

Statistique bayésienne avec R

Julien JACQUES

Préliminaires

Commençons par charger les libraires dont nous aurons besoin

```
library(bayess,quietly = T)
library(rjags,quietly = T)
library(BayesFactor,quietly = T)
```

Exemple 3 - données sleep

Les données sleep sont contenues dans R.

```
t.test(x = sleep$extra[sleep$group==1]-sleep$extra[sleep$group==2])

##
## One Sample t-test
##
## data:  sleep$extra[sleep$group == 1] - sleep$extra[sleep$group == 2]
## t = -4.0621, df = 9, p-value = 0.002833
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -2.4598858 -0.7001142
## sample estimates:
## mean of x
## -1.58
```

Il y a une différence significative entre les deux groupes. A noter qu'il aurait été préférable de faire un test de Wilcoxon ici, qui donne une p-value semblable.

```
wilcox.test(x = sleep$extra[sleep$group==1]-sleep$extra[sleep$group==2])

## Warning in wilcox.test.default(x = sleep$extra[sleep$group == 1] -
## sleep$extra[sleep$group == : cannot compute exact p-value with ties
## Warning in wilcox.test.default(x = sleep$extra[sleep$group == 1] -
## sleep$extra[sleep$group == : cannot compute exact p-value with zeroes
##
## Wilcoxon signed rank test with continuity correction
##
## data:  sleep$extra[sleep$group == 1] - sleep$extra[sleep$group == 2]
## V = 0, p-value = 0.009091
## alternative hypothesis: true location is not equal to 0
```

Faisons de même avec une approche bayésienne de choix de modèle à l'aide des facteurs de Bayes.

```
ttestBF(x = sleep$extra[sleep$group==1]-sleep$extra[sleep$group==2], rscale=sqrt(2)/2)
```

```
## Bayes factor analysis
## -----
## [1] Alt., r=0.707 : 17.25888 ±0%
##
## Against denominator:
##   Null, mu = 0
## ---
## Bayes factor type: BFoneSample, JZS
```

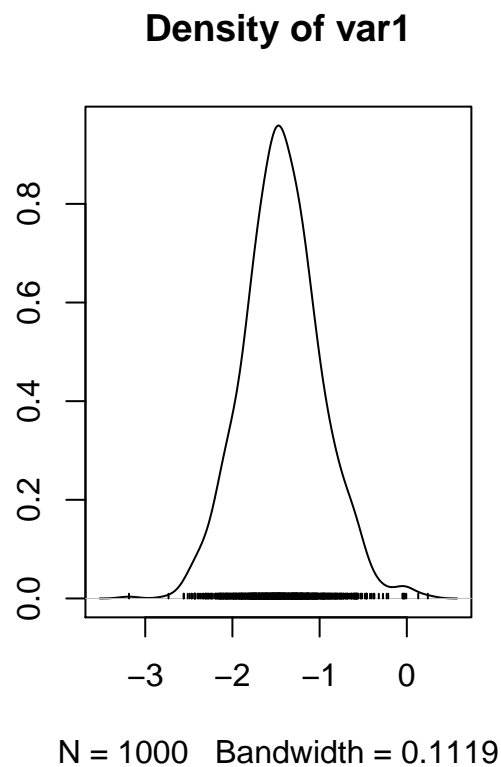
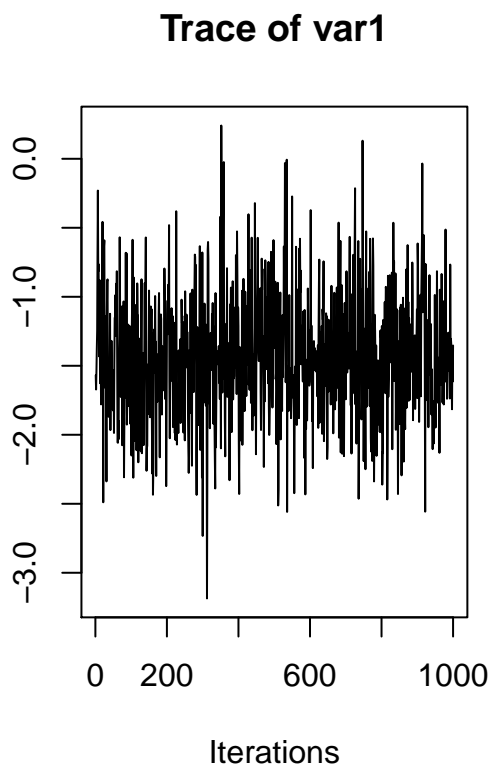
En calculant le log10 de BayesFactor, on conclut à une forte certitude que le modèle M1 (supposant l'absence de différence) est faux. Le résultat est le même qu'avec l'approche inférentielle classique.

```
log10(17.25888)
```

```
## [1] 1.237013
```

On peut aussi simuler l'espérance de la différence suivant la loi a posteriori

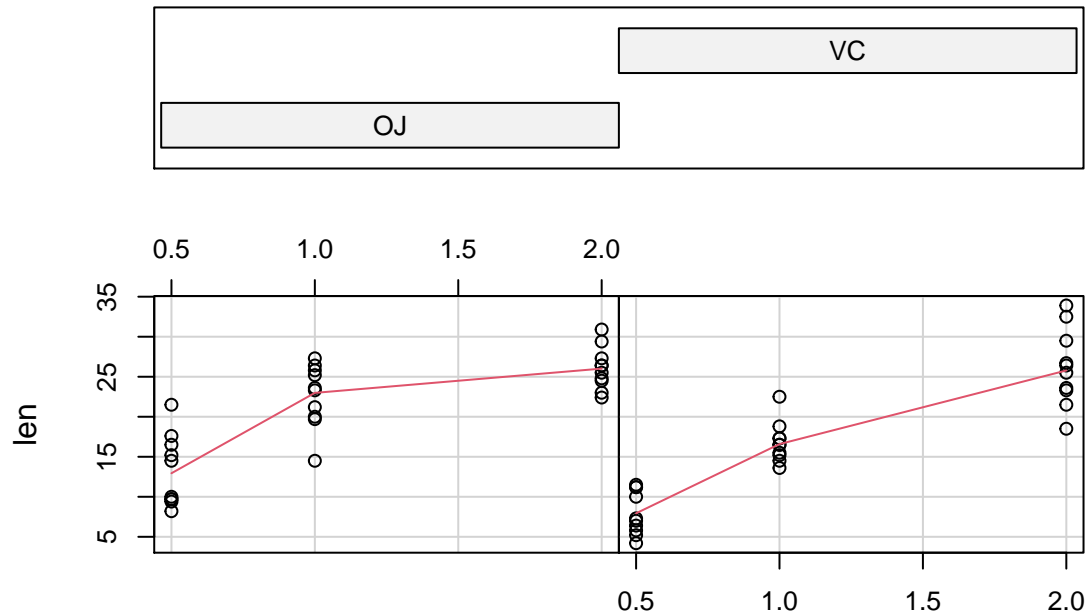
```
samples = ttestBF(x = sleep$extra[sleep$group==1]-sleep$extra[sleep$group==2], rscale=sqrt(2)/2, posterior = TRUE)
plot(samples[, "mu"])
```



Exercice 2 - données ToothGrowth

```
data(ToothGrowth)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```

Given : supp



ToothGrowth data: length vs dose, given type of supplement

```
ToothGrowth$dose = factor(ToothGrowth$dose)
levels(ToothGrowth$dose) = c("Low", "Medium", "High")

summary(aov(len ~ supp*dose, data=ToothGrowth))
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## supp       1  205.4    205.4   15.572 0.000231 ***
## dose       2 2426.4   1213.2   92.000 < 2e-16 ***
## supp:dose   2  108.3     54.2    4.107 0.021860 *
## Residuals  54  712.1     13.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

On constate un fort effet de la dose, un effet du type de vitamine C (jus d'orange ou acide ascorbic), peut-être quelque chose pour l'interaction... Avec une approche bayésienne :

```
bf = anovaBF(len ~ supp*dose, data=ToothGrowth, progress = F)
bf
```

```
## Bayes factor analysis
## -----
## [1] supp           : 1.198757      ±0.01%
## [2] dose          : 4.983636e+12 ±0%
## [3] supp + dose   : 2.945838e+14 ±1.58%
## [4] supp + dose + supp:dose : 7.845082e+14 ±2.95%
##
## Against denominator:
##   Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS
```

comme pour l'ANOVA, le modele complet est préféré.