

INFERENCE OF THE CONDITIONAL LUMINOSITY FUNCTION BY EXPLICIT MARGINALIZATION OVER ALL INDIVIDUAL GALAXY PARAMETERS

DAVID THOMAS,¹ PHIL MARSHALL,^{1,2} AND RISA WECHSLER^{1,2}
(LSST DARK ENERGY SCIENCE COLLABORATION)

¹*Kavli Institute for Particle Astrophysics & Cosmology, P. O. Box 2450, Stanford University, Stanford, CA 94305, USA*

²*SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*

ABSTRACT

We explore a hierarchical model for galaxy luminosities and halo masses whose hyperparameters govern the conditional luminosity function, and whose many individual object parameters are explicitly marginalized over. We investigate four techniques to perform the high dimensional marginalization, and comment on their performance. Furthermore, we analyze how these techniques scale in different scenarios.

Keywords: latex: templates, papers: awesome

1. INTRODUCTION

The large scale structure of the universe is dominated by the collective gravitational influence of dark matter particles and the repulsive force of dark energy. While there are ongoing efforts to infer the masses and locations of dark matter halos and improve measurements of dark energy parameters, observational progress has been gradual, encouraging alternative approaches. Hydrodynamic simulations have been an effective tool for bolstering our understanding of these contributors and their impact in structure formation and evolution. Given a cosmology, large N-body simulations can predict characteristics like clustering and the halo mass function with high precision (eg. [Springel, Frenk, & White 2006](#)). Virtual dark matter universes are readily available, and inspire us to explore ways to connect them to empirical observations.

The halo mass and galaxy luminosity relation is one such conduit between dark matter halos and galaxy observations. The conditional probability allows us to convert the dark matter halos from simulations into distributions of galaxy luminosities in different redshift bins. These distributions can be compared to observed galaxy luminosities to further constrain the halos and relation. There has been work to determine the exact form of the relation and different models for centrals and satellite galaxies have been proposed (eg. [Yang et al. 2005](#)).

Weak lensing is another technique that can characterize the halo composition of our universe. The gravitational influence of a massive dark matter halo causes the light from nearby galaxies to bend as it passes by, leading to a shearing effect. Over a swathe of sky, typi-

cally on the order of 1000 square arcminutes, the collective shears allow the masses and locations of halos to be weakly inferred. These inferences grow stronger as the density of galaxy observations increases. The ongoing Dark Energy Survey and upcoming Large Synoptic Survey Telescope will provide dense galaxy catalogs which will benefit weak lensing ([Dark Energy Survey Collaboration et al. 2016](#); [LSST Science Collaboration et al. 2009](#)).

Traditionally the mass luminosity relation and weak lensing have been pursued in isolation, each ignoring information the other could provide. It is our goal to harness the full constraining power by modelling both weak lensing and the mass luminosity relation in a self consistent, heirarchical framework. We employ the mass luminosity relation described in [Reddick \(2014\)](#) and the mass mapping infrastructure from [Marshall \(2006\)](#). Figure 1 outlines the dependencies of the model we aspire to build.

The model partitions galaxies into two groups. Background galaxies are sources that lie in a redshift bin immediately behind the three dimensional space of foreground galaxies. The background galaxies provide the observed shears which we use to infer the halo masses of foreground galaxies. We plan to use around 200,000 foreground galaxies and 1,600 background galaxies in our model and analysis. In the course of the heirarchical inference, each of the 200,000 galaxies contributes and has a corresponding mass and luminosity which must be marginalized out. Given the large size of our inference, we expect the primary challenges to be computational in nature - how can we achieve inference at this unprece-

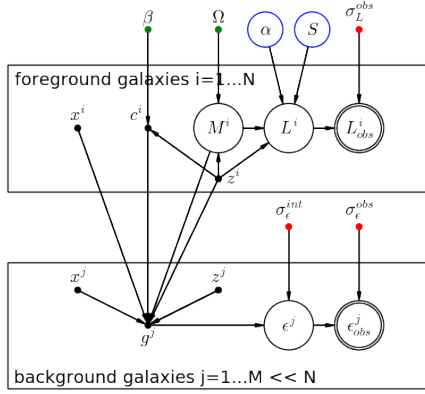


Figure 1. A probabilistic graphical model for the model we ultimately aspire to. There are two plates; one for foreground galaxies, one for background galaxies. The cosmology hyperparameters are colored green, the mass luminosity relation hyperparameters are colored blue, and the noise hyperparameters are colored red. β represents the form of the halo concentration. Ω represents the cosmology used to generate the halos in the Millennium Simulation. α represents four parameters that convert mass to mean luminosity, and S represents the scatter of the corresponding lognormal. M, L represent mass and luminosity respectively. The variables x, z denote angular position and redshift respectively. c is for the concentration, g is for the reduced shear, and ϵ is for the shear. Finally, σ represents various forms of noise and the superscript *obs* is used to denote an observational noise or observable.

dent scale? In order for us to rapidly iterate towards an answer to this question, we study a model with simplified dependencies. This condensed model, shown in Figure 2, is easier to test and retains many of the computational challenges.

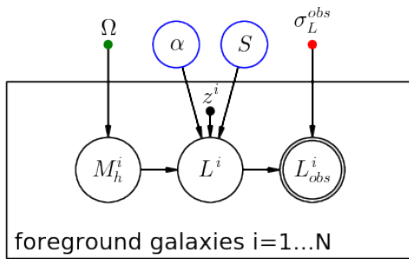


Figure 2. A probabilistic graphical model for the simplified model we will use for testing. The parameters are a subset of those in Figure 1.

The data we use ultimately comes from the Millennium Simulation (Springel et al. 2005). Our desire to eventually incorporate weak lensing components into our model drives our selection. Hilbert et al. (2009) have performed raytracing which we plan to make use of in later analysis. A friend-of-friend (FoF) group finding algorithm identifies halos from particle concentrations. We carve out a 40×40 square arcminute window of sky out to redshift 3.5, and use the interior halos as the foreground galaxies in our analysis.

The statistical inference we perform follows the canonical inference formula: posterior equals prior times likelihood. In the analysis that follows we use an overline, such as \bar{z} , to distinguish a vector over all foreground galaxies from the variable corresponding to a single halo, or z . Given the mass luminosity hyperparameters α, S ; a vector of observed luminosities \bar{L}^{obs} ; a vector of redshifts \bar{z} ; and the observational noise in luminosity measurements σ_L^{obs} - the posterior we seek is $P(\alpha, S | \bar{L}^{obs}, \bar{z}, \sigma_L^{obs}) = P(\alpha, S) P(\bar{L}^{obs} | \alpha, S, \bar{z}, \sigma_L^{obs})$. Using the probabilistic graphical model we factor this further to

$$P(\alpha, S | \bar{L}^{obs}, \bar{z}, \sigma_L^{obs}) = P(\alpha) P(S) \iint d\bar{M} d\bar{L} P(\bar{L}^{obs} | \bar{L}, \sigma_L^{obs}) P(\bar{L} | \bar{M}, \alpha, S, \bar{z}) P(\bar{M} | \bar{z})$$

The likelihood integral involves integrating over 400,000 variables, the mass and luminosity for each galaxy in our dataset. The Methods section of this paper describes various approaches to make such a large integral computationally tractable.

The integrand of the likelihood contains three probabilities. The first is the conditional observed luminosity probability. Since luminosities are usually reported in log-space, we assume that there are gaussian errors in the log-space luminosity measurements, or that the distribution is log-normal. We believe 5% errors seem reasonable and fix $\sigma_L^{obs} = 0.05$. Our conditional distribution is

$$P(\bar{L}^{obs} | \bar{L}, \sigma_L^{obs}) = \frac{1}{\bar{L}^{obs} \sigma_L^{obs} \sqrt{2\pi}} \exp \left(-\frac{(\ln \bar{L}^{obs} - \ln \bar{L})^2}{2\sigma_L^{obs}{}^2} \right)$$

The second factor in the likelihood integrand is the mass luminosity relation $P(\bar{L} | \bar{M}, \alpha, S, \bar{z})$. We use the form described in (reddick's thesis). As in numerous previous studies, Reddick divides the conditional luminosity into two parts. The central galaxy is described by a lognormal distribution and the satellite galaxies are described by a Schechter function. In our analysis, we

ignore the satellite contribution. This decision is made for convenience, and we can easily add the satellite contribution in future analyses. We use $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ to represent the parameters for computing the mean luminosity and S for the scatter. This gives

$$P(\bar{L}|\bar{M}, \alpha, S, \bar{z}) = \frac{1}{\bar{L}S\sqrt{2\pi}} \exp\left(-\frac{(\ln \bar{L} - \ln \bar{\mu}_L)^2}{2S^2}\right)$$

$$\mu_L = \exp(\alpha_1) \cdot \left(\frac{M}{\alpha_3}\right)^{\alpha_2} \cdot (1+z)^{\alpha_4}$$

The third factor in the likelihood integrand is the halo mass function $P(\bar{M}|\bar{z})$. For this we rely on the python software package `hmfcalc` (Murray, Power, & Robotham 2013). We use the functional form from (Tinker et al. 2008) and the cosmology from the Millennium Simulation.

2. METHODS

2.1. Test Dataset

2.2. Numerical Integration

2.3. Simple Monte Carlo

2.4. Importance Sampling

Plots showing why biased distribution makes sense.

2.5. Laplace Approximation

3. RESULTS

Compare single integrals.

Analyze how importance sampling and laplace approximation scale.

4. DISCUSSION

Mention obvious tradeoffs between methods.

Mention BigMaLI codebase for reproducibility.

Back of the envelope cluster scaling.

5. CONCLUSIONS

Short summary.

ACKNOWLEDGMENTS

Here is where you should add your specific acknowledgments, remembering that some standard thanks will be added via the `acknowledgments.tex` and `contributions.tex` files.

This is the text imported from `acknowledgments.tex`, and will be replaced by some standard LSST DESC boilerplate at some point.

Author contributions are listed below.

David Thomas: Initiated project, led development work.

Phil Marshall: Advised on statistics.

Risa Wechsler: Advised on galaxy formation and cosmology.

REFERENCES

- Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270
- Hilbert, S., Hartlap, J., White, S. D. M., & Schneider, P. 2009, A&A, 499, 31
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, ArXiv e-prints, arXiv:0912.0201
- Marshall, P. 2006, MNRAS, 372, 1289
- Murray, S. G., Power, C., & Robotham, A. S. G. 2013, Astronomy and Computing, 3, 23
- Reddick, R. 2014, PhD thesis
- Springel, V., Frenk, C. S., & White, S. D. M. 2006, Nature, 440, 1137
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629
- Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709
- Yang, X., Mo, H. J., van den Bosch, F. C., et al. 2005, MNRAS, 362, 711