Statistique et analyse des données

Yves Tille

Méthodes d'ajustement d'un tableau à des marges, propriétés des ajustements et choix des critères à optimiser

Statistique et analyse des données, tome 14, n° 3 (1989), p. 53-78

<http://www.numdam.org/item?id=SAD_1989__14_3_53_0>

Statistique et Analyse des Données
1989 - Vol. 14 n°3 - p. 53 - 78

METHODES D'AJUSTEMENT D'UN TABLEAU A DES MARGES, PROPRIETES DES AJUSTEMENTS ET CHOIX DES CRITERES A OPTIMISER

YVES TILLE

Laboratoire de Méthodologie du Traitement des Données
Institut de Sociologie
Université Libre de Bruxelles

Résumé : *Le but de cet article est d'étudier les méthodes d'ajustement de tableaux qui découlent de la minimisation de critères de dissimilarités sous les contraintes données par des marges fixées. Nous montrons d'abord qu'il existe une solution pour tous les critères de type quadratique. Ensuite, nous définissons un ensemble de propriétés pour les méthodes d'ajustement et nous étudions de manière systématique les propriétés que possèdent la plupart des méthodes proposées.*

Abstract : *This paper aims to study the table adjustment methods which are derived from the minimization of a dissimilarity criterium under the constraint given by the fixed marginal totals. First of all, we show that a solution exists for any quadratic criteria. After that, we define a set of properties for the adjustment methods and we search systematically the properties that the most proposed methods possess.*

Mots-clefs: *Ajustement de tableaux, Marges, Dissimilarités.*

Indice de classification STMA : *06-030*

Manuscrit reçu le 23 mars 1989, révisé le 15 février 1990

1. POSITION DU PROBLEME

Le problème est simple : on dispose d'un tableau de fréquences et on veut trouver un tableau "proche" de ce tableau qui possède des marges fixées; ce problème se résout généralement en minimisant un critère de dissimilarité entre le tableau de départ et le tableau recherché sous les contraintes que le tableau recherché possède bien les marges désirées. Il est possible dès lors de définir différentes méthodes d'ajustement en fonction du critère de dissimilarité choisi. La résolution de ce problème trouve de multiples applications particulièrement en théorie des sondages. Ainsi, les méthodes d'ajustement de tableaux permettent de "caler" les résultats d'un sondage sur ceux d'un recensement (stratification *a posteriori*), de modéliser des non-réponses et d'effectuer des projections de tableaux. Ce problème, déjà traité par Deming et Stephan [3] en 1940, a fait l'objet depuis lors de multiples publications, citons entre autres : Thionet [16] [17], Froment et Lenclud [8], Stemmelen [14], Benzécri [2] et Madre [11] [12].

Signalons également qu'il est possible d'ajuster un tableau à des marges en utilisant la formule de reconstitution de l'analyse factorielle des correspondances. Cette méthode, proposée initialement par Stemmelen [14], fut développée ensuite par Benzécri, Bougarit et Madre [1], Escofier [5], Fichet et Gbegan [6], et Tillé [18]. Nous nous bornerons ici à étudier les méthodes issues de la minimisation d'un critère de dissimilarité. Après un bref rappel des diverses méthodes existantes, nous allons montrer que nous pouvons trouver des solutions pour tous les critères de type quadratique. Ensuite, nous définirons plusieurs propriétés que peuvent posséder les méthodes d'ajustement de tableaux et nous examinerons systématiquement les propriétés des différentes méthodes d'ajustement.

2. LES METHODES EXISTANTES

Soit un tableau de fréquences $F = (f_{ij})$ de dimension nxp, tel que $\sum_i \sum_j f_{ij} = 1$ et $f_{ij} > 0$, $\forall$ i, j , dont les marges sont $f_{I.} = (f_{1.}, \ldots, f_{i.}, \ldots, f_{n.})$ et $f_{.J} = (f_{.1}, \ldots, f_{.j}, \ldots, f_{.p})$, nous cherchons un tableau G dont les marges sont $g_{I.} = (g_{1.}, \ldots, g_{i.}, \ldots, g_{n.})$ et $g_{.J} = (g_{.1}, \ldots, g_{.j}, \ldots, g_{.p})$. Le tableau G doit être proche de F, nous verrons que c'est en fonction du critère de dissimilarité défini entre F et G que nous obtiendrons des résultats distincts.

2.1 Solution avec la distance euclidienne des fréquences

La solution des moindres carrés fut la première à être proposée par Deming et Stephan [3] en 1940. Elle consiste à minimiser la distance euclidienne définie entre F et G :

$$d^2(G, F) = \sum_i \sum_j (g_{ij} - f_{ij})^2 \quad , \tag{2.1.1}$$

sous la contrainte que G ait bien les marges désirées. L'élément général du tableau ajusté vaut alors :

$$g_{ij} = f_{ij} + \frac{g_{i.} - f_{i.}}{p} + \frac{g_{.j} - f_{.j}}{n} \quad \forall \text{ i,j} \tag{2.1.2}$$

2.2 Solution avec le critère quadratique pondéré

Une deuxième solution, donnée par Madre [11], consiste à minimiser la distance quadratique pondéré :

$$C_{\rho}(G,F) = \sum_i \sum_j \frac{\left(g_{ij} - f_{ij}\right)^2}{\left[h\, f_{i.} + (1-h) g_{i.}\right]\left[h\, g_{j.} + (1-h) f_{.j}\right]} \qquad h \in]0,1[, \qquad (2.2.1)$$

l'élément général du tableau ajusté vaut alors :

$$g_{ij} = f_{ij} - f_{i.}\, f_{.j} + g_{i.}\, g_{.j} \qquad \forall \quad i,j . \qquad (2.2.2)$$

2.3 Utilisation du phi-deux de contingence

Une solution proposée également par Stephan [15] consiste à minimiser le phi-deux de contingence :

$$\phi^2(G,F) = \sum_i \sum_j \frac{\left(g_{ij} - f_{ij}\right)^2}{f_{ij}} \quad . \qquad (2.3.1)$$

La solution est donnée en résolvant le système suivant :

$$\begin{cases} g_{ij} = f_{ij}\ (\lambda_i + \mu_j) & \forall\ i=1,\ldots,n \quad j=1,\ldots,p\text{-}1 \\ g_{ip} = f_{ip}\ \lambda_i & \forall\ i=1,\ldots,n \\ \sum\limits_i g_{ij} = g_{.j} & \forall\ j=1,\ldots,p\text{-}1 \\ \sum\limits_j g_{ij} = g_{i.} & \forall\ i=1,\ldots,n\ . \end{cases} \qquad (2.3.2)$$

Le système peut être résolu soit comme un système linéaire soit par l'algorithme de Stephan [15]. La méthode A.S.AM. proposée par Durieux et Payen [4] aboutit (au cas où les variances sont proportionnelles aux effectifs) également à la minimisation du phi-deux.

2.4 Utilisation de la distance des rapports à l'indépendance

Cette distance vaut :

$$d_r^2(G,F) = \sum_i \sum_j \left[\frac{g_{ij}}{g_{i.}\, g_{.j}} - \frac{f_{ij}}{f_{i.}\, f_{.j}}\right]^2 \quad . \qquad (2.4.1)$$

Le problème d'ajustement est soluble et il existe une solution unique, ce problème a été étudié en détail par Fromend et Lenclud [8].

2.5 Critère du rapport uniforme

Nous utiliserons également le critère du rapport uniforme que nous définirons ainsi :

$$C_u(G,F) = \sum_i \sum_j \left[\frac{g_{ij}}{f_{ij}} - 1\right]^2 = \sum_i \sum_j \frac{\left(g_{ij} - f_{ij}\right)^2}{f_{ij}^2} \quad . \tag{2.5.1}$$

Nous verrons que l'on peut également résoudre le problème d'ajustement défini par ce critère, le tableau ajusté est l'unique solution du système :

$$\begin{cases} g_{ij} = f_{ij}\left[\,1 + f_{ij}\left(\lambda_i + \mu_j\right)\right] & \forall \; i = 1,\ldots,n \quad j = 1,\ldots,p\text{-}1 \\ g_{ip} = f_{ip}\left[\,1 + f_{ip}\,\lambda_i\right] & \forall \; i = 1,\ldots,n \\ \sum\limits_i g_{ij} = g_{.j} & \forall \; j = 1,\ldots,p\text{-}1 \\ \sum\limits_j g_{ij} = g_{i.} & \forall \; i = 1,\ldots,n \; . \end{cases} \tag{2.5.2}$$

2.6 Information de Kullback et algorithme R.A.S.

Enfin, la minimisation en G de l'information de Kullback :

$$I(G,F) = \sum_i \sum_j g_{ij} \log \frac{g_{ij}}{f_{ij}} \; , \tag{2.6.1}$$

sous les contraintes des marges, aboutit à l'algorithme R.A.S. Cet algorithme est décrit, par exemple, par Madre [11]. La solution de l'algorithme fournit l'unique solution du système :

$$\begin{cases} g_{ij} = f_{ij}\,\alpha_i\,\beta_j & \forall \; i = 1,\ldots,n \quad j = 1,\ldots,p\text{-}1 \\ g_{ip} = f_{ip}\,\alpha_i & \forall \; i = 1,\ldots,n \\ \sum\limits_i g_{ij} = g_{.j} & \forall \; j = 1,\ldots,p\text{-}1 \\ \sum\limits_j g_{ij} = g_{i.} & \forall \; i = 1,\ldots,n \; . \end{cases} \tag{2.6.2}$$

3. CRITERE QUADRATIQUE GENERALISE

3.1. Le problème

Nous allons montrer qu'il existe une solution pour un critère quadratique tout à fait général. Ce critère, nous le définirons ainsi :

$$C(F,G) = \sum_i \sum_j \frac{\left(a_{ij} f_{ij} - b_{ij} g_{ij}\right)^2}{c_{ij}} \tag{3.1.1}$$

où $a_{ij} > 0$, $b_{ij} > 0$ et $c_{ij} > 0$ $\forall$ i , j . Ce critère est différentiable et strictement convexe en G en effet :

$$\frac{\partial^2 C(F,G)}{\partial g_{ij}^2} = \frac{2\, b_{ij}^2}{c_{ij}} > 0 \quad \forall\ i, j,$$

$$\frac{\partial^2 C(F,G)}{\partial g_{ij} \partial g_{i'j'}} = 0 \quad \forall\ (i,j) \neq (i',j').$$

Minimisons ce critère par rapport à G sous les contraintes :

$$\begin{cases} \sum_j g_{ij} = g_{i.} & \forall\ i\ , \\ \sum_i g_{ij} = g_{.j} & \forall\ j\ . \end{cases} \tag{3.1.2}$$

3.2. Existence et unicité de la solution

Géométriquement, le problème consiste à projeter le tableau F sur la variété affine définie par les contraintes. La fonction à minimiser est convexe et les contraintes sont linéaires; nous sommes donc assurés de l'existence et de l'unicité de la solution.

3.3. Résolution

Remarquons d'abord que les n+p contraintes définies en (3.1.2) ne sont pas indépendantes puisque

$$\sum_i \sum_j g_{ij} = \sum_j \sum_i g_{ij} = 1\ .$$

Afin de pouvoir utiliser la technique des multiplicateurs de Lagrange qui requiert l'hypothèse d'indépendance des formes linéaires obtenues en dérivant les fonctions définissant les contraintes, il convient de supprimer une des équations du système (3.1.2) (nous supprimerons ici la dernière équation). L'unicité des multiplicateurs de Lagrange est alors assurée. La fonction lagrangienne définie par ce problème s'écrit alors :

$$\mathcal{L}(L) = \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{\left[a_{ij} f_{ij} - b_{ij} g_{ij}\right]^2}{c_{ij}} - 2 \sum_{i=1}^{n} \lambda_i \left[\sum_{j=1}^{p} g_{ij} - g_{i.}\right] - 2 \sum_{j=1}^{p-1} \mu_j \left[\sum_{i=1}^{n} g_{ij} - g_{.j}\right] \ . \tag{3.3.1}$$

En annulant les dérivées premières par rapport aux g_{ij}, λ_i et μ_j, nous obtenons le système à np + n + p - 1 équations et inconnues :

$$
\begin{cases}
\dfrac{b_{ij}^2}{c_{ij}}\, g_{ij} = \dfrac{a_{ij}\, b_{ij}}{c_{ij}}\, f_{ij} + \lambda_i + \mu_j & \forall \;\; i=1\ldots, n \;, j = 1,\ldots, p\text{-}1 \\[2ex]
\dfrac{b_{ip}^2}{c_{ip}}\, g_{ip} = \dfrac{a_{ip}\, b_{ip}}{c_{ip}}\, f_{ip} + \lambda_i & \forall \;\; i=1\ldots, n \\[2ex]
\sum\limits_{i} g_{ij} = g_{.j} & \forall \;\; j = 1, \ldots, p\text{-}1 \\[2ex]
\sum\limits_{j} g_{ij} = g_{i.} & \forall \;\; i=1\ldots, n
\end{cases}
\qquad (3.3.2)
$$

Afin de déterminer les valeurs des λ_i et μ_j. Nous allons résoudre le système suivant à n+p-1 équations et inconnues issu du système (3.3.2)

$$
\begin{cases}
g_{i.} = \sum\limits_{j=1}^{p} \dfrac{a_{ij} f_{ij}}{b_{ij}} + \lambda_i \sum\limits_{j=1}^{p} \dfrac{c_{ij}}{b_{ij}^2} + \sum\limits_{j=1}^{p-1} \dfrac{c_{ij}}{b_{ij}^2}\, \mu_j & \forall \;\; i = 1, \ldots, n \\[2ex]
g_{.j} = \sum\limits_{i=1}^{n} \dfrac{a_{ij} f_{ij}}{b_{ij}} + \sum\limits_{i=1}^{n} \dfrac{c_{ij}}{b_{ij}^2}\, \lambda_i + \mu_j \sum\limits_{i=1}^{n} \dfrac{c_{ij}}{b_{ij}^2} & \forall \;\; j = 1, \ldots, p\text{-}1
\end{cases}
\qquad (3.3.3)
$$

Ce système peut également s'écrire :

$$
\begin{cases}
\lambda_i = \dfrac{g_{i.} - \sum\limits_{j=1}^{p} \dfrac{a_{ij} f_{ij}}{b_{ij}}}{\sum\limits_{j=1}^{p} \dfrac{c_{ij}}{b_{ij}^2}} - \sum\limits_{j=1}^{p-1} \dfrac{\dfrac{c_{ij}}{b_{ij}^2}}{\sum\limits_{j=1}^{p} \dfrac{c_{ij}}{b_{ij}^2}}\, \mu_j & \forall \;\; i = 1, \ldots, n \\[4ex]
\mu_j = \dfrac{g_{.j} - \sum\limits_{i=1}^{n} \dfrac{a_{ij} f_{ij}}{b_{ij}}}{\sum\limits_{i=1}^{n} \dfrac{c_{ij}}{b_{ij}^2}} - \sum\limits_{i=1}^{n} \dfrac{\dfrac{c_{ij}}{b_{ij}^2}}{\sum\limits_{i=1}^{n} \dfrac{c_{ij}}{b_{ij}^2}}\, \lambda_i & \forall \;\; j = 1, \ldots, p\text{-}1
\end{cases}
\qquad (3.3.4)
$$

Si nous posons : $\lambda' = (\lambda_1, \ldots, \lambda_i, \ldots, \lambda_n)$, $\mu' = (\mu_1, \ldots, \mu_j, \ldots, \mu_{p-1})$, $\eta' = (\eta_1, \ldots, \eta_i, \ldots, \eta_n)$, $\varepsilon' = (\varepsilon_1, \ldots, \varepsilon_j, \ldots, \varepsilon_{p-1})$ où

$$\eta_i = \frac{g_{i.} - \sum_{j=1}^{p} \frac{a_{ij} f_{ij}}{b_{ij}}}{\sum_{j=1}^{p} \frac{c_{ij}}{b_{ij}^2}} \qquad \forall\ i = 1, \ldots, n \quad , \tag{3.3.5}$$

$$\varepsilon_j = \frac{g_{.j} - \sum_{i=1}^{n} \frac{a_{ij} f_{ij}}{b_{ij}}}{\sum_{i=1}^{n} \frac{c_{ij}}{b_{ij}^2}} \qquad \forall\ j = 1, \ldots, p-1 \quad , \tag{3.3.6}$$

et

$$H = [h_{ij}] \quad \text{où} \quad h_{ij} = \frac{\frac{c_{ij}}{b_{ij}^2}}{\sum_{j=1}^{p} \frac{c_{ij}}{b_{ij}^2}} \quad \forall\ i = 1, \ldots, n \quad j = 1, \ldots, p-1 \tag{3.3.7}$$

$$E = [e_{ij}] \quad \text{où} \quad e_{ij} = \frac{\frac{c_{ij}}{b_{ij}^2}}{\sum_{i=1}^{n} \frac{c_{ij}}{b_{ij}^2}} \quad \forall\ i = 1, \ldots, n \quad j = 1, \ldots, p-1 \ , \tag{3.3.8}$$

nous pouvons écrire le système (3.3.4) sous forme matricielle :

$$\begin{cases} \lambda = \eta - H\mu \\ \mu = \varepsilon - E'\lambda \ . \end{cases} \tag{3.3.9}$$

Nous voulons déterminer λ et μ. De (3.3.9) , nous tirons :

$$\mu = \varepsilon - E' (\eta - H \mu) \ , \tag{3.3.10}$$

ce qui nous donne :

$$(I - E'H) \mu = \varepsilon - E' \eta \ , \tag{3.3.11}$$

où I est la matrice identité. Posons :

$$W = I - E'H \ , \tag{3.3.12}$$

et

$$v = \varepsilon - E'\eta \ , \tag{3.3.13}$$

l'élément général de W est donné par :

$$w_{jk} = \delta_{(j=k)} - \sum_{i=1}^{n} \frac{\frac{c_{ij}}{b_{ij}^2}\ \frac{c_{ik}}{b_{ik}^2}}{\left(\sum_{i=1}^{n}\frac{c_{ij}}{b_{ij}^2}\right)\left(\sum_{j=1}^{p}\frac{c_{ij}}{b_{ij}^2}\right)} \quad \forall\ j = 1, .., p\text{-}1 \quad k = 1, .., p\text{-}1 \tag{3.3.14}$$

où $\delta_{(j=k)}$ est la fonction de Kronecker. Nous pouvons donc déterminer les valeurs de λ et μ qui fournissent une solution au système (3.3.9) :

$$\mu^* = W^{-1} v \ , \tag{3.3.15}$$

$$\lambda^* = \eta - H\mu^* \ .$$

Comme l'existence et l'unicité des multiplicateurs de Lagrange est assurée, W est toujours inversible. Connaissant μ^* et λ^*, nous pouvons déterminer le tableau G^* qui fournit la solution du système (3.3.2). La valeur du critère C(F,G*) vaut alors :

$$C(F,G^*) = \sum_{i=1}^{n}\sum_{j=1}^{p-1} \frac{c_{ij}\left[\lambda_i^* + \mu_j^*\right]^2}{b_{ij}^2} + \sum_{i} \frac{c_{ip}\lambda_i^{*2}}{b_{ip}^2} \ . \tag{3.3.16}$$

Montrons maintenant que G* est bien un minimum. Soit $G^{\sim}$ un autre tableau qui satisfait aux contraintes données par (3.1.2), on peut toujours écrire :

$$G^{\sim} = G^* + Z$$

où

$$Z = [z_{ij}] \ ,$$

et

$$\begin{cases} \sum_i z_{ij} = 0 & \forall\ j \\ \sum_j z_{ij} = 0 & \forall\ i\ . \end{cases}$$

On voit directement que

$$C(F,G^{\sim}) = C(F,G^*+Z) = C(F,G^*) + \sum_{i=1}^{n}\sum_{j=1}^{p} \frac{b_{ij}^2 z_{ij}^2}{c_{ij}} \tag{3.3.17}$$

et donc que

$$C(F,G^{\sim}) \geq C(F,G^*) \ .$$

La solution est relativement simple. La seule difficulté de calcul consiste en l'inversion de la matrice W. Il est clair que cette démarche reste valable si on transpose le tableau F, nous pouvons donc toujours faire en sorte que la dimension du tableau à inverser soit le minimum de n-1 et de p-1. Il est alors facile de programmer une solution pour la plupart des distances usuelles qui découlent de la distance quadratique généralisée. L'ajustement d'un tableau à des marges avec le critère quadratique généralisé est bien sûr une généralisation des méthodes exposées au § 2.1,2.2, 2.3, 2.4, 2.5. Il faut toutefois préciser que, pour le critère du rapport uniforme, le critère du phi-deux et l'information de Kullback, la fonction à minimiser n'est plus définie si un ou plusieurs éléments du tableau de départ sont nuls, l'existence et l'unicité de la solution ne sont alors plus nécessairement assurées. Pour que le problème d'ajustement ait encore une solution, il faut qu'il existe au moins un tableau possédant les mêmes éléments nuls que le tableau de départ et ayant bien les marges fixées.

Nous avons développé un logiciel en Turbo-Pascal permettant de réaliser ces divers ajustements. Ce logiciel est à la disposition de toutes personnes intéressées.

4. PROPRIETES DES METHODES D'AJUSTEMENT DE TABLEAUX

Si G est l'ajustement au tableau F, qui minimise le critère C entre ces deux tableaux sous les contraintes des marges $g_{I.}$ et $g_{.J}$, nous noterons :

$$G = Ajus_C (F , g_{I.} , g_{.J})$$

cette méthode d'ajustement. Un ajustement peut selon le choix du critère présenter ou non certaines propriétés. Les propriétés que nous énonçons ici se rapportent à des tableaux de fréquences. Nous supposerons que tous les éléments ces tableaux sont positifs. L'étude de ces propriétés nous permettra de nous guider dans le choix de la méthode d'ajustement.

4.1. Propriété d'agrégation forte

Une méthode d'ajustement possède la propriété d'agrégation forte si le tableau obtenu en agrégeant deux lignes (ou colonnes) avant ou après l'ajustement est toujours identique.

4.2. Propriété d'agrégation restreinte

La propriété d'agrégation faible s'exprime de la même manière que la propriété d'agrégation restreinte mais les lignes (ou colonnes) agrégées doivent être de même profil.

4.3. Persistance de l'indépendance

Si le tableau F présente une situation d'indépendance : $f_{ij} = f_{i.} f_{.j} \quad \forall$ i,j , et que les éléments du tableau G donné par :

$$G = Ajus_C(G , g_{I.} , g_{.J}) ,$$

peuvent toujours s'écrire sous la forme : $g_{ij} = g_{i.} g_{.j} \quad \forall$ i,j , la méthode d'ajustement possède la propriété de persistance de l'indépendance.

4.4. Réversibilité

L'ajustement est réversible quand, quelque soit le tableau F et les marges $g_{I.}$ et $g_{.J}$:

$$F = Ajus_C(Ajus_C(F , g_{I.} , g_{.J}) , f_{I.} , f_{.J}) .$$

4.5. Conservation des éléments positifs

L'ajustement conserve les éléments positifs si le tableau ajusté à des marges positives garde toujours tous ses éléments positifs ou nuls.

4.6. Conservation des éléments nuls

L'ajustement conserve les éléments nuls si, à tout élément nul du tableau à ajuster, correspond un élément nul du tableau ajusté.

4.7. Propriété d'additivité

Soit les marges $g_{I.}^{(1)}$, $g_{.J}^{(1)}$ et les marges $g_{I.}^{(2)}$, $g_{.J}^{(2)}$ qui sont telles que :

$$h\ g_{i.}^{(1)} + (1-h)\ g_{i.}^{(2)} = f_{i.} \quad \forall\ i \in I\ ,$$

$$h\ g_{.j}^{(1)} + (1-h)\ g_{.j}^{(2)} = f_{.j} \quad \forall\ j \in J \qquad h \in \left]0,1\right[\quad .$$

L'ajustement possède la propriété d'additivité si

$$F = h\ \mathrm{Ajus}_C(F, g_{I.}^{(1)}, g_{.J}^{(1)}) + (1-h)\ \mathrm{Ajus}_C(F, g_{I.}^{(2)}, g_{.J}^{(2)}).$$

4.8. Persistance des égalités de profils

Une méthode d'ajustement possède la propriété de persistance des égalités de profils si deux profils égaux avant l'ajustement restent toujours égaux après l'ajustement.

On remarquera immédiatement que la propriété d'agrégation forte implique la propriété d'agrégation faible et que la propriété de persistance des égalités de profils implique la propriété de persistance de l'indépendance. De même, la conservation des éléments nuls implique qu'il n'existe pas de solutions au problème d'ajustement quand le tableau de départ contient des éléments nuls répartis de telle manière qu'aucun tableau possédant les mêmes éléments nuls n'aient les marges fixées.

5. AJUSTEMENTS, CRITERES ET PROPRIETES

5.1. Distance euclidienne

Proposition 5.1.1

L'ajustement avec la distance euclidienne est réversible.

Ceci résulte directement de l'application de la formule d'ajustement (2.1.2).

Proposition 5.1.2

L'ajustement avec la distance euclidienne possède la propriété d'additivité.

Soient les marges $g_I^{(1)}$, $g_J^{(1)}$, $g_I^{(2)}$ et $g_J^{(2)}$ telles que

$$h\ g_{I.}^{(1)} + (1-h)\ g_{I.}^{(2)} = f_{I.}$$

$$h\ g_{.J}^{(2)} + (1-h)\ g_{J.}^{(2)} = f_{.J} \qquad h = [0,1]\ .\ ,$$

Soient les tableaux $G^{(1)}$ et $G^{(2)}$ obtenus en minimisant la distance euclidienne des fréquences sous la contrainte que les marges imposées soient : respectivement $g_{I.}^{(1)}$, $g_{.J}^{(1)}$ et $g_{I.}^{(1)}$, $g_{.J}^{(1)}$. Nous trouvons directement que

$$hg_{ij}^{(1)} + (1-h)g_{ij}^{(2)} = h\left[f_{ij} + \frac{g_{i.}^{(1)} - f_{i.}}{p} + \frac{g_{.j}^{(1)} - f_{.j}}{n}\right] + (1-h)\left[f_{ij} + \frac{g_{i.}^{(2)} - f_{i.}}{p} + \frac{g_{.j}^{(2)} - f_{.j}}{n}\right]$$
$$= f_{ij} \qquad \forall\ i,j.$$

5.2. Distance quadratique pondérée

Proposition 5.2.1

L'ajustement avec la distance quadratique pondérée possède la propriété d'agrégation forte. Soient k et k' deux lignes du tableau F, calculons par (2.2.2) la valeur d'un élément de la ligne obtenue en agrégeant ces deux lignes après l'ajustement :

$$g_{kj} + g_{k'j} = f_{kj} - f_{k.} f_{.j} + g_{k.} g_{.j} + f_{k'j} - f_{k'.} f_{.j} + g_{k'.} g_{.j} \qquad \forall j$$
$$= (f_{kj} + f_{k'j}) - (f_{k.} + f_{k'.}) f_{.j} + (g_{k.} + g_{k'.}) g_{.j} \qquad \forall j,$$

ce qui est bien l'élément général de la ligne obtenue en agrégeant ces deux lignes avant l'ajustement.

Proposition 5.2.2

L'ajustement avec la distance quadratique pondérée possède la propriété d'agrégation faible.

La propriété d'agrégation forte impliquant la propriété d'agrégation faible, cette proposition découle directement de la précédente.

Proposition 5.2.3

La situation d'indépendance persiste après l'ajustement avec la distance quadratique pondérée. En effet, si $f_{ij} = f_{i.} f_{.j} \quad \forall\ i,j$,

$$g_{ij} = f_{i.} f_{.j} - f_{i.} f_{.j} + g_{i.} g_{.j} = g_{i.} g_{.j} \qquad \forall\ i,j\ .$$

Proposition 5.2.4

L'ajustement avec la distance quadratique pondérée est réversible. En effet, si nous réajustons le tableau aux marges du tableau de départ, nous obtenons directement de (2.2.2) :

$$f_{ij} - f_{i.} f_{.j} + g_{i.} g_{.j} - g_{i.} g_{.j} + f_{i.} f_{.j} = f_{ij} \qquad \forall\ i,j\ .$$

5.3. Critère du phi-deux

Proposition 5.3.1

L'ajustement avec le critère du ϕ^2 possède la propriété d'agrégation restreinte.

Supposons que les deux dernières lignes (n - 1) et n du tableau F aient les mêmes profils. Si G est le tableau ajusté en minimisant le ϕ^2, les g_{ij}, λ_i et μ_j sont l'unique solution du système (2.3.2). Calculons d'abord l'élément général de la ligne obtenue en agrégeant ces deux lignes *après* l'ajustement :

$$g_{nj} + g_{(n-1)j} = f_{nj} (\lambda_n + \mu_j) + f_{(n-1)j} (\lambda_{(n-1)} + \mu_j) \qquad \forall\ j = 1, .., p-1$$

$$g_{np} + g_{(n-1)p} = f_{nj}\,\lambda_n + f_{(n-1)p}\,\lambda_{(n-1)} \quad .$$

Comme $\dfrac{f_{nj}}{f_{n.}} = \dfrac{f_{(n-1)j}}{f_{(n-1).}}$, nous avons

$$\left[g_{nj} + g_{(n-1)j}\right] = \left[f_{nj} + f_{(n-1)j}\right]\left[\frac{f_{n.}\lambda_n + f_{(n-1).}\lambda_{(n-1)}}{f_{n.} + f_{(n-1).}} + \mu_j\right] \quad \forall\ j = 1,\ldots,p\text{-}1$$

$$\left[g_{np} + g_{(n-1)p}\right] = \left[f_{np} + f_{(n-1)p}\right]\frac{f_{n.}\lambda_n + f_{(n-1).}\lambda_{(n-1)}}{f_{n.} + f_{(n-1).}} \qquad (5.3.1)$$

Si $G^{\sim}$ est le tableau obtenu en agrégeant les marges avant l'ajustement, les $g_{ij}^{\sim}$ sont l'unique solution du système :

$$\begin{cases} g_{ij}^{\sim} = f_{ij}\left(\lambda_i^{\sim} + \mu_j^{\sim}\right) & \forall j = 1,\ldots,p\text{-}1\ ,\ i = 1,\ldots,n-2 \\ g_{ip}^{\sim} = f_{ip}\,\lambda_i^{\sim} & \forall\ i = 1,\ldots,n-2 \\ g_{(n-1)j}^{\sim} = \left[f_{(n-1)j} + f_{nj}\right]\left[\lambda_{(n-1)}^{\sim} + \mu_j^{\sim}\right] & \forall j = 1,\ldots,p\text{-}1 \\ g_{(n-1)p}^{\sim} = \left[f_{(n-1)p} + f_{np}\right]\lambda_{(n-1)}^{\sim} & \\ \sum_j g_{ij}^{\sim} = g_{i.} & \forall i = 1,\ldots,n-2 \\ \sum_j g_{(n-1)j}^{\sim} = g_{n.} + g_{(n-1).} & \\ \sum_i g_{ij}^{\sim} = g_{.j} & \forall j = 1,\ldots,p\text{-}1. \end{cases} \qquad (5.3.3)$$

En prenant

$$\begin{cases} \mu_j^{\sim} = \mu_j & \forall j = 1,\ldots,p\text{-}1 \\ \lambda_i^{\sim} = \lambda_i & \forall i = 1,\ldots,n-2 \\ \lambda_{(n-1)}^{\sim} = \dfrac{f_{n.}\lambda_n + f_{(n-1)}\lambda_{(n-1)}}{f_{n.} + f_{(n-1)}} & , \end{cases}$$

nous obtenons une solution du système (5.3.2), cette solution est unique pour $G^{\sim}$, et le tableau obtenu en agrégeant les deux dernières lignes avant ou après l'ajustement est identique.

Proposition 3.2

L'ajustement avec le critère du phi-deux conserve les éléments nuls.

La première équation du système (2.3.2), auquel doivent satisfaire les g_{ij}, implique que si un des éléments de F est nul, l'élément correspondant de G est nul également.

Proposition 5.4.2
L'ajustement avec le critère du phi-deux possède la propriété d'additivité.
Soient les marges $g_I^{(1)}$, $g_J^{(1)}$ et $g_I^{(2)}$, $g_J^{(2)}$ telles que

$$\begin{cases} h\, g_I^{(1)} + (1-h)\, g_I^{(2)} = f_I \\ \\ h\, g_{.J}^{(1)} + (1-h)\, g_{.J}^{(2)} = f_{.J}\ . \end{cases} \tag{5.4.1}$$

Les tableaux $G^{(1)}$ et $G^{(2)}$, obtenus en minimisant le phi-deux sous les contraintes que leurs marges soient respectivement $g_{I.}$, $g_{.J}$ et $g_{I.}$, $g_{.J}$, sont les uniques solutions des systèmes,

$$\begin{cases} g_{ij}^{(q)} = f_{ij}\left(\lambda_i^{(q)} + \mu_j^{(q)}\right) & \forall\, i, j = 1, .., p-1 \\ g_{ip}^{(q)} = f_{ip}\ \lambda_i^{(q)} & \forall\, i \\ \sum_j g_{ij}^{(q)} = g_{i.}^{(q)} & \forall\, i \\ \sum_i g_{ij}^{(q)} = g_{.j}^{(q)} & \forall\, j = 1, .., p-1 \end{cases} \qquad q = 1, 2 \tag{5.4.2}$$

Calculons maintenant le tableau $F^{\sim}$ donné par la relation

$$F^{\sim} = G^{(1)}\, h + G^{(2)}\,(1-h)\ .$$

$F^{\sim}$ est une solution du système :

$$\begin{cases} f_{ij}^{\sim} = f_{ij}\left[h\lambda_i^{(1)} + (1-h)\,\lambda_i^{(2)} + h\,\mu_j^{(1)} + (1-h)\,\mu_j^{(2)} \right] & \forall\, i,\ j = 1, ..., p-1 \\ f_{ip}^{\sim} = f_{ip}\left[h\lambda_i^{(1)} + (1-h)\,\lambda_i^{(2)} \right] \quad \forall\, i \\ \sum_i f_{ij}^{\sim} = f_{.j} \qquad \forall\ j = 1, ..., p-1 \\ \sum_j f_{ij}^{\sim} = f_{i.} \qquad \forall\, i\ . \end{cases} \tag{5.4.3}$$

Or, nous savons que l'unique solution de ce système est telle que

$$\begin{cases} \tilde{f_{ij}} = f_{ij} & \forall\, i\,,j \\ h\,\lambda_i^{(1)} + (1-h)\,\lambda_i^{(2)} + h\,\mu_j^{(1)} - (1-h)\,\mu_j^{(2)} = 1 & \forall\ i\,, j = 1, \ldots, p\text{-}1 \\ h\,\lambda_i^{(1)} + (1-h)\,\lambda_i^{(2)} = 1 & \forall\ i\,. \end{cases}$$

5.4. Distance des rapports à l'indépendance

Proposition 5.4.1

L'ajustement avec la distance des rapports à l'indépendance conserve la situation d'indépendance.

En effet, si $f_{ij} = f_{i.}\, f_{.j}\ \forall\, i\,, j$, la distance est nulle si $g_{ij} = g_{i.}\, g_{.j}\ \forall\, i\, j$. La distance étant positive ou nulle, elle réalise dès lors le minimum; or, ce minimum est unique, donc l'unique solution pour G caractérise également la situation d'indépendance.

5.5. Critère du rapport uniforme

Proposition 5.2.1

L'ajustement avec le critère du rapport uniforme conserve les éléments nuls. Ceci découle directement de la première équation du système (2.5.2).

Proposition 5.2.2

L'ajustement avec le critères du rapports uniforme admet la propriété d'additivité. La démonstration se base sur le même raisonnement que dans le cas du critère du phi-deux.

5.6. Information de Kullback

Proposition 5.6.1

L'ajustement avec l'information de Kullback possède la propriété d'agrégation restreinte.

Soient F un tableau dont les deux dernières lignes sont de même profil, et G le tableau obtenu en ajustant, avec le critère d'information, F aux marges $g_{I.}$ et $g_{.J}$. Le tableau G est l'unique solution du système (2.6.2). Calculons l'élément général de la ligne obtenue en agrégeant les deux dernières lignes après l'ajustement;

$$g_{(n-1)j} + g_{nj} = f_{(n-1)j}\ \alpha_{(n-1)}\ \beta_j + f_{nj}\ \alpha_n\ \beta_j \qquad \forall\, j = 1, \ldots, p\text{-}1 \quad (5.6.1)$$

$$g_{(n-1)p} + g_{np} = f_{(n-1)p}\ \alpha_{(n-1)} + f_{np}\ \alpha_n \qquad ,$$

comme les lignes n et (n - 1) ont les même profil :

$$\left[g_{(n-1)j} + g_{nj}\right] = \left[f_{(n-1)j} + f_{nj}\right] \left[\frac{f_{(n-1).}\ \alpha_{(n-1)} + f_{n.}\ \alpha_n}{f_{(n-1).} + f_{n.}}\right] \beta_j \qquad \forall\, j = 1, \ldots, p\text{-}1 \qquad (5.6.2)$$

$$\left[g_{(n-1)p} + g_{np}\right] = \left[f_{(n-1)p} + f_{np}\right] \left[\frac{f_{(n-1).}\ \alpha_{(n-1)} + f_{n.}\ \alpha_n}{f_{(n-1).} + f_{n.}}\right] \quad .$$

Si $G^{\sim}$ est le tableau obtenu en agrégeant les deux dernières lignes avant l'ajustement, les

$g_{ij}^{\sim}$ sont l'unique solution du système :

$$
\left\{
\begin{array}{ll}
\tilde{g}_{ij} = f_{ij}\tilde{\alpha}_i \tilde{\beta}_j & \forall\, j = 1,\ldots,p\text{-}1 \;\; i = 1,\ldots,n-2 \\
\tilde{g}_{ip} = f_{ip}\tilde{\alpha}_i & \forall\;\; i = 1,\ldots,n-2 \\
\tilde{g}_{(n-1)j} = \left[f_{(n-1)j} + f_{nj} \right] \tilde{\alpha}_{(n-1)} \tilde{\beta}_j & \forall\, j = 1,\ldots,p\text{-}1 \\
\tilde{g}_{(n-1)p} = \left[f_{(n-1)p} + f_{np} \right] \tilde{\alpha}_{(n-1)} & \\
\sum\limits_j \tilde{g}_{ij} = g_{i.} \quad \forall\, i = 1,\ldots,n-2 & \\
\sum\limits_j \tilde{g}_{(n-1)j} = g_{n.} + g_{(n-1)\,.} & \\
\sum\limits_i \tilde{g}_{ij} = g_{.j} \quad \forall\;\; j = 1,\ldots,p\text{-}1 &
\end{array}
\right. \qquad (5.6.3)
$$

En prenant $\tilde{\beta}_j = \beta_j \qquad \forall\, j = 1,\ldots,p\text{-}1$

$$\tilde{\alpha}_i = \alpha_i \qquad \forall\, i = 1,\ldots,n-2$$

$$\tilde{\alpha}_{(n-1)} = \frac{f_{n.}\alpha_n + f_{(n-1).}\alpha_{(n-1)}}{f_{n.} + f_{(n-1).}} \quad ,$$

nous obtenons une solution pour le système (5.6.3), cette solution est unique et les tableaux obtenus en agrégeant les deux dernières lignes avant ou après l'ajustement sont identiques.

Proposition 5.6.2

L'ajustement avec le critère d'information conserve la situation d'indépendance.

Si $f_{ij} = f_{i.} f_{.j}$ le tableau G ajusté est l'unique solution du système :

$$
\left\{
\begin{array}{ll}
g_{ij} = \left(f_{i.}\alpha_i \right) \left(f_{.j}\beta_j \right) & \forall\, i, j = 1,\ldots,p\text{-}1 \\
g_{ip} = \left(f_{i.}\alpha_i \right) f_{.p} & \forall\, i \\
\sum\limits_i g_{ij} = g_{i.} & \forall\, i \\
\sum\limits_j g_{ij} = g_{.j} & \forall\, j = 1,\ldots,p\text{-}1 \quad .
\end{array}
\right. \qquad (5.6.4)
$$

En prenant $\alpha_i = \frac{g_{i.}\ g_{.p}}{f_{i.}\ f_{.p}}$ $\forall\, i$

$\beta_j = \frac{f_{.p}\ g_{.j}}{g_{.p}\ f_{.j}}$ $\forall\, j = 1, \ldots, p\text{-}1$ et donc

$g_{ij} = g_{i.} g_{.j}$ $\forall\, i, j$,

nous avons l'unique solution du système (5.6.4) .

Proposition 5.6.3

L'ajustement avec l'information de Kullback est réversible.

Le tableau G qui ajuste le tableau F aux marges $g_{I.}$ et $g_{.J}$ est l'unique solution du système (2.6.2) . Notons F~ le tableau obtenu en ajustant G aux marges de F. Le tableau F~ est l'unique solution du système :

$$\begin{cases} f_{ij}^{\sim} = f_{ij}\alpha_i \beta_j\ \alpha_i^{\sim} \beta_j^{\sim} & \forall\, i, j = 1, \ldots, p\text{-}1 \\ f_{ip}^{\sim} = f_{ip}\alpha_i\ \alpha_i^{\sim} & \forall\, i \\ \sum_i f_{ij}^{\sim} = f_{i.} & \forall\, i \\ \sum_j f_{ij}^{\sim} = f_{.j} & \forall\, j = 1, \ldots, p\text{-}1\ . \end{cases} \tag{5.6.5}$$

En prenant $\alpha_i^{\sim} = \frac{1}{\alpha_i}$ $\forall\, i$, $\beta_i^{\sim} = \frac{1}{\beta_j}$, $\forall\, j$, $f_{ij}^{\sim} = f_{ij}$ $\forall\, i, j$, nous trouvons l'unique solution pour F~ du système (5.6.5) .

Proposition 5.6.4

L'ajustement avec le critère d'information conserve les éléments nuls. Ceci découle directement de la première équation du système (2.6.2).

Proposition 5.6.5

L'ajustement avec le critère d'information conserve les éléments positifs. Cette propriété découle de l'algorithme R.A.S. En effet, à chaque itération, chaque élément du tableau, est multiplié par un nombre positif.

Proposition 5.6.6

L'ajustement avec l'information de Kullback conserve les égalités de profils.

Soient k et k' deux lignes de même profil. Calculons l'élément général du profil de chacune de ces deux lignes après l'ajustement ligne k : $$\frac{g_{kj}}{g_{k.}} = \frac{f_{kj}\,\beta_j}{\sum_{j=1}^{p-1} f_{kj}\,\beta_j + f_{kp}} \qquad \forall\, j = 1,..,p-1$$

$$\frac{g_{kp}}{g_{k.}} = \frac{f_{kp}}{\sum_{j=1}^{p-1} f_{kj}\,\beta_j + f_{kp}}$$

ligne k' : $$\frac{g_{k'j}}{g_{k'.}} = \frac{f_{k'j}\,\beta_j}{\sum_{j=1}^{p-1} f_{k'j}\,\beta_j + f_{k'p}} \qquad \forall\, j = 1,..,p-1 \ .$$

$$\frac{g_{k'p}}{g_{k'.}} = \frac{f_{k'p}}{\sum_{j=1}^{p-1} f_{k'j}\,\beta_j + f_{k'p}} \ .$$

Comme k et k' ont le même profil nous avons

$$\frac{f_{kj}}{f_{k.}} = \frac{f_{k'j}}{f_{k'.}} \qquad \forall\, j \ .$$

Nous obtenons

$$\frac{g_{k'j}}{g_{k'.}} = \frac{\frac{f_{k.}}{f_{k'.}}\, f_{k'j}\ \beta_j}{\frac{f_{k.}}{f_{k'.}} \left[\sum_{j=1}^{p-1} f_{k'j}\,\beta_j + f_{k'p}\right]} \qquad \forall\, j = 1,..,p-1$$

$$= \frac{g_{kj}}{g_{k.}} \qquad \forall\, j = 1,..,p-1$$

et

$$\frac{g_{k'p}}{g_{k'.}} = \frac{\frac{f_{k.}}{f_{k'.}}\, f_{k'p}}{\frac{f_{k.}}{f_{k'.}} \left[\sum_{j=1}^{p-1} f_{k'j}\,\beta_j + f_{k'p}\right]}$$

$$= \frac{g_{kp}}{g_{k.}}$$

5.7. Tableau récapitulatif

Pour montrer qu'une méthode d'ajustement ne possède pas une propriété, il nous a suffi de trouver un contre-exemple; ces contre-exemples sont présentés en annexe. La case qui contient un point d'interrogation indique que nous n'avons pas trouvé de démonstration permettant d'attribuer cette propriété à cette méthode d'ajustement, mais que nous n'avons pas non plus trouvé de contre-exemple.

Critère / Propriétés	dist. eucl.	dist quadr. pond.	phi-deux	dist. rapp. ind.	crit. rapp. unif.	Inf. Kull.
Agrégation forte	non	oui	non	non	non	non
Agrégation restreinte	non	oui	oui	non	non	oui
Persistance de l'indépendance	non	oui	non	oui	non	oui
Réversibilité	oui	oui	non	non	non	oui
Conservation des éléments nuls	non	non	oui	non	oui	oui
Conservation des éléments positifs	non	non	non	non	non	oui
Additivité	oui	non	oui	non	oui	non
Persistance des égalités de profils	non	non	non	?	non	oui

6. CONCLUSION

Aucune des méthodes basées sur le critère quadratique ne conserve les éléments positifs. On pourrait, pour ces méthodes, ajouter une contrainte d'inégalité ($g_{ij} \geq 0 \ \forall \ i,j$) au problème de minimisation, mais les méthodes d'ajustement perdent alors l'essentiel de leurs propriétés. Il est curieux de constater que, quoique l'on associe souvent phi-deux et information, les méthodes d'ajustement issues de ces deux critères possèdent des propriétés fondamentalement différentes. Les résultats obtenus par ces deux méthodes d'ajustement peuvent (voir annexe I.3 ,I.6, VI.3, VI.6) être également totalement différents et ceci particulièrement quand les marges imposées sont elles-mêmes très différentes des marges du tableau de départ. Le tableau récapitulatif nous montre que la méthode issue de l'information de Kullback est celle qui possède le plus de propriétés. Les résultats restent toujours cohérents, car tous les éléments du tableau ajusté restent positifs. De plus, avec cette méthode, les profils égaux restent égaux après l'ajustement ce qui nous semble fondamental car l'analyse d'un tableau de contingence passe toujours par la recherche des différences entre profils. La méthode issue de l'information de Kullback ne possède malheureusement pas toujours de solutions quand certains éléments du tableau de départ sont nuls mais aucune des méthodes admettant la persistance des élément nuls n'échappe à ce problème.

ANNEXE I

PROPRIETE D'AGREGATION FORTE

Tableau de départ

0.10	0.30
0.15	0.20
0.20	0.05

Tableau de départ avec les deux premières lignes agrégées

0.25	0.50
0.20	0.05

Marges imposées

colonne 0.1 0.7 0.2
ligne 0.3 0.7

colonne 0.8 0.2
ligne 0.3 0.7

Résultats obtenus après l'ajustement avec :

1. la distance euclidienne des fréquences (non)

-0.100	0.200	0.200	0.600
0.275	0.425	0.100	0.100
0.125	0.075		

2. la distance quadratique pondérée (oui)

-0.050	0.150	0.153	0.647
0.203	0.497	0.148	0.053
0.148	0.053		

3. le phi-deux (non)

-0.044	0.144	0.165	0.635
0.221	0.479	0.135	0.065
0.123	0.077		

4. la distance des rapports à l'indépendance (non)

0.014	0.086	0.186	0.614
0.174	0.526	0.114	0.086
0.111	0.089		

5. le critère des rapports uniformes (non)

0.000	0.100	0.152	0.648
0.165	0.535	0.148	0.052
0.135	0.065		

6. l'information de Kullback (non)

0.012	0.088	0.165	0.635
0.164	0.536	0.135	0.065
0.124	0.076		

ANNEXE II

PROPRIETE D'AGREGATION FAIBLE

Tableau de départ	Tableau de départ avec les deux premières lignes agrégées
0.10 0.20	0.25 0.50
0.15 0.30	0.20 0.05
0.20 0.05	

Marges imposées	Marges imposées
Colonne 0.3 0.3 0.4	Colonne 0.6 0.4
Ligne 0.5 0.5	Ligne 0.5 0.5

Résultats obtenus après ajustement avec :

1. *la distance euclidienne des fréquences* (non)

0.117	0.183	0.200	0.400
0.092	0.208	0.300	0.100
0.292	0.108		

2. *la distance quadratique pondérée* (oui)

0.115	0.185	0.213	0.388
0.098	0.203	0.287	0.112
0.287	0.113		

3. *le phi-deux* (oui)

0.094	0.206	0.184	0.416
0.090	0.210	0.316	0.084
0.316	0.084		

4. *la distance des rapports à l'indépendance* (non)

0.096	0.204	0.180	0.420
0.096	0.204	0.320	0.080
0.308	0.092		

5. *le critère des rapports uniformes* (non)

0.083	0.217	0.162	0.438
0.081	0.219	0.338	0.062
0.336	0.064		

6. *l'information de Kullback* (oui)

0.093	0.207	0.187	0.413
0.093	0.207	0.313	0.087
0.313	0.087		

On remarquera par la même occasion que pour cet exemple, seules les méthodes 4 et 6 conservent l'égalité des profils.

ANNEXE III

PROPRIETE DE PERSISTANCE DE L'INDEPENDANCE

<u>Tableau de départ</u>

0.05	0.15
0.15	0.45
0.05	0.15

<u>Marges imposées</u>

colonne 0.3 0.3 0.4
ligne 0.5 0.5

<u>Résultats obtenus après ajustement avec</u> :

1. la dist. eucl. fréq.

0.183	0.117	(non)
0.083	0.217	
0.233	0.167	

2. la dist. quadr. pond.

0.150	0.150	(oui)
0.150	0.150	
0.200	0.200	

3. le phi-deux

0.125	0.175	(non)
0.225	0.075	
0.150	0.250	

4. la dist. des rapp. l'ind.

0.150	0.150	(oui)
0.150	0.150	
0.200	0.200	

5. le crit. des rapp. unif.

0.083	0.217	(non)
0.325	-0.025	
0.093	0.307	

6. l'information de Kullback

0.150	0.150	(oui)
0.150	0.150	
0.200	0.200	

ANNEXE IV

PROPRIETE DE REVERSIBILITE DE L'AJUSTEMENT

<u>Tableau de départ</u>

0.10	0.30
0.15	0.20
0.20	0.05

<u>Marges imposées</u>

colonne 0.3 0.3 0.4
ligne 0.5 0.5

<u>Résultats obtenus après ajustement aux marges imposées puis réajustement aux marges initiales avec</u> :

1. la distance euclidienne des fréquences (oui)

0.067	0.233	0.100	0.300
0.142	0.158	0.150	0.200
0.292	0.108	0.200	0.050

2. la distance quadratique pondérée (oui)

0.070	0.230	0.100	0.300
0.143	0.158	0.150	0.200
0.287	0.113	0.200	0.050

3. le phi-deux (non)

0.066	0.234	0.095	0.305
0.119	0.181	0.149	0.201
0.315	0.085	0.206	0.044

4. la distance des rapports à l'indépendance (non)

0.070	0.230	0.108	0.292
0.124	0.176	0.152	0.198
0.306	0.094	0.190	0.060

5. le critère des rapports uniformes (non)

0.068	0.232	0.097	0.303
0.097	0.203	0.143	0.207
0.335	0.065	0.210	0.040

6. l'information de Kullback (oui)

0.069	0.232	0.100	0.300
0.120	0.180	0.150	0.200
0.312	0.088	0.200	0.050

ANNEXE V

PERSISTANCE DES ZEROS

<u>Tableau de départ</u>

0.2	0.2
0.0	0.3
0.1	0.2

<u>Marges imposées</u>

colonne 0.33 0.33 0.34
ligne 0.5 0.5

<u>Résultats obtenus après l'ajustement avec</u> :

1. la dist.eucl. de fréq. (non)

0.232	0.098
0.082	0.248
0.187	0.153

2. la dist. quadr. pond. (non)

0.245	0.085
0.075	0.255
0.180	0.160

3. le chi-deux (oui)

0.298	0.032
0.000	0.330
0.202	0.138

4. la dist. des rapp. l'ind.

0.252	0.078	(non)
0.056	0.274	
0.192	0.148	

5. le crit. des rapp. unif.

0.327	0.003	(oui)
0.000	0.330	
0.173	0.167	

6. l'information de Kullback

0.268	0.062	(oui)
0.000	0.329	
0.232	0.108	

ANNEXE VI
PERSISTANCE DES ELEMENTS POSITIFS

<u>Tableau de départ</u>

0.49	0.01
0.01	0.29
0.19	0.01

<u>Marges imposées</u>

colonne 0.15 0.15 0.8
ligne 0.1 0.9

Résultats obtenus après ajustement avec :

1. la dist.eucl. de fréq.

0.118	0.032	(non)
-0.312	0.362	
0.293	0.507	

2. la dist. quadr. pond.

0.160	-0.010	(non)
-0.192	0.242	
0.132	0.668	

3. le chi-deux

-0.127	0.277	(non)
-0.268	0.318	
0.495	0.305	

4. la dist. des rapp. l'ind.

0.021	0.129	(non)
-0.001	0.051	
0.080	0.720	

5. le crit. des rapp. unif.

-0.140	0.290	(non)
-0.270	0.320	
0.509	0.291	

6. l'information de Kullback

0.030	0.120	(oui)
0.000	0.050	
0.070	0.730	

ANNEXE VII
PROPRIETE D'ADDITIVITE

Tableau de départ

0.10	0.30
0.15	0.20
0.20	0.05

Marge imposée, colonne puis ligne

0.30 0.30 0.40
0.50 0.50

0.10 0.10 0.20
0.10 0.30

0.20 0.20 0.20
0.40 0.20

Résultats obtenus pour les trois ajustements avec :

1. la distance euclidienne des fréquences (oui)

0.067	0.233	-0.017	0.117	0.083	0.117
0.142	0.158	0.013	0.087	0.128	0.072
0.292	0.108	0.103	0.097	0.188	0.012

2. la distance quadratique pondérée (non)

0.070	0.230	-0.007	0.107	0.085	0.115
0.143	0.158	0.022	0.078	0.129	0.071
0.287	0.113	0.085	0.115	0.186	0.014

3. le phi-deux (oui)

0.066	0.234	-0.023	0.123	0.089	0.111
0.119	0.181	-0.012	0.112	0.130	0.070
0.315	0.085	0.135	0.065	0.181	0.019

4. la distance des rapports à l'indépendance (non)

0.070	0.230	0.007	0.093	0.095	0.105
0.124	0.176	0.019	0.081	0.123	0.077
0.306	0.094	0.074	0.126	0.181	0.019

5. le critère des rapports uniformes (oui)

0.068	0.232	-0.020	0.120	0.088	0.112
0.097	0.203	-0.040	0.140	0.137	0.063
0.335	0.065	0.160	0.040	0.175	0.025

6. l'information de Kullback (non)

0.069	0.232	0.006	0.095	0.089	0.110
0.120	0.180	0.012	0.088	0.129	0.071
0.312	0.088	0.083	0.117	0.181	0.019

BIBLIOGRAPHIE SOMMAIRE

[1] BENZECRI, J.-P., BOUGARIT C. et MADRE, J.-L., "Ajustement d'un tableau à des marges d'après la formule de reconstitution", *Les Cahiers de L'Analyse des Données,* 5, 1980, pages 163-172.

[2] BENZECRI, J.P., "Sur une généralisation du problème de l'ajustement d'une mesure à des marges", *Les Cahiers de L'Analyse des Données,* 8, 1983, pages 359-370.

[3] DEMING, W.E. et STEPHAN, F.F., "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known", *The Annals of Mathematical Statistics,* 11, 1940, pages 427-444.

[4] DURIEUX, B. et PAYEN, J.-F., "Ajustement d'une matrice sur des marges : la méthode ASAM", *Annales de l'INSEE,* 22-23, 1976, pages 311-337.

[5] ESCOFFIER, Br., "Traitement des questionnaires avec non-réponses, analyse des correspondances avec marge modifiée et analyse multicanonique avec contrainte", Publication Interne n°146, Institut National de la Recherche en Informatique et en Automatique (Laboratoire de Rennes), Rennes, 1981.

[6] FICHET, B et GBEGAN, A., "Analyse factorielle des correspondances sur des signes présence-absence", *Quatrièmes journées internationales : analyse de données et informatique,* tome 1, INRIA, Versailles, 1985, pages 29-53.

[7] FRIEDLANDER, D., "A technique for estimating a contingency table given the total marginals", *Journal of the Royal Statistical Society,* Series A, 123, 1961, pages 412-420.

[8] FROMENT, R. et LENCLUD, B., "Ajustement de tableaux statistiques", *Annales de l'INSEE,* 22-23, 1976, pages 29-53.

[9] GROSBRAS, J.-M., Méthodes statistiques des sondages, Economica, Paris, 1987.

[10] LEMEL, Y., "Une généralisation de la méthode du quotient pour le redressement des enquêtes par sondages", *Annales de l'INSEE,* 22-23, 1976, pages 273-281.

[11] MADRE, J.-L., "Méthodes d'ajustement d'un tableau à des marges", *Les Cahiers de L'Analyse des Données,* 5, 1980, pages 87-99.

[12] MADRE, J.-L., "Extrapolation des tableaux de la consommation par C.S.P.", *Consommation*, 4, 1982, pages 89-100.

[13] MARCOTORCHINO, F., Présentation des critères d'association en analyse des données qualitatives, Ecole de Commerce et CEME, Séminaire d'Analyse des Données et Processus Stochastiques, Université Libre de Bruxelles, Bruxelles, 1985.

[14] STEMMELEN, E., "Tableaux d'échanges, description et prévision", *Cahier du Bureau Universitaire de Recherche Opérationnelle,* 28, 1977.

[15] STEPHAN, F.F., "An iterative method of adjusting sample frequency tables when expected marginal totals are known", *The Annals of Mathematical Statistics*, 13-2, 1942, pages 168-178.

[16] THIONET, P., "L'ajustement des résultats des sondages sur ceux des dénombrements", *Revue de l'Institut international de Statistique*, 27(1/3), 1959, pages 8-25.

[17] THIONET, P., "Construction et reconstruction de tableaux statistiques", *Annales de l'INSEE,* 22, 1976, pages 5-27.

[18] TILLE, Y., "Une nouvelle méthode d'ajustement d'un tableau à des marges par la formule de reconstitution de l'analyse factorielle des correspondances", *Cahiers du C.E.R.O.*, 30, 1989.