Revue de statistique appliquée

J.-B. Lagrange

Analyse implicative d'un ensemble de variables numériques; application au traitement d'un questionnaire à réponses modales ordonnées

Revue de statistique appliquée, tome 46, n° 1 (1998), p. 71-93

<http://www.numdam.org/item?id=RSA_1998__46_1_71_0>

Rev. Statistique Appliquée, 1998, **XLVI** (1), 71-93

ANALYSE IMPLICATIVE D'UN ENSEMBLE DE VARIABLES NUMÉRIQUES; APPLICATION AU TRAITEMENT D'UN QUESTIONNAIRE À RÉPONSES MODALES ORDONNÉES

J.-B. Lagrange

Laboratoire de Didactique – Institut de Recherche Mathématique de Rennes (France)

RÉSUMÉ

Cet article présente une extension de l'analyse implicative. La «propension» est définie comme association non-symétrique de deux variables à valeurs dans [0; 1]. L'«intensité de propension» est introduite afin de décider si une association observée est ou non significative. Dans le cas de variables binaires, l'intensité de propension est égale à l'intensité d'implication. Dans les autres cas, l'intensité de propension est plus fiable que l'intensité d'implication. Les «filiations» sont introduites comme organisations des propensions dans un ensemble de variables.

Une application de l'extension proposée au traitement des réponses modales ordonnées à un questionnaire d'opinion est présentée et discutée. Certains résultats du traitement des réponses à un questionnaire passé dans le cadre d'une recherche en didactique sont présentés en conclusion.

Mots-clés : *Analyse implicative, variables numériques, questionnaires d'opinion, propension, filiations.*

ABSTRACT

This paper introduces an extension of implicative analysis. We define "propensity" as a non-symmetric association of two variables whose values belong to [0; 1]. We introduce "intensity of propensity" to decide whether an observed association is statistically significant. If the variables are binary, the intensity of propensity equals the intensity of implication. Otherwise, the intensity of propensity is more reliable than the intensity of implication. "Families" of variables are introduced to organize the propensities in a set of variables.

We then consider questionnaires with a four-way Likert scale answer grid and discuss the use of propensity theory to analyze the answers. We apply this method to a questionnaire, part of ongoing didactic research. Typical didactic and statistical findings of this research are discussed in the conclusion.

Keywords : *Implicative analysis, numerical variables, questionnaires, propensity.*

1. Introduction : propension et implication

Nous nous proposons de dégager des associations orientées significatives dans un ensemble de variables numériques. Ces associations doivent rendre compte de la notion intuitive de «propension», au sens où, par exemple, l'âge d'un individu peut avoir une influence significative sur sa propension à l'achat de tel bien de consommation. Nous cherchons de plus une organisation des associations susceptible de dégager des systèmes d'explications.

L'analyse implicative (Gras, 1979; Lerman, Gras, Rostram, 1981) constitue le point de départ de cette étude. Dans ce traitement, l'intensité d'implication quantifie une tendance à l'inclusion d'une partie A dans une partie B. Si l'on considère des variables binaires a et b, l'intensité d'implication de la partie A dont a est la fonction indicatrice vers la partie B dont b est la fonction indicatrice rend compte de la «propension» de a vers b. La modélisation probabiliste qui conduit à l'intensité d'implication est donc valable pour déterminer des associations entre variables binaires. Gras, Lahrer, (1993) ont conjecturé que l'intensité d'implication pouvait rendre compte plus généralement de l'association de variables non binaires. Pour éprouver cette conjecture, nous nous proposons de construire une modélisation valable pour des variables à valeur dans [0; 1], puis de comparer les associations issues de cette modélisation, à celles qui résulteraient de l'intensité d'implication[1] .

En analyse implicative on représente l'ensemble des implications significatives dans un ensemble de variables sous forme d'un graphe. Bailleul (1994) a montré par exemple comment un tel graphe pouvait aider à interpréter les relations statistiques entre opinions d'enseignants de Mathématiques sur l'enseignement de cette discipline. Néanmoins, l'interprétation d'un graphe implicatif pose problème à cause de la non-transitivité de la relation : des variables sans liaison significative se trouvent ainsi reliées dans un même chemin implicatif. Nous cherchons donc à isoler des «filiations» partielles ne présentant pas ce défaut.

Nous examinons ensuite comment les éléments théoriques ainsi dégagés peuvent être appliqués au traitement des réponses à modalités ordonnées dans un questionnaire d'opinion. Nous comparons à d'autres choix de traitement et nous présentons les résultats didactiques issus du dépouillement d'un questionnaire passé par 460 élèves dans le cadre d'une recherche en didactique sur l'utilisation d'un logiciel de calcul formel pour l'apprentissage des mathématiques.

2. La quantification de la «propension»

Dans cette partie, nous exposons une modélisation de la relation de propension entre deux variables à valeurs dans l'intervalle [0; 1]. Cette modélisation est conforme aux principes de l'analyse des données telle qu'exposée dans (Lerman, 1986). On détermine un indice de non-propension d'une variable vers l'autre. Si la valeur observée de cet indice est trop faible pour être statistiquement compatible avec une hypothèse d'absence de lien *a priori* entre les variables, on conclut que la relation de propension est établie. Concrètement, la modélisation conduit à définir une intensité

[1] Dans la suite, le terme «propension» sera employé quand il s'agira de la relation entre variables à valeur dans [0; 1]. Le terme «implication» sera employé pour qualifier la relation entre variables binaires. L'intensité de propension désignera la mesure de la propension, telle que nous la définissons dans ce texte. L'intensité d'implication désignera la mesure de l'implication, telle qu'introduite par (Lerman, Gras, Rostram, 1981) pour les variables binaires, et utilisée par (Gras, Lahrer, 1993) pour d'autres types de variables.

de propension entre deux variables s'exprimant comme une valeur de la fonction de répartition de la loi normale centrée réduite. Choisissant un seuil de risque ε, l'analyste conclura à une relation de propension, si cette intensité est supérieure à 1-ε.

2.1. Un indice de non-propension

Considérons deux variables a et b, et les vecteurs de réels de l'intervalle $[0;1]$ (a_i) et (b_i) représentant les valeurs de ces variables observées chez les individus i d'une population P de taille n.

Il y a propension de a vers b si on rencontre en moyenne peu d'individus pour lesquels (a_i) est fort et (b_i) faible. Nous considérerons donc la moyenne sur P des $a_i(1-b_i)$ comme un indice de «non-propension» de a vers b. Dans cet indice, pour un individu donné, une valeur faible de b est d'autant plus prise en compte que la valeur de a pour le même individu est forte.

On conclut à la propension de a vers b dans P si cette moyenne est significativement faible. La moyenne observée est donc comparée à celle qui résulterait d'une hypothèse d'absence de lien. Cette comparaison nécessite une modélisation afin de déterminer la loi de probabilité d'un indice aléatoire dont la moyenne $\dfrac{\sum_{i\in P} a_i(1-b_i)}{n}$ serait une valeur observée.

2.2. Détermination de l'indice aléatoire

Considérons un ensemble E «grand» de taille k, et un ensemble de variables aléatoires $(A_i)_{i\in E}$ (resp. $(B_i)_{i\in E}$), de même loi, représentative de la variable a (resp. b). Il n'y a pas de lien *a priori* entre a et b sur E, et donc nous supposons les $2\times k$ variables A_i, B_i indépendantes. Considérons aussi une partie aléatoire $\mathcal{P}$ de E, représentative de la population observée P.

Soit n la taille de l'ensemble P. Nous supposons les k événements $\{i\in\mathcal{P}\}$ indépendants et de probabilité $\dfrac{n}{k}$.

Ainsi, $Z=\dfrac{\sum_{i\in\mathcal{P}} A_i(1-B_i)}{n}$ est un indice aléatoire dont l'indice moyen de non propension $\dfrac{\sum_{i\in P} a_i(1-b_i)}{n}$ est la valeur observée.

Nous testons l'indépendance entre les événements $\{i\in\mathcal{P}\}$ et les variables $A_i(1-B_i)$, au regard de la valeur observée de l'indice aléatoire. Un seuil de risque ε étant fixé, supposons qu'il existe un réel α tel que, avec cette hypothèse d'indépendance, $P(Z<\alpha)=\varepsilon$. Si la valeur observée $\dfrac{\sum_{i\in P} a_i(1-b_i)}{n}$ est inférieure à α, elle est significativement faible, comparée à la distribution de Z avec l'hypothèse d'indépendance. Il y a donc dans ce cas propension de a vers b dans l'ensemble P.

Note 1 : L'indépendance entre les événements $\{i \in \mathcal{P}\}$ et les variables $A_i(1-B_i)$ n'est pas équivalente à l'indépendance entre les événements $\{i \in \mathcal{P}\}$ et les variables $B_i(1-A_i)$. Ainsi la définition de la propension n'est pas symétrique en a et b.

Note 2 : Posons $T_i = 1_{i\in\mathcal{P}} A_i(1-B_i)$. On a alors $Z = \dfrac{\sum_{i\in E} T_i}{n}$. Les variables $A_i(1-B_i)$ sont indépendantes par suite de l'indépendance des A_i, B_i. Les $\{i \in \mathcal{P}\}$ sont indépendants entre eux et indépendants des $A_i(1-B_i)$. Donc les T_i sont indépendantes.

Dans le paragraphe suivant, nous déterminons la loi commune des T_i, puis la loi asymptotique de Z.

2.3. La loi asymptotique de l'indice aléatoire

Proposition 1

Si les variables $A_i(1-B_i)$ ont une fonction caractéristique $\varphi(t)$ alors l'indice aléatoire Z suit asymptotiquement la loi normale d'espérance $E[A_i(1-B_i)]$ et de variance $\dfrac{E[(A_i(1-B_i))^2]}{n}$.

Lemme : *Les variables $T_i = 1_{i\in\mathcal{P}} A_i(1-B_i)$ ont pour fonction caractéristique $1 + \frac{n}{k}(\varphi(t)-1)$.*

Preuve du lemme :

$$\begin{aligned} E[e^{itT_i}] &= E[e^{itT_i}/i \in \mathcal{P}]P(i \in \mathcal{P}) + E[e^{itT_i}/i \notin \mathcal{P}]P(i \notin \mathcal{P}) \\ &= E[e^{itA_i(1-B_i)}/i \in \mathcal{P}]P(i \in \mathcal{P}) + P(i \notin \mathcal{P}) \end{aligned}$$

Par suite de l'hypothèse d'indépendance, pour chaque i, entre l'événement $\{i \in \mathcal{P}\}$ et la variable $A_i(1-B_i)$, $E[e^{itA_i(1-B_i)}/i \in \mathcal{P}] = E[e^{itA_i(1-B_i)}] = \varphi(t)$ et donc

$$E[e^{itT_i}] = \varphi(t)\frac{n}{k} + \left(1 - \frac{n}{k}\right) = 1 + \frac{n}{k}(\varphi(t)-1).$$

Preuve de la proposition :

Le nombre d'éléments de E étant k, et les T_i étant indépendants, la fonction caractéristique de l'indice aléatoire $Z = \sum_{i\in E} \frac{T_i}{n}$ est $\Psi(t) = \left(1 + \frac{n}{k}\left(\varphi\left(\frac{t}{n}\right) - 1\right)\right)^k$.

Posons $\lambda_1 = E[A_i(1-B_i)]$, $\lambda_2 = E[(A_i(1-B_i))^2]$. On a le développement limité $\varphi(t) = 1 + i\lambda_1 t - \frac{\lambda_2}{2}t^2 + t^2\varepsilon(t)$ et donc :

$$\begin{aligned}\mathrm{Ln}(\Psi(t)) &= k\,\mathrm{Ln}\left(1+\frac{n}{k}\left(i\lambda_1\frac{t}{n}-\frac{\lambda_2}{2}\left(\frac{t}{n}\right)^2+\left(\frac{t}{n}\right)^2\varepsilon\left(\frac{t}{n}\right)\right)\right)\\ &= k\,\mathrm{Ln}\left(1+\frac{t}{k}\left(i\lambda_1-\frac{\lambda_2}{2}\frac{t}{n}+\frac{t}{n}\varepsilon\left(\frac{t}{n}\right)\right)\right)\end{aligned}$$

Utilisons le développement limité $\mathrm{Ln}(1+u)=u-\frac{1}{2}u^2+u^2\varepsilon_1(u)$.

Comme $\left|\frac{t}{k}\right| \leqslant \left|\frac{t}{n}\right|$, on peut poser $\varepsilon_1\left(\frac{t}{k}\left(i\lambda_1-\frac{\lambda_2}{2}\frac{t}{n}+\frac{t}{n}\varepsilon\left(\frac{t}{n}\right)\right)\right)=\varepsilon_2\left(\frac{t}{n}\right)$.

De même $\left(i\lambda_1-\frac{\lambda_2}{2}\frac{t}{n}+\frac{t}{n}\varepsilon\left(\frac{t}{n}\right)\right)^2=-\lambda_1^2+\varepsilon_3\left(\frac{t}{n}\right)$. On a alors

$$\begin{aligned}\mathrm{Ln}(\Psi(t)) = k&\left(\frac{t}{k}\left(i\lambda_1-\frac{\lambda_2}{2}\frac{t}{n}+\frac{t}{n}\varepsilon\left(\frac{t}{n}\right)\right)\right.\\ &\left.+\left(\frac{t}{k}\right)^2\left(-\lambda_1^2+\varepsilon_3\left(\frac{t}{n}\right)\right)\left(-\frac{1}{2}+\varepsilon_2\left(\frac{t}{n}\right)\right)\right)\end{aligned}$$

En posant $\left(-\lambda_1^2+\varepsilon_3\left(\frac{t}{n}\right)\right)\left(-\frac{1}{2}+\varepsilon_2\left(\frac{t}{n}\right)\right)=\frac{\lambda_1^2}{2}+\varepsilon_4\left(\frac{t}{n}\right)$, et après simplification : $\mathrm{Ln}(\Psi(t))=i\lambda_1 t-\frac{\lambda_2}{2}\frac{t^2}{n}+\frac{t^2}{n}\varepsilon\left(\frac{t}{n}\right)+\frac{t^2}{k}\left(\frac{\lambda_1^2}{2}+\varepsilon_4\left(\frac{t}{n}\right)\right)$. Nous considérons que k étant grand devant n, $\frac{n}{k}=\varepsilon_5\left(\frac{1}{n}\right)$, et donc

$$\mathrm{Ln}(\Psi(t))=i\lambda_1 t-\frac{\lambda_2}{2}\frac{t^2}{n}+\frac{t^2}{n}\varepsilon\left(\frac{t}{n}\right)+\frac{t^2}{n}\varepsilon_5\left(\frac{1}{n}\right)\left(\frac{\lambda_1^2}{2}+\varepsilon_4\left(\frac{t}{n}\right)\right).$$

Pour t fixé, $\mathrm{Ln}(\Psi(t))=i\lambda_1 t-\frac{\lambda_2}{2}\frac{t^2}{n}+\frac{1}{n}\varepsilon_6\left(\frac{1}{n}\right)$

2.4. L'intensité de propension

Il résulte de ce qui précède, que, pour n assez grand, la loi de $\dfrac{Z-E[A_i(1-B_i)]}{\sqrt{\dfrac{E[(A_i(1-B_i))^2]}{n}}}$ est approchée par la loi normale $\mathcal{N}(0,1)$.

Par suite de l'indépendance de A_i et B_i on a $E[A_i(1-B_i)]=E(A_i)(1-E(B_i))$ et

$$\begin{aligned}E[(A_i(1-B_i))^2] &= E(A_i^2)E((1-B_i)^2)\\ &= (\mathrm{Var}(A_i)+E(A_i)^2)(\mathrm{Var}(B_i)+(1-E(B_i))^2)\end{aligned}$$

Φ désignant la fonction de répartition de la loi normale centrée réduite, et s l'indice observé $\dfrac{\sum_{i\in P} a_i(1-b_i)}{n}$,

$$P[Z < s] \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{q} e^{-\frac{t^2}{2}}\, dt = 1 - \Phi(-q)$$

avec $q = \dfrac{s - E(A_i)(1-E(B_i))}{\sqrt{\dfrac{(\mathrm{Var}(A_i)+E(A_i)^2)(\mathrm{Var}(B_i)+(1-E(b_i))^2)}{n}}}$. ε étant le seuil choisi, l'hypothèse d'absence de lien sera rejetée si $P[Z < s] < \varepsilon$, c'est-à-dire si $\Phi(-q) \geqslant 1-\varepsilon$.

Soit m_a (resp. m_b) la moyenne empirique de a (resp. b), et soit ν_a (resp. ν_b) leur variance empirique. n étant grand, m_a (resp. m_b) est peu différent de $E(A_i)$ (resp. $E(B_i)$), et ν_a (resp. ν_b) est peu différent de $\mathrm{Var}(A_i)$ (resp. $\mathrm{Var}(B_i)$) et donc $q \approx \dfrac{s - m_a(1-m_b)}{\sqrt{\dfrac{(\nu_a + m_a^2)(\nu_b + (1-m_b)^2)}{n}}}$.

Définitions :

Soit a (resp. b) une variable de moyenne empirique m_a (resp. m_b) et de variance empirique ν_a (resp. ν_b).

• $\tilde{q}(a,\overline{b}) = \dfrac{\dfrac{\sum_{i\in P} a_i(1-b_i)}{n} - m_a(1-m_b)}{\sqrt{\dfrac{(\nu_a + m_a^2)(\nu_b + (1-m_b)^2)}{n}}}$ est le coefficient de propension de a vers b.

• $\Phi(-\tilde{q}(a,\overline{b}))$ est l'intensité de propension de a vers b.

Remarque :

Le numérateur de $\tilde{q}(a,\overline{b})$ s'exprime aussi

$$\frac{\sum_{i\in P} a_i - \sum_{i\in P} a_i b_i}{n} - m_a(1-m_b) = m_a - \frac{\sum_{i\in P} a_i b_i}{n} - m_a(1-m_b) = -\frac{\sum_{i\in P} a_i b_i}{n} + m_a m_b.$$

Il est donc égal à l'opposé de la covariance empirique de a et b. Par conséquent l'intensité de propension de a vers b est supérieure à 0,5 si et seulement si leur covariance est positive.

3. Comparaison propension / implication

Dans ce paragraphe, nous montrons que la propension généralise aux variables à valeurs dans [0; 1] l'implication entre variables binaires. Pour différencier les variables à valeurs dans l'intervalle [0; 1] des variables binaires, nous introduisons le «rapport de variance». Ce rapport traduit la façon dont, statistiquement, la variable prend les valeurs intérieures de l'intervalle. En fonction de ce rapport, nous examinons dans quelle mesure la propriété de compatibilité avec l'ordre des moyennes, résultat remarquable de la relation implicative, reste vraie pour la relation de propension. Nous comparons de même les valeurs de l'intensité d'implication et de l'intensité de propension, de façon à savoir si l'intensité d'implication constitue ou non une approximation fiable de l'intensité de propension.

3.1. Les variables binaires

Une variable binaire est égale à son carré, et donc si a et b sont binaires, avec les notations ci-dessus : $\nu_a + m_a^2 = m_a$ et $\nu_b + (1 - m_b)^2 = m_b$ d'où $\tilde{q}(a, \overline{b}) = \dfrac{s - m_a(1 - m_b)}{\sqrt{\dfrac{m_a(1 - m_b)}{n}}}$. Dans ce cas, $\tilde{q}(a, \overline{b})$ coïncide avec le coefficient $q(a, \overline{b})$ utilisé en analyse implicative. Dans le cas binaire, l'intensité de propension est donc égale à l'intensité d'implication $\Phi(-q(a, \overline{b}))$ introduite par Gras (1979) et Lerman, Gras, Rostram, (1981).

3.2. La compatibilité de la relation avec l'ordre des moyennes

Gras, Lahrer (1993) observent que dans le cas de variables binaires a (resp. b) de moyenne m_a (resp. m_b) si $m_a \leqslant m_b$ alors $\Phi(-q(a, \overline{b})) \geqslant \Phi(-q(b, \overline{a}))$. Donc, si $m_a < m_b$ et si l'implication de b vers a est significative, alors l'implication de a vers b l'est aussi, avec une intensité supérieure. Dans ce cas, on considère seulement l'implication de a vers b, qui est la plus significative. Ce résultat conduit à considérer la relation implicative comme nécessairement orientée de la variable dont la moyenne est la moins forte vers celle dont la moyenne est la plus forte : nous disons que la relation implicative est *compatible* avec l'ordre des moyennes.

Nous examinons cette question dans le cas de variables à valeurs dans [0; 1], car elle conditionne la définition de la propension.

Proposition 2 :

Soient a et b deux variables à valeur dans [0; 1], de moyenne m_a (resp. m_b), de variance ν_a (resp. ν_b) et de covariance positive,

$$\Phi(-\tilde{q}(a, \overline{b})) \geqslant \Phi(-\tilde{q}(b, \overline{a})) \text{ si et seulement si } \frac{1 - 2m_b}{\nu_b + m_b^2} \leqslant \frac{1 - 2m_a}{\nu_a + m_a^2}.$$

Preuve : $\Phi(-\tilde{q}(a, \overline{b})) \geqslant \Phi(-\tilde{q}(b, \overline{a}))$ est équivalent à $\tilde{q}(a, \overline{b}) \leqslant \tilde{q}(b, \overline{a})$. Les numérateurs de ces deux coefficients de propension sont tous les deux égaux à l'opposé de la covariance (supposée positive) de a et b. Les coefficients sont donc dans l'ordre des dénominateurs, et donc

$$\begin{aligned}
&\Phi(-\tilde{q}(a,\overline{b})) \geqslant \Phi(-\tilde{q}(b,\overline{a}))\\
&\Leftrightarrow (\nu_a + m_a^2)(\nu_b + (1-m_b)^2) \leqslant (\nu_b + m_b^2)(\nu_a + (1-m_a)^2)\\
&\Leftrightarrow (\nu_a + m_a^2)(1-2m_b) \leqslant (\nu_b + m_b^2)(1-2m_a)
\end{aligned}$$

Conséquence :

La compatibilité de la relation de propension avec l'ordre des moyennes n'est pas systématiquement vérifiée.

Preuve : considérons par exemple une variable a binaire de moyenne m_a et une variable b de moyenne m_b et de variance $\dfrac{m_b(1-m_b)}{2}$.

$$\frac{1-2m_b}{\nu_b + m_b^2} \leqslant \frac{1-2m_a}{\nu_a + m_a^2} \Leftrightarrow \frac{2(1-2m_b)}{m_b(1+m_b)} \leqslant \frac{1-2m_a}{m_a} \Leftrightarrow m_a \leqslant \frac{m_b(1+m_b)}{2(m_b^2 - m_b + 1)}$$

Pour $m_b < \dfrac{1}{2}$, on a $\dfrac{m_b(1+m_b)}{2(m_b^2-m_b+1)} < m_b$. Donc on peut avoir $\dfrac{m_b(1+m_b)}{2(m_b^2-m_b+1)} < m_a < m_b < \dfrac{1}{2}$ et dans ce cas $m_a < m_b$ et $\Phi(-\tilde{q}(a,\overline{b})) < \Phi(-\tilde{q}(b,\overline{a}))$.

On peut cependant énoncer une condition suffisante pour que la relation de propension soit compatible avec l'ordre des moyennes. Pour cela posons la définition suivante :

Définition

Soit x une variable à valeur dans [0; 1], de moyenne m_x et de variance ν_x non nulle. On appelle «facteur de variance» le réel $\dfrac{m_x(1-m_x)}{\nu_x}$.

Propriété

Pour toute variable x à valeur dans [0; 1], de variance non nulle, le facteur de variance est supérieur ou égal à 1.

Preuve : on a $x^2 \leqslant x$ et donc $m_{x^2} \leqslant m_x$ d'où $\nu_x \leqslant m_x - m_x^2$.

On a alors la proposition suivante, qui généralise le cas binaire :

Proposition 3

Soient a et b deux variables à valeur dans [0; 1], de moyenne m_a (resp. m_b), de covariance positive et de même facteur de variance.

$$\Phi(-\tilde{q}(a,\overline{b})) \geqslant \Phi(-\tilde{q}(b,\overline{a})) \quad \text{si et seulement si } m_a \leqslant m_b$$

Preuve : Soit k le facteur de variance commun à a et b. D'après la proposition 2

$$\Phi(-\tilde{q}(a,\overline{b})) \geqslant \Phi(-\tilde{q}(b,\overline{a})) \Leftrightarrow \frac{1-2m_b}{\nu_b+m_b^2} \leqslant \frac{1-2m_a}{\nu_a+m_a^2}$$

$$\Leftrightarrow \frac{1-2m_b}{m_b(1+(k-1)m_b} \leqslant \frac{1-2m_a}{m_a(1+(k-1)m_a)}.$$

Il suffit ensuite de remarquer que la fonction $x \rightarrow \dfrac{1-2x}{x(1+(k-1)x)}$ est décroissante pour $k \geqslant 1$ et $0 \leqslant x \leqslant 1$.

Remarques et conséquences

• Si x est une variable à valeur dans [0; 1], de facteur de variance k et y une variable binaire de même moyenne que x. On a $\nu_x = \dfrac{\nu_y}{k}$. Le facteur de variance est donc représentatif du resserrement des valeurs de la variable, comparé à celui d'une variable binaire.

• Si pour un ensemble de variables, le resserrement est comparable, on observera la même compatibilité de la relation de propension avec l'ordre des moyennes que dans le cas binaire. Ce sera par exemple le cas si les variables mesurent des comportements comparables dans une population.

3.3. L'intensité d'implication est-elle une approximation fiable de l'intensité de propension?

Gras, Larher (1993) ont conjecturé que l'intensité d'implication $\Phi(-q(a,\overline{b}))$ pouvait rendre compte de l'association de variables binaires ou non. Nous avons vu ci-dessus qu'une modélisation de la relation de propension conduit à l'intensité de propension $\Phi(-\tilde{q}(a,\overline{b}))$.

Comparons ces deux intensités, en reprenant les notations du paragraphe 2. Les numérateurs des coefficients d'implication et de propension étant les mêmes, comparons les dénominateurs. Les variables a et $1-b$ sont à valeurs dans [0; 1], donc on a

$$(\nu_a+m_a^2)(\nu_b+(1-m_b)^2) = m_{a^2}m_{(1-b)^2} \leqslant m_a m_{(1-b)} = m_a(1-m_b).$$

Si l'on considère des relations significatives, les numérateurs sont négatifs et donc $\tilde{q}(a,\overline{b}) \leqslant q(a,\overline{b}) < 0$. Donc $\Phi(-\tilde{q}(a,\overline{b})) \geqslant \Phi(-q(a,\overline{b}))$, l'égalité étant obtenue seulement dans le cas binaire. L'intensité d'implication est donc inférieure à l'intensité de propension. Nous nous proposons d'étudier leurs valeurs dans des cas typiques. Si elles diffèrent peu, on pourra accepter la conjecture de Gras, Larher (1993). Dans le cas contraire, il faudra préférer l'intensité de propension pour une meilleure fiabilité.

Pour simplifier, nous nous plaçons dans le cas de variables ayant le même facteur de variance k. Selon la remarque ci-dessus, nous pouvons supposer que ce cas est le plus fréquent. Ceci nous permet de plus de poser l'hypothèse $m_a \leqslant m_b$, puisque nous nous intéressons à la propension la plus significative.

Proposition :

Soient a (resp. b) à valeur dans [0; 1], de moyenne m_a (resp. m_b), tels que $m_a \leqslant m_b$, et de même facteur de variance k. On a l'encadrement

$$\frac{1}{k} \leqslant \frac{q(a,\overline{b})}{\tilde{q}(a,\overline{b})} \leqslant \frac{k+1}{2k}$$

Preuve : En posant $\alpha = m_a$ et $\beta = m_b$

$$\tilde{q}(a,\overline{b}) = \frac{s-\alpha(1-\beta)}{\sqrt{\dfrac{\left(\dfrac{\alpha(1-\alpha)}{k}+\alpha^2\right)\left(\dfrac{\beta(1-\beta)}{k}+(1-\beta)^2\right)}{n}}}$$

$$= \frac{s-\alpha(1-\beta)}{\sqrt{\dfrac{\alpha(1-\beta)\left(\dfrac{(1-\alpha)}{k}+\alpha\right)\left(\dfrac{\beta}{k}+1-\beta\right)}{n}}}$$

Étudions les variations de $\lambda(\alpha,\beta) = \dfrac{q(a,\overline{b})}{\tilde{q}(a,\overline{b})}$. On a

$$\lambda(\alpha,\beta) = \sqrt{\left(\frac{(1-\alpha)}{k}+\alpha\right)\left(\frac{\beta}{k}+1-\beta\right)}$$

Posons $P(\alpha,\beta) = (1-\alpha+\alpha k)(\beta+k-\beta k)$ de façon que $\lambda(\alpha,\beta) = \dfrac{\sqrt{P(\alpha,\beta)}}{k}$

P étant croissant en m_a et décroissant en m_b, la valeur minimum de λ est : $\lambda(0,1) = \dfrac{1}{k}$

On a $P(\alpha,\beta) - P(\alpha,\alpha) = (\alpha-\beta)(1-\alpha+\alpha k)(k-1)$

Comme $\alpha \leqslant \beta$, on a $P(\alpha,\beta) \leqslant P(\alpha,\alpha) = -\alpha^2(k-1)^2+\alpha(k-1)^2+k$.

La valeur maximum du dernier membre est atteinte en $\alpha = \frac{1}{2}$.

La valeur maximum de λ est donc : $\lambda\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{k+1}{2k}$.

Conséquence : le Tableau 1 donne l'encadrement de λ pour les valeurs de k de 2 à 4, puis des encadrements de l'intensité d'implication pour une intensité de propension valant 0,9 puis 0,95. On constate que pour ces valeurs typiques d'intensités significatives, des valeurs du facteur de variance supérieures ou égales à 2 donnent une intensité d'implication notablement inférieure à l'intensité de propension. Il faudra donc, pour plus de fiabilité, considérer dans ce cas l'intensité de propension plutôt que l'intensité d'implication.

TABLEAU 1

Facteur de variance	Encadrement de λ	Encadrement de $\Phi(-q(a,\overline{b}))$ pour $\Phi(-\tilde{q}(a,\overline{b})) = 0,9$	Encadrement de $\Phi(-q(a,\overline{b}))$ pour $\Phi(-\tilde{q}(a,\overline{b})) = 0,95$
$k = 2$	$\frac{1}{2} \leqslant \lambda \leqslant \frac{3}{4}$	$0,74 \leqslant \Phi(-q(a,\overline{b})) \leqslant 0,83$	$0,79 \leqslant \Phi(-q(a,\overline{b})) \leqslant 0.89$
$k = 3$	$\frac{1}{3} \leqslant \lambda \leqslant \frac{2}{3}$	$0,66 \leqslant \Phi(-q(a,\overline{b})) \leqslant 0,80$	$0,71 \leqslant \Phi(-q(a,\overline{b})) \leqslant 0,86$
$k = 4$	$\frac{1}{4} \leqslant \lambda \leqslant \frac{5}{8}$	$0,63 \leqslant \Phi(-q(a,\overline{b})) \leqslant 0.79$	$0,66 \leqslant \Phi(-q(a,\overline{b})) \leqslant 0,85$

4. Les filiations

4.1. Les graphes implicatifs

Considérons non plus deux variables isolées, mais un ensemble de variables représentant certains comportements d'une population donnée. Gras (1979) a proposé de représenter les implications dans un ensemble de variables sous forme d'un graphe orienté où les variables sont reliées entre elles selon la relation implicative; on ne fait pas apparaître les transitivités et si pour deux variables a et b, les intensités $\Phi(-q(a,\overline{b}))$ et $\Phi(-q(b,\overline{a}))$ sont toutes les deux significatives, on considère seulement l'implication dont l'intensité est la plus forte. Du fait de la compatibilité de la relation avec l'ordre des moyennes, on ne peut pas avoir de cycles, en dehors de ceux qui joignent des variables «jumelles». Deux variables a et b sont jumelles si les intensités d'implication $\Phi(-q(a,\overline{b}))$ et $\Phi(-q(b,\overline{a}))$ sont égales et significatives. Dans ce cas, leurs moyennes sont égales.

Les «graphes implicatifs» ont permis des interprétations intéressantes, notamment par Bailleul (1994). Néanmoins, la non-transitivité de la relation implicative (Larher, 1991) complique l'interprétation de ces graphes, car, le long d'un même chemin peuvent se trouver des variables non consécutives qui ne sont pas en relation implicative. On peut même trouver sur ces chemins des variables dont l'intensité d'implication est inférieure à 0,5, et donc dont la covariance est négative.

Dans le cas de la relation de propension, une difficulté supplémentaire se présente, car, comme nous l'avons vu, la compatibilité de la relation avec l'ordre des moyennes (au sens de **3.2**) n'est pas vérifiée.

4.2. Les chemins de propension

Comme en analyse implicative, nous choisissons, lorsque deux propensions réciproques sont significatives, de conserver celle qui a l'intensité la plus forte. Comme nous l'avons vu au paragraphe 3.2, ce choix peut conduire à retenir une propension non compatible avec l'ordre des moyennes. On peut cependant penser que dans de nombreux exemples, les facteurs de variance des variables étant proches, ce cas ne se présentera pas.

D'autre part, nous imposons aux chemins de propension de relier seulement des variables ayant une covariance positive, et d'être maximaux. Enfin, nous considérons ensemble les chemins ayant la même extrémité, et nous appelons filiations les graphes ainsi formés.

L'existence de cycles serait contraire à la notion intuitive de propension, sauf dans le cas où le cycle est constitué de relations symétriques, c'est-à-dire que pour tous les couples de variables a et b consécutives de ce cycle, l'intensité de propension de a vers b est égale à l'intensité de propension de b vers a. Nous avons vu que, en analyse implicative, tous les cycles sont constitués de telles variables dites «jumelles». Nous montrons ci-dessous que cette propriété reste vraie pour les chemins de propension.

Définitions

On choisit un seuil de confiance ε inférieur à 0,5.

1) **Relation de propension** : aRb si et seulement si $\begin{cases} \Phi(-\tilde{q}(a,\overline{b})) \geqslant 1-\varepsilon \\ \Phi(-\tilde{q}(a,\overline{b})) \geqslant \Phi(-\tilde{q}(b,\overline{a})) \end{cases}$

2) **Chemins de propension** : ce sont les n-uplets $(a_0 \dots a_n)$ de variables distinctes tels que :

$$\begin{cases} \forall i \; a_i R \, a_{i+1} \\ \forall i,j \; \mathrm{cov}(a_i, a_j) > 0 \\ \forall c \notin \{a_0 \dots a_n\} non(cRa_0) \text{ ou } non \; (a_n Rc) \text{ ou } (\exists i \; \mathrm{cov}(c, a_i) \leqslant 0) \end{cases}$$

3) **Filiations** : une filiation est constituée des chemins de propension ayant une extrémité commune.

4) **Variables jumelles** : deux variables a, b sont jumelles si et seulement si

$$\Phi(-\tilde{q}(a,\overline{b})) = \Phi(-\tilde{q}(b,\overline{a})) \geqslant 1-\varepsilon$$

Propriétés

Nous formalisons et démontrons les propriétés avancées ci-dessus.

1) Si aRb alors $\text{cov}(a,b) > 0$

2) Si aRb et si a et b ont le même facteur de variance, alors $m_a \geqslant m_b$

3) Les chemins sont sans cycles, sauf entre variables jumelles.

Preuves :

1) Résulte de la remarque du paragraphe **2.4.**

2) Résulte de la proposition 3 du paragraphe **3.2.**

3) Considérons un cycle, c'est-à-dire une suite $(a_0 \dots a_n)$ où $\forall i\ a_i R\ a_{i+1}$ et $a_0 = a_n$.

Il résulte de la définition de R, que, alors, $\forall i\ \Phi(-\tilde{q}(a_i, \overline{a}_{i+1})) \geqslant \Phi(-\tilde{q}(a_{i+1}, \overline{a}_i))$. D'après la proposition 2 du paragraphe **3.2.** $\forall i\ \dfrac{1-2m_{a_{i+1}}}{\nu_{a_{i+1}} + m^2_{a_{i+1}}} \leqslant \dfrac{1-2m_{a_i}}{\nu_{a_i} + m^2_{a_i}}$ et donc tous ces quotients sont égaux. En appliquant la proposition 2, partie réciproque, $\forall i\ \Phi(-\tilde{q}(a_i, \overline{a}_{i+1})) = \Phi(-\tilde{q}(a_{i+1}, \overline{a}_i))$ et donc a_i et a_{i+1} sont jumelles.

5. Le traitement d'un questionnaire à réponses modales

5.1. Choix de la définition des variables et du traitement associé

Nous considérons ici un questionnaire constitué d'un ensemble O d'opinions, pour chacune desquelles les individus ont à choisir une modalité de réponse parmi les 4 suivantes : «*Pas du tout d'accord*», «*Plutôt en désaccord*», «*Plutôt d'accord*», «*Tout à fait d'accord*». Parmi les définitions possibles des variables, et les traitements associés voici deux options principales.

1) Considérer le choix d'un individu pour une opinion comme une variable qualitative à quatre modalités. Sur les variables ainsi définies, rechercher et analyser des associations dissymétriques (Abdesselam et Schektman, 1996).

2) Quantifier les différentes modalités de réponse de façon à obtenir une variable à valeur dans [0; 1]. Traiter les variables ainsi définies selon les principes définis dans cet article.

La quantification a nécessairement un côté arbitraire. A des fins de comparaison, nous nous proposons d'étudier deux options :

2.a) Une quantification binaire. On affecte 0 aux deux modalités de désaccord, et 1 aux deux modalités d'accord. On traite alors avec l'intensité d'implication.

2.b) Une quantification «nuancée». On affecte respectivement 0, 1/3, 2/3, 1 aux différentes modalités dans l'ordre croissant des degrés d'accord. On traite alors avec l'intensité de propension.

5.2. Comparaison

Considérons trois jeux de données simulées. Le Tableau 2 donne les effectifs croisés de réponses à deux questions x et y suivant les 4 modalités indiquées ci-dessus ($x1$ et $y1$ pour «*Pas du tout d'accord*», . . ., $x4$ et $y4$ pour «*Tout à fait d'accord*»). Le premier jeu de données $J1$ est symétrique en x et y. Nous l'avons construit pour avoir une valeur significative de l'intensité de propension. Le jeu de données $J2$ est obtenu en «nuançant» les réponses de 60 personnes à la question x : elles passent de $x4$ à $x3$, sans changer leur réponse à y. Le jeu de données $J3$ est obtenu en nuançant à nouveau les réponses de ces 60 personnes, mais cette fois à la question y.

TABLEAU 2

$J1$	$y1$	$y2$	$y3$	$y4$		$J2$	$y1$	$y2$	$y3$	$y4$		$J3$	$y1$	$y2$	$y3$	$y4$
$x1$	10	10	10	10		$x1$	10	10	10	10		$x1$	10	10	10	10
$x2$	10	10	10	10		$x2$	10	10	10	10		$x2$	10	10	10	10
$x3$	10	10	10	10		$x3$	10	10	10	70		$x3$	10	10	70	10
$x4$	10	10	10	70		$x4$	10	10	10	10		$x4$	10	10	10	10

Pour chaque jeu de données, le Tableau 3 présente

1) Le tau de Goodman-Kruskal, obtenu en considérant x et y comme deux variables à modalités qualitatives. En effet, Abdesselam et Schektman (1996) présentent ce tau comme une mesure possible de l'association dissymétrique entre deux variables de ce type.

2) L'intensité d'implication entre deux variables binaires obtenues en affectant zéro aux deux modalités de désaccord, et un aux deux modalités d'accord.

3) L'intensité de propension entre deux variables à valeurs dans $[0; 1]$, obtenues en affectant respectivement 0; 1/3, 2/3, 1 à $x1$ et $y1$, $x2$ et $y2$, $x3$ et $y3$, $x4$ et $y4$.

Le tau de Goodman est identique dans les trois jeux : en effet, on passe d'un jeu à l'autre par une permutation de modalités qui n'affecte pas l'association dissymétrique. Comme toutes les mesures de l'association dissymétrique de deux variables, le tau est une sommation sur les couples (i, j) de modalités en x et en y. Il est possible de repérer dans chacun des jeux de données, un terme prépondérant dans cette sommation et de l'interpréter comme «xi explique yj». On constate ainsi, que le même tau d'association dissymétrique dans les trois jeux s'interprète par des associations différentes de modalités.

L'intensité d'implication entre variables binaires est aussi la même dans les trois jeux. En effet, on nuance les réponses à l'intérieur de modalités d'accord, et donc, les variables binaires ne changent pas d'un jeu à l'autre.

L'intensité de propension entre variables quantifiées sur l'échelle 0, 1/3, 2/3, 1 varie fortement d'un jeu de données à l'autre. La propension est très significative pour le jeu $J1$, non significative pour le jeu $J2$, et quasi inexistante pour le jeu $J3$.

TABLEAU 3

		$J1$	$J2$	$J3$
variables	tau de Goodman	0,10	0,10	0,10
qualitatives	interprétation	$x4$ explique $y4$	$x3$ explique $y4$	$x3$ explique $y3$
quantification binaire	intensité d'implication	0,94	0,94	0,94
quantification nuancée	intensité de propension	0,97	0,77	0,59

On constate donc que, avec la quantification choisie, l'intensité de propension est très sensible au déplacement d'individus dans les nuances de réponses, à la différence des tau et intensité dans les autres choix. Si l'on fait le choix d'une optique qualitative, on devra, pour deux variables, étudier des associations parmi 16 couples de modalités telles que «*Plutôt en désaccord* avec x explique *Tout à fait d'accord* avec y», ce qui dépasse en finesse d'analyse ce que nous envisagions. Si l'on se place dans une optique quantitative, il est justifié d'employer une échelle du type 0, 1/3, 2/3, 1, plutôt que binaire[2] , qui ne respecte pas les nuances exprimées par les sujets.

6. Application à une recherche en didactique

Nous avons eu à traiter les réponses à un questionnaire passé par 460 élèves de l'enseignement secondaire. Ces élèves participaient à une expérience d'introduction du logiciel de calcul formel DERIVE dans l'enseignement des mathématiques (Lagrange, Drouhard, 1995) (Lagrange, 1996). Les professeurs, membres d'un groupe piloté par le ministère de l'Education, avaient établi une liste de 17 opinions sur la pertinence du logiciel pour l'apprentissage des mathématiques (Annexe 1, Tableau 7). Il s'agit dans ce paragraphe de montrer quels résultats statistiques et didactiques il est possible d'obtenir à partir du traitement décrit dans l'article.

6.1. Problématique

Voici un exemple de relation entre opinions que nous cherchions à tester. Le questionnaire proposait l'opinion numérotée 3, significative d'une résistance à l'utilisation du logiciel : «*Ça ne sert à rien de travailler avec DERIVE, puisque, aux contrôles et aux examens, il faut rédiger les calculs et les démonstrations*». Nous nous attendions à ce que les difficultés rencontrées dans l'utilisation du logiciel entraînent une «propension» à cette attitude. L'opinion numérotée 7 permettait à l'élève de témoigner de difficultés rencontrées : «*DERIVE, ça complique plus que ça n'aide pour apprendre des mathématiques*». Nous nous attendions à ce que cette opinion

[2] L'échelle adoptée pour la quantification «nuancée» reste bien sûr arbitraire. Nous montrons seulement qu'elle est préférable à une quantification binaire.

soit plus minoritaire que la première, et nous cherchions à quantifier la propension que donnerait cette seconde opinion à la première. Nous souhaitions savoir de plus si d'autres opinions sur DERIVE étaient, elles aussi, susceptibles de conduire à cette attitude de résistance, c'est pourquoi il nous fallait organiser l'ensemble des relations entre les 17 opinions.

6.2. Résultats

Quatre possibilités de réponse étaient proposées aux élèves comme au paragraphe 5.1. Nous avons fait le choix de la quantification «nuancée», avec les coefficients considérés ci-dessus : 0, 1/3, 2/3, 1. Nous avons calculé les moyennes et facteurs de variance des 17 opinions (Tableau 8, Annexe 2), puis les propensions des 17×17 couples de variables (Tableau 9). Nous étudions la distribution du facteur de variance dans les différentes opinions, puis quelques propensions et filiations à titre d'exemple, ainsi que l'interprétation didactique qu'on peut en faire.

Le facteur de variance

Au paragraphe 3.2, la notion de facteur de variance a été introduite comme représentative du regroupement des valeurs d'une variable à valeur dans $[0; 1]$, comparé à celui d'une variable binaire de même moyenne. Dans le cas d'un questionnaire d'opinion, le facteur de variance mesure la façon dont les individus utilisent les nuances intermédiaires. Nous avons démontré que la propriété remarquable des variables binaires, de compatibilité de la relation avec l'ordre des moyennes n'est pas conservée en cas de valeurs différentes des facteurs de variance. D'autre part, (paragraphe 3.3) l'intensité d'implication est d'autant moins une approximation fiable de l'intensité de propension que le facteur de variance est élevé.

Il est donc intéressant d'étudier ce facteur dans le cas de données réelles du questionnaire. Pour les 17 opinions, les valeurs du facteur de variance sont situées dans l'intervalle [1,7 ; 3,3] avec une moyenne de 2,5 et un écart type de 0,5 (Tableau 8). La valeur moyenne de 2,5 justifie l'utilisation de l'intensité de propension, plutôt que de l'intensité d'implication. La relative faiblesse de l'écart type permet de penser que l'on est dans une configuration analogue au cas binaire en ce qui concerne la compatibilité de la relation avec l'ordre des moyennes[3] .

Exemples de propensions et de filiations

Nous avons choisi un seuil de risque $\varepsilon = 0,08$. Ceci a permis de déterminer 19 propensions significatives parmi les 16×17 couples de variables du questionnaire (Annexe 2, Tableau 9). Nous présentons deux exemples parmi les quatre filiations déterminées et analysées dans (Lagrange et Drouhard, 1995) et (Lagrange, 1996). L'une se présente comme un système d'explication d'une opinion négative sur

[3] Une seule propension significative contredit cette compatibilité. Il s'agit de la propension d'intensité voisine de 1 de 16 vers 17. Les facteurs de variance de ces deux variables diffèrent de deux écarts types. Contrairement aux autres relations, cette propension ne s'intègre pas dans une filiation plausible.

l'emploi du logiciel. L'autre rassemble les opinions où l'élève perçoit les apports les plus pertinents du logiciel à l'enseignement.

La filiation «Refus de DERIVE»

Cette filiation (Tableau 4) organise trois opinions seulement. Les opinions 2 et 7 entraînent toutes les deux une propension à l'opinion 3, mais n'ont pas entre elles de relation de ce type.

On s'attendait à la propension de 7 vers 3 : elle indique que si un élève considère l'utilisation de DERIVE comme malaisée, cela peut le conduire à nier la pertinence de ce logiciel. Nous avons pu vérifier, en croisant avec une autre question, que près des trois quarts des élèves les plus concernés par cette propension rencontrent des difficultés d'utilisation du logiciel.

TABLEAU 4

«Refus de DERIVE»

7. DERIVE, ça complique plus que ça n'aide pour apprendre des mathématique	2. Avec DERIVE, il n'y a plus besoin d'apprendre à calculer, il le fait à notre place
↓	↓
3. Ca ne sert à rien de travailler avec DERIVE, puisque, aux contrôles et aux examens, il faut rédiger les calculs et les démonstrations	

La propension de 2 vers 3 était moins prévisible. L'opinion 2 est une vision optimiste des possibilités de DERIVE, opposée à l'opinion 7. Nous constatons que cette vision optimiste donne aussi une propension à nier la pertinence du logiciel. Cette propension s'explique pour des élèves qui accordent une grande importance à l'aspect technique ou calculatoire de l'activité mathématique. Certains ont besoin de faire les calculs eux-mêmes pour bien comprendre une notion. D'autres, ayant des difficultés quand il s'agit de raisonner, s'assurent une réussite minimum grâce aux tâches techniques. On peut voir dans leur refus tendanciel de DERIVE l'appréhension de voir disparaître les tâches qu'ils réussissent.

La filiation «Apports positifs de DERIVE à l'activité mathématique»

Cette filiation (Tableau 5) organise 5 opinions. L'opinion maximale (n° 16) est clairement favorable à l'emploi du logiciel pour l'enseignement des mathématiques. Les 4 autres viennent expliquer cette vision positive du logiciel.

Les opinions 5, 15 et 16 forment un chemin de propension. Pour des élèves auxquels DERIVE donne envie de faire des mathématiques, cette discipline présente

TABLEAU 5

«Apports positifs de DERIVE à l'activité mathématique»

5. DERIVE, ça donne envie de faire des mathématiques

6. DERIVE, ça aide à comprendre les mathématiques

14. Avec DERIVE, on peut résoudre des problèmes différents des exercices habituels

15. DERIVE, ça fait voir les maths autrement

16. DERIVE c'est bien car on peut travailler en même temps sur les équations et les graphiques

alors un intérêt nouveau, qui motive leur vision positive du logiciel. L'attrait de DERIVE résulte donc tendanciellement du changement dans la vision de la discipline que permet son introduction.

Dans cette organisation linéaire, les opinions 6 et 14 introduisent deux tendances. La première, avec l'opinion 6, insiste sur l'apport de DERIVE à la compréhension des mathématiques. Des élèves ont pu, grâce à un usage personnel du logiciel, s'approprier mieux certaines notions. La seconde tendance, avec l'opinion 14, met en avant la pratique de résolution de «vrais problèmes» souvent constatée avec le logiciel. Il s'agit de problèmes de modélisation souvent inspirés de situations extra-mathématiques. DERIVE permet que l'élève puisse les aborder en facilitant les nombreux calculs que, généralement, leur résolution entraîne. Ils diffèrent de la pratique habituelle où l'élève a surtout à résoudre des exercices d'application.

Un résultat intéressant dans cette filiation est l'absence de lien significatif entre les opinions 6 et 14. Il signifie que, parmi les raisons d'apprécier le logiciel, les deux tendances sont bien distinctes. Il permet de mieux comprendre les attitudes d'élèves vis-à-vis de l'emploi du logiciel. En effet, les professeurs associent recherche de nouveaux problèmes et meilleure compréhension : ils proposent ces tâches de modélisation pour que l'élève comprenne mieux les notions utilisées dans la résolution. Pour les élèves, c'est différent. Beaucoup d'entre eux apprécient DERIVE pour les nouveaux problèmes sur lesquels ils ont travaillé, mais ne pensent pas que cette nouvelle pratique des mathématiques améliore de façon décisive leur compréhension des notions. Les élèves qui ont pu améliorer leur compréhension des mathématiques grâce au logiciel l'ont fait dans des circonstances autres que la résolution de problèmes nouveaux.

6.3. Comparaison à un autre choix de quantification

Au paragraphe 5, nous avons comparé sur des données simulées la quantification «nuancée» adoptée pour le traitement du questionnaire DERIVE et une quantification binaire. Nous proposons d'effectuer ici cette comparaison sur les données réelles du questionnaire DERIVE. Pour simplifier, cette comparaison est limitée aux associations présentes dans les deux filiations présentées ci-dessus. Le Tableau 6 donne les intensités d'implication obtenues avec une quantification binaire pour les 7 modalités des deux filiations.

TABLEAU 6
Intensités d'implication dans une quantification binaire.
(Seules apparaissent les intensités supérieures à 0,5)

$\nearrow$	$q2$	$q3$	$q7$	$q5$	$q6$	$q14$	$q15$	$q16$
$q2$	1	0,99				0,56		0,88
$q3$		1						
$q7$	0,94	1	1					0,70
$q5$				1	1	0,93	1	0,88
$q6$					1	0,91	0,99	0,51
$q14$						1		0,91
$q15$						0,87	1	0,85
$q16$								1

En comparant au Tableau 9, on constate trois changements importants :

• l'association $7 \mapsto 2$ devient significative au seuil choisi.

• l'association $6 \mapsto 14$ se renforce, sans toutefois devenir significative.

• l'association $14 \mapsto 15$ n'existe plus. L'association inverse existe, mais n'est pas significative.

Le premier changement justifie à nos yeux la quantification «nuancée». En effet, contrairement à la quantification binaire, elle fait apparaître les opinions 7 et 2 comme deux explications bien distinctes du refus de DERIVE. Le second changement renforce ce jugement. En effet les opinions 6 et 14 constituent deux raisons bien distinctes d'apprécier DERIVE.

6.4. Apports didactiques

Une étude didactique vise à mieux comprendre des phénomènes d'enseignement et d'apprentissage. Les deux filiations rapportées ci-dessus sont utiles pour rendre compte de ce qui peut se passer lors de l'introduction d'un moyen technologique nouveau comme le calcul formel.

La filiation «Refus de DERIVE» montre que des difficultés d'utilisation du logiciel peuvent être à l'origine de phénomènes de refus. Elles montre aussi que ce n'est pas la seule raison de refus, et que tous les élèves ne sont pas prêts d'emblée à accepter qu'un logiciel prenne en charge des tâches techniques de calcul qu'ils estiment signifiantes.

La filiation «Apports positifs de DERIVE à l'activité mathématique», confrontée à des observations de classe, montre les limites de stratégies d'utilisation du logiciel pour la recherche de nouveaux problèmes. Certes, dans ces stratégies, les élèves cherchent activement, s'intéressent, mais l'utilisation du logiciel peut favoriser des recherches de solution par essai/erreur qui exploitent peu la compréhension des mathématiques. La question cruciale de l'apport d'un logiciel de calcul formel à la compréhension est probablement à poser en d'autres termes.

7. Pour conclure

Artigue (1996) montre comment les filiations dégagées ci-dessus complètent les données issues de l'observation directe d'un petit nombre de classes. On obtient ainsi des résultats didactiques issus à la fois d'une observation fine de situations de classe, et d'une enquête à large échelle.

Le calcul de l'intensité de propension et l'organisation en filiations sont donc susceptibles de répondre aux préoccupations suivantes :

• reposer sur une modélisation probabiliste fiable, basée sur des principes qui ont fait leurs preuves dans le cas de variables binaires,

• organiser les relations entre différentes opinions exprimées de façon nuancée par les sujets,

• permettre des interprétations didactiques.

Ces idées ont fait leur chemin parmi les utilisateurs de l'analyse implicative : le logiciel CHIC[4] incorpore maintenant le calcul de l'intensité de propension pour traiter les variables non binaires.

Au cours de l'élaboration théorique, nous avons rencontré la notion de «rapport de variance» et noté que les valeurs relatives de ce rapport jouent un rôle dans les propriétés de l'intensité de propension. Dans l'exemple du questionnaire DERIVE, les valeurs de ce rapport sont assez homogènes. Il resterait à voir si le traitement conserve son intérêt dans le cas de valeurs très différentes de ce rapport.

[4] CHIC 0.51, logiciel pour la Classification Hiérarchique Implicative et Cohésive. Association pour la Recherche en Didactique des Mathématiques.

Références bibliographiques

ABDESSALAM R., SCHEKTMAN Y., *Une analyse factorielle de l'association dissymétrique entre deux variables qualitatives*, Revue de Statistique Appliquée, **XLIV** (2), 1996, p. 5-34

ARTIGUE M., *Le logiciel «DERIVE» comme révélateur de phénomènes didactiques liés à l'utilisation d'environnements informatiques pour l'apprentissage*. Educational Studies in Mathematics, 33 (2), 1996, p. 133-169.

BAILLEUL M., *Analyse statistique implicative : variables modales et contribution des sujets. Application à la modélisation de l'enseignant dans le système didactique*. Thèse Univ. Rennes 1, 1994.

GRAS R., *Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques*. Thèse d'Etat. Univ. Rennes 1, 1979.

GRAS R., LARHER A., *L'implication statistique, une nouvelle méthode d'analyse de données*. Mathématiques, Informatique et Sciences Humaines, n° 120, 1993, p. 5-31

LAGRANGE J.B., DROUHARD J.-P., *L'intégration du système de mathématiques symboliques DERIVE*. Environnements Interactifs d'Apprentissage avec Ordinateurs, Eyrolles, Paris, 1995.

LAGRANGE J.B., *Analysing actual use of a computer algebra system in the teaching and learning of mathematics*. International DERIVE Journal. n°3.3, 1996, p. 91-107.

LARHER A., *Implication statistique et application à l'analyse de démarches de preuve mathématique*. Thèse Univ. Rennes 1, 1991.

LERMAN I.C., GRAS R., ROSTRAM H., *Elaboration et évaluation d'un indice d'implication pour des données binaires*. Mathématiques et Sciences Humaines, 1981, n°74 p.5-35 et n°75, p. 5-47.

LERMAN I.C., *Rôle de l'inférence statistique dans une approche de l'analyse classificatoire des données*. Journal de la Société de statistique de Paris. Tome 127, n°4, 1986.

8. Annexe 1

TABLEAU 7

Le questionnaire DERIVE

Pour chacunes des phrases suivantes, cochez la case qui correspond le plus à votre opinion

	Sans opinion	Pas du tout d'accord	Plutôt pas d'accord	Plutôt d'accord	Tout à fait d'accord
1. Avec DERIVE, même si on a des difficultés en algèbre, on peut faire des mathématiques	❑	❑	❑	❑	❑
2. Avec DERIVE, il n'y a plus besoin d'apprendre à calculer, il le fait à notre place	❑	❑	❑	❑	❑
3. Ça ne sert à rien de travailler avec DERIVE, puisque, aux contrôles et aux examens, il faut rédiger les calculs et les démonstrations	❑	❑	❑	❑	❑
4. Quand on utilise DERIVE, il faut bien organiser son travail, sinon on perd beaucoup de temps	❑	❑	❑	❑	❑
5. DERIVE, ça donne envie de faire des mathématiques	❑	❑	❑	❑	❑
6. DERIVE, ça aide à comprendre les mathématiques	❑	❑	❑	❑	❑
7. DERIVE, ça complique plus que ça n'aide pour apprendre des mathématiques	❑	❑	❑	❑	❑
8. Quand DERIVE ne donne pas de réponse, c'est que le problème n'a pas de solution	❑	❑	❑	❑	❑
9. DERIVE, c'est bien pour avoir une idée du résultat avant de se lancer dans les calculs	❑	❑	❑	❑	❑
10. DERIVE c'est bien pour contrôler ses réponses	❑	❑	❑	❑	❑
11. DERIVE, c'est bien pour découvrir des règles de calcul	❑	❑	❑	❑	❑
12. DERIVE, ça aide surtout quand les calculs sont longs et pénibles	❑	❑	❑	❑	❑
13. DERIVE, ça permet de résoudre des problèmes sans se perdre dans les calculs	❑	❑	❑	❑	❑
14. Avec DERIVE, on peut résoudre des problèmes différents des exercices habituels	❑	❑	❑	❑	❑
15. DERIVE, ça fait voir les maths autrement	❑	❑	❑	❑	❑
16. DERIVE c'est bien car on peut travailler en même temps sur les équations et les graphiques	❑	❑	❑	❑	❑
17. Si les données sont rentrées correctement, on peut faire confiance à 100 % au résultat donné par DERIVE	❑	❑	❑	❑	❑

9. Annexe 2

TABLEAU 8

Moyennes et facteurs de variance

	$q1$	$q2$	$q3$	$q4$	$q5$	$q6$	$q7$	$q8$	$q9$	$q10$	$q11$	$q12$	$q13$	$q14$	$q15$	$q16$	$q17$
moyenne	0,61	0,41	0,48	0,70	0,45	0,51	0,26	0,21	0,75	0,86	0,43	0,85	0,74	0,65	0,66	0,79	0,71
facteur de variance	3,21	1,64	2,03	2,49	2,27	2,75	2,54	1,74	2,42	3,17	2,32	2,83	2,41	2,87	2,88	3,33	2,27

TABLEAU 9

Intensités de propension

Seules apparaissent les intensités supérieures à 0,5.

$\nearrow$	$q1$	$q2$	$q3$	$q4$	$q5$	$q6$	$q7$	$q8$	$q9$	$q10$	$q11$	$q12$	$q13$	$q14$	$q15$	$q16$	$q17$
$q1$	1	0,79			0,73	0,75		0,69	0,69	0,85	0,64	0,94	0,8	0,79	0,85	0,91	0,91
$q2$	0,92	1	0,98				0,73	0,79	0,62	0,59		0,68	0,85	0,62		0,74	0,97
$q3$		0,95	1	0,65			0,9	0,77									0,58
$q4$			0,56	1	0,53	0,52	0,58		0,65	0,63	0,55	0,64		0,52	0,67	0,57	0,54
$q5$	0,83			0,57	1	0,99			0,66	0,73	0,82	0,77	0,86	0,93	1	0,93	0,8
$q6$	0,83			0,54	0,96	1			0,52	0,54	0,94	0,69	0,85	0,78	0,96	0,78	0,79
$q7$		0,83	0,99	0,76			1	0,59	0,68								
$q8$	0,88	0,87	0,89				0,58	1			0,77		0,58	0,55		0,54	0,93
$q9$	0,63	0,55		0,63	0,57	0,6	0,55		1	0,92	0,57	0,79	0,71	0,63	0,59	0,55	0,55
$q10$	0,66	0,52		0,57	0,56	0,52			0,8	1	0,52	0,87	0,81	0,56	0,63	0,79	
$q11$	0,71			0,63	0,83	0,97		0,7	0,52	0,56	1	0,87	0,94	0,96	0,92	0,76	0,74
$q12$	0,71	0,55		0,56	0,58	0,56			0,69	0,89	0,62	1	0,95	0,74	0,58	0,77	0,65
$q13$	0,72	0,63			0,69	0,74		0,52	0,71	0,94	0,75	0,99	1	0,83	0,83	0,85	0,52
$q14$	0,78	0,57		0,52	0,82	0,77		0,52	0,66	0,64	0,85	0,92	0,98	1	0,95	0,88	0,78
$q15$	0,82			0,69	0,94	0,88			0,68	0,76	0,78	0,66	0,86	0,93	1	0,98	0,74
$q16$	0,79	0,59		0,56	0,72	0,64		0,52	0,54	0,89	0,64	0,85	0,82	0,77	0,92	1	1
$q17$	0,84	0,8	0,54	0,54	0,67	0,68		0,69	0,56		0,62	0,75	0,53	0,74	0,72	0,84	1