KSP 3.0 Writeup

Nayan Telrandhe
nayantelrandhe150@gmail.com

A dataset of the population of Cepheid variable stars from SMC is given, obtained by OGLE survey. The data is available in .csv format with ID, RA (H:M:S), Decl (D:M:S), I band mag, V band mag and Period as parameters of the stars. The cepheid stars pulsate radially, resulting in change in their luminosity. And there exist a relationship between the period of pulsation and luminosity of these stars.

Data Cleaning

- The data in .csv file is imported as a Pandas DataFrame.
- The data is filtered with entries having V or I band magnitude as -99.990, such entries are removed.
- DataFrames use an explicit index for rows, after filtering, this index is reset for further convenience.
- RA and DEC values in given formats are to be converted into degrees. The entries in RA and Dec are strings, they are split and data type is changed appropriately for conversion. The conversion is done using the following relations. The result is stored in arrays RA and Dec

$$RA(Degrees) = \frac{360}{24}\left(H + \frac{M}{60} + \frac{S}{3600}\right)$$

$$Dec(Degrees) = D - \frac{M}{60} - \frac{S}{3600}$$

  There are negative sign corresponding to 2nd and 3rd terms as $D < 0$.

- For plotting the datapoints on sky grid using RA and Dec, Mollweide projection is used as suggested. The projection uses coordinates in radians such that,

$$RA(proj.) \rightarrow (-\pi, \pi]\,, Dec(proj.) \rightarrow [-\pi/2, \pi/2]$$

  So a proper conversion is done to get the projection coordinates as

$$Dec(proj.) = \frac{2.\pi.(Dec(deg.))}{360}$$

  $RA(proj.) = \frac{2.\pi.(RA(deg.))}{360}$ and if $RA(proj.) > \pi$, $RA = RA - 2\pi$

P-L Diagrams

- $m_\lambda = a_\lambda log(P) + I_\lambda$ (assuming the base of log is 10) is the given relation between apparent magnitude, period of pulsation of the stars and the Intercept.
- Plots are created using Matplotlib's scatter plot and Scipy's curve fit module (as suggested) is used to find a best fit curve for the given data with a_λ and I_λ being the free parameters.

Deviation

Division of data into regions

- The data is to be divided into 25 regions without outliers (based on position). The RA and Dec in projection coordinates are added as new columns in original Dataframe.
- Quantiles are calculated at 0.5 and 0.95 for RA and Dec (as suggested). Using these as boundaries of RA and Dec range. The rows in DataFrames are filtered lying in the above interval for RA and Dec.
- The range is further subdivided to form 5 equally spaced intervals for both RA and Dec using numpy.linspace().
- 25 regions can be defined based on cartesian product of RA and Dec intervals. The DataFrame is thus divided into 25 parts based on regions, all stored as elements of a list L.

Curve fit function

- A function $Cfit(l, m1, m2)$ is defined which has a dataframe and $m1$, $m2$, the slope values for V band and I band as inputs. The function does a curve fit using the relation

$$m_\lambda = a_\lambda log(P) + I_\lambda$$

  and returns a tuple with optimised values of intercept for V band and I band as output.
- The function $Cfit$ is used to find the V and I band Intercept values for all Data frames of list L. The outputs are stored in a list Ii.
- The values of $m1$ and $m2$ chosen are same as for the whole data before filtering based on RA and Dec values. The different values of intercepts are due to vertical offset for each of 25 set from the global fit done earlier.

Deviation for each set

- $$D_{V-I} = (GI_V - I_V) - (GI_I - I_I)$$

  The above quantity is calculated for each set using the required values stored in list G and Ii for global and set intercepts. The output values of Deviation are stored in a list Di.
- A Histogram is plotted for the Deviation values for all 25 sets using matplotlib.pyplot.hist().

Wesenheit Index

Wessenheit index is an extinction insensitive index. There might be deviations in data due to interstellar extinction that happens due to scattering of shorter wavelengths because of the presence of dust between the source and observer. This results in reddening of the observed light. Wesenheit Index is given as

$$W = I - 1.55(V - I)$$

- A copy of the DataFrame 'Data' was created before filtering based on RA-Dec quantiles. WI vs log(p) graph is plotted for this data frame.
- One more plot is created for DataFrame 'D' obtained after filtering.
- The WI vs log(p) scatter plots seem thinner and less scattered than initial plots for the data. It's better for a line to fit.