Stat 516, Homework 6

Name: Neal Marquez

Due date: Tuesday, November 28.

Note: Do this homework *individually.* Do not include any R code in your main handout – just include as an appendix, in compact form.

1. (a) As an illustration of rejection sampling, show how one can generate a draw from $N(0,1)$ by using a Cauchy proposal distribution. What is the acceptance rate of your sampler? What is the mean number of trials until acceptance of the Cauchy draw? (The Cauchy has density $1/\left[\pi(1+x^2)\right]$.)

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big) \\
g(x) &= \frac{1}{\pi(1+x^2)} \\
\frac{f}{g}(x) &= \frac{\pi(1+x^2)\exp\big(-\frac{x^2}{2}\big)}{\sqrt{2\pi}} \\
\frac{d}{dx} &= \frac{\pi}{\sqrt{2\pi}} - \exp\Big(-\frac{x^2}{2}\Big)x(x^2-1)
\end{aligned}
$$

set $\frac{d}{dx}$ to 0 to find x_M...

$$
\begin{aligned}
x_M &= \pm 1 \\
M &= \frac{2\pi}{\sqrt{2\pi}} \exp\left(\frac{-1}{2}\right) \approx 1.52 \\
p_a &= \frac{\int_{-\infty}^{\infty} f(x)}{M} = M^{-1} \approx .66
\end{aligned}
$$

Expected number of trails is then

$$
\begin{aligned}
E[X] &= p(X \geq 1) + p(X \geq 2) + p(X \geq 3) + \dots \\
&= 1 + (1-p_a) + (1-p_a)^2 + \dots \\
&= \frac{1}{p_a} = M
\end{aligned}
$$

(b) Again you wish to generate a draw from $N(0,1)$. But instead of using the standard Cauchy proposal from (a) you are using a scaled Cauchy with density $1/\left[\pi\gamma(1+(x/\gamma)^2)\right]$, where $\gamma > 0$ is a scale parameter. How would optimally choose γ?

$$\frac{f}{g}(x) = \frac{\pi\gamma(1+(x/\gamma)^2)\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi}}$$

find $\dfrac{d}{dx}$ and set to 0 to find x_M...

$$x_M = \sqrt{2-\gamma^2}$$

$$M = \frac{\pi\gamma(1+(2-\gamma^2)/\gamma^2)\exp\left(-\frac{2-\gamma^2}{2}\right)}{\sqrt{2\pi}}$$

find $\dfrac{d}{d\gamma}$ and set to 0 to find optimal γ...

$$\gamma = 1$$

(c) Is it possible to generate a draw from a Cauchy distribution using $N(0,1)$ as proposal distribution?
No because the integral of $\frac{f}{g}(x)$ is indefinite on the supoort for the standard normal distribution for any chosen cauchy distribution when g is the normal density and f is cauchy and furthermore M has no set value.

2. In this question we will analyze data on 10 power plant pumps using a Poisson gamma model. The number of failures Y_i is assumed to follow a Poisson distribution

$$Y_i|\theta_i \sim_{ind} \text{Poisson}(\theta_i t_i), \quad i = 1, ..., 10$$

where θ_i is the failure rate for pump i and t_i is the length of operation time of the pump (in 1000s of hours). The data is shown below.

Pump	1	2	3	4	5	6	7	8	9	10
t_i	94.3	15.7	62.9	126	5.24	31.4	1.05	1.05	2.1	10.5
y_i	5	1	5	14	3	19	1	1	4	22

A conjugate gamma prior distribution is adopted for the failure rates:

$$\theta_i|\alpha,\beta \sim_{iid} \text{Gamma}(\alpha,\beta), \quad i = 1, ..., 10,$$

with a hyperprior under which α and β are independent and

$$\alpha \sim \text{Exponential}(1), \qquad \beta \sim \text{Gamma}(0.1, 1).$$

(a) Carefully show, using Bayes theorem and the conditional independencies in the model description, that the posterior distribution is given by:

$$p(\alpha,\beta,\boldsymbol{\theta}|\mathbf{y}) \propto \prod_{i=1} \{\Pr(y_i|\theta_i) \times p(\theta_i|\alpha,\beta)\} \times \pi(\alpha,\beta) \tag{1}$$

where $\boldsymbol{\theta} = (\theta_1, ..., \theta_{10})$.

$$p(\alpha,\beta,\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\alpha,\beta,\boldsymbol{\theta}) \times p(\alpha,\beta,\boldsymbol{\theta})$$

$$p(\boldsymbol{y}|\alpha,\beta,\boldsymbol{\theta}) = \prod_{i=1} \frac{(t_i\theta_i)^{y_i} e^{-t_i\theta_i}}{y_i!} = \prod_{i=1} \Pr(y_i|\theta_i)$$

$$p(\alpha,\beta,\boldsymbol{\theta}) = \pi(\alpha,\beta)\prod_{i=1} \frac{\beta^\alpha \theta_i^{\alpha-1} e^{-b\theta_i}}{\Gamma(\alpha)} = \prod_{i=1} \{p(\theta_i|\alpha,\beta)\} \times \pi(\alpha,\beta)$$

(b) By using (1) write out the steps of a Metropolis-Hastings within Gibbs sampling algorithm to analyze these data.

Hint: First write down the forms for $(\alpha|\beta, \boldsymbol{\theta}, \mathbf{y})$, $(\beta|\alpha, \boldsymbol{\theta}, \mathbf{y})$, $(\theta_i|\boldsymbol{\theta}_{-i}, \alpha, \beta, \mathbf{y})$, for $i = 1, ..., 10$.

$$
\begin{aligned}
p(\theta_i|\boldsymbol{\theta}_{-i}, \alpha, \beta, \mathbf{y}) &\propto \prod_{i=1} \frac{(t_i\theta_i)^{y_i} e^{-t_i\theta_i}}{y_i!} \frac{\beta^\alpha \theta_i^{\alpha-1} e^{-b\theta_i}}{\Gamma(\alpha)} \\
&\propto \text{Gamma}(y_i + \alpha, \beta + t_i) \\
&\therefore \quad \theta_i \perp\!\!\!\perp \boldsymbol{\theta}_{-i}, \mathbf{y}_{-i}|y_i, \alpha, \beta \\
p(\beta|\alpha, \boldsymbol{\theta}, \mathbf{y}) &\propto \left\{ \prod_{i=1} \beta^\alpha e^{-\beta\theta_i} \right\} \Gamma(.1)^{-1} \beta^{-.9} e^{-\beta} \\
&= \beta^{10a-.9} e^{-\beta(1+\sum_{i=1}\theta_i)} \\
&\propto \text{Gamma}(10\alpha + .1, 1 + \sum_{i=1} \theta_i) \\
&\therefore \quad \beta \perp\!\!\!\perp \mathbf{y}|\boldsymbol{\theta}, \alpha \\
p(\alpha|\beta, \boldsymbol{\theta}, \mathbf{y}) &= e^{-\alpha}\beta^{10\alpha}\Gamma(\alpha)^{-10} \prod_{i=1} \theta_i^{\alpha-1} \\
&\therefore \alpha \perp\!\!\!\perp \mathbf{y}|\boldsymbol{\theta}\beta
\end{aligned}
$$

Algorithm...

i. Choose staring values for parameters $\boldsymbol{\theta}, \alpha, \beta$

ii. Update values of θ_i using last iteration of α and β with the distribution $\text{Gamma}(y_i + \alpha, \beta + t_i)$

iii. Update β using last iteration of α and current $\boldsymbol{\theta}$ with the distribution $\text{Gamma}(10\alpha + 1, 1 + \sum_{i=1} \theta_i)$

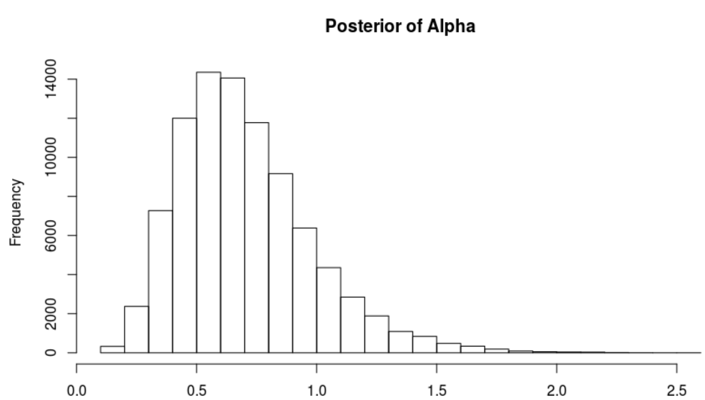
iv. simulate a value u which is distributed $\text{Uniform}(0, 1)$

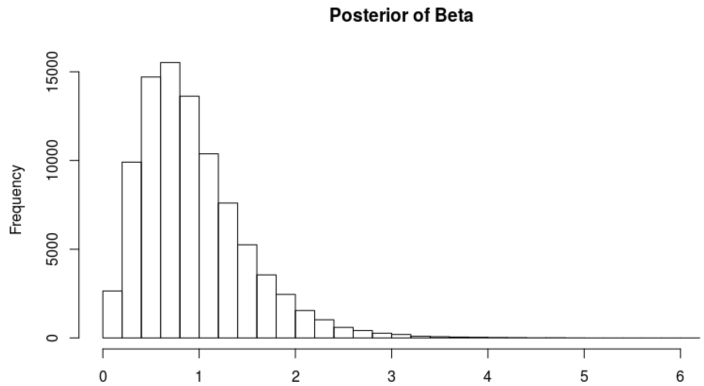
v. propose a new value of α, $\alpha^\star$, which is distributed $\mathcal{N}(\alpha, .2)$

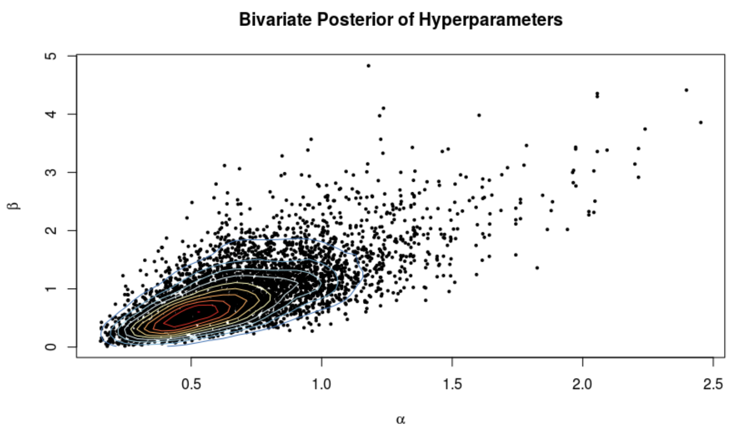
vi. accept $\alpha^\star$ as the new α if $\alpha^\star > 0$ and $u < \frac{p(\alpha^\star|\beta,\boldsymbol{\theta})}{p(\alpha|\beta,\boldsymbol{\theta})}$

vii. repeat steps ii-vi 1000000 times recording each new iteration of parameters $\boldsymbol{\theta}, \alpha, \beta$

(c) Apply your algorithm to the pump data and give histogram representations of the univariate posterior distributions for α and β and a scatterplot representing the bivariate posterior distribution.







(d) Analytically integrate θ_i, $i = 1, ..., 10$ from the posterior (1) and hence give the form, up to proportionality, of the posterior $p(\alpha, \beta|\mathbf{y})$.

$$
\begin{aligned}
p(\alpha, \beta|\mathbf{y}) &\propto \beta^{10a-.9} e^{-\beta(1+\sum_{i=1}\theta_i)} e^{-\alpha} \beta^{10\alpha} \Gamma(\alpha)^{-10} \prod_{i=1} \theta_i^{\alpha-1} \\
&= \beta^{20\alpha-.9} e^{-\beta(1+\sum_{i=1}\theta_i)-\alpha} \Gamma(\alpha)^{-10} \prod_{i=1} \theta_i^{\alpha-1}
\end{aligned}
$$

(e) Construct a Metropolis-Hastings algorithm, to provide a Markov chain with stationary distribution the posterior $p(\alpha, \beta|\mathbf{y})$.

i. Choose staring values for parameters α, β

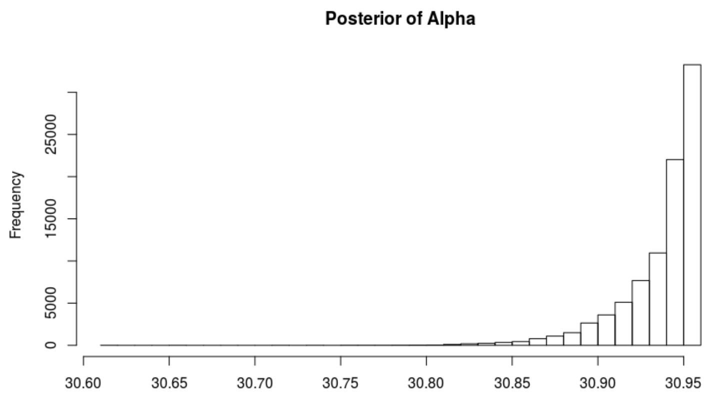
ii. simulate a value u which is distributed Uniform$(0, 1)$

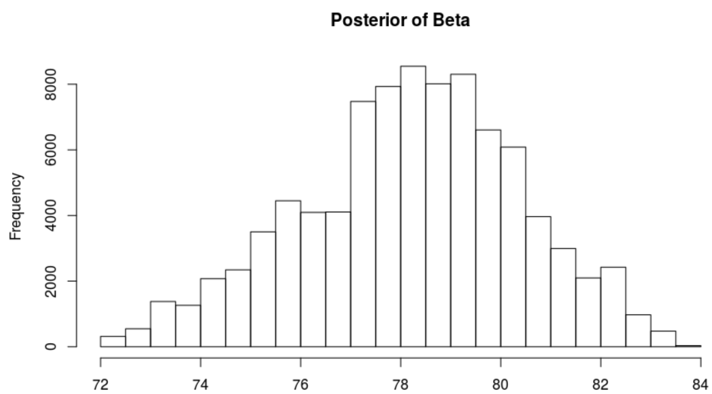
iii. propose a new value of α, $\alpha^\star$, which is distributed $\mathcal{N}(\alpha, .2)$ and β, $\beta^\star$, which is distributed $\mathcal{N}(\beta, .2)$

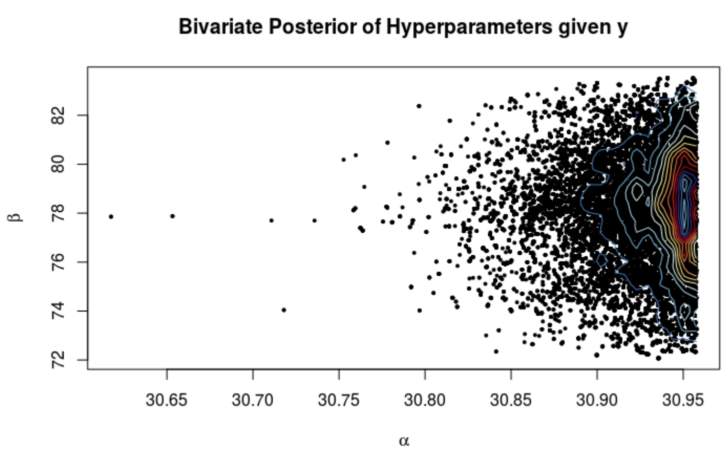
iv. accept $\alpha^\star$ and $\beta^\star$ as the new α and β if $\alpha^\star > 0$, $\beta^\star > 0$, and $log(u) < log(p(\alpha^\star, \beta^\star|\mathbf{y})) - log(p(\alpha, \beta|\mathbf{y}))$

v. repeat steps ii-iv 1000000 times recording each new iteration of parameters $\boldsymbol{\theta}, \alpha, \beta$

(f) Implement the algorithm of the previous part, and provide the same univariate and bivariate posteriors that were produced in part (c).







(g) How can you obtain samples from $p(\theta_i|\mathbf{y})$, based on samples from $p(\alpha, \beta|\mathbf{y})$?

You can sample from $p(\theta_i|\mathbf{y})$ by using a the samples directly from the smaples of $p(\alpha, \beta|\mathbf{y})$.

Code appendix

```
rm(list=ls())
library(dplyr)
library(ggplot2)
library(MASS)
library(RColorBrewer)
```

```r
Y <- c(5, 1, 5, 14, 3, 19, 1, 1, 4, 22)
E <- c(94.3, 15.7, 62.9, 126, 5.24, 31.4, 1.05, 1.05, 2.1, 10.5)
N <- length(Y)

nchain <- 100000
burnin <- 10000

proposalfunction <- function(param){
  return(rnorm(1, mean=param, sd=.2))
}

lambda.post <- matrix(1, nrow=N, ncol=nchain)
b0.post <- rep(1, nchain)
a0.post <- rep(2, nchain)
a0 <- 6
b0 <- 1

for(i in 2:nchain){
  lambda.post[,i] <- lambda <- rgamma(N, a0 + Y, b0 + E)
  b0.post[i] <- b0 <- rgamma(1, N * a0 + ha, sum(lambda) + hb)

  astar <- proposalfunction(a0)
  # generate a probability of accepting that is g(p*)/g(pi)
  paccept <- prod(lambda)^(astar-a0) * b0^(N * (astar-a0)) *
    (gamma(astar)/gamma(a0))^-N * exp(-astar+a0)
  if (astar > 0 & runif(1) < paccept){
    a0 <- astar
  }
  a0.post[i] <- a0
}

plot(a0.post, type="l")
plot(b0.post, type="l")

for(i in 1:N){
  title_ <- paste0("Posterior Density for y_", i)
  plot(density(lambda.post[i, burnin:nchain]), main=title_)
}

plot(density(b0.post[burnin:nchain]), main="Posterior of Beta")
plot(density(a0.post[burnin:nchain]), main="Posterior of Alpha")
hist(b0.post[burnin:nchain], nclass=30, main="Posterior of Beta", xlab="")
hist(a0.post[burnin:nchain], nclass=30, main="Posterior of Alpha", xlab="")
```

```r
k <- 11
my.cols <- rev(brewer.pal(k, "RdYlBu"))

z <- kde2d(a0.post[burnin:nchain], b0.post[burnin:nchain], n=50)

plot(a0.post[burnin:nchain], b0.post[burnin:nchain],
     xlab=expression(alpha), ylab=expression(beta), pch=19, cex=.4,
     main="Bivariate Posterior of Hyperparameters")
contour(z, drawlabels=FALSE, nlevels=k, col=my.cols, add=TRUE)


# sample the joint posterior of alpha and beta
b0joint.post <- rep(1, nchain)
a0joint.post <- rep(2, nchain)
a0 <- 1
b0 <- 1

posterior <- function(a, b, theta=lambda.post[,N]){
   20 * a*log(b) - log(b^.9) + (-b*(1 + sum(theta)) - a) +
      log(gamma(a)^(-10)) + log(prod(theta^(a-1)))
}

for(i in 2:nchain){
   astar <- proposalfunction(a0)
   bstar <- proposalfunction(b0)
   # generate a probability of accepting that is g(p*)/g(pi)
   paccept <- posterior(astar, bstar, theta=lambda.post[,i]) -
      posterior(a0, b0, theta=lambda.post[,i])
   if (astar > 0 & bstar > 0 & log(runif(1)) < paccept){
      a0 <- astar
      b0 <- bstar
   }
   a0joint.post[i] <- a0
   b0joint.post[i] <- b0
}

hist(b0joint.post[burnin:nchain], nclass=30,
     main="Posterior of Beta Given y", xlab="")
hist(a0joint.post[burnin:nchain], nclass=30,
     main="Posterior of Alpha Given y", xlab="")

z <- kde2d(a0joint.post[burnin:nchain], b0joint.post[burnin:nchain], n=50)

plot(a0joint.post[burnin:nchain], b0joint.post[burnin:nchain],
     xlab=expression(alpha), ylab=expression(beta), pch=19, cex=.4,
```

```
    main="Bivariate Posterior of Hyperparameters given y")
contour(z, drawlabels=FALSE, nlevels=k, col=my.cols, add=TRUE)
```