Exercise 3.3.3

Saurabh Steixner-Kumar
(social): - -

Contents

1 Question

Exercise 3.3.3 Compare the data sets k1 = 7, n1 = 10, k2 = 3, n2 = 10 and k1 = 5, n1 = 10, k2 = 5, n2 = 10. Make sure, following on from the previous question, that you understand why the comparison works the way it does.

2 Comments/Solution

Comparing the two datasets in this exercise, we see that the common latent probability drives the two binomial processes and hence as the overall probability of both the datasets is essentially the same i.e. k=10 and n=20, we see identical posterior estimates of θ in both. Please check out the posterior plots section to visually compare this.

In other words: The results for these data sets will be exactly the same. Because the model assumes a common rate, both data sets can in fact be re-described as having k = k1 + k2 = 10, n = n1 + n2 = 20.

The model used to calculate the required values and the plots is scripted below. Copy/pasting the given code will generate the same result on your own machine.

3 Code

3.1 libraries

The libraries required for the script and the plots.

```
# clears workspace
rm(list=ls())
```

```r
#load libraries
library(rstan)
library(ggplot2)
library(patchwork)
```

3.2 Data

The data required for this particular stan model.

```r
# data  initialization
k1 <- 7;n1 <- 10;k2 <- 3;n2 <- 10
# to be passed on to Stan
stan_data <- list(k1 = k1, n1 = n1, k2 = k2, n2 = n2)
#
k1 <- 5;n1 <- 10;k2 <- 5;n2 <- 10
# to be passed on to Stan
stan_data_1 <- list(k1 = k1, n1 = n1, k2 = k2, n2 = n2)
```

3.3 Stan code

Stan code, that can be written in R as such or in a separate new file with stan extension.

```r
write("// Stan code here in this section

// Inferring the common rate theta
data {
  int<lower=1> n1;
  int<lower=1> n2;
  int<lower=0> k1;
  int<lower=0> k2;
}
parameters {
  real<lower=0,upper=1> theta;
}
model {
  // Prior on Single Rate Theta
  theta ~ beta(1, 1);
  // Observed Counts
  k1 ~ binomial(n1, theta);
  k2 ~ binomial(n2, theta);
} // ",

"3_3_3.stan")
```

3.4 code in R to run stan

Running stan through R (with the required input parameters).

```r
myinits <- list(
  list(theta=.1),  # chain 1 starting value
  list(theta=.9))  # chain 2 starting value

# parameters to be monitored:
parameters <- c("theta")
```

```
# The following command calls Stan with specific options.
# For a detailed description type "?stan".
mod_fit <- stan(file="3_3_3.stan",
                data=stan_data,
                init=myinits,  # If not specified, gives random inits
                pars=parameters,
                iter=2000,
                chains=2,
                thin=1,
                warmup=100,  # Stands for burn-in; Default = iter/2
                seed=123  # Setting seed; Default is random seed
)
mod_fit_1 <- stan(file="3_3_3.stan",
                data=stan_data_1,
                init=myinits,  # If not specified, gives random inits
                pars=parameters,
                iter=2000,
                chains=2,
                thin=1,
                warmup=100,  # Stands for burn-in; Default = iter/2
                seed=123  # Setting seed; Default is random seed
)
```

4 Outputs

4.1 Model summary

In order of definition.

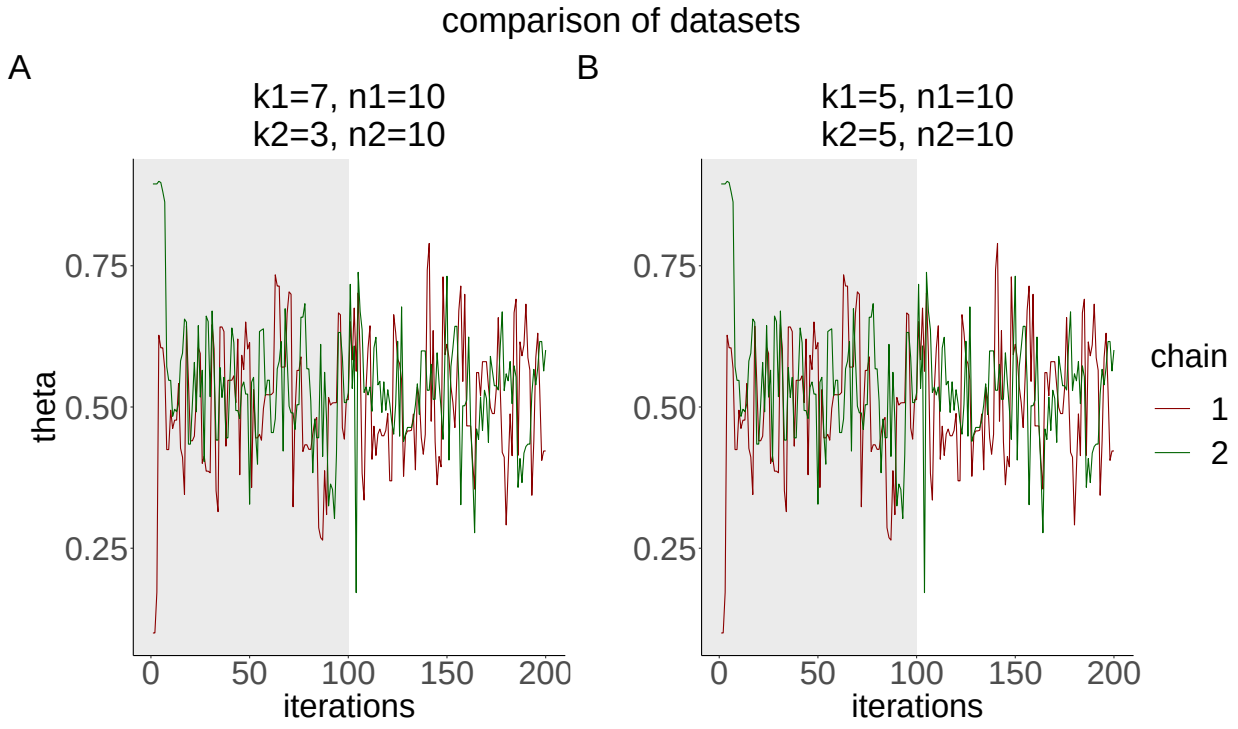
```
## Inference for Stan model: 3_3_3.
## 2 chains, each with iter=2000; warmup=100; thin=1;
## post-warmup draws per chain=1900, total post-warmup draws=3800.
##
##         mean se_mean  sd   2.5%    25%    50%    75%  97.5% n_eff Rhat
## theta   0.50    0.00 0.1   0.30   0.43   0.50   0.57   0.70  1513    1
## lp__  -15.76    0.02 0.7 -17.78 -15.93 -15.49 -15.30 -15.25  2169    1
##
## Samples were drawn using NUTS(diag_e) at Thu Nov 05 21:21:22 2020.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

```
## Inference for Stan model: 3_3_3.
## 2 chains, each with iter=2000; warmup=100; thin=1;
## post-warmup draws per chain=1900, total post-warmup draws=3800.
##
##         mean se_mean  sd   2.5%    25%    50%    75%  97.5% n_eff Rhat
## theta   0.50    0.00 0.1   0.30   0.43   0.50   0.57   0.70  1513    1
## lp__  -15.76    0.02 0.7 -17.78 -15.93 -15.49 -15.30 -15.25  2169    1
##
## Samples were drawn using NUTS(diag_e) at Thu Nov 05 21:21:22 2020.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

4.2 Plots

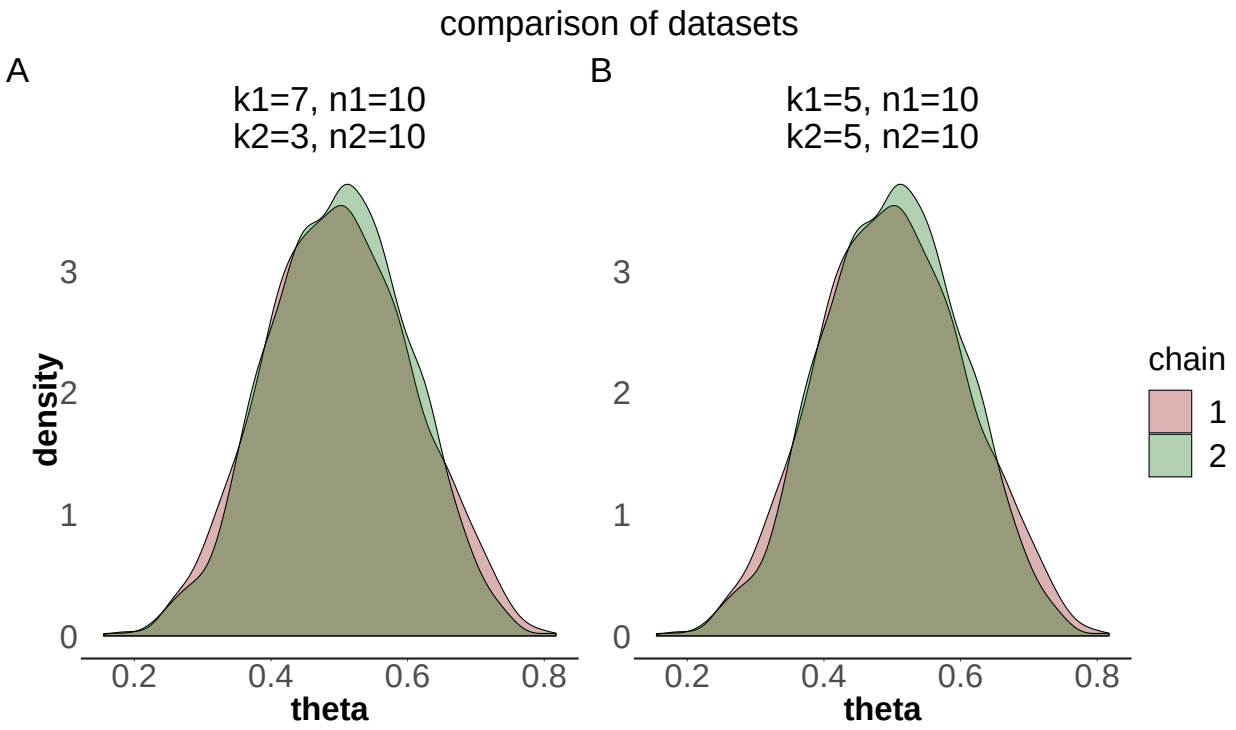
4.2.1 Plot (chains)

The initial movement of the chains are shown here (including the warmup phase). The two chains begin from the initial starting points of as defined in the input parameters of the stan model.



4.2.2 Plot (posterior)

The plot of the θ values per chain superimposed on each other.


comparison of datasets
A
k1=7, n1=10
k2=3, n2=10
B
k1=5, n1=10
k2=5, n2=10
density
theta
chain
1
2
0
1
2
3
0.2
0.4
0.6
0.8