UNIVERSIDADE ESTADUAL DE LONDRINA

MASTER'S DISSERTATION

Competing Likelihood Models for Mass Estimate of Galaxy Clusters from Weak Lensing Data

---

Author:
Caio Lima DE OLIVEIRA

Supervisor:
Dr. Sandro D. P. VITENTI

A thesis submitted in fulfillment of the requirements
for the degree of Master's

in the

Theoretical Physics Group
Physics Department

May 17, 2023

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

de Oliveira, Caio.
Competing Likelihood Models for Mass Estimate of Galaxy Clusters from Weak Lensing Data / Caio de Oliveira. - Londrina, 2022.
59 f. : il.

Orientador: Sandro Vitenti.
Dissertação (Mestrado em Física) - Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em Física, 2022.
Inclui bibliografia.

1. Cosmologia - Tese. 2. Cosmologia Numérica - Tese. 3. Cosmologia Observacional - Tese. 4. Estatística - Tese. I. Vitenti, Sandro. II. Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa de Pós-Graduação em Física. III. Título.

CDU 53

"Space is big. You just won't believe how vastly, hugely, mind-bogglingly big it is. I mean, you may think it's a long way down the road to the chemist's, but that's just peanuts to space."

Douglas Adams, The Hitchhiker's Guide to the Galaxy

UNIVERSIDADE ESTADUAL DE LONDRINA

Abstract

Center for Exact Sciences

Physics Department

Master's

Competing Likelihood Models for Mass Estimate of Galaxy Clusters from Weak Lensing Data

by Caio Lima DE OLIVEIRA

Weak gravitational lensing is an extremely promising probe for the total mass of galaxy clusters. With first light for the LSST on the horizon, we will now finally have accurate enough data for mass estimates from weak lensing. However, for this to be possible, we will need an excellent understanding of the physical and statistical background of this proxy. In this context, this work is a first step in that direction. We analyze the statistical assumptions on the intrinsic distribution of galaxy ellipticities by finding the implications on more common parameters for ellipses. We also study and test the likelihoods models already implemented on LSST's CLMM by testing the normality of the shear in-bin distribution through Shapiro-Wilk tests. Finally, we propose a novel approach on likelihood model building with Kernel Density Estimation and compare it to the previous work done on CLMM by creating a mock data set to which each likelihood can be benchmarked and running best fits and MCMC analysis. We conclude that, first, a more in-depth study of the intrinsic distribution of galaxies shape may be necessary, second, that binned likelihoods must not assume gaussianity for shear distributions, and third, that KDEs are a promising way of building a weak lensing likelihood, potentially more accurate and precise than current binned methods, and competing with unbinned ones.

Acknowledgements

Esse trabalho carrega meu nome mas é a culminância do esforço de muitos. É a esses nomes que eu devo tudo. Queria agradecer então aos meus pais, Gracy e Aluizio, aos meus irmãos, Hugo, Bernardo e Gabriela, e à Zélia, por todo o amor e suporte que me deram nos últimos 24 anos. Agradeço também aos meus amigos da Universidade de Brasília, que mesmo longe me ajudaram tanto durante o Mestrado. A exceção à distância é o Henrique, com quem dividi a casa e todas as experiências do Mestrado e de Londrina no último ano. A ele também agradeço muito. Nessa nota, também tenho que agradecer a todos que conheci nessa cidade e que me acompanharam nesse trabalho, entre eles, Luiz, Ricardo, Sander, Eduardo e Julia. Em especial, tenho que agradecer à Vitoria, que me encorajou a ir além do que sonhei durante esse ano.

Finalmente, tenho que agradecer muito ao Sandro e à Mariana. Sem eles, não teria conseguido navegar essa cidade nova, tampouco conquistar o que conquistei durante esse Mestrado. Todos os agradecimentos a eles são poucos e sou extremamente grato.

E é claro, agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, que financiou esse trabalho.

Contents

Abstract v

Acknowledgements vii

1 Introduction 1
- 1.1 Legacy Survey of Space and Time 1

2 Galaxy Clusters and Weak Gravitational Lensing 3
- 2.1 Clusters and Cosmology 3
  - 2.1.1 Mass Proxies 3
    - Sunyaev-Zel'dovich effect 3
    - Richness 4
    - Gravitational Lensing 4
- 2.2 Past efforts 6
- 2.3 Theoretical Background 6
  - 2.3.1 Density Profile 6
  - 2.3.2 Gravitational Lensing 7
    - Measuring the shear matrix 8
    - Elliptical approximation 9

3 Modeling, Analysis, and Results 13
- 3.1 Likelihood Models and Implementation 13
  - 3.1.1 Working with the full data 14
  - 3.1.2 Summarizing data with binning 15
  - 3.1.3 Summarizing data with KDE 17
- 3.2 Testing 19
  - 3.2.1 Source ellipticity distribution 19
  - 3.2.2 In-bin distribution of reduced shear 19
- 3.3 Best fit and MCMC 31
  - 3.3.1 Best fit 31
  - 3.3.2 Posterior 33

4 Conclusion 35

A Statistics 37
- A.1 Bayes' Theorem 37

A.2 Likelihood and Maximum Likelihood Estimators 38
A.3 Fisher Information and Covariance Matrix 38
A.4 MCMC 39
A.5 Kernel Density Estimation 39

List of Figures

1.1 View of the Vera C. Rubin Observatory 2

2.1 Galaxy cluster SMACS J0723.3-727 5
2.2 Lens diagram 8
2.3 Ellipticity components in different coordinate systems 10

3.1 Mass estimate flow chart 14
3.2 Front and side views of the lens depicting the binning process 16
3.3 Analysis of the physical implications of assumptions on the source ellipticity distribution 20
3.4 Scatter plot of Shapiro-Wilk tests for different z distributions 22
3.5 Scatter plot of Shapiro-Wilk tests for different bin sizes 23
3.6 Scatter plot of deviation from mean p_{value} for each bin size 24
3.7 Scatter plot of deviation from mean p_{value} for z distribution 25
3.8 Shear KDE reconstruction for different ϵ_t^s standard deviation 26
3.9 In-bin histograms of g_t for different N_{bins}. Redshift distribution is Chang13 27
3.10 In-bin histograms of g_t for different N_{bins}. Redshift distribution is Chang13 + photo-z 28
3.11 In-bin histograms of g_t for different N_{bins}. Redshift distribution is fixed $z = 0.7$. 29
3.12 In-bin histograms of g_t for different N_{bins}. Redshift distribution is fixed $z = 0.7$ + photo-z 30
3.13 Posterior distributions 32

List of Tables

3.1 Mock data set parameters 18
3.2 Best fit and error for $\log_{10} M_\Delta$ 31
3.3 Best fit and error for c_Δ 31

To all the dreams of space I had as kid.

Chapter 1

Introduction

The objective of this work, as well as one of the objectives of our project in the Legacy Survey of Space and Time (LSST), Cluster Working Group (WG), is to create statistical and computational pipelines to estimate the mass of individual clusters. So far, this pipeline relies on the Cluster weak Lensing Mass Modeling (CLMM) python library [1], which provides a framework for the analysis of weak lensing data and for the creation of mock data sets. We are thus testing their current implementation, as well as implementing alternative approaches and comparing them to CLMM's. These approaches may, if proven good, be implemented on modules on the Numerical Cosmology (NumCosmo) library [2], as well as in CLMM and Firecrown [3].

In that sense, we provide here in the Introduction a brief description of the LSST. Then, in Chapter 2 we display the physical background of galaxy cluster and gravitational lensing. In Chapter 3, we will explain how the likelihood models for weak lensing, including our own, have been built; what are the statistical assumptions involved in this process; and finally test our models and assumptions. In the end, Chapter 4, we write a brief conclusion with the summary of our findings, and what should our next steps be.

1.1 Legacy Survey of Space and Time

It is in the context of the LSST that our work takes place. With "a wide-field, ground-based telescope", located on the Vera C. Rubin Observatory (visible on Figure 1.1) in Cerro Pachón, Chile, its goals are to [4]

- Probe cosmological observables in order to study dark energy and dark matter;
- Take an inventory of the solar system;
- Explore the transient optical sky;
- Map the Milky Way.

Responsible for the handling of the data gathered by the LSST and its use to make high accuracy cosmology, the Dark Energy Science Collaboration (DESC) is divided into eighteen Working Groups, each with its own focus on the analysis and development of tools for DESC. We are currently working within DESC's Cluster Working Group, whose priorities are

- Accurate cosmological predictions;
- Cluster detection;
- Characterization of the survey observable-mass relation;
- Building a likelihood framework.

With first light expected on March 2024 and full system operation by the end of 2024, LSST will survey about 30000 deg^2 over 10 years, focusing on a main 18000 deg^2 area [4]. This main area will be 5 mag deeper than the Sloan Digital Sky Survey (SDSS), the current largest existing optical survey.

LSST will provide the scientific community with a wealth of weak gravitational lensing data in the coming years. These will enable high precision weak lensing analysis of clusters, and it is the responsibility of the Clusters Working Group to create the tools and framework for the processing and analysis of this data.

FIGURE 1.1: Vera C. Rubin Observatory in Cerro Pachón, Chile, as of 2021. Figure taken from [5].

Chapter 2

Galaxy Clusters and Weak Gravitational Lensing

2.1 Clusters and Cosmology

Galaxy clusters are the largest gravitationally bound structures of the Universe. Covering a scale we expect to be predominantly described by linear perturbation theory, clusters provide an observational window to the structure formation as predicted by linear perturbation. Their spatial and mass distributions allow us to study the large scale structure of our Universe, as well as use them as probes for other cosmological parameters.

It is not simple to estimate their mass, however, as it is not an observable that we can directly estimate. We need proxies, that is: physical quantities that can be measured and have (hopefully) strong correlation with the total cluster mass. Then, we can model the mass observable distribution, and use it to estimate the total mass. Multiple observables can be and are used alone or in conjunction, each one with its own advantages. These are discussed below.

2.1.1 Mass Proxies

Sunyaev-Zel'dovich effect

Cosmic Microwave Background (CMB) radiation photons interact with inter-cluster gas by Compton scattering, causing a spectral distortion, which can be measured as a temperature variation with respect to the CMB temperature T_{CMB}

$$\frac{\Delta T_{SZE}}{T_{CMB}} = f(x) \int n_e \frac{k_B T_e}{m_e c^2} \sigma_T \mathrm{d}l, \tag{2.1}$$

where n_e is the electron number density, $m_e c^2$ is the electron rest mass, T_e is the electron temperature, k_B is the Boltzmann constant, σ_T is the Thomson cross-section, $x \equiv \frac{h\nu}{k_B T_{CMB}}$ is the dimensionless frequency, and dl is the line of sight segment. $f(x)$ is the frequency dependence of the Sunyaev-Zel'dovich effect, given by

$$f(x) = \left(x \frac{e^x + 1}{e^x - 1} - 4 \right) (1 + \delta_{SZE}(x, T_e)) \,, \tag{2.2}$$

where $\delta_{SZE}(x, T_e)$ is the relativistic correction to the frequency dependence.

By integrating over a solid angle we are counting the number of electrons, weighted by temperature. This is proportional to the mass of electrons and, since we know the composition of the intercluster gas, to the mass of this gas

$$\int \Delta_{T_{SZE}} \mathrm{d}\Omega \propto \frac{M_{gas} \langle T_e \rangle}{D_{cl}^2}, \tag{2.3}$$

where $\langle T_e \rangle$ is the mean temperature of the electrons and D_{cl} is the angular diameter distance to the cluster.

Since the Sunyaev-Zel'dovich effect is caused by CMB photons interacting with the electrons in the inter-cluster gas, it provides a measurement only of the mass of gas. This can still be used to estimate the mass of the cluster by calibrating the observable-mass relation through the mass estimated by other proxies. For more information, see [6–8].

Richness

The richness λ of a cluster is the number of galaxies pertaining to that cluster, limited by a certain radius R_{cl} and a certain luminosity L_{cut}. This is not an easy number to obtain: there are also thousands of galaxies in the background and foreground of the cluster and the errors in the measurement of the observable parameters of each galaxy (position, redshift, luminosity, color, etc.) make it an observable with potential large variance and model dependency. Hence, measuring the richness of a cluster becomes a problem of statistical nature - what is the probability of a given galaxy being a member of the cluster? There are several algorithms to determine that and the richness λ. See [9, 10], for example, for information on the algorithm used by the redMaPPer project[1].

Gravitational Lensing

Gravitational lensing is a relativistic effect in which an object, due to its gravitational potential, bends the spacetime around it, leading to the deflection of light rays passing it. This results in shifted, distorted, and magnified images, which we can use to calculate the gravitational potential to which the light was submitted. When considering the weak-lensing regime, where the effects of lensing are small, this deflection is described by the lens equation

$$\beta^i = \theta^i - \alpha^i, \tag{2.4}$$

---

[1] The source for the code can be found on `https://github.com/erykoff/redmapper`.

where, in relation to the line of sight to the center of the lens, β^i is the angle of the source, θ^i is the angle of the image, and α^i is the reduced deflection angle, given by

$$\alpha^i = \int_0^{w_s} \frac{f_K(w_s - w)}{f_K(w_s) f_K(w)} \frac{\partial \Phi}{\partial \theta^i} (f_K(w)\theta^i, w) \mathrm{d}w. \tag{2.5}$$

Here, w is the comoving radial distance, w_s is the comoving radial distance of the source, f_K is the comoving angular diameter distance, and Φ is the Newtonian gravitational potential of the lens.

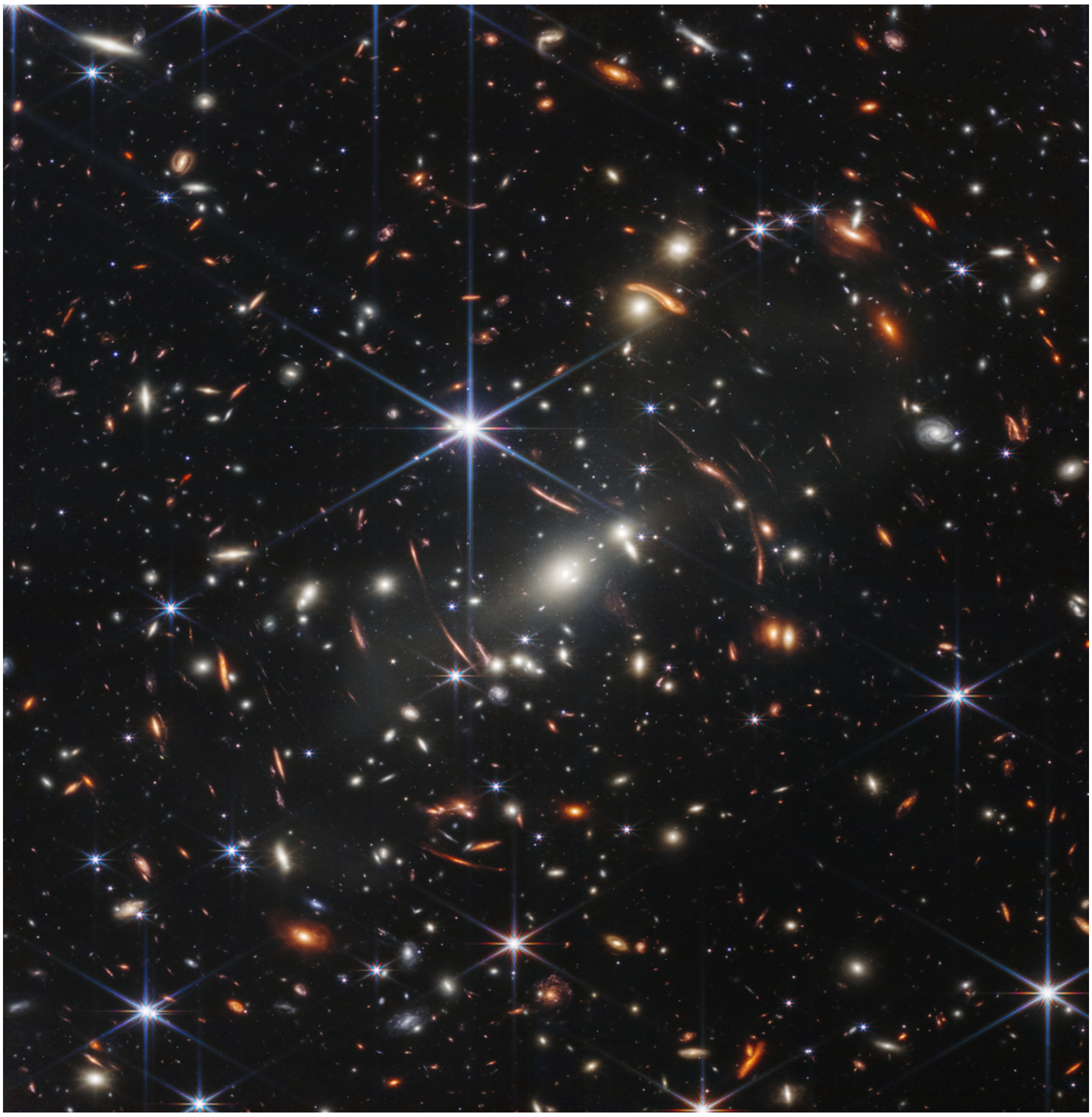


FIGURE 2.1: Image of galaxy cluster SMACS J0723.3-7327 as captured by James Webb Space Telescope. On it, it is possible to see both strong lensing effects (the elongated galaxies near the center of the cluster) and weak lensing effects - notice how galaxies seem to be aligned to circles around the cluster center. Figure taken from [11].

By measuring the deflection and/or other quantities affected by the deflection, we are indirectly measuring the gravitational potential, through which we can find the mass of the lens. We discuss this point further on Section 2.3.2. It is important to note that lensing is then a direct probe to the *total* mass of the cluster - the sum of the galactic, inter-cluster, and dark matter halo mass.

There are various methods to estimate cluster mass using gravitational lensing. They are usually divided into two main categories:

- **Strong Lensing**: mostly represented by more extreme showings of lensing effects, such as arcs and multiple images.
- **Weak Lensing**: contains less spectacular showings of lensing effects, such as small distortions on the images of the source objects and magnification of their brightness.

Each of these categories has its advantages. Strong lensing is usually easier to identify, but are statistically less likely to be observed. On the other side, weak lensing phenomena is extremely common on clusters, and thus more statistically powerful, but difficult to observe: powerful telescopes are necessary to observe the ellipticity of background galaxies. In this work we focus on a method that utilizes weak lensing to estimate the mass of clusters.

2.2 Past efforts

As a quick and short summary, in the past, strong and weak lensing phenomena have been used to observe high redshift galaxies [12, 13], estimate the mass of clusters [14–16], as well as measure cosmological parameters [17]. Most of the studies that relied on weak lensing were somewhat limited by the resolution of the telescopes of the time however, and as such, methods like the one presented on our work could not be used. It is only with recent projects like the Dark Energy Survey (DES) and upcoming ones like the Legacy Survey of Space and Time (LSST) and Euclid, that the true power of weak lensing analysis can finally be achieved.

2.3 Theoretical Background

2.3.1 Density Profile

In order to estimate the cluster mass for a certain shear distribution, we need a model for the density profile of the cluster. There have been a few different proposals for such a model [18–20], but we chose to focus this work on the proposed by [21, 22]. In 1996, J. Navarro, C. Frenk, and S. White, fitted an approximate radial model for the density profile of dark matter halos, which since then has been name NFW profile and shown to be a near universally good description[23]. It is given by

$$\rho(r) = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2}, \tag{2.6}$$

where ρ_s and r_s are the characteristic density and radius of the halo. They are defined as

$$\rho_s = \frac{1}{3}\frac{c^3\rho_{ref}\Delta}{\ln(1+c) - c/(1+c)}, \qquad r_s = \frac{r_\Delta}{c}. \tag{2.7}$$

Here, c is the concentration parameter, Δ is the overdensity parameter, ρ_{ref} is the reference density, and r_Δ is the radius within which the average density is ρ_Δ. The mass of such cluster is then

$$M_\Delta = \frac{4\pi\Delta}{3} r_\Delta^3 \rho_{ref}. \tag{2.8}$$

It is important to notice that this profile, parametrized by ρ_s and r_s, can be re-parametrized in terms of M_Δ and c:

$$\rho(r) = \frac{1}{4\pi}\frac{1}{\ln(1+c) - \frac{c}{(1+c)}}\frac{c^3 M_\Delta}{cr\left(\frac{3M_\Delta}{4\pi\rho_\Delta}\right)^{2/3} + 2c^2r^2\left(\frac{3M_\Delta}{4\pi\rho_\Delta}\right)^{1/3} + c^3r^3}. \tag{2.9}$$

2.3.2 Gravitational Lensing

Our goal in this section is to describe what is our observable of interest in gravitational lensing and how it relates to the mass M_Δ defined in the previous section. This description is rouhgly based on a few different sources [24–28].

As we mentioned before, the basic result in gravitational lensing is the lens equation,

$$\beta^i = \theta^i - \alpha^i, \tag{2.10}$$

where β^i is the source angle, θ^i is the image angle and α^i is the reduced deflection angle (see Figure 2.2). Given a Cartesian coordinate system on the lens plane given by $\vartheta_{1,2}$, the reduced deflection α^i can be written as the gradient of a lens potential,

$$\alpha^i = \frac{\partial\psi}{\partial\vartheta^i} = \int_0^{w_s} \frac{f_K(w_s - w)}{f_K(w_s)f_K(w)}\frac{\partial\Phi}{\partial\vartheta^i}(f_K(w)\vartheta^i, w)\mathrm{d}w, \tag{2.11}$$

with w the comoving radial distance, w_s the comoving radial distance of the source, f_K the comoving angular diameter distance, and Φ the Newtonian gravitational potential of the cluster.

For small deflections, α_i may be approximated to first order by

$$\alpha^i \approx M^i{}_j\theta^j, \tag{2.12}$$

where M^i_j is the shear matrix, which can be conveniently written as

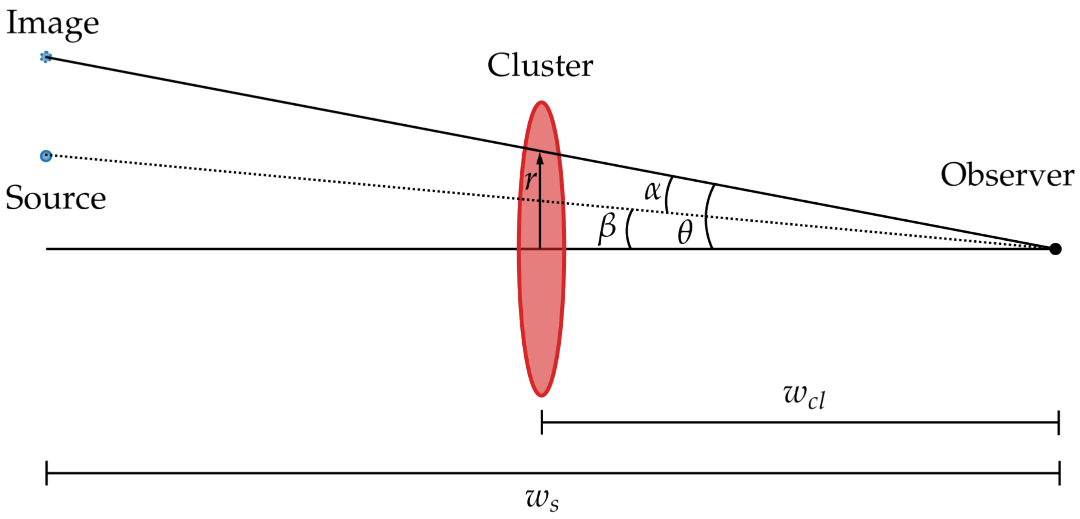


FIGURE 2.2: Diagram displaying the lens and the relevant angular distances. Here, θ is the angle of the observed image, β is the angle of the source, and α is the reduced deflection angle. The reader may also notice the distance of the image from the line of sight on the lens plane r, the comoving radial distance from the observer to the cluster w_{cl}, and the comoving radial distance from the observer to the source w_s.

$$M^i{}_j = \frac{\partial \alpha^i}{\partial \vartheta^j} = \begin{bmatrix} \kappa + \gamma_1 & \gamma_2 \\ \gamma_2 & \kappa - \gamma_1 \end{bmatrix}. \tag{2.13}$$

Here κ is the convergence and $\gamma_{1,2}$ is the shear field. These give, respectively, the isotropic and anisotropic behavior of the lens, and can be calculated from the lens potential by

$$\kappa = \frac{1}{2}\left(\frac{\partial^2\psi}{\partial\vartheta^1\partial\vartheta^1} + \frac{\partial^2\psi}{\partial\vartheta^2\partial\vartheta^2}\right), \quad \gamma_1 = \frac{1}{2}\left(\frac{\partial^2\psi}{\partial\vartheta^1\partial\vartheta^1} - \frac{\partial^2\psi}{\partial\vartheta^2\partial\vartheta^2}\right), \quad \gamma_2 = \frac{\partial^2\psi}{\partial\vartheta^1\partial\vartheta^2} = \frac{\partial^2\psi}{\partial\vartheta^2\partial\vartheta^1}. \tag{2.14}$$

Measuring the shear matrix

Our real observable is the surface brightness $I(\vartheta)$ of the image captured by our telescopes. We may describe the shape of an observed galaxy by its quadrupole moment of surface brightness

$$Q_{ij} \equiv \frac{\int \mathrm{d}^2\vartheta I(\vartheta)\vartheta_i\vartheta_j}{\int \mathrm{d}^2\vartheta I(\vartheta)}. \tag{2.15}$$

In the weak lensing regime, $|M_{ij}| \ll 1$, the source quadrupole moment Q^s_{ij} is mapped to an observed quadrupole Q_{ij} by the shear matrix through

$$Q_{ij} = ({\delta_i}^k + {M_i}^k)({\delta_j}^l + {M_j}^l)Q^s_{kl}, \tag{2.16}$$

$$Q_{ij} \approx Q^s_{ij} + {M_i}^k Q^s_{kj} + {M_j}^l Q^s_{il}. \tag{2.17}$$

The power of weak lensing surveys comes from the large number of observations, with which we can take the average of Q_{ij} and measure M_{ij}. Let us take a certain group of galaxies with similar image positions θ^i and redshift z. For this sample of galaxies, the shear matrix is approximately constant. Then, if the orientation of the source galaxies is random, the average $\langle Q^s_{ij}\rangle$ is proportional to the identity matrix, $\langle Q^s_{ij}\rangle = \frac{1}{2}\bar{Q}\delta_{ij}$, and the average of Q_{ij} is

$$\langle Q_{ij}\rangle = \bar{Q}\left(\frac{1}{2}\delta_{ij} + M_{ij}\right), \tag{2.18}$$

which gives us the shear matrix up to a constant $\bar{Q}$. This constant, however, is simply the average $\langle \mathrm{tr}\, Q^s\rangle = \bar{Q}$. Taking the average for the trace of Q_{ij} we have

$$\langle \mathrm{tr}\, Q\rangle = \bar{Q}\,(1 + \mathrm{tr}\, M)\,. \tag{2.19}$$

Finally, if we divide 2.18 by 2.19, we find to first order in M_{ij}

$$\frac{\langle Q_{ij}\rangle}{\langle \mathrm{tr}\, Q\rangle} = \frac{1}{2}\delta_{ij} + M_{ij} - \frac{1}{2}\delta_{ij}\,\mathrm{tr}\, M. \tag{2.20}$$

The isotropic component of this matrix of averages will simply be half of the identity matrix. The anisotropic component however will result in a measurement of the shear field $\gamma_{1,2}$, the anisotropic component of M_{ij}:

$$\gamma_1 = \frac{\langle Q_{11}\rangle - \langle Q_{22}\rangle}{\langle Q_{11} + Q_{22}\rangle}, \qquad \gamma_2 = \frac{\langle Q_{12}\rangle}{\langle Q_{11} + Q_{22}\rangle}. \tag{2.21}$$

Elliptical approximation

What is commonly done, and that is the current practice on CLMM, is to illustrate the shape of galaxies using ellipses, with complex ellipticity $\epsilon = \epsilon_1 + i\epsilon_2$, calculated from the quadrupole moment:

$$\epsilon_1 = \frac{Q_{11} - Q_{22}}{2N_Q}, \qquad \epsilon_2 = \frac{Q_{12}}{N_Q}, \qquad N_Q = \frac{1}{2}\,\mathrm{tr}\, Q + \sqrt{\det Q}. \tag{2.22}$$

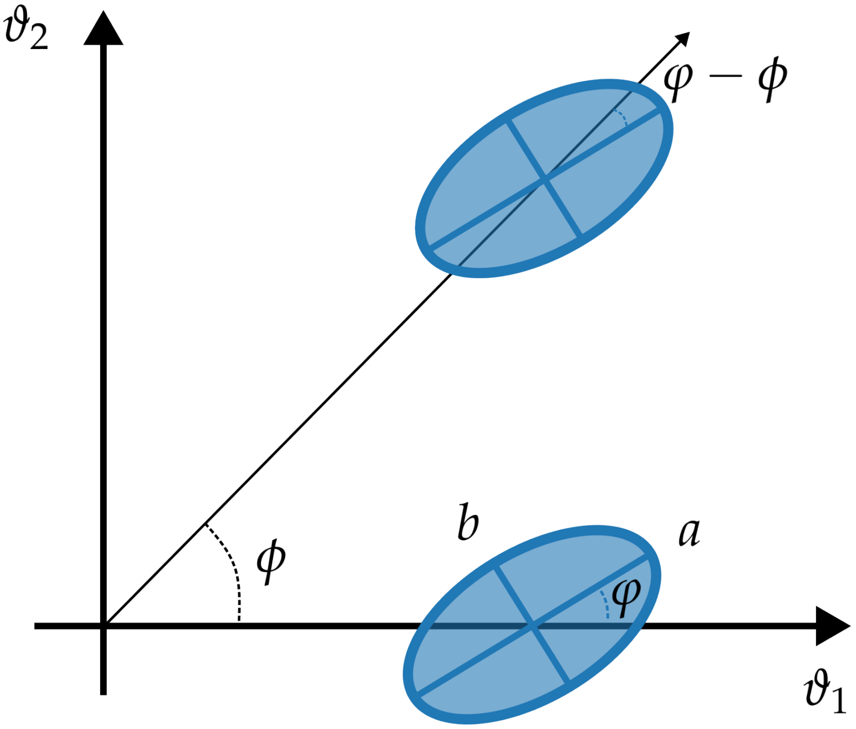


FIGURE 2.3: Ellipticity components in different coordinate systems

For an ellipse of major semi-axis a forming an angle φ in relation to ϑ_1 and minor semi-axis b, if we model the surface brightness to be 1 inside the ellipse and 0 outside, then the ellipticities become

$$\epsilon_1 = \frac{a-b}{a+b}\cos 2\varphi, \qquad \epsilon_2 = -\frac{a-b}{a+b}\sin 2\varphi, \tag{2.23}$$

or

$$\epsilon = \epsilon_1 + i\epsilon_2 = \frac{a-b}{a+b}e^{-2i\varphi} = \frac{1-r}{1+r}e^{-2i\varphi}, \qquad r \equiv \frac{b}{a}. \tag{2.24}$$

Note that these components are calculated in the Cartesian coordinate system $\vartheta_{1,2}$ - this is evident on the dependence on the angle φ. For an ellipse not laying on the axis ϑ_1, say, having an angle ϕ from it, it is most convenient to rotate the coordinate system by such angle ϕ and calculate the components in this rotated system. In that case, we simply multiply ϵ by $e^{2i\phi}$ to find the new components $\epsilon_{t,+}$

$$\epsilon_t = \epsilon_1 \sin 2\phi - \epsilon_2 \cos 2\phi, \qquad \epsilon_+ = -\epsilon_1 \cos 2\phi - \epsilon_2 \sin 2\phi, \tag{2.25}$$

which we call tangential and cross, respectively.

Parametrizing the quadrupole moment Q_{ij} by the complex ellipticity ϵ, the mapping from source to lensed data 2.16 becomes

$$\epsilon = \frac{\epsilon^s + g}{1 + g^* \epsilon^s}, \tag{2.26}$$

where we define g, the reduced shear, calculated from the shear matrix components by:

$$g = \frac{\gamma_1}{1-\kappa} + i\frac{\gamma_2}{1-\kappa}. \tag{2.27}$$

For $g \ll 1$, equation 2.26 can be approximated as

$$\epsilon \approx \epsilon^s + g. \tag{2.28}$$

The observational dilemma now becomes how to measure the reduced shear. The true observable here is the final ellipticity ϵ, but that differs from the reduced shear by the intrinsic ellipticity ϵ^s of the source galaxy, which is a random statistical variable. The relationship between those quantities become that of signal and noise - reduced shear and source ellipticity, respectively. Our job is then to somehow estimate the signal from the observable. To do this we need to build a certain model for the distribution of ϵ^s. Such model needs to have an expected value $\langle \epsilon^s \rangle = 0$, since we expect statistical isotropy. By looking at equation 2.28, in the case g is not a statistical variable but a perfectly defined number, then

$$\langle \epsilon \rangle = g. \tag{2.29}$$

Now, if we assume a probability distribution $\mathfrak{f}(\epsilon^s)$ for the source ellipticity, then the distribution $f(\epsilon)$ is the same, but shifted by the reduced shear g

$$f(\epsilon) \approx \mathfrak{f}(\epsilon - g). \tag{2.30}$$

We again raise the reader's attention to the fact that this is only true for g with no statistical distribution. We'll see in Chapter 3 (Modeling, Analysis, and Results) that this will not always be true, and will depend entirely on how we approach our likelihood model.

Finally, we note that for spherically symmetric density profiles, as is the case of the NFW profile used here, the cross component of the reduced shear vanishes. It is only necessary to care for the tangential component then, and so, we will only model our likelihood for ϵ_t.

Chapter 3

Modeling, Analysis, and Results

As stated in the Introduction, our goal for this work is to estimate the mass of galaxy clusters from weak lensing data. This process may be divided into two parts. The first one involves the physical aspects of the work: we need to make physical assumptions on clusters and weak lensing in order to create a physical model. This was already established on Chapter 2. The second part, which we describe now, is the statistical one, where we make our statistical assumptions on clusters, decide which tools to use, and build our likelihood model, all while testing these assumptions. The next step is to then use the likelihood to extract estimates and confidence regions for our parameters of interest (in this case, the cluster mass M_Δ) utilizing Maximum Likelihood Estimators (MLE) for point estimates of the parameters and Markov Chain Monte Carlo (MCMC) methods for the full exploration of the posterior distribution.

A basic description of this statistical process can be seen on Appendix A. Sections 3.1 and 3.2 describe our statistical assumptions, our likelihood model building and the testing done. Finally, Section 3.3 shows the results found for the models discussed.

3.1 Likelihood Models and Implementation

Now, our first assumption is that the source ellipticities $\epsilon^s_{1,2}$ are normally distributed with an average 0 and standard deviation σ_{ϵ^s}, that is

$$\epsilon^s_1 \propto \mathcal{N}(0, \sigma_{\epsilon^s}) \qquad \epsilon^s_2 \propto \mathcal{N}(0, \sigma_{\epsilon^s}), \tag{3.1}$$

which implies

$$\epsilon^s_t \propto \mathcal{N}(0, \sigma_{\epsilon^s}) \qquad \epsilon^s_+ \propto \mathcal{N}(0, \sigma_{\epsilon^s}). \tag{3.2}$$

This is not necessarily a good model for the distribution of source ellipticities, as we'll see in Section 3.2. However, CLMM draws the ellipticities of mock galaxies from such probability densities, so all our likelihoods will need to take that into account.

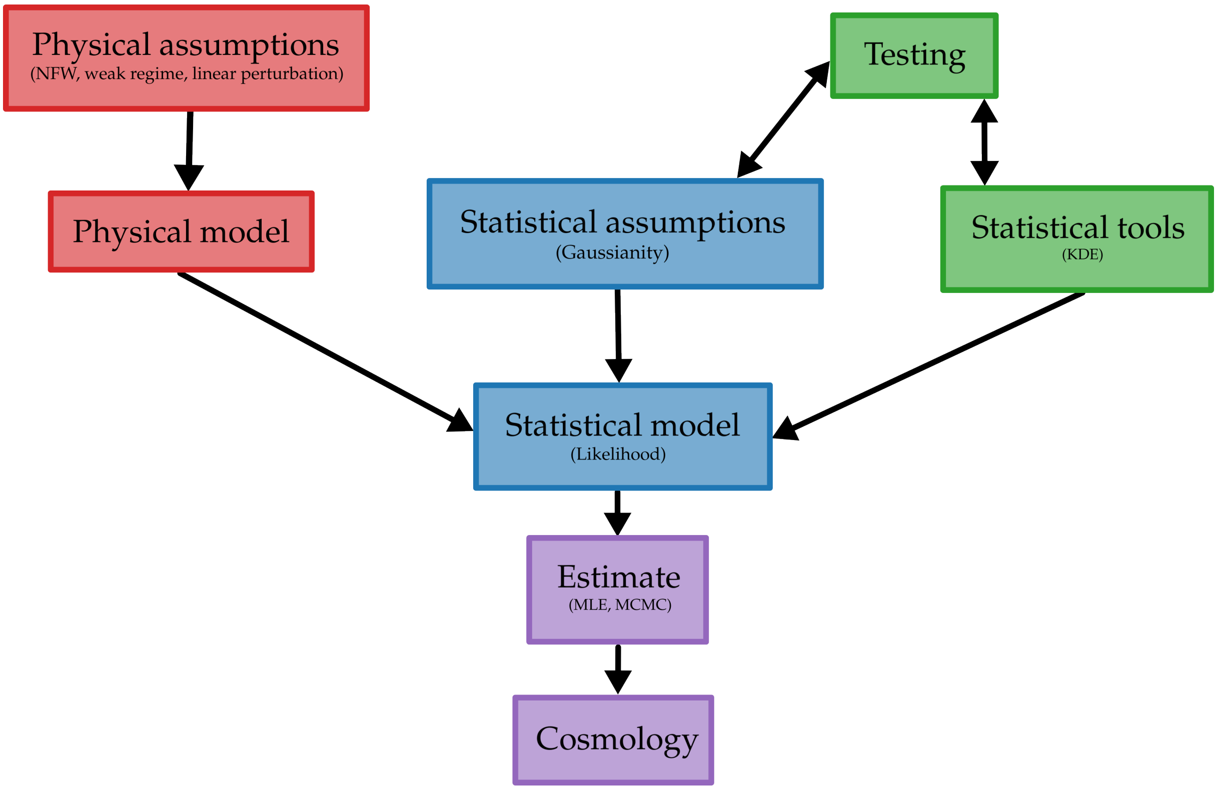

Physical assumptions
(NFW, weak regime, linear perturbation)
Physical model
Testing
Statistical assumptions
(Gaussianity)
Statistical tools
(KDE)
Statistical model
(Likelihood)
Estimate
(MLE, MCMC)
Cosmology

FIGURE 3.1: Entire mass estimate process. We start by making some physical assumptions to build a physical model, which, in combination with some statistical assumptions and tools, we build a statistical model. This model is then used to extract information on parameters of interest, which will then be used to inform Cosmology.

3.1.1 Working with the full data

At its simplest, our data on the background galaxies can be reduced into three variables: tangential ellipticity ϵ_t, angular radius θ, and redshift z. Remember that θ is the angle of the image in relation to the center of the lens (as can be seen on Figure 2.2). Both the tangential ellipticity and angular radius are measured with great precision and accuracy, being observed directly from the pixel brightness in the CCD. Their errors can be safely ignored then. The redshift measurement, though, can be a bit more complex. Depending on how it was measured, spectroscopically or photometrically, we might have to take their error into account.

Let us start with the simplest case: spectroscopic redshift. In this case, since both the angular radius and the redshift are known with almost no error, the reduced shear can be considered a deterministic variable, with no probability distribution. We can then use equations 2.30 and 3.2 to write the probability of measuring the ellipticity $\epsilon_{t,i}$ of the i-th galaxy given the reduced shear $g_{t,i}(M_\Delta, c_\Delta) = g_t(\theta_i, z_i, M_\Delta, c_\Delta)$ in the cluster as

$$\mathcal{P}\left(\epsilon_{t,i} | g_{t,i}\left(M_\Delta, c_\Delta\right)\right) = \frac{1}{\sqrt{2\pi}\sigma_{\epsilon^s}} e^{-\left(\epsilon_{t,i} - g_{t,i}(z,\theta,M_\Delta,c_\Delta)\right)^2/2\sigma^2_{\epsilon^s}}. \tag{3.3}$$

The likelihood of our N data points is then simply the product of all probabilities $\mathcal{P}\left(\epsilon_{t,i}\right)$

$$\mathcal{L}\left(\{\epsilon_t\}\,|M_\Delta, c_\Delta\right) = \prod_i^N \mathcal{P}\left(\epsilon_{t,i}|g_{t,i}\left(M_\Delta, c_\Delta\right)\right). \tag{3.4}$$

This is simple enough and easy to compute. Things become more difficult if the redshift data is photometric, though, as each galaxy will be assigned a photo-z distribution associated to the redshift measurement. In that case the reduced shear cannot be considered a deterministic variable, and we must take into account the probability $\mathcal{P}_i(z_p|z)$ of measuring the photometric redshift z_p given a real redshift z, **for a given specific galaxy**. The likelihood is then given by

$$\mathcal{L}\left(\epsilon_t, z_p|M_\Delta, c_\Delta\right) = \prod_i^N \int_{z_{cl}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_{\epsilon^s}} e^{-\left(\epsilon_{t,i}-g_{t,i}(M_\Delta,c_\Delta)\right)^2/2\sigma_{\epsilon^s}^2} \mathcal{P}_i(z_{p,i}|z)\mathrm{d}z. \tag{3.5}$$

Note that in the expression above we are computing the probability of observing $\epsilon_{t,i}$ and $z_{p,i}$ given the model and the true redshift z. When we integrate over z we are implicitly using a uniform prior on z representing our ignorance about the true redshift of each galaxy.

The computation time for this likelihood is much larger than the previous, and will vary depending on the distributions $\mathcal{P}_i(z_p|z)$. CLMM utilizes NumCosmo to implement this likelihood for the special case in which $\mathcal{P}_i(z_p|z)$ is Gaussian,

$$\mathcal{P}_i(z_p|z) = \frac{1}{\sqrt{2\pi}\sigma_z} e^{-\left(z_p - z\right)^2/2\sigma_z^2}, \qquad \sigma_z \equiv \sigma_{z0}(1+z), \tag{3.6}$$

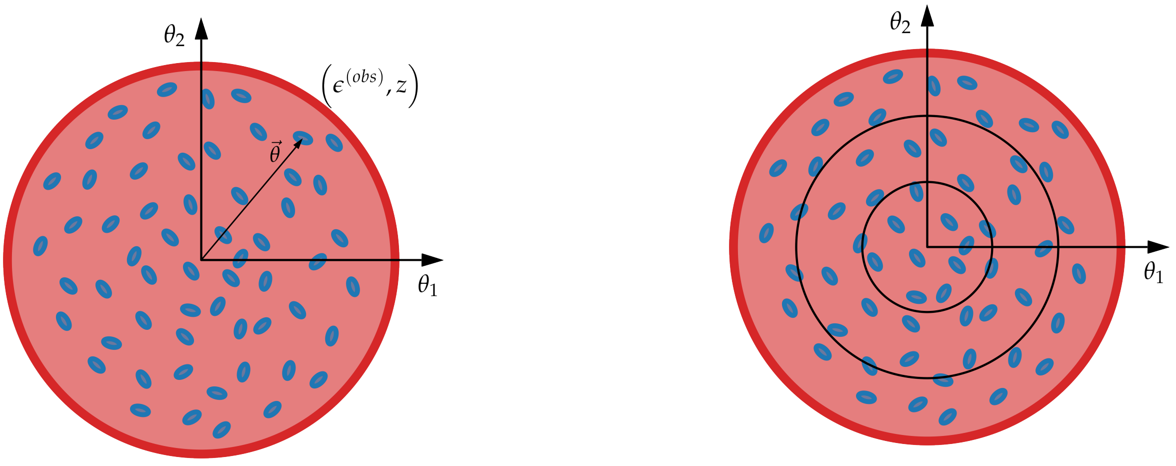
where σ_{z0} is a constant.

Note that even for a simple $\mathcal{P}_i(z_p|z)$ distribution, as the one given above, computation time may become unreasonable when we take into account that each cluster lenses something on the order of 10^5 background galaxies, and we may have to deal with thousands of clusters on LSST. We must then study ways of summarizing and approximating the data in a way that decreases computation time while retaining as much precision and accuracy as possible.

3.1.2 Summarizing data with binning

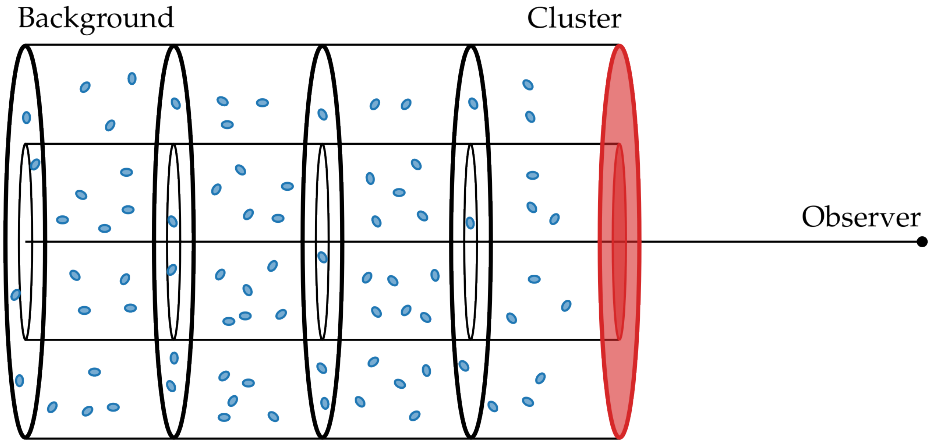
The simplest and most common way of summarizing the data is to bin it. The idea is that by grouping objects with closely related observables, we may approximate each group by their average measurements. We will then only have to deal with M data points, where M is the number of bins.

This technique is already used on some of CLMM's models, where the galaxies are binned by angular radius θ. The binning method can be seen on Figures 3.2a and 3.2b. Note that this process may not be as simple as taking an average though, as there are two complexities we must take into account. Let us consider the case of spectroscopic redshift. As there are no errors on θ and z, we may be tempted to simply calculate the reduced shear g_t at $(\langle\theta\rangle, \langle z\rangle)$ and then,



(A) Front view of the cluster. The coordinate system $\vartheta_{1,2}$ is shown along with the background galaxies and their associated 3 parameters: $\vec{\theta}$, ϵ, and z.

(B) Front view of the cluster. The radial binning is shown as black concentric circles. Note that each bin is composed of a group of galaxies, each with its own information on ϵ and z, with which we build the in-bin distribution.



(C) Side view of the cluster. It again shows the binning, highlighting the fact that the background galaxies are really distributed in three-dimensional space. Each bin is then a layer on a cylinder behind the cluster, again encompassing a group of galaxies with 3 associated parameters: $\vec{\theta}$, ϵ, and z (which may vary freely inside each bin).

FIGURE 3.2: Front and side views of the lens depicting the binning process. The red circle represents the projection of the spherical model of the cluster onto the image plane, while blue circles represent the background galaxies.

similarly to what was done on Equation 3.3, compute the bin likelihood as the probability of measuring the mean ellipticity $\langle\epsilon_t\rangle$ given the reduced shear $g_t\left(\langle\theta\rangle, \langle z\rangle\right)$, that is

$$\mathcal{L}_i\left(\langle\epsilon_t\rangle_i | M_\Delta, c_\Delta\right) = \frac{1}{\sqrt{2\pi}\sigma_{\epsilon^s}} e^{-\left(\langle\epsilon_t\rangle_i - g_t\left(\langle\theta\rangle_i, \langle z\rangle_i, M_\Delta, c_\Delta\right)\right)^2 / 2\sigma^2_{\epsilon^s}}. \tag{3.7}$$

There is a glaring problem here: when we bin by angular radius θ the galaxies within each bin will all have similar values of θ, but that is not true for the redshift z. Each bin will have a certain uncertainty on the redshift caused by the fact that the grouped galaxies will have completely different values of redshift (see Figure 3.2). We must then account for the in-bin true redshift distribution $\mathcal{P}_i(z)$ for the **group of galaxies** measured in that bin, when modeling the likelihood:

$$\mathcal{L}_i\left(\langle\epsilon_t\rangle_i|M_\Delta,c_\Delta\right)=\int_{z_{cl}}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_{\epsilon^s}}e^{-(\langle\epsilon_t\rangle_i-g_t(\langle\theta\rangle_i,z,M_\Delta,c_\Delta))^2/2\sigma^2_{\epsilon^s}}\mathcal{P}_i(z)\mathrm{d}z, \tag{3.8}$$

where the trade-off between having to deal with fewer data points and losing information on the redshift becomes apparent.

The second complexity comes from the already discussed problem of photometric redshift. Besides having to deal with the in-bin distribution of g_t, caused by the in-bin distribution of the measured photometric redshift z_p, we must also take into account the probability $\mathcal{P}_i(\{z_p\}|\mathcal{P}_i)$ of obtaining all photometric information of the group of galaxies in the bin given true in-bin distribution. Then we need the distribution of g_t given the same true in-bin distribution. Such likelihood may be written as

$$\mathcal{L}_i\left(\langle\epsilon_t\rangle_i|M_\Delta,c_\Delta\right)=\int_{z_{cl}}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_{\epsilon^s}}e^{-(\langle\epsilon_t\rangle_i-g_t(\langle\theta\rangle_i,z,M_\Delta,c_\Delta))^2/2\sigma^2_{\epsilon^s}}\mathcal{P}_i(z)\mathcal{P}_i(\{z_p\}|\mathcal{P}_i)\mathrm{d}z. \tag{3.9}$$

Currently, CLMM has implementations for the basic case of mean redshift and for spectroscopic redshift, where they acknowledge the redshift distribution and the reduced shear distribution consequently caused. There is a possible source of bias in their implementation, however. The in-bin distribution of reduced shear is being summarized by its average, essentially approximating this distribution as Gaussian - which might not be a good approximation. The bin likelihood is then

$$\mathcal{L}_i\left(\langle\epsilon_t\rangle_i|M_\Delta,c_\Delta\right)=\frac{1}{N_i}\sum_{j_i}^{N_i}\frac{1}{\sqrt{2\pi}\sigma_{\epsilon^s}}e^{-\left(\langle\epsilon_t\rangle_i-g_t(\langle\theta\rangle_i,z_{j_i},M_\Delta,c_\Delta)\right)^2/2\sigma^2_{\epsilon^s}} \tag{3.10}$$

Both these implementations can be found on CLMM's example notebooks.[1]

3.1.3 Summarizing data with KDE

An alternative we propose to the binning method is to utilize Kernel Density Estimation (see Appendix A). By considering the reduced shear a statistic variable to which we can assign a distribution, we are able to use Bayes' theorem to write

$$\mathcal{L}\left(\epsilon_{t,i}|M_\Delta,c_\Delta\right)=\int_{-\infty}^{\infty}\int_{-1}^{1}\mathcal{P}\left(\epsilon_{t,i}|g_{t,i}\left(M_\Delta,c_\Delta\right),\epsilon^s_{t,i}\right)\mathcal{P}\left(g_{t,i}\left(M_\Delta,c_\Delta\right)\right)\mathcal{P}\left(\epsilon^s_{t,i}\right)\mathrm{d}\epsilon^s_{t,i}\mathrm{d}g_{t,i}. \tag{3.11}$$

---

[1] Source code can be found on `https://github.com/LSSTDESC/CLMM/tree/main/examples`

From 2.28 we may write $\mathcal{P}\left(\epsilon_{t,i}|g_{t,i}\left(M_\Delta,c_\Delta\right),\epsilon^s_{t,i}\right) = \delta\left(\epsilon_{t,i}-g_t-\epsilon^s_{t,i}\right)$, which, combined with 3.2, leads us to

$$\mathcal{L}\left(\epsilon_{t,i}|M_\Delta,c_\Delta\right) = \int_{-\infty}^{\infty}\mathcal{P}\left(g_{t,i}\left(M_\Delta,c_\Delta\right)\right)\frac{1}{\sqrt{2\pi}\sigma_{\epsilon^s}}e^{-\left(\epsilon_{t,i}-g_{t,i}\left(M_\Delta,c_\Delta\right)\right)^2/2\sigma^2_{\epsilon^s}}. \tag{3.12}$$

Now, if we were to deal with photometric redshift, we would need to calculate the reduced shear g_t for each galaxy in order to find the distribution $\mathcal{P}\left(g_t\left(M_\Delta,c_\Delta\right)\right)$ and then find how the probability $\mathcal{P}_i(z_p|z)$ affects it. We have not done this of yet, and as so will be only dealing with spectroscopic redshift.

In this case, we can compute each $g_{t,i}$ and assign to it a Gaussian kernel with bandwidth h. By summing over all the kernels, we can find a reconstruction for $\mathcal{P}\left(g_t\left(M_\Delta,c_\Delta\right)\right)$. Since many of these data points are very close to each other, we can then "compactify" the KDE to instead use only a few of those points at a time, that is, we combine kernels that are too close in a single kernel.[2] More importantly, since we are using Gaussian kernels, the integral in 3.12 can be solved analytically to find

$$\mathcal{L}\left(\epsilon_{t,i}|M_\Delta,c_\Delta\right) = \frac{1}{N}\sum_j^N\frac{1}{\sqrt{2\pi\left(h^2+\sigma^2_{\epsilon^s}\right)}}e^{-\left(\epsilon_{t,i}-g_{t,j}\left(M_\Delta,c_\delta\right)\right)^2/2\left(h^2+\sigma^2_{\epsilon^s}\right)}, \tag{3.13}$$

that is, the likelihood simply becomes a KDE with data points given by $g_{t,i}$ and bandwidth $\sqrt{h^2+\sigma^2_{\epsilon^s}}$. This likelihood was implemented on NumCosmo.[3] Note that by utilizing all data points in our KDE, we are retaining a lot more information on the data than we would have if we had used the binned method.

TABLE 3.1: Parameters used when creating the mock data set.

Parameter	Value
H_0	$70\,\mathrm{km\,Mpc^{-1}\,s^{-1}}$
$\Omega_{dm,0}$	0.225
$\Omega_{b,0}$	0.045
$\Omega_{k,0}$	0.00
$\Omega_{\Lambda,0}$	0.730
z_{cl}	0.4
M_Δ	$10^{15}\,M_\odot$
c	4

---

[2] See more in `https://numcosmo.github.io/manual/NcmStatsDist1dEPDF.html`

[3] See source code in `https://github.com/NumCosmo/NumCosmo/blob/wl_kde_likelihood/numcosmo/lss/nc_galaxy_wl.c`

3.2 Testing

The first step in our testing procedure is to create a common mock data set with known parameters.[4] For this we use CLMM's `mock.generate_galaxy_catalog` and `GalaxyCluster` methods to create four clusters of mass $10^{15}M_\odot$ (we define our overdensity as $\Delta = 200$ times the reference density ρ_{ref}), concentration $c = 4$ and with 10^5 background galaxies. All parameters utilized are shown on Table 3.1. The difference between each of them is in the redshift distribution of the background galaxies and the standard deviation σ_{ϵ^s}. Both the first and second clusters have fixed redshift of $z = 0.7$, one being spectroscopic and the other photometric with a Gaussian error with standard deviation 0.05, while the third and the fourth both follow the distribution presented on [29], one being spectroscopic and the other also photometric with the same Gaussian error convoluted with the previous distribution. The standard deviation σ_{ϵ^s} took each value in (0.001, 0.01, 0.1, 1).

3.2.1 Source ellipticity distribution

Our first goal on this testing procedure was to understand what are the limitations of modeling the distribution of the source ellipticity ϵ^s as a Gaussian. Note that we are not able to find how faithful this model is: we are using mock data with this distribution baked into the creation process. We can, however, take an initial look into the physical implications of such model. We did this by drawing random values for ϵ_1^s and ϵ_2^s from a Gaussian with $\langle \epsilon_{1,2}^s \rangle = 0$ and $\sigma_{\epsilon^s} \in (0.001, 0.01, 0.1, 1)$, and then calculating the parameters φ and r, respectively the angle between the semi-major axis and ϑ_1 and the ratio between the semi-minor and semi-major axes. The results are shown on Figure 3.3.

The first thing to notice is that the φ distribution is uniform. That's a great sign, since we expect isotropy for the background galaxies ellipticities. Interpreting the results for r is a bit more difficult on the other hand. The only thing we can say for certain is that the standard deviation σ_{ϵ^s} must have an upper limit, since for $\sigma_{\epsilon^s} \approx 1$ the ratio r reaches values below 0, which is not physical: it may only happen for $a = 0$ or $b = \infty$. We then limit our following tests for $\sigma_{\epsilon^s} < 0.1$.

3.2.2 In-bin distribution of reduced shear

As was commented on subsection 3.1.2, a good binned model needs to account for the redshift distribution within each bin. The current implementations on CLMM do this by approximating the resulting reduced shear distribution as Gaussian. To test how well this approximation behaves we took each of the four mock clusters created and binned them from $\mathfrak{r} = 0.07$ to $\mathfrak{r} = 4$, where $\mathfrak{r} = \theta f_K(w_{cl})$ is the radial distance of the galaxies projected on the lens plane. In each case we binned with $N_{bins} = (73, 83, 94, 110, 132, 165, 220)$, which result in bins of width approximately $\Delta\mathfrak{r} = (0.45, 0.40, 0.35, 0.30, 0.25, 0.20, 0.15)$.

---

[4] See code in https://github.com/caiooliVV/masters/tree/main/CLMM

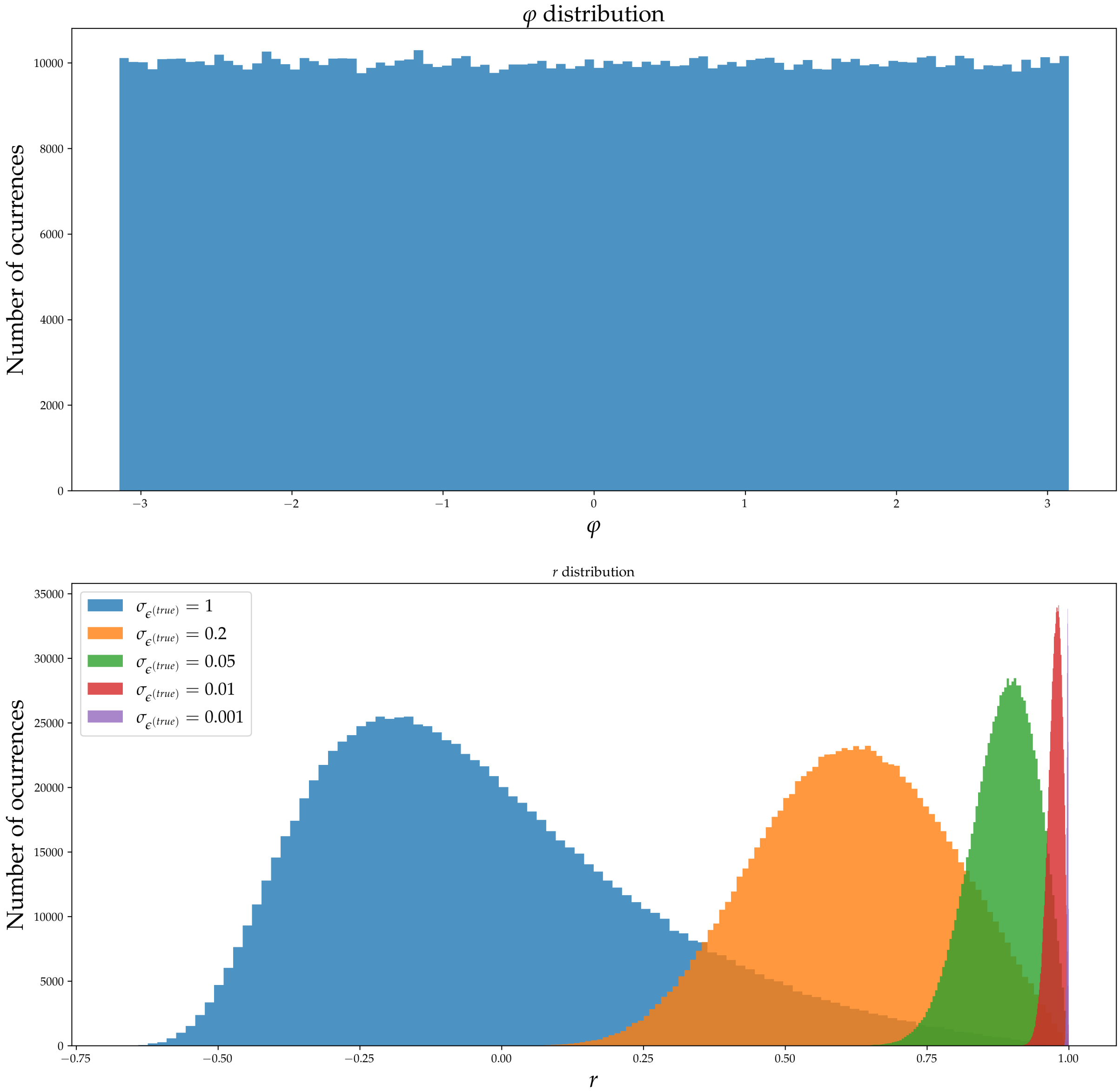


FIGURE 3.3: Analysis of the physical implications of assumptions on the source ellipticity distribution. Here, it is possible to see the uniform resultant distribution on the alignment of the galaxies to the x-axis and the highly skewed distributions for the ratio r depending on various values of σ_{ϵ^s}.

Our tests consisted in calculating through a Shapiro-Wilk [30] test the probability of the data being drawn from a Gaussian distribution. We did this for each redshift distribution, each number of bins N_{bins}, and each bin 10 times (i.e., generating 10 realizations). The results of these tests were plotted in Figures 3.4, 3.5, 3.7, and 3.6, in a way to emphasize the influence of the number of galaxies in the bin N_{gals}, the size of the bin, and the redshift distribution on the reduced shear distribution. Note the interpretation of the p_{value} here: it tells us the frequency in which a Gaussian distribution would produce that data. If we set our rejection limit at 2-σ significance, as we do, we are effectively deciding that samples of data points which cannot be drawn from a Gaussian more than 5% are not well approximated by a Gaussian. Any samples for which $p_{value} \gtrsim 0.05$ we cannot say anything about their normality, though.

Figure 3.4 shows scatter plots of the p_{value} of the test for each of the four redshift distributions. There are two things to notice here. The first one is that the p_{value} decreases with the number of galaxies in the bin N_{gals}. That is to be expected, as the test performs better with more data points to compare. The second is that there is a clear dependency on the redshift distribution, as can be seen for the case of fixed photometric redshift. This is supported by the scatter plot on Figures 3.5 and 3.6 in which samples with different redshift distributions are shown with different colors. It is easy to see on the first figure that the distributions created by a fixed photometric redshift cannot be rejected as normal on almost all cases, while the other redshift distributions can be rejected for most bins with $N_{gals} \gtrsim 80$. The second figure confirms how the redshift distribution affects the g_t distribution by looking at the deviation from the mean p_{value} between different redshifts.

We also tested the effect of the bin size on the p_{value} at each N_{gals}. It can be seen on Figure 3.7, which again shows the deviation from the mean for the different bin sizes. Unlike Figure 3.6, the p_{value} for each bin size seem to be randomly distributed around the mean, which shows that bin size does not seem to affect the p_{value}.

In any way, our results show that for realistic redshift distributions, the normality hypothesis can be rejected with 2-σ for any bin with $N_{gals} \gtrsim 50$. Considering that for any number of bins $N_{bins} \lesssim 200$ we can generally reject normality and that an estimation of optimal number of bins given by Sturge's rule [31] in our case is $N_{bins} \approx 32$, we may conclude that it is not possible to approximate the in-bin distribution of g_t as Gaussian.

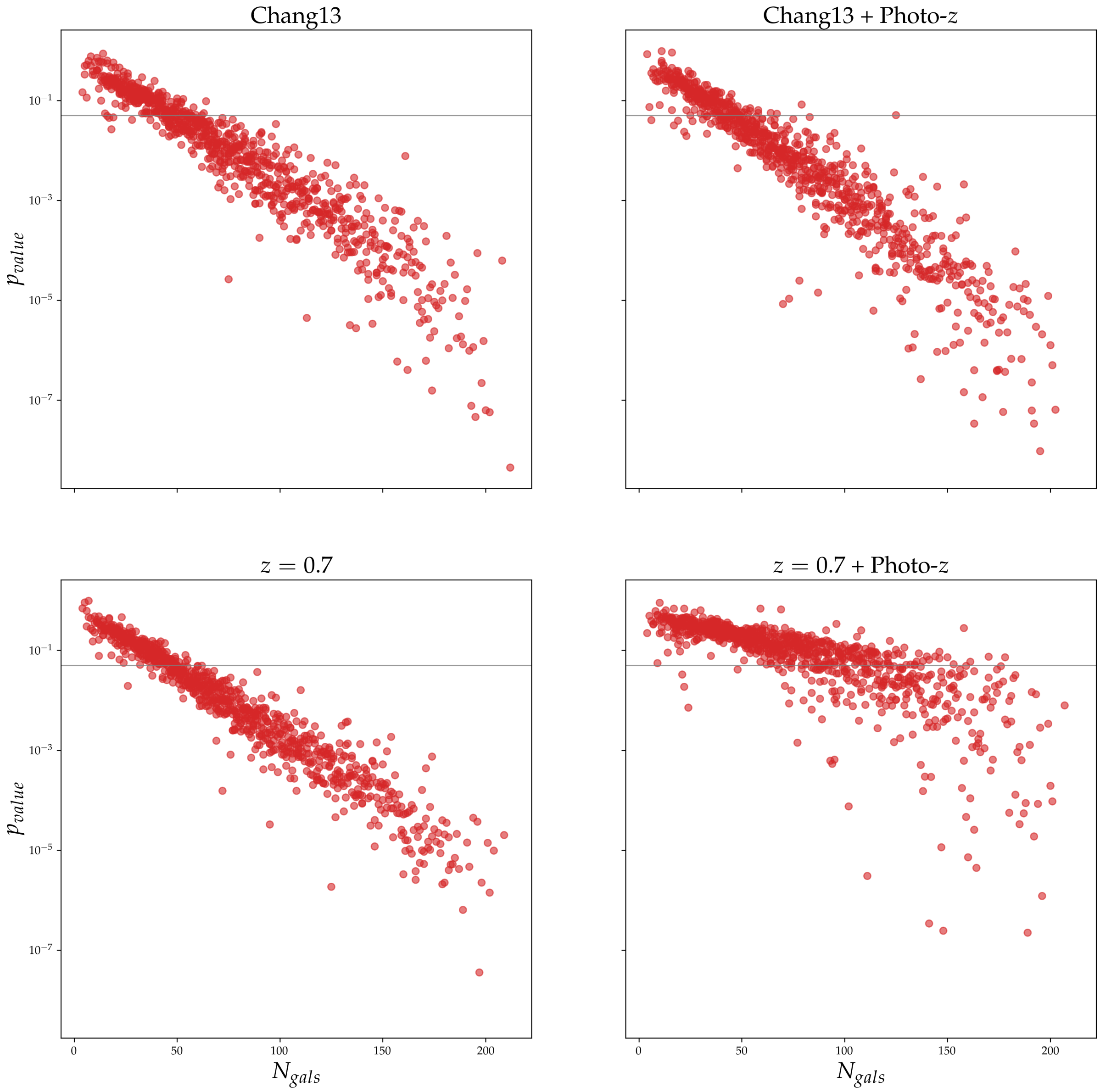


FIGURE 3.4: Scatter plot of Shapiro-Wilk tests. Each plot shows the tests done for each different redshift distribution, denoted on the title of the plot. The x-axis shows the number of galaxies in the bin tested, N_{gals}, while the y-axis shows the p_{value} of the test. The $p_{value} = 0.05$ line is shown in gray.

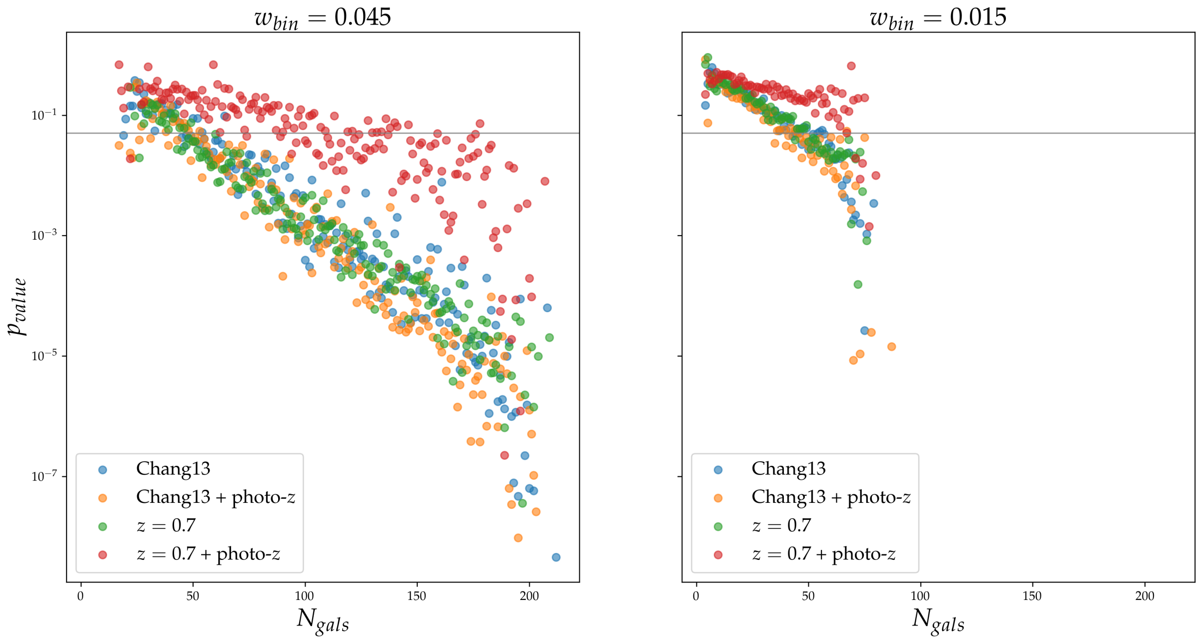


FIGURE 3.5: Scatter plot of Shapiro-Wilk tests. Each different colored point represents a different redshift distribution, while each plot shows the tests done for different bin sizes, denoted on the title of the plot. The x-axis shows the number of galaxies in the bin tested, N_{gals}, while the y-axis shows the p_{value} of the test. The $p_{value} = 0.05$ line is shown in gray.

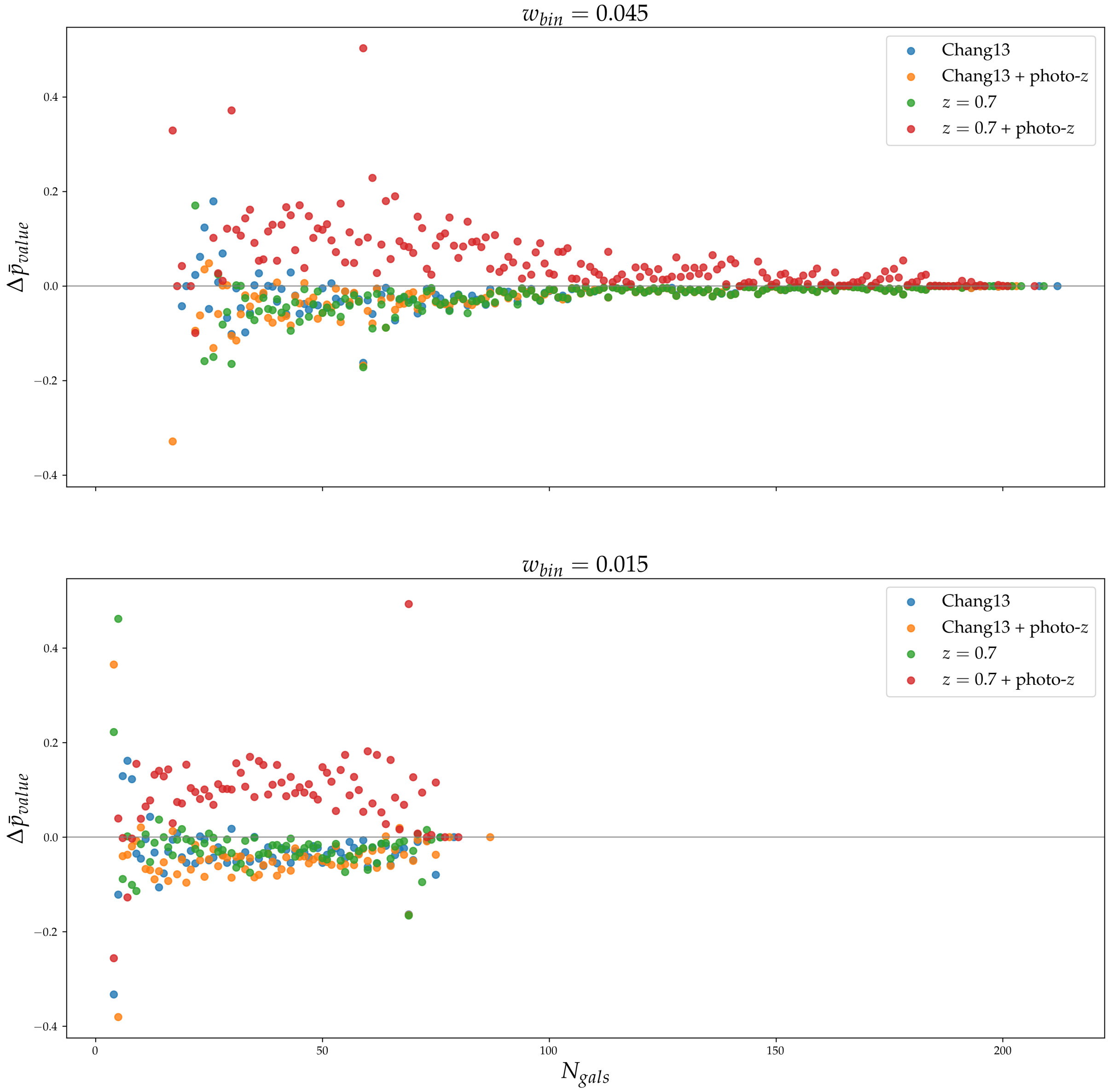


FIGURE 3.6: Scatter plot of deviation from mean p_{value}. Each different colored point represents a redshift distribution, while each plot shows a different bin size. The x-axis shows the number of galaxies in the bin tested, N_{gals}, while the y-axis shows the p_{value} of the test. The $\Delta p_{value} = 0$ line is shown in gray.

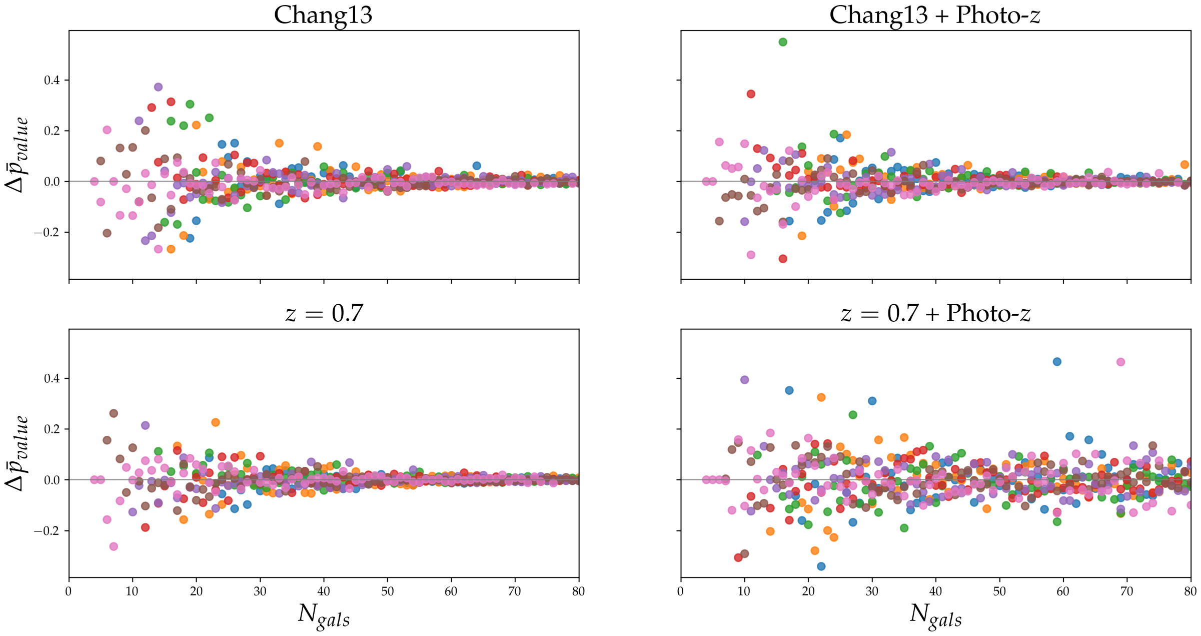


FIGURE 3.7: Scatter plot of deviation from mean p_{value}. Each different colored point represents a different bin size, while each plot shows a different redshift distribution. The x-axis shows the number of galaxies in the bin tested, N_{gals}, while the y-axis shows the p_{value} of the test. The $\Delta p_{value} = 0$ line is shown in gray.

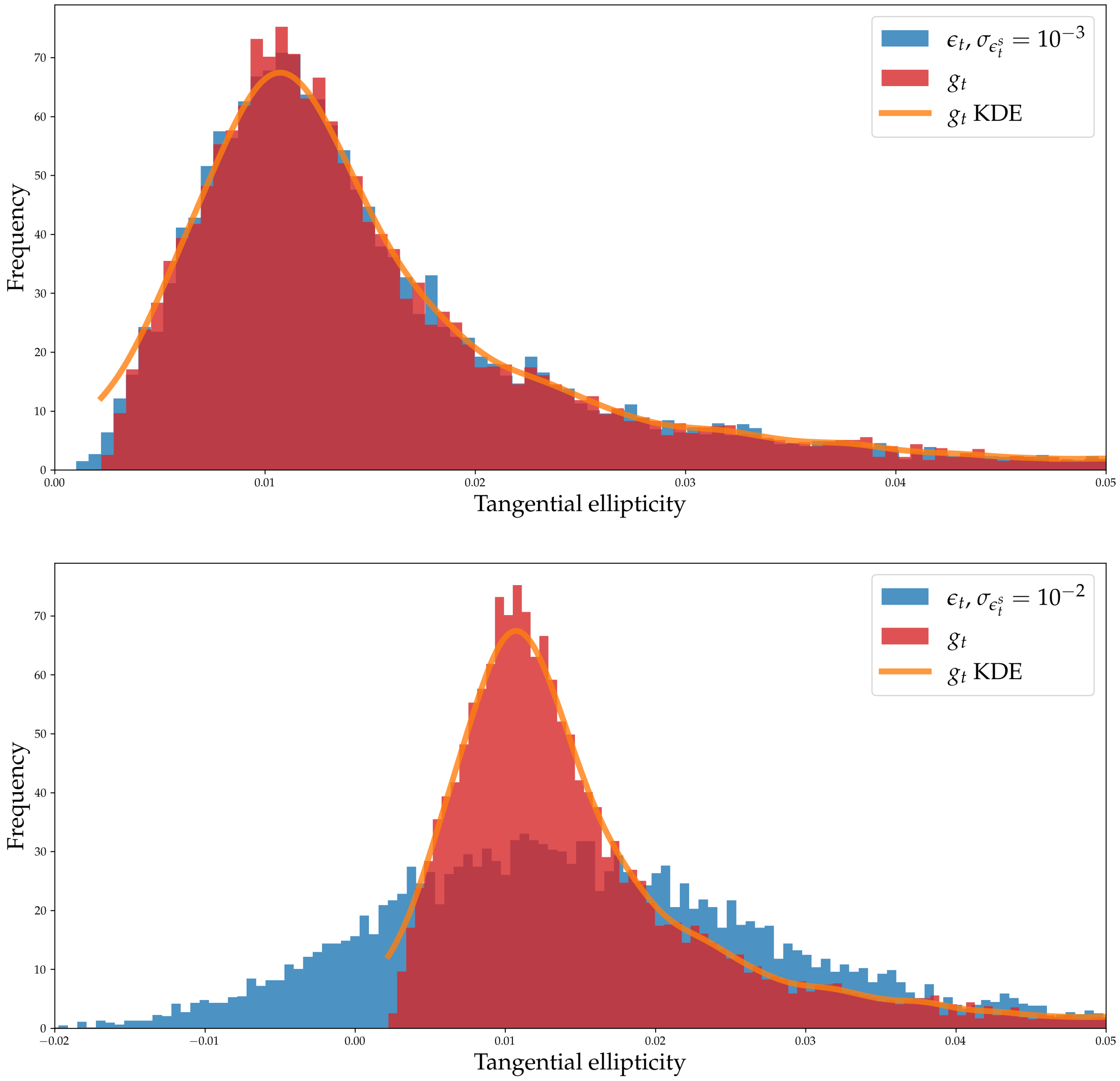


FIGURE 3.8: Both images show histograms for ϵ_t (in blue) and g_t (in red). The KDEs for g_t are also shown. The image on top uses $\sigma_{\epsilon_t^s} = 10^{-3}$, while the bottom one uses $\sigma_{\epsilon_t^s} = 10^{-2}$.

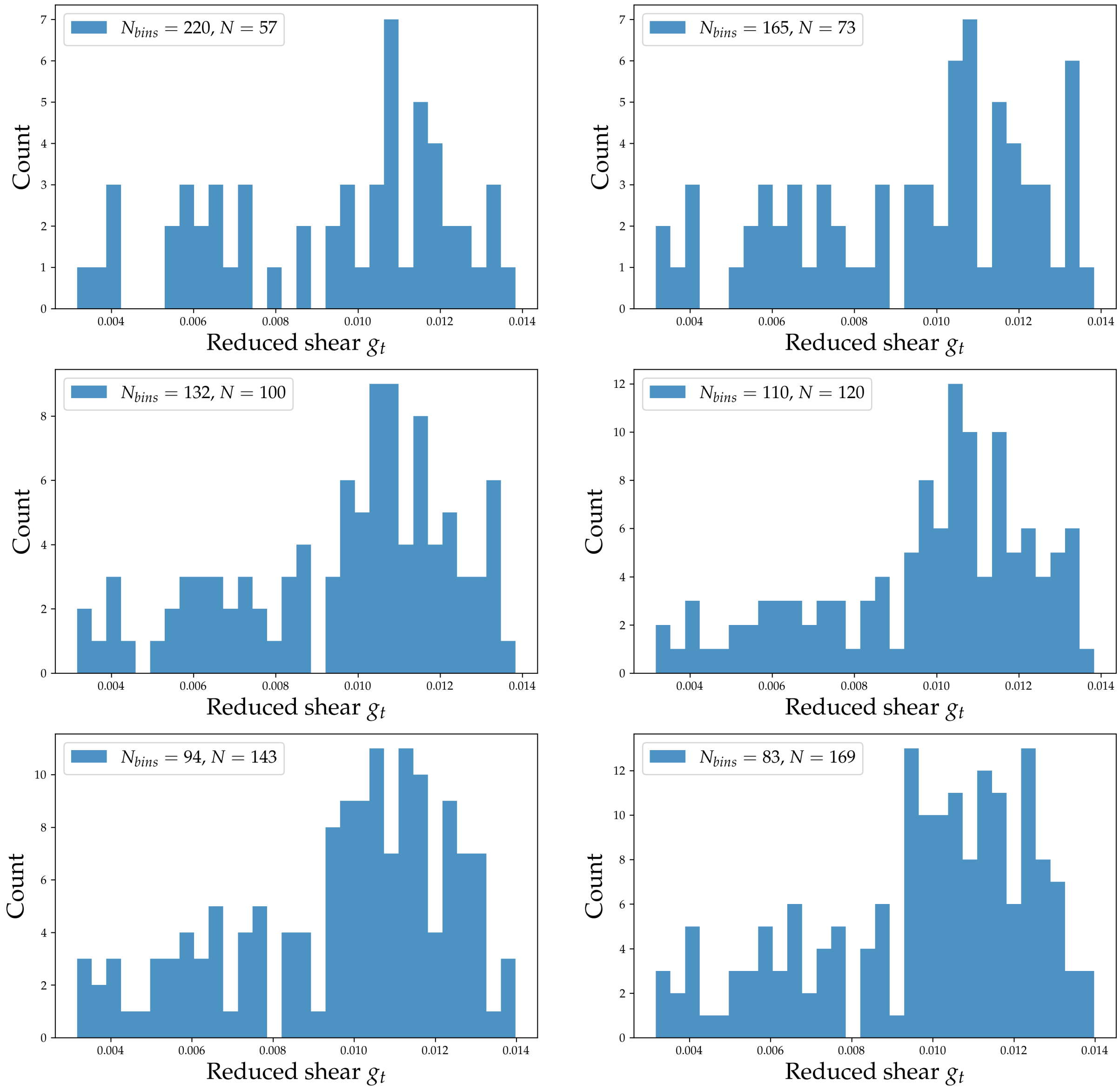


FIGURE 3.9: In-bin histograms of g_t for different N_{bins}. Redshift distribution is Chang13. Each plot shows the histogram of g_t at the last radial bin of galaxies for different N_{bins} and hence different bin sizes.

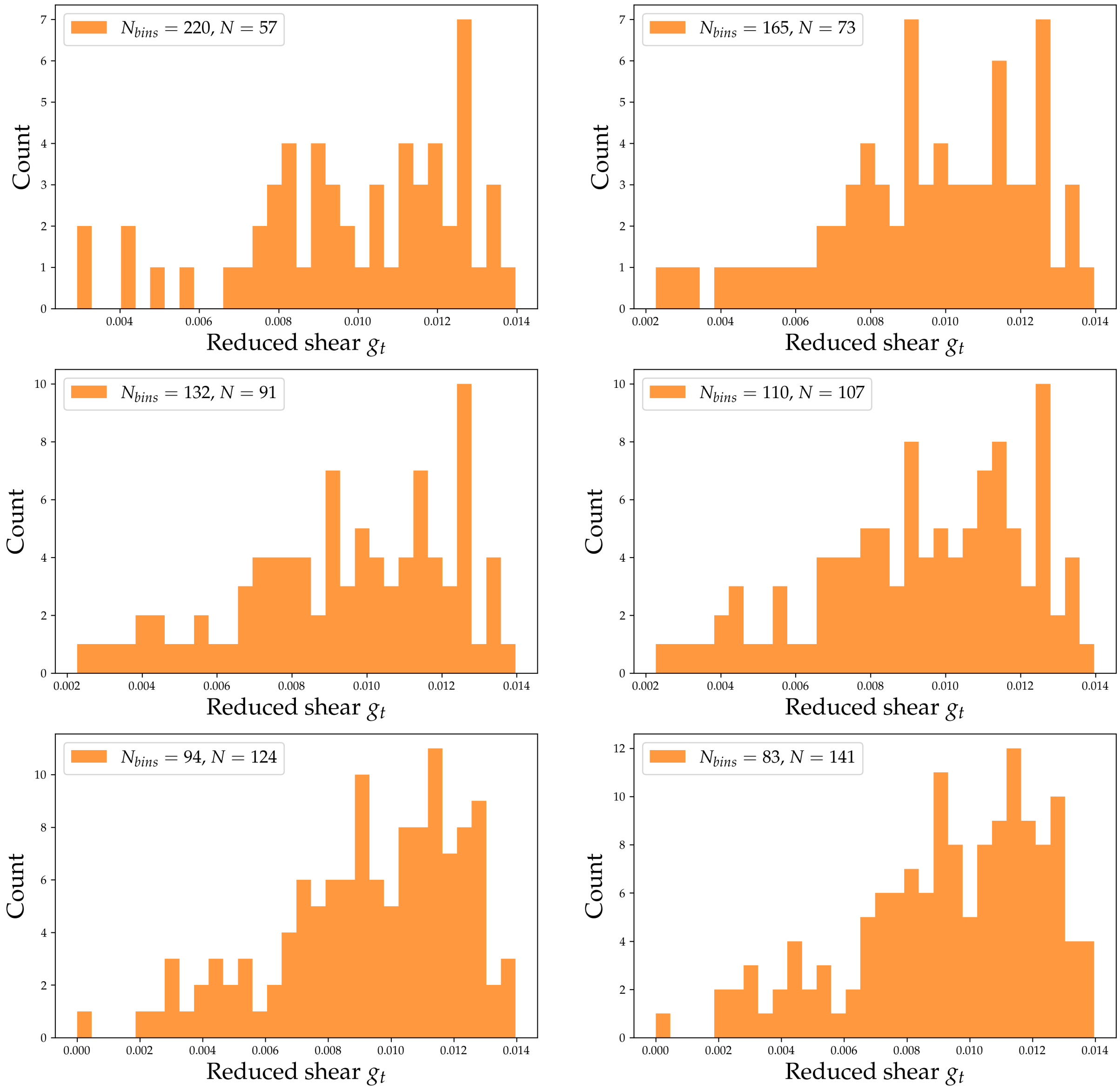


FIGURE 3.10: In-bin histograms of g_t for different N_{bins}. Redshift distribution is Chang13 + photo-z. Each plot shows the histogram of g_t at the last radial bin of galaxies for different N_{bins} and hence different bin sizes.

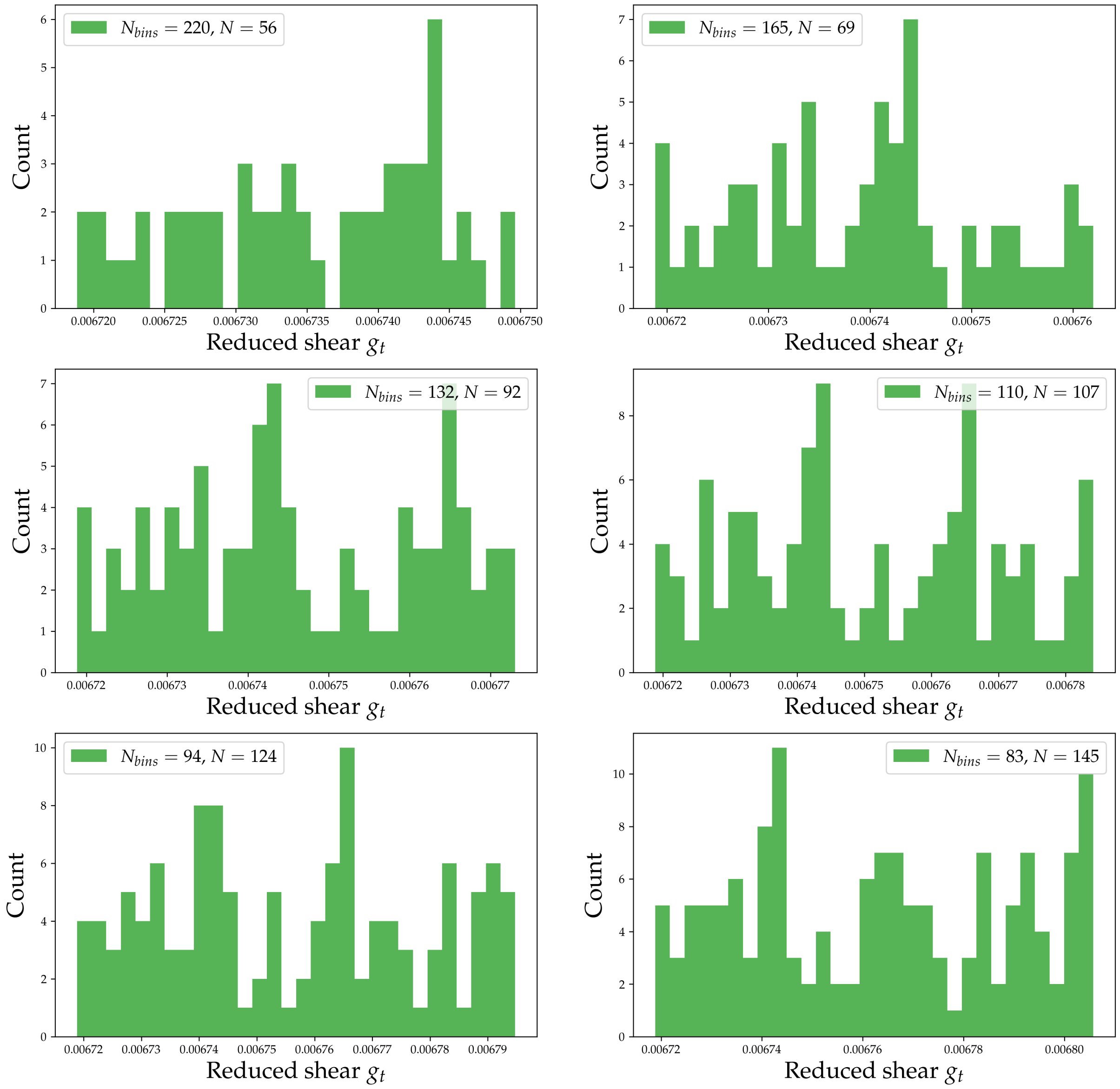


FIGURE 3.11: In-bin histograms of g_t for different N_{bins}. Redshift distribution is fixed $z = 0.7$. Each plot shows the histogram of g_t at the last radial bin of galaxies for different N_{bins} and hence different bin sizes.

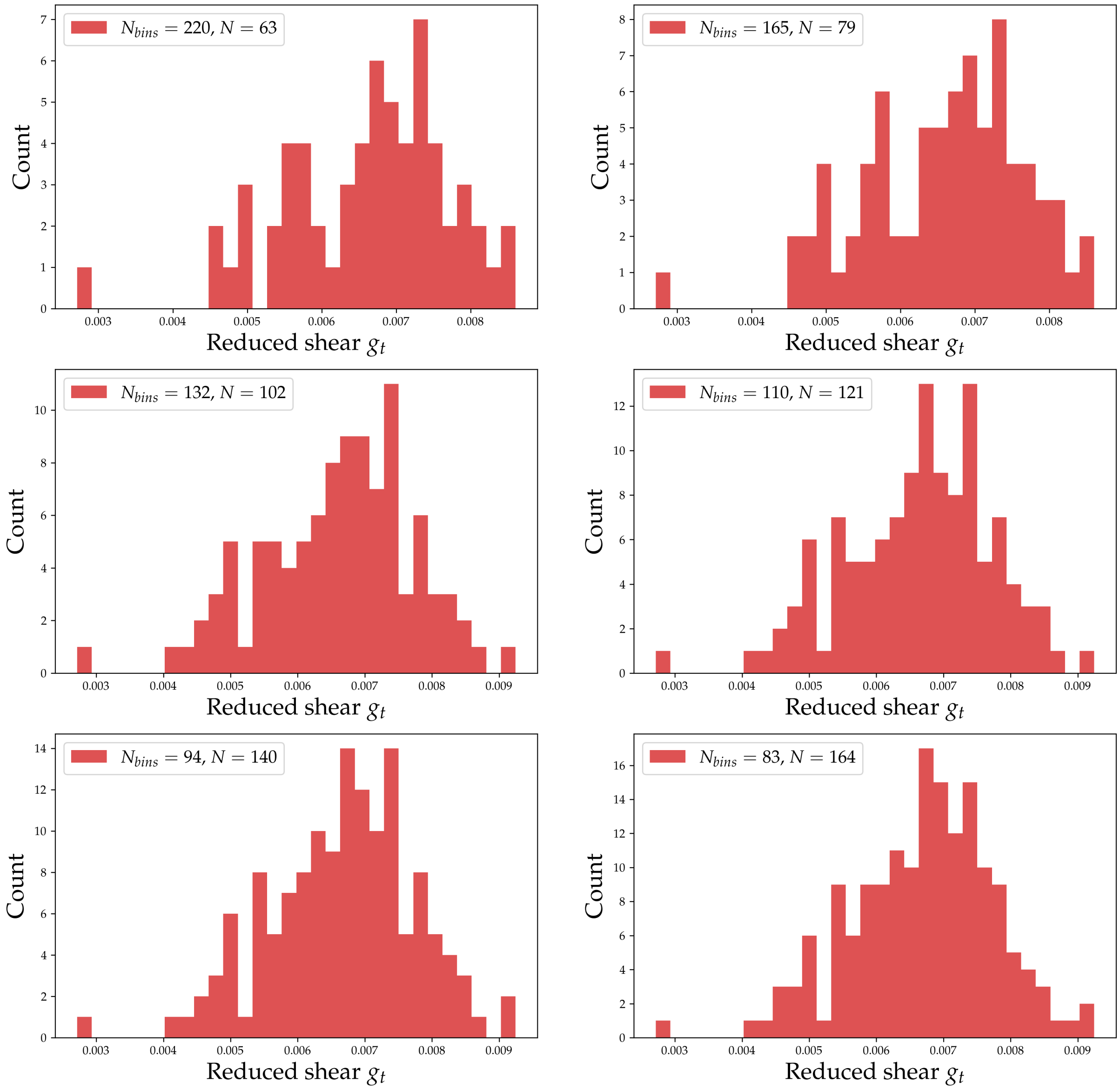


FIGURE 3.12: In-bin histograms of g_t for different N_{bins}. Redshift distribution is fixed $z = 0.7$ + photo-z. Each plot shows the histogram of g_t at the last radial bin of galaxies for different N_{bins} and hence different bin sizes.

3.3 Best fit and MCMC

Finally, we investigate the performance of each model by determining the best fit for M_Δ and c_Δ and finding the posterior distribution sampled by these likelihoods with a MCMC analysis.[5] This is possible for all models implemented through NumCosmo's framework. The only exception is the binned one, which was implemented independently. In that case we will use SciPy's `optimize.curve_fit` to find the best fit and error estimation only for M_Δ.

These results were taken by utilizing the fiducial model defined in the previous section with spectroscopic redshift distribution given by [29] and source ellipticity standard deviation σ_{ϵ^s} taking the values 0.2, 0.05, and 0.01.

3.3.1 Best fit

The results for the best fit of the mass M_Δ and the concentration c_Δ for each of the models are shown on Tables 3.2 and 3.3. There, the first column shows the name of the method used when building the likelihood, and each subsequent column shows the best fit and error for different values of σ_{ϵ^s}.

TABLE 3.2: Results for the best fit and error of $\log_{10} M_\Delta$ for each likelihood model. Columns 2 through 4 show the results for different values of σ_{ϵ^s}.

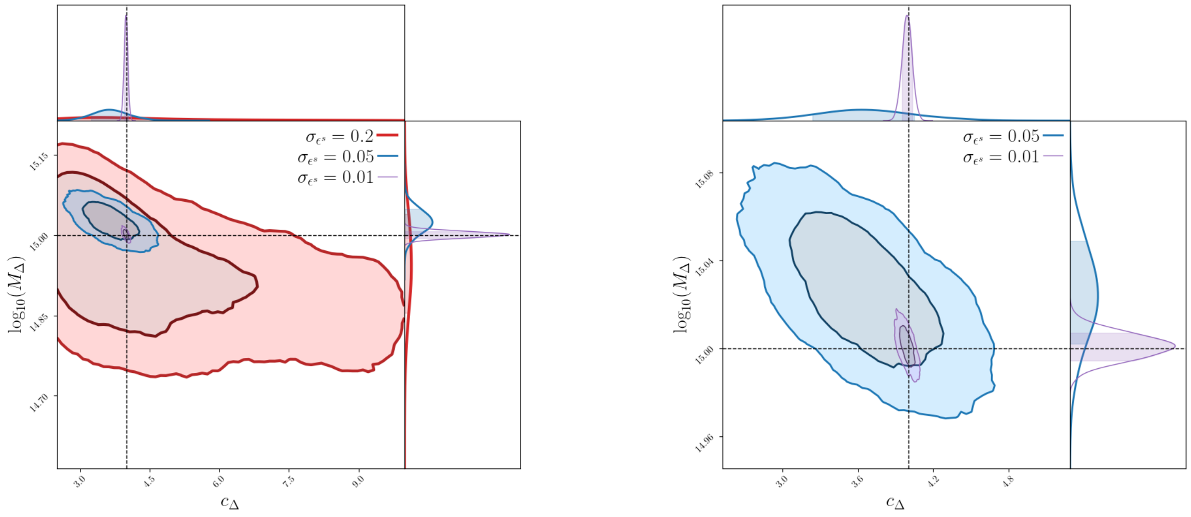
	$\log_{10} M_\Delta$		
Models	$\sigma_{\epsilon^s} = 0.2$	$\sigma_{\epsilon^s} = 0.05$	$\sigma_{\epsilon^s} = 0.01$
Unbinned	14.980 ± 0.073	15.020 ± 0.019	15.000 ± 0.004
Binned	14.865 ± 1.266	14.969 ± 0.277	14.995 ± 0.066
KDE	15.020 ± 0.145	15.030 ± 0.022	15.000 ± 0.004

TABLE 3.3: Results for the best fit and error of c_Δ for each likelihood model. Columns 2 through 4 show the results for different values of σ_{ϵ^s}.

	c_Δ		
Models	$\sigma_{\epsilon^s} = 0.2$	$\sigma_{\epsilon^s} = 0.05$	$\sigma_{\epsilon^s} = 0.01$
Unbinned	4.21 ± 0.43	3.97 ± 0.20	4.00 ± 0.03
KDE	2.50 ± 3.35	3.66 ± 0.36	3.99 ± 0.03

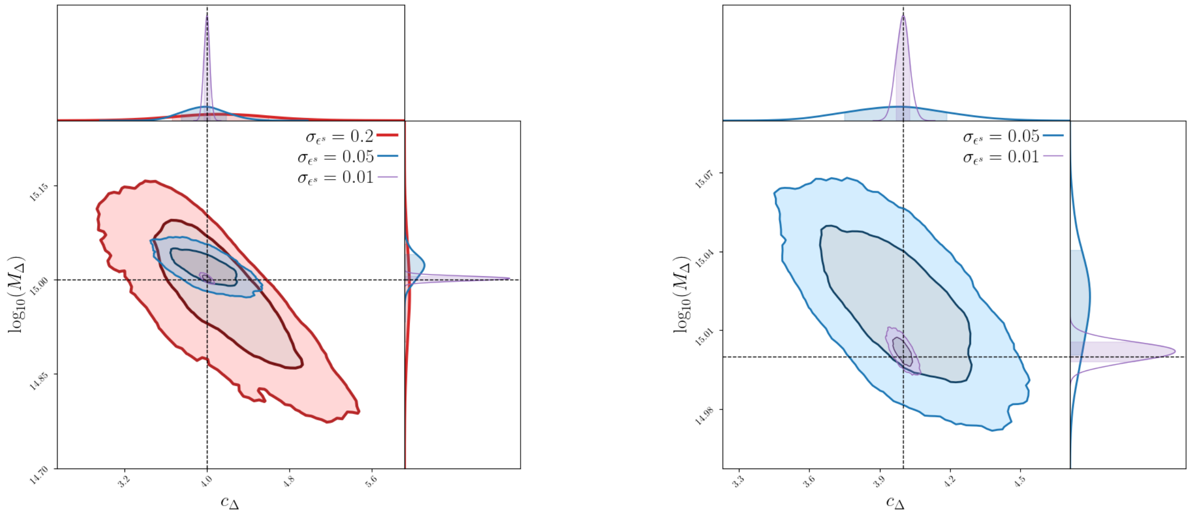
---

[5] See code in `https://github.com/caioolivv/masters/tree/main/CLMM`



(A) Posterior distribution for KDE likelihood sampled through a MCMC process. The distributions for all values of σ_{ϵ^s} are shown.

(B) Posterior distribution for KDE likelihood sampled through a MCMC process. The distributions for σ_{ϵ^s} equal only to 0.05 and 0.01 are shown.



(C) Posterior distribution for unbinned likelihood sampled through a MCMC process. The distributions for all values of σ_{ϵ^s} are shown.

(D) Posterior distribution for unbinned likelihood sampled through a MCMC process. The distributions for σ_{ϵ^s} equal only to 0.05 and 0.01 are shown.

FIGURE 3.13: Posterior distributions sampled for both the unbinned and KDE likelihood through a MCMC process. The y-axis shows $\log_{10} M_\Delta$ and the x-axis shows c_Δ. The corner plot displays the posterior distribution, while the top and side plots present the c_Δ and $\log_{10} M_\Delta$ distributions alone.

Let us start by analyzing the results for M_Δ. Note that, since our mock sample has spectroscopic redshift, the unbinned model has a huge advantage: it is both easy to compute (as we don't have to make any integral) and doesn't summarize the data. It is no surprise then that its fit is both the most accurate and precise. The slightly worse performance for the binned method is also to be expected, as we are summarizing the data and taking a Gaussian approximation that we have shown to be not nearly precise enough for this study. The interesting result here however is how well the KDE model performs in relation to the unbinned one. Even though we are not taking all the data into account our fit is as good as the unbinned, with a slightly bigger error bar. Of course, for spectroscopic redshift there is no reason to use any other method but the unbinned, but this result shows us there is merit in studying an equivalent approach for photometric redshift.

The results are a bit different for c_Δ though. It seems the concentration is specially sensitive to noise, as the unbinned likelihood presents a lower precision as it displayed for M_Δ. The problem is even worse for the KDE method, as the error estimation on the fit is at the same order of magnitude as the fit itself.

3.3.2 Posterior

The posterior distributions sampled are shown in Figure 3.13. These plots confirm our previous findings. The KDE model follows closely the unbinned likelihood results, as can be seen by how similar their posteriors are. They are both extremely precise, with the 2-σ region on low σ_{ϵ^s} on the order of 10^{-2} and 10^{-1} for M_Δ and c_Δ. Still, the 2-σ region of the concentration c_δ for high values of σ_{ϵ^s} covers an entire order of magnitude, which shows how sensitive this parameter is to the noise.

Chapter 4

Conclusion

On the beginning of this work, we discussed the implications and complications of modeling the intrinsic ellipticities of galaxies with normal distributions. The physical consequences of this model are not easily seen due to the fact that the ellipticities are complicated functions of the semi-major and semi-minor axes a and b. If we are to use this model, we must understand well the distribution of the ratio r of the axes, and find which are the acceptable values for its mean and standard deviation, specially because r is bounded both on the left and on the right ($r \in [0, 1]$).

There are ways in which we can study the nature of this ellipticity distribution (see for example [32]). There is an alternative to this approach, though, that we mentioned on Chapter 2. Instead of working with the complex ellipticity ϵ, we may study directly the quadrupole moment of surface brightness, from which ϵ is calculated. Working with this quantity may be advantageous, considering this is the real observable, and we intend to investigate this approach in our way forward.

Another important result regards binning methods for likelihood building. We've shown to 2-σ confidence that the shear data in each bin cannot be drawn from a Gaussian distribution, and as such it cannot be well summarized by an average and a standard deviation. Doing this leads to biased estimates for the mass, which we verified with the best fit in Table 3.2. It is our intention to share this work with other DESC researchers in order to implement better models on the libraries being developed at the Cluster and Weak Lensing working groups.

On that note, our proposed kernel density estimation model for the likelihood showed promising results. By producing mass estimates on par with the ones from the unbinned likelihood, which does not approximate or summarize the data, it has been shown to be a good alternative for binning methods. Of course, there are still two problems which we must deal with. The first is the fact that its estimate for the concentration is extremely sensitive to the noise, resulting in worse estimates than the binned likelihood as it has been implemented, which we know to be biased. The second complication is the lack of support for photometric redshifts. If it is to be used in LSST, and that is our intention, we must seriously work on these issues.

Finally, our future works have to address these topics. We hope to soon start reimplementing our current tools to both be able to better represent the underlying light distribution of background galaxies through the quadrupole moment Q_{ij}, and to take into account the photometric redshift distribution. These tools, of course, are to be readily available to the DESC community we are working with.

Appendix A

Statistics

Here we present a brief introduction of the statistical concepts and tools used in this work. The discussions here are loosely based on [33, 34].

A.1 Bayes' Theorem

Bayes' theorem is a simple and easily accepted statement about the conditional probabilities of two statistical variables. Let us look at the simplest example, in which we have two events A and B which can happen independently and in tandem. The probability of both A and B happening together, $\mathcal{P}(A \cap B)$, can be written as

$$\mathcal{P}(A \cap B) = \mathcal{P}(A|B)\mathcal{P}(B) = \mathcal{P}(B|A)\mathcal{P}(A), \tag{A.1}$$

where $\mathcal{P}(A|B)$ and $\mathcal{P}(B|A)$ are the probability of A happening given B has happened and vice versa. The previous equation can be rearranged to give us

$$\mathcal{P}(B|A) = \frac{\mathcal{P}(A|B)\mathcal{P}(B)}{\mathcal{P}(A)}. \tag{A.2}$$

This is the equations known as Bayes' theorem. Besides giving us a way to calculate the probabilities $\mathcal{P}(A|B)$ and $\mathcal{P}(B|A)$ from each other, it can be interpreted in a very particular way, as we will see later. The probability $\mathcal{P}(A|B)$ is usually called the likelihood of A given B, while $\mathcal{P}(B)$ and $\mathcal{P}(B|A)$ are respectively the prior and posterior probabilities.

Bayes' theorem can also be written for continuous statistical variables X and Y. In this case it becomes

$$\mathfrak{p}_{X|Y}(x) = \frac{\mathfrak{p}_{Y|X}(y)\mathfrak{p}_X(x)}{\mathfrak{p}_Y(y)}, \tag{A.3}$$

where $\mathfrak{p}_X(x)$ is the probability density function of X, and so on for the other terms. Since these p.d.f's are normalized, we can integrate on both sides to write

$$\mathfrak{p}_Y(y) = \int \mathrm{d}x \mathfrak{p}_{Y|X}(y)\mathfrak{p}_X(x). \tag{A.4}$$

A.2 Likelihood and Maximum Likelihood Estimators

Consider now the case in which we have a statistical variable x, dependent on a set of parameters θ, with realizations X. The likelihood $\mathcal{L}(X|\theta)$ is the probability of observing the realizations X given the parameters θ. We can use this likelihood function to estimate the parameters θ through a maximum likelihood estimator (MLE) of θ. The idea is that a good estimate of the parameters θ is the one that maximizes the likelihood of measuring the data X. This MLE is then defined as

$$\hat{\theta} = \arg\max{}_{\theta}\mathcal{L}(X|\theta). \tag{A.5}$$

These estimators have desirable properties at the limit in which the number of samples goes to infinity, such as consistency and efficiency. That is, as the number of samples N grows, these estimators generally approach the real value, meaning they are consistent, and have their variance approach the minimum variance bound, becoming efficient estimators.

A.3 Fisher Information and Covariance Matrix

A function closely related to the maximum likelihood estimator is the score, the partial derivative of the logarithm of the likelihood with respect to the parameter θ:

$$S = \frac{\partial}{\partial\theta} \ln \mathcal{L}(X|\theta). \tag{A.6}$$

Of course, calculated at the true value θ_0, the expected value of the score is 0. The Fisher information is defined as the variance of the score

$$\mathcal{I}(\theta) = \left\langle \left(\frac{\partial}{\partial\theta} \ln \mathcal{L}(X|\theta) \right)^2 \right\rangle, \tag{A.7}$$

which, if $\mathcal{L}(X|\theta)$ is twice differentiable with respect to θ, can be written as

$$\mathcal{I}(\theta) = -\left\langle \frac{\partial^2}{\partial\theta^2} \ln \mathcal{L}(X|\theta) \right\rangle. \tag{A.8}$$

For n parameters θ_i, we define the Fisher information matrix

$$\mathcal{I}\left(\theta\right)_{ij} = \left\langle \frac{\partial^2}{\partial\theta_i \partial\theta_j} \ln \mathcal{L}(X|\theta) \right\rangle. \tag{A.9}$$

If the data set is sufficiently large, the Fisher information matrix becomes related to the covariance matrix by [35]

$$C_{ij} = -\left(\frac{\partial^2}{\partial\theta^i\theta^j}\ln\mathcal{L}(X|\theta)\right)^{-1}, \tag{A.10}$$

from which we can estimate the confidence intervals for our maximum likelihood estimates.

A.4 MCMC

A particularly important interpretation of Bayes' theorem is that our knowledge of B can be updated through the likelihood $\mathcal{P}(A|B)$, by simply multiplying our prior knowledge $\mathcal{P}(B)$ to find the posterior knowledge $\mathcal{P}(B|A)$. We can repeat this process in steps by utilizing the posterior $\mathcal{P}(B|A)$ of the last step as the prior of the next. In this way, we increase our knowledge of B given A. Again, in the case in which we have a statistical variable x dependent on a set of parameters θ, we can use this process to find probable values of θ given our realizations X.

One such way of doing this is with Markov chain Monte Carlo methods. By generating a set of random samples p that have a certain probability $\mathcal{P}(p'|p)$ related to the likelihood $\mathcal{L}(x|\theta)$ of generating a new sample p', the MCMC process can converge to a sample of the posterior $\mathcal{P}(B|A)$.

A.5 Kernel Density Estimation

Kernel density estimation is a method of estimation of probability density functions [36]. It consists on assigning each sample of the data a kernel function K, that are to be summed over to form the estimation

$$\hat{p}(x) = \frac{1}{N}\sum_i^N K(x_i, x). \tag{A.11}$$

This kernel function is not defined a prior, and may be chosen dependent on the application. In this work we use a Gaussian kernel, leading us to write the KDE as

$$\hat{p}(x) = \frac{1}{N}\sum_i^N \frac{1}{\sqrt{2\pi}h} e^{\frac{-(x_i-x)^2}{2h^2}}, \tag{A.12}$$

where h is the smoothing parameter or bandwidth. As implied by its name, it controls how smooth the KDE is, and there are many possible ways to assign its value. We chose to use Silverman's rule of thumb [37], defined as

$$h = 0.9 \cdot min\left(\hat{\sigma}, \frac{IQR}{1.34}\right) \cdot N^{-1/5}, \tag{A.13}$$

where $\hat{\sigma}$ is the sample standard deviation and IQR is the interquartile range.

Bibliography

[1] M. Aguena et al. "CLMM: A LSST-DESC cluster weak lensing mass modeling library for cosmology". In: *Monthly Notices of the Royal Astronomical Society* 508 (4 July 2021), pp. 6092–6110. ISSN: 13652966. DOI: `10.1093/mnras/stab2764`. URL: `https://dx.doi.org/10.1093/mnras/stab2764`.

[2] Sandro D. P. Vitenti, M. Penna-Lima, and Cyrille. Doux. *Numerical Cosmology Library*. 2017. URL: `https://numcosmo.github.io/`.

[3] *LSST-DESC/firecrown: DESC Cosmology Likelihood Framework*. URL: `https://github.com/LSSTDESC/firecrown`.

[4] LSST Dark Energy Science Collaboration. "Large Synoptic Survey Telescope: Dark Energy Science Collaboration". In: 34 (Nov. 2012). DOI: `10.48550/arxiv.1211.0310`. URL: `https://arxiv.org/abs/1211.0310`.

[5] B. C. Quint. *Astrophotography at the Summit Facility*. URL: `https://gallery.lsst.org/bp/#/folder/10165313/127028907`.

[6] J. E. Carlstrom, G. P. Holder, and E. D. Reese. "Cosmology with the Sunyaev-Zel'dovich Effect". In: *Annual Review of Astronomy and Astrophysics* 40.1 (Sept. 2002), pp. 643–680. DOI: `10.1146/annurev.astro.40.060401.093803`. URL: `https://doi.org/10.1146/annurev.astro.40.060401.093803`.

[7] N. Huang et al. "Galaxy Clusters Selected via the Sunyaev-Zel'dovich Effect in the SPTpol 100-square-degree Survey". In: *The Astronomical Journal* 159.3 (Feb. 2020), p. 110. DOI: `10.3847/1538-3881/ab6a96`. URL: `https://doi.org/10.3847/1538-3881/ab6a96`.

[8] P. M. Motl et al. "The Integrated Sunyaev-Zeldovich Effect as a Superior Method for Measuring the Mass of Clusters of Galaxies". In: *The Astrophysical Journal* 623.2 (Mar. 2005), pp. L63–L66. DOI: `10.1086/430144`. URL: `https://doi.org/10.1086/430144`.

[9] E. S. Rykoff et al. "Robust Optical Richness Estimation with Reduced Scatter". In: *The Astrophysical Journal* 746.2 (Feb. 2012), p. 178. DOI: `10.1088/0004-637x/746/2/178`. URL: `https://doi.org/10.1088/0004-637x/746/2/178`.

[10] E. S. Rykoff et al. "redMaPPer I: Algorithm and SDSS DR8 Catalog". In: *The Astrophysical Journal* 785.2 (Apr. 2014), p. 104. DOI: `10.1088/0004-637x/785/2/104`. URL: `https://doi.org/10.1088/0004-637x/785/2/104`.

[11] *NASA's Webb Produces the Most Detailed Image of the Early Universe to Date*. URL: `https://webbtelescope.org/contents/news-releases/2022/news-2022-038` (visited on 09/22/2022).

[12] R. Pello et al. “Observing z>7 sources with the GTC”. In: *II International Workshop on Science with the GTC* (Apr. 2004). DOI: 10.48550/arxiv.astro-ph/0404131. URL: https://arxiv.org/abs/astro-ph/0404131v1.

[13] J. F. Le Borgne et al. “Spectroscopic follow up of arclets in AC114 with the VLT”. In: (2000). DOI: 10.48550/ARXIV.ASTRO-PH/0007154. URL: https://arxiv.org/abs/astro-ph/0007154.

[14] D. Clowe and P. Schneider. “Wide field weak lensing observations of A1689”. In: *Astronomy & Astrophysics* 379 (2 Nov. 2001), pp. 384–392. ISSN: 0004-6361. DOI: 10.1051/0004-6361:20011310. URL: https://www.aanda.org/miscs/aa/abs/2001/44/aah2933/aah2933.html.

[15] A. N. Taylor et al. “Gravitational Lens Magnification and the Mass of Abell 1689”. In: *The Astrophysical Journal* 501.2 (July 1998), pp. 539–553. DOI: 10.1086/305827. URL: https://doi.org/10.1086/305827.

[16] M. E. Gray et al. “Probing the Distribution of Dark Matter in the A901/902 Supercluster with Weak Lensing”. In: *The Astrophysical Journal* 568.1 (Mar. 2002), pp. 141–162. DOI: 10.1086/338763. URL: https://doi.org/10.1086/338763.

[17] G. Soucail, J.-P. Kneib, and G. Golse. “Multiple-images in the cluster lens Abell 2218: Constraining the geometry of the Universe?” In: *Astronomy & Astrophysics* 417.3 (Mar. 2004), pp. L33–L37. DOI: 10.1051/0004-6361:20040077. URL: https://doi.org/10.1051/0004-6361:20040077.

[18] L. Hernquist. “An Analytical Model for Spherical Galaxies and Bulges”. In: *The Astrophysical Journal* 356 (1990), pp. 359–364.

[19] J. Einasto. “On the Construction of a Composite Model for the Galaxy and on the Determination of the System of Galactic Parameters”. In: *Trudy Astrofizicheskogo Instituta Alma-Ata* 5 (Jan. 1965), pp. 87–100.

[20] B. Diemer and A. V. Kravtsov. “Dependence of the outer density profiles of halos on their mass accretion rate”. In: *The Astrophysical Journal* 789 (2014).

[21] J. F. Navarro, C. S. Frenk, and S. D. M. White. “The Structure of Cold Dark Matter Halos”. In: *The Astrophysical Journal* 462 (May 1996), p. 563. DOI: 10.1086/177173. URL: https://doi.org/10.1086/177173.

[22] J. F. Navarro, C. S. Frenk, and S. D. M. White. “A Universal Density Profile from Hierarchical Clustering”. In: *The Astrophysical Journal* 490.2 (Dec. 1997), pp. 493–508. DOI: 10.1086/304888. URL: https://doi.org/10.1086/304888.

[23] K. Umetsu et al. “CLASH: Weak-lensing Shear and Magnification Analysis of 20 Galaxy Clusters”. In: *The Astrophysical Journal* 795.2 (Oct. 2014), p. 163. DOI: 10.1088/0004-637x/795/2/163. URL: https://doi.org/10.1088/0004-637x/795/2/163.

[24] M. Bartelmann. “Gravitational lensing”. In: *Classical and Quantum Gravity* 27.23 (Nov. 2010), p. 233001. DOI: 10.1088/0264-9381/27/23/233001. URL: https://doi.org/10.1088/0264-9381/27/23/233001.

[25] M. Bartelmann and M. Maturi. “Weak gravitational lensing”. In: (2016). DOI: 10.48550/ARXIV.1612.06535. URL: https://arxiv.org/abs/1612.06535.

[26] P. Schneider, C. S. Kochanek, and J. Wambsganss. *Gravitational Lensing: Strong, Weak and Micro*. Vol. 33. Springer Berlin Heidelberg, 2006. ISBN: 978-3-540-30309-1. DOI: `10.1007/978-3-540-30310-7`. URL: `https://link.springer.com/10.1007/978-3-540-30310-7`.

[27] R. Narayan and M. Bartelmann. "Lectures on Gravitational Lensing". In: (1996). DOI: `10.48550/ARXIV.ASTRO-PH/9606001`. URL: `https://arxiv.org/abs/astro-ph/9606001`.

[28] S. Weinberg. *Cosmology*. Cambridge University Press, 2008, p. 616. ISBN: 9780198526827.

[29] C. Chang et al. "The effective number density of galaxies for weak lensing measurements in the LSST project". In: *Monthly Notices of the Royal Astronomical Society* 434.3 (July 2013), pp. 2121–2135. ISSN: 1365-2966. DOI: `10.1093/mnras/stt1156`. URL: `https://dx.doi.org/10.1093/mnras/stt1156`.

[30] S. S. Shapiro and M. B. Wilk. "An analysis of variance test for normality (complete samples)". In: *Biometrika* 52 (3-4 Dec. 1965), pp. 591–611. ISSN: 0006-3444. DOI: `10.1093/BIOMET/52.3-4.591`. URL: `https://academic.oup.com/biomet/article/52/3-4/591/336553`.

[31] H. A. Sturges. "The Choice of a Class Interval". In: *https://doi.org/10.1080/01621459.1926.10502161* 21 (153 2012), pp. 65–66. ISSN: 1537274X. DOI: `10.1080/01621459.1926.10502161`. URL: `https://www.tandfonline.com/doi/abs/10.1080/01621459.1926.10502161`.

[32] D. E. Applegate et al. "Weighing the Giants – III. Methods and measurements of accurate galaxy cluster weak-lensing masses". In: *Monthly Notices of the Royal Astronomical Society* 439.1 (Feb. 2014), 48–72. ISSN: 0035-8711. DOI: `10.1093/mnras/stt2129`. URL: `https://dx.doi.org/10.1093/mnras/stt2129`.

[33] R. Lupton. *Statistics in theory and practice*. Princeton University Press, 1993. ISBN: 9780691213194. URL: `https://press.princeton.edu/books/ebook/9780691213194/statistics-in-theory-and-practice`.

[34] R. Barlow. *Statistics : a guide to the use of statistical methods in the physical sciences*. Wiley, 1989, p. 204. ISBN: 9780471922957.

[35] Herman J. Bierens. "Introduction to the mathematical and statistical foundations of econometrics". In: (2005), p. 323.

[36] S. Weglarczyk. "Kernel density estimation and its application". In: *ITM Web of Conferences* 23 (2018), p. 00037. DOI: `10.1051/itmconf/20182300037`. URL: `https://doi.org/10.1051/itmconf/20182300037`.

[37] B. W. Silverman. *Density estimation: For statistics and data analysis*. CRC Press, Jan. 2018, pp. 1–175. ISBN: 9781351456173. DOI: `10.1201/9781315140919`. URL: `https://www.taylorfrancis.com/books/mono/10.1201/9781315140919/density-estimation-statistics-data-analysis-bernard-silverman`.