

Contains
RC 33 Newsletter

BMS

BULLETIN DE METHODOLOGIE SOCIOLOGIQUE

<i>RC33 Newsletter</i>	President's Report, Secretary's Report RC33 World Congress Sessions
Klaus G. Troitzsch	Mathematical Modelling and Computer Simulation of Social Processes: Problems and a New Solution
Philippe Cibois	Le PEM, pourcentage de l'écart maximum: un indice de liaison entre modalités d'un tableau de contingence
Saadi Lahlou	A Method for Measuring Network Effects in Scientific Cooperation

Centre. Spain, "Research Techniques" by Francisco Alvira Martín

Books/Livres. *Annual Review of Sociology, Professional Evaluation, Dictionary of Statistics, Multivariate Analysis, Scientométrie*, 48 Divers English, 7 Divers français

Brochures/Reviews. *ZUMA Nachrichten, ZA-Information, Measurement Models, Multilevel Models, Interviewers & Nonresponse, La Lettre du LASMAS*

Articles. Observation in French Sociology, Correspondence & Canonical Analysis, Time Sampling, Open-Ended Questions, Social Attitudes, Qualitative & Quantitative Data, 7 Divers

Computers/Ordinateurs. Les grands réseaux: "If Your Friends Won't Tell You, Who Will?", by Alden S. Klövdahl

Meetings/Réunions. Artificial Intelligence, Classification Societies, Analyse des données textuelles, Databases, New Europe, Social Structures, Categorical Data, 23 Divers

Past Meetings/Réunions passées. Sunbelt Network Conference, European Network Conference, American Sociological Association Meeting, 5 Divers

Calls. *Family Issues, Fondation Fyssen, IOPS Dissertation Training, Autonomie, Soutenabilité, Uncertainty, Vienna Dialogue, Society & Resource Management*, 9 Divers

N. 40 SEPTEMBER 1993

Bulletin de

LE PEM, POURCENTAGE DE L'ECART MAXIMUM : UN INDICE DE LIAISON ENTRE MODALITES D'UN TABLEAU DE CONTINGENCE

par

Ph.Cibois

(Université Paris V - René Descartes)

Résumé : On définit un indice de liaison entre modalités d'un tableau de contingence, le PEM ou Pourcentage de l'écart maximum. On montre deux exemples réels d'application de ce PEM : il permet de construire des profils, c'est à dire l'ensemble des modalités de réponse d'une enquête qui sont en attraction avec une modalité donnée. Il permet de construire également des styles de comportement, modalités qui ont un ensemble de PEM semblables quand on les croise avec des variables descriptives. PEM, profils, styles, Tri-deux.

Abstract : The PEM, percentage from maximum deviation : an index of tie between modalities of a contingency table. One define an index of tie between modalities of a contingency table, the PEM , percentage of maximum deviation. Two real exemples are shown : PEM permits to construct profiles, set of modalities of response from a survey which are in attraction with a modality. PEM permits also to build styles of comportement, modalities which have a set of PEM similar in a cross-tabulating with descriptive variables. PEM, profiles, styles, Tri-deux.

Si pour un tableau de contingence il existe plusieurs indices qui permettent de mesurer la dépendance entre lignes et colonnes (coefficients de Cramér, de Pearson, de Tschuprow), ce n'est que depuis peu de temps que la force de la liaison entre modalités de lignes et de colonnes est apparue au chercheurs comme méritant la création d'un indice spécifique.

On doit mettre ce fait en relation avec une attitude plus générale vis-à-vis des tableaux de contingence dont on peut dire globalement qu'ils sont examinés dans leur ensemble dans la littérature anglo-saxonne et qu'ils n'ont été pris en compte dans leur structure interne, au niveau de la structure des correspondances entre modalités que dans la perspective des méthodes d'Analyse des correspondances.

C'est dans cette perspective que l'on propose ici un indice de liaison entre modalités d'un tableau de contingence, le PEM, pourcentage de l'écart maximum. Avant de le définir on examinera un indice qui, sous divers noms, veut avoir cette fonction. On en donnera ensuite des applications qui permettront d'en juger l'intérêt.

LE RAPPORT ECART/THEORIQUE

Il consiste à calculer, pour une case donnée d'un tableau de contingence le rapport entre l'écart à l'indépendance (effectif observé - effectif théorique) et l'effectif théorique qu'il y aurait en cas d'indépendance.

Cet indice a été présenté une première fois par J.Bertin (1967 : 225) sous le nom de rapport ou de pourcentage de distorsion, puis repris sous le nom de "khi" (Bertin 1977 : 65).

Rouanet (1987 : 158) reprend cette indice sous le nom de "taux de liaison".

Ce rapport a les qualités suivantes :

- il est positif quand l'observé est plus grand que le théorique ce qui correspond à une attraction entre lignes et colonnes,
- il est nul quand l'observé est identique au théorique, c'est à dire quand on est dans la situation d'indépendance,
- il est négatif enfin quand l'observé est inférieur au théorique ce qui correspond à une situation de répulsion entre lignes et colonnes.

De ce fait le signe de l'indice est tout à fait interprétable d'un point de vue sociologique ainsi que sa nullité. On retrouve ici les propriétés de l'écart à l'indépendance.

- de plus l'indice a la bonne propriété mathématique d'être moyennable en cas de regroupement de cases.

On peut cependant critiquer cet indice pour deux raisons :

- il n'est pas borné ce qui ne nous permet pas de dire si une liaison est faible ou forte,
- il signifie autre chose que l'intensité de la liaison : il exprime d'abord le rapport entre l'écart à l'indépendance et l'effectif théorique.

Comme l'avait bien vu Bertin en le nommant Khi, il s'agit d'un coefficient de pondération que l'on multiplie par l'écart à l'indépendance et qui nous donne la contribution au Khi-deux. Si l'écart à l'indépendance est faible par rapport au théorique, cela signifie que la case apporte peu d'information et la contribution résultante au Khi-deux sera faible parce qu'inférieure à l'écart à l'indépendance. Si l'écart à l'indépendance est égal au théorique, l'indice sera égal à l'unité et la contribution au Khi-deux sera identique à l'écart à l'indépendance. Enfin, si l'indice est supérieur à un, cela signifie que l'écart à l'indépendance est plus grand que l'effectif théorique et le coefficient de pondération fera que la contribution au Khi-deux sera plus forte que l'écart à l'indépendance ce qui manifeste que cette case apporte beaucoup d'information. De ce fait, on peut considérer cet indice comme plus précisément qu'un indice de liaison, un indice de surprise, d'information apportée par une case.

LOGIQUE DU PEM

La logique de l'indice que nous présentons est dérivée du coefficient que Cramér a mis au point pour juger de l'intensité de la dépendance dans un tableau de contingence dans son entier. En effet celui-ci indique (Cramér 1946 : 282) que s'il y a dépendance parfaite d'une variable par rapport à l'autre, le Phi-deux sera alors égal à la plus petite dimension du tableau diminuée d'une unité. Disposant ainsi du maximum possible de la liaison, Cramér construit son coefficient en faisant le rapport du Phi-deux observé au Phi-deux maximum. Cet indice varie évidemment de zéro à un en allant de l'indépendance à la dépendance.

Si l'on cherche à appliquer la même logique dans le cas d'une case isolée d'un tableau de contingence, il est facile de trouver le minimum de la liaison puisqu'il correspond à l'indépendance, mais la question se pose pour le maximum : à quoi correspond pour une case la liaison maximum ?

Une réponse possible à cette question consiste à mettre dans cette case l'effectif le plus fort possible qui soit compatible avec les marges du tableau et à calculer l'écart à l'indépendance maximum qu'il y aurait dans ce cas. La force de la liaison sera simplement calculée en examinant le rapport entre l'écart à l'indépendance observé et l'écart maximum à l'indépendance calculé précédemment. Ce rapport étant mis en pourcentage, on parlera donc de Pourcentage de l'Ecart Maximum ou PEM.

PREMIER EXEMPLE : DESTINATION DES BACHELIERS 1975

A titre d'exemple, examinons le tableau suivant indiquant d'une manière simplifiée ce que sont devenus les bacheliers d'une année en fonction de la série du baccalauréat qu'ils avaient passé (bac littéraire section A, sciences BDD', mathématiques CE, technique FGH) selon qu'ils sont entrés à l'Université, dans des classes préparatoires aux grandes écoles ou dans d'autres orientations à finalité professionnelle plus immédiate.

	Université	Classes prep.	Autres	Total
Lettres	13	2	5	20
Sciences	20	2	8	30
Mathématiques	10	5	5	20
Technique	7	1	22	30

Total	50	10	40	100

Données simplifiées issues de : Note d'informamtion, 76-35 du 15 octobre 1976, Service des Etudes informatiques et statistiques du ministère de l'Education.

Pour la case série mathématique du bac allant dans les classes préparatoires aux grandes écoles, on a les résultats suivants :

effectif observé : 5

effectif théorique : $20 \times 10 / 100 = 2$

écart à l'indépendance : 3

taux de liaison (Rouanet) : $3 / 2 = 1,5$

effectif maximum compatible avec les marges : 10 (minimum des deux marges 10 et 20)

écart maximum à l'indépendance : $10 - 2 = 8$

Pourcentage de l'écart maximum : $3 / 8 \times 100 = 37,5\%$

D'une manière analogue, calculons les écarts à l'indépendance, les effectifs théoriques, le taux de liaison, la contribution au khi-deux et le PEM pour toutes les cases en attraction :

	écart à l'ind.	effectif théo.	taux de liaison	khi-deux	PEM
Math. - classes prép.	3	2	1,50	4,50	37,5%
Lettres - Université	3	10	0,30	0,90	30,0%
Sciences - Université	5	15	0,33	1,67	33,3%
Technique- autres	10	12	0,83	8,33	55,6%

Les deux premières cases ont même écart à l'indépendance mais ces mêmes écarts ne représentent pas la même chose dans la mesure où les séries mathématiques et la destination classes préparatoires sont les marges les plus faibles ce qui entraîne donc une attente d'un théorique faible, un rapport au théorique plus grand que un et donc un khi-deux plus élevé que l'écart. Par contre dans la deuxième case (Lettres - Université) un même écart de 3 en rapport avec une attente forte, car ces marges sont les plus fortes, entraîne un rapport écart/théorique inférieur à l'unité et un khi-deux plus faible que l'écart. Les "taux de liaison" nous indiquent simplement que le même écart apporte beaucoup d'information dans la première case et peu dans la seconde. Quant au PEM, il est du même ordre dans les deux cas et nous signifie que les deux liaisons sont à peu près à un tiers de leur maximum.

Si l'on oppose par contre les deux plus forts taux de liaison (Math. - Classes prep. = 1,5 et Techniques - autres orientations = 0,83) on voit que le PEM inverse l'ordre car l'attente est faible dans la première case et forte dans la deuxième alors que l'on est beaucoup plus proche de la liaison maximum dans la deuxième que dans la première.

On trouvera en annexe une formalisation complète de la recherche du PEM, y compris pour les écarts négatifs où le minimum ne peut toujours être zéro du fait de certaines contraintes de marges. On y montrera aussi quelle généralisation on peut en faire pour mettre au point un indice de liaison sur l'ensemble du tableau si on peut mettre à jour une structure d'ordre sur les lignes et les colonnes.

APPLICATION AUX "PROFILS"

On constate qu'il existe une pratique du sociologue qui est de regarder dans plusieurs tableaux croisés le comportement d'une modalité qui l'intéresse et ainsi d'arriver à en rendre compte, à l'expliquer par ses liaisons avec d'autres modalités. Cette pratique conduit en général à ce genre littéraire hautement soporifique qu'est le commentaire de tableaux croisés où une modalité est ainsi comparée de tableaux en tableaux.

Pour améliorer cette pratique, on utilise le PEM mais plusieurs décisions doivent être prises :

- définir l'univers de référence : quels sont les autres modalités de l'enquête que l'on souhaite voir être associées avec la modalité dont on veut rendre compte. En étant large, on peut fixer à environ 200 le nombre de modalités soit de comportement, soit de statut, que l'on désire croiser avec une modalité ;

- définir la fiabilité de la liaison observée : une liaison forte peut être due à des effectifs non significatifs. Pour avoir une idée de cette fiabilité, on calcule un Khi-deux (muni de la correction de Yates) et l'on indique si l'on atteint les seuils standards de significativité (10%, 5%, 1%).

Nous voudrions montrer un exemple d'utilisation (Singly 1993) de cette technique en comparant, à partir des données d'un panel d'élève, deux populations scolaires qui présentent un certain nombre de points communs puisqu'il s'agit d'élèves de faible niveau, mais qui n'ont pas la même attitude vis-à-vis de la lecture. En effet, comme on connaissait depuis longtemps la liaison qui existait entre le niveau scolaire et la lecture, il a paru intéressant de "décristalliser" cette liaison et de regarder les bons élèves qui, à l'encontre de ce qui est habituellement attendu, lisaient peu, ou inversement les élèves de faible niveau qui avaient une attitude plutôt favorable vis-à-vis de la lecture.

C'est en faisant cette exploration que nous avons vu que parmi les élèves qui étaient tout à fait d'accord pour dire que la lecture était classée comme leur loisir favori, se trouvaient des élèves d'origine étrangère défavorisée (originaires du maghreb, de turquie ou d'afrique). Pour expliquer cette liaison qui semblait surprenante, nous avons entrepris une comparaison systématique entre cette catégorie et une catégorie équivalente de Français, celle des élèves français de niveau scolaire équivalent à celui des étrangers. Du Panel DEP du Ministère de l'éducation qui suit des élèves depuis 1989 à partir de la 6e, on tire donc deux populations comparables par leur faible niveau scolaire mais qui les uns sont français d'origine et les autres sont étrangers.

La confection d'un "profil" consiste à croiser une modalité donnée (étrangers défavorisés puis français défavorisés), avec toutes les autres modalités d'un univers de questions qui englobe tous les domaines de l'enquête : les pratiques éducatives, les goûts, la relation à la télévision, à la lecture et aux loisirs, les occupations des jours de congé ainsi que toutes les variables de la vie scolaire concernant le cursus passé, le cursus en cours ainsi que la réussite.

Pour chaque croisement de la modalité dont on fait le profil avec toutes les autres modalités de l'univers de référence, on donne :

- le PEM centré sur 100 (PEM + 100) où 0 correspond à la liaison négative maximum, 100 à l'indépendance, et 200 à la liaison positive maximum,
- le Khi-deux associé à l'écart à l'indépendance de la case (avec la correction de continuité si un effectif théorique est inférieur à 5 et ramené à zéro si l'écart à l'indépendance est inférieur à 0.5),
- la mention de significativité du test (qui correspond toujours à un tableau 2 x 2 à 1 degré de liberté où l'on trouve la case observée et pour les autres cases les complémentaires aux marges). Il y a 3 options : "ns" pour non significatif, "*" pour significatif à .1, "***" pour significatif à .05 et "****" pour significatif à .01.

Etudions en premier lieu le profil des étrangers défavorisés :

Nom PEM Khi2 Seuil Intitulé

Q72 126 15.48 * un retard, c'est grave**

Il y a une liaison positive, elle représente 26% de la liaison maximum et elle est significative. Un retard en classe est considéré comme quelque chose de grave. Ceci est un indice d'une "bonne volonté scolaire" confirmée par la liaison avec l'intérêt fort pour le français, bien que le niveau apprécié par l'enseignant soit jugé faible et que le retard scolaire apparaisse déjà :

Q9D1 123 12.72 * intérêt français beaucoup**

NIV0 122 13.86 * appréciation : niveau faible**

AG64 116 23.02 * âge entrée 6e : 13 ans**

CL23 109 14.79 * ni 5e ni 6e normale en 2e année du panel**

CL32 121 15.39 * 5e en 3e année du panel**

Les parents, ouvriers, non diplômés, ne parlant pas français à la maison et ne sont pas à même d'intervenir dans la scolarité de leur enfant :

30 2 113 4.76 ** vérification par les parents : jamais

CS 6 147 56.79 * Catégorie socio-professionnelle : ouvriers**

NBE5 150 273.35 * 5 enfants et plus**

52P1 124 24.57 * père aucun diplôme**

52M1 113 4.66 ** mère aucun diplôme

A811 146 191.41 * on ne parle pas français à la maison**

A732 127 31.33 * l'enfant n'a pas une chambre**

La télévision joue un grand rôle :

36A1 109 4.15 ** télé le matin avant cours

36C1 121 8.00 * télé après les cours**

mais pas le sport :

50D0 136 8.56 * mercredi sport rarement ou jamais**

mais ce qui est plus surprenant, c'est la vision de la lecture où l'on retrouve le message de l'école, autre signe de bonne volonté scolaire :

38C1 113 9.85 *** aime encyclopédies beaucoup
39D1 122 5.97 ** lire c'est amusant : tout à fait d'accord
39E1 109 3.07 * lecture loisir favori : " " " "

De ce fait la bibliothèque est un lieu fréquenté assidument :

USB2 116 24.39 *** bibliothèque le mercredi : toujours ou souvent

A travers ce profil on devine le rôle positif en matière d'intégration que peut jouer l'école par le biais de la lecture et de la fréquentation de la bibliothèque, du fait de cette écoute attentive du message donné par l'école, message bien reçu, sans même la pression des parents et alors même que le niveau scolaire en français comme ailleurs est faible.

Il n'en est pas de même maintenant pour le profil des français défavorisés qui ont en commun d'avoir un faible niveau scolaire mais qui ont des attitudes toutes différentes :

Le niveau est faible :

NIV0 171 410.18 *** appréciation du niveau scolaire : faible
CL13 136 39.81 *** redouble 6e en 89

Le milieu est le même que précédemment :

CS 6 110 9.10 *** CSP ouvriers
52P1 110 5.30 ** père aucun diplôme

mais inversement les parents surveillent les horaires et les devoirs :

22 1 111 12.55 *** aide parents toujours ou souvent
30 1 109 4.04 ** vérification des parents tous les jours
17 1 119 7.21 *** parents décident travail
20 2 104 1.54 ns se couche avant 21h30

mais l'attitude de l'enfant vis-à-vis de l'école est beaucoup moins respectueuse des "valeurs" scolaires. Tant en ce qui concerne l'attitude générale :

Q4 1 112 11.22 *** oubli livre souvent ou parfois

qu'en ce qui concerne le français et la lecture :

Q9D4 129 12.28 *** intérêt français : pas du tout
32C0 118 35.46 *** pas inscrit bibliothèque

38B3	128	31.61	***	aime lire livre non
38C3	105	4.11	**	aime encyclopédies non
39A1	126	12.55	***	lire difficile tout à fait d'accord
39B1	129	14.89	***	lit pour étudier " " " " "
39D4	133	13.93	***	lire amusant pas du tout d'accord
40 2	125	14.40	***	n'a pas lu un livre depuis 3 mois
USB0	118	36.52	***	bibliothèque rarement ou jamais

Il n'y a pas de pratique de la lecture ou plutôt elle est associée à la pratique scolaire qui est surveillée par les parents.

Il y a vraiment une différence entre les deux profils sur la lecture entre les étrangers défavorisés qui "écoutent" le message véhiculé par l'école et y croient tandis que les français défavorisés ne l'entendent pas : seuls leurs parents croient à la réussite scolaire alors que les parents étrangers ne s'en soucient guère. A l'inverse ce sont les enfants étrangers qui considèrent que c'est grave d'arriver en retard alors que les élèves français oublient souvent leurs livres.

Les enfants étrangers croient à ce qu'on leur dit à l'école et passent à l'acte en ce qui concerne les recommandations de lecture alors que les français n'entendent rien de ce qui leur est dit, si ce n'est par le biais de la pression parentale. Pour les enfants français défavorisés, l'école est contre-productive en matière de lecture alors qu'elle est efficace pour les enfants équivalents étrangers.

On voit sur cette exemple que ce qui nécessitait un tableau ne nécessite plus qu'une ligne où l'on voit tout de suite s'il y a ou non une liaison, dans quel sens, et si cette liaison est significative. On peut ainsi déterminer un profil de chaque modalité qui est en quelque sorte "expliquée" par les liens qu'elle a avec les autres modalités de l'enquête. Cette explication n'est pas plus circulaire que dans un dictionnaire où un mot est expliqué par d'autres mots du dictionnaire. C'est le réseau sémantique qui est signifiant.

La fécondité de cette technique des profils vient de la facilité de sa mise en oeuvre logicielle (cf plus loin les problèmes d'implémentation) : elle permet cette exploration systématique que le sociologue a toujours envie de faire mais qui lui était rendue difficile par la masse des tableaux générés. Désormais, pour toute modalité, il est possible de faire une exploration systématique, sans a priori, qui se révèle rapidement fructueuse ou non.

ELABORATION DE STYLES DE COMPORTEMENT

La notion de "style de comportement" fait évidemment référence à la notion de style de vie, c'est à dire non pas à ces regroupements mal définis qui ont été popularisés sous des termes psychosociologiques (les "recentrés", les "décalés"), et qui ont été critiqués fort justement (Herpin 1986, Amaturio 1990), mais au concept de regroupement de comportements qui forment un tout intelligible, idéal-typique (Cibois 1983).

Pour arriver à cette notion de style de comportement nous partirons de la notion de "marquage" en disant qu'un comportement est souvent associé à des variables socio-démographiques. D'ailleurs, cette notion est déjà présente dans le sens commun : pour prendre l'exemple des émissions de télévision, il est tout à fait admis qu'il existe des émissions qui s'adressent à un public cultivé (comme *Apostrophes*) ou des émissions "populaires" (comme *Avis de recherche*). Cependant, le sens commun devient moins précis quand il parle d'émission "grand public" (comme *Grand échiquier*) : il désigne par ce terme une émission de qualité (culturelle) mais pas trop ésotérique pour pouvoir être appréciée par un vaste public. Le relais d'une définition rigoureuse s'impose.

Notre propos est donc de partir du sens commun qui "marque" des comportements sociaux comme étant en relation avec des groupes sociaux (le tricot est pour les femmes, le bricolage pour les hommes, *Océaniques* pour les intellectuels etc..), de voir comment se fait cette opération de marquage, quelle est la rationalité de cette sociologie spontanée et ensuite de la formaliser et de la généraliser.

On objectera à cette tentative qu'elle n'est pas en rupture avec le sens commun, qu'elle n'accomplit pas cette coupure épistémologique qui est le prérequis de la méthode sociologique. A cela nous répondrons que :

- la critique des prénotions n'est souvent que l'imposition des prénotions d'un autre groupe social, celui des intellectuels : s'ils ont le pouvoir légitime de distinguer le noble et le commun, le cultivé de l'ignorance, cette légitimité sociale réelle est une réalité à expliquer, non une source d'explication sociologique (Cf. Cibois 1989);

- la coupure épistémologique requise est bien présente quand, partant des notions du sens commun, on les formalise et on les généralise, donnant ainsi une rigueur formelle à des notions qui en sont dépourvues et les transformant ainsi en outils scientifiques.

Pour formaliser la pratique du sens commun qui "marque" socialement des comportements, nous utiliserons une analyse secondaire (Cibois 1992) d'une vaste enquête sur les pratiques culturelles des français (Donnat 1990) dont nous étudierons les données concernant les émissions de télévision préférées en 1989 et d'abord celles concernant les deux émissions que sont *Apostrophes* et *Avis de recherche*.

Que signifie qu'une émission soit jugée comme destinée à un public cultivé ou à un public populaire ? Si ces jugements font partie du sens commun, c'est que d'une certaine manière les publics correspondent bien aux jugements. On saura si c'est effectivement le cas en examinant en premier lieu le croisement entre la modalité "Bac et plus" et le fait d'affirmer qu'*Apostrophes* fait partie de ses émissions préférées :

	Apostrophe : émission préférée	non préférée	Total
Bac et plus	375	696	1071
autre	500	3426	3926
Total	875	4122	4997

Si cette émission est bien associée à un public cultivé, cela vient du fait que cette émission est préférée en moyenne par 17,5% des interrogés (875/4997) mais par 35,0% des "Bac et plus" (375/1071) : il y a donc surreprésentation de cette catégorie. C'est cette attraction qui est intériorisée dans les jugements qui associent à l'émission le qualificatif d'"intellectuel".

Si cette catégorie était neutre par rapport à l'émission, elle représenterait comme la moyenne 17,5% de la catégorie, c'est à dire 187,5 individus : c'est l'effectif qui correspond à une situation d'indépendance. L'écart à l'indépendance est de :

$$375 - 187,5 = 187,5$$

Inversement si, situation extrême, tous ceux qui préfèrent Apostrophes avaient le Bac ou plus, on aurait la distribution suivante qui nous sert de limite maximum de la liaison entre les deux modalités :

	Apostrophe : émission préférée	non préférée	Total
Bac et plus	875	196	1071
autre	0	3926	3926
Total	875	4122	4997

Avec cette distribution l'écart à l'indépendance serait cette fois de :

$$875 - 187,5 = 687,5.$$

On peut situer ainsi l'écart observé 187,5 entre un minimum nul qui correspondrait à la situation d'indépendance et un maximum qui correspondrait à la liaison maximum et qui serait de 687,5. On voit que l'observation correspond à une partie seulement du maximum et exactement à 27,3% (187,5/687,5).

Ce rapport de l'écart observé par rapport à l'écart maximum (PEM : pourcentage de l'écart maximum), nous permet de situer tout écart entre son minimum zéro et son maximum égal à cent.

Pour chaque modalité de pratique et pour chaque modalité de situation sociale, il est possible de faire la même démarche : on constate que selon les cas il y a écart positif, nul ou négatif à l'indépendance. Pour juger d'un écart négatif, prenons par exemple le croisement de la même modalité "Bac et plus" mais cette fois avec Avis de recherche :

Avis de recherche :	émission préférée	non préférée	Total
Bac et plus	59	1012	1071
autre	991	2935	3926
Total	1050	3947	4997

L'effectif théorique dans le cas de l'indépendance est pour la case "Bac et plus" - "Avis de recherche émission préférée" de:

$$1050 \times 1071 / 4997 = 225,0$$

et l'écart à l'indépendance de

$$59 - 225,0 = -166,0$$

Comme l'écart est négatif, la situation limite serait celle où aucun "Bac et plus" ne préférerait cette émission. On aurait la situation suivante :

Avis de recherche :	émission préférée	non préférée	Total
Bac et plus	0	1071	1071
autre	1050	2876	3926
Total	1050	3947	4997

Dans ce cas, l'écart maximum (dans le sens négatif) est égal à l'effectif théorique : $0 - 225,0 = -225,0$ et l'écart observé représente 73,8% de l'écart maximum ($-166,0/-225,0$).

On note ce pourcentage avec un signe négatif pour signaler qu'il s'agit d'une opposition et non d'une liaison.

Pour pouvoir comparer plus facilement pourcentages négatifs et positifs, on centre les pourcentages autour de 100 en ajoutant cette valeur à tous les PEM.

Le PEM avec signe de +27,3 aura la valeur centrée de 127

Le PEM avec signe de -73,8 aura la valeur centrée de 26

Une valeur de 100 correspond donc à l'indépendance : dans le cas où la variable est dichotomique (comme pour le sexe par exemple) à une valeur centrée de 110 pour une modalité (PEM avec signe de +10) correspondra un PEM centré de 90 (PEM avec signe de -10) pour l'autre (Cf. annexe).

Le fait qu'Apostrophes soit une émission perçue comme s'adressant à un public instruit est à mettre en relation avec le fait que son PEM pour "Bac et plus" est de 127. De même pour Avis de recherche où le PEM pour "Aucun diplôme ou CEP" est de 122.

On dispose ainsi, par le biais du PEM, d'un indicateur pour étudier la force (d'attraction ou de répulsion) entre une modalité de pratique et une modalité de description.

PROFILS, SIGNATURES, STYLES

Par ailleurs, une analyse factorielle préalable (Cibois 1992), montre que les modalités descriptives qui sont les principales sources d'opposition entre les modalités de pratique sont au nombre de 8 et sont définies de la manière suivant :

- Bac et plus (noté désormais BAC)
- Chef de famille ouvrier (OUV)
- Aucun diplôme ou CEP, certificat d'études primaires (CEP)

- Célibataire (CEL)
- Marié (MAR)
- Retraité (RTR)

- Masculin (MAS)
- Féminin (FEM)

Puisque ce sont ces modalités qui opposent le mieux les pratiques, ce sont elles dont nous calculerons le PEM avec chacune des modalités. On a les résultats suivants pour les deux émissions considérées :

	BAC	OUV	CEP	CEL	MAR	RTR	MAS	FEM
Avis de recherche	26	110	122	72	108	104	81	119
Apostrophes	127	48	66	84	105	100	92	108

Sur cette exemple on voit bien que chaque émission à sa "signature" particulière : toutes les deux plus féminines que masculines, toutes les deux touchant un public non jeune, mais avec l'opposition en ce qui concerne le niveau (par "niveau" on entend à la fois niveau scolaire et social puisque l'indicateur comporte les deux aspects). BAC étant lié à Apostrophes et OUV et CEP à Avis de recherche et inversement.

Si l'on regarde maintenant l'ensemble des émissions (et le fait de ne pas les regarder en n'ayant pas de téléviseur), on a les résultats suivants :

	BAC	OUV	CEP	CEL	MAR	RTR	MAS	FEM
pas de téléviseur	122	80	65	112	82	55	108	92
commissaire Maigret	74	91	118	47	95	117	98	102
ciné-club	120	80	52	119	81	37	93	107
émission A.Decaux	108	63	95	71	89	114	111	89
magazine Super Sexy	81	109	59	119	89	30	133	67
grand échiquier	119	47	95	49	102	115	81	119
avis de recherche	26	110	122	72	108	104	81	119
au théâtre ce soir	88	97	109	72	107	107	67	133
apostrophes	127	48	66	84	105	100	92	108
des chiff. & lettr	54	64	134	57	70	133	98	102
roue de la fortune	26	108	124	85	97	104	87	113
les enfants du rock	96	105	41	141	49	24	115	85
océaniques	123	77	57	104	99	69	98	102
doc. Cousteau	98	104	90	71	110	98	109	91
santa barbara	24	110	121	101	92	100	53	147
maguy	33	104	123	93	92	109	75	125
l'heure de vérité	116	48	94	79	104	108	128	72
ushuaia	106	109	53	117	108	46	110	90
émission Musicales	101	105	52	115	85	77	101	99
Stade 2	74	104	84	108	102	91	173	27

On voit que ces profils, ces signatures, sont tout à fait lisibles et compréhensibles individuellement mais que des rapprochements sont possible. Par exemple les deux profils suivants sont très proches :

	BAC	OUV	CEP	CEL	MAR	RTR	MAS	FEM
commissaire Maigret	74	91	118	47	95	117	98	102
des chiff. & lettr	54	64	134	57	70	133	98	102

Pour les 8 modalités descriptives, quand il y a opposition ou attraction pour une émission, il en est de même pour l'autre. Ces émissions sont associés à un faible niveau, des gens âgés, l'opposition des sexes étant trop faible pour pouvoir être notée.

Pour pouvoir classer l'ensemble, on simplifiera le problème en ne s'intéressant qu'aux "marques" positives, c'est à dire aux attractions : dans le cas précédent, il y a les deux fortes attraction notées. On pourra donc classer l'ensemble en regroupant tous les profils qui ont des attractions pour les mêmes modalités descriptives.

Reste à savoir à partir de quel niveau on considère qu'il y a attraction : si on prend un seuil trop haut, beaucoup d'émissions ne seront pas classées, si on prend un seuil trop bas, il y aura peu de regroupements.

Une manière de trancher est de chercher l'émission qui est la plus neutre, qui a le moins d'aspérité dans son profil : il s'agit de l'émission "Documentaire Cousteau" qui a le profil suivant :

	BAC	OUV	CEP	CEL	MAR	RTR	MAS	FEM
doc. Cousteau	98	104	90	71	110	98	109	91

Il y a des répulsions et des attractions peu marquées : le maximum du PEM est de 110, son minimum de 71. Cette émission peut être considérée comme l'émission la moins marquée, la plus fédérative et à partir de là, on prendra comme seuil pour noter les attractions simplement une valeur plus forte que la plus haute de ce profil, c'est à dire les PEM plus grand que 110. Avec ce critère on a le classement suivant ou seule cette émission ne sera pas marquée :

	BAC	OUV	CEP	CEL	MAR	RTR	MAS	FEM
14 1 pas de télé	+			+				
24B1 ciné-club	+			+				
24E1 grand échiquier	+					+		+
24P1 heure de vérité	+						+	
24H1 apostrophes	+							
24L1 océaniques	+							
24Q1 ushuaia				+				
24R1 "Musicales"				+				
24K1 enfants du rock				+			+	
24D1 Super Sexy				+			+	
24C1 émission A.Decaux					+	+		
24S1 Stade 2							+	
24G1 au théâtre ce soir								+
24M1 documentaire Cousteau								
24O1 maguy			+					+
24F1 avis de recherche			+					+
24J1 la roue de la fortune			+					+
24N1 santa barbara			+					+
24A1 commissaire Maigret			+			+		
24I1 chiffres & lettres			+			+		

On retrouve bien ce qui nous avait servi de point de départ c'est à dire le fait qu'*Apostrophes* soit marquée "Bac et plus" et *Avis de recherche* "Aucun diplôme ou CEP", mais des associations se sont faites et des précisions sont apportées. Par exemple les quatre émissions populaires : *Maguy*, *Avis de recherche*, *la Roue de la fortune* et *Santa Barbara* sont aussi marquées comme féminines, alors que *Commissaire Maigret* et *des Chiffres et des lettres*, également populaire sont marquées par l'âge de leur public.

Il y a quatre groupes d'émissions (ou de refus) de niveau social élevé mais marquées également par d'autres critères:

- jeunes : pas de téléviseur et ciné-club
- âgée et féminine : grand échiquier
- masculine : l'heure de vérité
- non marquée en plus : *apostrophes* et *océaniques*

Avec une absence de niveau social marqué, on a les groupes suivants :

- jeune seulement : *Ushuaia* et *Musicales*
- jeune et masculin : *enfants du Rock* et *super-sexy*
- âgé et masculin : émissions *A.Decaux*
- masculin : *Stade 2*
- féminin : *au théâtre ce soir*
- non classé : *Cousteau*

Pour les groupes d'émission marquées par un faible niveau social (CEP) on a, comme indiqué plus haut :

- les 4 émissions populaires (*Maguy*, *la Roue de la fortune*, *Avis de recherche* et *Santa Barbara*) marquées également comme féminines.
- des *Chiffres et des lettres* et *Commissaire Maigret* marquées "âgées".

Ces regroupements ne remettent pas en cause l'idée que l'on peut se faire spontanément des émissions de télévision, mais ils précisent et formalisent notre savoir. Ils permettent en particulier de trouver un fondement entre l'opposition classique qui a été faite entre les émissions qui se définissent par un contenu dit de qualité et les émissions qui recherchent la plus grande audience et qui de ce fait sont critiquées comme provoquant un nivellement par le bas (Mehl 1991).

Au vu des profils précédents, on comprend bien que cette opposition doit être repensée : toutes les émissions (sauf *Cousteau*) sont marquées vis-à-vis d'un certain public. Elles sont toutes "ciblées" mais toutes les cibles ne sont pas équivalentes ni en prestige ni en quantité : comme aux cibles les plus nobles (niveau élevé comme *Apostrophes* ou *Océaniques*) correspondent des effectifs minoritaires (Bac et plus) et aux cibles de faible niveau (*Avis de recherche*) les effectifs les plus forts (aucun diplôme ou CEP), on retrouve bien sous une forme d'opposition de milieux l'opposition ressentie entre culture et recherche d'audience.

Les promoteurs d'émission populaires ont le sentiment d'être "au service du public", de permettre à "tout le monde" de se retrouver dans une émission mais il y a dans cette attitude une fausse perception dans la mesure où "le public" est identifié socialement à ses composantes les plus nombreuses, ce qui entraîne les réactions critiques des fractions minoritaires du public qui ne s'y reconnaissent pas.

CONCLUSION

L'utilisation du PEM pour repérer des profils ou des styles s'inscrit dans une démarche de dépouillement d'enquête où l'on se sert de techniques factorielles pour se faire une idée globale sur les liaisons entre données et où l'on utilise des techniques comme celles-ci pour affiner, préciser et objectiver les résultats précédents. C'est ce que nous appelons une attitude post-factorielle dans la mesure où elle vient souvent après une analyse factorielle.

A travers les exemples précédents, il nous semble possible de montrer la pertinence du PEM comme indice de liaison entre modalités dans plusieurs types d'utilisation :

- l'utilisation du PEM en terme de profils ne fait que reprendre un usage du sociologue qui cherche à rendre compte d'une modalité en la comparant avec d'autres issues de nombreux tableaux croisés : le PEM permet de repérer les attractions les plus fortes. En l'associant à un test de signification, on se garantit contre le fait de se servir de liaisons fortes mais "insignifiantes" dans la mesure où ces liaisons porteraient sur des effectifs trop faibles pour qu'on puisse en inférer quelque chose.

- l'utilisation du PEM en termes de styles de comportement suppose que l'on ait au préalable repéré des modalités, en général socio-démographiques, qui décrivent le mieux ces comportements. On peut alors en examinant les signatures de PEM associées aux comportements, les classer par similitude de marques positives au-dessus d'un certain seuil. On construit ainsi d'une manière formalisée des styles qui regroupent les comportements communs à des individus de mêmes caractéristiques descriptives.

IMPLEMENTATION

Différents programmes du logiciel Tri-deux utilisent le PEM et en particulier PROFIL qui permet, pour une modalité donnée de trouver le profil des modalités en liaison avec elle, et STYLES qui, à partir de modalités repérées pour être sous-jacentes à des styles de comportement, permet de construire puis de regrouper des styles de comportement.

Le logiciel Tri-deux, est un logiciel libre (freeware) élaboré par l'auteur et centré sur le dépouillement d'enquête utilisant des techniques factorielles ou post-factorielles. Renseignements sur sa diffusion auprès de l'auteur.

**ANNEXE : ALGORITHME DE CALCUL
DU POURCENTAGE DE L'ECART MAXIMUM - PEM**

Pour tout croisement d'une modalité A d'une question avec une modalité B d'une autre question, on dispose d'un effectif observé OBS. On isole la case de croisement dans un tableau éventuellement plus vaste :

	Modalité A	autres modalités	Total
Modalité B	OBS		TOTB
Autres modalités			
Total	TOTA		TOTAL

La marge correspondant à la modalité A est TOTA

La marge correspondant à la modalité B est TOTB

Le total général est TOTAL

L'effectif théorique est $THEO = TOTA \times TOTB / TOTAL$

L'écart à l'indépendance est $ECART = OBS - THEO$

1) Cas d'une attraction entre A et B : l'écart à l'indépendance est positif.

Pour trouver l'écart maximum EMAX, on cherche à mettre la plus forte valeur possible MAX à la place de l'observation. Cette valeur doit cependant être compatible avec les marges et donc on ne peut prendre que la plus faible des deux marges. Cette valeur maximum $MAX = \min (TOTA, TOTB)$

L'écart positif à l'indépendance maximum EMAX est égal à l'effectif maximum qui correspondrait à la liaison maximum moins l'effectif théorique :

$$EMAX = MAX - THEO$$

Le PEM est alors égal au rapport entre l'écart observé et l'écart maximum (multiplié par 100 pour avoir un pourcentage) :

$$PEM = ECART / EMAX \times 100$$

2) Cas d'une opposition entre A et B : l'écart à l'indépendance est négatif.

On cherche cette fois l'effectif minimum MIN qu'il est possible de mettre à la place de l'observation.

Si l'on met zéro dans la case, il faut que le total des observations soient reporté dans les autres lignes et colonnes et cela n'est pas toujours compatible avec les marges comme par exemple dans le cas suivant :

	Modalité A	autres modalités	Total
Modalité B	0		7
Autres modalités			
Total	6		10

En effet comme 6 individus sont dans la première colonne, il n'en reste que 4 à mettre dans les autres, ce qui est incompatible avec le fait que les 7 de la première ligne doivent s'y trouver. On est dans le cas où le complément au total d'une marge est inférieur à l'autre marge. Dans ce cas, pour que le minimum mis dans la case soit compatible avec les marges, il faut qu'il soit plus grand que zéro. Ici il faut que l'effectif minimum soit suffisamment élevé pour que ce qui reste à répartir dans les colonnes autres que la première soit égal au complément de la première colonne c'est à dire 4. Cela suppose que sur la première ligne on ait 4 individus au maximum sur les colonnes autres que la première et donc 3 au minimum sur la première.

Cette valeur de 3 est la différence entre la marge ligne et le complément au total de la marge colonne.

D'une manière générale on a donc deux cas :

- si une marge est inférieure ou égale au complément au total de l'autre marge (et si c'est vrai pour l'une, c'est également vrai pour l'autre), alors on peut prendre pour effectif minimum zéro :

$$\text{MIN} = 0$$

L'écart maximum négatif EMAX à l'indépendance est alors :

$$\text{EMAX} = \text{MIN} - \text{THEO} = -\text{THEO}$$

et enfin

$$\text{PEM} = \text{ECART} / \text{EMAX} \times 100$$

(on notera qu'on retrouve dans ce cas le rapport bien connu écart/théorique : on voit que le PEM peut être considéré comme une généralisation de ce rapport)

- si une marge est supérieure au complément au total de l'autre marge (et réciproquement), alors on peut prendre pour effectif minimum leur différence :

$$\begin{aligned} \text{MIN} &= \text{TOTA} - (\text{TOTAL} - \text{TOTB}) = \text{TOTB} - (\text{TOTAL} - \text{TOTA}) \\ &= \text{TOTA} + \text{TOTB} - \text{TOTAL} \end{aligned}$$

On observe qu'il s'agit de la somme des marges moins le total.

$$\text{EMAX} = \text{MIN} - \text{THEO}$$

$$\text{PEM} = \text{ECART} / \text{EMAX} \times 100$$

D'une manière générale comme EMAX et ECART sont négatifs le pourcentage est positif et par convention on inverse son signe pour exprimer le sens de la liaison :

$$\text{PEM} = - \text{PEM}$$

3) Jusqu'à présent on a exprimé le sens de la liaison en donnant un signe au PEM : on va le centrer à 100 en lui ajoutant cette valeur.

$$\text{PEM} = \text{PEM} + 100$$

Dorénavant, un PEM centré de 100 correspond à l'indépendance, une valeur supérieure à 100 à une attraction entre modalités (200 correspondant à la liaison parfaite par exemple en croisant une modalité avec elle-même), une valeur inférieure à 100 à une opposition entre modalités (0 correspondant à l'opposition absolue).

GENERALISATION

Il est possible de généraliser en le PEM et de considérer non plus le PEM d'une case d'un tableau mais celui d'un tableau complet. C'était d'ailleurs la perspective du V de Kramer et nous étions parti de là en proposant dans L'analyse des données en sociologie (Cibois 1983) un Pourcentage du Khi-deux Maximum (PKM). Le principe (qui vaut pour un indice basé sur l'écart brut à l'indépendance ou sous sa forme pondérée de khi-deux) en est le suivant :

- on cherche à maximiser les attractions sur une diagonale du tableau : ceci n'est possible que s'il y a un ordre entre les lignes et entre les colonnes. Cet ordre peut être donné a priori mais s'il existe, le premier facteur d'une analyse factorielle le manifeste.

- on est donc conduit à faire au préalable une analyse factorielle du tableau pour trouver un ordre qui maximise les attractions sur une diagonale (respectivement les oppositions sur l'autre).

- déplacer lignes et colonnes selon cet ordre et maximiser les effectifs sur la diagonale d'attraction (Pour cet algorithme, cf. Cibois 1983 chap.1))

- calculer pour ce tableau maximum la somme des écarts positifs à l'indépendance.

- mettre en rapport en pourcentage la somme des écarts positifs observés avec la somme correspondants à la liaison maximum. On a ainsi un PEM qui indique la force de la liaison : il est égal à zéro si on est strictement dans le cas de l'indépendance et il peut aller à 100 si on croise une question avec elle-même (cas de la dépendance fonctionnelle).

On ne calcule pas le PEM correspondant à la diagonale d'opposition, c'est le même que l'on affecte donc du signe moins pour manifester le sens de la liaison.

Dans le cas d'un tableau 2 x 2, le PEM calculé avec cet algorithme est le même que celui calculé avec l'algorithme précédent : il est égal au PEM de chacune des cases en attraction.

On notera que dans le cas d'une case (ou d'un tableau 2 x 2) la solution est unique alors que dans le cas d'un tableau plus grand, comme il dépend de l'ordre des lignes et des colonnes et que cet ordre est donné par une analyse factorielle, il peut y avoir des variantes légères selon le type d'analyse utilisé. Pour notre part dans le programme REPFAC qui met en oeuvre cette méthode nous utilisons une analyse non pondérée. Cette technique évite en particulier plus souvent cette situation paradoxale que l'on rencontre quelque fois où l'ordre originel est meilleur que l'ordre proposé par l'analyse factorielle et où la somme des écarts positifs observée est plus grande que celle calculée avec l'algorithme de maximisation de la diagonale. On se trouve alors avec un PEM plus grand que 100. Ce cas paradoxal ne s'observe que sur des petits tableaux et l'on a intérêt dans ce cas à travailler case par case.

BIBLIOGRAPHIE

Amaturo, E. (1990). "Classi sociali et stili di vita", in Enrica Amaturo et Mauro Palumbo, *Classi sociali : stili di vita, coscienza e conflitto di classe*, Genova, ECIG, 1990, p.7-31.

Bertin, J. (1967). *Sémiologie graphique*, Mouton, Paris et La Haye.

Bertin, J. (1977). *La graphique et le traitement graphique de l'information*, Flammarion, Paris.

Cibois, Ph. (1981). "Analyse des données et sociologie", *L'année sociologique*, p.333-348.

Cibois, Ph. (1983). *L'analyse des données en sociologie*, Presses Universitaires de France, Paris.

Cibois, Ph. (1989). "Pour une science sociale synchronique", *Revue du MAUSS*, n°4, 1989, p.70-84.

Cibois, Ph. (1992). "Soixante styles de loisirs et de pratiques culturelles", rapport CERSOF, Université Paris V.

Cramér, H. (1946). *Mathematical methods of statistics*, Princeton University Press, Princeton.

Donnat, O. et Cogneau, D. (1990). *Les pratiques culturelles des français 1973-1989*, Paris, La Découverte et la Documentation française, 1990, 288p.

Herpin, N. (1986). "Socio-styles", *Revue française de sociologie*, XXVII, p.265-272.

Mehl, D. (1991). "Les deux lucarnes", *Esprit*, (10), oct., pp.82-96.

Rouanet, H., Le Roux, B. et Bert, M.-Cl. (1987). *Statistiques en sciences humaines : procédures naturelles*, Dunod, Paris.

Singly de, F. (1993). "Les jeunes et la lecture", *Les dossiers Educations et Formations*, n°24, Ministère de l'Education et de la Culture, Paris.

=====