PHYS40001: Computing Project

Measuring the Hubble 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 the Hubble–Lemaître law. The universe is expanding, which means the distance between objects is increasing with time. The Hubble–Lemaître 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 the Hubble 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 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, respectively, $c = 2.9979 \times 10^8 \mathrm{~m~s^{-1}}$ is the speed of light and v is the velocity of the emitter. One line that is often used to determine the redshift is the Hydrogen-beta spectral line (Hβ), a visible line with a wavelength of 486.1 nm that 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.

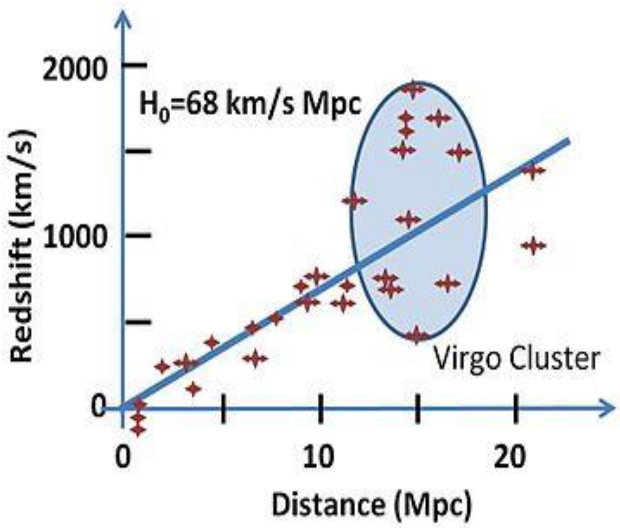

Fig. 1: A plot of velocity inferred from redshift versus distance for a range of galaxies. The slope of the line of best fit gives an estimate for H_0. (Image from Wikipedia)

Project objective

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

Data

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

The file called "dist_data.txt" contains the distance data. The first few rows are headers with information about the data. The first data column is the ID of the user who made the measurement. The second data column is a timestamp related to when the measurement was taken. The third data column is the observation number. The fourth data column is the measured distance to that galaxy in Mpc. The fifth and final data column is the instrument response: "1" if it is a valid measurement, "E" if the process returned an error.

The file "SpectralData_Hbeta.csv" contains data for the observed shift of the Hβ spectral line for each observation. The first few rows are headers with information about the data. The first row of the data indicates the observation number. For each observation number, the first column of data contains the frequency values (in Hz) and the second column contains the corresponding spectral intensity in arbitrary units for each frequency value. An example of the spectral data from one observation is plotted in fig. 2.

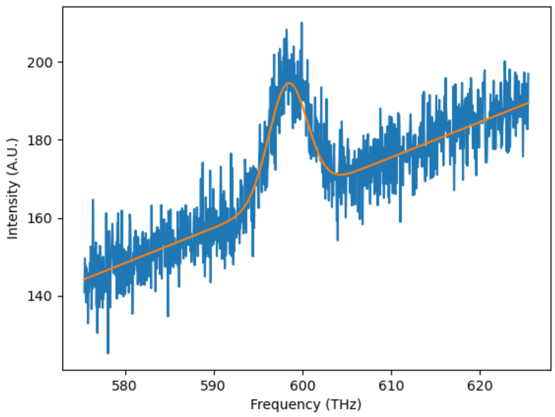

Fig. 2: A plot of spectral intensity vs. frequency. The blue line shows the observed data. The orange line is obtained by fitting a straight line + Gaussian function to the data.

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 scanning experiment fitting exercise of Core Worksheet 3. 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 Hβ 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 "E", indicating an error has occurred). 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. Observations with a bad instrument response should not be used in the calculation of H_0.

Generating a data set of velocities and distances in order to calculate H_0 involves matching each entry in the distance data file with the corresponding spectral data using the observation number. 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 some 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 in your code (following the # symbol). There is no strict format for such comments, but they should include the following information: what part of the code you are referring to (for example, "the next 2 lines of code"), how the source was used (taken directly, adapted, used as inspiration, etc.), and what the source is (with enough information for someone else to find exactly; if online, include a link).

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 no comments will not receive a passing mark.**

Project submission

The deadline for submission is **2pm on Friday the 21st of November**. On Blackboard, go to "PHYS40001 - Practical Physics: Laboratory, Computing and Problem Solving 2025-2026" → "Computing" → "Computing Project" → "Computing Project Submission". You are required to attach **two files**:

- your source code (.py file)
- the plot of redshift velocity vs. distance produced by running your source code (.png or .jpg file)

You must also include **your fitted value of Hubble's constant and its uncertainty**. This is done as follows:

1. Copy the text "Estimate of Hubble's constant: X +/- Y km/s/Mpc" into the Comments field
2. Replace X and Y with your own values