Lamp time-to-failure data analysis

Tetsuya Chau 2/18/2022

Task

-I am tasked to analyze the “LCD projector lamp failure” data (only the “Hours” column) from https://alysongwilson.github.io/BR/table42.txt. I will use a Gamma distribution for the likelihood/sampling distribution. An expert suggests that the shape parameter (α) should have a uniform(1, 1.5) prior distribution and the rate parameter (λ) is contained in the interval (0, 0.01) with 95% certainty.

```
#install.packages("R2jags")
library(R2jags)
```

```
## Loading required package: rjags

## Loading required package: coda

## Linked to JAGS 4.3.0

## Loaded modules: basemod,bugs

##
## Attaching package: 'R2jags'

## The following object is masked from 'package:coda':
##
##     traceplot
```

```
#Read in data
dataM <- read.table(url("https://alysongwilson.github.io/BR/table42.txt"),header=T)$H

show(dataM)
```

```
##  [1]  387  182  244  600  627  332  418  300  798  584  660   39  274  174   50
## [16]   34 1895  158  974  345 1755 1752  473   81  954 1407  230  464  380  131
```

```
## [31] 1205
```

Step 1

-Specify an informative exponential prior distribution for the λ parameter and explain the logic of how I obtained it.

I chose the prior to be alpha ~ dunif(1,1.5) because that was the prior that was given to us and lambda ~ dexp(299.57) because that’s the rate parameter that was contained in the interval (0,0.01) with a 95% certainty. I solved this by setting the CDF of the exponential distribution equal to .95, plugging in .01 into x, and solving for lambda.

Step 2

-Using JAGS to define the model, obtain 100,000 samples from the posterior distribution. Show histograms, traceplots and autocorrelation plots to show that the chains converged and are representative of the posterior distribution.

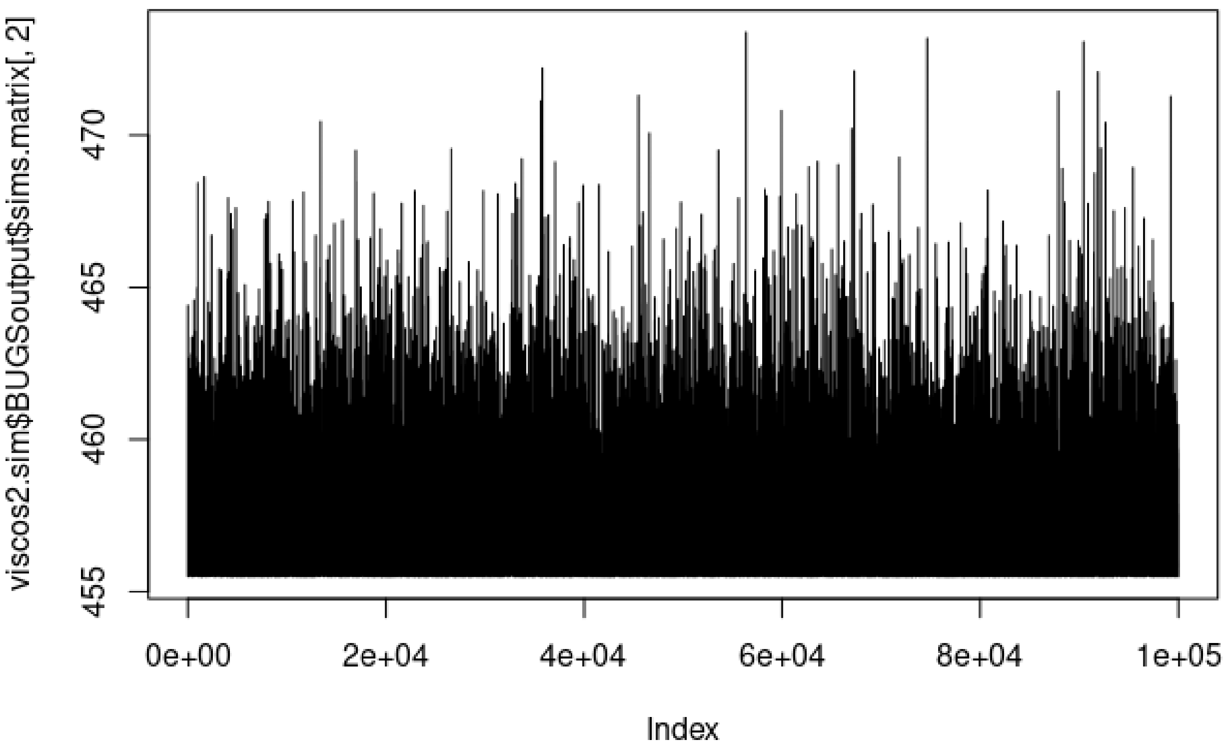
```
md2 <-
"model {
  for(i in 1:31){
    dataM[i] ~ dgamma(alpha,lambda);
  }
  lambda ~ dexp(299.57)
  alpha ~ dunif(1,1.5)
}
"

viscos2.sim <- jags(
  data=c('dataM'),
  parameters.to.save=c('lambda','alpha'),
  model.file=textConnection(md2),
  n.iter=27000,
  n.burnin=2000,
  n.chains=4,
  n.thin=1
)
```

```
## module glm loaded
```

```
## Compiling model graph
```

```
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 31
##    Unobserved stochastic nodes: 2
##    Total graph size: 36
##
## Initializing model
```

```
plot(viscos2.sim$BUGSoutput$sims.matrix[,2],type="l")
```




```
length(viscos2.sim$BUGSoutput$sims.matrix[,2])
```

```
## [1] 100000
```

```
effectiveSize(viscos2.sim$BUGSoutput$sims.matrix[,2])
```

```
##  var1
## 1e+05
```

```
gelman.diag(viscos2.sim$BUGSoutput)
```

```
## Potential scale reduction factors:
##
##          Point est. Upper C.I.
## alpha             1          1
## deviance          1          1
## lambda            1          1
##
## Multivariate psrf
##
## 1
```

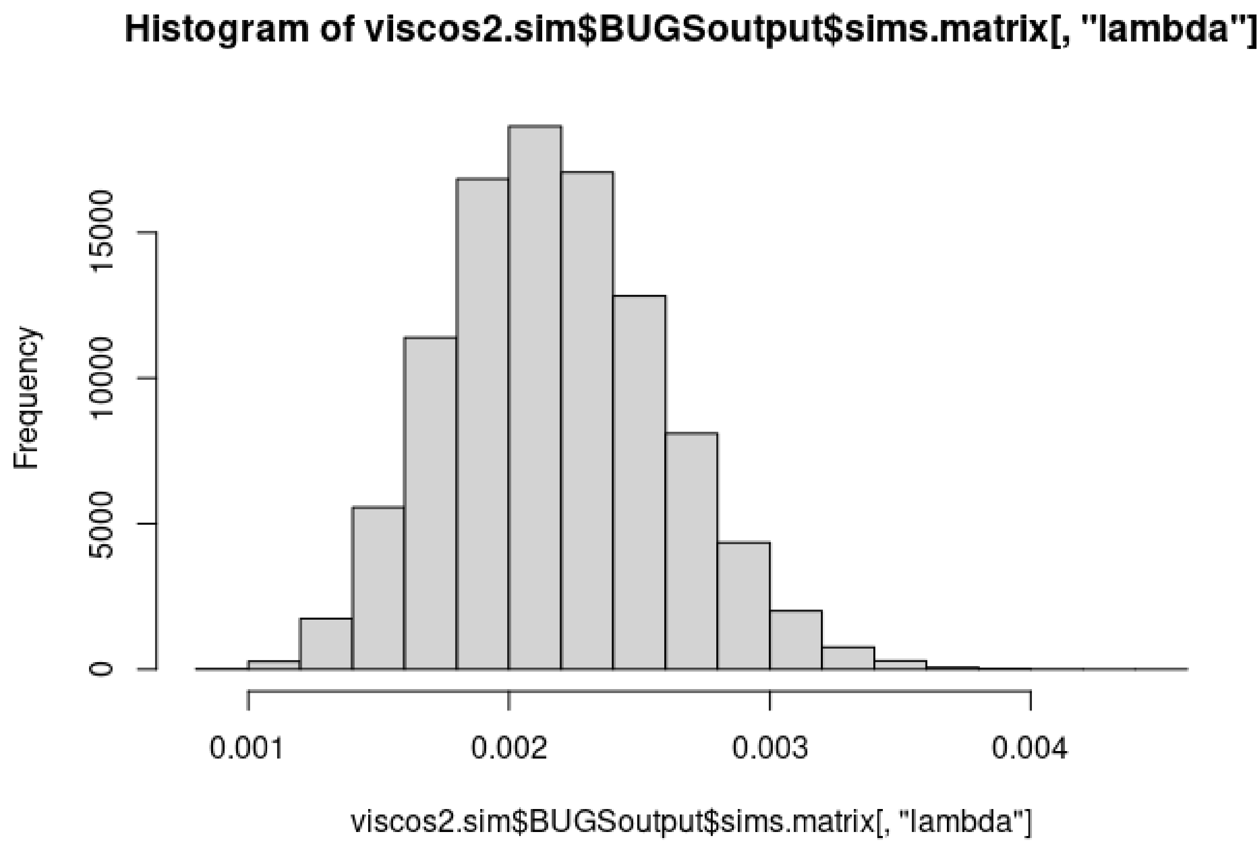
```
#head(viscos.sim$BUGSoutput$sims.matrix)

viscos2.sim$BUGSoutput$DIC
```

```
## [1] 457.7011
```

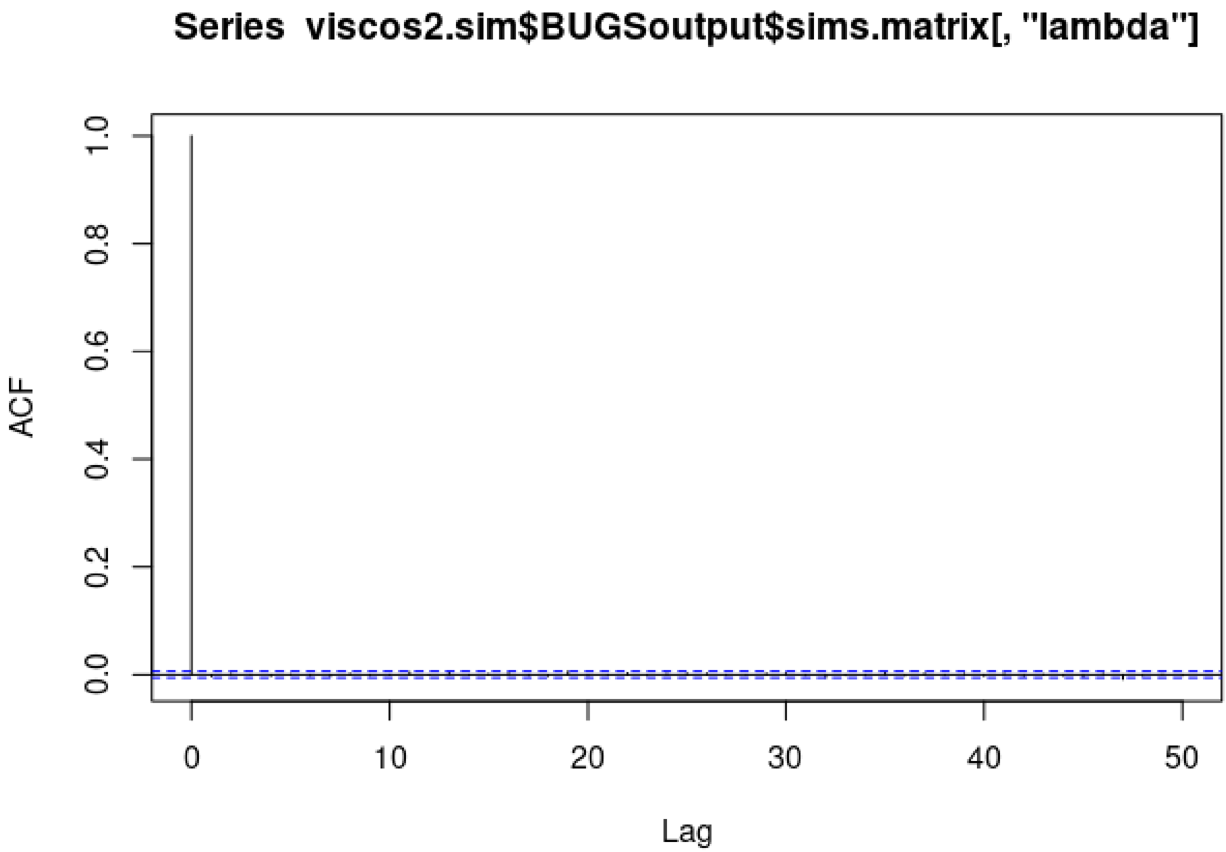
```
head(viscos2.sim$BUGSoutput$sims.matrix)
```

```
##         alpha deviance      lambda
## [1,] 1.199842 455.4986 0.002097713
## [2,] 1.063078 457.7616 0.002317023
## [3,] 1.041563 464.4105 0.001031755
## [4,] 1.408763 455.9107 0.002358674
## [5,] 1.270607 455.5335 0.002257039
## [6,] 1.169098 455.6280 0.001920906
```

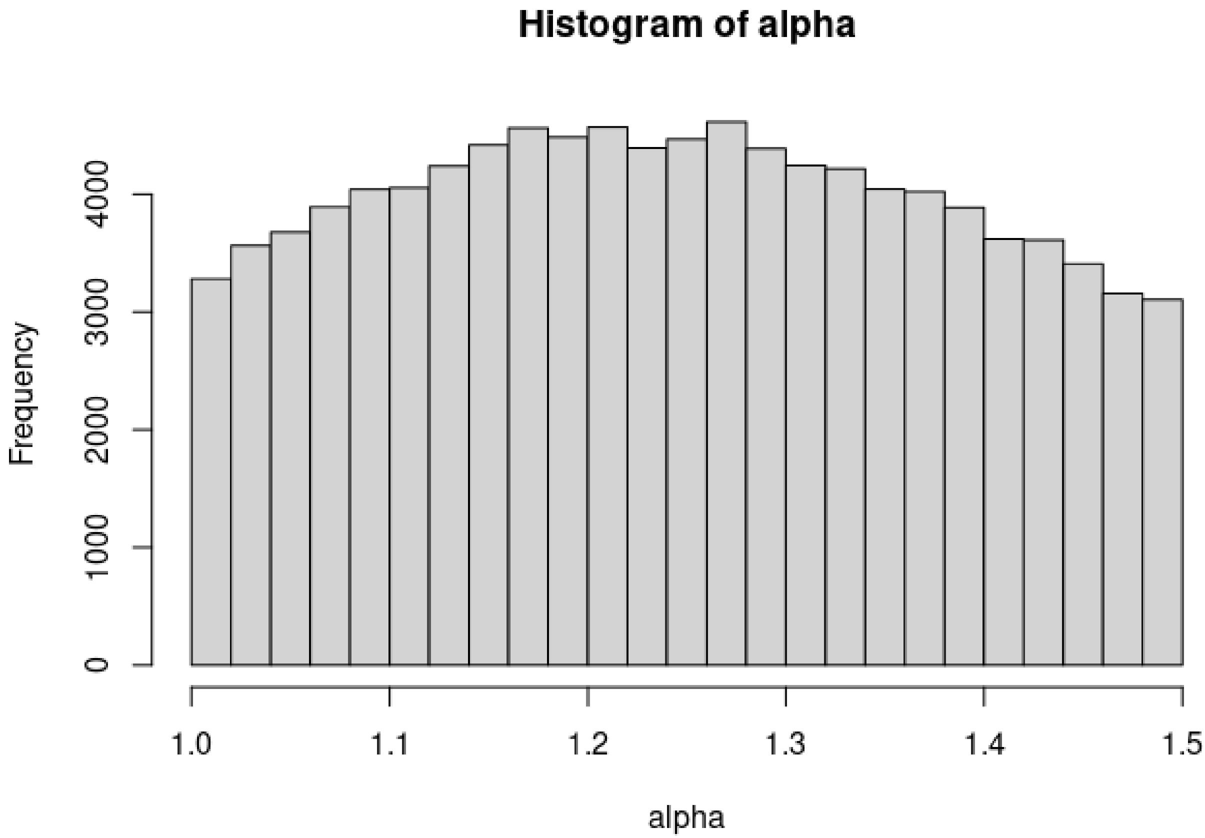
```
sims2 <- as.mcmc(viscos2.sim)
chains2 <- as.matrix(sims2)

#plot(chains2[,2],type="l")

hist(viscos2.sim$BUGSoutput$sims.matrix[,'lambda'])
```



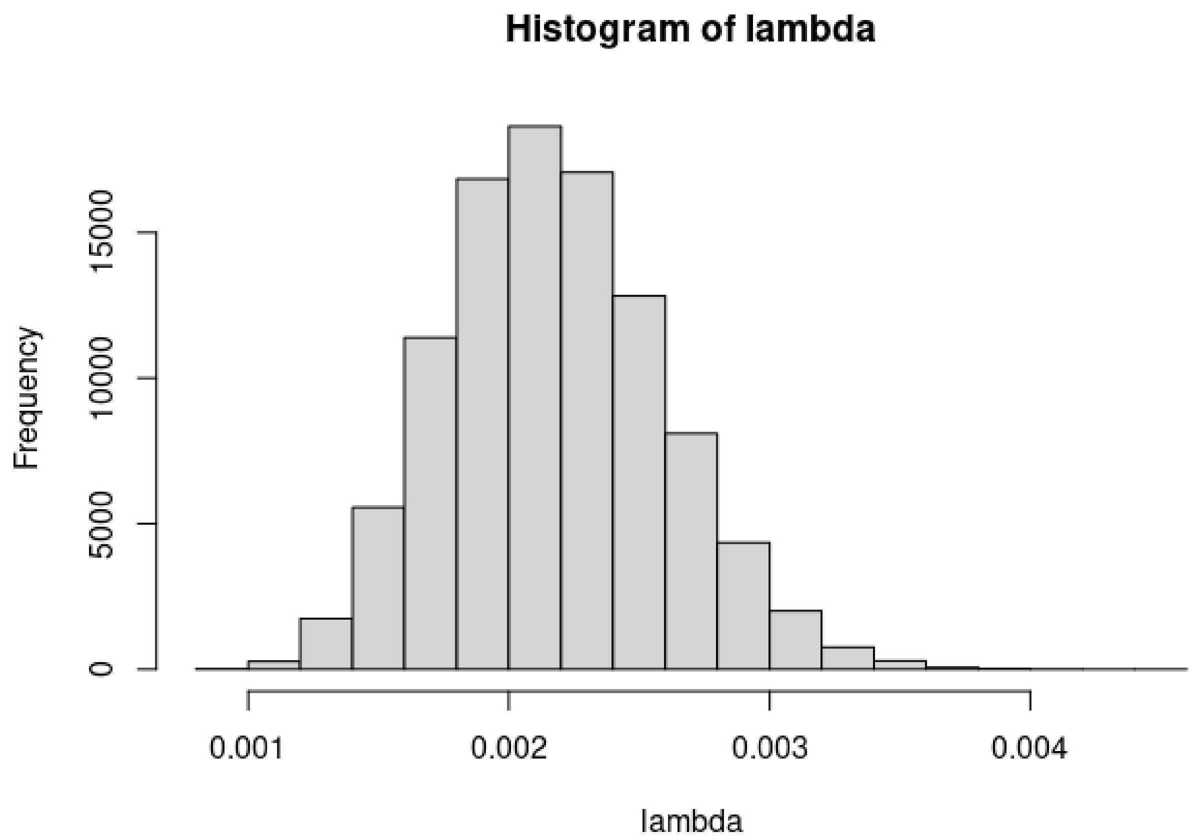

```
lambdas <- viscos2.sim$BUGSoutput$sims.matrix[,'lambda']
acf(viscos2.sim$BUGSoutput$sims.matrix[,'lambda'])
```




```
alpha <- viscos2.sim$BUGSoutput$sims.matrix[,1]
hist(alpha)
```




```
lambda <- viscos2.sim$BUGSoutput$sims.matrix[,3]
hist(lambda)
```



Step 3

-Find 100,000 samples from the posterior predictive distribution (PPD) and show a histogram. Use those samples to obtain an approximate 98% prediction interval for the next observation.

```
# finding quantities
#Mean
mean(chains2[,2])   # true value is 290/477.07 = 0.6078773
```

```
## [1] 456.7088
```

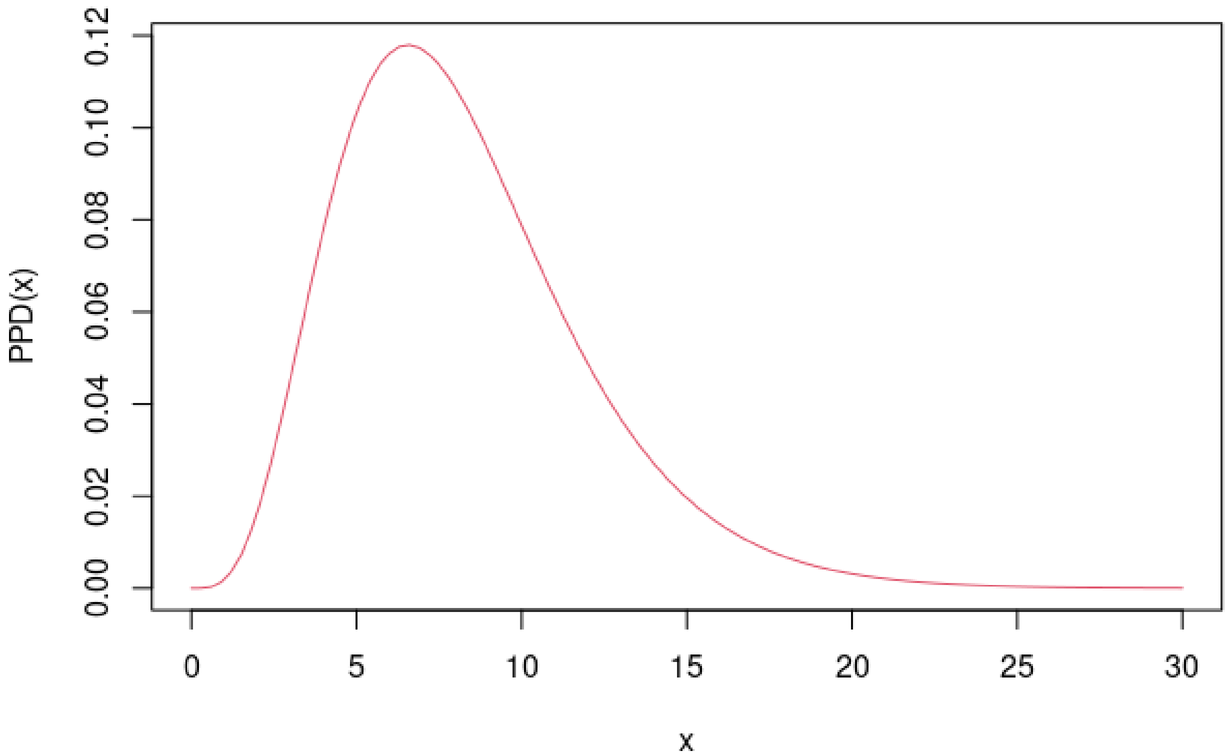
```
#Median
median(chains2[,2])   # true value is qgamma(0.5,290,477.07) = 0.6071787
```

```
## [1] 456.2345
```

```
#95% Credible Interval
quantile(chains2[,2],c(0.025,0.975))   # true value is qgamma(c(0.025,0.975),290,477.
```

```
##     2.5%    97.5%
## 455.5171 460.6528
```

```
#Posterior Predictive

#PPD_samp1 <- rgamma(length(alpha))
PPD_samp <- rgamma(length(chains2[,2]),alpha,chains2[,2])

#rgamma(100000,length(alpha),5,)

#rgamma()
PPD <- function(x) {
  out <- rep(NA,length(x))
  for (i in 1:length(x)) {
    out[i] <- exp(290*log(477.07)+4*log(x[i])-lgamma(290)-lgamma(5)
                  +lgamma(295)-295*log(x[i]+477.07))
  }
  return(out)
}

#hist(PPD_samp,br=50,xlim=c(0,30))
#par(new=T)
curve(PPD(x),xlim=c(0,30),col=2)
```




```r
quantile(PPD_samp,0.01)
```

```
##          1%
## 5.25223e-05
```

```r
quantile(PPD_samp,0.99)
```

```
##        99%
## 0.01128728
```

```r
quantile(PPD_samp,c(0.01,0.99))
```

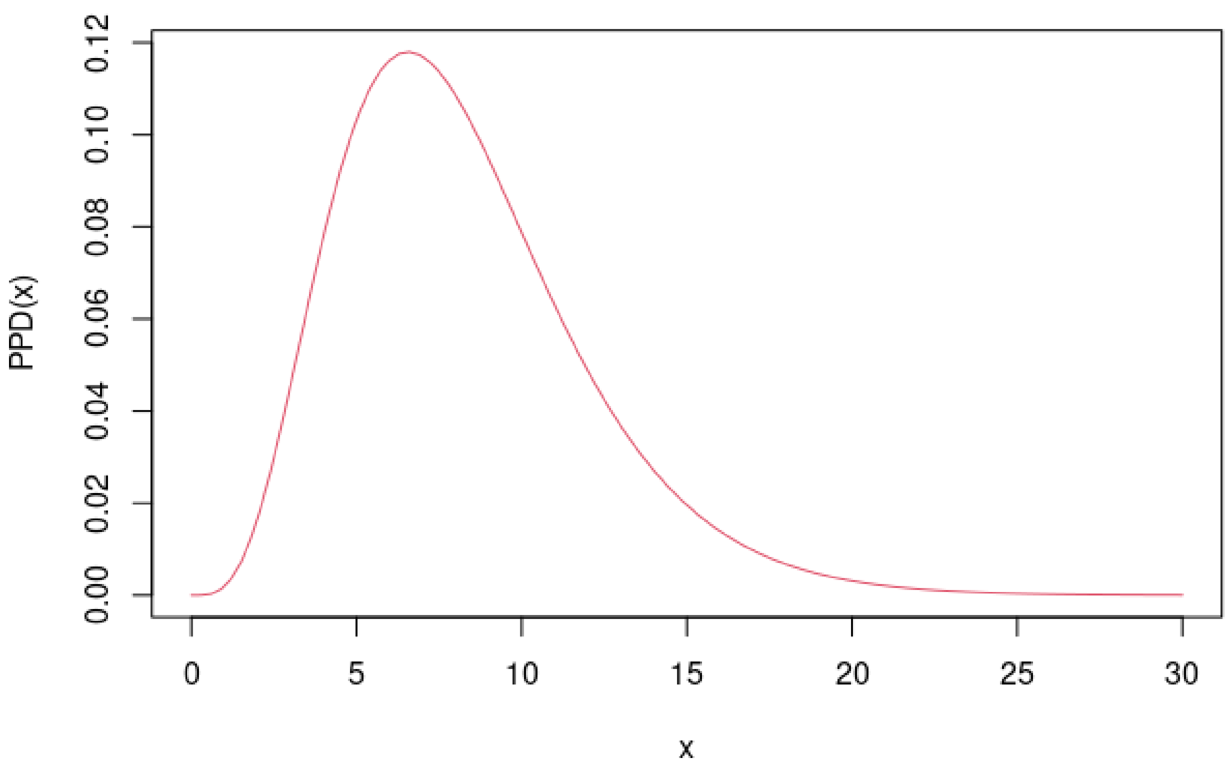
```
##           1%          99%
## 0.0000525223 0.0112872809
```

The 98% prediction interval of the next observation is between 5.3327e-05 and 1.137378e-02.

Step 4

-Determine and comment on how well this model fits the data. Justify the statement with evidence.

```
#hist(dataM,br=50,xlim=c(0,30),main="histogram of PPD",
        #xlab="",
        #ylab="",)
#par(new=T)
curve(PPD(x),xlim=c(0,30),col=2)
```



I plotted the histogram of the posterior predictive distribution and compared it to the histogram of the data and observed that the two histograms look similar with a mean of around 5 - 10. Thus, I concluded that the model fits the data well.

```
mean(PPD_samp)
```

```
## [1] 0.002722783
```

Step 5

-Report a point estimate and a 90% probability interval for the MTTF (mean time to failure).

```
EX <- alpha/lambda
mean(EX)
```

```
## [1] 587.6528
```

```
quantile(EX,c(0.05,0.95))
```

```
##       5%      95%
## 448.8930 761.5229
```

The point estimate is 586.9625 and the 90% probability interval for MTTF is (449.3526,760.2078).

Step 6

-Report a point estimate and a 95% probability interval for the median LCD lamp failure time and explain why there is a difference in the MTTF and the median.

```
medi <- qgamma(.5,alpha,lambda)
```

```
quantile(medi,c(0.025,0.975))
```

```
##     2.5%    97.5%
## 315.2211 602.1807
```

```
median(medi)
```

```
## [1] 431.4208
```

My point estimate for the median LCD failure time is 431.2719. The 95% probability interval is (315.1142,601.0926). The difference between the probability interval of the MTTF and the median is that the interval is wider for the probability interval for the median and that's because we found the 95% probability interval which generally outputs a wider interval than a 90% probability interval. The mean is also affected more by the outliers so that could explain why the mean time to failure has a greater point estimate than the point estimate based on the median.

Step 7

-Write a paragraph to a customer (a non-statistician) about the findings on the lifetime of the projector lamps. Specifically, recommend how often they will need to order new LCD projector lamps if they have 5 projectors (each operating independently for 20 hours per week).

Based on Bayesian analysis, I found the mean of the lamp failure time to be approximately 587 hours. I also observed that there is a 90% probability that the mean time to failure for the lamps lies between 449.3526 and 760.2078 hours given the data and prior knowledge. This means that the average time to failure for the lamps is in the interval of (449.3526,760.20178) with a 90% chance.

I also observed that the median of the lamp failure time is approximately 431.2719 hours with a 95% probability interval for the median LCD lamp failure time of (315.1142,601.0926). In other words, there is a 95% probability that the mean time to failure for the LCD lamps is in between 315.1142 and 760.20178 hours given the data and prior knowledge.

Based on these calculations, I can conclude that with the lamp usage of 20 hours per week, we can expect the lamps to last about 29 weeks or about 6 and a half months before we have to replace them. In other words, we can suggest that the 5 lamps that are operating independently should be replaced every half a year based on the average lamp failure time based on the mean time to failure. If we use the median as the point estimate, we can expect the lamps to last around 22 weeks or 5 months so the lamps should be replaced every 5 months based on the median of lamp failure time.