Progress Report: Week 1

Reading:

I read chapter 6 of *Think Bayes*. This chapter contains a, in-depth walkthrough of figuring out an optimal guess on a show like "The Price is Right". This chapter is, so far, my favorite chapter, because of the way that we walk through each step of one problem, as opposed to illustrating the same thing with many different problems.

I also read Silver's "How The FiveThrityEight Senate Forecast Model Works". This article contains a detailed description of all the factors that the FiveThrityEight senate forcast model takes into account, how those factors interact with one-another, and why a particular factor was taken into account. I'm not really interested in politics, but I thought that this model was fascinating, and was impressed by accurate their model was. This last point, I'm taking with a grain of salt because the author created the model, and would therefore want to convince the audience that the model works.

I was curious about this model and the disagreement between Silver and Wang, called "Kansas Is Key in a Fight among Statisticians" by Christian Triunfo. This post was posted on the statistics subredit. This post contained a discussion about how the author was really excited about seeing which one was right. I thought that this post was interesting from the perspective of seeing someone really excited about the election for the statistical side of it, but did think that there would be more talk of statistics in this blogpost.

I read chapter 7 of *Think Bayes*. This chapter was all about processes, mostly Poisson processes, and included a model of the probability that the Bruins would win the NHL finals. Again, I very much prefer walking through the solution to one problem as opposed to doing multiple problems that show the same thing. I really liked this chapter, and think that predicting the outcome/success of something like a game will be something I pursue in my case study.

I also read Chapter 1 of Cameron Davidson-Pilon, "Probablistic Programming and Bayesian Methods for Hackers". This chapter introduced most of the Bayesian concepts that we had learned so far, but with an approach greatly geared towards programmers. I thought that this approach was really helpful, and really helped me reinforce what we had learned so far, although I do not think it would have liked it as much had it been my first experience with Bayesian methods.

Exercises:

This week, I realized that I was a little shaky on some of the vocabulary, and that was making it a little hard for me to understand what was going on in class. As a result, I went back through

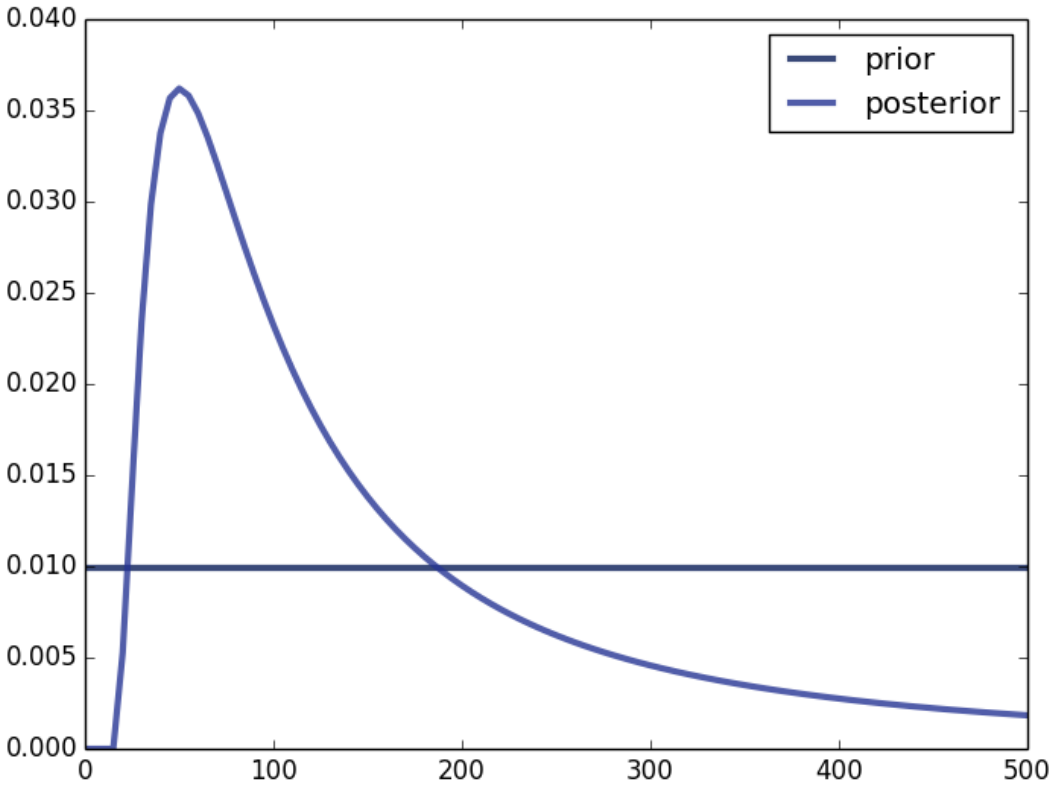
the chapters that we had already read and made this chart. Many of these definitions are the ones found in *Think Bayes*.

Term	**Definition**	**Notes**
Probability	Number between 0 and 1 representing degree of belief in fact or prediction	0 = False, 1 = True
Conjoint probability	Probability that two things are true	P(A and B)
Bayes Theorem	P(H|D) = P(H)*P(D|H)/P(D)	
Prior	What we believe before	
Posterior	What we believe after	
Likelihood	The probability of the data given a hypothesis	For a lot of hypotheses we'll probably have a likelihood function
Suite	A set of hypotheses	
Distribution	The probabilities of values	
Probability Mass Function (PMF)	The *Think Bayes* implementation of a distribution	Maps value to probability of that value
Credible Interval	The interval in which x% of the data falls	Has a percentage like 90% associated with it
Cumulative Distribution Function (CDF)	Maps values to the probability that the answer will be less than or equal to that value	Contains the same information as a PMF
Odds	Another way to represent probability	
Probability Density Function	A function that maps value to probability density, and when integrated maps to probability mass	
Process	Model of a physical system that contains some randomness	
Bernoulli process	Model of trials whose outcomes are success or failure	
Poisson Process	Continuous version of Bernoulli process	The trail can happen at any point in time

I also kept working on the Hyrax problem that we didn't finish from class on Tuesday. My likelihood function wound up looking like this (code: https://github.com/phiaseitz/ThinkBayes2/blob/master/code/hyrax.py) :

```
    def Likelihood(self, data, hypo):
        """Computes the likelihood of the data under the hypothesis.

        hypo: number of hyraxes
        data: the number already tagged the number caught, number tagged
        """
        (numberAlreadyTagged, numberCaught, numberTagged) = data
        #Assuming that for this case, the hypothesis is correct
        if (numberAlreadyTagged + (numberCaught - numberTagged))> hypo:
            like = 0
        else:
            like =
thinkbayes2.EvalBinomialPmf(numberTagged,numberCaught,(numberAlreadyTagged/hy
po))
        return like
```

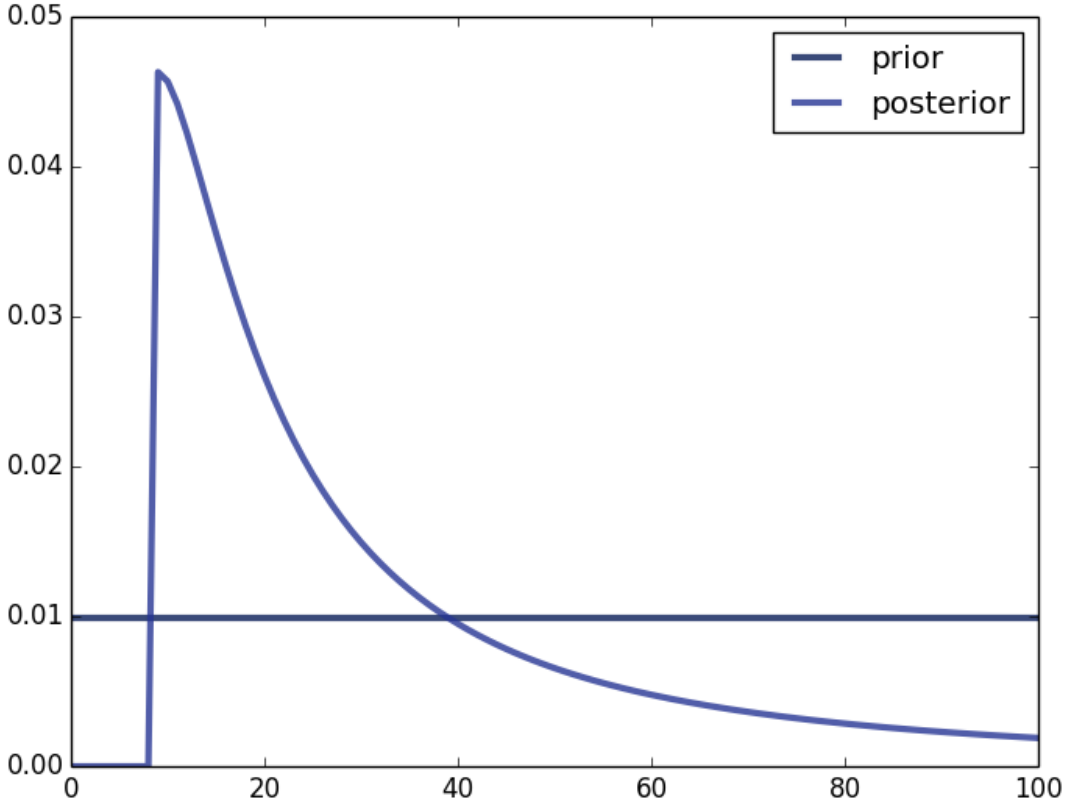
It took me a little while to figure out the likelihood function and using EvalBinomailPmf, but I think I finally understand it. Basically, if the number of hyraxes that are already tagged (plus the ones that we just caught that aren't tagged already) are greater than the hypothesis, then the likelihood of that hypothesis is 0. If not, then we evaluate the probability of finding 2 (or however many) tagged hyraxes out of 10, where the probability of that is the number of tagged hyraxes over the hypothesis. My graph looks like this.



This week, I also did the Fortune Cookie problem from a while back, to really make sure that I understood the ideas behind the Hyrax problem and these sorts of estimation problems. My likelihood function looked very similar (code: https://github.com/phiaseitz/ThinkBayes2/blob/master/code/FortuneCookie.py).

```
      def Likelihood(self, data, hypo):

            (numberOfCookies, inCommon) =  data

            if numberOfCookies > hypo:

                  like = 0

            else:

                  like =
thinkbayes2.EvalBinomialPmf(inCommon,numberOfCookies,inCommon/hypo)


            return like
```

And my graph looked like this:



Case Study:

This week, Claire Diehl and I sat down and tried to figure out the topic of our case study. We met and ideated on possible case study topics, and finally decided to do a case study on using Bayesian reasoning to predict the outcome of a crime television show. We talked about how we might go about implementing this, and we decided that we would look at the cast list for a given episode of whatever television show we wind up choosing (we'll be choosing the show and start trying to figure out how we might implement a likelihood function this weekend) and watch the first 2 or 3 minutes of the TV show. Then, we assign every member of the investigative team a 1% probability that they committed the crime. We assign the same probability to the victim of the crime, and everyone else on the cast list gets the left-over percentage evenly distributed. Then, every 5 or so minutes (whenever there was a new piece of data) we'll pause the show and update our priors. Right before the end of the show, when

everything is revealed, we'll see who we think committed the crime and then see if our thoughts match to who actually committed the crime.

Claire Diehl and I have not yet decided what TV show we'll use for this case study but we did come up with some requirements:

- There cannot be too much side plot
- Ideally, we would have some experience with the show, but would not have watched the episodes we're using for the case study
- The data can't be too hidden (so *Sherlock* would not be a good show)

Reflection:

This week, I feel much better about the work that I have done, mostly because I spent a lot of time trying to catch up my understanding on vocabulary terms that I missed or didn't have a good definition of before. Other than that, I did a lot of browsing the internet, mostly on the statistics subredit to try to come up with a case study that I thought would be interesting. This week, while I may not have produced a large amount of work, I feel much better about my understanding in the course than I have for the previous weeks. I also am glad that I now have a direction to go in with my case study and can start working on that.