Coding in advance of the Feb 12, 2018 class

For next time, I would like you to write codes to compute the chi squared, and the Poisson log likelihood, for one-dimensional data.

To do this, suppose that we have n data bins, the ith of which has d_i observed data counts, and that in a particular model we would expect m_i counts in bin i. Then we compute χ^2 using

$$\chi^2 = \sum_{i=1}^{n} \frac{(m_i - d_i)^2}{d_i} \ . \tag{1}$$

This formula *assumes* that the statistics are Gaussian, and that the standard deviation is $\sqrt{d_i}$. Note that Pearson's chi squared (which is what is often meant by "chi squared") actually assumes that the standard deviation is $\sqrt{m_i}$, but in astronomical applications it is common (but not correct!) to associate uncertainties with the data as we do above. So that's what we'll calculate.

The Poisson *likelihood* (not yet the log likelihood) is given by

$$\mathcal{L} = \prod_{i=1}^{n} \frac{m_i^{d_i}}{d_i!} e^{-m_i} \ . \tag{2}$$

As we'll learn when we discuss Bayesian statistics, unlike χ^2, the value $\mathcal{L}$ has no independent meaning; it is the *ratio* of $\mathcal{L}$ for one parameter combination to that of another that has meaning. Because $\prod_{i=1}^{n}(1/d_i!)$ is the same for any parameter combination, it therefore cancels in such ratios, and we can thus write

$$\mathcal{L} \propto \prod_{i=1}^{n} m_i^{d_i} e^{-m_i} \ . \tag{3}$$

This number can be either gigantic or tiny, and thus it is more convenient to work with its log, and to use differences in log likelihoods:

$$\ln \mathcal{L} = C + \sum_{i=1}^{n} \left[d_i \ln(m_i) - m_i \right] \ , \tag{4}$$

where C is a constant that never enters the calculations; thus we can set $C = 0$.

Please write codes to compute χ^2 and $\ln \mathcal{L}$ for any data set d_i and model expectations m_i; we will use these codes, and incorporate them into more involved codes, throughout the course. In particular, it will be important for you to use your χ^2 and $\ln \mathcal{L}$ codes in other codes, so please write them as easily-incorporated subroutines (where the inputs are the data vector d_i and the model vector m_i).

In particular, use the data sets on the website (data3_1.txt, data3_2.txt, data3_3.txt, and data3_4.txt) to estimate a single parameter for rolls of a die. Our model is that the probability of getting a 1 is $1-p$, and the probability for getting a 2, 3, 4, 5, or 6 is $p/5$. Let our prior probability density be uniform in p, from $p=0$ to $p=1$. Based on each data set, what is the posterior probability density for p if we use the Poisson likelihood? How about if we use Wilks' Theorem, where we look for $\Delta \ln \mathcal{L} = -0.5$ from the maximum for the 68.3% credible region? How about if we use χ^2, where we would use $\Delta\chi^2 = 1$ from the minimum for the 68.3% credible region? Note that the χ^2 calculation can in this case be performed analytically, but I recommend that you save time and do it numerically. What conclusions do you draw?