HW 2

Exercise 2.21

Data processing

```
pew <- foreign::read.dta("http://www.stat.columbia.edu/~gelman/book/data/pew_research_center_june_elect
Obama <- read.csv("http://www.stat.columbia.edu/~gelman/book/data/2008ElectionResult.csv")
```

```
library(usdata)
library(dplyr)

levels(pew$state)[9] <- 'District of Columbia'
pew$state <- state2abbr(pew$state)
pew <- pew[!is.na(pew$state), ]

exclude_states <- c('AK', 'HI', 'DC')

Obama$state <- state2abbr(Obama$state)
Obama <- Obama %>% filter(!state %in% exclude_states) %>%
  select(state, vote_Obama_pct) %>%
  mutate(vote_Obama_pct = vote_Obama_pct/100)
```

Calculate "very liberal adults" data in 47 states

```
res0 <- pew %>% group_by(state) %>%
  mutate(n = n()) %>%
  filter(!state %in% exclude_states,
                    age >= 18,
                    ideo == 'very liberal') %>%
  # Every state has at least 1 sample that is adult and...
  # ...very liberal so I will not fix it for the n=0 group drop issue.
  mutate(target = n(), target_ratio = target/n) %>%
  select(state, target, target_ratio, n) %>%
  unique()
```

```
library(ggplot2)

plt_data <- merge(res0, Obama)
plt1 <- ggplot(plt_data) +
   geom_text(aes(y=target_ratio, x=vote_Obama_pct, label=state), size = 2) +
   ggtitle('Target pct vs Obama vote pct') +
   theme(plot.title = element_text(size=10))
```

Posterior mean

Denote the target number (not the ratio) by y_j and the total respondent by n_j for state j. Assuming that $y_j \sim \text{Binomial}(n_j, \theta_j)$ with prior $\theta_j \sim \text{Beta}(\alpha, \beta)$. Notice that the prior is set for all $theta_j$.

To get a reasonable initial value of α and β, we use the method in Section 2.7. First notice that the predictive distribution is (for simplicity y_j is denoted by y)

$$
\begin{aligned}
p(y) &= \int p(y|\theta)p(\theta)d\theta && (1)\\
&= C\int \binom{n}{y}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}d\theta && (2)\\
&= C\binom{n}{y}\mathrm{B}(y+\alpha, n-y+\beta)\int \frac{1}{\mathrm{B}(y+\alpha, n-y+\beta)}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}d\theta && (3)\\
&= C\binom{n}{y}\mathrm{B}(y+\alpha, n-y+\beta), && (4)
\end{aligned}
$$

where C is a constant of y. This shows us that the predictive distribution is Beta-binomial(α, β, n). We use method of moments to estimate α and β. We have the formula as follows,

$$
\begin{aligned}
\hat{\alpha} &= \frac{nm_1 - m_2}{n(m_2/m_1 - m_1 - 1) + m_1} && (5)\\
\hat{\beta} &= \frac{(n-m_1)(n-m_2/m_1)}{n(m_2/m_1 - m_1 - 1) + m_1}, && (6)
\end{aligned}
$$

where m_1 and m_2 are the sample first and second moments and n is sample average of number of respondents in a state.

```r
n <- mean(res0$n)
m1 <- mean(res0$target)
m2 <- mean(res0$target^2)

alpha_hat <-  (n*m1 - m2)/(n*(m2/m1 - m1 - 1) + m1)
beta_hat <- (n - m1)*(n - m2/m1)/(n*(m2/m1 - m1 - 1) + m1)

alpha_hat
```

```
## [1] 0.8992001
```

```r
beta_hat
```

```
## [1] 18.23907
```

We are not going to call them $\hat{\alpha}$ and $\hat{\beta}$, instead we use these values as fixed hyperparameter for the prior.

The posterior is

$$
\begin{aligned}
p(\theta_j|y_j) &\propto p(y_j|\theta_j)p(\theta_j) && (7)\\
&\propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}, && (8)
\end{aligned}
$$

which indicates that $\theta_j|y_j \sim \text{Beta}(\alpha+y_j, \beta+n_j-y_j)$ and the posterior mean is nothing but $(\alpha+y_j)/(\alpha+\beta+n_j)$. The hyperparameters $\alpha = 0.8992001$ and $\beta = 18.23907$.

```r
res1 <- res0 %>% mutate(post_mean = (alpha_hat + target)/(alpha_hat + beta_hat + n))

res1
```

```
## # A tibble: 48 x 5
## # Groups:   state [48]
##    state target target_ratio     n post_mean
##    <chr>  <int>        <dbl> <int>     <dbl>
##  1 MA        46       0.0672   685    0.0666
##  2 RI         6       0.0458   131    0.0460
```

```
##   3 ME          9       0.0584   154    0.0572
##   4 VT          8       0.0696   115    0.0663
##   5 CT         14       0.0354   395    0.0360
##   6 NJ         34       0.0391   870    0.0393
##   7 NY        104       0.0611  1701    0.0610
##   8 PA         54       0.0339  1591    0.0341
##   9 DE          7       0.0588   119    0.0572
##  10 MD         24       0.0405   593    0.0407
## # ... with 38 more rows
```

```
plt_data <- merge(res1, Obama)
plt2 <- ggplot(plt_data) +
   geom_text(aes(y=post_mean, x=vote_Obama_pct, label=state), size = 2) +
   ggtitle('Post mean vs. Obama vote pct') + theme(plot.title = element_text(size=10))
```

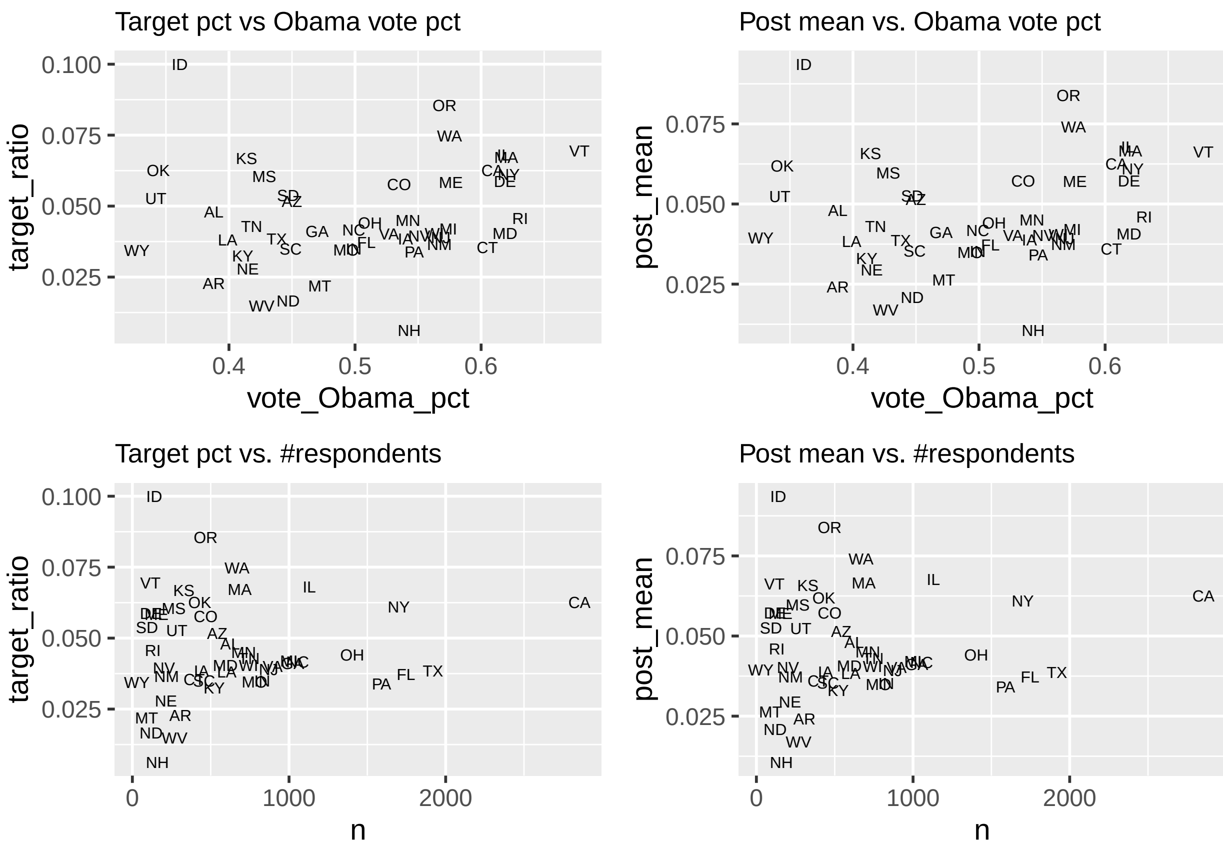
Plot in a single page

I don't know if I understand it correctly but to me the second part is drawing graphs using previous result against the number of respondents.

```
plt3 <- ggplot(plt_data) + geom_text(aes(y=target_ratio, x=n, label=state), size = 2) +
   ggtitle('Target pct vs. #respondents') +
   theme(plot.title = element_text(size=10))
```

```
plt4 <- ggplot(plt_data) + geom_text(aes(y=post_mean, x=n, label=state), size = 2) +
   ggtitle('Post mean vs. #respondents') +
   theme(plot.title = element_text(size=10))
```

```
library(gridExtra)
grid.arrange(plt1, plt2, plt3, plt4, ncol=2)
```

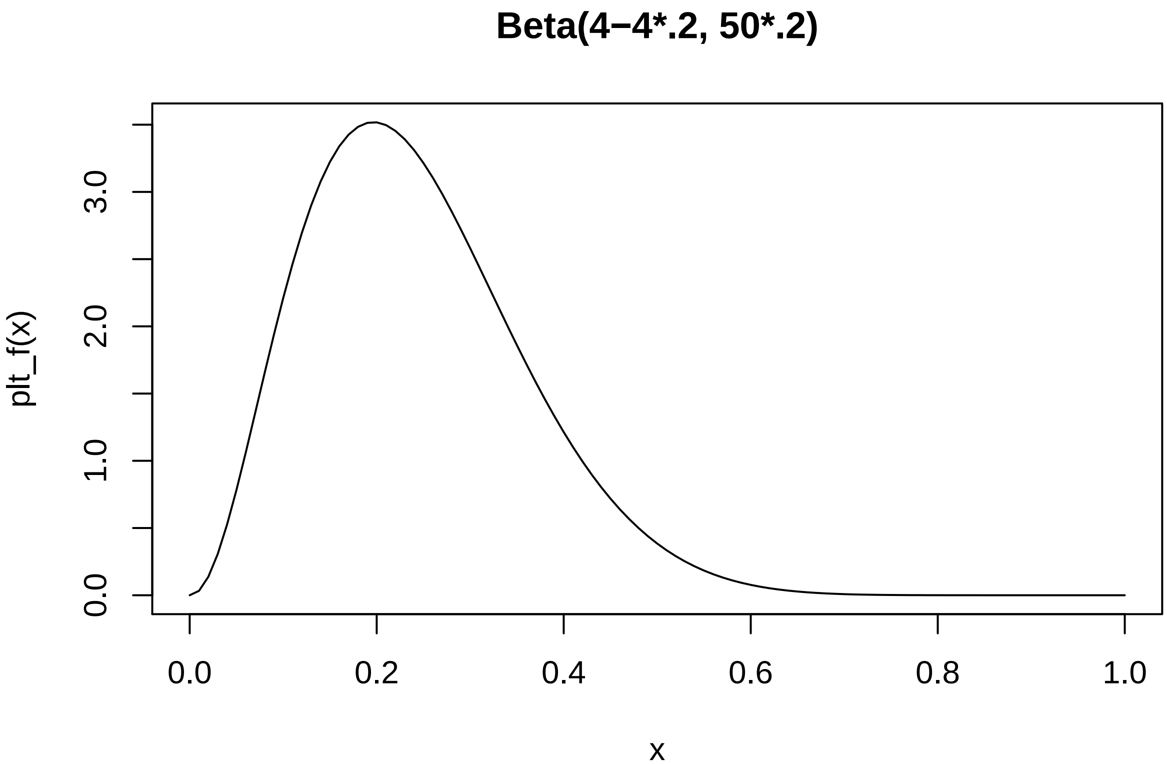


Exercise 2.22

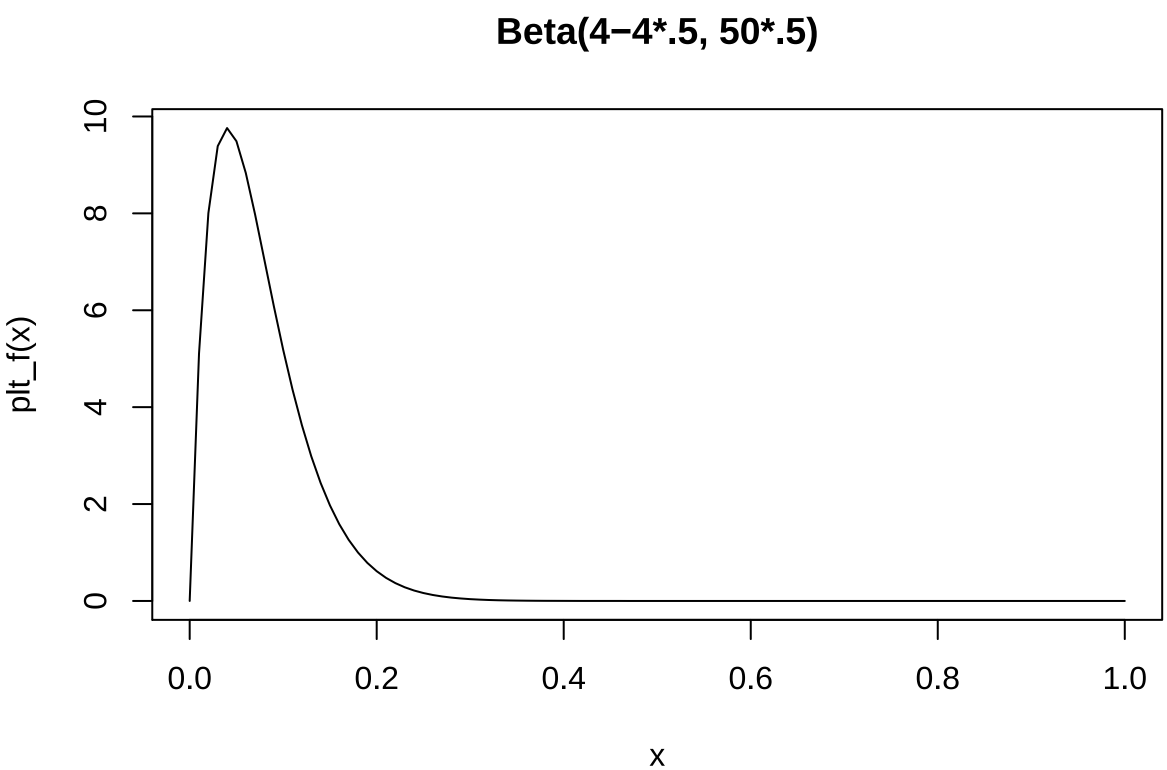
- Noninformative prior: $U(-1,1)$ is absolutely noninformative since it assigns equal prob. across support of θ.
- Subjective prior: $\text{Beta}(4-4p, 50p)$ where p is success prob. before training. By applying this prior we believe that the training improvement is not going to be negative and will be larger if you were bad at shooting before training.
- Weakly informative prior: $u(-.05, .2)$ since this prior gives information that the improvement are highly likely to be positive but the number will not be that large.

Below are two graphs to illustrate the subjective prior.

```
plt_f <- function(x) {dbeta(x, 4-4*.2, 50*.2)}
curve(plt_f, main="Beta(4-4*.2, 50*.2)")
```




```r
plt_f <- function(x) {dbeta(x, 4-4*.5, 50*.5)}
curve(plt_f, main="Beta(4-4*.5, 50*.5)")
```



Exercise 3.13

Starting from

$$p(\mu|y,\Sigma) \propto \exp\left(-\frac{1}{2}\Big((\mu-\mu_0)^T\Lambda_0^{-1}(\mu-\mu_0) + \sum_{i=1}^{n}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu)\Big)\right),$$

we have

$$\begin{aligned}
p(\mu|y,\Sigma) &\propto \exp\left(-\frac{1}{2}\Big(\mu^T\Lambda_0^{-1}\mu - 2\mu^T\Lambda_0^{-1}\mu_0 + C_1 + C_2 + \sum_{i=1}^{n}\mu^T\Sigma^{-1}\mu + \sum_{i=1}^{n}\mu^T\Sigma^{-1}y_i\Big)\right) && (9)\\
&\propto \exp\left(-\frac{1}{2}\Big(\mu^T(\Lambda_0^{-1}+n\Sigma^{-1})\mu - 2\mu^T(\Lambda_0^{-1}\mu_0 + \Sigma^{-1}\sum_{i=1}^{n}y_i)\Big)\right) && (10)\\
&\propto \exp\Big(-\frac{1}{2}\Big(\mu^T - (\Lambda_0^{-1}+n\Sigma^{-1})^{-1}(\Lambda_0^{-1}\mu_0 + \Sigma^{-1}\sum_{i=1}^{n}y_i)\Big)^T && (11)\\
&\quad (\Lambda_0^{-1}+n\Sigma^{-1})\Big(\mu^T - (\Lambda_0^{-1}+n\Sigma^{-1})^{-1}(\Lambda_0^{-1}\mu_0 + \Sigma^{-1}\sum_{i=1}^{n}y_i)\Big)\Big) && (12)\\
&\propto \exp\left(-\frac{1}{2}(\mu-\mu_n)^T\Lambda_n^{-1}(\mu-\mu_n)\right), && (13)
\end{aligned}$$

where $C_1 = -\mu_0^T\Lambda_0^{-1}\mu_0$ and $C_2 = \sum_{i=1}^{n} y_i^T\Sigma^{-1}y_i$ are constants of μ and are presented for explicity. Identity $\mu^T\Lambda_0^{-1}\mu_0 = tr(\mu^T\Lambda_0^{-1}\mu_0) = tr(\mu_0^T\Lambda_0^{-1}\mu) = \mu_0^T\Lambda_0^{-1}\mu$ is used in the first line.

Exercise 3.6

(a)

The prior is improper because the integral

$$\int p(\lambda,\theta) = \int_\theta\int_\lambda \lambda^{-1}$$

does not integrate to a finite number.

To derive $p(N,\theta)$, we have

$$\begin{aligned}
p(N,\theta) &= \int p(N|\theta,\lambda)p(\theta,\lambda)d\lambda && (14)\\
&= \int_0^\infty \frac{(\lambda/\theta)^N\exp(-\lambda/\theta)}{N!}\lambda^{-1}d\lambda && (15)\\
&= \int_0^\infty \frac{(\lambda/\theta)^N\exp(-\lambda/\theta)}{N!}(\lambda/\theta)^{-1}\frac{1}{\theta}d\lambda && (16)\\
&= \int_0^\infty \frac{(\lambda/\theta)^{N-1}\exp(-\lambda/\theta)}{N!}d(\lambda/\theta) && (17)\\
&= 1 && (18)
\end{aligned}$$

the last identity is because we form the pdf of Gamma$(N,1)$ so the integral is one.

It seems like the we choose prior $p(\lambda,\theta)=\lambda^{-1}$ because it gives this nice noninformative prior on (N,θ).

(b)

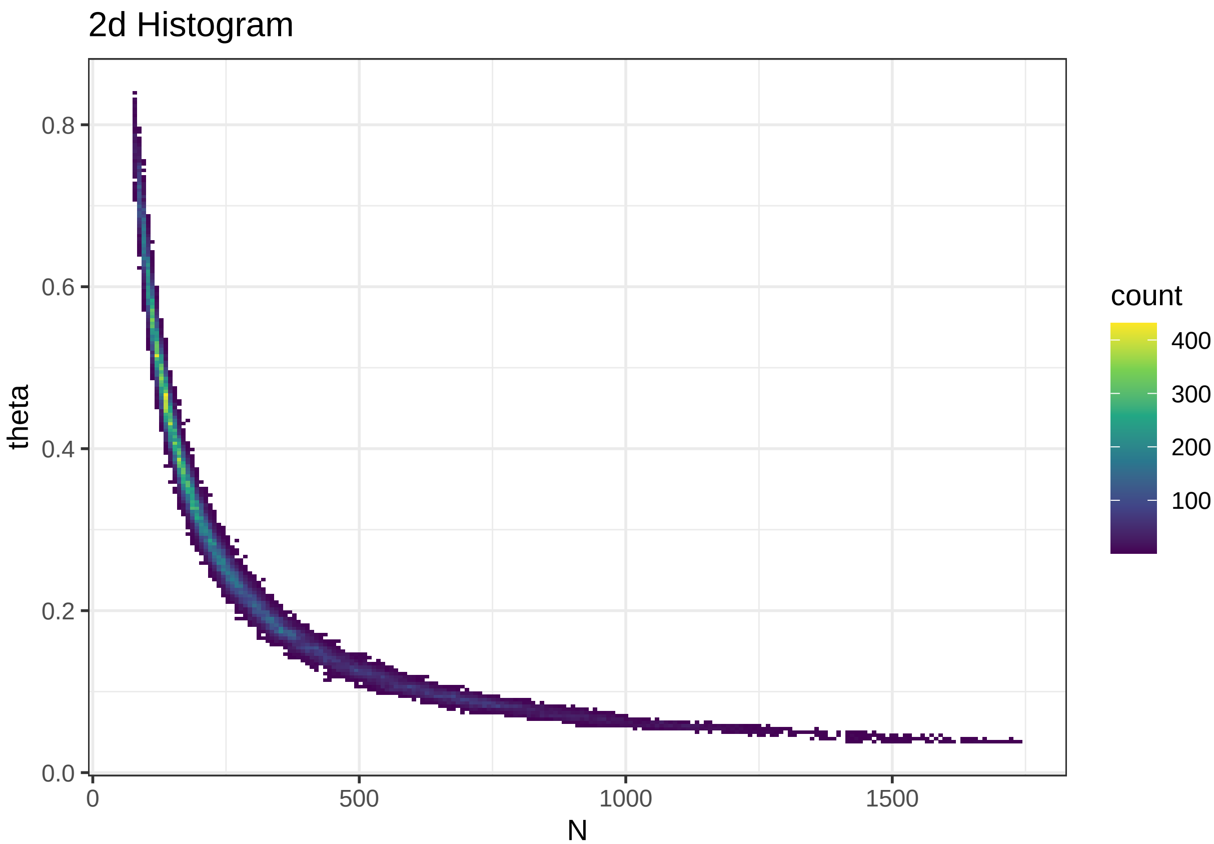
Thanks to the simple prior, the posterior is just

$$p(N, \theta|\overline{y}) \propto p(\overline{y}|N, \theta)p(N, \theta) \tag{19}$$

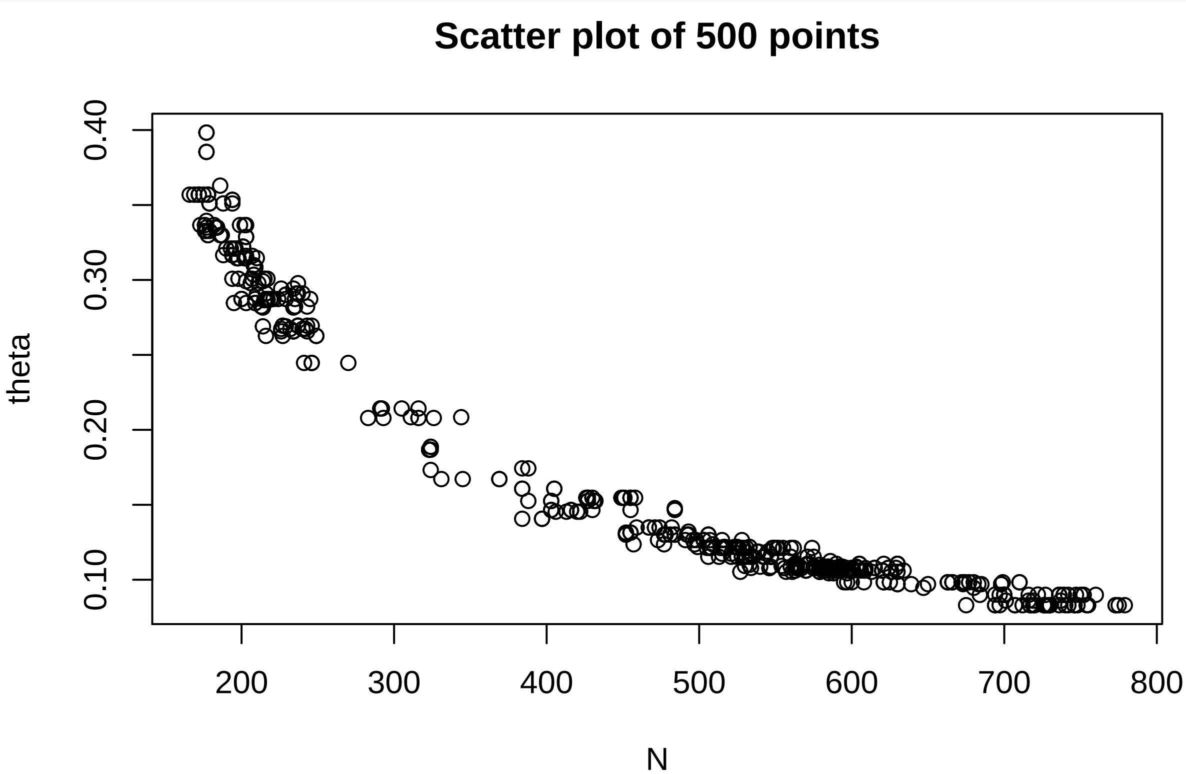
$$= \prod_{i=1}^{5} \binom{N}{y_i} \theta^{y_i}(1-\theta)^{N-y_i}. \tag{20}$$

```r
y <- c(53, 57, 66, 67, 72)
joint_density <- function(N, theta, y) {
  # y can be a vector
  prod(choose(N, y)*(theta^y)*(1-theta)^(N-y))
}

iter <- 1e5
res <- matrix(nrow=iter+1, ncol=2)
res[1, ] <- c(2*max(y), 1/2) # init
for (i in 1:iter) {
  temp_N <- rpois(1, lambda=res[i, 1]) # proposal
  log_ratio <- log(joint_density(temp_N, res[i, 2], y)) -
    log(joint_density(res[i, 1], res[i, 2], y)) +
    log(dpois(temp_N, lambda=res[i, 1])) -
    log(dpois(res[i, 1], lambda=temp_N))
  temp_log_U <- log(runif(1))
  if (temp_log_U <= log_ratio) {
    res[i+1, 1] <- temp_N
  } else {
    res[i+1, 1] <- res[i, 1]
  }

  temp_theta <- runif(1, min=max(0, res[i, 2]-.1), max=min(1, res[i,2]+.1)) # proposal
  log_ratio <- log(joint_density(res[i+1, 1], temp_theta, y)) -
    log(joint_density(res[i+1, 1], res[i, 2], y))
  # proposal ratio is always one so dropped
  temp_log_U <- log(runif(1))
  if (temp_log_U <= log_ratio) {
    res[i+1, 2] <- temp_theta
  } else {
    res[i+1, 2] <- res[i, 2]
  }
}
```

```r
library(ggplot2)
res1 <- res[1e4:1e5, ] # remove burn-in

plt_data <- as.data.frame(res1)
colnames(plt_data) <- c('N', 'theta')
ggplot(plt_data, aes(x=N, y=theta) ) +
  geom_bin2d(bins = 200) +
  scale_fill_continuous(type = "viridis") +
  theme_bw() + ggtitle('2d Histogram')
```




```r
plot(tail(plt_data$N, 500), tail(plt_data$theta, 500), xlab='N', ylab='theta', main='Scatter plot of 500
```



```r
print(paste('The posterior prob. of N>100 is', mean(res1[, 1] > 100)))
```

```
## [1] "The posterior prob. of N>100 is 0.931634092954523"
```

(c)

How do you choose the fix value of μ? It will not be a noninformative prior. Also, the posterior will be more complicated.