STAT 535 – Spring 2024 – Homework 3B

1. In class (for the Prussian cavalry example) we derived the posterior predictive distribution in the case where $Y_1, \ldots, Y_n \sim$ i.i.d. Poisson(λ), with the Gamma prior $\lambda \sim G(2, 4)$. Instead of deriving the posterior predictive distribution analytically, we could have sampled from it using Monte Carlo methods.

   (a) Randomly sample $\lambda^{[1]}, \ldots, \lambda^{[J]}$ from a Gamma(shape $= \sum y_i + 2$, rate $= n + 4$) distribution. Using these, sample $Y^{[1]}, \ldots, Y^{[J]}$ from Poisson($\lambda^{[j]}$) distributions, $j = 1, \ldots, J$.

   See R code on course webpage.

   (b) Plot the approximate posterior predictive distribution, similarly as our in-class example. How does it appear to compare to the observed Prussian army data distribution?

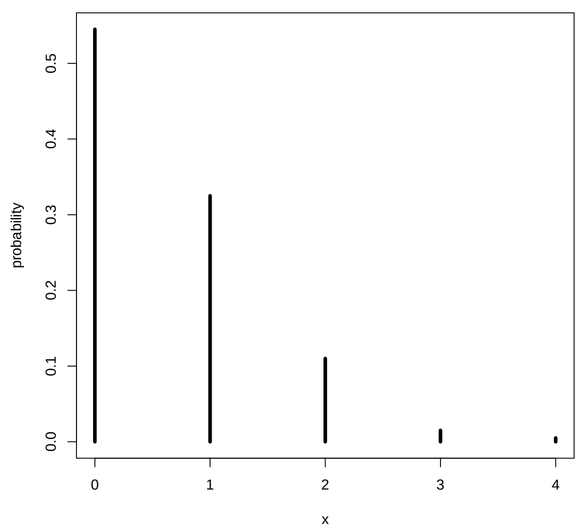

Figure 1: Problem 3: Plot of observed proportions of values in the Prussian horse kick example.

   The posterior predictive distribution looks very similar to the observed proportions. This indicates the Poisson model is a great fit to these data.

2. In a NASA experiment, 14 male rats were sent into space. When they returned, the red blood cell mass (in ml) of each rat was measured. In addition, 14 other male rats were kept on earth during the same period of time. Those rats also had their red blood cell mass measured. Assume the red blood cell masses for the two groups can be modeled with a normal distribution, with equal variances across the two groups. The data are:

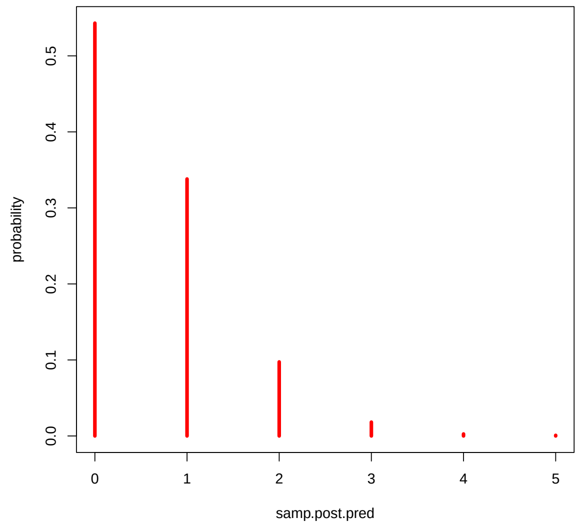

Figure 2: Problem 3: Plot of posterior predictive distribution.

```
Space rats: 8.59 8.64 7.43 7.21 6.87 7.89 9.79 6.85 7.00 8.80 9.30 8.03 6.39 7.24
Earth rats: 8.65 6.99 8.40 9.66 7.62 7.44 8.55 8.70 7.33 8.58 9.88 9.94 7.14 9.04
```

(a) Suppose the research question of interest was to test whether the mean red blood cell mass differed for the two groups. Answer this question based on a Bayesian hypothesis test. Clearly state your prior specifications (use $\mu_\Delta = 0, \sigma^2_\Delta = 1/5$). Give a posterior probability for each hypothesis being true.

The prior beliefs do not favor either hypothesis, since I choose $\mu_\Delta = 0$. My Bayes Factor was 0.83 for the null model that the mean red blood cell mass is the same for each group. So there is some evidence that the means differ, but it is not especially strong evidence. The posterior probability for H_0 (means are same) is 0.455, so the posterior probability for H_a (means differ) is 0.545.

(b) Suppose the research question of interest was to test whether the mean red blood cell mass for the space group was lower than the mean red blood cell mass for the control group. Answer this question based on a Bayesian hypothesis test. Clearly state your prior specifications. (The researcher believed *a priori* that the rat population as a whole might have average red blood cell mass somewhere around 7 ml, but was not at all sure about the effect of the space travel.) Give a posterior probability for each hypothesis being true.

I used a Gibbs sampling approach with a normal prior on μ (with mean 7) and a normal prior on τ (with mean 0, not favoring either group *a priori*). The posterior probability of H_a (mean is lower for space group) was about 0.87. So the posterior probability of

H_0 was about 0.13. I'd tend to believe H_a (mean is lower for space group), based on that. But that is a subjective decision of mine.

3. A physician is interested in determining whether the mean systolic blood pressure of a certain set of patients is less than 130. She takes a random sample of 17 patients and measured their systolic blood pressure. Assume the measurements follow a $N(\mu, \sigma^2)$ distribution, with μ unknown and known $\sigma^2 = 225$. (You can use normal-normal results from Chapter 5 to obtain the posterior distribution for μ.) The physician says *a priori* that she is 95% sure that the true mean systolic blood pressure is between 120 and 140. The data are:

   ```
   118 140 90 150 128 112 134 140 112 126 112 148 124 130 142 105 125
   ```

   (a) Conduct a Bayesian hypothesis test of $H_0 : \mu \geq 130$ vs. $H_a : \mu < 130$, basing your conclusions on the posterior distribution for μ. Clearly state your prior specifications.

   Based on the expert opinion, I use a normal prior on μ with mean 130 and standard deviation 5. The true σ is assumed to be 15. The conjugate analysis gives a normal posterior with mean 127.1539 and standard deviation 2.941742 (see R code for how to get this). Based on this, the posterior probability that H_0 is true is $P(\mu \geq 130|\mathbf{y}) = 0.167$. Based on this, I might conclude that $\mu < 130$, but this is my own judgment call.

   (b) Conduct a classical t-test using $\alpha = 0.05$. Are the substantive conclusions any different from those in part (a)?

   The classical t-test of $H_0 : \mu \geq 130$ vs. $H_a : \mu < 130$ has a P-value of 0.142. So using $\alpha = 0.05$, we would fail to reject H_0 and we would not conclude that $\mu < 130$. Answers will vary about whether this is different from the conclusion in part (a).

4. Do Problem 8.8 from the *Bayes Rules!* textbook. [You do not have to provide the sketches of the intervals on the posterior pdf. And you can use the `hpd` function in the `TeachingDemos` package to get the HPD intervals.]

   ```
   > # HPD:
   >
   > library(TeachingDemos)
   >
   > hpd(qgamma,shape=1,rate=5,conf=0.95)
   [1] 0.000000001268843 0.599146480087658
   >
   > # middle 95%:
   >
   > c(qgamma(0.025,shape=1,rate=5), qgamma(0.975,shape=1,rate=5) )
   [1] 0.005063562 0.737775891
   ```

```
>
> # HPD:
>
> library(TeachingDemos)
>
> hpd(qnorm,mean=-13,sd=2,conf=0.95)
[1] -16.919928  -9.080072
>
> # middle 95%:
>
> c(qnorm(0.025,mean=-13,sd=2), qnorm(0.975,mean=-13,sd=2) )
[1] -16.919928  -9.080072
>
```

We see the two methods produce different intervals with the gamma posterior density, since the gamma density is skewed. The HPD interval is much shorter and is probably to be preferred here.

With the normal posterior density, the two methods produce the same interval since the normal density is symmetric.

5. Do Problem 8.9 from the *Bayes Rules!* textbook.

```
> # posterior probability for Ha:
>
> post.prob.Ha <- 1-pbeta(0.4,4,3)
> print(post.prob.Ha)
[1] 0.8208
>
> # posterior odds for Ha:
>
> post.odds.Ha <- (post.prob.Ha)/(1-post.prob.Ha)
> print(post.odds.Ha)
[1] 4.580357
>
> # prior odds for Ha:
>
> prior.prob.Ha <- 1-pbeta(0.4,1,0.8)
> prior.odds.Ha <- (prior.prob.Ha)/(1-prior.prob.Ha)
> print(prior.odds.Ha)
[1] 1.98098
>
> # Bayes factor for Ha:
>
```

```
> (post.odds.Ha)/(prior.odds.Ha)
[1] 2.312168
>
> # Note the Bayes Factor for H0 would be the reciprocal of this.
>
```

After seeing the data, the odds that H_a is true are 2.3 times as great as they were before seeing the data.

Before seeing the data, I believed H_a was about twice as likely as H_0. After seeing the data, I now believe H_a is about 4.6 times as likely as H_0.