

# SDS II - HW 1

Braulio Villalobos-Quiros - Maticola 1999250

4/18/2022

## Question 1 - Rome car accidents

Importing the data...

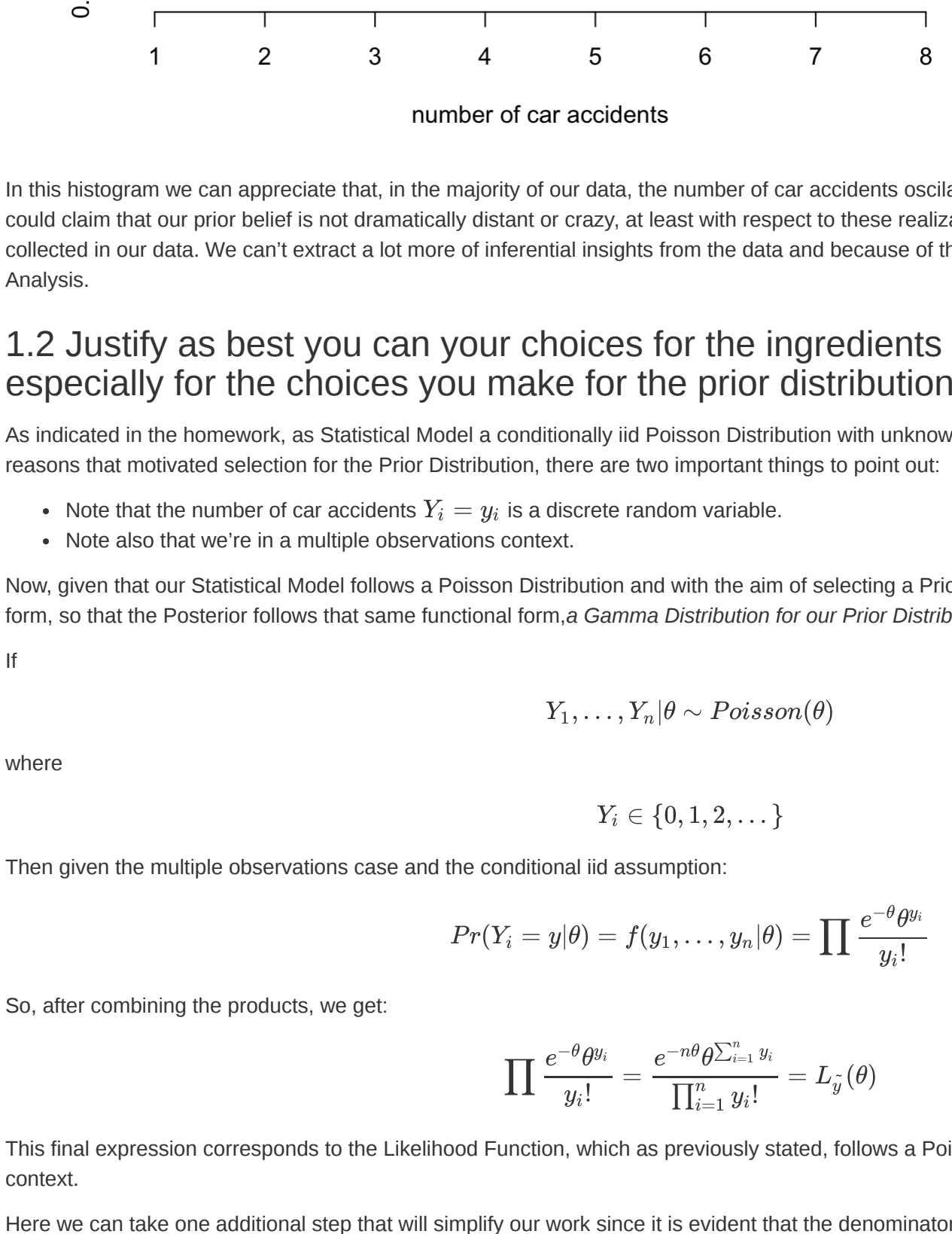
```
mydata <- subset(roma, subset=sign_up_number==184)
```

### 1.1 Describe your observed data

Data is composed of 5 variables and 19 observations.

- week: gives the number of the week in which the accidents took place.
- weekday: identifies the name of the day in which the accidents took place.
- hour: gives the hour at which the accidents took place.
- car\_accidents: gives the number of accidents that were registered on that day, week and hour.
- sign\_up\_number: unspecified but we only consider data with 104 in this variable.

```
hist(mydata$car_accidents, main = "Distribution of Car Accidents",
     xlab = "number of car accidents", probability = T)
```



In this histogram we can appreciate that, in the majority of our data, the number of car accidents oscilates between 1 and 3 car accidents. We could claim that our prior belief is not dramatically distant or crazy, at least with respect to these realizations of the random variable, which we have included in our data. We can't extract a lot more of inferential insights from the data and because of that we will carry out a Fully Bayesian Analysis.

### 1.2 Justify as best you can your choices for the ingredients of your Bayesian model especially for the choices you make for the prior distribution

As indicated in the homework, as Statistical Model a conditionally iid Poisson Distribution with unknown parameter is used. Before explaining the reasons that motivated selection for the Prior Distribution, there are two important things to point out:

- Note that the number of car accidents  $Y_i = y_i$  is a discrete random variable.
- Note also that we're in a multiple observations context.

Now, given that our Statistical Model follows a Poisson Distribution and with the aim of selecting a Prior Distribution that has the same functional form, so that the Posterior follows that same functional form, a *Gamma Distribution for our Prior Distribution* is selected. Following... the reasons:

If

$$Y_1, \dots, Y_n | \theta \sim \text{Poisson}(\theta)$$

where

$$Y_i \in \{0, 1, 2, \dots\}$$

Then given the multiple observations case and the conditional iid assumption:

$$Pr(Y_i = y_i | \theta) = f(y_i, \dots, y_n | \theta) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{y_i}}{y_i!}$$

So, after combining the products, we get:

$$\prod_{i=1}^n \frac{e^{-\theta} \theta^{y_i}}{y_i!} = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!} = L_{\vec{y}}(\theta)$$

This final expression corresponds to the Likelihood Function, which as previously stated, follows a Poisson Distribution for a multiple observations context.

Here we can take one additional step that will simplify our work since it is evident that the denominator of the previous Likelihood Functional doesn't depend on  $\theta$ . For this reason, we can treat the denominator as a multiplicative constant and use the proportionality such that:

$$L_{\vec{y}}(\theta) \propto \frac{e^{-n\theta} \theta^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!} \propto e^{-n\theta} \theta^{\sum_{i=1}^n y_i}$$

Now, based on this, we need to select a convenient Prior so that it shows the same functional form as the Likelihood we just derived and therefore will make our inference more direct and easier to interpret and update. Hence, keeping this in mind, consider the following if we select a Gamma Distribution for our Prior:

$$\theta \sim \text{Gamma}(\text{rate} = r, \text{shape} = s)$$

and

$$\Pi_{\vec{h}}(\theta) = \Pi(\theta); h = (r, s)$$

where  $h$  is the vector of components of the hyperparameter of the Gamma Distribution.

Hence, we have that:

$$\Pi_{\vec{h}}(\theta) = \frac{r^s}{\Gamma(s)} e^{-r\theta} \theta^{s-1} I_{(0, +\infty)}(\theta)$$

By selecting this prior, we can see the functional form is similar to the Poisson one, except for the proportionality constant in front of the expression. So, once again, taking this as a proportionality constant as it doesn't depend on  $\theta$ , we obtain the following:

$$\Pi(\theta) \propto e^{-r\theta} \theta^{s-1}$$

We now can see this appears to be quite similar to the Likelihood function up to a proportionality constant that we previously derived. So, as a quick recap we have that:

$$\Pi(\theta) \propto e^{-r\theta} \theta^{s-1} \text{ and } L_{\vec{y}}(\theta) \propto e^{-n\theta} \theta^{\sum_{i=1}^n y_i}$$

Therefore, combining this prior and the likelihood, we can obtain a **posterior** in the same functional form as the prior such that

$$\Pi(\theta | \vec{y}) \propto e^{-(n+r)\theta} \theta^{\sum_{i=1}^n y_i + s - 1} \text{ where } h^{\text{post}} = h^* = (n + r, \sum_{i=1}^n y_i + s)$$

### 1.3 Main inferential findings

Consider that the average number of hourly car accidents in Rome during the day is 3.22

Now, taking into account that the Prior Distribution is centered in 3.22 and we decide to take a variance of 2 because we decide that we never hear that a day passes by in Rome where there isn't at least one accident but there are also some days with a lot of accidents, then following is the explanation of how we derive the values of  $s$  and  $r$ .

In a Gamma Distribution we have:

$$\mu = \frac{\text{shape}}{\text{rate}} \text{ and } \sigma^2 = \frac{\text{shape}}{\text{rate}^2}$$

Then, solving for the Gamma parameters we obtain:

$$\text{shape} = \mu * \text{rate} \text{ and } \text{rate} = \frac{\mu}{\sigma^2}$$

Then if we substitute the values of  $\mu$  and  $\sigma^2$  in the formula of the rate, with the corresponding values that were previously defined according to our **prior belief**, we obtain:

$$\text{rate} = \frac{\mu}{\sigma^2} = \frac{3.22}{2}$$

And consequently,

$$\text{shape} = \mu * \text{rate} = 3.22 * \frac{3.22}{2} = \frac{10.3684}{2}$$

For this reason, we define the values of  $s = \text{shape}$  and  $r = \text{rate}$  for our Prior Distribution as:

```
s <- 10.3684/2
r <- 3.2209/2
```

If we simulate 100000 times from a Gamma distribution with shape = 10.3684/2 and rate = 3.2209/2, we obtain the following results

```
mean(rgamma(100000, shape=s, rate=r))
```

```
## [1] 3.220577
```

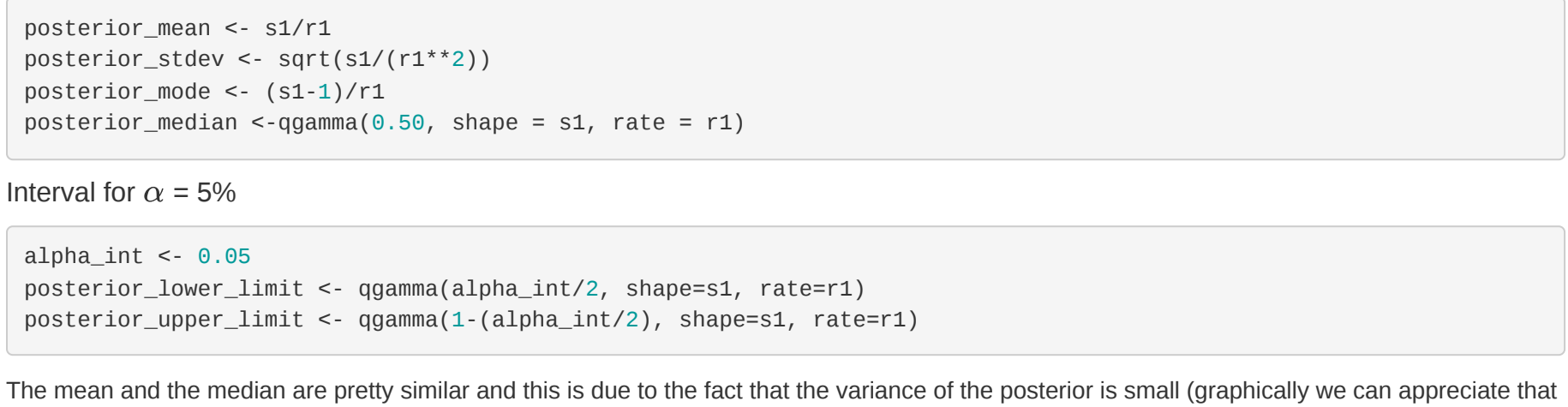
```
var(rgamma(100000, shape=s, rate=r))
```

```
## [1] 1.981943
```

By the Strong Law of Large Numbers, we know that in the limit, as the number of simulations increase, the random value will degenerate towards our mean and variance previously defined.

Prior Distribution

The following graph pictures the Gamma Prior Distribution defined by the  $\text{shape} = 10.3684/2$  and  $\text{rate} = 3.2209/2$  values previously defined and explained. This reflects our prior beliefs and what we think, before observing data, about the phenomena that we are trying to model and make inference from.



Posterior Distribution

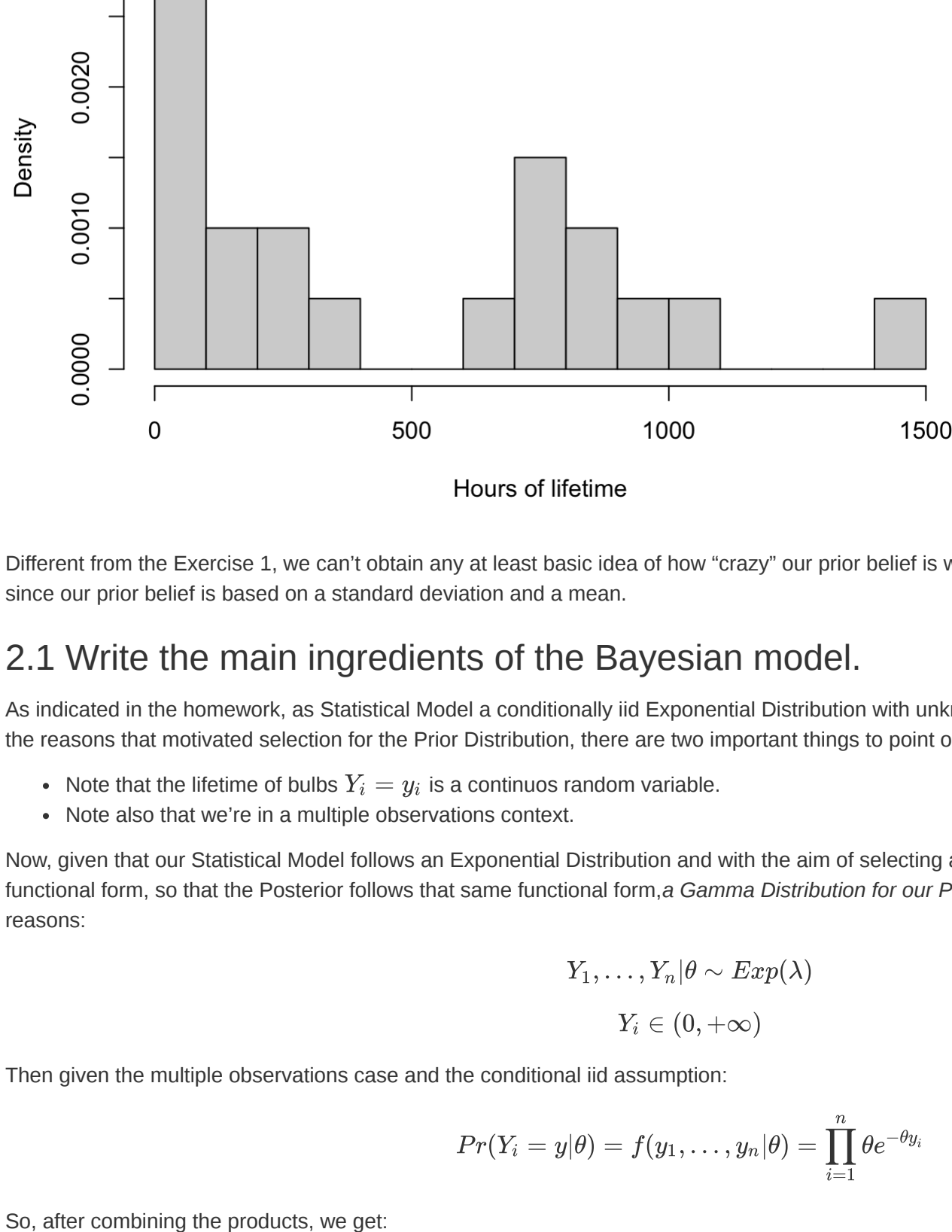
The following graph pictures the posterior distribution. This is obtained by means of **including the observed data along with the prior**. By using the  $h^*$  hyperparameter defined above, where the  $s^* = \sum_{i=1}^n y_i + s$  and  $r^* = n + r$ . In words, the new shape parameter corresponds to the sum of the old  $s$  parameter plus the sum of the observations of car accidents, while the new rate parameter corresponds to the old  $r$  parameter plus the total number of observations.

Hence, we obtain this by including the observed data and therefore update our prior belief

```
s1 <- s + sum(mydata$car_accidents)
r1 <- r + length(mydata$car_accidents)

curve(dgamma(x, rate=r1, shape=s1), from=0, to=10, col="blue",
      main=paste0("Gamma Prior and Posterior Distribution"), n=10000, col="cyan", type = "l", ylim=c(0, 1))
abline(v=c(1,1)), n=10000, col="cyan", type = "l")
```

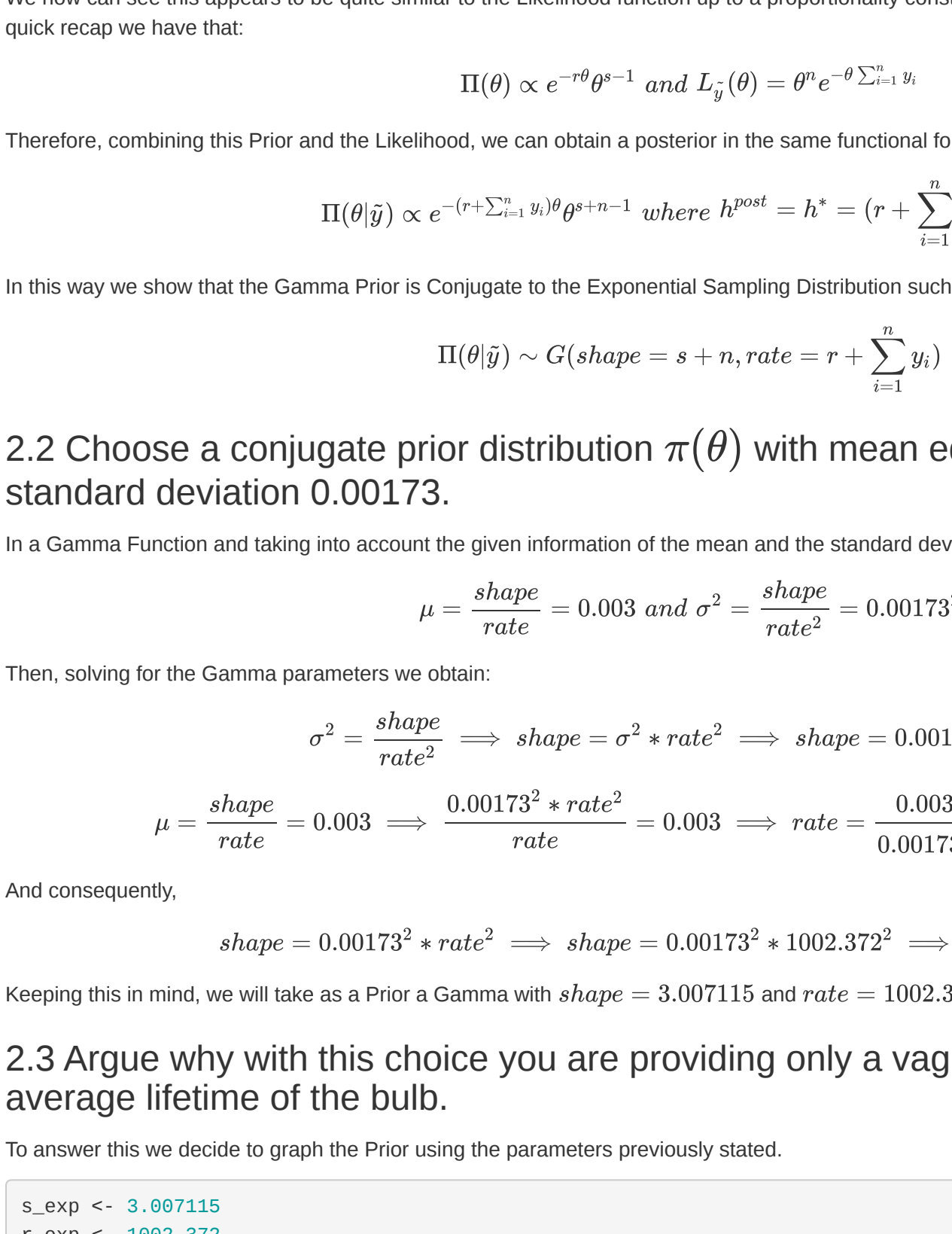
### Gamma Posterior Distribution with rate r=20.6 and shape s=79.2



Now, the following graph allows us to appreciate more clearly how the observed data influences our prior belief and how the posterior is nearer to the Likelihood, as we obtain more data.

```
curve(dgamma(x, rate=r, shape=s), from=0, to=10, xlab=expression(theta), ylab=expression(pi(theta)),
      main=paste0("Gamma Prior and Posterior Distribution"), n=10000, col="cyan", type = "l", ylim=c(0, 1))
curve(dgamma(x, rate=r1, shape=s1), from=0, to=10, col="blue", lty=2, add = T)
abline(v=mean(mydata$car_accidents), col = "red")
```

### Gamma Prior and Posterior Distribution



As we obtain more data, the prior belief becomes more and more negligible. We can appreciate this more clearly if we analyze the following formula for the Expected value of the Posterior:

$$\mu^* = \frac{s^*}{r^*} = \frac{s + \sum_{i=1}^n y_i}{r + n} = \frac{s}{r + n} + \frac{\sum_{i=1}^n y_i}{r + n} = \mu \frac{\psi}{r + n} + \bar{y}_n \frac{n}{r + n} = \mu w + \bar{y}_n (1 - w)$$

See that  $\mu^*$  has two components, the first one corresponds to the prior mean ( $\mu$ ) and the second one corresponds to the MLE in Poisson lid case  $\bar{y}_n$  and each of these terms have a weight. In other words, the updating of the new posterior expectation, which is the center of the posterior distribution, can be interpreted as a weighted average of two components.

Now, note that as  $n$  becomes bigger, the term  $w$  becomes smaller and therefore the term  $\mu w$  also becomes smaller. This means that as we collect more information, the prior becomes more and more negligible.

For this reason is that in the Gamma Prior and Posterior Distribution Graph, we can see that the Posterior (once we observe data) is way nearer to the Likelihood than the Prior (before seeing data). Our believe updates as we observe more data.

Posterior uncertainty

Regarding the posterior uncertainty if we analyze the Gamma Prior and Posterior Distribution Graph, we can already have a good understanding of how once we observe data the uncertainty of the Posterior decreases in comparison with the uncertainty of the parameter used for the prior. This decrease in the uncertainty can be appreciated by

- Mean: the mean of the Posterior (once data is observed) is nearer to the Likelihood than the mean of the Prior.
- Variance: the variance of the Posterior (once data is observed) is smaller than the variance of the Prior.

This can be seen graphically in the Gamma Prior and Posterior Distribution Graph or also on the following Comparison table where we can observe the mean of the Posterior is bigger than the one of the Prior, as equally as the variance of the Posterior is smaller than the one of the Prior.

```
set.seed(123)
mean_vec <- c(round(mean(rgamma(100000, shape=s, rate=r)), 3), round(mean(rgamma(100000, shape=s1, rate=r1)), 3))
var_vec <- c(round(var(rgamma(100000, shape=s, rate=r)), 3), round(var(rgamma(100000, shape=s1, rate=r1)), 3))

comp_df <- as.data.frame(rbind(mean_vec, var_vec))
colnames(comp_df) <- c("Prior", "Posterior")
comp_df
```

```
##      Prior Posterior
## mean_vec 3.223      3.842
## var_vec  2.093      0.186
```

### Possible alternative point estimates & Interval

Alternative Point Estimate

```
posterior_mean <- s1/r1
posterior_stdv <- sqrt(s1/(r1**2))
posterior_mode <- (s1-1)/r1
posterior_median <- qgamma(0.50, shape = s1, rate = r1)
```

Interval for  $\alpha = 5\%$

```
alpha_int <- 0.05
posterior_lower_limit <- qgamma(alpha_int/2, shape=s1, rate=r1)
posterior_upper_limit <- qgamma(1-alpha_int/2, shape=s1, rate=r1)
```

The mean and the median are pretty similar and this is due to the fact that the variance of the posterior is small (although we can appreciate that the Posterior is very concentrated, unlike the Prior). With respect to the lower and upper limit for an alpha = 5%, they allow us to be more secure of the decision we have made.

```
summary_table <- rbind(
  mode=posterior_mode,
  mean=posterior_mean,
  median=posterior_median,
  stdv=posterior_stdv,
  lower_limit=posterior_lower_limit,
  upper_limit=posterior_upper_limit)
colnames(summary_table) <- "Post"
summary_table
```

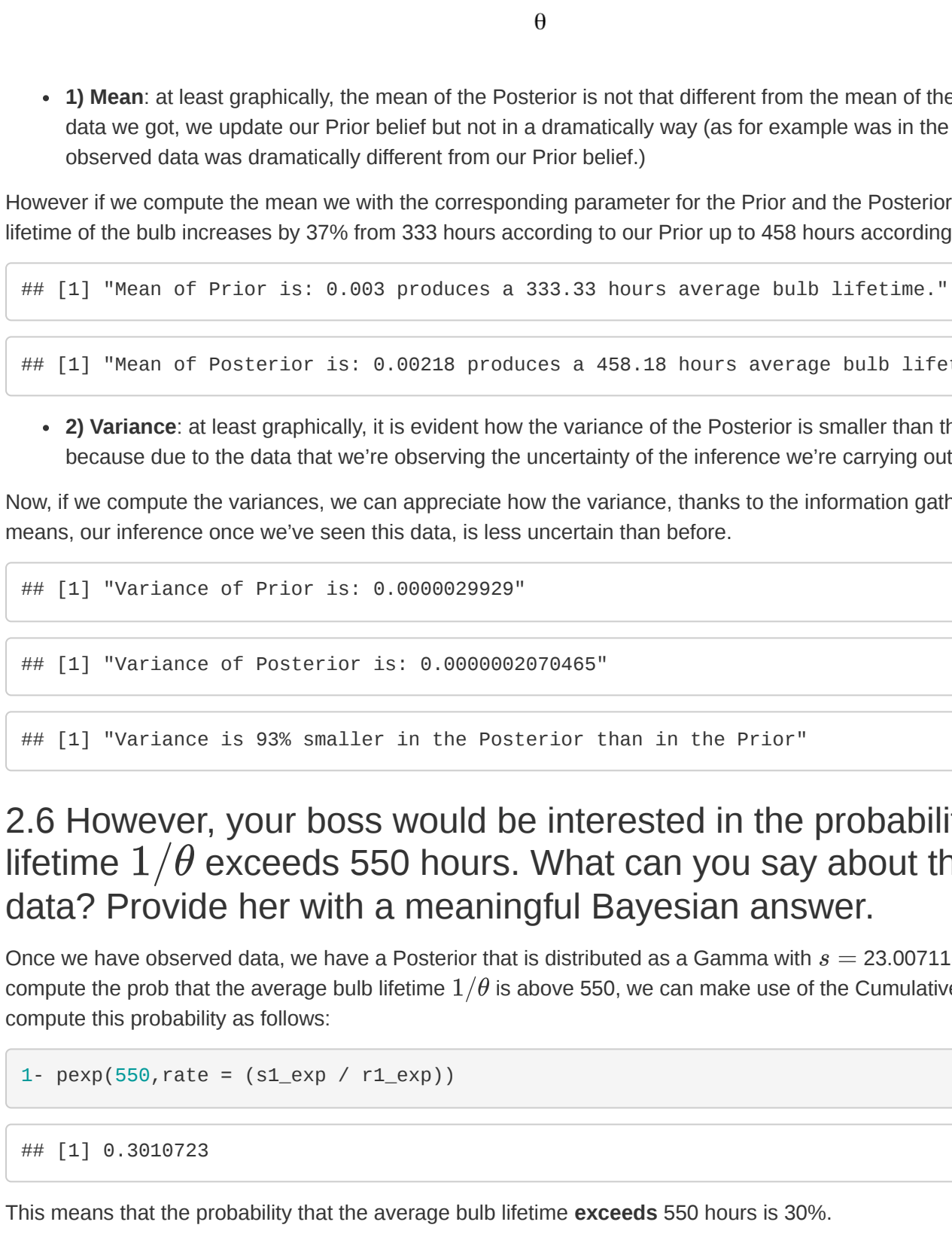
```
##      Post
## mode    3.7935080
## mean    3.8426281
## median  3.8258669
## stdv     0.4317589
## lower_limit 3.0426456
## upper_limit 4.7323288
```

## Question 2 - Bulb Lifetime

Importing data...

```
bulb_data <- c(1,13,27,43,73,75,154,196,220,297,344,619,734,783,796,845,859,992,1086,1471)
```

```
hist(bulb_data, main = "Distribution of Lifetime of Bulbs",
     xlab = "Hours of lifetime", probability = T, breaks = 16)
```



Different from the Exercise 1, we can't obtain any at least basic idea of how "crazy" our prior belief is with respect to these realizations of data, since our prior belief is based on a standard deviation and a mean.

### 2.1 Write the main ingredients of the Bayesian model.

As indicated in the homework, as Statistical Model a conditionally iid Exponential Distribution with unknown parameter is used. Before explaining the reasons that motivated selection for the Prior Distribution, there are two important things to point out:

- Note that the lifetime of bulbs  $Y_i = y_i$  is a continuous random variable.
- Note also that we're in a multiple observations context.

Now, given that our Statistical Model follows an Exponential Distribution and with the aim of selecting a Prior Distribution that has the same functional form, so that the Posterior follows that same functional form, a *Gamma Distribution for our Prior Distribution* is selected. Following... the reasons:

$$Y_1, \dots, Y_n | \theta \sim \text{Exp}(\lambda)$$

$$Y_i \in (0, +\infty)$$

Then given the multiple observations case and the conditional iid assumption:

$$Pr(Y_i = y_i | \theta) = f(y_i, \dots, y_n | \theta) = \prod_{i=1}^n \theta e^{-\theta y_i}$$

So, after combining the products, we get:

$$\prod_{i=1}^n \theta e^{-\theta y_i} = \theta^n e^{-\theta \sum_{i=1}^n y_i} = L_{\vec{y}}(\theta)$$

This final expression corresponds to the Likelihood Function, which as previously stated, follows an Exponential Distribution for a multiple observations context.

Now, based on this, we need to select a convenient Prior so that it shows the same functional form as the Likelihood we just derived and therefore will make our inference more direct and easier to interpret and update. Hence, keeping this in mind, consider the following if we select a Gamma Distribution for our Prior:

$$\theta \sim \text{Gamma}(\text{rate} = r, \text{shape} = s)$$

and

$$\Pi_{\vec{h}}(\theta) = \Pi(\theta); h = (r, s)$$

where  $h$  is the vector of components of the hyperparameter of the Gamma Distribution.

Hence, we have that:

$$\Pi_{\vec{h}}(\theta) = \frac{r^s}{\Gamma(s)} e^{-r\theta} \theta^{s-1} I_{(0, +\infty)}(\theta)$$

By selecting this prior, we can see the functional form is similar to the Exponential one, except for the proportionality constant in front of the expression. So, once again, taking this as a proportionality constant as it doesn't depend on  $\theta$ , we obtain the following:

$$\Pi(\theta) \propto e^{-r\theta} \theta^{s-1}$$

We now can see this appears to be quite similar to the Likelihood function up to a proportionality constant that we previously derived. So, as a quick recap we have that:

$$\Pi(\theta) \propto e^{-r\theta} \theta^{s-1} \text{ and } L_{\vec{y}}(\theta) = \theta^n e^{-\theta \sum_{i=1}^n y_i}$$

Therefore, combining this Prior and the Likelihood, we can obtain a posterior in the same functional form as the prior such that

$$\Pi(\theta | \vec{y}) \propto e^{-(r+\sum_{i=1}^n y_i)\theta} \theta^{s+n-1} \text{ where } h^{\text{post}} = h^* = (r + \sum_{i=1}^n y_i, s + n)$$

In this way we show that the Gamma Prior is Conjugate to the Exponential Sampling Distribution such that:

$$\Pi(\theta | \vec{y}) \sim G(\text{shape} = s + n, \text{rate} = r + \sum_{i=1}^n y_i)$$

### 2.2 Choose a conjugate prior distribution $\pi(\theta)$ with mean equal to 0.003 and standard deviation 0.00173.

In a Gamma Function and taking into account the given information of the mean and the standard deviation, we have:

$$\mu = \frac{\text{shape}}{\text{rate}} = 0.003 \text{ and } \sigma^2 = \frac{\text{shape}}{\text{rate}^2} = 0.00173^2$$

Then, solving for the Gamma parameters we obtain:

$$\sigma^2 = \frac{\text{shape}}{\text{rate}^2} \implies \text{shape} = \sigma^2 * \text{rate}^2 \implies \text{shape} = 0.00173^2 * \text{rate}^2$$

$$\mu = \frac{\text{shape}}{\text{rate}} = 0.003 \implies \frac{0.00173^2 * \text{rate}^2}{\text{rate}} = 0.003 \implies \text{rate} = \frac{0.003}{0.00173^2} \implies \text{rate} = 1002.372$$

And consequently,

$$\text{shape} = 0.00173^2 * \text{rate}^2 \implies \text{shape} = 0.00173^2 * 1002.372^2 \implies \text{shape} = 3.007115$$

Keeping this in mind, we will take as a Prior a Gamma with  $\text{shape} = 3.007115$  and  $\text{rate} = 1002.372$ .

### 2.3 Argue why with this choice you are providing only a vague prior opinion on the average lifetime of the bulb.

To answer this we decide to graph the Prior using the parameters previously stated.

```
s_exp <- 3.007115
r_exp <- 1002.372

curve(dgamma(x, rate=r_exp, shape=s_exp), from=0, to=0.01, xlab=expression(theta), ylab=expression(pi(theta)),
      main=paste0("Gamma Prior and Posterior Distribution"), n=10000, col="cyan", type = "l", ylim=c(0, 1000))
curve(dgamma(x, rate=r1_exp, shape=s1_exp), from=0, to=0.01, col="blue", lty=2, add = T)
abline(v=(1/mean(bulb_data)), col = "red")
```

### Gamma Prior Distribution with rate r=1002.4 and shape s=3.01



Posterior Distribution

The following graph pictures the posterior distribution. This is obtained by means of **including the observed data along with the prior**. Here the  $s^* = s + n$  and  $r^* = r + \sum_{i=1}^n y_i$ . In words, the new shape parameter corresponds to the sum of the old  $s$  parameter plus the total number of observations of bulbs, while the new rate parameter corresponds to the old  $r$  parameter plus the total lifetime on hours of the bulbs in the data we have.

Hence, we obtain this by including the observed data and therefore update our prior belief

```
s1_exp <- s_exp + length(bulb_data)
r1_exp <- r_exp + sum(bulb_data)

curve(dgamma(x, rate=r1_exp, shape=s1_exp), from=0, to=0.01, col="blue",
      xlab=expression(theta), ylab=expression(pi(theta)),
      main=paste0("Gamma Prior and Posterior Distribution"), n=10000, col="cyan", type = "l", ylim=c(0, 1000))
curve(dgamma(x, rate=r1_exp, shape=s1_exp), from=0, to=0.01, col="blue", lty=2, add = T)
abline(v=(1/mean(bulb_data)), col = "red")
```

### Gamma Posterior Distribution with rate r=10541.4 and shape s=23



Now, the following graph allows us to appreciate more clearly 2 basic things:

```
curve(dgamma(x, rate=r_exp, shape=s_exp), from=0, to=0.01, xlab=expression(theta), ylab=expression(pi(theta)),
      main=paste0("Gamma Prior and Posterior Distribution"), n=10000, col="cyan", type = "l", ylim=c(0, 1000))
curve(dgamma(x, rate=r1_exp, shape=s1_exp), from=0, to=0.01, col="blue", lty=2, add = T)
abline(v=(1/mean(bulb_data)), col = "red")
```

### Gamma Prior and Posterior Distribution



- **1) Mean:** at least graphically, the mean of the Posterior is not that different from the mean of the Prior. This could mean that by observing the data we got, we update our Prior belief but not in a dramatically way (as for example was in the exercise 1 where the mean once we observed data was dramatically different from our Prior belief).

However if we compute the mean we will with the corresponding parameter for the Posterior and the Posterior we have that by observing the average lifetime of the bulb increases by 37% from 333 hours according to our Prior up to 458 hours according to our Posterior.

```
## [1] "Mean of Prior is: 0.003 produces a 333.33 hours average bulb lifetime."
```

```
## [1] "Mean of Posterior is: 0.002128 produces a 458.18 hours average bulb lifetime."
```

- **2) Variance:** at least graphically, it is evident how the variance of the Posterior is smaller than the variance of the Prior. This is desired because due to the data that we're observing the uncertainty of the inference we're carrying out diminishes.

Now, if we compute the variances, we can appreciate how the variance, thanks to the information gathered on the 20 bulbs, decreases 93%. This means, our inference once we've seen this data, is less uncertain than before.

```
## [1] "Variance of Prior is: 0.0000029929"
```

```
## [1] "Variance of Posterior is: 0.000002070465"
```

```
## [1] "Variance is 93% smaller in the Posterior than in the Prior"
```

### 2.6 However, your boss would be interested in the probability that the average bulb lifetime $1/\theta$ exceeds 550 hours. What can you say about that after observing the data? Provide her with a meaningful Bayesian answer.

Once we have observed data, we have a Posterior that is distributed as a Gamma with  $s = 23.007115$  and  $r = 10541$ . Then if we want to compute the prob that the average bulb lifetime  $1/\theta$  is above 550, we can make use of the Cumulative Distribution Function (CDF) and the  $\theta$  to compute this probability as follows:

```
1- pexp(550, rate = (s1_exp / r1_exp))
```

```
## [1] 0.3010723
```

This means that the probability that the average bulb lifetime **exceeds** 550 hours is 30%.

#rmarkdown::render("1999250\_braulio\_villalobos\_sds2\_hw1.Rmd")