REVUE DE STATISTIQUE APPLIQUÉE

GILBERT SAPORTA

Dépendance et codages de deux variables aléatoires

Revue de statistique appliquée, tome 23, n° 1 (1975), p. 43-63

<http://www.numdam.org/item?id=RSA_1975__23_1_43_0>

DÉPENDANCE ET CODAGES DE DEUX VARIABLES ALÉATOIRES [1]

Gilbert SAPORTA
IUT Informatique - Université de Paris V

INTRODUCTION

Le point de départ de cette étude est la question classique suivante : une distribution statistique à deux dimensions est connue sous la forme d'un tableau de contingence m x p où les distributions marginales ne correspondent pas à des valeurs numériques mais à des modalités de caractère qualitatif (par ex : opinion politique et statut social) : comment attribuer à ces diverses modalités des valeurs numériques qui ne soient pas purement conventionnelles mais aient une signification ?

Il s'agit en somme de rendre quantitatives des variables qualitatives par le choix d'une paramétrisation adéquate des diverses modalités, ce qu'on appelle un codage [7] (en anglais "scores" [6]). Le codage cherché sera celui qui respecte au mieux la liaison entre les deux variables décrites par le tableau de contingence, ce qui implique le choix d'une mesure de liaison.

La résolution de ce problème conduit d'une part à étudier les diverses mesures de liaison statistique et à approfondir la notion de corrélation, d'autre part à étudier les transformations de variables aléatoires et l'influence de ces transformations sur la mesure de liaison afin d'obtenir une liaison maximale ce qui fournit les codages souhaités. Nous porterons notre attention sur des transformations donnant des variables centrées réduites car les mesures de liaison étant insensibles aux transformations linéaires, les résultats sont définis à un changement d'origine et un coefficient d'échelle près.

On constatera alors l'identité formelle entre cette démarche qui est celle de l'analyse canonique et l'analyse des correspondances.

La recherche de codages de variables qualitatives peut se réveler particulièrement fructueuse là où l'introduction de données qualitatives s'accorde mal avec une technique statistique essentiellement quantitative par ex : régression, analyse discriminante.

On trouvera des développements théoriques récents dans les travaux [9] de M. Masson, ainsi que dans un article de Pousse et Dauxois [8] dont nous avons eu connaissance ultérieurement.

---

(1) Article remis le 6/2/74 ; révisé le 20/5/74.

I – DEPENDANCE ENTRE DEUX VARIABLES ALEATOIRES, VARIABLES CANONIQUES

Nous supposerons dans cette partie que les variables X, Y étudiées sont continues, de densités marginales $f(x)$ et $g(y)$, la densité du couple étant notée $h(x, y)$. X et Y seront de plus supposées centrées réduites afin d'alléger les notations.

1/ Les indicateurs de liaison

Parmi tous les indicateurs possibles nous étudierons les deux plus importants : le coefficient de corrélation linéaire et le rapport de corrélation qui dépendent des valeurs des deux variables et le ϕ^2 de Karl Pearson qui constitue une mesure intrinsèque de la dépendance. Les rapports entre ces trois mesures seront établis au cours de cette étude.

a) L'indicateur de liaison linéaire ρ (ou ρ^2)

Le coefficient de corrélation linéaire ρ égal à $\int\int xy\, h(x, y)\, dx\, dy$ est la mesure de liaison la plus employée. On sait que $\rho^2 = 1$ entraîne que X et Y sont en relation linéaire (ici $X = Y$ car les variables sont centrées réduites) et que $\rho^2 = 0$ n'est pas un critère d'indépendance mais de non dépendance linéaire.

D'une manière générale ρ^2 mesure la qualité de l'approximation de Y (resp X) par une fonction linéaire X (resp de Y) : pour des variables centrées réduites la meilleure approximation linéaire de Y au sens des moindres carrés est fournie par $Y^* = \rho X$ et si on pose $Y = \rho X + \epsilon$ on a $E(\epsilon) = 0$ $V(\epsilon) = 1 - \rho^2$ et $V(Y^*) = \rho^2$.

ρ est un indicateur symétrique en X et Y mais est sensible aux changements de variable autres que linéaires comme le montre sa définition :

$$\rho(X, Y) \neq \rho(X, \Psi(Y)) \qquad \rho(X, Y) \neq \rho(\Psi(X), Y)$$

b) Les indicateurs de la dépendance en moyenne :

Les rapports de corrélation $\eta^2_{Y/X}$ et $\eta^2_{X/Y}$.

Si on recherche la meilleure approximation fonctionnelle et non plus linéaire de Y par X c'est-à-dire minimiser $E[(Y - f(X))^2]$.

La solution est donné par $f(X) = E(Y/X)$.

On pose alors

$$\eta^2_{Y/X} = \frac{V[E(Y/X)]}{V(Y)}$$

et on a en général $\eta^2_{Y/X} \neq \eta^2_{X/Y}$. Si on écrit comme pour l'approximation linéaire $Y = E(Y/X) + \epsilon$ on trouve $E(\epsilon) = 0$ et $V(\epsilon) = (1 - \eta^2_{Y/X})\, V(Y)$ ce qui prouve que :

$$\eta^2_{Y/X} \geqslant \rho^2$$

car l'ajustement par une fonction quelconque aboutit nécessairement à un résultat meilleur que l'ajustement par une fonction linéaire.

D'après la définition $0 \leqslant \eta^2_{Y/X} \leqslant 1$; si $\eta^2_{Y/X} = 1$

on a presque surement $Y = \varphi(X)$ ce qui n'implique pas nécessairement d'ailleurs $X = \psi(Y)$; si $\eta^2_{Y/X} = 0$ on ne peut conclure à l'indépendance mais seulement à la dépendance en moyenne car alors $E(Y/X) =$ constante.

Si Y est centré réduit $\eta^2_{Y/X} = V[E(Y/X)]$ et comme

$$E(Y/X) = x) = \int_{-\infty}^{+\infty} y \frac{h(x,y)}{f(x)} dy \quad \text{il vient} \quad \eta^2_{Y/X} = \int\int_{R^2} y^2 \frac{(h(x,y))^2}{f(x)} dxdy$$

on constate que $\eta^2_{Y/X}$ est insensible à un changement de variable sur X mais non sur Y :

$$\eta^2_{Y/X} = \eta^2_{Y/\varphi(X)} \neq \eta^2_{\psi(Y)/\varphi(X)} = \eta^2_{\psi(Y)/X}$$

si ψ est une transformation non linéaire.

Lorsque $\eta^2_{Y/X} = \rho^2$ c'est que $E(Y/X) = aX + b$

(régression linéaire) ce qui n'implique pas que $E(X/Y) = \alpha Y + \beta$.

On dit qu'il y a double régression linéaire si

$$E(Y/X) = aX + b \quad \text{et} \quad E(X/Y) = \alpha Y + \beta$$

dans ce cas $$\eta^2_{Y/X} = \eta^2_{X/Y} = \rho^2$$

ce qui est d'ailleurs le seul cas où les deux rapports de corrélation sont égaux. Pour des variables centrées réduites $E(Y/X) = \rho X$, $E(X/Y) = \rho Y$. Cette situation (double régression linéaire) se rencontre si (X, Y) suit une loi normale à deux dimensions mais ce n'est pas le seul cas. Toutefois on a le résultat suivant (Bernstein, Féron, Fourgeaud) :

"Si deux variables X et Y sont en double régression linéaire et si de plus les variables ϵ et ϵ' telle que

$$Y = \alpha X + \beta + \epsilon$$

$$X = \alpha' Y + \beta' + \epsilon' \qquad \text{sont}$$

indépendantes de X et de Y respectivement (condition dite des "erreurs identiques") alors (X , Y) soit une loi normale à deux dimensions".

(Référence : Kendall et Stuart Tome II, p. 352).

Le lien entre le coefficient de corrélation linéaire et le rapport de corrélation est précisé par le théorème suivant :

Théorème

Le rapport de corrélation de Y en X, $\eta^2_{y/x}$, est la valeur maximale que peut prendre le carré du coefficient de corrélation linéaire entre Y et n'importe quelle fonction de X :

$$\eta^2_{y/x} = \sup_f \rho^2 \ (Y \ ; f(X))$$

le sup étant atteint pour $f(X) = E(Y/X)$:

$$\eta^2_{y/x} = \rho^2 (Y \ ; E(Y/X)$$

Preuve :

Comme $\eta^2_{y/x} = \eta^2_{y/E} \ (Y/X)$ et qu'il y a alors regression linéaire de Y sur $E(y/x)$ c'est que $\eta^2_{y/x} = \rho^2 (Y, E(Y/X))$.

Par ailleurs $\eta^2_{Y/X} = \eta^2{}_{Y/f(X)} \geqslant \rho^2 (Y \ ; f(X))$ l'égalité n'étant atteinte que si $E(Y/f(X)) = f(X)$.

On peut aussi remarque que pour des variables centrées réduites

$$\rho(Y \ ; E(Y/X)) = \frac{E[Y\,E(Y/X)]}{\sigma_{E(Y/X)}} = \frac{E[(E(Y/X))^2]}{\eta_{Y/X}} = \frac{V[E(Y/X)]}{\eta_{Y/X}}$$

$$= \frac{\eta^2}{\eta} = \eta_{Y/X}$$

On peut donc dire que les rapports de corrélation constituent des mesures de la dépendance entre X et Y plus générales que le coefficient de corrélation inéaire.

c) Le ϕ^2 de Karl Pearson

Etant donné deux mesures de probabilité ρ et ν sur un même ensemble E on peut définir (Lancaster) [1] la proximité entre μ et ν par rapport à ν de la manière suivante :

pour une partition $E_1, E_2, \ldots, E_k$ de E on calcule la quantité

$$\sum_{i=1}^{k} \left(\frac{\nu(E_i)}{\mu(E_i)}\right)^2 \mu(E_i)$$

dont ont étudie la limite en passant à des partitions de plus en plus fines. Cette limite, si elle existe, est alors égale à

$$1 + \phi^2 = \int_E \left(\frac{d\nu}{d\mu}\right)^2 d\mu \quad \text{où} \quad \frac{d\nu}{d\mu}$$

est la dérivée de Radon-Nikodym de μ par rapport à ν. La quantité ϕ^2 n'est égale à 0 que si $\nu = \mu$ μ-presque partout.

Dans les problèmes de dépendance entre variables aléatoires on étudie en fait en quoi $h(x, y)$ diffère du produit $f(x)\,g(y)$, c'est-à-dire la proximité entre la mesure du couple et la mesure produit d'où la définition du ϕ^2 de dépendance entre X et Y par :

$$1 + \phi^2 = \iint_{R^2} \frac{[h(x,y)]^2}{f(x)\,g(y)}\,dx\,dy = \iint_{R^2} \left[\frac{h(x,y)}{f(x)\,g(y)}\right]^2 f(x)\,g(y)\,dx\,dy$$

soit

$$\phi^2 = \iint_{R^2} \frac{[h(x,y) - f(x)\,g(y)]^2}{f(x)\,g(y)}\,dx\,dy$$

où on reconnaît la généralisation en continue du $\frac{\chi^2}{n}$ des tables de contingences.

ϕ^2 prend ses valeurs entre 0 et ∞, et à la différence des autres mesures de dépendance n'est égal à zéro que si et seulement X et Y sont indépendantes. Lorsque Y dépend fonctionnellement de X (ou vice-versa) ϕ^2 est infini.

Pour obtenir un indicateur compris entre 0 et 1 à partir de ϕ^2 on considère généralement la quantité $\frac{\phi^2}{1+\phi^2}$ nommée "coefficient de contingence" qui fut introduite par K. Pearson dans ses études sur la loi normale à deux dimensions : ainsi que nous le verrons plus loin, pour une loi binormale, on a

$$\frac{\phi^2}{1+\phi^2} = \rho^2$$

On voit, sans difficulté, que ϕ^2 est invariant pour toute transformation continument différentiable $X \to \varphi(X)$ $Y \to \psi(Y)$ et est donc une mesure intrinsèque de la dépendance entre deux lois de probabilité : en effet posons $U = \varphi(X) V = \psi(Y)$ alors la densité $l(u, v)$ du couple aléatoire (U, V) vaut $\frac{h(x,y)}{\varphi'(x).\varphi'(y)}$ tandis que les densités marginales $i(u)$ et $j(v)$ valent respectivement

$$\frac{f(x)}{\varphi'(x)} \qquad \frac{g(y)}{\psi'(y)} \qquad \text{et} \qquad du = \varphi'(x)\,dx \quad dv = \psi'(y)\,dy$$

donc

$$\frac{[l(u,v)]^2}{i(u)\,j(v)}\,du\,dx = \left[\frac{h(x,y)}{\varphi'(x)\,\psi'(y)}\right]^2 \frac{\varphi'(x)\,\psi'(y)}{f(x)\;\;g(y)}\,\varphi'(x)dx\;\psi'(y)dy = \frac{[h(x,y)]^2}{f(x)\,g(y)}\,dx\,dy$$

Pour nous résumer, ϕ^2 constitue donc la mesure de dépendance entre phénomènes aléatoires la plus complète car elle ne dépend pas d'une représentation numérique particulière de ces phénomènes.

d) Géométrie de la dépendance entre variables aléatoires.

Considérons l'espace vectoriel L^2 des variables aléatoires[1] de deuxième moment fini $E(X^2) < \infty$ muni du produit scalaire $< X, Y) = E[X\,Y]$ et de la norme $||X||^2 = E(X^2)$.

---

(1) En toute rigueur dès classes de variables aléatoires presque partout égales pour la mesure de Lebesgue.

L'ensemble des variables aléatoires constantes est une droite D de cet espace passant par l'origine (puisque si $a \in D \lambda a \in D$) et les variables aléatoires centrées déterminent un sous-espace de L^2 orthogonal à D.

Dans ce sous-espace des variables centrées, le produit scalaire $E(XY)$ n'est autre que la covariance et les coefficient de corrélation linéaire ρ est le cosinus de l'angle formé par deux variables aléatoires centrées. Désignons par L^2_X et L^2_Y les sous-espaces de L^2 formés respectivement des variables centrées fonctions de X seul et de Y seul.

On peut établir alors les résultats suivants :

* $\eta^2_{Y/X}$ est le cosinus de l'angle entre Y et L^2_X

** La projection orthogonale de Y sur L^2_X est la variable aléatoire $E(Y/X)$

*** $L^2_X \cap L^2_Y = \{0\} \Longleftrightarrow X \notin L^2_Y$ et $Y \notin L^2_X$: non dépendance fonctionnelle.

**** X et Y ne sont indépendants que si et seulement si les projections de L^2_X sur L^2_Y et de L^2_Y sur L^2_X sont nulles.

Autrement dit L^2_Y est orthogonal à L^2_X et

$$\forall \varphi \; ; \; \forall \psi \quad \rho(\varphi(X) \; ; \varphi(Y)) = 0$$

***** soient π_x et π_y les opérateurs linéaires de projection sur L^2_X et L^2_Y : alors $\phi^2 = \text{Trace}(\pi_x \circ \pi_y) = \text{Trace}(\pi_y \circ \pi_x)$.

Les trois premiers points ne font que traduire dans un autre langage de propriétés statistiques bien connues, le quatrième point se démontre à partir du cinquième en effet $\phi^2 = 0$ (dont on sait qu'il est équivalent à l'indépendance) est donc équivalent à l'orthoganalité de L^2_X et L^2_Y. Le cinquième point sera démontré ultérieurement.

La double inégalité $\boxed{\phi^2 \geqslant \eta^2 \geqslant \rho^2}$ se déduit des remarques précédentes.

2/ Recherche d'une transformation simultanée maximisant la liaison de ces variables.

La problématique exposée dans l'introduction de cet article nous conduit à chercher, pour deux phénomènes aléatoires en dépendance stochastique, une représentation numérique de chacun d'eux, dite codage ou reparamétrisation, qui respecte au mieux leur dépendance, c'est-à-dire telle que leur liaison soit maximale. La non unicité des indicateurs de liaison (faut il choisir ρ ou η^2 ?) pourrait laisser croire à des solutions différentes selon le critère adopté ; on montre en fait qu'il n'en est rien et que ce problème possède une solution unique indépendante du critère retenu ; ces résultats s'étendent en cherchant une suite de transformations simultanées fournissant ce que l'on appelle les variables canoniques.

a) Unicité et propriétés de la transformation optimale

La maximisation de $\rho(\varphi(X), \psi(Y)$ peut sembler a priori plus restrictive que celle de $\eta^2_{\psi(Y)/\varphi(X)}$ ou $\eta^2_{\varphi(X)/\psi(Y)}$ car ρ ne mesure que la

linéarité de la dépendance. Le paragraphe précédent montre cependant l'équivalence entre ces deux critères puisque $\eta^2_{Y/X} = \sup_f \rho^2(Y, f(X))$ et alors maximiser $\rho^2(\varphi(X), \psi(Y))$; géométriquement, ce résultat est évident puisqu'il s'agit de déterminer les deux éléments $\varphi(X)$ et $\psi(Y)$ de L^2_X et L^2_Y qui font entre eux l'angle minimal (φ et ψ sont alors défini à une constante multiplicative près ce qui entraînera par la suite une condition de normalisation si on veut une solution unique). Toutefois les deux critères sont méthodologiquement différents car la maximisation de ρ entraîne une condition simultanée sur φ et ψ tandis que la maximisation de $\eta^2_{\psi(Y)/\varphi(X)}$ implique que la détermination de ψ n'est pas liée à celle de φ, remarque sur laquelle nous reviendrons lorsque l'un des codages est imposé.

Théorème :

Les transformations φ et ψ sont telles que les deux nouvelles variables $\varphi(X)$ et $\psi(Y)$ sont en double régression linéaire

$$E[\varphi(X)/\psi(Y)] = \lambda\, \varphi(Y)$$

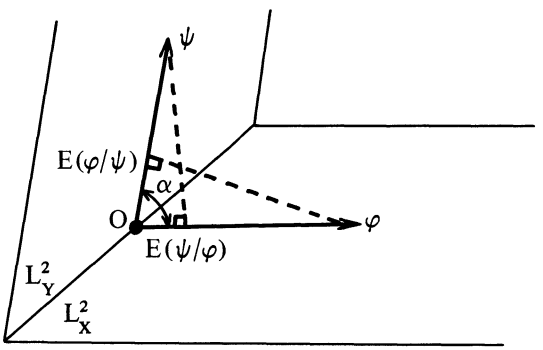
$$E[\psi(Y)/\varphi(X)] = \mu\, \varphi(X)$$

De plus

$$\lambda = \mu = \rho^2(\varphi(X) ; \psi(Y)) = \eta^2_{\psi(Y)/\varphi(X)} = \eta^2_{\varphi(X)/\psi(Y)}$$

si φ et ψ sont normalisées par $V[\varphi(X)] = V[\psi(Y)] = 1$

En effet si $\varphi(X)$ et $\varphi(Y)$ forment un angle minimal alors $\varphi(X)$ doit se projeter sur L^2_Y selon $\psi(Y)$ et réciproquement :



N.B. : Cette représentation imagée est abusive car L^2_X et L^2_Y n'ont en général que le point $\{0\}$ en commun.

La projection de $\varphi(Y)$ sur L^2_X est $E[\psi(Y)/X]$ donc

$$E[\psi(Y)/\varphi(X)] = \lambda\varphi(X) \quad \text{et} \quad E[\varphi(X)/\varphi(Y)] = \mu\,\varphi(Y)$$

Si $\varphi(X)$ et $\psi(Y)$ sont de variance unité il vient :

$$V[E[\psi(Y)/\varphi(X)]] = \lambda^2 = \eta^2_{\psi/\varphi}$$

d'où le résultat annoncé plus haut. On a alors :

$$\rho^2 = \eta^2 = \cos^2\alpha$$

On pouvait aussi remarquer que la minimisation de $\cos^2\alpha$ revient à celle de $E[[\varphi(X) - \varphi(Y)]^2]$ qui se résout si, à un facteur près, $\varphi(X)$ est la meilleure approximation fonctionnelle de $\psi(Y)$ et vice-versa.

Le théorème précédent constitue une condition nécessaire mais non suffisante d'optimalité : deux variables en double régression linéaire ne forment pas nécessairement le couple le plus lié, ainsi que le montre le contre exemple suivant, qui fournit au passage un cas de deux variables non gaussiennes en double régression linéaire :

Soit un couple (X, Y) de variables centrées réduites suivant une loi normale à deux dimensions avec un coefficient de corrélation ρ. Alors X^2 et Y^2 forment un couple en double régression linéaire : en effet :

$$E(X^2/Y^2) = \frac{1}{2} E[X^2/Y] + \frac{1}{2} E[X^2/-Y] = E(X^2/Y)$$

car la distribution de Y est symétrique par rapport à zéro. D'après une formule élémentaire $E(X^2/Y) = V(X/Y) + [E(X/Y)]^2$ et dans un couple normal centré réduit :

$$V(X/Y) = 1 - \rho^2 \quad \text{tandis que} \quad E(X/Y) = \rho Y \quad \text{donc}$$

$$E(X^2/Y^2) = 1 - \rho^2 + \rho^2 Y^2 = 1 + \rho^2(Y^2 - 1)$$

et

$$E(Y^2/X^2) = 1 + \rho^2(X^2 - 1)$$

Le coefficient de corrélation linéaire entre X^2 et Y^2 est égal à ρ^2 en identifiant les coefficients de la régression ($E(\psi/\varphi) = a\varphi + b$ entraîne que $a = \rho \frac{\sigma_\psi}{\sigma_\varphi}$). Cette corrélation est inférieure en valeur absolue à celle de (X, Y) ; le couple (X^2, Y^2) bien qu'en double régression linéaire n'est donc pas de corrélation maximale.

En revenant sur l'interprétation géométrique du théorème précédent on en déduit la propriété suivante :

π_x et π_y étant les opérateurs de projection sur L^2_X et L^2_Y respectivement, alors φ (resp ψ) est fonction propre de $\pi_x \circ \pi_y$ associé à sa plus grande valeur propre :

$$\pi_x \circ \pi_y(\varphi) = \lambda \varphi$$

L'expression des projecteurs π_x et π_y s'obtient en écrivant que π_x transforme toute variable aléatoire en son espérance conditionnée par X : appelons $k(x, u)$ la densité conjointe de X et U on a donc :

$$\pi_x(u) = \int_{-\infty}^{+\infty} u \frac{k(X,u)}{f(X)}\, du$$

$$\pi_y(u) = \int_{-\infty}^{+\infty} u \frac{l(u,Y)}{g(Y)}\, du$$

Si on considère les restrictions de π_x et π_y à L_Y^2 et L_X^2 respectivement :

$$\pi_x(\psi(y)) = \int_{-\infty}^{+\infty} \psi(y) \frac{h(X,y)}{f(x)}\, dy \quad \pi_y(\varphi(x)) = \int_{-\infty}^{+\infty} \varphi(x) \frac{h(x,Y)}{g(Y)}\, dx$$

d'où

$$\pi_x \circ \pi_y(\varphi(x)) = \int_{-\infty}^{+\infty} \Lambda(u) \frac{h(X,u)}{f(x)}\, du = \iint_{R^2} \varphi(v) \frac{h(v,u)\, h(X,u)}{f(x)\, g(u)}\, du\, dv$$

en posant :

$$\Lambda(Y) = \pi_y(\varphi(x))$$

et

$$\pi_y \circ \pi_x(\psi(y)) = \iint_{R^2} \psi(u) \frac{h(v,u)\, h(v,y)}{f(v)\, g(y)}\, du\, dv$$

La détermination de φ et ψ est celle des fonctions propres d'un opérateur intégral[1].

b) Variables canoniques d'une distribution à deux dimensions.

On peut généraliser les opérations précédentes en considérant que φ et ψ ne sont que les premiers éléments φ_1, ψ_1 d'une suite de transformation (φ_1, φ_1) ; (φ_2, φ_2) ... définies de la manière suivante où l'on notera désormais en abrégé φ pour $\varphi(x)$:

Les φ_i sont centrées-réduites et orthogonales à leur prédécesseurs dans chaque espace L_X^2 et L_Y^2 .

$$E(\varphi_i \varphi_j) = E(\psi_i \psi_j) = \delta_{ij}$$

(il s'agit donc de deux systèmes orthonormés de L_X^2 et L_Y^2 respectivement) et telles que pour chaque rang i la corrélation entre φ_i et ψ_i soit maximale $\rho_i = E(\varphi_i\, \psi_i)$

Les variables (φ_i, ψ_i) ainsi définies de proche en proche (à une convention de signe près) sont appelées variables canoniques et les ρ_i sont les corrélations canoniques. Comme on peut définir φ_0 et ψ_0 en les prenant égales à 1, il s'ensuit que les φ_i et ψ_i, orthogonales aux constantes, sont centrées de fait.

Propriété :

$$E[\varphi_i \psi_j] = 0 \quad \text{si} \quad i \neq j$$

---

(1) Si $\phi^2 < \infty$ $\pi_y \circ \pi_x$ et $\pi_x \circ \pi_y$ sont des opérateurs nucléaires (trace finie)

Les variables canoniques vérifient donc un autre système de relations d'orthogonalité.

Démonstration :

Supposons $j > i$. Par Hypothèse $\rho_i = E[\varphi_i \psi_i]$ était la plus grande corrélation possible au rang i parmi les variables orthogonales aux précédentes. Supposons $E[\varphi_i \psi_j] \neq 0 = \rho_i \operatorname{tg}\theta$; soit $Z = \psi_i \cos\theta + \psi_j \sin\theta$. La variable Z est orthogonale aux ψ_k pour $k < i$ car ψ_i et ψ_j le sont, de plus $E[Z^2] = 1$

$$\begin{aligned} E[Z\varphi_i] &= \cos\theta\, E[\varphi_i\psi_i] + \sin\theta\, E[\varphi_i\psi_j] \\ &= \cos\theta\ \rho_i + \sin\theta\ \operatorname{tg}\theta\ \rho_i \\ &= \frac{\rho_i}{\cos^2\theta} \end{aligned}$$

Si tg θ est non nul on a alors $E[Z\varphi_i] > \rho_i$ ce qui est en contradiction avec l'hypothèse que ρ_i est la $i^{ième}$ corrélation canonique.

Réciproque :

Si les φ_i et φ_j forment deux systèmes orthonormaux *complets* sur L^2_X et L^2_Y et sont tels que $E[\varphi_i \psi_j] = 0$ si $i \neq j$ alors ce sont les variables canoniques.

Démonstration :

Si les systèmes sont complets toute variable X' centrée fonction de X peut s'écrire de manière unique $X' = \sum a_i \varphi_i(x)$ et toute variable centrée Y' fonction de Y : $Y' = \sum_{i=1}^{\infty} b_i\ \psi_i(Y)$. Si ces variables sont réduites, c'est que:

$$\sum_{i=1}^{\infty} a_i^2 = \sum_{i=1}^{\infty} b_i^2 = 1$$

puisque les φ_i et ψ_i sont aussi réduites.

On a alors

$$\rho(X', Y') = E(X'Y') = \sum_{i=1}^{\infty} a_i b_i \rho_i$$

en posant $E[\varphi_i \psi_i] = \rho_i$. Si on suppose que la numérotation des φ_i et des ψ_i est telle que

$\rho_i \geqslant \rho_2 \geqslant \ldots \geqslant \rho_k \ldots \rho(X', Y')$ sera maximal si $a_1 = b_1 = 1$,

les autres coefficients étant nuls : en effet puisque

$$x^2 + y^2 \geqslant 2xy \quad \text{on a} \quad \sum (a_i^2 + b_i^2)\, \rho_i \geqslant 2 \sum a_i b_i \rho_i$$

donc

$$a_i b_i \rho_i \leqslant \frac{1}{2} \sum (a_i^2 + b_i^2)\rho_i \leqslant \frac{1}{2} \sum (a_i^2 + b_i^2)\rho_1 = \rho_1$$

l'égalité ne peut avoir lieu que si $a_1 = b_1 = 1$. Le couple (φ_1, ψ_1) est le premier couple de variables canoniques, on démontrerait de même que les couples (φ_2, ψ_2), (φ_3, ψ_3), sont les couples canoniques suivants.

Les variables canoniques sont les fonctions propre de $\pi_y \circ \pi_x$ et $\pi_x \circ \pi_y$ avec pour valeurs propres ρ_i^2 car, les variables φ_i et ψ_i étant à l'ordre i les variables les plus proches, on a

$$E[\varphi_i / \psi_i] = \rho_i \psi_i \quad \text{et} \quad E[\psi_i / \varphi_i] = \rho_i \varphi_i$$

Les variables canoniques permettent de reconstituer la densité du couple à partir des densités marginales si $\sum \rho_i^2 < \infty$ selon la formule suivante :

Formule de reconstitution de la densité conjointe :

$$h(x, y) = \left[1 + \sum_{i=1}^{\infty} \rho_i \varphi_i(x) \psi_i(y) \right] f(x)\, g(y)$$

On peut dire que les variables canoniques sont celles qui expliquent le mieux en quoi h (x, y) diffère du produit des densités marginales.

Démonstration

Supposons les ensembles orthonormaux $\{\varphi_i\}$ et $\{\psi_i\}$ *complets* pour L_X^2 et L_Y^2, alors on sait que l'ensemble $\{\varphi_i \cdot \psi_j \ \forall_{i,j}\}$ est complet pour l'ensemble des fonctions des deux variables, de carré vitégrable, muni de la mesure produit f (x) g (y) cherchons à rendre minimum l'intégrale

$$\iint_{R^2} \left[\frac{h(x, y)}{f(x)\, g(y)} - \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \lambda_{ij} \varphi_i(x) \psi_j(y) \right]^2 f(x)\, g(y)\, dx\, dy$$

en posant conventionnellement $\varphi_0 = \varphi_0 = 1$

En dérivant par rapport aux λ_{kl} il vient :

$$\int\int \left[\frac{h(x, y)}{f(x)\, g(y)} - \sum \sum x_{ij} \varphi_i(x) \psi_j(y) \right] \varphi_k(x) \psi_l(y) f(x)\, g(y)\, dx\, dy = 0$$

soit

$$\iint_{R^2} \varphi_k(x) \psi_l(y) h(x, y)\, dx\, dy =$$

$$\sum_0^{\infty} \sum_0^{\infty} \lambda_{ij} \iint_{R^2} \varphi_i(x) \psi_j(y) \varphi_k(x) \psi_l(y) f(x)\, g(y)\, dx\, dy$$

$$E[\varphi_k \psi_l] = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \lambda_{ij} \left[\int_{-\infty}^{+\infty} \varphi_i(x)\, \varphi_k(x)\, f(x)\, dx \right] \left[\iint_{-\infty}^{+\infty} \psi_j(y)\, \psi_l(y)\, g(y)\, dy \right]$$

D'après les conditions d'orthonormalisation il vient alors

$E[\varphi_k \psi_l] = \lambda_{kl}$ et $\lambda_{ij} = 0$ si $i \neq j$ soit $\lambda_{ii} = \rho_i$

avec $\lambda_{00} = \lambda_0 = 1$

et il s'ensuit que pour tout entier m

$$\iint \left[\frac{h(x,y)}{f(x)\, g(y)} - \sum_{i=0}^{m} \lambda_i \varphi_i(x)\, \psi_i(y) \right]^2 f(x)\, g(y)\, dx\, dy$$ est

minimal pour $\rho_i = \lambda_i$ ce qui entraîne le théorème par passage à la limite puisque l'ensemble $\{\varphi_i \times \psi_j\}$ est complet pour la mesure produit $f(x)\, g(y)$

Corollaire 1

ϕ^2 est égal à la somme des carrés des corrélations canoniques qui vaut Trace $(\pi_x \circ \pi_y)$ ou Trace $(\pi_y \circ \pi_x)$

en effet :

$$1 + \phi^2 = \iint \frac{[h(x,y)]^2}{f(x)\, g(y)}\, dx\, dy = \iint \left[1 + \sum_{i=r}^{\infty} \rho_i \varphi_i \psi_i \right] h(x,y)\, dx\, dy$$

$$= E\left[1 + \sum_{i=1}^{\infty} \rho_i \varphi_i \varphi_i \right] = 1 + \sum_{i=1}^{\infty} \rho_i\, E[\varphi_i \psi_i]$$

$$= 1 + \sum_{i=1}^{\infty} \rho_i^2$$

Corollaire 2

Deux variables sont indépendantes si et seulement si leur première corrélation canonique est nulle.

Cette propriété, évidente puisque

$$\phi^2 = 0 \Longrightarrow \rho_i = 0\ \forall_i \quad \text{et que} \quad \rho_1^2 \geqslant \rho_2^2 \geqslant \ldots,$$

pourrait suggérer de prendre ρ_1 comme indicateur de liaison entre X et Y puisqu'il aurait la propriété d'être compris entre 0 et 1, nul si et seulement si il y a une relation fonctionnelle $Y = \varphi(X)$ ou $X = \psi(Y)$; cela, à notre connaissance, n'a jamais été tenté, sans doute en raison des difficultés de résolution des équations intégrales fournissant les φ_i et les ψ_i (pour la résolution de ces équations, consulter la thèse de Naouri [3]).

On peut aussi démontrer les formules de dualité suivantes [1] :

$$\rho_i \varphi_i(x) \sqrt{f(x)} = \int_{-\infty}^{+\infty} \frac{h(x,y)}{\sqrt{f(x)\, g(y)}} \psi_i(y) \sqrt{g(y)}\, dy$$

$$\rho_i \psi_i(x) \sqrt{g(y)} = \int_{-\infty}^{+\infty} \frac{h(x,y)}{\sqrt{f(x)\, g(y)}} \varphi_i(x) \sqrt{f(x)}\, dx$$

On remarquera l'analogie entre les formules ci-dessus et les formules de l'analyse des correspondances.

c) Application à la loi normale à deux dimensions (d'après [2])

Soit (X, Y) un couple de variables centrées réduites suivant une loi normale à deux dimensions de coefficient de corrélation ρ.

Nous allons montrer que les variables canoniques ne sont autres que les polynomes de Hermite-Chebyshev réduits,définis sur X et Y par les conditions d'orthonormalisation suivantes :

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \varphi_i(x)\, \psi_j(x)\, e^{-\frac{1}{2}x^2}\, dx = 0 \quad \text{si} \quad i \neq j$$
$$= 1 \quad \text{si} \quad i \neq j$$

on a $$\psi_j(y) = \varphi_j(y)$$

Les polynomes de Hermite-Chebyshev non réduits $H_n(x)$ sont obtenus de la manière classique suivante :

$$e^{-\frac{1}{2}x^2} H_n(x) = (-1)^n \frac{d^n [e^{-\frac{1}{2}x^2}]}{d x^n} \quad \text{et vérifient}$$

les conditions

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} H_i(x)\, H_j(x)\, e^{-\frac{1}{2}x^2}\, dx = 0 \quad \text{si} \quad i \neq j$$
$$= i! \quad \text{si} \quad i = j$$

autrement dit : $V[H_i(x)] = i!$

On passe donc des $H_n(x)$ aux $\varphi_n(x)$ par la relation $\varphi_n(x) = \frac{H_n(x)}{\sqrt{n!}}$. Les φ_n se calculent de proche en proche, les premiers éléments étant

$$\varphi_0(x) = 1 \qquad \varphi_1(x) = x$$

$$\varphi_2(x) = \frac{x^2 - 1}{\sqrt{2}} \qquad \varphi_3(x) = \frac{x^3 - 3x}{\sqrt{6}}$$

$$\varphi_4(x) = \frac{x^4 + 6x^2 + 3}{\sqrt{24}} \qquad \varphi_5(x) = \frac{x^5 - 10\, x^3 + 15\, x}{\sqrt{120}}$$

Pour démontrer que les φ_i constituent les variables canoniques nous allons appliquer le réciproque de la propriété établie au paragraphe b) en prouvant que

$$E[\varphi_i \psi_j] = 0 \quad \text{si} \quad i \neq j$$

Théorème 1.

Les polynomes de Hermite-Chebyshev réduits sont tels que

$$E[\varphi_i \psi_j] = 0 \quad \text{si} \quad i \neq j \quad \text{et} \quad E[\varphi_i \psi_i] = \rho^1$$

les corrélations canoniques sont les puissances successives de ρ

Démonstration :

α) Les polynomes non réduits sont les coefficients de $\frac{t^n}{n!}$ dans de développement de Taylor de $e^{tx - \frac{1}{2}t^2}$: en effet posons

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{on a} \quad f(x - t) = e^{tx - \frac{t^2}{2}} f(x)$$

d'une part, et d'autre part

$$f(x - t) = \sum_{n=0}^{\infty} (-i)^n \frac{t^n}{n!} \frac{d^n f(x)}{dx^n}$$

d'après la définition de $H_n(x)$ on a :

$$H_n(x)\, f(x) = (-1)^n \frac{d^n f(x)}{dn^x}$$

β) L'espérance de $e^{tX - \frac{t^2}{2} + uY - \frac{u^2}{2}}$ vaut $e^{\rho tu}$ car la fonction génératrice du couple (X, Y) qui est $E[e^{tX + uY}]$ vaut $e^{-\frac{1}{2}(u^2 + 2\rho tu + t^2)}$

γ) Rapprochons les deux résultats précédents :

$$E[e^{tx - \frac{t^2}{2} + uy - \frac{u^2}{2}}] = E\left[\left(\sum_{m=0}^{\infty} \frac{t^m}{m!} H_m(x)\right)\left(\sum_{n=0}^{\infty} \frac{u^n}{n!} H_n(y)\right)\right] \text{ d'après } \alpha$$

$$= \sum_{i=0}^{\infty} \frac{(tu)^i}{i!} \rho^i \quad \text{d'après } \beta$$

En développant l'espérance et en identifiant terme à terme il s'ensuit que tous les termes où m est différent de n sont nuls :

$$E[H_m(x)\, H_n(y)] = 0 \quad \text{si} \quad m \neq n, \qquad \text{les autres donnant}$$

$$E[H_i(x)\, H_i(y)] = i!\, \rho^i \quad \text{ce qui démontre le théorème}$$

car

$$\varphi_i(x) = \frac{H_i(x)}{\sqrt{i!}}$$

On a alors :

Théorème 2 (Lancaster [2])

Pour un couple de variables centrées réduites suivant une loi normale à deux dimensions, il n'existe pas de transformation séparée de ces variables pouvant augmenter le coefficient de corrélation ρ. En effet $\varphi_1(X) = X$ et les polynomes de Hermite-Chebyshev forment un système complet orthonormé, les φ_i sont bien les variables canoniques.

Puisque les corrélations canonique ρ_i sont les puissances successives de ρ :

$$\rho_i^2 = \rho^{2i} \qquad \text{et la formule} \qquad \phi^2 = \sum_{i=1}^{\infty} \rho_i^2 \qquad \text{devient}$$

$$\phi^2 = \sum_{i=i}^{\infty} \rho^{2i} = \frac{\rho^2}{1 - \rho^2}$$

ce qui démontre la relation de Karl Pearson $\dfrac{\phi^2}{1 + \phi^2} = \rho^2$.

La formule de reconstitution de la densité du couple s'écrit alors :

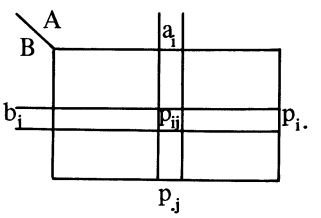
$$\frac{1}{2\pi\sqrt{1 - \rho^2}}\, e^{-\frac{1}{2}\left(\frac{x^2 + y^2 - 2\rho xy}{1 - \rho^2}\right)} = \left[1 + \sum_{i=1}^{\infty} \varphi_i(x)\, \psi_i(y)\, \rho^i\right] \frac{1}{2\pi}\, e^{-\frac{1}{2}(x^2 + y^2)}$$

expression connue sous le nom d'identité de Mehler (1866).

II – DISTRIBUTIONS DEFINIES PAR UNE TABLE DE CONTINGENCE A DEUX DIMENSIONS

Il s'agira tout aussi bien de la distribution de probabilité exacte p_{ij} d'un couple de v.a. à nombre fini de valeurs que d'un tableau de fréquence f_{ij} d'un échantillon de taille n (ce qui est le cas le plus usuel).

Comme l'annonce l'introduction il s'agit d'affecter à chaque ligne et chaque colonne du tableau une valeur numérique (les codages), indépendemment de toute information de nature quantitative sur les distributions marginales, en fonction seulement du tableau des p_{ij}



Nous cherchons donc deux systèmes de nombres réels $(a_1, a_2, \ldots a_m)$ et $(b_1, b_2, \ldots b_p)$ tels que les variables A et B rendues ainsi quantitatives soient en corrélation maximale.

On peut évidemment appliquer les résultats de I, le cas des tables de contingence n'étant qu'un cas particulier du cas général, les densités devenant des lois discrètes et les intégrales des sommes finies : il suffit d'identifier les opérateurs linéaires $\pi_x \circ \pi_y$ et $\pi_y \circ \pi_x$ qui se représentent alors par des matrices sur une certaine base et de les diagonaliser. Cependant, vu l'importance pratique du cas des tables de contingence, nous préferons donner une solution détaillée qui nous semble plus concrète.

1/ Codage optimal des marges d'un tableau de contingence.

En abusant quelque peu des notations, nous désignerons par A et B les deux variables en dépendance ainsi que les quantifications cherchées de ces variables.

Les codages devant être centrés réduits on a :

$$E(A) = E(B) = \sum_i p_{i.}\, a_i = \sum_j p_{.j}\, b_j = 0$$

$$V(A) = V(B) = \sum_i p_{i.} a_i^2 = \sum_\gamma p_{.j} b_j^2 = 1$$

Ainsi qu'en I 2a nous allons maximiser $\eta^2_{A/B}$ et $\eta^2_{B/A}$. On a pour des variables centrées réduites :

$$\eta^2_{A/B} = \sum_j \frac{1}{p_{.j}} \left(\sum_i p_{ij}\, a_i\right)^2 \qquad \eta^2_{B/A} = \sum_i \frac{1}{p_{i.}} \left(\sum_j p_{ij} b_j\right)^2$$

car

$$E(A \mid b_j) = \sum_i \frac{p_{ij}}{p_{.j}} a_i \qquad \text{et} \qquad \eta^2_{A/B} = V[E(A \mid B)] = \sum_j p_{.j} (E(A \mid b_j))^2$$

Nous nous occuperons de maximiser $\eta^2_{A/B}$, car la maximisation de $\eta^2_{A/B}$ s'en déduira aisément.

Le programme suivant est à résoudre :

$$\begin{cases} \max \sum_j \frac{1}{p_{.j}} \left(\sum_i p_{ij}\, a_i\right)^2 \\ \sum_i p_{i.}\, a_i = 0 \\ \sum_i p_{i.}\, a_i^2 = 1 \end{cases}$$

Introduisons deux multiplicateurs de Lagrange λ et μ correspondant aux deux contraintes et le programme se transforme en :

$$\max \left\{ \sum_j \frac{1}{p_{.j}} \left(\sum_i p_{ij}\, a_i \right)^2 - \lambda \sum_i p_{i.}\, a_i - \mu \sum_i p_{i.}\, a^2 \right\}$$

Le maximum s'botient en annulant les dérivés partielles par rapport aux a_k soit :

$$2 \sum_j \frac{1}{p_{.j}} \left(\sum_i p_{ij}\, a_i \right) p_{kj} - \lambda\, p_{k.} - 2\mu\, p_{k.}\, a_k = 0$$

ou

$$\sum_i \sum_j \frac{p_{ij}\, p_{kj}}{p_{.j}}\, a_i = \frac{\lambda}{2}\, p_{k.} + \mu\, p_{k.}\, a_k$$

En sommant membre à membre on obtient :

$$\sum_k \sum_i \sum_j \frac{p_{ij}\, p_{kj}}{p_{.j}}\, a_i = \frac{\lambda}{2} \qquad \text{car} \qquad \sum_k p_{k.}\, a_k = 0$$

si on écrit les sommations dans l'ordre i , j , k le premier membre s'écrit

$$\sum_i \sum_j p_{ij}\, a_i = \sum p_{i.}\, a_i = 0$$

On a donc $\lambda = 0$ ce qui signifie que la contrainte de centrage est automatiquement remplie, propriété signalée antérieurement pour les variables caoniques en I 2b.

Pour obtenir μ multiplions par a_k et sommons les deux membres de l'égalité restante :

$$\sum_i \sum_j \frac{p_{ij}\, p_{kj}}{p_{.j}}\, a_i = \mu\, p_{k.}\, a_k$$

soit :

$$\sum_k \sum_i \sum_j \frac{p_{ij}\, p_{kj}}{p_{.j}}\, a_i\, a_k = \mu \sum_k p_{k.}\, a_k^2 = \mu$$

dont le premier membre se met sous la forme

$$\sum_j \frac{1}{p_{.j}} \left[\sum_k \sum_i p_{ij}\, p_{kj}\, a_i\, a_k \right] = \sum_j \frac{1}{p_{.j}} \left(\sum_i p_{ij}\, a_i \right)^2 = \eta^2_{A/B}$$

donc $\mu = \eta^2_{A/B}$ Les a_k sont solutions du système d'équations

$$\sum_i \sum_j \frac{p_{ij}\, p_{kj}}{p_{.j}}\, a_i = \eta^2\, p_{k.}\, a_k$$

si on pose

$$\alpha_i = \sqrt{p_{i.}}\ a_i$$

il vient

$$\sum_i \sum_j \frac{p_{ij}\ p_{kj}}{p_{.j}\sqrt{p_{i.}\ p_{k.}}} = \alpha_i = \eta^2 \alpha_k$$

qui s'écrit matriciellement :

$$R\ R'\alpha = \eta^2 \alpha$$

ou $\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix}$ et R est la matrice du terme général $r_{ij} = \frac{p_{ij}}{\sqrt{p_{i.}\ p_{.j}}}$

α est donc vecteur propre de $R\ R'$ associé à sa plus grande valeur propre[1]. De même, en posant $\beta_j = \sqrt{p_{.j}}\ b_j$ on trouverait que β est vecteur propre de $R'\ R$ associé à la même valeur propre : on reconnaît ici les équations fournissant les facteurs de l'analyse des correspondances du tableau p_{ij}.

Le codage optimal de A est donc fourni par le premier facteur de l'analyse des correspondances, celui de B par le facteur dual : les valeurs numériques à attribuer aux n modalités de A sont les coordonnées des m points représentatifs des lignes du tableau sur le premier axe factoriel dans l'analyse où les individus sont les lignes et les caractères les colonnes ; réciproquement les valeurs à attribuer aux p modalités de B sont les coordonnées des points représentatifs des colonnes sur le premier axe factoriel de l'autre analyse.

Les relations de double régression linéaire $E(A/B) = \rho\ B$ et $E(B/A) = \rho\ A$ s'écrivent ici :

$$\sum_i \frac{p_{ij}}{p_{.j}} a_i = \sqrt{\lambda_1}\ b_j \quad \forall_j$$

$$\sum_j \frac{p_{ij}}{p_{i.}} b_j = \sqrt{\lambda_1}\ a_i \quad \forall_i$$

où λ_1 désigne la première valeur propre.

Ainsi que dans I 2b on peut chercher les couples de variables canoniques suivants $(A_2, B_2) \ldots (A_{m-1}, B_{m-1})$ si $m < p$ qui seront les couples de facteurs successifs normalisés de l'analyse des correspondances. La formule de reconstitution s'écrit alors :

$$p_{ij} = \left[1 + \sum_{k=1}^{m-1} \rho_k\ a_i^{(k)}\ b_j^{(k)} \right] p_{i.}\ p_{.j}$$

---

(1) Autre que la valeur propre triviale égale à 1 qui correspond à la variable canonique d'ordre 0 (constante égale à 1)

formule dont la première mention semble remonter à R.A. Fisher ainsi que l'atteste un article de Maung paru en 1942 et cité par Lancaster. Les résultats précédents traduisent le fait que l'analyse des correspondances peut s'interpréter comme une analyse canonique et constitue la méthode privilégiée d'analyse de la dépendance entre variables statistiques.

2/ Conséquences et applications

a) Recherche d'un seul codage

Bien que les deux codages optimaux A et B soient en relation, ils ont été obtenus séparément : le codage de A ne dépend en fait que du tableau des p_{ij} et non des valeurs que l'on peut attribuer aux modalités de B ; cette propriété est particulièrement intéressante lorsque le codage d'une des variables est soit imposé, soit sans intérêt.

α) On ne cherche pas à coder la variable B.

Le codage de A souhaitable est celui défini par le premier facteur de l'analyse des correspondances qui fournit la variable la plus proche de l'espace engendré par le découpage de B en p modalités.

Ainsi dans un problème de discrimination où B est le critère de classement ; il est sans intérêt de coder les modalités de B qui ne font qu'exprimer l'appartenance ou le non appartenance à une catégorie. Les méthodes classiques, analyse factorielle discriminante, métriques de Sebestyen, s'appliquent lorsque les variables explicatives sont toutes quantitatives et ne permettent pas l'introduction de prédicteurs qualitatifs à moins de quantifier ces prédicteurs, ce que l'on fait quelquefois à l'aide de codages arbitraires (nombres entiers consécutifs de 1 à m).

Les résultats précédents permettent alors de proposer un codage cohérent d'une ou de plusieurs variables qualitatives à introduire dans un modèle d'analyse discriminante : tout individu possédant la ième modalité d'une variable qualitative A se verra attribuer une valeur numérique égale à la projection du point représentatif de cette modalité sur le premier facteur de l'analyse des correspondances du tableau A × B. Ceci peut donner une solution au problème de "crédit scoring" par exemple : on cherche à discriminer entre bons et mauvais clients en fonction de variables telles que âge, revenu, catégorie sociale, sexe etc. . . et les praticiens de ce genre d'étude cherchent à affecter un certain nombre de points pour chaque modalité de chaque variable : selon la somme des points obtenus par un individu on décidera ou non de lui allouer un certain crédit. Le codage optimal permet de définir un système rationnel de "points" utilisable ensuite dans une fonction discriminante.

β) La variable B a un codage "naturel" qu'on ne veut pas changer.

C'est en particulier le cas des problèmes de régression multiple où on cherche à prendre en compte des prédicteurs qualitatifs, la variable à expliquer étant parfaitement définie.

La solution donnée par la première variable canonique s'oppose alors à la pratique usuelle qui est de coder les modalités de la variable explicative selon les moyennes conditionnelles de la variable expliquée :

Pour représenter la ième classe du préducteur qualificatif A on code par $\alpha_i = \sum_j \frac{p_{ij}}{p_{i.}} b_j$, le codage en α_i est en effet celui qui maximise le coefficient de corrélation linéaire ρ entre B et A codée par α et alors $\rho^2 = \eta^2_{B/A}$ tandis que le codage en a_i issu de la première variable canonique maximisé $\eta^2_{A/B}$ c'est-à-dire qu'il ne s'agit pas de la variable la plus proche de B mais la plus proche de toutes celles qu'engendre B.

Il peut alors sembler paradoxal de préférer le codage par la première variable canonique ou codage par les moyennes conditionnelles ; d'une part parce que la qualité de l'explication de B par A est inférieure en termes de variance expliquée, d'autre part parce que la maximisation de $\eta^2_{A/B}$ peut s'interpréter comme la recherche d'une explication de A par B alors qu'en régression c'est du contraire qu'il s'agit.

Ces deux objections sont cependant discutables : outre que la notion de causalité et le sens dans lequel elle s'exerce se révèlent souvent délicates à manier, le fait que le codage en a_i ne dépende pas de B peut devenir un avantage décisif si la quantification de B se révèle par la suite inadéquate (une échelle de mesure logarithmique par exemple peut se montrer plus appropriée qu'une échelle arithmétique). Le codage selon la première variable canonique (i.e premier facteur de l'analyse des correspondances) ne dépend que du tableau des p_{ij}, il est en quelque sorte intrinsèque à la dépendance entre les deux variables, tandis que le codage selon les moyennes conditionnelles fait pour ainsi dire la part trop belle à un certain modèle.

b) Analyse des données et variables normales.

Dans le cas où le tableau de contingence étudié peut être considéré comme représentatif d'un échantillon d'une loi normale bidimensionnelle on peut appliquer les résultats de I2C : la formule de Pearson $\phi^2 = \frac{\rho^2}{1 - \rho^2}$ conduit à une estimation de ρ^2 obtenue sans aucune information sur la quantification des marginales (on prend ici $\phi^2 = \sum_i \sum_j \frac{(p_{ij})^2}{p_{i.}\, p_{.j}} - 1$). Si les dimensions du tableau sont assez grande, de même que la taille de l'échantillon, les résultats de l'analyse de tableau de contingence doivent être assez proches de ceux obtenus pour la loi théorique : les variables canoniques (ou facteurs de l'analyse des correspondances) doivent être, aux fluctuations d'échantillonnage près, des polynomes de degré croissant en fonction de la première variable canonique : il s'agit alors d'un cas particulier de l'effet Guttman.

Commentant la propriété établie par Lancaster [2] (théorème 2 du I 2c) que dans une distribution normale à deux dimensions on ne peut améliorer le coefficient de corrélation ρ entre ces deux variables par une transformation quelconque, Kendall et Stuart [5] affirment :

"Si on cherche un codage séparé de deux variables maximisant leur corrélation, cela revient fondamentalement à fabriquer une distribution binormale en oérant sur les distributions marginales".

L'analyse des données ne serait elle donc qu'une tentative de normalisation ?

Cette propriété a en effet pour conséquence que s'il existe deux transformations $x \to \varphi(x)$ et $y \to \psi(y)$ telles que la distribution conjointe de (φ, ψ) soit binormale, alors (φ, ψ) constitue le premier couple des variables canoniques, les couples suivants étant des polynomes de degré croissant en φ et ψ dont les corrélations sont les puissances successives de la première.

L'analyse des correspondances devrait alors mener à une normalisation des données selon le premier facteur et à un effet Guttman systématique pour les autres facteurs, aux fluctuations d'échantillonnage près.

Rien ne dit cependant qu'une telle transformation soit toujours possible, on peut même fabriquer des distributions à deux dimensions telles que le premier couple de variables canoniques ne soit pas gaussien, mais la remarque de Kendall et Stuart méritait quelque attention et peut permettre à contrario, de rejeter l'hypothèse que certains phénomènes obéissent à une loi gaussienne sous-jacente.

c) Lien avec l'analyse canonique classique

Il est facile de montrer que les résultats précédents auraient pu être obtenu en faisant une analyse canonique opposant le groupe de n variables indicatrices des modalités de A à celui des p variables indicatrices des modalités de B [7].

Cette remarque permet de généraliser les procédures précédentes au cas de plusieurs variables et montre que les techniques de codage prennent place dans le cadre des analyse multivariées linéaires.

REFERENCES

[1] H.O. LANCASTER – "The Chi Square Distribution" John Wiley (1969) p. 85-100.

[2] H.O. LANCASTER – "Some properties of the bivariate normal distribution considered in the form of a contingency table". *Biometrika* 44 p. 289-292 (1957).

[3] J.C. NAOURI – "L'analyse factorielle des correspondances continues" thèse d'Etat Université de Paris VI (1971).

[4] J.P. BENZECRI – "L'analyse des données" Dunod Tome II p. 182-209 (1973).

[5] M.G. KENDALL et A. STUART – "The Advanced Theory of Statistics" Tome 2 Griffin p. 568-569 (1967).

[6] E.J. WILLIAMS – "Use of scores for the analysis of association in contingency tables". *Biometrika* 39 p. 274-289 (1952).

[7] C.E.E.E – Analyse des données multidimensionnelles . La Documentation Française (1971).

[8] A. POUSSE et J. DAUXOIS – "L'analyse canonique de deux tribus" Publication du Laboratoire de Statistique de l'Université Paul Sabatier - Toulouse (1974).

[9] M. MASSON – Thèse d'Etat Université de Paris VI (1974)