# ST495/590 Assignment 2 - Solutions

(1) *Write a function that uses Monte Carlo sampling to estimate the posterior mean and standard deviation of  given we observe $Y = y$.*

```
M <- 1000000
MCMC <- function(y,n,a,b){
  theta <- rbeta(M,a,b) # draw theta from beta(a,b)
  Y <- rbinom(M,n,theta) # draw Y|theta from binom(n,theta)

  # extract the index positions of the samples with Y = y.
  inx <- which(Y==y)
  # get those theta's.
  newtheta <- theta[inx]

  # get the mean and standard deviation of the newtheta.
  mn <- mean(newtheta)
  std <- sd(newtheta)

  return(list(mn=mn, std=std)) # wrap up the output in a list and return it.
}
```

(2) $n = 10$, $a = b = 1$

(3) $n = 10$, $a = b = 10$

(4) *Comment on the differences between the plots with $a = b = 1$ versus $a = b = 10$.*

As shown in the figure, the range of the posterior mean for $a = b = 1$ is $(0.08, 0.92)$, much larger than the range for $a = b = 10$, $(0.33, 0.66)$. Taking $y = 10$ as an example, the posterior mean is 0.92 when $a = b = 1$, while it's 0.66 when $a = b = 10$. $y = 10$ means in the 10 trials, we observe 10 successes. If we only have the observed data, we might conclude that the probability of the success in one trail would be exactly 1. If we only have prior distributions, the mean for a Beta$(a, b)$ distribution is $a/(a + b)$, so in both cases, the prior mean is 0.5. However, the variance is $ab/[(a + b)^2(a + b + 1)]$. So for $a = b = 1$, the variance is 0.083; while for $a = b = 10$, the variance is 0.012. This means $a = b = 10$ has smaller variance than $a = b = 1$, thus stronger effect.

For $a = b = 1$ case, when the prior effect is very small, and the data is dominating, so the estimated $\theta$ is close to the estimate got from data, i.e., 1. As $a$ and $b$ go larger, the prior will play a more important role, and the data or the observations will be less dominating, that's why the estimated $\theta$ is 0.66.

To explain it in more detail, as $\theta \sim \text{Beta}(a, b)$, and $Y|\theta \sim \text{Binomial}(n, \theta)$, the posterior $\theta|Y \sim \text{Beta}(a+Y, b+n-Y)$, and the posterior mean is $\hat{\theta} = E(\theta|Y) = \frac{Y+a}{n+a+b}$. When $Y = 10$, for $a = b = 1$, $\hat{\theta} = \frac{10+1}{10+1+1} = 0.92$; for $a = b = 10$, $\hat{\theta} = \frac{10+10}{10+10+10} = 0.67$, which corresponds with what we got from MCMC sampling.