How to obtain the redshift distribution from probabilistic redshift estimates

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

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-zs) 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 Because of the nontrivial inference involved in their determination, photo-z point estimates are being superseded by photo-z probability distribution functions (PDFs).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 the `Buzzard` simulations. `CHIPPR` yields a more 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

`CHIPPR` aims to answer the following questions:

- How are photo-z PDFs currently used in cosmology?

[1]Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, 9th floor, New York, NY 10003, USA

[2]Simons Center for Computational Astrophysics, 162 Fifth Avenue, 7th floor, New York, NY 10010, USA

[3]Center for Data Science, New York University, 60 Fifth Avenue, 7th floor, New York, NY 10003, USA

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

- Why should we challenge existing methods?
- How should photo-z PDFs be used in inference?
- How does the result of `CHIPPR` compare to established estimators of $n(z)$?

2. Method

2.1. Model

This is where I will present the DAG and corresponding math, along with an explicit discussion of the assumptions that must be satisfied in order for the `CHIPPR` approach to be valid.

2.2. Implementation

I will describe the `CHIPPR` code here.

2.3. Metrics

I will discuss the metrics that I use in 3 and include in the release version.

3. Validation

This is where the setup and results of the simulated test cases will be presented.

3.1. Variations on Photo-z Likelihoods

I think this is too many test cases!

- A fiducial case of Gaussian likelihoods with narrow, redshift-independent intrinsic scatter and no other systematics; also the fiducial case with larger intrinsic scatter
- The fiducial case with catastrophic outliers as seen in template-based photo-z estimators (attractor in the space of z_{phot}: some galaxies at a range of z_{spec} map to the same z_{phot} because their shared SED type does not have sufficiently strong features, leading galaxies of that type at many z_{spec} to have the same colors)

- The fiducial case with catastrophic outliers as seen in training-based photo-z estimators (attractor in the space of z_{spec}: some galaxies at a particular z_{spec} map to a range of z_{phot} because their shared SED features fall between filters, leading to many galaxies at particular z_{spec} to have the same colors)

3.2. Variations on Interim Priors

I will show that a really wacky interim prior, something very far from the true $n(z)$, will imprint itself on the stacked estimator but will have no effect on `CHIPPR` 's result.

3.3. Variations on True Hyperparameters

I will show that if $n(z)$ has significant features, stacking will not recover them but `CHIPPR` will.

4. Application

I will apply `CHIPPR` to a realistic simulated dataset (probably `Buzzard`) so the results can be compared to the truth. The photo-z PDFs should be derived using the `SDSS`-approved k nearest neighbors algorithm because it has the most straightforward interim prior.

5. Conclusion

I will review the questions from the introduction:

- Existing methods have shortcomings that propagate to inaccuracies in characterizing the cosmological parameters.
- Photo-z PDFs are probabilistic data products so must be handled in a mathematically consistent manner; `CHIPPR` is one such method, conditioned on some assumptions.
- In addition to coming with its own error distribution, the $n(z)$ estimator produced by `CHIPPR` is more accurate than established estimators; furthermore, propagation of the `CHIPPR` result leads to a quantifiable improvement in the constraints on cosmological parameters.

I may briefly discuss some obvious extensions of `CHIPPR` , such as a fully hierarchical inference of the cosmological parameters that avoids tomographic binning based on photo-z point estimates.

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

REFERENCES

This preprint was prepared with the AAS LaTeX macros v5.2.