The Astrophysical Journal, 821:110 (7pp), 2016 April 20 doi:10.3847/0004-637X/821/2/110

A NEW STATISTICAL PERSPECTIVE TO THE COSMIC VOID DISTRIBUTION

J-R Pycke and E. Russell
Division of Science and Mathematics, New York University Abu Dhabi, P.O. Box 129188, Abu Dhabi, UAE; jrp15@nyu.edu, er111@nyu.edu
Received 2015 November 23; accepted 2016 February 22; published 2016 April 19

ABSTRACT

In this study, we obtain the size distribution of voids as a three-parameter redshift-independent log-normal void probability function (VPF) directly from the Cosmic Void Catalog (CVC). Although many statistical models of void distributions are based on the counts in randomly placed cells, the log-normal VPF that we obtain here is independent of the shape of the voids due to the parameter-free void finder of the CVC. We use three void populations drawn from the CVC generated by the Halo Occupation Distribution (HOD) Mocks, which are tuned to three mock SDSS samples to investigate the void distribution statistically and to investigate the effects of the environments on the size distribution. As a result, it is shown that void size distributions obtained from the HOD Mock samples are satisfied by the three-parameter log-normal distribution. In addition, we find that there may be a relation between the hierarchical formation, skewness, and kurtosis of the log-normal distribution for each catalog. We also show that the shape of the three-parameter distribution from the samples is strikingly similar to the galaxy log-normal mass distribution obtained from numerical studies. This similarity between void size and galaxy mass distributions may possibly indicate evidence of nonlinear mechanisms affecting both voids and galaxies, such as large-scale accretion and tidal effects. Considering the fact that in this study, all voids are generated by galaxy mocks and show hierarchical structures in different levels, it may be possible that the same nonlinear mechanisms of mass distribution affect the void size distribution.

Key words: catalogs – galaxies: clusters: intracluster medium – large-scale structure of universe – methods: numerical – methods: statistical

1. INTRODUCTION

The seeds of the present-day large-scale structure of the universe are formed from random Gaussian density fluctuations in the early stages of its evolution. In these random density fluctuations, matter evolves from a linear to a highly nonlinear regime driven by gravitational instabilities. These instabilities in the primordial Gaussian density field form the complex structure of the universe that we observe today. As a result, it is natural to describe the probability distribution function (PDF) of the density fluctuations as a crucial statistical tool to identify different types of environments of the universe. For example, a Gaussian matter distribution function represents the linear regime, while the distribution function that deviates from Gaussian shows a highly nonlinear regime. In this framework, there are two important building blocks of the large-scale structure; galaxies/overdense regions and voids/underdense regions. Initially these two features are formed from the same primordial Gaussian density field. While voids are formed from the minima of the density field, galaxies are formed in the density maxima. In time when galaxies grow in mass, voids tend to empty in their mass due to their peculiar gravitational force through the nonlinear regime. Although there are different phenomenological models of the PDF of overdense regions (such as galaxies, galaxy clusters,..etc.) in nonlinear regimes (Saslaw 1985, p. 506; Gaztañaga & Yokoyama 1993; Lahav et al. 1993; Ueda & Yokoyama 1996), the statistical models of void probability functions (VPFs) (Fry 1986; Coles & Jones 1991; Elizalde & Gaztanaga 1992) are based on the counts in randomly placed cells following the prescription of White (1979). In addition to this VPF, the number density of voids is another key statistic to obtain the void distribution. Patiri et al. (2006) show that this number density can be estimated analytically by using numerical simulations or mock catalogs.

The phenomenological models of PDFs of overdense regions, especially the one-point and the two-point PDFs, are well studied. The two-point PDF is a tool for modeling the dark halo biasing as well as for obtaining the errors in the one-point statistics (Colombi et al. 1995; Szapudi & Colombi 1996), while the one-point PDF is a useful tool for showing the clustering of the universe with higher-order moments-statistics such as skewness and kurtosis (Kayo et al. 2001). Observations, as well as models based on cold dark matter numerical simulations of galaxy distributions, indicate that the density distribution is well-approximated by a log-normal rather than a Gaussian PDF (Hamilton 1985; Coles & Jones 1991; Bouchet et al. 1993; Kofman et al. 1994; Taylor & Watts 2000). Also, Bernardeau (1992, 1994) show that the PDF computed from perturbation theory in a weakly nonlinear regime approaches log-normal when the primordial power spectrum is proportional to an index $n = -1$. Later on, Kayo et al. (2001) found that the one-point log-normal PDF can describe the density distributions accurately, not only in the weakly nonlinear regime, but also in the highly nonlinear regime by using the nonlinear density fluctuations from *N*-body simulations with Gaussian initial conditions. Kayo et al. (2001) also indicate that this PDF is fairly independent of the shape and the power spectrum of density fluctuations. However, the underlying mechanism of the cosmological origin of the log-normal distribution remains unknown.

In this study we determine that a specific standard parametric model can fit samples of void radii from the Cosmic Void Catalog (CVC) of Sutter et al. (2012) after suitable estimation of the parameters in the model. The three mock catalogs under study, referred to by the nicknames given in the CVC, are *N*-body

Mock, Halo Occupation Distribution (HOD) Sparse, and HOD Dense. Using these samples we show that the system of three-parameter log-normal distribution provides a fairly satisfactory model of the size distribution of voids that is strikingly similar to the galaxy mass distribution of Kayo et al. (2001), assuming the void size is proportional to its mass content ($R^3 \approx M$). To do this, we follow three standard steps in the statistical analysis. These are the exploratory phase, the choice of a probability model, estimations of the parameters of the model, and finally the graphical and numerical assessment (see chapter 4, Sections 4.1 and 4.2 in Fisher et al. (1993) and the introduction in Sheskin (2011) for an overview of basic principles and some terminology employed in the fields of descriptive statistics, estimation, and goodness-of-fit procedures).

2. VOID CATALOG AND NUMERICAL DATA

To investigate the void size distribution function statistically in simulations, here we use the public CVC of Sutter et al. (2012). CVC has two main catalogs: a complete catalog that is suitable for void galaxy surveys, and a bias-free catalog of voids that provides a fair sampling of void shapes and alignments in which the void effective radii vary between 5 and 135 Mpc h^{-1} (Sutter et al. 2012). All the data samples we use here from the CVC are generated from a Λ cold dark matter (ΛCDM) N-body simulation by using an adaptive treecode N-body method called the 2HOT code (Sutter et al. 2014a, 2014b). Sutter et al. (2014a, 2014b) extract halos from the N-body simulation, and use the position and masses of halos to produce a HOD model. Therefore, galaxy catalogs are produced from a halo population by using the HOD code of Tinker et al. (2006) and the HOD model by Zheng et al. (2007). Sutter et al. (2014a, 2014b) generate three mock catalogs; HOD Dense, HOD Sparse, and N-body Mock. In these catalogs, voids are identified with by the modified version of the parameter-free void finder ZOBOV (Neyrinck 2008; Lavaux & Wandelt 2012; Sutter et al. 2012).

HOD Dense and HOD Sparse are produced by dark matter simulations of 1024^3 particles in a 1 Gpc h^{-1} box and all particles are kept in real space at $z = 0$ and are tuned to the observational HOD fits (Sutter et al. 2014a). As a result, the difference between these two catalogs comes from their resolutions as well as their tuned observational data sets. The HOD Sparse mock catalog consists of 1422 voids with 14 Mpc h^{-1} effective minimum radii ($R_{\rm eff,min} = 14$ Mpc h^{-1}) and this void catalog represents a relatively low-resolution galaxy sample with densities of 3×10^{-4} particles per cubic Mpc h^{-1}, matching the number density and clustering of the SDSS DR9 galaxy sample (Dawson et al. 2013) using the parameters found by Manera et al. (2013) ($\sigma_{\log M} = 0.596$, $M_0 = 1.2 \times 10^{13} h^{-1} M_\odot$, $M_1' = 10^{14} h^{-1} M_\odot$, $\alpha = 1.0127$, and $M_{\rm min}$ chosen to fit the mean number density). The HOD Dense catalog has 9503 voids with effective minimum radii $R_{\rm eff,min} = 7$ Mpc h^{-1} and includes relatively high-resolution galaxy samples with densities of 4×10^{-3} dark matter particles per cubic Mpc h^{-1}, matching the SDSS DR7 main sample (Strauss 2002) using one set of parameters found by Zehavi et al. (2011) ($\sigma_{\log M} = 0.21$, $M_0 = 6.7 \times 10^{11} h^{-1} M_\odot$, $M_1' = 2.8 \times 10^{13} h^{-1} M_\odot$, $\alpha = 1.12$). The N-body Mock catalog is a single HOD Mock in real space at $z = 0.53$, generated by a dark matter simulation of 4096^3 particles (with a particle mass resolution $7.36 \times 10^{10} h^{-1} M_\odot$) in a 4 Gpc h^{-1} box and is tuned to SDSS DR9 in full cubic volume by using the HOD parameters found in Manera et al. (2013). Although the N-body Mock catalog is processed slightly differently than HOD Sparse and HOD Dense, it is a HOD mock catalog and it uses *Planck* first-year cosmological parameters (Planck Collaboration 2014). The N-body Mock consists of 155, 196 voids (Sutter et al. 2014b).

It is particularly important to mention that both the N-body Mock and HOD Sparse catalogs represent voids generated from the relative low-density galaxy mocks, while HOD Dense consists of void populations generated from a relatively high-density galaxy mock catalog. As a result, these samples may help us to identify some possible relations between the void size distribution and the environment. Here we take into account the maximum tree depth as an indicator of an environment. Note that the maximum tree depth is the length from the root to the tip of the tallest tree in the hierarchy, and it indicates the amount of substructures in the most complex void in the sample (Sutter et al. 2014a). According to this, HOD Dense and Sparse have maximum tree depths of 10 and 4, respectively (Sutter et al. 2014a), while N-body Mock shows only the root voids, at the base of the tree hierarchy, and therefore they do not have parents, which indicates that the maximum tree depth of this sample is 0 (Sutter et al. 2014b).

Here the samples HOD Sparse, HOD Dense, and N-body Mock are denoted as $R^{(1)} = (r_i^{(1)})_{1 \leqslant i \leqslant 1,422}$, $R^{(2)} = (r_i^{(2)})_{1 \leqslant i \leqslant 9,503}$, and $R^{(3)} = (r_i^{(3)})_{1 \leqslant i \leqslant 155,196}$ respectively. While $R^{(i)}$ $(i = 1, 2, 3)$ represent the names of the samples, $r^{(i)}$ stand for data points.

3. EXPLORATORY PHASE

The first step in the exploratory phase consists of examining the basic graphical and numerical indicators of the samples as histograms, parameters of location (range, mean, median and mode) or dispersion (standard deviation), and shape (skewness, kurtosis).

Let us first consider the raw data plots. Informative features appear from a cursory inspection of the histograms. Whereas those of $R^{(1)}$ and $R^{(2)}$ (see Figure 3) present the features of samples drawn from a single parent population, an unexpected local mode occurs in $R^{(3)}$ about the value 50 Mpc h^{-1}; see the left panel of Figure 1. This suggests that N-body Mock, $R^{(3)}$, should be usefully considered as a sample drawn not from a single population but from two or more void populations with different central density values. In other words, it is reasonable to perform a task of classification. The latter is meant in the classical sense in mathematical statistics, as, e.g., in Section 44.1 of Kendall & Stuart 1977b); it consists of differentiating between two or more populations on the basis of multivariate measurements.

In this respect a closer inspection of the 14 remaining numerical characteristics of the voids suggests that the central density will provide us with a criterion of classification. Out of the 155, 196 voids under study, 147, 528 voids have a central density equal to zero. The sub-sample of $R^{(3)}$ consisting of the radii of these voids will be denoted as $R_0^{(3)}$. Remarkably, the local mode observed in $R^{(3)}$ no longer occurs in $R_0^{(3)}$. The latter satisfies the basic graphical properties expected from a sample drawn from a single population; see Figure 1 (right panel).

Let us now turn to the numerical study of our samples $R^{(1)}$, $R^{(2)}$, and $R_0^{(3)}$. The mean and the centered moments of a sample

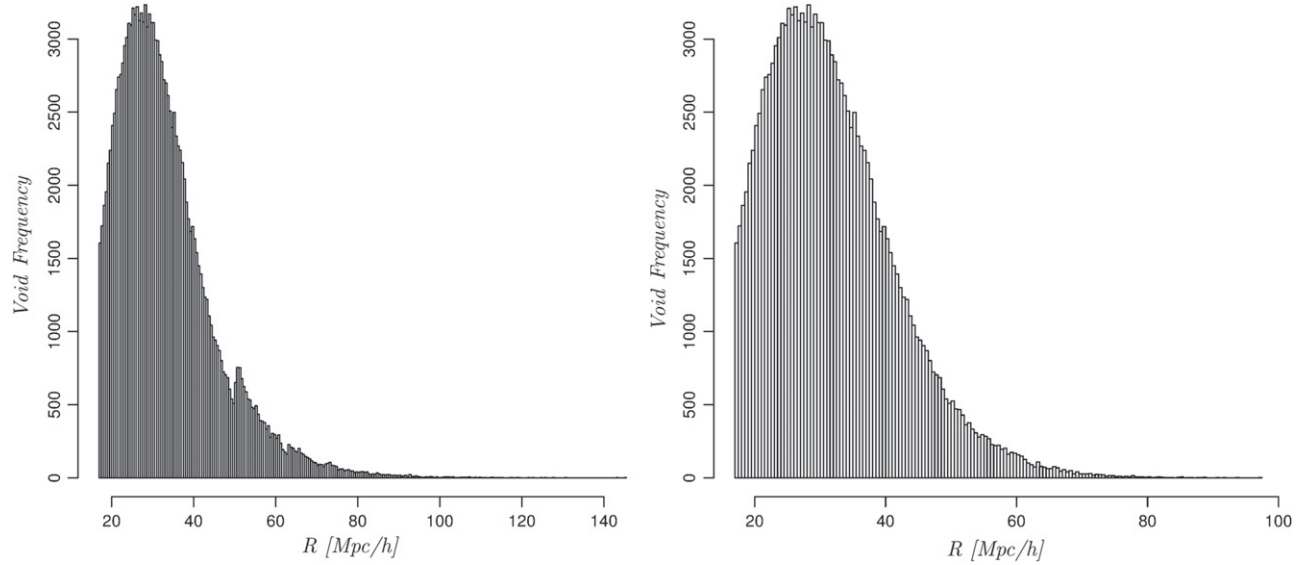

Figure 1. Void size distributions of the full N-body Mock data set, $R^{(3)}$ (left panel) and its sub-sample, $R_0^{(3)}$ consisting of the voids with only zero central density (right panel).

$R = (r_i)_{1 \leqslant i \leqslant N}$ are $\overline{R} = N^{-1} \sum_{i=1}^{N} r_i$ and $m_k = N^{-1} \sum_{i=1}^{N} (r_i - \overline{R})^k$, $k = 2, 3, \ldots$. The variance m_2 is the most common index of dispersion, whereas the widely used indices of the shape of a sample distribution are the skewness $b_1 = m_3^2/m_2^3$ and kurtosis $b_2 = m_4/m_2^2$ (see formulas (1.235)–(1.236) on p. 51 in Johnson et al. 1994 and (3.85)–(3.86) on p. 85 in Kendall & Stuart 1977a). The values of these moments computed from the samples are given in Table 1.

In Table 1, we see that the three empirical distributions share the property of being significantly positively skewed ($b_1 > 0$) and leptokurtic ($b_2 > 3$). Recall that the skewness b_1 measures the degree to which a distribution is asymmetrical. The kurtosis b_2 measures the curvature of a distribution. While a leptokurtic (resp. platykurtic) distribution is characterized by a high (resp. low) degree of peakedness, a mesokurtic distribution presents a peakedness considered medium whenever $b_2 = 3$, which is the kurtosis of a normal distribution with arbitrary average and variance (see Sheskin 2011 for more details). Among the classical parametric families of distributions, the density of the three-parameter log-normal random variable $\boldsymbol{R} = \mathrm{LN}(\theta, \zeta, \sigma)$, given by

$$p_{\theta,\zeta,\sigma}(r) = \frac{e^{-\frac{\{\log(r-\theta)-\zeta\}^2}{2\sigma^2}}}{(r-\theta)\sigma\sqrt{2\pi}}, \quad r > \theta, \tag{1}$$

(see (14.2)′ on p. 208 in Johnson et al. 1994) appears as a natural candidate to fit the size distributions of the samples HOD Dense, HOD Sparse, and N-body Mock. A random variable $\boldsymbol{R}$ follows the distribution of $\mathrm{LN}(\theta, \zeta, \sigma)$ if $\log(\boldsymbol{R} - \theta)$ follows a Gaussian distribution with a mean ζ and variance σ^2. This three-parameter log-normal distribution is strikingly similar to the galaxy distributions that Kayo et al. (2001) obtain from the N-body simulations. Our motive is confirmed by the close proximity of the pairs (b_1, b_2), computed from our sample, to the log-normal line observed in the skewness-kurtosis plane represented in Figure 2 (see Figure 12.3 in Johnson et al. 1994 for a similar pattern involving several parametric families of distributions). As is seen in Figure 2, the void distributions from the mock samples of CVC can be considered to behave as log-normal distributions with respect to their skewness and kurtosis.

Table 1
First Moments, Skewness b_1 and Kurtosis b_2 of the Sample Distributions

Sample	$R^{(1)}$	$R^{(2)}$	$R_0^{(3)}$
$\overline{r}$	40.415	16.671	31.978
m_2	235.752	40.052	96.034
m_3	4,416	386.278	898.965
m_4	306,994	11,109	38,131
b_1	1.488	2.322	0.912
b_2	5.523	6.925	4.134

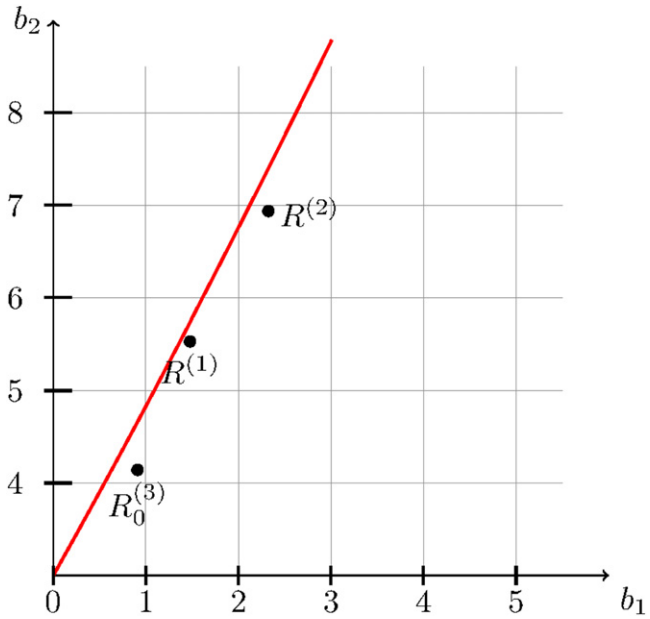

Figure 2. Kurtosis b_2 as ordinate against the skewness b_1 as abscissa for the log-normal distribution (red solid line). The dots represent the kurtosis and the skewness of the size distributions of voids for HOD Sparse $R^{(1)}$, HOD Dense $R^{(2)}$, and the sub-sample of N-body Mock $R_0^{(3)}$.

Table 2
Estimates Computed from the Sample Distributions

Sample	$R^{(1)}$	$R^{(2)}$	$R_0^{(3)}$
$\hat{\theta}$	0.770	3.284	0.223
$\hat{\zeta}$	3.610	2.493	3.412
$\hat{\sigma}$	0.373	0.449	0.301

Some characteristics of a random variable LN(θ, ζ, σ) are,

1. range: (θ, ∞);
2. mode: $\theta + e^{\zeta-\sigma^2}$;
3. median: $\theta + \exp\zeta$;
4. mean: $\theta + e^{\zeta+\sigma^2/2}$.

This concludes the exploratory phase of our study and we will now deal with the task of estimating the three parameters of the log-normal model for $R^{(1)}$, $R^{(2)}$, and $R_0^{(3)}$.

4. ESTIMATION

The task of estimation for a three-parameter log-normal distribution involves well-known computational difficulties related to the non-convergence of the maximum-likelihood estimator for the threshold parameter θ. This issue is discussed in Johnson et al. (1994, Section 4.2). Therefore the most convenient standard tool is the moment-method. This consists of equating the first three sample moments, $\overline{R}$, m_2, and m_3, to the corresponding population values. This estimation method is described in Johnson et al. (1994, Section 4.2), in particular formulas (14.45)–(14-46). It proceeds as follows. First, one has to compute the unique real-valued solution $\tilde{\omega}$ of the equation,

$$(\tilde{\omega} - 1)(\tilde{\omega} + 2)^2 = b_1 = m_3^2/m_2^3, \tag{2}$$

which is easily seen to be given by the explicit formula,

$$\tilde{\omega} = [1 + b_1/2 + \sqrt{(1 + b_1/2)^2 - 1}]^{1/3} + [1 + b_1/2 - \sqrt{(1 + b_1/2)^2 - 1}]^{1/3} - 1, \tag{3}$$

(see Johnson et al. 1994, particularly the last equality on p. 228). Once this auxiliary coefficient has been determined, the estimators $(\hat{\theta}, \hat{\zeta}, \hat{\sigma})$ of (θ, ζ, σ) are,

1. $\hat{\sigma} = \log(\tilde{\omega})^{1/2}$,
2. $\hat{\zeta} = \frac{1}{2}\log[m_2\tilde{\omega}^{-1}(\tilde{\omega} - 1)^{-1}]$,
3. $\hat{\theta} = \overline{R} - \tilde{\omega}^{1/2}e^{\hat{\zeta}}$,

(see the formulas of Section 4.2 in Johnson et al. 1994). The estimates computed from our samples are given in Table 2. The goodness-of-fit of our model is illustrated in two ways. First, it is illustrated numerically by Tables 3–5, where expected values and observed values are given for three void distributions. Second, it is illustrated graphically by Figure 3, which displays the sample histograms with the curves of the log-normal densities whose parameters are the estimates computed from the samples. In both cases the goodness-of-fit is evidently fairly good. Note that for $R^{(1)}$ the p-value associated with Kolmogorov, Cramér–von Mises, and Anderson–Darling statistics equals the remarkably high value of 0.99.

Table 3
Interval, Observed, and Expected Void Frequencies of the Data Set $R^{(1)}$ (HOD Sparse), LN(0.770, 3.610, 0.373)

Interval	Observed Frequencies	Expected Frequencies
[0,10)	0	0
[10,20)	56	57
[20,30)	334	320
[30,40)	429	424
[40,50)	296	306
[50,60)	163	168
[60,70)	77	81
[70,80)	37	37
[80,90)	16	16
⩾90	14	13

Table 4
Interval, Observed, and Expected Void Frequencies of the Data Set $R^{(2)}$ (HOD Dense), LN(3.284, 2.493, 0.449)

Interval	Observed Frequencies	Expected Frequencies
[0,5)	0	0
[5,10)	953	901
[10,15)	3448	3576
[15,20)	2947	2782
[20,25)	1267	1326
[25,30)	500	547
[30,35)	215	218
[35,40)	99	88
[40,45)	39	36
[45,50)	23	16
⩾50	12	12

Table 5
Interval, Observed, and Expected Void Frequencies of the Data Set, $R_0^{(3)}$ of the N-body Mock Catalog with the Estimates LN(0.223, 3.412, 0.301)

Interval	Observed Frequencies	Expected Frequencies
[0,5)	0	0
[0,20)	11 487	11 495
[20,25)	28 035	25 508
[25,30)	31 631	33 088
[30,35)	27 770	29 407
[35,40)	20 605	20 781
[40,45)	12 894	12 718
[45,50)	7 295	7 098
[50,55)	3 848	3 732
[65,70)	2 129	1 889
⩾60	1 834	1 812

5. CONCLUSIONS AND DISCUSSION

We now know that voids dominate the total observed volume of the large-scale structure (Kirshner et al. 1981; Geller & Huchra 1989; da Costa et al. 1994; Shectman et al. 1996) and that they are very sensitive to their environments, which can strongly affect their shapes as well their distributions (Sheth & van de Weygaert 2004; Russell 2013, 2014). Therefore the void size distribution functions may play an important role in understanding the dynamical processes affecting the

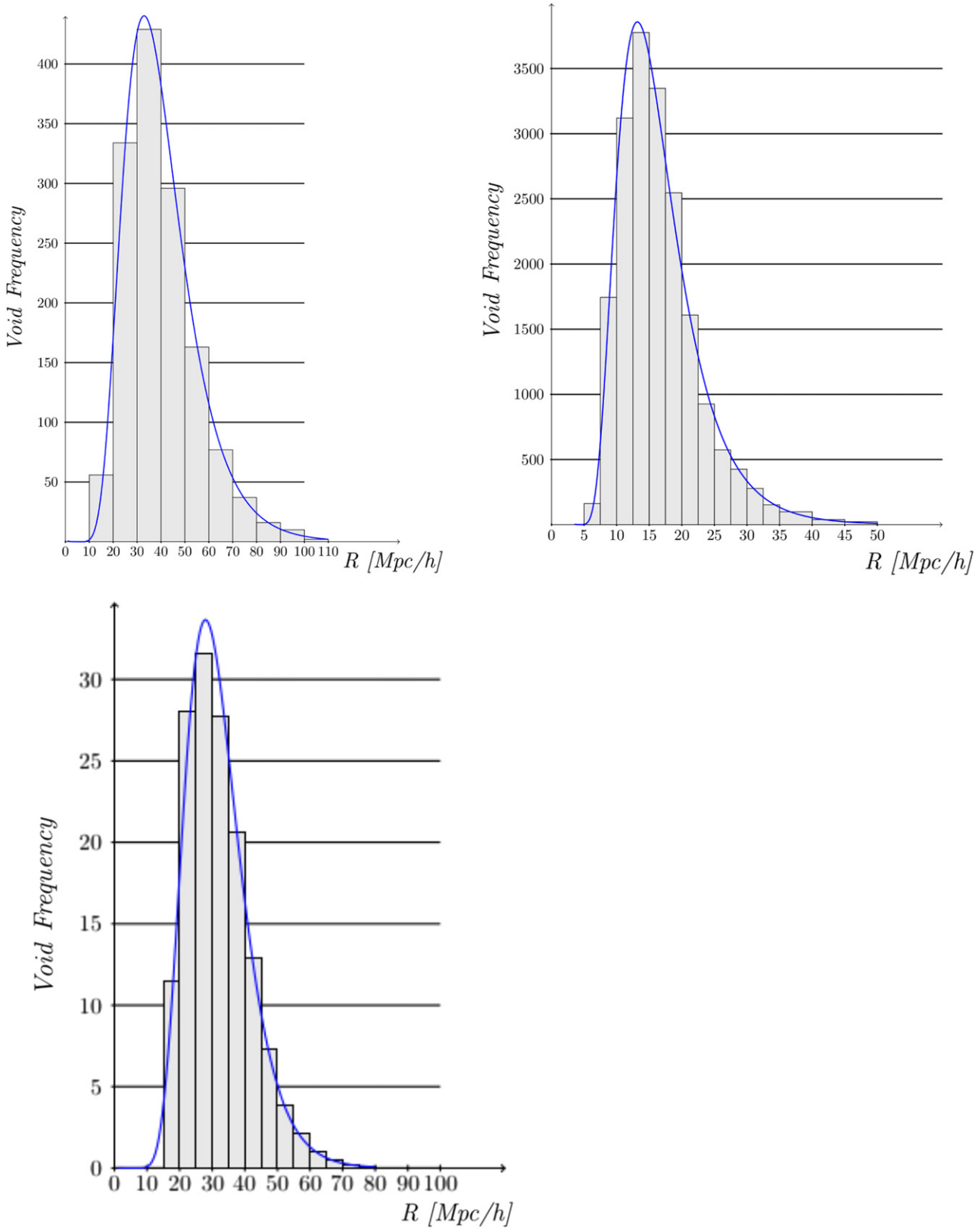

Figure 3. Void frequencies from the samples $R^{(1)}$, $R^{(2)}$, and $R_0^{(3)}$ with the density curves of LN(0.770, 3.610, 0.373), LN(3.284, 2.493, 0.449), and LN(0.223, 3.412, 0.301), respectively (from the top left, top right, and bottom left).

structure formation of the universe (Goldberg & Vogeley 2004; Croton et al. 2005; Hoyle et al. 2005).

In this study, we show that the system of three-parameter log-normal distributions gives a satisfactory model for the size distributions of voids obtained from the three different HOD mock catalogs of CVC; *N*-body Mock, HOD Sparse, and HOD Dense. These catalogs are especially important since they may allow us to compare HOD mock voids in different

environments. As mentioned previously, HOD Dense and HOD Sparse provide us with the data sets for obtaining the size distributions at $z = 0$ (Sutter et al. 2014a), while the N-body Mock catalog has 155, 196 voids at $z = 0.53$ in the full volume-set of mock galaxies (Sutter et al. 2014b). We find that the N-body Mock catalog, $R^{(3)}$, shows sub-void populations with respect to their central densities (left panel of Figure 1). From these
sub-populations we choose the population including voids with only zero central density. This sub-sample is named $R_0^{(3)}$, which consists of 147, 528 voids and a maximum tree depth of 0. The size distribution obtained from this sub-sample satisfies the three-parameter log-normal distribution of other catalogs. As is seen, this sub-sample has the smallest value of the maximum tree depth compared to the two catalogs we use in this study. This is especially important since it provides us with a third environment to compare the relationship between the maximum tree depth and the two shape parameters of the log-normal size distribution. Then we compare the void size distributions obtained from these mock data. As a result, we show that all three samples fit the three-parameter log-normal distribution. As is seen in Figure 2 and Table 3, the HOD Sparse sample, $R^{(1)}$, remarkably passes the Kolmogorov, Cramér–von Mises, and Anderson–Darling statistical tests with a high p-value of 0.99, which seldom occurs with a large sample.

Recalling that the skewness b_1 of the log-normal distribution indicates the asymmetry of the void size distribution, in Figure 2 and Tables 3–5, HOD Dense $R^{(2)}$, is the most asymmetric sample, with a skewness owith the density curves off $b_1 = 2.322$ compared to the void populations in N-body Mock, $R_0^{(3)}$, with $b_1 = 0.912$, and HOD Sparse, $R^{(1)}$, with $b_1 = 1.488$ (also see Table 1 to compare with the other parameters of the void size distributions between the three samples). Also, the void size distribution of N-body Mock, $R_0^{(3)}$, is the least asymmetric of the three samples. In addition, we find that the void distribution in HOD Dense indicates a higher kurtosis, $b_2 \approx 7$, than the void population in the HOD Sparse sample, $b_2 \approx 5.5$, while the N-body Mock sub-sample shows the lowest kurtosis $b_2 = 4.134$. As is seen, there may be a relation between the values of kurtosis, skewness, and the maximum tree depth of the void size distribution. Note that HOD Dense has the highest maximum tree depth, 10, as well as the highest values of kurtosis $b_2 \approx 7$ and skewness $b_1 = 2.322$ of the three samples, while the size distribution obtained from the N-body Mock sub-sample $R_0^{(3)}$ consists of the lowest values of the distribution shape parameters ($b_1 = 0.912$ and $b_2 = 4.134$) and a maximum tree depth of 0, taking into account only the root voids. Considering that the maximum tree depth is a measure of the amount of void substructures, it may be possible to determine a connection between the number of void substructures and how the void size distribution is skewed and peaked.

Note that all the void size distributions have positive skewness and the strength of the skewness may be related to the length of the tail of the distribution in radius (see Figure 3). In addition, it is natural to see a thin tail due to the shape of the log-normal distribution in any case. For example, HOD Dense ($R^{(2)}$) and HOD Sparse ($R^{(1)}$) show slightly longer tails on the large radius side compared to the distribution of the N-body Mock sub-sample ($R_0^{(3)}$), which has the smallest skewness (Table 1 and Figure 3). These thin tail formations of the log-normal void size distributions derived from the simulations are particularly interesting due to a possible connection of dynamical interplay between void size and galaxy distributions. Assuming that the void size is proportional to its mass content ($R^3 \approx M$), here we realized that the one-point galaxy distribution obtained by Kayo et al. (2001) from the N-body simulations seems to be strikingly similar to the log-normal void distribution that we obtain (see Equation (1)). On the other hand, the void log-normal distribution has one more additional parameter than the galaxy distribution of Kayo et al. (2001), which provides a better fit to the void data sets. From small-scale to large-scale matter distributions in the universe, i.e., from the interstellar to the intergalactic medium, there are models and observational studies to explain the skewed log-normal matter distribution. For example, Schneider et al. (2013) conclude that statistical density fluctuations, intermittency, and magnetic fields can cause excess from the log-normal distribution, and also that the core formation and/or global collapse of filaments and a non-isothermal gas distribution lead to a power-law tail. Pudritz & Kevlahan (2013) show that the observed skewed log-normal galaxy distribution could arise from repeated shock interactions, without the need for fully developed turbulence. Recently, it has been shown that it is possible to see large-scale bulk flows that indicate the possibility of the effects of large-scale turbulence on structure formation may not be ignored. Related to this, Wang et al. (1979) found a large-scale bulk flow of approximately a sphere of radius 170 Mpc h^{-1} produced by the massive structures associated with the SDSS Great Wall. Taking into account the fact that voids and galaxies/over-densities are formed from the same primordial density field, then the similarity between the void size and galaxy mass distributions may be expected. As a result, it is possible that the galaxy and void distributions are analogous to each other. This analogy may be caused by void substructures due to the inner as well as outer tidal streams, since it has been shown that the interiors of voids are filled with subvoids, galaxies, and even filaments by N-body simulations and observations (Benson et al. 1986; Mathis & White 2002; Gottlöber et al. 2003; Kreckel et al. 2011). Note that proving this connection between void and galaxy mass distribution deserves its own systematic statistical study, particularly using real data sets, therefore we leave the answer of this puzzle to a future study.

In addition to this, the peaks of the distributions are given by $\bar{r}$ in Table 1. $\bar{r}$ provides the most dominant void sizes in the void size distribution. Figure 3 and Table 1 show that $\approx$40.4 Mpc h^{-1} size voids dominate the HOD Sparse sample, $R^{(1)}$, while the size distributions of the HOD Dense sample, $R^{(2)}$, and the N-body Mock sample, $R_0^{(3)}$, are dominated by voids with radii $\approx$17 Mpc h^{-1} and $\approx$32 Mpc h^{-1}, respectively. This indicates that the size distributions of HOD Sparse and the N-body Mock sub-sample tend to have larger voids compared to the voids in HOD Dense due to their low-density environments. This result agrees with Watson et al. (2014), Jennings et al. (2013), and Russell (2013, 2014). On the other hand, as previously mentioned, the N-body Mock catalog has higher redshift, $z = 0.53$, than the other two samples, therefore the N-body Mock sample may imply larger size voids at $z = 0$ compared to HOD Sparse.

Although there have been some theoretical attempts to understand the dynamical, thermal, and chemical evolution of the void population, and the interplay between galaxies and the intergalactic medium (Sheth & van de Weygaert 2004; Shang et al. 2007; Viel et al. 2008; Russell 2013, 2014), a systematic

investigation of void size distributions is still lacking. Therefore, in this study we show that the void size distributions satisfy the three-parameter log-normal distribution and that this distribution may be a good candidate to show the large-scale environmental effects on voids. It also seems that there is a relation between the strength of the maximum tree depth and the three-parameter log-normal void size distribution parameters: skewness and kurtosis. On the other hand, an extended statistical study of the size distributions of voids in real data, as well as more simulated data, is crucial to fully understand the effects of the large-scale dynamical network between galaxies, filaments, and voids, and obtain more precise results.

The authors would like to thank Paul Sutter for insightful comments and suggestions regarding the Cosmic Void Catalogs. The three void catalogs used here can be found in the folder void$_{c}atalog_{2}$014.06.08 at http://www.cosmicvoids.net.

REFERENCES

Benson, A. J., Hoyle, F., Torres, F., & Vogeley, M. S. 1986, MNRAS, 340, 160
Bernardeau, F. 1992, ApJ, 392, 1
Bernardeau, F. 1994, A&A, 291, 697
Bouchet, F. R., Strauss, M. A., Davis, M., et al. 1993, ApJ, 417, 36
Coles, P., & Jones, B. 1991, MNRAS, 248, 1
Colombi, S., Bouchet, F. R., & Schaeffer, R. 1995, ApJS, 96, 401
Croton, D. J., Farrar, G. R., Norberg, P., et al. 2005, MNRAS, 356, 1155
da Costa, L. N., Geller, M. J., Pellegrini, P. S., et al. 1994, ApJL, 424, L1
Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10
Elizalde, E., & Gaztanaga, E. 1992, MNRAS, 254, 247
Fisher, N. I., Lewis, T., & Embleton, B. J. J. 1993, Statistical Analysis of Circular Data (Cambridge: Cambridge Univ. Press)
Fry, J. N. 1986, ApJ, 306, 358
Gaztañaga, E., & Yokoyama, J. 1993, ApJ, 403, 450
Geller, M. J., & Huchra, J. P. 1989, Sci, 246, 897
Goldberg, D. M., & Vogeley, M. S. 2004, ApJ, 605, 1
Gottlöber, S., Łokas, E. L., Klypin, A., & Hoffman, Y. 2003, MNRAS, 344, 715
Hamilton, A. J. S. 1985, ApJL, 292, L35
Hoyle, F., Rojas, R. R., Vogeley, M. S., & Brinkmann, J. 2005, ApJ, 620, 618
Jennings, E., Li, Y., & Hu, W. 2013, MNRAS, 434, 2167
Johnson, N. L., Kotz, S., & Balakrishnan, N. 1994, Continuous Univariate Distributions, Vol. 1 (2nd ed.; New York: Wiley)
Kayo, I., Taruya, A., & Suto, Y. 2001, ApJ, 561, 22
Kendall, M., & Stuart, A. 1977a, The Advanced Theory of Statistics, Vol 1 (2nd ed.; New York: Macmillan)
Kendall, M., & Stuart, A. 1977b, The Advanced Theory of Statistics, Vol 3 (2nd ed.; New York: Macmillan)
Kirshner, R. P., Oemler, A., Jr., Schechter, P. L., & Shectman, S. A. 1981, ApJL, 248, L57
Kofman, L., Bertschinger, E., Gelb, J. M., Nusser, A., & Dekel, A. 1994, ApJ, 420, 44
Kreckel, K., Platen, E., Aragón-Calvo, M. A., et al. 2011, AJ, 141, 4
Lahav, O., Itoh, M., Inagaki, S., & Suto, Y. 1993, ApJ, 402, 387
Lavaux, G., & Wandelt, B. D. 2012, ApJ, 754, 109
Manera, M., Scoccimarro, R., Percival, W. J., et al. 2013, MNRAS, 428, 1036
Mathis, H., & White, S. D. M. 2002, MNRAS, 337, 1193
Neyrinck, M. C. 2008, MNRAS, 386, 2101
Patiri, S. G., Betancort-Rijo, J., & Prada, F. 2006, MNRAS, 368, 1132
Planck Collaboration 2014, A&A, 571, A19
Pudritz, R. E., & Kevlahan, N. K.-R. 2013, RSPTA, 371, 20120248
Russell, E. 2013, MNRAS, 436, 3525
Russell, E. 2014, MNRAS, 438, 1630
Ryu, D., Kang, H., Hallman, E., & Jones, T. W. 2003, ApJ, 593, 599
Saslaw, W. C. 1985, Gravitational Physics of Stellar and Galactic Systems (Cambridge: Cambridge Univ. Press)
Schneider, N., André, P., Könyves, V., et al. 2013, ApJL, 766, L17
Shang, C., Crotts, A., & Haiman, Z. 2007, ApJ, 671, 136
Shectman, S. A., Landy, S. D., Oemler, A., et al. 1996, ApJ, 470, 172
Sheskin, D. J. 2011, Handbook of Parametric and Nonparametric Statistical Procedures (Boca Raton, FL: CRC Press)
Sheth, R. K., & van de Weygaert, R. 2004, MNRAS, 350, 517
Strauss, M. A. 2002, AJ, 124, 1810
Sutter, P. M., Lavaux, G., Hamaus, N., et al. 2014a, MNRAS, 442, 462
Sutter, P. M., Lavaux, G., Wandelt, B. D., et al. 2014b, MNRAS, 442, 3127
Sutter, P. M., Lavaux, G., Wandelt, B. D., & Weinberg, D. H. 2012, ApJ, 761, 44
Szapudi, I., & Colombi, S. 1996, ApJ, 470, 131
Taylor, A. N., & Watts, P. I. R. 2000, MNRAS, 314, 92
Tinker, J. L., Weinberg, D. H., & Zheng, Z. 2006, MNRAS, 368, 85
Ueda, H., & Yokoyama, J. 1996, MNRAS, 280, 754
Viel, M., Colberg, J. M., & Kim, T.-S. 2008, MNRAS, 386, 1285
Wang, H., Mo, H. J., Yang, X., & van den Bosch, F. C. 1979, MNRAS, 420, 1809
Watson, W. A., Iliev, I. T., Diego, J. M., et al. 2014, MNRAS, 437, 3776
White, S. D. M. 1979, MNRAS, 186, 145
Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, ApJ, 736, 59
Zheng, Z., Coil, A. L., & Zehavi, I. 2007, ApJ, 667, 760