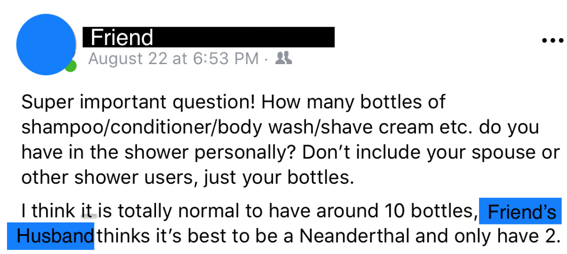
Stat 251: Assignment 5

Fall 2021

1. (This problem focuses on finding and interpreting the posterior distribution on a mean of a count.) The Orem City Department of Public Works is responsible for repairing potholes in city streets. It is reasonable to assume that the number of potholes that need to be repaired in a city block follow a Poisson(λ) distribution.

   a. The prior belief about λ is that it follows a Gamma(0.5, 0.5). What is the (*a priori*) expected number of potholes per block?
   b. After a particularly heavy winter, they repaired the following number of potholes in each block along a 20 block stretch of State St: 3, 3, 7, 1, 4, 6, 6, 7, 3, 1, 5, 5, 5, 3, 3, 0, 3, 1, 2, 2. We will assume that these data have a Poisson(λ) distribution. Orem City's prior belief about λ is a Gamma(0.5, 0.5). What is the posterior distribution of λ?
   c. Plot the posterior distribution and prior distribution on the same graph.
   d. What is their belief now (*a posteriori*) about the expected number of potholes they will repair in each block of State St?
   e. What is a 95% posterior interval for the parameter of the Poisson distribution? (Interpret.)
   f. What is the posterior probability that λ is greater than 4? (Interpret.)
   g. Orem randomly selects a block of state street (not previously examined). Find the posterior predictive probability that the actual number of potholes will be greater than 4. Why is this probability different than the probability found above?

2. (This problem focuses on comparing means from two populations.) In a lighthearted social media post, a friend started a heated discussion by posting the following comment[1]:



   The comical discussion that followed resulted in data that fit into our course; namely, counts. Thus, in a similarly lighthearted vein, we will add to my friend's discussion by conducting an analysis on this data.

   Many commenters identified differences between themselves and their spouse. Specifically, the wives tended to have more bottles than the husbands. In this problem, we will look at the averages from two populations: self-identified men and self-identified women. Additionally, we recognize that this is not *at all* a representative sample of men and women in general, but we will talk about it as though it is, just to make writing about the analysis easier.

   a. Let λ_M be the average number of bottles that men keep in the shower and λ_W be the average for women. We assume *a priori* that both λ_M and λ_W follow a Gamma$(8, 1.5)$. What are the prior expected values and variances for λ_M and λ_W?

[1](To which her husband responded: "I regret helping you spell 'Neanderthal.'")

b. The data from the comments are listed below. Assume the data from the two populations are independent and follow Pois(λ_W) and Pois(λ_M) distributions, respectively. What are the posterior distributions for λ_W and λ_M?

women:	12	9	10	8	9	4	10	15	3	5
	11	8	9	4	2	7	9	5	4	2
	3	12	10	2	9	8	13	9	7	6
	6	2	2	6	8					
men:	2	3	0	4	1	1	1	2	2	2
	0	3	2							

c. Compute 95% credible intervals for λ_W and λ_M.

d. Plot the posterior distribution on $d = \lambda_W - \lambda_M$. Given our data and prior knowledge, on average, how many more bottles do women keep in the shower than men?

3. (This problem focuses on the derivation of the posterior distribution. In class we used the fact that the posterior distribution is a valid pdf and identified its kernel as a "shortcut" for finding the posterior distribution. In this problem you will show that the "shortcut" gives the same approach as if we had derived the distribution without using the "shortcut.") Let

$$X_i|\lambda \overset{iid}{\sim} \text{Pois}(\lambda), \text{ for } i = 1, \ldots, n$$
$$\text{and } \lambda \sim \text{Gamma}(\gamma, \phi).$$

Derive $\pi(\lambda|\text{data})$ using Bayes Theorem by

a. solving for the normalizing constant, and

b. plugging the normalizing constant found in (a) into Bayes Theorem and simplifying to obtain the pdf of a Gamma($\gamma + \sum_{i=1}^{n} x_i, \phi + n$).