

B. ESCOFIER

D. DROUET

Analyse des différences entre plusieurs tableaux de fréquence

Les cahiers de l'analyse des données, tome 8, n° 4 (1983),
p. 491-499

http://www.numdam.org/item?id=CAD_1983__8_4_491_0

© Les cahiers de l'analyse des données, Dunod, 1983, 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/>

ANALYSE DES DIFFÉRENCES ENTRE PLUSIEURS TABLEAUX DE FRÉQUENCE

[ANA. PLUS. TAB.]

par B. Escofier (1), D. Drouet (2)

1 *Introduction* : Nous proposons ici une méthode de comparaison de plusieurs tableaux de fréquence définis sur le même couple d'ensembles I et J. Cette méthode a pour but de mettre en évidence systématiquement et d'analyser les distances entre les mêmes lignes des différents tableaux. Elle est dérivée de l'analyse des correspondances et permet une représentation simultanée de toutes les lignes de tous les tableaux et de l'ensemble J des colonnes. Des formules de type barycentrique lient ces deux représentations et permettent d'expliquer les résultats. On peut voir aussi cette méthode comme une analyse des différences entre les tableaux, après l'élimination de l'écart entre leur marge sur I, écarts qui peuvent masquer des phénomènes plus complexes.

Cette méthode a été utilisée par D. Drouet pour étudier, d'une part, l'évolution temporelle, et d'autre part les disparités géographiques des causes de mortalité par classe d'âge en Bretagne (Thèse de 3ème cycle). Les tableaux comparés croisent les classes d'âge et les causes de mortalité. Dans la première étude, ils sont définis chaque année pour l'ensemble de la Bretagne et pour chaque canton de Bretagne cumulé sur 20 ans dans la seconde étude. Un exemple simple destiné uniquement à illustrer la méthode est commenté dans (Généralisation de l'analyse des correspondances à la comparaison de tableaux de fréquence ; B. Escofier rapport IRISA n° 190). Nous donnons ici les principes de la méthode et les formules. Nous renvoyons aux deux publications ci-dessus pour les démonstrations complètes.

2 Notations

Nous notons I l'ensemble des lignes des tableaux et J l'ensemble de leurs colonnes. Les tableaux sont indicés par les éléments d'un ensemble T (évoquant le temps, qui apparaît fréquemment dans ce type de données). Le terme général du tableau t est noté $k(i, j, t)$, ses marges sur I et J sont notées $k(i, t)$ et $k(j, t)$ et son effectif total $k(t)$. Nous considérons aussi la somme de tous les tableaux de fréquence, que nous appelons aussi tableau moyen. Le terme général de cette somme est noté $k(i, j)$, ses marges $k(i)$ et $k(j)$ et son effectif total k .

3 Principe de la méthode dans R_J

3.1. Rappels et méthodes classiques

Les profils des lignes de tous les tableaux sont situés dans l'espace R_J . Nous notons $N(I, t)$ le nuage associé au tableau t, $N(I, T)$ l'union de tous ces nuages (qui est associé au tableau juxtaposition en lignes de tous ces tableaux) et $N(I)$ le nuage associé au tableau somme.

(1) Assistante IUT. Avenue Montaigne - Kercado 56000 VANNES

(2) Chercheur IRISA Campus de Beaulieu - 35042 RENNES CEDEX

Pour comparer les lignes de plusieurs tableaux de fréquence, il est classique de faire l'analyse du tableau somme en projetant les profils des lignes de chaque tableau en éléments supplémentaires. Ceci revient à projeter le nuage $N(I, T)$ sur les axes principaux d'inertie de $N(I)$. Les profils des lignes i du tableau somme étant situés aux centres de gravité des lignes homologues $\{(i, t); t \in T\}$, ces axes rendent maximum l'inertie "inter" de $N(I, T)$. Cette méthode permet de mettre en évidence la structure moyenne de tous les tableaux et les écarts entre les profils des lignes homologues ne sont vus que sur cette structure moyenne. Dans notre exemple d'évolution de la mortalité au cours du temps, les profils de mortalité des différentes classes d'âge d'une même année t , diffèrent beaucoup plus entre eux que ne diffère le profil d'une même classe d'âge au cours du temps. Dans ce cas, les différences entre ces derniers ne seront pas visibles sur les graphiques. Dans le cas extrême où les nuages $N(i, T)$ des profils d'une même ligne autour de leur centre de gravité i sont orthogonaux au sous-espace engendré par $N(I)$, leur dispersion même importante est totalement ignorée par cette analyse.

On peut aussi faire l'analyse du même tableau avec les lignes du tableau $k(I \times T) \times J$ en principales et celles du tableau somme $k(I \times J)$ en supplémentaires. Cette analyse traduira mieux que la précédente les différences entre les lignes homologues puisque les nuages sont projetés sur des axes qui rendent maximum l'inertie totale du nuage $N(I, T)$. L'inertie intra rentre donc en ligne de compte et dans la situation extrême d'orthogonalité évoquée ci-dessus, les différences entre lignes homologues apparaîtront. Mais si, comme dans l'évolution de la mortalité, l'inertie intra (différence entre profils homologues) est beaucoup plus faible que l'inertie inter (différence entre les profils de la somme), les axes obtenus seront proches des axes inter et les différences entre profils homologues ne seront guère plus visibles que dans l'analyse de $N(I)$.

Cette technique ne permet donc pas de mettre en évidence systématiquement et d'analyser les écarts entre les profils des lignes homologues. En posant le problème dans ce cadre géométrique et dans ces termes, la solution que nous proposons vient facilement à l'esprit : nous analyserons les différences entre les profils des lignes homologues en projetant leur nuage $N(I, T)$ sur des axes rendant maximum son inertie intra.

3.2. Maximisation de l'inertie intra

Il est bien clair que la métrique de l'espace R_n est la métrique diagonale $D = k/kJ$ qui est induite à la fois par le tableau juxtaposé en lignes $k((I \times T) \times J)$ et par le tableau somme $k(I \times J)$.

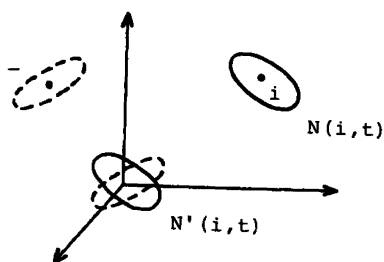
D'après le théorème de Huyghens l'inertie intra de $N(I, T)$ est la différence entre son inertie totale et son inertie inter (inertie de $N(I)$). Ceci est encore vrai au niveau des formes quadratiques d'inertie :

$$W_{\text{intra}} = W_{\text{totale}} - W_{\text{inter}}$$

Les projecteurs sur les axes rendant maximum la projection de l'inertie intra sont obtenus en diagonalisant le produit $W_{\text{intra}} \cdot D$ qui est la différence des deux matrices $W_{\text{totale}} \cdot D$ et $W_{\text{inter}} \cdot D$ associées respectivement au tableau juxtaposé et au tableau somme. Une variante du programme d'analyse des correspondances appliquée au tableau juxtaposé permet de les calculer. Mais on peut obtenir aussi (cf. ci-dessous) les axes d'inertie intra en analysant l'inertie totale d'un nouveau nuage.

3.3. Le nuage recentré : Analyse des correspondances d'un nouveau tableau

L'inertie intra du nuage $N(I, T)$, divisé en $\text{card} I$ nuages $N(i, T)$, est l'inertie cumulée de tous ces nuages en leur centre de gravité respectif i . Elle est égale à l'inertie totale d'un nuage obtenu en les recentrant tous à l'origine.



Nuage $N(I, T)$ et nuage $N'(I, T)$ obtenu en recentrant chaque sous nuage $N(i, T)$ à l'origine. Inertie intra de $N(I, T)$ = Inertie de $N'(I, T)$.

Notons $N'(I, T)$ le nuage ainsi recentré. A la ligne (i, t) est associée dans ce nuage la différence entre les deux profils k_{iJt}/k_{it} et k_{iJ}/k_i :

$$N'(I, T) = \{k(i, J, t)/k(i, t) - k(i, J)/k(i) \mid i \in J; t \in T\}$$

Or, ce nuage est exactement le nuage centré des profils des lignes d'un tableau $k'((IXT) \times J)$ ayant les mêmes marges $k(j)$ sur J et $k(i, t)$ sur IXT que le tableau juxtaposé et de terme général :

$$k'(i, j, t) = k(i, j, t) - k(i, j) \times k(i, t) / k(i) + k(j) \times k(i, t) / k$$

L'analyse des différences entre les lignes homologues des tableaux t , ou de l'inertie "intra" du nuage $N(I, T)$ se ramène donc à l'analyse des correspondances du tableau $k'((IXT) \times J)$. (cf. ANA. INTRACLASSE).

3.4. Projection des nuages $N(I, T)$ et $N'(I, T)$

Soient G_s les vecteurs propres, de norme $\sqrt{\lambda_s}$, de la matrice Wintra.D obtenus, soit dans l'analyse des correspondances du tableau $k'((IXT) \times J)$, soit dans "l'analyse intra" (cf. § 3.2.) du tableau $k((IXT) \times J)$. Ils correspondent aux axes qui maximisent la projection de l'inertie totale de $N'(I, T)$ et celle de l'inertie intra du nuage non recentré $N(I, T)$. Nous calculons et représentons les projections sur ces axes de ces deux nuages qui présentent chacun leur intérêt : le premier est exactement le nuage étudié dans le tableau juxtaposé avec les distances du χ^2 classique entre profils ; dans le second, on représente directement les écarts des profils (i, t) à leur moyenne (i) et on met en évidence, loin de l'origine ceux qui s'écartent le plus de leur moyenne.

Les projections $F_s(i, t)$ du nuage recentré $N'(I, T)$ se déduisent de G_s par les formules de transition du tableau k' :

$$F_s(i, t) = (1/\sqrt{\lambda_s}) \sum \{ (k(i, j, t)/k(i, t) - k(i, j)/k(j) + k(j)/k) G_s(j) \mid j \in J \}$$

Mais les facteurs G_s sont centrés et la formule précédente se simplifie sous la forme :

$$F_s(i, t) = (1/\sqrt{\lambda_s}) \sum \{ (k(i, j, t)/k(i, t) - k(i, j)/k(j)) G_s(j) \mid j \in J \}$$

Les projections $F_s(i, t)$ du nuage non recentré $N(I, T)$ et les projections $F_s(i)$ des centres de gravité i des sous nuages $N(i, T)$ se déduisent de G_s par les formules de transition classique. Pour les obtenir dans l'analyse des correspondances du tableau k' , il suffit de mettre en lignes supplémentaires le tableau juxtaposé et le tableau somme :

$$F_s(i, t) = (1/\sqrt{\lambda_s}) \sum \{ (k(i, j, t)/k(i, t)) G_s(j) \mid j \in J \}$$

et

$$F_s(i) = (1/\sqrt{\lambda_s}) \sum \{ k(i, j)/k(i) G_s(j) \mid j \in J \}$$

On a la relation :

$$F_S(i,t)' = F_S(i,t) - F_S(i)$$

Cette dernière formule permet de calculer $F_S(i,t)'$ dans l'analyse "intra".

3.5. Contribution et qualité de représentation

La contribution à l'inertie de F_S de (i,t) ou de $(i,t)'$ s'écrit :

$$CTR(i,t) = CTR(i,t)' = (k(i,t)/k) [F_S(i,t)']^2 = (k(i,t)/k) [F_S(i,t) - F_S(i)]^2$$

Un élément i du tableau somme ne contribue pas à l'inertie du facteur, mais on appellera contribution de i la contribution du nuage $N(i,T)$:

$$CTR(i) = \sum \{ CTR(i,t) \mid t \in T \}$$

De même, on appellera contribution du tableau t à l'inertie du facteur, la contribution de ses éléments :

$$CTR(t) = \sum \{ CTR(i,t) \mid i \in I \}$$

Un élément (i,t) contribuera beaucoup à l'inertie si le profil de cette ligne est très différent du profil de la ligne i du tableau moyen, et si son poids $k(i,t)$ est assez grand. La contribution d'un élément i à l'inertie sera grande si les profils des lignes $(i,t), t \in T$ diffèrent entre eux. Celle d'un tableau t sera grande si les profils des lignes de ces tableaux diffèrent de celles du tableau moyen.

Les qualités de représentation des (i,t) peuvent être calculées dans le nuage initial $N(I,T)$ ou dans le nuage recentré $N'(I,T)$. Dans le premier cas, c'est la qualité de représentation de l'écart de (i,t) à la moyenne générale, elle n'atteint pas forcément 1 pour l'ensemble des facteurs. Dans le second, c'est celle de l'écart de (i,t) au centre de gravité i de $N(i,T)$; elle atteint 1 pour l'ensemble des facteurs.

4 Élimination de l'effet des marges dans la comparaison des tableaux

4.1. Le problème des marges

Généralement, les marges des tableaux ne sont pas proportionnelles ce qui induit une différence entre les tableaux. Dans notre exemple d'évolution des causes de mortalité, la variation du pourcentage de chaque classe d'âge dans la mortalité totale évolue ; cette évolution est due en grande partie à celle du pourcentage des classes d'âge dans la population qui n'est pas notre propos. D'où la question : comment analyser les différences entre les tableaux en évitant que l'écart entre leur marge sur I ne masque des phénomènes plus fins ? La solution proposée ici est d'éliminer totalement l'effet de ces écarts : en construisant un modèle où les tableaux ne diffèrent que par leur marge sur I , et en analysant les différences entre les données réelles et ce modèle.

4.2. Le tableau modèle $k''((I \times T) \times J)$

Le modèle s'obtient en juxtaposant card T tableaux qui ne diffèrent entre eux que par leur marge sur I , $k(i,t)$. Chacun d'eux se déduit du tableau somme en multipliant son terme général $k(i,j)$ par le rapport $k(i,t)/k(i)$ entre la marge du tableau t et celle du tableau somme. Ce modèle qui a pour terme général $k''(i,j,t) = k(i,j)k(i,t)/k(i)$ a les mêmes marges $k(i,t)$ et $k(j)$ que le tableau de données juxtaposé.

4.3. Ecart entre le tableau de données et le modèle

En analysant l'inertie intra, on réalise l'élimination de l'influence des marges dans la comparaison des card T tableaux. En effet, le nouveau tableau $k'(i, j, t) = k(i, j, t) - k(i, j)k(i, t) / k(i) + k(j)k(i, t) / k$ dont on fait l'analyse des correspondances (cf. § 3.3.) est la différence entre le tableau de données et le tableau modèle ainsi défini, à laquelle se rajoute simplement le produit des marges $k(j)k(i, t) / k$. Ce dernier terme permet simplement d'obtenir un tableau dont les marges sont encore $k(j)$ et $k(i, t)$: les profils sont définis de la même manière et les centres de gravité des nuages sont confondus.

Le profil centré d'une ligne (i, t) de k' est donc la différence entre les 2 profils centrés des lignes homologues du tableau de données et du tableau modèle (les 3 centres de gravité sont situés en $k(J)/k$) :

$$k'(i, J, t) / k(i, t) = k(i, J, t) / k(i, t) - k(i, J) / k(i)$$

De même, le profil centré de la colonne j de k' est la différence entre les 2 profils centrés des colonnes homologues (les 3 centres de gravité sont situés en $k(I, T)/k$) :

$$k'(I, j, T) / k(j) = k(I, j, T) / k(j) - k(I, j)k(I, T) / k(I)k(j)$$

Dans l'analyse du tableau k' , on trouvera aux extrémités des axes les lignes et les colonnes du tableau de données juxtaposé dont les profils s'écartent le plus de ceux du tableau modèle (construit sous l'hypothèse que les card T tableaux ne diffèrent entre eux que par leur marge sur I). C'est une deuxième interprétation des résultats de cette analyse. Pour compléter la première, celle de l'analyse intra du nuage $N(I, T)$ il nous reste à démontrer que le nuage $N'(J)$ associé aux colonnes de k' se définit aussi en termes d'inertie intra ; ceci permet une interprétation des valeurs des facteurs G_S dans ces termes et assure la dualité de l'analyse intra. Montrons d'abord l'équivalence entre le tableau modèle et le tableau somme qui représentent tous deux l'inertie inter-tableaux.

4.4. Tableau modèle et tableau somme

Le nuage des profils des lignes (i, t) du tableau modèle $k(i, j)k(i, t) / k(j)$ se confond avec le nuage $N(I)$ associé au tableau somme : il y a card T profils égaux à $k(i, J) / k(i)$ dont les poids respectifs valent $k(i, t)$, donc de poids total $k(i)$.

Le nuage des profils des colonnes j du tableau modèle, situé dans R_{IXT} est isomorphe au nuage $N_{inter}(J)$ qui est associé au tableau somme dans R_I : les distances entre les profils de deux colonnes j et j' sont les mêmes dans les deux nuages comme nous le vérifions ci-dessous :

$$\begin{aligned} D^2(j, j') &= \sum \{ [k(i, j)k(i, t) / (k(i)k(j)) - k(i, j')k(i, t) / k(i)k(j')]^2 \\ &\quad / [k/k(i, t)] \quad i \in I ; t \in T \} \\ &= \sum \{ [k(i, t) / k(i)]^2 [k(i, j) / k(i) - k(i, j') / k(j')]^2 \\ &\quad / [k/k(i, t)] \quad i \in I ; t \in T \} \\ &= \sum \{ [k(i, j) / k(j) - k(i, j') / k(j')]^2 [k/k(i)] \quad i \in I \} \end{aligned}$$

Pour les lignes comme pour les colonnes, le tableau k' représente la différence entre le tableau juxtaposé et le tableau somme, donc la différence entre l'inertie totale et l'inertie inter.

5 Le nuage $N'(J)$ associé aux colonnes de k'

5.1. La décomposition de $N(J)$

Dans l'espace $R_{I \times T}$ sont situés 3 nuages de profils de colonnes : $N(J)$ associé au tableau juxtaposé ; $N''(J)$ associé au tableau modèle k'' et isomorphe à $Ninter(J)$; et $N'(J)$ associé à k' et différence entre les deux premiers. On montre que les sous espaces engendrés par les deux derniers sont orthogonaux. Les profils associés au tableau juxtaposé se décomposent sur ces deux sous espaces suivant les profils associés à k'' et k' ; les projections $N''(J)$ et $N'(J)$ de $N(J)$ sur ces deux sous espaces représentent respectivement son inertie inter tableaux et son inertie intra. Notons l'analogie entre la restriction à l'inertie intra et le centrage des variables en analyse en composantes principales qui se traduit par un changement d'origine pour les individus (comme dans $N(I, T)$) et une projection sur un sous espace pour les variables (comme dans $N(J)$).

Remarque: La décomposition des éléments de $N(J)$ en deux termes orthogonaux implique que les projections de $N(J)$ et de $N'(J)$ sur les axes d'inertie de ce dernier sont confondues.

5.2. Les distances dans $N'(J)$

D'après le théorème de Pythagore, à la décomposition orthogonale des profils, nous pouvons associer une décomposition des carrés des distances dans les 3 nuages :

$$D_{\text{totale}}^2(j, j') = D_{\text{inter}}^2(j, j') + D_{\text{intra}}^2(j, j')$$

Dans le nuage $N'(J)$, le carré de la distance entre deux points j et j' est la différence entre le carré de la distance induite par le tableau juxtaposé (D_{totale}^2) et le carré de la distance induite par le tableau somme (D_{inter}^2).

6 Lien avec les analyses partielles pour i fixé

Au lieu d'analyser l'inertie "intra", nous aurions pu analyser séparément chaque sous tableau $k(i, J, T)$, correspondant dans notre exemple à chaque classe d'âge. Quel est le lien entre ces analyses partielles et l'analyse intra que nous proposons ?.

6.1. Dans R_J

L'analyse de chaque sous tableau revient à considérer séparément chacun des petits nuages $N(i, T)$ autour de son centre de gravité, et avec sa propre métrique k_i/k_{iJ} . Dans l'analyse intra, nous choisissons une métrique commune k/k_J , et nous cumulons l'inertie de chacun de ces petits nuages.

6.2. Dans $R_{I \times T}$

L'espace $R_{I \times T}$ se décompose en une somme directe orthogonale de $\text{card } I$ sous espaces :

$$R_{I \times T} = \bigoplus_i R_{i \times T}$$

Le profil centré d'un point j associé à $N'(J)$ se décompose sur ces sous espaces, sa composante sur $R_{i \times T}$ est :

$$\frac{k(i, j, T) - k(j)k(i, T)/k(i)k(j)}{[k(i, j)/k(j)] [k(i, j, T) - k(i, T)/k(i)]}$$

On retrouve le profil centré de j dans l'analyse du sous tableau défini par i , multiplié par le coefficient $k(i,j)/k(j)$. Ainsi le profil de j apparaît comme une somme pondérée des profils associés à chaque card I sous tableaux.

L'inertie d'un profil j de $N'(J)$ s'écrit :

$$\sum \{ [k(j)/k(i,t)] [k(i,j,t)/k(j) - k(i,j)k(i,t)/k(j)k(i)]^2 \mid i \in I, t \in T \}$$

Dans le sous tableau i , l'inertie de j vaut

$$\sum \{ [k(i,j)/k(i,t)] [k(i,j,t)/k(i,j) - k(i,t)/k(i)]^2 \mid i \in I \}$$

Donc :

$$\text{Inertie de } j \text{ dans } N'(J) = \sum \{ (k(i,j)/k(j)) \text{Inertie de } j \text{ dans } i \mid i \in I \}$$

L'inertie d'un profil j dans l'analyse intra est donc une somme pondérée de l'inertie de j dans chacune des card I analyses. Le terme de pondération $k(i,j)/k(j)$ représente alors l'importance relative de la classe i pour la variable j . Si une variable j n'a une inertie importante que dans l'analyse d'une seule classe i , mais un poids relativement faible dans cette classe, l'inertie intra de j sera faible ; l'analyse intra apparaît comme une espèce de filtre par rapport aux variables ou ne conservant que celles dont l'inertie dans la plupart des analyses partielles est forte. (Dans l'exemple traité, certaines causes de mortalité de faible poids qui perturbent l'analyse d'une classe d'âge sont étouffées dans l'analyse intra).

7 Interprétation de la représentation simultanée

La représentation simultanée enrichit beaucoup l'interprétation des projections de chaque nuage. Nous commentons ici les formules de transition qui lient ces représentations en nous référant à l'exemple traité.

7.1. Projection de chacun des nuages

Projection de $N'(I,T)$: on trouve aux extrémités des axes et participant le plus à leur formation les profils k_{iJt}/k_{it} qui s'éloignent le plus du profil moyen k_{iJ}/k_i . (Les classes d'âge - cantons dont le profil de mortalité diffère du profil moyen de la même classe d'âge).

Projection de $N(I,T)$: on trouve aux extrémités des axes, (bien que ces axes ne soient pas ceux qui les mettent le mieux en évidence), les profils qui s'éloignent le plus de la moyenne générale k_J/k . (Les classes d'âge-cantons dont le profil de mortalité diffère du profil moyen, toute classe d'âge confondues).

Projection de $N'(J)$ ou $N(J)$: les j qui participent le plus à la formation des axes sont ceux qui ont une inertie importante dans les card I sous tableaux ; les causes de mortalité de probabilité les plus inégales dans les différents cantons à l'intérieur d'une ou de plusieurs classes d'âge.

7.2. Les formules de transition

Rappelons que :

$$F_S(i,t) = \sum \{ [k(i,j,t)/k(i,t) - k(i,j)/k(i)] G_S(j) \mid j \in J \}$$

Le profil recentré $(i,t)'$ est donc situé du côté des j auquel il s'associe plus que la moyenne i , et du côté opposé de ceux auquel il

s'associe moins. Un couple classe d'âge-cantons est situé du côté des causes de mortalité plus fréquentes dans ce canton qu'ailleurs dans cette classe d'âge.

La projection de $N(I, T)$ se déduit de G_s par la formule de transition classique.

$$F_s(i, t) = \sum \{ [k(i, j, t) / k(i, t)] G_s(j) \mid j \in J \}$$

Le profil (i, t) est situé du côté des j auxquels il s'associe le plus, une classe d'âge-cantons est située du côté des causes de mortalité qui y sont les plus fréquentes.

La projection de $N'(J)$ se déduit de celle de $N'(I, T)$ par la formule classique puisque les projections de $N(J)$ et de $N'(J)$ sont confondues.

$$G_s(j) = \sum \{ [k(i, j, t) / k(j)] F_s(i, t) \mid i \in I ; t \in T \}$$

Elle s'écrit en fonction du nuage non recentré :

$$G_s(j) = \sum \{ [k(i, j, t) / k(j) - (k(i, j) / k(i)) (k(i, t) / k(j))] F_s(i, t) \mid i \in I ; t \in T \}$$

Une cause de mortalité j est située du côté des profils recentrés cantons-classe d'âge dans lesquels elle apparaît le plus. Elle est située du côté des profils non recentrés cantons-classe d'âge auxquels elle s'associe plus que dans le modèle où les card T tableaux ne diffèrent que par des proportions différentes des classes d'âge.

7.3. Éléments supplémentaires : représentation d'un tableau

Les formules de transition permettent de projeter des lignes et des colonnes supplémentaires ; en particulier, la somme des lignes d'un même tableau dans k' , i.e. le centre de gravité des profils de ce tableau dans le nuage recentré $N'(I, T)$, le tableau t est alors représenté par le profil :

$$k(J, t) / k(t) - \sum \{ [k(i, j) / k(i)] [k(i, t) / k(t)] \mid i \in I \}$$

On retrouve une idée de J. Daudin (Analyse factorielle des dépendances partielles R.S.A., 1981, n°2) pour étudier la liaison entre T et J "détachée de I " en éliminant la dépendance "attachée" à I . Pour cela il propose une analyse des correspondances du tableau de dimension $J \times T$ et de terme général $k(j, t) / k - \sum \{ k(i, j) k(i, t) / k(i) \mid i \in I \} + k(j) k(t) / k$. Il considère donc des profils qui sont exactement ceux avec lesquels nous représentons la moyenne des tableaux t . Il travaille ainsi sur une moyenne de toutes les classes i , alors que nous les "cumulons".

8 Interprétation en termes de correspondance ternaire

Dans l'analyse intra, à un profil j nous associons une différence entre 2 mesures de probabilités : la première est la mesure de probabilité conditionnelle sachant j . La seconde est le produit d'une mesure sur I et d'une mesure sur J qui peuvent aussi s'écrire comme des probabilités conditionnelles :

$$j \rightarrow P\{(i, t) / j\} - P\{i / j\} P\{t / i\}$$

L'analyse de $N'(J)$ est donc l'analyse des écarts entre ces deux mesures induites sur $I \times T$. En termes de correspondance ternaire, la première est définie par une "tranche" du tableau ternaire $I \times J \times T$. Les deux autres sont définies par les marges sur $I \times J$ et $I \times T$ de ce tableau.

Le point j sera éloigné du centre de gravité si la "tranche" induite par j ne peut être reconstituée à partir des marges sur $I \times J$ et sur

IxT, s'il existe donc un terme d'interaction ternaire.

Remarquons que ce ne sont pas les termes d'interaction ternaire qui sont analysés. Dans ces derniers I, J, T jouent un rôle symétrique, ce qui n'est pas le cas ici. Nous avons précisé que la comparaison s'orientait uniquement selon la dimension T et, ceci posé, nous avons encore disymétrisé le problème sur I et sur J en faisant jouer un rôle différent aux lignes et aux colonnes.

9 Conclusion

La méthode développée ici d'analyse de l'inertie intra-lignes de plusieurs tableaux de fréquence croisant le même couple d'ensembles I et J permet la comparaison de ces tableaux en mettant en évidence des différences même très petites entre les profils des lignes des tableaux à comparer. Elle s'est montrée particulièrement appropriée pour mettre en évidence l'évolution d'un même tableau au cours du temps. Un autre usage de cette méthode est la comparaison de plusieurs tableaux, en tenant compte de l'effet des différences de marges des tableaux, sur un des ensembles. En dernier lieu, cette méthode peut s'interpréter comme l'étude du lien entre 2 variables T et J, après avoir éliminé l'effet sur ce lien d'une troisième variable I.

Facile à programmer, elle devrait pouvoir s'intégrer dans la panoplie des méthodes actuellement disponibles pour comparer plusieurs tableaux, en particulier l'analyse des tableaux juxtaposés et du tableau somme ; elle ne peut d'ailleurs s'utiliser séparément de l'une de ces 2 méthodes. L'analyse du tableau-somme (ou de l'inertie inter-lignes) permettant de mettre en évidence la structure commune à tous les tableaux, alors que l'analyse de l'inertie-intra met en évidence les écarts des tableaux au tableau moyen.

On pourrait alors concevoir une chaîne de programmes qui, lorsqu'on veut comparer plusieurs tableaux de fréquence, analyse successivement le tableau-somme, les tableaux juxtaposés en ligne, l'inertie intra-lignes homologues, et (si c'est judicieux) les tableaux juxtaposés en colonnes et l'inertie-intra-colonnes homologues ; ceci procurerait une comparaison complète, ou du moins sous plusieurs points de vue différents des tableaux de fréquence.