HubbleConstant

May 8, 2019

1 Assignment 2:

2 The Expansion Rate and Age of the Universe

2.1 Executive Summary

Your goal in this assignment is to obtain estimates of the expansion rate and age of the Universe from actual astrophysical data.

All of your data analysis should be carried out in *Python*, using the methods you have learned in this course.

Your submission should consist of:

- the code you have written to carry out the data analysis
- a brief write-up of your analysis that presents your results and explains what you did

2.2 Required Analysis

The steps required to obtain estimates of the expansion rate and age of the Universe are:

1. determine the Cepheid period-luminosity (PL) relation

    - Cepheid variables are pulsating stars whose pulsation period is related to their intrinsic brightness
    - given a well-observed sample of nearby Cepheids, you can determine the form of this relationship

2. estimate the distances for a small set of nearby galaxies

    - given a sample of Cepheid variables in a galaxy, you can determine the distance towards that galaxy from your PL relation
    - our data set for this assignment includes samples of Cepheids in 8 nearby galaxies, so you can estimate distances for all of these

3. estimate the expansion rate of the Universe

    - we know the recession velocities of the galaxies in our sample, so you can estimate the "Hubble Constant", i.e. the constant of proportionality between recession velocity and distance

4. estimate the age of the Universe

   - since the age of the Universe is approximately equal to the inverse of Hubble's constant, you can easily estimate one from the other

2.2.1 Step 1: The Cepheid Period-Luminosity Relationship

The file "MW_Cepheids.dat" contains data for a sample of 10 Cepheid stars in the Milky Way. The table contains 8 columns:

1. The name of the star
2. its *parallax* in milli-arcsecond (see below)
3. the error on the parallax
4. its pulsation period in days
5. its V-band (visual) magnitude (see below)
6. its I-band (infrared) magnitude (see below)
7. the V-band extinction towards the star in magnitudes (see below)
8. the error on the V-band extinction in magnitudes

Magnitudes The brightness of astronomical objects is usually measured in *magnitudes*. Magnitudes can be a counterintuive measure of brightness when first encountered, for two reasons: (1) they are *logarithmic*, and (2) they are defined so that *brighter objects have smaller magnitudes*. Specifically, the magnitude scale is designed in such a way that if object 1 is 100 times brighter than object 2, and if object 2 has a magnitude of m_2, then the magnitude of object 1 is exactly $m_1 = m_2 - 5$.

All that said, luckily, the details of all this need not concern us here. This is because we will work in magnitude units throughout, so we never have to convert back and forth between magnitudes and other measures of brightness. The only other thing you should be aware of is that magnitudes are usually measured in distinct *bands*, such as a "blue" (B), "visible" (V) and "infrared" (I). In this assignment, we will be using magnitudes measured in the V and I bands.

Parallax The parallax of a star is its angular displacement relative to distant background stars over the course of a year (see e.g. here). The closer the star is to Earth, the larger its parallax.

More specifically, there is a simple relationship between the parallax of a star (p_{mas}, measured in milli-arcseconds) and its distance from Earth (d_{pc}, measured in *parsecs*):

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

The conversion from parsecs to meters is $1\,\mathrm{pc} = 3.086 \times 10^{16}\,\mathrm{m}$. For reference, $1\,\mathrm{pc} \simeq 3.3$ light years.

Apparent vs absolute magnitudes; distance modulus Suppose we observe two identical objects that are located at two different distances from us. Then obviously the one further away from us will appear to be fainter. This is no different than saying that the 100 Watt light bulb in my office seems a lot brighter to me than the same light bulb in another office across campus, which I can barely make out through my window. It is therefore important to distinguish between the *intrinsic* brightness of an object -- something that measures the fact that the light bulb is 100 Watts -- and the *apparent* brightness of an object -- which, for a given intrinsic brightness, depends only on the distance towards the object.

In astronomy, we use *absolute magnitudes* to talk about *intrinsic brightness* and *apparent magnitude* to talk about *apparent brightness*. In order to distinguish between them, we denote an *absolute magnitude* by an *upper case M*, and *apparent magnitudes* by lower case *m*. And since we usually talk about brightness measured in a specific band, we use subscripts to tell us which band. So, for example, m_V is the apparent magnitude of an object in the V (visual) band, while M_I denotes the absolute magnitude of an object in the I (infrared) band. For reference, the absolute magnitude of an object is formally defined as the apparent magnitude it would have if it were located at a distance of exactly 10 pc.

Finally, the relationship between absolute and apparent magnitudes obviously has to involve the distance from the object. And since magnitudes are logarithmic units, the difference between apparent and absolute magnitudes scales with the logarithm of the distance. This difference is called the *distance modulus μ* and defined as:

$$\mu = m - M = 5 \log d_{pc} - 5.$$

Extinction   When we observe an object in the Milky Way or in a distant galaxy, there is often a substantial amount of gas and dust between us and the object. This gas and dust can obscure our view of the object, i.e. it can make the object appear *dimmer* than it would if this stuff weren't in the way. This implies that the relationship between apparent and absolute magnitudes we've just seen usually needs to be corrected for this effect. We do this by defining the *extinction* in a given band -- e.g. A_V or A_I -- as the dimming effect of the intervening "stuff", measured in units of magnitudes. We then have the "corrected" version of the equations above:

$$m_V - M_V = 5 \log d_{pc} - 5 + A_V = \mu + A_V$$

$$m_I - M_I = 5 \log d_{pc} - 5 + A_I = \mu + A_I$$

Note that extinctions are always positive, and we have to *add* them to the right-hand side of the equation. For a given object, the extinction in the I band is always less than that in the V band. In particular, a reasonable approximation linking the two extinctions is given by

$$A_I \simeq 0.556 \, A_V.$$

Period-luminosity relation   The intrinsic brightness of a pulsating Cepheid variable star depends on the period of its pulsation. The physics of this is interesting, but needn't concern us here. The dependence of brightness on period is usually modelled via a simple linear relationship between the magnitude in some band and the logarithm of the pulsation period. So, for example, in the V and I bands, the models can be written as:

$$M_V = \alpha_V \log P + \beta_V$$

$$M_I = \alpha_I \log P + \beta_I$$

These are the "V-band and I-band period-luminosity relations".

Putting it all together The data in the file "MW_Cepheids.dat", together with the information provided above, is sufficient to estimate the parameters defining the V-band and I-band period luminosities (and their uncertainties). Specifically, your goal is to obtain optimal estimates (and errors) for α_V and β_V, as well as for α_I and β_I. You should also address the correlation or lack thereof between the αs and the βs.

In order to achieve this, you will first need to estimate the absolute magnitudes for all Cepheids in the sample (using their apparent magnitudes, the parallaxes and the extinction estimates). The data includes the pulsation periods, so you can then fit the absolute magnitudes, using the period-luminosity relations described above as your model. This will give you the α and β values.

Hints and tips Here are a few hints and tips you may want to think about as you do this (and address in your write-up).

- where errors are provided, you may assume that they are Gaussian

- where errors are *not* provided (e.g. on apparent magnitudes and periods), it usually means that the relevant quantities is very precisely known, so the error on it can be ignored

- in order to estimate errors on the absolute magnitudes, you will need to propagate the errors on the distance and extinction somehow; make sure to think about this error propagation carefully

- since the simple, linear period-luminosity relation is really just an approximation to fairly complex underlying physics, it is possible that there may be extra scatter around the relationship (in excess of that expected from the observational errors); if this happens -- i.e. if one or both of your fits to the period-luminosity relation are not formally acceptable (in terms of its χ^2), you should still "accept" the relationship but explicitly allow for this extra "intrinsic dispersion".

- If your best-fit α and β values are correlated, you might want to consider using the strategy discussed in Lesson 15 to reduce or remove that correlation).

Step 1: The Cepheid Period-Luminosity Relation

$$M_V = \alpha_V \log P + \beta_V$$

2.2.2 Step 2: Distances for a Set of Nearby Galaxies

The files "hst_gal1_cepheids.dat" through "hst_gal8_cepheids.dat" contain data obtained by the Hubble Space Telescope on Cepheid variables discovered in 8 nearby galaxies. Each file contains the following information:

- Header Row 1: Name of the galaxy

- Column 1: Name of the Cepheid

- Column 2: logarithm of the Cepheid period in days

- Column 3: apparent magnitude of the Cepheid in the V-band

- Column 3: apparent magnitude of the Cepheid in the I-band

4

In addition, the file "galaxy_data.dat" provides information about each of these galaxies. Specifically:

- Column 1: Name of the Galaxy
- Column 2: the galaxy's recession velocity (in km/s)
- Column 3: $A_{V,MW}$, an estimate of the V-band extinction in the Milky Way towards the galaxy

Now that $A_{V,MW}$ only accounts (approximately) for the dimming due to materials in *our* galaxy (the Milky Way). Each Cepheid may additionally suffer from some extinction in its own host galaxy, but we have not been given estimates of this here.

Using the relationship between apparent and absolute magnitude, extinction and distance modulus, obtain an optimal estimate of the distance modulus (and its uncertainty) for each of the 8 galaxies. Also convert these estimates to distances in parsecs.

Hints and tips

- you should at least think about -- and discuss in your write-up -- how extinction in each Cepheid's host galaxy may affect your distance estimates

- propagate all relevant uncertainties correctly into your final errors for the distance moduli and distances; make sure to think about whether any of the uncertainties you are propagating are correlated with each other.

- you might want to think about and/or investigate whether the resulting errors on the distance moduli and the distances are Gaussian or not

2.2.3 Step 2: Distances for a Set of Nearby Galaxies

$$m_V - M_V = 5\log d_{pc} - 5 + A_V$$

2.2.4 Step 3: Expansion Rate of the Universe

You now have distances (and associated errors) for your 8 galaxies. As described above, the table "galaxy_data.dat" also provides the recession velocity for each galaxy. The recession velocity is the speed with which the galaxy appears to move away from us.

This recession velocity is actual due to the expansion of the universe. A simple way to picture this is to think of galaxies as raisins in a loaf of bread being baked in an oven. As the bread loaf expands, all raisins move apart from each other, and since there is more expanding dough between distant raisins thant nearby raisins, distance raisins move away from each other faster.

Thus there is -- or should be -- a simple linear relationship (called *Hubble's Law*) between the distance of a galaxy from us, D_{gal} and its apparent recession speed, v_{rec}:

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

Note that there is no constant term in this relationship, and the slope of the relationship H_0 -- known as the *Hubble Constant* -- is usually measured in units of (km/s) per Mpc (1 $Mpc = 10^6$ pc).

By combining the distances you have measured for our 8 galaxies with the recession velocities provided for them, you should be able to obtain an optimal estimate for H_0, along with an estimate of the error on this.

Hints and tips

- No errors were provided for us on the recession velocities. Since these galaxies are fairly nearby, their recession velocities are actually fairly hard to measure (since their overall velocities are also affected by "local" effects, i.e. effects other than just the expansion of the universe). We should therefore not be surprised if the recession velocities do suffer from significant errors.

- Instead of using the relationship between v_{rec} and D_{gal} exactly as given above, you might consider taking the logarithm of both sides and using this in your analysis. Think about -- and and perhaps comment in your write-up -- as to why that might be a good idea.

2.2.5 Step 3: Expansion Rate of the Universe

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

2.2.6 Step 4: Age of the Universe

If you think about the form of Hubble's Law and the units of Hubble's Constant, you can see that it allows you to estimate the age of the universe.

If the whole universe is expanding, and all of it started in a single "point", then the time it took an object moving with v_{rec} to move to a distance D_{gal} is given by

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

So $1/H_0$ is an estimate for the age of the expanding universe!

Strictly speaking, this is only an approximation, since the detailed recession speed can be a function of time, with the exact functional form depending on the "contents" of the universe. However, it is a very good approximation, and certainly good enough for our purposes here.

Using your estimate of H_0 from above, estimate the age of the universe (in Gigayears, i.e. in units of 10^9 years), along with the uncertainty on your estimate.

2.2.7 Step 4: Age of the Universe

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

2.3 Due data

Deadline for submission: 6:00pm on Friday, May 10, 2019

2.4 Late submission penalties

Late submissions will incur a penalty of 10% of the awarded mark times the number of working days, or part thereof, after the published submission deadline.

2.5 Extensions

Requests for extensions on the basis of illness or other special circumstances should be made via the normal special considerations process. The special considerations team in the faculty office (pa-scb@soton.ac.uk), or your personal tutor, should be able to help you with this.

2.6 Submission method

Completed programs should be submitted via the Blackboard site (under "Assignments"). If you encounter any problems during submission, email me (c.knigge@soton.ac.uk) before the deadline and include your submission as an attachment.

2.7 Report format and length

The written report on your analysis should provide a clear and concise explanation of what you have done and what your results are. You do **not** have to provide an introduction into the underlying (astro-)physics, nor any discussion or conclusions regarding the astrophysical implications of the results. (Do make sure to clearly define any symbols you are using, however.)

A good way to think about your report is as the data analysis section of a larger paper in which the observations and results that have been obtained would be described in detail.

The main thing we will be looking for in your report is evidence that you have fully understood the principles and methods you have used in order to obtain your results. Thus you should make sure to explain any assumptions or approximations you make, as well as any non-standard methods you apply.

A simple and sensible structure for your report could be:

- Outline of the overall analyis steps
- Step 1: the Cepheid period-luminosity relation
- Step 2: distances for nearby galaxies
- Step 3: the Hubble constant
- Step 4: the age of the universe

There is no need to pad your report, but do make sure to clearly explain what you have done. A rough *upper limit* on the report length is 3000 words.

2.8 Assessment and marking

All submitted programs and reports will be read/tested by teaching staff associated with the course.

Marks will be awarded in the following three basic categories:

1. The submitted code (Weight: 30%)

    - are all of the necessary steps included in the code?
    - are sensible algorithms used?
    - are the algorithms implemented sensibly and correctly?
    - are reasonable comments and docstrings provided?
    - is the code easy to read and understand?

2. The data analysis (Weight: 40%)

    - has the data analysis been carried out correctly?
    - have errors been estimated and propagated correctly?

3. The report (Weight: 30%)

    - does the write-up clearly and concisely explain what has been done?

- does it explain and justify any approximations that were made?
- does it explain and justify any non-standard approaches that were used?
- does it demonstrate a good understanding of the statistical principles used?
- is the report well-written and structured?

The mark achieved for this assignment will count 60% towards the final module mark.