# How long does it take me to fall asleep?

For Christmas 2018, I got a Google Home Mini. Not long after that, I started experimenting with the types of programming that you could do with the speaker. After a while I decided that I wanted a way to estimate how long it took me to fall asleep. Everyone made fun of me and told me I should just get a smartwatch (I have a Fitbit now), but it sounded like fun so I decided to go for it.

After a few hours of tinkering, I got a system set up as follows: when I tell my smart speaker goodnight, it triggers an app on my phone called Tasker, Tasker uses a random number generator to choose a random amount of minutes to wait, which roughly followed an exponential distribution. (Note: exponential was preferred because of it's "memoryless" property. The probability that it went off in the next minute given that it still hasn't gone off stays constant, so it reduces anticipation that the timer is going to go off soon). After the time expires the app triggers my phone's flash to light up my room for 5 seconds (I experimented with vibration and sound, but found light to be very noticeable but not disruptive if I was asleep). If I was awake when my room lit up I would tell my speaker "I'm awake." The speaker recorded the start time, when it alerted me, and whether I responded into a spreadsheet automatically.

Another reason I liked this problem is that because the thing of interest, exactly when we fell asleep, is never directly observable (in fact, it's even questionable if a single moment that crosses you from awake to asleep exists). Instead, we learn about that sleep threshold by censored information, if I respond to the light we know it happened after, if I didn't we know it happened before. After collecting data, I attempted to answer the question, "How long does it take me to fall asleep?" using two different types of analysis: censored Bayesian modeling and logistic regression.

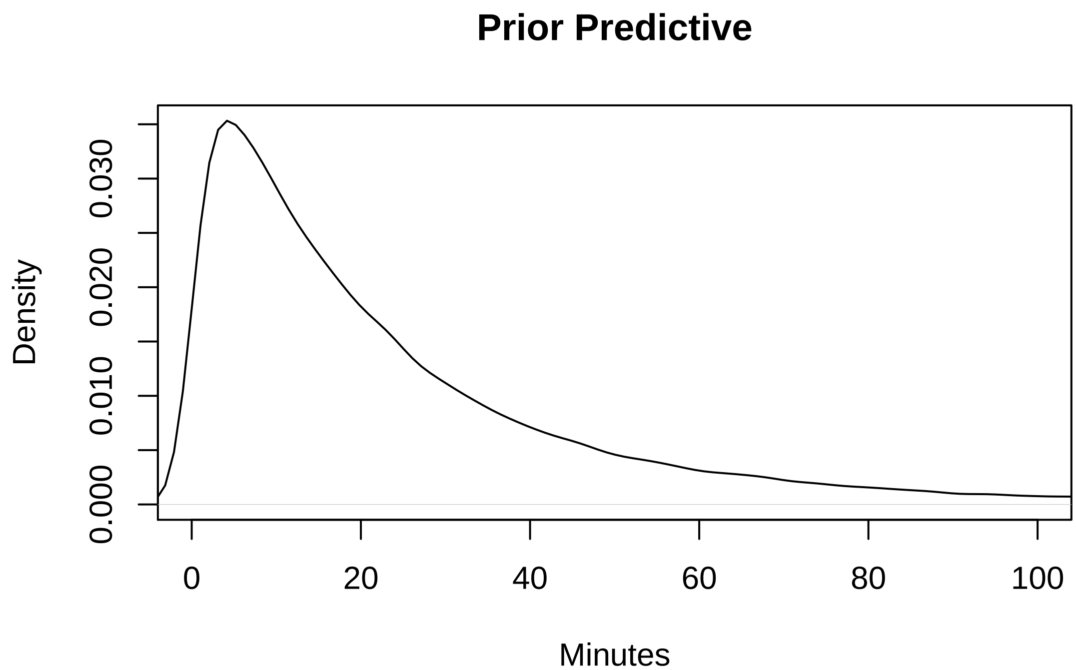
### Censored Bayes

Censoring is when you recognized that you didn't observe your data directly, instead you observe that it was higher or lower than some point, in this case that point being the amount of time that had passed until my phone lit up. Instead of the likelihood being a product of independent pdfs, a censored observation's contribution to the likelihood is the CDF or the reliability function (1 - the CDF) depending on whether it was right or left censored. This is not a tutorial on censored data, but more info can be found online.

### Choosing a likelihood and prior

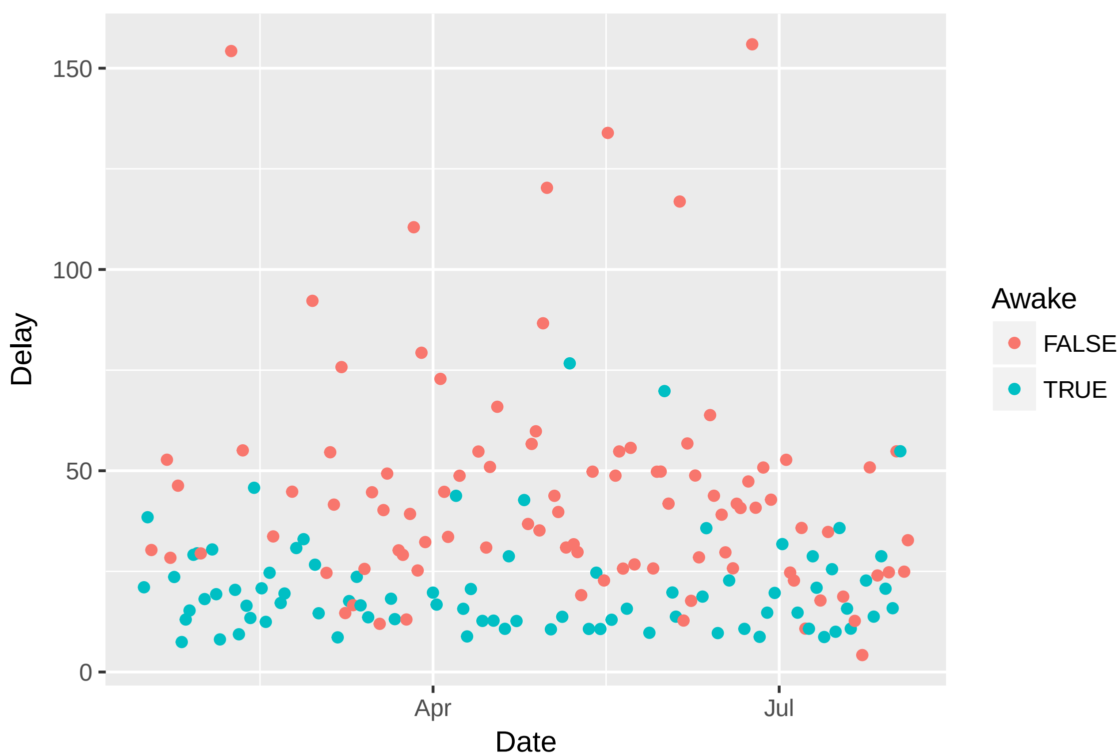
To build a Bayesian model, one must specify the likelihood and prior. For the likelihood, a natural choice seemed to be the gamma distribution. It had many properties that seemed to describe falling asleep: no negative times (I rarely fall asleep before going to bed), right skewed (for those long nights when you lay awake), and it is a very flexible distribution.

I chose to go with a fairly informative prior. Prior to gathering data, I happened to have a lot of prior experience of laying in bed at night and falling asleep (I had done it over 8,500 times!). I thought it took me about 20 minutes on average to fall asleep, a far cry from the "seven minutes" you read as about as normal online. I used gamma priors on alpha and beta and used the prior predictive distribution to hone in on this one, the one that captured my belief:

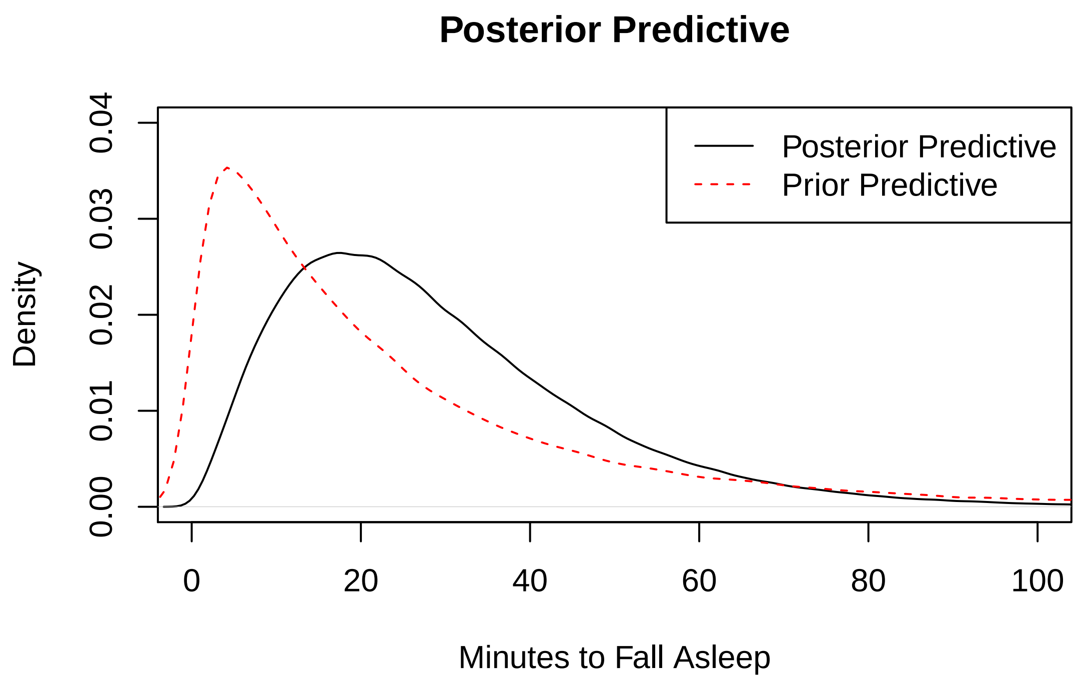



That was made by giving the prior on the shape and scale parameters $\alpha$ and $\beta$ gamma distributions both with a shape of 2.5 and a scale of 2.

Now that we have our model, we can fast-forward 177 days (plus some extras where I wasn't able to run the experiment due to not being home), and we can look at our data:



Setting up a Metropolis-Hastings algorithm in order to get posterior draws on my gamma parameters (see Analysis.R), we can obtain the coveted posterior predictive distribution, or the distribution that answers the question, "When I lay down to fall asleep tonight, what's the probability distribution over how long it will take me to fall asleep?"

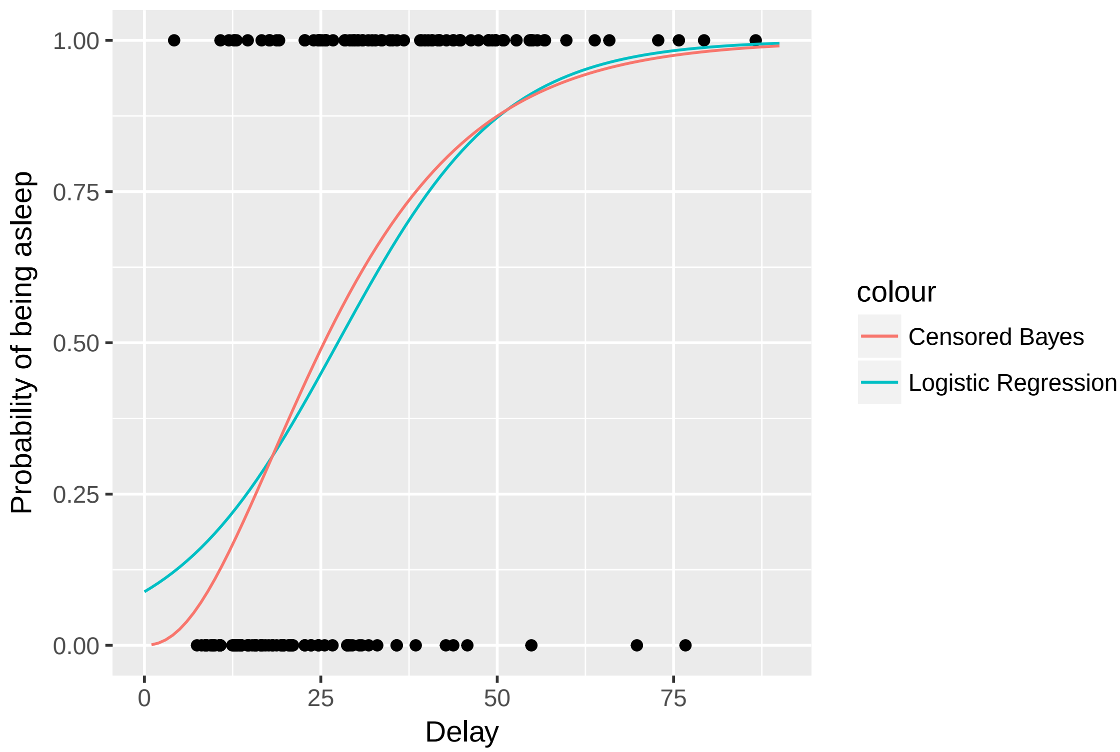



Analyzing the graph we can see pretty clearly that I took longer to fall asleep than I thought. My point estimate on my average time to sleep is 29.15 minutes, 95% credible interval of (25.49, 33.40). My posterior variance of falling asleep is quite wide ranging from (4.87 to 74.69) minutes.

**Logistic Regression**

About halfway through the experiment, I realized that instead of looking at each point as a censored observation of the truth, I could look at each observation as a success (asleep) or a failure (awake), and use logistic regression to answer the question "What is the probability of being asleep after x minutes?"

Using an extrememly simple logistic regression (modeling success just as a function of delay), I am able to compare the two methods in terms of their predicted probability of being asleep:




While the results are fairly similar, clearly the biggest difference is in the left tail. Logistic regression predicts an almost 9% chance of falling asleep instantaneously, something I have never experienced. The Bayesian approach puts far less probability on quick fall asleeps, something that I think lines up better with my day to day reality.

### Conclusion and Future Work

This experiment made me much more concious of sleep and the nuances involved. A huge assumption of this experiment was that sleep is fairly static: you transition to being asleep at one point in time and then stay asleep for the rest of the night. This is obviously a useful, but not totally accurate simplification of how sleep works. Moving forward I'm going to transition away from measuring if I've fallen asleep yet to "am I awake at this random time during the night?" by adjusting the timer to light up my room randomly a few times a night.