STATISTIQUE ET ANALYSE DES DONNÉES

PHILIPPE BESSE
CLAUDE VIDAL

Analyse des correspondances et codage par une probabilité de transition

Statistique et analyse des données, tome 7, nº 3 (1982), p. 1-25.
<http://www.numdam.org/item?id=SAD_1982__7_3_1_0>

Statistiques et Analyse de données
1982 - Vol. 7 - n° 3 - pp. 1-25

ANALYSE DES CORRESPONDANCES ET CODAGE PAR UNE PROBABILITE DE TRANSITION

Philippe BESSE[(1)] et Claude VIDAL[(2)]

Laboratoire de Statistique et Probabilités
E.R.A. - C.N.R.S. n° 591
Université Paul Sabatier - 31062 TOULOUSE Cédex

(1) I.U.T. de Statistique, Université Toulouse-le-Mirail
(2) Ecole Nationale d'Ingénieurs des Techniques et Industries Agro-Alimentaires, Nantes.

Résumé : *Dans cet article, la notion de T-codage ou codage par une transition est utilisée dans le but d'adapter les modèles classiques de l'Analyse Factorielle des Correspondances (A.F.C.) et de l'A.F.C. généralisée à certaines situations concrètes (A.F.C. d'un couple de variables aléatoires (Y_1,Y_2) connaissant un couple de v.a. (X_1,X_2) et la loi conjointe, prise en compte des incertitudes et des erreurs de mesures présentes dans un corpus de donnée ...). Ceci conduit alors à définir le T-codage sur des espaces produits, à étudier les problèmes d'indépendance entre les codages des différentes variables, leurs conséquences sur les analyses et enfin à aborder les questions d'estimation et de convergence des codages, puis des analyses.*

Abstract : *In this paper, the notion of scaling through a transition probability will be used in order to adapt the classical patterns of Correspondence Factorial Analysis (C.F.A.) and of Multiple Correspondence Analysis to some practical experiences (C.F.A. of a random variables pair (Y_1,Y_2), given the R.V. pair (X_1,X_2) and the joint distribution, after taking into account the dubiousness as well as the errors of measurement inherent in any data set ...). This can enable thus to define on spaces products the scaling through a transition probability study the problems of independence between the scaling of different R.V., their consequences on analyses, and finally raise up the questions of the estimation and convergence of a transition probability and of analyses.*

Mots clefs : *Analyse Factorielle des Correspondances, Codage, Probabilité de Transition.*

1 - INTRODUCTION

Il est fréquent, en pratique, d'avoir à étudier un phénomène décrit par une variable aléatoire (v.a.) Y inobservable (ou bien trop coûteuse à observer). Pour pallier à cet inconvénient l'étude de ce phénomène se fait au travers d'une v.a. X qui est -elle- observable et dont on peut estimer la liaison avec Y (en laboratoire ou sur un échantillon-test). Dans le cas où Y est une variable qualitative multidimensionnelle, une étude descriptive peut conduire à l'utilisation de l'Analyse Factorielle des Correspondances (A.F.C.) si Y est de la forme (Y_1, Y_2) où l'A.F.C. généralisée lorsque Y est un m-uple $(Y_1, Y_2, \ldots, Y_m)$. Pour déterminer cette analyse, il suffit de connaître (ou tout du moins d'estimer) la loi de Y. Celle-ci est déterminée (peut être estimée) dès que l'on connaît (sait estimer) la loi de X et la loi conditionnelle de Y à X.

Cette approche peut être utilisée en particulier dans le problème concret (à l'origine de cet article) exposé par P. BESSE [1] et que l'on peut résumer ainsi : des témoins observent un ensemble de phénomènes représentés par une variable qualitative multidimensionnelle Y dont ils font une estimation X (narration des témoins). Le problème est alors le suivant : comment rendre compte de Y à partir de l'étude faite sur X ?

En utilisant les outils introduits par J.F. MARTIN [8] sous l'appellation de codage "flou" (ou codage par une probabilité de transition), ce travail s'inscrit dans un cadre plus large visant à intégrer aux analyses factorielles les erreurs de mesure, de classement, les problèmes d'arrondi ou de lissage ... qui peuvent apparaître lors de l'observation des variables.

Ainsi, après avoir été sommairement rappelée, la notion de codage est étendue à des espaces produits de façon à pouvoir être appliquée à des couples (ou des m-uples) de variables. Ceci conduit alors à des problèmes d'indépendance (indépendance des erreurs entre elles, indépendance entre l'erreur sur une variable et les autres variables ...) qui sont discutés. Enfin, les problèmes d'estimation et donc de convergence sont résolus.

<u>Remarque</u> : Ce travail se rapportant à l'A.F.C. classique, toutes les tribus considérées par la suite sont de cardinal fini.

2 - T-CODAGE OU CODAGE PAR UNE TRANSITION

En pratique, les analyses factorielles non-linéaires utilisent essentiellement des fonctions indicatrices de modalités (pour des variables qualitatives) ou d'intervalles (pour des variables quantitatives), pour modéliser les situations à analyser ; ceci revenant à employer le codage disjonctif complet. La simplicité d'emploi et de mise en oeuvre de cet outil (qui permet d'exhiber simplement une base orthonormée) est aussi cause de sa rigidité, de son inadaptabilité à certains problèmes spécifiques. C'est pourquoi certains auteurs ont introduit d'autres formes de codage faisant appel à des fonctions spline (D. LAFAYE de MICHEAUX [6] , J.O. RAMSAY - S. WINSBERG [11] et [12] , J. VAN RIJCKEVORSEL [13]) ou bien à la notion de codage flou (J.P. BORDET [2] , J.F. MARTIN [8]) - notions reprises très récemment par J.L. MALLET [7] et J.M. GAUTIER , G. SAPORTA [5] -. On peut encore citer D.M. TITTERINGTON [14] qui a développé des outils similaires - estimation de densités de probabilités discrètes par la méthode des noyaux - pour l'étude de données catégorielles.

Parmi ces divers modèles, l'approche probabiliste due à J.F. MARTIN [8] (codage par une probabilité de transition) a été choisie, car elle semble la plus adaptée, et ce pour différentes raisons :

* elle est synthétique et englobe les autres approches.
* le cadre précis dans lequel elle est développée permet de donner une signification au codage, et donc d'interpréter les résultats.
* enfin et surtout, ce n'est que dans ce cadre qu'il est possible de parler d'indépendance de codages (des différentes variables à analyser cf. 4.1.).

2.1 - rappels et notations

On rappelle ici sommairement les notions introduites par J.F. MARTIN [8] , auquel on renvoie pour ce qui est des démonstrations.

Le codage par une (probabilité de) transition permet de tenir compte de certains problèmes parasites qui peuvent survenir lors de la saisie des données : erreurs de mesure, erreurs de classement, "bruit" qu'il faut lisser, erreurs systématiques d'arrondi ou encore, comme dans le cas qui nous intéresse, l'analyse d'une variable Y à l'aide d'une autre variable X.

Le principe adopté est alors d'associer à une observation $x = X(\omega)$ une probabilité (i.e. un codage)

P_x sur un espace $(F,\mathcal{F})$ servant à apprécier l'imprécision de la mesure. S'il n'y a pas d'imprécision (ou si l'on ne veut pas en tenir compte), il suffit d'employer le codage disjonctif classique.

Définitions

Soient $(E,\mathcal{E})$ et $(F,\mathcal{F})$ deux espaces probabilisables et P une transition sur $(E,\mathcal{E}) \times \mathcal{F}$, c'est-à-dire une application.

$$(E,\mathcal{E}) \times \mathcal{F} \longrightarrow ([0,1],\mathcal{B}_{[0,1]})$$
$$(x,B) \longrightarrow P_x(B)$$

telle que :

(i) $\forall x \in E$ $\quad B \longrightarrow P_x(B)$ est une probabilité sur $(F,\mathcal{F})$

(ii) $\forall B \in \mathcal{F}$ $\quad x \longrightarrow P_x(B)$ est mesurable de $(E,\mathcal{E})$ dans $([0,1],\mathcal{B}_{[0,1]})$.

Si X est une v.a. définie sur $(\Omega,\mathcal{A})$ à valeurs dans $(E,\mathcal{E})$, alors l'application P_X :

$$(\Omega,\mathcal{A}) \times \mathcal{B} \longrightarrow ([0,1],\mathcal{B}_{[0,1]})$$
$$(\omega,B) \longrightarrow P_{X(\omega)}(B)$$

définit une transition sur $(\Omega,\mathcal{A}) \times \mathcal{F}$.

Définition 1 :

On appelle T-codage sur $(E,\mathcal{E})$ relatif à $\mathcal{F}$ (resp. T-codage de X relatif à $\mathcal{F}$) une transition P sur $(E,\mathcal{E}) \times \mathcal{F}$ (resp. une transition P_X sur $(\Omega,\mathcal{A}) \times \mathcal{F}$).

Définition 2 :

Un T-codage sur $(E,\mathcal{E})$ relatif à $\mathcal{F}$ définit pour tout B de $\mathcal{F}$ une application mesurable ϕ_B définie par :

$$\phi_B : (E,\mathcal{E}) \longrightarrow ([0,1],\mathcal{B}_{[0,1]}) \;.$$
$$x \longrightarrow \phi_B(x) = P_x(B)$$

est appelée fonction de codage (associée à B).

Exemples

* Au codage disjonctif classique est associée la transition

$P = \delta_{\cdot}$: $(x,B) \longrightarrow P_x(B) = \delta_x(B)$ (où δ_x désigne la distribution de Dirac au point x) et, dans ce cas, les fonctions de codage sont les indicatrices :

$\phi_B(x) = \mathbb{I}_B(x)$.

Le codage conditionnel : Soit (X,Y) un couple de v.a. définies sur $(\Omega,\mathcal{A},\mu)$ à valeurs dans $(E \times F, \mathcal{E} \otimes \mathcal{F})$; pour tout B de $\mathcal{F}$, la fonction :

$$\begin{aligned} \phi_B : (E,\mathcal{E}) &\longrightarrow ([0,1],\mathcal{B}_{[0,1]}) \\ x &\longrightarrow \phi_B(x) = E_\mu^{X=x}(\mathbb{1}_B \circ Y) \end{aligned}$$

(où $E_\mu^{X=x}(\mathbb{1}_B \circ Y)$ est encore la probabilité conditionnelle $\mu[Y \in B/_{X=x}]$) est une fonction mesurable.

Si, de plus, il existe une version régulière de la probabilité conditionnelle de Y à X, l'application :

$$\begin{aligned} (E,\mathcal{E}) \times \mathcal{F} &\longrightarrow ([0,1],\mathcal{B}_{[0,1]}) \\ (x,B) &\longrightarrow P_x(B) = \phi_B(x) = E_\mu^{X=x}(\mathbb{1}_B \circ Y) \end{aligned}$$

est une transition sur $(E,\mathcal{E}) \times \mathcal{F}$ et on appelle T-codage conditionnel de Y en X la transition sur $(\Omega,\mathcal{A}) \times \mathcal{F}$:

$$P_X : (\omega,B) \longrightarrow P_{X(\omega)}(B) = [E_\mu^X (\mathbb{1}_B \circ Y)] (\omega).$$

Remarque : La fonction de codage ϕ_B n'est autre que la fonction de régression de $\mathbb{1}_B \circ Y$ en X.

T-codage d'une probabilité

Lorsque $(E,\mathcal{E})$ est muni d'une probabilité ν, le T-codage définit une probabilité sur $(F,\mathcal{F})$ par :

$$\begin{aligned} \mathcal{F} &\longrightarrow [0,1] \\ B &\longrightarrow E_\nu[P.(B)] = \int_E P_x(B)\, d\nu(x) = E_\nu(\phi_B). \end{aligned}$$

Elle est appelée probabilité codée de ν par P et notée $E_\nu(P)$ (ceci correspond à la notation νP de J. NEVEU [10], image à gauche de la probabilité ν par P considéré comme un opérateur sous-markovien).

Avec les notations des paragraphes précédents, si μ_X désigne la loi de X, la probabilité codée de μ_X par P est égale à la probabilité codée de μ par P_X :

$$E_{\mu_X}(P) = E_\mu(P_X) .$$

Exemples :

* Si P est la transition associée au codage disjonctif, on a :

$$E_{\mu_X}(P) = E_{\mu_X}(\delta) = \mu_X .$$

Si P est la transition associée au codage conditionnel, on trouve :

$$* \ \forall B \in \mathcal{F} \quad [E_{\mu_X}(P)](B) = E_{\mu_X}[P_{\cdot}(B)]$$
$$= \int_E P_X(B)\, d\mu_X(x)$$
$$= \int_\Omega P_X(\omega)(B)\, d\mu(\omega) = \mu_Y(B).$$

2.2 - T-codage sur un espace produit

Position du problème

La variable X à laquelle est appliqué le T-codage peut être un couple (X_1,X_2) ou même un m-uple $(X_1,\dots,X_m)$ prenant ses valeurs sur un espace produit $(E_1 \times \dots \times E_m, \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_m)$. Comment peut-on construire le codage de X à partir de celui des X_i $(i=1,\dots,m)$?

Soit donc $(E,\mathcal{E},\nu)$ un espace probabilisé et $(F,\mathcal{F})$ un espace probabilisable tels que :

$$E = E_1 \times E_2 \ , \ F = F_1 \times F_2 \quad \text{et} \quad \mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2 \ , \ \mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2 \ .$$

On note P une transition sur $(E,\mathcal{E}) \times \mathcal{F}$.

T-codage marge

Pour tout x de E, P_x est une probabilité sur $(F,\mathcal{F})$ et on peut considérer ses marges sur chacun des espaces $(F_i,\mathcal{F}_i)_{i=1,2}$; soit ${}^iP_x$ ces probabilités. On a :

$$\forall x \in E \ , \ \forall B_1 \in \mathcal{F}_1 \ , \ {}^1P_x(B_1) = \sum_{\ell=1}^{s} P_x(B_1 \times B_2^\ell)$$
$$\forall B_2 \in \mathcal{F}_2 \ , \ {}^2P_x(B_2) = \sum_{k=1}^{r} P_x(B_1^k \times B_2)$$

où $\{B_1^k\}_{k=1,\dots,r}$ (resp. $\{B_2^\ell\}_{\ell=1,\dots,s}$) désigne une partition de F_1 (resp. de F_2) engendrant $\mathcal{F}_1$ (resp. $\mathcal{F}_2$).

$P_{\cdot}(B)$ étant une application mesurable pour tout B de $\mathcal{F}$, les applications ${}^iP_{\cdot}(B_i)$ $(i=1,2)$ le sont aussi, comme sommes finies d'applications mesurables. Ainsi , ${}^iP$ est une transition sur $(E,\mathcal{E}) \times \mathcal{F}_i$.

Définition 3 ;

On appelle T-codages marge de P, les transitions ${}^iP$ sur $(E,\mathcal{E}) \times \mathcal{F}_i$, telles que pour tout x de E, les lois ${}^iP_x$ sur $(F_i,\mathcal{F}_i)$ soient les marges de P_x.

Proposition 1

La loi ν codée par un T-codage marge ${}^{i}P$ de P, est la marge de la probabilité codée de ν par P, soit :

$${}^{i}[E_{\nu}(P)] = E_{\nu}({}^{i}P).$$

La démonstration est évidente par interversion des signes de sommation.

Codage produit

Produit simple

Si P^{i} désigne un T-codage sur $(E_i, \mathcal{E}_i)$ relatif à $\mathfrak{F}_i$ et si $\mathcal{S}$ est le semi-anneau des "rectangles" :

$$\mathcal{S} = \{B_1 \times B_2 \;/\; B_1 \in \mathfrak{F}_1 \text{ et } B_2 \in \mathfrak{F}_2\}$$, on définit l'application

$P = P^{1} \otimes P^{2}$ sur $(E, \mathcal{E}) \times \mathcal{S}$ par :

$$P : (E, \mathcal{E}) \times \mathcal{S} \longrightarrow ([0,1], \mathfrak{B}_{[0,1]})$$
$$(x = (x_1, x_2), B_1 \times B_2) \longrightarrow P_x(B_1 \times B_2) = P^{1}_{x_1}(B_1) \times P^{2}_{x_2}(B_2).$$

Proposition 2

P est un T-codage sur $(E, \mathcal{E})$ relatif à $\mathfrak{F} = \mathfrak{F}_1 \otimes \mathfrak{F}_2$.

Démonstration :

$\mathfrak{F}_1$ et $\mathfrak{F}_2$ étant finies, on peut décomposer tout élément de $\mathfrak{F}_1 \otimes \mathfrak{F}_2$ en une réunion finie d'ensembles minimaux deux à deux disjoints et ainsi définir P sur $E \times \mathcal{S}$.

Soit :

$$P_x(B) = P_x\Big(\bigcup_{i=1}^{n} B_1^{i} \times B_2^{i}\Big) = \sum_{i=1}^{n} P_x(B_1^{i} \times B_2^{i})$$

$$= \sum_{i=1}^{n} P^{1}_{x_1}(B_1^{i}) . P^{2}_{x_2}(B_2^{i}).$$

D'autre part, pour tout B de $\mathfrak{F}$, l'application : $x \to P_x(B)$ est mesurable comme somme finie d'applications mesurables. □

Produit des marges

Si ${}^{i}P$ est un T-codage sur E relatif à $\mathfrak{F}_i$, on définit :

$$\forall x \in E,\ \forall B_1 \times B_2 \in \mathcal{S} ,\ P_x(B_1 \times B_2) = {}^{1}P_x(B_1) \ .\ {}^{2}P_x(B_2)$$
$$= ({}^{1}P \otimes {}^{2}P)_x\ (B_1 \times B_2).$$

${}^1P \otimes {}^2P$ est de même un T-codage appelé codage produit de ses marges.

Remarque

Si Card $E_1 = p$, Card $E_2 = q$ et si les tribus $\mathcal{F}_1$ et $\mathcal{F}_2$ sont engendrées par des partitions de cardinaux respectifs r et s, on voit alors que la définition d'un T-codage sur E relatif à $\mathcal{F}$ nécessite la connaissance (ou du moins dans la pratique l'estimation) de p q r s termes.

On aura donc tout intérêt à se ramener, si cela est possible, à des codages produits plus simples, puisque :

* pour le produit simple, il suffit de connaître pr + qs termes.
* pour le produit des marges, il suffit de connaître pq(r+s) termes.

3 - ANALYSES ET T-CODAGES

3.1 - Rappels et notations

L'A.F.C., peut être introduite de différentes manières équivalentes entre elles. Nous en utiliserons deux au cours de cet exposé :

Analyse canonique (A.C.) d'une probabilité sur un espace produit

Soit (X_1,X_2) un couple de v.a. défini sur $(\Omega, \mathcal{A}, \mu)$ et à valeurs dans $(E_1 \times E_2), \mathcal{E}_1 \otimes \mathcal{E}_2)$. On désigne par $\mathcal{A}_k$ (k=1,2) les tribus complétées de $X_k^{-1}(\mathcal{E}_k)$ et, respectivement, par $\{\alpha_1^i\}_{i=1,\dots,p}$ et $\{\alpha_2^j\}_{j=1,\dots,q}$ les partitions finies de Ω qui les engendrent. On pose :

$$p_{ij} = \mu(\alpha_1^i \cap \alpha_2^j) \ , \quad p_{i.} = \sum_{j=1}^{q} p_{ij} = \mu(\alpha_1^i)$$

$$p_{.j} = \sum_{i=1}^{p} p_{ij} = \mu(\alpha_2^j).$$

On note $\mu_X = \mu_{(X_1,X_2)}$ l'image de μ par (X_1,X_2) sur $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$, μ_{X_1} et μ_{X_2} ses marges et λ le produit $\mu_{X_1} \otimes \mu_{X_2}$.

μ_X admet relativement à λ une densité qu'il suffit d'expliciter sur les pq modalités $(\{x_1^i\} \times \{x_2^j\})$ de (X_1,X_2) ;

$$\frac{d\mu_{(X_1,X_2)}}{d\lambda}(\{x_1^i\} \times \{x_2^j\}) = \frac{\mu(\alpha_1^i \cap \alpha_2^j)}{\mu(\alpha_1^i)\cdot\mu(\alpha_2^j)} = \frac{p_{ij}}{p_{i.} \times p_{.j}}$$

On sait (cf. J. DAUXOIS et A. POUSSE [4]) que l'A.F.C. de (X_1,X_2) peut alors être définie comme l'A.C. de μ_X. Elle est obtenue par l'analyse spectrale de l'opérateur $E^{X_1} \circ E^{X_2}{}_{/L^2(\mathcal{Q}_1)}$ où l'on a :

$$E^{X_1} : g \in L^2(\mathcal{Q}_2) \longrightarrow h = E^{X_1}(g) \qquad \text{tq} : h(i) = \sum_{j=1}^{q} \frac{p_{ij}}{p_{i.}}\, g(j)$$

$$E^{X_2} : f \in L^2(\mathcal{Q}_1) \longrightarrow k = E^{X_2}(f) \qquad \text{tq} : k(j) = \sum_{i=1}^{p} \frac{p_{ij}}{p_{.j}}\, f(i)$$

où $f(i)$ (resp. $g(j)$) désigne la valeur prise sur α_1^i (resp. α_2^j) pour $i=1,\dots,p$ (resp. $j=1,\dots,q$).

Ces opérateurs ont pour matrices, relativement aux bases $\{\mathbb{I}_{\alpha_1^i}\}_{i=1,\dots,p}$ et $\{\mathbb{I}_{\alpha_2^j}\}_{j=1,\dots,q}$ de $L^2(\mathcal{Q}_1)$ et $L^2(\mathcal{Q}_2)$, respectivement $D_1^{-1}T$ et $D_2^{-1}\,{}^tT$, où T est la matrice de terme général p_{ij} et D_1 (resp. D_2) la matrice $\mathrm{diag}(p_{i.})_{i=1,\dots,p}$ (resp. $\mathrm{diag}(P_{.j})_{j=1,\dots,q}$).

L'A.F.C. est alors obtenue par l'analyse spectrale de la matrice : $D_1^{-1}\, T\, D_2^{-1}\, {}^tT$.

A.F.C. en tant qu'A.C.P.

Si on note encore (on confond les notations par souci de simplification) : $X_1 = \{\mathbb{I}_{\alpha_1^i}\}_{i=1,\dots,p}$ et $X_2 = \{\mathbb{I}_{\alpha_2^j}\}_{j=1,\dots,q}$ les paquets d'indicatrices, une façon plus classique d'introduire l'A.F.C. de (X_1,X_2) est de la définir (cf. par exemple CAZES et all. [3]) comme l'Analyse en Composantes Principales (A.C.P.) du tableau $Z = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ obtenu en superposant les deux paquets d'indicatrices. Le schéma de dualité associé s'écrit :

$$\begin{tikzcd}[column sep=huge]
\mathbb{R}^{p+q} \arrow[d, shift right, "M"'] & L^2(\Omega, \ ,\mu) \simeq \mathbb{R}^n \arrow[l, "Z = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}"'] \arrow[d, shift right, "W"'] \\
\mathbb{R}^{p+q} \arrow[u, shift right, "V"'] \arrow[r, "{}^tZ"] & L^2(\Omega, \ ,\mu) \arrow[u, shift right, "D_n = \frac{1}{n} I_n"']
\end{tikzcd}$$

où $V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$ et $M = \begin{bmatrix} V_{11}^{-1} & 0 \\ 0 & V_{22}^{-1} \end{bmatrix}$ avec $V_{ij} = X_i\, D_n\, {}^tX_j$.

Cette A.C.P. est obtenue par l'analyse spectrale de l'opérateur $W \circ D_n$ représenté par la matrice :

$$ {}^tX_1 \, V_{11}^{-1} \, X_1 \, D_n + {}^tX_2 \, V_{22}^{-1} \, X_2 \, D_n \; . $$

3.2 - A.F.C. codée

Introduction

J.F. MARTIN [8] a déjà décrit quelques applications du T-codage (ou codage flou) aux analyses factorielles (A.C.P. d'une probabilité codée, A.C. non linéaire d'un couple de v.a. codées ...) permettant entre autre d'introduire dans le modèle des notions d'erreurs de mesure, d'arrondi, ou encore de lissage de variables quantitatives. Un autre type d'application est proposé ici, concernant cette fois les variables qualitatives et donc l'A.F.C. . Il permet de tenir compte des erreurs de classement des individus à l'intérieur de chaque modalité ou encore d'estimer l'analyse de la v.a. Y "coûteuse" à observer par l'intermédiaire d'une variable X plus facilement accessible.

A.C. *T-codée d'une probabilité*

Définition 4 :

Soit ν *une probabilité sur* $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ *et* P_m *un T-codage relatif à* $\mathcal{F}_1 \otimes \mathcal{F}_2$ *, on appelle analyse canonique T-codée de* ν*, l'A.C. de la loi* $E_\nu(P)$ *sur l'espace* $(F_1 \times F_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$.

A.F.C. *codée d'un couple de v.a.*

De même, si $X = (X_1, X_2)$ est une v.a. définie sur $(\Omega, \mathcal{A}, \mu)$ à valeurs dans $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ muni de la loi image μ_X de μ par X, on a :

Définition 5 :

Si P est un T-codage sur $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ *relatif à* $\mathcal{F}_1 \otimes \mathcal{F}_2$*, on appelle A.F.C. T-codée de X, l'A.C. de la loi* $E_{\mu_X}(P)$ *sur l'espace produit* $(F_1 \times F_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$.

Remarque :

Si ${}^1P$ et ${}^2P$ désignent les T-codages marges de P, cette analyse est obtenue à partir de l'explicitation de la densité :

$$\frac{dE_{\mu}(P_X)}{d[{}^1E_{\mu}(P_X) \otimes {}^2E_{\mu}(P_X)]} = \frac{dE_{\mu_X}(P)}{d[E_{\mu_X}({}^1P) \otimes E_{\mu_X}({}^2P)]}$$

3.3 - m-A.F.C. codée

Définition

Une généralisation possible de l'A.F.C. à plus de deux variables qualitatives s'établit à partir de la définition donnée en 3.1.

Considérons m variables ayant chacune p_i modalités, on est conduit au schéma de dualité suivant :

$$\begin{tikzcd}[column sep=huge, row sep=huge]
\mathbb{R}^{\sum_{i=1}^{m} p_i} \arrow[d, shift right, "M"'] & L^2(\Omega,\mathcal{A},\mu) \simeq \mathbb{R}^n \arrow[l, "Z = \begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix}"'] \arrow[d, shift right, "W"'] \\
\mathbb{R}^{\sum_{i=1}^{m} p_i} \arrow[u, shift right, "V"'] \arrow[r, "{}^tZ"] & \mathbb{R}^n \arrow[u, shift right, "D_n = 1/n\, I_n"']
\end{tikzcd}$$

où $V = \begin{bmatrix} V_{ij} \end{bmatrix}_{\substack{i=1,\dots,m \\ j=1,\dots,m}}$ avec $V_{ij} = X_i\, D_n\, {}^tX_i$ et $M = \mathrm{diag}(V_{ii}^{-1})_{i=1,\dots,m}$.

L'analyse cherchée est alors obtenue par l'analyse spectrale de l'opérateur ${}^tZMZD_n$ ou encore par celle de l'opérateur $M \circ V$.

Remarque :

Dans ce type de généralisation de l'A.F.C., seuls interviennent les tableaux V_{ij} croisant toutes les variables deux à deux ; c'est-à-dire qu'il n'est pas nécessaire de connaître la loi de $X = (X_1,\dots,X_m)$ sur toute la tribu $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_m$, mais seulement ses "marges d'ordre deux".

m-A.F.C. codée

La notion de T-codage sur un espace produit introduite en 2.2 - se généralise sans difficulté à plus de deux espaces ; on obtient encore une transition sur $(E = E_1 \times \dots \times E_m\ ,\ \mathcal{E} = (\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_m)) \times (\mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_m)$.

D'autre part, compte tenu de la remarque précédente, la m-analyse ne nécessite que la connaissance des "marges d'ordre deux" de la loi $E_{\mu_X}(P)$ pour le calcul de la densité, soit :

$$\forall (i,j) \in \{1,\ldots,m\}^2 \quad , \quad {}^{ij}[E_{\mu_X}(P)] = E_{\mu_X}({}^{ij}P)$$

(propriété analogue à la proposition 1)
et donc il suffit de connaître les T-codages marges ${}^{ij}P$ sur $(E,\mathcal{E})$ relatifs à $\mathcal{F}_i \otimes \mathcal{F}_j$.

4 - APPLICATION AU PROBLEME CONCRET

4.1 - Codage conditionnel sur un espace produit

Position du problème

On souhaite étudier une variable Y inobservable, ou trop coûteuse à observer, à l'aide d'une variable X facilement observable mais entachée d'erreurs aléatoires (non nécessairement indépendantes ou additives). Il est alors impossible de déterminer une fonction de transfert de X en Y, mais on suppose connue (i.e., par exemple, estimable sur un échantillon-test) la loi de probabilité conditionnelle de Y à X qui permet de déterminer, par l'intermédiaire du codage, la loi de Y connaissant celle de X ($E_{\mu_X}(P) = \mu_Y$ cf. 2.1). On peut ainsi calculer l'A.F.C. de $Y = (Y_1,Y_2)$ qui ne dépend que de μ_Y.

Le T-codage conditionnel

Dans le cas qui nous intéresse, les variables qualitatives X et Y sont des couples et ont un nombre fini de modalités. Soit donc $E = E_1 \times E_2$ et $F = F_1 \times F_2$ avec :

$$E_1 = \{x_1^i / i \in I = \{1,\ldots,p\}\} \quad , \quad E_2 = \{x_2^j / j \in J = \{1,\ldots,q\}\}$$
$$F_1 = \{y_1^k / k \in K = \{1,\ldots,r\}\} \quad , \quad F_2 = \{x_2^\ell / \ell \in L = \{1,\ldots,s\}\}.$$

On définit alors le T-codage conditionnel de Y en X par :

$$\forall (i,j,k,\ell) \in I \times J \times K \times L \qquad \phi_{ijk\ell} = P_{(x_1^i,x_2^j)}(\{y_1^k\} \times \{y_2^\ell\})$$
$$= \mu[(Y_1,Y_2)=(y_1^k,y_2^\ell)/(X_1,X_2)=(x_1^i,x_2^j)]$$

et la fonction de codage se met sous la forme d'une matrice :

$$\Phi = [\phi_{ijk\ell}]_{\substack{(i,j)\in I\times J \\ (k,\ell)\in K\times L}}$$

$\phi_{ijk\ell}$ exprime encore la probabilité de bon ou de mauvais classement : c'est la probabilité que $Y(\omega)$ prenne la modalité (y_1^k, y_2^ℓ) sachant que $X(\omega)$ est observée en (x_1^i, x_2^j).

Hypothèses de simplification

On a vu en 2.2 - que l'on pouvait diminuer le nombre de termes à estimer pour déterminer le T-codage P en l'exprimant sous la forme d'un produit. Afin de simplifier les notations on pose :

$$\alpha_1^i = \overset{-1}{X_1}(\{x_1^i\}),\ \alpha_2^j = \overset{-1}{X_2}(\{x_2^j\}) \ , \quad \beta_1^k = \overset{-1}{Y_1}(\{y_1^k\}) \ ; \quad \beta_2^\ell = \overset{-1}{Y_2}(\{y_2^\ell\}) \ ,$$

et ainsi :

$$\phi_{ijk\ell} = \mu[\beta_1^k \cap \beta_2^\ell /_{\alpha_1^i \cap \alpha_2^j}].$$

Considérons alors les hypothèses suivantes (valables pour tout (i,j,k,ℓ) de $I \times J \times K \times L$) :

$$(H_1) \qquad \mu[\beta_1^k \cap \beta_2^\ell /_{\alpha_1^i \cap \alpha_2^j}] = \mu[\beta_1^k /_{\alpha_1^i \cap \alpha_2^j}] \cdot \mu[\beta_2^\ell /_{\alpha_1^i \cap \alpha_2^j}]$$

$$(H_2) \quad \begin{cases} \mu[\beta_1^k /_{\alpha_1^i \cap \alpha_2^j}] = \mu[\beta_1^k /_{\alpha_1^i}] \\ \text{et} \\ \mu[\beta_2^\ell /_{\alpha_1^i \cap \alpha_2^j}] = \mu[\beta_2^\ell /_{\alpha_2^j}] \end{cases}$$

Proposition 3

(i) $(H_1) \Longleftrightarrow P = {}^1P \otimes {}^2P$

(ii) (H_1) *et* $(H_2) \longleftrightarrow P = P^1 \otimes P^2$

<u>Démonstration</u> :

(i) En explicitant,

$P = {}^1P \otimes {}^2P \Longleftrightarrow \forall x \in E,\ \forall B \in \mathcal{B} \quad P_x(B) = {}^1P_x(B_1)\ .\ {}^2P_x(B_2)$

$\Longleftrightarrow \forall(i,j,k,\ell) \in I\times J\times K\times L \quad \mu[\beta_1^k \cap \beta_2^\ell / \alpha_1^i \cap \alpha_2^j] = \mu[\beta_1^k / \alpha_1^i \cap \alpha_2^j]\ .\ \mu[\beta_2^\ell / \alpha_1^i \cap \alpha_2^j]$

$\Longleftrightarrow (H_1)$

(ii) Supposons (H_1) et (H_2), alors : $\forall(i,j,k,\ell) \in I\times J\times K\times L$

$$P_{(x_1^i,x_2^j)}(\{y_1^k\}\times\{y_2^\ell\}) = \mu[\beta_1^k \cap \beta_2^\ell / \alpha_1^i \cap \alpha_2^j] = \mu[\beta_1^k / \alpha_1^i]\ .\ \mu[\beta_2^\ell / \alpha_2^j]$$

$$= P^1_{x_1^i}(\{y_1^k\})\ .\ P^2_{x_2^j}(\{y_2^\ell\})$$

Réciproquement, si P est égal à $P^1 \otimes P^2$ on a :

$\forall x \in E\ ,\quad \forall B \in \mathcal{B} \qquad P_x(B) = P^1_{x_1}(B_1)\ .\ P^2_{x_2}(B_2)$

en prenant $x = (x_1^i, x_2^j)$ et successivement $B = \{y_1^k\} \times F_2$ et $B = F_1 \times \{y_2^\ell\}$ on obtient (H_2),

puis en prenant $x = (x_1^i, x_2^j)$ et $B = \{y_1^k\} \times \{y_2^\ell\}$ et en utilisant (H_2), on obtient (H_1). □

Signification des hypothèses

(H_1) ou $[(H_1)$ et $(H_2)]$ sont des hypothèses d'indépendance conditionnelle permettant de simplifier l'expression de P.

En écrivant les variables sous la forme :

$Y_i = \psi_i(X_i, \varepsilon_i)$ pour $i=1,2$,

où ψ_i est une fonction mesurable et ε_i une v.a. "mesurant" l'erreur, on remarque que les hypothèses signifient :

* <u>(H_1)</u> : X_1 et X_2 étant connues, Y_1 et Y_2 sont indépendantes ou encore, les erreurs de mesure (ou de classement) sont indépendantes entre elles.

* (H_1) et (H_2) : X_1 et X_2 étant connues ; pour $i=1,2$, Y_i ne dépend que de X_i ou encore, ε_i est indépendante de la valeur de X_{3-i} et, ε_1 et ε_2 sont indépendantes entre elles.

4.2 - Détermination pratique de l'analyse

Explicitation de la densité

Comme il est décrit en 3.1 - , le calcul de l'analyse de Y ne nécessite pas l'observation de Y, mais la connaissance de la loi de Y et, plus précisément, de la densité de cette loi par rapport au produit de ses marges dont l'existence est assurée car E et F sont de dimension finie.

L'expression de cette densité est plus ou moins complexe suivant les hypothèses qu'il est possible d'admettre.

Posons :

$$p_{ij} = \mu_X(\{x_1^i\} \times \{x_2^j\}) = \mu(\alpha_1^i \cap \alpha_2^j).$$

avec pour marges $p_{i.}$ et $p_{.j}$ et :

$$q_{k\ell} = \mu_Y(\{y_1^k\} \times \{y_2^\ell\}) = \mu(\beta_1^k \cap \beta_2^\ell)$$

avec pour marges $q_{k.}$ et $q_{.\ell}$.

La densité s'exprime alors :

$$\frac{dE_{\mu_X}(P)}{d[E_{\mu_X}({}^1P)\otimes E_{\mu_X}({}^2P)]}(\{y_1^k\}\times\{y_2^\ell\}) = \frac{q_{k\ell}}{q_{k.}\times q_{.\ell}} = \frac{\sum\limits_{i,j} \phi_{ijk\ell}\, p_{ij}}{\sum\limits_{\ell=1}^{s}\left(\sum\limits_{i,j}\phi_{ijk\ell}p_{ij}\right)\times\sum\limits_{k=1}^{r}\left(\sum\limits_{i,j}\phi_{ijk\ell}p_{ij}\right)}$$

$$= \frac{\sum\limits_{i,j}\phi_{ijk\ell}p_{ij}}{\left(\sum\limits_{i,j}\phi_{ijk.}\,p_{ij}\right)\times\left(\sum\limits_{i,j}\phi_{ij.\ell}\,p_{ij}\right)}$$

où, par analogie,

$$\phi_{ijk.} = \sum_{\ell=1}^{s}\phi_{ijk\ell} \quad \text{et} \quad \phi_{ij.\ell} = \sum_{k=1}^{r}\phi_{ijk\ell}$$

Avec les mêmes notations on a la :

Proposition 4

(*i*) $$P = {}^1P \otimes {}^2P \Rightarrow \frac{dE_{\mu_X}(P)}{d[E_{\mu_X}({}^1P)\otimes E_{\mu_X}({}^2P)]}(\{y_1^k\}\times\{y_2^\ell\}) = \frac{\sum_{i,j} \phi_{ijk.}\,\phi_{ij.\ell}\,p_{ij}}{(\sum_{i,j}\phi_{ijk.}\,p_{ij})\times(\sum_{i,j}\phi_{ij.\ell}\,p_{ij})}$$

(*ii*) $$P = P^1 \otimes P^2 \Rightarrow \frac{dE_{\mu_X}(P)}{d[E_{\mu_X}({}^1P)\otimes E_{\mu_X}({}^2P)]} = \frac{dE_{\mu_X}(P)}{d[E_{\mu_X}(P^1)\otimes E_{\mu_X}(P^2)]} = \frac{dE_{\mu_X}(P)}{dE_{\mu_{X_1\otimes X_2}}(P)}\,.$$

<u>Démonstration</u> :

(i) Evident

(ii) Si P est de la forme $P^1 \otimes P^2$ on peut poser :

$$\phi_{ijk\ell} = \phi^1_{ik}\,.\,\phi^2_{j\ell}$$

et remarquant que :

$$\forall(i,j) \in I \times J \quad , \quad \sum_{k=1}^{r} \phi^1_{ik} = 1 \quad \text{et} \quad \sum_{\ell=1}^{s} \phi^2_{j\ell} = 1,$$

il vient alors :

$$\frac{dE_{\mu_X}(P)}{d[E_{\mu_X}({}^1P)\otimes E_{\mu_X}({}^2P)]}(\{y_1^k\}\times\{y_2^\ell\}) = \frac{\sum_{i,j} \phi^1_{ik}\,\phi^2_{j\ell}\,p_{ij}}{(\sum_i \phi^1_{ik}\,p_{i.})\times(\sum_j\phi^2_{j\ell}\,p_{.j})} = \frac{\sum_{i,j} \phi^1_{ik}\,\phi^2_{j\ell}\,p_{ij}}{\sum_{i,j} \phi^1_{ik}\,\phi^2_{j\ell}\,p_{i.}\times p_{.j}}$$

□

A.F.C. *de* (Y_1,Y_2)

On cherche donc l'A.F.C. de Y connaissant X. Pour ce faire il faut calculer l'A.C. de $E_{\mu_X}(P)$ qui est la loi de Y si P est le T-codage conditionnel. En fait les calculs présentés ci-dessous sont valables pour tout T-codage P pourvu qu'il se mette sous la forme élémentaire : $P = P^1 \otimes P^2$ (le codage de X est égal au produit des codages de X_1 et X_2) ; c'est-à-dire, dans le cas du codage conditionnel, si les hypothèses (H_1) et (H_2) sont vérifiées.

On note :

$${}^t\Phi_1 = \left[\phi^1_{ik}\right]_{\substack{i=1,\dots,p\\k=1,\dots,r}} \quad \text{et} \quad {}^t\Phi_2 = \left[\phi^2_{j\ell}\right]_{\substack{j=1,\dots,q\\\ell=1,\dots,s}}$$

$$C = X_1 \, D_n \, {}^tX_2 = \left[p_{ij}\right]_{\substack{i=1,\dots,p \\ j=1,\dots,q}} \qquad \text{et} \qquad C' = \left[q_{k\ell}\right]_{\substack{k=1,\dots,r \\ \ell=1,\dots,s}}$$

C étant la table de contingence associée à X, C' est celle associée à Y et s'écrit aussi :

$$C' = \Phi_1 C \, {}^t\Phi_2 \ .$$

Le schéma de dualité associé à l'A.C. de $E_{\mu_X}(P)$ (équivalente à l'A.F.C. du couple (Y_1,Y_2)) est le suivant :

$$\begin{tikzcd}
\mathbb{R}^r \arrow[d, "N_1"'] & \mathbb{R}^p \arrow[l, "\Phi_1"'] & \mathbb{R}^n \arrow[l, "X_1"'] \arrow[r, "X_2"] & \mathbb{R}^q \arrow[r, "\Phi_2"] & \mathbb{R}^s \arrow[d, "N_2"] \\
\mathbb{R}^r \arrow[r, "{}^t\Phi_1"] & \mathbb{R}^p \arrow[r, "{}^tX_1"] & \mathbb{R}^n \arrow[u, "D_n = Y_n I_n"'] & \mathbb{R}^q \arrow[l, "{}^tX_2"'] & \mathbb{R}^s \arrow[l, "{}^t\Phi_2"']
\end{tikzcd}$$

où $N_1 = \mathrm{diag}[(E_{\mu_X}({}^1P)(\{y_1^k\}))^{-1}]_{k=1,\dots,r}$

et $N_2 = \mathrm{diag}[(E_{\mu_X}({}^2P)(\{y_2^\ell\}))^{-1}]_{\ell=1,\dots,s}$

L'A.F.C. cherchée est obtenue par l'analyse spectrale d'un opérateur ayant pour représentation matricielle :

$$N_1 \, \Phi_1 \, X_1 \, D_n \, {}^tX_2 \, {}^t\Phi_2 \, N_2 \, \Phi_2 \, X_2 \, D_n \, {}^tX_1 \, {}^t\Phi_1 = N_1 \, C' \, N_2 \, {}^tC' .$$

En pratique, et ce quelles que soient les hypothèses admises, il suffit donc de faire exécuter un programme classique d'A.F.C. sur le tableau C' de terme général $q_{k\ell}$.

Sous les hypothèses (H_1) et (H_2), cette matrice s'exprime simplement par : $\Phi_1 \, C \, {}^t\Phi_2$. L'analyse obtenue est équivalente à l'A.C.P. du tableau :

$$\begin{bmatrix} \Phi_1 & 0 \\ 0 & \Phi_2 \end{bmatrix} \times \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

les espaces étant munis des métriques N_1 et N_2.

Remarque 1 :

Dans le cas très particulier où ${}^t[\mu(\{x_1^i\})]_{i=1,\dots,p}$ et ${}^t[\mu(\{x_2^j\})]_{j=1,\dots,q}$ sont respectivement vecteurs propres associés à la valeur propre 1 des matrices ${}^t\Phi_1$ et ${}^t\Phi_2$, J.F. MARTIN [8] a montré que μ_X était invariante par P, i.e. :

$$\mu_Y = E_{\mu_X}(P) = \mu_X .$$

Dans ce cas, et dans ce cas seulement, les analyses de X et de Y sont les mêmes, et nous dirons que l'A.F.C. de X est invariante par P.

Remarque 2 :

On peut appliquer cette méthode aux processus prévisionnels.

Dans l'hypothèse où X_1 et X_2 sont des chaînes de Markov de matrices de transition Φ_1 et Φ_2 , le point de vue adopté ici rejoint celui de A. YOUSFATE [13] : connaissant l'état initial de (X_1,X_2), on peut calculer les A.F.C. aux différents instants.

4.3 - Cas de l'A.F.C. généralisée

Position du problème

On se propose de généraliser ce qui précède à la situation suivante : soit $Y = (Y_1,\dots,Y_m)$ une v.a. définie sur $(\Omega,\mathcal{Q},\mu)$ à valeurs dans $(F = F_1 \times \dots \times F_m$, $\mathcal{F} = \mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_m)$ et $X = (X_1,\dots,X_m)$ à valeurs dans $(E = E_1 \times \dots \times E_m$, $\mathcal{E} = \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_m)$. Soit pour tout i, Card $E_i = p_i$ et nous utiliserons la définition de la m-A.F.C. codée donnée en 3.3 .

Cette analyse ne pose pas de problème théorique spécifique mais rend encore plus délicate la mise en oeuvre pratique à cause de la multiplication des hypothèses à considérer.

Hypothèses de simplification

Comme l'analyse de Y ne tient pas compte des relations (interactions) intervenant entre plus de deux variables, elle ne nécessite pas la connaissance du T-codage conditionnel P, mais seulement celle des marges ${}^{ij}P$ et donc une hypothèse du type de (H_1) n'est plus nécessaire.

On cherche donc à simplifier l'expression des m^2 transitions :

$$\begin{aligned} {}^{ij}P : (E,\mathcal{E}) \times (\mathcal{F}_i \otimes \mathcal{F}_j) &\longrightarrow ([0,1] , \mathcal{B}_{[0,1]}) \\ (x,B_{ij}) &\longrightarrow {}^{ij}P_x(B_{ij}) = \mu[Y \in B_{ij/X=x}]. \end{aligned}$$

5 - ESTIMATION DU CODAGE ET CONVERGENCE

5.1 - Estimation

En pratique, la loi conditionnelle de Y à X est inconnue et il est alors nécessaire de l'estimer sur un échantillon-test. Pour simplifier les notations nous considérons X et Y au lieu des couples (X_1,X_2), (Y_1,Y_2) et nous noterons $\Phi = [\phi_{ik}]$ (au lieu de $[\phi_{ijk\ell}]$) la matrice de codage.

On suppose que l'on observe un échantillon-test $(X_k,Y_k)_{k=1,\ldots,m}$ de (X,Y), puis un échantillon $(X_{k'})_{k'=m+1,\ldots,m+n}$ de X, indépendant du précédent.

Si M_{ik} désigne l'effectif de la modalité (x_i,y_k) de la variable (X,Y) pour l'échantillon de taille m et N_i l'effectif de la modalité x_i pour l'échantillon de taille m+n, l'estimateur du maximum de vraisemblance de ϕ_{ik} (voir J.F. MARTIN [8]) s'exprime :

$$\hat{\phi}_{ik} = \phi^m_{ik} = P^m_{x_i}(\{y_k\}) \begin{cases} = \dfrac{\sum_{h=1}^{m} \mathbb{1}_{\{x_i\}\times\{y_k\}} \circ (X_h,Y_h)}{\sum_{h=1}^{m} \mathbb{1}_{\{x_i\}} \circ X_h} & \text{si } \sum_{h=1}^{m} \mathbb{1}_{\{x_i\}} \circ X_h \neq 0 \\ = 0 \text{ sinon} \end{cases}$$

c'est-à-dire
$$\hat{\phi}_{ik} = \frac{M_{ik}}{M_{i.}}$$

L'estimateur de μ_X est alors : $\hat{\mu}_X(\{x_i\}) = \dfrac{M_{i.}+N_i}{m+n}$.

Ces deux estimateurs permettant de construire celui de la loi de Y :
$$\hat{\mu}_Y(\{y_k\}) = \sum_{i=1}^{p} \hat{\phi}_{ik}\, \hat{\mu}_X(\{x_i\}) = \frac{1}{m+n} \sum_{i=1}^{p} M_{ik} \frac{M_{i.}+N_i}{M_{i.}} .$$

5.2 - Convergence de T-codages

Position du problème

Soit P un T-codage sur $(E,\mathcal{E})$ relatif à $\mathcal{F}$ et $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a. définies sur $(\Omega,\mathcal{A},\mu)$, à valeurs dans $(E,\mathcal{E})$, indépendantes et de même loi que X. Pour tout n de $\mathbb{N}^*$, $(X_1,\ldots,X_n)$ est un n-échantillon de X auquel on peut associer le n-uplet de probabilités de transition $(P_{X_1},\ldots,P_{X_n})$ appelé échantillon T-codé.

On s'intéresse donc à la convergence, en un sens à préciser, de la suite $(\frac{1}{n}\sum_{i=1}^{n} P_{X_i})_{n\in\mathbb{N}^*}$ des moyennes empiriques des échantillons T-codés vers la probabilité T-codée $E_\mu(P_X)$.

De plus, comme le T-codage P est lui-même estimé, il faut étudier la double convergence de la suite

$$\left[\frac{1}{n}\left(\sum_{i=1}^{n} P^m_{X_i}\right)\right]_{(m,n)\in N^2_*} .$$

Convergence vague et convergence faible

Soient P et $(P^m)_{m\in N^*}$ des T-codages sur $(E,\mathcal{E},\nu)$ relatifs à $\mathcal{F}$. On peut définir comme le fait J.F. MARTIN [8] une "convergence vague ν-presque sûre (ν-p.s.)".

Définition 6 :

On dit que la suite $(P^m)_{m\in N^*}$ converge vaguement ν-p.s. vers P si pour tout B de $\mathcal{F}$ la suite de v.a. $(P^m(B))_{m\in N^*}$ converge ν-p.s. vers la v.a. P(B), i.e. :

(1) $\forall B \in \mathcal{F} \quad \nu\{x \in E / \lim_{m\to+\infty} P^m_x(B) = P_x(B)\} = 1.$

Généralement (1) est moins exigeante que la propriété :

(2) $\nu\{x \in E / \forall B \in \mathcal{F} \ \lim_{m\to+\infty} P^m_x(B) = P_x(B)\} = 1.$

Cependant, ici $\mathcal{F}$ étant une tribu finie, il est clair que (1) et (2) sont équivalentes. On utilisera donc (2) et on a la :

Définition 7 :

$(P^m)_{m\in N^*}$ converge faiblement ν-p.s. vers P si pour presque tout x *de* E *la suite de probabilités $(P^m_x)_{m\in N^*}$ converge faiblement vers P_x (i.e. (2)).*

Remarque :

Nous utilisons la notion de convergence faible puisque, $\mathcal{F}$ étant finie, les indicatrices $\mathbb{I}_B$ sont continues.

5.3 - Convergence de l'analyse

Convergence du T-codage conditionnel

En reprenant les notations de 5.1 -, le T-codage conditionnel est déterminé par :

$$P_{x_i}(\{y_k\}) = \frac{1}{\mu_X(\{x_i\})} \int_{[X=x_i]} \mathbb{I}_{\{y_k\}} \circ Y \; d\mu ;$$

Soit (H_{ij}) : ${}^{ij}P_X(B_{ij}) = \mu[Y \in B_{ij}/(X_i,X_j)=(x_i,x_j)]$

Ce sont des hypothèses du même type que (H_2) permettant de définir la transition ${}^{ij}P$ seulement sur l'espace $E_i \times E_j$ au lieu de E tout entier. Sommairement (H_{ij}) signifie que l'erreur commise sur la "mesure" de (X_i,X_j) est indépendante des valeurs prises par les autres variables X_k $(k \neq i$ et $k \neq j)$.

Ainsi si pour tout couple (i,j), l'hypothèse (H_{ij}) est vérifiée, le problème se ramène alors à m^2 fois la situation précédente de l'A.F.C. d'un couple de v.a..

Pour chacun des couples (X_i,X_j), on peut considérer les hypothèses (H_1^{ij}) et (H_2^{ij}) comme définies en 4.1 - et évaluer alors :

$$V'_{ij} = E_{\mu_X}({}^{ij}P)$$

(dont la simplicité relative dépend de l'admissibilité de (H_1^{ij}) et (H_2^{ij})).

m-analyse

L'ensemble de ces blocs V_{ij} permet de construire la matrice à diagonaliser :

$$M' \circ V' = \begin{bmatrix} V'^{-1}_{11} & & 0 \\ & \ddots & \\ 0 & & V'^{-1}_{mm} \end{bmatrix} \times \begin{bmatrix} V'_{ij} \end{bmatrix}_{(i,j) \in I^2}$$

Sous le jeu complet des hypothèses :

$$\forall (i,j) \in I^2 \quad (H_{ij}) \wedge (H_1^{ij}) \wedge (H_2^{ij})$$

et en se référent au paragraphe 3.3 - , on obtient le schéma de dualité :

$$\begin{tikzcd}
\mathbb{R}^{\sum_{i=1}^m q_i} \arrow[d, shift right, "M'"'] & & \mathbb{R}^{\sum_{i=1}^m p_i} \arrow[ll, "\Phi = \begin{bmatrix} \Phi_1 & & 0 \\ & \ddots & \\ 0 & & \Phi_m \end{bmatrix}"'] & & \mathbb{R}^n \arrow[ll, "Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_m \end{bmatrix}"'] \\
\mathbb{R}^{\sum_{i=1}^m q_i} \arrow[u, shift right, "V'"'] \arrow[rr, "{}^t\Phi"] & & \mathbb{R}^{\sum_{i=1}^m p_i} \arrow[rr, "{}^tZ"] & & \mathbb{R}^n \arrow[u, "D_n"']
\end{tikzcd}$$

où $V' = \Phi\, V\, {}^t\Phi$.

Ce schéma correspond à l'A.C.P. des variables $(\Phi_1 X_1, \dots, \Phi_m X_m)$ lorsque l'espace des "individus" est muni de la métrique M'.

codage que l'on peut estimer sur l'échantillon-test $(X_i,Y_i)_{i=1,\dots,m}$ par :

$$P^m_{x_i}(\{y_k\}) = \begin{cases} \dfrac{\sum_{j=1}^{m} \mathbb{1}_{\{x_i\}\times\{y_k\}} \circ (X_j,Y_j)}{\sum_{j=1}^{m} \mathbb{1}_{\{x_i\}} \circ X_j} & \text{si } \sum_{j=1}^{m} \mathbb{1}_{\{x_i\}} \circ X_j \neq 0 \\ 0 & \text{sinon.} \end{cases}$$

Les tribus $\mathcal{E}$ et $\mathcal{F}$ étant finies, la suite de probabilités $(P^m_x)_{m\in\mathbb{N}^*}$ converge faiblement vers P_x uniformément en x sur $\mathcal{E}$ et on peut donc appliquer la proposition 6 p. 74 de J.F. MARTIN [8] montrant ainsi que :

$$\lim_{\substack{m\to+\infty \\ n\to+\infty}} \frac{1}{n} \sum_{i=1}^{n} P^m_{x_i} = E_\mu(P_X) \text{ faiblement } \mu\text{-p.s.}$$

Convergence de l'analyse

En utilisant le résultat ci-dessus pour $X = (X_1,X_2)$ et $Y = (Y_1,Y_2)$, on peut appliquer la proposition 9 p. 291 de J. DAUXOIS et A. POUSSE [4]. On en déduit la convergence uniforme p.s. de la suite des analyses obtenues par échantillonnage vers celles de Y et ainsi , la convergence des éléments propres (valeurs propres et facteurs canoniques de l'analyse) .

6 - CONCLUSION

Les outils développés permettent donc bien de résoudre théoriquement le problème concret posé en [1]. En pratique il faut être conscient que les difficultés dues à l'estimation de la probabilité de transition sont délicates à traitées. On a vu que pour simplifier ces questions il fallait pouvoir décomposer le codage P en un produit de ses marges, ceci ne peut se justifier alors que par la considération d'hypothèses de type probabiliste. Ainsi apparaît la nécessité d'introduire des outils adéquats (probabilités de transition) pour modéliser la notion de codage dans le cadre des analyses factorielles non-linéaires.

D'autres types d'application que celle proposée en [1] peuvent être suggérées. Il est fréquent de vouloir réactualiser une enquête ou un sondage sans pour autant interroger un échantillon complet. Il suffirait alors de ne réenquêter qu'un sous-échantillon de l'échantillon initial permettant d'estimer une probabilité de transition et donc, d'actualiser les représentations factorielles.

Dans un autre ordre d'idée, on considère les hypothèses proposées par YOUSFATE 15 (chaîne de Markov homogène d'ordre 1) dans le but de prévoir un processus qualitatif. Alors, connaissant les transitions de 2 ou plusieurs de ces processus on peut prévoir les représentations factorielles (A.F.C. ou m-A.F.C.) associées à ces processus.

REFERENCES

[1] - BESSE P. : *"Recherche statistique d'une typologie des descriptions de phénomènes aérospatiaux non identifiés".*
CNES/GEPAN - note technique n° 4 - TOULOUSE (1981)

[2] - BORDET J.P. : *"Etude de données géophysiques. Modélisations statistiques par régression factorielle".*
Thèse de 3° cycle - Université de PARIS VI,(1973).

[3] - CAZES P. - BAUMEDER A. - BONNEFOUS S. - PAGES J.P. : *"Codage et tableaux des Données logiques - Introduction à la pratique des variables qualitatives".*
Cahiers du B.U.R.O. n° 27 - PARIS (1977).

[4] - DAUXOIS J. - POUSSE A. : *"Les analyses factorielles en calcul des probabilités et en statistique : essai d'étude synthétique".*
Thèse - Université de TOULOUSE III (1976).

[5] - GAUTIER J.M. - SAPORTA G. : *"About fuzzy discrimination".*
COMPSTAT 82 - Proceedings in Computational Statistics, Physica-Verlag, WIEN (1982).

[6] - LAFAYE DE MICHEAUX D. : *"Approximation d'analyses canoniques non-linéaires de variables aléatoires et analyses factorielles privilégiantes".*
Thèse de docteur ingénieur - Université de NICE (1978).

[7] - MALLET J.L. : *"Propositions for fuzzy characteristic functions in data analysis".*
COMPSTAT 82 - Proceedings in Computational Statistics - Physica-Verlag, WIEN (1982).

[8] - MARTIN J.F. : *"Le codage flou et ses applications en statistique".*
Thèse de 3° cycle - Université de PAU et des pays de l'Adour (1980).

[9] - MARTIN J.F. : *"Le codage aléatoire - Utilisation en statistique".*
Journées de Statistiques - NANCY (1981).

[10] - NEVEU J. : *"Bases mathématiques du calcul des probabilités".*
MASSON - PARIS (1964).

[11] - RAMSAY J.O. - WINSBERG S. : *"Monotonic transformations to additivity using splines".*
Biométrika - 1980, 67, 3, pp. 669-74.

[12] - RAMSAY J.O. - WINSBERG S. : *"Analysis of pairwise preference data using B-splines".*
Psychometrika - vol. 46, N° 2, june, 1981. pp 171-186

[13] - RIJCKEVORSEL (VAN) J. : *"Canonical analysis with B-splines".*
COMPSTAT 82 - Proceedings in Computational Statistics, Physica-Verlag, WIEN (1982).

[14] - TITTERINGTON D.M. : *"A comparative study of kernel-based density estimates of categorical data".*
TECHNOMETRICS - vol. 22, n° 2, mai 1980. pp 259-268

[15] - YOUSFATE A. : *"Analyses factorielles des processus qualitatifs de type Markovien. Description et prévision".*
Thèse de 3° cycle - Université de TOULOUSE III (1981).