Homework 8

Tyler Poelking

4/3/2018

Question 1) Provide interpretations for all of the unknown parameters in the model above in the context of the problem.

μij represents the mean change in packed cell volume from each breed j given the ith's cat dose and type.

σ2 represents the variability of the mean change in packed cell volume about the breed dose and type specific mean.

aj represents the change in the mean change in packed cell volume due to the given breed of the cat.

B1 represents the baseline of the mean change in packed cell volume of secondary erythrocytosis without taking breed and dose into account.

B2 represents the change in mean change in pakced cell value as the dose increases by one.

B3 represents the change in mean change in pakced cell value as the type goes from secondary erythrocytosis to primary erythrocytosis.

γ (gamma) represents the mean of the change in the mean change in packed cell volume of domestic (breed = A,B,or C) cats. Note aj (see interpretation above) has a mean of zero when the cat is non-domestic (breed = D or E).

τ2α represents the variability of the domestic specific change in the mean change in packed cell volume about the domestic specific mean.

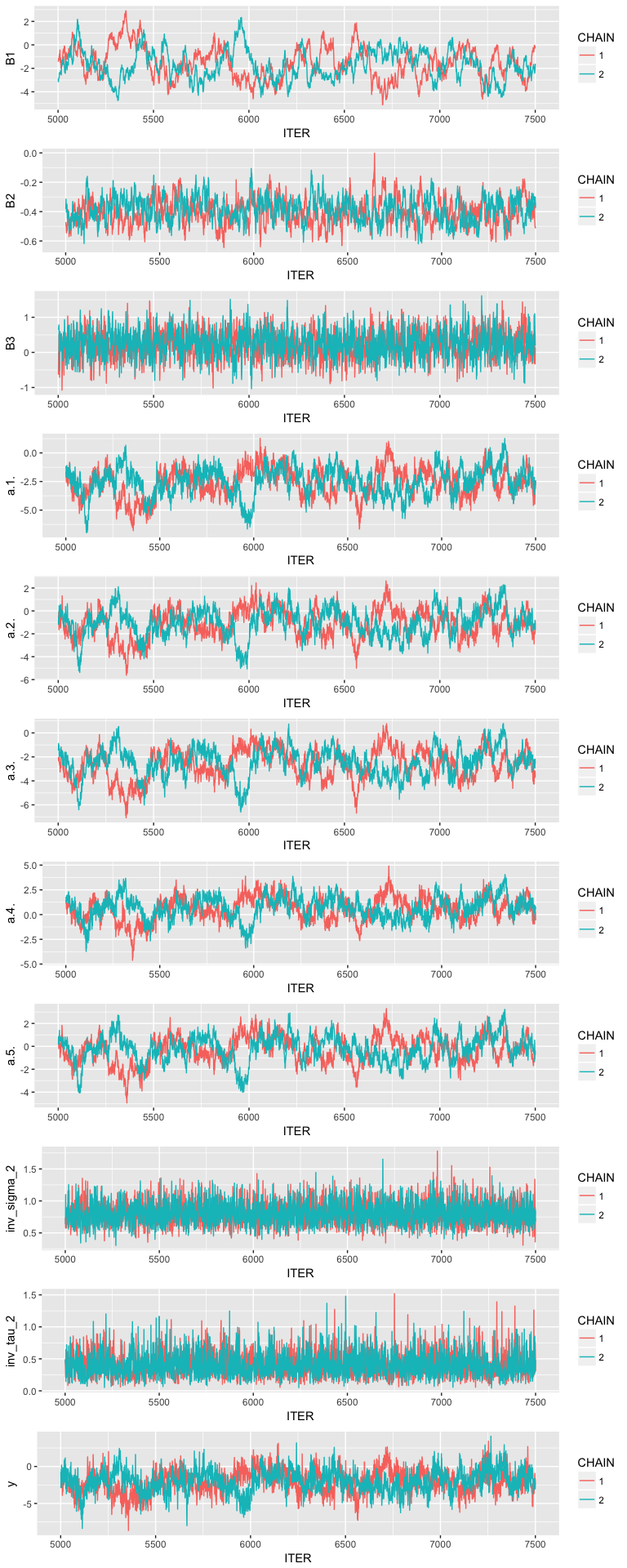
Question 2)

cats = read.table("~/Downloads/cats.txt", header=T)  
  
#add domestic column  
cats <- transform(cats, Domestic= ifelse(Breed%in% c('A', 'B','C'), 1, 0))  
#View(cats)  
attach(cats)  
  
n = length(DeltaPCV)  
  
breedNumeric <- as.numeric(Breed)  
typeNumeric <- as.numeric(Type)  
domestic<-c(1,1,1,0,0)  
  
breedUniq <-length( unique( Breed ) )  
typeUniq <-length( unique( Type ) )  
  
  
# create objects for JAGS  
dataList <- list( "DeltaPCV" = DeltaPCV,  
 "breedUniq" = breedUniq,  
 "n" = n,  
 "Dose" = Dose,  
 "Breed" = breedNumeric,  
 "Domestic" = domestic,  
 "Type" = typeNumeric)  
  
# list of parameters to be monitored   
parameters <- c( "a",   
 "B1",  
 "B2",  
 "B3",  
 "y",  
 "inv\_sigma\_2",  
 "inv\_tau\_2")  
  
  
# set initial values  
initsValues <- list( "a" =rep(0, breedUniq),   
 "B1" = 1,  
 "B2"=1,  
 "B3"=1,  
 "y" = 1,  
 "inv\_sigma\_2"=.1 ,  
 "inv\_tau\_2"= .1)  
  
# number of iteration for "tuning"   
adaptSteps <- 5000   
  
# number of iterations for "burn-in"   
burnInSteps <- 5000   
  
# number of chains to run  
nChains <- 2   
  
# total number of iterations to save  
numSavedSteps <- 5000   
  
# "thinning" (1 = keep every interation)  
thinSteps <- 1   
  
# iterations per chain  
ITER <- ceiling( (numSavedSteps \* thinSteps) / nChains )   
  
# -------------  
# Run JAGS  
# -------------  
  
# create, initialize, and adapt the model  
jagsModel <- jags.model("~/Desktop/All Stuff/School Stuff/STATS/3303/Homework8/cats\_model.txt",   
 data = dataList,   
 inits = initsValues,   
 n.chains = nChains,   
 n.adapt = adaptSteps)

## Compiling model graph  
## Resolving undeclared variables  
## Allocating nodes  
## Graph information:  
## Observed stochastic nodes: 40  
## Unobserved stochastic nodes: 11  
## Total graph size: 226  
##   
## Initializing model

# burn-in the algorithm  
update( jagsModel,   
 n.iter = burnInSteps )  
  
# run algorithm to get interations for inference  
codaSamples <- coda.samples( jagsModel,   
 variable.names = parameters,   
 n.iter = ITER,   
 thin = thinSteps )  
  
# -------------  
# Look at posterior samples  
# -------------  
  
# make a dataframe with the posterior samples  
mcmcChainDF <- data.frame( as.matrix( codaSamples,   
 iters = T,   
 chains = T ) )  
  
# create a vector with the variable names  
varNames <- names( mcmcChainDF )[3:( dim( mcmcChainDF )[2] )]  
  
# number of variables  
nVars <- length( varNames )  
  
mcmcChainDF$CHAIN <- as.factor(mcmcChainDF$CHAIN)  
  
# construct trace plots  
p <- list()  
for( k in 1:nVars )  
{  
 plot\_frame <- mcmcChainDF  
 plot\_frame$dep\_var <- mcmcChainDF[ , varNames[k]]  
 p[[k]] <- ggplot( plot\_frame,   
 aes( x = ITER,   
 y = dep\_var)) +  
 geom\_line( aes( color = CHAIN ) ) +   
 labs( y = varNames[k] )  
}

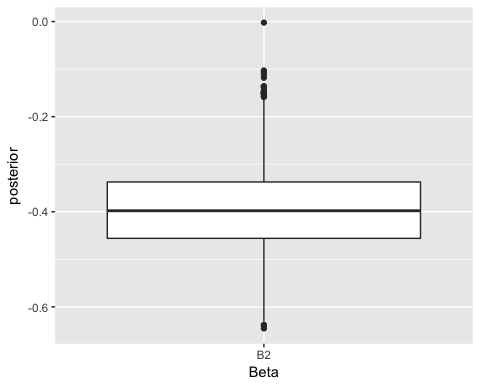
do.call( grid.arrange, c( p, list("ncol" = 1) ) )



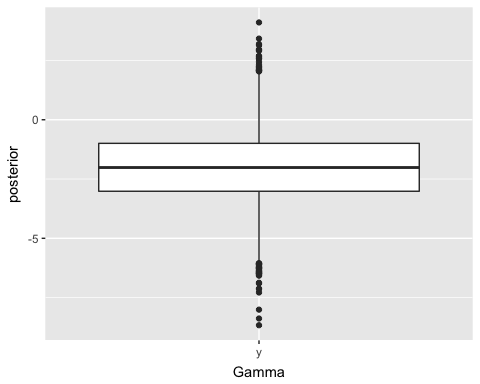
The initial values can be found in the code above. The number of iterations for tuning and burn-in were set to 5000. Two chains ran for an additional 2,500 iterations each. The trace plots above show both chains apear to be sampling from the same distributions, thus providing evidence the algorithm converged.

Question 3)

#For part a  
postDFreshape <- melt( mcmcChainDF,   
 id.vars = "ITER",  
 measure.vars = c("B2"))  
ggplot(postDFreshape,   
 aes(x = variable, y = value )) +  
 geom\_boxplot() +  
 scale\_x\_discrete( labels = c( "B2" )) +  
 ylab( "posterior" ) +  
 xlab( "Beta" )



#For part b  
postDFreshape <- melt( mcmcChainDF,   
 id.vars = "ITER",  
 measure.vars = c("y"))  
ggplot(postDFreshape,   
 aes(x = variable, y = value )) +  
 geom\_boxplot() +  
 scale\_x\_discrete( labels = c( "y" )) +  
 ylab( "posterior" ) +  
 xlab( "Gamma" )



b2 = mcmcChainDF$B2  
mean\_b2 = mean(b2)  
sd\_2\_b2 = sd(b2)  
  
mean\_b2 + sd\_2\_b2 \* qnorm(.95)

## [1] -0.2533297

mean\_b2 - sd\_2\_b2 \* qnorm(.95)

## [1] -0.5369375

y = mcmcChainDF$y  
mean\_y = mean(y)  
sd\_2\_y = sd(y)  
  
mean\_y + sd\_2\_y \* qnorm(.95)

## [1] 0.549883

mean\_y - sd\_2\_y \* qnorm(.95)

## [1] -4.590829

1. Is a higher dose of hydroxyurea associated with a greater reduction in packed cell volume?

B2 has a 95% credible interval of (-0.2435039, -0.5351103). This leads to the conclusion that a higher dose of hydroxyurea is associated with a greater reduction in packed cell volume. In other words, there is evidence that the medication worked! We can be confident in this conclusion because the lower bound = -.24 which is still a fair way away considering the entire interval is .29. The box plot above further demonstrates this conclusion since the third quartile is below zero.

1. Controlling for the dose of hydroxyurea administered and the type of erythrocytosis, do domestic breeds of cats have a different expected change in pack cell volume than non-domestic cats?

y (gamma) has a 95% credible interval of (0.682959, -4.877819). Because zero is included in this credible interval, this is evidence that domestic breeds of cats have do not have a different expected change in pack cell volume compared to non-domestic cats. We can be fairly confidence in this interval but now as confident as we are in the conclusion above. This is because the lower bound of the CI is only 0.68 away from zero, where the higher bound is 4.87 away from zero. Therefore, we do not know if y is positive or negative but it is more likely to be negative (which implies domestic breeds lean towards a greater reduction in packed cell volume). The box plot above further demonstrates this conlusion since the third quartile is less than zero.