Exercise 3.4.1

Saurabh Steixner-Kumar
(social):

Contents

1 Question

Exercise 3.4.1 Make sure you understand the prior, posterior, prior predictive and posterior predictive distributions, and how they relate to each other (e.g., why is the top panel of Figure 3.9 a line plot, while the bottom panel is a bar graph?). Understanding these ideas is a key to understanding Bayesian analysis. Check your understanding by trying other data sets, varying both k and n.

2 Comments/Solution

The top panel represents the density (positive hits) of the parameters while the bottom panel is the number of counts of the prior and posterior positive hits. This is based on the given data for 5 positive hit out of 15. Look at the plots section (below) to visually inspect it further. You will also see the trace chain for the theta_prior all over the place, showing a flat prior (density plot clears it further).

Line plots are for continuous quantities (e.g., rate parameter θ) and bar plots are for discrete quantities (e.g., success counts of data).

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> theta;
  real<lower=0,upper=1> thetaprior;
}
model {
  // Prior on theta
  theta ~ beta(1, 1);
  thetaprior ~ beta(1, 1);
  // Observed Data
  k ~ binomial(n, theta);
}
generated quantities {
  int<lower=0> postpredk;
  int<lower=0> priorpredk;

  // Posterior Predictive
  postpredk = binomial_rng(n, theta);
  // Prior Predictive
  priorpredk = binomial_rng(n, thetaprior);
} // ",

"3_4_1.stan")
```

3.4 code in R to run stan

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

```
myinits <- list(
  list(theta=.5,thetaprior=.5),  # chain 1 starting value
  list(theta=.5,thetaprior=.5))  # chain 2 starting value

# parameters to be monitored:
parameters <- c("theta", "thetaprior", "postpredk", "priorpredk")

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

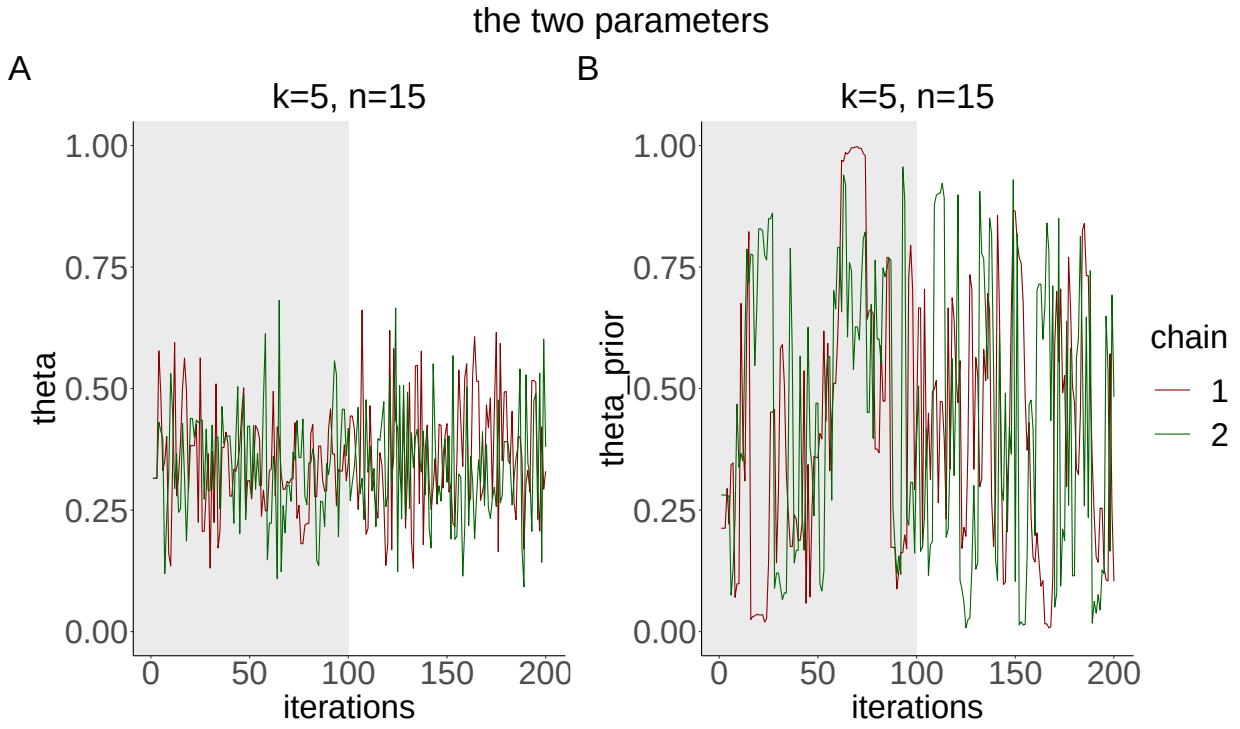
In order of definition.

```
## Inference for Stan model: 3_4_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
## theta        0.35    0.00 0.11   0.15   0.27   0.34   0.43   0.59  4232
## thetaprior   0.49    0.01 0.29   0.02   0.24   0.50   0.75   0.97  1211
## postpredk    5.29    0.04 2.45   1.00   4.00   5.00   7.00  10.00  3846
## priorpredk   7.43    0.13 4.63   0.00   3.00   7.00  11.00  15.00  1356
## lp__       -13.57    0.04 1.14 -16.76 -14.00 -13.23 -12.77 -12.45   778
##            Rhat
## theta      1.00
## thetaprior 1.00
## postpredk  1.00
## priorpredk 1.00
## lp__       1.01
##
## Samples were drawn using NUTS(diag_e) at Thu Nov 05 21:22:48 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

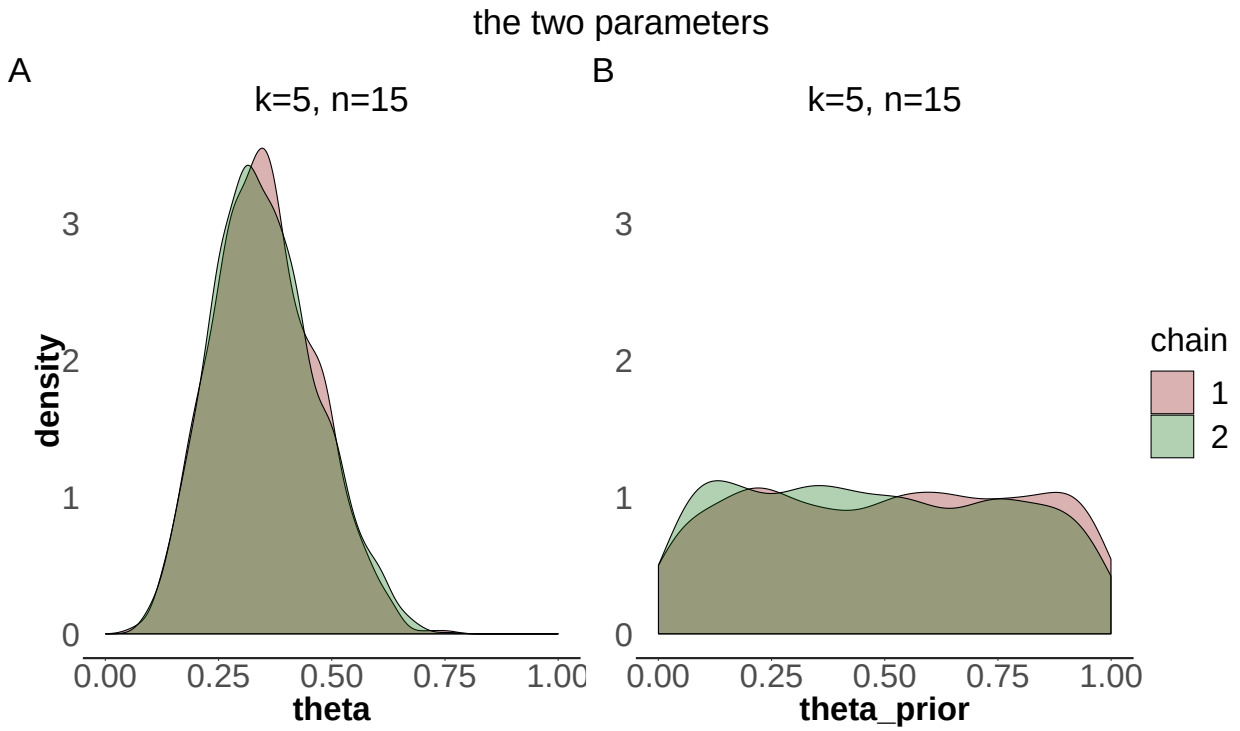
4.2.1 Plot (chains)

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.



4.2.3 Plot (different distributions)

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