Redshift estimates for samples of galaxies using correlation functions without a spectroscopic reference sample

Ben Hoyle, Markus Rau, DES
Max Plank Extragalactic Astrophysics
University Sternwarte Munich
The biggest problem facing current large area galaxy surveys: Validating redshifts for samples of galaxies.
The biggest problem facing current large area galaxy surveys: Validating redshifts for samples of galaxies.

Colors of Galaxies → map → photometric-z dn/dz → validate → dn/dz spec-z → True-dn/dz

Dark Energy Survey SV Bonnet & DES (2016)
The biggest problem facing current large area galaxy surveys: Validating redshifts for samples of galaxies.

Bonnet & DES (2016)

Dark Energy Survey SV
The biggest problem facing current large area galaxy surveys: Validating redshifts for samples of galaxies.

We extract 1-d spectra from simulations (known redshift), added realistic noise. Ask OzDES observers to redshift the spectra, using their common analysis tools.

We cannot validate photo-z performance on data which is biased w.r.t the truth.
The biggest problem facing current large area galaxy surveys: Validating redshifts for samples of galaxies.

The solution:
The self-consistent correlation redshift method.
The biggest problem facing current large area galaxy surveys: Validating redshifts for samples of galaxies.

The solution:
The self-consistent correlation correlation redshift method.
The biggest problem facing current large area galaxy surveys: Validating redshifts for samples of galaxies.

The solution: The self-consistent correlation redshift method.

Results using idealised simulations with known redshifts.

How well can we recover redshifts distributions?

Colors of Galaxies
pos. on sky

Measured CIs

theory CIs

consistent dn/dz

True dn/dz

If the model is correct!
The self-consistent correlation redshift method

Step 1. Choose some [e.g. 12] sub-samples of galaxies in observational space
e.g. Radio emission, galaxy color/magnitude, photo-z=> color cell.
The self-consistent correlation redshift method

Step 1. Choose some [e.g. 12] sub-samples of galaxies in observational space
e.g. Radio emission, galaxy color/magnitude, photo-z=> color cell.

Step 2. View their distribution on the Sky

- Color (g-r)
- Magnitude (r)
Step 3. Measure auto- cross- correlations, and galaxy-galaxy lensing, and galaxy-CMB lensing.

The self-consistent correlation redshift method
Step 3. Measure auto-cross-correlations, and galaxy-galaxy lensing, and galaxy-CMB lensing.

We don’t need to know any redshift information, galaxy-dark matter bias, or the cosmological model parameters to make these measurements.
Step 4. Model the Cls. Parameterise our uncertainties, for each color cell.

\[
C_l^{xy} = 4\pi \int \frac{dk}{k} P_{\text{ini}}(k) \Delta^x_l(k) \Delta^y_l(k)
\]

\[
P_{xy}(k) \approx b_x(k, z) b_y(k, z) P(k)
\]

\[
b(k, z) = k^{b_k} (b_0 + b_1 \times (1 - 1/(1 + z))
\]

\[
\Delta^x_l(k) = \int dz b_x(k, z) \frac{dn_x}{dz} j_l(k \chi(z)) D(k, z)
\]
The self-consistent correlation redshift method

Step 4. Model the Cls. Parameterise our uncertainties, for each color cell.

\[ C_{l}^{xy} = 4\pi \int \frac{dk}{k} P_{ini}(k) \Delta_{l}^{x}(k) \Delta_{l}^{y}(k) \]

\[ P_{xy}(k) \approx b_{x}(k, z) b_{y}(k, z) P(k) \]

\[ b(k, z) = k^{b_{k}} (b_{0} + b_{1} \times (1 - 1/(1 + z))) \]

\[ \Delta_{l}^{x}(k) = \int dz b_{x}(k, z) \frac{dn_{x}}{dz} j_{l}(k\chi(z)) D(z) \]

We choose a flexible 5 Gaussian Mixture model, to describe the redshift distribution in each of the 12 color cells.
We choose a flexible 5 Gaussian Mixture model, to describe the redshift distribution in each of the 12 color cells.

$$C_{i}^{xy} = 4\pi \int \frac{dk}{k} P_{\text{ini}}(k) \Delta_{i}^{x}(k) \Delta_{i}^{y}(k)$$

$$P_{xy}(k) \approx b_{x}(k, z) b_{y}(k, z) P(k)$$

$$b(k , z) = k^{b_{k}} \left( b_{0} + b_{1} \times \left( 1 - 1/(1 + z) \right) \right)$$

$$\Delta_{i}^{x}(k) = \int dz \; b_{x}(k, z) \frac{dn_{x}}{dz} j_{i}(k\chi(z)) \mathcal{D}(i)$$

We choose a flexible 5 Gaussian Mixture model, to describe the redshift distribution in each of the 12 color cells.

+3*5 Gauss. Mix. model parameters per color cell

= 18 parameters per color cell to estimate.
Step 5. Make some random (or reasonable) guesses for each of the parameter values in the model.

Ask CAMB/CLASS for model Cls.

\[
\chi^2 = \sum_{i=1}^{N} \sum_{l} \left( \frac{C_i^D - C_i^M}{\sigma_l} \right)^2
\]
The self-consistent correlation redshift method

Step 5. Make some random (or reasonable) guesses for each of the parameter values in the model.
Ask CAMB/CLASS for model Cls.

\[ \chi^2 = \sum_{i=1}^{N} \sum_{l} \left( \frac{C_{l}^{\text{D}} - C_{l}^{\text{M}}}{\sigma_{l}} \right)^2 \]

Roll down the 18*12 dimensional parameter space, trying to find suitable dn/dz and gal-dm bias model param. values for each color cell, such that the model Cls are consistent with the measured Cls.

\[ \chi^{2}_{R}(\text{DoF: } \sim 2304) \]
Overview of results for color-cell: #4 / 12

True: \( <z> = 0.2849 \)  \( \sigma_{68,95,99.5} = 0.12, 0.24, 0.32 \)

Initial Guess: \( \Delta z = 0.32 \)

Final: \( \Delta z = -2.7 \times 10^{-4} \) (-0.1\%)  \( |\delta \sigma_{68,95,99.5}| = 0\%, 2\%, 5\% \)
Overview of results for color-cell: #4 / 12

True: $\langle z \rangle = 0.2849$, $\sigma_{68,95,99.5} = 0.12, 0.24, 0.32$

Initial Guess: $\Delta z = 0.32$

Final: $\Delta z = -2.7e-04$ (-0.1%) $|\delta \sigma_{68,95,99.5}| = 0\%, 2\%, 5\%$

Statistics of the true $dn/dz$
Overview of results for color-cell: #4 / 12

True: $\langle z \rangle = 0.2849$, $\sigma_{68,95,99.5} = 0.12, 0.24, 0.32$

Initial Guess: $\Delta_z = 0.32$

Final: $\Delta_z = -2.7 \times 10^{-4}$ (-0.1%) $|\delta \sigma_{68,95,99.5}| = 0\%, 2\%, 5\%$

Statistics of the true dn/dz

True dn/dz

Initial Random guess [draw from prior]
Overview of results for color-cell: #4 / 12

True: $<z>=0.2849$  $\sigma_{68,95,99.5}=0.12, 0.24, 0.32$

Initial Guess: $\Delta_z=0.32$

Final: $\Delta_z=-2.7e-04$ (-0.1%)  $|\delta\sigma_{68,95,99.5}|=0\%, 2\%, 5\%$

Statistics of the true $dn/dz$

Walking through parameter space

True $dn/dz$

Initial Random guess [draw from prior]
Overview of results for color-cell: #4 / 12

True: \( \langle z \rangle = 0.2849 \) \( \sigma_{68,95,99.5} = 0.12, 0.24, 0.32 \)

Initial Guess: \( \Delta z = 0.32 \)

Final: \( \Delta z = -2.7 \times 10^{-4} \) (-0.1\%) \( |\delta \sigma_{68,95,99.5}| = 0\%, 2\%, 5\% \)

Walking through parameter space

Draw from posterior

Initial Random guess [draw from prior]

Statistics of the true \( dn/dz \)
Overview of results for color-cell: #4 / 12

True: $\langle z \rangle = 0.2849$, $\sigma_{68,95,99.5} = 0.12, 0.24, 0.32$

Initial Guess: $\Delta z = 0.32$

Final: $\Delta z = -2.7e-04$ (-0.1%) $|\delta \sigma_{68,95,99.5}| = 0\%, 2\%, 5\%$

- Walking through parameter space
- Draw from prior
- Draw from posterior
- Initial Random guess
- Statistics of the true $dn/dz$
Overview of results for color-cell: #4 / 12

- True: $<z> = 0.2849, \sigma_{68,95,99.5} = 0.12, 0.24, 0.32$
- Initial Guess: $\Delta z = 0.32$
- Final: $\Delta z = -2.7 \times 10^{-4} (-0.1\%)$  $|\delta \sigma_{68,95,99.5}| = 0\%, 2\%, 5\%$

- Statistics of the true dn/dz
- Accuracy of posterior draw
- Gal-DM bias model truths & posterior
- Walking through parameter space
- Initial Random guess [draw from prior]
- Draw from posterior

Redshift
$\Delta_z$  $\sigma_{68}[\%]$  $\sigma_{95}[\%]$  $\sigma_{99.5}[\%]$  $b_0 [\sigma]$  $b_1 [\sigma]$  $b_k [\sigma]$  $b_0 + b_1 \times \bar{z} [\sigma]$

$-0.0007 \pm 0.0007$  $1.1 \pm 1.8$  $2.5 \pm 2.1$  $6.4 \pm 3.6$  $-5.9 \pm 6.7$  $6.5 \pm 6.2$  $0.0 \pm 0.4$  $2.2 \pm 2.9$
Conclusions

No longer rely on spec-z to validate population redshifts (dn/dz).

Our simple method requires no redshift knowledge to estimate dn/dz.

\[ \sigma(\Delta_z = (\langle z_{true} \rangle - \langle z_{pred} \rangle)) \approx 0.0007 \]

=> meet Euclid & DES requirements.

Prior knowledge can be used as starting points for the MCMC chains.

Codes take a long time to converge, Gibbs sampling in 250D space!

Only population redshifts for galaxy sub-samples.

Adding cosmology is just a few (e.g. 10) extra parameters. CMB TT (etc.)

can also be included at no extra computational cost.

The process must be repeated for all extensions to the cosmological
model explored, or changes to galaxy-dm bias model.