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

Lydia Rickett, Contact: Lydia.Rickett@sainsbury-laboratory.ac.uk

- In R studio open the file *example2.R* in the folder "Program".

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 code *Bayesdouble.R* and import and plot the data *LmH_411.csv* and *M126_50.csv* by running the relevant lines of *example2.R*.

```
source("Bayesdouble.R")
data1 <-read.csv("data/LmH_411.csv",header=TRUE,sep =",",
                 na.strings=c("ND","NA"))
data2 <-read.csv("data/M126_50.csv",header=TRUE,sep =",",
                 na.strings=c("ND","NA"))
plot(data1,ylim=c(0,10))
points(data2,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 will take a few minutes to complete. Some typical output is shown below.

```
results_H1 <- Bayescompare(data1, data2, hyp = "H1", model = "Bar4par")
```

```
## # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
## Model =  Bar4par
## mu_max1 prior type = Uniform , mu_max2 prior type = Uniform
## Hypothesis =  H1 ( data curves are replicates )
## # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
##
## log evidence =  -40.65
## Means and standard deviations:
## Log cell counts at time 0, y_01 = y_02 1.57 +/- 0.2936
## Log final cell counts, y_max1 = y_max2 = 4.081 +/- 0.1827
## Growth rates, mu_max1 = mu_max2 = 0.04435 +/- 0.01511
## h_01 = h_02 = 2.845 +/- 1.804
## Noise level, sigma1 = 0.2568 +/- 0.23
## Noise level, sigma2 = 3.739 +/- 0.4173
```

- 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="red",lwd=2)
```

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

Now that we have the results, we can compare the three hypotheses.

- Extract the log evidences for the three hypotheses by running the relevant lines in the section *"Compare hypotheses using Bayes' factor"*. An example is shown below.

```
logevidence_H1 <- results_H1$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)
cat("Bayes' factor for first vs. second hypothesis = ", Bayes, '\n')

## Bayes' factor for first vs. second hypothesis =  8.169e-06

cat("Log Bayes' factor = ", log(Bayes), '\n')

## Log Bayes' factor =  -11.72
```

- 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?

1.1 Extension: plotting posterior samples

The last section of *example2.R* again plots the curves fitted by the posterior parameter samples alongside the combined data and the two curves fitted using the mean values of the parameter samples.

- Run the lines in the section *"Plot the curves fitted by posterior samples"* to plot the posterior curves for hypothesis 3.
- Try changing the hypothesis for the results that we plot to hypothesis 1 and 2 in turn to see how the plot changes in each case.

2 Comparing two data curves with two points

Next we will consider experimental data with just two data points.

- Uncomment and run the lines to import and plot the data *Tas1387.csv* and *B174_Ah.csv*.

Since we only have two data points we now use the linear model and let the noise level be prescribed at 0.3.

- By uncommenting and running the relevant lines perform the analysis for each hypothesis using the linear model and again plot the fitted curves in each case. An example is shown below.

```
results_H1 <- Bayescompare(data1,data2,hyp="H1",model="linear",
                           inf.sigma1=FALSE,inf.sigma2=FALSE)
```

- Again extract the log evidences for each hypothesis and test using Bayes' factor to see which is the most likely, and by how much. You could also take a look at the posterior curve plots for this example (the y axis scale will again need to be changed for these).