

Homework 2

Kevin Ayala and Aaron Bales

Due on October 21, 2018 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.

Cancer Research in Laboratory Mice

1. 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 θ_A and θ_B . Based on previous research you settle on the following prior distribution:

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

- (a) 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.

We expect to see group A to have higher average incidents of cancer because we have a stronger prior to group A “Type A mice have been well studied...”, also for group A, we have 10 as the sample whereas for group B the sample is 1 which is questionable.

- (a) After you complete the experiment, you observe the following tumor counts for the two populations:

$$y_A = (12, 9, 12, 14, 13, 13, 15, 8, 15, 6)$$

$$y_B = (11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7)$$

Write down the posterior distributions, posterior means, posterior variances and 95% quantile-based credible intervals for θ_A and θ_B

```
y_a_data<- c(12,9,12,14,13,13,15,8,15,6)
k_ya <- sum(y_a_data)
sample1size <- length(y_a_data)
```

```
y_b_data<- c(11,11,10,9,9,8,7,10,6,8,8,9,7)
k_yb<- sum(y_b_data)
sample2size <- length(y_b_data)
```

```
# We have seen in class that the posterior of a gamma for a poisson distribution is
# Gamma ~ (alpha+k, n+beta )
```

```
#For type 1 Rat, our distribution is
# Gamma~(120+k_ya, 10 + sample1size)
# Gamma~(120+117, 10+10)
# Gamma ~ (237 , 20) for Type 1 rats, distribution is as follows
```

```

mean_ya <- (120+k_ya)/(10+sample1size)
mean_ya

## [1] 11.85
#For type 2 Rat, our distribution is
# Gamma~(12+k_yb, 1 + sample2size)
# Gamma~(12+113, 1+13)
# Gamma ~ (125 , 14) for Type 1 rats, distribution is as follows
mean_yb <- (12+k_yb)/(1+sample2size)
mean_yb

## [1] 8.928571
Var_ya <- (120+k_ya)/((10+sample1size)^2)
Var_ya

## [1] 0.5925
Var_yb <- (12+k_yb)/((1+sample2size)^2)
Var_yb

## [1] 0.6377551
#credible Interval for Group A
qgamma(.975, 237, 20)

## [1] 13.40545
qgamma(.025, 237, 20)

## [1] 10.38924
# 95% credible interval for lab A rats is (10.38924, 13.40545)

#credible Interval for Group B
qgamma(.975, 125, 14)

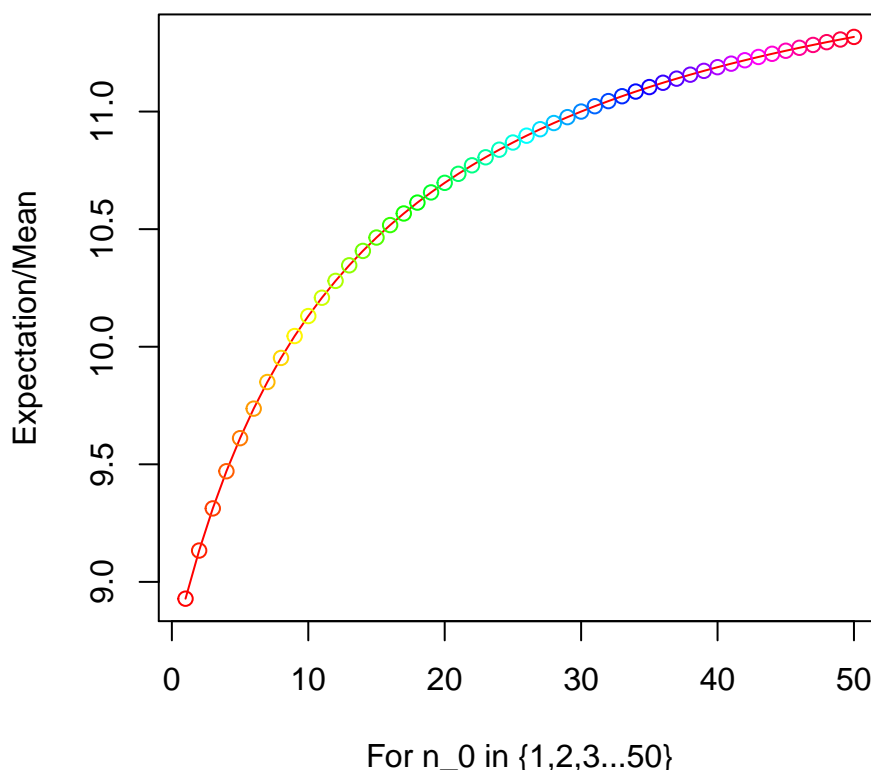
## [1] 10.56031
qgamma(.025, 125, 14)

## [1] 7.432064
#95% credible Interval for LAB B rats is (7.432064, 10.56031)

#. Compute and plot the posterior expectation of  $\theta_B$  given  $y_B$  under the prior distribution  $\theta \sim \text{Gamma}(12, 10)$ 
n_0 <- c(1:50) #for n_0 in {1,2,3,...50}
#given posterior from previous, then gamma~ (12*n_0+113, n_0 + 10)
#expectation = mean of distribution, (12*n_0+113/n_0 + 10)
expectation <- (12*n_0+113)/(n_0 + 13)
plot(n_0,expectation, type = "o", xlab="For n_0 in {1,2,3,...50}"
     , ylab="Expectation/Mean", col = rainbow(50), main = "Posterior Expectation of Theta")

```

Posterior Expectation of Theta



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

They are different populations of rats so it makes sense that the populations are independent of one another since tumors are not contagious. Thus multiplying them together by law of independence is appropriate.

A Mixture Prior for Heart Transplant Surgeries

1. 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 λ is the rate of deaths/patient and ν 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 $p_1(\lambda) = \text{gamma}(3, 2000)$ is a reasonable prior distribution. The second expert thinks that $p_2(\lambda) = \text{gamma}(7, 1000)$ 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: $p(\lambda) = 0.5 * \text{gamma}(3, 2000) + 0.5 * \text{gamma}(7, 1000)$

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

Expert 1 believes that the rate is .0015 (3/2000) and the second expert believes the rate of his prior is .007 (7/1000). Let us see what their 95% credible intervals are using their priors.

```
upperlim <- qgamma(.975, 3, 2000)
upperlim <- qgamma(.025, 3, 2000)

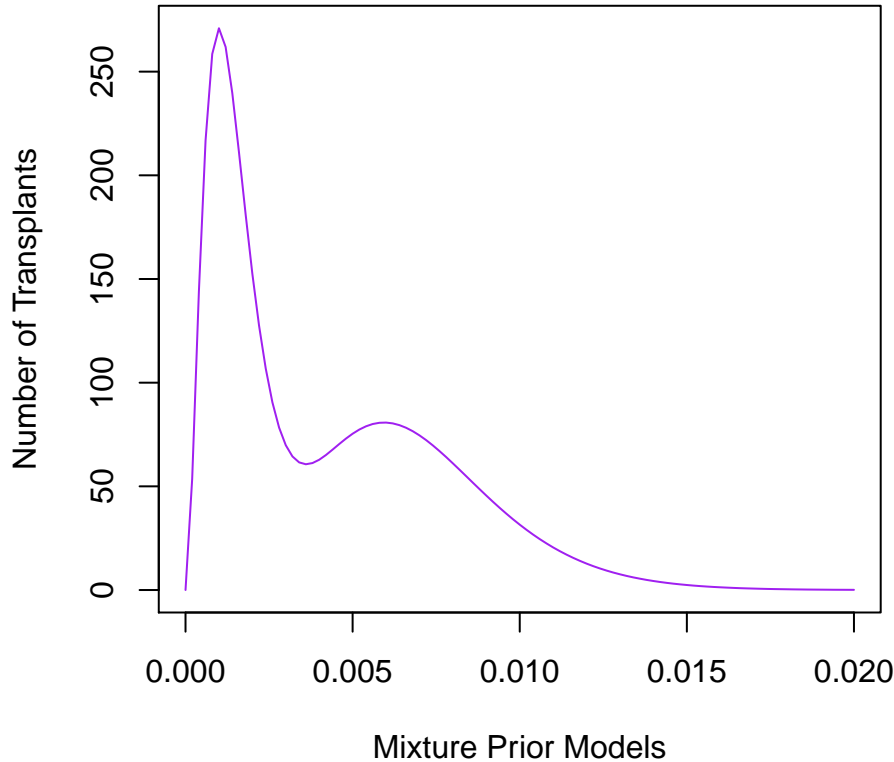
upperlim2 <- qgamma(.975, 7, 1000)
```

```
lowerlim2 <- qgamma(.025, 7, 1000)
```

Expert 1 confidence interval is (.0003, .003) and expert 2 interval is (.0028, .013). Comparing these intervals, we notice that expert 2 is more confident because he/she has a tighter interval indicating confidence.

(a) Plot the mixture prior distribution.

```
curve(.5*dgamma(x, shape = 3, rate=2000) + .5*dgamma(x, shape = 7, rate=1000), from=0, to=.02, xlab="Mixture Prior Models")
```



```
#set limits small, cant see clearly when its from 0 to 1
```

#. Suppose the hospital has $y=8$ deaths with an exposure of $\nu=1767$ surgeries performed. Write the The likelyhood function of a Poisson distribution with ignored constants is:

$$L(\lambda | y = 8) = \lambda^y e^{-\lambda \nu} = \lambda^8 e^{-1767\lambda}$$

The leading constant for Poisson Likelihood does not matter hence we will not consider it here, or set it equal to one. We can do this because when we take the derivative of the likelihood, the leading constant goes away and does not play any role on the MLE. We can also ignore denominator since it contains no lambda.

$$p(\lambda | y = 8) \propto \lambda^8 e^{-1767\lambda} (\lambda^2 e^{-2000\lambda} + \lambda^6 e^{-1000\lambda})$$

#. Let $K = \int_0^\infty L(\lambda | y) p(\lambda) d\lambda$ be the integral of the proportional posterior. Then $K = \int_0^\infty L(\lambda | y) \times p(\lambda) d\lambda$

Since we know a true density integral evaluates to 1, then by multiplying both sides by the reciprical of K:

$$\frac{1}{K} \times K = \frac{1}{K} \int_0^\infty L(\lambda | y) \times p(\lambda) d\lambda$$

Thus the following is true:

$$1 = \int_0^{\infty} \frac{L(\lambda|y) \times p(\lambda)}{K} d\lambda$$

Then:

$$1 = \frac{1}{K} \int_0^{\infty} (\lambda^{10} e^{-3767\lambda} + \lambda^{14} e^{-2767\lambda}) d\lambda$$

$$1 = \frac{1}{K} \left(\int_0^{\infty} \lambda^{10} e^{-3767\lambda} d\lambda + \int_0^{\infty} \lambda^{14} e^{-2767\lambda} d\lambda \right)$$

This is equivalent to a mixture model of a gamma posterior distribution:

$$\text{gamma}(\alpha, \beta) = f(\lambda) = \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)}$$

With exposure:

$$\text{gamma}(\alpha, \beta) = f(\nu\lambda) = \frac{\beta^\alpha \nu \lambda^{\alpha-1} e^{-\beta\nu\lambda}}{\Gamma(\alpha)}$$

Then if we set the integral with limits of 0 to infinity, then we can let every constant be proportion to K:

$$f(\lambda\nu) = \text{Gamma}(\alpha, \beta) = \frac{1}{K} \int_0^{\infty} \lambda^{\alpha-1} e^{-\lambda\nu}$$

Thus in integral number one, we set $\alpha=11$, and $\nu=3767$. Likewise for integral 2, let $\alpha=15$ and $\nu = 2767$

Thus, from the mixed posterior model of the two experts in part c with constants removed, we know the posterior distribution is $\text{Gamma}(11, 3767) + \text{Gamma}(15, 2667)$

#. Plot the posterior distribution. Add vertical lines clearly indicating the prior means from each expert.

```
curve(dgamma(x, shape = 11, rate = 3767) + dgamma(x, shape = 15, rate = 2767),
      from = 0, to = .01, col = "purple")
abline(v = .002775, col = "red")
abline(v = (3/2000), col = "green") #expert 1 mle from part a since MLE is the mean.
abline(v = (7/1000), col = "blue") #expert 2 mle from part a since mean is MLE
```

$\text{amma}(x, \text{shape} = 11, \text{rate} = 3767) + \text{dgamma}(x, \text{shape} = 15, \text{rate} =$

