CAHIER DU LAMSADE

Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision
(Université de Paris-Dauphine)
Equipe de Recherche Associée au C.N.R.S. N° 656

LES TABLEAUX A n ENTREES
RAPPEL DE QUELQUES QUESTIONS
ET SOLUTION D'UN PROBLEME DE CARACTERISATION

CAHIER N° 39
mars 1982

V. COHEN

SOMMAIRE

A CHARACTERIZATION OF n-WAY TABLES WHEN SOME MARGINS ARE GIVEN

ABSTRACT

A significant number of statistical studies is devoted to data tables, particularly to the "filling up" of such tables under constraints.

A characterization of n-way tables with fixed marginal totals is given here, in the general case by means of an easy-to-use algorithm (which is described and proved in an annex).

KEY WORDS : Statistical adjustment, n-way tables with fixed marginal totals.

LES TABLEAUX A n ENTREES
RAPPEL DE QUELQUES QUESTIONS
ET SOLUTION D'UN PROBLEME DE CARACTERISATION

RESUME

De nombreuses études statistiques sont consacrées aux tableaux de données et en particulier au problème du "remplissage" de tables sous certaines contraintes.

C'est dans cette perspective qu'est recherchée ici une caractérisation des tableaux à un nombre quelconque d'entrées lorsque certaines de leurs marges sont imposées. Dans le cas général, on aura recours pour cela à un algorithme de mise en oeuvre aisée (qui est décrit et justifié en annexe).

MOTS-CLES : Ajustement statistique, tableaux à n entrées à marges fixées.

INTRODUCTION

L'intérêt porté par les Statisticiens aux tableaux numériques,à deux ou plusieurs entrées, bien qu'ancien (cf.par exemple [4]) ne semble pas fléchir (comme en témoigne ,entre autres, une rencontre prévue à ROME en Juin 1981 sur le thème: "Analysis of Multidimensional Contingency Tables").C'est qu'il s'agit là d'une représentation:

- d'emploi souple:elle convient dans les cas de variables quantitatives aussi bien que qualitatives, dès qu'un nombre fini de classes est constitué;
- de mise en oeuvre simple : il suffit de trier les données;
- de compréhension aisée: les variables qui seules interviennent sont celles qui ont été mesurées (et non ,contrairement à de nombreuses analyses, certaines de leurs transformées).

Cependant, dans la mesure même où cette représentation est peu réductrice, de tels tableaux ne constituent souvent que le premier maillon d'une chaîne de traitements visant à faciliter l'exploration des données recueillies.

On peut distinguer deux classes de problèmes se posant à propos de tableaux numériques:

A.L'analyse de tels tableaux,supposés entièrement remplis, visant:
A.1.soit à tester une hypothèse structurelle sur la population étudiée (par exemple l'indépendance en probabilité des variables observées sur celle-ci);
A.2. soit à une représentation plus "commode" des données (souvent de nature géométrique).

B.Le remplissage de tableaux incomplètement garnis ,sous des contraintes à préciser dans chaque cas.

Toutes ces questions ont donné lieu à une très abondante littérature;ainsi:
.A.1. relève typiquement de la statistique probabiliste:s'y rattachent en particulier les tests du χ^2 dans les tables de contingence ,en vue d'éprouver la conformité à un "modèle" de référence;citons en référence [10,14] parmi d'autres ,fort nombreuses.Notons également ,dans cette sous-classe,

une question du type: tel tableau est-il symétrique, à des écarts aléatoires près ?[3]

. A.2. couvre la majeure partie des procédures d'analyse des données (analyse en composantes principales,des correspondances, des proximités etc)[2].

C'est à la classe B que la présente étude est consacrée ,venant s'ajouter à de très nombreuses autres citées en référence et dont certaines ,telle que [15] ,comportent une abondante bibliographie.Après avoir mentionné quelques problèmes concrets qui s'y rattachent, nous présenterons essentiellement une caractérisation des tableaux à n entrées (c'est-à-dire à n variables observées) et dont certaines marges sont fixées ,ce terme de "marge" prenant un sens que nous préciserons dans le cas où $n \geqslant 2$.

1. Quelques problèmes de remplissage de tableaux à marges fixées

1.1. Un problème classique de transport

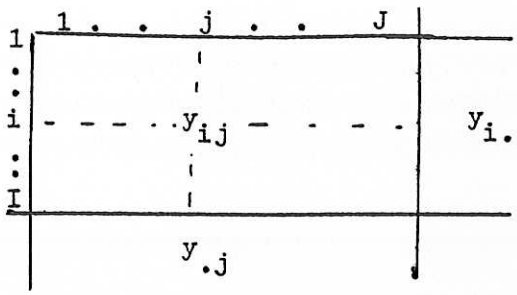
Un produit (charbon,céréale ou autres) est distribué à J "clients" indicés par j (j=1,...,J) à partir de I dépôts indicés par i (i=1,...,I);chaque dépôt i contient une quantité a_i du produit à distribuer;chaque client j a commandé une quantité b_j de ce produit et pourra être servi à partir d'un ou de plusieurs dépôts.Il s'agit de constituer le tableau $Y=(y_{ij})$ des quantités à livrer à partir des différents dépôts de manière ,si possible,à satisfaire les commandes tout en minimisant le coût global du transport.Cette dernière fonction de coût $f(y_{11},\ldots,y_{ij},\ldots,y_{IJ})$ étant supposée connue,on voit qu'il s'agit ici de rechercher:

$$\begin{cases} \text{MIN } f(y_{11},\ldots,y_{ij},\ldots,y_{IJ}) \\ \text{sous les contraintes:} \\ (\forall\ i,j)\ ,\quad y_{ij} \geqslant 0 \\ (\forall\ i\ )\ ,\quad \sum_j y_{ij} \leqslant a_i \\ (\forall\ j\ )\ ,\quad \sum_i y_{ij} = b_j \end{cases}$$

Rappelons que même si la fonction f est supposée (abusivement) linéaire, le problème de programmation ainsi posé peut être de résolution longue en raison du nombre IJ ,éventuellement très élevé, d'inconnues.

1.2. Une étude de "mobilité sociale"

Comment évolue la répartition en CSP (Catégories Socio-Professionnelles) des individus d'un pays au cours des générations successives ? Il conviendrait de dresser le tableau à double entrée suivant:



y_{ij} désignant le nombre d'individus appartenant à la CSP j tandis que leurs parents appartenaient à la CSP i .

Si les marges ($y_{i.}$,i=1...I; $y_{.j}$,j=1...J) d'un tel tableau peuvent être (relativement) bien connues grâce aux recensements effectués au cours des générations successives, les éléments y_{ij} seront en général mal connus (et éventuellement approchés par sondages)d'où le problème ainsi posé du remplissage d'un tableau "équilibré" (c'est-à-dire dont les marges données a priori sont bien égales aux sommes correspondantes de lignes et de colonnes) tenant compte des différentes sources d'information disponibles.

1.3.Une question de Comptabilité Nationale:les tableaux d'échanges

L'économie d'un pays étant par exemple divisée en "branches",il importe de connaître les échanges s'effectuant entre elles:ici encore,si l'on connaît relativement bien le total des sorties ("output") pour chaque branche ,ou encore celui des entrées ("input"),-donc les marges d'un tableau d'échanges, ses éléments mêmes ne seront que très approximativement évalués:d'où un problème de construction de tableau en tenant compte des informations détenues (essentiellement sur ses marges et ,approximativement, sur ses éléments).

1.4. Une interrogation post-électorale: les reports de voix

Prenons l'exemple de l'élection présidentielle française à deux tours:au premier tour,les suffrages exprimés se répartissent sur I candidats (en 1981, I=10),puis au deuxième tour, sur J=2 candidats seulement.Si l'on connaît bien ,après dépouillement, ces deux répartitions,on ignore,en principe, comment se sont effectués les reports (combien d'électeurs ayant voté pour le candidat i au premier tour ont voté pour j au second?)Cette question revient à s'interroger sur les éléments d'un tableau de format IxJ dont les marges sont bien connues alors que ceux-ci ne sont qu'estimés par des enquêtes par sondage complétées d'analyses politiques.

Les différents problèmes présentés ici concernent le "chercheur opérationnel" intéressé par le transport et la logistique (1.1),le sociologue (1.2), l'économiste (1.3),le politologue (1.4).Cette question a préoccupé également des mathématiciens attirés par la statistique.

1.5. Le problème de FRECHET

Sous ce terme l'on entend l'étude des tableaux (en particulier de corrélation) dont les marges sont fixées et soumis éventuellement à des contraintes supplémentaires (par exemple à éléments compris entre des bornes données), ce sujet ayant donné lieu à plusieurs publications de M.FRECHET s'échelonnant sur une dizaine d'années.Elles établissent en particulier des relations dites "conditions de FRECHET",certaines très naturelles,d'autres qui le sont moins et de démonstration plus difficile:si l'on envisage les tableaux $Y=(y_{ij})$ de format IxJ , à éléments entiers naturels et de marges notées $y_{i.}$ et $y_{.j}$, alors:

1)Si les marges $y_{i.}$ et $y_{.j}$ (pour i=1...I et j=1...J) sont des entiers naturels arbitraires soumis cependant à la condition:

$$\sum_{i=1}^{I} y_{i.} = \sum_{j=1}^{J} y_{.j} \quad (= y_{..}) \qquad (1)$$

il existera au moins un tableau Y d'entiers naturels admettant pour marges les entiers choisis a priori [6,7].

2)Si l'on impose des majorants m_{ij} tels que:

$$(\forall\, i,j)\ ,\quad y_{ij} \leq m_{ij}$$

des conditions nécessaires et suffisantes d'existence d'un tel tableau sont:

$$\forall\ A \subset \{1,\ldots,i,\ldots,I\}\quad ,\ \forall\ B \subset \{1,\ldots,j,\ldots,J\}$$

$$\sum_{i \in A} y_{i.} + \sum_{j \in B} y_{.j} - y_{..} \leq \sum_{j \in B} \sum_{i \in A} m_{ij}$$

Connues sous le nom de "conditions (C) de FRECHET",elles imposent en général la vérification d'un assez grand nombre d'inégalités;ainsi [8] est consacré à la détermination de toutes les solutions (entières) qui existent dans ce cas.

1.6. Evaluation de la probabilité d'événements "rares"

Nous allons aborder un problème se rattachant directement au précédent et concernant l'évaluation de la probabilité d'événements rarement observés, mais auxquels on s'intéresse en raison des conséquences graves de leur réalisation éventuelle.

Plus précisément,supposons qu'un événement e_{ij} se réalise si et seulement si un événement a_i et un événement b_j sont conjointement réalisés.Il peut se faire qu'à partir d'observations disponibles on soit en mesure d'estimer convenablement les probabilités respectives de réalisation de a_i et de b_j, soit $p_1(a_i)$ et $p_2(b_j)$. On ne sera pas pour autant en mesure d'estimer correctement la probabilité de réalisation de e_{ij} ,soit $p(e_{ij})$.Certes,on peut affirmer que:

$$p(e_{ij}) \leq \min(\, p_1(a_i),\, p_2(b_j)\,)$$

Mais le second membre peut apparaître comme un majorant trop large de $p(e_{ij})$. Une autre hypothèse adoptée, à titre de simplification (abusive), est parfois celle de l'indépendance stochastique:elle conduit à estimer $p(e_{ij})$ par:

$$\tilde{p}(e_{ij}) = p_1(a_i)\ p_2(b_j)$$

ce qui risque de sous-évaluer gravement $p(e_{ij})$.(Si $p_1(a_i)$ et $p_2(b_j)$ sont faibles,on voit même parfois négliger ,selon cette logique , des termes tels que $p(e_{ij})$ comme étant du "second ordre"). En fait,ce problème conduit à envisager les probabilités sur un espace produit compatibles avec des probabilités marginales connues,ou du moins supposées convenablement estimées.Donc,dans le cas d'ensembles finis d'événements, nous serons amenés,comme dans le problème de FRECHET, à essayer de caractériser les tableaux admettant des marges connues.

2. Caractérisation des tableaux à n entrées dont les marges sont fixées

Avant d'aborder ce problème,nous passerons rapidement en revue les méthodes de remplissage de tableaux à marges fixées ,en nous limitant au cas n=2 qui est celui habituellement traité.

2.1. Construction de tableaux à marges fixées pour n=2

2.1.1. Pour les problèmes du type (1.1.) ,les éléments du tableau sont choisis de façon à minimiser (ou maximiser) une certaine fonction (économique) sous contraintes de marges.Dans le cas le plus simple,il en résulte un programme linéaire traité par un algorithme donnant la solution exacte (type "simplexe") ,ou,s'il y a trop d'inconnues, améliorant une solution de départ par une procédure de transfert circulaire (STEPPING STONE) ,d'ailleurs interrompue dès que le gain procuré à chaque étape n'en semble plus justifier la poursuite.

2.1.2. Pour les problèmes (1.2),(1.3),(1.4),une première méthode ("probabiliste") consiste à choisir une forme de distribution bidimensionnelle admettant des marges fixées et dépendant de paramètres qu'il faudra estimer :c'est la procédure utilisée dans [9] et [15] en s'appuyant sur la loi de GUMBEL :

$$H(x,y) = F(x)\ G(y)\ (1 + \alpha\ (1-F(x))\ (1- G(y))\ )$$

où F(x) et G(y) désignent les lois de distribution des marges et α un paramètre à estimer.

Une autre méthode consiste à adopter un tableau de référence ayant des

marges quelconques et d'essayer de s'en rapprocher (en un sens à préciser) sous les contraintes de marges fixées.Selon la "distance" choisie,cette méthode sera plus ou moins commode et conduira à une solution explicite ou à un algorithme de résolution;ajoutons que la "distance"en question n'est pas toujours explicitement donnée ,alors même que la procédure proposée de construction du tableau correspond en fait implicitement à un tel choix:c'est le cas ,comme il est dit dans (15), de la méthode classique RAS et de l'information de KULLBACK.

2.2. Détermination de l'ensemble des tableaux admettant des marges fixées

Comme il n'est pas question,sauf dans des cas particuliers,de construire effectivement tous les tableaux admettant des marges fixées,il convient en fait de définir une procédure de caractérisation .
Celle que nous choisissons a l'avantage de permettre une résolution commode des problèmes rappelés en 2.1.2.,pour une classe de distances assez générale et surtout dans le cas d'un nombre quelconque n d'entrées.
Il sera établi que tout tableau Y à n entrées à marges fixées pourra s'écrire sous la forme d'une somme:

$$Y = Y^{\circ} + Z$$

où:

. Y° et Z seront de même format que Y;

. Y° sera un certain tableau ayant les marges imposées à Y et qui sera déterminé explicitement;

. Z sera un tableau dont toutes les marges imposées à Y seront astreintes à être nulles.

Contrairement à Y°, le tableau Z ne sera pas unique et comportera même une large part d'arbitraire ,restreinte seulement par la contrainte que nous venons d'indiquer et éventuellement d'autres, découlant par exemple (c'est souvent le cas) du caractère positif que l'on imposerait aux éléments de Y.

L'intérêt de cette décomposition tiendra en particulier au fait que cet Y° sera "orthogonal" à tous les Z en ce sens que:

$$Y^\circ \otimes Z = \sum_k y^\circ_k \, z_k = 0$$

(la sommation étant étendue à tous les éléments de Y° et Z

2.2.1. Remarque importante :qu'est-ce qu'une "marge"?
Dans le cas d'un tableau à 2 entrées,il s'agit de l'un ou l'autre des ensembles de valeurs obtenues par sommation sur l'indice de ligne ou de colonne, soit pour $Y = (y_{ij})$:

$$\left\{ y_{i.} \;\middle|\; y_{i.} = \sum_{j=1}^{J} y_{ij} \;\; ; \; i=1\ldots I \right\}$$

$$\left\{ y_{.j} \;\middle|\; y_{.j} = \sum_{i=1}^{I} y_{ij} \;\; ; j=1\ldots J \right\}$$

Dans le cas du "parallélotope" correspondant à n=3, les marges pourront désigner aussi bien les faces (obtenues par sommation sur l'un des 3 indices) que les arêtes (obtenues par sommation sur 2 des 3 indices). Plus généralement,pour un tableau à n entrées,les marges seront des "facettes" ayant elles-mêmes la forme de tableaux à n-k entrées si les sommations ont été effectuées sur k ($1 \leq k \leq n$) indices.

2.2.2. Interprétation géométrique

On pourrait interpréter un tableau numérique de format IxJ comme I vecteurs de $\mathbb{R}^J$.Mais cette interprétation ne se prête guère à une généralisation au cas de plus de 2 entrées.C'est pourquoi nous considèrerons plutôt qu'un tableau Y est un vecteur $\vec{Y}$ de l'espace vectoriel euclidien de dimension égale au nombre de ses éléments,soit IJ.Imposer des marges, c'est imposer la projection $\vec{Y^\circ}$ de $\vec{Y}$ sur un certain sous-espace vectoriel H :celui-ci est engendré par les I+J vecteurs suivants :

exprimés dans la base canonique($(e_{ij}$;i=1...I;j=1...J) où chaque e_{ij} est constitué de I suites mises bout à bout et comportant chacune J éléments tous nuls à l'exception du jème élément de la ième suite auquel est attribué la valeur 1), I de ces vecteurs auront la forme:

$$V_1 = \frac{1}{\sqrt{J}} \; (\underbrace{1\,1\,\ldots\,1}_{J \text{ termes}} \underbrace{0\,0\,\ldots\,0}_{(I-1)J \text{ termes}})$$

$$V_2 = \frac{1}{\sqrt{J}} \; (\underbrace{0\,0\,\ldots\,0}_{J \text{ termes}} \underbrace{1\,1\,\ldots\,1}_{J \text{ termes}} \underbrace{0\,0\,\ldots\,0}_{(I-2)J \text{ termes}})$$

$$\vdots$$

$$V_I = \frac{1}{\sqrt{J}} \; (\underbrace{0\,0\,\ldots\,0}_{(I-1)J \text{ termes}} \underbrace{1\,1\,\ldots\,1}_{J \text{ termes}})$$

tandis que les J suivants seront:

$$W_1 = \frac{1}{\sqrt{I}} \; \underbrace{(\underbrace{1\,0\,0\,\ldots\,0}_{J}\,\underbrace{1\,0\,0\,\ldots\,0}_{J}\,\underbrace{1\,0\,0\,\ldots\,0}_{J}\,\underbrace{1\,0\,\ldots\,0}_{J})}_{IJ \text{ termes}}$$

$$W_2 = \frac{1}{\sqrt{I}} \; (\underbrace{0\,1\,0\,\ldots\,0}_{J \text{ termes}}\,\underbrace{0\,1\,0\,\ldots\,0}_{J \text{ termes}}\,0\,1\,0\ldots\,0\,\underbrace{0\,1\,0\ldots 0}_{J \text{ termes}})$$

$$\vdots$$

$$W_J = \frac{1}{\sqrt{I}} \; (\underbrace{0\,0\,\ldots\,1}\,\underbrace{0\,0\,\ldots\,0\,1}\,\underbrace{0\,0\,\ldots\,0\,1}\,\underbrace{0\,\ldots\,0\,1})$$

Notons que ces I+J vecteurs n'engendrent qu'un espace H de dimension I+J-1.En effet,soit une combinaison linéaire

$$L = \sum_{i=1}^{I} \lambda_i \, V_i + \sum_{j=1}^{J} \mu_j \, W_j \qquad (\lambda_i, \mu_j \in \mathbb{R})$$

Alors: $\{L = 0\} \Leftrightarrow (\forall i, \forall j) \; \frac{\lambda_i}{\sqrt{J}} + \frac{\mu_j}{\sqrt{I}} = 0$

D'où : $(\forall i), \; \lambda_i = K\sqrt{J}$; $(\forall j) \; \mu_j = -K\sqrt{I}$; K constante arbitraire.

Donc ,à un facteur près,une et une seule combinaison linéaire est nulle.

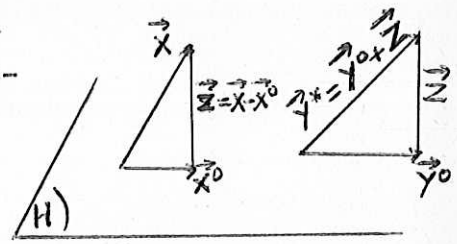
La décomposition: $Y = Y^o + Z$

correspond alors simplement à la décomposition vectorielle:

$\vec{Y} = \vec{Y^o} + \vec{Z}$ où $\vec{Y^o} \in H$ et $\vec{Z} \in H^{\perp}$ (supplémentaire orthogonal de H).

2.2.3. Conséquence: solution d'un problème d'ajustement

Pour construire un tableau Y de marges fixées et approchant au mieux un tableau donné X au sens des moindres carrés, il suffira de considérer leurs projections orthogonales sur le sous-espace H, donc de déterminer (comme il sera précisé plus loin) les tableaux Y^o et X^o respectivement associés à Y^* (ou plutôt à ses marges connues) et à X.



Le tableau Y^* cherché sera alors simplement :

$$Y^* = Y^o + (X - X^o)$$

le vecteur $\vec{X} - \vec{X^o}$ (ou tableau $X - X^o$) jouant le rôle du vecteur $\vec{Z}$ (élément du supplémentaire $H^{\perp}$ de H dans l'espace $\mathbb{R}^{IJ}$) le mieux adapté dans ce problème d'ajustement:

parmi les tableaux ayant les marges que l'on s'est fixées, Y^* est ainsi celui qui approche X au mieux (au sens des moindres carrés; une extension au cas d'une distance quadratique quelconque étant d'ailleurs possible en utilisant une procédure analogue). A l'optimum :

$$\sum_{i,j} (y^*_{ij} - x_{ij})^2 = \sum_{i,j} (y^o_{ij} - x^o_{ij})^2$$

2.2.4. Cas particulier du tableau à 2 entrées

Les considérations géométriques précédentes conduisent directement à la décomposition suivante:

$$Y = Y^o + Z$$

avec : $Y^o = \left(y^o_{ij} = \frac{y_{i.}}{J} + \frac{y_{.j}}{I} - \frac{y_{..}}{IJ} \right)$ (2.2.4.)

tandis que Z est un tableau à marges nulles de même format que Y.

Cette formule (2.2.4.) peut s'établir comme suit:

si nous cherchons Y° sous la forme:

$$Y^\circ = \sum_{i=1}^{I} a_i V_i + \sum_{j=1}^{J} b_j W_j$$

comme :

$$(\forall\, i \in \{1,\ldots,I\}),\ (Y - Y^\circ).\ V_i = 0$$

et

$$(\forall\, j \in \{1,\ldots,J\}),\ (Y - Y^\circ).\ W_j = 0$$

il apparaît que:

$$\frac{1}{\sqrt{J}}\ y_{i.} = a_i + \frac{1}{\sqrt{IJ}} \sum_{j=1}^{J} b_j \qquad (i=1,\ldots,I)$$

et:

$$\frac{1}{\sqrt{I}}\ y_{.j} = b_j + \frac{1}{\sqrt{IJ}} \sum_{i=1}^{I} a_i \qquad (j=1,\ldots,J)$$

Il en résulte que:

$$\frac{1}{\sqrt{J}} \sum_{j=1}^{J} b_j + \frac{1}{\sqrt{I}} \sum_{i=1}^{I} a_i = \frac{1}{\sqrt{IJ}}\ y_{..}$$

Remplaçant les V_i et les W_j en fonction des vecteurs de base e_{ij} ,

il vient:

$$Y^\circ = \frac{1}{\sqrt{J}} \sum_{i,j} a_i\, e_{ij} + \frac{1}{\sqrt{I}} \sum_{i,j} b_j\, e_{ij}$$

et par suite,utilisant les relations ci-dessus liant les a_i et les b_j :

$$Y^\circ = (\ y^\circ_{ij} = y_{i.} / J + y_{.j} / I - y_{..} / IJ\ ).$$

Application numérique : cherchons le tableau Y^* de format (2x3) ,admettant les mêmes marges que le tableau Y ci-dessous:

$Y =$

7	1	1	9
1	5	9	15
8	6	10	24

et le plus proche (au sens quadratique) de:

$X =$

2	1	3	6
6	5	1	12
8	6	4	18

Solution :on calcule d'abord Y° par application de la formule précédente:

$Y^\circ =$

3	2	4	9
5	4	6	15
8	6	10	24

et semblablement:

$X^\circ =$

3	2	1	6
5	4	3	12
8	6	4	18

Il en résulte immédiatement(cf. 2.2.3.):

$Y^* = Y^\circ + (X - X^\circ) =$

3	2	4	9
5	4	6	15
8	6	10	24

$+$

-1	-1	2	0
1	1	-2	0
0	0	0	0

$=$

2	1	6	9
6	5	4	15
8	6	10	24

Notons que le tableau ainsi ajusté ne sera pas forcément à éléments tous positifs ,ce qui,dans certains cas ,est fort regrettable:une contrainte supplémentaire de ce type conduira en général ,si la solution précédente ne la vérifie pas spontanément, à un programme quadratique:on ajoute en effet à Y^* un tableau $T = (t_{ij})$ de même format réalisant:

$$\begin{cases} \text{MIN} \sum_{i,j} t_{ij}^2 \\ \text{sous les contraintes:} \\ (\forall i, \forall j),\ y^*_{ij} + t_{ij} \geqslant 0 \\ (\forall i),\ \sum_j t_{ij} = 0 \ ;\ (\forall j),\ \sum_i t_{ij} = 0 \end{cases}$$

Remarquons que notre problème d'optimisation se décompose car le tableau T est cherché dans le sous-espace H^1,donc orthogonal à $(Y^\circ - X^\circ)$.

Ainsi,reprenant les données précédentes,il apparaît que le tableau X^* admettant les mêmes marges que X et le plus proche possible de Y est:

$$X^* = \begin{array}{|c|c|c|} \hline 7 & 1 & -2 \\ \hline 1 & 5 & 6 \\ \hline \end{array}$$

qui a donc un élément négatif. On établit que le tableau T^* solution du programme précédent est:

$$T^* = \begin{array}{|c|c|c|c|} \hline -1 & -1 & 2 & 0 \\ \hline 1 & 1 & -2 & 0 \\ \hline 0 & 0 & 0 & 0 \\ \hline \end{array}$$

Le tableau à éléments positifs le plus proche de Y et de mêmes marges que X est ainsi:

$$\hat{X} = \begin{array}{|c|c|c|c|} \hline 6 & 0 & 0 & 6 \\ \hline 2 & 6 & 4 & 12 \\ \hline 8 & 6 & 4 & 18 \\ \hline \end{array}$$

et il est clair que dans ce cas:

$$d(Y^*, X) \neq d(\hat{X}, Y)$$

(d désignant la distance euclidienne canonique).

2.2.5. Extensions

Il est tentant de chercher à étendre la formule (2.2.4.) au cas de 3 puis de n entrées.

2.2.5.1. Cas de 3 entrées

Si $Y = (y_{ijk} \quad / i=1...I;j=1...J;k=1...K)$

et si toutes les faces sont imposées,alors la décomposition

$$Y = Y^\circ + Z \qquad (Z \text{ étant à faces toutes nulles})$$

fait intervenir:

$$Y^\circ = (y^\circ_{ijk} = \frac{y_{.jk}}{I} + \frac{y_{i.k}}{J} + \frac{y_{ij.}}{K} - \frac{y_{i..}}{JK} - \frac{y_{.j.}}{IK} - \frac{y_{..k}}{IJ} + \frac{y_{...}}{IJK}) \quad (2.2.5.1)$$

Si ce sont les arêtes qui sont imposées, la décomposition $Y=Y^{\circ\circ} + Z$ (avec Z à arêtes nulles) fera intervenir:

$$Y^{\circ\circ} = (y^{\circ\circ}_{ijk} = \frac{y_{i..}}{JK} + \frac{y_{.j.}}{IK} + \frac{y_{..k}}{IJ} - \frac{2\,y_{...}}{IJK})$$

2.2.5.2.Cas général de n entrées

Il convient d'abord d'adopter des notations commodes:le terme général du tableau Y sera noté:

$$y_{r_1 \ldots r_i \ldots r_n} \qquad \text{avec} \quad r_i = 1 \ldots R_i \quad \text{(càd:la ième variable a } R_i \text{ modalité}$$

Posons: $I = \{ 1,\ldots,i,..,n \}$.
Dire que les faces de Y sont imposées revient à fixer toutes les valeurs obtenues par sommation sur un seul indice.Dans ces conditions, la formule (2.2.5.1.) ci-dessus s'étend en:

$$y^{\circ}_{r_1 \ldots r_n} = - \sum_{\substack{e \in \mathcal{P}(I) \\ e \neq I}} (-1)^{\mathrm{Card}(I \setminus e)} \frac{y_{r_e}}{\prod_{k \in I \setminus e} R_k} \qquad (2.2.5.2)$$

où y_{r_e} désigne le terme général de la "facette" obtenue par sommation par rapport à tous les indices r_i tels que $i \in I \setminus e$ (c'est-à-dire encore que $y_{r_e} = y_{\ldots r_i \ldots}$ où i décrit la partie e de I).

Démonstration

Le tableau Y° ainsi constitué sera parfaitement déterminé à partir des faces de Y ,puisque la partie $e = I$ est exclue de la sommation.D'autre part,ses faces seront précisément égales à celles de Y.En effet,soit:

$$y^{\circ}_{r_1 \ldots r_{i-1} \bullet r_{i+1} \ldots r_n} = y^{\circ}_{r_{(I \setminus i)}}$$

le terme général de la face obtenue par sommation sur l'indice r_i . Alors:

$$y^{\circ}_{r_{(I \setminus i)}} = - \sum_{r_i=1}^{R_i} \left(\sum_{\substack{e \in \mathcal{P}(I) \\ e \neq I}} (-1)^{\mathrm{Card}(I \setminus e)} \frac{y_{r_e}}{\prod_{k \in I \setminus e} R_k} \right) \qquad (2.2.5.3)$$

Dans la parenthèse du second membre, distinguons, dans l'ensemble des e intervenant dans la sommation:

. le sous-ensemble E_i des e, parties de I telles que $i \in e$;

. du sous-ensemble $\bar{E}_i$ (complémentaire de E_i) ayant pour éléments les e, parties de I telles que $i \notin e$.

Exceptons d'abord le terme $e_o = I \setminus i$, élément de $\bar{E}_i$; on peut alors associer à tout élément $a \in \bar{E}_i \setminus e_o$ l'élément $a' = a \cup i \in E_i$: nous excluons l'élément e_o car $e_o \cup i = I$; or e=I est exclu de la sommation dans 2.2.5.2.

En groupant les termes correspondant aux couples (a,a') ainsi formés, nous constatons qu'ils s'annulent car ils sont opposés (en effet: $(-1)^{Card(I \setminus a)} = - (-1)^{Card(I \setminus a')}$) .Reste seulement le terme correspondant à $e_o = I \setminus i$ .Donc:

$$y^{\circ}_{r_1 \cdots r_{i-1} \bullet r_{i+1} \cdots r_n} = - \sum_{r_i=1}^{R_i} (-1)^{Card \{i\}} \frac{y_{r_1 \cdots r_{i-1} \bullet r_{i+1} \cdots r_n}}{R_i}$$

$$= y_{r_1 \cdots r_{i-1} \bullet r_{i+1} \cdots r_n}$$

■

Généralisation: fixation de "marges" quelconques pour un tableau à n entrées

Si ce sont des "facettes", c'est-à-dire des distributions marginales quelconques, qui sont imposées à un tableau Y à n entrées, la caractérisation sous forme:

$$Y = Y^{\circ} + Z$$

exige, pour déterminer Y°, la mise en oeuvre d'un algorithme donné (ainsi que sa justification) en ANNEXE .

Remarques générales et conclusion

1.Un problème de "boîte noire"

Notre tableau à n entrées peut être comparé à une boîte opaque et que l'on ne peut ouvrir ,mais à la"surface" extérieure de laquelle il est permis d'effectuer des mesures (mesures sur des arêtes (dimension 1), des faces (dimension n-1), ou plus généralement des facettes(dimension n-k)).La caractérisation proposée conduit à considérer que la boîte noire est la somme d'une boîte parfaitement connue en fonction des mesures effectuées en surface et d'une boîte soumise seulement à des conditions liées à la nature des facettes sur lesquelles les mesures ont été effectuées ,mais non aux valeurs trouvées sur celles-ci (ce que l'on peut aussi rapprocher -mais uniquement à titre d'image-de l'expression de la solution générale d'une équation différentielle linéaire avec second membre,comme somme d'une solution particulière et de la solution générale de l'équation sans second membre associée).

2.Cette décomposition en somme facilite ,on l'a vu,certaines études;mentionnons,en plus des problèmes d'ajustement,quelques autres exemples:

. Si les n variables qui interviennent dans la construction du tableau sont des variables aléatoires ,elle permettra d'exprimer la fonction caractéristique du tableau en fonction de celles de ses marges;

.si l'on dispose d'un échantillon décrit par un tableau ,mais qui n'est pas représentatif de la population sur n variables de contrôle,la décomposition proposée permet l'extraction d'un sous-échantillon redressé par rapport à celles-ci et de taille maximale;

. si l'on arrondit les marges d'un tableau équilibré (par exemple à des entiers voisins), elle facilite l'arrondissage des éléments du tableau de façon qu'il reste équilibré.

En tant que premier instrument de représentation en analyse multidimensionnelle,le tableau à n entrées mérite sans doute une attention particulière.On peut parfois y résoudre certains problèmes (moins inextricables qu'il ne paraît de prime abord)avant de songer à en obtenir des vues simplifiées mais dont l'interprétation reste parfois délicate.

ANNEXE

UNE ETUDE DES TABLEAUX A n ENTREES

Soit un tableau à n entrées $Y = (y_{r_1 \ldots r_i \ldots r_n})$ où $r_i = 1 \ldots R_i$, c'est-à-dire que la ième variable a R_i modalités.
Nous poserons: $I = \{1, \ldots, i, \ldots, n\}$
et noterons $\mathcal{E}(I)$ le treillis constitué par l'ensemble des parties de I ordonné par la relation d'inclusion.
La sommation sur un indice r_i donne une "face" du tableau Y, elle-même tableau à n-1 entrées ;la sommation sur n-1 indices conduit à une "arête" de Y.
Plus généralement,la sommation sur k indices conduit à une "facette" d'ordre n-k ou tableau à n-k entrées(celles par rapport auxquelles la sommation n'a pas été effectuée).
Une facette sera désignée par (e) si e est une partie de I et que la sommation a été effectuée par rapport à tous les indices constituant $I \setminus e$ .Son terme général sera noté:

y_{r_e} pour représenter $y_{\ldots r_i \ldots}$ où i décrit e .

Supposons que soit imposé un ensemble $L = \{L_1, \ldots, L_j, \ldots, L_J\}$ de J facettes, chaque facette (L_j) correspondant à une partie L_j de I.
Il est clair que la connaissance d'une facette (L_j) entraîne celle de toutes les facettes (h) telles que $h \subset L_j$,de sorte qu'imposer l'ensemble (L) de facettes revient à imposer toutes les facettes associées à l'ensemble:

$$E = \bigcup_{j=1}^{J} \mathcal{P}(L_j)$$

E ,élément de $\mathcal{P}(I)$, est,dans le treillis $\mathcal{E}(I)$, la partie héréditaire engendrée par L.

Objet de l'étude : constituer un tableau Y° ,de même format que Y,qui soit fonction linéaire des facettes (e /e $\in$ E) de Y, soit:

$$y^{\circ}_{r_1 \ldots r_i \ldots r_n} = \sum_{e \in E} A_e \; y_{r_e} \qquad (A_e \text{ constantes} \in \mathbb{R})$$

et qui admette les facettes imposées: $\forall L_j \in L, \; y^{\circ}_{r_{L_j}} = y_{r_{L_j}}$.

Il se pose un problème d'existence ,d'unicité et de détermination de l'ensemble des coefficients $\{A_e / e \in E\}$.

Imposer une facette (L_j) conduit à identifier $y_{r_{L_j}}$ à T_j où:

$$T_j = \sum_{i \notin L_j} \sum_{r_i=1}^{R_i} \sum_{e \in E} A_e \; y_{r_e} \qquad (1)$$

Cette identification conduit,pour chaque L_j , à un système (S_j) d'équations en A_e .

Etude du système d'équations associé à une facette imposée (L_j)

Cette étude passe par l'établissement de cinq lemmes simples.

Lemme 1

Après sommation, T_j comporte tous les termes y_{r_e} tels que $e \subset L_j$ et ceux-là seulement.

En effet: $e \subset L_j \Rightarrow e \in E$ puisque $\mathcal{P}(L_j) \subset E$ .Comme la sommation est effectuée pour $i \notin L_j$,un tel y_{r_e} subsiste dans T_j .Si $e \cap L_j = e' \neq e$, cette sommation réduit y_{r_e} à $y_{r_{e'}}$.

Conséquence :lemme 2

Chaque système (S_j) comporte $2^{\text{Card } L_j}$ équations et fait intervenir,dans son ensemble, une fois et une fois seulement chacune des $^{\text{Card } E}$ inconnues A_e .

Lemme 3

Si l'on pose: $a_e = A_e \prod_{i \in I \setminus e} R_i$

chaque système S_j devient un système s_j d'inconnues a_e,de même structure (mêmes nombres d'inconnues et d'équations,chaque a_e intervenant une fois et une fois seulement dans le système) mais dont chaque équation est constituée d'une somme de termes a_e (réduite éventuellement à un terme unique) tandis que les seconds membres sont tous égaux à 0 ,sauf un seul,égal à 1.

En effet, l'identification (1) se traduit par:

1° Une équation égalant à 1 la somme des coefficients des termes en $y_{r_{L_j}}$ de T_j. Or ces termes de T_j ne sont obtenus qu'à partir des termes en y_{r_e} où $L_j \subset e \in E$ puisque la sommation est effectuée sur tout $i \notin L_j$. Alors:

- pour i décrivant $e \setminus L_j$, les sommations sur r_i transforment $A_e\, y_{r_e}$ en $A_e\, y_{r_{L_j}}$;
- pour i décrivant $I \setminus e$, les sommations sur r_i transforment ce terme $A_e y_{r_{L_j}}$ en $A_e \prod_{i \in I \setminus e} R_i\ y_{r_{L_j}}$.

Il en résulte que pour tout $e \supset L_j$, le terme en $A_e\, y_{r_{L_j}}$ de T_j est exactement $a_e y_{r_{L_j}}$. Par suite, une équation de (s_j) sera:

$$\sum_{e \in E/\ e \supset L_j} a_e = 1 \qquad (2)$$

2° Chacune des $2^{\text{Card } L_j} - 1$ autres équations du système (s_j) est formée en égalant à 0 la somme des coefficients des $y_{r_{e'}}$ de T_j pour e' strictement inclus dans L_j. Il faut prendre en compte, dans les termes intervenant dans T_j, tous ceux qui correspondent initialement à un e tel que: $e \cap L_j = e'$.
Les sommations sur r_i pour i décrivant $e \setminus L_j$ transforment $A_e y_{r_e}$ en $A_e\, y_{r_{e'}}$ puis pour i décrivant $I \setminus (e \cup L_j)$ en:

$$A_e \prod_{i \in I \setminus (e \cup L_j)} R_i\ y_{r_{e'}}$$

Par suite, l'équation du système (s_j) égalant à 0 le coefficient dans T_j de $y_{r_{e'}}$ sera:

$$\sum_{e \in E\ /\ e \cap L_j = e'} A_e \prod_{i \in I \setminus (e \cup L_j)} R_i = 0$$

soit, puisque e' est fixé:

$$0 = \prod_{i \in L_j \setminus e'} R_i \sum_{e \in E/ e \cap L_j = e'} A_e \prod_{i \in I \setminus (e \cup L_j)} R_i = \sum_{e \in E/ e \cap L_j = e'} A_e \prod_{i \in I \setminus e} R_i$$

D'où:

$$\sum_{e \in E/\ e \cap L_j = e' \subsetneq L_j} a_e = 0 \qquad (3)$$

Cette équation (3) est celle qui, dans le système (s_j), est associée à la partie (strictement incluse) e' de L_j.
A partir des équations (2) et (3) obtenues pour chaque système $s_j \in s$, nous utiliserons la procédure suivante pour déterminer les a_e.

Procédure de détermination des a_e.

Dans le treillis $\mathcal{C}(I)$, nous considérons les générations successives de points $e \in E$:

1° si un e n'a pas d'ascendant dans E, il se confond avec un L_j et d'après (2) : $a_e = 1$.

Inversement d'ailleurs, si un L_j a un ascendant dans E, celà signifie que la facette imposée correspondante est contenue dans une facette également imposée $(L_{j'})$: dans ce cas doit être vérifiée la condition de compatibilité :

$$\sum_{i \in L_{j'} \setminus L_j} \sum_{r_i=1}^{R_i} y_{r_{L_{j'}}} = y_{r_{L_j}}$$

Si l'on supprime les contraintes ainsi "redondantes" (telles que celle relative à L_j vis-à-vis de celle relative à $L_{j'}$) pour ne conserver que la partie "exhaustive" $\tilde{L}$ de L, alors on aura :

$$\forall \tilde{L}_j \in \tilde{L} \quad , \quad a_{\tilde{L}_j} = 1 .$$

2° Descendant dans le treillis $\mathcal{C}(I)$, considérons la première génération de points admettant des ascendants dans E. Soit e' l'un de ces points. Il existe au moins un L_j tel que :

$$e' \subsetneq L_j$$

Donc, dans le système (s_j), l'équation (3) associée à e' permettra de déterminer $a_{e'}$: en effet, toutes les autres inconnues a_e de cette équation correspondent à des $e \supset e'$: ces e appartiennent à des générations antérieures, donc les a_e correspondants ont déjà été déterminés dans la procédure.
On détermine de même tous les a_e pour tous les $e \in E$ de cette génération, puis on applique la procédure aux générations successives, jusqu'au coefficient a_ϕ correspondant au point minimal ϕ du treillis $\mathcal{C}(I)$.

Cependant,on n'aura pas utilisé au cours de la procédure toutes les équations du système (s) ,car un e' peut être inclus dans plusieurs L_j .Le choix dans ce cas du système (s_j) permettant de calculer $a_{e'}$ aura-t-il une influence sur la détermination de ce coefficient? La réponse (négative) est donnée par les lemmes suivants:

Lemme 4

Soit un $e' \in E$ et soit $s' \subset s$ l'ensemble des systèmes s_j tels que: $e' \subset L_j$.**Toute** équation d'un système $s_j \in s'$ faisant intervenir un a_e au moins tel que e soit ascendant de e' (càd $e \supset e'$) ne comporte que de tels ascendants.

En effet,soit un système $s_j \in s'$ et dans s_j une équation comportant un a_e tel que $e \supset e'$. Cette équation est nécessairement associée à une partie e" de L_j telle que $e'' \supset e'$ et elle aura soit la forme (2):

$$\sum_{e \in E / e \cap L_j = L_j} a_e = 1$$

soit la forme (3):

$$\sum_{e \in E / e \cap L_j = e'' \subsetneq L_j} a_e = 0$$

équations qui ne pourront comporter de terme a_e tel que $e \not\supset e'$ puisque $e' \subset e''$.

Lemme 5

Les a_e sont déterminés par la procédure décrite d'exploration du treillis $\mathcal{E}(I)$.

En effet,le résultat est vrai pour tout $a_{e'}$ tel que e' n'ait qu'un seul ascendant L_j dans L.

Si e' admet plusieurs ascendants L_j:dans chaque système s_j associé, il suffira d'additionner membre à membre toutes les équations faisant intervenir au moins l'un de ses ascendants. On obtient ainsi:

$$\sum_{e \in E / e \cap L_j = L_j} a_e + \sum_{e'' / e'' \supset e'} \sum_{e \in E / e \cap L_j = e'' \subsetneq L_j} a_e = 1+0$$

soit, plus simplement:

$$\sum_{e \in E / \ e \supset e'} a_e = 1 \qquad (4)$$

Or cette équation (4) est indépendante de l'ascendant particulier $L_j \in L$ de e'.

Ainsi la procédure permet de déterminer chaque a_e de manière unique ,quel que soit l'ascendant auquel on le rattache dans L, si l'on a le choix.Chaque $a_{e'}$ s'exprime uniquement en fonction de termes a_e tels que $e \not\supseteq e'$, donc en fonction de termes déterminés antérieurement dans la procédure:

$$a_{e'} = 1 - \sum_{e \in E/\ e' \not\subseteq e} a_e \qquad (4)$$

En particulier:

$$a_{\phi} = 1 - \sum_{e \in E / e \neq \phi} a_e$$

<u>En résumé</u> : comme toutes les équations de (s) ont pu être utilisées tour à tour et ont donné ,pour chaque a_e , une valeur unique, nous énoncerons:

<u>Proposition</u>

Le terme général d'un tableau $Y= (y_{r_1 \ldots r_i \ldots r_n})$ à n entrées dont la ième variable d'entrée r_i prend R_i modalités et dont J facettes sont imposées (précisément celles qui correspondent aux sous-ensembles d'indices $L_1,\ldots,L_j,\ldots,L_J$ tels que $L_j \subset I = \{1,\ldots,n\}$) s'exprime sous la forme d'une somme:

$$y_{r_1 \ldots r_i \ldots r_n} = \sum_{e \in E} a_e \frac{y_{r_e}}{\prod_{i \in I \setminus e} R_i} + z_{r_1 \ldots r_i \ldots r_n} \qquad (5)$$

où E est la partie héréditaire $\bigcup_{j=1}^{J} \mathcal{P}(L_j)$ engendrée par l'ensemble $L= \{ L_j / j=1 \ldots J \}$ dans le treillis $\mathcal{E}(I)$ des parties de I ordonné par l'inclusion,le tableau $Z= (z_{r_1 \ldots r_i \ldots r_n})$ étant astreint à avoir ses J facettes correspondantes nulles:

$$\forall\ L_j \in L\ ,\quad z_{r_{L_j}} = 0\ .$$

Les coefficients a_e sont des entiers rationnels calculés de proche en proche par (4).

Remarques

1) Le tableau Y° dont le terme général est le 1er terme du membre de droite de (5) peut s'exprimer uniquement en fonction des facettes imposées puisque : $\forall e, \exists j$ tel que $e \subset L_j$; mais on sera amené à effectuer des sommations sur certains indices r_i .

2) En tant que vecteurs de $\mathbb{R}^{\prod_{i=1}^{n} R_i}$, Y° et Z sont orthogonaux en ce sens que:

$$\sum_{i=1}^{n} \sum_{r_i=1}^{R_i} y^\circ_{r_1 \ldots r_i \ldots r_n} \; z_{r_1 \ldots r_i \ldots r_n} = 0$$

On pourra donc aisément déterminer le tableau de format $(R_1 \times \ldots R_i \times \ldots R_n)$ dont certaines facettes sont fixées et qui est le plus proche (au sens de la distance euclidienne canonique) d'un tableau donné de même format.

Application numérique

Soit un tableau à 5 variables d'entrée notées par commodité d'écriture i,j,k,l,m (au lieu de $r_1,\ldots,r_5$) et qui prennent respectivement I,J,K,L,M modalités (au lieu de $R_1,\ldots,R_5$).

Imposons les 4 facettes: ikl;jkl;im;km.

Alors $E = (ikl, jkl, ik, il, jk, jl, kl, im, km, i, j, k, l, m, \phi)$.

La procédure décrite donne successivement:

$a_{jkl} = a_{ikl} = 1-0 = 1$; $a_{jk} = 1-a_{jkl} = 0$; $a_{jl} = 1-a_{jkl} = 0$; $a_{kl} = 1-a_{jkl}-a_{ikl} = -1$

$a_{ik} = 1-a_{ikl} = 0$; $a_{il} = 1-a_{ikl} = 0$; $a_{im} = 1-0 = 1$; $a_{km} = 1-0 = 1$;

$a_i = 1-a_{ikl}-a_{ik}-a_{il}-a_{im} = -1$; $a_j = 1-a_{jkl}-a_{jk}-a_{jl} = 0$;

$a_k = 1-a_{jkl}-a_{ikl}-a_{jk}-a_{kl}-a_{ik}-a_{km} = -1$; $a_l = 1-a_{jkl}-a_{ikl}-a_{jl}-a_{kl}-a_{il} = 0$;

$a_m = 1-a_{im}-a_{km} = -1$; $a_\phi = 1 - \sum_{e \neq \phi} a_e = 1$.

Soit: $$y^\circ_{ijklm} = \frac{y_{jkl}}{IM} + \frac{y_{ikl}}{JM} - \frac{y_{kl}}{IJM} + \frac{y_{im}}{JKL} + \frac{y_{km}}{IJL} - \frac{y_i}{JKLM} - \frac{y_k}{IJLM} - \frac{y_m}{IJKL} + \frac{y_\bullet}{IJKLM}$$

Il est aisé de vérifier que: $y^\circ_{ikl} = y_{ikl}$; $y^\circ_{jkl} = y_{jkl}$; $y^\circ_{im} = y_{im}$; $y^\circ_{km} = y_{km}$.

BIBLIOGRAPHIE

[1] BACHARACH M.(1965) Estimating Nonnegative Matrices from marginal data. Intern. Econ. Rev. Sept. 1965,pp.294-310

[2] BENZECRI JP (1973) Analyse des Données,T I et II, DUNOD

[3] CAUSSINUS H. (1976) Quelques points de vue sur l'analyse des tableaux d'échanges. Annales de l'INSEE,n° 22-23.

[4] DEMING W.E.,STEPHAN F.F.(1940) On a least-squares adjustment of a sampled frequency table when the expected marginal totals are known.The Annals of Math. Stat.,Vol.11,n°4, pp.427-44.

[5] ESCOUFIER Y.(1981) L'analyse des tableaux de contingence simples et multiples. Publ. du CRIG,Montpellier.

[6] FRECHET M. (1951) Sur les tableaux de corrélation dont les marges sont données .Ann. Univ.de LYON,III^{e} Série,Section A, pp.53-77.

[7] ———— (1956) Même titre CRAS T.242,p.2426.

[8] ———— (1960) Sur les tableaux dont les marges et des bornes sont données .Rev.Inst.Intern. Stat. ,Vol.28,1/2,pp.10-32.

[9] KAMINSKI Ph.(1975) Utilisation de l'entropie pour remplir un tableau à partir de ses marges :problèmes connexes.Note INSEE, Service Régional de BESANÇON .

[10]RAO C.R.(1952) Advanced Statistical Methods in Biometric Research.WILEY.

[11]THIONET P.(1959) L'ajustement des résultats des sondages sur ceux des dénombrements.Rev.Inst.Intern. Stat.,Vol.27,1/3 ,pp.8-25.

[12]————(1961) Sur le remplissage d'un tableau à double entrée. Journ. de la Soc. Stat. de PARIS,n°102,4ème trim.,pp.331-345.

[13] ———— (1964) Note sur le remplissage d'un tableau à double entrée. Journ. de la Soc. Stat.de PARIS,n°105,4ème trim.,pp.228-47.

[14] ———— (1973) Sur la distribution exacte du chi-carré de PEARSON d'une table de contingence. Rev. de Stat.Appl.,XXI,n°4,pp.5-23.

[15] ——— (1976) Construction et reconstruction de tableaux statistiques. Annales de l'INSEE n° 22-23.