

# CSI-674 Assignment 4

## Jericho McLeod

[Code ▾](#)

1. This problem continues analysis of the automobile traffic data from Assignment 3. Transforming the arrival times to counts of cars in each 15-second interval gives the following table of counts:

[Hide](#)

```
cars <- seq(0,5,1)
occur <- c(3,5,7,3,3,0)
car_data <- data.frame('Number of Cars' = cars, 'Number of Occurences' = occur)
car_data
```

Number.of.Cars <dbl>	Number.of.Occurences <dbl>
0	3
1	5
2	7
3	3
4	3
5	0

6 rows

1-a) Assume a Poisson likelihood and a uniform prior distribution for the unknown rate  $\Lambda$ . Find the posterior distribution for  $\Lambda$ .

[Hide](#)

```
# The data
obs <- c(0,0,0,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,4,4,4)

alpha0=2; theta0=Inf          # prior shape and scale
alpha1=alpha0+sum(obs)         # posterior shape
theta1=1/(1/theta0+length(obs)) # posterior scale

print(paste0("Posterior Alpha: ", alpha1))
```

```
[1] "Posterior Alpha: 42"
```

[Hide](#)

```
print(paste0("Posterior Beta: ", theta1))
```

```
[1] "Posterior Beta: 0.0476190476190476"
```

**1-b) Find the mean, standard deviation, median and mode of the posterior distribution.**

$$\mu = \alpha \cdot \theta = \alpha \cdot \frac{1}{\beta}$$

$$\sigma = \alpha^{0.5} \cdot \theta$$

$$mode = (\alpha - 1) \cdot \theta$$

Hide

```
require(invgamma)
mean <- alpha1*theta1
median <- 1/qinvgamma(1-0.5,shape=alpha1,scale=theta1)
mode <- (alpha1-1)*theta1
sd <- alpha1^0.5 *theta1
print(paste0("Mean: ", mean))
```

```
[1] "Mean: 2"
```

Hide

```
print(paste0("Median: ", median))
```

```
[1] "Median: 1.98414957498933"
```

Hide

```
print(paste0("Mode: ", mode))
```

```
[1] "Mode: 1.95238095238095"
```

Hide

```
print(paste0("Standard Deviation: ", sd))
```

```
[1] "Standard Deviation: 0.308606699924184"
```

**1-c) Find a 95% symmetric tail area posterior credible interval for  $\Lambda$ .**

Hide

```
top5 <- 1/qinvgamma(1-0.975,shape=alpha1,scale=thetal)
bottom5 <- 1/qinvgamma(1-0.025,shape=alpha1,scale=thetal)
post_credible_interval<- c(bottom5,top5)

print("95% Posterior Credible Interval for Lambda:")
```

```
[1] "95% Posterior Credible Interval for Lambda:"
```

Hide

```
print(post_credible_interval)
```

```
[1] 1.441424 2.648625
```

**1-d) Compare your results with the results from Assignment 3. (Solutions to Assignment 3 will be posted by Wednesday and will be visible to everyone who has turned in Assignment 3.)**

In assignment 3, the 95% credible interval is [1.4, 2.6], compared to here where it is [1.441, 2.648]. These are very similar, and the difference is due to the use of a Gamma distribution rather than a Poisson distribution, giving us a continuous line and more exact answers.

**2. Suppose a highway engineer provided the following prior judgments about the rate of traffic on the stretch of highway (before seeing the data).**

- The rate is equally likely to be above or below 15 cars per minute (or 3.75 cars every 15 seconds).
- There is a 90% chance that the rate is fewer than 28 cars per minute (or 7 cars every 15 seconds).
- There is a 90% chance that the rate is greater than 6 cars per minute (or 1.5 cars every 15 seconds).

**Find a Gamma prior distribution that matches these judgments reasonably well. What are the parameters of this Gamma distribution? How well does it match the engineer's judgments? Comment on whether you think it is reasonable to use this Gamma distribution as a prior distribution for the Poisson rate parameter.**

Hide

```
# Fit using the function in rriskDistributions
require(rriskDistributions)
params <- get.gamma.par(p=c(0.1,0.5,0.9),q=c(1.5,3.75,7))
```

```
The fitting procedure 'L-BFGS-B' was successful!
```

```

$par
[1] 3.2058986 0.7789808

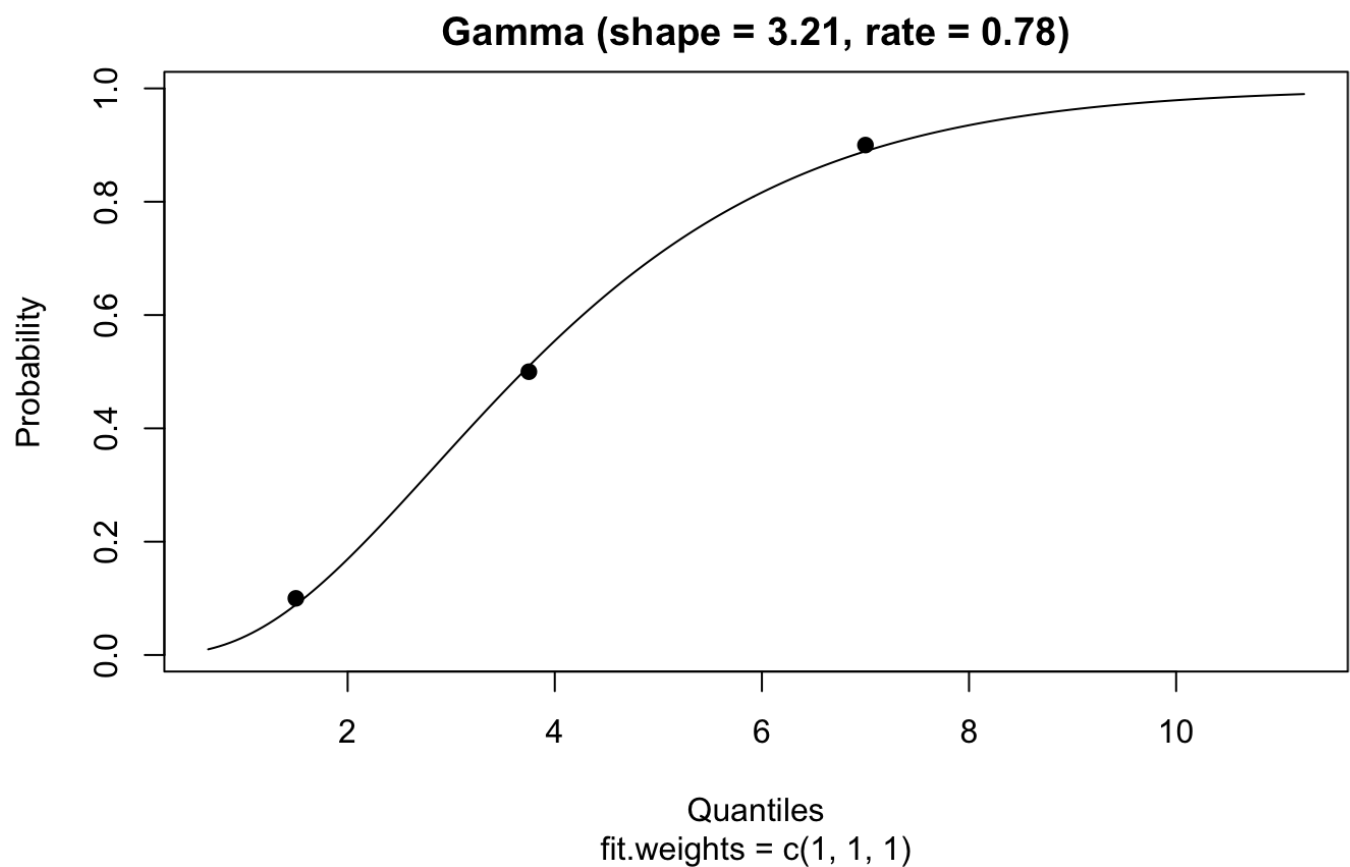
$value
[1] 3.817275e-05

$counts
function gradient
      25      25

$convergence
[1] 0

$message
[1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"

```



Hide

```
print(params)
```

```

      shape      rate
3.2058986 0.7789808

```

Hide

```
#Verify that it matches input reasonably well
# Note: this function returns rate, not shape, thus the inverse is used in the previousl
y used
# invgamma function. This can also be done without all the inversions.
alpha_test <- params[1]
beta_test <- params[2]
median_test <- 1/qinvgamma(1-0.5,shape=alpha_test,scale=(1/beta_test))
top10 <- 1/qinvgamma(1-0.9,shape=alpha_test,scale=(1/beta_test))
bottom10 <- 1/qinvgamma(1-0.1,shape=alpha_test,scale=(1/beta_test))
print(paste0(bottom10," ",median_test," ",top10))
```

```
[1] "1.57859683314186 3.69641444022323 7.19761180136962"
```

The gamma distribution for the prior fits pretty well with inputs  $\alpha = 3.2059$  and rate, or  $\beta = 0.7790$ . The rate and shape are relatively distance from what we expect, but it using this as priors does not impact the posterior gamma distribution very much. I would surmise that so long as the estimate was a reasonably one given a professional duty of care, using it as a prior would be acceptable, and perhaps even preferable, to my initial inputs.

### 3. Repeat Problem 1 using the prior distribution from Problem 2. Compare your results with Problem 1.

Hide

```
comparison_df <-data.frame('Alpha'=c(alpha1),'Theta'=c(theta1),'Mean'=c(mean),'Median'=m
edian,'Mode'=mode,'Standard Deviation'=sd,'95% Cred Interval-Lower'=c(bottom5),'95% Cred
Interval-Upper'=c(top5))

alpha0=alpha_test; theta0=beta_test
alpha1=alpha0+sum(obs)
theta1=1/(1/theta0+length(obs))
mean <-alpha1*theta1
median <- 1/qinvgamma(1-0.5,shape=alpha1,scale=theta1)
mode <- (alpha1-1)*theta1
sd <- alpha1^0.5 *theta1
top5 <- 1/qinvgamma(1-0.975,shape=alpha1,scale=theta1)
bottom5 <- 1/qinvgamma(1-0.025,shape=alpha1,scale=theta1)
post_credible_interval<- c(bottom5,top5)

comparison_df[nrow(comparison_df) + 1,] = list('Alpha'=c(alpha1),'Theta'=c(theta1),'Mea
n'=c(mean),'Median'=median,'Mode'=mode,'Standard Deviation'=sd,'95% Cred Interval-Lower'
=c(bottom5),'95% Cred Interval-Upper'=c(top5))

print(comparison_df)
```

Alpha <dbl>	Theta <dbl>	Mean <dbl>	Median <dbl>	Mode <dbl>	Standard.Deviation <dbl>	X95..Cred.Interva
1 42.0000	0.04761905	2.000000	1.984150	1.952381	0.3086067	1
2 43.2059	0.04487579	1.938899	1.923961	1.894023	0.2949739	1

2 rows | 1-8 of 8 columns

The results using the expert opinion for priors is more concise as an estimate. This makes sense; I initially started with an infinite scale parameter and a very low alpha.

**4. Make a triplot of the prior distribution, normalized likelihood and posterior distribution for Problem 3. Discuss the plot.**

Hide

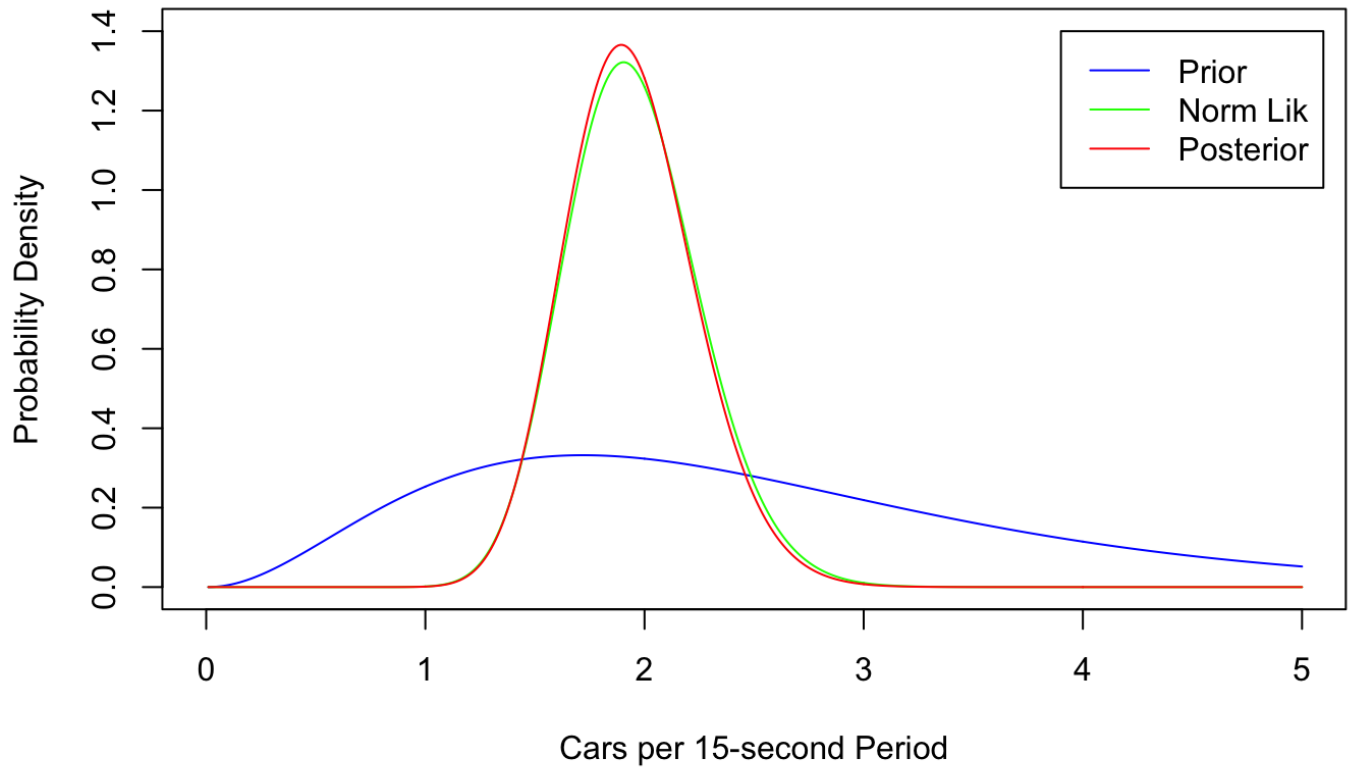
```
# Do a triplot
lambda=seq(length=500,from=0.01,to=5)
priorDens=dgamma(lambda,shape=alpha0,scale=theta0)
postDens=dgamma(lambda,shape=alpha1,scale=theta1)
normLikC=dgamma(lambda,
                 shape=sum(obs)+1,scale=1/length(obs))

plot(lambda,priorDens,type="l",col="blue",main="Triplot for Car Arrival Rates",
     xlab="Cars per 15-second Period",ylab="Probability Density",
     xlim=c(0,5),ylim=c(0,1.4))
lines(lambda,normLikC,col="green")
```

Hide

```
lines(lambda,postDens,col="red")
legend(3.9,1.4,c("Prior","Norm Lik","Posterior"),col=c("blue","green","red"),lty=c(1,1,1))
```

## Triplot for Car Arrival Rates



The triplot shows the prior distribution- in this case from an expert- in blue. This distribution is relatively dispersed. Then, in green, is the normalized likelihood given the data, which is much more narrowly distributed around the mean. Finally, the posterior Gamma distribution, which takes both the prior and normalized likelihoods into account, is even more tightly distributed around its mean, which is between the means of the prior and normalized likelihoods.