LES CAHIERS DE L'ANALYSE DES DONNÉES

V. CHOLAKIAN

Relations entre les termes d'interaction supérieurs et les traces des correspondances binaires associées à une correspondance multiple

Les cahiers de l'analyse des données, tome 10, n° 1 (1985), p. 85-90

<http://www.numdam.org/item?id=CAD_1985__10_1_85_0>

Les Cahiers de l'Analyse des Données
Vol. X - 1985 - nº 1 - pp. 85-90

RELATIONS ENTRE LES TERMES D'INTERACTION SUPÉRIEURS ET LES TRACES DES CORRESPONDANCES BINAIRES ASSOCIÉES A UNE CORRESPONDANCE MULTIPLE

[TRACE INT. MULT.]

*par V. Cholakian**

1 Le cas d'une correspondance ternaire

1.1 Présentation des relations dans le cas d'une correspondance ternaire

On rencontre fréquemment, notamment dans les statistiques économiques des tableaux de correspondance ternaire I.J.T. Ces données sont traitées en analysant d'une part les tableaux de marges binaires IJ, IT, JT ; d'autre part le tableau ternaire lui-même considéré comme un tableau binaire, e.g. (IT).J ; ce dernier tableau pouvant être soit analysé par lui-même, soit adjoint en lignes supplémentaires au tableau I.J. Ces multiples analyses suggèrent des comparaisons entre les traces, c'est-à-dire entre les inerties des divers nuages. Par exemple aux trois tableaux IJ, TJ, (IT)J dont l'ensemble des colonnes est J, sont associés trois nuages de profils sur J : N(I), N(T), N(IT) (avec dans R_J la même métrique du chi2 de centre F_J), l'inertie de chacun des nuages étant la trace d'une correspondance binaire, (trace que dans la suite on notera "int" ; interaction) : une première conjecture serait que l'inertie de N(IT) soit la somme de celles de N(I) et de N(T) ; i.e. avec la notation "int" :

$$\text{int}(IT,J) =? \text{int}(I,J) + \text{int}(T,J).$$

Ainsi qu'il apparaîtra bientôt, cette formule n'est vraie que sous des conditions très restrictives ; en revanche nous donnerons d'une part une formule approchée (où le terme ternaire int(I,J,T) sera défini ultérieurement)

$$\text{int}(IT,J) \approx \text{int}(I,J) + \text{int}(T,J) + \text{int}(I,J,T),$$

formule qui est même une identité si int(I,T) = 0 ; et d'autre part une formule explicite, plus complexe, qui sera calculée dans certains cas (§ 3).

Quant aux démonstrations, il nous a paru avantageux de les faire dans le cas général d'une correspondance multiple (§ 2) en sorte que l'exposé du cas ternaire (sur lequel on reviendra au § 3), n'est qu'une introduction ; écrite avec des notations plus légères mais moins puissantes que celles requises pour le cas général.

L'auteur remercie Monsieur le Professeur J.P. Benzécri pour ses conseils.

Ce travail a été subventionné par le conseil de Recherche en Sciences Naturelles et en Génie du Canada et par le conseil de Recherche de l'Université de Moncton.

(*) *Docteur 3° cycle en statistique. Université de Moncton. Canada.*

1.2 Notations et rappels sur la décomposition des interactions :

L'objet étudié est une loi de probabilité f_{IJT} sur le produit de trois ensembles finis I,J,T.

$$f_{IJT} = \{f_{ijt} | i \in I \ ; \ j \in J \ ; \ t \in T\}.$$

Les lois marginales simples (sur un ensemble) sont notées f_I, f_J, f_T ; et de même f_{IJ}, f_{IT}, f_{JT}, pour les marges binaires :

$$f_i = \Sigma\{f_{i,j,t} | j \in J, \ t \in T\} \ ; \quad f_I = \{f_i | i \in I\} \ ;$$

$$f_{jt} = \Sigma\{f_{i,j,t} | i \in I\} \ ; \qquad f_{JT} = \{f_{JT} | j \in J \ ; t \in T\}$$

Par densité d'une loi, on entend toujours : "densité par rapport au produit des marges simples"; e.g. :

$$d^{ijt} = f_{ijt}/(f_i f_j f_T) \ ; \ d^{ij} = f_{ij}/(f_i f_j) \ ; \ \text{etc..}$$

Si l'on fait choix sur les ensembles I, J, T des bases de fonctions orthonormées (par rapport au produit scalaire défini par les lois simples f_I, f_J, f_T ; la fonction constante 1 étant la première fonction $\varphi_\circ$ de base ; e.g.

$$\Sigma\{\varphi^i_\alpha \ \varphi^i_{\alpha'} \ f_i | \ i \in I\} = \delta_{\alpha\alpha} \ ; \ \Sigma\{ \varphi^i_\alpha \ f_i | \ i \in I\} = \delta_{\alpha\circ}.),$$

toute fonction réelle sur IJT (ou sur le produit de deux ensembles de base peut être décomposée en un terme constant (la moyenne), des termes simples (dépendant d'une seule variable) ; des termes doubles (dépendant de deux variables) ; et enfin un terme triple : par exemple pour une fonction :

$$\psi^{ijt} = \psi^{\circ\circ\circ} + \Sigma_\alpha \ \psi^\alpha \varphi^i_\alpha + \Sigma_\beta \ \psi^\beta \varphi^j_\beta + \Sigma_\gamma \ \psi^\gamma \varphi^t_\gamma +$$

$$\Sigma_{\alpha\beta} \ \psi^{\alpha\beta} \varphi^i_\alpha \ \varphi^j_\beta + \Sigma_{\beta\gamma} \psi^{\beta\gamma} \ \varphi^j_\beta \ \varphi^t_\gamma + \Sigma_{\gamma\alpha} \psi^{\gamma\alpha} \ \varphi^t_\gamma \varphi^i_\alpha + \Sigma_{\alpha\beta\gamma} \psi^{\alpha\beta\gamma} \varphi^i_\alpha \varphi^j_\beta \varphi^t_\gamma \ ;$$

dans ces sommes, les indices α, β, γ sont toujours supposés différents de 0 ; en sorte que, e.g., $\varphi^j_\beta \ \varphi^t_\gamma$ (dont le coefficient est $\psi^{\beta\gamma}$) est une fonction produit qui dépend véritablement des deux variables j et t. Comme le démontre A. Bener (cf. [*INTER. CORR. MULT.*], *in* CAD, Vol VII, n° 1 ; 1982) les huit termes ne dépendent pas des bases de fonctions particulières choisies, seules interviennent les lois marginales simples f_I, f_J, f_T. Nous utiliserons ces décompositions de fonctions uniquement pour la densité d^{ijt} et les densités des lois marginales binaires. Dans ces cas, précisément parce que les mesures de références f_I, f_J, f_T sont les marges de ces lois, il n'y a pas de terme dépendant d'une seule variable (l'analogue de $\Sigma_\alpha \psi^\alpha \varphi^i_\alpha$ est nul) et de plus (parce que f_{IJ} est marge f_{IJT} ; etc.) le terme en ij de d^{ijt} n'est autre que celui de d^{ij} et de même pour les autres couples de variables. On écrit donc :

$$d^{ijt} = 1 + r^{ij} + r^{jt} = r^{ti} + r^{ijt} \ ;$$

$$d^{ij} = 1 + r^{ij} \ ; \ d^{jt} = 1 + r^{jt} \ ; \ d^{ti} = 1 + r^{ti} \ ;$$

la lettre r a été choisie comme l'initiale de "relation" ; r^{ij} représente l'écart à l'indépendance pour la loi marginale f_{IJ} sur IJ ; et la trace de cette correspondance n'est autre que :

$$\| f_{IJ} - f_I f_J \|^2_{fIfJ} = \mathrm{int}(I,J) = \Sigma\{(r^{ij})^2 f_i f_j] i \in I \ ; \ j \in J\} \ ;$$

de même pour r^{it} et r^{ti} (notons au passage qu'on écrira indifféremment dans la suite r^{ti} ou r^{it} selon l'ordre où sont considérés les ensembles I, J, T) pour le terme ternaire on écrirá de même :

$$\mathrm{int}(I,J,T) = \Sigma\{(r^{ijt})^2 \ f_i f_j f_t \,|\, i \in I \ ; \ j \in J \ ; \ t \in T\}.$$

Enfin nous rappelons que, considérés comme des fonctions sur $I \times J \times T$ (muni de la loi produit $f_I f_J f_T$) les différents termes r^{ij}, r^{ti}, r^{jt}, r^{ijt} sont orthogonaux entre eux) i.e. leurs produits deux à deux ont moyenne nulle), ce qui en bref, résulte du fait que les φ^I_α, φ^J_β, φ^T_γ sont des fonctions de moyenne nulle .

1.3 *Calcul de int(IT,J)* : On a par définition :

$$\begin{aligned} \mathrm{int}(IT,J) &= \| f_{IT,J} - f_{IT}.f_J \|^2_{fIT.fJ} \\ &= \Sigma\{(f_{ijt} - f_{it} f_j)^2/(f_{it} f_j) \,|\, i \in I \ ; \ j \in J \ ; \ t \in T\} \\ &= \Sigma\{f_i f_j f_t (d^{ijt} - d^{it})^2/d^{it} \,|\, i \in I \ ; \ j \in J \ ; \ t \in T\} \\ &= \Sigma\{f_i f_j f_t (r^{ijt} + r^{ij} + r^{tj})^2/(1+r^{it}) \,|\, i \in I \ ; \ j \in J \ ; \ t \in T\}. \end{aligned}$$

Dans le calcul de cette expression, nous considérerons divers cas particuliers :

1.3.1 *Indépendance entre I et T* : On a alors $f_{it} = f_i f_t$, ou encore $r^{it} = 0$; il n'y a plus de dénominateur $(1+r^{it})$; d'où :

$$\begin{aligned} \mathrm{int}(IT,J) &= \Sigma\{f_i f_j f_t (r^{ijt} + r^{ij} + r^{tj})^2 \,|\, i \in I \ ; \ j \in J \ ; \ t \in T\}. \\ &= \mathrm{int}(I,J) + \mathrm{int}(T,J) + \mathrm{int}(I,J,T). \end{aligned}$$

(égalité qui résulte de ce que, comme on l'a rappelé les termes r^{ijt}, r^{ij}, r^{tj} sont orthogonaux entre eux.

1.3.2 *Développement à l'ordre 2* : La même formule se retrouve si on suppose que tous les termes r de d^{ijt} sont des infiniments petits d'ordre 1, et qu'on cherche le terme principal du développement de int(IT,J) : alors $(r^{ijt} + r^{ij} + r^{tj})^2$ est un infiniment petit d'ordre 2, et le coefficient $(1 + r^{it})^{-1}$ est équivalent à 1. On a donc :

$$\mathrm{int}(IT,J) = \mathrm{int}(I,J) + \mathrm{int}(T,J) + \mathrm{int}(I,J,T) + 0^3(r) = \mathrm{int}^2(IT,J) + 0^3(r)$$

1.3.3 *Calcul du terme d'ordre 3* : Le développement usuel $(1 + r)^{-1} = (1 - r + r^2 \ldots)$, donne ici :

$$\begin{aligned} \mathrm{int}(IT,J) = {} & \mathrm{int}(I,J) + \mathrm{int}(T,J) + \mathrm{int}(I,J,T) - \\ & \Sigma\{ f_i f_j f_t (r^{ijt} + r^{ij} + r^{tj})^2 r^{it} \ i,j,t \} + 0^4(r) \ ; \end{aligned}$$

en développant le carré, on trouve six termes, certains multipliés par r^{it} ont moyenne nulle ; par exemple $(r^{ij})^2 r^{it}$ a moyenne nulle en t, comme r^{it} ; finalement il reste, comme terme du 3-ème ordre, int^3:

$$int^3(IT,J) = - \Sigma\{f_i f_j f_t\, r^{it}(2r^{ij}\, r^{tj} + r^{ijt}(r^{itj}+2(r^{ij}+r^{tj})))\,|\, i,j,t\}$$

Cette formule ne contient plus qu'un terme dans le cas particulier où l'intersection ternaire est absente (i.e. $\forall$ i, j, t : $r^{ijt} = 0$). Au § 3 nous ferons ces calculs dans des cas très simples de correspondance ternaire, afin de vérifier que ces termes peuvent effectivement être non nuls . Dès maintenant on peut noter que puisque au-delà de int^2 on a un terme de 23, ce dernier peut avoir un nombre quelconque (positif ou négatif).

2 *Cas général d'une correspondance multiple*

2.1 *Notations d'appel* : Comme A. Bener (*loc. cit.*), nous prenons comme exemple typique d'une correspondance multiple un questionnaire Q , avec pour chaque question q , un ensemble Jq de modalités de réponse. On note donc :

$$JQ = \Pi\{Jq\,|\,q \in Q\} \quad ; \quad jQ = \{jq\,|\,q \in Q\} \;;$$

un élément jQ de JQ n'est autre qu'un système de réponse à toutes les questions (réponse jq à la question q ; etc.). On considère une loi de probabilité f_{JQ}, qui donne la probabilité relative f_{jQ} de chaque système de réponses. Il y a sur chaque Jq une loi marginale f_{Jq} et des lois marginales sur chaque produit partiel d'un sous ensemble X du questionnaire ; on note :

$$X \subset Q \;:\; JX = \Pi\{Jq\,|\,q \in X\} \;;\; jX = \{jq\,|\,q \in X\} \;;$$

$$d^{jX} = f_{jX}/\, \Pi\{f_{jq}\,|\,q \in X\}.$$

Au § 1, on a considéré $IJT = IT \times J$, ici plus généralement on prendra :

$$Q = A \cup B \;;\; A \cap B = \emptyset \;;\; JQ = JA\,.\,JB\;;$$

par exemple, s'il y a 7 questions q , JA pourra être le produit de 3 des Jq ; et JB le produit des 4 autres. La décomposition des densités en termes dépendant d'un sous-ensemble déterminé de variables se fait de manière unique, comme au § 1.2 :

$$d^{jQ} = \Sigma\{r^{jX}\,|\,X \subset Q\}$$

dans cette formule, le terme correspondant à $X = \emptyset$ (terme qui ne dépend d'aucune variable) est la constante 1 ; et les termes dépendant d'une seule variable ($X = \{q\}$) sont nuls ; on peut donc récrire :

$$d^{jQ} = 1 + \Sigma\{r^{jX}\,|\,X \subset Q \;:\; \mathrm{Card}\, X \geq 2\} \;;$$

avec entre les r^{jX} les propriétés d'orthogonalité usuelles.

2.2 *Calcul de int(JA,JB)* : Il s'agit simplement de la trace de f_{JQ} considérée comme une correspondance binaire usuelle sur le produit des ensembles JA et JB :

$$int(JA,JB) = \Sigma\{(f_{jQ} - f_{jA}f_{jB})^2/(f_{jA}f_{jB})\,|\, jQ \in JQ\}.$$

Comme au § 1.3, on transforme la formule en introduisant les densités totale et partielle :

$$int(JA,JB) = \Sigma\{\Pi\{f_{jq} | q \in Q\} \; (d^{JQ} - d^{JA} d^{JB})^2 (d^{JA} d^{JB})^{-1} | jQ \in JQ\}.$$

Pour le développement en r, on note que le dénominateur $(d^{JA}d^{JB})$ est équivalent à 1 ; tandis que la fonction élevée au carré, qui constitue le numérateur est d'ordre 1 en r ; et peut s'exprimer comme suit, en écartant les termes de $^{\underline{o}}$ 2 en r :

$$d^{JQ} - d^{JA} d^{JB} \approx \Sigma\{r^{jV} | V \subset Q \; ; \; V \cap B \neq \emptyset \; ; \; V \cap A \neq \emptyset\}$$

en effet les termes r^{jV} pour V inclus dans A, (ou $V \subset B$), disparaissent parce qu'ils sont à la fois dans d^{JQ} et $d^{JA}d^{JB}$. On a donc :

$$int(JA,JB) \approx \Sigma\{\|r^{JV}\|^2 \mid V \subset Q \; ; \; V \cap B \neq \emptyset \; ; \; V \cap A \neq \emptyset\} + 0^3(r).$$

(où on a noté $\|r^{JV}\|^2 = \Sigma\{\Pi\{f_{jq} \; q \in V\}(r^{jV})^2 | jV \in JV\})$

Le calcul du terme d'ordre 3 peut se faire, mais il est plus complexe qu'au § 1.3 : parce que, ici il faut tenir compte d'une part de ce que $(d^{JA} d^{JB})^{-1} = 1 + 0^1(r)$ et d'autre part de ce que $(d^{JQ} - d^{JA}d^{JB})$ comporte des termes du degré 2 donc $(d^{JQ} - d^{JA} d^{JB})$ a des termes du $^{\underline{o}}$ 3 ; (au § 1.3, l'analogue de d^{JB} est 1 !).

Comme au § 1.3, int(JA,JB) se réduit à son terme d'ordre 2 en r s'il y a indépendance à l'intérieur de chacun des deux groupes de variables A et B , i.e. si : $d^{JA} \equiv 1$; $d^{JB} \equiv 1$.

3 *Exemples de correspondance ternaire*

3.1 *Correspondance sur le cube* : On pose :

$$I = J = T = \{+1, -1\} \; ; \; \forall \, i,j,t \; : \; f_i = f_j = f_t = 1/2.$$

Dans ce cas sur I, J, T, il y a une base de deux fonctions orthonormées : d'une part la constante 1 : $\varphi_0 = (1,1)$; d'autre part la fonction de moyenne nulle $\varphi_1 = (+1,-1)$; comme nous avons identifié les i,j,t à des nombres (+1 ou -1), on peut écrire :

$$d^{ijt} = 1 + a\,i\,j + b\,j\,t + c\,t\,i + e\,i\,j\,t$$

où a, b, c, e sont des nombres réels (auxquels on incorpore seulement d'être tels que la densité soit partout positive ou nulle). Voici , à titre de précision le tableau des $d^{ijt} = 8\, f_{ijt}$:

	it	j = 1	j = -1
IT	++	1 + a + b + c + e	1 - a - b + c - e
	+-	1 + a - b - c - e	1 - a + b - c + e
	-+	1 - a + b - c - e	1 + a - b - c + e
	--	1 - a - b + c + e	1 + a + b + c - e

Il vient pour l'intersection :

$$int(IT,J) = (1/2)\,(((a+b)^2/(1+c)) + ((a-b)^2/(1-c)) + (2e^2/(1-c^2)))$$

au second ordre en r , i.e. ici en a, b, c, e, il vient :

$$\mathrm{int}(IT,J) \approx a^2 + b^2 + e^2 + 0^3(r) ;$$

où on reconnaît : $a^2 = \mathrm{int}(I,J)$; $b^2 = \mathrm{int}(T,J)$; $e^2 = \mathrm{int}(I,J,T)$; le terme du 3-ème ordre ne dépend pas de e : on a :

$$\mathrm{int}(IT,J) \approx a^2 + b^2 + e^2 - 2abc + 0^4(r).$$

En se reportant au § 1.3, on voit que le terme n° 3, (-2 abc), n'est autre que $-\Sigma\{f_i f_j f_t\, 2\, r^{it}\, r^{ij}\, r^{tj}\}$; ici r^{it} n'est autre que aij, etc.; et $r^{it}\, r^{ij}\, r^{tj}$ est $abc\; i^2 j^2 t^2$, c'est-à-dire la constante abc puisque les i, j, t valent +1 ou -1. Les autres termes de int^3 sont nuls : par exemple $r^{it}\, r^{ijt}\, r^{ij} = c\,e\,a\; i^3 j^2 t^2 = c\,e\,a\; i$, dont la moyenne est nulle.

3.2 *Autre exemple* : Il est facile de voir que les annulations de termes observées dans la correspondance sur le cube ne se produisent pas pour toute correspondance ternaire : c'est-à-dire que dans le calcul de int^3, r^{ijt} peut apporter une contribution effective. Par exemple on pose :

$$J = T = \{+1,-1\} \;;\; f_j = f_t = 1/2 \;;\; \text{mais :}$$

$$I = \{i0 \;;\; i' \;;\; i"\} \;; f_{i0} = f_{i'} = f_{i"} = 1/3$$

sur I on a les fonctions de base :

$$\varphi_0 \equiv 1 \;;\; \varphi_1 = \sqrt{3/2}(0\,;\,1\,;\,-1) \;;\; \varphi_2 = \sqrt{1/2}(-2\,;\,1\,;\,1) \;;$$

on prend pour la loi ternaire :

$$r^{it} = \varepsilon\varphi_1^i.t \;;\; r^{ij} = \varepsilon\varphi_1^i.j \;;\; r^{jt} = 0 \;;\; r^{ijt} = \varepsilon\varphi_2^i.j.t$$

d'où pour $\mathrm{int}^3(IT,J)$ d'après la formule du § 1.3 (en remplaçant j^2 et t^2 par 1)

$$\mathrm{int}^3 = -\varepsilon^3\, \Sigma\{f_i f_j f_t\; \varphi_1^i.t(\varphi_2^i.jt)(\varphi_2^i jt + 2\varphi_1^i\, j)\}$$

$$= -\varepsilon^3\, \Sigma\{f_i f_j f_t\; \varphi_1^i\; \varphi_2^i\,(\varphi_2^i\, t + 2\varphi_1^i)\}$$

$$= -2\varepsilon^3\, \Sigma\{f_i f_j f_t (\varphi_1^i)^2 \varphi_2^i\} = -\varepsilon^3\sqrt{2} \;;$$

résultat non nul, comme annoncé.