Coding in advance of the Feb 26, 2018 class

For next time, I would like you to write codes to compute the Bayes factor between two models of the probabilities of different numbers that appear when a die is rolled. These models are the same as in the second case we treat in Lecture 5. In each case, please normalize your model so that the expected total number of rolls is the actual total number of rolls.

In Model 1 the probability is equal to 1/6 for every number. Given our assumed normalization, this model has no parameters.

In Model 2 the probability of a 1 is $1 - p$, and the probability of a 2, 3, 4, 5, or 6 is $p/5$. Given our assumed normalization, this model has a single parameter, p, and our prior will be that p can with equal probability be anywhere from 0 to 1.

Your task is to write a code that will output the Bayes factor $\mathcal{B}_{12}$ for any set of rolls, and then to apply that to find the Bayes factors for the four data sets (which are the same as they were for coding task 3, but which are reproduced here for convenience). Note that although you could in principle expand the polynomial as we did in the notes, when there are many 1's in the rolls, the expansion of $(1 - p)^n$ becomes unwieldy. Thus you will need to find a better way.

As a second approach, we will use χ^2. To do this:

1. Compute the χ^2 for the data using Model 1.
2. Compute the χ^2 for the data using the value of p that *minimizes* χ^2 for Model 2.
3. Take the difference, and call that $\Delta\chi^2$. Note, as in the Lecture 5 notes, that Model 2 *cannot* have a larger minimum than Model 1, because Model 2 contains Model 1 (just set $p = 1/6$).
4. For one parameter, just take the square root of $\Delta\chi^2$ to find the significance in units of σ. For example, if $\Delta\chi^2 = 9$, then you need the extra parameter at the 3σ level. Note that this works only for *nested* models; if the two models were distinct from each other, you could not use this approach. Also note that if you have two or more parameters then the value of $\Delta\chi^2$ for a given significance is different; for example, if you have three parameters then $\Delta\chi^2 = 3.53$ is needed for 1σ.
5. For larger and larger numbers of rolls, how do the two types of model comparisons compare with each other?