

# Introduction to Bayesian Statistics with R

## 5: Exercise solutions

Jack Kuipers

29 November 2022

First we load the tidyverse, brms and set a seed.

```
library(tidyverse); options(dplyr.summarise.inform = FALSE) # suppress summarise warnings
library(brms)
set.seed(42)
```

### Exercise 5.1 - a Bayesian analysis

For the data (*lung\_data.csv*) from Exercise 1.1 (from a small clinical trial on asthma patients measuring the lung function of a control group on a placebo and a treatment group on a new drug)

- run a Bayesian analysis to explore the difference in function between the two groups
- visualise the output, especially for the component of interest
- compare the 95% credible interval to the 95% confidence interval of Exercise 1.1.

Did you run the analogue of the equal/unequal variance *t*-test, and did you check the posterior?

We first read in the data

```
lung_data <- read.csv("./data/lung_data.csv")
```

which we can plug straight into a *brms* model. Since in Exercise 1.1 we used the default *t*-test which has unequal variance, we'll work analogously here. Following the lecture example, we use the helper function from the Exercise sheet to run

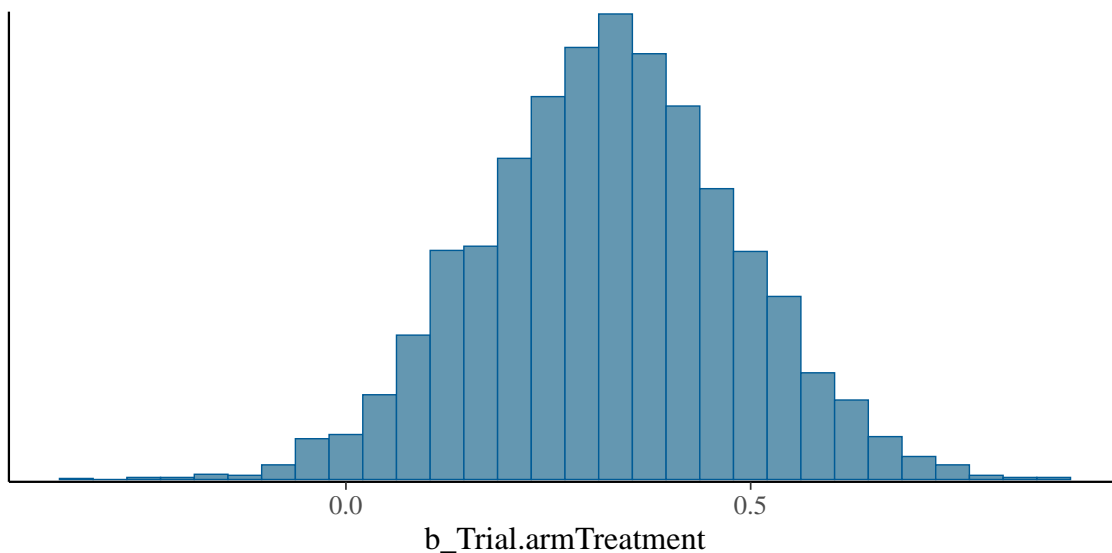
```
brmfit_t_ex5 <- run_model(brm(bf(Lung.function ~ Trial.arm, sigma ~ Trial.arm), lung_data),
                          "./brm_models_exercises/t_test_ex5")
summary(brmfit_t_ex5)
```

```
## Family: gaussian
## Links: mu = identity; sigma = log
## Formula: Lung.function ~ Trial.arm
##          sigma ~ Trial.arm
## Data: lung_data (Number of observations: 40)
## Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup draws = 4000
##
## Population-Level Effects:
##              Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS
## Intercept          10.15      0.12   9.92  10.38 1.00     3631
## sigma_Intercept     -0.70      0.17  -1.01  -0.35 1.00     3881
```

```
## Trial.armTreatment      0.32      0.16      0.02      0.63 1.00      3355
## sigma_Trial.armTreatment -0.02      0.24     -0.50      0.45 1.00      3754
##                               Tail_ESS
## Intercept              2416
## sigma_Intercept        2624
## Trial.armTreatment       2942
## sigma_Trial.armTreatment 2925
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

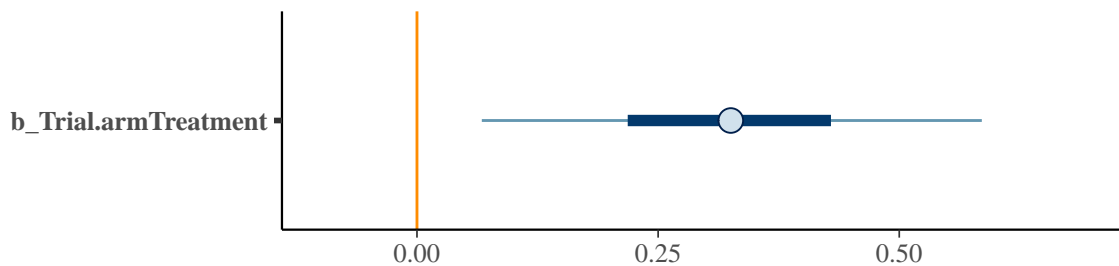
Now that we have run the model, we can visualise the output, and we focus on the coefficient of the Trial.arm.

```
mcmc_plot(brmfit_t_ex5, variable = "b_Trial.armTreatment", type = "hist")
```



Looks like a distribution some way from 0, but let's check the quantiles

```
mcmc_plot(brmfit_t_ex5, variable = "b_Trial.armTreatment", type = "intervals") +
  geom_vline(xintercept = 0, color = "darkorange") + xlim(-0.1, 0.7) # add line at 0
```



Here we need to remember that the default intervals shown are the 50% (bold) and 90% (thin) with the posterior median as the dot. The 90% credible interval excludes 0, but what about the 95% one? That was already printed out in the `summary` above, and we can also extract it directly

```
fixef(brmfit_t_ex5)["Trial.armTreatment", ]
```

```
## Estimate Est.Error      Q2.5      Q97.5
## 0.32417665 0.15912535 0.01524432 0.63261832
```

and we can see that it also (just) excludes 0. To compare to the confidence interval of the  $t$ -test, we quickly compute

```
t.test(Lung.function ~ Trial.arm, lung_data)$conf.int
```

```
## [1] -0.63003538 -0.01996462
## attr("conf.level")
## [1] 0.95
```

and can see that (apart from the reflection of sign) it is numerically extremely similar to the credible interval. Of course the interpretation is somewhat different. In the Bayesian version, the coefficient of interest is a random variable and based on our prior (more on that in the next session!), data and modelling assumptions our current belief in its distribution and its uncertainty is represented in the plots (and credible interval) above. The confidence interval is a range centred on the observed difference in means (our best estimate for this difference) with a width corresponding to the typical uncertainty we would expect for the mean difference (based on the standard deviations of this sample data).

For the final questions of the Exercise, we ran the unequal variance case, but for the equal variance case we might run either of the two following options

```
brm(Lung.function ~ Trial.arm, lung_data)
brm(bf(Lung.function ~ Trial.arm, sigma ~ 1), lung_data)
```

where the difference will be that the second uses the log-link for  $\sigma$  (like for our unequal variance case above). And we haven't checked the posterior yet. Let's look at the posterior predictive check

```
pp_check(brmfit_t_ex5, ndraws = 40)
```

