STATS 451 Homework 7

Yuzhou Peng

2024-04-04

```
library(ggplot2)
library(gridExtra)
library(tidyr)


y <- c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2,
2, 1, 5, 2, 5, 3, 2, 7, 7, 3, 3, 2, 9, 10, 4, 4, 4, 4, 4, 4, 4, 10, 4, 4, 4, 5,
11, 12,  5, 5, 6, 5, 6, 6, 6, 6, 16, 15, 15, 9, 4)
n <- c(20, 20, 20, 20, 20, 19, 19, 19, 19, 18, 17, 20, 20, 20, 20, 19, 19, 18, 18, 25, 24,
23, 20, 20, 20, 20, 20, 20, 10, 49, 19, 46, 27, 17, 49, 47, 20, 20, 13, 48, 50, 20,
20, 20, 20, 20, 20, 48, 19, 19, 19, 22, 46, 49, 20, 20, 23, 19, 22, 20, 20, 20,
52, 46, 47, 24, 14)
```

Problem 1

Let n_i be the total number of rats in lab i, and y_i be the number of rats that developed tumor in lab i.
From the problem setting, we construct a Beta-binomal conjucate model for $\theta_i, i = 1, 2...68$.
$$\theta_i|\alpha, \beta \sim Beta(\alpha, \beta) = Beta(1.4, 8.6)$$
The posterior distribution is given by $\theta_i|n_i, y_i \sim Beta(1.4 + y_i, 8.6 + n_i - y_i)$

The joint posterior density is givne by $\mathbb{P}(\theta_1 \ldots \theta_{68}|\alpha, \beta, n, y) = \prod_{i=1}^{68} \mathbb{P}(\theta_i|\alpha, \beta, n_i, y_i) = \prod_{i=1}^{68} \frac{\Gamma(\alpha + \beta + n_i)}{\Gamma(\alpha + y_i)\Gamma(\beta + n_i - y_i)} \theta_i^{\alpha + y_i - 1}(1 - \theta_i)^{\beta + n_i -}$

```
x <- seq(0.0001, 0.9999, length.out = 1000)
# helperf function to evaluate density over observations
bdens <- function(n, y, x){
  dbeta(x, y+1.4, n-y+8.6)
}
df_sep <- mapply(bdens, n, y, MoreArgs = list(x = x)) %>%
  as.data.frame() %>% cbind(x) %>% gather(ind, p, -x)
labs1 <- paste('posterior of', c('theta_j', 'theta_68'))
# plot the separate model
plot_sep <- ggplot(data = df_sep) +
  geom_line(aes(x = x, y = p, color = (ind=='V68'), group = ind)) +
  labs(x = expression(theta), y = '', title = 'Separate model', color = '') +
  scale_y_continuous(breaks = NULL) +
  scale_color_manual(values = c('blue','red'), labels = labs1) +
  theme(legend.background = element_blank(), legend.position = c(0.8,0.9))
```

```
## Warning: A numeric `legend.position` argument in `theme()` was deprecated in ggplot2
## 3.5.0.
## ℹ Please use the `legend.position.inside` argument of `theme()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
plot_sep
```

Separate model

- posterior of theta_j
- posterior of theta_68
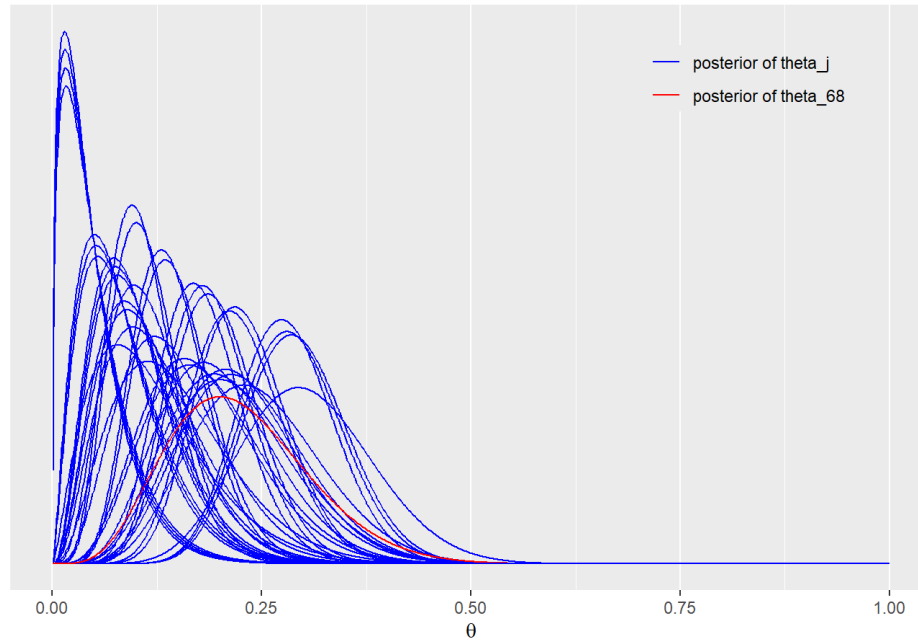
Problem 2

$$\mathbb{P}(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$$

The joint posterior distribution is given by $\mathbb{P}(\theta_1 \dots \theta_{68}, \alpha, \beta | n, y) \propto \mathbb{P}(\alpha, \beta) \prod_{i=1}^{68} \mathbb{P}(\theta_i | \alpha, \beta) \prod_{i=1}^{68} \mathbb{P}(y_i, n_i | \theta_i)$

$$= (\alpha + \beta)^{-5/2} \prod_{i=1}^{68} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_i^{\alpha-1}(1-\theta_i)^{\beta-1} \prod_{i=1}^{68} \theta_i^{y_i}(1-\theta_i)^{n_i - y_i}$$

The marginal posterior is given by $\mathbb{P}(\alpha, \beta | y, n) \propto \dfrac{\mathbb{P}(\theta_1 \dots \theta_{68}, \alpha, \beta | n, y)}{\prod_{i=1}^{68} \mathbb{P}(\theta_i | \alpha, \beta, n_i, y_i)} = \dfrac{(\alpha+\beta)^{-5/2} \prod_{i=1}^{68} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_i^{\alpha-1}(1-\theta_i)^{\beta-1} \prod_{i=1}^{68} \theta_i^{y_i}(1-\theta_i)^{n_i}}{\prod_{i=1}^{68} \frac{\Gamma(\alpha+\beta+n_i)}{\Gamma(\alpha+y_i)\Gamma(\beta+n_i-y_i)} \theta_i^{\alpha+y_i-1}(1-\theta_i)^{\beta+n_i-y_i-1}}$

$$= (\alpha+\beta)^{-5/2} \prod_{i=1}^{68} \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+y_i)\Gamma(\beta+n_i-y_i)}{\Gamma(\alpha+\beta+n_i)\Gamma(\alpha)\Gamma(\beta)}$$

Problem 3

```
# Compute the marginal posterior of alpha and beta in hierarchical model
# Use grid
A <- seq(0.5, 6, length.out = 100) ## alpha
B <- seq(3, 33, length.out = 100) ## beta
# make vectors that contain all pairwise combinations of A and B
cA <- rep(A, each = length(B))
cB <- rep(B, length(A))

# Use logarithms for numerical accuracy!
lpfun <- function(a, b, y, n) log(a+b)*(-5/2) +
  sum(lgamma(a+b)-lgamma(a)-lgamma(b)+lgamma(a+y)+lgamma(b+n-y)-lgamma(a+b+n))

lp <- mapply(lpfun, cA, cB, MoreArgs = list(y, n))
df_marg <- data.frame(x = cA, y = cB, p = exp(lp - max(lp)))

# Subtract maximum value to avoid over/underflow in exponentation
title1 <- 'The marginal posterior of alpha and beta in hierarchical model'
# create a plot of the marginal posterior density
postdensityalphabeta = ggplot(data = df_marg, aes(x = x, y = y)) +
  geom_raster(aes(fill = p, alpha = p), interpolate = T) +
  geom_contour(aes(z = p), colour = 'black', size = 0.2) +
  coord_cartesian(xlim = c(1,5), ylim = c(4, 26)) +
  labs(x = 'alpha', y = 'beta', title = title1) +
  scale_fill_gradient(low = 'yellow', high = 'red', guide = F) +
  scale_alpha(range = c(0, 1), guide = F)
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

obtain random samples from p(alpha,beta|data)

(samp_A, samp_B) contains 100 samples of alpha and beta
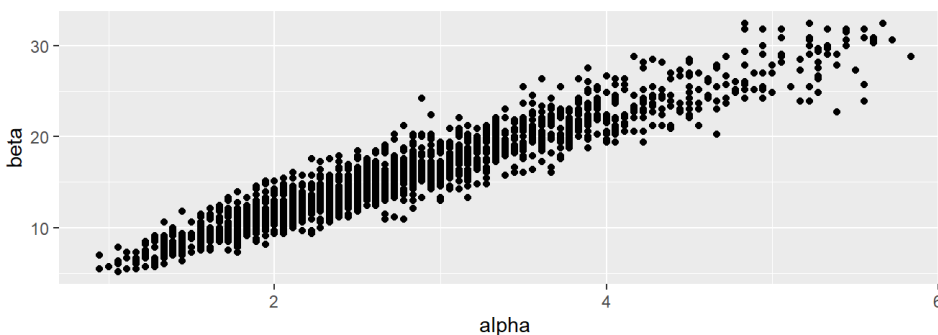
```
nsamp <- 100
samp_indices <- sample(length(df_marg$p), size = nsamp,
                       replace = T, prob = df_marg$p/sum(df_marg$p))

samp_A <- cA[samp_indices[1:nsamp]]
samp_B <- cB[samp_indices[1:nsamp]]
```

visualize the samples

```
nsamp2 <- 2000
samp_indices2 <- sample(length(df_marg$p), size = nsamp2,
                        replace = T, prob = df_marg$p/sum(df_marg$p))
samplesalphabeta = data.frame(alpha = cA[samp_indices2[1:nsamp2]], beta = cB[samp_indices2[1:nsamp2]])
scatteralphabeta = ggplot(samplesalphabeta, aes(x=alpha, y=beta)) + geom_point()
grid.arrange(scatteralphabeta, postdensityalphabeta)
```

```
## Warning: The `guide` argument in `scale_*()` cannot be `FALSE`. This was deprecated in
## ggplot2 3.3.4.
## i Please use "none" instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```


The marginal posterior of alpha and beta in hierarchical model
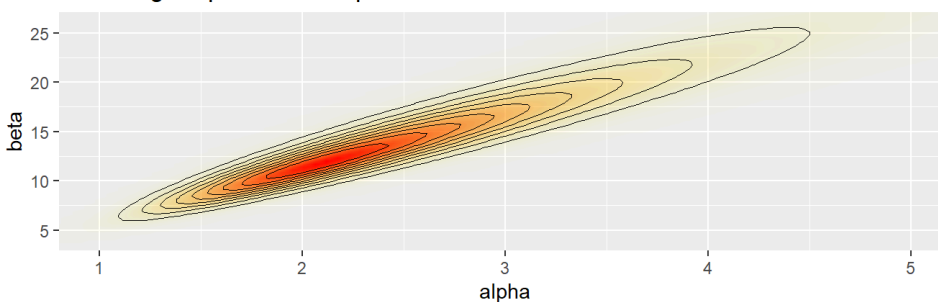
Problem 4

```
# helper function to convert ind to numeric for subsetting
indtonum <- function(x) strtoi(substring(x,2))

# sample from posterior distribution of theta_j

nsamp <- 10000
samp_indices <- sample(length(df_marg$p), size = nsamp,
                       replace = T, prob = df_marg$p/sum(df_marg$p))

samp_A <- cA[samp_indices[1:nsamp]]
samp_B <- cB[samp_indices[1:nsamp]]
samplestheta <- matrix(0, nsamp, length(y))
for(j in 1:length(y)){
  samplestheta[, j] = sapply(1:nsamp, function(k) rbeta(1, samp_A[k]+y[j], samp_B[k]+n[j]-y[j]))
}
breakpoints <- seq(0, 1, length.out = 200)
xx <- hist(samplestheta[, 1], breaks = breakpoints, plot = FALSE)$mids
densitytheta <- matrix(0, length(breakpoints) - 1, length(y))
for(j in 1:length(y)){
  densitytheta[, j] <- hist(samplestheta[, j], breaks = breakpoints, plot = FALSE)$density
}
df_hier_samp_full <-densitytheta %>%
  as.data.frame() %>% cbind(xx) %>% gather(ind, p, -xx)
plot_hier7_samp <- ggplot(data = df_hier_samp_full) +
  geom_line(aes(x = xx, y = p, color = (ind=='V49'), group = ind)) +
  labs(x = expression(theta), y = '', title = paste('Hierarchical model given the posterior samples of ', bquote(alpha),",", ex
pression(beta)), color = '') +
  scale_color_manual(values = c('blue', 'red'), guide = F) +
  scale_y_continuous(breaks = NULL) +
  theme(legend.background = element_blank(), legend.position = c(0.8,0.9))


plot_hier7_samp_part <- ggplot(data = subset(df_hier_samp_full, indtonum(ind)%%7==0)) +
  geom_line(aes(x = xx, y = p, color = (ind=='V49'), group = ind)) +
  labs(x = expression(theta), y = '', title = paste('(Part of )Hierarchical model given the posterior samples of ', bquote(alph
a),",", expression(beta)), color = '') +
  scale_color_manual(values = c('blue', 'red'), guide = F) +
  scale_y_continuous(breaks = NULL) +
  theme(legend.background = element_blank(), legend.position = c(0.8,0.9))


grid.arrange(plot_hier7_samp_part, plot_hier7_samp)
```
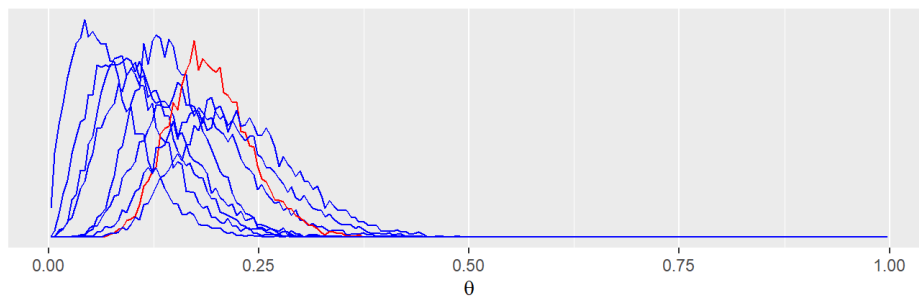
(Part of)Hierarchical model given the posterior samples of alpha , beta

Hierarchical model given the posterior samples of alpha , beta