Tutorial07 Solutions (Part 2)

```
data.single <- generate_data.onecolor()
table(data.single$item)
```

```
##
##   1   2
## 200 200
```

```
data.mixed <- generate_data.mixed_proportion()
```

```
## [1] 5 1 0
## [1] 5 1 1
## [1] 5 1 2
## [1] 5 1 3
## [1] 5 1 4
## [1] 5 2 5
## [1] 5 2 6
## [1] 5 2 7
## [1] 5 2 8
## [1] 5 2 9
## [1]  5  3 10
## [1]  5  3 11
## [1]  5  3 12
## [1]  5  3 13
## [1]  5  3 14
## [1]  5  4 15
## [1]  5  4 16
## [1]  5  4 17
## [1]  5  4 18
## [1]  5  4 19
```

3. A model without overhypotheses

Exercise: Edit the model declaration in `model.stan`, or create a copy of that file so that you get the non-hierarchical, single level model in Figure 3b. Run this truncated model with both the single-color-bags and the mixed-color-bags settings. Compare `theta_new` under both data settings - how does this differ from the full model?

Hints:

- Variable assignments (+ simultaneous type declarations) can be made in the `model` block, but not the `parameters` block.
- Commenting out is done with `//` (comment out, don't delete, so you can restore things later!)
- For beta, you will get a syntax error if you try to declare it as a simplex (i.e. as a vector that has to sum to 1); use `vector` instead.
- `rep_vector(real, int)` is a function that creates a vector (size int) that consists of `[int]` copies of `[real]`.
- make sure your lines end with `;`

Once you are done restore the model to the original (full, hierarchical) structure.

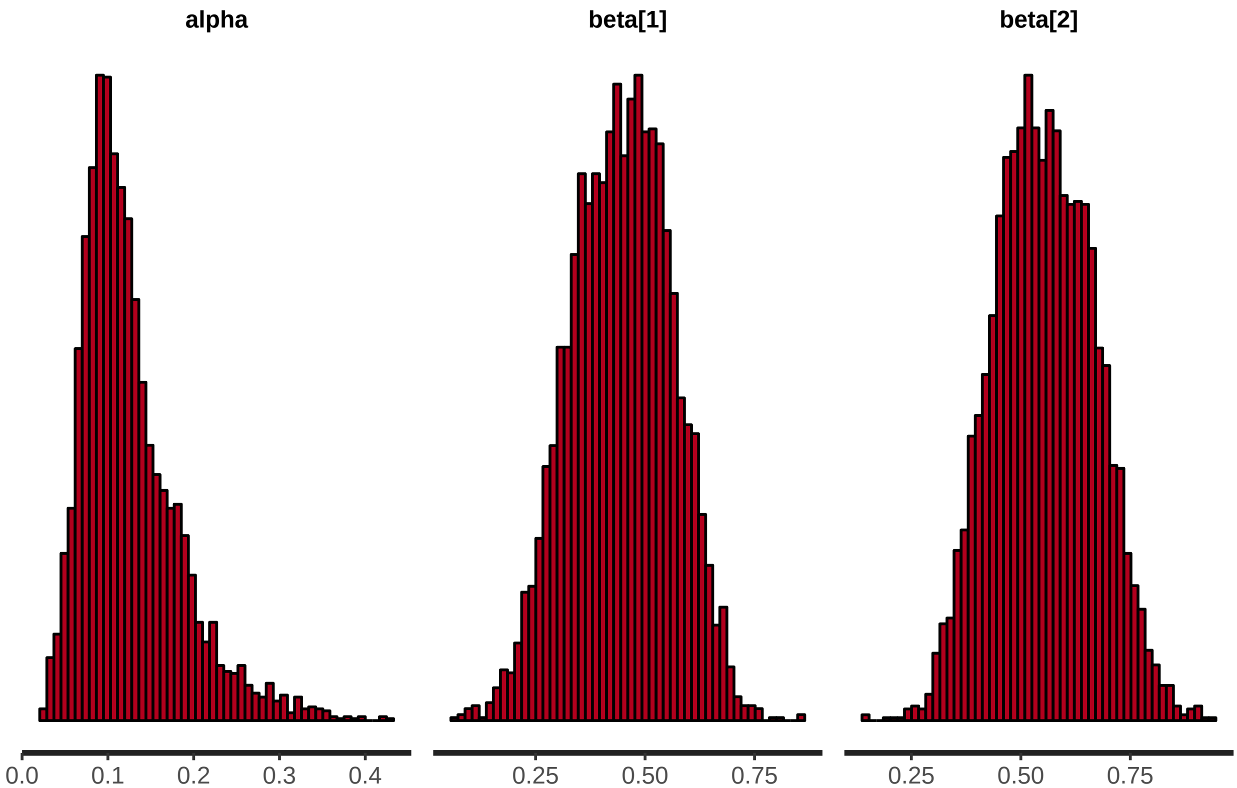
```
fitsinglenoH <- stan(file = 'modelNH.stan', data =data.single , verbose = FALSE,
            iter = 2000, chains = 4, control=list(adapt_delta=0.99))
```

```
## Warning: There were 729 divergent transitions after warmup. Increasing adapt_delta above 0.99 may hel
## http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

## Warning: There were 2644 transitions after warmup that exceeded the maximum treedepth. Increase max_
## http://mc-stan.org/misc/warnings.html#maximum-treedepth-exceeded

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

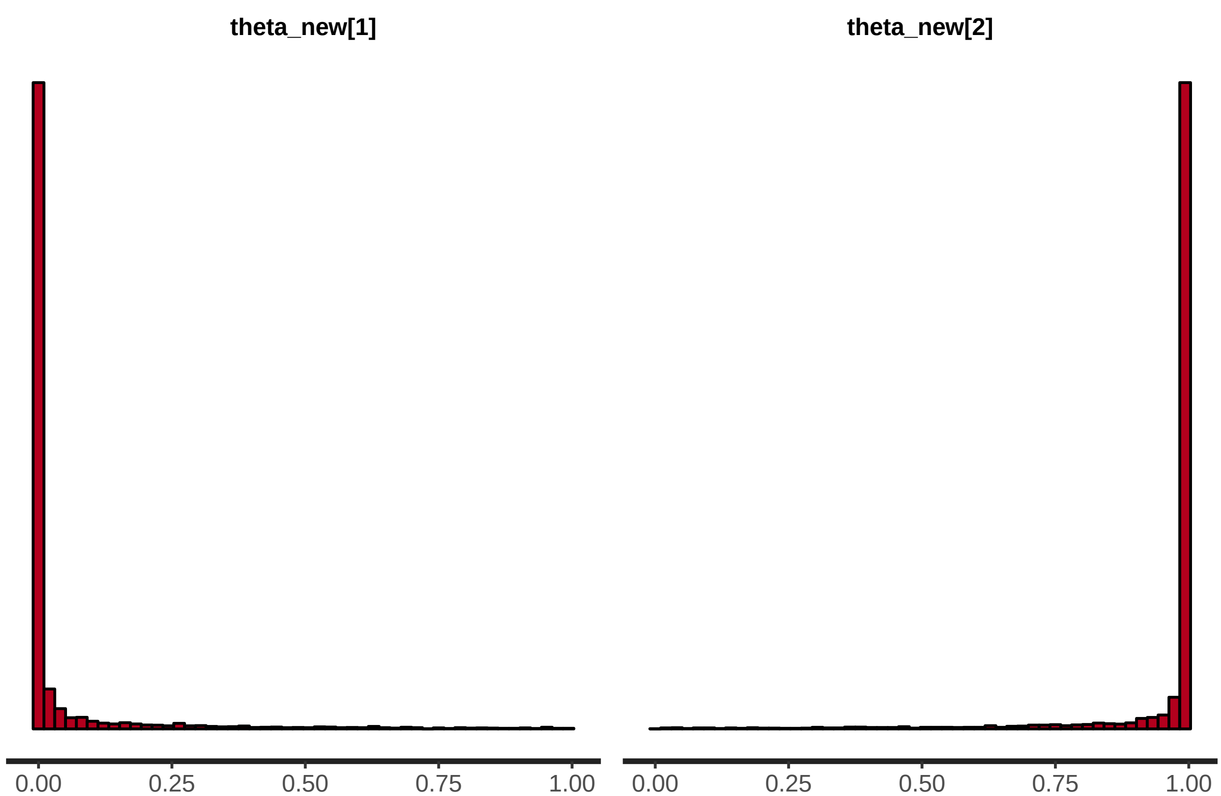
```r
sNHsumm <- summary(fitsinglenoH)
kable(sNHsumm$summary,  digits=2)
```

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
lambda	1.86	0.16	1.34	0.36	0.97	1.52	2.36	5.35	69.45	1.01
mu	1.72	0.18	1.34	0.23	0.75	1.35	2.34	5.23	57.01	1.08
alpha	0.13	0.01	0.06	0.05	0.09	0.11	0.15	0.28	59.72	1.06
beta[1]	0.45	0.02	0.11	0.22	0.36	0.45	0.53	0.66	51.42	1.07
beta[2]	0.55	0.02	0.11	0.34	0.47	0.55	0.64	0.78	51.42	1.07
theta[1,1]	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02	1103.87	1.00
theta[1,2]	1.00	0.00	0.01	0.98	1.00	1.00	1.00	1.00	1103.87	1.00
theta[2,1]	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02	946.12	1.00
theta[2,2]	1.00	0.00	0.01	0.98	1.00	1.00	1.00	1.00	946.12	1.00
theta[3,1]	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02	1910.47	1.00
theta[3,2]	1.00	0.00	0.01	0.98	1.00	1.00	1.00	1.00	1910.47	1.00
theta[4,1]	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	1672.70	1.00
theta[4,2]	1.00	0.00	0.01	0.99	1.00	1.00	1.00	1.00	1672.70	1.00
theta[5,1]	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	1771.99	1.01
theta[5,2]	1.00	0.00	0.01	0.99	1.00	1.00	1.00	1.00	1771.99	1.01
theta[6,1]	1.00	0.00	0.01	0.98	1.00	1.00	1.00	1.00	513.74	1.02
theta[6,2]	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02	513.74	1.02
theta[7,1]	1.00	0.00	0.01	0.98	1.00	1.00	1.00	1.00	976.24	1.00
theta[7,2]	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02	976.24	1.00
theta[8,1]	1.00	0.00	0.01	0.99	1.00	1.00	1.00	1.00	580.47	1.01
theta[8,2]	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	580.47	1.01
theta[9,1]	1.00	0.00	0.01	0.98	1.00	1.00	1.00	1.00	490.90	1.00
theta[9,2]	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02	490.90	1.00
theta[10,1]	1.00	0.00	0.01	0.98	1.00	1.00	1.00	1.00	1529.88	1.01
theta[10,2]	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02	1529.88	1.01
theta_new[1]	0.05	0.00	0.15	0.00	0.00	0.00	0.00	0.59	3662.88	1.00
theta_new[2]	0.95	0.00	0.15	0.41	1.00	1.00	1.00	1.00	3662.88	1.00
lp__	-53.91	0.51	3.91	-63.88	-55.93	-53.60	-51.35	-47.18	58.93	1.05

```r
stan_hist(fitsinglenoH, pars = c("alpha", "beta"), bins=50)
```




```
stan_hist(fitsinglenoH, pars = c("theta_new"), bins=50)
```

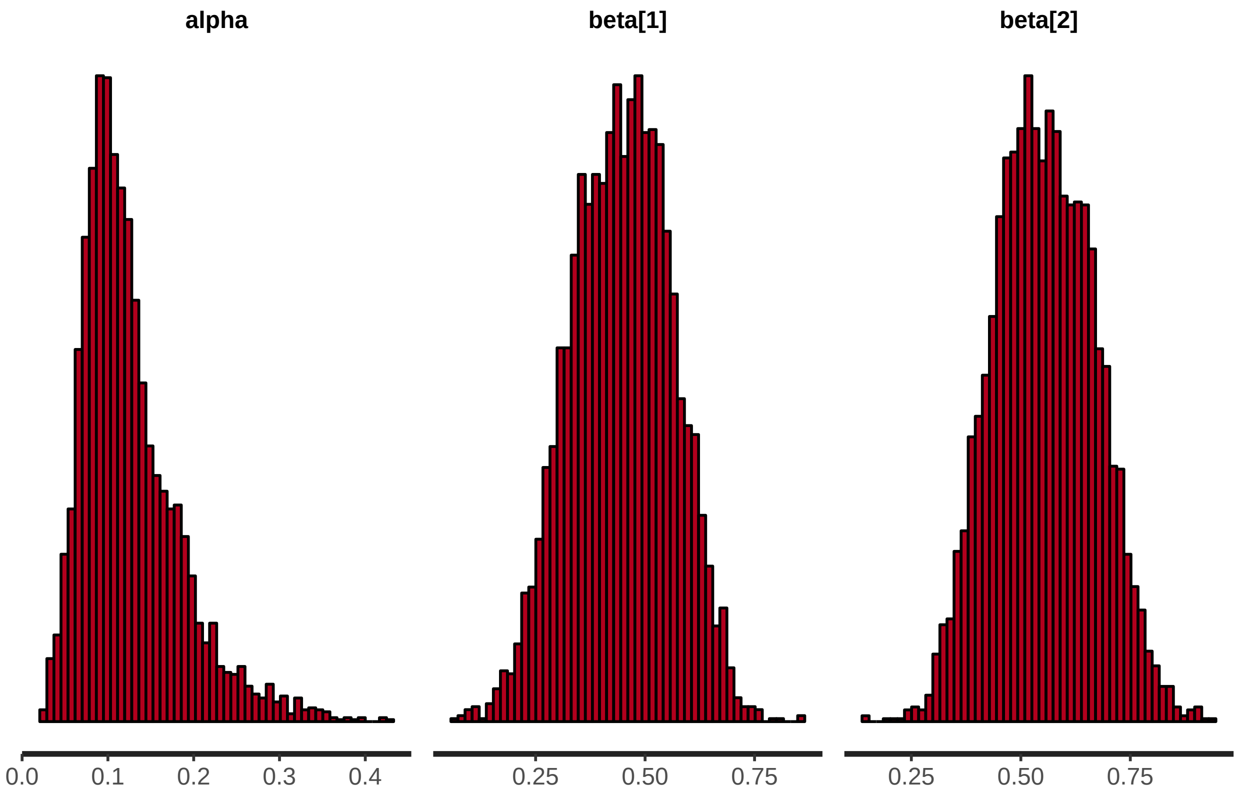



```
fitmixedNH <- stan(file = 'modelNH.stan', data =data.mixed , verbose = FALSE,
            iter = 2000, chains = 4, control=list(adapt_delta=0.99))
```

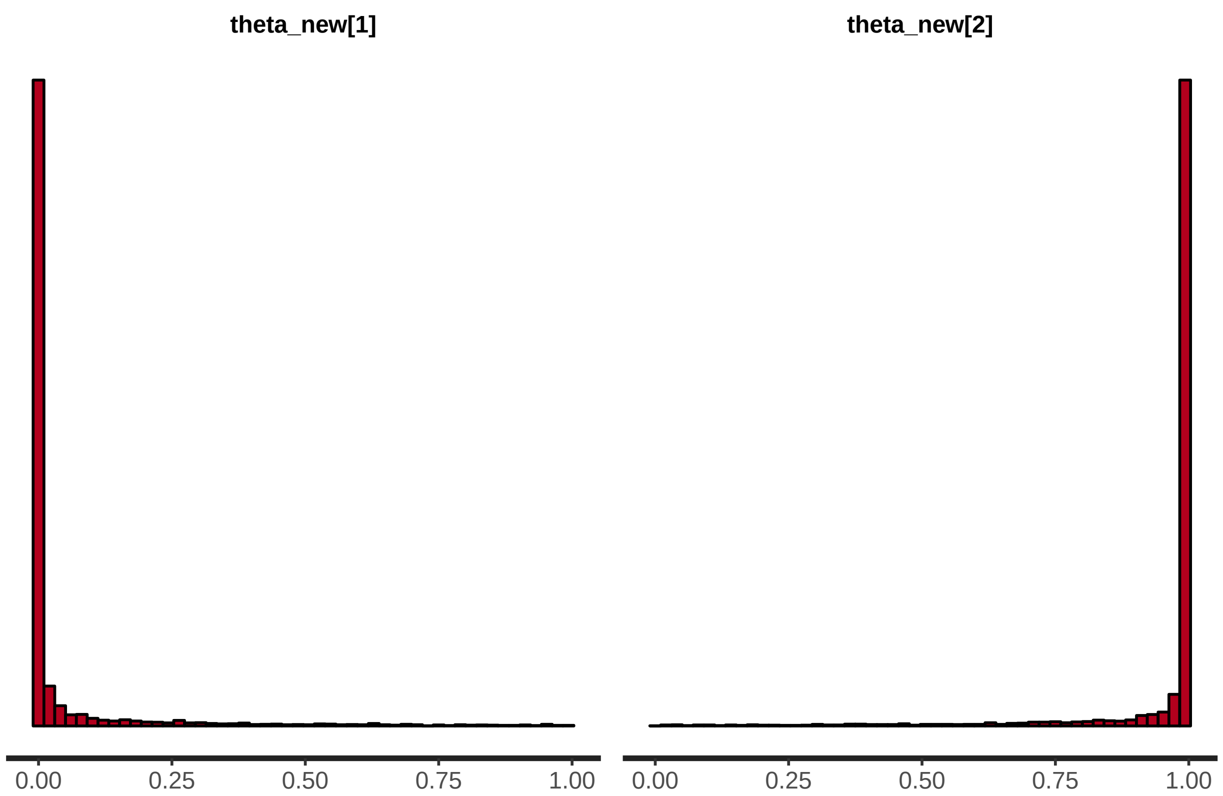
```r
mixNHs <- summary(fitmixedNH)
kable(mixNHs$summary,  digits=2)
```

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
lambda	0.08	0.00	0.09	0.00	0.02	0.05	0.10	0.32	1315.57	1.01
mu	1.28	0.02	0.99	0.15	0.56	1.02	1.73	3.83	3994.05	1.00
alpha	57.47	6.10	101.69	8.35	17.93	29.73	53.20	305.99	278.02	1.02
beta[1]	0.25	0.00	0.03	0.20	0.24	0.25	0.27	0.31	2122.69	1.00
beta[2]	0.75	0.00	0.03	0.69	0.73	0.75	0.76	0.80	2122.69	1.00
theta[1,1]	0.19	0.00	0.06	0.07	0.15	0.19	0.23	0.31	1509.26	1.00
theta[1,2]	0.81	0.00	0.06	0.69	0.77	0.81	0.85	0.93	1509.26	1.00
theta[2,1]	0.19	0.00	0.06	0.08	0.15	0.19	0.23	0.32	1486.41	1.01
theta[2,2]	0.81	0.00	0.06	0.68	0.77	0.81	0.85	0.92	1486.41	1.01
theta[3,1]	0.19	0.00	0.06	0.07	0.15	0.19	0.23	0.31	1412.08	1.00
theta[3,2]	0.81	0.00	0.06	0.69	0.77	0.81	0.85	0.93	1412.08	1.00
theta[4,1]	0.19	0.00	0.06	0.07	0.15	0.19	0.24	0.32	1340.38	1.01
theta[4,2]	0.81	0.00	0.06	0.68	0.76	0.81	0.85	0.93	1340.38	1.01
theta[5,1]	0.19	0.00	0.06	0.08	0.15	0.19	0.24	0.31	1430.64	1.01
theta[5,2]	0.81	0.00	0.06	0.69	0.76	0.81	0.85	0.92	1430.64	1.01
theta[6,1]	0.23	0.00	0.06	0.12	0.19	0.23	0.27	0.36	3085.89	1.00
theta[6,2]	0.77	0.00	0.06	0.64	0.73	0.77	0.81	0.88	3085.89	1.00
theta[7,1]	0.23	0.00	0.06	0.11	0.19	0.23	0.27	0.37	3101.73	1.00
theta[7,2]	0.77	0.00	0.06	0.63	0.73	0.77	0.81	0.89	3101.73	1.00
theta[8,1]	0.23	0.00	0.06	0.12	0.19	0.23	0.27	0.36	3442.70	1.00
theta[8,2]	0.77	0.00	0.06	0.64	0.73	0.77	0.81	0.88	3442.70	1.00
theta[9,1]	0.23	0.00	0.06	0.12	0.19	0.23	0.27	0.35	3313.47	1.00
theta[9,2]	0.77	0.00	0.06	0.65	0.73	0.77	0.81	0.88	3313.47	1.00
theta[10,1]	0.23	0.00	0.06	0.12	0.19	0.23	0.27	0.36	4266.71	1.00
theta[10,2]	0.77	0.00	0.06	0.64	0.73	0.77	0.81	0.88	4266.71	1.00
theta[11,1]	0.27	0.00	0.07	0.16	0.23	0.27	0.31	0.42	3876.94	1.00
theta[11,2]	0.73	0.00	0.07	0.58	0.69	0.73	0.77	0.84	3876.94	1.00
theta[12,1]	0.27	0.00	0.06	0.16	0.23	0.27	0.31	0.42	4297.37	1.00
theta[12,2]	0.73	0.00	0.06	0.58	0.69	0.73	0.77	0.84	4297.37	1.00
theta[13,1]	0.27	0.00	0.06	0.16	0.23	0.27	0.31	0.41	4367.83	1.00
theta[13,2]	0.73	0.00	0.06	0.59	0.69	0.73	0.77	0.84	4367.83	1.00
theta[14,1]	0.27	0.00	0.06	0.16	0.23	0.27	0.31	0.41	4315.18	1.00
theta[14,2]	0.73	0.00	0.06	0.59	0.69	0.73	0.77	0.84	4315.18	1.00
theta[15,1]	0.27	0.00	0.06	0.15	0.23	0.27	0.31	0.41	4031.26	1.00
theta[15,2]	0.73	0.00	0.06	0.59	0.69	0.73	0.77	0.85	4031.26	1.00
theta[16,1]	0.31	0.00	0.07	0.19	0.26	0.31	0.36	0.46	1840.22	1.00
theta[16,2]	0.69	0.00	0.07	0.54	0.64	0.69	0.74	0.81	1840.22	1.00
theta[17,1]	0.31	0.00	0.07	0.19	0.26	0.30	0.35	0.48	2094.78	1.00
theta[17,2]	0.69	0.00	0.07	0.52	0.65	0.70	0.74	0.81	2094.78	1.00
theta[18,1]	0.31	0.00	0.07	0.19	0.26	0.30	0.35	0.47	1928.66	1.00
theta[18,2]	0.69	0.00	0.07	0.53	0.65	0.70	0.74	0.81	1928.66	1.00
theta[19,1]	0.31	0.00	0.07	0.19	0.26	0.30	0.35	0.48	2214.01	1.00
theta[19,2]	0.69	0.00	0.07	0.52	0.65	0.70	0.74	0.81	2214.01	1.00
theta[20,1]	0.31	0.00	0.07	0.19	0.26	0.30	0.35	0.47	2007.70	1.00
theta[20,2]	0.69	0.00	0.07	0.53	0.65	0.70	0.74	0.81	2007.70	1.00
theta_new[1]	0.24	0.00	0.08	0.09	0.19	0.24	0.29	0.43	3690.87	1.00
theta_new[2]	0.76	0.00	0.08	0.57	0.71	0.76	0.81	0.91	3690.87	1.00
lp__	-238.34	0.39	6.98	-249.95	-243.09	-239.11	-234.63	-221.50	315.92	1.02

```
stan_hist(fitsinglenoH, pars = c("alpha", "beta"), bins=50)
```




```
stan_hist(fitsinglenoH, pars = c("theta_new"), bins=50)
```



4. Hyperparameter settings

Exercise: Play with changing the value of the hyperparameter (ω). Try three different values, fairly far apart — try different orders of magnitude. To do this, you will have to specify the `omega_param` when you call `generate_data`.

Questions:

- Does this change have an effect on the final posterior, for example on theta_new?
- Does it matter which data (mixed or homogeneous) you use?
- Does any change that you see make sense, given what you know about the distributions?

Hints:

- Draws from Exponentials with smaller parametrisations (i.e, smaller values of ω) will tend to be larger, since the mean/expected values is $E[X] = 1/\omega$ (see https://en.wikipedia.org/wiki/Exponential_distribution)
- For alpha, this end up cancelling out (you're drawing from exp(exp(omega)), with expected value $\frac{1}{1/\omega} = \omega$ so smaller omega correspond to (expected) smaller α;
- For β: smaller values of ω will lead to larger values of μ, and $\beta \sim Dirichlet(\mu)$; remember that here μ is playing the role of 'concentration parameter' (so same as α on the layer below), while the 'base distribution' is uniform over all categories.
- The concentration parameter governs how concentrated draws from the Dirichlet will be to the base distribution. Small values will lead to greater spread, while large values lead to most draws staying close to the base distribution.

Note: The sampler may start giving you warnings about divergences; these are important indicators that the hyperparameter values you've given aren't ideal, since it makes it harder for the sampler to find the correct posterior distribution. However, you can ignore them for now, as long as the Rhat values look OK.

Solution

This is meant to be be an exercise for you to explore how to setup, modify and evaluate a Bayesian model in Stan. The results will vary depending on the values you pick. In general smaller ω values should lead to smaller α and β values and thus no strong prior for thetas to be close to beta/base distribution. Larger values for ω results in larger α and β values, more distant from a uniform distribution. This results in a strong prior for θ to be skewed as well.

5. Learning at different levels

a) Inspect Figure 6 in Kemp, Perfors, and Tenenbaum (2007), which shows how the model can learn more confident, i.e., tighter, predictions at different levels, depending on the data it's exposed to. Again, first, look at the figure and read the figure caption; make sure you understand the reason for the shapes of the posteriors.
b) Now recreate the figures with your model, by giving it the different datasets (have a look at `data.R` for the generating functions).
c) What effect will the tighter distributions over α, β in scenario b(iii) have on theta? Does this pattern of thetas in b(iii) depend on the values of α, β, or just the certainty? (I.e. if the peaks of the posterior α and β were elsewhere, but the shape was the same, would the pattern of theta look the same?)

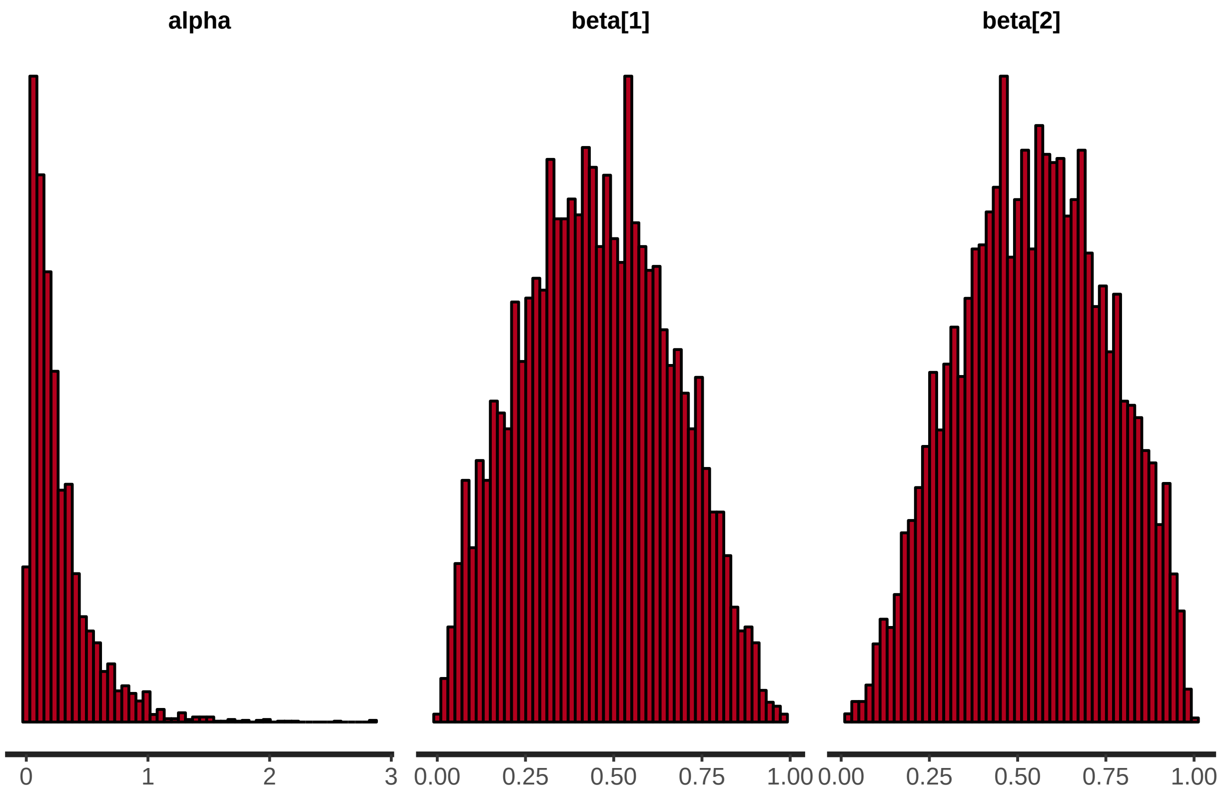
Solution

```r
data.big_bags <- generate_data.onecolor(64, 2,2, 1)
data.many_bags <- generate_data.onecolor(64, 32,2, 1)
data.mix <- generate_data.sets(64, c(5,5,22), 1)
```

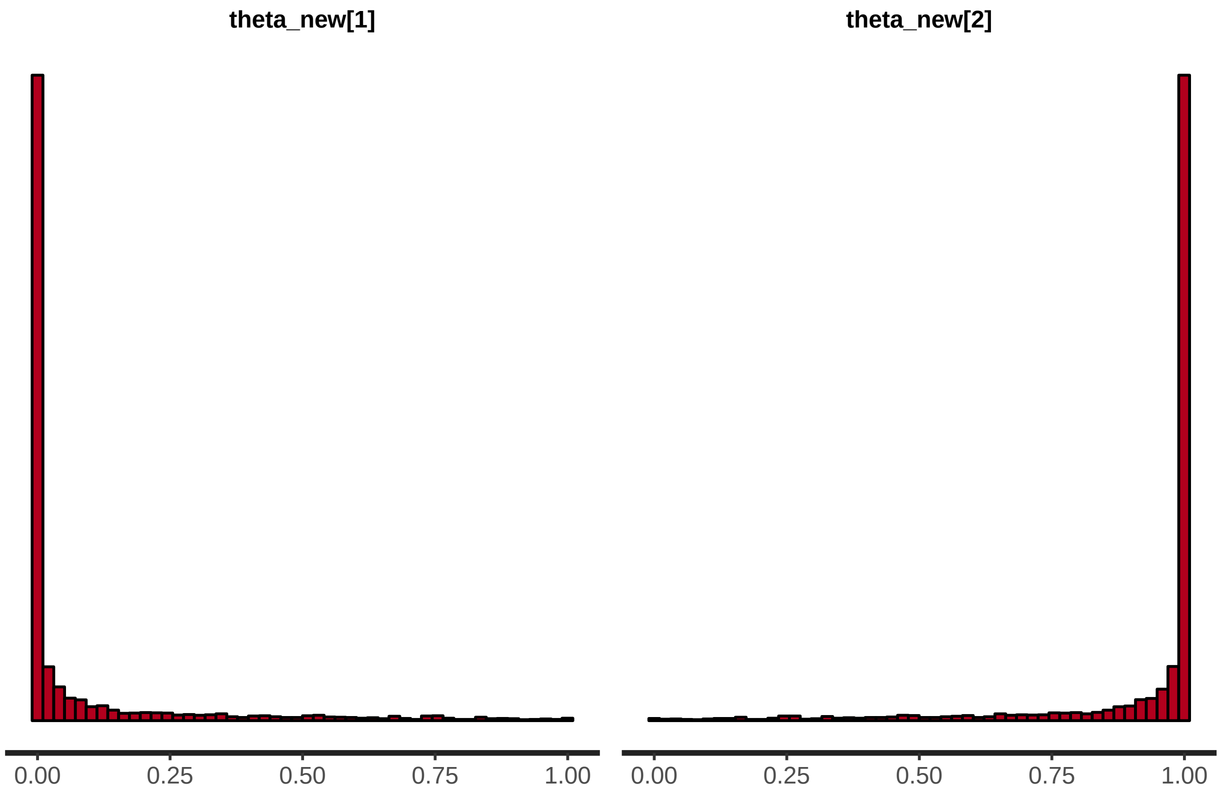
```r
fitbb <- stan(file = 'model.stan', data=data.big_bags, verbose = FALSE,
              iter=2000, chains=4, control=list(adapt_delta=0.95))
```

```
## Warning: There were 290 divergent transitions after warmup. Increasing adapt_delta above 0.95 may hel
## http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

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

```r
stan_hist(fitbb, pars = c("alpha", "beta"), bins=50)
```



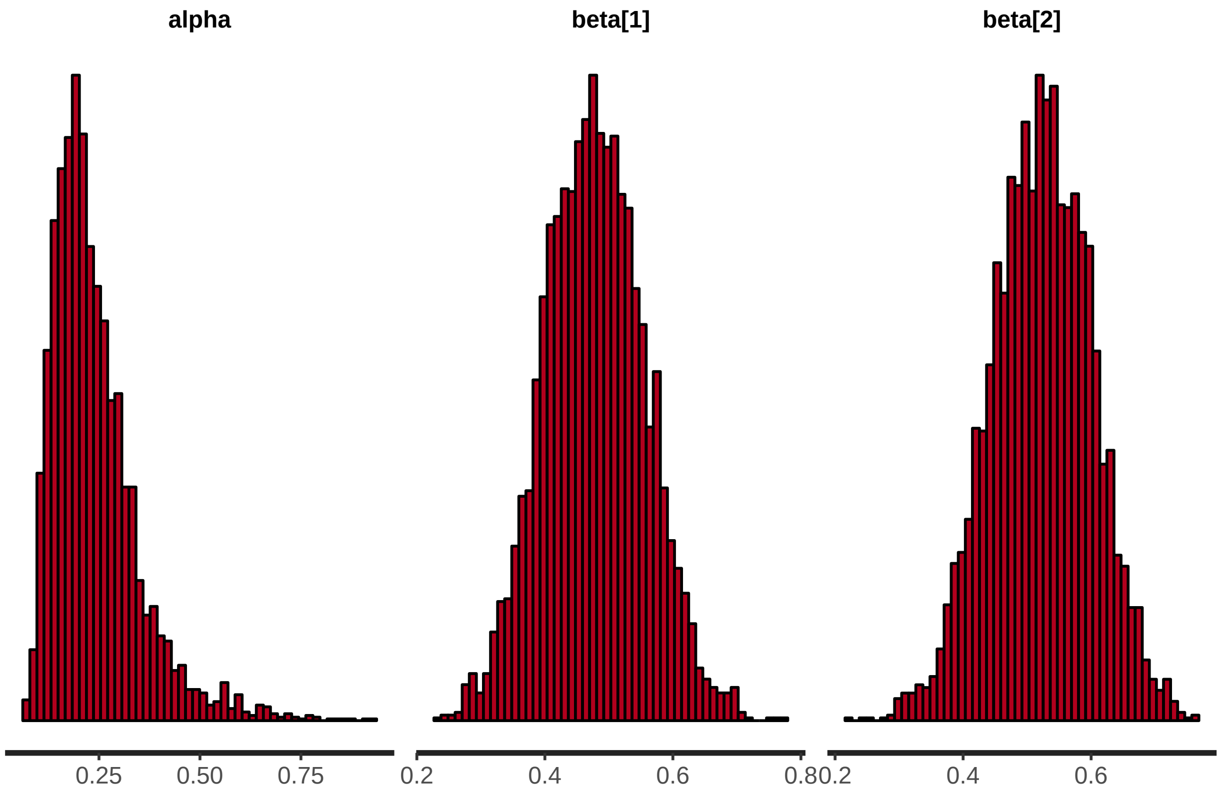

```r
stan_hist(fitbb, pars = c("theta_new"), bins=50)
```



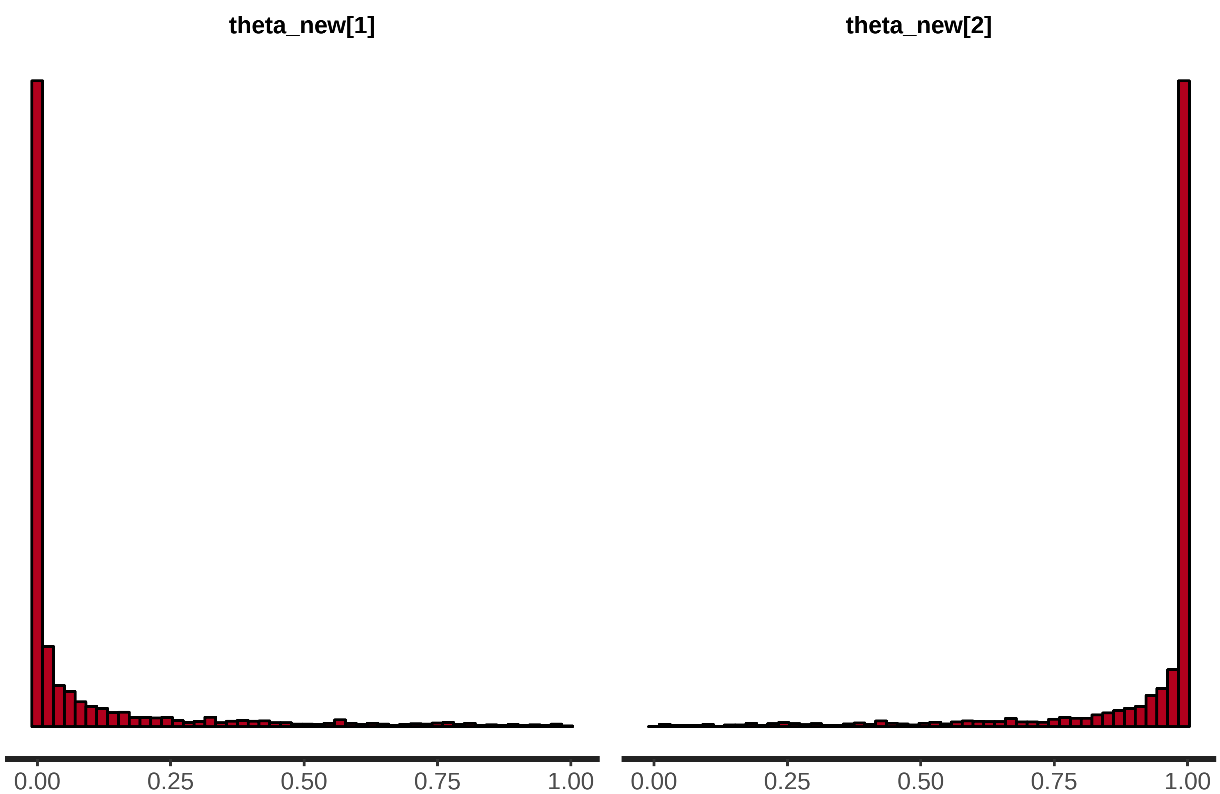

```r
fitmb <- stan(file = 'model.stan', data=data.many_bags, verbose = FALSE,
            iter = 2000, chains = 4, control=list(adapt_delta=0.95))
```

```
## Warning: There were 1273 divergent transitions after warmup. Increasing adapt_delta above 0.95 may he
## http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

## Warning: There were 810 transitions after warmup that exceeded the maximum treedepth. Increase max_tr
## http://mc-stan.org/misc/warnings.html#maximum-treedepth-exceeded

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

```r
stan_hist(fitmb, pars = c("alpha", "beta"), bins=50)
```



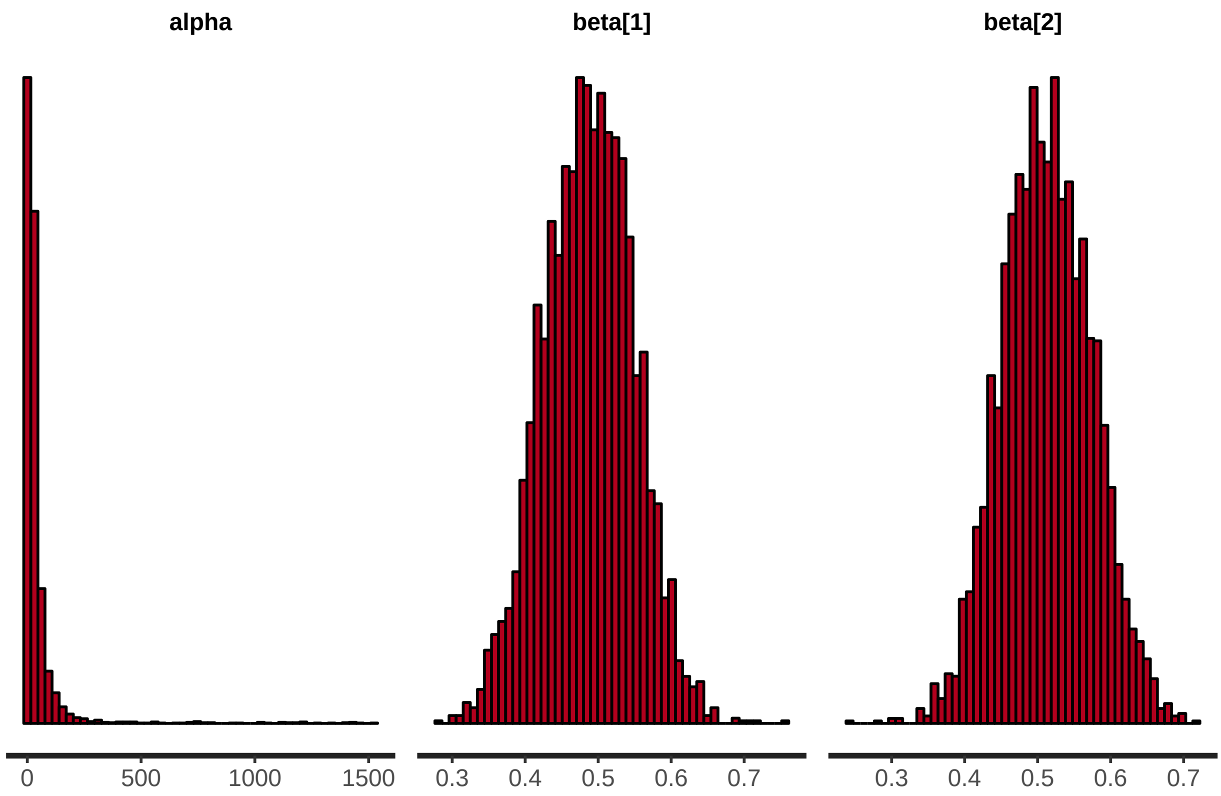

```
stan_hist(fitmb, pars = c("theta_new"), bins=50)
```



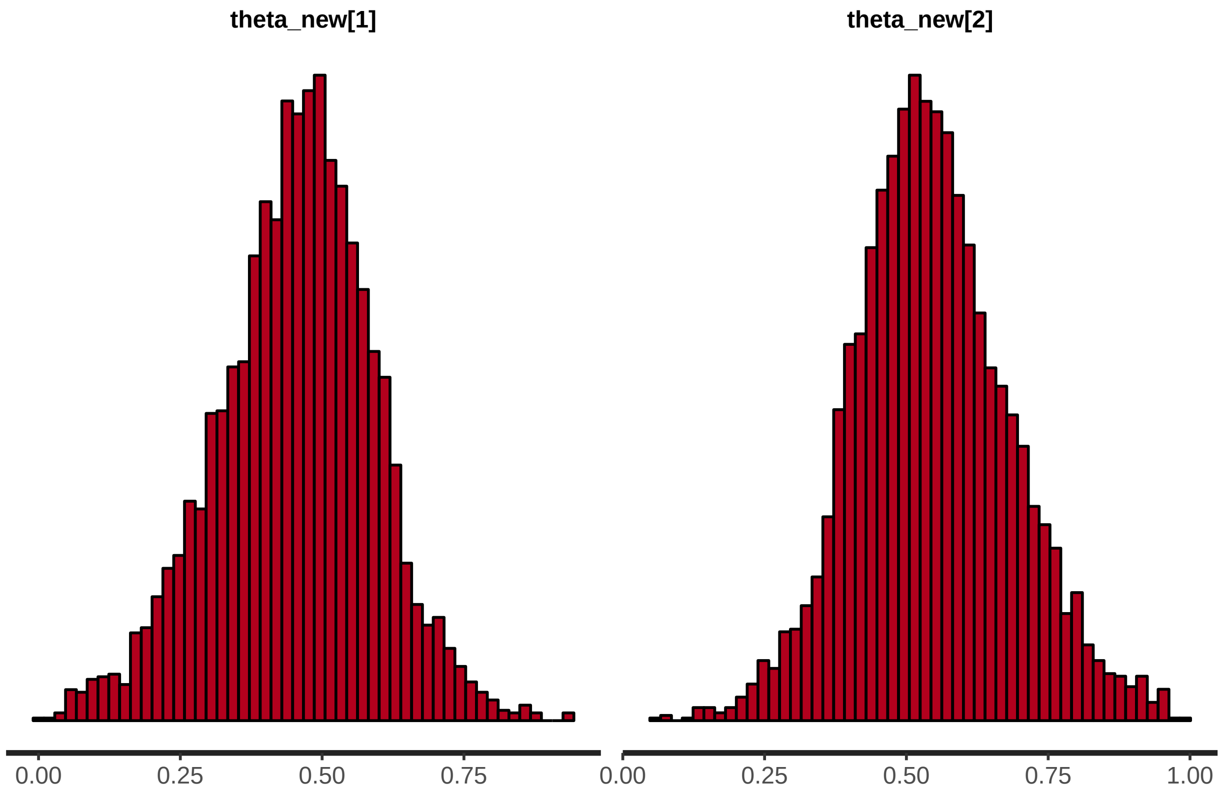

```
fitmix <- stan(file = 'model.stan', data=data.mix, verbose = FALSE,
               iter = 2000, chains = 4, control=list(adapt_delta=0.95))
```

```
## Warning: There were 1 divergent transitions after warmup. Increasing adapt_delta above 0.95 may help
## http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

## Warning: There were 4 chains where the estimated Bayesian Fraction of Missing Information was low. Se
## http://mc-stan.org/misc/warnings.html#bfmi-low

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

```r
stan_hist(fitmix, pars = c("alpha", "beta"), bins=50)
```




```r
stan_hist(fitmix, pars = c("theta_new"), bins=50)
```



The fact that the distributions of α and β are peaked on >1 values for α and values for the uniform(.5,.5) distribution for β means that the thetas will have strong priors towards a uniform distribution as well (as the plot shows). The value (>1) of alpha matters here, because it encourages flat and uniform θ, in contrast to 6bi and ii, where small alpha values result in moving away from the uniform distribution, towards the very skewed thetas seen in the figure.

References

Kemp, Charles, Amy Perfors, and Joshua Tenenbaum. 2007. "Learning Overhypotheses with Hierarchical Bayesian Models." *Developmental Science* 10. Wiley Online Library.