

Travaux Pratiques de Régression Bayésienne

2026-01-22

Contents

| | | |
|----------|--|----------|
| 1 | Travaux Pratique 1 | 2 |
| 1.1 | Simulation des données | 2 |
| 1.2 | Mise en place des fonctions utiles | 2 |
| 1.3 | Utilisation du package <code>rrBLUP</code> | 3 |
| 1.4 | La régression bayésienne A | 4 |
| 1.5 | La régression bayésienne LASSO | 4 |
| 1.6 | La méthode de sélection bayésienne SSVS | 4 |

1 Travaux Pratique 1

L'objectif de ce TP est de comparer le modèle de régression aléatoire, à la régression bayésienne, la régression bayésienne LASSO et la méthode de sélection bayésienne *SSVS*. Le modèle linéaire étudié est le suivant :

$$Y = \mu \mathbb{I} + X\beta + \varepsilon$$

1.1 Simulation des données

Afin de comparer ces différentes méthodes, des données simulées seront utilisées. Pour cela, on génère un jeu de données de 200 observation qui sont réparties en données d'apprentissage pour la construction des modèles et des données de test qui serviront à comparer les modèles en terme de performance. L'échantillon sera coupé en deux : 100 observations pour l'apprentissage et 100 observations pour le test. On générera 300 variables indépendantes et identiquement distribuées (*i.i.d.*), chacune suivant une loi uniforme entre -5 et 5 .

```
simuVarExpl <- matrix(rep(0, 300*200), ncol=300, nrow=200)
for (i in 1 :300){
  simuVarExpl[,i] <- runif(200, -5, 5)
  simuVarExpl <- as.data.frame(simuVarExpl)
}
```

Dans le lot des variables indépendantes générées, nous n'utiliserons que cinq (05) pour la simulation de la variable d'intérêt. Il s'agira de celles d'indices 10, 20, 30, 40 et 50. Nous utiliserons le vecteur de paramètres $\beta = (+1, -1, +2, -2, +3)$ et nous supposons que $\mu = 0$.

```
trueInd <- c(10, 20, 30, 40, 50)
beta <- c(1, -1, 2, -2, 3)
ySimu <- as.matrix(simuVarExpl)[,trueInd]%% beta + rnorm(200,0,2)
```

1.2 Mise en place des fonctions utiles

Deux principales fonctions seront utiles pour la réalisation de ce TP. Il s'agit des fonctions de prédiction de la variable d'intérêt et de sélection de variables explicatives.

1.2.1 Prédictions

En se basant sur des variables explicatives et des estimateurs de μ et β , la fonction suivante permet de faire des prédictions de la variable d'intérêt.

```
predictions <- function(maTableTest, muChap, betaChap){
  yChap <- muChap * rep(1, dim(maTableTest)[1] + as.matrix(maTableTest[,]) %% betaChap)
  return(yChap)
}
```

1.2.2 Sélections

La fonction suivante sera utile pour la sélection de variables ; elle donne un sous-ensemble de variables pertinentes à sélectionner.

```
subsetSelection <- function(resAlgo, varExpl, mini, maxi){
  numselected <- c(which(resAlgo < mini), which(resAlgo > maxi))
  selected <- character()
  valeurs <- numeric()
  for (i in 1:length(numselected)) {
    selected[i] <- names(varExpl)[numselected[i]]
    valeurs[i] <- resAlgo[numselected[i]]
    subset <- cbind(selected, valeurs)
    subset <- as.data.frame(subset)
    subset$valeurs <- as.numeric(as.vector(subset$valeurs))
  }
  return(subset)
}
```

1.3 Utilisation du package rrBLUP

La fonction `mixed.solve` du package `rrBLUP` nous permet d'estimer les paramètres β et μ du modèle aléatoire

```
library(rrBLUP)
resBLUP <- mixed.solve(ySimu[1:100], Z = as.matrix(simuVarExpl[1:100, ]),
                      X = as.matrix(rep(1, 100)), method = "REML")
muChap <- resBLUP$beta
betaChap <- resBLUP$u
```

```
resBLUP$Ve
```

```
## [1] 21.89501
```

```
resBLUP$Vu
```

```
## [1] 0.05684796
```

L'estimation de la variance résiduelle vaut 16.43825 ; ce qui révèle une forte variabilité au sein de celle-ci.

1.3.1 Estimation

1.3.2 Prédiction

1.3.3 Sélection

1.4 La régression bayésienne A

1.4.1 Fonction

1.4.2 Estimation

1.4.3 Prédiction

1.4.4 Sélection

1.5 La régression bayésienne LASSO

1.5.1 Estimation

1.5.2 Prédiction

1.6 La méthode de sélection bayésienne SSVS