LES CAHIERS DE L'ANALYSE DES DONNÉES

K. BEN SALEM

Le critère de la trace minima pour estimer les données manquantes dans un tableau de correspondance

Les cahiers de l'analyse des données, tome 17, nº 1 (1992), p. 97-112

<http://www.numdam.org/item?id=CAD_1992__17_1_97_0>

Les Cahiers de l'Analyse des Données
Vol XVII 1992 - n°1 pp. 97 112.

LE CRITÈRE DE LA TRACE MINIMA POUR ESTIMER LES DONNÉES MANQUANTES DANS UN TABLEAU DE CORRESPONDANCE

[TRAC. MANQ.]

*K. Ben SALEM**

1 Estimation des données manquantes suivant un critère global

Ainsi que nous l'avons exposé ailleurs, il est fréquent de rencontrer, dans la pratique de la statistique multidimensionnelle, des tableaux qui paraissent contenir des informations sûres, et suffisantes pour éclairer le problème qu'on a en vue, mais avec quelques vides sporadiques qui interdisent d'entreprendre immédiatement l'analyse.

Les méthodes utilisées jusqu'ici pour combler de tels vides s'apparentent plus ou moins à la régression multiple tout en perfectionnant celle-ci, notamment par la recherche des plus proches voisins et le recours à la formule de reconstitution comme modèle d'ajustement. Nous avons pu, en suivant cette voie, obtenir des résultats très satisfaisants dans la reconstitution des lacunes de nombreux tableaux.

On se propose ici de procéder autrement: au lieu de considérer chaque donnée manquante comme une inconnue à estimer pour elle-même, prendre pour objet mathématique inconnu le tableau complet lui-même qu'on doit créer; ce tableau complet étant caractérisé par la propriété de minimiser une quantité critère, Crit, convenablement choisie.

De façon précise, soit k un tableau donné considéré comme une fonction positive ou nulle sur le produit $I \times J$ de deux ensembles finis I et J; A une partie de $I \times J$ où, par hypothèse, k n'est pas défini et qu'on appellera ensemble des cases manquantes. (On peut, par commodité de langage, supposer que k a des zéros dans les cases de A.) Notons T(k) l'ensemble des tableaux t (ou fonctions positives ou nulles sur $I \times J$) coïncidant avec k sur $I \times J$ (et assujétis, éventuellement, à des conditions complémentaires qui conviennent au problème

(*) Département des Sciences de l'Informatique; Faculté des Sciences de Tunis.

particulier qu'on a en vue). Alors, dans notre optique, pour combler les lacunes de k, on cherchera un tableau $t \in T(k)$ rendant minimum Crit(t).

Ainsi que l'annonce le titre de l'article, la quantité Crit choisie est la trace, ce terme étant entendu au sens de l'analyse des correspondances, comme l'inertie totale de l'un et l'autre des nuages N(I) et N(J):

$$\mathrm{Crit}(t) = \sum\{\ t(i, j)^2 / (t(i).t(j)) \mid i \in I\ ;\ j \in J\ \} - 1\ \ ;$$

formule où t(i) et t(j) désignent respectivement (comme d'usage) le total de la ligne i ou de la colonne j. On sait que Crit est ≥0, et ne s'annule que si t est une loi produit sur $I \times J$.

Ce critère nous paraît avantageux parce que minimiser la trace revient, en bref, à faire en sorte qu'il y ait au sein des lignes, comme au sein des colonnes, la similitude la plus grande possible. En particulier, si, comme c'est souvent le cas dans la pratique, la plupart des lignes et des colonnes sont sans lacune, en minimisant la trace, on comble les lacunes de telle manière que les lignes complétées s'assimilent au nuage des lignes complètes; et de même pour les colonnes. Or tel est le principe commun à toutes les méthodes usuelles de reconstitution des données manquantes.

Comme dans tout problème de minimum, la question se posera de l'existence et de l'unicité d'une solution. Ici, l'existence sera généralement assurée du fait de la compacité de T(k) : il faudra cependant préciser le domaine de définition de Crit, ce que nous ferons au §2. Nous n'avons, en revanche, que des résultats partiels quant à l'unicité. Dans certains cas, certes très simples mais d'un réel intérêt pour des applications, il est facile d'obtenir un minimum et d'en démontrer l'unicité. En général, on pourrait recourir à une méthode itérative suggérée en conclusion (cf. §7).

Dans la suite, après avoir levé l'indétermination que recèle la définition de la trace (§2), nous considérerons deux cas complètement résolus: celui où les marges (plus précisément les lois marginales) du tableau complet t sont imposées (§§3 & 5); et celui d'une seule donnée manquante (§4). Puis nous compléterons des démonstrations et examinerons des exemples (§6); et conclurons en posant des questions relatives à l'ensemble A des lacunes et à la recherche itétative d'une solution (§7).

2 Définition de la trace et recherche du minimum

2.1 Extension du domaine de définition de la trace

La formule de la trace comporte des quotients. Pour un tableau t donné, le calcul n'est bien défini que si tous les totaux marginaux t(i) et t(j) sont non nuls. Sous l'hypothèse générale que t ne comporte que des nombres positifs ou nuls, une division par zéro ne peut se rencontrer qu'avec un numérateur nul; car si t(i) ou t(j) s'annule, t(i, j) est certainement nul (comme étant l'un des termes, tous ≥ 0, rentrant dans les sommes à effectuer pour calculer t(i) ou t(j)). Plus précisément, le terme tr(i, j), défini par:

$$\mathrm{tr}(i, j) = t(i,j)^2 / (t(i).t(j)) \quad ,$$

est certainement nul si l'un des deux nombres t(i) ou t(j) est non nul et que l'autre est nul; mais si t(i) = t(j) = 0, on peut, par passage à la limite, assigner à tr(i, j) toute valeur de l'intervalle (0, 1). C'est ce qu'on voit sur l'exemple très simple du tableau carré ci-dessous:

$$t(e) = \begin{vmatrix} 1 & xe \\ & \\ ye & ze \end{vmatrix} \quad ; \qquad \mathrm{tr}(2,\ 2) = z.z/((z+x)(z+y)) \ ;$$

Quand e tend vers 0, x, y et z, restant constants, tr(2, 2) garde la même valeur qui est aussi la valeur limite de Crit(t(e)).

Bien que certains calculs se fassent le plus simplement en se plaçant dans l'espace des fonctions ≥ 0 sur $I \times J$ (et c'est ce qu'on vient de faire pour t(e)), les propriétés de convergence apparaissent mieux si l'on se restreint aux tableaux de nombres ≥ 0 ayant pour somme 1, c'est-à-dire au simplexe S = S($I \times J$) des lois de fréquence sur $I \times J$. Mais, même sur ce simplexe, la fonction Crit n'est définie et continue que si l'on se restreint au sous-ensemble S* des lois dont les deux marges sont stictement positives.

Afin de traiter Crit comme une fonction définie et continue, on va construire un compact S° dont S* est un ouvert partout dense. De façon précise, notons:

$$m = \mathrm{Card} I \quad ; \quad n = \mathrm{Card} J \quad ; \quad \mathbf{K} = (0,\ 1)^{3mn} \quad ;$$

K est un cube d'arête 1 en dimension 3mn; à toute loi t sur $I \times J$, on associe un point K(t) de **K**, dont les 3mn coordonnées sont, successivement, les composantes de la loi elle-même et celles des transitions qui lui sont associées:

$$K(t) = \{t_{IJ} \ ; t_J{}^I \ ; t_I{}^J\}$$

$$= \{t_{11}, t_{12}, \ldots, t_{mn} \ ; t_1{}^1, t_2{}^1, \ldots, t_n{}^m \ ; t_1{}^1, t_1{}^2, \ldots, t_m{}^n\} \quad ;$$

où, selon l'usage, $t_j{}^i = t_{ij}/t_i$; $t_i{}^j = t_{ij}/t_j$.

Sur le cube **K** tout entier, on peut définir une fonction continue Krit comme étant la somme de mn termes dont chacun est le produit d'une coordonnée de rang mn+h par la coordonnée de rang 2mn+h, h variant de 1 à mn. On a ainsi:

$$\mathrm{Crit}(t) = \mathrm{Krit}(K(t)) = \ \Sigma\{t_j{}^i \, t_i{}^j \mid i \in I \ ; j \in J\} - 1 \quad ;$$

Ceci étant, il suffit de prendre pour S° la fermeture de K(S*) dans **K**; et l'on obtient un domaine compact où Krit prolonge par continuité le critère Crit défini sur S*. Les mn premières coordonnées de **K** nous ont servi pour que K(S*) soit une image biunivoque et bicontinue de S*; et les deux blocs suivants de mn coordonnées ont permis de prolonger la fonction critère.

Il importe de comparer S°, fermeture dans **K** de K(S*) $\approx$ S*, et S, fermeture proprement dite de S*. Les mn premières coordonnées de **K** permettent d'associer à tout élément s° $\in$ S°, une loi s $\in$ S; mais à une seule loi s est associée une infinité de s°, correspondant aux différents systèmes limites $s_J{}^I$ et $s_I{}^J$ qu'on peut obtenir en tendant vers s. On dit que, dans S°, s est *éclaté.*

Le tableau carré 2 $\times$ 2, t(e), considéré plus haut, offre un exemple simple d'éclatement: selon la direction du vecteur (x, y, z) on a, dans S°, une limite diférente de k(t(e)) pour e $\rightarrow$ 0.

2.2 Recherche générale du minimum

Revenons maintenant à la recherche d'un tableau complété réalisant le minimum du critère. Convenons de noter f_{IJ}, ou simplement f, la loi de fréquence associé au tableau k donné, les lacunes étant comblées par des zéros. À chaque lacune, a = (i,j) $\in$ A, est associée une loi singulière sur I $\times$ J, a_{IJ}, dont la masse 1 est concentrée en a, avec 0 ailleurs. Au §1, on a noté T(k) l'ensemble des lois susceptibles d'être obtenues en comblant les lacunes de k. Dans le cas général (sans conditions complémentaires) T(k), considéré comme ensemble de lois sur I $\times$ J, n'est autre que le simplexe S(k) ayant pour sommets, d'une part, f_{IJ} et, d'autre part, l'ensemble des a_{IJ} pour a $\in$ A.

Or, une loi t $\in$ S(k) rentre dans la formule générale:

$$t_{IJ} = (1-\alpha)\, f_{IJ} + \alpha\, \mu_{IJ} \quad ,$$

où μ désigne une loi de probabilité (mesure de masse 1) dont le support est A. Si l'on a trouvé dans S(k) une loi t_{IJ} réalisant le minimum du critère, il est facile d'en déduire un tableau k complété: si |k| désigne la masse totale du tableau k (autrement dit la somme des k(i, j) donnés), il suffira, pour combler les lacunes, d'ajouter au tableau k donné le tableau |k| $(\alpha/(1\text{-}\alpha)\ \mu_{IJ}$.

Cette formule appelle notre attention sur l'éventualité que α soit égal à 1. Dans ce cas, on rencontre une division par 0; le tableau reconstitué contiendrait des quantités infinies. En fait, l'obstacle apparaît avant ce calcul. Il est légitime de postuler que les marges de f ne comportent pas de zéros (ceci revient à éliminer *a priori* les lignes ou colonnes de zéros dont les lacunes éventuelles sont naturellement comblées par des 0); mais tel n'est aucunement le cas pour μ_{IJ}.

Le critère Crit(t) n'est lui-même certainement défini que pour $\alpha < 1$; mais si le domaine S(k) est restreint à son intersection S*(k) avec S*, on n'a que des lois t pour lesquelles Crit est défini. En se plaçant dans S°, on trouvera toujours au moins un point de la fermeture de S*(k) où se réalise le minimum de Krit; ce point correspondra à une limite de tableaux t. Mais sans imposer à A de conditions particulières, nous ne savons pas démontrer, en général, que cette limite corresponde à $\alpha < 1$. Toutefois, cette inégalité sera satisfaite dans les applications des §§3, 4 & 5.

Si, outre le fait de coïncider avec k sur $I \times J$, on impose au tableau complété d'autres conditions, S*(k) sera restreint à une partie T*(k), définie, éventuellement, par des conditions linéaires affines: tel est bien le cas (cf. §3) si les lois marginales t_I et t_J sont fixées *a priori*. Il faut seulement veiller à ce que la partie T*(k), de S*(k), satisfaisant aux conditions complémentaires, ne soit pas vide. Alors, à T*(k) correspond, dans **K**, une fermeture T°(k); et si l'on a trouvé dans T°(k) une loi t_{IJ} réalisant le minimum du critère, il est facile d'en déduire un tableau k complété.

3 Cas où les marges du tableau complété sont imposées

3.1 Algorithme général

Si les données à reconstituer ne sont assujéties à aucune condition *a priori*, les lois marginales de la loi de fréquence (mesure de masse totale 1) associée à t ne sont pas connues: il y a dans T*(k) des tableaux qui ont des lois marginales différentes. On ne rentre donc dans le cas du présent § que si, moyennant l'adjonction de conditions complémentaires, on impose les lois marginales de t. Or, du fait de ces conditions complémentaires, T*(k) est fermé (car, en bref, toute limite de tableaux ayant des marges fixées a ces mêmes marges): T*(k) est homéomorphe à ce qu'on a noté T°(k). Nous écrirons donc, simplement, T(k).

De plus, si les marges t_I et t_J sont fixées, la fonction Crit a sur T(k) une forme très simple. On sait qu'en général, la trace d'une loi g_{IJ} est égale au carré de la distance du χ^2 entre g_{IJ} et le produit $g_I \times g_J$ de ses marges, la distance étant calculées entre mesures sur $I \times J$, avec pour centre la loi produit $g_I \times g_J$ elle-même. Dans le cas présent, Crit, restreint à T(k), n'est donc autre que la distance au carré entre t_{IJ} et $t_I \times t_J$, qui est fixé, pour une métrique également fixée. Le minimum de Crit est donc unique, et s'obtient en projetant $t_I \times t_J$ orthogonalement sur T(k), muni dune base orthonormée par l'algorithme de SCHMIDT et HILBERT (cf. *l'Anal. des Données* TII Bn°12 §1.3).

3.2 Modification de la diagonale d'un tableau symétrique

L'application la plus utile du présent résultat nous paraît être la suivante. Outre les données manquantes fortuites, résultant d'une erreur de mesure ou de la perte d'un document, il y a des données manquantes systématiques si, pour certaines paires a = (i, j), k(i, j) ne peut être défini comme il l'est, en général, pour les autres paires. Tel est, en particulier, le cas de tableaux carrés dont la diagonale manque.

Un article récent ([MARTINIQUE FLUX], in *CAD*, Vol XVI, n°3) a rappelé l'attention sur le fait que, dans une matrice de flux (en l'occurrence une matrice de flux migratoires de commune à commune entre deux recensements consécutifs), les termes diagonaux manquent. On peut certes considérer que ce qui reste en place au point i (dans l'article cité: la population non migrante) constitue un flux de i vers i; mais cette assimilation engendre une diagonale très chargée qui perturbe l'analyse, avec de fortes valeurs propres et des facteurs qui n'expriment nettement que l'association de tel ou tel point i avec lui-même.

Dans l'article cité, un raisonnement plausible fournit des termes diagonaux d'un ordre de grandeur acceptable; et l'analyse est très satisfaisante; mais on souhaiterait disposer d'une méthode générale pour créer la diagonale.

La matrice d'incidence d'un graphe (orienté ou non) offre un exemple semblable. Il est naturel de poser k(i, i') = 1 si une arête va de i vers i', et zéro sinon; mais la valeur des k(i, i) ne s'impose pas, en général. Fixer k(i, i) à 1 (voire à 0) peut, comme dans le cas des flux, perturber l'analyse de la matrice d'incidence; analyse à laquelle il est cependant naturel de demander une image globale de la structure du graphe; image sur laquelle apparaîtraient des amas de sommets fortement reliés entre eux.

Reste à fixer judicieusement les marges qu'on imposera à la matrice complétée des flux (ou des arêtes).

Supposons d'abord que la matrice k donnée est symétrique; et qu'on a choisi la marge f_I imposée (marge unique, du fait de la symétrie). Parmi les tableaux de nombres ≥ 0 coïncidant avec k en dehors de la diagonale et ayant la loi marginale f_I, il en est un de plus faible poids. Notons f_{II} la loi correspondante; alors, dans l'espace des lois sur $I \times I$, T(k) est le segment joignant la loi de fréquence f_{II} à la loi portée par la diagonale et ayant pour marge f_I. La recherche, sur T(k), de la loi de trace minima est aisée. Si k n'est pas symétrique, il se peut même que ses deux marges sur I ne soient pas égale. On considérera donc la matrice k' symétrisée de k; on reconstituera pour k' une diagonale; et on attribuera cette diagonale à k.

Pour fixer f_I, nous proposons deux voies.

Première voie: imposer que t ait mêmes marges que k; en d'autre terme, que la diagonale de t soit proportionnelle à la marge de k : f_{II} (cf. *supra*) est alors la loi de k (complété par des zéros sur la diagonale) et f_I est la marge de f_{II}. La voie est facile, mais en butte à une grave objection. Si la diagonale de k peut être choisie de telle sorte qu'on obtienne une loi produit (donc de trace nulle; et, *a fortiori*, minima), cette solution ne sera généralement pas atteinte par cette voie.

Supposons en effet qu'il existe une loi a_I et une constante h telles que, pour i≠i', on ait k(i, i') = $h.a_i.a_{i'}$; alors, la somme des termes extradiagonaux de la ligne i (ou de la colonne i) sera: $h_i = h.(a_i - (a_i)^2)$; et le système des $\{h_i\}$ n'est proportionnel à celui des $\{a_i\}$ que si les a_i sont tous égaux entre eux.

Voici donc la deuxième voie. Les h_i étant connus (par sommation de termes extradiagonaux du tableau k donné), on cherchera une constante h telle qu'ait pour somme 1 le système des $\{a_i\}$ déterminés chacun en résolvant l'équation du second degré: $(a_i)^2 - a_i + (h_i/h) = 0$; et ces $\{a_i\}$ fourniront la marge à imposer.

Nous nous bornerons à affirmer ici l'existence et l'unicité de la constante h; renvoyant au §5 pour les démonstrations et pour la recherche de h.

Outre son utilité pour les applications, la reconstitution d'un tableau avec des marges imposées a l'intérêt de suggérer des hypothèses générales quant aux propriétés du critère Crit: Crit est-il une fonction concave sur T(k)? A-t-il au moins un minimum unique sur tout segment de droite inclus dans le simplexe des lois sur I × J ? Nous verrons, au §6, que ces hypothèses ne se vérifient pas, ce qui rend d'autant plus difficile la démonstration d'un théorème général d'unicité du minimum.

4 Reconstitution d'une donnée manquante unique

Reprenons les notations du §2.2: f_{IJ}, est la loi de fréquence associée au tableau k donné, la lacune unique étant comblée par 0. Il est commode de désigner cette lacune par a = (i°, j°), ou, plus brièvement, a = (o, o); dans la suite les lettres i ou j serviront, respectivement, à désigner un élément autre que i°=o ou j°=o; et le signe Σ désignera des sommes dont o est exclu. Le simplexe S(k) est ici réduit au segment dont les extrémités sont f_{IJ} et a_{IJ}, (loi de masse 1, concentrée en a, avec 0 ailleurs); et la forme générale d'une loi t ∈ S(k) est:

$$t_{IJ} = (1- x)\, f_{IJ} + x\, a_{IJ} \quad .$$

En toute rigueur, il faut se restreindre à S*(k), i.e. à x ∈ [0, 1[; mais un calcul, analogue à celui fait au §2.1 pour t(e), donne pour la trace une fonction Cr(x) dont la limite est bien définie pour x → 1; et S°(k) s'identifie à S(k).

Dans la somme, sur I × J, qui donne Cr(x) = Crit(t), nous distinguerons 4 membres: le terme en (o, o); les termes en (o, j) (ligne i°); les termes en (i, o) (colonne j°); et le bloc des termes en (i, j) (i.e. I × J, privé de la ligne i° et de la colonne j°). Les termes de ce dernier bloc ne dépendent pas de x; car on a:

$$(t_{ij})^2 / (t_i\, t_j) = ((1\text{-}x)\, f_{ij})^2 / ((1\text{-}x)f_i\, (1\text{-}x)f_j) = (f_{ij})^2 / (f_i\, f_j) \quad ;$$

nous noterons simplement Y la somme de ces termes qui ne jouent aucun rôle dans la recherche du minimum de Cr(x).

Convenons de noter:

$$\Sigma\{f_{oj} \mid j \neq o\} = H \quad ; \quad \Sigma\{f_{io} \mid i \neq o\} = G \quad ;$$

le terme en (o, o) de la trace s'écrit alors:

$$(t_{oo})^2 / (t_o\, t_o) = x^2 / ((x + (1\text{-}x)\, H)\, (x + (1\text{-}x)\, G)) \quad ;$$

pour les termes en (o, j), de la ligne i°, on a la somme:

$$(1\text{-}x)\, \Sigma\{(f_{oj})^2/f_j \mid j \neq o\} / (x + (1\text{-}x)\, H) \quad ;$$

et, de même, pour les termes en (i, o), de la colonne o:

$$(1\text{-}x)\, \Sigma\{(f_{io})^2/f_i \mid i \neq o\} / (x + (1\text{-}x)\, G) \quad .$$

Puisque f_{oj} est l'un des termes de la colonne j dont f_j est la somme, on a:

$$f_{oj} \leq f_j \quad ; \quad \text{d'où : } (f_{oj})^2/f_j \leq f_{oj} \quad ; \text{ et de même :}$$

$$f_{io} \leq f_i \quad ; \quad \text{d'où : } (f_{io})^2/f_i \leq f_{io} \quad ;$$

d'où pour les sommes, désignées ci-après par h et g, les inégalités:

$$h = \Sigma\{(f_{oj})^2/f_j \mid j \neq o\} \leq H \quad ; \quad g = \Sigma\{(f_{io})^2/f_i \mid i \neq o\} \leq G \quad .$$

Les égalités, h = H et g = G, ne sont vérifiées simultanément que si les f_{ij} non nuls (autres que ceux situés dans la ligne i° ou la colonne j°) sont tels que $f_{i^\circ j}=f_{ij^\circ}=0$. Dans ce cas exceptionnel, le tableau t_{IJ} est décomposé en deux blocs diagonaux dont l'un a pour trace Y-1; et l'autre a pour trace une fonction décroissante de x qui s'annulle si x tend vers un; auquel cas la trace Cr(x) atteint son minimum, Y. Nous supposerons désormais qu'on a l'inégalité stricte: $h+g < H+G$.

Avec les notations déjà posées, Cr(x) s'exprime assez simplement:

$$Cr(x) = x^2 / ((H+x(1\text{-}H))\, (G+x(1\text{-}G)))$$
$$+ (1\text{-}x)h / (H+x(1\text{-}H)) + (1\text{-}x)g / (G+x(1\text{-}G)) + Y - 1 \quad ;$$

ou, en notant : $z = 1\text{-}x$; $U = 1\text{-}H$; $V = 1\text{-}G$:

$$Cr(x) = (Y-1) + (1-z)^2/((1-zU)(1-zV)) + zh/(1-zU) + zg/(1-zV)$$
$$= (Y-1) + ((1-z)^2 + z(h+g) - z^2(hV+gU)) / ((1-zU)(1-zV)) \quad ;$$

Ainsi, Cr(x) est le quotient de deux trinômes en x; le dénominateur a deux zéros négatifs (ou un zéro double, si H=G); pour x = 0, la dérivée de Cr est $-((g/G^2)+(h/H^2)) < 0$; pour x = 1, elle vaut H+G -(h+g) > 0 ; Cr' a donc un nombre impair de zéros entre 0 et 1; or, du fait de la forme algébrique de Cr, Cr' a au plus 2 zéros; donc Cr' a, entre 0 et 1, un seul zéro, qui ne peut être que l'unique minimum de Cr. Ce qui établit qu'avec le critère choisi, une donnée manquante isolée peut être reconstituée de façon unique.

5 Recherche de la diagonale et de la loi marginale d'un tableau symétrique sous l'hypothèse d'indépendance

Reprenons le problème et les notations du §3.2: f_{II} est la loi d'un tableau k donné dont la diagonale est nulle; on désigne par $b_I = \{b_i\}$ la loi marginale de f_{II}; on cherche une loi a_I telle que, pour une valeur convenable de h, on ait

$$\forall\, i \in I \;:\; (a_i)^2 - a_i + (b_i/h) = 0 \quad ;$$

a_I fournira la marge à imposer; le tableau (1/h) f_{II} sera complété par la diagonale c_I :

$$\forall\, i \in I \;:\; c_i = a_i - (a_i)^2 \quad ;$$

et le tableau ainsi complété sera combiné avec le tableau diagonal dont la marge est a_I afin de minimiser la trace.

Nous nous proposons de montrer qu'un a_I existe et qu'il est unique; dans la démonstration, joue un rôle essentiel le fait que chacun des b_i est compris entre 0 et 1/2. cette dernière propriété résulte immédiatement du fait que, dans un tableau symétrique dont la diagonale est nulle, le poids d'une colonne (e.g. la colonne 1) est inférieur (ou égal) à la somme des poids des autres colonnes; car, en bref, chacun des éléments non nuls, de rang i ($i \neq 1$), de la colonne 1, se retrouve, par symétrie, dans la colonne i, sur la ligne 1.

Partons du simplexe S_I des lois de probabilités a_I sur I; en général, une loi a_I définit une mesure positive c_I (dont la masse totale est inférieure à 1); et la loi de probabilité associée à c_I sera notée b_I: on vient de voir que b_I est contenu dans la partie B_I de S_I :

$$B_I = \{b_I \mid b_I \in S_I \;;\; \forall\, i \in I : b_i \leq 1/2\} \;;$$

reste à montrer que (à des particularités près, que nous préciserons, relatives à la frontière) la correspondance entre $a_I \in S_I$ et $b_I \in B_I$ est biunivoque.

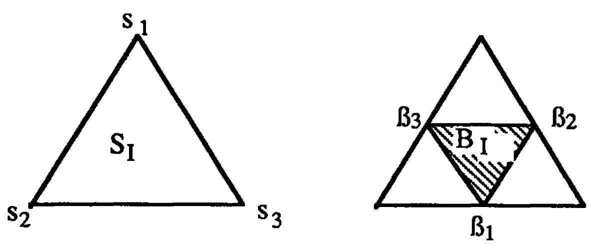
Pour préparer la démonstration, partons des cas particuliers où cardI, noté n, vaut 2, 3 et 4. Si n=2, on a, nécessairement:

$$a_2 = 1-a_1 \quad ; \quad c = c_1 = c_2 = a_1 - (a_1)^2 = a_2 - (a_2)^2 \quad ; \quad b_1 = b_2 = 1/2 \quad ;$$

c est compris entre 0 et 1/4, et les a_i ne sont autres que les deux racines de l'équation du second degré en a:

$$a^2 - a + c = 0 \quad ; \quad a = (1 \pm \sqrt{(1 - 4c)}\,)\,/\,2 \;;$$

ces deux racines seront notées, dans la suite, $a\ (c)<1/2$ et $a_+(c)>1/2$.



Si n=3, S_I est un triangle dont les trois sommets sont les mesures ponctuelles de masse 1: $s^1 = \{1, 0, 0\}$, $s^2 = \{0, 1, 0\}$, $s^3 = \{0, 0, 1\}$; d'après ce qu'on a vu pour n=2, au côté (s^1, s^2) de S_I , correspond l'unique point $ß^3 = \{0, 1/2, 1/2\}$; et B_I n'est autre que le triangle $(ß^1, ß^2, ß^3)$.

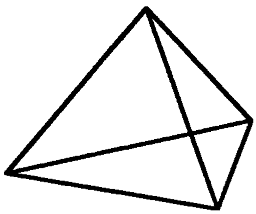
Il importe de préciser comment, au voisinage du sommet s^1 de S_I, correspond le côté $(ß^2, ß^3)$ du triangle B_I. Soit $a_I = (1-\varepsilon, \varepsilon', \varepsilon'')$, où les ε sont des quantités petites, positives, telles que $\varepsilon = \varepsilon' + \varepsilon''$. On a:

$$c_1 = (1-\varepsilon) - (1-\varepsilon)^2 = \varepsilon - \varepsilon^2 \approx \varepsilon \quad ; \quad c_2 = \varepsilon' - \varepsilon'^2 \approx \varepsilon' \quad ; \quad c_3 = \varepsilon'' - \varepsilon''^2 \approx \varepsilon'' \quad ;$$

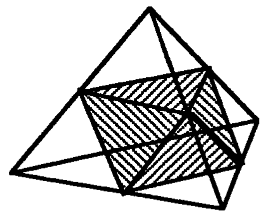
$$b_1 \approx \varepsilon / (2\varepsilon) = 1/2 \quad ; \quad b_2 \approx \varepsilon' / (2\varepsilon) \quad ; \quad b_3 \approx \varepsilon'' / (2\varepsilon) \quad ;$$

(où $b_1 < 1/2$); ainsi, à la limite, à chaque sommet de S_I, lieu de a_I, il correspond une arête de B_I ; et à chaque arête, un sommet.

Si n=4, S_I est un tétraèdre; la condition que chacune des 4 masses b_i soit $\leq 1/2$ délimite un octaèdre B_I ; à chaque arête de S_I, correspond un point $ß_I$

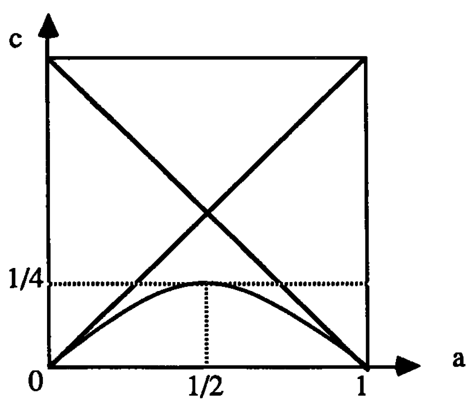


tétraèdre S_I



octaèdre B_I

unique qui n'est autre que le milieu de cette arête; si a_I parcourt une face de S_I, e.g., la face (s^1, s^2, s^3), définie par la condition a_4=0, b_I parcourt le triangle, inscrit dans la face (b_4=0), et ayant pour sommets les milieux des arêtes qui sont les côtés de cette face; au voisinage d'un sommet, e.g. s^4, correspond la face triangulaire de l'octaèdre B_I, ensemble des points b_I du simplexe assujétis à la condition b_4=0.



Reste à préciser, en général, la correspondance entre S_I et B_I.

Nous ferons une remarque préliminaire: si les a_i satisfont, entre eux, à certaines relations d'ordre, les c_i, et donc les b_i, satisfont aux mêmes relations d'ordre. Si tous les a_i sont inférieurs à 1/2, c'est clair: on a, pour tout i, a_i=a (c_i) et la fonction a réalise une correpondance biunivoque croissante entre les deux intervalles c $\in$ (0, 1/4) et a $\in$ (0, 1/2). Supposons maintenant que l'un des a_i, e.g. a_1, est supérieur à 1/2; on a alors a_1=a_+(c_1); les autres a_i, qui se partagent le reste de la masse, 1-a_1=a (c_1), sont nécessairement inférieurs à ce reste, *a fortiori* <1/2; on a donc pour ces a_i: a_i=a (c_i); et, d'après ce qu'on a dit de la fonction a , il y a les mêmes relations d'ordre entre les nombres de $\{1-a_1,\ldots,a_i\}$, ou de $\{a_1,\ldots,a_i\}$, et ceux de $\{c_1,\ldots,c_i\}$ ou de $\{b_1,\ldots,b_i\}$.

Dans la suite, on supposera que le plus grand des a_i est a_1; on dira qu'on est dans le cas (+) si $1/2<a_1$; et dans le cas (-) si $a_1<1/2$. On notera $\hat{a}_1=a_1$ dans le cas (-) et $1-a_1$ dans le cas (+). De la concavité de la fonction a , dont la courbe représentative est un arc de parabole, il résulte qu'il y a, entre les nombres $\{\hat{a}_1,\ldots,a_i\}$ et $\{c_1/\hat{a}_1,\ldots,c_i/a_i\}$ des relations d'ordre inversées; en particulier, $c_1/\hat{a}_1$ est inférieur à chacun des c_i/a_i.

Cette dernière remarque confirme que les b_i sont $\leq 1/2$ (l'égalité n'étant réalisée que si tous les b_i sont nuls, exceptés deux d'entre eux, qui valent 1/2; ce qui, comme dans les cas n=2 ou n=3, correspond à a_I situé sur une arête et b_I, milieu de cette arête; dans la suite, nous laissons au lecteur d'effectuer les restrictions qu'implique ce cas limite). En effet, on retrouve ici que c_1 est inférieur à la somme des autres c_i; car on sait déjà que $\hat{a}_1$ est inférieur ou égal à la somme des autres a_i: dans le cas (-) c'est clair, puisque $\hat{a}_1=a_1<1-a_1=\Sigma\{a_i|i\neq 1\}$; et dans le cas (+), $\hat{a}_1=1-a_1=\Sigma\{a_i|i\neq 1\}$; l'inégalité $c_1<\Sigma\{c_i|i\neq 1\}$résulte alors de ce que $(c_1/\hat{a}_1)$ est inférieur aux (c_i/a_i).

Soit maintenant à déterminer a_I, b_I étant donné. Selon la convention faite, désignons par b_1 le plus grand des b_i. On va voir que le choix de a_1 détermine les autres a_i ; en sorte qu'il ne restera plus qu'à vérifier que la somme de a_1 et des a_i, ainsi déterminés, ne vaut exactement 1 que pour une seule valeur de a_1, prise entre 0 et 1.

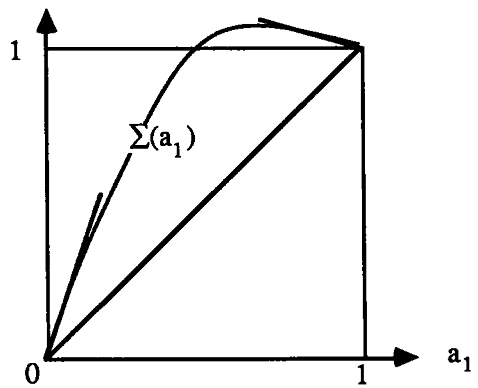
De façon précise, a_1 étant choisi, on a $c_1=c(a_1)=a_1-(a_1)^2$; on notera $x=x(a_1)=c_1/b_1$; et on prendra pour chacun des a_i la valeur $a_i(x)=a\,(x.b_i)$. Ceci posé, on peut noter:

$$\Sigma(a_1) = a_1 + \Sigma\{a_i(x) \mid i \neq 1\} = a_1 + \Sigma\{a\,(b_i/b_1.(a_1-(a_1)^2)) \mid i \neq 1\} \;;$$

où il est légitime d'utiliser la fonction a pour $i\neq 1$, parce que ces a_i sont $\leq 1/2$.

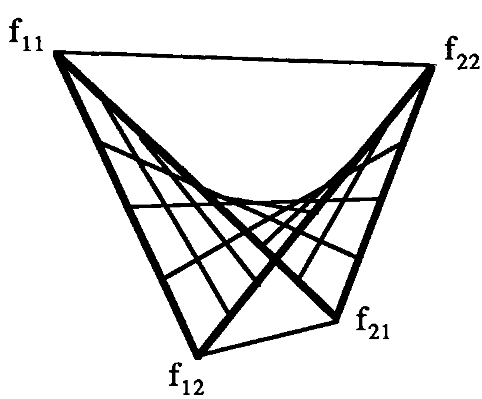
Dans $\Sigma(a_1)$, le terme initial, a_1, varie de 0 à 1; chacun des termes de rang $i\neq 1$ est une fonction convexe de a_1 symétrique (le graphe en est un arc d'ellipse), i.e. prenant la même valeur pour $a_1=u$, quelconque, et pour $a_1=1-u$; maxima et de dérivée nulle pour $a_1=1/2$; croissante et de dérivée (b_i/b_1) pour $a_i=0$; donc décroissante et de dérivée $-(b_i/b_1)$ pour $a_i=1$. De ce que $b_1<1/2$, ou $b_1<\Sigma\{b_i|i\neq 1\}$, il résulte que la dérivée de la somme des termes en i de Σ est supérieure à 1 pour $a_1=0$ et inférieure à -1 pour $a_1=1$.

En somme, $\Sigma(a_1)$ est une fonction convexe dont la dérivée est >2 pour $a_1=0$; égale à 1 pour $a_1=1/2$; négative pour $a_1=1$: ceci suffit à établir que $\Sigma(a_1)$



prend une fois et une seule la valeur 1 quand a_1 est dans [0, 1[; et achève la démonstration. On notera que, selon que $\Sigma(1/2)$ est >1 ou <1, on est dans le cas (-) ou dans le cas (+) pour a_1.

6 Variation de la trace pour un tableau carré 2 × 2



f_{11}	f_{12}
f_{21}	f_{22}

Dans le cas où cardI = cardJ = 2, l'ensemble des lois sur I×J s'identifie à un tétraèdre dont les sommets correspondent aux 4 masses ponctuelles (i, j), donc aussi aux 4 termes de la matrice. Sur la figure, on a marqué f_{ij} le sommet qui représente une loi dont tous les termes sont nuls, excepté f_{ij} lui-même, qui vaut 1. Les lois de trace nulle ne sont autres que celles pour lesquelles s'annule le déterminant $(f_{11}\, f_{22}) - (f_{12}\, f_{21})$.

La surface décrite par ces lois est donc une portion de paraboloïde hyperbolique (PH), limitée par le quadrilatère gauche (en gras) dont les arêtes successives sont (f_{11}, f_{12}), (f_{12}, f_{22}), (f_{22}, f_{21}), (f_{21}, f_{11}). Ces arêtes correspondent aux couples de termes constituant une ligne ou une colonne de la matrice: en effet, celle-ci a déterminant nul si s'annule l'une de ses lignes ou l'une de ses colonnes; e.g., l'arête (f_{11}, f_{12}) contient les lois représentées par une matrice dont la 2-ème ligne est nulle: $f_{21} = f_{22} =0$.

En revanche, la trace est non nulle (elle vaut 1) sur l'arête (f_{11}, f_{22}) qui représente l'ensemble des lois diagonales (pour lesquelles $f_{12} = f_{21} = 0$); et de même pour l'arête opposée (f_{12}, f_{21}).

Les deux familles de génératrices rectilignes du PH, sont définies, respectivement par la condition que le rapport entre les deux lignes, ou les deux colonnes de la matrice ait une valeur déterminée. Par exemple, sur une génératrice dont la direction est parallèle à la drection de plan définie par (f_{11}, f_{12}), arête lieu des lois dont la 2-ème ligne est nulle, et (f_{22}, f_{21}), arête lieu des lois dont la 1-ère ligne est nulle, le rapport de la 1-ère à la 2-ème ligne a une valeur déterminée.

Sur la figure, on a représenté, dans le tétraèdre, quelques–unes de ces génératrices. Par un heureux effet de perspective, ces génératrices font voir la forme en selle du PH: on peut imaginer un cavalier assis entre f_{11} et f_{22}, tandis que la courbure entre f_{12} et f_{21} repose sur le dos de sa monture!

Le cas d'une seule donnée manquante, cas étudié au §4, est ici très simple. Supposons, e.g., que la donnée manquante soit f_{11}; la connaissance des trois autres f_{ij} définit un point f_{IJ} sur la face (f_{12}, f_{22}, f_{21}): cette face est, en effet, le lieu géométrique des lois pour lesquelles f_{11} est nul; et elle est opposée au sommet f_{11} du tétraèdre ; on cherche donc le minimum de la trace sur le segment joignant ce sommet à f_{IJ}; ce minimum est zéro, car le segment recoupe le PH.

$$3.t(x) = \begin{vmatrix} 3x & 1-x \\ & \\ 1-x & 1-x \end{vmatrix}$$

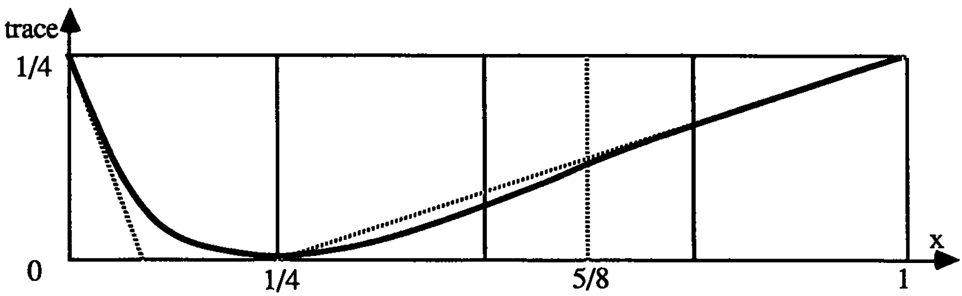
Mais il vaut la peine de reprendre ici, dans un cas particulier, les calculs du §4. Il vient:

$$\text{trace}(x) = ((7x^2 + x + 1) / (1 + 2x)^2) - (3/4) \;;$$

$$\text{trace}(x)' = 3\,(4x - 1) / (1 + 2x)^3 \;;$$

$$\text{trace}(x)'' = 6\,(5 - 8x) / (1 + 2x)^4 \;;$$

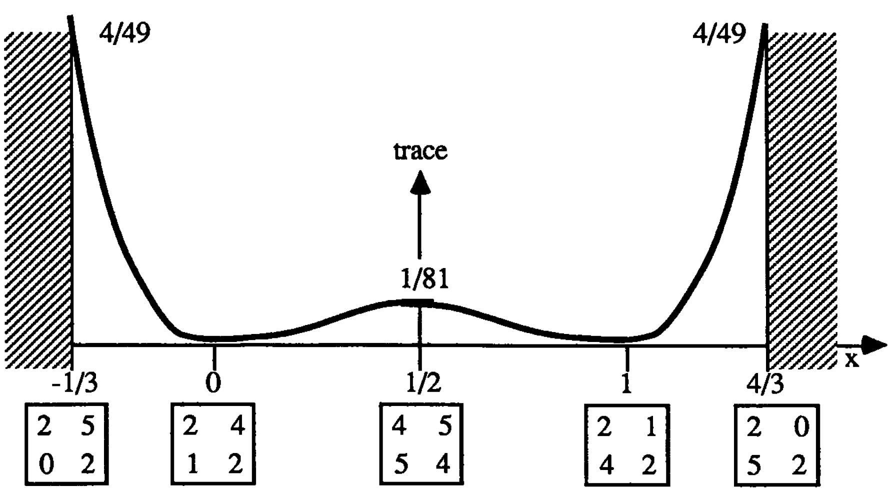
la valeur minima est atteinte pour x=1/4: les quatre termes de t(x) sont alors égaux, et la trace est nulle.



On a représenté sur un graphique la variation de trace(x): on remarque que, pour x=5/8, la courbe de la trace présente une inflexion: la trace n'est donc pas une fonction concave.

On voit, d'autre part, aisément sur un autre exemple que le critère peut avoir deux minima sur un segment de droite du simplexe S(I × J). Il suffit de choisir un segment qui rencontre en deux points la portion de PH sur laquelle la trace s'annule. Dans la figure publiée ici, on a, sous chacune des cinq valeurs de x {-1/3, 0, 1/2, 1, 4/3}, placé la matrice qui lui correspond: à ceci près que, pour plus de simplicité, sont écrites, non des lois de probabilité, mais des matrices entières qui leur sont proportionnelles. Il y a deux minima nuls de la trace, pour x=0 et x=1; un maximum local pour x=1/2; et des maxima à la frontière du simplexe, quand s'annule l'une des 4 composantes.

$$9.t(x) = \begin{vmatrix} 2 & 4-3x \\ 1+3x & 2 \end{vmatrix}$$



7 Problèmes ouverts

Il convient de montrer, en conclusion, les recherches à poursuivre.

$$\left| \begin{array}{ccc} 1 & 0 & ? \\ & & \\ 0 & 1 & ? \end{array} \right|$$

Partons de l'exemple du tableau ci-dessus, où deux données manquantes sont signlées par '?': il est clair que, si l'on met des quantités finies, la trace du tableau reconstitué sera strictement positive, les deux lignes n'étant pas proportionnelles; cependant, si l'on met, dans les deux cases, un même x, arbitrairement grand, la trace sera arbitrairement voisine de zéro. Nous dirons, en général, que le problème de reconstitution n'admet qu'une solution triviale, si, en reprenant la formule du §2.2, $t_{IJ} = (1-\alpha) f_{IJ} + \alpha\, \mu_{IJ}$, où μ_{IJ} désigne une loi de probabilité dont le support est l'ensemble A des lacunes, le minimum du critère n'est atteint que pour α tendant vers l'infini, avec μ_{IJ} convenable.

On est certainement dans ce cas trivial si, comme dans l'exemple, il manque une colonne entière, (ou une ligne entière): car il suffit, pour faire tendre la trace vers zéro, de remplacer la colonne manquante par le produit de la somme des autres colonnes par un nombre arbitrairement grand; en effet, le nuage N(J) est alors de plus en plus concentré en son centre de gravité (où est la colonne créée), sans que les distances varient. Il convient donc de chercher, en général, à quelles conditions doit satisfaire l'ensemble A des lacunes pour que le problème de la reconstitution admette une solution non triviale.

Dans ce cas, nous pensons que le problème de la reconstitution pourra être résolu par un algorithme analogue à celui de la méthode du gradient conjugué. La structure particulière du problème laisse, d'autre part, espérer une solution rapide si l'on applique itérativement la méthode donnée au §4 pour le cas d'une seule donnée manquante. Supposons, en effet, que les éléments de A soient numérotés de a_1 jusqu'à a_n: on reconstituera d'abord, successivement, une par une, les données manquantes, de la 1-ère à la n-ème; et l'on fera autant de cycles qu'il convient pour que, dans sa décroissance, le critère soit devenu stationnaire à un seuil près qu'on aura choisi.

Quant à l'unicité, dans le cas non trivial, les exemples du §6 nous interdisent de la postuler à la légère.