Instructions

(1) Homework must be typed and answered in the order given (problem 1(a)(b)(c)(d) first, problem 2(a)(b)... second, etc...)

(2) Undergrads and grads will answer all questions.

(3) Include in each part of the homework only the answer. R code and R output (without mistakes), must be included in the appendix to the question. For example, for question 1.a, write only the answer and your comments. The code and output for that part of the question will be in the appendix (the last part of question 1).

(4) No late homework under any circumstances.

(5) Write your name and ID this way: Last name, first name, UCLA ID, date, Homework number.

(6) Do not just give a number as an answer. For example, if asked for probability that posterior proportion is larger than 0.7, write $Prob(p > 0.7) = 0.3$, say and write comments or explanations if needed.

(7) The homework must be turned in in lecture (no mail box, no e-mail).

Homework

Problem 1. (Hoff, chapter 8, problem 3). The files `school1.dat` through `school8.dat` give weekly hours spent on homework for students sampled from eight different schools. Obtain posterior distributions for the true means for the eight different schools using a hierarchical normal model with the following prior parameters:

$$\mu_0 = 7; \gamma_0^2 = 5; \tau_0^2 = 10, \eta_0 = 2; \sigma_0^2 = 15; \nu_0 = 2$$

(a) Run a Gibbs sampling algorithm to approximate the posterior distribution of $(\theta, \sigma^2, \mu, \tau^2)$. Assess the convergence of the Markov chain, and find the effective sample size for $\left\{\sigma^2, \mu, \tau^2\right\}$. Run the chain long enough so that the effective sample sizes are all above 1000.

(b) Compute posterior mean and 95% confidence regions for $\left\{\sigma^2, \mu, \tau^2\right\}$. Also, compare the posterior densities to the prior densities, and discuss what was learned from the data.

(c) Plot the posterior density of $R = \frac{\tau^2}{\sigma^2+\tau^2}$ and compare it to a plot of the prior density of R. Describe the evidence for between-school variation.

(d) Obtain the posterior probability that θ_7 is smaller than θ_6, as well as the posterior probability that θ_7 is the smallest of all the $\theta' s$.

(e) Plot the sample averages $\bar{y}_1, ..., \bar{y}_8$ against the posterior expectations of $\theta_1, ..., \theta_8$, and describe the relationship. Also compute the sample mean of all observations and compare it to the posterior mean of μ.

(f) Attach your code at the end of the problem, well labelled, saying which section is which, and with comments.

Solution 1. (a) The effective sample sizes for MST (μ, τ^2, σ^2), respectively are (4189.287, 4485.125, 3904.720). So running the Gibbs is like obtaining that many iid random draws from the marginal posterior distribution of each of these parameters.

I assess the stationarity and convergence of the chains using box plots of every 500 observations. As we can see in Figure 1, the chains converge very quickly and are stationary. All medians in the box plots for μ are close to 7.5. All medians in plot of σ^2 are close to 14.3. And all medians around τ^2 are close to 5.

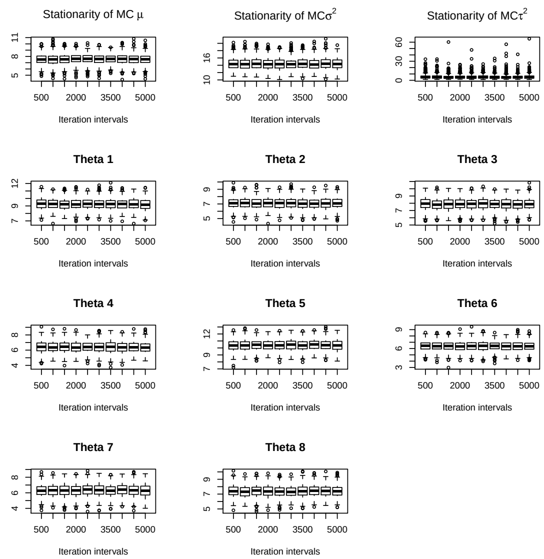

Figure 1: Checking stationarity of Markov Chains for μ, σ^2 and τ^2

(b) The posterior mean of μ is 7.56682 with 95% posterior interval (5.979894, 9.151025). Prior mean of μ is 7, and the prior 95% interval is (2.617387, 11.382613), much wider than the posterior one. Thus, the data made us increase our certainty about where the value of μ lies as compared to what we believed before observing the data.

The posterior mean of σ^2 is 14.50886 with 95% posterior interval (11.71345, 17.88898). The prior distribution of the precision of the distribution of the data,$1/\sigma^2$, is Gamma(1,15). Thus the prior mean is 14.92537. The 95% prior interval for σ^2 is (4.066275, 592.468353), much wider. Thus, we are much more certain after seeing the data of where the within school variable σ^2 is. We also notice that the average of the between school variability moved from prior value 14.92537 to posterior value 14.50. They are pretty close, though.

The posterior mean of τ^2 (between school variability) is 5.444953 with 95% posterior interval (1.867804, 14.473082). Thus, a posteriori, there is more within school variability than between school variability. The prior distribution of the precision of the distribution of the thetas, $1/\tau^2$, is Gamma(1,10), so the prior mean of the precision is 0.1. Thus the prior mean of τ^2 is 10. The 95% prior interval for τ^2 is (2.710850,394.978902), much wider than the posterior one. Thus, we are much more certain after seeing the data of where the between school variability is, and we have also noticed that the posterior average intraclass variability is smaller than before we saw the data (from 10 to 5.44).

(c) As we can see in Figure 2,the prior for R was very diffused, almost uniform across the (0,1) range. But the posterior distribution shows that the most likely values for R are between 0.15 and 0.3 after seeing the data. Thus about 20% of the total variability in the data is due to the across school variability, which is an indication that there is not that much difference among the groups.

(d)

$$prob(\theta_7 < \theta_6) = 0.5384$$

$$prob(\theta_7 = min(\theta_{1:8})) = 0.3328$$

(e)

(f)
```
############### Read the data from the author's web site ##############
y1<-scan("http://www.stat.washington.edu/~hoff/Book/Data/hwdata/school1.dat")
```

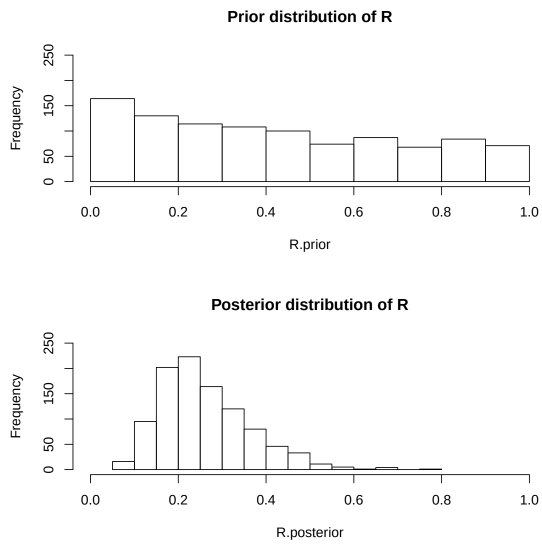

Figure 2: Markov Chain, Q3, hwk 4, initial $\beta = 1, \lambda = (0.1, 0.1, ..., 0.1$

```
y2<-scan("http://www.stat.washington.edu/~hoff/Book/Data/hwdata/school2.dat")
y3<-scan("http://www.stat.washington.edu/~hoff/Book/Data/hwdata/school3.dat")
y4<-scan("http://www.stat.washington.edu/~hoff/Book/Data/hwdata/school4.dat")
y5<-scan("http://www.stat.washington.edu/~hoff/Book/Data/hwdata/school5.dat")
y6<-scan("http://www.stat.washington.edu/~hoff/Book/Data/hwdata/school6.dat")
y7<-scan("http://www.stat.washington.edu/~hoff/Book/Data/hwdata/school7.dat")
y8<-scan("http://www.stat.washington.edu/~hoff/Book/Data/hwdata/school8.dat")

########Make a matrix Y with the school number and score, as in program p.137 Hoff ###

Y=matrix(        c(   rep(  c(1,2,3,4,5,6,7,8),c(length(y1),length(y2),length(y3),length(y4),
length(y5), length(y6), length(y7), length(y8)  ) )   ,
y1,y2,y3,y4,y5,y6,y7,y8), ncol=2         )


#####
#  Some descriptive plots of the data to get an idea of variability
#   in the data.  First, boxplots (without ranking as in Fig 8.4)
#  The histogram of means and scatter plot of sample size and sample mean given later.
#  (as in Fig 8.5). This graph not shown in the answer key.
#########

#pdf('hwk4q1a,pdf')  This is just for me to know where I saved the plot
boxplot(Y[,2]~Y[,1], xlab="School", ylab="Weekly hours doing homework",
main="Homework habits of students" )   # see boxplots of weekly hours in all schools
#dev.off()    #  again, just for me, to close the plot device


#####
```

```
#### MCMC analysis for school data
###### Program on page 137.
##  with some comments

# Model in the chapter and the lecture notes:
#        y_ij ~ N(thetaj, sigma^2)
#        theta_j ~ N(mu, tau^2)
#        mu ~    N(mu_0, gamma_0^2)
#        1/tau^2 ~ Gamma(eta_0/2, eta_0 tau_0^2/2)
#        1/sigma^2 ~Gamma(nu_0/2, nu_0 sigma_0^2/2)


### weakly informative priors
nu0<-2  ; s20<-15  # parameter values in dist of 1/sigma^2
eta0<-2 ; t20<-10  # parameter values in dist of 1/tau^2
mu0<-7 ; g20<-5   # parameter values in dist of mu
###

### starting values
m<-length(unique(Y[,1]))   # finds out m (number of schools)

n<-sv<-ybar<-rep(NA,m)   # create vectors for n_j, sv_j, ybar_j
for(j in 1:m)
{
  ybar[j]<-mean(Y[Y[,1]==j,2]) # put sample mean (ybar) of each school
  sv[j]<-var(Y[Y[,1]==j,2]) # put sample variance (sv) of each school
  n[j]<-sum(Y[,1]==j) # put sample size of each school
}

### Some plots that summarize the data. Not shown in the answer key

par(mfrow=c(1,2))

hist(ybar,xlab="sample mean", main="Empirical distribution of sample means")
 plot(n,ybar,xlab="sample size",ylab="sample mean", main="relationship between sample mean and sampl

# After copy pasting the picture, type dev.off() to close the window.

#### set theta_j^(0), sigma^2(0),tau^2(0), mu(0), all hyperparameters

theta<-ybar  # for init values of thetaj in each school, use ybar_j
sigma2<-mean(sv)  # for init sigma^2 use mean of variances of all schools
mu<-mean(theta)  # for init for mu use the mean of the ybars
tau2<-var(theta) # for init for tau^2 use the variance of the y-bars


### setup MCMC
S<-5000  # we will do 5000 iterations
THETA<-matrix( nrow=S,ncol=m)  # 5000 iterations of theta_j j =1-9
```

```
MST<-matrix( nrow=S,ncol=3)  # 5000 iterations of mu,tau^2, sigma^2
###

### MCMC algorithm
###  The loops here are based on the conditional distributions
##  See class notes or full conditional on pages 134 and 135 of
### Hoffłs book

for(s in 1:S)    # Repeat S times
{

  # sample new values of the thetas from conditional for theta_j
  for(j in 1:m)
  {
    vtheta<-1/(n[j]/sigma2+1/tau2)  # var of condit. dist of theta_j
    etheta<-vtheta*(ybar[j]*n[j]/sigma2+mu/tau2) #mean of dist theta_j
    theta[j]<-rnorm(1,etheta,sqrt(vtheta)) # draw random number of theta_j
   }

  #sample new value of sigma2 from conditional for sigma^2
  nun<-nu0+sum(n)
  ss<-nu0*s20
  for(j in 1:m){ss<-ss+sum((Y[Y[,1]==j,2]-theta[j])^2)}
  sigma2<-1/rgamma(1,nun/2,ss/2)  # draw sigma^2

  #sample a new value of mu from conditional for mu
  vmu<- 1/(m/tau2+1/g20)  #variance of mu
  emu<- vmu*(m*mean(theta)/tau2 + mu0/g20)  # mean of mu
  mu<-rnorm(1,emu,sqrt(vmu)) # draw random number from dist of mu

  # sample a new value of tau2
  etam<-eta0+m
  ss<- eta0*t20 + sum( (theta-mu)^2 )
  tau2<-1/rgamma(1,etam/2,ss/2)

  #store results
  THETA[s,]<-theta   # each row of THETA matrix contains 100 thetas
  MST[s,]<-c(mu,sigma2,tau2) # each row of MST matrix
       # contains a mu, a sigma^2 and a tau^2

}

#############
#   Check convergence with plots of traces
#############

par(mfrow=c(3,4))
for(i in 1:8){
   plot(1:nrow(THETA), THETA[,i], xlab="Iteration", ylab=paste("THETA[,",i,"]"),
   main=paste("theta ",i," "),  type="l"  )
 }
```

```
 for(i in 1:3){
   plot(1:nrow(MST), MST[,i], xlab="Iteration", ylab=paste("THETA[,",i,"]"),
   main=paste("MST ",i," "),  type="l"  )
 }


##############
# Q 1(a) Hwk 4.Find the effective sample sizes.  The effective sample size estimates the
#  number of independent iterations  necessary to give the same precision as the MCMC
#  samples.  REf. See Hoff Chapter 6
# I also check the convergence using box plots of the traces.
#############

library(coda)
effectiveSize(MST )  #gives effective sample size for each column of MST


### Hoff's function to create stationarity plots to assess convergence

stationarity.plot = function(x,...){
S = length(x)
scan = 1:S
ng = min(round(S/100),10)
group=S*ceiling(ng*scan/S)/ng
boxplot(x~group,...)
}

### stationarity plots for columns of MST matrix
pdf("hwk4q1stationarity.pdf")
par(mfrow=c(4,3))
stationarity.plot(MST[,1],xlab="Iteration intervals",main=expression(paste("Stationarity of MC ", mu
stationarity.plot(MST[,2],xlab="Iteration intervals",main=expression(paste("Stationarity of MC",  si
stationarity.plot(MST[,3],xlab="Iteration intervals",main=expression(paste("Stationarity of MC", tau
for(i in 1:8){
stationarity.plot(THETA[,i],xlab="Iteration intervals",main=paste("Theta", i ) )
}
dev.off()




############
#  Question 1.b, hwk 4
#############

for(j in 1:3){
print(mean(MST[,j]))
print(quantile(MST[,j], c(0.025,0.975)))
}
```

```
# To compare with prior densities (Q 1.b)
#  we generate random numbers from the
# prior densities


######### for prior interval for mu

qnorm(c(0.025,0.975),7,sqrt(5))

####### for prior interval for tau^2

1/qgamma(c(0.025,0.975), 1,10)

######### for prior interval for sigma^2

1/qgamma(c(0.025,0.975), 1,15)

##############################
# Question 1.c  Prior and posterior for R
###############################

#### prior R

sigma2 =1/rgamma(1000, 1,15)
tau2=1/rgamma(1000, 1,10)
R.prior = tau2/(tau2+sigma2)

#### posterior R

 R.posterior =  MST[500:1500,3]/(MST[500:1500,3]+MST[500:1500,2])

#### Plot the two distributions together

pdf("hwk4q1c.pdf")
par(mfrow=c(2,1))
hist(R.prior,xlim=c(0,1), ylim=c(0,250), xlab="R.prior", main="Prior distribution of R")
hist(R.posterior,xlim=c(0,1),ylim=c(0,250), xlab="R.posterior",main="Posterior distribution of R")
dev.off()

################################
#  Question 1.d
################################

mean(THETA[,7] < THETA[,6] )
x=rep(0,S)

for(i in 1:S){
 if(THETA[i,7] == min(THETA[i,]) )   x[i]=1 }
mean(x)
```

Problem 2. (Hoff, Chapter 9, problem 3). The file `crime.data` contains crime rates and data on 15 explanatory variables for 47 U.S. states, in which both the crime rates and the explanatory variables have been centered and scaled to have variance 1. A description of the variables can be obtained by typing `library(MASS); ?UScrime` in R

(a) Fit a regression model $y = X\beta + \epsilon$ using the g-prior with $g = n; \nu_0 = 2$; and $\sigma_0^2 = 1$. Obtain marginal posterior means and 95% confidence intervals for β, and compare to the least squares estimates. Describe the relationships between crime and the explanatory variables. Which variables seem strongly predictive of crime rates?

```
library(MASS)
data(UScrime)
 attach(UScrime)
 y=y
 X=cbind(M,So,Ed,Po1,Po2,LF,M.F,Pop,NW,U1,U2,GDP,Ineq,Prob)

 g=length(y); nu0=2; s20=1
S=1000

 ##  data: y, X
 ## prior parameters: g, nu0, s20
 ## number of independent samples to generate: S

 n = dim(X)[1] ; p=dim(X)[2]
 Hg =(g/(g+1))*X%*%solve(t(X)%*%X)%*%t(X)
 SSRg= t(y)%*%(diag(1,nrow=n) -Hg)%*%y

 s2 = 1/rgamma(S, (nu0+n)/2, (nu0*s20+SSRg)/2)

 Vb=g*solve(t(X)%*%X)/(g+1)
 Eb=Vb%*%t(X)%*%y

 E = matrix(rnorm(S*p, 0, sqrt(s2)),S,p)
 beta=t( t(E%*%chol(Vb)) + c(Eb))
```

(b) Lets see how well regression models can predict crime rates based on the $X-$ variables. Randomly divide the crime roughly in half, into a training set $\{y_{tr}, X_{tr}\}$ and a test set $\{y_{te}, X_{te}\}$.

newmat ¡- mainmat[sample(yourmatrixsize, size=howmuchtosample, replace=F),]

(i) Using only the training set, obtain least squares regression coefficients $\hat{\beta_{OLS}}$. Obtain predicted values for the test data by computing $\hat{y}_{OLS} = X_{te}\hat{\beta}_{OLS}$. Plot $\hat{y}_{OLS}$ versus y_{te} and compute the prediction error $\frac{1}{n_{te}} \sum(y_{i,te} - \hat{y}_{i,ols})^2$.

(ii) Now obtain the posterior mean $\hat{\beta}_{Bayes} = E[\beta \mid y_{tr}]$ using the $g-$prior described above and the training data only. Obtain predictions for the test set $\hat{y}_{Bayes} = X_{test}\hat{\beta}_{Bayes}$. Plot versus the test data, compute the prediction error, and compare to the OLS prediction error. Explain the results

(iii) Attach your code at the end of the problem, well labelled, saying which section is which, and with comments.

Problem 3. The number of pump failures Y_i over time periods t_i in 10 power plants is given below:

Y_i	5	1	5	14	5	19	1	1	4	22
t_i	94.320	15.72	62.880	125.760	5.240	31.440	1.048	1.048	2.096	10.480

We consider a hierarchical event rate model with a Poisson likelihood

$$Y_i \sim Poi(\lambda_i t_i)$$

and a prior model

$$\lambda_i \sim Ga(\alpha, \beta) \quad i = 1, ..., 10$$
$$\beta \sim Ga(c, d)$$

where (α, c, d) are fixed hyperparameters.

(a) Write down the joint posterior distribution for the relevant parameters

(b) Show that the full conditional distributions needed to do a Gibbs sampling are:
If $\lambda_j \quad (j \neq i)$ and β are given,

$$\begin{aligned} P(\lambda_i \mid \lambda_j, \beta, y) &\propto \lambda_i^{\alpha+y_i-1} e^{-\lambda_i(\beta+t_i)} \\ &\sim Ga(\alpha + y_i, \beta + t_i) \end{aligned}$$

If all the λ_i's are given, $i = 1, ..., 10$

$$\begin{aligned} P(\beta \mid \lambda_1, \lambda_2, ..., \lambda_1 0, y) &\propto \beta^{(n\alpha+c-1)} e^{-\beta(\sum \lambda_i + d)} \\ &\sim Ga(c + n\alpha, d + \sum \lambda_i) \end{aligned}$$

(c) Use the code given to you in the lab, modified if you need to, to do the remaining questions in this problem. First, run your gibbs 1000 times to simulate from the posterior distribution of the parameters using the following values for the hyperparameters: $c = 0.1, d = 0.1, \alpha = 1$ and the following initial values for the sampler:

$$\beta = 1, \lambda = (0.1, 0.1, ..., 0.1$$

Give the 5 number summary for your parameters (without the burn in) and comment on your findings. Include 95% posterior intervals. Check that the sampler has converged. Show trace plot of the 1000 iterations for each parameter.

(d) Run again the Gibbs sampler 1000 times (notice the problem said 5000 times, I run it another 1000 times) , but now with initial values

$$\beta = 2, \lambda = (0.5, 0.5, ..., 0.5$$

Give the 5 number summary for your parameters (without the burn in) and comment on your findings. Check that the sampler has converged. Show trace plot of the 1000 iterations for all parameters. Compare with the results you got with the previous initial values.

(e) Plot the posterior distributions of the parameters in one graph containing boxplots for all the parameters. Comment on the results.

(f) Attach your code indicating question number, etc...

Solution 2. (a)

$$\begin{aligned} P(\beta, \lambda_1, ..., \lambda_{10}) &\propto \left[\text{Poisson}(y_i \mid \lambda_i t_i) \prod_{i=1}^{n} \text{Gamma}(\lambda_i \mid \alpha, \beta)\right] \text{Gamma}(\beta \mid c, d) \qquad (1) \\ &= \left[\prod_{i=1}^{n} e^{-\lambda_i t_i} \frac{(\lambda_i t_i)^{y_i}}{y_i!} \quad \frac{\beta^\alpha}{\Gamma(\alpha)} (\lambda_i)^{\alpha-1} e^{-\beta\lambda_i} \quad \right] \frac{d^c}{\Gamma(c)} \beta^{c-1} e^{-d\beta} \qquad (2) \end{aligned}$$

(b) If λ_j, $(j \neq i)$, and β are given,

$$P(\lambda_i \mid \lambda_j, i \neq j, \beta, Y) \propto \lambda_i^{\alpha-1} e^{-\beta\lambda_i} e^{-\lambda_i t_i} \left(\frac{(\lambda_i t_i)^{y_i}}{y_i!} \right) \quad (3)$$

$$\propto \lambda_i^{\alpha+y_i-1} e^{-\lambda_i(\beta+t_i)} \quad (4)$$

$$\sim \text{Gamma}(\alpha + y_i, \beta + t_i) \quad (5)$$

If all the λ_i's are given, $i = 1, ..., 10$,

$$P(\beta \mid \lambda_1, ..., \lambda_{10}, y) \propto \left[\prod_{i=1}^n \beta^\alpha e^{-\beta\lambda_i}\right] \beta^{c-1} e^{-d\beta} \quad (6)$$

$$\propto \beta^{(n\alpha+c-1)} e^{-\beta(\sum \lambda_i)} \quad (7)$$

$$\sim \text{Gamma}(c + n\alpha, d + \sum \lambda_i) \quad (8)$$

(c) The traces of the Markov chain can be seen in figure 4 The chains converge after a few iterations. I report the summaries and posterior intervals for the second chain. See the next question. They are not too different from those found in the first chain.

(d) The traces of the Markov chain can be seen in 5. All chains converge after a few iterations.

The summaries of the parameters can be seen in the Table below. The box plots convey similar information.

As we can see, we are more uncertain about the average number of failures of pumps number 8 and 9. The box plots show large variability of the simulated averages. These are pumps that were observed very briefly, with one failure. In general, the plot shows that pumps observed over a small period of time (pumps 5,7,8,9) have less accurate estimates (wider variability in the estimates) than those observed over a longer period of time. The table below and the box plots reveal that the highest median number of failures are in pump 5, 7, 8, 9,10. It is better to use the median as a summary given the skewness of the marginal posterior distributions.

Coefficient	Min.	1st Qu.	Median	Mean	3rd Qu	Max	Post. Interval
λ_1 ch 2	0.007834	0.044260	0.058440	0.062120	0.076050	0.500	(0.022, 0.117)
λ_2 ch 2	0.002644	0.057680	0.098440	0.115800	0.155700	0.771	(0.0123, 0.314)
λ_3 ch 2	0.01596	0.06512	0.09025	0.09563	0.12060	0.50	(0.035, 0.188)
λ_4 ch 2	0.04851	0.09439	0.11480	0.12000	0.14060	0.50	(0.063, 0.193)
λ_5 ch 2	0.1819	0.6130	0.8294	0.8844	1.1030	2.375	(0.329, 1.597)
λ_6 ch 2	0.2465	0.5025	0.5935	0.6020	0.6857	1.197	(0.349, 0.932)
λ_7 ch 2	0.02619	0.37120	0.66580	0.80100	1.08200	4.133	(0.104, 2.104)
λ_8 ch 2	0.02083	0.36690	0.63380	0.79770	1.03500	4.226	(0.095, 2.366)
λ_9 ch 2	0.1462	0.9119	1.3060	1.4180	1.8170	5.104	(0.479, 3.082)
λ_{10}	0.500	1.630	1.883	1.920	2.186	3.485	(1.167, 2.850)
β ch 2	0.3929	1.0960	1.4330	1.5270	1.8650	4.7770	0.6694405 2.8688984

Notice that the mean of the β is 1.52. This, combined with the given $\alpha = 1$ means that the posterior mean of the population average for the λ_i's is $\frac{\alpha}{\beta} = \frac{1}{1.52} = 0.657$. You may see in the boxplots that pumps 1,2,3,4,6 have most simulated values below the mean of the population of lambdas, whereas the other pumps do not.

Conclusion: lacking more data, we conclude that power plants 1,2,3,4,6 are the safest of the 10 power plants given in this problem.

(e)

(f)
```
########################################
# Read the data in y, t
```

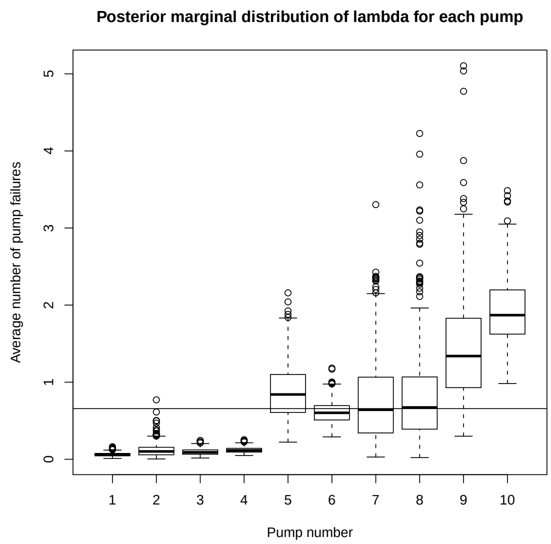

Figure 3: Marginal Posterior Distributions question 3, hwk 4

```
######################################

y = c(5,1,5,14,5,19,1,1,4,22)
ti= c(94.320, 15.72, 62.880, 125.760, 5.240, 31.440,
1.048,1.048,2.096,10.480)
n=length(y)


######################################
# Initialize parameters and do the Gibbs
# Part (c) of question 3 in Homework 4
######################################

alpha=1.0 ; c=0.1; d=0.1
beta=1
lambda=c(rep(0.1,10))

TH=c(lambda,beta)    #
T0=1000
for(i in 1:T0){
beta1=rgamma(1, (c+n*alpha))
beta=beta1/(d+sum(lambda))
lambda1=rgamma(10,(alpha+y))
lambda=lambda1/(beta+ti)
tht = c(lambda,beta)
TH=rbind(TH, tht)
}

#######################################
```

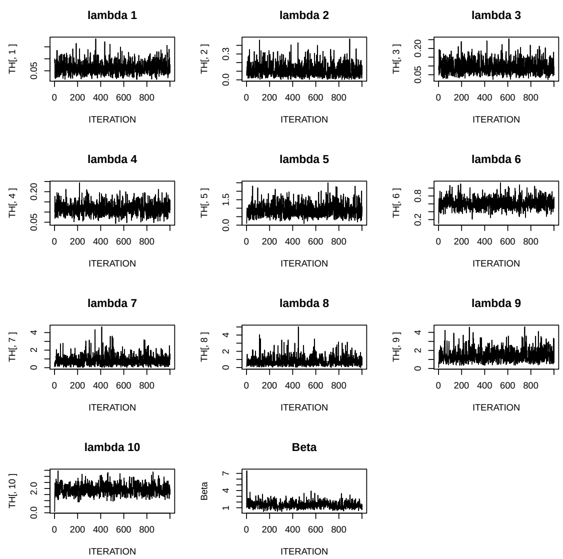

Figure 4: Markov Chain, Q3, hwk 4, initial $\beta = 1, \lambda = (0.1, 0.1, ..., 0.1$

```
# Plot the trajectories
######################################

 pdf("hwk4q3partc.pdf")
 par(mfrow=c(4,3))
 for(i in 1:10)
 {
 plot(1:nrow(TH), TH[,i],xlab="ITERATION",ylab=paste("TH[,",i,"]"), type="l", main=paste("lambda", i
 }
 plot(1:nrow(TH), TH[,11], xlab="ITERATION", ylab="Beta",type="l",main="Beta")
 dev.off()

#######################################
# Part (d) Question 3 Hwk 4
# Second chain. Different initial values
#######################################

TH2 = NULL
alpha=1; c=0.1; d=0.1 ; beta=2.0
lambda=c(rep(0.5,10))

TH2=c(lambda, beta)
T0=1000
for(i in 1:T0){
beta1=rgamma(1,(c+n*alpha))
beta=beta1/(d+sum(lambda))
lambda1=rgamma(10,(alpha+y))
lambda=lambda1/(beta+ti)
tht=c(lambda,beta)
```

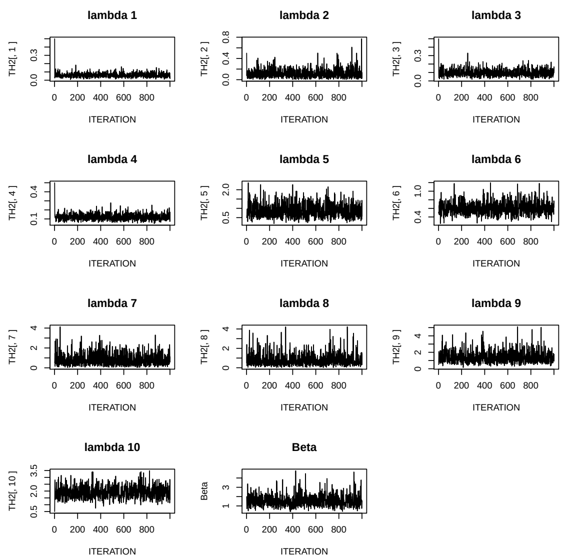

Figure 5: Markov Chain, Q3, hwk 4, initial $\beta = 2, \lambda = (0.5, 0.5, ..., 0.5$

```
TH2=rbind(TH2,tht)
}

#######################################
# Plot the new trajectories
#######################################

 pdf("hwk4q3partcnew.pdf")
 par(mfrow=c(4,3))

for(i in 1:10)
{
plot(1:nrow(TH2), TH2[,i],xlab="ITERATION",ylab=paste("TH2[,",i,"]"), type="l", main=paste("lambda",
}
plot(1:nrow(TH2), TH2[,11], xlab="ITERATION", ylab="Beta",type="l",main="Beta")

dev.off()

########################################
#  Plot both chains together
########################################

pdf("hwk4q3together.pdf")
par(mfrow=c(4,3))
for(i in 1:10)
{
TT = min(nrow(TH),nrow(TH2))
matplot(1:TT,cbind(TH[1:TT,i],TH2[1:TT,i]), xlab="ITERATION", ylab=paste("TH[,",i,"]"), type="l", ma
}
```

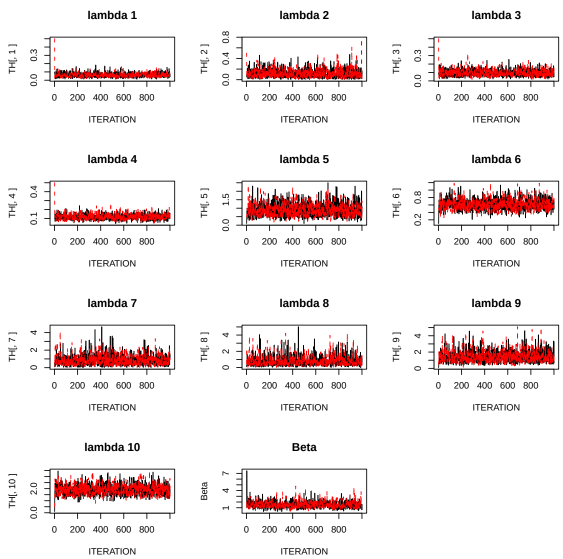

Figure 6: The two Markov chains, Q3, Hw 4

```
matplot(1:TT,cbind(TH[1:TT,11],TH2[1:TT,11]), xlab="ITERATION", ylab="Beta", type="l", main="Beta")
dev.off()

########################################
# Summary 5 number summary of the parameter values
# and obtain posterior 95 percent intervals
########################################

for(i in 1:11)
{
print(summary(TH2[500:1000,i]))
}

for(i in 1:11)
{
print(quantile(TH2[500:1000,i],c(0.025,0.975)))
}


#########################################
#  Part e, Q3, Hwk 4. Box plots of parameters simulated
#########################################

pdf("hwk4q3marginals.pdf")

boxplot(TH2[500:1000,1:10],xlab="Pump number",ylab="Average number of pump failures", main="Posterio
abline(h=0.657)
dev.off()
```