Exercise 3.1.1

Saurabh Steixner-Kumar
(social): - -

Contents

1 Question

Exercise 3.1.1 Carefully consider the posterior distribution for θ given k = 5 successes out of n = 10 trials. Based on a visual impression, what is your estimate of the probability that the rate θ is higher than 0.4 but smaller than 0.6? How did you arrive at your estimate?

2 Comments/Solution

Looking at the posterior distribution (from the plots section below), one can estimate that θ has a very high probability of being between the values of 0.4 and 0.6. We can calculate the exact value from the estimates as:

```
length(theta[theta>0.4 & theta<0.6])/length(theta)
```

The value is given in the last section of this page under 'calculation'. 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
k <- 5
n <- 10
# to be passed on to Stan
stan_data <- list(k = k, n = n)
```

3.3 Stan code

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

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

// Inferring theta
data {
  int<lower=1> n;
  int<lower=0> k;
}
parameters {
  real<lower=0,upper=1> theta;
}
model {
  // Prior Distribution for theta
  theta ~ beta(1, 1);

  // Observed Counts
  k ~ binomial(n, theta);
} // ",

"3_1_1.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_1_1.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
)
```

4 Outputs

4.1 Model summary

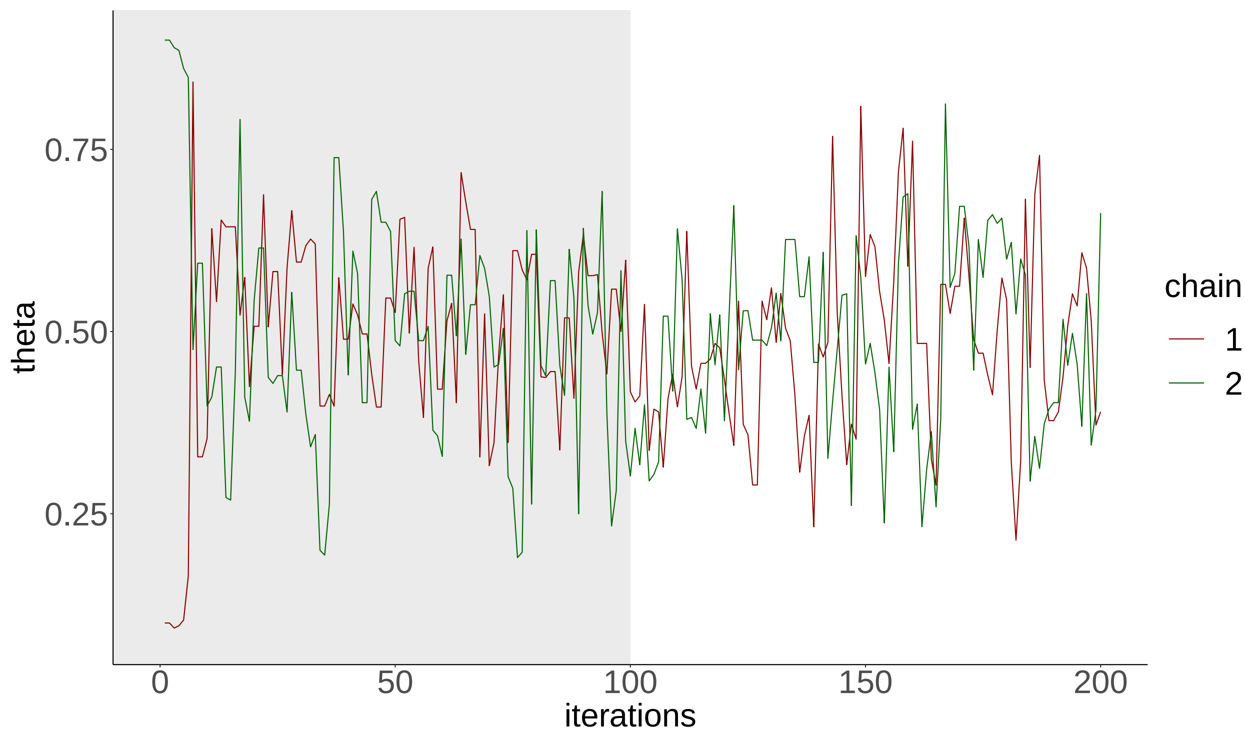
In order of definition.

```
## Inference for Stan model: 3_1_1.
## 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  0.49    0.00 0.14   0.23  0.40  0.49  0.59  0.76  1464    1
## lp__  -8.83    0.02 0.73 -10.88 -8.98 -8.55 -8.37 -8.32  1475    1
##
## Samples were drawn using NUTS(diag_e) at Wed Oct 21 15:54:59 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

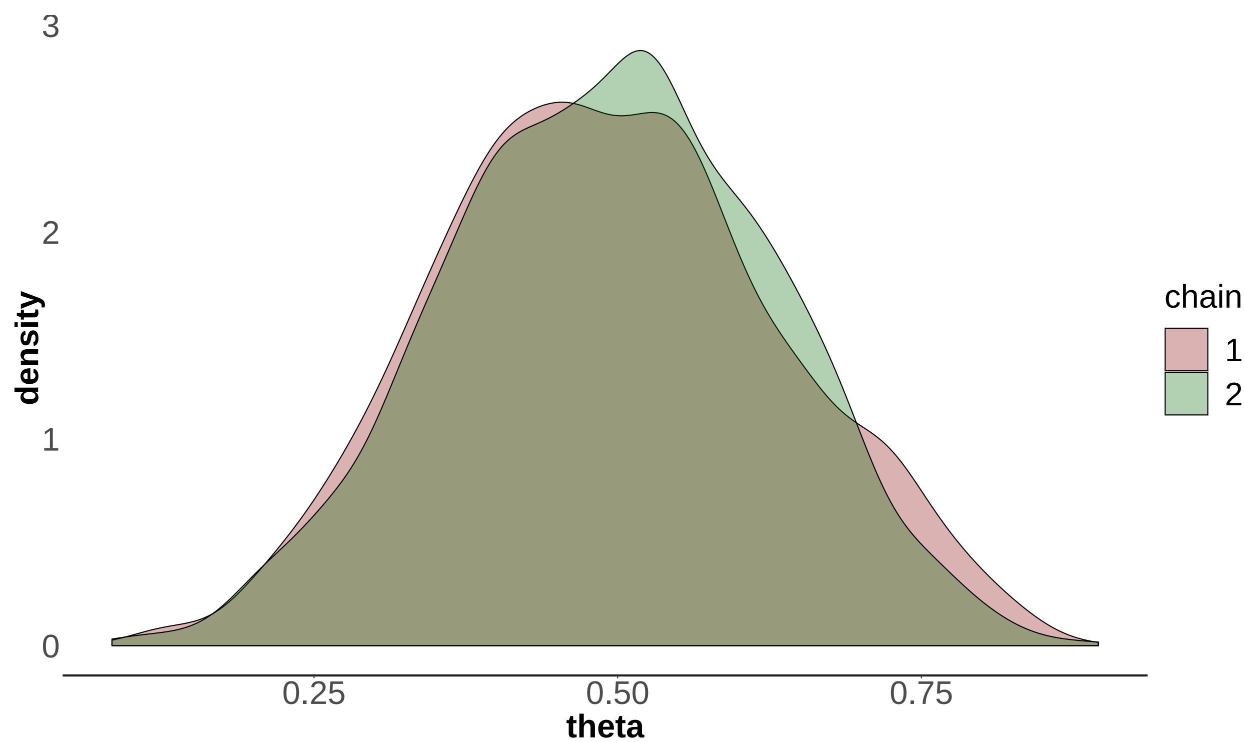
4.2.1 Plot (chains)

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.



5 Calculation

Calculation from the comment/solution.

```
length(theta[theta>0.4 & theta<0.6])/length(theta)
```

```
## [1] 0.5171053
```