PHYS40001: Computing Project

Measuring Hubble's Constant

Introduction

In the field of cosmology, which studies the origin and evolution of the universe, one of the most fundamental observations is that of Hubble's Law. The universe is expanding, which means the distance between objects is increasing with time. Hubble's Law states that the velocity at which objects, such as galaxies, are moving away from the Earth is proportional to their distance from Earth

$$v = H_0 D$$

where v is the velocity, D is the distance and H_0 is the value of Hubble's constant.

The distance to a galaxy is inferred from the brightness of the galaxy. The velocity is inferred from the Doppler-shift of light emitted by the galaxy. For a galaxy that is moving away from the observer, the light will appear to the observer to be "redshifted" to longer wavelengths. The relationship between the velocity of the galaxy and the shift in wavelengths is given by the redshift formula

$$\frac{\lambda_o}{\lambda_e} = \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}$$

Where λ_o and λ_e are the observed and emitted wavelengths of light, $c = 2.9979 \times 10^8 \, m \, s^{-1}$ is the speed of light and v is the velocity of the emitter. The Hydrogen-alpha spectral line (Hα) is a deep-red visible line with wavelength $656.28 \, nm$. This line is commonly observed in the emission spectra from galaxies.

Figure 1 illustrates how H_0 can be inferred from measurements of distance and the "redshifted" velocity.

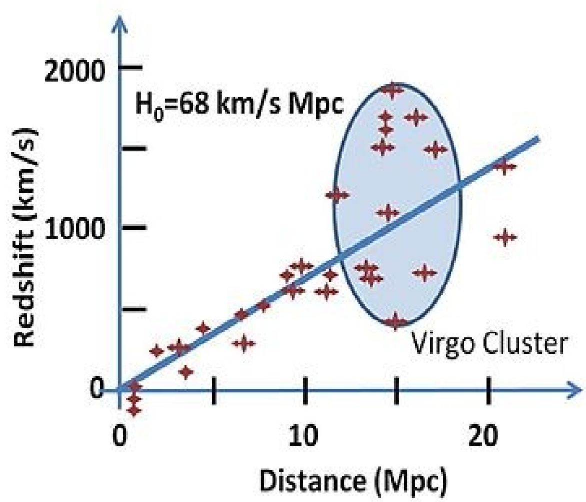

Fig. 1: A plot of velocity inferred from redshift versus distance for a range of galaxies. The slope of the line of best fit gives an estimate for H_0.

Project Objective

Given 30 observations of the redshifted Hα line from different galaxies and the distance to each galaxy, write a code in Python to calculate a value for Hubble's constant. The output from your code should be a plot of velocity inferred from redshift vs. distance for each galaxy, the value of Hubble's constant inferred from fitting this plot and its uncertainty.

Data

Here is a description of the data that you are given.

The file called "Halpha_spectral_data" contains data for the observed shift of the Hα spectral shift. This file is in .csv format. It consists of 4 header rows, below which are 31 rows of data. The first of these rows is frequency in Hz. The first column lists the observation numbers. The remaining rows and columns contain spectral intensity in arbitrary units, with each row corresponding to a different observation. An example of an observation is plotted in fig. 2. The first row of the header contains the Date at which the observation occurred and the name of the person who made the observation. The second row contains the observation number for each column of spectral intensity data.

The file called "distances_Mpc" is a .txt file in which the first row is a header. It has three columns. The first column is the observation number, the second column is the measured distance to that galaxy in Mpc and the third column is the instrument response. Note that the order of the observation numbers in the "distances_Mpc" and "Halpha_spectral_data" files are different.

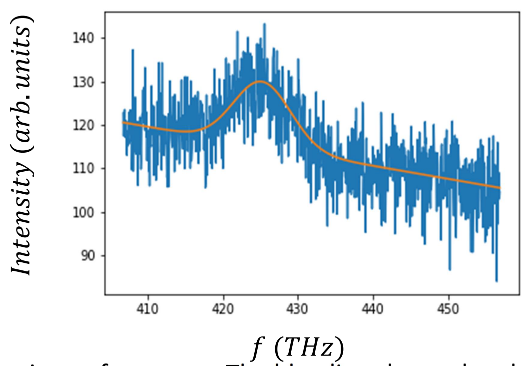

Fig. 2: A plot of spectral intensity vs. frequency. The blue line shows the observed data. The orange line is obtained by fitting a straight line + Gaussian function to the data.

Project Notes

The data for spectral intensity versus frequency is noisy, as can be seen in fig. 2. In order to use it to calculate the velocity, the data should first be fitted with a combination of a straight line and a Gaussian function. A similar fitting procedure was covered in the final exercise of Core Worksheet 2. Once the data is fitted, the mean value of the fitted Gaussian is a good estimate for the observed frequency, from which an observed wavelength, λ_o, of the shifted Hα line can be calculated.

The instrument response for each observation is either good (corresponding to a value of 1) or bad (corresponding to a value of 0). A bad instrument response can occur if, for example, we have low signal intensity, there is atmospheric interference of the signal or there is a drift in wavelength calibration due to thermal expansion of the instrument. These observations should not be used in the calculation of H_0.

Finding the distance for each observation involves matching the observation number in the spectral data file with the observation number in the distance data file since the order of observation numbers is different in both files. You should write Python code to do this matching.

Commenting your code & plagiarism

The project must be your own work and a full plagiarism check will be carried out on all submissions. It is acceptable for you to use pieces of code from online sources, your colleagues, etc. This is good practice for computer programmers. However, when you do so, you need to clearly state the source of the piece of code using a comment (#).

More generally, we expect you to effectively comment your code such that another programmer could easily understand what the code does, e.g. brief descriptions of functions, definitions of variables, etc. Submissions that contain little or no comments will be closely scrutinised for plagiarism.

Project Submission

You are required to submit your code (.py file), a plot of redshift velocity vs. distance and the fitted value of Hubble's constant and its uncertainty. The deadline for submission is 5pm on Friday 26th November. Full details on how to submit will be given out shortly.