Estimating the Age of the Universe

Surya Sathujoda

1. Overview of Analysis

It is the objective of this work to develop a model to estimate the Hubble Constant and henceforth the age of the Universe, analysing Astrophysical data measured by the Hubble Space Telescope.

The steps taken to be able to estimate these quantities are as follows:

1. Determining the Cepheid period-luminosity (PL) Relation
2. Estimating the distance of a small set of nearby galaxies
3. Estimating the expansion rate of the Universe
4. Estimating the Age of the Universe

This report goes through the statistical methodologies used in each step of the procedure (without concerning itself with the physical derivations or astrophysical implications of the same), with details of the results from each section.

2. Cepheid Period-Luminosity Relation

The Cepheid PL relation relates the pulsation period of a star to its intrinsic brightness. In this section, data of nearby stars in the Milky Way Galaxy are used for calibration to determine the estimate of the two free parameters in the linear model relating the pulsation period and the absolute magnitude of the object, the equation for which is given below,

$$M = \alpha log P + \beta$$

where M is the absolute magnitude of the object and P is the pulsation period.

The data analysed for this is the parallax p_{mas}, the period P, the apparent magnitude m, and the extinction A. The uncertainties for p_{mas} and A are provided. The uncertainties in P and m are neglected as they are assumed to be far smaller than the actual values due to the precision of measurement available for these quantities.

First, d_{pc} (the distance of the star from Earth) in *parsecs* is calculated using the following relation,

$$d_{pc} = 1000/p_{mas}$$

where the resulting uncertainty in d_{pc} is given by,

$$\sigma_{d_{pc}}{}^2 = \left(\frac{-1000}{p_{mas}{}^2}\right)^2 \sigma_{p_{mas}}{}^2$$

The results of this are visualised in Figure 2.1.

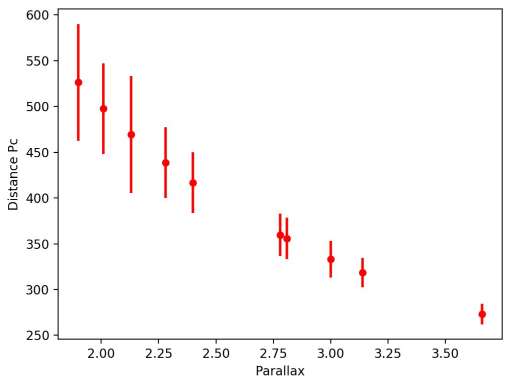

Figure 2.1

Tangentially, we derive the quantity of μ, this distance modulus which is given by the difference of m and M.

$$\mu = m - M$$
$$\sigma_\mu{}^2 = \sigma_M{}^2$$

(The uncertainty in m is neglected for the reasons mentioned previously). μ is also related to d_{pc} and A in the manner described below.

$$\mu = m - M = 5log_{10}d_{pc} - 5 + A$$

$$\sigma_\mu{}^2 = \left(\frac{\partial\mu}{\partial d_{pc}}\right)^2 \sigma_{d_{pc}}{}^2 + \left(\frac{\partial\mu}{\partial A}\right)^2 \sigma_A{}^2$$

$$\sigma_\mu{}^2 = \left(\frac{5}{d_{pc} log 10}\right)^2 \sigma_{d_{pc}}{}^2 + \sigma_A{}^2$$

After rearranging the above equation we get,

$$M = m - \mu$$

And the values calculated from this are used as the expected values to fit a linear model of the form described at the start of this section. The numerical algorithm used to fit this model is the *Levenberg-Marquardt* algorithm implemented in the *scipy* package.

After fitting the data from the 10 stars, the optimal values of α and β which the algorithm arrived at were

$$\alpha = -2.4006676240432157 \pm 0.23218943559966743$$
$$\beta = -1.6075025286776639 \pm 0.2023551114613707$$

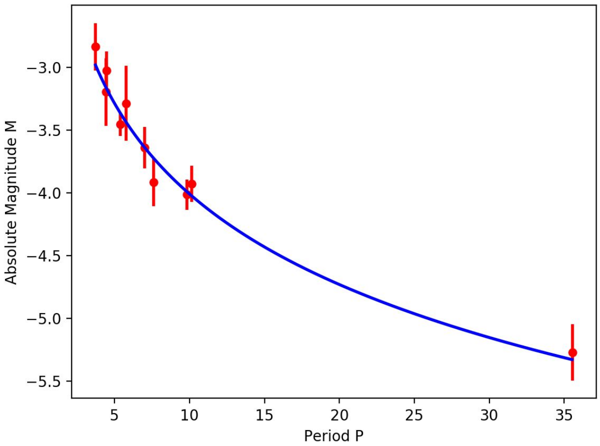

Figure 2.2

The correlation between α and β, which was derived from the covariance matrix of the *Levenberg-Marquardt* algorithm, was observed to be,

$$Correlation = -0.0455633067128161$$

which shows no strong dependency or anti-dependency between the two parameters so no further adjustment is done to nullify this.

Figure 2.2 shows the observed values of the absolute magnitude M with the predicted values by the model having the α and β values from above. (The plot of just the predicted values of M for each of the given periods with prediction uncertainty is given in the code run attached Figure 2 in code). To test the goodness of fit for this model, the Chi-squared and the Reduced Chi-squared statistic is used.

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - y_{m,i})^2}{\sigma_i^2}$$

$$\chi_\nu^{\ 2} = \frac{\chi^2}{N}$$

Here, y_m is the predicted model value, ν is the degrees of freedom. In this case, $\nu = N - 2$ since we have two free parameters in our model.

The values of Chi-squared and Reduced Chi-squared for the model fit with the aforementioned parameters is:

$$\chi^2 = 4.279452671456862$$

$$\chi_\nu^{\ 2} = 0.5349315839321077$$

Ideally, for a ‘good’ fit, the values of $\chi^2 \approx 8$ and $\chi_\nu^{\ 2} \approx 1$ (since we have 10 data points and 2 free variables). Clearly, the values for this model are below the good fit value, implying that the model fits too well for the given data. This could be due to many possibilities including overly optimistic uncertainties or lack of high volume of data. Since the period-luminosity relation is just an approximation of more complex physical phenomena, it is expected to have a fair amount of scatter which isn’t taken into account in the model. It can be easy to overfit a model if the number of data points are sparse, especially as the degrees of freedom increases. Acknowledging this flaw in our model however, we still accept it as it produces strong results for its given sample size.

Finally, it is important to determine the confidence in our model by looking at the calculated uncertainties in the free parameters. The uncertainties are calculated by finding the square-root of the diagonal entries of the covariance matrix from the *Levenberg-Marquardt* algorithm. These values give us the $1-\sigma$ *confidence intervals* for our free parameters.

$$\sigma_\alpha = 0.23218943559966743$$

$$\sigma_\beta = 0.2023551114613707$$

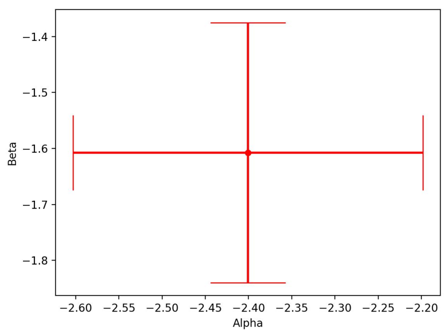

Figure 2.3

3. Distances for Nearby Galaxies

With the calibration values from the Cepheid PL relation, we work on estimating the distances of nearby galaxies given the $log_{10}P$ (pulsation period) and m (apparent magnitude) values of numerous sources within 8 different galaxies and also the A (extinction) of each of these galaxies caused from within the Milky Way.

Firstly, a method of estimating the distance modulus μ, of a given galaxy is required, using the multiple available sources from that galaxy. Recall that,

$$\mu = m - M$$

$$M = \alpha log_{10}P + \beta$$

Since we are given the m values and $log_{10}P$ values for the sampled stars in the galaxy we want to estimate the distance for, we can apply the calibrated α and β to estimate M and hence calculate μ. The uncertainty in the μ for each case is given by

$$\sigma_\mu^2 = (log_{10}P)^2\sigma_\alpha^2 + \sigma_\beta^2$$

Clearly, σ_μ^2 for each star inside a galaxy is not the same due to varying $log_{10}P$ for each star. Hence, it does not make sense to simply use the mean value of μ of all the stars in the galaxy as an overall estimate for the μ of the whole galaxy. Instead, we use the *inverse-variance weighted mean* to take into account the different uncertainties for each contributing term.

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i/\sigma_i^2}{\sum_{i=1}^{N} 1/\sigma_i^2}$$

Here, $\bar{x}$ is the *inverse-variance weighted mean* of the x_i terms with uncertainties σ_i. The uncertainty in the mean term is given by,

$$\sigma_{\bar{x}} = \sqrt{\frac{1}{\sum_{i=1}^{N} 1/\sigma_i^2}}$$

With these two values, we are able to estimate the μ (distance modulus) of each of the galaxies along with our uncertainty in the values. This can then be used to find the distance of the galaxy as follows,

$$\mu = 5log_{10}d_{pc} - 5 + A$$

$$d_{pc} = 10^{(\mu + 5 - A)/5}$$

The corresponding uncertainty in distance is,

$$\sigma_{d_{pc}}^2 = (\frac{\partial d_{pc}}{\partial \mu})^2\sigma_\mu^2 + (\frac{\partial d_{pc}}{\partial A})^2\sigma_A^2$$

$$\sigma_{d_{pc}}^2 = (\frac{d_{pc}log10}{5})^2\sigma_\mu^2$$

The terms for the uncertainty in A are dropped as they are assumed to be negligible compared to those of μ.

Figure 3.1 shows the results of following the above process to estimate the distance of various nearby galaxies. Some uncertainties are too small to be visible on the plot so for more comprehensive results, refer to the attached code output.

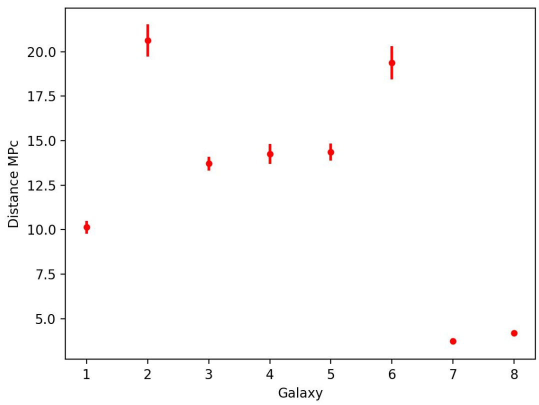

Figure 3.1

4. Estimating the Expansion Rate of the Universe

Since the distances of the galaxies have been approximated in the previous step, Hubble's model that the Recession Velocity v_{rec} and the distance of the galaxy D_{gal} are linearly dependant through a constant term (Hubble's constant H_0) is fitted to the data.

$$v_{rec} = H_0 D_{gal}$$

The same fitting algorithm is used as in the previous sections however here, the uncertainty in the values of v_{rec} are not known. The uncertainties on this term are non-negligible because it can be very difficult the measure v_{rec} for galaxies close to the observer. Hence, we are tasked with coming up with estimates for this term ourselves.

Since no information is given whatsoever regarding the uncertainties, we start off by assuming that the uncertainties are equal for all the data points. An arbitrary value of $\sigma_{v_{rec}} = 100$ is assigned for each data point (this value does not have a significant impact on the overall model as we will adjust this later to produce a desirable χ^2 value). Next, the model is fit using the dataset for v_{rec}, D_{gal} and $\sigma_{v_{rec}}$. The value of H_0 is estimated from this model, resulting in

$$H_0 = 70.75290738055529$$
$$\sigma_{H_0} = 2.5538539211708096$$

Now, the χ^2 and the $\chi_\nu{}^2$ with $\nu = 7$ is calculated giving the result below,

$$\chi^2 = 21.623424432037034$$
$$\chi_\nu{}^2 = 3.0890606331481476$$

Clearly, these values are too large as we want $\chi^2 \approx 7$ and $\chi_\nu{}^2 \approx 1$. We assume our model to be 'correct' however and adjust the value of $\sigma_{v_{rec}}$ to get the desired statistics. The formula to do this is,

$$\sigma'_{v_{rec}} = \sigma_{v_{rec}}\sqrt{\chi_\nu{}^2}$$

This results in the new uncertainty in v_{rec} to be

$$\sigma'_{v_{rec}} = 175.7572369249172$$

Finally, the model is refit with the new uncertainty to give the Hubble Constant and uncertainty to be

$$H_0 = 70.75290736601407$$
$$\sigma_{H_0} = 4.488583106503261$$

With $\chi^2 \approx 7$ and $\chi_\nu{}^2 \approx 1$. If the *scipy.optimize* module was used directly to use *curve_fit* on unweighted uncertainties, the same results are acquired.

Figure 4.1 shows the qualitative performance of the fit with the given H_0 value and also illustrates the 1σ *confidence interval* on the curve.

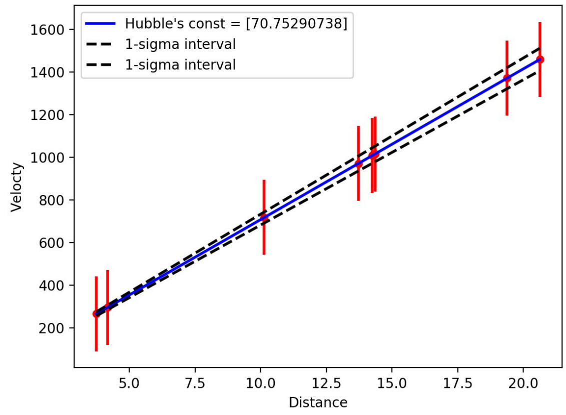

Figure 4.1

5. Estimating the Age of the Universe

Finally, the age of the Universe is pretty straightforward to predict given the above information. Some of the assumptions of this prediction is that the Hubble Constant is not time dependent or not a function of the contents of the Universe. That being said, the relationship between the age of the Universe and Hubble's Constant is given as

$$\tau = \frac{D_{gal}}{v_{rec}} = \frac{1}{H_0}$$

The uncertainty in this value therefore is

$$\sigma_\tau{}^2 = \left(\frac{-1}{H_0{}^2}\right)^2 \sigma_{H_0}{}^2$$

Since we are interested in finding the age of the Universe in Billions of years as units, we must first convert the Hubble's Constant to suitable units.

$$H_0\ km/s/MPc = 1.02271120232 * 10^{-12}\ H_0\ years$$

$$\tau\ gigayears = \frac{1}{1.02271129232 * 10^{-12} H_0} * 10^{-9}$$

The final results of the analysis give

$$\tau = 13.81982986 \pm 0.87673365\ gigayears.$$