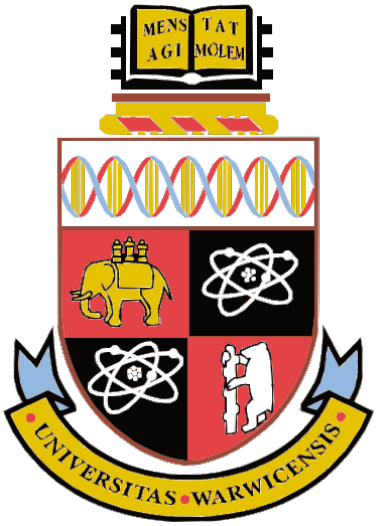


Applications of Extreme Value Statistics to massive galaxy clusters

by

Tom Swallow

MA4K9 Dissertation

Submitted to The University of Warwick

Mathematics Institute

April, 2022

Contents

Abstract

We will use the Generalised Extreme Value approach of Extreme Value Statistics to predict the mass of the largest galaxy cluster in the Universe. By considering a variety of mass functions, we will construct a model that will determine the distribution of cluster masses. We will then use the Generalised Extreme Value distribution to write down a formula for the tail of this distribution, and compare our results to several surveys in order to validate our prediction.

1 Introduction

"What is the most massive thing in the world?" — a question so simple, a child might ask it. One might expect this question had an equally simple answer. Let us consider some candidates for "most massive thing". The Earth is massive, the Solar system is more so. The Solar System is a small part of the Milky Way galaxy, which is itself also quite massive. Galaxies themselves are found in galaxy clusters. These are some of the most massive gravitationally bound objects we know of — collections of thousands of galaxies, all orbiting each other.

Though these clusters are incredibly large, space is vast, and so we are never going to be able to observe all of them. We may know the mass of the largest observed cluster, but what if there exists a greater one, just out of sight, waiting to be found. Indeed, we may never observe them all, so how do we answer our question?

Consider this example — imagine that we have observed thousands upon thousands of galaxy clusters, and the top 50 largest ones all have a very similar mass. We may feel justified in saying that, though we have probably not actually seen the most massive cluster in the Universe, there seems to be some sort of limit on how massive it could be. It would be peculiar if we suddenly found a cluster that was one hundred times as massive as our top 50, so we may expect that the largest cluster in existence is probably just a bit more massive than our observed largest.

This paper will aim to capture this notion. We will construct a model that will allow us to plot the distribution of cluster masses in a given volume. Next, we will apply the theory of Extreme Value Statistics in order to model the tail of this distribution — that is, we will be able to gain insight into the distribution of these most massive clusters. Finally, we will compare this to observed data, and hopefully find a reasonable answer to our question.

2 Preliminaries

2.1 Assumptions

In order to start the construction of our model of the Universe, it is necessary to make some assumptions. In particular, we will be relying on the Cosmological Principle. That

is, the Universe has two important properties:

1. Homogeneity - On large enough scales, the Universe looks the same at every point, and there is no "preferred location".
2. Isotropy - On large enough scales, the Universe looks the same in every direction, and there is no "preferred orientation".

Upon first viewing, these may seem like very strong and unrealistic assumptions. After all, in my current position and orientation I am looking at my computer. However, if I turned around and moved a few metres to my left, I'd be floating in mid-air above a busy A road. These are very different circumstances. On a slightly larger scale, if I was somewhere in the solar system, in one direction I would be blinded by the sun, and in the other I'd be looking into the darkness of space.

It is only on the largest of scales where these assumptions start to become appropriate. For example, the Cosmic Microwave Background is very nearly isotropic, with deviations of about one part in one hundred thousand. Though there is no concrete cut-off where these assumptions suddenly become applicable, they seem to hold on scales above 100 Mpc [26], which makes them useful for our study here.

Finally, we will be assuming that the Universe is flat, and so has zero curvature. This will simplify our equations, as we can ignore Ω_k terms, and is backed up in the literature [3].

2.2 Units

Due to the subject of study in this project, we will be dealing with incomprehensibly vast distances and unfathomably large masses. We will therefore measure the mass of galaxy clusters in terms of solar masses, denoted $M_\odot$. We will rarely be considering masses under $10^{14} M_\odot$.

Distances are typically measured in megaparsecs - that is, 1 million parsecs. However, using Hubble's law, we can rewrite the formula for redshift, denoted by z, as

$$z = \frac{v}{c} = \frac{H_0 r}{c}. \tag{1}$$

Hence, we can use a galaxy cluster's redshift as a way to measure its distance, as these two quantities are proportional.

This calculation is made even simpler if we take $c = 1$. Factors of c appear frequently in our model, so having it be a dimensionless unit is a convenient re-scaling.

Finally, we have used H_0 to denote the current value of the Hubble parameter. We will parameterise the Hubble constant as $H_0 = 100h$ km s^{-1}Mpc^{-1}, where $h = 0.72$ [13, p. 46]. As the only way to measure the distance to distant galaxy clusters is through their

recession velocity, and this value depends on the observed value of h, it is common to write distances with this quantity built-in. For example, later on we will consider the distance $8h^{-1}$ Mpc, where the value 8 is exact, but the h^{-1} is not.

2.3 Symbols and Constants

Symbol	Description
z	Redshift
V	Volume
R	Distance
M	Mass
f_{sky}	Proportion of sky observed

Table 1: A key to all symbols used in this paper, and their values where appropriate

Constant	Description	Value (units)
$M_\odot$	Solar mass	1.988×10^{30} kg
H_0	Hubble constant	$100h$ km s^{-1}Mpc^{-1}
h	Parameterisation of H_0	0.72
Ω_c	Density parameter of cold dark matter	0.23
Ω_b	Density parameter of baryons	0.046
Ω_r	Density parameter of radiation	8.6×10^{-5}
Ω_m	Density parameter of all matter	$\Omega_c + \Omega_b$
Ω_Λ	Density parameter of dark energy	$1 - \Omega_m - \Omega_r$
ρ_{crit}	Critical density	$3H_0^2/8\pi G$
ρ_m	Time-dependant energy density of all matter	$\Omega_m \times \rho_{crit}$
n_s	Scalar Spectral Index	0.96

Table 2: A key to all constants used in this paper, and their values where appropriate

3 Model

3.1 Computing the Mass Dispersion

We will begin by briefly summarising some important results from cosmological perturbation theory. For a more detailed review of these topics, see [18] and [12].

Firstly, we need to capture the notion of homogeneity and isotropy. We do this by defining the Friedmann-Robertson-Walker (FRW) metric, as found in [16],

$$d\ell^2 = a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right] \tag{2}$$

where $a(t)$ denotes the scale factor, $d\ell$ is the distance between two points, and k is the constant curvature of this universe. As we are assuming flatness, this can be written more simply as

$$d\ell^2 = a^2(t)\left[dr^2 + r^2(d\theta^2 + \sin^2\theta\ d\phi^2)\right] \tag{3}$$

Now, if we let $\rho_m(\mathbf{x}, z)$ denote the matter energy density at a certain position in space and redshift, and using $\langle\rho_m\rangle$ to denote the mean of this value, we can define the density fluctuation field, $\delta(\mathbf{x}, z)$, as

$$\delta(\mathbf{x}, z) = \frac{\rho_m(\mathbf{x}, z) - \langle\rho_m(z)\rangle}{\langle\rho_m(z)\rangle}. \tag{4}$$

To aid readability, we shall write $\delta(\mathbf{x})$ in place of $\delta(\mathbf{x}, z)$. We can write down the Fourier decomposition of this function, as in [6], to yield

$$\delta(\mathbf{x}) = \int d\mathbf{k}\ \delta(\mathbf{k}) \exp(i\mathbf{k}\cdot\mathbf{x}) \tag{5}$$

It is now our aim to find an alternative parameterisation for this δ — we do this by perturbing the FRW metric. As shown in [18, p. 217], the Newtonian gravitational potential Φ gives us the amplitudes of linear perturbations in this metric. We can then use the cosmological Poisson equation, as in [16], to write

$$\delta(\mathbf{k}) = \mathcal{A}(k, z)\Phi(\mathbf{k}) \tag{6}$$

where

$$\mathcal{A}(k, z) = \frac{2}{3\Omega_m}\left(\frac{k}{H_0}\right)^2 T(k_{\mathrm{EH}})D(z). \tag{7}$$

Here, we see the transfer function T, and the linear growth function D. These two functions summarise the evolution of these perturbations from their initial amplitudes as given by Φ, to their current values as given by δ.

Instead of using the actual transfer function $T(k)$, we shall be using a fitted formula, as approximated by Weinberg [27]:

$$T(k) = \frac{\ln(1 + (0.124k)^2)}{(0.124k)^2}\left(\frac{1 + (1.257k)^2 + (0.4452k)^4 + (0.2197k)^6}{1 + (1.606k)^2 + (0.8568k)^4 + (0.3927k)^6}\right)^{1/2} \tag{8}$$

This approximation leads to an error of less that 2% [27, p. 307], so is appropriate to use here.

Note that we do not evaluate the transfer function at k, but instead at the point k_{EH}, as we incorporate the baryonic correction as described by Eisenstein and Hu [10, p. 612]. Here, we find the paramaterisation

$$k_{\mathrm{EH}} = \frac{k\Omega_r^{1/2}}{H_0\Omega_m}\left(\alpha + \frac{1-\alpha}{1+(0.43ks)^4}\right)^{-1} \tag{9}$$

where

$$\alpha = 1 - 0.328\ln(431\Omega_m h^2)\frac{\Omega_b}{\Omega_m} + 0.38\ln(22.3\Omega_m h^2)\left(\frac{\Omega_b}{\Omega_m}\right)^2 \tag{10}$$

and

$$s = \frac{44.5\ln(9.83/\Omega_m h^2)}{\sqrt{1+10(\Omega_b h^2)^{3/4}}}\ \mathrm{Mpc}. \tag{11}$$

Finally, we consider the linear growth factor $D(z)$, as calculated by Lyth and Liddle [15]. We see that, if we let

$$\Omega_m(z) = \Omega_m\frac{(1+z)^3}{1-\Omega_m+(1+z)^3\Omega_m} =: \Omega_{m,z} \tag{12}$$

as detailed on [15, p. 64], then we can use this to compute this growth factor, by way of

$$D(z) = \frac{5}{2}\Omega_{m,z}\left(\frac{1}{70} + \frac{209\Omega_{m,z}}{140} - \frac{\Omega_{m,z}^2}{140} + \Omega_{m,z}^{4/7}\right)^{-1}. \tag{13}$$

This formula is described on [15, p. 110], where Lyth and Liddle have also computed the present value of this growth factor ($D(0) \approx 0.76$). The fact that a specific value of this function was given was convenient, as it allowed me to check that our model was producing the results expected.

Now that have an expression for δ, we can define the Power Spectrum, $P(k)$, using the two-point correlation of δ, as found on [9, p. 16]

$$\langle\delta(\mathbf{k}_1), \delta(\mathbf{k}_2)\rangle = \delta_D(\mathbf{k}_1 - \mathbf{k}_2)P(k), \tag{14}$$

where δ_D is the 3-dimensional Dirac delta function. In [9], we note that in this formulation, $P(k)$ has dimensions of k^{-3}, and hence a dimensionless quantity $\mathcal{P}(k)$ is often defined as $\mathcal{P}(k) = k^3P(k)$, whilst [6] reminds us that in linear perturbation theory, it is often useful to assume that $P(k)$ is proportional to k^{n_s}, with n_s being the spectral index. The results found on [4, p. 9] give a tight constraint of $n_s = 0.96$.

Following the derivation found on [16, p. 197], we are then able to write

$$\mathcal{P}(k) \propto (\mathcal{A}(k,z))^2 \left(\frac{k}{H_0}\right)^{n_s-1} \tag{15}$$

We are now able to compute the smoothed variance of linear density fluctuations on a scale R, as given by the formula

$$\sigma_R^2 = 4\pi \int_0^\infty \frac{dk}{k} (W(kR))^2 \mathcal{P}(k). \tag{16}$$

Here, we choose w to be the spherical top-hat function of radius R, given by

$$w_R(x) = \begin{cases} \frac{3}{4\pi R^3} & x < R \\ 0 & x \geq R \end{cases}, \tag{17}$$

and then define W to be the Fourier transform of this function, as found in [14]. That is,

$$W(kR) = 3\left(\frac{\sin(kR)}{(kR)^3} - \frac{\cos(kR)}{(kR)^2}\right). \tag{18}$$

Combining all this, we are able to compute different values of σ_R. However, we must remember that $\mathcal{P}(k)$ was only proportional to $(\mathcal{A}(k,z))^2(k/H_0)^{n_s-1}$, and so we may be off by some constant factor. We can avoid this issue, as σ_R is known for certain values of R. In particular, the mass dispersion on a scale of $8h^{-1}$ Mpc is given in [6] - we will take $\sigma_8 = 0.801$. This allows us to compute the normalising constant required for our model.

Finally, it is often required to express σ_R as σ_M, for some mass M. These are equivalent, as we can relate mass to radius via the formula found in [6], as the mass of matter enclosed in a top-hat window of radius R is given by:

$$M = \frac{4}{3}\pi R^3 \rho_m \approx 1.16 \times 10^{12} \Omega_m \left(\frac{R}{h^{-1}\text{Mpc}}\right)^3 h^{-1} M_\odot \tag{19}$$

3.2 Computing the Mean Number Density

Now that we are able to compute σ_R, our next goal is to use this to derive the mean number density of clusters above a certain mass. That is, we are seeking a function $n(>m,z)$ that can calculate the number of objects of mass larger than m, and at a given redshift z.

Firstly, we define the critical overdensity δ_c, which is assumed to be $\sqrt{a} \times 1.686$, where $\sqrt{a} = 0.9$ is a "fudge factor", as described in [25].

Now we have done this, we are able to write down an expression for the differential number density, as found in [20]:

$$\frac{dn}{dM} = F(\nu)\frac{\rho_m}{M}\frac{d\ln\sigma_M^{-1}}{dM}. \tag{20}$$

Here, $\nu = \delta_c/\sigma_M$, and $F(\nu)$ is mass function. In this paper, we will be considering three standard mass functions.

The first is the Press-Schechter mass function, also described in [20]

$$F_{PS}(\nu) = \sqrt{\frac{2}{\pi}}\nu\exp(-\nu^2/2), \tag{21}$$

whilst the second is the Sheth-Tormen mass function, whose statement can be found in [22]

$$F_{ST}(\nu) = 0.322\sqrt{\frac{2a}{\pi}}\nu\exp\left(-\frac{a\nu^2}{2}\right)\left(1+(a\nu^2)^{-0.3}\right), \quad a = 0.707. \tag{22}$$

Finally, we have the Tinker et al. model, which is a more fitted equation, and tends to lie between the two previous estimates, as we will see.

$$F_{Tinker}(\nu) = 0.368\left(1+(\beta\nu)^{-2\phi}\right)\nu^{2\eta+1}\exp\left(-\frac{\gamma\nu^2}{2}\right) \tag{23}$$

$$\beta = 0.589(1+z)^{0.2}, \quad \phi = -0.729(1+z)^{-0.08}, \tag{24}$$

$$\eta = -0.243(1+z)^{0.27}, \quad \gamma = 0.864(1+z)^{-0.01}. \tag{25}$$

Once we fix any of these mass functions, we are then able to compute the differential number density $\frac{dn}{dM}$ using equation 20. Computing the mean number density is then just a matter of integrating:

$$n(>m,z) = \int_m^\infty \frac{dn}{dM}\, dM. \tag{26}$$

Now we have a density, computing the total number count is done by simply multiplying by a volume element. We find such an element expressed in [6],

$$\frac{dV}{dz} = f_{sky}\frac{4\pi}{H(z)}\left(\int_0^z \frac{d\tau}{H(\tau)}\right)^2, \tag{27}$$

where f_{sky} is the proportion of the sky we are surveying. While in practice we would only observe some portion of the entire sky, we are trying to find the most massive cluster in the whole Universe, and so f_{sky} will be equal to 1 for the duration of this paper. As we are assuming a flat Universe, we can approximate $H(z)$ — the Hubble parameter — as in [19, p. 100]

$$H(z) = H_0 \left(\Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda\right)^{1/2} = H_0 \left(\Omega_m(1+z)^3 + \Omega_\Lambda\right)^{1/2}, \qquad (28)$$

since we are assuming the curvature parameter $\Omega_k = 0$.

Putting all this together, we have all the necessary formulae to define the number count

$$N(> m, z) = \frac{dV}{dz} \times \int_m^\infty \frac{dn}{dM}\, dM, \qquad (29)$$

as we required.

3.3 Running the Model

Now that we have created our model, we are in a position to start generating data from it. We would like to determine how many clusters above a given mass there are in different redshift bins. That is, we will be computing $N(> m, z)$ for masses of the form $M = 10^k M_\odot$ (for $k = 12, 13, 14, 15, 16$). Our first redshift bin will be all those clusters in the range $0 \leq z < 0.1$, our second will be $0.1 \leq z < 0.2$, and so on.

Computing all these values, we are able to produce Figure 1, with the y axis showing the number count of clusters with the specified mass and redshift. Note that, in this instance, we have chosen to use the Press-Schechter mass function, as defined in equation (21).

As expected, increasing the mass of the cluster leads to less clusters being observed. In fact, the expected number of clusters of mass $10^{16} M_\odot$ is less than 1 at all redshifts, suggesting that we do not expect to see any of this size at all.

For each mass, the number count increases with redshift until we reach a maximum count. From this point on, further increasing redshift leads to less clusters of the given mass being observed. Increasing the mass appears to make this peak count happen at lower redshifts. This implies that, if we found the largest mass that had a number count larger than 1 at any redshift, the redshift where this occurred would be small, and so this most massive cluster would be relatively nearby.

3.4 A note on the Code

Though this paper mainly contains the theory underpinning the project, a majority of my hours were dedicated to the creation of the scripts required to define and run the model. I was faced with implementation issues that, whilst being necessary to overcome, are not immediately obvious from the equations defined above.

The first major revision came from the sheer size of the numbers being worked with. 10^{15} is a large number [1], and as such is computationally time consuming to deal with, as well as being prone to rounding and overflow errors. It became convenient to work with the

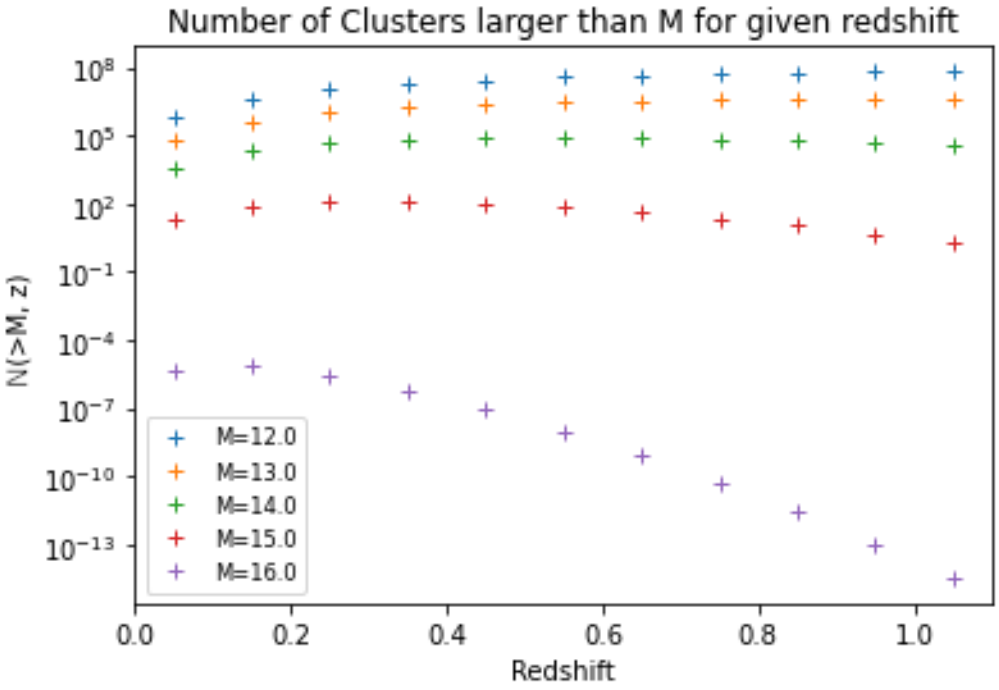


Figure 1: Number count for different masses and redshifts

substitution $u = \log_{10}(m)$, where m is the "true" mass. Because of the chain rule, this lead to factors of dm/du appearing in equations like (20), as I was actually wanted to compute dn/du, not dn/dm as stated. It became important to stay on top of which quantities were in this log space and which weren't, as well as being able to convert between these easily.

I was concerned about certain functions becoming inaccurate for particularly small values. In particular, I noticed that the top-hat function (18) was unstable for small values of k. I solved this by Taylor expanding , and using this approximation if k fell below a certain pre-defined threshold. In particular, I wrote:

$$\begin{aligned}
W(kR) &= \frac{3}{(kR)^3}\left(\sin(kR) - kR\cos(kR)\right) \\
&\approx \frac{3}{(kR)^3}\left[kR - \frac{(kR)^3}{3!} + \frac{(kR)^5}{5!} - \frac{(kR)^7}{7!} + \frac{(kR)^9}{9!} - \frac{(kR)^{11}}{11!}\right. \\
&\qquad \left. -kR\left(1 - \frac{(kR)^2}{2!} + \frac{(kR)^4}{4!} - \frac{(kR)^6}{6!} + \frac{(kR)^8}{8!} - \frac{(kR)^{10}}{10!}\right)\right] \\
&= \frac{3}{(kR)^3}\left[(kR)^3\left(\frac{1}{2!} - \frac{1}{3!}\right) + (kR)^5\left(\frac{1}{5!} - \frac{1}{4!}\right) + (kR)^7\left(\frac{1}{6!} - \frac{1}{7!}\right)\right. \\
&\qquad \left. + (kR)^9\left(\frac{1}{9!} - \frac{1}{8!}\right) + (kR)^{11}\left(\frac{1}{10!} - \frac{1}{11!}\right)\right] \\
&= 1 - \frac{1}{10}(kR)^2 + \frac{1}{280}(kR)^4 - \frac{1}{15120}(kR)^6 + \frac{1}{1330560}(kR)^8.
\end{aligned}$$

This improved stability without noticeably increasing computation time.

Finally, I noticed that the integral defined in (16) sometimes failed to converge. To solve this problem, I split the integral into two parts:

$$\sigma_R^2 = 4\pi \int_0^\infty \frac{dk}{k}(W(kR))^2\mathcal{P}(k) = 4\pi \int_0^1 \frac{dk}{k}(W(kR))^2\mathcal{P}(k) + 4\pi \int_1^\infty \frac{dk}{k}(W(kR))^2\mathcal{P}(k), \tag{30}$$

and then used the substitution $u = 1/k$ in order to write this at the sum of two integrals from 1 to infinity. This was to avoid small values of k leading to numerical imprecision. In particular, we have that

$$4\pi \int_0^1 \frac{dk}{k}(W(kR))^2\mathcal{P}(k) = 4\pi \int_\infty^1 \frac{dk}{1/k}(W(R/k))^2\mathcal{P}(1/k) \times \frac{-1}{k^2} \tag{31}$$

$$= 4\pi \int_1^\infty \frac{dk}{k}(W(R/k))^2\mathcal{P}(1/k) \tag{32}$$

and so we are able to write σ_R^2 as one integral from 1 to infinity

$$\sigma_R^2 = 4\pi \int_1^\infty \frac{1}{k}(W(kR))^2\mathcal{P}(k) + \frac{1}{k}(W(R/k))^2\mathcal{P}(1/k)\ dk. \tag{33}$$

This solved the numerical issues, without leading to noticeable errors or runtime increases.

A link to the code, as well as a short note on all the outputs, routines and runtime, can be found in the Appendix 7.1.

4 Extreme Value Statistics

4.1 Extreme Value Theorem

Now that we have set up our model, we shall enlist the help of Extreme Value Statistics (EVS) in order to quantify the rarity of the largest galaxy clusters. In particular, we will be applying the Generalised Extreme Value (GEV) approach (also known as the Block Maxima approach).

The main idea behind this approach is to split up your observations into N "blocks" of a set size, and then normalise appropriately in order to achieve one of three distributions as N tends to infinity: the Gumbel, Fréchet or Weibull distributions. In our case, these blocks will correspond to the red-shift bins we outlined in Section 3.3.

In particular, we will be relying on the Extreme Value Theorem of statistics:

Theorem 4.1 (Fisher-Tippett-Gnedenko)**.** *The class of extreme value distributions is*

$G_\gamma(ax+b)$, *with* $a>0,\ b\in\mathbb{R}$, *where*

$$G_\gamma(x)=\exp\left(-(1+\gamma x)^{-1/\gamma}\right),\qquad 1+\gamma x>0, \tag{34}$$

where $\gamma\in\mathbb{R}$ *and where* $\gamma=0$ *on the right hand side is to be interpreted as* $\exp\left(-e^{-x}\right)$.

In particular,

1. $\gamma=0$ corresponds to the Gumbel distribution,
2. $\gamma>0$ corresponds to the Fréchet distribution,
3. $\gamma<0$ corresponds to the Weibull distribution.

We will be using this theorem without proof, though De Haan and Ferreira give an easy to follow account on [8, p. 7].

Following the approach laid out in [7], we first note the general form of the GEV distribution:

$$G_\gamma(y)=\exp\left(-(1+\gamma y)^{-1/\gamma}\right)\Leftrightarrow -\ln G_\gamma(y)=(1+\gamma y)^{-1/\gamma} \tag{35}$$

where $y=(u-\alpha)/\beta$, assuming that $\gamma\neq 0$. Here, γ is referred to as the shape parameter, α the location or position parameter, and β the scale parameter. Our goal now is to determine these constants.

As before, let $n(>m)$ be the mass function (i.e. the number density of clusters at least as big as m), and let V be a volume element. We make use of the substitution $u=\log_{10}(m)$ for ease, as introduced in Section 3.4, with u_{max} denoting the mass of the largest cluster within our volume element.

Let $p_G(u_{max})$ be the probability distribution function of the values taken by u_{max}. In other words, $p_G(u_{max})$ is the probability that the largest cluster in our volume element is of size u_{max}. Then, the cumulative distribution function is given by

$$\mathbb{P}(u_{max}\leq u)=\int_0^u p_G(u_{max})du_{max} \tag{36}$$

If we introduce the notation $P_0(u)$ to be the probability that the volume we are looking at has zero clusters above mass u, we see that these probabilities coincide, so

$$\mathbb{P}(u_{max}\leq u)=P_0(u). \tag{37}$$

This would allow us to determine $p_G(u)$, as we would then have that

$$p_G(u) = \frac{dP_0}{du} \tag{38}$$

However, on large enough scales such that we can apply the cosmological principle ([26] suggests scales larger than 100 h^{-1}Mpc), we can ignore clustering between galaxy clusters, as well as any clusters straddling the boundary of our volume. This would allow us to instead write P_0 as a Poisson distribution

$$P_0(u) = \exp(-n(> u)V). \tag{39}$$

Now, assuming that we can in fact model P_0 by G, for some collection of parameters, we can Taylor expand both functions about the peaks of the underlying PDFs — denoted by u_0 — as such:

$$P_0(u) = P_0(u_0) + (u - u_0)\left.\frac{dP_0(u)}{du}\right|_{u_0} + \dots \tag{40}$$

$$G(u) = G(u_0) + (u - u_0)\left.\frac{dG(u)}{du}\right|_{u_0} + \dots \tag{41}$$

If we carry out this expansion, these first two terms are actually enough to determine γ, α and β just through comparing coefficients. In particular, we have that

$$\gamma = n(> m_0)V - 1, \quad \beta = \frac{(1+\gamma)^{(1+\gamma)}}{\left.\frac{dn}{dm}\right|_{m_0} V m_0 \ln 10}, \quad \alpha = \log_{10} m_0 - \frac{\beta}{\gamma}((1+\gamma)^{-\gamma} - 1) \tag{42}$$

where m_0 denotes is the mass such that $dP_0/du = \ln(10)\ m\ p_G(m)$ peaks. Therefore, to determine these parameters, we need to compute m_0.

4.2 Application of EVS to our Model

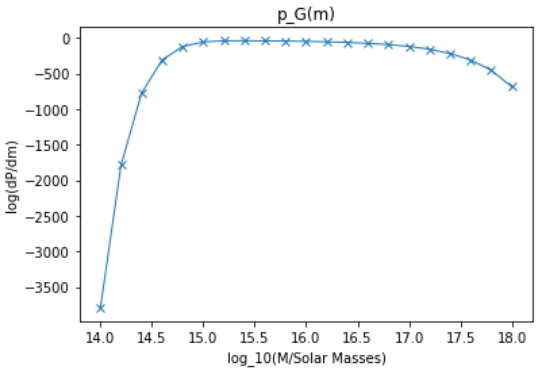
We do this by considering the following:

$$P_0(m) = \exp\left(-n(> m)V\right) \Rightarrow \frac{dP_0(m)}{dm} = -\exp\left(-n(> m)V\right)\frac{dn(m)}{dm}V \tag{43}$$

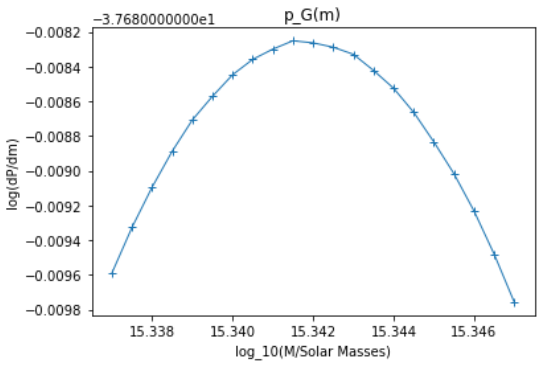
$$\log\left[\frac{dP_0(m)}{dm}\right] = -\log V + \log\frac{dn(m)}{dm} - n(> m)V \tag{44}$$

As the model already contains functions to simulate $\frac{dn}{dm}$ and $n(> m)$, we can simply plot this function without too much more work. Log is monotonic and so preserves the location

of peaks, so we can simply read off u_0. Figure 2 contains two example outputs, at two different scales, which allow us to obtain the value $u_0 = 15.325$.



(a) A plot of p_G over a range of masses. The shape indicates that this quantity should peak for some value m_0



(b) The same plot, zoomed in on the peak. This enables us to read off the value that maximises this distribution.

Figure 2: Numerically approximating m_0

We then use this to numerically compute γ, α and β. Note that all these parameters depend solely on γ, so γ actually defines the distribution entirely. We see that $\gamma = -0.11$, $\beta = 0.003$ and $\alpha = 15.325$.

This justifies the assumption we made earlier in equation (35) that $\gamma \neq 0$. We also see that α, the location parameter, is very close to our m_0, which is to be expected, as it was chosen to be the value that caused the probability distribution p_G to peak.

The fact that $\gamma < 0$ means that the limiting CDF of our normalised block maxima is a member of the Weibull (Type III) family of extreme value distributions.

I repeated this process for each redshift bin — I would fix my value of z, determine the mass that caused p_G to peak, and then use the formulas stated in equation (42) to determine the parameters describing the corresponding distribution. The exact values can be found in Table 3.

As we predicted, m_0 appears to be peaking at lower redshifts, meaning that this is where we expect the larger masses to be. However, we will need to plot all of these distributions to confirm this.

All the γ values are less than zero, meaning that all the extreme distributions are Weibull. Had any of these been non-negative, the distribution would have changed type — this would have dramatically affected the mass of the largest predicted cluster in that bin. Instead of this, we see m_0 varying relatively smoothly, with no sharp jumps or discontinuities. This is good, as the bin width was chosen arbitrarily, and rapid variation would require a much finer grid to truly capture the dynamics.

Redshift bin	m_0 ($10^{15} M_\odot$)	γ	β (10^{-3})	α
$0 \leq z < 0.1$	2.198	-0.109	2.55	15.342
$0.1 \leq z < 0.2$	2.698	-0.098	2.10	15.431
$0.2 \leq z < 0.3$	2.698	-0.087	1.93	15.431
$0.3 \leq z < 0.4$	2.545	-0.085	1.84	15.404
$0.4 \leq z < 0.5$	2.312	-0.080	1.79	15.364
$0.5 \leq z < 0.6$	2.075	-0.076	1.76	15.317
$0.6 \leq z < 0.7$	1.803	-0.089	1.76	15.256
$0.7 \leq z < 0.8$	1.633	-0.083	1.72	15.213
$0.8 \leq z < 0.9$	1.435	-0.068	1.72	15.157
$0.9 \leq z < 1$	1.264	-0.082	1.71	15.102
$1 \leq z < 1.1$	1.109	-0.069	1.71	15.045
$1.1 \leq z < 1.2$	0.975	-0.073	1.72	14.989

Table 3: The parameters describing the extreme distribution for each redshift bin

5 Comparison with Data

5.1 Data Sources and Analysis

We now have the required information to parameterise the extreme distribution of clusters in each bin, which means that we can now directly compute the mass of the largest expected cluster in each interval. However, it is not immediately obvious how we would define the mass of a real-world galaxy cluster. After all, there is no clearly defined boundary, where you can say the cluster definitely ends.

One approach is to compare the density of a region you think might be in the cluster to the density of "empty" space. If the density of this region is above a certain threshold, the region can be declared as part of the cluster. This is a definition still open to interpretation though — what threshold do you pick?

For simplicity, consider a spatially flat universe consisting only of non-relativistic matter, with no radiation or cosmological constant to speak of. Such a universe is called an Einstein–de Sitter universe (fully described in [21, p. 80]), and is an appropriate case study here, as such a universe is necessarily matter-dominated.

In such a model, the critical overdensity at virialisation (as a proportion of the background density) is equal to $\Delta_c = 18\pi^2 \approx 178$, where virialisation is the point where the cluster has collapsed under its own gravity enough to reach a dynamic equilibrium [28]. As we are making approximations anyway, we might as well round up to $\Delta_c = 200$. Hence, M_{200} is used to denote the mass of a galaxy cluster.

Though we have seen the theory backing up this choice, it was still slightly arbitrary. In some surveys, alternatives such as M_{500} or even M_{1000} are used. I was unable to find any logic behind behind these choices — I imagine they doubled to get 400, and then rounded up to 500. We see that, for any given cluster, $M_{500} \leq M_{200}$, as the M_{200} value will include

all the matter covered by M_{500}, but will also contain the mass found in lower density areas.

For instance, the Planck Survey [2] was one such experiment that used M_{500} as its way of defining the mass of its observed galaxy clusters. In this survey, they measured the mass and redshift of thousands of galaxy clusters — I went through the data they collected, and found the largest galaxy cluster in each of our redshift bins, as we defined earlier in Section 3.3. Table 4 contains a summary of this data. Note that if a cluster has mass M, an upper error of E and a lower error of e, this means that our 68% confidence interval is $(M - e, M + E)$.

Redshift bin	Name	M_{500} $(10^{14}M_{\odot})$	Upper error $(10^{14}M_{\odot})$	Lower error $(10^{14}M_{\odot})$
$0 \leq z < 0.1$	G044.20	8.771	0.186	0.239
$0.1 \leq z < 0.2$	G149.75	8.859	0.323	0.320
$0.2 \leq z < 0.3$	G006.76	16.116	0.297	0.292
$0.3 \leq z < 0.4$	G286.98	14.693	0.3917	0.417
$0.4 \leq z < 0.5$	G205.93	12.250	0.525	0.55
$0.5 \leq z < 0.6$	G180.25	11.487	0.535	0.548
$0.6 \leq z < 0.7$	G209.79	10.727	0.630	0.664
$0.7 \leq z < 0.8$	G138.61	9.481	0.669	0.530
$0.8 \leq z < 0.9$	G297.97	10.754	0.478	0.472
$0.9 \leq z < 1$	G266.54	6.774	0.487	0.535

Table 4: The largest cluster in each redshift bin, as observed by the Planck survey [2]

Another source of data I used was compiled by Harrison and Coles, and can be found in [11]. I have reproduced the most relevant values in Table 5, with the upper and lower errors to be read as in Table 4.

Name	Redshift	Mass $(10^{14}M_{\odot})$	Upper error $(10^{14}M_{\odot})$	Lower error $(10^{14}M_{\odot})$
A2163	0.203	30.4	8.7	6.7
A370	0.375	26.2	8.7	6.7
RXJ1347	0.451	21.4	6.0	4.8
ACT-CL J0102	0.87	18.5	4.2	3.3
PLCK G266	0.94	14.5	2.7	2.0
SPT-CL J2106	1.132	11.1	2.4	2.0
SPT-CL J0546	1.067	7.80	1.27	0.90
XXMU J2235	1.4	6.82	1.52	1.23
XXMU J0044	1.579	4.02	0.88	0.73

Table 5: A collection of large galaxy clusters, and their redshifts, as compiled by Harrison and Coles [11]

The immediate difference we notice is that the clusters in Table 5 are much more massive than in the previous table. This is to be expected, as the second table contains the largest galaxies encountered over multiple surveys, including the Planck survey itself. Additionally, whilst the Planck survey used M_{500}, Harrison and Coles considered M_{200},

which we established earlier this section could only increase a galaxy cluster's mass.

This mass discrepancy can also be explained by the fact that Harrison and Coles followed the lead of Waizmann, Ettori, and Moscardini in [26], and Mortonson, Hu, and Huterer in [17] by correcting for Eddington bias. That is, our mass functions as laid out in Section 3.2 are very steep at large masses, and shallower on the lower mass end. This means that it is more likely for lower-mass clusters to scatter up than it would be for higher-mass clusters to scatter down, leading to a systematic bias in the data. This has been corrected for in Table 5, as the masses presented there are actually m_{Edd}, related to the observed mass m by the equation

$$\ln m_{Edd} = \ln m + \frac{1}{2}\varepsilon\sigma^2_{\ln m}, \tag{45}$$

where ε is the local slope of the mass function, and $\sigma^2_{\ln m}$ is the uncertainty in the cluster's mass.

We also notice that the error bounds found in Table 5 are quite large, and much larger than those given by the Planck survey. This could be due to the fact that the observed masses are larger, or as a side effect of the bias correction. It is also worth noting that some of these observations were made years before the Planck survey took place, so could have been using less precise equipment.

5.2 Comparison of Mass Functions

Now that we have a good selection of data , we can start generating predictions from our model. Recall that, using Table 3, we are able to describe a different Weibull distribution for each redshift bin. I plotted these distributions, and computed their peak — remember, the mode of this distribution is the most likely maximum cluster mass in this redshift range.

I did this for both the Press-Schechter mass function, as found in equation (21), and also the Sheth-Tormen mass function, which is located in equation (22). I then used the in-built Cubic Spline function to interpolate between these points. This allowed me to create the graph found in Figure 3.

Both functions predict very similar masses for the smallest redshifts. However, as we look at further distances, Sheth-Tormen starts to predict higher masses than Press-Schechter. The discrepancy between these two functions increases, until both peak just below $z = 0.3$. From this point onward, both functions decrease at a similar rate, such that the average difference between them remains roughly constant. This holds up to the highest redshifts I modeled here.

Both functions peak in the $0.2 \leq z < 0.3$ bin, so this is where we the model suggests we

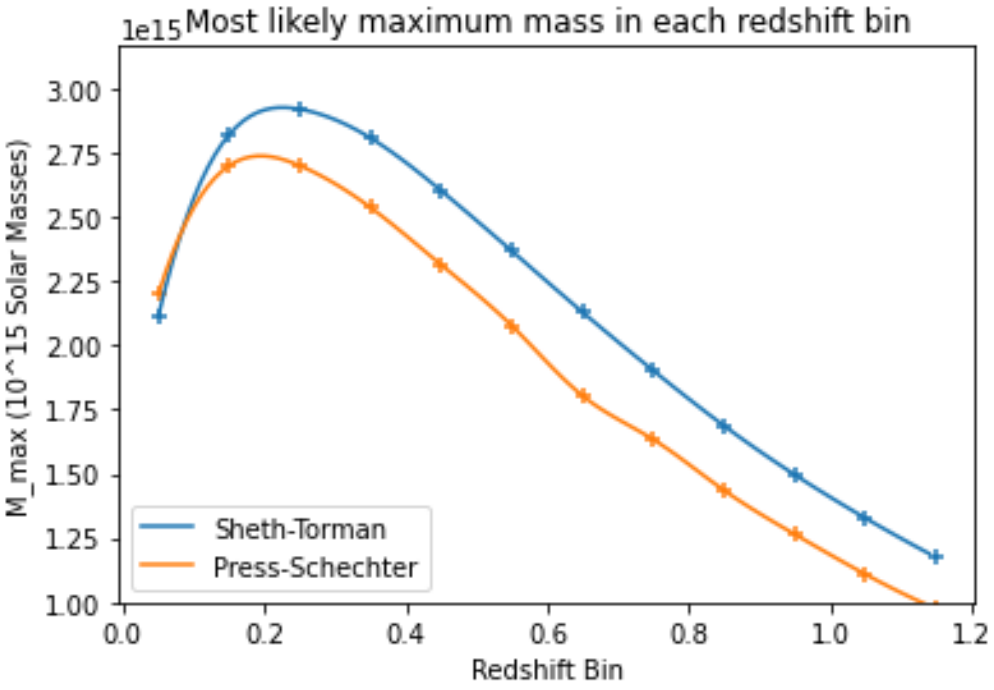


Figure 3: The most likely mass of the largest cluster in each redshift bin

should concentrate our efforts if we wanted to find the largest galaxy cluster in the whole Universe. Though I did not plot the model proposed by Tinker et al., as their function was fitted to observed data, it would likely lie somewhere between thee two functions.

At first glance, Figure 3 implies that the most massive galaxy cluster ought to have mass just below $3 \times 10^{15} M_{\odot}$. However, we can do better by considering intervals around the peaks of these Weibull distributions. To do this, we will need to derive the inverse of the cumulative distribution of the Weibull function. This can be done fairly easily analytically:

$$y = \exp\left(-\left(1+\gamma\frac{x-\alpha}{\beta}\right)^{-1/\gamma}\right) \tag{46}$$

$$\Leftrightarrow \quad \log y = -\left(1+\gamma\frac{x-\alpha}{\beta}\right)^{-1/\gamma} \tag{47}$$

$$\Leftrightarrow \quad (-\log y)^{-\gamma} = 1+\gamma\frac{x-\alpha}{\beta} \tag{48}$$

$$\Leftrightarrow \quad x = \alpha + \frac{\beta}{\gamma}\left((-\log y)^{-\gamma} - 1\right) =: F^{-1}(y). \tag{49}$$

We now have everything required to fully test our model against these observations.

5.3 Interpretation

Our aim is to compare our model to the data collected in Tables 4 and 5. Now that we have the inverse function as defined in equation (49), we can determine intervals that we expect our largest mass to lie in — in order to compute the 95th percentile of the distribution, we simply work out $F^{-1}(0.95)$. Computing these values for each Weibull distribution allows us to produce the plots shows in Figures 4 and 5.

In these plots, the red data corresponds to the Planck survey as detailed in Table 4, whilst the purple data shows the data curated by Harrison and Coles in Table 5.

In each plot there are three lines, though perhaps only two are distinguishable. These represent the upper and lower bounds for the 95th percentile. I have also sketched the peak probability line, as was displayed in Figure 3. Due to the shape of the underlying Weibull distributions, the peak of the distribution is actually at a relatively low mass, with a large tail stretching to greater masses. This means that most of our interval is above the peak, and so the lower bound is tight to the peak.

As expected, the Sheth-Tormen mass function tends to predict higher masses than Press-Schechter; in fact, the data may suggest that Press-Schechter actually under-predicts. This is especially noticeable at higher redshifts, where the data points from Harrison and Coles are mostly above our predicted values. In particular, the cluster at $z = 0.87$ has error bars that do not even intersect our given interval, it is so much more massive than expected.

On the other hand, the Sheth-Tormen mass function fits the data well, with all error bars overlapping our given interval. In fact, the centre of these intervals tends to be slightly below our prediction. Though we have observed many clusters, we have not observed all of them, and then we are not guaranteed to have seen the largest one, and so it is not an issue if our predicted is slightly above the largest observed value.

I have not included the Tinker mass function plot, but as discussed in Section , we expect this lie between these two plots. This would likely cause it to intersect all the error bars, but it would leave the high redshift measurements with little leeway.

If we now compare the two data sets, we see that the points from the Planck survey tend to be lower than the both the other data set, and the predicted maximum line. Though the Planck survey did observe over 1500 clusters, this is apparently not enough to give us a representative sample of all the clusters in space. We do see, however, the general shape of these points mimic the profile of the expected maximum line. This justifies the drop in expected mass at lower redshifts that the Harrison and Coles data set couldn't confirm for us. In a similar fashion, it also covers the gap in this data set for redshift values between 0.5 and 0.8.

In order to determine the size of the most massive cluster in the Universe at any redshift,

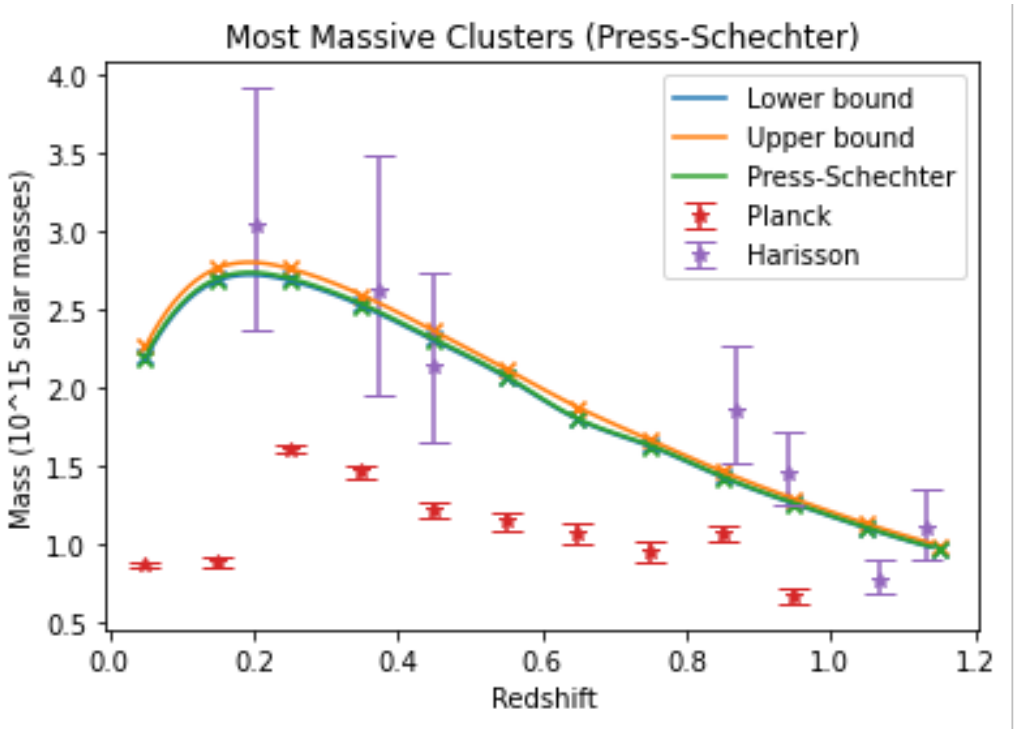


Figure 4: Predicted vs Observed most massive clusters (Press-Schechter mass function)

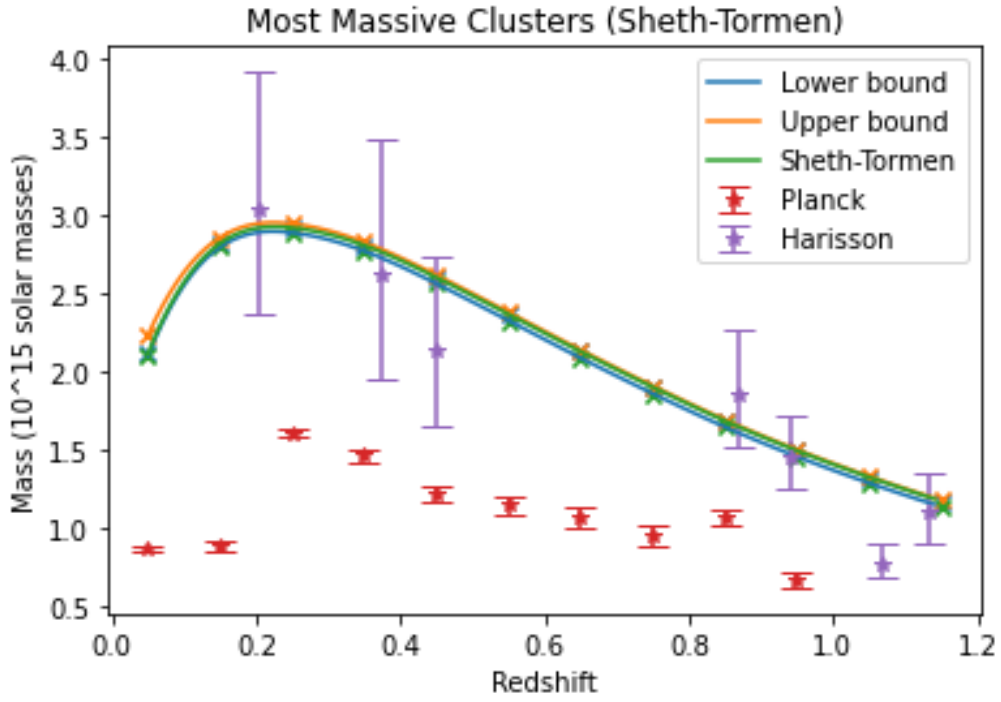


Figure 5: Predicted vs Observed most massive clusters (Sheth-Tormen mass function)

we take the consider 95th percentile for the Sheth-Tormen mass function over all redshifts bins, and take that bin that maximises this value — this quantity peaks at $2.95 \times 10^{15} M_{\odot}$.

Given the large number of surveys of the nights sky that have been made, I believe that, whilst we may have yet to see the true biggest galaxy cluster, we should have seen clusters very similar mass. The data shown here does not give us reason to believe that the Sheth-Tormen mass function is severely underestimating cluster masses systematically, as it is larger than or on par with all the observed data points considered here.

6 Conclusion

6.1 Summary of Main Results

After settling on some base assumption about the homogeneity and isometry of the Universe, we spent a lot of time defining and implementing our model of the Universe. Our first major milestone was successfully computing σ_R, which we did considering a top-hat window, and then integrating the power spectrum over this range. We then normalised this using the known value of σ_8.

Once we had obtained this function, we were able to compute the differential number density dn/dM. This required us to choose a mass function out of the three we considered. Integrating this from a fixed mass m to infinity gave us the mean number density of clusters of mass larger than this. We then defined a volume element — multiplying the density by this volume allowed us to compute the number count of clusters of mass larger than m.

Once we had our model, we stated and applied the Extreme Value Theorem of statistics to narrow down the extreme distribution to one of three forms. We were able to write down the probability of the largest mass in a redshift bin being below some value as a Poisson distribution.

After Taylor expanding and comparing coefficients, we were left with a method of determining the parameters defining the extreme distribution. These all ended up being Weibull distributions, with slightly different parameters.

We then numerically approximated the peak of all these distributions, in order to calculate percentiles for the size of the most massive cluster in a redshift bin. This information is summarised in Figures 4 and 5.

We then looked at some real-world data. Some of this was taken from the Planck survey, whilst the rest contained the most massive clusters from many different experiments. We then used this to contextualise our findings, and make conclusions as the suitability of our model and approach. In general, the data lined up well with our predictions, specifically when we considered the Sheth-Tormen mass function.

Though we can never know whether we have observed the largest galaxy cluster in the

Universe, our results suggest that the most massive cluster would not be too much larger than the largest ones already observed, and would most likely be somewhere in the range $0.2 \leq z < 0.3$. We expect the 95th percentile of this distribution to have mass not larger than $3 \times 10^{15} M_{\odot}$.

6.2 Further Research

Though this project manage to achieve its main goal of predicting the mass of the most massive galaxy cluster in the Universe, there is always more work that could be done.

In Section 4.2, we were tasked with computing the peak mass m_0 for each redshift bin, in order to parameterise the distribution for this range. It was computationally intensive to do this, and I was not able to effectively automate this process. In particular, it took significant time to achieve the required precision on these values. It was decided to forgo repeating this process for the Tinker et al. mass function, as we knew that this would only lie between the other two distributions anyway. Given more time, it would have been satisfying to explore all mass function outlined in this paper in this way, even if it would not have affected our end result.

The intensive nature of these calculations also meant that we chose to restrict our search to redshifts below $z = 1.2$. However, Table 5 contains clusters with redshifts up to $z = 1.6$. Though we would expect the maximum mass to continue to decrease as we increase the redshift — and indeed that is what these data points suggest will be the case — it would have been insightful to see how large the difference between predicted and observed masses gets as z gets large.

As noted in [7] and [5], the application of Extreme Value Statistics to cosmology has seeing a resurgence in recent years. This is with good reason — it is an incredibly powerful tool, that has yet to be applied to many situations. Though we chose the Block Maxima approach, there is a second method, known as the "Peaks Over Threshold", or "Generalised Pareto" approach. This method concerns itself with computing the probability that an observed quantity exceeds a pre-determined threshold. As with our approach, given a high enough threshold, this probability can be described by a singular distribution.

Throughout this project, I had to make some assumptions about the Universe in order to derive or simplify the equations. Had I had more time, it would have been interesting to see what would happened if we relaxed some of these. In particular, in equation (28), we assume a flat universe for simplicity, as well as when we define the FRW metric, but as there is experimental error when measuring this curvature, this is not necessarily a completely accurate statement. It would be interesting to vary this curvature to see what effects this had on our final results.

It would also be interesting to run this model on other universes. In particular, altering the values of Ω_m and Ω_Λ could allow us to predict most massive cluster possible in a matter

dominated universe. Alternatively, if we knew the values of the density parameters for some time in the past, such as before matter–Λ equivalence, we could potentially compute the mass of the most massive cluster ever to exist.

Given boundless time, I would have liked to model the extreme distributions of other quantities. Applying these approaches elsewhere could allow us to determine the expected mass of the largest black holes or dark matter halos. Whilst the model would need to be re-defined, the theory behind the application of EVS would remain largely the same.

Bibliography

References

[1] https://en.wikipedia.org/wiki/Names_of_large_numbers.

[2] PAR Ade et al. "Planck 2015 results-XXVII. The second Planck catalogue of Sunyaev-Zeldovich sources". In: *Astronomy & Astrophysics* 594 (2016). http://cdsarc.u-strasbg.fr/ftp/cats/J/A+A/594/A27/, A27.

[3] Nabila Aghanim et al. "Planck 2018 results-VI. Cosmological parameters". In: *Astronomy & Astrophysics* 641 (2020), A6.

[4] Yashar Akrami et al. "Planck 2018 results-X. Constraints on inflation". In: *Astronomy & Astrophysics* 641 (2020), A10.

[5] Siri Chongchitnan and Joseph Silk. "Extreme-value statistics of the spin of primordial black holes". In: *Physical Review D* 104.8 (2021), p. 083018.

[6] Sirichai Chongchitnan and Joseph Silk. "Primordial non-Gaussianity and extreme-value statistics of galaxy clusters". In: *Physical Review D* 85.6 (2012), p. 063508.

[7] Olaf Davis et al. "Most massive haloes with Gumbel statistics". In: *Monthly Notices of the Royal Astronomical Society* 413.3 (2011), pp. 2087–2092.

[8] Laurens De Haan and Ana Ferreira. *Extreme value theory: an introduction*. Vol. 21. Springer, 2006.

[9] Scott Dodelson. *Modern cosmology*. Elsevier, 2003.

[10] Daniel J Eisenstein and Wayne Hu. "Baryonic features in the matter transfer function". In: *The Astrophysical Journal* 496.2 (1998), p. 605.

[11] Ian Harrison and Peter Coles. "Testing cosmology with extreme galaxy clusters". In: *Monthly Notices of the Royal Astronomical Society: Letters* 421.1 (2012), pp. L19–L23.

[12] Hideo Kodama and Misao Sasaki. "Cosmological perturbation theory". In: *Progress of Theoretical Physics Supplement* 78 (1984), pp. 1–166.

[13] Andrew Liddle. *An introduction to modern cosmology*. John Wiley & Sons, 2015.

[14] Zarija Lukić et al. "The halo mass function: High-Redshift evolution and universality". In: *The Astrophysical Journal* 671.2 (2007), p. 1160.

[15] David H Lyth and Andrew R Liddle. *The primordial density perturbation: Cosmology, inflation and the origin of structure*. Cambridge University Press, 2009.

[16] Houjun Mo, Frank Van den Bosch, and Simon White. *Galaxy formation and evolution*. Cambridge University Press, 2010.

[17] Michael J Mortonson, Wayne Hu, and Dragan Huterer. "Simultaneous falsification of Λ CDM and quintessence with massive, distant clusters". In: *Physical Review D* 83.2 (2011), p. 023015.

[18] Viatcheslav F Mukhanov, Hume A Feldman, and Robert Hans Brandenberger. "Theory of cosmological perturbations". In: *Physics reports* 215.5-6 (1992), pp. 203–333.

[19] Phillip James Edwin Peebles and Phillip J Peebles. *Principles of physical cosmology*. Vol. 27. Princeton university press, 1993.

[20] William H Press and Paul Schechter. "Formation of galaxies and clusters of galaxies by self-similar gravitational condensation". In: *The Astrophysical Journal* 187 (1974), pp. 425–438.

[21] Barbara Ryden. *Introduction to cosmology*. Cambridge University Press, 2017.

[22] Ravi K Sheth and Giuseppe Tormen. "Large-scale bias and the peak background split". In: *Monthly Notices of the Royal Astronomical Society* 308.1 (1999), pp. 119–126.

[23] Tom Swallow. `https://github.com/Tomtomsatnav101/EVS`.

[24] Jeremy L Tinker et al. "The large-scale bias of dark matter halos: numerical calibration and model tests". In: *The Astrophysical Journal* 724.2 (2010), p. 878.

[25] Christian Wagner and Licia Verde. "N-body simulations with generic non-Gaussian initial conditions II: Halo bias". In: *Journal of Cosmology and Astroparticle Physics* 2012.03 (2012), p. 002.

[26] J-C Waizmann, Stefano Ettori, and Lauro Moscardini. "An application of extreme value statistics to the most massive galaxy clusters at low and high redshifts". In: *Monthly Notices of the Royal Astronomical Society* 420.2 (2012), pp. 1754–1763.

[27] Steven Weinberg. *Cosmology*. OUP Oxford, 2008.

[28] Martin White. "The mass of a halo". In: *Astronomy & Astrophysics* 367.1 (2001), pp. 27–32.

7 Appendices

7.1 Appendix A - The Code

All of the coding work has been compiled into one python script, which is available to view, download and modify here [23]. This file contains a lot of routines that are never called. These were all used at some point, mainly for testing and producing graphs, but aren't actually vital for the code to work. Upon running the code, many graphs are generated, most of which I didn't include in this paper. These were mainly for my benefit, to determine what the program was doing at various points.

Stability and numerical precision were favoured over execution time. Having said this, I did not go out of may way to slow the code down, and I tried to write my program in a relatively efficient way. On my laptop, the main routine takes less than two minutes to run. However, some functions — like the one that generated Figure 1 — took closer to half an hour to complete, so be aware of this.