

I. C. LERMAN
Université de Rennes

6. ANALYSE ORDINALE D'UNE CLASSE D'ECHELLES OU ANALYSE HIERARCHIQUE

I. INTRODUCTION

L'objet de cet article consiste à exposer dans ses grandes lignes et à illustrer sur un exemple réel notre méthode d'analyse hiérarchique.

Cette méthode a pris naissance en 1966, elle a été programmée puis appliquée à différents jeux de données réelles (cf. [8], [9], [10], [11 chap. 8] et [4]).

C'est **improprement** que le terme d'analyse hiérarchique a pu désigner une analyse des données résultant d'une hiérarchie de classifications ; il s'agit, comme le titre l'indique, d'une analyse ordinaire d'une classe d'échelles. Chaque échelle d'attitude qui représente un caractère dont l'ensemble des modalités est totalement ordonné, que nous appellerons également ci-dessous "item total", définit un préordre total sur l'ensemble défini par la population étudiée. Relativement à une même classe d'échelles recouvrant une même "dimension" sous-jacente, le problème de l'analyse hiérarchique unidimensionnelle consiste, à partir du comportement de la population, à ordonner totalement l'ensemble de toutes les modalités non minimales des diverses variables, formant ainsi une échelle fine d'attitude.

Le Psycho-Sociologue posait ce problème à propos d'une classe d'échelles établie a priori, pour laquelle il postulait de l'existence d'une même "variable sous-jacente" aux différentes échelles.

Il est moins contestable de poser le problème de l'analyse hiérarchique unidimensionnelle, relativement à une classe d'échelles reconnue comme définissant une forte tendance comportementale de la population étudiée.

Nous découvrons une telle classe comme sous-tendue par un nœud “significatif” de notre arbre des classifications qui organise par proximité en une structure en classes et sous-classes un vaste ensemble d’items observés sur un échantillon de la population concernée. Par conséquent, dans la pratique, notre méthode de classification hiérarchique intervient de façon préalable à une analyse hiérarchique proprement dite. C’est pour cette raison et au risque de déplacer le centre d’intérêt de l’article que nous chercherons à l’exprimer schématiquement avant l’application sur des données réelles. La méthode de classification hiérarchique est d’ailleurs illustrée par R. Gras 7 ; elle est également illustrée dans [12] où il s’agit d’un exemple réel de pédagogie mathématique. Dans ces deux cas les variables de description peuvent être assimilées à des attributs ; alors qu’il s’agit dans notre problème de la situation où les variables descriptives sont des échelles d’attitude. I. Cohen (cf. [4]) a précisément considéré ce cas dans le cadre d’une vaste enquête de Psycho-Pédagogie de l’enfant, élaborée par le Docteur A. Berge et sa collaboratrice G. Denjean (cf. [2]). Nous reprendrons ci-dessous un aspect des résultats qui concerne la méthode présentée ici.

Les spécialistes de l’analyse hiérarchique se sont longtemps et surtout préoccupés du cas des items dichotomiques (i.e. échelles à deux modalités). Ce travail se situe d’emblée dans le cas d’échelles à nombre quelconque de modalités : chaque échelle étant représentée par un ordre total sur l’ensemble des modalités, le support de l’information est un produit direct d’ordres totaux finis. Un même sommet de ce treillis sera appelé, lorsque nous emprunterons le langage des praticiens de ces méthodes, “patron de réponse” ; patron étant la traduction adaptée du mot anglais “pattern”.

La possibilité d’erreur dans la réponse d’un sujet à un item pose le problème du choix de la “bonne” échelle hiérarchique permettant au “mieux” de ranger totalement l’ensemble des divers comportements élémentaires proposés à travers les différents items. Une première partie de notre travail aboutira à la proposition d’un algorithme ayant un caractère optimal, qui, en un “petit” nombre de pas, détermine l’échelle hiérarchique qui s’ajuste le “mieux” à l’ensemble des patrons de réponse observés. Cet algorithme repose naturellement sur une certaine idée de la “distance” entre deux patrons de réponse et entre un patron de réponse et une échelle hiérarchique ; cette idée de la distance se traduisant mathématiquement au niveau du treillis de représentation. Nous avons proposé trois distances et adapté l’algorithme à deux d’entre elles ; l’adaption à la distance que nous désignons par d_3 (la plus classique) a été effectuée par M. Barbut (cf. [1]). Nous montrons ainsi, par rapport à l’état de la question tel que rapporté dans l’ouvrage de B. Matalon (cf. [13]), comment remplacer le tâtonnement par la recherche systématique et comment distinguer un degré d’erreur dans la réponse d’un sujet à un item par le choix judicieux, tenant compte du problème particulier traité, de l’une des trois distances.

II. Définitions et représentation mathématique

1. DEFINITIONS GENERALES

Nous allons présenter ici, en le précisant, le vocabulaire de l'analyse hiérarchique.

On appelle "item" la donnée d'un ensemble fini de comportements proposés. Une modalité de l'item est un élément de l'ensemble ; c'est-à-dire un comportement proposé. On distingue habituellement les items dichotomiques des items non dichotomiques. Un item est dichotomique, si le nombre de ses modalités est deux, et ne l'est pas si le nombre de ses modalités est strictement supérieur à deux. L'item est dit **présentant k modalités** si le nombre de ses modalités est k.

Nous supposons qu'est définie sur l'item, c'est-à-dire sur l'ensemble des comportements proposés, une relation d'ordre ; cette relation d'ordre est en général une relation de préférence par rapport à un but donné de caractère psychologique social ou autre. Si l'item est totalement ordonné par la relation d'ordre définie, il sera dit **total**.

Nous nous intéressons seulement aux items totaux ; nous ne restreignons pas ainsi la généralité de notre étude parce qu'on peut toujours se ramener à ce cas en réalisant une partition de l'item non total en sous items totaux. Notons $(a_0, a_1, a_2, \dots, a_{k-1})$ la suite des modalités de l'item total a, où on a

$$a_0 < a_1 < \dots < a_j < \dots < a_{k-1},$$

la modalité a_j est dite avoir le "code" j.

Exemples :

Dans le cadre de l'enquête de Psycho-Pédagogie à laquelle nous faisons allusion ci-dessus, la question : "L'enfant a-t-il eu des vomissements motivés par des causes extérieures à l'alimentation : contrariété, émotion,..." définit un item dichotomique dont les deux modalités sont :

$$a_0 = \text{non} \text{ et } a_1 = \text{oui}.$$

D'autre part, la question : "L'enfant parle-t-il de l'école à la maison ?"

	Code
Oui, avec plaisir	<input type="checkbox"/> 0
Oui, incidemment	<input type="checkbox"/> 1
Oui, avec hostilité	<input type="checkbox"/> 2
Non	<input type="checkbox"/> 3

définit un item total à quatre modalités.

Pour cette question qui vise l'affection qu'a un enfant pour son école, une réponse "Non" est jugée par le Psychologue la plus défavorable, car elle reflète un aspect de démission de l'enfant vis-à-vis de son école.

2. VARIABLE SOUS JACENTE A UN ITEM TOTAL

Soit un item total a présentant k modalités : $a_0 < a_1 < \dots < a_{k-1}$.
On admet l'existence d'une échelle dense (*) de valeurs α que nous représenterons par un demi-axe orienté de gauche à droite. La variable α est liée à l'item a de la façon suivante : Un sujet donné répond à l'item a par la modalité a_j si et seulement si la valeur de la variable, mesurée sur le sujet, est comprise entre deux bornes α_j et α_{j+1} , **indépendantes du sujet** ; $\alpha_j \leq \alpha < \alpha_{j+1}$.

On pose $\alpha_0 = 0$ et $\alpha_k = +\infty$; d'où

$$0 = \alpha_0 < \alpha_1 < \dots < \alpha_j < \dots < \alpha_k = \infty$$

L'item a définit ainsi sur l'axe de α une subdivision $(\alpha_0, \alpha_1, \dots, \alpha_{k-1}, \alpha_k)$



La variable α est dite sous-jacente à l'item a .

Bien qu'elles nous soient inconnues, les bornes α_j sont supposées fixées par la donnée de l'item. Par la suite nous donnerons un rapide aperçu de la manière de déterminer statistiquement les nombres α_j dans le cadre d'un modèle probabiliste qui suppose *continue* l'échelle α .

3. ENSEMBLE D'ITEMS TOTAUX SE REFERANT A UNE MEME VARIABLE

3.1. Introduction

Pour fixer les idées considérons un ensemble de deux items totaux a et b présentant respectivement h et k modalités :

$$a_0 < a_1 < \dots < a_i < \dots < a_{(h-1)} \quad \text{et} \quad b_0 < b_1 < \dots < b_j < \dots < b_{(k-1)}$$

Désignons a priori par α la variable sous-jacente à l'item a et par β , celle sous-jacente à l'item b . On a, avec des notations que l'on comprend à partir du paragraphe précédent,

$$0 = \alpha_0 < \alpha_1 < \dots < \alpha_i < \dots < \alpha_{h-1} < \alpha_h = \infty \quad \text{et} \\ 0 = \beta_0 < \beta_1 < \dots < \beta_j < \dots < \beta_{k-1} < \beta_k = \infty$$

(*) Une échelle est dite ici dense dans le sens que Cantor donnait à ce mot pour les types d'ordres : entre deux éléments distincts quelconques x et y de l'échelle, on peut toujours en insérer un troisième z (et par suite une infinité).

Exemples d'échelles denses : l'ensemble ordonné \mathbb{Q} des rationnels, l'ensemble ordonné \mathbb{D} des nombres décimaux.

Définition

Les deux items a et b sont dits relatifs à une même variable sous-jacente si on a pu définir une relation d'ordre total R sur la réunion des deux items, compatible avec les relations d'ordre définies respectivement sur chacun des items et pour laquelle : $a_i R b_j \iff \alpha_i \leq \beta_j$.

Pratiquement, la relation R est une relation de préférence vis-à-vis d'un but donné ; celui pour lequel ont précisément été établis chacun des deux items a et b.

Chacun des deux items définit, sur le demi-axe de la variable commune que nous noterons alors γ , une subdivision. Celle définie par l'item a est $(0, \alpha_1, \alpha_2, \dots, \alpha_{h-1}, \infty)$ et celle définie par b est $(0, \beta_1, \beta_2, \dots, \beta_{k-1}, \infty)$.

L'item a sera dit plus *fin* que l'item b si la subdivision définie par a est plus fine que celle définie par b ; en d'autres termes si pour tout j, il existe i tel que $\alpha_i = \beta_j$.

3.2. Composition d'items relatifs à une même variable sous-jacente

Soient deux items a et b ; reprenons ceux que nous venons de considérer ci-dessus. Désignons par $(0, \gamma_1, \gamma_2, \dots, \gamma_{h+k-2}, \infty)$ la suite ordonnée croissante des valeurs α_i et β_j (voir figure 1).

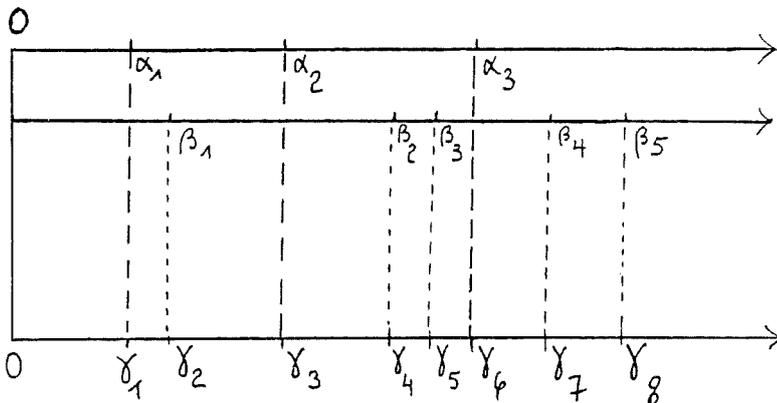


Figure 1

Au couple d'items (a,b) faisons correspondre l'item $c = a*b$ défini comme suit : si $\gamma_l = \alpha_i$ (resp. β_j), la modalité de code l de l'item c est a_i (resp. b_j). L'item c est dit **composé** des items a et b. La loi de composition ainsi définie entre items totaux se référant à une même variable sous-jacente est associative et commutative.

En n'admettant pas d'erreur possible de la part d'un individu donné, la condition nécessaire et suffisante pour que cet individu réponde à l'item c par la modalité c_l est que la valeur de la variable sous-jacente γ

mesurée sur le sujet soit comprise entre γ_ℓ et $\gamma_{\ell+1}$. Dans le cas où on a pour cette valeur mesurée $\alpha_i \leq \gamma < \alpha_{i+1}$ et $\beta_j \leq \gamma < \beta_{j+1}$, alors $\gamma_\ell = \sup(\alpha_i, \beta_j)$ et $\gamma_{\ell+1} = \inf(\alpha_{i+1}, \beta_{j+1})$ car, les deux intervalles $[\alpha_i, \alpha_{i+1}[$ et $[\beta_j, \beta_{j+1}[$ ayant une intersection non vide, il n'existe pas de points α_ℓ ou β_m entre $\sup(\alpha_i, \beta_j)$ et $\inf(\alpha_{i+1}, \beta_{j+1})$.

Considérons le cas particulier important de la composition d'items dichotomiques. Soient b^1, b^2, \dots, b^{m-1} , $(m-1)$ items dichotomiques se référant à une même variable sous-jacente. L'item $c = b^1 * b^2 * \dots * b^{(m-1)}$ est un item total à m modalités en général ; la subdivision définie par l'item b_j étant $(0, \beta_j, \infty)$, celle définie par c est $(0, \beta_1, \beta_2, \dots, \beta_j, \dots, \beta_{m-1}, \infty)$

Exemple :

b^1, b^2, \dots, b^{m-1} peuvent être définis à partir de $(m-1)$ questions d'un problème de mathématique s'échelonnant par difficulté croissante : la j -ème question fait intervenir toute l'aptitude requise pour la solution de la question $(j-1)$. La modalité codée 0 de l'item b_j est une réponse incorrecte à la question n° j . L'item $c = b^1 * b^2 * \dots * b^{(m-1)}$ contient la suite totalement ordonnée des modalités suivantes :

	Code
réponse incorrecte à la question n° 1	→ 0
réponse correcte à la question n° 1 et incorrecte à la question n° 2	} → 1
réponse correcte à la question n° 2 et incorrecte à la question n° 3	} → 2
⋮	
⋮	
⋮	
réponse correcte à la question $(m-1)$	→ $(m-1)$

Réciproquement, on peut aisément définir une décomposition d'un item total présentant m modalités en $(m-1)$ items dichotomiques totaux. En effet, à partir de l'item a présentant m modalités totalement ordonnées : $a_0 < a_1 < \dots < a_j < \dots < a_{m-1}$, on définit la suite $(b^1, b^2, \dots, b^j, \dots, b^{m-1})$ des items dichotomiques où l'item b^j consiste en la partition en deux classes R_j et S_j de l'ensemble des comportements proposés par l'item a , où $R_j = \{a_0, a_1, \dots, a_{j-1}\}$ et $S_j = \{a_j, a_{j+1}, \dots, a_{m-1}\}$. La modalité de code 0 (resp. 1) pour l'item dichotomique b_j correspond à une réponse à l'item a par un comportement appartenant à R_j (resp. S_j). Si $(0, \alpha_1, \dots, \alpha_j, \dots, \alpha_{m-1}, \infty)$ est la subdivision définie par l'item a , celle définie par l'item b_j est $(0, \alpha_j, \infty)$.

Nous allons abandonner maintenant cette approche avant de la reprendre plus richement mais très brièvement (cf. § III.3) dans le cadre d'une hypothèse générale de la distribution de la variable sous-jacente

sur la population étudiée et d'un modèle probabiliste cohérent de l'erreur dans la réponse d'un sujet à un item. Cette construction nous permettra alors de déterminer statistiquement les points de la subdivision sur l'axe de la variable sous-jacente à l'item.

4. TREILLIS DE REPRESENTATION

En vue de l'analyse du comportement d'une population donnée de sujets relativement à un but fixé, on suppose établi un ensemble fini d'items totaux

$$\{a^j / 1 \leq j \leq m\}.$$

Chaque a^j est un ensemble fini totalement ordonné ; à $\{a^j / 1 \leq j \leq m\}$, associons le produit fini d'ensembles finis totalement ordonnés

$$\prod_{1 \leq j \leq m} a^j = a^1 \times a^2 \times \dots \times a^m, \quad (1)$$

qui sera l'ensemble de représentation.

A a^j associons l'ensemble $\Omega_j = \{0, 1, 2, \dots, r_j\}$ des codes de ses modalités ($\text{card}(a^j) = r_j + 1$) que nous représentons géométriquement par l'intervalle commençant $[0, r_j]$ de \mathbb{N} dont nous ne retiendrons que la structure ordinale.

A l'ensemble $\prod_{1 \leq j \leq m} a^j$ se trouve associé l'ensemble produit

$$\Omega = \prod_{1 \leq j \leq m} \Omega_j = \Omega_1 \times \Omega_2 \times \dots \times \Omega_m, \quad (2)$$

dont les éléments sont les points

$$\omega = (\omega_1, \omega_2, \dots, \omega_j, \dots, \omega_m)$$

où ω_j est le code d'une modalité de l'item a^j ; $\omega_j \in \Omega_j$, $0 \leq \omega_j \leq r_j$.

Ω sera dans ces conditions représenté géométriquement par le pavé de \mathbb{N}^m : $[0, r_1] \times [0, r_2] \times \dots \times [0, r_m]$.

Sur Ω , on définit la relation d'ordre produit $\omega \leq \omega'$:

$$(\omega_1, \dots, \omega_j, \dots, \omega_m) \leq (\omega'_1, \omega'_2, \dots, \omega'_j, \dots, \omega'_m) \Leftrightarrow (\forall j), \omega_j \leq \omega'_j \quad (3)$$

ω' succède à ω si $\omega \leq \omega'$ et si, pour tout j , $\omega_j = \omega'_j$, sauf pour un seul j_0 où $\omega'_{j_0} = \omega_{j_0} + 1$.

Les deux lois de composition interne \vee et \wedge (supremum et infimum)

$$\omega \vee \omega' = \sup(\omega, \omega') \quad \text{et} \quad \omega \wedge \omega' = \inf(\omega, \omega'),$$

confèrent à Ω une structure de treillis distributif dont le minorant universel est le point $(0, 0, \dots, 0)$ et le majorant universel, le point (r_1, r_2, \dots, r_m) .

Une chaîne du treillis est une suite de sommets $(\omega^{(1)}, \omega^{(2)}, \dots, \omega^{(k)})$ de Ω totalement ordonnée par la relation d'ordre (3) :

$$\omega^{(1)} \leq \omega^{(2)} \leq \dots \leq \omega^{(k)}.$$

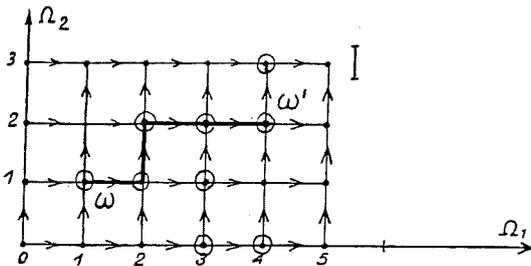
La chaîne est **maximale** si pour tout $i, 1 \leq i \leq (k-1), \omega^{(i+1)}$ succède à $\omega^{(i)}$; la **longueur** d'une chaîne maximale ayant k sommets est $(k-1)$. Toutes les chaînes maximales reliant deux sommets fixés de Ω ont même longueur. La longueur commune de toutes les chaînes maximales joignant le point origine $(0,0,\dots,0)$ au sommet ω sera appelée (pour nous conformer notamment au vocabulaire des praticiens de l'Analyse hiérarchique) "score" de ω . La fonction "score" **gradue** le treillis.

On peut naturellement associer à Ω muni de la relation d'ordre (3) un graphe, au sens de Cl. Berge (cf. [3]), $G = (\Omega, U)$ dont l'ensemble des sommets est Ω et l'ensemble des arcs U ; du point ω sont issus les arcs qui le relient à l'ensemble $\Gamma(\omega)$ de ses successeurs :

$$\Gamma(\omega) = \{(\omega_1, \dots, \omega_{(j-1)}, \omega_j + 1, \omega_{(j+1)}, \dots, \omega_m), \omega_j < r_j / 1 \leq j \leq m\} \quad (4)$$

où on a noté $\omega = (\omega_1, \omega_2, \dots, \omega_m)$.

Dans notre représentation géométrique, nous matérialiserons l'arc joignant $(\omega_1, \dots, \omega_j, \dots, \omega_m)$ à $(\omega_1, \dots, \omega_{(j-1)}, \omega_j + 1, \omega_{(j+1)}, \dots, \omega_m)$, par le vecteur v_j .



Représentation géométrique du graphe associé à $a^1 \times a^2$ où a_1 et a_2 sont des items totaux présentant respectivement 6 et 4 modalités.

Figure 2

La notion de chaîne maximale dans le treillis est celle de **chemin** dans le graphe. Les deux sommets O et I de scores 0 et $r_1 + r_2 + \dots + r_m$ seront appelés extrémités du graphe et tout chemin les reliant sera dit **extrémal**.

Bien que le langage des treillis suffise, nous utiliserons également le vocabulaire, parfois plus souple, des graphes; le contexte sera assez clair pour savoir de quoi nous parlons.

Chaque "patron de réponse" $\omega = (\omega_1, \dots, \omega_m)$ étant représenté par le sommet correspondant du treillis, la représentation de la population de sujets étudiée est une mesure entière positive sur $\Omega : \{n_\omega / \omega \in \Omega\}$ où n_ω est le nombre d'individus de la population dont le patron de réponse est ω . Un sommet ω du treillis est dit "patron de réponse **observé**" si la mesure n_ω qui lui est affectée est différente de 0 .

Une **échelle relative à l'ensemble des items totaux** $\{a^j / 1 \leq j \leq m\}$ est représentée par une chaîne du treillis; il s'agit d'un ensemble de sommets que peut recouvrir un chemin du graphe $G = (\Omega, U)$.

On dit que "le comportement d'une population donnée de sujets est **unidimensionnel** par rapport à l'ensemble des items totaux $\{a^j / 1 \leq j \leq m\}$ " si l'ensemble des patrons de réponse **observés** forme une échelle.

On montre aisément qu'une condition suffisante d'unidimensionnalité du comportement de **toute** population est que l'ensemble des items $\{a^j / 1 \leq j \leq m\}$ se réfère à une même variable sous-jacente (cf. §3 ci-dessus). Désignons par $(\alpha^j) = (0 = \alpha_0^j, \alpha_1^j, \dots, \alpha_{r_j}^j)$ la subdivision sur l'axe de la variable sous-jacente définie par l'item a^j , $1 \leq j \leq m$; et soit $(0 = \gamma_0, \gamma_1, \dots, \gamma_{r_1+r_2+\dots+r_m})$ la subdivision associée à la composition $c = a^1 * a^2 * \dots * a^m$ des différents items a^j . En considérant qu'un même point ne peut appartenir à plus d'une subdivision (α^j) , considérons la chaîne maximale reliant les extrémités du treillis $(\omega^{(0)}, \omega^{(1)}, \dots, \omega^{(\ell)}, \dots, \omega^{(r_1+\dots+r_m)})$, définie de la façon suivante : $\omega^{(0)}$ est le sommet dont toutes les coordonnées sont nulles ; si le ℓ -ème point γ_ℓ de la subdivision définie par c est un point provenant de la subdivision définie par le j -ème item a^j , alors le ℓ -ème sommet $\omega^{(\ell)}$ de la chaîne se déduit du précédent $\omega^{(\ell-1)}$ en ajoutant 1 à la j -ème composante :

$$\omega_h^{(\ell)} = \omega_h^{(\ell-1)} \text{ (resp. } = \omega_h^{(\ell-1)} + 1) \text{ si } h \neq j \text{ (resp. si } h = j).$$

Cette chaîne sera dite l'"échelle **définie** par la suite des items totaux $\{a^j / 1 \leq j \leq m\}$ ", laquelle est sensée supporter toute la mesure définie sur Ω lorsqu'on soumet à une population donnée le questionnaire formé par la suite de ces items.

Nous avons pris soin de définir l'unidimensionnalité du comportement d'une population donnée par rapport à un ensemble d'items indépendamment de l'hypothèse d'une même variable sous-jacente. En effet, par rapport à un même ensemble d'items, une population donnée peut se comporter de façon unidimensionnelle alors que ce ne sera pas le cas pour une autre, peut-être d'ailleurs plus nombreuse. Inversement, relativement à un ensemble d'items se référant à une même variable (on dit également «formant une échelle»), l'ensemble des patrons de réponse observés d'une population donnée peut, en raison de réponses erronées, ne pas former une échelle ; d'où le problème de l'analyse hiérarchique unidimensionnelle.

Le support de la mesure $\{n_\omega / \omega \in \Omega\}$ sur l'ensemble Ω des sommets du treillis, c'est-à-dire $\{\omega \in \Omega / n_\omega \neq 0\}$, est une partie d'un produit fini d'ordres totaux finis ; il s'agit d'un ordre partiel sur un ensemble fini. Or tout ensemble fini partiellement ordonné Λ peut apparaître comme une partie d'un produit fini d'ensembles finis totalement ordonnés. Relativement à Λ on définit, fondamentalement, deux nombres entiers. Le nombre δ , **dimension de l'ordre**, est le nombre minimum d'ordres totaux dont Λ peut être considéré comme une partie de leur produit ; le **nombre de stabilité** β qui est le nombre minimum de chaînes totalement ordonnées pouvant recouvrir Λ .

Ainsi, en se référant à l'exemple fourni par la figure 2 précédente, considérons l'ensemble Λ des sommets marqués \circ :

$$\Lambda = \{(0,3), (0,4), (1,1), (1,2), (1,3), (2,2), (2,3), (2,4), (3,4)\}$$

muni de l'ordre produit défini par (3) ci-dessus. Pour cet ensemble ordonné, la dimension est 2 et le nombre de stabilité est 3.

La notion de multidimensionnalité en Analyse hiérarchique recouvre la seconde notion et non la première.

Le comportement de la population étudiée par rapport à l'ensemble des items est **k-dimensionnel** si le nombre minimum d'échelles recouvrant l'ensemble $\{\omega \in \Omega / n_{\omega} \neq 0\}$ des patrons de réponse observés est k.

Il peut y avoir plus d'un système de k chaînes (échelles) recouvrant le support de la mesure ; le problème se pose de reconnaître celui des systèmes dont chaque élément représente une échelle d'attitude pour une partie de la population étudiée. Nous proposerons un algorithme qui tient compte de la présence éventuelle de patrons de réponse erronés.

III. Analyse hiérarchique unidimensionnelle

1. POSITION DU PROBLEME

La question principale que se pose l'Analyse hiérarchique est la suivante : "Peut-on admettre que les items totaux a^j , $1 \leq j \leq m$, se réfèrent à une même variable ?" Or nous avons défini l'unidimensionnalité d'un ensemble d'items en supposant qu'il n'existe pas de possibilité d'erreur dans le choix par le sujet de l'une des modalités d'un item (cf. § II.2.). On se rend compte du caractère déterministe d'une telle hypothèse qui ne tient pas compte de la présence éventuelle de patrons de réponse « erronés ». Il y a lieu de remplacer cette hypothèse par une autre, à caractère probabiliste, dans le cadre de laquelle la probabilité d'un patron erroné est petite sans être exactement nulle.

Dans le cas d'une réponse positive à la question posée, l'ensemble des patrons observés non erronés est une échelle, qu'il faut alors déterminer. Pour l'estimation de l'échelle nous proposerons deux approches. La première, qui suppose la définition d'une métrique sur Ω , est algorithmique, elle consiste à déterminer celle des échelles qui "s'ajuste le mieux" à l'ensemble pondéré des patrons de réponse observés ; en d'autres termes, à « distance » minimum d'une mesure supportée par une partie du treillis de représentation. Il peut exister plus d'une solution ; d'autre part, une telle estimation est indépendante de l'existence d'une même variable sous-jacente à l'ensemble des items.

La seconde approche consiste à proposer un modèle probabiliste de l'erreur dans le choix par un sujet d'une modalité d'un item. Un tel modèle, dans l'hypothèse d'une certaine distribution (qu'on peut se fixer sans restreindre la généralité) de la variable sous-jacente sur la

population étudiée, permet d'une part de déterminer l'échelle, d'autre part, de définir les probabilités des différents patrons de réponse possibles. D'où la possibilité de tester le modèle.

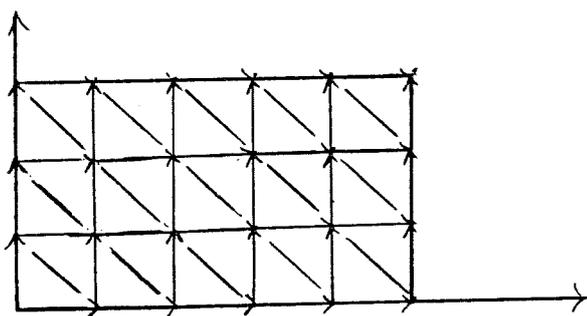
2. ALGORITHME DE CONSTRUCTION D'UNE ECHELLE

2.1. Distances adoptées sur Ω

En vue de la définition de l'«écart» d'un patron de réponse observé à une échelle, nous allons introduire trois distances sur Ω en utilisant le graphe $G = (\Omega, U)$ de représentation. Considérons pour cela les deux graphes suivants liés à $G = (\Omega, U)$.

1) Le graphe G' , dit symétrisé de G , obtenu en adjoignant à tout arc de G , l'arc inverse ; les deux sommets extrémités sont alors reliés par une arête.

2) Le graphe \bar{G} , dit complété de G , obtenu à partir de G en reliant d'une arête tout couple de sommets ayant un même prédécesseur, c'est-à-dire tels que $\omega = (\omega_1, \dots, \omega_m)$, $\omega' = (\omega'_1, \dots, \omega'_m)$ où $\omega_j = \omega'_j$ pour tout j sauf pour deux indices h et h' pour lesquels $\omega_h = \omega'_h + 1$ et $\omega_{h'} = \omega'_{h'} - 1$.



Graphe G associé à un couple d'items présentant respectivement 6 et 4 modalités.

Figure 3

Nous adaptons sur Ω l'une des distances

- a) d_1 ; $d_1(\omega, \omega') = 1$ (resp. $= 0$) si $\omega' \neq \omega$ (resp. si $\omega' = \omega$)
 - b) d_2 ; $d_2(\omega, \omega')$ est la **longueur** du plus court chemin de \bar{G} joignant ω à ω' .
 - c) d_3 ; $d_3(\omega, \omega')$ est la **longueur** du plus court chemin de G' joignant ω à ω' .
- On rappelle que la longueur d'un chemin est le nombre d'arcs qu'il contient.

2.2. «Ecart» d'un patron de réponse à une échelle

Ayant adopté, à partir de considérations concrètes, l'une des trois distances d_1 , d_2 ou d_3 , la manière qui nous semble la plus « naturelle » pour définir l'écart d'un patron de réponse observé représenté par un

sommet ω du graphe à une échelle représentée par un chemin ξ , est de prendre la distance au sens topologique du sommet ω au sous-ensemble de points ξ ; soit

$$e_i(\omega, \xi) = \min_{\bar{\omega} \in \xi} d_i(\omega, \bar{\omega}), \quad (1)$$

où $i = 1, 2$ ou 3 .

Toutefois, pour pouvoir adapter l'algorithme, que nous présenterons bientôt, au cas de la distance d_3 , il est nécessaire (cf. [1]) de définir la distance de ω à ξ comme étant la somme pondérée des distances d_3 des différents sommets de ξ à ω ; soit

$$e'_3(\omega, \xi) = \sum_{\bar{\omega} \in \xi} n_{\bar{\omega}} d_3(\omega, \bar{\omega}) \quad (2)$$

Exemple :

Considérons le cas du système représentatif $a^1 \times a^2$, où a^1 et a^2 présentent respectivement 6 et 4 modalités, et l'échelle ξ dont les points sont :

$(0,0), (1,0), (2,0), (2,1), (3,1), (3,2), (4,2), (5,2)$ et $(5,3)$.

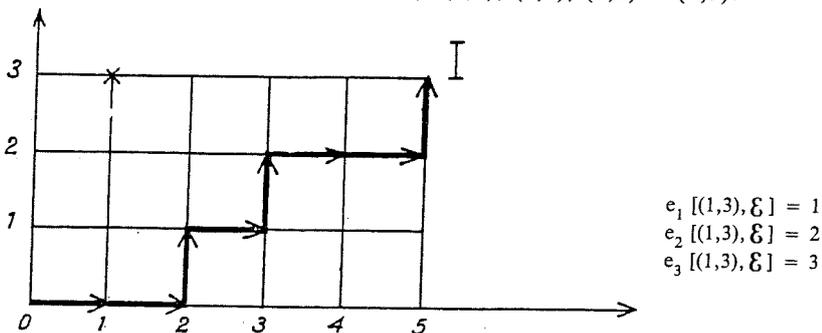


Figure 4

Supposons qu'à partir des réponses de la population étudiée, la suite des pondérations affectée à la suite des sommets de l'échelle est

7, 5, 12, 11, 23, 15, 10, 8, 2; on a

$$e'_3[(1,3), \xi] = 7 \times 4 + 5 \times 3 + 12 \times 4 + 11 \times 3 + 23 \times 4 + 15 \times 3 + 10 \times 4 + 8 \times 5 + 2 \times 4.$$

Cherchons quelque indication intuitive sur le choix de l'une des distances. Si on admet que les différents patrons de réponse observés erronés ont le même "degré d'erreur", on adoptera la distance d_1 .

Si on admet que le "degré d'erreur" du patron de réponse erroné est proportionnel à son « éloignement » du sommet de l'échelle de même score (i.e. situé sur le même niveau du treillis), on adoptera la distance

d_2 , laquelle donne lieu à la forme analytique suivante de l'écart du sommet ω à l'échelle \mathcal{E} ;

$$e_2(\omega, \mathcal{E}) = \frac{1}{m} \sum_{1 \leq j \leq m} |\omega_j - \bar{\omega}_j|, \quad (3)$$

où $\bar{\omega} = (\bar{\omega}_1, \dots, \bar{\omega}_m)$ est le sommet de l'échelle de même score que $\omega = (\omega_1, \dots, \omega_m)$ qui représente le patron de réponse erroné. Cette expression se réduit, dans le cas particulier de deux items, à $(|\omega_1 - \bar{\omega}_1| + |\omega_2 - \bar{\omega}_2|)$ qu'il est naturel de considérer ; en effet, un sujet donné, normalement situé au point $(\bar{\omega}_1, \bar{\omega}_2)$ de l'échelle, et répondant trop fortement à l'un des items, a tendance pour récupérer, à répondre faiblement à l'autre item.

La distance mathématiquement la plus classique, celle la plus conforme à la structure du treillis de représentation, est la distance d_3 , laquelle donne lieu à la forme analytique suivante de l'écart du sommet ω à l'échelle \mathcal{E} ;

$$e_3(\omega, \mathcal{E}) = \min_{\bar{\omega} \in \mathcal{E}} \sum_{1 \leq j \leq m} |\omega_j - \bar{\omega}_j|, \quad (4)$$

Ici, l'erreur dans la réponse d'un sujet à un item est considérée indépendamment, de l'erreur dans la réponse du sujet à un autre item.

Les deux critères d'adéquation d'une échelle obtenus respectivement à partir de l'une des distances d_2 ou d_3 doivent conduire sensiblement au même résultat et ce, d'autant que l'échelle \mathcal{E} s'impose statistiquement.

2.3. Critère d'adéquation d'une échelle

Une échelle \mathcal{E} est d'autant « meilleure » que la somme **pondérée** des écarts des sommets de Ω (i.e. des patrons de réponse) à \mathcal{E} est plus petite.

Ayant adopté l'une des distances d_1 , d_2 ou d_3 , si on considère la forme (1), la plus naturelle de l'écart d'un patron de réponse observé à une échelle, la quantité critère s'écrit

$$\sum_{\omega \in \Omega} n_\omega e_i(\omega, \mathcal{E}) = \sum_{\omega \in \Omega} n_\omega \left\{ \min_{\bar{\omega} \in \mathcal{E}} d_i(\omega, \bar{\omega}) \right\}, \quad (5)$$

Si on considère la forme (2) de l'écart, seulement adaptée lorsque la distance choisie sur Ω est d_3 , la quantité critère se met sous la forme

$$\sum_{\omega \in \Omega} n_\omega e'_3(\omega, \mathcal{E}) = \sum_{\omega \in \Omega} \sum_{\bar{\omega} \in \mathcal{E}} n_\omega n_{\bar{\omega}} d_3(\omega, \bar{\omega}), \quad (6)$$

La quantité critère représente une « mesure » du désaccord entre les résultats observés et l'hypothèse d'unidimensionnalité. Notons avec intérêt que le critère d'adéquation d'un système de « noyaux » à une partition, utilisé par la suite par E. Diday dans sa méthode des « nuées dynamiques » (cf.5) est de même nature que ceux que nous présentons.

Remarquons enfin que pour $i = 1$, (5) représente ce que les praticiens de l'analyse hiérarchique désignent sous le nom de « coefficient de reproductibilité ».

2.4. Algorithme

•

Il s'agit de déterminer l'échelle, représentée par une chaîne maximale reliant les extrémités du treillis de représentation, autrement dit, par un chemin du graphe G reliant ses extrémités O et I , qui rend minimum le critère adopté. Pour cela, la méthode la plus directe consiste à affecter à chaque chemin de G joignant O et I , la valeur du critère d'adéquation défini et à retenir celui ou ceux des chemins pour lesquels cette valeur est minimum (on peut trouver plus d'une échelle répondant à la question). Dans ces conditions, pour chacun des chemins extrémaux, on aura à calculer la somme pondérée des écarts (5) ou (6) ; or, le nombre total de tels chemins est $(r_1 + r_2 + \dots + r_m)! / r_1! r_2! \dots r_m!$. On voit bien que le volume des calculs que nécessite une telle méthode est, en général, particulièrement important, même pour un puissant ordinateur ; d'où le caractère crucial d'un algorithme optimal dont le nombre de pas soit de l'ordre du nombre de sommets du treillis. Cet algorithme, que nous allons exprimer, ne peut s'adapter au cas de la distance d_3 avec la quantité critère (5). On peut dans ces conditions proposer l'algorithme sous-optimal dont la première étape consiste, par une recherche directe, à déterminer l'échelle $b^{(2)}$ relative au couple d'items (a^1, a^2) ; et la i -ème étape, à former l'échelle $b^{(i+1)}$ associée au couple d'échelles $(b^{(i)}, a^{i+1})$; $1 < i < m$. L'échelle ainsi obtenue, relativement à la suite des items (a^1, a^2, \dots, a^m) , peut dépendre de l'ordre dans lequel ont été rangés les items ; elle n'en dépend pas et est optimale si sa restriction à tout couple d'échelles $(a^j, a^{j'})$ est l'échelle optimale résumant a^j et $a^{j'}$.

Nous allons à présent exposer l'algorithme qui en un « petit » nombre de pas permet de déterminer les échelles optimales. Cet algorithme revient à construire, à partir d'une pondération positive des sommets du graphe $G = (\Omega, U)$:

$$\{p_\omega / p_\omega \in \mathbf{R}^+, \omega \in \Omega\},$$

le ou les chemins extrémaux les moins pesants ($\sum_{\omega \in \mathcal{E}} p_\omega$ minimum).

Pour fixer les idées nous nous plaçons dans le cas où $\Omega = \Omega_1 \times \Omega_2$ est un produit de deux ordres totaux finis (associé à un couple (a, b) d'items). Chaque sommet ω de Ω est défini, dans la représentation géométrique du treillis ou du graphe G , par ses coordonnées ; $\omega = (\alpha, \beta)$. A ω est attaché la masse $p(\alpha, \beta)$: nombre réel positif dépendant de ω et de n_ω .

La première étape de l'algorithme consiste à affecter aux sommets de Ω une nouvelle pondération $\{r_\omega / \omega \in \Omega\}$ où $r(\alpha, \beta)$ sera le poids du chemin le moins pesant reliant l'origine $(0, 0)$ au point $\omega = (\alpha, \beta)$. L'algorithme démarre à partir de l'origine où $r(0, 0) = p(0, 0)$.

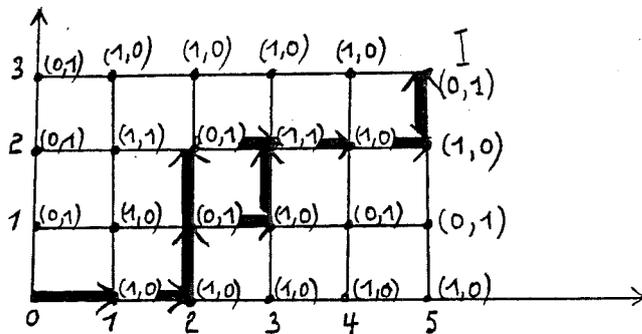


Figure 5

Il s'agit de la représentation d'un couple d'items (a,b) présentant 6 et 4 modalités respectivement. Chaque sommet (α, β) est affecté de son indicateur de précédence $i(\alpha, \beta)$. On trouve dans ce cas deux échelles représentées par deux chaînes qui ont une section commençante (resp. finissante) commune. Ces deux chaînes sont

$$(0,0), (1,0), (2,0), (2,1), (3,1), (3,2), (4,2), (5,2), (5,3) \text{ et} \\ (0,0), (1,0), (2,0), (2,1), (2,2), (3,2), (4,2), (5,2), (5,3).$$

La première correspond à l'ordre suivant sur l'ensemble des modalités des deux items

$$a_0 \sim b_0 < a_1 < a_2 < b_1 < a_3 < b_2 < a_4 < a_5 < b_3$$

et la seconde, à

$$a_0 \sim b_0 < a_1 < a_2 < b_1 < b_2 < a_3 < a_4 < a_5 < b_3$$

Chaque sommet a un ou deux prédécesseurs. Le nouveau poids $r(\alpha, \beta)$ de (α, β) est la somme de $p(\alpha, \beta)$ et du poids q de son prédécesseur le moins lourd ; soit :

$$r(\alpha, \beta) = p(\alpha, \beta) + \min \{r(\alpha-1, \beta), \text{ si } (\alpha, \beta) \text{ a deux prédécesseurs} \\ = p(\alpha, \beta) + r(\alpha-1, \beta), \text{ si le seul prédécesseur de } (\alpha, \beta) \text{ est } (\alpha-1, \beta) \\ = p(\alpha, \beta) + r(\alpha, \beta-1), \text{ si le seul prédécesseur de } (\alpha, \beta) \text{ est } (\alpha, \beta-1)$$

De plus, à (α, β) nous attacherons l'indicateur $i(\alpha, \beta)$ qui est un vecteur logique à deux composantes 0 ou 1 ; la composante égale à 1 indiquant le sommet le moins lourd au sens de r , précédant (α, β) ; soit

$$i(\alpha, \beta) = (1,0) \text{ si } r(\alpha-1, \beta) < r(\alpha, \beta-1) \text{ ou si le seul prédécesseur de } (\alpha, \beta) \text{ est } (\alpha-1, \beta) \\ = (0,1) \text{ si } r(\alpha-1, \beta) > r(\alpha, \beta-1) \text{ ou si le seul prédécesseur de } (\alpha, \beta) \text{ est } (\alpha, \beta-1) \\ = (1,1) \text{ si } r(\alpha-1, \beta) = r(\alpha, \beta-1).$$

De la sorte, le point I, extrémité du treillis ou du graphe, sera affecté du poids du chemin le moins pesant reliant les extrémités du graphe ; d'autre part la suite, à partir de I, des indicateurs permettra, de proche en proche, de retrouver le ou les chemins extrémaux de pondération minimum.

La généralisation de ce qui précède au cas d'un produit de plus de deux ordres totaux ne présente aucune difficulté théorique. Cependant, l'application de l'algorithme devient délicate s'il y a plusieurs échelles optimales.

Nous allons à présent pour chacune des distances d_1 , d_2 ou d_3 définir la pondération $\{p_\omega / \omega \in \Omega\}$ de telle sorte que le poids r du chemin extrémal représente la valeur du critère d'adéquation de l'échelle définie par ce chemin.

a) Cas de la distance d_1

Dans ce cas $p_\omega = (n - n_\omega)$ où $n = \sum_{\omega \in \Omega} n_\omega$ est l'effectif de la population étudiée.

En effet ; de la sorte, l'algorithme minimise $\sum_{\omega \in \mathcal{E}} (n - n_\omega)$, où \mathcal{E} est un chemin extrémal définissant une échelle ; cette quantité est égale, à une constante additive près, à $\sum_{\omega \in \Omega} n_\omega e_1(\omega, \Omega)$ (cf. § 2.3.).

b) Cas de la distance d_2

Désignons par $s(\omega)$ la longueur de la chaîne maximale d'extrémité ω et d'origine, celle du treillis et soit $T(s)$ le niveau du treillis où la fonction "score" prend la valeur s ; en utilisant la représentation géométrique, on a

$$T(s) = \{\omega \in \Omega / \omega_1 + \omega_2 + \dots + \omega_m = s\}$$

où $(\omega_1, \dots, \omega_m)$ est la représentation du sommet ω .

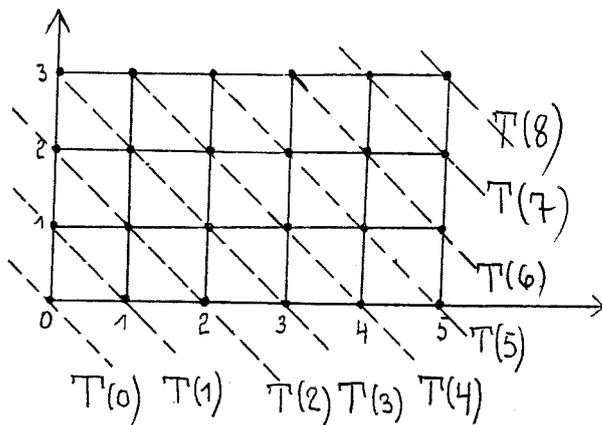


Figure 6

A $\omega \in \Omega$, on affectera le poids

$$p_\omega = \sum_{\bar{\omega} \in T(s(\omega))} n_{\bar{\omega}} \sum_{1 \leq j \leq m} |\bar{\omega}_j - \omega_j|$$

car de la sorte une échelle optimale minimisera $\sum_{\omega \in \Omega} p_\omega$; c'est-à-dire, $\sum_{\omega \in \Omega} n_\omega e_2(\omega, \varepsilon)$, (cf. § 2.3.).

c) Cas de la distance d_3

Ici il s'agit d'affecter la pondération suivante des sommets de Ω (cf. [1])

$$p_\omega = \sum_{\bar{\omega} \in \Omega} n_{\bar{\omega}} d_3(\bar{\omega}, \omega)$$

On montre d'ailleurs, moyennant la relation caractérisant une mesure μ

$$\mu(\omega) + \mu(\omega') = \mu(\omega \vee \omega') + \mu(\omega \wedge \omega'),$$

qu'il suffit de calculer la valeur de p sur chacun des axes représentant Ω_j , pour la déterminer en tout point de Ω .

Avec une telle pondération une échelle optimale minimise $\sum_{\omega \in \Omega} n_\omega e'_3(\omega, \varepsilon)$, (cf. § 2.3.).

Les méthodes, que nous venons d'exposer, de détermination de l'échelle sont essentiellement métriques et combinatoires.

Dans la pratique, il arrive assez souvent, surtout lorsque la tendance comportementale sous-jacente n'est pas très forte, que l'algorithme extraie plus d'une seule échelle optimale. Toutefois, les différentes échelles optimales restent très « voisines », au sens par exemple d'une distance entre ordres totaux finis, définie par le « nombre d'inversions ».

Considérons l'ensemble des sommets du treillis qui se trouvent sur la réunion des chaînes définies par les différentes échelles optimales. En saturant cet ensemble au moyen des opérations \wedge et \vee ; on trouve de façon opérationnelle une généralisation au cas d'un ensemble d'items totaux (où le nombre de modalités par item est quelconque et non nécessairement le même d'un item à l'autre) de la notion de *tresse* introduite par Cl. Flament (cf. [6]) dans le cas d'un ensemble d'items dichotomiques. C'est seulement dans ce dernier cas que la caractérisation donnée par Y. Kergall (cf. [7]) de la notion de tresse à partir d'une correspondance de Galois s'avère, à notre avis, pertinente.

3. ECHELLE A PARTIR D'UN MODELE PROBABILISTE

Nous avons associé à un item total une subdivision d'une variable sous-jacente qu'on supposera ici continue (cf. § II.2. dont on reprend ici les notations). On peut dans ces conditions introduire un modèle probabiliste qui tient compte d'une possibilité d'erreur dans le choix d'une

modalité de l'item par le sujet et qui permettra la détermination statistique des points de la subdivision. L'application de ce modèle à chacun des éléments d'une classe d'items totaux se référant à une même variable sous-jacente donnera la subdivision associée à la composition des différents items de la classe (cf. § II.3.2.) et par conséquent, l'échelle hiérarchique ordonnant totalement l'ensemble des modalités non minimales des différents items.

Ce modèle suppose d'une part la donnée de la fonction de répartition $G(\xi)$ de la variable sous-jacente sur la population étudiée et d'autre part, la donnée, pour un individu quelconque dont la valeur de la variable sous-jacente est ξ , de la probabilité $F_\xi(\alpha_j)$ d'observer une valeur inférieure à α_j ; c'est-à-dire d'avoir une réponse de l'individu par l'une des modalités a_0, a_1, \dots, a_{j-1} .

Des considérations intuitives et techniques nous ont conduit à proposer le modèle où $G(\xi)$ et une fonction de répartition eulérienne :

$$G(\xi) = \frac{1}{\Gamma(\lambda)} \int_0^\xi e^{-x} x^{\lambda-1} dx$$

et où

$$F_\xi(\alpha_j) = \exp[-k\xi / \alpha_j]$$

avec $k = \text{Log}_e 2$.

λ est un paramètre d'échelle qu'on peut se fixer une fois pour toutes. Quant au paramètre k , on peut le déterminer par la condition "naturelle" suivante : "la probabilité d'observer sur un sujet dont la valeur de la variable sous-jacente est $\xi = \alpha_j$, une valeur inférieure à α_j est égale à 1/2".

Ce modèle permet la détermination en fonction des α_j ($1 \leq j \leq r$) de la fréquence théorique des différentes modalités de l'item a_j ; ce qui permet d'identifier les abscisses α_j des points de la subdivision.

L'approche esquissée ici montre une voie par laquelle peut se faire le raccord entre analyse hiérarchique au sens de L. Guttman et analyse de la structure latente au sens de P. Lazarfeld (voir dans [13] et dans [16]). Toutefois, nous n'allons pas nous y attarder car nous n'avons pas eu à appliquer cette technique dans les données dont il sera bientôt question.

Terminons ce paragraphe en signalant que l'algorithme du paragraphe 2 précédent peut être utilisé dans le cas de l'analyse hiérarchique multidimensionnelle (cf. [11], chap. 8) à laquelle nous avons fait allusion à la fin du paragraphe II.4. Une telle utilisation fait appel à des propriétés fondamentales de la théorie des ordres finis (cf. [5], [14], [15] et [17]).

IV. Présentation schématique de la méthode de classification hiérarchique

Comme nous le signalions dans l'introduction, dans la pratique, nous ne nous permettons de poser le problème de l'analyse hiérarchique unidimensionnelle que par rapport à une classe d'échelles reconnue de bonne cohésion dans une classification hiérarchique sur un ensemble plus vaste d'échelles pouvant recouvrir la dimension recherchée. On ne s'étonnera donc pas de voir ci-dessous (cf. § V.) les schémas d'organisation des classes d'échelles qu'on analysera. Ces derniers sont extraits de l'arbre des classifications portant sur l'ensemble des items du questionnaire. L'application des algorithmes d'analyse hiérarchique permet d'une part, une analyse plus complète de la classe d'échelles et surtout (c'est dans cette étude l'objectif central), de fournir au spécialiste une échelle ordinale de "mesure" fine.

Nous allons par conséquent rappeler ici très brièvement les grandes lignes de notre méthode de classification hiérarchique qui est basée sur une notion générale de vraisemblance du lien par rapport à une hypothèse d'absence de relation statistique.

Face à un tableau de données variables descriptives \times Individus ou Objets, la méthode permet aussi bien une classification de l'ensemble des variables V que de l'ensemble E des objets et ceci quelle que soit la nature des variables. Si la classification des variables permet la découverte des principales tendances de comportement de la population étudiée, à travers un échantillon de cette dernière, la classification de E permet de découvrir une typologie de l'ensemble des objets conformément aux ressemblances perçues à travers V .

Si L est l'ensemble à classifier (L est soit l'ensemble V , soit l'ensemble E), le point de départ de la méthode est la définition d'un indice de proximité entre éléments de L . Cet indice qui se réfère à une échelle de probabilité est basé sur la vraisemblance du lien mesuré par un indice « brut » de proximité qui s'impose en général.

Il s'agit dans notre problème du cas de la classification d'un ensemble V de variables descriptives où chaque variable définit un item total (cf. [4]). Si ces différents items sont dichotomiques, en associant à la modalité haute (i.e. codée 1) de chaque item un attribut descriptif, l'ensemble V peut être assimilé à un ensemble A d'attributs de description. C'est la situation envisagée par Gras dans 7 et celle considérée dans [12] par rapport à laquelle nous allons nous exprimer.

Relativement à un couple (a, b) d'attributs, on introduit l'indice brut $s(a, b) = \text{card}(E_a \cap E_b)$ où E_a (resp. E_b) est l'ensemble des sujets possédant a (resp. b). Conformément au principe ci-dessus énoncé, l'indice définitif prend la forme

$$P(a, b) = \text{Prob.}^N \{S < s(a, b)\}, \quad (1)$$

où S est la variable aléatoire associée à s dans une hypothèse N d'absence de lien, adéquate : qui tient compte des caractéristiques de cardinalité des structures à comparer ; soit, ici $n_a = \text{card}(E_a)$ et $n_b = \text{card}(E_b)$. L'indice $P(a, b)$ s'obtient en fait à partir de la formule

$$P(a, b) = \phi[Q(a, b)], \quad (2)$$

où ϕ est la fonction de répartition de la loi normale centrée réduite et où $Q(a, b)$ s'obtient en « centrant » et en « réduisant » s par rapport à l'hypothèse N

$$Q(a, b) = \frac{s - \mu_{ab}}{\sigma_{ab}}, \quad (3)$$

Cependant il y a trois formes « voisines » de l'hypothèse d'absence de lien ayant un caractère symétrique par rapport à a et b , qui conduisent à trois expressions, d'ailleurs proches, de l'indice $Q(a, b)$. Mais à travers les nombreuses expériences menées, l'hypothèse N_1 qui a fourni les résultats les plus cohérents dans leurs nuances conduit à l'expression suivante de l'indice :

$$Q_1(a, b) = \frac{s - \frac{n_a n_b}{n}}{\sqrt{\frac{n_a n_b}{n}}} \quad (4)$$

Cette notion de proximité entre deux attributs descriptifs est généralisée dans la méthode à la notion de proximité entre deux variables descriptives de même type (au sens de la structure induite sur E), quel que soit ce dernier.

La notion de proximité entre deux éléments est alors étendue à celle entre deux classes C et D d'éléments où le rôle de l'indice brut de proximité sera joué par

$$p(C, D) = \max \{P(c, d) / (c, d) \in C \times D\} ; \quad (5)$$

celui, définitif, où on se réfère à une hypothèse d'absence de lien adéquate, prend la forme suivante

$$P(C, D) = (p(C, D))^{\ell m}, \quad (6)$$

où $\ell = \text{card}(C)$ et $m = \text{card}(D)$.

L'« algorithme de la vraisemblance du lien » (A.V.L.) établit à partir de cette proximité entre classes, l'arbre détaillé des classifications où à chaque niveau les paires de classes les plus voisines sont réunies. Cet arbre est en général quasiment binaire et comporte presque autant de niveaux que l'ensemble à classer a d'éléments.

Une étape décisive de la méthode consiste à condenser l'arbre aux niveaux où se produit un « nœud significatif » et ce, au moyen d'une statistique de proximité, obéissant au principe ci-dessus énoncé, entre

une forme adéquate de l'information quant aux ressemblances des éléments de l'ensemble à classifier et l'association correspondant au nœud. On introduit en fait, relativement à un même niveau i de l'arbre des classifications, deux statistiques : l'une « globale » Σ_i , qui rend compte de l'adéquation de la partition obtenue au niveau i , et l'autre, « locale » τ_i qui « mesure » le degré de signification de l'association qui se produit au niveau i , par rapport à l'ensemble des paires restant séparées à ce niveau. D'ailleurs le comportement de τ_i le long de la suite des niveaux est quasiment le même que celui du taux d'accroissement $\theta_i = (\Sigma_i - \Sigma_{i-1})$ de la statistique globale. Les nœuds significatifs correspondent aux associations qu'accompagnent des maxima locaux de τ_i (resp. θ_i).

Terminons ce point en précisant les expressions de Σ et de τ . Pour se ramener à la comparaison de deux structures de même type dans la recherche d'une mesure d'adéquation entre une partition π (éventuellement produite à un niveau de l'arbre) et l'information quant aux ressemblances de l'ensemble A à classifier, on ne retient de la définition de l'indice de proximité sur A que le préordre total associé sur l'ensemble B des paires d'éléments distincts de A (i.e. des parties à deux éléments de A), appelé préordonnance sur A et défini de la façon suivante :

$$(\forall (p, q) \in B \times B), p < q \iff Q(p) < Q(q), \quad (7)$$

Pour l'indice Q que nous adoptons (cf. formule (3)), la préordonnance associée $\omega(A)$, se réduit le plus souvent dans la pratique à une ordonnance, ordre total sur B . $\omega(A)$ sera représentée par la partie $gr(\omega)$ de $B \times B$ définie par

$$gr(\omega) = \{(p, q) \in B \times B / p < q \text{ et non } q < p \text{ pour } \omega(A)\} \quad (8)$$

La partition π sur A est regardée comme définissant un préordre total sur B à deux classes $S(\pi)$ et $R(\pi)$ où $S(\pi)$ est l'ensemble des paires séparées et où $R(\pi)$ est celui des paires réunies par la partition π . $S(\pi) < R(\pi)$ pour l'ordre quotient ; ainsi π sera représentée dans $B \times B$ par le rectangle

$$S(\pi) \times R(\pi) \quad (9)$$

L'indice brut de proximité entre la partition et la préordonnance sera dans ces conditions

$$s(\omega, \pi) = \text{card}\{gr(\omega) \cap (S(\pi) \times R(\pi))\}, \quad (10)$$

l'indice définitif prend la forme

$$\Sigma = [s(\omega, \pi) - r(\pi) \times s(\pi) / 2] / \sqrt{r(\pi) \times s(\pi) (b+1) / 12}, \quad (11)$$

où $r(\pi) = \text{card}(R(\pi))$, $s(\pi) = \text{card}(S(\pi))$ et $b = r(\pi) + s(\pi) = \text{card}(B)$.

Σ est obtenu en centrant et en réduisant $s(\omega, \pi)$ par rapport à l'hypothèse d'absence de lien où π est un élément aléatoire dans l'ensemble, muni d'une probabilité uniformément répartie, de toutes les partitions d'un même type. Dans le cadre d'une telle hypothèse, nous démontrons que Σ peut être considérée comme une réalisation d'une variable aléatoire $\mathcal{N}(0, 1)$.

De la même façon, l'indice brut de proximité qui conduit à τ_i se met sous la forme

$$\sigma_i = \text{card}\{\text{gr}(\omega) \cap (S(\pi) \times R'(\pi_i))\} \quad (12)$$

où $R'(\pi_i)$ est l'ensemble des paires réunies pour la première fois au niveau i .

Pour mieux contrôler l'interprétation, nous associons à chaque variable son degré de « neutralité » par rapport à une visée classificatoire, qu'on « mesure » par la petitesse de sa variance des proximités aux autres variables selon la formule

$$V(a) = \frac{1}{[\text{card}(A) - 1]} \sum_{c \in A - \{a\}} (Q(a,c) - Q(a))^2, \quad (13)$$

où $Q(a)$ est la moyenne des $\{Q(a,c)/c \in A - \{a\}\}$.

La présence d'attributs neutres dans une classe peut expliquer la difficulté de son interprétation.

V. Application à des données réelles

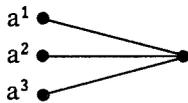
Les données proviennent d'une vaste et très riche enquête sur la Psycho-Pédagogie de l'enfant à laquelle nous avons fait allusion dans l'introduction (cf. [2] et [4]). Le questionnaire comprend une centaine d'items totaux où le nombre de modalités par item varie entre 2 et 10 ; le nombre total de modalités est de 250. Quatre mille parents ont répondu sur différents aspects du développement de leur enfant en relation avec le milieu psychologique où il évolue. Les deux principaux aspects visés sont la "Scolarité" et le "Comportement Alimentaire". Nous nous limiterons ici à l'analyse des tendances associées à l'**utilisation des connaissances**. Rappelons que chacune des tendances nous apparaît sous-tendue par un nœud significatif de l'arbre des classifications organisant par proximité l'ensemble des items et que l'algorithme mis en œuvre pour la recherche de l'échelle hiérarchique est celui du paragraphe III.2.4., avec usage de la distance d_2 .

La première classe regroupe les items suivants :

- a^1 : *Est-il plutôt*
- a_0^1 : consciencieux et méthodique
- a_1^1 : normalement appliqué pour son âge
- a_2^1 : tendance à bâcler son travail
- a^2 : *Dans son travail scolaire, est-il*
- a_0^2 : ordonné
- a_1^2 : non ordonné
- a_2^2 : brouillon

- a^3 : Pour son travail scolaire, est-il
- a_0^3 : méticuleux
- a_1^3 : soigneux
- a_2^3 : assez soigneux
- a_3^3 : peu soigneux

L'organisation de l'ensemble des trois précédents items dans l'arbre des classifications est la suivante



L'échelle hiérarchique qui représente la "concordance entre l'ordre et le soin" est la suivante :

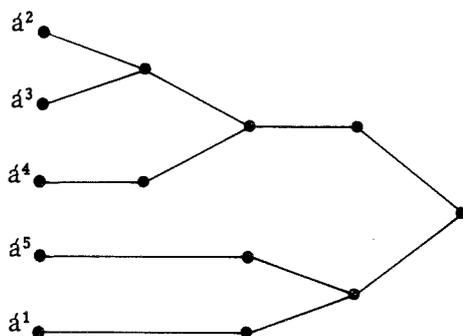
- e_0 : consciencieux et méthodique, travail scolaire ordonné, travail scolaire méticuleux.
- $e_1 = a_1^3$: travail scolaire soigneux
- $e_2 = a_1^1$: normalement appliqué pour son âge
- $e_3 = a_2^3$: assez soigneux
- $e_4 = a_1^2$: non ordonné
- $e_5 = a_1^2$: tendance à bâcler son travail
- $e_7 = a_3^3$: peu soigneux
- $e_8 = a_2^2$: brouillon.

La deuxième classe regroupe les items suivants :

- a^1 : Votre enfant, à la maison, parle-t-il de l'école ou du lycée ?
- a_0^1 : oui, avec plaisir
- a_1^1 : oui, incidemment
- a_2^1 : oui, avec hostilité
- a_3^1 : non.
- a^2 : Est-il intéressé par son travail scolaire ?
- a_0^2 : oui
- a_1^2 : partiellement
- a_2^2 : non.
- a^3 : Ses résultats scolaires actuels sont-ils ?
- a_0^3 : bons
- a_1^3 : assez bons
- a_2^3 : moyens
- a_3^3 : mauvais.

- a^4 : Ses résultats ont-ils été ?
 a_0^4 : satisfaisants jusqu'à présent.
 a_1^4 : satisfaisants pour certains cycles d'étude seulement
 a_2^4 : toujours insuffisants.
 a^5 : Pensez-vous ?
 a_0^5 : qu'il sait bien tirer parti de ce qu'il a appris
 a_1^5 : qu'il retient ce qu'il apprend
 a_2^5 : qu'il apprend mais oublie
 a_3^5 : qu'il a une mauvaise mémoire.

L'organisation de l'ensemble des cinq précédents items dans l'arbre des classifications est la suivante :



L'échelle hiérarchique obtenue qui représente "la relation entre la mémoire, le travail et les résultats scolaires" est la suivante :

e_0 : intéressé par son travail scolaire, résultats toujours satisfaisants, résultats actuels bons, tire parti de ce qu'il apprend, parle de l'école avec plaisir.

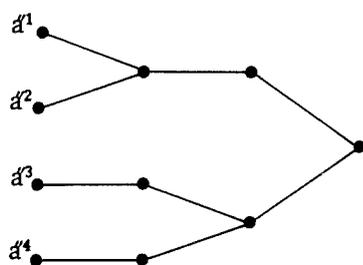
- $e_1 = a_1^5$: retient ce qu'il apprend
 $e_2 = a_1^3$: résultats actuels assez bons
 $e_3 = a_1^4$: résultats satisfaisants selon certains cycles
 $e_4 = a_2^3$: résultats actuels moyens
 $e_5 = a_1^1$: parle de l'école incidemment
 $e_6 = a_1^2$: partiellement intéressé par son travail scolaire
 $e_7 = a_2^5$: apprend mais oublie
 $e_8 = a_2^4$: résultats scolaires toujours insuffisants
 $e_9 = a_3^5$: a mauvaise mémoire

- $e'_{10} = a'_3$: résultats actuels mauvais
- $e'_{11} = a'_2$: parle de l'école avec hostilité
- $e'_{12} = a'_3$: ne parle pas de l'école
- $e'_{13} = a'_2$: non intéressé par son travail scolaire.

La troisième et dernière sous-classe de variables considérée est formée des items suivants :

- d^1 : *Est-il intéressé par d'autres occupations que le travail scolaire ?*
- d^1_0 : oui
- d^1_1 : parfois
- d^1_2 : non.
- d^2 : *Réussit-il mieux dans les travaux nécessitant une initiative personnelle ?*
- d^2_0 : oui, d^2_1 : non.
- d^3 : *Réussit-il mieux dans les travaux nécessitant une réflexion personnelle ?*
- d^3_0 : oui, d^3_1 : non.
- d^4 : *A-t-il des réussites dans des activités pratiques hors de l'école ?*
- d^4_0 : oui, d^4_1 : non.

L'organisation de l'ensemble des quatre items précédents dans l'arbre des classifications est la suivante :



L'échelle hiérarchique obtenue qui représente "l'intérêt pour les activités extra-scolaires" est la suivante :

- e''_0 : intéressé par des activités extra-scolaires, réussit mieux dans les travaux nécessitant une initiative personnelle, réussit mieux dans les travaux nécessitant une réflexion personnelle, a des réussites dans des activités pratiques hors de l'école.
- $e''_1 = d^3_1$: ne réussit pas mieux dans les travaux nécessitant une réflexion personnelle

- $e_2'' = d_1^2$: ne réussit pas mieux dans les travaux nécessitant une initiative personnelle
- $e_3'' = d_1^4$: n'a pas de réussites dans des activités pratiques hors de l'école
- $e_4'' = d_1^1$: parfois intéressé par d'autres occupations que le travail scolaire
- $e_5'' = d_2^1$: non intéressé par des activités extra-scolaires.

Les différentes échelles extraites ont parfaitement satisfait le spécialiste.

A partir de ces méthodes le pédagogue - mathématicien peut ordonner par complexité croissante une suite de tests mathématiques, dont chacun définit un item total dichotomique, relatifs tous à l'appropriation d'un concept donné. Cet ordre fournit une échelle ordinale fine de « mesure » permettant de situer de façon unidimensionnelle tout sujet examiné.

Bibliographie

(cf. aussi [10] de la biblio. générale)

- [1] BARBUT M. - "Echelles à distances minimum d'une partie donnée d'un treillis distributif fini" ; dans "Ordres totaux finis", (Aix-en-Provence, juillet 67) ; Gauthier-Villars, Mouton ; Paris, 1971.
- [2] BERGE A. et DENJEAN G. - "Comportement digestif et fonctionnement intellectuel" ; *Revue de Neuropsychiatrie Infantile*, 1974.
- [3] BERGE Cl. - "Graphes et Hypergraphes", Dunod, Paris, 1970.
- [4] COHEN HALLALEH I. - "Classification d'une famille d'échelles au moyen d'un nouvel indice. Comparaison avec le traitement par l'Analyse des correspondances. Application à des données en Psycho-Pédagogie et en Sociologie rurale. Thèse de 3ème cycle soutenue le 7/02/77 à l'Université Paris VI.
- [5] DILWORTH R.P. - "A decomposition theorem for partially ordered sets", *Ann. of Math.*, 60, 2, 1954 (359-364).
- [6] FLAMENT Cl. - "Tresses de Gutmann" ; dans "Ordres totaux finis" (Aix-en-Provence, juillet 67) ; Gauthier-Villars, Mouton ; Paris, 1971.