Honors Seminar 259 — Extraterrestrial Life — Fall 2008

Guidelines for the mid–term project

For your mid–term project, you will undertake a quantitative analysis of data that I will provide; you will then write up a 2–3 page summary of your analysis. There are three options (described below), of which you should choose one; if you decide you would like to change projects, please let me know by October 17th, so that I can provide you with a new set of data. No two sets of data will be exactly identical. You may consult with classmates who are working on projects similar to yours, but you should write up your analysis on your own.

The written summary of your analysis should include several components:

1. A description of the data and your goal in analyzing them (in your own words).
2. A review of the steps you went through in your analysis. This section should address what software you wrote or used for your project (e.g., Excel, Mathematica, Maple, your own C++ code); how you applied it to the data; and how you overcame any notable hurdles.
3. A statement of your results (including "unexciting" negative results, if you found them; these are often scientifically valuable!).
4. An assessment of any *uncertainties* in your results. Possible questions to consider: Does your periodogram show more than one significant peak? Does the "trend" you see in your scatter plot depend heavily on the position of one, isolated point that might be wrong? Are your conclusions about the history of exoplanet detections limited by inconsistencies in how different groups have published their data?
5. A discussion of the conclusions that you can draw from the results– what possible explanation(s) can you think of for the pattern(s) that you see in the data?

Your summary should be complete but does not need to contain an overwhelming amount of detail (e.g., pages and pages of numbers). Take the 2–3 page target seriously, and concentrate your efforts on the analysis itself. **Very important:** you should also email me an electronic version of whichever spreadsheet, script, or program you use to analyze the data, so that I will be able to see which formulas and (if not included in your written report) plots you used to draw your conclusions. Both the written summary and the electronic supporting material are due **October 29th**. *Please do not wait until the last minute to start your analysis.* I am willing to help you with this assignment, but there is only so much I can do on the evening of October 28th.

Option 1: Planet detection in radial velocity data

Figures 2 and 3 of Marcy & Butler (2006) illustrate one of the main challenges of determining the parameters of exoplanets from radial velocity data. Plotting radial velocity vs. *time* as in Figure 2 (imagined without the fit to the data points) will initially show a confusing mess; it is only after determining an orbital period that astronomers can produce the much cleaner radial velocity vs. *orbital phase* plot of Figure 3. But how is an orbital period determined? The most straightforward strategy is to calculate the *periodogram* of the data, which reveals how strong a periodicity exists within the measured velocities at a range of possible orbital frequencies. Mathematically, if we have N measurements of velocities $\{v_i\}$ at times $\{t_i\}$ (where i ranges from 1 to N), then the periodogram at a particular frequency ω is defined as the sum

$$P(\omega) = \frac{1}{N}\left[\Big(\sum_{i=1}^{N} v_i(t_i)\cos\omega t_i\Big)^2 + \Big(\sum_{i=1}^{N} v_i(t_i)\sin\omega t_i\Big)^2\right] \tag{1}$$

A frequency ω will have typically have units of day^{-1} for $\{t_i\}$ measured in days. As discussed by Scargle (1982, *Astrophysical Journal*, 1982, **263**, 835), the appropriate range in frequency over which to compute the periodogram is from $\omega_{\rm min} = 2\pi\,(t_N - t_1)^{-1}$ to $\omega_{\rm max} = \pi\min(\Delta t)$, where $\min(\Delta t)$ is the smallest separation between successive velocity measurements $t_{i+1} - t_i$. The reason for considering this range of frequencies is that from a given dataset, we cannot draw any conclusions about orbits that change on timescales longer than that over which the data were collected, or shorter than that sampled by the data.

For this project, you will receive a list of ordered pairs $\{(t_i, v_i)\}$, which describe radial velocity measurements for a star whose name I will not provide. You can be sure that there is *at least* one planet orbiting the star (there may in fact be more). As part of your analysis, you should

- calculate and plot the periodogram for your data;
- identify a peak in the periodogram that corresponds to a possible orbital period;
- replot the radial velocities as a function of *orbital phase* for this putative orbital frequency, and decide whether you have found a planet or not.

If the plot of velocity vs. orbital phase looks clean and the curve is fairly sinusoidal, you may wish to consider subtracting a model for the orbit and looking at the residuals (i.e., the differences between the data points and the model curve). Going through the above steps for a single planet will count as successful completion of this project.

Option 2: Properties of stars that have planets

One of the key terms in the Drake Equation is the fraction of stars that have planets orbiting around them. The most obvious way to estimate this fraction is to select a large sample of *nearby* stars and determine how many of them have planets. However, since more distant stars are fainter and more difficult to observe directly, we would also like to have a way to guess whether a given star has planets based simply on the properties of the star itself (i.e., without having to go to the trouble of an exhaustive planet search). But are there any such properties, and if they exist, which are they?

For this project, you will receive a table containing for each of a large number of stars

1. the star's name;
2. the star's "effective temperature" (roughly speaking, the temperature of its surface), in K;
3. the logarithm of the star's surface gravity (in units of $\mathrm{cm\,s^{-2}}$);

4. the logarithm of the star's iron abundance relative to the iron abundance in the Sun (this would have a value of 0.00 if the iron abundance matches that of the Sun, 0.48 if the iron abundance is a factor of 3 larger than in the Sun, and −0.70 if the iron abundance is a factor of 5 smaller than in the Sun);

5. the star's apparent rotation velocity, $v \sin i$, in units of $\mathrm{km\,s^{-1}}$ (here i is the inclination of the *star's* rotation axis relative to our line of sight, not the inclination of a planet's orbit);

6. the star's mean radial velocity towards or away from us (positive or negative, respectively), in $\mathrm{km\,s^{-1}}$; and

7. the number of planets known to orbit the star.

As part of your analysis, you should

- determine whether the probabability that a star has at least one planet depends on effective temperature, surface gravity, iron abundance, apparent rotation velocity, and/or mean radial velocity;

- for each such relationship that you identify, create a plot that illustrates it, and assess whether the relationship is robust (i.e., would it go away altogether if you removed one or two key data points?); and

- determine whether effective temperature, surface gravity, iron abundance, apparent rotation velocity, and mean radial velocity are in any way *related to each other* within this dataset.

Some of the stars in your table may have multiple planets; you may therefore also wish to investigate whether any of a star's properties suggest it is likely to have *more than one* planet, although this exercise is not required for the successful completion of this project.

Option 3: Properties of exoplanets as a function of discovery date

As we have discussed in class, the first two exoplanets detected by radial velocity measurements had properties that were surprisingly different from those of the planets in our own solar system: they were "hot Jupiters", massive planets orbiting at very small radii around their central stars. These discoveries raised the question of whether our solar system is somehow unusual relative to other planetary systems, or whether these first two exoplanetary systems are somehow unusual relative to other planetary systems. One strategy for addressing this question is to consider how the properties of the set of "all known exoplanets" have changed as the number of "all known exoplanets" has grown larger.

For this project, you will receive a table containing for each of a large number of exoplanets

1. a name or equivalent identifier;

2. the year in which the planet was discovered; and

3. several parameters describing the properties of the planet and its orbit. The format of this section of the table will vary, but parameters are likely to include the period in days, the peak line-of-sight orbital velocity in $\mathrm{km\,s^{-1}}$, the eccentricity of the orbit (0 for perfectly circular, $0 < e < 1$ for elliptical), the minimum mass $M \sin i$ (for i the inclination of the planet's orbit) in $M_{\odot}$, and the semimajor axis of the planet's orbit in astronomical units. You will receive a detailed description of the contents of the table when you receive the data.

As part of your analysis, you should

- determine which if any parameters of the exoplanets and their orbits depend on the year of discovery;
- determine which if any parameters of the exoplanets and their orbits are related *to each other*; and
- for each relationship you identify, create a plot that illustrates it, and assess whether the relationship is robust (i.e., would it go away altogether if you removed one or two key data points?).

Several of the exoplanets in your table will comprise multiple-planet systems; you may wish to consider in particular whether and how the properties of these planets are related to the order in which they were discovered, although this exercise is not required for the successful completion of this project.