Les Cahiers de l'Analyse des Données
Vol. XXI - 1996 - n°3 - pp.271-274.

RAPPEL : CALCUL TENSORIEL ET CALCUL MATRICIEL EN ANALYSE DES DONNÉES

[MATR. TENS.]

*N. El KHOURY**
*T. AKIKI**

Dans l'enseignement de l'analyse des données, il est de règle de mettre en préliminaire un exposé du calcul tensoriel des mesures et fonctions, en vue des applications à l'étude des lois de probabilité sur un produit de deux ensembles finis. Il nous est cependant apparu que la pratique du calcul pouvait d'autant plus faire oublier ces principes de base que le calcul matriciel, avec ses ambiguïtés, jouit en statistique d'une possession invétérée. D'où des erreurs que le présent rappel vise à prévenir.

Les exposés en forme étant entre les mains des lecteurs de *CAD*, nous préférons ici nous appuyer sur un exemple où les calculs numériques, dans leur simplicité, montrent concrètement de quoi il s'agit.

1 Un exemple de tableau: lois et transitions associées

tableau

3	ja	jb	jc
iA	1	1	1
iB	1	2	2
iC	2	1	1

Le présent tableau met en correspondance deux ensembles I et J, dont chacun comprend trois éléments: I = {iA, iB, iC}; J = {ja, jb, jc}.

En le considérant comme une mesure sur le produit I × J, on peut, pour plus de clarté, noter ce tableau : $k_{IJ} = \{k_{ij} \mid i \in I, j \in J\}$; en particulier, l'élément k(iA, jc) qui, dans la présentation matricielle est à l'intersection de la ligne iA et de la colonne jc, s'écrit : $k_{iA,jc} = 1$; où la virgule est introduite afin de séparer les deux indices iA et jc.

Du point de vue de l'analyse des données, d'après ce tableau de correspondance, k_{IJ}, il convient de calculer la loi de probabilité, p_{IJ} ; qui s'en déduit en multipliant chacune des composantes par un coefficient qui est l'inverse du total de celles-ci; soit, ici, (1/12). Par exemple, $p_{iA,jc} = 1/12$.

(*) Université Libanaise ; Beyrouth ; LIBAN. télécopie : 961.1.491765 .

(p_{IJ})

	ja	jb	jc
iA	1/12	1/12	1/12
iB	1/12	2/12	2/12
iC	2/12	1/12	1/12

Du fait de la présentation en tableau, les ensembles I et J, reçoivent des rôles dissymétriques. Mais, pour le calcul tensoriel, tous deux ont même valeur: car p_{IJ} est un élément du produit tensoriel de l'espace, R_I, des mesures sur I, par l'espace, R_J, des mesures sur J : $p_{IJ} \in R_I \otimes R_J$.

À la loi p_{IJ}, sont associées une ligne de marge, p_J, et une colonne de marge, p_I, qui sont, respectivement, la loi marginale sur J et la loi marginale sur I: e.g. : $p_I = \{p_{iA}, p_{iB}, p_{iC}\} = \{1/4, 5/12, 1/3\} \in R_I$.

Cependant dans la définition et le calcul des facteurs, un rôle essentiel est dévolu à deux autres entités tensorielles, indicées également par le produit $I \times J$, et qu'on peut présenter sur un tableau dont les lignes et colonnes reçoivent leur nom de J ou de I; pourvu que cette présentation ne fasse pas oublier de quel tenseur il s'agit.

$(p_J{}^I)$

	iA	iB	iC
ja	1/3	1/5	1/2
jb	1/3	2/5	1/4
jb	1/3	2/5	1/4

Dans la transition, $p_J{}^I$, de I vers J, sous la présentation adoptée ici, chaque colonne, $p_J{}^i$, qui n'est autre que le quotient, par son total, d'une ligne de la loi p_{IJ}, ou du tableau de base k_{IJ}, est une loi sur l'ensemble J, associée à l'élément particulier, i, de I : $p_J{}^i \in R_J$.

Globalement on écrira: $p_J{}^I \in R_J \otimes R^I$; pour signifier que $p_J{}^I$ est à la fois mesure relativement à I et fonction relativement à J: il s'agit, en bref, d'une mesure sur J: $p_J{}^i \in R_J$, qui est fonction de $i \in I$.

Compte tenu des habitudes acquises dans la multiplication des matrices, on a cru bon de présenter dans le tableau de transition, les mesures en colonne. Mais ce choix ne peut suffire à éviter de faire des calculs absurdes: d'ailleurs, pour p_{IJ}, qui est doublement mesure, à la fois sur I et sur J, on ne pouvait présenter simultanément en ligne les mesures p_{iJ} et les p_{Ij} qui sont ici, respectivement, lignes et colonnes (pour ne rien dire des tenseurs à plus de deux indices!).

Il convient donc, quelle que soit la disposition des composantes, de garder présente à l'esprit la nature tensorielle de chaque tableau.

$(p_I{}^J)$

	ja	jb	jc
iA	1/4	1/4	1/4
iB	1/4	1/2	1/2
iC	1/2	1/4	1/4

De même dans la transition, $p_I{}^J$, de J vers I, chaque colonne, $p_I{}^j$, qui n'est autre que le quotient, par son total, d'une colonne de la loi p_{IJ}, ou du tableau de base k_{IJ}, est une loi sur l'ensemble I, associée à l'élément particulier, j, de J : $p_I{}^j \in R_I$.

2 Formules générales de composition des transitions mesures et fonctions

Une transition de I vers J peut servir, d'une part, à transporter toute mesure μ_I (sur I) pour obtenir une mesure π_J sur J; et, d'autre part, par voie contravariante, à associer à toute fonction G^J (sur J) une fonction F^I sur I.

De façon précise, on pose:

$$\pi_J = p_J{}^I \circ \mu_I \;;\; \pi_J = \{\pi_j \mid j \in J\} \;;\; \forall\, j \in J : \pi_j = \Sigma\{p_j{}^i \,.\, \mu_i \mid i \in I\} \quad ;$$

$$F^I = G^J \circ p_J{}^I \;;\; F^I = \{F^i \mid i \in I\} \;;\; \forall\, i \in I : F^i = \Sigma\{G^j \,.\, p_j{}^i \mid j \in J\} \quad ;$$

Sans reprendre un exposé complet, nous noterons que, dans les formules, on somme par rapport à un indice, i ou j, qui, dans l'expression à sommer, figure deux fois: d'une part, en position supérieure, et, d'autre part, en position inférieure. Ceci correspond au principe de dualité suivant lequel une fonction s'intègre par rapport à une mesure. Une sommation par rapport à un indice répété deux fois en haut, ou deux fois en bas, n'a pas de sens invariant; mis à part le cas, considéré à la fin du §3, où, du fait d'une structure euclidienne, on a pu prendre des coordonnées orthonormées sur l'espace, identifié à son dual.

Compte tenu des deux interprétations, ci-dessus proposées, de $p_J{}^I$ comme application linéaire, on peut écrire:

$$p_J{}^I \;\in\; R_J \otimes R^I \;\approx\; L(R_I, R_J) \;\approx\; L(R^J, R^I) \quad ;$$

et définir, de deux façon différentes, la composition des transitions $p_I{}^J$ et $p_J{}^I$:

$$\begin{aligned}
c_I{}^I &= p_I{}^J \circ p_J{}^I \quad ; \quad c_i{}^{i'} = \Sigma\{p_i{}^j \,.\, p_j{}^{i'} \mid j \in J\} \quad ; \\
c_J{}^J &= p_J{}^I \circ p_I{}^J \quad ; \quad c_j{}^{j'} = \Sigma\{p_j{}^i \,.\, p_i{}^{j'} \mid i \in I\} \quad ; \\
c_I{}^I &\in R_I \otimes R^I \approx L(R_I, R_I) \approx L(R^I, R^I) \quad ; \\
c_J{}^J &\in R_J \otimes R^J \approx L(R_J, R_J) \approx L(R^J, R^J) \quad ;
\end{aligned}$$

en particulier, la transition $c_I{}^I$ peut servir à définir les facteurs F^I sur I, et G^J sur J par des équations aux vecteurs propres:

$$F^I \circ c_I{}^I = \lambda \,.\, F^I \quad ; \quad G^J \circ c_J{}^J = \lambda \,.\, G^J \quad ;$$

avec entre les deux goupes de facteurs les relations:

$$G^J \circ p_J{}^I = \sqrt{(\lambda)} \,.\, F^I \quad ; \quad F^I \circ p_I{}^J = \sqrt{(\lambda)} \,.\, G^J \quad ;$$

On notera que dans toutes ces formules, comme au §1, l'on somme par rapport à un indice répété deux fois; et qu'un indice ne subsiste dans le résultat que s'il n'a pas fait l'objet d'une sommation.

$$(F^{iA} \quad F^{iB} \quad F^{iC}) \begin{vmatrix} c_{iA}{}^{iA} & c_{iA}{}^{iB} & c_{iA}{}^{iC} \\ c_{iB}{}^{iA} & c_{iB}{}^{iB} & c_{iC}{}^{iC} \\ c_{iC}{}^{iA} & c_{iC}{}^{iB} & c_{iC}{}^{iC} \end{vmatrix} = \lambda \quad (F^{iA} \quad F^{iB} \quad F^{iC})$$

3 Transitions composées et équation des facteurs pour l'exemple

	iA	iB	iC
iA	1/12 + 1/12 + 1/12	1/20 + 1/10 + 1/10	1/8 + 1/16 + 1/16
iB	1/12 + 1/6 + 1/6	1/20 + 1/5 + 1/5	1/8 + 1/8 + 1/8
iC	1/6 + 1/12 + 1/12	1/10 + 1/10 + 1/10	1/4 + 1/16 + 1/16

Voici, dans le cas de l'exemple du §1, le calcul complet de la matrice de la transition $c_I{}^I$. On vérifie que cette matrice admet pour valeurs propres (1, 1/15, 0); i.e. d'abord la valeur propre triviale; et ensuite, une seule valeur propre non nulle, parce que le tableau de base est dégénéré (de rang 2). En prenant garde que, pour respecter les lois de sommation entre indices haut et bas, il s'agit, selon le langage du calcul matriciel, d'une multiplication de ligne (facteur: à gauche) par colonne (de la transition: à droite), on trouve sans peine que {0 , 4 , -5} est facteur relatif à la valeur propre non triviale (1/15). Ce facteur est, comme il se doit, de moyenne nulle pour la loi marginale p^I = {(1/4), (5/12), (1/3)}; et on peut l'amener à avoir la variance λ = (1/15).

Si l'on cherche un vecteur colonne qui soit (à droite) vecteur propre de $c_I{}^I$ pour la valeur propre (1/15), on trouve {0, 1, -1}; qui ne peut être un facteur, car ce n'est pas une fonction, mais une mesure sur I. Toutefois, l'on vérifie que la densité de cette mesure relativement à la loi marginale, p_I, est une fonction proportionnelle au facteur déjà trouvé.

Dans la pratique de l'analyse numérique, on procède autrement qu'on ne l'a fait ici. Au lieu de la matrice non symétrique, $c_I{}^I$, on diagonalise la matrice symétrique des $s(i, i') = c_i{}^{i'}. \sqrt{(p_{i'})}/\sqrt{(p_i)}$. Du fait de la symétrie, il n'y a plus lieu de distinguer entre vecteurs propres à droite et à gauche. Mais les vecteurs propres obtenus $\{f(i)\}$ ne sont pas exactement les facteurs cherchés. Il reste à les diviser par $\sqrt{(p_i)}$: on a, à une constante près: $F^i \approx f(i)/\sqrt{(p_i)}$. Du point vue géométrique, les changements de coordonnées effectués ici, avec, pour coefficients, les $\sqrt{(p_i)}$, servent à rejoindre la structure euclidienne, dans laquelle la variance du facteur F^I n'est autre que la somme des carrés des $f(i)$.

Pour les lecteurs souhaitant vérifier, voici, enfin, le listage du facteur!

SIGI	QLT	PDS	INR	F1	CO2	CTR
iA	0	250	0	0	0	0
iB	1000	417	444	283	1000	444
iC	1000	333	556	-354	1000	556

SIGJ	QLT	PDS	INR	F1	CO2	CTR
ja	1000	333	667	-387	1000	667
jb	1000	333	167	194	1000	167
jc	1000	333	167	194	1000	167