**Cullen Blair**

**5380 Homework 3**

**Due Wednesday March 24 by 2 pm**

* **Give the R code and output for the problems.**
* **Use jags or stan by taking 100000 samples for 1 chain.**

**P.3.1 P.5.14 (p 81, 137) Consider where the data is provided on the next page.**

1. **Do part (a) stated in the text. Identify the parts of the hierarchical Poisson-Gamma model to write out the joint posterior using proper notation. Use Gamma(0.001,0.001) distributions for both hyperparameters.**

Where *i*=1,…,10 ; *j*=1,2 ;

1. **Obtain the posterior summaries for which is the average of for the 10 blocks. Carefully interpret the corresponding 95% credible interval for this application.**

dd=100000  
y = c(58,90,48,57,103,57,86,112,273,64)  
n=length(y)  
library(R2jags)

samp.out.vec=function(samp) {  
 mean=apply(samp,2,mean);sd=apply(samp,2,sd)  
 qq=t(apply(samp,2,quantile,probs=c(0.25,0.5,0.75,0.975)))  
round(cbind(mean,sd,qq),digits=5)}  
samp.out = function(d){   
 c(mean = mean(d), sd = sd(d), quantile(d, probs = c(0.025, 0.975)))  
}  
mod.str="model {  
 for (i in 1:n) { # precision! #  
 y[i] ~ dpois(theta[i]) # likelihood #  
 theta[i] ~ dgamma(phi[1],phi[2]) } #stage 1 prior  
 for (j in 1:2) {phi[j]~dgamma(0.001,0.001) #stage 2 prior  
}}  
"  
dat.l = list(n=n,y=y)  
parms = c("theta","phi") # initialize tau #  
init = function() list("theta"=y,"phi"=c(1,1)) #start by a/b for phi.j by mome  
jags.mod = R2jags::jags(data=dat.l,parms,inits=init,n.iter=dd,  
 n.chains=1,n.thin=1,n.burnin=0,model.file=textConnection(mod.str))

post.samp = as.mcmc(jags.mod)[[1]][,c(-1)]

sum=samp.out.vec(post.samp)sum[c(3,5:12,4),]

## mean sd 25% 50% 75% 97.5%  
## theta[1] 59.33623 7.60727 54.06540 58.97998 64.28953 75.11264  
## theta[2] 90.18982 9.30520 83.79810 89.87891 96.22967 109.43446  
## theta[3] 49.64302 6.97978 44.75125 49.34013 54.18263 64.19868  
## theta[4] 58.32103 7.52752 53.08768 57.97470 63.16870 73.95592  
## theta[5] 102.73248 9.97130 95.86826 102.44233 109.26022 123.24480  
## theta[6] 58.36725 7.52320 53.09623 58.06205 63.22601 74.09699  
## theta[7] 86.29989 9.12204 79.98321 85.96263 92.26593 105.03748  
## theta[8] 111.40535 10.35545 104.27834 111.11878 118.20100 132.49132  
## theta[9] 266.77781 16.47076 255.51876 266.34195 277.73737 299.92383  
## theta[10] 65.07894 7.94078 59.48956 64.76582 70.24791 81.49338

sum2=apply(post.samp, 1, samp.out)  
samp.out(sum2)

## mean sd 2.5% 97.5%   
## 93.4021240 81.7742306 0.4799083 240.3016734

There is 95% posterior probability that the population mean amount of total traffic for all blocks i is in [0.4799083,240.3016734] which corresponds to the equal tails interval.

1. **Obtain the posterior summaries for the parameters of the Gamma distribution in the stage 1 prior. Carefully interpret corresponding Bayes estimates for this application.**

sum[c(1,2),]

## mean sd 25% 50% 75% 97.5%  
## phi[1] 3.48617 1.64218 2.29891 3.20294 4.35445 7.42367  
## phi[2] 0.03681 0.01849 0.02346 0.03365 0.04669 0.08117

The Bayes estimate of the population parameter of shape (α) for the gamma distribution in the stage 1 prior is 3.48617 which corresponds to the posterior mean.

The Bayes estimate of the population parameter of rate (β) for the gamma distribution in the stage 1 prior is 0.03681 which corresponds to the posterior mean.

**P.3.2 Consider the linear regression problem described on the next page. Assume the improper prior (p 355). Be careful to use σ or where appropriate!**

1. **Obtain the posterior summaries for . Carefully assess the importance of these 4 predictors in the model using 95% credible intervals.**

#savings rate (sr) - personal saving divided by disposable income (y)  
#pop15 - percent population under age of 15 (x1)  
#pop75 - percent population over age of 75 (x2)  
#dpi - per-capita disposable income in dollars (x3)  
#ddpi - percent growth rate of dpi (x4)  
dat=read.csv("~/Downloads/savings.csv")  
y = dat$sr; n = length(y)  
x1 = dat$pop15;  
x2 = dat$pop75;  
x3 = dat$dpi;  
x4 = dat$ddpi  
fit=lm(y~x1+x2+x3+x4)  
sfit=summary(fit);bh.o=sfit$coefficient  
cov.bh=sfit$cov.unscaled;mse=sfit$sigma^2;

X=model.matrix(fit)[1:nrow(model.matrix(fit)),];p=ncol(X)  
#round(cbind(sfit$coefficients,confint(fit)),digits = 4) #check

mod.str="model {  
 for (i in 1:n) { # precision! #  
 y[i] ~ dnorm(mu[i],tau) # likelihood #  
 mu[i] <- beta[1]+beta[2]\*x1[i]+beta[3]\*x2[i]+beta[4]\*x3[i]+beta[5]\*x4[i]  
}  
 for (j in 1:p) { # beta prior #  
 beta[j] ~ dnorm(0,1e-06)  
}  
 tau ~ dgamma(1e-5,1e-5) # tau prior #   
 sigma.squared=1/(tau) # monitor sigma #  
}  
"  
dat.l = list(n=n,p=p,y=y,x1=x1,x2=x2,x3=x3,x4=x4)  
parms = c("beta","sigma.squared") # initialize tau #  
init = function() list("beta"=bh.o[,1],"tau"=1/mse)  
jags.mod = R2jags::jags(data=dat.l,parms,inits=init,n.iter=dd,  
 n.chains=1,n.thin=1,n.burnin=0,model.file=textConnection(mod.str))

post.samp = as.mcmc(jags.mod)[[1]][,c(-6)]   
sum=samp.out.vec(post.samp)  
rownames(sum)=c("B0","B1","B2","B3","B4","sigma.squared")  
sum #check w mse

## mean sd 25% 50% 75% 97.5%  
## B0 28.52326 7.52311 23.55497 28.53514 33.49467 43.33748  
## B1 -0.46048 0.14806 -0.55858 -0.46084 -0.36249 -0.16845  
## B2 -1.69035 1.10909 -2.42958 -1.68821 -0.95199 0.48156  
## B3 -0.00033 0.00096 -0.00096 -0.00033 0.00030 0.00156  
## B4 0.41047 0.20120 0.27619 0.41089 0.54396 0.80790  
## sigma.squared 15.15149 3.36279 12.78242 14.69685 17.00587 23.03195

According to the 95% posterior credible intervals above, there is evidence that percent of the population under 15, per-capita disposable income, and percent growth rate of dpi all indicate that their population linear regression trend on national savings rate is not zero as their credible intervals do not contain 0. However, percent of population over 75 does not indicate that it has a non-zero linear regression trend on national savings rate as its credible interval contains 0.

1. **Obtain the posterior predictive summaries for (United States). Interpret the Bayes estimate and uncertainty with respect to this application.**

sigma.squared=sum[6,1]  
beta=sum[1:p,1]  
set.seed(0)  
y.44=rnorm(dd,X[44,]%\*%beta,sqrt(sigma.squared))  
samp.out(y.44)

## mean sd 2.5% 97.5%   
## 8.671262 3.894067 1.028279 16.288926

The predictive savings rate for the United States from the population of such nations () has baye estimate of 8.671262% which is the posterior predictive mean, and a Bayes uncertainty of 3.894067 which is corresponds to the poster predictive standard deviation.

1. **Compute the PP-P with where is the Shapiro-Wilk test statistic value evaluated with the standardized residuals. Interpret the resulting PP-P value and what it indicates about the normality assumption of the model.**

set.seed(23) # 10000 x n #  
r.obs = pred.s = y.rep = r.rep = matrix(NA,dd,n);  
for (i in 1:dd) { # scale by std dev! #  
 pred.s[i,] = X%\*%post.samp[i,1:p]# p 362 #  
 r.obs[i,] = (y-pred.s[i,])/sqrt(post.samp[i,p+1])  
 y.rep[i,] = rnorm(n,pred.s[i,],sqrt(post.samp[i,p+1]))  
 r.rep[i,] = (y.rep[i,]-pred.s[i,])/sqrt(post.samp[i,p+1]) }  
pred.bayes = apply(pred.s,2,mean) # Bayesian predicted #  
res.bayes = apply(r.obs,2,mean) # Bayesian residuals #  
  
shap.func= function(x) -1\*shapiro.test(x)$statistic  
w.pred.s=apply(r.rep, 1, shap.func)  
w.pred.obs=apply(r.obs,1,shap.func)  
ppp=sum(w.pred.s>=w.pred.obs)/dd  
ppp

## [1] 0.79225

pred.freq = fit$fitted # predicted values #  
res.freq = fit$residual/sfit$sigma # # standard residuals  
#shapiro.test(res.freq)# # Frequentist Check 0.8524

The PP-P of .79225 means that the replicated values of the Shapiro-Wilk test statistic with standardized residuals exceed the observed values of the Shapiro-Wilk test statistic 79.225% of the time. Thus, observed values of the Shapiro-Wilk test statistic are consistent with corresponding replicated values; no indication of model discrepancy with respect to the normality of the model.