ASTP 720 - Homework 6 - The Cepheid Period-Luminosity-Metallicty Relation

Due Date: October 15th, 2020

In 1908, after upon examination of 1777 classical (Type I) Cepheid stars in the Magellanic Clouds, Henrietta Swan Leavitt discovered a relationship between the star's luminosity and the regular periodic oscillation of their lightcurves. The work was published with Edward Pickering in a 1912 paper, where she showed the following relation on a specific examination of 25 Cepheids in the Small Magellanic Cloud:

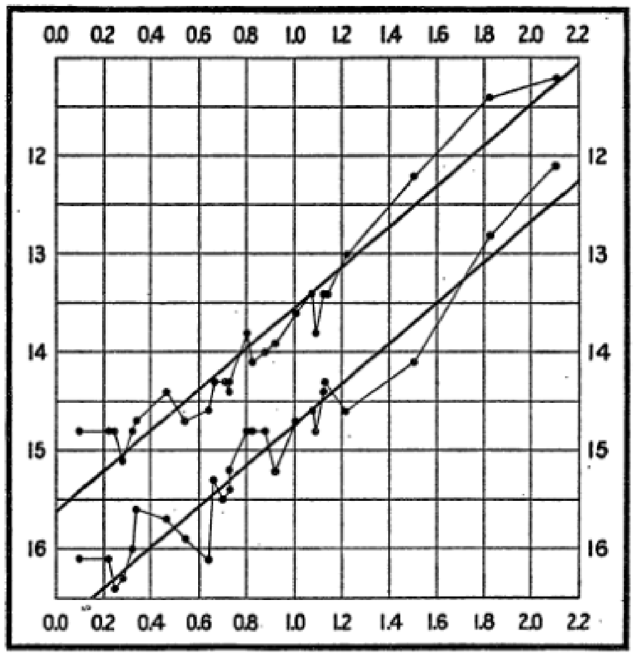


FIG. 2.

Leavitt and Pickering (1912), Harvard College Observatory Circular, 173, 1. The figure is unlabeled but in the text it is given that the x-axis is the logarithm of the period (in days) and the y-axis is the absolute magnitude (presumably in V band given the photographic plates but it is unclear). The two curves denote the maximum and the minimum of the lightcurve values rather than showing one mean value.

In the paper, they state "The logarithm of the period increases by about 0.48 for each increase of one magnitude in brighteness." By eye then, a possible model to describe the relationship is (using the mean value at $\log P = 0.4$ to be $(16+14.75)/2)$:

$$M \approx -\frac{1.0}{0.48}\log P + 16.2, \tag{1}$$

where M is the absolute magnitude (proportional to the logarithm of the luminosity) and P is the period. The 0.48 is in the denominator based on the description in the text being backwards from how we would usually describe a y vs x relation, and the minus sign is present since magnitudes are in a reversed scale. You can verify that when $\log P = 2$, we get $M = 12$, which is slightly off but pretty close to the mean magnitude.

Over a century later, we are still working on these relations since Cepheids are excellent standard candles for distance determination, so knowing the precise relation is key. We now know that the luminosity also dependends on the metallicity of the star as well as the period, i.e., one can extend the model to be:

$$M = \alpha + \beta \log P + \gamma \log Z. \tag{2}$$

Using a data set of 452 classical Cepheids accumulated by Groenewegen (2018, A&A, 619, A8), you will determine the parameters of the PLZ relation. In the work, he uses [Fe/H], the logarithm of the iron-to-hydrogen abundance compared to the Sun, as the metallicity, and therefore he writes the model as

$$\boxed{M = \alpha + \beta \log P + \gamma [\mathrm{Fe/H}]}. \tag{3}$$

This data set has been compiled for you in `cepheid_data.txt`[1] In this file, you will find the star name, the pulsation period in days, the distance in kiloparsecs, *apparent* magnitude values in four bands, the color excess, and the metallicity. Recall that the distance modulus for a photometric band denoted by λ is given as

$$m_\lambda - M_\lambda = 5 \log d - 5 + A_\lambda, \tag{4}$$

where m_λ is the apparent magnitude, M_λ is the absolute magnitude, d is the distance in parsecs, and A_λ is the interstellar extinction. Rather than give this value, you are provided with E_{B-V}, the color excess or *reddening*:

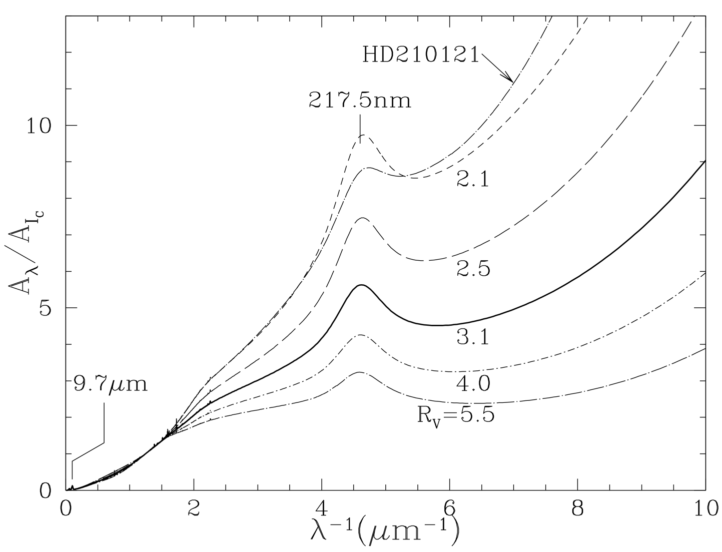
$$E_{B-V} = (B-V) - (B-V)_0 = A_B - A_V, \tag{5}$$

the difference between the observed $B-V$ color and the intrinsic $(B-V)_0$ color.

To apply the appropriate color correction to obtain absolute magnitudes, we can look at extinction curves through the Galaxy, shown below. The slope of the extinction curves at visible wavelengths can be parameterized by a dimensionless ratio:

$$R_V \equiv \frac{A_V}{A_B - A_V} = \frac{A_V}{E_{B-V}}. \tag{6}$$

It can be shown that with fits to the curves and this quantity alone, we can estimate the values of A_λ from the infrared to ultraviolet:



Draine, 2011, *Physics of the Interstellar and Intergalactic Medium*, Figure 21.2. Plotted are extinction curves at an inverse wavelength λ^{-1} relative to the Cousins I band (802.0 nm). The value of $R_V \approx 3.1$ is the average for the Milky Way.

[1] In Table 1 of Groenewegen (2018), which is available electronically, he does not provide the *Gaia* distances. These distances were instead queried using the SIMBAD database. However, there is significantly more scatter in the data than shown in his work. Nonetheless, a trend can be discerned.

The values of A_λ/A_V can be calculated from Table 21.1 of Draine (2011) and are given as:

Band	λ (μm)	A_λ/A_V
V	0.5470	1
J	1.22	0.271
H	1.65	0.175
K	2.19	0.117

Therefore, given E_{B-V} as in the data file, and assuming $R_V = 3.1$, one can determine A_λ and then make the appropriate estimate of M_λ.

Your Tasks:

[1.] Using linear least squares, estimate the parameters in the boxed equation above, α, β, and γ, and their uncertainties. You can pick your favorite photometric band.

[2.] Please plot your fit against the data. You can do this in a number of ways, I leave that to you as long as you can demonstrate that the fit is good (this is also a good check for you!).

[3.] Assume that the errors on the apparent magnitudes are all $\sigma_m = 0.1$ mag. Repeat the estimate of the parameters and their uncertainties.

You can also replace any previous homework code with `numpy` code, e.g., your `Matrix` class. Obviously, do not use `scipy.optimize.leastsq()` or a similar such code.

[Bonus] Again assume that the errors on the apparent magnitudes are all $\sigma_m = 0.1$ mag. Repeat your least squares estimate and perform it again when you fit a nested model of the form $M = \alpha + \beta \log P$. Recall that the F distribution results from the ratio of two chi-squared-distributed numbers (the distributions can have different degrees of freedom), then the F statistic is given by

$$F \equiv \frac{\left(\dfrac{\chi^2_{\text{nested}} - \chi^2_{\text{full}}}{\nu_{\text{full}} - \nu_{\text{nested}}}\right)}{\left(\dfrac{\chi^2_{\text{full}}}{\nu_{\text{full}}}\right)}. \tag{7}$$

To see where this F value "sits" in probability space (drawn in class as along the PDF), you can use `scipy`'s `stats` module:

```
return_value = stats.f.cdf(F, numerator_dof, demoninator_dof)
```

Think about what goes into those arguments and what the output value is saying and discuss what this means in terms of the significance of the parameter γ. To be clear, you can get partial bonus points for completing portions of this!