

# STA 602 Lab 4

Ryan Tang

04 October, 2022

---

## Exercise 1

Below are the means and [2.5%, 97.5%], 95% posterior confidence interval on  $\beta$  and  $\alpha$ . The posteriors are diffuse because of two. First, we do not have too many data points and used a flat prior. In other words, flat prior placed our beliefs equally throughout the place. Hence, with only 5 data points, it is not yet enough to rule out the extreme.

```
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  2.04     0.45 5.34 -7.57   1.08   2.49  3.84 10.49   140 1.01
## beta  -0.80     0.34 4.35 -7.51 -2.20 -1.07  0.12   7.30   163 1.01
##
## Samples were drawn using NUTS(diag_e) at Tue Oct  4 01:20:14 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

Below are the means and [2.5%, 97.5%], 95% posterior confidence interval on  $\beta$  and  $\alpha$ . Given that we know that the true  $\alpha$  is 1, the  $\mathcal{N}(0,1)$  works a bit well than the Uniform prior. No more extreme values, but it is hard to determine the adequacy of the prior. However, the posterior is still spread out with around 0.5 standard deviations. And the effective sample size increased to 625.

```
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  1.03     0.03 0.74 -0.43   0.54  1.05  1.56   2.44   748   1
## beta   0.03     0.03 0.65 -1.26 -0.42  0.02  0.47   1.31   644   1
##
## Samples were drawn using NUTS(diag_e) at Tue Oct  4 01:20:36 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).
```

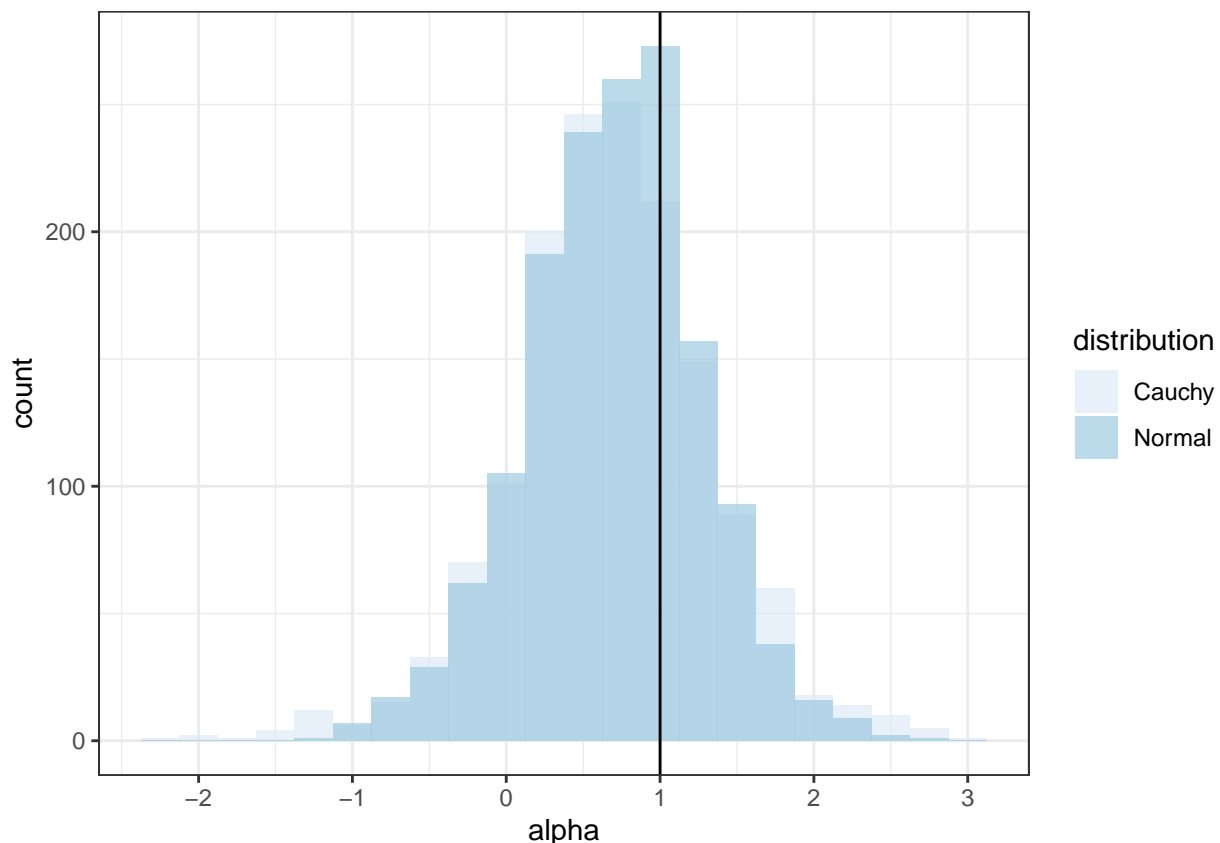
### Exercise 3

Theoretically, Cauchy prior puts more weights on the tail. It is not necessarily less informative than the normal distribution. And the comparison between the two posterior does not show significant differences. However, two things to note are that the Cauchy posterior only has a 465 effective sample size. It contains less information than the normal prior. And, it puts more mass on the extreme value than the normal distribution making it more sensitive to the wrong parameterization of the prior.

### Exercise 4

Here, first, we set the true  $\alpha = 1$  and compares of using  $\mathcal{N}(0, 1)$  and  $\text{Cauchy}(0, 1)$  as a prior. When the prior mean is close to the mean  $\alpha$ , the two prior performed similarly despite having only 5 data points. Both posterior distributions cover the true  $\alpha$  in their right tail. In this limited case, we can't conclude which one is more appropriate.

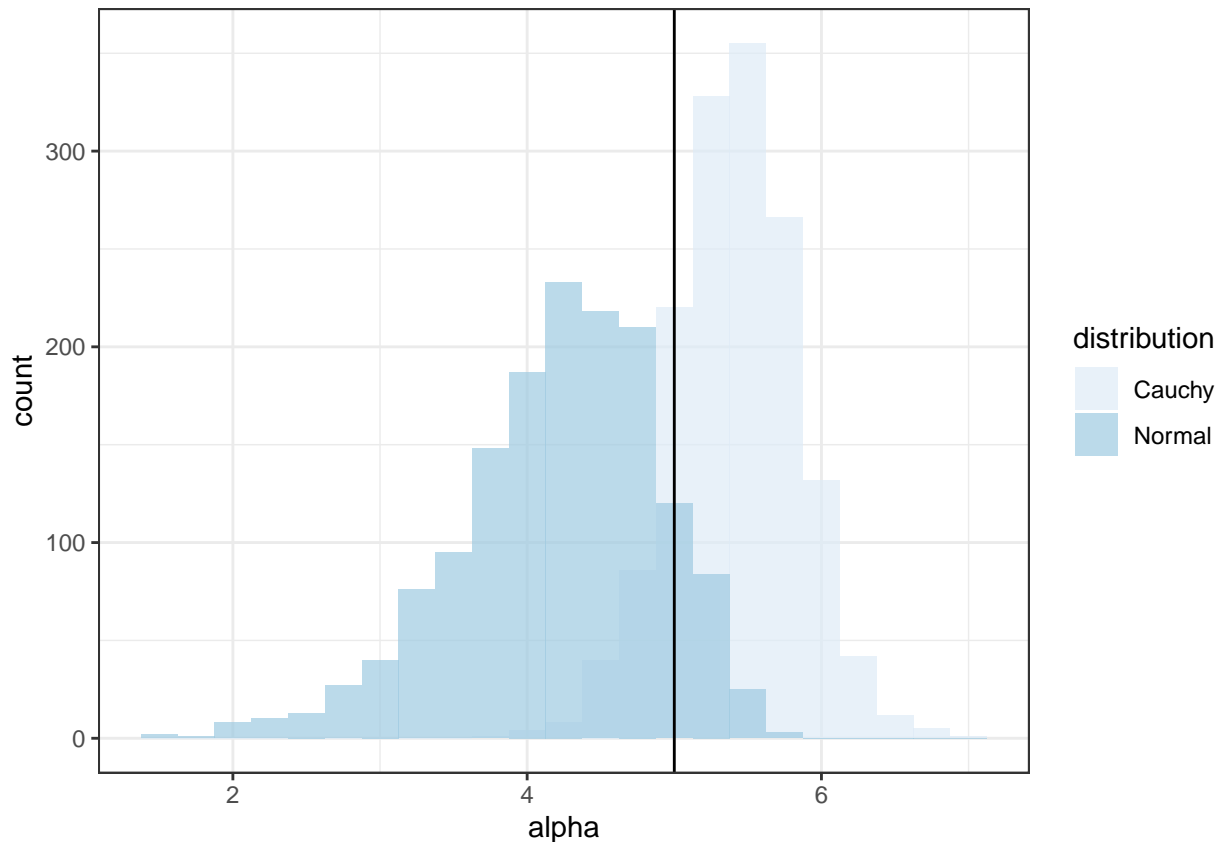
```
plot_dat <- create_df(alpha.norm, alpha.cauchy) %>%  
  mutate(distribution = if_else(distribution == "posterior", "Normal", "Cauchy"))  
  
ggplot(plot_dat, aes(alpha, fill = distribution)) +  
  geom_histogram(binwidth = 0.25, alpha = 0.7, position = "identity") +  
  geom_vline(xintercept = alpha) +  
  scale_fill_brewer()
```



Now, if we changed the true  $\alpha$  from 1 to 5, the differences in Prior started to take effect. Because the Prior mean is far from the actual value; thus, the sensitivity began to play a more significant role here in determining the posterior distribution. We can see that, in this case, Cauchy performed much better than the Normal prior by just having slightly more mass on the far tail.

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

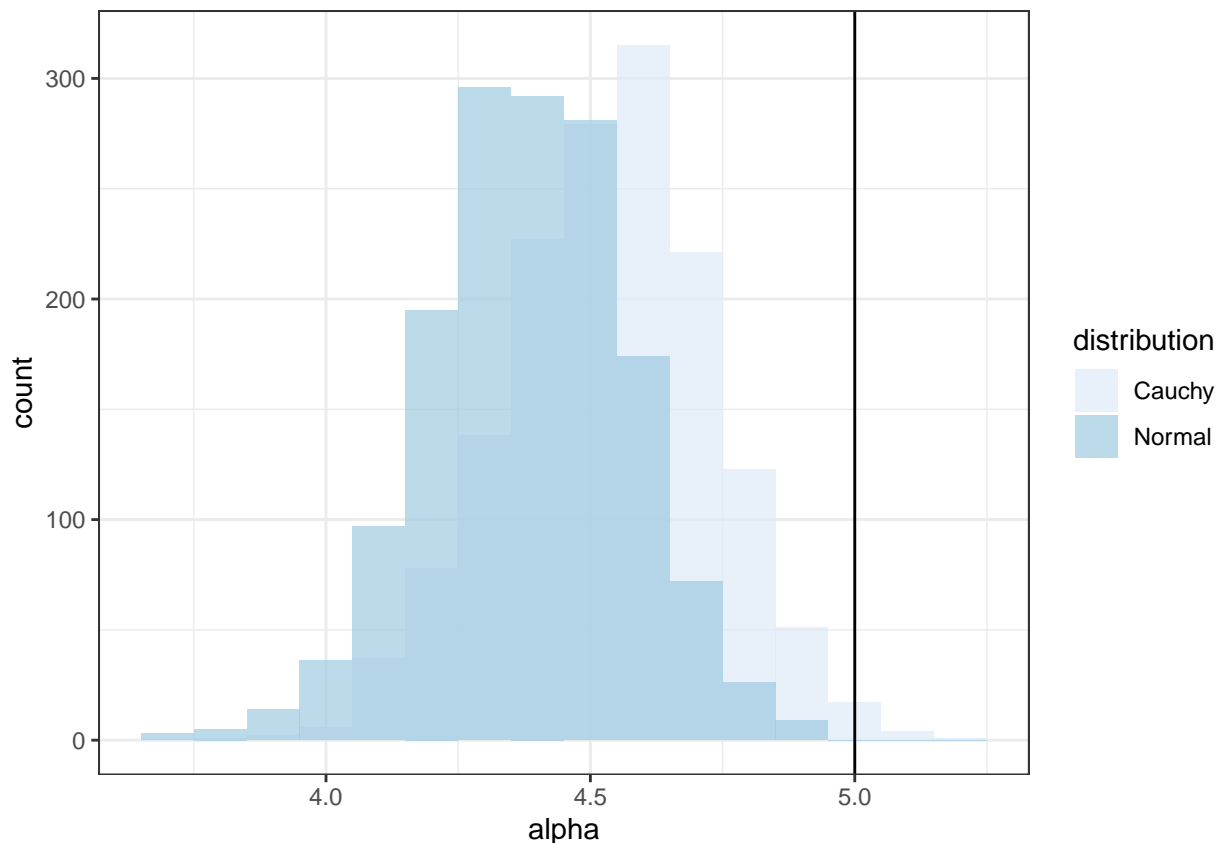
ggplot(plot_dat, aes(alpha, fill = distribution)) +
  geom_histogram(binwidth = 0.25, alpha = 0.7, position = "identity")+
  geom_vline(xintercept = alpha) +
  scale_fill_brewer()
```



Lastly, if we keep the  $\alpha$  as 5 but increase the sample size, we will notice that the posterior resulting from using the Normal prior started to catch up. Although, at a much slower pace, two posteriors tend to converge to a similar posterior when we have large samples.

```
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 = 0.1, alpha = 0.7, position = "identity")+
  geom_vline(xintercept = alpha) +
  scale_fill_brewer()
```



### Exercise 5

Lighter-tailed priors are much less sensitive than their Heavy-tailed counterparts. We can use the Lighter-tailed priors when we are reasonably sure about the phenomenon. Vice versa, if we don't hold a strong prior belief, we can opt for heavy-tailed priors to cover the cases when the observations are pretty off our expectations. In noisy observations, strong prior beliefs are insensitive to outliers but can also be slow to adopt the real norm.

Concerning the sample size, if we have an abundance of data, neither of them will matter much. However, when we have limited data, Heavy-tailed priors can help when we are not certain about our beliefs. And, Lighter-tailed priors can guardrail any outliers by imposing stronger bias.

### Exercise 6

We can keep using the actual scale if we know and interpret it. If not, we can standardize the data, impose a standardized prior, and analyze the scale in unit scale. This has to do with the data.

### Exercise 7

We are imposing a  $Beta(1,1)$  prior on the parameter  $\theta$ , with the following dataset  $Y$ , we expect a  $Beta(5,7)$  posterior. Hence, the theoretical mode is 0.4. It is pretty close to what we sampled a bit off. See below for a histogram.

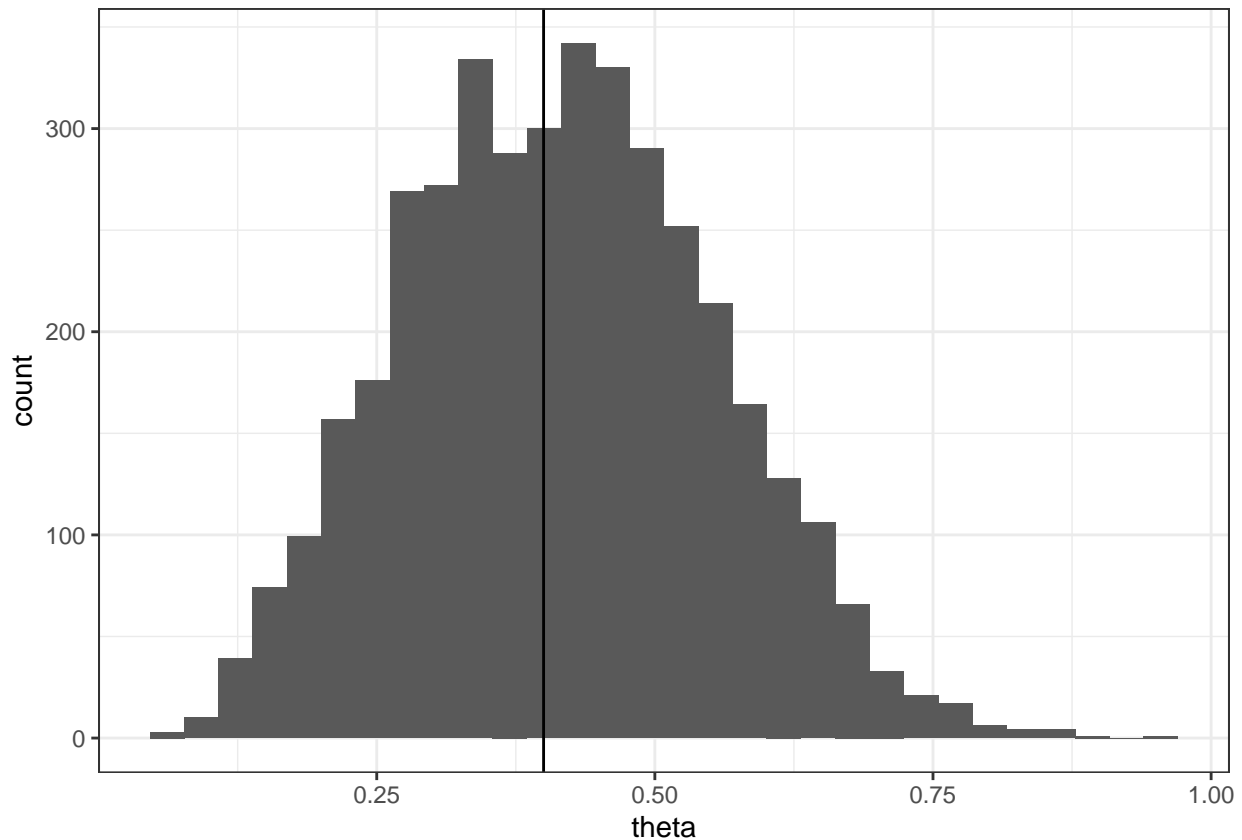
```
print(y) # sample data
```

```
## [1] 0 1 0 1 1 0 0 1 0 0
```

```
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  0.41    0.00 0.14  0.16  0.31  0.41 0.51  0.69 1537   1
## eta   -0.38    0.02 0.62 -1.64 -0.79 -0.35 0.04  0.79 1547   1
##
## Samples were drawn using NUTS(diag_e) at Tue Oct  4 01:21:24 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).

postSample <- as.data.frame(as.matrix(fit.bayes.prob, pars=("theta")))
ggplot(postSample, aes(theta)) +
  geom_histogram(bins=30) +
  geom_vline(xintercept = 0.4)
```



## Exercise 8

$Beta(0, 0)$  is not a proper distribution, the denominator  $Beta(0, 0)$  is infinity. Well, it works in computers with finite precision, for example,  $1e-4$  used in the STAN model. But, as a distribution, it is pure theoretically. Although, it can be used as an uninformative prior for the Binomial in our case.

```
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  0.13  0.29  0.40 0.5  0.70 1575   1
## eta    -0.45    0.02 0.69 -1.88 -0.89 -0.42 0.0  0.87 1538   1
##
## Samples were drawn using NUTS(diag_e) at Tue Oct  4 01:21:45 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 9

We will have  $p(\theta) \propto (1 - \theta)^{-2}$  as the induced prior from a uniform  $\pi$ . It definitely does not integrate to 1 between the range of  $\in [0, 1]$ , nor is it a uniform prior.

---