ASTM21: Project 4

Robin Emanuelsson

January 2020

Introduction

Periodic motion is a widespread phenomenon in physics. From a simple harmonic oscillator to planets orbiting stars. Usually when one samples data from such a phenomena it's evenly spaced data but this need not to be the case. One such case is when you want to measure the radial-velocity of stars. The measurements could be spaced unevenly from a number of different reasons such as the weather at the telescope that was intended to measure the radial-velocity. When dealing with unevenly spaced data one technique is the Lomb-Scargle periodogram.

In this report we will take data from three stars and estimate the period and velocity amplitude, which tells us how strongly the star is tugged upon by its planet companion, of these signals.

Theory

The derivation for the Lomb-Scargle periodogram formula is long and not necessary for this report. For a full derivation see [1]. The Lomb-Scargle normalized periodogram (spectral power as a function of angular frequency) is

$$P_N(\omega) = \frac{1}{2\sigma^2}\left[\frac{\left(\sum_i y_i \cos(\omega(t_i-\tau))\right)^2}{\sum_i \cos^2(\omega(t_i-\tau))} + \frac{\left(\sum_i y_i \sin(\omega(t_i-\tau))\right)^2}{\sum_i \sin^2(\omega(t_i-\tau))}\right] \tag{1}$$

where σ is the variance, h_i is the data and t_i the corresponding time.
τ is given by

$$\tan(2\omega\tau) = \sum_i \frac{\sin(2\omega t_i)}{\cos(2\omega t_i)} \tag{2}$$

[2]. However this formula does not take into an account the individual errors. For this we need the generalised Lomb-Scargle periodogram. Again, the whole derivation of this formula is beyond the scope of this report, but it can be found here[3]. To account for the individual errors we need the normalized weights

$$w_i = \frac{1}{W}\frac{1}{\sigma^2}, \quad \left(W = \sum_i \frac{1}{\sigma_i^2}, \quad \sum_i w_i = 1\right) \tag{3}$$

The following abbreviations will be crucial

$$Y = \sum_i w_i y_i \tag{4}$$

$$C = \sum_i w_i \cos(w_i t_i) \tag{5}$$

$$S = \sum_I w_i \sin(w_i t_i) \tag{6}$$

$$YY = \hat{YY} - YY, \quad \hat{YY} = \sum w_i y_i^2 \tag{7}$$

$$YC = \hat{YC} - YC, \quad \hat{YC} = \sum w_i y_i \cos(\omega_i t_i) \tag{8}$$

$$CC = \hat{CC} - CC, \quad \hat{CC} = \sum w_i \cos^2(w_i t_i) \tag{9}$$

$$YS = \hat{YS} - YS, \quad \hat{YS} = \sum w_i y_i \sin(\omega_i t_i) \tag{10}$$

$$SS = \hat{SS} - SS, \quad \hat{SS} = \sum w_i \sin^2(w_i t_i) \tag{11}$$

Using these abbreviations the periodogram will be

$$p(\omega) = \frac{N-1}{2} \frac{1}{YY} \left[\frac{YC_\tau^2}{CC_\tau} + \frac{YS_\tau^2}{SS_\tau} \right] \tag{12}$$

where $\frac{N-1}{2}$ is for normalization and N is the number of data points. The subscript τ indicates that t in all the abbreviations have become $(t_i - \tau)$, where τ is give by the same relation as before[3].

When computing the periodogram one need to look in some frequency interval. Since the data is unevenly space it's not obvious what range this should be. Eyer and Bartholdi, in 1999, proved that

Let p be the largest value such that each t_i can be written $t_i = t_0 + n_i p$, for integers n_i. The Nyquist frequency then is $f_{Ny} = 1/(2p)$

This is not easy and in fact impossible if spacings are irrational. A simple but not as accurate way is simply to take the the harmonic mean of the sampling intervals[1].

As always when dealing with statistics you need a way of verifying the results and in this case a way to measure how significant a peak is in the power spectrum. A convenient property of the periodogram is that the null hypothesis, that is that the data values are independent Gaussian random values, can be tested rigorously. It can be shown that the probability that the power spectrum is between some z and $z + dz$ is $exp(-z)dz$. For M independent frequencies, the probability that none give values larger than z is $1 - (1 - e^{-z})^M$. So we get

$$P(> Z) = 1 - (1 - e^{-z})^M \tag{13}$$

Now we also need to estimate M. There are number of ways to estimate it but the easiest, since we have such a low amount of data points, is to say that M is equal to the number of data points.[2]

Results

For all the stars a spacing with a 10000 steps in the frequency domain was used.

Star: hd000142

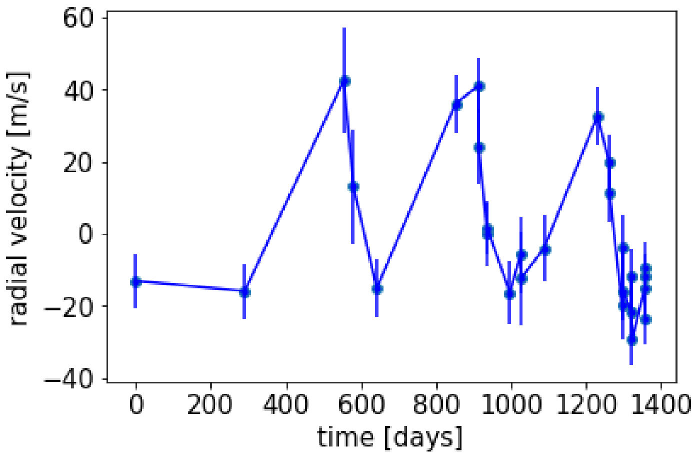

Figure 1: Plot of the radial-velocity as function of time, with the data points corresponding error bars

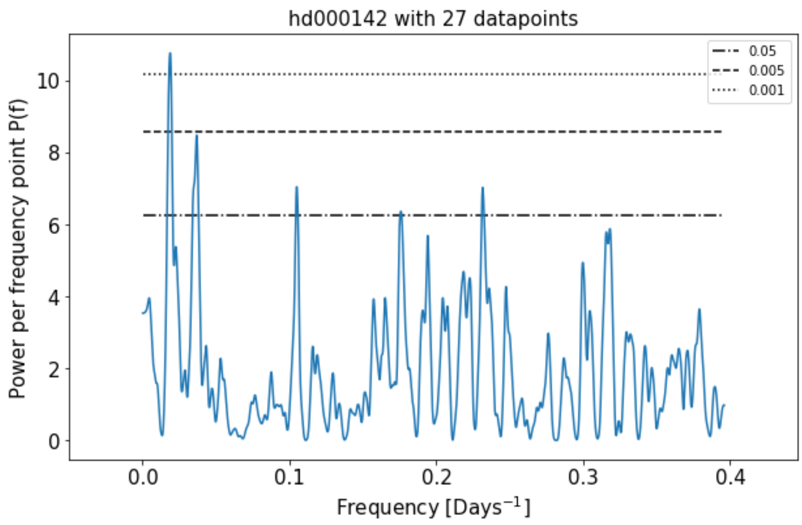

Figure 2: Plot of power spectrum as function of frequency. The dotted lines are the corresponding significance levels.

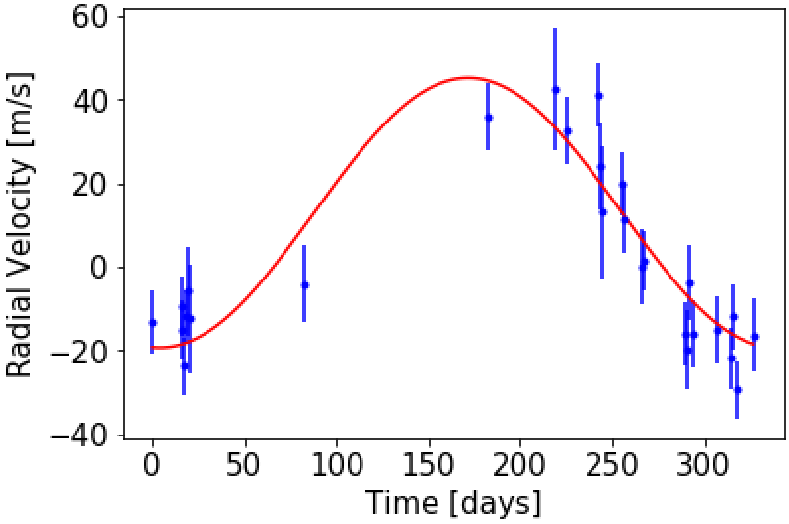

Figure 3: Plot of the folded data and a fitted sinus wave

Figure 1 shows the radial-velocity data points distributed, in an uneven fashion, over some days. Figure 2 shows the generalized Lomb-Scargle periodogram. There's a clear peak at roughly 0.02, over the significance level $p < 0.001$. The period is 53.34 days up to precision of two significant digits. Figure 3 shows the folded data and their corresponding errors and with a fitted sinus wave. The fitted waves amplitude is $32.3 ^m/_s$.

Star: hd027442

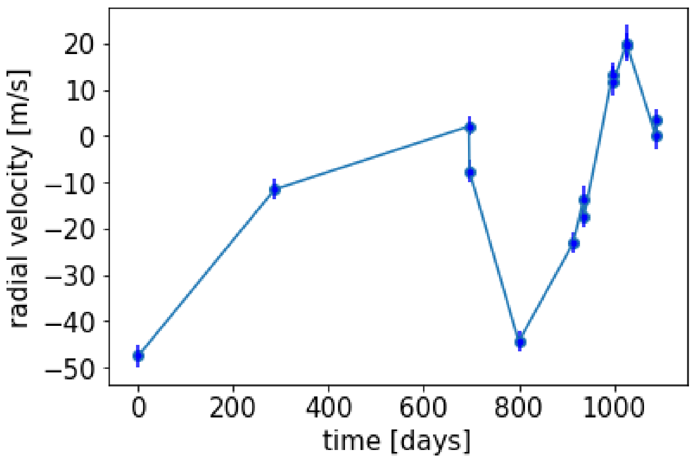

Figure 4: Plot of the radial-velocity as function of time, with the data points corresponding error bars

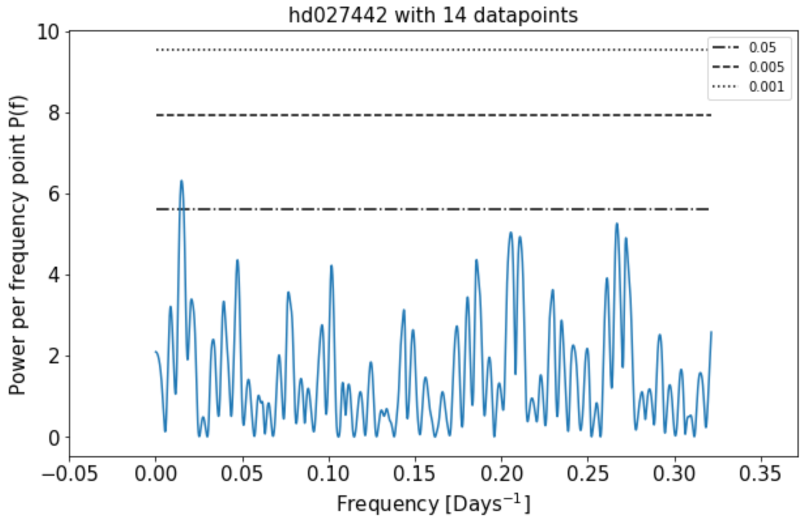

Figure 5: Plot of power spectrum as function of frequency. The dotted lines are the corresponding significance levels.

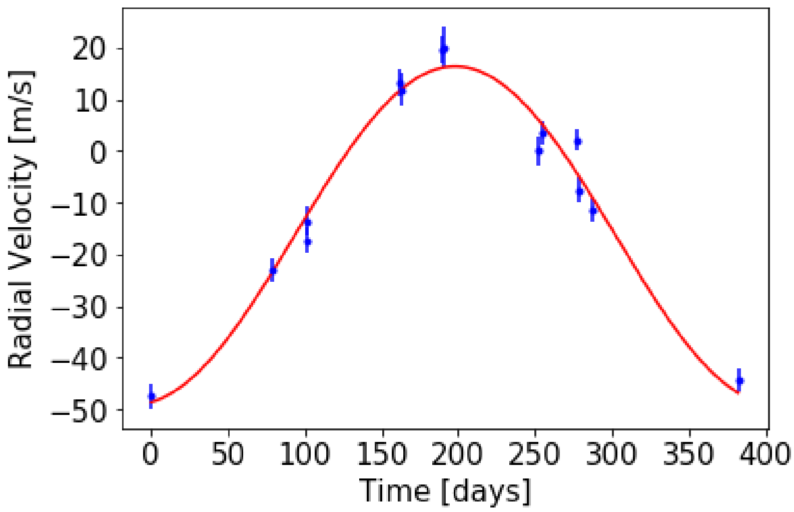

Figure 6: Plot of the folded data and a fitted sinus wave

Figure 4 shows the radial-velocity data points distributed, in an uneven fashion, over some days. Figure 5 shows the generalized Lomb-Scargle periodogram. The most significant peak is roughly at 0.0015 just slightly over the significance level $p < 0.05$. The period is 66.48 days up to precision of two significant digits. Figure 6 shows the folded data and their corresponding errors and with a fitted sinus wave. The fitted waves amplitude is $32.6^{m}/_{s}$.

Star: hd102117

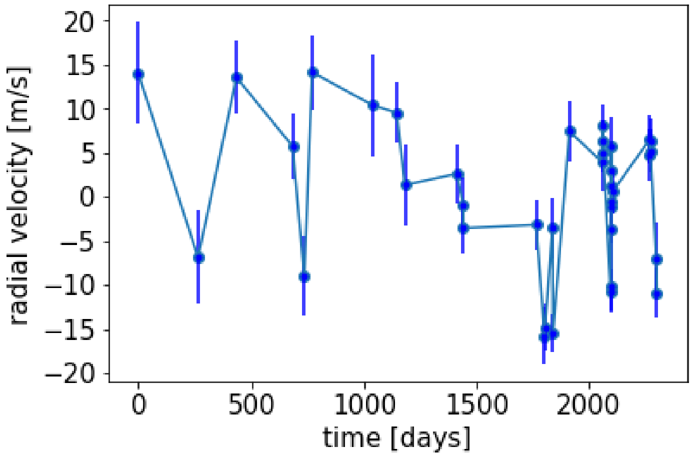

Figure 7: Plot of the radial-velocity as function of time, with the data points corresponding error bars

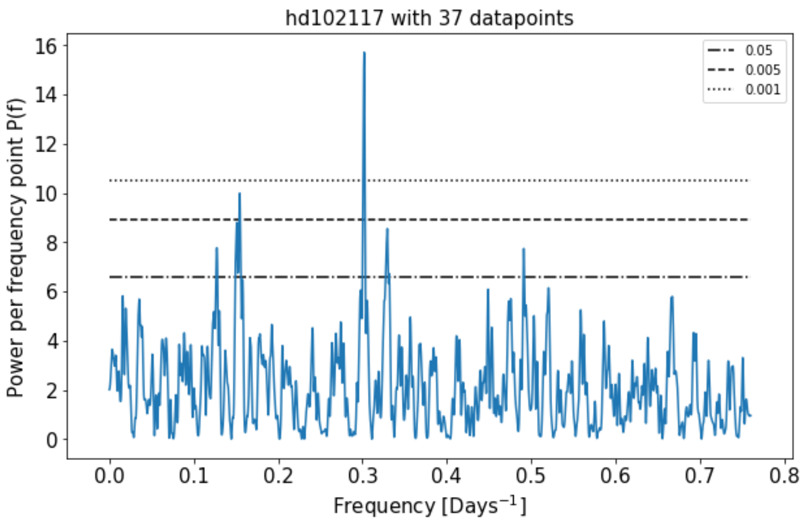

Figure 8: Plot of power spectrum as function of frequency. The dotted lines are the corresponding significance levels.

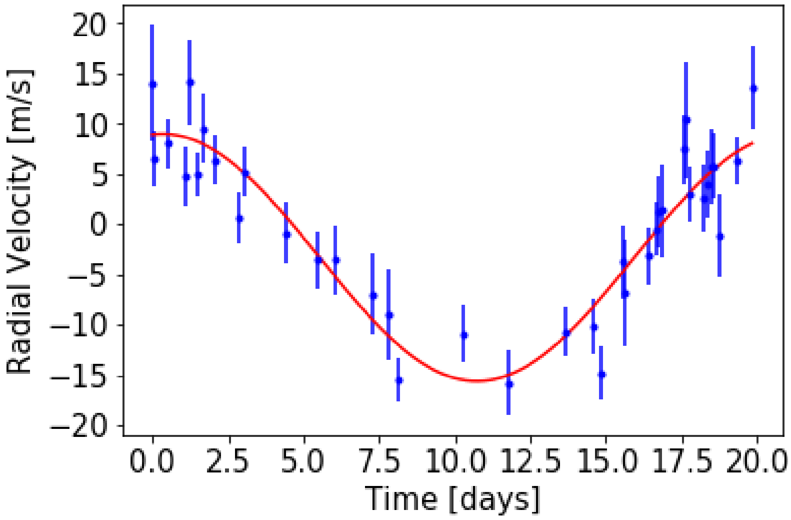

Figure 9: Plot of the folded data and a fitted sinus wave

Figure 7 shows the radial-velocity data points distributed, in an uneven fashion, over some days. Figure 8 shows the generalized Lomb-Scargle periodogram. There's a clear peak at roughly 0.3, significantly over the significance level $p < 0.001$. The period is 3.31 days up to precision of two significant digits. Figure 9 shows the folded data and their corresponding errors and with a fitted sinus wave. The fitted waves amplitude is $12.2^{m}/_{s}$.

Discussion

The results in term of their significance seems to be, atleast in this case, very correlated with the number of data points. For one of the stars the were only 14 data points and in the periodogram there was no obvious clear peak like for the rest of the stars, even though its errors were relatively small. And the highest peak was not much larger than the other peaks and did not have a very high significance. Another interesting thing with the star with few data points is that the most significant peak changed when using the normal Lomb-Scargle periodogram instead of the generalize one. See Appendix A. This goes to show that the individual is something that really needs to be considered.

Speaking of the significance levels, which depends on M. How M was chosen was not chosen in a cleaver way, like the number of peaks, in stead just the number of data points. The reason for this is that there was found (not presented in this report) that there was no real significant difference if M was chosen in such a manner like counting the number of peak.

For two of the starts the estimated frequency was very low which indicates that the star is not moving much and one star stood out, with respect to the other two, and had significantly larger frequency then the other two. However all the amplitudes where in the same order or magnitude.

References

[1] J. T. VanderPlas, 'Understanding the Lomb–Scargle Periodogram', University of Washington, eScience Institute, Seattle, (2018).

[2] W. H. Press, et. al., 'Numerical Recipes', Cambridge University Press, (2007)

[3] M. Zechmeister, et. al., 'The generalised Lomb-Scargle periodogram', Max-Planck-Institut für Astronomie, (2009)

Appendix A

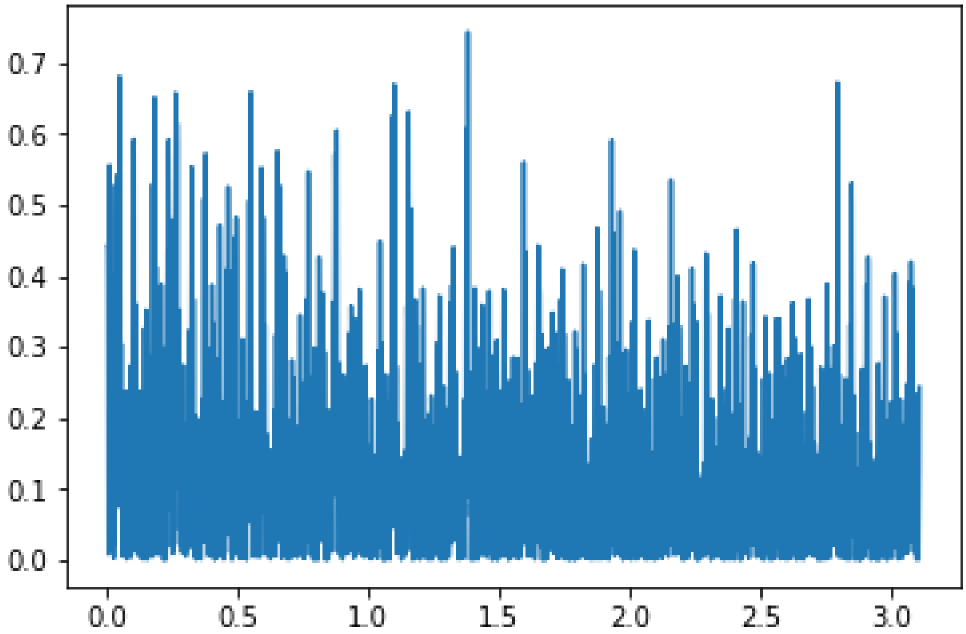

Figure 10: Lomb-Scargle periodogram of the hd027442 star using the scipy package