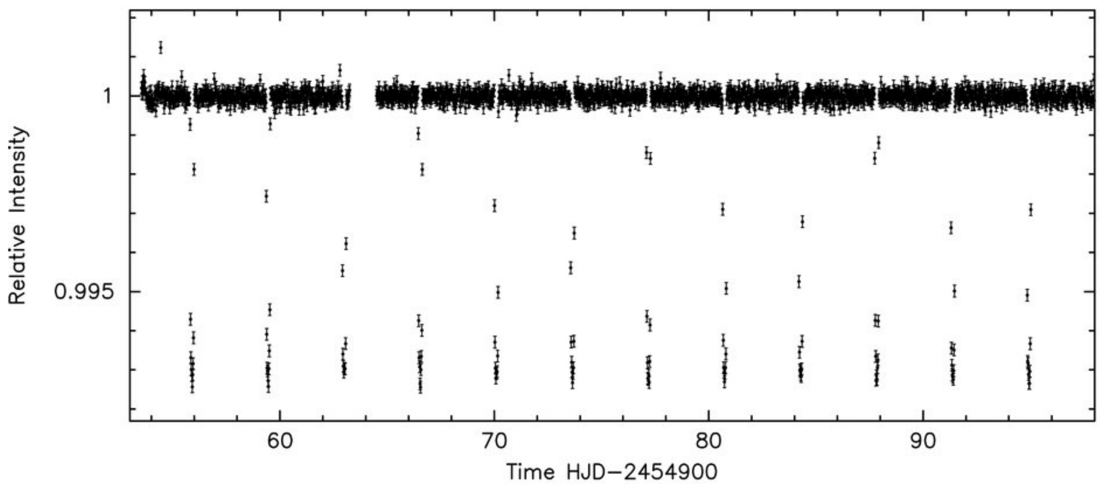
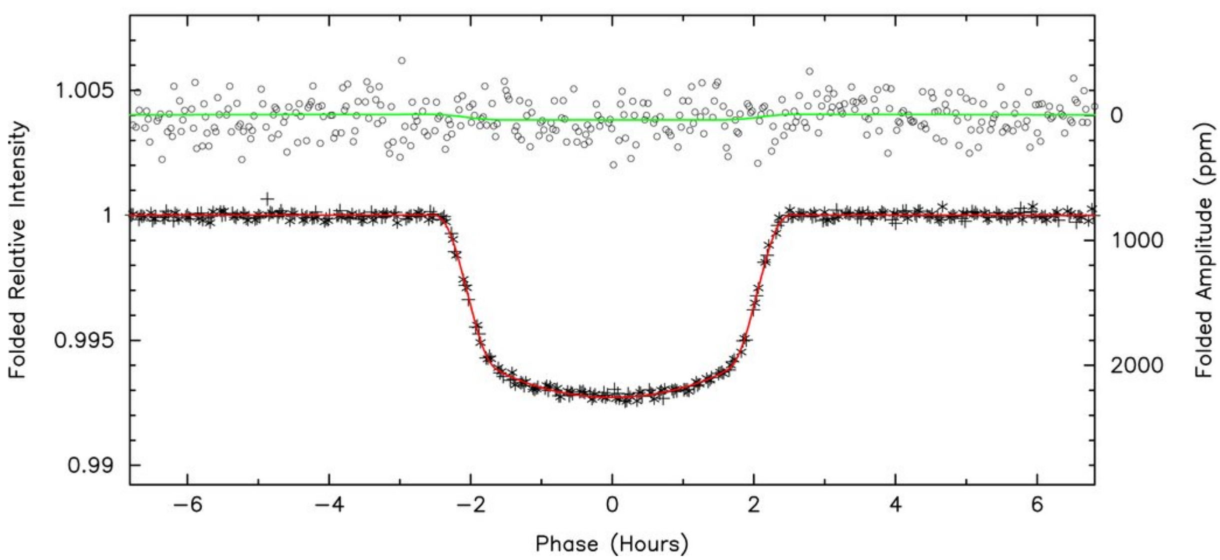
ASTP 720 - Homework 7 - The Exoplanet Kepler-5b

Due Date: October 22nd, 2020

You're going to figure out some of the properties of the planet Kepler-5b, one of the first exoplanets discovered by NASA's *Kepler* mission. *Kepler* looked at a single spot in the sky and continuously monitors stars in the region. The focus was originally on stars similar to the Sun, in order to find planets hopefully similar to Earth, but we are now finding a wide variety of planets around a wide variety of stars in a wide variety of configurations.

Below are two plots of real data of the planet Kepler-5b (Koch et al. 2010). You will use a similar data set but the figure illustrates the rough idea of what you'll need to do.

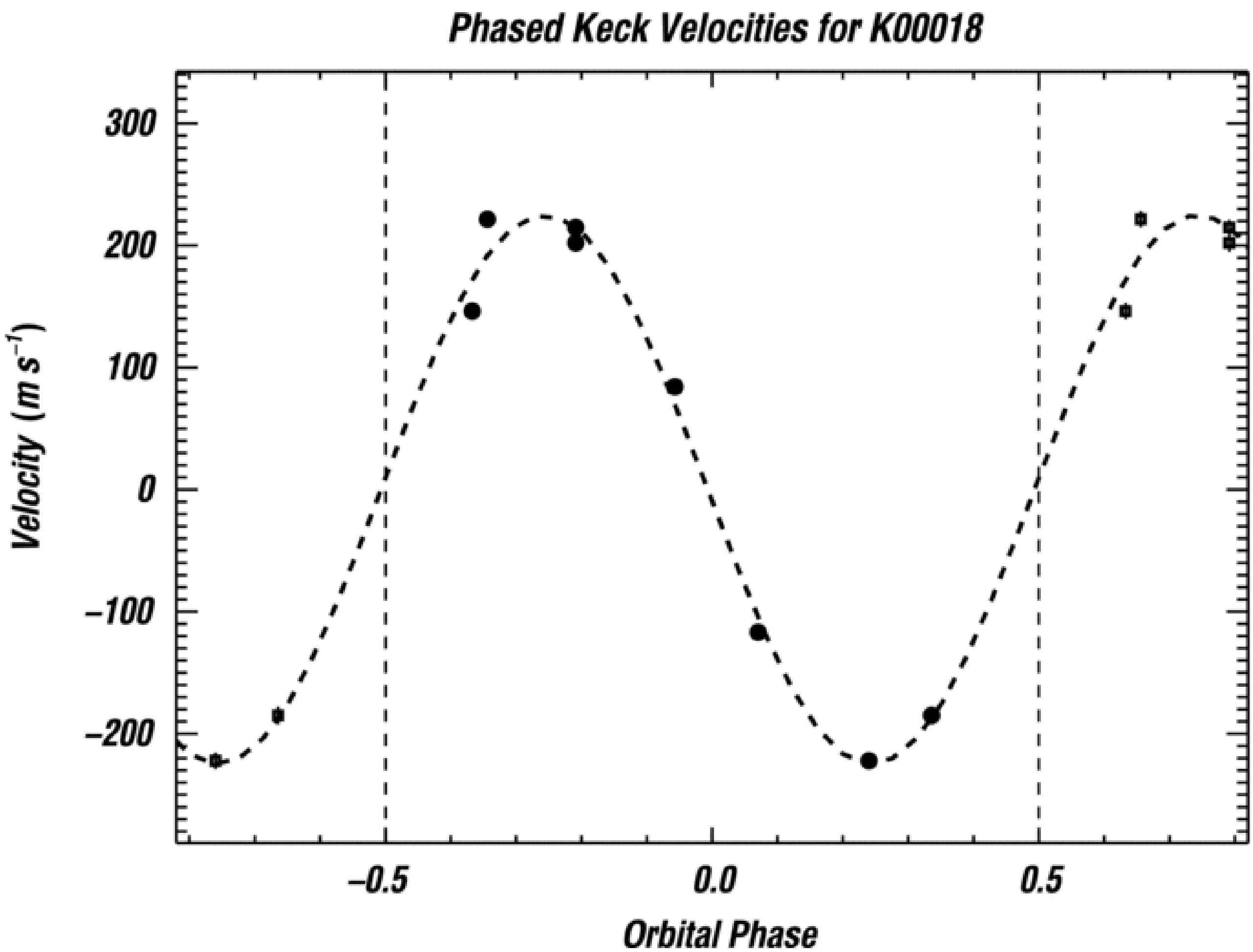




The above plot shows 12 transits of the planet Kepler-5b. The upper plot shows the lightcurve versus time, referenced to Heliocentric Julian Day, a heliocentric correction to the Julian Day (MJD equals JD - 2400000.5, for reference). The downward spikes are very obvious indications of a regular transit.

If we average the lightcurve by *folding* it at the appropriate orbital period, we see the plot on the bottom. The lower plot shows the average lightcurve of a transit (averaged over the 12 measured). The top line is an indication of the errors (on a different scale), which you can ignore. You can see that their model, in red, fits the average transit quite well.

In addition to transit data, Koch et al. (2010) showed the radial velocity of the start as a function of orbital phase. That is, they have already folded the measurements with the known period.



Using both transit and radial velocity measurements, one can determine the mass and radius of the planet, thus setting the planet's density and thus compositon. In addition, one obtains the orbital radius of the planet, setting it's location in the star system.

For your homework, your goal will be to determine the properties of Kepler-5b given a *Kepler* lightcurve data set (`lightcurve_data.txt`), and a Keck-HIRES radial velocity data set (`RV_data.txt`) for bonus points.

Kepler Lightcurve Data

Your primary analysis will be on the lightcurve. From the folded lightcurve, you can see the effects of the ingress/egress from the slanted "walls" along with the bowl shape due to limb darkening. Instead, you should consider a much simpler model, one in which you have a (negative) rectangular/boxcar function[1] at the time of every transit. That is, you might have a piecewise defined function that is equal to some negative value during the dip and zero outside of the dip.

Therefore, you can model the lightcurve signal as follows. The primary parameters that you need to consider are the period P, the width of the transit τ, the time of the first transit $t_{\rm ref}$ (defined to the start or the center of the dip or some such), and the drop in intensity (the amplitude of the dip) ΔI. So, given data in the form of times t_i and intensities I_i, a rough sketch of the likelihood might be

$$\mathcal{L}(P,\tau,t_{\rm ref},\Delta I|t_i,I_i) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\left(\frac{I_i - s(t_i|P,\tau,t_{\rm ref},\Delta I)}{\sigma}\right)^2\right], \tag{1}$$

where the signal $s(t|P,\tau,t_{\rm ref},\Delta I)$ is the periodic dip function that you define.

While one can search over the period, to simplify this problem for you, I will give you the period of the orbit: $P = 3.5485$ days. In order to make a folded lightcurve, what you can do is take the timeseries, subtract off the first time, and then take the time modulo the period (in Python, using the `%` symbol), i.e., you can define a new set of *folded* times as $t_i' = (t_i - t_0) \mod P$. This will get you something that looks kind of like the bottom plot of Figure 1, though not centered on 0.

The lightcurve measurements do not have error bars, so you may treat these in one of a few ways in order to find the best parameters. One, you could simply ignore them ($\sigma = 1$) since by eye they are roughly equal anyway and you only need to maximize $\mathcal{L}$. Two, you could take a sample of the data when the transit is not occuring, calculate the root-mean-square value, and set that equal to σ. Three, you could actually let σ be a *parameter* to *model* your the noise. In this case, the definition of the likelihood would remain the same on the right-hand side but the left would be

$$\mathcal{L}(P,\tau,t_{\rm ref},\Delta I,\sigma|t_i,I_i). \tag{2}$$

I leave this to you, but you should describe your method in your write-up.

The drop in intensity is related to the planet's radius R_p and the star's radius R_* as the ratio of the areas of the disks, i.e.,

$$\frac{\Delta I}{I} = \frac{\pi R_p^2}{\pi R_*^2}. \tag{3}$$

Therefore, the fractional dip size tells you the radius of the planet if you know the radius of the star. For reference, Kepler-5 is an F star with mass $1.35 M_\odot$ and radius $1.79 R_\odot$.

[1] If instead we fit a trapezoid, that is, we had a different width for the top of the transit and the bottom of the transit, we could measure the inclination of the orbit. I'm happy to show notes if you are interested.

[1.] Your first task is to solve for the parameters you choose (it is your signal to model!) via MCMC. Your end goal here is to determine the radius of Kepler 5-b. If you are feeling adventurous, you can try to make one of those fancy corner plots, but if you can show the one-dimensional marginalized distribution for the main variable that determines the true physics of the system (ΔI) and quote your measurement and its uncertainty (MAP? median? some confidence interval range of your choosing?), that will be perfectly fine. Again, please just describe what approach(es) you have taken.

[Bonus.] Keck-HIRES Radial Velocity Data

The radial velocity data can be fit with a sine wave with period P (which I have given), phase offset ϕ, and amplitude $v_{\max}$ (assuming a perfectly edge-on system). From conservation of momentum, we know that the momentum of the star is equal to the momentum of the planet, which is most easily determined when the star is moving towards or away from us. The mass of the star is given above. The velocity of the star is what is measured. The velocity of the planet can be determined from Kepler's Third Law (first solving for the semi-major axis of the orbit, which you can assume is circular). Thus, this technique will yield the mass of the planet, and combining both measurements will yield the density of the planet.

Rather than perform an independent fit to the radial velocity data, from above, you have information on what the period is of the orbit, and also when the planet should be at different parts of its orbit (from $t_{\rm ref}$, but think about the geometry!). Instead of analyzing the likelihood, use this *prior* information to find the best parameters of the posterior distribution. That is, for example, consider some prior on P (constant!) and ϕ. Feel free to use your empirical distribution what ϕ might be from $t_{\rm ref}$, or an approximation (Gaussian, given the estimated value of $t_{\rm ref}$ and its uncertainties), but in your write-up, please just state what you have done.

The end question you should attempt to answer is: what kind of planet is Kepler-5b?