Assignment 3 - Question 4

Question 4

The observed breaking strengths (in grams) of 20 pieces of yarn randomly sampled from spinning machines in a certain production area are as follows:

$$46 \quad 58 \quad 40 \quad 47 \quad 47 \quad 53 \quad 43 \quad 48 \quad 50 \quad 55 \quad 49 \quad 50 \quad 52 \quad 56 \quad 49 \quad 54 \quad 51 \quad 50 \quad 52 \quad 50$$

Assume that breaking strengths are known to follow a Normal distribution, and that the mean breaking strength of yarn produced by machines in this area is known to be 51 grams. We are trying to estimate the variance, which we will now call θ.

(a) What is the conjugate family of prior distributions for a Normal variance (not precision) when the mean is known?

For a Gaussian distribution with a known mean but unknown variance we use an Inverse Gamma. So, to obtain a posterior for θ in full we combine $IG(\alpha, \beta)$

$$p(\theta) \propto (\theta)^{-(\alpha+1)} e^{-\beta/\theta},$$

with the likelihood of the Gaussian $\sim N(\bar{y} = 51, \theta)$

$$p(y \mid \theta) \propto (\theta)^{-n/2} e^{\sum_{i=1}^{n}(y_i - \bar{y})^2/2\theta}$$

to get a posterior Inverse Gamma $\sim IG(n + 2\alpha/2, \beta + \sum_{i=1}^{n}(y_i - \bar{y})^2/2)$

$$p(\bar{y}, \theta \mid y) \propto (\theta)^{-\left(\frac{n+2\alpha}{2}+1\right)} e^{-\left(\frac{\beta}{\theta} + \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{2\theta}\right)}$$

(b) Suppose previous experience suggests that the expected value of θ is *12* and the variance of θ is *4*. What parameter values are needed for the prior distribution to match these moments?

Because θ follows an Inverse Gamma such that $p(\theta) \sim IG(\alpha, \beta)$, we have

$$E[\theta] = \frac{\beta}{\alpha - 1}, \qquad \alpha > 1 \quad E[\theta^2] = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)}, \qquad \alpha > 2$$

Assuming the derived values $E[\theta] = 12$ and $E[\theta^2] = 4$, we solve simultaneously to get

$$12 = \frac{\beta}{\alpha - 1}$$
$$\implies \quad \beta = 12(\alpha - 1)$$
$$\implies \quad 4 = \frac{12^2(\alpha - 1)^2}{(\alpha - 1)^2(\alpha - 2)}$$
$$144 = 4(\alpha - 2)$$
$$36 = \alpha - 2$$
$$\implies \quad \alpha = 38$$
$$\implies \quad \beta = 444$$

This gives us a prior $\sim I.G.(38, 444)$.

(c) What is the posterior distribution $p(\theta \mid y)$ for this data under the prior from the previous step?

If we sub $\alpha = 38$ and $\beta = 444$ into the posterior derived in part (a), we get

$$p(\theta \mid y) \propto (\theta)^{-\left(\frac{n+76}{2}+1\right)} e^{-\left(\frac{444}{\theta} + \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{2\theta}\right)}$$

which is $\sim IG\left(\frac{n+76}{2}, 444 + \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{2}\right)$

We can simulate from this using the following code and then plot the resulting density.

```
invgamm<-function(x,a,b){
  lg <- a*log(b) - lgamma(a) - (a+1)*log(x) - b/x
  return(exp(lg))
}

n <- length(strdata)
A <- (n + 76)/2
B <-  444 + ((n-1)*var(strdata))/2
x <- seq(0.01,30,by=0.01)
q4data <- invgamm(x,A,B)
q4data <- as.data.frame(cbind(q4data,x))
names(q4data) <- c("dens","x")
ggplot(q4data, aes(x=x,y=dens)) +
  geom_area() +
  labs(x = "",y = "Density",title = "Inverse Gamma Posterior") +
  theme_grey()
```
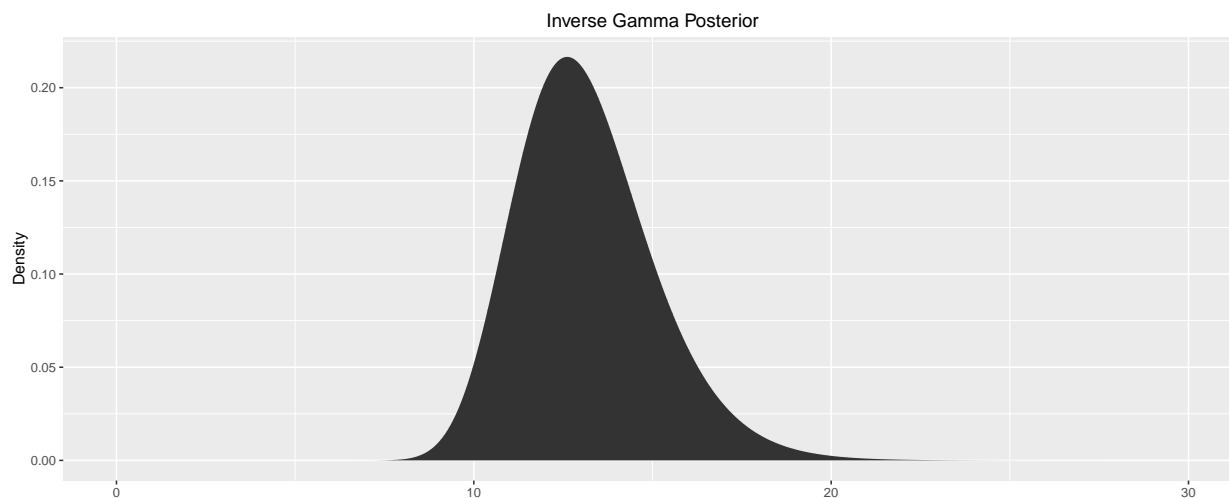

(d) Find the posterior mean and variance of θ.

Using the values for **A** and **B** we obtained earlier, we can simulate using *rgamma()*, take the inverse and then calculate the required statistics.

```
g <- rgamma(10000,A,B)
CI <- quantile(1/g,c(0.05,0.95))
# Confidence Intervale for theta
```

```
CI
##        5%       95%
## 10.27494 16.71264
# Posterior Mean for theta
mean(1/g)
## [1] 13.17787
# Posterior Variance for theta
var(1/g)
## [1] 3.839052
```

(e) Now use WinBUGS to carry out the analysis. You will have to specify the model in terms of the precision.

Here we will use $\mu = 51$ (which was given) and $\alpha = 38; \beta = 444$ (which we calculated) to calculate the posterior mean and variance of θ in WinBUGS.

```
strdata <- c(46,58,40,47,47,53,43,48,50,55,49,50,52,56,49,54,51,50,52,50)
#define the model
strmodel <- function() {
  for (i in 1:n) {x[i] ~ dnorm(mu,tau) }  # likelihood with unkown tau, known mu
  tau ~ dgamma(alph,bet)                  # prior with alpha and beta calculated
  theta <- 1/tau                          # variance (expressed as inverse of precision)
}
# write the model code out to a file
write.model(strmodel, "strmodel.txt")
model.file1 = paste(getwd(),"strmodel.txt", sep="/")
coda.file = paste(getwd(),"Q4Coda", sep="/")
file.show("strmodel.txt")

#prepare the data for input into OpenBUGS
x <- strdata
mu <- 51
alph <- 38
bet <- 444
n <- 20
data <- list("x","n","mu","alph","bet")

#initialization of variables
inits <- function(){
  list(tau=0)
}

WINE="/opt/local/bin/wine"
WINEPATH="/opt/local/bin/winepath"
OpenBUGS.pgm=paste0("/Users/benjamin/Applications/wine/",
                    "drive_c/ProgramFiles/OpenBUGS/OpenBUGS323/OpenBUGS.exe")

#these are the parameters to save
parameters = c("theta")

#run the model
str.sim <- bugs(data,
                inits,
                model.file = model.file1,
```
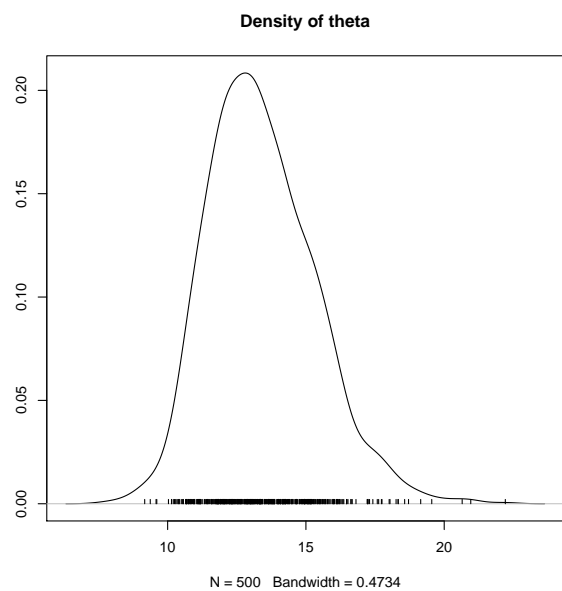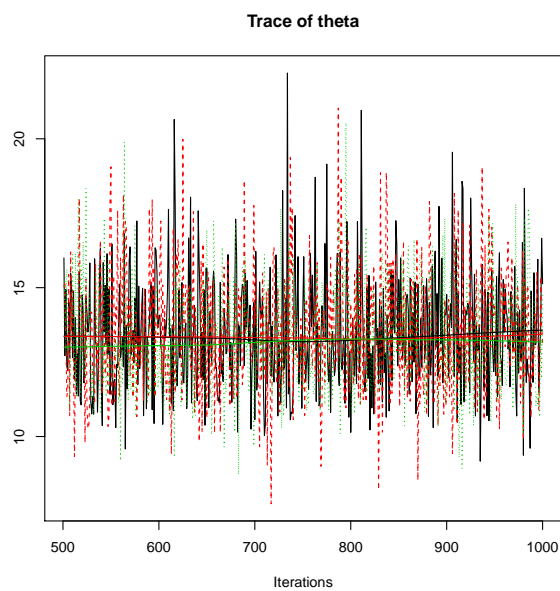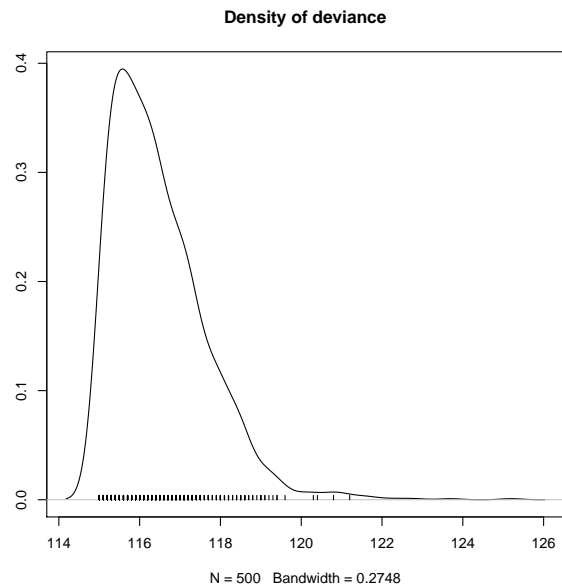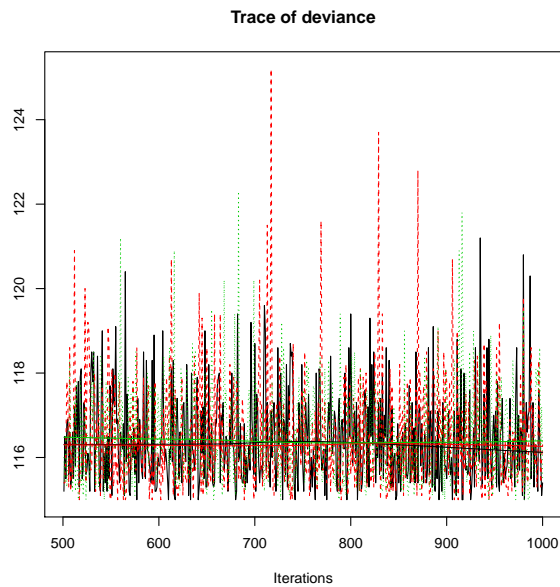
```
                parameters=parameters,
                n.chains = 3,
                n.iter = 1000,
                OpenBUGS.pgm=OpenBUGS.pgm,
                WINE=WINE,
                WINEPATH=WINEPATH,
                useWINE=T,
                codaPkg = T,
                working.directory = coda.file,
                debug = F)
```

```
samples <- read.openbugs(paste0(coda.file,"/"))
mutheta <- mean(unlist(samples[,2]))
sdtheta <- sd(unlist(samples[,2]))
vartheta <- var(unlist(samples[,2]))
```

```
plot(samples)
```

Trace of deviance

Density of deviance

N = 500   Bandwidth = 0.2748

Trace of theta

Density of theta

N = 500   Bandwidth = 0.4734

```
summary(samples)
##
## Iterations = 501:1000
## Thinning interval = 1
## Number of chains = 3
## Sample size per chain = 500
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##            Mean    SD Naive SE Time-series SE
## deviance 116.5 1.185  0.03059        0.02988
## theta     13.4 1.928  0.04978        0.04848
##
```

```
## 2. Quantiles for each variable:
##
##            2.5%    25%   50%    75%  97.5%
## deviance 115.00 115.60 116.2 117.10 119.20
## theta     10.27  12.03  13.2  14.64  17.67
```

From the results we can read that the expected value of θ is given as 13.3983447; the standard deviation is given as 1.9281636; and the variance is therefore 3.7178149. These come pretty close to the answers we obtained by via the other estimation method in part (c).

End of Assignment 3