Analysis of sorting data using multiple correspondence analysis and a related method

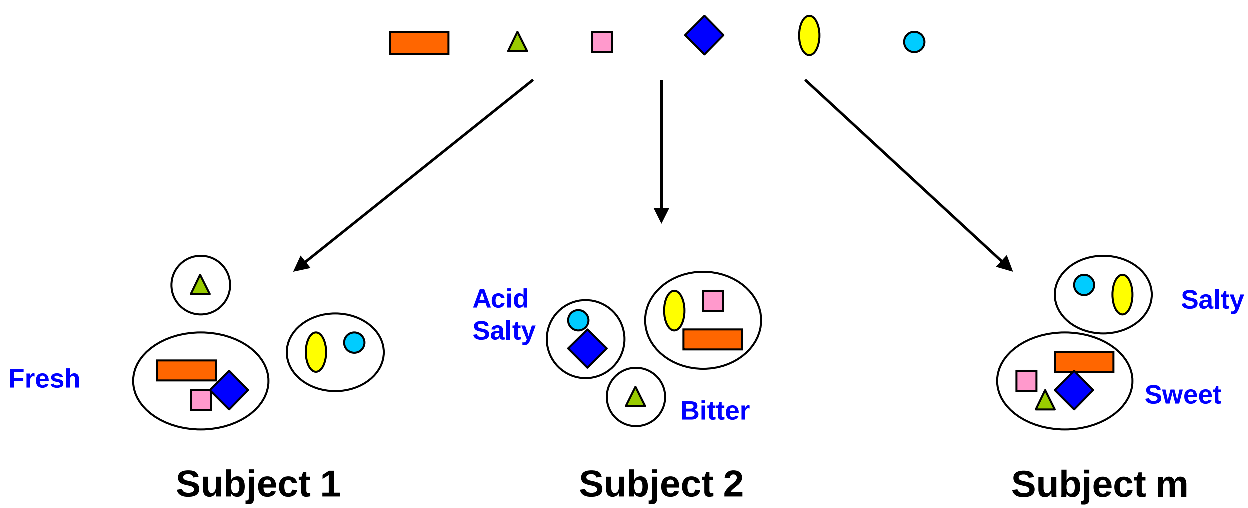
E.M. Qannari   Ph. Courcoux
V. Cariou

ONIRIS, Nantes, F-44322, France

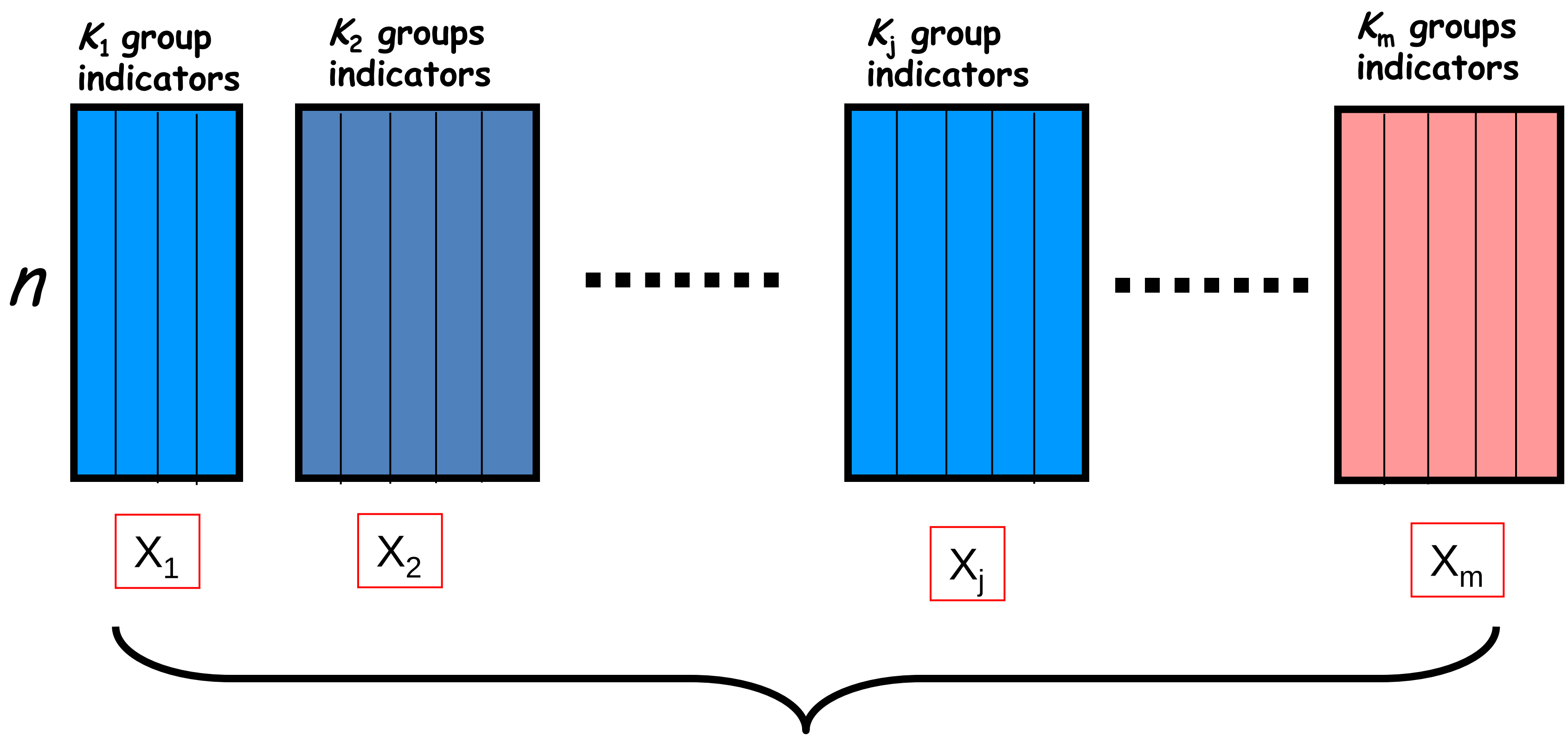
Sorting data : Procedure

n **stimuli evaluated by m subjects:**

"Please, sort the stimuli in as many groups as you consider necessary with the understanding that stimuli in the same group are perceived as similar"



General setting and notations



***m* categorical variables**

(represented by their indicator variables)

Beer data

Data from Abdi H., Chollet S., Valentin D. and Chréa C. (2007) Analysing assessors and products in sorting tasks: DISTATIS,theory and applications. *Food Quality and Preference.*

Data from Abdi et al. (2007)

- **The data relate to an experiment where ten consumers were instructed to sort eight commercial beers.**

#	Beer	Subj1	Subj2	Subj3	Subj4	Subj5	Subj6	Subj7	Subj8	Subj9	Subj10
1	Affligen	1	4	3	4	1	1	2	2	1	3
2	Budweiser	4	5	2	5	2	3	1	1	4	3
3	BucklerBlonde	3	1	2	3	2	4	3	1	1	2
4	Killian	4	2	3	3	1	1	1	2	1	4
5	StLandelin	1	5	3	5	2	1	1	2	1	3
6	BucklerHighland	2	3	1	1	3	5	4	4	3	1
7	FruitDefendu	1	4	3	4	1	1	2	2	2	4
8	EKU28	5	2	4	2	4	2	5	3	4	5

Discrimination indices and MCA

- Given a (quantitative) variable z and let's consider (categorical) variable X_j:

  $\eta^2(z/j)$: discrimination index : the between groups to total variance ratio associated with z and X_j.

- We seek z so as to maximize :

$$I(z) = \sum_{j=1}^{m} \eta^2(z/j)$$

- It is know that this problem leads to MCA

- Subsequent z variables (factors) are sought following the same strategy, under orthogonality constraints.

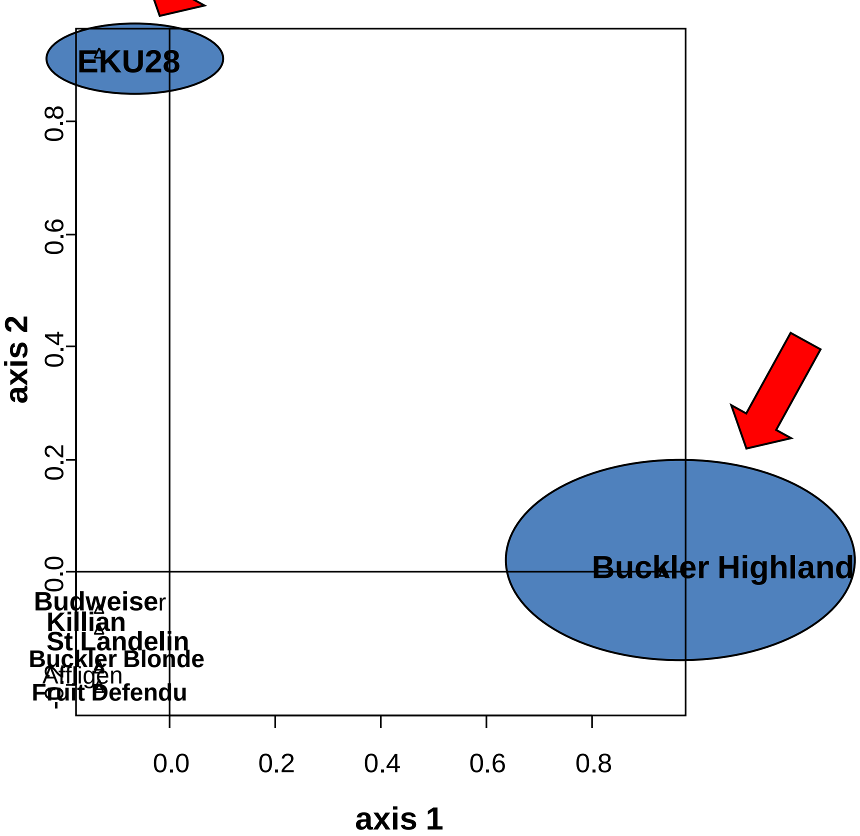
Standardized MCA

- **Alternatively:**

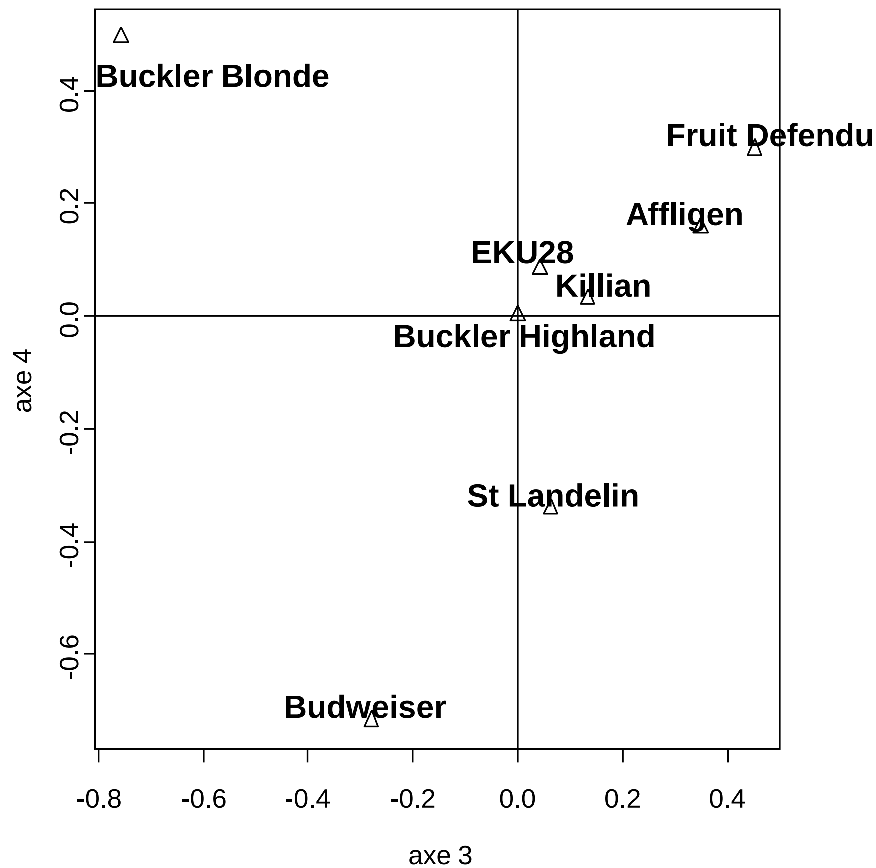
$$I(z) = \sum_{j=1}^{m} \frac{1}{\sqrt{K_j}} \eta^2(z/j)$$

MCA applied to beer data

Reprsentation of the beers axes 1&2



Reprsentation of the beers axes 3&4



Alternative method: maximizing the between groups variances

- $X=[X_1 , X_2 , \ldots, X_m]$ (the indicator variables supposed to be centered)

- Let $z=Xu$ and denote by $B(z/j)$ the <u>between groups variance</u> of z with respect to X_j.

- We define the total between groups variance as:

$$B(z) = \sum_{j=1}^{m} B(z / j)$$

An alternative method to MCA

- **We can show that the vector of loadings u is an eigenvector of the matrix (associated with the largest eigenvalue).**

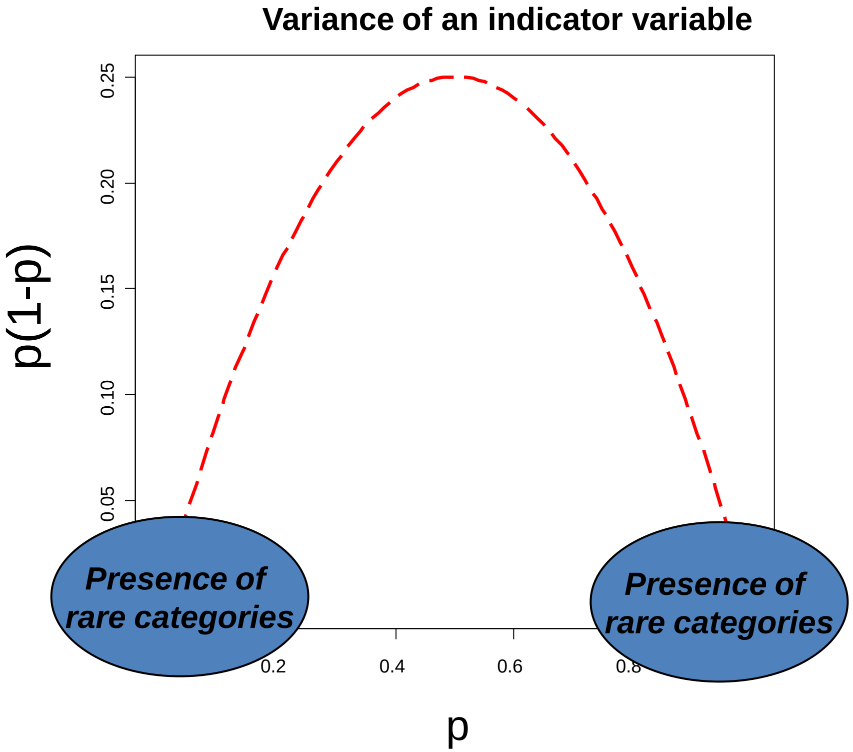
$$X^T\left(\sum_{j=1}^{m} X_j\left(X_j^T X_j\right)^{-1} X_j^T\right)X = X^T P X$$

$$\text{with } P = \sum_{j=1}^{m} X_j\left(X_j^T X_j\right)^{-1} X_j^T$$

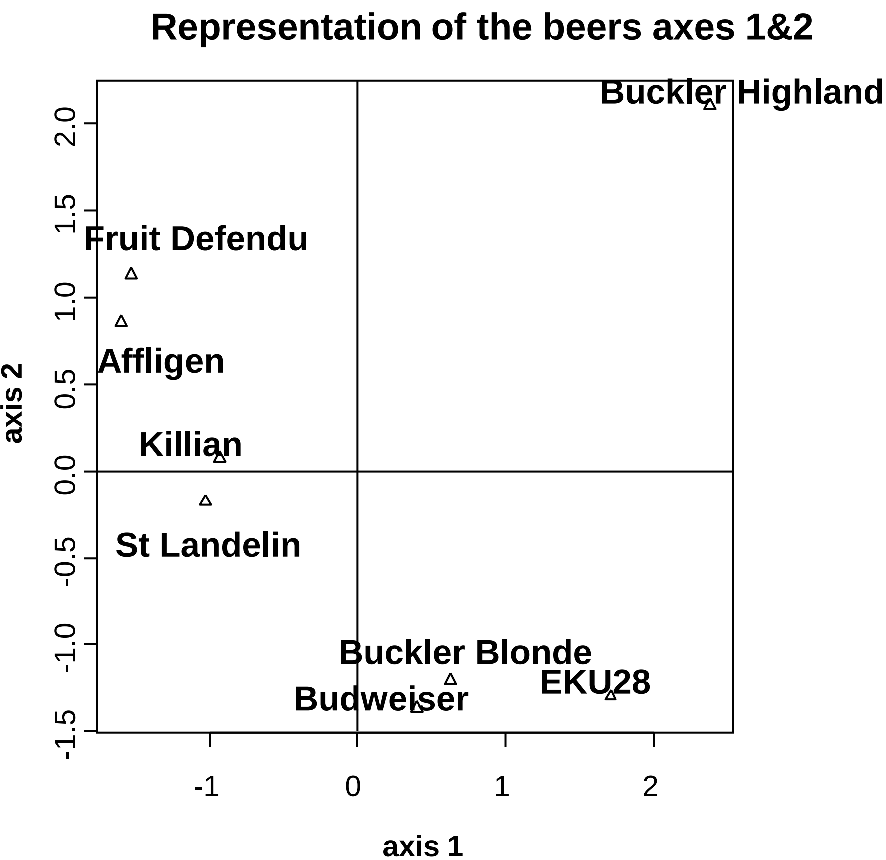
- **Subsequent z variables can be sought following the same strategy, under orthogonality constraints.**

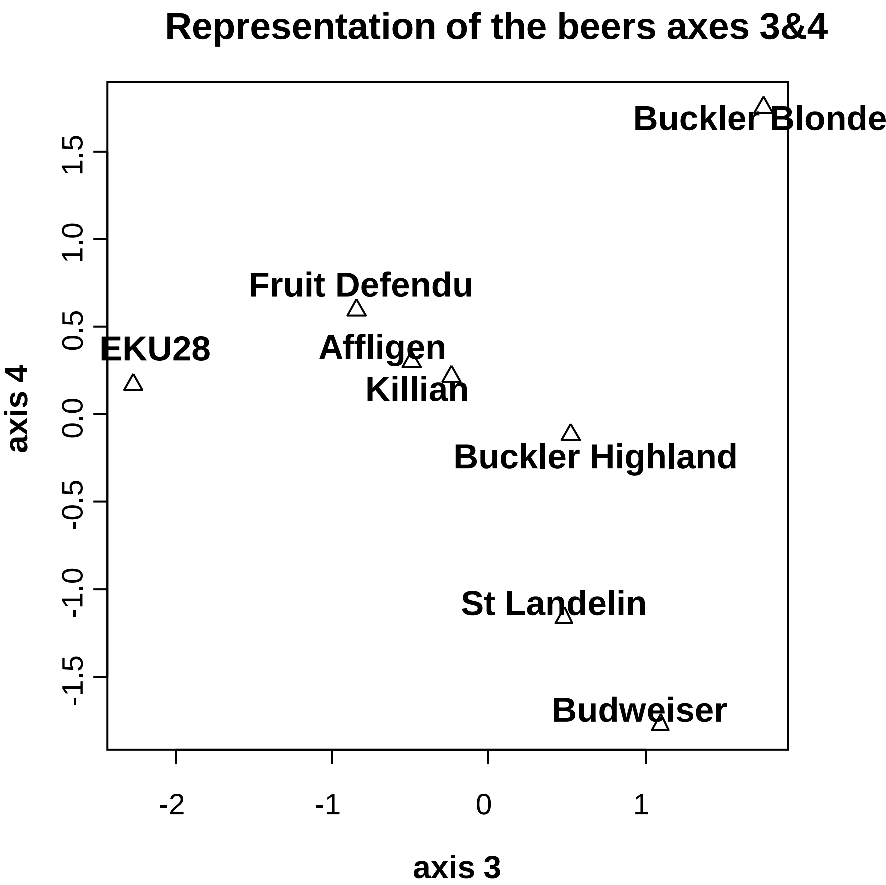
The rationale behind the method of analysis

- **In addition to investigating the relationships between the categorical variables, we take account of the variances of the indicator variables.**
- **VAR(Indicator)=p*(1-p)**



Alternative method applied to beer data





A continuum approach

- **MCA**

  z=Xu with u eigenvetor of : $$(X^T X)^{-1} X^T PX$$

- **Alternative method**

  z=Xu with u eigenvetor of :

  $$X^T PX$$

- **Regularized MCA:**

  z=Xu with u eigenvetor of :

  $$\left\langle (1-\lambda) X^T X + \lambda I \right\rangle^{-1} X^T PX$$

continuum approach and Ridge Regularization

The eigenvectors of :

$$\left\langle (1-\lambda) X^T X + \lambda I \right\rangle^{-1} X^T P X$$

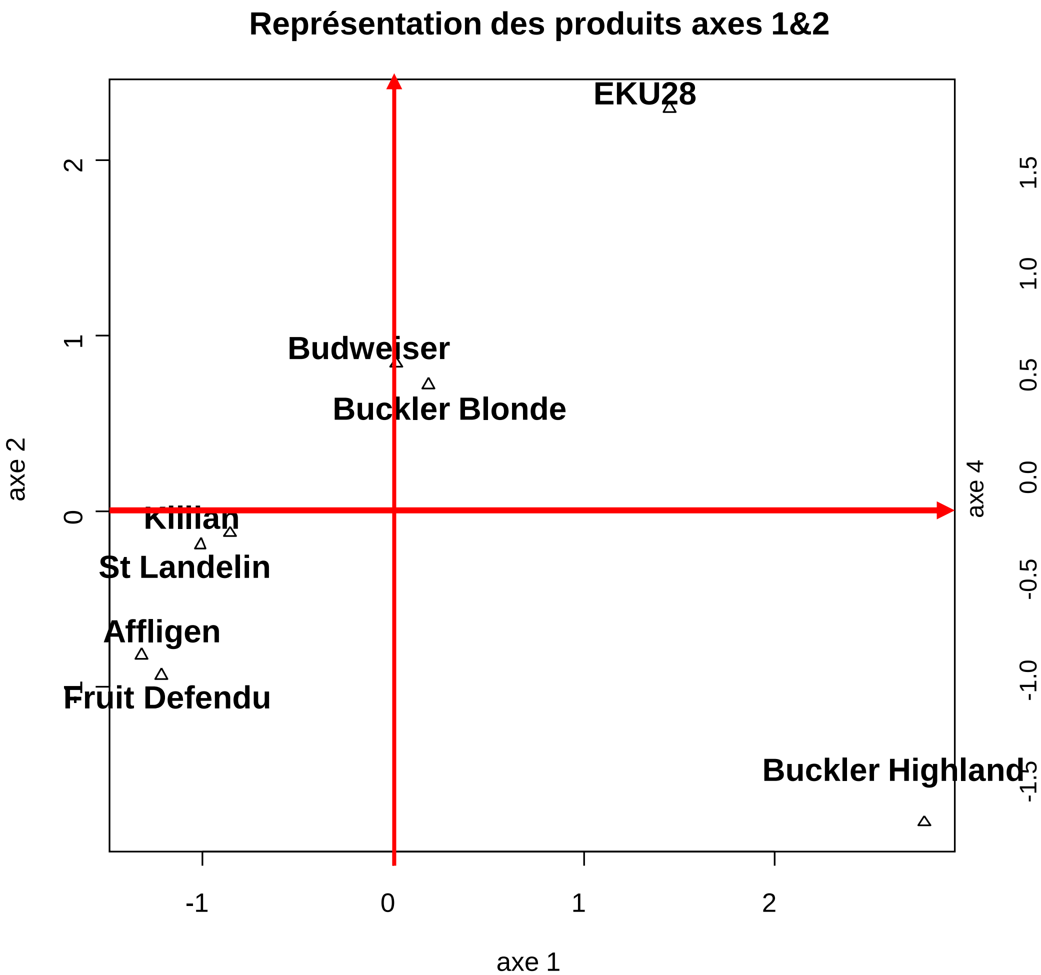
are also eigenvectors of :

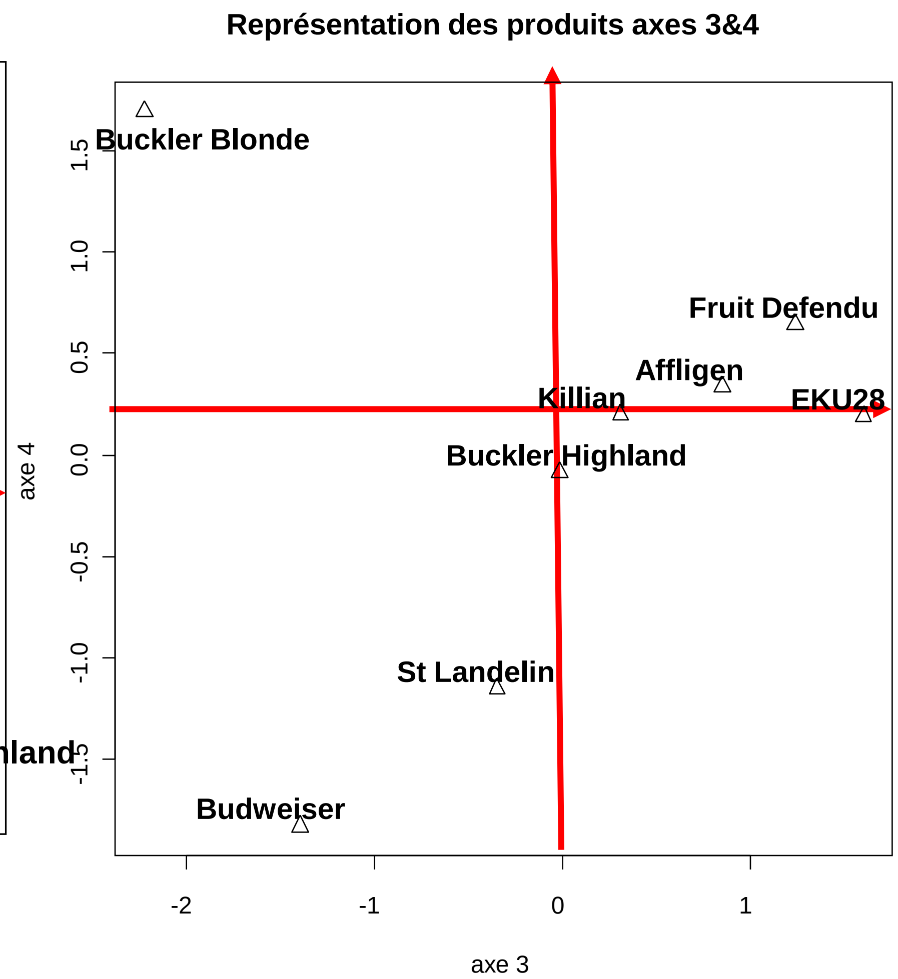
$$\left\langle X^T X + kI \right\rangle^{-1} X^T P X$$

$$with \quad k = \frac{\lambda}{(1-\lambda)}$$

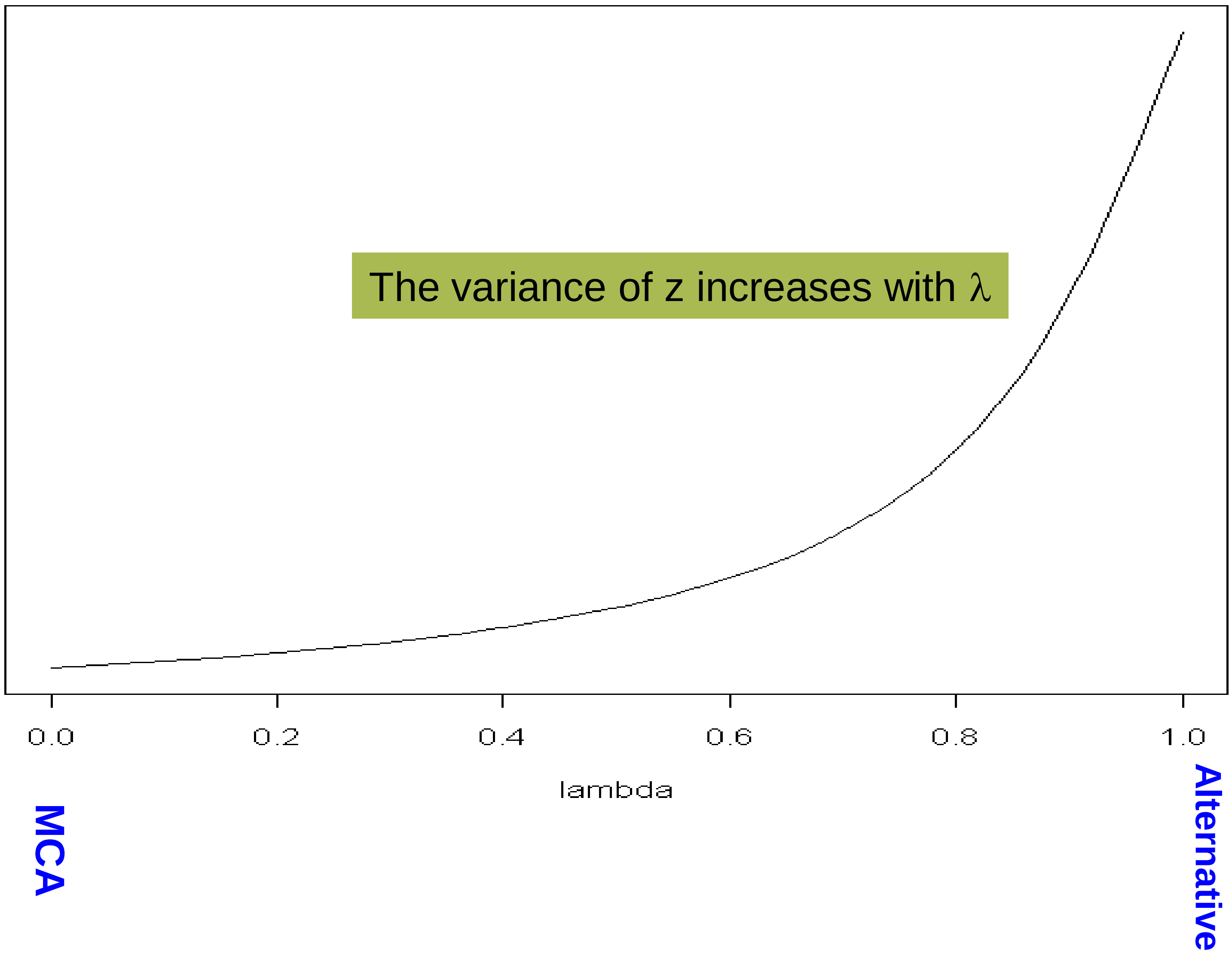
→ Ridge regularization

RMCA (lambda=0.95)

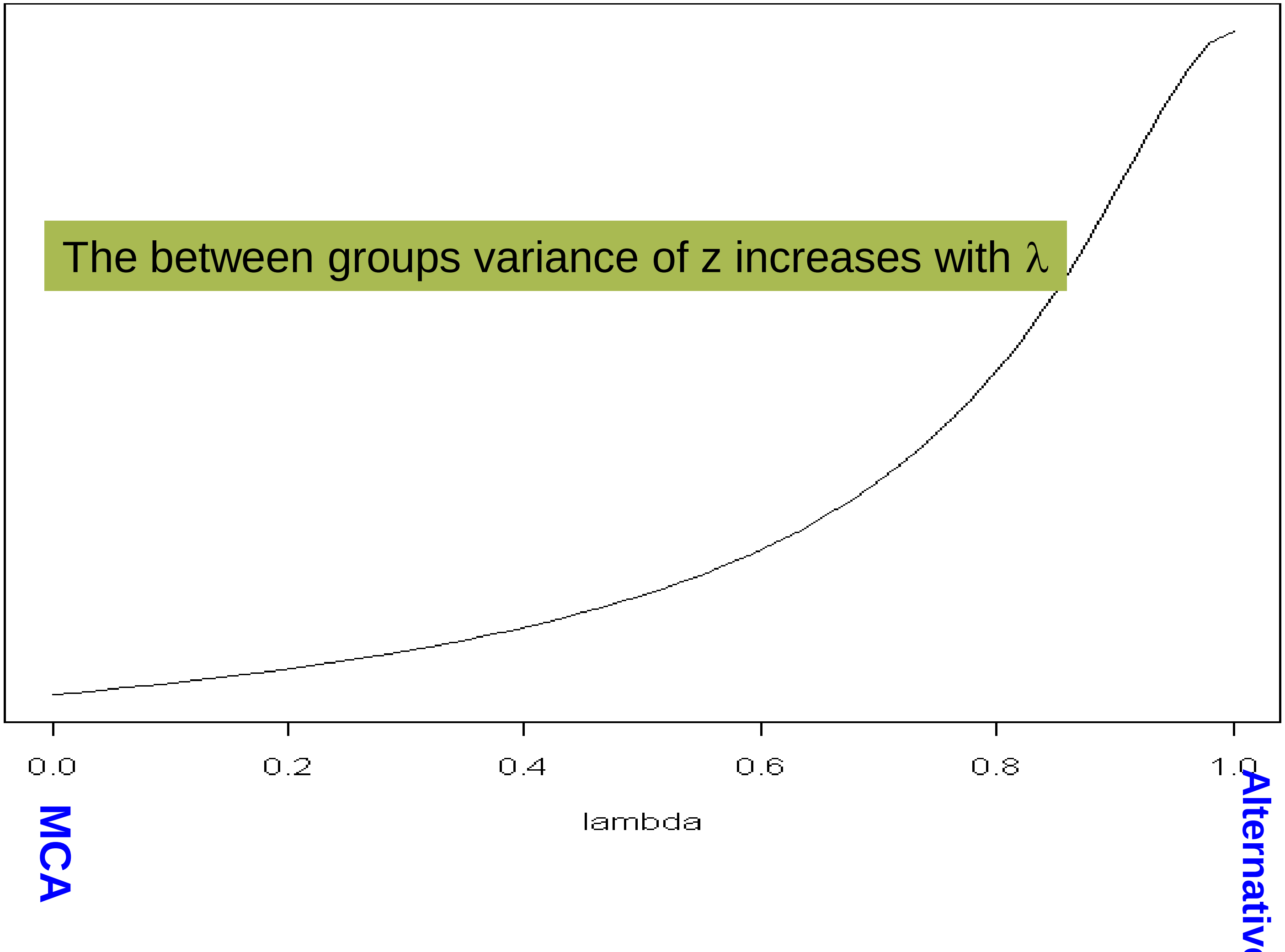




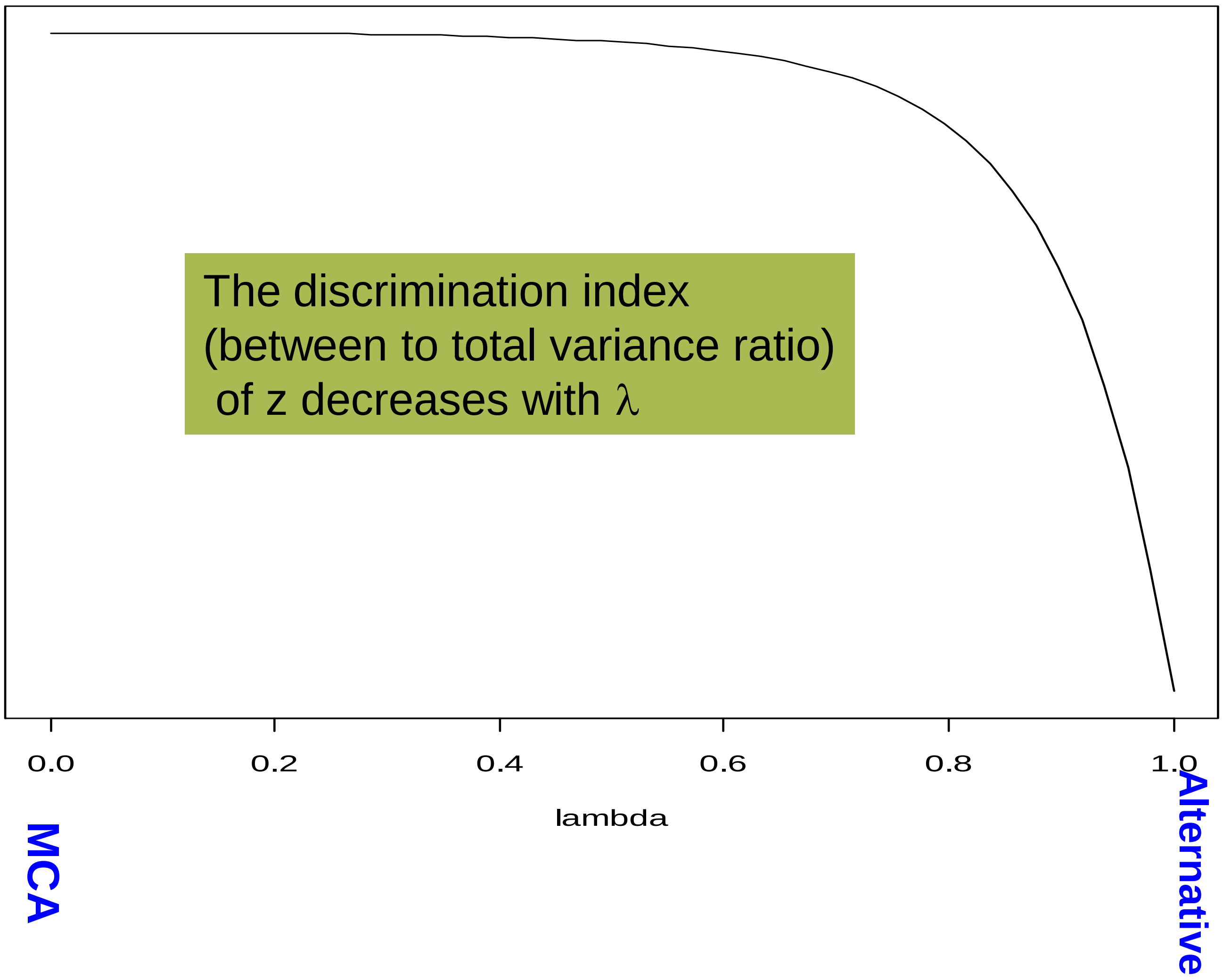
Property 1 illustrated on beer data



Property 2 illustrated on beer data



Property 3 illustrated on beer data



Conclusion

- Proposition of an alternative method that handles the problem of rare categories
- Further research work is needed to investigate this alternative method.
- Proposition of a continuum approach whose end points are MCA and the alternative method.
- This approach enjoys interesting properties and can easily be extended to the framework of Generalized Canonical Correlation Analysis.
- See how it relates to Regularized MC by Takane and Hwang.

TRUGAREZ!

Co-occurrence matrix

Beers		Beers							
		1	2	3	4	5	6	7	8
	1	10	1	1	5	6	0	8	0
	2	1	10	3	2	5	0	0	1
	3	1	3	10	2	2	0	0	0
	4	5	2	2	10	5	0	5	1
	5	6	5	2	5	10	0	4	0
	6	0	0	0	0	0	10	0	0
	7	8	0	0	5	4	0	10	0
	8	0	1	0	1	0	0	0	10