Exercise 3.2.2

Saurabh Steixner-Kumar
(social):

Contents

1 Question

Exercise 3.2.2 Try the data k1 = 0, n1 = 1, and k2 = 0, n2 = 5. Can you explain the shape of the posterior for δ.

2 Comments/Solution

Here we compare the dataset k1 = 0, n1 = 1, and k2 = 0, n2 = 5 to a default dataset of k1 = 80, n1 = 100, and k2 = 70, n2 = 100 side by side to better understand the effect. We see (posterior plots below) that as the number of data samples are lower the posterior is very wide and uncertain. A higher number of n will make it narrower.

The key to understanding the posterior is that you can be relatively sure that θ_2 is small, but you cannot be so sure about the value of θ_1. This means $\theta_1 - \theta_2$ could be a large positive value, because θ_1 could be large and θ_2 small. But $\theta_1 - \theta_2$ cannot be a large negative value, since θ_2 is small. The asymmetry in the uncertainty about θ_1 and θ_2 creates the asymmetry evident in the posterior for the difference.

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.

```r
# clears workspace
rm(list=ls())
#load libraries
library(rstan)
library(ggplot2)
library(patchwork)
```

3.2 Data

The data required for this particular stan model.

```r
# data  initialization
k1 <- 80;n1 <- 100;k2 <- 70;n2 <- 100
# to be passed on to Stan
stan_data <- list(k1 = k1, n1 = n1, k2 = k2, n2 = n2)
#
k1 <- 0;n1 <- 1;k2 <- 0;n2 <- 5
# 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 delta through theta1 and theta2
data {
  int<lower=1> n1;
  int<lower=1> n2;
  int<lower=0> k1;
  int<lower=0> k2;
}
parameters {
  real<lower=0,upper=1> theta1;
  real<lower=0,upper=1> theta2;
}
transformed parameters {
  real<lower=-1,upper=1> delta;
  delta = theta1 - theta2;
}
model {
  // Prior Distribution for Rate Theta
  theta1 ~ beta(1, 1);
  theta2 ~ beta(1, 1);
  // Observed Counts
  k1 ~ binomial(n1, theta1);
  k2 ~ binomial(n2, theta2);
} // ",

"3_2_2.stan")
```

3.4 code in R to run stan

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

```
myinits <- list(
  list(theta1=.1,theta2=.9),  # chain 1 starting value
  list(theta1=.9,theta2=.1))  # chain 2 starting value

# parameters to be monitored:
parameters <- c("delta", "theta1", "theta2")

# The following command calls Stan with specific options.
# For a detailed description type "?stan".
mod_fit <- stan(file="3_2_2.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_2_2.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_2_2.
## 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
## delta     0.10    0.00 0.06   -0.02    0.06    0.10    0.14    0.22  3801
## theta1    0.79    0.00 0.04    0.71    0.77    0.80    0.82    0.87  3720
## theta2    0.70    0.00 0.05    0.60    0.67    0.70    0.73    0.78  4058
## lp__   -115.58    0.02 1.06 -118.38 -116.01 -115.24 -114.82 -114.53  1798
##        Rhat
## delta     1
## theta1    1
## theta2    1
## lp__      1
##
```

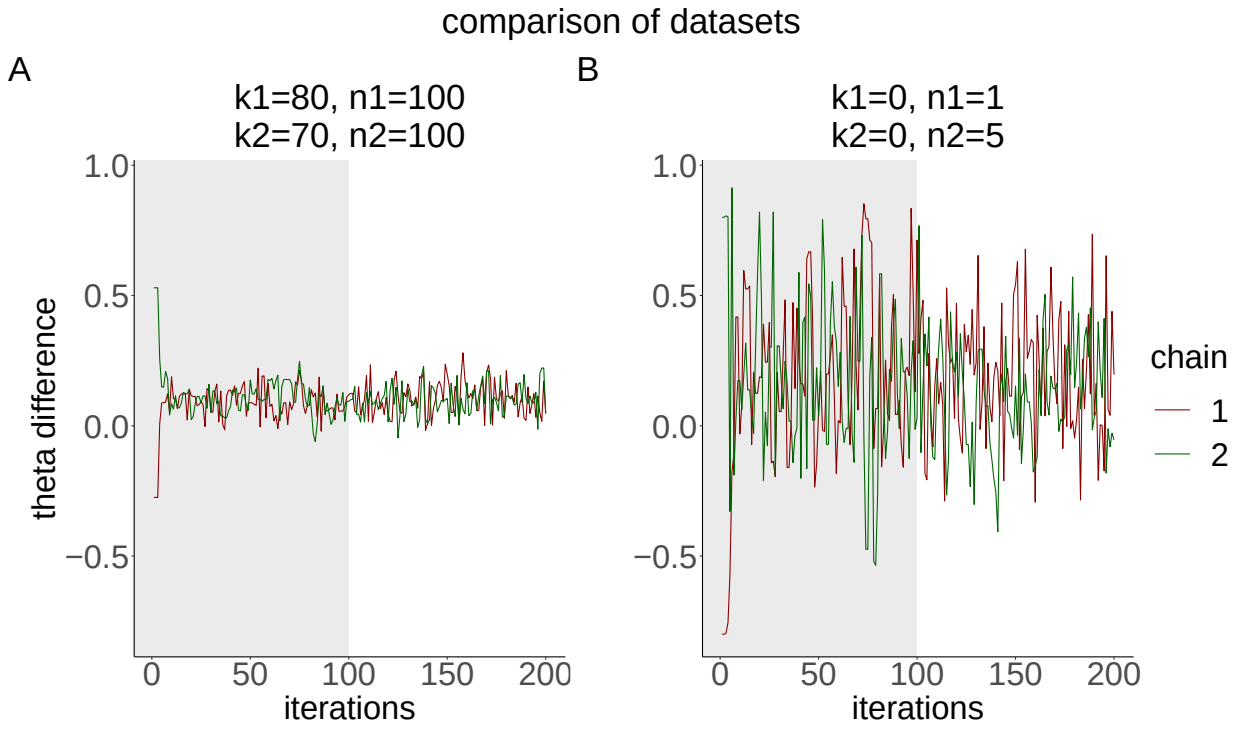
```
## Samples were drawn using NUTS(diag_e) at Thu Nov 05 21:17:00 2020.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

```
## Inference for Stan model: 3_2_2.
## 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
## delta   0.19    0.01 0.26 -0.29  0.00  0.16  0.36  0.74  2370    1
## theta1  0.33    0.00 0.23  0.01  0.13  0.29  0.49  0.84  2246    1
## theta2  0.14    0.00 0.13  0.00  0.04  0.11  0.21  0.47  3295    1
## lp__   -5.99    0.03 1.20 -9.20 -6.47 -5.64 -5.13 -4.81  1193    1
##
## Samples were drawn using NUTS(diag_e) at Thu Nov 05 21:17:01 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

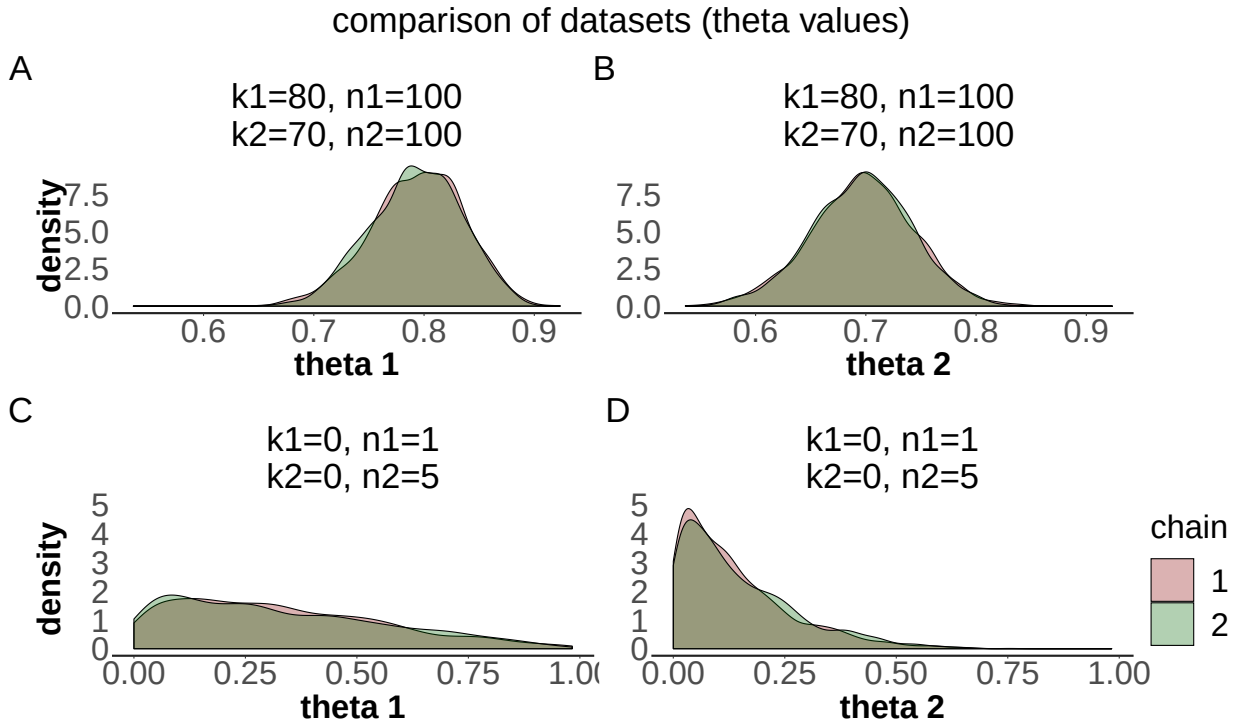
4.2.1 Plot (chains)

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_1 and θ_2 values per chain superimposed on each other.



4.2.3 Plot (posterior)

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