Exercise 5.3.3

Saurabh Steixner-Kumar
(social): - -

Contents

1 Question

Exercise 5.3.3 Rare Disease Suppose you are testing a cheap instrument for detecting a rare medical condition. After 170 patients have been screened, the test results show that 157 did not have the condition, but 13 did. The expensive ground truth assessment subsequently revealed that, in fact, none of the patients had the condition. These data correspond to a = 0, b = 0, c = 13, d = 157. Apply the kappa graphical model to these data, and reach a conclusion about the usefulness of the cheap instrument. What is special about this data set, and what does it demonstrate about the Bayesian approach?

2 Comments/Solution

The posterior mean for kappa is approximately .05, with a 95% credible interval (ETI) of approximately (0, .23). The data are noteworthy because the disease has never been observed, so there are two zero cells, and a zero column sum. This poses a challenge for frequentist estimators. In order to deal with the problem of zero counts a frequentist may add a "1" to each cell in the design, but this amounts to fabricating data. An attractive property of the Bayesian approach is that it is always possible to do the analysis.

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

3.2 Data

The data required for this particular stan model.

```r
# data  initialization
# Rare Disease
y <- c(0, 0, 13, 157)
# to be passed on to Stan
stan_data <- list(y=y)
```

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

// Kappa Coefficient of Agreement
data {
  int<lower=0> y[4];
}
parameters {
  // Underlying Rates
  // Rate Objective Method Decides 'one'
  real<lower=0,upper=1> alpha;
  // Rate Surrogate Method Decides 'one' When Objective Method Decides 'one'
  real<lower=0,upper=1> beta;
  // Rate Surrogate Method Decides 'zero' When Objective Method Decides 'zero'
  real<lower=0,upper=1> gamma;
}
transformed parameters {
  simplex[4] pi;
  real xi;
  real psi;
  real kappa;
  // Probabilities For Each Count
  pi[1] = alpha * beta;
  pi[2] = alpha * (1 - beta);
  pi[3] = (1 - alpha) * (1 - gamma);
  pi[4] = (1 - alpha) * gamma;

  // Derived Measures
  // Rate Surrogate Method Agrees With the Objective Method
  xi = alpha * beta + (1 - alpha) * gamma ;
  // Rate of Chance Agreement
  psi = (pi[1] + pi[2]) * (pi[1] + pi[3]) + (pi[2] + pi[4]) * (pi[3] + pi[4]);
```

```
  // Chance-Corrected Agreement
  kappa = (xi - psi) / (1 - psi);
}
model {
  alpha ~ beta(1, 1);  // could be removed
  beta ~ beta(1, 1);  // could be removed
  gamma ~ beta(1, 1);  // could be removed
  // Count Data
  y ~ multinomial(pi);
} // ",

"5_3_3.stan")
```

3.4 code in R to run stan

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

```r
myinits <- list(
  list(alpha=.5, beta=.5, gamma=.5),  # chain 1 starting value
  list(alpha=.5, beta=.5, gamma=.5))  # chain 2 starting value

# parameters to be monitored:
parameters <- c("kappa", "xi", "psi", "alpha", "beta", "gamma", "pi")

# The following command calls Stan with specific options.
# For a detailed description type "?stan".
mod_fit <- stan(file="5_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
)
```

4 Outputs

4.1 Model summary

In order of definition.

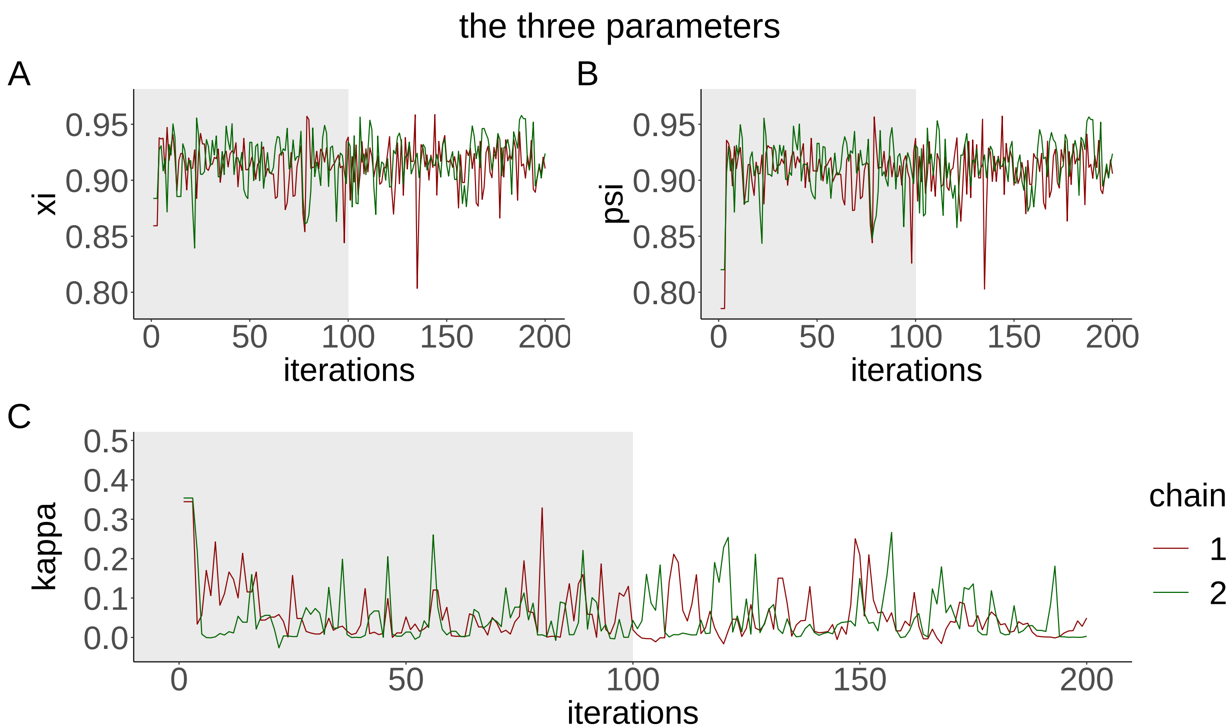
```
## Inference for Stan model: 5_3_3.
## 2 chains, each with iter=2000; warmup=100; thin=1;
## post-warmup draws per chain=1900, total post-warmup draws=3800.
##
##          mean se_mean   sd  2.5%   25%   50%   75% 97.5% n_eff Rhat
## kappa    0.05    0.00 0.06  0.00  0.01  0.03  0.07  0.23  2150    1
## xi       0.92    0.00 0.02  0.87  0.90  0.92  0.93  0.95  4359    1
## psi      0.91    0.00 0.02  0.86  0.90  0.91  0.93  0.95  4041    1
## alpha    0.01    0.00 0.01  0.00  0.00  0.00  0.01  0.02  2264    1
## beta     0.50    0.01 0.28  0.03  0.26  0.50  0.75  0.98  1543    1
## gamma    0.92    0.00 0.02  0.87  0.91  0.92  0.93  0.95  4462    1
```

```
## pi[1]    0.00    0.00 0.00   0.00   0.00   0.00   0.00   0.01  2203    1
## pi[2]    0.00    0.00 0.00   0.00   0.00   0.00   0.00   0.01  1790    1
## pi[3]    0.08    0.00 0.02   0.05   0.07   0.08   0.09   0.13  4462    1
## pi[4]    0.91    0.00 0.02   0.87   0.90   0.91   0.93   0.95  4238    1
## lp__   -57.72    0.05 1.32 -61.07 -58.32 -57.37 -56.74 -56.20   794    1
##
## Samples were drawn using NUTS(diag_e) at Thu Nov 05 21:28:33 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

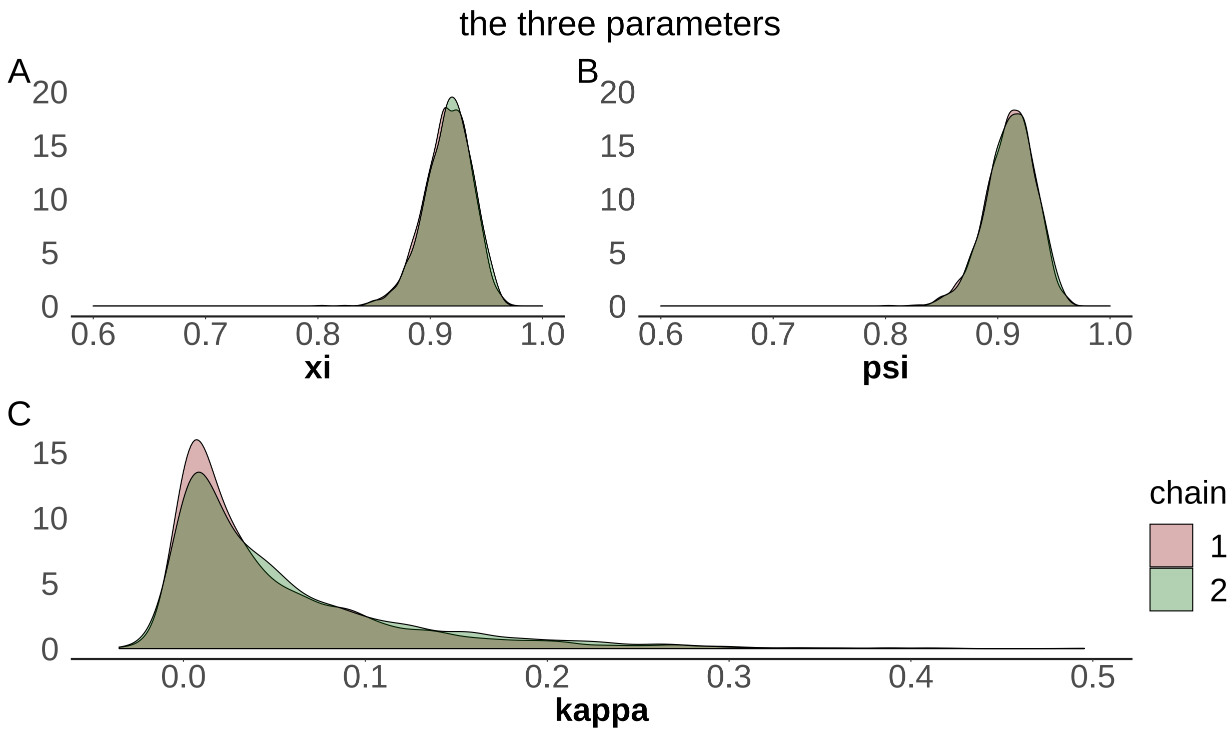
4.2.1 Plot (chains)

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.



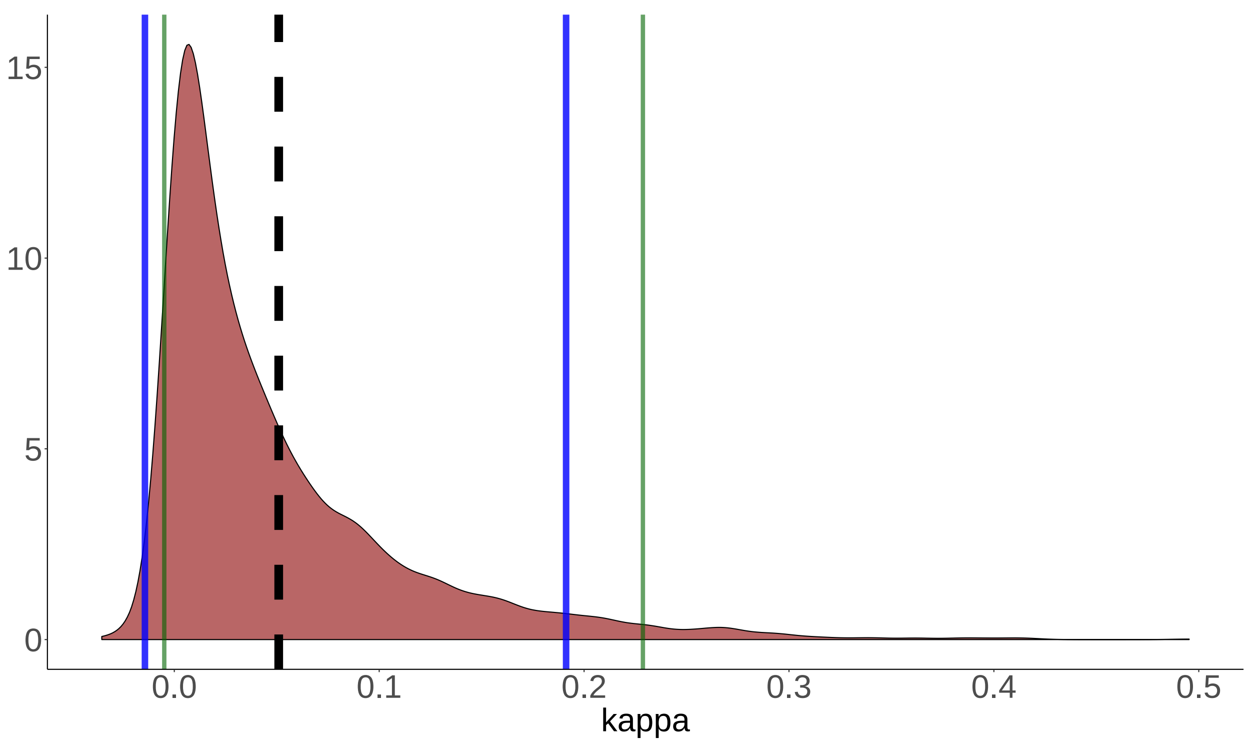
4.2.2 Plot (posterior)

The plot of the rate of agreement (ξ), rate of agreement by chance (ψ) and the chance-corrected agreement rate (κ) values per chain superimposed on each other.



4.2.3 Plot (posterior with CI)

The plot of the combined κ value with confidence intervals. (HDI CI : blue; ETI CI : green; mean : dashed black)



4.3 Confidence interval value and the mean value

The confidence interval values.

```
## # Highest Density Interval
##
##        95% HDI
##  [-0.01, 0.19]
```

```
## # Equal-Tailed Interval
##
##        95% ETI
##  [-0.00, 0.23]
```

The mean value

```
## [1] 0.05095879
```