Python Assignment

In this assignment you will use `emcee` in python (`http://dfm.io/emcee/current/`) or on github. You will simulate a periodic data set and fit a function to it. This could be e.g. a photometric dataset from Kepler, or a series of radial velocity points. Some skeleton code (with many gaps!) is included on the course website.

1. Create a function using python and numpy that simulates data that take a periodic function with a form:

$$v = a_0 + a_1 t + a_2 \sin(a_4 t) + a_3 \cos(a_4 t) \tag{1}$$

   You should simulate data at a number of random times over an interval, and include Gaussian errors for the data. The inputs a_i should take the form of a 1-dimensional python array.

2. Setting $a_0 = 0$, $a_1 = 1$, $a_2 = 1$, $a_3 = 1$ and $a_4 = 0$, simulate a data set from times $t = 20$ to $t = 35$, containing 100 points with Gaussian errors with uncertainty 0.5.

3. Use emcee to fit to this dataset. Plot histograms of the fitted parameters - do the results make sense? Are any of the parameter fits correlated?

4. (advanced) Show that the following is a re-parameterisation of Equation 1:

$$v = a_0 + a_1 t + a_2 \sin(a_3 t + a_4) \tag{2}$$

   Which is better - equation (1) or equation (2) for a reliable run of `emee`, and why? Is there a way the equation could have been re-parameterised to remove the correlation between a_0 and a_1?

5. (extra mark) If Equation (2) is your model with uniform priors in all parameters but Equation (1) is used in `emcee` instead, this produces an implicit prior on a_2. What is it?

Include all python code in your assignment.