INDICE DE SIMILARITÉ ET PRÉORDONNANCE ASSOCIÉE

I. C. LERMAN (1)

RÉSUMÉ

On se donne un ensemble fini E *d'objets :* $x, y, z, \ldots$ *et un ensemble* A *de* T *attributs :* $a_1, a_2, \ldots, a_i, \ldots, a_T$ *établis pour décrire les éléments de* E. *Un objet donné* x *est défini par la donnée du sous-ensemble* $(X \subset A)$ *des attributs qu'il possède ; il est représenté dans* $\{0, 1\}^T$ *par le point*

$$a(x) = (x_1, x_2, \ldots, x_i, \ldots, x_T)$$

ou x_i, $i = 1, 2, \ldots, T$ *est une variable logique qui vaut 1 si l'objet* x *possède l'attribut* a_i *et 0 sinon.*

A deux objets quelconques x *et* y, *nous associons les cardinaux suivants :*

s (resp. t) *cardinal du sous-ensemble des attributs possédés en commun* (resp. *non possédés par aucun des deux objets*),

u (resp. v) *cardinal du sous-ensemble des attributs possédés par l'objet* x (resp. y) *et non possédés par* y (resp. *par* x).

Nous appelons indicateur *du couple* (x, y), $I(x, y)$, *le triplet* (s, u, v) *qui est considéré comme la donnée de base de la mesure de la ressemblance des deux objets. De la sorte l'importance accordée à la présence* (resp. *à l'absence*) *d'un attribut donné est la même pour tout attribut.*

Un indice de similarité *est une fonction réelle positive* S *définie sur l'ensemble* $E \times E$ *qui se présente sous la forme*

$$(x, y) \rightarrow S(x, y) = \mathcal{S}[I(x, y)] = \mathcal{S}(s, u, v)$$

ou la fonction $\mathcal{S}(s, u, v)$, *définie sur le sous-ensemble de* N^3 : $\{(s, u, v) ; s + u + v \leqslant T\}$, *est croissante par rapport à* s, *symétrique en* u *et* v, *et décroissante par rapport à* u.

---

(1) Centre de Mathématiques Appliquées et de Calcul, M. S. H.

Notre définition est plus restrictive que celle habituellement adoptée, la restriction consiste exactement à imposer à l'indice de similarité d'être une fonction du triplet (s, u, v).

Notre définition reste assez générale puisqu'elle inclut les différents indices de similarité, proposés, qui sont, à une élévation au carré près, des fractions rationnelles en s, $u + v$ *et* uv.

Le choix d'un indice de similarité S *définit sur l'ensemble des paires d'objets distincts de* E, F, *un préordre :* $\{ x, y \} < \{ z, t \} \Leftrightarrow S(x, y) \leqslant S(z, t)$.

Un tel préordre, noté w_S, *est appelé « préordonnance sur* E *associé à* S » La donnée de base de certaines méthodes de classification est précisément w_S. *Devant l'incertitude, ou on se trouve, lorsqu'on a à choisir l'un ou l'autre des indices, on peut se demander dans* quelle mesure, la préordonnance w_S varie lorsqu'on remplace un indice par un autre.

Pour un ensemble E *d'objets, donné, deux indices de similarité sont équivalents, si et seulement si, les préordonnances sur* E, *respectivement associées, sont identiques.*

Nous montrons que si le nombre d'attributs possédés par un même objet de E est invariable, tous les indices de similarité sont équivalents.

Dans le cas où deux indices S *et* S′ *ne sont pas équivalents,* nous définissons, naturellement, l' « écart » entre w_S et $w_{S'}$ par le nombre d'inversions que présente $w_{S'}$ par rapport à w_S.

Parmi les indices de similarité qui se présentent sous la forme, assez générale, $\mathcal{S}(s, u + v)$, *les deux indices, pour lesquels, les préordonnances respectivement associées sont les plus écartées, sont* $S(x, y) = s$ *et* $S'(x, y) = s + t$. *Nous montrons que si le nombre d'attributs possédés par un même objet de* E, *peut prendre, l'une des deux valeurs consécutives* (n *ou* $n + 1$), *le nombre d'inversions que présente* $w_{(s+t)}$ *par rapport à* w_s, *est nul. Plus généralement, si la variance dans* E *du nombre d'attributs possédés par un même objet est petite, le nombre d'inversions est petit. Nous avons par ailleurs déterminé l'espérance mathématique de ce nombre d'inversions pour un échantillon* E *d'objets, dont chacun est considéré comme une réalisation indépendante d'un sous-ensemble de* A, *selon la loi de probabilité :*

$$Pr\{ x_i = 1 \} = p \qquad \text{et} \qquad Pr\{ x_i = 0 \} = 1 - p,$$

où x_i *est la variable logique relative à* a_i *et ou les différents attributs sont indépendants en probabilité.*

Dans la pratique, le plus souvent (questionnaires, codes descriptifs d'objets) le nombre d'attributs possédés par un même objet, dans la population étudiée, est sinon invariable, du moins de variance faible. Cette circonstance confère aux méthodes basées sur la donnée d'une préordonnance, un caractère intrinsèque, c'est-à-dire ne dépendant pas du choix de l'indice de similarité.

1. INTRODUCTION

On se donne un ensemble fini E d'objets : $x, y, z, \ldots$; et un ensemble A de T attributs : $a_1, a_2, \ldots, a_T$ établis pour décrire les éléments de E. Un objet donné x est défini par la donnée du sous-ensemble X, ($X \subset A$) des attributs qu'il possède ; il est représenté dans $\{0, 1\}^T$ par le point

$$a(x) = (x_1, x_2, \ldots, x_i, \ldots, x_T)$$

où x_i, $i = 1, 2, \ldots, T$ est une variable logique qui vaut 1 si l'objet x possède l'attribut a_i et 0 sinon.

A deux objets quelconques x et y de E, nous associons les cardinaux suivants :

s (resp. t) cardinal du sous-ensemble des attributs possédés en commun (resp. non possédés par aucun des deux objets),

u (resp. v) cardinal du sous-ensemble des attributs possédés par l'objet x (resp. y) et non possédés par y (resp. par x).

Désignons, par X et Y les sous-ensembles d'attributs possédés respectivement par x et y ; et par $(x_1, x_2, \ldots, x_i, \ldots, x_T)$ et $(y_1, y_2, \ldots, y_i, \ldots, y_T)$ les points dans $\{0, 1\}^T$ représentant respectivement x et y. On a

$$s = |X \cap Y| = \sum_{i=1}^{T} x_i y_i, \quad t = |X^c \cap Y^c| = \sum_{i=1}^{T} (1 - x_i)(1 - y_i)$$

X^c (resp. Y^c) étant le complémentaire dans A de X (resp. Y)

$$u = |X \cap Y^c| = \sum_{i=1}^{T} x_i(1 - y_i), \quad v = |X^c \cap Y| = \sum_{i=1}^{T} (1 - x_i) y_i$$

On a $s + u = |X|$ et $s + v = |Y|$.

Les cardinaux s, u, v et t sont liés par la relation $s + u + v + t = T$.

2. DÉFINITION D'UN INDICE DE SIMILARITÉ

2.1. **Définition.** — Nous appelons indicateur du couple (x, y), $I(x, y)$, le triplet (s, u, v) où s, u et v sont relatifs au couple (x, y).

Nous considérons $I(x, y) = (s, u, v)$, comme la donnée de base de la mesure de la ressemblance des deux objets. De la sorte l'importance accordée à la présence (resp. à l'absence) d'un attribut donné est la même pour tout attribut.

2.2. **Définition** 1. — Un « indice de similarité » est une fonction réelle positive S définie sur l'ensemble E × E qui se présente sous la forme

$$(x, y) \to S(x, y) = \mathscr{S}I(x, y) = \mathscr{S}(s, u, v)$$

où $\mathscr{S}(s, u, v)$ est une fonction définie sur le sous-ensemble de N^3 : $\{ (s, u, v) ; s + u + v \leqslant T \}$, croissante par rapport à s, symétrique en u et v, et décroissante par rapport à u. Nous allons comparer notre définition avec celle habituellement adoptée.

2.3. **Définition** 2. — Un « indice de similarité » est une fonction réelle positive S définie sur l'ensemble E × E qui satisfait aux conditions :

1) $\forall (x, y) \in E \times E : S(x, y) = S(y, x)$
2) $\forall (x, y) \in E \times E ; S(x, y) \leqslant S(x, x)$
3) Si deux objets x et y diffèrent seulement du point de vue de l'attribut a_k, $(x_i = y_i, i \neq k$ et $x_k \neq y_k)$, et si z est un objet quelconque de E, pour lequel :

$$| z_k - x_k | < | z_k - y_k |,$$

Alors $S(x, z) \geqslant S(y, z)$.

2.4. **Proposition.** — Tout indice de similarité satisfaisant aux axiomes de la définition 1, satisfait aux axiomes de la définition 2.

Réciproquement, tout indice de similarité, $S(x, y)$, satisfaisant aux axiomes de la définition 2, qui se met sous la forme $S(x, y) = \mathscr{S}(s, u, v)$, satisfait aux axiomes de la définition 1.

Il est aisé de démontrer la première partie de la proposition. Occupons-nous de la réciproque. Si un indice de similarité, $S(x, y)$, se met sous la forme : $S(x, y) = \mathscr{S}I(x, y) = \mathscr{S}(s, u, v)$, l'axiome 1 de la « définition 2 » implique la symétrie de la fonction $\mathscr{S}(s, u, v)$ en u et v : $\mathscr{S}(s, u, v) = \mathscr{S}(s, v, u)$. D'autre part l'axiome 3 de la « définition 2 » s'exprime par l'inégalité

$$\mathscr{S}(s, u, v) \geqslant \mathscr{S}(s - \varepsilon^2, u + (1 - \varepsilon)^2, v + \varepsilon^2) \tag{1}$$

où $\varepsilon = 0$ ou 1.

En faisant $\varepsilon = 0$, on obtient la condition de décroissance par rapport à u de la fonction $\mathscr{S}(s, u, v)$.

En faisant $\varepsilon = 1$, on obtient :

$$\mathscr{S}(s, u, v) \geqslant \mathscr{S}(s - 1, u, v + 1) \tag{2}$$

On a nécessairement

$$\mathscr{S}(s, u, v) \geqslant \mathscr{S}(s - 1, u, v) \tag{3}$$

Car sinon on aurait

$$\mathscr{S}(s - 1, u, v) > \mathscr{S}(s, u, v)$$

ce qui implique en raison de (2)

$$\mathscr{S}(s-1, u, v) > \mathscr{S}(s-1, u, v+1)$$

ce qui est impossible en vertu de la décroissance de $\mathscr{S}(s, u, v)$ par rapport à u et de la symétrie de $\mathscr{S}(s, u, v)$ en u et v, déjà établies ci-dessus. La condition (3) exprime que la fonction $\mathscr{S}(s, u, v)$ est croissante par rapport à s.

C. Q. F. D.

Notre définition est donc plus restrictive que celle habituellement adoptée, la restriction consiste exactement à imposer à l'indice de similarité d'être une fonction du triplet (s, u, v). Notre définition reste assez générale puisqu'elle inclut les différents indices de similarité proposés, qui sont, à une élévation au carré près, des fractions rationnelles en s, $u + v$ et uv.

3. PRÉORDONNANCE ASSOCIÉE A UN INDICE DE SIMILARITÉ

Le choix d'un indice de similarité S définit sur l'ensemble des paires d'objets distincts de E, F, un préordre.

$$(\{x, y\}, \{z, t\}) \in F \times F ; \quad \{x, y\} < \{z, t\} \Leftrightarrow S(x, y) \leqslant S(z, t).$$

Un tel préordre, est appelé, « préordonnance sur E associée à S et noté w_s. D'autre part nous notons W_s, le graphe dans F × F de ce préordre. *La donnée de base de certaines méthodes de classification est précisément* w_s.

Sous ce point de vue, on saisit l'arbitraire qui consiste à prendre pour indice de similarité, précisément une fraction rationnelle en s, $u + v$ et uv. Devant l'incertitude où on se trouve lorsqu'on a à choisir l'un ou l'autre des indices, on peut se demander *dans quelle mesure, la préordonnance* w_s *varie lorsqu'on remplace un indice de similarité par un autre.*

3.1. **Définition.** — Deux indices de similarité sont équivalents sur un ensemble E d'objets donné, si et seulement si, les préordonnances respectivement associées sur E, sont identiques.

$S(x, y)$ et $S'(x, y)$ désignant deux indices de similarité, il y a lieu de remarquer que s'il existe une fonction numérique strictement croissante, f, telle que $S'(x, y) = f[S(x, y)]$; S et S′ sont équivalents sur tout ensemble, E, d'objets.

3.2. **Proposition.** — Si le nombre d'attributs possédés par un même objet de E est invariable tous les indices de similarité sont équivalents sur E.

La fonction, $\mathscr{S}(s, u, v)$, où $S(x, y) = \mathscr{S}(s, u, v)$, est supposée strictement croissante en s, ou, non exclusivement, strictement décroissante en u.

Cette condition est pratiquement vérifiée pour tous les indices de similarité qui ont été proposés.

Dans les conditions de la proposition, l'indicateur d'un couple (x, y) de E × E se présente sous la forme :

$$I(x, y) = (s, n - s, n - s),$$

où $s = \sum_{i=1}^{T} x_i y_i$ et n est le nombre d'attributs communs à tous les éléments de E.

Il en résulte qu'un indice de similarité quelconque, $S(x, y)$, se met sous la forme : $S(x, y) = \mathscr{S}(s, u, v) = \mathscr{S}(s, n - s, n - s) = R(s)$; où $R(s)$ est une fonction numérique, à valeurs positives, définie sur 0, 1, 2, ..., T et *strictement croissante.*

C. Q. F. D.

Expression de différentes mesures de similarité proposées par différents auteurs

(1) Russel et Rao (1940) $S(x, y) = \frac{s}{T}$

(2) M. G. Kendall; Sokal-Michener (1958) $S(x, y) = 1 - \frac{u + v}{T}$

(3) Roger et Tanimoto (1960) $S(x, y) = \frac{T - (u + v)}{T + (u + v)}$

(4) Hamman (1961) $S(x, y) = 1 - \frac{2(u + v)}{T}$

(5) Jaccard (1908) $S(x, y) = \frac{s}{s + (u + v)}$

(6) Kulezynski (1927) $S(x, y) = \frac{s}{u + v}$

(7) Dice (1945) $S(x, y) = \frac{s}{s + \frac{1}{2}(u + v)}$

(8) Sokal et Sneath $S(x, y) = \frac{s}{s + 2(u + v)}$

(9) Kulezynski $S(x, y) = \frac{1}{2}\left[\frac{s}{s + u} + \frac{s}{s + v}\right]$

(10) Ochiai (1957) $S(x, y) = \dfrac{s}{\sqrt{(s+u)(s+v)}}$

(11) Yule (1911) $S(x, y) = \dfrac{st - uv}{st + uv}$

(12) Pearson $S(x, y) = \dfrac{st - uv}{\sqrt{(s+u)(s+v)(t+u)(t+v)}}$

où $I(x, y) = (s, u, v)$.

Notons que la similarité (1) est une fonction de s seulement, que les indices (2), (3) et (4) sont respectivement des fonctions de $d = u + v$ et que (5), (6), (7) et (8) sont des similarités dont chacune est une fonction du rapport $s/(u + v)$. Remarquons encore que la similarité (9) est la moyenne algébrique de $s/(s + u)$ et de $s/(s + v)$ et que (10) en est la moyenne géométrique.

Notons enfin que $S(x, x) = \mathscr{S}(s, 0, 0) = 1$ pour ces différentes mesures de similarité proposées à l'exception de (1) et de (6) où

$$S(x, x) = \frac{1}{T}\sum_{i=1}^{T} x_i^2$$

pour (1) et $S(x, x) = \infty$ pour (6).

Relativement aux indices (1) et (2), $1 - \dfrac{s}{T}$ et $\dfrac{u+v}{T}$ définissent deux distances sur l'ensemble des parties de A.

Ces similarités ont été proposées dans le cas où les différents caractères sont bivalents et où l'ensemble des attributs, A, est déterminé en retenant pour chacun des caractères celle des deux modalités la plus significative.

4. ÉCART ENTRE DEUX PRÉORDONNANCES ASSOCIÉES A DEUX INDICES DE SIMILARITÉ

Dans le cas où deux indices S et S′ ne sont pas équivalents, il y a lieu de définir l' « écart » entre les deux préordonnances, respectivement associées, w_s et $w_{s'}$, sur un ensemble E donné.

4.1. **Définitions.** — 1) W_s et $W_{s'}$ désignant les graphes dans $F \times F$ des préordres w_s, $w_{s'}$, nous appelons « désaccord » de W_s et $W_{s'}$ le sous-ensemble de $F \times F$.

$$\Delta(W_s, W_{s'}) = \{(p, q) \in F \times F ;\ (p, q) \notin W_s \text{ et } (q, p) \notin W_{s'} \text{ ou } (p, q) \notin W_{s'} \text{ et } (q, p) \notin W_s\}$$

Remarquons que :

$$(p, q) \in \Delta(W_s, W_{s'}) \Leftrightarrow (q, p) \in \Delta(W_s, W_{s'})$$

2) Nous appelons « écart » entre W_s et $W_{s'}$ le nombre entier

$$\frac{1}{2} | \Delta(W_s, W_{s'}) |,$$

$| \Delta(W_s, W_{s'}) |$ est un entier pair en vertu de la remarque faite.

$\Delta(W_s, W_{s'})$ est l'ensemble des couples (p, q) pour lesquels p et q ne sont pas comparables pour l'intersection des deux préordres. Dans le cas où W_s et $W_{s'}$ sont des ordres totaux, $\Delta(W_s, W_{s'})$ est la différence symétrique de W_s et $W_{s'}$. *L'« écart » entre* W_s *et* $W_{s'}$ *est le nombre d'inversions que présente* $W_{s'}$ *par rapport à* W_s. Le couple $(\{ x, y \}, \{ z, t \})$ donne lieu à une inversion si et seulement si $S(x, y) < S(z, t)$ et $S'(x, y) > S'(z, t)$ ou $S(x, y) > S(z, t)$ et $S'(x, y) < S'(z, t)$.

4.2. **Notations.** — E étant l'ensemble d'objets donné de cardinal N. nous désignerons, dans la suite, par ω et Ω les préordonnances sur E. respectivement associées aux indices de similarité $S(x, y) = s$ et $S'(x, y) = s + t$ et par Δ le « désaccord » de ω et Ω ; $\Delta = \Delta(\omega, \Omega)$. Nous allons, dans la suite, restreindre notre attention à l'ensemble des indices de similarité qui se présentent sous la forme générale $\mathscr{S}(s, u + v)$.

4.3. **Proposition.** — α et β étant les préordonnances respectivement associées aux deux indices de similarité $\varphi(s, u + v)$ et $\sigma(s, u + v)$, on a : $\Delta(\alpha, \beta) \subset \Delta$.

En effet, soit $(\{ x, y \}, \{ z, t \})$ un élément quelconque de $\Delta(\alpha, \beta)$ désignons par (s, u, v) et $(s + r, u + k, v + h)$, les indicateurs des couples (x, y) et (z, t). Dans ces conditions, $\varphi(s + r, u + v + (k + h)) - \varphi(s, u, v)$ et $\sigma(s + r, u + v + (k + h)) - \sigma(s, u, v)$ sont non nuls et de signes contraires. Il en résulte que r et $(k + h)$ sont non nuls et de même signe, donc $(\{ x, y \}, \{ z, t \})$ appartient à Δ.

C. Q. F. D.

Remarque. — Il résulte de cette proposition que, parmi les indices de similarité qui se présentent sous la forme $\mathscr{S}(s, u + v)$, les deux indices, pour lesquels, les préordonnances respectivement associées ont l' « écart » le plus grand, sont $S(x, y) = s$ et $S'(x, y) = s + t$.

4.4. **Proposition.** — Si le nombre d'attributs possédés par un même objet de E, prend l'une de deux valeurs consécutives (n ou $n + 1$), l'ensemble Δ est vide : $\Delta = \emptyset$.

En effet si Δ n'est pas vide, il existerait un élément de $F \times F$, $(\{x, y\}, \{z, t\})$ pour lequel

$$(1) \qquad \sum_i x_i y_i > \sum_i z_i t_i \quad \text{et} \quad \sum_i (x_i + y_i) > \sum_i (z_i + t_i) + 2\left(\sum_i x_i y_i - \sum_i z_i t_i\right).$$

La relation (1) implique

$$(2) \qquad \sum_i (x_i + y_i) - \sum_i (z_i + t_i) \geqslant 3$$

Or

$$\sum_i (x_i + y_i) \leqslant 2n + 2 \quad \text{et} \quad \sum_i (z_i + t_i) \geqslant 2n,$$

ce qui rend (2) impossible.

C. Q. F. D.

Remarque. — Dans les conditions de la proposition précédente, le nombre d'inversions que présente ω par rapport à Ω, est nul.

4.5. **Proposition.** — V désignant la variance dans E du nombre d'attributs possédés par un même objet, on a :

$$\frac{|\Delta|}{|F \times F|} < \frac{4V}{9}$$

Si n_x désigne le nombre d'attributs possédés par l'objet x, $x \in E$, on a

$$V = \frac{1}{N} \sum_{x \in E} (n_x - \bar{n})^2 \qquad \text{où} \qquad \bar{n} = \frac{1}{N} \sum_{x \in E} n_x$$

L'énoncé de la proposition exprime *que l'écart entre ω et Ω est strictement majoré* par $|F|^2 \times \frac{2V}{9}$.

Désignons par $\Phi(p, q)$, $(p, q) \in F \times F$, la fonction indicatrice de l'ensemble Δ. On a :

$$(1) \qquad |\Delta| = \sum_G \Phi(\{x, y\}, \{z, t\}) + \sum_H \Phi(\{x, y\}, \{x, z\})$$

La première somme (de gauche à droite), est étendue à l'ensemble G, $G \subset F \times F$, des couples de paires sans composante commune, et la seconde

somme à l'ensemble H, $H \subset F \times F$, des couples de paires ayant une composante commune. On a pour les cardinaux de G et de H :

$$|G| = N(N-1)(N-2)(N-3)/4 \tag{2}$$
$$|H| = N(N-1)(N-2)$$

Posons :

$$p = \frac{|G \cap \Delta|}{|G|} \quad \text{et} \quad p' = \frac{|H \cap \Delta|}{|H|}$$

Dans ces conditions la relation (1) peut être écrite sous la forme

$$|\Delta| = \frac{1}{4}N(N-1)(N-2)(N-3)p + N(N-1)(N-2)p'. \tag{3}$$

Munissons l'ensemble E de la loi uniforme, le nombre n_x d'attributs possédés par un objet x pris dans E selon le modèle (équiprobabilité d'apparition de chacun des objets), est une variable aléatoire de moyenne $\bar{n}$ et de variance V. Dans ce cadre, considérons un échantillon de 4 objets indépendants, x, y, z et t de E ; p est précisément la probabilité de l'événement : $(\{x, y\}, \{z, t\}) \in \Delta$. Un tel événement implique :

$$\{|(n_n + n_y) - (n_x + n_t)| \geqslant 3\}.$$

Donc

$$p \leqslant Pr\{|(n_x + n_y) - (n_z + n_t)| \geqslant 3\}.$$

Or, en vertu de l'inégalité de Tchebychef

$$Pr\{|(n_x + n_y) - (n_z + n_t)| \geqslant 3\} \leqslant \frac{1}{3^2}E[(n_x + n_y) - (n_z + n_t)]^2 = \frac{4V}{9}.$$

D'autre part, pour un échantillon de 3 objets indépendants, x, y et z de E, p' est précisément la probabilité de l'événement : $(\{x, y\}, \{x, z\}) \in \Delta$. Un tel événement implique :

$$\{|n_y - n_z| \geqslant 3\}.$$

Donc

$$p' \leqslant Pr\{|n_y - n_z| \geqslant 3\} \leqslant \frac{1}{3^2}E[n_y - n_z]^2 = \frac{2V}{9}.$$

Il en résulte que $|\Delta|$, défini en (3), est majoré par

$$\frac{1}{9}N(N-1)(N-2)(N-3)V + \frac{2}{9}N(N-1)(N-2)V$$

soit

$$\frac{1}{9}(N-2)(N-1)^2N.V. \tag{4}$$

Le cardinal de $F \times F$ étant $\frac{1}{4}(N-1)^2N^2$, on a

$$\frac{|\Delta|}{F \times F} \leqslant \frac{4}{9} \times \frac{N-2}{N} V < \frac{4V}{9} \tag{5}$$

4.6. **Conclusion.** — Donc si la variance dans E du nombre d'attributs possédés par un même objet est petite, le nombre d'inversions est petit. Dans la pratique, le plus souvent (questionnaires, codes descriptifs d'objets) le nombre d'attributs possédés par un même objet, dans la population étudiée, est, sinon invariable, du moins de variance faible. Cette circonstance confère aux méthodes basées sur la donnée d'une préordonnance associée à un indice de similarité, un caractère intrinsèque, c'est-à-dire ne dépendant pas du choix de l'indice.

BIBLIOGRAPHIE

[1] J. P. BENZECRI, *Classification Automatique et reconnaissance des formes*, cours I. S. U. P., 1968-1969.

[2] I. C. LERMAN, *Les bases de la Classification Automatique*. Gauthier-Villars. Collection Programmation. Paris, 1970.

[3] G. ROUX et M. ROUX, A propos de quelques méthodes de classification en phytosociologie. *Rev. de Stat. Appl.*, vol. XV, n° 2, 1967.

[4] R. R. SOKAL et P. H. A. SNEATH, *Principles of numerical taxonomy*, San Francisco and London, Freeman and Co, 1963.

[5] W. F. DE LA VEGA, Techniques de classification automatique utilisant un indice de ressemblance. *Rev. Franç. de Sociologie*, décembre 1967.