UNIVERSITY OF CAMBRIDGE

MPhil in Data Intensive Science

Submission: 11:59pm on Friday the 21st of June

- Coursework Assignment - Applications of Data Science to Exoplanets

Dr Vinesh Maguire Rajpaul

The coursework consists of two parts. Candidates should answer both.

The coursework needs to be submitted via the GitLab repository we have created for you. This should be accompanied by a short report of not more than 3,000 words describing the project and its development. The report should be in pdf format and placed in a folder named `report`. *You should ensure your report is logically structured and discusses the points in bold, mentioned below in the coursework description.*

The main requirement of the code associated with the coursework is for us, the lecturers, is to be able to follow clearly what steps you have done to achieve your results. We will not judge the level of sophistication/complexity of the code.

You are expected to submit code and associated material that demonstrates good software development practices as covered in the Research Computing module.

Your report should not exceed 3000 words (including tables, figure captions and appendices but excluding references); please indicate its word count on its front cover.

You are reminded to comply with the requirements given in the Course Handbook regarding the use of, and declaration of use of, autogeneration tools for both the code and report.

Search for planet candidate

Relevant files: `ex1_tess_lc.txt, ex1_stars_image.png`
Examine the lightcurve of the TESS target. You will see strong variability (rotational modulation in a spotty star, plus flares) which may be hiding planetary candidates.

1. **Analyse the lightcurve to identify the strongest period**
2. **Filter the lightcurve to search for any planet candidate(s).**
3. **Derive estimates of the properties of any of the systems that you find**
4. **Summarise your findings.**

Data

`ex1_tess_lc.txt` contains >31,000 epochs of TESS photometry for the star system. There are 4 columns of data (listed in the header):
column1: time (BJD-2457000)
column2: flux (relative)
column3: flag
column4: flux error
BJD stands for Barycentric Julian Day. The flags for this lightcurve are all set to 0 (i.e. there are no bad data).
`ex1_stars_image.png` is an HST high-resolution image of the target system, and shows there are 3 stars (spanning about 10 arcseconds) contributing to the light curve in the (unresolved) TESS image (TESS pixel size is 21 arcseconds). Analysis of the centroid shifts of the source during the transit confirms that the component A is the host of the candidate. The apparent magnitudes of the three sources, converted to the TESS bandpass are:
A 8.88 ± 0.02
BC 8.80 ± 0.02 (combined)
B 9.27 ± 0.07
C 9.92 ± 0.07
Parameters of component A (derived from K-band photometry, with Gaia parallax)
$M = 0.257 \pm 0.014 M_{\odot}$
$R = 0.268 \pm 0.027 R_{\odot}$

Planet detection and characterisation via RVs

Examine the Doppler observations of CB 01223.

1. **Do the RVs show any evidence of long-term trends or periodic variability? What about the stellar activity indicators (FWHM, bisector span)?**
2. **Make a plot to illustrate your exploratory findings, and suggest possible origins for any apparent (quasi-)periodic variability you find, e.g. planetary, stellar, sampling effects, etc.**

Next, implement models to address quantitatively the question of whether the Doppler data support the detection of any planets.

1. **Assume that there might be** 0, 1, **or** 2 **planetary signals present, and account for stellar variability (if applicable) using any technique(s) you deem appropriate. Be sure to justify all model parameter priors.**
2. **Produce posterior summaries for key parameters in your favoured model, and briefly discuss your overall findings. E.g., if you conclude that a planet(s) is detected, what can you say about its orbital and physical properties? If not, what do you infer about the nature and origin of the RV variability you've modelled?**

Data

Relevant files: `ex2_RVs.txt`.
Suppose we have obtained a series of high-resolution Doppler spectrographic measurements of a K-dwarf star called CB 01223[1].
The file `RVs.txt` is a CSV file containing 84 rows, each row corresponding to one observation of CB 01223, and each column (listed in the header) as follows.
column1: time [BJD; arbitrary zero point]
column2: radial velocity [km/s]
column3: —"— uncertainty [km/s]
column4: FWHM of the CCF [km/s]
column5: —"— uncertainty [km/s]
column6: bisector span of the CCF [km/s]
column7: —"— uncertainty [km/s]
column8: instrument name
As you will see in `RVs.txt`, the observations were made using four different spectrographs over a baseline of nearly twenty years. Assume instrumental systematics are negligible such that the observations can be treated as a single data set with heteroscedastic uncertainties.

[1]This name is entirely arbitrary and does not correspond to any real star. But it mirrors the typical (rather dry though systematic and catalogue-derived) names given to most stars. It seems easier here to refer to such a name than, say, "some hypothetical K-dwarf star".

Additionally, CB 01223 has been observed photometrically by TESS (data not needed for this exercise), on the basis of which a rotation period of $P_{\mathrm{rot}} = 31 \pm 10$ d, a mass of $M_* = 0.69 \pm 0.01\ M_\oplus$, and an effective stellar temperature of $T_{\mathrm{eff}} = 4500$ K have been derived.