# HW 2

*Nicholas Marey*

*September 12, 2017*

## 1.

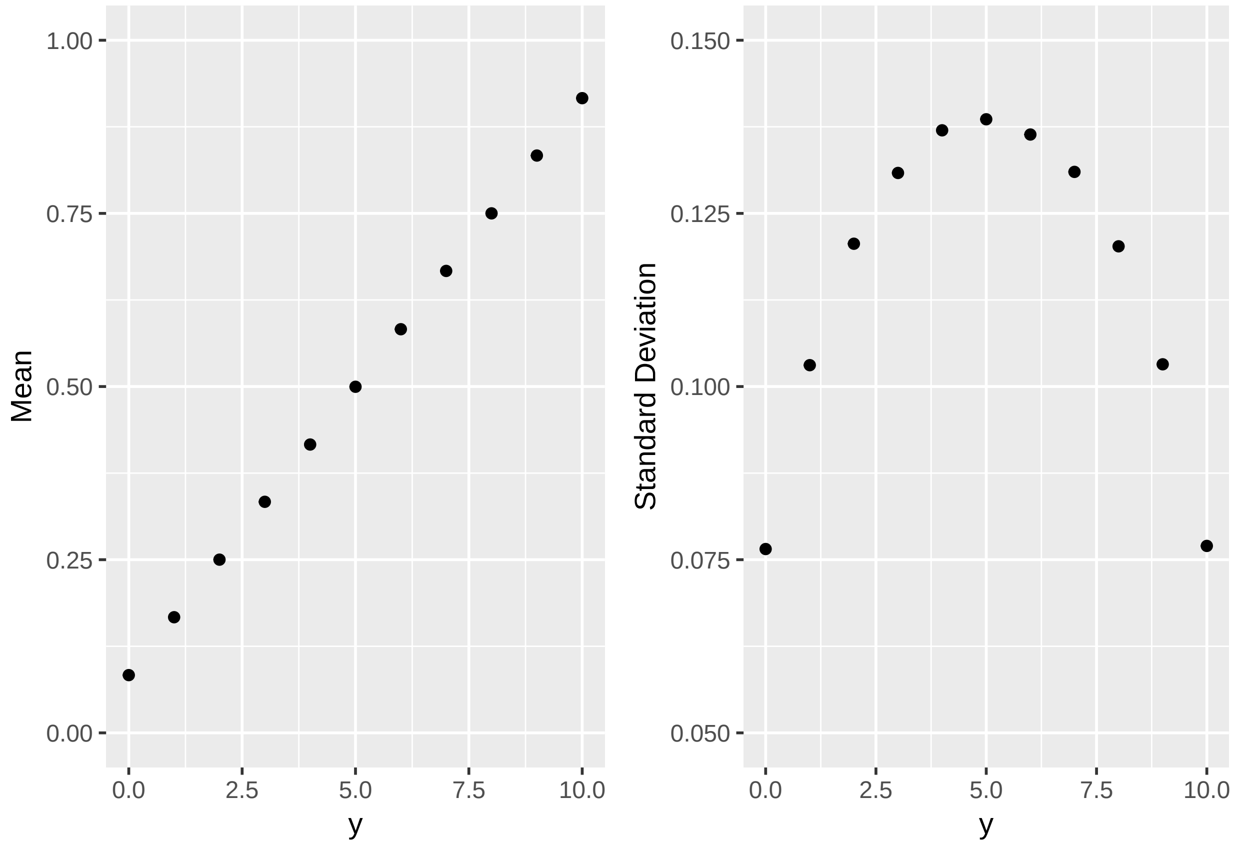
**Assume the Bayesian model with likelihood $Y|\theta \; Binomial(n, \theta)$ and prior $\theta \; Beta(a, b)$. Write a function that uses Monte Carlo sampling to estimate the posterior mean and standard deviation of $\theta$¸ given we observe $Y = y$. The function should take inputs $y, n, a, andb$. Given these inputs, the function should generate 1, 000, 000 samples of $(\theta, Y)$ (by first drawing $\theta$ from a beta distribution and then $Y|\theta$ from a binomial distribution), extract the samples with $Y = y$, and return the mean and standard deviation of $\theta$ for these samples. Include code for this function in your write-up.**

```
montecarlo <- function(y, n, a, b) {
    theta <- rbeta(1e+06, a, b)
    like <- rbinom(1e+06, n, theta)
    post <- rbeta(1e+06, like + a, n - like + b)  #posterior probability distribution
    post <- as.data.frame(cbind(post, like))
    avg <- as.data.frame(tapply(post$post, post$like, mean))
    sd <- as.data.frame(tapply(post$post, post$like, sd))
    values <- unique(post$like)
    values <- values[order(values)]
    dat <- cbind(values, avg, sd)
    colnames(dat) <- c("k", "mean", "sd")
    dat1 <- dat[dat$k %in% y, ]
    colnames(dat1) <- c("y", "Mean", "Standard Deviation")
    return(dat1)
}
```

## 2.

**Use the code from (1) with $n = 10$ and $a = b = 1$ to compute the posterior mean and standard deviation for $\theta$, for all $y = 0, 1, ..., n$ and plot the posterior mean and standard deviation as a function of $y$.**

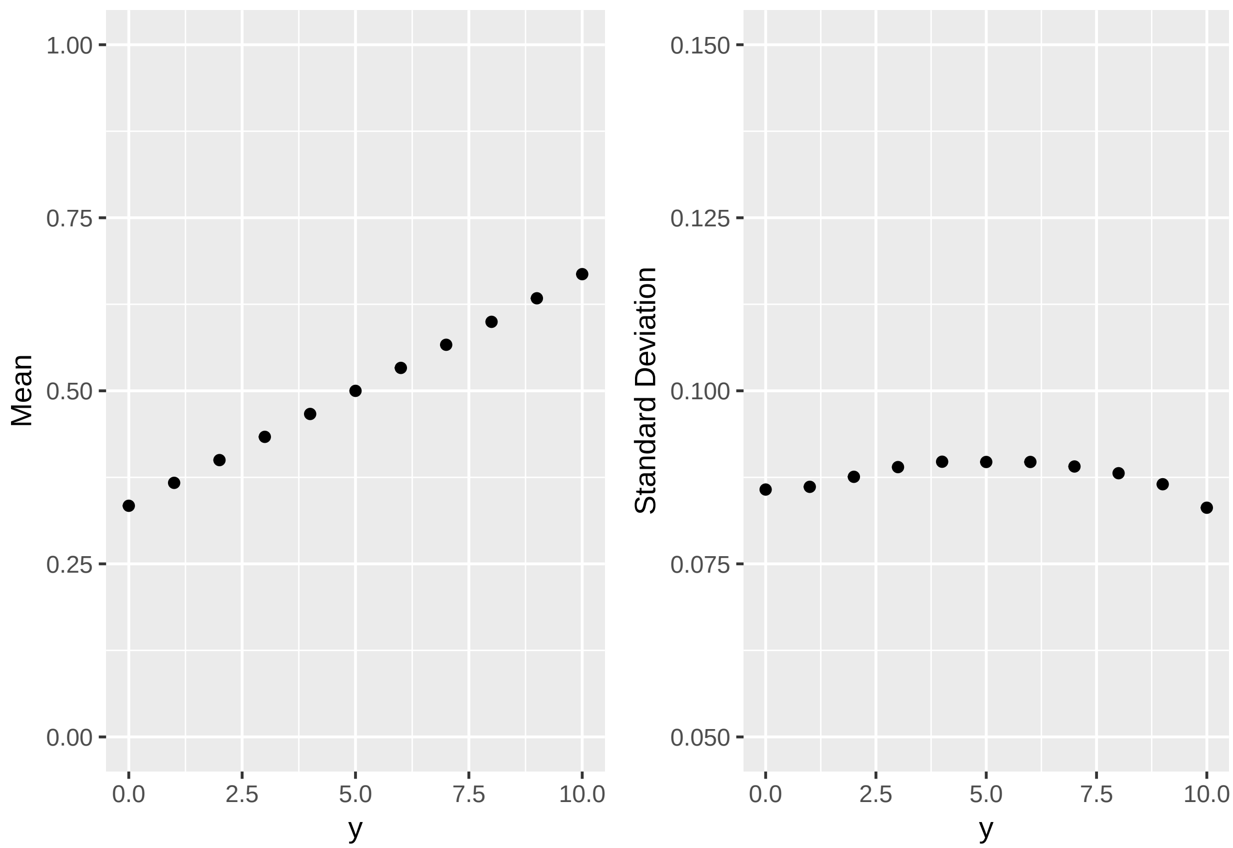
```
library(ggplot2)
library(gridExtra)
set.seed(1)
p2 <- montecarlo(0:10, 10, 1, 1)
plot1 <- ggplot(p2, aes(y, Mean)) + geom_point() + scale_y_continuous(limits = c(0,
    1))
plot2 <- ggplot(p2, aes(y, `Standard Deviation`)) + geom_point() + scale_y_continuous(limits = c(0.05,
    0.15))
grid.arrange(plot1, plot2, ncol = 2)
```




### 3.

**Use the code from (1) with $n = 10$ and $a = b = 10$ to compute the posterior mean and standarddeviation for $\theta$ for all $y = 0, 1, ..., n$ and plot the posterior mean and standard deviation as a function of $y$.**

```
set.seed(1)
p3 <- montecarlo(0:10, 10, 10, 10)
plot3 <- ggplot(p3, aes(y, Mean)) + geom_point() + scale_y_continuous(limits = c(0,
    1))
plot4 <- ggplot(p3, aes(y, `Standard Deviation`)) + geom_point() + scale_y_continuous(limits = c(0.05,
    0.15))
grid.arrange(plot3, plot4, ncol = 2)
```




## 4.

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

When both a and b equal 1, the mean of theta varies greatly as y increases from 0 to 10. This can be seen in question two, where the mean of theta when y equals 0 is approximately 0.08, while the mean of theta when y equals 10 is approximately 0.92. Meanwhile, when a and b both equal 10, the mean of theta starts to flatten. When this situation is present, the mean of theta only varies between approximately 0.33 and 0.67.

Additionally, when a and b both equal 1, the standard deviation also varies greatly, with theta ranging from 0.077 to 0.139. When a and b both equal 10, the range for the standard deviation of theta has shrunk to between 0.0831 and 0.0897.

This means that as a and b increased from 1 to 10, the distribution took a much tighter form. When both a and b equaled 1, the values of theta were wider resulting in a greater standard deviation as compared to when a and b both equaled 10. When a and b both were 10, the posterior distribution had a higher peak and the values of theta do not vary that much. This results in smaller standard deviations for theta. Furthermore, not only did the standard deviations of theta tighten up as a and b increased, but so did the means of theta. As a and b increased to 10, the means of theta became closer to 0.5.