Calculating Hubble's Constant from Redshifted Galaxies

Katherine Mora

The initial purpose of this project was to analyze the redshifts of type Ia supernovae spectra in distant galaxies in comparison to known spectra of elements like hydrogen. However, finding data on distinct supernovae did not fit the scope of the project as the aim was to calculate Hubble's constant which requires galaxy spectra as well. Thus instead of finding standard candles in addition to galaxy spectra, only galaxy spectra from the DESI Early Data Release (EDR) were used to determine the value of Hubble's constant, and the project is as follows.

1. **Theoretical Background & Method Description**

In order to find, filter, and analyze the redshift data of numerous galaxies to determine the approximate value of Hubble's constant, H_0, the rate of the universe's expansion, the redshifts of numerous distant galaxies will be plotted against their distance as viewed from Earth. Initially the data will be loaded as an ascii data table with the use of astropy, in which it will be further reduced into useful values. In order to ensure the initial data was of high accuracy, any objects with a delta chi squared value less than 100 can be removed. Since this data will be used for calculating Hubble's constant, only spiral galaxies with redshift values $z << 1$ and no parallax can be used for calculations. Since parallax can only be used accurately to a relatively close distance, any galaxies close enough to Earth to have a parallax are not far enough to have a substantial cosmological redshift. On top of that, Hubble's Law, $v = H_0 d$, is only accurate with redshift values $z << 1$. Thus, given the z value for galaxies in the data, the recessional velocity, v, can be found by multiplying the redshift by the speed of light, c. So, Hubble's Law can be rewritten as $z * c = H_0 d$, and further manipulated into an equation to find Hubble's constant: $H_0 = \frac{z*c}{d}$. However, the distance of a galaxy can only be found through a few methods, and since the spectra and size of the galaxies will not be included in the data, the distance must be found through a number of assumptions. In this case the distance formula $d = \frac{s}{\theta}$ will be used, where d, s, and θ represent the distance from Earth to the galaxy, the actual size of the galaxy, and the observed angular size of the galaxy respectively. For the purpose of this project, it is assumed that all redshifts in the data are cosmological redshifts, which are redshifts due to the expansion of space between the observed galaxy and Earth. In addition to that, another large assumption that must be considered is that all galaxies being worked with are roughly the same size, due to the fact that the size of a galaxy cannot be easily found when the distance is unknown. So, if the longest axis of all the galaxies is assumed to be about 22 kiloparsec, only the angular size of the galaxy needs to be calculated to find the distance. Finding the observed angular size also comes with its own assumptions, since the data is given as a table and not as observable images. Thus, since the number of pixels along the longest edge of the galaxy is given, each pixel is taken to be about 1.18 arcseconds long. Assuming this, the angular size can be converted into milliradians by multiplying the length of the galaxy (in arcseconds) by a conversion factor. This conversion factor also takes into account the total viewing window, which in this case is assumed to be about 10,000 pixels in order to ensure that the full length of every galaxy can be viewed inside that box. The size of the galaxy in milliradians can then be used as θ to calculate the distance from Earth to the galaxy. Therefore, H_0 can be found by graphing the relationship between the recessional velocity (dependent variable) and the distance (independent variable) of a galaxy.

2. **Test Data**

The test data used to ensure the accuracy of the model's performance was data used in Astronomy C10's Hubble Law Lab. The redshifts and angular size of the galaxies had been manually measured and computed along with a graph of distance versus the recessional velocities of time to find Hubble's constant. Thus, test data was not generated as the original data was found on DESI's EDR, and special access is required to access the data and test data. The test data proved that the model performed as expected, and returned an approximate value of $H_0 = 57.88 \pm 12.05\ (km/s)/Mpc$. While the model was slightly adjusted to accommodate for a different type of data input, the form was exactly the same, and was fitted using the scipy.optimize curve fitting package.

3. **Data Fitting with Error**

The DESI EDR data used was plotted using the matplotlib.pyplot package, and later fitted with the scipy.optimize package. Initially, the data showed no correlation between distance (Mpc) and recession velocity (km/s) as the slope, the value of Hubble's constant, was found to be 0.014 ± 0.002 despite the model having performed as intended with the test data. There are a number of reasons why this could have occurred, but the most likely cause for the lack of correlation is the computation of the distance. Since the images of the galaxies were not a part of the data table, the observed size relative to the viewing box could not be measured itself. Instead, the number of pixels was given in the table as 'NPIXELS' and taken to be the number of pixels along the longest axis of the galaxy, or otherwise the observed number of pixels. All of the galaxies were roughly 7000 pixels across, so the size of the viewer was taken to be 10,000 to ensure that all of the galaxies fit within the theoretical viewing box. From there, each pixel was converted to arcseconds with the assumption that each pixel was about 1.18 arcseconds. While this assumption is not terribly out of bounds, the details of the binning and pixel size of the original data taking device is unknown. Thus, it is entirely possible that the arcseconds-to–pixel ratio is wrong and causing the distance to be calculated incorrectly, which would be reflected when plotting distance versus the recessional velocity. On top of the assumptions about the pixels and binning, the size of all of the galaxies were taken to be 22 kpc, the same as the test data. Spiral galaxy sizes can range from 5 kpc to 100 kpc along their longest axis, and 22 kpc is a reasonable size to assume the galaxies to be. However, seeing as the correlation between distance and recessional velocity was completely lost, other galaxy sizes were also tried with the data and proved to have no effect on improving the correlation.

4. **Explanation of Model Fit**

The model used for the purpose of this experiment found the distance to the galaxies by indexing through the data and performing the following procedure for each individual object. The model finds the number of pixels along the longest axis of a galaxy, converts said pixels to arcseconds, and thus multiplies the found angle in arcseconds by a milliradians to arcseconds ratio of the viewing box. This process then returns the observed angular size of the galaxy which can then be used to find the distance to the galaxy by dividing the assumed galaxy size, in this case 22 kpc, by the angular size. The graph of the distance versus recessional velocity was fitted using scipy.optimize, which allowed more accurate parameters to be found. This package also allows for the use of a covariance matrix, which provides the error of the refined slope when the square root is taken of the respective values for each variable's error. Since the correlation should have been linear, the only variable and parameter was the slope.

5. **Conclusion**

While the output was not what was expected, the model proved to work well with data that was chosen very particularly. In the future, this model could be revisited and refined to provide a more accurate value for Hubble's constant if the method of finding the distance can be improved upon by making fewer assumptions. Moreover, this experiment could be further improved upon if the distances to the galaxies were found using standard candles such as Type Ia supernovae spectra. Overall, this project was critical in furthering my understanding of several different filtering, fitting, and data analysis methods.