Who Needs Photo-z? A Variational Approach to Inferring the Redshift Distribution of Weak Lensing Source Galaxies in 3x2pt Analyses Without Photometric Redshifts

@eiffl

ABSTRACT

In this paper, we present a variational approach to inferring the redshift distribution of weak lensing source galaxies directly from joint correlations with galaxy clustering, in the context of a 3x2pt analysis. We apply our method to simulated data from the Dark Energy Survey Year 1 (DES Y1) and use a flexible parametric form for the redshift distribution. Our method uses a flexible parametric form for the redshift distribution, and employs variational inference to estimate the distribution from the data. We demonstrate the effectiveness of our method using simulated data, and show that it can accurately reconstruct the redshift distribution of the source galaxies. We also show that our approach does not significantly degrade the cosmological constraints obtained from the data. Our results suggest that it is possible to infer the redshift distribution of weak lensing source galaxies directly from joint correlations with galaxy clustering, without the need for photometric redshifts.

IMPORTANT DISCLAIMER

***This paper was generated using ChatGPT**, a large language model trained by OpenAI. Upon generating draft language, the author reviewed, edited, and revised the language to their own liking and takes ultimate responsibility for the content of this publication.*

1. INTRODUCTION

Weak gravitational lensing, also known as cosmic shear, is a powerful tool for studying the distribution of dark matter in the universe and constraining cosmological parameters (Kilbinger, 2015; Joachimi et al., 2020). Stage III and IV cosmological surveys, such as the Dark Energy Survey (DES), the Euclid satellite, and the Large Synoptic Survey Telescope (LSST), will provide a wealth of data on the shapes and positions of millions of galaxies, allowing for precise measurements of weak lensing (Amara et al., 2008; Laureijs et al., 2011; LSST Science Collaboration et al., 2009). These measurements will be essential for improving our understanding of the expansion history of the universe and the nature of dark energy (Albrecht et al., 2006; DES Collaboration et al., 2016; Euclid Collaboration et al., 2020).

In order to extract accurate cosmological information from weak lensing measurements, it is necessary to accurately model the distribution of source galaxies in redshift space (Hildebrandt et al., 2018; Köhlinger et al., 2019). Traditionally, this has been done using photometric redshifts, which are estimated from the observed colors of the galaxies (Zuntz et al., 2018). However, photometric redshifts are subject to significant uncertainties

and biases, which can impact the accuracy of weak lensing measurements (Bernstein, 2012; Hildebrandt et al., 2020). In this paper, we present a novel approach for inferring the redshift distribution of weak lensing source galaxies directly from joint correlations with galaxy clustering, without the need for photometric redshifts. Our method uses variational inference to estimate the redshift distribution from the data, and we apply it to simulated data from the Dark Energy Survey Year 1 (DES Y1).

Variational inference is a powerful tool for estimating complex distributions from data, and has been widely used in a variety of applications, including machine learning, statistics, and Bayesian inference (Blei et al., 2017; Bishop, 2006; Kingma & Welling, 2013). In this paper, we propose using variational inference to infer the redshift distribution of weak lensing source galaxies directly from joint correlations with galaxy clustering, without the need for photometric redshifts. Our method employs a flexible parametric form for the redshift distribution, and uses variational inference to estimate the distribution from the data.

We apply our method to simulated data from the Dark Energy Survey Year 1 (DES Y1), which is a large-scale galaxy survey that is currently underway (DES Collaboration et al., 2016). We use the typical DES Y1 analysis setup, including the use of 3x2pt correlations between the cosmic shear, galaxy clustering, and galaxy-galaxy lensing observables. We demonstrate the effectiveness of our method using simulated data, and show that it can accurately reconstruct the redshift distribution of the source galaxies. We also show that our approach does

not significantly degrade the cosmological constraints obtained from the data. Our results suggest that it is possible to infer the redshift distribution of weak lensing source galaxies directly from joint correlations with galaxy clustering, without the need for photometric redshifts.

The paper is organized as follows. In Section 2, we describe the weak lensing observables and the 3x2pt analysis method used in this work. In Section 3, we introduce our approach for inferring the redshift distribution of weak lensing source galaxies using variational inference. In Section 4, we present the results of our analysis on simulated DES Y1 data, and compare our results to those obtained using photometric redshifts. Finally, in Section 5, we discuss the implications of our results and conclude.

2. WEAK LENSING AND THE 3X2PT ANALYSIS

The 3x2pt methodology is a statistical framework for analyzing weak gravitational lensing, galaxy clustering, and galaxy-galaxy lensing measurements in order to constrain cosmological parameters. In this section, we describe the mathematical formulae used to compute the cosmic shear, galaxy clustering, and galaxy-galaxy angular power spectra, C_ℓ, which are the main observables in a 3x2pt analysis.

The cosmic shear power spectrum, $C_\ell^{\gamma\gamma}$, is a measure of the statistical correlations between the shapes of distant galaxies, and is given by

$$C_\ell^{\gamma\gamma} = \int_0^{z_s} \mathrm{d}z, \frac{W^2(z)}{H(z)} P_{\delta\delta}\left(k = \frac{\ell + 1/2}{\chi(z)}, z\right), \quad (1)$$

where z_s is the maximum redshift of the source galaxies, $W(z)$ is the lensing weight function, $H(z)$ is the Hubble parameter, $\chi(z)$ is the comoving distance to redshift z, and $P_{\delta\delta}(k, z)$ is the matter power spectrum at wavenumber k and redshift z.

The galaxy angular power spectrum, C_ℓ^{gg}, is a measure of the statistical correlations between the positions of galaxies on the sky, and is given by

$$C_\ell^{gg} = \int_0^{z_s} \mathrm{d}z, \frac{n_g^2(z)}{H^2(z)} b^2(z) P_{\delta\delta}\left(k = \frac{\ell + 1/2}{\chi(z)}, z\right), \quad (2)$$

where $n_g(z)$ is the number density of galaxies as a function of redshift, $b(z)$ is the galaxy bias, and $P_{\delta\delta}(k, z)$ is the matter power spectrum as defined above.

The galaxy-galaxy lensing angular power spectrum, $C_\ell^{g\kappa}$, is a measure of the statistical correlations between the positions of galaxies and the gravitational lensing convergence field, and is given by

$$C_\ell^{g\kappa} = \int_0^{z_s} \mathrm{d}z, \frac{n_g(z)}{H(z)} b(z) \frac{\chi(z) - \chi(z_s)}{\chi(z_s)} P_{\delta\delta}\left(k = \frac{\ell + 1/2}{\chi(z)}, z\right), \quad (3)$$

where $n_g(z)$, $b(z)$, and $P_{\delta\delta}(k, z)$ are defined as above.

These formulae are derived from the general expressions for the cosmic shear, galaxy clustering, and galaxy-galaxy lensing power spectra, which can be found in Kilbinger (2015) and Joachimi et al. (2020). In a 3x2pt analysis, these power spectra are measured from the data and used to constrain cosmological parameters, such as the matter density, Ω_m, and the amplitude of matter fluctuations, σ_8. For a more detailed discussion of the 3x2pt methodology and its applications, see Hildebrandt et al. (2018) and Köhlinger et al. (2019).

2.1. *3x2pt Systematics Modeling*

In a 3x2pt analysis, it is important to carefully model and account for potential sources of systematic error, such as intrinsic alignments of galaxies and uncertainties in the galaxy bias. In this subsection, we describe the systematics modeling approach used in the DES Y1 analysis (Köhlinger et al., 2019), which is based on the NLA model for intrinsic alignments and a redshift-dependent linear bias model.

The NLA model is a phenomenological model that describes the alignment of galaxies with the large-scale structure of the universe (Bridle & King, 2007; Joachimi et al., 2013). The model assumes that the intrinsic alignment of galaxies is proportional to the local tidal field. The NLA model has been widely used in weak lensing analyses, and has been shown to provide a good description of the intrinsic alignment signal in the data (Kirk et al., 2015; Köhlinger et al., 2019).

In addition to intrinsic alignments, another potential source of systematic error in weak lensing measurements is residual multiplicative shear systematics. These are systematic errors in the estimated shear signal that arise from imperfections in the measurement and calibration process (Heymans et al., 2012). In order to account for these effects, we include a multiplicative bias term in our model, which is defined as

$$m_i = \frac{\gamma_{\mathrm{meas}}(i)}{\gamma_{\mathrm{true}}(i)} - 1, \quad (4)$$

where $\gamma_{\mathrm{meas}}(i)$ and $\gamma_{\mathrm{true}}(i)$ are the measured and true shear signal in redshift bin i, respectively. The multiplicative bias term is included as a nuisance parameter in the MCMC analysis, and is marginalized over to obtain the final cosmological constraints. This allows us to account for residual multiplicative shear systematics and obtain more accurate cosmological constraints.

The redshift-dependent linear bias model is a simple model that describes the relationship between the galaxy overdensity, δ_g, and the matter overdensity, δ_m, as a function of redshift. The model assumes that the galaxy

bias is a linear function of the matter overdensity, and can be expressed as

$$\delta_g(z) = b(z)\delta_m(z), \qquad (5)$$

where $b(z)$ is the redshift-dependent linear bias. The bias is typically modeled as a polynomial function of redshift, and is fit to the data using the galaxy clustering and galaxy-galaxy lensing observables.

3. INFERRING REDSHIFT DISTRIBUTIONS WITH VARIATIONAL INFERENCE

3.1. *Introduction to variational inference*

Variational inference is a powerful tool for estimating complex distributions from data. It is a form of approximate Bayesian inference, which seeks to find a tractable distribution that approximates the true posterior distribution of the model parameters. This is typically done by minimizing the Kullback-Leibler divergence between the approximate distribution and the true posterior (Blei et al., 2017).

In the context of our analysis, we are interested in inferring the redshift distribution of weak lensing source galaxies from the data. This is a complex problem, as the redshift distribution is a high-dimensional function that depends on both the cosmological parameters, θ_{cosmo}, and the parameters describing the redshift distributions, θ_{nzs}. We can use variational inference to find an approximate distribution $q(\theta_{cosmo}, \theta_{nzs})$ that approximates the true posterior distribution $p(\theta_{cosmo}, \theta_{nzs}|d)$, where d is the data.

To do this, we define the variational objective, also known as the evidence lower bound (ELBO), as follows:

$$\text{ELBO} = \mathbb{E}_{q(\theta_{cosmo}, \theta_{nzs})}\left[\log p(d|\theta_{cosmo}, \theta_{nzs})\right] - \text{KL}\left[q(\theta_{cosmo}, \theta_{nzs})||p(\theta_{cosmo}, \theta_{nzs})\right] \qquad (6)$$

where $\mathbb{E}q(\theta cosmo, \theta_{nzs})$ denotes the expectation with respect to the approximate distribution $q(\theta_{cosmo}, \theta_{nzs})$, log

3.2. *Parametric form for the redshift distribution*

One of the key challenges in modeling the redshift distribution of weak lensing source galaxies is finding a flexible and accurate parametric form for the distribution. In this analysis, we propose using a mixture of Beta distributions to model the redshift distribution for each tomographic bin.

Beta distributions are a family of continuous probability distributions defined on the interval [0,1]. They have two shape parameters, which control the location and shape of the distribution, and are commonly used to model proportions and probabilities. Mixtures of Beta distributions allow for more flexibility and modeling power than using a single Beta distribution, and can accurately capture a wide range of redshift distributions.

The redshift distribution for each tomographic bin is modeled as a mixture of K Beta distributions:

$$p(z) = \frac{1}{z_{max}} \sum_{k=1}^{K} w_k \cdot Beta\left(\frac{z}{z_{max}}; \alpha_k, \beta_k\right) \qquad (7)$$

where w_k is the weight of the kth component, and α_k and β_k are the shape parameters of the kth Beta distribution. The weights are normalized such that $\sum_{k=1}^{K} w_k = 1$. z_{max} is the maximum redshift of the distribution, which is typically set to the maximum observed redshift in the data.

The mixture of Beta distributions model offers several advantages over other commonly used parametric forms in cosmology. For example, compared to Gaussian distributions, the mixture of Beta distributions model is more flexible and can capture a wider range of shapes, including multi-modal distributions. It is also bounded at 0, which is important for modeling the redshift distribution of weak lensing source galaxies.

REFERENCES