Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **481,** 4588–4601 (2018)
Advance Access publication 2018 September 26

doi:10.1093/mnras/sty2613

Density distribution of the cosmological matter field

Anatoly Klypin,[1]★ Francisco Prada,[2] Juan Betancort-Rijo[3,4] and Franco D. Albareti[5]†

[1]*Astronomy Department, New Mexico State University, Las Cruces, NM 88003, USA*
[2]*Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, E-18080 Granada, Spain*
[3]*Instituto de Astrofísica de Canarias, C/Via Lactea s/n, E-38205, La Laguna, Tenerife, Spain*
[4]*Universidad de La Laguna, Dpto. Astrofísica, C/Astrofísico Francisco Sanchez s/n, E-38206 La Laguna, Tenerife, Spain*
[5]*Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, Cantoblanco, E-28049 Madrid, Spain*

Accepted 2018 September 18. Received 2018 September 18; in original form 2017 September 14

ABSTRACT

The one-point probability distribution function (PDF) of the matter density field in the universe is a fundamental property that plays an essential role in cosmology for estimates such as gravitational weak lensing, non-linear clustering, massive production of mock galaxy catalogues, and testing predictions of cosmological models. Here we make a comprehensive analysis of the dark matter PDF, using a suite of ∼7000 *N*-body simulations that covers a wide range of numerical and cosmological parameters. We find that the PDF has a simple shape: it declines with density as a power-law $P \propto \rho^{-2}$, which is exponentially suppressed on both small and large densities. The proposed double-exponential approximation provides an accurate fit to all our *N*-body results for small filtering scales $R < 5$ h^{-1} Mpc with *rms* density fluctuations $\sigma > 1$. In combination with the spherical infall model that works well for small fluctuations $\sigma < 1$, the PDF is now approximated with just few per cent errors over the range of 12 orders of magnitude – a remarkable example of precision cosmology. We find that at ∼5−10 per cent level the PDF explicitly depends on redshift (at fixed σ) and on cosmological density parameter $\Omega_{\rm m}$. We test different existing analytical approximations and find that the often-used lognormal approximation is always 3–5 times less accurate than either the double-exponential approximation or the spherical infall model.

Key words: methods: numerical – galaxies: haloes – dark matter – cosmology: Large-scale structure.

1 INTRODUCTION

The one-point probability distribution function (PDF) of the matter density field in the universe, and its related statistics the distribution of galaxy counts, have a long and somewhat patchy history in cosmology and extragalactic astronomy. It was Edwin Hubble almost a century ago who found that the counts of about 44 000 extra-galactic nebulae distributed over a large area of the sky have a probability distribution that is not Gaussian but can be approximated by a lognormal distribution (Hubble 1934). The statistics of galaxy counts in the Lick survey, in projected cells of size 10 arcmin × 10 arcmin, was studied by Soneira & Peebles (1978), who also discovered that the distribution of the counts is much broader than the Poisson PDF.

The *rms* of galaxy counts σ in cells of size R is an integral over the power spectrum of the galaxy distribution (e.g. Peebles 1980, section 36). As such, in former times, a count-in-cells analysis of the IRAS redshift galaxy survey was performed by Efstathiou et al. (1990), who used the counts as a measure of the 2-point clustering statistics on different scales. Once methods to estimate the correlation function and the power spectrum were developed and new large-scale galaxy surveys were available, the count-in-cells as clustering statistics started to play a secondary role. Higher moments of cell counts depend on correlation functions of order larger than 2. This means that the whole PDF has information not only on the 2-point clustering but also on higher order statistics, which by itself is very valuable information.

At present, a precise description and modeling of the underlying matter density distribution – and biasing prescription that connects the dark matter field with the galaxy distribution – are fundamental to extract cosmological information from current and upcoming large-scale redshift and lensing galaxy surveys (e.g. Taruya et al. 2002; Takahashi et al. 2011; Manera et al. 2013; Carron, Wolk & Szapudi 2015; Kitaura et al. 2016; Clerkin et al. 2017). For this reason in the last years there has been a rejuvenated interest in the cosmic density distribution from both cosmological *N*-body simulations and galaxy surveys.

Wild et al. (2005) estimated the PDF of galaxies in the 2dF redshift survey, using about 200 000 galaxies. Because of a relatively small volume, their analysis was done only for large cells of size

★ E-mail: aklypin@nmsu.edu
† 'la Caixa'-Severo Ochoa Scholar

10–30 Mpc. They found that the lognormal distribution fits the data reasonably well, but the noise in the data did not allow them to make accurate measurements of the PDF. The situation was improved by Hurtado-Gil et al. (2017), using the count-in-cells statistics for galaxies in the SDSS main sample. They used ∼100 000 galaxies and estimated the PDF for spheres of radius $R = (8–24)$ h^{-1} Mpc. They found that the lognormal distribution was very inaccurate (a factor of ∼2 errors) for spheres of $R = 8$ h^{-1} Mpc. A modification of the lognormal distribution (called lognormal + bias) somewhat improved the fits, but it still had ∼50 per cent errors at a low number of galaxy counts. The negative binomial distribution was a much better fit for all filtering scales. At hight redshift, Bel et al. (2016) studied the count-in-cells distribution of ∼30 000 galaxies in the VIPERS redshift survey with the typical average number of galaxies per cell of 0.5–5 and spherical cells of radius $R = (4–8)$ h^{-1} Mpc. They found that the skewed lognormal distribution (a modification of the lognormal distribution with four more free parameters) was not accurate enough to fit the results of observations. Instead, they found that the negative binomial distribution was much more accurate. Yet, Clerkin et al. (2017), using the DES science verification data, confirmed that the lognormal model is a good fit to both the galaxy density contrast and weak lensing convergence PDFs on scales of (3–10) Mpc at median redshift $z = 0.3$. In spite of the fact that at present these seems to be the best observational results, the errors and noise in the data are still substantial.

On the theoretical side the situation is also complicated. There are two types of approaches: models that start with some dynamical description of the non-linear evolution of the density field and proceed to make predictions of the matter PDF (e.g. Betancort-Rijo 1991; Bernardeau 1994; Kofman et al. 1994; Betancort-Rijo & López-Corredoira 2002; Ohta, Kayo & Taruya 2003; Lam & Sheth 2008a), and then there are phenomenological approximations that assume a specific analytical form of the PDF (Coles & Jones 1991; Gaztañaga, Fosalba & Elizalde 2000; Lee et al. 2017; Shin et al. 2017), and find best-fitted parameters of this distribution function to simulation data.

Theoretical models based on the non-linear dynamics typically use either some variant of the spherical infall model (e.g. Ohta et al. 2003; Lam & Sheth 2008a; Neyrinck 2016) or the Zeldovich approximation (e.g. Kofman et al. 1994; Betancort-Rijo & López-Corredoira 2002). These models have made substantial progress and now can make very accurate predictions for relatively large smoothing scales $R \gtrsim 5$ h^{-1} Mpc and small $\sigma \lesssim 1$ (Lam & Sheth 2008a) giving errors less than ∼10 per cent. Not surprisingly, as expected, the models are not very useful and start to fail at larger *rms* fluctuations $\sigma \gtrsim 1$ (Lam & Sheth 2008a; Neyrinck 2016).

One of the disadvantages of the dynamical models is their complexity. They typically require some manipulation of the linear power spectrum, analytical approximations for different terms, and can be quite cumbersome to deal with. This is not a serious impediment to their use, but it is a nuisance. Simple analytical functions can serve as an alternative to more complicated dynamical models.

The lognormal distribution is an example of this approach. It was heavily advocated by Coles & Jones (1991) and is often used for relatively large smoothing scales. There is little justification why the density distribution function should be lognormal. Coles & Jones (1991) argue that under the assumption that the divergence of the peculiar velocity field in Eulerian coordinates grows as the velocities themselves (as given by linear theory) the density field can be expressed as the exponential of a Gaussian field. But while their assumption is acceptable for the Lagrangian divergence, for the Eulerian one there is an additional growth roughly proportional to the cubic root of the normalized density. This leads to a density field that is equal to a Gaussian field to the third power, whose PDF is quite different from a lognormal distribution. Here is the main argument of Coles & Jones (1991): 'The lognormal is one of the simplest ways of defining a fully self-consistent random field that always has $\rho > 0$ and, most importantly, is one of the few non-Gaussian random fields for which interesting properties are calculable analytically.' This says that the PDF should be lognormal because it can be handled analytically – hardly a serious argument. Another argument is of the same caliber: the lognormal distribution is well-known and frequently used in other fields of science (Ohta et al. 2003). Similar arguments were used for other phenomenological models (Gaztañaga et al. 2000; Lee et al. 2017).

The only real justification for the existing phenomenological approximations (including the lognormal) is that they make a fit to *N*-body results. This is the reason why cosmological *N*-body simulations are important for the field. In this paper we use a very large suite of cosmological simulations to produce accurate estimates of the dark matter distribution functions. Our simulations cover a wide range of numerical and cosmological parameters. We use the estimates to test different dynamical models and approximations for the PDF and to study its dependence on redshift and cosmological parameters.

One of our goals in this paper is to make a comprehensive study of the different effects that can be associated with *N*-body results regarding the matter density distribution, such as mass and force resolution, size of the box, shot noise, and cosmic variance. In this regard we find that systematic errors in the PDF can be important. For example, noise related to the discreteness of the density probed by particles is the leading factor of seriously wrong estimates of the PDF in underdense regions.

Generation of mock galaxy catalogues provides a motivation for our study of the density distribution function. One needs to produce thousands of realizations of the dark matter density and velocity fields. This can be done by carefully tuning parameters of simulations and limiting their resolution to a fraction of a megaparsec (see e.g. Tassev, Zaldarriaga & Eisenstein 2013; Chuang et al. 2015; Klypin & Prada 2018). A biasing prescription then connects the dark matter with galaxies. This path requires knowledge of the distribution of dark matter mass on very small scales 100 h^{-1} kpc^{-1} h^{-1} Mpc. This is a challenge because the resolution of these simulations is not sufficient to resolve individual haloes and subhaloes, making it difficult to apply existing tools such as Halo Abundance Matching and Halo Occupation Distribution. A path to solve the problem is to map dark matter to galaxies using a biasing scheme (e.g. Kitaura et al. 2016) that requires the understanding details of the density distribution function and finding limitations to its estimates.

Unfortunately, only very few studies in the literature provide PDF results for small smoothing scales $\lesssim 1$ h^{-1} Mpc (Bouchet, Schaeffer & Davis 1991; Bouchet & Hernquist 1992; Platen 2009; Pandey et al. 2013; Lee et al. 2017). So, we will make an effort to study this regime too. Bouchet et al. (1991), Bouchet & Hernquist (1992), and Platen (2009) find that in the regime of small smoothing and large density the PDF has a power-law shape with a slope of ≈ -2, which is similar to what we find in this work.

This paper is organized as follows. In Section 2, we define quantities related with the PDF and provide details of some analytical approximations in Section 3.1. The spherical infall and the double-exponential models are introduced in Section 3.2. Numerical simulations used in this paper are discussed in

Section 4. Section 5 also presents main features of the PDF. Accuracy of different approximations are discussed in Section 6. Summary of results is given in Section 7. Finally, numerical effects are discussed in Appendix A, and Appendix B presents tables of parameters and fits for the double-exponential approximation.

Results of simulations – tables of PDFs for thousands of realizations at different redshifts and filtering scales – are available on *skiesanduniverses.org* (Klypin, Prada & Comparat 2017).

2 DEFINITIONS

In order to estimate the density distribution function $P(\rho)$ from N-body simulations, we split the computational volume L^3, where L is the box size, with a $3D$ mesh of size $N_{\rm cell}^3$ and use the Cloud-In-Cell (CIC) density assignment scheme to estimate the density ρ at each grid point of the mesh. The cell size of the grid $\Delta x = L/N_{\rm cell}$ defines the smoothing length. The density is normalized to the average matter density $\rho_{av} \equiv \Omega_m \rho_{\rm cr}$, i.e.

$$\rho \equiv 1 + \delta_{\rm NL} = \rho_{\rm DM}/\,\Omega_m \rho_{\rm cr}, \tag{1}$$

where $\delta_{\rm NL}$ is the matter density contrast or overdensity. The index *NL* highlights the fact that ρ is a non-linear quantity and it can be distinguished from the density contrast δ as estimated by the linear theory. Throughout the paper we use the quantity ρ as 'density' in spite of the fact that it is really a normalized density – a dimensionless quantity as shown in equation (1). This is done for convenience to avoid repeating $1 + \delta_{\rm NL}$ in most plots and equations.

The values of density ρ are binned using logarithmically spaced bins with width $\Delta\log_{10}(\rho) = 0.025$–$0.050$. The density distribution function – PDF of the cosmic density field – is then defined as a normalized number of cells with density in the range $[\rho, \rho + \Delta\rho]$:

$$P(\rho) = \frac{\Delta N_{\rm cell}}{N_{\rm cell}^3 \Delta\rho}. \tag{2}$$

The PDF can have a surprisingly large range of values. For example, density can reach values larger than 10^5 for hundreds of cells when we use a large mesh of $\sim 3000^3$ cells in high-resolution simulations. That gives $P(\rho) \sim 10^{-12}$. At the same time the number of cells at low densities can be millions for a small density bin leading to a large PDF value $P(\rho) \gtrsim 1$. In order to avoid a large dynamical range of quantities, we typically plot $\rho^2 P(\rho)$.

By design, the density distribution function is normalized to have the total volume and total mass equal to unity:

$$\int_0^\infty P(\rho){\rm d}\rho = 1, \quad \int_0^\infty \rho P(\rho){\rm d}\rho = 1. \tag{3}$$

The second moment of $P(\rho)$ gives the *rms* density fluctuation of the density field σ, and as such it is related to the non-linear power spectrum $P_{\rm NL}(k)$ of density perturbations:

$$\sigma^2 = \int_0^\infty (\rho - 1)^2 P(\rho){\rm d}\rho = \frac{1}{2\pi^2}\int_0^{k_{Ny}} P_{\rm NL}(k) W^2(k\Delta x) k^2 {\rm d}k, \tag{4}$$

where $W^2(k\Delta x)$ is the power spectrum of the CIC filter with the width Δx, and the integral is truncated at the Nyquist frequency of the mesh $k_{\rm Ny} = \pi/\Delta x$.

Fig. 1 shows our first results on the structure of the PDF. More detailed discussion is given later in Section 5. The plot demonstrates the main trend of the shape of the PDF. In the linear regime of the growth of fluctuations, when $\sigma \ll 1$, the PDF is a Gaussian distribution that quickly acquires a skewed shape as σ increases. With the further increase of σ, the PDF becomes progressively wider and becomes a power law with the slope close to -2 that is smoothly suppressed on small and large densities.

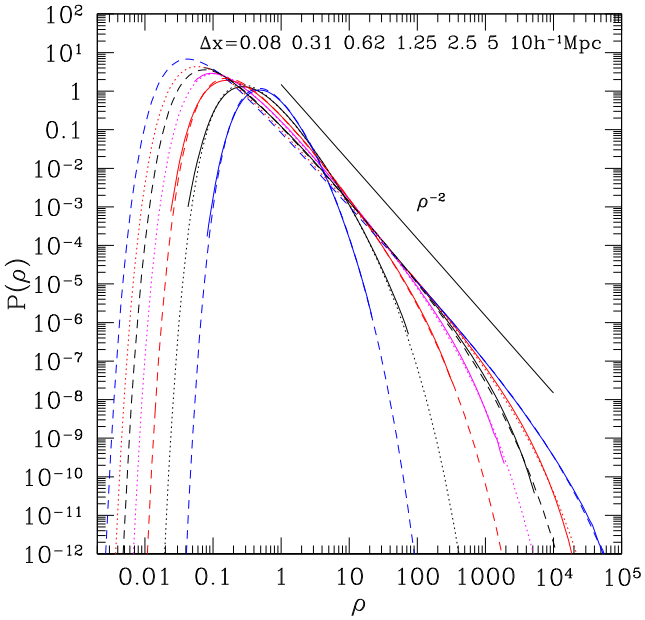

Figure 1. Density distribution function at $z = 0$ for different filtering scales indicated in the plots. Full curves show the results from our simulations. Double-exponential models are presented by dashed and dotted curves. The full line shows the power-law behaviour with the slope -2. As the filtering scale decreases in value, the PDF becomes wider and approaches the power law.

3 METHODOLOGY

Different analytical approximations and theoretical models, as mentioned in Section 1, are used to fit and make predictions for the PDF estimates obtained from numerical simulations. Here we describe both approaches.

3.1 PDF analytical approximations: lognormal, negative binomial, and generalized extreme value

We introduce the lognormal, negative binomial, and the generalized extreme value (GEV) distributions that have been traditionally adopted as analytical approximations for the PDF in many works.

The **log-normal** (LN) distribution function $P_{\rm LN}$ is defined as

$$\rho P_{\rm LN}(\rho) = \frac{1}{\sqrt{2\pi\sigma_{\rm LN}^2}} \exp\left(-\frac{[\ln(\rho) + \sigma_{\rm LN}^2/2]^2}{2\sigma_{\rm LN}^2}\right), \tag{5}$$

where

$$\sigma_{\rm LN}^2 = \ln[1 + \sigma^2]. \tag{6}$$

is the only free parameter. $\sigma_{\rm LN}^2$ can be obtained from results of simulations, and thus should be considered as fixed.

Because the lognormal distribution does not provide accurate fits to numerical simulations, a number of modifications have been proposed (Hamilton 1985; Colombi 1994; Shin et al. 2017). None of those modifications extend the approximation to large densities, hence we do not discuss them in this paper.

The **negative binomial** (NBN) distribution (Betancort-Rijo 2000; Gaztañaga et al. 2000; Bel et al. 2016) is defined as a

discrete distribution. It is the probability $P_N(V)$ to find N particles in a cell of volume V with the average number of particles $\bar{N}$. It can be re-written as a distribution function of density contrast $\rho = N/\bar{N}$:

$$P_{\rm NBN}(\rho) = \frac{1}{\bar{N}} \cdot \frac{\Gamma(N+1/g)}{\Gamma(1/g)\Gamma(N+1)} \cdot \frac{(g\bar{N})^N}{(1+g\bar{N})^{N+1/g}}, \quad (7)$$

where g is a parameter that is defined by *rms* fluctuations of counts: $g = (\sigma_N^2 - \bar{N})/\bar{N}^2$. The average number of particles per cell $\bar{N}$ in general is not an integer number and is defined by the average density and cell volume: $\bar{N} = \bar{\rho}V$. So the two parameters g and $\bar{N}$ that define the NBN distribution are not free and can be fixed from simulations. However, for PDFs with not too small *rms* fluctuations ($\sigma \gtrsim 1$), the fits produced by this distribution are not very accurate. As a result, we decided to treat both $\bar{N}$ and g as free parameters.

While the negative bimodal distribution formally has two parameters, there is little change in $P_{\rm NBN}$ when $\bar{N} > 10$. For cases where NBN makes some reasonable fits, our simulations have typical values of $\bar{N}$ of many hundreds. So, in practice, the NBN PDF depends only on one parameter g, which defines the width of the distribution function: the larger is g, the wider is $P(\rho)$.

At large average number of objects in a cell $\bar{N} \gg 1$ and for large $N > \bar{N}$ the NBN approximation predicts that the density distribution changes with density $\rho = N/\bar{N}$ as follows

$$P_{\rm NBN}(\rho) \approx \frac{1}{\bar{N}^{1/g}\Gamma(1/g)} \rho^{1-1/g} \exp\left(-\frac{\rho}{g}\right). \quad (8)$$

This expression is very different compared to the behaviour of the PDF observed in N-body simulations for large ρ and σ: in that regime $P(\rho) \propto \rho^{-2}\exp(-C\rho^{0.5})$. Thus, the NBN approximation predicts too steep a decline with density and lacks the power-law regime of the N-body PDF.

A GEV distribution as an approximation for the density distribution function was used by Lee et al. (2017) and Repp & Szapudi (2018). It can be written as

$$\rho P_{\rm GEV}(\rho) = \frac{1}{\ln(10)\beta} \frac{\exp(-z^{-1/k})}{z^{1+1/k}}, \; z \equiv 1 + \frac{k}{\beta} \lg_{10}\left(\frac{\rho}{\rho_0}\right), \quad (9)$$

where k, ρ_0, and β are free parameters.

3.2 Spherical infall and double-exponential PDF models

Approximations discussed so far were not based on any dynamical models. They simply make a guess regarding the functional form of $P(\rho)$ and then proceed to finding parameters that produce the best fit. The guess is not based on any insights from the dynamics of clustering either. **Spherical infall models** are different because they are theoretical predictions for the density distribution function that are based on simplified approximations of the non-linear evolution of the density field. Here we closely follow the theoretical framework developed by Betancort-Rijo & López-Corredoira (2002) and Lam & Sheth (2008a). We assume that the linear density field was smoothed with a top-hat filter with radius R_f and corresponding mass M. The variance of the smoothed field is equal to

$$\sigma_L^2(M) = \frac{1}{2\pi^2} \int k^2 {\rm d}k P(k) W^2(kR). \quad (10)$$

In the spherical infall model the mapping from linear density contrast δ_L to the non-linear overdensity ρ is approximated by the following relation (Betancort-Rijo 1991; Bernardeau 1994):

$$\rho = \left(1 - \frac{\delta_L}{\delta_c}\right)^{-\delta_c}, \quad \rho \equiv \frac{M}{\bar{M}}, \quad \bar{M} = \rho_b \Delta x^3, \quad (11)$$

where δ_c is the linear theory prediction for the critical overdensity of collapse. Here we will use $\delta_c = 5/3$ as suggested by Betancort-Rijo & López-Corredoira (2002). For the initial Gaussian fluctuations, this model gives the density distribution function:

$$\rho^2 P(\rho) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\delta_L^2}{2\sigma_L^2}\right) \frac{{\rm d}(\delta_L/\sigma_L)}{{\rm d}\ln\rho}. \quad (12)$$

This is the same as equation (6) in Lam & Sheth (2008a). We also use a modification of the spherical infall model, which is based on the excursion set model (Sheth 1998; Lam & Sheth 2008a):

$$\rho^2 P(\rho) = \frac{1}{\sqrt{2\pi\sigma_L^2}} \exp\left(-\frac{\delta_L^2}{2\sigma_L^2}\right) \left[1 - \delta_L\left(\frac{1}{\delta_c} - \frac{\gamma}{3}\right)\right], \quad (13)$$

where $\gamma(\rho)$ is

$$\gamma = -3\frac{d\ln\sigma_L^2}{d\ln M}. \quad (14)$$

Note that in these equations σ_L is a function of filtering mass M, which in turn depends on density $\rho = M/\bar{M}$. So, $\sigma_L = \sigma_L(\rho)$.

In order to apply the models, we need to adjust the top-hat filtering scale $R_{\rm f}$ used in the spherical infall model so that it matches the Cloud-In-Cell filtering scale used in our simulations. This is done by matching the power spectra of both filters at wavenumbers $k < 0.7k_{\rm Ny}$ by applying $R_{\rm f} = \Delta x/\sqrt{1.3}$, where $k_{\rm Ny} = \pi/\Delta x$ is the Nyquist frequency of the density grid used for density assignment, and Δx is the grid size. Specifically, for every bin with density ρ we find the mass $M = \rho\bar{M}$ and then the top-hat filtering radius is found as $R_{\rm f} = \sqrt{2}\,[3M/4\pi\rho_{\rm cr}\Omega_m]^{1/3}$. When applying the relations given in equations (10–14), we integrate numerically equation (10) (4) with the top-hat filter $W(kR_f)$ and use finite differences to estimate the derivatives in equation (12) and equation (14).

As was found previously (e.g. Betancort-Rijo & López-Corredoira 2002; Lam & Sheth 2008a,b; Neyrinck 2016), the spherical infall model provides good approximations for the PDF in those cases with large filtering scales where the *rms* fluctuation in the simulation box $\sigma(M)$ is relatively small $\sigma \lesssim 1$. This is consistent with our results, which will be presented later. At larger σ (large densities), the model provides results that are much less accurate. Even so, the spherical infall model predicts trends that track our N-body results. This is somewhat unexpected because at large overdensities $\rho \gtrsim 10^2$ the model is clearly outside of the limits of its dynamical applicability. After all, a simplistic treatment of non-linear evolution used by the model cannot be valid for densities that are appropriate for collapsed and virialized haloes.

It is interesting and instructive to find what makes the spherical infall model much better than expected. The last factor in equation (12) is just a correction to the leading terms that are the power-law $P \propto \rho^{-2}$ (see also Fig. 1) and the exponential term on the right hand side of equation (12). The exponential term originates from the assumption that the density distribution function of primordial fluctuations is Gaussian. The ρ^{-2} term comes from integrating the Gaussian distribution function over mass and then by writing it in a differential form. This is basically the same logic as in the Press–Schechter derivation of the mass function of dark matter haloes.

More in detail, we see that the exponential term provides the truncation of the $P \propto \rho^{-2}$ behaviour both on the low-density $\rho < 1$ and on the high-density $\rho \gg 1$ regimes. At low densities (small masses, large σ_L), the truncation is mostly due to large negative values of δ_L. At large ρ (large masses and small σ_L) the decline is related to the combination of decreasing σ_L and increasing δ_L. However, the increase of δ_L is limited: it cannot exceed $\delta_{L,\,\rm max} = \delta_{\rm c} = 5/3$, and the decline in σ_L is not strong enough by itself

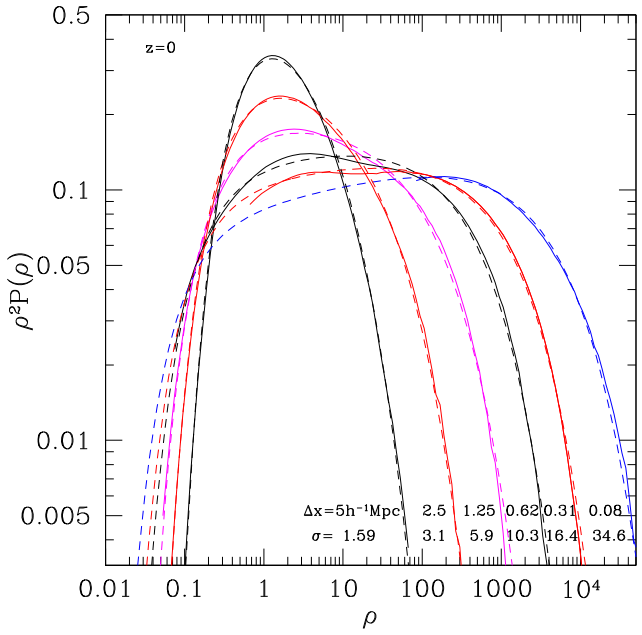

Figure 2. Density distribution function scaled with ρ^2 at $z = 0$ for different filtering scales indicated in the plot. Full curves show the results from our simulations. Double-exponential models are presented by dashed and dotted curves. As the filtering scale decreases in value, the PDF becomes wider and approaches the power law.

to produce a substantial suppression of the PDF. So, the $\rho^2 P(\rho)$ shape becomes nearly flat at very large densities $\rho > 100$ and small smoothing scales $\Delta x \lesssim 1$ h^{-1} Mpc. This is clearly seen in Fig. 2.

At very small smoothing scales $\Delta x \ll 1$ h^{-1} Mpc and large densities there is another regime that sheds light on how the density distribution function should behave at very large densities. When density is larger than ∼100, we are likely to deal with interiors of collapsed dark matter haloes. In this regime the distribution function $P(\rho)$ is a sum over the distribution functions of individual haloes. Assuming that the density profile of a halo can be approximated by the Navarro–Frank–White (NFW) profile, and for a very small filtering scale, we can can derive the PDF provided by a single halo. If $\rho(r)$ is the halo density profile, then the density distribution function given in equation (2) can be written as

$$P(\rho)\mathrm{d}\rho = \mathrm{d}V/V, \tag{15}$$

where dV/V is a fraction of volume with density in the range (ρ, $\rho + d\rho$). The density dependence on radius can be inverted to give us the radius at a given density $r = r(\rho)$. Then PDF in equation (15) takes the form,

$$P(\rho) = -\frac{4\pi r^2(\rho)}{V}\frac{\mathrm{d}r(\rho)}{\mathrm{d}\rho}. \tag{16}$$

This can be applied, for example, to the NFW profile as was extensively done by Pandey et al. (2013). It is easy to see the trend, if the density profile is a power-law $\rho \propto r^{-\alpha}$ with the slope α. In this case

$$\rho^2 P(\rho) \propto \rho^{1-3/\alpha}. \tag{17}$$

In the outer regions of a dark matter halo the density declines as a power law with the slope $\alpha \approx 3$, which implies that $\rho^2 P(\rho) \approx$ constant. It is easy to invert the NFW profile numerically. Results show that for the halo masses in the range $M = (10^{12}-10^{15})$ h^{-1}M$_\odot$ the product $\rho^2 P(\rho)$ is nearly constant for overdensities $\rho = 10^2-10^4$ and declines at much larger densities. This decline is consistent with the fact that the slope α becomes smaller at radii comparable with the characteristic scale radius r_s of the NFW profile. Adding results for many haloes with different masses and concentrations will change the behaviour of the $\rho^2 P(\rho)$ trend at $\rho > 10^4$, but not for the range $\rho = 10^2-10^4$, where it should remain flat because it is also flat for each halo.

In summary, both the spherical infall model and the dark matter halo profiles indicate that the leading term in the density distribution function should be $P(\rho) \propto \rho^{-2}$. The trend should be modified by adding suppression on large and small scales.[1] However insightful, the spherical infall model or the NFW results at large densities cannot be used to produce accurate results for the density distribution function. We use those hints to construct our own approximation for the PDF $P(\rho)$.

Motivated by these results and by our simulations, we design our own model. It is nearly a power-law $P(\rho) \propto \rho^{-\alpha}$ with the slope $\alpha \approx 2$ that is truncated with exponents on both the small and large densities. We call this model **double-exponential**. We tested different shapes for the exponential terms and find that the following expression provides errors less than few per cent at all redshifts, smoothing scales, and cosmologies that we study:

$$P(\rho) = A\rho^{-\alpha}\exp\left[-\left(\frac{\rho_0}{\rho}\right)^{1.1}\right]\exp\left[-\left(\frac{\rho}{\rho_1}\right)^{0.55}\right], \tag{18}$$

where A, α, ρ_0, and ρ_1 are free parameters. As noticed above, the slope $\alpha \approx 2$. The slopes in the exponential terms 0.55 and 1.1 are results of the fitting of numerical PDFs at different smoothing scales and redshifts. One may expect that adding two more free parameters to the approximation (i.e. the slopes in the exponential terms) may further improve the quality of the fits. We find that this is not the case: the data prefer the same slopes, regardless of the value of σ.

The double-exponential model has four formal free parameters. One may use three constraints to limit the parameters: the total mass and volume should be equal to unity (see equation 3), and the second moment of the PDF should be equal to σ measured in simulations (see equation 4). Note that there must be a degree of freedom left after fixing the constraints otherwise the model would not be able to reproduce the numerical results that show that the PDF is not defined solely by σ, and depends on both the redshift and Ω_m. The double-exponential model has this additional degree of freedom.

In practice, we use all four parameters to fit the numerical data. We typically find that the best-fitting parameters provide PDF approximations that within 1–2 per cent conserve the mass and match well the numerical value of σ measured in the simulations. The volume is conserved within 1–5 per cent accuracy.

4 SIMULATIONS

Numerical parameters of our simulations are presented in Table 1, which gives box size, the number of particles, mass of a particle m_p, the number of mesh points N_g^3 (if relevant), cell size of the

[1]Our results are in broad agreement with analysis of the density distribution function in Millenium simulations by Pandey et al. (2013) though the comparison is complicated by the fact that Pandey et al. (2013) used adaptive kernel to find density at position of each particle while we are using a constant kernel (a cube) with volume-weighted PDF. Mass-weighted PDF, in combination with volume assigned to each particle $V \propto 1/\rho$, gives the difference −2 in the slopes of $P(\rho)$. Indeed, Pandey et al. (2013) find that for a wide range of densities $\rho \sim 10^{-1}-10^4$ their PDF is nearly constant. That correspods to $P(\rho) \propto \rho^{-2}$ for our definition of PDF.

Table 1. Numerical and cosmological parameters of different simulations. The columns give the simulation identifier, the size of the simulated box in h^{-1} Mpc, the number of particles, the mass per simulation particle m_p in units $h^{-1}\,M_\odot$, the mesh size N_g^3, the gravitational softening length ε in units of h^{-1} Mpc, the number of time-steps N_s, the amplitude of perturbations σ_8, the matter density Ω_m, the number of realizations N_r.

Simulations	Box	particles	m_p	N_g^3	ε	N_s	σ_8	Ω_m	N_r
A0.5	500^3	1200^3	6.16×10^9	2400^3	0.208	181	0.822	0.307	680
A1.5	1500^3	1200^3	1.66×10^{11}	2400^3	0.625	136	0.822	0.307	4513
A2.5	2500^3	1000^3	1.33×10^{12}	2000^3	1.250	136	0.822	0.307	1960
B0.5	500^3	1600^3	2.66×10^9	3200^3	0.156	271	0.828	0.307	5
D0.5	500^3	1600^3	2.33×10^9	3200^3	0.156	271	0.828	0.270	5
MDPL1	1000^3	3840^3	1.5×10^9	–	0.010	–	0.828	0.307	1
BolshoiP	250^3	2048^3	1.5×10^8	–	0.001	–	0.828	0.307	1

density/force mesh ε, the number of time-steps N_s, cosmological parameters σ_8 and Ω_m, and number of realizations N_r.

In order to estimate the density distribution function, we split each simulation box with a $3-D$ mesh with the cell size Δx. We then use the Cloud-In-Cell (CIC) density assignment to generate the density field. Many filtering sizes Δx were used for each simulation and snapshot.

Different codes were used to make those simulations. The MultiDark 1*Gpc/h* simulation (MDPL1) (Klypin et al. 2016) was done with the GADGET-2 code (Springel 2005). The ART code (Kravtsov, Klypin & Khokhlov 1997) was used to produce the BolshoiP simulation (Klypin, Trujillo-Gomez & Primack 2011). These two simulations have the largest resolution and the largest number density of particles. However, there are only two of these simulations because they are very expensive computationally. Other simulations were carried out with the parallel Particle-Mesh code GLAM (Klypin & Prada 2018). Because the GLAM code is much faster, we have many realizations of the same cosmological and numerical parameters. Simulations B0.5 and D0.5 were designed for testing possible dependence of the PDF on the matter density Ω_m. These simulations have the same random seeds to make comparisons of results easier.

All the simulations were started at initial redshift $z_{init} = 100$ using the Zeldovich approximation. The simulations span three orders of magnitude in mass resolution, a factor of hundred in force resolution, and differ by a factor of 500 in effective volume (see Table 1). Altogether, we use about 7000 simulations to study the density distribution function. To our knowledge, this is the largest set of simulations available today for this type of analysis.

Analysis of different numerical effects in the estimates of the PDF is presented in Appendix A. In summary, our results are mostly dominated by systematics, not by the finite-volume simulation variance. When dealing with individual simulations such as MDPL1 or BolshoiP, we use only bins with more than $N > 100$ cells per bin. For the large sets of simulations A0.5, A1.5, A2.5 we accept bins with more than 10 cells. The discreteness of density assignment may become an issue at small densities while the force resolution may affect the high-density tail of the PDF. We find limits on numerical parameters that should be satisfied to produce the $P(\rho)$ with errors less than few per cent: (1) the filtering scale Δx must be resolved with not fewer than 8 force resolution elements: $\Delta x > 8\varepsilon$, and (2) the number of particles per filtering cell should be not less than 10–20.

5 PDF: MAIN TRENDS

The overall dependence of $P(\rho)$ on filtering scale and *rms* density fluctuation σ is illustrated in Figs 1 and 2. Full curves in the plots show the results from our simulations. At small σ and large Δx the PDF has a peak at $\rho \approx 1$ that shifts to smaller densities as σ increases (and smoothing scale Δx decreases). At the same time the distribution function becomes extremely wide and develops a distinct ρ^{-2} power-law trend that expands to both large and small densities. By scaling out the ρ^{-2} dependence (Figs 2), we reduce the dynamical range and can see better details of the PDF. In particular, we find a very steep decline of $\rho^2 P(\rho)$ on both high- and low-density tails. The double-exponential model equation (18) was tuned to find the shape of both declines. Our results for different filtering scales and different redshifts show that the decline at the large-density limit is $\propto \exp[-(\rho/\rho_1)^\nu]$ and at the small-density limit it is $\propto \exp[-(\rho/\rho_0)^{-2\nu}]$ with $\nu \approx 0.55$.

Note that at very large $\sigma \gtrsim 10$ the peak of $\rho^2 P(\rho)$ has a nearly constant amplitude $\rho^2 P(\rho) \approx 0.11$ but the position of the peak shifts to larger values of ρ. Because the total mass must be preserved ($\int \rho P(\rho) \mathrm{d}\rho = 1$), this implies that at intermediate scales $\rho \approx 1$–100 the PDF $P(\rho)$ should decline when σ increases, and that the slope α in equation (18) must become slightly shallower with increasing σ.

It is often taken for granted that the PDF depends only on the amplitude of the density perturbations on a given filtering scale σ. Indeed, this is the dominant behaviour of $P(\rho)$. However, this is not exactly correct. Our simulations have such a good accuracy that now we can test the dependence of the PDF on redshift at fixed σ and on cosmological parameters.

We first select redshifts and smoothing scales in such a way that σ for two different redshifts are nearly identical. Fig. 3 presents two examples of such cases – one for relatively low $\sigma \approx 1$ and another for larger $\sigma \approx 4$. The differences between PDFs at the same σ and different z are not large: 5–10 per cent depending on the density where the differences are measured. Nevertheless, the differences clearly exist. We also estimate the differences using the spherical infall model and present the results in the right-hand panel of Fig. 3.

The PDF also slightly depends on parameters of the cosmological model. In the left-hand panel of Fig. 4 we compare $z = 0$ results for the B0.5Gpc and D0.5Gpc simulations that differ only by the matter density parameter Ω_m. Again, differences are small but clearly present at $\sim 5-10$ per cent level. The right-hand panel shows predictions for the spherical infall model with the same basic conclusion: PDF does depend on Ω_m.

One can understand why the PDF depends on z and Ω_m, if one realizes that at any given density ρ the value of $P(\rho)$ is formally a functional on the non-linear power spectrum. This means that it depends not only on $\sigma(R)$ but also on the whole shape of the power spectrum. We can analyse the situation by assuming that $P(\rho)$ depends just on $\sigma_L(R)$ (see equation 10) and on its local logarithmic derivative γ at scale R as defined by equation (14). Then the dependence of the PDF on redshift for a fixed value of $\sigma_L(R)$ can be

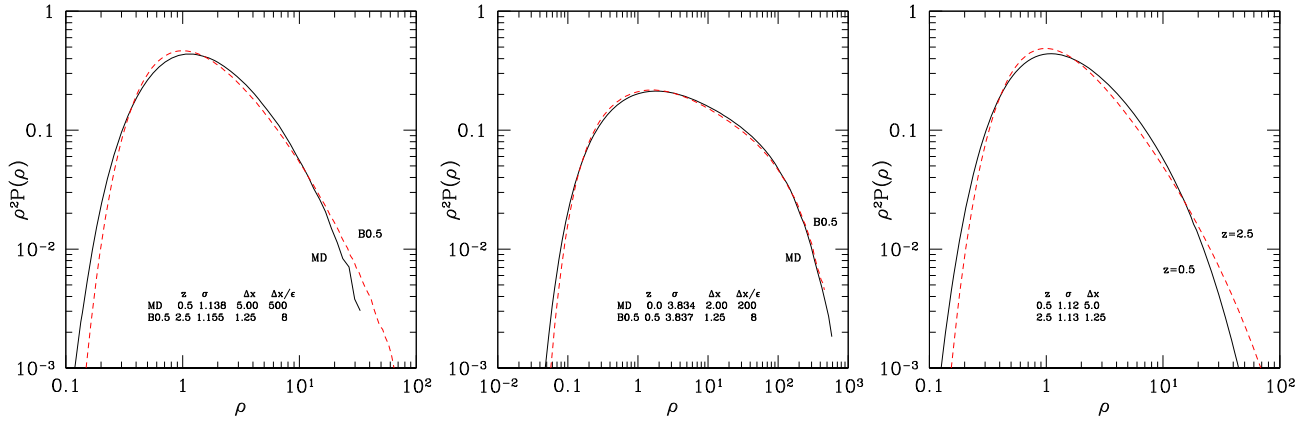

Figure 3. Dependence of $P(\rho)$ on redshift at fixed *rms* amplitude of perturbations σ. Two left-hand panels show examples for different redshifts and for different σ in the *N*-body simulations. In each case we select nearly identical σ. The left-hand panel is for relatively small $\sigma \approx 1$ and for $z > 0.5$. The middle panel is for high σ and $z < 0.5$. The distribution function clearly depends on redshift though the differences are relatively minor for low redshifts. The right-hand panel presents analytical estimates based on the spherical infall model for the same parameters selected for the left-hand panel. The model reproduces the same trend of $P(\rho)$ with redshift.

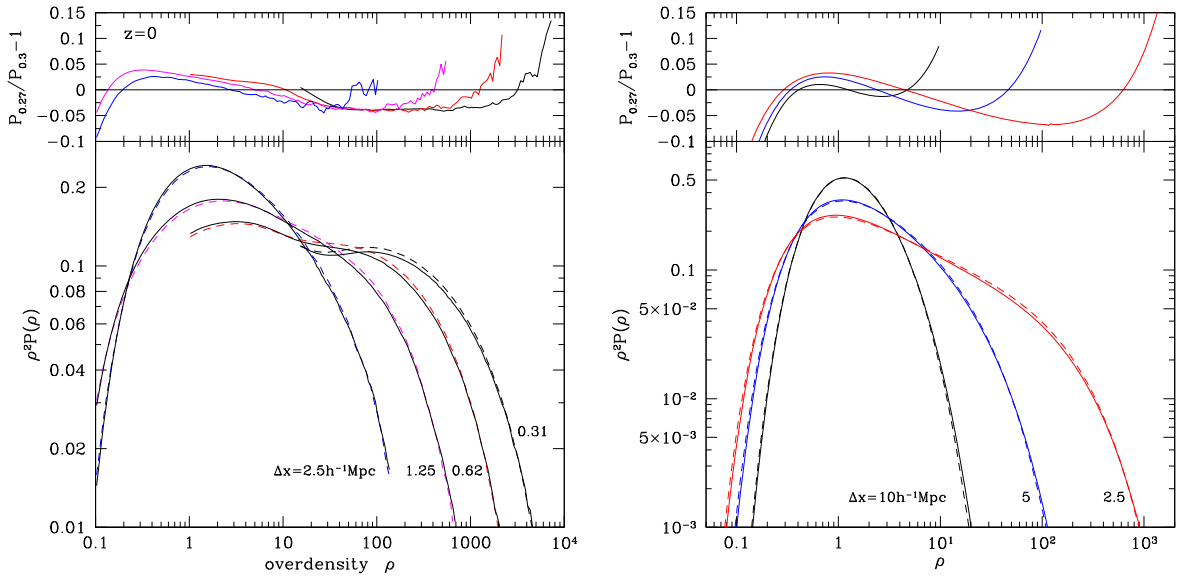

Figure 4. Dependence of $P(\rho)$ on the cosmological matter density $\Omega_{\rm m}$. Dashed (full) curves show results for models with $\Omega_{\rm m} = 0.307$ ($\Omega_{\rm m} = 0.270$) at $z = 0$ for different filtering scales. Other cosmological parameters are the same for both models. The left-hand panel presents comparison of *N*-body simulations B0.5 and D0.5. The right-hand panel shows results of the spherical infall model. The density distribution function weakly but systematically depends on $\Omega_{\rm m}$ with ∼5 per cent deviations at different densities.

explained, because at larger redshits the given value of σ_L is attained at a smaller scale R, where $\sigma_L(R)$ is less steep for a CDM-type power spectrum.

If we look at different terms for $\rho^2 P(\rho)$ in equation (12), then we note that σ_L in the argument of the exponential term is also a function of ρ (falling faster with ρ as γ grows). So, for larger values of γ the PDF will be smaller for large ρ values, where the exponential behaviour dominates, while for small values of ρ it will be larger. The derivative in the right hand side of equation (12) also depends on γ: it will be larger for larger γ values. However, it is only important for the intermediate values of density, where the behaviour can qualitatively be explained by the behaviour in the extremes and the conservation of probability. In short, this all implies a smaller PDF in the low ρ limit and a larger one in the large ρ limit as compared to the PDF at smaller redshift.

The same qualitative behaviour would be observed for smaller values of $\Omega_{\rm m}$ when studying the dependence of the PDF on $\Omega_{\rm m}$ at a given redshift: as $\Omega_{\rm m}$ decreases the power spectrum flattens (up to scales of the order of the horizon at the start of matter domination), leading to a less steep σ_L.

6 PDF: TESTING DIFFERENT APPROXIMATIONS AND MODELS

We start our analysis of different approximations by testing the spherical infall model and the lognormal distribution. Both models are expected to work and typically used for relatively low rms fluctuations $\sigma \lesssim 1$.

Fig. 5 shows results for two modifications of the spherical infall model. We select five configurations with different filtering scales, σ and redshifts. There are some differences between the pure spherical infall model in equation (12) and the excursion set model in equation (13). For example, the excursion set model produces smaller errors at $\rho < 1$ and $\sigma \approx 1$. At the same time, it makes visibly larger errors at $\rho > 1$. For this reason we prefer the standard spherical infall model. It provides smaller than 10 per cent errors for points that are larger than 0.1 of the maximum of the PDF. The error increases substantially in the peripheral regions. If better accuracy is required for the small values of the PDF, one would need to use results of *N*-body simulations, not the approximations.

Results for the lognormal distribution are shown in Fig. 6, where in the left-hand panel we present $z = 0$ results, and results for different redshifts are shown in the right-hand panel. It is clear that the lognormal distribution produces much worse fits as compared with the spherical infall model. For example, errors in the central region are less than 10 per cent only for $\sigma < 0.5$. They become dramatically worse for even slightly larger σ. For $\sigma > 1$ the lognormal distribution makes typically ∼50 per cent errors and, which is even worse, predicts a wrong shape of the PDF. It predicts a wrong position of the maximum; slopes of the declining PDF are not correct.

The lognormal distribution has two advantages as compared with the spherical infall model: (a) It is simple and does not require the machinery of handling the power spectrum and numerical derivatives. (b) It makes mediocre predictions that do not totally fail. We definitely recommend it for quick-and-dirty applications, but not for accurate estimates.

Left-hand panels in Fig. 7 show results for the Negative Binomial distribution given by equation (7). For the large smoothing scale $\Delta x = 20$ h^{-1} Mpc it provides a reasonable fit with errors ∼10 per cent for densities $\rho = 0.4$–2.5. However, it is clear that it has a wrong shape: too steep at large densities and too shallow at small densities. This becomes a serious issue for large values of σ. For example, we could not find any good fit for $\Delta x = 5$ h^{-1} Mpc shown in the bottom panel. It may not be fair to use the NBN for the dark matter PDF because we are in a regime that was not favourable for the NBN: study configurations with a very large number of particles per cell while the NBN was designed to handle very small number $\bar{N}$.

The GEV approximation equation (9) scored much better, as illustrated in the right-hand panels of Fig. 7. Indeed, it provided excellent fits for $\sigma \lesssim 2$ with errors less than 10 per cent and even for larger σ it gives very good accuracy, but it starts to fail catastrophically at very large densities. Lee et al. (2017) tested this approximation for PDF for the Bolshoi simulation. Their results are compatible with ours. However, it seems that Lee et al. (2017) did not pay attention to the situation at large densities, where the GEV becomes unacceptable. It still can be useful for low densities, but the problem is that there is no a priori estimate at what density and σ the GEV fails.

Comparison of the double-exponential model with the *N*-body results has been already presented in Fig. 1. More detailed analysis is shown in Fig. 8. Parameters of the fits for different smoothing scales and redshifts are given in Appendix B. Of all approximations studied in this paper the double exponential model is by far the best. It provides very accurate (few per cent) fits for $\sigma \gtrsim 1$ with densities $\rho \approx 10^{-1}$–10^{5}. With somewhat larger errors it still works down to $\sigma \approx 0.5$. To some degree the success of the approximation is not surprising because it was designed to reproduce the main features of the spherical infall model and the PDF of individual NFW halo profiles that predict the $P(\rho) \propto \rho^{-2}$ trend for large densities. And, of course, the parameters of the approximation – the two power-law slopes – were tuned to produce best fits.

However, it is unexpected that the exponential terms in equation (18) should have the same power-law slopes 0.55 and 1.11 for all densities and smoothing scales, respectively. Indeed, we tried different combinations of the slopes – even with changing values of the slopes for different σ – and did not find them to improve the fits. There is one problem with the constant slopes, though the approximation cannot work for a very small σ where the PDF must become a Gaussian. Equation (18) does not allow a transition to a Gaussian distribution. This is not a serious issue because at $\sigma \lesssim 1$ the spherical infall model provides an adequate approximation for the dark matter density distribution function.

7 CONCLUSIONS

Using a large suite of cosmological *N*-body simulations, we study the shape and the evolution of the dark matter 1-point PDF. Unlike most of other studies, we cover a very large range of smoothing scales $R = 100\,\mathrm{h^{-1}\,kpc} - 20\,\mathrm{h^{-1}\,Mpc}$ and *rms* density fluctuations σ. We find that as σ increases, the PDF becomes a power-law $P \propto \rho^{-2}$, which is truncated with exponential terms on both small and large densities. The same qualitative results were previously found by Bouchet et al. (1991), Bouchet & Hernquist (1992), and Platen (2009). This trend is consistent with the extrapolation of both the spherical infall model and the PDF expected at high densities for the NFW density profile of dark matter haloes.

The PDF weakly depends on redshift (at fixed σ) and on matter density Ω_m. The effect is relatively small (∼5 − 10 per cent) but is clearly observed in simulations. The spherical infall model also has the same trend. This behaviour contradicts analytical approximations such as the lognormal or the GEV that assume that the PDF should depend only on σ.

The basic trend $P \propto \rho^{-2}$ gives us a motivation to construct a new model given in equation (18), which we call double-exponential distribution. It formally has four free parameters of which two can be fixed by requiring that the total volume and mass must be equal to unity. The model works only for $\sigma \gtrsim 1$ and does not allow the transition to a Gaussian distribution as expected for $\sigma \ll 1$. Nevertheless, for $\sigma \gtrsim 1$ the model gives the best performance of all approximations that we tested in this work, with errors of just few per cent for $P(\rho)$ when the PDF changes by 12 orders of magnitude. Parameters of the double-exponential model are provided in Appendix B. The model potentially may be modified by adding few extra parameters to allow for accurate treatment at small σ. We did not try to do it for two reasons: (1) the spherical infall model gives accurate enough treatment for this regime and (2) it is cheap to make adequately accurate *N*-body simulations for $\sigma < 1$ if needed.

The spherical infall model provides accurate predictions for the *N*-body PDF results for low values of $\sigma < 1$, but it becomes less reliable for larger σ, which is expected for this model. The combination of the double-exponential model at large *rms* fluctuation with the spherical infall model in the small *rms* regime yields a remarkably accurate density distribution for all regimes of clustering of the cosmological matter field.

We also tested different analytical approximations. The often-used lognormal distribution, as was shown before, does not make good fits and needs significant modifications before it can provide accurate results. It is pretty much useless for large densities because it does not provide a path to explain the main trend at large densities, i.e. the power-law trend $P \propto \rho^{-2}$. It clearly has an advantage of being simple, and it does not fail catastrophically as shown in Fig. 6. Our results for $\sigma < 1$ demonstrate that the lognormal approximation always made significantly worse fits as compared with the spherical infall model predictions. In addition, the lognormal

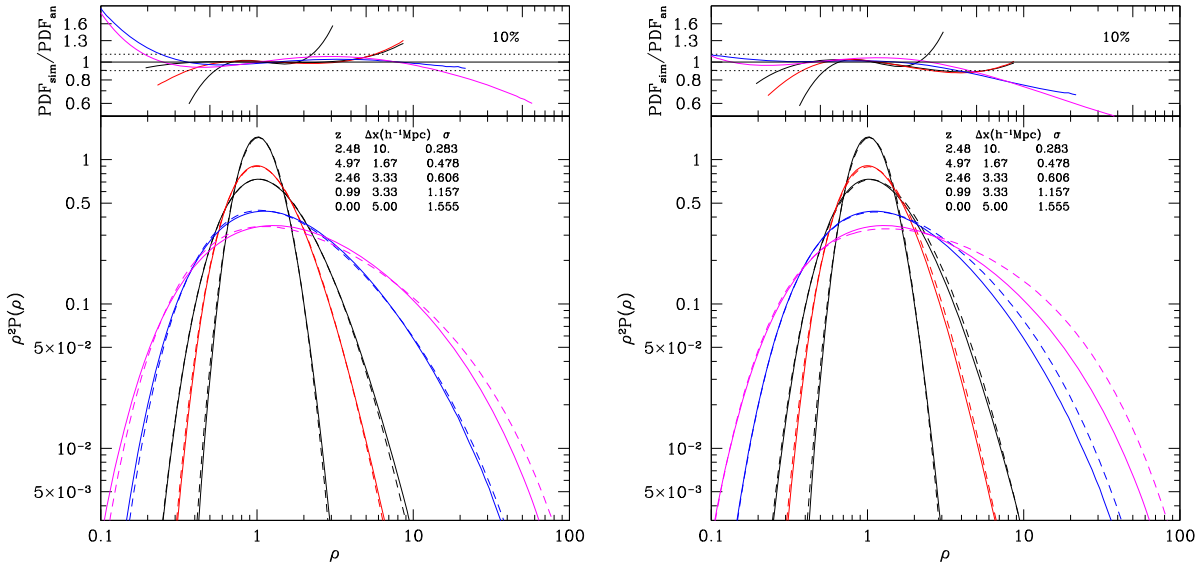

Figure 5. Accuracy of spherical infall model in the regime of $\sigma \lesssim 1$. The left-hand panel presents results for the approximation equation (12). The right-hand panel is for equation (13). We select different redshifts and different filtering scales. Full curves in the bottom panels show results of simulations, while the dashed curves are for the analytical approximation. The top panels present relative errors of the approximations.

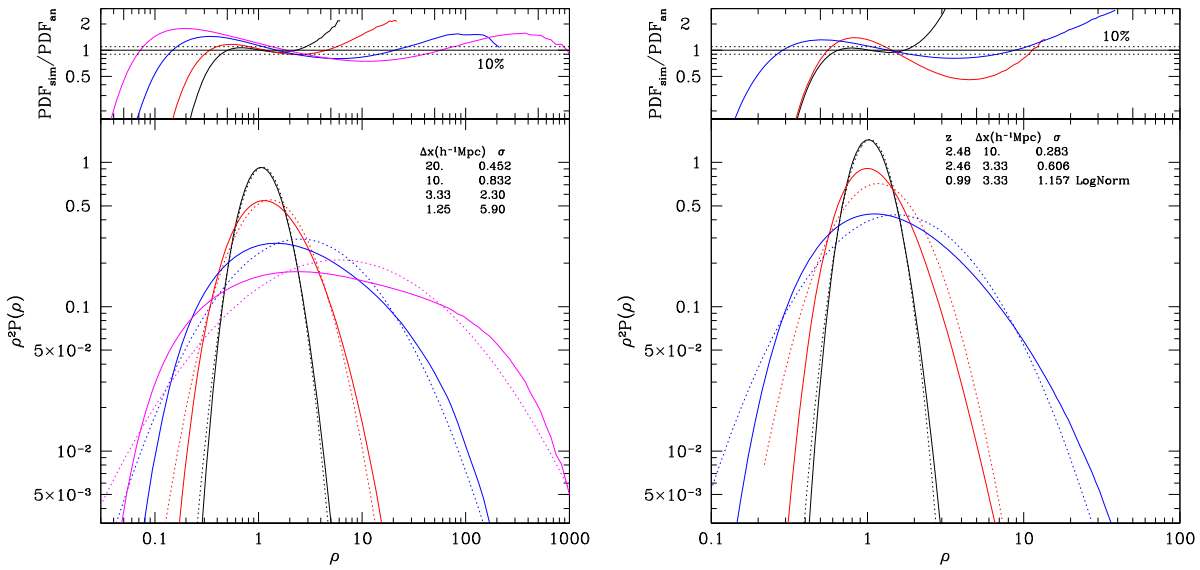

Figure 6. Accuracy of the lognormal distribution. *Left:* Results for different smoothing scales at $z = 0$. *Right:* Results for different redshifts. The lognormal distribution provides accurate fits (less than 10 per cent errors) only for a very limited range of densities and σ. Comparison with the predictions of the spherical infall model in Fig. 5 shows that the lognormal fits for $\sigma < 1$ always give errors 2–3 times larger than the spherical infall model. To make things worse, the lognormal distribution has a wrong shape. It predicts wrong position of the maximum; slopes on both ends of the PDF are also incorrect. The only advantages of the lognormal fits are that it is very simple and that it never fails catastrophically.

approximation cannot accomodate the dependence of the PDF on redshift and $\Omega_{\rm m}$, while the spherical infall model nicely reproduces the effect.

The GEV approximation scores much better than the lognormal distribution. Even for a large $\sigma \approx 10$ it gives remarkably accurate results for densities up to $\rho \approx 10^3$. However, the approximation fails catastrophically at larger densities, and there is no obvious way to predict at what density it still works or fails.

While it is useful and insightful to have analytical models for the PDF, one does not really need them if ∼1 per cent accuracy is a requirement and $\sigma \gtrsim 0.5$. Then there is no alternative to *N*-body simulations: one can get very accurate, fast, and cheap results. The cost of one GLAM A0.5 simulation is just 3.5 h of wall clock on a modest data server (Klypin & Prada 2018). For the B0.5 run it is 10 h. In this paper we analyse thousands of realizations, but this was to prove that one does not need those to make accurate PDF: just a few realizations are enough. However, one needs to be careful about numerical effects when using these fast Particle-Mesh simulations. Klypin & Prada (2018) provide detailed description of constraints for the simulations, and Appendix A in this paper gives prescriptions on how to use the *N*-body simulations to reach ∼1 per cent accuracy in the PDF.

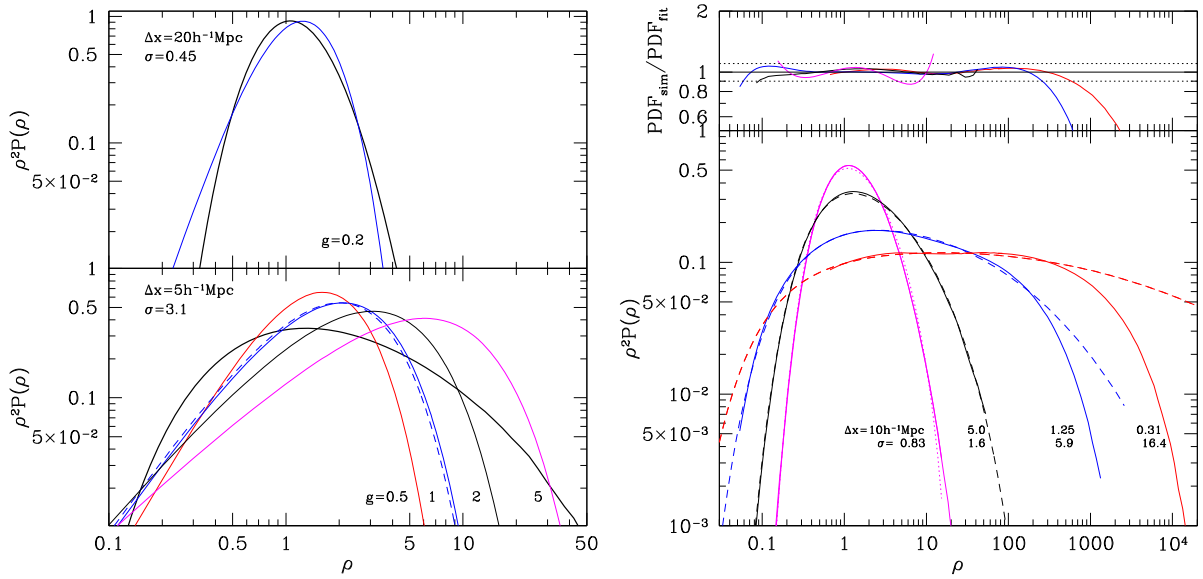

Figure 7. Accuracy of different analytical approximations. PDF of *N*-body simulations are taken at $z = 0$. *Left:* Negative BiModal (NBN) fits (equation 7) are shown as thin cuves labeled by values of g parameter used to make the fits. *N*-body results are the thick lines. NBN provides a reasonable fit (∼10 per cent accuracy for $\rho = 0.4$–3) for large cell sizes corresponding to low *rms* fluctuations $\sigma < 0.5$. The bottom panel shows an example of NBN fits for large σ. It does not provide a good fit, regardless of what value of g is used. *Right:* Results of fitting with the GEV distribution. Full cuves in the bottom panel show *N*-body results; GEV fits are the dashed curves. GEV provides much more accurate fits as compared with the lognormal or NBN approximations. However, it fails at very large densities for small cell sizes.

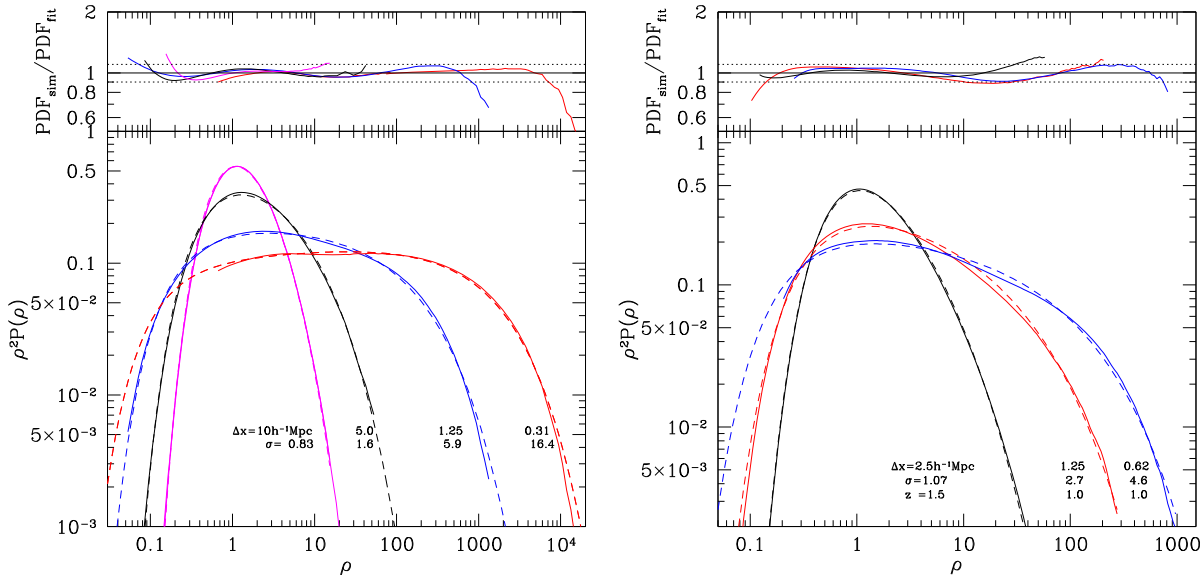

Figure 8. Accuracy of the double-exponential model given by equation (18). The left-hand panels show results for $z = 0$. Results of fits for different redshifts are presented in the right-hand panel.

ACKNOWLEDGEMENTS

A.K. acknowledges support of the Fulbright Foundation and support of the Instituto de Fisica Teorica, CSIC, Madrid, Spain. F.P. acknowledges support from the Spanish MINECO grant AYA2014-60641-C2-1-P. FDA acknowledges financial support from 'la Caixa'-Severo Ochoa doctoral fellowship. We thank Johan Comparat (IFT, Madrid), Claudia Scoccola (IAC, Tenerife), and Sergio Rodriguez-Torres (UAM, Madrid) for comments and fruitful discussions. The PPM-GLAM simulations have been performed on the FinisTarrae II supercomputer at CESGA in Galicia, supported by the Xunta de Galicia, CISIC, MINECO, and EU-ERDF.

REFERENCES

Bel J. et al., 2016, A&A, 588, A51
Bernardeau F., 1994, A&A, 291, 697
Betancort-Rijo J., 1991, MNRAS, 251, 399
Betancort-Rijo J., 2000, J. Stat. Phys., 98, 3
Betancort-Rijo J., López-Corredoira M., 2002, ApJ, 566, 623
Bouchet F. R., Hernquist L., 1992, ApJ, 400, 25
Bouchet F. R., Schaeffer R., Davis M., 1991, ApJ, 383, 19
Carron J., Wolk M., Szapudi I., 2015, MNRAS, 453, 450
Chuang C.-H. et al., 2015, MNRAS, 452, 686
Clerkin L. et al., 2017, MNRAS, 466, 1444
Coles P., Jones B., 1991, MNRAS, 248, 1

Colombi S., 1994, ApJ, 435, 536
Efstathiou G., Kaiser N., Saunders W., Lawrence A., Rowan-Robinson M., Ellis R. S., Frenk C. S., 1990, MNRAS, 247, 10P
Gaztañaga E., Fosalba P., Elizalde E., 2000, ApJ, 539, 522
Hamilton A. J. S., 1985, ApJ, 292, L35
Hubble E., 1934, ApJ, 79, 8
Hurtado-Gil L., Martínez V. J., Arnalte-Mur P., Pons-Bordería M.-J., Pareja-Flores C., Paredes S., 2017, A&A, 601, A40
Kitaura F.-S. et al., 2016, MNRAS, 456, 4156
Klypin A., Prada F., 2018, MNRAS, 478, 4602
Klypin A., Prada F., Comparat J., 2017, preprint (arXiv:1711.01453)
Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Kofman L., Bertschinger E., Gelb J. M., Nusser A., Dekel A., 1994, ApJ, 420, 44
Kravtsov A. V., Klypin A. A., Khokhlov A. M., 1997, ApJS, 111, 73
Lam T. Y., Sheth R. K., 2008a, MNRAS, 386, 407
Lam T. Y., Sheth R. K., 2008b, MNRAS, 389, 1249
Lee C. T., Primack J. R., Behroozi P., Rodríguez-Puebla A., Hellinger D., Dekel A., 2017, MNRAS, 466, 3834
Manera M. et al., 2013, MNRAS, 428, 1036
Neyrinck M. C., 2016, MNRAS, 455, L11
Ohta Y., Kayo I., Taruya A., 2003, ApJ, 589, 1
Pandey B., White S. D. M., Springel V., Angulo R. E., 2013, MNRAS, 435, 2968
Peebles P. J. E., 1980, The Large-scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ
Platen E., 2009, PhD thesis, Kapteyn Institute, Groningen, the Nertherlands
Repp A., Szapudi I., 2018, MNRAS, 473, 3598
Sheth R. K., 1998, MNRAS, 300, 1057
Shin J., Kim J., Pichon C., Jeong D., Park C., 2017, ApJ, 843, 73
Soneira R. M., Peebles P. J. E., 1978, AJ, 83, 845
Springel V., 2005, MNRAS, 364, 1105
Takahashi R., Oguri M., Sato M., Hamana T., 2011, ApJ, 742, 15
Taruya A., Takada M., Hamana T., Kayo I., Futamase T., 2002, ApJ, 571, 638
Tassev S., Zaldarriaga M., Eisenstein D. J., 2013, J. Cosmol. Astropart. Phys., 6, 036
Wild V. et al., 2005, MNRAS, 356, 247

APPENDIX A: NUMERICAL EFFECTS: FINITE-VOLUME VARIANCE, MASS, AND FORCE RESOLUTION

Numerical effects may play an important role for the estimates of the density distribution function. We start with the analysis of the effects of the variance due to the fine volume of the simulations. We use the GLAM simulations A0.5 with small cell $\Delta x = 0.83\,h^{-1}$ Mpc and A1.5 with larger cell $\Delta x = 0.5\,h^{-1}$ Mpc to estimate the level of fluctuations in different realizations. Bottom panels in Fig. A1 show the average values of $\rho^2 P(\rho)$ and the statistical fluctuations of *a single realization*. As expected, the fluctuations due to the fine-volume simulation variance are larger for very large densities and become very small for $\rho \approx 0.5$–10.

The middle panels present relative fluctuations $\Delta P/\langle P\rangle$, where $\langle P\rangle$ is the average PDF over an ensamble of realizations and ΔP is the *rms* deviation. The vertical dotted lines at large ρ show the density bin with 100 cells in a single realization. At this density the level of statistical fluctuations $\Delta P/P$ is about 0.1, which is consistant with the expected shot noise. We clarify this situation by plotting in the top panels the relative fluctuations scaled with $N^{1/2}$, where N is the number of cells of given density in a bin. Indeed, the fluctuations are defined by the number of cells in a bin for large densities.

The situation is different at small densities $\rho \lesssim 1$, where the fluctuations become substantially stronger than the Gaussian $\Delta P/P = N^{-1/2}$. This is likely related with the increasing noise in the density field due to too few particles per cell n. The vertical dotted lines on low densities show the bin at which $n = 10$.

In spite of being strongly non-Poissonian at small densities, the errors are still very small. For example, at $\rho = 0.1$ for the A0.5 simulations the errors are just ∼2 per cent for a single realization (right-hand panels). Note that the fluctuations plotted in

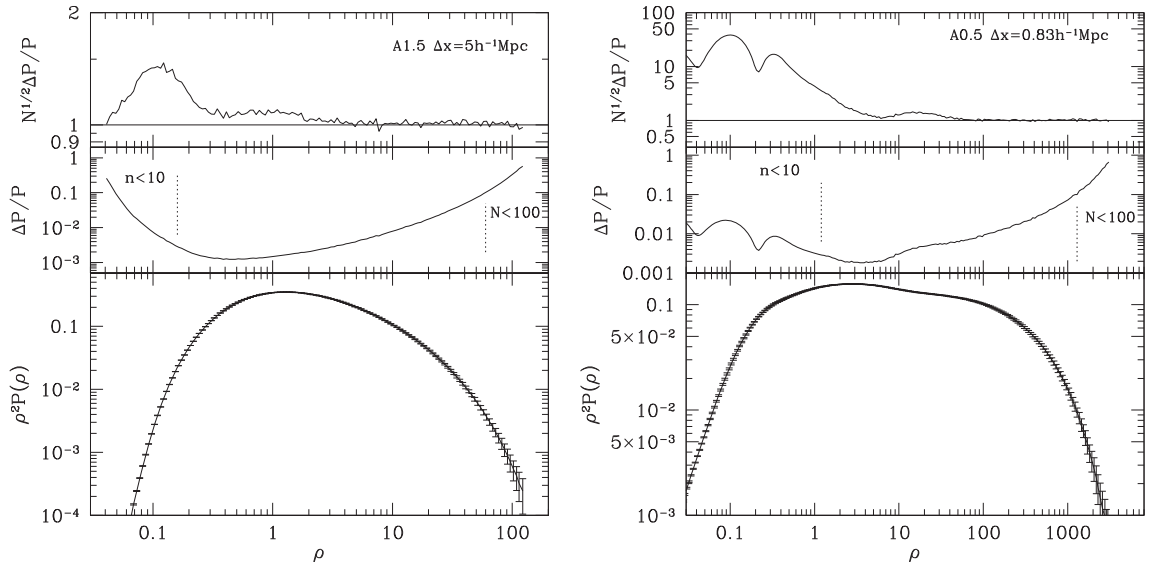

Figure A1. Statistical errors of the density distribution function $P(\rho)$ due to the finite-volume simulation variance. Results are shown at $z = 0$ for the GLAM A1.5 (left) and A0.5 (right) simulations with different cell sizes. The middle panels present the *rms* fluctuations $\Delta P/P$ of a single realization. The errors of the mean $P(\rho)$ are significantly lower because these simulations have a very large number of realizations. The top panels show the *rms* deviations scaled with $\sqrt{N}$, where N is the number of cells in a bin of $P(\rho)$. For densities that are probed with a large number of particles, the *rms* fluctuations are defined by the number of cells per bin of ρ. The fluctuations are substantially non-Gaussian for bins with a number of particles n per bin less than 10.

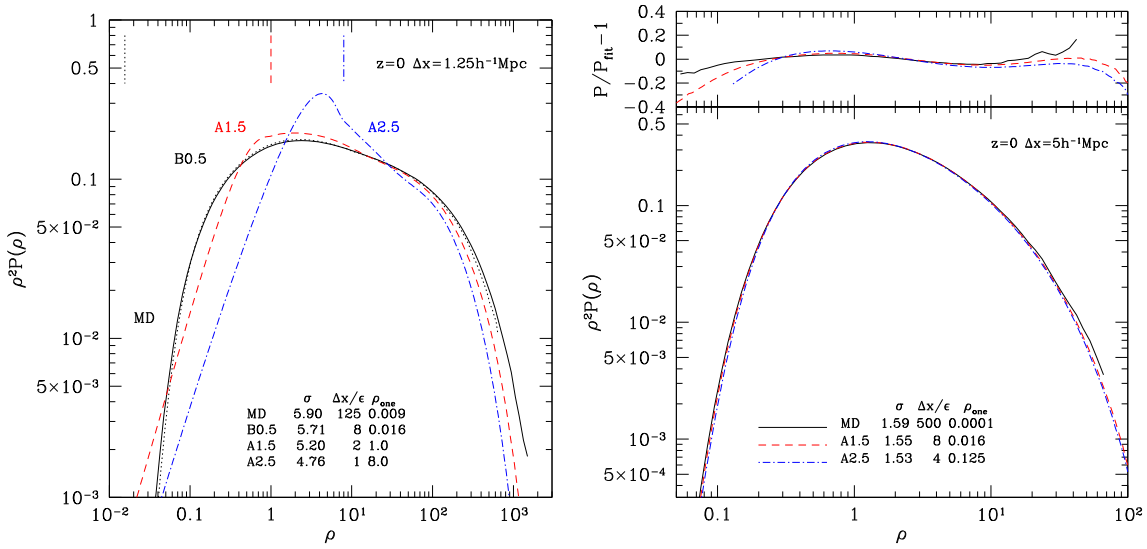

Figure A2. Comparison of density distribution functions estimated in simulations with different box sizes and resolutions. *Left:* Example of numerical convergence for the cell size $\Delta x = 1.25$ h^{-1} Mpc at $z = 0$. Insufficient force resolution reduces the amplitude of $P(\rho)$ at large densities. As the force resolution increases, the results converge. In order to have errors less than few per cent the ratio of the cell size Δx to the force resolution ε should be larger than $\sim$8. At small densities the noise due to the finite number-density of particles can produce large errors. The vertical lines at the top of the panel show the density $\rho_{\rm one}$ that a single particle produces when placed at the center of the cell. The bump in the A2.5 curve at $\rho \lesssim \rho_{\rm one}$ is due to large discreteness effects: for this cell size the A2.5 simulations did not have enough particles. *Right:* The same as for the left-hand panel, but now for better force resolution $\Delta x/\varepsilon$ and for larger number-densities of particles. To facilitate detailed comparisons, the top panel shows deviations of $P(\rho)$ from the same analytical fit. Results indicate that the box size does not affect the density distribution. However, the force resolution has a tendency to reduce the amplitude of $P(\rho)$ at very large density ρ.

Fig. A1 provide average deviations of a single realization from the ensamble average. Errors of the average $P(\rho)$ are significantly smaller. For example, for the A1.5 simulations (left-hand panels) the error of the mean at $\rho = 60, N = 100$ is just $\sim$0.1 per cent.

In summary, our results are mostly dominated by systematics, not by the variance. When dealing with individual simulations such as MDPL1 or BolshoiP, we use only bins with more than $N > 100$ per bin. For large sets of simulations A0.5, 1.5, 2.5 we accept bins with more than 10 cells.

In order to evaluate other possible numerical effects, we select two filtering scales $\Delta x = 1.25$ h^{-1} Mpc and $\Delta x = 5$ h^{-1} Mpc and analyse $P(\rho)$ at $z = 0$ obtained from different simulations. The two filtering scales probe different dynamical regimes. For $\Delta x = 1.25$ h^{-1} Mpc the *rms* density fluctuation is large $\sigma \approx 5$. So, we are testing very non-linear regime with densities up to 1000. The larger filtering scale $\Delta x = 5$ h^{-1} Mpc probes more modest fluctuations with $\sigma \approx 1.5$ and densities $\rho = 0.1$–100.

By comparing these results, we test the effects of the finite box size (ranging from 500 h^{-1} Mpc to 2500 h^{-1} Mpc), the effects of force resolution $\Delta x/\varepsilon$ (ranging from 1 to 500) and the discreteness effects associated with the finite number of dark matter particles. The later can be characterized by the density $\rho_{\rm one}$ produced by a single particle placed at the node of a cell:

$$\rho_{\rm one} = \left(\frac{L}{\Delta x N_p}\right)^3, \qquad \text{(A1)}$$

where $N_{\rm p}^3$ is the number of particles and L is the size of the simulation box. Vertical lines in the left-hand panel of Fig. A2 show $\rho_{\rm one}$ for different simulations. Values of different parameters are also given in Fig. A2.

We use the left-hand panel in Fig. A2 to demonstrate two numerical effects. At large densities $\rho \gtrsim 50$ the discreteness effects are small even for the A2.5 simulations and the results are dominated by the force resolution. Here we find a trend that is expected for simulations with low force resolution: the PDF increases with increasing force resolution. However, there is little difference between simulations when the resolution becomes sufficiently high: the PDF for B0.5 simulations with eight resolution elements across one cell ($\Delta x/\varepsilon = 8$) is nearly the same as for the much high-resolution simulation MDPL1 with $\Delta x/\varepsilon = 125$. This is a signature that the results have converged.

The discreteness of density assignment becomes an issue at small densities as is clearly seen in Fig. A2 for $\rho \lesssim 10$. The A2.5 simulations provide a good example on how the particle noise affects the PDF. There is a large bump in $\rho^2 P(\rho)$ at densities slightly below $\rho_{\rm one}$. At much smaller densities $\rho \approx 0.1\rho_{\rm one}$ the PDF falls much below the real one. The effects of the particle noise extend to densities above $\rho_{\rm one}$ but quickly die out beyond $\rho \approx 10\rho_{\rm one}$.

The right-hand panel in Fig. A2 shows better convergence of $P(\rho)$ because we select simulations that have better force resolution and smaller level of the particle noise. In order to see the differences more clearly, we make a fit to the data, and on the top panel plot we show only the deviations from the fit. The A2.5 data still fall below the more accurate MDPL1 results at both small densities (effects of the particle noise) and large densities (effects of the force resolution). The A1.5 data show much smaller errors.

Using the results presented in Fig. A2 and similar results of the comparison between different simulations (e.g. comparison of BolshoiP and MDPL1), we find limits on numerical parameters that should be satisfied to produce a PDF $P(\rho)$ with errors less than a few per cent:

(a) the filtering scale Δx must be resolved with not less than 8 force resolution elements: $\Delta x > 8\varepsilon$, and (b) the number of particles per filtering cell should be not less than 10–20: $\rho > (10\text{–}20)\rho_{\rm one}$.

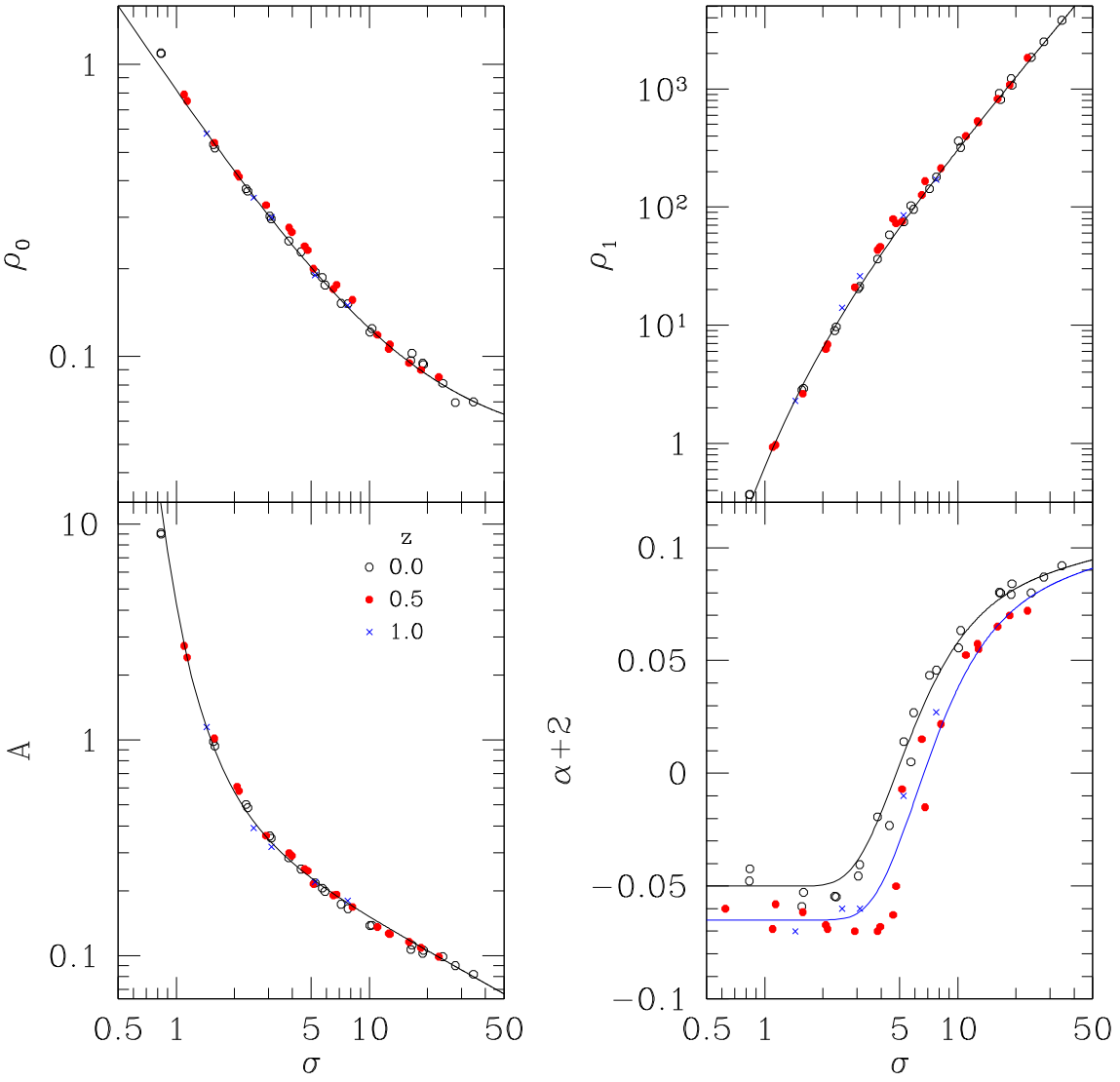

Figure A3. Dependence of the parameters of the double-exponential approximation, equation (18), on the *rms* density fluctuation σ. Different symbols represent results of fitting the PDFs for different simulations at redshifts $z = 0$–1. The only parameter that depends on redshift is the slope α. Curves in the plots are approximations B1-5 in Appendix B.

APPENDIX B: PARAMETERS OF THE DOUBLE-EXPONENTIAL MODEL FOR THE DARK MATTER DENSITY DISTRIBUTION FUNCTION

Tables B1 and B2 list the parameters for the double-exponential model given in equation (18) for our set of simulations, filtering scales and redshifts discussed in the main text. The simulation data and PDF tables are available at *www.skiesanduniverses.org* (Klypin et al. 2017)

Fig. A3 shows the dependence of the double-exponential approximation parameters on the *rms* fluctuation σ for redshifts $z = 0, 0.5, 1$. We find that with the exception of the slope α all other three parameters do not show any signs of evolution with redshift. The slope slightly evolves with redshift with the tendency to produce smaller values of α at higher redshifts at constant σ. Yet, the effect is small and is limited only to $z < 0.5$.

For practical reasons it is convenient to have fitting expressions for the dependence of those four parameters as a function of σ and redshift. Fig. A3 clearly indicates that for a given redshift there are tight relations of those parameters with σ. We tried different analytical expressions and find that the following relations provide accurate fits for $\sigma > 0.7$:

$$A = \frac{0.47}{\sqrt{\sigma}} \exp\left(\frac{2.2}{\sigma^2}\right) \tag{B1}$$

$$\rho_0 = 0.048 + \left(\frac{0.77}{\sigma}\right) \tag{B2}$$

$$\rho_1 = 4.7\sigma^{1.9} \exp\left(-\frac{2}{\sigma}\right) \tag{B3}$$

$$\alpha + 2 = -\frac{0.05}{f(z)}\left[1 - 2.4\sigma^{0.05} \exp\left(-\left[\frac{4.7}{f(z)\sigma}\right]^2\right)\right] \tag{B4}$$

$$f(z) = 0.75 + 0.25/(1+z)^5 \tag{B5}$$

Table B1. Best-fit parameters for the double-exponential model, equation (18), for redshift $z = 0$. Different columns give (1) name of the simulation, (2) cell size in h^{-1} Mpc, (3) *rms* density fluctuation, (4–7) parameters for the double-exponential model, (8) the number of cells in 1D, (9) relative force resolution – the number of force resolution elements per each density assignment cell.

Simulation (1)	Cell Size (2)	σ (3)	A (4)	$\alpha + 2$ (5)	ρ_0 (6)	ρ_1 (7)	$N_{\rm cell}$ (8)	$\Delta x/\varepsilon$ (9)
BolshoiP	0.081	34.62	8.20E-02	0.092	0.070	3800.0	3100	80
BolshoiP	0.125	27.92	9.00E-02	0.087	0.069	2500.0	2000	125
MDPL1	0.167	24.02	9.90E-02	0.080	0.081	1850.0	6000	17
MDPL1	0.250	19.07	1.06E-01	0.084	0.094	1070.9	4000	25
BolshoiP	0.250	18.89	1.03E-01	0.079	0.095	1223.0	1000	250
BolshoiP	0.313	16.42	1.07E-01	0.080	0.097	916.00	800	313
MDPL1	0.313	16.63	1.12E-01	0.080	0.102	811.29	3200	31
MDPL1	0.500	12.15	1.30E-01	0.064	0.125	451.21	2000	50
BolshoiP	0.625	10.08	1.38E-01	0.056	0.121	362.30	400	616
MDPL1	0.625	10.34	1.38E-01	0.063	0.125	319.20	1600	62
A0.5	0.833	7.75	1.65E-01	0.045	0.151	180.09	600	4
MDPL1	1.000	7.14	1.73E-01	0.044	0.152	143.08	1000	100
MDPL1	1.250	5.90	1.98E-01	0.027	0.175	95.484	800	125
B0.5	1.250	5.71	2.05E-01	0.005	0.186	103.01	400	8
MDPL1	1.429	5.24	2.17E-01	0.014	0.195	74.855	700	143
A0.5	1.667	4.42	2.51E-01	− 0.023	0.228	58.198	300	8
MDPL1	2.000	3.83	2.84E-01	− 0.019	0.248	36.300	500	200
MDPL1	2.500	3.10	3.50E-01	− 0.040	0.295	21.230	400	250
B0.5	2.500	3.04	3.58E-01	− 0.045	0.302	20.465	200	16
A0.5	3.333	2.29	5.02E-01	− 0.054	0.375	9.010	150	16
MDPL1	3.333	2.34	4.85E-01	− 0.055	0.368	9.694	300	333
A1.5	5.000	1.55	9.81E-01	− 0.059	0.532	2.829	300	8
MDPL1	5.000	1.58	8.90E-01	− 0.050	0.495	3.100	200	500
A1.5	10.000	0.83	8.97E + 00	− 0.042	1.090	0.368	150	16
A2.5	10.000	0.83	9.12E + 00	− 0.047	1.097	0.368	250	8
A2.5	20.000	0.45	2.19E + 03	− 0.034	2.508	0.052	125	16

Table B2. Best-fitting parameters for the double-exponential model, equation (18), for redshifts $z \approx 0.5$–1. Different columns give (1) name of the simulation, (2) cell size in h^{-1} Mpc, (3) *rms* density fluctuation, (4–7) parameters for the double-exponential model, (8) number of cells in 1D, (9) relative force resolution – the number of force resolution elements per each density assignment cell, (10) redshift.

Simulation (1)	Cell Size (2)	σ (3)	A (4)	$\alpha + 2$ (5)	ρ_0 (6)	ρ_1 (7)	$N_{\rm cell}$ (8)	$\Delta x/\varepsilon$ (9)	redshift (10)
MDPL1	0.167	16.050	1.160E-01	0.065	0.10	820.00	6000	16.7	0.492
MDPL1	0.250	12.780	1.260E-01	0.055	0.11	520.40	4000	25.0	0.492
BolshoiP	0.313	11.000	1.361E-01	0.052	0.12	399.62	800	300	0.500
MDPL1	0.500	8.177	1.683E-01	0.022	0.16	213.10	2000	50.0	0.492
B0.5	0.625	6.510	1.907E-01	0.015	0.17	126.33	800	4	0.510
BolshoiP	0.625	6.765	1.920E-01	− 0.015	0.18	165.64	400	625	0.500
A0.5	0.833	5.136	2.151E-01	− 0.007	0.20	75.60	600	4	0.510
MDPL1	1.000	4.793	2.476E-01	− 0.050	0.23	72.81	1000	100	0.492
MDPL1	1.250	3.961	2.909E-01	− 0.068	0.27	46.15	800	125	0.492
B0.5	1.250	3.837	2.993E-01	− 0.081	0.28	45.38	400	8	0.492
A0.5	1.667	2.923	3.882E-01	− 0.082	0.33	19.24	300	8	0.510
B0.5	2.500	2.068	6.078E-01	− 0.067	0.42	6.282	200	16	0.492
MDPL1	2.500	2.112	5.801E-01	− 0.069	0.41	6.892	400	250	0.492
A0.5	3.333	1.575	1.015E + 00	− 0.062	0.55	2.635	150	16	0.510
MDPL1	5.000	1.138	2.413E + 00	− 0.058	0.75	0.974	200	500	0.492
A2.5	5.000	1.099	2.731E + 00	− 0.069	0.79	0.930	500	4	0.497
A2.5	10.00	0.625	7.163E + 01	− 0.053	1.63	0.141	250	8	0.497
A2.5	20.00	0.347	1.409E + 05	− 0.026	3.57	0.024	125	16	0.497
B0.5	1.25	2.671	4.516E-01	− 0.060	0.3612	11.71	400	8	0.99
A0.5	1.667	2.066	6.462E-01	− 0.062	0.4374	5.37	300	8	0.99
B0.5	2.50	1.475	1.248E + 00	− 0.070	0.5876	2.10	200	16	0.99
A0.5	3.333	1.157	2.484E + 00	− 0.056	0.7563	0.94	150	16	0.99

This paper has been typeset from a TEX/LATEX file prepared by the author.