REVUE DE STATISTIQUE APPLIQUÉE

MICHEL TENENHAUS

La régression qualitative

Revue de statistique appliquée, tome 27, nº 2 (1979), p. 5-21

<http://www.numdam.org/item?id=RSA_1979__27_2_5_0>

LA RÉGRESSION QUALITATIVE

Michel TENENHAUS

CESA 1, rue de la Libération – 78350 JOUY-en-JOSAS

RESUME

La régression multiple peut se généraliser au cas où la variable dépendante et les variables indépendantes sont de n'importe quel type (nominal, ordinal ou numérique).

La méthode utilisée consiste à coder les variables qualitatives, en respectant leurs structures nominales ou ordinales, de manière à maximiser la corrélation multiple entre la variable dépendante et les variables indépendantes. Il est bien connu que cette méthode permet de retrouver l'analyse de variance ainsi que l'analyse discriminante. Nous montrons qu'elle permet également de retrouver l'analyse de variance monotone (MONANOVA) proposée par J.B. KRUSKAL en 1965.

Nous étudions également dans cet article les programmes ADDALS et MORALS mis au point en 1976 par F.W. YOUNG, J. De LEEUW et Y. TAKANE.

L'objectif de ces programmes est de résoudre le problème général de la régression qualitative (la variable dépendante et les variables indépendantes sont de n'importe quel type) en utilisant la méthode décrite plus haute.

Enfin, nous avons soigneusement étudié les bases géométriques communes à ces méthodes et montrons que celles-ci sont finalement toutes équivalentes à une analyse canonique avec des contraintes sur le signe de certains coefficients.

INTRODUCTION

La régression multiple peut se généraliser au cas où la variable dépendante Y et les variables indépendantes $X_1, \ldots, X_k$ sont de n'importe quel type (nominal, ordinal ou numérique).

La méthode utilisée consiste à coder les variables qualitatives, en respectant leurs structures nominales ou ordinales, de manière à maximiser la corrélation multiple entre la variable dépendante et les variables indépendantes.

Il est bien connu que cette méthode permet de retrouver l'analyse de variance ainsi que l'analyse discriminante. Nous montrons qu'elle permet également de retrouver l'analyse de variance monotone (MONANOVA) proposée par J.B. KRUSKAL [6] en 1965.

Nous étudions également dans cet article les programmes ADDALS et MORALS mis au point en 1976 par F.W. YOUNG, J. De LEEUW et Y. TAKANE [3] et [11].

L'objectif de ces programmes est de résoudre le problème général de la régression qualitative (la variable dépendante et les variables indépendantes sont de n'importe quel type) en utilisant la méthode décrite plus haut. Le programme MORALS est général tandis que le programme ADDALS suppose un plan d'expérience orthogonal.

Mots-clés : **Régression multiple, régression monotone, analyse de variance, analyse de variance monotone, codage optimal.**

Par ailleurs nous étudions soigneusement les bases géométriques communes à ces méthodes et montrons que celles-ci sont finalement toutes équivalentes à une analyse canonique avec des contraintes sur le signe de certains coefficients.

Enfin nous présentons un exemple d'utilisation de MORALS.

Donnons le plan de cet article :

1) Codage d'une variable qualitative
2) Equivalence entre différents critères utilisés pour déterminer des codages optimaux
3) Un critère général utilisé en régression qualitative
4) Trois programmes de régression qualitative : MONANOVA, ADDALS, MORALS.
5) Un exemple d'utilisation de MORALS.

I. CODAGE D'UNE VARIABLE QUALITATIVE

I.1. Définition

Considérons une population E de taille n sur laquelle est définie une variable qualitative X prenant ses valeurs dans l'ensemble $\mathcal{X} = \{1, \ldots, p\}$ des modalités.

On appelle codage δ des modalités toute application $\delta : \mathcal{X} \to \mathbf{R}$ et codage X^* de la variable X l'application $X^* = \delta \circ X : E \to \mathbf{R}$.

Si la variable X est nominale ($\mathcal{X}$ ne possède pas de structure particulière) on impose en général à X^* d'être centrée ou centrée-réduite.

Si la variable X est ordinale ($\mathcal{X}$ est muni d'une structure d'ordre à priori que nous supposons identique à l'ordre naturel $1, 2, \ldots, p$) on impose, en plus des éventuelles contraintes ci-dessus, les contraintes $\delta(1) \leqslant \delta(2) \leqslant \ldots \leqslant \delta(p)$.

I.2. Interprétation géométrique dans $\mathbf{R}^n$

I.2.1. *X nominale*

Notons $X_1, \ldots, X_p$ les variables indicatrices associées à X.

Nous avons $X^* = \delta \circ X = \sum_{i=1}^{p} \delta(i) X_i$.

Ainsi si nous nous plaçons maintenant dans $\mathbf{R}^n$ (X^* et les X_i sont assimilées à des vecteurs de R^n) nous avons :

$X^* \in \mathcal{L}(X)$ sous espace vectoriel de R^n engendré par $X_1, \ldots, X_p$.

La contrainte X^* centrée s'écrit $X^* \in \mathcal{L}(X) \cap 1^{\perp}$ où $1^{\perp}$ est le sous espace de $\mathbf{R}^n$ orthogonal au vecteur 1 dont toutes les composantes sont égales à un.

La contraintes X^* centrée réduite s'écrit $X^* \in \mathcal{L}(X) \cap 1^{\perp} \cap S$ où S est la sphère centrée sur 0 et de rayon $\sqrt{n}$.

I.2.2. *X ordinale*

Remarquons que tout codage δ vérifiant les contraintes

$$\delta(1) \leqslant \delta(2) \leqslant \cdots \leqslant \delta(p)$$

peut s'écrire

$$\delta(1) = -\lambda_1 + \lambda_2$$

$$\delta(2) = -\lambda_1 + \lambda_2 + \lambda_3$$

$$\vdots$$

$$\delta(p) = -\lambda_1 + \lambda_2 + \lambda_3 + \cdots + \lambda_{p+1}$$

où les λ_i, $i = 1, \ldots, p + 1$, sont positifs ou nuls.

La variable codée X^* associée à la variable ordinale X peut alors s'écrire

$$X^* = \sum_{i=1}^{p} \delta(i) X_i = \sum_{i=1}^{p} \left(-\lambda_1 + \sum_{j=2}^{i+1} \lambda_j\right) X_i$$

$$= \sum_{k=1}^{p+1} \lambda_k \widetilde{X}_k$$

où les λ_k sont positifs ou nuls et où les $\widetilde{X}_k$ sont définies par $\widetilde{X}_1 = -1$ et $\widetilde{X}_k = \sum_{i=k-1}^{p} X_i$ pour $k = 2, \ldots, p + 1$.

Autrement dit $\widetilde{X}_1$ représente la variable codée X^* telle que toutes les modalités sont codées -1 et $\widetilde{X}_k$ représente la variable codée X^* telle que les modalités $1, \ldots, k - 2$ sont codées 0 et les modalités $k - 1, \ldots, p$ sont codées 1.

Ainsi si nous nous plaçons maintenant dans $\mathbf{R}^n$ (les $\widetilde{X}_k$ sont assimilées à des vecteurs de R^n) nous avons :

$X^* \in \mathcal{C}(X)$ cône polyédrique convexe de $\mathbf{R}^n$ engendré par $\widetilde{X}_1, \ldots, \widetilde{X}_{p+1}$.

La contrainte X^* centrée s'écrit $X^* \in \mathcal{C}(X) \cap 1^\perp$ et la contrainte X^* centrée réduite s'écrit $X^* \in \mathcal{C}(X) \cap 1^\perp \cap S$.

II. EQUIVALENCES

Les propositions suivantes sont utiles pour montrer que les méthodes proposées par J.B. KRUSKAL [6] puis par F.W. YOUNG, J. De LEEUW et Y. TAKANE [3] et [11] possèdent une base géométrique commune et utilisent un critère commun.

Rappelons tout d'abord quelques propriétés des cônes polyédriques convexes (J. STOER et C. WITZGALL [10], chapitre 2).

Soit $\mathcal{C}$ un cône polyédrique convexe de $\mathbf{R}^n$. Tout vecteur M de $\mathbf{R}^n$ possède une projection orthogonale sur $\mathcal{C}$, notée $P_{\mathcal{C}}(M)$, réalisant le $\underset{X \in \mathcal{C}}{\text{Minimum}} \|M - X\|^2$. Cette projection est unique.

On définit le cône polaire $\mathcal{C}^p$ de $\mathcal{C}$ par

$$\mathcal{C}^p = \{Y \in \mathbf{R}^n \mid Y' X \leqslant 0 \quad \text{pour tout} \quad X \in \mathcal{C}\}.$$

On a alors la propriété suivante pour tout $M \in \mathbf{R}^n$:

$$M = P_{\mathcal{C}}(M) + P_{\mathcal{C}^p}(M)$$

avec $P_{\mathcal{C}}(M)$ et $P_{\mathcal{C}^p}(M)$ orthogonaux.

Et cette décomposition orthogonale de M en un vecteur de $\mathcal{C}$ plus un vecteur de $\mathcal{C}^p$ est unique.

Cette dernière propriété implique donc :

1) $(M - P_{\mathcal{C}}(M)' P_{\mathcal{C}}(M) = 0$

2) $(M - P_{\mathcal{C}}(M))' X \leqslant 0$ pour tout $X \in \mathcal{C}$.

Proposition 1

Soit $\mathcal{C}$ un cône polyédrique convexe de $\mathbf{R}^n$.

Soit M un vecteur de $\mathbf{R}^n$ n'appartenant pas à $\mathcal{C}^p$.

Soit :

$$A_1 : \underset{A \in \mathcal{C}}{\text{Minimum}} \|M - A\| = \|M - A_1\|$$

$$A_2 : \underset{A \in \mathcal{C} \cap S}{\text{Minimum}} \|M - A\| = \|M - A_2\|$$

$$A_3 : \underset{A \in \mathcal{C}}{\text{Maximum}} \cos(A, M) = \cos(A_3, M)$$

$$A_4 : \underset{A \in \mathcal{C} \cap S}{\text{Maximum}} \cos(A, M) = \cos(A_4, M)$$

alors

$$\frac{\sqrt{n}}{\|A_1\|} A_1 = A_2 = A_3 = A_4$$

Démonstration

1) L'identité $\|M - A\|^2 = \|M\|^2 + \|A\|^2 - 2\|M\| \|A\| \cos(M, A)$

implique $A_2 = A_4$.

2) Montrons que $\dfrac{\sqrt{n}}{\|A_1\|} A_1 = A_2$.

Soit $A \in \mathcal{C} \cap S$.

On a, puisque $(M - A_1)' A_1 = 0$, $P_S(A_1) = \dfrac{\sqrt{n}}{\|A_1\|} A_1$ (où $P_S(A_1)$ est la projection orthogonale de A_1 sur S) et

$(M - A_1)'(A - A_1) \leqslant 0$, les relations suivantes :

$$\|M - \frac{\sqrt{n}}{\|A_1\|} A_1\|^2 = \|M - A_1\|^2 + \|A_1 - \frac{\sqrt{n}}{\|A_1\|} A_1\|^2 \leqslant \|M - A_1\|^2 + \|A_1 - A\|^2$$

$$\leqslant \|M - A_1\|^2 + \|A_1 - A\|^2 + 2(M - A_1)'(A_1 - A) = \|M - A\|^2$$

Et par conséquent $\quad A_2 = \dfrac{\sqrt{n}}{\|A_1\|} A_1$.

3) Le cosinus étant indépendant de la norme des vecteurs on peut choisir $A_3 = A_4$.

Proposition 2

Soient $\mathcal{C}_1$ et $\mathcal{C}_2$ deux cônes polyédriques convexes de $\mathbf{R}^n$.

Soit $\mathcal{M} = \mathcal{C}_2 \cap S$.

Supposons que $\mathcal{M}$ ne soit pas entièrement contenu dans $\mathcal{C}_1^p$.

Soit A_1 , M_1 : $\underset{\substack{A \in \mathcal{C}_1 \\ M \in \mathcal{M}}}{\text{Minimum}} \|M - A\| = \|M_1 - A_1\|$.

Alors $\underset{\substack{A \in \mathcal{C}_1 \\ M \in \mathcal{M}}}{\text{Maximum}} \cos (A , M) = \cos (A_1 , M_1)$,

Démonstration

Soit A_2 , M_2 : $\underset{\substack{A \in \mathcal{C}_1 \\ M \in \mathcal{M}}}{\text{Maximum}} \cos (A , M) = \cos (A_2 , M_2)$.

Nous avons $A_1 = P_{\mathcal{C}_1} (M_1)$. L'hypothèse $\mathcal{M} \not\subset \mathcal{C}_1^p$ implique $A_1 \neq 0$. La proposition 1 nous permet de choisir $A_2 = P_{\mathcal{C}_1} (M_2)$.

Par conséquent nous avons

$$\|M_1 - A_1\|^2 = n (1 - \cos^2 (M_1 , A_1))$$

et

$$\|M_2 - A_2\|^2 = n (1 - \cos^2 (M_2 , A_2))$$

D'où la proposition puisque $\|M_1 - A_1\|^2 \leqslant \|M_2 - A_2\|^2$ implique $\cos^2 (M_1 , A_1) \geqslant \cos^2 (M_2 , A_2)$.

Dans la proposition suivante et son corollaire nous reprenons, en l'approfondissant, la note théorique écrite par J.B. KRUSKAL dans le mode d'emploi du programme MONANOVA [7].

Proposition 3

Soit M un vecteur de $\mathbf{R}^n$.

Soit $\mathcal{A}$ un sous-espace vectoriel de $\mathbf{R}^n$

Soit $S (A , M) = \dfrac{\|M - A\|}{\|A\|}$ (Stress de Kruskal lorsque $\bar{A} = 0$).

Soit A_1 : $\underset{A \in \mathcal{A}}{\text{Minimum}}\ S (A , M) = S(A_1 , M)$.

Alors

1) $\underset{A \in \mathcal{A}}{\text{Maximum}} \cos (A , M) = \cos (A_1 , M)$

2) $S (A_1 , M) = \sqrt{1 - \cos^2 (A_1 , M)}$

3) $A_1 = \dfrac{1}{\cos^2 (P_{\mathcal{A}} (M) , M)} P_{\mathcal{A}}(M)$

Démonstration

Nous avons :

$$\cos^2(A, M) = 1 - \frac{\underset{\alpha}{\operatorname{Min}} \|\alpha M - A\|^2}{\|A\|^2} = 1 - \underset{\alpha}{\operatorname{Min}} \frac{\|M - \alpha A\|^2}{\|\alpha A\|^2}$$

$$= 1 - \underset{\alpha}{\operatorname{Min}} S(\alpha A, M)^2$$

Soit A_2 tel que $\underset{A \in \mathcal{A}}{\operatorname{Maximum}} \cos(A, M) = \cos(A_2, M)$

Montrons que $\cos(A_1, M) = \cos(A_2, M)$ en vérifiant que $\cos(A_1, M) \geqslant \cos(A_2, M)$:

$$\cos^2(A_1, M) = 1 - \underset{\alpha}{\operatorname{Min}} S(\alpha A_1, M)^2 = 1 - S(A_1, M)^2$$

$$\geqslant 1 - \underset{\alpha}{\operatorname{Min}} S(\alpha A_2, M)^2 = \cos^2(A_2, M).$$

D'où les relations (1) et (2).

La relation (1) implique $A_1 = \lambda_1 P_{\mathcal{A}}(M)$.

Montrons que $\lambda_1 = \dfrac{1}{\cos^2(P_{\mathcal{A}}(M), M)}$:

$$S(A_1, M) = \underset{\lambda}{\operatorname{Minimum}} S(\lambda P_{\mathcal{A}}(M), M)$$

$$= \underset{\lambda}{\operatorname{Minimum}} \frac{\|M - \lambda P_{\mathcal{A}}(M)\|}{\|\lambda P_{\mathcal{A}}(M)\|}$$

$$= \underset{\lambda}{\operatorname{Minimum}} \frac{\|\frac{1}{\lambda} M - P_{\mathcal{A}}(M)\|}{\|P_{\mathcal{A}}(M)\|}$$

Il faut donc rechercher le $\underset{\lambda}{\operatorname{Minimum}} \|\frac{1}{\lambda} M - P_{\mathcal{A}}(M)\|$

D'où $$\frac{1}{\lambda_1} = \frac{M' P_{\mathcal{A}}(M)}{\|M\|^2} = \frac{\cos(M, P_{\mathcal{A}}(M)) \|P_{\mathcal{A}}(M)\|}{\|M\|} = \cos^2(P_{\mathcal{A}}(M), M)$$

et la relation (3) est démontrée.

Corollaire

Soit $\mathcal{A}$ un sous-espace vectoriel de $\mathbf{R}^n$ et $\mathcal{C}$ un cône polyédrique convexe.

Soit A_1, M_1 : $\underset{\substack{A \in \mathcal{A} \\ M \in \mathcal{C}}}{\operatorname{Minimum}} S(A, M) = S(A_1, M_1)$

Nous avons $M_1 = P_{\mathcal{C}}(A_1)$ et $A_1 = \dfrac{1}{\cos^2(P_{\mathcal{A}}(M_1), M_1)} P_{\mathcal{A}}(M_1)$

Démonstration

1) $S(A_1, M_1) = \underset{M \in \mathcal{C}}{\text{Minimum}} \dfrac{\|M - A_1\|}{\|A_1\|}$. D'où $M_1 = P_{\mathcal{C}}(A_1)$.

2) $S(A_1, M_1) = \underset{A \in \mathcal{A}}{\text{Minimum}}\ S(A, M_1)$.

D'où
$$A_1 = \frac{1}{\cos^2 (P_{\mathcal{A}}(M_1), M_1)} P_{\mathcal{A}}(M_1).$$

Proposition 4

Soit $\mathcal{C}(Y)$ le cône polyédrique convexe associé à une variable ordinale Y.

Si M est un vecteur centré on a $P_{\mathcal{C}(Y)}(M) = P_{\mathcal{C}(Y) \cap 1^{\perp}}(M)$.

Démonstration

Il suffit de montrer que si M est centré, alors sa projection sur $\mathcal{C}(Y)$ l'est aussi.

Or les vecteurs 1 et -1 appartiennent à $\mathcal{C}(Y)$.

Par conséquent

$$(M - P_{\mathcal{C}(Y)}(M))' 1 = - P_{\mathcal{C}(Y)}(M)' 1 \leqslant 0$$

et
$$(M - P_{\mathcal{C}(Y)}(M))' (-1) = P_{\mathcal{C}(Y)}(M)' 1 \leqslant 0$$

D'où nous déduisons que $P_{\mathcal{C}(Y)}(M)$ est centré.

III – UN CRITERE GENERAL UTILISE EN REGRESSION QUALITATIVE

Les variables $Y, X_1, \ldots, X_k$ sont d'un type quelconque (nominal, ordinal ou numérique).

Notons $\mathcal{C}, \mathcal{C}_1, \ldots, \mathcal{C}_k$ les sous-espaces vectoriels ou les cônes polyédriques convexes de $\mathbf{R}^n$ associés à ces variables.

Un critère général consiste à rechercher les codages $Y^* \in \mathcal{C}, X_1^* \in \mathcal{C}_1, \ldots X_k^* \in \mathcal{C}_k$ maximisant la corrélation multiple $R(Y^* ; X_1^*, \ldots, X_k^*)$ entre Y^* et les X_i^*.

Pour éviter des solutions dégénérées on élimine les codages constants.

Comme $Y^* \in \mathcal{C}$ implique $\dfrac{Y^* - \bar{Y}^*}{S_Y} \in \mathcal{C}$, et de même pour les X_i^*, le critère général peut aussi s'écrire :

$$\underset{\substack{Y^* \in \mathcal{C} \cap 1^{\perp} \cap S \\ X_i^* \in \mathcal{C}_i \cap 1^{\perp} \cap S \\ i=1,\ldots,k}}{\text{Maximiser}} R(Y^* ; X_1^*, \ldots, X_k^*)$$

On recherche donc les codages centrés-réduits des variables $Y, X_1, \ldots, X_k$ maximisant la corrélation multiple.

Le critère général se ramène à une analyse canonique sous contrainte.

En effet, en utilisant les résultats de la partie I de cet article, nous constatons que tout codage d'une variable qualitative s'écrit comme combinaison linéaire de variables bien définies :

$$X^* = \sum_{i=1}^{m} \alpha_i \check{X}_i$$

puisque, si X est nominale, $X^* = \sum_{i=1}^{p} \delta(i) X_i$

et, si X est ordinale, $X^* = \sum_{i=1}^{p+1} \lambda_i \widetilde{X}_i$ avec $\lambda_1, \ldots \lambda_{p+1} \geqslant 0$.

Les coefficients α_i sont donc quelconques si X est nominale et positifs si X est ordinale.

Le critère général est donc équivalent à :

$$\underset{\alpha_j, \tau_i, \beta_{il}}{\text{Maximiser}} \ \text{cor} \left(\sum_{j=1}^{m_0} \alpha_j \check{Y}_j, \sum_{i=1}^{k} \tau_i \sum_{l=1}^{m_i} \beta_{il} \check{X}_{il} \right)$$

avec les contraintes de positivité adéquates.

Si on pose $\gamma_{il} = \tau_i \beta_{il}$ le critère devient

$$\underset{\alpha_j, \gamma_{il}}{\text{Maximiser}} \ \text{cor} \left(\sum_{j=1}^{m_0} \alpha_j \check{Y}_j, \sum_{i=1}^{k} \sum_{l=1}^{m_i} \gamma_{il} \check{X}_{il} \right)$$

avec les contraintes

– les α_j associés à une variable Y ordinale sont positifs

– les γ_{il} associés à une variable X_i ordinale sont de même signe.

Remarque :

En analyse des préférences on étudie souvent le problème d'expliquer une variable ordinale Y à l'aide de variable ordinales X_i et on impose aux τ_i d'être positifs.

Ce problème d'analyse multicritère se ramène donc au problème ci-dessus avec les contraintes plus simples car linéaires : les α_j et les γ_{il} doivent être positifs.

IV. TROIS PROGRAMMES DE REGRESSION QUALITATIVE

Nous allons étudier dans ce paragraphe les programmes MONANOVA, ADDALS et MORALS.

Nous montrerons que le critère général s'applique exactement à MONANOVA et à MORALS et, avec une contraintes supplémentaire, à ADDALS.

IV.1. MONANOVA*

Le programme MONANOVA mis au point par J.B. KRUSKAL [6] et J.B. KRUSKAL et F. CARMONE [7] s'applique au cas où Y est ordinale et les X_i nominales.

IV.1.1. Le critère utilisé dans MONANOVA est équivalent au critère général

Notons :

$$\mathcal{M} = \mathcal{C}(Y) \cap 1^{\perp} \cap S$$

l'ensemble des codages centrés réduits de Y,

$$\mathcal{A}_i = \mathcal{L}(X_i) \cap 1^{\perp}$$

l'ensemble des codages centrés de X_i,

$$\mathcal{A} = \overset{k}{\underset{i=1}{\oplus}} \mathcal{A}_i$$

somme directe des $\mathcal{A}_i$ (c'est l'ensemble des combinaisons linéaires de X_i^* appartenant à $\mathcal{A}_i$).

J.B. KRUSKAL définit le Stress $S(A,M) = \dfrac{||M - A||}{||A||}$

L'objectif de MONANOVA est de rechercher le

$$\underset{\substack{A \in \mathcal{A} \\ M \in \mathcal{M}}}{\text{Minimum}} \; S(A, M)$$

La proposition suivante montre que ce critère est équivalent au critère général.

Proposition 5

Soit M_1, A_1 tels que $\underset{\substack{M \in \mathcal{M} \\ A \in \mathcal{A}}}{\text{Minimum}} \; S(A,M) = S(A_1, M_1)$.

Alors $\underset{\substack{M \in \mathcal{M} \\ A \in \mathcal{A}}}{\text{Maximum}} \cos(A, M) = \cos(A_1, M_1)$

et de plus, si on pose $\quad A_1 = \sum_{i=1}^{k} A_{1i} \quad$ où $\quad A_{1i} \in \mathcal{A}_i$,

on a :

$$\underset{\substack{Y^* \in \mathcal{M} \\ X^* \in \mathcal{A}_i \\ i=1,\dots,k}}{\text{Maximum}} \; R(Y_1^* ; X_1^* \dots X_k^*) = R(M_1 ; A_{11}, \dots, A_{1k})$$

Démonstration

1) Montrons que

$$\underset{\substack{M \in \mathcal{M} \\ A \in \mathcal{A}}}{\text{Maximum}} \; \cos(A, M) = \cos(A_1, M_1)$$

(*) MONANOVA = Monotone Analysis of Variance

D'après la proposition 3 nous avons :

$$\underset{A \in \mathcal{A}}{\text{Minimum}} \; S(A, M) = \sqrt{1 - \underset{A \in \mathcal{A}}{\text{Maximum}} \; \cos^2 (A, M)}$$

D'où $\cos (A_1, M_1) = \underset{\substack{M \in \mathcal{M} \\ A \in \mathcal{A}}}{\text{Maximum}} \; \cos (A, M)$

et
$$S(A_1, M_1) = \sqrt{1 - \cos^2 (A_1, M_1)}$$

2) La corrélation multiple $R(Y^* ; X_1^*, \ldots, X_k^*)$ étant égale au

$$\cos (Y^* ; P_{\mathcal{L}(X_1^*, \ldots, X_k^*)} (Y^*))$$

où
$$Y^* \in \mathcal{M} \text{ et } P_{\mathcal{L}(X_1^*, \ldots, X_k^*)} (Y^*) \in \mathcal{A}$$

on a bien le résultat annoncé dans la proposition.

IV.1.2. L'algorithme

L'algorithme est itératif et permet de construire une suite de vecteurs A^t, M^t tels que la suite $S(A^t, M^t)$ soit décroissante.

La suite peut converger vers un optimum local et l'optimum atteint dépend sans doute de la solution initiale choisie.

Décrivons l'algorithme.

Initialisation

On choisit arbitrairement $M^\circ \in \mathcal{M}$, codage centré réduit de Y.

Etape t

A l'étape $t - 1$ on a obtenu les vecteurs

$$M^{t-1} \in \mathcal{M}, \; A_1^{t-1} \in \mathcal{A}_1, \ldots, A_k^{t-1} \in \mathcal{A}_k \quad \text{codages de} \quad Y, X_1, \ldots, X_k .$$

Et on construit

$$A^{t-1} = \sum_{i=1}^{k} A_i^{t-1}$$

Détermination des A_i^t

Il s'agit de rechercher le $\underset{A \in \mathcal{A}}{\text{Minimum}} \; S(A, M^{t-1})$.

J.B. KRUSKAL utilise une procédure du gradient pour obtenir A^t.

Il nous semble préférable d'utiliser la proposition 3 donnant :

$$A^t = \frac{1}{\cos^2 (M^{t-1}, P_{\mathcal{A}} (M^{t-1}))} P_{\mathcal{A}} (M^{t-1})$$

$$= \sum_{i=1}^{k} A_i^t .$$

Détermination de M^t

Il s'agit de rechercher le $\underset{M \in \mathcal{M}}{\text{Minimum}} \; S(A^t, M)$.

Comme $S(A^t, M) = \frac{||M - A^t||}{||A^t||}$ il suffit de rechercher le $\underset{M \in \mathcal{M}}{\text{Minimum}} ||M - A^t||$.

D'après la proposition 1 ce minimum s'obtient en deux étapes :

1) Rechercher le $\underset{M \in \mathcal{C}(Y) \cap 1^{\perp}}{\text{Minimum}} ||M - A^t||$.

2) Normaliser le résultat obtenu.

Par ailleurs nous avons vu (Proposition 4) que si un vecteur A est centré sa projection sur $\mathcal{C}(Y)$ l'est aussi.

Le minimum cherché peut donc s'obtenir en modifiant l'étape 1.

Soit :

1) Rechercher le $\underset{M \in \mathcal{C}(Y)}{\text{Minimum}} ||M - A^t||$.

2) Normaliser le résultat obtenu.

D'où :
$$M^t = \frac{\sqrt{n}}{||P_{\mathcal{C}(Y)}(A^t)||} P_{\mathcal{C}(Y)}(A^t).$$

L'algorithme permettant d'obtenir $P_{\mathcal{C}(Y)}(A^t)$ est décrit dans [1] et [5].

Convergence de l'algorithme

La suite $S(A^t, M^t)$ est décroissante :

$$S(A^t, M^t) \leqslant S(A^t, M^{t-1}) \leqslant S(A^{t-1}, M^{t-1}).$$

L'algorithme s'arrête lorsque le critère $S(A^t, M^t)$ se stabilise.

IV.2. ADDALS*

Le programme ADDALS, mis au point en 1976 par J. De LEEUW, F.W. YOUNG et Y. TAKANE [3] s'applique au cas où Y est de type quelconque et les X_i nominales ou ordinales. De plus le plan d'expérience associé aux X_i doit être orthogonal.

IV.2.1. Le Critère utilisé dans ADDALS est équivalent au critère général avec une contrainte supplémentaire

Notons $\mathcal{M} = \mathcal{C}(Y) \cap 1^{\perp} \cap S$ l'ensemble des codages centrés-réduits de la variable Y.

$\mathcal{C}(Y)$ est un sous-espace vectoriel ou un cône polyédrique convexe de $\mathbf{R}^n$ suivant la nature de Y.

Notons :

$$\mathcal{B}_i = \begin{cases} \mathcal{A}_i = \mathcal{L}(X_i) \cap 1^{\perp} & \text{si } X_i \text{ est nominale} \\ \mathcal{C}_i = \mathcal{C}(X_i) \cap 1^{\perp} & \text{si } X_i \text{ est ordinale} \end{cases}$$

On suppose que le plan d'expérience associé aux X_i est orthogonal :

$$\mathcal{A}_i \perp \mathcal{A}_j \quad \text{pour} \quad i \neq j.$$

(*) ADDALS = Additivity analysis by alternating least Squares.

On pose $\mathcal{A} = \overset{k}{\underset{i=1}{\oplus}} \mathcal{A}_i$ et $\mathcal{B} = \overset{k}{\underset{i=1}{\oplus}} \mathcal{B}_i$

Le critère utilisé dans ADDALS est :

$$\lambda(M, B) = ||M - B||^2$$

et l'objectif est de rechercher le

$$\underset{\substack{M \in \mathcal{M} \\ B \in \mathcal{B}}}{\text{Minimum}} \ \lambda(M, B)$$

D'après la proposition 2 il est équivalent de rechercher le

$$\underset{\substack{M \in \mathcal{M} \\ B \in \mathcal{B}}}{\text{Maximum}} \ \cos(M, B),$$

soit de rechercher le

$$\underset{\substack{Y^* \in \mathcal{M} \\ X_1^* \in \mathcal{B}_1, \ldots, X_k^* \in \mathcal{B}_k}}{\text{Maximum}} \quad \text{cor}\,(Y^*, \sum_{i=1}^{k} X_i^*).$$

Ce critère est bien équivalent au critère général avec la contrainte supplémentaire : les coefficients de régression des variables ordinales sont positifs.

IV.2.2. L'algorithme

L'algorithme est itératif et permet de construire une suite de vecteurs M^t, B^t tels que la suite $\lambda(M^t, B^t)$ soit décroissante.

Comme pour MONANOVA cette suite peut probablement converger vers un optimum local et l'optimum atteint dépend sans doute de la solution initiale choisie.

Décrivons l'algorithme.

Initialisation

On choisit arbitrairement $M^\circ \in \mathcal{M}$, codage centré réduit de Y.

Etape t

A l'étape t $-$ 1 on a obtenu les vecteurs $M^{t-1} \in \mathcal{M}$, $B_1^{t-1} \in \mathcal{B}_1, \ldots, B_k^{t-1} \in \mathcal{B}_k$ codages de $Y, X_1, \ldots, X_k$. Et on construit $B^{t-1} = \sum_{i=1}^{k} B_i^{t-1}$.

Détermination des B_i^t

La procédure est basée sur l'identité suivante. Pour tout $A = \sum_{i=1}^{k} A_i$, avec $A_i \in \mathcal{A}_i$, nous avons :

$$||M - A||^2 = ||M - P_{\mathcal{A}}(M)||^2 + ||P_{\mathcal{A}}(M) - A||^2$$

$$= ||M - P_{\mathcal{A}}(M)||^2 + \sum_{i=1}^{k} ||P_{\mathcal{A}_i}(M) - A_i||^2$$

puisque les $\mathcal{A}_i$ sont orthogonaux.

D'où nous déduisons que rechercher le

$$\underset{B \in \mathcal{B}}{\text{Minimum}} \ ||M^{t-1} - B||^2$$

revient à rechercher, pour $i = 1, \ldots, k$, le

$$\underset{B_i \in \mathcal{B}_i}{\text{Minimum}} \ ||P_{\mathcal{A}_i}(M^{t-1}) - B_i||^2$$

Par conséquent

$$B_i^t = \begin{cases} P_{\mathcal{A}_i}(M^{t-1}) & \text{si } X_i \text{ est nominale} \\ P_{\mathcal{C}_i}(M^{t-1}) & \text{si } X_i \text{ est ordinale} \end{cases}$$

puisque $$P_{\mathcal{B}_i} \circ P_{\mathcal{A}_i}(M^{t-1}) = P_{\mathcal{B}_i}(M^{t-1}).$$

Et $$B^t = \sum_{i=1}^{k} B_i^t.$$

Détermination de M^t

Il s'agit de rechercher le $\underset{M \in \mathcal{M}}{\text{Minimum}} \ ||M - B^t||^2$.

D'où,d'après la proposition 1 et les propriétés de la projection,

$$M^t = P_{\mathcal{M}}(B^t) = \frac{\sqrt{n}}{||P_{\mathcal{C}(Y)}(B^t)||} P_{\mathcal{C}(Y)}(B^t).$$

Convergence de l'algorithme

La suite $\lambda(M^t, B^t)$ est décroissante :

$$||M^t - B^t||^2 \leqslant ||M^{t-1} - B^t||^2 \leqslant ||M^{t-1} - B^{t-1}||^2.$$

L'algorithme s'arrête lorsque le critère $\lambda(M^t, B^t)$ se stabilise.

IV – 3. MORALS*

Le programme MORALS mis au point en 1976 par F.W. Young, J. De Leeuw et Y. Takane [11] s'applique au cas où les variables $Y, X_1, \ldots, X_k$ sont de type quelconque.

IV – 3.1. Le critère utilisé dans MORALS est équivalent au critère général

Notons $\mathcal{M}$, $\mathcal{A}_1, \ldots, \mathcal{A}_k$ les espaces associés aux variables $Y, X_1, \ldots, X_k$ représentant les codages centrés-réduits de ces différentes variables.

Notons $\mathcal{A} = \sum_{i=1}^{k} \mathcal{L}(\mathcal{A}_i)$ le sous-espace vectoriel engendré par les $\mathcal{A}_i$.

Le critère utilisé dans MORALS est $\lambda(M, A) = ||M - A||^2$ et l'objectif de

(*) MORALS = Multiple optimal regression by alternating least squares.

l'algorithme est de recherche le

$$\underset{\substack{M \in \mathcal{M} \\ A \in \mathcal{A}}}{\text{Minimum}} \; ||M - A||^2 .$$

D'après la proposition 2 il est équivalent de rechercher le

$$\underset{\substack{M \in \mathcal{M} \\ A \in \mathcal{A}}}{\text{Maximum}} \cos (M , A)$$

Il s'agit donc de rechercher le

$$\underset{\substack{Y^* \in \mathcal{M} \\ X_i^* \in \mathcal{A}_i \\ i=1,\ldots,k}}{\text{Maximum}} \; R\,(Y^* \; ; \; X_1^*, \ldots, X_k^*).$$

IV – 3.2. L'algorithme de MORALS

Comme dans MONANOVA et ADDALS l'algorithme est itératif et permet de construire une suite de vecteurs M^t, A^t tels que la suite $\lambda(M^t , A^t)$ soit décroissante.

La convergence peut être locale et dépendre de la solution initiale choisie.

Décrivons l'algorithme.

1) *Initialisation*

On choisit M^0 arbitraire appartenant à $\mathcal{M}$, codage centré-réduit de Y.

2) *Etape t*

A l'étape $t - 1$ on a défini $M^{t-1}, A_1^{t-1}, \ldots, A_k^{t-1}$ appartenant respectivement à $\mathcal{M}, \mathcal{A}_1, \ldots, \mathcal{A}_k$.

On définit également $A^{t-1} = P_{\mathcal{L}(A_1^{t-1}, \ldots, A_k^{t-1})} (M^{t-1}) = \sum_{i=1}^{k} \alpha_i^{t-1} A_i^{t-1}$.

a) *Détermination de* M^t

Il s'agit de rechercher le $\underset{M \in \mathcal{M}}{\text{Minimum}} \; ||M - A^{t-1}||^2$.

Donc $M^t = P_{\mathcal{M}}(A^{t-1})$.

b) *Détermination des* A_i^t

On détermine progressivement les A_i^t.

1) *Détermination de* A_1^t

On recherche le $\underset{A_1 \in \mathcal{A}_1}{\text{Minimum}} \; ||M^t - \alpha_1^{t-1} A_1 - \sum_{i=2}^{k} \alpha_i^{t-1} A_i^{t-1}||^2$.

Donc $A_1^t = P_{\mathcal{A}_1} \left(\frac{1}{\alpha_1^{t-1}} \left(M^t - \sum_{i=2}^{k} \alpha_i^{t-1} A_i^{t-1}\right)\right)$.

2) *Détermination de* A_2^t.

On recherche de la même manière le

$$\underset{A_2 \in \mathcal{A}_2}{\text{Minimum}} \; ||M^t - \alpha_1^{t-1} A_1^t - \alpha_2^{t-1} A_2 - \sum_{i=3}^{k} \alpha_i^{t-1} A_i^{t-1}||^2$$

D'où
$$A_2^t = P_{\mathcal{A}_2} \left(\frac{1}{\alpha_2^{t-1}} \left(M^t - \alpha_1^{t-1} A_1^t - \sum_{i=3}^{k} \alpha_i^{t-1} A_i^{t-1} \right) \right)$$

3) On procède de la même manière pour déterminer $A_3^t, \dots, A_k^t$.

c) *Détermination de* A^t

On choisit $A^t = P_{\mathcal{L}(A_1^t, \dots, A_k^t)} (M^t) = \sum_{i=1}^{k} \alpha_i^t A_i^t$.

3) *Convergence de l'algorithme*

La suite $\lambda (M^t, A^t) = ||M^t - A^t||^2$ est décroissante :

$$||M^t - A^t||^2 \leqslant ||M^t - \sum_{i=1}^{k} \alpha_i^{t-1} A_i^t||^2 \leqslant ||M^t - \sum_{i=1}^{k-1} \alpha_i^{t-1} A_i^t - \alpha_k^{t-1} A_k^{t-1}||^2$$

$$\leqslant \dots \leqslant ||M^t - A^{t-1}||^2 \leqslant ||M^{t-1} - A^{t-1}||^2.$$

Remarque

Nous n'avons pas décrit toutes les possibilités des programmes ADDALS et MORALS et nous renvoyons le lecteur intéressé aux articles originaux et à la documentation des programmes.

V – UN EXEMPLE D'UTILISATION DE MORALS

Nous avons utilisé le programme MORALS sur les données de G. de Pesloüan [9] concernant les femmes ingénieurs diplômées en France.

Cet exemple a déjà été traité par là segmentation [2]. Il s'agissait d'analyser la liaison entre une note d'ambition et des facteurs socioéconomiques.

MORALS permet de résoudre le même problème que la segmentation, mais de manière plus rapide et plus synthétique.

Précisons les variables utilisées dans cet exemple donné à titre illustraitf.

Y : *Ambition (ordinale)*

1 beaucoup d'ambition

.

.

.

5 très peu d'ambition

X_1 : *Ecole (nominale)*

1 école polytechnique féminine

2 école d'ingénieurs mixte

X_2 : *Enfants (numérique)*

0 enfant
1 enfant
2 enfants
3 enfants
4 enfants
5 enfants et plus

X_3 : *Emploi (nominale)*

1 recherche
2 bureau d'études
3 laboratoire
4 informatique
5 autres

Le programme MORALS nous a permis d'obtenir les codages suivants des variables qualitatives :

Y^* : *Ambition*	X_1^* : *Ecole*	X_3^* : *Emploi*
$1 \to -0{,}62$	$1 \to -1$	$1 \to -0{,}48$
$2 \to -0{,}62$	$2 \to +1$	$2 \to 1{,}86$
$3 \to -0{,}62$		$3 \to -0{,}13$
$4 \to 1{,}6$		$4 \to -1{,}52$
$5 \to 1{,}6$		$5 \to 0{,}69$

La régression multiple entre Y^* et X_1^*, X_2, X_3^* donne

$$\hat{Y}^* = -0{,}08 + 0{,}19\,X_1^* + 0{,}08\,X_2 + 0{,}16\,X_3^*$$

avec
$$R\,(Y^* \ ;\ X_1^*, X_2, X_3^*) = 0{,}29.$$

Ainsi une femme sortie de l'E.P.F., ayant 2 enfants et travaillant en informatique a une note d'ambition estimée à

$$-0{,}08 + 0{,}19 \times (-1) + 0{,}08 \times 2 + 0{,}16 \times (-1{,}52)$$

$$= -0{,}35$$

Nous la classerions dans la catégorie des femmes ambitieuses.

CONCLUSION

La régression qualitative nous permet donc de généraliser la régression multiple au cas où les variables $Y, X_1, \ldots, X_k$ sont de n'importe quel type.

Lorsque toutes les variables sont quantitatives ou qualitatives nominales le critère général proposé dans la troisième partie de cet article est en effet équivalent