Exercise 3.4.2

Saurabh Steixner-Kumar
(social): - -

Contents

1 Question

Exercise 3.4.2 Try different priors on θ, by changing $\theta \sim \text{Beta}(1,1)$ to $\theta \sim \text{Beta}(10,10)$, $\theta \sim \text{Beta}(1,5)$, and $\theta \sim \text{Beta}(0.1,0.1)$. Use the figures produced to understand the assumptions these priors capture, and how they interact with the same data to produce posterior inferences and predictions.

2 Comments/Solution

The posterior distributions are dependent on the priors given for the inference. The chose prior can have huge implications as they may mean different things altogether. If the observed data are infinite (or at least a very huge number), then all posterior distribution will always converge to the true value. But, practically this is not the scenario, we do not have the resources to observe infinite data and hence we should carefully choose our prior. In this exercise, we have tested 4 different priors and see different outcomes for the posterior. For our data and number of iterations in the model fit, we observe slightly similar posteriors. Please have a look at the plots and specially the last plot comparing the 4 different posteriors in one plot.

One of the nice properties of using the $\theta \sim \text{Beta}(\alpha, \beta)$ prior distribution for a rate θ, is that it has a natural interpretation. The α and β values can be thought of as counts of "prior successes" and "prior failures", respectively. This means, using a $\theta \sim \text{Beta}(3,1)$ prior corresponds to having the prior information that 4 previous observations have been made, and 3 of them were successes. Or, more elaborately, starting with a $\theta \sim \text{Beta}(3,1)$ is the same as starting with a $\theta \sim \text{Beta}(1,1)$, and then seeing data giving two more successes (i.e., the posterior distribution in the second scenario will be same as the prior distribution in the first). As always in Bayesian analysis, inference starts with prior information, and updates that information—by changing the probability distribution representing the uncertain information—as more information becomes available. When a type of likelihood function (in this case, the Binomial) does not change the type of distribution (in this case, the Beta) going from the posterior to the prior, they are said to have a "conjugate" relationship. This is valued a lot in analytic approaches to Bayesian inference, because it makes for tractable calculations. It is not so important for that reason in computational approaches, because sampling methods can handle easily much more general relationships between parameter distributions and likelihood functions. But conjugacy is still useful in computational approaches because of the natural semantics it gives in setting prior distributions.

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(tidyr)
library(ggplot2)
library(patchwork)
```

3.2 Data

The data required for this particular stan model.

```r
# data  initialization
k <- 5;n <- 15
# 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 separate new file with stan extension.

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

// Prior and Posterior Prediction
data {
  int<lower=1> n;
  int<lower=0> k;
}
parameters {
  real<lower=0,upper=1> theta1;
  real<lower=0,upper=1> theta2;
  real<lower=0,upper=1> theta3;
  real<lower=0,upper=1> theta4;
  real<lower=0,upper=1> thetaprior1;
  real<lower=0,upper=1> thetaprior2;
  real<lower=0,upper=1> thetaprior3;
  real<lower=0,upper=1> thetaprior4;
}
model {
  // Prior on theta
  theta1 ~ beta(1, 1);
  theta2 ~ beta(10, 10);
```

```
  theta3 ~ beta(1, 5);
  theta4 ~ beta(0.1, 0.1);
  thetaprior1 ~ beta(1, 1);
  thetaprior2 ~ beta(10, 10);
  thetaprior3 ~ beta(1, 5);
  thetaprior4 ~ beta(0.1, 0.1);
  // Observed Data
  k ~ binomial(n, theta1);
  k ~ binomial(n, theta2);
  k ~ binomial(n, theta3);
  k ~ binomial(n, theta4);
}
generated quantities {
  int<lower=0> postpredk1;
  int<lower=0> postpredk2;
  int<lower=0> postpredk3;
  int<lower=0> postpredk4;
  int<lower=0> priorpredk1;
  int<lower=0> priorpredk2;
  int<lower=0> priorpredk3;
  int<lower=0> priorpredk4;

  // Posterior Predictive
  postpredk1 = binomial_rng(n, theta1);
  postpredk2 = binomial_rng(n, theta2);
  postpredk3 = binomial_rng(n, theta3);
  postpredk4 = binomial_rng(n, theta4);
  // Prior Predictive
  priorpredk1 = binomial_rng(n, thetaprior1);
  priorpredk2 = binomial_rng(n, thetaprior2);
  priorpredk3 = binomial_rng(n, thetaprior3);
  priorpredk4 = binomial_rng(n, thetaprior4);
} // ",

"3_4_2.stan")
```

3.4 code in R to run stan

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

```
myinits <- list(
  list(theta1=.5,theta2=.5,theta3=.5,theta4=.5,thetaprior1=.5,thetaprior2=.5,thetaprior3=.5,thetaprior4
  list(theta1=.5,theta2=.5,theta3=.5,theta4=.5,thetaprior1=.5,thetaprior2=.5,thetaprior3=.5,thetaprior4

# parameters to be monitored:
parameters <- c("theta1", "theta2", "theta3", "theta4", "thetaprior1", "thetaprior2", "thetaprior3", "t

# The following command calls Stan with specific options.
# For a detailed description type "?stan".
mod_fit <- stan(file="3_4_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
)
```

4 Outputs

4.1 Model summary

In order of definition.

```
## Inference for Stan model: 3_4_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
## theta1          0.35    0.00 0.11   0.15   0.27   0.34   0.43   0.58  4879
## theta2          0.43    0.00 0.08   0.27   0.37   0.43   0.48   0.59  3957
## theta3          0.29    0.00 0.10   0.11   0.21   0.28   0.35   0.50  4691
## theta4          0.34    0.00 0.12   0.14   0.25   0.33   0.42   0.59  4684
## thetaprior1     0.51    0.01 0.29   0.03   0.25   0.51   0.77   0.97  2548
## thetaprior2     0.50    0.00 0.11   0.28   0.42   0.50   0.58   0.71  4523
## thetaprior3     0.17    0.00 0.13   0.01   0.06   0.13   0.24   0.50  3250
## thetaprior4     0.52    0.02 0.46   0.00   0.00   0.59   1.00   1.00   574
## postpredk1      5.34    0.04 2.42   1.00   4.00   5.00   7.00  10.00  3805
## postpredk2      6.43    0.04 2.27   2.00   5.00   6.00   8.00  11.00  3913
## postpredk3      4.23    0.04 2.25   1.00   3.00   4.00   6.00   9.00  3836
## postpredk4      5.06    0.04 2.55   1.00   3.00   5.00   7.00  10.00  4220
## priorpredk1     7.62    0.09 4.61   0.00   4.00   8.00  12.00  15.00  2634
## priorpredk2     7.43    0.04 2.49   3.00   6.00   7.00   9.00  12.00  4071
## priorpredk3     2.54    0.04 2.46   0.00   1.00   2.00   4.00   9.00  3552
## priorpredk4     7.73    0.28 6.87   0.00   0.00   9.00  15.00  15.00   583
## lp__          -79.89    0.07 2.16 -84.96 -81.19 -79.58 -78.30 -76.63  1034
##               Rhat
## theta1           1
## theta2           1
## theta3           1
## theta4           1
## thetaprior1      1
## thetaprior2      1
## thetaprior3      1
## thetaprior4      1
## postpredk1       1
## postpredk2       1
## postpredk3       1
## postpredk4       1
## priorpredk1      1
## priorpredk2      1
## priorpredk3      1
## priorpredk4      1
## lp__             1
##
## Samples were drawn using NUTS(diag_e) at Thu Nov 05 21:24:10 2020.
```

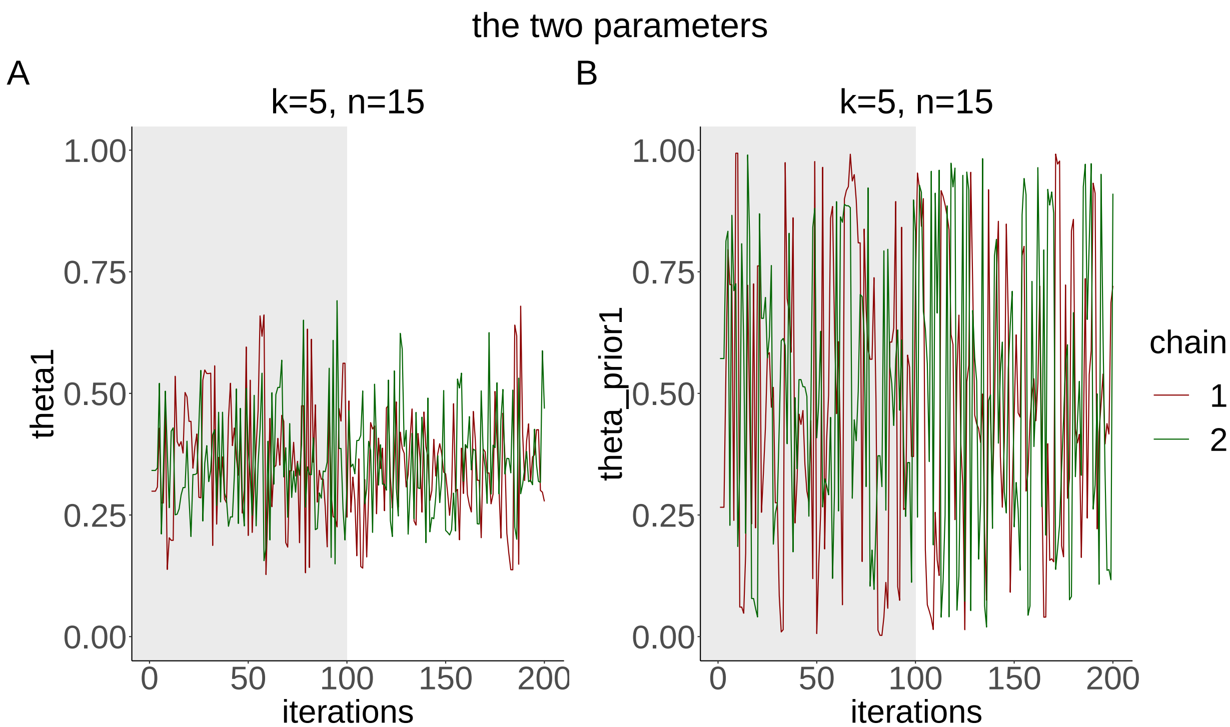
```
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

4.2 Plots

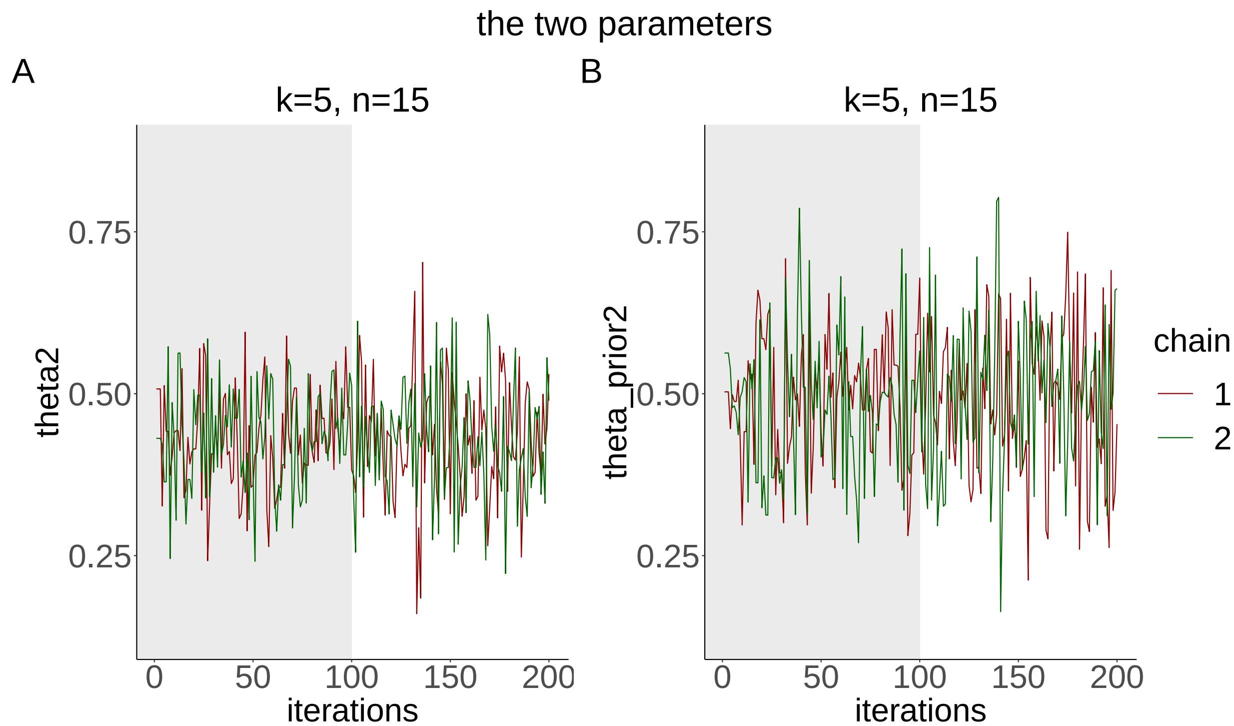
4.2.1 Plot (chains)

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.

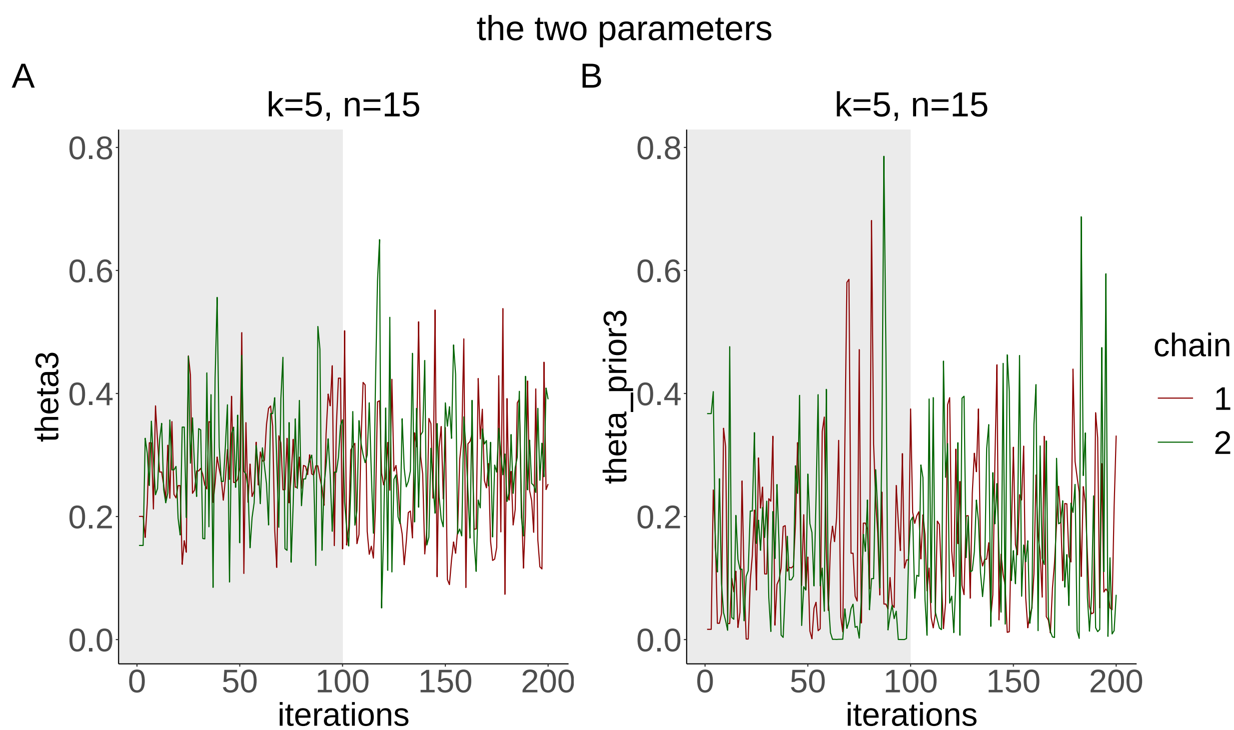
4.2.1.1 plot for θ ~Beta(1,1)



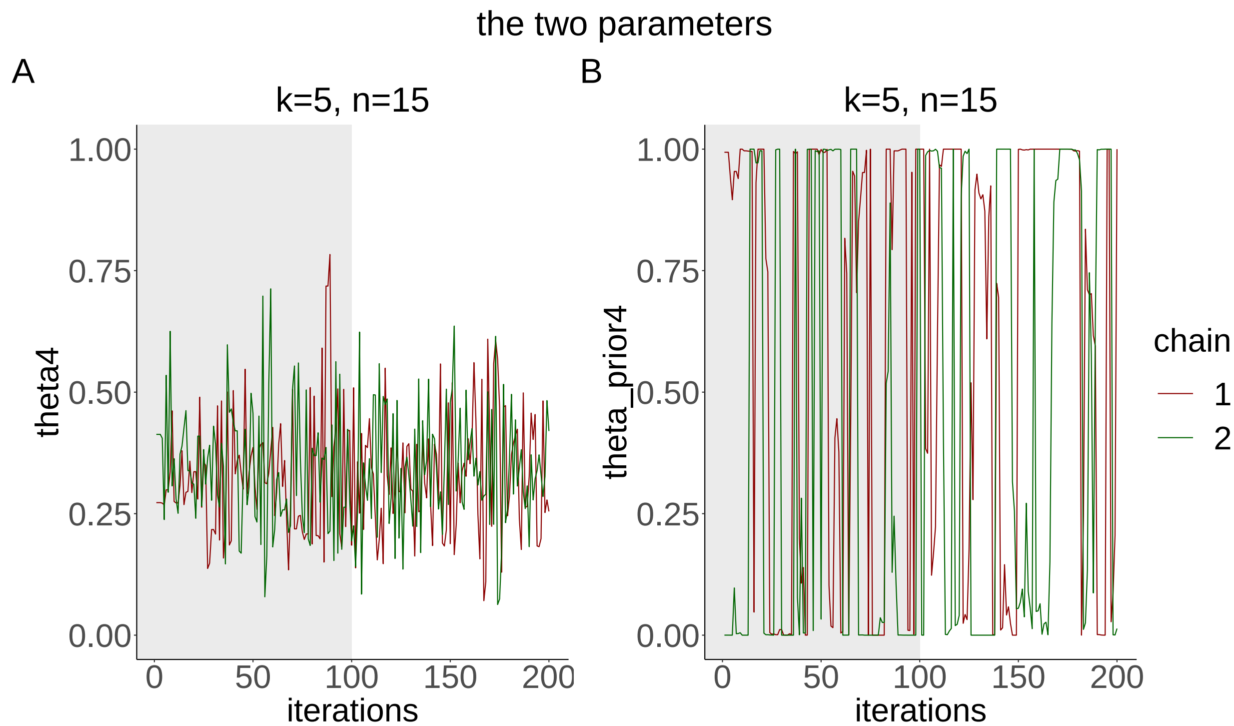
4.2.1.2 plot for θ ~Beta(10,10)



4.2.1.3 plot for θ ~Beta(1,5)



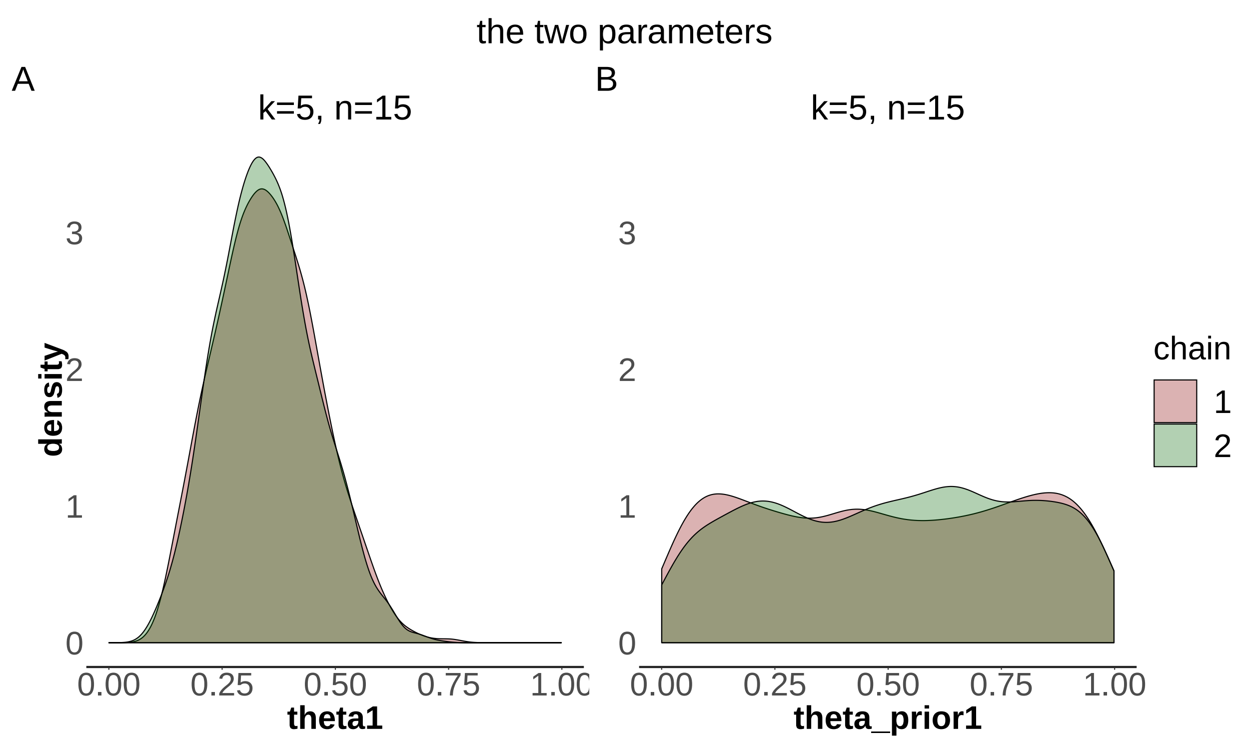
4.2.1.4 plot for θ ~Beta(0.1,0.1)



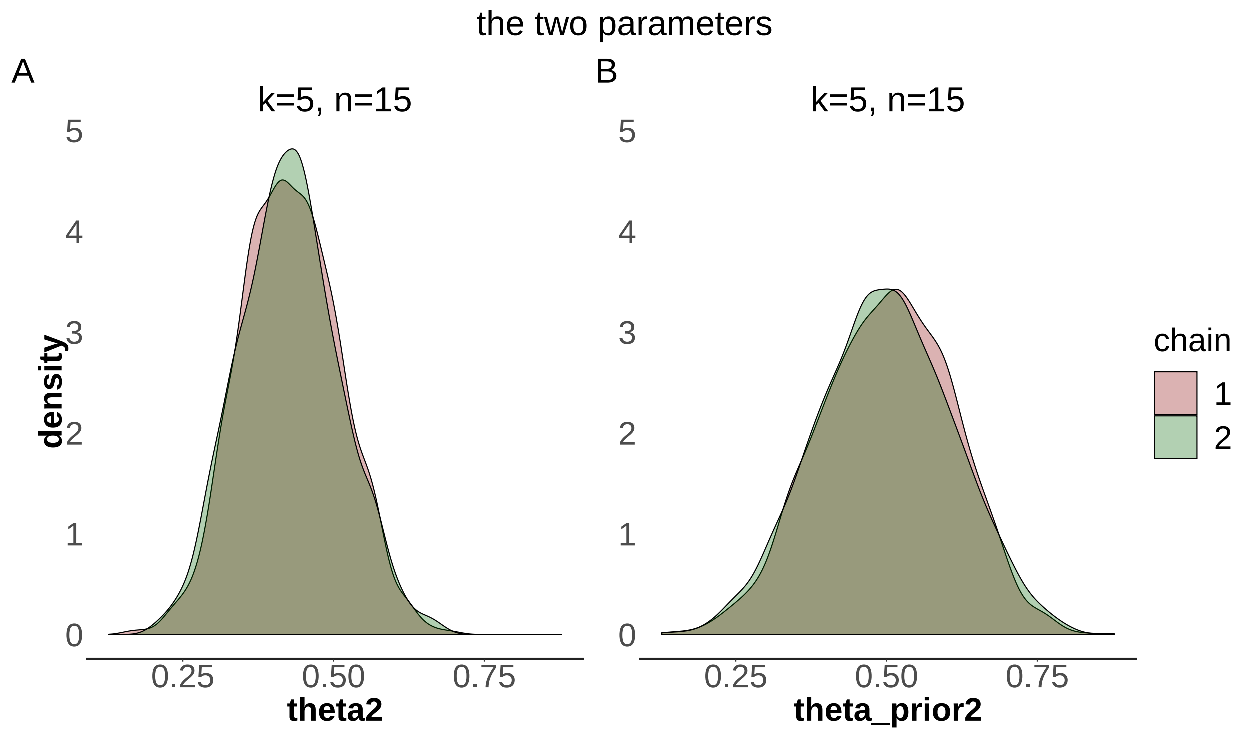
4.2.2 Plot (posterior)

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

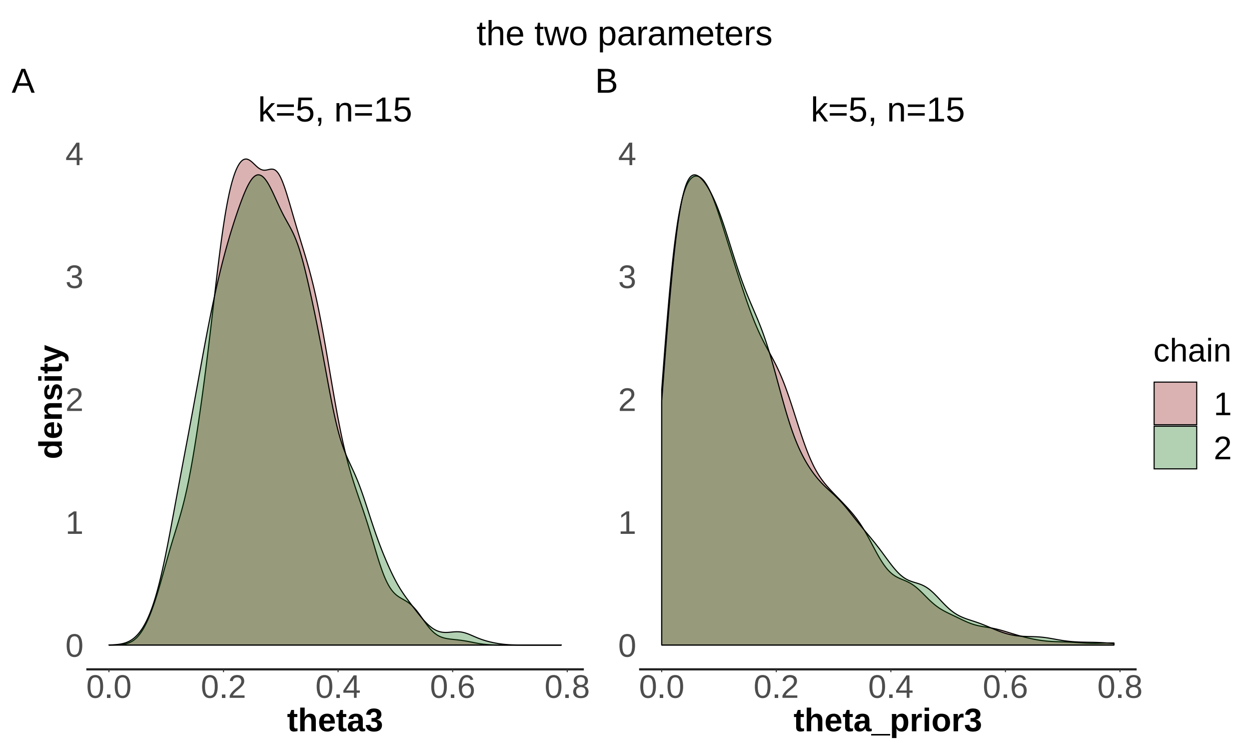
4.2.2.1 plot for θ ~Beta(1,1)



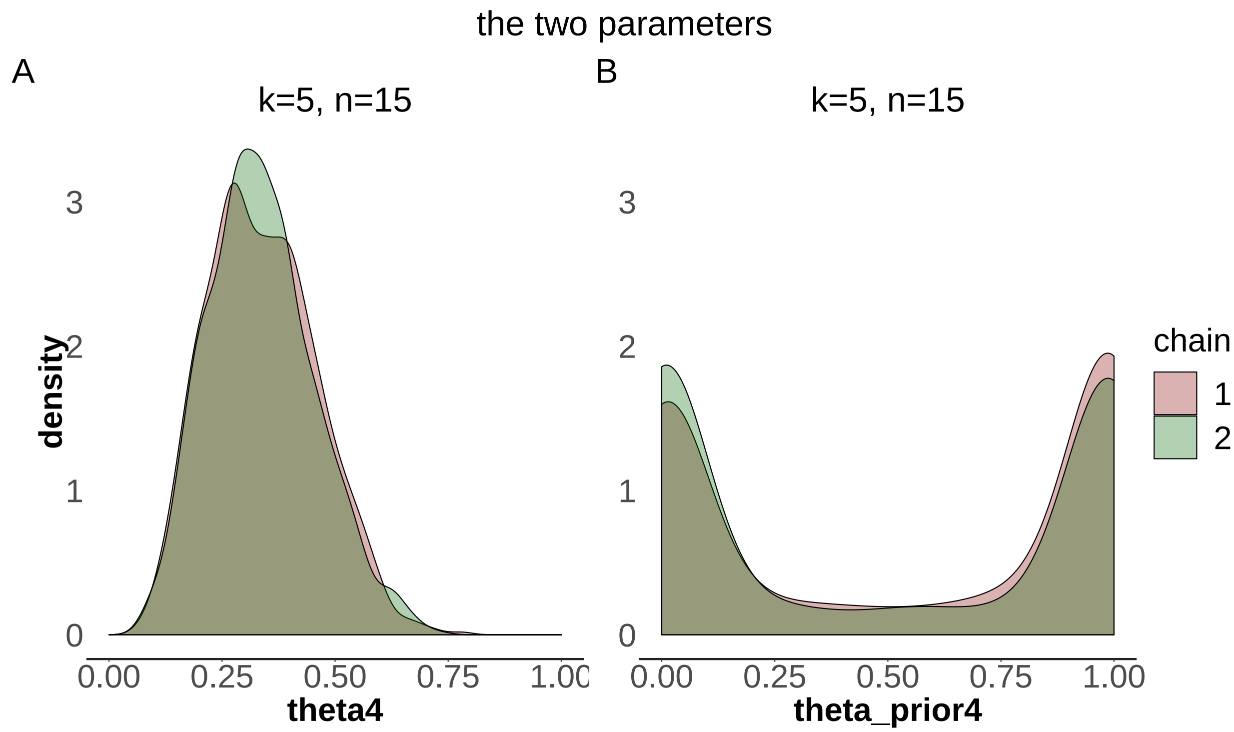
4.2.2.2 plot for θ ~Beta(10,10)



4.2.2.3 plot for θ ~Beta(1,5)



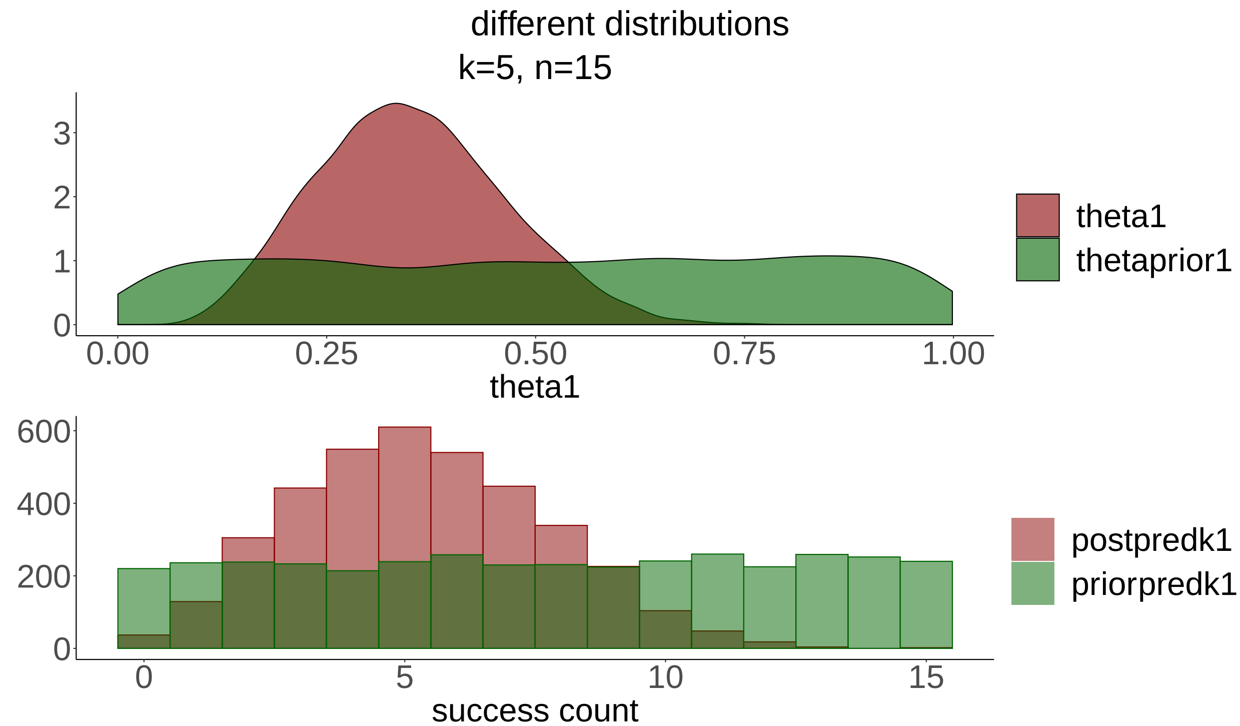
4.2.2.4 plot for θ ~Beta(0.1,0.1)



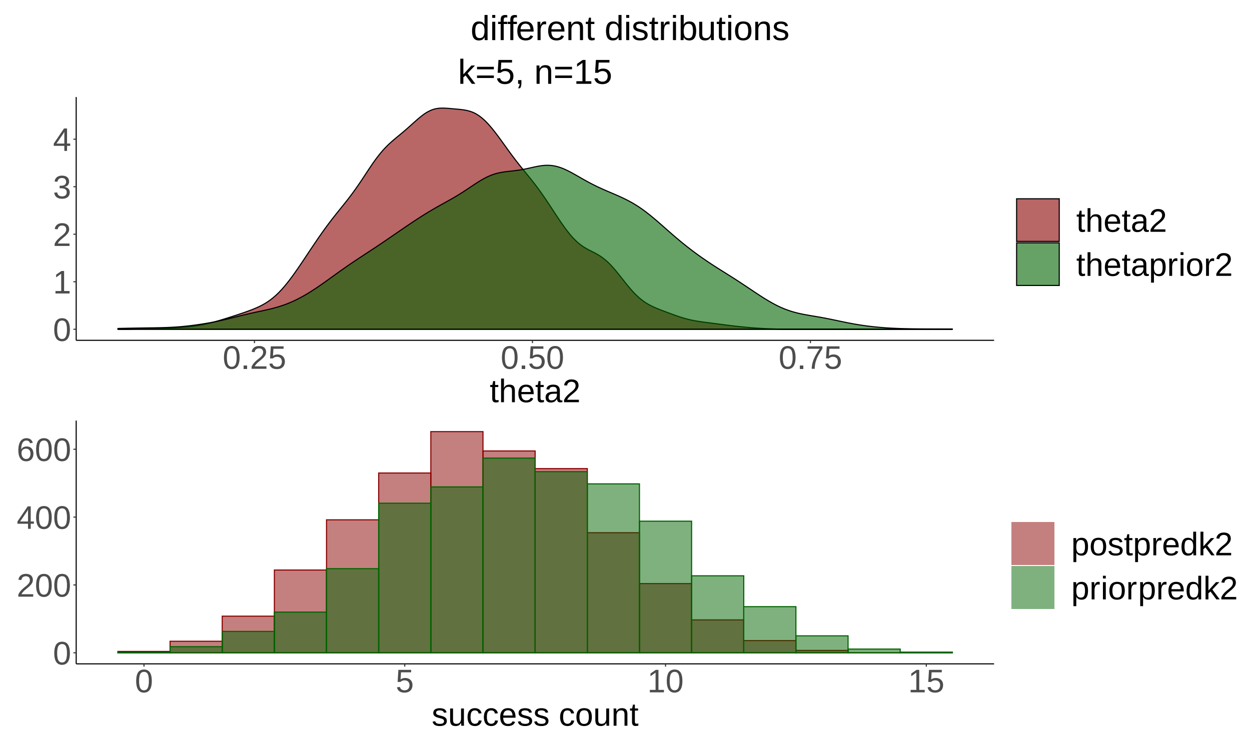
4.2.3 Plot (different distributions)

The combined plot of the different distributions superimposed on each other.

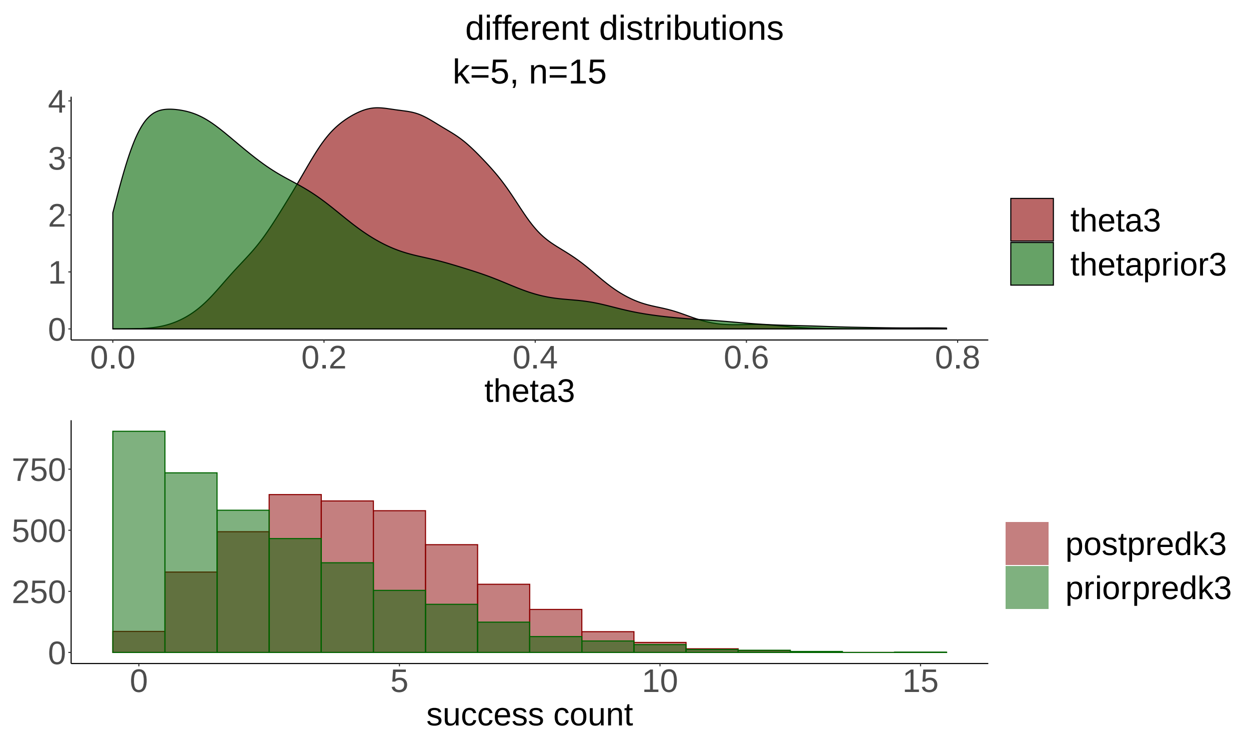
4.2.3.1 plot for θ ~ Beta(1,1)



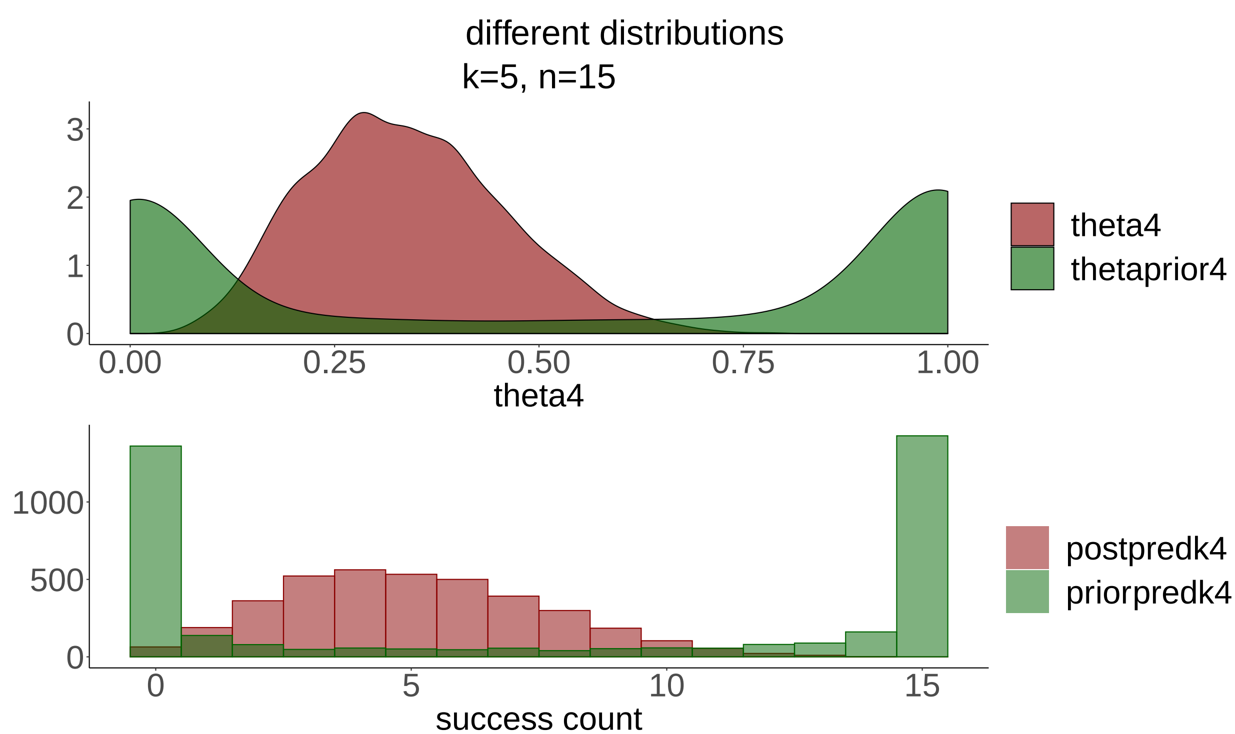
4.2.3.2 plot for $\theta \sim$ Beta(10,10)



4.2.3.3 plot for $\theta \sim$ Beta(1,5)



4.2.3.4 plot for $\theta \sim$ Beta(0.1,0.1)



4.2.4 Plot (different posteriors together)

The combined plot of the different posterior distributions superimposed on each other. These are for $\theta \sim$ Beta(1,1), $\theta \sim$ Beta(10,10), $\theta \sim$ Beta(1,5) and $\theta \sim$ Beta(0.1,0.1)