

```
#####
#####
## Script R pour reproduire les analyses de la présentation:
##                                     Matthias Studer
##      Analyse de dispersion de séquences d'états:
##      analyser les liens entre trajectoires et variables explicatives
## Rencontres méthodes et recherche, FORS, Lausanne, 24 janvier 2012
#####
#####

#####
## N'hésitez pas à poser vos questions!
## Mailing list TraMineR: http://mephisto.unige.ch/traminer/contrib.shtml
## !!ATTENTION!!! Pour la représentation graphique de l'arbre
## Il faut installer GraphViz au préalable avant de lancer R!!
## Téléchargeable gratuitement pour toutes les plateformes ici:
## http://www.graphviz.org
#####

### Chargement des librairies
library(TraMineR)
library(cluster)
library(vcd)

#####
## Préparation des données / Recodage
## Les variables dans le jeu de données originale sont
## codées sous formes de dummies, ici nous les regroupons
## en variables à plusieurs catégories.
#####

## Chargement des données exemples
data(mvad)

## Recodage de la variables régions.
region <- character(nrow(mvad))

for(r in c("Belfast", "N.Eastern", "Southern", "S.Eastern", "Western")){
  region[mvad[,r]=="yes"] <- r
}
mvad$region <- factor(region)
sr <- c(yes="Male", no="Female")
mvad$sex <- factor(sr[as.character(mvad$male)])
sr <- c(yes="Catholic", no="Protestant")
mvad$religion <- factor(sr[as.character(mvad$catholic)])

levels(mvad$gcse5eq) <- c("Lower qualif.", "Higher qualif.")

#####
## Construction des objets "séquences d'états" avec seqdef
#####
mvad.lab <- c("Employment", "Further Education", "Higher Education",
             "Joblessness", "School", "Training")
mvad.shortlab <- c("EM", "FE", "HE", "JL", "SC", "TR")
mvad.seq <- seqdef(mvad, 17:86, states=mvad.shortlab, labels=mvad.lab, xtstep=6)

#####
### Graphique des 10 premières séquences
#####
```

```

seqiplot(mvad.seq, border=NA, title="Dix premières séquences")

#####
### Chronogramme
#####
seqdplot(mvad.seq, border=NA)

#####
### Calcul des dissimilarités entre séquences.
#####

## Coûts de substitution entre états tels que définis dans l'article
subm.custom <- matrix(c(0,1,1,2,1,1, 1,0,1,2,1,2, 1,1,0,3,1,2, 2,2,3,0,3,1,
1,1,1,3,0,2, 1,2,2,1,2,0), nrow=6, ncol=6, byrow=TRUE,
dimnames=list(mvad.shortlab, mvad.shortlab))
subm.custom
## Calcul de la matrice des distances entre séquences.
mvad.dist <- seqdist(mvad.seq, method="OM", indel=1.5, sm=subm.custom)

## Clustering en trois groupes avec l'algorithme PAM
pamclust <- pam(as.dist(mvad.dist), diss=TRUE, cluster.only=TRUE, k=3)
## Ajout des labels des groupes
clust.labels <- c("Emploi", "Étude supérieur", "Sans emploi")
pamclustfac <- factor(pamclust, labels = clust.labels)

#####
### Chronogramme des types obtenus.
#####
seqdplot(mvad.seq, group=pamclustfac, border=NA)

#####
### Graphique des trajectoires selon la variable "funemp"
### Père sans emploi
#####
seqdplot(mvad.seq, group=mvad$funemp, border=NA)

#####
### Test du chi-carré et v de Cramer pour chaque variable
#####

allvars <- c("sex", "region", "religion", "funemp", "fmpr", "livboth",
"Grammar", "gcse5eq")
## Pour chaque variable "v"
for( v in allvars){
  ## Affiche le nom de la variable
  print(v)
  ## Tableau croisé entre "v" et "pamclust" (typologie)
  tb <- table(pamclust, mvad[, v])
  ## Test du chi-carré
  print(chisq.test(tb))
  ## Mesure d'association dont le v de Cramer
  print(assocstats(tb))
}

```

```
#####
## MDS colorié en fonction du clustering
#####

## MDS: représentation dans le plan des séquences.
mds <- cmdscale(mvad.dist, k=2)

plot(mds, col=pamclust, pch=pamclust, main="Clustering", xlab="Dim 1", ylab="Dim
2")
legend("bottomleft", legend=clust.labels, fill=1:3)

#####
## Test des dispersions.
## R^2: association entre trajectoire et facteur explicatif
## Levene: test des différences de trajectoires
#####

## Pour chaque variable "v"
for( v in allvars){
  ## Affiche le nom de la variable
  print(v)
  ## Ajuste le seed pour que les résultats soient reproductibles
  (reproduction des mêmes permutations aléatoires).
  set.seed(1)
  ## Calcul de l'association avec la variable "v", R=5000 permutations
  ds <- dissassoc(mvad.dist, mvad[,v], R=5000)
  ## Affiche les résultats
  print(ds)
}

## Calcul de l'association avec la variable "livboth"
## Ajuste le seed pour que les résultats soient reproductibles (reproduction des
mêmes permutations aléatoires).
set.seed(1)
ds <- dissassoc(mvad.dist, mvad[, "livboth"], R=5000)
## Représentation de la courbe de densité de la distribution nul des R^2
hist(ds, test="Pseudo R2", main="Vivre avec ses deux parents", lwd=3, txtcex=1.2)

#####
### Evolution de l'association
#####
mvad.diff <- seqdiff(mvad.seq, group=mvad$funemp, seqdist_arg=list(method="OM",
indel=1.5, sm=subm.custom))

#####
### Graphique de l'évolution du R^2
#####
plot(mvad.diff, lwd = 3, stat=c("Pseudo R2"))

#####
### Modèle multifacteur
#####
## Ajuste le seed pour que les résultats soient reproductibles (reproduction des
mêmes permutations aléatoires).
set.seed(1)
da.mfac <- dissmfacw(mvad.dist ~ male + Grammar + funemp +
gcse5eq + fmpr + livboth, data=mvad, R=1000)
print(da.mfac)

#####
```

```
## Arbre de séquences
#####
## Ajuste le seed pour que les résultats soient reproductibles (reproduction des
mêmes permutations aléatoires).
set.seed(1)
st <- seqtree(mvad.seq~gcse5eq+Grammar+funemp+sex, data=mvad, R=5000,
diss=mvad.dist)

## Représentation graphique de l'arbre
## !!ATTENTION!!! Il faut installer GraphViz avant de lancer R
## Téléchargeable gratuitement pour toutes les plateformes ici:
## http://www.graphviz.org
seqtreedisplay(st, type="d", border=NA)

#####
## Test des différences de dispersion déjà montré auparavant
## (avec la fonction dissassoc et le test de Levene)
#####

#####
## Evolution des différences de dispersion
#####
plot(mvad.diff, lwd = 3, stat=c("Pseudo R2", "Levene"))
## Evolution de la dispersion
plot(mvad.diff, lwd = 3, stat="discrepancy", legendposition="bottomleft")
```