Homework 2

Gaurav Kandlikar

October 14, 2016

1a: Algebraically derive the posterior p(lambda|y1) given a single observation y1. Specify your answer (a) in terms of a named distribution with parameters (ie Gamma(a, b), and specify a and b), and give the actual density formula.

1

Getting the posterior for Poisson dataset y and a Gamma prior:

$$P(\lambda|y) = \frac{\overset{①}{P(y|\lambda)}\ \overset{②}{P(\lambda)}}{③\ P(y)}$$

① $$P(y|\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}$$

② $$P(\lambda) = \frac{\beta^{\alpha}\lambda^{\alpha-1}e^{-\lambda(\beta)}}{\Gamma(\alpha)}$$ for $Gamma(\lambda; \alpha, \beta)$

③ $$P(y) = \int_0^\infty P(y|\lambda)\ p(\lambda)\ d\lambda$$

$$= \int_0^\infty \frac{\beta^{\alpha}\lambda^{\alpha-1}e^{-\lambda\beta}}{\Gamma(\alpha)} \cdot \frac{e^{-\lambda}\lambda^{y}}{y!}\ d\lambda = \int_0^\infty \frac{\beta^{\alpha}\lambda^{\alpha+y-1}e^{-\lambda(\beta+1)}}{y!\ \Gamma(\alpha)}\ d\lambda$$

We multiply the integral by two convenience terms:

$$\int_0^\infty \frac{\beta^{\alpha}\lambda^{\alpha+y-1}e^{-\lambda(\beta+1)}}{y!\ \Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+y)}{\Gamma(\alpha+y)} \cdot \frac{(1+\beta)^{\alpha+y}}{(1+\beta)^{\alpha+y}}\ d\lambda$$

and we factor out some terms that are constant WRT λ:

$$\frac{\beta^{\alpha}\ \Gamma(\alpha+y)}{y!\ \Gamma(\alpha)\ (1+\beta)^{\alpha+y}} \int_0^\infty \frac{(\beta+1)^{\alpha+y}\lambda^{\alpha+y-1}e^{-\lambda(\beta+1)}}{\Gamma(\alpha+y)}\ d\lambda$$

As the term inside the integral is the PDF of $Gamma(\lambda; \alpha+y, \beta+1)$, integrating across the range of $0 \to \infty$ yields 1.

(3) Con't

Once we set the integral equal to 1, the denominator term (3) reduces to

$$\frac{\beta^{\alpha}\,\Gamma(\alpha+y)}{y!\,\Gamma(\alpha)\,(1+\beta)^{\alpha+y}}$$

Putting together (1), (2), and (3):

$$\frac{\dfrac{e^{-\lambda}\lambda^{y}}{y!}\cdot\dfrac{\beta^{\alpha}\lambda^{\alpha-1}e^{-\lambda\beta}}{\Gamma(\alpha)}}{\dfrac{\beta^{\alpha}\,\Gamma(\alpha+y)}{y!\,\Gamma(\alpha)\,(1+\beta)^{\alpha+y}}}$$

$$= \frac{e^{-\lambda}\lambda^{y}}{\cancel{y!}}\,\frac{\cancel{\beta^{\alpha}}\lambda^{\alpha-1}e^{-\lambda\beta}}{\cancel{\Gamma(\alpha)}}\cdot\frac{\cancel{y!}\,\cancel{\Gamma(\alpha)}\,(1+\beta)^{\alpha+y}}{\cancel{\beta^{\alpha}}\,\Gamma(\alpha+y)}$$

$$= \frac{(1+\beta)^{\alpha+y}\,\lambda^{\cancel{\alpha+y}\,\alpha+1}\,e^{-\lambda(\beta+1)}}{\Gamma(\alpha+y)}$$

which is the PDF of Gamma $(\lambda;\,\alpha+y,\,\beta+1)$

1b: What is the support (place where density/function is non-negative) of: (i) prior, (ii) posterior, (iii) sampling density, (iv) likelihood (v) prior predictive?

1c: Algebraically derive the posterior given a sample yi , i = 1,. . . ,n of size n. Again specify it both in terms of a named distribution with parameters, and give the actual density formula.

Getting the posterior for Poisson distributed dataset $Y: [y_1, y_2 \dots y_n]$ and a Gamma prior:

$$P(\lambda|Y) = \frac{^{①}P(Y|\lambda)\, P(\lambda)^{②}}{^{③}P(Y)}$$

① $$P(Y|\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} = \frac{e^{-n\lambda}\lambda^{\sum_i y_i}}{\prod_{i=1}^{n} y_i!}$$

② $$P(\lambda) = \frac{\beta^{\alpha}\lambda^{\alpha-1}e^{-\lambda\beta}}{\Gamma(\alpha)} \quad \text{for } P(\lambda;\alpha,\beta)$$

③ $$\int_0^{\infty} \frac{\beta^{\alpha}\lambda^{\sum_{i=1}^{n} y_i+\alpha-1} e^{-\lambda(\beta+n)}}{\Gamma(\alpha)\prod_{i=1}^{n} y_i!}\, d\lambda$$

We multiply the integral by two convenience terms:

$$= \int_0^{\infty} \frac{\beta^{\alpha}\lambda^{\sum_{i=1}^{n} y_i+\alpha-1} e^{-\lambda(\beta+n)}}{\Gamma(\alpha)\prod_{i=1}^{n} y_i!} \cdot \frac{\Gamma(\alpha+\sum_i y_i)}{\Gamma(\alpha+\sum_i y_i)} \cdot \frac{(\beta+n)^{\alpha+\sum y_i}}{(\beta+n)^{\alpha+\sum y_i}}\, d\lambda$$

Factor out some terms that are constant WRT λ:

$$\frac{\beta^{\alpha}\cdot\Gamma(\alpha+\sum_i y_i)}{\Gamma(\alpha)\prod_{i=1}^{n} y_i!\cdot(\beta+n)^{\alpha+\sum y_i}} \int_0^{\infty} \frac{(\beta+n)^{\alpha+\sum y_i}\cdot\lambda^{\sum_{i=1}^{n} y_i+\alpha-1} e^{-\lambda(\beta+n)}}{\Gamma(\alpha+\sum_i y_i)}\, d\lambda$$

as the term inside integral is the PDF of Gamma$(\lambda;\alpha+\sum y_i, \beta+n)$, integrating over the whole range of (λ), ie $0\to\infty$, will yield 1

③, con't

Once we set the integral to 1, the denominator term ③ is:

$$\frac{\beta^\alpha \Gamma(\alpha + \sum_i y_i)}{\Gamma(\alpha) \cdot \prod_{i=1}^{n} y_i! \cdot (\beta+n)^{\sum_i y_i + \alpha}}$$

Putting together ①, ②, and ③:

$$\frac{\overset{①}{\dfrac{e^{-n\lambda} \lambda^{\sum y_i}}{\prod_{i=1}^{n} y_i!}} \cdot \overset{②}{\dfrac{\beta^\alpha \lambda^{\alpha-1} e^{-\lambda\beta}}{\Gamma(\alpha)}}}{\overset{③}{\dfrac{\beta^\alpha \Gamma(\alpha + \sum_i y_i)}{\Gamma(\alpha) \cdot \prod_{i=1}^{n} y_i! \, (\beta+n)^{\sum_i y_i + \alpha}}}}$$

$$= \frac{\dfrac{\beta^\alpha \lambda^{\sum y_i + \alpha - 1} e^{-\lambda(\beta+n)}}{\prod_{i=1}^{n} y_i! \; \Gamma(\alpha)}}{\dfrac{\beta^\alpha \Gamma(\alpha + \sum y_i)}{\Gamma(\alpha) \cdot \prod_{i=1}^{n} y_i! \; (\beta+n)^{\sum y_i + \alpha}}}$$

$$= \frac{\cancel{\beta^\alpha} \lambda^{\sum y_i + \alpha - 1} e^{-\lambda(\beta+n)}}{\cancel{\prod_{i=1}^{n} y_i!} \; \cancel{\Gamma(\alpha)}} \cdot \frac{\cancel{\Gamma(\alpha)} \cdot \cancel{\prod_{i=1}^{n} y_i!} \; (\beta+n)^{(\sum y_i + \alpha)}}{\cancel{\beta^\alpha} \; \Gamma(\alpha + \sum y_i)}$$

$$= \frac{(\beta+n)^{\sum y_i + \alpha} \lambda^{\sum y_i + \alpha - 1} e^{-\lambda(\beta+n)}}{\Gamma(\alpha + \sum_i y_i)}$$

which is the pdf of $\text{Gamma}\left(\lambda; \alpha + \sum_{i=1}^{n} y_i, \beta+n\right)$

1d: In the prior gamma(a, b), which parameter acts like a prior sample size? (Hint: look at the posterior from problem (1c), how does n enter into the posterior density?) You will need this answer inthe second part of problem 1(g)iv or you will do that problem incorrectly.

In the prior gamma(a, b), the parameter b approximates the prior sample size s_0. This is because the gamma prior for lambda when we have $\mathbf{Y} = [y_1, y_2, ...y_n]$, $b = b_0 + n$.

1f: Name your store, and the date and time.

I will visit the Santa Monica REI on Friday, October 14 2016 and watch for people coming in between 10:15-10:20 AM.

1g: We are now going to specify the parameters a and b of the gamma prior density.

We will do this in two different ways, giving two different priors. We designate one set of prior parameters as a1 and b1; the other set of prior parameters are a2 and b2.

1. Before you visit the store, make a guess as to the mean number of customers entering the store in one minute. Call this m0. This is the mean of your prior distribution for lambda

   I estimate a prior mean m_0 of 7 people entering the store per minute.

2. Make a guess s0 of the prior sd associated with your estimate m0. This s0 is the standard deviation of the prior distribution for λ. Note: most people underestimate s0.

   I estimate a standard deviation s_0 of 3 people: sometimes nobody may come in (after all I am going right after opening time on a weekday); other times, a bunch may come in (this store is right by the Santa Monica promenade, so maybe lots of tourist groups?)

3. Separately from the previous question 1(g)ii, estimate how many data points n0 your prior guess is worth. That is, n0 is the number (strictly greater than zero) of data points (counts of 5 minutes) you would just as soon have as have your prior guess of m0.

   For the purpose of this exercise, I will assume that **my prior is worth just 1 data point**. I don't frequent stores and am not tracking customer dynamics when I do, so I am going with what I think is a conservative approach.

4. Solve for a1 and b1 based on m0 and s0.

$$E[\lambda] = \frac{\alpha}{\beta}; Var[\lambda] = \frac{\alpha}{\beta^2}$$

$$7 = \frac{\alpha}{\beta}; 9 = \frac{\alpha}{\beta^2}$$

$$\alpha_1 = 49/9; \beta_2 = 7/9$$

5. Separately solve for a2 and b2 using m0 and n0 only. You usually will not get the same answer each time. This is ok and is NOT wrong. (Note: if you do get the same answer, then please specify a second choice of a2 , b2 to use with the remainder of this problem!)

   We can set $\beta = n_0$, as the value of n modifies the rate parameter in our posterior derived above. Therefore, $\beta = 1$, and we can use this value to calculate α:

$$E[\lambda] = \frac{\alpha}{\beta}$$

$$7 = \frac{\alpha}{1}$$

$$\alpha = 7; \beta = 1$$

The variance of this would also be 7, as $\frac{\alpha}{\beta^2} = \frac{7}{1^2} = 7$.

1h: Suppose we need to have a single prior, rather than two priors. Suggest 2 distinct methods to settle on a single prior.

DO SOMETHING

1i: Go to your store and collect your data as instructed in 1e. Report it here.

Minute 1	Minute 2	Minute 3	Minute 4	Minute 5
3	6	1	0	9

1j: Update both priors algebraically using your 5 data points. Give the two posteriors.

Prior 1

$$\alpha_{posterior} = \alpha_{prior} + \sum_{i=1}^{n} y_i = \frac{49}{9} + 3 + 6 + 1 + 0 + 9 = 24.444$$

$$\beta_{posterior} = \beta_{prior} + n = \frac{7}{9} + 5 = 5.778$$

So, the posterior for prior 1 is of the form $Gamma(\alpha = 24.4444, \beta = 5.778)$

Prior 2

$$\alpha_{posterior} = \alpha_{prior} + \sum_{i=1}^{n} y_i = 7 + 3 + 6 + 1 + 0 + 9 = 26$$

$$\beta_{posterior} = \beta_{prior} + n = 1 + 5 = 6$$

So, the posterior for prior 2 is of the form $Gamma(\alpha = 26, \beta = 6)$

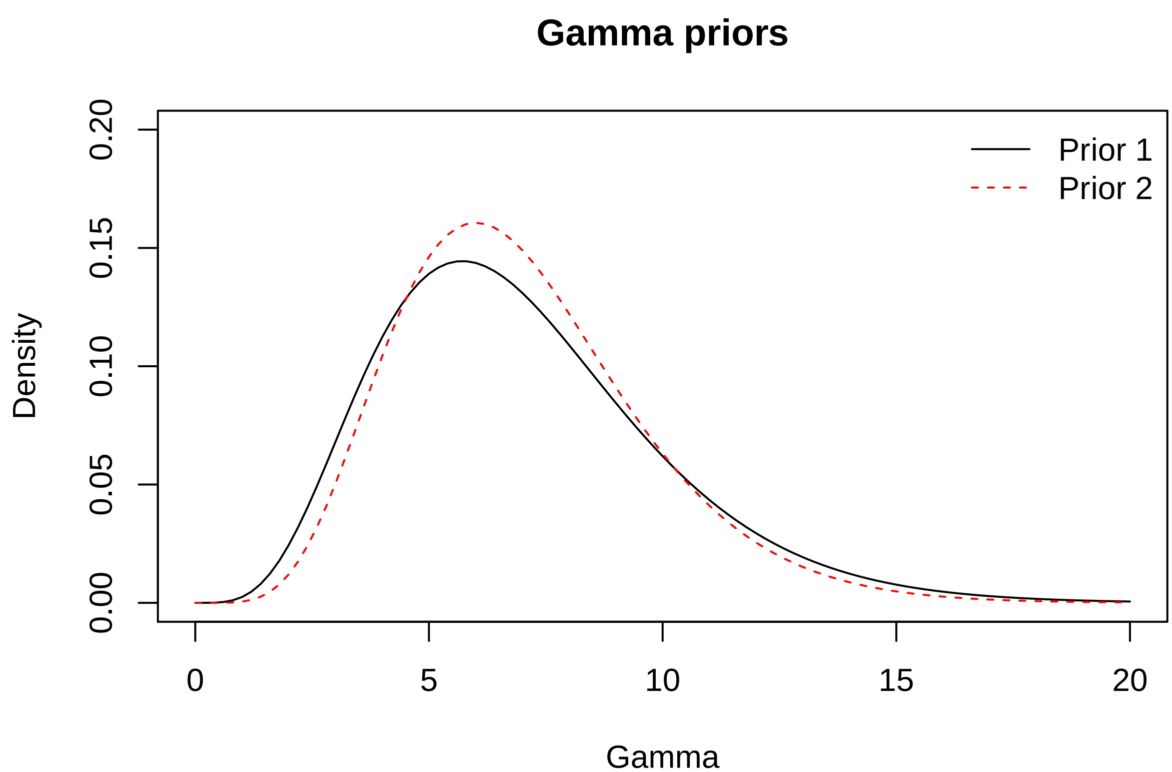
1k: Give the posterior mean and variance for your two posteriors.

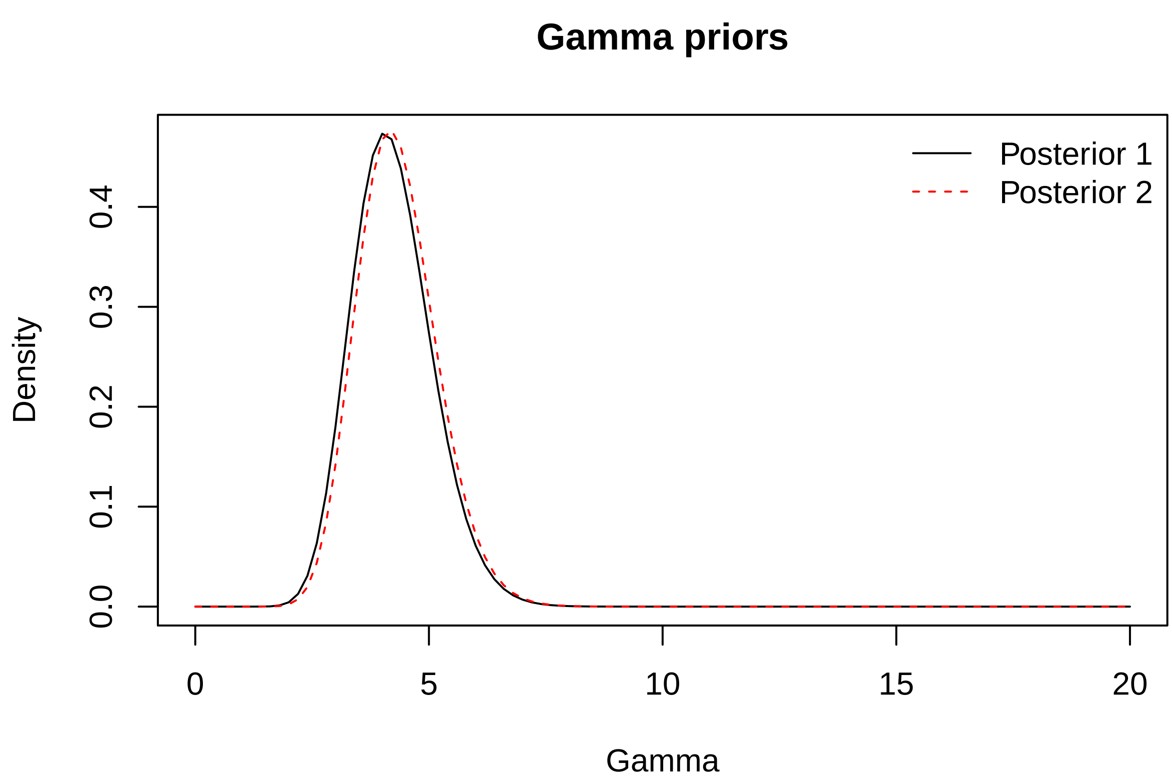
Prior 1

$$\mu_{posterior} = \frac{\alpha_{posterior}}{\beta_{posterior}} = \frac{24.4444}{5.778} = 4.231$$

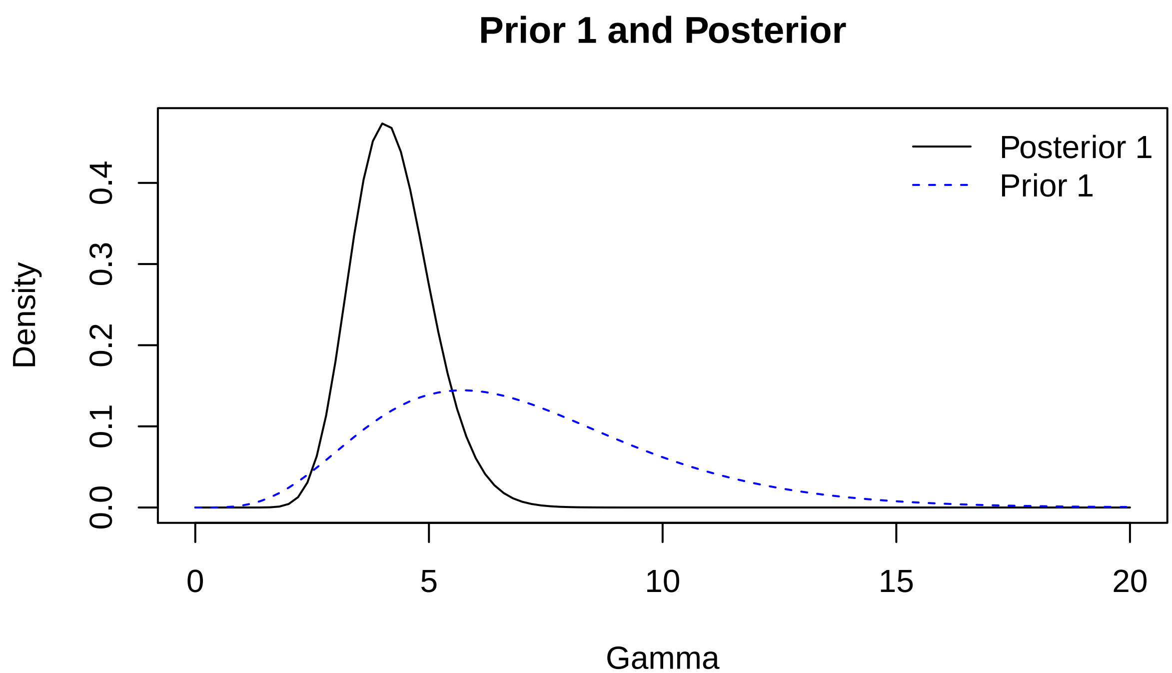
$$Var_{posterior} = \frac{\alpha_{posterior}}{\beta^2_{posterior}} = \frac{24.4444}{5.778^2} = 0.732$$

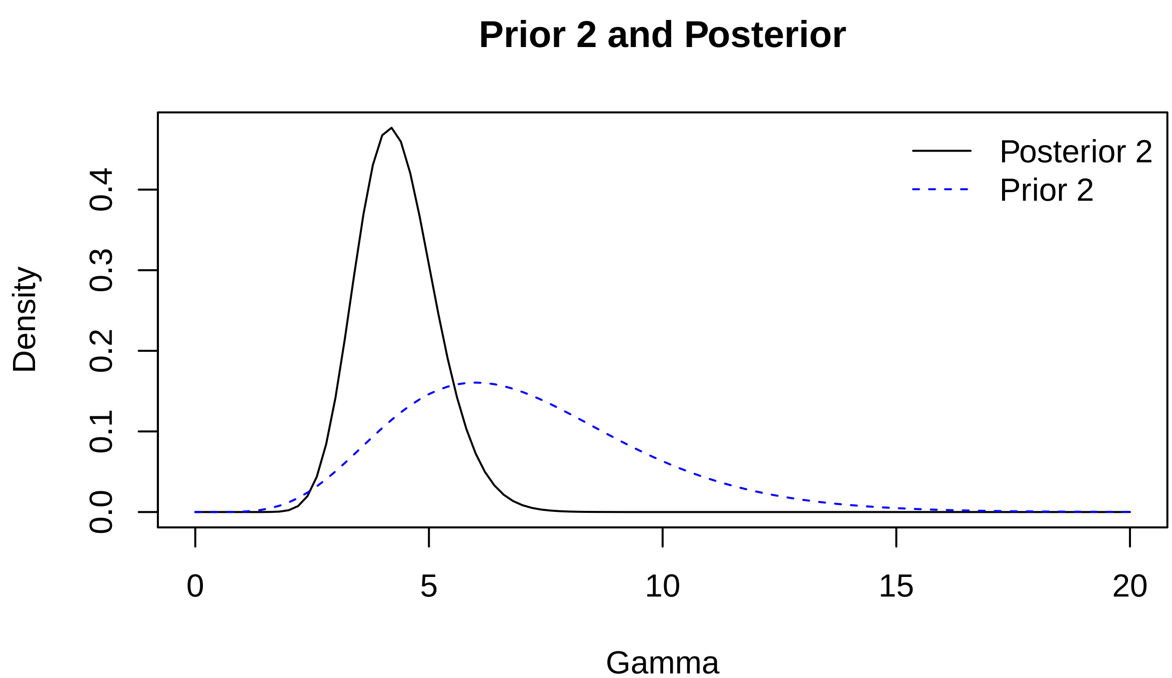
1l: Plot your two prior densities on one graph. Plot your two posterior densities in another graph. In one sentence for each plot, compare the densities.





1m: Plot each prior density/posterior density pair on the same graph. For each plot, compare the two densities in one sentence.



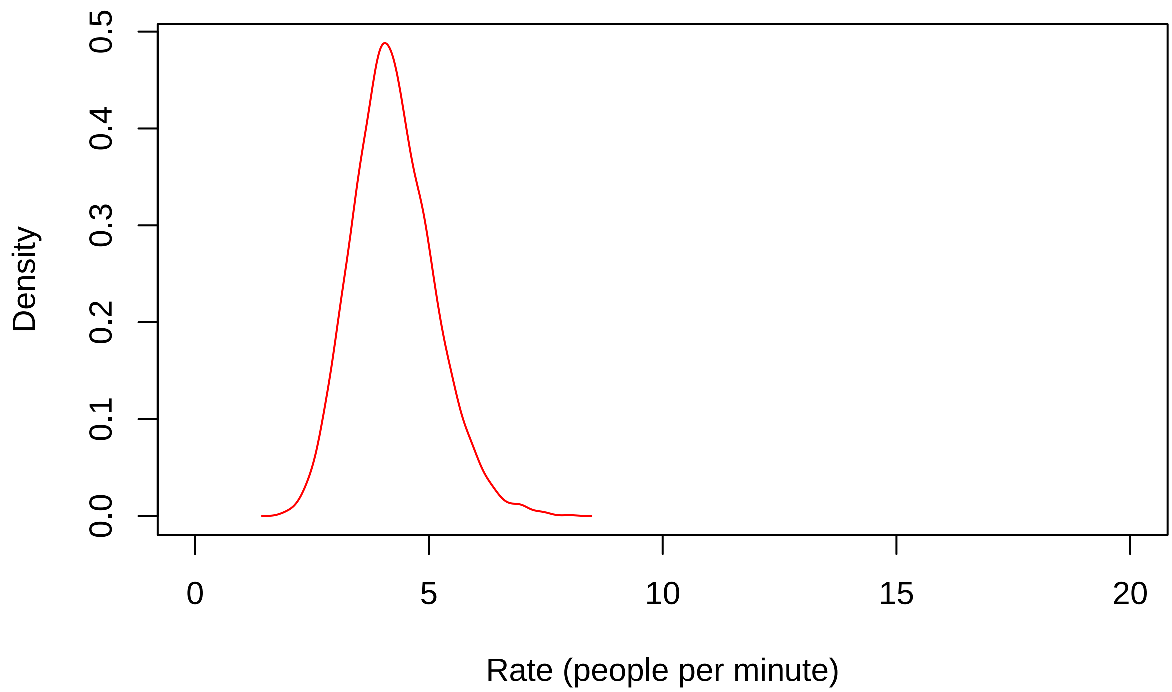


1n: Use WinBUGS (twice) to update your two priors with your data to get your two posteriors. Compare summary statistics between the two posteriors.

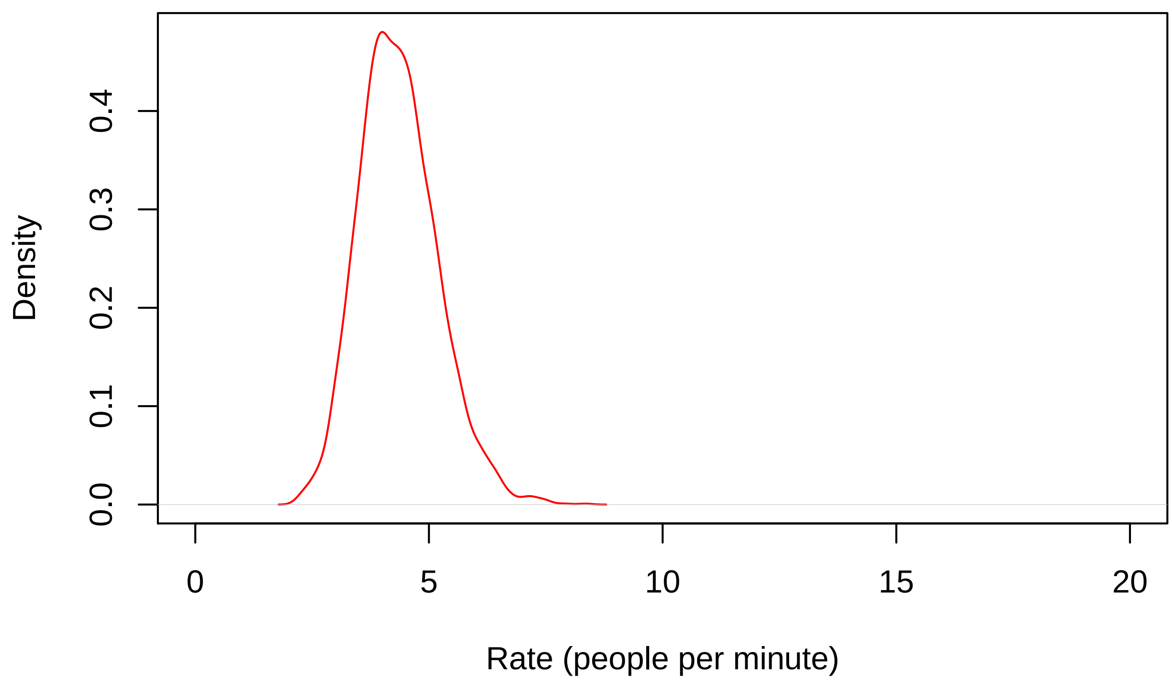
```r
# Write the model
sink("model_hw2.txt")
cat("model {
  # prior
  lambda ~ dgamma(alpha, beta)

  for (i in 1:N){
    x[i] ~ dpois(lambda)
  }


}", fill = TRUE)
sink()
```

```
## module glm loaded

## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 5
##    Unobserved stochastic nodes: 1
##    Total graph size: 9
##
## Initializing model

## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 5
##    Unobserved stochastic nodes: 1
##    Total graph size: 9
##
## Initializing model
```

JAGS simulation of Posterior assuming Priors 1



JAGS simulation of Posterior 1 assuming Priors 2



1o: How close are the WinBUGS numerical calculations to the actual algebraically calculated posterior means?

Table 2: Comparing calculted vs. simulated Posterior statistics

	mean	2.5%	97.5%	sd
Calculated Posterior 1	4.2308	NA	NA	0.8557
Simulated Posterior 1	4.2388	2.7420	6.0942	0.8617
Calculated Posterior 2	4.3333	NA	NA	0.8498
Simulated Posterior 2	4.3522	2.9326	6.1441	0.8335

For Prior 1, the simulated and calculated posterior Mean and SD were very close (10^{-3} accuracy) to the algebraically calculated Mean and SD. That said, the 95% CI around the mean was quite wide in the simulations, ranging from 50% to 150% of the calculated mean.

Question 2

2a: Give algebraic formulas for the relationships between (i) lambda5 and lambda1, (ii) the prior mean of lambda5 and lambda1, (iii) prior variances, (iv) prior standard deviations, (v) prior a-parameters, and (vi) b-parameters.

1. Relationship between Lambda5 and Lambda1

   stuff

2. Prior mean of Lambda5 and Lambda1

   Prior mean of λ_5 = 5*(Prior mean of λ_1)

3. Prior variances of Lambda5 and Lambda 1

   Prior Var of $\lambda_5 = 5^2$*(Prior Var of λ_1)

4. Prior SD of Lambda5 and Lambda1

   Prior SD of λ_5 = 5*(Prior SD of λ_1)

5. Prior alpha of Lambda5 and Lambda1

   $\alpha_{\lambda_5} = 5 * SD_{\lambda_1} * \alpha_{\lambda_1}$

6. Prior beta of Lambda5 and Lambda1

   $\beta_{\lambda_5} = \frac{1}{5} * \beta_{\lambda_1}$

2b: Give the two priors for the parameter lambda5 that correspond to your priors for lambda1

Prior 1

$$\lambda_5 \sim Gamma(5 * 3 * \frac{49}{9}, 3 * \frac{7}{9}) = Gamma(81.667, 2.3333)$$

Prior 2

$$\lambda_5 \sim Gamma(5 * 3 * 7, 3 * 1) = Gamma(105, 3)$$

2c: Give the two resulting posteriors for lambda5 .

Posterior 1

$$\lambda_5 \sim Gamma(5 * 3 * \frac{49}{9} + 19, 3 * \frac{7}{9} + 1) = Gamma(100.67, 3.3333)$$

Posterior 2

$$\lambda_5 \sim Gamma(5 * 3 * 7 + 19, 3 * 1 + 1) = Gamma(124, 4)$$

2d: Explain the relationship between the posterior means of lambda5 and lambda1. Repeat for the posterior variance, posterior standard deviation,posterior a-parameters and finally posterior b parameters.

2e: Do you need to redraw your plots (of priors and posteriors) that you drew in the previous problem? How could you alter them without redrawing to make them conform to the new data structure?

We don't strictly need to redraw the plots of posteriors and priors, since the shapes of the densities should remain the same. However, we will need to change the scale of both axes: the X-mean will now be around 5*old_mean, and the Y-axis will end up being shorter, since the density values at any single value of X will be lower.

2f: Do your conclusions change if you consider your data as a single 5 minute observation or as 5 one minute observations? That is, do your recommendations to the store on staffing levels change?