On the comparison and representation of fuzzy partitions

François Bavaud, University of Lausanne

Abstract

A fuzzy partition assigns to each among n objects a distribution over a categories. Elementary linear algebraic methods permit to introduce and investigate concepts and properties such as a) variance and inertia decomposition; b) coarse- and fine-graining (nestedness); c) iteration of fuzzy partitions; d) stability of a group in regard to another partition; e) (euclidean embeddable) dissimilarities between objects; f) (euclidean embeddable) dissimilarities between partitions. Unweighted (R) or weighted (T, P) object similarities are further investigated, and found to be related to the chi-square as well as to the indices of Gini, variety and Mirkin-Cherny-Rand. Weighted versions T and P differ for fuzzy partitions, allowing various non-equivalent constructions characterizing differing aspects of fuzzy partitions and possessing no formal analog at the crisp level[1].

1 Introduction and notations

Partitioning (deterministically) n objects consists in assigning each object i to a group j, among a possible groups; see e.g. Saporta pp. 210-224 (1990) or Mirkin pp. 229-246 (1996) for a classical, formal approach. A *fuzzy* partition consists of a probabilistic assignment of object i to group j, specified with z_{ij} = "probability that object i belongs to group j", obeying $z_{ij} \geq 0$, $\sum_{j=1}^{a} z_{ij} = 1$ and $\sum_{i=1}^{n} z_{ij} > 0$ (absence of empty groups); see e.g. Bezdek (1981) for a presentation of the fuzzy context.

Elementary algebra allows characterizing the combination, iteration or nesting of fuzzy partitions; associated operators, whose projective or Markov-like properties are exploited, possess simple interpretations in terms of dissimilarities between objects, yielding in turn euclidean embeddable dissimilarities between objects and even between partitions themselves.

The present general framework suggests a certain view of the multivariate analysis of fuzzy partitions (=fuzzy categorical variables), that is of *multiple fuzzy correspondence analysis.*

2 Membership matrices

Definition 1 A (fuzzy) partition $\mathcal{A}$ of a set of n objects in a groups is defined by a $(n \times a)$

[1]The work has benefited from stimulating discussions with M.Rajman in the framework of the joint UNIL-EPFL "Clavis" project (2001).

(fuzzy) *membership* or *indicator* matrix such that $z_{ij}^{\mathcal{A}} \geq 0$, $\sum_{j=1}^{a} z_{ij}^{\mathcal{A}} = 1$ for all $i = 1, \ldots, n$ and $n_j^{\mathcal{A}} := \sum_{i=1}^{n} z_{ij}^{\mathcal{A}} > 0$ for all $j = 1, \ldots, a$.

Definition 2 a) A deterministic or *crisp* partition obtains when $z_{ij} = 1$ or $z_{ij} = 0$ for all i, j, or equivalently $z_{ij}^2 = z_{ij}$. In the case of a crisp partition, $j(i)$ will denote the group to which i belongs.

b) A partition is said to be *full* if $\mathrm{Rank}(Z) = a$, and *defective* if $\mathrm{Rank}(Z) < a$.

Crisp partitions are full (since $n_j > 0$). The *uniform partition* in a groups $\mathcal{U}(a)$ is defined by the $(n \times a)$ matrix $z_{ij}^{\mathcal{U}(a)} = \frac{1}{a}$ for all i and $j = 1, \ldots, a$. Uniform partitions are defective for $a \geq 2$; the full case $a = 1$ defines the *one-group partition* $\mathcal{O} = \mathcal{U}(1)$, with associated $(n \times 1)$ membership matrix $z_{i1}^{\mathcal{O}} = 1$. The *n-groups partition* $\mathcal{N}$ is defined by the $(n \times n)$ identity matrix $z_{ij}^{\mathcal{N}} = \delta_{ij}$ (or a permutation of it).

2.1 Variance decomposition

Let X be a numerical variable with scores x_i, $i = 1, \ldots, n$. Define the (fuzzy) average for the j-th group as $\bar{x}_j := \sum_{i=1}^{n} \frac{z_{ij}}{n_j} x_i$, and the total average as $\bar{x} = \sum_{j=1}^{a} f_j \, \bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_i$, where $f_j := \frac{n_j}{n}$. Define the total, within- and between-groups variances as

$$\mathrm{var}(x) := \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad \mathrm{var}_W(x) := \frac{1}{n} \sum_{j=1}^{a} \sum_{i=1}^{n} z_{ij} (x_i - \bar{x}_j)^2 \qquad \mathrm{var}_B(x) := \sum_{j=1}^{a} f_j (\bar{x}_j - \bar{x})^2 \tag{1}$$

Then the (fuzzy) variance decomposition formula $\mathrm{var}(x) = \mathrm{var}_W(x) + \mathrm{var}_B(x)$ holds. In particular, for fixed values $\{x_i\}$, $\mathrm{var}_W(x)$ is maximum for $\mathcal{A} = \mathcal{U}(a)$ (for any a), and minimum for $\mathcal{A} = \mathcal{O}$.

2.2 Connected components

Define the $(a \times a)$ matrices $B = (b_{jj'})$ and $N = (n_{jj'})$ as

$$B := Z'Z \quad \text{i.e.} \quad b_{jj'} := \sum_i z_{ij} z_{ij'} \qquad\qquad N := \mathrm{diag}(\mathbf{1}'Z) \quad \text{i.e.} \quad n_{jj'} := \delta_{jj'} \, n_j \tag{2}$$

where $\mathbf{1}$ is the $(n \times 1)$ unit vector. Note that $Z\,\mathit{1} = \mathbf{1}$ and $Z'\mathbf{1} = N\mathit{1}$, where $\mathit{1}$ is the $(m \times 1)$ unit vector. Also, B^{-1} exists iff $\mathcal{A}$ is full.

$b_{jj'} \geq 0$ constitutes an index of overlapping between groups j and j' and measures their common sharing of objects. Distinct groups j and j' with $b_{jj'} > 0$ are said to be *adjacent*. Distinct groups j and j' related by a path $b_{jk_1} \, b_{k_1 k_2} \ldots b_{k_l j'} > 0$ of adjacent groups are *connected*. A set of connected groups constitutes an (irreducible) *component*, indexed by $J = 1, \ldots, c(\mathcal{A})$, where $c(\mathcal{A}) \leq m$ is the number of irreducible components of the partition $\mathcal{A}$, or, equivalently, the number of irreducible blocks of Z. One has:

$$
\begin{array}{rcccc}
c(\mathcal{A}) = m & \Leftrightarrow & \mathcal{A} \text{ is crisp} & \Leftrightarrow & B = N \\
\text{Rank}(Z) = m & \Leftrightarrow & \mathcal{A} \text{ is full} & \Leftrightarrow & B^{-1} \text{ exists}
\end{array}
$$

2.3 Iterated partitions

In view of the previous section, the matrix $G := N^{-1}B$ is the identity iff is $\mathcal{A}$ crisp. In general, G generates *iterated partitions*:

Definition 3 The r-th iterated membership $Z^{(r)}$ defining partition $\mathcal{A}^{(r)}$ obtains as

$$
Z^{(r)} := Z\, G^{r-1} \qquad G := N^{-1}B = (g_{jj'}) \qquad g_{jj'} = \frac{1}{n_j}\sum_{i=1}^{n} z_{ij}\, z_{ij'} \tag{3}
$$

Indeed, identity $G\,1 = 1$ ensures the normalization $Z^{(r)}\,1 = 1$: that is, $Z^{(r)}$ is the membership matrix associated to some (fuzzy) partition denoted $\mathcal{A}^{(r)}$.

$G\,1 = 1$ with $g_{jj'} \geq 0$ also shows G to be the $(a \times a)$ transition matrix of a Markov chain among groups $j = 1, \ldots, a$: $g_{jj'}$ is the probability that, starting from group j in which one selects an individual i, one precisely gets group j' when further selecting a group from individual i. Identity $\mathbf{1}'ZN^{-1}Z'Z = \mathbf{1}'Z$ ensures $n_j^{(r)} := \sum_i z_{ij}^{(r)} = n_j$: the group sizes are thus unchanged by iteration.

G is doubly stochastic, and made up of $J = 1, \ldots, c(\mathcal{A})$ irreducible doubly stochastic matrices $G^{(J)}$, each with stationary distribution $f_j^{(J)} = n_j/n_J$ where $n_J := \sum_{j \in J} n_j$. Iterating partitions mixes the objects i among the various classes j of the *same* connected component J; for instance, $z_{ij}^{(2)} = \sum_{i'j'} z_{ij'}\, z_{i'j'}\, z_{i'j}/n_{j'}$. In the limit $r \to \infty$, objects inside the same component J possess the same group membership:

$$
g_{jj'}^{(\infty)} = \frac{n_{j'}\, I(j' \in J(j))}{n_{J(j)}} \qquad \text{implying} \qquad z_{ij}^{(\infty)} = \frac{n_j\, I(i \in J(j))}{n_{J(j)}} \tag{4}
$$

where $I(E)$ denotes the characteristic function for event E, and $J(j)$ denotes the component to which group j belongs. Partition $\mathcal{A}^{(\infty)}$ thus obtains by

1. first assigning individuals i to their component $J(i)$; we denote this partition as $\mathcal{A}^{(0)}$, with $(n \times c(\mathcal{A}))$ associated membership matrix $z_{iJ}^{(0)} = I(i \in J)$
2. then choosing group $j \in J(i)$ with probability $n_j/n_{J(i)} = f_j/f_{J(i)}$; the $(c(\mathcal{A}) \times a)$ membership matrix associated to this *component-group* partition is $z_{Jj}^{\text{cg}} = \frac{n_j}{n_J} I(j \in J)$.

By construction

$$
Z^{(\infty)} = Z^{(0)}\, Z^{\text{cg}} \qquad Z^{(0)} = Z\, Z^{\text{gc}} \qquad \text{where} \quad z_{iJ}^{\text{gc}} := I(j \in J) \tag{5}
$$

Partition $\mathcal{A}^{(\infty)}$ is defective iff $\mathcal{A}$ is fuzzy (since $\text{Rank}(Z^{(\infty)}) = c(\mathcal{A}) < a$), and full iff $\mathcal{A}$ is crisp. Crisp partitions are characterized by $g_{jj'} = g_{jj'}^{(r)} = \delta_{jj'}$ and $z_{ij} = z_{ij}^{(r)} = I(i \in j)$.

Example 1 Consider the fuzzy partition $\mathcal{A}$ of $n = 5$ objects in $a = 4$ classes with

$$Z^{\mathcal{A}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.2 & 0.8 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0.2 & 0.8 \end{pmatrix} \qquad N = \begin{pmatrix} 1.2 & 0 & 0 & 0 \\ 0 & 1.8 & 0 & 0 \\ 0 & 0 & 0.8 & 0 \\ 0 & 0 & 0 & 1.2 \end{pmatrix} \qquad B = \begin{pmatrix} 1.04 & 0.16 & 0 & 0 \\ 0.16 & 1.64 & 0 & 0 \\ 0 & 0 & 0.4 & 0.4 \\ 0 & 0 & 0.4 & 0.8 \end{pmatrix}$$

$$G = \begin{pmatrix} \frac{13}{15} & \frac{2}{15} & 0 & 0 \\ \frac{4}{45} & \frac{41}{45} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix} \qquad G^{(\infty)} = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0.4 & 0.6 \end{pmatrix} \qquad Z^{(\infty)} = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0.4 & 0.6 \end{pmatrix}$$

3 Object comparisons

3.1 Object similarities

Let $S = (s_{ii'})$ denote a general $(n \times n)$ similarity matrix between objects, obeying $s_{ii'} \geq 0$, $s_{ii'} = s_{i'i}$ and $s_{ii'} \leq \sqrt{s_{ii}\, s_{i'i'}}$. Three natural candidates for S are provided by the $(n \times n)$ matrices $R := ZZ'$, $T := ZN^{-1}Z'$ and (assuming the partition to be full, that is B^{-1} exists) $P := ZB^{-1}Z'$, namely[2]

$$r_{ii'} := \sum_{j=1}^{a} z_{ij}\, z_{i'j} \qquad t_{ii'} := \sum_{j=1}^{a} \frac{z_{ij}\, z_{i'j}}{n_j} \qquad p_{ii'} := \sum_{j,j'=1}^{a} z_{ij}\, b_{jj'}^{(-1)}\, z_{i'j'} \tag{6}$$

- $R = (r_{ii'})$ (the *relation* similarity matrix) yields, for a crisp classification, the indicator matrix of the relation "objects i and i' belong to the same group".
- $T = (t_{ii'})$ (the *transition* similarity matrix) satisfies $\sum_{i'} t_{ii'} = 1$: it is thus the transition matrix among objects of a doubly stochastic Markov chain; $t_{ii'}$ is the probability to jump from object i to object i' when first selecting class j with probability z_{ij} and then selecting object i' inside class j with probability $z_{i'j}/n_j$.
- $P = (p_{ii'})$(the *projection* similarity matrix) is a projection matrix (see theorem (1)). Also, $P = T$ iff $\mathcal{A}$ is crisp; in other words, similarity matrices can be made simultaneously markovian and projective for crisp partitions only. Note that $p_{ii'} \geq 0$ can be violated (since B^{-1} possesses negative components); however, $|p_{ii'}| \leq \sqrt{p_{ii}\, p_{i'i'}}$ holds.

Theorem 1 *a) T is a Markov transition matrix, with stationary uniform distribution $\pi_i = 1/n$; its iterate obeys $T^2 = T$ iff $\mathcal{A}$ is crisp (that iff $T = P$ as well).*

[2]Equations (2) and (6) are somewhat reminiscent of the "Burt-Condorcet" duality in multiple correspondence analysis (see e.g. Marcotorchino (2000)). Recall however the latter to refer to $p \geq 2$ *crisp partitions*, rather than *one fuzzy partition* as in the present case.

b) In general, however, $1 \leq \mathrm{Tr}(T^2) \leq \mathrm{Tr}(T) \leq a$, with $\mathrm{Tr}(T) = 1$ iff $\mathcal{A}$ is the uniform partition $\mathcal{U}(a)$ (for any a), and $\mathrm{Tr}(T) = a$ iff $\mathcal{A}$ is crisp, in which case $T = P$.

c) P exists iff $\mathcal{A}$ is full, in which case $P^2 = P$ and $\mathrm{Tr}(P) = a$.

Proof of theorem 1 a) obtains from c) below; recall $T = P$ in the crisp case.

b) by definition, $\mathrm{Tr}(T) = \sum_{ij} \frac{z_{ij}^2}{n_j}$. $\mathrm{Tr}(T) = a$ holds as a consequence of $z_{ij}^2 \leq z_{ij}$ with equality iff $\mathcal{A}$ is crisp. $\mathrm{Tr}(T) = 1$ obtains from Jensen's inequality $\frac{1}{n}\sum_i z_{ij}^2 \geq \{\frac{1}{n}\sum_i z_{ij}\}^2$ with equality iff $\mathcal{A}$ is the uniform partition.

Inequality $(z_{ij} - z_{i'j})^2 z_{ij'} z_{i'j'} \geq 0$ holds in general, while $(z_{ij} - z_{i'j})^2 z_{ij'} z_{i'j'} = 0$ for all i, i', j, j' iff $\mathcal{A}$ is crisp. Summing the latter yields

$$\mathrm{Tr}(T^2) = \sum_{ii'jj'} \frac{z_{ij} z_{i'j} z_{ij'} z_{i'j'}}{n_j \, n_{j'}} \leq \sum_{ii'jj'} \frac{z_{ij}^2 z_{ij'} z_{i'j'}}{n_j \, n_{j'}} = \sum_{ij} \frac{z_{ij}^2}{n_j} = \mathrm{Tr}(T)$$

which demonstrates that $T^2 \neq T$ if $\mathcal{A}$ is not crisp. On the other hand, $T = P$ if $\mathcal{A}$ is crisp, and thus $T^2 = T$.

c) $P^2 = ZB^{-1}Z'ZB^{-1}Z' = ZB^{-1}BB^{-1}Z' = ZB^{-1}Z' = P$; also, $\mathrm{Tr}(P) = \mathrm{Tr}(ZB^{-1}Z') = \mathrm{Tr}(B^{-1}Z'Z) = \mathrm{Tr}(B^{-1}B) = \mathrm{Tr}\, I = a$.

3.2 Iterated object similarities

Higher order similarities can be constructed as $R^{(r)} := Z^{(r)}(Z^{(r)})'$, $T^{(r)} := Z^{(r)}(N^{(r)})^{-1}(Z^{(r)})'$ and (for a full partition) $P^{(r)} := Z^{(r)}\,(B^{(r)})^{-1}\,(Z^{(r)})'$, where $Z^{(r)} := Z\,G^{r-1}$, $B^{(r)} := (Z^{(r)})'\,Z^{(r)}$ and $N^{(r)} := \mathrm{diag}(\mathbf{1}'Z^{(r)})$.

Theorem 2 For $r \geq 0$, $T^{(r)} = T^{2r-1}$ and (for a full partition) $P^{(r)} = P$.

Proof of theorem 2

$$\begin{aligned} P^{(r+1)} &= Z^{(r+1)}\,(B^{(r+1)})^{-1}\,(Z^{(r+1)})' = Z^{(r)}N^{-1}B[BN^{-1}B^{(r)}N^{-1}B]^{-1}BN^{-1}(Z^{(r)})' = \\ &= Z^{(r)}N^{-1}BB^{-1}N(B^{(r)})^{-1}NB^{-1}BN^{-1}(Z^{(r)})' = Z^{(r)}\,(B^{(r)})^{-1}\,(Z^{(r)})' = P^{(r)} \end{aligned}$$

Using $N^{(r)} = N$, identity $T^{(r)} = T^{2r-1}$ is proved similarly.

3.3 Object distances

Matrices R, T and P are three instances of positive-definite similarity matrices $S = (s_{ii'})$ between objects i and i', from which a *squared euclidean distance* can be constructed as $D^S_{ii'} := (d^S_{jj'})^2 = s_{ii} + s_{i'i'} - 2s_{ii'}$ (Schoenberg 1935; Gower 1982). Explicitly

$$D^R_{ii'} = \sum_j (z_{ij} - z_{i'j})^2 \qquad D^T_{ii'} = \sum_j \frac{(z_{ij} - z_{i'j})^2}{n_j} \qquad D^P_{ii'} = \sum_{jj'} (z_{ij} - z_{i'j}) b^{(-1)}_{jj'} (z_{ij'} - z_{i'j'}) \tag{7}$$

Theorem 3 : for a crisp partition:

$$r_{ii'} = 1 \quad t_{ii'} = p_{ii'} = \frac{1}{n_j} \qquad D^R_{ii'} = D^T_{ii'} = D^P_{ii'} = 0 \qquad \text{for } i, i' \in j$$

$$r_{ii'} = t_{ii'} = p_{ii'} = 0 \qquad D^R_{ii'} = 2 \quad D^T_{ii'} = D^P_{ii'} = \tfrac{1}{n_j} + \tfrac{1}{n_{j'}} \qquad \text{for } i \in j,\ i' \in j' \text{ with } j \neq j'$$

In particular, $r^{\mathcal{N}}_{ii'} = t^{\mathcal{N}}_{ii'} = p^{\mathcal{N}}_{ii'} = \delta_{ii'}$; $r^{\mathcal{O}}_{ii'} = 1$ and $t^{\mathcal{O}}_{ii'} = p^{\mathcal{O}}_{ii'} = \frac{1}{n}$.

Proof of theorem 3 : straightforward.

Let $a_j \geq 0$ with $\sum_j a_j = 1$ be the membership profile of some object a; for instance, $g_j = \frac{1}{n}\sum_i z_{ij} = \frac{n_j}{n}$ represents the membership profile of the gravity center g. Then squared distances D^S_{ia} can be defined by the substitution $z_{i'j} \to a_j$ in (7). Define

$$I^S_2 =: \frac{1}{2n^2}\sum_{i,i'} D^S_{ii'} \quad \text{(pair inertia)} \qquad I^S_1(a) =: \frac{1}{n}\sum_i D^S_{ia} \quad \text{(central inertia with center } a\text{)} \tag{8}$$

Then, for $S = R, T$ or P,

$$I^S_1(a) = I^S_1(g) + D^S_{ag} \quad \text{(strong Huygens principle)} \qquad I^S_2 = I^S_1(g) \quad \text{(weak Huygens principle)} \tag{9}$$

(see e.g. Bavaud (2002)). The pair inertia $I^S_2 = I^S_1(g)$ constitutes an index of classificatory diversity; for crisp partitions $\mathcal{A}$, one gets $I^R_2 = \sum_{j=1}^a f_j(1-f_j)$ (Gini diversity index) and $I^T_2 = I^P_2 = (a-1)/n$.

As it it well known (classical MDS), coordinates $x^S_{i\alpha}$ realizing an euclidean representation of the objects $i = 1, \ldots, n$ in dimensions $\alpha = 1, \ldots, a-1$ (that is satisfying $D^S_{ii'} = \sum_\alpha (x^S_{i\alpha} - x^S_{i'\alpha})^2$) can be obtained as $x^S_{i\alpha} := \sqrt{\lambda^S_\alpha}\, u^S_{i\alpha}$, where the λ^S_α are the eigenvalues and the $u^S_{i\alpha}$ the eigenvectors occurring in the spectral decomposition $S = U^S \Lambda^S (U^S)'$.

Example 1, continued: the (5×5) corresponding similarity matrices are

$$R = \begin{pmatrix} 1 & .2 & 0 & 0 & 0 \\ .2 & .68 & .8 & 0 & 0 \\ 0 & .8 & 1 & 0 & 0 \\ 0 & 0 & 0 & .52 & .44 \\ 0 & 0 & 0 & .44 & .68 \end{pmatrix} \quad T = \begin{pmatrix} .83 & .17 & 0 & 0 & 0 \\ .17 & .39 & .44 & 0 & 0 \\ 0 & .44 & .56 & 0 & 0 \\ 0 & 0 & 0 & .58 & .42 \\ 0 & 0 & 0 & .42 & .58 \end{pmatrix} \quad P = \begin{pmatrix} .98 & .12 & -.10 & 0 & 0 \\ .12 & .40 & .48 & 0 & 0 \\ -.10 & .48 & .62 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

with associated squared distances *between objects*

$$D^R = \begin{pmatrix} 0 & 1.28 & 2 & 1.52 & 1.68 \\ 1.28 & 0 & .08 & 1.2 & 1.36 \\ 2 & .08 & 0 & 1.52 & 1.68 \\ 1.52 & 1.2 & 1.52 & 0 & .32 \\ 1.68 & 1.36 & 1.68 & .32 & 0 \end{pmatrix} \quad D^T = \begin{pmatrix} 0 & .89 & 1.39 & 1.42 & 1.42 \\ .89 & 0 & .06 & .97 & .97 \\ 1.39 & .06 & 0 & 1.14 & 1.14 \\ 1.42 & .97 & 1.14 & 0 & .33 \\ 1.42 & .97 & 1.14 & .33 & 0 \end{pmatrix} \quad D^P = \begin{pmatrix} 0 & 1.14 & 1.79 & 1.98 & 1.98 \\ 1.14 & 0 & .07 & 1.40 & 1.40 \\ 1.79 & .07 & 0 & 1.62 & 1.62 \\ 1.98 & 1.40 & 1.62 & 0 & 2 \\ 1.98 & 1.40 & 1.62 & 2 & 0 \end{pmatrix}$$

Spectral decomposition of S yields the corresponding (5×4) coordinates $X^S = (x^S_{i\alpha})$:

$$X^R = \begin{pmatrix} .24 & .97 & 0 & 0 \\ .82 & 0 & 0 & 0 \\ .97 & -.24 & 0 & 0 \\ 0 & 0 & .66 & -.30 \\ 0 & 0 & .79 & .25 \end{pmatrix} \quad X^T = \begin{pmatrix} .58 & -.71 & 0 & 0 \\ .58 & .24 & 0 & 0 \\ .58 & .47 & 0 & 0 \\ 0 & 0 & -.71 & .29 \\ 0 & 0 & -.71 & -.29 \end{pmatrix} \quad X^P = \begin{pmatrix} -.12 & .98 & 0 & 0 \\ .61 & .19 & 0 & 0 \\ .79 & -.24 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

4 Nested partitions: coarser and finer

Definition 4 Partition $\mathcal{B}$ (defined by the $(n \times b)$ membership matrix $Z^{\mathcal{B}}$) is *coarser* than partition $\mathcal{A}$ (defined by the $(n \times a)$ membership matrix $Z^{\mathcal{A}}$), or, equivalently, $\mathcal{A}$ is *finer* than $\mathcal{B}$, noted $\mathcal{B} \leq \mathcal{A}$, if $Z^{\mathcal{B}} = Z^{\mathcal{A}}\, W^{\mathcal{AB}}$ where $W^{\mathcal{AB}} = (w^{\mathcal{AB}}_{jk})$ is a $(a \times b)$ *class membership matrix* such that $w^{\mathcal{AB}}_{jk} \geq 0$ and $\sum_{k=1}^{b} w^{\mathcal{AB}}_{jk} = 1$.

Theorem 4 a) The relation "$\mathcal{B} \leq \mathcal{A}$" is a partial order

b) its minimal element is the *one-group partition* $\mathcal{O}$

c) its maximal element is the *n-groups partition* $\mathcal{N}$

d) if $\mathcal{B} \leq \mathcal{A}$ (with $\mathcal{A}$ and $\mathcal{B}$ both full) then $P^{\mathcal{A}}P^{\mathcal{B}} = P^{\mathcal{B}}P^{\mathcal{A}} = P^{\mathcal{B}}$.

Proof of theorem 4 a) By definition, $\mathcal{A} \leq \mathcal{A}$ (with $w^{\mathcal{AA}}_{jk} = \delta_{jk}$). Also, $\mathcal{B} \leq \mathcal{A}$ and $\mathcal{C} \leq \mathcal{B}$ entail $\mathcal{C} \leq \mathcal{A}$ (with $W^{\mathcal{AC}} = W^{\mathcal{AB}}\, W^{\mathcal{BC}}$).

b) for any $\mathcal{A}$, $Z^{\mathcal{O}} = Z^{\mathcal{A}}\, W^{\mathcal{AO}}$ with $w^{\mathcal{AO}}_{j} = 1$ for all $j = 1, \ldots, a$.

c) for any $\mathcal{B}$, $Z^{\mathcal{B}} = Z^{\mathcal{N}}\, W^{\mathcal{NB}}$ with $w^{\mathcal{NB}}_{ij} = z^{\mathcal{B}}_{ij}$.

d) $Z^{\mathcal{A}}W^{\mathcal{AB}} = Z^{\mathcal{B}}$ and $B^{\mathcal{A}} = (Z^{\mathcal{A}})'Z^{\mathcal{A}}$ yield

$$P^{\mathcal{A}}P^{\mathcal{B}} = Z^{\mathcal{A}}(B^{\mathcal{A}})^{-1}(Z^{\mathcal{A}})'Z^{\mathcal{B}}(B^{\mathcal{B}})^{-1}(Z^{\mathcal{B}})' =$$
$$= Z^{\mathcal{A}}\underbrace{(B^{\mathcal{A}})^{-1}(Z^{\mathcal{A}})'Z^{\mathcal{A}}}_{I}W^{\mathcal{AB}}(B^{\mathcal{B}})^{-1}(Z^{\mathcal{B}})' = Z^{\mathcal{B}}(B^{\mathcal{B}})^{-1}(Z^{\mathcal{B}})' = P^{\mathcal{B}}$$

Identity $P^{\mathcal{B}}P^{\mathcal{A}} = P^{\mathcal{B}}$ is demonstrated analogously.

Theorem 5 For $r \geq 1$, the sequence of partitions $\mathcal{A}^{(r)}$ associated with the iterated memberships $Z^{(r)}$ (definition 3) is decreasing (that is $\mathcal{A}^{(r+1)} \leq \mathcal{A}^{(r)}$). Its limit $\mathcal{A}^{(\infty)}$ is given by the membership matrix $Z^{(\infty)}$ defined in (4). Also, $\mathcal{A}^{(\infty)} \leq \mathcal{A}^{(0)} \leq \mathcal{A}$.

Proof of theorem 5 The first two assertions follow from $Z^{(r+1)} = Z^{(r)}\, G$ (equation (3)), where G is a $(a \times a)$ non-negative matrix obeying $\sum_{j'=1}^{a} g_{jj'} = 1$ together with properties listed in section (2.3). The last assertion is a direct consequence of (5).

5 Comparison and representation of partitions

5.1 The general case: euclidean visualization

Definition 5 Let $S^{\mathcal{A}} = (s^{\mathcal{A}}_{ii'})$, respectively $S^{\mathcal{B}} = (s^{\mathcal{B}}_{ii'})$, be the $(n \times n)$ similarity matrix associated to partition $\mathcal{A}$, respectively partition $\mathcal{B}$. The corresponding *(squared) distance* $D^S_{\mathcal{A},\mathcal{B}}$ between two partitions $\mathcal{A}$ and $\mathcal{B}$ is defined as

$$D^S_{\mathcal{A},\mathcal{B}} := \sum_{ii'} (s^{\mathcal{A}}_{ii'} - s^{\mathcal{B}}_{ii'})^2 = \mathrm{Tr}(S^{\mathcal{A}} - S^{\mathcal{B}})^2 = \mathrm{Tr}((S^{\mathcal{A}})^2) + \mathrm{Tr}((S^{\mathcal{B}})^2) - 2\mathrm{Tr}(S^{\mathcal{A}}\, S^{\mathcal{B}}) \qquad (10)$$

The distance $D^S_{\mathcal{A},\mathcal{B}}$ possesses all the properties of a squared euclidean distance, in particular the embeddability property. Then classical MDS applied on matrix D^S yields a low-dimensional euclidean visualization of the distances between partitions, *each partition being represented by a point* (see figure 1).

Example 2 Consider $n = 5$ objects. Define

- $\mathcal{A}$ as the fuzzy partition of example 1
- $\mathcal{B}$ as the partition $\mathcal{A}^{(0)}$ in connected components (namely $(123; 45)$)
- $\mathcal{C}$ as the crisp partition $(12; 345)$
- $\mathcal{D} \equiv \mathcal{N}$ as the n-groups partition $(1; 2; 3; 4; 5)$
- $\mathcal{E} \equiv \mathcal{O}$ as the one-group partition (12345)
- $\mathcal{F} \equiv \mathcal{A}^{(\infty)}$ as the limiting iterated partition:

$$Z^{\mathcal{A}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.2 & 0.8 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0.2 & 0.8 \end{pmatrix} \qquad Z^{\mathcal{B}} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \qquad Z^{\mathcal{C}} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

$$Z^{\mathcal{D}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \qquad Z^{\mathcal{E}} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \qquad Z^{\mathcal{F}} = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0.4 & 0.6 \end{pmatrix}$$

The corresponding dissimilarities (10) D^S *between partitions* (listed in alphabetical order) are[3]

$$D^R = \begin{pmatrix} 0 & 4.42 & 7.62 & 2.18 & 16.42 & 1.43 \\ 4.42 & 0 & 8 & 8 & 12 & 3.00 \\ 7.62 & 8 & 0 & 8 & 12 & 7.16 \\ 2.18 & 8 & 8 & 0 & 20 & 3.32 \\ 16.42 & 12 & 12 & 20 & 0 & 15.00 \\ 1.43 & 3.00 & 7.16 & 3.32 & 15.00 & 0 \end{pmatrix}$$

$$D^T = \begin{pmatrix} 0 & 0.63 & 1.37 & 1.74 & 1.74 & 0.63 \\ 0.63 & 0 & 1.11 & 3 & 3 & 0 \\ 1.37 & 1.11 & 0 & 3 & 3 & 1.11 \\ 1.74 & 3 & 3 & 0 & 0 & 3 \\ 1.74 & 3 & 3 & 0 & 0 & 3 \\ 0.63 & 0 & 1.11 & 3 & 3 & 0 \end{pmatrix} \qquad D^P = \begin{pmatrix} 0 & 2 & 2.63 & 1 & 1 \\ 2 & 0 & 1.11 & 3 & 3 \\ 2.63 & 1.11 & 0 & 3 & 3 \\ 1 & 3 & 3 & 0 & 0 \\ 1 & 3 & 3 & 0 & 0 \end{pmatrix}$$

5.2 The full case for $S = P$

When $\mathcal{A}$ and $\mathcal{B}$ are both full, $P^{\mathcal{A}}$ and $P^{\mathcal{B}}$ are well defined and theorem (1) yields $D^P_{\mathcal{A},\mathcal{B}} = \mathrm{Tr}(P^{\mathcal{A}} + P^{\mathcal{B}} - 2\, P^{\mathcal{A}}\, P^{\mathcal{B}})$. If $\mathcal{B} \le \mathcal{A}$ in addition, the distance further expresses (theorem (4)) as $D^P_{\mathcal{A},\mathcal{B}} = \mathrm{Tr}(P^{\mathcal{A}}) - \mathrm{Tr}(P^{\mathcal{B}}) = a - b$: the distance between two nested, full partitions is measured by the difference of their number of groups. In particular

- $D^P_{\mathcal{A},\mathcal{C}} = D^P_{\mathcal{A},\mathcal{B}} + D^P_{\mathcal{B},\mathcal{C}}$ if $\mathcal{C} \le \mathcal{B} \le \mathcal{A}$ or $\mathcal{A} \le \mathcal{B} \le \mathcal{C}$
- for $\mathcal{A}$ full, $D^P_{\mathcal{O},\mathcal{A}} = a - 1$ and $D^P_{\mathcal{N},\mathcal{A}} = n - a$
- for $\mathcal{A}$ full, $D^P_{\mathcal{A}^{(0)},\mathcal{A}} = a - c(\mathcal{A})$

5.3 The crisp case: chi-square and Mirkin-Cherny-Rand indices

Let $\mathcal{A}$ and $\mathcal{B}$ be two crisp partitions possessing respectively a and b non-empty classes. Let $n^{\mathcal{A}}_j := \sum_i z^{\mathcal{A}}_{ij}$ be the number of objects in class j of $\mathcal{A}$, $n^{\mathcal{B}}_k := \sum_i z^{\mathcal{B}}_{ik}$ be the number of objects in class k of $\mathcal{B}$, and define $n^{\mathcal{A},\mathcal{B}}_{jk} := \sum_i z^{\mathcal{A}}_{ij}\, z^{\mathcal{B}}_{ik}$ as the number of objects both in class j of $\mathcal{A}$ and k of $\mathcal{B}$.

Definition 6 : $N_{\mathcal{A},\mathcal{B}} := \sum_{ii'} \sum_{jj'} z^{\mathcal{A}}_{ij}\, z^{\mathcal{A}}_{i'j}\, z^{\mathcal{B}}_{ij'}\, z^{\mathcal{B}}_{i'j'} = \sum_{ii'} r^{\mathcal{A}}_{ii'}\, r^{\mathcal{B}}_{ii'} = \sum_{jk} (n^{\mathcal{A},\mathcal{B}}_{jk})^2$ denotes the *number of pairs* (distinct or not) which are simultaneously classified in the same group j of $\mathcal{A}$ and k of $\mathcal{B}$.

[3]as $\mathcal{F}$ is defective, $P^{\mathcal{F}}$ is not defined.

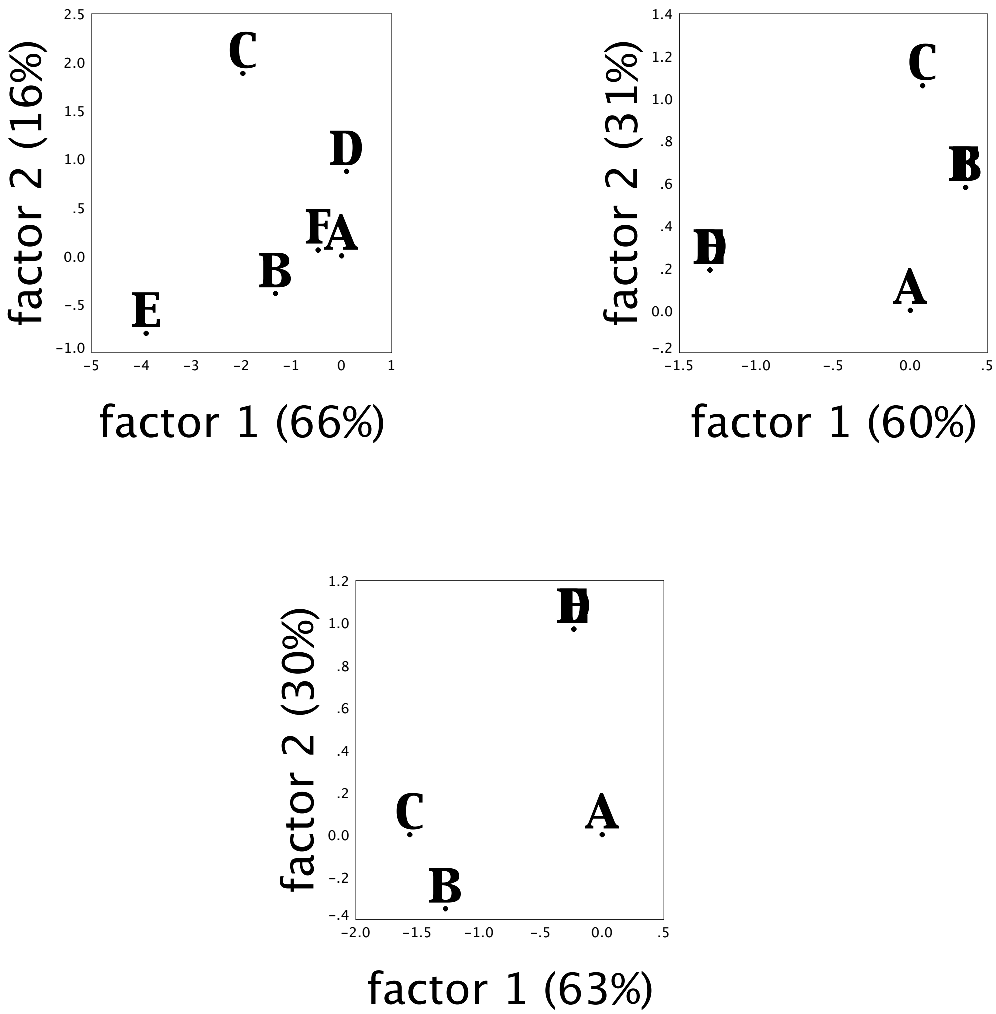

Figure 1: Euclidean visualization (classical MDS) of the distances between partitions $D^S_{\mathcal{A},\mathcal{B}}$, for $S = R$ (top left), $S = T$ (top right) and $S = P$ (down). Coordinates for $\mathcal{D}$ and $\mathcal{E}$ are identical in the $T-$ and $P-$representation; also, coordinates for $\mathcal{B}$ and $\mathcal{F}$ are identical in the $T-$representation. Recall that $\mathcal{F}$ is defective and hence not $P-$representable.

$n_{jk}^{\mathcal{A},\mathcal{B}} \leq n_{jj}^{\mathcal{A},\mathcal{A}} = (n_j^{\mathcal{A}})^2$, and thus $N_{\mathcal{A},\mathcal{B}} \leq N_{\mathcal{A},\mathcal{A}} = \sum_j (n_j^{\mathcal{A}})^2$ (and $N_{\mathcal{A},\mathcal{B}} \leq N_{\mathcal{B},\mathcal{B}} = \sum_k (n_k^{\mathcal{B}})^2$), with equality iff the two (crisp) partitions $\mathcal{A}$ and $\mathcal{B}$ are identical.

Theorem 6

$$D^R_{\mathcal{A},\mathcal{B}} = N_{\mathcal{A},\mathcal{A}} + N_{\mathcal{B},\mathcal{B}} - 2N_{\mathcal{A},\mathcal{B}} \qquad D^T_{\mathcal{A},\mathcal{B}} = D^P_{\mathcal{A},\mathcal{B}} = (a-1) + (b-1) - \frac{2}{n}\,\chi^2_{\mathcal{A},\mathcal{B}} \quad (11)$$

where $\chi^2_{\mathcal{A},\mathcal{B}} := \sum_{j=1}^a \sum_{k=1}^b \frac{(n_{jk} - \frac{n_{j\bullet} n_{\bullet k}}{n})^2}{\frac{n_{j\bullet} n_{\bullet k}}{n}}$ is the chi-square associated to the contingency table $n_{jk} = n_{jk}^{\mathcal{A},\mathcal{B}}$.

The quantity $\frac{1}{n^2}\, D^R_{\mathcal{A},\mathcal{B}}$ is called "relative symmetric-difference distance" by Mirkin and Cherny (1970). Its complement to unity[4] is known as the "Rand similarity index" (Rand 1971).

Proof of theorem 6 : the first identity follows from

$$D^R_{\mathcal{A},\mathcal{B}} = \sum_{ii'} (r^{\mathcal{A}}_{ii'} - r^{\mathcal{B}}_{ii'})^2 = \sum_{ii'} (\sum_j z^{\mathcal{A}}_{ij} z^{\mathcal{A}}_{i'j} - \sum_k z^{\mathcal{B}}_{ik} z^{\mathcal{B}}_{i'k})^2 =$$

$$= \sum_{ii'} [\sum_{jj'} z^{\mathcal{A}}_{ij} z^{\mathcal{A}}_{i'j} z^{\mathcal{A}}_{ij'} z^{\mathcal{A}}_{i'j'} + \sum_{j'k'} z^{\mathcal{B}}_{ik} z^{\mathcal{B}}_{i'k} z^{\mathcal{B}}_{ik'} z^{\mathcal{B}}_{i'k'} - 2 \sum_{jk} z^{\mathcal{A}}_{ij} z^{\mathcal{A}}_{i'j} z^{\mathcal{B}}_{ik} z^{\mathcal{B}}_{i'k}] =$$

$$= \sum_{jj'} \delta_{jj'}\, n^{\mathcal{A}}_j \delta_{jj'}\, n^{\mathcal{A}}_j + \sum_{kk'} \delta_{kk'}\, n^{\mathcal{B}}_k \delta_{kk'}\, n^{\mathcal{B}}_k - 2 \sum_{jk} (n^{\mathcal{AB}}_{jk})^2 = \sum_j (n^{\mathcal{A}}_j)^2 + \sum_k (n^{\mathcal{B}}_k)^2 - 2 \sum_{jk} (n^{\mathcal{AB}}_{jk})^2$$

and the second from $D^P_{\mathcal{A},\mathcal{B}} = \mathrm{Tr}(P^{\mathcal{A}}) + \mathrm{Tr}(P^{\mathcal{B}}) - 2\mathrm{Tr}(P^{\mathcal{A}} P^{\mathcal{B}}) = a + b - 2\,\mathrm{Tr}(P^{\mathcal{A}} S^{\mathcal{B}})$ and

$$\mathrm{Tr}(P^{\mathcal{A}} P^{\mathcal{B}}) = \sum_{ii'} \sum_{jk} \frac{z^{\mathcal{A}}_{ij} z^{\mathcal{A}}_{i'j}}{n^{\mathcal{A}}_j} \frac{z^{\mathcal{A}}_{i'k} z^{\mathcal{A}}_{ik}}{n^{\mathcal{B}}_k} = \sum_{jk} \frac{(n^{\mathcal{A},\mathcal{B}}_{jk})^2}{n^{\mathcal{A}}_j n^{\mathcal{B}}_k} = 1 + \frac{1}{n}\,\chi^2_{\mathcal{A},\mathcal{B}}$$

5.4 The crisp case: instability of a group relatively to another partition

Let $\mathcal{A}$ and $\mathcal{B}$ be two crisp partitions whose non-empty groups are respectively indexed by $j = 1, \ldots, a$ and $k = 1, \ldots, b$; let $n_j := n^{\mathcal{A}}_j > 0$ denote the number of objects $i \in j$.

Theorem 7

$$D^R_{\mathcal{A},\mathcal{B}} = \sum_{j=1}^a \rho^{\mathcal{B}}_j \qquad \rho^{\mathcal{B}}_j := n_j^2 - 2\alpha^{\mathcal{B}}_j + \beta^{\mathcal{B}}_j \qquad \alpha^{\mathcal{B}}_j := \sum_{i,i' \in j} r^{\mathcal{B}}_{ii'} \qquad \beta^{\mathcal{B}}_j := \sum_{i \in j; i'} r^{\mathcal{B}}_{ii'} \qquad (12)$$

$$D^T_{\mathcal{A},\mathcal{B}} = D^P_{\mathcal{A},\mathcal{B}} = \sum_{j=1}^a \tau^{\mathcal{B}}_j \qquad \tau^{\mathcal{B}}_j := 1 - 2\gamma^{\mathcal{B}}_j + \delta^{\mathcal{B}}_j \qquad \gamma^{\mathcal{B}}_j := \frac{1}{n_j} \sum_{i,i' \in j} p^{\mathcal{B}}_{ii'} \qquad \delta^{\mathcal{B}}_j := \sum_{i \in j} p^{\mathcal{B}}_{ii} \quad (13)$$

[4]at least in the variant restricted to the contribution of *distinct* pairs only.

$\rho_j^{\mathcal{B}}$ and $\tau_j^{\mathcal{B}}$ constitute measures of the *instability* of group j (of partition $\mathcal{A}$) relatively to partition $\mathcal{B}$; by construction, their sum over the groups $j = 1, \ldots, a$ yields the (squared) distance between partitions $\mathcal{A}$ and $\mathcal{B}$. Note that:

- n_j^2 is the number of (distinct or not) pairs of objects in j
- $\alpha_j^{\mathcal{B}}$ is the number of pairs in j which are also classified in the same group k of $\mathcal{B}$
- $\beta_j^{\mathcal{B}}$ is the number of pairs classified in the same group k of $\mathcal{B}$, such that the first object of the pair belongs to j.

As $n_j^2 \geq \alpha_j^{\mathcal{B}}$ and $\beta_j^{\mathcal{B}} \geq \alpha_j^{\mathcal{B}}$, one has $\rho_j^{\mathcal{B}} \geq 0$ with equality iff $n_j^2 = \alpha_j^{\mathcal{B}}$ (all pairs in j are pairs for $\mathcal{B}$) and $\beta_j^{\mathcal{B}} = \alpha_j^{\mathcal{B}}$ (all pairs (i, i') for $\mathcal{B}$ such that $i \in j$ satisfy $i' \in j$). Also:

- $\gamma_j^{\mathcal{B}}$ is a measure of the pair cohesion in $\mathcal{B}$ "as seen from j"
- $\delta_j^{\mathcal{B}}$ is a measure of the fineness of groups of $\mathcal{B}$ "as seen from j".

Properties $p_{ii'}^{\mathcal{B}} \leq p_{ii}^{\mathcal{B}}$ and $\sum_{i'} p_{ii'}^{\mathcal{A}} = 1$ entail $\gamma_j^{\mathcal{B}} \leq \delta_j^{\mathcal{B}}$ and $\gamma_j^{\mathcal{B}} \leq 1$, and thus $\tau_j^{\mathcal{B}} \geq 0$.

Proof of theorem 7

$$D^R_{\mathcal{A},\mathcal{B}} = \sum_{ii'} (r_{ii'}^{\mathcal{A}} - r_{ii'}^{\mathcal{B}})^2 = \sum_j \sum_{i \in j} \sum_{i'} [r_{ii'}^{\mathcal{A}} - 2 r_{ii'}^{\mathcal{A}} r_{ii'}^{\mathcal{B}} + r_{ii'}^{\mathcal{B}}] = \sum_j [n_j^2 - 2 \sum_{i,i' \in j} r_{ii'}^{\mathcal{B}} + \sum_{i \in j; i'} r_{ii'}^{\mathcal{B}}]$$

$$D^T_{\mathcal{A},\mathcal{B}} = \sum_{ii'} (p_{ii'}^{\mathcal{A}} - p_{ii'}^{\mathcal{B}})^2 = \sum_j [\sum_{i \in j} p_{ii}^{\mathcal{A}} - 2 \sum_{i \in j} p_{ii'}^{\mathcal{A}} p_{ii'}^{\mathcal{B}} + \sum_{i \in j} p_{ii}^{\mathcal{B}}] = \sum_j [1 - \frac{2}{n_j} \sum_{i,i' \in j} p_{ii'}^{\mathcal{B}} + \sum_{i \in j} p_{ii}^{\mathcal{B}}]$$

where $(r_{ii'}^{\mathcal{A}})^2 = r_{ii'}^{\mathcal{A}}$ and $\sum_{i'} (p_{ii'}^{\mathcal{A}})^2 = p_{ii}^{\mathcal{A}} = 1/n_{j(i)}$ have been used.

Example 3 : the instability of class j with regard to the n-groups partition $\mathcal{B} = \mathcal{N}$ is:

- $\rho_j^{\mathcal{N}} = n_j \, (n_j - 1)$ (*large* groups are unstable)
- $\tau_j^{\mathcal{N}} = n_j - 1$ (*large* groups are unstable)

Its instability with regard to the one-group partition $\mathcal{B} = \mathcal{O}$ is:

- $\rho_j^{\mathcal{O}} = (n - n_j) \, n_j$ (*medium* groups are unstable)
- $\tau_j^{\mathcal{O}} = \frac{n - n_j}{n} = 1 - f_j$ (*small* groups are unstable).

6 References

- BAVAUD, F. (2002): "Quotient Dissimilarities, Euclidean Embeddability, and Huygens' Weak Principle". In K. Jajuga and al. (Eds.): *Classification, Clustering, and Data Analysis.* Springer, 195-202
- BEZDEK, J. (1981): "Pattern recognition with Fuzzy Objective Function Algorithms", Plenum Press
- GOWER, J.C. (1982): "Euclidean distance geometry". *The Mathematical Scientist*, *7*, 1-14
- MARCOTORCHINO, J.F. (2000): "Dualité Burt-Condorcet: relation entre analyse factorielle des correspondances et analyse relationnelle". In J.Moreau and al. (Eds.): *L'analyse des correspondances et les techniques connexes.* Springer, 211-227
- MIRKIN, B. (1996): "Mathematical Classification and Clustering", Kluwer
- MIRKIN, B. and CHERNY, L.B. (1970): "On a distance measure between partitions of a finite set", *Automation and Remote Control*, *31*, 786-792
- RAND, W.M. (1971): "Objective criteria for the evaluation of clustering methods", *Journal of the American Statistical Association*, *66*, 846-850
- SAPORTA, G. (1990): "Probabiltés, analyse de données et statistique", Editions Technip, Paris
- SCHOENBERG, I.J. (1935): "Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espaces vectoriels distancés applicables vectoriellement sur l'espace de Hilbert"". *Annals of Mathematics*, *36*, 724-732