CPMA 573 — Homework #5

Exercise 1: Normal model inference. Let $Y_1, \ldots, Y_{43}$ be iid normal random variables from the density

$$f(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\} .$$

Realizations from this data model are loçated at

`www.mathcs.duq.edu/~kern/hw5.dat`

It is your goal to make inference on the values of μ and σ^2 used to generate these data.

a. Assuming μ and σ^2 are a priori independent, with prior densities $\pi(\mu) \propto 1$ and $\pi(\sigma^2) \propto \sigma^{-2}$, obtain 25000 draws from the marginal posteriors $\pi(\mu|\vec{y})$ and $\pi(\sigma^2|\vec{y})$ using

Method 1: The Gibbs sampler. (Be sure to check for zero autocorrelation using 'acf' plots.)

Method 2: Independent draws directly from the (theoretical) marginal distribution of σ^2 in conjunction with the full conditional distribution for μ.

For both methods, provide trace plots ('ts.plot') and histograms of your μ and σ^2 realizations. Just for kicks, superpose the theoretical marginal density of σ^2 on both histograms of σ^2 realizations. The four plots for Method 1 can be produced in **R** as follows:

```
par(mfrow=c(2,2)) #Splits the plotting window into two rows and two columns
ts.plot(mu1,xlab=''Iteration'') #mu1 represents Method 1 realizations of mu
ts.plot(sigsq1,xlab=''Iteration'') #sigsq1 as with mu1
hist(mu1,probability=T)
hist(sigsq1,probability=T)
lines(yy, IGdens(yy)) #Here yy is a vector, and IGdens is your own
                      #inverse gamma density function
```

b. Use the `quantile` function in **R** in conjunction with your posterior draws to find the values (b_1, b_2) and (c_1, c_2) that satisfy the following posterior probabilities:

- $\Pr(b_1 < \mu < b_2) = k$ [with $\Pr(\mu < b_1) = (1-k)/2$]
- $\Pr(c_1 < \sigma^2 < c_2) = k$ [with $\Pr(\sigma^2 < c_1) = (1-k)/2$]

for $k = \{0.95, 0.99\}$.

c. Provide your estimates of μ and σ^2, along with corresponding 95% credible intervals. (Credible intervals are the Bayesian analog to confidence intervals. They are named differently because their interpretation is different.)

Exercise 2: Posterior predictive distribution. Use the 25000 (μ, σ^2) pairs generated in the previous problem—from either method—to generate 25000 predicted y-values. Based on these predicted y-values, answer the following:

a. What is the chance that the next (44th) observation is greater than 10?

b. What is the shortest interval that has a 95% chance of containing the next observation?

see which values of $\dfrac{y^* > 10}{\text{total } y^*}$

use quantile 2.5%, 97.5%

Use method 2