# Rmixmod: A MIXture MODelling R package

**R. Lebret**[a,1] and **S. Iovleff**[a,2] and **F. Langrognet**[b]

[a]Laboratoire de mathématiques Paul Painlevé
Université Lille 1 - U.M.R. CNRS 8524
Cité Scientifique - 59655 Villeneuve d'Ascq Cedex - FRANCE
[1]remi.lebret@math.univ-lille1.fr
[2]serge.iovleff@univ-lille1.fr

[b]Laboratoire de mathématiques de Besançon
Université de Franche-Comté - U.M.R. CNRS 6623
16 route de Gray - 25030 Besançon - FRANCE
florent.langrognet@univ-fcomte.fr

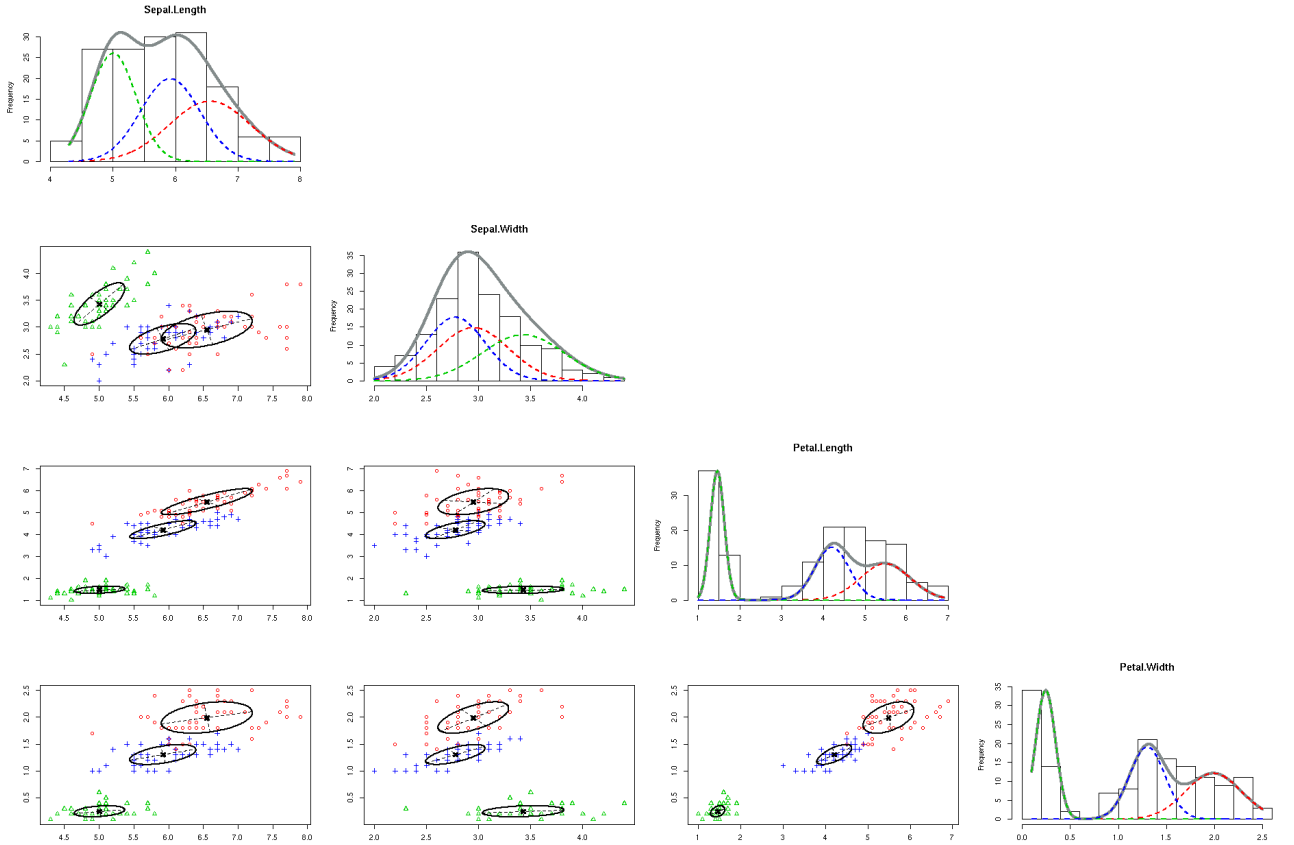
**Keywords** : model-based clustering, classification, discriminant analysis, R, Rmixmod

**Abstract** : Due to its interpretabilities, the model-based clustering approach for fitting a mixture model of multivariate gaussian or multinomial components to a given data set, is an increasing prefered tool. Mixmod is a well established software able to perform in a fast and efficient way this task. It is written in C++ and the core library have been interfaced with scilab and matlab. It lacks an interface with the R program.
The Rmixmod package provide a bridge between the core library of Mixmod and the R statistical computing environment. Both clustering analysis and discriminant analysis can be performed using Rmixmod. Rmixmod is dealing with multivariate Gaussian mixture models for quantitative data and multivariate multinomial mixture models for qualitative data.

**An example of clustering in a quantitative case** : To illustrate the outputs and graphs of Rmixmod, we use the well-known iris flower data set. `iris` is a data frame with 150 cases (rows) and 5 variables (columns) named `Sepal.Length`, `Sepal.Width`, `Petal.Length`, `Petal.Width`, and `Species`. The first fourth variables are quantitative and the `Species` variable is qualitative with 3 modalities. So we know that there are three clusters and we want to find the best gaussian model to fit the data set. That can be done by invoking the `mixmodCluster()` method:

```
R> library(Rmixmod)
R> xem <- mixmodCluster(iris[1:4], 3, models=mixmodGaussianModel())
R> summary(xem)
*************************************************************
* Number of samples    =  150
* Problem dimension    =  4
*************************************************************
* Number of cluster =  3
*         Criterion =  BIC(553.4052)
*        Model Type =  Gaussian_p_Lk_Dk_A_Dk
*        Parameters =  list by cluster
*        Cluster  1 :
              Proportion =  0.3333
                   Means =  6.5516 2.9510 5.4909 1.9904
               Variances = |     0.4282     0.1078     0.3310     0.0630 |
                           |     0.1078     0.1155     0.0879     0.0606 |
```

```
                              |      0.3310      0.0879      0.3585      0.0831 |
                              |      0.0630      0.0606      0.0831      0.0847 |
*         Cluster  2 :
               Proportion =  0.3333
                    Means =  5.0060 3.4280 1.4620 0.2460
                Variances = |      0.1328      0.1089      0.0192      0.0116 |
                              |      0.1089      0.1545      0.0120      0.0100 |
                              |      0.0192      0.0120      0.0283      0.0058 |
                              |      0.0116      0.0100      0.0058      0.0107 |
*         Cluster  3 :
               Proportion =  0.3333
                    Means =  5.9177 2.7781 4.2106 1.3022
                Variances = |      0.2274      0.0761      0.1487      0.0441 |
                              |      0.0761      0.0807      0.0744      0.0347 |
                              |      0.1487      0.0744      0.1692      0.0510 |
                              |      0.0441      0.0347      0.0510      0.0342 |
*    Log-likelihood =  -186.5112
************************************************************
R> plot(xem)
```




**Références**

[1] Lebret R., Iovleff S., Langrognet F., (2012). Rmixmod: A MIXture MODelling R package. To appear in *Journal of Statistical Software*.

[2] Biernacki C., Celeux G., Govaert G., Langrognet F., (2006). Model-Based Cluster and Discriminant Analysis with the MIXMOD Software. *Computational Statistics and Data Analysis*, vol. 51/2, pp. 587-600.