Elements of validity in Multiple Factor Analysis

Marine Cadoret, Sébastien Lê, Jérôme Pagès

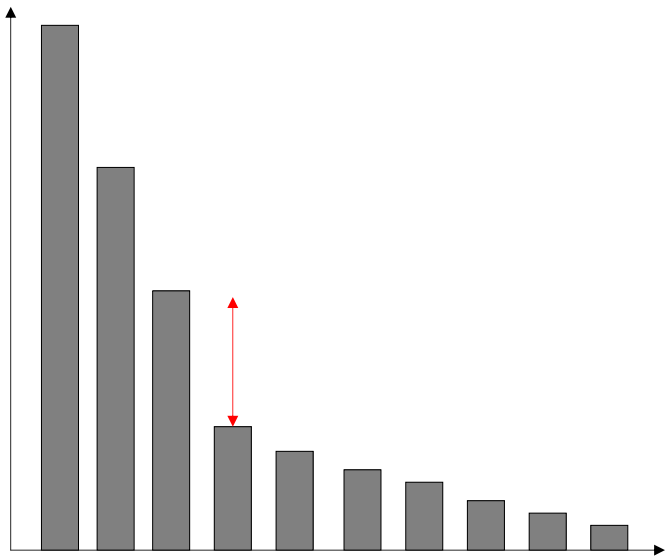
Applied Mathematics Department, Agrocampus Rennes, France

Caserta, june 11th 2008

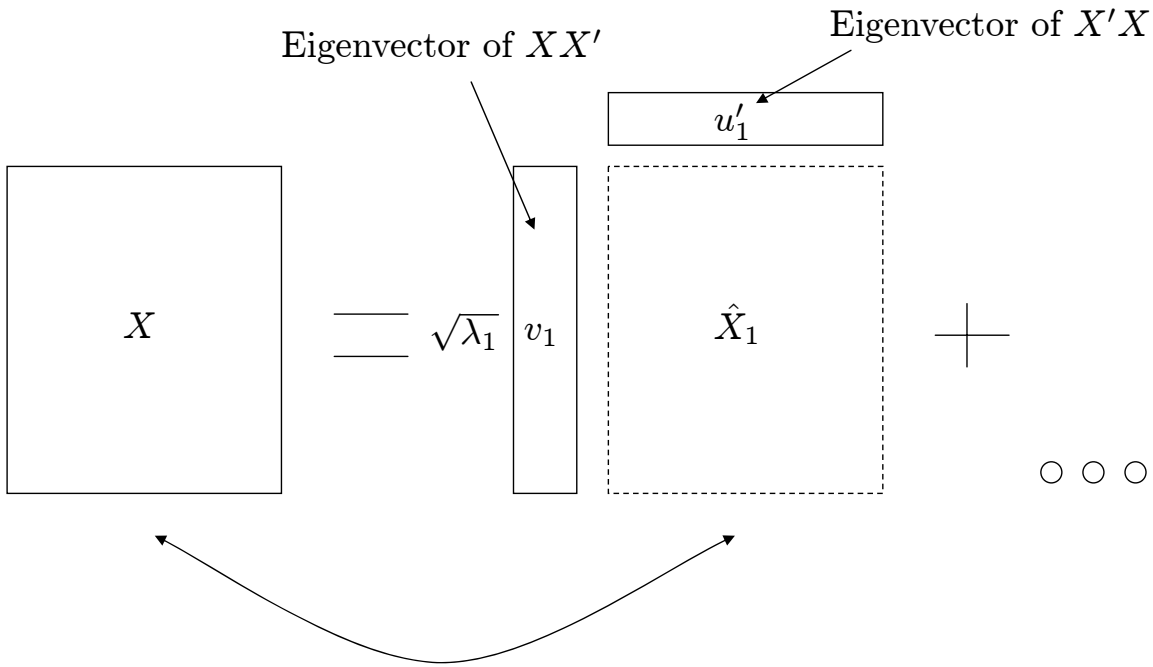
Problem

Selection of the number of dimensions in Principal Component Analysis (PCA) :

- Bar plot of the eigenvalues
- Visual test : Cattell criterion
- Stability in spite of perturbations in the dataset

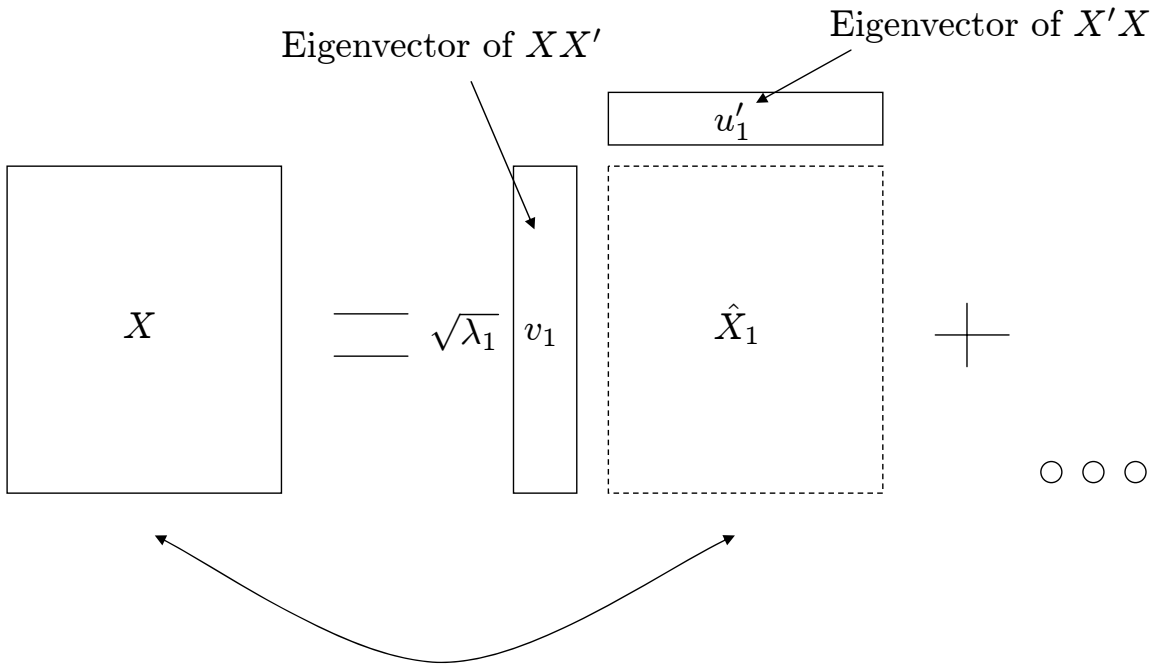


Dray, 2007 : first dimension



- Is the data reconstituted from the first dimension ($\hat{X}_1$) closer to the one of original data (X) than a random table?
- Measure of similarity : RV coefficient (Escouffier, 1973)

Dray, 2007 : first dimension



- Is the observed *RV* coefficient large?
- H_0 : Absence of structure among variables
- Procedure based on permutation tests

First dimension : permutation tests

Calculate the p-value associated to the observed RV :

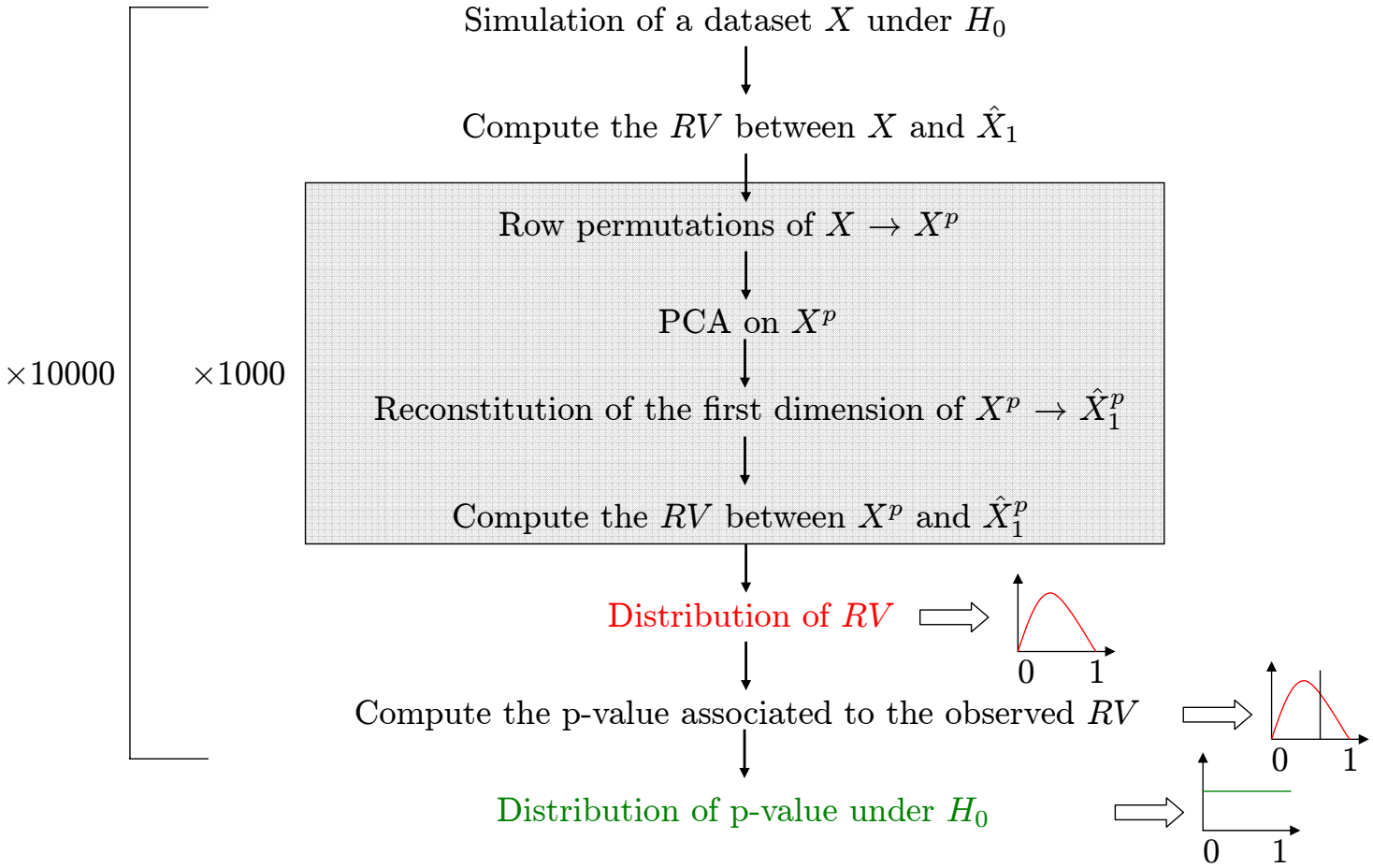
1. Repeat a large number of times :
   1. Independent row permutations within each column of $X \longrightarrow X^p$
   2. PCA on X^p
   3. Reconstitution of X^p from the first dimension of the PCA on $X^p \longrightarrow \hat{X}_1^p$
   4. Calculate $RV(X^p, \hat{X}_1^p)$
2. Distribution of RV coefficient under H_0
3. Identify the observed value in this distribution to get the p-value

Evaluation of Dray's procedure

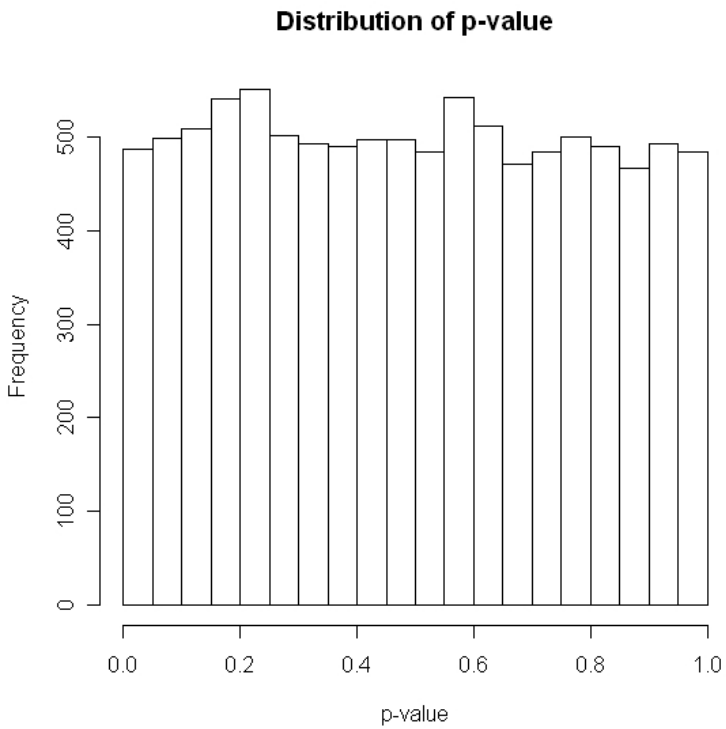
- Behavior of the procedure under the alternative hypothesis (Dray)
- Behavior of the procedure under the null hypothesis

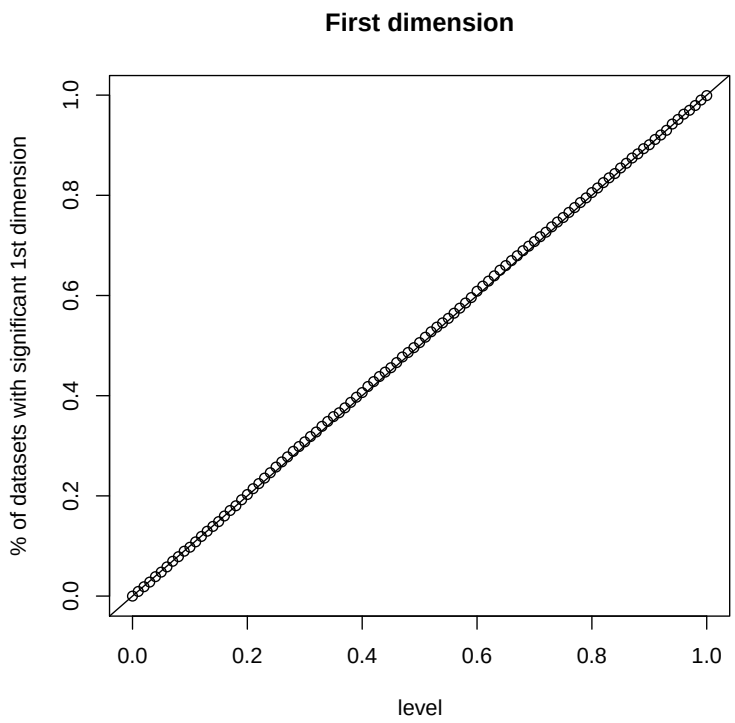
Behavior of the procedure under H_0 : first dimension

- simulation algorithm



Behavior of the procedure under H_0 : first dimension

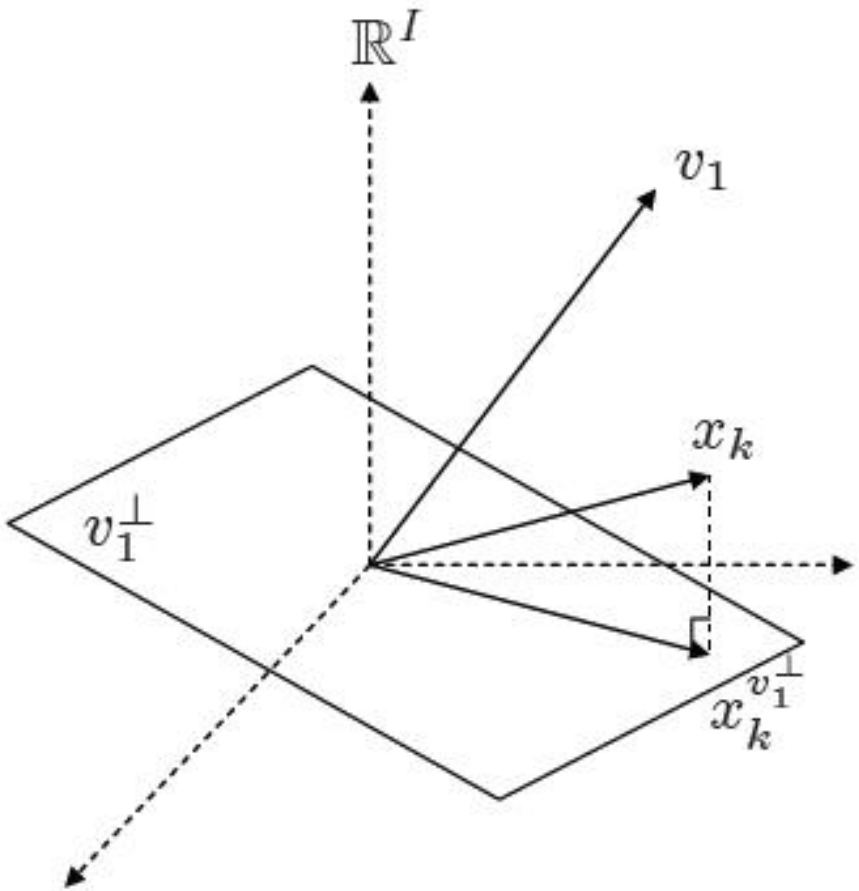




$\Rightarrow$ For a significant level of α%, we observe α% of data tables having a significant first dimension

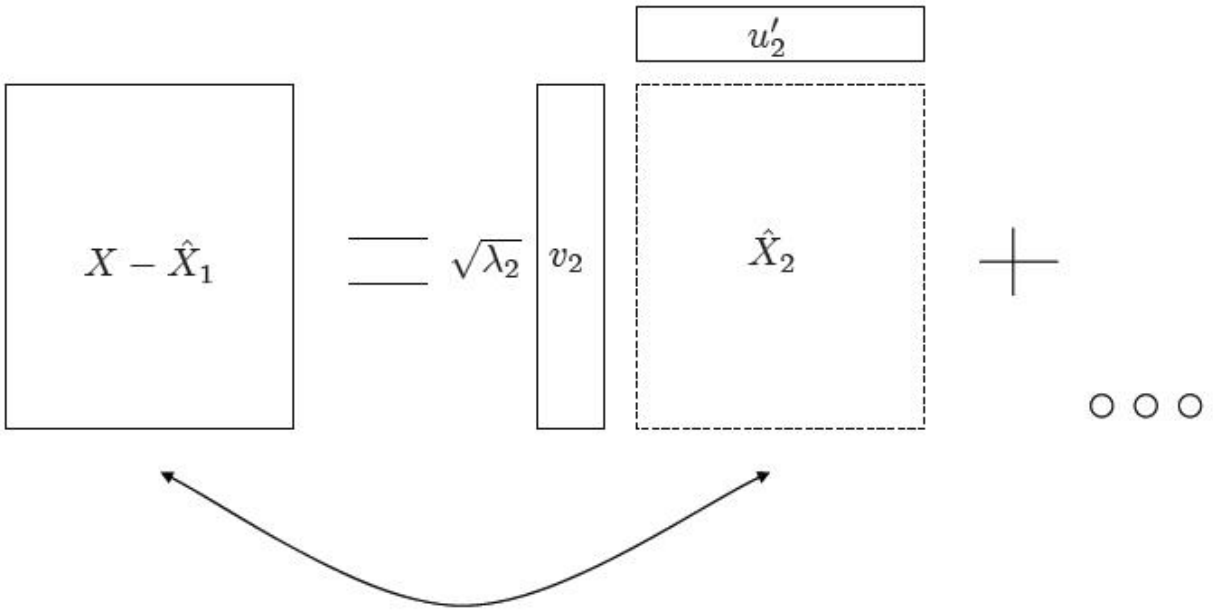
Dray, 2007 : second dimension

We are in the space orthogonal to the first dimension



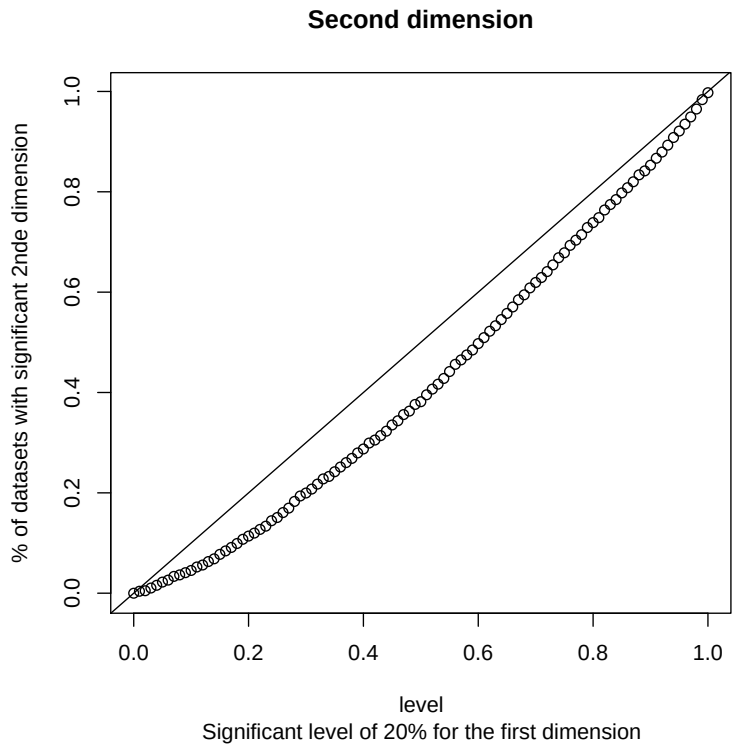
Dray, 2007 : second dimension

We use the same methodology that for the first dimension : we calculate the RV coefficient between $X - \hat{X}_1$ and $\hat{X}_2$.

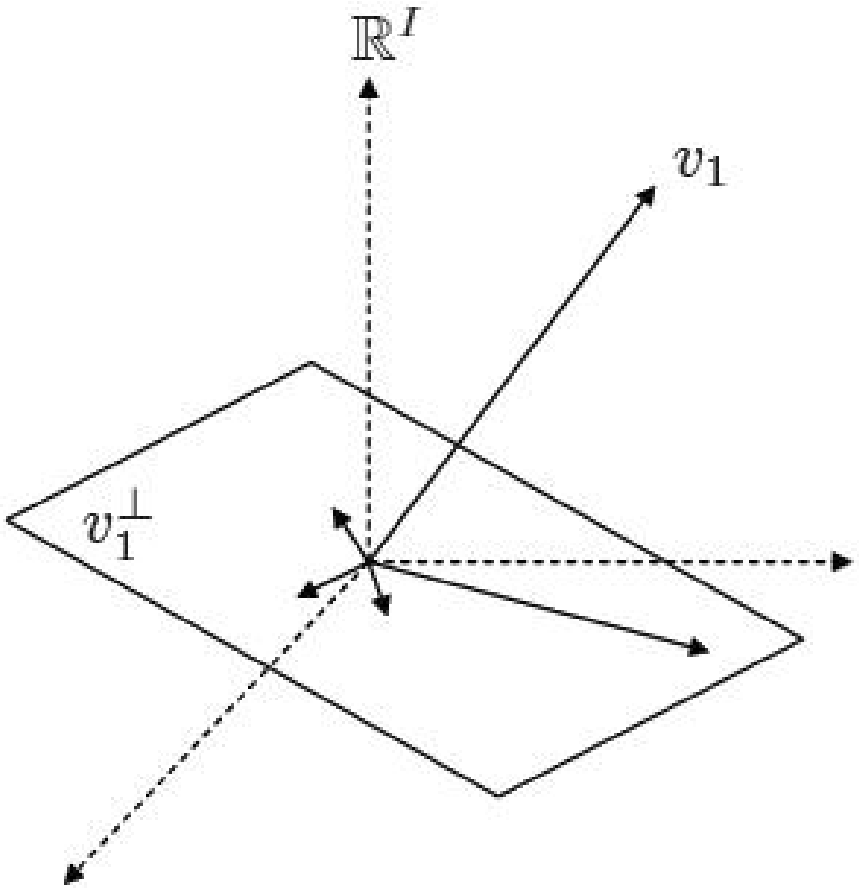


Behavior of the procedure under H_0 : second dimension

- Same simulation procedure



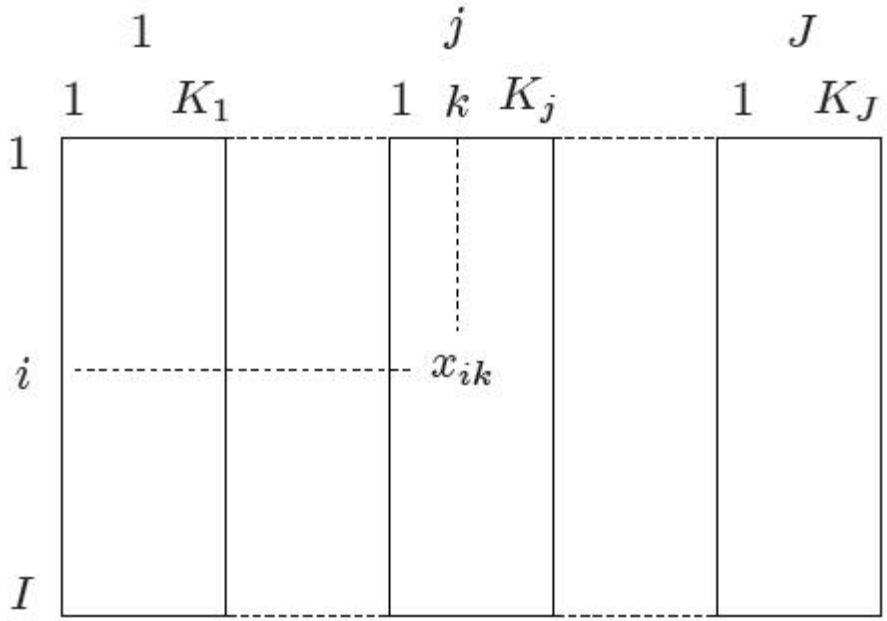
Particular case



$\Rightarrow$ Stability $\neq$ Significant structure

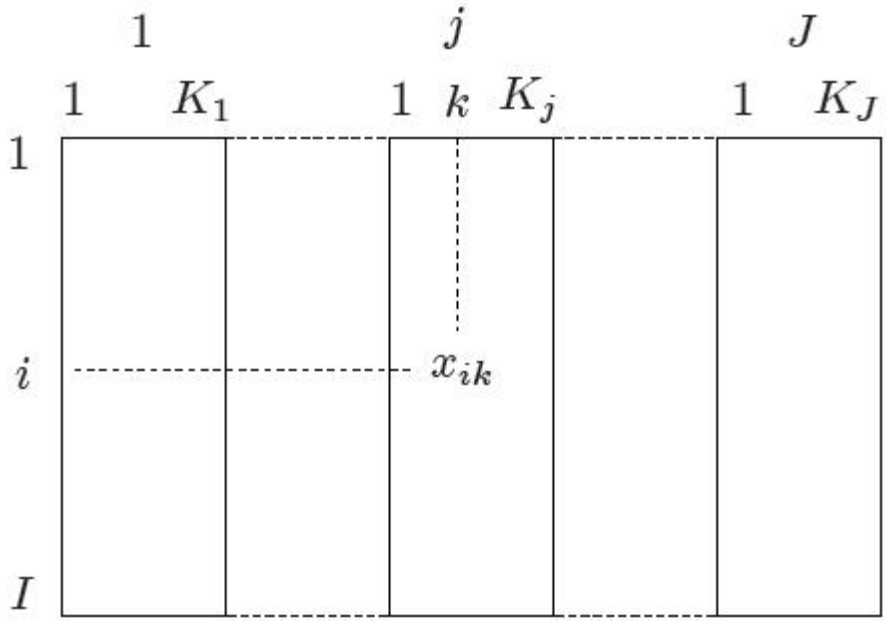
Multiple Factor Analysis

Multiple Factor Analysis deals with data tables in which a set of individuals (I) is described by several groups of variables (J)



MFA highlights a structure common to all the groups, to some groups or specific to a group.

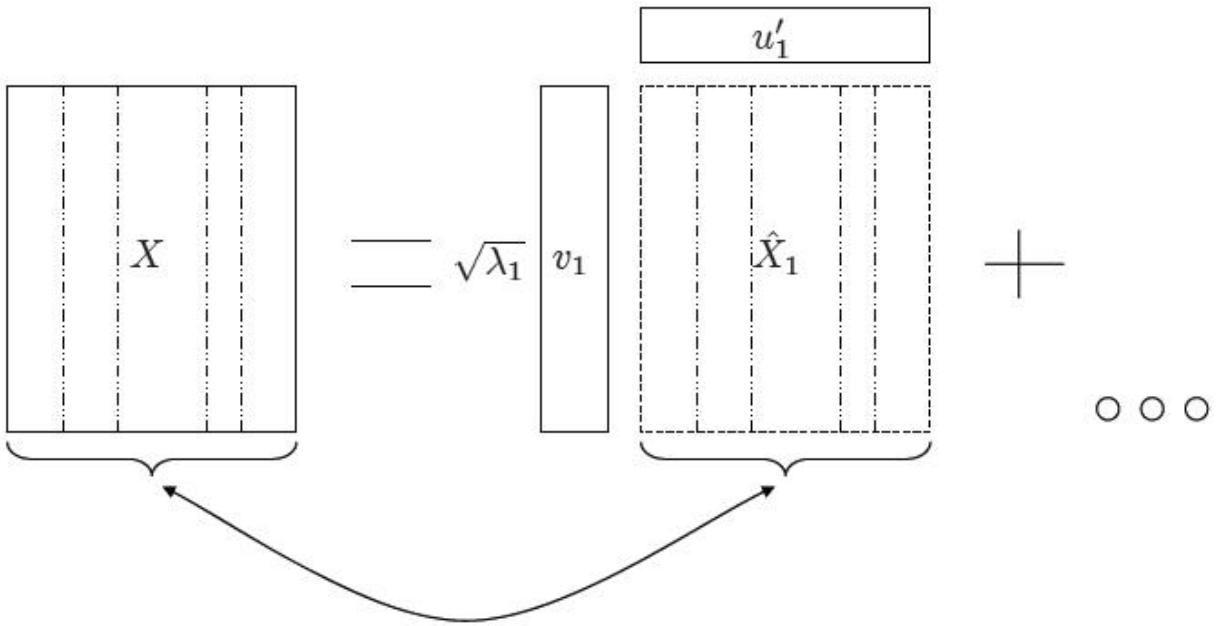
2 main questions



- Does the dimension s correspond to a structure common to several groups?
- In this case, which groups contribute to this common structure?

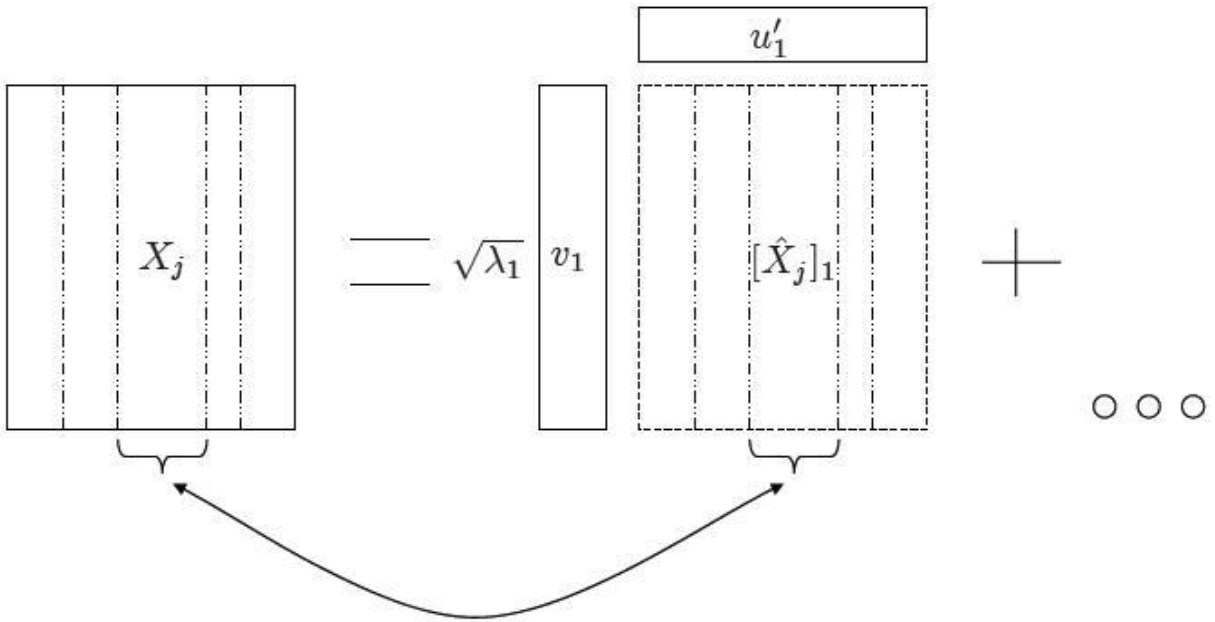
Existence of a common structure in MFA

- H_0 : Absence of common structure (no links between groups)
- Row permutations within each group
- First dimension : Calculate the RV coefficient between X and $\hat{X}_1$



Contribution of groups to the common structure

- H_0 : No contribution of the group j to the common structure
- First dimension : Calculate the RV coefficient between X_j and $[\hat{X}_j]_1$

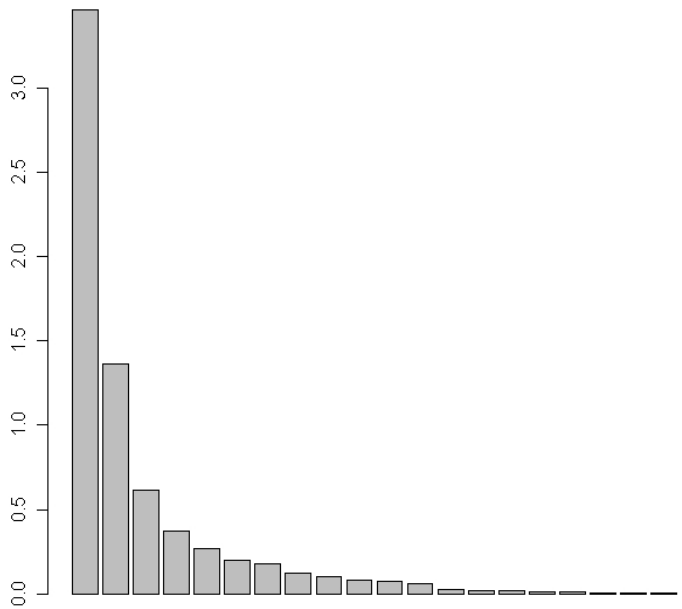


Application

- Classical example of MFA (INRA Angers, Agrocampus Rennes, Spad, FactoMineR)
- 21 wines described by 27 variables gathered into 4 groups :
  - Olfaction before shaking : 5 variables
  - Vision : 3 variables
  - Olfaction after shaking : 10 variables
  - Gustation : 9 variables
- Expected results :

	Dim.1	Dim.2	Dim.3	Dim.4
Olfaction before shaking	×	×	×	
Vision	×			
Olfaction after shaking	×	×	×	
Gustation	×	×		

Application : Number of dimensions



	λ	P-value
Dim.1	3.46	< 0.001
Dim.2	1.37	< 0.001
Dim.3	0.62	0.004
Dim.4	0.37	0.15

Application : Contribution of the groups

- Contribution

	Dim.1	Dim.2	Dim.3	Dim.4
Olfaction before shaking	0.78	0.62	0.37	0.17
Vision	0.85	0.04	0.01	0.05
Olfaction after shaking	0.92	0.47	0.18	0.10
Gustation	0.90	0.24	0.05	0.05
Sum	3.46	1.37	0.62	0.37

- P-value

	Dim.1	Dim.2	Dim.3	Dim.4
Olfaction before shaking	0.02	0.174	0.038	0.127
Vision	0.007	0.104	0.387	0.149
Olfaction after shaking	< 0.001	0.004	< 0.001	0.638
Gustation	< 0.001	0.002	0.278	0.39

Conclusion, perspective

- Dray's procedure extended to MFA
- Ambiguity between stability and significant structure
- Implementation of systematic simulations in MFA

FACTOMINER

http://factominer.free.fr

R package dedicated to exploratory analysis
written by Applied Mathematics Department of Agrocampus