

STA 601/360 Homework 5

Yifei Wang

03 October, 2019

1. Hoff 5.2. Sensitivity analysis: Thirty-two students in a science classroom were randomly assigned to one of two study methods, A and B , so that $n_A = n_B = 16$ students were assigned to each method. After several weeks of study, students were examined on the course material with an exam designed to give an average score of **75** with a standard deviation of **10**. The scores for the two groups are summarized by $\{\bar{y}_A = 75.2, s_A = 7.3\}$ and $\{\bar{y}_B = 77.5, s_B = 8.1\}$. Consider independent, conjugate normal prior distributions for each of θ_A and θ_B , with $\mu_0 = 75$ and $\sigma_0^2 = 100$ for both groups. For each $(\kappa_0, \nu_0) \in \{(1, 1), (2, 2), (4, 4), (8, 8), (16, 16), (32, 32)\}$ (or more values), obtain $\Pr(\theta_A < \theta_B \mid y_A, y_B)$ via Monte Carlo sampling. Plot this probability as a function of $(\kappa_0 = \nu_0)$. Describe how you might use this plot to convey the evidence that $\theta_A < \theta_B$ to people of a variety of prior opinions.

a. Do this via Monte Carlo as the book recommends.

According to the text book, given the sampling model and priors

$$\begin{aligned} y_1, \dots, y_n \mid \theta, \sigma^2 &\sim \text{Normal}(\theta, \sigma^2) \\ 1/\sigma^2 &\sim \text{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0}{2} \sigma^2\right) \\ \theta \mid \sigma^2 &\sim \text{Normal}\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right) \end{aligned}$$

We have the posterior

$$\begin{aligned} p(\theta, \sigma^2 \mid y_1, \dots, y_n) &= p(\theta \mid y_1, \dots, y_n, \sigma^2) p(\sigma^2 \mid y_1, \dots, y_n) \\ p(\theta \mid y_1, \dots, y_n, \sigma^2) &\sim \text{Normal}\left(\mu_n, \frac{\sigma^2}{\kappa_n}\right) \\ p(\sigma^2 \mid y_1, \dots, y_n) &\sim \text{Gamma}\left(\frac{\nu_n}{2}, \frac{\nu_n}{2} \sigma_n^2\right) \end{aligned}$$

where

$$\begin{aligned} \kappa_n &= \kappa_0 + n \\ \mu_n &= \frac{\kappa_0}{\kappa_n} \mu_0 + \frac{n}{\kappa_n} \bar{y} \\ \nu_n &= \nu_0 + n \\ \sigma_n^2 &= \frac{1}{\nu_n} \left[\nu_0 \sigma_0^2 + (n-1) s^2 + \frac{\kappa_0 n}{\kappa_n} (\bar{y} - \mu_0)^2 \right] \\ \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i \\ s^2 &= \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \end{aligned}$$

```

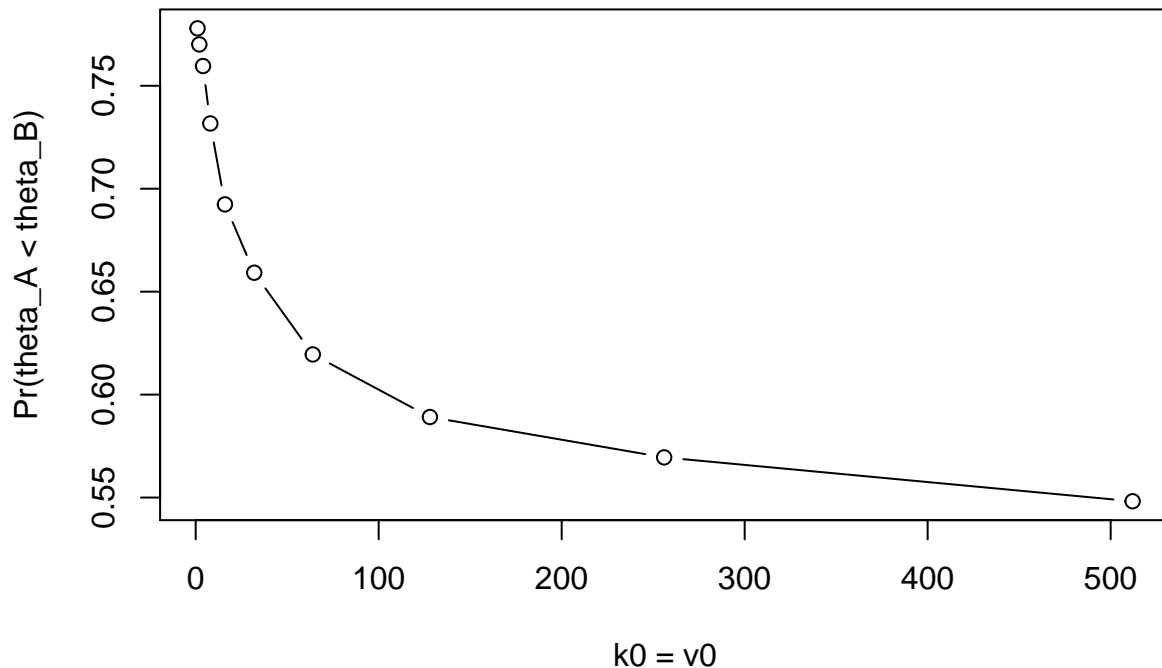
# data
n_A = n_B = 16
mu_A = 75.2; s_A = 7.3
mu_B = 77.5; s_B = 8.1

N = 10
Pr = c(rep(NA, N))
for (i in 1:N) {
  # prior
  mu0 = 75; s0 = 10
  k0 = v0 = 2^(i-1)
  # posterior
  k_n_A = k0 + n_A
  k_n_B = k0 + n_B
  mu_n_A = (k0*mu0 + n_A*mu_A) / k_n_A
  mu_n_B = (k0*mu0 + n_B*mu_B) / k_n_B
  v_n_A = v0 + n_A
  v_n_B = v0 + n_B
  s2_n_A = (v0*s0^2 + (n_A-1)*s_A^2 + k0*n_A*(mu_A-mu0)^2/k_n_A) / v_n_A
  s2_n_B = (v0*s0^2 + (n_B-1)*s_B^2 + k0*n_B*(mu_B-mu0)^2/k_n_B) / v_n_B
  # MC sampling
  S = 10000
  sigma2_A = 1/rgamma(S, v_n_A/2, s2_n_A*v_n_A/2)
  sigma2_B = 1/rgamma(S, v_n_B/2, s2_n_B*v_n_B/2)
  theta_A = rnorm(S, mu_n_A, sqrt(sigma2_A/k_n_A))
  theta_B = rnorm(S, mu_n_B, sqrt(sigma2_B/k_n_B))
  # Pr
  Pr[i] = mean(theta_A < theta_B)
}

plot(2^(0:(N-1)), Pr, "b", xlab = "k0 = v0", ylab = "Pr(theta_A < theta_B)",
     main="Probability of theta_A < theta_B with different prior using MC")

```

Probability of $\theta_A < \theta_B$ with different prior using MC



The sample mean in group B is greater than the sample mean in group A , and both of them are greater than the prior mean. Priors here are “dragging” the posterior mean from data mean. The greater the prior sample size ν_0 and κ_0 , the greater the “dragging” effect. As shown in the plot, the probability of $\theta_A < \theta_B$ is greater than 0.5 and converges to 0.5 for larger prior size. This means θ_A is greater than θ_B with a large confidence, unless we have a huge prior sample size (compared to $n_A = n_B = 16$).

b. Repeat the above problem using the STAN code from lab. Comment on the quality of estimation when you run HMC for only a few hundred iterations. (You might need to edit the stan file to get the priors you want)

```
# generate data for stan
restandardize <- function(x, mean, sd){
  z <- (x - mean(x))/sd(x)
  return(sd*z + mean)
}

nA = nB = 16
ybarA = 75.2
sA = 7.3
ybarB = 77.5
sB = 8.1

yA = rnorm(nA) %>%
  restandardize(ybarA, sA)
yB = rnorm(nB) %>%
  restandardize(ybarB, sB)

# compile stan model
library(rstan)
```

```

## Loading required package: StanHeaders
## rstan (Version 2.19.2, GitRev: 2e1f913d3ca3)
## For execution on a local, multicore CPU with excess RAM we recommend calling
## options(mc.cores = parallel::detectCores()).
## To avoid recompilation of unchanged Stan programs, we recommend calling
## rstan_options(auto_write = TRUE)

##
## Attaching package: 'rstan'

## The following object is masked from 'package:magrittr':
##
##      extract

## The following object is masked from 'package:tidyr':
##
##      extract

hmc_model = stan_model("STA601_hw5_yw_hmc_norm.stan")

# data
n_A = n_B = 16
mu_A = 75.2; s_A = 7.3
mu_B = 77.5; s_B = 8.1

# HMC sampling
N = 10
Pr = c(rep(NA, N))
for (i in 1:N) {
  # prior
  mu0 = 75; s0 = 10
  k0 = v0 = 2^(i-1)
  # posterior
  k_n_A = k0 + n_A
  k_n_B = k0 + n_B
  mu_n_A = (k0*mu0 + n_A*mu_A) / k_n_A
  mu_n_B = (k0*mu0 + n_B*mu_B) / k_n_B
  v_n_A = v0 + n_A
  v_n_B = v0 + n_B
  s2_n_A = (v0*s0^2 + (n_A-1)*s_A^2 + k0*n_A*(mu_A-mu0)^2/k_n_A) / v_n_A
  s2_n_B = (v0*s0^2 + (n_B-1)*s_B^2 + k0*n_B*(mu_B-mu0)^2/k_n_B) / v_n_B

  sam = rstan::sampling(hmc_model, data = list(S = 1000,
                                                yA = yA,
                                                yB = yB,
                                                a_A = v_n_A/2,
                                                a_B = v_n_B/2,
                                                b_A = s2_n_A*v_n_A/2,
                                                b_B = s2_n_B*v_n_B/2,
                                                mu_A = mu_n_A,
                                                mu_B = mu_n_B),
                        chains = 1,
                        iter = 800,
                        warmup = 100,
                        verbose = F,

```

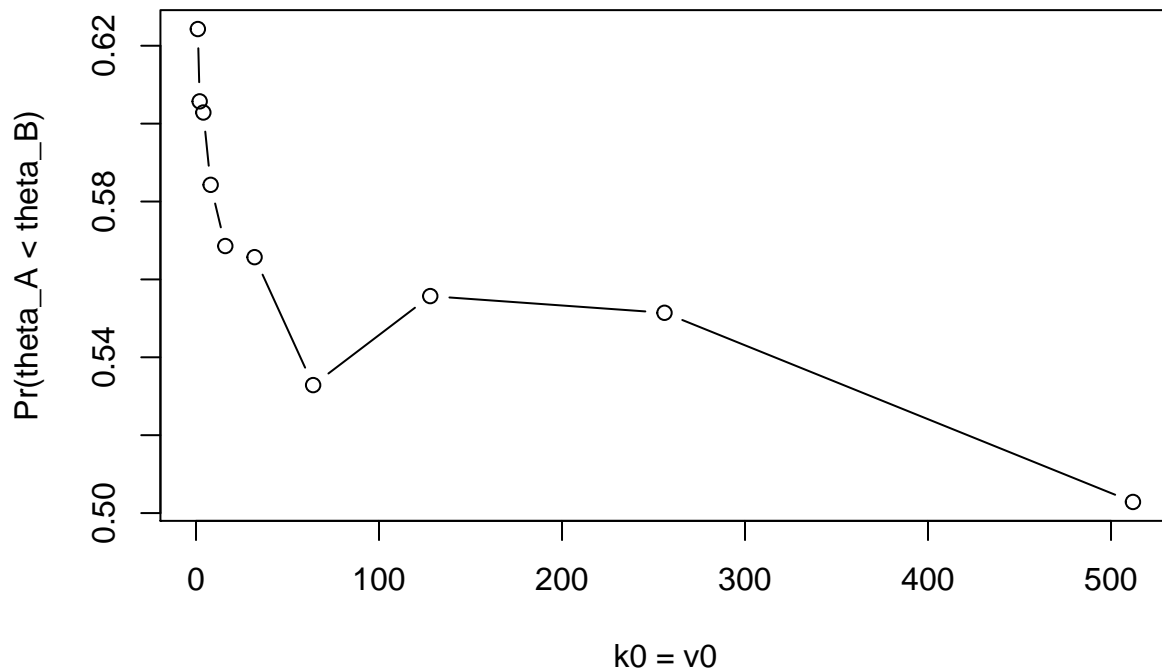
```

refresh = 0)
Pr[i] = mean(as.matrix(sam, pars = "theta_A") < as.matrix(sam, pars = "theta_B"))
}

plot(2^(0:(N-1)), Pr, "b", xlab = "k0 = v0", ylab = "Pr(theta_A < theta_B)",
     main="Probability of theta_A < theta_B with different prior using HMC")

```

Probability of $\theta_A < \theta_B$ with different prior using HMC



As shown in this plot, we can find that the result curve for HMC using only a few hundreds iterations is not that smooth as the result using MC with 10000 iters. It fluctuates much more than the MC result in part (a) and even yields unreasonable results for some runs. HMC uses a numeric approach to solve and sample the posterior distribution, which requires more iterations to fully explore the posterior parameter space. When the iterations are small, it might give weird results for some runs.

2. Jeffreys' prior for a normal (jointly for θ and σ^2) is proportional to $(\sigma^2)^{-3/2}$. Imagine you have n data points from a mean θ , variance σ^2 normal distribution. Derive the posterior under this prior and state whether it is proper. You can either integrate directly, or try to put it into a form of a distribution you can try to recognize (while in class we decided we won't use the Normal-Gamma and Normal-Inverse-Gamma distributions for sampling or discussions, if you can recognize the pdf of those in the posterior you can verify that you have "legal" parameters and then conclude whether the posterior is proper or not).

As shown in class, we know that the Jeffreys' prior for Gaussian distribution is $\sqrt{\det(I(\theta, \sigma^2))}$, which is $(\sigma^2)^{-3/2}$, given by the problem. Suppose we have some constant c that give our prior $p(\theta, \sigma^2) = c \times (\sigma^2)^{-3/2}$. Then we have,

$$\begin{aligned}
p(\theta, \sigma^2 | y_1, \dots, y_n) &\propto p(\theta, \sigma^2) p(y_1, \dots, y_n | \theta, \sigma^2) \\
&\propto (\sigma^2)^{-3/2} p(y_1, \dots, y_n | \theta, \sigma^2) \\
&\propto (\sigma^2)^{-3/2} (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta)^2 \right\} \\
&\propto (\sigma^2)^{-1/2} (\sigma^2)^{-(n+2)/2} \exp \left\{ -\frac{\sum_{i=1}^n y_i^2 - 2\theta \sum_{i=1}^n y_i + n\theta^2}{2\sigma^2} \right\} \\
&\propto \frac{1}{\sqrt{\sigma^2}} \left(\frac{1}{\sigma^2} \right)^{\frac{n}{2}+1} \exp \left\{ -\frac{(\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}) + n(\theta - \frac{\sum_{i=1}^n y_i}{n})^2}{2\sigma^2} \right\} \\
&\propto \text{N-}\Gamma^{-1}(\mu, \lambda, \alpha, \beta)
\end{aligned}$$

where

$$\begin{aligned}
\mu &= \frac{\sum_{i=1}^n y_i}{n} \\
\lambda &= n \\
\alpha &= \frac{n}{2} \\
\beta &= \frac{1}{2} \left(\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n} \right)
\end{aligned}$$

The posterior distribution for this Jeffreys' prior is a Normal-Inverse-Gamma distribution, with parameters $(\mu, \lambda, \alpha, \beta) = (\frac{\sum_{i=1}^n y_i}{n}, n, \frac{n}{2}, \frac{1}{2}(\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}))$. Thus, it is a proper posterior distribution.