Yeojin

ESC 2019 Fall – HW #3

Chapter 8

8.1 Components of variance: Consider the hierarchical model where

$$\theta_1, \ldots, \theta_m | \mu, \tau^2 \sim \text{i.i.d. normal}(\mu, \tau^2)$$
$$y_{1,j}, \ldots, y_{n_j,j} | \theta_j, \sigma^2 \sim \text{i.i.d. normal}(\theta_j, \sigma^2) .$$

For this problem, we will eventually compute the following:
$\text{Var}[y_{i,j}|\theta_i, \sigma^2]$, $\text{Var}[\bar{y}_{\cdot,j}|\theta_i, \sigma^2]$, $\text{Cov}[y_{i_1,j}, y_{i_2,j}|\theta_j, \sigma^2]$
$\text{Var}[y_{i,j}|\mu, \tau^2]$, $\text{Var}[\bar{y}_{\cdot,j}|\mu, \tau^2]$, $\text{Cov}[y_{i_1,j}, y_{i_2,j}|\mu, \tau^2]$
First, lets use our intuition to guess at the answers:

a) Which do you think is bigger, $\text{Var}[y_{i,j}|\theta_i, \sigma^2]$ or $\text{Var}[y_{i,j}|\mu, \tau^2]$? To guide your intuition, you can interpret the first as the variability of the Y's when sampling from a fixed group, and the second as the variability in first sampling a group, then sampling a unit from within the group.

b) Do you think $\text{Cov}[y_{i_1,j}, y_{i_2,j}|\theta_j, \sigma^2]$ is negative, positive, or zero? Answer the same for $\text{Cov}[y_{i_1,j}, y_{i_2,j}|\mu, \tau]$. You may want to think about what $y_{i_2,j}$ tells you about $y_{i_1,j}$ if θ_j is known, and what it tells you when θ_j is unknown.

c) Now compute each of the six quantities above and compare to your answers in a) and b).

d) Now assume we have a prior $p(\mu)$ for μ. Using Bayes' rule, show that

$$p(\mu|\theta_1, \ldots, \theta_m, \sigma^2, \tau^2, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_m) = p(\mu|\theta_1, \ldots, \theta_m, \tau^2).$$

(a) <intuition>

$Var(y_{i,j} | \mu, \tau^2)$ 가 더 크다. (이미 문제에서 답 다 알려줌)

→ $var(y_{i,j} | \theta_i, \sigma^2)$ 의 경우 이미 i^{th} group으로 group이 fixed 되어있고 그 그룹에서 y를 sampling 할 때 변동성

→ $var(y_{i,j} | \mu, \tau^2)$ $N(\mu, \tau^2)$ 을 따르는 $\theta_1, \ldots, \theta_m$ 에서 하나의 group을 선택하고 corresponding θ_i 에 대해 y를 sampling

: 위에 비해 그룹 간 변동성이 추가되어 더 크다

(b) → $Cov(y_{i1,j}, y_{i2,j} \mid \theta_j, \sigma^2) = 0$

∵ $\{Y_{1,j}, \ldots, Y_{nj,j} \mid \phi_j\} \stackrel{iid}{\sim} p(y \mid \phi_j)$ $\phi_j = \{\theta_j, \sigma^2\}$

θ_j가 known 이면 $y_{i1,j}$ 와 $y_{i2,j}$는 independently distributed

→ $Cov(y_{i1,j}, y_{i2,j} \mid \mu, \tau^2) \neq 0$ probably $> 0 \ldots ?$

∵ $y_{i1,j}$를 아는 것이 θ_j를 update하고 $y_{i2,j}$에 대해 정보를 제공한다. 같은 θ_j에서 온 $y_{i1,j}$ 와 $y_{i2,j}$는 아마 비슷할 것이므로 positively correlated

(c) $\mathrm{Var}[y_{i,j}|\theta_i, \sigma^2]$, $\mathrm{Var}[\bar{y}_{\cdot,j}|\theta_i, \sigma^2]$, $\mathrm{Cov}[y_{i_1,j}, y_{i_2,j}|\theta_j, \sigma^2]$
$\mathrm{Var}[y_{i,j}|\mu, \tau^2]$, $\mathrm{Var}[\bar{y}_{\cdot,j}|\mu, \tau^2]$, $\mathrm{Cov}[y_{i_1,j}, y_{i_2,j}|\mu, \tau^2]$

※ 활용!
$\phi_j = \{\theta_j, \sigma^2\}$, $p(y|\phi_j) = N(\theta_j, \sigma^2)$
$\psi = \{\mu, \tau^2\}$, $p(\theta_j|\psi) = N(\mu, \tau^2)$

① $Var(y_{ij} \mid \theta_i, \sigma^2) = \sigma^2$

② $Var(\bar{y}_{\cdot,j} \mid \theta_i, \sigma^2) = \dfrac{\sigma^2}{n_j}$ sample mean의 variance

③ $Var(y_{ij} \mid \mu, \tau^2) = Var(\underbrace{E(y_{ij} \mid \theta_j, \sigma^2)}_{=\theta_j} \mid \mu, \tau^2)$
$+ E(\underbrace{Var(y_{ij} \mid \theta_j, \sigma^2)}_{=\sigma^2} \mid \mu, \tau^2)$
$= Var(\theta_j \mid \mu, \tau^2) + E(\sigma^2 \mid \mu, \tau^2)$
$= \tau^2 + \sigma^2$

* (a)에서의 결론과 같다! $Var(y_{i,j} \mid \mu, \tau^2) > Var(y_{i,j} \mid \theta_i, \sigma^2)$

④ $Var(\bar{y}_{\cdot,j} \mid \mu, \tau^2) = Var(E(\bar{y}_{\cdot j} \mid \theta_j, \sigma^2) \mid \mu, \tau^2)$
$+ E(Var(\bar{y}_{\cdot j} \mid \theta_j, \sigma^2) \mid \mu, \tau^2)$
$= Var(\theta_j \mid \mu, \tau^2) + E(\frac{\sigma^2}{n_j} \mid \mu, \tau^2)$

$$= \tau^2 + \frac{\sigma^2}{n_j}$$

⑤ $Cov(y_{i_1 j}, y_{i_2 j} \mid \theta_j, \sigma^2)$

$= E(y_{i_1 j} y_{i_2 j} \mid \theta_j, \sigma^2) - E(y_{i_1 j} \mid \theta_j, \sigma^2) E(y_{i_2 j} \mid \theta_j, \sigma^2)$

$= 0$

⑥ $Cov(y_{i_1 j}, y_{i_2 j} \mid \mu, \tau^2)$ by law of total covariance

$= E(\cancel{Cov(y_{i_1 j}, y_{i_2 j} \mid \theta_j, \sigma^2)} \mid \mu, \tau^2)$ → 0 (iid에 의해)

$+ Cov(E(y_{i_1, j} \mid \theta_j, \sigma^2), E(y_{i_2, j} \mid \theta_j, \sigma^2))$

$= Cov(\theta_j, \theta_j)$

$= \tau^2$

* 이 역시 위에서 생각한대로 $Cov(y_{i_1 j}\ y_{i_2 j} \mid \theta_j, \sigma^2) = 0$

$Cov(y_{i_1 j}, y_{i_2 j} \mid \mu, \tau^2) > 0$

(d) assume prior $p(\mu)$

$p(\mu \mid \theta_1, \ldots, \theta_m, \sigma^2, \tau^2, y_1, \ldots, y_m)$

$= p(\mu \mid \underset{\sim}{\theta}, \sigma^2, \tau^2, \underset{\sim}{y})$

$$= \frac{p(\mu, \underset{\sim}{\theta}, \sigma^2, \tau^2, \underset{\sim}{y})}{\int p(\mu, \underset{\sim}{\theta}, \sigma^2, \tau^2, \underset{\sim}{y})\, d\mu} = \frac{p(\mu) p(\tau^2) p(\sigma^2) p(Y \mid \underset{\sim}{\theta}, \sigma^2) p(\underset{\sim}{\theta} \mid \mu, \tau^2)}{\int \quad '' \quad d\mu}$$

= 분모의 적분에서 μ를 갖고 있지 않는 항은 밖으로 빼줌

$$= \frac{p(\mu)\, \cancel{p(\tau^2)}\, \cancel{p(\sigma^2)}\, \cancel{p(Y \mid \underset{\sim}{\theta}, \sigma^2)}\, p(\underset{\sim}{\theta} \mid \mu, \tau^2)}{\cancel{p(\tau^2)}\, \cancel{p(\sigma^2)}\, \cancel{p(Y \mid \underset{\sim}{\theta}, \sigma^2)} \int p(\mu) p(\underset{\sim}{\theta} \mid \mu, \tau^2)\, d\mu} = p(\underset{\sim}{\mu} \mid \theta, \tau^2)$$

cf) Bayes Rule : $P(A \mid B) = \dfrac{P(B \mid A)\, P(A)}{P(B)}$

ESC 2019 Fall HW#3

Yeojin Jung

2019 11 11

(a).

Gibbs Sampler- obtain posterior distributions /check convergence of Markov chain /find effective sample size

(For further information, read 8.4.1 of the textbook.)

We need to find posterior distributions of $\theta, \sigma^2, \mu, \tau^2$.

Load Data

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(ggplot2)

schools.list = lapply(1:8, function(i) {
  s.tbl = paste0('http://www.stat.washington.edu/people/pdhoff/Book/Data/hwdata/school', i, '.dat') %>%
    url %>%
    read.table

  data.frame(
    school = i,
    hours = s.tbl[, 1] %>% as.numeric
  )
})
schools.raw = do.call(rbind, schools.list)
Y = schools.raw
str(Y)
```

```
## 'data.frame':    180 obs. of  2 variables:
##  $ school: int  1 1 1 1 1 1 1 1 1 1 ...
##  $ hours : num  2.11 9.75 13.88 11.3 8.93 ...
```

```
head(Y)
```

```
##   school hours
## 1      1  2.11
## 2      1  9.75
```

```
## 3      1 13.88
## 4      1 11.30
## 5      1  8.93
## 6      1 15.66
```

Assign the prior values given in the problem.

```
mu0 = 7
g20 = 5
t20 = 10
eta0 = 2
s20 = 15
nu0 = 2

m=length(unique(Y[,1])) #m is the number of groups(schools)
```

Obtain initial values of theta, sigma2, mu, tau2

(Note that theta is a vector, while the others are constants.)

```
# Starting values - use sample mean and variance
n = sv = ybar = rep(NA, m)
for (j in 1:m) {
  Y_j = Y[Y[, 1] == j, 2]
  ybar[j] = mean(Y_j)
  sv[j] = var(Y_j)
  n[j] = length(Y_j)
}

# Let initial theta estimates be the sample means
#Let initial values of sigma2, mu, and tau2 be "sample mean and variance"
theta = ybar
sigma2 = mean(sv)
mu = mean(theta)
tau2 = var(theta)
```

Conduct MCMC

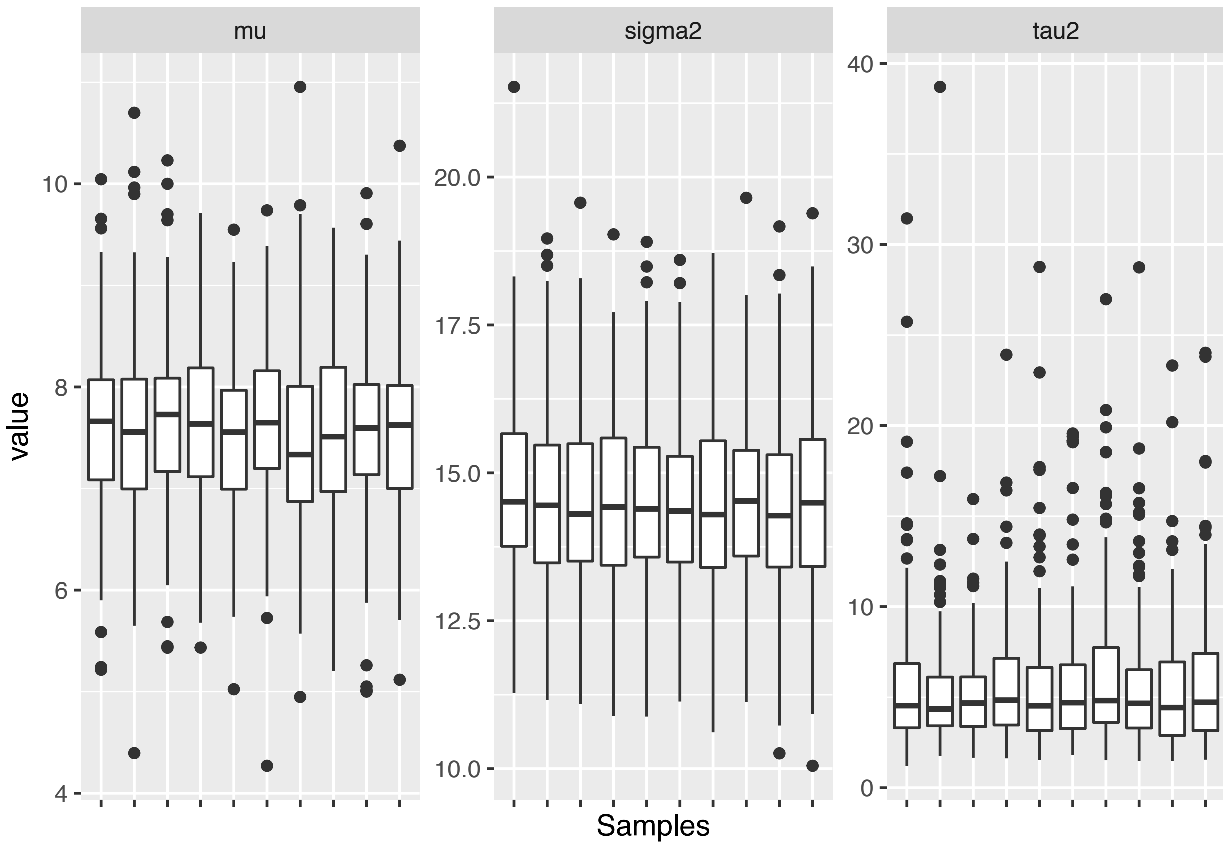
```
# MCMC
S = 1500
THETA = matrix(nrow = S, ncol = m)
# Storing sigma, mu, theta together
SMT = matrix(nrow = S, ncol = 3)
colnames(SMT) = c('sigma2', 'mu', 'tau2')


for (s in 1:S) {
  # Sample thetas(normal)
  for (j in 1:m) {
    vtheta = 1 / (n[j] / sigma2 + 1 / tau2)
    etheta = vtheta * (ybar[j] * n[j] / sigma2 + mu / tau2)
    theta[j] = rnorm(1, etheta, sqrt(vtheta))
  }
```

```r
  # Sample sigma2(inverse gamma)
  nun = nu0 + sum(n)
  ss = nu0 * s20
  # Pool variance
  for (j in 1:m) {
    ss = ss + sum((Y[Y[, 1] == j, 2] - theta[j])^2)
  }
  sigma2 = 1 / rgamma(1, nun / 2, ss / 2)

  # Sample mu(normal)
  vmu = 1 / (m / tau2 + 1 /g20)
  emu = vmu * (m * mean(theta) / tau2 + mu0 / g20)
  mu = rnorm(1, emu, sqrt(vmu))

  # Sample tau2(inverse gamma)
  etam = eta0 + m
  ss = eta0 * t20 + sum((theta - mu)^2)
  tau2 = 1 / rgamma(1, etam / 2, ss / 2)

  # Store params
  THETA[s, ] = theta
  SMT[s, ] = c(sigma2, mu, tau2)
}
```

Check convergence of Markov Chain

The chain seems to have achieved stationarity. We will move on.

```r
smt.df = data.frame(SMT)
colnames(smt.df) = c('sigma2', 'mu', 'tau2')
smt.df$s = 1:S
cut_size = 10
smt.df = smt.df %>%
  tbl_df %>%
  mutate(scut = cut(s, breaks = cut_size)) %>%
  gather('variable', 'value', sigma2:tau2)
ggplot(smt.df, aes(x = scut, y = value)) +
  facet_wrap(~ variable, scales = 'free_y') +
  geom_boxplot() +
  theme(axis.text.x = element_blank()) +
  xlab('Samples')
```



Find effective sample size

```
library(coda)
effectiveSize(SMT[, 1])
```

```
## var1
## 1500
```

```
effectiveSize(SMT[, 2])
```

```
##    var1
## 1001.97
```

```
effectiveSize(SMT[, 3])
```

```
##     var1
## 1091.324
```

(b).

Compute poterior means & 95% confidence regions for sigma2, mu, tau2 / compare posterior and prior densities

Posterior means & 95% confidence intervals

```
t(apply(SMT, MARGIN = 2, FUN = quantile, probs = c(0.025, 0.975)))
```

```
##              2.5%     97.5%
```

```
## sigma2 11.790907 17.852995
## mu      5.897132  9.223577
## tau2    1.897006 15.700606
```

```
apply(SMT,2, mean)
```

```
##    sigma2        mu      tau2
## 14.517564  7.576914  5.645859
```

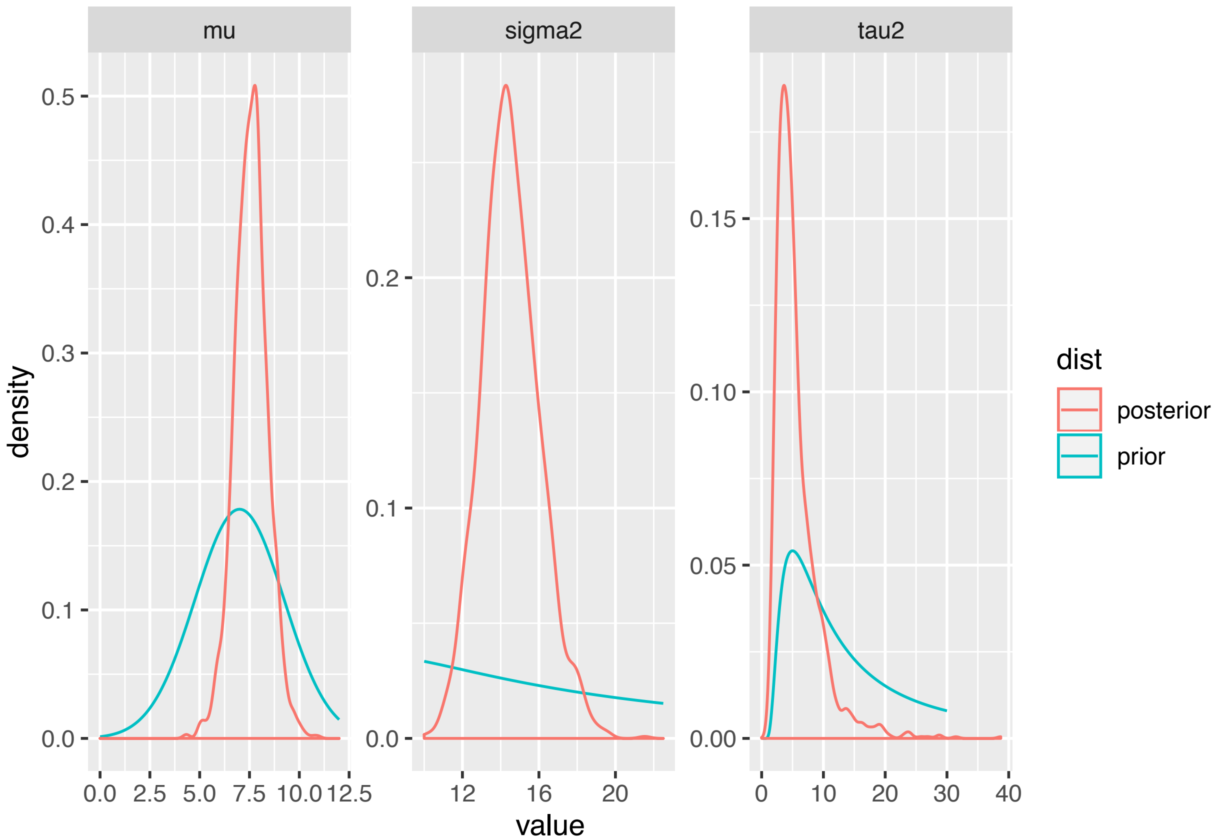
Compare Prior and Posterior Densities

We can see that prior estimate for sigma2 was very off

```
library(MCMCpack)
```

```
## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

## ##
## ## Markov Chain Monte Carlo Package (MCMCpack)

## ## Copyright (C) 2003-2019 Andrew D. Martin, Kevin M. Quinn, and Jong Hee Park

## ##
## ## Support provided by the U.S. National Science Foundation

## ## (Grants SES-0350646 and SES-0350613)
## ##
```

```
sigma2_prior = data.frame(
  value = seq(10, 22.5, by = 0.1),
  density = dinvgamma(seq(10, 22.5, by = 0.1), nu0 / 2, nu0 * s20 / 2),
  variable = 'sigma2'
)
tau2_prior = data.frame(
  value = seq(0, 30, by = 0.1),
  density = dinvgamma(seq(0, 30, by = 0.1), eta0 / 2, eta0 * t20 / 2),
  variable = 'tau2'
)
mu_prior = data.frame(
  value = seq(0, 12, by = 0.1),
  density = dnorm(seq(0, 12, by = 0.1), mu0, sqrt(g20)),
  variable = 'mu'
)
priors = rbind(sigma2_prior, tau2_prior, mu_prior)
priors$dist = 'prior'
smt.df$dist = 'posterior'
ggplot(priors, aes(x = value, y = density, color = dist)) +
  geom_line() +
  geom_density(data = smt.df, mapping = aes(x = value, y = ..density..)) +
  facet_wrap(~ variable, scales = 'free')
```

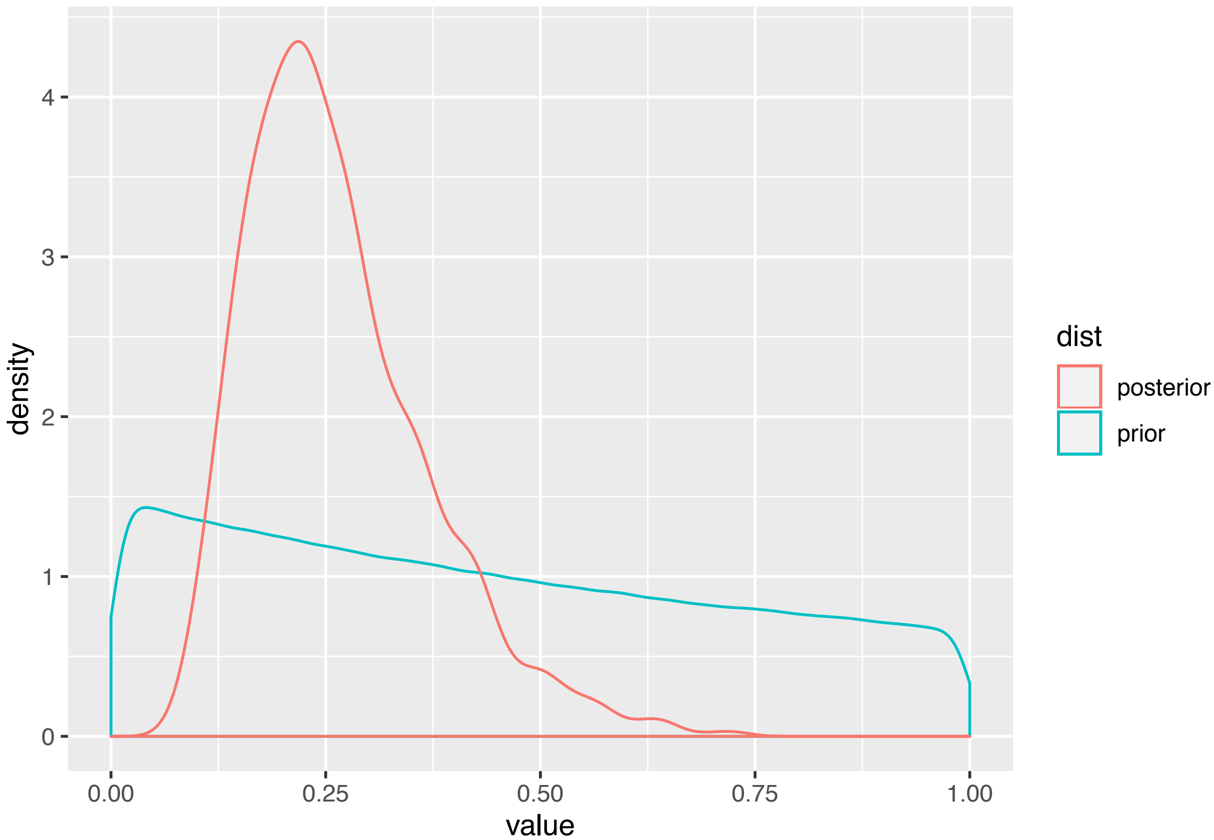


(c).

Plot posterior and prior density of R/ Describe evidence of between school variation

$$R = \frac{\tau^2}{\sigma^2 + \tau^2}$$

```
t20_prior = (1 / rgamma(1e6, eta0 / 2, eta0 * t20 / 2))
s20_prior = (1 / rgamma(1e6, nu0 / 2, nu0 * s20 / 2))
R_prior = data.frame(
  value = (t20_prior) / (t20_prior + s20_prior),
  dist = 'prior'
)
R_post = data.frame(
  value = SMT[, 'tau2'] / (SMT[, 'tau2'] + SMT[, 'sigma2']),
  dist = 'posterior'
)
ggplot(R_prior, aes(x = value, y = ..density.., color = dist)) +
  geom_density(data = R_prior) +
  geom_density(data = R_post)
```




```r
mean(R_post$value)
```

```
## [1] 0.2632715
```

R denotes the proportion of between group variability to total variability. The peak of posterior density is formed at 0.25, so after inference, we expect between group variability to be around 25% of total variability.

(d).

Obtain posterior probability that theta 7 is smaller than theta 6 & posterior probability that theta 7 is the smallest among all theta's.

```r
theta7_lt_6 = THETA[, 7] < THETA[, 6]
mean(theta7_lt_6)
```

```
## [1] 0.5213333
```

```r
theta7_smallest = (THETA[, 7] < THETA[, -7]) %>%
  apply(MARGIN = 1, FUN = all)
mean(theta7_smallest)
```

```
## [1] 0.3073333
```

(e).

Plot the sample averages against posterior expectations of thetas/ compare mu and sample mean of all observations

```r
relationship = data.frame(
  sample_average = ybar,
  post_exp = colMeans(THETA),
  school = 1:length(ybar)
)
relationship
```

```
##   sample_average  post_exp school
## 1       9.464000  9.239876      1
## 2       7.033478  7.093397      2
## 3       7.953000  7.891648      3
## 4       6.232083  6.359439      4
## 5      10.765833 10.362508      5
## 6       6.205000  6.357737      6
## 7       6.132727  6.315325      7
## 8       7.381000  7.410154      8
```

```r
ggplot(relationship, aes(x = sample_average, y = post_exp, label = school)) +
  geom_text() +
  geom_abline(slope = 1, intercept = 0) +
  geom_hline(yintercept = mean(schools.raw[, 'hours']), lty = 2) +
  annotate('text', x = 10, y = 7.9, label = paste0("Pooled sample mean ", round(mean(schools.raw[, 'hou
  geom_hline(yintercept = mean(SMT[, 'mu']), color = 'red') +
  annotate('text', x = 10, y = 7.4, label = paste0("Posterior exp. mu ", round(mean(SMT[, 'mu']), 2)), 
```