05 Model comparison and hypothesis testing

Shravan Vasishth

25-29 March 2019

Introduction

Bayes' rule can be written with reference to a specific statistical model M_1. D refers to the data. θ is the parameter, or vector of parameters.

$$P(\theta \mid D, M_1) = \frac{P(D \mid \theta, M_1)P(\theta \mid M_1)}{P(D \mid M_1)} \tag{1}$$

Introduction

$P(D \mid M_1)$ is the likelihood, and is a single number that tells you the likelihood of the observed data D given the model M_1 (and only in the discrete case, it tells you the probability of the observed data D given the model).

Introduction

Obviously, you would prefer a model that gives a higher likelihood. For example, and speaking informally, if you have data that were generated from a Normal(0,1) distribution, then the likelihood of the data given that $\mu = 0$ will be higher than the likelihood given some other value like $\mu = 10$.

Introduction

The higher likelihood is telling us that the underlying model is more likely to have produced the data. So we would prefer the model with the higher likelihood: we would prefer Normal(0,1) over Normal(10,1) as the presumed distribution that generated the data.

Introduction

```
x<-rnorm(100)
sum(dnorm(x,mean=0,sd=1,log=TRUE))
```

```
## [1] -144.2104
```

```
sum(dnorm(x,mean=10,sd=1,log=TRUE))
```

```
## [1] -5210.068
```

Introduction

One way to compare two models M_1 and M_2 is to use the Bayes factor:

$$BF_{12} = \frac{P(D \mid M_1)}{P(D \mid M_2)} \tag{2}$$

In essence, the Bayes factor is a likelihood ratio. (If you have studied frequentist linear modeling, you would have encountered the likelihood ratio test, otherwise known as the ANOVA.)

Introduction

How to compute the likelihood? Consider the simple binomial case where we have a subject answer 10 questions, and they get 9 right. That's our data.

Introduction

Discrete example

Assuming a binomial likelihood function, $Binomial(n, \theta)$, the two models we will compare are

- M_1, the parameter has a point value $\theta = 0.5$ with probability 1 (a very sharp prior), and
- M_2, the parameter has a vague prior $\theta \sim Beta(1, 1)$. Recall that this $Beta(1, 1)$ distribution is $Uniform(0, 1)$.

Introduction

Discrete example

The likelihood under M_1 is:

$$\binom{n}{k}\theta^9(1-\theta)^1 = \binom{10}{9}0.5^{10} \tag{3}$$

We already know how to compute this:
[1] 0.009765625

Introduction

Discrete example

The marginal likelihood under M_2 involves solving the following integral:

$$P(D \mid M_2) = \int P(D \mid \theta, M_2) P(\theta \mid M_2) \, d\theta \qquad (4)$$

The integral is simply integrating out ("summing over") all possible values of the parameter θ.

Introduction

Discrete example

To see what summing over all possible values means, first consider a discrete version of this:

suppose we say that our θ can take on only these three values: $\theta_1 = 0, \theta_2 = 0.5, \theta_3 = 1$, and each has probability $1/3$. Then, the marginal likelihood of the data given this prior specification of θ would be:

$$
\begin{aligned}
P(D \mid M) =& P(\theta_1)P(D \mid \theta_1) + P(\theta_2)P(D \mid \theta_2) + P(\theta_3)P(D \mid \theta_3) \\
=& \sum P(D \mid \theta_i, M)P(\theta_i \mid M)
\end{aligned}
\tag{5}
$$

Introduction

Discrete example

In our discrete example, this evaluates to:

```
res<-(1/3)* (choose(10,9)* (0)^9 * (1-0)^1) + (1/3)*
  (choose(10,9)* (0.5)^9 * (1-0.5)^1) + (1/3)* (choose(10,9)*
res
```

```
## [1] 0.003255208
```

This may be easier to read in mathematical form:

$$
\begin{aligned}
P(D \mid M) =& P(\theta_1)P(D \mid \theta_1) + P(\theta_2)P(D \mid \theta_2) + P(\theta_3)P(D \mid \theta_3) \\
=& \frac{1}{3}\left(\binom{10}{9}0^9(1-0)^1\right) + \frac{1}{3}\left(\binom{10}{9}0.5^9(1-0.5)^1\right) \\
& + \frac{1}{3}\left(\binom{10}{9}1^9(1-1)^1\right) \\
=& 0.003
\end{aligned}
\tag{6}
$$

Introduction

Discrete example

Essentially, we are computing the marginal likelihood $P(D \mid M)$ by averaging the likelihood across possible parameter values (here, only three possible values), with the prior probabilities for each parameter value serving as a weight.

Introduction

Discrete example

The Bayes factor for Model 1 vs Model 2 would then be

```
0.0097/0.003
```

```
## [1] 3.233333
```

Model 1, which assumes that θ has a point value 0.5, is approximately three times more likely than the Model 2 with the discrete prior over θ ($\theta_1 = 0, \theta_2 = 0.5, \theta_3 = 1$, each with probability 1/3).

Introduction

Continuous example

The integral shown above does essentially the calculation we show above, but summing over the entire continuous space that is the range of possible values of θ:

$$P(D \mid M_2) = \int P(D \mid \theta, M_2) P(\theta \mid M_2) \, d\theta \tag{7}$$

Introduction

Continuous example

Let's solve this integral analytically. We need to know only one small detail from integral calculus:

$$\int_a^b x^9 \, dx = [\frac{x^{10}}{10}]_a^b \tag{8}$$

Similarly:

$$\int_a^b x^{10} \, dx = [\frac{x^{11}}{11}]_a^b \tag{9}$$

Having reminded ourselves of how to solve this simple integral, we proceed as follows.

Introduction

Continuous example

Our prior for θ is $Beta(\alpha = 1, \beta = 1)$:

$$
\begin{aligned}
P(\theta \mid M_2) =& \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} \theta^{\beta-1} \\
=& \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} \theta^{1-1} \theta^{1-1} \\
=& 1
\end{aligned}
\tag{10}
$$

Introduction

Continuous example

So, our integral simplifies to:

$$
\begin{aligned}
P(D \mid M_2) &= \int_0^1 P(D \mid \theta, M_2) \, d\theta \\
&= \int_0^1 \binom{10}{9} \theta^9 (1 - \theta)^1 \, d\theta \\
&= \int_0^1 \binom{10}{9} (\theta^9 - \theta^{10}) \, d\theta \\
&= 10 \left[\frac{\theta^{10}}{10} - \frac{\theta^{11}}{11} \right]_0^1 \\
&= 10 \times \frac{1}{110} = \frac{1}{11}
\end{aligned} \tag{11}
$$

Introduction

Continuous example

So, when Model 1 assumes that the θ parameter is 0.5, and Model 2 has a vague prior $Beta(1,1)$ on the θ parameter, our Bayes factor will be:

$$BF_{12} = \frac{P(D \mid M_1)}{P(D \mid M_2)} = \frac{0.00977}{1/11} = 0.107 \tag{12}$$

Introduction

Continuous example

Thus, the model with the vague prior (M2) is about 9 times more likely than the model with $\theta = 0.5$:

$$\frac{1}{0.10742} = 9.309 \tag{13}$$

Introduction

Continuous example

We could conclude that we have some evidence against the guessing model M1 in this case. Jeffreys (1998) has suggested the following decision criterion using Bayes factors. Here, we are comparing two models, labeled 1 and 2.

- $BF_{12} > 100$: Decisive evidence
- $BF_{12} = 32 - 100$: Very strong
- $BF_{12} = 10 - 32$: Strong
- $BF_{12} = 3 - 10$: Substantial
- $BF_{12} = 2 - 3$: Not worth more than a bare mention

Introduction

Prior sensitivity

The Bayes factor is sensitive to the choice of prior. It is therefore important to do a sensitivity analysis with different priors.

Introduction

Prior sensitivity

For the model M_2 above, consider the case where we have a prior on θ such that there are 10 possible values for θ, 0.1, 0.2, 0.3,...,1, and the probabilities of each value of θ are $1/10$.

```
theta<-seq(0.1,1,by=0.1)
w<-rep(1/10,10)

prob<-rep(NA,length(w))
for(i in 1:length(theta)){
prob[i]<-(1/w[i])*choose(10,9)*theta[i]^9*(1-theta[i]^1)
}
## Likelihood for model M2 with
## new prior on theta:
sum(prob)
```

```
## [1] 8.287079
```

Introduction

Prior sensitivity

Now the Bayes factor for M1 compared to M2 is:

```
0.0097/sum(prob)
```

[1] 0.001170497

Now, model M2 is decisively more likely compared to model M1:

```
1/(0.0097/sum(prob))
```

[1] 854.3381

This toy example illustrates the effect of prior specification on the Bayes factor. It is therefore very important to display the Bayes factor under both uninformative and informative priors for the parameter that we are interested in.

One should never use a single 'default' prior and report a single Bayes factor.

Introduction

The Bayes factor is the ratio of posterior to prior odds

The Bayes factor is really the ratio of posterior odds vs prior odds for any given pair of models:

$BF = \frac{\text{posterior odds}}{\text{prior odds}}$

In the context of our problem:

$$\underbrace{\frac{P(M_1 \mid D)}{P(M_2 \mid D)}}_{\substack{\uparrow \\ posterior\ odds}} = \underbrace{\frac{P(D \mid M_1)}{P(D \mid M_2)}}_{\substack{\uparrow \\ BF_{12}}} \underbrace{\frac{P(M_1)}{P(M_2)}}_{\substack{\uparrow \\ prior\ odds}} \tag{14}$$

Introduction

The Bayes factor is the ratio of posterior to prior odds

So, when the prior odds for M_1 vs M_2 are 1 (i.e., when both models are a priori equi-probable), then we are just interested in computing the posterior odds for the two models.

The Savage-Dickey method

This method consists of computing the Bayes factor by dividing the height of the posterior for the parameter of interest, θ, by the height of the prior for θ at the specific point corresponding to some null hypothesis value $\theta = \theta_0$. Because we call the baseline model the null model, we label it M_0.

The Savage-Dickey method

The Savage-Dickey method is based on a theorem whose proof appears in several published works (Verdinelli and Wasserman 1995).

Savage-Dickey Density ratio

Suppose that M_1 is a model with parameters $\theta = (\phi, \omega)$, and M_0 is a model that is a restricted version of M_1 with $\omega = \omega_0$ and free parameter ϕ. Suppose that the priors in the two models satisfy

$$f(\phi \mid M_0) = f(\phi \mid \omega = \omega_0, M_1) \tag{15}$$

[The above holds if ϕ and ω are independent under M_1, that is, if $f(\phi, \omega \mid M_1) = f(\phi \mid M_1)f(\omega \mid M_1)$.]

Savage-Dickey Density ratio

Then, Bayes factor of M_0 can be written as

$$BF_{01} = \frac{P(D|H_0)}{P(D|H_1)} = \frac{f(\omega = \omega_0 \mid D, M_1)}{f(\omega = \omega_0 \mid M_1)} \tag{16}$$

Savage-Dickey Density ratio

Computing Bayes Factors using the Savage-Dickey method

- This example is taken from Lee and Wagenmakers (2013). Suppose we have within-subjects data for two conditions.
- The data represent increase in recall performance in a memory task from the same subject, once in winter and once in summer.
- Suppose one theory says that increase in recall performance is higher in summer, but an alternative theory claims that there is no difference between the two seasons.
- We will test the null vs alternative hypotheses using Bayes factors.

Savage-Dickey Density ratio

Computing Bayes Factors using the Savage-Dickey method

```
# Read data:
Winter <- c(-0.05,0.41,0.17,-0.13,0.00,-0.05,0.00,0.17,0.29,0.
            0.17,0.08,-0.04,-0.04,0.04,-0.13,-0.12,0.04,0.21,0
            0.33,0.04,0.04,0.04,0.00,0.21,0.13,0.25,-0.05,0.29
            0.04,0.25,0.12)

Summer <- c(0.00,0.38,-0.12,0.12,0.25,0.12,0.13,0.37,0.00,0.50
            -0.37,-0.25,-0.12,0.50,0.25,0.13,0.25,0.25,0.38,0.
            0.00,0.00,0.25,0.13,-0.25,-0.38,-0.13,-0.25,0.00,0
            0.00,0.50,0.00)
```

Savage-Dickey Density ratio

Computing Bayes Factors using the Savage-Dickey method

Let's say we want to compare the evidence for two hypotheses: the difference between the two conditions (Winter and Summer) is $H_0 : \delta = 0$ and $H_0 : \delta \neq 0$.

Savage-Dickey Density ratio

Computing Bayes Factors using the Savage-Dickey method

Normally, we would do a paired t-test. We get a non-significant result:

```
##
##  Paired t-test
##
## data:  Winter and Summer
## t = 0.78608, df = 40, p-value = 0.4365
## alternative hypothesis: true difference in means is not equ
## 95 percent confidence interval:
##  -0.05364719  0.12193987
## sample estimates:
## mean of the differences
##              0.03414634
```

Savage-Dickey Density ratio

Computing Bayes Factors using the Savage-Dickey method

Equivalently, one can do a one sample test after taking the pairwise differences in scores:

```
##
##  One Sample t-test
##
## data:  d
## t = 0.78608, df = 40, p-value = 0.4365
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.05364719 0.12193987
## sample estimates:
##  mean of x
## 0.03414634
```

Savage-Dickey Density ratio

Computing Bayes Factors using the Savage-Dickey method

We will now compute the Bayes factor, using the Savage-Dickey method. This will allow us to test the null against the alternative hypothesis.

Savage-Dickey Density ratio

Example 1

Prepare data:

```
d <- d / sd(d)        # standardize the paired difference of s
ndata <- length(d)  # number of subjects

data <- list(x=d, ndata=ndata)  # to be passed on to Stan
```

Savage-Dickey Density ratio

Example 1

We will now compute, using Stan, the Bayes Factor for the two hypotheses $H_0 : \delta = 0$ and $H_1 : \delta \neq 0$.

The model is:

- $\delta \sim Cauchy(0, 1)$
- $\sigma \sim Cauchy(0, 1)_{I(0, \infty)}$
- $\mu \leftarrow \delta\sigma$
- $x_i \sim Normal(\mu, \sigma^2)$

Savage-Dickey Density ratio

Example 1

```
model_example1 <- "
data {
  int<lower=0> ndata;
  vector[ndata] x;
}
parameters {
  real sigmatmp;
  real delta;
}
transformed parameters {
  real mu;
  real<lower=0> sigma;
  sigma = fabs(sigmatmp);
  mu = delta * sigma;
}
```

Savage-Dickey Density ratio

Example 1

```r
library(rstan)
# Parameters to be monitored
parameters <- c("delta")

samples <- stan(model_code=model_example1,
                data=data,
                iter=20000,
                chains=4)

# Collect posterior samples across all chains:
delta.posterior <- extract(samples,pars=parameters)$delta
```
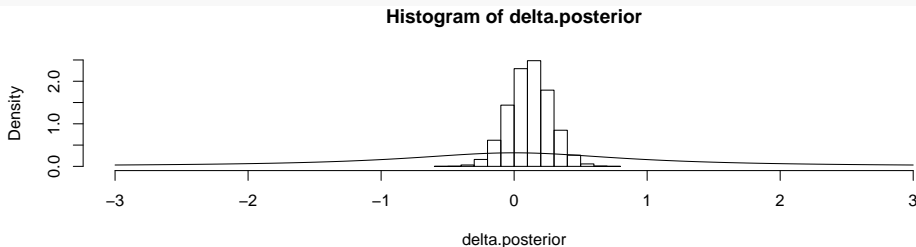
Savage-Dickey Density ratio

Example 1

```
hist(delta.posterior,freq=FALSE,xlim=c(-3,3))
x<-seq(-3,3,by=0.01)
lines(x,dcauchy(x))
```

Histogram of delta.posterior

Savage-Dickey Density ratio

Example 1

```r
#BFs based on logspline fit
library(polspline)
fit.posterior <- logspline(delta.posterior)

# 95% confidence interval:
x0 <- qlogspline(0.025,fit.posterior)
x1 <- qlogspline(0.975,fit.posterior)

posterior <- dlogspline(0, fit.posterior) # this gives the pd
prior      <- dcauchy(0)                  # height of order-restricted
(BF01      <- posterior/prior)
```

```
## [1] 6.11385
```

Savage-Dickey Density ratio

Example 1

The odds of H_0 being true compared to H_1 are $6:1$.
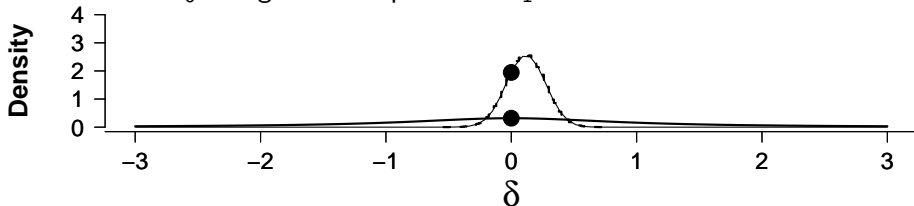
Figure 1: Shown are the prior and posterior densities on delta. The null hypothesis was that delta is 0, and we see that delta=0 has a value 6 times larger under the posterior compared to the prior. This means that the evidence for the null hypothesis that delta=0 is 6 times more than the alternative.

Two methods for computing Bayes factors with brms

First, set up data as a data-frame:

```r
y<-c(Winter,Summer)
#length(Winter)
n<-length(Summer)

cond<-factor(c(rep("winter",n),
        rep("summer",n)))
subject<-rep(rep(1:n),2)
dat<-data.frame(y,cond,subject)
```

Two methods for computing Bayes factors with brms

```
##       y   cond subject
## 1 -0.05 winter       1
## 2  0.41 winter       2
## 3  0.17 winter       3
## 4 -0.13 winter       4
## 5  0.00 winter       5
## 6 -0.05 winter       6
```

Two methods for computing Bayes factors with brms

Set priors:

```
priors0 <- c(set_prior("cauchy(0, 1)", class = "Intercept"),
                    set_prior("cauchy(0, 1)",
                              class = "sd"),
                    set_prior("cauchy(0, 1)",
                              class = "sigma"))

priors <- c(set_prior("cauchy(0, 1)", class = "Intercept"),
                    set_prior("cauchy(0, 1)",
                              class = "b"),
                    set_prior("cauchy(0, 1)",
                              class = "sd"),
                    set_prior("cauchy(0, 1)",
                              class = "sigma"))
```

Two methods for computing Bayes factors with brms

Using Savage-Dickey method (the hypothesis function in brms):

```
m_full <- brm(y ~ cond + (1|subject),
              data = dat,
              prior = priors,
              sample_prior = TRUE,
              iter = 10000,
         control=list(adapt_delta=0.99))
```

SAMPLING FOR MODEL '5c22c60911e928624280d59a52e89f48' NOW (CHAIN 1). Chain 1: Chain 1: Gradient evaluation took 5.9e-05 seconds Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 0.59 seconds. Chain 1: Adjust your expectations accordingly! Chain 1: Chain 1: Chain 1: Iteration: 1 / 10000 [0%] (Warmup) Chain 1: Iteration: 1000 / 10000 [10%] (Warmup) Chain 1: Iteration: 2000 / 10000 [20%] (Warmup) Chain 1: Iteration: 3000 / 10000 [30%] (Warmup) Chain 1: Iteration: 4000 / 10000 [40%] (Warmup) Chain 1: Iteration: 5000 / 10000

Two methods for computing Bayes factors with brms

Using Savage-Dickey method (the hypothesis function in brms):

```
# H0: No effect of cond
BF_brms_m <- brms::hypothesis(m_full,
                        "condwinter = 0")
## Evidence for NULL model vs FULL model:
BF_brms_m$hypothesis$Evid.Ratio
```

```
## [1] 7.405533
```

Two methods for computing Bayes factors with brms

```
normalpriors <- c(set_prior("normal(0, 1)", class = "Intercept
                  set_prior("normal(0, 1)",
                            class = "b"),
                  set_prior("normal(0, 1)",
                            class = "sd"),
                  set_prior("normal(0, 1)",
                            class = "sigma"))

m_full <- brm(y ~ cond + (1|subject),
              data = dat,
              prior = normalpriors,
              sample_prior = TRUE,
              iter = 10000,
        control=list(adapt_delta=0.99))

#summary(m_full)
```

Two methods for computing Bayes factors with brms

```
# H0: No effect of cond
BF_brms_m <- brms::hypothesis(m_full,
                        "condwinter = 0")
## Evidence for NULL model vs FULL model:
BF_brms_m$hypothesis$Evid.Ratio

## [1] 7.405533
```

Two methods for computing Bayes factors with brms

Using the bayes_factor function in brms:

Using the bayes_factor function in brms:

```
bayes_factor(m0,m1)$bf
```

Iteration: 1 Iteration: 2 Iteration: 3 Iteration: 4 Iteration: 5 Iteration: 6
Iteration: 1 Iteration: 2 Iteration: 3 Iteration: 4 Iteration: 5 [1] 21.32731

Notice that if you flip the order of the models in the function, the evidence is for the first model:

```
bayes_factor(m1,m0)$bf
```

Iteration: 1 Iteration: 2 Iteration: 3 Iteration: 4 Iteration: 5 Iteration: 1
Iteration: 2 Iteration: 3 Iteration: 4 Iteration: 5 Iteration: 6 [1] 0.04672283

Bayes factor using Stan

Example 2

We will now compute, using Stan, the Bayes Factor for the two hypotheses $H_0 : \delta = 0$ and $H_1 : \delta \sim Cauchy(0,1)_{I(-\infty,0)}$.

The Bayesian model is:

- $\delta \sim Cauchy(0,1)_{I(-\infty,0)}$
- $\sigma \sim Cauchy(0,1)_{I(0,\infty)}$
- $\mu \leftarrow \delta\sigma$
- $x_i \sim Normal(\mu, \sigma^2)$

Bayes factor using Stan

Example 2

Bayes factor using Stan

Example 2

```
## samples from model:
samples <- stan(model_code=model_example2,
                data=data,
                #init=myinits,
                pars=parameters,
                iter=30000,
                chains=4,
                control = list(adapt_delta = 0.99,max_treedept

# Collect posterior samples across all chains:
delta.posterior <- extract(samples)$delta
```

Bayes factor using Stan

Example 2

**Posterior distribution
and prior (line)**

Bayes factor using Stan

Example 2

Now we compute the Bayes Factor, comparing the two hypotheses.

```
fit.posterior <- logspline(delta.posterior)

fit.posterior <- logspline(delta.posterior,ubound=0) # NB. no

# 95% confidence interval:
x0 <- qlogspline(0.025,fit.posterior)
x1 <- qlogspline(0.975,fit.posterior)

posterior <- dlogspline(0, fit.posterior) # this gives the pd
prior     <- 2*dcauchy(0)        # height of order--restricted
(BF01     <- posterior/prior)
## [1] 14.22946
```

Bayes factor using Stan

Example 2

According to this analysis, the null hypothesis $H_0 : \delta = 0$ being true is 14 times more likely than $H_1 : \delta \sim Cauchy(0,1)_{I(-\infty,0)}$.

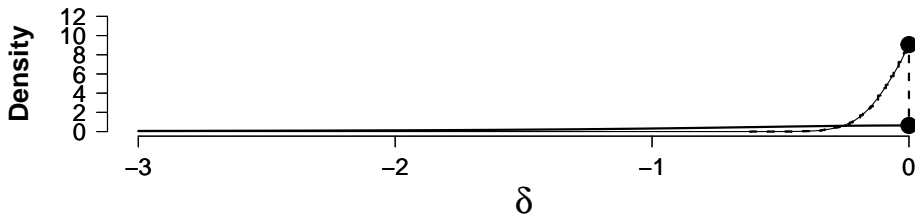**Figure 2:** The prior and posterior densities.

Class Exercise 1

Refit examples 1 and 2 with a different prior for σ. Does the Bayes Factor change in either case?

What happens when the prior for δ is changed to a suitable normal distribution?

Class Exercise 2

Estimate the Bayes factor for the hypotheses: $H_0 : \delta = 0$, and
$H_1 : \delta \sim Cauchy(0,1)_{I(0,\infty)}$.

Example 3

Here is an example of a two-sample test.

The data:

```
x <- c(70,80,79,83,77,75,84,78,75,75,78,82,74,81,72,70,75,72,7
y <- c(56,80,63,62,67,71,68,76,79,67,76,74,67,70,62,65,72,72,6

n1 <- length(x)
n2 <- length(y)

# Rescale
y <- y - mean(x)
y <- y / sd(x)
x <- (x - mean(x)) / sd(x);

data <- list(x=x, y=y,
             n1=n1, n2=n2)
```

Example 3

We could do an unpaired two-sample t-test:

```
##
##  Welch Two Sample t-test
##
## data:  x and y
## t = 4.4708, df = 33.139, p-value = 8.625e-05
## alternative hypothesis: true difference in means is not equ
## 95 percent confidence interval:
##  0.980942 2.618788
## sample estimates:
##      mean of x      mean of y
## -1.405838e-15 -1.799865e+00
```

We see strong evidence against the null hypothesis.

Example 3

The Stan model:

```
model_example3 <- "
// Two-sample Comparison of Means
data {
  int<lower=1> n1;
  int<lower=1> n2;
  vector[n1] x;
  vector[n2] y;
}
parameters {
  real mu;
  real sigmatmp;
  real delta;
}
transformed parameters {
  real<lower=0> sigma;
```
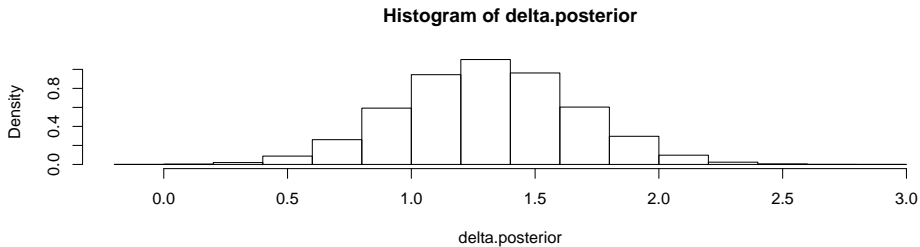
Example 3

```r
# Parameters to be monitored
parameters <- c("delta")

# The following command calls Stan with specific options.
# For a detailed description type "?rstan".
samples <- stan(model_code=model_example3,
                data=data,
                iter=20000,
                chains=4)

# Collect posterior samples across all chains:
delta.posterior <- extract(samples,pars=parameters)$delta
```

Example 3

Histogram of delta.posterior

Example 3

```
#=========== BFs based on logspline fit =================
library(polspline) # this package can be installed from within
fit.posterior <- logspline(delta.posterior)

# 95% confidence interval:
x0 <- qlogspline(0.025,fit.posterior)
x1 <- qlogspline(0.975,fit.posterior)

posterior <- dlogspline(0, fit.posterior) # this gives the pd
prior     <- dcauchy(0)           # height of  prior at delta =
(BF01 <- posterior/prior)

## [1] 0.003425931
```
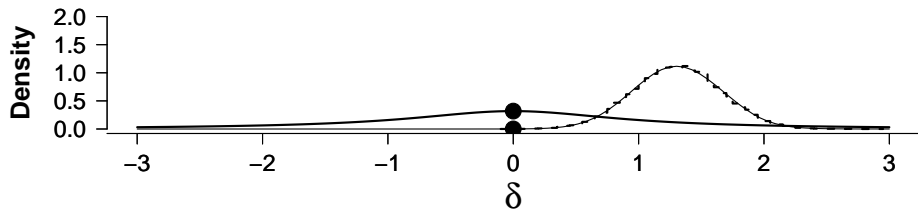
Example 3

The odds in favor of the alternative are 292 : 1. (The Lee and Wagenmakers (2013) book has a higher number, in the 400s.)

Example 3

References

Jeffreys, Harold. 1998. *The Theory of Probability*. Oxford University Press.

Lee, Michael D, and Eric-Jan Wagenmakers. 2013. *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press.

Verdinelli, Isabella, and Larry Wasserman. 1995. "Computing Bayes factors using a generalization of the Savage-Dickey density ratio." *Journal of the American Statistical Association* 90 (430). Taylor & Francis: 614–18.