STA 602 Lab 4

Yicheng Shen

03 October, 2022

```r
create_df <- function(post_draws, prior_draws){
  post_draws <- data.frame(post_draws)
  post_draws$distribution <- "posterior"

  prior_draws <- data.frame(prior_draws)
  colnames(prior_draws) <- "alpha"
  prior_draws$distribution <- "prior"

  dat <- rbind(post_draws, prior_draws)
  return(dat)
}
set.seed(689934)

alpha <- 1
beta <- -0.25
sigma <- 1

N <- 5
x <- array(runif(N, 0, 2), dim=N)
y <- array(rnorm(N, beta * x + alpha, sigma), dim=N)
```

```r
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())
```

```r
stan_dat <- list(y = y, x=x, N=N)
fit.flat <- stan(file = "lab-04-flat_prior.stan", data = stan_dat,
                 chains = 1, refresh = 0, iter = 2000, warmup = 500, seed=48)
```

```r
alpha.flat <- as.matrix(fit.flat, pars = "alpha")
beta.flat <- as.matrix(fit.flat, pars = "beta")
ggplot(alpha.flat %>% as.data.frame, aes(x = alpha)) +
  geom_histogram(bins = 30) +
  labs(title = "Posterior distribution of alpha under the flat prior")
```

Posterior distribution of alpha under the flat prior
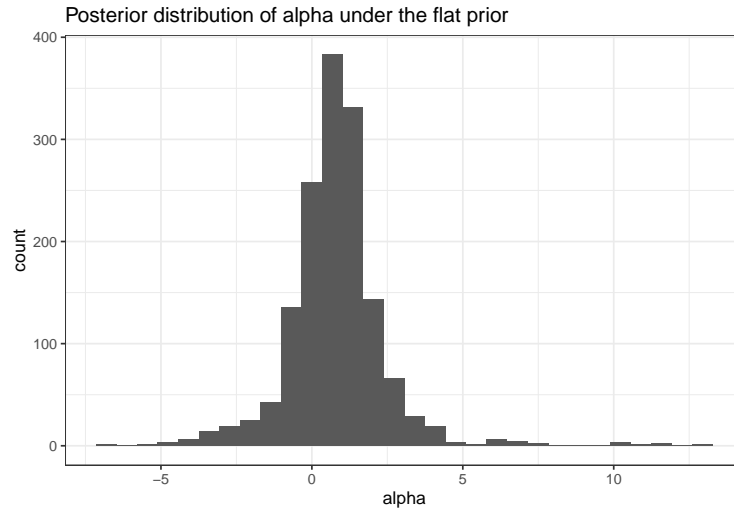

```
print(fit.flat, pars = c("alpha", "beta"))
```

```
## Inference for Stan model: lab-04-flat_prior.
## 1 chains, each with iter=2000; warmup=500; thin=1;
## post-warmup draws per chain=1500, total post-warmup draws=1500.
##
##       mean se_mean   sd  2.5%   25% 50%  75% 97.5% n_eff Rhat
## alpha 0.78    0.09 1.61 -2.61  0.03 0.8 1.47  3.96   315    1
## beta  0.32    0.07 1.24 -2.03 -0.18 0.3 0.86  2.79   335    1
##
## Samples were drawn using NUTS(diag_e) at Sun Oct  2 18:52:45 2022.
## 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).
```

Exercise 1

$\pi(\alpha) \propto 1, \pi(\beta) \propto 1$

$\pi(\theta|y) \propto \pi(\theta)\pi(y|\theta) \propto 1 \times N(X_i^\top \theta, \sigma^2)$

Then we integrate the multivariate normal distribution to the marginal of α and β posterior.

```
solve(t(x) %*% x) %*% t(x) %*% y
```

```
##           [,1]
## [1,] 0.7773957
```

With posterior draws above, we can show that

Posterior mean for α is 0.78, with 95% CI from -2.61 to 3.96.

Posterior mean for β is 0.32, with 95% CI from -2.03 to 2.79.

This is indeed pretty diffused. It is also not surprising since we are using a flat prior and with low sample size data.

Then we try the uniform prior (similiar behaviors)

```r
stan_dat <- list(y = y, x=x, N=N, lb = -10, ub = 10)
fit.unif <- stan(file = "lab-04-unif_prior.stan", data = stan_dat,
                 chains = 1, refresh = 0, iter = 2000, warmup = 500, seed=48)
```

```
## Warning: There were 3 divergent transitions after warmup. See
## https://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
## to find out why this is a problem and how to eliminate them.
```

```
## Warning: Examine the pairs() plot to diagnose sampling problems
```

```r
alpha.unif <- as.matrix(fit.unif, pars = c("alpha"))
beta.unif <- as.matrix(fit.unif, pars = c("beta"))
ggplot(alpha.unif %>% as.data.frame, aes(x = alpha)) +
  geom_histogram(bins = 30) +
  labs(title = "Posterior distribution of alpha under the Uniform(-10, 10) prior")
```


Posterior distribution of alpha under the Uniform(−10, 10) prior
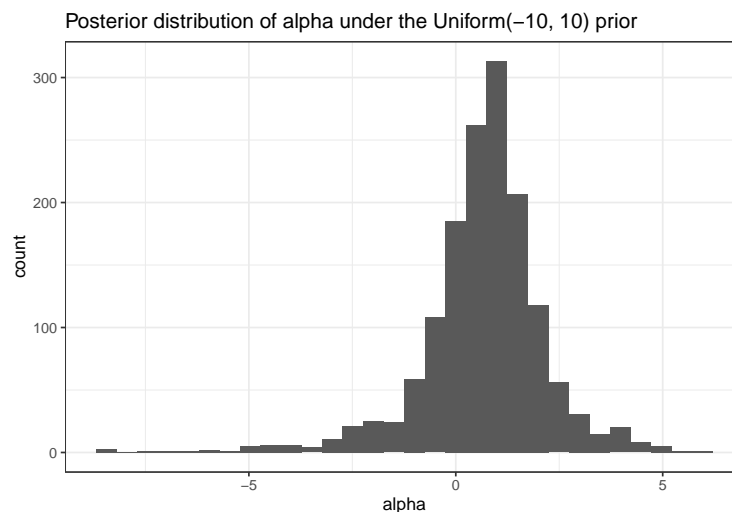
```r
print(fit.unif, pars = c("alpha"))
```

```
## Inference for Stan model: lab-04-unif_prior.
## 1 chains, each with iter=2000; warmup=500; thin=1;
## post-warmup draws per chain=1500, total post-warmup draws=1500.
##
##       mean se_mean   sd  2.5%  25%  50%  75% 97.5% n_eff Rhat
## alpha 0.65    0.1 1.51 -2.82 0.05 0.78 1.44  3.54   214    1
##
## Samples were drawn using NUTS(diag_e) at Mon Oct  3 11:23:59 2022.
## 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).
```

Exercise 2

```
stan_dat <- list(y = y, x=x, N=N)
fit.norm <- stan(file = "lab-04-normal_prior.stan", data = stan_dat,
                 chains = 1, refresh = 0, iter = 2000, warmup = 500, seed=49)
```

```
alpha.norm<- as.matrix(fit.norm, pars = c("alpha"))
ggplot(alpha.norm %>% as.data.frame, aes(x = alpha)) +
  geom_histogram(bins = 30) +
  labs(title = "Posterior distribution of alpha under N(0,1) weakly informative prior")
```


Posterior distribution of alpha under N(0,1) weakly informative prior
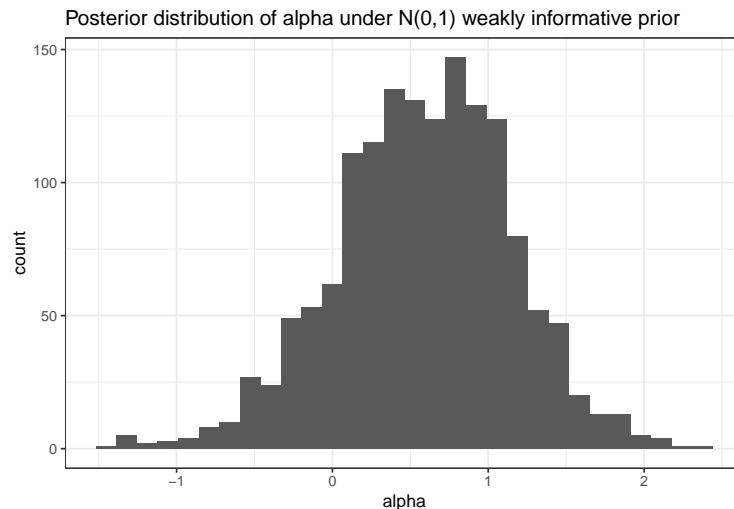
```
print(fit.norm, pars = c("alpha", "beta"))
```

```
## Inference for Stan model: lab-04-normal_prior.
## 1 chains, each with iter=2000; warmup=500; thin=1;
## post-warmup draws per chain=1500, total post-warmup draws=1500.
##
##        mean se_mean   sd  2.5%  25%  50%  75% 97.5% n_eff Rhat
## alpha 0.58    0.02 0.57 -0.56 0.22 0.60 0.98  1.65   625    1
## beta  0.36    0.02 0.43 -0.47 0.08 0.36 0.63  1.26   644    1
##
## Samples were drawn using NUTS(diag_e) at Mon Oct  3 11:24:01 2022.
## 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).
```

We can do similar tricks as above. Below is the result from the posterior draws.

Posterior mean for α is 0.58, with 95% CI from -0.56 to 1.65.

Posterior mean for β is 0.36, with 95% CI from -0.47 to 1.26.

Now we have tighter CI from posterior. This seems to be a more concentrated normal distribution. So our prior is informative (again since the sample size is tiny).

Exercise 3

```r
stan_dat <- list(y = y, x=x, N=N)
fit.cauchy <- stan(file = "lab-04-cauchy_prior.stan",data = stan_dat,
                   chains = 1, refresh = 0, iter = 2000, warmup = 500, seed=55)
```

```r
alpha.cauchy<- as.matrix(fit.cauchy, pars = c("alpha"))
ggplot(alpha.cauchy %>% as.data.frame, aes(x = alpha)) +
  geom_histogram(bins = 30) +
  labs(title = "Posterior distribution of alpha under Cauchy(0,1) weakly informative prior")
```
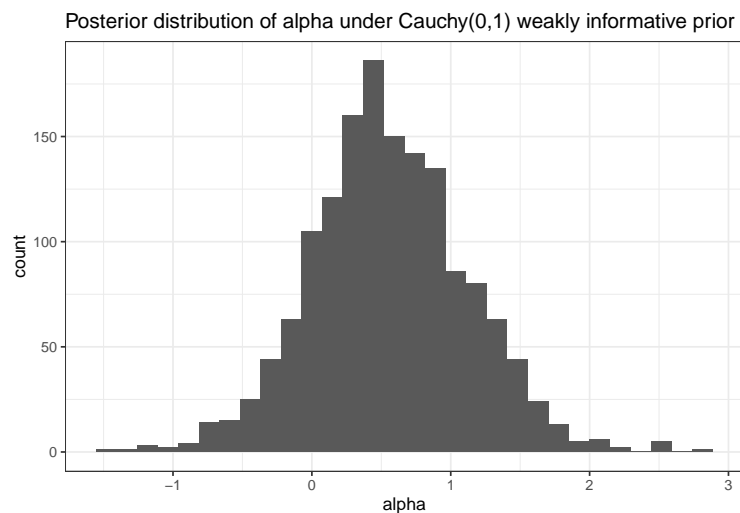


```r
print(fit.cauchy, pars = c("alpha", "beta"))
```

```
## Inference for Stan model: lab-04-cauchy_prior.
## 1 chains, each with iter=2000; warmup=500; thin=1;
## post-warmup draws per chain=1500, total post-warmup draws=1500.
##
##        mean se_mean   sd  2.5%  25%  50%  75% 97.5% n_eff Rhat
## alpha 0.56    0.03 0.56 -0.54 0.21 0.53 0.91  1.66   465    1
## beta  0.37    0.02 0.42 -0.47 0.09 0.37 0.64  1.19   371    1
##
## Samples were drawn using NUTS(diag_e) at Mon Oct  3 11:24:02 2022.
## 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).
```

```r
plot_dat <- create_df(alpha.norm, alpha.cauchy) %>%
  mutate(distribution = if_else(distribution == "posterior", "Normal","Cauchy"))

ggplot(plot_dat, aes(alpha, fill = distribution)) +
  geom_histogram(binwidth = 0.25, alpha = 0.7, position = "identity")+
  geom_vline(xintercept = alpha) +
  scale_fill_brewer()
```
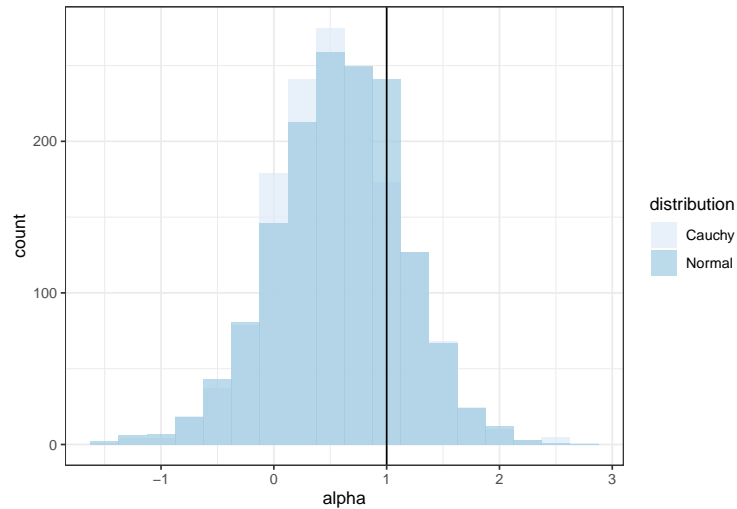
Cauchy is more informative in a way it behaves heavy tails and puts more probability mass in the tails. But the difference is tiny.

```r
alpha <- 5
N <- 10
x <- runif(N, 0, 2)
y <- rnorm(N, beta * x + alpha, sigma)
```

```r
stan_dat <- list(y = y, x = x, N = N)
fit.norm <- stan(file = "lab-04-normal_prior.stan", data = stan_dat,
                 chains = 1, refresh = 0, iter = 2000, warmup = 500, seed=49)
```

```r
alpha.norm<- as.matrix(fit.norm, pars = c("alpha"))
prior_draws <- rnorm(1000, 0, 1)
plot_dat <- create_df(alpha.norm, prior_draws)
ggplot(plot_dat, aes(alpha, fill = distribution)) +
  geom_histogram(binwidth = 0.25, alpha = 0.7, position = "identity")+
  geom_vline(xintercept = alpha) + scale_fill_brewer()
```
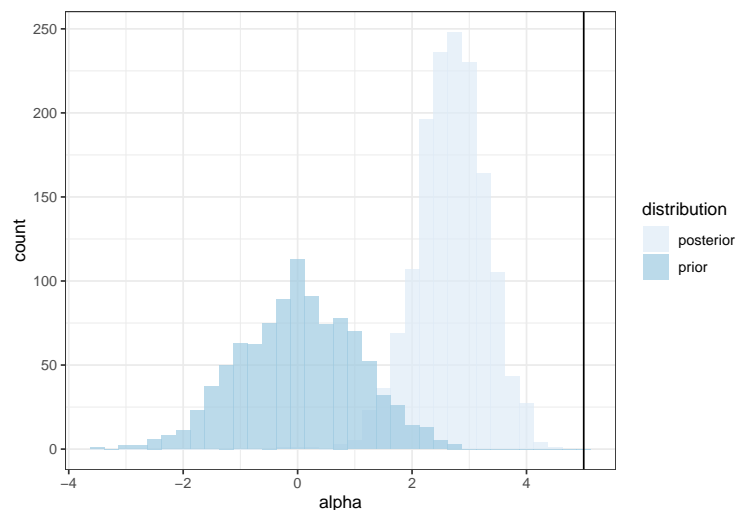
```
stan_dat <- list(y = y, x = x, N = N)
fit.cauchy <- stan(file = "lab-04-cauchy_prior.stan",data = stan_dat,
                    chains = 1, refresh = 0, iter = 2000, warmup = 500, seed=55)
```

```
alpha.cauchy<- as.matrix(fit.cauchy, pars = c("alpha"))
prior_draws <- rcauchy(1000, 0, 1)
prior_draws <- prior_draws[abs(prior_draws) < 25]
plot_dat <- create_df(alpha.cauchy, prior_draws)
ggplot(plot_dat, aes(alpha, fill = distribution)) +
  geom_histogram(binwidth = .5, alpha = 0.7, position = "identity")+
  geom_vline(xintercept = alpha) +
  scale_fill_brewer()
```
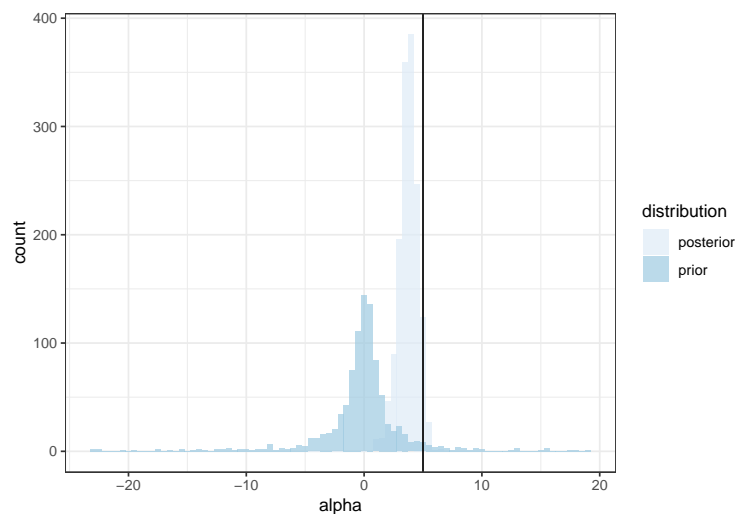

Exercise 4

In $N = 5, \alpha = 1$ case, $N(0,1)$ and $Cauchy(0,1)$ behaves similarly. Both priors are appropriate guesses considering the true α value is 1. Since the sample size is small compared with prior belief, the posterior is dominated by the prior and it is able to roughly cover the true value. The two posteriors are similar.

When $N = 10, \alpha = 5$, $Cauchy(0,1)$ is a much better choice than $N(0,1)$. The true α value is now away from the prior guesses, so priors are not that appropriate. And as the sample size grows relatively, $Cauchy(0,1)$ has heavy tails that allow the posterior to be adjusted by data and to capture the true value.

As sample size grows to be large enough, both priors are not that strong and the posterior would be dominated by the likelihood (data).

Exercise 5

The heavier-tailed priors usually put more probability mass on the extreme values, thus allowing the posterior probably move far from the prior center.

If the sample size is small and we don't have strong prior knowledge, we should consider to have heavier-tailed priors to allow the posterior to be adjusted by data. If we do have some firm belief, we might consider using light-tailed priors to strengthen our prior.

Again, if the sample size is pretty large, the different priors matter less since posterior is dominated by data.

Exercise 6

We should construct priors based on how much prior knowledge we have and whether we strongly trust our prior knowledge. If we lack a strong belief, we should prefer heavy-tailed priors in case that the true value is far away from our prior guess.

Exercise 7

```r
theta <- runif(10000,0,1)
hist(theta)
```

Histogram of theta
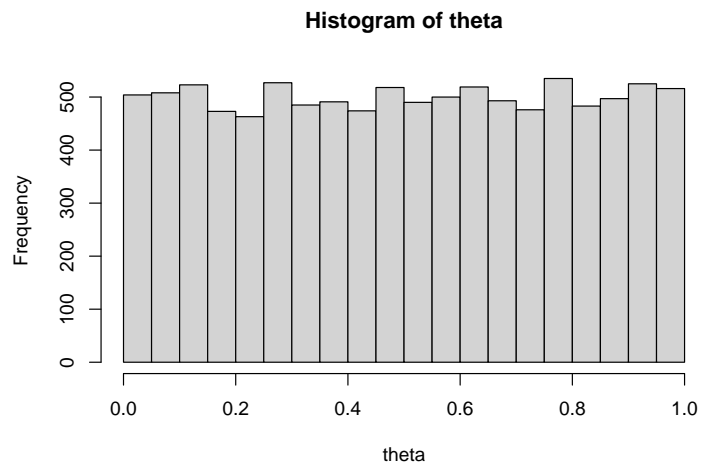

```r
logit <- function(x){
  ret <- log(x/(1-x))# finish function for log odds
  return(ret)
}
eta <- logit(theta)
hist(eta)
```

Histogram of eta
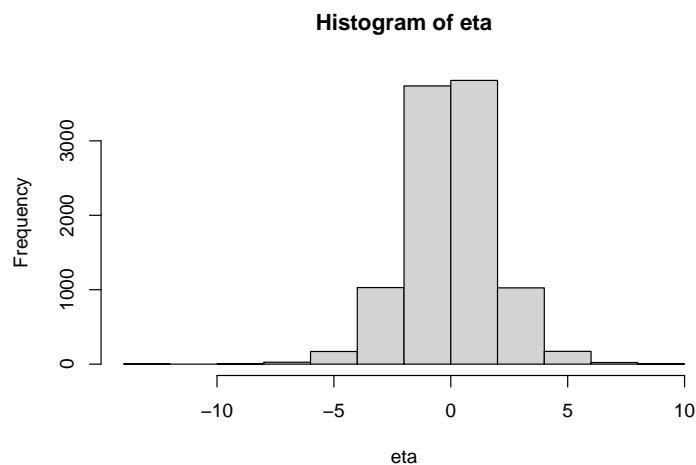
```
set.seed(123);
theta <- 0.3;
N <- 10;
y <- rbinom(N, 1, theta)
theta.mle <- sum(y)/N
```

```
stan_dat <- list(y = y,N=N)
fit.bayes.prob <- stan(file = "lab-04-prob.stan", data = stan_dat, refresh = 0, iter = 2000)
```

```
print(fit.bayes.prob, pars = c("theta", "eta"))
```

```
## Inference for Stan model: lab-04-prob.
## 4 chains, each with iter=2000; warmup=1000; thin=1;
## post-warmup draws per chain=1000, total post-warmup draws=4000.
##
##         mean se_mean   sd  2.5%   25%   50%  75% 97.5% n_eff Rhat
## theta   0.42    0.00 0.14  0.16  0.32  0.42 0.51  0.69  1533    1
## eta    -0.37    0.02 0.63 -1.64 -0.78 -0.34 0.06  0.80  1542    1
##
## Samples were drawn using NUTS(diag_e) at Mon Oct  3 11:24:07 2022.
## 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).
```

```
hist(as.matrix(fit.bayes.prob, pars = "theta"), breaks = 50)
```


Histogram of as.matrix(fit.bayes.prob, pars = "theta")
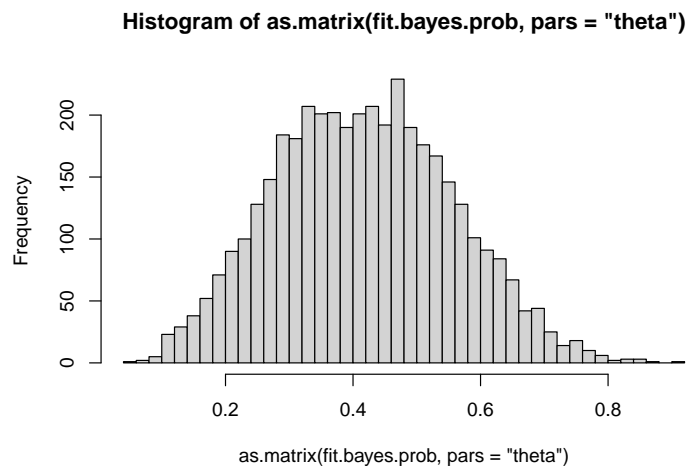
The output above verifies that posterior mean for θ is 0.42.

The mode of beta distribution $(1+4, 1+6)$ is $\frac{a-1}{a+b-2} = \frac{5-1}{5+7-2} = 0.4$.

We can thus see that both theoretical and actual posterior draws show the mode is around 0.4 (simulated draws might vary a bit).

Exercise 8

Beta(0,0) is an improper prior (cannot integrate to 1), which usually leads to an improper posterior and that would not be appropriate to do Bayesian.

Nevertheless, sometimes we can use improper prior as long as we make sure that the posterior is proper.

Exercise 9

```
fit.logodds <- stan(file = "lab-04-log_odds.stan", data = stan_dat, refresh = 0, iter = 2000)
```

```
## recompiling to avoid crashing R session
```

```
## Trying to compile a simple C file
```

```
print(fit.logodds, pars = c("theta", "eta"))
```

```
## Inference for Stan model: lab-04-log_odds.
## 4 chains, each with iter=2000; warmup=1000; thin=1;
## post-warmup draws per chain=1000, total post-warmup draws=4000.
##
##         mean se_mean   sd  2.5%    25%    50%  75% 97.5% n_eff Rhat
## theta   0.40    0.00 0.15  0.14   0.30   0.40 0.50  0.71  1453    1
## eta    -0.43    0.02 0.68 -1.82  -0.86  -0.42 0.01  0.89  1383    1
##
## Samples were drawn using NUTS(diag_e) at Mon Oct  3 11:24:35 2022.
## 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).
```

We have $\pi = \frac{\theta}{1-\theta}$ and $\theta = \frac{\pi}{1+\pi}$ then $\frac{\partial \pi}{\partial \theta} = \frac{1}{(\theta-1)^2}$

Then $p(\theta) = 1 \times \frac{1}{(\theta-1)^2}$, which is still an improper posterior.