Assignment 9

Jake Whalen

April 18, 2018

Table of Contents

[Problem 1 1](#_Toc512162205)

[Problem 2 3](#_Toc512162206)

[Problem 3 5](#_Toc512162207)

#Packages  
library(tidyverse)  
library(magrittr)  
library(coda)

# Problem 1

This assignment uses the reaction time data from Assignment 8, taken from the Gelman et al. reference and found at this url: <http://www.stat.columbia.edu/~gelman/book/data/schiz.asc>.

As with Assignment 8, we will consider only the non-schizophrenic subjects (rows 1 to 11).

#Data URL  
www <- "http://www.stat.columbia.edu/~gelman/book/data/schiz.asc"  
#Read and Wrangle  
nschiz <- read.table(www, skip = 5) %>%   
 mutate(condition = rep(c("non-schizophrenics", "schizophrenics"), c(11, 6))) %>%  
 filter(condition == 'non-schizophrenics') %>%   
 select(-condition) %>%  
 mutate(subject = sprintf("p%s", 1:11)) %>%  
 gather(reaction, ms, -subject) %>%  
 mutate(lms = log(ms))

* As with Assignment 8, assume the logarithms of the response times are independent and identically distributed normal random variables with person-specific mean .
* Following Gelman, et al., assume all the observations have the same standard deviation .
* The 11 means are independent and identically distributed normal random variables with mean and standard deviation .
* The unknown mean has a normal distribution with mean 5.52 and standard deviation 0.22. This reflects a prior 95% credible interval of [162, 385] ms for the population average reaction time, which is consistent with the literature on reaction times.
* The inverse variance has a gamma distribution with shape 0.5 and scale 50. This reflects weak prior information focused on a value of 25 for , or 0.2 for .
* The inverse variance has a gamma distribution with shape 0.5 and scale 50. This reflects weak prior information focused on a value of 25 for , or 0.2 for .

Note that this model is similar to the math test scores example in Unit 7. Using the formulas in the Unit 7 notes as a model, find the posterior distribution for each of the unknown parameters given the other parameters.

* Given (), is normally distributed with:
* Given (), is distributed with a gamma distribution with:
* Given (), is distributed with a gamma distribution with:
* Given () each is normally distributed with:

# Problem 2

Use Gibbs sampling to draw 5000 samples from the posterior distribution of the parameters , , , and . Find posterior credible intervals for each of these parameters.

init.nschiz <-   
 nschiz %>%  
 group\_by(subject) %>%  
 summarise(yBar = mean(lms),  
 ySSQ = sum(lms^2),  
 nSub = n(),  
 ySD = sd(lms))  
  
nSub <- length(unique(nschiz$subject))  
nSubL <- init.nschiz$nSub  
ySSQ <- init.nschiz$ySSQ  
yBar <- init.nschiz$yBar  
ySD <- init.nschiz$ySD

Set the prior hyperparameters:

mu.mean.0 = 5.52  
mu.SD.0 = 0.22  
tau.sh.0 = sig.sh.0 = 1/2  
tau.sc.0 = sig.sc.0 = 50

Initialize Gibbs sample variables:

numSim = 5000  
thetaG = matrix(nrow = numSim, ncol = nSub)  
thetaG[1,] = yBar # initial value for theta  
muG = mean(yBar) # initial value for mu  
tauG = sd(yBar) # initial value for tau  
sigG = mean(ySD) # initial value for sigma

Run Gibbs sampler:

for (k in 2:numSim) {  
 mu.mean.1 = (mu.mean.0/mu.SD.0^2+sum(thetaG[k-1,])/tauG[k-1]^2)/(1/mu.SD.0^2+nSub/tauG[k-1]^2)  
 mu.SD.1 = (1/mu.SD.0^2+nSub/tauG[k-1]^2)^-(1/2)  
 #Sample new mu  
 muG[k] = rnorm(1,mu.mean.1, mu.SD.1)  
   
 tau.sh.1 = tau.sh.0 + 0.5\*nSub # this is the same every iteration  
 tau.sc.1 = (1/tau.sc.0 + 0.5\*sum((thetaG[k-1,]-muG[k])^2))^-1  
 #Sample new tau  
 tauG[k] = rgamma(1,shape=tau.sh.1,scale=tau.sc.1)^(-1/2)  
   
 sig.sh.1 = sig.sh.0 + 0.5\*sum(nSubL) # this is the same every iteration  
 sig.sc.1 = (1/sig.sc.0 + 0.5\*sum(ySSQ - nSubL\*yBar^2 + nSubL\*(yBar - thetaG[k-1,])^2))^-1  
 #Sample new sigma  
 sigG[k] = rgamma(1,shape=sig.sh.1,scale=sig.sc.1)^(-1/2)  
   
 mu.1 = (muG[k]/tauG[k]^2+yBar\*nSubL/sigG[k]^2)/(1/tauG[k]^2+nSubL/sigG[k]^2)  
 tau.1 = (1/tauG[k]^2+nSubL/sigG[k]^2)^(-1/2) # vector of posterior SDs  
 #Sample new theta  
 thetaG[k,] = rnorm(nSub,mu.1,tau.1)  
}

95% CI on Parameters ()

|  |  |  |
| --- | --- | --- |
| Parameter | Lower | Upper |
| Mu | 5.6274449 | 5.7932086 |
| Tau | 0.0857699 | 0.2105276 |
| Sigma | 0.1484364 | 0.1733463 |

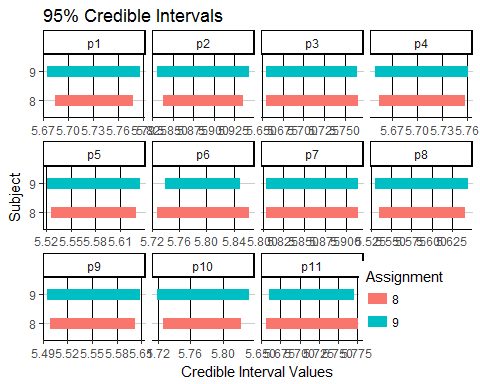
95% CI for person-specific

theta.hat = colMeans(thetaG)  
th.lower=th.upper=NULL  
for (i in 1:nSub) {  
 th.lower[i] = quantile(thetaG[,i],0.025)  
 th.upper[i] = quantile(thetaG[,i],0.975)  
}

|  |  |  |  |
| --- | --- | --- | --- |
| Subject | Lower | Upper | Mean |
| p1 | 5.675378 | 5.786607 | 5.731062 |
| p2 | 5.829983 | 5.943383 | 5.886129 |
| p3 | 5.654438 | 5.765291 | 5.710458 |
| p4 | 5.649328 | 5.760658 | 5.705381 |
| p5 | 5.521701 | 5.634008 | 5.577798 |
| p6 | 5.738778 | 5.849968 | 5.793891 |
| p7 | 5.803857 | 5.915108 | 5.858491 |
| p8 | 5.530267 | 5.644245 | 5.587792 |
| p9 | 5.495595 | 5.608370 | 5.550455 |
| p10 | 5.718368 | 5.832081 | 5.774531 |
| p11 | 5.660067 | 5.771438 | 5.716454 |

# Problem 3

Discuss your results. Compare your intervals for with the results from Assignment 8. The lines in the chart below represent the person-specific credible intervals for from assignment 8 (pink) and 9 (blue).



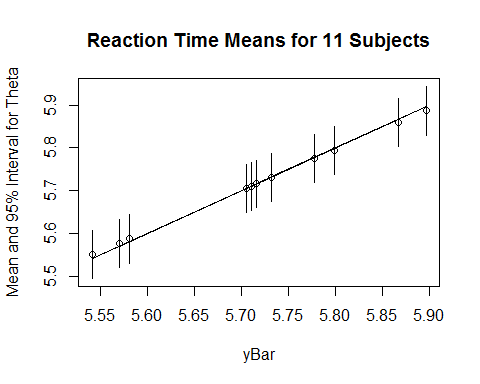
Most of the CIs from assignment 8 are narrower then those from assignment 9 but not by much. All of the CIs look to be centered around the same value between both problems.

Since all eleven standard deviations for the log reaction times are equal the widths of the CIs for each are roughly the same in problem 9. The table below shows the width of the CIs from 8 and 9 for each subject.

|  |  |  |
| --- | --- | --- |
| subject | 8 | 9 |
| p1 | 0.0945745 | 0.1112297 |
| p2 | 0.0979664 | 0.1133999 |
| p3 | 0.1116778 | 0.1108530 |
| p4 | 0.1034933 | 0.1113302 |
| p5 | 0.1032364 | 0.1123070 |
| p6 | 0.1363198 | 0.1111908 |
| p7 | 0.1113367 | 0.1112513 |
| p8 | 0.1063485 | 0.1139784 |
| p9 | 0.1041249 | 0.1127749 |
| p10 | 0.0953442 | 0.1137126 |
| p11 | 0.1211357 | 0.1113712 |

In assignment 9 all of the differences are close to the hundreths place (0.11). In assignment 8 the CI widths vary from subject to subject.

The plot below shows the shrinkage for the 11 subjects values.



Since the slope of the line is slightly less then 1 we can tell that the lower values for correspond to slightly higher values for and higher values for correspond to slightly lower values for (Hoff, 140).

Effective sample sizes for parameters :

|  |  |  |  |
| --- | --- | --- | --- |
| numSim | Mu | Tau | Sigma |
| 5000 | 4877.03 | 3673.13 | 4583.45 |

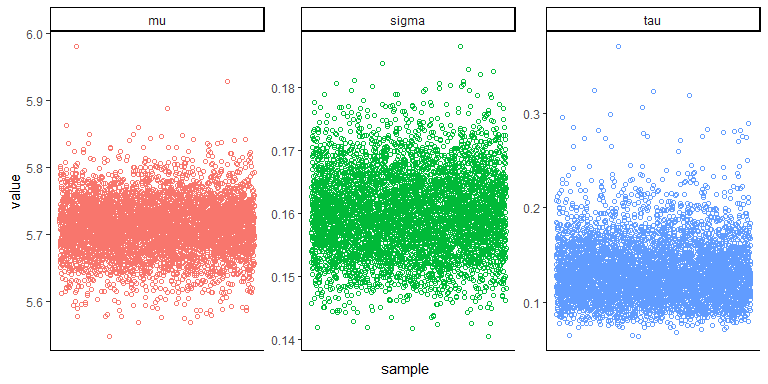
The effective sample sizes are close the the number of gibbs samples drawn. Thus we effectively have a large sample after using gibbs sampling.

Effective sample sizes for person-specific means :

|  |  |  |
| --- | --- | --- |
| Subject | numSim | Effective.Sample.Size |
| p1 | 5000 | 5000.00 |
| p2 | 5000 | 4564.68 |
| p3 | 5000 | 4700.17 |
| p4 | 5000 | 5000.00 |
| p5 | 5000 | 5000.00 |
| p6 | 5000 | 5000.00 |
| p7 | 5000 | 4669.96 |
| p8 | 5000 | 4708.82 |
| p9 | 5000 | 5000.00 |
| p10 | 5000 | 5000.00 |
| p11 | 5000 | 5000.00 |

We also get large effective sample sizes for the person-specific means.

Next I review the traceplots for the parameters in gibbs sampling.



All of the parameters traceplots look stationary thus there is no issue with the samples being taken.