ON THE INCOMPLETENESS OF THE MOMENT AND CORRELATION FUNCTION HIERARCHY AS PROBES OF THE LOGNORMAL FIELD

JULIEN CARRON

Institute for Astronomy, ETHZ, CH-8093 Zurich, Switzerland; jcarron@phys.ethz.ch

Received 2011 April 25; accepted 2011 June 10; published 2011 August 16

ABSTRACT

We trace with analytical methods and in a model parameter-independent manner the independent bits of Fisher information of each of the moments of the lognormal distribution as a now standard prescription for the distribution of the cosmological matter density field as it departs from Gaussian initial conditions. We show that, when entering the regime of large fluctuations, only a tiny, dramatically decaying fraction of the total information content remains accessible through the extraction of the full series of moments of the field. This is due to the known peculiarity of highly tailed distributions that they cannot be uniquely recovered given the values of all of their moments. Under this lognormal assumption, cosmological probes such as the correlation function hierarchy or, equivalently, their Fourier transforms, are rendered fundamentally limited once the field becomes nonlinear, for any parameter of interest. We show that the fraction of the information accessible from two-point correlations decays to zero following the inverse squared variance of the field. We discuss what general properties of a random field’s probability density function are making the correlation function hierarchy an efficient or inefficient, complete or incomplete, set of probes of any model parameter.

Key words: cosmology: theory – large-scale structure of universe

1. INTRODUCTION

The cosmological matter density field is becoming more and more directly accessible to observations with the help of weak lensing (Schneider et al. 1992; Bartelmann & Schneider 2001; Refregier 2003; Munshi et al. 2008). Its statistical properties are the key element in trying to optimize future large galaxy surveys aimed at answering actual fundamental cosmological questions, such as the nature of the dark components of the universe (Caldwell & Kamionkowski 2009; Frieman et al. 2008). To this aim, Fisher’s measure of information on parameters (Fisher 1925; Rao 1973; van den Bos 2007) has naturally become of standard use in cosmology. It provides indeed a handy framework within which it is possible to evaluate in a quantitative manner standard probes of any model parameter.

In low-dimensional settings. More recently, it has been used as a starting point for a tentatively better description of clustering (Kitaura 2010). The lognormal assumption is itself. The first evaluation of the former within the framework of perturbation theory appeared recently (Wang et al. 2011). Lognormal statistics (see Aitchison & Brown 1957 for a textbook presentation) are not innocuous. More specifically, the lognormal distribution is only one among many distributions that lead to the very same series of moments. This fact indicates that, going from the distribution to the moments, one may be losing information in some way or another. A fundamental limitation of the correlation function hierarchy in extracting the information content of the field in the nonlinear regime could therefore exist, if its statistics are indeed similar to the lognormal. This important fact was already mentioned qualitatively in Coles & Jones (1991), but it seems that no quantitative analysis is available at present. It is the purpose of this paper to provide for the first time answers to these issues, in terms of Fisher information, looking at the details of the structure of the information within the lognormal field. It is built out of two main parts. The first deals exclusively with the case of a single lognormal variable, illustrating the main aspects we want to point out in this work. We begin by presenting how to identify the independent bits of matter field in these conditions is the lognormal distribution, various properties of which are discussed in detail in an astrophysical context in Coles & Jones (1991). It was later shown to be reproduced accurately, both from the observational point of view as well as in comparison with standard perturbation theory (SPT) and N-body simulations (Bernardeau & Kofman 1995; Bernardeau 1994; Kayo et al. 2001; Taylor & Watts 2000; Wild et al. 2005), in low-dimensional settings. More recently, it has been used as a starting point for a tentatively better description of clustering (Kitaura 2010). The lognormal assumption is also highly compatible with numerical works (Neyrinck et al. 2009, 2011) showing that the spectrum of logarithm of the field ln(1 + δ) carries much more information than the spectrum of δ itself. The first evaluation of the former within the framework of perturbation theory appeared recently (Wang et al. 2011). Lognormal statistics (see Aitchison & Brown 1957 for a textbook presentation) are not innocuous. More specifically, the lognormal distribution is only one among many distributions that lead to the very same series of moments. This fact indicates that, going from the distribution to the moments, one may be losing information in some way or another. A fundamental limitation of the correlation function hierarchy in extracting the information content of the field in the nonlinear regime could therefore exist, if its statistics are indeed similar to the lognormal. This important fact was already mentioned qualitatively in Coles & Jones (1991), but it seems that no quantitative analysis is available at present. It is the purpose of this paper to provide for the first time answers to these issues, in terms of Fisher information, looking at the details of the structure of the information within the lognormal field. It is built out of two main parts. The first deals exclusively with the case of a single lognormal variable, illustrating the main aspects we want to point out in this work. We begin by presenting how to identify the independent bits of information that are contained in the successive moments of a distribution, with the help of orthogonal polynomials. We discuss the properties of this decomposition that are relevant for our purposes. The procedure is very similar to the decompositions presented in Jarrett (1984), to which we refer for a more
and the associated covariance matrix with
\[ \Sigma_{ij} = m_{i+j} - m_i m_j. \]

Since \( p(x, \alpha) \) is normalized to unity for any value of the parameter, we have
\[ \frac{\partial m_0}{\partial \alpha} = 0 = \langle s(x, \alpha) \rangle, \]
where \( s(x, \alpha) = \delta_{\alpha} \ln p(x, \alpha) \) is the score function. Fisher’s measure of information (Fisher 1925; van den Bos 2007) on the parameter \( \alpha \) is then defined as the variance of the score function
\[ F_\alpha = \langle s^2(x, \alpha) \rangle. \]

The Fisher information density
\[ s^2(x, \alpha)p(x, \alpha)dx = \frac{(\partial_\alpha p(x, \alpha))^2}{p(x, \alpha)}dx \]
is the amount of information associated with observations of realizations of the variable in the range \((x, x+dx)\). According to the Cramér-Rao inequality (Rao 1973) \( 1/F_\alpha \) bounds the accuracy with which the parameter \( \alpha \) can be extracted from the data \( x \) with the help of unbiased estimators (e.g., Tegmark et al. 1997).

2.1. Fisher Information and Orthogonal Polynomials

The decomposition of the information in independent pieces associated with each moment relies on the approximation of the score function through orthonormal polynomials. For each natural number \( n, \) \( \mathcal{P}_n \) is a polynomial of degree \( n \) and
\[ \langle \mathcal{P}_m(x)\mathcal{P}_n(x) \rangle = \delta_{mn}. \]

These polynomials can always be constructed for a given distribution and are unique up to a sign, which is fixed by requiring the coefficient of \( x^n \) in \( \mathcal{P}_n \) to be positive. We refer to textbooks (Szegö 2003; Freud 1971) for the general theory of orthogonal polynomials. These polynomials can be written in the monomial basis with the help of a triangular transition matrix \( C \) that we will use later in the paper,
\[ \mathcal{P}_n(x) = \sum_{m=0}^{n} C_{nm} x^m. \]

According to Equation (6), it holds that the non-constant orthogonal polynomials average to zero. As the value of the model parameter changes, these averages take non-vanishing values at a rate equal to the component of the score function parallel to these polynomials:
\[ \frac{\partial \langle \mathcal{P}_n(x) \rangle}{\partial \alpha} = \langle s(x, \alpha) \mathcal{P}_n(x) \rangle =: s_n, \]
where the relation \( \partial_\alpha p(x, \alpha) = s(x, \alpha)p(x, \alpha) \) was used. We argue that \( s_n^2 \) is precisely the independent information content of the moment of order \( n \). This can be seen in the following: For any natural number \( n \), it is not difficult to show that the inverse covariance matrix of size \( n \) is given by
\[ [\Sigma^{-1}]_{ij} = [C^T C]_{ij}, \quad i, j = 1, \ldots, n. \]
Therefore, noting that from Equation (7) and from definition (8) of the information coefficients \( s_n \) we can write

\[
s_n = \sum_{k=1}^{n} C_{nk} \frac{\partial m_k}{\partial \alpha},
\]

and the following relation holds:

\[
\sum_{i=1}^{n} s_i^2 = \sum_{i,j=1}^{n} \frac{\partial m_i}{\partial \alpha} \left( \Sigma^{-1} \right)_{ij} \frac{\partial m_j}{\partial \alpha}.
\]

This expression, weighting the sensitivity of the moments to the parameter by the covariance matrix, is the amount of information present in the first \( n \) moments, taking all correlations into account. For instance, this is exactly the amount of information available, if the first \( n \) moments were to be extracted with the help of unbiased, Gaussian-distributed estimators (e.g., Tegmark et al. 1997). Also, it is simple to show that these coefficients are invariant under any linear transformation of the field. This allows us to identify unambiguously the \( n \)th squared coefficient as the independent information contained in the moment of order \( n \).

If the set of orthonormal polynomials forms a complete basis, the partial sums

\[
\sum_{n=1}^{N} s_n p_n(x)
\]

will tend, with increasing \( N \), to reproduce accurately the score function. By the Parseval identity, the full amount of Fisher information can be written as

\[
F_\alpha = \sum_{n=1}^{\infty} s_n^2.
\]

This last equation implies that the information contained in the full set of moments is identical to the total amount of information. This is certainly in perfect agreement with expectations. A well-known result of M. Riesz (Riesz 1923) in the theory of moments states, namely, that if the moment problem associated with the moments \( \{ m_i \}_{i=0}^{\infty} \) is determinate (i.e., the distribution giving rise to these moments is uniquely determined by their values), then the set of associated orthonormal polynomials is complete. Since the distribution is uniquely determined, common sense would then require that the total amount of information contained in the distribution to be the same as that contained in the full set of moments.

However, moment problems are not always determinate, and orthonormal polynomials associated with weight functions do not always form complete sets. Therefore, the series

\[
f_\alpha := \sum_{n=1}^{\infty} s_n^2
\]

may not always converge to the total amount of Fisher information, and, if not, will always underestimate it. We have, namely, instead of Parseval’s identity, the Bessel inequality,

\[
0 \leq \left( \int \left( \sum_{n=1}^{\infty} s_n p_n(x) - s(x, \alpha) \right)^2 \right) = F_\alpha - f_\alpha.
\]

In other words, the mean squared error in approximating the score function with polynomials is the amount of Fisher information absent from the full set of moments. As already emphasized in an astrophysical context by Coles & Jones (1991) and stated in our introduction, the moments of the lognormal distribution are precisely an example of an indeterminate moment problem. In fact, a whole family of distribution, given explicitly in Heyde (1963), has the very same series of moments. In light of these considerations, our subsequent results cannot be considered surprising. Before turning to the actual calculation of the coefficients \( s_n \) of the lognormal distribution, let us state that when the score function is itself a polynomial, it is clear that series (12) actually terminates,

\[
s_k = 0, \quad k > n,
\]

where \( n \) is the order of the polynomial representing the score function. The prime example is the Gaussian distribution, for which \( n = 2 \), with associated orthonormal polynomials, the Hermite polynomials. The well-known fact that the mean and the variance completely determine Gaussian variables becomes, within our framework, that for such variables only \( s_1 \) and \( s_2 \) are non-zero.

### 2.2. Basic Properties of the Lognormal Distribution

The variable \( X \) has a lognormal distribution when \( Y = \ln X \) is a normal variable, with mean \( \mu_Y \) and variance \( \sigma_Y^2 \). The dependency on a model parameter \( \alpha \) can enter one or both of these parameters. The moments of \( X \) are given by Gaussian integrals and read explicitly

\[
m_n = \exp \left( n \mu_Y + \frac{1}{2} n^2 \sigma_Y^2 \right).
\]

The mean and variance of \( Y \) relate therefore to the mean \( \mu \) and variance \( \sigma^2 \) of \( X \) according to

\[
\begin{align*}
\mu_Y &= \ln \mu - \frac{1}{2} \ln \left( 1 + \sigma_{\mu}^2 \right), \\
\sigma_Y^2 &= \ln \left( 1 + \sigma_{\mu}^2 \right).
\end{align*}
\]

where \( \sigma_{\mu}^2 \) is the variance of the fluctuations of \( X \),

\[
\sigma_{\mu}^2 = \frac{\sigma^2}{\mu^2}.
\]

Since \( Y \) is Gaussian and Fisher’s measure is invariant under invertible transformations, the total Fisher information content of \( X \) is given by the well-known expression for the Gaussian distribution:

\[
F_\alpha = \frac{1}{\sigma^2} \left( \frac{\partial \mu}{\partial \alpha} \right)^2 + \frac{1}{2\sigma^4} \left( \frac{\partial \sigma^2}{\partial \alpha} \right)^2.
\]

The key parameter throughout this part of this work will be the quantity \( q \), defined as

\[
q := e^{-\sigma^2} = \frac{1}{1 + \sigma_{\mu}^2}.
\]

Note that \( q \) is strictly positive and smaller than unity. The regime of small fluctuations, where the lognormal distribution is very close to the Gaussian distribution, is described by values of \( q \) close to unity. Deep in the nonlinear regime, it tends to zero. These two regimes are conveniently separated at \( q = 1/2 \), corresponding to fluctuations of unit variance. We note the following convenient property of the moments for further reference:

\[
m_{i+j} = m_i m_j q^{-ij}.
\]
2.3. Information Coefficients

From Equations (10), (17), and (18), we see that the \( n \)th information coefficient \( s_n \) is given by

\[
s_n = \frac{\partial \ln \mu}{\partial \alpha} \sum_{k=0}^{n} C_{nk} m_k k + \frac{1}{2 (1 + \sigma^2)} \frac{\partial \sigma^2}{\partial \alpha} \sum_{k=0}^{n} C_{nk} m_k (k-1). \tag{23}
\]

Evaluation of the above sums can proceed in different ways. Notably, it is possible to get an explicit formula for the orthonormal polynomials, and therefore of the matrix \( \mathbf{C} \), for the lognormal distribution. These are essentially the Stieltjes–Wigert polynomials (Wigert 1923; Szegő 2003). We will, namely, use their explicit expression for the matrix elements (Wigert, 1923; Szegő 2003), as

\[
(t : q)_n = \prod_{k=0}^{n-1} (1 - t q^k), \quad (t : q)_0 := 1 \tag{24}
\]

with \( t \) being a real number, and prove in Appendix A that the following curious identity holds,

\[
\langle P_n(t x) \rangle = (-1)^n \frac{q^{n/2}}{\sqrt{(q : q)_n}} (t : q)_n. \tag{25}
\]

By virtue of

\[
\langle P_n(t x) \rangle = \sum_{k=0}^{n} C_{nk} m_k t^k, \tag{26}
\]

it follows from our identity (25) that the sums given in the right-hand side of Equation (23) are proportional to the first and, respectively, the second derivative of the \( q \)-Pochammer symbol evaluated at \( t = 1 \). Besides, matching the powers of \( t \) on both sides of Equation (25) will provide us immediately with the explicit expression for the matrix elements \( C_{nk} \).

We distinguish explicitly two situations, labeled by an index \( a \) taking either value \( \mu \) or \( \sigma \), where only one of the two parameters of the lognormal distribution actually depends on \( \alpha \). The general case is reconstructed trivially from these two cases.

Case \( a = \mu \). We assume in this case that the parameter enters the mean of the distribution only:

\[
\frac{\partial \sigma^2}{\partial \alpha} = 0. \tag{27}
\]

From Equation (23), we see that the derivative of \( \mu \) with respect to \( \alpha \) only plays the role of an overall normalization constant. Since we will deal exclusively with ratios, it is irrelevant for our purposes. We choose for convenience

\[
\frac{\partial \ln \mu}{\partial \alpha} = 1. \tag{28}
\]

The total amount of information in the distribution becomes, from Equations (20) and (18),

\[
F^\mu_a := \frac{1}{\ln (1 + \sigma^2)}. \tag{29}
\]

Case \( a = \sigma \). The parameter enters the variance of the distribution only, and we pick again a convenient normalization of its derivative:

\[
\frac{1}{2 (1 + \sigma^2)} \frac{\partial \sigma^2}{\partial \alpha} = 1, \quad \frac{\partial \ln \mu}{\partial \alpha} = 0. \tag{30}
\]

This situation is the most common in cosmology, for instance, for any model parameter entering the matter power spectrum. The exact amount of information becomes, again from Equations (20) and (18),

\[
F^\sigma_a := \frac{1}{\ln (1 + \sigma^2)} \left( 1 + \frac{2}{\ln (1 + \sigma^2)} \right). \tag{31}
\]

In both of these situations, we obtain the information coefficients (23) by differentiating once, respectively, twice our relation (25) with respect to the parameter \( t \), and evaluating these derivatives at \( t = 1 \). The result is, in the case \( a = \mu \),

\[
s^\mu_n = (-1)^{n-1} \sqrt{\frac{q^n}{1 - q^n}} (t : q)_{n-1} \tag{32}
\]

and for \( a = \sigma \),

\[
s^\sigma_n = -2 s^\mu_n \sum_{k=1}^{n-1} \frac{q^k}{1 - q^k}, \quad n > 1, \tag{33}
\]

whereas \( s^\sigma_{n=1} \) is easily seen to vanish from its definition.

2.4. Incompleteness of the Information in the Moments

The series

\[
f^a_a := \sum_{n=1}^{\infty} \left( s^a_n \right)^2, \quad a = \mu, \sigma \tag{34}
\]

are the total amount of information contained in the full series of moments, in the respective cases described above. The ratios \( \epsilon_\mu \) and \( \epsilon_\sigma \), defined as

\[
\epsilon_a := \frac{f^a_a}{F^a_a}, \quad a = \mu, \sigma, \tag{35}
\]

are the fraction of the information that can be accessed by extraction of the full set of moments of \( X \). The two asymptotic regimes of very small and very large fluctuation variance \( \sigma_q \) can be seen without difficulty. In both cases, it is seen that the first non-vanishing term of the corresponding series completely dominates its value. For very small variance, or equivalently \( q \) very close to unity, \( \epsilon_a \) tends to unity, illustrating the fact the distribution becomes arbitrary close to Gaussian: all the information is contained in the first two moments. The large variance regime is more interesting, and, even though the information coefficients decay very sharply as well, the series (34) are far from converging to the corresponding expressions (29) and (31) showing the total amount of information. Considering only the dominant first term in the relevant series and setting \( q \to 0 \), one obtains

\[
\epsilon_\mu \to \frac{1}{\sigma_\mu^2} \ln (1 + \sigma_\mu^2) \tag{36}
\]
and a much more dramatic decay of \( \epsilon_\sigma \):

\[
\epsilon_\sigma \to \frac{4}{\sigma_\delta^2} \ln \left( 1 + \frac{\sigma_\delta^2}{2} \right).
\]

Both series given in Equation (34) are quickly convergent and well suited for numerical evaluation. Figure 1 shows the accessible fractions \( \epsilon_\sigma \) of information through extraction of the full series moments. Figure 2 shows the repartition of this accessible fraction among the first 10 moments. Most relevant from a cosmological point of view in Figure 1 is the solid line, dealing with the case of the parameters of interest entering the variance only. These figures show clearly that the moments, as probes of the lognormal matter field, are penalized by two different processes. First, as soon as the field shows nonlinear features, following Equations (36) and (37), almost all the information content can no longer be accessed by extracting its successive moments. Within a range of one magnitude in the variance, the moments go from very efficient to highly inefficient probes. Second, as shown in Figure 2, as the variance of the field approaches unity, this accessible fraction gets transferred quickly from the variance alone to higher order moments. This repartition of the information within the moments is built out of two different regimes. First, for large variance, or large \( n \), we see easily from the above expressions (32) and (33) that in both cases the information coefficients decay exponentially,

\[
s_n \propto \left( 1 + \sigma_\delta^2 \right)^{-n}, \quad n \ln \left( 1 + \sigma_\delta^2 \right) \gg 1.
\]

On the other hand, if the variance or \( n \) is small enough, we can set \( 1 - q^n \approx -n \ln q \), and obtain, very roughly,

\[
s_n^2 \propto \left[ n \ln \left( 1 + \sigma_\delta^2 \right) \right]^n, \quad n \ln \left( 1 + \sigma_\delta^2 \right) \ll 1,
\]

explaining the trend with variance seen in Figure 2, which places more importance on higher order moments as the variance grows. Note that the latter regime can occur only for small enough values of the variance. Deeper in the nonlinear regime, the trend is therefore reversed, obeying Equation (38) for all values of \( n \), with a steeper decay for higher variance.

2.5. A q-analog of the Logarithm

These results show clearly that large parts of the information become invisible to the moments. However, it does not tell us what is responsible for this phenomenon. It is therefore of interest to look more deeply into the details of this missing piece of information. As we have seen, these are due to the inability of the polynomials to reconstruct precisely the score function. In the case \( a = \mu \), the score function of the lognormal distribution is easily shown to take the form of a logarithm in base \( q \),

\[
s(x) = -\frac{1}{2} - \ln_q \left( \frac{x}{\mu} \right).
\]

Therefore, the series

\[
s^\mu(x) := \sum_{n=0}^{\infty} s_n^\mu P_n(x)
\]

will represent some function, very close to a logarithm for \( q \to 1 \) over the range of \( p(x, \alpha) \). It will, however, fail to reproduce some of its features at lower \( q \)-values. This is hardly surprising, since it is well known that the logarithm function does not have a Taylor expansion over the full positive axis. For this reason, the approximation \( s^\mu(x) \) of \( s(x) \) through polynomials can indeed only fail when the fluctuation variance becomes large enough. In Appendix B, we show that \( s^\mu(x) \) takes the form

\[
s^\mu(x) = -\sum_{k=1}^{\infty} \frac{q^k}{1-q^x} \ln \left( 1 + \left( 1 - q^\mu \right) \frac{q^{(k-1)}}{(q^\mu)^k} \left( \frac{x}{\mu} \right)^k \right).
\]

It is interesting to note that this series expansion is almost identical to the one of the \( q \)-analog of the logarithm \( S_q(x) \) defined by E. Koelink and W. Van Assche, with the only difference allowing, for instance from Equation (11), the direct evaluation of the polynomials to reconstruct precisely the score function. In particular, \( S_q(x/\mu) \) has, making it a real \( q \)-analog of the logarithm, such as \( S_q(q^{-n}) = n \), for positive integers. The qualitative behavior of \( s^\mu \), however, stays close to \( S_q \). Notably, its behavior in underdense regions, \( x/\mu \ll 1 \), seen from Equation (42) \( s^\mu \) tends to a finite value, which is very different from a logarithm.

This calculation can be performed as well in the case \( a = \sigma \), with similar conclusions. Since it is rather tedious and not very enlightening, we do not reproduce it in these pages. We show in Figure 3 the information density, Equation (5), of the lognormal distribution (dashed line), and its approximation by the orthogonal polynomials (solid line),

\[
p(x, \alpha) \left( \sum_{n=0}^{\infty} s_n^\mu P_n(x) \right)^2,
\]

when the fluctuation variance \( \sigma^2 \) is equal to unity. It is clear from this figure that in this regime, while most of information is located within the underdense regions of the lognormal field, the moments are unable to catch it. To check the correctness of our numerical and analytical calculations, we compared the total information content as evaluated from integrating the information densities in Figure 3 to that given by Equation (31), respectively, Equation (34), with essentially perfect agreement.

2.6. Comparison to Standard Perturbation Theory

For any distribution, knowledge of its first \( 2n \) moments allows, for instance from Equation (11), the direct evaluation of
the independent information content of the first $n$ moments. This is true even if the exact shape of the distribution is not known or is too complicated. In particular, we can use the explicit expressions for the first six moments of the density fluctuation field within the framework of SPT provided by Bernardeau (1994) in order to compare $s_2$ and $s_3$ as given from SPT to their lognormal analogs. We note that a comparison with Bernardeau (1994) can only be very incomplete and, to some extent, can only fail. It is indeed part of the approach in Bernardeau (1994), when producing functional forms for the distribution of the fluctuation field, to invert the relation between a moment generating function and its probability density function. For such an inversion to be possible it is of course necessary that the probability density is uniquely determined by its moments. This is not the case for the lognormal distribution. Therefore, that approach can never lead to an exact lognormal distribution, or to any distribution for which the moment hierarchy forms an incomplete set of probes. However, such a comparison can still lead to conclusions relevant for many practical purposes, such as those dealing with the first few moments.

The variance of the field is explicitly given as an integral over the matter power spectrum,

$$\sigma_8^2 = \frac{1}{2\pi^2} \int_0^\infty dk \, k^2 P(k, \alpha) |W(kR)|^2, \quad (44)$$

where $W(kR)$ is the Fourier transform of the real space top hat filter of size $R$, and any cosmological parameter $\alpha$ entering the power spectrum $P(k)$. In the notation of Bernardeau (1994), the moments of the fluctuation field $m_n = \langle \delta^n \rangle$ are given by the disconnected, or Gaussian, components, together with the connected components $\langle \delta^n \rangle_c$ in terms of parameters $S_n$:

$$\langle \delta^n \rangle_c = \sigma_8^{2n-1} S_n$$

$$m_n^{\text{Gauss}} = \begin{cases} 0 & n \text{ odd} \\ \sigma_8^n (n-1)!! & n \text{ even}. \end{cases} \quad (45)$$

The parameters $S_n$ contain a leading, scale-independent coefficient, and deviation from this scale independence is given in terms of the logarithmic derivative of the variance,

$$\gamma_i = \frac{d^i \ln \sigma_8^2}{d \ln R^i}, \quad i = 1, \ldots. \quad (46)$$

Neglecting the very weak dependence of $S_n$ on cosmology, from Equation (45) we can write

$$\frac{\partial m_n}{\partial \alpha} = \frac{\partial \sigma_8^2}{\partial \alpha} \begin{cases} 0, & n = 1 \\ 1, & n = 2 \\ 2m_3/\sigma_8^2, & n = 3. \end{cases} \quad (47)$$

With the coefficients $S_n$ up to $n = 6$ given in Bernardeau (1994, page 703), and the above relations, we perform a straightforward evaluation of the information coefficients $s_2^2$ and $s_3^2$ using Equation (11). The variance is obtained from Equation (44) within a flat $\Lambda$CDM universe ($\Omega_0 = 0.7$, $\Omega_m = 0.3$, $\Omega_b = 0.045$, $h = 0.7$), with power spectrum parameters ($\sigma_8 = 0.8$, $n = 1$). We use the transfer function from Eisenstein and Hu (Eisenstein & Hu 1998). The needed derivatives $\gamma_i$, $i = 1, \ldots, 4$ are obtained numerically through finite differences.

In Figure 4, we show the ratio

$$\left(\frac{s_3^2}{s_2^2}\right)^{\frac{1}{2}}, \quad (48)$$

i.e., the relative importance of the third moment with respect to the second, as a function of the variance, both for the lognormal distribution and the SPT predictions. This ratio is identically zero for a Gaussian distribution. The models stand in good agreement over many orders of magnitude. It is striking that both models consistently predict the entrance of the nonlinear regime; this ratio takes a maximal value close to unity. Surely, the SPT curve for larger values of the variance is hard to interpret, since it is out of its domain of validity.

### 3. Several Variables

So far, we have considered only one variable. We now extend our analysis to the more interesting multidimensional case. It is possible to derive formally a general expression for the independent information content of the $n$-point correlations of any distribution. This proceeds in strict analogy with the one-dimensional case, where an expansion of the score function...
in polynomials of several variables is made. It is presented in greater detail in Appendix C. For the lognormal field, a given model parameter can only enter via the means of the logarithm of the field at each point or through the elements of its two-point correlation matrix. However, we could not transform the corresponding expressions in a useful, easily evaluated form in a general situation, as in the one-dimensional case. We therefore focus on two more restricted but tractable situations. First, in analogy with the \( a = \mu \) case, we consider a parameter that enters the mean of the field \( \rho \) and no elements of the correlation matrix. In this case, the complete amount of information can be extracted via the mean of \( \ln \rho \), and we compare that amount to the one obtained by extracting the mean of \( \rho \) only. In a second step, we consider the extraction of the correlation amplitude \( \xi(r) \) between independent pairs of cells separated by that distance \( r \). In this case, the full amount of information on any parameter is related to those of \( \ln \rho \) through

\[
\xi_{\ln \rho}(r) = \ln[1 + \xi(r)].
\]

On the other hand, the means obey

\[
\ln \rho = \ln \bar{\rho} - \frac{1}{2} \xi_{\ln \rho}(0) = \ln \bar{\rho} - \frac{1}{2} \ln (1 + \sigma_b^2).
\]

The Fisher information content of \( \rho \) is again given by the standard expression for the Gaussian field \( \ln \rho \). It splits into the part coming from the observation of \( \ln \rho \) and the one coming from the correlations \( \xi_{\ln \rho} \),

\[
F_{\alpha} = \frac{1}{2} \text{Tr} \left[ \xi_{\ln \rho}^{-1} \frac{\partial \ln \rho}{\partial \alpha} \xi_{\ln \rho}^{-1} \frac{\partial \ln \rho}{\partial \alpha} \right] + \frac{\partial \ln \rho}{\partial \alpha} \xi_{\ln \rho}^{-1} \frac{\partial \ln \rho}{\partial \alpha}.
\]

We denote by \( Q_N \) the square matrix defined as

\[
[Q_N]_{ij} = Q_{ij}, \quad |i|, |j| \leq N.
\]

We note that by virtue of Equations (54) and (56), its matrix elements read

\[
Q_{ij} = \prod_{k,l=1}^d \left( 1 + \xi(x_k - x_l) \right)^{j_i k_l}.
\]

A general expression for the independent information content \( s^2_n \) of the correlations of order \( n \), Equation (C8), and its link to the completeness of the orthogonal polynomials is presented for completeness in Appendix C. This machinery is, however, not compulsory for the following considerations, which are restricted to the two lowest order correlations. We will only use the analog of Equation (11), which gives the total amount of information contained in the correlations up to order \( N \),

\[
\sum_{n=1}^N s^2_n = \sum_{|j|, |j| \leq N} \frac{\partial m_j}{\partial \alpha} \Sigma^{-1}_{ij} \frac{\partial m_j}{\partial \alpha},
\]

where \( \Sigma \) is the covariance matrix

\[
\Sigma_{ij} = m_{ij} - m_i m_j, \quad |i|, |j| \leq N.
\]

For the lognormal distribution, the property (54) allows us to rewrite this last expression in the equivalent form

\[
\sum_{n=1}^N s^2_n = \sum_{|i|, |j| \leq N} \frac{\partial \ln m_j}{\partial \alpha} \left( Q_N^{-1} \right)_{ij} \frac{\partial \ln m_j}{\partial \alpha}.
\]

### 3.2. Extraction of the Mean

In this case, we set

\[
\frac{\partial \xi(r)}{\partial \alpha} = 0
\]

for any argument. Again, the actual value of the derivative of the mean with respect to \( \alpha \) will play no role. This condition (64) implies that \( \delta_{\alpha} \xi_{\ln \rho} \) vanishes as well. We can read out from Equation (58) that the total amount of information is given by

\[
F_{\alpha} = \left( \frac{\partial \ln \bar{\rho}}{\partial \alpha} \right)^2 \sum_{i,j=1}^d \xi_{\ln \rho}^{-1} [i]_{ij}.
\]
We stress that since $\ln \rho$ is Gaussian, the information is accessible in its entirety by extraction of the mean of $\ln \rho$. On the other hand, the amount extracted by looking at the mean of $\rho$ itself is Equation (56) with $N = 1$. Using the definition of $\xi$ in Equation (55), it becomes

$$s_1^2 = \left( \frac{\partial \ln \hat{\rho}}{\partial \alpha} \right)^2 \sum_{i,j=1}^d [\xi^{-1}]_{ij}. \quad (66)$$

In the limit of a continuous sample $d \to \infty$, the sum

$$\sum_{i,j=1}^d [\xi^{-1}]_{ij} \to \int_V \frac{V}{P_\rho(k = 0)} \quad (67)$$

becomes a double integral over space, which can be performed by Fourier transformation, and is the inverse of the power spectrum of the field at zero argument,

$$P_\rho(0) = \int_V d^d r \xi(r) \quad (68)$$

and similarly for $\hat{\xi}_{\ln \rho}$. We conclude that the loss of information by looking at the mean of the field only is given straightforwardly by the ratio of the power of the fields $\ln \rho$ and $\rho$ at zero argument:

$$\epsilon := \frac{s_1^2}{F_\alpha} = \frac{P_{\ln \rho}(0)}{P_\rho(0)}. \quad (69)$$

From the explicit representation of $P_{\ln \rho}(0)$,

$$P_{\ln \rho}(0) = \int_V d^d r \ln(1 + \xi(r)), \quad (70)$$

and the fact that $\ln(1 + x)$ is strictly smaller than $x$ whenever $x \neq 0$, it follows that the loss of information always occurs, but is substantial only if the correlation function takes substantial values. However, the information loss $\epsilon$ is roughly insensitive to the presence of some correlation scale. We see that the presence of correlations does not alter the main conclusions drawn in the first part of this work.

### 3.3. Extraction of Correlations

We suppose now that the parameter $\alpha$ enters the correlation function $\xi$ for some argument $r$. Since the field is lognormal, measurement of $\hat{\xi}_{\ln \rho}(r)$ captures all the information on $\alpha$. We want to compare this amount to the one extracted by measuring $\xi(r)$ itself. We suppose further that $\xi(r)$ is extracted from a number of independent pairs of points separated by that distance $r$. The independence of the pairs allows us to simplify drastically the problem, since by additivity of the information the information loss will be independent of the number of pairs. Our problem thus becomes two dimensional. Our assumptions for the impact of the parameter $\alpha$ are, more explicitly,

$$\frac{\partial \rho}{\partial \alpha} = 0, \quad \frac{\partial \sigma_2}{\partial \alpha} = 0, \quad \frac{\partial \xi(r)}{\partial \alpha} \neq 0. \quad (71)$$

We point out that this is very different from the $\alpha = \sigma$ case that we treated earlier, since here the variance $\sigma_2^2$ only acts as a noise source and not as a source of information. The correlation matrix of a pair of points of the homogeneous Gaussian field $\ln \rho$ separated by $r$ reads, according to Equation (56),

$$\xi_{\ln \rho} = \begin{pmatrix} \ln (1 + \sigma_2^2) & \ln (1 + \xi(r)) \\ \ln (1 + \xi(r)) & \ln (1 + \sigma_2^2) \end{pmatrix}. \quad (72)$$

The positivity of the matrix constrains, at a fixed variance $\sigma_2^2$, the values of $\xi(r)$ to the following range:

$$1 + \xi(r) = (1 + \sigma_2^2)^\eta, \quad \eta \in (-1, 1). \quad (73)$$

Clearly, vanishing correlations correspond to $\eta = 0$, positive correlations to $\eta > 0$, and negative correlations to $\eta < 0$. By assumption, the parameter $\alpha$ does enter $\xi(r)$ only. Therefore,

$$\frac{\partial \xi_{\ln \rho}}{\partial \alpha} = \begin{pmatrix} -\eta (1 + \xi(r)) \frac{\partial \sigma_2^2}{\partial \alpha} & 0 \\ 0 & 1 \end{pmatrix}. \quad (74)$$

Putting Equation (74) into Equation (58), we obtain that the total amount of information for $\alpha$ is

$$F_\alpha = \left( \frac{\partial \xi_{\ln \rho}}{\partial \alpha} \right)^2 \frac{1 + \eta^2}{1 + \xi(r)} \frac{1}{1 - \eta^2} \ln^2 \left( 1 + \sigma_2^2 \right). \quad (75)$$

For the information obtained by extracting $\xi$, we first note that under our assumption (71) the mean carries no information,

$$s_1 = 0. \quad (76)$$

For this reason, $s_2^2$ is given by Equation (63), with $N = 2$. The only non-zero element of the vector of derivatives $\partial \ln m_i, |i| \leq 2$, is in our configuration (71) for the multi-index $i = (1, 1)$. From Equations (52) and (56), we obtain

$$\frac{\partial m_{i(1,1)}}{\partial \alpha} = \frac{\partial \xi(r)}{\partial \alpha} \frac{1}{1 + \xi(r)}. \quad (77)$$

It follows immediately that $s_2^2$ is given by

$$s_2^2 = \left( \frac{\partial \xi(r)}{1 + \xi(r)} \right)^2 \left[ Q_2^{-1} \right]_{(1,1)(1,1)}. \quad (78)$$

Of special interest is the limit of low correlations, where the exact result

$$\frac{s_2^2}{F_\alpha} = \frac{1}{\sigma_2^2} \ln^2 \left( 1 + \sigma_2^2 \right), \quad (79)$$

can be obtained making profit of the simple structure of the $Q_2$ matrix. We note that $Q_2$ is in our two-point configuration a six-dimensional matrix, where, as seen from its representation (60), all of its elements are products and powers of $1 + \sigma_2^2$ and $1 + \xi(r)$. The needed inverse matrix element can be written as the ratio of determinants,

$$\left[ Q_2^{-1} \right]_{(1,1)(1,1)} = \frac{\det \hat{Q}_2}{\det Q_2}. \quad (80)$$

where $\hat{Q}_2$ is the five-dimensional matrix originating from $Q_2$ where the row and column corresponding to the multi-index $(1, 1)$ have been taken out. Both of these determinants are therefore clearly polynomials in $1 + \sigma_2^2$ and $1 + \xi(r)$. The asymptotic behavior of the accessible information for large
variance can thus be obtained by noting that \( \eta \to 0 \) for any value of \( \xi \). Looking then at the leading coefficients of the two polynomials entering Equation (80), we obtain

\[
\det Q_2 \to (\sigma_3^2)^2 (1 + \xi(r))^2 \\
\det \hat{Q}_2 \to (\sigma_3^2)^{10}.
\]

Therefore, the information loss

\[
\epsilon := \frac{\lambda_2}{F_{\alpha}}
\]

(82)
tends to, for asymptotic values of the variance,

\[
\epsilon \to \frac{1}{\frac{3}{2} \ln^2 (1 + \sigma_3^2)} \\
= \frac{\epsilon (\xi = 0)}{(1 + \xi(r))^2}.
\]

The second line follows from the first using the exact result for vanishing correlations given in Equation (79). The efficiency of \( \xi \) in extracting the information for any parameter therefore always goes to zero, following approximately the inverse squared variance. The presence of substantial positive correlations, generic for a field generated by gravitational instability on a wide range of scales, only makes the information loss worse. This is illustrated in Figure 5, where the dotted line shows the loss of information at the nonlinearity scale \( \xi(r) = 1 \), evaluated numerically from Equations (75) and (78), together with the exact result (79) for vanishing correlations (solid line).

4. SUMMARY AND CONCLUSION

We have investigated in detail the structure of the information within the moments of the univariate lognormal distribution as a model for the matter density field. We have provided exact expressions, Equations (32) and (33), for the independent information content of each moment. Using these expressions, we have shown that the moments become dramatically incomplete probes in the nonlinear regime. In the cosmologically relevant case of the parameter entering the power spectrum of the fluctuations, the fraction of the information that is accessible from the moments is close to 1/4 at \( \sigma_3 = 1 \), and decays following the inverse fourth power of the variance. We showed that it is mainly due to the inability of the moments to probe the information located in the underdense regions of the distribution. Additionally, a comparison with SPT for the lower order moments showed that both approaches are consistent and predict that the third-order moment becomes as important as the variance itself when entering the nonlinear regime. In a second step, we extended our results to the multivariate case, showing that the mean of the field and two-point correlations become in a very similar manner very inefficient probes for any parameter of interest. With the help of two simplified situations we have shown that the presence of correlations only makes the information loss even worse. More specifically, we have shown that the extraction of the two-point correlation function for any argument provides access to a fraction of the information, which is generically well below unity at the entrance of the nonlinear regime, and decays like the inverse squared variance.

These results, making clear that the information content of the lognormal field is not only transferred to higher order point functions, but also becomes largely inaccessible to the correlation function hierarchy in the nonlinear regime, confirm to full extent qualitative suspicions raised in Coles & Jones (1991, Section 4), to which we refer for a more complete discussion of physical arguments that may cause such behavior. We can understand for any random field whether the hierarchy members are promising probes from the following considerations: Let \( p(\phi, \alpha) \) be the probability density function for a realization \( \phi \) of the field. As we have seen, the information content of the first \( N \)-correlation functions is based on the approximation of the score function \( \partial_\alpha \ln p(\phi) \) through polynomials up to order \( N \), over the range of \( p(\phi) \). Let us assume, for instance, that for any value of the model parameter \( \ln p(\phi) \) can be expanded in a low-order Taylor series in the field,

\[
\ln p(\phi, \alpha) = \sum_{n=0}^{N} \int dx_1 \cdots dx_n \\
\lambda_n(x_1, \ldots, x_n, \alpha) \phi(x_1) \cdots \phi(x_n).
\]

This class of distributions, including Gaussian fields for which \( N = 2 \), can arise notably as maximal entropy distributions for fixed values of the first \( N \)-correlation functions. The coefficient \( \lambda_n \) is called in this framework the potential associated with the \( n \)th correlation function (see Jaynes 1983; Caticha 2008, e.g.). Since \( \partial_\alpha \ln p \) is itself a polynomial of order \( N \), all the information is contained in the first \( N \)-correlation functions. Of course, the relative importance of each one of these will be modulated by the sensitivity of the potentials to the parameter \( \alpha \) and their covariances. This situation is certainly the one where the correlation function hierarchy is the probe of choice, since only a finite number of these grasp the entire information content. Two different processes may render the hierarchy inefficient or incomplete. First of all, a large number of terms may be needed in expansion (84) to reproduce accurately the score function \( \partial_\alpha \ln p \). In this case, one would need to go deep down the hierarchy in order to catch the information. This is certainly not desirable. The last case occurs when \( \partial_\alpha \ln p \) has no Taylor expansion at all over the relevant range. It is then simply not possible to represent accurately the score function. Parts of the information (given in the field analog of Equation (15)) become invisible to the correlation function hierarchy. The lack of a Taylor expansion for the logarithm function is the reason for the failure of the moments and correlation functions to catch...
the information of a lognormal field in the nonlinear regime, when the range of the probability density function becomes very large. We emphasize, as in Coles & Jones (1991), that these peculiar dynamics of the information are not due to a pathological character of the lognormal distribution. This should be expected for any distribution decaying slowly at infinity. We can add to their discussion that this is so because ln p cannot be well reproduced by polynomials under this condition.

Given the very high amplitude of these effects within the lognormal assumption, we believe that in order to get the best out of future galaxy survey data, it is crucial to understand better these issues. The present work presents first steps toward this aim. It is also very consistent and brings strong support to the recent studies started in Neyrinck et al. (2009, 2011), and Wang et al. (2011), making in the nonlinear regime the logarithm of the field rather than the field itself the central quantity of interest. However, it remains to be seen to what extent the approach presented in this work is able to provide quantitative predictions for the statistical power of higher point functions, or for power spectrum extraction. We leave these for future work.

We thank Adam Amara, Simon Lilly, Alexander Szalay, and Mark Neyrinck for useful discussions, and acknowledge the support of the Swiss National Science Foundation.

APPENDIX A

DERIVATION OF RELATION (25)

To prove Equation (25), we note that both sides of the equation are polynomials of degree n in t and that the zeros of the right-hand side are given by

\[ t = q^{-i}, \quad i = 0, \ldots, n - 1. \]  

(A1)

We first show that the left-hand side evaluated at these points does vanish as well, so that the two polynomials must be proportional. We then find the constant of proportionality by requiring \( P_n \) to have the correct normalization. The first step is performed by noting that

\[ \langle P_n(q^{-i} x) \rangle = \frac{1}{m_i} \langle P_n(x)x^i \rangle, \quad i = 0, 1, \ldots. \]  

(A2)

an identity that is proven by expanding \( P_n \) on both sides of the equation in terms of the transition matrix \( C \), and using relation (22) between the moments. Since \( P_n \) is by construction orthogonal to any polynomial of strictly lower degree, we have

\[ \langle P_n(q^{-i} x) \rangle = 0, \quad i = 0, \ldots, n - 1. \]  

(A3)

This implies

\[ \sum_{k=0}^{n} C_{ni} m_k t^k = \alpha_n (t : q)_n \]  

(A4)

for some constant of proportionality \( \alpha_n \). To find it, we note that by expanding the normalization condition of \( P_n \),

\[ 1 = \langle P_n^2(x) \rangle, \]  

(A5)

using again property (22), it must hold that

\[ 1 = \sum_{i,j=0}^{n} C_{ni} m_j C_{nj} m_j q^{-ij}. \]  

(A6)

The sums can be performed using Equation (A4), leading to the following equation for \( \alpha_n \):

\[ 1 = (-1)^{i} \alpha_n^2 q^{n(n-1)/2}(q^{-n} : q)_n. \]  

(A7)

This expression simplifies to

\[ \alpha_n^2 = \frac{q^n}{(q : q)_n}, \]  

(A8)

and the sign of \( \alpha_n \) must be \(-1^n\) in order to have a positive matrix element \( C_{mn} \). This concludes the proof of Equation (25).

APPENDIX B

DERIVATION OF THE REPRESENTATION (42)

In order to get the explicit series representation of Equation (42), we first obtain from relation (25) the exact expression of the transition matrix \( C \). The expansion of the \( q \)-Pochhammer symbol on the right-hand side of Equation (42) is in powers of \( r \) the Cauchy binomial theorem,

\[ (t : q)_n = \sum_{k=0}^{n} \binom{n}{k}_q q^{k(k-1)/2}(-t)^k, \]  

(B1)

where

\[ \binom{n}{k}_q = \frac{(q : q)_n}{(q : q)_k(q : q)_{n-k}} \]  

(B2)

is the Gaussian binomial coefficient. Matching powers of \( t \) in Equation (25), we obtain the explicit form

\[ C_{nk} = (-1)^{n-k} \frac{q^{n/2}}{\sqrt{(q : q)_n}} \binom{n}{k}_q q^{x^2 \mu^{-k}}. \]  

(B3)

Therefore, interchanging the \( n \) and \( k \) sums in Equation (42), it holds

\[ s^{\mu}(x) = -\sum_{n=1}^{\infty} \frac{q^n}{1-q^n} + \sum_{k=1}^{\infty} q^{x^2} \left(-\frac{x}{\mu}\right)^k \sum_{n=k}^{\infty} \frac{q^n}{1-q^n} \binom{n}{k}_q. \]  

(B4)

With algebra the following identity is not difficult to show:

\[ \sum_{n=k}^{\infty} q^n \binom{n}{k}_q = \frac{1}{(q : q)_k 1-q^k}, \quad k \geq 1. \]  

(B5)

Consequently, the series expansion of \( s^{\mu}(x) \) is given by

\[ s^{\mu}(x) = -\sum_{k=1}^{\infty} \frac{q^k}{1-q^k} \left[1 + (-1)^k q^{x^2(k-1)/k} \left(-\frac{x}{\mu}\right)^k\right]. \]  

(B6)

APPENDIX C

SEVERAL VARIABLES

We first need a little bit of notation. For a variable \( X \) taking values \( x = (x_1, \ldots, x_d) \), we use the multi-index notation

\[ x^n = x_1^{n_1} \cdots x_d^{n_d}, \quad n = (n_1, \ldots, n_d), \quad n_i = 0, 1, \ldots \]

(C1)
where $|n|$ is the order of the multi-index $n$. A moment of order $N$ is given by

$$m_n = \langle x^n \rangle, \quad |n| = N, \quad \text{(C2)}$$

and the covariance between the moments is

$$m_{n+m} - m_n m_m =: \Sigma_{nm}. \quad \text{(C3)}$$

In this notation, the decomposition of the information in independent bits of order $N$ proceeds by strict analogy with the one-dimensional case. We refer to Dunkl & Xu (2001) for the general theory of orthogonal polynomials in several variables. A main difference is that at a fixed order $N$ there are not one but $\binom{N+d-1}{d}$ independent orthogonal polynomials, which are not uniquely defined. The orthogonality of the polynomials of the same order is not essential for our purposes; requiring the following condition is enough,

$$\langle P_n(x) P_m(x) \rangle = 0, \quad |n| \neq |m|$$

$$\langle P_n(x) P_m(x) \rangle = [H_n]_{mm}, \quad |n| = |m| = n \quad \text{(C4)}$$

for some matrices $H_n$. The component of some function $f$ parallel to the polynomial $P_n$ is

$$s_n := \langle f(x) P_n(x) \rangle, \quad \text{(C5)}$$

and the expansion of $f$ in terms of these polynomials reads, in the notation of Dunkl & Xu (2001, Section 3.5),

$$S_N(f)(x) = \sum_{n=0}^{N} \sum_{|n| = |m|} s_n [H_n^{-1}]_{nm} P_m(x). \quad \text{(C6)}$$

It will converge to the actual function $f$ for $N \to \infty$ if the set of polynomials is complete, whereas it may not if it is not complete. The expansion is also independent of the freedom there is in the choice of the orthogonal polynomials in Equations (C4). Writing the orthogonal polynomials in terms of a triangular transition matrix

$$P_n(x) = \sum_{|m| \leq |n|} C_{nm} x^m \quad \text{(C7)}$$

and taking $f$ as the score function $s(x, \alpha)$, it is simple to see that the independent bits of information of order $n$ are given by

$$s_n^2 = \sum_{|n|, |m| = |n|} s_n [H_n^{-1}]_{nm} s_m$$

$$= \sum_{|i|, |j| \leq n} \left[ C^T H_n^{-1} C \right]_{ij} \frac{\partial m_i}{\partial \alpha} \frac{\partial m_j}{\partial \alpha}. \quad \text{(C8)}$$

and the strict analog of Equation (11) holds for each $N$,

$$\sum_{n=1}^{N} s_n^2 \sum_{i,j=1}^{N} \frac{\partial m_i}{\partial \alpha} \frac{\partial m_j}{\partial \alpha}. \quad \text{(C9)}$$

REFERENCES

Aitchison, J., & Brown, J. C. 1957, The Lognormal Distribution (Cambridge: Cambridge Unv. Press)

Albrecht, A., et al. 2006, arXiv:astro-ph/0605951

Amara, A., & Réfrégier, A. 2007, MNRAS, 381, 1018

Andrews, G. E., Askey, R., & Roy, R. 1999, Special Functions (Cambridge: Cambridge Unv. Press)

Bartelmann, M., & Schneider, P. 2001, Phys. Rep., 340, 291

Bernardeau, F. 1994, A&A, 291, 697

Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, Phys. Rep., 367, 1

Bernardeau, F., & Kofman, L. 1995, ApJ, 443, 479

Bernstein, G. M. 2009, ApJ, 695, 652

Calderwood, R., & Kamionkowski, M. 2009, Nature, 458, 587

Caticha, A. 2008, arXiv:physics.data-an/0808.0012

Coles, P., & Jones, B. 1991, MNRAS, 248, 1

Dunkl, C. F., & Xu, Y. 2001, Orthogonal Polynomials of Several Variables (Cambridge: Cambridge Unv. Press)

Eisenstein, D. J., & Hu, W. 1998, ApJ, 496, 605

Fisher, R. A. 1925, Math. Proc. Camb. Phil. Soc., 22, 700

Freud, G. 1971, Orthogonal Polynomials (Oxford: Pergamon)

Friedman, J. A., Turner, M. S., & Huterer, D. 2008, ARA&A, 46, 385

Gautschi, W. 2008, J. Comput. Appl. Math., 219, 408

Hayde, C. C. 1963, J. R. Stat. Soc. Ser. B (Methodological), 25, 392

Hu, W., & Jain, B. 2004, Phys. Rev. D, 70, 043009

Hu, W., & Tegmark, M. 1999, ApJ, 514, L65

Jarrett, R. G. 1984, Biometrika, 71, 101

Jaynes, E. T. 1983, in Papers On Probability, Statistics and Statistical Physics, ed. R. Rosenkranz (Dordrecht: Reidel)

Kac, J., & Cheung, P. 2001, Quantum Calculus (New York: Springer)

Kayo, I., Taruya, A., & Suto, Y. 2001, ApJ, 561, 22

Kitaura, F. S. 2010, arXiv:1012.3168

Kolenci, E., & Van Assche, W. 2009, Proc. Am. Math. Soc., 137, 1663

Matsubara, T. 2011, Phys. Rev. D, 83, 083518

Munshi, D., Valageas, P., Van Waerbeke, L., & Heavens, A. 2008, Phys. Rep., 462, 117

Neyrinck, M. C., Szapudi, I., & Szalay, A. S. 2009, ApJ, 698, L90

Neyrinck, M. C., Szapudi, I., & Szalay, A. S. 2011, ApJ, 731, 116

Parkinson, D., Blake, C., Kunz, M., Bassett, B. A., Nichol, R. C., & Glazebrook, K. 2007, MNRAS, 377, 185

Peebles, P. J. E. 1980, The Large-Scale Structure of the Universe (Princeton: Princeton University Press)

Rao, C. 1973, Lineare Statistische Methoden und ihre Anwendungen (Berlin: Academic)

Refregier, A. 2003, ARA&A, 41, 645

Riesz, M. 1923, Acta Syeged Sect. Math., 1, 209

Schneider, P., Ehlers, J., & Falco, E. E. 1992, Gravitational Lenses (Berlin: Springer)

Szego, G. 2003, Orthogonal Polynomials (4th ed.; Providence, RI: American Mathematical Society)

Taylor, A. N., & Watts, P. I. R. 2000, MNRAS, 314, 92

Tegmark, M. 1997, Phys. Rev. Lett., 79, 3806

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

van den Bos, A. 2007, Parameter Estimation for Scientists and Engineers (New York: Wiley)

Wang, X., Neyrinck, M., Szapudi, I., Szalay, A., Chen, X., Lesgourgues, J., Riotto, A., & Sloth, M. 2011, ApJ, 735, 32

Wigert, S. 1923, Ark. Mat., Astron. Fys., 17, 15

Wild, V., et al. 2005, MNRAS, 356, 247