

A Stochastic Process Model of Quasar Spectroscopy

Abstract

We present a method to measure the red-shift of quasars from photometric observations. Our method treats the unknown spectrum of a quasar as a latent stochastic process and uses statistical inference to infer the unknown structure. Our model leverages a small number of existing examples of full quasar spectra with known red-shift to build a structured stochastic process prior distribution over unknown spectra. We then use Bayesian inference to infer red-shift from a typical 5-band photometric sample of astronomical imagery, the so called “photo-z” problem.

1. Introduction

Enormous amounts of astronomical data are collected to [cite sloan digital sky survey](#), [other astronomical surveys](#). Among this collection are measurements of spectral energy distribution (SED) of a light source (e.g. a star, galaxy, quasar). The spectrum of an object carries information about properties of the particular object, including effective temperature, type, red-shift, and chemical makeup.

However, measurements of astronomical spectra are produced by instruments at widely varying resolutions. Spectroscopic measurements can resolve noisy measurements of the spectral energy distribution (SED) of an object (e.g. a star, galaxy, or quasar) in finer detail than broadband photometric measurements. For example, the Baryonic Oscillation Spectroscopic Survey ([Dawson et al., 2013](#)) samples measurements at over four thousand wavelengths between 3,500 and 10,500 Å. In contrast, the photometry from the Sloan Digital Sky Survey (SDSS) [cite sloan](#), gathers broadband photometric measurements in the u,g,r,i, and z bands. These measurements are the weighted average response over a large swath of the spectrum [refer to plot](#).

Photometric measurements, however, are available for a larger number of sources, including objects that are fainter and possibly at extremely high red-shift (farther away). This work focuses on extracting information from observa-

tions of light sources by jointly modeling spectroscopic and photometric data. In particular, we focus on measuring the red-shift of quasars for which we only have photometric observations. Quasars, or quasi-stellar radio sources, are extremely distant and energetic sources of electromagnetic radiation that can exhibit high red-shift [cite something here](#). Identifying and measuring the red-shift of a quasars from photometric data is a necessary task due to the widespread availability of large photometric surveys. Photometric estimates of red-shift have the potential to guide the study of certain quasars with higher resolution instruments. Furthermore, accurate models can aid identification and classification of faintly observed quasars in such a large photometric survey. Study of distant quasars allow astronomers to observe the universe as it was many billions of years ago [need lots of references.](#)

In this paper, we describe a probabilistic model that jointly describes both high resolution spectroscopic data and low resolution photometric observations of quasars in terms of their latent spectral energy distribution, luminosity, and red-shift. We model a quasar’s spectral energy as a hidden variable, and describe a fully Bayesian inference procedure to compute the marginal probability distribution of a quasar’s red-shift given observed photometric fluxes and their uncertainties.

2. Background

The spectral energy distribution of an object describes the distribution of energy it radiates as a function of wavelength. For example, most stars are well modeled as a blackbody, so their spectral radiance is well modeled by Planck’s law, which describes a parametric form for the spectral energy distribution. Resolvable stars, however, tend to be close enough that they are never observed at a measurable redshift. Quasars, on the other hand, have a complicated spectral energy distribution characterized by some salient features ([mention Ly- \$\alpha\$ and Lyman forest?](#)). Furthermore, quasars can be much more luminous and at a much higher red-shift. The red-shift effectively stretches an object’s rest frame spectral energy distribution, skewing its mass toward higher wavelengths. Denoting the rest-frame SED of a quasar n as a function, $f_n^{(rest)} : \mathcal{W} \rightarrow \mathbb{R}_+$, the affect of red-shift on our observations is summarized by the

relationship

$$f_n^{(obs)}(\lambda) \propto f_n^{(rest)}(\lambda \cdot (1 + z_n)). \quad (1)$$

An example of an observed quasar spectrum and its de-redshifted rest frame spectrum can be found in [figure with an example](#)

2.1. Related work

The problem of estimating the red-shift of a source (galaxy or quasar) is known as “photo- z ”. There have been many statistical and machine learning methods developed to tackle this problem. These roughly break down into template-based and regression-based methods (Walcher et al., 2011). Regression-based (or empirical) methods for photo- z fit a functional relationship between a set of photometric features and red-shift value.

Probabilistic models for inferring red-shift from photometry has also had some success. (Benítez, 2000) presents a thorough summary of Bayesian methods for photometric red-shift estimation from spectral templates.

(Budavari et al., 2001) and (Richards et al., 2001) go into specific detail for SED model based photometric redshifts for quasars. They compare a bunch of methods. Particularly, they describe an algorithm for reconstruction a quasar spectrum template from photometric observations and spectroscopic redshifts. It seems sorta like a dynamic K-means/EM algorithm (they add spectral types as needed), and does a decent job reconstructing bumps where the emission lines are.

Others to mention:

- (Bovy et al., 2012) - extreme deconvolution method
- (Suzuki, 2006) (QUASAR SPECTRUM CLASSIFICATION WITH PRINCIPAL COMPONENT ANALYSIS (PCA): EMISSION LINES IN THE Ly α FOREST)
- (Brescia et al., 2013) use a multi-layer perceptron (four layers) regression setting on a combination of SDSS (from the DR7QSO dataset, I believe), UKIDSS, and WISE photometric datasets, comparing photo- z performance on the following intersections:
- (Budavári, 2009) (Unified photo- z paper)

3. Model

This section describes the probabilistic model for spectroscopic and photometric observations.

3.1. Stochastic Process Model of Spectra

The SED of a quasar is a nonnegative valued function of wavelength. We model it in rest-frame as a linear combination of a set of basis functions. We place a log-Gaussian process prior on each of these basis functions. The generative process for quasar spectra is

$$\beta_k(\cdot) \sim \mathcal{GP}(0, K_\theta) \quad (2)$$

$$B_k(\cdot) = \frac{\exp(\beta_k(\cdot))}{\int_{\mathcal{W}} \exp(\beta_k)} \quad (3)$$

$$\mathbf{w} \sim p(\mathbf{w}), \text{ s.t. } \sum_{w_k} = 1 \quad (4)$$

$$f_n^{(rest)} = \sum_k w_k B_k \quad (\text{SED}) \quad (5)$$

$$\tilde{f}_n^{(rest)} = \ell_n \sum_k w_k B_k \quad (6)$$

Warp input for varying lengthscale.

We view the spectrum of a quasar as a (scaled) probability distribution itself, and thus model it as a random measure. Specifically, we specify a model for the *rest frame* spectrum of a quasar, and explain observations by warping the input measurements by the appropriate red-shift values.

As the spectrum of a source characterizes much of the information of interest, we take special care creating a structured prior over $f_n(\cdot)$ using available observations of rich spectra. Furthermore, due to the fact that the red-shift value effectively scales the input of this function, we cannot rely on a fixed grid of $\lambda_1, \dots, \lambda_P$ values, but must define a stochastic process to coherently define a probabilistic model. The rest of this section outlines a model for these spectra and subsequent photometric observations.

4. Experiments

5. Discussion

References

- Ball, Nicholas M, Brunner, Robert J, Myers, Adam D, Strand, Natalie E, Alberts, Stacey L, and Tchong, David. Robust machine learning applied to astronomical data sets. iii. probabilistic photometric redshifts for galaxies and quasars in the sdss and galex. *The Astrophysical Journal*, 683(1):12, 2008a.
- Ball, Nicholas M, Loveday, Jon, and Brunner, Robert J. Galaxy colour, morphology and environment in the sloan digital sky survey. *Monthly Notices of the Royal Astronomical Society*, 383(3):907–922, 2008b.

220	Benitez, Narciso. Bayesian photometric redshift estimation. <i>The Astrophysical Journal</i> , 536(2):571, 2000.	275
221		276
222		277
223	Berk, Daniel E Vanden, Richards, Gordon T, Bauer, Amanda, Strauss, Michael A, Schneider, Donald P, Heckman, Timothy M, York, Donald G, Hall, Patrick B, Fan, Xiaohui, Knapp, GR, et al. Composite quasar spectra from the sloan digital sky survey. <i>The Astronomical Journal</i> , 122(2):549, 2001.	278
224		279
225		280
226		281
227		282
228		283
229		284
230	Bovy, Jo, Myers, Adam D, Hennawi, Joseph F, Hogg, David W, McMahon, Richard G, Schiminovich, David, Sheldon, Erin S, Brinkmann, Jon, Schneider, Donald P, and Weaver, Benjamin A. Photometric redshifts and quasar probabilities from a single, data-driven generative model. <i>The astrophysical journal</i> , 749(1):41, 2012.	285
231		286
232		287
233		288
234		289
235		290
236		291
237	Brescia, M, Caviuoti, S, D'Abrusco, R, Longo, G, and Mercurio, A. Photometric redshifts for quasars in multi-band surveys. <i>The Astrophysical Journal</i> , 772(2):140, 2013.	292
238		293
239		294
240		295
241	Budavári, Tamás. A unified framework for photometric redshifts. <i>The Astrophysical Journal</i> , 695(1):747, 2009.	296
242		297
243		298
244	Budavari, Tamas, Csabai, Istvan, Szalay, Alexander S, Connolly, Andrew J, Szokoly, Gyula P, Berk, Daniel E Vanden, Richards, Gordon T, Weinstein, Michael A, Schneider, Donald P, Benitez, Narciso, et al. Photometric redshifts from reconstructed quasar templates. <i>The Astronomical Journal</i> , 122(3):1163, 2001.	299
245		300
246		301
247		302
248		303
249		304
250		305
251	Dawson, Kyle S, Schlegel, David J, Ahn, Christopher P, Anderson, Scott F, Aubourg, Éric, Bailey, Stephen, Barkhouser, Robert H, Bautista, Julian E, Beifiori, Alessandra, Berlind, Andreas A, et al. The baryon oscillation spectroscopic survey of sdss-iii. <i>The Astronomical Journal</i> , 145(1):10, 2013.	306
252		307
253		308
254		309
255		310
256		311
257		312
258	Kind, Matias Carrasco and Brunner, Robert J. Somz: photometric redshift pdfs with self-organizing maps and random atlas. <i>Monthly Notices of the Royal Astronomical Society</i> , 438(4):3409–3421, 2014.	313
259		314
260		315
261		316
262		317
263	Myers, Adam D, White, Martin, and Ball, Nicholas M. Incorporating photometric redshift probability density information into real-space clustering measurements. <i>Monthly Notices of the Royal Astronomical Society</i> , 399(4):2279–2287, 2009.	318
264		319
265		320
266		321
267		322
268		323
269	Richards, Gordon T, Weinstein, Michael A, Schneider, Donald P, Fan, Xiaohui, Strauss, Michael A, Berk, Daniel E Vanden, Annis, James, Burles, Scott, Laubacher, Emily M, York, Donald G, et al. Photometric redshifts of quasars. <i>The Astronomical Journal</i> , 122(3):1151, 2001.	324
270		325
271		326
272		327
273		328
274		329