Lab 1: Gaia, RR Lyrae stars, and Galactic Dust

Astro 128 / 256 (UC Berkeley)

Feb 4, 2019

Problem 0

Note: you likely already did some or all of Problem 0 for Lab 0.

a) For this lab, you will have to write a lot of ADQL queries. If you run into difficulties, you can consult the following tutorials.

1. Basic ADQL syntax
2. More extensive introduction to ADQL
3. Full *Gaia* ADQL cookbook. This one is quite long. It should serve mainly as a reference of example queries from which to construct new queries.

These tutorials are for your own edification, so spend as much or as little time as you need on them.

b) Make yourself an account on the Gaia Archive by going to the *Gaia* archive and clicking "sign in" and "register new user" in the upper right-hand corner. You don't need an account to query the Gaia catalogs, but having an account will allow you to save your previous queries and to upload and query your own tables.

c) Download Astroquery. The astroquery.utils.tap.core.TapPlus utility will allow you to combine ADQL queries with Python code.
Note: For Lab 0, some people used a different package, `astroquery.gaia.Gaia`. Feel free to use whatever packages you like to submit queries, but be aware that this one seems to time out faster than the `TapPlus` package.

Problem 1

In this problem, we'll explore the sample of RR Lyrae stars curated by Gaia DR2. The final goal is to build a dust map of the Milky Way. We'll make use of the fact that RR Lyrae's are "standardizable candles" – that is, they follow a period-luminosity trend (the Leavitt Law) that can be used to infer intrinsic luminosity using measurements of brightness and periodicity. This, in turn, can be used to infer distances.

The catalogs we'll need can all be found on the Gaia archive. To see which catalogs are available, go to the *Gaia* archive, click "search", and click on the "Advanced (ADQL)" tab on the upper left. Available catalogs are listed on the left side of the page in nested drop-down menus.

a) The `gaiadr2.vari_rrlyrae` catalog contains the results of RR Lyrae classification done by the Gaia data processing and analysis consortium for more than 10^5 RR Lyrae stars. Write an ADQL query to download the first 100 rows of the table for which a fundamental pulsation frequency ("pf") had been measured. Submit your query using Astroquery, and display the first 10 rows of the thus-obtained catalog.

To find the meaning of individual columns, you can search for them in the *Gaia* data model. To learn about the meaning of RR Lyrae–specific parameters and how they were derived, see Clementini et al. 2016.

b) The table you downloaded above contains the results of fitting the light curves of RR Lyrae stars, but it does not contain the raw light curves. The raw light curves can be accessed as described in the Datalink and light curves tutorial.

Download the light curves (i.e., G-band magnitude vs. time, with magnitude uncertainties) for the 100 RR Lyrae stars in your table. This is most easily accomplished with a Python utility for fetching web urls. `Urllib` is one good option. Plot one of the light curves.

c) Estimate the period and mean G-band magnitude of the 100 light curves you downloaded above.

A simple estimate of the period can be obtained using a Lomb-Scargle periodogram. An implementation is available in `astropy.stats.LombScargle`.

Plot the periodogram for one light curve, marking your estimate of the period on the plot.

d) Compare the periods and mean magnitudes you computed from the 100 light curves to the values reported in the `vari_rrlyrae` catalog. Comment on your results.

d.1) Grad students only: *If your Lomb-Scargle period-estimation routine works as ours does, you should have found that your period is in good agreement with the value in the* `vari_rrlyrae` *catalog in roughly 75% of cases, but that it disagrees significantly in the other cases.*
Read about how periods are derived in Clementini et al. 2016, and comment on what might drive the difference between your periods and those in the catalog. Then test your hypothesis:

- *Plot the periodograms of a few objects where your estimate of the period disagrees from that in the catalog. Overplot both your estimate of the period and that from the Gaia catalog. Does the Gaia estimate correspond to a secondary peak in the periodogram?*
- *Are your estimates of the period typically integer multiples of the Gaia estimates? This would indicate aliasing.*
- *Does the agreement rate with the Gaia estimates get better in cases where the light curve is better sampled (ie., more observations)?*

Given your analysis, can you improve your period estimation routine to better reproduce the Gaia period estimates?

e) Now that you've looked at the light curves for a few objects, we'll assume for the rest of the problem that the the periods reported in the Gaia catalog are reliable. We'll now use the data in the `vari_rrlyrae` table to infer the RR Lyrae period-luminosity relation in different bandpasses.

First, we'll need to use Gaia distances to estimate the absolute magnitude of RR Lyraes. This will only work if there isn't a lot of dust between us and the RR Lyraes. Explain why this is.

It turns out that most of the dust in the Milky Way is in the disk, at low Galactic latitude. Write an ADQL query to select RR Lyrae stars that (a) have accurately measured distances, with parallax errors of less than 20%, (b) are above or below the disk, with $|b| < 30$ degrees, where b is Galactic latitude, and (c) are relatively nearby, with distances less than 4 kpc. To do this, you'll need to join the `vari_rrlyrae` and `gaia_source` catalogs. You should find about 500 objects.

f) We'll also want an estimate of the distance that uses our knowledge of Galactic structure to place a prior on the distance inferred from parallaxes. A catalog of distances obtained in this way is described in Bailer-Jones et al. 2018 and is available in the Gaia archive as `external.gaiadr2_geometric_distance`. Write another ADQL query that gets these distance estimates and uncertainties for the targets obtained in your query in part (e).

g) Plot the distribution of targets you obtained in Galactic coordinates. Verify that your ADQL query has removed stars in the Galactic disk.

Compare the distance estimate from the Bailer-Jones catalog to the naive distance estimate from $d = 1/\varpi$ (where ϖ is the parallax in appropriate units). Explain the trend you see.

h) Plot period vs. absolute G-band magnitude for all stars returned by your query in part (g).

i) You should see that the majority of the stars have similar absolute magnitude, but a non-negligible fraction of them scatter far off the median relation. This is mostly due to incorrectly measured parallaxes. Apply the quality cuts in Equations C1 and C2 of Lindegren et al. 2018 to the sample. Then plot the period-luminosity relation again. Has the scatter decreased?

j) Most of the "bad" objects should have been removed by the above cut, but a few will remain. You can crudely remove these outliers by removing objects with absolute G magnitudes greater than some threshold of your choice. You can also try more sophisticated outlier rejection, if you wish. Plot the period-magnitude relation, including error bars on absolute magnitude due to distance uncertainties, from the resulting cleaned sample.

k) Fit a line to the G-band period vs. absolute magnitude relation. Assume a model $M_G = a \times \log\left[P/\mathrm{day}\right] + b$, where a and b are free parameters. Allow for intrinsic scatter in your relation. (This means your fit should include a "scatter" nuisance parameter).

Clearly state your likelihood function and priors. Use an MCMC sampler to explore the covariances between different model parameters. The "emcee" sampler and its tutorial may be useful.

l) Plot the traces for your MCMC walkers and convince yourself that the fit has converged. Use the "corner" package to visualize constraints on the posterior for the fit.

m) Plot ∼50 random, independent samples from the posterior over the data. Does the spread between samples as a function of period seem consistent with what you'd expect given the data? Explain.

n) Many of the RR Lyrae stars identified by Gaia were also observed by the `WISE` survey, which observed the whole sky in the near-infrared. Cross-match your sample of clean RR-Lyrae stars with WISE. The catalogs `gaiadr2.allwise_best_neighbour` and `gaiadr1.allwise_original_valid`, available on Gaia archive, will be useful.

o) Repeat steps k–m, now using the WISE "W2" magnitude rather than the G band. You can simply use the magnitude reported in the Wise catalog; you don't need to average any light curves.

p) Comment on the differences between your inferred period-luminosity relations in the optical and in the near-infrared. In which band is the period-luminosity relationship steeper?

Grad students only: *explain physically why the period-luminosity relation is steeper in one bandpass than in the other. This empirical fact was first pointed out by Longmore et al. 1986.* Also, explain why using the WISE magnitude in (o) above is a valid approximation.

q) Compare your derived period-luminosity relations to results in the literature, which can be found in Beaton et al. 2018 or Klein & Bloom 2014. You can compare your G-band relation to their V-band relation. If there are systematic differences between your results and the literature, what might account for them?

r) Following a similar procedure to the one in steps k-m, derive a period-color relation for RR Lyrae stars in the Gaia bands. That is, a relation between log(period) and the Gaia $G_{\mathrm{BP}} - G_{\mathrm{RP}}$ color. You will need to calculate color uncertainties using the uncertainties in BP and RP flux.

s) Write an ADQL query to download the full Gaia RR Lyrae catalog (i.e., no longer excluding stars with imprecise parallaxes or low galactic latitude) and cross-match it with the Gaia source catalog.

t) Calculate the *color excess,* $E\left(G_{BP}-G_{RP}\right) = \left(G_{BP}-G_{RP}\right)_{\text{observed}} - \left(G_{BP}-G_{RP}\right)_{\text{intrinsic}}$, for all RR Lyraes in the catalog. From this, calculate A_G, the G-band extinction, for each star. You may assume

$$R_G \equiv \frac{A_G}{E\left(G_{BP}-G_{RP}\right)} = 2.0$$

t.1 Grad students only*: Derive why $R_G = 2.0$ is a good approximation, assuming $R_V = A_V/E\left(B-V\right) = 3.1$. You may use the values in Table 6 of Schlafly & Finkbeiner 2011.*

u) Compare your calculated A_G values to the "G_absorption" value provided in the RR Lyrae catalog.

v) Plot a 2-d map of $E\left(G_{BP}-G_{RP}\right)$ as a function of galactic longitude and latitude, using an Aitoff projection. A simple way to do this is to plot each RR Lyrae as a point in a scatter plot with semi-transparent points, coloring each point by its A_G.

When you first make the plot, you will find that some large-scale structure is clearly visible, but also that there are some points for which the reddening looks clearly wrong; i.e., the color excess in one point is very different from the adjacent points. Construct appropriate quality cuts to remove these objects. You may want to cut on photometric signal-to-noise and/or BP/RP excess.

You will notice that the distribution of RR Lyrae stars in the catalog is not at all uniform. Why is this?

w) The most widely used Galactic dust map is the "SFD" map produced by Schlegel, Finkbeiner, and Davis 1998. Incidentally, all three authors have strong ties to UC Berkeley.

Compare the attenuation map you produced above to the SFD map. You can query SFD using the "dustmaps" Python package. You only need the SFD map; don't worry about installing the larger maps (some of which are several GB).

Plot the SFD optical reddening map, $E\left(B-V\right)$, sampled at the same positions as your RR Lyrae map. Does the general structure of your map agree with SFD? What about the small-scale details?

w.1) Grad students only: *Derive a conversion factor between $E\left(B-V\right)_{\text{SFD}}$ and $E\left(G_{BP}-G_{RP}\right)$. Table 6 of Schlafly & Finkbeiner 2011 may again be useful. Now plot the difference between $E\left(B-V\right)_{\text{SFD}}$ and $E\left(B-V\right)$ inferred from your map. Comment on the origin of any discrepancies.*
Hint: think about how the SDF map is constructed. What are the differences between what it is measure and what your map measures?