Progress Report: Week 1

Reading:

This week, I read chapter 4 of *Think Bayes*. This chapter dealt primarily with how to estimate priors if they are not equally likely. I thought that this chapter made logical sense, and I like that the chapters of this book aren't that separated from each other in discrete topics like some other textbooks, but that we just do more and more complicated, and realistic, problems and learn what we need to do those problems.

I also read Downey's blog post "Bayesian solution to the Lincoln index problem". This blog post included a solution to the Lincoln index (the scenario where two testers search for bugs and find some in common) using Bayesian reasoning. I found the blog post easy to follow, but I'm still a little curious about whether a Bayesian solution is always the best one.

In this vein, I did some googling, and found the stack exchange question "List of situations where a Bayesian approach is simpler, more practical, or more convenient". (Link: http://stats.stackexchange.com/questions/41394/list-of-situations-where-a-bayesian-approach-is-simpler-more-practical-or-more) This post contains numerous examples of when a Bayesian approach is a good idea. I am not sure whether this necessarily answered my question of whether or not Bayesian is always better, but it did provide a lot of good examples when Bayesian reasoning was a good way to solve a problem.

I was also curious about how Bayesian statistics could be applied. In his guest blog post, (Link: http://andrewgelman.com/2014/01/21/everything-need-know-bayesian-statistics-learned-eight-schools/) Phil uses a Bayesian approach to model the Radon distribution across various counties. I thought that this was an interesting approach to using Bayesian statistics, and would recommend this post as an example of how Bayesian statistics can be applied to more real-world problems.

I read chapter 5 of *Think Bayes.* This chapter talked mostly about odds. I thought that all the examples in this chapter were informative, and that they were easier to understand. I think that this was partially because this was a conceptually easier chapter, but also because we had visited a lot of these examples in class.

Exercises:

I completed exercise 4-1 from *Think Bayes*. The class I implemented was:

```
class Euro (thinkbayes2.Suite):

    def __init__(self, hypos, y):

        """Initialize self.

        hypos: sequence of string bowl IDs
```

```
        """
        self.y = y
        thinkbayes2.Pmf.__init__(self)
        for hypo in hypos:
            self.Set(hypo, 1)
        self.Normalize()

    def Likelihood(self, data, hypo):
        x = hypo
        if data == 'H':
            return self.y*(x/100.0)
        else:
            return self.y*(1 - x/100.0)
```

Here y is the probability that the instrument misreported whether something was heads or tails. It turns out that no matter what y is, the posterior distribution is still the same. I think that this is because the same number of reported heads and tails that are falsely reported cancel each other out.

Case Study:

This week, I have been doing some poking around trying to find an interesting topic for a case study. I'm interested in estimation problems, so for example how scientists, or others, estimate the populations of endangered species, and whether this might be better done with a Bayesian approach. To that effect, I read this website (http://www.cals.ncsu.edu/course/fw353/Estimate.htm) which has information about how people might estimate the populations of species in a number of different ways. I'm not sure, based on some other estimation problems that we have done that this is currently done in the best way possible, so this might be a promising case study.

I am also playing around with perhaps asking a question something along the lines of "If your flight is delayed by some time, at what point is it a better idea to try to get on a different flight than to wait for

your now-delayed flight?" I'm not entirely sure about this question yet, but I'm interested in seeing if there is a variation of this question that might lend itself well to a Bayesian approach.

Reflection:

This week, I feel like I did a little less work than I did last week. I think that's mostly because I felt so lost last week, that I had to do a bunch of extra to even feel like I was starting to understand. This week, I fell more in command of what is happening, but still a little lost. I also spent quite a bit of time browsing the internet looking for possible case study material and trying to understand the situations in which Bayesian can be reasonably be applied. I guess I'm not quite used to an open class structure like this one, yet and I have to get used to being able to choose to budget my time however I want. This semester, I'd like to spend the semester getting better at git and using version control.