# Worksheet 2; Comparing two data curves using the function *Bayescompare*

## 1 Comparing two data curves with many points

In this example we begin by importing two sets of experimental data with many points.

Load the *babar* library and import and plot the data *LmH_411.csv* and *M126_50.csv* by running the lines below.

```
library(babar)

LmH_411.file <- system.file("extdata", "LmH_411.csv", package = "babar")
LmH_411.data <- read.csv(LmH_411.file, header=TRUE, sep =",",
                  na.strings=c("ND","NA"))
M126_50.file <- system.file("extdata", "M126_50.csv", package = "babar")
M126_50.data <- read.csv(M126_50.file, header=TRUE, sep =",",
                  na.strings=c("ND","NA"))
plot(LmH_411.data, ylim=c(0,10))
points(M126_50.data, pch=0)
```

The three hypotheses that we can use for testing the differences between two data curves are summarised in table 1.

| Hypothesis name | Hypothesis |
|---|---|
| H1 | "data curves are replicates" |
| H2 | "data curves have same growth rate" |
| H3 | "all data curve parameters are different" |

Table 1: The three hypotheses available for testing in the function.

We can use Bayesian analysis to fit curves to the combined data for each of the three different hypotheses using the *Bayescompare* function. We begin by using the 4 parameter Baranyi model and inferring the noise level.

Perform the analysis for hypothesis 1. This may take a couple of minutes or so to complete. Some typical output is shown below.

```
set.seed(11)  ## for reproducibility
results_H1 <- Bayescompare(LmH_411.data, M126_50.data, hyp = "H1", model = "Bar4par")

## # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
## Model =  Bar4par
```

```
## mu_max1 prior type = Uniform , mu_max2 prior type = Uniform
## Hypothesis =  H1 ( data curves are replicates )
## # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
##
## log evidence =  -41.1382
## Means and standard deviations:
## Log cell counts at time 0, y_01 = y_02 1.595486 +/- 0.328752
## Log final cell counts, y_max1 = y_max2 = 4.081193 +/- 0.207032
## Growth rates, mu_max1 = mu_max2 = 0.04265824 +/- 0.01524108
## h_01 = h_02 = 2.665055 +/- 1.839037
## Noise level, sigma1 = 0.2962434 +/- 0.2556993
## Noise level, sigma2 = 3.704077 +/- 0.4342604
```

Plot the fitted curves alongside the data.

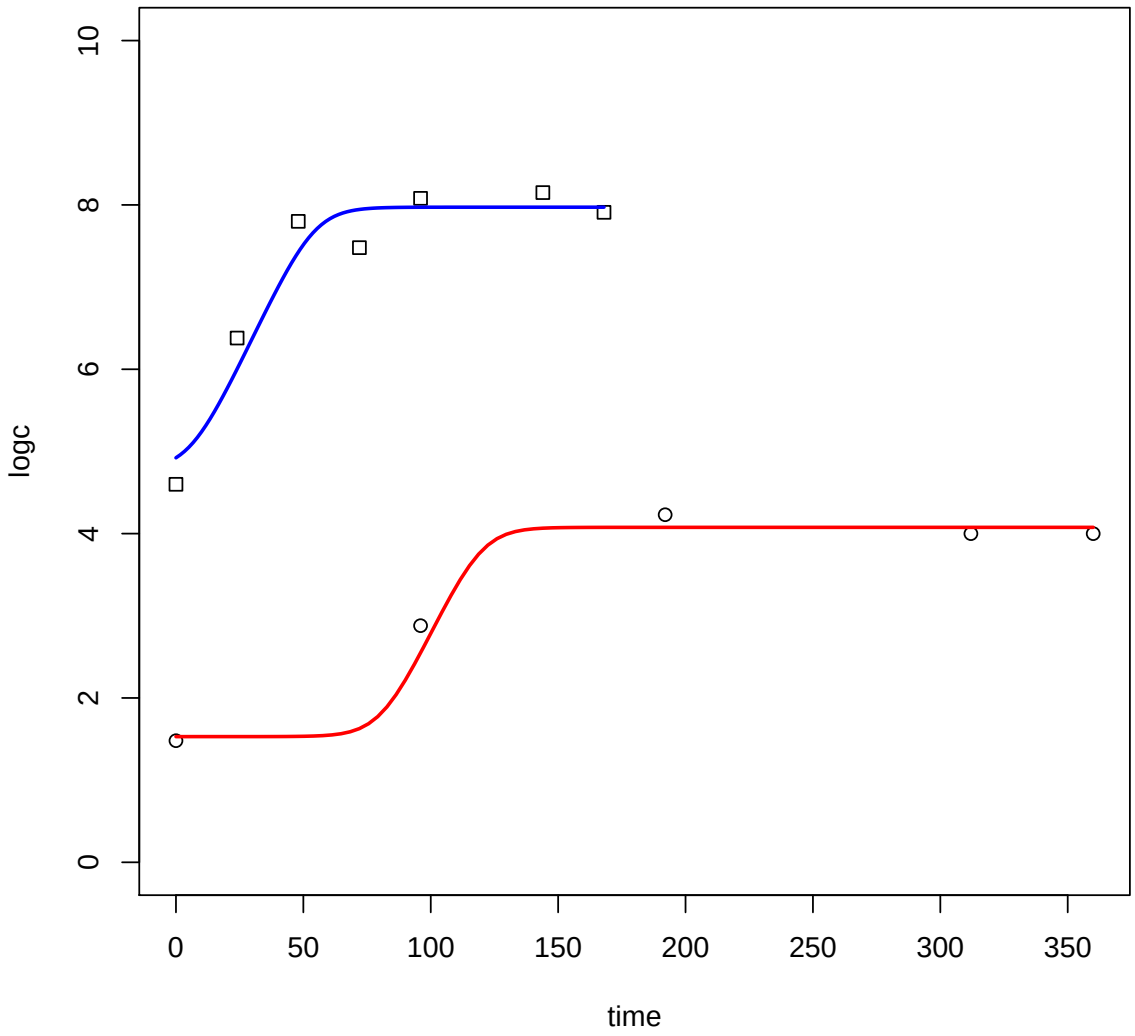
```
t1 <- results_H1$fit.t1; y1 <- results_H1$fit.y1mean;
t2 <- results_H1$fit.t2; y2 <- results_H1$fit.y2mean
lines(t1,y1,col="red",lwd=2)
lines(t2,y2,col="blue",lwd=2)
```

Do the same for hypotheses 2 and 3 by running the relevant lines of the code.

```
set.seed(11)  ## for reproducibility
results_H2 <- Bayescompare(LmH_411.data, M126_50.data, hyp = "H2", model = "Bar4par")
set.seed(11)  ## for reproducibility
results_H3 <- Bayescompare(LmH_411.data, M126_50.data, hyp = "H3", model = "Bar4par")
```

Now that we have the results, we can compare the three hypotheses, but first let's just check what the curves look like from hypothesis 2 — "data curves have same growth rate".

```
t1 <- results_H2$fit.t1; y1 <- results_H2$fit.y1mean
t2 <- results_H2$fit.t2; y2 <- results_H2$fit.y2mean
plot(LmH_411.data, ylim=c(0,10))
points(M126_50.data, pch=0)
lines(t1,y1,col="red",lwd=2)
lines(t2,y2,col="blue",lwd=2)
```




Extract the log evidences for the three hypotheses by running the relevant lines, for example as shown below.

```
logevidence_H1 <- results_H1$logevidence
logevidence_H2 <- results_H2$logevidence
logevidence_H3 <- results_H3$logevidence
```

Calculate and print the results for the Bayes' factor for hypothesis 1 versus hypothesis 2. The results can again be interpreted using table 2 on worksheet 1.

```
Bayes <- exp(logevidence_H1)/exp(logevidence_H2)

## Bayes' factor for first vs. second hypothesis =  2.690874e-06
## Log Bayes' factor =  -12.82564
```

Try changing the hypotheses that we are comparing (by changing the hypothesis names) to see how the Bayes' factor changes. Which is the preferred hypothesis? To what degree is it preferred over the second most likely hypothesis?