MCMC Demo 2

Exercise

Consider the mixture distribution described on p. 99 (Hoff). This distribution is a joint probability distribution of a discrete variable $\delta = \{1, 2, 3\}$, denoting which mixture component the mass comes from and a continuous variable θ. The target density is $\{Pr(\delta = 1), Pr(\delta = 2), Pr(\delta = 3)\} = (.45, .10, .45)$ and $p(\theta|\delta = i) \sim N(\theta; \mu_i, \sigma_i^2)$ where $\{\mu_1, \mu_2, \mu_3\} = (-3, 0, 3)$ and $\sigma_i^2 = 1/3$ for $i \in \{1, 2, 3\}$.

1. Generate 1000 samples of θ from this distribution using a Monte Carlo procedure. (Hint: first generate $\delta^{(i)}$ from the marginal distribution $p(\delta)$ and then generate $\theta^{(i)}$ from $p(\theta|\delta)$.) Plot your samples in a histogram form and superimpose a curve of the density function. Comment on your samples, do they closely match the true distribution?

```
num_sims <- 1000
delta <- theta <- rep(0, num_sims)
mu_vals <- c(-3,0,3)
sigma_vals <- rep(sqrt(1/3),3)

for (iter in 1:num_sims){
  delta[iter] <- sample(1:3, size = 1, prob = c(.45,.1,.45))

  # Now sample theta | delta
}
```

2. Next, generate samples from a Gibbs sampler using the full conditional distributions of θ and δ. You already know the form of the full conditional for θ from above. The full conditional distribution for δ is given below:

$$Pr(\delta = d|\theta) = \frac{Pr(\delta = d) \times p(\theta|\delta = d)}{\sum_{d=1}^{3} Pr(\delta = d) \times p(\theta|\delta = d)}$$

Hint: for $p(\theta|\delta = d)$ evaluate θ from a normal distribution with parameters $\{\mu_d, \sigma_d^2\}$. Intialize θ at 0.

a. Generate 100 samples using this procedure. Plot your samples as a histogram with the true density superimposed on the plot. Also include a plot of your θ value on the y-axis and the iteration number on the x-axis. This is called a trace plot, and allows your to visualize the movement of your MCMC *particle*. Comment on how close your samples match the true density. What does the trace plot reveal about the position of θ over time (the iterations)? Does the proportion of the time the sample spends in each state (δ) match the true probabilities?

```
num_sims <- 100
delta <- rep(2, num_sims)
theta <- rep(0, num_sims)

for (iter in 2:num_sims){
  delta_probs <- c(.45 *  dnorm(theta[iter-1], mean = mu_vals[1], sd  = sigma_vals[1]),
                   .10 *  dnorm(theta[iter-1], mean = mu_vals[2], sd  = sigma_vals[2]),
                   .45 *  dnorm(theta[iter-1], mean = mu_vals[3], sd  = sigma_vals[3]))
  delta_probs <- delta_probs / sum(delta_probs)
  delta[iter] <- sample(1:3, size = 1, prob = delta_probs)

  # Now sample theta | delta
}
```

b. Repeat for 1000 samples.
c. Repeat for 10000 samples.