The Copula of the Cosmological Matter Density Field is Non-Gaussian

Jian Qin\textsuperscript{1,2} $\otimes$, Yu Yu\textsuperscript{1,2} $\otimes$, and Pengjie Zhang\textsuperscript{1,2,3}

\textsuperscript{1}Department of Astronomy, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, 200240, People’s Republic of China; yuyu22@sjtu.edu.cn
\textsuperscript{2}Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai, 200240, People’s Republic of China; zhangpj@sjtu.edu.cn
\textsuperscript{3}Division of Astronomy and Astrophysics, Tsung-Dao Lee Institute, Shanghai Jiao Tong University, Shanghai, 200240, People’s Republic of China

Received 2020 April 13; revised 2020 May 15; accepted 2020 May 18; published 2020 July 7

Abstract

Non-Gaussianity of the cosmological matter density field can be largely reduced by a local Gaussianization transformation (and its approximations, such as the logarithmic transformation). Such behavior can be recast as the Gaussian copula hypothesis (GCH), and has been verified to very high accuracy at a two-point level. On the other hand, statistically significant non-Gaussianities in the Gaussianized field have been detected in simulations. We point out that this apparent inconsistency is caused by the very limited degrees of freedom in the copula function, which make it misleading as a diagnosis of residual non-Gaussianity in the Gaussianized field. Using the copula density and at the two-point level, we highlight the departure from Gaussianity. We further quantify its impact in the predicted $n$th ($n \geq 2$) order correlation functions. We explore a remedy of the GCH, which alleviates but does not completely solve the above problems.

Unified Astronomy Thesaurus concepts: N-body simulations (1083); Large-scale structure of the universe (902); Dark matter distribution (356); Non-Gaussianity (1116)

1. Introduction

The large-scale structure (LSS) fields in the late-time universe, in particular the matter density field, are often significantly non-Gaussian, due to the nonlinear evolution of the universe (Bernardeau et al. 2002). Describing the non-Gaussianity accurately and extracting the encoded cosmological information efficiently is an active area of research (e.g., Hamilton 2000; Zhang et al. 2011; Yu et al. 2012a). An interesting finding is that, despite the vast possibility of non-Gaussian behaviors, the non-Gaussianity of the matter density field induced by nonlinearity takes a specific form of simplicity. It has been known for a long time that a local monotonic transformation $\delta(x) \rightarrow y(x) = f(\delta(x))$ can significantly reduce the non-Gaussianity. Namely, the non-Gaussianity is largely encoded in the one-point probability distribution function (PDF), and the field $(y(x))$ after the above local transformation is close to Gaussian. In the literature, various approximations such as the logarithmic transform (Coles & Jones 1991; Neyrinck et al. 2009), the rank-order transform (Weinberg 1992; Neyrinck 2011; McCullagh et al. 2016), the Box–Cox transform (Joachimi et al. 2011), and the clipping method (Simpson et al. 2011), along with the exact Gaussianization transformation (Yu et al. 2011), have been investigated. All are able to significantly reduce the non-Gaussianity and enhance the information content encoded in the two-point statistics.

Copula provides an alternative description on the above findings. The $n$-point PDF $f(\delta_1, \delta_2, \ldots, \delta_n)$ $(n = 1, 2, \ldots)$ completely describes the statistics of the density field. They can be equivalently described by the combination of a one-point PDF $f(\delta)$, and all the $n$-point copula $C(u_1, u_2, \ldots, u_n)$ $(n \geq 2)$. Here $u(\delta) \equiv F(\delta) \equiv \int_{-\infty}^{\delta} f(\delta')d\delta'$, and $F(\delta)$ is the cumulative distribution function of the density field. Copula has a nice property: it is invariant under any local monotonic transformation. This makes it convenient to describe the residual non-Gaussianity in the $y$ field. For example, if the copula is found to depart from the Gaussian form, then no local transformation can render the density field Gaussian. Scherrer et al. (2009) found through $N$-body simulations that the two-point copula for all the investigated spatial separation is Gaussian to extremely high accuracy. This motivates the authors to postulate the Gaussian copula hypothesis (GCH) that all $n$-point copulas are Gaussian. The GCH, along with the one-point (non-Gaussian) PDF, provides a convenient and close form description of the non-Gaussian density field. It has been applied to study the covariance matrix of the lensing power spectrum and other statistics (e.g., Sato et al. 2010, 2011; Takeuchi 2010; Lin & Kilbinger 2015; Yu et al. 2016; Zhan. ng 2018).

However, direct investigation of the Gaussianized field (the $y$ field) shows the existence of residual non-Gaussianity. For example, non-vanishing bispectra (Yu et al. 2011) and off-diagonal covariance matrix elements of the power spectrum (Yu et al. 2016) have been detected robustly. An intuitive illustration of the residual non-Gaussianity is the strong anisotropic structures in the two-dimensional visualizations of the Gaussianized density fields (e.g., Figure 1 in Neyrinck et al. 2009).

What causes the above inconsistency? One possibility is the non-Gaussianity in higher-order copula (e.g., $n \geq 3$). Here we point out an alternative possibility. The two-point copula investigated in Scherrer et al. (2009) is indeed nearly Gaussian, as we have verified independently with our simulations. However, this nearly Gaussian copula is misleading due to the built-in nature of copulas. We take the two-point copula $C(u, v)$ as an example. It is subject to the following constraints:

$$u, v \in [0, 1]; \quad C(u, v) \in [0, 1]; \quad C(u, v) = C(v, u);$$

$$\frac{\partial C(u, v)}{\partial u} \geq 0; \quad \frac{\partial C(u, v)}{\partial v} \geq 0;$$

$$C(u, 0) = C(0, v) = 0; \quad C(u, 1) = u; \quad C(1, v) = v. \quad (1)$$

So it has very limited degrees of freedom. It monotonically increases with both $u$ and $v$. It has fixed values when $u = 0, 1$ or $v = 0, 1$. So different copulas may look similar. This has two implications. (i) For example, from the viewpoint of LSS, a field with significant spatial correlation, and a random field of vanishing
The copula density is defined as
\[
c(u_1, u_2, ..., u_n) = c(u_1, u_2, ..., u_n) = \frac{\partial^n C(u_1, u_2, ..., u_n)}{\partial u_1 \partial u_2 ... \partial u_n},
\]

(4)

It is related to the joint PDF (JPDF) \( f(\delta_1, \delta_2, ..., \delta_n) \) by
\[
c(u_1, u_2, ..., u_n) = \frac{f(\delta_1, \delta_2, ..., \delta_n)}{\prod_{i=1}^{n} f(\delta_i)}. \]

(5)

2.1. General Properties

By the above definitions and results, we can derive some basic properties of copula. For brevity, we demonstrate them with the two-point copula \( C(u, v) \).

1. \( u, v \in [0, 1] \), and \( C \in [0, 1] \).
2. \( C(u, v) = C(v, u) \). Note that this holds for statistically homogenous fields such as the cosmological matter density field. It does not hold for general fields.
3. \( C(u, 0) = 0 \) and \( C(u, 1) = u \), as shown by Equation (2).
4. Due to this constraint, different values, while keeping the Gaussian form.
4. Note that this holds for statistically homogenous fields, statistically homogenous or not.
5. Invariance of copula under monotonically increasing transformation \( y = f(\delta) \). This is obvious, since \( F_{\delta}(y_1, y_2, ..., y_n) = F_{\delta}(\delta_1, \delta_2, ..., \delta_n) \).

For special cases, the copula function has an analytical expression. One is the case of two uncorrelated variables \( \delta_1, \delta_2 \). Since \( F_{\delta}(\delta_1, \delta_2) = F_{\delta}(\delta_1)F_{\delta}(\delta_2) \), we have
\[
C(u, v) = uv, \quad c(u, v) = 1.
\]

(6)

Another case is the Gaussian copula for the Gaussian field, as detailed below.

2.2. Gaussian Copula

To distinguish the Gaussian CDF from a general CDF, we denote it as \( \Phi_{\rho,n}(\delta_1, ..., \delta_n) = F_G(\delta_1, ..., \delta_n) \). The covariance matrix,
\[
\rho_{ij} \equiv \langle \delta_i \delta_j \rangle, \quad \text{with} \quad \sigma_i^2 \equiv \langle \delta_i^2 \rangle,
\]

(7)

completely fixes the statistics of the Gaussian field. From Equation (2), we derive the Gaussian copula (see, e.g., Malevergne & Sornette 2003; Neyrinck 2011)
\[
C_G(u_1, ..., u_n) = \Phi_{\rho,n}(\Phi^{-1}_1(u_1), ..., \Phi^{-1}_n(u_n)).
\]

(8)

Here \( \Phi_1 \) is the marginal Gaussian CDF,
\[
\Phi_1(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi} \sigma_i} \exp \left( -\frac{x^2}{2\sigma_i^2} \right) dx.
\]

(9)
The Gaussian copula density is
\[
c_{G}(u_1, u_2, \ldots, u_n) = \frac{\sigma_1 \sigma_2 \ldots \sigma_n}{\sqrt{\det \rho}} \times \exp \left( -\frac{1}{2} \left( \rho^{-1} \cdot \text{diag}(\rho^{-1}) \right) y_{(a)} \right).
\] (10)

Here \( y_{(a)} = (\Phi_{1}^{-1}(u_1), \Phi_{1}^{-1}(u_2), \ldots, \Phi_{1}^{-1}(u_n))^T \) and
\[
\text{diag}(\rho^{-1}) = \begin{pmatrix}
\sigma_1^{-2} & 0 & \cdots & 0 \\
0 & \sigma_2^{-2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma_n^{-2}
\end{pmatrix}
\] (11)
is the diagonal part of \( \rho^{-1} \). One can further verify that, as we expect,
\[
f(\delta_1, \delta_2, \ldots, \delta_n) = c_{G}(u_1, u_2, \ldots, u_n) \prod_{i=1}^{n} f(\delta_i) = \frac{1}{\sqrt{(2\pi)^n \det \rho}} \exp \left( -\frac{1}{2} \left( \frac{\sigma_i^T}{2} \rho^{-1} \frac{\sigma_i}{2} \right) \right).
\] (12)

Since the copula is invariant under local monotonically increasing transformation,
\[
C_{G}(u_1, \ldots, u_n) = \Phi_{L}^{-1}(u_1, \ldots, \Phi_{L}^{-1}(u_n)).
\] (13)

Now \( \Phi_{L}^{-1} \) is the JCD of \( \delta_i/\sigma_i \) (\( i = 1, 2, \ldots, n \)) and \( \Phi \) is the one-point Gaussian CDF with unit variance. The Gaussian copula density can then be simplified to
\[
c_{G}(u_1, u_2, \ldots, u_n) = \frac{1}{\sqrt{\det \rho}} \exp \left( -\frac{1}{2} \left( \frac{\gamma_i}{2} \rho^{-1} \frac{\gamma_i}{2} \right) \right).
\] (14)

Any Gaussian copula/copula density can be written in the form of Equations (13) and (14). For simplicity, we use these two forms to illustrate Gaussian copula (densities) in this work.

### 3. Testing the GCH

The GCH (Scherrer et al. 2009) states that, despite the strong non-Gaussianity of one-point PDF \( f(\delta) \), the copula of the cosmological matter density field is Gaussian, as expressed by Equation (13). It also predicts that the covariance matrix \( \rho \) is that of the transformed field \( y = \Phi^{-1}(u = F(\delta)) \). Namely, \( y \) is the Gaussianization of \( \delta \) such that its one-point PDF is Gaussian. For the two-point copula, the only free parameter there is \( r \), i.e., the cross correlation coefficient between \( y_1 \) and \( y_2 \). The normalized covariance matrix
\[
\rho = \begin{pmatrix}
1 & r \\
r & 1
\end{pmatrix}.
\] (15)

We test GCH with a \( \Lambda \)CDM \( N \)-body simulation. The simulation was run with 3072\(^3 \) particles in a box of side length 600 Mpc \( h^{-1} \), and a flat cosmology specified by \( \Omega_m = 0.268, \Omega_{\Lambda} = 0.732, H_0 = 71 \text{ km s}^{-1} \text{Mpc}^{-1}, \sigma_8 = 0.83, n_s = 0.968 \). The details of the simulation are described in Jing et al. (2007) & Jing (2019). The density fields are sampled at redshifts \( z = 17, 5, 1, 0 \) with the pixel size 1 Mpc \( h^{-1} \). The mean number of particles per pixel is \( \sim 134 \), so we can safely neglect the effect of shot noise. As explained earlier, the copula may be misleading in revealing the non-Gaussianity. So we measure both the copula and the copula density.

We restrict our investigation on the two-point copula (densities), which can be measured from the JCDF and PDF. To measure the joint distributions, we sample \( n = 600^3 \) of \( \delta_1 \) at position \( x_1 \) and the associated \( \delta_2 \) at position \( x_2 = x_1 + s \). \( s \) is the pair separation vector, and we investigate the cases of \( s = 2 \text{ Mpc} h^{-1} \) and 6 Mpc \( h^{-1} \), respectively. We follow the procedure described in Scherrer et al. (2009), and utilize the transformation invariant property of copulas. We rank \( \delta_1 \) and adopt the monotonic transformation \( y_1 = R(\delta_1)/n \). Namely, the lowest \( \delta_1 \) is mapped to \( y_1 = 1/n \) and the highest \( \delta_1 \) corresponds to \( y_1 = 1 \). We do the same for \( \delta_2 \) to obtain \( y_2 \). \( y \) has the uniform PDF, \( F(R(\delta_1)/n) = R(\delta_1)/n, f(R(\delta_1)/n) = 1 \). Then
\[
c_{G}(R(\delta_1)/n, R(\delta_2)/n) = F(R(\delta_1)/n, R(\delta_2)/n),
\] (16)
\[
c_{D}(R(\delta_1)/n, R(\delta_2)/n) = f(R(\delta_1)/n, R(\delta_2)/n).
\] (17)

Here, the subscript “D” denotes the data simulation. In other words, the joint distributions of density ranks (divided by the number of points) give the two-point copula (density), which is called the “empirical copula.”

### 3.1. Direct Comparison

Figure 1 shows the copula at four redshifts. One finding is the lack of evolution in redshifts, implying that the rank order of the density field roughly conserves under gravitational evolution (Weinberg 1992). We also overplot the Gaussian copula predicted by GCH. The curves almost completely overlap with the simulation result, for all four redshifts and two spatial separations investigated. This confirms the finding of Scherrer et al. (2009). However, as we argued earlier, this is very misleading. To demonstrate this point, we overplot the copula of vanishing \( \delta_1 \delta_2 \). From the viewpoint of LSS, this one is fundamentally different. Nevertheless, it overlaps with the simulated one when \( u, v \to 0 \), or \( u, v \to 1 \). Therefore, even tiny differences in the copula may lead to significant differences in LSS statistics.

Figure 2 shows the copula density. In contrast to the case of copula, now departures from Gaussianities are clearly revealed (black solid curves versus black dash curves), at low redshifts or small separation. The next step is to quantify its impact on commonly used LSS statistics.

### 3.2. GCH-induced Bias in LSS Statistics

The two-point copula density determines all correlation functions of the following form:
\[
\xi_{mn} = \langle \delta_1^m \delta_2^n \rangle = \int_{-1}^{1} \int_{-1}^{1} \delta_1^m \delta_2^n f(\delta_1, \delta_2) d\delta_1 d\delta_2
\]
\[
= \int_{0}^{1} \int_{0}^{1} \delta_1^m \delta_2^n c(u_1, u_2) du_1 du_2.
\] (18)

The GCH fixes \( c \) and therefore makes a unique prediction of \( \langle \delta_1^m \delta_2^n \rangle \). Inaccuracies in GCH can then be quantified by the bias in \( \langle \delta_1^m \delta_2^n \rangle \), with respect to the simulated (true) value. As shown in Figure 3, \( \xi_{mn} \) predicted by GCH is accurate only at high redshift. Significant bias has developed even at \( z = 5 \). Therefore, despite (almost) invisible deviation from GCH in copula, the induced bias in \( \xi_{mn} \) can be significant.
3.3. The Alternative Gaussian Copula Approximation

A surprising finding above is that GCH even fails to predict $\langle \delta_1 \delta_2 \rangle$ at low redshift. Since the prediction is completely fixed by $r$ in the covariance matrix $\rho$, this motivates us to check whether we can choose another $r$ to improve the prediction of not only $\langle \delta_1 \delta_2 \rangle$, but $\langle \delta_i \delta_j \rangle$ in general. $r$’s in GCH are fixed by

Figure 1. Empirical two-point copula $C_D(u, v)$ for the simulated dark matter density distributions at the two-point separations 2 Mpc $h^{-1}$ (top panels) and 6 Mpc $h^{-1}$ (bottom panels). The four columns show results at redshifts $z = 17, 5, 1, 0$. Solid curves are the contours corresponding to (from lower left to upper right) $C_D(u, v) = 0.1, 0.3, 0.5, 0.7, 0.9$. The dashed curves show the Gaussian copula measured based on GCH. Note that they are almost indistinguishable from the solid curves. The dotted curves show the copula of spatially uncorrelated fields. Note the similarity in the copula, especially in the vicinity of $u \to 0$, $v \to 0$, $1$. Such similarity implies limitations of copulas in describing LSS, and in particular the LSS non-Gaussianity.

Figure 2. Empirical two-point copula densities $c_D(u, v)$ for the simulated dark matter density distributions at the two-point separations 2 Mpc $h^{-1}$ (top panels) and 6 Mpc $h^{-1}$ (bottom panels). The four columns show results at redshifts $z = 17, 5, 1, 0$. Solid curves are the contours corresponding to $c_D(u, v) = 1.8, 1.5, 1.2, 0.9, 0.6, 0.3$ (top panels) and $c_D(u, v) = 1.20, 1.12, 1.04, 0.96, 0.88, 0.80$ (bottom panels). The dashed curves give the Gaussian copula densities measured based on GCH, and the dotted curves give the results of the alternative Gaussian copula approximation.

4
the field $\Phi^{-1}(F(\delta))$.\(^5\) We show them in Figure 4. An alternative $r$ can be fixed by requiring that $\langle \delta_1 \delta_2 \rangle$ predicted by Equation (18) agrees with the simulated one. Such $r$’s are also shown in Figure 4. The two sets of $r$ do show visible difference at low redshifts. To distinguish from the GCH copula, we call the copula with the new set of $r$ as the “alternative Gaussian copula approximation.”

Copulas under the alternative Gaussian copula approximation gives unbiased results of $\langle \delta_1 \delta_2 \rangle$. However, they do not give a better match for the copula densities (Figure 2). Furthermore, they improve the accuracy of predicted $\langle \delta_1^m \delta_2^n \rangle$ ($m+n>1$), but not significantly (Figure 5). Biases in $\langle \delta_1 \delta_2^2 \rangle$ vary from 5% to 15%. Biases in $\langle \delta_1^m \delta_2^n \rangle$ ($m+n=4$) are larger, ranging from 10% to ~60%. Such biases are too large for precision cosmology. Therefore, even the alternative Gaussian copula approximation has limited usage in precision cosmology.

4. Summary

We have revealed the otherwise hidden non-Gaussianity of the copula of the (3D) cosmological matter density field, via the copula density statistics and the accuracy in the predicted $n$-point correlation functions. The found non-Gaussianity shows that the nonlinear statistics of the 3D density field are more complicated than the GCH suggests. This further verifies our previous finding that the $y$ field after local Gaussianization has detectable non-Gaussianity. One remaining question is the information encoded in the non-Gaussian part of the $y$ field, and another question is whether we can conveniently describe and capture such non-Gaussianity. These are for future works. On the other hand, Gaussianization of 2D density field (e.g., the weak lensing convergence field) is much more accurate, and has valuable applications (Joachimi et al. 2011; Yu et al. 2011, 2012b, 2016; Munshi et al. 2014; Chen et al. 2020). It is also interesting to check whether the nonlocal Gaussianization (Hall & Mead 2018) results in Gaussian copulas.

The copula is a promising tool because of its advantageous mathematical properties, but its misuse can be misleading. For example, the misuse of Gaussian copulas in econometric modeling was blamed for the 2008 global financial crisis. To make full and correct use of copulas, we need knowledge and judgment beyond that used in traditional statistical measures. There are methodologies developed by mathematicians for other applications that we can learn from. For example, in geology, Gräler (2014) found that vine copula allow inclusion of the extremal behavior of a spatial random field and capture of the distribution of heavily skewed spatial random field, where Gaussian copulas failed. In structural engineering, Wang & Li (2018) found that, while the specification of random fields in terms of the marginal distributions and correlation structure

---

5 According to the transformation invariance of the copula, the GCH indicates that $r$ equals the correlation coefficient of $\Phi^{-1}(F(h_1)), \Phi^{-1}(F(h_2))$. We also checked that the value of $r$ determined by $\Phi^{-1}(F(h))$ matches the value produced by the Spearman rank correlation procedure in Scherrer et al. (2009).
is incomplete, the non-Gaussian dependence structure is a real phenomenon in engineering practice and they found the D-vine copulas are more suitable for representing one-dimensional stationary random field. In ecology, Prates et al. (2015) transformed the margins of a Gaussian Markov random field to desired marginal distributions, which accommodate the asymmetry and heavy tail needed in many ecological circumstances.

This work was supported by the National Key Basic Research and Development Program of China (No. 2018YFA0404504), and the National Science Foundation of China (11621303, 11653003, 11773048, 11890691).

ORCID iDs
Jian Qin @ https://orcid.org/0000-0003-0406-539X
Yu Yu @ https://orcid.org/0000-0002-9359-7170

References
Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, PhR, 367, 1
Chen, Z., Yu, Y., Liu, X., & Fan, Z. 2020, arXiv:2001.10765
Coles, P., & Jones, B. 1991, MNRAS, 248, 1
Gräler, B. 2014, Spatial Stat., 10, 87
Hall, A., & Mead, A. 2018, MNRAS, 473, 3190
Hamilton, A. J. S. 2000, MNRAS, 312, 257
Jing, Y. 2019, SCPMA, 62, 19511
Jing, Y. P., Suto, Y., & Mo, H. J. 2007, ApJ, 657, 664
Joachimi, B., Taylor, A. N., & Kiessling, A. 2011, MNRAS, 418, 145
Lin, C.-A., & Kilbinger, M. 2015, A&A, 583, A70
Malevergne, Y., & Sornette, D. 2003, QuantFin, 3, 231
McCullagh, N., Neyrinck, M., Norberg, P., & Cole, S. 2016, MNRAS, 457, 3652
Munshii, D., Coles, P., & Kilbinger, M. 2014, JCAP, 2014, 004
Neyrinck, M. C. 2011, ApJ, 742, 91
Neyrinck, M. C., Szapudi, I., & Szalay, A. S. 2009, ApJL, 698, L90
Prates, M. O., Dey, D. K., Willig, M. R., & Yan, J. 2015, Spatial Stat., 14, 382
Sato, M., Ichiki, K., & Takeuchi, T. T. 2010, PhRvL., 105, 251301
Sato, M., Ichiki, K., & Takeuchi, T. T. 2011, PhRvD, 83, 023501
Scherrer, R. J., Berlind, A. A., Mao, Q., & Mcbride, C. K. 2009, ApJ, 708, 9
Simpson, F., James, J. B., Heavens, A. F., & Heymans, C. 2011, PhRvL, 107, 271301
Sklar, A. 1959, Publications de l’Institut de statistique de l’Université de Paris, 8, 229
Takeuchi, T. T. 2010, MNRAS, 406, 1830
Wang, F., & Li, H. 2018, Struct. Safety, 75, 75
Weinberg, D. H. 1992, MNRAS, 254, 315
Yu, H.-R., Harnois-Déraps, J., Zhang, T.-J., & Pen, U.-L. 2012a, MNRAS, 421, 832
Yu, Y., Zhang, P., & Jing, Y. 2016, PhRvD, 94, 083520
Yu, Y., Zhang, P., Lin, W., Cui, W., & Fry, I. J. 2011, PhRvD, 84, 023523
Yu, Y., Zhang, P., Lin, W., Cui, W., & Fry, I. J. 2012b, PhRvD, 86, 023515
Zhang, P. 2018, ApJ, 869, 74b
Zhang, T.-J., Yu, H.-R., Harnois-Déraps, J., MacDonald, I., & Pen, U.-L. 2011, ApJ, 728, 35

Figure 5. Same as Figure 3, but for the Gaussian copulas determined by the alternative Gaussian copula approximation.