The CMB Bispectrum, Trispectrum, non-Gaussianity, and the Cramer-Rao Bound

Marc Kamionkowski

California Institute of Technology, Mail Code 350-17, Pasadena, CA 91125

Tristan L. Smith

Berkeley Center for Cosmological Physics, Physics Department, University of California, Berkeley, CA 94720

Alan Heavens

SUPA, Institute for Astronomy, School of Physics, University of Edinburgh,
Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, U.K.

(Dated: October 5, 2010)

Minimum-variance estimators for the parameter $f_{\rm nl}$ that quantifies local-model non-Gaussianity can be constructed from the cosmic microwave background (CMB) bispectrum (three-point function) and also from the trispectrum (four-point function). Some have suggested that a comparison between the estimates for the values of $f_{\rm nl}$ from the bispectrum and trispectrum allow a consistency test for the model. But others argue that the saturation of the Cramer-Rao bound by the bispectrum estimator implies that no further information on $f_{\rm nl}$ can be obtained from the trispectrum. Here we elaborate the nature of the correlation between the bispectrum and trispectrum estimators for $f_{\rm nl}$. We show that the two estimators become statistically independent in the limit of large number of CMB pixels and thus that the trispectrum estimator *does* indeed provide additional information on $f_{\rm nl}$ beyond that obtained from the bispectrum. We explain how this conclusion is consistent with the Cramer-Rao bound. Our discussion of the Cramer-Rao bound may be of interest to those doing Fisher-matrix parameter-estimation forecasts or data analysis in other areas of physics as well.

PACS numbers:

I.   INTRODUCTION

Observations of the cosmic microwave background (CMB) have confirmed a now 'standard' cosmological model [1]. A key aspect of this model is that primordial fluctuations are a realization of a Gaussian random field. This implies that CMB fluctuations are completely characterized by their two-point correlation function $C(\theta)$ in real space, or equivalently, the power spectrum C_ℓ in harmonic space. All higher-order N-point correlation functions with even N can be written in terms of the two-point function, and all N-point correlation functions with odd N are zero.

But while the simplest single-field slow-roll (SFSR) inflationary models assumed in the standard cosmological model predict departures from Gaussianity to be undetectably small [2], several beyond-SFSR models predict departures from Gaussianity to be larger [3], and possibly detectable with current or forthcoming CMB experiments. While the range of predictions for non-Gaussianity is large, the local model for non-Gaussianity [4]—that which appears in arguably the simplest beyond-SFSR models—has become the canonical model for most non-Gaussianity searches. The non-Gaussianity is parametrized in these models by a non-Gaussian amplitude $f_{\rm nl}$ to be defined more precisely below.

Most efforts to measure $f_{\rm nl}$ have relied on an estimator constructed from the CMB bispectrum, the three-point correlation function in harmonic space. However, the local model also predicts a non-zero trispectrum (the

harmonic-space four-point function) [5–9], and efforts have recently been mounted to determine $f_{\rm nl}$ from the trispectrum [10]. It has been suggested, moreover, that a comparison of the values of $f_{\rm nl}$ obtained from the bispectrum and trispectrum can be used as a consistency test for the local model [8, 10, 11].

However, it can be shown that the bispectrum estimator for $f_{\rm nl}$ saturates the Cramer-Rao bound, and it has been argued that this implies that no new information on the value of $f_{\rm nl}$, beyond that obtained from the bispectrum, can be obtained from the trispectrum [12, 13]. Ref. [13] further outlines the nature of the correlation between the bispectrum and trispectrum $f_{\rm nl}$ estimators implied by this conclusion.

Here we show that the trispectrum *does* provide additional information on $f_{\rm nl}$; i.e., it is *not* redundant with that from the bispectrum. We show that there is indeed a correlation between the bispectrum and trispectrum $f_{\rm nl}$ estimators, elaborating the arguments of Ref. [13]. However, we show with analytic estimates and numerical calculations that this correlation becomes weak in the high-statistics limit. We explain, with a simple example, how additional information on $f_{\rm nl}$ can be provided by the trispectrum given that the bispectrum estimator for $f_{\rm nl}$ saturates the Cramer-Rao bound. Put simply, the Cramer-Rao inequality bounds the variance with which a distribution can be measured, but there may be additional information in a distribution, about a theory or its parameters, beyond the distribution variance. The discussion of the Cramer-Rao bound and the examples we work out in Section II may be of interest to a much broader audience of readers than just those interested in

CMB non-Gaussianity.

The outline of this paper is as follows: We begin in Section II with our discussion of the Cramer-Rao bound. The aim of the rest of the paper is to illustrate explicitly the nature of the correlation between the bispectrum estimator for $f_{\rm nl}$ and the trispectrum estimator for $f_{\rm nl}{}^2$ and to show that the correlation becomes small in the high-statistics limit. In Section III we introduce our conventions for the bispectrum and trispectrum. In Section IV we derive the minimum-variance estimators for $f_{\rm nl}$ from the bispectrum and trispectrum and evaluate the noises in each. We also write down approximations for the estimators and noises valid for the local model. In Section V we explain the nature of the correlation between the bispectrum and trispectrum estimators for $f_{\rm nl}$. We then show that this correlation becomes weak (scaling with $(\ln N_{\rm pix})^{-1}$) as the number $N_{\rm pix}$ of pixels becomes large. We conclude in Section VI. Appendix A details the correspondence between continuum and discrete Fourier conventions for power spectra, bispectra, and trispectra, and Appendix B provides describes the numerical evaluation of the correlation.

II. THE CRAMER-RAO BOUND

In the Sections below we will demonstrate that the estimators for $f_{\rm nl}$ and $f_{\rm nl}{}^2$ becomes statistically independent with sufficiently good statistics. However, the bispectrum estimator for $f_{\rm nl}$ saturates the Cramer-Rao bound, and it has been argued that this saturation implies that no further information about $f_{\rm nl}$, beyond that obtained from the bispectrum, can be obtained from the trispectrum [12, 13]. Here we explain that the Cramer-Rao inequality bounds only the variance with which $f_{\rm nl}$ can be measured; additional information, beyond the variance, can be obtained from measurement of $f_{\rm nl}{}^2$ from the trispectrum.

To illustrate, consider, following Ref. [12], the analogous problem of determining $f_{\rm nl}$ and $f_{\rm nl}{}^2$ from a one-dimensional version of the local model. Suppose we have a random variable X written in terms of a Gaussian random variable x of zero mean ($\langle x \rangle = 0$) and unit variance ($\langle x^2 \rangle = 1$) as $X = x + \epsilon(x^2 - 1)$. Here, ϵ parametrizes the departure from the null hypothesis $\epsilon = 0$. The PDF for X, for a given ϵ, is

$$P(X|\epsilon) = \frac{1}{\sqrt{2\pi}} \left[\frac{e^{-x_+^2/2}}{1 + 2\epsilon x_+} + \frac{e^{-x_-^2/2}}{1 + 2\epsilon x_-} \right], \quad (1)$$

where

$$x_\pm = \frac{1}{2\epsilon} \left[\pm\sqrt{1 + 4\epsilon(X + \epsilon)} - 1 \right]. \quad (2)$$

The logarithm of the PDF can then be Taylor expanded about $\epsilon = 0$ as

$$\ln P(X|\epsilon) = -\frac{X^2}{2} + \epsilon I_1(X) - \frac{\epsilon^2}{2} I_2(X) + \mathcal{O}(\epsilon^3), \quad (3)$$

where $I_1(X) \equiv X^3 - 3X$, and $I_2(X) = 5X^4 + 5 - 14X^2$. It will be useful below to note that the expectation values of these quantities in the weakly non-Gaussian limit are $\langle I_1 \rangle = 6\epsilon + \mathcal{O}(\epsilon^3)$ and $\langle I_2 \rangle = 6 + 272\,\epsilon^2 + \mathcal{O}(\epsilon^4)$.

Now suppose we have a realization consisting of N data points X_i, each drawn independently from the distribution in Eq. (1), and let's arrange these data points into a vector $\mathbf{X}$. The PDF for this realization, for a given ϵ, is

$$\ln P(\mathbf{X}|\epsilon) = \sum_i \left[-\frac{X_i^2}{2} + \epsilon I_1(X_i) \right.$$
$$\left. -\frac{\epsilon^2}{2} I_2(X_i) + \mathcal{O}(\epsilon^3) \right]. \quad (4)$$

The Cramer-Rao inequality states that the smallest variance $\mathrm{Var}(\widehat{\epsilon}) \equiv \langle \widehat{\epsilon}^2 \rangle - \langle \widehat{\epsilon} \rangle^2$ to an estimator $\widehat{\epsilon}$ is

$$\mathrm{Var}(\widehat{\epsilon}) \geq \frac{1}{F}, \quad (5)$$

where

$$F = \int \left[\frac{\partial \ln P(\mathbf{X}|\epsilon)}{\partial \epsilon} \right]^2 P(\mathbf{X}|\epsilon) d\mathbf{X}$$
$$\equiv \left\langle \left[\frac{\partial \ln P(\mathbf{X}|\epsilon)}{\partial \epsilon} \right]^2 \right\rangle, \quad (6)$$

is the Fisher information. Here, the angle brackets denote an expectation value with respect to the null-hypothesis ($\epsilon = 0$) PDF. Applying Eq. (6) to Eq. (4), we find

$$F = \sum_i \left\langle [I_1(X_i)]^2 \right\rangle = 6N, \quad (7)$$

from which we infer

$$\mathrm{Var}(\widehat{\epsilon}) \geq \frac{1}{6\,N}. \quad (8)$$

This model predicts a skewness $\langle I_1 \rangle = \langle X^3 - 3X \rangle = 6\epsilon$, and so we can construct an estimator for ϵ from the measured skewness as follows:

$$\widehat{\epsilon}_s = \frac{1}{6N} \sum_i (X_i^3 - 3X_i). \quad (9)$$

The variance to this estimator is $\mathrm{Var}(\widehat{\epsilon}_{\rm s}) = (6N)^{-1}$, and so this estimator saturates the Cramer-Rao bound.

In retrospect, this saturation should come as no surprise. According to Eqs. (4) and (6), the Fisher information—and thus the minimum variance with which ϵ can be measured—is determined entirely by the term in $\ln P(\mathbf{X}|\epsilon)$ linear in ϵ which, in this case, is precisely the skewness. Thus, the terms in $\ln P(\mathbf{X}|\epsilon)$ that are higher order in ϵ contribute nothing to the Fisher information. And since the term linear in ϵ multiplies the skewness, $\widehat{\epsilon}_s$ saturates the Cramer-Rao bound.
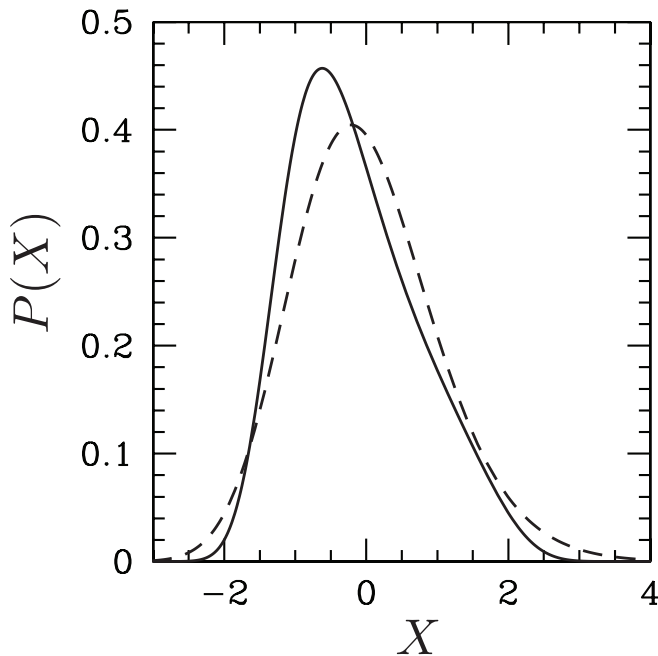
FIG. 1: Here we plot two probability distribution functions that share the same skewness but with two different values for the kurtosis.

But this does *not* mean that there is no information about ϵ from these higher-order terms. Consider, for example, a more general PDF,

$$\ln P_\alpha(X|\epsilon, \epsilon_1^2) = -\frac{X^2}{2} + \epsilon I_1(X)$$
$$-\frac{\epsilon_1^2}{2} I_2(X) + \mathcal{O}(\epsilon^3), \quad (10)$$

parametrized by ϵ_1^2, in addition to the parameter ϵ. This PDF differs from the PDF in Eq. (3) in the coefficient of $I_2(X)$. In the weakly non-Gaussian limit, the skewness of this PDF is $\langle I_1(X) \rangle = 6\epsilon$, and its "kurtosis" is $\langle I_2(X) \rangle = 6 + 846\,\epsilon^2 - 574\,\epsilon_1^2 + 18(\epsilon^2 - \epsilon_1^2)$.[1] If we fix ϵ, we then have a family of PDFs, parametrized by ϵ_1, that all have the same skewness but with different values of the kurtosis. Fig. 1 shows two PDFs that have the same skewness but different kurtoses. These are clearly two very different distributions; qualitatively, the large-X tails are suppressed as ϵ_1 is increased.

The estimator in Eq. (9) once again gives us the optimal estimator for ϵ in this new PDF, but we can now also measure from the data the kurtosis, the expectation value of $I_2(X)$, which provides an estimator for $846\,\epsilon^2 - 574\,\epsilon_1^2 + 18(\epsilon^2 - \epsilon_1^2)$. This can then be used in

---

[1] In this paper we use the term "kurtosis" to denote the expectation value of $I_2(X)$. This is qualitatively similar to, but slightly different, than the usual kurtosis, which is usually defined to be the expectation value of $X^4 - 6X^2 + 3$.

combination with the skewness estimator for ϵ to obtain an estimator for ϵ_1^2. According to the Cramer-Rao inequality, the smallest variance to ϵ_1^2 that can be obtained is

$$\mathrm{Var}(\epsilon_1^2) = \left\{ \int \left[\frac{\partial \ln P(\mathbf{X}|\epsilon, \epsilon_1^2)}{\partial(\epsilon_1^2)} \right]^2 \right.$$
$$\left. \times P(\mathbf{X}|\epsilon, \epsilon_1^2)\, dX \right\}^{-1} = \frac{1}{278\,N}. \quad (11)$$

Note that we cannot apply the Cramer-Rao bound to the parameter ϵ_1, rather than ϵ_1^2, as $\partial P(\mathbf{X}|\epsilon, \epsilon_1^2)/\partial \epsilon_1$ is zero under the null hypothesis $\epsilon_1 = 0$, thus violating one of the conditions for the Cramer-Rao inequality to apply. Since ϵ_1^2, not ϵ_1, is determined by the data, the distribution function for ϵ_1^2 (not ϵ_1) will approach a Gaussian distribution in the large-N limit.

The covariance between ϵ and ϵ_1^2 is zero, as the former is odd in X and the latter even. Still, this does not necessarily imply that the two are statistically independent, as there is still a covariance between ϵ^2 and ϵ_1^2. However, this becomes small as N becomes large. The correlation coefficient in this example is $r \equiv \mathrm{Cov}(\epsilon^2, \epsilon_1^2)/\sqrt{\mathrm{Var}(\epsilon^2)\mathrm{Var}(\epsilon_1^2)} \simeq 6\,N^{-1/2}$. Thus, for large N, ϵ and ϵ_1^2 are two statistically independent quantities that can be obtained from the data and then compared with the local-model prediction that $\epsilon_1^2 = \epsilon^2$. In brief, the skewness and kurtosis are two different quantities that can be obtained from a measured distribution. In the limit of large N, no measurement of the skewness, no matter how precise, can tell us anything about the kurtosis, and *vice versa*.

In this example, a one-sigma excursion in ϵ from a measurement with N data points is $\mathrm{Var}^{1/2}(\epsilon) = (6N)^{-1/2}$, and this is smaller than $\mathrm{Var}^{1/4}(\epsilon_1^2) = (278\,N)^{-1/4}$, the square root of the one-sigma excursion in ϵ_1^2, for any $N \gtrsim$ few. Thus, the skewness will provide better sensitivity if we are simply trying to detect a departure from the null hypothesis $\epsilon = 0$; measurement of ϵ_1^2 will not add much in this case. Still, if ϵ is measured with high statistical significance from the skewness, then measurement of ϵ_1^2 can, with sufficient statistics, provide a statistically independent determination of ϵ^2 and/or an independent test of the theory.

Now consider another PDF,

$$\ln P_{\mathrm{small}}(X|\epsilon) = -\frac{X^2}{2} + 10^{-2}\epsilon I_1(X)$$
$$-\frac{\epsilon^2}{2} I_2(X) + \mathcal{O}(\epsilon^3), \quad (12)$$

that differs from the local-model PDF in the suppression we have inserted for the term linear in ϵ, which thus suppresses the skewness. Application of the Cramer-Rao inequality in this case tells us that the smallest value of ϵ that can be distinguished from the null hypothesis ($\epsilon = 0$) is $10^2/\sqrt{6N}$, and we know from the discussion

above that this variance is obtained via measurement of the skewness. However, ϵ^2, the coefficient of the second term in the expansion—that obtained from measurement of the kurtosis—can be obtained with the variance given above. Thus, in this case, estimation of ϵ^2 via measurement of the kurtosis, provides a more sensitive probe of a departure from the null hypothesis $\epsilon = 0$ than does estimation of ϵ from measurement of the skewness, as long as $N \lesssim 10^7$. Note that the Cramer-Rao bound is not violated in this case, as measurement of ϵ^2, which does not discriminate between positive and negative values of ϵ, does not provide any further information on $\text{Var}(\epsilon)$. The *apparent* violation of the Cramer-Rao bound arises in this case because one of the conditions for the validity of the Cramer-Rao bound—that $\partial \ln P / \partial \epsilon$ be non-zero at $\epsilon = 0$ (under the null hypothesis)—is becoming invalid as the numerical coefficient of ϵ in $\ln P$ is made smaller. Had we chosen that coefficient to be zero, rather than 10^{-2}, then the Cramer-Rao inequality would have given a nonsensical bound for $\text{Var}(\epsilon)$.

A. Summary

Suppose we have a theory that predicts new effects parametrized by a quantity ϵ, with $\epsilon = 0$ representing the null hypothesis. A general PDF for the data $\mathbf{X}$ given ϵ (or likelihood for ϵ for given data $\mathbf{X}$) can be expanded in ϵ as $\ln P(X|\epsilon) = \ln P_0(X) + \epsilon g(X) + \epsilon^2 h(X) + \cdots$, where $P_0(X)$ is the PDF under the null hypothesis $\epsilon = 0$ and $g(X)$ and $h(X)$ are functions that describe the theory. Estimation of ϵ can be obtained through measurement of the mean value of $g(X)$, and an independent estimation of ϵ^2 can, with sufficiently good statistics, be obtained from measurement of the mean value of $h(X)$. If $\left\langle [g(X)]^2 \right\rangle^2 \gtrsim \left\langle [h(X)]^2 \right\rangle$, where the expectation value is with respect to P_0, then measurement of the mean value of $g(X)$ will provide a more sensitive avenue for detection of a value of ϵ that departs from the null hypothesis than measurement of the mean value of $h(X)$. If $\left\langle [g(X)]^2 \right\rangle^2 \lesssim \left\langle [h(X)]^2 \right\rangle$, then measurement of the mean value of $h(X)$ will provide a more sensitive test for detection of a value of ϵ that departs from the null hypothesis. If the two are comparable, then both tests will be comparable. In the case of a statistically-significant detection, there may be, given sufficient statistics, independent information on the values of ϵ and ϵ^2 from measurement of both moments. Care must be taken in interpreting results of measurement of ϵ^2 from $h(X)$, to note that the distribution of the $h(X)$ estimator for ϵ^2 is Gaussian in ϵ^2, not ϵ.

B. Local-model bispectrum and trispectrum

Similar arguments apply, *mutatis mutandis*, to measurement of the bispectrum and trispectrum, generalizations of the skewness and kurtosis: the estimator for f_{nl}

obtained from the bispectrum is statistically independent (for sufficiently large N_{pix}) from the estimator for f_{nl}^2 obtained from the trispectrum. If the variance to f_{nl} obtained from the bispectrum is comparable to the square root of the variance to f_{nl}^2 obtained from the trispectrum [6, 8], both will have roughly comparable sensitivities toward detection of a departure from the null hypothesis $f_{\text{nl}} = 0$. If there is a statistically significant detection, both can provide, with sufficiently good statistics, independent information on f_{nl} and f_{nl}^2, even if the bispectrum estimator for f_{nl} saturates the Cramer-Rao bound. We stop short of verifying these claims with the full likelihood for the local model. However, the arguments given explicitly for the one-dimensional analog above also apply to the skewness and kurtosis in the local model, the three- and four-point functions at zero lag, respectively. While the skewness and kurtosis are not optimal estimators for f_{nl} or f_{nl}^2, they are statistically independent quantities that are derived from the bispectrum and trispectrum, respectively.

C. Another example

Here we provide another example where statistically-independent information can be provided for estimators for ϵ and ϵ^2, where ϵ is a parameter that quantifies a departure from a null hypothesis. Suppose we want to test a theory in which the decay product from a polarized particle is predicted to have an angular distribution $P(\theta) \propto P_0(\theta) + \epsilon P_1(\theta) + \epsilon^2 P_2(\theta)$, where P_n are Legendre polynomials, and ϵ parametrizes the departure from the null hypothesis. In this case, measurement of the dipole, the mean value of $P_1(x)$, provides an estimator for ϵ, and measurement of the quadrupole, the mean value of $P_2(x)$, provides a statistically-independent (with sufficiently high statistics) estimator for ϵ^2. Thus, measurement of both the dipole and quadrupole can be used to test the data, even though the Cramer-Rao inequality tells us that $\text{Var}(\epsilon)$ is bounded by the value obtained from the dipole.

III. DEFINITIONS AND CONVENTIONS

We have argued above that the bispectrum estimator for f_{nl} and the trispectrum estimator for f_{nl}^2 may provide statistically independent information. The aim of the rest of the paper will be to evaluate explicitly the correlation between the bispectrum estimator for f_{nl} and the trispectrum estimator for f_{nl}^2. We will find that it is nonzero, but that it becomes small in the large-l_{max} limit.

We assume a flat sky to avoid the complications (e.g., spherical harmonics, Clebsch-Gordan coefficients, Wigner $3j$ and $6j$ symbols, etc.) associated with a spherical sky, and we further assume the Sachs-Wolfe limit. We denote the fractional temperature perturbation at posi-

tion $\vec{\theta}$ on a flat sky by $T(\vec{\theta})$, and refer to it hereafter simply as the temperature.

The temperature in the local model is written,

$$T(\vec{\theta}) = t(\vec{\theta}) + f_{\mathrm{nl}}[t(\vec{\theta})]^2, \tag{13}$$

in terms of a Gaussian random field $t(\vec{\theta})$. Note that our f_{nl} is three times the definition, in terms of the gravitational potential, used in most of the literature. We use this alternative definition to simplify the equations, but the difference should be noted if comparing our quantitative results with others. The field $t(\vec{\theta})$ has a power spectrum C_l given by

$$\left\langle t_{\vec{l}_1} t_{\vec{l}_2} \right\rangle = \Omega \delta_{\vec{l}_1+\vec{l}_2,0} C_l, \tag{14}$$

where $\Omega = 4\pi f_{\mathrm{sky}}$ is the survey area (in steradian), $t_{\vec{l}}$ is the Fourier transform of $t(\vec{\theta})$, and $\delta_{\vec{l}_1+\vec{l}_2,0}$ is a Kronecker delta that sets $\vec{l}_1 = -\vec{l}_2$. In the limit $f_{\mathrm{nl}}T \ll 1$ (current constraints are $f_{\mathrm{nl}}T \lesssim 10^{-3}$), C_l is also the power spectrum for $T(\vec{\theta})$.

The bispectrum $B(l_1, l_2, l_3)$ is defined by

$$\left\langle T_{\vec{l}_1} T_{\vec{l}_2} T_{\vec{l}_3} \right\rangle = \Omega \delta_{\vec{l}_1+\vec{l}_2+\vec{l}_3,0} B(l_1, l_2, l_3). \tag{15}$$

The Kronecker delta insures that the bispectrum is defined only for $\vec{l}_1 + \vec{l}_2 + \vec{l}_3 = 0$; i.e., only for triangles in Fourier space. Statistical isotropy then dictates that the bispectrum depends only on the magnitudes l_1, l_2, l_3 of the three sides of this Fourier triangle. The bispectrum for the local model is,

$$B(l_1, l_2, l_3) = 2 f_{\mathrm{nl}}[C_{l_1} C_{l_2} + C_{l_1} C_{l_3} + C_{l_2} C_{l_3}]. \tag{16}$$

Likewise, the trispectrum is defined by

$$\left\langle T_{\vec{l}_1} T_{\vec{l}_2} T_{\vec{l}_3} T_{\vec{l}_4} \right\rangle = \Omega \delta_{\vec{l}_1+\vec{l}_2+\vec{l}_3+\vec{l}_4,0} \mathcal{T}(\vec{l}_1, \vec{l}_2, \vec{l}_3, \vec{l}_4), \tag{17}$$

and for the local model,

$$\begin{aligned} \mathcal{T}(\vec{l}_1, \vec{l}_2, \vec{l}_3, \vec{l}_4) =\ & f_{\mathrm{nl}}{}^2 \left[P_{l_3 l_4}^{l_1 l_2}(|\vec{l}_1 + \vec{l}_2|) \right. \\ & \left. + P_{l_2 l_4}^{l_1 l_3}(|\vec{l}_1 + \vec{l}_3|) + P_{l_2 l_3}^{l_1 l_4}(|\vec{l}_1 + \vec{l}_4|) \right], \end{aligned} \tag{18}$$

where

$$\begin{aligned} P_{l_3 l_4}^{l_1 l_2}(|\vec{l}_1 + \vec{l}_2|) =\ & 4 C_{|\vec{l}_1+\vec{l}_2|} \left[C_{l_1} C_{l_3} + C_{l_1} C_{l_4} \right. \\ & \left. + C_{l_2} C_{l_3} + C_{l_2} C_{l_4} \right]. \end{aligned} \tag{19}$$

Again, the trispectrum is nonvanishing only for $\vec{l}_1 + \vec{l}_2 + \vec{l}_3 + \vec{l}_4 = 0$, that is, only for quadrilaterals in Fourier space.

IV. MINIMUM-VARIANCE NON-GAUSSIANITY ESTIMATORS

We now review how to measure f_{nl} from the bispectrum and the trispectrum. To keep our arguments clear (and since the current goal is simply detection of a departure from non-Gaussianity, rather than precise evaluation of f_{nl}), we assume the null hypothesis $f_{\mathrm{nl}} = 0$ in the evaluation of noises and construction of estimators. The generalization to nonzero f_{nl} is straightforward [13].

A. The bispectrum

From Eqs. (15) and (16), each triangle $\vec{l}_1 + \vec{l}_2 + \vec{l}_3 = 0$ gives an estimator,

$$(\widehat{f_{\mathrm{nl}}{}^b})_{123} = \frac{T_{\vec{l}_1} T_{\vec{l}_2} T_{\vec{l}_3}}{\Omega B(l_1, l_2, l_3)/f_{\mathrm{nl}}}, \tag{20}$$

with variance [using Eq. (14)],[2]

$$\frac{\Omega^3 C_{l_1} C_{l_2} C_{l_3}}{[\Omega B(l_1, l_2, l_3)/f_{\mathrm{nl}}]^2}. \tag{21}$$

The minimum-variance estimator is constructed by adding all of these estimators with inverse-variance weighting. It is

$$\widehat{f_{\mathrm{nl}}{}^b} = \sigma_b^2 \sum \frac{T_{\vec{l}_1} T_{\vec{l}_2} T_{\vec{l}_3} B(l_1, l_2, l_3)/f_{\mathrm{nl}}}{\Omega^2 C_{l_1} C_{l_2} C_{l_3}}, \tag{22}$$

and it has inverse variance,

$$\sigma_b^{-2} = \sum \frac{[B(l_1, l_2, l_3)/f_{\mathrm{nl}}]^2}{\Omega C_{l_1} C_{l_2} C_{l_3}}. \tag{23}$$

The sums in Eqs. (22) and (23) are taken over all *distinct* triangles with $\vec{l}_1 + \vec{l}_2 + \vec{l}_3 = 0$. We may then take $\vec{L} \equiv \vec{l}_3$ to be the shortest side of the triangle—i.e., $l_1, l_2 > L$—and re-write the estimator as,

$$\begin{aligned} \widehat{f_{\mathrm{nl}}{}^b} =\ & \frac{1}{2} \sigma_b^2 \sum_{\vec{L}} \frac{1}{C_L} \\ & \times \sum_{\vec{l}_1+\vec{l}_2=-\vec{L},\, l_1,l_2>L} \frac{T_{\vec{l}_1} T_{\vec{l}_2} T_{\vec{L}} B(l_1, l_2, L)/f_{\mathrm{nl}}}{\Omega^2 C_{l_1} C_{l_2}}, \end{aligned} \tag{24}$$

---

[2] Here we ignore the negligible contributions from triangles and for the trispectrum below, quadrilaterals, where two sides have the same length. We do, however, include these configurations in the numerical analysis described in Appendix B and verify that this assumption is warranted.

and the inverse-variance as

$$\sigma_b^{-2} = \frac{1}{2} \sum_{\vec{L}} \frac{1}{C_L} \sum_{\vec{l}_1 + \vec{l}_2 = -\vec{L}, \, l_1, l_2 > L} \frac{[B(l_1, l_2, L)/f_{\rm nl}]^2}{\Omega C_{l_1} C_{l_2}}.$$

(25)

The factor of $1/2$ is included to account for double counting of identical triangles, those with $\vec{l}_1 \leftrightarrow \vec{l}_2$.

1. Approximation to the Bispectrum Estimator

Now consider the variance σ_b^2 with which $f_{\rm nl}$ can be measured from the bispectrum. Take $C_l = A/l^2$ for the power spectrum, where $A \simeq 6 \times 10^{-10}$ is the power-spectrum normalization. The bispectrum in Eq. (16) is maximized for squeezed triangles, those with $L \ll l_1, l_2$, and thus with $l_1 \simeq l_2$. In this limit, the bispectrum can be approximated $B(l_1, l_2, L) \simeq 4A^2 f_{\rm nl} L^{-2} l_1^{-2}$. Then, from Eq. (25) the inverse variance (and thus the signal-to-noise) is dominated by squeezed triangles, and it is furthermore dominated by those triangles with the modes $\vec{L}$ of the *smallest* magnitudes L.
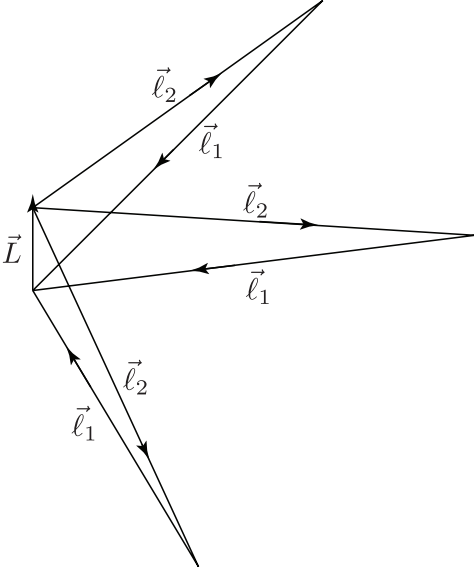

FIG. 2: Three triangles that all share a shortest side $\vec{L}$.

More precisely, let us evaluate the contribution $(\sigma_b^{-2})_{\vec{L}}$ to the inverse variance obtained from all triangles that share the same shortest side $\vec{L}$, as shown in Fig. 2. Since this contribution is dominated by modes with $\vec{l}_1 \simeq \vec{l}_2$, the inverse-variance from these triangles is,

$$(\sigma_b^{-2})_{\vec{L}} \simeq \frac{1}{2\Omega} \frac{L^2}{A} \sum_{\vec{l}_1} \frac{(4C_L C_{l_1})^2}{C_{l_1}^2} = \frac{8A}{\Omega L^2} \sum_{\vec{l}} 1$$

$$\simeq \frac{8A}{L^2} \frac{1}{2\pi} \int_L^{l_{\rm max}} l \, dl \simeq \frac{2A}{\pi L^2} l_{\rm max}^2,$$

(26)

where we have used $\sum_{\vec{l}} = \Omega \int d^2 l / (2\pi)^2$ in the last line.

The full estimator then sums over all $\vec{L}$ as in Eq. (24). The full inverse-variance is then

$$\sigma_b^{-2} = \sum_{\vec{L}} (\sigma_b^{-2})_{\vec{L}} = \Omega \int \frac{d^2 L}{(2\pi)^2} (\sigma_b^{-2})_{\vec{L}}$$

$$\simeq \frac{A\Omega}{\pi^2} l_{\rm max}^2 \ln \frac{L_{\rm max}}{L_{\rm min}}$$

$$\simeq \frac{4A f_{\rm sky} l_{\rm max}^2}{\pi} \ln \frac{L_{\rm max}}{L_{\rm min}},$$

(27)

in agreement with Ref. [14].

To summarize: (1) the signal-to-noise is greatly dominated by triangles with one side much shorter than the other two. (2) The signal-to-noise is dominated primarily by those with the smallest short side. (3) The contribution to the full signal-to-noise is equal per logarithmic interval of L, the magnitude of the smallest mode in the triangle. (4) Even if there is a huge number of triangles that enter the estimator, the error in the estimator is still dominated by the cosmic variance associated with the values of $T_{\vec{L}}$ for the $\vec{L}$ modes of the smallest L.

Since the variance is dominated by squeezed triangles, we can approximate the estimator, Eq. (24), as

$$\widehat{f_{\rm nl}}^b = \frac{2\sigma_b^2}{A\Omega^2} \sum_{\vec{L}} T_{\vec{L}} X_{\vec{L}},$$

(28)

where

$$X_{\vec{L}} \equiv \sum_{\vec{l}} T_{\vec{l}} T_{-\vec{L}-\vec{l}} \, l^2.$$

(29)

B. The trispectrum

Now consider the trispectrum. Each distinct quadrilateral $\vec{l}_1 + \vec{l}_2 + \vec{l}_3 + \vec{l}_4 = 0$ gives an estimator for the trispectrum with some variance. Adding the individual estimators with inverse-variance weighting gives the minimum-variance estimator,[3]

$$\widehat{(f_{\rm nl}^2)}^t = \sigma_t^2 \sum \frac{T_{\vec{l}_1} T_{\vec{l}_2} T_{\vec{l}_3} T_{\vec{l}_4} \mathcal{T}(\vec{l}_1, \vec{l}_2, \vec{l}_3, \vec{l}_4)/f_{\rm nl}^2}{\Omega^3 C_{l_1} C_{l_2} C_{l_3} C_{l_4}},$$

(30)

and the inverse variance,

$$\sigma_t^{-2} = \sum \frac{\left[\mathcal{T}(\vec{l}_1, \vec{l}_2, \vec{l}_3, \vec{l}_4)/f_{\rm nl}^2 \right]^2}{\Omega^2 C_{l_1} C_{l_2} C_{l_3} C_{l_4}}.$$

(31)

The sums here are over all distinct quadrilateral $\vec{l}_1 + \vec{l}_2 + \vec{l}_3 + \vec{l}_4 = 0$, and we again neglect quadrilaterals where two or more sides are the same.

---

[3] Strictly speaking, one must subtract the connected part of the trispectrum. We omit this term to keep our expression compact, but it is included in the analytic and numerical calculations of the variances and covariances discussed below.

Each quadrilateral will have a smallest diagonal, which we call $\vec{L}$. The quadrilateral is then described by two triangles that each share their smallest side $\vec{L}$; the two sides of the first triangle will be $\vec{l}_1$ and $\vec{l}_2$ and the two sides of the second triangle will be $\vec{l}_3$ and $\vec{l}_4$. We can then re-write the sums in Eqs. (30) and (31) as

$$\sum_{\vec{L}} \sum_{\vec{l}_1 + \vec{l}_2 = \vec{L}} \sum_{\vec{l}_3 + \vec{l}_4 = -\vec{L}} . \qquad (32)$$

The sum here is only over combinations of $\{\vec{l}_1, \vec{l}_2, \vec{l}_3, \vec{l}_4\}$ where the lengths of the two other diagonals, $|\vec{l}_1 + \vec{l}_4| = |\vec{l}_2 + \vec{l}_3|$ and $|\vec{l}_2 + \vec{l}_4| = |\vec{l}_1 + \vec{l}_3|$, are both $> L$, so that L is the shortest diagonal [cf. Eq. (18)].
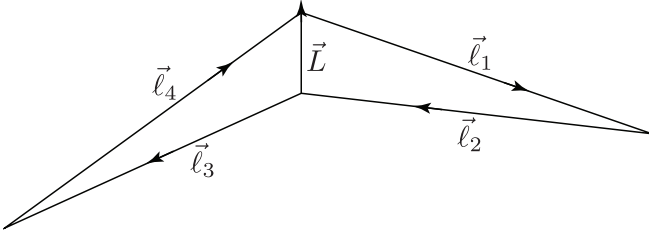

FIG. 3: An example of an elongated quadrilateral with a shortest diagonal $\vec{L}$. Note that it is equivalent to two elongated triangles that share the same shortest side $\vec{L}$.

Let's now consider the local-model trispectrum given in Eqs. (18) and (19). The three terms in Eq. (18) sum over the three diagonals of the quadrilateral. Eq. (19) then shows that each of these terms is the product of the power spectrum C_L evaluated for the diagonal (e.g., $\vec{L} = \vec{l}_1 + \vec{l}_2 = -\vec{l}_3 - \vec{l}_4$) times a sum of products of power spectra evaluated for each of the quadrilateral sides. The quadrilateral is thus maximized for highly elongated quadrilaterals, those with $l_i \gg L$, with one short diagonal, as shown in Fig. 3. The trispectrum for these elongated quadrilaterals may be approximated as $\mathcal{T}(\vec{l}_1, \vec{l}_2, \vec{l}_3, \vec{l}_4) \simeq 16 f_{\rm nl}{}^2 C_L C_{l_1} C_{l_3}$.

Now consider the contribution $(\sigma_t^{-2})_{\vec{L}}$ to the inverse variance from all quadrilaterals that share the same shortest diagonal $\vec{L}$. Using Eq. (31) and approximating the trispectrum by the squeezed limit, this is

$$\begin{aligned}
(\sigma_t^{-2})_{\vec{L}} &\simeq \frac{1}{8} \sum_{\vec{l}_1} \sum_{\vec{l}_3} \frac{(16 C_L C_{l_1} C_{l_3})^2}{\Omega^2 (C_{l_1} C_{l_3})^2} \\
&= \frac{32 A^2}{\Omega^2 L^4} \left(\sum_{\vec{l}} 1 \right)^2 = \frac{2}{\pi^2} \frac{A^2}{L^4} l_{\rm max}^4. \quad (33)
\end{aligned}$$

The factor $1/8$ in the first line accounts for the $\vec{l}_1 \leftrightarrow \vec{l}_2$ and $\vec{l}_3 \leftrightarrow \vec{l}_4$ symmetries and the symmetry under interchange of the $(\vec{l}_1, \vec{l}_2)$ and $(\vec{l}_3, \vec{l}_4)$ triangles. Again, the full variance is obtained by summing over $\vec{L}$ modes. Thus,

$$\sigma_t^{-2} \simeq \frac{2 f_{\rm sky}}{\pi^2} \frac{A^2}{L_{\rm min}^2} l_{\rm max}^4. \qquad (34)$$

Note that we obtain the $l_{\rm max}^{-4}$ scaling of the variance noted in Ref. [8]. Recall that σ_t^2 is a variance to $f_{\rm nl}{}^2$ (rather than $f_{\rm nl}$). Thus, the ratio of the smallest $f_{\rm nl}$ detectable via the trispectrum to the smallest detectable via the bispectrum is $\sqrt{\sigma_t/\sigma_b^2} \simeq 1.7 f_{\rm sky}^{1/4} [L_{\rm min} \ln(L_{\rm max}/L_{\rm min})]^{1/2}$. For reasonable values of $L_{\rm min}$ and $L_{\rm max}$, the smallest $f_{\rm nl}$ detectable with the bispectrum is smaller, by a factor of order a few, than that detectable with the trispectrum [6–8].

We can now derive an approximation for $\widehat{(f_{\rm nl}{}^2)^t}$ noting that the variance, and thus the signal-to-noise, is dominated by equilateral triangles. From Eq. (30), and using the squeezed limit for the trispectrum, we find,

$$\widehat{(f_{\rm nl}{}^2)^t} = \frac{2}{3} \sigma_t^2 \sum_{\vec{L}} \frac{1}{L^2} X_{\vec{L}}^2, \qquad (35)$$

where $X_{\vec{L}}$ is the quantity given in Eq. (29). Comparing with the estimator, Eq. (28), we see that *this estimator is constructed from precisely the same sums of triangles as the bispectrum estimator*. Strictly speaking, the bispectrum estimator for $f_{\rm nl}$ involves a sum over a huge number of triangles; the number of such triangles scales as $N_{\rm pix}^2/6$ with the number of pixels in the map. Likewise, the trispectrum estimator for $f_{\rm nl}{}^2$ involves a sum over all quadrilaterals, and the number of these scales as $N_{\rm pix}^3/24$. Thus, one naively expects the correlation between the estimators to be extremely weak, given the huge number of bispectrum and trispectrum configurations. Eqs. (28) and (35) show, however, that the quadrilateral configurations that dominate the trispectrum estimator for $f_{\rm nl}{}^2$ are very closely related to the triangle configurations that dominate the bispectrum estimator for $f_{\rm nl}$.

V. CORRELATION BETWEEN BISPECTRUM AND TRISPECTRUM ESTIMATORS FOR $f_{\rm nl}$

Since the bispectrum and trispectrum estimators for $f_{\rm nl}$ are both constructed from the same CMB map, it is expected that there should be some correlation between the two estimators. Eqs. (28) and (35) help clarify the nature of the correlation. Clearly, if we use for the bispectrum estimator only triangles that share a single shortest side $\vec{L}$ and for the trispectrum estimator only quadrilaterals with the same $\vec{L}$ as the shortest diagonal, then the two estimators provide the same quantity, modulo the difference between the magnitude $|T_{\vec{L}}|^2$ (from the bispectrum estimator) and its expectation value A/L^2 (from the trispectrum estimator).

However, we have not only triangles/quadrilaterals from a single $\vec{L}$ shortest side/diagonal, but those constructed from many $\vec{L}$'s. The correlation between the bispectrum and trispectrum estimators should thus decrease as the number of $\vec{L}$ modes increases in the same way that the means $\langle x \rangle$ and $\langle x^2 \rangle$ measured with a large

number N of data points x_i will become uncorrelated as N becomes large.

Of course since $\widehat{f_{\mathrm{nl}}}^{\,b}$ is linear in $T_{\vec{L}}$, the covariance between $\widehat{f_{\mathrm{nl}}}^{\,b}$ and $\widehat{(f_{\mathrm{nl}}^2)}^t$ will be zero. However, the correlation between $(\widehat{f_{\mathrm{nl}}}^{\,b})^2$ and $\widehat{(f_{\mathrm{nl}}^2)}^t$ will be nonzero. We thus now estimate the magnitude of the correlation coefficient, which we define as

$$r \equiv \frac{\left\langle \Delta\left((\widehat{f_{\mathrm{nl}}}^{\,b})^2\right)\Delta\left(\widehat{(f_{\mathrm{nl}}^2)}^t\right)\right\rangle}{\left\langle \left[\Delta\left((\widehat{f_{\mathrm{nl}}}^{\,b})^2\right)\right]^2\right\rangle^{1/2}\left\langle \left[\Delta\left(\widehat{(f_{\mathrm{nl}}^2)}^t\right)\right]^2\right\rangle^{1/2}},\quad (36)$$

where $\Delta(Q) \equiv Q - \langle Q\rangle$. To simplify the equations, we can drop the prefactors in Eqs. (28) and (35) and deal with quantities,

$$F \equiv \sum_{\vec{L}} T_{\vec{L}} X_{\vec{L}}, \qquad G \equiv \sum_{\vec{L}} \frac{A}{L^2}X_{\vec{L}}^2. \qquad (37)$$

The desired correlation coefficient is then

$$r = \frac{\left\langle \Delta(F^2)\Delta G\right\rangle}{\left\langle [\Delta(F^2)]^2\right\rangle^{1/2}\left\langle (\Delta G)^2\right\rangle^{1/2}}. \qquad (38)$$

We begin by noting that $X_{\vec{L}}$ is a random variable with zero mean. In the large-$l_{\max}$ limit, it will be well approximated by a Gaussian random variable, in which case $\left\langle X_{\vec{L}}^4\right\rangle = 3\left\langle X_{\vec{L}}^2\right\rangle^2$. Some other useful relations include,

$$\left\langle F^2\right\rangle = \sum_{\vec{L}_1,\vec{L}_2}\left\langle T_{\vec{L}_1}T_{\vec{L}_2}X_{\vec{L}_1}X_{\vec{L}_2}\right\rangle = \Omega\sum_{\vec{L}}\frac{A}{L^2}\left\langle X_{\vec{L}}^2\right\rangle, \qquad (39)$$

$$\left\langle G\right\rangle = \sum_{\vec{L}}\frac{A}{L^2}\left\langle X_{\vec{L}}^2\right\rangle = \left\langle F^2\right\rangle/\Omega, \qquad (40)$$

$$\left\langle G^2\right\rangle = \sum_{\vec{L}_1,\vec{L}_2}\frac{A^2}{L_1^2 L_2^2}\left\langle X_{\vec{L}_1}^2 X_{\vec{L}_2}^2\right\rangle = \left\langle G\right\rangle^2 + 2\sum_{\vec{L}}\frac{A^2}{L^4}\left\langle X_{\vec{L}}^2\right\rangle^2, \qquad (41)$$

$$\left\langle F^2 G\right\rangle = \sum_{\vec{L}_1}\sum_{\vec{L}_2}\sum_{\vec{L}_3}\frac{A}{L_1^2}\left\langle T_{\vec{L}_2}T_{\vec{L}_3}\right\rangle\left\langle X_{\vec{L}_1}^2 X_{\vec{L}_2}X_{\vec{L}_3}\right\rangle$$
$$= \Omega\sum_{\vec{L}_1,\vec{L}_2}\frac{A}{L_1^2}\frac{A}{L_2^2}\left\langle X_{\vec{L}_1}^2 X_{\vec{L}_2}^2\right\rangle = \Omega\left\langle G^2\right\rangle. \qquad (42)$$

Also, since F is a sum over (approximately) Gaussian random variables, it is also well approximated by a Gaussian random variable, and so $\left\langle F^4\right\rangle \simeq 3\left\langle F^2\right\rangle^2$.

From these relations, it follows that

$$\left\langle \Delta(F^2)\Delta G\right\rangle = \left\langle F^2 G\right\rangle - \left\langle F^2\right\rangle\left\langle G\right\rangle = \Omega\left[\left\langle G^2\right\rangle - \left\langle G\right\rangle^2\right], \qquad (43)$$

and thus that

$$r = \frac{\Omega\left\langle (\Delta G)^2\right\rangle^{1/2}}{\sqrt{2}\left\langle F^2\right\rangle} = \frac{\left(\sum_{\vec{L}}L^{-4}\right)^{1/2}}{\sum_{\vec{L}}L^{-2}}$$
$$= \left[2\sqrt{\pi f_{\mathrm{sky}}}L_{\min}\ln\left(L_{\max}/L_{\min}\right)\right]^{-1}. \qquad (44)$$

Thus, if $L_{\max}$ is small, then the correlation will be large. However, the correlation coefficient decreases as $[\ln(L_{\max})]^{-1}$, and it will become negligible in the limit that $L_{\max}$ is large.

Strictly speaking, the $X_{\vec{L}}$ are not entirely statistically independent, as we have assumed here, as many are constructed from the same measurements. They are also not perfectly Gaussian, as we have assumed. However, as we discuss in Appendix B, we have checked with a full numerical calculation of the correlation coefficient that the basic conclusions—and particularly the scaling of the correlation coefficient r with $L_{\max}$—are sound.

VI. CONCLUSIONS

A large body of recent work has focused on tests of the local model for non-Gaussianity that can be performed with measurement of the CMB trispectrum and bispectrum. Here we have clarified how the bispectrum and trispectrum may provide statistically independent information on the local-model non-Gaussianity parameter f_{nl}, even if the bispectrum estimator for f_{nl} saturates the Cramer-Rao bound. The basic point is that the Cramer-Rao inequality puts a lower limit to the variance with which a given parameter can be measured. If the likelihood function is precisely Gaussian, then the likelihood is described entirely by the variance. However, if the likelihood function is not precisely Gaussian, then there is more information in the likelihood beyond the variance (see, e.g., Section VI in Ref. [15]). In the current problem, this is manifest in that a statistically-independent measurement of f_{nl}^2 can be obtained from the trispectrum without contributing to the variance of f_{nl}.

We then built on an observation of Ref. [13] to illustrate the nature of the correlation between the bispectrum estimator for f_{nl} and the trispectrum estimator of f_{nl}^2. This analysis demonstrates that the two estimators do indeed become statistically independent in the large-$l_{\max}$ limit.

Throughout we have made the null hypothesis $f_{\mathrm{nl}} = 0$ to estimate the variances with which f_{nl} can be measured from the bispectrum and with which f_{nl}^2 can be measured from the trispectrum. This is suitable if one is simply searching the data for departures from the null hypothesis. However, as emphasized by Ref. [13], the minimum-variance estimators constructed under the null hypothesis are no longer optimal if there is a strong signal. If so, then forecasts of signal-to-noise made with the null hypothesis are no longer valid in the limit of large signal-to-noise, and this calls into question claims

[8] that the trispectrum will provide a better probe of the local model in the large-S/N limit. In this limit, a new bispectrum estimator can be constructed to saturate the Cramer-Rao bound [13], and an analogous optimal trispectrum estimator can in principle be found. Still, the observation that the bispectrum and trispectrum estimators in the local model are constructed from the same sums of triangles suggests that the precisions with which $f_{\rm nl}$ can be measured, in the high-S/N limit, from the bispectrum and trispectrum will be roughly comparable.

Although we assumed the null hypothesis to argue that the bispectrum and trispectrum estimators for $f_{\rm nl}$ are independent, the same arguments should also apply in the high-S/N limit. For example, if the bispectrum estimator finds $f_{\rm nl}$ to be different from zero, with best-fit value $\bar{f}_{\rm nl}$, then the likelihood can be re-parametrized in terms of a quantity $\epsilon = f_{\rm nl} - \bar{f}_{\rm nl}$ that quantifies the departure from the new null hypothesis $f_{\rm nl} = \bar{f}_{\rm nl}$. Measurement of ϵ with the trispectrum can then be used to provide a statistically independent consistency check of the model. Or, in simpler terms, the skewness and kurtosis are still two statistically independent quantities that can be obtained from a measured distribution, even if the skewness (or kurtosis) of that distribution is nonzero.

Throughout, we have made approximations and simplifications to make the basic conceptual points clear, and we have restricted our attention simply to the local model, which we have here defined to be $\Phi = \phi + f_{\rm nl}(\phi^2 - \langle \phi^2 \rangle)$. However, inflationary models predict a wider range of trispectra [16]. Likewise, analysis of real data will introduce a number of ingredients that we have excised from our simplified analysis. Still, we hope that the points we have made here may assist in the interpretation and understanding of experimental results and perhaps elucidate statistical tests of other, more general, non-Gaussian models.

Acknowledgments

Appendix A: The continuum-discretuum connection

In this paper we have chosen to work with discrete Fourier transforms where the calculations of variances and covariances are more straightforward. Here we show how to derive the expressions for power spectra, bispectra, and trispectra for this discrete formalism to the continuum analysis discussed in most of the theoretical literature.

Following Ref. [13], we note that

$$T_{\vec{l}} = \int d^2\vec{\theta}\, e^{-i\vec{l}\cdot\vec{\theta}} T(\vec{\theta}) \simeq \frac{\Omega}{N_{\rm pix}} \sum_{\vec{\theta}} e^{-i\vec{l}\cdot\vec{\theta}} T(\vec{\theta}), \quad (\text{A1})$$

where $\Omega = 4\pi f_{\rm sky}$ is the area of sky (in steradians) surveyed, from which we infer the correspondence $\sum_{\vec{\theta}} \Leftrightarrow (N_{\rm pix}/\Omega)\int d^2\vec{\theta}$. Likewise,

$$T(\vec{\theta}) = \int \frac{d^2\vec{l}}{(2\pi)^2} e^{i\vec{l}\cdot\vec{\theta}} T_{\vec{l}} \simeq \frac{1}{\Omega}\sum_{\vec{l}} e^{-i\vec{l}\cdot\vec{\theta}} T(\vec{\theta}), \quad (\text{A2})$$

from which we infer the correspondence $\sum_{\vec{l}} \Leftrightarrow \Omega \int d^2\vec{l}/(2\pi)^2$. The Dirac delta function is then written in the discrete formalism as a Kronecker delta as follows:

$$(2\pi)^2 \delta(\vec{l}-\vec{l}') = \int d^2\vec{\theta} e^{i\vec{\theta}\cdot(\vec{l}-\vec{l}')} \simeq \frac{\Omega}{N_{\rm pix}}\sum_{\vec{\theta}} e^{i\vec{\theta}\cdot(\vec{l}-\vec{l}')} = \Omega\delta_{\vec{l},\vec{l}'}.$$
$$(\text{A3})$$

The definitions in Section III of the power spectrum, bispectrum, and trispectrum follow from this relation.

One advantage of this formulation is that equations can be checked for consistency using dimensional analysis. Recalling that θ has units $[\theta^2]$ =sterad and that temperature has units $[T(\vec{\theta})]$=K, it follows, for example, that $[T_{\vec{l}}]$ =K-sterad, $[C_l]$ =K^2-sterad, $[f_{\rm nl}]$ =K^{-1}, $[B]$ =K^3-sterad^2, and $[\mathcal{T}]$ =K^4-sterad^3. As another check, the variance and covariances should have an appropriate scaling with $f_{\rm sky}$ if factors of Ω are carried properly through the calculation.

Appendix B: Full correlation between trispectrum and bispectrum estimators

As discussed in the text, the minimum-variance bispectrum and trispectrum estimators for $f_{\rm nl}$ are given by

$$\widehat{f_{\rm nl}}^b = \sigma_{\rm b}^2 \sum_{\vec{l}_1+\vec{l}_2+\vec{l}_3=0} \frac{B(l_1,l_2,l_3)}{3!\Omega^2 C_{l_1}C_{l_2}C_{l_3}} T_{\vec{l}_1}T_{\vec{l}_2}T_{\vec{l}_3}, (\text{B1})$$

$$\widehat{(f_{\rm nl}^2)}^t = \sigma_{\rm t}^2 \sum_{\vec{l}_1+\vec{l}_2+\vec{l}_3+\vec{l}_4=0} \frac{\mathcal{T}(\vec{l}_1,\vec{l}_2,\vec{l}_3,\vec{l}_4)}{4!\Omega^3 C_{l_1}C_{l_2}C_{l_3}C_{l_4}}$$
$$\times\, T_{\vec{l}_1}T_{\vec{l}_2}T_{\vec{l}_3}T_{\vec{l}_4}. \quad (\text{B2})$$

where $\sigma_{\rm b,t}^2$ are the variances of the bispectrum and trispectrum estimator. Here we sum over all triangles and quadrilaterals (not just those with no equal sides), and the factors of 3! and 4! take into account double counting of degenerate terms in the sum and permutation factors for triangles and quadrilaterals with equal sides. In Sec. V we used the squeezed-limit approximation to estimate the correlation coefficient between $(\widehat{f_{\rm nl}}^b)^2$ and $\widehat{(f_{\rm nl}^2)}^t$. In this Appendix we derive the full expression for this correlation coefficient and verify that the approximations made in Sec. V are valid.

The covariance will consist of a weighted sum of the 10-point function. However, because of the fact that no two indices in the trispectrum or each bispectrum estimator can add to zero we know that two of the bispectrum indices must combine. The rest of the covariance will then be diagonal leading to

$$\left\langle \widehat{(f_{\rm nl}^{\,b})^2}\, \widehat{(f_{\rm nl}^{\,2})^t} \right\rangle = \frac{\sigma_b^4 \sigma_t^2}{4}$$

$$\times \sum_{\vec{l}_1+\vec{l}_2=-\vec{L},\vec{l}_3+\vec{l}_4=\vec{L}} \frac{B(L,l_1,l_2)B(L,l_3,l_4)\mathcal{T}(\vec{l}_1,\vec{l}_2,\vec{l}_3,\vec{l}_4)/f_{\rm nl}^{\,4}}{\Omega^2 C_L C_{l_1} C_{l_2} C_{l_3} C_{l_4}}.$$

$$(B3)$$

Finally, we need to compute the variance of $\widehat{(f_{\rm nl}^{\,b})^2}$. To do this we must compute the 12-point function

$$\left\langle T_{\vec{l}_1} T_{\vec{l}_2} T_{\vec{l}_3} | T_{\vec{l}_4} T_{\vec{l}_5} T_{\vec{l}_6} | T_{\vec{t}_1} T_{\vec{t}_2} T_{\vec{t}_3} | T_{\vec{t}_4} T_{\vec{t}_5} T_{\vec{t}_6} \right\rangle, \qquad (B4)$$

where all temperatures within each group of three separated by a '|' have zero covariance. The variance takes the form

$$\left\langle \left(\widehat{(f_{\rm nl}^{\,b})^2} \right)^2 \right\rangle = 3\sigma_b^4 \qquad (B5)$$

$$+ \sigma_b^8 \sum_{\{\vec{l},\vec{t}\}} \frac{\mathcal{B}_{\vec{t}_1 \vec{t}_2 \vec{t}_3}^{\vec{l}_1 \vec{l}_2 \vec{l}_3}}{\Omega^2 C_{l_1} C_{l_2} C_{l_3} C_{t_1} C_{t_2} C_{t_3}},$$

where

$$\mathcal{B}_{\vec{t}_1 \vec{t}_2 \vec{t}_3}^{\vec{l}_1 \vec{l}_2 \vec{l}_3} \equiv B(l_1,l_2,l_3)B(l_1,t_1,t_2)B(l_2,t_1,t_3)B(l_3,t_2,t_3)$$

$$\times \ \delta_{\vec{l}_1+\vec{l}_2+\vec{l}_3,0}\delta_{-\vec{l}_1+\vec{t}_1+\vec{t}_2,0}\delta_{-\vec{l}_2-\vec{t}_1+\vec{t}_3,0}\delta_{\vec{l}_3+\vec{t}_2+\vec{t}_3,0}$$

$$+ \ \frac{3}{2}B(l_1,l_2,l_3)B(l_1 l_2 t_1)B(t_2,t_3,l_3)B(t_2,t_3,t_1)$$

$$\times \ \delta_{\vec{l}_1+\vec{l}_2+\vec{l}_3,0}\delta_{-\vec{l}_1-\vec{l}_2+\vec{t}_1,0}\delta_{\vec{t}_2+\vec{t}_3-\vec{l}_3,0}\delta_{\vec{t}_1+\vec{t}_2+\vec{t}_3,0}.$$

$$(B6)$$

Numerically evaluating the sum in Eq. (B6) shows that for $l_{\rm max} \gtrsim 100$ the second (non-Gaussian) term contributes less than 1% to the variance of $\widehat{(f_{\rm nl}^{\,b})^2}$. We have moreover numerically evaluated the exact expression for the correlation coefficient and verified that, as our estimates indicate, the correlation is of order $\lesssim 10\%$ for $l_{\rm min} = 2$ and $l_{\rm max} \gtrsim 100$.

[1] E. Komatsu *et al.*, arXiv:1001.4538 [astro-ph.CO].
[2] T. Falk, R. Rangarajan and M. Srednicki, Astrophys. J. **403**, L1 (1993) [arXiv:astro-ph/9208001]; A. Gangui *et al.*, Astrophys. J. **430**, 447 (1994) [arXiv:astro-ph/9312033]; A. Gangui, Phys. Rev. D **50**, 3684 (1994) [arXiv:astro-ph/9406014].
[3] T. J. Allen, B. Grinstein and M. B. Wise, Phys. Lett. B **197**, 66 (1987); L. A. Kofman and D. Y. Pogosian, Phys. Lett. B **214**, 508 (1988); D. S. Salopek, J. R. Bond and J. M. Bardeen, Phys. Rev. D **40**, 1753 (1989); A. D. Linde and V. F. Mukhanov, Phys. Rev. D **56**, 535 (1997) [arXiv:astro-ph/9610219]; P. J. E. Peebles, Astrophys. J. **510**, 523 (1999) [arXiv:astro-ph/9805194]; P. J. E. Peebles, Astrophys. J. **510**, 531 (1999) [arXiv:astro-ph/9805212]; L. M. Wang and M. Kamionkowski, Phys. Rev. D **61**, 063504 (2000) [arXiv:astro-ph/9907431]; S. Mollerach, Phys. Rev. D **42**, 313 (1990); A. D. Linde and V. F. Mukhanov, Phys. Rev. D **56**, 535 (1997) [arXiv:astro-ph/9610219]; D. H. Lyth and D. Wands, Phys. Lett. B **524**, 5 (2002) [arXiv:hep-ph/0110002]; T. Moroi and T. Takahashi, Phys. Lett. B **522**, 215 (2001) [Erratum-ibid. B **539**, 303 (2002)] [arXiv:hep-ph/0110096]; D. H. Lyth, C. Ungarelli and D. Wands, Phys. Rev. D **67**, 023503 (2003) [arXiv:astro-ph/0208055]; K. Ichikawa *et al.*, arXiv:0802.4138 [astro-ph]; A. L. Erickcek, M. Kamionkowski and S. M. Carroll, Phys. Rev. D **78**, 123520 (2008) [arXiv:0806.0377 [astro-ph]]. G. R. Dvali

and S. H. H. Tye, Phys. Lett. B **450**, 72 (1999) [arXiv:hep-ph/9812483]; P. Creminelli, JCAP **0310**, 003 (2003) [arXiv:astro-ph/0306122]; M. Alishahiha, E. Silverstein and D. Tong, Phys. Rev. D **70**, 123505 (2004) [arXiv:hep-th/0404084]; A. Vilenkin and E. P. S. Shellard "Cosmic Strings and other Topological Defects," (Cambridge University Press, Cambridge, 1994); R. Durrer, New Astron. Rev. **43**, 111 (1999); R. Durrer, M. Kunz and A. Melchiorri, Phys. Rept. **364**, 1 (2002) [arXiv:astro-ph/0110348]; A. H. Jaffe, Phys. Rev. D **49**, 3893 (1994) [arXiv:astro-ph/9311023]; A. Silvestri and M. Trodden, Phys. Rev. Lett. **103**, 251301 (2009) [arXiv:0811.2176 [astro-ph]]; M. Hindmarsh, C. Ringeval and T. Suyama, Phys. Rev. D **80**, 083501 (2009) [arXiv:0908.0432 [astro-ph.CO]]; M. Hindmarsh, C. Ringeval and T. Suyama, Phys. Rev. D **81**, 063505 (2010) [arXiv:0911.1241 [astro-ph.CO]]. D. M. Regan and E. P. S. Shellard, arXiv:0911.2491 [astro-ph.CO]; N. Turok and D. N. Spergel, Phys. Rev. Lett. **66**, 3093 (1991).
[4] X. c. Luo, Astrophys. J. **427**, L71 (1994) [arXiv:astro-ph/9312004]; L. Verde, L. M. Wang, A. Heavens and M. Kamionkowski, Mon. Not. Roy. Astron. Soc. **313**, L141 (2000) [arXiv:astro-ph/9906301]; E. Komatsu and D. N. Spergel, Phys. Rev. D **63**, 063002 (2001) [arXiv:astro-ph/0005036].
[5] M. Kunz, A. J. Banday, P. G. Castro, P. G. Ferreira and K. M. Gorski, Astrophys. J. **563**, L99 (2001)

[arXiv:astro-ph/0111250].

[6] W. Hu, Phys. Rev. D **64**, 083005 (2001) [arXiv:astro-ph/0105117].

[7] T. Okamoto and W. Hu, Phys. Rev. D **66**, 063008 (2002) [arXiv:astro-ph/0206155].

[8] N. Kogo and E. Komatsu, Phys. Rev. D **73**, 083007 (2006) [arXiv:astro-ph/0602099].

[9] D. M. Regan and E. P. S. Shellard, Phys. Rev. D **82**, 023520 (2010) [arXiv:1004.2915 [astro-ph.CO]].

[10] J. Smidt, A. Amblard, C. T. Byrnes, A. Cooray, A. Heavens and D. Munshi, Phys. Rev. D **81**, 123007 (2010) [arXiv:1004.1409 [astro-ph.CO]]; D. Munshi, P. Coles, A. Cooray, A. Heavens and J. Smidt, arXiv:1002.4998 [astro-ph.CO]; J. Smidt, A. Amblard, A. Cooray, A. Heavens, D. Munshi and P. Serra, arXiv:1001.5026 [astro-ph.CO]; D. Munshi, A. Heavens, A. Cooray, J. Smidt, P. Coles and P. Serra, arXiv:0910.3693 [astro-ph.CO];

[11] C. T. Byrnes, M. Sasaki and D. Wands, Phys. Rev. D **74**, 123519 (2006) [arXiv:astro-ph/0611075].

[12] D. Babich, Phys. Rev. D **72**, 043003 (2005) [arXiv:astro-ph/0503375].

[13] P. Creminelli, L. Senatore and M. Zaldarriaga, JCAP **0703**, 019 (2007) [arXiv:astro-ph/0606001].

[14] D. Babich and M. Zaldarriaga, Phys. Rev. D **70**, 083005 (2004) [arXiv:astro-ph/0408455].

[15] G. Jungman, M. Kamionkowski, A. Kosowsky and D. N. Spergel, Phys. Rev. D **54**, 1332 (1996) [arXiv:astro-ph/9512139].

[16] D. Babich, P. Creminelli and M. Zaldarriaga, JCAP **0408**, 009 (2004) [arXiv:astro-ph/0405356]; P. Creminelli *et al.*, JCAP **0605**, 004 (2006) [arXiv:astro-ph/0509029]; P. Creminelli *et al.*, JCAP **0703**, 005 (2007) [arXiv:astro-ph/0610600]; X. Chen, M. x. Huang, S. Kachru and G. Shiu, JCAP **0701**, 002 (2007) [arXiv:hep-th/0605045]; X. Chen, R. Easther and E. A. Lim, JCAP **0706**, 023 (2007) [arXiv:astro-ph/0611645]; R. Holman and A. J. Tolley, JCAP **0805**, 001 (2008) [arXiv:0710.1302 [hep-th]]; P. D. Meerburg, J. P. van der Schaar and P. S. Corasaniti, JCAP **0905**, 018 (2009) [arXiv:0901.4044 [hep-th]]; L. Senatore, K. M. Smith and M. Zaldarriaga, JCAP **1001**, 028 (2010) [arXiv:0905.3746 [astro-ph.CO]]; D. G. Figueroa, R. R. Caldwell and M. Kamionkowski, Phys. Rev. D **81**, 123504 (2010) [arXiv:1003.0672 [astro-ph.CO]].