

Galaxy Clusters as Cosmological probes

A Thesis

submitted to

Indian Institute of Science Education and Research Pune

in partial fulfillment of the requirements for the

BS-MS Dual Degree Programme

by

Arindam Sharma



Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2020

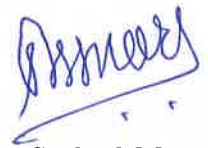
Supervisor: Dr. Surhud More

© Arindam Sharma 2020

All rights reserved

Certificate

This is to certify that this dissertation entitled 'Galaxy Clusters as Cosmological probes' towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Arindam Sharma at the Inter-University Centre for Astronomy and Astrophysics, Pune under the supervision of Dr. Surhud More, Associate Professor, Inter-University Centre for Astronomy and Astrophysics, during the academic year 2019-2020.



Dr. Surhud More

Committee:

Dr. Surhud More

Dr. Suneeta Vardarajan

This thesis is dedicated to my parents, Iswar and Gayatree Sharma

Declaration

I hereby declare that the matter embodied in the report entitled 'Galaxy Clusters as Cosmological probes' are the results of the work carried out by me at the Inter-University Centre for Astronomy and Astrophysics, Pune, under the supervision of Dr. Surhud More and the same has not been submitted elsewhere for any other degree.

Arindam Sharma

Arindam Sharma

Acknowledgments

First and foremost, I would like to express my heartfelt gratitude to my supervisor and mentor, Dr. Surhud More for helping and guiding me throughout the project. The countless hours I have spent in his office debugging codes and brainstorming ideas have genuinely helped me to achieve the goals of my project. I am forever indebted to his infinite patience with my rather frequent mistakes. His kind words have always helped me stay motivated throughout the project and I am also grateful for his critical evaluation of presentation and writing skills. In every sense, he is the ideal mentor and I feel fortunate to have him as my supervisor. I would also like to thank IUCAA for providing me with the opportunity to work in the serene and peaceful environment there.

I want to thank my thesis advisory committee member, Dr. Suneeta Vardarajan. I would also like to acknowledge Amit Kumar and Divya Rana for helping me with the implementation of emcee in Python.

It's been an excellent opportunity to study at the Indian Institute of Science Education and Research (IISER), Pune. I am grateful to all the professors at IISER Pune who have taught me since their beautiful teaching further increased my passion for pursuing physics. I would also like to acknowledge my faculty advisor at IISER Pune, Prof. Ramakrishna G. Bhat, whose constant support has helped me through the initial years at IISER. My time at IISER has only been incredible only because of my wonderful friends. I want to thank all of them while refraining from naming them as it will take half a page to do so. Last but not least, I want to thank my parents from the bottom of my heart as I am indebted to them forever. They have always encouraged me to pursue a career in science, the passion for which was built by my father since I was very young. Their unconditional love and unbelievable faith in my calibre never let me feel down in my life.

Abstract

Galaxy clusters form the largest gravitationally bound structures in the Universe. Galaxy clusters form at rare peaks in the initial density fluctuation field in the Universe. Therefore their abundance is very sensitive to the cosmological parameters which describe our Universe. The goals of this thesis are to understand how to characterize the abundance of galaxy clusters from observations and constrain cosmological parameters from these observations.

The primary cosmological parameters of our interest are Ω_m , the matter-density parameter and σ_8 , which is related to the amplitude of the linear power spectrum of density fluctuations in the early Universe. The abundance of massive dark matter halos in which galaxy clusters form is very sensitive to these cosmological parameters. We develop an analytical and numerical framework to infer the cosmological parameters of interest from the X-ray observations of galaxy clusters after marginalizing over the halo mass-observable scaling relations and their scatter.

Our modeling scheme uses inputs from N-body simulations and follow-up observations of galaxy clusters to fit the abundances of galaxy clusters selected using the ROSAT survey data. We use a Bayesian framework to infer the posterior distribution of cosmological parameters of interest given the observations using Monte Carlo Markov Chain (MCMC) techniques. Although our framework is general, in this thesis, we present results using certain simplistic priors on the halo-mass observable relations of galaxy clusters. We obtain $\Omega_m = 0.280^{+0.052}_{-0.046}$ and $\sigma_8 = 0.721^{+0.030}_{-0.033}$. We also observe the well-known degeneracy between Ω_m and σ_8 in the posterior distributions. In the future, we will improve on this estimate of the cosmological parameters by marginalizing over the scatter of the mass-observable relation, appropriately. Future work will involve using weak lensing data to better calibrate the masses used in the mass-observable relation. Upcoming X-ray surveys like those conducted with the X-ray telescope eROSITA, combined with follow-up studies of these galaxy clusters using the Large Synoptic Survey Telescope (LSST) to constrain the mass-observable relationship can place much stronger constraints on the cosmological parameters.

Contents

| | |
|--|-----------|
| Abstract | xi |
| 1 Galaxies and Galaxy Clusters | 6 |
| 1.1 Dark Matter Halos | 7 |
| 1.2 The Cosmological Parameters | 8 |
| 1.3 Time dependence of the scale factor | 11 |
| 1.4 Evolution of the Hubble Parameter | 15 |
| 2 Structure Formation and the Cosmological Parameters | 16 |
| 2.1 Jeans Instability | 17 |
| 2.2 Linear Perturbation theory | 18 |
| 2.3 The Power Spectrum as a Cosmological Probe | 20 |
| 2.4 The Mass function and its relation to the Power Spectrum | 23 |
| 2.5 Press-Schechter Mass Function | 24 |
| 2.6 Luminosity function | 26 |
| 3 Modelling the Luminosity function and Comparing it with Data | 28 |
| 3.1 Theoretical Model of the Luminosity function | 28 |
| 3.2 Luminosity function from data | 32 |

| | | |
|----------|---|-----------|
| 3.3 | Getting the error bars: Jackknife | 35 |
| 4 | Monte Carlo Markov Chain | 38 |
| 4.1 | Random Sampling | 39 |
| 4.2 | Bayesian Inference | 40 |
| 4.3 | Introduction to emcee | 40 |
| 4.4 | Implementing emcee | 41 |
| 4.5 | Autocorrelation Analysis | 45 |
| 5 | Results and Discussion | 46 |
| 6 | Conclusion | 52 |
| A | Hydrostatic Equilibrium | 54 |
| B | More Monte Carlo Methods | 55 |
| B.1 | Importance sampling | 55 |
| B.2 | Rejection sampling | 56 |
| B.3 | Metropolis-Hastings algorithm | 57 |
| B.4 | Gibbs sampling | 58 |
| C | χ^2 Distribution | 60 |

List of Figures

| | | |
|-----|--|----|
| 1.1 | Three different types of galaxies | 6 |
| 1.2 | Schematic representation of the time dependence of the scale factor | 14 |
| 2.1 | Evolution of the matter power spectrum of density fluctuations during the radiation dominated era | 22 |
| 2.2 | The Power Spectrum measured from different observations | 23 |
| 2.3 | Distribution of primordial overdensities | 24 |
| 2.4 | Mass function obtained from numerical simulations | 26 |
| 2.5 | X-ray Luminosity function in Stanek et al. (2006) | 27 |
| 3.1 | Luminosity function model | 31 |
| 3.2 | Luminosity-redshift distribution of REFLEX clusters of galaxies | 33 |
| 3.3 | Observed Luminosity function from REFLEX data | 34 |
| 3.4 | Area of the sky covered by the REFLEX survey | 36 |
| 3.5 | Luminosity function from REFLEX data plotted with errorbars calculated using jackknife resampling | 37 |
| 4.1 | Luminosity-redshift distribution of BCS clusters | 44 |
| 5.1 | The model with scatter and observed luminosity function obtained from REFLEX data in the range $L = 10^{42} - 10^{46} h_{70}^{-2}$ erg/s | 47 |

| | | |
|-----|--|----|
| 5.2 | The simple model (without the scatter) and the observed X-ray luminosity function compared to the luminosity function given in Allen et al. (2003) . . . | 49 |
| 5.4 | MCMC chains | 49 |
| 5.3 | Posterior probability distribution of Ω_m and σ_8 | 50 |
| 5.5 | The X-ray luminosity function obtained for a model with scatter compared to Allen et al. (2003) | 51 |
| B.1 | Metropolis-Hastings algorithm in two dimensions | 57 |
| B.2 | Gibbs sampling | 59 |

List of Tables

| | | |
|-----|---|----|
| 5.1 | The cluster counts of REFLEX and combined BCS samples | 48 |
|-----|---|----|

Introduction

Dark matter halos form the backbone of the large scale structures of the Universe. The most massive dark matter halos ($> 10^{14} h^{-1} M_{\odot}$) are hosts to galaxy clusters, an extensive collection of galaxies bound by the gravitational potential of dark matter. Galaxy clusters act as signposts of otherwise invisible dark matter halos, which can be as massive as $10^{15} M_{\odot}$ [1]. Structure formation proceeds hierarchically, where small dark matter halos form first and merge to form bigger and more massive dark matter halos [1]. Baryonic processes such as star formation within dark matter halos lead to the formation of galaxies. Clusters of galaxies can be readily identified from galaxy observations in the optical or observations of hot gas in X-rays [1]. These massive structures can act as beaming beacons of the cosmological model that describes our Universe. The inference of cosmological parameters from various observations is crucial to verify the theoretical framework of modern cosmology. The cosmological parameters are of paramount importance and give information about the expansion history and the evolution of structure in the Universe, as shown in section 1.3. They are also crucial to infer distances to far-away galaxies from their redshifts [2].

After the Big Bang, the Universe was a hot dense plasma consisting of a soup of fundamental particles like quarks, which later combined to form the protons and neutrons [3]. The extremely high temperatures of the early Universe didn't allow protons and neutrons to combine with electrons to form neutral atoms. This meant that photons were easily scattered by free electrons via Compton scattering [4]. Thus, photons were strongly coupled to matter for around 370,000 years until the expansion of the Universe cooled it down sufficiently for neutral atoms to form, leading to radiation decoupling from matter [4]. This radiation permeates all of space today and is detected as the Cosmic Microwave Background (CMB). The CMB can place strong constraints fraction of energy densities of different components like radiation, dark matter and dark energy and this has been achieved up to a great extent

by the Planck satellite [5].

Galaxy clusters provide a path to an independent measurement of these parameters. To determine the cosmological parameters from galaxy clusters, we resort to the halo mass function, which is the number density of dark matter halos as a function of their masses [1]. It can be used to infer information about the cosmological parameters of interest after marginalizing over the mass-observable relation because the mass function is closely related to the linear power spectrum of density perturbations of the early Universe, which led to the formation of large scale structures [1]. The shape of the power spectrum is highly sensitive to the cosmological parameters, particularly the matter density parameter, Ω_m and the σ_8 . σ_8 is the average amplitude of primordial density fluctuations and is directly proportional the amplitude of the linear power spectrum. Thus, the mass function can be used to infer the cosmological parameters [6]. But measuring the masses of clusters is quite a daunting task. Therefore, we use measurements of X-ray luminosities of galaxy clusters to obtain the number density of halos as a function of their luminosities, also known as the X-ray luminosity function. The X-ray luminosity function is then related to the mass function through a mass-luminosity relation [7].

In this project, we make theoretical predictions of the X-ray luminosity function from a mass function obtained from N-body simulations. This theoretical model is then used to fit the observed luminosity function obtained from the X-ray luminosity measurements of the REFLEX (ROSAT-ESO Flux-Limited X-ray) [8] survey of galaxy clusters and the RASS BCS (ROSAT All-Sky Survey Extended Brightest Cluster Sample) [9]. Both of these are sensitive to the cosmological parameters. We then use a Monte Carlo Markov Chain (MCMC) technique, which is used in fitting models to data, to place constraints on the parameters of interest [10]. MCMC techniques are widely used in cosmological parameter estimation and are usually based on the Metropolis-Hastings algorithm. Simply put, MCMC methods consist of random sampling from the parameter space to find a probability distribution of the parameters. This sampling is done through a random walk in the parameter space in the Metropolis-Hastings algorithm. We use emcee for sampling, which differs from the Metropolis-Hastings algorithm by having multiple random walkers instead of one [11].

Although Planck has been successful in measuring cosmological parameters to significantly high accuracy, there are some systematics and tensions within the Planck dataset [5]. Since galaxy clusters provide an independent measurement of these parameters, they can

act as powerful cosmological probes. Galaxy clusters have huge potential wells full of hot X-ray emitting gas. X-ray observatories that have conducted extensive sky surveys can be used to detect galaxy clusters by looking for such X-ray emission [12]. ROSAT¹ is one such observatory. It is an X-ray space telescope launched by the German Aerospace Centre. It has provided us with one of the earliest and most extensive all-sky surveys in the soft X-ray band (0.1-2.4 keV). Surveys like XMM-Newton and Chandra², launched after the success of ROSAT, have provided us with a deeper and higher resolution X-ray data of the night sky covering even a larger area than ROSAT. Current missions like eROSITA³, built by the Max Planck Institute for Extraterrestrial Physics (MPE) and launched in 2019, hope to provide strong constraints on cosmological parameters, especially dark energy, in the future.

The first chapter of this thesis gives a brief review of the background of this project. In Chapter 2, the theory of structure formation and the relation between the cosmological parameters and the mass function is explored in greater detail. In Chapter 3, we describe the details behind obtaining both the theoretical model of the X-ray luminosity function and its observational counterpart. In Chapter 4, we throw light on the MCMC methods used to obtain our final constraints on the cosmological parameters. The fifth chapter presents our results and the discussion following them. The sixth chapter concludes the thesis with an outline for future work.

¹For details, refer to <https://heasarc.gsfc.nasa.gov/docs/rosat/rosat.html>

²For details, refer to <https://cxc.harvard.edu/xraysurveys/>

³For details, refer to <https://www.mpe.mpg.de/eROSITA>

Chapter 1

Galaxies and Galaxy Clusters

The night sky is filled with wonderful objects called galaxies, which are gravitationally bound systems of stars, dust, gases and dark matter. The number of stars in a galaxy can go up to a billion and their sizes are of the order of a few kpc (kiloparsecs) to hundreds of kpc (1 parsec = 3.26 light year). Most galaxies have a supermassive black hole at the centre [12].



(a) Milky Way

(b) M87

(c) Large Magellanic Cloud

Courtesy: NASA

Figure 1.1: Three different types of galaxies

On the largest scales of the cosmos, structures like galaxy clusters exist. As their name suggests, clusters are large collections of galaxies bound together by gravity with their masses ranging from $10^{14} - 10^{15}$ solar masses in a volume with the radius of the order of few megaparsecs [1]. Smaller aggregates of galaxies are classified as groups. One such example of a group is the Local Group, of which the Milky Way is a part. The Local Group is a part

of a supercluster called Virgo, which is a part of a bigger supercluster called the Laniakea Supercluster [13]. Massive surveys coming in the next decade, like the Large Synoptic Survey Telescope (LSST)¹ will help us in mapping out the structures in our cosmos to even higher redshifts.

1.1 Dark Matter Halos

A large fraction of our Universe is composed of dark matter, a mysterious form of matter that does not interact with electromagnetic radiation and hence it can't be seen. But its gravitational effects are significant, which is how it was discovered in the first place [14].

A halo is a region that contains matter which is bound sufficiently strongly by gravity that it has decoupled from the cosmic expansion. In our current models of galaxy formation, every galaxy forms within a dark matter halo, with the halo extending much further than the optical extent of the galaxy [15]. A crucial part of clusters is the hot intracluster medium (ICM), which consists of hot gas between galaxies in the clusters with a peak temperature of the order of 10^7 K, making the ICM emit strongly in X-ray bands. This temperature is due to in-falling gas in the gravitational potential of the halo and due to line emission of heavy elements in the ICM [12].

Multiple models have been proposed to understand the particle nature of dark matter ranging from baryonic (protons and neutrons) objects like Massive Compact Halo Objects (MaCHOs), white, or brown dwarfs to non-baryonic matter like axion-like particles, Weakly Interacting Massive Particles (WIMPs), etc. MaCHOs have almost been ruled out now and WIMPs and other non-baryonic models are being investigated as a possible candidate for dark matter [16].

1.1.1 History of Dark Matter

One of the earliest evidence for dark matter was obtained by Fritz Zwicky, who obtained an unexpectedly high-velocity dispersion of eight galaxies in the Coma Cluster ($\sigma = 1019 \pm 360$ km s⁻¹) [17]. From this, he concluded that the mean density of the Coma Cluster would have

¹For more details, refer to <https://www.lsst.org/>

to be much higher than what is derived from luminous matter. Similarly, it was observed that the Virgo cluster had unexpectedly high mass [17]. In 1939, Babcock obtained the rotation curve (velocity as a function of radius) of the outer arms of the Andromeda galaxy (M31). He observed that these curves weren't falling off as expected and instead remained flat [17].

Many similar observations led to the hypothesis of dark matter. Another way to explain such anomalies is by modifying the theory of gravity itself. One such theory is Modified Newtonian Dynamics or MOND for short [14]. MOND has now been mostly discredited due to one powerful observation, the Bullet Cluster, which was formed due to the collision of two giant galaxy clusters. From weak lensing, it was observed that there was an 8-sigma significant spatial offset between the center of total mass and baryonic mass [18]. This could not be explained by alteration to gravity and hence it provided a strong piece of evidence for the existence of dark matter, although modified gravity theories have not entirely been ruled out yet [19].

1.2 The Cosmological Parameters

It is a well-known fact now that the Universe is expanding. This expansion of the Universe can be characterized by the scale factor, $a(t)$. At present, its value is set to 1 and as we go backward in time, its value decreases. More specifically, $a(t)$ characterizes how physical distances evolve with time, with the physical distances given by $a(t)d_0$, where d_0 is the distance measured at the present time.

In 1915, Einstein proposed his General Theory of Relativity, which revolutionized how we thought about gravity. His equations, $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi G T_{\mu\nu}$ are the foundations of modern cosmology [2]. With just the simplifying assumptions of homogeneity and isotropy, one can derive the Friedmann-Lemaître-Robertson-Walker metric, also known as FLRW metric, for our Universe [2]:

$$ds^2 = -c^2 dt^2 + a^2(t) \left(\frac{dr^2}{1 - k r^2} + r^2 d\Omega^2 \right) \quad (1.1)$$

Here, k characterizes the geometry of the universe. There are three possibilities: open ($k = -1$), closed ($k = +1$) and flat ($k = 0$). An open Universe has a saddle-like geometry, a closed Universe has a sphere-like geometry and a flat Universe has a Euclidean geometry.

To understand the evolution of the Universe, it is necessary to understand how the scale factor evolves with time. This is done so by the Friedmann equations, which can be directly derived from Einstein's equations by putting in the FLRW metric in the left-hand side of the equation and a perfect fluid's energy-momentum tensor on the right-hand side. A perfect fluid's energy-momentum tensor is given by $T_{\mu\nu} = (\rho + p)U_\mu U_\nu + p g_{\mu\nu}$ [2]. The Friedmann equations are [2]

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \quad (1.2)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) \quad (1.3)$$

Here, $\left(\frac{\dot{a}}{a}\right)$ is the Hubble parameter, H . It tells us about the rate of expansion of the Universe. It has dimensions of time^{-1} . If we take the first Friedmann equation (1.2) and divide it by H^2 , we get the following,

$$\frac{8\pi G}{3H^2}\rho - \frac{k}{a^2 H^2} = 1$$

We can define $\frac{3H^2}{8\pi G}$ to be a quantity called the critical density, ρ_c (for reasons explained later), so that we have a dimensionless parameter, $\Omega = \frac{\rho}{\rho_c}$.

$$\Omega - \frac{k}{a^2 H^2} = 1 \quad (1.4)$$

The density ρ in this equation can include the density of baryonic matter, dark matter, radiation and dark energy which can be represented by Ω_m , Ω_b , Ω_r , Ω_Λ respectively. So, $\Omega = \Omega_m + \Omega_b + \Omega_r + \Omega_\Lambda$. Sometimes, the term $-\frac{k}{a^2 H^2}$ is referred to as the curvature parameter, Ω_k [2]. We therefore have,

$$\sum_i \Omega_i = 1 \quad (1.5)$$

Since the sum over all Ω_i is 1, these quantities directly tell us about the fractional distribution of dark matter, baryonic matter, dark energy and radiation. The curvature parameter tells us about the intrinsic geometry of the Universe. Note that Ω_k is not a density parameter.

$$\begin{aligned}\Omega_k < 0 &\implies k > 0 \implies \text{closed} \\ \Omega_k > 0 &\implies k < 0 \implies \text{open} \\ \Omega_k = 0 &\implies k = 0 \implies \text{flat}\end{aligned}$$

Measurements from the Cosmic Microwave Background (CMB) have shown us that we live in a flat universe, as $\Omega_k \approx 0$ [5].

These Ω_i are conventionally referred to as the density parameters of the different components of the energy density in the Universe. All cosmological observations seem to point to a cosmological model consisting of just six parameters, referred to as the **cosmological parameters** [5]:

$$\begin{aligned}\Omega_m &= \text{Matter density parameter} \\ \Omega_k &= \text{Curvature parameter} \\ \omega &= \text{Dark energy equation of state parameter} \\ h &= \text{Hubble parameter} = H_0/100 \\ \theta_{CMB} &= \text{CMB Temperature} \\ \sigma_8 &= \text{Amplitude of the linear power spectrum} \\ n_s &= \text{Power spectrum index}\end{aligned}$$

Of these parameters our primary focus is Ω_m and σ_8 , as the galaxy cluster observations we will model in this thesis are primarily sensitive to Ω_m and σ_8 in the context of a flat Λ CDM model. As defined earlier, Ω_m is the fraction of matter density in the Universe. Roughly, σ_8 is the amplitude of primordial density fluctuations. After linearly growing the primordial density fluctuations to present time, if we take spheres of radius 8 Mpc randomly in the Universe and calculate the mean overdensities for each sphere, then these mean overdensities roughly follow a Gaussian distribution. The width of this Gaussian distribution is defined as the σ_8 parameter.

1.3 Time dependence of the scale factor

The energy conservation equation gives us a relation between energy density and the scale factor [2]. It is given by

$$\begin{aligned}\nabla_\mu T^\mu{}_0 &= 0 \\ -\dot{\rho} - 3\frac{\dot{a}}{a}(\rho + p) &= 0\end{aligned}\tag{1.6}$$

For perfect fluids, there is a relation between density and pressure of the form $p = w\rho$. This is known as the **equation of state** of the fluid. Using this, (1.6) becomes,

$$\frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{a}}{a}\tag{1.7}$$

which can be integrated to obtain

$$\rho \propto a^{-3(1+w)}\tag{1.8}$$

In the cosmological context, matter is defined to be any collection of non-relativistic, collision-less particles with essentially zero pressure, $p_M = 0$ [2]. Thus, $w = 0$ for a matter-dominated universe and its energy density scales with the scale factor as

$$\rho_M \propto a^{-3}\tag{1.9}$$

The equation of state for relativistic particles and radiation is given as $p_R = \rho/3$ [2]. Hence, the energy density for radiation, ρ_R is given by

$$\rho_R \propto a^{-4}\tag{1.10}$$

According to quantum field theory, the vacuum has a non-zero expectation value of energy known as the vacuum energy. Assuming, ρ_{vac} to be a constant overall space, its contribution to the energy-momentum tensor is given by $-\rho_{vac}g_{\mu\nu}$ [4]. This constant is

called the cosmological constant Λ . The negative sign is there so that T_{00} is positive since it is energy density. Now, assuming a perfect fluid, we can derive the equation of state for vacuum energy,

$$\begin{aligned}
 (\rho_{vac} + p) U_i U_i + p g_{ii} &= -\rho_{vac} g_{ii} \\
 p g_{ii} &= -\rho_{vac} g_{ii} \\
 p &= -\rho_{vac}
 \end{aligned} \tag{1.11}$$

This means $w = -1$ and thus, $\rho_{vac} \propto a^0$. Thus, we have the dependence of ρ on the scale factor in various epochs such as radiation-dominated, matter-dominated and vacuum dominated epochs. The Universe was radiation dominated initially [4]. But since ρ_R dies off faster than ρ_M , the Universe became matter-dominated. Now, ρ_M is also very small and the Universe is vacuum dominated.

1.3.1 Solving the first Friedmann equation

Now, that we have obtained a relation between the scale factor and energy density, $\frac{\rho}{\rho_0} = \left(\frac{a_0}{a}\right)^n$ we can solve the first Friedmann equation.

Case I: $k=0$, Flat Universe

$$\begin{aligned}
 \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G \rho_0}{3} \left(\frac{a_0}{a}\right)^n \\
 \frac{1}{a} \frac{da}{dt} &\propto a^{-n/2} \\
 a &\propto t^{2/n}
 \end{aligned} \tag{1.12}$$

For a flat universe, it is quite easy to solve the equations. We know $n = 3$ for a matter-dominated universe and it is equal to 4 for a radiation-dominated universe. Hence, $a \propto t^{2/3}$ for a matter-dominated universe and $a \propto t^{1/2}$ for a radiation-dominated universe.

For a vacuum-dominated universe with Λ as the cosmological constant

$$\begin{aligned} \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3} \frac{\Lambda}{8\pi G} \\ a &= a_0 e^{\sqrt{\Lambda/3}t} \end{aligned} \quad (1.13)$$

Thus, the scale factor grows exponentially fast in a vacuum-dominated universe.

Case II: $k=+1$, Closed Universe

Let us first consider a matter dominated universe. Then, $\rho \propto a^{-3}$

$$\begin{aligned} \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G\rho_0}{3} \left(\frac{a_0}{a}\right)^3 - \frac{1}{a^2} \\ \frac{da}{dt} &= \sqrt{\frac{8\pi G\rho_0 a_0^3}{3a} - 1} \end{aligned}$$

We can define a quantity called as conformal time, η , where $d\eta = \frac{dt}{a}$

$$\begin{aligned} \frac{da}{d\eta} &= \sqrt{\frac{8\pi G\rho_0 a_0^3}{3} a - a^2} \\ a &= A(1 - \cos\eta) \end{aligned} \quad (1.14)$$

Now, $dt = ad\eta$. Therefore,

$$t = \int_0^\eta A(1 - \cos\eta')d\eta' = A(\eta - \sin\eta) \quad (1.15)$$

We can obtain similar equations for a radiation-dominated universe. These are, ($A_0 = 2A$)

$$a = \sqrt{A_0} \sin\eta \quad (1.16)$$

$$t = \sqrt{A_0} (1 - \cos\eta) \quad (1.17)$$

The parametric equations we obtained we obtained for a and t are equations for a cycloid for both the radiation and matter-dominated cases.

Case III: $k=-1$, Open Universe

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho_0}{3} \left(\frac{a_0}{a}\right)^3 + \frac{1}{a^2}$$

$$a = A(\cosh \eta - 1) \tag{1.18}$$

Now,

$$\int dt = \int a d\eta = t = A(\sinh \eta - \eta) \tag{1.19}$$

Similarly, for a radiation-dominated universe, we obtain the following equations for a and t

$$a = \sqrt{A_0} \sinh \eta \tag{1.20}$$

$$t = \sqrt{A_0} (\cosh \eta - 1) \tag{1.21}$$

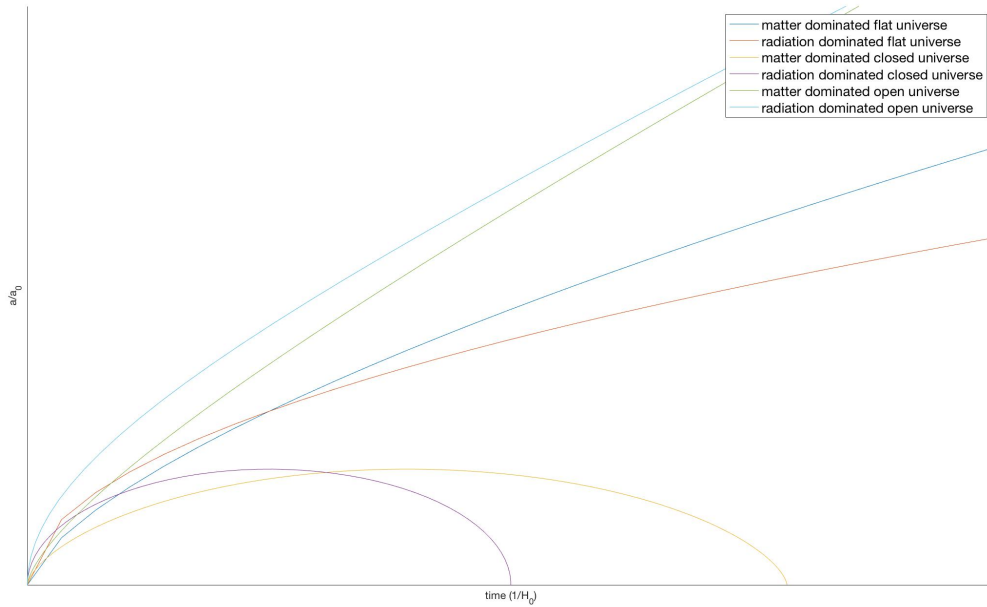


Figure 1.2: Schematic representation of the time dependence of the scale factor

From figure 1.2, we can see that the evolution of the scale factor depends on the curvature and densities of various constituents of the Universe. Interestingly, for a closed universe, the scale factor always turns around and goes back to zero. Thus, the Universe starts contracting

after some point in time and ends in a possible singularity like the Big Bang. This is known as the Big Crunch. For other cases, we can see that the scale factor continues to grow.

We derived the dependence of the scale factor on time, assuming contribution from only a single component, i.e., matter or radiation or vacuum. But in reality, the energy density has contributions from all components. Then the Friedmann equations can be solved numerically to obtain the actual expansion history of the Universe. Hence, measuring the values of the cosmological parameters is essential, in this case, to obtain the expansion history of the Universe.

1.4 Evolution of the Hubble Parameter

We have shown $\rho_m = \rho_{0m} \left(\frac{a_0}{a}\right)^3$ and ρ_Λ remains constant. Since, $\frac{a}{a_0} = \frac{1}{1+z}$, we have $\rho_m = \rho_0(1+z)^3$. From the first Friedmann equation (1.2) we obtain,

$$\begin{aligned} H^2 &= \frac{8\pi G}{3} [\rho_m + \rho_\Lambda] \\ \frac{H^2}{H_0^2} &= \frac{8\pi G}{3 H_0^2} [\rho_{0m}(1+z)^3 + \rho_\Lambda] \\ H^2 &= H_0^2 [\Omega_m(1+z)^3 + \Omega_\Lambda] \end{aligned} \tag{1.22}$$

where, H_0 is the hubble parameter at redshift $z = 0$. Sometimes, $[\Omega_m(1+z)^3 + \Omega_\Lambda]$ is called as the evolution factor, $E(z)$ and is used to quantify the evolution of the Hubble parameter with redshift [2].

Chapter 2

Structure Formation and the Cosmological Parameters

To understand how galaxies, galaxy clusters and everything else around us formed from tiny density fluctuations in the early Universe, we need to study the theory of structure formation. This is a dynamic and elegant theory where we have to take into account the competing effects of the expansion of the Universe and gravitational attraction. As can be seen from the Cosmic Microwave Background, the early Universe was highly homogeneous with small density fluctuations. The origin of these fluctuations is thought to be of quantum in nature and were supposed to have been amplified by inflation. These small density fluctuations started accumulating more matter and eventually grew to form the large-scale structures that we see today. Since the initial fluctuations are tiny, their evolution in time can be studied using the linear perturbation theory in a background expanding universe.

The study of these perturbations leads us to the linear power spectrum, the shape of which determines Ω_m as will be explained later. In this chapter, we first discuss the principle of Jeans instability in a homogeneous, static universe. Then we proceed to describe the evolution of perturbations in an expanding universe in section 2.2. We then introduce the power spectrum and its role in the measurement of cosmological parameters in section 2.3. The relation of the power spectrum to the mass function is discussed in section 2.4. The mass function by Press-Schechter [20] is outlined in section 2.5. In section 2.6, we discuss the luminosity function and its relation to the mass function.

2.1 Jeans Instability

At the beginning of the twentieth century, Sir James Jeans demonstrated how studying perturbations to a fluid in a gravitational potential could be used to study the formation of stars and planets. At that point in time, the fact that the Universe is expanding was unknown and he carried out his calculations in the context of a static background fluid [21]. He showed how fluctuations in density, $\delta\rho$ and velocity, v could evolve in time.

Let us consider a collisional fluid in a static background. Then the equations of motion are given by [21],

$$\frac{\partial}{\partial t}\rho + \nabla \cdot (\rho\mathbf{v}) = 0 \quad (\text{Continuity equation}) \quad (2.1)$$

$$\frac{\partial}{\partial t}\mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{v} + \frac{\nabla p}{\rho} + \nabla\phi = 0 \quad (\text{Euler equation}) \quad (2.2)$$

$$\nabla^2\phi = 4\pi G\rho \quad (\text{Poisson equation}) \quad (2.3)$$

Now, let us perturb the density, velocity, pressure and potential from their mean values ρ_0 , \mathbf{v}_0 , p_0 , ϕ_0 to ρ , \mathbf{v} , p , ϕ respectively, assuming the perturbations (ρ_1 , \mathbf{v}_1 , p_1 , ϕ_1) are small so that we can ignore terms of second order or higher in the equations later.

$$\begin{aligned} \rho &= \rho_0 + \rho_1 & \mathbf{v} &= \mathbf{v}_0 + \mathbf{v}_1 \\ p &= p_0 + p_1 & \phi &= \phi_0 + \phi_1 \end{aligned}$$

Then we have the following linearised equations,

$$\frac{\partial}{\partial t}\rho_1 + \rho_0\nabla\mathbf{v}_1 = 0 \quad (2.4)$$

$$\frac{\partial}{\partial t}\mathbf{v}_1 + \frac{\nabla p_1}{\rho_0} + \nabla\phi_1 = 0$$

$$\implies \frac{\partial}{\partial t}\mathbf{v}_1 + \frac{v_s^2\nabla\rho_1}{\rho_0} + \nabla\phi_1 = 0 \quad (\because \text{Sound speed, } v_s^2 = \frac{p}{\rho}) \quad (2.5)$$

$$\nabla^2\phi_1 = 4\pi G\rho \quad (2.6)$$

Differentiating (2.4) with respect to time and (2.5) with respect to space and then sub-

tracting the the two equations gives us $\frac{\partial^2 \rho_1}{\partial t^2} - v_s^2 \nabla^2 \rho_1 - 4\pi G \rho_0 \rho_1 = 0$. The Fourier transform of this equation gives us

$$\frac{\partial^2 \tilde{\rho}_1(k, t)}{\partial t^2} + (v_s^2 k^2 - 4\pi G \rho_0) \tilde{\rho}_1(k, t) = 0 \quad (2.7)$$

where, $\tilde{\rho}_1(k, t) = \int \rho_1(x, t) e^{-i\mathbf{k}\cdot\mathbf{x}} d^3k$. Here, we have used the fact that the Fourier transform of $\nabla^2 \rho_1$ is $-k^2 \tilde{\rho}_1$.

Let $\omega^2 = (v_s k)^2 - 4\pi G \rho_0$. We see that (2.7) admits oscillatory plane wave solutions if $\omega^2 < 0$ and exponentially decaying or growing solution if $\omega^2 > 0$. Thus, $\omega^2 = 0$ sets a characteristic scale, λ_J which is known as the **Jeans length** [21]. It is given by

$$\begin{aligned} k_J &= \sqrt{\frac{4\pi G \rho_0}{v_s}} \\ \lambda_J &= \sqrt{\frac{\pi}{G \rho_0}} v_s \end{aligned} \quad (2.8)$$

Thus, for $\lambda > \lambda_J$ ($k < k_J$) we have unstable exponentially decaying or growing solution and for $\lambda < \lambda_J$ ($k > k_J$) we have a stable oscillating solution.

2.2 Linear Perturbation theory

From the Cosmic Microwave Background, we see that the Universe had small density variations during the epoch of recombination. These variations later grew, due to gravitational instabilities, to form the structure that we see today, including the galaxy clusters. Since these inhomogeneities were small, we can use linear perturbation theory to study their evolution [3].

We consider a perturbation of $g_{\alpha\beta}(x)$ and the source $T_{\alpha\beta}$ in the form $(g_{\alpha\beta} + \delta g_{\alpha\beta})$ and $(T_{\alpha\beta} + \delta T_{\alpha\beta})$, where $\delta g_{\alpha\beta}$ and $\delta T_{\alpha\beta}$ denote the perturbation. Perturbations may grow at different rates when we change the system of coordinates. To solve this for length scales much smaller than the Hubble radius (d_H), we can use Newtonian gravity for which this is not a problem [3]. In the early Universe, the wavelength of the modes was larger than the

Hubble radius. But as the Universe expanded, the wavelength of the modes, λ became equal to d_H at one point of time, t_{enter} . As the Universe expanded further, the modes become much smaller than the Hubble radius, d_H [3].

Since Newtonian analysis can't be used for modes with wavelengths greater than d_H , i.e., for times $t < t_{enter}$, thus, the early evolution of any mode needs to be treated using general relativity by either sticking to one coordinate system or constructing quantities which are linear combinations of various perturbed physical variables [3].

2.2.1 Case I

Let us consider the case, $\lambda \gg d_H$ ($t < t_{enter}$). Here, we consider a spherical region of radius λ and $k = +1$ containing matter with mean density ρ_1 embedded in a $k = 0$ Friedmann Universe of density ρ_0 such that $\rho_1 = \rho_0 + \delta\rho$ [3]. Then,

$$H_1^2 + 1/a_1^2 = \frac{8\pi G}{3}\rho_1; \quad H_0^2 = \frac{8\pi G}{3}\rho_0 \quad (2.9)$$

where, $H_0 = \frac{\dot{a}_0}{a_0}$ and $H_1 = \frac{\dot{a}_1}{a_1}$, a_0 and a_1 being the corresponding scale factors. Let $H_0 = H_1$. We then get,

$$\begin{aligned} \frac{8\pi G}{3}(\rho_1 - \rho_0) &= 1/a^2 \\ \frac{(\rho_1 - \rho_0)}{\rho_0} &= \frac{\delta\rho}{\rho_0} = \frac{3}{8\pi G\rho_0 a_1^2} \end{aligned}$$

If $\delta\rho/\rho_0 \ll 1$ then $a_0 \approx a_1 \approx a$.

Since, $\rho_0 \propto a^{-4}$ for $t < t_{eq}$ and $\rho_0 \propto a^{-3}$ for $t > t_{eq}$, where t_{eq} corresponds to the time of matter-radiation equality, we get

$$\frac{\delta\rho}{\rho} = \begin{cases} a^2 & , t < t_{eq} \\ a & , t > t_{eq} \end{cases} \quad (2.10)$$

2.2.2 Case II

Let us now consider the case of $\lambda < d_H$. Now, gravitational collapse or amplitude growth is opposed by pressure support, specifically [3],

$$\begin{aligned}
 t_{\text{pressure re-adjustment}} &< t_{\text{gravitational collapse}} & (2.11) \\
 \frac{\lambda}{v} &< \frac{1}{\sqrt{G\rho}} \\
 \lambda &< \lambda_J
 \end{aligned}$$

where, $\frac{1}{\sqrt{G\rho}}$ is the free fall time, v is the velocity dispersion of the perturbed component, ρ is the density of the most gravitationally dominant component and $\lambda_J = \frac{\sqrt{\pi}v}{\sqrt{G\rho}}$ is the Jean's length. As we derived in the earlier section, modes with $\lambda < \lambda_J$ don't grow.

Another reason which prevents the growth of perturbations is that the Universe may be expanding faster than the rate of gravitational collapse which causes the dominant species to be smoothly distributed [3]. In this case,

$$\begin{aligned}
 t_{\text{expansion}} &< t_{\text{grav}} < t_{\text{pressure}} & (2.12) \\
 \implies \frac{1}{\sqrt{G\rho_R}} &< \frac{1}{\sqrt{G\rho_{DM}}} < \frac{\lambda}{v}
 \end{aligned}$$

where, ρ_R is the radiation density, and ρ_{DM} is the dark matter density. From (2.12), we can see that expansion of the Universe will prevent growth in all modes with $\lambda < d_H$ in the radiation dominated phase since $\rho_R > \rho_{DM}$ then. Thus, in the radiation dominated phase, only modes with $\lambda > d_H$ and they grow as a^2 (from 2.10). In the matter-dominated phase, since we can ignore pressure effects for all $\lambda < d_H$, the analysis leading to (2.10) is still valid. Thus, the amplitude grows as a in this case.

Thus, we have heuristically shown that in the radiation-dominated phase, only modes with $\lambda < d_H$ and modes with $\lambda > d_H$ don't grow. In the matter-dominated phase, all modes grow.

2.3 The Power Spectrum as a Cosmological Probe

Overdensities in the early universe refer to the change in the density at a particular point as compared to the average density of the Universe [22]. The overdensity field in the early Universe is given by,

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}_m}{\bar{\rho}_m} \quad (2.13)$$

where, $\bar{\rho}_m$ is the mean density of the Universe. It is assumed that $\delta(\mathbf{x})$ is a homogeneous and isotropic Gaussian random field [22]. We can take a Fourier transform of $\delta(\mathbf{x})$ which is $\delta(\mathbf{k})$ and the power spectrum is given by [22],

$$P(\mathbf{k}) \propto \langle |\delta(\mathbf{k})|^2 \rangle \quad (2.14)$$

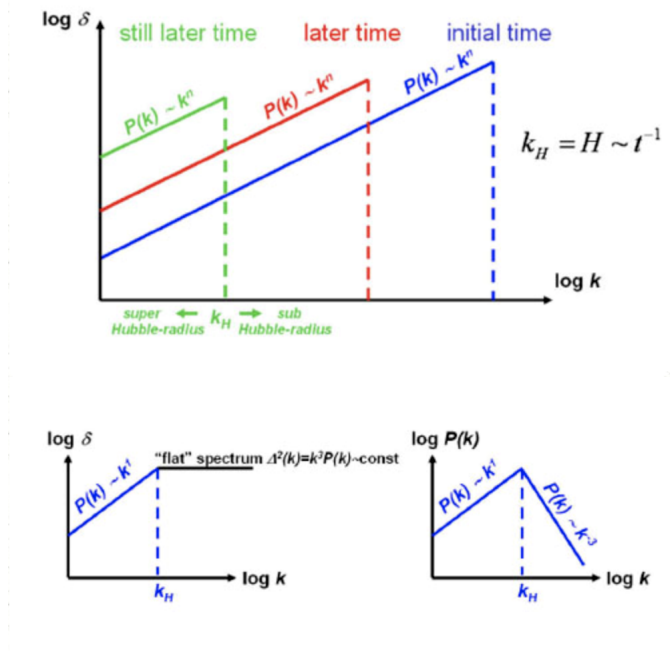
The two-point correlation function is defined to be $\xi = \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{R}) \rangle$. It tells us given a random galaxy what is the probability another galaxy will be found within \mathbf{R} . The Fourier transform of the correlation function also gives us the power spectrum [22],

$$P(\mathbf{k}) = \int \xi(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} d^3\mathbf{r} \quad (2.15)$$

From the Fourier transform of the Poisson's equation, $\nabla^2\delta\phi = 4\pi G \bar{\rho}\delta(\mathbf{x})$ [1], we have

$$\begin{aligned} k^2\delta\phi &\propto \delta(\mathbf{k}) \\ k^4P_\phi &\propto P(\mathbf{k}) \end{aligned} \quad (2.16)$$

where P_ϕ is the power spectrum of the potential. From (2.15), we can see that $P(\mathbf{k})$ has dimensions of volume and hence k^3P_ϕ is dimensionless and also turns out to be scale invariant [1]. Since, k^3P_ϕ is scale invariant, we conclude (from 2.16) $P(\mathbf{k})$ scales as k just after inflation for k beyond the horizon. The horizon at an epoch is defined to be the maximum distance any particle could have travelled from inflation to that epoch [1]. For k within the horizon, $P(\mathbf{k})$ scales as k^{-3} since $k^3P(\mathbf{k})$ becomes constant at this scale.

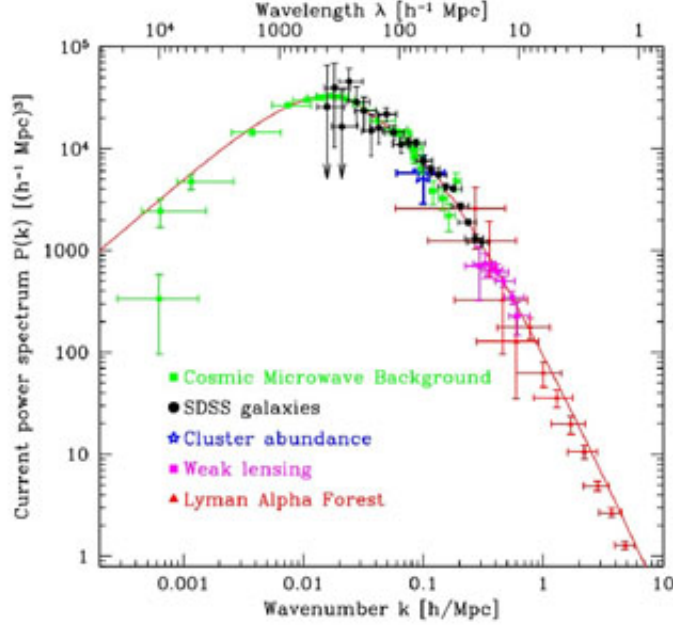


Source: NED (NASA/IPAC Extragalactic database)

Figure 2.1: Evolution of the matter power spectrum of density fluctuations during the radiation dominated era

In the radiation dominated phase, there were two kinds of fluctuations: ones which were within the horizon and ones which were outside the horizon. Fluctuations beyond the horizon grow and those within the horizon don't grow because the expansion timescale of the Universe is larger than the timescale of collapse (Section 2.2). Also, with time the horizon keeps growing due to the expansion of the Universe. On the other hand, in the matter-dominated phase, all scales grow equally. So, the shape of the power spectrum remains constant after the Universe becomes matter-dominated and as a result, the power spectrum forms a characteristic shape, as shown in fig. 2.2. This shape eventually smoothens due to different rates of growth of different components before and after matter-radiation equality [3]. The peak of the power spectrum characterizes the matter-radiation equality and thus, the shape of power spectrum is dependent on the fraction of matter and radiation in the Universe making it sensitive to the cosmological parameters, specifically Ω_m and σ_8 .

Observing the power spectrum directly is not trivial. The clustering of galaxies as a function of luminosity can be observed, which gives the shape of the power spectrum (which is sensitive to Ω_m), but not to the amplitude of density fluctuations (σ_8). To put constraints on this parameter, one can resort to abundances of galaxy clusters as a function of their luminosity, also known as the luminosity function.



Source: NED (NASA/IPAC Extragalactic database)

Figure 2.2: The Power Spectrum measured from different observations

2.4 The Mass function and its relation to the Power Spectrum

The luminosity function depends on the power spectrum through the **mass function**, which is the number density of galaxy clusters as a function of their mass. The mass function is then related to the variance of the density field, which relates to the power spectrum (as shown below) and thus acts as a probe for cosmological parameters.

The power spectrum can be directly related to the variance in the following way [1]:

$$\sigma^2(R) = \frac{1}{(2\pi)^3} \int P(k) |\tilde{W}(\mathbf{k}, R)|^2 d^3k \quad (2.17)$$

$$\sigma(R) = \int \delta(\mathbf{x} - \mathbf{r}) W(\mathbf{r}, R) d^3r \quad (2.18)$$

Here, $W(\mathbf{r}, R)$ is a window function defined as [1]

$$W(\mathbf{r}, R) = \begin{cases} \frac{3}{4\pi R^3}, & r < R \\ 0 & r > R \end{cases}$$

The variance is just the average of the overdensity field squared: lower the variance, higher the probability of forming massive clusters.

Let us choose a random point in our Universe. At high redshifts, mass and radius are interchangeable according to the relation $M = 4\pi/3 \bar{\rho} R^3$ [22]. As we increase R , M increases and $\sigma^2 \rightarrow 0$ because $\rho \rightarrow \bar{\rho}$ and as a result $\delta \rightarrow 0$. Conversely, for small R , the the variance can be quite high if we are centred around a density peak.

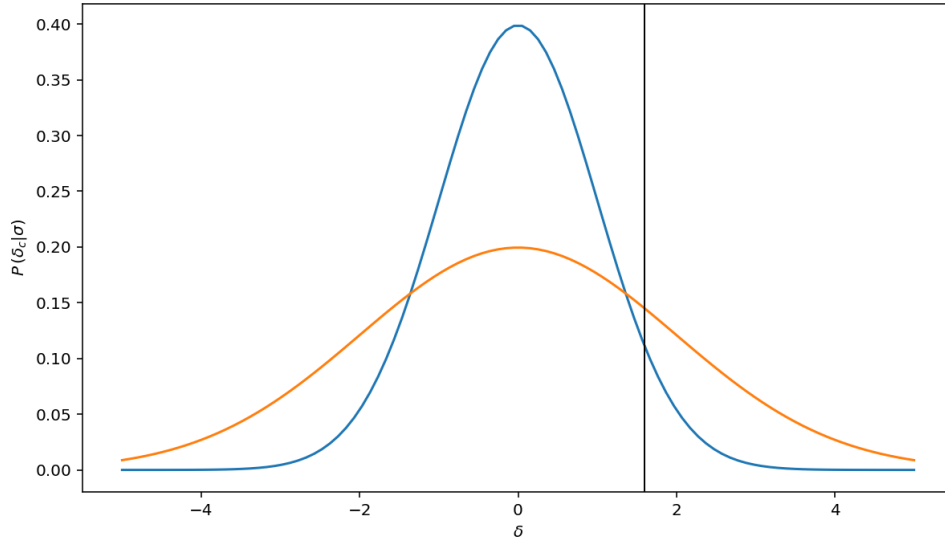


Figure 2.3: Distribution of primordial overdensities. The orange curve has a higher variance than the blue curve. The black vertical represents the critical overdensity δ_c

The densities at early times in the Universe can be approximated by a Gaussian distribution peaked at the mean density $\bar{\rho}_m$, where $\delta = 0$. For clusters to form, the overdensity at a point must be higher than a critical overdensity, δ_c [1]. Now, since clusters with low mass have higher variance than more massive clusters, the tails of their probability distribution beyond δ_c have a higher area under the curve compared to clusters with high mass, which have a lower area under the curve beyond δ_c , as can be seen in fig. 2.3. Hence we have a large number of low mass clusters and a small number of very massive clusters. This gives us the characteristic shape of the mass function, as can be seen in fig. 2.4.

2.5 Press-Schechter Mass Function

The first attempt to understand the abundance of galaxies was done by Press & Schechter (1974) [20]. It is one of the simplest mass functions to derive. If we assume a Gaussian

random field as our initial overdensity field and evolve it linearly from $\delta(0)$ to $\delta(t)$ then, the probability that the δ is greater than the critical density, δ_c required to form haloes is given by

$$\mathcal{P}(> \delta_c(t)) = \frac{1}{\sqrt{2\pi}\sigma(M)} \int_{\delta_c}^{\infty} \exp\left[-\frac{\delta^2}{2\sigma^2(M)}\right] d\delta = \frac{1}{2} \operatorname{erfc}\left[\frac{\delta_c(t)}{\sqrt{2}\sigma(M)}\right] \quad (2.19)$$

where, erfc is the complementary error function. As we know, overdensities above the critical density, $\delta_c(t)$ will collapse to form halos. Therefore, Press and Schechter argued that $\mathcal{P}(> \delta_c(t))$ is equal to the mass fraction of collapsed objects above a particular mass M , given by $F(> M)$. However, there is the following caveat to this argument. As $M \rightarrow 0$ or equivalently $R \rightarrow 0$ and hence $\sigma(R) \rightarrow \infty$, $\mathcal{P}(> \delta_c(t)) \rightarrow 1/2$. However, it is expected that $\mathcal{P}(> \delta_c(t)) \rightarrow 1$ because if the mean overdensity, $\sigma(R)$, is infinite then it is always greater than $\delta_c(t)$ and hence, clusters will always be formed. Thus, to account for this, Press and Schechter added an adhoc ‘fudge factor’ of 2, i.e., $F(> M) = 2 \mathcal{P}(> \delta_c(t))$.

Then the number density as a function of mass, $n(M)$, can be obtained using properties of the error function [20]

$$n(M) = \frac{\rho_m}{M} \frac{\partial F(> M)}{\partial M} = \sqrt{\frac{2}{\pi}} \frac{\rho_m \delta_c}{M \sigma} \exp\left(-\frac{\delta_c^2}{2\sigma^2}\right) \left| \frac{d \ln \sigma}{d \ln M} \right| \quad (2.20)$$

Press and Schechter proposed this mass function model in 1974. Now, the mass function is obtained numerically from N-body simulations and the Press-Schechter mass function qualitatively reproduces the mass function in the simulations, although it shows quite some deviations, as seen in fig. 2.4. Since 1974, several advancements have been made in our computational power and therefore, more refined mass functions have been proposed (Jenkins et al. (2001) [23], Tinker et al. (2008) [24]).

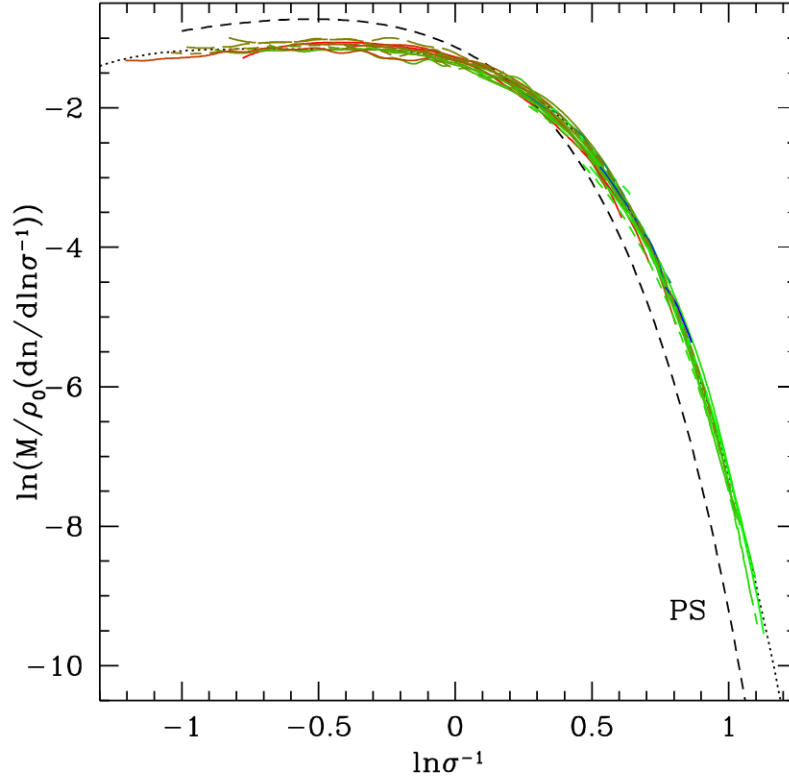


Figure 2.4: Mass function obtained from numerical simulations (green and red) [23]. The dotted line represents the mass function fit from Jenkins et al. (2001) and the dashed line represents the Press-Schechter fit to the mass function.

2.6 Luminosity function

The luminosity function is a very important quantity in the study of galaxies and galaxy clusters and is our primary observable. The luminosity function $\Phi(L)dL$ is the number density of galaxy clusters within the luminosity range $L \pm dL/2$ in a specific wavelength band [25],

$$\Phi(L) = \frac{1}{V_{\max}} \frac{dn}{dL} \simeq \frac{1}{V_{\max}} \frac{\Delta n}{\Delta L} \quad (2.21)$$

$$\Phi(\ln L) = \frac{1}{V_{\max}} \frac{dn}{d \ln L} \simeq \frac{1}{V_{\max}} \frac{\Delta n}{\Delta \ln L} \quad (2.22)$$

The luminosity function is calculated as a histogram with the weights as $1/V_{\max}$. V_{\max} is the co-moving volume corresponding to the maximum distance a cluster with a particular

luminosity can be detected by our telescopes. Here, L is the mean luminosity of each bin, Δn is the number counts in that bin, ΔL (or $\Delta \ln L$ in logspace) is the bin width. For more details on V_{\max} refer to section 3.2.2.

The luminosity is commonly fitted with a Schechter function of the form

$$\Phi(L) dL = \Phi^* \left(\frac{L}{L^*} \right)^\alpha \exp \left(-\frac{L}{L^*} \right) \frac{dL}{L^*} \quad (2.23)$$

This was proposed by Paul Schechter in 1976 [26]. Here, L^* , α and Φ^* are parameters of the model. As shown below (fig. 2.5), this function fits observed data of luminosities of the REFLEX clusters of galaxies quite well.

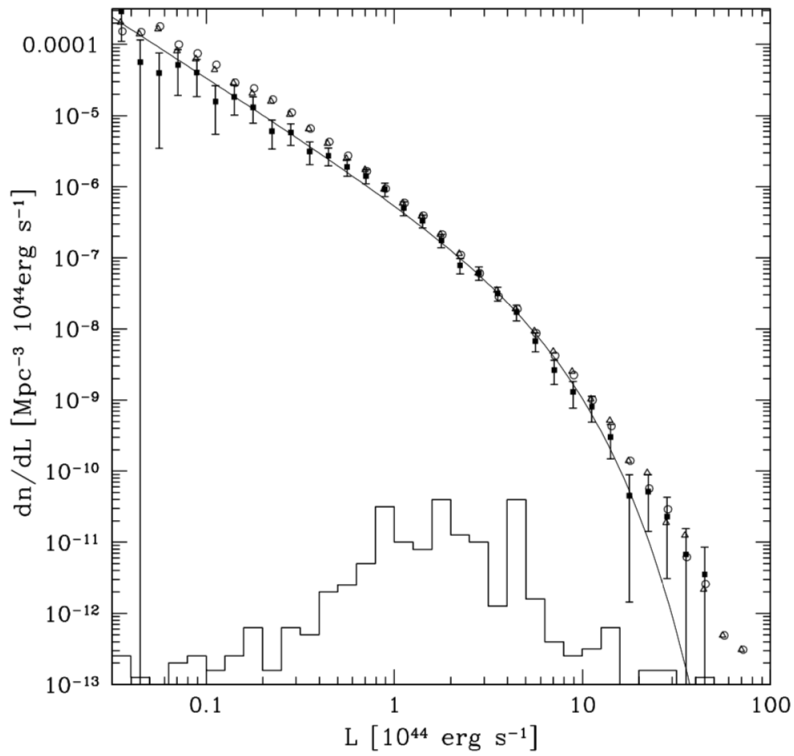


Figure 2.5: X-ray Luminosity function in Stanek et al. (2006) [7]. The filled squares represent data from REFLEX survey of clusters and the line is the Schechter fit to the luminosity function.

The luminosity function is related to the mass function through a mass-luminosity relationship and hence also related to the power spectrum. Thus, the luminosity function is a good measure of the cosmological parameters.

Chapter 3

Modelling the Luminosity function and Comparing it with Data

In the previous chapter, we have established how initial density perturbations grew to form galaxies and galaxy clusters we see today and how the power spectrum of density fluctuations is related to their abundances. The power spectrum is a powerful cosmological probe as its shape and amplitude characterizes our cosmological parameters of interest, Ω_m and σ_8 . But we do not directly measure the power spectrum. Our observable is the luminosity function, which can be related to the power spectrum via the mass function. The luminosity function is related to the mass function with the help of a mass-luminosity relation and the mass function is then related to the power spectrum via the variance in the initial density field.

In this chapter, we will describe how we can obtain the luminosity function from the mass function obtained from numerical N-body simulations while assuming a particular cosmology, i.e., a particular set of cosmological parameters (section 3.1). This acts as our theoretical model, to which we will try to fit the luminosity function obtained from surveys of galaxy clusters (REFLEX survey, in our case). In section 3.2, we describe the process of obtaining the luminosity function from luminosity data from surveys (we use and the caveats associated with it, like the flux-limit and V_{\max}). Finally, in section 3.3, we discuss the jackknife resampling technique we use to obtain the error bars in the luminosity function obtained from observations.

3.1 Theoretical Model of the Luminosity function

Our theoretical model consists of a mass function obtained from N-body simulations while assuming a particular cosmology. The mass function is then related to the luminosity function

via a mass-luminosity relation.

A relation between mass and luminosity is expected from our understanding of non-linear models of structure formation, particularly the self-similar model [22]. Thus, this relation is expected to hold across different clusters, assuming they evolved via the self-similar model. We use the following relation between the cluster mass, M and luminosity \bar{L} [7],

$$\bar{L} = L_{15} \left(\frac{M}{M_0} \right)^p \quad (3.1)$$

From the self-similar model, p is expected to be around 4/3 [7]. We use the logarithm of this relation, $\ln \bar{L} = \ln L_{15}(z) + p (\ln M - \ln M_0)$. Here, $\ln L_{15}$ is the intercept, $M_0 = 10^{15} h^{-1} M_\odot$ and p is the slope. There is a redshift dependence on the intercept,

$$\ln L_{15}(z) = \ln L_{15,0} + s \ln \left(\frac{\rho_c(z)}{\rho_c(0)} \right) \quad (3.2)$$

where, $\rho_c(z)$ is the critical density at redshift z and s is the self-similar evolution parameter, with $s = 7/6$ for self-similar evolution and $s = 0$ for no evolution [7].

But, the above mass-luminosity relation, (3.1) not an exact relation. There is a scatter around this mean relation which is assumed to be lognormal. For any given M , the scatter in L is

$$P(\ln L | \ln M, z) = \frac{1}{\sqrt{2\pi}\sigma_{\ln L}} \exp \left[-\frac{(\ln L - \ln \bar{L})^2}{2\sigma_{\ln L}^2} \right] \quad (3.3)$$

The standard deviation in mass at a fixed luminosity is $\sigma_{\ln M} = 0.37$ [7]. From this, we can obtain the standard deviation in luminosity at a fixed mass as $\sigma_{\ln L} = \sigma_{\ln M} p$.

Using the mass-luminosity relation, we can find the luminosity function, $\Phi(L)$. For each luminosity L , we can add all the clusters of different masses within a volume to get the number density of clusters for that luminosity. We can do this for a range of luminosities and take the continuous limit to change the summation to integration,

$$\Phi(L) = \int n(M, z) P(L|M, z) dM \quad (3.4)$$

In log scale we have, $\Phi(\ln L) = \int n(\ln M, z) P(\ln L | \ln M, z) d \ln M$ where, $n(M, z)$ and $n(\ln M, z)$ are related as follows,

$$\begin{aligned} n(M, z) dM &= n(\ln M, z) d \ln M \\ n(\ln M, z) &= n(M, z) M \ln 10 \end{aligned} \quad (3.5)$$

$\Phi(\ln L)$ is then converted to $\Phi(L)$ in the following way:

$$\begin{aligned}\Phi(L) dL &= \Phi(\ln L) d \ln L \\ \Phi(L) &= \frac{\Phi(\ln L)}{L}\end{aligned}$$

Thus, we finally have

$$\Phi(L) = \frac{1}{L} \int n(M, z) M P(\ln L | \ln M, z) d \ln M \quad (3.6)$$

3.1.1 Obtaining the model

The mass-luminosity relation has a strong dependence on the cosmological model used. In our case, we assume a flat Λ CDM cosmology with $\Omega_m = 0.3$, $\sigma_8 = 0.8$ and $H_0 = 70$ km/s/Mpc. We use the mass-luminosity relation parameters (slope, intercept and scatter), as mentioned in Stanek et al. (2006). They use the counts and scaling relations of low-redshift clusters in the HIFLUGCS survey (Reiprich & Böhringer, 2002) [27] to obtain the mass-luminosity relation. For mass measurements via the assumption of hydrostatic equilibrium (refer to Appendix A), measurements of the X-ray gas temperature and density profiles are used. They have also provided the values of the mass-luminosity relation parameters for three different values of Ω_m ($\Omega_m = 0.24, 0.30, 0.36$) between which we can linearly interpolate. Then we have the following values of the slope, intercept and scatter from Stanek et al. (2006) [7],

$$p = 1.59 \pm 0.05 \quad \ln L_{15,0} = 1.34 \pm 0.09 \quad \sigma_{\ln M} = 0.37 \pm 0.05$$

We later compare our model luminosity function to the the luminosity function obtained from data of REFLEX survey of galaxy clusters (Section 3.2), which are at a mean redshift of $z_{mean} = 0.085$. From (3.2), we can calculate $\ln L_{15}$. Since $\rho_c(z)/\rho_c(0) = E(z)^2$, we can use (1.22) to calculate the evolution factor, $E(z)$ and it turns out to be 1.08 for a mean redshift of 0.085. This gives $\ln \mathbf{L}_{15} = \mathbf{1.53}$ for the REFLEX clusters of galaxies.

Using a python class¹ written by Dr. Surhud More, which has a function that returns the mass function for a given mass and redshift and using the mass-luminosity parameters written above, we obtain the luminosity function model by numerically integrating (3.6) from 10^{11} to $10^{16} h^{-1} M_\odot$ in the log scale. The halo mass used here is M_{200} , which is the mass within a sphere whose mean interior density is $\Delta = 200 \rho_c(z)$. We use the mass function as

¹For details, refer to <http://surhudm.github.io/aum/code.html#module-cosmology>

given in Tinker et al. (2009) [24]

$$n(M) = \frac{dn}{dM} = f(\sigma) \frac{\bar{\rho}_m}{M^2} \frac{d \ln \sigma^{-1}}{d \ln M} \quad (3.7)$$

where, $f(\sigma)$ is the mass fraction and σ is the variance in the primordial density field as described in section 2.4. The mass fraction is universal to changes in redshift and cosmology. It is parametrized by

$$f(\sigma) = A \left[\left(\frac{\sigma}{b} \right)^{-a} + 1 \right] e^{-c/\sigma^2} \quad (3.8)$$

The parameters A , a , b , c are calibrated by simulations. These parameters are sensitive to the redshift and the overdensity Δ used to define the halos [24]

$$A(z) = A_0 (1+z)^{-0.14} \quad (3.9)$$

$$a(z) = a_0 (1+z)^{0.06} \quad (3.10)$$

$$b(z) = b_0 (1+z)^{-\alpha} \quad (3.11)$$

$$\log \alpha(\Delta) = - \left(\frac{0.75}{\log(\Delta/75)} \right)^{1.2} \quad (3.12)$$

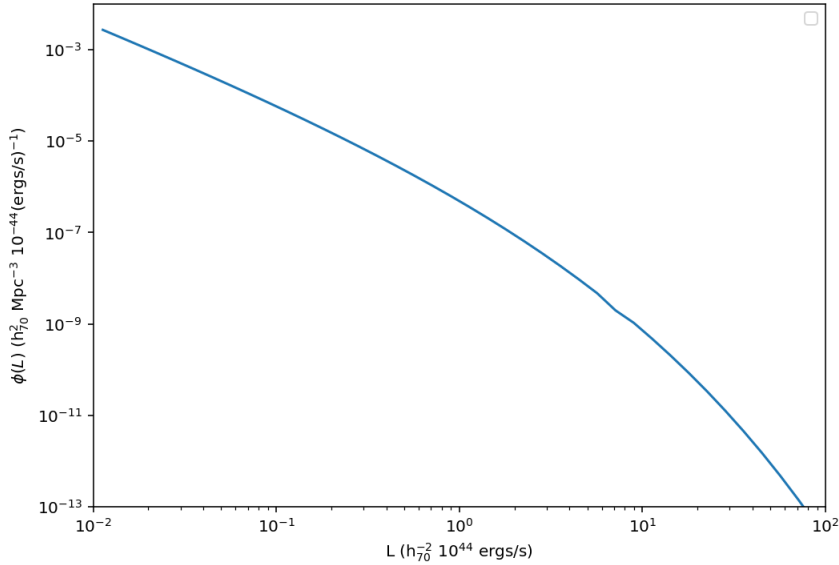


Figure 3.1: Luminosity function model

3.2 Luminosity function from data

We use the REFLEX survey of galaxy clusters as our dataset. It consists of 447 clusters and covers an area $\Omega = 4.24$ sr in the southern sky below a declination of 2.5° . Due to obstruction by the galactic plane, a region of $\pm 20^\circ$ around it is excluded. As mentioned before, the mean redshift of the clusters is 0.085 and the brightest clusters are towards a redshift of 0.35 [28].

3.2.1 Flux limit

Flux is the energy per unit area per unit time received by a detector. Luminosity is the total energy per unit time emitted by an object. If $d_L(z)$ is the luminosity distance, the flux is then related to luminosity as [4]

$$F = \frac{L}{4\pi d_L^2(z)} \quad (3.13)$$

Like most X-ray telescopes, the ROSAT telescope has a lower bound below which it can't measure the flux. The clusters were selected by their X-ray emissions in the ROSAT band (0.1-2.4 keV) down to a flux-limit of $F_{x \text{ lim}} = 3 \times 10^{-12}$ ergs s⁻¹ cm⁻² (3×10^{-15} W m⁻²) [28]. This means that any cluster whose flux is less than $F_{x \text{ lim}}$ won't be detected.

3.2.2 Luminosity limit and V_{max}

To obtain the luminosity function from data, we first created 40 logarithmically spaced bins from 10^{44} to 10^{46} h₇₀⁻² ergs/s as that is the range of luminosity function we are trying to reproduce.

From the flux limit defined earlier, we can find a luminosity limit, i.e., for each luminosity, there is a maximum distance $d_{L \text{ max}}$ and hence a maximum redshift up to which a cluster can be detected, as shown in fig. 3.2. Because of this, we can detect only very bright clusters at higher redshifts. The luminosity limit can be defined as

$$L_{x \text{ lim}} = F_{x \text{ lim}} 4\pi d_{L \text{ max}}^2(z) \quad (3.14)$$

To find the luminosity function, we can first create a histogram of the REFLEX luminosity data. But the luminosity function is not just the number of clusters in each bin. It is, in fact, the number of clusters per unit volume in each bin. This volume, V_{max} is defined as the maximum volume proportional to the cube of the maximum distance a cluster can be pushed to, from where its flux equals the flux limit $F_{x \text{ lim}}$. In other words, we can use $d_{L \text{ max}}$

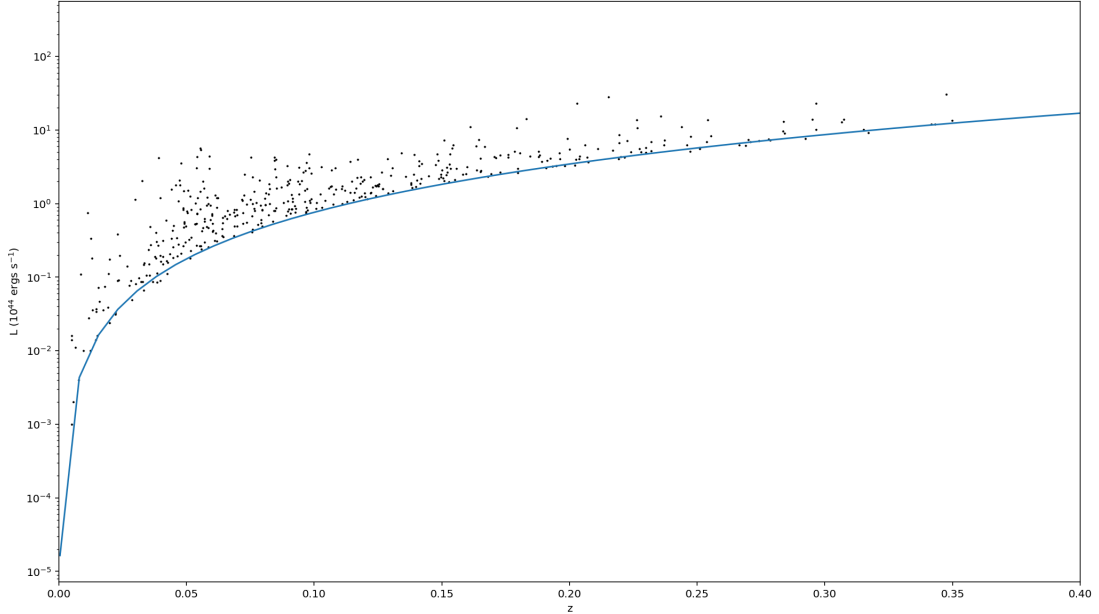


Figure 3.2: The REFLEX clusters (*filled circles*) plotted in luminosity-redshift space. The blue line represents the luminosity limit

to find the maximum co-moving distance, $d_{c \text{ max}} = \frac{d_L}{1+z}$ which defines a co-moving volume,

$$V_{\text{max}} = \frac{\Omega}{3} d_{c \text{ max}}(z_{\text{max}})^3 \quad (3.15)$$

Note that we use $\Omega = 4.24$ and not 4π to define the volume, since the REFLEX survey only includes an area of 4.24 sr and not the entire sky [28].

To find V_{max} , we fit a spline to the luminosity limit (3.14) so that we can invert it to find the z_{max} for each cluster, given the luminosity of the cluster. This z_{max} can be used to find a $d_{L \text{ max}}$, which in turn can give us V_{max} (from (3.15)).

3.2.3 Towards a luminosity function

After calculating V_{max} , a histogram is calculated with the the weights as $1/V_{\text{max}}$. This gives us $\frac{1}{V_{\text{max}}} \left(\frac{dn}{d \log L} \right) d \log L$. So, we divide by $d \log L$, which is the logarithmic bin width to obtain $\frac{1}{V_{\text{max}}} \left(\frac{dn}{d \log L} \right)$.

To go from $\frac{dn}{d \log L}$ to $\frac{dn}{dL}$, we use the chain rule,

$$\frac{dn}{dL} = \frac{dn}{d \log L} \frac{d \log L}{dL} = \frac{dn}{d \log L} \frac{1}{L \ln 10}$$

giving us the final luminosity function, $\Phi(L)$ as

$$\Phi(L) = \frac{1}{V_{\max}} \frac{dn}{d \log L L \ln 10} \quad (3.16)$$

Thus, we obtain the luminosity function from data, as shown in fig. 3.3. As we can see, the luminosity function we obtained (*Black*) closely matches the luminosity function obtained by Stanek et al. (2006) [7] and Böhringer et al. (2002) [25].

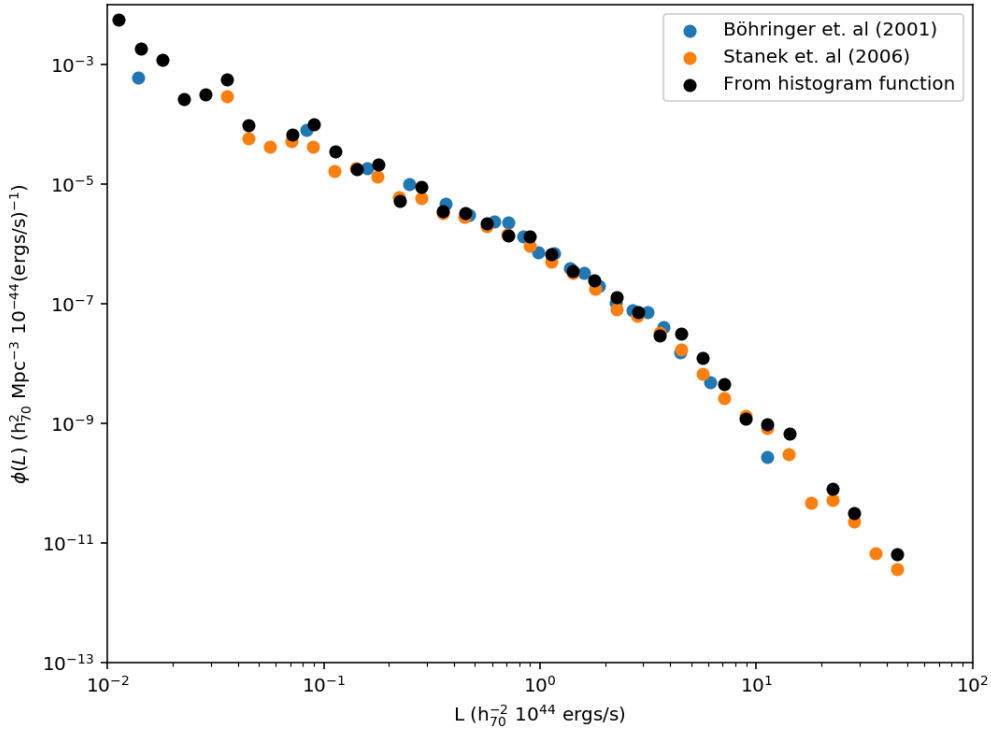


Figure 3.3: Observed Luminosity function from REFLEX data (*black*) compared with luminosity function from Böhringer et. al.[25] and Stanek et. al.[7]

3.3 Getting the error bars: Jackknife

Jackknife resampling is a statistical technique, which we use to determine error bars in our data. In general, the procedure involves removing one element from the sample set and constructing the observable from the remaining samples [29]. The scatter in the observable obtained after resampling multiple times gives the error bars.

In our case, we are constructing the luminosity function from the luminosity observations of REFLEX clusters in the southern sky. What we have to do is take N *equal* sized patches of the sky and remove one patch and construct the luminosity function from the remaining patches. This returns a luminosity function slightly different from the full luminosity function, which was obtained earlier without deleting any patch of the sky. After repeating this procedure N times, we get N luminosity functions, which slightly differ from each other. For each bin of the luminosity function, the standard deviation times \sqrt{N} of the N resampled data points for that bin gives us the error bar on that bin. In our case, we use $N = 16$.

3.3.1 RA, DEC

Right ascension (RA) and Declination (DEC) corresponds to the equivalent of longitude and latitude, respectively, in the celestial sphere. The zero of RA is chosen to be line perpendicular to the equatorial plane and passing through the point of intersection of the earth's equatorial plane and its orbital plane around the sun. RA is in units of hours:minutes:seconds, where 360° corresponds to 24 hours, hence $1 \text{ hour} = 15^\circ$. DEC is usually in units of deg:arcmin:arcsec.

$$1 \text{ deg} = 1^\circ \quad ; \quad 1 \text{ min} = \frac{1}{60}^\circ \quad ; \quad 1 \text{ sec} = \frac{1}{3600}^\circ$$

3.3.2 Creating equal-sized patches for Jackknife

It is challenging to create equal-sized patches in the sky because simply taking equally spaced bins of RA and DEC won't work since we have a 2D projection of a 3D spherical map. Also, obtrusion by the Milky Way and Magellanic clouds is another problem.

To achieve this, first, we have to have to change our coordinates from RA, DEC (in deg.) to ϕ (in rad.) and θ ,

$$\phi = \text{RA} \frac{\pi}{180} \quad ; \quad \theta = (90^\circ - \text{DEC}) \frac{\pi}{180}$$

The map, shown in fig. 3.4, shows all the points surveyed by the telescope. Note, these are not the clusters. An equal number of points correspond to equal areas. To make 16 patches:

- First we plot $\cos(\theta)$ vs ϕ as shown in fig. 3.4.
- We sort the data along the x-axis, i.e., ϕ .
- Then we divide the x-axis into four equal bins, with each bin containing the same number of points and call these as ϕ bins.
- We then make a 5×9 matrix to store the points such that the first column contains all the coordinates of the ϕ bins.
- We sort the data along y-axis, i.e., $\cos(\theta)$.
- Then we divide each ϕ bin into four patches with an equal number of points and call these as the $\cos(\theta)$ bins.
- To enter the $\cos(\theta)$ bins in the matrix, for each ϕ bin edge (each row of the matrix), we write down the coordinates of the $\cos(\theta)$ bins for that ϕ bin.

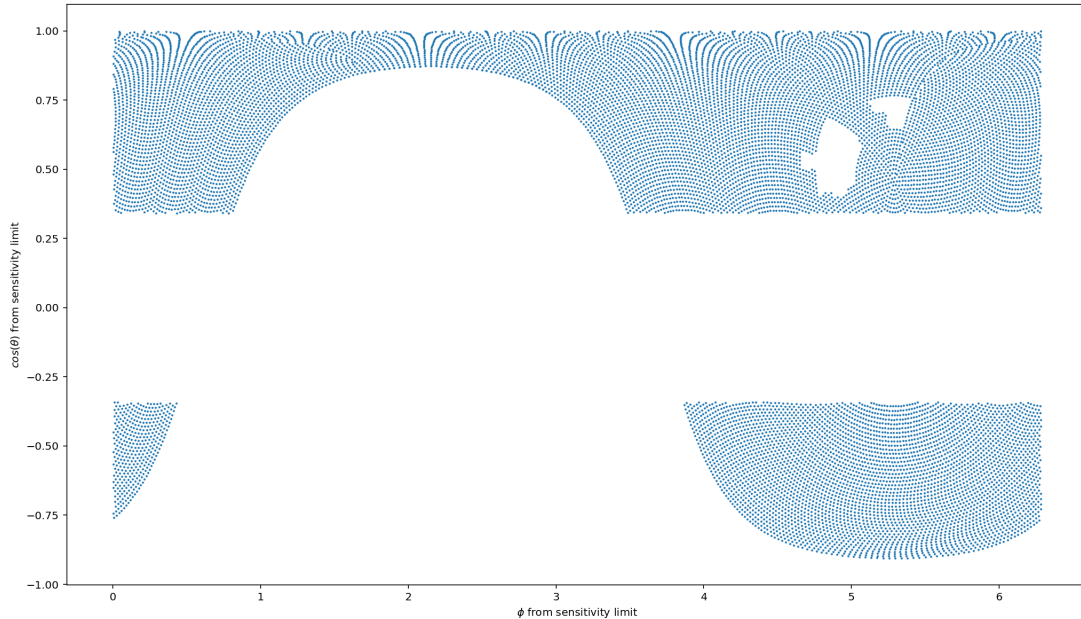


Figure 3.4: Area of the sky covered by the REFLEX survey. The points represent the areas surveyed by the ROSAT telescope.

This matrix contains the coordinates of all the patches. Now we can go over all the galaxy clusters and if check if it lies in a particular patch. If it does, we can delete the luminosity data of that cluster and thus implement jackknife resampling.

3.3.3 Implementing Jack Knife

To implement Jackknife, we select one patch and delete the data of all galaxy clusters that lie in it. Then from the remaining luminosity data, we construct a luminosity function. We then repeat this procedure for all patches such that we get multiple luminosity functions (N , if there are N patches) scattered around the original one. We calculate the standard deviation of all the new luminosity functions for each bin and multiply it by \sqrt{N} to get the error on each bin.

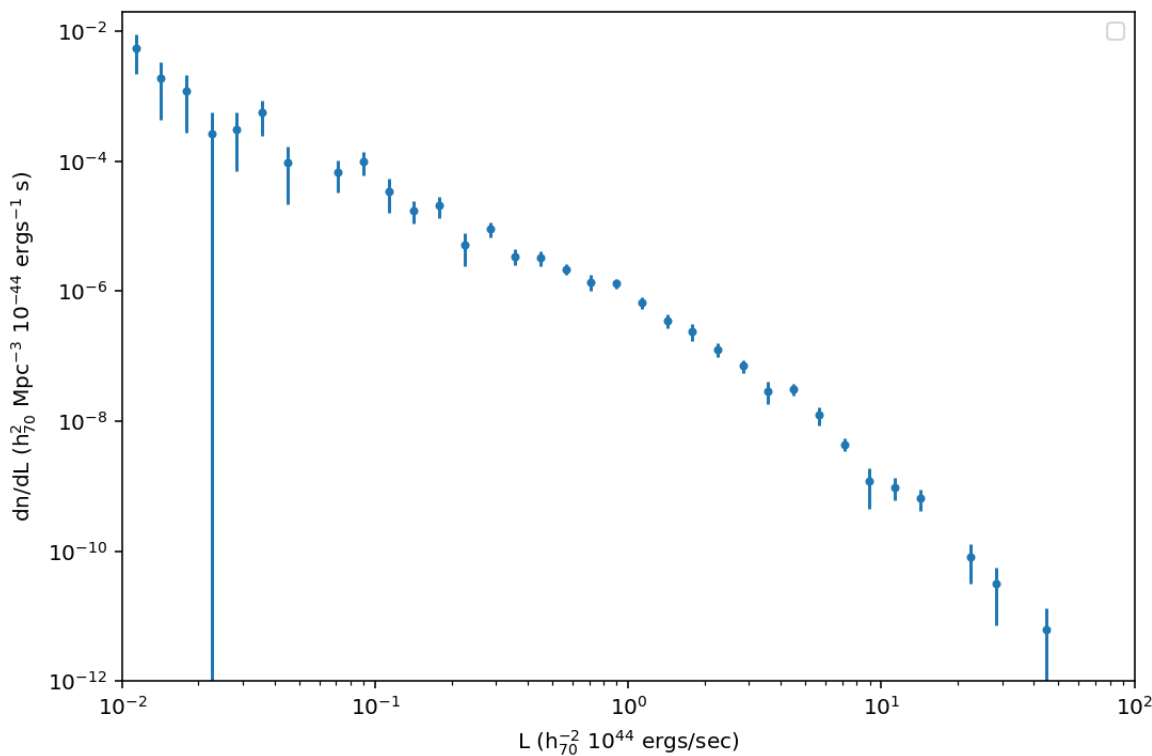


Figure 3.5: Luminosity function from REFLEX data plotted with errorbars calculated using jackknife resampling

Chapter 4

Monte Carlo Markov Chain

In the previous chapter, we have described how to get the theoretical model of the luminosity function from the mass function obtained from simulations and on obtaining the luminosity function from luminosity data. To see how well the model matches the data, refer to chapter 5. We have also described the jackknife resampling technique to obtain the error bars on the luminosity function we obtained from data.

Monte Carlo Markov Chain (MCMC) techniques, mostly based on the Metropolis algorithm, are widely used in the estimation of cosmological parameters. We employ Bayesian inference to obtain the posterior probability distribution of the cosmological parameters of interest (Ω_m and σ_8). Using the MCMC algorithm ‘emcee’, we have tried to reproduce the cosmological constraints obtained by Allen et al. (2003) [30]. emcee is a Python-based implementation of the Affine-Invariant MCMC Ensemble Sampler by Goodman & Weare (2010) [31].

In this chapter, we first describe random sampling from a probability distribution and the problems associated with it. Then, in section 4.2, we outline the principles behind bayesian likelihood analysis to obtain the posterior probability distribution. In section 4.3, we briefly summarise the mathematical principles behind emcee algorithm. In section 4.4, we describe in detail the implementation of emcee, including the details of the likelihood, priors, model and the observed luminosity function. These functions must be defined correctly before running the MCMC chains. Finally, we end this chapter with a brief discussion on the autocorrelation analysis of MCMC chains.

4.1 Random Sampling

In statistics, sampling refers to the selection of a subset of a population to infer properties about the entire population, since the entire population may be too large to practically study. For example, we may want to find out the average height of an Indian male. But it is not possible to measure the heights of all men of India. Therefore, we select a subset from the Indian male population and find their mean height, assuming the sample mean is equal to the mean of the entire population. To reduce bias, we may want to randomly select people from all states of India. This is known as random or uniform sampling. Random sampling occurs when each subset has an equal probability of being selected to minimize selection bias. But this has its own problems. We may end up selecting a subset that doesn't reflect the true distribution of the population as we may not have selected a subset that can have a significant contribution to the sample mean. Coming back to our example, we may not end up sampling a small population of extremely tall men from a particular region. In this case, our assumption of the sample mean being equal to the population mean is being violated.

Suppose we are given a probability distribution, $P(\mathbf{x})$, whose exact functional form is unknown or complicated. Then our problem is to generate samples from this distribution, $P(\mathbf{x})$, also referred to as the **target distribution**. Since we can't possibly sample from every \mathbf{x} , particularly if \mathbf{x} is a higher-dimensional vector, then it becomes imperative to sample from a smaller subset of the whole sample space. In this case, uniform sampling seems like the most straightforward solution. But uniform sampling can have the following problems [10]

1. We might not know the normalization, Z of $P(\mathbf{x})$, $Z = \int P^*(\mathbf{x})d^N x$, where $P^*(\mathbf{x})$ is the unnormalized distribution. If N is large, calculating Z becomes an immensely difficult task.
2. Even if we know Z drawing from $P(\mathbf{x})$ can be quite tough because the sampling space might be so huge (especially in higher dimensions like the Ising model with thousand spins which has a 1000 dimensions with each dimension taking values +1 or -1) that we won't be able to sample all possible states in the state space. In this case, we'll have to *uniformly* sample and for our samples to give accurate statistical results (like mean, median, etc.) close to the true statistics we'll have to ensure we're not missing those regions with high probability density. So, our number of samples should be high enough that we don't miss the typical set, i.e., the set where most of the high probability region is concentrated. How high should this be? For the Ising model with 1000 spins, it is about $2^{500} \approx 10^{150}$.

Thus, uniform or random sampling is not the best choice we have to sample from a probability distribution, especially in higher dimensions. We can use other techniques like importance sampling, rejection sampling, Metropolis-Hastings algorithm and Gibbs sampling, the details of which are briefly discussed in Appendix B. We use `emcee` for this purpose.

4.2 Bayesian Inference

A common problem in science is establishing how well a model function with given parameters ‘fits’ our observed data, or in other words, we want to determine the values of the model parameters for which the difference between the observed data and our model function is minimized. Thus, we would like to determine the probability distribution of the parameters given our data.

We use likelihood analysis based on the Bayes’ theorem for this purpose. Let θ represent the set of parameters and \mathbf{y} represents the measured outcome to which we have to fit the model. In our case, $\mathbf{y} = \Phi(L)$, the luminosity function and $\theta = \{\Omega_m, \sigma_8\}$. Then the posterior distribution of the parameters, $P(\theta|\mathbf{y})$ is given by,

$$P(\theta|\mathbf{y}) \propto \mathcal{L}(\mathbf{y}|\theta)p(\theta) \quad (4.1)$$

where \mathcal{L} is the likelihood function. It tells us that, given our model parameters, how well does the model match the data. $p(\theta)$ is the prior distribution. Just as it sounds, it enforces some range of values that the parameters are allowed to have, based on ‘prior knowledge’ of the parameters, say from previous experiments or theory.

MCMC methods are sampling methods that allow us to sample from the posterior distribution, $P(\theta|\mathbf{y})$, given the likelihood \mathcal{L} and priors $p(\theta)$.

4.3 Introduction to emcee

In the Metropolis-Hastings algorithm, we have one random walker sampling the entire space. Instead, we can have multiple random walkers so that the ensemble of walkers denoted by θ converges faster to the target distribution, $P(\theta|\mathbf{y})$. This is what is done in emcee.

In emcee, to update the position of a walker θ_j ($j \neq k$), the walker is selected randomly and a new position is proposed [11]:

$$\theta_k(t) \rightarrow \theta' = \theta_j + Z [\theta_k(t) - \theta_j]$$

where, Z is a random variable. Then, the proposal is accepted with probability

$$q = \min \left(1, Z^{-1} \frac{P(\theta')}{P(\theta_k(t))} \right) \quad (4.2)$$

where, N is the dimension of the parameter space. This process is repeated for each walker in the series, completing one iteration.

4.4 Implementing emcee

To implement emcee, we need to define our likelihood function, \mathcal{L} (Section 4.4.1) and the priors (Section 4.4.2). We use log likelihoods and log priors to get a posterior probability distribution, $P(\theta | \mathbf{y})$ up to a normalization.

$$\ln P(\theta | \mathbf{y}) \approx \ln \mathcal{L} + \ln p(\theta) \quad (4.3)$$

Then, we define functions which predict our model luminosity function (Section 4.4.3) and return the observed X-ray luminosity function (Section 4.4.4), both of which are sensitive to the cosmological parameters as described earlier. After that, we establish the initial positions of the n random walkers in the 2-dimensional parameter space, where the position of the i^{th} walker is given by θ

$$\theta = \begin{pmatrix} \Omega_{\text{m}}^{(i)} \\ \sigma_8^{(i)} \end{pmatrix} \quad i = 1, \dots, n \quad (4.4)$$

After establishing the initial positions of the random walkers and defining all the required functions, we can run emcee using the emcee Ensembler in Python¹ [11]. At each iteration, the random walkers sample the parameter space within the constraints set by the priors. From \mathcal{L} and the priors, the poster probability distribution, $P(\theta | \mathbf{y})$ is calculated. Based on the algorithm described above, each new step is accepted or rejected, based on just the previous step. Thus, the positions of the walkers form a Markov chain. These chains are the samples from the parameter space, which can be used to create a posterior probability distribution, which is shown in Chapter 5.

4.4.1 Likelihood

The log likelihood, $\ln \mathcal{L}$ is given by a χ^2 estimator (refer to Appendix C)

$$\ln \mathcal{L}(L_{\text{x}}, \Phi_{\text{data}}, \Phi_{\text{err}}) = -\frac{1}{2} \sum \left(\frac{\Phi_{\text{data}}(L_{\text{x}}) - \Phi_{\text{model}}(L_{\text{x}})}{\Phi_{\text{err}}} \right)^2 \quad (4.5)$$

where, $\Phi_{\text{data}}(L_{\text{x}})$ refers to the luminosity function from data and $\Phi_{\text{model}}(L_{\text{x}})$ refers to the model luminosity function. Φ_{err} is the measurement error on the luminosity function from data and L_{x} is the mean luminosity of each bin of $\Phi(L)$, both of which we obtain from Table 1 of Allen et al. (2003) [30].

¹refer to <https://emcee.readthedocs.io/en/stable/user/install/>

4.4.2 Prior distribution

We use flat priors on the cosmological parameters, as given in Allen et al. (2003)

$$p(\Omega_m) = \begin{cases} 1, & 0.05 \leq \Omega_m < 1 \\ 0, & \text{otherwise} \end{cases} \quad \ln p(\Omega_m) = \begin{cases} 0, & 0.05 \leq \Omega_m < 1 \\ -\infty, & \text{otherwise} \end{cases}$$

$$p(\sigma_8) = \begin{cases} 1, & 0.2 < \sigma_8 < 1.5 \\ 0, & \text{otherwise} \end{cases} \quad \ln p(\sigma_8) = \begin{cases} 0, & 0.2 < \sigma_8 < 1.5 \\ -\infty, & \text{otherwise} \end{cases}$$

4.4.3 Model

In this section, we describe the model for the luminosity function we use while trying to reproduce the MCMC results from Allen et al. (2003) [30]. We use the same mass function as used by Allen et al. (2003), i.e., the mass function from Evrard et al. (2002) [32] to marginalize over the mass-luminosity scaling relation and obtain the luminosity function. The mass fraction in Evrard et al. (2002) is given by [32]

$$f(\sigma^{-1}) = A \exp[-|\ln \sigma^{-1} + B|^\epsilon] \quad (4.6)$$

where, for $\Omega_m = 0.3$ and $z = 0$, $A = 0.22$, $B = 0.73$, $\epsilon = 3.86$. To obtain A , B , ϵ for other values of Ω_m and z , we use simple interpolation provided by Evrard et al. (2002). As defined earlier, the mass function is then given by $\frac{dn}{dM} = f(\sigma) \frac{\bar{\rho}_m}{M^2} \frac{d \ln \sigma^{-1}}{d \ln M}$.

The mass-luminosity relation as given by Allen et al. (2003) is $E(z)M = M_0 \left[\frac{L}{E(z)} \right]^p$ or

$$\log(E(z) M) = p \log \left[\frac{L}{E(z)} \right] - 2p \log(7/5) + \log M_0 \quad (4.7)$$

where, $E(z)$ is the evolution factor as described in section 1.4. We have included the factor of $2p \log(7/5)$ to account for the fact that we use different units of luminosity from Allen et al. (2003), who use units of $h_{50}^{-2} 10^{44}$ erg/s. The values of the slope, p and intercept $\log M_0$ as obtained by Allen et al. by are

$$\log \left[\frac{M_0}{h^{-1} M_\odot} \right] = 13.99^{+0.20}_{-0.23} \quad p = 0.76^{+0.16}_{-0.13} \quad (4.8)$$

Without accounting for the scatter in the mass-luminosity relation, we can directly obtain a simplistic model of the luminosity function (in units of $h_{70}^5 \text{ Mpc}^{-3} (10^{44} \text{ erg/s})^{-1}$) from the

mass function, as shown below

$$\begin{aligned}\Phi(L) &= \frac{dn}{dL} = \frac{dn}{dM} \frac{dM}{dL} \\ \Phi(L) &= \frac{dn}{dM} \frac{p M}{L}\end{aligned}\tag{4.9}$$

Accounting for the scatter gives us a more refined model for luminosity function which we have already shown in section 3.1, leading to (3.6).

4.4.4 Luminosity function from data

Allen et al. (2003) focus on low-redshift ($z < 0.3$) and highly luminous ($L_x > 5 \times 10^{44} \text{ h}_{70}^{-2} \text{ erg/s}$) clusters from the REFLEX (ROSAT ESO Flux-limited Survey) and the extended BCS (Brightest Cluster Sample) studies. The restriction to high luminosities is done to reduce systematic uncertainties and also because that is the region where the mass-luminosity relation is well-calibrated [30].

The BCS is the first truly X-ray selected cluster sample to emerge from the ROSAT All-Sky Survey (RASS). The original sample has 203 clusters (Ebeling et al. 1998 [33]) in the northern sky, covering an area of 4.14 sr. The extension (Ebeling et al. 2000 [9]) consists of 107 additional clusters within an area inside one-sixth of the original BCS area. The original BCS clusters have fluxes $> 4.4 \times 10^{-15} \text{ W m}^{-2}$ and the newer extended sample consists of fluxes between $2.8 \times 10^{-15} \text{ W m}^{-2}$ and $4.4 \times 10^{-15} \text{ W m}^{-2}$. Their distribution in the luminosity-redshift space is shown in fig. 4.1. We use the combined sample for our analysis. The mean redshift of the combined sample is 0.21.

After obtaining the cluster counts for REFLEX and the combined BCS samples (results shown in Chapter 5) while restricting to the redshift limit of 0.3, we proceed to obtain the X-ray luminosity function from data. Since, REFLEX is a flux-limited survey with a flux-limit of $3 \times 10^{-15} \text{ W m}^{-2}$, we obtain V_{max} and then $\Phi(L)$ for REFLEX as shown earlier in section 3.2.

The luminosities of the extended BCS clusters are reported in an Einstein de-Sitter cosmology ($\Omega_m = 1$ and $H_0 = 50 \text{ km/s Mpc}^{-1}$). So, we have to transform the luminosities back to our flat Λ CDM cosmology with $\Omega_m = 0.3$ and $H_0 = 70 \text{ km/s Mpc}^{-1}$ to have consistent results. This dependence on cosmologies occur since the measurement of the luminosity distance, $d_L(z)$ depends on Ω_m and $L \propto d_L^2(z)$ [2]. The luminosities in our cosmology $L_{\Lambda\text{CDM}}$ is then given by

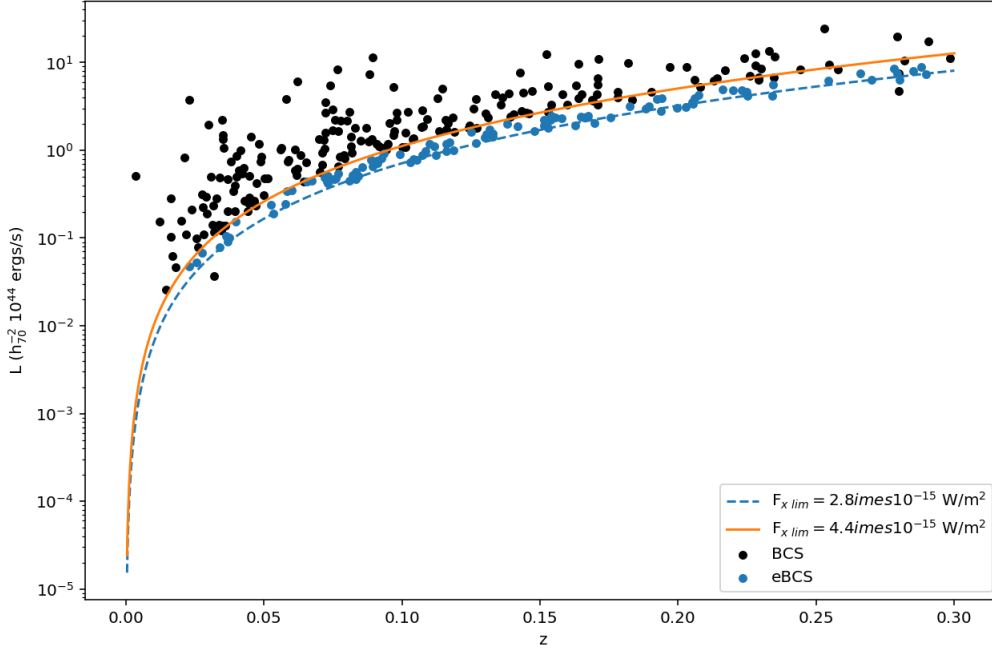


Figure 4.1: The original BCS (black) and the extended BCS (blue) clusters shown in luminosity-redshift space

$$L_{\Lambda\text{CDM}} = L_{\text{Einstein}} \left[\frac{d_L^{\Lambda\text{CDM}}(z)}{d_L^{\text{Einstein}}(z)} \right]^2 \quad (4.10)$$

where, $d_L^{\Lambda\text{CDM}}(z)$ and $d_L^{\text{Einstein}}(z)$ are the distances measured in the ΛCDM and Einstein de-Sitter cosmologies respectively.

After obtaining the transformed luminosities while restricting ourselves to the aforementioned redshift cut $z_{\text{max}} = 0.3$, we obtain the cluster counts which are shown in Chapter 5. Since, Ebeling et al. (2000) place a redshift cut, we use that cut to define a $\tilde{V} = \Omega/3 d_{\text{c max}}^3(z_{\text{max}})$, with $\Omega = 4.14 \text{ sr}$ [9]. If our calculated V_{max} from the flux-limit of the extended BCS clusters ($F_{\text{x lim}} = 2.8 \times 10^{-15} \text{ W m}^{-2}$) is greater than \tilde{V} , we use \tilde{V} as our V_{max} . Else, we use the V_{max} calculated from $F_{\text{x lim}}$.

We use the REFLEX and BCS bins given in Allen et al. (2003) to obtain the weighted histogram, as shown in section 3.2. Finally, we obtain the X-ray luminosity functions for REFLEX and BCS datasets from the counts using (3.16).

4.5 Autocorrelation Analysis

One important question here is for how long should we run these chains to be confident that the walkers have converged to the required posterior probability distribution? For this, we need to find the autocorrelation time, which essentially tells us the number of iterations required by series to “forget” where it started. But first, we must define the autocorrelation function $\hat{C}(\tau)$ [34],

$$\hat{C}(\tau) = \frac{1}{\sigma_{\theta_k}^2 (N - \tau)} \sum_{n=1}^{N-\tau} [\theta_k(n) - \mu_{\theta_k}] [\theta_k(n + \tau) - \mu_{\theta_k}] \quad (4.11)$$

where, $\mu_{\theta_k} = \frac{1}{N} \sum \theta_k(n)$ is the sample mean of the chains and N is the total number of iterations. Here, the ‘chains’ $\theta_k(n)$ refer to the time series of the positions of the walkers in the parameter space of each parameter indexed by k . The autocorrelation function tells us about the correlation of the emcee chains at a current time with the itself at some previous time, τ . Here, ‘chains’ refers to the positions of the walkers in the parameter space at any iteration. We expect $\hat{C}(\tau)$ to fall off with τ , as chains more separated in time are expected to have lesser correlation. Heuristically, the autocorrelation time is the time at which the $\hat{C}(\tau)$ falls to $1/e$ of its initial value, which is 1 because the correlation function defined above is normalized.

The integrated autocorrelation time, τ_{θ_k} is then obtained by summing over the autocorrelation function over all times [34].

$$\tau_{\theta_k} = \sum_{-\infty}^{\infty} \hat{C}(\tau) \approx 1 + 2 \sum_{\tau=1}^N \hat{C}(\tau) \quad (4.12)$$

But this is not a very good idea as the summing all the way to N will give us a very noisy estimate of τ_{θ_k} . Instead, we can sum upto some $M \ll N$. This results in a decreased variance of the estimator with the cost of some added bias [34]. Thus, it is recommended to run the chains for longer than $50 \tau_{\theta_k}$ for emcee to remove the bias².

In practice, we run the emcee chains for a small number of iterations to estimate the autocorrelation time. This is known as the ‘burn-in’ phase. After using the built-function of emcee to estimate the autocorrelation time, τ_{θ_k} for each θ_k , we run the final emcee Ensemble for more than $50 \tau_{\theta_k}$ for the largest θ_k . This ensures that our walkers have converged reasonably well enough.

²refer to <https://emcee.readthedocs.io/en/stable/tutorials/autocorr/>

Chapter 5

Results and Discussion

We have shown in previous chapters the methods for obtaining the luminosity function from data as well as the theoretical model of the luminosity function from the mass function. In this chapter, we present our results and outline the prospects for future work.

In Chapter 3, we have shown how to obtain the X-ray luminosity function from REFLEX data in the luminosity range of $10^{42} - 10^{46} h_{70}^{-2}$ erg/s. We obtained the luminosity function for 40 logarithmically spaced bins in the given luminosity range with the V_{max} obtained from the flux-limit of $3 \times 10^{-15} \text{ W m}^{-2}$. We also obtain the theoretical model from the Tinker et al. (2009) [24] mass function by integrating over the scatter $\sigma_{\ln L}$ in the mass luminosity relation. Both the model and the observed luminosity function are constructed assuming a flat Λ CDM cosmology with $\Omega_m = 0.3$, $\sigma_8 = 0.8$ and $H_0 = 70 \text{ km/s Mpc}^{-3}$. The model of the X-ray luminosity function and the one obtained from data is compared to that obtained by Böhringer et al. (2001) [28] and Stanek et al. (2006) [7]. The errorbars on the observed luminosity function are obtained by jackknife resampling as described earlier in section 3.3.

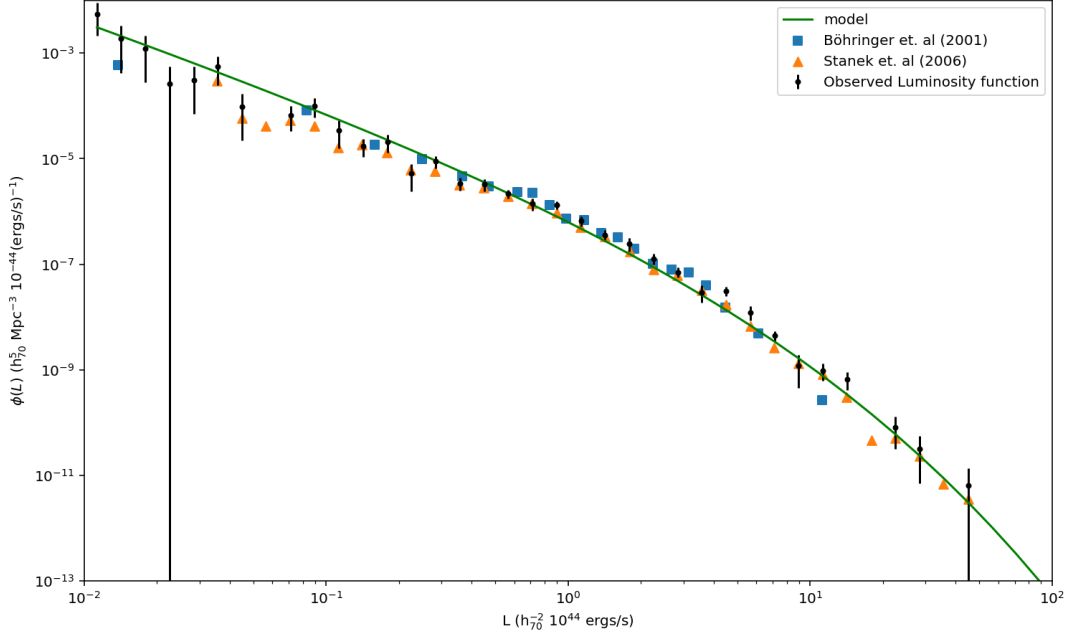


Figure 5.1: The theoretical model and the observed X-ray luminosity function obtained from REFLEX data in the luminosity range $10^{42} - 10^{46} h_{70}^{-2} \text{ erg/s}$ compared to the luminosity functions obtained by Böhringer et al. (2001) [28] and Stanek et al. (2006) [7]

As we can from fig. 5.1 the model fits the data well when a flat Λ CDM cosmology with $\Omega_m = 0.3$, $\sigma_8 = 0.8$ is used.

To sample from the posterior probability distribution of the cosmological parameters, we use the MCMC method called emcee [11]. Here, we try to reproduce the constraints obtained by Allen et al. (2003) [30], who use data from both the REFLEX (Böhringer et al. (2002) [25]) and the combined BCS (Ebeling et al. (1998) [33] and Ebeling et al. (2000) [9]) studies.

We use the bins as given in Table 1 of Allen et al. (2003). Limiting our cluster samples to $z < 0.3$ and $L_x > 5.1 \times 10^{44} h_{70}^{-2} \text{ erg/s}$, we obtain the number of galaxy clusters in each bin, shown in table 5.1. We observe that the counts in each bin closely match the counts given in Allen et al. except for the last REFLEX bin. This can be due to the fact Allen et al. use an older catalogue of the REFLEX clusters of galaxies (Böhringer et al. (2002) [25]), whereas the data currently available is an updated catalogue (Böhringer et al. (2004) [8]). In the updated catalogue 7 clusters have been removed [8], which is why our counts are possibly lower than expected in the last REFLEX bin.

We also observe that the mean luminosities of the clusters of each bin are close to the

mean luminosities obtain by Allen et al. According to Allen et al., the mean redshift of the BCS clusters is $z = 0.21$, which we also confirm.

Table 5.1: The cluster counts of REFLEX and combined BCS clusters ($L_x > 5 \times 10^{44} h_{70}^{-2}$ erg/s). Column 2 shows the mean X-ray luminosity of each bin in units of $h_{50}^{-2} 10^{44}$ erg/s and the error bars give the boundaries of bins. Column 3 shows the number of clusters in each bin. Column 4 shows the mean luminosities each bin as obtained by Allen et al. (2003), with the error bars giving the bin boundaries. Column 5 gives the number of clusters in each bin as given in Allen et al. (2003)

| | L_x | n | L_x Allen | n_{Allen} |
|--------|-------------------------|----|-------------------------|--------------------|
| BCS | $11.77^{+1.62}_{-1.73}$ | 17 | $11.73^{+1.62}_{-1.73}$ | 17 |
| | $15.66^{+2.20}_{-2.30}$ | 18 | $15.65^{+2.20}_{-2.30}$ | 17 |
| | $25.19^{+28.8}_{-6.06}$ | 16 | $23.91^{+28.8}_{-6.06}$ | 17 |
| REFLEX | $11.41^{+2.19}_{-1.19}$ | 21 | $11.25^{+2.19}_{-1.19}$ | 20 |
| | $15.58^{+4.61}_{-2.83}$ | 22 | $16.27^{+4.61}_{-2.83}$ | 20 |
| | $31.18^{+76.2}_{-9.07}$ | 12 | $29.95^{+76.2}_{-9.07}$ | 20 |

After reproducing the counts of Allen et al. we obtain the model and observed luminosity function as explained in section 4.4.4. Note, that the V_{max} is calculated differently for the combined BCS and the REFLEX samples. To obtain the model, we directly use the mass-luminosity relation to obtain $\Phi(L)$ directly from dn/dM , as shown in (4.9). This is a simplistic model without accounting for the scatter in the mass-luminosity relation. It is used to obtain preliminary constraints on the cosmological parameters. The observed luminosity function is obtained as described in section 4.4.4. The model without the scatter and the observed luminosity function are then plotted, along with the luminosity function given by Allen et al., as shown in fig. 5.2. A flat Λ CDM cosmology with $\Omega_m = 0.3$, $\sigma_8 = 0.7$ and $H_0 = 70$ km/s/Mpc is assumed. As we can see the both our model without the scatter and the observed luminosity function are within the error bars of the luminosity function given by Allen et al.

This sets a clear starting point to run the MCMC chains. After defining functions for the model without scatter, observable, likelihood and priors, we can run emcee in our two-dimensional parameter space of Ω_m and σ_8 using the likelihood and priors described in section 4.4.1 and 4.4.2 respectively. This returns the samples from the posterior probability distributions of Ω_m and σ_8 , as shown in fig. 5.3.

We obtain autocorrelation times, τ of 34.19 and 33.88 for Ω_m and σ_8 respectively, as a result of which we run the emcee chains for 50000 iterations which is much greater than

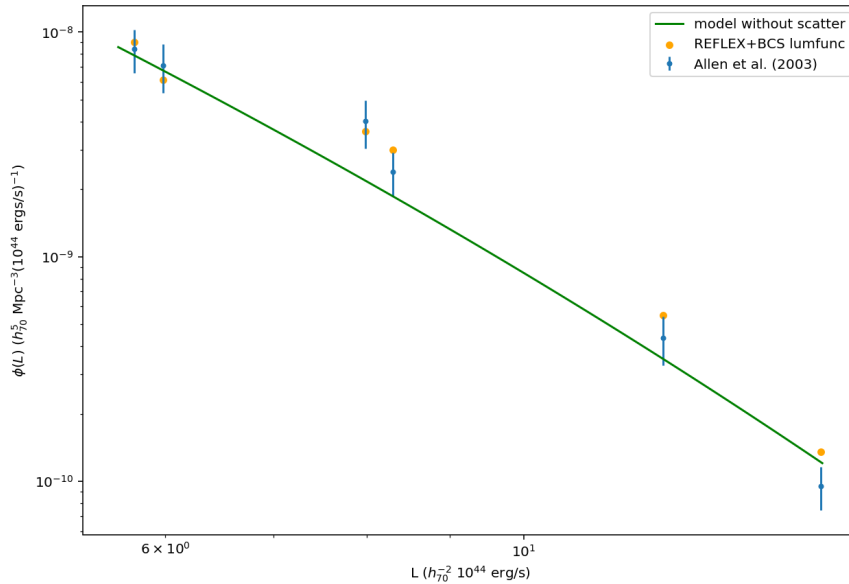


Figure 5.2: The X-ray luminosity function. A simple model is used, obtained without integrating over the scatter. The blue points with error bars represent the luminosity function from Allen et al. (2003). Yellow points represent the luminosity function obtained from REFLEX+BCS data

the required time of 50τ (Section 4.5). As we can see from fig. 5.4, the emcee chains have reasonably converged.

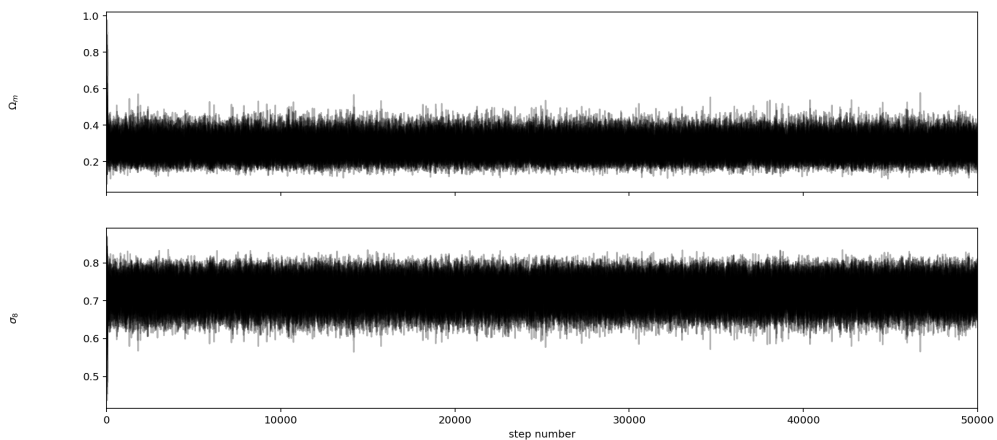


Figure 5.4: MCMC chains for the parameters Ω_m and σ_8 for a run of 50000 iterations

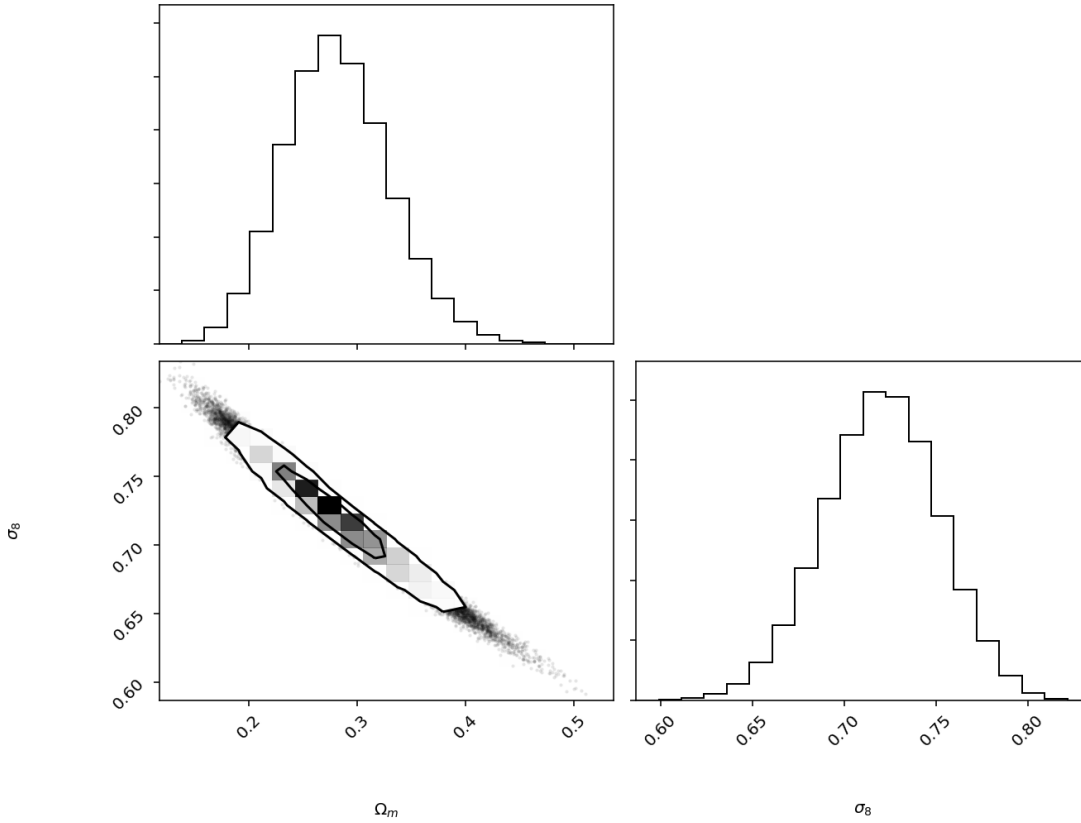


Figure 5.3: The 68 and 95 percent confidence contours in the posterior distribution of Ω_m and σ_8 obtained from the combined BCS and REFLEX luminosity function (Section 4.4.4) fitted with a model without taking the scatter in the mass-luminosity relation into account (Section 4.4.3). A flat Λ CDM cosmology is assumed.

In the MCMC analysis, we have used a simpler model without accounting for the scatter in the mass-luminosity relation to fit the luminosity function obtained from REFLEX and combined BCS data. We use an intercept $\log\left(\frac{M_0}{h^{-1} M_\odot}\right) = 13.99$ and slope $p = 0.76$ in our model, as given by Allen et al. (2003). When neglecting the effects of scatter in the mass-luminosity relation, a MCMC run in the parameter space of Ω_m and σ_8 returns their mean values as $\Omega_m = 0.280^{+0.052}_{-0.046}$ and $\sigma_8 = 0.721^{+0.030}_{-0.033}$. For a model without the scatter and while ignoring the uncertainties in the slope and intercept of the mass-luminosity relation, Allen et al. report $\Omega_m = 0.23$ and $\sigma_8 = 0.74$ [30]. These values are within the error bars we have obtained for the measurements of Ω_m and σ_8 . The posterior probability contours of Ω_m and σ_8 exhibit the well-known degeneracy between these parameters (fig. 5.3).

We ran our MCMC chains with our parameters only at Ω_m and σ_8 . In principle, we

should run the MCMC chains for 5 parameters, Ω_m , σ_8 , the slope, intercept and scatter in the mass-luminosity relation. In that case, we should not use the simple model without the scatter (4.9) but then integrate over the scatter using (3.6)

$$\Phi(L) = \frac{1}{L} \int n(M, z) M P(\ln L | \ln M, z) d \ln M$$

The attempts at using the model with scatter have failed as this model doesn't match the observed luminosity function and the luminosity function from Allen et al., as shown in fig. 5.5. We believe the main reason for this is that the mass-observable relation quoted in Allen et al. is the mean of the distribution $P(\ln M | \ln L)$ instead of $P(\ln L | \ln M)$ that we need in our framework. In the absence of scatter, the two relations are straightforward inverses of each other. But they differ in the case of a non-trivial scatter and are related by the Bayes theorem. In the near future, we plan to modify our framework to use the mass-observable relation as quantified by the mean of the $P(\ln M | \ln L)$ as an additional constraint and constrain cosmological parameters.

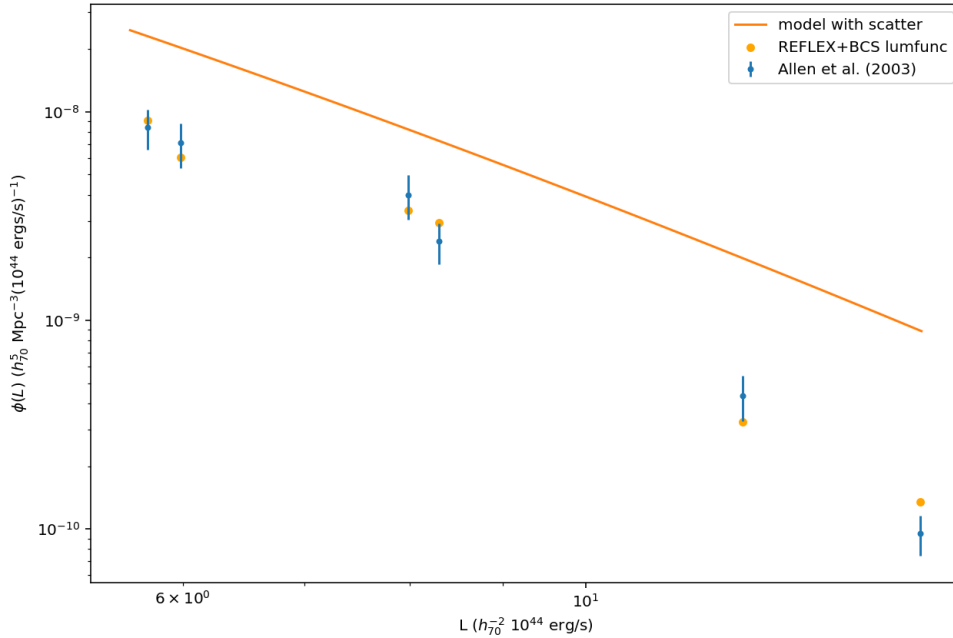


Figure 5.5: The X-ray luminosity function. The model is obtained by intergrating over the scatter. The blue points represent the luminosity function from Allen et al. (2003). Yellow points represent the luminosity function obtained from REFLEX+BCS data

Chapter 6

Conclusion

In this thesis, we have shown how the matter power spectrum of density fluctuations is sensitive to the cosmological parameters. The power spectrum, $P(k)$ can be obtained from the variance of the primordial density field, which relates it to the halo mass function (Section 2.3). The halo mass function is directly related to our observable, the X-ray luminosity function, via a mass-luminosity scaling relation.

We have also used a theoretical model to predict the X-ray luminosity function of galaxy clusters (Section 3.1). We have also shown the procedure for obtaining the luminosity function from the observations of luminosities and redshifts of galaxy clusters (Section 3.2). In a flat Λ CDM cosmology, the model matches the observed luminosity function obtained from REFLEX data and also the luminosity function by Böhringer et al. (2002) and Stanek et al. (2006) in the luminosity range of $10^{44} - 10^{46} h_{70}^{-2}$ erg/s. This preliminary analysis increases our confidence in the model.

Using a simple model which does not marginalize over the scatter in the mass-luminosity relation, we obtain $\Omega_m = 0.280^{+0.052}_{-0.046}$ and $\sigma_8 = 0.721^{+0.030}_{-0.033}$, which lies within the errorbars of similar results obtained by Allen et al. (2003) [30]. As seen in Fig. 5.3, the 68 and 95 per cent confidence contours of Ω_m and σ_8 show the familiar degeneracy between these parameters.

We must note that the mean values of the parameters and confidence contours obtained with MCMC are obtained for a simple model, without accounting for the scatter in the mass-luminosity relation. For a more refined analysis, one must integrate over the scatter to obtain the luminosity function from the mass function. We also ran our MCMC chains for only two parameters, assuming the values of the slope and intercept given by Allen et al. (2003). In principle, one should run MCMC for a 5-parameter model, Ω_m , σ_8 , slope, intercept and scatter in the mass-luminosity relation.

In future, one can use weak lensing data by experiments like the Hyper-Suprime Cam (HSC)¹ of the Subaru telescope, to calibrate the mass-luminosity relationship for wider halo mass range. Future surveys by telescopes like the Large Synoptic Survey Telescope (LSST)² and eROSITA³, will play a crucial role in estimations of cosmological parameters. eROSITA has a sensitivity which is about 30 times higher than ROSAT [35]. It can also survey up to higher redshifts (> 1) and an energy band of 0.2-10 keV, covering a much broader energy range than ROSAT [35]. Thus, it hopes to create the largest catalog of galaxy clusters with its deep sky X-ray survey.

The LSST aims to conduct the deepest and widest survey of the southern sky ever done. In addition to carrying out extensive surveys in multiple wavelengths for cluster catalogs, LSST, with its weak lensing and photometric-redshift capabilities, promises to play a crucial role in the estimation of cosmological parameters, particularly dark energy and its equation of state [36].

¹For more details, refer to <https://www.subarutelescope.org/Observing/Instruments/HSC/index.html>

²For more details, refer to <https://www.lsst.org/>

³For more details, refer to <https://www.mpe.mpg.de/eROSITA>

Appendix A

Hydrostatic Equilibrium

The assumption of hydrostatic equilibrium helps in the measurement of masses of galaxies and clusters, which can be used to characterize mass-luminosity scaling relations. Hydrostatic equilibrium states that the potential gradient $\nabla\phi$ is balanced by the pressure gradient, $\nabla p(\mathbf{x})$ [22] assuming spherical symmetry and the equation of state of an ideal gas, $p = \rho kT/\mu m_p$

$$\begin{aligned} \frac{1}{\rho} \frac{dp}{dr} &= -\frac{GM}{r^2} \\ \frac{k_b}{\rho \mu m_p} \left[\frac{d\rho}{dr} T + \frac{dT}{dr} \rho \right] &= -\frac{GM}{r^2} \\ \frac{-r^2 k_b T}{G \mu m_p} \left[\frac{d\rho}{dr} \frac{1}{\rho} + \frac{dT}{dr} \frac{1}{T} \right] &= M \\ M(< r) &= -\frac{r k_b T}{G \mu m_p} \left[\frac{d \ln \rho}{d \ln r} T + \frac{d \ln T}{d \ln r} \rho \right] \end{aligned} \tag{A.1}$$

Thus, M tells us about the cluster mass within a radius r . Note that, Eq. (A.1) is valid only for collisional baryonic matter. Also, a cluster may not reach complete equilibrium due to occurrence of mergers and continuous accretion [22].

Appendix B

More Monte Carlo Methods

B.1 Importance sampling

Suppose we have random variable \mathbf{x} with a probability distribution $P(\mathbf{x})$ from which it is too difficult to sample from, due to higher dimensions or $P(\mathbf{x})$ being too complicated a function. To find the expectation value of a function X , we would use $\langle \mathbf{x} \rangle = \sum \mathbf{x} P(\mathbf{x})$. But since the summation might contain a large number of terms due to large number of terms in the state space of $P(\mathbf{x})$, evaluating this might be extremely difficult. Instead we can randomly sample from $P(\mathbf{x})$ and use that to approximate a mean

$$\langle \mathbf{x} \rangle \approx \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i^{(P)}$$

where, the superscript indicates we are drawing the samples from $P(\mathbf{x})$. As we saw earlier, this might be very difficult too since $P(\mathbf{x})$ might be too complicated to sample from directly (we might not hit the typical set). So, instead of sampling from $P(\mathbf{x})$ we can sample from a simpler distribution $Q(\mathbf{x})$ which must be a close approximation of $P(\mathbf{x})$. Now, we can reduce N , the number of times which we sample. But how do get $\langle \mathbf{x} \rangle$ from $Q(\mathbf{x})$?

$$\begin{aligned} \langle \mathbf{x} \rangle &= \sum \mathbf{x} P(\mathbf{x}) \\ &= \sum \mathbf{x} \frac{P(\mathbf{x})}{Q(\mathbf{x})} Q(\mathbf{x}) \end{aligned}$$

We can see this as calculating the expectation value of $\mathbf{x} \frac{P(\mathbf{x})}{Q(\mathbf{x})}$ with respect to $Q(\mathbf{x})$ which can be approximated as [10]

$$\langle \mathbf{x} \rangle \approx \frac{1}{N} \sum_{i=1}^N x_i^{(Q)} \frac{P(x_i)}{Q(x_i)}$$

where $x_i^{(Q)}$ is being sampled from $Q(\mathbf{x})$ now. This considerably simplifies our problem as our sampling distribution has now changed from the complicated $P(\mathbf{x})$ to the much simpler $Q(\mathbf{x})$.

In higher dimensions, obtaining samples that lie in the typical set of P may take a long time unless Q is a good approximation of P . This is a major problem associated with importance sampling.

B.2 Rejection sampling

Rejection sampling works very well if Q is a good approximation of P . It is assumed we know a constant c such that $cQ(x) > P(x), \forall x$. If this is the case, then we can proceed as follows:

1. Generate a sample x_0 from $Q(x)$
2. Generate another random number u such that $u \in [0, cQ(x_0)]$
3. If $u > P(x_0)$ then x_0 is rejected, else it is accepted and added to our set of samples $\{x\}$

Since cQ completely encompasses P , all accepted points lie inside P as can be seen from the above procedure.

In higher dimensions, rejection sampling doesn't work very well. We can see this if we consider a pair of normalized Gaussians with the same mean as P and Q with $\sigma_P < \sigma_Q$ in N dimensions.

$$c = \frac{P(0)}{Q(0)} = \left(\frac{2\pi\sigma_Q^2}{2\pi\sigma_P^2} \right)^{N/2} = \left(\frac{\sigma_Q}{\sigma_P} \right)^N$$

The acceptance rate is given by volume of curve under P by the volume under cQ . Since, P and Q are both normalized, it is just $1/c$, i.e., acceptance rate = $\left(\frac{\sigma_Q}{\sigma_P} \right)^{-N}$, which can be quite small for larger N .

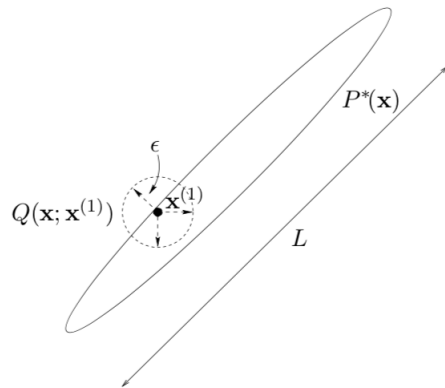
B.3 Metropolis-Hastings algorithm

In the Metropolis-Hastings algorithm, Q depends on the current state \mathbf{x}_t . This works better now because $Q(\mathbf{x})$ doesn't have to be similar to $P(\mathbf{x})$, unlike importance sampling or rejection sampling. $Q(\mathbf{x}; \mathbf{x}_t)$ can be a simple function from which we can draw samples, where the notation means that Q is a function of \mathbf{x} centered on the current state \mathbf{x}_t .

This algorithm works in the following way. A new state \mathbf{x}' is generated from Q . Then [10]

$$a = \frac{P(\mathbf{x}') Q(\mathbf{x}_t; \mathbf{x}')}{P(\mathbf{x}_t) Q(\mathbf{x}'; \mathbf{x}_t)} \quad (\text{B.1})$$

is evaluated. In practice, the proposal distribution is often symmetric, i.e., $Q(\mathbf{x}_t; \mathbf{x}') = Q(\mathbf{x}'; \mathbf{x}_t)$. Then, $a = \frac{P(\mathbf{x}')}{P(\mathbf{x}_t)}$. If $a \geq 1$ then new state is accepted, else it is accepted with probability a [10]. This means that a random number, r , should be sampled from a uniform distribution such that $0 < r < 1$. If $r < a$ then accept the new state, else reject it. If accepted $\mathbf{x}_{t+1} = \mathbf{x}'$, else $\mathbf{x}_{t+1} = \mathbf{x}_t$. This forms a Markov chain.



Source:[10]

Figure B.1: Metropolis-Hastings algorithm in two dimensions, showing a proposal density, $Q(x_t; x')$ with a length scale ϵ and a target distribution unnormalized $P^*(\mathbf{x})$ with a length scale L

There are problems with the Metropolis-Hastings algorithm too. Consider the example shown in fig.B.1. The random walker is at $\mathbf{x}^{(t)}$ trying to sample from the 2D probability distribution $P^*(x)$. Due to the elongated unidirectional shape of $P^*(x)$, if the step size is too large, most of the states will be rejected. If the step size is too small, it will take a long time to sample from the entire distribution.

If the no. of steps is T with step size ϵ , it can be shown that random walks follow a

Gaussian statistical distribution in the large T limit. Its mean displacement, L is then

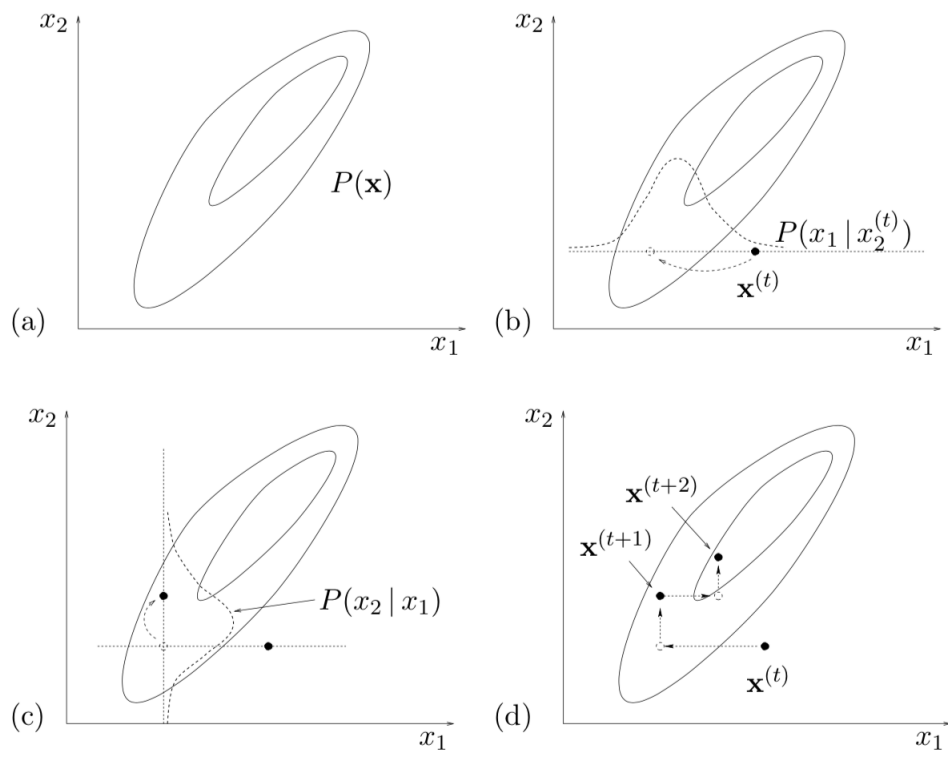
$$\begin{aligned} L &\approx \sqrt{T}\epsilon \\ T &\approx \left(\frac{L}{\epsilon}\right)^2 \end{aligned} \tag{B.2}$$

This applies to higher N -dimensional state spaces too. Let us consider the target distribution as an N -dimensional Gaussian with σ_n as the standard deviation in the n^{th} direction and the proposal distribution as a spherical Gaussian with standard deviation ϵ . If we assume ϵ is small enough such that each new state is almost always accepted, then the random walker can take step sizes of ϵ . Then, $T \approx \left(\frac{\sigma_{max}}{\epsilon}\right)^2$, since σ_{max} is the largest length scale in the problem. Now, ϵ can be increased to σ_{min} to obtain an optimal no. of iterations as $T \approx \left(\frac{\sigma_{max}}{\sigma_{min}}\right)^2$. ϵ can't be larger than σ_{min} because the acceptance rate will fall sharply then. At least this doesn't have any catastrophic dependence on the dimension N , although it can still be quite large due to a quadratic dependence on length scale ratios.

B.4 Gibbs sampling

In Gibbs sampling, we need at least a two dimensional distribution to work with. It is useful if drawing from the conditional distribution $P(x_i|\{x_j\}_{j \neq i})$ is easier than drawing from the actual distribution $P(\mathbf{x})$. We start with a point $\mathbf{x}^{(t)}$ in the parameter space. Using $P(x_1|x_2^{(t)})$ we sample x_1 along the line shown in fig.B.2 and then sample along x_2 from $P(x_2|x_1)$ which gives us $\mathbf{x}^{(t+1)}$. This completes one iteration and continuing this till $t \rightarrow \infty$, will exactly reproduce $P(x)$ from the distribution of the samples.

There are problems with this method, which are similar to the problems faced in the Metropolis-Hastings algorithm. If we have two strongly correlated variables x_1 and x_2 , with the length scales of the target and conditional distributions as L and ϵ respectively, then it takes $(L/\epsilon)^2$ iterations to sample reliably from the target distribution.



Source:[10]

Figure B.2: Gibbs sampling

Appendix C

χ^2 Distribution

The χ^2 distribution is obtained from the gamma distribution. It is given as [37],

$$f(x) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{p/2-1} e^{-x/2} \quad , \quad 0 < x < \infty \quad (\text{C.1})$$

The χ^2 distribution has an interesting property. Let X_1, X_2, \dots, X_n be samples from a normal distribution, $n(\mu, \sigma^2)$. Then [37],

$$S^2 = \sum_i \frac{(X_i - \bar{X})^2}{\sigma^2}$$

has a χ^2 distribution with $n - 1$ degrees of freedom. A χ^2 test is used to test how well a given model fits our data. This is a method of hypothesis testing with the null hypothesis that the model is true. Then, given the expected observations (from the model) and the actual observations, we can calculate a χ^2 value and compare it with a critical χ^2 . If the calculated χ^2 is less than the critical one, then our hypothesis is true, otherwise its false. The χ^2 value for a random variable, X is given by

$$\chi^2 = \sum_i \left(\frac{X_i^{\text{observed}} - X_i^{\text{model}}}{\Delta X_i^{\text{observed}}} \right)^2 \quad (\text{C.2})$$

The critical chi-squared value, χ_c^2 is for a certain significance level, α and a certain number of degrees of freedom is given by

$$\alpha = \int_{\chi_c^2}^{\infty} f(x) dx \quad (\text{C.3})$$

χ_c^2 is usually obtained from commonly available χ^2 tables.

Bibliography

- [1] Houjun Mo, Frank C. van den Bosch, and Simon White. *Galaxy Formation and Evolution*. 2010.
- [2] Sean M. Carroll. *Spacetime and Geometry: An Introduction to General Relativity*. Cambridge University Press, 2019.
- [3] T. Padmanabhan. *Structure Formation in the Universe*. Cambridge University Press, 1993.
- [4] S. Weinberg. *Cosmology*. Cosmology. OUP Oxford, 2008.
- [5] Planck Collaboration et al. Planck 2018 results. vi. cosmological parameters. *ArXiv e-prints*, page 2, 2018.
- [6] Steven W. Allen, August E. Evrard, and Adam B. Mantz. Cosmological parameters from observations of galaxy clusters. *Annual Review of Astronomy and Astrophysics*, 49(1):409–470, Sep 2011.
- [7] Rebecca Stanek, A. E. Evrard, H. B Bohringer, P. Schuecker, and B. Nord. The x-ray luminosity-mass relation for local clusters of galaxies. *Astrophys. J.*, 648:956–968, 2006.
- [8] H. Böhringer, P. Schuecker, L. Guzzo, C. A. Collins, W. Voges, R. G. Cruddace, A. Ortiz-Gil, G. Chincarini, S. De Grandi, A. C. Edge, H. T. MacGillivray, D. M. Neumann, S. Schindler, and P. Shaver. The ROSAT-ESO Flux Limited X-ray (REFLEX) Galaxy cluster survey. V. The cluster catalogue. *A&A*, 425:367–383, October 2004.
- [9] H. Ebeling, A. C. Edge, S. W. Allen, C. S. Crawford, A. C. Fabian, and J. P. Huchra. The rosat brightest cluster sample — iv. the extended sample. *Monthly Notices of the Royal Astronomical Society*, 318(2):333–340, Oct 2000.
- [10] David J. C. MacKay. *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, New York, NY, USA, 2002.
- [11] Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, and Jonathan Goodman. emcee: The MCMC Hammer. *PASP*, 125(925):306, Mar 2013.

- [12] Linda S. Sparke and John S. Gallagher, III. *Galaxies in the Universe: An Introduction*. Cambridge University Press, 2 edition, 2007.
- [13] R. Brent Tully, Hélène Courtois, Yehuda Hoffman, and Daniel Pomarède. The laniakea supercluster of galaxies. *Nature*, 513(7516):71–73, Sep 2014.
- [14] Gianfranco Bertone and Dan Hooper. History of dark matter. *Reviews of Modern Physics*, 90(4):045002, Oct 2018.
- [15] Risa H. Wechsler and Jeremy L. Tinker. The connection between galaxies and their dark matter halos. *Annual Review of Astronomy and Astrophysics*, 56(1):435–487, Sep 2018.
- [16] K. Freese, B. Fields, and D. Graff. Limits on stellar objects as the dark matter of our halo: nonbaryonic dark matter seems to be required. *Nuclear Physics B Proceedings Supplements*, 80:03/05, January 2000.
- [17] Sidney van den Bergh. The early history of dark matter. *Publications of the Astronomical Society of the Pacific*, 111(760):657–660, jun 1999.
- [18] Douglas Clowe, Maruša Bradač, Anthony H. Gonzalez, Maxim Markevitch, Scott W. Randall, Christine Jones, and Dennis Zaritsky. A direct empirical proof of the existence of dark matter. *The Astrophysical Journal*, 648(2):L109–L113, Aug 2006.
- [19] Jose A. R. Cembranos. Modified gravity and dark matter. *Journal of Physics: Conference Series*, 718:032004, may 2016.
- [20] William H. Press and Paul Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187:425–438, February 1974.
- [21] P.P. Coles and F. Lucchin. *Cosmology: The Origin and Evolution of Cosmic Structure*. Wiley, 2002.
- [22] Andrey Kravtsov and Stefano Borgani. Formation of galaxy clusters. 2012.
- [23] A. Jenkins, C. S. Frenk, S. D. M. White, J. M. Colberg, S. Cole, A. E. Evrard, H. M. P. Couchman, and N. Yoshida. The mass function of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 321(2):372–384, Feb 2001.
- [24] Jeremy Tinker, Andrey V. Kravtsov, Anatoly Klypin, Kevork Abazajian, Michael Warren, Gustavo Yepes, Stefan Gottlöber, and Daniel E. Holz. Toward a halo mass function for precision cosmology: The limits of universality. *The Astrophysical Journal*, 688(2):709–728, Dec 2008.
- [25] H. Böhringer, C. A. Collins, L. Guzzo, P. Schuecker, W. Voges, D. M. Neumann, S. Schindler, G. Chincarini, S. De Grandi, R. G. Cruddace, A. C. Edge, T. H. Reiprich, and P. Shaver. The ROSAT-ESO Flux-limited X-Ray (REFLEX) Galaxy Cluster Survey. IV. The X-Ray Luminosity Function. *ApJ*, 566(1):93–102, February 2002.

- [26] P. Schechter. An analytic expression for the luminosity function for galaxies. *ApJ*, 203:297–306, January 1976.
- [27] Thomas H. Reiprich and Hans Böhringer. The Mass Function of an X-Ray Flux-limited Sample of Galaxy Clusters. *ApJ*, 567(2):716–740, March 2002.
- [28] H. Böhringer, P. Schuecker, L. Guzzo, C. A. Collins, W. Voges, S. Schindler, D. M. Neumann, R. G. Cruddace, S. De Grandi, G. Chincarini, and et al. The rosat-eso flux limited x-ray (reflex) galaxy cluster survey. i. the construction of the cluster sample. *Astronomy & Astrophysics*, 369(3):826–850, Apr 2001.
- [29] B. Efron and C. Stein. The jackknife estimate of variance. *Ann. Statist.*, 9(3):586–596, 05 1981.
- [30] S. W. Allen, R. W. Schmidt, A. C. Fabian, and H. Ebeling. Cosmological constraints from the local x-ray luminosity function of the most x-ray-luminous galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 342(1):287–298, Jun 2003.
- [31] Jonathan Goodman and Jonathan Weare. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5(1):65–80, January 2010.
- [32] A. E. Evrard, T. J. MacFarland, H. M. P. Couchman, J. M. Colberg, N. Yoshida, S. D. M. White, A. Jenkins, C. S. Frenk, F. R. Pearce, J. A. Peacock, and P. A. Thomas. Galaxy Clusters in Hubble Volume Simulations: Cosmological Constraints from Sky Survey Populations. *ApJ*, 573(1):7–36, July 2002.
- [33] H. Ebeling, A. C. Edge, H. Bohringer, S. W. Allen, C. S. Crawford, A. C. Fabian, W. Voges, and J. P. Huchra. The ROSAT Brightest Cluster Sample - I. The compilation of the sample and the cluster log N-log S distribution. *MNRAS*, 301(4):881–914, December 1998.
- [34] A. Sokal. *Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms*, pages 131–192. Springer US, Boston, MA, 1997.
- [35] G. Chon and H. Böhringer. Cluster science from rosat to erosita. *Astronomische Nachrichten*, 334(4-5):478–481, May 2013.
- [36] J.A. Tyson, D.M. Wittman, J.F Hennawi, and D.N. Spergelb. Lsst: a complementary probe of dark energy. *Nuclear Physics B - Proceedings Supplements*, 124:21–29, Jul 2003.
- [37] George Casella and Roger Berger. *Statistical Inference*. Duxbury Resource Center, June 2001.