

J. P. BENZÉCRI

**Histoire et préhistoire de l'analyse des données.
Partie IV La psychométrie**

Les cahiers de l'analyse des données, tome 1, n° 4 (1976),
p. 343-366

http://www.numdam.org/item?id=CAD_1976__1_4_343_0

© Les cahiers de l'analyse des données, Dunod, 1976, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

HISTOIRE ET PRÉHISTOIRE DE L'ANALYSE DES DONNÉES (*)

Partie IV - La Psychométrie

par J. P. Benzécri (1)

2.4. L'école psychométrique américaine et l'analyse factorielle :

Stimulée par la pensée de Darwin, la biométrie, ou étude des dimensions physiques des êtres vivants, a considéré de nombreux systèmes de grandeurs naturellement corrélées entre elles : ainsi se fortifia la statistique multidimensionnelle. Mais l'analyse des données telle que nous l'entendons : c'est-à-dire la recherche inductive de dimensions cachées définies par combinaisons de mesures primaires, ne se développa qu'avec la psychométrie. Car de ce que depuis des millénaires les hommes ont appelé intelligence, mémoire, imagination, patience ... aucune mesure physique immédiate n'est possible; d'où la nécessité d'une construction statistique, si jamais les chiffres peuvent mesurer l'âme. De cette construction, l'école biométrique de Galton et Pearson, avait suggéré le dessin et posé les fondements (cf § 2.2.5); l'oeuvre fut poursuivie en Grande Bretagne même, par Cyril Burt; elle intéressa les psychologues du monde entier; et prit de monumentales proportions avec l'école américaine dont nous choisissons de présenter ici quelques images.

2.4.1. Définition et mesure objective de "l'intelligence générale" :

Tel est le titre du mémoire fameux (**) par lequel Charles Spearman fonda l'école américaine d'analyse factorielle. Expliquons brièvement quelles recherches convergent dans l'oeuvre du psychologue américain.

Dès le XVIII^e siècle, des physiciens et astronomes (français notamment : Guilford, 1936, p. 115 et 1954 p. 23 (***) , cite Sauveur 1700, Bouguer 1760, Lambert 1764 et Delezenne 1826; on remarquera qu'après tant d'autres disciplines, comme la méthode des moindres carrés, cf § 1.5.2, la psychophysique est née du désir de pénétrer le ciel) étudièrent les erreurs introduites par les sens de l'homme dans l'emploi des instruments de mesure physique. Ainsi, étudiant le seuil différentiel dans la perception des intensités lumineuses (i.e. la plus petite différence que l'on puisse percevoir) Bouguer et Lambert avaient reconnu que ce seuil est proportionnel à la grandeur physique que l'on fait varier (e.g. pour prendre exemple dans un domaine où la terminologie physique

(*) Suite de l'article paru sous le même titre dans les cahiers n° 1, 2 et 3.

(1) Professeur, Laboratoire de Statistique. Université Pierre et Marie Curie, Paris.

(**) "General intelligence" objectively determined and measured; in *American Journal of Psychology*. Vol 15 pp. 201-292 (1904).

(***) *Psychometric Methods* : MacGraw-Hill N.-Y.; 1° ed. 1936; 2° ed. 1954.

nous est plus familière qu'en photométrie : si en soupesant une masse de 100 g on perçoit une variation de 10 g, sur une masse de 1000 g on perçoit une variation de 100 g). Cette loi de la relativité des erreurs est si naturelle que, de même que pour la loi de Boyle-Mariotte ("doublez la pression, le volume est divisé par deux"), on en conçoit mal d'abord l'importance. Et ce n'est qu'au milieu du XIX^e siècle que, sur un domaine dont l'étude était déjà commencée, la psychophysique se constitua en une discipline autonome. Sous l'impulsion de savants allemands et notamment du fougueux Fechner (il forgea, croyons-nous, le nom même de psychophysique; et les contemporains évoquent souvent son ardeur de polémiste à défendre la science nouvelle !) on en vint à considérer le jeu imprécis des sens non en tant que cause d'erreur perturbant les mesures physiques, mais comme une source précieuse d'informations quantitatives sur un des domaines de la nature jusque-là rebelle aux chiffres : le psychique. Si la *loi des erreurs relatives* est aujourd'hui associée aux noms de Weber et de Fechner, c'est qu'après que Weber en eut le premier marqué la généralité, Fechner en tira sa fameuse formule selon laquelle la sensation est proportionnelle au logarithme de l'excitation : car en bref, si le sujet distingue une variation $\Delta E = kE$ de l'excitation physique E (un poids, ou une brillance; etc.) il est plausible qu'il y corresponde une augmentation constante $\Delta I = s$ de l'impression psychique I(E) produite par E; d'où la loi logarithmique, par intégration de la relation différentielle $dI/dE = s/(kE)$. Seuil absolu (plus petite excitation perçue) et seuil différentiel (plus petite différence perçue; cf *supra*) ne sont pas constants, mais fluctuantes; d'où des modèles probabilistes où l'on ne s'étonnera pas de trouver la loi normale (sur les distinctions, souvent mal exposées, que font ces modèles nous renvoyons à notre étude : *Histoire et critique de la notion de seuil*, où nous nous sommes efforcé d'être fidèle aux sources). De plus avec la vision des couleurs, la psychophysique pénètre dans le domaine multidimensionnel (un stimulus lumineux uniforme a trois dimensions : intensité lumineuse - gradation du clair au sombre; saturation - gradation du rouge au blanc par le rose; et ton - ou dominante spectrale, violet, indigo, bleu...). Mais nous ne considérons ici la psychophysique que comme une étape vers la psychométrie; celle-ci se distinguant de celle-là par l'absence de références physiques mesurables.

A la fin du XIX^e siècle l'école biométrique anglaise (F. Galton puis K. Pearson), s'applique à mesurer les aptitudes mentales (cf *supra* § 2.2.5). Aux Etats-Unis dès 1890 Cattell (qui fut à Londres assistant de Galton) et ses élèves tentent d'appliquer des tests psychologiques (séries d'épreuves simples où l'on dénombre succès et échecs d'un sujet) pour prédire les succès scolaires : mais selon H. Gulliksen (*Theory of Mental Tests*; ed. J. Wiley, N.-Y.; 1950) le coefficient de corrélation entre note de test et note scolaire ne dépassait pas la valeur négligeable de 0,19. Tout changea avec l'*échelle métrique de l'intelligence* de Binet et Simon (*Bourrellet, Paris 1957*). "Durant une longue période d'incubation, écrit Charles Spearman (*), les tests avaient été cultivés dans la retraite de plusieurs laboratoires de psychologie. Et soudain, Binet transforma ce travail théorique en une vivante pratique. Les maîtres trouvèrent dans le test d'intelligence un outil maniable; le public eut l'impression que cela lui était accessible". Très pragmatiquement nous semble-t-il, Binet avait constitué pour chacune des années de l'enfance une batterie d'épreuves diverses, (cinq composantes pour l'âge de trois ans, dont deux se rapportent à la mémoire, note sans complaisance Spearman; alors que pour l'âge de sept ans il y a quatre tests; dont aucun n'a la mémoire pour objet), réussie à 70% par les sujets de cet âge; d'où par référence à ces batteries détermination de l'*âge mental* de tout sujet et de son Q.I. (quotient intellectuel : i.e. quotient de l'âge mental par l'âge réel : selon Guilford, l'idée de calculer ce quotient semble due au psychologue allemand Stern).

(*) *The abilities of man; their nature and measurement; MacMillan ed; Londres 1927.*

C'était raviver le problème de la définition de l'intelligence. Reprenant avec vingt ans de recul le principe de ses recherches sur *les aptitudes de l'homme (op. laud)*, C. Spearman consacre à l'exposé de doctrines rivales plusieurs chapitres qu'il intitule : Doctrine monarchique, l'Intelligence; (e.g. Terman : "un individu est intelligent dans la mesure où il est capable de pensée abstraite"); Doctrine oligarchique : les facultés formelles (c'est notamment la doctrine scolastique, remontant à Aristote; pour qui l'activité de l'esprit humain n'est pas celle d'une pure intelligence, mais d'une intelligence spirituelle qui conjugue des facultés où le corps a sa part : sensation, imagination, mémoire); Doctrine oligarchique : les types (doctrines caractérologiques attentives à la diversité des hommes); Doctrine anarchique : "Niveau Général", "Moyenne", "Echantillonnage" (l'esprit est une mosaïque de facultés; dont chaque tâche requiert une partie; ce qu'on appelle globalement intelligence est une résultante; ici Spearman place l'oeuvre de Binet). Puis Spearman apporte ses découvertes sous l'enseigne modeste de : Doctrine éclectique; les deux facteurs. Philosophiquement la thèse du psychologue américain est simple : toute activité mentale (telle que la mesure un test) requiert qu'à l'aptitude polyvalente du sujet (mesurée par le *facteur général*) s'ajoute une aptitude particulière à cette activité (mesurée par un *facteur spécifique*). L'intérêt est pour nous dans la démonstration statistique de la thèse. Afin de critiquer celle-ci nous la placerons dans une doctrine plus générale, qui remonte en substance à l'école biométrique anglaise (dont Spearman avait été l'élève; cf §§ 2.2.5 & 2.2.7) mais est généralement citée d'après l'exposé qu'en a fait L. Thurstone, à partir de 1930.

2.4.2. Facteurs communs et facteurs spécifiques :

Thurstone (*) voit en géomètre et écrit des formules matricielles; il sera donc facile de traduire ses pensées dans le langage qui nous est familier. Soit I un ensemble de sujets; J un ensemble d'épreuves (ou tests) auxquelles ont été soumis les sujets; n_j^i la note du sujet i à l'épreuve j. Plus précisément, il s'agit de notes normalisées (ramenées par changement d'échelle à avoir moyenne 0, variance 1; et distribution approximativement normale). Thurstone regarde le système $n_j^I = \{n_j^i | i \in I\}$ des notes obtenues par tous les sujets à l'épreuve j, comme un vecteur de norme 1 dans R^I (muni de la norme définie par la somme des carrés des coordonnées divisée par le nombre card I des sujets); ainsi le coefficient de corrélation $r_{jj'}$ entre deux épreuves n'est autre que le produit scalaire $\langle n_j^I, n_{j'}^I \rangle$ des vecteurs associés. Outre les vecteurs n_j^I , Thurstone introduit pour expliquer ceux-ci d'autres vecteurs unitaires : les facteurs; parmi lesquels on distingue : l'ensemble A des *facteurs communs* $\varphi_\alpha^I, (\alpha \in A)$; les *facteurs spécifiques* ψ_j^I ; et les termes d'erreur ϵ_j^I . Ceci posé le modèle multifactoriel de Thurstone s'écrit :

$$n_j^I = \Sigma \{a_j^\alpha \varphi_\alpha^I | \alpha \in A\} + b_j \psi_j^I + e_j \epsilon_j^I ;$$

(*) Nous avons consulté : *Multiple-Factor Analysis, The University of Chicago Press, 1947; ainsi que l'article posthume des actes du colloque de 1955 sur L'analyse factorielle et ses applications : C.N.R.S., Paris, 1956.*

le résultat de l'épreuve j est une combinaison linéaire des facteurs communs (ainsi appelés parce qu'ils participent, à des degrés divers, à toutes les épreuves), complétée d'un terme spécifique (propre à l'épreuve) et d'un terme d'erreur. Pour Thurstone (op. laud. 1947, p. 75) il n'y a pas entre facteur spécifique et facteur commun de différence absolue : la notion de facteur spécifique est relative à la batterie (ensemble J) d'épreuves considérée; est spécifique un facteur qui représente une aptitude psychique (e.g. le sens du rythme) qui n'intervient que dans une seule épreuve de cette batterie. De ce point de vue, à une seule épreuve j il peut correspondre plusieurs facteurs spécifiques, ou n'en correspondre aucun : si nous avons uniformément noté un facteur ψ_j^I , c'est pour simplifier les notations. Initialement, Thurstone suppose que les φ_α^I , ψ_j^I et ϵ_j^I sont deux à deux orthogonaux (non-corrélés) : quitte à modifier ensuite les φ_α^I (cf § 2.4.5); il appelle donc :

$h_j^2 = \Sigma \{ (a_j^\alpha)^2 \mid \alpha \in A \}$: *communauté* (en anglais *communality*) de l'épreuve j ;

b_j^2 : *spécificité* de l'épreuve j ;

$u_j^2 = b_j^2 + e_j^2$: *unicité* ("uniqueness") ou variance propre de l'épreuve j . La variance totale (1 par normalisation) est ainsi la somme de la *communauté* et de l'*unicité*. Reste à extraire des données (des notes n_j^i) la structure demandée par le modèle. Aux termes propres près (facteur spécifique et erreur) les vecteurs des épreuves n_j^I sont combinaison linéaire des facteurs communs φ_α^I : on désire expliquer les corrélations mutuelles par un nombre card A de facteurs communs aussi faible que possible. Géométriquement, le sous-espace de R^I engendré par les φ_α^I , est encore celui engendré par les $n_j^I = n_j^I - b_j \psi_j^I - e_j \epsilon_j^I$: et la dimension de ce sous-espace est le rang de la matrice des produits scalaires $r_{jj}^I = \langle n_j^I, n_j^I \rangle$. Pour $j \neq j'$, on a $r_{jj'}^I = r_{j'j}^I$, de par les propriétés d'orthogonalité des facteurs et des erreurs; pour $j = j'$ on a $r_{jj}^I = h_j^2$, la *communauté* de l'épreuve j ou variance des notes expliquées par les facteurs communs. On conçoit dès lors que l'analyse factorielle part de la matrice r_{jj} , (matrice des corrélations entre notes) dont la diagonale est modifiée; en retranchant de $1 = r_{jj}$, la variance d'erreur e_j^2 (dont l'ordre de grandeur au moins est accessible, si l'on soumet plusieurs fois les mêmes sujets à l'épreuve j ou à des épreuves parallèles), et la variance spécifique b_j^2 (beaucoup plus cachée ! en fait les divers auteurs estiment plutôt directement h_j^2 d'après la corrélation de l'épreuve j aux autres épreuves de la batterie : ces corrélations étant, selon le modèle de Thurstone d'autant plus élevées, que l'épreuve j est plus riche en facteurs communs). Nous reviendrons sur cette analyse (cf §§ 2.4.4 & 2.4.5) qui selon nous finit par s'enliser, mais recourt à certains bons principes : ici nous ne donnons la philosophie de Thurstone que pour servir de cadre à celle de Spearman, qui l'a précédée.

2.4.3. Le facteur général :

Dans le langage du modèle de Thurstone, il est facile d'énoncer la thèse de Spearman : pour Spearman, il n'y a qu'un seul facteur commun φ^I , appelé puisqu'il est unique, *facteur général*. Toute épreuve peut

être décomposée suivant la formule :

$$n_j^I = a_j \varphi^I + b_j \psi_j^I + c_j \varepsilon_j^I ;$$

formule qui ne comporte outre l'erreur ε_j que deux facteurs proprement dits : le facteur général, et le facteur spécifique à l'épreuve j :

C'est pourquoi Spearman dit : "doctrine éclectique, les deux facteurs".

Algébriquement on dira que la matrice des corrélations :

$$r_{jj'} = \langle n_j^I, n_{j'}^I \rangle = a_j a_{j'} + \delta_{jj'} (b_j^2 + c_j^2)$$

(où on note $\delta_{jj'} = 0$ si $j \neq j'$; $\delta_{jj} = 1$) est de rang 1, aux termes diagonaux près dont il faut retrancher la variance propre ($b_j^2 + c_j^2$) pour ne garder que la *communauté* $h_j^2 = a_j^2$. Soit donc u, v, u', v' quatre épreuves différentes (toutes dans J); le mineur (ou déterminant du sous-tableau de corrélations $\{u, v\} \times \{u', v'\}$ est nul :

$$(r_{uu'} \times r_{vv'}) - (r_{uv'} \times r_{vu'}) = 0$$

Spearman appelle *différence de tétrade* ce mineur où interviennent quatre épreuves distinctes; la condition de Spearman est que toutes les différences soient nulles.

En réalité les différences ne sont pas nulles. Spearman le sait aussi bien que Thurstone, et il est instructif de voir comment Spearman a pu concevoir que les différences de tétrades fussent nulles et persévérer à l'affirmer. Il y a deux raisons *a priori* pour que les différences ne soient pas nulles : 1°, les coefficients de corrélations empiriques diffèrent des coefficients vrais par des fluctuations d'échantillonnage; 2°, si les épreuves j et j' sont voisines, elles mettent en jeu à peu près le même facteur spécifique, ce qui accroît d'autant leur corrélation.

C'est en tenant compte des fluctuations d'erreurs que Spearman, ayant trouvé les différences de tétrade issues d'un tableau de corrélation entre tests étroitement resserrées autour de zéro, a pensé avoir découvert qu'elles étaient, au fond, nulles; car au contraire les différences issues d'un tableau de mesures somatiques s'écartent manifestement de zéro (cf *The abilities of man*, pp. 140 sqq.). Et il n'affirme la nullité des différences qu'à condition que les deux paires d'épreuves $\{u, v\}$ et $\{u', v'\}$ soient éloignées l'une de l'autre (en sorte que la corrélation entre un j et un j' ne puisse comporter une part due à leur spécificité voisine; Spearman dit encore à un facteur propre au groupe auquel toutes deux appartiennent). D'ailleurs Spearman n'a pas débuté par le calcul des différences de tétrade, mais justement par l'étude des corrélations entre deux épreuves u et v représentant deux versions d'une même tâche j ; et deux autres épreuves u', v' représentant deux versions d'une autre tâche; et ayant calculé par une habile compensation des fluctuations d'erreur (*) la corrélation entre ce que

(*) La formule de Spearman est :

$$s_{jj'} = (r_{uu'} r_{uv'} r_{vu'} r_{vv'})^{1/4} (r_{uv} r_{u'v'})^{-1/2}$$

ce qu'on démontrera en posant, selon un modèle analogue à celui de

$$\text{Thurstone} : n_u^I = a_u \theta^I + c_u \varepsilon_u^I ; n_v^I = a_v \theta^I + c_v \varepsilon_v^I$$

$$n_{u'}^I = a_{u'} \theta'^I + c_{u'} \varepsilon_{u'}^I ; n_{v'}^I = a_{v'} \theta'^I + c_{v'} \varepsilon_{v'}^I,$$

formules où θ^I, θ'^I sont unitaires et proportionnelles à la part des facteurs communs dans les tâches j et j' respectivement; tandis que les ε^I sont quatre termes d'erreur orthogonaux entre eux et aux θ^I, θ'^I ; d'où pour $s_{jj'}$, = $\text{corr}(\theta^I, \theta'^I)$ la valeur annoncée.

Thurstone appellerait la part des facteurs communs dans la tâche $j(\sum_{\alpha} a_j^{\alpha} \varphi_{\alpha}^I)$ et la part des facteurs communs dans la tâche $j'(\sum_{\alpha} a_j^{\alpha} \varphi_{\alpha}^I)$, il s'est étonné de trouver cette corrélation s_{jj} , proche de 1, comme s'il n'y avait qu'un seul facteur ($\sum_{\alpha} a_j^{\alpha} \varphi_{\alpha}^I$, n'est en toute généralité parallèle à $\sum_{\alpha} a_j^{\alpha} \varphi_{\alpha}^I$, que s'il n'y a qu'un seul φ_{α}^I). Pour Thurstone (Colloque C.N.R.S. 1955 déjà cité) "Il semble étonnant que le problème de Spearman n'a jamais été formulé en fonction du rang de la matrice de corrélation, bien que Spearman ait eu la collaboration de mathématiciens compétents. Si la question avait été formulée ainsi..., l'analyse multifactorielle se serait développée immédiatement". A la vérité, le modèle multidimensionnel avait précédé le modèle unidimensionnel. La théorie du facteur général de Spearman n'était née que pour affirmer que les dimensions mentales se distinguaient des dimensions somatiques par une structure plus simple. Un quart de siècle après Spearman, un autre auteur américain, Thurstone, se trouvait par la force des choses, ramené aux conceptions de Galton et de Pearson; et en formulait une théorie sans référence précise à ces auteurs.

2.4.4. Composantes principales et facteurs :

G. Kendall écrit (*) : "Dans l'analyse en composantes on part des observations et l'on cherche des composantes, avec l'espoir de réduire le nombre des dimensions suivant lesquelles varient les données et aussi parfois d'interpréter les composantes : on va des données vers un modèle hypothétique. En analyse factorielle, c'est l'inverse; partant d'un modèle, on doit chercher s'il s'accorde avec les données et, si tel est le cas, on estime les paramètres". Maint statisticien pour qui analyse en composantes principales est à peu près synonyme d'analyse factorielle, s'étonnera de voir opposer ces deux méthodes comme si la première seule était véritablement inductive. Or pour Kendall, l'analyse en composantes (le terme de composantes principales vient de Hotelling) est ce qu'usant d'une analogie mécanique nous appelons recherche des axes principaux d'inertie; tandis que par analyse factorielle il entend l'ensemble des procédés auxquels recourt Thurstone et ses émules pour interpréter la structure (parfois hiérarchisée) d'un ensemble d'épreuves, notamment d'après le modèle des facteurs communs et spécifiques. A cette distinction de principe, Kendall sait qu'il ne correspond pas d'opposition pratique absolue : car, surtout avant l'avènement des ordinateurs, algorithmes de calculs des facteurs, et hypothèses d'interprétation s'entremêlent. Mais il est juste de considérer d'abord (§ 2.4.4) les principes à la fois géométriques (axes d'un nuage de points) et algébriques (réduction d'un tenseur) de l'analyse inductive d'un tableau de mesures, principes désormais indispensables à l'analyse des données; et de critiquer ensuite (§ 2.4.5) les règles toujours discutées que les psychologues ont tenté de formuler comme des méthodes mathématiques conduisant à la découvertes des structures.

Pour définir les axes principaux d'inertie (ou les composantes principales) associés au tableau de notes normalisées défini au § 2.4.2, on peut considérer soit le nuage $N(I)$ des individus (Kendall, *op. laud.*) soit une sorte de nuage $N(J)$ des épreuves (Thurstone). Nous commencerons par $N(I)$, dont la construction nous est très familière (cf [Repr. Eucl.] TII B n° 2).

L'espace R_J des suites de notes indicées par $j \in J$ (une coordonnée ou note par épreuve) est muni de la norme euclidienne usuelle, somme des carrés des coordonnées. A tout individu i est associé le point

(*) A course in multivariate analysis; *Griffin ed.*; Londres (1957); p. 37.

$n_j^i = \{n_j^i | j \in J\}$, dont les coordonnées sont les notes qu'il a obtenues; ce point étant muni de la masse $1/\text{card } I$ (inverse du nombre de sujets). La forme quadratique d'inertie σ_{JJ} du nuage $N(I)$ ainsi construit n'est autre que la matrice des corrélations entre épreuves : $\sigma_{jj'} = r_{jj'}$, (car les notes sont normalisées). Les formes linéaires coordonnées sur les axes principaux d'inertie de $N(I)$ sont définies par le système :

$$\forall j \in J : \varphi_\alpha^j = \lambda_\alpha \sum \{\varphi_\alpha^{j'} r_{j,j'} | j' \in J\};$$

(où λ_α est le moment principal d'inertie); et la coordonnées de i sur l'axe α est donc :

$$\varphi_\alpha^i = \sum \{\varphi_\alpha^j n_j^i | j \in J\}; \text{ d'où } \varphi_\alpha^I = \sum \varphi_\alpha^j n_j^I.$$

Soit maintenant dans R^I l'ensemble des vecteurs unitaires n_j^I représentant les épreuves j de J . Thurstone recherche un premier axe puis un plan, puis un sous-espace de dimension 3, etc. passant par l'origine et tel que la somme des carrés des projections sur lui des vecteurs n_j^I soit maxima (ou encore soit minima la somme des carrés des distances aux axes des extrémités des vecteurs n_j^I : il s'agit, ici encore d'un problème de moindres carrés). Si l'on désigne par $\varphi_\alpha^I = \sum \varphi_\alpha^j n_j^I$ un tel vecteur axial, la recherche des φ_α^j équivaut à l'étude en axes obliques (les n_j^I formant entre eux des angles dont les cos sont les r_{jj}) du nuage des n_j^I eux-mêmes, rapporté non à son centre de gravité mais à l'origine. Dans ce système de coordonnées, on a pour matrice d'inertie σ , la matrice unité; et pour matrice de la métrique m la matrice des cosinus r_{jj} ; ; donc pour les φ_α^j l'équation $\varphi_0 \sigma_0 m = \lambda \varphi$ (cf [Repr. Eucl.]) n'est autre que celle trouvée déjà à partir du nuage $N(I)$: $\varphi_\alpha^j \cdot r_{JJ} = \lambda_\alpha \varphi_\alpha^j$.

Quand en 1931 (*), Thurstone (à l'instigation d'un astronome, le professeur W. Bartky; cf *Multiple factor analysis*, p. 474) entreprit d'appliquer la méthode des moindres carrées à l'analyse factorielle, les moyens de calcul étaient rudimentaires; et les algorithmes de diagonalisation eux-mêmes peu étudiés : c'est en 1933 que Hotelling, ayant en vue l'analyse factorielle, proposa de rechercher les vecteurs propres par itération et orthogonalisation. Longtemps les facteurs furent calculés non pour satisfaire à une définition mathématique rigoureuse, mais par des tâtonnements dont il suffira de donner un aperçu en décrivant brièvement la méthode du *centroïde* (**). Généralement les notes n_j^I mesurent des aptitudes estimables, et sont toutes corrélées positivement au premier facteur φ_1^I , qui n'est autre que le *facteur général* de Spearman; de plus les n_j^I sont corrélées positivement entre elles, et forment dans R^I comme un faisceau de vecteurs assez resserré à l'intérieur duquel est φ_1^I (cf fig. 2-7).

(*) Cf. *Multiple Factor Analysis in Psychological Review* V. 38 pp. 406-27 (1931); première publication de L. Thurstone, bientôt suivie d'un livre : *Theory of Multiple Factors*, Ann Arbor, (1932).

(**) Ainsi nommée par Thurstone, mais pratiquée déjà par Burt (cf *Colloque C.N.R.S.* 1955; p. 84)

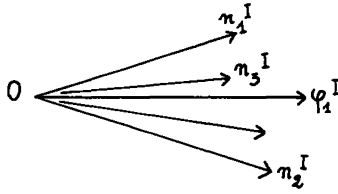


Figure 2-7: le faisceau des notes n_j^I est resserré autour du facteur général φ_1^I

D'où l'idée de déterminer rapidement φ_1^I comme étant la moyenne des n_j^I (donc l'extrémité du vecteur φ_1^I est au centre de gravité, en anglais *centroid*, des extrémités des vecteurs n_j^I). Pour déterminer φ_2^I , on se place dans l'hyperplan H_1^I de R^I orthogonal à φ_1^I , et on projette sur cet hyperplan les notes n_j^I (ce qui équivaut à retrancher des n_j^I leur composante dans la direction du premier facteur φ_1^I). Dans H_1^I , les projections $n_{j;1}^I$ des n_j^I ne forment pas un faisceau étroit (comme les n_j^I eux-mêmes dans R^I); leurs corrélations mutuelles peuvent être négatives (i.e. en langage géométrique, deux $n_{j;1}^I$ peuvent former un angle obtu). Mais par tâtonnement on change les signes de certains des $n_{j;1}^I$ afin d'avoir dans l'ensemble le plus possible de fortes corrélations positives : on détermine alors φ_2^I comme une moyenne (cf fig. 2-8); et ainsi de suite. On conçoit que sur des facteurs ainsi mis en place, les psychologues s'adonnent sans retenue à toutes sortes de rotations et passages en axes obliques.

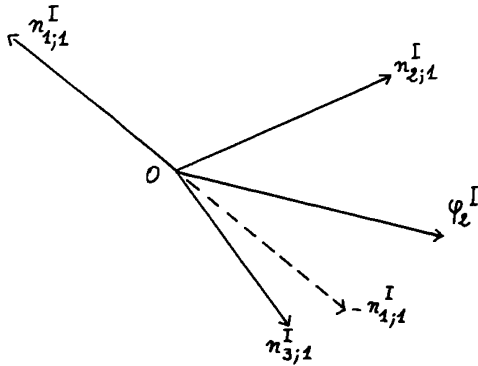


Figure 2-8 : principe de la recherche du second facteur par la méthode du centroïde, dans le cas linéaire où il n'y a que trois épreuves, on a changé le signe de $n_{1;1}^I$ afin d'avoir un faisceau assez étroit définissant φ_2^I .

2.4.5. L'interprétation des structures :

Voici comment L.L. Thurstone condense en un paragraphe (*op. laud.*, (1947), p. ix) les principes qu'il croit indispensables à l'analyse factorielle (*) :

"Bien que l'analyse multifactorielle ait suscité de profondes controverses, nous croyons probable que la grande majorité des études récentes ont accepté les points suivants :

formulation matricielle du problème,
interprétation des théorèmes de Spearman comme se rapportant au cas particulier d'une matrice de rang 1,
concept de structure simple,
nécessité des rotations pour l'interprétation scientifique,
effort pour donner, dans l'interprétation, un sens aux facteurs primaires,
recours à des systèmes obliques d'axes de référence,
principe d'invariance configurationnelle".

A quoi nous ajouterons, l'estimation des *communautés* (cf § 2.4.2) sur laquelle le psychologue américain insiste souvent. En expliquant ces thèses de Thurstone, qui, à la vérité, n'ont pas toutes triomphé dans les *controverses*, nous les confronterons à notre pratique de l'analyse des correspondances.

Quant à la formulation matricielle, au recours à la géométrie multidimensionnelle, il faut rendre à Thurstone un chaleureux hommage; tout en affirmant que les méthodes algébriques aujourd'hui enseignées partout, permettent d'aller plus loin qu'il ne le fit (cf §§ 3.2.3 & 3.5.1).

Il est incontestable que les calculs de Spearman se réfèrent à un tableau de corrélation $\{r_{jj}\}$ qui est le rang 1 (à la nécessité près de modifier sa diagonale pour remplacer les 1 par des *communautés*). Mais le modèle de Spearman continue d'éclairer l'interprétation d'une analyse, même s'il est nécessaire de recourir à plusieurs facteurs. Dans de nombreuses analyses et particulièrement en psychologie, le premier facteur a une interprétation distinguée; il se détache des autres comme la valeur propre λ_1 se détache de la suite dense des $\lambda_2, \lambda_3, \lambda_4 \dots$. Cependant si l'on se permet de ne pas conserver le système des axes principaux du nuage, qu'on fait tourner ce système et qu'on le déforme; on pourra éliminer le facteur général. Prenons pour fixer un exemple la batterie des épreuves du concours d'admission à l'Ecole Polytechnique (cf F. Nakhlé, thèse et [Dédou] § 3.2 (**)) dans le plan des deux premiers axes principaux on a schématiquement la figure 2-9 : l'axe 1 est un axe de niveau général; sur l'axe 2 les aptitudes techniques s'opposent aux aptitudes littéraires; mais on peut introduire deux axes obliques (en tireté sur la figure) : technique et lettres.

Pour décider entre ces deux principes d'interprétation (les axes principaux tels quels; ou : de nouveaux axes) il faut s'efforcer de saisir le nuage $N(J)$ dans toute sa complexité : $N(J)$ n'est pas seulement un vague ellipsoïde qu'on se bornera à projeter sur le sous-espace engendré par les premiers axes principaux; $N(J)$ a une forme et il a un sens dont la découverte conduit inductivement à des modèles structuraux (cf *infra* § 3.4.5) et [Prat. Corr.] § 3.2). C'est ce que Thurstone a en vue

(*) Dans la traduction, nous avons pris la liberté d'aller à la ligne, afin de rendre plus sensible l'énumération de ces principes.

(**) L'analyse des notes obtenues par les candidats à l'Ecole Polytechnique est publiée dans le présent cahier pp. 367-379.

quand il parle de structure simple (*). Mais les axes principaux sont-ils la meilleure charpente pour découvrir et souligner ces structures : Thurstone ne le pense pas.

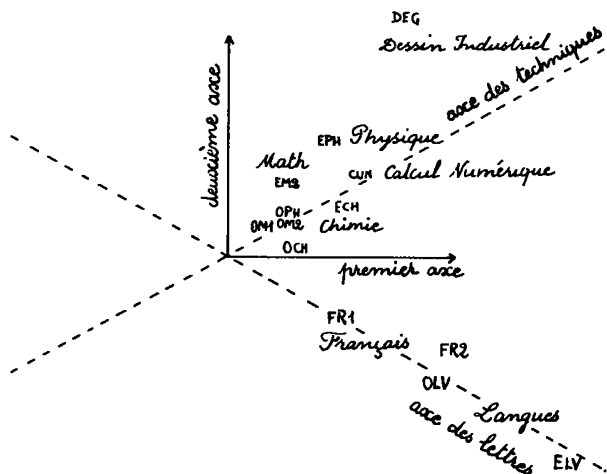


Figure 2-9: Concours d'admission à l'École Polytechnique. Analyse de correspondance du tableau dédoublé (cf § 3.7.4) des notes obtenues par 900 candidats à 20 épreuves écrites (E) ou orales (O); sur le graphique simplifié on a omis les sigles des notes complémentaires (9"; cf § 3.7.4) qui occupent le demi-plan $F_3 < 0$.

Certes les moyens de calcul dont dispose Thurstone ne suffisant pas en général à fixer les véritables axes principaux, il lui est difficile d'argumenter fermement ! Mais il y a plus : Thurstone ne croit pas possible d'appliquer en analyse factorielle le principe statistique de l'échantillonnage (uniforme et exhaustif; au moins dans un domaine d'étude judicieusement défini).

A telle enseigne qu'il prend pour cas modèle des vicissitudes de l'interprétation des axes en analyse factorielle, l'étude d'ensembles d'objets physiques (par exemple des cylindres) caractérisés par un ensemble J de paramètres géométriques et mécaniques : $J = \{ \text{hauteur } h, \text{ rayon } R, \text{ volume } V, \text{ masse volumique } \mu, \text{ masse totale } m, \text{ moment d'inertie } I, \dots \}$; (paramètres qui sont liés entre eux : e.g. $m = \pi \times \mu \times R^2 \times h$) l'échantillon I n'étant lui-même défini par aucune condition naturelle (il pourra considérer un ensemble I de 1000 cylindres de hauteur 10, 20, ..., 100; rayon 1, 2, ..., 10; masse volumique 10, 20, ..., 100; aussi bien que, un ensemble I' de 1000 cylindres de hauteur 1, 2, ..., 10; rayon 1, 2, ..., 10; masse volumique 1, 2, ..., 10). Et comme d'une analyse à l'autre, d'un échantillon à un autre (i.e. quand on passe du tableau $I \times J$ au tableau $I' \times J$) le nuage $N(J)$ se transforme mais demeure reconnaissable (car, par exemple le premier axe principal qui passait par V vient à passer par m , mais il y a toujours proximité entre V et m et opposition entre h et I), il parle d'invariance configurationnelle. Ainsi en serait-il quand en psychologie on passe d'un ensemble I d'étudiants grecs résidant à Paris à un ensemble I' de lycéens français, la

(*) cf Multiple Factor analysis p. 328 : "si l'on peut trouver un système d'axes tel que tout vecteur représentant une épreuve [vecteur n_j dans nos notations] appartienne à un ou plusieurs des hyperplans de coordonnées [sous-espace engendré par des axes] on dit qu'il y a structure simple : de plus, en déplaçant les axes, Thurstone et ses émules procèdent souvent à une classification des variables : un axe étant défini comme vecteur moyen d'un îlot de variables voisines.

batterie J pouvant elle-même quelque peu varier : l'interprétation devrait donc considérer $N(J)$ comme n'étant pas informe, sans être rigide; donc rechercher la forme et la signification suivant d'autres lignes que les axes principaux.

C'est ici que nous croyons qu'il faut mieux faire que Thurstone ! Quelque difficile que soit la définition de la base $I \times J$ de la collecte des données (et aussi du codage spatial en nuages $N(I)$ et $N(J)$) il est très fructueux de s'appliquer à en bannir l'arbitraire (cf *infra* §§3.7.1 *sqq*) pour parvenir à définir par le calcul des axes principaux directement interprétables (*). Pour Thurstone (Colloque C.N.R.S., 1955, p. 33) "En commençant une analyse factorielle, le chercheur doit décider lequel des deux objectifs suivants il va s'efforcer d'atteindre 1) reproduire les notes individuelles et 2) reproduire les corrélations". Ces deux objectifs - que nous interpréterons : 1) condenser les données et 2) les comprendre - ne lui semblent pas exactement compatibles : la condensation résulte au mieux de l'analyse en composantes principales de la matrice des corrélations empiriques; l'interprétation requiert l'estimation des *communautés* et des variances propres (cf § 2.2) et force rotations et changements d'axes. Selon nous, le modèle même de Thurstone est défectueux; en requérant des variables normées il mutile les données; en ne considérant que des tableaux de notes continues, il écarte les données les plus sûres : les tableaux de contingence, dont Pearson après Galton avait vu l'importance (cf § 2.2.7) et pour lesquels a été conçue l'analyse des correspondances (cf § 3).

2.4.6. Les écoles d'analyse factorielle :

Dans ce § nous évoquons d'abord la diversité des conceptions auxquelles se prêtent les formules mêmes utilisées par Thurstone; nous décrivons ensuite une construction mathématique différente due à Hotelling; nous terminons sur les rapports de l'analyse multifactorielle - méthode de traitement des tableaux de données qui ne s'écarte jamais beaucoup de la réduction d'une matrice carrée par diagonalisation - avec la statistique prônée par Fisher, laquelle requiert des hypothèses paramétriques précises et vise à déterminer les distributions exactes des grandeurs que l'on calcule.

L'analyse multifactorielle proposée par Louis Léon Thurstone intéressa de nombreux chercheurs qui construisirent des représentations multidimensionnelles, sans pour autant adopter quant à l'interprétation des structures, tous les principes du psychologue américain (cf § 2.4.5). Dans sa préface, Thurstone cite successivement Holzinger & Harman, Kelley, Burt, Thomson : tel accepte les axes obliques sans faire grand cas des structures simples; tel refuse les *communautés*; tel tente des rotations, sans cependant attendre rien de profond de l'interprétation des facteurs. Nous ne reprendrons pas en détail ces distinctions mais aurons une mention spéciale pour C. Burt; qui fut l'un des fondateurs de l'analyse factorielle (cf *supra* § 2.4, début; et *Colloque C.N.R.S.*; 1955; pp. 79 *sqq*) et a extrait des facteurs du tableau de contingence recensant les cooccurrences de l'ensemble des modalités de réponse aux questions successives d'un questionnaire (cf § 3.7.3). Il s'en faut toutefois d'un coefficient $p_j^{1/2}$ que les facteurs que C. Burt extrait de ce tableau soient ceux-mêmes de l'analyse de correspondance : ainsi au facteur trivial constant et égal à 1 (cf § 3.5.2), toujours éliminé en analyse de correspondance, répond chez Burt un premier facteur Φ^J , tel que

(*) Il faut avouer qu'un échantillonnage très particulier, une expérimentation, sur la nature contrainte per vexationes artis disait Bacon, peut conduire aux plus géniales découvertes : cf *supra* §§ 2.2.4 & 2.3.5; mais jusqu'à présent, dans les sciences de l'homme de tels essais n'ont rien produit de tel; et la respectueuse observation de la nature dans son équilibre propre nous semble mériter tous nos soins.

$$\phi(j) = p_j^{1/2}.$$

Nous avons dit le rôle de Harold Hotelling dans la recherche des composantes principales, connues aujourd'hui sous le nom qu'il leur a donné (§ 2.2.4) Hotelling est aussi l'auteur d'une méthode des corrélations canoniques qu'on a souvent mise en parallèle avec l'analyse factorielle, et que nous décrirons brièvement, car les constructions géométriques de Hotelling se retrouvent en analyse des correspondances (cf § 3.4.3). Le problème statistique que Hotelling entend résoudre, relève plutôt de la régression : c'est (cf *infra*) la recherche dans deux sous-espaces linéaires, des couples de fonctions les plus corrélées entre elles. Mais les équations que Hotelling doit résoudre sont finalement analogues à celles d'une analyse factorielle. Esquissions cette méthode en reprenant les notations adoptées ci-dessus. L'ensemble J des épreuves est partagé en deux sous-ensembles disjoints J_1 et J_2 : $J = J_1 \cup J_2$; $J_1 \cap J_2 = \emptyset$. Notons L_1^I l'espace vectoriel des combinaisons linéaires des notes J_1 et de même L_2^I pour J_2 . Chaque vecteur $y_1^{J_1} \in R^{J_1}$ définit une combinaison $y_1^I \in L_1^I$: sur l'ensemble des individus I on a :

$$y_1^I = \sum \{y_1^j n_j^I | j \in J_1\};$$

$$L_1^I = \{\sum \{y^j n_j^I | j \in J_1\} | y^{J_1} \in R^{J_1}\}.$$

On recherche des couples de combinaisons (y_1^I, y_2^I) l'une dans L_1^I , l'autre dans L_2^I dont le coefficient de corrélation soit maximum. Géométriquement, ceci revient à étudier dans l'espace euclidien R^I la figure formée par les deux sous-espaces L_1^I et L_2^I ; et à chercher en particulier les couples de vecteurs formant un angle minimal (donc dont le cosinus est maximal). Ce problème se résout en considérant les opérateurs π_1^2 (projection orthogonale de L_2 sur L_1) et π_2^1 (projection orthogonale de L_1 sur L_2). Les couples (y_1^I, y_2^I) cherchés ne sont autres que les couples de vecteurs propres des applications $\pi_1^2 \circ \pi_2^1$ (de L_1 dans L_1) et $\pi_2^1 \circ \pi_1^2$ (de L_2 dans L_2); les droites support de y_1^I et y_2^I étant projections orthogonales l'une de l'autre; on a :

$$y_1^I = r \pi_1^2 y_2^I; \quad y_2^I = r \pi_2^1 y_1^I,$$

où r est le cosinus ou coefficient de corrélation dit canonique, entre y_1 et y_2 . Pour Hotelling les deux groupes de notes J_1 et J_2 ont conjointement une distribution normale multidimensionnelle; et partant de l'hypothèse nulle que les deux groupes J_1 et J_2 sont non-corrélés entre eux; il s'interroge sur le seuil à partir duquel les coefficients de corrélation canoniques trouvés sont assez élevés pour entraîner le rejet de cette hypothèse d'indépendance (cf *Biometrika* T. 28 pp. 321-377; 1936).

A la vérité de semblables problèmes statistiques peuvent être posés à propos des diverses variantes de l'analyse factorielle. Dès 1928, Wishart détermine la distribution de la matrice de variance-covariance d'un échantillon issu d'une loi normale multidimensionnelle donnée (matrice considérée comme un élément aléatoire multidimensionnel). On conçoit qu'en prolongeant de telles recherches on puisse obtenir des épreuves de validité pour l'analyse en composantes principales : la validité du premier facteur correspondant e.g. au rejet de l'hypothèse nulle que les notes sont deux à deux non corrélées entre elles. Plus

difficile à aborder est le modèle même de Thurstone qui comporte pour chaque note outre la variance due aux facteurs communs une variance propre qui doit être estimée soit par la statistique, soit *a priori*. Sous les efforts de Lawley, Maxwell et Joreskog, le modèle s'est pourtant plié aux estimations et aux épreuves de validité.

Selon nous, les divers modèles normaux considérés en analyse factorielle ont pu servir à fixer sur un exemple précis les premières explorations multidimensionnelles de nos devanciers. Il en est resté des formules intéressantes, dans la mesure où tout calcul sur une loi normale, est aussi un calcul sur la matrice d'inertie d'un nuage, matrice qui existe en dehors de toute hypothèse paramétrique. Mais ici plus encore qu'ailleurs les épreuves de validité fondées sur des distributions probabilistes exactes nous sont suspectes. Thurstone lui-même s'efforçait d'échapper aux griffes de la statistique mathématique (cf *Multiple Factor Analysis*, préface pp. X sqq). "Factor analysis has not been generally accepted by mathematical statisticians... Il faut d'abord reconnaître qu'un savant qui cherche une théorie pour couvrir les phénomènes d'un certain domaine, une description de l'ordre sous-jacent à celui-ci, ne se soucie pas des distributions d'échantillonnage"... et après avoir rappelé que rien ne devait manquer dans le modèle mathématique; des *communautés*, rotations etc... auxquelles il est attaché (*), il émet des doutes sur l'universalité de la loi normale et répète que mieux vaut une mesure significative sans distribution d'échantillonnage, qu'une mesure triviale ou irrelevante choisie parce que la distribution en est connue".

Les libertés que non sans bon sens et mauvaise humeur revendiquait Thurstone, l'analyse des données, après lui, les a prises ...

2.5. L'analyse des données non-métriques.

2.5.1. Expansion du domaine de la statistique :

L'étude mathématique de la corrélation, (par quoi il faut entendre non le calcul d'un coefficient qui ne mesure celle-ci qu'imparfaitement; mais la liaison même entre deux grandeurs entachées d'aléas qui varient simultanément) aurait pu débiter dès le seuil du XIX^e siècle avec Laplace et Gauss : pourtant cette étude initiée par Bravais (§ 2.2.3) ne s'est développée qu'avec l'Ecole biométrique anglaise fondée par F. Galton (§ 2.2). L'analyse factorielle (§ 2.4), dont les premiers traits se trouvent chez Bravais et Schols (§ 2.2.3) et dont K. Pearson a vu tout le dessin à propos de données anthropométriques (§ 2.2.5), n'a pris son essor que dans la psychométrie; laquelle a pour objet les grandeurs mentales qui ne peuvent être atteintes que comme sous-jacentes à un faisceau de mesures. L'analyse multidimensionnelle des données non-métriques, (c'est-à-dire d'un système de variables aléatoires dont chacune prend pour valeurs non des nombres - mesures ou notes- mais les éléments d'un ensemble fini) est en germe dans les réflexions sur la contingence que suggère à K. Pearson l'examen d'un tableau rectangulaire (§ 2.2.7); on trouve chez Fisher et tel de ses contemporains Hirschfeld ou Maung (cf §§ 2.3.5. & 3.5.2), les premières formules de l'analyse des correspondances. Mais c'est une fois de plus, l'expansion du domaine de la statistique qui a déterminé le progrès de celle-ci : ce que nous appelons analyse des données est issu premièrement de la considération des données verbales, réponses à des questionnaires ou corpus de textes (§ 3.2). Chez les psychologues et les sociologues américains on voit l'analyse des réponses se substituer graduellement à

(*) Sans adopter la méthode de Thurstone nous soulignons que la difficulté majeure en analyse multidimensionnelle n'est pas de rejeter l'hypothèse nulle, de mesurer globalement par un nombre, une interaction..., mais de donner de cette interaction une expression mathématique interprétable : c'est peu de reconnaître l'existence de liens significatifs, il faut en déchiffrer la signification; cf § 3.8.5.

l'analyse des notes. Nous suivrons principalement cette histoire d'après les ouvrages de synthèse de Torgerson et de Coombs (§§ 2.5.2 & 2.5.5), qui dans leur discipline ont fait époque. Sans oublier que vers l'ordinateur, outil indispensable de la statistique multidimensionnelle convergent aussi les constructions taxinomiques des écologistes et les rêves ambitieux de la reconnaissance automatique des formes (§ 2.5.6)

2.5.2. La construction des échelles :

Dans son ouvrage *Theory and methods of scaling* (J. Wiley & Sons, Inc.; N.-Y.; 1958), W.S. Torgerson part de la construction d'échelles psychophysiques associées à une grandeur physique unidimensionnelle, pour aboutir à la construction d'échelles multidimensionnelles rendant compte de l'évaluation des qualités les plus diverses, d'après des données générales; donc à ce que nous appelons l'analyse des données.

On sait (cf § 2.4.1) que les constructions des psychophysiciens reposent sur l'élaboration statistique de jugements rendus par des sujets quant à l'égalité des deux stimulus, la perception d'une différence (seuil), l'égalité de deux variations, etc ... Pour Torgerson (p. 205) ces expériences sont comprises dans un modèle général qu'il attribue en substance à Thurstone : "A l'attribut [qualité, grandeur] étudié, est sous-jacent un continuum [i.e. une droite, ou plus généralement un espace]. Chaque fois qu'un stimulus est présenté à un sujet, se produit comme la perception distincte (discrimination; *discriminal process* dit T.) d'un point sur ce continuum. Mais du fait de causes diverses, à des présentations répétées d'un même stimulus, ce n'est pas toujours le même point qui est associé... On suppose que les points (valeurs) associées à un même stimulus sont distribués normalement sur le continuum. Divers stimulus peuvent différer quant à la moyenne de cette distribution (leur valeur sur l'échelle) et aussi quant à l'écart-type (ou dispersion du processus de discrimination)". Donc (dans le cas unidimensionnel) la perception associée à un stimulus s aura la valeur entachée d'aléa :

$$x = m(s) + \sigma(s)\xi$$

où $m(s)$ est la valeur moyenne, $\sigma(s)$ l'écart type et ξ une grandeur aléatoire normale de variance 1; et la comparaison de deux stimulus s et s' présentés ensemble se fera d'après la différence :

$$x - x' = (m(s) - m(s')) + \sigma(s)\xi - \sigma(s')\xi';$$

formule où il faut prendre garde que les termes d'erreur ξ et ξ' sont généralement corrélés.

De cette *loi des jugements de comparaison*, Torgerson passe à la *loi des jugements de catégorie* (p. 206). En bref il s'agit d'expériences où le sujet décide si le stimulus s rentre dans la catégorie c : on peut dire que pour répondre, le sujet doit comparer s aux deux bornes inférieures et supérieures de l'intervalle associé à c ; donc il s'agit de comparaisons. Par généralisations successives on comprendra dans ce modèle, des jugements tels que : un tel (stimulus s) est poète (catégorie c)...

Pour le cas multidimensionnel, Torgerson remarque d'abord (p. 249) qu'un stimulus participe généralement d'un grand nombre de dimensions que le sujet n'a pas à considérer : par exemple si l'on compare quant au sens les mots aimable, charmant, avenant etc..., il ne s'agit ni du nombre de lettres, ni de la prononciation, ni de la proximité dans le dictionnaire etc. Reste qu'il y a des situations vraiment multidimensionnelles (comme la couleur en psychophysique; cf § 2.4.1). Les expériences relatives aux stimulus multidimensionnels sont analogues à celles du cas unidimensionnel; mettant en jeu égalités, seuils, variations ... Quant à leur élaboration, elle requiert d'une part une conception géométrique du continuum spatial sous-jacent (qui n'est plus la droite);

d'autre part un modèle expliquant les pourcentages des divers jugements rendus par les sujets, par la disposition des points-stimulus sur ce continuum (e.g. le taux de confusion entre deux stimulus est relié à leur distance et à un écart-type, ou seuil, universel...).

La géométrie multidimensionnelle peut utiliser de multiples formules de distance : par exemple, dans R^n , la distance L_1 :

$$\|x - y\|_1 = \sum \{|x^i - y^i| \mid i = 1, \dots, n\},$$

somme des valeurs absolues des différences de coordonnées. Cette distance associée en psychologie au nom d'Attneave (cf Dimensions of similarity; in *Amer. J. Psychol.*, T. 63 pp. 516-556; 1950) peut s'accorder aux résultats de certaines expériences (nous avons exposé sur ce point les vues de R.N. Shepard : cf Sur l'analyse des matrices de confusion § 4; in *Revue de Stat. Appl.* T. 18 pp. 5-63; 1970). Mais la clarté et la simplicité des propriétés mathématiques (particularly neat and simple mathematical properties, écrit T. ; p. 252) nous ramène nécessairement à la géométrie euclidienne; et au problème fondamental suivant : étant donné un ensemble I d'éléments et leurs distances deux à deux d_{ii} , donner

dans un espace euclidien E une représentation de l'ensemble I respectant ces distances. On sait que ce problème admet une solution exacte (sous réserve de conditions de positivité; dont la plus simple est l'inégalité du triangle) dans un espace E de dimension Card $I - 1$; et une solution approchée (liée à la recherche des axes principaux d'inertie) si on borne *a priori* la dimension de E : en substance dans son livre ou dans ses mémoires Torgerson donne la plupart des résultats que nous exposons dans la leçon [Repr. Eucl.] TII B, n° 2, à ceci près qu'il ne considère pas des points de masses inégales.

Dans les derniers chapitres de son livre, Torgerson propose de placer sur une même échelle (ou continuum spatial sous-jacent; soit à une, soit à plusieurs dimensions) non seulement les stimulus, mais les sujets auxquels ceux-ci sont présentés. Les données requises pour cela sont autres que les comparaisons, distinctions et différences familières à la psychophysique : la tâche des sujets est d'indiquer leur propre attitude vis-à-vis des stimulus proposés. Torgerson parle de données catégorielles (*categorical data*) : il s'agit en fait principalement du dépouillement du tableau $I \times J$ des réponses d'un ensemble I de sujets à un questionnaire dont l'ensemble des modalités de réponse est J . Par exemple si chaque question admet les trois modalités de réponse Oui, Non, Ne sait Pas, on aura pour p question : Card $J = 3p$; c'est ce que nous appelons codage sous forme disjonctive complète (cf § 3.7.3). Ici les travaux les plus remarquables que puisse citer Torgerson, sont dus à L. Guttman; dont nous exposerons les vues en anticipant sur l'analyse des correspondances, objet du § 3.

2.5.3. Les modèles de Guttman :

C'est au début de ses recherches (cf the quantification of a class of attributes, in P. Horst & Coll. *The Prediction of Personal Adjustment* N.-Y. 1941), que L. Guttman écrit les formules les plus générales, dans l'esprit de l'analyse des données. De là, faute de moyens de calcul croyons-nous, il passa à des modèles - scalogramme, simplexe, circumplex, radex ... qu'il conçut pour embrasser une diversité de structure de plus en plus grande, mais sans parvenir à retrouver la généralité autrement qu'au prix de constructions ingénieuses mais arbitraires.

Torgerson (p. 301) montre un tableau, dont les lignes - numérotées de 1 à N - se rapportent chacune à un sujet; et les colonnes vont par groupes de 3 ou 4, chacun afférant à une question : si le sujet i a répondu à la question q par la modalité c , la i -ème ligne comporte des zéros dans toutes les colonnes du groupe q sauf la colonne q_c où est 1. Pour un tel tableau (cf pp. 338 sqq.). Torgerson écrit d'après Guttman

(1941) les équations des *composantes principales* qui ne sont autres que les facteurs de l'analyse des correspondances; avec pour critère (cf infra §§ 3.4.3 & 3.5.2) "maximiser le rapport de la variance entre individus à la variance totale; ou puisque ce rapport n'est autre que le carré d'un coefficient de corrélation, maximiser le coefficient de corrélation".

En 1941, de telles équations ne peuvent être résolues que dans des cas particuliers. Guttman en distingue bientôt un qui est d'importance : celui des échelles ou scalogrammes associés à une structure d'ordre stricte à la fois sur les colonnes - modalités de réponse - et sur les lignes - systèmes (on dit souvent *patron*; en anglais *pattern*) de réponses d'un sujet-type. C'est le tableau en parallélogramme dont nous rappelons le dessin, (cf tableau 2-1). On voit que la connaissance du rang

		OUI				NON			
		1	2	3	4	1	2	3	4
sujets-type	I	1	1	1	1				
	Q		1	1	1	1			
				1	1	1	1		
					1	1	1	1	
						1	1	1	1

Tableau 2-1 : exemple de scalogramme de Guttman, à quatre questions numérotées de 1 à 4 et cinq sujets-type; allant de l'acceptation générale au non absolu; le Oui à la question q impliquant nécessairement le Oui à toute question q' de rang supérieur à q . On notera que les deux modalités de réponse à une même question - e.g. Oui 2 et Non 2 - ne sont pas juxtaposées, mais séparées.

d'un sujet dans la suite des sujets-type (cinq systèmes de réponse sont permis dans l'exemple du tab. 2-1) permet de reconstituer complètement les réponses données par ce sujet : c'est ce que Guttman appelle la *reproductibilité* (*). En fait, il est rare que les réponses suivent exactement le modèle du scalogramme parfait : mais pour approcher cet idéal, il faut jalonner le thème qu'on se propose d'étudier (e.g. la confiance du soldat en l'armée : Guttman travaille aux U.S.A. pendant la deuxième guerre mondiale) par une suite unidimensionnelle de questions graduées. L'un des mérites de la méthode est justement de servir de guide pour la construction d'un tel questionnaire.

Répétons que Guttman ne dispose pas d'un ordinateur : il traite le tableau des données par essais de permutation des lignes et des colonnes, matérialisées sur un support ingénieusement chevillé, afin de faire apparaître autant qu'il est possible la forme parallélogrammatique (cette technique est toujours préconisée en France par J. Bertin qui l'a perfectionnée). Mais pour un scalogramme parfait, des calculs algébriques généraux (qui relèvent de la théorie des équations aux différences finies et des polynômes orthogonaux), fournissent à Guttman la suite des composantes principales, ou facteurs : il se trouve que sur I (ensemble des lignes ou sujets-type) les facteurs de rang 2, 3 etc... s'expriment en fonction du premier par des polynômes de degré 2, 3 etc. C'est ce que nous avons appelé l'effet Guttman (cf TII B, n° 7, § 3).

(*) Signalons au passage qu'aux exigences de reproductibilité formulées par L. Guttman, répond en analyse des correspondances la formule de reconstitution des données en fonction des facteurs.

Dans le modèle foncièrement unidimensionnel du scalogramme, le premier facteur suffit à exprimer toute la structure : pourtant Guttman s'intéresse aussi aux facteurs suivants. Ainsi le deuxième F_2 ($F_2 = F_1^2 - cte$) qui est négatif au centre de la suite des types et fortement positif à ses extrémités représente l'*intensité* (des convictions du sujet) etc ... Mais dans le cas général qui s'écarte du modèle du scalogramme n'y-a-t-il rien de plus à faire que serrer au plus près ce modèle, notamment en rangeant lignes puis colonnes dans l'ordre du premier facteur ? Torgerson exprime l'opinion généralement reçue quand il écrit (en 1958) des facteurs suivants : "they are all extraneous solutions", des solutions parasites; mais il ajoute en italique : "*as for as the problem posed is concerned*", pour autant qu'on se borne au problème posé (construire une échelle). La conception multidimensionnelle de l'analyse factorielle n'est donc pas tout à fait perdue de vue.

Cependant c'est dans une tout autre voie que le Guttman des années 50 cherche la description de données plus riches. Comme Thurstone (cf § 2.4.2), Guttman postule d'abord que les notes reçues par un sujet à un ensemble d'épreuves sont des variables aléatoires dont la distribution conjointe est normale multidimensionnelle. Mais il considère un ensemble potentiellement infini d'épreuves, formant un continuum qu'il appelle *univers de continu* : c'est poser opportunément le problème (souvent évoqué dans ces leçons : cf § 3.7.1) de la représentativité d'un questionnaire, considéré comme échantillon d'un domaine à explorer (l'*univers* de L.G.). Dans le modèle le plus simple - celui du *simplexe* - l'*univers* est une demi-droite : aux points abscisse $t_1 < t_2 < t_3$, correspondent des notes n_1, n_2, n_3 , de variances respectives t_1, t_2, t_3 ; les différences $(n_2 - n_1)$ et $(n_3 - n_2)$ étant elles-mêmes des variables normales *indépendantes* entre elles de variances $(t_2 - t_1)$ et $(t_3 - t_2)$: c'est introduire en psychologie la fonction aléatoire du mouvement brownien linéaire !

Le modèle du simplexe représente bien - au moins analogiquement - l'*univers* de contenu associé au progrès de l'esprit dans une direction déterminée : e.g. l'étude d'une seule science. Mais Guttman conçoit que s'offrent à l'esprit plusieurs voies : la musique, l'éloquence, la grammaire, la logique, les mathématiques; que ces voies soient comme les rayons d'un même cercle (e.g. la logique serait entre la grammaire et les mathématiques; l'éloquence entre la grammaire et la musique;) - c'est le modèle du *circumplex* -; et enfin que dans chacune de ces voies le progrès se fasse suivant la loi du *simplexe* : c'est le modèle du *radex*, ou expansion radiale. Torgerson (P. 346), citant Bennett, remarque judicieusement que deux aptitudes peuvent contribuer à une même tâche de façon soit conjonctive (comme la connaissance de la langue française et celle de l'histoire simultanément requises pour répondre à des questions d'histoire posées en français) soit disjonctive (comme quand un même problème peut être résolu par la géométrie ou par l'algèbre)...

Tout cela donne à réfléchir; mais n'aboutit pas à un système pratique d'analyse et d'interprétation des données. Un tel système était au contraire en germe dans les composantes principales de Guttman, que nous devions retrouver dans l'analyse des correspondances (§ 3).

2.5.4. Les structures latentes de P. Lazarsfeld :

Pour Torgerson les modèles de Guttman ne rendent pas explicitement compte de la part d'aléa dans les réponses d'un sujet : tel est bien le cas du scalogramme parfait; et aussi du simplexe : car quand on dit que l'ensemble des notes $\{n_j\}$ obtenues par le sujet i est distribué normalement on ne se réfère pas à la variabilité dont sont susceptibles les réponses d'un même sujet, mais au nuage même des sujets, chacun représenté par le point $\{n_j\}$. Aussi après le chapitre intitulé *Modèles déter-*

ministres pour les données catégorielles, où règne L. Guttman, Torgerson présente-t-il les idées de P. Lazarsfeld sous le titre de *Modèles probabilistes pour les données catégorielles*.

Selon nous, pas plus que le *circumplex* et le *radex* de L. Guttman, les *structures latentes* de P. Lazarsfeld n'offrent de procédé défini susceptible d'être reçu dans la pratique constante après les calculs de moyenne ou de coefficient de corrélation et les compilations d'histogrammes ou de tableaux de contingence (toutes choses dont on peut critiquer l'abus; mais non bannir l'usage). Mais touchant à la statistique paramétrique, à l'analyse factorielle et à la classification automatique, les structures latentes intéressent la genèse de l'analyse des données.

Considérons un questionnaire Q dont chaque question q admet un ensemble J_q de modalités de réponse; la réponse d'un sujet i à Q sera un point $\rho(i)$ de l'ensemble produit $J_Q = \prod \{J_q | q \in Q\}$. Lazarsfeld postule que pour un sujet i donné, $\rho(i)$ est aléatoire (i.e. n'est pas déterminée totalement par l'attitude générale du sujet; mais dépend aussi de diverses causes contingentes); que les réponses rendues par i aux diverses questions q sont indépendantes les unes des autres; et que les lois de ces réponses dépendent d'un paramètre $\theta(i)$ paramètre latent (caché : non directement mesurable) qui caractérise l'attitude du sujet i . En formules la loi de $\rho(i)$ s'écrit donc comme une loi produit :

$$p_{JQ}^{\theta} = \prod \{p_{Jq}^{\theta} | q \in Q\}.$$

L'hypothèse d'indépendance est peu vraisemblable (en terme de statistique classique ceci revient à postuler que les corrélations partielles entre questions sont nulles sous la condition $\theta = \text{cte}$) : elle n'est prise que pour restreindre l'indétermination du modèle et permettre de confronter celui-ci au réel.

Le paramètre θ pourrait *a priori* varier dans un domaine multidimensionnel Θ : ici encore mieux vaut se restreindre et faire varier θ sur une droite ou un intervalle d'une droite. Reste à faire sur la dépendance de la loi p_{Jq}^{θ} par rapport à θ , une hypothèse paramétrique; bornons-nous au cas où la question q n'a que deux issues Oui et Non : il suffit d'exprimer la probabilité de répondre Oui par une fonction de θ : p_{q+}^{θ} (fonction que Lazarsfeld appelle *trace-line*, ce que nous traduirions par *profil*). Une première idée est de rejoindre la loi des jugements de comparaison (cf § 2.5.2) : on postule que p_{q+}^{θ} est égale à la probabilité que soit positive une variable aléatoire normale de variance $\sigma(q)$ et de moyenne $m(q) + \theta$: ainsi p_{q+}^{θ} croît avec θ (les sujets à $\theta = +\infty$ répondent Oui à toutes questions q de Q : sous réserve que l'unidimensionnalité soit acceptée, il faudra peut-être pour rejoindre ce modèle renverser certaines questions). On peut encore songer à faire de p_{q+}^{θ} une fonction linéaire de θ : dans ce cadre, la recherche des paramètres du modèle équivaut à un ajustement linéaire; et, les moindres carrés aidant, rejoint l'analyse factorielle. Enfin on peut postuler une fonction de trace très simple avec un seuil $\theta(q)$ et deux niveaux $p_0 | p_1$: si θ est inférieur à $\theta(q)$, $p_{q+}^{\theta} = p_0$; au delà du seuil ($\theta > \theta(q)$) : $p_{q+}^{\theta} = p_1$. Dans ce cas, comme il n'y a qu'un nombre fini de questions, le comportement d'un sujet est caractérisé par la position de son paramètre $\theta(i)$ par rapport à un nombre fini de seuils $\theta(q_1), \theta(q_2) \dots$.

Autrement dit les points $\theta(q)$ divisent l'axe de la variable latente θ en une suite d'intervalles sur lesquels le comportement est constant; ces intervalles successifs (en nombre fini répétons-le) définissent ce que P. Lazarsfeld appelle les *classes latentes*. La loi conjointe p_{JQ} des réponses $\rho(i)$ fournies par l'ensemble des sujets (répétons que $\rho(i) \in J_Q = \Pi\{j_q | q \in Q\}$) se trouve ainsi décomposée en une somme de tranches :

$$p_{JQ} = \sum \{p_c p_{JQ}^{\theta_c} | c \in C\},$$

formule où C désigne l'ensemble des classes latentes; θ_c une valeur de θ comprises dans l'intervalle correspondant de Θ ; p_c la probabilité qu'un individu appartienne à la classe c (ait son paramètre $\theta(i)$ dans l'intervalle afférent à cette classe). Or selon l'hypothèse fondamentale des structures latentes les $p_{JQ}^{\theta_c}$ sont des lois produits; mais il est clair que la loi globale p_{JQ} n'est pas une loi produit (déjà peu acceptable au niveau d'un seul sujet, ou des sujets d'une même classe; l'hypothèse d'indépendance entre les réponses à toutes les questions est absolument insoutenable sur l'ensemble de la population interrogée). On a donc un problème mathématique précis : décomposer (approximativement) une loi p_{JQ} en somme de lois produits. Et cette décomposition une fois faite permet de diviser en classes l'ensemble des sujets.

2.5.5. Une théorie des données :

Tel est le titre d'un intéressant ouvrage de C.H. Coombs : *A theory of Data* (J. Wiley ed.). Ecrit dans la ferveur de l'unité découverte entre tant de méthodes de traitement et d'interprétation des données le livre paru en 1964 se donne pour le complément de celui de Torgerson. Sans reprendre l'exposé détaillé des pratiques éprouvées de la psychophysique, ni même l'élaboration plus récente des questionnaires, il passe en revue des méthodes nouvelles et surtout propose une synthèse, une philosophie... De l'ouvrage de C.H. Coombs, nous retiendrons plusieurs aspects selon nous de très inégale valeur. Souvent nous reprenons les propres paroles de l'auteur, mais sans nous astreindre à le citer, afin de pouvoir condenser sa pensée.

a) *Généralité du point de vue géométrique* : Pour Coombs, la théorie des données se présente comme une analyse des fondements de la psychométrie (psychological measurement); elle suit le processus de mesure depuis son origine, l'étude du comportement jusqu'à son terme, la classification des stimuli et des individus. Le livre considère exclusivement les modèles géométriques - échelles unidimensionnelles ou cartes multidimensionnelles - qui servent à cette classification. De ce point de vue géométrique, on découvre la parenté qui unit des modèles divers communément présentés chacun dans un contexte expérimental spécifique : ainsi faire un diagnostic, juger des étudiants, estimer le poids d'un objet, c'est toujours mettre en relation deux ensembles d'éléments : malades avec syndromes; étudiants avec adjectifs; objets pesants avec nombres (poids estimés).

b) *Idéalisme* : Coombs se déclare persuadé que les données (data) sont en partie un produit de l'intellect de l'observateur : entre les observations que l'expérimentateur a choisi de recueillir sur un domaine du monde réel auquel il s'intéresse, et les données d'où une analyse mathématique appropriée tirera une typologie géométrique, s'interpose un choix théorique, et ce choix reste en partie arbitraire. Mesurer, définir une échelle, requiert au moins une théorie en miniature : même si

l'observation elle-même est un nombre - nombre de gouttes de salive, nombre de réponses exactes... - une autre interprétation de l'observation comme une donnée, demeure possible; même alors, il n'y a pas d'interprétation qui s'impose de toute nécessité. Reste valide le principe général suivant lequel toute connaissance résulte d'une théorie : l'information ne s'acquiert qu'au prix d'hypothèses (assumptions) : comme tout autre *fait*, données (data), mesures et échelles sont des *inférences*.

c) *Multipllicité des sortes de données* : Coombs reconnaît quatre sortes de données, et fonde sur cette distinction le plan de son ouvrage en quatre parties.

c₁) *Données de choix et de préférence* : à la base, un ensemble I d'individus et un ensemble J d'objets; les informations élémentaires recueillies ont toutes le format : l'individu i préfère l'objet j à l'objet j'. Mais la collecte des données peut varier : soit qu'on présente à l'individu i l'ensemble J tout entier qu'on lui demande d'ordonner selon ses préférences; soit qu'on lui propose seulement les objets par deux (voire par trois ou par quatre) pour qu'il formule un choix; procédure qui peut faire apparaître des cercles (i préfère j à j', j' à j" et j" à j!). Pour Coombs, ces données deviennent intelligibles par l'hypothèse qu'il existe un espace (joint space) dans lequel sujets et objets sont représentés par des points; le sujet i préférant l'objet j à l'objet j' si le point i est plus proche du point j que du point j'.

c₂) *données à un seul stimulus* : toujours un ensemble I d'individus et un ensemble J d'objets; seul change le format des informations élémentaires recueillies : le sujet i répond par approbation ou refus (à un degré variable) à la présentation des objets j pris un par un (d'où le terme de *single stimulus data*). Coombs se demande comment le modèle précédemment conçu pour rendre compte des choix entre deux stimulus peut être adapté aux réponses données à un stimulus isolé. Et il n'a pas de peine à suggérer qu'un sujet répond favorablement aux stimuli dont il est proche.

c₃) *Comparaisons entre stimulus* : il s'agit premièrement du problème fondamental de la psychophysique - et de son extension à la psychométrie : ranger un ensemble J d'objets suivant un certain critère (e.g. la clarté, ou la douceur) en une série ordonnée qui aux fluctuations près doit s'accorder avec les jugements individuels dont elle est la résultante : on sait (cf § 2.4.1) que l'unité métrique sur l'ensemble J est fournie par la notion complexe de seuil. En fait dès la psychophysique (ou mesure subjective des qualités dont l'étude physique est hors de conteste) le problème n'est pas exclusivement unidimensionnel : comme on le voit sur l'exemple des couleurs qui forment un continuum tridimensionnel (tonalité chromatique : rouge ≠ bleu; mais aussi saturation : rouge ≠ rose et intensité lumineuse : gris rose ≠ blanc rosé).

c₄) *données de similarité* : ici encore les variantes individuelles passent au second plan : l'information élémentaire se rapporte à deux paires d'objet (j, j') et (j'', j'''); et à la forme d'une inégalité : j est plus proche (ou encore moins proche) de j' que j'' ne l'est de j'''; ce qu'en formule on écrira $d(j, j') < d(j'', j''')$; mais la spécification d'une distance reste dans le vague; il convient de la préciser par une construction géométrique.

d) *Simplicité de l'outil mathématique* : beaucoup des méthodes recensées par Coombs procèdent par tâtonnement pour disposer un petit nombre de points de telle sorte que les inégalités entre distances correspondent à peu près aux informations recueillies. Coombs connaît la puissance de l'algorithme de diagonalisation des matrices et l'analyse en composante principale (diagonalisation d'une matrice symétrique conçue comme une matrice de corrélation, même si cette interprétation n'est pas immédiate) ne lui est pas étrangère. Cependant il écrit "Shepard a

bien une méthode (il s'agit de l'analyse des proximités cf § 3.4.1) mais celle-ci est impraticable sans ordinateur". Bien que publié en 1964, l'ouvrage de Coombs s'accommode fort bien des techniques de calcul antérieures à l'avènement des ordinateurs.

Notre prétention n'étant pas de communiquer la richesse de l'expérience psychométrique de Coombs, nous nous contenterons de critiquer ses positions méthodologiques que nous venons de caractériser très brièvement. La place faite à la géométrie, le rôle unificateur que lui attribue Coombs nous satisfait pleinement. Mais son idéalisme (sans être des plus exagérés : un expérimentateur se permet difficilement d'aller jusqu'à nier l'existence du monde extérieur, comme l'ont fait des penseurs auxquels nous refusons le titre de sage !) nous paraît arrêter à mi-chemin l'unification qu'il entreprend : là où subsiste un arbitraire lié à l'intellect de l'observateur la diversité des méthodes tend à dépasser ce que requiert le réel. Pour nous le passage de ce que Coombs appelle observation à ce qu'il appelle données relève plutôt du codage (cf § 3.6) que du choix d'un modèle théorique. D'ailleurs la place faite aux données de préférence nous semble très exagérée : les situations de choix sont souvent des situations forcées qui révèlent imparfaitement les principes (la vision spatiale) sous-jacents au comportement ordinaire des individus; et il est particulièrement fâcheux de traduire en *data* de préférences un tableau d'observation brut tel le tableau de contingence "qui-cite-qui" (nombre de fois que le journal *j* est cité dans le journal *j'*) ou la matrice de confusion entre signaux Morse de E. Rothkopf : il s'impose de soumettre telle quelle une matrice de confusion à l'analyse de correspondance; quant au tableau des "qui-cite-qui", il est préférable de le doubler; en mettant sur une ligne *j* (journal *j*) après les informations relatives aux journaux qu'il cite, celles (également contenues dans la colonne *j*) relatives aux journaux qui le citent; e.g. avec deux journaux *a*, *b*.

	a	b	a'	b'
a	10	5	10	3
b	3	7	5	7

(e.g. d'après la première ligne *a* cite *a* 10 fois et cite *b* 5 fois; *a* est cité par *a* 10 fois et est cité par *b* 3 fois). Il ne s'agit pas tant de concevoir le tableau des observations d'après un modèle; que de réduire un tableau de nombres par des opérations algébriques qui s'imposent. Mais ici l'idéalisme de Coombs n'est pas seul en cause; il n'est qu'une conception philosophique accommodée à la faiblesse des moyens mathématiques dont il use.

Ceci dit sans ambages, l'ouvrage de Coombs, écrit dans une période de transition, a le grand mérite à nos yeux de témoigner que des psychologues américains tout adonnés qu'ils soient à l'ajustement de modèles de toute sorte, aspirent à l'unité de la méthode inductive et font confiance à la géométrie pour exprimer cette unité.

2.5.6. Vers l'avènement des ordinateurs : Créer hors de la nature un être qui voit et qui pense, recréer la vie : ce projet orgueilleux égare l'homme qui sait seulement en imitant la Création, concevoir des méthodes et forger des outils. Mais il faut avouer que souvent seule la folie a vaincu notre paresse. La patiente statistique n'analyse les données qu'elle compile, que grâce à l'ordinateur, machine qu'on n'aurait sans doute pas construite, sans le rêve de faire un robot.

Ouvrons le recueil *Mathematical Thinking in the Social Sciences* (Free Press, 1954; et Russel & Russel, 1969) édité par P. Lazarsfeld : on y trouve à côté de l'exposé mathématique de ses modèles par Guttman (cf § 2.5.4) et d'une initiation aux structures latentes par Lazarsfeld (cf § 2.5.5), tel exposé d'introduction aux probabilités appliquées dont le ton nous rappelle qu'il y a moins d'un quart de siècle un sociologue

-fût-il américain - n'acceptait pas de voir s'enchaîner des formules mathématiques sans protester (cf Herbert Simon : op. laud. P. 388) que "le but d'une langue est de dire quelque chose et non simplement de parler de la langue elle-même". Mais brillent aussi les ambitieuses inventions de Rashevski qui ayant imité par des systèmes d'équations différentielles les réseaux nerveux, se propose aussi de représenter par les mêmes calculs les phénomènes de conduction au sein du tissu social. Et H. Simon, après avoir noté que le principe mathématique classique de la recherche d'un optimum ne fournit pas de base complètement rationnelle aux décisions de plusieurs hommes en compétition, reconnaît volontiers que de telles décisions sont imparfaitement rationnelles; mais il doit citer le principe du *minimax* par lequel Von Neumann et Morgenstern ont poussé le plus loin l'explication rationnelle des choix en attribuant aux hommes une habileté dont on trouve peu d'exemples en dehors des joueurs de Poker les plus déliés ("at the cost of attributing to human beings a cleverness they have perhaps not often exhibited outside the more successful poker-playing circles"; *ibid* p. 396).

Dans les années 50 en effet, cependant qu'à partir des machines prototypes à la conception desquelles participa Von Neumann, se développent les ordinateurs aujourd'hui communément en usage, l'idée cybernétique (issue de N. Wiener) d'explication et d'imitation de la vie selon des équations différentielles régissant des processus aléatoires inspire plus d'un savant - des neurologues aux électroniciens en passant par les sociologues. Dans cette direction, le projet de la reconnaissance automatique des formes intéresse de près l'analyse des données.

Au départ (≈ 1955), on espérait que la machine à percevoir (à lire l'écriture manuscrite; à reconnaître les paroles et les visages ...) fonctionnerait bientôt. Le *Perceptron* de Rosenblatt est une machine qui reçoit sur une couche de cellules sensibles ou rétine d'entrée, des images simples appartenant à plusieurs classes (e.g. représentant deux lettres différentes); et grâce à une organisation très ingénieuse modifiant automatiquement ses connexions internes dans un *apprentissage* au cours duquel lui sont présentées des images avec l'indication (e.g. numérique) de leur classe, parvient à fournir des réponses exactes. Ce succès limité à des cas parfois spectaculaires mais non pratiques, n'intéresse plus aujourd'hui : un recueil intitulé *Methodologies of Pattern recognition* (Actes du symposium de Honolulu 1968; édité par S. Watanabe; Acad. Press. 1969) cite 14 fois le nom de Rosenblatt : mais le mot de perceptron n'est pas à l'indice systématique ! On convient généralement aujourd'hui que l'objectif premier n'est pas de fabriquer des machines spécialisées, mais d'étudier en toute généralité le problème suivant : étant donné un ensemble I d'individus, répartis en classes $I_1, \dots, I_q, \dots, I_k$ et chacun décrit par un vecteur de R^n ; trouver une règle simple pour décider d'après sa description de l'affectation d'un individu nouveau à l'une des n classes. C'est ce qu'en statistique on appelle analyse discriminante (cf § 2.3.5). Mais il faut ici préciser plusieurs points. Le problème est insoluble si la description des individus n'en comporte pas les traits pertinents : d'où nécessité préalable du codage familier à l'analyse des données (cf § 3.7). Les distributions dans R^n des classes réelles n'ont pas la forme qu'imagine la statistique paramétrique fêrue de lois normales multidimensionnelles : mieux vaut donc accepter l'échantillon des données tel quel comme un substitut des distributions probabilistes (méthode dite Néobayésienne cf § 2.3.3). Les algorithmes de construction de cloison considérés depuis Fisher sont le plus utiles après que par analyse factorielle la dimension du vecteur de description a été réduite. Enfin souvent il n'y a pas de classification *a priori* valide : le statisticien doit découper en classes un nuage unique I .

C'est ce qu'en reconnaissance des formes on appelle "unsupervised learning" apprentissage sans maître (qui donne le n° de la classe des

individus présentés; rappelons-nous le Perceptron); et en statistique classification automatique. Discipline dont les fondements ont été posés par les écologistes attelés dès le début du XX^e siècle à de grands tableaux de données (cf TI B n° 1 & 2; C n° 2, 3, 3'; et le traité classique de Sokal et Sneath : *Principles of Numerical Taxonomy*; Freeman; 1963).

Nous voici parvenu dans notre historique à l'avènement des ordinateurs. Impossible de faire un inventaire même rapide de ce que le troisième quart du XX^e siècle a produit, qui touche à l'analyse des données ! Certes tout ce que la statistique avait calculé ou projeté de calculer avant 1950 se calcule désormais sur ordinateur par des programmes ordinaires de bibliothèque. Mais cherchant à faire le bilan de la rencontre entre Statistique et calcul électronique, D. Wallace écrit avec sagesse mais non sans mélancolie (cf *Computers in the teaching of Statistics : Where are the main effects*; in *Statistical Computation* ed. by K.C. Milton & J.A. Nelder; Acad. Press; 1969) : "Computers in statistics have had a slow and uneven course" ... Les ordinateurs ont traversé la statistique comme un fleuve au cours lent et irrégulier; intérêt certain et multiforme ...; mais rôle presque inexistant des statisticiens dans le progrès de l'analyse numérique; ce qui contraste avec les traditions de calcul propres à la statistique... et il conclut en substance "For better or worse, statistical practice is being revolutionized by computers". Pour le meilleur ou pour le pire, les ordinateurs mettent en révolution la statistique... Avec ou sans nous, les praticiens des sciences sociales et les naturalistes parviendront à analyser leurs données ... Si notre enseignement ne porte pas sur ces problèmes, la statistique universitaire s'atrophiera en une branche maussade et inélégante des mathématiques. Je ressens profondément que la révolution des ordinateurs nous propose une tâche urgente et exaltante ...

Venu à la statistique après avoir appris la géométrie, pour faire au sein de l'ordinateur la synthèse de données linguistiques (§ 3.2), nous nous permettrons de laisser à D. Wallace ce dernier mot dans notre chronique des écoles statistiques anglo-saxonnes.

Appendice : La reconnaissance des formes et le perceptron

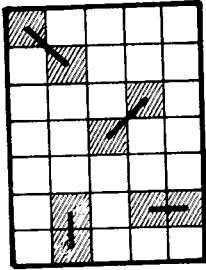


Figure 2-10
Exemples d'éléments
de trait diversement
orientés sur une ré-
tine de 7x5 cellules.

Soit par exemple le domaine optique. On suppose ordinairement les objets reçus sur une rétine, ou grille rectangulaire dont la maille n'est guère fine : par exemple $7 \times 5 = 35$ cellules. Chaque cellule étant désignée par un indice j qui est un couple d'entiers (le numéro de la ligne avec celui de la colonne) un objet ou stimulus visuel i sera décrit par un vecteur $k(i, j)$, l'intensité lumineuse apportée par l'image i sur la cellule j . Mais ces informations primaires décrivent mal une forme : il est déjà meilleur de chercher des traits, (ou suite de deux ou trois cellules contiguës intensément stimulées), orientés soit suivant les axes de la rétine soit suivant leurs bissectrices (cf Fig. 2-10).

On a espéré naguère, par un automate comprenant outre la rétine d'entrée une ou plusieurs couches de cellules internes reliées par des liens plus ou moins intenses, élaborer une réponse manifestée par l'excitation de telle ou telle cellule de sortie (e.g. l'allumage de l'une des lampes d'une série). L'appareil à chaque présentation d'un objet aurait reçu d'abord l'indication de la réponse, d'où renforcement des liens entre la sortie de celle-ci et les diverses cellules principalement excitées par celui-là (cf Fig. 2-11). Cette phase d'apprentissage terminée, l'appareil aurait été capable de répondre par lui-même. L'analogie entre cette structure en couches et les aires de projections du système nerveux est manifeste ; non moins que les conceptions psychologiques qui inspirent les règles d'apprentissage de l'automate. Le système, appelé *Perceptron* par son inventeur *Rosenblatt*, a connu d'abord des succès assez spectaculaires : mais on n'espère plus aujourd'hui qu'il puisse être utile.

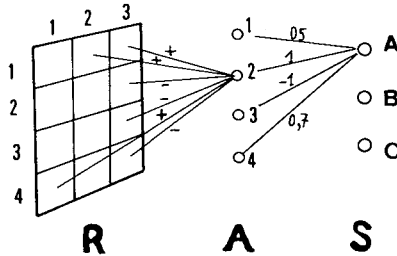


Figure 2-11 : Schéma d'un perceptron : entre les cellules d'entrée de la rétine (R) et les cellules de sortie du niveau (S), s'interpose la couche d'association A. Chaque unité d'association reçoit la somme des excitations de certaines cellules d'entrée affectées de signes choisis au hasard ; si la somme ainsi reçue est supérieure à un seuil, l'unité d'association est activée ; les unités d'association étant très nombreuses, leur activité décèle en quelque sorte la présence de tous les traits ou autres combinaisons graphiques que peut porter la rétine. Les influx émanant des cellules de la couche (A) sont combinés linéairement avec des coefficients - appelés forces des liens - pour produire une réponse au niveau (S) ; seuls les liens entre (A) et (S) sont modifiés en cours d'apprentissage.