Exercise #1 (assigned 4 February, due 11? February)

Download `exercise1.dat` from the course web page, and use it to answer the following questions. You should also download the Burgasser paper from the course web page. Note: For this exercise, I want you to use equations and write things out and solve for things — don't just use canned computer or calculator routines.

(1) You know nothing about the data in `exercise1.dat` other than that it is a measurement of a phenomenon that probably contains some kind of periodic signal.

To make a measurement with this experimental apparatus costs money. The cost is related to the sampling frequency in the following way:

$$c = 10/\Delta t$$

where c is the cost in thousands of dollars and Δt is the time elapsed since the last measurement. Clearly, the cost goes down the longer you wait.

Show that the best use of your grant funding to determine the approximate signal in the data is none of fine sampling, medium sampling, or coarse sampling, but logarithmic sampling. By "show" I don't mean with a logical proof — I mean that you should use the provided data and sample it at these various frequencies, and see how well you can determine the underlying signal.

This data set has a true signal, a systematic error, and random errors. What are the magnitudes of these three types of errors in this data set?

One way to understand the properties of the random errors is to remove the signal(s) from the data, leaving *residuals*. What is the standard deviation of the residual (random) signal?

What is the probability of having a measurement that is 1σ from the true signal? What about 2σ? 2.5σ? 3.2σ? In this dataset, how many measurements should exist at each of these sigma levels? Make a histogram of the deviations of the random errors from true signal in this data set. Does it look like you expected? Why or why not?

Now imagine that instead of 100 measurements for this data set you had 10,000,000 measurements. (Your funding portfolio has done really well this year.) At what sigma would you believe an outlier to be real, that is, not simply due to random statistical fluctuations? What is the number of measurements that you would expect between 3σ and 4σ for this much larger data set? How many $+3\sigma$ measurements will there be, and how many -3σ?

(2) Adam Burgasser did what few other people do, which is present and discuss a well-reasoned expression for errors in a scientific paper (Burgasser et al. 2003, ApJ, 586, 512; see the appendix). Burgasser et al. present the binary fraction in their paper as $0.20^{+0.17}_{-0.07}$. Using the formalism laid out in that appendix, what would the binary fraction and error bars be if they had found 1 binary among their sample of 10 targets? What if they had found zero binaries among their sample of 10 targets?

What would the error bars be if they had found 4 binaries out of 20 targets? What about 8 binaries out of 40 targets? At what sample size does binomial become within 1% of gaussian?