How to estimate the redshift density distribution

Alex Malz[1] & David W. Hogg[1,2,3,4]

aimalz@nyu.edu

ABSTRACT

The redshift density function $n(z)$ is a crucial ingredient for weak lensing cosmology. Spectroscopically confirmed redshifts for the dim and numerous galaxies observed by past, present, and future weak lensing surveys are inaccessible, making photometric redshifts (photo-zs) the next best alternative. Due to their intrinsic scatter and susceptibility to catastrophic outliers, photo-z point estimates are being superseded by photo-z probability distribution functions (PDFs). However, analytic methods for utilizing these new data products in cosmological inference are still evolving.

This paper presents a novel approach to estimating the posterior distribution over $n(z)$ from a survey of galaxy photo-z PDFs based upon a probabilistic graphical model. We present the Cosmological Hierarchical Inference with Probabilistic Photometric Redshifts (CHIPPR) code implementing this technique, as well as its validation on mock data and testing on a subset of BOSS DR10. CHIPPR yields a more comprehensive and accurate characterization of $n(z)$ than traditional procedures. The code is easily extensible to other one-point statistics of redshift that are informative to galaxy evolution and large-scale structure.

Subject headings: methods: analytical "___" methods: data analysis "___" methods: statistical "___" techniques: photometric "___" galaxies: statistics

1. Introduction

The redshift density function $n(z)$ is necessary for calculating two-point statistics of galaxy properties used to determine the cosmological parameter values that inform our understanding of the evolution of large-scale structure in the universe (Masters et al. 2015). Inaccurate estimates of

[1]Center for Cosmology and Particle Physics, Department of Physics, New York University, 4 Washington Pl., room 424, New York, NY 10003, USA

[2]Simons Center for Data Analysis, 160 Fifth Avenue, 7th floor, New York, NY 10010, USA

[3]Center for Data Science, New York University, 726 Broadway, 7th floor, New York, NY 10003, USA

[4]Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

$n(z)$ can significantly impact the constraining power of a galaxy survey (Hildebrandt et al. 2016), biasing recovery of the cosmological parameters.

Though the redshift density function has traditionally been determined from spectroscopically observed redshifts, modern galaxy surveys seek to obtain two-point statistics of redshift from samples of galaxies for which spectroscopic redshifts are unavailable, either due to their large numbers or their low brightnesses. For decades (Baum 1962), photometrically estimated redshifts (photo-zs) have been the leading alternative to spectroscopically observed redshifts, though they suffer from issues of precision in the form of an intrinsic scatter and accuracy in the form of catastrophic outliers, as well as other systematics imparted by the properties of the survey, data reduction pipeline(s), and assumptions underlying the analysis.

These weaknesses are illuminated by a probabilistic interpretation of photo-zs; if these nontrivial uncertainties were expressed as a probability distribution function (PDF) over redshift, photo-zs could be replaced by photo-z PDFs containing more information than a simple point estimate (Koo 1999). However, methods for using photo-z PDFs in cosmological inference remain underdeveloped, with many survey pipelines reducing them to familiar point estimators that are compatible with existing technology or engaging with them heuristically in a manner inconsistent with their probabilistic nature.

It is desirable to create robust methods for using photo-z PDFs in inference, beginning with the simplest one-point statistic of redshift, $n(z)$. In the spirit of Hogg et al. (2010); Foreman-Mackey et al. (2014), this paper derives a mathematically rigorous approach to inferring $n(z)$ in Sec. 2. The presentation of a public code implementing this novel technique is given in Sec. 3. The code is validated on mock data and tested on a subset of BOSS DR10 data in Sec. 4. We conclude and discuss future directions in Sec. 5.

2. Method

3. Model specifics

4. Experiments

5. Discussion

AIM thanks Mohammadjavad Vakili for insightful input on statistics, Geoffrey Ryan for assistance in debugging code, and Boris Leistedt for helpful comments provided in the preparation of this paper.

REFERENCES

Baum, W. A. 1962, Proc. from IAU Symp.

Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, Astrophys. J., 795, 64

Hildebrandt, H., Viola, M., Heymans, C., et al. 2016, Monthly Notices of the Royal Astronomical Society, doi:10.1093/mnras/stw2805

Hogg, D. W., Myers, A. D., & Bovy, J. 2010, Astrophys. J., 725, 2166

Koo, D. C. 1999, in Photometric Redshifts and the Detection of High Redshift Galaxies, Vol. 191, eprint: arXiv:astro-ph/9907273, 3

Masters, D., Capak, P., Stern, D., et al. 2015, Astrophys. J., 813, 53