

$$Var\left(\frac{I_2(\Psi/\mu)}{6}\right) = \frac{287}{9N}, \qquad (28)$$

thus we need $N \gg 287/9$ for our estimator to be optimal. This number is simply due to combinatorial factors and therefore we do not expect it to drastically change when we consider the full problem with radiative transfer and a scale-invariant distribution. Moreover WMAP observes approximately 10^6 pixels, so we expect this approximation to be quite good.

To check our conclusions, we should still calculate the Fisher matrix elements and compare them with the variance of $\hat{f}_{NL}$. After scaling our results, since we have N IID samples, we find to $\mathcal{O}(f_{NL}^3 \mu^3)$

$$F = N \begin{pmatrix} 6\mu^2 + 368 f_{NL}^2 \mu^4 & -8 f_{NL} \mu \\ -8 f_{NL} \mu & 2/\mu^2 + 20 f_{NL}^2 \end{pmatrix}, \qquad (29)$$

where $\boldsymbol{\lambda} = (f_{NL}, \mu)$. The errors on the parameters are given by the inverse of the Fisher matrix

$$F^{-1} = \frac{1}{N} \begin{pmatrix} 1/6\mu^2 - 28 f_{NL}^2/3 & 2 f_{NL} \mu/3 \\ 2 f_{NL} \mu/3 & \mu^2/2 - 7 f_{NL}^2 \mu^4/3 \end{pmatrix}, \qquad (30)$$

thus the Cramer-Rao bound on f_{NL} is smaller than the variance of $\hat{f}_{NL}$ except when our underlying map is Gaussian.

The fractional discrepancy caused by underlying non-Gaussianity is

$$\frac{Var(\hat{f}_{NL}) - (F_{ff})^{-1}}{Var(\hat{f}_{NL})} = 170 f_{NL}^2 \mu^2 \leq 2.3 \times 10^{-3} \quad (31)$$

assuming, $f_{NL}\mu \leq 3.5 \times 10^{-3}$, which is the best current constraint. Clearly we are justified in ignoring the $\mathcal{O}(f_{NL}^2 \mu^2)$ corrections and in considering the three-point function to be optimal even when the underlying non-Gaussianity is nonzero.

We can extend this discussion by removing the assumption that μ is known *a priori*, then we must simultaneously estimate f_{NL} and μ. Dropping this assumption we will search for an estimator of f_{NL} that jointly estimates μ. Actually it is most natural to estimate $\lambda = f_{NL}\mu^4$, using this variable we find that the coupled unbiased estimator is

$$\hat{\lambda} = \frac{1}{6} \frac{N}{N-3} \left(\frac{1}{N} \sum_i \Psi_i^3 - \frac{3}{N^2} \sum_{ij} \Psi_i^2 \Psi_j \right), \qquad (32)$$

with variance

$$Var(\hat{\lambda}) = \frac{\mu^6}{6N} \frac{N^2 - 3N + 12}{(N-3)^2}. \qquad (33)$$

At $N = 100$, the variance of $\hat{\lambda}$ is increased by just 3% with respect to $\hat{f}_{NL}$. In general the simultaneous estimation of

parameters that are correlated will increase the estimator variance. However the variance will asymptotically approach the single parameter estimator variance as the size of the data sample grows.

IV. SCALE-INVARIANT LOCAL MODEL NON-GAUSSIAN DISTRIBUTIONS

The purpose of this section is to determine whether an optimal estimator for f_{NL} exists when we change the underlying distribution from Poisson to scale-invariant. Observations of the CMB and large scale structure imply that the primordial spectrum of curvature perturbations is scale-invariant; therefore, we must adapt the discussion of the previous section. The Poisson distribution covariance matrix is diagonal in all bases; however, the real-space basis is convenient because the "local model" definition of the non-Gaussian field does not mix different modes in this basis. Therefore in the real-space basis we were able to analyze a single pixel and appropriately scale the final results. For a scale-invariant distribution the covariance matrix is no longer diagonal in a real-space basis, but only in a Fourier basis. Moreover we observe the primordial curvature perturbations projected on the sky as CMB anisotropies after hydrodynamical and gravitational evolution. The covariance matrix of the CMB anisotropies is diagonal in a spherical harmonic basis.

Taking these features into account we will determine if the bispectrum, the three-point correlation function in the spherical harmonic basis, contains all of the non-Gaussian information of a CMB map. This problem will be divided into two parts: (1) We will analyze a 2D scale-invariant distribution without radiative transfer in a spherical harmonic basis; (2) We will include the effects of radiative transfer and show that the WMAP non-Gaussianity estimator is an optimal estimator for f_{NL}.

A. Sachs-Wolfe effect

At first we will ignore the effects of radiative transfer and consider scale-invariant primordial curvature perturbations projected onto the sky. This essentially occurs on large angular scales where the Sachs-Wolfe effect directly maps the curvature perturbations onto temperature anisotropies. However, the main purpose of this subsection is simply to show the changes in the PDF as we switch from an underlying Poisson distribution to a scale-invariant one.

The two-point correlation function of the projected Gaussian curvature field is

$$C \equiv \langle \Phi^*_{lm} \Phi_{l'm'} \rangle = \delta_{l,l'} \delta_{m,m'} \frac{2}{\pi} \int k^2 dk P(k) \frac{j_l^2(k\tau_D)}{9}, \tag{34}$$

where τ_D is the distance to the surface of last scattering. For a scale-invariant power spectrum, $P(k) = A/k^3$, can be

exactly evaluated as

$$\langle \Phi^*_{lm} \Phi_{l'm'} \rangle \equiv \delta_{l,l'} \delta_{m,m'} D_l = \delta_{l,l'} \delta_{m,m'} \frac{A}{9\pi} \frac{1}{l(l+1)}. \tag{35}$$

Now defining the Gaunt integral as

$$\mathcal{G}^{l_1 l_2 l_3}_{m_1 m_2 m_3} = \int d^2 \hat{n} Y_{l_1 m_1}(\hat{n}) Y_{l_2 m_2}(\hat{n}) Y_{l_3 m_3}(\hat{n}) \tag{36}$$

$$= \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\times \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \tag{37}$$

and the average of the quadratic curvature fluctuation as

$$\langle \Phi^2(\mathbf{x}) \rangle = \mu^2 = \int \frac{d^3 \mathbf{k}}{(2\pi)^3} P(k), \tag{38}$$

we can follow the above steps to expand the non-Gaussian PDF in powers of $f_{NL}\mu$. However, we must first calculate the Jacobian transformation of the delta-function

$$J_{lm,ab} \equiv \frac{\partial g_{lm}}{\partial \Phi_{ab}} = -\delta_{l,a} \delta_{m,b} - f_{NL}(-1)^m \sum_{l_1,l_2,m_1,m_2} \mathcal{G}^{ll_1 l_2}_{mm_1 m_2}$$
$$\times \frac{\partial}{\partial \Phi_{a,b}} \Phi_{l_1 m_1} \Phi_{l_2 m_2}, \tag{39}$$

$$= -\delta_{l,a} \delta_{m,b} - 2f_{NL}(-1)^m \sum_{l_1,m_1} \mathcal{G}^{ll_1 a}_{-mm_1 b} \Psi_{l_1 m_1}$$
$$+ \mathcal{O}(f_{NL}^2 \mu^2), \tag{40}$$

where g_{lm} is the argument of the delta-function, used to define the local model PDF in Eq. (4), expressed in a spherical harmonic basis.

For the local model non-Gaussian PDF we find

$$\log P(\Psi | f_{NL}, C) = -\frac{\Psi^T C^{-1} \Psi}{2} + f_{NL} \Psi^T C^{-1}(\Psi^2 - \mu^2)$$
$$- \log \det \frac{\partial g(\Phi)}{\partial \Phi} + \mathcal{O}(f_{NL}^2 \mu^2), \tag{41}$$

which expressed in a spherical harmonic basis is

$$\log P(\Psi | f_{NL}, C) = -\frac{1}{2} \sum_{lm} \frac{\Psi^*_{lm} \Psi_{lm}}{D_l} + f_{NL} \sum_{(l,m)} \mathcal{G}^{l_1 l_2 l_3}_{m_1 m_2 m_3}$$
$$\times \frac{\Psi_{l_1 m_1}}{D_{l_1}} \Psi_{l_2 m_2} \Psi_{l_3 m_3} - f_{NL} \frac{\mu^2 \Psi_{00}}{D_0}$$
$$- \log \det \frac{\partial g(\Phi)}{\partial \Phi} + \mathcal{O}(f_{NL}^2 \mu^2). \tag{42}$$

The notation (l, m) is meant to imply that the sum is over all three l_i and m_i. It is clear that rotational invariance, enforced through the Gaunt integral selection rules, forces

the terms linear in Ψ to only contain the monopole term, Ψ_{00}.

We can simplify matters by examining the derivative of the expression in Eq. (41); using the standard matrix identity, $\log \det A = \mathrm{Tr} \log A$, to rewrite the term that comes from the Jacobian transformation of the delta-function, we find

$$\frac{\partial \log P(\Psi|f_{NL}, C)}{\partial f_{NL}} = \Psi^T C^{-1}(\Psi^2 - \mu^2)$$
$$- Tr\left[J^{-1}\frac{\partial J}{\partial f_{NL}}\right] + \mathcal{O}(f_{NL}\mu), \tag{43}$$

which expressed in the spherical harmonic basis is

$$\frac{\partial \log P(\Psi|f_{NL}, C)}{\partial f_{NL}} = \sum_{(l,m)} \mathcal{G}^{l_1 l_2 l_3}_{m_1 m_2 m_3} \frac{\Psi_{l_1 m_1}}{D_{l_1}} \Psi_{l_2 m_2} \Psi_{l_3 m_3}$$
$$- \frac{\mu^2 \Psi_{00}}{D_0} - 2\Psi_{00} + \mathcal{O}(f_{NL}\mu). \tag{44}$$

Using our intuition from the Poisson case we can identify

$$f_{NL}(\Psi) = \sum_{(l,m)} \mathcal{G}^{l_1 l_2 l_3}_{m_1 m_2 m_3} \frac{\Psi_{l_1 m_1}}{D_{l_1}} \Psi_{l_2 m_2} \Psi_{l_3 m_3}$$
$$- \frac{\Psi_{00}}{D_0}(2\mu^2 + D_0), \tag{45}$$

as the scale-invariant generalization of the estimator for f_{NL}. There are notable differences between the scale-invariant case and the Poisson case. Fundamentally, these differences are a result of expressing the PDF in a spherical harmonic basis instead of a real-space basis. The assumption of scale-invariance only affects the form of D_l.

Rotational invariance forces the linear terms to be proportional to the monopole anisotropy modes. Recall that the CMB monopole anisotropy is unmeasurable in principle and the primordial dipole anisotropy is dominated by the Doppler effect due to the local kinematic motion of the galaxy, so these modes are typically removed from the data. Labelling the original data as Ψ^0 and the new data with the monopole and dipole modes eliminated as Ψ^1, the projection operator will change the PDF of the original data as

$$P(\Psi^1) = \int d^N \Psi^0 \delta^{(N-4)}(\Psi^1 - M\Psi^0) P(\Psi^0), \tag{46}$$

where M is the relevant projection matrix. For all non-monopole and dipole modes the effect of the functional integration will simply be to replace Ψ^0 with Ψ^1. Since the projection matrix eliminates the monopole and dipole modes in the argument of the delta-function, the monopole and dipole terms are simply integrated out. This functional integration simply eliminates the term in $P(\Psi)$ which is linear in Ψ_{00}.

Moreover this functional integration also removes all terms from the cubic piece that contain any Ψ_{00} or Ψ_{1m} modes. It is clear that any term linear or cubic in a monopole or dipole mode will vanish. However it is possible that terms such as $\Psi_{00}\Psi_{00}\Psi_{lm}$ or $\Psi_{11}\Psi_{1-1}\Psi_{lm}$, which survive the functional integration, might introduce a new linear term containing higher modes. Fortunately the symmetry properties of the Gaunt integral eliminate these terms since $\mathcal{G}^{lll'}_{m-m0} \propto \delta_{l'0}$ (see Appendix B of [11]). Therefore the projection of the monopole and dipole eliminates all terms from the PDF which contain any monopole or dipole modes and importantly does not introduce a new linear term. After performing this projection the new estimator is

$$f_{NL}(\Psi) = \sum_{(l,m)} \mathcal{G}^{l_1 l_2 l_3}_{m_1 m_2 m_3} \frac{\Psi_{l_1 m_1}}{D_{l_1}} \Psi_{l_2 m_2} \Psi_{l_3 m_3}, \tag{47}$$

where now the sum excludes all modes with $l_i \le 1$.

We must decide whether this estimator contains all the information of the modified PDF. In analogy to the Poisson case, the f^2_{NL} term is an even function of the Ψ_{lm} modes. Therefore its expectation value will have a purely Gaussian piece which is fixed by the regularity condition to be precisely the correct value for the scale-invariant PDF to satisfy Eq. (11) and admit an optimal estimator. Again we conclude that $f_{NL}(\Psi)$ is an optimal estimator in the weakly non-Gaussian limit.

B. Radiative transfer

Now we will include the effects of radiative transfer and we will find that the 2D CMB version of our optimal estimator is the Weiner filter estimator used by the WMAP science team [22]. The process of radiative transfer alters the amplitudes of the 3D curvature perturbations and projects them onto 2D CMB temperature anisotropies. This process will relate the CMB PDF to the local model PDF as

$$P(a|f_{NL}) = \int d^N \Psi \delta^{(M)}(a_{lm}$$
$$- \int r^2 dr \alpha_l(r) \Psi_{lm}(r)) P(\Psi|f_{NL}), \tag{48}$$

where $M < N$ because of the projection. The additional spurious degrees of freedom, which do not affect the observable CMB anisotropies, can be integrated out. Here we define

$$\alpha_l(r) = \frac{2}{\pi} \int k^2 dk j_l(kr) \Delta_l(k), \tag{49}$$

where $\Delta_l(k)$ is the standard radiation transfer function [31]. We can connect the present discussion with previous case of the Sachs-Wolfe effect by noting that the formulae of the previous subsection can be reproduced by choosing

$$\alpha_l(r) = \frac{2}{\pi} \int k^2 dk j_l(kr) \frac{1}{3} j_l(k\tau_D). \qquad (50)$$

By substituting this expression in the following equations the results in the previous subsection can be derived. This formula is valid on large angular scales when the wavelength of the relevant primordial curvature perturbation is much larger than the size of the sound horizon.

To calculate this functional integral we will use the exponentiated form of the delta-function

$$\delta^{(M)}(a_{lm} - \int r^2 dr \Delta_l(r) \Psi_{lm}(r))$$

$$= \int d^M B e^{-i\sum_{lm}(-1)^m B_{l-m}(a_{lm} - \int r^2 dr \alpha_l(r) \Psi_{lm}(r))}, \qquad (51)$$

here B_{lm} is simply a "dummy variable." Using this representation we can "complete the square" and perform the functional Gaussian integrations.

First we must find the 3D local model PDF, the covariance matrix between two primordial 3D curvature perturbations can be calculated as

$$C = \langle \Phi^*_{l_1 m_1}(r_1) \Phi_{l_2 m_2}(r_2) \rangle = \delta_{l_1, l_2} \delta_{m_1, m_2} D_{l_1}(r_1, r_2) \quad (52)$$

$$= \delta_{l_1, l_2} \delta_{m_1, m_2} \frac{2}{\pi} \int k^2 dk P(k) j_{l_1}(kr_1) j_{l_1}(kr_2). \qquad (53)$$

Since we will need the inverse of C in order to express the PDF of Φ, we can symbolically define the inverse of $D_l(r_1, r_2)$ as

$$\int r^2 dr D_l(r_1, r) D_l^{-1}(r, r_2) = \frac{\delta(r_1 - r_2)}{r_1^2}. \qquad (54)$$

Using these definitions the non-Gaussian PDF can be expanded as

$$\log P(\Psi | f_{NL}, C) = \log P_G(\Psi | f_{NL}, C)$$
$$+ \log P_{NG}(\Psi | f_{NL}, C) \qquad (55)$$

$$= -\frac{1}{2} \sum_{l,m} \int r_1^2 dr_1 r_2^2 dr_2 (-1)^m \Psi_{l-m}(r_1) D_l^{-1}(r_1, r_2) \Psi_{lm}(r_2)$$

$$+ f_{NL} \sum_{(l,m)} \int r_1^2 dr_1 r_2^2 dr_2 G^{l_1 l_2 l_3}_{m_1 m_2 m_3} \Psi_{l_1 m_1}(r_1) D_{l_1}^{-1}(r_1 r_2)$$

$$\times \Psi_{l_2 m_2}(r_2) \Psi_{l_3 m_3}(r_2) + \mathcal{O}(f_{NL}^2 \mu^2),$$

where we have ignored the irrelevant constant piece and the linear terms since they will ultimately vanish when we project out the monopole mode.

At this point it will be convenient to introduce the following function

$$\beta_l(r) = \frac{2}{\pi} \int k^2 dk P(k) j_l(kr) \Delta_l(k), \qquad (56)$$

which is defined such that

$$\int r^2 dr \alpha_l(r) \beta_l(r) = C_l. \qquad (57)$$

Using the definition of $D_l^{-1}(r_1, r_2)$ we can show that

$$\alpha_l(r_1) = \int r_2^2 dr_2 D_l^{-1}(r_1, r_2) \beta_l(r_2), \qquad (58)$$

$$\beta_l(r_1) = \int r_2^2 dr_2 D_l(r_1, r_2) \alpha_l(r_2) \qquad (59)$$

these properties will be very useful in what follows. Note that the functions $\alpha_l(r)$ and $\beta_l(r)$ are equivalent to $b_l^{lin}(r)$ and $b_l^{nl}(r)$ defined in [32].

Now we are ready to perform the functional integration and find the non-Gaussian CMB PDF. We must "complete the square" of the following term

$$-\frac{1}{2} \sum_{l,m} (-1)^m \left[\int r_1^2 dr_1 r_2^2 dr_2 (-1)^m \Psi_{l-m} D_l^{-1}(r_1, r_2) \Psi_{lm}(r_2) \right.$$

$$\left. - 2i B_{l-m} \int r^2 dr \alpha_l(r) \Psi_{lm}(r) \right]. \qquad (60)$$

This can be rewritten as

$$-\frac{1}{2} \sum_{l,m} (-1)^m \left[\int r_1^2 dr_1 r_2^2 dr_2 (\Psi_{l-m}(r_1) \right.$$

$$- \zeta_{l-m}(r_1)) D_l^{-1}(r_1, r_2) (\Psi_{lm}(r_2) - \zeta_{lm}(r_2))$$

$$\left. + B_{l-m} B_{lm} C_l \right], \qquad (61)$$

where

$$\zeta_{lm}(r) = -i \beta_l(r) B_{lm}. \qquad (62)$$

Once more we must "complete the square" for the following term

$$-\frac{1}{2} \sum_{l,m} (-1)^m [C_l B_{l-m} B_{lm} + 2i B_{l-m} a_{lm}]. \qquad (63)$$

This can be rewritten as

$$-\frac{1}{2} \sum_{l,m} (-1)^m \left[C_l(B_{l-m} - \eta_{l-m})(B_{lm} - \eta_{lm}) + \frac{a_{l-m} a_{lm}}{C_l} \right] \qquad (64)$$

where

$$\eta_{lm}(r) = \frac{i a_{lm}}{C_l}. \qquad (65)$$

Thus we are left with the correct Gaussian piece for the CMB PDF. Performing the two Gaussian functional integrations is equivalent to performing the substitution

$$\Psi_{lm}(r) \rightarrow \frac{\beta_l(r)}{C_l} a_{lm}. \qquad (66)$$

into the non-Gaussian cubic portion of the PDF.

This substitution is equivalent to the Weiner filter solution proposed by [22] in order to estimate the underlying curvature perturbations from an observed CMB anisotropy map. The process of projection and radiative transfer is not invertible, but nevertheless we can regularize the inversion process by requiring that the reconstructed potential minimizes the variance. This approach was adopted in [22], where it was shown that the optimal inversion procedure was to Weiner filter the CMB map as

$$\Psi_{lm}(r) = \frac{\beta_l(r)}{C_l} a_{lm}. \tag{67}$$

Here we have demonstrated that a straightforward functional integration of the local model PDF gives the same result.

Performing the functional integration, or equivalently substituting the Weiner filter solution for Ψ into Eq. (55), we find that the non-Gaussian cubic term in the PDF

$$\begin{aligned}
\log P_{NG}(a|f_{NL}, C) &= \sum_{(l,m)} \mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} \int r_1^2 dr_1 r_2^2 dr_2 \Psi_{l_1 m_1}(r_1) D_{l_1}^{-1}(r_1, r_2) \Psi_{l_2 m_2}(r_2) \Psi_{l_3 m_3}(r_2) \\
&= \sum_{(l,m)} \mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} \frac{a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3}}{C_{l_1} C_{l_2} C_{l_3}} \int r_1^2 dr_1 r_2^2 dr_2 \beta_{l_1}(r_1) D_{l_1}^{-1}(r_1, r_2) \beta_{l_2}(r_2) \beta_{l_3}(r_2) \\
&= \sum_{(l,m)'} \mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} \frac{a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3}}{C_{l_1} C_{l_2} C_{l_3}} b_{l_1 l_2 l_3},
\end{aligned} \tag{68}$$

where we define the reduced CMB Bispectrum as

$$\begin{aligned}
b_{l_1 l_2 l_3} = 2 \int r^2 dr [\alpha_{l_1}(r)\beta_{l_2}(r)\beta_{l_3}(r) &+ \beta_{l_1}(r)\alpha_{l_2}(r)\beta_{l_3}(r) \\
&+ \beta_{l_1}(r)\beta_{l_2}(r)\alpha_{l_3}(r)],
\end{aligned} \tag{69}$$

and the prime indicates that we restrict the sum over l_i such that $l_1 \leq l_2 \leq l_3$. It is standard to separate the piece of the CMB three-point correlation function fixed by rotational invariance. We do this by introducing the Gaunt integral, Eq. (36), and the reduced CMB bispectrum, which are related to the CMB three-point correlation function as

$$\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle = \mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} b_{l_1 l_2 l_3}. \tag{70}$$

Combining these results we find that the scale-invariant non-Gaussian PDF for the CMB anisotropies, ignoring the constant piece, is

$$\begin{aligned}
\log P(a|f_{NL}) = -\frac{1}{2}\sum_{l,m} \frac{a_{lm}^* a_{lm}}{C_l} + f_{NL} \sum_{(l,m)'} \frac{\mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} b_{l_1 l_2 l_3}}{C_{l_1} C_{l_2} C_{l_3}} \\
\times a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} + \mathcal{O}(f_{NL}^2 \mu^2).
\end{aligned} \tag{71}$$

Again this has the correct form to admit an optimal estimator, within the weakly non-Gaussian approximation, which is

$$\hat{f}_{NL}(a) = \frac{1}{S_{\text{norm}}} \sum_{(l,m)'} \frac{\mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} b_{l_1 l_2 l_3}}{C_{l_1} C_{l_2} C_{l_3}} a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3}, \tag{72}$$

where the normalization constant is

$$\begin{aligned}
S_{\text{norm}} = \sum_{(l,m)'} &\frac{(2l_1 + 1)(2l_2 + 1)(2l_3 + 1)}{4\pi} \\
&\times \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}^2 \frac{b_{l_1 l_2 l_3}^2}{C_{l_1} C_{l_2} C_{l_3}}.
\end{aligned} \tag{73}$$

A procedure based on this estimator has been implemented by the WMAP science team in their analysis of primordial non-Gaussianity. Their interpretation of the estimator is in terms of Weiner filtered estimates of the primordial curvature. First the Weiner filter $\beta_l(r)/C_l$ is used to estimate the primordial curvature fluctuation $\Psi_{lm}(r)$ from the CMB anisotropy a_{lm}, as in Eq. (67). Then the two estimates of the primordial curvature perturbations are used to estimate the quadratic piece of the primordial curvature field according to the definition of the local model and $\alpha_l(r)$ is used to calculate the CMB temperature anisotropy due to this quadratic piece. This quadratic CMB template is correlated with the observed CMB anisotropies to give an estimate, once properly normalized, of the non-Gaussianity of the observed CMB anisotropies. We have shown that this intuitive procedure results in an estimator that is optimal.

V. GENERAL NON-GAUSSIAN MODELS

In the previous section we applied the basic ideas of estimation theory to show that the local model non-Gaussian CMB PDF admitted an optimal estimator in the weakly non-Gaussian limit. As mentioned in the Introduction the local model is a physically motivated idealization, however most primordial bispectra calculated within models of the early universe contain additional terms. Starting with the definition of the local model, Eq. (1), we derived the resulting CMB PDF. Since it is not possible to find the general non-Gaussian PDF which corresponds to a general non-Gaussian bispectrum, we are unable to retrace the above steps in order to extend our analysis. However we will argue that the non-Gaussian CMB PDF in Eq. (71) is simply the Edgeworth expansion, which holds for arbitrary non-Gaussian CMB PDF, and therefore our conclusion that the bispectrum estimator is optimal holds for arbitrary model of the primordial non-

Gaussianity. Because of the arguments in the Introduction we assume that a model for primordial non-Gaussianity is best characterized by its bispectrum and that we should try to constrain the amplitude of the bispectrum. We choose to define the amplitude of a general bispectrum as the coefficient of the bispectrum evaluated in the equilateral configuration (see [4] for a full discussion of this point).

The Edgeworth expansion is a way to express the non-Gaussianity of a PDF in the form of a series expansion (see [2] and references there within). This allows one to explicitly write down a non-Gaussian PDF if its lowest order moments or correlation functions are known. The Edgeworth expansion of a 1D PDF is a simple expansion in Hermite polynomials; a multivariate generalization has been found [33,34]. Adopting the notation relevant for CMB anisotropies the Edgeworth expansion is

$$
P(a) = \left[1 - \sum_{(l,m)'} \langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle \frac{\partial}{\partial a_{l_1 m_1}} \frac{\partial}{\partial a_{l_2 m_2}} \frac{\partial}{\partial a_{l_3 m_3}} \right]
$$
$$
\times \prod_{lm} \frac{e^{-\frac{a_{lm}^* a_{lm}}{2C_l}}}{\sqrt{2\pi C_l}}, \tag{74}
$$

where $\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle$ can still be decomposed into the gaunt integral and the reduced bispectrum. Here the general reduced bispectrum can be calculated as

$$
b_{l_1 l_2 l_3} = \left(\frac{2}{\pi} \right)^3 \int k_1^2 dk_1 k_2^2 dk_2 k_3^2 dk_3
$$
$$
\times \int x^2 dx j_{l_1}(k_1 x) j_{l_2}(k_2 x) j_{l_3}(k_3 x)
$$
$$
\times B(k_1, k_2, k_3) \Delta_{l_1}(k_1) \Delta_{l_2}(k_2) \Delta_{l_3}(k_3), \tag{75}
$$

where $B(k_1, k_2, k_3)$ is the general primordial bispectrum [4]. When the primordial bispectrum is calculated according to the local model the CMB reduced bispectrum is given by Eq. (69).

Performing the functional differentiation in the Edgeworth expansion we find

$$
P(a) = \prod_{lm} \frac{e^{-\frac{a_{lm}^* a_{lm}}{2C_l}}}{\sqrt{2\pi C_l}} \left[1 + \sum_{(l,m)'} b_{l_1 l_2 l_3} \mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} \left[\frac{a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3}}{C_{l_1} C_{l_2} C_{l_3}} \right. \right.
$$
$$
- (-1)^{m_3} \left(\frac{a_{l_1 m_1}}{C_{l_1} C_{l_3}} \delta_{l_2, l_3} \delta_{m_2, -m_3} + \frac{a_{l_2 m_2}}{C_{l_2} C_{l_3}} \delta_{l_1, l_3} \delta_{m_1, -m_3} \right.
$$
$$
\left. \left. \left. + \frac{a_{l_3 m_3}}{C_{l_2} C_{l_3}} \delta_{l_1, l_2} \delta_{m_1, -m_2} \right) \right] \right]. \tag{76}
$$

Again the properties of the gaunt integral force the linear terms to be proportional to the monopole anisotropy mode, a_{00}. After projecting out the monopole anisotropy the form of the non-Gaussian cubic term in the Edgeworth expansion is identical to the form of non-Gaussian cubic term in the local model CMB PDF, Eq. (71).

In the previous section we argued that estimators based on the bispectrum are optimal if the PDF can be expressed

as in Eq. (71). This form is not unique to the local model, but is simply part of the Edgeworth expansion which is relevant regardless the form of primordial bispectrum. The only necessary condition is that the level of non-Gaussianity is small. Thus we conclude that estimators based on the bispectrum are optimal for any form of primordial non-Gaussianity characterized by its bispectrum provided the amplitude of the non-Gaussianity is sufficiently small.

VI. DISCUSSION

We analyzed the standard model for primordial non-Gaussianity with the tools of estimation theory and found that the estimator constructed out of the CMB three-point correlation function is an optimal estimator for f_{NL}, the amplitude of the primordial non-Gaussianity. Our conclusion is only true within the weakly non-Gaussian approximation, which implies that we ignore non-Gaussian contributions to four-point correlation function compared to the much larger Gaussian contributions. In our calculations this is equivalent to ignoring terms proportional to $f_{NL}^2 \langle \Phi(\boldsymbol{x})^2 \rangle$, which is already constrained by the WMAP data to be extremely small. Therefore we can consider the standard WMAP estimator, which is based on the CMB three-point correlation function, to be optimal in practice. This property only depends on the weak level of non-Gaussianity, not on any assumed form for the primordial bispectrum. Therefore we argued that the amplitude of a general primordial bispectrum can be optimally estimated.

Our calculations demonstrated that the WMAP estimator was optimal to constrain f_{NL} and therefore contained all the information in the observed data on this parameter. Future work can now focus on practical implementations of this optimal estimator. Some of the practical concerns that affect the implementation of an estimator for f_{NL} include biasing due to non-Gaussianity from secondary anisotropies, nonuniform instrument noise and the need to jointly estimate f_{NL} and the basic cosmological parameters from the same data which will degrade the performance of the estimator. While we demonstrated that the WMAP estimator is optimal for the estimation of f_{NL}, there still is a need to find a quick method to optimally estimate individual three-point correlation function modes from CMB maps with Galactic foreground contamination and inhomogeneous instrument noise [24,28]. We implicitly assumed that the three-point correlation function modes could simply be calculated from observed CMB maps. For maps with inhomogeneous noise, this procedure results in suboptimal error bars on the individual bispectra modes. An optimal procedure, which is quite slow for large data sets, has been developed [28]; a quicker procedure must be found if it can be realistically applied to the WMAP and Planck data sets.

Also, the non-Gaussianity of secondary anisotropies is not well known on the arcminute scales relevant for current

and future CMB experiments. Fortunately on intermediate scales there is little contamination of the primordial non-Gaussianity estimator by secondary anisotropies; however we simply do not know if this is true on the extremely small scales relevant for upcoming CMB experiments. Also we do not know how errors in the basic cosmological parameter, from which we calculate the radiation transfer functions and the Weiner filters, will affect this analysis. This topic should also be investigated.

While much progress has been made in both the theoretical understanding and practical analysis of the signatures of primordial non-Gaussianity in CMB maps, there still is much more work needed to be done before the field will become fully developed. However the tremendous potential for insight into the production mechanism of the primordial curvature perturbations makes this work worthwhile.

ACKNOWLEDGMENTS

APPENDIX: SIGNAL-TO-NOISE RATIO OF HIGHER-ORDER ESTIMATORS

In this appendix we demonstrate the features of the S/N of an estimator based on an nth-order correlation function needed to argue that the estimator based on the three-point correlation function would have the largest S/N. We will do our calculation within a toy model that should retain all of the important features of the problem. In what follows we ignore radiative transfer and the curvature of the sky, so we will simply observe the underlying modes within the flat-sky approximation [35].

Extending the basic idea of the "local model" we will assume that the observed non-Gaussian field Ψ can be expanded in the Gaussian field Φ as

$$\Psi = \Phi + f_2[\Phi^2 - \langle\Phi^2\rangle] + f_3[\Phi^3 - \langle\Phi^3\rangle] + \cdots$$
$$+ f_n[\Phi^n - \langle\Phi^n\rangle] + \cdots. \tag{A1}$$

The various n-point correlations function of Ψ can be calculated by substituting our definition of Ψ, Eq. (A1), and using Wick's theorem. It is important to note that we are interested in connected correlation functions, which contain a single delta-function. In what follows we ignore all $\mathcal{O}(1)$ factors.

Working within our model, the power spectrum is evaluated as [35]

$$C(l) \sim \frac{A}{l^2}, \tag{A2}$$

and the connected n^{th}-order correlation function as

$$T(\boldsymbol{l}_1, \boldsymbol{l}_2, \cdots, \boldsymbol{l}_n) \sim (f_{n-1} + f_{n-2}^2 + \cdots + f_s^{n-s}$$
$$+ \cdots f_2^{n-2})A^{n-1}\frac{l_1^2 + l_2^2 + \cdots + l_n^2}{l_1^2 l_2^2 \cdots l_n^2}, \tag{A3}$$

where A is the amplitude of the primordial curvature power spectrum, $P(k) = A/k^3$. Note that there is no contribution to n^{th}-order correlation function from f_n.

We also define the total $(S/N)^2$ of the estimator as

$$\left(\frac{S}{N}\right)^2 \sim \int d^2\boldsymbol{l}_1 d^2\boldsymbol{l}_2 \cdots d^2\boldsymbol{l}_n \frac{[\delta^{(2)}(\boldsymbol{l}_1 + \boldsymbol{l}_2 + \cdots + \boldsymbol{l}_n)T(\boldsymbol{l}_1, \boldsymbol{l}_2, \cdots, \boldsymbol{l}_n)]^2}{C(l_1)C(l_2)\cdots C(l_n)} \tag{A4}$$

Substituting the results for C_l and $T(\boldsymbol{l}_1, \boldsymbol{l}_2, \cdots, \boldsymbol{l}_n)$ into Eq. (A4) we find

$$\left(\frac{S}{N}\right)^2 \sim f_{\text{sky}} f_{\text{tot}}^2 A^{n-2} \int d^2\boldsymbol{l}_1 d^2\boldsymbol{l}_2 \cdots d^2\boldsymbol{l}_n \delta^{(2)}$$
$$\times (\boldsymbol{l}_1 + \boldsymbol{l}_2 + \cdots + \boldsymbol{l}_n)\frac{[l_1^2 + l_2^2 + \cdots + l_n^2]^2}{l_1^2 l_2^2 \cdots l_n^2}, \tag{A5}$$

where f_{sky} is the fraction of the observed sky and $f_{\text{tot}} = f_{n-1} + f_{n-2}^2 + \cdots + f_s^{n-s} + \cdots f_2^{n-2}$ is the total coupling cofficient that results from considering all contribution to a given correlation function.

The upper bound of integration for all l_i is l_{max} so we can rescale the integral to be "dimensionless," then we find

$$\left(\frac{S}{N}\right)^2 \sim f_{\text{sky}} f_{\text{tot}}^2 A^{n-2} l_{\text{max}}^2 = f_{\text{tot}}^2 A^{n-2} N_{\text{pix}}, \tag{A6}$$

where we defined N_{pix} to be the total number of observed pixels. In addition to numerical factors, the integral may also contribute factors of the "Coulomb logarithm," i.e. $\ln(l_{\text{max}}/l_{\text{min}})$, as found in the bispectrum calculation [35]. These additional correction are small, so we can argue that to $\mathcal{O}(1)$, the S/N is simply the product of the characteristic amplitude and the square-root of the number of observed pixels. We have established the necessary results used in the Introduction where we argued that the S/N of the estimator constructed out of the three-point correlation functions will be dominant.

[1] J. Maldacena, J. High Energy Phys. 05 (2003) 013.
[2] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, Phys. Rep. **367**, 1 (2002).
[3] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, Phys. Rep. **402**, 103 (2004).
[4] D. Babich, P. Creminelli, and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 08 (2004) 009.
[5] T. Pyne and S. M. Carroll, Phys. Rev. D **53**, 2920 (1996).
[6] N. Bartolo, S. Matarrese, and A. Riotto, J. Cosmol. Astropart. Phys. 01 (2004) 003.
[7] P. Creminelli and M. Zaldarriaga, Phys. Rev. D **70**, 083532 (2004).
[8] D. M. Goldberg and D. N. Spergel, Phys. Rev. D **59**, 103002 (1999).
[9] A. R. Cooray and W. Hu, Astrophys. J. **534**, 533 (2000).
[10] M. Zaldarriaga, Phys. Rev. D **62**, 063510 (2000).
[11] E. Komatsu, Ph.D. thesis, Tohuku University, 2001.
[12] D. H. Lyth, C. Ungarelli, and D. Wands, Phys. Rev. D **67**, 023503 (2003).
[13] G. Dvali, A. Gruzinov, and M. Zaldarriaga, Phys. Rev. D **69**, 023505 (2004).
[14] G. Dvali, A. Gruzinov, and M. Zaldarriaga, Phys. Rev. D **69**, 083505 (2004).
[15] E. Komatsu *et al.* Astrophys. J. Suppl. Ser. **148**, 119 (2003).
[16] M. Zaldarriaga (private communication).
[17] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall PTR, Englewood Cliffs, 1993).
[18] S. Zacks, *Parametric Statistical Inference: Basic Theory and Modern Applications* (Pergamon Press, New York, 1981).
[19] H. V. Poor, *An Introduction to Signal Detection and Estimation* (Springer-Verlag, New York, 1994).
[20] S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959).
[21] M. Kendall and A. Stuart, *The Advanced Theory of Statistics* (Macmillan, New York, 1979), Vol. 2.
[22] E. Komatsu, D. N. Spergel, and B. D. Wandelt, astro-ph/ 0305189.
[23] S. Smith *et al.*, Mon. Not. R. Astron. Soc. **352**, 887 (2004).
[24] M. G. Santos *et al.*, Mon. Not. R. Astron. Soc. **341**, 623 (2003).
[25] H. K. Eriksen, *et al.*, Astrophys. J. **612**, 64 (2004).
[26] L. Y. Chiang and P. D. Naselsky, astro-ph/0407395.
[27] L. Y. Chiang and P. Naselsky, and P. Coles, Astrophys. J. **602**, L1 (2004).
[28] A. F. Heavens, Mon. Not. R. Astron. Soc. **299**, 805 (1998).
[29] J. R. Bond, A. H. Jaffe, and L. Knox, Phys. Rev. D **57**, 2117 (1998).
[30] S. P. Oh, D. N. Spergel, and G. Hinshaw, Astrophys. J. **510**, 551 (1999).
[31] C. P. Ma and E. Bertschinger, Astrophys. J. **455**, 7 (1995).
[32] E. Komatsu and D. N. Spergel, Phys. Rev. D **63**, 063002 (2001).
[33] A. Taylor and P. Watts, Mon. Not. R. Astron. Soc. **328**, 1027 (2001).
[34] F. Bernardeau and L. Kofman, Astrophys. J. **443**, 479 (1995).
[35] D. Babich and M. Zaldarriaga, Phys. Rev. D **70**, 083005 (2004).