PHYSICAL REVIEW D **85,** 063508 (2012)

Primordial non-Gaussianity and extreme-value statistics of galaxy clusters

Sirichai Chongchitnan* and Joseph Silk

Department of Physics, University of Oxford, and Beecroft Institute for Particle Astrophysics and Cosmology, Denys Wilkinson Building, 1 Keble Road, Oxford, OX1 3RH, United Kingdom

(Received 27 July 2011; published 9 March 2012)

What is the size of the *most* massive object one expects to find in a survey of a given volume? In this paper, we present a solution to this problem using extreme-value statistics, taking into account primordial non-Gaussianity and its effects on the abundance and the clustering of rare objects. We calculate the probability density function (PDF) of extreme-mass clusters in a survey volume, and show how primordial non-Gaussianity shifts the peak of this PDF. We also study the sensitivity of the extreme-value PDFs to changes in the mass functions, survey volume, redshift coverage, and the normalization of the matter power spectrum, σ_8. For ''local'' non-Gaussianity parametrized by $f_{\rm NL}$, our correction for the extreme-value PDF due to the bias is important when $f_{\rm NL} \gtrsim 100$, and becomes more significant for wider and deeper surveys. Applying our formalism to the massive high-redshift cluster XMMUJ0044.0-2-33, we find that its existence is consistent with $f_{\rm NL} = 0$, although the conclusion is sensitive to the assumed values of the survey area and σ_8. We also discuss the convergence of the extreme-value distribution to one of the three possible asymptotic forms, and argue that the convergence is insensitive to the presence of non-Gaussianity.

DOI: 10.1103/PhysRevD.85.063508 PACS numbers: 98.80.Cq

I. INTRODUCTION

The statistics of the primordial seeds that grew into the observed large-scale structures holds a wealth of information about the physics of the primordial Universe. In the simplest models of inflation, the primordial density fluctuations obey an almost Gaussian statistics (see [1] for a review). Tiny deviations from Gaussianity may be quantified, amongst other ways,[1] by the ''local'' non-Gaussianity parameter, $f_{\rm NL}$, defined via the expansion of the nonlinear Newtonian potential

$$\Phi = \phi + f_{\rm NL}(\phi^2 - \langle\phi^2\rangle) + \ldots, \tag{1}$$

where ϕ is a Gaussian random field. This form of non-Gaussianity arises in simple models of single and multi-field inflation [2–4] as well as some curvaton models [5–7]. Observational constraints on $f_{\rm NL}$ from the cosmic microwave background (CMB) anisotropies are currently consistent with $f_{\rm NL} = 32 \pm 42$ (2σ) [8]. However, if $f_{\rm NL}$ is in fact much smaller, its effects on the CMB would be difficult to extract and distinguished from non-Gaussianity arising from secondary sources such as gravitational lensing and instrumental noise [9,10].

The statistics of large-scale structures offers a complementary probe of non-Gaussianity on much smaller scales than the CMB. Indeed, it is possible that $f_{\rm NL}$ measured on Gpc scales may be quite different from that measured on Mpc scales. In wave number space, this translates to a possible k-dependence of $f_{\rm NL}$, which have been hinted at by the numerous observations of massive high-redshift clusters [11–15]. These massive clusters exist, according to some, in greater abundances than expected from Gaussian statistics. Some authors have concluded that the level of non-Gaussianity on the Mpc scale required to explain the existence of certain rare clusters is $f_{\rm NL} =$ a few $\times 10^2$ [16,17]. In contrast, some have argued that these claims are based on a misinterpretation of data, and that the occurrences of these rare objects are in fact consistent with Gaussian statistics [18–20].

In this work, we offer our opinion on this debate by presenting an approach to calculating the probability of observing rare objects based on extreme-value statistics. We begin by asking: what is the probability distribution of the *most* massive clusters found within a given volume at a given redshift range? Our technique relies on a basic application of the so-called void probability distribution introduced by White [21]. This approach was successfully used to study the abundances of massive clusters given a Gaussian initial condition in [22,23]. In this work, we extend the groundwork laid by these authors to study the effect of $f_{\rm NL}$ on the distribution of extreme-mass objects. For other cosmological applications of extreme-value theory, see [24–31].

Previous approaches to extreme-value statistics of clusters have so far either neglected the clustering, or bias, of galaxy clusters [31,32], or considered it in the Gaussian case [22,29]. In this work, however, we have included the effects of the bias in the presence of non-Gaussianity. Whilst Davis *et al.* [22] have previously reported that the effects of the bias on the extreme-value distribution are small in the Gaussian case, it remains to be shown if this

*siri@astro.ox.ac.uk

[1] In general, a non-Gaussian probability density function (PDF) can have divergent moments (e.g. the Cauchy distribution). In this work we assume that the primordial density distribution has finite moments up to third order (i.e. finite skewness).

also holds in the presence of non-Gaussianity, which can introduce a strong scale dependence in the bias [33]. We investigate this problem in this work using the formalism of Valageas [34,35], who showed how the bias can be calculated in real space for a given $f_{\rm NL}$. As we shall see later, the contribution from the bias can indeed be significant if $f_{\rm NL}$ is sufficiently large.

II. THE PRIMORDIAL DENSITY FLUCTUATIONS

In this section, we introduce the parameters needed to describe the primordial density fluctuations statistically. Some of our present conventions (such as the Fourier transform and the moments of the density fluctuations) slightly differ from our earlier work [36,37]. In particular, smoothing by a window function will be kept explicit, in contrast with other work in which overdensities are defined to be implicitly smoothed.

Let ρ_c, ρ_b, ρ_r, ρ_Λ be the time-dependent energy densities of cold dark matter, baryons, radiation, and dark energy. Let $\rho_m = \rho_c + \rho_b$. We define the density parameter for species i as

$$\Omega_i \equiv \frac{\rho_i(z=0)}{\rho_{\rm crit}}, \tag{2}$$

where $\rho_{\rm crit}$ is the critical density defined as $\rho_{\rm crit} \equiv 3H_0^2/8\pi G$. The Hubble constant, H_0, is parametrized by h via the usual formula $H_0 \equiv 100h\ {\rm km\,s^{-1}\,Mpc^{-1}}$. Results from a range of astrophysical observations are consistent with $h \simeq 0.7$, $\Omega_c \simeq 0.23$, $\Omega_b \simeq 0.046$, $\Omega_r \simeq 8.6\times10^{-5}$, and $\Omega_\Lambda = 1 - \Omega_m - \Omega_r$ (see e.g. [8,38]).

The density fluctuation field, δ, is defined at redshift z as

$$\delta(\mathbf{x}, z) \equiv \frac{\rho_m(\mathbf{x}, z) - \langle\rho_m(z)\rangle}{\langle\rho_m(z)\rangle}, \tag{3}$$

where $\langle\rho_m\rangle$ is the mean matter energy density. To make the notation less cumbersome, we shall write $\delta(\mathbf{x})$ to mean $\delta(\mathbf{x}, z)$. The Fourier decomposition of $\delta(\mathbf{x})$ is given by

$$\delta(\mathbf{x}) = \int d\mathbf{k}\,\delta(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{4}$$

The gravitational Newtonian potential, Φ, is related to the density fluctuation by the cosmological Poisson equation

$$\delta(\mathbf{k}) = \mathcal{A}(k, z)\Phi(\mathbf{k}), \tag{5}$$

$$\mathcal{A}(k, z) \equiv \frac{2}{3\Omega_m}\left(\frac{k}{H_0}\right)^2 T(k_{\rm EH})D(z), \tag{6}$$

where T is the transfer function and D is the linear growth factor calculated using the fitting formula of [39,40] with $D(0) \approx 0.76$. We follow the approach outlined in [41] and use the transfer function

$$T(k) = \frac{\ln[1 + (0.124k)^2]}{(0.124k)^2} \times\left[\frac{1 + (1.257k)^2 + (0.4452k)^4 + (0.2197k)^6}{1 + (1.606k)^2 + (0.8568k)^4 + (0.3927k)^6}\right]^{1/2}. \tag{7}$$

In addition, we also incorporate the baryonic correction of Eisenstein and Hu [42], whereby the transfer function is evaluated at

$$k_{\rm EH} = \frac{k\Omega_r^{1/2}}{H_0\Omega_m}\left[\alpha + \frac{1-\alpha}{1 + (0.43ks)^4}\right]^{-1}, \tag{8}$$

with

$$\alpha = 1 - 0.328\ln(431\Omega_m h^2)\frac{\Omega_b}{\Omega_m} + 0.38\ln(22.3\Omega_m h^2)\left(\frac{\Omega_b}{\Omega_m}\right)^2,$$

and

$$s = \frac{44.5\ln(9.83/\Omega_m h^2)}{\sqrt{1 + 10(\Omega_b h^2)^{3/4}}}\ {\rm Mpc}.$$

The power spectrum, $P(k)$, can be defined via the two-point correlation function in Fourier space as

$$\langle\delta(\mathbf{k_1}), \delta(\mathbf{k_2})\rangle = \delta_D(\mathbf{k_1} + \mathbf{k_2})P(k), \tag{9}$$

where δ_D is the three-dimensional Dirac delta function. In linear perturbation theory, it is usually assumed that inflation laid down a primordial spectrum of the form $P(k) \propto k^{n_s}$, where n_s is the scalar spectral index (assumed to be 0.96 in this work).

The variance of linear density fluctuations smoothed on scale R is given by

$$\sigma_R^2 = 4\pi\int_0^\infty \frac{dk}{k}W^2(kR)\mathcal{P}(k), \tag{10}$$

where $\mathcal{P}(k) \equiv k^3P(k) \propto [\mathcal{A}(k, z)]^2(k/H_0)^{n_s-1}$.

We choose W to be the spherical top-hat function of radius R. In Fourier space, we have

$$W(kR) = 3\left[\frac{\sin(kR)}{(kR)^3} - \frac{\cos(kR)}{(kR)^2}\right]. \tag{11}$$

The mass, M, of matter enclosed by a top-hat window of radius R is given by

$$M \equiv \tfrac{4}{3}\pi R^3\rho_m \approx 1.16\times10^{12}\Omega_m\left(\frac{R}{h^{-1}\ {\rm Mpc}}\right)^3 h^{-1}M_\odot. \tag{12}$$

With the above relation, the smoothed variance, σ_R, can be equivalently expressed as σ_M. Finally, the normalization of $\mathcal{P}(k)$ is such that

$$\sigma_8 \equiv \sigma(R = 8h^{-1}\ {\rm Mpc}, z = 0) = 0.801. \tag{13}$$

III. CLUSTER NUMBER COUNTS

The mean number density, n, of objects with mass greater than m, at redshift z can be calculated by

$$n(>m,z) = \int_m^\infty \frac{dn}{dM} dM, \tag{14}$$

where dn/dM is the differential number density. In the presence of local non-Gaussianity MVJ [43] used a saddle-point expansion (assuming that the deviation from Gaussianity is sufficiently weak for such an expansion to converge) to derive a correction factor for dn/dM of the form

$$\mathcal{R} = \exp\left(\frac{S_3\delta_c^3}{6\sigma_M^2}\right)\left[\frac{\delta_c^2}{6\Delta}\cdot\frac{dS_3}{d\ln\sigma_M} + \Delta\right], \tag{15}$$

$$\Delta \equiv \sqrt{1 - \frac{\delta_c S_3}{3}}, \tag{16}$$

where the third cumulant, S_3, is given by $S_3 = \langle\delta^3\rangle$, and is assumed to be almost independent of the smoothing mass-scale, M. This latter assumption indeed holds on cluster scales (see (18) below). S_3 can be calculated either from a three-dimensional integral

$$S_3(M) = \frac{6 f_{R;RR}(0)}{\sigma_M^4}, \tag{17}$$

with $f_{R;RR}$ defined in Eq. (34), or from the fitting formula [36]

$$S_3(M) = \frac{3.15\times10^{-4} f_{\rm NL}}{\sigma_M^{0.838}}. \tag{18}$$

The critical overdensity, δ_c, is taken to be[2]

$$\delta_c = \sqrt{a}\times 1.686, \tag{19}$$

with the "fudge factor" $\sqrt{a} = 0.9$ as recommended by [45], although there is still debate over its value [44,46].

There are alternative forms of the correction factor, $\mathcal{R}$, given by LoVerde *et al.*[47] based on a low-order Edgeworth expansion, and by Paranjape *et al.*[48] based on resumming terms in the saddle-point expansion of the mass function. We tested both of these alternative corrections and found that, in the range of parameters used in this paper, there are only small differences between the various prescriptions and our main results are unaffected by the choice of the correction factor. In the rest of this work, we shall use only the MVJ correction factor (see [45] for a comparison between the correction factors).

In summary, we shall consider the non-Gaussian differential abundance of the form

$$\frac{dn}{dM} = \mathcal{R}\times F(\nu)\frac{\rho_m}{M}\frac{d\ln\sigma^{-1}}{dM}, \tag{20}$$

where $\nu \equiv \delta_c/\sigma_M$ and $F(\nu)$ is one of the following three standard mass-functions

press-schechter[49] $$F_{\rm PS} = \sqrt{\frac{2}{\pi}}\nu e^{-\nu^2/2}, \tag{21}$$

Sheth-Torman[50] $$F_{\rm ST} = 0.322\sqrt{\frac{2a}{\pi}}\nu\exp\left(-\frac{a\nu^2}{2}\right)[1+(a\nu^2)^{-0.3}], \qquad a = 0.707, \tag{22}$$

Tinker etal.[51,52] $$F_{\rm Tinker} = 0.368[1+(\beta\nu)^{-2\phi}]\nu^{2\eta+1}e^{-\gamma\nu^2/2},$$
$$\beta = 0.589(1+z)^{0.2}, \quad \phi = -0.729(1+z)^{-0.08},$$
$$\eta = -0.243(1+z)^{0.27}, \quad \gamma = 0.864(1+z)^{-0.01}. \tag{23}$$

The Press-Schechter and Sheth-Tormen mass functions can be derived by considering the overdensity, δ, as a stochastic function of the smoothing mass scale, M, and associating trajectories [in the (M,δ) plane] that overshoot a barrier, δ_c, with a collapsed object. One can show that a spherical collapse can be associated with a barrier of constant height, resulting in the Press-Schechter mass function [49], while an ellipsoidal collapse can be associated with a drifting barrier, $\delta_c = 1.686(1+\alpha(\delta/\sigma)^\beta)$ (α, β constant), giving the Sheth-Tormen mass function [50]. The Tinker mass function belongs to a family of so-called universal mass functions derived from a suite of N-body simulations, with the functional form deviating from simulation results by $\lesssim 5\%$ in the redshift range considered here (for detail see [51,52], and also [53–55]).

[2]Note that δ_c is taken to be constant, while the redshift dependence is carried by the factor $\mathcal{A}(k,z)$. This convention agrees with [34,35] but is different from the "excursion-set" convention in which the redshift dependence is carried by δ_c, with σ extrapolated to $z = 0$ (e.g. [44]). We believe our present convention will facilitate comparison with [34,35], whose results will be used in the next section.

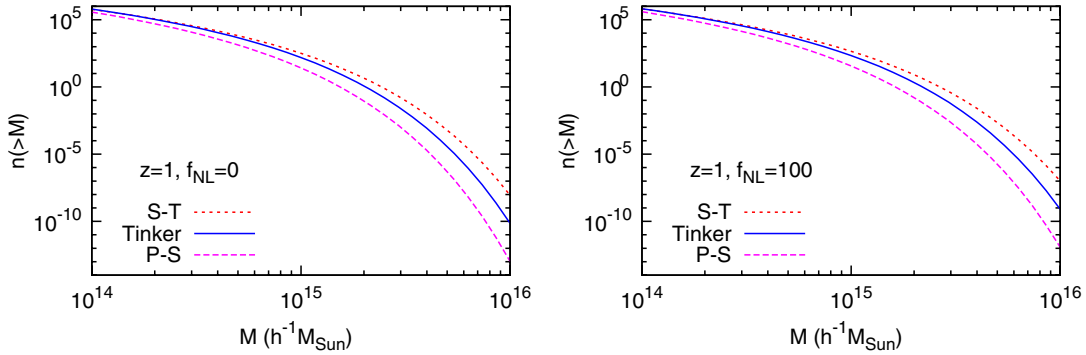

FIG. 1 (color online). The number of objects above mass M with $f_{\rm NL} = 0$ (left) and 100 (right) at $z = 1$, calculated over the full sky using the Press-Schechter (P-S), Sheth-Tormen (S-T), and Tinker *et al.* mass functions. With $f_{\rm NL} = 100$, the number count increases by roughly an order of magnitude at the high-mass end compared with $f_{\rm NL} = 0$.

The number of objects with mass above M expected at redshift z is given by the integral

$$\frac{dV}{dz} \times \int_M^\infty \frac{dn}{dm} dm, \tag{24}$$

where the volume element dV/dz satisfies

$$\frac{dV}{dz} = f_{\rm sky} \frac{4\pi}{H(z)} \left(\int_0^z \frac{dz'}{H(z')} \right)^2, \tag{25}$$

$$H(z) \approx H_0[\Omega_m(1+z)^3 + \Omega_\Lambda]^{1/2}, \tag{26}$$

and $f_{\rm sky}$ is the fraction of the sky covered by the survey. The number count for $z = 1$, $f_{\rm sky} = 1$, and $f_{\rm NL} = 0$ or 100 is shown in Fig. 1. Comparing the mass functions, we see that the Sheth-Tormen gives the highest number count, followed by the Tinker and the Press-Schechter mass functions. Changing $f_{\rm NL}$ to 100 (right panel) increases the number count at the high-mass end by roughly an order of magnitude. See e.g. [46,56] for more comparisons between various mass functions.

IV. BIAS

In the seminal work of Dalal *et al.* [33], it was shown quantitatively how non-Gaussianity gives rise to characteristic changes in the clustering of density peaks corresponding to rare objects. At leading order, it is common to define the bias in Fourier space as the ratio of the power-spectra

$$b^2(k) = \frac{P_{\rm halo}(k)}{P_m(k)}, \tag{27}$$

which represents the amplitude at which density peaks ($P_{\rm halo}$) trace the underlying dark-matter distribution (P_m). The Fourier-space formalism was used by a majority of papers on non-Gaussian bias (e.g. [45,57–60]).

However, an arguably more intuitive measure of the bias is in real space, where the density fluctuation in peaks (i.e. luminous objects) is expressed as a nonlinear function of the local dark-matter density fluctuation. On linear scales, the bias is given by the ratio of the correlation functions [61,62]

$$b^2(r) = \frac{\xi_{\rm pk}(r)}{\xi(r)}, \tag{28}$$

where r is the comoving length in Eulerian space (throughout this work quantities with a subscript "pk" are associated with density peaks). The correlation function, ξ, is defined as

$$\xi(\mathbf{x_1}, \mathbf{x_2}) = \langle \delta(\mathbf{x_1}), \delta(\mathbf{x_2}) \rangle, \qquad r = |\mathbf{x_1} - \mathbf{x_2}|. \tag{29}$$

On linear scales where $\mathcal{P}(k)$ is of a power-law form parametrized by $n_{\rm s}$, we can write

$$\xi(r) = 4\pi \int_0^\infty \frac{dk}{k} \mathcal{P}(k) j_0(kr), \tag{30}$$

where $j_0(x) = \sin x/x$. The real-space bias tells us directly about the clustering amplitude of density peaks separated by distance r. We shall refer to r as the separation length.

Unfortunately, when comparing (27) and (28), we see that the real-space bias, $b(r)$, and the Fourier-space bias, $b(k)$, are not simply related via a Fourier transform but rather a complicated convolution. In [37], we avoided this problem by interpreting (28) as a ratio of joint probabilities of finding overdensities at two points distance r apart, and then applying a bivariate Edgeworth expansion. Because of the algebraic nature of the Edgeworth expansion, this technique was readily applied to non-Gaussianity parametrized by the cubic order parameter, $g_{\rm NL}$, but surprisingly the application is much less straightforward for $f_{\rm NL}$.

An alternative method for calculating the real-space bias in the presence of $f_{\rm NL}$ was presented by Valageas [34,35] in which he showed that analytic calculations could be made as long as the separation length is sufficiently large. In this work, we shall follow this formalism, of which we give a simplified account here.

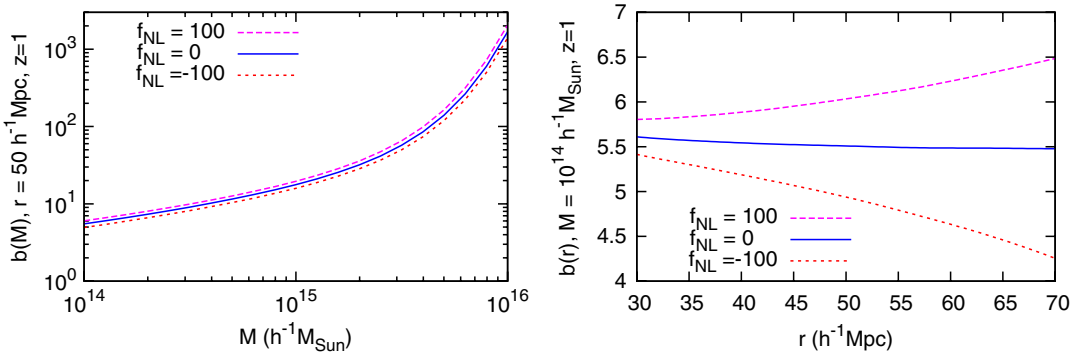

FIG. 2 (color online). The effect of $f_{\rm NL}$ on the real-space bias, b, at $z = 1$. In the panel on the left, the bias is shown as a function of smoothing mass-scale, M, with separation length $r = 50h^{-1}$ Mpc. The other panel shows the effect of varying r with $M = 10^{14}h^{-1}M_\odot$, illustrating the scale dependence of the bias when $f_{\rm NL} = \pm 100$.

A crucial element in the real-space approach is the mapping between the separation length, s, in Lagrangian coordinates (associated with linear density fluctuations) and that in Eulerian coordinates (associated with nonlinear fluctuations). This relation is given by

$$s \simeq r\left(1 + \frac{2\delta_R(r)}{3}\right), \tag{31}$$

accurate at large distances where $\delta_R(r) \ll 1$. Here $\delta_R(r)$ can be interpreted as the radial profile of the linear density contrast from the center of the halo. The profile is given by

$$\delta_R(r) = \frac{\delta_c}{\sigma_R^2}\sigma_{R,0}^2(r) + \frac{\delta_c^2}{\sigma_R^4}\Big[f_{0;RR}(r) + 2g_{R;0R}(r) - 3\frac{\sigma_{R,0}^2(r)}{\sigma_R^2}f_{R;RR}(0)\Big]. \tag{32}$$

In this expression, the functions $\sigma_{R_1,R_2}(r)$, $f_{R;R_1R_2}(r)$, and $g_{R;R_1R_2}(r)$ are defined by the following integrals[3]

$$\sigma_{R_1,R_2}^2(r) = 4\pi\int_0^\infty \frac{dk}{k}\mathcal{P}(k)W(kR_1)W(kR_2)j_0(kr), \tag{33}$$

$$f_{R;R_1R_2}(r) = 8\pi^2 D(0)f_{\rm NL}\int_0^\infty \frac{dk_1}{k_1}\mathcal{P}(k_1)W(k_1R_1) \times \int_0^\infty \frac{dk_2}{k_2}\mathcal{P}(k_2)W(k_2R_2)\int_{-1}^1 d\mu W(kR) \times \frac{\mathcal{A}(k)}{\mathcal{A}(k_1)\mathcal{A}(k_2)}j_0(kr), \tag{34}$$

[3]In this paper $f_{\rm NL}$ is defined in the "LSS" convention. The "CMB" convention, as used in [34], satisfies $f_{\rm NL}^{\rm CMB} = D(0)f_{\rm NL}^{\rm LSS}$.

$$g_{R;R_1R_2}(r) = 8\pi^2 D(0)f_{\rm NL}\int_0^\infty \frac{dk_1}{k_1}\mathcal{P}(k_1)W(k_1R_1)j_0(k_1r) \times \int_0^\infty \frac{dk_2}{k_2}\mathcal{P}(k_2)W(k_2R_2)\int_{-1}^1 d\mu W(kR) \times \frac{\mathcal{A}(k)}{\mathcal{A}(k_1)\mathcal{A}(k_2)}, \tag{35}$$

where μ is the cosine of the angle between $\mathbf{k_1}$ and $\mathbf{k_2}$, and $k = \sqrt{k_1^2 + k_2^2 + 2k_1k_2\mu}$.

With these definitions, Valageas showed via a saddle-point expansion that the bias for objects mass M is given by

$$b^2(M, r) = \frac{1}{\sigma_{0,0}^2(r)}[(1 + \delta_R(s))e^{\Delta(s)} - 1], \tag{36}$$

where

$$\Delta(s) = \frac{\sigma_{R,R}^2(s)\delta_c^2}{u\sigma_R^2} + \frac{2\delta_c^3}{u^3}\Big[f_{R;RR}(s) + 2g_{R;RR}(s) + \Big(1 - \frac{u^3}{\sigma_R^6}\Big)f_{R;RR}(0)\Big], \tag{37}$$

$$u = \sigma_R^2 + \sigma_{R,R}^2(s). \tag{38}$$

Figure 2 shows the real-space bias for a range of M and r. Keeping r fixed and varying M (left panel), non-Gaussianity shifts $b(M)$ up or down (depending on the sign of $f_{\rm NL}$). On the other hand, keeping M fixed and varying r, we see how nonzero $f_{\rm NL}$ introduces a scale dependence on $b(r)$ ($b(r)$ is roughly constant on large scale if $f_{\rm NL} = 0$). This scale dependence is similar to that seen in [37] for $g_{\rm NL}$.

In the limit of large separation length, the non-Gaussian bias $b(r)$ (with M fixed) follows a simple scaling relation. Given $f_{\rm NL}$, we can define the deviation from the Gaussian bias as $\Delta b \equiv b(f_{\rm NL}) - b(f_{\rm NL} = 0)$. For $r \gg R$, it was shown that [34]

$$\Delta b \propto f_{\rm NL}b(f_{\rm NL} = 0)\Big[\frac{r}{h^{-1}\ {\rm Mpc}}\Big]^2, \tag{39}$$

so that the overall scaling is $\Delta b \sim r^2$, since $b(f_{\rm NL} = 0)$ is approximately constant for large r.

In later work, it will be necessary to define the effective bias associated with a comoving volume. For a spherical region of comoving radius L, we define such a bias as

$$b_L(M) = \frac{1}{V^2}\int_V \mathbf{dx_1}\int_V \mathbf{dx_2} b(M, |\mathbf{x_1} - \mathbf{x_2}|), . \quad (40)$$

It will also be useful to define the effective bias for objects of mass $>M$

$$b(>M, r) \equiv \frac{1}{n(>M)}\int_M^\infty b(m, r)\frac{dn}{dm}dm. \quad (41)$$

Combining the averaging processes (40) and (41), we can define

$$\beta(L, M) \equiv b_L(>M), \quad (42)$$

which, as shown in the Appendix, simplifies in the limit $r \gg R(M)$ to

$$\beta(L, M) \approx F(L)G(M), \quad (43)$$

where

$$F(L) = 1 + \tfrac{6}{5}K(z)f_{\rm NL}\left[\frac{L}{h^{-1}\ {\rm Mpc}}\right]^2, \quad (44)$$

$$G(M) = \frac{1}{n(>M)}\int_M^\infty b(m, f_{\rm NL} = 0)\frac{dn}{dm}dm. \quad (45)$$

An example of the effective bias, β, with $L = 100h^{-1}$ Mpc is shown in Fig. 3, in which we set $z = 1$ and use the Tinker mass function for dn/dM, although using a different mass function only results in small differences. Comparing this graph with that of $b(M)$ (left panel of Fig. 2), we see that while β retains the overall shape of the curves, it is clearly more sensitive to non-Gaussianity. This quantity will be especially useful in the next section in which we consider the clustering of massive objects within a specified volume in the presence of non-Gaussianity.

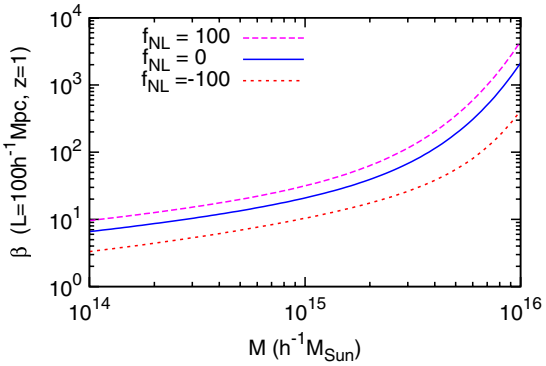

FIG. 3 (color online). The effective bias, β [defined by Eq. (43)] associated with objects of mass greater than M in a spherical volume of radius $L = 100h^{-1}$ Mpc. Compared with Fig. 2, the effect of non-Gaussianity on β is much more apparent.

We conclude this section with a brief comparison between the non-Gaussian imprints in the bias and in the cluster counts. Because the non-Gaussian imprint on the clustering of biased objects is significant only on large scales, surveys covering a large volume ($\sim \mathcal{O}(10)$ Gpc^3) will be required. Preliminary forecasts have shown good prospects of achieving $f_{\rm NL} \sim \mathcal{O}(1)$ constraints from measurements of the bias with upcoming surveys such as Dark Energy Survey,[4] Euclid,[5] and Large Synoptic Survey Telescope[6] [63,64]. On the other hand, cluster number counts, while not requiring a large-volume survey, are almost completely insensitive to the shape of non-Gaussianity [47,65]. The bias probes correlation between scales and is therefore sensitive to the shape of non-Gaussianity, particularly the local shape, whereas the equilateral shape shows up only weakly in the bias [66]. This suggests that a combination of these probes will be required to constrain both the amplitude and the shape of non-Gaussianity.

V. EXTREME-VALUE DISTRIBUTIONS

In this section, we present the calculation of the distribution[7] of extreme-mass clusters. The necessary ingredients are the non-Gaussian number density and real-space bias calculated in the previous sections.

A. Distribution function

White [21] derived the following expression for the cumulative probability that a region of volume V contains no object of mass M and above

$$P(M) = \exp\left[\sum_{k=1}^\infty \frac{(-n(>m))^k}{k!} \times \left(\prod_{i=1}^k \int_V \mathbf{dx_i}\right)\xi_k^{\rm pk}(\mathbf{x_1}, \mathbf{x_2}, \ldots \mathbf{x_k})\right], \quad (46)$$

where $\xi_1^{\rm pk} \equiv 1$, $n(>M)$ is given by (14) and $\xi_k^{\rm pk}$ is the k-point correlation function of density peaks in V associated with halos of mass $>M$. As in [22,23], we shall at times refer to V as a "patch." If we take the patch to be a sphere of comoving radius L, the volume-averaged correlation then simplifies to the cumulant (connected moment) smoothed by a top-hat window of radius L as follows

$$\left(\prod_{i=1}^k \int_V \frac{\mathbf{dx_i}}{V}\right)\xi_k^{\rm pk}(\mathbf{x_1}, \mathbf{x_2}, \ldots \mathbf{x_k}) = \langle \delta_{\rm pk}^k\rangle_c(L) = (\sigma^{\rm pk}(L))^{2k-2}S_k^{\rm pk}(L) = (\beta(L, M)\sigma_L)^{2k-2}S_k^{\rm pk}(L), \quad (47)$$

[4] http://www.darkenergysurvey.org
[5] http://sci.esa.int/euclid
[6] http://www.lsst.org
[7] We use the word "distribution" in the strict sense, referring to the cumulative distribution and not the PDF.

where β is given by (42). The cumulants for density peaks have been calculated in the context of hierarchical structure formation with Gaussian initial condition [67–70]. In the presence of non-Gaussianity, however, the perturbation theory required to calculate the cumulants for density peaks becomes much more complex (see e.g. [70–72]). To make analytic progress, we shall consider only k up to 3 in the sum (46). The terms $k = 1$ and 2 correspond to well-known results previously found in [23,29], namely

$$-nV + \tfrac{1}{2}(nV\beta\sigma_L)^2, \tag{48}$$

where we have used the reduced cumulant $S_2 = 1$. Given a weakly non-Gaussian initial condition, the third cumulants for density peaks are expected to be dominated by nonlinear gravitational effects, since primordial nonvanishing cumulants decay at the rate $S_k/D^{k-2}(z)$ [73]. Neglecting these effects, we can use the expression for the cumulant of the lognormal distribution [74]

$$S_3^{\rm pk}(L) = 3 + \sigma_L^2, \tag{49}$$

which was found to be in fair agreement with N-body simulations of non-Gaussian models with $|f_{\rm NL}|$ as large as 1000 (at least in the quasilinear regime with $\sigma \simeq 1$) [44]. This approximation is sufficient for the range of cluster masses ($\geq 10^{14}h^{-1}M_\odot$) examined in this work.

Collecting these results, we find the extreme-value distribution

$$\ln P(M) \approx -X + \frac{1}{2}X^2Y^2 - \frac{1}{6}X^3Y^4S_3^{\rm pk}(L), \quad \text{where } X \equiv n(>M)V, \quad Y \equiv \beta(L,M)\sigma_L. \tag{50}$$

Setting (50) equal $\log(1/2)$ (i.e. the median value of $M_{\rm max}$) gives an estimate of the modal value of $M_{\rm max}$, at least in weakly non-Gaussian distributions (see [29] for the Gaussian case). However, the equation is nonlinear in $M_{\rm max}$ and the approximate $M_{\rm max}$ dependences in these terms are not intuitive. Instead, we look for the peak in the derivative of (50), i.e. the probability density function. Nevertheless, the shape of the distribution function holds valuable statistical information to which we shall return when we consider the extremal-type distributions in Appendix B.

B. PDF of extreme-mass objects

We can obtain the PDF for the most massive objects expected in a volume by differentiating the distribution function (50) with respect to M, noting that the only dependence on M is in the number density, n, and the bias, b. The result is

$$\begin{aligned} p(M) &= \frac{dP}{dM} \\ &= VP(M)\Big[-\frac{dn}{dM}\Big(-1 + nV\beta^2\sigma_L^2 \\ &\quad - \tfrac{1}{2}n^2V^2\beta^4\sigma_L^4 S_3(L)\Big) \\ &\quad + n^2V\beta\sigma_L^2\frac{d\beta}{dM}\Big(1 - \tfrac{2}{3}nV\beta^2\sigma_L^2 S_3(L)\Big)\Big], \end{aligned} \tag{51}$$

where dn/dM is given by (20) [note the subtlety that $dn(>M)/dM = -dn/dM$]. Here, we see explicitly that the PDF of extreme-mass objects not only depends on the bias, but also on its mass variation, $d\beta/dM$.

C. $f_{\rm NL}$ and extreme-value PDF

The main results of this paper are shown in Fig. 4. The panels show the probability density function (51) for the three mass functions at redshift $z = 1$, 1.6, and 3 (corresponding to the left, middle, and right column) with $f_{\rm NL} =$ 200, 100, and 0 (top, middle, and bottom row, respectively). The survey volume is taken to be a sphere of radius $100h^{-1}$ Mpc. To display the correct scaling on the horizontal log scale, we plot $dP/d\log M_{\rm max}$ on the vertical axis while the actual value of the PDF is $dP/dM_{\rm max}$. From these graphs, we make the following observations:

(a) Going from the bottom row to the top, we see that increasing $f_{\rm NL}$ increases the height of the PDF while positively skewing it (i.e. lifting the positive tail). This has the effect of increasing the mass of the most probable extreme objects in a given volume.

(b) Going from the first column to the third, we see that at higher redshifts, the PDFs are more peaked and the peaks are located at lower $M_{\rm max}$.

(c) The Sheth-Tormen mass function predicts the largest mass of extreme objects, followed by the Tinker and the Press-Schechter mass functions. This is a consequence of their predicted number densities as seen in Fig. 1.

(d) The differences between the mass functions become much more pronounced at high redshifts. In the third column, we see a clear separation of the peaks for different mass functions, with non-Gaussianity further enhancing the differences.

(e) In Fig. 5, we show the effect of varying the patch radius, L from 100 to $500h^{-1}$ Mpc (with $z = 1$ and $f_{\rm NL} = 100$). By increasing L, the peak of the PDF shifts significantly to higher $M_{\rm max}$. The PDF also becomes more peaked with increasing L. This is simply due to the fact that as the sample size, L, approaches the population size, repeating the sampling will yield almost identical maxima in the samples.

Finally, we investigate the relative importance of the three terms on the right-hand side of Eq. (50). We consider the extreme-value distributions in following cases

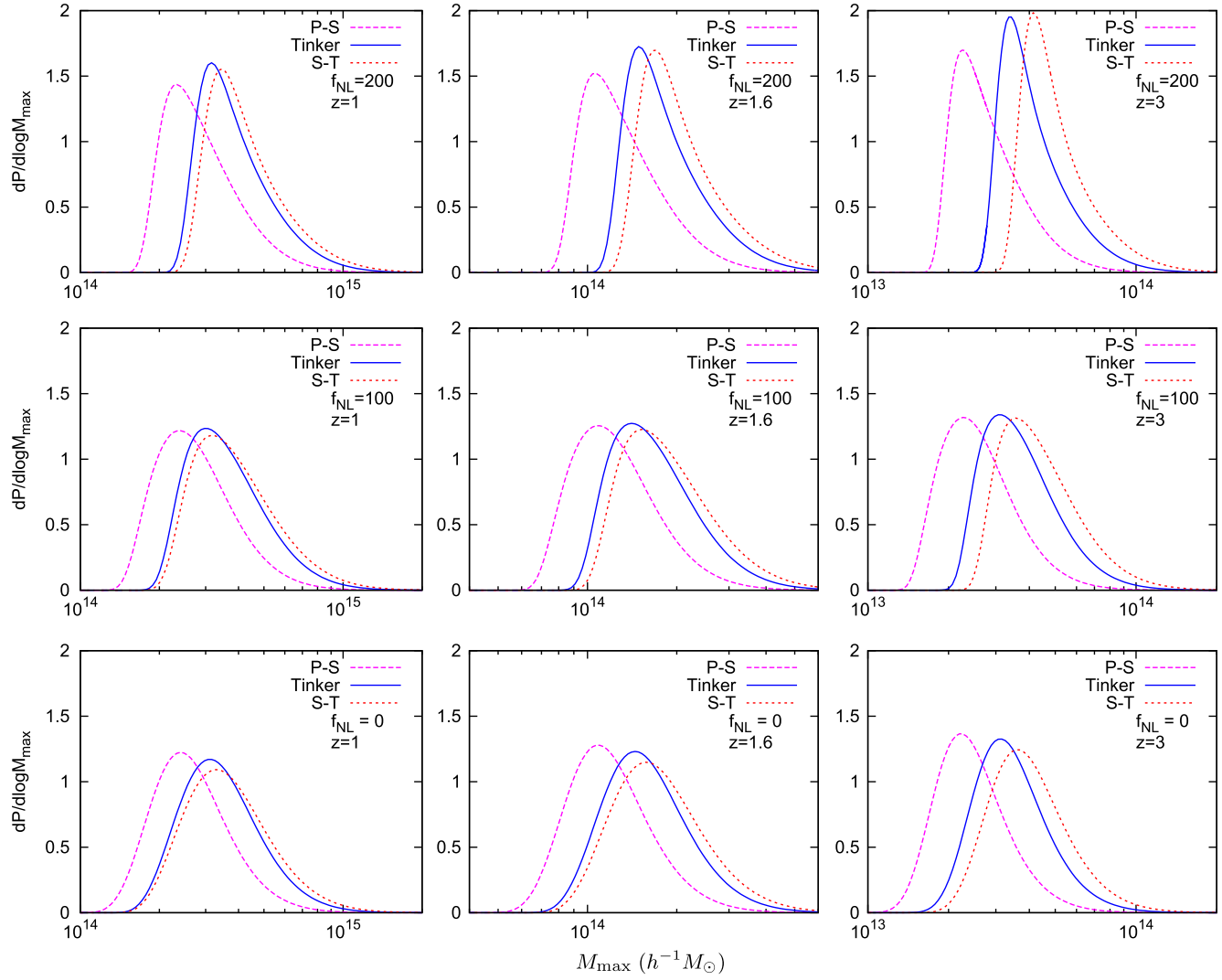

FIG. 4 (color online). The probability density function of the maximum mass, $M_{\rm max}$, of objects in a spherical volume of radius $L = 100h^{-1}$ Mpc. In each panel, the mass functions used are Press-Schechter (dashed line), Tinker (solid line), and Sheth-Tormen (dotted line). *Top row:* $f_{\rm NL} = 200$ with $z = 1$, 1.6, and 3 (from left to right). The PDF at the same redshifts are shown for $f_{\rm NL} = 100$ (middle row) and $f_{\rm NL} = 0$ (bottom row). The non-Gaussian effects are most visible in the third column in which the peaks can be seen to move to higher $M_{\rm max}$ with increasing $f_{\rm NL}$.

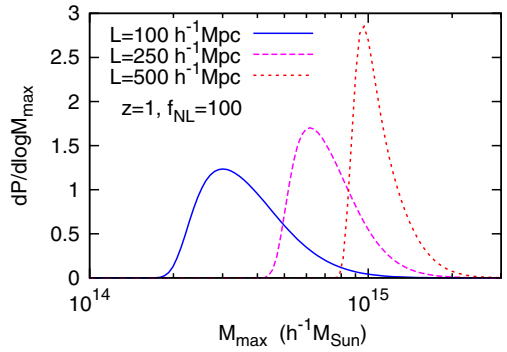

FIG. 5 (color online). The extreme-value PDF for patch sizes $L = 100h^{-1}$ Mpc (solid line), $250h^{-1}$ Mpc (dashed line), and $500h^{-1}$ Mpc (dotted line). The Tinker mass function is used, with $z = 1$, and $f_{\rm NL} = 100$. We see that the location of the peak is clearly very sensitive to changes in L.

(A) $f_{\rm NL} = 100$, $L = 100h^{-1}$ Mpc, $z = 1$,
(B) $f_{\rm NL} = 200$, $L = 500h^{-1}$ Mpc, $z = 3$,
(C) $f_{\rm NL} = 0$, $L = 500h^{-1}$ Mpc, $z = 3$.

Figure 6 shows the resulting distributions when one, two, or three terms on the right-hand side of (50) are taken into account (using the Tinker mass function). In case (A), we see that the Poisson approximation (keeping only the first term in (50)) is fairly close to the 3-term result. Generally, this holds as long as the non-Gaussian effects are small (i.e. at small volume, low redshift). However, in case (B), we see that the Poisson approximation underestimates the extreme cluster masses. In this case the peak of the PDF, though merely shifted by $\lesssim 10\%$, is much narrower and rises to a higher maximum value (as can be estimated by eye from the slope around $P = 0.5$). In both cases,

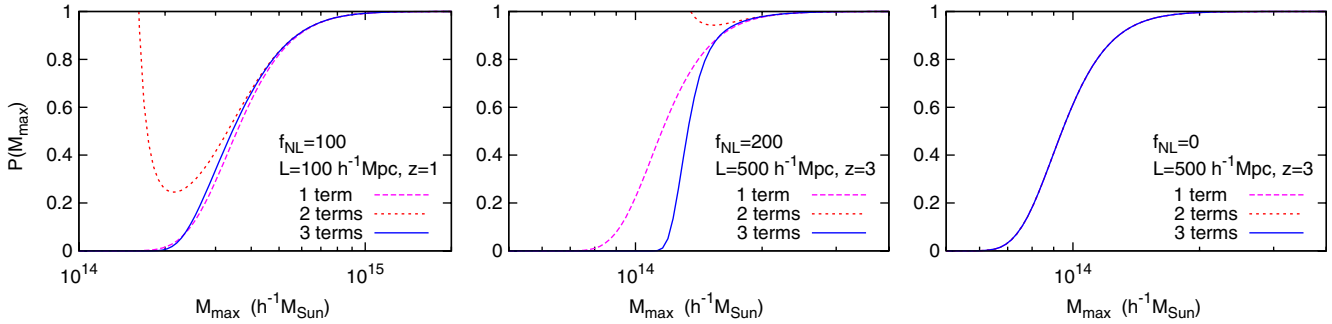

FIG. 6 (color online). The contributions of the 3 terms in (50) towards the extreme-value distributions, using 3 sets of parameters Left panel: (A) $f_{\rm NL} = 100$, $L = 100h^{-1}$ Mpc, $z = 1$. Middle panel: (B) $f_{\rm NL} = 200$, $L = 500h^{-1}$ Mpc, $z = 3$. Right panel: (C) $f_{\rm NL} = 0$, $L = 500h^{-1}$ Mpc, $z = 3$. In case (B), the Poisson approximation ("1-term") is clearly inaccurate.

however, it is inconsistent to include the bias but ignore the third term (representing the skewness of the halo distribution in the sample). This is shown in the upturns of the 2-term distribution functions. In case (C), where $f_{\rm NL} = 0$, the 2nd and 3rd correction terms are negligible and the Poisson approximation is very good indeed.

Strictly speaking, the calculations here are valid only for $L \gg r$. In actual applications, we will be interested in the cases where $L = \mathcal{O}$ (a few Gpc). In such cases, the redshift variation within the patch must be taken into account (as emphasized in [22]). This requires replacing the number density, bias, and cumulants by their average within a comoving volume. We shall demonstrate this in the next section.

VI. APPLICATIONS

A. A massive cluster at $z = 1.579$: A problem for ΛCDM?

Santos *et al.* [14] recently reported the discovery of a cluster XMMUJ0044.0-2-33 (hereafter XMMUJ0044) at $z = 1.579$, detected in the x-ray data of the XMM-Newton telescope and later followed up spectroscopically. The cluster mass was estimated to be $\sim 3.5 - 5 \times 10^{14} M_\odot$, far greater than the previous x-ray cluster of mass 5.7×10^{13} at $z = 1.62$ reported by Tanaka, Finoguenov and Ueda [15]. We shall now use extreme-value statistics to study the probability of finding XMMUJ0044 as the maximum-mass cluster. In particular, we ask, is the existence of XMMUJ0044 consistent with $f_{\rm NL} = 0$?

B. Eddington bias

We take the mass of XMMUJ0044 to be[8] $M_{\rm obs} = (4.46 \pm 0.79) \times 10^{14} M_\odot (= (3.12 \pm 0.55) \times 10^{14} h^{-1} M_\odot)$ and ask: what is the probability that this cluster is the most massive one in the redshift range $1.579 \lesssim z \lesssim 2.2$? Here the redshift upper bound is consistent with the highest redshift probed by the XMM survey.

The reported cluster mass must first be corrected for Eddington bias, which refers to the apparent boost in the cluster mass due to the fact that it is more likely for lower-mass objects to scatter to high luminosity than it is for rarer massive objects to scatter to lower luminosity. We account for this effect by the correction [20]

$$\ln M = \ln M_{\rm obs} + \tfrac{1}{2}\gamma\sigma^2_{\rm lnM}, \tag{52}$$

where $\sigma_{\rm lnM} \sim 0.3$ is the error estimated from the observation and γ is the local slope of the mass function determined using the relation $dn/d\ln M \propto M^\gamma$, and therefore satisfies

$$\gamma = \frac{d^2 n}{du^2} \Big/ \frac{dn}{du}, \qquad u \equiv \ln M. \tag{53}$$

The final masses with Eddington-bias corrections are listed in Table I. The corrected mass depends on $n(>M)$ and hence it also depends on the mass function used. There is also a weak dependence on $f_{\rm NL}$ [entering via the MVJ correction (16)]. With $f_{\rm NL} = 100$, the changes in the corrected masses are less than a percent and it is reasonable to neglect this correction as long as $|f_{\rm NL}| \lesssim \mathcal{O}(10^2)$. We quote the corrected mass with $f_{\rm NL} = 0$ in Table I.

C. Redshift averaging

The patch of interest is now a spherical shell whose thickness is determined by the redshift band Δz. To account for the redshift variation within the shell, we perform the following modifications to the variables X and Y in the distribution function (50)

$$X = \langle n(>M)\rangle V = \int_{\Delta z} dz \int_M^\infty dm \frac{dn}{dm}\frac{dV}{dz}, \tag{54}$$

$$Y = \beta_{\rm shell}(M)\langle\sigma\rangle. \tag{55}$$

[8]The mass of XMMUJ0044 published in [14] is $(4.25 \pm 0.75) \times 10^{14} M_\odot$, given with respect to the critical density. Assuming NFW cluster profile, Waizmann *et al.* [30] showed that, with respect to the mean background density, the value increases to $(4.46 \pm 0.79) \times 10^{14} M_\odot$. We thank J. C. Waizmann for bringing this point to our attention.

TABLE I. Observed and Eddington-corrected mass for the cluster XMMUJ0044, in units of $10^{14}h^{-1}M_\odot$.

Observed mass	Eddington-corrected mass		
$(10^{14}h^{-1}M_\odot)$	Press-Schechter	Sheth-Tormen	Tinker
3.12 ± 0.55	2.48 ± 0.40	2.62 ± 0.43	2.56 ± 0.41

The bias averaged within the shell, $\beta_{\rm shell}$, is derived in Appendix A (Eq. (A17)) and is given by

$$\beta_{\rm shell} = \frac{L^3\langle F(L)\rangle + \ell^3}{L^3 - \ell^3}\langle G(M)\rangle. \tag{56}$$

Here L and ℓ are comoving lengths corresponding to redshifts 2.2 and 1.579, respectively. $\langle\sigma\rangle$, $\langle F(L)\rangle$ and $\langle G(M)\rangle$ are the redshift averages of (10), (44), and (45). We define the redshift average of a quantity, $Q(z)$, by

$$\langle Q\rangle = \frac{1}{V}\int_{\Delta z} dz Q(z)\frac{dV}{dz}. \tag{57}$$

D. Results

We are interested in the probability that a cluster of mass in the range $[\bar{M}+\sigma, \bar{M}-\sigma]$ is the maximum-mass object observed in a survey with a given $f_{\rm sky}$ and redshift range. Denoting this probability as Π, we can express it as the difference in the distribution function (50) evaluated at $\bar{M}\pm\sigma$,

$$\Pi = P(\bar{M}+\sigma) - P(\bar{M}-\sigma). \tag{58}$$

We take the mass range to be those in shown in Table I. In addition to the choice of mass function explored in the previous sections, here we consider three other factors that affect Π.

1. Dependence on $f_{\rm sky}$

The XDCP survey covers a sky area of approximately $80{\rm deg}^2$ [14]. However, the value of $f_{\rm sky}$ appropriate for our calculation must take into account all previous surveys that have explored the redshift interval in other parts of the sky, regardless of whether a positive detection is reported.

In Fig. 7 (left panel), we show Π as a function of survey area in square degree. Here Π is calculated using the Tinker mass function and $f_{\rm NL}=0$. We see that the observation of an extreme object such as XMMUJ0044 is most likely in a survey area of around $50{\rm deg}^2$ (where $\Pi\simeq 0.5$). At wider coverages, we expect the most probable extreme mass to be larger. In fact, taking $f_{\rm sky}=1$ as the most conservative limit, we find the most likely extreme object in this redshift range to be a cluster of mass $\sim 7-8\times 10^{14}h^{-1}M_\odot$ (Fig. 7, right panel), well above the Eddington-corrected mass of XMMUJ0044. On the other hand, taking $f_{\rm sky}=80{\rm deg}^2$, the most probable extreme-mass object is consistent with XMMUJ0044, as the peak of the extreme-value PDF lies within the mass estimate (vertical contours in Fig. 7).

It is difficult to estimate what is the correct value of $f_{\rm sky}$ is needed in such cases and, unfortunately, the question of whether non-Gaussianity is needed to explain the existence of certain clusters depends sensitively on the value of $f_{\rm sky}$ assumed. Making the most conservative interpretation using $f_{\rm sky}=1$ and assuming that there are good prospects for discovering many more massive high-redshift clusters in

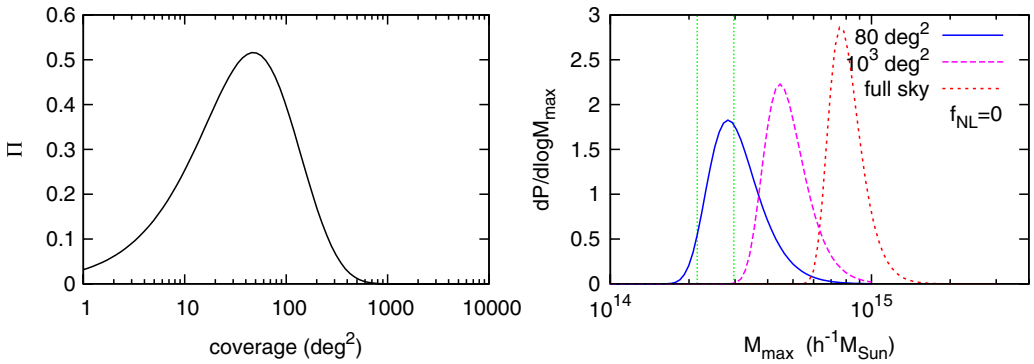

FIG. 7 (color online). Left panel: The probability Π pEq. (58)] that XMMUJ0044 is the most massive cluster in $1.579 < z < 2.2$ plotted as a function of sky coverage (in square degree). Using the Tinker mass function, the probability is maximized to $\approx 50\%$ when $f_{\rm sky}$ is of order $50{\rm deg}^2$. Right panel: The extreme-value PDF with coverages (from left to right) $80{\rm deg}^2$, $1000{\rm deg}^2$ and full sky (using the Tinker mass function and $f_{\rm NL}=0$). Note from Table I that the cluster lies at $2.56\pm 0.41h^{-1}M_\odot$ (vertical dotted lines).

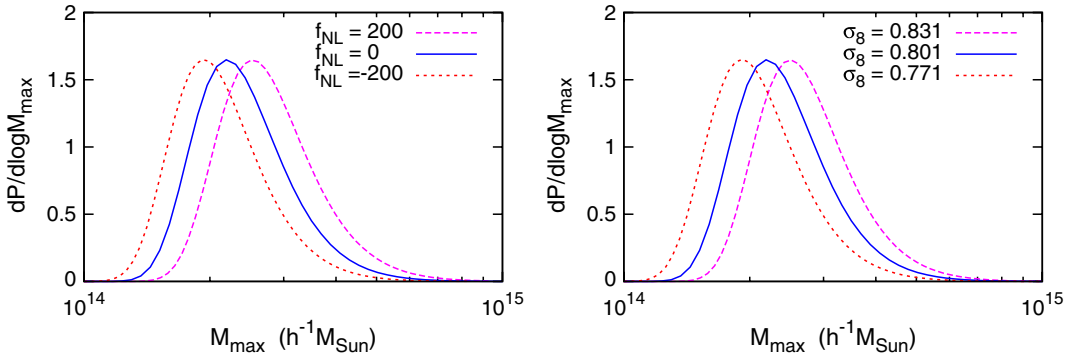

FIG. 8 (color online). The extreme-value PDF for objects in the range $1.579 < z < 2.2$ with a fiducial value for $f_{\rm sky} = 1$, using the Tinker mass function. Left panel: $f_{\rm NL}$ is varied from -150 to 150 with $\sigma_8 = 0.801$ Right panel: σ_8 is varied in the range $\sigma_8 = 0.801 \pm 0.03$ [8] with $f_{\rm NL} = 0$. The degeneracy is discussed in the text.

the future, we conclude that XMMUJ0044 presents no tension with ΛCDM (see also the conclusions of [20,31]).

2. Dependence on $f_{\rm NL}$.

In Fig. 8 (left panel) we show the effect of introducing $f_{\rm NL} = \pm 150$ to the extreme-value PDF. Here we take a fiducial value of $f_{\rm sky} = 1$, and use the Tinker mass function. As expected, the PDF is shifted left or right depending on the sign of $f_{\rm NL}$.

3. Dependence on σ_8

The panel on the right of Fig. 8 shows the equivalent effect when σ_8 is varied in the range $\sigma_8 = 0.801 \pm 0.03$ (taken from the WMAP 7-year constraint [8]), while $f_{\rm NL}$ is fixed to 0. The shift of the PDF towards more massive extreme objects follows from the fact that a greater σ_8 introduces a larger spread in the mass range of cluster in the survey volume.

The similarity between the two panels of Fig. 8 reflects the well-known degeneracy between $f_{\rm NL}$ and σ_8 [75,76]. This degeneracy can be easily broken, for instance, by the constraint on the galaxy power spectrum (which is sensitive to σ_8 but not $f_{\rm NL}$) or the CMB temperature anisotropies [19].

In summary, the degenerate effects between $f_{\rm sky}$, $f_{\rm NL}$ and σ_8 shown here imply that it is very difficult to deduce information on non-Gaussianity from the extreme-value distributions alone. The most sensible approach is combine the results from several cluster surveys (to achieve $f_{\rm sky} = 1$) with probes of the CMB (to break the σ_8 degeneracy), assuming selection effects and error in the mass determination can be kept in check.

E. The most massive object in the Universe

As a consistency check, we plot the extreme-value PDF for an extended redshift range $z > 0$ and $f_{\rm sky} = 1$ in Fig. 9. This gives the extreme-value PDF for the most massive object in the Universe. With $\sigma_8 = 0.801$, we find this to be

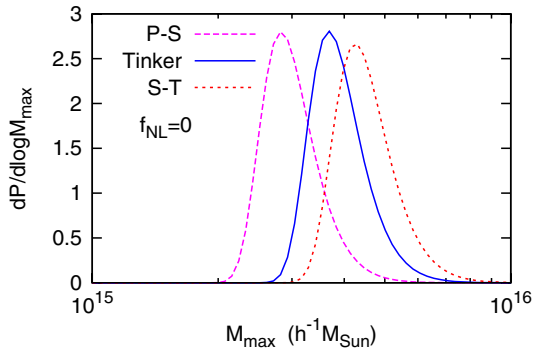

FIG. 9 (color online). The extreme-value PDF for the most massive object in the Universe (with $f_{\rm sky} = 1$ and $f_{\rm NL} = 0$) for various mass functions.

an object of mass $2 - 5 \times 10^{15} h^{-1} M_\odot$, depending on the mass function. The result using the Tinker mass function is $M_{\rm max} \approx 3.5 \times 10^{15} h^{-1} M_\odot$, which agrees broadly with those reported by [22,77]. The effect of $f_{\rm NL} = 100$ increases this value by less than 10%.

F. Extremal types

Extreme-value distributions obey a limit theorem analogous to the central limit theorem. This is the so-called *extremal types theorem*, which roughly states that extreme-value distributions converge to one of three possible types in the large-sample limit (see Appendix B). This beautiful theorem has found applications in areas such as meteorology, engineering, and finance, where a large volume of data allows extreme-value statistics to be modeled by fitting only one or two parameters of an extremal type (analogous to fitting the mean and variance of the normal distribution) [78,79].

In cluster cosmology, the dearth of cluster data at present is not ideal for application of the limit theorem, although there have been attempts to apply it to simulated data (see [22,23]). The question of which extremal-type extreme-mass clusters converge to remains unclear, although there

is some evidence from simulations that $f_{\rm NL}$ appears to play no role in the convergence [28,80]. We concur with this latter view and present the detail our investigation in Appendix B.

VII. CONCLUSION AND DISCUSSION

In summary, we have investigated quantitatively how the statistics of extreme-mass clusters is affected by uncertainties in the mass function, non-Gaussian corrections of the mass function and bias, Eddington correction, $f_{\rm sky}$, redshift, $f_{\rm NL}$, and σ_8. More specifically:

(1) We have presented a procedure to calculate the statistics of extreme-mass galaxy clusters in the presence of primordial non-Gaussianity parametrized by $f_{\rm NL}$. Our main results are the expressions for the cumulative probability distribution (50) and the probability density function (51) for the most massive object in a survey of a given sky coverage and redshift range. These expressions enable us to deduce the most probable extreme-mass cluster in a survey of a given specification. The effects of changing the mass function and varying the value of $f_{\rm NL}$, survey volume, and redshift are summarized in Figs. 4 and 5.

(2) Our correction terms for the extreme-value distribution [second and third terms on the right-hand side of (50)] are significant when considering a large volume, high redshift, or large non-Gaussianity (see Fig. 6). For non-Gaussianity with $f_{\rm NL} = \mathcal{O}(1)$, the first term of (50) (Poisson approximation) suffices.

(3) Next, we applied our formalism to investigate the extreme-value properties of cluster XMMUJ0044 0-2-33 ($M_{\rm obs} \simeq 3.12 \times 10^{14} h^{-1} M_\odot$ at $z = 1.579$). We find that the probability that the cluster is the most likely extreme-mass cluster in the survey depends sensitively on the assumed sky coverage, but is consistent with $f_{\rm NL} = 0$ (assuming $\sigma_8 = 0.801$). More conservatively, with $f_{\rm sky} = 1$, the most probable extreme-mass cluster is expected to be much larger and perhaps this will be confirmed by future x-ray cluster surveys.

(4) We show that the effect of $f_{\rm NL}$ in shifting the extreme-mass cluster to higher values is degenerate with an increase in σ_8 (Fig. 8). The degeneracy can be broken by combining cluster data with CMB constraints.

An important ingredient in our calculation is the mass function. In the presence of primordial non-Gaussianity, it remains to be seen what the correct mass function should be. Our investigation showed that the Press-Schechter, Sheth-Tormen, and Tinker mass functions give similar extreme-value statistics at low redshift, but there are large differences at high redshift and large $f_{\rm NL}$. The understanding of the correct form of the mass function appropriate for these extreme-mass objects is important since the uncertainty in the distribution of extreme-mass clusters due to the mass function is comparable with that from the mass determination (typically $\sigma_{\ln M} \sim 0.3$). Thus, it remains for further numerical simulations along the lines of [45,46] to establish the validity of the various mass functions and non-Gaussian correction factors in the presence of non-Gaussianity.

Another crucial ingredient is the bias which, in this work, was calculated using the real-space formalism given in [34,35] combined with our averaging procedure outlined in Appendix A. As pointed out in these papers, it is possible to extend the calculation to other types of non-Gaussianity (nonlocal or higher-order local type). It will be an interesting extension to study extreme-value statistics in the presence of different types of non-Gaussianity.

ACKNOWLEDGMENTS

We are indebted to the anonymous referee, whose many insightful comments led to a major improvement of the paper. We are also grateful to Olaf Davis for helpful discussions in the initial stages, and to Christopher Gordon, Aseem Paranjape, Shaun Hotchkiss and, in particular, Jean-Claude Waizmann for their comments on an early version of the manuscript. S. C. supported by Lincoln College, Oxford.

APPENDIX A: VOLUME-AVERAGED BIAS AND ITS APPROXIMATION

Let $b^G(M)$ be the real-space bias associated with objects of mass M with $f_{\rm NL} = 0$. In the text, we have seen that b^G is independent of r to a good approximation. In the presence of non-Gaussianity, the large-r behavior of $b(M, r)$ is given by

$$b(M, r) \approx f(r) g(M), \quad \text{where} \tag{A1}$$

$$f(r) \equiv 1 + K(z) f_{\rm NL} \left[\frac{r}{h^{-1}\ \text{Mpc}} \right]^2, \tag{A2}$$

$$g(M) \equiv b(M, f_{\rm NL} = 0), \tag{A3}$$

and $K(z)$ is independent of r, M, and $f_{\rm NL}$. This approximation allows the averaging (40) and (41) to be performed separately on $f(r)$ and $g(M)$.

Our goal is to perform the averaging (40) within a given volume. In the analysis of clusters lying within a redshift range Δz, the associated volume is a spherical shell whose thickness is proportional to Δz. Whilst the integration (40) could, in principle, be evaluated using a six-dimensional Monte Carlo integration, in this Appendix we show how (40) could be reduced to a triple integral. The results presented here are clearly applicable to other fields in which volume averages are required.

To begin, let us first consider an integral of the form

$$I(\mathbf{x_1}) = \int_V \mathbf{dx_2} f(|\mathbf{x_1} - \mathbf{x_2}|), \tag{A4}$$

where $\mathbf{x_1}$ is a fixed vector and V is a sphere.

1. Integration within a given sphere

If $\mathbf{x_1}$ lies inside a given sphere of radius L, we rotate the coordinate axes so that $\mathbf{x_1}$ lies along with the z-axis. We then translate the origin to the tip of $\mathbf{x_1}$. Let the spherical coordinates centered on this new origin be given by (u, θ, ϕ). One can show that the equation of the surface of the sphere is given by

$$u = -x_1 \cos\theta + \sqrt{L^2 - x_1^2 \sin^2\theta}, \qquad x_1 = |\mathbf{x_1}|. \tag{A5}$$

Hence, the integral (A4) can be written as

$$\int_V \mathbf{dx_2} f(|\mathbf{x_1} - \mathbf{x_2}|) = \int_0^\pi \sin\theta d\theta \int_0^U u^2 du \int_0^{2\pi} d\phi f(u, \theta, \phi), \tag{A6}$$

$$= 2\pi \int_{-1}^1 d\mu \int_0^U du u^2 f(u), \tag{A7}$$

where the integration limit is given by

$$U = -x_1 \mu + \sqrt{L^2 - x_1^2(1 - \mu^2)}. \tag{A8}$$

2. Integration outside a given sphere

If $\mathbf{x_1}$ lies outside a given sphere of radius ℓ, the same transformation gives

$$\int_V \mathbf{dx_1} f(|\mathbf{x_1} - \mathbf{x_2}|) = 2\pi \int_{-1}^{-\sqrt{1-\ell^2/x^2}} d\mu \int_{u^-}^{u^+} du u^2 f(u), \tag{A9}$$

where

$$u^\pm = -x_1 \mu \pm \sqrt{\ell^2 - x_1^2(1 - \mu^2)}. \tag{A10}$$

3. Averaging within a shell

Denote $r = |\mathbf{x_1} - \mathbf{x_2}|$. The net contribution of $f(r)$ within a spherical shell of inner radius, ℓ, and outer radius, L, is obtained by subtracting (A9) from (A7). Finally, the average of f over the entire shell is obtained by integrating $\mathbf{x_1}$ over the sphere, and then dividing by the volume of the sphere,

$$\bar{f}_{\text{shell}} = \frac{1}{V_{\text{shell}}^2} \int_V \mathbf{dx_1} \int_V \mathbf{dx_2} f(r), \tag{A11}$$

$$= \frac{4\pi}{V_{\text{shell}}^2} \int_\ell^L x_1^2 dx_1 [(A7)\text{-}(A9)], \tag{A12}$$

where

$$V_{\text{shell}} = \frac{4\pi}{3}(L^3 - \ell^3). \tag{A13}$$

4. Applications to the bias

Let us first apply (A12) to calculate the bias averaged within a sphere radius L (Eqs. (43)–(45)). Substituting $\ell = 0$ and $f(r)$ as in (A2) we find

$$\bar{f}_{\text{sphere}} = 1 + \tfrac{6}{5} K(z) f_{\text{NL}} \left[\frac{L}{h^{-1}\ \text{Mpc}}\right]^2. \tag{A14}$$

Therefore, the bias averaged over V for mass $>M$ [Eq. (42)] becomes

$$\beta(L, M) \approx \bar{f}_{\text{sphere}} G(M), \tag{A15}$$

where

$$G(M) \equiv \frac{1}{n(>M)} \int_M^\infty b(m, f_{\text{NL}} = 0) \frac{dn}{dm} dm. \tag{A16}$$

More generally, with nonzero ℓ, we find

$$\bar{f}_{\text{sphere}} = \frac{L^3 \bar{f}_{\text{sphere}}(L) + \ell^3}{L^3 - \ell^3}. \tag{A17}$$

Note that by setting $\ell = 0$, we recover $\bar{f}_{\text{sphere}}$. To include the averaging over redshift, one performs the replacement

$$\bar{f}_{\text{shell}} \to \frac{1}{V} \int_{\Delta z} dz \bar{f}_{\text{shell}} \frac{dV}{dz}. \tag{A18}$$

APPENDIX B: EXTREMAL TYPES

The shape of the extreme-value distribution function holds valuable information about the statistical nature of galaxy clusters. The following theorem, which roughly states that extreme-value distributions converge to one of only three possible types, lies at the heart of extreme-value theory.

Extremal Types Theorem—Let X_i be independent and identically distributed random variables. Define the block maximum as $M_n \equiv \max_{1 \le i \le n}\{X_i\}$. If, for some constants $a_n > 0$, b_n, we have

$$P(a_n M_n + b_n \le x) \to G(x) \quad \text{as } n \to \infty,$$

(in other words, if the distribution of rescaled maxima converges to a distribution G for large-sample size), then G is one of the following distributions:

(I) Gumbel type,

$$G(x) = \exp(-e^{-y}),$$

(II) Fréchet type,

$$G(x) = \begin{cases} 0, & x \le b \\ \exp(-y^{-\alpha}), & x > b \end{cases},$$

(III) Weibull type,

$$G(x) = \begin{cases} \exp(-(-y)^{\alpha}), & x \le b \\ 1, & x > b \end{cases},$$

where $y = ax + b$, a, b, c are constants, $a > 0$ and $\alpha > 0$.

See, for example, [81] for the proof. In this Appendix, we investigate which of these extremal types do the distributions of extreme-mass clusters belong to.

The following function will be useful in distinguishing between the three cases:

$$g(x) = -\ln(-\ln P(x)). \tag{B1}$$

In the case of the Gumbel distribution, $g(x) = P^{-1}(x) = \inf\{y: P(y) \ge x\}$, which means that $g(x)$ is the x-quantile of P. We shall refer to $g(x)$ as the quantile function [79,82].

To see which extremal type a given extreme-value distribution, $P(x)$, belongs to, one simply plots the quantile function and analyze its curvature for increasing patch size L. If the quantile is a linear, the distribution is of Gumbel type. If it concaves up (i.e. $g''(x) > 0$), the distribution is of Weibull type. If the quantile concaves down, it is of Fréchet type. Note that the quantiles must be plotted on linear and not logarithmic scales.

Figure 10 shows the quantile plot, $g(M_{\max})$, of the distribution of extreme-mass clusters, $P(M_{\max})$, with $f_{\rm NL}$ in the range 0–200. The parameters for each curve are those listed in the cases (A), (B), and (C) in Sec. V C, and the Tinker mass function is used. The concavity of these graphs clearly shows that the distribution of extreme-mass clusters are of the Fréchet type, although the tails of the quantile graphs show an almost linear (i.e. Gumbel) behavior.

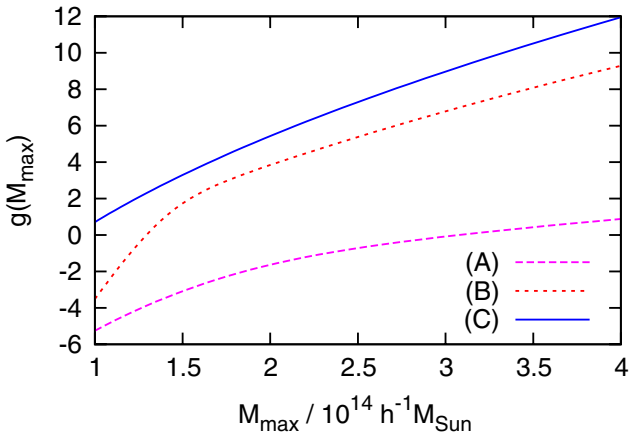

FIG. 10 (color online). The quantile plots for the distributions of extreme-mass clusters for the three cases: (A) $f_{\rm NL} = 100$, $L = 100h^{-1}$ Mpc, $z = 1$; (B) $f_{\rm NL} = 200$, $L = 500h^{-1}$ Mpc, $z = 3$; and (C) $f_{\rm NL} = 0$, $L = 500h^{-1}$ Mpc, $z = 3$. The Tinker mass function was used. The concavity of these curves suggests that they belong to the Fréchet class of distribution though they approach the Gumbel distribution at the high-mass end. This behavior is insensitive to changes in all other parameters.

The Fréchet distribution[9] arises in situations when there is a natural lower limit in the distribution function ($P(x) = 0$ for $x \le$ some constant). In our case, the definition of a galaxy cluster (e.g. via M_{200}) translates to a loose lower bound on $M_{\max}$ and this may explain why the distribution of extreme-mass clusters appears to be of the Fréchet type. If only the high-mass tail of the distribution is taken into account, the Gumbel distribution is a reasonable approximation. As pointed out in [23], if the underlying distribution is exactly Gaussian, the distribution can be shown to converge to the Gumbel type, albeit very slowly. In any case, we find that $P(M_{\max})$, for all practical purposes, belongs to the Fréchet family.

This conclusion is remarkably robust against changes in $f_{\rm NL}$, mass function, redshift, and patch size. It may be possible that this insensitivity stems from the truncation of the series (50). A more thorough approach to studying the extremal-type convergence is to fit the distribution to some functional form (e.g. see [28,80] in which the extreme-value distributions are modeled as a generalized extreme-value distribution) or prove the convergence using one of the criteria given, for example, in [78,81]. Like [28,80], we find the convergence insensitive to the value of $f_{\rm NL}$.

We note that, contrary to the observation in [23], we found no combination of parameters which give rise to a Weibull distribution, which arises when there is a natural upper bound for the distribution function. Moreover, it is worth noting that if the PDFs such as those in Figs. 4 and 5 are well-approximated by "skew-symmetric" functions (e.g. an Edgeworth expansion) then the distribution cannot converge to the Weibull type as proven in [83].

Finally, we point out an interesting fact that if the coefficients in the expansion (46) conspire to make $P(x)$ an exactly Poissonian distribution

$$P(x, \lambda) = e^{-\lambda} \sum_{k=0}^{x} \frac{\lambda^k}{k!}, \tag{B2}$$

then the limiting distribution $G(x)$ will completely degenerate to $G = 1$ or 0. This is one of the rare examples where the extreme-value distribution does not converge to any of the three standard distributions. Of course, we do not expect a realistic distribution of galaxy clusters to be exactly Poissonian.

[9]Some applications of the Fréchet distribution to environmental sciences are summarized in [78,79].

[1] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, Phys. Rep. **402**, 103 (2004).
[2] J. M. Maldacena, J. High Energy Phys. 05 (2003) 013.
[3] G. I. Rigopoulos, E. P. S. Shellard, and B. J. W. van Tent, Phys. Rev. D **73**, 083522 (2006).
[4] C. T. Byrnes and K. Choi, Adv. Astron. **2010**, 724525 (2012).
[5] N. Bartolo, S. Matarrese, and A. Riotto, Phys. Rev. D **69**, 043503 (2004).
[6] M. Sasaki, J. Väliviita, and D. Wands, Phys. Rev. D **74**, 103003 (2006).
[7] K. A. Malik and D. H. Lyth, J. Cosmol. Astropart. Phys. 9 (2006) 008.
[8] E. Komatsu *et al.*, Astrophys. J. Suppl. Ser. **192**, 18 (2011).
[9] A. Cooray, D. Sarkar, and P. Serra, Phys. Rev. D **77**, 123006 (2008).
[10] B. Yu and T. Lu, Phys. Rev. D **78**, 063008 (2008).
[11] M. J. Jee *et al.*, Astrophys. J. **704**, 672 (2009).
[12] J. P. Stott *et al.*, Astrophys. J. **718**, 23 (2010).
[13] M. Brodwin *et al.*, Astrophys. J. **721**, 90 (2010).
[14] J. S. Santos *et al.*, Astron. Astrophys. **531**, L15 (2011).
[15] M. Tanaka, A. Finoguenov, and Y. Ueda, Astrophys. J. Lett. **716**, L152 (2010).
[16] K. Enqvist, S. Hotchkiss, and O. Taanila, J. Cosmol. Astropart. Phys. 4 (2011) 017.
[17] B. Hoyle, R. Jimenez, and L. Verde, Phys. Rev. D **83**, 103502 (2011).
[18] S. Hotchkiss, J. Cosmol. Astropart. Phys. 07 (2011) 18.
[19] L. Cayon, C. Gordon, and J. Silk, Mon. Not. R. Astron. Soc. **415**, 849 (2011).
[20] M. J. Mortonson, W. Hu, and D. Huterer, Phys. Rev. D **83**, 023015 (2011).
[21] S. D. M. White, Mon. Not. R. Astron. Soc. **186**, 145 (1979).
[22] O. Davis, J. Devriendt, S. Colombi, J. Silk, and C. Pichon, Mon. Not. R. Astron. Soc. **413**, 2087 (2011).
[23] S. Colombi, O. Davis, J. Devriendt, S. Prunet, and J. Silk, Mon. Not. R. Astron. Soc. **414**, 2436 (2011).
[24] J. P. Bernstein and S. P. Bhavsar, Mon. Not. R. Astron. Soc. **322**, 625 (2001).
[25] T. Antal, F. Sylos Labini, N. L. Vasilyev, and Y. V. Baryshev, Europhys. Lett. **88**, 59001 (2009).
[26] L. Dobos and I. Csabai, Mon. Not. R. Astron. Soc. **414**, 1862 (2011).
[27] S. P. Bhavsar and J. D. Barrow, Mon. Not. R. Astron. Soc. **213**, 857 (1985).
[28] G. Mikelsons, J. Silk, and J. Zuntz, Mon. Not. R. Astron. Soc. **400**, 898 (2009).
[29] R. K. Sheth and A. Diaferio, arXiv:1105.3378.
[30] J.-C. Waizmann, S. Ettori, and L. Moscardini, Mon. Not. R. Astron. Soc. **418**, 456 (2011).
[31] J.-C. Waizmann, S. Ettori, and L. Moscardini, Mon. Not. R. Astron. Soc. **420**, 1754 (2012).
[32] I. Harrison and P. Coles, arXiv:1111.1184.
[33] N. Dalal, O. Dore, D. Huterer, and A. Shirokov, Phys. Rev. D **77**, 123514 (2008).
[34] P. Valageas, Astron. Astrophys. **514**, A46 (2010).
[35] P. Valageas, Astron. Astrophys. **508**, 93 (2009).
[36] S. Chongchitnan and J. Silk, Astrophys. J. **724**, 285 (2010).
[37] S. Chongchitnan and J. Silk, Phys. Rev. D **83**, 083504 (2011).
[38] O. Lahav and A. R. Liddle, arXiv:1002.3488.
[39] S. M. Carroll, W. H. Press, and E. L. Turner, Annu. Rev. Astron. Astrophys. **30**, 499 (1992).
[40] O. Lahav, P. B. Lilje, J. R. Primack, and M. J. Rees, Mon. Not. R. Astron. Soc. **251**, 128 (1991).
[41] S. Weinberg, *Cosmology* (Oxford University, New York, 2008).
[42] D. J. Eisenstein and W. Hu, Astrophys. J. **496**, 605 (1998).
[43] S. Matarrese, L. Verde, and R. Jimenez, Astrophys. J. **541**, 10 (2000).
[44] M. Grossi, E. Branchini, K. Dolag, S. Matarrese, and L. Moscardin, Mon. Not. R. Astron. Soc. **390**, 438 (2008).
[45] C. Wagner and L. Verde, arXiv:1102.3229.
[46] A. Pillepich, C. Porciani, and O. Hahn, Mon. Not. R. Astron. Soc. **402**, 191 (2010).
[47] M. LoVerde, A. Miller, S. Shandera, and L. Verde, J. Cosmol. Astropart. Phys. 04 (2008) 014.
[48] A. Paranjape, C. Gordon, and S. Hotchkiss, Phys. Rev. D **84**, 023517 (2011).
[49] W. H. Press and P. Schechter, Astrophys. J. **187**, 425 (1974).
[50] R. K. Sheth and G. Tormen, Mon. Not. R. Astron. Soc. **308**, 119 (1999).
[51] J. Tinker *et al.*, Astrophys. J. **688**, 709 (2008).
[52] J. L. Tinker *et al.*, Astrophys. J. **724**, 878 (2010).
[53] A. Jenkins *et al.*, Mon. Not. R. Astron. Soc. **321**, 372 (2001).
[54] M. S. Warren, K. Abazajian, D. E. Holz, and L. Teodoro, Astrophys. J. **646**, 881 (2006).
[55] D. Reed *et al.*, Mon. Not. R. Astron. Soc. **346**, 565 (2003).
[56] S. Bhattacharya *et al.*, Astrophys. J. **732**, 122 (2011).
[57] S. Matarrese and L. Verde, Astrophys. J. **677**, L77 (2008).
[58] V. Desjacques, D. Jeong, and F. Schmidt, Phys. Rev. D **84**, 063512 (2011).
[59] F. Schmidt and M. Kamionkowski, Phys. Rev. D **82**, 103002 (2010).
[60] S. Shandera, N. Dalal, and D. Huterer, J. Cosmol. Astropart. Phys. 03 (2011) 017.
[61] J. N. Fry and E. Gaztanaga, Astrophys. J. **413**, 447 (1993).
[62] N. Kaiser, Astrophys. J. **284**, L9 (1984).
[63] C. Fedeli, C. Carbone, L. Moscardini, and A. Cimatti, Mon. Not. R. Astron. Soc. **414**, 1545 (2011).
[64] C. Carbone, O. Mena, and L. Verde, J. Cosmol. Astropart. Phys. 7 (2010) 020.
[65] C. Wagner, L. Verde, and L. Boubekeur, J. Cosmol. Astropart. Phys. 10 (2010) 022.
[66] L. Verde and S. Matarrese, Astrophys. J. **706**, L91 (2009).
[67] M. Manera and E. Gaztanaga, Mon. Not. R. Astron. Soc. **415**, 383 (2011).
[68] R. Juszkiewicz, F. R. Bouchet, and S. Colombi, Astrophys. J. **412**, L9 (1993).
[69] A. Cooray and R. Sheth, Phys. Rep. **372**, 1 (2002).
[70] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro, Phys. Rep. **367**, 1 (2002).
[71] T. Giannantonio and C. Porciani, Phys. Rev. D **81**, 063530 (2010).
[72] T. Matsubara, Phys. Rev. D **83**, 083518 (2011).

[73] J. N. Fry and R. J. Scherrer, Astrophys. J. **429**, 36 (1994).
[74] I. Kayo, A. Taruya, and Y. Suto, Astrophys. J. **561**, 22 (2001).
[75] J. Robinson, E. Gawiser, and J. Silk, arXiv:astro-ph/9805181.
[76] A. L. B. Ribeiro, C. M. Coelho, A. P. A. Andrade, and M. S. Dantas, Astron. Astrophys. **468**, 19 (2007).
[77] D. E. Holz and S. Perlmutter, arXiv:1004.5349.
[78] S. Kotz and S. Nadarajah, *Extreme Value Distributions: Theory and Applications* (Imperial College Press, London, 2000).
[79] R.-D. Reiss and M. Thomas, *Statistical Analysis of Extreme Values* (Birkhauser Verlag, Basel, 2007), 3rd ed..
[80] I. Harrison and P. Coles, Mon. Not. R. Astron. Soc. Lett. **418**, L20 (2011).
[81] M. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes* (Springer-Verlag, New York, 1982).
[82] E. Jondeau, S.-H. Poon, and M. Rockinger, *Financial Modeling Under Non-Gaussian Distributions* (Springer-Verlag, London, 2007).
[83] S.-M. Chang and M. G. Genton, Commun. Stat., Theory Methods **36**, 1705 (2007).