L'adaptation de l'analyse factorielle exploratoire à des données discrètes

Gilles Raîche, UQAM

Séminaire du Collectif sur le développement et les applications en mesure et évaluation

22 février 2012

CONTENU DE L'ATELIER

- Problématique
- Données sur les pratiques d'évaluation en salle de classe
- Détermination du nombre de facteurs
- Comparaisons des méthodes
- Conclusions
- Références et logiciels

PROBLÉMATIQUE

- L'analyse factorielle exploratoire
- Bref historique
- Modèle linéaire de base: sujet *i*, variable manifeste *j*, facteur *k*

$$x_{ij} = \sum_{k=1}^{p} \lambda_j f_{jk} + e_j$$

- Non adapté à des données discrètes
- Utilisation des corrélations tétrachoriques et polychoriques
- Analyse factorielle à information complète (*full information factor analysis*)

Données sur les pratiques d'évaluation en salle de classe

- 1608 sujets étudiants : primaire, secondaire, collégial et universitaire
- 24 items à 4 choix de réponses de type fréquence : de jamais à toujours
- 4 dimensions théoriques : intégration, équité, authenticité et planification
- Exemple :
  - Mon enseignant utilise des tâches d'évaluation qui sont utiles pour moi.

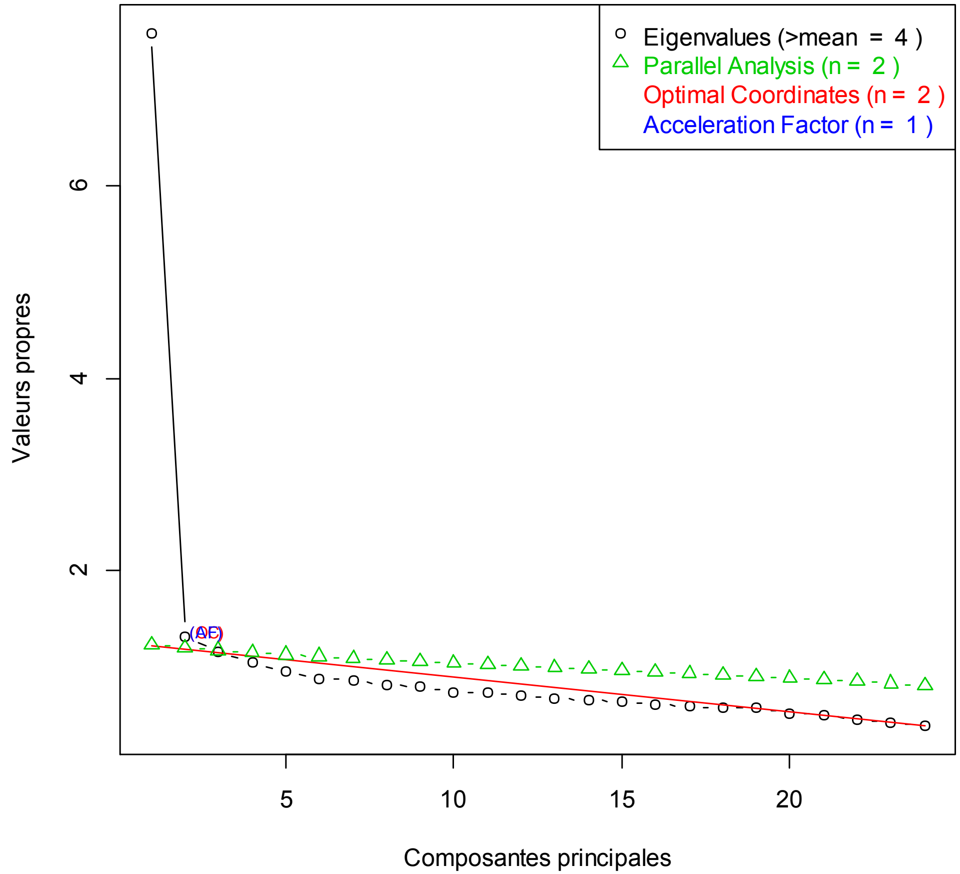
DONNÉES SUR LES PRATIQUES D'ÉVALUATION EN SALLE DE CLASSE

```
# Librairies nécessaires
> require(nFactors); require(psych); require(mirt)
> X <- read.table(file="http://www.camri.uqam.ca/camri/data/PRATIQUES/ETUDIANTS/RAICHE_CDAME_DATA_2012_02_22.dat", header=TRUE)
> X <- data.frame(X[,-1]) # Retrait du facteur Ordre d'enseignement

# Informations descriptives sur les données
> dim(X)
[1] 1608   24
> summary(X[,1:3])
CP1              CP2              CP3
Min.   : 1.000   Min.   : 1.000   Min.   : 1.000
1st Qu.: 3.000   1st Qu.: 3.000   1st Qu.: 3.000
Median : 3.000   Median : 4.000   Median : 3.000
Mean   : 3.203   Mean   : 3.484   Mean   : 3.219
3rd Qu.: 4.000   3rd Qu.: 4.000   3rd Qu.: 4.000
Max.   : 4.000   Max.   : 4.000   Max.   : 4.000
NA's   :29.000   NA's   :32.000   NA's   :33.000
```

DÉTERMINATION DU NOMBRE DE FACTEURS SELON UNE ANALYSE DITE CLASSIQUE

- Critère de Kaiser, test de l'éboulis et analyse parallèle

	Valeur propre	%
1	7,60	32 %
2	1,31	5 %



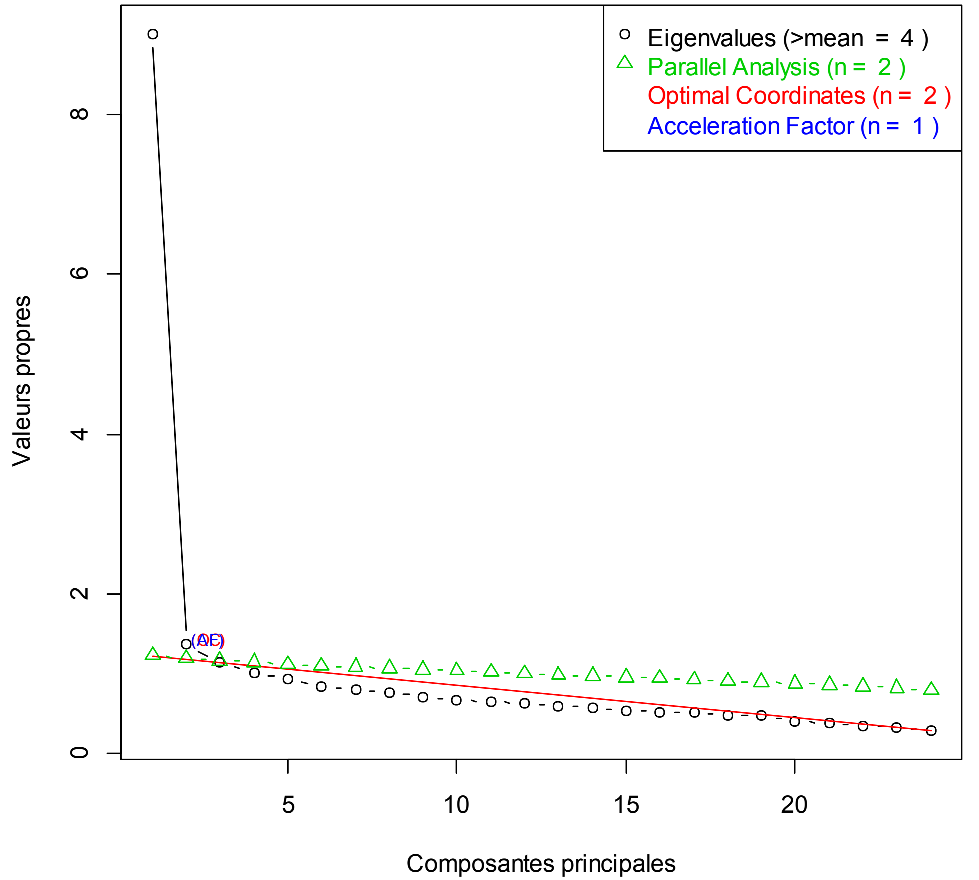
Détermination du nombre de facteurs selon une analyse dite classique

```
# Analyse factorielle exploratoire classique
> corX        <- cor(X, use="pairwise.complete.obs")

> eigenValues <- eigenComputes(x=X, use="pairwise.complete.obs")

> aParallel   <- parallel(subject = dim(X)[1], var = dim(X)[2],
                          rep = 1000, cent = 0.50)$eigen$mevpea

> nFact       <- nScree(eig=eigenValues, x=eigenValues,
                        aparallel=aParallel, cor=TRUE,
                        model="components")

> summary(nFact, digits=2)

> plot(nFact, xlab="Composantes principales", ylab="Valeurs
     propres", main="")
```

DÉTERMINATION DU NOMBRE DE FACTEURS SELON UNE ANALYSE BASÉE SUR LES VARIABLES SOUS-JACENTES

- Critère de Kaiser, test de l'éboulis et analyse parallèle

	Valeur propre	%
1	9,01	38 %
2	1,37	6 %



DÉTERMINATION DU NOMBRE DE FACTEURS SELON UNE ANALYSE BASÉE SUR LES VARIABLES SOUS-JACENTES

```
# Analyse factorielle exploratoire avec variables sous-jacentes
> polyX       <- polychoric(X, smooth=TRUE, global=TRUE)
> eigenPoly <- eigenComputes(polyX[[1]], cor=TRUE])

> aParallel   <- parallel(subject = dim(X)[1], var = dim(X)[2],
                             rep = 1000, cent = 0.50)$eigen$mevpea

> nFactPoly   <- nScree(eig=eigenPoly, x=eigenPoly,
                   aparallel=aParallel, cor=TRUE, model="components")
> summary(nFactPoly, digits=2)

> plot(nFactPoly, xlab="Composantes principales", ylab="Valeurs
     propres", main="")
```

DÉTERMINATION DU NOMBRE DE FACTEURS SELON UNE ANALYSE FACTORIELLE À INFORMATION TOTALE

- Critères: ANOVA, AIC et BIC

anova(mod4, mod5)

X2 = 138.754 (SE = 0.31), df = 20, **p = 0**

AIC difference = 98.754 (SE = 0.31)

BIC difference = **-8.901** (SE = 0.31)

DÉTERMINATION DU NOMBRE DE FACTEURS SELON UNE ANALYSE FACTORIELLE À INFORMATION TOTALE

```
> mod1 <- polymirt(data=X, nfact=1)
> mod2 <- polymirt(data=X, nfact=2)
> mod3 <- polymirt(data=X, nfact=3)
> mod4 <- polymirt(data=X, nfact=4)
> mod5 <- polymirt(data=X, nfact=5)
> anova(mod1, mod2)
> anova(mod2, mod3)
> anova(mod3, mod4)
> anova(mod4, mod5)
Chi-squared difference:
X2 = 138.754 (SE = 0.31), df = 20, p = 0
AIC difference = 98.754 (SE = 0.31)
BIC difference = -8.901 (SE = 0.31)
```

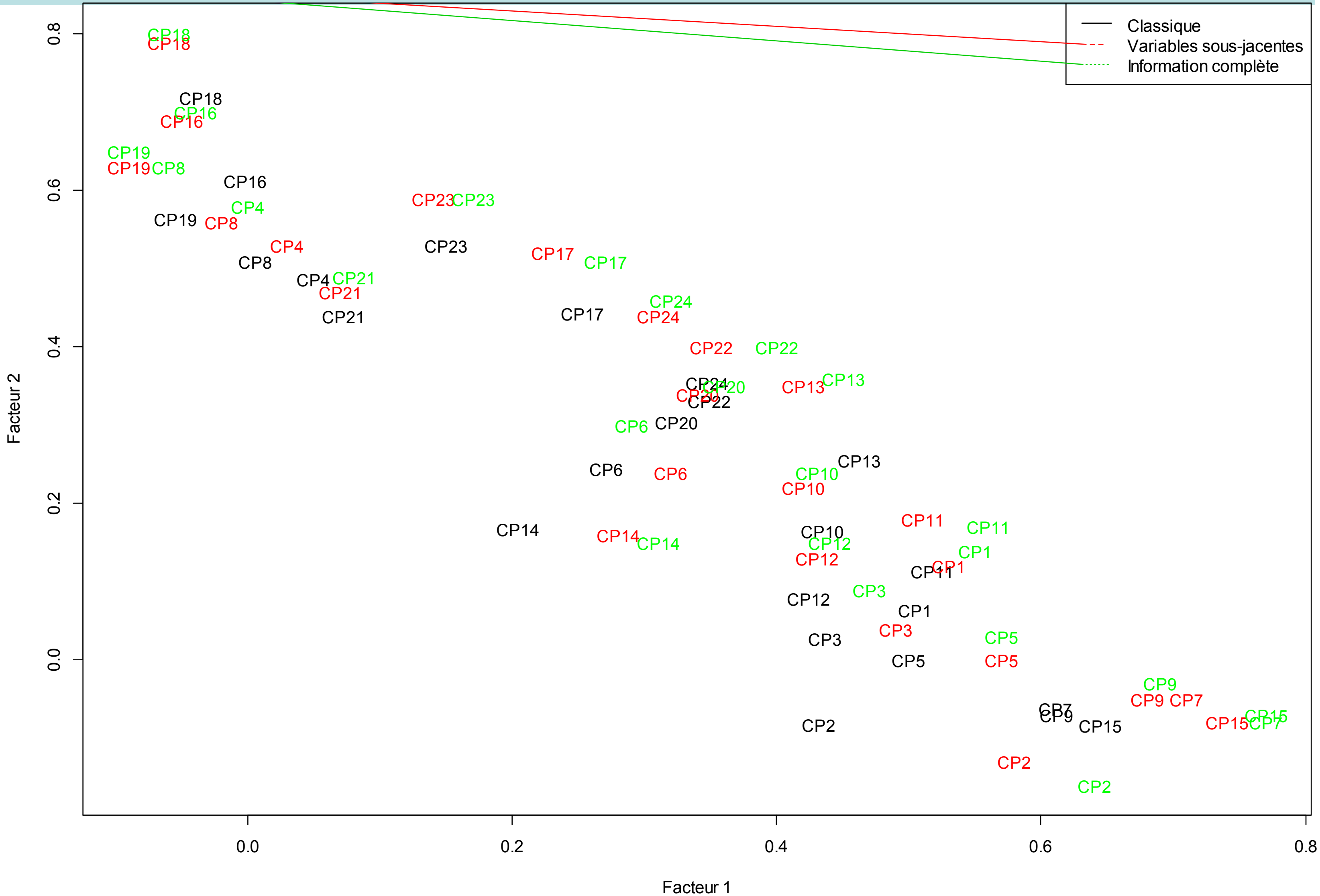
Comparaisons des méthodes (saturations)

Item	Classique (-0,74)			Variables sous-jacentes (-0,74)			Information totale (-0,75)		
	F1	F2	h^2	F1	F2	h^2	F1	F2	h^2
1	0,50	0,06	0,25	0,53	0,12	0,30	0,55	-0,14	0,32
2	0,43	-0,08	0,19	0,58	-0,13	0,35	0,64	0,16	0,44
3	0,44	0,03	0,19	0,49	0,04	0,24	0,47	-0,09	0,23
4	0,05	0,49	0,24	0,03	0,53	0,28	0,00	-0,58	0,34
5	0,50	0,00	0,25	0,57	0,00	0,32	0,57	-0,03	0,33
6	0,27	0,25	0,14	0,32	0,24	0,16	0,29	-0,30	0,17
7	0,61	-0,06	0,38	0,71	-0,05	0,51	0,77	0,08	0,60
8	0,01	0,51	0,26	-0,02	0,56	0,31	-0,06	-0,63	0,40
9	0,61	-0,07	0,38	0,68	-0,05	0,46	0,69	0,03	0,48
10	0,43	0,17	0,21	0,42	0,22	0,22	0,43	-0,24	0,24
11	0,52	0,11	0,28	0,51	0,18	0,29	0,56	-0,17	0,34
12	0,42	0,08	0,18	0,43	0,13	0,20	0,44	-0,15	0,22
13	0,46	0,26	0,28	0,42	0,35	0,30	0,45	-0,36	0,33
14	0,20	0,17	0,07	0,28	0,16	0,10	0,31	-0,15	0,12
15	0,64	-0,08	0,42	0,74	-0,08	0,55	0,77	0,07	0,60
16	0,00	0,61	0,37	-0,05	0,69	0,48	-0,04	-0,70	0,49
17	0,25	0,44	0,26	0,23	0,52	0,32	0,27	-0,51	0,33
18	-0,04	0,72	0,52	-0,06	0,79	0,63	-0,06	-0,80	0,64
19	-0,06	0,56	0,32	-0,09	0,63	0,40	-0,09	-0,65	0,43
20	0,32	0,30	0,19	0,34	0,34	0,23	0,36	-0,35	0,25
21	0,07	0,44	0,20	0,07	0,47	0,23	0,08	-0,49	0,25
22	0,35	0,33	0,23	0,35	0,40	0,28	0,40	-0,40	0,32
23	0,15	0,53	0,30	0,14	0,59	0,37	0,17	-0,59	0,38
24	0,35	0,35	0,24	0,31	0,44	0,29	0,32	-0,46	0,31

COMPARAISONS DES MÉTHODES (SATURATIONS)

```
> corX        <- cor(X, use="pairwise.complete.obs")
> classique <- factanal(factors=2, covmat=corX,
                         rot="promax")
> PHIclass  <- solve(classique$rot) %*%
               t(solve(classique$rot))
> classique; PHIclass

> polyX          <- polychoric(X, smooth=TRUE, global=TRUE)$rho
> sousjacentes <- factanal(factors=2, covmat=polyX[[1]],
                            rot="promax")
> PHIpoly      <- solve(sousjacentes $rot) %*%
                  t(solve(sousjacentes $rot))

> sousjacentes; PHIpoly


> mod2 <- polymirt(data=X, nfact=2)

> summary(mod2, rot="promax")
```

COMPARAISONS DES MÉTHODES (SATURATIONS)



COMPARAISONS DES MÉTHODES (SATURATIONS)

```
> loadFull <- round(unclass(summary(mod2, digits=2,
                            rotate='promax')[[1]]), 2)
➤ loadFull[,2] <- -loadFull[,2]

> cex=1
> plot(loadFull, type="n",xlab="Facteur 1",
       ylab="Facteur 2")
> text(loadings(sousjacentes),
       rownames(loadings(sousjacentes)), cex=cex)
> text(loadings(classique),
       rownames(loadings(classique)), cex=cex,
       col="red")
> text(loadFull, rownames(loadPoly), cex=cex,
       col="green")
> legend("topright", legend=c("Classique", "Variables
         sous-jacentes", "Information complète"),
         lty=1:3, col=c(1:3))
```

Conclusions

- Le nombre de facteurs à retenir varie selon le critère utilisé
- L'interprétation des facteurs extraits ne varie pas énormément
- Toutefois, la méthode classique est celle qui présente la plus petite communauté
- Tandis que la méthode par information complète présente la communauté la plus élevée
- Moins le nombre de catégories utilisée est grand, moins l'approche classique est appropriée

RÉFÉRENCES ET LOGICIELS

Bartholomew, D. J., Steele, F., Moustaki, I. et Galbraith, J. I. (2002). *The analysis and interpretation of multivariate data for social scientists*. Boca Raton, Floride : Chapman et Hall.

Cai, L. (2010). High-Dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika, 75*, 33-57.

Chalmers, P. (2011). *mirt 0.1.19 – Multidimensional item response theory*. CRAN.

Mulaik, S. A. (2010). *Foundations of factor analysis*. 2e édition. Boca Raton, Floride : Chapman et Hall.

Raîche, G. et Magis, D. (2011). *nFactors 2.12 – Application d'études de dimensionnalité en analyse factorielle*. Montréal, Québec : Université du Québec à Montréal.

Raîche, G. (2006). L'intégration des pratiques d'évaluation des apprentissages aux pratiques pédagogiques dans le contexte des approches par compétences. *Vivre le primaire, 19*(2), 43-45.

CONTACT

- Gilles Raîche
  - http://www.Cdame.uqam.ca
  - Raiche.Gilles@uqam.ca