A Stochastic Process Model of Quasar Spectroscopy

Andrew Miller ACM@SEAS.HARVARD.EDU
Albert Wu AWU@COLLEGE.HARVARD.EDU
Ryan Adams RPA@SEAS.HARVARD.EDU
Harvard University, 33 Oxford St, Cambridge, MA, 02138 USA

David Schlegel DJSCHLEGEL@LBL.GOV
Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA, 94720 USA

Abstract

We describe a joint model of two disparate sources of astronomical data, spectroscopy and photometry, which carry information about sources light at extremely different resolutions. Our model treats the energy distribution of a source's radiation as a latent variable, which hierarchically generates both photometric and spectroscopic observations. We place a structured stochastic process prior over the spectral energy distribution of a light source (e.g. star, galaxy, or quasar) that admits a physically interpretable decomposition and allows us to tractably perform fully Bayesian inference using Markov Chain Monte Carlo (MCMC). We use our model to compute the distribution of red-shift of quasars from five-band (low-resolution) photometric data, the so called photometric red-shift estimation ("photo-z") problem. Our method shows that tools from machine learning and Bayesian statistics can aid in the joint modeling of

Our method leverages a small number of existing examples of high resolution quasar spectra with known red-shift to build a structured prior distribution over unknown spectra.

1. Introduction

Enormous amounts of diverse astronomical data are cite sloan digital sky survey, other astronomical surveys. Among this collection are measurements of spectral energy distribution (WED) of a light source (e.g. a star, galaxy, quasar). The spectrum of a source 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 (BOSS) (**?**) 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) (**?**), 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. The two methods of spectral information collection are compared in Figure 2.

Photometric measurements, however, are much more widely available and exist for a larger number of sources, including objects that are fainter and possibly at extremely high redshift (farther away). This work focuses on extracting information from observations of light sources by jointly modeling spectroscopic and photometric data. In particular, we 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...

Proceedings of the 31st International Conference on Machine Learning, Lille, France, 2015. JMLR: W&CP volume 37. Copyright 2015 by the author(s).

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 latent 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 SED of an object describes the distribution of energy it radiates as a function of wavelength. For example, most stars are well-modeled as blackbodies, so their spectral radiance closely follows 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 non-zero red-shift. Quasars, on the other hand, have a complicated spectral energy distribution characterized by some salient features (mention Ly-α and Lyman forest?). Furthermore, quasars can be much more luminous and at a much higher red-shift than stars and galaxies. The effect of red-shift is the stretching of the input space of wavelengths, $\lambda \in \Lambda$, of an object's *rest-frame* SED toward higher values, skewing its mass toward higher (redder) wavelengths. Denoting the rest-frame SED of a quasar n as a function, $f_n^{(\text{rest})} : \Lambda \rightarrow \mathbb{R}_+$, the effect of red-shift on our observations is summarized by the relationship

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

Observed quasar spectra and their "de-red-shifted" rest frame spectra are depicted in Figure 1.

2.1. Related work

The problem of estimating the red-shift of a source (galaxy or quasar) is known as photometric red-shift estimation, or "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. Template-based methods use prior knowledge about spectral energy distributions and match these with observed quasar spectra (?). 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 have also had some success. (?) presents a thorough summary of Bayesian methods for photometric red-shift estimation from spectral templates.

(?) and (?) go into specific detail on SED model-based photometric redshift estimation for quasars. They compare many methods. Particularly, they describe an algorithm for reconstruction a quasar spectrum template from photometric observations and spectroscopic redshifts. The method is very similar to a dynamic K-means or expectation-maximization algorithm, which spectral types added as needed, and does a decent job reconstructing the locations of emissions lines in the spectrum.

(?) use a multi-layer perceptron with a combination of SDSS, UKIDSS, and WISE photometric fluxes datasets, to compute a regression function for red-shift. Though predictions are accurate, we do not have an idea of the distribution over redshifts, i.e., uncertainty in our predictions.

Other models blur the line between regression-based and generative models. (?) develops a method termed $XDQSOz$ to use a large dataset of astronomical objects to simultaneously infer redshift and classify quasars. They do this by inferring a joint distribution over object type, fluxes, and redshift. In particular, they factor this joint distribution into one part that describes the distribution over star brightness, which involves binning based on a single-channel flux, and one part that describes the distribution of relative fluxes (as compared to this channel) and the red-shift. The latter distribution is represented by a mixture of up to 60 Gaussians. However, we see that, due to the binning, this approach is not fully probabilistic or Bayesian.

(?) unifies template-based and regression-based approaches into a single probabilistic framework, distinguishing methods based on the assumptions they impose on probability distributions over photometric fluxes. Instead of treating the spectral distribution and physical properties as disparate, the authors represent physical properties as a random variable whose conditional distributions over observations such as spectra can be computed.

Others to mention: (?) does unsupervised learning on the UV spectra of quasars using principal components analysis (PCA). Through this analysis, the authors found that 96% of the total variance was accounted for by the first 3 spectral components. As a result, they created a classification scheme based on the first two component's coefficients that separates quasars into five different classes. This classification scheme allows researchers to understand quasars better from a qualitative perspective.

This doesn't seem that relevant. What do you think, Andy?

3. Model

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

3.1. Stochastic Process Model of Spectra

The SED of a quasar is a nonnegative valued function $f : \Lambda \rightarrow \mathbb{R}^+$, where Λ denotes the range of wavelengths

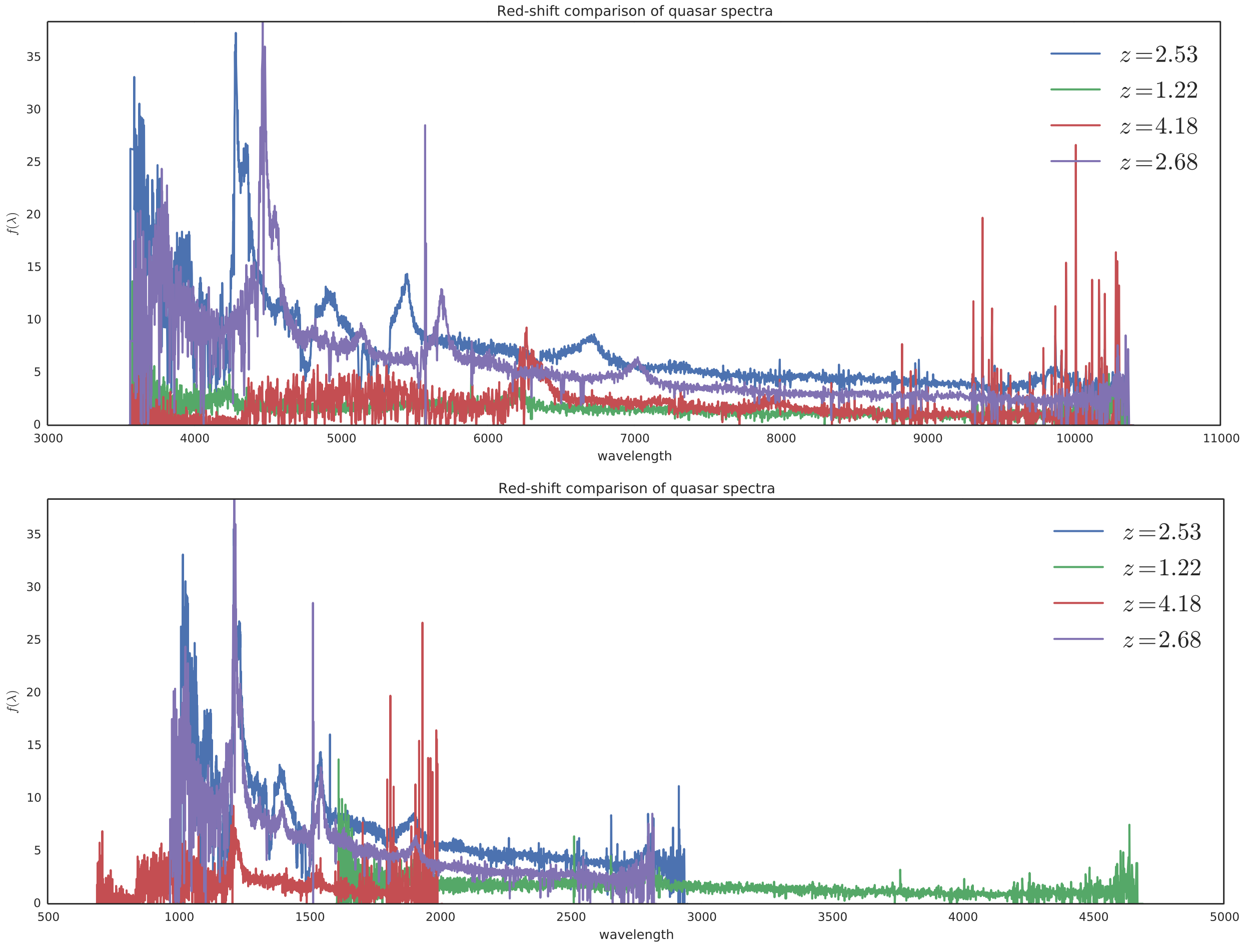

Figure 1. High fidelity spectroscopy of multiple quasars at different red-shifts, z. The top graphic depicts the spectrograph in the observation frame, which can be intuitively thought of as "stretched" by a factor $(1+z)$. The lower figure depicts the "de-redshifted" version of the same quasar spectra. This effectively squashes observations, and reconstructs the quasar spectra as it would be seen in the quasar's rest frame. The salient feature of this operation is the alignment of the emission and absorption lines (albeit at different scales). This alignment will provide the information about z that we can infer from SDSS photometric projections of the very same spectra.

and $\mathbb{R}^+$ are nonnegative real-valued numbers. However, the types of spectra we observe from quasars is highly structured, and we use an additive decomposition of a small number of positive bases (or templates) to capture this structure. We model a quasar's SED *in rest-frame* as a positive linear combination of a set of positive basis functions. Interpreting the spectra as a scaled probability distribution, we model it as a random measure. We place a log-Gaussian process prior on each of these basis functions, and a prior over positive weight values for each quasar. Todo: include brief GP background

The generative procedure for quasar spectra is first generate a shared positive basis

$$\beta_k(\cdot) \sim \mathcal{GP}(0, K_\theta)\,, k = 1, \dots, K \tag{2}$$

$$B_k(\cdot) = \frac{\exp(\beta_k(\cdot))}{\int_\Lambda \exp(\beta_k(d\lambda))} \tag{3}$$

and for each quasar, n

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

$$m_n \sim p(m) \text{ s.t. } m_n > 0 \tag{5}$$

$$f_n^{(\text{rest})}(\cdot) = \sum_k w_k B_k(\cdot) \qquad \text{(SED)} \tag{6}$$

$$\tilde{f}_n^{(\text{rest})}(\cdot) = m_n \sum_k w_k B_k(\cdot) \tag{7}$$

$$z_n \sim p(z) \qquad \text{(red-shift)} \tag{8}$$

where $p(\mathbf{w})$ is a prior over weights that sum to one, and $p(z)$ is a prior over red-shifts.

Each positive basis function, B_k is normalized to integrate to one, and each quasar's weight vector $\mathbf{w}_n$ also sums to one. This allows us to interpret the $f_n^{(\text{rest})}(\cdot)$ function as a density, scaled by m_n. Todo: Warp input for varying lengthscale.

For each quasar n, we observe noisy samples of the red-shifted and scaled spectral energy distribution at a grid of wavelengths $\lambda \in \{\lambda_1, \dots, \lambda_P\}$. Our *observation frame* samples are modeled

$$x_{n,\lambda} \sim \mathcal{N}\left(\frac{1}{(1+z_n)}\tilde{f}_n^{(\text{rest})}(\lambda \cdot (1+z_n)), \sigma^2_{n,\lambda}\right) \tag{9}$$

where $\sigma^2_{n,\lambda}$ is known measurement variance from the instruments used to make the observations. The BOSS spectra (and our rest-frame basis) are stored in units ergs/cm^2/s/Å.

3.2. Photometric Observations

Photometric observations summarize the number of photon observations over a large swatch of the wavelength spectrum. Roughly, a photometric flux measures the number of photons hitting the instrument's lens, filtered by a band-specific sensitivity curve, over the duration of an exposure. We will express measurements of flux in nanomaggies reference, a linear unit of flux, in our method and experiments. The photometric data are measured in the SDSS $ugriz$ bands, giving us a vector, $\mathbf{y}_n$, of five flux values and their variances, $\tau_{n,b}$ for $b \in ugriz$, which are computed from images. PSFFLUX - those are straight from pixels???. Each band measures photon observations at each wavelength in proportion to a known filter sensitivity, $S_{band}(\lambda)$. The filter sensitivities for the SDSS $ugriz$ bands are depicted in Figure 2, with an example (observation frame) quasar spectrum overlaid. The actual measured fluxes can be computed by integrating the full object's spectrum, $m_n \cdot f_n(\lambda)$ against the filters. For a band $b \in \{u, g, r, i, z\}$

$$\mu_b = \int f_n^{(\text{obs})}(\lambda) S_b(\lambda) C(\lambda) d\lambda \tag{10}$$

$$\equiv I_b(f_n^{(\text{rest})}, z_n) \tag{11}$$

where $C(\lambda)$ is a conversion factor to go from the units of $f_n(\lambda)$ to nanomaggies. This projection onto SDSS bands results in five fluxes, which are modeled as independent Gaussian random variables with known variance (derived from a pixel to flux inference procedure)

$$y_{n,b} \sim \mathcal{N}(I_b(f_n^{(\text{rest})}, z_n), \tau_{n,b}) \tag{12}$$

Given a low dimensional summary of $f_n^{(\text{rest})}$, we can rewrite I as a function of those parameters.

3.3. Joint model

Given a sample of M noisy full spectra and their sample locations, $\{\mathbf{x}_m, \lambda_m^{(\text{obs})}\}_{m=1}^M$, and a set of N photometric fluxes, $\{\mathbf{y}_n\}_{n=1}^N$, our full likelihood in terms of $\{\mathbf{w}_m\}_{m=1}^M, B_1, \dots, B_K, \{w_n\}_{n=1}^N$, and $\{z_m\}, \{z_n\}$ is

$$\begin{aligned} L(\{\mathbf{w}_m, z_m\}, &\{B_k\}, \{\mathbf{w}_n, z_n\}) \\ &= \prod_{m=1}^M p(\mathbf{x}_m | \mathbf{w}_m, z_m, \{B_k\}) \\ &\times \prod_{n=1}^N p(\mathbf{y}_n | \mathbf{w}_m, z_m, \{B_k\}) \end{aligned}$$

where the probability distribution of the first term in the product is given by 9, and the distribution for the second term in the product is given by 12.

We express the joint prior distribution over the quasar weights $\mathbf{w}_n$ and $\mathbf{w}_m$ and basis

$$p(\{\mathbf{w}_m, z_m\}, \{B_k\}, \{\mathbf{w}_n, z_n\}) \tag{13}$$

$$= p(\{B_k\}) p(\{\mathbf{w}_m, z_m\}) \tag{14}$$

$$p(\{\mathbf{w}_n, z_n\} | \{\mathbf{w}_m, z_m\}) \tag{15}$$

where we condition the photometric weights on the spectroscopically fit weights.

4. Inference

The "photo-z" tasks requires that we compute posterior marginal distributions of z, $\mathbf{w}$ and m. To simplify notation, let $\mathbf{X} = \{\mathbf{x}_m, \lambda_m^{(obs)}\}_{m=1}^M$ and $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)^\intercal$. For example, we wish to compute the posterior marginal distribution for

$$p(z_n | \mathbf{y}_n, \mathbf{X}) = \int p(z_n, \mathbf{w}_n, B | \mathbf{y}_n, \mathbf{X}) d\mathbf{w}_n dB \tag{16}$$

$$= \int p(z_n, \mathbf{w}_n | \mathbf{y}_n, B) p(B | \mathbf{X}) d\mathbf{w}_n dB \tag{17}$$

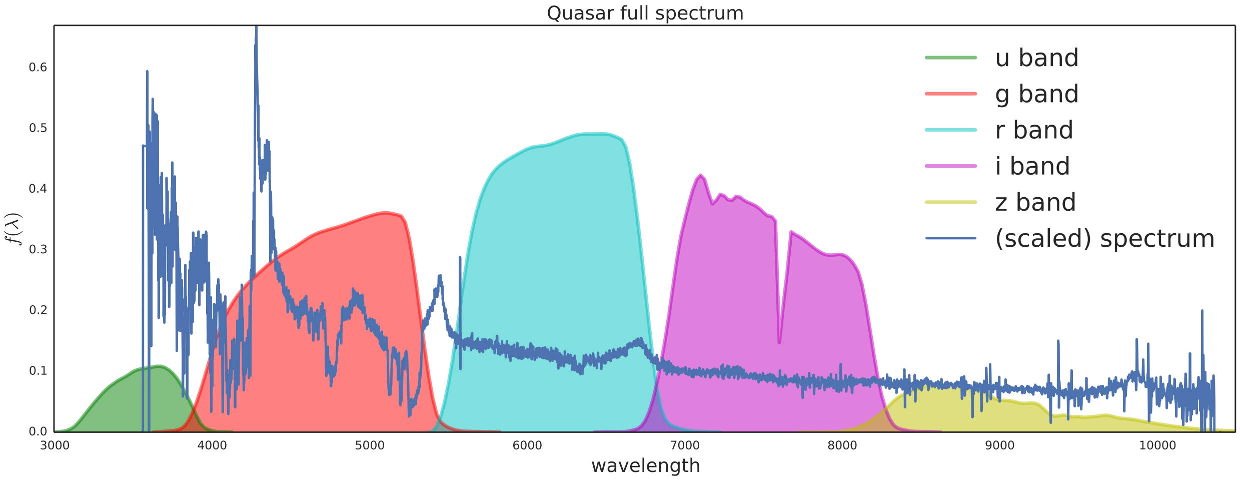

Figure 2. Example of a (scaled) quasar spectrum with SDSS *ugriz* band filters, $S_\beta(\lambda)$, overlaid. Computing red-shift given the full spectrum is often straightforward and often quite precise, however full spectra are more difficult to obtain. Photometry data, while far easier to obtain, only measures band specific flux, a weighted average of of the unobserved spectrum of the object.

This section outlines our MCMC procedure to compute posterior samples of $\mathbf{w}_n, z_n, m_n$ and B given the sample of spectra, $\mathbf{X}$, and photometric fluxes $\mathbf{y}_{ugriz}$.

4.1. Sampling B

To accelerate computation, we use only information present in $\mathbf{X}$ to draw samples of $B_1, \ldots, B_K$. That is, we approximate the full conditional distribution

$$p(B_1, \ldots, B_k | \mathbf{X}, \mathbf{Y}) \approx p(B_1, \ldots, B_k | \mathbf{X}) \tag{18}$$

We expect this to have little effect on the distribution, as the amount of information about B present in $\mathbf{X}$ is expected to dwarf that of $\mathbf{Y}$.

To generate samples from the posterior distribution $p(B|\mathbf{X})$, we use elliptical slice sampling. add some details about elliptical slice sampling.

$$\begin{aligned} p(B|\mathbf{X}) &= L(\beta; \mathbf{X}) p(\beta) && (19) \\ &= L(\beta; \mathbf{X}) \prod_k \mathbf{N}(\beta_k | 0, K_{\theta_k}) && (20) \end{aligned}$$

4.2. Sampling $\mathbf{w}_n, m_n$, and z_n

Conditioned on a basis $B_k, k = 1, \ldots, K$, we can draw posterior samples of $\mathbf{w}_n$ and z_n independently for each n

$$\begin{aligned} & p(\mathbf{w}_n, m_n, z_n | B, \mathbf{y}_n) && (21) \\ & \propto p(\mathbf{y}_n | \mathbf{w}_n, m_n, z_n, B) p(\mathbf{w}_n, m_n, z_n) && (22) \\ & = p(\mathbf{y}_n | \mathbf{w}_n, m_n, z_n, B) p(\mathbf{w}_n) p(m_n, z_n) && (23) \end{aligned}$$

We assume that $\mathbf{w}_n$ is independent of m_n, z_n. Indeed, our prior for $\mathbf{w}_n$ will introduce $\gamma_n \sim \mathcal{N}(0, \mathbb{I})$ and then a softmax function:

$$w_{ni} = \frac{\exp(\gamma_{ni})}{\sum_i \exp(\gamma_{ni})},$$

which enforces non negativity and the fact that $\sum \mathbf{w}_n = 1$.

5. Experiments

We use the DR10QSO dataset cite dr10, which includes spectroscopically confirmed red-shifts from over 150,000 quasar spectra.

To compute samples of our basis, we include information from one thousand randomly sampled spectra.

Experiments/model output

- z_{spec} vs. z_{phot} plot
- Individual marginals of z_n, w_n
- Clustering of w_n (BAL vs. Non BAL?)
- Comparison w/ or without prior over w_n?

6. Discussion