Stellar Populations
Homework #1 - Bayesian Statistics

Ned Molter

September 8, 2016

Problem 1

Question: M&M color distribution changed from 1994 to 1996. These distributions given as percentages for both years. Suppose two bags of M&Ms, one from 1994 and one from 1996. You are given one M&M from each bag. One is yellow, one is green. What is the relative prob. that the yellow one is from the 1994 bag?

Solution This problem is easiest if you treat the two draws as two independent experiments, and then just multiply them later. That is, reaching blindfolded into one of the two bags (without knowing which bag is which) and pulling out a single M&M is an experiment. Thus the hypotheses just ask 'which bag did that M&M come from?' rather than anything about M&M color. In this framework, the draw in which we pulled a yellow M&M can be represented by the following table:

possible values θ	prior $p(\theta)$	likelihood $p(x\|\theta)$	prior × likelihood $p(\theta)p(x\|\theta)$	posterior $p(\theta\|x\|)$
1994 bag	0.5	0.2	0.1	0.59
1996 bag	0.5	0.14	0.07	0.41

Table 1: Draw of yellow M&M from an unknown bag

The prior probabilities are just 50-50 because we have no idea which of the two bags we drew from. The likelihood that a particular bag was drawn given that we got a yellow M&M (the likelihood) is then just given by the fraction of yellow M&Ms in that bag.

Applying the same framework to the other draw, which produced a green M&M, we have

possible values θ	prior $p(\theta)$	likelihood $p(x\|\theta)$	prior × likelihood $p(\theta)p(x\|\theta)$	posterior $p(\theta\|x\|)$
1994 bag	0.5	0.1	0.05	0.33
1996 bag	0.5	0.2	0.1	0.66

Table 2: Draw of green M&M from an unknown bag

Now we can simply multiply the posterior probabilites from the two tables, since the draws were independent. We are told that one M&M came from each bag, so we only need to consider two scenarios: yellow from '94 × green from '96 and green from '94 × yellow from '96. These are $0.59 \times 0.66 = 0.39$ and $0.41 \times 0.33 = 0.14$, respectively. Normalizing, we get $p = 0.74$ and $p = 0.26$ respectively. This means that **the yellow M&M is roughly three times more likely to have come from the 1994 bag.**

See the .ipynb for problems 2 and 3