Statistical Challenges in 21st Century Cosmology
Proceedings IAU Symposium No. 306, 2014
A. F. Heavens, J.-L. Starck & A. Krone-Martins, eds.

doi:10.1017/S1743921314011144

New Challenges in Cosmology Posed by the Sloan Digital Sky Survey Quasar Data

Adrija Banerjee and Arnab Kumar Pal

Dept.: Physics and Mathematics Unit, Indian Statistical Institute
203, Barrackpore Trunk Road, Kolkata, West Bengal- 700108, India
email: adibanhere@gmail.com, arnabandstats@gmail.com

Abstract. For SDSS quasar data (2005) we have truncated data structure whereas for the survey of 2007 the data is no longer truncated. This calls for development or use of completely different statistical methodology to study the data for the evolution of the same objects like quasars. These different methodologies suggest different interpretation for a particular phenomenon in nature. This leads to the issue of validation of the data. More intriguing and challenging issue crops up as, given all of the data, what can be said about the laws of physics that have been operating over the universe? Over here we have used the concept of Neural Network to model the relationship between redshift and apparent magnitude.

Keywords. galaxies: distances and redshifts

1. Introduction

According to the Hubble Law, under the assumption of the universe being homogeneous and isotropic, the galaxies appear to be receding with a velocity v proportional to their distance d from the observer: $v = H_o d$, where H_o is called the Hubble constant. There are two luminosity functions, absolute luminosity and apparent luminosity. Normally astronomers work with absolute magnitude M and apparent magnitude m, where $d \approx m - M$ is known as distance modulus and is related to the luminosity distance. Therefore, the relation between the distance modulus and $\log z$ is considered to test Hubble law. Roy *et al.* studied the bivariate distribution of redshift and apparent magnitude observed in Sloan Digital Sky Survey(SDSS) for quasar redshift (2005) in which the data was truncated. However, the redshift data in SDSS quasar survey of 2007 was no longer truncated. Mukhopadhayay *et al.* (2009) discussed a general framework to study data for redshift and apparent magnitude in quasar sample and found a nonlinearity of the fitting curve for Hubble like relation. Now the intriguing issue is, for SDSS quasar data (2005) we have truncated data structure whereas for the survey of 2007 the data is no longer truncated. This calls for development or use of completely different statistical methodology to study the data for the evolution of the same objects like quasars. These two different methodologies suggest different interpretation for a particular phenomenon in nature. This leads to the issue of validation of the data. In order to take a closer look into the debate we have taken up the Tenth Data Release of Sloan Digital Sky Survey on Quasars and investigate the validation of Hubble Law on it.

2. Data

Our massive cosmological data set, consists of 96307 data points on logarithm of redshift (z) and apparent magnitude (m) for Qsasars (qsasi-stellar objects) collected from SDSS data. We have cleaned the data by removing outliers and those values of for which ln z turns out to be undefined.

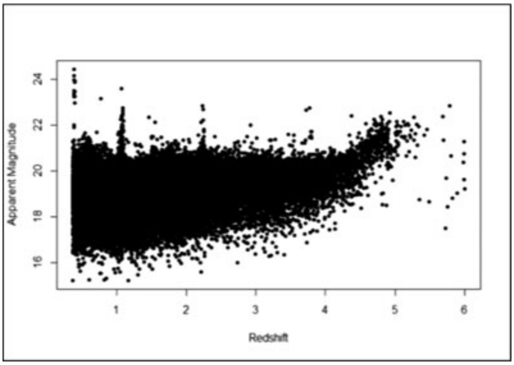

Figure 1. Scatter plot of z vs m

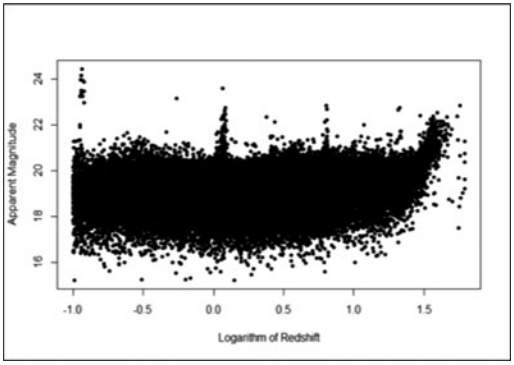

Figure 2. Scatter plot of inz vs m

From the Scatter Plot of z vs m (Fig 1) it is clear that the data is non-truncated.

3. Methodology

The huge and rich in information data as the SDSS Quasar data require cutting edge statistical technologies for their analysis. Thus over here we have used the concept of Neural Network Model Fitting. A Neural Network Model is very similar to non- linear regression model, with the exception that the former can handle an incredibly large amount of model parameters. For this reason, Neural Network Models are said to have the ability to approximate any continuous function.

At first to understand the overall pattern of the data we have binned the data, i.e., we have divided the whole support of lnz into 558 equal classes of width 0.005 and have taken the mean of corresponding variables. Fig 3 illustrates the scatter plot of the above data after binning. Thereafter on the binned data we have carried out Neural Networking and obtained the predicted values (See Fig 4). In order to deduce the upper and lower prediction lines we have taken the respective maximum and minimum points to each point of the binned data (See Fig 5) and fit Neural Networking on them in similar fashion (See Fig 6). Next in Fig 7 we have constructed the prediction curves and modified data after binning with the data as a whole.

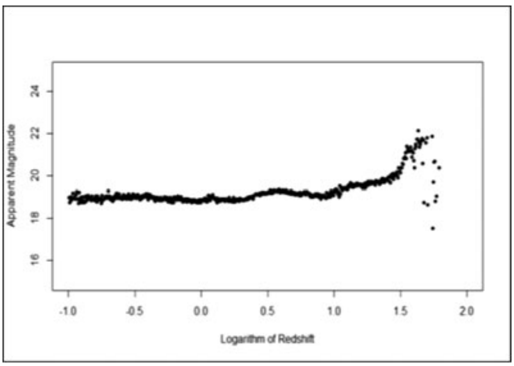

Figure 3. Scatter plot of Binned Data

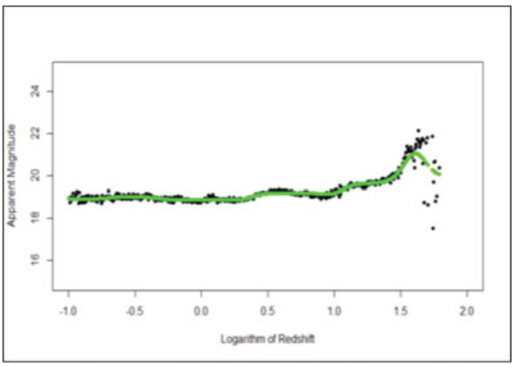

Figure 4. Predicated Values

Fig 4 illustrates the scatter plot of the above data after binning. Thereafter on the binned data we have carried out Neural Networking and obtained the predicted values (See Fig 5).

https://doi.org/10.1017/S1743921314011144 Published online by Cambridge University Press

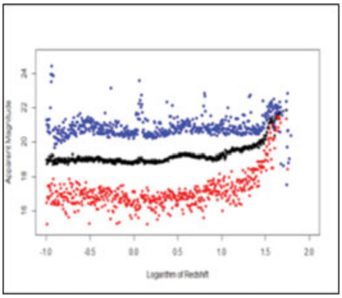

Figure 5. Maximum and Minimum Points Corresponding to Each of the Binned Data

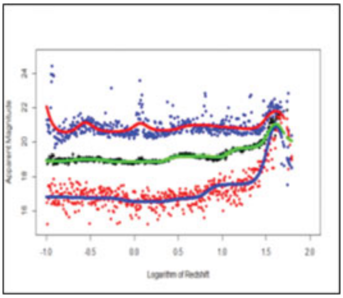

Figure 6. Fitted Upper and Lower Prediction Lines and Modified Data after Binning

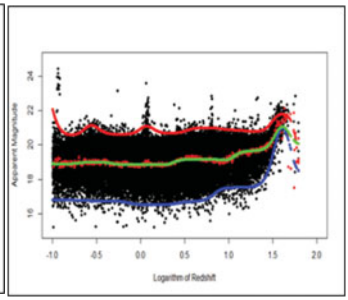

Figure 7. Prediction Curves Prediction Lines and Modified With Data as a Whole

4. Results

It is evident from the fitted values that for $lnz < 1$ the relationship between lnz and m is almost linear (Refer Fig 7). Thus, In order to validate Hubbles Law, we have tested for linearity when $lnz < 1, i.e., z \in [0, 2.718281]$. On constructing the test for linearity for the Upper, Mean and Lower Data Clouds we get the following results:

For Upper Boundary Data Cloud				For Lower Boundary Data Cloud				For Mean Constructed Data Cloud			
Coefficients	Est.	Std. Error	p-value	Coefficients	Est.	Std. Error	p-value	Coefficients	Est.	Std. Error	p-value
Intercept	20.840	0.0006	<2E-16	Intercept	16.700	0.0007	<2E-16	Intercept	19.010	0.0004	<2E-16
β	0.064	0.0012	<2E-16	β	0.119	0.0014	<2E-16	β	0.198	0.0008	<2E-16
Adjusted R^2		0.032		Adjusted R^2		0.087		Adjusted R^2		0.467	

5. Possible Implication

It is found from our analysis of the Tenth Data Release of Sloan Digital Sky Survey on Quasars that the Hubble relation between m and log z is valid for small values of redshift (z) , i.e., for the range: $z \in [0, 2.718281]$ For higher values of z, Hubble Law breaks down because we get a non-linear curve. Now the deviation from Hubble relation may be due to some other mechanism for redshift. It may be pointed out that the environmental effects for the Quasars may be taken into consideration to explain the deviation. This kind of environmental effect has been modeled in Doopler like mechanism. The deviation may also be a hint at the presence of dark energy or vacuum energy.

6. Acknowledgment

We would like to acknowledge and thank Prof. Sisir Roy of Indian Statistical Institute for his precious guidance and support.

https://doi.org/10.1017/S1743921314011144 Published online by Cambridge University Press

References

Abell, G. O.(1958) *Astrophys.* J.Suppl. 3, 211-288

Coles, P. & Lucchin, F., (2002) *Cosmology : The Origin and Evolution of Cosmic Structure.* 2nd Edition, John Wiley & Sons, Ltd. p.77

Datta, S., Roy, S., Roy, M., & Moles, M. (1998), *Phys.Rev.A*, 58,720

Efron, B. & Petrosian, V. (1992) *Astrophys. J.*, 399, 345-352

Hoessel, J. G., Gunn, J. E. & Thuan, T.X. (1980) *Astrophys.*, 241,486-492

Korany, D. M. & Strauss, M. A. (1996) *Testing the Hubble Law with the IRAS 1.2 Jy redshift Survey.* http://xxx.lanl.gov/astro-ph/9610034

McLaren, C., Wagstaff, M., Brittegram, G., & Jacobs, A. (1991) *Biometrics* 47, 607-708

Roy, S., Datta, S., Roy, M., & Kafatos, M. (2006) *Statistical Analysis of Quasar Data and Hubble Law*, arXiv:astro-ph/0605356v1

Roy, S., Kafatos, M., & Datta, S. (1999), *Phys. Rev. A*, 60, 273

Segal, I. E. & Nicoll, J. F. (1992) *Proc. Nat. Acad. Sci.* (USA) 89,11669-11672

Veron-Cetty, M. P. & Veron, P. (2001) *A&A*, 374, 92

Wolf, E. & James, D. F. V. (1996), *Rep. Prog. Phys.*, 59, 771