

# Homework 2

PSTAT 115, Fall 2020

**Due on October 25, 2020 at 11:59 pm**

**Note:** If you are working with a partner, please submit only one homework per group with both names and whether you are taking the course for graduate credit or not. Submit your Rmarkdown (.Rmd) and the compiled pdf on Gauchospace.

## 1. Cancer Research in Laboratory Mice

A laboratory is estimating the rate of tumorigenesis (the formation of tumors) in two strains of mice, A and B. They have tumor count data for 10 mice in strain A and 13 mice in strain B. Type A mice have been well studied, and information from other laboratories suggests that type A mice have tumor counts that are approximately Poisson-distributed. Tumor count rates for type B mice are unknown, but type B mice are related to type A mice. Assuming a Poisson sampling distribution for each group with rates  $\theta_A$  and  $\theta_B$ . Based on previous research you settle on the following prior distribution:

$$\theta_A \sim \text{gamma}(120, 10), \theta_B \sim \text{gamma}(12, 1)$$

**1a.** Before seeing any data, which group do you expect to have a higher average incidence of cancer? Which group are you more certain about a priori? Your answers should be based on the priors specified above.

*Type your answer here, replacing this text.*

**1b.** After you complete the experiment, you observe the following tumor counts for the two populations:

$$y_A = (12, 9, 12, 14, 13, 13, 15, 8, 15, 6)$$
$$y_B = (11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7)$$

Compute the posterior parameters, posterior means, posterior variances and 95% quantile-based credible intervals for  $\theta_A$  and  $\theta_B$ . Save them in the appropriate variables in the code cell below. You do not need to show your work, but you cannot get partial credit unless you do show work.

```
. = ottr::check("tests/q1b.R")
```

**1c.** Compute and plot the posterior expectation of  $\theta_B$  given  $y_B$  under the prior distribution  $\text{gamma}(12 \times n_0, n_0)$  for each value of  $n_0 \in \{1, 2, \dots, 50\}$ . As a reminder,  $n_0$  can be thought of as the number of prior observations (or pseudo-counts).

```
# YOUR CODE HERE
```

```
posterior_means = NULL # YOUR CODE HERE
```

```
# YOUR CODE HERE
```

```
. = ottr::check("tests/q1c.R")
```

**1d.** Should knowledge about population A tell us anything about population B? Discuss whether or not it makes sense to have  $p(\theta_A, \theta_B) = p(\theta_A) \times p(\theta_B)$ .

*Type your answer here, replacing this text.*

## 2. A Mixture Prior for Heart Transplant Surgeries

A hospital in the United States wants to evaluate their success rate of heart transplant surgeries. We observe the number of deaths,  $y$ , in a number of heart transplant surgeries. Let  $y \sim \text{Pois}(\nu\lambda)$  where  $\lambda$  is the rate of deaths/patient and  $\nu$  is the exposure (total number of heart transplant patients). When measuring rare events with low rates, maximum likelihood estimation can be notoriously bad. We'll take a Bayesian approach. To construct your prior distribution you talk to two experts. The first expert thinks that  $p_1(\lambda)$  with a  $\text{gamma}(3, 2000)$  density is a reasonable prior. The second expert thinks that  $p_2(\lambda)$  with a  $\text{gamma}(7, 1000)$  density is a reasonable prior distribution. You decide that each expert is equally credible so you combine their prior distributions into a mixture prior with equal weights:  $p(\lambda) = 0.5 * p_1(\lambda) + 0.5 * p_2(\lambda)$

**2a.** What does each expert think the mean rate is, *a priori*? Which expert is more confident about the value of  $\lambda$  *a priori* (i.e. before seeing any data)?

*Type your answer here, replacing this text.*

**2b.** Plot the mixture prior distribution.

*# YOUR CODE HERE*

**2c.** Suppose the hospital has  $y = 8$  deaths with an exposure of  $\nu = 1767$  surgeries performed. Write the posterior distribution up to a proportionality constant by multiplying the likelihood and the prior density. *Warning:* be very careful about what constitutes a proportionality constant in this example.

*Type your answer here, replacing this text.*

**2d.** Let  $K = \int L(\lambda; y)p(\lambda)d\lambda$  be the integral of the proportional posterior. Then the proper posterior density, i.e. a true density integrates to 1, can be expressed as  $p(\lambda | y) = \frac{L(\lambda; y)p(\lambda)}{K}$ . Compute this posterior density and clearly express the density as a mixture of two gamma distributions.

*Type your answer here, replacing this text.*

**2e.** Plot the posterior distribution. Add vertical lines clearly indicating the prior means from each expert. Also add a vertical line for the maximum likelihood estimate.

*# YOUR CODE HERE*