School of Physics
and Astronomy

Senior Honours Project:
Simulating without Simulations

Seán Smith

January - March 2019

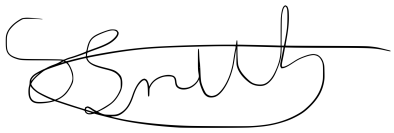
Abstract

The effort to constrain cosmological parameters requires a fundamental understanding of the large-scale structure of the universe. Describing this is the dark matter halo function, which is a pivotal cosmological property that can aid our understanding of galactic distribution and evolution. As the mass function is founded upon primordial perturbations of the early universe, understanding the connection between galaxies and their dark matter halos can provide key insight into the cosmology that governs our universe. The standard method for obtaining such data is N-body dark matter halo simulations. However, due to computational limitations, this method can only provide a limited volume of precise data. This project looks to explore a more economical technique. Random dark matter halo mass and redshift data are sampled from derived halo mass function distributions. Fitting functions from the literature are employed to obtain observable galactic properties from this data. The objective here was to determine whether such a method is viable through the comparison of derived scattered and non-scattered observable galactic data. Through correlation analysis, it was found that this method was feasible within some constraints. This technique may offer an alternative approach to harvesting large volumes of galactic data.

Supervisor: Dr B. Gillis

Declaration

I, Sean Smith, confirm that all this work is my own except where indicated by citation, reference or acknowledgement.

Signature: 

Contents

1 Introduction

1.1 Cosmology

Figure 1: Portrait of Edwin Hubble

The field of physical cosmology is a branch of astronomy which is concerned with origin, evolution and the ultimate fate of the universe. Our understanding of the cosmos has developed by observing the natural phenomena of our environment and empirically formulating these observations into a generalised theory which is then tested on new data. This scientific method has been applied as far back as ancient Greece. Since this time our picture of the universe has developed extensively from the geocentric solar system-sized universe to the uncomprehendingly large and inconceivably complex one we (don't) know today. It is only within the past century that Edwin Hubble discovered that the universe goes beyond the Milky Way. Hubble found that many of the 'spiral nebulae' in the Milky Way contained cepheids, from which he was able to conclude that these bodies were too far away to be contained within our galaxy (Hubble, 1926). Furthermore, he found that these extra-galactic bodies appeared to be moving away from us at a rate proportional to their distance (Hubble, 1929). Hubble's discovery implies that the universe must be expanding and consequently, it must have had a beginning. This evidence lead to the prevailing model of our observable universe known as The Big Bang theory.

The Big Bang theory gives a comprehensive description of the universe we see today with key evidence such as the cosmic microwave background (CMB). To understand what occurred at the beginning of the universe we adopt the method described above. That is, observe the large-scale structure of the cosmos and ask why is the universe the way it is? What happened to drive it to its current state?

1.2 Cosmic Microwave Background

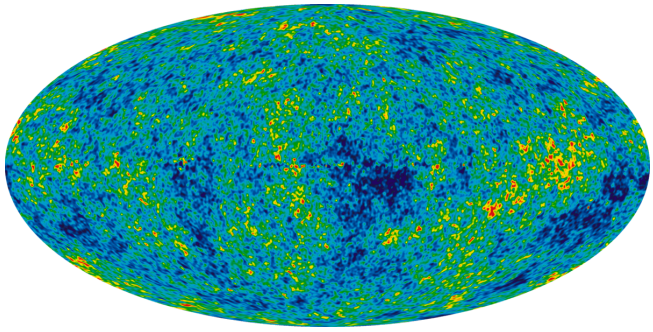


Figure 2: An all sky image of the CMB. The colour differences represent temperature fluctuations on the μK scale.

A corollary of The Big Bang Theory is that the universe is homogeneous and isotropic. Clearly this isn't the case here on Earth which is a very over dense region, however, this is believed to hold at much larger scales. This begs the question of why is the universe not perfectly even? What is responsible for the matter density variations in the first place? Again the CMB can provide some insight into this question.

The CMB was discovered by Penzias and Wilson (1965) and it allows us to see the universe as it once was at the epoch of recombination[1] when it was only ∼400kyr old. This is the youngest age we can directly observe the universe and everything before this must be inferred. Nevertheless, in 1989 the COBE satellite was launched and found that the CMB had a temperature of $T = 2.725 \pm 0.002$K. In regards to the structure of the universe, the most interesting part of this result is the uncertainty or more specifically its fluctuations. These are known as primary anisotropies and they illustrate that the universe was (and therefore is) somewhat inhomogeneous. Of course, it is possible that the anisotropies may be a consequence of some interaction that occurred along the line of sight. This is true, however, for the most part, these can be ignored (Aghanim et al., 2008). The primary anisotropies provide a direct link to fluctuations in the density field of the early universe. For example, consider a photon leaving an over-dense region at the epoch of recombination. This photon will become marginally redshifted relative to other photons as it climbs out of the potential well of this region. This perturbation in the density field directly caused a variation in the energy and therefore the temperature of the CMB.

1.3 Dark Matter

Along with the great vastness of the universe, the conclusive existence of dark matter (DM) is a very modern discovery. In the 1970s many papers were published on the topic of spiral galaxy rotation curves which provided

[1] The time of which the plasma of protons and electrons became bound to form atoms. It is also the time that the universe became transparent and hence the CMB was born.

strong evidence for DM (Rubin, 1979; Roberts and Rots, 1973). Assuming circular orbits and that galaxies are spherically symmetric, the rotation velocity, v, can be given by,

$$v^2 = \frac{GM(<r)}{r} \tag{1}$$

Where r is the radial distance from the galactic centre, M is the enclosed mass and G is the universal gravitational constant. It was expected that stars within the galaxy would orbit more slowly the further they were from the galactic centre. In other words, the velocity, v, would display a Keplerian fall-off $v \propto r^{-1/2}$. However, results exhibited a near constant velocity irrespective of r. Using this result and by consulting equation 1 it can be shown that $M(<r) \propto r$ in these spiral galaxies. Clearly, this relation will come to a halt at the galactic edge otherwise the mass would diverge. Nonetheless, it appeared as if the mass was increasing with the galactic radius. This was in contrast to the amount of light observed as a function of r. The only explanation for this is that there must be matter present that could not be detected. This matter is thought to take an ellipsoidal shape around the galaxies and are called dark matter halos (DMH). The rotation curves towards the edge of these galaxies imply that the mass-to-light ratio is extremely high, $M/L \sim 1000$, where L is luminosity. To compare, the typical mass to light ratios of stars is $M/L \leqslant 10$. This suggested that the visible matter only makes up about 1% of the mass of the galaxy.

Furthermore, there are other independent methods that are in close agreement with these results. One example is gravitational lensing - the curved path light takes as it passes massive bodies. This heavily involves Albert Einstein's theory of General Relativity and so this won't be discussed in detail. Nonetheless, these agreements make it difficult to argue that our understanding of gravity is at fault rather than we are wrong about the existence of DM.

It should be clearly stated that it is believed that DM is not simply 'regular' matter that is faint to detect. Rather it is matter which is composed of particles which currently aren't understood by the Standard Model (non-baryonic).

As mentioned in section 1.2, the universe appears to be inhomogeneous on smaller scales. However, this diagnoses was entirely founded upon the uneven distribution of baryonic (visible) matter. Baryonic matter constitutes a very small portion of the mass-energy density of the universe (see table 1). Where the present-day DM density parameter is $\Omega_{d,0} = \Omega_{m,0} - \Omega_{b,0}$ and therefore $\Omega_{d,0} \gg \Omega_{b,0}$. Consequently, the distribution of DM is a very important property when discussing the large-scale structure of the universe. As previously discussed, DMHs envelope galaxies and account for the vast majority of mass in the system. It is believed that DMHs formed under gravitational collapse in over-dense regions of the early universe. Subsequently, baryonic matter fell into the halo's potential well to form stars, galaxies and all other celestial bodies we see today. Usually, the halo encompasses one central galaxy and often have smaller satellite galaxies (Rodríguez-Puebla et al., 2016).

1.4 Inflation

The very early universe ($< 10^{-12}$ s) is poorly understood and there is very little knowledge of the events that occurred in this time. Despite this, the conditions of the very early universe had an undoubtedly important part to play in its development. Inflation theory is the leading postulate for this time and is the idea that the universe expanded exponentially for a very short period at the very beginning of the universe. This is thought to have been a consequence of oscillations in a potential called the inflation field. Much like the electrostatic potential energy of the electric field, the inflation field has a state of pure potential energy. This corresponds to the vacuum (dark) energy which drives the expansion of the universe. Another attractive property of inflation theory is that it inevitably predicts inhomogeneities in the universe. It is reasonable to assume that the inflation field is governed by quantum mechanics and therefore exhibits quantum fluctuations. These fluctuations would be a consequence of the Heisenberg uncertainty principle and result in minor perturbations in the inflation field. These perturbations were then 'stretched' to macroscopic scales during the inflationary period. Consequently, this would result in vacuum energy perturbations and subsequently, the universe will have expanded unequally in different directions. Thus, since $\rho \propto R^{-3}$ the inflationary period produced perturbations in the density field. This is the seed for DMH collapse. Again, this would produce fluctuations in the CMB as described previously. It is now clear that the magnitude of the CMB perturbations could be key in predicting the events of the early universe.

1.5 Large-Scale Structure

In order to understand how matter is distributed throughout the universe, it is useful to construct a DMH mass function which describes the number of structures per unit comoving volume per mass interval. In this project, the mass function is founded upon linear perturbation theory, the spherical collapse model and adopts the Sheth-Torman formalism (STF).

1.5.1 Density Perturbation Field

Before discussing the dynamics of structure formation, it is constructive to understand how the density perturbation distribution of the universe can be entirely described by its power spectrum. To begin, we quantify the density field perturbations discussed previously.

$$\delta \equiv \delta(\vec{x}) = \frac{\rho(\vec{x}) - \bar{\rho}}{\bar{\rho}} \tag{2}$$

This is the dimensionless density perturbation field[2] (DPF) at some given time. Where $\rho(\vec{x})$ and $\bar{\rho}$ are the density at some position $\vec{x}$ and the average density of the universe, respectively. Across the entire universe, the perturbations will oscillate about $\bar{\rho}$. Consequently, this is is best handled using Fourier analysis.

$$\delta = \sum \delta_{\vec{k}}\, e^{i\vec{k}\cdot\vec{x}} \tag{3}$$

This describes the density perturbation field as a sum over all wave modes, k, in some volume of the which the universe is considered periodic. Attempting to describe the DPF at every point would be a fruitless task. Instead, seeking an expectation value would be more useful as the universe will exhibit random variations about the average density. Of course the perturbations will average to zero ($\langle\delta\rangle = 0$), however, the variance, σ^2, of the DPF is non-zero and is a useful property.

$$\sigma^2 \equiv \langle\delta^2\rangle = \frac{1}{V}\int \left|\sum \delta_{\vec{k}}\, e^{i\vec{k}\cdot\vec{x}}\right|^2 dV \tag{4}$$

Where V is the periodic volume which the wave modes are averaged over. This gives,

$$\sigma^2 = \sum |\delta_{\vec{k}}|^2. \tag{5}$$

A similar analysis can be applied to the correlation function which describes the probability of two bodies being separated by a distance, r.

$$\xi(\vec{r}) \equiv \langle\delta(\vec{x})\delta(\vec{x}+\vec{r})\rangle \tag{6}$$

The Fourier transform of the correlation function is called the power spectrum, $P(k)$, and vice versa.

$$\xi(\vec{r}) = \frac{1}{(2\pi)^3}\int P(k)e^{i\vec{k}\cdot\vec{r}}d^3\vec{k} \tag{7}$$

Moreover, notice that when $r = 0$ then $\xi = \sigma^2$. Therefore,

$$\sigma^2 = \xi(0) = \frac{1}{(2\pi)^3}\int P(k)d^3\vec{k} \tag{8}$$

Now, because the variation of the DPF is as equally probable in every direction, the variance can be expressed in terms of spherical polars as,

$$\sigma^2 = \frac{1}{2\pi^2}\int P(k)k^2dk \tag{9}$$

This is a very useful result as the DPF can be entirely described by the power function.

In order to get a complete description of the variance of the DPF, the integral in equation 9 should be taken to infinity. Consequently, it is clear to see that the variance will diverge - this suggests that there is a lot of structure on small scales. The solution to this is to introduce a window function, $W(x;R)$, to filter out large values of k. Normally, window functions meet some basic conditions. These are $W(0) = 1$, $\int W(x)dV = 1$ and they should be reasonably integrable. The top-hat window function is commonly used in cosmology and is used in this project. This is described in real and k-space below.

$$\begin{aligned} W(\vec{x};R) &= \begin{cases} \frac{3}{4\pi R^3}, & r \leqslant R \\ 0, & r > R \end{cases} \\ \tilde{W}(kR) &= \frac{3}{(kR)^3}(\sin(kR) - kR\cos(kR)) \end{aligned} \tag{10}$$

By implementing this in equation 9, we have a final result for the variance of the DPF:

$$\sigma^2 = \frac{1}{2\pi^2}\int P(k)\tilde{W}^2(kR)k^2dk \tag{11}$$

An important observation is that provided the background density of the universe, $\bar{\rho}$, is known, a volume of space can be as easily described by its mass such that $\sigma(R) = \sigma(M)$.

There is a conventional cosmological parameter denoted σ_8. It is defined as the variance of the DPF at a scale of $8h^{-1}(\approx 11.8)$ Mpc and at $z = 0$. This parameter is used to characterise the normalisation of the power spectrum. Currently, the universal value of this parameter is thought to be ~ 0.8.

1.5.2 Spherical Model and Linear Growth

As previously discussed, the period of inflation caused density perturbations throughout the universe. Clearly, this will result in over and under-dense regions and will ultimately prompt gravitational collapse to form a DMH. A simplified yet reasonable analysis of this mechanism is the spherical model. Consider an isolated and over-dense region of space that can be treated as a sub-universe. The equation of motion of the matter within the sphere is $\ddot{r} = -\frac{GM}{r^2}$; note that r is the proper radius of the sphere. The matter outside of the sphere will have no gravitational effect on the over-dense region as it is expected that the universe is homogeneous. Integrating this equation yields $\frac{\dot{r}^2}{2} = \frac{GM}{r} + E$ where E is the energy per unit mass of the shell. If $E = 0$, this gives a solution of the form $r \propto t^{2/3}$. According to Einstein-de Sitter cosmology[3], this expansion rate is equal to that of the universe.

An over-dense shell is gravitationally bound; this corresponds to $E < 0$. Solutions for this situation are best given in parametric form:

$$\begin{aligned} r &= A(1-\cos\theta) \\ t &= B(\theta - \sin\theta) \end{aligned} \tag{12}$$

Where $A^3 = GMB^2$ from $\ddot{r} = -\frac{GM}{r^2}$ and θ is $\in [0, 2\pi]$. Figure 3 illustrates the radius of the over-dense shell ex-

[2] This density refers to matter only and does not include the vacuum or radiation energy density

[3] This is a highly simplified version of cosmology, however, this is a good approximation at high red-shifts

panding with the universe and then turning around to collapse under gravitational force.

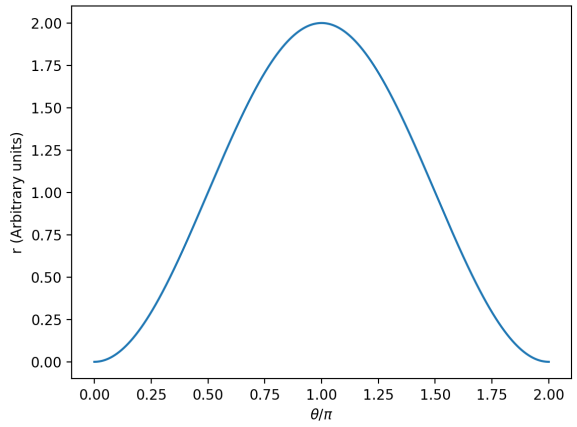


Figure 3: The radius of an over-dense sub-universe as it varies with θ (time). The shell reaches a maximum radius as the universe expands and then collapses due to gravity.

Like the radius of the shell, according to linear theory, the density perturbations will also grow at the same rate ($\delta \propto t^{2/3}$). Therefore, it is essential to determine the correct initial conditions of the perturbations at recombination ($\theta \ll 1$). By Taylor expanding the relations in equation 12 and substituting into the mass-density relationship of the shell (see Appendix I), the density perturbation can be expressed as

$$\delta = \frac{3}{20}\left(\frac{6\pi t}{t_{max}}\right)^{2/3}. \tag{13}$$

Where t_{max} is the time of maximum r. From this equation, the linear theory critical density, δ_c, can now be found. This is the point when an over-dense region should have collapsed. By consulting figure 3, the shell collapses when $t/t_{max} = 2$. By substituting this into equation 13, the critical density can be found.

$$\delta_c \simeq 1.686 \tag{14}$$

This result isn't entirely general as linear theory breaks down as $|\delta|$ increases. Linear theory assumes that as a region collapses it becomes denser and in turn, attracts more mass and so the region becomes even denser. In other words, the region grows linearly in the absence of other physical processes. However, as these perturbations grow, other mechanisms besides gravity will become more significant. For example, as a DMH collapses it will reach virial equilibrium where the potential energy and kinetic energy will balance ($E_p = -2E_k$). Furthermore, $\delta \nless -1$ as this corresponds to a region being more than 100% less dense than the average density. This is nonphysical and therefore the linear evolution of the perturbation ceases. At this point, a non-linear model is required to describe the collapse. Furthermore, the symmetric collapse model used here is highly idealised as the DMH comes to a singularity. Despite these assumptions, they are reasonable approximations and applying them is in the nature of this project.

The critical density isn't constant throughout redshift (time). Instead, it will evolve as described below.

$$\delta_c(z) = \frac{1.686}{D(z)} \tag{15}$$

Where $D(z)$ is the linear growth factor which describes the growth of structure as a function of time.

1.5.3 Press-Schecter Theory

Linear theory and the spherical collapse model have provided a critical density perturbation that describes when an over-dense region will form a DMH. The goal now is to assign a mass to those halos using the statistical variation of the DPF. The general idea is to search and count the locations of the DPF where the density perturbation surpasses the critical density. Bardeen et al. (1986) explored this idea but ran into a critical issue of over-counting peaks by counting ones which were part of other larger peaks. Instead, another approach made by Press and Schechter (1974) is much more successful. They postulated that at any redshift, the probability of $\delta > \delta_c$ in a region of mass M, is equal to the mass fraction $> M$ that would be contained within the DMH. This expressed as Gaussian random field is

$$F(>M,z) = P(\delta > \delta_c) = \frac{1}{\sqrt{2\pi}\sigma_M}\int_{\delta_c}^{\infty} \exp\left[-\frac{\delta^2}{2\sigma_M^2}\right]d\delta \tag{16}$$

Where $\sigma_M \equiv \sigma(M)$. There is a further issue with this postulate in that only half of the matter in the universe will be part of a halo as its equally likely that a region will be over or under-dense. The problem with this is that an under-dense region could be enclosed within a larger over-dense region. This means that under-dense regions have a non-zero probability of being part of a halo. To tackle this, a fudge factor of 2 was introduced into the probability such that $F = 2P$.

By defining the mass function as $n(M,z)dM$, the number of DMH with masses in the range of M to $M + dM$ per comoving volume. $\frac{\partial F}{\partial M}dM$ is the fraction of mass within a DMH in the same range. Therefore, by multiplying by the average density of the universe, $\bar{\rho}$, gives the total mass per comoving volume within the DMH. And finally, by dividing by the mass of the halo, this gives the mass function.

$$n(M,z)dM = \frac{\bar{\rho}}{M}\frac{\partial F}{\partial M}dM = \frac{2\bar{\rho}}{M}\frac{\partial P}{\partial M}dM \tag{17}$$

By applying the chain-rule, $\frac{\partial P}{\partial M} = \frac{\partial P}{\partial \sigma}|\frac{d\sigma}{dM}|$ the full expression for the mass function can be given by

$$n(M,z)dM = \sqrt{\frac{2}{\pi}}\frac{\bar{\rho}\ \delta_c(z)}{\sigma_M M^2}\exp\left(-\frac{\delta_c^2(z)}{2\sigma_M^2}\right)\left|\frac{d\ln\sigma_M}{d\ln M}\right|dM \tag{18}$$

By defining the variables of $\nu \equiv \delta_c/\sigma_M$ and the Press-Schecter formalism (PSF), $f_{PS}(\nu) = \sqrt{\frac{2}{\pi}}\nu\exp(-\nu^2/2)$, the mass function can be condensed to,

$$n(M,z)dM = \frac{\bar{\rho}}{M^2}f_{PS}(\nu)\left|\frac{d\ln\nu}{d\ln M}\right|dM \tag{19}$$

Notice that time enters the mass function exclusively through δ_c. Furthermore, as δ_c decrease with time, the mass at which the exponential cut-off becomes dominant increases. By consulting equation 18 and holding M constant, as time increases (z decreases) the exponential term increases giving a higher number of objects at that mass. This makes sense, as more time passes it is increasingly likely that more massive structures will have formed.

1.5.4 Sheth-Torman Formalism

The PSF proved to be to relatively successful and showed fair agreement to numerical solutions (Raig et al., 2001). Despite this, the formalism clearly conveyed that there was something awry. One of the largest pitfalls of PSF is that it assumes that the halo collapse is spherical. Sheth et al. (2001) found that if the collapse is treated as ellipsoidal, the discrepancy between the resulting mass function and numerical simulations decreased significantly. Moreover, the spherical model only has the condition that the density perturbation must be greater than a critical value for a collapse to occur. This means that the probability of a collapse occurring is independent of the size of the over-dense region. Subsequently, as the size is related to the mass, the critical value is also independent of the final mass of the region. To counter this, it was proposed that δ_c should also be a function of σ_M which is a function of mass. This new relation asserted that δ_c decreased with mass because more massive regions are inherently more gravitationally attractive. This amendment to the mass function can be applied by simply replacing $f_{PS}(\nu)$ with $f_{ST}(\nu)$.

$$f_{ST}(\nu) = 0.322\left(1+\frac{1}{\tilde{\nu}^{0.6}}\right)f_{PS}(\tilde{\nu}) \tag{20}$$

Where $\tilde{\nu} = 0.84\nu$. This formalism is in much better agreement with numerical solutions and was used for the construction of the mass function in this project.

1.6 Objective

The mass function is one of the most important properties in cosmology. It describes the hierarchical structure of the whole universe and as a result governs the processes of galaxy formation and distribution (Frenk et al., 2001; Reed et al., 2007). Therefore, by extrapolating observational galactic properties from it, the mass function can aid in our understanding of the astrophysical mechanisms that fashioned the structure of the universe. A great effort has gone into identifying the connection between DMHs and their host galaxies (Wechsler and L. Tinker, 2018). This link can provide estimates of both the density perturbations of the early universe and the development of fundamental cosmological parameters such as Ω_m (Murray et al., 2013). In other words, the mass function holds key knowledge to the primordial universe and the events that occurred then.

The conventional method for predicting the mass function is by investigating N-body simulations which produce a DMH population. From here semi-analytical methods are used to determine galactic properties (Bower et al., 2006). Modelling galaxy formation is a complex task as the underlying physics isn't well known. This means that a generalised equation can't be used to determine galactic properties (Frenk et al., 2012). Moreover, due to computational limitations, simulations of DMH populations have to be compromised between the parameters of volume and resolution. Consequently, relevant processes such as star formation are on too small a scale to be directly simulated. To understand how these structure formations evolve with time, often an 'observer' is placed inside the simulated volume. A lightcone is then constructed to determine which of the galaxies' light has had sufficient time to be visible to the observer (Frenk et al., 2012). This method is extremely intensive and computationally expensive if precise results are desired.

This project takes a different approach by simply employing a dependency tree of fitting functions from the literature to obtain observational galactic properties. This includes using a fitting function for the power spectrum to describe the mass function. It is in the nature of this project to apply more economical methods than described above at every stage of the dependency tree and in the physics of the mass function. In order to obtain galactic properties, probability density functions were derived for redshift and mass from the halo mass function. From this, pseudo-random values of redshift and mass were sampled and propagated through the dependency tree, adding scatter where required. The objective was to determine whether this approach is viable by comparing virtual images with and without additional scatter. This will determine whether this technique is too simplistic and accretes too much uncertainty. This method was applied for 1 million DMH from the present-day universe out to a redshift of $z = 1$.

2 Methods

2.1 Choice of Cosmology

For this project, a flat ΛCDM model of the universe was chosen. That being a universe which contains vacuum energy, dark matter which moves at velocities $\ll c$ (cold), baryonic matter and is geometrically flat ($\Omega_k = 0$). This

model is a very reasonable approximation of the universe out to $z = 1$. All relevant cosmological parameters used in the project can be found in table 1.

Parameter	Value
$\Omega_{m,0}$	0.308
$\Omega_{b,0}$	0.048
H_0	67.81
σ_8	0.815
n_s	0.968

Table 1: Cosmological parameters used for constructing the dark matter halo function taken from the Planck 2015 Collaboration (Ade et al., 2016). Where H_0 has units of km s^{-1} Mpc^{-1}

2.2 Mass Function

The mass function was constructed using `python` program (see Appendix VII). As discussed in section 1.5.1, the variance of the DPF can be completely described by the power spectrum. Ma (1998) presents a fitting function for a linear power spectrum that was applicable to cosmology used in this project.

$$P(k,z) = \frac{Ak^{n_s}[D(z)/D_0]^2[\ln(1+\alpha_1 q/(\alpha_1 q)]^2}{[1+\alpha_2 q+(\alpha_3 q)^2+(\alpha_4 q)^3+(\alpha_5 q)^4]^{1/2}} \tag{21}$$

This function was founded upon other previous work. Bardeen et al. (1986) presented pioneering analysis into the statistical approach of the large-scale structure by treating the density field as Gaussian. It was from this research that provided the α constants of the power spectrum presented in table 2.

α_i	Value
1	2.34
2	3.89
3	16.1
4	5.46
5	6.71

Table 2: α constants used in the fitting function of the power spectrum.

The normalisation constant, A, was fixed such that $\sigma_8 = 0.815$. $q = k/(\Gamma h)$, where Γ is the shape parameter well approximated by $\Gamma = \Omega_{m,0} h \exp[-\Omega_{b,0}(1+1/\Omega_{m,0})]$ (Bunn and White, 1997). $D(z)$ is the linear growth factor and $D_0 \equiv D(0)$. n_s is the spectral index which describes how structure changes with scale. Finally, k is the wavenumber in units of Mpc^{-1}.

By substituting the power spectrum into equation 11 and integrating, the variance of the primordial DPF was determined. From here σ was applied to the Sheth-Torman mass function within a mass range of $(10^8 - 10^{16})M_\odot$ at eleven redshifts up to $z = 1$. This gave the number of DMHs per unit volume per unit mass, $n(M, z)$.

2.3 Random Data

Drawing out a realistic number of masses for each mass interval initially relies on z. This is because the mass function changes with time. For this reason, a probability density function (PDF) of redshift was required to generate z values. This would, in turn, draw mass values from the mass function. In order to construct such a function, the number of DMHs (counts) as a function of redshift needed to be determined. At each redshift, the mass function can be summed together to marginalise over mass. This produces the total count per unit volume per unit redshift. To remove the volume dependency, the relationship between the volume and redshift has to be considered. The redshift affects distance and subsequently volume in two different ways. Firstly, a constant angle will subtend a larger transverse distance at a larger radial distance (up to a certain z). Secondly, the distances between interval redshifts are not identical at different redshifts as a result of the variable rate of expansion.

In Euclidean space, the radial distance, angular size and the transverse distance the angle subtends, are simply related by basic trigonometry. However, this relation is more complex in a ΛCDM universe. As one may expect, a celestial object will appear smaller at greater distances, however, this only holds up until a certain point. Because the universe is expanding, light emitted by a source at earlier times will have been part of a smaller universe. In other words, the further back in time that light was emitted, the more expansion the universe has gone through. Therefore, parallel photons from either side of a source are pulled further apart such that the object will appear to subtend a greater angle in the sky to a perpendicular observer. The expansion can be accounted for by the angular diameter distance, $d_A(z)$ (see Appendix IV). The angular diameter distance accounts for the expansion of the universe perpendicular to an observer. As the transverse expansion will occur in two dimensions relative to the observer, the count per volume per must be multiplied by $d_A^2(z)$.

Furthermore, because the Hubble parameter, $H(z)$, is not constant with time, the distance between interval redshifts changes with time (redshift). Therefore, the rate of expansion with respect to redshift needs to be determined. This can be done by taking the derivative of the comoving distance with respect to redshift, $d'_C(z)$ (see Appendix V). By multiplying the summed mass function at each redshift by these two quantities the number of DMHs per steradian per unit redshift, Ξ, can be found.

$$\Xi(z) = d_A^2(z)\ d'_T(z)\ \sum_i^N n(M_i, z) \tag{22}$$

From here a probability density function for redshift, P_z, was constructed with Ξ such that,

$$P_z = \frac{1}{A}\,\Xi(z), \quad \text{where} \quad A = \int_0^1 \Xi(z)dz \tag{23}$$

A cumulative distribution function (CDF), C_z, was then calculated from the PDF. This is defined as,

$$C_x = \int_{-\infty}^{x} P_{x'}\, dx'. \tag{24}$$

As $P_z = 0$ for $z \leqslant 0$ and only redshifts up to $z = 1$ are considered, the limit of integration can be fixed as the following.

$$C_z = \int_0^1 P_{z'}\, dz' \tag{25}$$

A pseudo-random Monte Carlo method was implemented to draw out random z values from the inverse of C_z.

As mentioned previously, eleven mass functions were calculated at different redshift intervals. Consequently, data for $n(z)$ for some fixed M was only available at 0.1 intervals. In an effort to use the true randomly determined values of z, the mass functions were interpolated between redshifts. Next, for every z a new mass function was interpolated at that value. A similar technique was then applied such that a CDF, C_M, was constructed for each new mass function from which m values were drawn. These values of m and z were then partnered to determine observable galactic properties from the randomly generated DMH.

2.4 Determining Observable Data

To create a virtual image, observable properties were calculated with the halo mass and redshift data. Each step in the flow chart (figure 4) from halo mass required a fitting function or well known astrophysical relations to determine the property below. An effort was made to search for literature data which provided a stepping stone to each property across the entire mass range. However, most work provided fitting functions which were constrained to narrower ranges.

[4]The luminosity analogue of the mass function

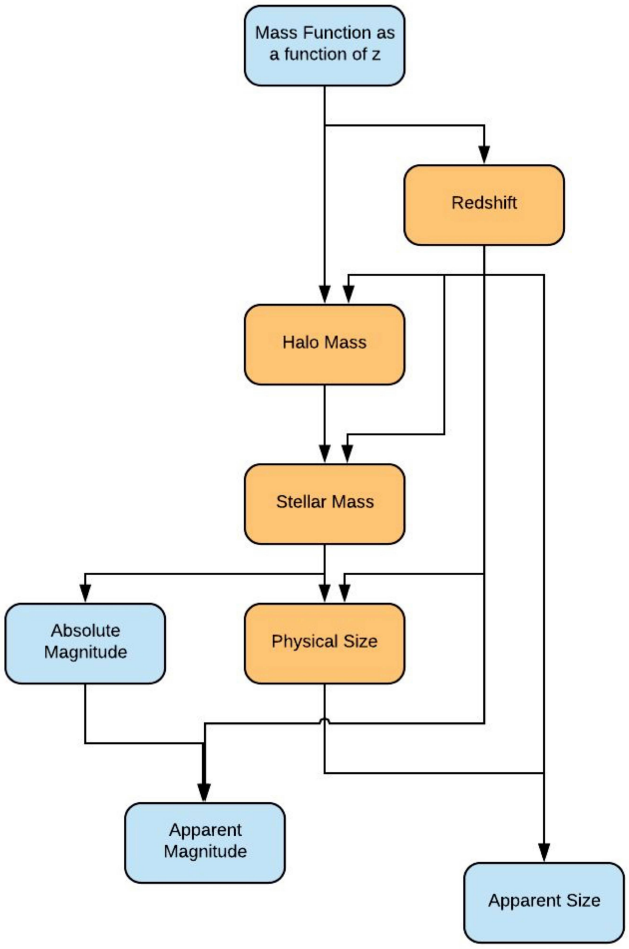


Figure 4: Flow chart describing the process of determining observable galactic properties from randomly generated mass and redshift.

2.4.1 Stellar Mass

As previously mentioned, DMH surrounds its host galaxy which is frequently accompanied with smaller satellite galaxies. In keeping in the nature of this project, these satellites were ignored and only the central host galaxy was considered. When searching for a link between two astronomical properties, it is often resourceful to determine their ratio. Benson et al. (2003) illustrated that the luminosity function[4] is steeper than the halo mass function at higher masses and less so at lower masses (their figure 1). This suggests that the baryonic mass of a galaxy does not follow the same power-law as the DMH mass and that there is a characteristic mass at which the slope changes. Yang et al. (2003) developed a parametrisation of the mass-to-light ratio with four free parameters. This idea was then adopted by Moster et al. (2010) to construct a fitting function between the stellar mass and the halo mass.

$$\frac{m(M)}{M} = 2A_0\left[\left(\frac{M}{M_1}\right)^{-\beta} + \left(\frac{M}{M_1}\right)^{\gamma}\right]^{-1} \tag{26}$$

Where m is the stellar mass, M is the halo mass, A_0 is the normalisation and M_1 is a characteristic mass. In order to determine the four free parameters, a χ^2 fit was applied to their simulation data. The results are shown in table 3.

Parameter	Value
$\log_{10} M_1$	11.884 ± 0.027
A_0	0.02820 ± 0.00057
β	1.057 ± 0.05
γ	0.556 ± 0.007

Table 3: Free parameters values of equation 26 (Moster et al., 2010)

Moster et al. (2010) also found a redshift parameterisation for each free parameter which were applied to equation 26. The redshift equations can be found in Appendix II.

2.4.2 Physical Size

As one might expect, the stellar mass-size ratio of a galaxy will be dependent on its morphology. Early-type galaxies take a more simplified spherical shape compared to the asymmetric nature of late-types. Despite this, if the radius of a galaxy is defined at the point where 80% of the stellar light is contained (r_{80}), the morphological properties of galaxies become negligible (Miller et al., 2019). Before this stellar mass-size relations have been treated as a function of morphology (Shen et al., 2003). Mowla et al. (2019) adopted the r_{80} definition to present a stellar mass-size relation irrespective of morphology.

$$r_{80} = r_p\left(\frac{M}{M_p}\right)^{\alpha}\left[\frac{1}{2}\left(1 + \left(\frac{M}{M_p}\right)^{6}\right)\right]^{(\beta-\alpha)/6} \tag{27}$$

Where M_p is a characteristic mass, r_p is a normalisation factor, α and β are free parameters. The relationship between the stellar mass-size relation and redshift has been studied extensively (Sulentic et al., 2013; Grado et al., 2018). Despite this, there does not appear to be a paramerterisation that incorporates redshift. Mowla et al. (2019) offers parameter changes over redshift bins. The bins over the concerned redshift of 0 to 1, show very little change in the parameters. Nonetheless, these changes were applied to the parameters with respect to redshift in the `python` program. The values of these parameters are given in table 4.

Parameter	$z < 0.5$	$z \geqslant 0.5$
$\log_{10} M_p$	10.2	10.5
r_p	8.6	8.7
α	0.17	0.17
β	0.50	0.61

Table 4: Free parameters values of equation 27 (Moster et al., 2010)

2.4.3 Apparent Size

The tool for finding the apparent size has already been discussed in section 2.3 - the angular diameter distance, d_A. This is the distance between an object and an observer whilst accounting for universal expansion. Thus, the apparent size, θ, can be given as the following.

$$\theta = \frac{r_{80}}{d_A(z)} \tag{28}$$

Where r_{80} is the radius of the object that contains 80% of the light.

2.4.4 Absolute Magnitude

The attempt to link stellar mass to luminosity is a historic campaign in astronomy. For a long time, it has been clear that the relation between these galactic properties is multivariate (Pskovskii, 1965). This is mainly because stellar M/L values are not constant but instead functions of type, age, metallicity and so on. Furthermore, galactic luminsoities depend on star formation rate and colour (Bell and de Jong, 2001) as well as rotational velocities in late-type galaxies (Tully and Fisher, 1977) and velocity dispersion in early-type galaxies (Faber and Jackson, 1976). It is therefore difficult to establish a simple fitting function to relate these two properties. The streamlined approach would then be to find some average M/L value. As mentioned before, typical M/L values at the centre of a galaxy are less than 10. Therefore, the average value was calculated by finding the logarithmic mean between 1 and 10. This gives $M/L \approx 3.9$ which was used for every virtual galaxy.

Deriving the absolute magnitude from luminosity is very straightforward using the following equation.

$$M = -2.5\log_{10}\left(\frac{L}{L_\odot}\right) + M_\odot \tag{29}$$

2.4.5 Apparent Magnitude

Calculating the apparent magnitude, m, from M is fairly trivial. However, at high redshifts, there is a complication. Again, this is a consequence of the expanding universe.

$$m = M + 5\log_{10}\left(\frac{d_L}{10\text{pc}} + K(z)\right) \tag{30}$$

Where d_L is the luminosity distance which is defined as the distance from an object with luminosity L such

that the flux is given by $F = \sqrt{L/(4\pi d_L^2)}$. This can also be expressed as $d_L = (1+z)^2 d_A$. K is the K-correction which converts a redshifted magnitude of an object to its equivalent in its rest frame. It can be defined as,

$$K(z) = 2.5(\alpha - 1)\log_{10}(1+z) \tag{31}$$

2.4.6 Scatter

Each observable property contained within an orange box in the dependency tree was treated with scatter. This scatter was provided from the literature of each respective fitting function and was applied to every free parameter. The properties in the blue boxes received no additional scatter as the physics connecting these parameters are well known. However, there was one exception. As mentioned previously M/L relations are extremely complex and sensitive to many variables. An *ad hoc* gamma distribution was constructed to produce scatter when determining galactic luminosity. The details of this distribution are discussed later.

3 Results & Analysis

To understand the reliability of the somewhat economical method used in this project, a detailed discussion of the simplifications is required. Naturally, fitting functions are very *ad hoc* and normally are only applicable to a constrained range of a variable. The large mass range used here often suffered from this which produced incongruous results at the range's extremes. However as one would expect, the fitting functions were developed for a parameter range with a high abundance of data. Consequently, the fitting functions were applicable in the most suitable parameter range to create a virtual image.

3.1 Mass Function

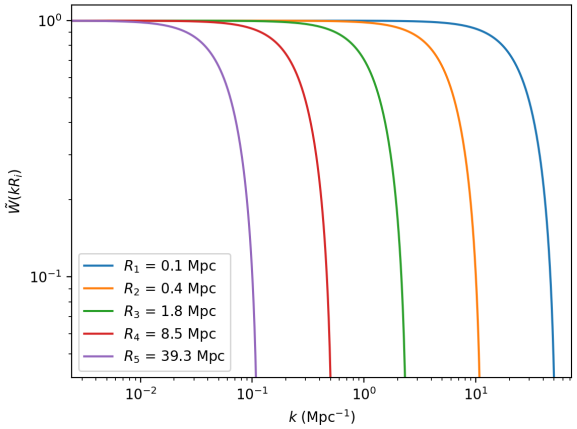


Figure 5: Each line represents the window function cut-off point for values of k at a fixed value of R. The cut-off moves from right to left as R increases. This is at $z = 0$.

The mass function was heavily dependent on the power spectrum of the DPF and therefore the fitting function played a pivotal role in its construction. The largest shortcoming of the power spectrum was its disregard of non-linear processes. Ma (1998) also provided a non-linear power spectrum fitting function. The linear and non-linear fitting functions began to deviate at $k \sim 1$ Mpc^{-1} (their figure 1) at the point where linear theory starts to break down. Moreover, the errors in the linear $P(k)$ reach 50% when $k \sim 10$ Mpc^{-1}. Despite this, the window function filtered out most of the large values of k (figure 5) which prevented the less accurate values of the power spectrum to contribute to value of σ^2 (see equation 11). By consulting figure 5, it is clear that as R decreases, less values of $k \geqslant 1$ Mpc^{-1} are filtered by the window function. From this, it is clear that non-linear effects become more significant on smaller scales. Subsequently, as $M \propto R$, the errors of the linear power spectrum will affect lower masses of the mass function more significantly. Additionally, as z increases the universe becomes denser. This means that regions of space require less volume (radius) to contain the same mass. Therefore, $\tilde{W}$ permits larger values of k to contribute before the cut-off. It can be concluded that as redshift increases the reliability of the mass function will decline. A similar plot in figure 13 at $z = 1$ can be found in Appendix III.

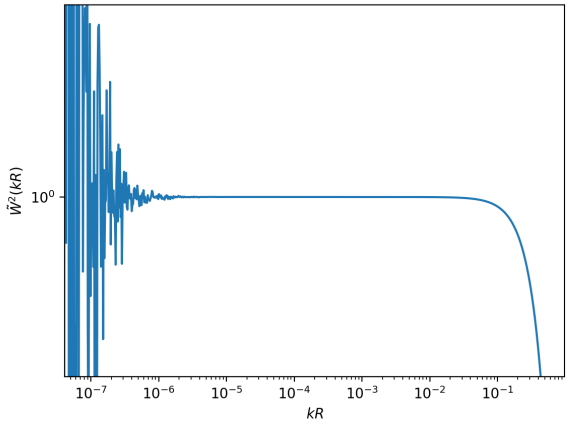


Figure 6: The window function is constant at 1 for values $< R$. At $kR \ll 1$ it appears as if `python` isn't able to handle the function.

The analytical calculation of σ^2 requires that k is integrated from 0 to ∞. Clearly, this isn't possible numerically and so a suitable range had to be determined. Figure 5 shows that values of $k \gtrsim 10^2$ Mpc^{-1} have very little contribution to the integration. This is also true up to $z = 1$ (Appendix III). For this reason, it was reasonable to set the upper limit of integration to 100. Moreover, $\tilde{W}$ appears to become poorly defined at $kR < 10^{-6}$ as illustrated in figure 6. Consequently, a lower limit of the integration was set to $k = 10^{-5}$ as R very rarely dropped below ~ 0.1.

The resulting dark matter halo mass function was in good agreement with DM simulations from previous work (Sheth et al., 2001; Eales, 2015; Reed et al., 2007).

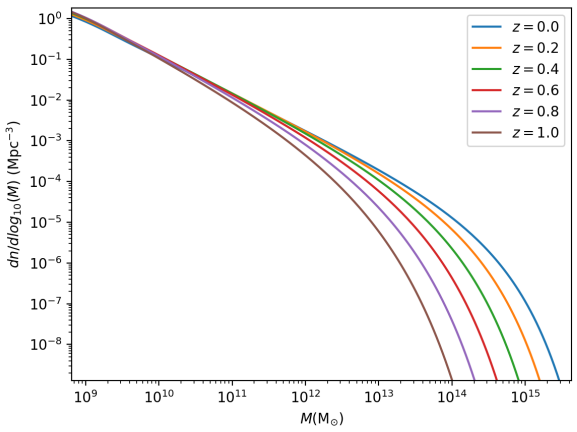


Figure 7: The Sheth-Torman mass function with respect to the comoving volume plotted at multiple redshifts.

3.2 Virtual Image

The pseudo-random redshifts and masses produced were most common at $z \sim 1$ and $m \sim 10^8 M_\odot$, respectively. This was to be expected as a higher redshift corresponds to a greater volume for an observer and less massive objects are much more common. Figure 8 conveys the rarity of massive halos and the exceptionally high abundance of low mass halos.

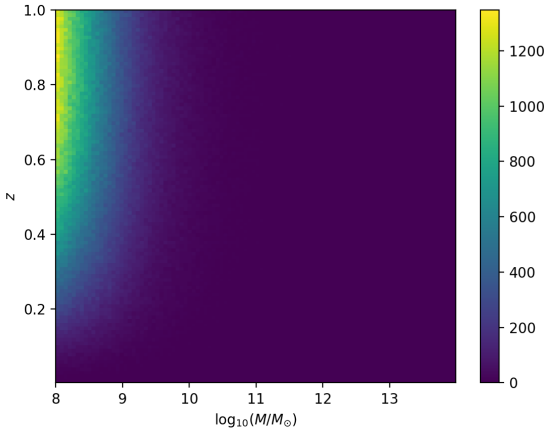


Figure 8: Drawn samples from the 2-dimensional probability density function of redshift, z, and halo mass, M.

An issue immediately arose from this data. The abundance of low mass halos meant that most of the data had a mass value that was too small for the stellar mass fitting function. This produced very low mass galaxies ($\sim 10^4 M_\odot$) and went on to generate nonsensical results for galactic radius and magnitude. To counter this, the mass range was reduced such that the lowest mass was $10^{10} M_\odot$.

Once this was implemented the stellar masses of the virtual galaxies were much more realistic. The random data was well distributed around the fitting function ($z = 0$), however, there was a larger variation below the theoretical line. The data was pulled in this direction as a consequence of higher redshifts being more likely which in turn means that less massive galaxies are more probable. Scatter was then introduced into the calculation of stellar mass as provided by Moster et al. (2010). This increased the stellar mass uncertainty by a fairly significant degree, especially at lower masses.

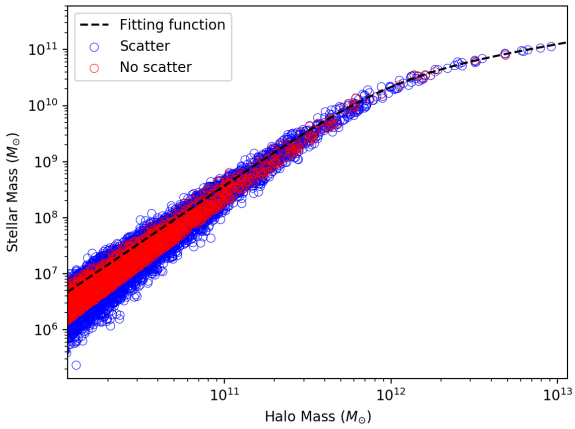


Figure 9: Dark matter halo mass to galactic stellar mass relation. Additional scatter is more effective at lower masses.

Next, the galactic radii were calculated. At this point, the uncertainties started to become more significant than the variation from the random data. Furthermore, figure 10 shows that the additional scatter is now in multiple dimensions that will carry through to other observable properties in different ways. Again, the uncertainties on the values are much more significant at lower mass and radii.

As discussed previously, a straightforward fitting function to link stellar mass and luminosity was not available in the literature. Consequently, a rough average value of their ratio was chosen to relate the two properties. An effort was made to add some uncertainty to this value. A strict range of $1 \leqslant M/L \leqslant 10$ and an average of 3.9 was chosen for the scatter. This function was best represented by a gamma distribution (see Appendix VI). Random values were drawn from this distribution and were treated as the mass-luminosity ratio for each respective galaxy.

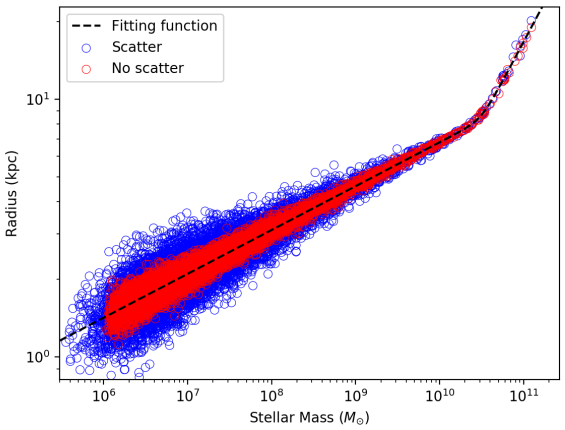


Figure 10: The stellar mass to galactic radius relation. Again, additional scatter is more significant at lower values of mass and radius. The additional scatter is as significant as the randomisation uncertainties.

From here, the methods for determining the observable properties are very well known and the physics is well understood. Thus, no more scatter was added when determining the properties in the blue boxes in figure 4. Once the apparent magnitude data had been calculated it became clear that the galaxies were exceptionally faint. The vast majority of the galaxies had an apparent magnitude of ~ 32. This level of brightness is essentially undetectable and subsequently, this data doesn't produce a useful virtual image. The obvious way to resolve this was to yet again reduce the mass range by increasing the minimum mass to $10^{12}\mathrm{M}_\odot$. Consequently, the most abundant halo mass shifted to this value and figure 9 and 10 were shifted up two orders of magnitude on the x-axes. Moreover, this reduced the variation of the additional scatter as this reduced as massed increased (around the broken point in the power-laws). Apparent size and magnitude were calculated for the new pseudo-random data. This produced apparent magnitudes of ~ 22 which is much more reasonable.

To examine the reliability of the observable fitting functions were, a correlation test of the random apparent magnitude distribution, $\mathcal{M}$, and size, Θ, were conducted with and without scatter. Initially, the covariance of the two sets of data was calculated using the `numpy` module.

$$\mathrm{cov}(\Theta, \mathcal{M}) = \frac{1}{n}\sum_{i=1}^{n} E\Big[(\theta_i - E[\Theta])(m_i - E[\mathcal{M}])\Big] \quad (32)$$

Where n is the number of data values, θ_i and m_i are the discrete values of apparent size and magnitude, respectively. Subsequently, the correlation can be calculated and is defined as,

$$\rho = \frac{\mathrm{cov}(\Theta, \mathcal{M})}{\sigma_\Theta \sigma_\mathcal{M}} \quad (33)$$

where σ represents the standard deviation of the random distributions. By defining ρ_n as the correlation coefficient of the non-scattered data and ρ_s for the scattered data, the following results were found.

$$\rho_n = -0.75 \qquad \rho_s = -0.73$$

This conveys that both sets of data are well correlated. As expected the non-scattered data shows a stronger correlation but only marginally. This suggests that the additional scatter of the observable parameters have a negligible effect on the parameter's uncertainties. Two 2-dimensional histograms of the data were plotted to illustrate the significance of the additional scatter in figure 11 and 12. Figure 11 displays a tight relation between apparent magnitude and size with most of the data following a narrow line which corresponds to galaxies of $\sim 10^{12}\mathrm{M}_\odot$ at $z \sim 1$. However, higher mass and less distant galaxies deviate from this line which produces a significant dispersion. As well as the covariance, `numpy` provides the variance of each individual parameter (table 5).

Variation	No scatter	Scatter
$\sigma^2(\Theta)$	2.40	2.42
$\sigma^2(\mathcal{M})$	1.92	2.00

Table 5: A comparison of how the variation of apparent size and magnitude changes with additional scatter.

This further verifies that the additional scatter is insignificant in producing uncertainty. Despite the scattered data being more evenly spread, the histograms convey that the range of variance of each parameter is similar for both sets of data. Therefore it can be concluded that the source of the covariance is the mass and redshift range rather than the additional scatter.

It should be mentioned that the correlation coefficient assumes that parameters are related linearly. Figure 11 shows that there is a slight upward curve towards higher magnitudes. Consequently, the correlation coefficient will be somewhat inaccurate at these values. This will affect the non-scattered data more significantly as the additional scatter shifts the data such that the linearity of the relationship increases.

Furthermore, the randomisation variance is dependent on the mass and redshift range. Therefore, in order for this economical method to viable, the mass and redshift range have to be sufficiently large such that it outweighs the contribution of additional scatter. As discussed previously, the additional scatter was dependent on the value of the galactic properties. For instance, the scatter appeared to become more significant at lower masses and radii (see figure 10). Therefore, the reliability of this method is alos dependent on the mass and redshift values contained within the range.

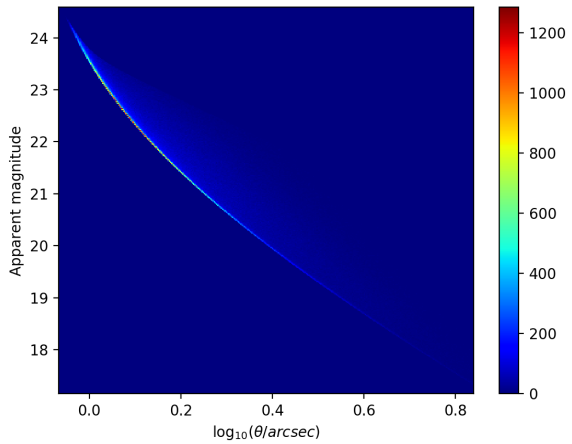


Figure 11: The apparent magnitude and size of galaxies in the mass range of $(10^{12} - 10^{16})\mathrm{M}_{\odot}$ out to $z = 1$. This data has an absence of scatter from the fitting functions.

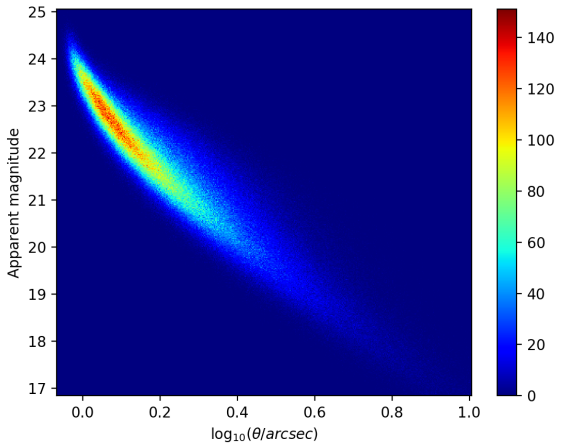


Figure 12: The apparent magnitude and size of galaxies in the mass range of $(10^{12} - 10^{16})\mathrm{M}_{\odot}$ out to $z = 1$. This data contains additional scatter from the observable fitting functions.

It would appear that this streamlined method of producing a virtual image is viable within certain conditions. However, it would be constructive to examine how realistic the observable parameters are. The method was far too simplistic and contained too many oversights to yield precise results. However, a rudimentary test was applied to the Andromeda galaxy to identify the accuracy of the observable parameters derived in this project. Andromeda has a mass of $\sim 10^{12}\mathrm{M}_{\odot}$ and therefore would be a very common galaxy in the data used here. By fixing the distance of Andromeda to $z = 1$, it is reasonable to presume that it would appear in a common bin of figure 12. Typically, a galaxy of this mass and redshift has an apparent magnitude of ~ 22. using these values and ignoring the K-correction, the absolute magnitude of Andromeda can be calculated using equation 30. Where $d \approx 3400$ Mpc for $z = 1$ according to the `astropy` cosmology used in this project. This gives an absolute magnitude of -21.12 which is within remarkable agreement with its true value.

4 Future Application

In order to constrain cosmological parameters such as the ones used in this project, deep space lensing surveys are required to observe the large-scale structure of the universe at high redshift. In the advent of these surveys, statistical models of cosmic shear are in high demand such that galactic shape measurement algorithms can be developed to high precision. Moreover, to better understand how galactic parameters (including shape) evolve with time, large quantities of galactic data are required to develop distributions of these parameters. Therefore, a high abundance of simulated virtual images is needed. As discussed in the objective, the typical, method of producing such data is through dark matter simulations which require enormous computational power. Subsequently, it may be imperative to develop economical techniques such as the one used in this project in order to develop the galactic shape algorithms in time for the upcoming surveys.

5 Conclusions

Our understanding of the foundation of the universe still remains exceptionally limited. So far only 5% of the universal structure can be accounted for by the Standard Model. Despite this, there has been an extraordinary and somewhat successful effort to infer the mechanics and properties of the unknown universe. An example of this is the dark matter halo function. This is an essential cosmological feature which can provide insights to the events of the early universe. In an effort to increases our knowledge of cosmology, we must look deep into space to observe the universe as it once was and identify patterns to describe the way it changes. However, light is obscured, scattered and perturbed by its contents. To understand the mechanisms that cause such effects, a high abundance of simulated galactic data is required. Computational limitations constrain the volume of data we can produce to insufficient levels.

This project employed a less rigorous and computationally expensive approach. In summary, this method developed galactic properties from pseudo-random mass and redshift data sampled from PDFs that were derived from the halo mass function. This data was cast down a dependency tree of well known astrophysical relations and fitting functions from the literature to obtain observable galactic data. This data was then examined to determine whether such a method is viable.

In this project, the mass function was constructed based upon the statistical variations of the density pertur-

bation field, linear theory and incorporated the ellipsoidal collapse of the Sheth-Torman formalism. Although this doesn't employ the more complex mechanisms such as non-linear collapse, it was in the nature of this project to apply a more streamlined approach. The constructed mass function was in good agreement with the literature.

1 million pseudo-random DMHs were generated. Using the halo's mass and redshift, the observable properties of the halo's host galaxy were determined. Scatter from the literature was added for the derivation of each respective property except when the physics is well understood. Scattered and non-scattered data of apparent magnitude and apparent size were compared to determine if the economical method employed in this project was viable. It was found that the covariance (correlation) of these galactic properties were minimally increased (decreased) when additional scatter was added to the fitting functions. The main source of the uncertainty in the galactic properties was the randomisation of the DMH mass and redshift values. Therefore, it can be concluded that this method is feasible and could provide sufficiently accurate data at a far less computational expense. However, some caveats are partnered with this conclusion. A sufficient range of mass and redshift is required such that the scattering is outweighed by the randomisation. Furthermore, the magnitude of scatter is dependent on parameter values contained within this range. However, this scatter is usually greater at the values which fall outwith the applicable range of the fitting functions.

In our search to define the cosmology that governs the universe, it is essential that we develop the tools required to decipher high redshift data that is soon to come. To do so, the employment of this method may prove to be imperative in the near future of cosmology.

6 Acknowledgements

I would like to express my utmost gratitude to my supervisor Dr Bryan Gillis for his assistance, guidance and most importantly his patience throughout this project.

I would also like to extend my gratitude to Professor Bob Mann for his cosmology course notes which aided me in my understanding of physical cosmology.

This project was entirely founded upon the `python` programming language. I would like to show my appreciation for `python`'s modules `numpy`, `scipy` (Jones et al., 2001–) and `astropy` (Price-Whelan et al., 2018) for their assistance with data handling, numerical integration as well as interpolation and units, respectively.

I would like to acknowledge that the theory in sections 1.2, 1.4 and 1.5 was based upon the textbooks of Serjeant (2010) and Liddle and Lyth (2000).

Lastly, I would like to thank Associate Professor Frank Van Den Bosch of Yale University for his lectures on Theory of Galaxy Formation[5]. These lectures were extremely helpful and fortified my understanding of large-scale structure and the dark matter halo mass function. Sections 1.5.3 and 1.5.4 were based on this content.

[5] `http://campuspress.yale.edu/vdbosch/teaching/`

Bibliography

P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, and et al. Planck 2015 results. XIII. Cosmological parameters. , 594:A13, Sept. 2016. doi: 10.1051/0004-6361/201525830.

N. Aghanim, S. Majumdar, and J. Silk. Secondary anisotropies of the CMB. *Reports on Progress in Physics*, 71(6):066902, may 2008. doi: 10.1088/0034-4885/71/6/066902. URL https://doi.org/10.1088%2F0034-4885%2F71%2F6%2F066902.

J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. The Statistics of Peaks of Gaussian Random Fields. *Astrophys. J.*, 304:15–61, 1986. doi: 10.1086/164143.

E. F. Bell and R. S. de Jong. Stellar mass-to-light ratios and the tully-fisher relation. *The Astrophysical Journal*, 550(1):212–229, mar 2001. doi: 10.1086/319728. URL https://doi.org/10.1086%2F319728.

A. J. Benson, R. G. Bower, C. S. Frenk, C. G. Lacey, C. M. Baugh, and S. Cole. What shapes the luminosity function of galaxies? *The Astrophysical Journal*, 599(1):38–49, dec 2003. doi: 10.1086/379160. URL https://doi.org/10.1086%2F379160.

R. G. Bower, A. J. Benson, R. Malbon, J. C. Helly, C. S. Frenk, C. M. Baugh, S. Cole, and C. G. Lacey. Breaking the hierarchy of galaxy formation. , 370:645–655, Aug. 2006. doi: 10.1111/j.1365-2966.2006.10519.x.

E. F. Bunn and M. White. The 4 Year COBE Normalization and Large-Scale Structure. , 480:6–21, May 1997. doi: 10.1086/303955.

S. Eales. Practical Cosmology with Lenses. *Mon. Not. Roy. Astron. Soc.*, 446:3224–3234, 2015. doi: 10.1093/mnras/stu2214.

S. M. Faber and R. E. Jackson. Velocity dispersions and mass-to-light ratios for elliptical galaxies. , 204: 668–683, Mar. 1976. doi: 10.1086/154215.

C. S. Frenk, S. Cole, A. Jenkins, J. M. Colberg, N. Yoshida, S. D. M. White, A. E. Evrard, and H. M. P. Couchman. The mass function of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 321(2):372–384, 02 2001. ISSN 0035-8711. doi: 10.1046/j.1365-8711.2001.04029.x. URL https://doi.org/10.1046/j.1365-8711.2001.04029.x.

C. S. Frenk, C. M. Baugh, C. G. Lacey, C. d. P. Lagos, J. C. Helly, P. Norberg, R. Bielby, R. G. Bower, S. Cole, V. Gonzalez-Perez, A. I. Merson, and A. J. Benson. Lightcone mock catalogues from semi-analytic models of galaxy formation â I. Construction and application to the BzK colour selection. *Monthly Notices of the Royal Astronomical Society*, 429(1):556–578, 12 2012.

A. Grado, E. Puddu, F. Getman, F. L. Barbera, M. Brescia, N. R. Napolitano, M. A. Raj, N. Roy, S. Cavuoti, C. Vellucci, G. Longo, M. Capaccioli, V. Amaro, G. Covone, E. Valentijn, G. VerdoesÂ Kleijn, M. Radovich, C. Tortora, and K. Kuijken. Evolution of galaxy sizeâstellar mass relation from the Kilo-Degree Survey. *Monthly Notices of the Royal Astronomical Society*, 480(1):1057–1080, 07 2018. ISSN 0035-8711. doi: 10.1093/mnras/sty1917. URL https://doi.org/10.1093/mnras/sty1917.

E. Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences*, 15(3):168–173, 1929. ISSN 0027-8424. doi: 10.1073/pnas.15.3.168. URL https://www.pnas.org/content/15/3/168.

E. P. Hubble. Extragalactic nebulae. , 64, Dec. 1926. doi: 10.1086/143018.

E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. URL http://www.scipy.org/. [Online; accessed 2019].

A. R. Liddle and D. H. Lyth. *Cosmological inflation and large scale structure*. 2000. ISBN 0521575982, 9780521575980, 9780521828499.

C.-P. Ma. Analytical Approximation to the Nonlinear Power Spectrum of Gravitational Clustering. , 508: L5–L8, Nov. 1998. doi: 10.1086/311711.

T. B. Miller, P. van Dokkum, L. Mowla, and A. van der Wel. A new view of the size–mass distribution of galaxies: Using r 20 and r 80 instead of r 50. *The Astrophysical Journal*, 872(1):L14, feb 2019. doi: 10.3847/2041-8213/ab0380. URL https://doi.org/10.3847%2F2041-8213%2Fab0380.

B. P. Moster, R. S. Somerville, C. Maulbetsch, F. C. van den Bosch, A. V. Macciò, T. Naab, and L. Oser. CONSTRAINTS ON THE RELATIONSHIP BETWEEN STELLAR MASS AND HALO MASS AT LOW AND HIGH REDSHIFT. *The Astrophysical Journal*, 710(2):903–923, jan 2010. doi: 10.1088/0004-637x/710/2/903. URL https://doi.org/10.1088%2F0004-637x%2F710%2F2%2F903.

L. Mowla, A. van der Wel, P. van Dokkum, and T. B. Miller. A mass-dependent slope of the galaxy size–mass relation out to z $\sim$ 3: Further evidence for a direct relation between median galaxy size and median halo mass. *The Astrophysical Journal*, 872(1):L13, feb 2019. doi: 10.3847/2041-8213/ab0379. URL https://doi.org/10.3847%2F2041-8213%2Fab0379.

S. G. Murray, C. Power, and A. S. G. Robotham. How well do we know the halo mass function? *Monthly Notices of the Royal Astronomical Society: Letters*, 434(1):L61–L65, 07 2013. ISSN 1745-3925. doi: 10.1093/mnrasl/slt079. URL https://doi.org/10.1093/mnrasl/slt079.

A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. , 142: 419–421, July 1965. doi: 10.1086/148307.

W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. , 187:425–438, Feb. 1974. doi: 10.1086/152650.

A. M. Price-Whelan, B. M. Sipőcz, Günther, and A. Contributors. The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. , 156:123, Sept. 2018. doi: 10.3847/1538-3881/aabc4f.

Y. P. Pskovskii. The Mass-Luminosity Relation for Galaxies and the Mean Density of Metagalactic Matter. , 9: 253, Oct. 1965.

A. Raig, E. Salvador-SolÃ©, and G. GonzÃ¡lez-Casado. Testing the modified Press-Schechter model against N-body simulations. *Monthly Notices of the Royal Astronomical Society*, 327(3):939–948, 11 2001. ISSN 0035-8711. doi: 10.1046/j.1365-8711.2001.04808.x. URL `https://doi.org/10.1046/j.1365-8711.2001.04808.x`.

D. S. Reed, R. Bower, C. S. Frenk, A. Jenkins, and T. Theuns. The halo mass function from the dark ages through the present day. , 374:2–15, Jan. 2007. doi: 10.1111/j.1365-2966.2006.11204.x.

M. S. Roberts and A. H. Rots. Comparison of Rotation Curves of Different Galaxy Types. , 26:483–485, Aug. 1973.

A. Rodríguez-Puebla, J. R. Primack, P. Behroozi, and S. M. Faber. Is main-sequence galaxy star formation controlled by halo mass accretion? , 455:2592–2606, Jan. 2016. doi: 10.1093/mnras/stv2513.

V. C. Rubin. Rotation curves of high-luminosity spiral galaxies and the rotation curve of our galaxy. *Symposium - International Astronomical Union*, 84:211â220, 1979. doi: 10.1017/S0074180900014492.

S. Serjeant. *Observational Cosmology*. Sept. 2010.

S. Shen, H. J. Mo, S. D. M. White, M. R. Blanton, G. Kauffmann, W. Voges, J. Brinkmann, and I. Csabai. The size distribution of galaxies in the Sloan Digital Sky Survey. , 343:978–994, Aug. 2003. doi: 10.1046/j.1365-8711.2003.06740.x.

R. K. Sheth, H. J. Mo, and G. Tormen. Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 323(1):1–12, 2001. ISSN 0035-8711.

J. Sulentic, L. Verdes-Montenegro, M. Argudo-FernÃ¡ndez, and M. FernÃ¡ndez Lorenzo. The stellar massâsize relation for the most isolated galaxies in the local Universeâ. *Monthly Notices of the Royal Astronomical Society*, 434(1):325–335, 07 2013. ISSN 0035-8711. doi: 10.1093/mnras/stt1020. URL `https://doi.org/10.1093/mnras/stt1020`.

R. B. Tully and J. R. Fisher. A new method of determining distances to galaxies. , 54:661–673, Feb. 1977.

R. Wechsler and J. L. Tinker. The connection between galaxies and their dark matter halos. *Annual Review of Astronomy and Astrophysics*, 56, 04 2018. doi: 10.1146/annurev-astro-081817-051756.

Wikipedia contributors. Angular diameter distance, 2017. URL `https://en.wikipedia.org/wiki/Angular_diameter_distance`. [Online; accessed 30-03-2019].

X. Yang, H. J. Mo, and F. C. van den Bosch. Constraining galaxy formation and cosmology with the conditional luminosity function of galaxies. , 339:1057–1080, Mar. 2003. doi: 10.1046/j.1365-8711.2003.06254.x.

Appendices

Appendix I

Taylor expansions of relations in equation 12:

$$\lim_{\theta\to 0} r(\theta) = A\left(\frac{\theta^2}{2} + \frac{\theta^4}{24}\right)$$
$$\lim_{\theta\to 0} t(\theta) = B\left(\frac{\theta^3}{6} + \frac{\theta^5}{120}\right) \tag{34}$$

factorise t term,

$$\lim_{\theta\to 0} t(\theta) = \frac{B\theta^3}{6}\left(1 + \frac{\theta^2}{20}\right) \tag{35}$$

substitute in dominant t term for θ^2,

$$\lim_{\theta\to 0} t(\theta) = \frac{B\theta^3}{6}\left(1 + \frac{1}{20}\left(\frac{6t}{B}\right)^{2/3}\right) \tag{36}$$

rearrange for θ and substitute into $r(\theta)$,

$$r(\theta \ll 1) \approx \frac{A}{2}\left(\frac{6t}{B}\right)^{2/3}\left(1 - \frac{1}{20}\left(\frac{6t}{B}\right)^{2/3}\right). \tag{37}$$

Then by substituting r into $\rho = \frac{M}{(4/3)\pi r^3}$ and dropping higher order terms,

$$\rho \approx \frac{1}{6\pi t^2 G}\left(1 + \frac{3}{20}\left(\frac{6t}{B}\right)^{2/3}\right) \tag{38}$$

where $A^3 = GMB^2$. From this, it is clear that the perturbation is

$$\delta \approx \frac{3}{20}\left(\frac{6t}{B}\right)^{2/3} \tag{39}$$

the time at which r is at its maximum value is when $\theta = \pi$ (figure 3) and therefore, $B = t_{max}/\pi$ (equation 12). Finally,

$$\delta = \frac{3}{20}\left(\frac{6\pi t}{t_{max}}\right)^{2/3} \tag{40}$$

Appendix II

Each redshift parameterisation for the free parameters of equation 26 are presented here.

$$\log M_1(z) = \log M_1(0)(z+1)^{0.019} \tag{41}$$

$$A_0(z) = \log A_0(0)(z+1)^{0.0282} \tag{42}$$

$$\gamma(z) = \gamma(0)(z+1)^{-0.26} \tag{43}$$

$$\beta(z) = \beta(0) + 0.17z \tag{44}$$

All $z = 0$ values are the ones given in table 3.

Appendix III

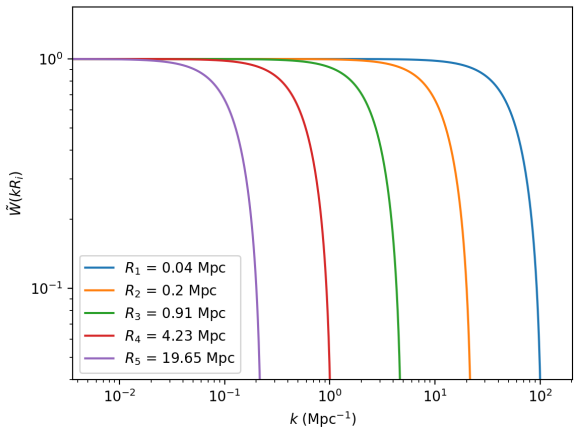


Figure 13: Each line represents the window function cut-off point for values of k at a fixed value of R. The cut-off moves from right to left as R increases. This is plotted using the power spectrum at $z = 1$.

Appendix IV

The following is the definition of angular diameter distance, d_A. This was implemented in the project by using `astropy`.

$$d_A = \frac{d_C}{1+z}, \qquad \Omega_k = 0 \tag{45}$$

Where d_C is the comoving distance defined by,

$$d_C = \int_0^z \frac{dz'}{\sqrt{\Omega_m(1+z')^3 + \Omega_\Lambda}}, \qquad \Omega_r = 0 \tag{46}$$

where Ω_r is the energy density of radiation. How d_A changes with redshift is illustrated below.

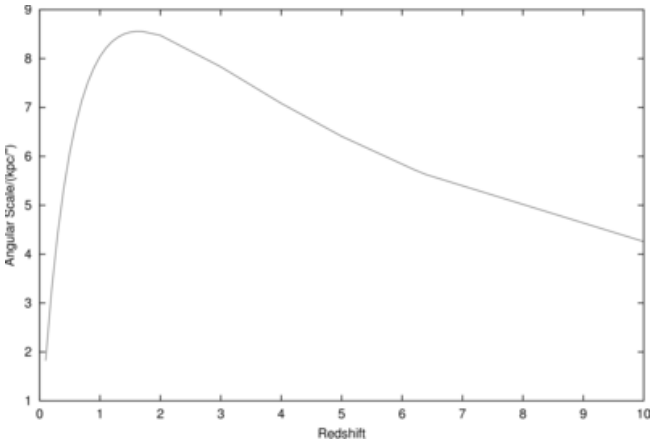


Figure 14: Intuitavley, d_A increases with redshift (distance) up until $z \sim 1$. It then decreases due to the fact the expansion of the universe has become more significant than r^{-2} scaling law. Consequently, the aparent size of an object increases with distance after this turning point. (Wikipedia contributors, 2017).

Appendix V

The light-travel time describes the time it takes for a photon to reach an observer from a source whilst accounting for the expansion of the universe. The light-travel distance is simply the light-travel time multiplied by c.

$$d_T = \frac{c}{H_0} \int_0^z \frac{dz'}{(1+z')\sqrt{\Omega_m(1+z')^3 + \Omega_\Lambda}} \tag{47}$$

The rate at which this distance changes with respect to redshift is,

$$\frac{d}{dz}(d_T) = \frac{c}{H_0(1+z)\sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}} \tag{48}$$

Appendix VI

The following gamma function has an average of 3.9 and a scale of 0.25.

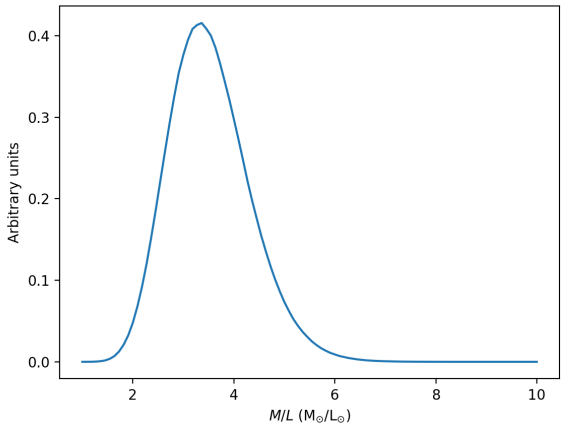


Figure 15: A gamma function which was constructed to produce scatter for the stellar mass to luminosity ratio.

Appendix VI

This is a link to the GitHub repository containing the code used in this project: `https://github.com/Smithy3001/97-design.git`