

REVUE DE STATISTIQUE APPLIQUÉE

PIERRE THIONET

Sur un problème de Lucien Féraud concernant des tableaux de marges données (tableaux de Fréchet)

Revue de statistique appliquée, tome 25, n° 3 (1977), p. 75-82

http://www.numdam.org/item?id=RSA_1977__25_3_75_0

© Société française de statistique, 1977, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) 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/>

SUR UN PROBLÈME DE LUCIEN FÉRAUD CONCERNANT DES TABLEAUX DE MARGES DONNÉES (Tableaux de Fréchet)

Pierre THIONET
Laboratoire de Statistique
Université Paris IX Dauphine

PRESENTATION

Considérons une population de salariés distribués suivant l'âge (lignes) et le niveau de salaire (colonnes). Les deux distributions marginales sont supposées données, mais non l'intérieur du tableau. A chacun des tableaux possible (sans changer les marges) correspond le salaire moyen de chaque groupe d'âge. Ces salaires moyens sont astreints à un système de contraintes beaucoup plus contraignant en fait que la simple obligation (évidente) d'être compris entre les niveaux minimum s et maximum S de salaires.

Nous donnons une représentation géométrique de ces contraintes, sous forme d'un polyèdre convexe, ensemble des positions possibles du point moyen des salaires. La représentation est explicitée dans un cas simple, sur des valeurs numériques.

1 — Un tableau à double entrée représente la distribution d'une certaine population suivant deux caractères croisés. Par exemple, ce sera une population humaine, de salariés, répartie suivant l'âge et le niveau de salaire. Tel est le point de départ d'un problème que M.L. Féraud, Professeur Honoraire à l'Université de Genève, s'est posé : problème de remplissage d'un tableau (à éléments non négatifs) sous certaines contraintes. Féraud suppose donnée une des deux marges et impose le salaire moyen de chaque tranche d'âge (d'effectif imposé) ; l'autre marge (distribution par niveau de salaire) n'est pas donnée mais se trouve soumise à des contraintes qu'il met en évidence.

Nous étudions ici un problème un peu différent : nous supposons les deux marges imposées, comme dans un problème classique de Fréchet de type I. Fréchet a donné les conditions d'existence d'une solution à un tel problème : "total des sommes de lignes égale total des sommes de colonnes" ; on supposera cela vérifié. Fréchet a donné en outre les conditions d'existence d'une solution à un problème (Type II) beaucoup plus difficile, où l'on impose un majorant à chaque terme du tableau. Nous avons eu jusqu'ici à appliquer ses principes au cas (très fréquent en pratique) où le majorant est zéro ou l'infini. M. Féraud nous a conduit à étudier des contraintes d'un autre type.

2 – LES NOUVELLES CONTRAINTES

On considère un second tableau, de même format que le premier, et de colonnes proportionnelles à celles du premier ; les marges horizontales des deux tableaux sont donc dans des rapports (s_j) donnés. Les marges verticales se trouvent donc dans des rapports (r_i) définis par les (s_j) et par le premier tableau. Supposons les (s_j) fixes.

Si les (r_i) sont également maintenus constants, il est clair que le tableau initial est beaucoup moins libre de varier que lorsque ses seules marges sont imposées. On pourrait étudier ce sous-ensemble de tableaux de l'ensemble défini par le problème de Type I de Fréchet. Nous ne le ferons pas.

Nous désignons par (x) un tableau de type I, par \mathcal{X} son ensemble – en principe non vide puisque la condition d'existence est postulée. Soit (r) le vecteur des (r_i). Nous étudierons l'ensemble \mathcal{R} des (r) pour des (s_j) donnés.

3 – FORMULATION ADOPTÉE

Désignons par ($l \times c$) le format de (x). (x) est un point de \mathbf{R}^{lc} \mathcal{X} est un sous-ensemble de \mathbf{R}^{lc} . Une certaine application $f(s)$ de \mathbf{R}^c dans \mathbf{R}^l applique (x) en (r), donc \mathcal{X} en \mathcal{R} .

Les marges de (x) étant imposées, il s'agira en fait de

$$\mathbf{R}^{(l-1) \cdot (c-1)} \quad \text{et} \quad \mathbf{R}^{(l-1)}$$

Proposition : Ces sous-ensembles sont des *polyèdres convexes*.

Pour le vérifier, précisons les *notations* :

$$(x) = (x_{ij}) \quad , \quad i = 1, 2, \dots, l \quad ; \quad j = 1, 2, \dots, c .$$

$$(r) = (r_i) \quad , \quad (s) = (s_j) .$$

$$(y) = (y_{ij}) = (x_{ij} \cdot s_j)$$

Contrainte des marges : $x_{i.} = a_i$, $x_{.j} = b_j$, $a_{.} = b_{.} = x_{..}$

Définition de (r) : $r_i = y_{i.} / a_i = \sum_j x_{ij} s_j / \sum_j x_{ij}$

On voit que : \mathcal{X} est *borné*, car tout x_{ij} est borné par $\min(a_i, b_j)$.

De plus, \mathcal{X} est *convexe*, car (si p est réel, compris entre 0 et 1) si (x) et (x') font partie de \mathcal{X} , alors $p \cdot (x) + (1 - p) \cdot (x')$ en fait aussi partie (vérification immédiate).

Comme les r_i sont des moyennes de s_j pondérés par les x_{ij} , \mathcal{R} est aussi *borné*. Sa *convexité* se voit en posant : $p \cdot (x) + (1 - p) \cdot (x') = (x'')$, $p \cdot (r) + (1 - p) \cdot (r') = (r'')$; on a : $a_i r_i'' = \sum_j x_{ij}'' s_j$. Enfin, \mathcal{X} et \mathcal{R} ne peuvent être fermés que par des plans. On entend par là les *inégalités linéaires* imposées aux x_{ij} et r_i .

A vrai dire, il faut distinguer entre inégalités *effectives*, qui constituent une contrainte sur un ensemble réel non vide de points (x) ou (r), — et inégalités *non-effectives*, — dominées partout par des contraintes plus fortes ; sans oublier les *redondances*. On sait que Fréchet, pour le problème de type II, n'est pas en mesure d'isoler à coup sur les inégalités effectives, c'est-à-dire les faces du polyèdre. Quant aux sommets (mode usuel de définition d'un polyèdre), leur calcul serait extrêmement pénible.

La suite de cette étude sera consacrée à la recherche des faces du polyèdre \mathcal{R} . Il sera commode d'en illustrer chaque étape au moyen d'un exemple numérique, d'ailleurs très particulier. Alors les sommets seront effectivement calculés.

4 – L'EXEMPLE NUMERIQUE

On s'en tiendra à un tableau (x) de format : 4×3 . Par suite, (r) = (r_i) est un vecteur à 4 dimensions, mais la condition de Fréchet relative au second

tableau (y) lie linéairement les r_i $\sum_i y_i \cdot r_i = \sum_j y_j \cdot r_j = y_{..}$ équivaut à :

$$\sum_i a_i r_i = \sum_j b_j s_j$$

Nous représenterons donc \mathcal{R} dans \mathbf{R}^3 (vue cavalière).

$$a_i = 3 \ 6 \ 9 \ 12 \quad ; \quad b_j = 10 \ 10 \ 10 \quad ; \quad s_j = 1 \ 2 \ 33$$

Il sera commode pour la suite d'imposer aux s_j de former une suite non décroissante, et de distinguer le cas des s_j tous distincts.

Le plus petit et le plus grand des s_j (soit s₁ et s_c) vont jouer des rôles importants ; on les désignera par s et S.

A l'origine, L. Féraud voyait dans s_j un niveau de salaire et il y avait donc lieu de les supposer (strictement) croissants. Les r_i sont des salaires moyens par groupe d'âge i, il les nomme r_i = \bar{s}_i . On ne fera plus état ici de cette interprétation.

5 – NATURE DES FACES

S'inspirant du Type II de Fréchet, on recherchera min r_i et max r_i, i = 1, 2, 3. Puis min r_i + r_h et max r_i + r_h pour les paires (ih) et ainsi de suite. D'ailleurs, pour l = 4, on n'ira pas plus loin. Par contre, min a_i r_i + a_h r_h (max. aussi) apporte aussi des faces.

5.1. Faces de type r_i = min r_i ou max r_i :

Exemple : Les r_i étant moyennes pondérées des s_j sont nécessairement compris entre s et S, ici entre 1 et 3. Mais il s'agit de bornes atteintes si et seulement si la masse totale a_i peut se porter sur 1 (: n_{i1} = a_i) ou bien sur 3 (: n_{i3} = a_i).

Comme n_{ij} est borné par $\min(a_i, b_j)$, et que $b_j = 10$, il s'en suit :

une possibilité pour $i = 1, 2, 3$ ($a_i = 3, 6, 9$) ;

une impossibilité pour $i = 4$: ($a_4 = 12$).

Ainsi, le polyèdre \mathcal{R} est enfermé dans le cube $1 \leq r_i \leq 3$ ($i = 1, 2, 3$) et a en commun avec lui une partie des faces (6).

Mais le cube est entamé à coup sur par :

$$\min r_4 \leq r_4 \leq \max r_4$$

sachant que :

$$3r_1 + 6r_2 + 9r_3 + 12r_4 = 10 + 20 + 30 = 60$$

Le minimum effectif de r_4 est atteint en garnissant le plus possible la case x_{41} (poids $s = 1$) et en reportant le surplus sur x_{42} ($s_2 = 2$). D'où :

$$10 \times 1 + 12 \times 10 \times 2 = 14 \quad \text{pour} \quad 12 r_4 .$$

Le maximum effectif correspond à $n_{42} = 2$, $n_{43} = 10$; d'où 34. D'où :

$$\begin{array}{ccc} 60 - 14 \geq 3r_1 + 6r_2 + 9r_3 \geq 60 - 34 \\ 46 & & 26 \end{array}$$

Ces 2 nouvelles faces s'ajoutent aux 6 premières, mais ne ferment pas le polyèdre (comme on le voit en dessinant). Passons au 2° type.

5.2. Faces de type $a_i r_i + a_h r_h = \min$ (ou \max).

Exemple : On considère successivement les paires de lignes (12) (13) et (23). Dans le cas (12) on voit que $r_1 = r_2 = 1$ (et aussi bien 3) correspondent à des tableaux (x) existant vraiment ; il n'y a pas à envisager pour $3r_1 + 6r_2$ d'autre minimum que 9, ni d'autre maximum que 27.

C'est totalement faux pour les paires (13) et (23). A ce propos, il faut noter que les cas $r_i \max$ et $r_h \min$ (ensemble) sont sans intérêt pour \mathcal{R} .

On va procéder comme suit. On partitionne les lignes, disons (13) et (24) ; ou bien (23) et (14) ; donc on agrège les lignes i et h , puis les autres lignes. Posons $x_{ij} + x_{hj} = x_{Ij}$, $a_i + a_h = a_I$:

Le minimum cherché est celui de $x_{Ij} s_j$ et s'obtient en gonflant x_{I1} d'abord, puis x_{I2} etc. Le maximum s'obtient à l'opposé.

CAS (13) :

$a_I = 3 + 9 = 12$; la répartition (10 2 0) minimise, la répartition (0 2 10) maximise.

CAS (23) :

$a_I = 6 + 9 = 15$: la répartition (10 5 0) minimise, la répartition (0 5 10) maximise

Il reste à opérer la ventilation des x_{Ij} entre lignes i et j . Il y a une infinité de solutions possibles, laissant inchangé le total $r_i a_i + r_h a_h$, avec les deux extrêmes : $r_i \min$ (et $r_h \max$) et l'inverse.

Pour les lignes complémentaires de (ih), il existe une autre infinité de solutions. Ces deux infinités constituent une "face".

Ainsi :

$$14 \leq 3r_1 + 9r_3 \leq 34 \quad ; \quad 20 \leq 6r_2 + 9r_3 \leq 40$$

constituent les quatre dernières faces de notre polyèdre.

5.3. Cas général :

a) *Faces du type* : r_i minimum (ou maximum).

Obtention du min r_i :

Ou bien $r_i = s$ est possible ; c'est que $a_i \leq b_1$ ($n_{i1} = a_i$) ;

Si non :

mais si $b_1 < a_i \leq b_1 + b_2$,

alors : $a_i \min r_i = b_1 s + (a_i - b_1) s_2$

puisqu'on a supposé les s_j croissants (le cas de s_j égaux est disjoint).

Si non :

si $b_1 + b_2 < a_i \leq b_1 + b_2 + b_3$,

alors : $a_i \min r_i = b_1 s + b_2 s_2 + (a_i - b_1 - b_2) s_3$.

etc.

Le maximum de r_i s'obtiendra par un procédé symétrique (à droite).

b) *Faces du type* : $a_i r_i + a_h r_h$ minimum (ou maximum)

Obtention du minimum :

Agrégeant les deux lignes, l'expression à minimiser n'est autre que :

$$x_{I1} s + x_{I2} s_2 + \dots, \quad \text{d'où il suit que :}$$

si $a_i \leq b_1$, le minimum est $a_i s$;

si $b_1 < a_i \leq b_1 + b_2$, c'est $b_1 s + (a_i - b_1) s_2$

et ainsi de suite comme pour le 1° cas.

Il en serait de même pour l'obtention du maximum.

Avec 4 lignes, on a donc 3 opérations sur paires, rien de plus.

Avec 5 lignes, on aurait 10 opérations sur paires.

Avec 6 lignes, 15 opérations sur paires et 10 sur triplets de lignes, etc.

6 – RECHERCHE DES SOMMETS (Exemple numérique seul)

Les faces contigües se coupent suivant des arêtes, les sommets sont communs à au moins 3 arêtes. Aucun dénombrement ne semble possible.

Certaines arêtes du cube subsistent en partie, seuls deux sommets du cube subsistent : soit A et A'.

On appellera AB, AC, AD, les portions d'arête issues de A.

De même pour A'B'C'D'.

Deux autres segments d'arête (EF et E'F') subsistent.

On nomme G H J K G' H' J' K' les autres sommets.

La matrice d'incidence Sommets \times Faces est donnée en Annexe.

Chaque sommet est défini par un certain tableau (x) ; en outre il est défini par ses coordonnées $r_i, i = 1, 2, 3$.

Les coordonnées sont rassemblées en annexe, ainsi que les tableaux.

Le polyèdre est représenté en Annexe en perspective cavalière.

7 – COMPLEMENT : Cas des s_j non tous distincts

Il suffira d'agréger les colonnes de (x) de même s_j pour se ramener au cas traité. On voit facilement que (r) ne s'en trouve pas modifié et qu'il n'est pas nécessaire de désagréger ensuite.



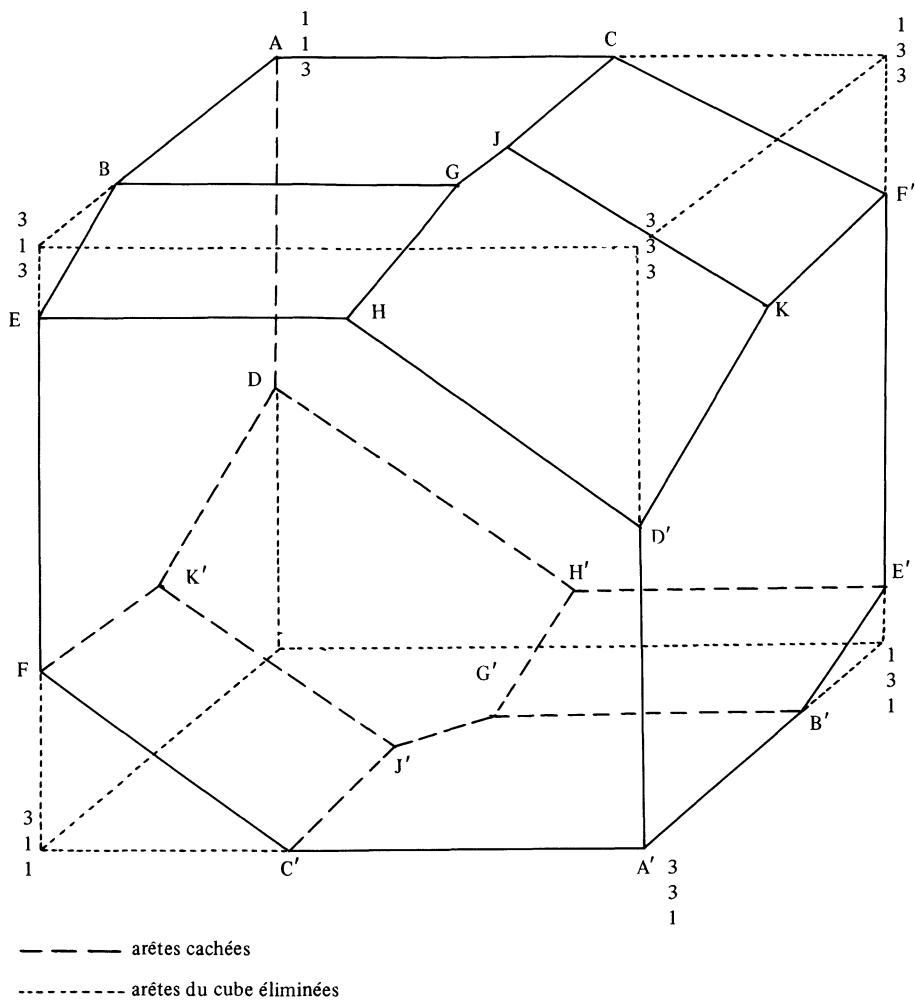
ANNEXE 2

MATRICE D'INCIDENCE ; SOMMETS \times FACES

A	B	C	D	E	F	G	H	J	K	K'	J'	H'	G'	F'	E'	D'	C'	B'	A'	: sommets	
1	1	1					1		1												face : $r_3 = 3$
1	1		1	1	1					1											face : $r_2 = 1$
1		1	1								1		1	1							face : $r_1 = 1$
						1	1	1	1						1						face : $r_4 = 14/12$
	1			1	1	1	1														face : $3r_1 + 9r_3 = 34$
		1						1	1				1								face : $6r_2 + 9r_3 = 40$

Compléter par symétrie pour les 6 autres faces du polyèdre.

ANNEXE I



ANNEXE 3

COORDONNEES DES SOMMETS

A	1	1	3	A'	3	3	1
B	7/3	1	3	B'	5/3	3	1
C	1	13/6	3	C'	3	11/6	1
D	1	1	17/9	D'	3	3	19/9
E	3	1	25/9	E'	1	3	11/9
F	3	1	14/9	F'	1	3	22/9
G	7/3	2	3	G'	5/3	2	1
H	3	2	25/9	H'	1	2	11/9
J	2	13/6	3	J'	2	11/6	1
K	2	3	22/9	K'	2	1	14/9

ANNEXE 4

TABLEAUX (x) DES SOMMETS

A			B			C			D		
3	0	0	0	2	1	3	0	0	3	0	0
6	0	0	6	0	0	0	5	1	6	0	0
0	0	9	0	0	9	0	0	9	1	8	0
1	10	1	4	8	0	7	5	0	0	2	10
E			F			G			H		
0	0	3	0	0	3	0	2	1	0	0	3
6	0	0	6	0	0	0	6	0	0	6	0
0	2	7	4	5	0	0	0	9	0	2	7
4	8	0	0	5	7	10	2	0	10	2	0
			J			K					
			0	3	0	0	3	0			
			0	5	1	0	0	6			
			0	0	9	0	5	4			
			10	2	0	10	2	0			

Les Tableaux A' B' C' D' E' F' G' H' J' K' s'obtiennent de même.

REFERENCES

- FERAUD (L.). — Correspondance privée relative à la thèse d'actuariat de J.P. LACAM.
- FRECHET (M.). — Sur les tableaux de corrélation dont les marges sont données. *Ann. Université de Lyon*, III^o Série Sci. A, 53-77, 1951.
- FRECHET (M.). — Sur les tableaux dont les marges et des bornes sont données. *Revue Inst. Intern. Statistique*, 10-32, 1960.
- THIONET (P.). — Sur le Remplissage d'un tableau à double entrée, *Journ. Soc. Statistique de Paris*, 102 (4^o trim. 1961) 331-345.
- THIONET (P.). — *Ibidem*, 105, (4^o trim. 1964) 228-247.