Applied Bayesian Data Analysis

Assignment 4

Fredrik Nyström

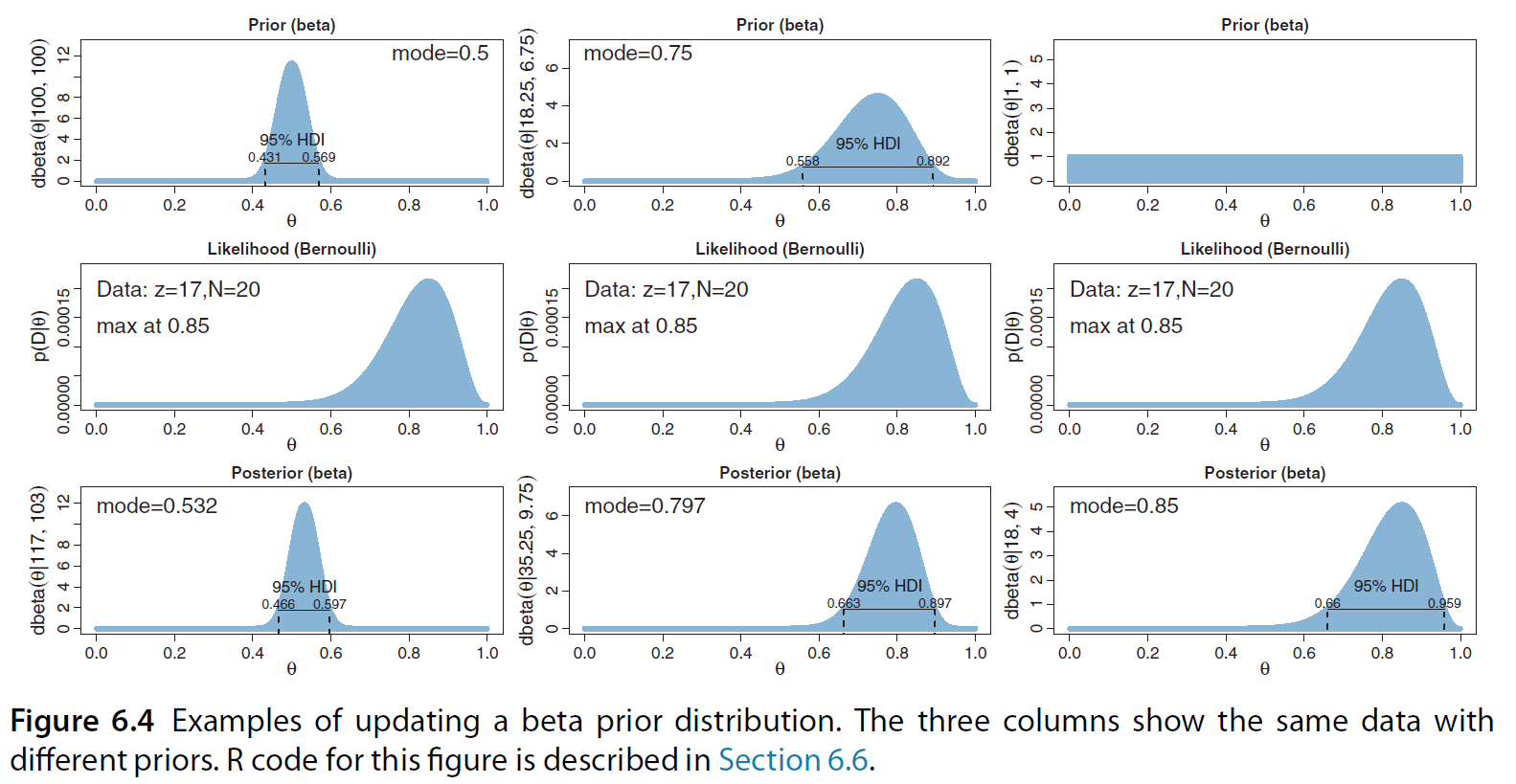
2019-10-28

## Table of Contents

* [Task A.1](#a1)
  + [Slice sampler](#slicesampler)

## [Task A.2](#a2)

* name: a1 ## Task 1 Recreate Figure 6.4 using STAN and/or slice sampling.
* Priors, beta(100, 100), beta(18.25, 6.75), beta(1, 1)
* Bernoulli likelihood
* Data, z = 17, N = 20

.center[]

## Stan model code: Prior & Likelihood

// prior  
parameters {  
 real<lower = 0, upper = 1> theta;  
}  
model {  
 theta ~ beta(`${a}`, `${b}`);  
}

// likelihood  
data {  
 int<lower = 0> N;  
 int y[N];  
}  
  
parameters {  
 real<lower = 0, upper = 1> theta;  
}  
model {  
 y ~ bernoulli(theta);  
}

## String interpolation for prior parameters

# Specify the shape parameters for the beta priors,   
# used for string interpolation  
beta\_left <- list(a = 100, b = 100)  
beta\_middle <- list(a = 18.25, b = 6.75)  
beta\_right <- list(a = 1, b = 1)  
  
# Prepare the prior code for each prior by string interpolation  
left\_prior\_code <- str\_interp(prior\_code, beta\_left)  
middle\_prior\_code <- str\_interp(prior\_code, beta\_middle)  
right\_prior\_code <- str\_interp(prior\_code, beta\_right)  
  
# Prepare the posterior code for each prior by string interpolation  
left\_posterior\_code <- str\_interp(posterior\_code, beta\_left)  
middle\_posterior\_code <- str\_interp(posterior\_code, beta\_middle)  
right\_posterior\_code <- str\_interp(posterior\_code, beta\_right)

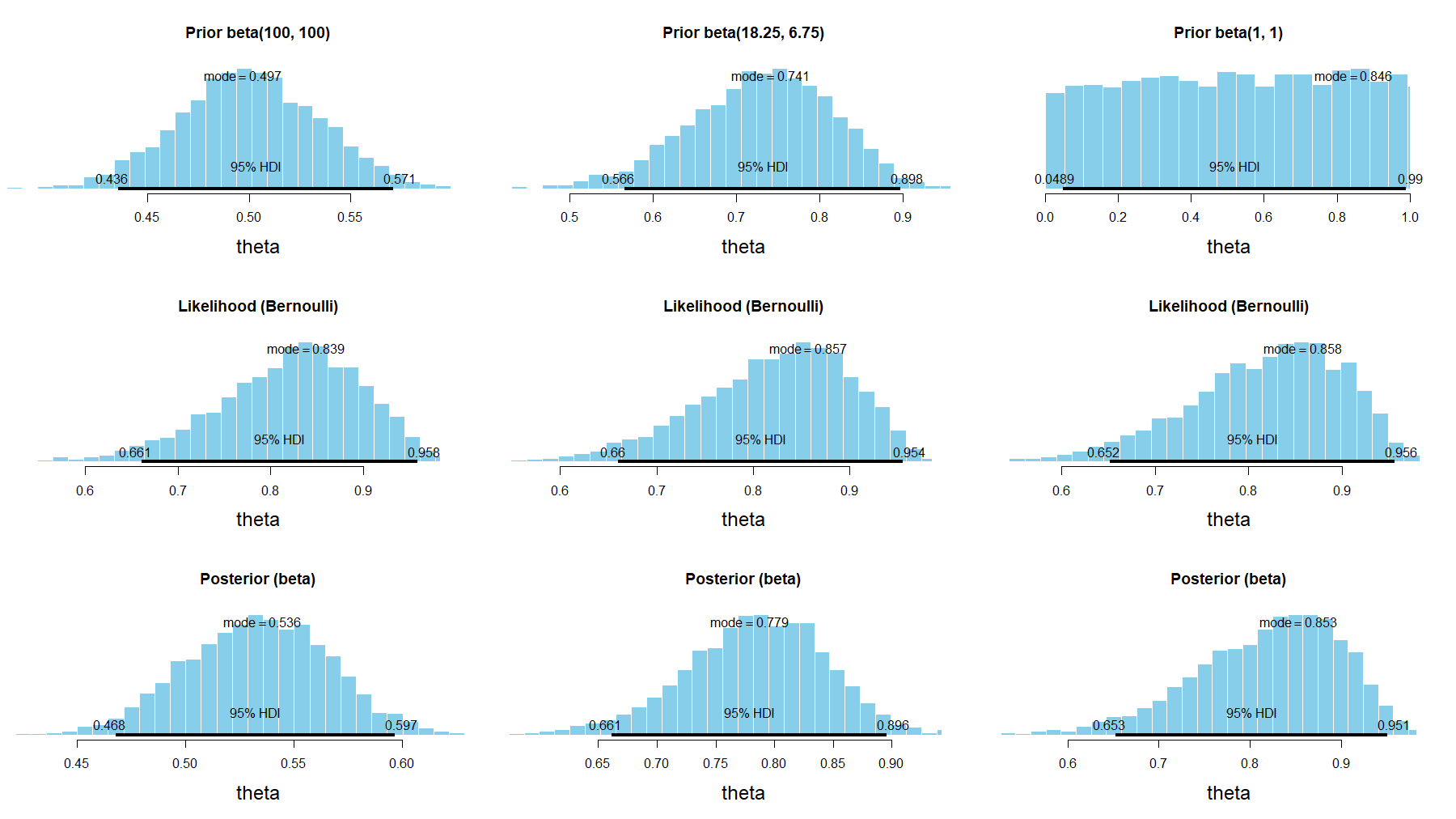
## Stan compilation

# Compile the code. Sampling from the likelihood and the posterior needs the   
# data supplied as an argument  
left\_prior <- stan(model\_code = left\_prior\_code)  
left\_likelihood <- stan(model\_code = likelihood\_code, data = get\_data(y))  
left\_posterior <- stan(model\_code = left\_posterior\_code, data = get\_data(y))  
  
middle\_prior <- stan(model\_code = middle\_prior\_code)  
middle\_likelihood <- stan(model\_code = likelihood\_code, data = get\_data(y))  
middle\_posterior <- stan(model\_code = middle\_posterior\_code, data = get\_data(y))  
  
right\_prior <- stan(model\_code = right\_prior\_code)  
right\_likelihood <- stan(model\_code = likelihood\_code, data = get\_data(y))  
right\_posterior <- stan(model\_code = right\_posterior\_code, data = get\_data(y))

## Construct the plots

left\_prior\_plot <- ~plotPost(left\_prior\_samples, showMode = TRUE, cex = 1,  
 main = "Prior beta(100, 100)", xlab = "theta")  
left\_likelihood\_plot <- ~plotPost(left\_likelihood\_samples, showMode = TRUE, cex = 1,  
 main = "Likelihood (Bernoulli)", xlab = "theta")  
left\_posterior\_plot <- ~plotPost(left\_posterior\_samples, showMode = TRUE, cex = 1,  
 main = "Posterior (beta)", xlab = "theta")  
  
middle\_prior\_plot <- ~plotPost(middle\_prior\_samples, showMode = TRUE, cex = 1,  
 main = "Prior beta(18.25, 6.75)", xlab = "theta")  
middle\_likelihood\_plot <- ~plotPost(middle\_likelihood\_samples, showMode = TRUE, cex = 1,  
 main = "Likelihood (Bernoulli)", xlab = "theta")  
middle\_posterior\_plot <- ~plotPost(middle\_posterior\_samples, showMode = TRUE, cex = 1,  
 main = "Posterior (beta)", xlab = "theta")  
  
right\_prior\_plot <- ~plotPost(right\_prior\_samples, showMode = TRUE, cex = 1,  
 main = "Prior beta(1, 1)", xlab = "theta")  
right\_likelihood\_plot <- ~plotPost(right\_likelihood\_samples, showMode = TRUE, cex = 1,  
 main = "Likelihood (Bernoulli)", xlab = "theta")  
right\_posterior\_plot <- ~plotPost(right\_posterior\_samples, showMode = TRUE, cex = 1,  
 main = "Posterior (beta)", xlab = "theta")

## Recreated Figure 6.4 using MCMC with Stan

.center[]

## Slice sampler

slice\_sampling <- function(posterior, initial\_point, scale\_estimates, iter = 2000, warmup = floor(iter/2)) {  
 chain\_length <- iter - warmup  
 samples <- data.frame(matrix(vector(), iter, length(initial\_point)))  
   
 samples[1,] <- slice\_sample(f = posterior, x\_0 = initial\_point, w = scale\_estimates)  
 for (i in 2:iter) {  
 samples[i, ] <- slice\_sample(f = posterior, x\_0 = unlist(samples[i - 1,]), w = scale\_estimates)   
 }  
   
 colnames(samples) <- paste0("param", seq\_along(colnames(samples)))  
 return(tail(samples, n = chain\_length))  
}

name: a2 ## Task A.2

Given the following measurements .

1. What is the probability of getting a head? Give a 95% credible interval of this probability.
2. What is the probability that ?

–

data {  
 int<lower = 0> N;  
 int y[N];  
}  
  
parameters {  
 real<lower = 0, upper = 1> theta;  
}  
  
model {  
 y ~ bernoulli(theta);  
}  
  
generated quantities {  
 int prob05;  
 prob05 = theta > 0.5;   
}

## Summary from Stan

samples

## Inference for Stan model: f26aeaff8c4fa11f5fab3854aaa025d5.  
## 4 chains, each with iter=12000; warmup=6000; thin=1;   
## post-warmup draws per chain=6000, total post-warmup draws=24000.  
##   
## mean se\_mean sd 2.5% 25% 50% 75% 97.5% n\_eff Rhat  
## theta 0.75 0.00 0.11 0.52 0.68 0.76 0.83 0.92 9698 1  
## prob05 0.98 0.00 0.14 1.00 1.00 1.00 1.00 1.00 16978 1  
## lp\_\_ -9.51 0.01 0.73 -11.60 -9.68 -9.23 -9.05 -9.00 9691 1  
##   
## Samples were drawn using NUTS(diag\_e) at Wed Oct 30 15:02:32 2019.  
## 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).

## Visualizing the posterior

params %>% ggplot(aes(x = theta)) + geom\_histogram(binwidth = 0.01) +   
 geom\_vline(xintercept = mean(params[["theta"]]), color = "blue") +   
 geom\_vline(xintercept = map\_estimate(params[["theta"]]), color = "green") +   
 geom\_vline(xintercept = median(params[["theta"]]), color = "red")

## 

## Expected probability of getting a head?

We know since Chapter 6 with a prior of beta(1, 1), bernoulli likelihood and data of z = 11, N = 14, that the posterior will be a beta(1+11, 1 + 14 - 11) = beta(12, 4).

Expected value of a beta distribution:

mean(params[["theta"]])

## [1] 0.7485882

Other estimators:

map\_estimate(params[["theta"]])

## MAP = 0.79

median(params[["theta"]])

## [1] 0.7595861

## What is the probability that

generated quantities {  
 int prob05;  
 prob05 = theta > 0.5;   
}

glimpse(params["prob05"])

## Observations: 24,000  
## Variables: 1  
## $ prob05 <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...

mean(params[["prob05"]])

## [1] 0.9809583

sum(params[["theta"]] > 0.5)/length(params[["theta"]])

## [1] 0.9809583

## Sampling from the MCMC

y <- c(1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1)  
z <- c(1, 0, 0, 0, 0, 0, 0, 1, 1, 0)  
data <- list(  
 y = y, Ny <- length(y),  
 z = z, Nz <- length(z)  
)  
  
# Compile the stan model  
model <- stan\_model(file = model\_file, model\_name = "Task A.2")  
samples <- sampling(model, data = data, iter = 12000)

## Sampling from the MCMC

params <- rstan::extract(samples)

1. Calculating the probability that .
2. Creating a new variable , and calculating a 95% CI.

mean(params[["y\_gr\_z"]])

## [1] 0.9901667

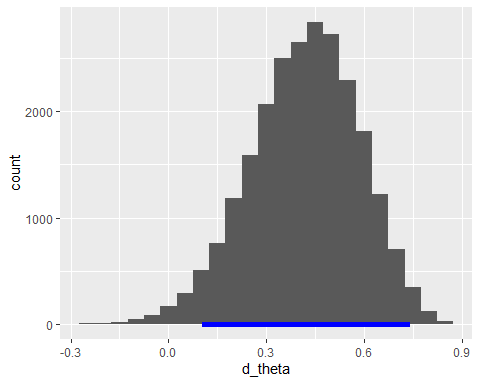
hdi(params[["d\_theta"]], ci = 0.95)

## # Highest Density Interval  
##   
## 95% HDI  
## [0.10, 0.74]

## Histogram of

1. Plot a histogram of . Is it beta distributed?

d\_theta\_hdi <- hdi(params[["d\_theta"]], ci = 0.95)  
params %>% ggplot(aes(x = d\_theta)) + geom\_histogram(binwidth = 0.05) +   
 geom\_segment(x = d\_theta\_hdi$CI\_low, xend = d\_theta\_hdi$CI\_high, y = 0, yend = 0, color = "blue", size = 2)

.pull-left[ ]

.pull-right[ A beta distribution is defined on the interval . The distribution for is not bounded on that interval, therefore it cannot be a beta distribution.]