Assignment 2

Eveline Surbakti - 19200629

3/5/2020

Exercise 1

(a) Write a JAGS misprints.model file and misprints.R script to fit the Poisson-Gamma misprints example of Lecture 4, with x = 3, 4, 2, 1, 2, 3 assumed to each be Poisson distributed with rate lambda and a prior that lambda ~ Gamma(9, 6).

(b) Show that the result is approximately equal to the Gamma(24,12) posterior obtained using conjugacy. Do this by plotting a density estimate of your sampled lambda values together with the Gamma(24,12)

Answer:

The model file is on the separate file and here is the detail for part a and b. The JAGS and theoretical results are approximately equal as the plot of both lines are laying closer to each other with the similar shape.

```
#Data List
misprints <- list(N=6, x = c(3, 4, 2, 1, 2, 3),alpha=9,beta=6)
#JAGS Posterior
jmodel=jags.model(file="misprints.model", data=misprints)

## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 6
##    Unobserved stochastic nodes: 1
##    Total graph size: 10
##
## Initializing model

samps=jags.samples(jmodel,"lambda",n.iter=1e3)
plot(density(samps$lambda),col=3,lwd=2,xlab="x", xlim=c(0,5), ylim=c(0.0,1.2)
)

#Theoretical Posterior
alpha=24
beta=12
curve(dgamma(x, alpha, beta),add=T,col=1,lwd=2,lty=2)
legend("topright", c("JAGS Posterior","Theoretical Posterior"), lty=c(1,2), c
ol=c(3,1), lwd=c (2,2))
```

density.default(x = samps$lambda)
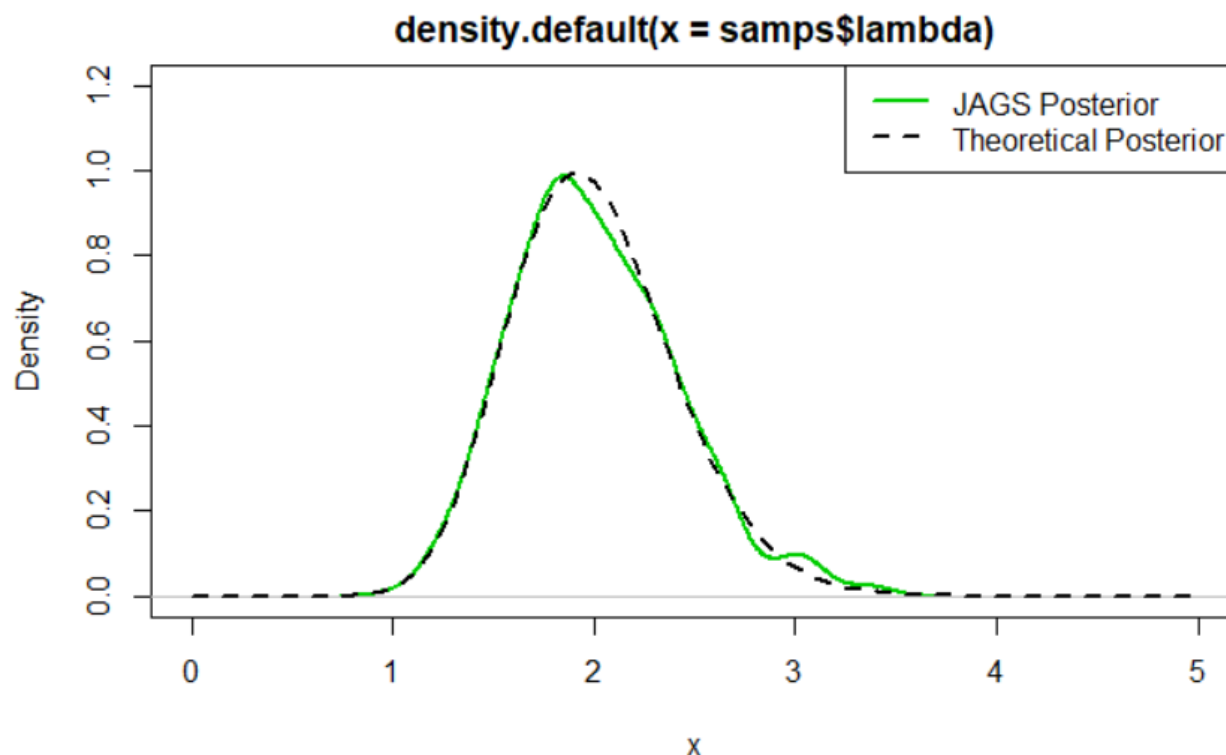
#they are the same because of the conjugacy

Exercise 2

Some more misprints are observed on a further 10 pages as 2, 2, 3, 5, 2, 5, 6, 4, 3, 1.

(a) Use the posterior distribution Gammma(24,12) from Exercise 1 as a prior distribution for this new data and derive the Gamma posterior. Then show that this posterior distribution is the same as that if the data had all been observed simultaneously and a Gamma(9,6) prior is used.

Answer:

In Exercise 1, we have:

Number of data: 6 Sum of data: 15

Prior: Gamma(9,6) then become Posterior: Gamma(9+15,6+6)*

After we observe ten more data, the previous distribution become Prior: Gamma(24,12)* Posterior: Gamma(24+33,12+10) which is Gamma(57,22)**

this posterior distribution is the same as that if the data had all been observed simultaneously and a Gamma(9,6) prior is used such as Prior: Gamma(9,6) Posterior: Gamma(9+48,6+16) which is Gamma(57,22)**

(b) Now use JAGS to sample lambda based on the Gamma(24,12) and these 10 pages of counts.

Answer:

```r
tenmisprints <- list(N=10, x = c(2, 2, 3, 5, 2, 5, 6, 4, 3, 1),alpha=24,beta=
12)

x = c(2, 2, 3, 5, 2, 5, 6, 4, 3, 1)
total1=sum(2, 2, 3, 5, 2, 5, 6, 4, 3, 1)
n1=10

jmodel=jags.model(file="misprints.model", data=tenmisprints)

## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 10
##    Unobserved stochastic nodes: 1
##    Total graph size: 14
##
## Initializing model

samps=jags.samples(jmodel,"lambda",n.iter=1e3)

plot(density(samps$lambda),col=3,lwd=2,xlab="x", xlim=c(0,5),ylim=c(0,1.2))

alpha3=24+total1
beta3=12+n1
#posterior distribution gamma 24,12 which gamma(24+33,12+10) < theoretical po
sterior
curve(dgamma(x, alpha3, beta3),add=T,col=1,lwd=2,lty=2)
legend("topright", c("JAGS Posterior","Theoretical Posterior"), lty=c(1,2), c
ol=c(3,1), lwd=c (2,2))
```
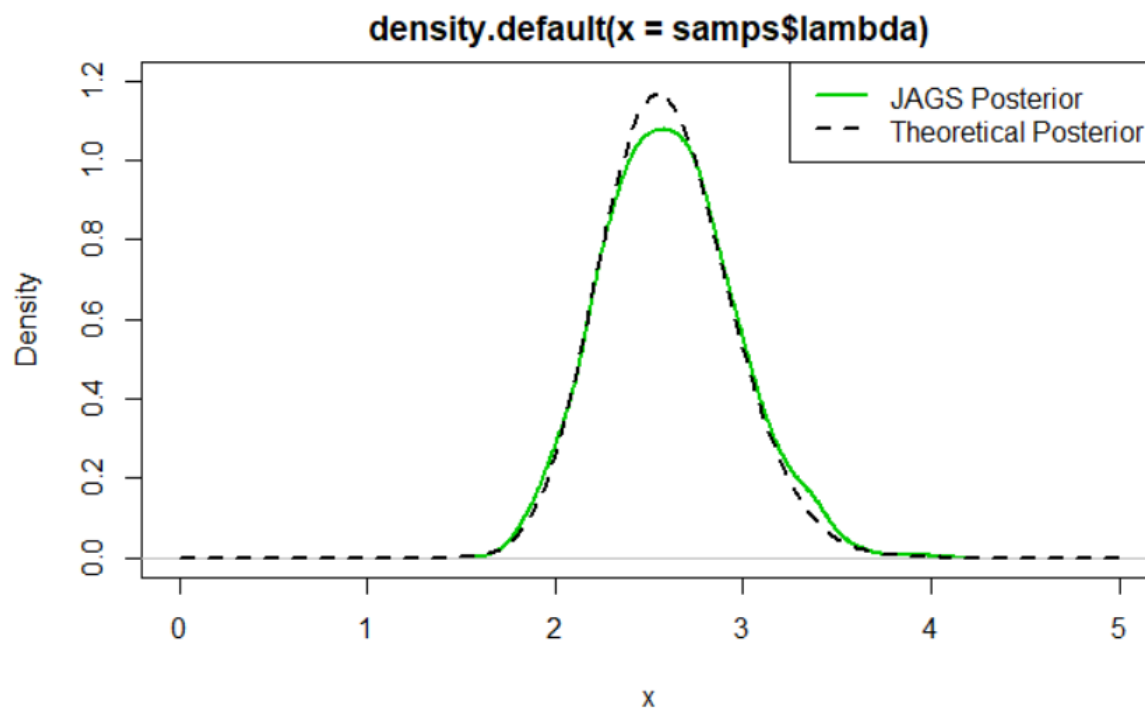
density.default(x = samps$lambda)

Exercise 3

Assess the absolute model fit using the posterior predictive distribution.

Sample an ~x of length 16 using the Poisson likelihood with rate parameter lambda-sample for each sampled lambda-sample. Once you have created 1000 simulated ~x vectors (each of length 16) calculate: (i) the minimum value, (ii) maximum value, (iii) mean, and (iv) standard deviation across the 16 simulated pages. You should then have 1000 of these 4 summary statistics, one for each simulated dataset.

Answer:

```
set.seed(16)
list1 = list() # Make an empty list to save the output
for (i in 1:1000) { # 1000 iterations
  list1[[i]] = rpois(16, lambda = samps$lambda[i]) # Save output in list for
each iteration
}

ccc<-(t(simplify2array(list1))) #transform the list

meansam<- apply(ccc,1,mean)
minsam<- apply(ccc,1,min)
maxsam<- apply(ccc,1,max)
sdsam<- apply(ccc,1,sd)

sixteen<-c(misprints$x,tenmisprints$x) #all observed data
```
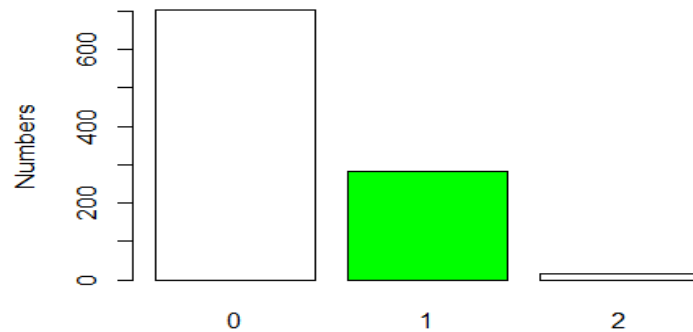
```
#barplot of the minimum
min(sixteen) #minimum of observed data

## [1] 1

colour <- c("white","green","white")
barplot(table(minsam), ylab = "Numbers", col=colour)
```
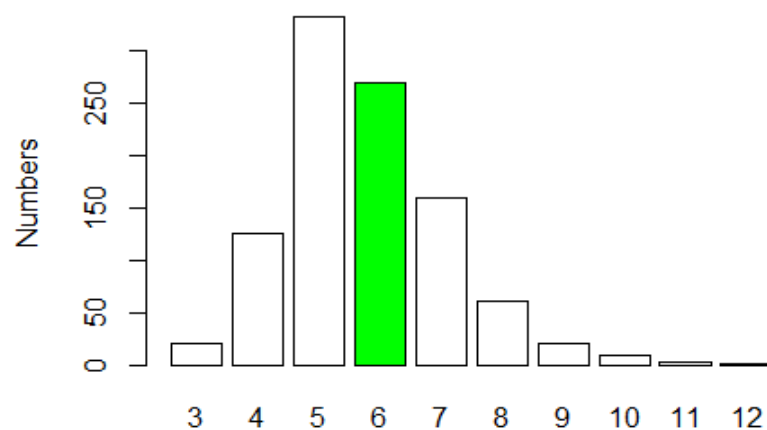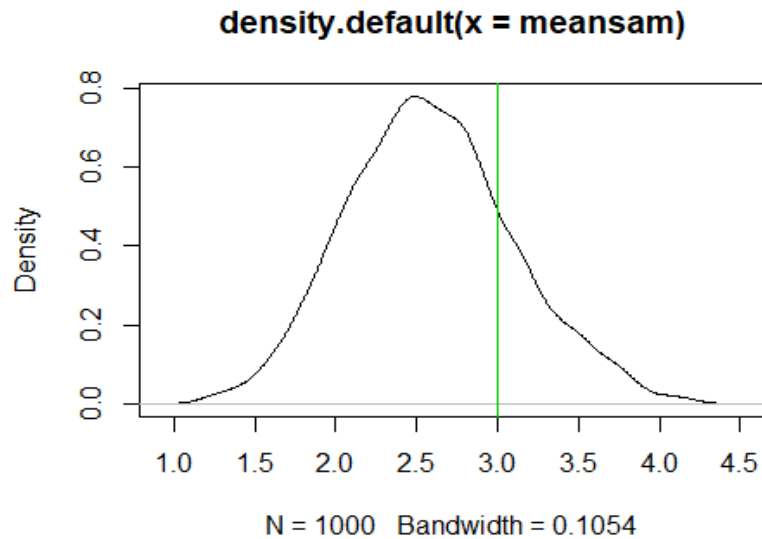


```
#barplot of the maximum
max(sixteen) #maximum of observed data

## [1] 6

colours <- c("white","white","white","green","white","white","white","white")
barplot(table(maxsam), ylab = "Numbers", col=colours)
```
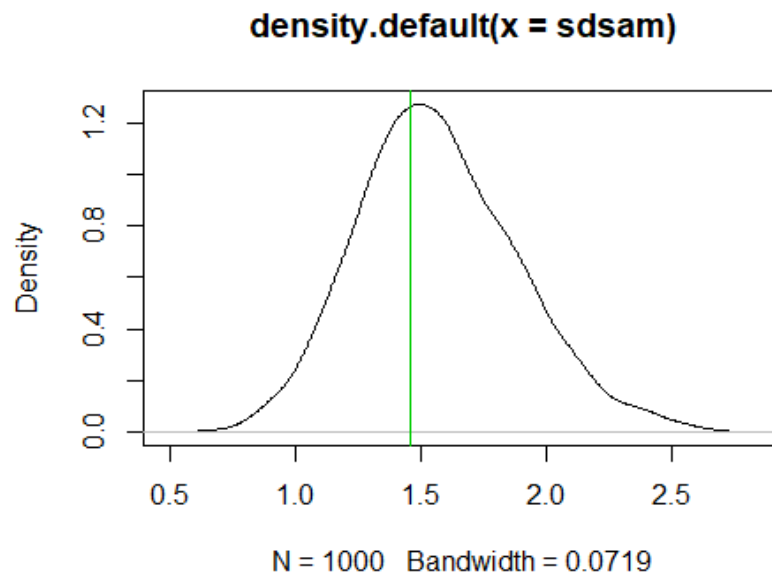
```
#histogram of the mean
plot(density(meansam))
abline(v=mean(sixteen),col=3) #v is the mean of observed data
```

density.default(x = meansam)

N = 1000   Bandwidth = 0.1054

```
#histogram of the standard deviation
plot(density(sdsam))
abline(v=sd(sixteen),col=3) #v is standard deviation of observed data
```

density.default(x = sdsam)

N = 1000   Bandwidth = 0.0719

(b) The posterior predictive approach to assessing model fit compares features of the observed data to the same features of replicate data generated under the model. In this exercise, none of the observed data laying too close to the tails (either barplots or histograms) so there is no evidence of poor fit. Further inspection is probably needed.