STA 360 HW 9

Ken Ye

# 1

a) $$P(z \mid y, X) = \frac{p(y \mid X, z)\, p(z \mid X)}{p(y \mid X)} \qquad z \in \{0,1\}^p = Z$$

$$= \frac{p(y \mid X, z)\, p(z)}{\sum_{z \in Z} p(y \mid X, z)\, p(z)}$$

$$= \frac{e^{l(z)}\, p(z)}{\sum_{z \in Z} e^{l(z)}\, p(z)}$$

b) Assume all models have equal prior probability

$\Rightarrow$ $p(z)$'s are equal for all $z \in Z$, $p(z) = \frac{1}{2^p}$

Therefore, $$p(z \mid y, X) = \frac{e^{l(z)}\, \cancel{p(z)}}{\sum_{z \in Z} e^{l(z)}\, \cancel{p(z)}}$$

$$= \frac{e^{l(z)}}{\sum_{z \in Z} e^{l(z)}}$$

# 2

$$z \in \{0,1\}^3$$

a) $P(z_1 \mid z_2, z_3, y, X) \propto p(y \mid X, z)\, p(z_1)$

$P(z_1 = 1 \mid z_2, z_3, y, X) \propto P(y \mid X, z_1 = 1, z_2, z_3)\, \pi_1$

$P(z_1 = 0 \mid z_2, z_3, y, X) \propto P(y \mid X, z_1 = 0, z_2, z_3)\,(1 - \pi_1)$

$\Rightarrow P(z_1 \mid z_2, z_3, y, X) \sim$ Bernoulli $(p)$.

where

$$p = \frac{\pi_1\, p(y \mid X, z_1 = 1, z_2, z_3)}{\pi_1\, p(y \mid X, z_1 = 1, z_2, z_3) + (1 - \pi_1)\, p(y \mid X, z_1 = 0, z_2, z_3)}$$

b) Assume $\pi_j = \frac{1}{2}$ for all $j$

$p(z_1 \mid z_2, z_3, y, X) \sim$ Bernoulli $(p')$

where

$$p' = \frac{\cancel{\tfrac{1}{2}}\, p(y \mid X, z_1 = 1, z_2, z_3)}{\cancel{\tfrac{1}{2}}\, p(y \mid X, z_1 = 1, z_2, z_3) + \cancel{\tfrac{1}{2}}\, p(y \mid X, z_1 = 0, z_2, z_3)}$$

$$= \frac{p(y \mid X, z_1 = 1, z_2, z_3)}{p(y \mid X, z_1 = 1, z_2, z_3) + p(y \mid X, z_1 = 0, z_2, z_3)}$$

#4

a) prior :

$$\mathbb{E}[\theta_j] = \frac{\alpha}{\alpha/\mu} = \mu$$

$$V[\theta_j] = \frac{\alpha}{(\alpha/\mu)^2} = \frac{\mu^2}{\alpha}$$

posterior :

$$p(\theta_j \mid \alpha, \mu, \underset{\sim}{y}_1 \cdots \underset{\sim}{y}_j) \underset{\theta_j}{\propto} p(\underset{\sim}{y}_j \mid \theta_j, \alpha, \mu)\, p(\theta_j \mid \alpha, \mu)$$

$$\propto \left\{ \prod_{i=1}^{n} \frac{\theta_j^{y_{ij}} e^{-\theta_j}}{\cancel{y_{ij}!}} \right\} \left\{ \theta_j^{\alpha-1} e^{-\frac{\alpha}{\mu}\theta_j} \right\}$$

$$= \theta_j^{\Sigma y_{ij}} e^{-n\theta_j}\, \theta_j^{\alpha-1} e^{-\frac{\alpha}{\mu}\theta_j}$$

$$= \theta^{\alpha+\Sigma y_{ij}-1} e^{-(\frac{\alpha}{\mu}+n)\theta_j}$$

$$\sim \text{Gamma}\left(\alpha + \sum_{i=1}^{n} y_{ij},\ \frac{\alpha}{\mu}+n\right)$$

$$\mathbb{E}[\theta_j \mid \alpha, \mu, \underset{\sim}{y}_1 \ldots \underset{\sim}{y}_j] = \frac{\alpha + \sum_{i=1}^{n} y_{ij}}{\frac{\alpha}{\mu}+n} = \frac{\frac{\alpha}{\mu}\mu + n\bar{y}_j}{\frac{\alpha}{\mu}+n} = \frac{\frac{\alpha}{\mu}}{\frac{\alpha}{\mu}+n}\mu + \frac{n}{\frac{\alpha}{\mu}+n}\bar{y}_j$$

$$V[\theta_j \mid \alpha, \mu, \underset{\sim}{y}_1 \ldots \underset{\sim}{y}_j] = \frac{\alpha + n\bar{y}_j}{(\frac{\alpha}{\mu}+n)^2}$$

The posterior expectation is a weighted avg. of prior mean and sample mean, the weights depend on $n$, the sample size. The posterior also involves the sample mean and sample size, compared to the prior.

m = 16 for this problem

b) $$p(\mu \mid \alpha, \theta, Y) \underset{\mu}{\propto} p(\theta \mid \alpha, \mu)\, p(\mu)$$

$$= \prod_{j=1}^{m} p(\theta_j \mid \mu, \alpha)\, p(\mu)$$

$$\underset{\mu}{\propto} \left\{ \prod_{j=1}^{m} \left(\frac{\alpha}{\mu}\right)^{\alpha} e^{-\frac{\alpha}{\mu}\theta_j} \right\} \left\{ \left(\frac{1}{\mu}\right)^{a-1} e^{-\frac{b}{\mu}} \right\}$$

$$\propto \left(\frac{1}{\mu}\right)^{m\alpha} e^{-\frac{\alpha \sum \theta_j}{\mu}} \left(\frac{1}{\mu}\right)^{a-1} e^{-\frac{b}{\mu}}$$

$$= \left(\frac{1}{\mu}\right)^{m\alpha + a - 1} e^{-\left(\frac{\alpha \sum \theta_j + b}{\mu}\right)}$$

$$\sim \text{inv-gam}\left(m\alpha + a,\ \alpha \sum_{j=1}^{m} \theta_j + b\right)$$

# STA360 Homework 9 (Ken Ye)

```r
library(latex2exp)
library(ggplot2)
library(MASS)
library(ggrepel)
library(mvtnorm)
set.seed(0)
```

# Question 3 (Book Exercise 9.2)

```r
# load diabetes data
diabetes <- read.table("azdiabetes.dat", header =  TRUE)
diabetes <- diabetes[,-8]
y <- diabetes[-1,2]
X <- diabetes[-1,-2]
y <- as.matrix(y)
X <- as.matrix(X)
y <- y - mean(y)
for (col in col(X)){
  X[, col] <- X[, col] - mean(X[, col])
}
n <- dim(X)[1]
p <- dim(X)[2]
```

# Part a

```
# MC

# priors
g <- length(y)
nu0 <- 2
s20 <- 1
S <- 1000

Hg <- (g / (g + 1)) * X %*% solve(t(X) %*% X) %*% t(X)
SSRg <- t(y) %*% (diag(1, nrow = n) - Hg) %*% y
s2 <- 1 / rgamma(S, (nu0 + n) / 2, (nu0 * s20 + SSRg) / 2)
Vb <- g * solve(t(X) %*% X) / (g + 1)
Eb <- Vb %*% t(X) %*% y
E <- matrix(rnorm(S*p,0 ,sqrt(s2)),S ,p)
beta <- t(t(E %*% chol(Vb)) + c(Eb))
```

```
# posterior confidence intervals
for (col in 1:ncol(beta)) {
  print(quantile(beta[, col], c(0.025, 0.975)))
}
```

```
##       2.5%      97.5% 
## -1.6523894  0.3591608 
##        2.5%       97.5% 
## -0.02570705  0.42447760 
##       2.5%      97.5% 
## -0.1385321  0.5060806 
##     2.5%    97.5% 
## 0.149078 1.165557 
##      2.5%     97.5% 
##  3.484193 17.869760 
##      2.5%     97.5% 
## 0.4443995 1.0598537 
```

The 95% posterior confidence intervals for “npreg”, “bp”, “skin”, “bmi”, “ped”, and “age” are (-1.6523894, 0.3591608), (-0.02570705, 0.42447760), (-0.1385321, 0.5060806), (0.149078, 1.165557), (3.484193, 17.869760), and (0.4443995, 1.0598537), respectively.

## Part b

```r
# function to compute marginal prob
lpy.X <- function(y, X, g = length(y), nu0 = 1,
                  s20 = try(summary(lm(y ~ -1 + X))$sigma^2, silent = TRUE)) {
n <- dim(X)[1]
p <- dim(X)[2]

if(p == 0) {
  Hg <- 0
  s20 <- mean(y^2)
}

if(p > 0) {
  Hg <- (g / (g + 1)) * X %*% solve(t(X) %*% X) %*% t(X)
}

SSRg <- t(y) %*% (diag(1, nrow = n) - Hg) %*% y

-0.5 * (n * log(pi) + p * log(1 + g) + (nu0 + n) * log(nu0 * s20 + SSRg) - nu0 * log(nu0 * s20)) +
lgamma((nu0 + n) / 2) - lgamma(nu0 / 2)
}
```

```
lm.gprior <- function(y, X, g = dim(X)[1], nu0 = 1, s20 = try(summary(lm(y~-1+X))$sigma^2,silent=TRUE),S=1000)
{
  n <- dim(X)[1] ; p<-dim(X)[2]
  Hg <- (g/(g+1)) * X%*%solve(t(X)%*%X)%*%t(X)
  SSRg <- t(y)%*%( diag(1,nrow=n)  - Hg ) %*%y

  s2 <- 1/rgamma(S, (nu0+n)/2, (nu0*s20+SSRg)/2 )

  Vb <- g*solve(t(X)%*%X)/(g+1)
  Eb <- Vb%*%t(X)%*%y

  E <- matrix(rnorm(S*p,0,sqrt(s2)),S,p)
  beta <- t(t(E%*%chol(Vb)) +c(Eb))

  list(beta=beta,s2=s2)
}
```

```r
# MCMC

# starting values
z <- rep(1, dim(X)[2])
lpy.c <- lpy.X(y, X[, z == 1, drop = FALSE])
S <- 10000
Z <- matrix(NA, S, dim(X)[2])
BETA <- matrix(NA, S, dim(X)[2])

# Gibbs sampler
for (s in 1 : S) {
  for (j in sample(1 : dim(X)[2])){
    # update each z
    zp <- z
    zp[j] <- 1 - zp[j]
    lpy.p <- lpy.X(y, X[, zp == 1, drop = FALSE])
    r <- (lpy.p - lpy.c) * (-1)^(zp[j] == 0)
    z[j] <- rbinom(1, 1, 1 / (1 + exp(-r)))
    if (z[j] == zp[j]) {
      lpy.c <- lpy.p
    }
  }

  beta <- z
  if (sum(z) > 0){
    beta[z == 1] <- lm.gprior(y, X[, z == 1, drop = FALSE], S = 1)$beta
  }

  Z[s,] <- z
  BETA[s,] <- beta
}
```

```r
# prob beta_j not equal to 0
for (col in 1:ncol(BETA)) {
  print(mean(BETA[, col] != 0))
}
```

```
## [1] 0.0889
## [1] 0.1681
## [1] 0.0904
## [1] 0.9849
## [1] 0.7016
## [1] 1
```

```
# posterior confidence intervals
for (col in 1:ncol(BETA)) {
  print(quantile(BETA[, col], c(0.025, 0.975)))
}
```

```
##      2.5%     97.5%
## -0.922899  0.000000
##      2.5%     97.5%
## 0.0000000 0.3241354
##      2.5%     97.5%
## 0.0000000 0.3576535
##      2.5%     97.5%
## 0.4266099 1.3257586
##     2.5%    97.5%
##  0.00000 16.92295
##      2.5%     97.5%
## 0.4836515 0.9971252
```

Using Gibbs sampling, the 95% posterior confidence intervals for “npreg”, “bp”, “skin”, “bmi”, “ped”, and “age” are (-0.922899, 0.000000), (0.0000000, 0.3241354), (0.0000000, 0.3576535), (0.4266099, 1.3257586), (0.00000, 16.92295), and (0.4836515, 0.9971252), respectively.

Comparing to the results in part a, we see that for “npreg”, “bp”, and “skin”, both approaches’ 95% posterior confidence intervals contains 0, meaning the coefficient for these three variables could be 0 (no effect on “glu” at all). However, Gibbs sampling resulted in (0.00000, 16.92295) as the 95% posterior confidence intervals for “ped”, which includes 0, whereas the approach in part a resulted in (3.484193, 17.869760), which doesn’t contain 0. This suggests that the Gibbs sampling approach indicates “ped” may not affect “glu” but the part a approach indicates the opposite, that “ped” has a positive effect on “glu”.

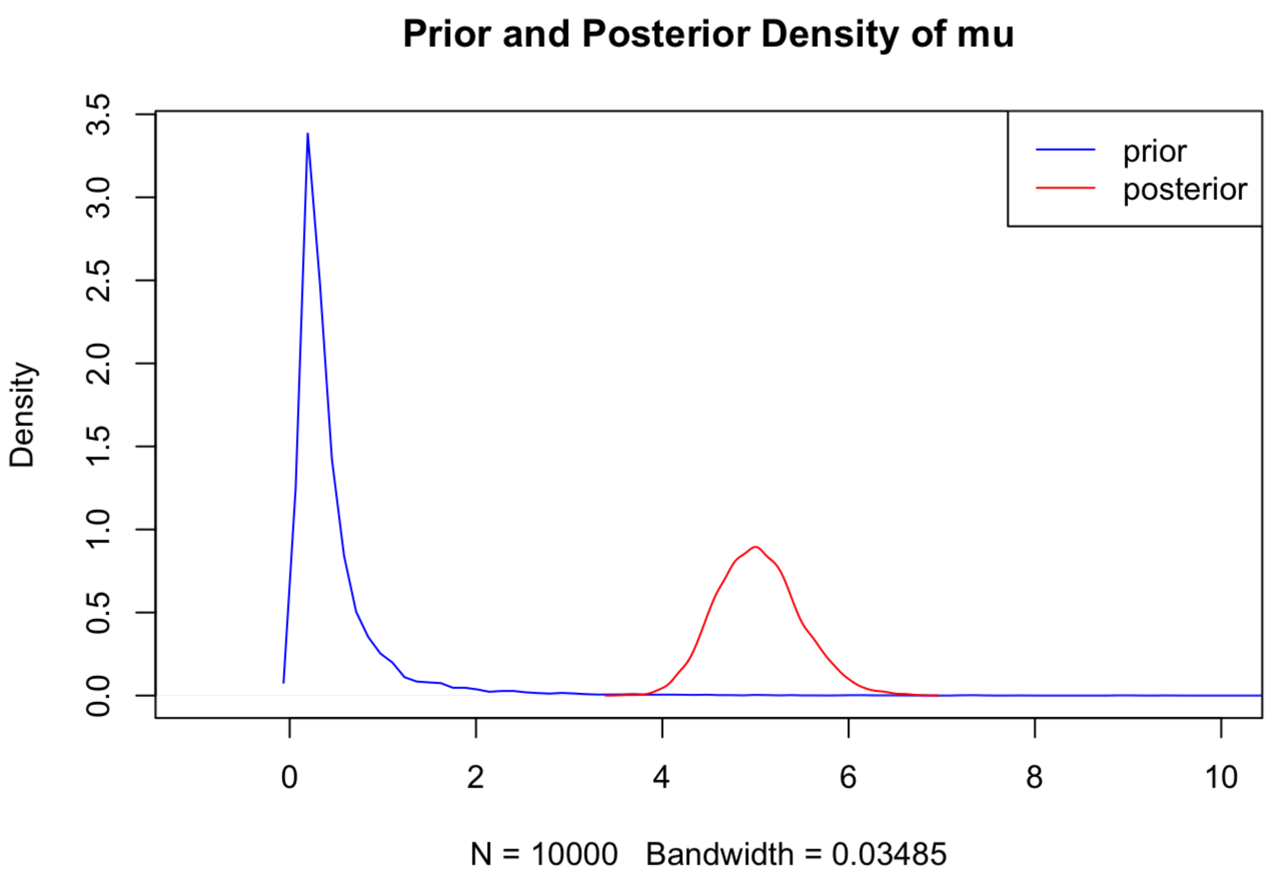
# Question 4

```
# load bird count data
bird <- readRDS("birdCount.rds")
```

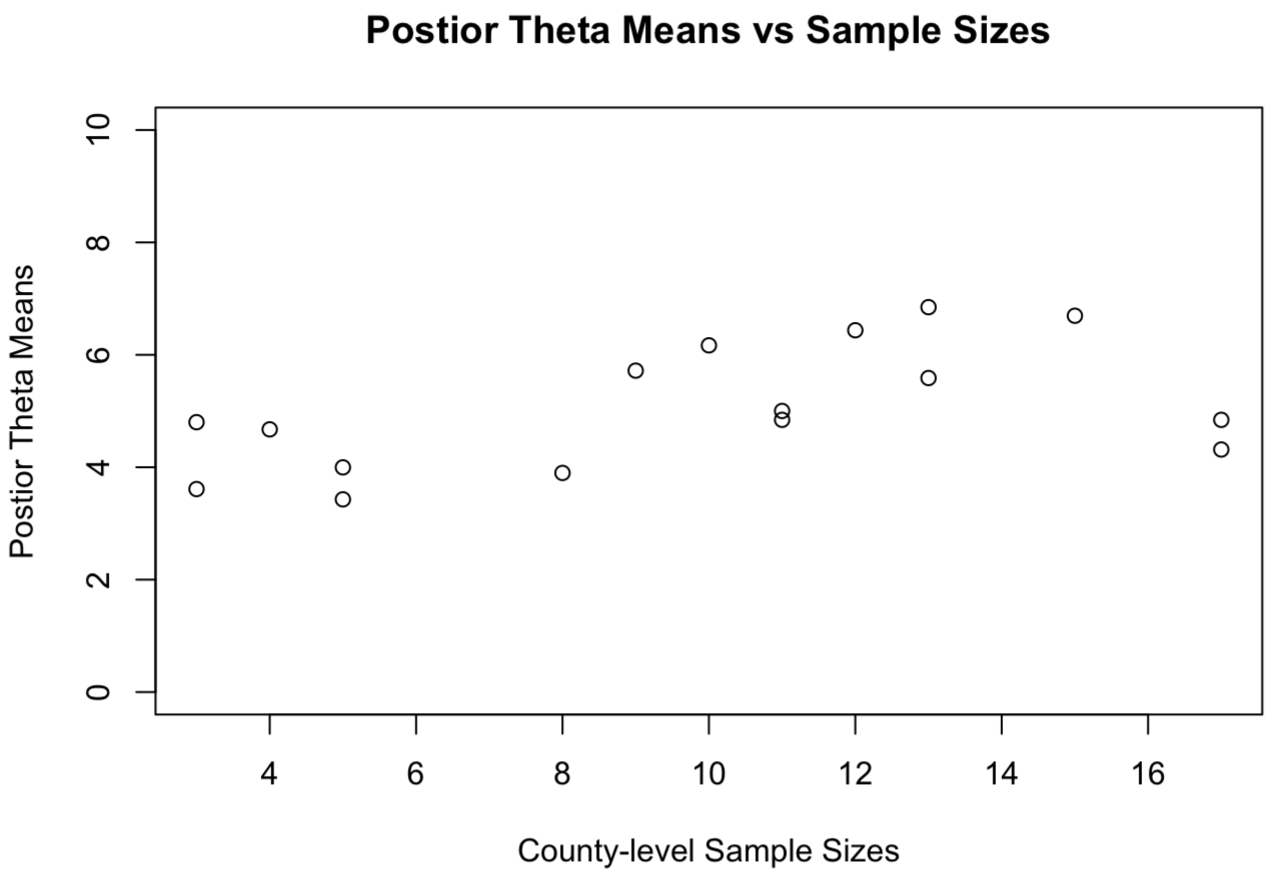
## Part c

```
# MCMC

# start values
alpha <- 10
a <- 2
b <- 1/2
mu <- mean(bird[,2])
S <- 10000

# posterior storage
THETA <- matrix(nrow = S, ncol = 16)
MU <- rep(0, S)

# Gibbs sampling
for (s in 1 : S) {
  # update theta
  theta <- rep(0, 16)
  for (j in 1:16){
    y_ij <- bird[bird[,1] == j, 2]
    n <- length(y_ij)
    theta[j] <- rgamma(1, alpha + sum(y_ij), alpha / mu + n)
  }
  THETA[s,] <- theta

  # update mu
  mu <- (1/rgamma(1, 16 * alpha + a, alpha * sum(theta) + b))
  MU[s] <- mu
}
```

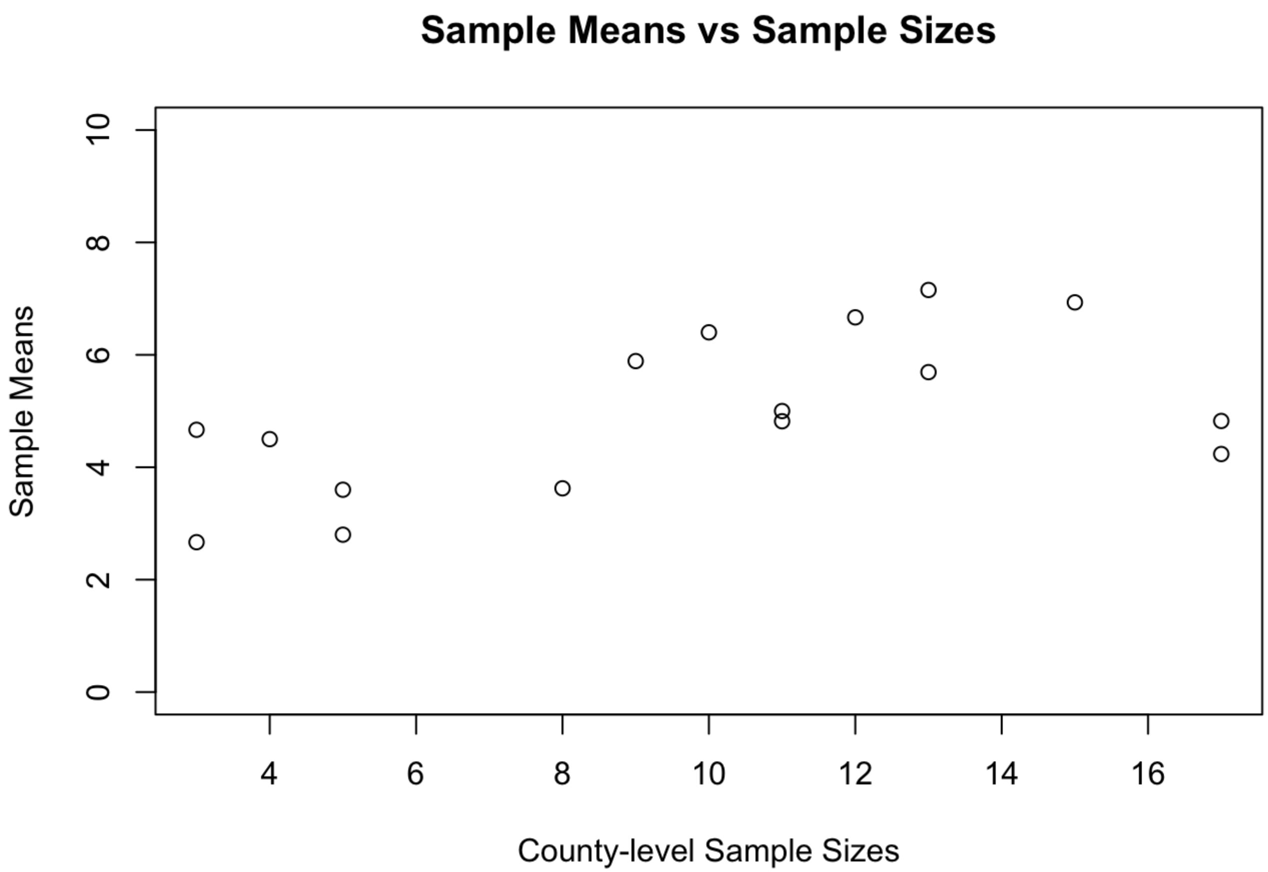
```r
# plot of the prior and posterior density of mu
mu_prior <- 1 / rgamma(S, a, b)
plot(density(mu_prior), col = "blue", xlim = c(-1, 10), main = "Prior and Posterior Density of mu")
lines(density(MU), col = "red")
legend(x = "topright", legend = c("prior", "posterior"), col = c("blue", "red"), lty = 1:1)
```



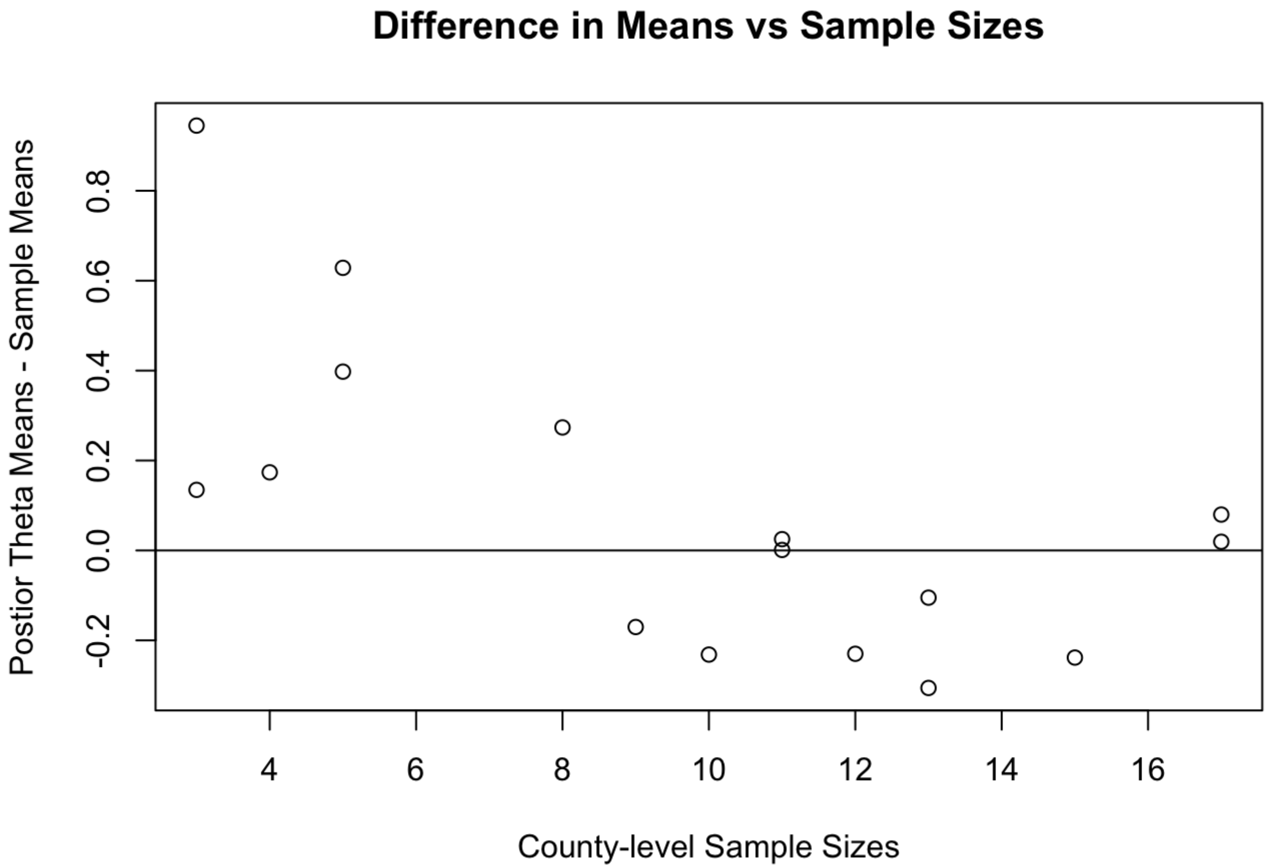
```r
# plot of the posterior means of the theta_j's versus the county-level sample sizes (the n_j's)
theta_j.means <- apply(THETA, MARGIN = 2, mean)
N <- rep(0, 16)
for (j in 1:16){
  y_ij <- bird[bird[,1] == j, 2]
  N[j] <- length(y_ij)
}

plot(N, theta_j.means,
     main = "Postior Theta Means vs Sample Sizes",
     xlab = "County-level Sample Sizes",
     ylab = "Postior Theta Means",
     ylim = c(0,10))
```




```r
# plot of the sample means (y_j's) versus sample size
sample.means <- rep(0, 16)
for (j in 1:16){
  y_ij <- bird[bird[,1] == j, 2]
  sample.means[j] <- mean(y_ij)
}

plot(N, sample.means,
     main = "Sample Means vs Sample Sizes",
     xlab = "County-level Sample Sizes",
     ylab = "Sample Means",
     ylim = c(0,10))
```




```
# hard to see difference between two graphs, so plot difference
plot(N, theta_j.means - sample.means,
     main = "Difference in Means vs Sample Sizes",
     xlab = "County-level Sample Sizes",
     ylab = "Postior Theta Means - Sample Means")
abline(h = 0)
```




We can see from the above graph that as as county-level sample sizes increases, the difference between postior theta means and sample means decreases. In fact, for smaller county-level sample sizes (e.g. size = 3), the posterior theta means are much higher than the sample means at a relative scale.