A study of Quasar Colors using the SDSS

Poruri Sai Rahul

July 31, 2013

Contents

List of Figures

List of Tables

Abstract

The aim of the project is to understand and replicate the results of the paper by Richards et al. (2001a) titled "Colors of 2625 quasars at $0 \leq z \leq 5$ measured in the SDSS photometric system". To summarize the contents of the paper, the authors calculate the u-g, g-r, r-i and i-z colors of 2625 quasars and plot these colors as a function of redshift. Considerable structure can be seen in the color-redshift relation of these quasars and the reasons as to why this structure arises have been explained. As part of my project, I've only been able to partially replicate the results of this paper but I've been successful in extending the results of this paper thanks to a much larger sample set of quasars I've acquired from the SDSS DR9. The results of the original paper have been extended using the data pertaining to 146,659 quasars - in contrast to the 2625 quasars in the original data set, that were acquired via SQL query of the DR9. In the process, interesting features have been found in the color-redshift plot for the samples in the new data set and these outliers are being studied further.

This report has been broken down into of three sections -

- An introduction to Quasars and The Sloan Digital Sky Survey.
- Data Acquisition, Analysis & Results
- Further work

1 An introduction to Quasars and The Sloan Digital Sky Survey

1.1 An Introduction to Quasars

Quasars, short for Quasi-Stellar Radio Sources were first observed by radio telescopes thanks to their variability in the radio spectrum. Follow up observations by visible, IR and UV telescopes showed that these objects were much more luminous in the UV than in the Radio. Follow up surveys turned up quite a few objects with properties similar to quasars but these objects were radio quiet unlike the first ones discovered. Henceforth, the objects were categorized simply as Quasi Stellar Objects - QSOs. The name QSOs is synonymous with quasars now-a-days. QSOs belong to a much bigger class of objects known as Active Galactic Nuclei - AGN. The other types of AGN are Seyfert galaxies, Broad Absorption Line (BAL) galaxies, Compact Emission Line (CEL) galaxies, Blazar & BL Lac objects.

Quasars are tremendously luminous across the EM spectrum and the light of a pulsar has implications on the formation of galaxies and on the thermal history of the universe. The size of a quasar is comparable to the size of our solar system - this compact size is what leads us to believe that a quasar is a massive black hole and the energetic emissions are accredited to surrounding mass being accreted, compressed and accelerated. The accreted mass slowly falls into the black hole thanks to viscosity and get heated up in the process due to the shocks. These bursts of energy are short lived and if this rate was static, the host galaxy would be consumed over the course of the quasar's evolution as the rate of mass accretion is in terms of M_o per year. In reality, the rate of consumption slows down and quasars are rare today - the histogram of quasar redshift observed shows this clearly. Further evidence for accretion disks is the fact that molecular masers have been observed from the cores of certain active galaxies. It is inferred that the accretion disks are dense enough to shield gas from the radiation of the quasar. Had there not been an accretion disk, the molecules would've disassociated thanks to the intense radiation from the quasar.

The study of evolution of a quasar is done by looking at various quasars at different epochs.

1.2 The Sloan Digital Sky Survey

The SDSS, short for Sloan Digital Sky Survey is carried out under the guidance of the Astrophysical Research Consortium, funded by the Alfred P. Sloan Foundation. It is carried out from the Apache Point Observatory, New Mexico situated at 32^{deg} N latitude. The SDSS has gone through 9 data releases and is currently guided by the SDSS - III collaboration. The surveys currently being carried out by the SDSS are -

- APOGEE, the Apache Point Observatory Galactic Evolution Experiment,

- BOSS, the Baryon Oscillation Spectroscopic Survey,
- MARVELS, the Multi-Object APO Radial Velocity Exoplanet Large-area Survey,
- SEGUE-2, the Sloan Extension for Galactic Understanding and Exploration-2

Coming to the instruments used for the survey, the SDSS uses a 2.5-meter telescope at APO with a 120 Mega pixel camera and a fiber fed spectrograph capable of measuring the spectra of 640 objects, simultaneously.

The primary camera is actually made up of 30 smaller CCDs, grouped in a 5x6 matrix. These 5 rows with 6 CCDs are each in different colors i.e these 5 rows are covered with the 5 filters that SDSS uses for photometry which are - u, g, r, i, z. The effective wavelengths of these filters are [1]

- u - 3590 A°
- g - 4810 A°
- r - 6230 A°
- i - 7640 A°
- z - 9060 A°

The SDSS scanning method vertical i.e the objects move from the upper row to the lower row. To summarise the scan method, a single portion of the sky is scanned in 5 different filters, as the region passes from one filter to the other. The imaging survey is taken in drift-scan mode - time delay and integrate or TDI, where the telescope doesn't track the objects. Instead the earth's intrinsic rotation moves the objects across the CCD and data is acquired continuously. The effective integration time per filter is 54.1 seconds. For further information regarding the SDSS, you can refer to the the technical summary of the SDSS survey [2].

1.3 Quasar Properties

The spectrum of a quasar is characteristic of a gas that has disintegrated into e^{-}, p^{+}, n° and of emission from hot dust. A Quasar's spectrum can be modelled as a negative power law spectrum i.e

$$f_{\nu} = \nu^{\alpha_{\nu}} \tag{1.1}$$

and

$$\alpha_{\lambda} = -(2 + \alpha_{\nu}) \tag{1.2}$$

[1] from for Fukugita et al. 1996 and Gunn et al. 1998.

[2] www.sdss.org/dr7/instruments/technicalPaper/index.html

By creating a composite quasar spectrum using 2200 quasars observed by the SDSS, α_ν is estimated to be -0.5 $\pm$ 0.65 (95% confidence)[3]. In reality, a quasar spectrum has absorption and emission features on top of this pure power law spectrum. If the spectrum of a quasar were just a power law, we wouldn't observe any of the characteristic variations in the color-redshift plots shown above. It is because of these emission and absorption lines on the power law continuum that the color varies with respect to redshift. The standard set of lines that the SDSS searches for in a spectrum while identifying quasars are.[4], [5]

[3]Vanden Berk et al. 2001

[4]http://www.sdss.org/dr2/algorithms/speclinefits.html

[5]http://www.sdss.org/dr6/algorithms/linestable.html

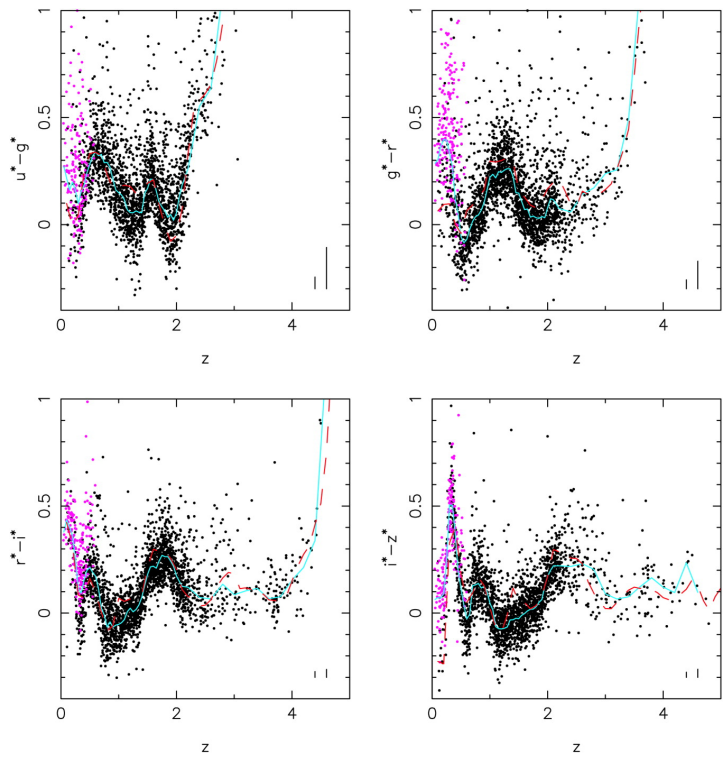

Figure 1: The color-redshift plots from the paper Richards et al. (2001a).

For a thorough description of which specific emission or absorption lines lead to the specific features on the color-redshift diagram, refer to the Section 4.3 - Empirical Colours from the Richards et al. (2001a).

To summarize this section, the color u'-g' decreases i.e becomes more blue if an emission line is present in the u band. With increasing redshift, this line moves into the g band, at which point the value u'-g' increases i.e becomes more red. This process is continuous as there are multiple emission and absorption lines that come into and leave the bands. Towards higher redshift we see a continuous increase. This is because of the fact that at these high redshifts, the lyman α leaves the u' band and the Gunn-Peterson Trough[6] moves in. Going further, especially in the u'-g' color-redshift plot, the color decreases again and we see a plateau like feature beyond z=5 - These two features are much more obvious in my analysis of the data, where the redshift interval is from 0-5 instead of 0-4 like in the case of the paper. It is speculated that at these high red shifts, the lymanα line moves out of the g band and we begin to see emission lines in the gunn-peterson trough in the u band.

The emission lines that significantly affect the broad band colors include the Lyman α, CIV, MgII, Hα, Hβ and the λ3000 bump.

6

λ_{vac}	$weight_{galaxy}$	$weight_{quasar}$	species
1033.30	0.0	1.0	O VI
1215.24	0.0	9.0	Lyα
1239.42	0.0	3.0	N V
1305.53	0.0	0.0	O I
1335.52	0.0	0.0	C II
1397.61	0.0	0.0	Si IV
1399.8	0.0	1.0	Si IV + O IV
1545.86	0.0	8.0	C IV
1637.85	0.0	0.0	He II
1665.85	0.0	0.0	O III
1857.4	0.0	0.0	Al III
1908.27	0.0	7.0	C III
2326.0	0.0	0.5	C II
2439.5	0.0	0.0	Ne IV
2800.32	1.0	8.0	Mg II
3346.79	0.0	0.0	Ne V
3426.85	0.0	0.0	Ne VI
3728.30	5.0	1.0	O II
3889.0	0.0	0.0	He I
3971.19	0.0	0.0	Hζ
4072.3	0.0	0.0	S II
4102.89	0.5	2.0	Hδ
4341.68	1.0	3.0	Hγ
4364.436	0.0	0.0	O III
4862.68	2.0	4.0	Hβ
4932.603	0.0	0.0	O III
4960.295	2.0	2.0	O III
5008.240	3.0	2.0	O III
6302.046	0.0	0.0	O I
6365.536	0.0	0.0	O I
6529.03	0.0	0.0	N I
6549.86	3.0	0.0	N II
6564.61	8.0	8.0	Hα
6585.27	3.0	0.0	N II
6718.29	3.0	0.0	S II
6732.67	3.0	0.0	S II

Table 1: Emission Lines the SDSS data pipeline looks for in a spectra. As per convention O III refers to doubly ionized Oxygen and so on. Regarding the weights mentioned, those with positive weights are used to estimate the emission line redshifts for Galaxies or QSOs.

λ_{vac}	$weight_{galaxy}$	$weight_{quasar}$	species
3934.777	-1.0	0.0	K
3969.588	-1.0	0.0	H
4305.61	-1.0	0.0	G
5176.7	-1.0	0.0	Mg
5895.6	-1.0	0.0	Na

Table 2: Absorption Lines

λ_{vac}	$weight_{galaxy}$	$weight_{quasar}$	species
5578.5	0.0	0.0	Sky
5894.6	0.0	0.0	Sky
6301.7	0.0	0.0	Sky
7246.0	0.0	0.0	Sky

Table 3: Sky Lines

2 Data Acquisition, Analysis & Results

2.1 Reproducing the results

Results from the paper Richards G.T. et al (2000) have been partially reproduced and extended using the SDSS DR9 quasar sample set. A Quasar sample set has been queried and acquired from the DR9 of SDSS catalog - samples which conform to the original limits on AB magnitude limits set by the above mentioned paper. Specifically, the magnitude limits for the different filters are

- u - 22.3
- g - 22.6
- r - 22.7
- i - 22.4
- z - 20.5

Reproduction of results from the original paper was only partial as the full 2625 quasar sample data set could not be acquired. Partial sample set with data corresponding to 898 quasars has been found in NED and SIMBAD, thanks to the links mentioned in ADS by the authors. Data pertaining to the rest of the quasars is speculated to be present on the Early Data Release of the SDSS but this data isn't easily accessible using the SQL Query, as the query options are very primitive and limited. This speculation regarding the whereabouts of the rest of the data set arise from the fact that none of

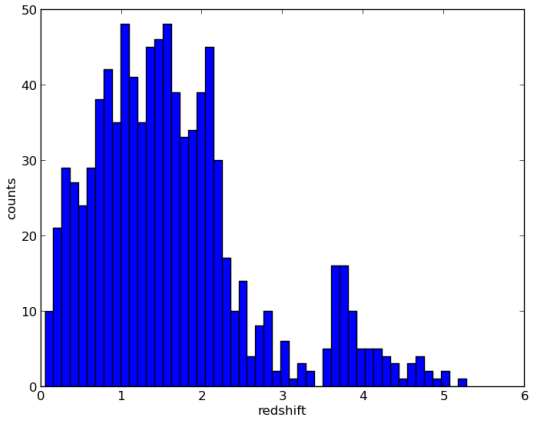

Figure 2: A histogram of the quasar redshift. It is clearly seen that the number of quasar dies off sharply at z=2, with a small peak in the number at z=4

the quasars on the data set available for download are named SDSS - though a few are named SDSSp. It is speculated that the authors chose to not add data pertaining to the new quasars discovered by the SDSS - which could've been acquired through DR1. The quasars for which data is made available for download are quasars discovered by earlier quasars surveys such as the LBQS - Large Bright Quasar Survey, 2MASS - the 2 Micron All Sky Survey, the VAL FIRST survey etc.

Therefore, the color vs redshift plots shown here pertain only to the data for these 898 quasars and so are the median-color vs redshift plot.

The method to calculate the median color, as mentioned in the paper[7] is to bin the color-red shift plot at 0.05 intervals on the redshift axis where the width of each bin is -

- 0.075 for z $\leq$ 2.15
- 0.2 for 2.15 $\leq z \leq$ 2.5
- 0.5 for $z >$ 2.5

In practice, the median color has been calculated using a bin width of 0.075 over the full whole redshift range! This step, taken to make the calculation easier, has led to some interesting features in the median-color vs redshift plot. These features and the reason behind these features are described further. This simplification has also led to understand why the authors chose to vary the bin size with increasing redshift.

[7]Photometric redshifts of quasars, Richards et al 2001

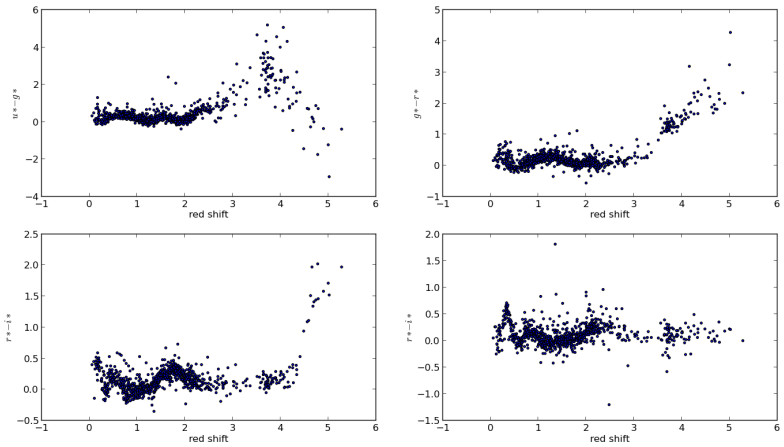

Figure 3: 4 color vs redshift plots. Anomalous increase in the u-g color after z=4 is speculated to be due to the *Lymanα* line having left the g band and the Gunn-Peterson trough's presence. The red shift dependent trend of the 4 colors is discussed further.

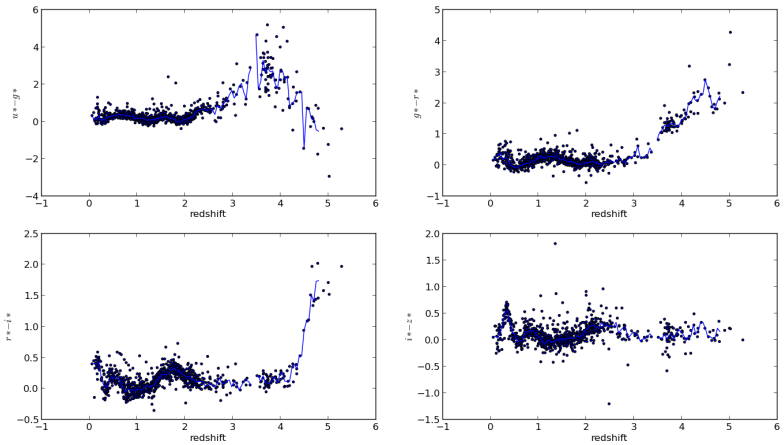

Figure 4: The color-redshift plots with the median color relation in light blue. The large fluctuations in the median color beyond z=3, especially u-g median color is a consequence of the bin size chosen to be constant. Further explanation follows.

The fluctuations in the median color, especially the u-g median color are very large beyond $z = 3$ as is obvious from the median color vs redshift plot. It is speculated that the reason for these fluctuations is the fixed bin width of 0.075 used while binning and calculating the median color over the whole redshift interval. As less number of quasars are available beyond $z = 3$, as is obvious from the histogram, the number of quasars that fall into these redshift bins reduces thereby effecting the smoothness of these median color - redshift relation. This is also the reason, I believe, the authors choose a varying bin width with respect to increasing redshift *i.e* in order to account for the reducing number of quasars at higher redshifts. On the other hand, how the authors arrived upon the exact values of $0.075, 0.2$ and 0.5 is still puzzling. Also, the apparent dip in the u-g color beyond $z = 4$ is of interest as no particular reason exists that would lead to a decrease in the u-g color once the lyman α lines crosses the u band into the g band. It can be speculated that by this point on the redshift axis, the lymanα line has crossed the g band into the r band and the appearance of emission lines in the gunn-peterson trough are behind this apparent decrease in the u-g color.

2.2 Extending the results

Moving further, as mentioned earlier, 146,659 quasars have been obtained from the DR9 of SDSS catalog, all of which conform with the AB magnitude limits mentioned above. SQL query has been used to acquire this quasar sample set. The 4 Color - redshift plots have been made and calculating the median color is due process. In the meanwhile, interesting features have been found in these color-redshift plots.

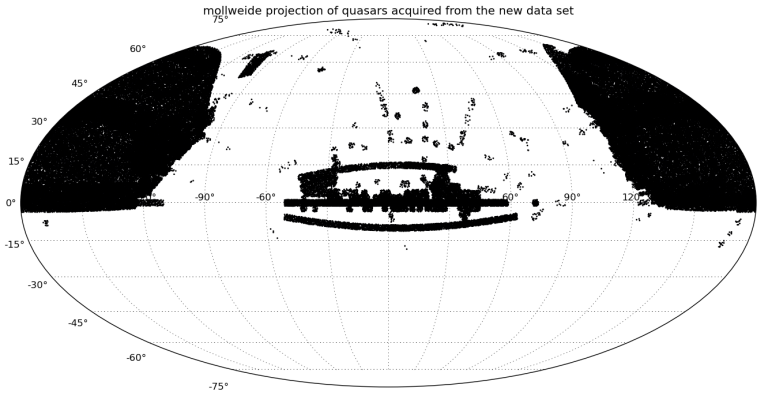

Figure 5: A mollweide projection of the new quasars in RA & Dec. The Apache Point Observatory, from where the SDSS is pursued is at 32 deg N latitude. An estimated declination range of ± 30 is obvious from the projection.

As will be obvious from the following 4 color-redshift figures, it is quite interesting that many a quasar can be found in the vertical features and these quasars belong to one of two specific locations in the sky! It is also interesting that most of these quasars have 6 specific plate ID - mjd[8] combinations which are

- 1250-52930
- 1251-52964
- 1256-52902
- 1257-52944
- 2555-54265
- 2565-54329

These quasars are being studied further - either the observations on these specific dates are corrupted and are thus redder than usual or these quasars are intrinsically reddened. Either way, further analysis of the other objects observed from these same plates on the specific mean julian data should tell us if the data from these particular dates is corrupted or if we are indeed looking at something interesting and new!

[8]Mean Juliand Date

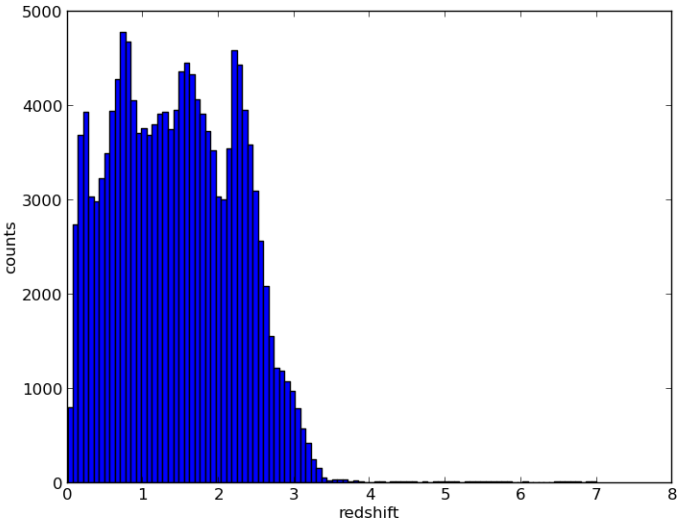

Figure 6: A histogram of the redshift of quasars from the new data set. Again, we can see that most of the quasars lie within redshift 3 and the number of quasars falls off rapidly afterwards.

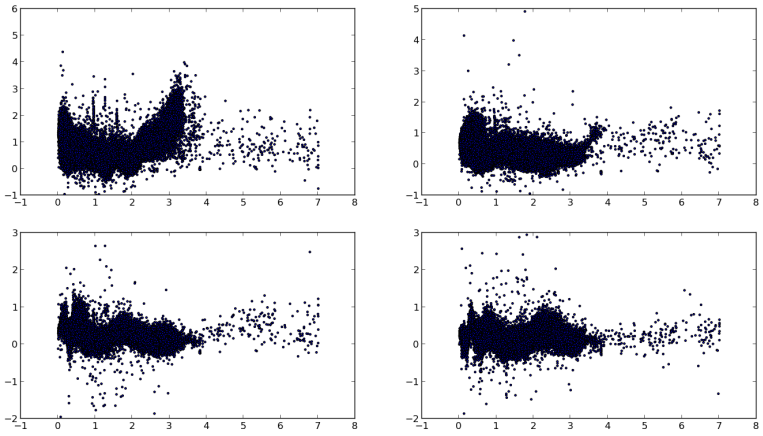

Figure 7: The 4 color - redshift plots for quasars in the new data set. A trend similar to that followed by the quasar colors of the original data set can be seen.

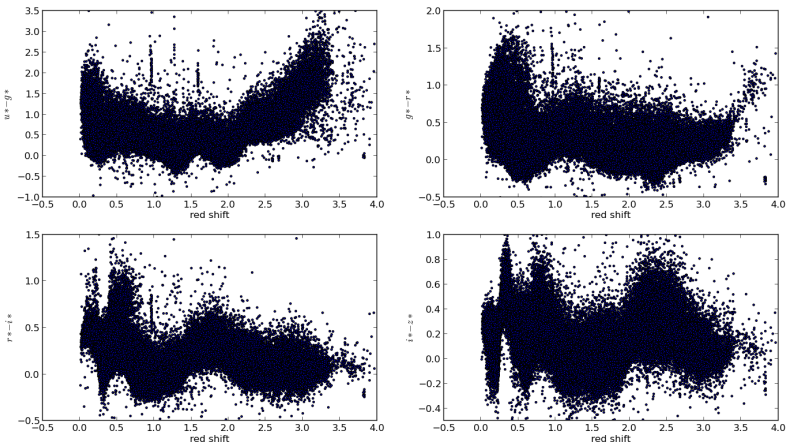

Figure 8: Enhanced version of the previous plot, with redshift range limited to 4. The trend in the color with increasing redshift is much more obvious here. Note the vertical features at z = 0.97 and z = 1.59 in the color-redshift plots, clearly visible in the u-g color-redshift plot. These vertical features are being studied further.

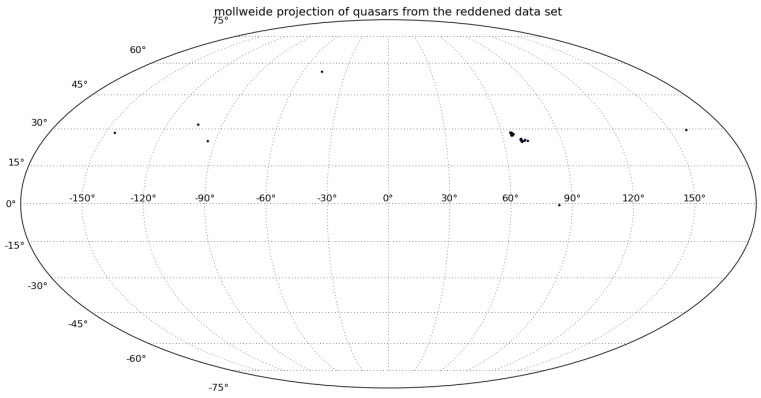

Figure 9: Positions of quasars from the first vertical feature at z-0.97.

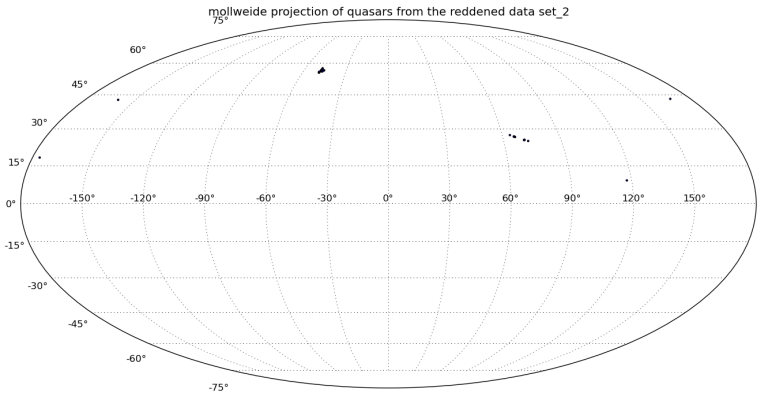

Figure 10: Positions of quasars from the second vertical feature at z=1.59. It is clearly obvious that there are two specific regions in the sky where quasars from these two vertical features seem to lie in.

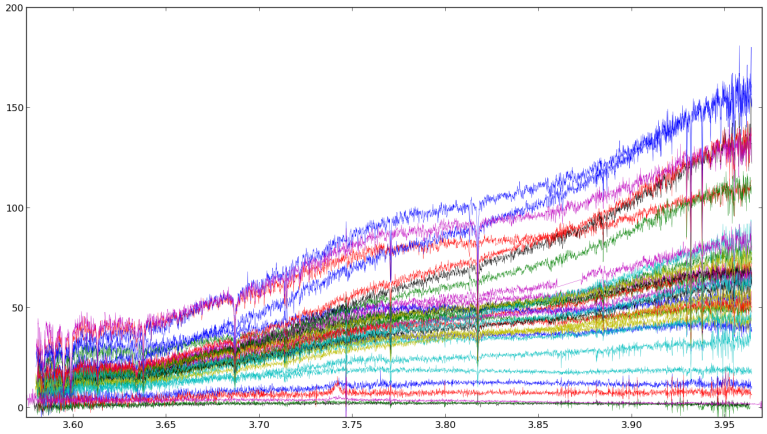

Figure 11: Spectra of the 33 quasars that lie on the first vertical feature at z=0.97. Again, we can see a band in the middle and a few quasars above and below this band with extremely large and extremely small fluxes correspondingly.

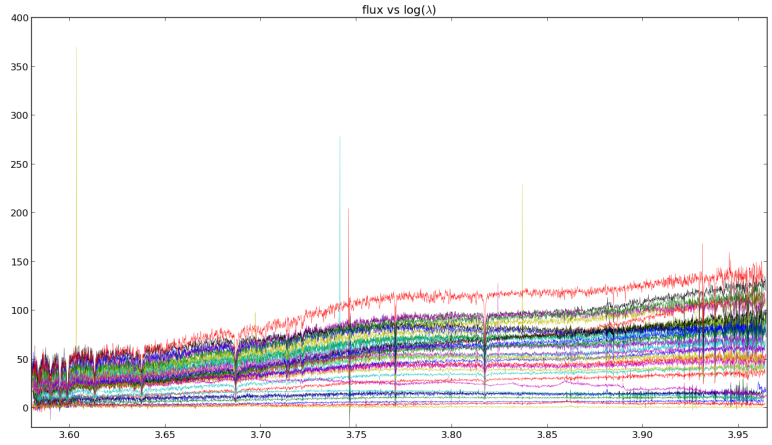

Figure 12: Spectra of the 34 quasars that lie on the second vertical feature at z=1.59. Similar features as that of the previous plot are visible.

3 Future Work

As mentioned above, the vertical features in the new color-redshift relation are being studied as their behaviour seem peculiar.

A follow-up paper by Richards et al. (2001) titled "The Photometric Redshifts of Quasars" is under study. The paper describes a method to estimate the redshift of a quasar using their photometric magnitudes. This method is extremely useful as Spectroscopy is a time-consuming and arduous task in comparison to Photometry. Therefore, photometric redshift estimation can help us with candidate selection for further spectroscopic studies. The paper describes a basic χ^2 minimization technique where we calculate χ^2_z as follows

$$\chi^2_z = \frac{(u'-g') - (u'-g')_z}{\sigma_{(u'-g')} - \sigma_{(u'-g')_z}} + C_{gr} + C_{ri} + C_{iz} \tag{3.1}$$

where C_{gr} and the others are analogous to the term for u'-g'. $(u'-g')_z$ is the median color wrt redshift, $\sigma_{(u'-g')}$ is the 1σ variance of the $(u'-g')$ color and $\sigma_{(u'-g')_z}$ is the 1σ variance in the median color.

We now look for the z where the χ^2_z is a minimum. Further details regarding the χ^2 minimization technique and the various hurdles in the process are described in Richards et al. (2001b). The paper also describes other methods that can be implemented to estimate photometric redshifts.

This technique of estimating the photometric redshifts of quasars is implemented on every single quasar discovered by the SDSS before it is confirmed spectroscopically. And

as I mentioned, not every single quasar is observed spectroscopically, so for the rest of the quasars we apply this method.

The paper by Fan et al. (1999) titled "Simulation of stellar objects in SDSS color space" is also under study. The paper describes the method to simulate the color-redshift relation using a composite quasar spectrum. Various composite spectra - differing in emission and absorption features and put through the simulation to look for differences in the color-redshift relation of a quasar caused by these spectral features.

The χ^2 minimization technique can also be implemented by using these simulated median color values - instead of the median color values calculated based on richards et al. (2001). The outcome of this test should be interesting as the simulated median color values are independent of the size and contents of the data set being worked on where as in the other case, the better the data set, the better the median color obtained.

Appendix A Relevant Codes

A.1 SQL Query

```
select
        plateX.plate, plateX.mjd,
        PhotoObj.ra, PhotoObj.dec, PhotoObj.ObjID
        PhotoObj.modelMag_u, PhotoObj.modelMag_g,
        PhotoObj.modelMag_r, PhotoObj.modelMag_i, PhotoObj.modelMag_z,
        specObj.fiberID, specObj.z,

from
        PhotoObj, specObj, plateX

where
        specObj.bestObjid = PhotoObj.ObjID
        AND plateX.plateID = specObj.plateID
        AND specObj.class = 'qso'
        AND specObj.zWarning = 0
        AND PhotoObj.modelMag_u < 22.3
        AND PhotoObj.modelMag_g < 22.6
        AND PhotoObj.modelMag_r < 22.7
        AND PhotoObj.modelMag_i < 22.4
        AND PhotoObj.modelMag_z < 20.5
```

A.2 Python

A.2.1 The Color-Redshift Plots

```
>>> import pandas as pd

>>> u,g,r,i,z = [], [], [], [], []
>>> redshift = []

>>> data = pd.read_csv(file)
>>> data

>>> u,g,r,i,z = data['modelMag_u'], data['modelMag_g'], data['modelMag_r'],
                        data['modelMag_i'], data['modelMag_z']
>>> redshift = data['z']

>>> u_g, g_r, r_i, i_z = [], [], [], [], []

>>> u_g, g_r, r_i, i_z = data['modelMag_u'] - data['modelMag_g'],
                        data['modelMag_g'] - data['modelMag_r'],
                        data['modelMag_r'] - data['modelMag_i'],
                        data['modelMag_i'] - data['modelMag_z']

>>> plt.hold(True)

>>> subplot(221)
>>> scatter(redshift, u_g, s=5)
>>> subplot(221)
>>> scatter(redshift, g_r, s=5)
>>> subplot(221)
>>> scatter(redshift, r_i, s=5)
>>> subplot(224)
>>> scatter(redshift, i_z, s=5)
```

A.2.2 The Median Color

```
>>> a = np.linspace(0,5,100,endpoint=False)

>>> import pandas as pd

>>> data = pd.read_csv(file)

# having already run the color_plot() command and
# set read redshift

>>> mid_u_g = []
>>> median_u_g = []

>>> for i in xrange(len(a)):
        for j in xrange(len(redshift)):
                if a[i]-0.0375 < red[j] < a[i]+0.0375:
                        tmp = u_g[j]
                        mid_u_g.append(tmp)
        tmp = np.median(mid_u_g)
        median_u_g.append(tmo)
        mid_u_g = []
>>> plot(a,mid_u_g)
```

A.2.3 The Spectra

```
>>> import pyfits as pf
>>> import astropy.io.ascii.read as ap.read
>>> import matplotlib.pyplot as plt

>>> plt.hold(True)

>>> def specplot(file):
        hdulist = pf.open(file)
        print hdulist.info()
        print len(hdulist[1].data)
        flux = []
        loglam = []
        for j in xrange(len(hdulist[1].data)):
                tmp = (hdulist[1].data)[j][0:1]
                flux.append(tmp)
                temp = (hdulist[1].data)[j][1:2]
                loglam.append(temp)
        print len(flux)
        print len(loglam)
        plot(loglam, flux, lw=0.3)

>>> list = ap.read('fits_list_file')
>>> for j in xrange(len(list)):
                temp = list[j][0]
                specplot(temp)
```

Appendix B Sample Data

Here is a sample from the data set used. The numbers have been truncated to be able to show the various fields in the data set while in reality the magnitudes and RA, dec are F8.6 and F10.8 integers respectively.

- Mag_u - 20.050003
- Mag_z - 15.770048
- RA - 64.66438286
- dec - 28.69054366

plate	mjd	fiberID	Mag_u	Mag_g	Mag_r	Mag_i	Mag_z	ra	dec	z	ObjID
1257	52944	91	19.02	16.99	15.85	15.26	14.87	71.32	25.64	0.967008	1237660760537100000
4997	55738	704	21.68	20.67	19.95	19.76	19.51	256.87	32.02	0.969408	1237655501888550000
1251	52964	27	20.06	17.89	16.65	15.99	15.57	66.13	27.99	0.967218	1237660553302510000
1256	52902	470	19.85	17.72	16.50	15.85	15.44	69.05	25.90	0.967194	1237660559209400000
1251	52964	24	19.74	17.70	16.57	16.00	15.67	65.83	28.24	0.96773	1237660750872310000