

J. P. BENZÉCRI

Ajustement d'un tableau à des marges sous l'hypothèse d'absence d'interaction ternaire

Les cahiers de l'analyse des données, tome 8, n° 2 (1983),
p. 227-232

http://www.numdam.org/item?id=CAD_1983__8_2_227_0

© Les cahiers de l'analyse des données, Dunod, 1983, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

AJUSTEMENT D'UN TABLEAU A DES MARGES SOUS L'HYPOTHÈSE D'ABSENCE D'INTERACTION TERNAIRE

[AJUT. MARGES INTER.]

par J.P. Benzécri (1)

1 Le problème de l'ajustement : En 1980, la présente revue a publié trois articles consacrés à ce problème : ce sont [METH. AJUS. MARGES] (J.L. Madre ; CAD Vol V n° 1 pp 87-99) ; [AJUS. MARGES FAC.] (J.P. B. ; Ch. Bourgarit ; J.L. Madre ; ibid n° 2 pp 163-172) ; et [EXEMP. AJUS. MARGES] (V. Cholakian ; ibid pp 173-176). Aux critères considérés dans ces articles on propose ici d'adjoindre une autre méthode (qui aboutit d'ailleurs à des formules classiques) : en bref on considère que le tableau cherché forme avec le tableau donné un tableau à deux étages, auquel on impose l'absence d'interaction ternaire ; (au sens de A. Bener : [INTER. CORR. MULT.] in CAD Vol VII n° 1 pp 25-32 ; 1982). Avant de donner explicitement les formules de cette méthode, nous rappelons pour la commodité du lecteur, l'énoncé du problème dont il s'agit ; les notations étant choisies pour la méthode proposée.

Etant donné un tableau de correspondance f_{IJ}^1 (ou loi de probabilité sur le produit des ensembles finis I et J), et deux lois de probabilité f_I^2 et f_J^2 ; trouver un tableau de correspondance f_{IJ}^2 ayant pour marges les lois données f_I^2 et f_J^2 , et aussi voisin que possible du tableau f_{IJ}^1 (le sens du mot "voisin" restant à préciser).

2 Construction d'une correspondance multiple sous l'hypothèse d'absence d'interaction : En l'absence d'interaction ternaire, un tableau de correspondance f_{IJT} s'exprime en fonction de ses marges simples et binaires par la formule :

$$f_{ijt} = f_{ij} f_t + f_{it} f_j + f_{jt} f_i - 2f_i f_j f_t.$$

Cette formule ne sert pas seulement à calculer les f_{ijt} en fonction des marges données ; elle peut permettre de calculer l'ensemble des f_{ijt} , en partant d'une connaissance partielle des f_{ijt} et des marges. Supposons par exemple que soient données les marges f_{IT} et f_{JT} (satisfaisant à la condition de compatibilité d'avoir une même marge f_T) ainsi que l'étage f_{IJ1} , du tableau ternaire correspondant à la valeur $t = 1$; et satisfaisant de même aux conditions de compatibilité (avec les marges f_{IT} et f_{JT}) :

(1) Professeur de statistique . Université P. et M. Curie.

$$\forall i \in I : \Sigma\{f_{ij1} | j \in J\} = f_{i1} ;$$

$$\forall j \in J : \Sigma\{f_{ij1} | i \in I\} = f_{j1} .$$

On a pour déterminer la marge f_{IJ} , (sous l'hypothèse d'absence d'interaction ternaire) les équations :

$$f_{ij1} = f_{ij} f_1 + f_{i1} f_j + f_{j1} f_i - 2f_i f_j f_1 ; \text{ i.e. :}$$

$$f_{ij} = (1/f_1)(f_{ij1} - f_{i1} f_j - f_{j1} f_i + 2f_i f_j f_1) ;$$

et à partir des marges binaires désormais toutes trois connues, on calcule l'ensemble du tableau ternaire suivant la formule initiale (laquelle, en particulier, redonne pour l'étage f_{IJ1} les valeurs de départ).

Il est facile d'imaginer des généralisations du procédé à la construction de correspondances d'ordre supérieur à 3 (e.g. quaternaires), données par certaines marges et certains étages, sous l'hypothèse que sont nulles certaines interactions.

La seule difficulté est que le tableau construit peut comporter des valeurs négatives : mais à la vérité, comme il n'y a pas d'autre solution aux équations linéaires que celle que nous avons calculée, cette difficulté est inévitable ; au moins si on s'en tient strictement aux conditions posées. Autrement, il est possible, d'une manière ou d'une autre, de modifier le tableau construit pour éliminer les valeurs négatives (cf. *infra*).

3 Application au problème de l'ajustement : On considère le tableau donné f_{IJ}^1 et le tableau cherché f_{IJ}^2 , comme les profils de deux étages d'un tableau ternaire f_{IJT} , avec $T = \{1, 2\}$; et (par exemple) $f_1 = f_2 = 1/2$. Alors les marges binaires f_{IT} et f_{JT} sont connues, car on a pour $t \in \{1, 2\}$:

$$f_{it} = f_i^t f_t = (1/2)f_i^t ; \quad f_{jt} = f_j^t f_t = (1/2)f_j^t .$$

et de même l'étage f_{IJ1} n'est autre que le tableau binaire donné f_{IJ}^1 multiplié par $f_1 = (1/2)$. La méthode du § 2 permet de calculer l'étage f_{IJ2} , lequel n'est autre que f_{IJ}^2 à un coefficient $(1/2)$ près. On aboutit aux résultats suivants :

$$f_{ij} = f_{ij}^1 + (1/2)f_i^2 f_j^2 - (1/2)f_i^1 f_j^1 ;$$

$$f_{ij}^2 = f_{ij}^1 + f_i^2 f_j^2 - f_i^1 f_j^1 .$$

Cette formule est classique : comme l'indique J.L. Madre (cf. [METH. AJUS. MARGES] § 2.2) elle est obtenue par Froment et Lenclud en imposant à f_{IJ}^2 d'être aussi proche que possible de f_{IJ}^1 pour une distance quadratique convenable (qui n'est autre que la distance du χ^2 de centre $f_I f_J$, entre mesures sur $I \times J$; cf. §§ 5 et 6).

Quant aux valeurs négatives éventuellement présentes dans f_{IJ}^2 , on peut les remplacer par des zéros ; et si cette correction altère trop les marges appliquer une méthode itérative pour réduire cet écart (cf. J.L Madre *op. laud.* § 1.4).

4 Extrapolation temporelle : Au § 2 on propose de construire tous les étages f_{IJt} à partir d'un seul d'entre eux f_{IJ1} et des marges f_{IT} , f_{JT} : dans la pratique on connaît souvent plusieurs étages f_{IJt} ; voire tous à l'exception d'un seul, le dernier qu'on désire prédire. En général l'analyse factorielle montre clairement l'existence d'une interaction ternaire. Si par exemple on désigne par T' l'ensemble des temps pour lesquels on a des données complètes et par f_{IJ} le tableau de marge binaire associé à f_{IJT} , on peut analyser f_{IJ} en y adjoignant en supplémentaire (comme lignes et colonnes) d'une part les tableaux réels f_{IJt} ; et d'autre part des tableaux fictifs f'_{IJt} reconstitués sous l'hypothèse d'absence d'interaction ternaire. De ce point de vue pour tenir compte des interactions ternaires on est conduit à des méthodes de prédiction fondées sur une extrapolation indépendante des marges f_{It} , f_{Jt} et des facteurs pour les profils f_J^{it} , f_I^{jt} ; suivie de l'application de la formule de reconstitution en fonction des facteurs.

4' Remarque : décomposition des interactions et ajustement à des marges : Dans la présente note, on a proposé de réaliser l'ajustement à des marges, en construisant un tableau à deux étages pour lequel s'annule l'interaction ternaire. Réciproquement la décomposition des interactions peut être associée à la construction d'une suite de tableaux ajustés à leurs marges simples, binaires etc. : construction plus complexe que la décomposition linéaire faite dans [INTER. CORR. MULT.], mais qui présente l'avantage décisif de ne pas introduire de valeurs négatives. Expliquons cela.

La présence d'interaction ternaire dans un tableau $I \times J \times T$, se manifeste en ce que le tableau $I \times J \times T$ diffère de ce que laissent attendre ses marges IJ , JT , TI : "ce que laissent attendre"... ce peut être la formule de f_{ijt} rappelée au § 2 ; mais ce peut être également un tableau ternaire ajusté aux marges binaires suivant un algorithme de type R.A.S. (cf. Madre ; *op. laud.*) ; avec comme état initial :

$$f_{ijt}^{(0)} = f_i f_j f_t$$

et comme cet algorithme itératif d'ajustement aux marges binaires :

$$f_{ijt}^{(n+1)} = f_{ijt}^{(n)} \left(\frac{f_{ij}}{f_{ij}^{(n)}} \right) \left(\frac{f_{jt}}{f_{jt}^{(n)}} \right) \left(\frac{f_{ti}}{f_{ti}^{(n)}} \right).$$

éventuellement si, comme c'est souvent le cas dans la pratique, les interactions entre i et t ou j et t sont faibles relativement à celles entre i et j , on prendra pour état initial $f_{ij} f_t$).

Ceci fait, on dira que l'existence d'interaction ternaire se manifeste entre $f_{ijt}^{(\infty)}$ et f_{IJT} ; i.e. entre le tableau reconstitué par ajustement à partir des seules marges binaires et le tableau réel. Cette différence pourra être rendue sensible si à l'analyse du tableau binaire f_{IJ} , on adjoint p en éléments supplémentaires (comme lignes ou colonnes suppl.) les étages t du tableau ternaire $f^{(\infty)}$ et ceux du tableau réel : ainsi on comparera les points figuratifs des profils f_J^{it} des lignes réelles à ceux des profils $f_J^{(\infty)it}$ des lignes reconstituées.

On peut, dans cette même voie, définir l'interaction quaternaire dans un tableau IJST, comme la différence entre ce tableau réel et celui qu'on peut reconstituer par ajustement à l'ensemble de ses marges ternaires.

5 Un énoncé d'exercice

Il ne nous a pas paru utile de publier en détail les calculs des constructions et démonstrations de la présente note. Nous préférons en donner le plan sous la forme d'un exercice, qui pourra aider le lecteur désireux de vérifier nos assertions ; ou servira à l'enseignement de l'analyse des données.

Exercice : On rappelle qu'une correspondance ternaire, ou loi de probabilité f_{IJT} sur le produit de trois ensembles finis I, J, T est dite *dépourvue d'interaction ternaire*, si elle s'exprime en fonction de ses marges simples et binaires par la formule (valant quels que soient i, j, t) :

$$f_{ijt} = f_i f_{jt} + f_j f_{it} + f_{ij} f_t - 2 f_i f_j f_t$$

Dans la suite on supposera toujours que f_{IJT} désigne une correspondance dépourvue d'interaction ternaire.

5.1 Soit un élément de l'ensemble T : on se propose de déterminer complètement f_{IJT} en fonction de l'étage f_{IJ1} et des marges f_{IT} et f_{JT} .

5.1 a) i et j étant des éléments de I et J , déterminer f_{ij} en fonction de $f_{ij1}, f_{i1}, f_{j1}, f_i, f_j, f_1$.

5.1 b) i, j, t étant des éléments de I, J, T , déterminer f_{ijt} en fonction de $f_{ij1}, f_{i1}, f_{it}, f_{j1}, f_{jt}, f_i, f_j, f_1, f_t$.

5.2 On suppose désormais que l'ensemble T ne comporte que deux éléments : $T = \{1, 2\}$; et que de plus $f_1 = f_2 = 1/2$. Pour chacun des étages t on note :

$$f_{IJ}^t = (1/f_t) f_{IJt} ; \text{ i.e. } f_{ij}^t = f_{ijt}/f_t$$

et on désigne par f_I^t et f_J^t les lois marginales sur I et J de la loi de probabilité f_{IJ}^t sur $I \times J$.

5.2 a) Trouver des coefficients réels v, w, x, y, z tels que

$$f_{IJ} = vf_{IJ}^1 + wf_I^1 \otimes f_J^1 + xf_I^1 \otimes f_J^2 + yf_I^2 \otimes f_J^1 + zf_I^2 \otimes f_J^2.$$

(on rappelle qu'on note $g_I \otimes g_J$ la mesure h_{IJ} sur $I \times J$ telle que :
 $\forall i, j : h_{ij} = g_i g_j$).

5.2 b) Trouver des coefficients réels $\hat{v}, \hat{w}, \hat{x}, \hat{y}, \hat{z}$ tels que :

$$f_{IJ}^2 = \hat{v}f_{IJ}^1 + \hat{w}f_I^1 \otimes f_J^1 + \hat{x}f_I^1 \otimes f_J^2 + \hat{y}f_I^2 \otimes f_J^1 + \hat{z}f_I^2 \otimes f_J^2$$

5.2 c) Trouver des coefficients réels a, b, c, d tels que :

$$2(f_{IJ}^1 - f_{IJ}^2) =$$

$$(f_I^1 + f_I^2) \otimes (af_J^1 + bf_J^2) + (f_I^1 - f_I^2) \otimes (cf_J^1 + df_J^2)$$

5.3 Soit g_{IJ} une mesure sur $I \times J$ dont les deux marges soient nulles ; i.e. :

$$\forall i \in I : \Sigma\{g_{ij} | j \in J\} = g_i = 0 ;$$

$$\forall j \in J : \Sigma\{g_{ij} | i \in I\} = g_j = 0 .$$

Calculer dans la métrique du χ^2 de centre $f_I \otimes f_J$, le produit scalaire $\langle (f_{IJ}^1 - f_{IJ}^2), g_{IJ} \rangle$. (On se souviendra dans ce calcul de 5.2c).

5.4 Plus généralement soit g_{IJ} une mesure quelconque sur $I \times J$. Exprimer le produit scalaire :

$$\langle (f_{IJ}^1 - f_{IJ}^2), g_{IJ} \rangle_{f_I \otimes f_J},$$

en fonction des produits scalaires :

$$\langle g_I, f_I^1 \rangle_{f_I} ; \langle g_I, f_I^2 \rangle_{f_I} ; \langle g_J, f_J^1 \rangle_{f_J} ; \langle g_J, f_J^2 \rangle_{f_J} .$$

5.5 On note L le sous-espace des mesures sur $I \times J$ ayant pour lois marginales f_I^2 et f_J^2 : déterminer quel est le point de L le plus proche de f_{IJ}^1 , au sens de la métrique du χ^2 de centre $f_I \otimes f_J$.

6 *Solution abrégée*

6.1 a, b : cf. § 2.

6.2 a : $(v, w, x, y, z) = (1, -1/2, 0, 0, 1/2)$

6.2 b : $(\hat{v}, \hat{w}, \hat{x}, \hat{y}, \hat{z}) = (1, -1, 0, 0, 1)$.

6.2 c : $(a, b, c, d) = (1, -1, 1, 1)$

6.3 le produit scalaire demandé est nul

6.4 le produit scalaire demandé s'exprime sous la forme :

$$\langle g_I, f_I^1 \rangle - \langle g_I, f_I^2 \rangle + \langle g_J, f_J^1 \rangle - \langle g_J, f_J^2 \rangle$$

6.5 le point demandé n'est autre que f_{IJ}^2 : en effet f_{IJ}^2 est la projection orthogonale de f_{IJ}^1 sur L ; car si h_{IJ} désigne un point quelconque de L, la différence $g_{IJ} = h_{IJ} - f_{IJ}^2$ a ses deux marges nulles ; donc, (cf. 3°) le produit scalaire $\langle g, f^1 - f^2 \rangle$ est nul. Autrement dit la droite joignant f_{IJ}^1 à f_{IJ}^2 est perpendiculaire à toute droite de L passant par f_{IJ}^2 . C'est le résultat de Froment et Lenclud cité par J.L Madre ; et rappelé ci-dessus au § 3.