

Data Analytics & Communication Assignment 6

Micky Labreche s4021290 & Yann Amado s5091128

15-01-2023

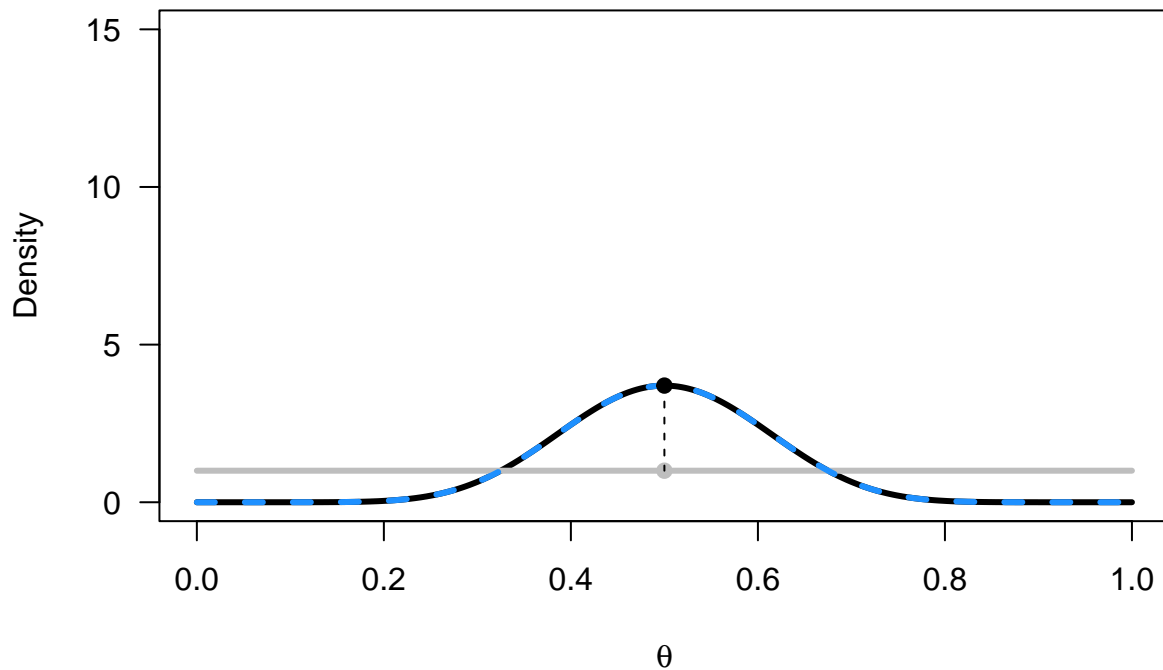
1 Priors, posteriors and likelihoods

```
H0<-0.5 #Set the point null hypothesis you want to calculate the Bayes Factor for
n<-20 #set total trials
x<-10 #set successes
aprior<-1 #Set the alpha for the Beta distribution representing the prior
bprior<-1 #Set the beta for the Beta distribution representing the prior

alikelikelihood<-x+1 #Calculate the alpha for the Beta distribution for the likelihood
blikeliikelihood<-n-x+1 #Calculate the beta for the Beta distribution for the likelihood
aposterior<-aprior+alikelikelihood-1 #Calculate the alpha for the Beta distribution for the
# posterior
bposterior<-bprior+blikeliikelihood-1 #Calculate the beta for the Beta distribution for the
# posterior

theta<-seq(0,1,0.001) #create theta range from 0 to 1
prior <- dbeta(theta, aprior, bprior)
likelihood <- dbeta(theta, alikelikelihood, blikeliikelihood)
posterior <- dbeta(theta, aposterior, bposterior)
plot(theta, posterior, ylim=c(0, 15), type = "l", lwd = 3, xlab = bquote(theta),
      ylab = "Density", las = 1)
lines(theta, prior, col="grey", lwd = 3)
lines(theta, likelihood, lty = 2, lwd = 3, col="dodgerblue")
# define a Bayes Factor comparing prior and posterior
BF<-dbeta(H0, aposterior, bposterior)/dbeta(H0, aprior, bprior)
points(H0,dbeta(H0, aposterior, bposterior), pch = 19)
points(H0,dbeta(H0, aprior, bprior), pch = 19, col="grey")
segments(H0, dbeta(H0, aposterior, bposterior), H0, dbeta(H0, aprior, bprior), lty=2)
title(paste('Bayes Factor:',round(BF,digits=2)))
```

Bayes Factor: 3.7



a) The code calculates the probability density function of the prior, likelihood and posterior functions, and it does this by calculating the beta distribution of each taking into account the corresponding parameters. The graph is then a result of those calculations, the prior belief is that the probability distribution is equally likely, hence its constant shape; the likelihood takes into consideration the amount of successes and failures to then construct using the beta distribution a likelihood pdf, which looks like a standard normal distribution as expected of the flipping of a normal coin. Finally, the posterior pdf is calculated by summing the prior and the likelihood, which is equivalent to the multiplication of the previous graphs.

From this graph, it's possible to conclude that the null hypothesis is, according to the Bayes Factor, 3.7 times more likely than the alternative hypothesis, suggesting that a substantial difference exists.

b)

```
H0<-0.675 #Set the point null hypothesis you want to calculate the Bayes Factor for
n<-20 #set total trials
x<-10 #set successes
aprior<-1 #Set the alpha for the Beta distribution representing the prior
bprior<-1 #Set the beta for the Beta distribution representing the prior

alikehood<-x+1 #Calculate the alpha for the Beta distribution for the likelihood
blikehood<-n-x+1 #Calculate the beta for the Beta distribution for the likelihood
aposterior<-aprior+alikehood-1 #Calculate the alpha for the Beta distribution for the
# posterior
bposterior<-bprior+blikehood-1 #Calculate the beta for the Beta distribution for the
# posterior

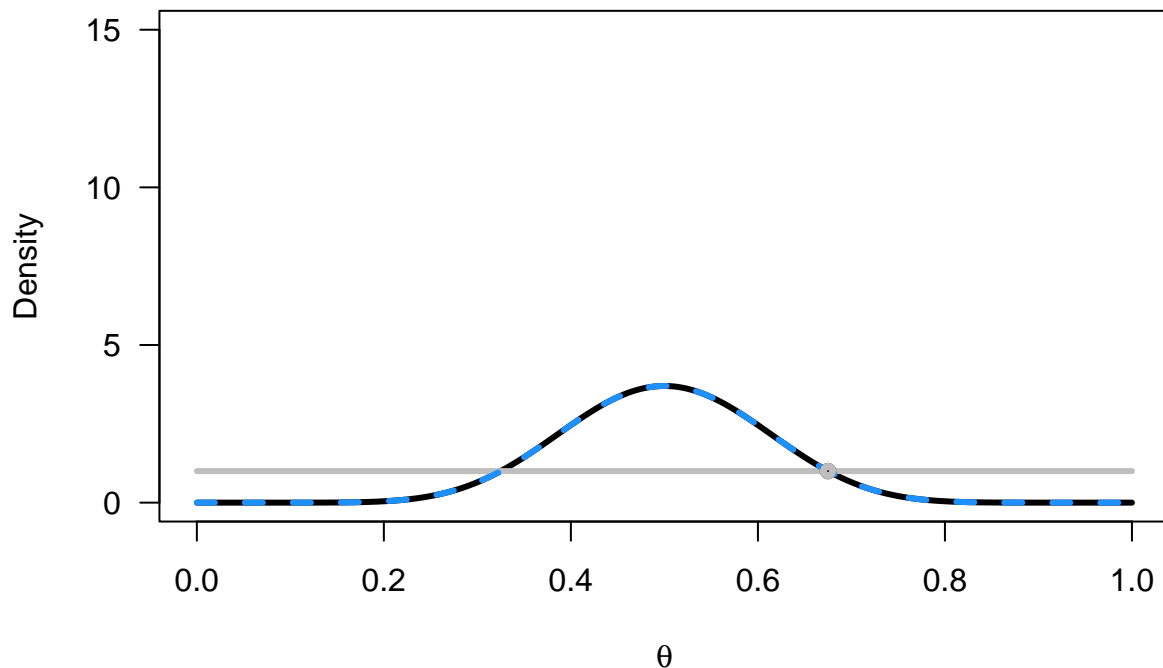
theta<-seq(0,1,0.001) #create theta range from 0 to 1
```

```

prior <- dbeta(theta, aprior, bprior)
likelihood <- dbeta(theta, alikelihood, blikelihood)
posterior <- dbeta(theta, aposterior, bposterior)
plot(theta, posterior, ylim=c(0, 15), type = "l", lwd = 3, xlab = bquote(theta),
      ylab = "Density", las = 1)
lines(theta, prior, col="grey", lwd = 3)
lines(theta, likelihood, lty = 2, lwd = 3, col="dodgerblue")
# define a Bayes Factor comparing prior and posterior
BF<-dbeta(H0, aposterior, bposterior)/dbeta(H0, aprior, bprior)
points(H0,dbeta(H0, aposterior, bposterior), pch = 19)
points(H0,dbeta(H0, aprior, bprior), pch = 19, col="grey")
segments(H0, dbeta(H0, aposterior, bposterior), H0, dbeta(H0, aprior, bprior), lty=2)
title(paste('Bayes Factor:',round(BF,digits=2)))

```

Bayes Factor: 1



With the null hypothesis as 67.5% heads, it's not possible to conclude which hypothesis is more likely since the bayesian factor has a value of 1, meaning that both are equally likely. As the null hypothesis value directly dictates the probability that it has generated the data, the posterior probability of the null hypothesis will change accordingly, affecting the final outcome of the Bayes Factor since it's the ratio between the posterior of the alternative and null hypothesis.

c)

```

H0<-0.5 #Set the point null hypothesis you want to calculate the Bayes Factor for
n<-20 #set total trials
x<-10 #set successes
aprior<-100 #Set the alpha for the Beta distribution representing the prior

```

```

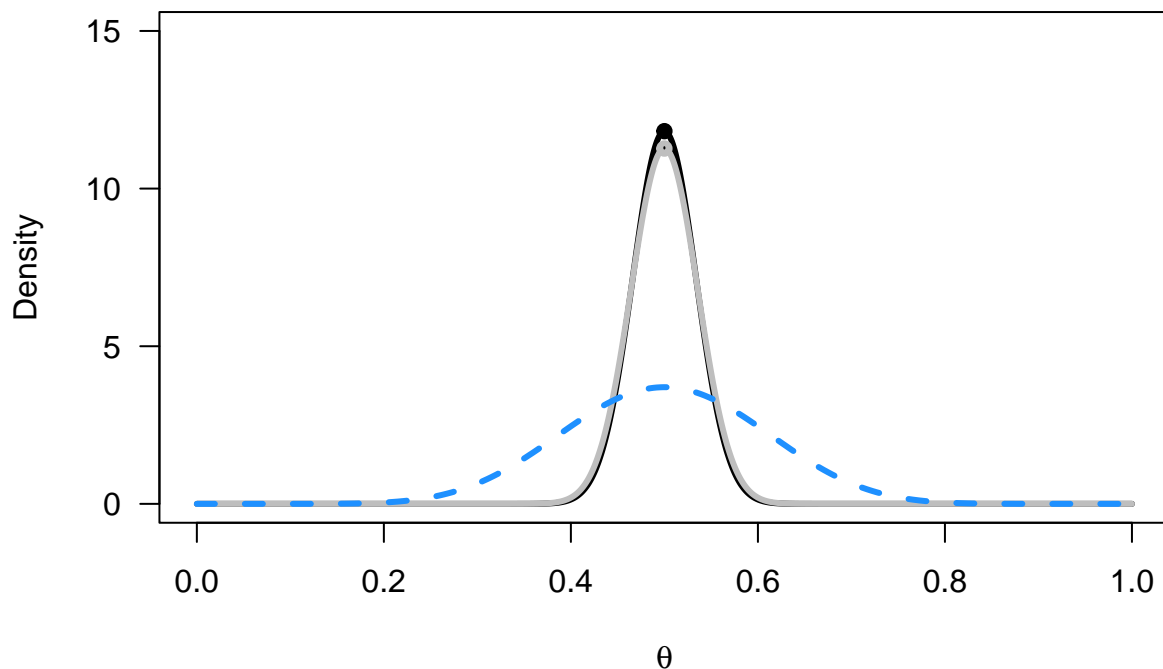
bprior<-100 #Set the beta for the Beta distribution representing the prior

alikehood<-x+1 #Calculate the alpha for the Beta distribution for the likelihood
blikehood<-n-x+1 #Calculate the beta for the Beta distribution for the likelihood
aposterior<-aprior+alikehood-1 #Calculate the alpha for the Beta distribution for the
# posterior
bposterior<-bprior+blikehood-1 #Calculate the beta for the Beta distribution for the
# posterior

theta<-seq(0,1,0.001) #create theta range from 0 to 1
prior <- dbeta(theta, aprior, bprior)
likelihood <- dbeta(theta, alikehood, blikehood)
posterior <- dbeta(theta, aposterior, bposterior)
plot(theta, posterior, ylim=c(0, 15), type = "l", lwd = 3, xlab = bquote(theta),
      ylab = "Density", las = 1)
lines(theta, prior, col="grey", lwd = 3)
lines(theta, likelihood, lty = 2, lwd = 3, col="dodgerblue")
# define a Bayes Factor comparing prior and posterior
BF<-dbeta(H0, aposterior, bposterior)/dbeta(H0, aprior, bprior)
points(H0,dbeta(H0, aposterior, bposterior), pch = 19)
points(H0,dbeta(H0, aprior, bprior), pch = 19, col="grey")
segments(H0, dbeta(H0, aposterior, bposterior), H0, dbeta(H0, aprior, bprior), lty=2)
title(paste('Bayes Factor:',round(BF,digits=2)))

```

Bayes Factor: 1.05



now more strongly believe, due to more prior data points being collected for example, that the distribution should have a certain shape; because of this, the likelihood derived from the data will not have the same effect as it did before, and instead, will have a slight effect on the update of our posterior belief. This shape of the prior accounts for the skeptic since this one will have a stronger belief of their previous information, and will not change his beliefs as much; which is the case for when several previous data points have formed our current prior, and then a few extra data points form our likelihood and subsequent prior.

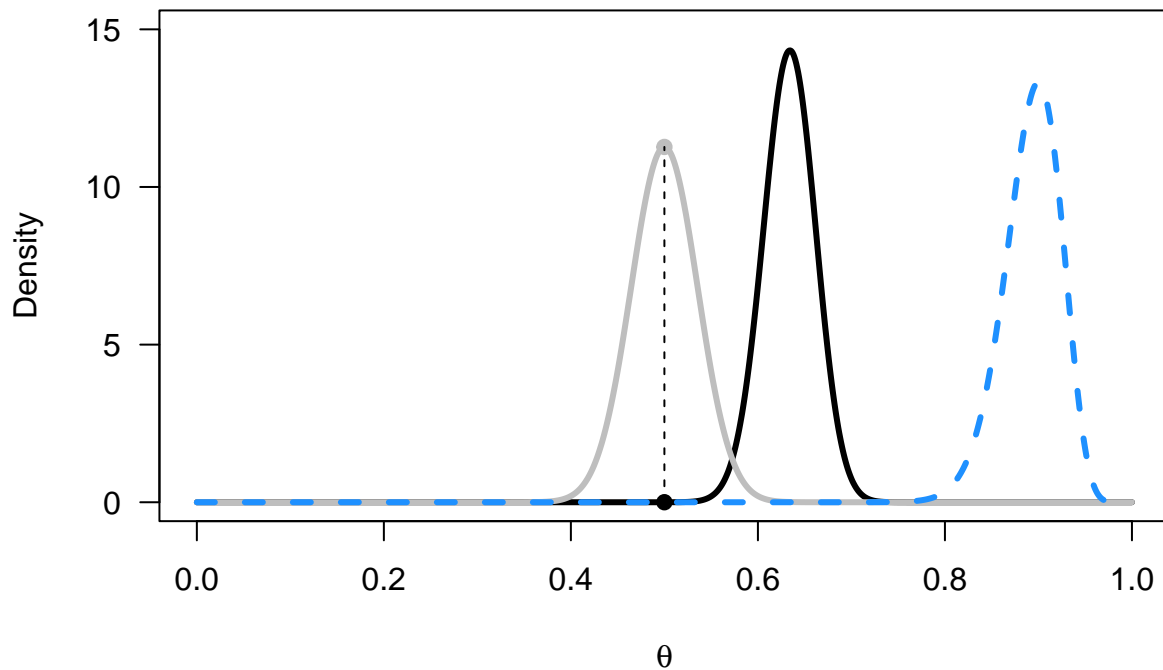
d)

```
H0<-0.5 #Set the point null hypothesis you want to calculate the Bayes Factor for
n<-100 #set total trials
x<-90 #set successes
aprior<-100 #Set the alpha for the Beta distribution representing the prior
bprior<-100 #Set the beta for the Beta distribution representing the prior

alikehood<-x+1 #Calculate the alpha for the Beta distribution for the likelihood
blikehood<-n-x+1 #Calculate the beta for the Beta distribution for the likelihood
aposterior<-aprior+alikehood-1 #Calculate the alpha for the Beta distribution for the
# posterior
bposterior<-bprior+blikehood-1 #Calculate the beta for the Beta distribution for the
# posterior

theta<-seq(0,1,0.001) #create theta range from 0 to 1
prior <- dbeta(theta, aprior, bprior)
likelihood <- dbeta(theta, alikehood, blikehood)
posterior <- dbeta(theta, aposterior, bposterior)
plot(theta, posterior, ylim=c(0, 15), type = "l", lwd = 3, xlab = bquote(theta),
      ylab = "Density", las = 1)
lines(theta, prior, col="grey", lwd = 3)
lines(theta, likelihood, lty = 2, lwd = 3, col="dodgerblue")
# define a Bayes Factor comparing prior and posterior
BF<-dbeta(H0, aposterior, bposterior)/dbeta(H0, aprior, bprior)
points(H0,dbeta(H0, aposterior, bposterior), pch = 19)
points(H0,dbeta(H0, aprior, bprior), pch = 19, col="grey")
segments(H0, dbeta(H0, aposterior, bposterior), H0, dbeta(H0, aprior, bprior), lty=2)
title(paste('Bayes Factor:',round(BF,digits=2)))
```

Bayes Factor: 0



The skeptic would conclude that the alternative hypothesis is strongly more likely than the null hypothesis and accordingly updating it's posterior beliefs by a considerable amount, but still not entirely fitting the data and instead shifting their beliefs towards $\theta = 0.9$. The graph illustrates that conclusion, by showing the shift of the prior represented by the gray line towards what the data dictates, represented by the blue line, with the final posterior being the black line.

```

H0<-0.5 #Set the point null hypothesis you want to calculate the Bayes Factor for
n<-100 #set total trials
x<-90 #set successes
aprior<-1 #Set the alpha for the Beta distribution representing the prior
bprior<-1 #Set the beta for the Beta distribution representing the prior

alikehood<-x+1 #Calculate the alpha for the Beta distribution for the likelihood
blikehood<-n-x+1 #Calculate the beta for the Beta distribution for the likelihood
aposterior<-aprior+alikehood-1 #Calculate the alpha for the Beta distribution for the
# posterior
bposterior<-bprior+blikehood-1 #Calculate the beta for the Beta distribution for the
# posterior

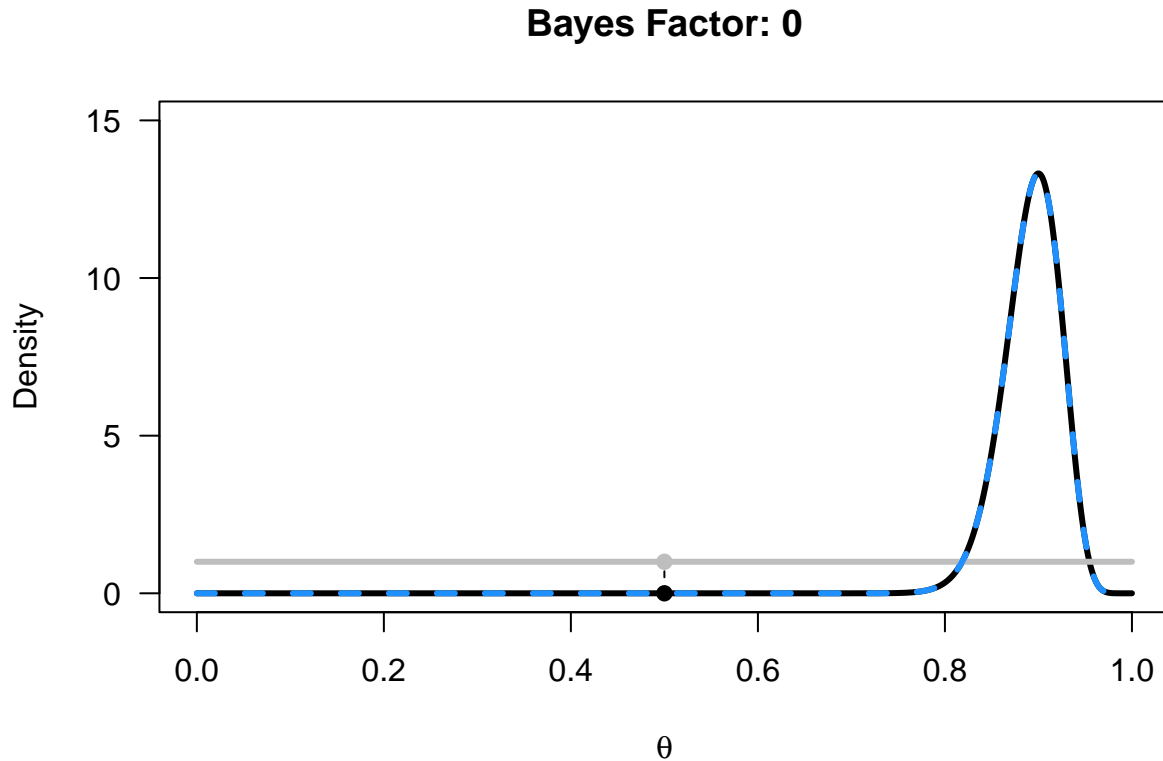
theta<-seq(0,1,0.001) #create theta range from 0 to 1
prior <- dbeta(theta, aprior, bprior)
likelihood <- dbeta(theta, alikehood, blikehood)
posterior <- dbeta(theta, aposterior, bposterior)
plot(theta, posterior, ylim=c(0, 15), type = "l", lwd = 3, xlab = bquote(theta),
      ylab = "Density", las = 1)
lines(theta, prior, col="grey", lwd = 3)

```

```

lines(theta, likelihood, lty = 2, lwd = 3, col="dodgerblue")
# define a Bayes Factor comparing prior and posterior
BF<-dbeta(H0, aposterior, bposterior)/dbeta(H0, aprior, bprior)
points(H0,dbeta(H0, aposterior, bposterior), pch = 19)
points(H0,dbeta(H0, aprior, bprior), pch = 19, col="grey")
segments(H0, dbeta(H0, aposterior, bposterior), H0, dbeta(H0, aprior, bprior), lty=2)
title(paste('Bayes Factor:',round(BF,digits=2)))

```



As for someone with prior distributions in a , since they did not believe as strongly in their priors as the skeptic, they'll update their posterior beliefs to one that will completely fit the data, and will now believe that heads is 9 times more likely than tails.

2 Doing a Bayesian Data Analysis

a)

```

onlyDotsAll <- dat[dat$isDots==1,]

# selecting only correct trials
onlyDotsAll <- onlyDotsAll[onlyDotsAll$ER != 1,]

# removing outliers
cleaned <- onlyDotsAll %>%
  filter(!(abs(RT - median(RT)) > 2*sd(RT)))

```

```

average <- onlyDotsAll %>%
  group_by(cohFac, subjNo) %>%
  summarise_all(median)

cohFacHighAll <- average %>% filter(cohFac == 1)
cohFacLowAll <- average %>% filter(cohFac == 0)

#t.test(cohFacLowAll$RT, cohFacHighAll$RT, paired=TRUE)
ttestBF(x = cohFacLowAll$RT, y = cohFacHighAll$RT, paired=TRUE)

## Bayes factor analysis
## -----
## [1] Alt., r=0.707 : 6460.968 ±0%
##
## Against denominator:
##   Null, mu = 0
## ---
## Bayes factor type: BFoneSample, JZS

```

The Bayes factor for the t-test is 6460.97. This means that there's a strong difference in the response times between a high coherence of dots and a low coherence of dots when comparing using all participants.

- b) Yes. Frequentist t-test only provides evidence to reject the null-hypothesis and does not say anything about supporting the alternative hypothesis; in contrast, Bayesian t-test actively provides evidence in favor of the alternative hypothesis. In other words, while both methods arrive at the same result (choosing the alternative hypothesis), the interpretation of those results is different: for the frequentist method, we only reject the null hypothesis, as for bayesian, we actively provide evidence for the alternative. This can be summarized by “it can't be the null hypothesis so it must be the alternative” against “the alternative hypothesis is more likely of being right when compared to the null hypothesis”.
- c) Two of the benefits of the Bayes factor method in relation to Frequentist statistics are: It can provide evidence in favor of the null hypothesis, instead of just rejecting it; as well as the capability of comparing the two hypothesis against each other by considering the data under both of them, which makes it robust to cases where even though the null hypothesis is unlikely, the alternative is even less likely.

#Bootstrap We use the decision.dat dataset to do this next exercise.

```

onlyDots <- dat %>% filter(isDots == 1)

diffRT <- onlyDots %>%
  group_by(cohFac, subjNo) %>%
  summarise_all(median) %>%
  group_by(subjNo) %>%
  summarise_all(diff) %>%
  mutate(RT.abs = abs(RT)) %>%
  pull(RT.abs)

```

- (a) We only use data in the dots condition and not the arrows control trials. We first have to create a vector of differences which. We need to get the difference between the mean RT in two conditions for each participant.


```
diffRT.boot <- boot(diffRT,function(x,i) mean(x[i]),R=1000)
```

In the difference vector we have 23 entries, one for each participant. Because we have a relatively small sample size we decide to use a bootstrap. This randomly sub-samples from our sample like it is the population. Eventually we get a distribution with a larger sample size. We can use this for statistical analysis. The bootstrap samples 1000 times and will calculate the mean.

```
boot.ci(boot.out = diffRT.boot, conf = c(0.95, 0.99), type = "perc")
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 1000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = diffRT.boot, conf = c(0.95, 0.99), type = "perc")
##
## Intervals :
## Level      Percentile
## 95%      ( 0.1582,  0.3007 )
## 99%      ( 0.1423,  0.3260 )
## Calculations and Intervals on Original Scale
## Some percentile intervals may be unstable
```

- (b) We can use the confidence interval to see if the difference between the low and high coherence conditions is different from zero. We calculate the 95% percentile and the 99% percentile. We can see from the results that even with 99% confidence there is still a difference that is more than zero.

```
z = diffRT.boot$t0/sd(diffRT.boot$t)

pnorm(-z, lower.tail = TRUE)
```

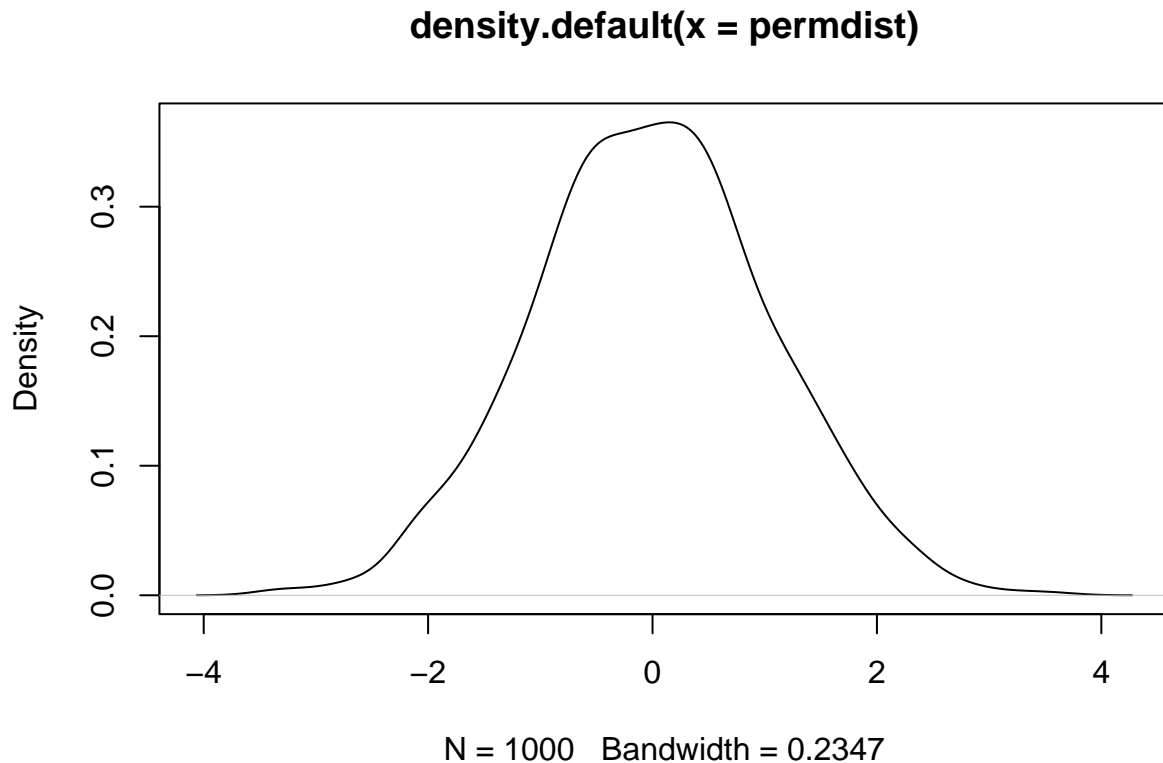
```
## [1] 4.032751e-10
```

- (c) We calculate the z value. This value is the mean divided by the standard error. This lets us see how many standard deviations the mean is from 0. This is the null hypothesis, which says there is no significant difference. If the z value is very large than the difference is really significant.
- (d) From the z value we can calculate the p-value using pnorm. We compare the resulting significant value to the p-value of the t-test done in the previous assignment (1d). The p-value there was 5.575e-06. The significance with the bootstrap was 4.296e-09. Both are highly significant.

3 Permutation / randomization test

```
one.test<-function(x){
  flippedSigns<-sign(rbinom(length(x),1,0.5)-0.1)
  t.test(x*flippedSigns)$statistic
}

permdist <- replicate(1000, one.test(diffRT))
plot(density(permdist))
```



(a) The code above generates a vector with a random distribution of 1 and -1. Then this vector is multiplied with our input vector. Each difference value is made positive or negative. Effectively we assign a random sign to each value. This is the randomization test where we assign a random label to each value and see if it makes a difference. This results in a null distribution. If this distribution would be the same as the bootstrapped distribution, there would be no significant difference.

```
z = 5.9424 - mean(permdist)/sd(permdist)
pnorm(-z, lower.tail = TRUE)
```

```
## [1] 1.374542e-09
```

(b) The empirical statistic from the t-test in the previous assignment was $t=5.94$. As we can see from the approximate null distribution, this value is very far into the right tail of the bell curve, which means really significant. We can calculate a p-value the same way as before. This results in a low value, $p=1.226e-09$.

#Contributions Yann - Questions 1 and 2, Micky - Questions 3 and 4