

# How to obtain the redshift distribution from probabilistic redshift estimates

Alex Malz<sup>1</sup> & David W. Hogg<sup>1,2,3,4</sup>

aimalz@nyu.edu

## ABSTRACT

The redshift distribution  $n(z)$  is a crucial ingredient for weak lensing cosmology. Spectroscopically confirmed redshifts for the dim and numerous galaxies observed by weak lensing surveys are expected to be inaccessible, making photometric redshifts (photo- $z$ s) the next best alternative. Because of the nontrivial inference involved in their determination, photo- $z$  point estimates are being superseded by photo- $z$  probability distribution functions (PDFs).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 of hierarchical inference. 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 an accurate characterization of  $n(z)$  containing information beyond the best-fit estimator produced by traditional procedures. The publicly available code is easily extensible to other one-point statistics that depend on redshift.

*Subject headings:* catalogs — cosmology: cosmological parameters — galaxies: statistics — gravitational lensing: weak — methods: analytical — methods: data analysis — methods: statistical — techniques: photometric

## 1. Introduction

The redshift distribution  $n(z)$  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

---

<sup>1</sup>Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, 9th floor, New York, NY 10003, USA

<sup>2</sup>Simons Center for Computational Astrophysics, 162 Fifth Avenue, 7th floor, New York, NY 10010, USA

<sup>3</sup>Center for Data Science, New York University, 60 Fifth Avenue, 7th floor, New York, NY 10003, USA

<sup>4</sup>Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

$n(z)$  can significantly impact the constraining power of a galaxy survey, biasing recovery of the cosmological parameters. For example, if the  $n(z)$  used in analysis is offset from the true  $n(z)$  by a small, negative constant, it results in underestimating  $w_0$  and overestimating  $\sigma_8$  (Samuroff et al. 2017).

Though the redshift density function has traditionally been determined from spectroscopically observed redshifts, modern galaxy surveys including DES, LSST, Euclid, and WFIRST 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, photometrically estimated redshifts (photo- $z$ s) 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 (Baum 1962). These weaknesses are illuminated by a probabilistic interpretation of photo- $z$ s; if these nontrivial uncertainties were expressed as a probability distribution function (PDF) over redshift, photo- $z$ s could be replaced by photo- $z$  PDFs containing more information than a simple point estimate (Koo 1999). Such data products have been commonly released by photometric galaxy surveys since SDSS DR7 (Abazajian & Survey 2009) using a great variety of methods.

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. Stacking photo- $z$  PDFs, or other mathematically equivalent methods, to obtain an estimator of  $n(z)$  is especially popular (Cunha et al. 2009; Sheldon et al. 2012).

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 assis-

tance in debugging code, and Boris Leistedt for helpful comments provided in the preparation of this paper.

## REFERENCES

- Abazajian, K., & Survey, f. t. S. D. S. 2009, The Astrophysical Journal Supplement Series, 182, 543, arXiv: 0812.0649
- Baum, W. A. 1962, Proc. from IAU Symp.
- Cunha, C. E., Lima, M., Oyaizu, H., Frieman, J., & Lin, H. 2009, Mon. Not. R. Astron. Soc., 396, 2379
- Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, Astrophys. J., 795, 64
- 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
- Samuroff, S., Troxel, M. A., Bridle, S. L., et al. 2017, Monthly Notices of the Royal Astronomical Society: Letters, 465, L20
- Sheldon, E. S., Cunha, C. E., Mandelbaum, R., Brinkmann, J., & Weaver, B. A. 2012, Astrophys. J. Suppl. Ser., 201, 32