REVUE DE STATISTIQUE APPLIQUÉE

J. P. BENZÉCRI

Description mathématique des classifications

Revue de statistique appliquée, tome 20, nº 3 (1972), p. 23-56

<http://www.numdam.org/item?id=RSA_1972__20_3_23_0>

DESCRIPTION MATHÉMATIQUE DES CLASSIFICATIONS

J. P. BENZÉCRI *

Dans la présente leçon, nous donnons une description mathématique de notions dont la leçon consacrée à la taxinomie zoologique et botanique nous a fourni des exemples.

Ainsi, il ne s'agira nullement ici des classifications par recouvrement, ou systèmes de classes empiétantes, mais seulement des hiérarchies de classes emboitées (ou des simples partitions).

Le numéro 1 traite des structures d'ordres et plus particulièrement des arbres.

Le numéro 2 définit diverses notions de classifications arborescentes.

Le numéro 3 montre que la structure de hiérarchie stratifiée, est un cas particulier de la structure de préordonnance (structure plus faible que la structure métrique, en ce qu'elle ne spécifie pas des distances numériques, mais des inégalités entre distances) : l'importance de la structure est on le sait apparue dans les travaux de R.N. Shepard (1962).

Le numéro 4 est consacré à la géométrie ultramétrique, autre structure étroitement associée aux hiérarchies taxinomiques.

Bien qu'il soit difficile de séparer la description mathématique des classifications des critères et des algorithmes utilisés pour construire une classification sur un ensemble, nous nous réservons de traiter ailleurs des critères et algorithmes.

1 - Arbres

1.1 - Relations d'ordre

Le but de ce rappel est de préciser nos situations, tout en proposant quelques exercices aux statisticiens peu familiers avec les formules logiques.

Communément, on note une relation d'ordre par un signe placé entre l'élément inférieur et l'élément supérieur : ainsi on écrit, pour deux nombres réels : $x < y$; ou, pour deux ensembles : $a \subset b$. Lorsqu'on introduit une nouvelle relation d'ordre, il faut un nouveau signe, qui peut être α. Pour éviter les difficultés typographiques, on peut aussi noter fonctionnellement la relation d'ordre, et c'est ce que nous ferons ici ; appliquée à l'inclusion ensembliste, cette méthode conduirait à écrire, par exemple

----

* Laboratoire de Statistique Mathématique, Université de Paris VI

Inc (a, b) au lieu de : $a \subset b$.

Soit A un ensemble. Rappelons qu'une relation binaire H sur A est dite relation de préordre si elle satisfait aux deux axiomes de transitivité, H_{tr}, et de réflexivité, H_r.

$$H_{tr} : \forall a, b, c \in A : H(a, b) \wedge H(b, c) \Rightarrow H(a, c) ;$$

$$H_r : \forall a \in A : H(a, a).$$

Un préordre est dit total, s'il permet de comparer tout couple d'éléments ; plus précisément s'il satisfait à l'axiome H_{to} :

$$H_{to} \ \forall a, b \in A : H(a, b) \vee H(b, a)$$

Ce dernier axiome n'exclut nullement la possibilité que les deux relations H (a, b) et H (b, a) soient simultanément vraies pour certains couples. On dit que H est une relation d'ordre, si, outre les axiomes H_{tr} et H_r, elle satisfait à l'axiome H_{or} :

$$H_{or} : \forall a, b \in A : H(a, b) \wedge H(b, a) \Rightarrow (a = b).$$

On remarquera, à titre d'exercice que l'on peut réunir les axiomes H_r et H_{or} en un seul en posant :

$$\forall a, b \in A : H(a, b) \wedge H(b, a) \Leftrightarrow (a = b).$$

Un ordre est dit total s'il satisfait à l'axiome H_{to}. A tout préordre H est associée une relation d'équivalence Q_H (ou, en abrégé Q) sur l'ensemble A, telle que le quotient A/Q est naturellement muni d'une relation d'ordre (laquelle est totale si le préordre H est total). Il est commode de noter fonctionnellement la relation d'équivalence Q, en posant Q (a) = Q (b) si a et b sont Q-équivalents, i.e. ont même classe d'équivalence Q (a) = Q (b). Ceci posé la relation d'équivalence Q est définie par :

$$\forall a, b \in A \ : Q(a) = Q(b) \Leftrightarrow H(a, b) \wedge H(b, a)$$

Il résulte des axiomes H_r et H_{tr} que la relation (Q (a) = Q (b) est bien réflexive et transitive. Rappelons que l'ordre H_Q sur le quotient Q (A) ou A/Q...) est défini par :

$$\forall q, q' \in Q(A) : H_Q(q, q') \Leftrightarrow \exists a, a' \in A : Q(a) = q ; Q(a') = q' ; H(a,a').$$

(Dans cette formule, suivant un usage commun, on a remplacé le signe Λ de la conjonction logique, par un point et virgule). On vérifiera en exercice que l'ordre H possède la propriété suivante :

$$\forall a, b \in A : H(a, b) \Leftrightarrow H_Q(Q(a), Q(b))$$

L'ordre quotient H_Q est total si et seulement si le préordre H est total.

Soit f une application de l'ensemble A dans un ensemble B muni d'un préordre H : on définit sur A un préordre H_f, appelé image de H par f, par la condition :

$$\forall a, a' \in A : H_f(a, a') \Leftrightarrow H(f(a), f(a')).$$

Si H est un ordre, ou est total, il en est de même de H_f, à condition toutefois que f soit injective. Dans le cas particulier où A est une partie de B et où f est l'inclusion (application identité) de A dans B, H_f est appelé ordre induit par H sur la partie A de B et on écrit généralement simplement H pour H_f. Soit A un ensemble fini muni d'un préordre total H : H peut être défini comme l'image réciproque par une application f convenable de A dans la droite réelle R (ou dans un intervalle, e.g. (0,1) de R) de l'ordre naturel de R : on a :

$$\forall (a, a') \in A : H (a, a') \Leftrightarrow f (a) < f (a').$$

(l'application f range les éléments de A sur la droite R dans l'ordre H). On a là une méthode commode pour noter un préordre total : au lieu du tableau carré $A \times A$ de la relation H (fonction logique qui vaut 1 si H (a, b) et 0 sinon), on a une suite $\{f (a) \mid a \in A\}$ de nombres indicée par A. Dans la suite on raisonnera souvent sur une fonction f (indice d'ordre) en ne faisant usage que de la relation H qu'elle définit, et non des valeurs mêmes prises par f.

1.2 - Arbres

Avant d'énoncer l'axiome de la structure d'arbre, posons quelques notations valables pour un ensemble A muni d'une structure d'ordre, (voire de préordre quelconque) H, qu'on lira H (a, b) : b précède a, ou est au dessus de a ; a succède à b, est sous b. On note :

Som (A), ou Som A ; l'ensemble des éléments maximaux, ou sommets de A :

$$\mathrm{Som}\ (A) = \{a \mid a \in A ; \forall b \in A : H (a, b) \Rightarrow a = b\}.$$

Ter (A), ou Ter A : l'ensemble des éléments minimaux, ou terminaux de A :

$$\mathrm{Ter}\ (A) = \{a \mid a \in A ; \forall b \in A : H (b, a) \Rightarrow a = b\}.$$

Nod (A), ou Nod A : l'ensemble des noeuds vrais de A, i.e. des éléments non minimaux :

$$\mathrm{Nod}\ (A) = A - \mathrm{Ter}\ (A).$$

Pour tout élément a de A, on note :

Pred (a ; A), ou Pred (a) : l'ensemble des prédécesseurs de a, (a, y compris) :

$$\mathrm{Pred}\ (a) = \{b \mid b \in A ; H (a, b)\}$$

Su (a ; A), ou Su (a), l'ensemble des successeurs de a :

$$\mathrm{Su}\ (a) = \{b \mid b \in A ; H (b, a)\}$$

Sui (a ; A), ou Sui (a) : l'ensemble des successeurs immédiats de a :

$$\mathrm{Sui}\ (a) = \ldots$$

$$\{b \mid b \in A ; \{u \mid u \in A ; H (b, u) ; H (u, a)\} = \{a, b\}\}.$$

l'ensemble Sui (a) est non-vide si et seulement si $a \in \mathrm{Nod}\ (A)$.

Il va de soi que toutes les définitions ci-dessus étant relatives à l'ordre H, il convient de mentionner cette relation (e.g. en indice : Som_H, Sui_H etc...) si une confusion est à craindre.

Ceci posé, l'ensemble A ordonné par H est appelé un arbre si est satisfait l'axiome H_a ; H est alors appelé un ordre hiérarchique.

H_a : Quel que soit a dans A, l'ensemble Pred (a ; A), muni de l'ordre induit par H, est totalement ordonné.

A titre d'exercice on récrit cet axiome comme une formule logique :

$$\forall a, u, v \in A : H(a, u) \wedge H(a, v) \Rightarrow H(u, v) \vee H(v, u).$$

On vérifiera, autre exercice, que, si l'ensemble A est fini, il est équivalent de demander que tout élément $a \in A - \text{Som}(A)$ admette un prédécesseur immédiat unique.

Un arbre est dit connexe s'il satisfait à l'axiome H_c

H_c : card Som (A) = 1 ;

On note alors som (A) l'unique sommet de A : {som (A)} = Som (A). Un arbre est dit binaire s'il satisfait à l'axiome H_b :

$$H_b : \forall a \in \text{Nod}(A) : \text{card Sui}(a) = 2.$$

Un arbre est dit fourchu s'il satisfait à l'axiome H_f :

$$H_f : \forall a \in \text{Nod}(A) : \text{card Sui}(a) > 2.$$

Une partie A' d'un arbre A, munie de la structure d'ordre induite par celle de A, satisfait nécessairement à l'axiome H_a : c'est donc un arbre. On peut dire que A' est un sous-arbre de A, et que A est un sur-arbre de A'. En particulier, soit $a \in \text{Nod}(A)$; on définit le sous-arbre élémentaire A (a) de A ayant pour sommet a.

$$\forall a \in \text{Nod}(A) : A(a) = \{a\} \uplus \text{Sui}(a ; A).$$

On a :

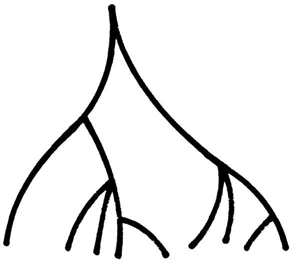
$$\text{som } A(a) = a \text{ et : Ter } A(a) = \text{Sui}(a).$$

Rappelons ici, pour aider l'intuition que, dans la représentation graphique d'un arbre fini il est d'usage de déployer la hiérarchie depuis le haut vers le bas : (i.e , si on a H (a, b), l'abscisse (ou cote) de b sera supérieure à celle de a) ; afin de préciser la relation H, on doit encore relier chaque noeud par un trait rectiligne à chacun de ses successeurs immédiats (cf fig 1) (parfois, cf figure 1', on dessine comme un peigne les sous-arbres élémentaires). Alors le sommet som (A) d'un arbre connexe apparait comme une souche ou un tronc ; pour chaque noeud $a \in \text{Nod}(A)$, l'ensemble Su (a) des successeurs de a, est comme la branche issue de a ; cette branche se divise à partir de a en autant de rameaux que a a de successeurs immédiats ; il y a toujours plus d'un rameau si A est fourchu. Il faut aviser que l'ordre H est entendu de telle sorte que les branches sont inférieures au tronc ; en ce sens, dans l'usage mathématique, l'arbre est renversé. Le dessin d'un arbre non-connexe comprend en effet plusieurs graphiques séparés : c'est une sorte de forêt dont chaque arbre Su (s) a pour souche un sommet $s \in \text{Som}(A)$. L'axiome H_a correspond à la propriété que les branches une fois qu'elles ont divergé, ne se referment plus : il existe donc un parcours ordonné de l'ensemble des prédécesseurs d'un noeud quelconque.

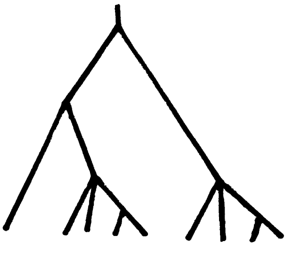
Dans certains cas, la hiérarchie est déployée non de haut en bas, mais de bas en haut ; ou de droite à gauche (cf listage de sortie de programmes de classification : les éléments terminaux y sont à la marge, à gauche).

L'axiome H_a peut fort bien s'appliquer à un ensemble ordonné infini : toutefois l'arbre généalogique continu dessiné pour illustrer la leçon "La classification dans les sciences de la nature" (fig.3), ne pourrait y satisfaire parce que ses branches ont une épaisseur. Mais si les branches sont réduites à des lignes, on a un arbre au sens de H (fig.1).

Les arbres probabilisés (munis d'une loi de probabilité sur l'ensemble Ter (A)), finis ou infinis, sont sous-jacents à des modèles hiérarchiques usités en statistique.



arbre infini



arbre fini

Figure 1

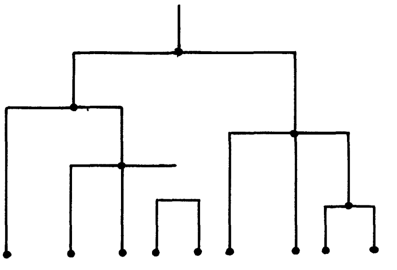


Figure 1'

1.3 - Arbres multiordonnés

Dans les applications, un arbre est souvent muni, outre la structure hiérarchique H, d'une ou deux structure d'ordre total, compatibles avec H en un certain sens. Ces structures, qui apparaissent naturellement sur le dessin plan d'un arbre sont communément appelés par nous : ordre de stratification (correspondant à l'axe vertical du dessin), et ordre séquentiel (correspondant à l'axe horizontal). L'ordre de stratification est usité en taxinomie ; tandis que nous ne rappelons ici que pour mémoire l'ordre séquentiel qui apparait en linguistique mathématique, en théorie des opérations...

1.3.1

Une relation binaire Str est une stratification sur l'arbre A (muni de l'ordre hiérarchique H) si c'est un préordre total satisfaisant à l'axiome H_{str}.

H_{str} : $\forall$ a, b $\in$ A : H (a, b) $\wedge$ a $\neq$ b $\Rightarrow$ Str (a, b) $\wedge$ Str (b, a), i.e. si a est strictement en dessous de b pour l'ordre hiérarchique, a est strictement en dessous de b pour l'ordre de stratification. L'ensemble A muni des relations H et Str est appelé un arbre stratifié. L'ensemble (A/Q_{str}) quotient de A par la relation d'équivalence associée à la relation de préordre Str (cf 11) est appelé ensemble des niveaux ou ensemble des strates ; cet ensemble est totalement ordonné par l'ordre quotient Str_Q (qui provient de Str, cf 11).

Suivant la remarque qui termine le numéro 11, il est commode de définir l'ordre Str par une application f de A dans R : pour que soit satisfait l'axiome H_{str}, il faut et il suffit que la fonction f appelée indice de stratification - satisfasse à l'axiome H_{ind} :

H_{ind} : $\forall$ a, b $\in$ A : H (a, b) $\wedge$ a $\neq$ b $\Rightarrow$ (f (a) < f (b)) $\wedge$ (f (a) $\neq$ f (b)), i.e. l'indice f doit être une fonction strictement croissante pour l'ordre hiérarchique. L'ensemble A muni de H et de f est appelé un arbre indicé. Sur le graphique de l'arbre, l'indice de stratification est, tout simplement, représenté par l'ordonnée. Pour définir une stratification, on peut au lieu de l'ensemble R, utiliser un ensemble totalement ordonné quelconque T : c'est ce qu'on fait classiquement dans la hiérarchie linéenne avec l'ensemble ordonné des termes :

T = {espèce, ..., règne} ; espèce < genre < ... < règne.

L'ensemble des strates (ou niveaux) de A s'identifie alors avec T. Nous reviendrons sur cet exemple aux numéros 22 et 23.

1.3.2

Quant à l'ordre séquentiel S, on demande qu'il satisfasse aux deux axiomes H_{su} et H_h :

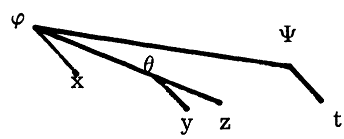
H_{su} : quel que soit a dans Nod (A), Sui (a), muni de l'ordre induit par S est totalement ordonné ; de plus Som (A) est totalement ordonné par S.

H_h : $\forall$ a, b, a', b' $\in$ A : H (a', a) $\wedge$ H (b', b) $\wedge$ S (a, b) $\Rightarrow$ S (a', b') :

On peut dire que l'axiome H_h exprime que les éléments a' et b' héritent de la relation d'ordre séquentielle vérifiée entre a et b ; sur le graphique l'axiome H_h implique que les branches ne se croisent pas entre elles. Aux axiomes H_{su} et H_h, on peut en ajouter un troisième qui achève d'imposer que l'ordre S soit un ordre total : (de façon précise la donnée d'un ordre total sur chacun des ensembles Som (A) et Sui (a, A) a $\in$ Nod (A), définit sur A un ordre total unique S satisfaisant aux axiomes H_h et H_{pr})

$$H_{pr} : \forall a, b \in A : H (a, b) \Rightarrow S (a, b)$$

L'axiome H_{pr}, qui, on va le voir, correspond à la notation préfixe des opérateurs fonctionnels, n'est pas toujours employé. Bornons-nous à un exemple. A l'expression φ (x, θ (y, z), Ψ (t)), on associe un arbre qui a pour noeuds les trois opérateurs fonctionnels φ, θ, ψ et pour éléments terminaux les quatre arguments x, y, z, t : l'ordre hiérarchique est défini en plaçant sur chaque opérateur les variables et les fonctions qui lui servent d'arguments ; l'ordre séquentiel est celui de l'écriture fonctionnelle, à condition d'entendre S (a, b), (b précède a), comme b est à gauche de a (sens de l'écriture latine). On a le graphique de la figure 1" :



φ(x, (θ(y , z), ψ(t))

Figure 1"

Répétons que l'ordre séquentiel ne sera pas employé en taxinomie.

2 - Classifications hiérarchiques

Sous le titre général de classifications hiérarchiques, nous étudions successivement plusieurs notions communément utilisées dans les sciences de la nature.

2.1 - Hiérarchie de parties dans un ensemble

Soit I un ensemble fini. On dit qu'un ensemble A de parties non-vides de I, (A $\subset$ Part I - $\{\emptyset\}$), est une hiérarchie de parties dans I si il satisfait aux deux axiomes H_i et H_u :

H_i : Axiome d'intersection : deux éléments de A, non comparables pour la relation d'inclusion, ont une intersection vide :

$$a, b \in A : a \cap b \in \{a, b, \emptyset\}.$$

H_u : axiome de réunion : tout élément de A qui n'est pas minimal (pour la relation d'inclusion) est la réunion des éléments distincts de lui et inclus dans lui :

$$\forall a \in A : \cup \{b \mid b \in A ; b \neq a ; b \subset a\} \in \{a, \emptyset\}.$$

Il résulte de l'axiome H_i que A, muni pour ordre hiérarchique de la relation d'inclusion, (H (a, b) $\Leftrightarrow$ a $\subset$ b), est un arbre. En effet, soit u et v deux prédécesseurs (pour l'inclusion) de a $\in$ A : l'intersection u $\cap$ v est non-vide, car elle contient a ; par conséquent u et v sont comparables (i.e. on a soit u $\subset$ v, soit v $\subset$ u). Désormais, nous considérerons donc A comme un arbre.

Ex Des axiomes H_i et H_u, il résulte que tout noeud $a \in$ Nod (A) est la réunion disjointe de ses successeurs immédiats : et que tout a de A est la réunion disjointe des éléments terminaux inclus dans lui :

$$\forall a \in A : a = \cup \{t \mid t \in \text{Ter } A ; t \subset a\}.$$

$$\forall t, t' \in \text{Ter } A : t \neq t' \Rightarrow t \cap t' = \emptyset.$$

$$\forall a \in \text{Nod } A : a = \cup \{b \mid b \in \text{Sui } (a)\}.$$

$$\forall a \in \text{Nod } A, \forall b, b' \in \text{Sui } (a) : b \neq b' \Rightarrow b \cap b' = \emptyset.$$

De plus l'arbre A est nécessairement fourchu, car si un noeud $a \in$ Nod (A), n'avait qu'un seul successeur immédiat a', on devrait avoir a = a',pour que a soit réunion de ses successeurs immédiats : or a ne peut être son propre successeur immédiat. Si l'arbre A est binaire (satisfait à l'axiome H_b cf 12) on dit que A est une hiérarchie dichotomique.

On appelle support d'une hiérarchie A de parties, l'ensemble noté sup (A), qui est la réunion de ces parties :

$$\text{sup } (A) = \cup \{a \mid a \in A\}$$

On montre que l'on a :

$$\text{sup } (A) = \cup \{t \mid t \in \text{Ter } (A)\} = \cup \{s \mid s \in \text{Som } (A)\} ;$$

plus précisément, il s'agit de réunions disjointes, autrement dit Som (A) et Ter (A) sont des partitions de sup (A). (On rappelle qu'une partition P d'un ensemble J est famille de parties non vides, deux à deux disjointes de J dont la réunion est J). Quand l'arbre A est connexe, on a simplement : sup (A) = som (A). Si le support de A est I tout entier, on dit que A est une hiérarchie de parties sur I (cf la notation d'application surjective, ou sur); dans le cas général, on parle seulement de hiérarchie dans I, comme nous l'avons fait jusqu'ici.

Une hiérarchie de parties sera dite fine si chacun de ses éléments terminaux est un ensemble à un élément ; ce qu'on énonce en un axiome H_t :

$$H_t : \forall t \in \text{Ter } A, \exists i \in I : t = \{i\}.$$

Dans la pratique, on s'intéresse particulièrement aux hiérarchies fines de sommet I : c'est ce que nous appellerons des hiérarchies totales.

Ex On montrera en exercice qu'un ensemble A de parties non-vides de l'ensemble I est une hiérarchie totale sur I si et seulement si il satisfait à l'axiome H_i et de plus aux deux conditions suivantes :

$$I \in A ; \forall i \in I : \{i\} \in A.$$

Dans le cours de 1965, nous définissons par ces conditions les hiérarchies totales, appelées là, simplement, hiérarchies de parties).

Soit A une hiérarchie totale ; soit $B \subset A$; si B satisfait à l'axiome H_u, c'est aussi une hiérarchie de parties ; mais une telle sous-hiérarchie de A n'est pas en général une hiérarchie totale.

Voici, représenté graphiquement, un exemple de hiérarchie totale de parties sur l'ensemble]8] des entiers de 1 à 8.

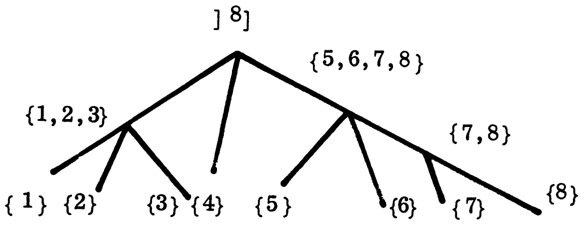


Figure 2

On notera qu'il est superflu d'indiquer sur les éléments non terminaux (ou noeuds), de quelle partie de I il s'agit, puisque, chaque noeud est la réunion des terminaux inclus dans lui (cf supra).

2.2 - Hiérarchie dans un ensemble

Soit A un arbre : à A est associé une hiérarchie fine A^{Ter} sur l'ensemble Ter (A) : à chaque élément a de A on associe la partie $a^{Ter} \subset$ Ter (A) définie par :

$$a^{Ter} = \{t \mid t \in \text{Ter}(A) : H(t, a)\}$$

(en particulier, si $t \in$ Ter (A), t^{Ter} n'est autre que $\{t\}$, ensemble dont t est le seul élément).

On pose :

Ex $$A^{Ter} = \{a^{Ter} \mid a \in A\}.$$

On vérifiera que A^{Ter} est bien une hiérarchie fine de support Ter (A).

Il faut noter que A et A^{Ter} ne se correspondent biunivoquement que si l'arbre A est fourchu sinon, soit a un noeud ayant un successeur immédiat unique a' : on aura : $a^{Ter} = a'^{Ter}$.

Plus généralement, étant donné un arbre A, un ensemble I, et une surjection d'une partie I' de I sur l'ensemble Ter (A) on a une hiérarchie A^I de parties de I (dont le support est I') en posant :

$$a^I = \{i \mid i \in I' ; H(\varphi(i), a)\}$$
$$A^I = \{a^I \mid a \in A\} ;$$

On a imposé que φ soit une surjection pour qu'aucune des parties a^I ne soit vide comme ci-dessus A^I et A se correspondent biunivoquement si et seulement si A est un arbre fourchu. Soit $t \in$ Ter (A) : t^I n'est autre que l'image réciproque $\varphi^{-1}(t)$: on peut aussi, par abus de langage écrire : $a^I = \varphi^{-1}(a)$.

Soit A un arbre ; supposons associé à chaque élément a de A une partie non vide a^I d'un ensemble fini I de telle sorte que soient satisfaites les conditions suivantes :

H' : Si a et b ne sont pas comparables pour l'ordre hiérarchique de A, $a^I \cap^i b^I$ est vide :

$$\forall\ a, b \in A : \neg (H (a, b) \wedge H (b, a)) = a \cap b = \emptyset.$$

$$H'_u : \forall\ a \in \text{Nod } A : a^I = \bigcup\ \{b^I \mid b \in \text{Sui } (a)\}.$$

(plus précisément, il résulte de l'axiome H'_i que les b^I forment une partition de a^I). Notons φ l'application sur Ter (A) de l'ensemble $I' = \bigcup\ \{t^I \mid t \in \text{Ter } A\}$, qui sur chacun des t^I est constante et égale à t. Alors on voit que $a^I = \varphi^{-1}(a)$.

Lorsque à chaque élément a d'un arbre A est associée une partie a^I d'un ensemble I, de telle sorte que soient satisfaits les axiomes H'_i et H'_u, on dit qu'on s'est donné une hiérarchie dans I ; (on a une hiérarchie sur I si la réunion des a^I est I). Une hiérarchie dans I peut être donnée de manière unique par un triple (A, I, φ), où φ est une surjection d'une partie I' sur l'ensemble Ter A des éléments terminaux de l'arbre A.

Pour représenter graphiquement une hiérarchie dans I, on représentera l'arbre A, et on indiquera sous chaque $t \in$ Ter (A), la partie t^I qui lui correspond : ces indications suffisent car :

$$\forall\ a \in A : a^I \cup \{t^I \mid t \in \text{Ter } A \ ; H (t, a)\}$$

Considérons comme exemple la hiérarchie linnéenne, A est l'ensemble des noms de toutes les divisions taxinomiques ; comme : les végétaux, les échinodermes, les oiseaux, les tubulidentés (ou, en latin, tubulidentata), les félidés, félis (nom d'un genre), ou panthera leo (nom d'une espèce) : Ter (A) est l'ensemble de tous les noms d'espèces. L'ensemble I n'est pas un ensemble au sens mathématique usuel du terme, car c'est un ensemble potentiel soit qu'on le considère du point de vue de l'extension (ensemble de tous les vivants qui auront existé..), soit du point de vue de la compréhension (ensemble des descriptions de tous les vivants possibles..). Cependant, on peut dire que φ est l'application qui à tout individu i fait correspondre son espèce. L'arbre A n'est pas fourchu parce qu'il existe des classes monotypes, des éléments a ayant un successeur immédiat unique.

2.3 - Hiérarchie stratifiée

Soit A une hiérarchie de parties dans (ou sur) un ensemble I : A, muni d'une stratification (compatible au sens du numéro 131 avec son ordre hiérarchique qui est l'ordre d'inclusion des parties de I) est appelé une hiérarchie stratifiée de parties dans (ou sur) I.

Soit de même (A, I, φ) une hiérarchie dans (ou sur) I : (A, I, φ), muni d'une stratification est appelé hiérarchie stratifiée dans (ou sur) I.

2.3.1

Comme nous l'avons dit au numéro 13, on peut définir une stratification Str sur A, par une application f de A dans un ensemble totalement ordonné T. La hiérarchie linnéenne de base utilise l'ensemble :

T = {espèce, genre, famille, ordre, classe, embranchement, règne}.

Dire que a est, (e.g.), un ordre, s'exprime dans notre formalisme par l'équation : f (a) = "ordre". Outre les conditions de compatibilité avec la structure arborescente de A (cf n° 13), la fonction f de la hiérarchie linnéeenne de base satisfait à la condition suivante :

$$H_{lin} : \forall\ t \in \text{Ter } (A) : f\ (\text{Pred } (t)) = T$$

Ex Ici, (cf n° 12), Pred (t) n'est autre que l'ensemble des a tels que a^I contient t^I. L'axiome H_{ind} (cf 131) selon lequel la fonction f est strictement croissante, implique que f met en correspondance biunivoque T et Pred (t): donc si t est un élément terminal de A, f (t) = "espèce" (t est une espèce) et l'on trouve au dessus de t tous les niveaux de la terminologie linnéenne de base T. Parfois, l'on fait usage d'une terminologie T' plus fine que la terminologie linnéenne de base : l'ensemble T' est construit à partir de T en utilisant les préfixes sur et sous, et aussi de nouveaux termes tels que tribu (cf Simpson, cité au n° 13 de la leçon sur les sciences de la nature). On sait que, afin de ne pas multiplier les classes intermédiaires on n'impose plus alors que l'application f' (de A dans T') prenne dans tout Pred (t) toutes les valeurs possibles ; on impose seulement les conditions suivantes:

$$H'_{lin} : \forall\ t \in \text{Ter } (A) : T \subset f'\ (\text{Pred } (t)) \subset T'$$

$$H''_{lin} : \forall\ a,\ b,\ b' \in A : \{b',\ b\} \subset \text{Sui } (a) \Rightarrow f'\ (b) = f'\ (b').$$

Autrement dit, d'une part, au dessus de chaque noeud terminal (qui n'est pas nécessairement une espèce : on peut avoir f' (t) = "sous espèce") on doit trouver tous les niveaux principaux (ceux de T) ; d'autre part tous les successeurs immédiats d'un noeud donné reçoivent même nom par f'. Pour un exemple d'emploi des niveaux supplémentaires, on se reportera à la leçon C.S.N., figure 2.

2.3.2

En taxinomie numérique on utilise le plus communément les hiérarchies indicées par une fonction f à valeur réelle. Qu'il s'agisse d'une hiérarchie de partie A, ou d'une hiérarchie (A, I, φ) la fonction f n'est assujettie d'abord qu'à la condition de compatibilité H_{ind} (n° 131). Toutefois, particulièrement dans le cas où A est une hiérarchie totale de parties de T, (i.e. $I \in A$ et $\{\{i\} \mid i \in I\} \subset A$), on s'intéresse aux fonctions d'indice f satisfaisant aux conditions supplémentaires suivantes :

$$H_{diam} : f\ (I) < 1\ ;\ f\ (\{\{i\} \mid i \in I\}) = \{0\}.$$

f est nulle sur les classes $\{i\}$ réduites à un élément, et vaut au plus 1 sur l'ensemble I tout entier ; ailleurs, (étant croissante), elle prend des valeurs comprises entre 0 et f (I). On dira alors que l'indice est un indice de diamètre, et on le désignera par la lettre d (les classes réduites à un seul élément ont pour diamètre 0 etc..) ; la fonction s = (1 - d) qui est au contraire maxima et égale à 1 pour les parties $\{i\}$ mesure la similitude interne des éléments d'une classe et est appelée indice de similitude. Sur la figure 2', on a représenté une hiérarchie indicée (par un indice de diamètre) sur l'ensemble]5] des 5 premiers entiers. Dans la leçon C.A.H. numéro 13'

13', nous donnons deux exemples de programme de tracé d'arbres indicés. Les graphiques fournis par la machine, notamment d'après le programme du numéro 13', différent de ceux dessinés ici ; mais on y reconnait sans peine les mêmes informations.

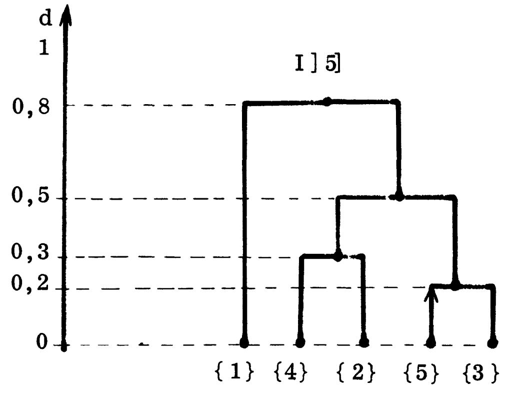


Figure 2'

Soit A une hiérarchie de parties indicée par une fonction réelle d : à tout nombre réel x est associée une hiérarchie A_x ensemble des a de A dont l'indice est inférieur ou égal à x. Considérons, pour fixer l'intuition, le cas d'une hiérarchie totale A sur I, et d'un indice de diamètre d.

$$A_x = \{a \mid a \in A \ ; \ d(a) < x\} \ ;$$

A_x (en tant qu'ensemble de parties) est une fonction croissante de x .

On a $A_o = \{\{i\} \mid i \in I\}$: A est la partition la plus fine de I ;

$A_1 = A_{d(I)} = A$.

Ex Pour $x > 0$, Som (A_x) est une partition de l'ensemble I, d'autant plus fine que x est plus petit ; disons que, en un sens qui sera précisé au numéro 3, la partition Som (A_x) est l'ensemble des classes de A maximales parmi celles qui sont du diamètre inférieur ou égal à x. La hiérarchie A_x (et donc aussi la partition Som (A_x) ne varie que lorsque x traverse une valeur prise par l'indice d pour $a \in A$.

Dans l'exemple de la figure 2' on a les partitions Som (A_x) par des coupes horizontales du graphique :

$0 < x < 0{,}2$: Som $A_x = \{\{i\} \mid i \in]5]\}$;

$0{,}2 < x < 0{,}3$: Som $A_x = \{\{1\}, \{4\}, \{2\}, \{5,3\}\}$;

$0{,}3 < x < 0{,}5$: Som $A_x = \{\{1\}, \{4,2\}, \{5\ 3\}\}$;

$0{,}5 < x < 0{,}8$: Som $A_x = \{\{1\}\ \{4,2,5,3\}\}$;

$0{,}8 < x$: Som $A = \{]5]\}$.

(Dans le cours de 1965, figure le même exemple ; mais on y note P_{1-x} la partition notée ici Som A_x : c'est à dire que l'on fait usage de l'indice de similitude 1 - d ; et l'arbre est dessiné avec l'ensemble I en bas).

L'intérêt de la famille des sous-hiérarchies A_x, vient de ce que un grand nombre d'algorithmes taxinomiques construisent les hiérarchies à partir du bas : à partir de la partition la plus fine Som A_x, on forme par réunion successive des classes dont le diamètre est de plus en plus grand.

Notons ici que pour Jardine et Sibson (1967) une hiérarchie taxinomique (hiérarchie pouvant comporter des classes monotypes ; notion que nous formalisons par les hiérarchies (A, I, φ) sur un ensemble I) est défini par la donnée sur un ensemble I d'une suite finie de partitions $\{P_0, P_1, \ldots, P_n\}$ d'autant moins fines que l'indice en est plus élevé. (Rappelons qu'une partition P' du même ensemble si :

$$\forall p \in P \ \exists p' \in P' : p \subset p',$$

Ex i.e. si toute classe p de P est incluse dans une classe p' de P'). Une division taxinomique est pour eux une paire (a, b) formée d'un entier h (le rang) et d'une partie a de I qui est une classe de P_h. Ainsi la même partie a peut être considérée à deux niveaux différents sous deux noms différents; on retrouve exactement les hiérarchies linnéennes de base : définies au numéro 231, à ceci près qu'à l'ensemble T de termes est substituée la suite ordonnée [n] des entiers de 0 à n. Toutefois, le système de J.J. & S. nous paraît présenter deux inconvénients : d'une part, il oblige à considérer uniformément au dessus de chaque classe de rang 0, n classes de rang 1, 2.. n, ce qui ne correspond pas exactement à ce que font les taxinomistes (cf supra : Simpson, et axiome H'_{lin}) ; d'autre part, et c'est là le plus fâcheux, la structure arborescente de la hiérarchie n'est pas en évidence dans le formalisme mathématique de J.J. & S.

Les constructions faites dans ce numéro sont évidemment possibles dans le cas général d'une hiérarchie stratifiée dans un ensemble I, (A, I, φ ; Str). Nous ne les répéterons pas puisque dans la pratique on a avantage à calculer avec un indice ; il faut seulement veiller à ce que, dans la mesure du possible, les constructions faites ne dépendent que de la stratification définie par cet indice (soient invariantes si on substitue à f un nouvel indice f' qui est une fonction croissante de f). Notons seulement qu'à la condition d $(\{i\}) = 0$ imposée aux indices de diamètre correspond la condition suivante pour la stratification :

$$H'_{diam} : \forall i \in I, \forall a \in A : \mathrm{Str}\ (\{i\}, a).$$

i.e. toutes les parties à un élément d'une hiérarchie totale ont le même niveau, qui est nécessairement inférieur à celui de tout noeud a $\in$ Nod (A).

2.4. - Complété binaire d'un arbre

Par complété binaire d'un arbre A, on entend un surarbre binaire B coterminal à A i.e. un B tel que :

A $\subset$ B ; l'ordre hiérarchique A est induit par celui de B ;

Som A = Som B ; Ter A = Ter B.

On peut montrer (cf e.g. J.P. Benzécri 1968 décomposition des mesures et fonctions etc...) que tout arbre fourchu admet un complété binaire. Sans donner ici de démonstration en forme, indiquons que la construction de P se fait au niveau de chaque noeud non binaire de A : soit a $\in$ Nod A ; Card Sui (a) = n > 2 : on intercalera entre a et les n éléments de Sui (a), n - 2 noeuds supplémentaires dont nous nous bornons à indiquer les deux dispositions possibles pour le cas d'un noeud ternaire.

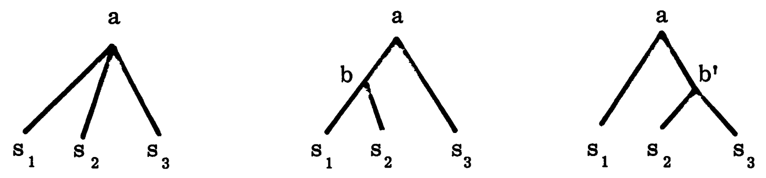


Figure 3

La même construction fournit une hiérarchie dichotomique complétée d'une hiérarchie quelconque de parties : en effet au noeud supplémentaire b on associera la partie $s_1 \cup s_2$; à b' on associera $s_2 \cup s_3$.

Dans le cas d'une hiérarchie indicée, ou stratifiée, on pourra choisir librement le niveau des noeuds supplémentaires introduits, sous réserve que soient satisfaites les conditions de compatibilité énoncées au numéro 13 : e.g. b devra être strictement au dessous de a et strictement au dessus de s_1 et s_2. Mais si l'on manque à satisfaire à cette condition, un artifice de notation intéressant se présente pour les arbres non-binaires. Il est particulièrement simple de noter en machine un arbre binaire : il suffit de ranger comme deux fonctions du numéro de chaque noeud, le numéro de son premier et celui de son second successeur, (le choix de quoi sera premier quoi second étant souvent irrelevant). Avec un arbre non binaire, on rencontre la difficulté de ranger des données de format variable tel noeud aura 3 successeurs, tel autre 7 ; on ne sait a priori quel sera le maximum de Card Sui (a) sur un arbre A que l'on construit comme solution d'un problème etc... Mais considérons un arbre binaire B indicé par une fonction d (a), (croissante) non strictement croissante : on associera naturellement à D son sous-arbre non-binaire A obtenu en supprimant les noeuds situés au même niveau qu'un ou plusieurs de leurs prédécesseurs.

$$A = B - \{b \mid b \in B \ ; \ \exists \ a \in \text{Pred (b)} : d \ (b) = d \ (a)\}.$$

Réciproquement en construisant un surarbre indicé tel que B d'un arbre A (non binaire) donné, on aura une notation binaire commode d'un arbre quelconque. Ainsi procèdent les algorithmes (cf A M n° 1).

3 - Ordonnances et hiérarchies :

Il arrive souvent dans la pratique, que sans disposer d'une formule naturelle pour calculer des distances entre éléments d'un ensemble I, on puisse déterminer de façon quasi-incontestable ce que devraient être les inégalités entre ces distances. D'où l'intérêt de la notion d'ordonnance, que nous mettons ici en relation avec celle de hiérarchie stratifiée.

3.1 - Ordonnance et indice de distance

Soit I un ensemble fini ; on appelle préordonnance sur I, une relation de préordre Ord sur I × I satisfaisant aux deux axiomes suivants :

$$H_{point} : \forall i_1,\ i_2,\ i_3 \in I : Ord\ ((i_1,\ i_1),\ (i_2,\ i_3)).$$

$$H_{sym} : \forall i_1,\ i_2 \in I \quad : Ord\ ((i_1,\ i_2),\ (i_2,\ i_1))$$

$$H_{+} : \forall i_1,\ i_2,\ i_3 \in I : Ord\ ((i_2,\ i_3),\ (i_1,\ i_1)) \Rightarrow i_2 = i_3$$

La raison d'être de ces axiomes apparait immédiatement si l'on précise que la relation Ord $((i_1,\ i_2),\ (i_3,\ i_4))$ se lit : i_1 est plus proche de i_2 que i_3 ne l'est de i_4. On voit maintenant que l'axiome H_{point} exprime que tout point est plus proche de soi-même que deux points quelconques ne le sont l'un de l'autre ; tandis que l'axiome H_{sym} exprime la symétrie de la notion de proximité : quels que soient i_1 et i_2, i_1 est aussi proche de i_2 que i_2 l'est de i_1. (Ce dernier axiome ne s'impose pas dans toutes les applications ; le voyage d'aller peut être plus difficile que le voyage de retour... ; mais les distances dissymétriques sont peu étudiées en mathématiques). De la transitivité du préordre et de l'axiome H_{sym} il résulte que l'on a :

Ex $$\forall\ i_1,\ i_2,\ i_3,\ i_4 \in I : Ord\ ((i_1,\ i_2,\ i_3,\ i_4)) = \ldots$$

$$\ldots\ Ord\ ((i_2,\ i_1),\ (i_3,\ i_4)) \wedge Ord\ ((i_1,\ i_2),\ (i_4,\ i_3)).$$

Enfin l'axiome H_+ correspond à la condition imposée à une distance d (i, i') d'être positive si $i \neq i'$: il est impossible que i soit aussi proche de i_3, dont il est distinct que i_1 est proche de lui-même. Si on supprime la condition H_+ on est conduit à définir les semi-ordonnances, analogues aux semi-normes (éventuellement nulles pour un vecteur non-nul) utilisées en analyse ; nous y reviendrons au numéro 32.

Notons J l'ensemble des parties à deux éléments de l'ensemble I. La donnée d'une préordonnance sur I équivaut à la donnée d'un préordre sur J (préordre qui n'est assujéti à aucune condition). On notera de la même manière :

$$Ord\ ((i_1,\ i_2),\ (i_3,\ i_4) \Leftrightarrow Ord\ (\{i_1,\ i_2\},\ \{i_3,\ i_4\})\ ;$$

(où on suppose $i_1 \neq i_2$ et $i_3 \neq i_4$). Dans la suite, on considérera la préordonnance comme une relation soit sur I^2, soit sur J. Si Ord est un préordre total, un ordre, un ordre total sur J, on parlera de préordonnance totale, d'ordonnance, d'ordonnance totale.

Il est clair que la donnée sur l'ensemble fini I d'une structure d'espace métrique définit sur l'ensemble I une ordonnance totale (ou une préordonnance totale dans le cas particulier où il a des distances égales) : il suffit de poser :

$$\forall\ i_1,\ i_2,\ i_3,\ i_4 \in I \ldots$$

$$\ldots\ Ord\ ((i_1,\ i_2),\ (i_3,\ i_4)) \Leftrightarrow d\ (i_1,\ i_2) < d\ (i_3,\ i_4),$$

(où on a noté d la fonction de la distance). Evidemment une ordonnance donnée est compatible avec une infinité de structures métriques (de fonctions de distance). Cependant, rappelons ici pour mémoire un remarquable résultat découvert par R.N. Shepard (1962). Soit I une partie de l'espace R^p muni de la métrique euclidienne usuelle : la connaissance de l'ordonnance de la figure I, et de la dimension p détermine en général approximativement les rapports entre les distances qui séparent les points de I ; et ce avec une précision d'autant meilleure que le cardinal (nombre de points) de I est plus élevé. Autrement dit, la connaissance de l'ordonnance permet de reconstruire approximativement la figure I à une similitude près. Nous avons donné une formule approchée simple (utilisant la loi du χ^2 à p degré de liberté) pour recalculer (à un coefficient d'échelle près) les distances en fonction de leur rang ordinal ; ce qui permet dans le cas p = 2 (figure plane) de reconstruire rapidement au compas (en tatonnant) une figure d'une dizaine de points dont on ne connait que l'ordonnance.

Appelons indice de distance sur l'ensemble I une fonction d sur $I \times I$ satisfaisant aux conditions suivantes :

$$H_+^i : \forall\, i, i' \in I : i \neq i' \Rightarrow d\,(i, i') > 0 \;;\; d\,(i, i') \neq 0.$$

$$H_o^i : \forall\, i \in I : d\,(i, i) = 0$$

$$H_s^i : \forall\, i, i' \in I : d\,(i, i') = d\,(i', i)$$

On voit que la définition d'un indice de distance est moins restrictive que celle d'une distance, en ce que l'on n'impose pas que l'indice satisfasse à l'inégalité triangulaire :

$$\forall\, i, i', i'' \in I : \delta\,(i, i'') < \delta\,(i, i') + \delta\,(i', i'') \;;$$

Il est clair que la donnée d'un indice de distance définit une ordonnance sur I ; l'ordonnance ne change pas si l'on substitue à l'indice d'un nouvel indice d' = f (d) où f est une fonction strictement croissante (et telle que f (0) = 0). Puisque nous nous intéressons ici aux seules propriétés ordinales l'inégalité du triangle (qui n'est pas conservée par les changements d'indices d' = f (d) et qui, quel que soit l'indice d sur un ensemble fini I, est vérifiée par d' = f (d), pour un choix convenable de f) ne nous importe pas.

Si on impose de plus à l'indice d de prendre ses valeurs dans l'intervalle (0,1) (ce que nous ferons dans la suite, sauf indication contraire) on peut dire que la fonction s = 1 - d, qui prend elle aussi ses valeurs dans (0,1) est un indice de similitude tel que $\forall\, i \in I : s\,(i, i) = 1$ (similitude maxima). On retrouve utilisés dans la littérature soit des indices de similitude, soit des indices de distance : il est évidemment facile de passer d'une notion à l'autre. Et en fait ces indices ne servent qu'à définir de façon commode, une ordonnance sur I :

$$\text{Ord}\,((i_1, i_2), (i_3, i_4)) \Leftrightarrow d\,(i_1, i_2) \Leftrightarrow d\,(i_3, i_4) \quad \ldots$$

$$\ldots : \Leftrightarrow s\,(i_3, i_4) < s\,(i_1, i_2)\,.$$

3.2 - Ordonnances et partitions

Soit I un ensemble fini muni d'une partition P. Il est naturel d'associer à P une notion de similitude à l'intérieur de l'ensemble I : deux éléments appartenant à une même classe de P étant plus semblables entre eux que deux éléments qui appartiennent à des classes distinctes. D'où les définitions suivantes pour un indice de distance d et une préordonnance Ord associées à P. L'indice d est défini par les deux conditions :

$$C_1 : \forall i, i' \in I : d(i, i') = 1 \Leftrightarrow \rceil [\exists p \in P : i \in p ; i' \in p];$$

$$C_2 : \forall i, i' \in I : (i \neq i') \wedge (d(i, i') \neq 1) \Leftrightarrow d(i, i') = 0,5 ;$$

(i.e. d (i, i') vaut 1 si et seulement si il n'existe pas de classe p de P qui contienne à la fois i et i' ; d (i, i') vaut 0,5 si i et i' sont différents mais appartiennent à une même classe ; enfin (inutile de le préciser), selon l'axiome H_o^1, d (i, i) = 0 quel que soit $i \in I$).

A l'indice de distance d correspond, comme il est général, une préordonnance totale définie par la formule déjà donnée plus haut :

$$\forall i_1, i_2, i_3, i_4 \in I : \text{Ord}((i_1, i_2), (i_3, i_4)) \Leftrightarrow d(i_1, i_2) < d(i_3, i_4).$$

Dans un beau travail qui, ayant attiré notre attention sur les problèmes de taxinomie, se trouve à l'origine de ce cours, S. Regnier utilise la fonction d' caractéristique de la partition P : d' (i, i') vaut 1 quand d (i, i') vaut 1 ; et d' (i, i') vaut 0 quand d (i, i') vaut 0 ou 0,5 (autrement dit d', à l'encontre de d, est nulle pour i et i' différents mais appartenant à une même classe). Il importe de noter ici que la fonction d' a l'inconvénient de n'être pas un indice de distance, puisque d' (i, i') peut être nul pour $i \neq i'$. On peut, il est vrai, dire que d' est un indice de semi-distance, et que la relation de préordre correspondante est une semi-ordonnance ; mais celà complique un peu les définitions axiomatiques. A titre d'exercice, nous donnons ici les axiomes modifiés ; mais dans la suite du cours nous n'utiliserons jamais ces axiomes modifiés.

De même qu'en analyse on utilise des semi-normes (qui peuvent être nulles pour un vecteur non nul), on peut définir des indices de semi-distance. Un tel indice satisfait aux axiomes H_o^i et H_s^i, à une forme affaiblie de l'axiome H_+^i :

$$H'^i_+ : \forall i, i' \in I : d(i, i') \geqslant 0,$$

(on demande seulement que d (i, i') soit positif ou nul, même si $i \neq i'$), et de plus à l'axiome H'^i_e suivant :

$$H'^i_e : \forall i, i', i'' \in I : d(i, i') = 0 \Rightarrow d(i, i'') = d(i', i'').$$

Ex Cet axiome implique que, comme il est naturel, la relation d (i, i) = 0 (i est à une distance nulle de i') est une relation d'équivalence ; et que sur l'ensemble quotient de I par cette relation la fonction d induit un véritable indice de distance (Si i est à une distance nulle de i', alors i et i' sont à une même distance de tout autre élément i'' : dans le cas d'une semi-norme, cette propriété résulte de l'inégalité du triangle, mais ici il faut l'imposer par un axiome supplémentaire).

On définit de même une semi-ordonnance sur un ensemble I. Il faut que la semi-ordonnance définisse une ordonnance (ou une préordonnance) sur le quotient de I par une relation d'équivalence convenable. C'est ce que l'on peut exprimer en disant que la semi-ordonnance Ord (considérée comme une relation de préordre sur I × I) doit en plus des axiomes H_{point} et H_{sym} satisfaire à l'axiome H'_e :

$$H'_e : \forall\ i, i', i'' \in I : \text{Ord}\ ((i, i'), (i, i)) \Rightarrow \text{Ord}\ ((i, i''), (i, i''))$$

Ex En effet il est clair d'abord que la relation Q_{Ord} (notée en abrégé Q), définie par :

$$Q\ (i) = Q\ (i') \Leftrightarrow \text{Ord}\ ((i, i'), (i, i)),$$

est une relation d'équivalence sur I, relation que l'axiome H_{point} permet d'écrire sous une forme plus symétrique :

$$Q\ (i) = Q\ (i') \Leftrightarrow \forall\ i_o \in I : \text{Ord}\ ((i, i'), (i_o, i_o)),$$

ou encore sous la forme équivalente :

$$Q\ (i) = Q\ (i') \Leftrightarrow \exists\ i_o \in I : \text{Ord}\ ((i, i'), (i_o, i_o)).$$

De plus l'axiome H'_e permet de montrer que la relation Ord est compatible avec la relation Q_{Ord} au sens suivant : Soit (i_1, i'_1), (i_2, i'_2), (i_3, i'_3), (i_4, i'_4) quatre paires d'éléments deux à deux équivalents ($Q\ (i_1) = Q\ (i'_1)$ etc..) : alors on a l'équivalence logique :

$$\text{Ord}\ ((i_1, i_2), (i_3, i_4)) \Leftrightarrow \text{Ord}\ ((i'_1, i'_2), (i'_3, i'_4))$$

Il en résulte que Ord définit une préordonnance sur le quotient I/Q_{Ord}.

3.3 - Ordonnances et hiérarchies stratifiées

Soit I un ensemble fini ; soit A une hiérarchie totale de parties sur I (i.e. on suppose que la hiérarchie A a pour sommet I et contient pour éléments minimaux toutes les parties de A réduites à un seul élément : Ter A = $\{\{i\} \mid i \in I\}$; cette restriction n'est pas génante et elle est commode) ; soit d un indice de diamètre sur A (i.e. un indice de stratification, cf numéro 131, nul pour les parties à un élément). On va définir sur I un indice de distance associé à d, et noté également d (puisqu'aucune confusion ne semble à craindre). Il apparaitra que d est une véritable distance, et même, cf numéro 4, une distance ultramétrique.

La distance d (i, i') est le diamètre de la plus petite partie de la hiérarchie A qui contienne à la fois i et i' : pour définir cette distance par une formule, il est commode de définir une application (notée a) de I × I dans A :

$$\forall\ i, i' \in I : a\ (i, i') = \inf \{a \mid a \in A\ ;\ i \in a\ ;\ i' \in a\}$$

(i.e. a (i, i') est la plus petite partie contenant i et i'). On a alors :

$$\forall\ i, i' \in I : d(i, i') = d(a(i, i')).$$

On peut encore définir directement la distance d par une seule formule en posant :

$$\forall\ i, i' \in I : d(i, i') = \dots .$$

$$\dots \inf \{x \mid x \in R\ ;\ \exists\ a \in A : i \in a\ ;\ i' \in a\ ;\ d(a) = x\}$$

Bornons-nous à vérifier ici l'inégalité ultramétrique (plus forte que l'inégalité triangulaire) sur laquelle nous reviendrons au numéro 4. Il faut montrer que :

$$\forall\ i, i', i'' \in I : d(i', i'') \leqslant \sup\{d(i, i'), d(i, i'')\},$$

(où on note sup (u, v), le plus grand des nombres u et v). En effet considérons a (i, i') et a (i, i'') : tous deux contiennent i et appartiennent à la hiérarchie A ; d'après l'axiome H_i des hiérarchies (cf n° 21) l'un des deux ensembles a (i, i') ou a (i, i'') contient l'autre, et par conséquent contient à la fois i' et i'' ; l'inégalité annoncée en résulte.

Ex A titre d'exercice, nous donnons ici le tableau de la distance d calculée sur l'exemple de la figure 2'.

	1	2	3	4	5
1	0	0,8	0,8	0,8	0,8
2	0,8	0	0,5	0,3	0,5
3	0,8	0,5	0	0,5	0,2
4	0,8	0,3	0,5	0	0,5
5	0,8	0,5	0,2	0,5	0

Ex Voici un autre exercice ; soit P une partition de l'ensemble I ; A (P) la hiérarchie totale sur I définie par $A = \{I\} \cup P \cup \{\{i\} \mid i \in I\}$; soit d l'indice de diamètre sur A (P) défini par : d (I) = 1, d (P) = {0,5};$\forall\ i \in I$; d ({i}) = 0 ; alors la distance sur I associé à l'indice de diamètre d, n'est autre que celle définie au numéro 32.

Supposons maintenant défini par la hiérarchie A (totale sur I) un ordre de stratification Str (tel que toutes les parties réduites à un élément appartiennent à la même strate) : l'ordre Str peut être défini par une infinité d'indices de diamètres qui ne diffèrent entre eux que par une transformation croissante, et auxquels correspondent sur I des distances (ultramétriques) définissant toute la même préordonnance totale. Cette préordonnance, que nous noterons Ord (Str) (ou en abrégé Ord) peut être définie directement, sans recourir à un indice de diamètre : on posera :

$$\forall\ i_1, i_2, i_3, i_4 : \mathrm{Ord}\ ((i_1, i_2), (i_3, i_4)) \Leftrightarrow \mathrm{Str}\ (a(i_1, i_2), a(i_3, i_4)).$$

(où a (i, i') a le même sens que ci-dessus) ; intuitivement on peut dire que la distance entre i et i' est au niveau de a (i, i'). La préordonnance totale Ord peut encore se définir en une seule formule, sans recourir à la notation a (i, i') :

Ex $$\forall\ i_1,\ i_2,\ i_3,\ i_4 : \mathrm{Ord}\ ((i_1,\ i_2,\ i_3,\ i_4)) \Leftrightarrow \ldots$$

$$\ldots\ \exists\ a \in A : [\{i_1,\ i_2\} \subset a\ ;\ \forall\ a' \in a : \{i_3,\ i_4\} \subset a' \Rightarrow \mathrm{Str}\ (a,\ a')]$$

(i.e. il existe un a contenant i_1 et i_2 et tel que tout a' contenant i_3 et i_4 soit à un niveau supérieur à celui de a).

Remarquons pour finir que l'on peut aussi parler de distance et de préordonnance totale définies sur I par la donnée d'une hiérarchie indicée ou stratifiée sur I (et non plus d'une hiérarchie de parties) : il suffit, dans les constructions ci-dessus, de remplacer la définition de a (i, i') par (cf n° 22 pour la notation a^I) :

$$a\ (i,\ i') = \inf \{a \mid a \in A\ ;\ i \in a^I\ ;\ i' \in a^I\}.$$

Dans la suite de ce cours, en taxinomie numérique, nous utilisons presque exclusivement des hiérarchies de parties.

4 - STRUCTURE ULTRAMETRIQUE ET STRUCTURE METRIQUE :

4.1 - Résumé de géométrie ultramétrique

Une distance d sur un ensemble I (fini ou infini) est dite ultramétrique si elle satisfait à l'axiome H_{ult} :

$$H_{ult} : \forall\ i,\ i',\ i'' \in I : d\ (i',\ i'') \leqslant \sup \{d\ (i,\ i'),\ d\ (i,\ i'')\}\ ;$$

cet axiome qui peut s'énoncer : dans un triangle tout côté (dans ce numéro nous entendons par triangle : ensemble de trois points, ou sommets ; et par côté distance entre deux de ces points) est inférieur ou égal au plus grand des deux autres, est plus fort que l'axiome triangulaire usuel. L'axiome H_{ult} a été, croyons-nous, posé pour la première fois vers 1930 par M. Krassner dans des recherches arithmétiques (cf exemples in fine).

L'axiome H_{ult} est encore équivalent à l'énoncé géométrique suivant : tout triangle est soit équilatéral, soit isocèle ; et s'il est isocèle c'est la base qui est le plus petit côté (elle est plus petite que chacun des deux côtés égaux). Pour démontrer ce résultat, considérons un triangle {i, i', i''} dont le plus grand côté est, e.g., (i', i'') : d'après H_{ult} l'un au moins des deux autres côtés (i, i') ou (i, i'') a même longueur que (i', i'') ; et le troisième côté s'il n'est pas égal à (i', i'') lui est inférieur. Réciproquement si le triangle {i, i', i''} a au moins un second côté égal à (i', i''), l'axiome H_{ult} est bien vérifié.

Rappelons la définition de la boule ouverte B_o (i, ρ) et de la boule fermée B_F (i, ρ) de centre i et de rayon ρ (ces définitions sont celles usitées dans tout espace métrique) :

$$B_0 (i, \rho) = \{i' \mid i' \in I ; d (i, i') < \rho ; d (i, i') \neq \rho \}$$

$$B_F (i, \rho) = \{i' \mid i' \in I ; d (i, i') \leqslant \rho\}$$

On va montrer que l'on a :

$$\forall i' \in B_0 (i, \rho) : B_0 (i', \rho) = B_0 (i, \rho) ;$$

$$\forall i' \in B_F (i, \rho) : B_F (i', \rho) = B_F (i, \rho) ;$$

Autrement dit : tout point d'une boule (ouverte ou fermée) peut être pris pour centre de celle-ci. Démontrons, e.g. la proposition dans le cas des boules fermées : il suffit d'établir que l'on a l'implication :

$$d (i, i') < \rho \Rightarrow [\forall i'' \in I : d (i, i'') \leqslant \rho \Leftrightarrow d (i', i'') \leqslant \rho]$$

or celà résulte immédiatement de l'inégalité ultramétrique H_{ult}

Ex Soit alors B une partie d'un espace ultramétrique, dont, sans précision de centre ni de rayon, on sait qu'elle est une boule : proposons-nous de chercher un centre et un rayon pour B. Il résulte d'abord de ce qui précède que le centre est indéterminé : pour déterminer le rayon ρ on peut donc partir d'un centre quelconque $i \in B$ et chercher pour quelle valeurs de ρ : $B = B_0 (i, \rho)$ ou $B_F (i, \rho)$. Comme on le voit sur des exemples où I est un ensemble fini, il se peut que l'on ait $B (i, \rho) = B (i, \rho')$ pour des valeurs ρ' formant un intervalle; nous nous intéresserons donc au plus petit nombre ρ qui puisse convenir. On voit, en se bornant au seul cas utilisé dans la suite que si I est fini on a : $B = B_F (i, \rho)$ ou :

$$\rho = \sup \{d (i, i') \mid i' \in B\} = \sup \{d (i, i') \mid i, i' \in B\} ;$$

en d'autres termes en géométrie ultramétrique on doit prendre pour rayon d'une boule B ce que l'on appelle classiquement le diamètre de l'ensemble B.

Il est maintenant clair que, en géométrie ultramétrique, si deux boules B et B' (dont chacune peut être soit ouverte soit fermée) ont une intersection non vide, l'une des deux boules est incluse dans l'autre. En effet soit i un point de l'intersection $B \cap B'$: i peut être considéré comme le centre à la fois de B et de B' ; l'on est ramené à montrer que de deux boules concentriques l'une est incluse dans l'autre ; or celà est vrai dans tout espace métrique.

En géométrie métrique, on a l'inégalité suivante, vraie quels que soient les n + 1 points $i^\circ, \ldots, i^h, i^n$:

$$d (i^\circ, i^n) \leqslant \Sigma \{d (i^{h-1}, i) \mid h = 1, \ldots, n\}.$$

(cf le principe bien connu : la ligne droite (i° i) est plus courte que toute ligne brisée). En géométrie ultramétrique, on a une inégalité plus forte :

$$d (i^\circ, i^n) \leqslant \sup \{d (i^{h-1}, i) \mid h = 1, \ldots, n\},$$

cette inégalité peut s'énoncer : la distance qui sépare les deux extrémités i° et i^n d'une chaine est au plus égale au plus long de ses maillons ($i^{h-1} i^h$).

L'inégalité se démontre par récurrence sur n : pour n = 2, on a l'axiome H_{ult} ; supposons l'inégalité démontrée pour n - 1 on a :

$$d\,(i°,\ i^n) \leqslant \sup\,\{d\,(i°,\ i^{n-1}),\ d\,(i^{n-1},\ i^n)\}$$

$$\leqslant \sup\,\{\sup\,\{d\,(i^{h-1},\ i^h) \mid h = 1,\ \ldots,\ n-1\},\ d\,(i^{n-1},\ i^n)\}$$

$$\leqslant \sup\,\{d\,(i^h,\ i^{h-1}) \mid h = 1, \ldots, n\}\ ;$$

c'est l'inégalité annoncée.

L'inégalité ultramétrique sous la forme "dans un triangle tout côté est inférieur ou égal au plus grand des deux autres" se généralise naturellement aux préordonnances. On posera l'axiome H°_{ult} :

$$H^{\circ}_{ult} : \forall\ i,\ i',\ i'' \in I : \mathrm{Ord}\ ((i',\ i''),\ (i,\ i'))\ \vee\ \mathrm{Ord}\ ((i',\ i''),\ (i,\ i'')).$$

Ex On montre que si l'indice de distance d définit sur I une préordonnance ultramétrique (satisfaisant à H°_{ult}), d est nécessairement une véritable distance ultramétrique. (Car il vérifie H_{ult}). On voit que l'axiome H_{ult} est purement ordinal, ce que l'on exprimera encore comme suit. Soit f une fonction réelle positive, strictement croissante telle que f (0) = 0, soit d une distance ultramétrique : f (d) est aussi une distance ultramétrique. Réciproquement si d est une distance ultramétrique, tout indice de distance d' définissant la même ordonnance que d est une distance ultramétrique de la forme d' = f (d), (où f satisfait aux conditions déjà dites).

Ex Terminons par deux exemples classiques de distance ultramétrique Soit p un nombre premier ; soit n un nombre entier ; notons ordre $_p$ (n) l'exposant (entier positif ou nul) de p dans la décomposition en facteurs premiers de n ; de même si r = n/m est un nombre rationnel notons ordre$_p$ (r) = ordre $_p$ (n) - ordre $_p$ (m). Alors sur l'ensemble Q des nombres rationnels la fonction :

$$d\,(r,\ r') = (1/2) \uparrow \text{ordre}_{\,p}\,(r - r'),$$

(où on a noté a $\uparrow$ b pour a^b) définit une distance ultramétrique appelée distance p - adique.

Soit S (x) une série formelle : S (x) = $\Sigma\ \{a_n\ x^n \mid n \in N\}$; On appelle ordre de S (n) le rang de son premier terme non nul :

$$\text{ordre}\ (S) = \inf\,\{n \mid n \in N\ ;\ a_n \neq 0\}$$

(la série identiquement nulle est donc d'ordre infini). On a une distance ultramétrique entre séries formelles en posant :

$$d\,(S,\ S') = (1/2) \uparrow \text{ordre}\ (S - S').$$

4.2 - Hiérarchie indicée et structure ultramétrique

On va montrer qu'il est équivalent de se donner sur un ensemble fini I soit une métrique ultramétrique, soit une hiérarchie totale indicée de parties. Nous utiliserons les notations suivantes (justifiées par les résultats de géométrie rappelés au n° 4.1) pour un ensemble fini I muni d'une distance ultramétrique d :

$$\text{Boules } (I, d) = \{B \mid B \subset I ; \exists \, i \in I, \exists \, \rho \in I^{+} : B = B_f \, (i, \rho)\} ;$$

sur Boules (I, d) est définie la fonction (positive ou nulle) rayon (qu'on peut aussi appeler diamètre) :

$$\text{ray } (B) = \text{diam } (B) = \sup \{d \, (i, i') \mid i \in B, i' \in B\} ;$$

On a vu que l'on a :

$$\forall \, B \in \text{Boules}, \forall \, i \in B : B = B_F \, (i, \text{ray } (B)).$$

Soit maintenant sur l'ensemble fini I une hiérarchie totale A de parties, munie d'un indice de diamètre d : on a montré au numéro 3.3 que d définit sur I une distance ultramétrique, notée aussi d. On a les relations :

$$A = \text{Boules } (I, d)$$

$$\forall \, a \in A : d \, (a) = \text{ray } (a)$$

Ex En effet, soit $i \in I$, $\rho \in R^{+}$: la boule fermée de centre i et de rayon ρ n'est autre que la plus grande partie a de A, contenant i, et telle que d (i) $\leqslant \rho$; et l'on a bien ray (a) = d (a), car : $\forall \, i \in a, \exists \, i' \in a : a \, (i, i') = a$; soit par exemple a non réduit à un point : dans la hiérarchie A, Sui (a) est une partition de a : il suffit de prendre i' dans une autre classe de cette partition que celle qui contient i).

Ex Réciproquement soit sur l'ensemble fini I une distance ultramétrique d (à laquelle on peut imposer la restriction de prendre ses valeurs dans l'intervalle (0,1). L'ensemble Boules (I, d) est ensemble de parties de I, qui est une hiérarchie totale sur I ; et la fonction ray (B) est un indice de diamètre sur cette hiérarchie. En effet, on voit d'abord que l'on a les inclusions :

$$I \in \text{Boules } (I, d) \; ; \; \{\{i\} \mid i \in I\} \subset \text{Boules } (I, d)$$

(I est la boule de rayon maximum ; et les {i} sont les boules de rayon nul) ; il reste donc pour établir que l'on a bien une hiérarchie totale (cf n° 2.1, exercice) à montrer que Boules (I, d) satisfait à l'axiome H_i : or c'est bien là une propriété de géométrie ultramétrique rappelée au numéro 4.1 (l'intersection de deux boules est vide ou est l'une des deux boules). Pour établir que ray est un indice de diamètre il suffit (cf n° 2.3 et n° 1.3.1, axiome H_{ind}) de vérifier qu'il est une fonction croissante sur l'ensemble des boules : or il est clair que si b est une boule incluse dans b' et distincte de b', ray (b) est strictement inférieur à ray (b'). Enfin l'indice de diamètre ray, redonne bien (par la définition du n° 3.3) la distance d qui a servi à le définir.

On établit de même qu'il est équivalent de se donner sur un ensemble fini I soit une hiérarchie totale stratifiée (où les parties à un élément {i} sont toutes à la même strate), soit une préordonnance totale ultramétrique (cf n° 4.1). Il est possible de démontrer cette équivalence en raisonnant uniquement sur les relations de préordre Strat et Ord ; mais il est plus simple de se ramener aux résultats que nous venons de démontrer. En effet à une préordonnance ultramétrique, correspond une distance ultramétrique, défini à un automorphisme croissant de la demi-droite R^+ près, et de même à une hiérarchie stratifiée correspond un indice de diamètre défini lui aussi à une transformation croissante de R^+ près.

4.3 - Ultramétriques associées à une métrique

Lorsqu'on entreprend d'établir une classification des éléments d'un ensemble fini I, on commence souvent par utiliser les informations disponibles (description des éléments de l'ensemble I, etc.) pour calculer un indice de distance. Or nous venons de voir que la notion de hiérarchie indicée totale ne se distingue pas de celle de distance ultramétrique. Ce qui permet d'énoncer comme suit, en termes géométriques, le problème de la classification : transformer, au prix d'une déformation minima, un indice de distance ultramétrique.

Nous traiterons ailleurs ce problème, en étudiant les critères de classification et les algorithmes. Ici nous donnons seulement quelques exemples d'ultramétriques naturellement associées à un indice de distance, et nous soulignerons que les constructions faites ne reposent que sur des propriétés ordinales (peuvent être faites sur des préordonnances).

4.3.1 - L'ultramétrique inférieure maxima

Soit I un ensemble fini ; D une famille de distances ultramétriques sur D dont nous supposerons (e.g.) qu'elles prennent leurs valeurs dans l'intervalle (0,1) ($\forall$ i, i' $\in$ I, $\forall$ d $\in$ D : d (i, i') $\leqslant$ 1) : alors on va montrer que la fonction $d_{\sup D}$ définie ci-dessous est une distance ultramétrique sur I :

$$\forall\ i,\ i' \in I : d_{\sup D}\ (i,\ i') = \sup\ \{d\ (i,\ i')\ |\ d \in D\}.$$

Il nous suffit de vérifier que $d_{\sup D}$ satisfait à l'axiome $H_{u1\ell}$ i.e., que, quels que soient i, i', i'' :

$$\sup\ \{d\ (i',\ i'')\ |\ d \in D\} \leqslant \sup\ \{\sup\ \{d\ (i,\ i')\ |\ d \in D\},$$

$$\sup\ \{d\ (i,\ i'')\ |\ d \in D\}\}\ ;$$

Ex le calcul du sup dans le membre de droite pouvant se faire en regroupant autrement les quantités d (i, i') et d (i, i''), on est amené à démontrer l'inégalité :

$$\sup\ \{d\ (i',\ i'')\ |\ d \in D\} \leqslant \sup\ \{\sup\ \{d\ (i,\ i'),\ d\ (i,\ i'')\}\ |\ d \in D\}\}\ ;$$

or cette dernière inégalité résulte immédiatement de ce que les fonctions d définissent des distances ultramétriques.

Soit maintenant d un indice de distance sur l'ensemble I. Considérons avec Jardine & Sibson (1967) l'ensemble, noté inf (d), des ultramétriques inférieures ou égales à d en tant que fonction sur l'ensemble I × I (une ultramétrique d' appartient à inf (d) si et seulement si :

$$\forall\ i, i' \in I : d'(i, i') \leqslant d(i, i')).$$

La distance d_{uiM} (d)=$d_{\sup\ \inf(d)}$(définie comme ci-dessus) est elle-même une ultramétrique partout inférieure ou égale à d ; d' est l'ultramétrique inférieure à d maxima.

(J.J. & S., 1967, et Roux, 1968, l'appellent l'ultramétrique sous-dominante).

(On peut dire, en termes imagés, que d_{uiM} est la meilleure approximation ultramétrique de d, par le dessous). Il est intéressant de noter avec J.J. & S. que l'ultramétrique d_{uiM} dépend continument de d.

Nous donnerons ailleurs la méthode utilisée par Roux pour calculer d_{uiM} (d) ainsi qu'un autre algorithme taxinomique (e.g. Johnson 1967) qui conduit à la hiérarchie indicée définie par d_{uiM} (d). Donnons seulement ici une définition par chaines de l'ultramétrique inférieure maxima. Notons ici $C^n(i, i')$ l'ensemble des chaines à n maillons allant de I dans i à i' :

$$C^n(i, i') = \{i^h \mid h \in [n]\} \mid \forall\ h \in [n] : i^h \in I\ ;\ i^\circ = i^n\ ;\ i^{n} = i'\}$$

où [n] est la suite 0,1, ... n). Soit $i^{[n]} = \{i^h\}$ un élément on définit le saut maximum saut $(i^{[n]})$:

$$\text{saut}\ (i^{[n]}) = \sup\ \{d\ (i^{h-1}, i^h) \mid h \in\ ]n]\}$$

où]n] est la suite 1, ..., n)

Alors la distance $d_{uiM}(i, i')$ est la plus petite valeur du saut sur l'ensemble des chaines (à un nombre quelconque de maillons) reliant i à i' :

$$d_{uiM}(i, i') = d_{saut}(i, i') = \inf\ \{\text{saut}\ (i^{[n]}) \mid n \in N\ ;\ i_{[n]} \in C^n(i, i')\}$$

Pour établir que $d_{uiM} = d_{saut}$ il suffit de montrer que d_{uiM} est inférieur ou égal à d_{saut}, et que d_{saut} est une ultramétrique inférieure à d.

D'abord d_{saut} est inférieur à d : il suffit de considérer $i^{[1]} = \{i, i'\}$, on a saut $(i^{[1]}) = d(i, i')$. Pour montrer que d_{saut} est une ultramétrique, considérons deux chaines $i^{[n]} \in C(i', i)$ et $i^{[m]} \in C(i, i'')$ telles que :

$$d_{saut}(i', i) = \text{saut}\ (i^{[n]})\ ;\ d_{saut}(i, i'') = \text{saut}\ (i^{[m]})$$

(il existe de telles chaines parce que I est un ensemble fini) ; alors la chaine $i^{[n+m]}$ obtenue en mettant bout à bout $i^{[n]}$ et $i^{[m]}$ va de i' à i'' et on a :

$$d_{saut}(i', i'') \leqslant \text{saut}\ (i^{[n+m]}) = \sup\ \{\text{saut}\ (i^{[n]}),\ \text{saut}\ (i^{[m]})\}$$

$$d_{saut}(i', i'') \leqslant \sup\ \{d_{saut}(i', i),\ d_{saut}(i, i'')\}$$

c'est bien l'égalité ultramétrique. (En résumé, la distance minima qu'il faut être capable de sauter pour aller de i' à i'' en prenant un chemin quelconque, est inférieure ou égale à celle qu'il faut sauter si l'on va de i' à i'' en passant par i).

Reste à voir que d_{uiM} et, plus généralement, toute ultramétrique d' inférieure à d est inférieure à d_{saut} : cela résulte immédiatement de ce que (cf. n° 4.1) :

$$\forall\ i^{[n]} \in C^n (i, i') : d' (i, i') \leqslant \sup \{d' (i^{h-1}, i^h) \mid h = 1, \ldots, n\}$$

$$\leqslant \sup \{d (i^{h-1}, i^h) \mid h \in [n]\} = \text{saut } (i, i')$$

4.3.2. - Les ultramétriques supérieures minimales

Soit encore d un indice de distance sur I. Considérons, toujours avec J.J. & S., l'ensemble sup (d) des ultramétriques d' qui sont supérieures à d (i.e. $\forall$ i, i' $\in$ I : d' (i, i') $\geqslant$ d (i, i')) : on pourrait désirer chercher dans sup (d) un minimum, qui sera la meilleure approximation ultramétrique de d par le dessus. Mais on voit sans peine (cf. J.J. & S.,) sur un exemple que sup (d) peut ne pas contenir d'élément minimum : soit en effet ;

$$I = \{1, 2, 3\} ; d (1, 2) = 3 ; d (1, 3) = 2 ; d (2, 3) = 1 ;$$

il faut, sans faire décroître aucun côté, faire de I un triangle isocèle, dont la base soit le plus petit côté. D'après le côté choisi pour base, les ultramétriques d' supérieures à d se répartissent en trois classes :

$$\sup_1 (d) = \{d' \mid d' (1, 2) = d' (1, 3) \geqslant 3 ; 1 \leqslant d' (2, 3) \leqslant d' (1, 2)\}$$

$$\sup_2 (d) = \{d' \mid d' (1, 2) = d' (2, 3) \geqslant 3 ; 2 \leqslant d' (1, 3) \leqslant d' (1, 2)\}$$

$$\sup_3 (d) = \{d' \mid d' (1, 3) = d' (2, 3) \geqslant 3 ; 3 \leqslant d' (1, 2) \leqslant d' (1, 3)\}$$

l'intersection de deux de ces classes correspondant aux triangles équilatéraux (des trois distances égales entre elles et supérieures ou égales à 3). On voit que l'on a :

$$\inf \sup_1 (d) = d'_1 : d'_1 (1, 2) = d'_1 (1, 3) = 3 ; d'_1 (2, 3) = 1 ;$$

$$\inf \sup_2 (d) = d'_2 : d'_2 (1, 2) = d'_2 (2, 3) = 3 ; d'_2 (1, 3) = 2 ;$$

toute instance d' appartenant à la classe $\sup_3$ (d) est supérieure à la fois à d'_1 et d'_2 : mais il n'existe pas d'ultramétrique qui soit à la fois inférieure à d'_1 et d'_2 et supérieure à d.

A défaut d'avoir dans sup (d) un élément minimum unique on peut y rechercher des éléments minimaux (i.e. des d' tels qu'il n'existe pas d'ultramétrique inférieure strictement à d', en tant que fonction sur I $\times$ I, et supérieure à d) qui dans notre exemple simple sont d'_1 et d'_2. C'est ce que fait d'une façon particulière, Johnson qui n'a d'ailleurs pas vu le problème posé par J.J. & S., et se laisse seulement guider par le désir de construire des classes de diamètre minimum. En bref, Johnson rend tout triangle isocèle en lui donnant pour base son plus petit côté : ce qui, dans l'exemple ci-dessus, conduit à l'ultramétrique supérieure minimale d'_1. Appliqué à un triangle isocèle dont la base est le plus grand côté, l'algorithme de Johnson ne conduit pas à une solution unique ; cette indétermination est liée à ce que, comme l'indiquent J.J. & S., (cf. notre figure 4), on ne peut trouver d'application continue F qui à une distance d, fasse correspondre une ultramétrique d' qui soit un élément minimal de sup (d) (par continuité on entend que d' considéré comme vecteur de R^J :

$$d' = \{d'(i, i') \mid \{i, i'\} \in J\}$$

dépend continument de d, considéré aussi comme vecteur).

Nous reverrons au numéro 4.3.4 ces délicates questions de continuité.

4.3.3. - La hiérarchie d'Apresjan

Soit toujours I muni d'indice de distance d. Dans des recherches de classification en linguistique, Ju. D Apresjan considère les parties a de I satisfaisant à la condition :

$$\forall\ i, i', i'' \in a : i \in a \wedge i' \in a \wedge d(i, i') \leqslant d(i, i') \Rightarrow i'' \in a$$

i.e. si a contient i et i', il contient la boule de centre i et de rayon d (i, i'): la raison de cette condition est que si i' est dans la même classe que i, i'' qui ne s'éloigne pas de i plus que ne le fait, i' doit aussi être dans la même classe que i. On va voir que, relativement à la métrique d elle-même, toute classe d'Apresjan a peut être considérée comme une boule de centre l'un quelconque de ses points. En effet soit i $\in$ I ; notons :

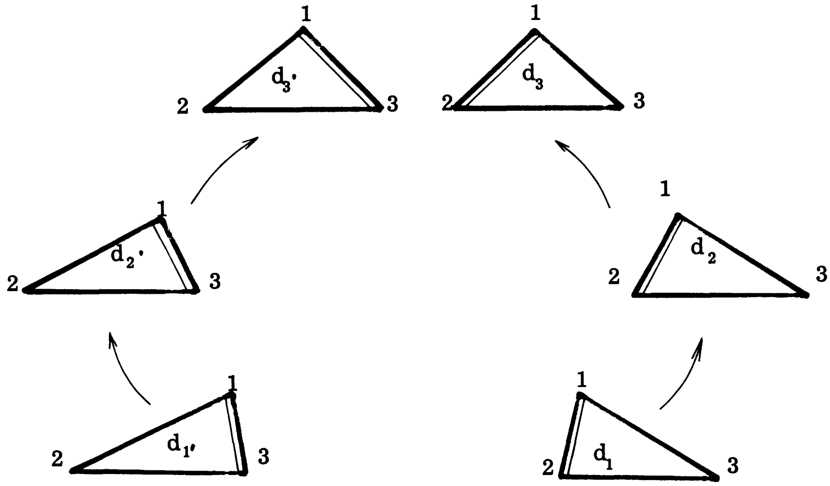


Figure 4

Chacun des triangles de cette figure représente l'ensemble {1, 2, 3} muni d'une métrique (d_1, $d_{1'}$ etc...) correspondant aux longueurs des côtés dessinés. L'un des côtés est un trait double : il correspond au droit d'une ultramétrique supérieure minimale d' pour laquelle il est la base : par exemple à la métrique d_2 est associée l'ultramétrique supérieure minimale d'_2:

$$d'_2(1, 2) = d_2(1, 2) \; ; \; d'_2(1, 3) = d'_2(2, 3) = d_2(2, 3).$$

Ex On veut montrer qu'il est impossible de choisir d' fonction continue de d sur l'ensemble de tous les triangles. En effet : on a nécessairement $d_1 = d'_1$ et $d_{1'} = d'_{1'}$; car pour les triangles isocèles dont la base est petite, il y a une seule ultramétrique supérieure minimale, qui n'est autre que la métrique donnée. Par continuité on est ainsi conduit à mettre sur un triangle isocèle dont la base est grande chacune des deux ultramétriques supérieures minimales possibles d'_3 et $d'_{3'}$. .

$$\rho\ (a\ ;\ i) = \sup \{d\ (i,\ i') \mid i' \in a\}\ ;$$

$$B\ (i,\ \rho\ (a,\ i)) = \{i' \mid i' \in I\ ;\ d\ (i,\ i') \leqslant \rho\ (a,\ i)\}.$$

Par définition de ρ (a, i), B (i, ρ (a, i)) contient a ; et par définition des classes d'Apresjan, si $i \in a$, la boule B (i, ρ (a, i)) est contenue dans a : donc a est une boule de centre l'un quelconque de ses points. Cependant contrairement à ce qui se passe pour les boules ultramétriques, ρ (a, i) peut varier avec i). De cette propriété on déduit (comme au n° 4.1) que l'ensemble A_p des classes d'Apresjan est une hiérarchie totale de parties sur I. Mais il est facile de voir sur des exemples que souvent cette hiérarchie se réduit à $\{I\} \cup \{\{i\} \mid i \in I\}$, i.e. ne comporte que des classes triviales. C'est pourquoi Apresjan propose de définir sa hiérarchie non à partir de la distance donnée d, mais à partir de la distance ultramétrique inférieure à d maxima (que Apresjan définit par les chaines cf 4.3.1) : il retrouve alors la hiérarchie considérée par J.J. & S., Johnson etc..., car si d est une ultramétrique toute boule est une classe d'Apresjan (et réciproquement cf ci-dessus).

On va définir, sur la hiérarchie d'Apresjan, deux fonctions simples d_d et d_r qui sont des indices de stratification :

$$d_d\ (a) = \sup \{\rho\ (a,\ i) \mid i \in a\} = \sup \{d\ (i,\ i') \mid i,\ i' \in a\}\ ;$$

$$d_r\ (a) = \inf \{\rho\ (a,\ i) \mid i \in I\}\ ;$$

(où ρ (a, i) est défini comme ci-dessus) ; d_d n'est autre que le diamètre de l'ensemble a, et d_r est le plus petit rayon d'une boule fermée contenant a.

Ex On vérifie qu'une fonction est un indice en montrant qu'elle est strictement croissante sur la hiérarchie A_p. Pour d_d la démonstration est simple : soit a', $a \in A_p$, $a' \subset a$; si $d_d\ (a) = d_d\ (a')$ il existe un point i de a' tel que $\rho\ (a',\ i) = d_d\ (a)$; on a donc aussi $\rho\ (a,\ i) = \rho\ (a',\ i)$ et d'après ce qu'on a vu plus haut, a et a' coïncident tous deux avec la boule de centre i et de rayon d_d (a) : ceci suffit à établir que d_d est strictement croissant sur A_p.

Pour d_r, nous ne connaissons pas de démonstration aussi simple.

Montrons d'abord que :

$$\forall\ a \in A_p,\ \rho\ i \in I : \rho\ (a,\ i) = d_r\ (a) \Rightarrow i \in a$$

(si a est une classe d'Apresjan, le minimum de ρ (i, a) ne peut être atteint qu'en un point i de a). En effet soit $i' \in a$; on a :

$$\rho\ (a,\ i') \geqslant \rho\ (a,\ i) \geqslant d\ (i,\ i'),$$

d'où :

$$i \in B\ (i',\ P\ (a,\ i')) \subset a.$$

Ex Soit maintenant a' une classe d'Apresjan incluse dans a ($a' \subset a$), et telle que $d_r(a) = d_r(a')$. Soit $i \in a$; $\rho(a, i) = d_r(a)$; on a donc :

$$B(i, \rho(a, i)) = a.$$

Mais on a aussi : $d_r(a') \leqslant \rho(a', i)$; $\rho(a', i) \leqslant \rho(a, i) = d_r(a)$;

d'où l'on tire que $\rho(a', i) = d_r(a)$ et $a' = B(i, d_r(a)) = a$.

Ce qui établit que d_r est strictement croissante.

Nous avons dit qu'il y a en général fort peu de classes d'Apresjan. En revanche, il semble que ces classes les plus naturelles et que toute hiérarchie, plus fine que A , que l'on pourra construire doive inclure A_p. Dans ce sens voici deux résultats.

Soit $a \in A_p$; pour la distance ultramétrique inférieure maxima d_{uiM}, a est une boule de rayon au plus égal à $d_r(a)$. En effet, soit i tel que $\rho(a,i) = d_r(a)$: on vérifie immédiatement en considérant la chaine i'ii''que deux points quelconques i', i'' $\in$ a, sont à une distance d_{uiM} inférieure ou égale à $d_r(a)$. D'autre part toute chaine reliant un point i de a à un point i_e de I - a, comporte nécessairement d'après la définition des classes d'Apresjan, un maillon supérieur strictement à $d_r(a)$;

$$(\text{si } i' \in a,\ i'' \in I - a, \text{ on a } d_r(a) \leqslant \rho(a, i') < d(i', i'')).$$

Considérons maintenant l'exemple I = {1, 2, 3} du numéro 4.3.2 : avec la métrique d, les classes d'Apresjan sont entre les classes triviales {1}, {2}, {3} et I, le côté {2, 3}. On voit que {2, 3} est une boule pour l'ultramétrique supérieure minimale d'_1 mais non pour d'_2 ; on ne peut donc dire que les classes d'Apresjan se retrouvent dans toute hiérarchie définie par une ultramétrique supérieure minimale. En revanche nous montrerons que A_p est incluse dans toute hiérarchie construite par le programme de Johnson (cf la leçon consacrée aux algorithmes).

4.3.4. - Préordonnances ultramétriques associées à une préordonnance totale

Rappelons les diverses classes de hiérarchies indicées, ou ce qui est équivalent d'ultramétriques, que nous avons associées à une métrique d donnée sur un ensemble fini I :

1) l'ultramétrique inférieure maxima : d_{uiM} ;

2) les ultramétriques supérieures minimales $\{d_{usm}\}$, parmi lesquelles on distingue :

2' Les ultramétriques supérieures minimales de Johnson :

$$\{d_{usm}\}_J \subset \{d_{usm}\}.$$

La classe des ultramétriques supérieures construites par Johnson, contient au moins une métrique (pour d donnée) et peut contenir plusieurs métriques (si les distances d (i, i') ne sont pas toutes différentes deux à deux)

3) Les deux ultramétriques d_{ar} et d_{ad} associées à la hiérarchie d'Apresjan munis de l'indice d_r ou d_d. Toutes ces classes sont invariantes par un au-

tomorphisme croissant f de R^+ (ou plutôt de (0,1) où d prend ses valeurs : (f (0) = 0, (f (1) = 1 : f monotone), au sens suivant : on a, e.g.,

$$\text{si } d^f = f(d) \; ; \; d^f_{uiM} = f(d_{uiM}).$$

Par conséquent on peut, étant donnée une préordonnance totale Ord, y associer une préordonnance ultramétrique Ord_{uiM}, comme on associe d_{uiM} à d, et de même on construira les classes de préordonnance ultramétriques : $\{Ord_{usm}\}$, $\{Ord_{usm}\}$ J et les préordonnances d'Apresjan Ord_{ar} et Ord_{at} Mais le rapport entre, e.g., Ord_{uiM} et Ord ne nous parait plus si clair qu'entre d_{uiM} et d : approcher par le dessous une préordonnance ne veut rien dire ; puisque toute préordonnance peut être représentée par une infinité de métriques différant par des changements d'échelle f. Le point de vue naturel pour comparer à la préordonnance Ord, donnée sur I, une préordonnance ultramétrique Ord' (construite à partir de Ord) est selon nous de considérer Ord et Ord' comme des parties de J × J (J ensembles des paires {i, i'}) à Ord on associe :

$$\{(j, j') \mid j \in J, j' \in J \; ; \; Ord\,(j, j')\}$$

ensemble des paires ordonnées (j, j') de paires (j et j') d'éléments de I, entre lesquels Ord met une inégalité. Nous reviendrons sur ce point de vue dans les leçons consacrées aux critères et aux algorithmes.

On a dit au numéro 4.3.1 que, (cf J.J. & S.).d_{uiM} dépend continument de d : examinons, à titre d'exercice quelle est la propriété correspondante pour Ord_{uiM}, considérée comme fonction de la préordonnance totale Ord. Notons D l'espace de tous les indices de distances (positifs bornés par 1) sur l'ensemble I : D n'est autre que $(0, 1)^J$, J espaces des paires d'éléments de I. Notons F le groupe des automorphismes (correspondances biunivoques croissantes) de (0, 1) : F opère aussi sur D, comme on l'a vu : la distance f (d) est donnée par f (d) (i, i') = f (d (i, i')). Notons ORD l'ensemble (fini) de toutes les préordonnances totales possibles sur I. Il y a une projection naturelle p de D sur ORD qui à un indice d associe la préordonnance totale Ord défini par d ; l'image réciproque p^{-1} (Ord) est un polyèdre défini par les inégalités de l'ordonnance ; ainsi D apparait muni d'une décomposition polyédrale, dont chaque cellule est l'image réciproque d'une préordonnance ; au bord d'une cellule se trouvent les cellules définies en ajoutant des inégalités supplémentaires. Le groupe F opère transitivement sur chaque facette : i.e. si d_1 et d_2 ont même ordonnance, il existe f (non unique d'ailleurs) tel que $f(d_1) = d_2$. Soit maintenant une application φ de D dans D, qui (comme l'application $d \to d_{uiM}$) soit continue et commute avec les opérations de F, au sens suivant :

$$\forall f \in F, \forall d \in D : \varphi\,(f(d)) = f\,(\varphi(d)).$$

L'application φ passe de l'ensemble ORD, (qui est le quotient de D par la projection p) : il existe $\varphi°$, unique application de ORD dans ORD, telle que :

$$\forall d \in D : \varphi°\,(p(d)) = p\,(\varphi(d))$$

(e.g. si φ est l'application $d \to d_{uiM}$ $\varphi°$ est l'application $Ord \to Ord_{uiM}$).

Il est exclu que $\varphi°$, application de l'ensemble fini ORD dans lui-même soit continue : mais $\varphi°$ hérite de φ la propriété suivante : soit $Ord_1, Ord_2 \in ORD$; si la facette $p^{-1}(Ord_2)$ appartient à la fermeture de la facette $p^{-1}(Ord_1)$ la facette $p^{-1}(\varphi(Ord_2))$ appartient à la fermeture de la facette de $p^{-1}(\varphi\, Ord_1)$). Ce qui, au niveau des préordonnances, systèmes d'inégalités, signifie simplement ceci : si l'ordonnance Ord_2 est obtenue à partir de Ord_1 en ajoutant des inégalités supplémentaires (ce qui donnera des égalités là où Ord_1 a des inégalités), $\varphi°(Ord_2)$ s'obtient de même à partir de $\varphi°(Ord_1)$. Nous avons ainsi énoncé une propriété qui au niveau des préordonnances est l'analogue de la continuité.

L'équivalence entre distance ultramétrique et que nous appelons dans ce cours hiérarchie indicée est démontrée par Benzécri (1965 b), J.J. & S (1967) et Johnson (1967) ; J.J. & S. ont étudié les ultramétriques inférieures ou supérieures à une métrique donnée et montré leur intérêt taxinomique ; les ultramétriques correspondant aux hiérarchies construites par Johnson sont ainsi le mieux définies par J.J. & S. ; J.J. & S. ne citent pas Johnson et ne donnent pas d'algorithme, mais leur travail a inspiré un algorithme de Roux. Outre une hiérarchie considérée par J.J. & S., Johnson et Roux, Apresjan en donne une autre que nous n'avons trouvé chez aucun autre auteur.

Regnier et al. s'intéressent à la recherche globale d'une partition s'accordant le mieux avec une distance ; puis suivant une voie indiquée par Benzécri (1965 a) ils adaptent leur algorithme à la recherche d'une partition s'accordant le mieux avec une distance. L'importance de la notion même d'ordonnance apparait dans les remarquables recherches de Shepard (1962), complétées par Benzécri (1964 - 1965) ; Johnson (1967) s'inspire, lui aussi, de Shepard.

Les publications suivantes : F. Benzécri (1965), J.P. Benzécri (1965 ; 1968), Friant (1966), R.D. Guedj (1966), qui sont disponibles au laboratoire, contiennent des applications linguistiques de la théorie des arbres.

BIBLIOGRAPHIE

Ju. D. APRESJAN - Un algorithme pour construire des classes d'après une matrice de distances ; in Mashinnyi perevod : prikladnaja lingvistika : numéro 9 pp. 3 - 18 ; Institut Maurice Thorez, Moscou (1966).

F. BENZECRI - Sur la profondeur des arbres de dérivation ; Rennes 1965.

J.P. BENZECRI - Sur l'analyse factorielle des proximités, in Publications de l'Institut de statistique de l'Université de Paris, (1964 - 1965)

J.P. BENZECRI - Sur les algorithmes de classification, Rennes (1965 a) et Paris, cours ISUP (1965 - 1966).

J.P. BENZECRI - Problèmes et méthodes de la taxinomie, Rennes (1965 b) et Paris, cours ISUP (1965 - 1966).

J.P. BENZECRI - Symbol manipulation and general algebraic structures (1966)

J.P. BENZECRI - Lois de probabilité sur un ensemble produit : Décomposition des mesures et fonctions sur un ensemble muni d'une classification arborescente : ISUP (1968).

F.J. FRIANT - Les langages "context sensitive" ; Thèse Paris (1966) et annales de l'I.H.P., 1967, pp 35 - 120.

R.D. GUEDJ - Grammaires de constituants généraux ; Thèse Paris (1966)

C.J. JARDINE ; N. JARDINE & R. SIBSON - The structure and construction of taxinomic hierarchies ; in Mathematical Biosciences Vol 1, pp 171 - 179 ; (1967) (cité : J.J. & S.)

S.C. JOHNSON - Hierarchical clustering schemes ; in Psychometrika, Vol 12 numéro 3, pp. 241 - 254 ; (1967).

M. ROUX - Un algorithme pour construire une hiérarchie particulière ; Thèse Paris ISUP (1968).

R.N. SHEPARD - Analysis of proximities : Multidimensional scaling with an unknown distance function ; in Psychometrika, Vol 27 pp. 125 - 140 & 219 - 246 ; (1962).

W.F. de la VEGA ; M. RENAUD ; S. REGNIER - Techniques de la classification automatique ; diffusé par le centre de calcul de la Maison des Sciences de l'Homme ; (1964 - 1966).

D.M.Cl. Indice