Lab4Q2-report

Jaskirat S Marar

11/25/2021

Assignment2

We are given the following:

$$n = number\ of\ observations$$

$$\vec{\mu} = (\mu_1, ..., \mu_n)\ are\ unknown\ parameters$$

$$Y_i \sim N(\mu_i, variance = 0.2),\ i = 1, ..., n$$

$$Prior:\ p(\mu_1) = 1\ ;\ p(\mu_{i+1}|\mu_i) = N(\mu_i, 0.2),\ i = 1, ..., (n-1)$$

We'll be interested in deriving the posterior i.e. P(μ|Y) using the Bayes' theorem as follows:

$$posterior = likelihood\ *\ prior$$
$$P(\vec{\mu}|Y) = \frac{P(Y|\vec{\mu}) * P(\vec{\mu})}{\int_\mu P(Y|\vec{\mu})P(\vec{\mu})d\mu}$$

Since the denominator is constant w.r.t μ, we can drop it in favour of proportionality

$$posterior \propto likelihood\ *\ prior$$
$$P(\vec{\mu}|Y) \propto P(Y|\vec{\mu}) * P(\vec{\mu})$$

First we define the likelihood:

$$L(Y_i|\mu_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp\Big(-\frac{1}{2\sigma^2}(Y_i - \mu_i)^2\Big)$$

Since, we are only interested in the terms dependent on the parameter μ we can drop the term at the start of the expression and the likelihood then becomes:

$$L(Y_i|\mu_i) \propto exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \mu_i)^2\Big)$$

Similarly, now for the prior using chain rule,

$$P(\vec{\mu}) = P(\mu_1).P(\mu_2|\mu_1).P(\mu_3|\mu_2)...P(\mu_n|\mu_{n-1}) \qquad ...(a)$$

$$= \prod_{i=1}^{n-1} P(\mu_{i+1}|\mu_i) \ \sim \ N(\mu_i, \sigma^2_{\mu_i})$$

$$= \prod_{i=1}^{n-1} \frac{1}{\sqrt{2\pi\sigma^2}} exp\Big(- \frac{1}{2\sigma^2_{\mu}}(\mu_{i+1} - \mu_i)^2\Big)$$

$$\propto exp\Big(- \frac{1}{2\sigma^2_{\mu}} \sum_{i=1}^{n-1}(\mu_{i+1} - \mu_i)^2\Big)$$

Hence, we can now write our posterior as the following:

$$P(\vec{\mu}|Y) \propto exp\Big(- \frac{1}{2\sigma^2} \sum_{i=1}^{n}(Y_i - \mu_i)^2\Big) \ * \ exp\Big(- \frac{1}{2\sigma^2_{\mu}} \sum_{i=1}^{n-1}(\mu_{i+1} - \mu_i)^2\Big)$$

$$\propto exp\Big[- \frac{1}{2\sigma^2} \sum_{i=1}^{n}(Y_i - \mu_i)^2 - \frac{1}{2\sigma^2_{\mu}} \sum_{i=1}^{n-1}(\mu_{i+1} - \mu_i)^2\Big] \quad ...(b)$$

Since, σ is given to be same everywhere in the problem we can use common notation across the expression.

Moving to the next part to find conditional probability $(\mu_k|\mu_{-k})$, we know that by the definition of conditional probability, the joint probability of the prior, P($\vec{\mu}$), can be expanded into the conditional components of each individual μ_k and can be expressed as

$$P(\vec{\mu}) \propto P(\mu_k|\vec{\mu}) * P(\mu_1, ..., \mu_{k-1}, \mu_{k+1}, ..., \mu_n)$$

Or in a more condensed form as,

$$P(\mu_i|\vec{\mu}_{-i}, \vec{Y}) = \frac{P(\vec{\mu}|\vec{Y})}{P(\vec{\mu}_{-i}|\vec{Y})} \qquad ...(c)$$

$$= \frac{P(\vec{\mu}|\vec{Y})}{\int P(\vec{\mu}|\vec{Y})d\mu}$$

We now solve for μ_1

$$P(\mu_1|\vec{\mu}_{-1}, \vec{Y}) = \frac{P(\vec{\mu}, \vec{Y})}{\int P(\vec{\mu}, \vec{Y})d\mu_1}$$

$$= \frac{\prod_{i=2}^{n} P(Y_i|\mu_i) \prod_{i=3}^{n} P(\mu_i|\mu_{i-1}) P(\mu_2|\mu_1) P(Y_1|\mu_1) P(\mu_1)}{\prod_{i=2}^{n} P(Y_i|\mu_i) \prod_{i=3}^{n} P(\mu_i|\mu_{i-1}) \int P(\mu_2|\mu_1) P(Y_1|\mu_1) P(\mu_1) d\mu_1}$$

Common product terms cancel out in numerator and denominator, and the integral is constant w.r.t. μ_1

$$P(\mu_1|\vec{\mu}_{-1}, \vec{Y}) \propto P(\mu_2|\mu_1) P(Y_1|\mu_1) P(\mu_1)$$

$$\propto exp\Big(- \frac{1}{2\sigma^2}(\mu_2 - \mu_1)^2\Big) \ * \ exp\Big(- \frac{1}{2\sigma^2}(Y_1 - \mu_1)^2\Big) \ * \ 1$$

$$\propto exp\Big(- \Big(\frac{1}{2\sigma^2}\Big)\Big((\mu_2 - \mu_1)^2 + (Y_1 - \mu_1)^2\Big)\Big)$$

using *Hint B*

$$P(\mu_1|\vec{\mu}_{-1}, \vec{Y}) \propto exp\Big(- \frac{(\mu_1 - (Y_1 + \mu_2)/2)^2}{\sigma^2}\Big) \ \sim \ N\Big((\frac{Y_1 + \mu_2}{2}), \frac{\sigma^2}{2}\Big)$$

Similarly, we now check the case for μ_n. Here we use the posterior we defined in (b) and the relationship we established in (c)

$$P(\mu_n|\vec{\mu}_{-n},\vec{Y}) = \frac{P(\vec{\mu}|\vec{Y})}{P(\vec{\mu}_{-n}|\vec{Y})}$$
$$\propto \frac{exp\Big(-\sum_{i=1}^{n-1}\frac{(\mu_{i+1}-\mu_i)^2}{2\sigma^2}-\sum_{i=1}^{n}\frac{(Y_i-\mu_i)^2}{2\sigma^2}\Big)}{exp\Big(-\sum_{i=1}^{n-2}\frac{(\mu_{i+1}-\mu_i)^2}{2\sigma^2}-\sum_{i=1}^{n-1}\frac{(Y_i-\mu_i)^2}{2\sigma^2}\Big)}$$
$$\propto exp\Big(-\Big(\frac{1}{2\sigma^2}\Big)\Big((\mu_n-\mu_{n-1})^2+(Y_n-\mu_n)^2\Big)\Big)$$

again using *hint B*

$$P(\mu_n|\vec{\mu}_{-n},\vec{Y}) \propto exp\Big(-\frac{(\mu_n-(Y_n+\mu_{n-1})/2)^2}{\sigma^2}\Big) \ \sim\ N\Big((\frac{Y_n+\mu_{n-1}}{2}),\frac{\sigma^2}{2}\Big)$$

Now for the tedious case of μ_i where is i $\notin$ (1,n)

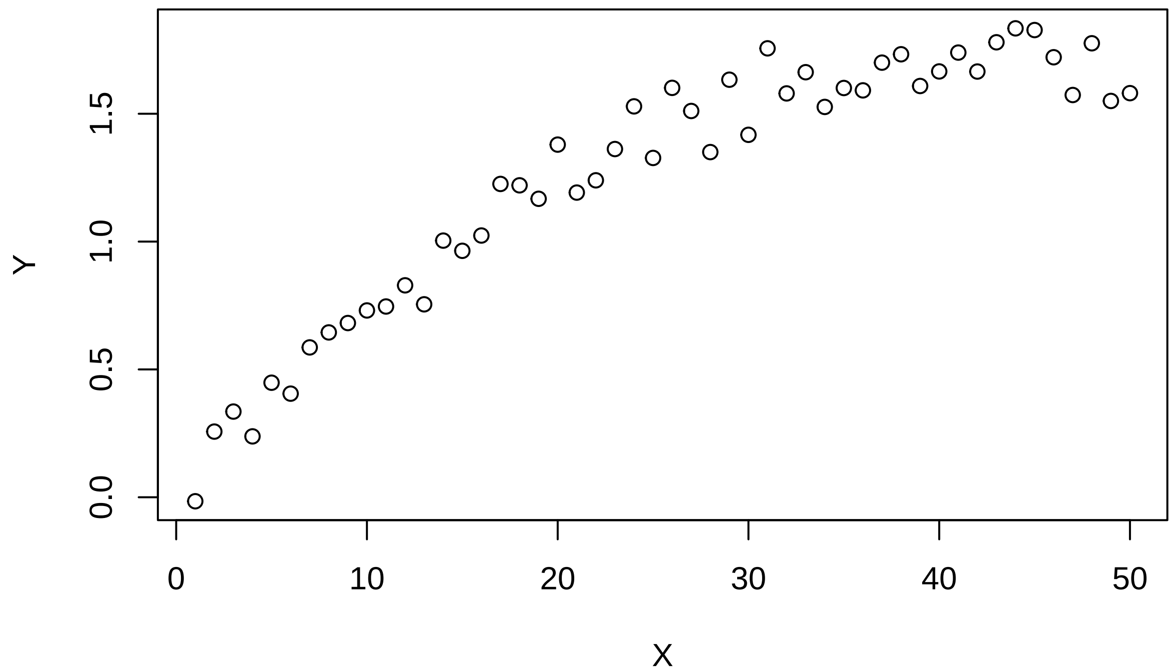
$$P(\mu_i|\vec{\mu}_{-i},\vec{Y}) \propto P(\mu_{i+1}|\mu_i)P(Y_i|\mu_i)P(\mu_i|\mu_{i-1}) \propto exp\Big[-\Big(\frac{1}{2\sigma^2}\Big)\Big((\mu_{i+1}-\mu_i)^2+(Y_i-\mu_i)^2+(\mu_i-\mu_{i-1})^2\Big)\Big]$$

Using *hint C*

$$P(\mu_i|\vec{\mu}_{-i},\vec{Y}) \propto exp\Big[-\frac{(\mu_i-(Y_i+\mu_{i-1}+\mu_{i+1})/3)^2}{2\sigma^2/3}\Big] \ \sim\ N\Big((\frac{Y_i+\mu_{i-1}+\mu_{i+1}}{3}),\frac{\sigma^2}{3}\Big)$$

We now use these derivations to design a Gibbs sampler which will estimate $\vec{\mu}$ for 1000 values of μ

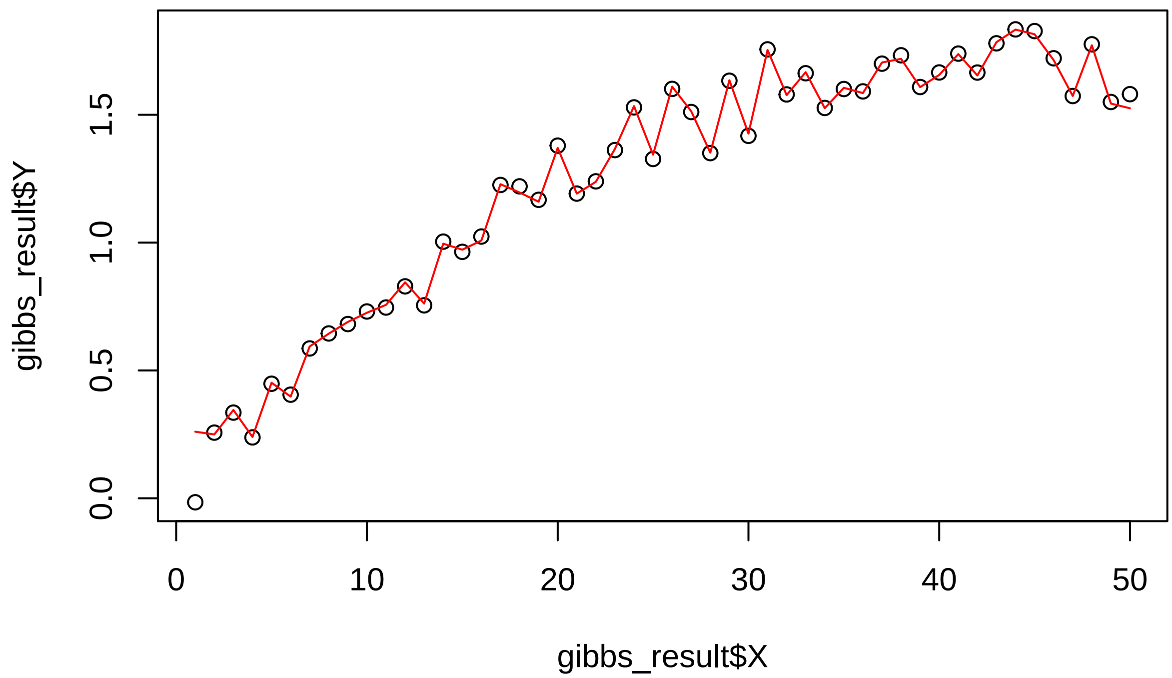
```
#load the data
load("chemical.RData")
#converting data into a DF
chemical_df <- data.frame(X, Y)
#initial plot of data to understand distribution
plot(X, Y)
```




```r
#3 inputs to sampler: # of time steps, Y to be estimated, known sigma
mc_gibbs <- function(step, Y, sig) {
  mu_i <- 0
  #initialize matrix to store results of sampler in each iteration
  result_matrix <- matrix(0, nrow = 50, ncol = step)
  for (tstep in 1:(step-1)) {
    for (i in 1:50) {
      #Case for mu_1
      if (i==1) {
        mu_i <- mean(result_matrix[i+1,tstep],Y[i])
        result_matrix[i,tstep+1] <- rnorm(1, mean = mu_i, sd = sig/sqrt(2))
      }
      #Case for mu_n
      else if (i == 50) {
        mu_i <- mean(result_matrix[i-1,tstep+1],Y[i])
        result_matrix[i,tstep+1] <- rnorm(1, mean = mu_i, sd = sig/sqrt(2))
      }
      #Case for mu_i
      else {
        mu_i <- mean(Y[i],result_matrix[i-1,tstep+1], result_matrix[i+1,tstep])
        result_matrix[i,tstep+1] <- rnorm(1, mean = mu_i, sd = sig/sqrt(3))
      }
    }
  }
  return(result_matrix)
}
```

Plotting the results of the sampler:

```
#function call with 1000 steps and variance 0.2
gibbs_output <- mc_gibbs(1000, chemical_df$Y, sqrt(0.2))
#append expected values of mu to original DF
gibbs_result <- cbind(chemical_df, rowMeans(gibbs_output))
names(gibbs_result) <- c("X", "Y", "expected_mu")
#Plot original points
plot(gibbs_result$X, gibbs_result$Y)
lines(gibbs_result$expected_mu, col = "red")
```

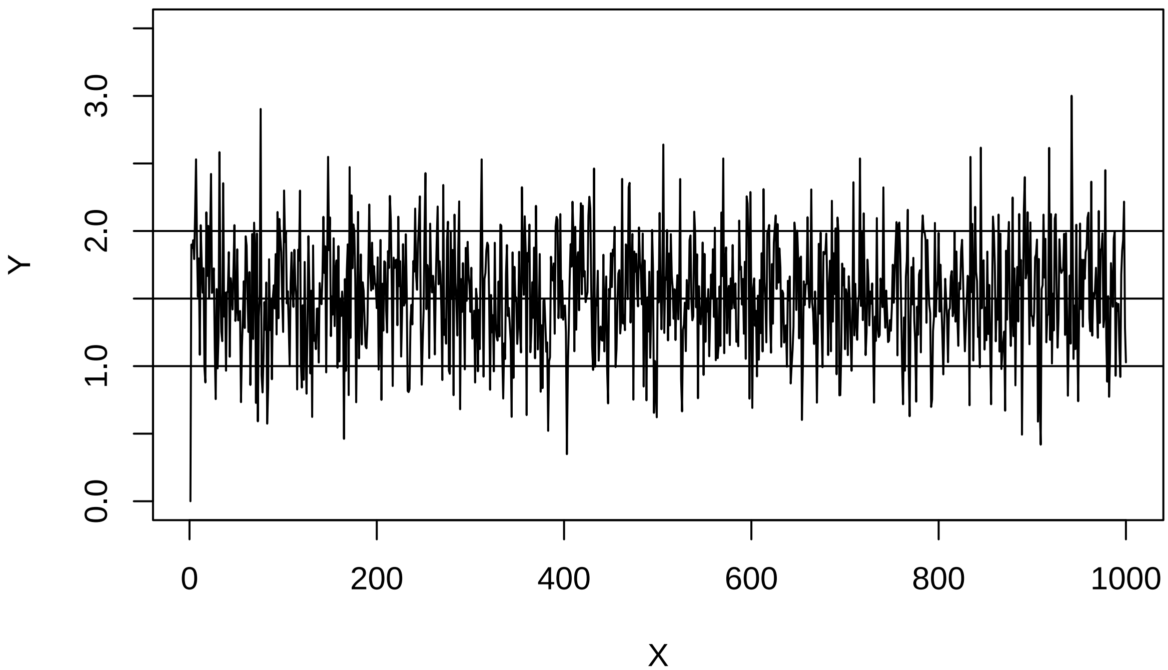



```
#Expected values as a line
```

The result shows that our expected values have followed a similar result as the original data. Implying that not much improvement has happened in terms of reducing the measurement noise of the concentration. But the noise reduction at the start and end is quite pronounced, especially at the start, thus, suggesting that our burning period may have been very short.

Trace Plot:

```
vN <- 1:ncol(gibbs_output)
vX <- gibbs_output
plot(vN,vX[50,],pch=19,cex=0.3,col="black",xlab="X",ylab="Y", ylim=c(0,3.5), type = 'l')
abline(h=1)
abline(h=2)
abline(h=1.5)
```



Our above conclusion is supported by the trace plot, where we see that the chain doesnt converge and keeps producing noisy values, but the burning period is evidently quite small. But this can also be because the data set that we are working with is quite small (only 50 observations). To expect a good reduction in noise, we would ideally want a larger data set.