PARTIE 1

Textes inventoriés, coordonnés, mis en forme ou reformulés par

Régis Gras* et Jean-Claude Régnier**

*LINA– Ecole Polytechnique de l'Université de Nantes, UMR 6241
La Chantrerie BP 60601 44306 Nantes cedex
regisgra@club-internet.fr
** Université de Lyon - UMR 5191 ICAR
ENS-LSH 15, Parvis René Descartes BP 7000 69342 LYON cedex 07
jean-claude.regnier@univ-lyon2.fr

FONDEMENTS THÉORIQUES DE L'ANALYSE STATISTIQUE IMPLICATIVE

Résumé. La partie 1 vise à exposer en 9 chapitres, la théorie de l'Analyse Statistique implicative (ASI). Nous cherchons à y définir le plus précisément possible les concepts et les théorèmes de cette théorie ainsi que leurs fondements épistémologiques et méthodologiques. Parmi ceux-ci, citons : relation de quasi-implication, coefficient et indice d'implication, de propension, graphe implicatif, classification hiérarchique orientée, etc. De plus, chaque concept est illustré par un exemple.

Chapitre 1 : Analyse implicative des variables binaires. Intensité implicative. Intensité entropique.

1 Approche épistémologique de l'ASI

Les connaissances opératoires que les êtres humains construisent sur le monde se constituent principalement selon deux ordres : celui des faits et celui des règles entre les faits ou entre règles elles-mêmes. Ce sont leurs apprentissages qui, à travers leur culture et au travers d'expériences sociales ou singulières, leur permettent une élaboration progressive de ces formes de connaissances, en dépit des régressions, des remises en cause, des ruptures qui surgissent au détour d'infirmations décisives. Cependant, on sait que celles-ci contribuent dialectiquement à lui assurer un équilibre opératoire. Or les règles se forment inductivement de façon relativement stable dès lors que le nombre de succès, quant à leur qualité explicative ou anticipatrice, atteint un certain niveau de confiance à partir duquel elles seront susceptibles d´être mises en oeuvre. En revanche, quand ce niveau subjectif demeure non atteint, l´économie de l´individu le fera résister, dans un premier temps, à son abandon ou à sa critique. En effet, il est coûteux de substituer à la règle initiale une autre règle lors de

l´apparition d´un faible nombre d'infirmations, dans la mesure où elle aurait été confortée par un nombre important de confirmations. Un accroissement de ce nombre d'instances négatives, fonction de la qualité de robustesse du niveau de confiance en la règle, conduira peut-être à un réajustement de celle-ci, voire à son abandon. Laurent Fleury (Fleury 1996), dans sa thèse, cite avec pertinence l´exemple de la règle tout à fait admissible : « toutes les Ferrari sont rouges ». Cette règle, très robuste, ne sera pas pour autant abandonnée lors de l´observation d´un seul ou de deux contre-exemples. D´autant plus qu´elle ne manquerait pas d´être rapidement re-confortée.

Ainsi, à l´opposé de ce qui est légitime en mathématiques, où aucune règle (théorème) ne souffre d´exception et où le déterminisme est total, les règles dans les sciences humaines et sociales, plus généralement dans les sciences dites "molles", sont acceptables et donc opératoires tant que le nombre de contre-exemples restera "supportable" en rapport à la fréquence de situations où elles seront positives et efficaces. Le problème dans le champ de l'analyse des données, est alors d´établir un critère numérique, relativement consensuel, pour définir la notion de niveau de confiance ajustable au niveau d´exigence de l´usager de la règle. Qu´il soit alors établi sur des bases statistiques peut alors ne pas surprendre. Qu´il possède une propriété de résistance non linéaire au bruit (faiblesse du ou des premiers contre-exemples) peut également paraître naturel, conforme au sens "économique" évoqué plus haut. Qu´il s´effondre si les contre-exemples se répètent semble aussi devoir guider notre choix dans la modélisation des critères recherchés.

Différentes approches théoriques ont été adoptées pour modéliser l'extraction et la représentation de règles d'inférence imprécises (ou partielles) entre variables binaires (ou attributs ou caractères) décrivant une population d'individus (ou sujets ou objets). Mais les situations de départ et la nature des données ne modifient pas la problématique initiale. Il s'agit de découvrir des règles inductives non symétriques pour modéliser des relations du type *"si a alors presque b"*. C'est, par exemple, l'option des réseaux bayésiens (Amarger et al 1991, Pearl 1988) ou des treillis de Galois (Simon 2000). Mais le plus souvent, l'indice de corrélation linéaire et le test du χ^2, s'avérant inadaptés du fait de leur caractère symétrique, la probabilité conditionnelle (Loevinger 1947, Agrawal et al. 1993, Gras et al. 2004) reste le moteur de la définition de l'association, même quand l'indice de cette association retenu est de type multivarié (Bernard et Poitrenaud 1999).

De plus et à notre connaissance à ce jour, d'une part, les différents développements intéressants se centrent le plus souvent sur des propositions d'un indice d'implication partielle pour des données binaires (Lerman et al. 2004) ou (Lallich et al 2005) et, d'autre part, cette notion n'est pas étendue à d'autres types de variables que les variables binaires, à l'extraction et à la représentation selon un graphe de règles ou selon une hiérarchie de méta-règles en tant que structures visant l'accès à la signification d'un tout non réduit à la somme de ses parties[1], c'est-à-dire fonctionnant comme un système complexe non linéaire. Par exemple, on sait fort bien, par l'usage même, que la signification d'une phrase ne passe pas complètement par le sens de chacun des mots qui la compose. Par ailleurs il semblerait que, dans la littérature consacrée à ces questions, la notion d'indice d'implication ne soit pas non plus étendue à la recherche de sujets et de catégories de sujets responsables des associations. Ni même que

[1] C'est ce que souligne le philosophe L. Sève : « …dans le passage non additif, non linéaire des parties au tout, il y a *apparition de propriétés* qui ne sont d'aucune manière *précontenues* dans les parties et qui ne peuvent donc s'expliquer par elles » (*Émergence, complexité et dialectique* , Odile Jacob, mai 2005).

cette responsabilité soit quantifiée et conduise, de ce fait, à une structuration réciproque de l'ensemble des sujets, conditionnée par leurs relations aux variables.

2 Situation fondamentale et fondatrice de l'approche classique.

Une population E d'individus, objets ou sujets, est croisée avec des variables (caractères, critères, réussites, etc.) que l'on explore de la façon suivante : "*dans quelle mesure peut-on considérer qu'instancier la variable*[2] *a implique instancier la variable b ?* Autrement dit, *les individus ont-ils tendance à être b si l'on sait qu'ils sont a ?"*. Dans les situations habituelles de la vie humaine ou dans les domaines des sciences de la vie ou des sciences humaines et sociales, où les théorèmes (si a alors b) au sens déductif du terme ne peuvent être établis du fait des exceptions qui les entachent, il est important pour le chercheur comme pour le praticien de "*fouiller dans ses données*" afin de dégager, malgré tout, des règles suffisamment fiables (des sortes de "théorèmes partiels", des inductions) pour pouvoir conjecturer[3] une possible relation causale ou pour le moins quasi-causale, par exemple, pour décrire, structurer une population et faire l'hypothèse d'une certaine stabilité à des fins descriptives et, si possible, prédictives. Ainsi Hervé Londeix[4] recourant à l'ASI dans un cadre de psychologie différentielle met en évidence un ordre partiel des stades piagétiens à partir d'épreuves proposées à de jeunes enfants Mais cette fouille exige la mise au point de méthodes pour la guider et pour la dégager du tâtonnement et de l'empirisme.

3 Modélisation mathématique de l'approche classique.

Pour ce faire, à l'instar de la méthode de mesure de la similarité de I.C. Lerman (1981), à l'instar de la démarche classique dans les tests non paramétriques (ex. Fischer, Wilcoxon, etc.), nous définissons (Gras 1979, Gras et al. 1996 c) la mesure de qualité confirmatoire de la relation implicative $a \Rightarrow b$ à partir de l'invraisemblance de l'apparition, dans les données, du nombre de cas qui l'infirment, c'est-à-dire pour lesquels a est vérifié sans que b ne le soit. Ceci revient à comparer l'écart entre le contingent et le théorique si seul le hasard intervenait[5]. Mais, dans le cadre de l'analyse de données, c'est cet écart qui est pris en compte et non pas l'énoncé d'un rejet ou de l'admissibilité d'hypothèse nulle. Cette mesure est relativisée par le nombre de données vérifiant respectivement a et non b, circonstance dans laquelle l'implication est précisément mise en défaut. Elle quantifie "*l'étonnement"* de

[2] Ici, le mot « variable » désigne aussi bien une variable isolée en prémisse (ex. : « être blonde ») qu'une conjonction de variables isolées (ex. : « être blonde **et** avoir moins de 30 ans **et** habiter Paris »)

[3] « L'exception confirme la règle » nous dit l'adage populaire qui devrait être pris au sens où il n'y aurait pas d'exceptions s'il n'y avait pas de règle. Dans le contexte du raisonnement déductif, il faudrait bien sûr dire « L'exception infirme la règle. »

[4] Londeix H. (1983) *Approche génétique et différentielle du développement intellectuel*. Thèse de doctorat. Université de Bordeaux II.

[5] « … [en accord avec Jung] si la fréquence des coïncidences n'excède pas de façon significative la probabilité qu'on peut leur calculer en les attribuant au seul hasard à l'exclusion de relations causales cachées, nous n'avons certes aucune raison de supposer l'existence de telles relations. », Atlan H., (1986) *A tort et à raison. Intercritique de la science et du mythe*, Paris : Seuil.

l'expert devant le nombre invraisemblablement petit de contre-exemples sous l'hypothèse d'une indépendance entre les variables eu égard aux effectifs en jeu.

Précisons plus avant la modélisation. Un ensemble fini V de v variables, désignées par des lettres a, b, c, etc., est donné. Dans la situation paradigmatique classique, il s'agit des performances (réussite-échec) à des items d'un questionnaire de connaissances. A un ensemble fini E de n sujets désignés x, on associe, par abus d'écriture, les fonctions du type : $x \rightarrow a(x)$ où a(x)=1 (ou a(x)=vrai) si x satisfait ou possède le caractère a et 0 (ou a(x)=faux) sinon. En intelligence artificielle, on dira que x est un exemple ou une instance pour a si a(x)=1 et un contre-exemple dans le cas contraire.

La règle $a \Rightarrow b$ est logiquement vraie si pour tout x de l'ensemble E, b(x) n'est nul que dans le cas où a(x) l'est aussi, autrement dit si l'ensemble A des x pour lesquels a(x)=1 est contenu dans l'ensemble B des x pour lesquels b(x)=1. Cependant, cette inclusion stricte n'est qu'exceptionnellement observée dans les expériences réelles. Dans le cas d'un questionnaire de connaissances, on pourrait en effet observer quelques rares élèves réussissant un item a et ne réussissant pas l'item b, sans que ne soit contestée la *tendance* à réussir b quand on a réussi a. Relativement aux cardinaux de E (soit n), de A (soit n_a) et de B (soit n_b), c'est le poids des contre-exemples (soit $n_{a \wedge \bar{b}}$) qu'il faut donc prendre en compte pour accepter statistiquement de conserver ou non la **quasi-implication** ou **quasi-règle** $a \Rightarrow b$. Ainsi, c'est à partir de la dialectique entre les exemples et les contre-exemples que la règle apparaît comme le dépassement de la contradiction.

4 Formalisation de la quasi-règle implicative dans l'approche classique.

Pour formaliser cette quasi-règle, nous considérons, comme le fait I.C. Lerman pour la similarité, deux parties quelconques X et Y de E, choisies aléatoirement et indépendamment (absence de lien a priori entre ces deux parties) et de mêmes cardinaux respectifs que A et B. Soit $\overline{Y}$ et $\overline{B}$ les ensembles complémentaires respectifs de Y et de B dans E de même cardinal $n_{\bar{b}} = n - n_b$.

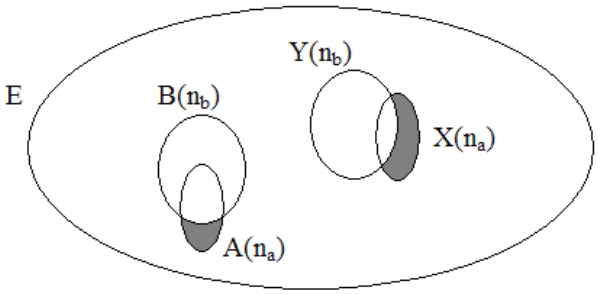


FIG. 1- *représentation par les diagrammes d'Euler*

Nous dirons alors :

Définition 1: la quasi-règle a $\Rightarrow$ b est *admissible au niveau de confiance* 1-α si et seulement si $\Pr\left[Card(X \cap \overline{Y}) \leq card(A \cap \overline{B})\right] \leq \alpha$]

Intuitivement et qualitativement, ceci signifie que la quasi-implication a $\Rightarrow$ b sera admissible à l'issue d'une expérience si le nombre d'individus de E la contredisant est invraisemblablement petit par rapport au nombre d'individus attendu sous une hypothèse d'absence de lien. Par exemple, si Card E = 100, Card A = 36, Card B = 50, alors $card(A \cap \overline{B}) = 3$ est "invraisemblablement petit" sous l'hypothèse d'une absence de lien entre a et b. On constate, en effet, que A est "presque" contenu dans B, alors que, sans liaison de A et B, on pourrait s'attendre à ce qu'environ la moitié des éléments de A soient aussi dans B.

Définition 2: On appelle intensité d'implication de la quasi-règle $a \Rightarrow b$, le nombre $\varphi(a,b) = 1 - \Pr\left[Card(X \cap \overline{Y}) \leq Card(A \cap \overline{B})\right]$ si $n_b \neq n$ et $\varphi(a,b) = 0$ si $n_b = n$

L'intensité d'implication est une valeur probabiliste, et non une fréquence, qui fonde la décision de retenir ou non une relation de quasi-implication entre les variables binaires a et b.

Cette modélisation de la quasi-implication est pertinente pour mesurer l'étonnement face au constat de la petitesse du nombre des contre-exemples en regard du nombre surprenant des instances de l'implication. Il s'agit d'une mesure de la qualité inductive et informative de l'implication. Par conséquent, si la règle est triviale, comme dans le cas où B est très grand ou coïncide avec E, cet étonnement devient petit. Nous démontrons (Gras R., 1996 c) d'ailleurs que cette trivialité se traduit par une intensité d'implication très faible, voire nulle : *Si, n_a étant fixé et A étant inclus dans B, n_b tend vers n (B "croît" vers E), alors $\varphi(a,b)$ tend vers 0*. C'est pourquoi nous définissons par « continuité »: $\varphi(a,b) = 0$ si $n_b = n$. De même, si $A \subset B$, $\varphi(a,b)$ peut être inférieure à 1 dans le cas où la confiance inductive, mesurée par l'étonnement statistique, est insuffisante.

5 Différents modèles pour évaluer l'intensité d'implication.

La détermination de l'intensité d'implication dépend du modèle retenu pour définir la loi de probabilité de la variable aléatoire « nombre de contre-exemples ».

5.1 Modèle de Poisson

Reprenant une idée abordée dans (Bodin et al., 1997), au lieu de considérer la situation statistique figée par les données du croisement sujets-variables, envisageons un processus temporel transactionnel discret, durant lequel, à certains instants, apparaît une transaction. Celle-ci représente, par exemple, tout aussi bien un sujet, une feuille d'enquête qu'une opération bancaire. Par exemple, si la transaction *s* satisfait (resp. ne satisfait pas) la variable *a* (par exemple un attribut) de *V*, nous l'instancions par 1 (resp. 0) à l'intersection de la ligne s et la colonne *a*. Les transactions peuvent être considérées comme tirées d'une population-mère supposée infinie. Le choix de cette approche dynamique, en harmonie avec la philosophie de l'ASI, puise sa justification dans l'éventail de situations illustratives où se

produisent des arrivées successives d'informations, de données, d'observations, capitalisées à un moment T fixé, dans un tableau de croisement E x V.

Dans la contingence, au bout d'un nombre *n* d'instants (donc *n*=T par convention), on a observé *n* transactions dont, parmi elles, n_a et n_b transactions de types respectifs *a* et *b*. Le nombre de contre-exemples à l'événement « si *a* alors *b* » est $n_{a\wedge\bar{b}}$ où $\bar{b}$ désigne l'événement « non *b* ».

Afin de mesurer la qualité de la quasi-règle $a \Rightarrow b$, comme nous l'avons déjà dit, nous construisons alors un modèle aléatoire en comparant ce nombre de contre-exemples à celui qu'aurait donné le seul hasard si les types *a* et *b* apparaissaient, de façon indépendante, au cours d'un processus respectant des hypothèses « raisonnables » que nous préciserons plus loin. Pour évaluer cette qualité, les événements aléatoires qui nous intéressent sont ceux où seraient réalisées les apparitions de *a* et de non *b* parmi les n transactions.

Notons A l'événement réalisant la variable *a*, soit [*a*=1], au cours du processus. De même notons B l'événement [*b*=1]. Du fait que dans la contingence, nous observons n_a fois *a* et n_b fois *b*, nous attribuons dans le modèle aléatoire à A (resp. B) la probabilité estimée par $\frac{n_a}{n}$ (resp. $\frac{n_b}{n}$). De plus, A et B devant être indépendants par hypothèse, la réalisation simultanée de A ([*a*=1]) et non B ([*b*=0]), lors d'une transaction aléatoire, y aura pour probabilité estimée $\frac{n_a}{n}.\frac{n_{\bar{b}}}{n}$. On pourrait alors ici s'arrêter sur le modèle binomial permettant de préciser la loi de la variable aléatoire « nombre de contre-exemples » qui pourrait s'imposer d'évidence. Nous y reviendrons plus loin.

Pour préciser et légitimer le modèle de processus d'extraction des transactions spécifiées, nous énonçons les hypothèses sémantiquement admissibles suivantes, relativement à la réalisation de l'événement : A et non B = [*a*=1 et *b*=0].:

- h1 : les temps d'attente successifs d'un événement [A et non B] sont des variables aléatoires indépendantes. Cette hypothèse est légitimée par l'indépendance a priori de A et B ;
- h2 : la loi du nombre d'événements survenant dans un intervalle de temps de durée T ne dépend que de T indépendamment de l'origine de temps ; ceci nécessite que le dépouillement des données soit régulier, ce qui est une moindre exigence ;
- h3 : deux tels événements ne peuvent arriver simultanément, ce qui est le cas du dépouillement des transactions qui est séquentiel

On démontre alors (Saporta, 2006) que le nombre d'événements, se produisant pendant une période de durée *n* fixée, suit une loi de Poisson de paramètre *c.n* où *c* est appelé cadence du processus d'apparitions de [*a*=1 et *b*=0] pendant l'unité de temps. Par suite, dans notre modèle, la cadence choisie est estimée par $c = \frac{n_a}{n}.\frac{n_{\bar{b}}}{n}$. Ainsi pour une durée de temps n, les apparitions de l'événement [A et non B] suivent une loi de Poisson de paramètre λ dont nous connaissons une estimation : $\lambda_{estimé} = \frac{n_a.n_{\bar{b}}}{n}$

Par suite $\forall s \in \{0,1,2,...,n\}$ Pr[Card($X \cap \overline{Y}$) = s]= $\frac{\lambda^s}{s!} e^{-\lambda}$ dans laquelle λ est remplacé par $\hat{\lambda} = \lambda_{\text{estimé}}$.

Remarquons que la nullité du nombre de contre-exemples à $a \Rightarrow b$ est équivalente à l'inclusion de X dans Y, en tant que parties aléatoires extraites de la population-mère supposée infinie.

En conséquence, la probabilité pour que le hasard conduise, sous l'hypothèse d'absence de lien a priori entre *a* et *b*, à au plus le nombre de contre-exemples observé, est estimée par :

$$\Pr\left[Card(X \cap \overline{Y}) \leq n_{a \wedge \overline{b}}\right] = \sum_{s=0}^{n_{a \wedge \overline{b}}} \frac{\hat{\lambda}^s}{s!} . e^{-\hat{\lambda}}$$

L'intensité de l'implication estimée, notée $\varphi(a,b)$ est alors :

$$\varphi(a,b) = 1 - \Pr\left[Card(X \cap \overline{Y}) \leq n_{a \wedge \overline{b}}\right] = 1 - \sum_{s=0}^{n_{a \wedge \overline{b}}} \frac{\hat{\lambda}^s}{s!} . e^{-\hat{\lambda}}$$

Définition 3: On appelle indice d'implication de la quasi-règle $a \Rightarrow b$, la variable aléatoire, notée $Q(a,\overline{b})$, déduite de la variable aléatoire $Card(X \cap \overline{Y})$ par centrage réduction.

Pour λ≥5 (λ paramètre estimé par $\frac{n_a n_{\overline{b}}}{n}$, rappelons-le), la variable « indice d'implication », notée : $Q(a,\overline{b}) = \dfrac{Card(X \cap \overline{Y}) - \frac{n_a n_{\overline{b}}}{n}}{\sqrt{\frac{n_a n_{\overline{b}}}{n}}}$ qui résulte du centrage-réduction de la variable de Poisson, Card(X ∩ $\overline{Y}$), peut être approchée par la variable gaussienne centrée réduite N(0 ;1).

Si nous considérons la **valeur empirique de l'indice** $q(a,\overline{b}) = \dfrac{n_{a \wedge \overline{b}} - \frac{n_a n_{\overline{b}}}{n}}{\sqrt{\frac{n_a n_{\overline{b}}}{n}}}$, alors l'intensité d'implication estimée de la quasi-règle $a \Rightarrow b$, est approximativement :

$$\varphi(a,b) = 1 - \Pr\left[Q(a,\overline{b}) \leq q(a,\overline{b})\right] = \frac{1}{\sqrt{2\pi}} \int_{q(a,\overline{b})}^{\infty} e^{-\frac{t^2}{2}} dt$$

Insistons sur le sens de cette intégrale. Elle représente la probabilité gaussienne pour que le nombre de transactions observées satisfaisant la quasi-règle $a \Rightarrow b$, soit supérieur à celui qui serait observable sous l'hypothèse d'indépendance de *a* et b. Autrement dit, $\Pr[Q(a,\overline{b}) \leq q(a,\overline{b})]$ est la **p-value** du test visant à réfuter l'hypothèse de l'indépendance de *a* et *b* au profit d'une relation de type quasi-implication

5.2 Modèle binomial

Examiner la qualité de la quasi-règle a⇒b, dans le cas où les variables sont binaires, revient à mesurer de façon équivalente celle de l'inclusion du sous-ensemble des transactions satisfaisant *a* dans le sous-ensemble des transactions satisfaisant *b*. Les contre-exemples relatifs à l'inclusion sont en effet les mêmes que ceux qui sont relatifs à l'implication exprimée par : « toute transaction satisfaisant a satisfait aussi b ». Dans cette optique ensembliste, dès que $n_a \leq n_b$, la qualité de la quasi-règle a⇒b, ne peut qu'être sémantiquement meilleure que celle de $b \Rightarrow a$. Nous supposerons donc, par la suite, que $n_a \leq n_b$ lors de l'étude de a⇒b. Dans ce cas, la population-mère est finie et *Card E = n.*

La modélisation binomiale fut chronologiquement la première adoptée (Gras 1979 chap. 2). Elle fut comparée à d'autres modélisations dans (Lerman et al. 1981). Rappelons brièvement en quoi consiste **le modèle binomial**. Avec les notations adoptées, *X* et *Y* sont deux sous-ensembles aléatoires, indépendamment choisis dans l'ensemble des parties de E, respectivement de mêmes cardinaux n_a et n_b que les sous-ensembles de réalisations de *a* et de *b*. La valeur observée $n_{a\wedge\bar{b}}$ peut être considérée comme la réalisation d'une variable aléatoire Card ($X \cap \bar{Y}$) qui représente le nombre aléatoire de contre-exemples à l'inclusion de *X* dans *Y*, contre-exemples observés au cours de *n* tirages successifs indépendants. De là, Card ($X \cap \bar{Y}$) peut être considérée comme une variable binomiale de paramètres *n* et π où π est elle-même estimée par $p = \frac{n_a}{n}\frac{n_{\bar{b}}}{n}$. Ainsi :

$$\Pr\left[Card(X \cap \bar{Y}) = k\right] = C_n^k \left(\frac{n_a n_{\bar{b}}}{n^2}\right)^k \left(1 - \frac{n_a n_{\bar{b}}}{n^2}\right)^{n-k}$$

La variable centrée réduite estimée $Q(a,\bar{b})$ admet alors comme réalisation :

$$q(a,\bar{b}) = \frac{n_{a\wedge\bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}\left(1 - \frac{n_a n_{\bar{b}}}{n^2}\right)}}$$

Comme précédemment, nous obtenons l'intensité d'implication empirique estimée :

$$\varphi(a,b) = 1 - \Pr\left[Q(a,\bar{b}) \leq q(a,\bar{b})\right] = 1 - \sum_0^{n_{a\wedge\bar{b}}} C_n^k \left(\frac{n_a n_{\bar{b}}}{n^2}\right)^k \left(1 - \frac{n_a n_{\bar{b}}}{n^2}\right)^{n-k}$$

La loi de probabilité de $Q(a,\bar{b})$ peut être approchée par celle de la loi de Laplace-Gauss centrée réduite N(0,1). Généralement, l'intensité calculée dans le modèle de Poisson est plus « sévère » que l'intensité découlant du modèle binomial au sens où $\varphi(a,b)_{Poisson} \leq \varphi(a,b)_{binomiale}$

Remarque : Nous pouvons noter que l'indice d'implication est nul si et seulement les deux variables a et b sont indépendantes. En effet

$$q(a,\bar{b}) = \frac{n_{a\wedge\bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}(1-\frac{n_a n_{\bar{b}}}{n^2})}} = 0 \Leftrightarrow n_{a\wedge\bar{b}} - \frac{n_a n_{\bar{b}}}{n} = 0$$

$$q(a,\bar{b}) = 0 \Leftrightarrow n_{a\wedge\bar{b}} = \frac{n_a n_{\bar{b}}}{n} \quad \text{ou encore} \quad q(a,\bar{b}) = 0 \Leftrightarrow \frac{n_{a\wedge\bar{b}}}{n} = \frac{n_a}{n}\frac{n_{\bar{b}}}{n}$$

Cette dernière relation traduit la propriété d'indépendance statistique.

5.3 Modèle hypergéométrique.

Rappelons brièvement la 3ème modélisation proposée dans (Lerman et al. 1981) et (Gras et al 1996 c). Nous reprenons la même démarche : A et B sont les parties de E représentant les individus satisfaisant respectivement a et b et dont les cardinaux sont card (A)=n_a et card (B)=n_b. Puis considérons, deux parties aléatoires indépendantes X et Y telles que card (X)=n_a et card (Y)=n_b. La variable aléatoire Card ($A\cap\overline{Y}$) représente le nombre aléatoire d'éléments de *E* qui, étant dans A ne sont pas dans Y. Cette variable suit une loi hypergéométrique et l'on a pour tout $k \leq n_a$:

$$\Pr\left[Card(A\cap\overline{Y}) = k\right] = \frac{C_{n_a}^{k} C_{n-n_a}^{n-n_b-k}}{C_n^{n-n_b}} = \frac{n_a! n_{\bar{a}}! n_b! n_{\bar{b}}!}{k! n! (n_a-k)! (n_{\bar{b}}-k)! (n_b-n_a+k)!} =$$

$$\frac{C_{n-n_b}^{k} C_{n_b}^{n_a-k}}{C_n^{n_a}} = \Pr\left[Card(X\cap\overline{B}) = k\right]$$

Ceci montre, en échangeant le rôle de *a* et *b* que l'indice d'implication empirique $Q(a,\bar{b})$ correspondant à la quasi-règle a⇒b, est le même que celui correspondant à la réciproque soit $Q(b,\bar{a})$**.** Nous obtenons ainsi la même intensité pour la quasi-règle a⇒b et pour la quasi-règle réciproque b⇒a.

5.4 Choix des modèles pour évaluer l'intensité d'implication.

Si la modélisation binomiale reste compatible avec la sémantique de l'implication, relation binaire non symétrique, il n'en est plus de même pour la modélisation hypergéométrique puisqu'elle ne distingue pas la qualité d'une quasi-règle de celle de sa réciproque et présente un faible caractère pragmatique. En conséquence, nous ne retiendrons que le modèle de Poisson et le modèle binomial comme modèles adaptés à la sémantique de l'implication entre variables binaires.

La coexistence légitimée de trois modélisations différentes de notre problématique de mesure de qualité d'une quasi-règle n'est pas incohérente : elle tient au mode de prise en compte du tirage une à une de transactions (loi de Poisson) ou d'ensembles de transactions groupées (loi binomiale ou loi hypergéométrique). Par ailleurs, nous savons que, lorsque le nombre total de transactions devient très grand, les trois modèles convergent vers le même modèle gaussien. Dans (Lallich S. et al. 2005), on trouve, à titre de généralisation, une paramétrisation des trois indices obtenus par ces modélisations, qui permet d'évaluer l'intérêt des règles obtenues en les comparant à un seuil donné.

6 Quelques propriétés de l'indice d'implication et de l'intensité d'implication.

6.1 Stabilité de l'indice d'implication et de l'intensité d'implication

Le problème de la sensibilité aux faibles perturbations des paramètres en jeu, donc de la stabilité des indices de mesure de qualité des règles d'association se pose dès lors que les données sont susceptibles d'être bruitées. Trois méthodes nous semblent appropriées dans le but d'examiner cette sensibilité des règles, en particulier celles de la forme $a \Rightarrow b$ où a et b sont des variables observées sur un ensemble de sujets :

1- la simulation consistant à partir de fichiers plus ou moins artificiels à travers lesquels sont modifiés les paramètres intervenant dans la définition des indices (Gras et al. 2004) ;

2- la méthode du bootstrap consistant à effectuer des changements de certaines valeurs des paramètres, tout en conservant constantes certaines d'entre elles dont en particulier l'effectif de la population des individus ;

3- une méthode mathématique consistant à étudier, par l'analyse, les variations des paramètres en examinant leurs dérivées partielles et donc le gradient de l'indice global. (Gras 2005), (Lenca et al. 2006) et (Vaillant et al. 2006).

C'est cette dernière méthode que nous retenons ici. Nous porterons notre attention de façon privilégiée sur l'indice d'implication à la base de l'ASI et comparerons les résultats obtenus à ceux dérivant d'autres indices retenus pour des mesures de qualité de règles.

6.2 Analyse des variations de l'indice d'implication en fonction des cardinaux

Étudier la stabilité de l'indice d'implication q, revient à examiner ses petites variations au voisinage des 4 valeurs entières observées (n, n_a, n_b, $n_{a\wedge\bar{b}}$). Pour ce faire, il est possible d'effectuer différentes simulations en croisant ces 4 variables entières dont q dépend (Fleury, 1996, Gras et al., 2004). Mais, considérons ces variables comme nombres réels et q comme une fonction continûment différentiable par rapport à ces variables contraintes à respecter les inégalités : $0 \le n_a \le n_b$ et $n_{a\wedge\bar{b}} \le \inf\{n_a, n_b\}$ et $\sup\{n_a, n_b\} \le n$. Il suffit alors d'examiner la différentielle de q par rapport à ces variables et d'en conserver la restriction aux valeurs entières des paramètres de la relation $a \Rightarrow b$. La différentielle de q s'exprime de la façon suivante :

$$dq = \frac{\partial q}{\partial n} dn + \frac{\partial q}{\partial n_a} dn_a + \frac{\partial q}{\partial n_b} dn_b + \frac{\partial q}{\partial n_{a\wedge\bar{b}}} dn_{a\wedge\bar{b}}$$

Si l'on veut étudier comment varie q en fonction de $n_{\bar{b}}$, il suffit de remplacer n_b par $n-n_b$ et donc changer le signe de la dérivée de n_b dans la dérivée partielle. En fait, l'intérêt de cette différentielle réside dans l'estimation de l'accroissement (positif ou négatif) de q que nous notons Δq par rapport aux variations respectives Δn, Δn_a, Δn_b, $\Delta n_{\bar{b}}$ et $\Delta n_{a\wedge\bar{b}}$.

Si nous examinons le cas où seuls varient n_b et $n_{a\wedge\bar{b}}$, c'est à dire où les dérivées partielles de n et n_a sont nulles, on obtient alors :

$$\frac{\partial q}{\partial n_b} = \frac{1}{2} n_{a\wedge\bar{b}} (\frac{n_a}{n})^{-\frac{1}{2}} (n-n_b)^{-\frac{3}{2}} + \frac{1}{2} (\frac{n_a}{n})^{\frac{1}{2}} (n-n_b)^{-\frac{1}{2}} > 0$$

$$\frac{\partial q}{\partial n_{a\wedge\bar{b}}} = \frac{1}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}} = \frac{1}{\sqrt{\frac{n_a (n-n_b)}{n}}} > 0$$. Ainsi, si les accroissements Δn_b et $\Delta n_{a\wedge\bar{b}}$ sont positifs, l'accroissement de $q(a,\bar{b})$ est également positif. Ceci s'interprète ainsi : si le nombre d'exemples de b et celui des contre-exemples de l'implication augmentent alors l'intensité d'implication diminue pour n et n_a constants. Autrement dit, cette intensité d'implication est maximum aux valeurs observées n_b et $n_{a\wedge\bar{b}}$ et minimum aux valeurs $n_b+\Delta n_b$. et $n_{a\wedge\bar{b}}+\Delta n_{a\wedge\bar{b}}$.

Si nous examinons le cas où seul n_a varie, nous obtenons la dérivée partielle de q par rapport à n_a est :

$$\frac{\partial q}{\partial n_a} = -\frac{1}{2\sqrt{n_{\bar{b}}/n}} \cdot \frac{n_{a\wedge\bar{b}}}{n} \cdot (\frac{n}{n_a})^{\frac{3}{2}} - \frac{1}{2}\sqrt{\frac{n_{\bar{b}}}{n_a}} < 0$$

Ainsi, sur [0, n_b], la fonction indice d'implication $q(a,\bar{b})$ est toujours décroissante par rapport à n_a et est donc minimum pour $n_a = n_b$. Par suite, l'intensité d'implication y est croissante et maximum pour $n_a = n_b$

6.3 Analyse des variations de l'indice d'implication en fonction des fréquences

On examine maintenant les variations de q en fonction des fréquences relatives des variables précédentes où le référentiel a pour cardinal n. Ainsi, on note $f_i = \frac{n_i}{n}$ chacune des fréquences des variables respectives n, n_a, $n_{\bar{b}}$ (que nous privilégions par rapport à n_b pour des raisons de calculs) et $n_{a\wedge\bar{b}}$. Dans ces conditions $q(a,\bar{b})$ s'écrit alors :

$$q(a,\bar{b}) = \sqrt{n}\frac{(f_{a\wedge\bar{b}} - f_a f_{\bar{b}})}{\sqrt{f_a f_{\bar{b}}}} = \sqrt{n}\frac{f_{a\wedge\bar{b}}}{\sqrt{f_a f_{\bar{b}}}} - \sqrt{n}\sqrt{f_a f_{\bar{b}}}$$

On étudie alors la stabilité à partir des dérivées partielles de q par rapport aux 4 variables fréquentielles :

$$dq = \frac{\partial q}{\partial n} dn + \frac{\partial q}{\partial f_{a\wedge\bar{b}}} df_{a\wedge\bar{b}} + \frac{\partial q}{\partial f_a} df_a + \frac{\partial q}{\partial f_{\bar{b}}} df_{\bar{b}} = \vec{grad}\, q. \begin{bmatrix} dn \\ df_{a\wedge\bar{b}} \\ df_a \\ df_{\bar{b}} \end{bmatrix}$$

Or les dérivées partielles respectives sont :

$$\frac{\partial q}{\partial n} = \frac{1}{2\sqrt{n}} \frac{f_{a\wedge\bar{b}} - f_a f_{\bar{b}}}{\sqrt{f_a f_{\bar{b}}}} = \frac{1}{2} \frac{f_{a\wedge\bar{b}} - f_a f_{\bar{b}}}{\sqrt{n f_a f_{\bar{b}}}} \; ;$$

$$\frac{\partial q}{\partial f_{a\wedge\bar{b}}} = \sqrt{n} \frac{1}{\sqrt{f_a f_{\bar{b}}}} > 0 \; ;$$

$$\frac{\partial q}{\partial f_a} = -\frac{\sqrt{n}}{2} \left[\frac{f_{a\wedge\bar{b}}}{\sqrt{f_{\bar{b}}}} f_a^{-\frac{3}{2}} + \sqrt{f_{\bar{b}}} f_a^{-\frac{1}{2}} \right] < 0$$

$$\frac{\partial q}{\partial f_{\bar{b}}} = -\frac{\sqrt{n}}{2} \left[\frac{f_{a\wedge\bar{b}}}{\sqrt{f_{\bar{a}}}} f_{\bar{b}}^{-\frac{3}{2}} + \sqrt{f_a} f_{\bar{b}}^{-\frac{1}{2}} \right] < 0$$

Nous pouvons remarquer qu'en calculant $\frac{\partial \mathrm{q}}{\partial \mathrm{f_b}}$ au lieu de $\frac{\partial \mathrm{q}}{\partial \mathrm{f_{\bar{b}}}}$, on constate que cette dérivée partielle est alors positive. En effet : $f_b = 1 - f_{\bar{b}}$ et $\frac{\partial q}{\partial f_b} = \frac{\partial q}{\partial f_{\bar{b}}} \frac{\partial f_{\bar{b}}}{\partial f_b} = -\frac{\partial q}{\partial f_{\bar{b}}}$

Par ailleurs, à n et n_a constants, la vitesse d'accroissement de q (en valeur absolue) quand s'accroît le nombre de contre-exemples $a \wedge \bar{b}$, est inversement proportionnelle à celle de la racine carrée de $n_{\bar{b}}$. Autrement dit : si $n_{\bar{b}}$ décroît, par exemple deux fois plus petit, cette vitesse est accélérée et multipliée par 2. Conséquence, l'intensité d'implication diminue et la qualité de l'implication devient moins bonne.

6.4 Examen d'autres indices d'implication

Contrairement à l'indice d'implication q de base et à l'intensité d'implication qui mesure la qualité à travers une probabilité, d'autres indices parmi les plus courants se veulent eux-mêmes directement des mesures de qualité . Nous examinons leurs sensibilités respectives aux variations des paramètres retenus dans la définition de ces indices. Nous conservons les notations adoptées et choisissons des indices qui sont rappelés dans (Lenca et al., 2004).

L'indice Lift

Il s'exprime par : $l = \frac{n.n_{a\wedge b}}{n_a n_b}$. Cette expression peut encore s'écrire pour mettre en évidence le nombre de contre-exemples : $l = \frac{n(n_a - n_{a\wedge\bar{b}})}{n_a.n_b}$.

Pour étudier la sensibilité de l aux variations des paramètres, nous formons :

$$\frac{\partial l}{\partial n_{a\wedge\bar{b}}} = -\frac{n}{n_a n_b}$$

Ainsi, la variation de l'indice Lift est indépendante de celle du nombre de contre-exemples. C'est une constante qui ne dépend que des variations des occurrences de a et de b. L'indice Lift l décroît donc lorsque le nombre de contre-exemples croît, ce qui

sémantiquement, est acceptable mais la vitesse de décroissance ne dépend pas de la vitesse de croissance de $n_{a \wedge \bar{b}}$.

L'indice m multiplicateur de cote

Cet indice d'implication s'exprime ainsi : $m = \frac{n_a - n_{a \wedge \bar{b}}}{n_b n_{a \wedge \bar{b}}} n_{\bar{b}}$ (Lallich et al, 2004). Remarquons qu'en étant indépendant de n, il n'a pas un sens statistique aussi intéressant. Sa dérivée partielle par rapport au nombre de contre-exemples est :

$$\frac{\partial m}{\partial n_{a \wedge \bar{b}}} = -\frac{n_a n_{\bar{b}}}{n_b} (\frac{1}{n_{a \wedge \bar{b}}})^2 .$$

L'indice m multiplicateur de cote décroît donc lorsque $n_{a \wedge \bar{b}}$ croît et la vitesse de décroissance est même plus rapide qu'avec l'indice Lift et qu'avec l'indice d'implication q de base dans l'intensité d'implication. Il ne résiste pas à l'instabilité du nombre de contre-exemples.

L'indice c, confiance

Cet indice c est le plus connu et, historiquement, après celui de J. Lovinger, le plus utilisé grâce à la caisse de résonance dont dispose une publication anglo-saxonne (Agrawal et al. 1993). Il est à l'origine de plusieurs autres indices communément employés qui n'en sont que des variantes satisfaisant telle ou telle exigence sémantique. De plus, il est simple et s'interprète aisément et immédiatement.

$$c = \frac{n_{a \wedge b}}{n_a} = 1 - \frac{n_{a \wedge \bar{b}}}{n_a}$$

Cet indice c s'interprète comme une fréquence conditionnelle des exemples de b quand a est connu. La sensibilité de cet indice aux variations des occurrences des contre-exemples se lit avec la dérivée partielle :

$$\frac{\partial c}{\partial n_{a \wedge \bar{b}}} = -\frac{1}{n_a} < 0$$

Par conséquent, la confiance c croît quand $n_{a \wedge \bar{b}}$ décroît ce qui est sémantiquement acceptable, mais la vitesse de variation est constante, indépendante de la vitesse de décroissance de cette quantité ainsi que des variations de n et de n_b. Le gradient de c ne s'exprime que par rapport à $n_{a \wedge \bar{b}}$ et à n_a. Ceci peut apparaître comme une restriction du rôle des paramètres dans l'expression de la sensibilité de l'indice.

6.5 Coefficient de corrélation linéaire et indice d'implication

La quasi-implication définie par l'indice d'implication $q(a, \bar{b})$ non symétrique ne coïncide pas avec le coefficient de corrélation $\rho(a,b)$ qui est symétrique et qui rend compte d'une liaison linéaire entre les variables a et b. En effet, nous démontrons la proposition suivante :

Proposition 1 si $\rho(a,b) \neq 0$ alors $\frac{q(a,\bar{b})}{\rho(a,b)} = -\sqrt{\frac{n_b n_{\bar{a}}}{n}}$

En effet, d'une part, $q(a,\bar{b}) = \frac{n_{a\wedge\bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}} = \frac{n_a n_b - n n_{a\wedge b}}{\sqrt{n n_a n_{\bar{b}}}}$ et, d'autre part,

$\rho(a,b) = \frac{n n_{a\wedge b} - n_a n_b}{\sqrt{n_a n_b n_{\bar{a}} n_{\bar{b}}}} = \frac{n_{a\wedge b} n_{\bar{a}\wedge\bar{b}} - n_{a\wedge\bar{b}} n_{\bar{a}\wedge b}}{\sqrt{n_a n_b n_{\bar{a}} n_{\bar{b}}}}$ d'où la relation annoncée

Proposition 2 $q(a,\bar{b}) = 0 \Leftrightarrow \rho(a,b) = 0$ et $\rho(a,b) \geq 0 \Leftrightarrow \varphi(a,b) \geq 0{,}5$.

Ceci signifie que quasi-implication et corrélation linéaire vont plutôt "*dans le même sens*". Cependant, on peut observer une croissance de l'implication en même temps qu'une décroissance de la corrélation. Ce qui montre bien, qu'outre la dépendance aux effectifs n, n_a et n_b, l'expression du rapport $\frac{\rho}{q}$ indique la non-coïncidence des deux concepts et par conséquent une différence dans le sens de l'information apportée. L'ASI n'est pas une étude de la dépendance au sens statistique habituel où elle est essentiellement symétrique.

	b	$\bar{b}$	marge
a	82	18	100
$\bar{a}$	45	55	100
marge	127	73	4000

TAB. 1 – *exemple n°1*

L'indice d'implication $q(a,\bar{b}) = -3{,}06$

L'intensité d'implication est $\varphi(a,b) = 0{,}9994$

Le coefficient de corrélation linéaire $\rho(a,b)=0{,}3842$

	b	$\bar{b}$	marge
a	78	22	100
$\bar{a}$	49	51	100
marge	127	73	4000

TAB. 2 – *exemple n°2*

L'indice d'implication $q(a,\bar{b}) = -2{,}4000$

L'intensité d'implication est $\varphi(a,b) = 0{,}9931$

Le coefficient de corrélation linéaire $\rho(a,b)=0{,}3011$

On constate que, d'une part, a et b sont moins corrélées dans l'exemple 1 que dans le second, mais que, d'autre part l'intensité d'implication est plus forte dans le premier que dans le second cas..

6.6 Mesure du χ^2 d'indépendance et indice d'implication

La pratique fréquente du test d'indépendance à partir de tableaux 2x2 par la méthode du χ^2 nous conduit à montrer les apports respectifs des concepts de χ^2 d'indépendance et d'indice d'implication.

Dans la théorie des tests inférentiels, la mesure du χ^2 sert d'appui au raisonnement pour rejeter l'hypothèse d'indépendance entre deux variables a et b, à un seuil déterminé à l'avance. Dans la théorie de l'ASI, l'indice d'implication permet, non seulement d'assurer ce rejet, mais aussi, d'expliciter le **sens de la relation de dépendance**. Pragmatiquement, questionnés sur l'implication de la variable a sur b, certains chercheurs utilisent la mesure du χ^2 pour étudier et rejeter l'indépendance, puis constatant la faiblesse numérique de $n_{a \wedge \bar{b}}$, à partir du tableau de croisement de a et b, décident de retenir que a implique b. Or, par cette stratégie décisionnelle qualitative, prise à partir de nombres bruts, même à bon escient, ils négligent le rôle des valeurs relatives des nombres du tableau et ne peuvent, en outre, énoncer le niveau de qualité de l'implication conjecturée.

La proposition suivante met en évidence la relation fonctionnelle entre les deux concepts en jeu

Proposition 3 Considérons la mesure χ^2 et l'indice d'implication $q(a,\bar{b})$ entre les variables binaires a et b alors $\dfrac{\chi^2}{q(a,\bar{b})^2} = \dfrac{n^2}{n_a n_{\bar{b}}}$

En effet, $q(a,\bar{b}) = \dfrac{n_{a\wedge\bar{b}} - \dfrac{n_a n_{\bar{b}}}{n}}{\sqrt{\dfrac{n_a n_{\bar{b}}}{n}}} = \dfrac{n_a n_b - n n_{a\wedge b}}{\sqrt{n n_a n_{\bar{b}}}}$ et $\chi^2 = \dfrac{n(n_{a\wedge b} n_{\bar{a}\wedge\bar{b}} - n_{\bar{a}\wedge b} n_{a\wedge\bar{b}})^2}{n_a n_b n_{\bar{a}} n_{\bar{b}}}$

La démonstration est directe dès que l'on remarque que $\chi^2 = n\rho^2$. On constate ainsi que les deux concepts χ^2 et indice d'implication ne se superposent pas, ce qui, compte tenu de leur définition analytique n'a pas lieu de surprendre. Nous pouvons interpréter la valeur de l'indice d'implication $q(a,\bar{b})$ comme la contribution absolue de la case $(a,\bar{b})$ du tableau 2x2 croisant les deux variables binaires a et b au χ^2.

A. Totohasina (1992) a étudié de plus près les relations entre le χ^2 d'indépendance (ici à 1 degré de liberté) et l'indice d'implication. En particulier, les valeurs limites qui permettent de réfuter l'indépendance mutuelle entre deux variables, dans le cas où q est négatif, c'est-à-dire lorsque $q(a,\bar{b}) = -\sqrt{\dfrac{n_a n_{\bar{b}}}{n^2}\chi^2}$ peuvent éventuellement servir à définir les valeurs limites d'acceptabilité de la règle de quasi-implication de a sur b.

6.7 Mesure du χ^2 de Mac Nemar et indice d'implication

Si nous voulons comparer deux séries successives de données binaires de type présence-absence ou échec-réussite relevées sur le même échantillon d'individus comme cela est le cas

en ASI, nous pouvons aussi utiliser le test du χ^2 de **Mac Nemar**. Comme nous l'avons déjà présenté, l'information est alors résumée dans un tableau 2x2 dont nous donnons un exemple ci-dessous qui sera repris plus tard (Partie 2 Chap. 10). Pour rester congruent au mode de présentation des recherches de liens génériquement notés a⇒b dans le contexte ASI qui présuppose que N(a)≤N(b), nous représentons systématiquement la variable binaire a en ligne et la variable binaire b en colonne.

		Variable b = Épreuve Finale		
		1 = Réussite	0 = Échec	**Total**
Variable a = Épreuve Initiale	1 = Réussite	56	6	62
	0 = Échec	14	26	40
	Total	70	32	102

TAB. 3- *Tableau de contingence du type AVANT/APRES*

La question que nous nous posons alors est de savoir si les fréquences de réussite aux deux épreuves sont significativement différentes ou non.

L'idée de Mac Nemar pour étudier ce type de lien entre les deux épreuves est qu'il est plus pertinent de ne prendre en compte que les discordances entre les deux épreuves. Dans le tableau ci-dessus, ce sont les deux effectifs 14 et 6 correspondant aux couples (A_Echec, B_Réussite) et (A_Réussite, B_Echec) qui sont considérés comme des informations majeures. Cette idée n'est pas rendue par le test du χ^2 d'indépendance que nous avons déjà évoqué précédemment (Partie 1 Chap. 1-6.6) en établissant la relation algébrique entre l'indice d'implication et la mesure du χ^2.

Si nous nous remettons dans le contexte de l'ASI, le tableau de référence est donc celui-ci :

		Variable b		
		1	0	**Total**
Variable a	1	$n(a \wedge b)$	$n(a \wedge \bar{b})$	$n(a)$
	0	$n(\bar{a} \wedge b)$	$n(\bar{a} \wedge \bar{b})$	$n(\bar{a})$
	Total	$n(b)$	$n(\bar{b})$	n

TAB 4 - *Tableau de contingence avec les notations ASI*

Dans l'hypothèse d'une équivalence entre les deux épreuves, la fréquence de ceux qui sont passés de l'état 1 à l'état 0 parmi ceux qui ont changé d'état est égale à la fréquence de ceux qui sont passés de l'état 0 à l'état 1 parmi ceux qui ont changé d'état, c'est à dire égale à 0,5. D'une certaine manière, cela revient à comparer une fréquence observée à une fréquence théorique de 0,5.

Mac Nemar a montré qu'il suffisait de prendre comme indice, la mesure suivante que nous nommerons χ^2 de Mac Nemar, $\chi^2_{MacNemar} = \frac{(n(\bar{a} \wedge b) - n(a \wedge \bar{b}))^2}{n(\bar{a} \wedge b) + n(a \wedge \bar{b})}$ dont la loi de probabilité est approximativement celle de la variable de Pearson χ^2 de degré de liberté ddl=1.

Nous ne chercherons pas ici à expliciter une relation algébrique en $Q(a,\bar{b})$ et $\chi^2_{MacNemar}$.

Dans le cas présenté, nous calculons la valeur empirique comme suit $\chi^2_{MacNemar} = \frac{(14-6)^2}{14+6} = \frac{8^2}{20} = 3{,}2$ et nous la confrontons à la valeur critique au niveau de risque α. Si nous choisissons un niveau de risque de 0,05, la valeur critique est alors de 3,84. Comme 3,2<3,84, nous ne rejetons pas l'hypothèse d'équivalence des deux épreuves que nous considérons comme telle avec un risque de 2ème espèce β inconnu.

En résumé les 4 étapes de la démarche de ce test sont les suivantes :

- Étape 1 : formulation des hypothèses :

H_0 : symétrie des changements d'état entre les deux épreuves
H_1 : non-symétrie des changements d'état entre les deux épreuves

- Étape 2 : calcul de la valeur empirique du χ^2 (Mac Nemar)
- Étape 3 : lecture de la valeur critique dans la table du χ^2 de Pearson de ddl=1 pour un risque α donné
- Étape 4 : décision statistique rejet ou non rejet de H_0

Si nous revenons à la perspective de recherche de lien par rejet de l'indépendance en appliquant le test du χ^2 d'indépendance, nous trouvons une valeur empirique de 34,56 qui est très largement supérieure à la valeur critique de 3,84 pour un niveau de risque α=0,05 et même à la valeur critique 6,63 pour un niveau de risque α=0,01. Au sens du test du χ^2 d'indépendance, il existe donc un lien fort entre les deux variables.

Si nous nous plaçons dans la perspective de recherche de lien au sens de l'ASI, le calcul de l'intensité d'implication $\varphi_P(a,b)$ avec le modèle de Poisson et le calcul de l'intensité d'implication $\varphi_{BIN}(a,b)$ avec le modèle binomial

	χ^2 =34,56	χ^2_{MC} =3,2		Intensités d'implication
a \ b	b=1	b=0		$\varphi_P(a,b)$
a=1	56	6	62	0,9996
a=0	14	26	40	$\varphi_{BIN}(a,b)$
	70	32	102	0,9998

TAB. 4- *analyse selon les trois perspectives:* χ2 d'indépendance, χ2 Mac Nemar, ASI,

Les valeurs qui figurent dans le tableau ci-dessus, indiquent un niveau de confiance en l'implication statistique (a)⇒(b) supérieur à 0,99.

Face à ce qui semble paradoxal dans la mesure où le même tableau de contingence est susceptible d'être interprété de manière contradictoire, il y a tout lieu de considérer les logiques à l'œuvre dans ces trois approches : ASI, χ^2 de Mac Nemar, χ^2 d'indépendance.

Comme nous avons pu le voir au travers des propos tenus tout au long de ce qui précède, le raisonnement s'appuie sur un point de vue soutenu par I.-C. Lerman (1992) appliqué à l'étude d'une certaine relation de dépendance orientée entre des variables descriptives. Ce point de vue oppose la logique des tests statistiques, comme celui dit du χ^2 d'indépendance

ou encore celui du χ^2 de Mac Nemar, à celle des méthodes classificatoires de la manière suivante : pour les premiers, dit I.-C. Lerman, « relativement à l'existence d'un lien, on a FAUX, jusqu'à preuve du contraire » par le rejet de l'hypothèse nulle ; pour les secondes, « pour l'optique des données, on a VRAI, jusqu'à preuve du contraire », c'est-à-dire vrai selon une certaine échelle de probabilité du lien.

6.8 Indice de similarité et indice d'implication

Étudions maintenant la **relation** qui ne peut manquer d'exister entre **l'indice de similarité de I.-C. Lerman et l'indice d'implication** tel que nous le définissons dans le modèle de Poisson. Nous rappelons que l'indice de similarité poissonnien, défini sous la condition de la vraisemblance du lien tout comme l'indice d'implication, est donné par la formule :

$$s(a,b) = \frac{n_{a \wedge b} - \frac{n_a n_b}{n}}{\sqrt{\frac{n_a n_b}{n}}}$$

Proposition 4 : L'indice de similarité et l'indice d'implication sont liés par la relation suivante :

$$\frac{q(a,\bar{b})}{s(a,b)} = -\sqrt{\frac{n_b}{n_{\bar{b}}}} = -\sqrt{\frac{n_b}{n - n_b}}$$

On peut note que ce rapport ne dépend que de la réalisation de la variable b, de la réalisation concomitante de son contraire et non pas de celle de a. Par ailleurs nous pouvons aussi interpréter $s(a,b) = \frac{n_{a \wedge b} - \frac{n_a n_b}{n}}{\sqrt{\frac{n_a n_b}{n}}} = q(a,b)$ comme étant l'indice d'implication de $a \Rightarrow \bar{b}$. Ainsi, la similarité et l'implication vont, comme la corrélation et la dépendance, plutôt dans le même sens mais, bien entendu, ne coïncident pas. En effet, voici un exemple où la similarité ne change pas alors que l'implication varie très sensiblement:

	b	$\bar{b}$	marge
a	10	0	10
$\bar{a}$	70	20	90
marge	80	20	100

TAB. 5 – *exemple n°1*

L'indice d'implication $q(a,\bar{b})$ = -1,414

L'intensité d'implication φ(a,b)=0,864

L'indice de similarité est s(a,b) = 0,707

	b	$\bar{b}$	marge
a	3	7	10
$\bar{a}$	17	73	90
marge	20	80	100

TAB. 6– *exemple n°1*

L'indice d'implication $q(a,\bar{b})$ = -3,54

L'intensité d'implication φ(a,b)=0,547

L'indice de similarité est s(a,b) = 0,707

6.9 Comparaison avec d'autres approches de l'implication statistique.

Approche de J. Loevinger (1947)

L'approche de J. Loevinger fonde l'analyse de la quasi-implication de a sur b sur l'indice H qui prend ses valeurs sur tout]-∞,1] :

$$H(a,b) = 1 - \frac{n_{a\wedge\bar{b}}}{\frac{n_a n_{\bar{b}}}{n}} = 1 - \frac{n n_{a\wedge\bar{b}}}{n_a n_{\bar{b}}} = \frac{\frac{n_a n_{\bar{b}}}{n} - n_{a\wedge\bar{b}}}{\frac{n_a n_{\bar{b}}}{n}}$$

Si H(a,b) est *assez proche* de 1, la quasi-implication est *presque satisfaite*. Rappelons que cet indice, assez naturel il est vrai, avait été dans un premier temps "redécouvert" par R. Gras en 1978 à la suite de tâtonnements numériques, comme il l'a été également par A. Bodin en 1985 sous la forme suivante : notons x = P[B/A], probabilité conditionnelle de B sachant A et p = P[B]. Alors, H(a,b)= $\frac{x-p}{1-p}$. Mais cet indice présente l'inconvénient, en ne se référant pas à une échelle de probabilité, de ne pas fournir de seuil de vraisemblance et d'être invariant dans toute dilatation de E, A, B et A∩nonB.

Proposition 5 : L'indice H de Loevinger et l'indice d'implication sont liés par la relation suivante :

$$\frac{q(a,\overline{b})}{H(a,b)} = -\sqrt{\frac{n_a n_{\bar{b}}}{n}}$$

Dans l'approche de J. Pearl (1988), de S. Acid (Acid et *al*., 1991), de A. Gammerman et Z. Luo (Gammerman et Luo, 1991), c'est l'écart entre la distribution conjointe de a et b, et non pas celle de a et de non-b, et la distribution produit qui tient lieu de critère de comparaison. Cet écart est évalué par une expression de la forme :

$$\left|\Pr ob[A \cap \overline{B}] - \Pr ob[A]\Pr ob[\overline{B}]\right|$$

En ce qui concerne le système GENRED (Ralambrodrainy 1991) conçu dans une perspective de génération de règles d'inférence en intelligence artificielle, il est considéré tout simplement qu'une règle est pertinente dès lors que pour deux seuils α et β donnés par

l'utilisateur, le nombre de contre-exemples $card[A \cap \overline{B}]$ et celui des exemples $card[A \cap B]$ vérifient les conditions $card[A \cap \overline{B}] \leq \alpha$ et $card[A \cap B] \geq \beta$

Une autre façon comparable d'aborder cette question est (Sebag et Schoenauer, 1991) est de ne retenir qu'une règle au seuil α par une condition sur le rapport entre le nombre d'exemples et celui des contre-exemples : $\frac{card[A \cap B]}{card[A \cap \overline{B}]} \geq \alpha$ ou encore au travers d'une relation équivalente dans son principe :

$$\frac{card[A \cap B] - card[A \cap \overline{B}]}{card[A \cap B]} \geq \alpha$$

Comme nous l'avons déjà évoqué, la probabilité conditionnelle P[B/A] est aussi fréquemment utilisée comme indice de référence pour juger la plausibilité de la règle a=>b (Schektman, Trejos et Troupe, 1992) (Diday et Menessier, 1991). Par exemple, dans le domaine de l'apprentissage dans les bases de connaissances (Ganascia 1991), l'incertitude sur l'implication a=>b est évaluée par l'indice : $2\Pr ob[B|A] - 1$ et s'applique même aux classes de variables. Parmi les inconvénients, notons que ce dernier indice ne sépare pas, numériquement, deux implications dont l'une serait triviale et l'autre hautement informative.

Examinons deux situations-limites qui nous semblent probantes.

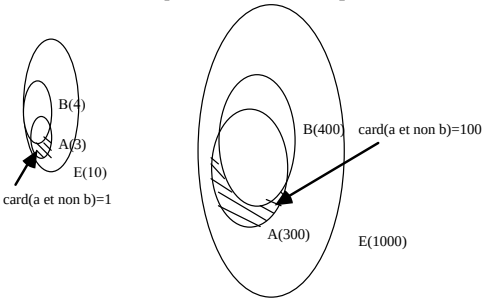


FIG. 2– *Deux situations cardinales*

Dans les deux cas, P[B/A]= 0,667, alors que, dans le premier cas, l'indice d'implication de a vers b est de l'ordre de $\varphi(a,b) \approx 0{,}7$, dans le second cas $\varphi(a,b)$ est très proche de 1. Le crédit que nous pouvons effectivement accorder, dans ce dernier cas, à une relation étroite et *invraisemblable* entre a et b y est nettement plus grand du fait de la taille des exemples la confirmant. Notre indice en rend compte alors que l'indice conditionnel ne change pas et ne souligne pas, de ce fait, la densité du phénomène.

7 Situation fondamentale et fondatrice de l'approche entropique.

Deux raisons nous ont conduits à améliorer le modèle formalisé par l'intensité d'implication dans l'approche classique:

- lorsque les tailles des ensembles d'individus traités augmentent, atteignant des effectifs de l'ordre du millier ou plus, l'intensité d'implication φ(a,b) a tendance à ne plus être suffisamment discriminante car ses valeurs peuvent être très voisines de 1, alors que l'inclusion dont elle cherche à modéliser la qualité, est loin d'être satisfaite. Ce phénomène a été déjà signalé par A. Bodin (1997) dont les travaux traitent avec des ensembles de grande taille d'élèves impliqués dans des enquêtes internationales;

- le modèle classique de la quasi-implication retient essentiellement la mesure de l'intensité de la quasi-règle $a \Rightarrow b$. Or comme nous l'avons abordé en introduction à propos de la causalité, la prise en compte concomitante de la contraposée de l'implication de $\bar{b} \Rightarrow \bar{a}$ est indispensable pour renforcer l'évaluation de la qualité suffisamment bonne de la relation de quasi-implication, voire quasi-causale, de a sur b[6]. En même temps, elle pourrait permettre de corriger la difficulté évoquée en relation à la taille des ensembles en jeu. En effet si A et B sont des ensembles de petite taille par rapport à E, leurs complémentaires seront importants et réciproquement.

8 Formalisation de la quasi-règle implicative dans l'approche entropique.

La solution[7] que nous apportons utilise à la fois l'intensité d'implication et un autre indice qui rend compte de la dissymétrie entre les situations S_1=(a et b) et S'_1=(a et non b) qui concerne la quasi-règle a⇒b ainsi que celle entre les situations S_2=(non a et non b) et S'_2 = (a et non b) qui concerne la quasi-règle contraposée. Notons que ce sont les mêmes instances qui contredisent la quasi-implication et sa contraposée. Les valeurs relatives de ces instances sont fondamentales dans notre approche.

8.1 Construction d'indice d'inclusion

Pour rendre compte de l'incertitude liée à un éventuel pari de l'appartenance à une des deux situations S_1 ou S'_1, (resp. S_2 ou S'_2), nous avons choisi le concept **d'entropie de Shannon** (1949).

Ainsi nous déterminons **l'entropie conditionnelle** relative à S_1 et S'_1 quand a est réalisée

$$H(b/a) = -\frac{n_{a\wedge b}}{n_a}\log_2\frac{n_{a\wedge b}}{n_a} - \frac{n_{a\wedge \bar{b}}}{n_a}\log_2\frac{n_{a\wedge \bar{b}}}{n_a}$$

puis l'entropie conditionnelle relative à S_2 et S'_2 lorsque non b est réalisée ou encore b n'est pas réalisée

[6] Ce phénomène est signalé par Y. Kodratoff dans son article publié dans les Actes du Colloque « Fouille dans les données par la méthode implicative », IUFM de Caen, juin 2000. Nous avons aussi abordé cette question du paradoxe de Hempel en Introduction de l'ouvrage

[7] J. Blanchard apporte dans (Blanchard J. et al. 2005) une réponse à ce problème par une mesure de « l'écart à l'équilibre ».

$$H(\bar{a}/\bar{b}) = -\frac{n_{a\wedge\bar{b}}}{n_{\bar{b}}}\log_2\frac{n_{a\wedge\bar{b}}}{n_{\bar{b}}} - \frac{n_{\bar{a}\wedge\bar{b}}}{n_{\bar{b}}}\log_2\frac{n_{\bar{a}\wedge\bar{b}}}{n_{\bar{b}}}$$

Ces entropies conditionnelles sont à valeurs dans [0,1] et devraient être simultanément faibles. En conséquence, les dissymétries entre les situations S_1 et S'_1 (resp. S_2 et S'_2) devraient être simultanément fortes si l'on souhaite disposer d'un bon critère d'inclusion de A dans B. En effet les entropies conditionnelles représentent l'**incertitude** moyenne des expériences qui consistent à observer si b (resp. non a) est réalisé lorsque l'on a observé a (resp. non b). Le complément à 1 de cette incertitude représente donc **l'information** moyenne recueillie par la réalisation de ces expériences. Plus cette information est importante, plus forte est la garantie de la qualité simultanée de l'implication et de sa contraposée. Nous devons maintenant adapter ce critère numérique entropique au modèle attendu dans les différentes situations cardinales.

Pour que le modèle ait la signification attendue, il doit satisfaire, selon nous, les contraintes épistémologiques suivantes :

1° il devra intégrer les valeurs de l'entropie et même, pour amplifier les contrastes, prendre le carré de ces valeurs ;

2° ce carré varie aussi de 0 à 1, pour rendre compte du déséquilibre, c'est-à-dire de l'inclusion en s'opposant à l'entropie, c'est-à-dire à l'incertitude, la valeur retenue sera le complément à 1 de son carré tant que le nombre de contre-exemples restera inférieur à la moitié des observations de a (resp. de non b). Si ce nombre dépasse la moitié, , nous affecterons la valeur 0 au critère compte tenu du fait que sémantiquement les implications perdent leur sens inclusif ;

3° afin de prendre en compte les deux informations propres à la quasi-implication et à sa contraposée, c'est le produit des valeurs que nous retiendrons. Le produit a la propriété de s'annuler dès que l'un de ses termes s'annule, i.e. dès que cette qualité s'efface ;

4° enfin, le produit ayant une dimension 4 par rapport à l'entropie, nous prendrons sa racine quatrième pour revenir à la même dimension.

Posons $\alpha = \frac{n_a}{n}$ la fréquence de a, $\overline{\beta} = \frac{n_{\bar{b}}}{n}$ la fréquence de non b et $t = \frac{n_{a\wedge\bar{b}}}{n}$ la fréquence des contre-exemples. Nous construisons alors deux fonctions h_1 et h_2 définies dans [0 ;1] comme suit :

$$h_1(t) = H(b/a) = -(1-\frac{t}{\alpha})\log_2(1-\frac{t}{\alpha}) - \frac{t}{\alpha}\log_2\frac{t}{\alpha}$$ si $t\in[0,\frac{\alpha}{2}[$ et $h_1(t) = 1$ si $t\in[\frac{\alpha}{2}, \alpha]$

$$h_2(t) = H(\bar{a}/\bar{b}) = -(1-\frac{t}{\overline{\beta}})\log_2(1-\frac{t}{\overline{\beta}}) - \frac{t}{\overline{\beta}}\log_2\frac{t}{\overline{\beta}}$$ si $t\in[0,\frac{\overline{\beta}}{2}[$ et $h_2(t) = 1$ si $t\in[\frac{\overline{\beta}}{2}, \overline{\beta}]$

De là, nous proposons la définition suivante permettant de déterminer le critère entropique.

Définition 4: **L' indice d'inclusion** de A, support de a, dans B, support de b, est le nombre :

$$i(a,b) = \left(\left[1-h_1^2(t)\right]\left[1-h_2^2(t)\right]\right)^{\frac{1}{4}}$$

qui intègre l'information délivrée par la réalisation du nombre de contre-exemples, d'une part à la quasi-règle $a \Rightarrow b$ et, d'autre part, à la quasi-règle $\bar{b} \Rightarrow \bar{a}$

8.2 Construction d'un indice d'implication-inclusion

Pour prendre en compte à la fois, ce que nous avons appelé l'étonnement statistique et la qualité de l'inclusion de l'ensemble A des instances de a dans celui B des instances de b, nous proposons une nouvelle mesure de la qualité inductive suivante que nous nommons intensité d'implication-inclusion ou intensité entropique.

Définition 5: **L'intensité d'implication-inclusion ou intensité entropique**, est le nombre suivant:

$$\psi(a,b) = [i(a,b)\ \varphi(a,b)]^{1/2}$$

La fonction ψ de la variable t admet une représentation qui a la forme indiquée par la figure ci-dessous, pour n_a et n_b fixés. On remarquera sur cette figure, la différence de comportement de la fonction ψ par rapport à la probabilité conditionnelle P(B/A), indice fondamental des autres modélisations de la mesure des règles, comme par exemple chez Agrawal et son école. Outre son caractère linéaire, donc peu nuancé, cette dernière probabilité conduit à une mesure qui décroît trop vite dès les premiers contre-exemples et résiste ensuite trop longuement lorsque ceux-ci apparaissent en nombre important.

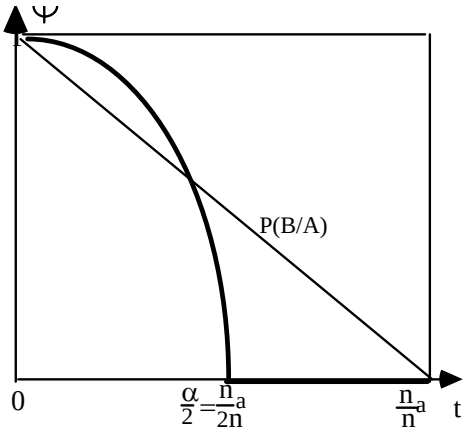


FIG. 3- *représentation graphique de la fonction ψ*

On constate que cette représentation de la fonction ψ continue de t traduit les propriétés attendues du critère d'inclusion :

1- réaction lente aux premiers contre-exemples (résistance au bruit),

2- accélération du rejet de l'inclusion au voisinage de l'équilibre soit $\frac{n_a}{2n}$,

3- rejet au-delà de $\frac{n_a}{2n}$ ce que n'assurait pas à elle seule, l'intensité d'implication φ(a,b).

8.3 Exploration succincte des propriétés de ψ

	b	$\bar{b}$	marge
a	200	400	600
$\bar{a}$	600	2800	3400
marge	800	3200	4000

TAB. 7 – *exemple n°1*

L'indice d'implication $q(a,\bar{b}) = -3,65$ (modèle de Poisson)

L'intensité d'implication est φ(a,b) = 0,999869

L'indice d'implication $q(a,\bar{b}) = -3,89$ (modèle binomial)

L'intensité d'implication est φ(a,b) = 0,999950

Les valeurs entropiques de l'expérience sont $h_1 = 1$ et $h_2 = 1$

La valeur du coefficient modérateur est donc : i(a,b) = 0

Par suite ψ(a,b) = 0 alors que P(B/A) = 0,33333

Ainsi, les fonctions entropiques modèrent l'intensité d'implication dans ce cas où justement l'inclusion est médiocre.

	b	$\bar{b}$	marge
a	400	200	600
$\bar{a}$	1000	2400	3400
marge	1400	2600	4000

TAB. 8 – *exemple n°2*

L'intensité d'implication est φ(a,b) = 1 pour un indice d'implication $q(a,\bar{b}) = -9,621$ (modèle de Poisson)

L'intensité d'implication est φ(a,b) = 1 pour un indice d'implication $q(a,\bar{b}) = -10,127$ (modèle binomial)

Les valeurs entropiques de l'expérience sont $h_1 = 0,918$ et $h_2 = 0,391$

La valeur du coefficient modérateur est donc : i(a,b) = 0,6036

Par suite ψ(a,b) = 0,777 alors que P(B/A) = 0,66666

	b	$\bar{b}$	marge
a	40	20	60
$\bar{a}$	100	240	340
marge	140	260	400

TAB. 9 – *exemple n°3*

L'intensité d'implication est φ(a,b) = 0,9988 pour un indice d'implication $q(a,\bar{b}) = -3,04$ (modèle de Poisson)

L'intensité d'implication est φ(a,b) = 0,9993 pour un indice d'implication $q(a,\bar{b})$ = -3,20 (modèle binomial)

Les valeurs entropiques de l'expérience sont h_1 = 0,918 et h_2 = 0,391

La valeur du coefficient modérateur est donc : i(a,b) = 0,603

Par suite ψ(a,b) = 0,776 alors que P(B/A) = 0,66666

Ainsi, φ(a,b) a diminué du 2ème au 3ème exemples, puisque le cardinal de l'ensemble de référence E a crû dans l'homothétie cardinale de rapport 1/10. Mais i(a,b) a augmenté de même que ψ(a,b).

Notons que, dans les deux cas, la probabilité conditionnelle ne change pas.

Pour plus de précisions, nous renvoyons à (Lenca et al, 2004) pour une étude comparative, très fouillée, des indices d'association pour des variables binaires. En particulier, les intensités d'implication classique et entropique (inclusion) présentées y sont confrontées à d'autres indices selon une entrée « utilisateur ».

Chapitre 2 : Représentation des règles d'implication et graphe implicatif

1 Problématique

A l'issue des calculs des intensités d'implication, que ce soit dans le modèle classique ou celui entropique, nous disposons d'un tableau p x p qui croise les p variables entre elles, quelle que soit leur nature, et dont les éléments sont les valeurs de ces intensités d'implication, nombres de l'intervalle [0 ;1]. Force est de constater que la structure sous-jacente de l'ensemble de ces variables est loin d'être explicite et demeure largement inapparente. L'utilisateur reste aveugle face à un tel tableau carré de taille p^2. Il ne peut embrasser simultanément les multiples enchaînements éventuels des règles qui sous-tendent la structure globale de l'ensemble des p variables. Afin de faciliter une extraction plus claire des règles et d'en examiner leur structure, nous avons associé à ce tableau, et pour un seuil d'intensité donné, un **graphe implicatif,** orienté, pondéré par les intensités d'implication, sans cycle dont l'utilisateur peut contrôler la complexité de la représentation en fixant lui-même le seuil de prise en compte de la qualité implicative des règles. Chaque arc de ce graphe représente une règle : si $n_a < n_b$, l'arc $a \rightarrow b$ représente la règle $a \Rightarrow b$; si $n_a = n_b$, alors l'arc $a \leftrightarrow b$ représentera la double règle $a \Leftrightarrow b$, en d'autres mots, l'équivalence entre ces deux variables. En faisant varier le seuil d'intensité d'implication, il est évident que le nombre d'arcs varie dans le sens opposé : pour un seuil fixé à 0,95, le nombre d'arcs est inférieur ou égal à ceux qui constitueraient le graphe au seuil 0, 90. Nous en reparlerons plus loin.

2 Algorithme

La relation définie par l'implication statistique, si elle est réflexive et non symétrique, donc sans cycle, n'est pas **transitive** bien évidemment, comme l'induction et au contraire de la déduction. Or nous voulons qu'elle modélise la relation d'ordre partiel entre deux variables. Ainsi en est-il de la question des réussites évoquées dans notre exemple fondamental.

Proposition 6 : Par convention, si $a \Rightarrow b$ et si $b \Rightarrow c$, il y a **fermeture transitive** $a \Rightarrow c$ si et seulement si $\varphi(a,c) \geq 0,5$, c'est-à-dire si la relation implicative de a sur c, qui traduit une certaine dépendance entre a et c, est meilleure que sa réfutation. Notons que, pour tout couple de variables (x ; y), l'arc $x \rightarrow y$ est pondéré par l'intensité d'implication $\varphi(x,y)$.

Prenons un exemple formel en supposant qu'entre les 7 variables *a, b, c, d, e f et g* existent, au seuil supérieur à 0,5, les règles suivantes :

$e \Rightarrow c \quad e \Rightarrow a \quad e \Rightarrow f \quad e \Rightarrow b$
$c \Rightarrow a \quad c \Rightarrow f$
$b \Rightarrow a \quad b \Rightarrow f$
$g \Rightarrow d \quad g \Rightarrow f$
$a \Rightarrow f$

On pourra alors traduire cet ensemble de relations par le graphe suivant[8] :

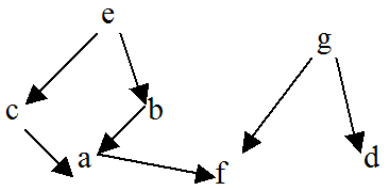


FIG. 4- *Un exemple de graphe implicatif*

Une des difficultés liées à la représentation graphique tient au fait que le graphe n'est pas planaire. L'algorithme qui en permet la construction, doit le prendre en compte et, en particulier, se doit de « décroiser » les chemins du graphe afin d'en permettre une lisibilité acceptable pour l'expert qui va l'analyser.

Le nombre d'arcs du graphe peut être réduit (*resp*. augmenté) si nous élevons (*resp.* abaissons) le seuil d'acceptation des règles, le niveau de confiance en les règles retenues. Corrélativement, des arcs peuvent apparaître ou disparaître selon les variations du seuil. Rappelons que ce graphe est nécessairement sans cycle, qu'il n'est pas un treillis puisque, par exemple la variable a n'implique pas la variable (a ou non a) dont le support est E. A fortiori, il ne peut être un treillis de Galois. Des options du logiciel C.H.I.C. de traitement automatique des données sous l'A.S.I., comme nous le verrons (Partie 2, chap. 11), permettent de supprimer à volonté des variables, de déplacer leur image dans le graphe afin de décroiser les arcs ou de se centrer sur certaines variables dites sommets d'une sorte de « **cône » dont les deux « nappes** » sont constituées respectivement des variables « **parents** » et des variables « **enfants** » de cette variable sommet (Partie 2, chap. 11). Nous désignons les extrémités des arcs par le terme « **nœuds** ». A un nœud d'un graphe donné, correspond une et une seule variable ou une conjonction de variables. Le passage d'un nœud S1 à un nœud S2 est appelé aussi « transition » qui est représentée par un arc du graphe.

3 Un exemple numérique à des fins didactiques

Nous partirons de la situation d'un tableau simulé joint ci-dessous croisant 30 individus (de i_1 à i_30) et 5 variables binaires (V1, V2, V3, V4 et V5). Deux autres variables (F s et H s) seront prises en compte comme variables supplémentaires que nous aborderons plus loin (Partie 1 Chap. 5). Ce tableau constitue un fichier dénommé RAF.

[8] Les traitements automatiques des calculs et des graphiques sont exécutés à l'aide du logiciel C.H.I.C. (acronyme de **C**lassification **H**iérarchique **I**mplicative et **C**ohésitive) disponible sous Windows 95, 98, NT, XP et Vista. Ce logiciel, à partir d'une première version établie par R. Gras et H. Rostam, révisée sous Pascal par S. Ag Almouloud, (Ag Almouloud S. 1992), est maintenant développé par R. Couturier (Couturier et Gras. 2005), constamment étendu par lui aux nouveaux concepts et nouveaux algorithmes et entretenu pour sa convivialité

	V1	V2	V3	V4	V5	F s	G s
i1	1	1	1	0	0	1	0
i2	1	1	1	0	0	1	0
i3	1	1	1	0	0	1	0
i4	1	1	1	0	0	1	0
i5	1	1	1	0	0	1	0
i6	1	1	1	1	0	1	0
i7	1	1	0	1	1	1	0
i8	1	1	0	1	1	1	0
i9	1	1	1	1	1	1	0
i10	1	1	1	1	1	1	0
i11	1	0	1	1	0	1	0
i12	1	0	1	1	0	1	0
i13	1	0	1	1	0	1	0
i14	1	0	1	1	1	0	1
i15	1	0	1	1	0	1	0

	V1	V2	V3	V4	V5	F s	G s
i16	1	0	1	1	1	0	1
i17	1	0	1	1	1	0	1
i18	1	0	1	1	0	0	1
i19	1	0	1	1	0	0	1
i20	1	0	1	0	1	1	0
i21	1	0	1	1	0	0	1
i22	1	0	1	0	0	0	1
i23	1	0	0	1	1	0	1
i24	1	0	0	0	0	0	1
i25	0	0	0	0	0	1	0
i26	0	0	0	0	0	0	1
i27	0	0	0	0	0	0	1
i28	0	0	0	1	0	0	1
i29	0	0	1	1	1	0	1
i30	0	0	0	1	0	0	1

TAB. 10- *tableau de données du fichier Raf*

Afin de prendre pleine conscience des concepts de l'implication statistique définis précédemment (Partie 1 Chap. 1) nous proposons la situation suivante dans laquelle les calculs seront réalisés une calculatrice ou un tableur :

- Étape n°1 : calculer les indices d'implication des couples de variables puis les intensités d'implication en se plaçant selon le modèle de Poisson puis dans le modèle recourant à l'approximation gaussienne de la loi de la variable $Q(a,\bar{b})$
- Étape n°2 : réaliser la représentation graphique du graphe des règles.

Pour ne pas avoir à retourner aux formules déjà présentées (Partie 1 Chap. 1), pour évaluer la quasi-implication a⇒b, nous les redonnons ici. En utilisant le modèle de Poisson de paramètre estimé $\hat{\lambda} = \frac{n_{a\wedge\bar{b}}}{n}$, l'intensité d'implication est calculée ainsi :

$$\varphi(a,b) = 1 - \sum_{s=0}^{n_{a\wedge\bar{b}}} \frac{\hat{\lambda}^s}{s!}.e^{-\hat{\lambda}} = 1 - e^{-\hat{\lambda}}(\hat{\lambda} + \frac{\hat{\lambda}^2}{2} + ... + \frac{\hat{\lambda}^{n_{a\wedge\bar{b}}}}{(^{n}{}_{a\wedge\bar{b}})!})$$

En utilisant l'algorithme de calcul de l'intensité d'implication par l'approximation gaussienne, nous obtenons :

En posant : $q(a,\bar{b}) = \dfrac{n_{a\wedge\bar{b}} - \dfrac{n_{a\wedge\bar{b}}}{n}}{\sqrt{\dfrac{n_{a\wedge\bar{b}}}{n}}}$, on a $\varphi(a,b) = \dfrac{1}{\sqrt{2\pi}} \int_{q(a,\bar{b})}^{\infty} e^{-\frac{t^2}{2}} dt$

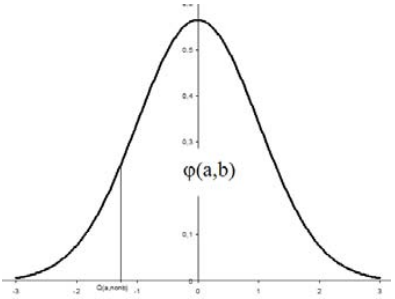


FIG. 5- *Image de l'intensité d'implication dans le cas du calcul fondé sur l'approximation gaussienne*

Il nous faut établir les 25 croisements entre les 5 variables binaires dont nous pouvons réduire les résultats en 10 tableaux suivants. Nous y avons indiqué les valeurs respectives du χ^2 dont seules celles correspondant aux tableaux V3-V1, V2-V3 indiquent une liaison significative au seuil de risque α=0,05 pour lequel la valeur critique est de 3,84 . La situation du tableau V2-V1 est proche de la liaison significative avec une valeur empirique de 3,75. Abordé d'une autre façon, nous pourrions préciser que dans ce cas la p-value serait de p=0,052. Comme nous avons pu le voir progressivement au travers des propos tenus tout au long de ce qui précède, notre perspective s'appuie sur un point de vue soutenu par I.-C. Lerman (1992) appliqué à l'étude d'une certaine relation de dépendance orientée entre des variables descriptives. Ce point de vue oppose les tests d'indépendance, comme celui bien connu dit du χ^2, aux méthodes classificatoires de la manière suivante : rappelons-le, pour les premiers, dit I.-C. Lerman, « on a, relativement à l'existence d'un lien, FAUX, jusqu'à preuve du contraire » par le rejet de l'hypothèse nulle ; pour les secondes, on a VRAI, jusqu'à preuve du contraire », c'est-à-dire vrai selon une certaine échelle de probabilité. Cette posture permet de lever une observation qui aurait pu paraître paradoxale.

Nous avons présenté les tableaux de façon à ce que la variable-ligne corresponde à la variable générique a et la variable-colonne, à la variable b, avec le respect de la contrainte $n_a \leq n_b$. Nous portons dans chaque tableau, la valeur de l'indice d'implication Q= $Q(a,\bar{b})$, les valeurs des intensités d'implication $\varphi_P(a,b)$ et $\varphi_{LG}(a,b)$ correspondant respectivement au modèle de Poisson et au calcul par l'approximation gaussienne. Pour donner plus de précision sur la mise en œuvre des algorithmes de calcul, nous développons le cas du tableau V2-V1.

D'une part $n_{V2} = 10 < n_{V1} = 24$ et $n_{V2\wedge\overline{V1}} = 0$ ce qui conduit à :

$$Q(V2,\overline{V1}) = \frac{n_{V2\wedge\overline{V1}} - \frac{n_{V2}n_{\overline{V1}}}{n}}{\sqrt{\frac{n_{V2}n_{\overline{V1}}}{n}}} = \frac{0-\frac{10x6}{30}}{\sqrt{\frac{10x6}{30}}} = -\sqrt{\frac{10x6}{30}} = -\sqrt{2} \approx -1{,}414$$

D'autre part $\hat{\lambda} = \frac{n_{V2}n_{\overline{V1}}}{n} = 2$ et $\varphi_P(V2,V1) = 1 - \sum_{s=0}^{0} \frac{\hat{\lambda}^s}{s!}.e^{-\hat{\lambda}} = 1 - e^{-2} \approx 0{,}86466472$

Si nous utilisons l'approximation gaussienne, nous obtenons

$$\varphi_{LG}(V2,V1) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{2}}^{\infty} e^{-\frac{t^2}{2}} dt \approx 0{,}921319$$

χ^2 = 3,75		Q= -1,41			χ^2 = 1,14		Q= 0,696		
V2-V1	V1oui	V1non		φ_P(V2,V1)	V2-V4	V4oui	V4non		φ_P(V2,V4)
V2oui	10	0	10	**0,864**	V2oui	5	5	10	0,165
V2non	14	6	20	φ_{LG}(V2,V1)	V2non	14	6	20	φ_{LG}(V2,V4)
	24	6	30	0,921		19	11	30	0,243
χ^2 = 10,15		Q= -1,56			χ^2 = 0,3		Q= -0,258		
V3-V1	V1oui	V1non		φ_P(V3,V1)	V2-V5	V5oui	V5non		φ_P(V2,V5)
V3oui	20	1	21	**0,922**	V2oui	4	6	10	0,499
V3non	4	5	9	φ_{LG}(V3,V1)	V2non	6	14	20	φ_{LG}(V2,V5)
	24	6	30	0 ,940		10	20	30	0,60
χ^2 = 0,574		Q= -0,41			χ^2 = 0,334		Q= -0,293		
V4-V1	V1oui	V1non		φ_P(V4,V1)	V4-V3	V3oui	V3non		φ_P(V4,V3)
V4oui	16	3	19	0,526	V4oui	14	5	19	0,505
V4non	8	3	11	φ_{LG}(V4,V1)	V4non	7	4	11	φ_{LG}(V4,V3)
	24	6	30	0,659		21	9	30	0,615
χ^2 = 0,937		Q= -0,70			χ^2 = 0		Q= 0		
V5-V1	V1oui	V1non		φ_P(V5,V1)	V3-V5	V5oui	V5non		φ_P(V3,V5)
V5oui	9	1	10	0,593	V3oui	7	3	10	0 , 352
V5non	15	5	20	φ_{LG}(V5,V1)	V3non	14	6	20	φ_{LG}(V3,V5)
	24	6	30	0,760		21	9	30	0,5
χ^2 = 0,714		Q= - 0,57			χ^2 = 4,59		Q= -1,39		
V2-V3	V3oui	V3non		φ_P(V2,V3)	V5-V4	V4oui	V4non		φ_P(V5,V4)
V2oui	8	2	10	**0,576**	V5oui	9	1	10	0,880
V2non	13	7	20	φ_{LG}(V2,V3)	V5non	10	10	10	φ_{LG}(V4,V5)
	21	9	30	0,718		19	11	30	0,918

TAB. 11

On remarquera que les valeurs de l'intensité données par l'approximation gaussienne sont plus élevées que celles données par la valeur « vraie » calculée dans le modèle de Poisson.

Mais comme l'ordre entre les valeurs est inchangé, ce qui est essentiel, le graphe correspondant reste le même quel que soit le modèle choisi.

En fonction du seuil choisi avec le modèle de Poisson, on obtient deux graphes différents.

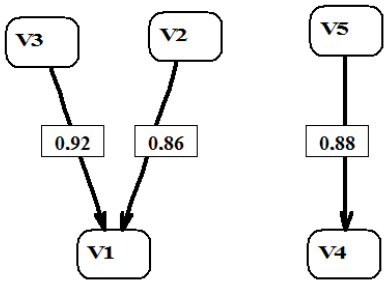


FIG. 6- *Graphe implicatif au seuil de confiance de 0,85 (Intensité calculée directement dans le modèle de Poisson)*

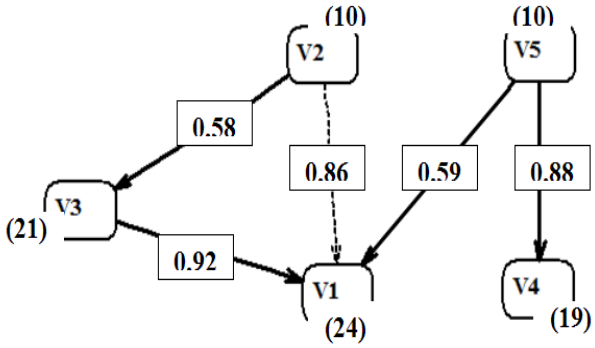


FIG. 7- *Graphe implicatif au seuil de confiance de 0,57 (Intensité calculée directement dans le modèle de Poisson)*

Modifiant le seuil à 0,57, la liaison entre V2 et V3 peut s'établir et permettre de créer un chemin transitif entre V2 et V1. L'arc V2 → V1 pourrait même être supprimé car il devient redondant en raison de l'information V2 → V3. C'est pourquoi nous l'avons mis en ligne pointillée.

En fonction de ce seuil choisi 0,57 et en procédant au calcul de l'intensité d'implication en recourant à l'approximation gaussienne, on obtient le graphe suivant.

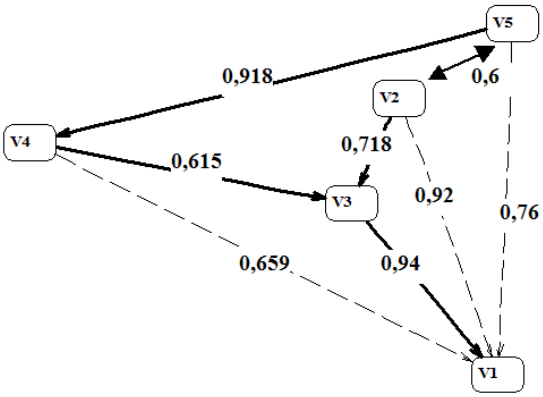


FIG. 8- *Graphe implicatif au seuil de confiance de 0,57 (Intensité calculée avec l'approximation gaussienne)*

Nous observons l'apparition d'une équivalence statistique entre V2 et V5 repérée par l'arc double V2↔V5. Par ailleurs nous avons marqué les fermetures transitives par les flèches en pointillé lesquelles pourraient être retirées pour alléger la représentation du graphe comme nous le présentons ci-dessous :

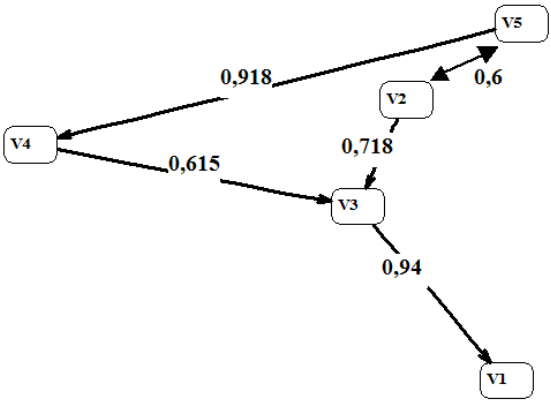


FIG. 9- *Graphe implicatif au seuil de confiance de 0,57 sans les fermetures transitives)*

4 Question sur la transition d'un nœud d'un graphe implicatif à un suivant

Nous avons déjà abordé précédemment (Partie 1, Chap. 1 – 6.2, 6.3) quelques propriétés de l'indice d'implication et de l'intensité d'implication. Nous reprenons ici l'étude de leurs

variations en fonction des variables effectifs (n, n_a, n_b, $n_{\bar{b}}$, $n_{a\wedge\bar{b}}$) ou plutôt des variables fréquences (f_a, f_b, $f_{\bar{b}}$, $f_{a\wedge\bar{b}}$). Ce problème de contrôle des variations de l'indice et de l'intensité d'implication se présente lors de fouille de règles d'association généralisées (cf. Chap 4) ou encore pour l'alignement de deux ontologies[9] (David J. et al., 2006) et (Gras et al. 2007 a).

Pour illustrer l'étude visée et lui justifier son sens, partons de l'exemple suivant où 3 variables attributs binaires (d'un repas) sont en jeu : a = {manger des huîtres} ; b_1 = {boire du vin blanc} ; b_2 = {boire du muscadet}.

En général, au cours d'un repas, on observe la règle au sens de l'ASI : $r_1 = [a \Rightarrow b_1]$, mais aussi la règle $r_2 = [a \Rightarrow b_2]$. Dans une taxonomie (ou **ontologie**) des vins, on a $b_2 \Rightarrow b_1$ parce que « muscadet » est plus spécifique que « vin blanc ». Cette dernière règle est stricte au sens de la logique mathématique. On remarque aisément que la règle r_1 est plus générale (en terme de support la réalisant) que la règle r_2. On peut ainsi se demander si la règle r_2 apporte plus d'information prédictive que r_1, autrement dit si r_2 apporte un réel gain de force ou d'intensité implicative que celle attendue par rapport à r_1. Pour cela nous proposons d'étudier la variation d'intensité d'implication entre r_1 et r_2.

On voit alors que le problème peut se généraliser dans la recherche d'élimination de redondances ou de superfluité de la façon suivante : sachant que $a \Rightarrow b_i$ pour un même attribut a et un certain nombre d'attributs spécifiques b_i, n'existe-t-il pas une règle $a \Rightarrow b_0$ qui conduise à une intensité d'implication optimale parmi les règles découlant de taxonomies T_i et qui permette l'élimination de règles moins riches en information inductive ?

4.1 Variations de l'indice d'implication q

Dans les hypothèses ci-dessus, les occurrences n de la population et les occurrences n_a de l'attribut a étant fixées, étudions les conditions de gain quand on élimine la règle r_1 au profit de la règle r_2 en fonction des paramètres associées. Désignons par b_1 (resp. b_2) les extrémités (ou nœuds) des arcs représentant respectivement ces deux règles. Supposons, comme dans l'exemple, que n_b varie en décroissant selon que l'on échange b_1 en b_2, Autrement dit, $n_{b1} > n_{b2}$. En conséquence, dans le même temps et nécessairement, le nombre de contre-exemples $n_{a\wedge\bar{b}}$ croît en passant de b_1 à b_2.

Nous voulons alors comparer les deux états contingents b_1 et b_2 associés respectivement aux règles r1 et r_2. Pour ce faire, il suffit de calculer les dérivées partielles par rapport aux fréquences variables $f_{\bar{b}}$ et $f_{a\wedge\bar{b}}$ au nœud b_1.

$$\Delta q(b_1; b_2) \approx \frac{\partial q}{\partial f_{a\wedge\bar{b}}}(b_1)\Delta f_{a\wedge\bar{b}} + \frac{\partial q}{\partial f_{\bar{b}}}(b_1)\Delta f_{\bar{b}} > 0$$

avec $\Delta f_{a\wedge\bar{b}} = f_{a\wedge\bar{b}_2} - f_{a\wedge\bar{b}_1} > 0$ et $\Delta f_{\bar{b}} = f_{\bar{b}_2} - f_{\bar{b}_1} > 0$

[9] De façon raccourcie, une **ontologie** est une théorie logique qui décrit des termes (concepts) d'un domaine et leurs propriétés (relations entre concepts). Une ontologie peut être plus ou moins formelle (et exprimée de manière plus ou moins expressive) Souvent ces concepts sont organisés en taxonomie (du concept le plus général au plus spécifique).

et pour $k \in \{1;2\}, q(b_k) = q(a, \overline{b}_k) = \sqrt{n}\,\dfrac{(f_{a\wedge\overline{b_k}} - f_a f_{\overline{b_k}})}{\sqrt{f_a f_{\overline{b_k}}}}$

D'où $\dfrac{\partial q}{\partial f_{a\wedge\overline{b_k}}} = \sqrt{n}\,\dfrac{1}{\sqrt{f_a f_{\overline{b_k}}}}$ et $\dfrac{\partial q}{\partial f_{\overline{b_k}}} = -\dfrac{\sqrt{n}}{2}\left[\dfrac{f_{a\wedge\overline{b_k}}}{\sqrt{f_{\overline{a}}}} f_{\overline{b_k}}^{-\frac{3}{2}} + \sqrt{f_a}\, f_{\overline{b_k}}^{-\frac{1}{2}}\right]$

Comme les valeurs de l'indice d'implication $q(b_k)$ sont négatives dans le cas où n_{a_k} est inférieur à n_{b_k}, il suffit alors de comparer la variation observée $q(b_2) - q(b_1)$ que l'on souhaite négative, c'est-à-dire qu'il y ait une meilleure intensité d'implication en b_2 qu'en b_1, à la variation attendue par le gradient calculé. Ainsi donc c'est le signe de la différence $\Delta q(b_1; b_2) - [q(b_2) - q(b_1)]$ qui va nous informer sur l'amélioration ou non de l'intensité d'implication au cours de la transition T de b_1 vers b_2.

Si cette différence est positive, l'amélioration observée est plus intéressante pour l'intensité d'implication qu'elle n'aurait été si l'évolution d'un nœud b_1 au suivant b_2 avait suivi le gradient de q en b_1.

4.2 Variations de l'intensité d'implication

Nous cherchons maintenant à déterminer les variations de $\varphi(a,b)$ lors de la transition T du nœud S_1 vers le nœud S_2. Pour cela, nous calculons l'intensité d'implication $\varphi(T)$, en recourant à l'approximation gaussienne de la loi de la v.a. centrée réduite $Q(a,\overline{b})$ établie à partir des contre-exemples à l'implication, afférente à la variation théorique attendue du gradient de q, à savoir

$$q(T) = q(a,\overline{b_1}) + \Delta(S_1, S_2) \text{ et } \varphi(T) = \frac{1}{\sqrt{2\pi}} \int_{q(T)}^{\infty} e^{\frac{-t^2}{2}} dt$$

Pour estimer le gain dû à la transition T, il suffit alors de comparer cette intensité à celle qui a été observée $\varphi(a,b_2) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{q(a,\overline{b_2})}^{\infty} e^{\frac{-t^2}{2}} dt$.

Le gain sera positif ou négatif selon que l'intensité observée sera supérieure ou égale à l'intensité attendue du gradient. Il sera exprimé en pourcentage de $\varphi(T)$ par le rapport :

$$\frac{\varphi(a,b_2) - \varphi(T)}{\varphi(T)}$$

.

Nous utiliserons alors la formulation suivante que le gain lié à la transition T est de g% de l'intensité attendue de l'observation en S_1 et du gradient de q en ce nœud.

Notons que cette méthode peut être généralisée quelles que soient les variables en jeu, c'est-à-dire dans les cas où d'autres variables peuvent être modifiées. Il suffit d'avoir recours à la différentielle de q selon les 4 variables actives (f_a, f_b, $f_{\overline{b}}$, $f_{a\wedge\overline{b}}$).

Remarque. Considérons l'intensité d'implication φ comme fonction de $q(a,\bar{b})$:

$$\varphi(q) = \frac{1}{\sqrt{2\pi}} \int_q^{\infty} e^{-t^2/2} dt$$

On peut alors examiner comment φ(q) varie lorsque q varie au voisinage d'une valeur donnée (a,b), sachant comment q varie lui-même en fonction des 4 paramètres qui le déterminent. Par dérivation de la borne d'intégration, on obtient :

$$\frac{d\varphi}{dq} = -\frac{1}{\sqrt{2\pi}} e^{\frac{-q^2}{2}} < 0$$

Ce qui confirme bien que l'intensité croît lorsque q décroît, mais la vitesse de croissance est précisée par la formule, ce qui permet d'étudier avec plus de précision les variations de φ(q).

.

Chapitre 3 : Extension de l'Analyse Statistique Implicative aux variables non binaires

1 L'A.S.I. des variables modales, variables fréquentielles et variables numériques

Nous abordons dans ce chapitre la question du traitement de variables autres que les variables binaires.

1.1 Situation fondatrice de la relation de propension entre deux variables modales

La thèse de Marc Bailleul (1994) porte en particulier, sur la représentation que se font les enseignants de mathématiques de leur propre enseignement. Afin de la mettre en évidence, des mots significatifs leur sont proposés qu'ils doivent hiérarchiser. Leurs choix ne sont donc plus binaires, les mots retenus par un enseignant quelconque sont ordonnés du moins au plus représentatif. L'interrogation de M. Bailleul se centre alors sur des questions du type : « si je choisis tel mot avec telle importance alors je choisis tel autre mot avec une importance au moins égale ». Pour analyser les données, il a fallu étendre la notion d'implication statistique à des variables autres que binaires. C'est le cas des **variables modales** qui sont associées à des phénomènes où les valeurs a(x) sont des nombres de l'intervalle [0,1] et qui décrivent des degrés d'appartenance ou de satisfaction comme le sont en logique floue, par exemple, les modificateurs linguistiques "peut-être", "un peu", "quelquefois", etc.. Cette problématique se retrouve également dans des situations où la fréquence d'une variable traduit un préordre sur les valeurs attribuées par les sujets aux variables qui leur sont présentées. Il s'agit de variables fréquentielles qui sont associées à des phénomènes où les valeurs de a(x) sont des réels positifs quelconques. On trouve une telle situation lorsque l'on considère le pourcentage de réussite d'un élève à une batterie de tests portant sur des domaines distincts.

1.2 Formalisation de la relation de propension entre variables modales

J.B. Lagrange (1998) a construit dans le cas des variables modales, un indice de propension (auparavant M. Bailleul (1994) fit autrement) entre variables modales qui généralise l'indice d'implication entre variables binaires. En posant la définition suivante :

Définition 6:

- si a(x) et $\overline{b}(x)$ sont les valeurs prises en x par les variables modales a et $\overline{b}$, où $\overline{b}(x) = 1 - b(x)$

- si s_a^2 et $s_{\overline{b}}^2$ sont les variances empiriques des variables a et $\overline{b}$

alors

$$\tilde{q}(a,\bar{b}) = \frac{\sum_{x\in E} a(x)\bar{b}(x) - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{(n^2 s_a^2 + n_a^2)(n^2 s_{\bar{b}}^2 + n_{\bar{b}}^2)}{n^3}}}$$ est l'indice de propension de variables modales.

Cette solution apportée au cas modal est aussi applicable au cas des **variables fréquentielles**, voire des **variables numériques** positives, à condition d'avoir normalisé les valeurs observées sur les variables, telles que a et b, la normalisation dans [0,1] étant faite à partir du maximum de la valeur prise respectivement par a et b sur l'ensemble E.

1.3 Modélisation et propriétés de l'indice de propension et de l'intensité de propension entre variables non binaires

Nous nous intéressons désormais aux variables qui ne sont plus nécessairement binaires mais à valeurs réelles normalisées sur l'intervalle [0, 1]. Elles sont observées sur un ensemble de transactions E de cardinal n. Parmi ces variables *numériques*, les variables dites *modales* admettent un nombre fini de modalités respectivement *ordonnées* sur l'intervalle [0,1].

A l'instar de J.B. Lagrange (1998) et de S. Guillaume (2000), nous utilisons l'expression : propension (ou tendance) de a vers b, si l'on rencontre généralement dans E peu de transactions $i \in E$ pour lesquels $a_i > b_i$, pour la relation d'ordre sur [0, 1] où a_i et b_i sont respectivement les valeurs de *a* et *b* observées en *i*.

Notons $\bar{b}_i$ le complément à 1 de b_i : $\bar{b}_i = 1\text{-}b_i$. On choisit donc, comme pour l'implication entre variables binaires, l'indice $\sum_{i\in E} a_i \bar{b}_i$ - qui prend la valeur $n_{a\wedge\bar{b}}$ dans le cas binaire - comme indice de non-propension (ou de non-tendance) de *a* vers *b*. Ainsi, intuitivement et grossièrement, plus cet indice sera petit, plus on pourra s'attendre à une propension de *a* vers *b* (Régnier et Gras 2004)). Nous préciserons ce point plus loin. `

Cet indice présente un caractère voisin de l'indice de corrélation linéaire. Mais, d'une part, il n'est pas centré et, d'autre part, notre intérêt portera surtout sur ses valeurs fortes obtenues dans les cas où la variable *b* domine la variable *a*

Nous reprenons la démarche de J.B. Lagrange (1998), mais à travers une modélisation qui sera différente. En effet, il visait une *modélisation de Poisson* dans la restriction au cas binaire. Nous en avons cité le résultat principal dans l'article (Gras et al., 2001). En revanche, nous visons ici une modélisation de l'indice de propension en tant qu'extension stricte de la *modélisation binomiale* dans le cas des variables binaires.

Considérons pour cela, *n* couples de variables aléatoires X_i et Y_i indépendantes et identiquement distribuées, pour tout sujet *i*, dont la moyenne et la variance estimées sont celles des observations empiriques de E. Les réalisations de ces variables sont respectivement les n couples (a_i, b_i) de valeurs de $[0, 1]^2$.

Ainsi, par exemple, pour tout *i*, l'espérance E(X_i) peut être estimée par $m_a = \frac{1}{n}\sum_{k\in E} a_k$ et la variance Var(X_i) peut l'être par la réalisation de la variance empirique de *a* :

$$\upsilon_a = \frac{1}{n}\sum_{k\in E}(a_k - m_a)^2$$

On note également $\overline{Y_i} = 1 - Y_i$.

On pose : $M_1 = E[X_i \overline{Y_i}]$, espérance de la variable aléatoire produit $X_i \overline{Y_i}$ et $M_2 = E[(X_i \overline{Y_i})^2]$, moment d'ordre 2 de la variable aléatoire $X_i \overline{Y_i}$.

Par suite, M_1 est égale à $E[X_i]E[\overline{Y_i}]$ en raison de l'indépendance a priori entre X_i et $\overline{Y_i}$ et M_2 est égale à $E[(X_i^2)]E[(\overline{Y_i})^2]$. On a donc : $M_1 = m_a m_{\bar{b}}$ et $M_2 = (v_a + m_a^2)(v_{\bar{b}} + m_{\bar{b}}^2)$ où $m_{\bar{b}} = 1 - m_b$. Il est aisé de démontrer que $v_{\bar{b}} = v_b$.

La propension (ou tendance) sera alors mesurée par l'écart entre ce qui est attendu sous l'hypothèse d'indépendance a priori entre les variables numériques X_i et Y_i et ce qui a été réellement observé à travers les réalisations a_i et b_i. Plus la moyenne des n observations $a_i(1-b_i)$ est petite par rapport à la moyenne attendue, plus la propension sera grande.

En utilisant le théorème de la limite centrale dite de Lindeberg-Lévy puisque l'indice est une moyenne de variables aléatoires indépendantes identiquement distribuées, on démontre que Z suit approximativement, pour n grand, la loi normale d'*espérance* $E(Z) = M_1$ et de *variance* $\frac{1}{n}(M_2 - M_1^2)$.

Autrement dit, la loi de la variable « indice de propension empirique » $\tilde{Q}(a,\bar{b}) = \dfrac{Z - M_1}{\sqrt{\frac{1}{n}(M_2 - M_1^2)}}$ est approximativement la loi gaussienne centrée réduite N(0 ; 1).

Plus explicitement, pour obtenir la valeur estimée de la réalisation de la variable « indice de propension empirique », notons m_a la moyenne des observations a_i et m_b la moyenne des observations b_i ; la moyenne M_1 est alors égale à $m_a(1-m_b)$, ou $m_a . m_{\bar{b}}$ où $m_{\bar{b}}$ est la moyenne des $\overline{b_i}$.

Notons également v_a la variance des a_i et v_b celles des b_i, le moment M_2 d'ordre 2 des observations a_i et b_i est égal à $(v_a + m_a^2)(v_b + m_{\bar{b}}^2)$ puisque les observations b_i et $\overline{b_i}$ ont la même variance.

L'indice empirique d'implication devient :

$$\tilde{q}(a,\bar{b}) = \frac{\frac{1}{n}\sum_{i\in E} a_i \bar{b}i - m_a m_{\bar{b}}}{\sqrt{\frac{v_a(v_b + m_{\bar{b}}^2) + m_a^2 v_b}{n}}}$$

Quant à l'estimation de l'intensité de propension, elle est encore obtenue par :

$$\varphi(a,b) = 1 - \Pr\left[\tilde{Q}(a,\bar{b}) \le \tilde{q}(a,\bar{b})\right] = \frac{1}{\sqrt{2\pi}} \int_{\tilde{q}(a,\bar{b})}^{\infty} e^{-\frac{t^2}{2}} dt$$

1.4 Restriction de l'indice de propension aux variables binaires et indice d'implication.

Si *a* et *b* sont deux variables binaires, les lois des variables aléatoires associées X_i et Y_i indépendantes sont celles de Bernoulli.

Par ailleurs, $n_a = \sum_{i \in E} a_i$ et $n_b = \sum_{i \in E} b_i$ et $n_{a \wedge \bar{b}} = \sum_{i \in E} a_i \bar{b}_i$.

Par suite, la réalisation de $M_1 = E(X_i)E(\bar{Y}_i)$ peut être estimée par $\frac{n_a}{n} \frac{n_{\bar{b}}}{n}$. D'autre part pour les variables de Bernoulli, X_i et Y_i, nous avons $X_i^2 = X_i$ et $\bar{Y}_i^{\,2} = \bar{Y}_i = 1 - Y_i$.

Ainsi $M_2 = E[X_i^2]E[\bar{Y}_i^{\,2}] = E[X_i]E[\bar{Y}_i]$ peut aussi être estimée par $\frac{n_a}{n} \frac{n_{\bar{b}}}{n}$. Donc les valeurs estimées de M_2 et M_1 sont égales à $\frac{n_a}{n} \frac{n_{\bar{b}}}{n}$. Par suite, il en résulte que l'estimation de $\frac{1}{n}(M_2 - M_1^2) = \frac{1}{n}(M_1 - M_1^{\,2})$ vaut :

$$\frac{1}{n}\left[\frac{n_a n_{\bar{b}}}{n^2} - \left(\frac{n_a n_{\bar{b}}}{n^2}\right)^2\right] = \frac{1}{n^2}\frac{n_a n_{\bar{b}}}{n^2}\left[1 - \frac{n_a n_{\bar{b}}}{n^2}\right]$$

Ainsi l'estimation de la réalisation de Z est : $\frac{1}{n}\sum_{i \in E} a_i \bar{b}_i = \frac{1}{n} n_{a \wedge \bar{b}}$ ce qui conduit à ce que les estimations respectives des réalisations des variables aléatoires $\tilde{Q}(a, \bar{b})$ et $Q(a, \bar{b})$ coïncident. Dit autrement,

$$\tilde{q}(a,\bar{b}) = \frac{\frac{n_{a \wedge \bar{b}}}{n} - \frac{n_a n_{\bar{b}}}{n^2}}{\frac{1}{n}\sqrt{\frac{n_a n_{\bar{b}}}{n}\left(1 - \frac{n_a n_{\bar{b}}}{n^{2)}}\right.}} = \frac{n_{a \wedge \bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}\left(1 - \frac{n_a n_{\bar{b}}}{n^{2)}}\right.}} = q(a,\bar{b})$$

Ce modèle adopté correspond strictement à une extension de la notion d'implication statistique fondée sur des variables binaires, à celle de propension statistique fondée sur les variables modales à valeurs dans [0,1]. En effet l'indice de propension statistique coïncide exactement avec l'indice d'implication statistique en considérant une variable binaire comme une variable modale dégénérée ne prenant que les valeurs extrêmes 0 et 1. Ceci permet d'utiliser dans C.H.I.C. les mêmes algorithmes que les variables soient binaires ou numériques.

2 L'ASI des variables sur intervalles et variables-intervalles

2.1 Situation fondatrice de l'ASI des variables sur intervalles

On recherche, par exemple, à extraire d'un ensemble de données biométriques, la règle suivante, en estimant sa qualité : « si un individu pèse entre 65 et 70 kg alors en général il mesure entre 1.70 et 1.76 m ». Une situation comparable se présente dans la recherche de relation entre des intervalles de performances d'élèves dans deux disciplines différentes. La situation plus générale s'exprime alors ainsi : deux variables réelles a et b prennent un certain nombre de valeurs sur 2 intervalles finis $[a_1 ; a_2]$ et $[b_1 ; b_2]$. Soit A (resp. B) l'ensemble des valeurs de a (resp. b) observées sur $[a_1 ; a_2]$ (resp. $[b_1 ; b_2]$. Par exemple ici, a représente les poids d'un ensemble de n sujets et b les tailles de ces mêmes sujets.

Deux problèmes se posent :

1° peut-on définir, et comment le faire, des sous-intervalles adjacents de $[a_1 ; a_2]$ (resp. $[b_1 ; b_2]$.) afin que la partition la plus fine obtenue respecte au mieux la distribution des valeurs observées dans $[a_1 ; a_2]$ (resp. $[b_1 ; b_2]$.) ?

2° peut-on trouver, et comment le faire, les partitions respectives de $[a_1 ; a_2]$ et $[b_1 ; b_2]$ constituées de réunions des sous-intervalles adjacents précédents, partitions qui maximisent l'intensité d'implication moyenne des sous-intervalles de l'un sur des sous-intervalles de l'autre appartenant à ces partitions ?

Nous répondons à ces deux questions dans le cadre de notre problématique en faisant choix des critères à optimiser pour satisfaire l'optimalité attendue dans chaque cas. A la première question, de nombreuses solutions ont été apportées dans d'autres cadres (par exemple : Lahanier-Reuter, 1998).

2.2 Caractérisation et propriétés des variables sur intervalles.

Approche du premier problème

On va s'intéresser à l'intervalle $[a_1 ; a_2]$ en le supposant muni d'une partition initiale triviale de sous-intervalles de même longueur, mais pas nécessairement de même distribution des fréquences observées sur ces sous-intervalles.

Notons $P_0 = \{A_{01}, A_{02},..., A_{0p}\}$, cette partition en p sous-intervalles. On cherche à obtenir une partition de $[a_1 ; a_2]$ en p sous-intervalles $A_{q1}, A_{q2},..., A_{qp}$ de telle façon qu'au sein de chaque sous-intervalle, on ait une bonne homogénéité statistique (faible **inertie intra-classe**) et que ces sous-intervalles présentent une bonne hétérogénéité mutuelle (forte **inertie inter-classe**). On sait que si l'un des critères est vérifié, l'autre l'est nécessairement en vertu de l'application du théorème de Koenig-Huyghens. Pour ce faire, nous adoptons une méthode directement inspirée de la méthode des **nuées dynamiques** conçue par Edwin Diday (1972), exposée par Lebart (Lebart et al. 2006) et que nous avons adaptée à la situation présente. Nous obtenons ainsi la partition optimale visée.

Approche du second problème

On suppose maintenant que les intervalles $[a_1 ; a_2]$ et $[b_1 ; b_2]$ sont munis de partitions optimales P et Q, respectivement, au sens des nuées dynamiques. Soit p et q les nombres respectifs de sous-intervalles composant P et Q. A partir de ces deux partitions, il est possible d'engendrer 2^{p-1} et 2^{q-1} partitions obtenues par réunions itérées de sous-intervalles adjacents respectivement de P et de Q [10].

On calcule les intensités d'implication respectives de chaque sous-intervalle réuni ou non à un autre de la première partition sur chaque sous-intervalle réuni ou non à un autre de la seconde, puis les valeurs des intensités des implications réciproques.

Il y a donc au total $2.2^{p-1}.2^{q-1}$ familles d'intensités d'implication, chacune d'entre elles nécessitant le calcul de tous les éléments d'une partition de $[a_1 ; a_2]$ sur tous les éléments d'une des partitions de $[b_1 ; b_2]$ et réciproquement.

On choisit comme *critère d'optimalité* la moyenne géométrique des intensités d'implication, moyenne associée à chaque couple de partitions d'éléments, réunis ou non, définies inductivement. On note les deux maxima obtenus (implication directe et sa réciproque) et on retient les deux partitions associées en déclarant que l'implication de la variable-sur-intervalle a sur la variable-sur-intervalle b est optimale lorsque l'intervalle $[a_1 ; a_2]$ admet la partition correspondant au premier maximum et que l'implication réciproque optimale est satisfaite pour la partition de $[b_1 ; b_2]$ correspondant au deuxième maximum.

Comme nous allons le voir dans le sous-chapitre suivant, cette approche est applicable à la notion de variable-intervalle étudiée par ailleurs par E. Diday et ses collaborateurs. En effet, les modalités sont nominales ou numériques. Dans le premier cas, il est possible de calculer les implications des modalités ou des réunions de modalités de l'une sur les modalités ou les réunions de modalités de l'autre, comme ci-dessus. Dans le second cas, il peut être intéressant et utile de redéfinir les partitions des variables sous-jacentes en optimisant ces partitions comme nous l'avons fait dans le premier problème posé dans le cadre des variables sur intervalles.

2.3 Situation fondatrice de l'ASI des variables-intervalles

On dispose des données fournies par une population de n individus (qui peuvent être chacun ou certains des ensembles d'individus, par ex. une classe d'élèves) selon p variables (par ex. notes sur une année en français, math, physique, …, mais aussi bien : poids, tailles, tour de poitrine, …). Les valeurs prises par ces variables selon chaque individu sont des intervalles de réels positifs. Par exemple, l'individu x donne la valeur [12 ; 15,50] à la variable note de math. E. Diday parlerait à ce sujet de p **variables symboliques** à valeurs intervalles définies sur la population.

On cherche à définir une implication d'intervalles, relatifs à une variable a, constitués eux-mêmes des intervalles observés, vers d'autres intervalles pareillement définis et relatifs à une autre variable b. Ceci permettra de mesurer l'association implicative, donc non

[10] Il suffit de considérer l'arborescence dont A_1 est la racine, puis de le réunir ou non à A_2 qui lui-même sera ou non réuni à A_3, etc. Il y a donc 2^{p-1} branches dans cette arborescence.

symétrique, de certain(s) intervalle(s) de la variable a avec certain(s) intervalle(s) de la variable b, ainsi que l'association réciproque à partir de laquelle on retiendra la meilleure pour chaque couple de sous-intervalles en jeu, comme nous l'avons fait avec les variables-sur-intervalles. Par exemple, on dira que le sous-intervalle [2 ; 5,5] de notes de mathématiques implique généralement le sous-intervalle [4,25 ; 7,5] de notes de physique, ces deux sous-intervalles appartenant à une partition optimale au sens de la variance expliquée des intervalles respectifs de valeurs [1 ; 18] et [3 ; 20] prises dans la population. De même, on dira que [14,25 ; 17,80] en physique implique le plus souvent [16,40 ; 18] en mathématiques.

2.4 Caractérisation et propriétés des variables-intervalles.

En suivant la démarche de E. Diday et ses collaborateurs, si les valeurs prises selon les sujets par les variables a et b sont de nature symbolique, en l'occurrence des intervalles de R^+, il est possible d'étendre les algorithmes ci-dessus (Gras, 2001a). Par exemple, à la variable a sont associés des intervalles de poids et à la variable b des intervalles de tailles, intervalles dus à une imprécision des mesures. Effectuant la réunion des intervalles I_X et J_X décrits par les sujets x de E selon respectivement chacune des variables a et b, on obtient deux intervalles I et J recouvrant toutes les valeurs possibles de a et de b. Sur chacun d'eux on peut définir une partition en un certain nombre d'intervalles respectant comme plus haut un certain critère d'optimalité. Pour cela, les intersections des intervalles tels que I_X et J_X avec ces partitions seront munies d'une distribution prenant en compte les étendues des parties communes. Cette distribution peut être uniforme ou d'un autre type discret ou continu. Mais ainsi, nous sommes ramenés à la recherche de règles entre deux ensembles de variables-sur-intervalles qui prennent leurs valeurs sur [0 ; 1] et à partir desquelles nous pouvons chercher les implications optimales.

Chapitre 4 : Extension de l'Analyse Statistique Implicative à des hiérarchies de règles[11].

1 Introduction

Les développements théoriques de l'analyse de données offrent des retombées enrichissantes pour l'E.C.D. (Extraction des Connaissances dans des Données) et sa vitalité n'est pas étrangère aux échanges induits. Par exemple, la construction d'indices permettant d'affecter une mesure non symétrique à des règles d'inférence partielle fournit des points d'application à l'extraction et à la représentation de règles d'association imprécises entre attributs binaires décrivant une population. Les démarches fondamentales convergent vers une problématique commune aux deux domaines ; il s'agit, avons-nous vu, de découvrir et de quantifier des règles non symétriques pour modéliser des relations du type *"si a alors presque b"*. C'est, par exemple, un objectif majeur des réseaux bayésiens (Pearl 1988) ou de certains travaux utilisant les treillis de Galois (Simon, 2000 ; Bernard 1999). Le plus souvent, notamment en Fouille de Données (Agrawal 1993), la probabilité conditionnelle est l'indice fondamental de l'association, même dans l'approche multivariée. Cela peut être aussi la corrélation comme dans (Brin 1997). Cependant à notre connaissance, d'une part, les développements s'arrêtent à la proposition d'un indice d'implication partielle pour des données binaires, et d'autre part, cette notion n'est pas étendue à l'extraction de règles de règles où les prémisses et les conclusions peuvent être elles-mêmes des règles. Nous proposons ici ces prolongements en formalisant la notion de hiérarchie orientée, en charge de la représentation graphique de la structuration de l'ensemble des variables selon ces règles de règles, donc d'un niveau conceptuel supérieur, dites règles généralisées. Cette formalisation peut nous rappeler la notion d'abstraction réfléchissante selon J. Piaget qui fait passer de la strate « objet » à celle des opérations sur les objets, à celle des opérations sur ces opérations, etc.. « Elle est réfléchissante aux deux sens suivants : elle transpose sur un plan supérieur de conceptualisation ce qu´elle emprunte à un palier précédent » écrit Sylvie Lucas dans Le Tome 52 du Bulletin de Psychologie, juillet-août 1999.

Rappelons qu'une représentation structurée des relations implicatives dans l'ensemble des variables instanciées a été obtenue, dans le cadre de l'A.S.I., par un graphe implicatif sans cycle pondéré par les intensités, fermé transitivement à un seuil donné. Il est utile dans des situations d'évolution, par exemple, en psychologie cognitive, pour interpréter des chemins de ce graphe, constitués de suites d'arcs qui lient certaines variables, en termes de genèses différentielles. Mais la hiérarchie que nous présentons maintenant va doubler les informations fournies par les règles d'association en organisant leur ensemble selon une structure ordonnée en méta-règles, en méta-méta-règles, etc..

[11] Une partie de ce chapitre a été publiée sous une forme voisine dans une sélection RNTI-C-1 des Actes des 11èmes Rencontres de la Société Francophone de Classification sous le titre : « Significativité des niveaux d'une hiérarchie orientée en analyse statistique implicative » (septembre 2004) par Régis Gras, Pascale Kuntz et Jean-Claude Régnier

2 Hiérarchie de classes de variables

Afin de soutenir l'intuition, nous baserons le modèle de la **hiérarchie orientée cohésitive** sur la métaphore linguistique suivante :

1. *les variables* (ou attributs) de l'ensemble V (card V=p) constitueront l'ensemble les *lettres de l'alphabet V,*
2. *les classes de k variables,* éléments de V^k, par exemple $(a_1, a_2,\ldots, a_k)$, constitueront les *syllabes du vocabulaire,*
3. *les classes maximales,* i.e. telles qu'aucune variable ne la complète, constitueront *les mots* du vocabulaire,
4. *l'organisation hiérarchique* de l'ensemble des classes constituera une *phrase,* structurée par des propositions incises.

D'autres métaphores peuvent illustrer le modèle que nous allons construire comme, par exemple, l'ensemble des séquences constituant le génome ou encore une théorie mathématique organisée en théorèmes et corollaires. Mais nous verrons que ces métaphores ne satisfont pas totalement le modèle.

On va également constater que ce modèle hiérarchique, où l'ordre intervient, ne s'apparente pas au modèle classique d'une hiérarchie ascendante, par exemple celle basée sur un indice de similarité entre attributs, car les classes d'une telle hiérarchie sont des sous-ensembles de variables et non pas des k-uplets.

2.1 Hiérarchie orientée. Définitions. Propriétés

Définition 7: On appelle hiérarchie orientée **H** sur l'ensemble des variables V, une suite d'arrangements (au sens de la combinatoire) des éléments de V, vérifiant les axiomes 1. 2 et 3 énoncés ci-dessous. Ces arrangements sont appelés classes de H.

Par exemple, {(j), (f,g), (b,c), (e,f,g), (b,c,d), (h,i), (a,b,c,d), (e,f,g,h,i)} est une hiérarchie orientée sur V={a,b,c,d,e,f,g,h,i} et (a,b,c,d) est une classe de cette hiérarchie.

Définition 8: On appelle classe C de degré k de la hiérarchie **H** un arrangement de k éléments de V appartenant à **H**. On notera $\prec$ la relation d'ordre induite sur **C** par le tirage d'un arrangement.

Par exemple, $(a_1, a_2,\ldots, a_k)$, pour $k \leq p$, est une classe de degré k et $a_1 \prec a_2 \prec \ldots \prec a_k$. Mais également, par convention, (a) est une classe de degré 1. Elle est dite élémentaire

Définition 9: On appelle **troncation de C**, tout sous-arrangement des éléments de **C** respectant la structure d'ordre $\prec$ et la consécutivité.

Par exemple, si **C** = $(a_1, a_2,\ldots, a_k)$, la classe **C'** = $(a_i, a_{i+1},\ldots, a_j)$ où $1 \leq i$ et $j \leq k$, est une troncation de **C**.

Définition 10: On note **C'** $\widehat{\subset}$ **C** si et seulement si **C'** est une troncation de **C**. Une classe est dite maximale s'il n'existe pas de classe qui la contienne dans **H**. Elle est dite minimale si elle ne contient aucune classe de la hiérarchie **H**. En particulier, une classe élémentaire est donc minimale (mais elle peut être aussi maximale).

On peut comparer, sans la confondre cependant, cette relation à l'inclusion ensembliste. Dans l'exemple initial, les classes (a,b,c,d), (e,f,g,h,i) et (j) sont maximales. Cette dernière est aussi minimale.

Définition 11: La trace de **C** sur **C'** est constituée d'éléments communs à **C** et **C'** et elle respecte la structure d'ordre $\prec$ et la consécutivité. La trace est une opération commutative notée $\widehat{\cap}$.

Ainsi on peut comparer, sans la confondre, cette opération à l'intersection ensembliste.

Par exemple, (d,f,g,a,e) $\widehat{\cap}$ (f,g,a,e,b,h)= (f,g,a,e)

Définition 12: Si les deux classes quelconques **C'** et **C''** ont une trace vide (**C'** $\widehat{\cap}$ **C''**= $\varnothing$), la concaténation de **C'** et **C''** notée **C'** $\widehat{\cup}$ **C''** est la classe **C** dont les éléments appartiennent à **C'** et **C''** et à elles exclusivement. Elle respecte les ordres au sein de **C'** et **C''**et le plus grand élément de **C'** précède le plus petit de **C''**. On dira que **C'** et **C''** sont des classes génératrices de **C'** $\widehat{\cup}$ **C''.**

Cette opération, comparable à la concaténation ordinaire, ainsi qu'à la réunion ensembliste sans se confondre avec elle, est non commutative.

Par exemple, si **C'** = (d,f,g,a) et **C''** = (b,u,r,p,y), **C'** $\widehat{\cup}$ **C''** = (d,f,g,a,b,u,r,p,y), alors que **C''** $\widehat{\cup}$ **C'** = (b,u,r,p,y,d,f,g,a)

2.2 Axiomes d'une hiérarchie orientée

Axiome 1 : $\forall$ **C** et $\forall$ **C'** classes de **H** , **C'** $\widehat{\cap}$ **C''** $\in$ **{** $\varnothing$, **C, C'}**

Axiome 2 : $\forall$ **C** $\in$ **H**, si **C** n'est pas élémentaire ou minimale, elle est la concaténation de classes de **H**

Axiome 3 : Il existe une permutation des éléments de V qui coïncide avec la concaténation de toutes les classes maximales de **H**

2.3 Algorithme de construction de l'ensemble des classes

Nous définissons ci-dessous un critère algébrique en vue de nous permettre de construire de façon ascendante, la hiérarchie organisatrice de l'ensemble V des variables et qui respecte les trois axiomes d'une hiérarchie orientée.

2.3.1 Critères algébriques

Définition 13: La **cohésion d'une classe** de degré 2, correspondant au couple (a,b) est définie, à partir de l'entropie au sens de Shannon, par la formule,

$coh(a,b) = \left(1 - \left[-p(\log_2(p)) - (1-p)(\log_2(1-p))\right]^2\right)^{\frac{1}{2}}$ où p=φ(a,b)≥0,50 et $coh(a,b) = 0$ si p=φ(a,b)<0,50.

Définition 14: La cohésion d'une classe **C** = (a_1, a_2,…, a_r) de degré r, est définie par la formule : $coh(C) = \left[\prod_{i=1,...,r-1}^{j=2,...,r;\, j>i} coh(a_i, a_j)\right]^{\frac{2}{r(r-1)}}$

Définition 15: La cohésion d'une classe **C** = (a) de degré 1, est définie par $coh(a) = 1$

2.3.2 Algorithme de construction de la hiérarchie

Niveau 0
Les classes sont élémentaires et toutes les cohésions sont égales à 1

Niveau 1
On compare toutes les cohésions des arrangements 2 à 2 de V.
On conserve celle, notée C_1, qui correspond au maximum, soit par ex. C_1=(a,b).

Définition 16: On appelle **nœud** 1, la règle $a \Rightarrow b$

Niveau 2
On compare toutes les cohésions des classes à 2 éléments, sauf C1, à celles des classes à 3 éléments du type (x, a, b) et (a, b, x).
On conserve celle, notée C_2, correspondant au maximum obtenu.
Le nœud 2 sera :

1. soit une classe à 2 éléments, et dans ce cas le nœud sera du type : c=>d
2. soit une classe à 3 éléments, et dans ce cas le nœud sera noté (a=>b)=>c ou c=>(a=>b).

Ces dernières règles sont dites **composées ou généralisées ou R-règles**. Pour restituer l'ordre dans lequel est constituée la classe, on notera, par exemple ici, ((a,b),c) ou, dans l'autre cas, (c,(a,b)).

Niveau k
On compare toutes les cohésions des concaténations de 2 des classes déjà formées aux niveaux inférieurs, du type C_i et C_j avec i<k et j<k. On conserve celle C_k=$C_i \cup C_j$ qui satisfait le maximum et est la concaténation de C_i et C_j.
Le nœud k correspondant sera noté par extension C_i=>C_j. Mais on peut expliciter des nœuds correspondant à la formation des troncations respectives et génératrices de C_i et C_j aux niveaux inférieurs.
Par exemple, la règle composée $((f \Rightarrow (e \Rightarrow u)) \Rightarrow ((a \Rightarrow b) \Rightarrow (c \Rightarrow d))$ est l'explicitation d'un nœud particulier. Afin de faire apparaître les classes formées à des niveaux successifs, on notera aussi la règle sous la forme : ((f(eu))((ab)(cd)))

L'algorithme s'arrête, au plus tard, au niveau p-1, lorsque toute concaténation conduirait à une classe de cohésion nulle ou à une permutation de V. Les classes formées au niveau ultime et qui n'admettent pas de classes qui les contiennent sont donc maximales. Certaines classes maximales peuvent aussi être élémentaires. La hiérarchie est composée de l'ensemble des classes maximales et de toutes leurs parties.

2.4 Conformité de la construction aux axiomes d'une hiérarchie orientée

La hiérarchie ainsi construite vérifie les trois axiomes d'une hiérarchie orientée. En effet :

Axiome 1 :
Deux classes de **H**, **C'** et **C''** étant données,

1. ou bien elles sont associées dans une même concaténation et la constituent entièrement, alors $\mathbf{C'} \hat{\cap} \mathbf{C''} = \varnothing$
2. ou bien l'une est la concaténation de l'autre et d'une troisième et alors $\mathbf{C'} \hat{\subset} \mathbf{C''}$ ou $\mathbf{C''} \hat{\subset} \mathbf{C'}$
3. ou bien elles ne sont pas associées dans une concaténation et alors elles sont des arrangements sans élément commun, donc $\mathbf{C'} \hat{\cap} \mathbf{C''} = \varnothing$

Axiome 2 :
$\forall C \in H$

1. ou bien elle est constituée d'un élément et c'est une classe élémentaire
2. ou bien elle est constituée de plus d'un élément et elle est alors la concaténation de deux ou plusieurs classes par construction.

Axiome 3 :
On range toutes les classes maximales par ordre croissant de la cohésion ; les classes élémentaires seront les éléments maximaux de cet ordre. Toutes les classes sont 2 à 2 disjointes et tous les éléments de V appartiennent à l'une et l'une seulement des classes. La concaténation de leur ensemble constitue alors une permutation particulière de tous les variables de V.
Notons qu'à une permutation de V peuvent correspondre plusieurs hiérarchies.

Exemple 1: Si l'on range les classes maximales de la hiérarchie donnée au début du texte par ordre croissant de la cohésion, on obtient par exemple :
$coh(e,f,g,h,i) \leq coh(a,b,c,d) \leq coh(j)$ et (e,f,g,h,i,a,b,c,d,j) est une permutation de V.
Mais à cette permutation, peut aussi correspondre la hiérarchie :
{(g,h), (b,c), (e,f), (a,b,c), (g,h,i), (d,j), (a,b,c,d,j)} dont les classes maximales sont (e,f), (g,h,i) et (a,b,c,d,j).

Exemple 2 : Reprenant encore l'exemple initial, l'autre hiérarchie {(j), (f,g), (b,c), (e,f,g), (b,c,d), (h,i), (a,b,c,d), (h,i,e,f,g)}ne coïncide pas avec la première. La permutation correspondante de V est (h,i,e,f,g,a,b,c,d,j).

La figure ci-dessous montre la hiérarchie obtenue, artificiellement, à partir de 7 variables. Des interprétations de telles règles généralisées sont quelquefois complexes, comme par exemple, la règle $(x \Rightarrow (y \Rightarrow z)) \Rightarrow (t \Rightarrow v)$. Mais quelques règles sont réductibles à des assemblages plus aisément interprétables. Par exemple, la règle $x \Rightarrow (y \Rightarrow z)$ se ramènerait, dans le cas formel, à $x \wedge y \Rightarrow z$. La règle $(d \Rightarrow b) \Rightarrow (a \Rightarrow f)$ ou ((db)(af)), illustrée par la figure (FIG 10), peut s'interpréter comme : le « théorème » $d \Rightarrow b$ a généralement pour conséquence le « théorème » $a \Rightarrow f$. Cette figure montre aussi que la variable *e* n'a ni prémisse ni conclusion.

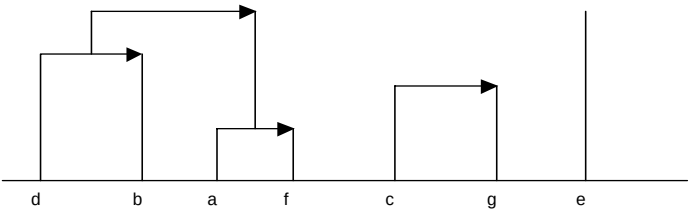


FIG. 10 *Un exemple de hiérarchie orientée*

3 La hiérarchie cohésitive basée sur une distance ultramétrique

On part de l'algorithme de construction de **H** déjà défini et par lequel la cohésion de la classe en voie de formation, à un niveau donné, est inférieure à la valeur de la cohésion au niveau immédiatement inférieur et est supérieure à celle du niveau immédiatement supérieur. C'est donc une fonction décroissante des niveaux et, a fortiori, avec l'inclusion des parties de **H**. C'est la propriété de la cohésion qui va permettre de définir la **distance ultramétrique** d, pour laquelle tous les triangles sont isocèles. Cette propriété d'ultramétricité de la cohésion va justifier, a posteriori, le bien-fondé de l'expression « hiérarchie » employée pour la construction.

Il suffit de choisir pour un couple quelconque de variables (x, y) :

$$d(x,y) = 1 - coh(C_{(x,y)})$$

où $coh(C_{(x,y)})$ est la cohésion de la plus petite classe $C_{(x,y)}$ contenant x et y. Rappelons les propriétés suivantes de la hiérarchie :

1. quelles que soient les classes C et C' de H, ou bien $C \hat{\subset} C'$ ou bien $C \hat{\supset} C'$ ou bien $C \hat{\cap} C' = \varnothing$
2. si $C \hat{\subset} C'$, alors $coh(C) \geq coh(C')$ par construction.

Vérifions alors que les trois axiomes d'ultramétricité sont bien valides sur l'ensemble V des variables

Axiome 1 :

Pour tout $x \in V$, d(x,x) = 0 par construction de d car $coh(x,x) = 1$

Axiome 2 :

Pour tout couple (x,y)∈AxA, d(x,y)=d(y,x) par construction

Axiome 3 :

(x,y,z)∈AxAxA, d(x,y)≤sup[d(x,z), d(y,z)] (1)

La définition adoptée respecte aussi l'axiome 3. En effet, soit $C_{x,y}$, $C_{x,z}$, $C_{y,z}$ les plus petites classes contenant respectivement x et y, x et z, y et z.

Alors $z \in C_{(x,z)} \hat{\cap} C_{(y,z)} \neq \varnothing$ d'où $C_{(x,z)} \hat{\subset} C_{(y,z)}$ ou bien $C_{(x,z)} \hat{\supset} C_{(y,z)}$ d'après les propriétés des classes de H. Supposons alors: $C_{(x,z)} \hat{\supset} C_{(y,z)}$. On en déduit d(x,z)≥d(y,z) et par suite d(x,z)=sup[d(x,z) ; d(y,z)] (2). De plus, on a à la fois : $x \in C_{(x,z)}$ et , par conséquent $C_{(x,y)} \hat{\subset} C_{(x,z)}$

Comme l'indice d croît avec l'inclusion en raison de la décroissance de la cohésion, alors : d(x,z)≥d(x,y) (3)

Par (2) et (3) on obtient donc (1) :d(x,y)≤sup[d(x,z), d(y,z)]

H est donc bien une **hiérarchie indicée** par la distance d au sens strict de hiérarchie mathématique.

4 Présentation d'une application de l'approche par hiérarchie cohésitive

Une enquête de l'Association française des Professeurs de Mathématiques de l'Enseignement Public (APMEP) a été proposée récemment aux professeurs de mathématiques de classes terminales de l'enseignement secondaire dans différentes filières : S (à dominante sciences dures) et ES (à dominante scientifique et sociale), littéraires L et technologique T. Ces variables constitueront des variables supplémentaires de l'analyse, donc n'entrent pas directement dans la constitution des règles. Nous avons recueilli et analysé (Bodin et Gras 1999) les réponses de 311 professeurs, à des questions portant sur les objectifs (15 ont été retenus) qu'ils assignent à leur enseignement et sur leurs opinions relatives à 11 phrases susceptibles d'être communément énoncées. Des objectifs jugés non pertinents sont ajoutés aux variables supplémentaires (cf. chap. 5) « filières ». Nous présentons ci-après le questionnaire mis dans un format conforme au présent ouvrage :

QUESTIONNAIRE -Professeurs de Terminale

Q1- Au nom de quelle série répondez-vous ? :…….

(*Si plusieurs séries, utiliser un questionnaire par série*)

Q2-Objectifs de la formation mathématique

A votre avis, quels sont les objectifs essentiels de la mission d'un professeur de mathématiques dans la série pour laquelle vous répondez ?

(*Choisir 6 objectifs et ranger par ordre préférentiel décroissant de 1 à 6*)

Code	*Objectifs*	Code	*Objectifs*
A-	acquisition de connaissances	B-	préparation à la vie professionnelle
C-	préparation à la vie civique et sociale	D-	préparation aux examens, concours, au passage dans l'enseignement supérieur
E-	développement de l'imagination et la créativité	F-	développement de la capacité à prouver et valider sa preuve
G-	développement de la capacité d'accepter des points de vue différents	H-	développement de la volonté et la persévérance
I-	développement de l'esprit critique	J-	développement de la capacité à communiquer avec objectivité, clarté et précision par des modes de représentation divers
K-	développement de compétences utiles dans les autres disciplines	L-	développement de la pratique de calculs formels, donc sans nécessité de signification
M-	développement de la capacité à mathématiser et à formaliser	N-	acquisition de savoir-faire
O-	participation au développement d'une culture générale		

Rang	1	2	3	4	5	6
Objectif						

Les objectifs ci-dessus vous paraissent-ils pertinents (PER): OUI NON (entourez votre choix)
Si non, précisez lesquels en utilisant le codage proposé

Q3- Votre opinion sur des opinions

Voici quelques opinions entendues dans la salle des professeurs. Entourez votre choix :

Code	**Opinion** 1= D'accord ; 2= Peu d'accord ; 3= Pas d'accord	**Choix**
OP1	C'est vrai que les math constituent un instrument de sélection excessif.	1 2 3
OP2	Au bac, je préfère qu'il y ait un grand problème avec plusieurs parties plutôt qu'un ensemble de petits problèmes indépendants.	1 2 3
OP3	Dans ma notation, j'attache plus d'importance à la démarche qu'au résultat.	1 2 3
OP4	Quand je corrige, j'aime bien un barème très détaillé sur les résultats à obtenir.	1 2 3
OP5	La démonstration est la seule façon rigoureuse de faire des mathématiques.	1 2 3
OP6	Je préfère des programmes bien définis indiquant ce que je dois et ce que je ne dois pas faire.	1 2 3
A la sortie de la terminale de la série sur laquelle vous répondez, un élève devrait pouvoir ou avoir...		
OP7	reconnaître si un nombre entier écrit dans la base 10 est divisible par 4	1 2 3
OP8	donner un exemple ou un contre-exemple personnels à l'affirmation : "si deux applications *f* et g sont strictement croissantes sur un intervalle, l'application produit *fxg* y est également croissante".	1 2 3
OP9	avoir appris à faire un test statistique pour pouvoir réfuter ou accepter l'hypothèse d'adéquation d'une loi théorique à une distribution empirique	1 2 3
OPX	estimer à vue, à 30% près, le périmètre et l'aire du plancher ainsi que le volume de la salle de classe	1 2 3

Les 26 variables correspondantes ne sont pas binaires, mais ordinales, elles prennent des valeurs décimales sur [0 ;1]. Ainsi l'analyse intègre l'intensité des attitudes, le choix prioritaire d'un objectif pondérant de façon différente un choix plus secondaire, voire non retenu. Pour ce faire, les enseignants font choix de 6 objectifs parmi 15 qu'ils assignent à leurs enseignements (ex : A : « Acquisition de connaissances », B : « Préparation à la vie professionnelle ») et d'opinions relatives à dix phrases communément énoncées (par exemple : OP 1 : « les maths constituent un instrument de sélection excessif ») (Bodin et Gras, 1999). Les poids des 26 variables figurent dans le tableau ci-dessous, compte tenu des pondérations décimales accordées aux variables ordinales (rangs de 6 choix pondérés par 1, 0.8, 0.6, etc. et accords modulés 1, 0.5, 0 suivant l'accord avec les opinions).

A	B	C	D	E	F	G	H		
105.7	8.8	9.7	140.0	21.8	138.7	19.5	44.8		
I	J	K	L	M	N	O	PER		
83.1	108.4	77.6	4.6	90.2	66.6	33.2	254		
OP1	OP2	OP3	OP4	OP5	OP6	OP7	OP8	OP9	OPX
81.5	147.5	242.5	229.0	190.0	240.0	200.0	165.0	98.0	207.0

TAB. 12 – *Occurrences des variables de l'enquête sur les professeurs de mathématiques*

La hiérarchie orientée obtenue structure les 26 variables en plusieurs classes qui définissent des R-règles de longueur, d'interprétation et d'intérêt variés. Une aide à l'interprétation peut être apportée si l'on se souvient de la tautologie en logique formelle :

$$(a \Rightarrow (b \Rightarrow c)) \Leftrightarrow ((a \wedge b) \Rightarrow c)$$

De plus, relativement à chaque classe maximale, le logiciel C.H.I.C. indique quelle variable supplémentaire contribue le plus à la formation de la classe. Cette information permet d'améliorer la compréhension et la signification de la classe.

Voici une partie de la hiérarchie où nous nous limitons dans un souci de clarté à trois classes maximales. Le logiciel CHIC fournit cette hiérarchie par une symétrie orthogonale par rapport à sa base.

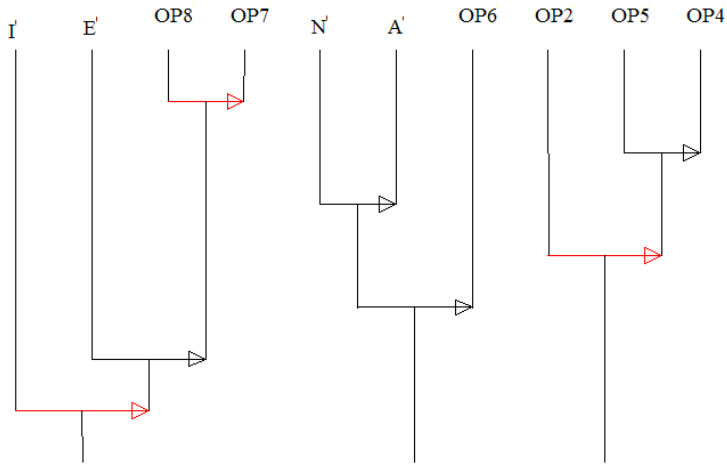


FIG. 11 *Hiérarchie orientée pour le questionnaire*

Des règles généralisées apparaissent sur cette hiérarchie et sont significatives :

- **(si N alors A) alors OP6**. Cette R-règle peut être lue ainsi : « si l'acquisition de savoir-faire (N) doit s'accompagner de celle des connaissances (A), alors le professeur demande que les programmes soient bien définis (OP6). On peut faire l'hypothèse que cette relation soit de type causal : « je suis très attaché aux contenus d'enseignement, mais inscrits dans l'institution, je demande que celle-ci définisse et précise ses choix ». Autrement dit, toute centration sur les savoirs exige un encadrement préalable de l'institution à travers des programmes. On observe alors que la règle généralisée permet de donner une signification plus synthétique aux règles qui la constituent : on passe des comportements, au sens behavioriste, à une conduite supérieure qui piloterait le comportement. Les enseignants qui considèrent que n'est pas pertinent l'objectif C ("contribuer à la préparation à la vie civique et sociale") sont les principaux responsables de cette règle ; ces enseignants possèdent donc une représentation de la formation mathématique très fermée sur la matière, dont l'enseignement, conformément aux programmes, n'est pas discutable.

- **si OP2 alors (si OP5 alors OP4).** Explicitement, cette R-règle se lit de la façon suivante : si l'on considère un grand problème indispensable à l'examen (OP2), alors, considérant que la démonstration est le seule façon rigoureuse de faire des mathématiques (OP5), le barème de correction doit être précis. On a là une évidente relation de type causal induite par une conception d'une catégorie d'enseignants très soumise à l'institution et conservatrice dans ses choix pédagogiques. Ne soyons pas surpris, la démonstration en France est le fondement de l'activité mathématique (pays de Descartes), tout en étant difficile à évaluer ; le grand problème en est le critère d'évaluation. On retrouve ici une image plus synthétique des règles d'association qui sous-tendent la R-règle, à savoir une conception de l'enseignement très classique qui exige un soutien institutionnel explicite et libérateur. Les enseignants de ES contribuent à cette association plus que les autres enseignants. Nous

reparlerons dans le chapitre suivant du critère dit de contribution qui nous permet d'établir cette affirmation.

-si I alors (si E alors (si OP8 alors OP7)) qui peut s'interpréter ainsi : si un enseignant choisit I (développement de l'esprit critique) et E (imagination et créativité) alors en général il considère que pour que l'élève découvre un caractère de divisibilité par 4 (OP7), il suffit qu'il ait été entraîné à trouver lui-même exemple et contre-exemple (OP8). D'une philosophie éducative ouverte vont découler des savoirs spécifiques laissant l'élève en situation de découverte, de construction personnelle. Cette règle est d'ailleurs constituée à un niveau significatif de la hiérarchie. Ce sont les enseignants des classes S qui contribuent le plus à son instauration. Elle met l'accent sur la relation entre des comportements non dogmatiques de l'enseignant et, en conséquence, la volonté de placer l'élève en situation de recherche personnelle. Ainsi, nous pouvons interpréter cette R-règle comme l'indice d'une conception d'ouverture didactique.

Insistons sur l'accroissement, dans chacun de ces cas, de la richesse de l'analyse obtenue par l'association des règles d'association en des R-règles. Ce ne sont plus seulement des faits ou des comportements isolés qui sont extraits, mais plutôt des conduites générales, révélatrices elles-mêmes de phénomènes plus globaux, moins singuliers ou de représentations psychologiques profondes. Une typologie comme en fournissent les classifications traditionnelles (donc symétriques), ne pourrait pas rendre compte de la dynamique des faits ou comportements sous-jacents. C'est pourtant cette dynamique restituée par les règles généralisées qui, appuyée sur des nécessités (les prémisses des règles), conduit à des élucidations vives d'un fragment de théorie, éventuellement en voie de construction.

En conclusion, quelles que soient les difficultés d'interprétation des règles généralisées, on constate le changement significatif d'information fournie par les classes orientées formées par rapport aux seules règles binaires de type $a \Rightarrow b$. Il s'agit ici de plonger l'ensemble des attributs dans une sorte de théorie globale, organisatrice des attributs où à chaque nœud on définit un théorème complexe dont la signification permet d'entrer plus profondément dans l'extraction des cohérences locales.

La construction d'une hiérarchie de règles généralisées par l'analyse implicative est implémentée dans le logiciel C.H.I.C.. Outre sa construction, un indice statistique, inspiré par celui que I.C. Lerman (1981b) a défini pour la hiérarchie des similarités, permet de mesurer la qualité de la hiérarchie globalement (ensemble des classes) et localement (nœuds). Nous avons construit un nouvel indice sur d'autres bases plus combinatoires (voir § 5 suivant).

Afin de permettre à l'expert qui doit faire l'analyse de la hiérarchie en intégrant la sémantique des variables, il semble absolument nécessaire de lui indiquer quelles sont le classes les plus pertinentes de la hiérarchie. C'est-à-dire celles où il doit porter son attention maximale, où l'interprétation serait plus cohérente avec le phénomène statistique qui a conduit, par les concepts jugés pertinents dans la théorie, à la formation de la classe. C'est dans cette intention que nous avons ensuite porté notre attention modélisante sur la construction successive des niveaux de la hiérarchie et, pour ce faire, défini un critère de significativité.

5 Significativité des niveaux d'une hiérarchie orientée

5.1 Critère de cohérence des niveaux

Une classe C de la hiérarchie orientée H formée au niveau k est considérée comme ***cohérente*** pour un seuil α, s'il y a conformité ou quasi-conformité au seuil α entre l'ordre – ou le préordre- ω_0 dans lequel s'organisent les attributs de C selon la cohésion et l'ordre –ou le préordre- théorique ω_t défini par leurs intensités d'implication mutuelles. Pour évaluer précisément cette conformité, nous nous basons sur une propriété de l'intensité d'implication (Gras et Larher, 1992): si le nombre d'occurrences de a_i est inférieur au nombre d'occurrences de a_j, alors la qualité de $a_i \Rightarrow a_j$ au sens de φ est meilleure que celle de sa réciproque $a_j \Rightarrow a_i$. Ainsi, l'ordre théorique ω_t défini par les intensités d'implication mutuelles coïncide avec celui défini par les occurrences des attributs. Nous comparons la conformité entre ω_0 et ω_t avec celle entre un ordre aléatoire ω^* et ω_t. Nous mesurons la conformité par le nombre d'inversions entre les différents ordres : i est le nombre d'inversions observées entre ω_0 et ω_t et I est le nombre d'inversions entre ω^* et ω_t. Le nombre d'inversions entre deux ordres est simplement défini ici par le nombre de paires d'attributs (a_i, a_j) telles que a_i est avant a_j dans le premier ordre et après dans le second.

Intuitivement cela signifie que, si α est petit, la conformité entre ω_0 et ω_t est invraisemblablement très grande puisqu'il paraît exceptionnel que le hasard « fasse mieux » que ce qui est observé.

Définition 17: La ***cohérence*** *o(C)* d'une classe C d'une hiérarchie orientée est définie par la probabilité $Pr(I > i)$.

Ainsi, plus le nombre d'inversions est faible, eu égard à la cardinalité de la classe, plus grande est la cohérence de la classe. De plus, pour un même nombre d'inversions observées pour deux classes *C'* et *C''*, si la classe *C'* contient plus de variables que la classe *C''*, la cohérence de *C'* est meilleure que celle de *C''*.

Exemple 1 : Considérons une classe C d'une hiérarchie orientée H constituée de cinq variables, a_i, i =1 à 5 structurée selon l'ordre $\omega_0 = \{a_1, a_4, a_3, a_2, a_5\}$. On suppose d'autre part que leurs occurrences sont telles que $n_{a_1} < n_{a_2} < \ldots < n_{a_5}$, ce qui induit selon la propriété rappelée ci-dessus, un ordre théorique $\omega_t = \{a_1, a_2, a_3, a_4, a_5\}$ pour les intensités d'implication. On vérifie aisément que le nombre d'inversions i entre ω_0 et ω_t est *3* (échanges de a_3 et a_2, a_4 et a_3, a_4 et a_2). Afin d'évaluer la cohérence de la classe, il faut déterminer la loi de la variable I. Pour 5 variables on peut obtenir pas à pas la distribution en énumérant les cas où chacune des variables est minimale dans l'ordre ω^* (TAB.13). Ici, on a :

$$\Pr(I > i) = \Pr(I > 3) = \frac{91}{120} \approx \frac{3}{4}$$

Nombre d'inversions	0	1	2	3	4	5	6	7	8	9	10
a_1 minimal	1	3	5	6	5	3	1	0	0	0	0
a_2 minimal	0	1	3	5	6	5	3	1	0	0	0
a_3 minimal	0	0	1	3	5	6	5	3	1	0	0
a_4 minimal	0	0	0	1	3	5	6	5	3	1	0
a_5 minimal	0	0	0	0	1	3	5	6	5	3	1
Total des permutations	*1*	*4*	*9*	*15*	*20*	*22*	*20*	*15*	*9*	*4*	*1*

TAB. 13 – *Détermination de la distribution de I dans l'exemple*

D'une façon générale, la mise en œuvre de la cohérence définie en 3.1. nécessite de déterminer la loi de *I*, que nous noterons I_m dans la suite puisqu'elle dépend du nombre *m* de variables. Notons que le recours à la variable aléatoire I_m donnant le nombre d'inversions entre deux permutations est présent dans le calcul du coefficient de corrélation des rangs τ de Kendall qui peut effectivement s'écrire

$$1-\tau=\frac{4I_m}{m(m-1)} \qquad \textbf{(1)}$$

Cependant, à notre connaissance, la loi de I_m n'est ni explicitement donnée ni formalisée (Kendall et Stuart, 1991). Nous proposons et établissons, dans la suite, une formule de récurrence permettant de calculer ses valeurs dans l'indice de cohérence.

5.2 Loi de la variable I_m, nombre d'inversions dans une permutation

Sous l'hypothèse d'équiprobabilité des permutations, nous considérons la variable aléatoire $N(I_m(k))$ donnant le nombre total de permutations aléatoires correspondant à un nombre d'inversions avec ω_t égal à *k* pour un nombre de variables égal à *m*. Notons que l'on a trivialement $Pr(I_m = 0) = 1/m\ !$ puisque le nombre d'inversions est nul si et seulement si ω_t coïncide avec ω^*.

Proposition 7 Pour tout $k < m$, on a

$$N(I_m(k)) = \sum_{j=0}^{k} N\left(I_{m-1}(j)\right) \qquad \textbf{(2)}$$

et, pour tout $k \geq m$, on a

$$N(I_m(k)) = \sum_{j=k-m+1}^{k} N\left(I_{m-1}(j)\right) \quad \textbf{(3)}$$

Preuve.

Remarquons tout d'abord, que pour tout *k*, on a

$$N\left(I_m(k)\right) = \sum_{i=1}^{m} N\left(I_m(k); a_i\right) \qquad \textbf{(4)}$$

où $N(I_m(k) ; a_i)$ est le nombre total de permutations lorsque la variable a_i est minimale dans l'ordre aléatoire ω^* et placé au $i^{ème}$ rang dans l'ordre théorique ω_t.

Supposons maintenant que k < m. Pour tout i de 1 à m, la place minimale de a_i dans entraîne (i-1) inversions ; les k-(i-1) autres inversions sont donc provoquées par les m–1 autres variables. Ainsi, pour tout i de 1 à (k +1) on a

$$N(I_m(k)) = \sum_{i=1}^{k+1} N(I_{m-1}(k-(i-1)) = \sum_{i=0}^{k} N(I_{m-1}(k-i)) = \sum_{j=0}^{k} N(I_{m-1}(j)) \quad \textbf{(5)}$$

et, pour tout i de k+2 à m, $N(I_m(k) ; a_i) = 0$ puisque dans ce cas la variable a_i étant minimale il y a au moins i-1 inversions auxquelles aucune permutation ne peut conduire.

La preuve de la formule de récurrence pour le cas k≥l est basée sur un raisonnement similaire.

Les relations (2) et (3) de la proposition précédente permettent de calculer les lois des variables I_m selon une formule de récurrence. En effet, connaissant la distribution de I_{m-1} on peut déterminer celle de I_m, et les valeurs initiales sont directement calculables. On vérifie aisément que $N(I_2(0) ; a_1) = 1$ (c'est la permutation de a_1 et a_2), $N(I_2(1) ; a_1) = 0$; $N(I_2(0) ; a_2) = 0$, $N(I_2(1) ; a_2) = 1$, d'où $N(I_2(0)) = 1$ et $N(I_2(1)) = 1$. On en déduit ainsi la loi de I_2 : $Pr(I_2 = 1) = Pr(I_2 = 0) = 0.5$, puis celle de I_3, etc..

Proposition 8 : Pour tout k<m, on a $N(I_m(k))=N(I_{m-1}(k))+N(I_m(k-1))$ et, pour tout k ≥m, on a $N(I_m(k))=N(I_m(k-1))+N(I_{m-1}(k))-N(I_{m-1}(k-m))$.

Cette proposition se déduit d'arguments similaires à ceux employés dans la proposition précédente (Gras, 1997 a) et de la relation (4).

On peut ainsi déduire de façon récurrente les différentes valeurs de la loi de I_m utiles pour le calcul de la cohérence. En effet, pour k=1 et m>1, on déduit l'équation linéaire aux différences d'ordre 1 en m suivante : $N(I_m(1))= N(I_m(0))+N(I_{m-1}(1))=1+N(I_{m-1}(1))$.

D'où, $N(I_m(1))-N(I_{m-1}(1))=1$ dont la solution avec second membre est $N(I_m(1))=m-1$.

Par conséquent, rappelant que l'on a pour tout m>1, $Pr(I_m=0)=1/m$! on obtient :

$$Pr(I_m = 1) = \frac{m-1}{m!} \quad \textbf{(6)}$$

Ainsi, par exemple si m = 2, la cohérence associée à une situation sans inversion est :

$$Pr(I_2>0)=Pr(I_2=1)=0{,}5.$$

De la même façon, pour m > 2, on obtient en utilisant d'une part, le résultat donnant $N(I_m(1))$ et d'autre part, le fait que $N(I_m(1 ; a_m))=0$, la relation $N(I_m(2))=N(I_{m-1}(2))+m-1$. D'où, $N(I_m(2)) = m^2/2 – m/2 –1$, et pour m > 2, on a donc

$$Pr(I_m = 2) = \frac{m^2 - m - 2}{2m!} \quad \textbf{(7)}$$

Puis, comme $N(I_m(3))=N(I_{m-1}(3))+\frac{m^2}{2}-\frac{m}{2}-1$, on obtient pour m> 3 :

$$Pr(I_m(3)) = \frac{m^2 - 7}{6(m-1)!} \quad \textbf{(8)}$$

Par exemple, on a $Pr(I_6 \leq 2) = 0.028$. Par suite, $Pr(I_6 > 2)=0.972$ est la valeur de la cohérence d'une classe de *6* variables dans laquelle on observerait *2* inversions entre l'ordre associé à la classe ω_0 et l'ordre théorique ω_t .

Remarque 1. Pour une classe *C* réduite à un singleton, sa cohérence ne peut se déduire des relations précédentes. Nous posons dans ce cas *o(C)*=0,5 du fait que l'absence d'inversion n'apporte aucune information puisqu'elle est nécessaire.

Remarque 2. Lorsque deux variables d'une classe *C* ont le même nombre d'occurrences et ont donc ainsi le même rang dans le préordre ω_t , le nombre d'inversions qui leur sont relatives est calculé comme si les variables avaient un rang distinct.

Proposition 9 . L'espérance de la variable aléatoire I_m vaut $E(I_m) = \frac{m(m-1)}{4}$ et sa variance vaut $V(I_m) = \frac{m(m-1)(2m+5)}{72}$. Et la loi de probabilité de I_m converge vers une loi normale quand m tend vers l'infini. L'espérance et la variance peuvent se déduire, par la relation (1), des résultats connus de la statistique τ dont l'espérance est $E(\tau)=0$ et la variance $V(\tau) = \frac{2(2m+5)}{9m(m-1)}$. De plus, dans son article séminal (Kendall, 1938), Kendall expose les grandes lignes d'une démonstration permettant de déduire que, quand *m* tend vers l'infini, la variable

$$Z = \frac{\tau}{\sqrt{\frac{2(2m+5)}{9m(m-1)}}} \qquad \textbf{(9)}$$

est asymptotiquement distribuée comme une variable de Laplace-Gauss centrée réduite, dont il donne la table, à partir de ses moments centrés d'ordre pair, sous l'hypothèse d'équiprobabilité des permutations.

$$\mu_{2k} \approx \frac{(2k)!}{2^k k!}(\mu_2)^k \qquad \textbf{(10)}$$

Compte tenu de la relation entre I_m et τ, il est clair que la distribution de I_m est également asymptotiquement distribuée comme une variable gaussienne, dont l'approximation est acceptable à partir de m=10 (Siegel, 1956). A titre d'exemple, voici une comparaison :

$Pr(I_{20} = 50) \approx 0.0003197$ et $Pr(I_{20} > 50) \approx 0.9985028$.

Avec la loi de Laplace Gauss on obtient :

Pr[49.5<LG(95 ;15.41)<50.5]≈0.0003647 et Pr(LG(95 ; 15,41> 50,5) ≈ 0.998059

5.3 Vers un nouvel indice de significativité

A un niveau *k*>0 donné, une hiérarchie orientée H_A présente plusieurs classes déjà formées et associées à des *R*-règles de degré supérieur strictement à 0, et éventuellement, quelques variables non encore associés. Nous cherchons maintenant à quantifier la significativité d'une classe, ainsi que la qualité de la hiérarchie à ce niveau.

Afin de restituer l'information maximale relative à l'ensemble des classes constituées, cette significativité doit intégrer deux paramètres majeurs :

- les cohésions des classes dont, par construction de H_A, les valeurs décroissent avec la croissance des niveaux de la hiérarchie,

- les cohérences des classes qui peuvent croître ou décroître selon les niveaux en fonction de la probabilité associée à la variable aléatoire I_m, eu égard aux inversions observées et à la taille de la classe.

Le concept que nous proposons pour associer ces deux paramètres satisfait aux quatre contraintes suivantes liées à la « sémantique » de la significativité :

1. être fonction de la cohérence et de la cohésion majorant les valeurs de la cohérence ;
2. conserver l'aspect probabiliste que possède la cohérence;
3. pondérer la cohérence, indice de « bon ordre » des attributs dans la classe selon l'implication par un facteur qui pourrait être qualifié d'affaiblissement de la cohésion et visant selon les cas à prendre en compte :
   (a) favorablement le fait que la classe formée au niveau k + 1 ait une cohésion peu différente de la classe formée à niveau k,
   (b) défavorablement le fait que la différence étant élevée, cela affecte la crédibilité de la classe formée en k + 1, même si elle a une bonne cohésion.
4. diminuer la significativité d'une classe au niveau k +1 qui, bien qu'ayant une bonne cohérence, a une cohésion qui décroît entre k et k + 1.

Définition 18 : L'indice *co* de **cohésion-cohérence** qui mesure la significativité de la classe C_{k+1} formée au niveau $k + 1$ est défini par

$$co(C_{k+1}) = \frac{c(C_{k+1})}{c(C_k)}.o(C_{k+1}) \quad (11)$$

Par convention, $co(C_0) = 1$. Un niveau k de la hiérarchie H_A est *significatif* s'il correspond à un maximum local de l'indice de cohésion-cohérence de la classe formée à ce niveau.

En effet, l'indice *co* n'étant pas une fonction monotone, il apparaît des maxima locaux correspondant d'une part à une meilleure adéquation entre les restrictions, à la classe formée à ce niveau, des préordres théorique ω_t et contingent ω_0, et d'autre part à une bonne cohésion.

Définition 19: La *qualité* de l'ensemble des niveaux h, $0 \le h \le k$, est définie par

$$q_k(H_A) = \left(\prod_{i=1}^{k} co(C_i) \right) \quad (12)$$

où C_i désigne la classe formée au niveau i. La hiérarchie orientée H_A est *significative* au niveau k si sa qualité $q_k(H_A)$ admet un minimum local.

5.4 Retour sur l'application

Conservant cette fois toutes les variables en jeu, la hiérarchie orientée obtenue avec le logiciel CHIC (Couturier et Gras 2005) comporte *16* niveaux (figure ci-dessous). Le tableau ci-dessous donne les cohésions des *R*-règles correspondantes. Par exemple, au niveau 13, la *R*-règle met l'accent sur la relation dérivée des comportements d'ouverture des élèves (I : esprit critique, E : imagination et créativité) vers des situations mathématiques les réalisant : OP8 : exemple et contre-exemple personnels, OP7 : test de réfutation). Cette interprétation globale est difficile par le seul emploi du graphe implicatif qui opère de façon binaire. Ainsi, il y a complémentarité et non redondance entre les deux approches.

Niveaux	R-règles	cohésion	Maxima locaux de l'indice co (cohésion-cohérence)
1	OP8 → OP9	0.998	0.499
2	OP5 → OP4	0.981	
3	N → A	0.955	
4	OP2 → (OP5 → OP4)	0.941	0.821
5	OP9 → OPX	0.92	
6	H → PER	0.92	0.5
7	F → OP3	0.903	
8	(N → A) → OP6	0.865	0.8
9	B → K	0.858	
10	E → (OP8 → OP7)	0.856	0.831
11	G → OP1	0.783	
12	J → (OP9 → OPX)	0.752	
13	I → (E → (OP8 → OP7))	0.707	0.752
14	C → O	0.669	
15	M → (F → OP3)	0.661	0.823
16	L → (J → (OP9 → OPX))	0.404	

TAB. 14 – *Cohésion des R-règles associées aux niveaux de la hiérarchie et maxima locaux de l'indice* co *de cohésion-cohérence*

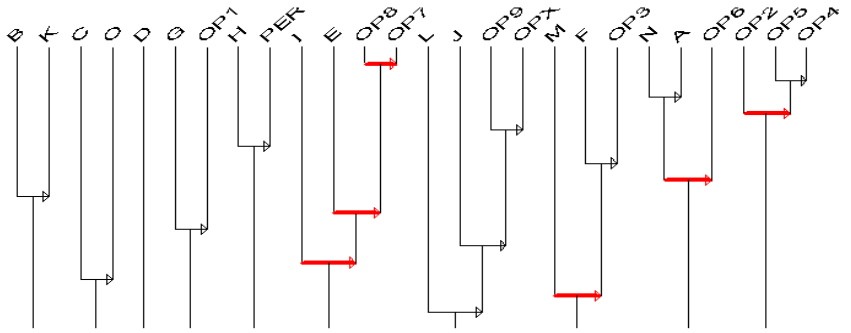


FIG. 12– *Hiérarchie orientée pour l'enquête auprès des professeurs de mathématiques*

Le calcul des cohérences à l'aide de l'algorithme implémenté dans CHIC conduit aux probabilités suivantes : $Pr(I_2 > 0) = Pr(I_2 \geq 1) = Pr(I_3 > 1) = 0.5$, $Pr(I_3 > 0) = 0.833$ et $Pr(I_4 > 2) = 0.625$. Une inversion seulement, par rapport aux occurrences, est observée pour les classes des niveaux *12*, *13* et *16*. Les maxima locaux de la cohésion-cohérence sont indiqués dans TAB 15. On observe également des maxima de l'indice de qualité *q* aux niveaux 1, *4*, *8*, *10*, *13* et *16*. Les niveaux significatifs, indiqués en gras sur la figure ci-dessus, à l'exclusion du niveau intéressant 6 (déclarer la non-pertinence des objectifs, c'est choisir de secondariser le développement de la volonté et la persévérance), avaient déjà été obtenus précédemment avec la méthode globale inspirée des travaux de Lerman (1981). Cependant, ce résultat n'a

pas valeur de généralité. Dans la situation expérimentale, la sémantique semble bien respectée dans les deux cas.

En conclusion, nous avons ici développé une approche complémentaire pour évaluer la significativité des niveaux d'une hiérarchie orientée et la qualité d'une hiérarchie orientée partielle qui tient compte du préordre défini sur les attributs de chaque *R*-règle constituée à chaque niveau de la hiérarchie. Cette approche ne nécessite pas, contrairement à une approche globale précédemment employée, la détermination d'une préordonnance sur l'ensemble des couples selon le critère de cohésion. De plus, lorsque le nombre *m* d'attributs « classés » devient grand, les calculs du nouveau critère peuvent être simplifiés par le recours à l'approximation à une loi de Laplace Gauss. Cette approche pourrait être généralisée à la recherche d'une mesure de distorsion entre deux permutations, prolongeant ainsi des travaux de Kendall. Mais, de nouvelles mises à l'épreuve sur des données réelles et, en particulier, des corpus de grande taille tels qu'on les trouve en fouille de données, permettront une comparaison plus robuste sur le plan de l'information restituée au cours des analyses que pourraient en faire des experts des domaines étudiés.

Chapitre 5 : Dualité entre variables actives et variables supplémentaires : typicalité et contribution[12]

1 Introduction

Après avoir étudié la fonction structurante (graphe, hiérarchie) du comportement des individus de E sur l'ensemble V des variables, nous nous interrogeons maintenant au sujet du rôle d'identificateurs de ces individus sur les éléments des différentes structures obtenues dans l'ensemble des variables. Afin d'y parvenir, dans ce chapitre, nous chercherons à établir une correspondance entre des éléments de certaines partitions de E et une structure de V du corpus de données, que ce soit un graphe ou une hiérarchie. Comme nous le verrons, cette correspondance devrait permettre d'opérer dans les deux sens individus↔variables au moyen de critères quantitatifs.

Cette analyse implicative de la correspondance entre deux ensembles E et V, une fois munis de métriques liées entre elles, est prometteuse et mériterait des développements que nous ne ferons qu'aborder ici. Nous mettrons cependant en évidence une dualité entre deux structures comme il est fait en Analyse Factorielle des Correspondances (A.F.C.), mais ici de façon plus modeste. Nous construirons pour cela, et dans un premier temps, une mesure visant à quantifier la « responsabilité » d'individus ou de groupes d'individus à l'élaboration des structures diversement et graduellement construites[13] sur l'ensemble des variables. Ces groupes homogènes rassemblent en une partition des individus sur la base d'un lien qui les identifie et les discrimine des autres comme, par exemple, le feraient une classe de collège ou une tranche d'âge. La « responsabilité » s'appliquera sur une structure, peut-être causale, obtenue dans l'ensemble des variables par l'A.S.I.. Inversement, sur la base de cette première correspondance, il sera possible d'identifier quelle(s) variable(s) peut ou peuvent caractériser tel individu ou tel groupe d'individus par rapport à cette structure dissymétrique donnée par l'A.S.I.. On cherchera enfin à détecter le couple individu(s)-variable(s) le plus spécifique du croisement des structures retenues sur E et sur V.

Pour ce faire, nous introduisons, en élargissant la problématique, la notion de **variable supplémentaire** en A.S.I. à l'instar de la même notion définie en A.F.C. (Benzecri, 1973). Il s'agit d'une variable exogène, un descripteur par exemple, n'intervenant pas directement dans l'établissement et la représentation (graphe ou hiérarchie) des liaisons exprimées par la classification et l'arborescence entre les variables dites principales de V. Il lui correspond, le plus souvent, une partition de l'ensemble des individus. Par exemple, une variable supplémentaire pourra représenter une catégorie de individus (âge, sexe, attitude, catégorie socio-professionnelle, etc.), mais aussi bien, si nécessaire, une des variables jugée principale, dans un premier temps..

---

[12] Ce chapitre a été présenté sous une forme voisine en atelier lors du Congrès EGC 6 sous le titre : « Typicalité et contribution des sujets et des variables supplémentaires en Analyse Statistique Implicative » avec pour auteurs : Régis Gras , Jérôme David, Jean-Claude Régnier, Fabrice Guillet. La publication de l'atelier figure dans *Volume 2, Cépaduès Editions(2006)*

[13] …graduellement construites car liées au seuil retenu par exemple dans l'élaboration des graphes implicatifs

Au cours de l'analyse, à un niveau quelconque de la hiérarchie se forme une classe C de cohésion non nulle ou un chemin C au seuil choisi. Notre objectif, particulièrement dans le cas d'un nœud significatif de la hiérarchie, est de définir un critère permettant d'identifier un ou des individus, puis la catégorie ou le groupe de individus, ou tout autre variable supplémentaire, qui seraient associés à l'apparition de cette classe ou de ce chemin, à savoir :

- ou bien des individus ou des variables supplémentaires **typiques** du comportement global de la population ; en d'autres termes, le comportement de ces individus ou de ces variables sera ainsi en harmonie avec le comportement statistique de la population à l'origine de la classe C,
- ou bien des individus ou des variables supplémentaires les plus **contributives**[14] c'est-à-dire *contribuant* formellement le plus à l'agrégation conduisant à C, c'est-à-dire en se référant à une population respectant formellement les règles constitutives de C, en d'autres termes, plus ou moins responsables de l'agrégation conduisant à C.

D'autres auteurs ont cherché à quantifier la relation qu'un individu et/ou un groupe d'individus entretiennent avec une ou des variables, par exemple, (Lerman,1981a) pour l'analyse des similarités, mais au moyen d'une modélisation et de concepts différents.

Notre approche se ramène à trouver des réponses originales aux questions suivantes :

1. quoi retenir de la forêt des règles qui sous-tendent un certain chemin d'un graphe ou une certaine classe d'une hiérarchie afin de conserver la « charpente » constituée des règles les plus consistantes ?
2. comment quantifier les positions respectives des éléments de E et de V eu égard à cette dominance ?
3. quel critère statistique permettrait de retenir avec un risque d'erreur minimal la variable supplémentaire la plus « responsable » de la « charpente » ?

Nous nous appuierons pour illustrer notre propos, sur l'exemple développé dans le chapitre 4 de la présente Partie 1. Nous rapportons les deux représentations graphiques :

[14] Les concepts et leurs propriétés définis ici diffèrent de ceux donnés dans (Gras et al., 1996 c) où l'on n'y distingue pas les deux notions « typicalité » et « contribution ».

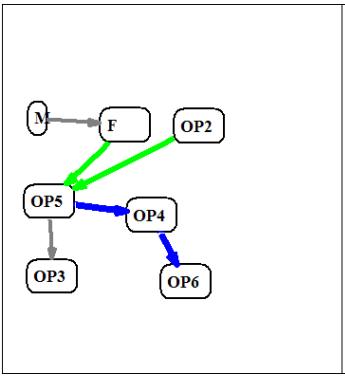


FIG. 13- *Graphe implicatif à 7 variables*

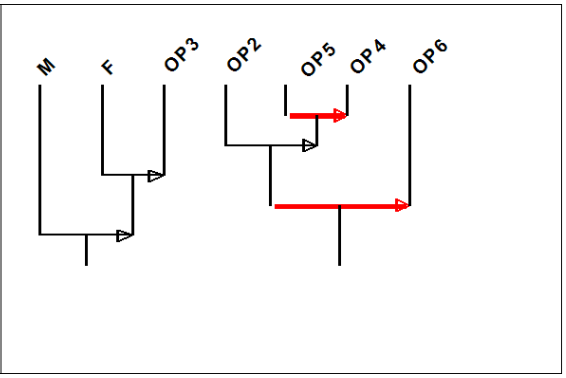


FIG. 14 *Hiérarchie cohésitive à 7 variables. Ces représentations portent sur l'exemple du questionnaire « enseignants » du chapitre 4*

2 Puissance implicative de classe et de chemin

2.1 Couples génériques

L'idée directrice suivie consiste à porter notre attention sur les « lignes de force », (ou, selon une autre métaphore : les « lignes de crête » ou la « charpente ») des règles d'association, plutôt que de les retenir toutes avec le risque afférent d'être submergé par leur nombre et de brouiller l'essentiel par les bruits confus qui les accompagnent. Plaçons-nous à un niveau k de la hiérarchie où viennent de se réunir, pour former C, deux classes $\underline{A}$ et $\underline{B}$ telles que $\underline{A} \Rightarrow \underline{B}$. Ainsi dans la figure ci-dessus rapportant la hiérarchie cohésitive, nous pouvons dire que :

- au niveau 2, on lirait $\underline{A} = OP2$ et $\underline{B} = (OP5, OP4)$;
- au niveau 4, on lirait : $\underline{A} = (\mathrm{OP2}, (\mathrm{OP5}, \mathrm{OP4}))$ et $\underline{B} = OP6$

Définition 20: Étant donné les intensités d'implication $\varphi(i, j)$, le couple (a,b) tel que:

$\forall i \in \underline{A}, \forall j \in \underline{B}, \varphi(a,b) \geq \varphi(i,j)$ est appelé **couple générique** de C[15], qui présente un caractère dominant sur le plan implicatif. En cas d'ex-æquos, on choisit le couple selon le critère des occurrences maximales.

Définition 21: Le nombre $\varphi(a,b)$ est appelé **intensité générique de la classe** C. C'est sa force qui domine les autres intensités des couples non retenus.

[15] C'est ce couple, généralement unique, qui intervient par le sup. dans le calcul de l'implication de $\underline{A} \Rightarrow \underline{B}$ (Gras et al, 1996 c).

Mais, dans chaque sous-classe de C, il existe également un couple générique. Précisément, si C est constituée de g (g≤k) sous-classes (C comprise), il y a g couples génériques à l'origine de C et g intensités maximales d'implication notées $\varphi_1, \varphi_2, \ldots, \varphi_g$, qui leur correspondent.

Dans le cas d'un chemin ch(i) du graphe implicatif, chemin fermé transitivement (chaque arc de la fermeture admet une intensité d'implication au moins égale à 0,50), composé de g nœuds, ch(i) présente $\frac{g(g-1)}{2}$ arcs transitifs. A chacun de ces arcs, par ex. (a,b), on associe, comme pour une classe, l'intensité d'implication de la règle correspondante.

Définition 22: Parmi les $\frac{g(g-1)}{2}$ intensités, l'une d'entre elles, au moins, est maximale et est appelée encore **intensité générique du chemin**.

Définition 23: Le vecteur $(\varphi_1, \varphi_2, ..., \varphi_g)$ de $[0,1]^g$, est appelé **vecteur puissance implicative** de la classe ou du chemin. Il traduit une force implicative interne à la classe ou au chemin. Ce vecteur a la propriété, en ne retenant, métaphoriquement, que les lignes de force (ou de crête) de représenter une sorte de « flux » implicatif au sein de la classe ou du chemin. Ce vecteur occupe une place stratégique par rapport à la classe ou au chemin. Toujours métaphoriquement, c'est un indicateur de la visibilité maximale d'un certain « courant » qui transporterait le flux implicatif au sein de la classe ou du chemin.

2.2 Puissance implicative d'un individu sur une classe ou sur un chemin du graphe implicatif et distance à cette classe ou à ce chemin

Dans le cas où les variables V sont binaires, un individu x quelconque respecte ou non l'implication du couple générique d'une classe ou d'un arc de chemin avec un ordre de qualité comparable. Associant logique formelle et considération sémantique, nous noterons $\varphi_x(a,b)$ cette qualité de respect en x de l'implication a=>b, par exemple et en fonction des valeurs prises en a et b par l'individu x :

$$\varphi_x(a,b) = 1 \text{ si a=1 ou a=0 et b=1}$$

$$\varphi_x(a,b) = 0 \text{ si a=1 et b=0}$$

$$\varphi_x(a,b) = p \text{ si a= b=0 avec } p \in]0;1]$$

Dans nos premières expériences, nous choisissions p=0,50, valeur neutre[16]. Ainsi, à l'individu x, nous pouvons associer g nombres notées $\varphi_{x,1}, \varphi_{x,2}, \ldots, \varphi_{x,g}$ correspondant aux g valeurs respectivement prises par x selon les g règles génériques de la classe ou du chemin.

[16] Dans le logiciel CHIC, pour des variables **modales** ou **numériques**, le calcul des typicalités (et des contributions) se fait cependant en modulant ces valeurs, à l'aide d'une fonction ad hoc, munie de propriétés adaptées afin de mieux prendre en compte la sémantique des valeurs attribuées par x à a et à b. Par exemple, pour a=0 et b=1, la fonction prend, dans CHIC, la valeur 0.682.

Définition 24 : Le vecteur $(\varphi_{x,1},\varphi_{x,2},...,\varphi_{x,g})$ est appelé **vecteur contingent générique** de l'individu x ou puissance implicative de x sur la classe ou le chemin.

Définition 25 : L'individu fictif, théorique x_t qui admettrait $(\varphi_1,\varphi_2,...,\varphi_g)$ comme vecteur contingent générique est appelé **individu typique optimal**.

En effet, on peut interpréter ce vecteur comme étant celui d'un individu « typique » des règles génériques puisque les valeurs prises par cet individu selon ces règles sont exactement celles de l'ensemble de la population. Cet individu, image conforme d'un individu fictif typique de E, n'existe pas réellement en général mais, s'il existe, il peut ne pas être unique. Dans ces conditions, on peut munir l'espace du produit $[0,1]^g$ d'une métrique afin d'obtenir un contraste accentuant les effets de fortes intensités génériques ou, réciproquement, minorant les effets d'une faible intensité générique.

Définition 26 : On appelle **distance de typicalité** d'un individu quelconque x à la classe ou au chemin C, le nombre $d(x,C)=\left[\frac{1}{g}\sum_{i=1}^{i=g}\frac{[\varphi_i-\varphi_{x,i}]^2}{1-\varphi_i}\right]^{\frac{1}{2}}$

Ce nombre, qui vérifie formellement les trois axiomes d'une distance, n'est autre également que la distance du type χ^2 entre les deux distributions $\{1-\varphi_i\}$ et $\{1-\varphi_{x,i}\}$ pour i=1 à g, qui expriment les écarts entre les implications génériques contingentes et l'implication stricte. Elle exprime, aussi et en particulier, l'écart observé sur les règles génériques entre l'individu x considéré et l'individu théorique typique optimal, écart nuancé par ces intensités. Elle exprime aussi la place que prend l'individu x par rapport à la classe ou au chemin C. C'est pour cette raison que nous avons choisi le mot **typicalité** pour quantifier le comportement de l'individu x selon les règles génériques. Nous allons le préciser plus loin.

Remarque 1 : Lorsque φ_i=1, une légère correction sur cette valeur permet d'éviter la division par zéro (par exemple, prendre φ_i = 0,99999999) ce qui ne change pas fondamentalement la distance.

Remarque 2 : Une classe C étant donnée, on peut définir une structure d'espace métrique sur E par la donnée de la distance indicée par C entre deux individus quelconques de E, distance qui mesure la différence de comportement des individus x et y à l'égard de C :

$$d_C(x,y)=\left[\frac{1}{g}\sum_{i=1}^{i=g}\frac{[\varphi_{x,i}-\varphi_{y,i}]^2}{1-\varphi_i}\right]^{\frac{1}{2}}$$

Cette distance établit une première correspondance entre l'ensemble des individus et l'ensemble des variables structuré par un graphe implicatif ou une hiérarchie cohésitive. On voit alors que la distance de typicalité donnée plus haut n'est que la spécification de d_C aux individus respectivement x et x_t. La distance d_C permet de conférer à E une d_C-structure topologique discrète. Cette topologie est équivalente à celle qui serait définie sur l'ensemble des vecteurs contingents $\vec{X}=(\varphi_{x,1},\varphi_{x,2},...,\varphi_{x,g})$, sous-ensemble d'un espace vectoriel normé de dimension g et de norme : $\left\|\vec{X}-\vec{Y}\right\|=d_C(x,y)$. L'opérateur symétrique associé à la forme

quadratique qui conduit à cette distance, a pour matrice, la matrice diagonale d'éléments $(g(1-\varphi_i))^{-1}$ pour i=1,…, g. Toutefois, l'interprétation de la somme de deux tels vecteurs n'a de sens que dans cadre théorique mathématique, c'est-à-dire hors du contexte dans lequel nous travaillons en A.S.I..

Une application intéressante peut consister à déterminer le ou les individus appartenant à une boule de diamètre donné et de centre l'un des individus pré-désignés, comme par exemple, l'individu optimal. En prolongement de cette approche métrique, le problème de complétion des données manquantes pourrait y puiser une solution originale. Nous aborderons (Partie 2, Chapitre 4) cette question, toutefois sous un angle un peu différent. Nous restons par ailleurs convaincus que les perspectives de recherche sur ce sujet sont nombreuses et stimulantes.

2.3 Typicalité, spécificité et contribution d'un individu ou d'une variable supplémentaire à une classe d'une hiérarchie cohésitive ou à un chemin d'un graphe implicatif

2.3.1 La notion de typicalité

Nous définissons la mesure de **typicalité** à partir du rapport entre la distance de typicalité relative à l'individu considéré et la distance à C, classe ou chemin, la plus grande dans l'ensemble des individus. Cette distance maximale est celle des individus y dont les $\varphi_{y,i}$ sont tous nuls ou très faibles. Ces individus sont donc ceux les plus opposés aux règles génériques. La typicalité d'un individu est alors d'autant plus grande qu'il s'écarte de ces mêmes individus, donc qu'il manifeste un comportement comparable à celui de l'individu théorique optimal. La typicalité d'une catégorie d'individus ou d'une variable supplémentaire G[17] s'en déduit alors :

Définition 27 : La typicalité de l'individu x à la classe ou au chemin C est mesurée par :

$$\gamma(x,C) = 1 - \frac{d(x,C)}{\max_{y \in E}\{d(y,C)\}}$$

Définition 28 : La typicalité de la variable supplémentaire G à la classe ou au chemin C est mesurée par :

$$\gamma(G,C) = \frac{1}{cardG}\sum_{x \in G}\gamma(x,G)$$

Afin de donner au chercheur le moyen de savoir ou de vérifier rapidement si telle catégorie d'individus qui l'intéresse est statistiquement déterminante dans la constitution d'une classe ou d'un chemin transitif, un algorithme a été élaboré en s'appuyant sur les deux notions que l'on définit ci-dessous : groupe optimal et catégorie déterminante.

Définition 29 : du **groupe optimal** d'une classe ou d'un chemin.

Soit E l'ensemble des individus étudiés. Un groupe optimal, noté GO(C), d'une classe de la hiérarchie cohésitive ou d'un chemin du graphe implicatif, noté C, est le sous-ensemble de

[17] Les deux mots « catégorie » et « variable supplémentaire » seront utilisés indifféremment, le premier ayant une charge sémantique plus forte que le second.

E qui accorde à C une typicalité plus grande que le complémentaire de GO(C) et qui forme avec celui-ci une partition en deux groupes maximisant la variance inter-classe de la série statistique des typicalités individuelles des individus les constituant (Ratsimba-Rajohn, 1992). Une telle partition est dite *significative*.

L'existence de ce groupe optimal est également démontrée dans (Gras R. et al., 1996 b, 1996 c). Les propriétés utilisées sont aussi celles qui le sont pour établir l'algorithme sur lequel se basent les modules des programmes informatiques qui construisent, automatiquement dans C.H.I.C., chaque sous-groupe optimal.

Considérons une partition $\{G_i\}_i$ de E. Cette partition peut être définie par une variable supplémentaire correspondant à une variable catégorielle, à un descripteur de E, admettant deux modalités binaires ou plus, par exemple, la variable « catégories socio-professionnelles ». Soit X_i une partie aléatoire de E de même cardinal que G_i et Z_i la variable aléatoire $Z_i = card(X_i \cap GO(C))$. Selon un modèle équiprobable admissible, Z_i suit une loi binomiale de paramètres : $card(G_i)$ et $\frac{card(GO(C))}{cardE}$ qui est la fréquence du groupe optimal de la classe ou du chemin C dans l'ensemble E.

Définition 30 : On appelle **variable supplémentaire**, ou catégorie, **la plus typique** de la classe ou du chemin C, celle qui minimise l'ensemble $\{p_i\}_i$ des probabilités p_i telles que:

$$\forall i, p_i = \Pr ob\{card(G_i \cap GO(C)) < Z_i\}$$

Ainsi, établir que G_j est la catégorie la plus typique revient à déceler, parmi les catégories, celle dont le nombre d'individus appartenant en même temps à celle-ci et au groupe optimal, est le *plus étonnamment grand eu égard à son cardinal*. Nous retrouvons ici la philosophie sous-jacente à la construction de l'indice d'implication.

Définition 31 : Une catégorie G_0 est dite déterminante au niveau de risque ou au seuil α si la probabilité associée p_0 est inférieure à α. Autrement dit, le risque de se tromper en affirmant cette propriété est donc au plus égale à α.

Par suite, la signification d'une classe ou d'un chemin ayant été donnée par l'expert, il lui associera le sous-ensemble le plus porteur de ce sens, celui correspondant au niveau de risque qu'il juge acceptable.

Groupe optimal et indice de similarité. Par ailleurs, nous pouvons associer au groupe optimal, une variable binaire définie par la fonction indicatrice de ce sous-ensemble de E. De la même façon, nous pouvons également associer à chaque catégorie G_i ou bien à la variable supplémentaire correspondante, une variable binaire dont l'indice de similarité $s = \frac{n_{a\wedge b} - \frac{n_a n_b}{n}}{\sqrt{\frac{n_a n_b}{n}}}$, au sens de I.C. Lerman, vérifie la condition $p_i = \Pr ob\{S \geq s\}$ où S est la variable aléatoire dont s est une réalisation. Ainsi, minimiser l'ensemble des probabilités $\{p_i\}_i$ revient à maximiser l'indice de similarité entre les variables binaires, indicatrices de

sous-ensembles, associées l'une au groupe optimal GO(C) et les autres aux différentes catégories $\{G_i\}_i$.

Cette remarque permet d'étendre efficacement la notion de variable supplémentaire la plus typique à des variables numériques, prenant leurs valeurs sur l'intervalle [0 ;1]. Il suffit alors d'extraire la plus forte des valeurs de similarité entre la variable binaire indicatrice définie par le groupe optimal et les différentes variables numériques placées en variable supplémentaire, l'indice de similarité étant calculé selon le principe retenu en analyse statistique implicative pour les variables numériques. Nous savons que sa restriction au cas binaire coïncide avec sa valeur s dans le cas où les 2 variables sont binaires.

Ainsi en résumé, il est possible de dégager à la fois les individus et les groupes d'individus typiques d'une règle ou d'un ensemble (classe ou chemin) de règles généralisées. Ce sont donc ceux qui sont le plus en accord avec la qualité de ces liaisons au sein de l'ensemble E considéré. Si, par exemple, la liaison implicative entre les variables a et b est quantifiée par $\varphi(a,b) = 0{,}92$, les individus x qui lui attribuent la valeur $\varphi_x(a,b) = 0{,}90$ sont plus typiques que ceux qui lui attribuent la valeur 0,98. Ceux-ci sont à une distance plus grande que les premiers pour le comportement statistique de l'ensemble E. La nuance entre cette notion et celle de contribution définie plus loin prend tout son sens dans l'étude des variables modales ou numériques.

2.3.2 La notion de spécificité

Si à chaque classe ou chemin C_j, on peut associer au moins un groupe typique, il est pertinent de mettre en évidence le couple (variable supplémentaire G_i , classe ou chemin C_j) remarquable quant à l'optimalité de sa conjugaison. D'où le recours à la notion de spécificité que nous introduisons par la définition suivante :

Définition 32 : La variable supplémentaire G_i étant donnée, le couple (G_i, C_j) est dit **mutuellement spécifique** lorsque G_i est la variable la plus spécifique de la règle associée à C_j et que la probabilité (le risque)

$$p_i^k = \Pr ob\left\{card(G_i \cap GO(C_k)) < Z_{i,k}\right\}$$

de G_i par rapport aux C_k, autres classes de la hiérarchie cohésitive ou autres chemins du graphe implicatif est supérieure à un seuil β (arbitrairement fixé par l'utilisateur).

Une analyse étant donnée, il peut n'exister aucun couple mais il peut aussi apparaître un ou plusieurs couples mutuellement spécifiques. Ce ou ces couples offrent l'intérêt de faire porter l'attention de l'expert sur les plus fortes associations prenant origine dans une variable supplémentaire.

Définition 33 : De façon analogue, un individu x étant donné, le couple (x, C_j) est mutuellement spécifique lorsque cet individu appartient au groupe optimal relatif à la règle associée à C_j et que la mesure de typicalité à C_j est maximale par rapport à toutes les autres valeurs de typicalité aux classes de la hiérarchie cohésitive ou aux chemins du graphe implicatif.

2.3.3 La notion de contribution

A ce jour, nous distinguons cette notion de celle de typicalité, ce que nous ne faisions pas en 1996. Cette distinction se manifeste dans la manière dont nous examinons la responsabilité des individus, puis des variables supplémentaires qui peuvent en être des descripteurs, *dans l'existence* d'une règle ou d'une règle généralisée entre variables principales.

Supposons que deux variables *a* et *b* (*resp.* plusieurs variables sur un chemin du graphe implicatif ou bien deux classes de la hiérarchie) soient réunies par un arc sur un graphe à un certain seuil (*resp.* en un chemin transitif C du graphe ou bien en une classe C dans une hiérarchie à un certain niveau). Connaissant la valeur $\varphi_{x,i}$ attribuée par l'individu x à la règle i : $a \Rightarrow b$ (resp. règle i du chemin C ou bien de la classe C constituée de g règles génériques) supposée admissible, nous posons alors la définition suivante.

Définition 34 : On appelle **distance de contribution** de x à l'arc (a,b) ou à C :

$$d(x,C) = \left[\frac{1}{g} \sum_{i=1}^{i=g} \left[1 - \varphi_{x,i}\right]^2 \right]^{\frac{1}{2}} \text{ où g=1 dans le cas de l'arc (a,b)}$$

Cette distance, de type euclidien, mesure l'écart entre le vecteur contingent générique de x : $(\varphi_{x,1}, \varphi_{x,2}, ..., \varphi_{x,g})$ et le vecteur à g composantes (1,1,1,…,1). Ce dernier est le vecteur d'un *individu théorique optimal qui satisferait strictement toutes les règles génériques.*

C'est donc en ce sens que les notions de typicalité et de contribution sont distinctes.

Toutefois à l'instar de ce que nous avons fait avec la notion de typicalité, nous pouvons définir sur E une topologie discrète d'espace normé dont la norme est associée à la distance entre deux individus quelconques suivante :

$$d_C(x,y) = \left[\frac{1}{g} \sum_{i=1}^{g} (\varphi_{x,i} - \varphi_{y,i})^2 \right]^{\frac{1}{2}}$$

Définition 35 : On appelle **contribution de x à C** le nombre : $\gamma(x,C) = 1 - d(X,C)$

Cette définition est la restriction de celle de la typicalité au cas où, cette fois, on compare l'individu x aux « pires » individus par rapport aux règles génériques : leur comportement s'oppose à l'implication de chaque règle (1 pour la prémisse et 0 pour la conclusion). Cette contribution a pour maximum 1 dans le cas où l'individu x a donné la valeur 1 à toutes les règles i. Ceci permet de concilier la sémantique avec la définition formelle. En effet, plus la différence est importante, plus l'individu observé a un comportement voisin de celui de l'individu théorique optimal et plus il s'éloigne de ceux qui réfutent les règles génériques. Nous pouvons donc dire qu'en contribuant à l'émergence de la classe, ils en sont aussi responsables.

La suite des définitions et des algorithmes de calcul (contribution d'une catégorie ou d'une variable supplémentaire G, groupe optimal d'individus, catégorie ou variable supplémentaire la plus contributive, couple mutuellement spécifique) se transpose immédiatement à partir des principes explicités pour la typicalité et la spécificité. Mais, dans les situations réelles, nous observons la nuance entre les deux concepts ce qui enrichit

l'information exploitable par l'utilisateur. Notons cependant que le concept de contribution est plus volontiers retenu pour l'interprétation dans une perspective inductive.

3 Application à l'étude du fichier Raf du chapitre 2

Reprenons l'exemple présenté dans le chapitre 2 dont le fichier RAF (30 sujets, 5 variables binaires, 2 variables supplémentaires : Fs et Gs) est rapporté dans le tableau (TAB 11). A titre didactique, nous allons conduire les calculs à la main pour que chaque définition et algorithme du présent chapitre soient clairement développés. Nous calculerons donc la mesure de typicalité des sujets à l'égard de la règle $V_3 => V_1$, c'est-à-dire tout autant à l'égard du chemin C ou arc $V_3 \rightarrow V_1$ qu'à celui de la classe $C=(V_3, V_1)$ puisque dans ce cas simplifié les deux représentations coïncident.

	V1	V3	F s	G s		V1	V3	F s	G s		V1	V3	F s	G s
i1	1	1	1	0	i11	1	1	1	0	i21	1	1	0	1
i2	1	1	1	0	i12	1	1	1	0	i22	1	1	0	1
i3	1	1	1	0	i13	1	1	1	0	i23	1	0	0	1
i4	1	1	1	0	i14	1	1	0	1	i24	1	0	0	1
i5	1	1	1	0	i15	1	1	1	0	i25	0	0	1	0
i6	1	1	1	0	i16	1	1	0	1	i26	0	0	0	1
i7	1	0	1	0	i17	1	1	0	1	i27	0	0	0	1
i8	1	0	1	0	i18	1	1	0	1	i28	0	0	0	1
i9	1	1	1	0	i19	1	1	0	1	i29	0	1	0	1
i10	1	1	1	0	i20	1	1	1	0	i30	0	0	0	1

TAB. 15- *extrait du tableau de données du fichier Raf*

Nous avons trouvé l'intensité d'implication (modèle de Poisson) de la règle $\varphi(V_3,V_1)$ =0,92.

Notons $v_x(V_k)$ la valeur que l'individu x prend en V_k, c'est-à-dire 1 ou 0 suivant que V_k est observée ou non en x.

Cas	$v_X(V_3)$=	$v_X(V_1)$=	$\varphi_x(V_3,V_1)$ =	Nombre d'individus
(a)	1	1	1	20
(b)	0	1	0,68 (voir note 16)	1
(c)	0	0	0,50	5
(d)	1	0	s0	4

TAB. 16

Ce qui est justifié par les données fournies dans le tableau ci-après :

	V3	V1	Cas		V3	V1	Cas		V3	V1	Cas
i1	1	1	(a)	i11	1	1	(a)	i21	1	1	(a)
i2	1	1	(a)	i12	1	1	(a)	i22	1	1	(a)
i3	1	1	(a)	i13	1	1	(a)	i23	0	1	(b)
i4	1	1	(a)	i14	1	1	(a)	i24	0	1	(b)
i5	1	1	(a)	i15	1	1	(a)	i25	0	0	(c)
i6	1	1	(a)	i16	1	1	(a)	i26	0	0	(c)
i7	0	1	(b)	i17	1	1	(a)	i27	0	0	(c)
i8	0	1	(b)	i18	1	1	(a)	i28	0	0	(c)
i9	1	1	(a)	i19	1	1	(a)	i29	1	0	(d)
i10	1	1	(a)	i20	1	1	(a)	i30	0	0	(c)

TAB. 17-

L'individu fictif x_t tel que $\varphi_{x_t}(V_3,V_1) = 0,92$ est un individu typique optimal

On rappelle que la distance implicative d'un sujet x à la classe C ayant g sous-classes, le nombre:

$$d(x,C) = \left[\frac{1}{g} \sum_{i=1}^{i=g} \frac{[\varphi_i - \varphi_{x,i}]^2}{1-\varphi_i} \right]^{\frac{1}{2}}$$

Dans ce cas $\varphi_i = \varphi(V_3,V_1)$ et $\varphi_{x,i} = \varphi_x(V_3,V_1)$ avec g=1 car il n' y a pas de sous-classe.

Nous obtenons donc les valeurs suivantes :

$v_x(V3)=$	$v_x(V1)=$	$\varphi_x(V_3,V_1) =$	d(x,C)
1	1	1	$\left[\frac{(1-0,92)^2}{1-0,92} \right]^{\frac{1}{2}} = 0,28$
0	1	0,68	$\left[\frac{(0,68-0,92)^2}{1-0,92} \right]^{\frac{1}{2}} = 0,85$
0	0	0,50	$\left[\frac{(0,50-0,92)^2}{1-0,92} \right]^{\frac{1}{2}} = 1,48$
1	0	0	$\left[\frac{(0,92)^2}{1-0,92} \right]^{\frac{1}{2}} = 3,25$

TAB. 18

Ainsi, $\max_{y \in E} (d(y,C)) = 3,25$

La mesure de la typicalité d'un individu x relativement à la classe (V_3,V_1) est donnée par la formule :

$$\gamma(x,C) = 1 - \frac{d(x,C)}{\max_{y \in E}\{d(y,C)\}}$$

(v_x(V1) ; v_x(V3))	(1 ; 1)	(0 ; 1)	(0 ;0)	(1 ; 0)
γ(x, C)=	$1-\frac{0,28}{3,25} \approx 0,913$	$1-\frac{0,85}{3,25} \approx 0,738$	$1-\frac{1,48}{3,25} \approx 0,544$	$1-\frac{3,25}{3,25} = 0$

La valeur moyenne de la mesure de typicalité d'un individu x relativement à la classe (V3,V1) est alors $\bar{\gamma} = \frac{1}{30}\left[(20.0,913)+(4.0,738)+(5.0,544)+1.0\right] \approx \frac{2,953}{30} \approx 0,798$

Catégorie la plus typique.

GO((V3,V1))=
{i1; i2; i3; i4; i5; i6; i9; i10; i11; i12; i13; i14; i15; i16; i17; i18; i19; i20; i21; i22}

	V1	V3	F s	G s		V1	V3	F s	G s
i1	1	1	1	0	i13	1	1	1	0
i2	1	1	1	0	i15	1	1	1	0
i3	1	1	1	0	i20	1	1	1	0
i4	1	1	1	0	i14	1	1	0	1
i5	1	1	1	0	i16	1	1	0	1
i6	1	1	1	0	i17	1	1	0	1
i9	1	1	1	0	i18	1	1	0	1
i10	1	1	1	0	i19	1	1	0	1
i11	1	1	1	0	i21	1	1	0	1
i12	1	1	1	0	i22	1	1	0	1

TAB. 19

Dans ce groupe, on trouve 13 sujets F_s ou filles parmi les 16 filles du fichier, et 7 sujets G_s ou garçons parmi les 14 garçons. La variable Z_1 qui représente le cardinal de l'intersection d'une partie aléatoire X_1 (de cardinal g_1=card(F)=16) avec le groupe optimal E_1 = GO((V3,V1)) de cardinal 20 suit la loi binomiale $B(16;\frac{20}{30})$. Le caractère typique de F relativement à E_1 est mesuré par la probabilité de dépasser dans une expérience aléatoire le nombre d'observations de filles dans le groupe optimal. Plus ce nombre est faible, plus il est "surprenant" de constater un tel effectif dans E_1.

$$Prob\{Z_1 > 13\} = \sum_{k=14}^{16} C_{16}^{k} (\frac{2}{3})^{k} (\frac{1}{3})^{16-k} \approx 0,0593$$

La variable Z_2 qui représente le cardinal de l'intersection d'une partie aléatoire X_2 (de cardinal g_2= card(G) = 14) avec le groupe optimal E_1 de cardinal 20 suit la loi binomiale $B(14;\frac{20}{30})$.

$$\Pr ob\{Z_2 > 7\} = \sum_{k=8}^{14} C_{14}^{k} (\frac{2}{3})^k (\frac{1}{3})^{14-k} \approx 0{,}8505$$

Par suite, la catégorie F est la plus typique relativement à la relation implicative $V_3 => V_1$, avec un niveau de risque de 0,0593 (niveau de risque de se tromper en affirmant qu'il ne s'agit pas de cette catégorie F = probabilité de trouver plus de 13 filles parmi les 16 dans un groupe de 20 sujets si la répartition fille-garçon se faisait au hasard dans E_1).

Remarquons que si nous prenions l'inégalité au sens large, nous obtiendrions les valeurs suivantes :

$$\Pr ob\{Z_1 \geq 13\} = \sum_{k=13}^{16} C_{16}^{k} (\frac{2}{3})^k (\frac{1}{3})^{16-k} \approx 0{,}166$$

$$\Pr ob\{Z_2 \geq 7\} = \sum_{k=7}^{14} C_{14}^{k} (\frac{2}{3})^k (\frac{1}{3})^{14-k} \approx 0{,}942$$

4 Application au questionnaire « Professeurs de Terminale »

Rappelons brièvement la situation développée dans le chapitre 4 de cette partie 1. Dans le cadre d'une enquête de l'Association des Professeurs de Mathématiques de l'Enseignement Public (APMEP) auprès de professeurs de mathématiques de classes terminales (séries scientifiques S et ES, littéraires LI et technologiques TE sont les variables supplémentaires), nous avons recueilli et analysé (Bodin et Gras., 1999) les réponses de 311 professeurs, à des classements (de 1 à 6) portant sur quinze objectifs qu'ils assignent à leur enseignement (A, B, C, …, O)[18] et sur leurs opinions relatives à dix phrases susceptibles d'être communément énoncées (OP1, OP2, ..., OPX)[19]. La variable PER donne la possibilité de désigner les objectifs jugés non pertinents. Les 26 variables correspondantes ne sont pas binaires, sauf PER, mais ordinales (valeurs {1, 0.8, 0.6, 0.4, 0.2, 0.1, 0} pour les objectifs et {1, 0.5, 0} pour les opinions). Ainsi l'analyse intègre l'intensité des attitudes, d'un choix prioritaire d'un objectif à un choix plus secondaire, voire non retenu.

Les variables supplémentaires sont : S(cientifique), ES(économique et sociale), LI(ttéraire), TE(chnologique).

Les occurrences des 30 variables sont les suivantes :

[18] Par exemple, E symbolise l'objectif : « développement de l'imagination et de la créativité »

[19] Par exemple, OP4 symbolise : « Pour corriger, j'aime bien un barème très détaillé sur les résultats à obtenir »

A	B	C	D	E	F	G	H	S s	ES s
105.7	8.8	9.7	140.0	21.8	138.7	19.5	44.8	155	68
I	J	K	L	M	N	O	PER	LI s	TE s
83.1	108.4	77.6	4.6	90.2	66.6	33.2	254	22	66
OP1	OP2	OP3	OP4	OP5	OP6	OP7	OP8	OP9	OPX
81.5	147.5	242.5	229.0	190.0	240.0	200.0	165.0	98.0	207.0

TAB. 20 – *Occurrences des variables de l'enquête sur les professeurs de mathématiques*

La hiérarchie cohésitive obtenue par CHIC à partir d'un nombre réduit des variables, afin de conserver les niveaux les plus significatifs, est donnée par la figure ci-dessous.

Considérons la classe C = [E ⇒ (OP8 ⇒ OP7)] ⇒ OPX. Son sens, analysé plus en détail dans (Bodin et Gras., 1999), est fortement marqué par l'importance accordée à l'imagination et à la recherche personnelle, par les enseignants d'accord avec ces objectifs et ces opinions, La variable la plus typique pour cette classe est S (série Scientifique) avec un niveau de risque de : 0,00393.

En effet, 116 des enseignants de S parmi les 155 de cette série qui ont répondu au sondage, figurent dans le groupe optimal (GO) de cardinal 201 relatif à C. Soit X une partie aléatoire de même cardinal (155) que S et Z la variable aléatoire égale au cardinal de l'intersection de X et du groupe optimal GO. Selon un modèle équiprobable de distribution des enseignants, Z suit la loi binomiale de paramètres 155 et 201/311, soit 0,656. La probabilité pour que Z soit plus grande que 116 est le risque annoncé, soit 0,00393. Mais pour S, c'est le couple (S, (OP8, OP7)) qui est mutuellement spécifique au seuil $\beta = 2.10^{-5}$. On retrouve une telle spécificité mutuelle pour TE avec le couple (TE, (B,K)) à un seuil $\beta = 5.10^{-7}$ nous confirmant, sans surprise, que les enseignants des sections techniques (TE) considèrent que les mathématiques doivent être utiles à la vie professionnelle (B) et, en conséquence, aux autres disciplines (K) et y sont les plus attachés.

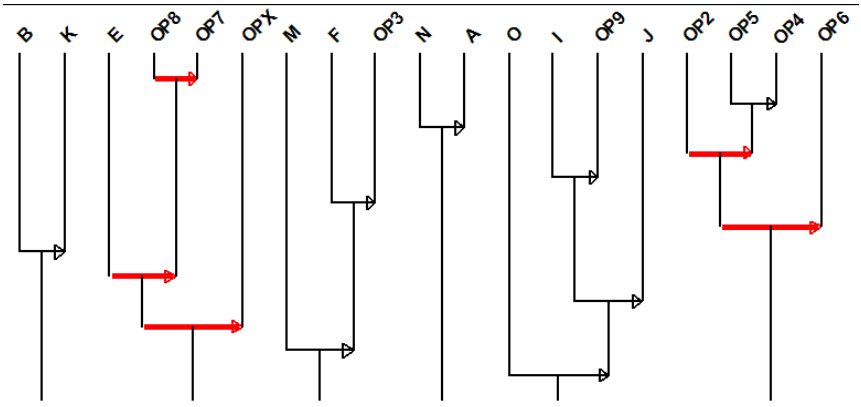


FIG. 15 - *Hiérarchie cohésitive significative*

Les calculs de contribution à la classe C montrent que, cette fois, 111 enseignants sur les 311 sondés, participent au groupe optimal. Le nombre d'enseignants de S a diminué (il passe de 116 à 67) et, surtout, sa proportion est bien moindre que précédemment dans le GO. Ceci

se ressent dans le seuil qui est 0,0251, soit un risque 6 fois plus élevé que pour la typicalité. Ce sont les enseignants sondés de S qui sont les plus typiques, c'est-à-dire « conformes » au comportement général de la population elle-même sondée. Mais ils sont moins contributeurs dans les relations strictes entre les 4 variables constituant C. Cette remarque nous montre les nuances apportées par les deux concepts : typicalité et contribution**.**

Certaines liaisons apparues et commentées ci-dessus se retrouvent dans le graphe. Les contributions calculées dans CHIC montrent encore que les enseignants de la série S contribuent le plus au chemin : $(E \Rightarrow (OP8 \Rightarrow OP7)) \Rightarrow OPX$ avec un risque d'erreur de 0,00746. La transitivité le long de ce chemin est assurée au niveau 0,75.

Ainsi on peut voir que les différents concepts mis en place ont permis d'établir une véritable dualité d'échanges d'informations d'un espace (celui des sujets) à un autre (celui des variables). Identifier le sujet ou le groupe le plus responsable de la formation de classe ou de chemin se ramène à l'identification de la variable ou du groupe de variables le ou les caractérisant.

Chapitre 6 : Règle et R-règle d'exception en Analyse Statistique Implicative ou encore l'exception confirme-t-elle la règle ?[20]

1 Introduction

On suppose une extraction de règles effectuée sur un ensemble de données binaires. Si l'on étudie localement avec attention les relations obtenues, on peut découvrir une situation parmi les associations qui défie l'intuition. C'est ainsi qu'il arrive que nous observions, entre trois variables (par exemple, des attributs) a, b et c, éventuellement conjonctions de variables binaires dans l'étude présente, les règles suivantes :a⇒c et b⇒c. Et que dans ce cas exceptionnel, on n'ait pas (a et b)⇒c, (que l'on peut aussi noter, $a \wedge b \Rightarrow c$) comme le bon sens l'attendrait, mais plutôt (a et b)⇒non(c) (que l'on peut aussi noter, $a \wedge b \Rightarrow \bar{c}$)). Cette dernière règle sera, de façon naturelle, appelée ici **règle d'exception**.

Rappelons que dans des travaux antérieurs (Suzuki et Kodratoff, 1999 ; Suzuki et Zytkow, 2005) les auteurs considèrent comme situation d'exception, la situation suivante :

a⇒c (dite règle de **sens commun**), non(b⇒c') (dite règle de **référence**) et (a et b)⇒c' (dite règle d'**exception**) où c≠c' et où *a* et *b sont respectivement des conjonctions (a = a_1 et a_2 et…et a_m)* et *(b= b_1 et b_2 et… et b_p)*. Notre définition de règle d'exception se distingue ainsi de celle-ci, mais présente, comme chez E. Suzuki et Y. Kodratoff, un caractère surprenant. Son expression plus simple lui permet d'être mieux saisie par l'intuition.

Or, il existe, comme nous le verrons en donnant des exemples, des situations naturelles où un caractère exceptionnel associe les trois variables. Pour le prendre en compte et en étudier un modèle, nous étendons ici le sens précédent en accentuant ainsi le caractère surprenant (le caractère *d'exception*) d'une règle dérivée de deux règles simples.

Pour illustrer ce type de règle, nous faisons référence tout d'abord au cas de l'incompatibilité de groupes sanguins en ce qui concerne le facteur Rhésus. Certaines femmes, non primo-parturientes, dont les globules rouges sont porteurs de deux allèles Rh- et dont l'immunisation anti-Rh+ est active, possèdent alors le phénotype Rh- (caractère *a*). Quel que soit le père en général, l'enfant qu'elles portent ne présentera pas, à la naissance, de problème sur le plan sanguin (caractère *c*). Nous sommes en présence de la règle : *a⇒c*.

Un homme, de génotype Rh+ et Rh+, possède le phénotype Rh+ (caractère *b*). Quelle que soit la mère en général, l'enfant qu'il engendrera n'aura pas de problème à sa naissance (caractère *c*). C'est la situation où la règle *b⇒c* est valide.

En revanche, un couple où la femme est Rh- et remplit les conditions *a* et l'homme est Rh+ (caractère *b*) pourra donner naissance à un enfant qui présentera un risque important du fait de l'incompatibilité Rhésus (caractère $\bar{c}$). Dans des cas exceptionnels, en effet, la mère s'immunisant contre le facteur Rh du fœtus, fabrique des anticorps, qui détruisent les globules rouges de l'enfant. Même si la conjugaison des caractères *a* et *b* est rare, on

[20] Ce chapitre a aussi été publié dans les Actes de ASI 4, Castellon, sous une forme et un contenu voisins, avec le titre : "Règle et R-règle d'exception en Analyse Statistique Implicative", Régis Gras, Einoshin Suzuki, Pascale Kuntz.

rencontre cependant la réalisation de la règle, que nous appelons « règle d'exception », (*a* et *b*) $\Rightarrow \overline{c}$. On sait d'ailleurs que des précautions sont prises pour éviter ce problème dès que sont connus les phénotypes des parents grâce à une prévention adaptée (par ex. l'exsanguino-transfusion).

On trouve une situation comparable d'apparition de règle d'exception dans l'étude des phénomènes d'interférences lumineuses, par exemple, dans l'expérience classique des franges de Young (Bruhat, 1959). La même source lumineuse franchissant deux fentes identiques (*a* et *b*) conduit à des franges d'interférences où alternent des zones d'intensité lumineuse (*c*) variable susceptible de faiblir et/ou s'annuler ($\overline{c}$).

Ces deux exemples montrent l'intérêt d'examiner le fonctionnement, la représentation et surtout les conditions d'apparition des règles d'exception afin de se prémunir en A.S.I. des inférences un peu rapides sur la stabilité implicative de la conjonction de règles simples et, comme nous allons le constater de règles généralisées ou R-règles.

2 Interprétation et illustration des règles d'exception

Soient *A*, *B*, *C* et $A \cap B$ respectivement les sous-ensembles d'individus de *E* qui satisfont respectivement les variables *a*, *b*, *c* et (*a* et *b*). Dans la situation illustrée ici, elles sont binaires, mais l'A.S.I. permet de considérer également d'autres types de variables (Gras, 2005).

2.1 Deux approches pour la caractérisation des règles d'exception

Supposons la situation prototypique des règles d'exception : $a \Rightarrow c$, $b \Rightarrow c$ et (*a* et *b*) $\Rightarrow \overline{c}$ (alors que (*a* et *b*) $\Rightarrow c$ est de piètre qualité). Elle s'exprime, en termes ensemblistes, par une quasi-inclusion des ensembles d'instances à savoir : *A* et *B* sont presque contenus dans C, mais $A \cap B$ est plutôt contenu dans le complémentaire de *C*. L'illustration ci-dessous rend compte de la situation ensembliste.

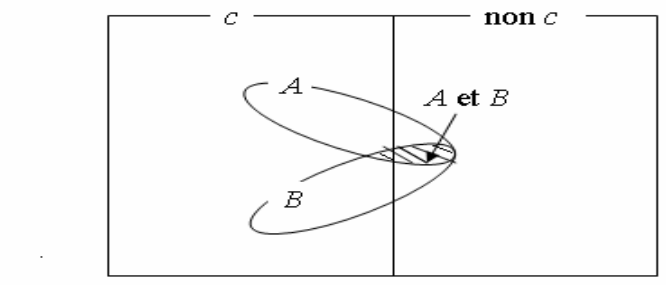


FIG. 16- – *Apparition d'une règle d'exception ensembliste*

Dans le cadre de l'A.S.I., deux approches pourraient nous permettre de mettre en évidence cette situation.

2.1.1 La première approche

Elle est basée sur l'analyse de l'intensité d'implication φ(*a*,*b*) selon la théorie présentée dans le chapitre 1 de cette partie 1. Elle nous permet de conclure au rejet de (a et b)⇒c et, a contrario, à l'apparition d'une intensité, non négligeable quelquefois, de (*a* et *b*)⇒$\bar{c}$ qui en justifie la prise en compte en tant que règle d'exception. Une représentation en graphe des relations implicatives entre règles élémentaires ci-dessus sera illustrée plus loin.

2.1.2 La deuxième approche

Elle est basée sur l'extension, que nous avons proposée et exposée au chapitre 4, des règles en *R*-règles (règles de règles) de type *R*⇒*R'* où *R* et *R'* sont elles-mêmes des règles (Gras et Kuntz, 2005). Rappelons la métaphore intuitive : ces règles sont comparables à celles qui apparaissent en mathématiques où un théorème *R* a pour conséquence un autre théorème *R'* ou est suivi d'un corollaire *R'*. Bien évidemment, il ne s'agit que d'une métaphore puisque en A.S.I. on considère des règles partielles, qui ne sont donc pas strictes et ne relèvent de la logique formelle qu'exceptionnellement. Nous avons vu qu'elles étaient construites selon un algorithme récursif utilisant un indice appelé « cohésion » et que celui-ci rendait compte de la qualité des liaisons implicatives des variables de la règle *R* avec les variables de la règle *R'*.

Rappelons, qu'en logique formelle, la règle généralisée, règle de règles, ou *R*-règle, *a*⇒(*b*⇒*c*), composée de la règle R_1 =(*b*⇒*c*) et R_2 = *a*⇒R_1 est vraie en même temps que (a et b)⇒c, donc lui est logiquement équivalente, où les variables *a, b* et *c* peuvent être elles-mêmes des règles (Gras et Kuntz., 2005). Or, nous avons vu, dans l'examen des règles élémentaires de la forme α⇒β que (*a* et *b)* ⇒ $\bar{c}$, règle d'exception, est, généralement, en contradiction sémantique avec (*a*⇒*c* et *b*⇒*c*) et que cette conjonction est plutôt formellement compatible avec (*a* et *b*) ⇒*c.*

De la même façon, la R-règle *a*⇒ *(b*⇒*c*) est en contradiction formelle avec (*a* et *b*) ⇒ $\bar{c}$. Mais comme nous sommes dans le cadre de l'A.S.I. où les règles sont partielles, cette dernière règle peut apparaître bien qu'elle soit inattendue. Nous dirons alors, comme précédemment, que (*a* et *b*)⇒$\bar{c}$ est une règle d'exception de la R-règle *a*⇒*(b*⇒*c*). Un arbre hiérarchique, présenté plus loin, permet d'illustrer cette approche par des R-règles.

2.2 Exemple numérique à partir de données fictives

Nous avons construit un fichier fictif de 200 sujets sur lesquels nous observons les variables binaires : a, b , a ∧ b, c et non(c) dont nous rapportons un extrait des 20 premiers sujets.

sujets	a	b	a ∧ b	c	non(c)
1	1	1	1	0	1
2	1	1	1	0	1
3	1	1	1	0	1
4	0	1	0	0	1
5	1	0	0	1	0
6	0	1	0	1	0
7	1	0	0	1	0
8	0	1	0	1	0
9	1	0	0	1	0
10	0	1	0	1	0

sujets	a	b	a ∧ b	c	non(c)
11	1	0	0	1	0
12	0	1	0	1	0
13	1	0	0	1	0
14	0	1	0	1	0
15	1	0	0	1	0
16	0	1	0	1	0
17	1	0	0	1	0
18	0	1	0	1	0
19	1	0	0	1	0
20	0	1	0	1	0
….	….	….	….	….	….

TAB. 21

Ce sont les 4 premiers qui vont principalement intervenir dans l'apparition de la règle d'exception. Les valeurs associées des différentes intensités sont données dans TAB 23. Elles sont obtenues par le logiciel CHIC (Couturier et Gras, 2005) qui permet les calculs et les représentations graphiques des ensembles de règles extraites des instances,

	a	*b*	*c*	$\bar{c}$	$a \wedge b$
a	0	.79	.89	.08	.89
b	.79	0	.84	.10	.88
c	.68	.67	0	0	.36
$\bar{c}$	.32	.33	0	0	.64
$a \wedge b$	1.00	1.00	.03	.97	0

TAB. 22–*Intensités d'implication associées à un jeu de données*

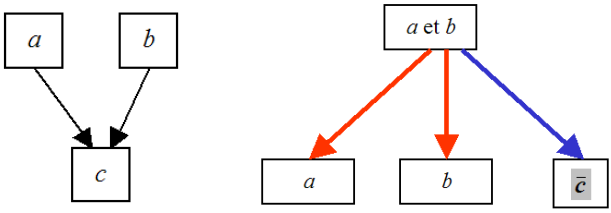
Notons les fréquences des occurrences des variables : $n_a = n_b = 12$; $n_{a \wedge b} = 7$; $n_c = 50$. Les intensités d'implication associées sont :

$$\varphi(a,c) = 0{,}89 \text{ ; } \varphi(b,c) = 0{,}84 \text{ ; } \varphi((a \text{ et } b), c) = 0{,}03$$

alors que $\varphi((a \text{ et } b), \bar{c}) = 0{,}97$; ce qui confirme la présence d'une règle d'exception.

2.2.1 Selon la première approche de règles élémentaires

Une analyse par CHIC sur le tableau complet donne le graphe implicatif et l'on constate la bonne qualité d'implication de *a* et de *b* sur *c*. On constate aussi que l'on a bien $a \Rightarrow c$ et $b \Rightarrow c$. Lorsque CHIC conjoint les variables, on obtient cette fois le phénomène lié à l'existence d'une règle d'exception.



Règles simples *Apparition de la règle d'exception*

FIG. 17– *Graphe implicatif avec les règles simples puis conjointes*

D'une façon générale en A.S.I., trois conditions nous semblent favorables à l'apparition d'une règle d'exception de $a \wedge b$ sur non c :

4. Une certaine qualité d'implication de a et de b sur c ; cette condition de bon sens conduit à ce que la règle $(a \wedge b) \Rightarrow c$ soit attendue et non pas $(a \wedge b) \Rightarrow \bar{c}$ qui en définitive va l'être;
5. Une mauvaise qualité de ressemblance entre $(a \wedge b)$ et c ($n_{a \wedge b \wedge c}$ est faible) ;
6. Une bonne qualité de ressemblance entre a et b si le référentiel devient $\bar{c}$ ($n_{a \wedge b \wedge \bar{c}}$ est grand relativement à $n_{a \wedge b}$).

2.2.2 Selon la deuxième approche

Selon la deuxième approche, relativement à cet exemple numérique, $a \Rightarrow b$, faiblement (.81), et $a \Rightarrow c$ et $b \Rightarrow c$ un peu plus fortement (.85 et .89). Par suite, la R-règle $a \Rightarrow (b \Rightarrow c)$ est validée par l'ASI et restituée au moyen du logiciel CHIC. Les arbres cohésitifs suivants illustrent respectivement d'une part la règle généralisée $a \Rightarrow (b \Rightarrow c)$ où l'on voit que a et b n'ont aucune relation implicative avec non(c), d'autre part que la règle d'exception $(a$ et $b) \Rightarrow \bar{c}$ est obtenue lorsque l'on conjoint a et b.

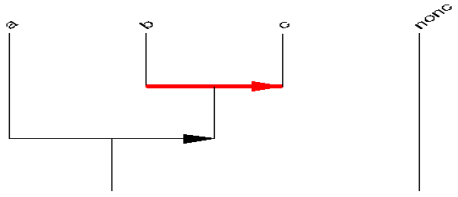


FIG. 18- *Représentation hiérarchique de la R-règle* a⇒(b⇒c)

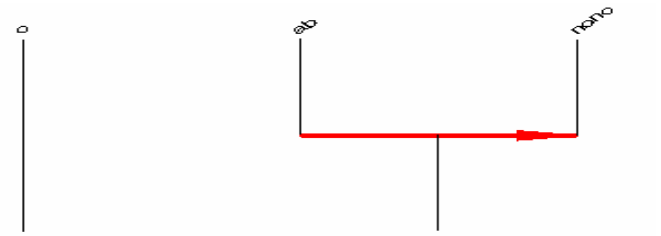


FIG. 19- *Représentation hiérarchique de la R-règle d'exception (a et b) $\Rightarrow \bar{c}$*

Autrement dit, l'absence de cohérence entre l'arbre extrait des données et l'arbre après conjonction des variables témoigne de l'apparition de la règle d'exception. Celle-ci a pu être observée de façon analogue à travers l'absence de cohérence entre les deux représentations des graphes implicatifs.

Ainsi, comme pour l'approche graphique ci-dessus, la double construction de la hiérarchie implicative : variables élémentaires puis variables conjointes, permet par examen de la non-cohérence de repérer l'existence d'une R-règle d'exception.

3 Relation entre les intensités d'implication de *a et b* sur *c* et sur non(*c*)

Rappelons ce nous avons exposé en détail dans le chapitre 1 de cette partie 1, que, en A.S.I., nous modélisons l'implication de *a* sur *b* de deux manières :

1. par une loi de Poisson de paramètre estimé $\lambda = \frac{n_a n_{\bar{b}}}{n}$;
2. par une loi binomiale de paramètres *n* et $p = \frac{n_a n_{\bar{b}}}{n^2}$

Une modélisation hypergéométrique est écartée car elle n'induit pas de différence entre une implication et sa réciproque (Gras et al., 1996 c)

Établissons pour chacun de ces deux modèles retenus les intensités d'implication de la conjonction $a \wedge b$ sur les variables *c* et non(*c*) (encore notée $\bar{c}$). Nous utiliserons la relation simple : $n_{a \wedge b \wedge \bar{c}} = n_{a \wedge b} - n_{a \wedge b \wedge c}$.

3.1 L'intensité d'implication de la conjonction dans le modèle de Poisson

Il convient d'expliciter le calcul des intensités d'implication mettant en jeu des conjonctions de variables.

3.1.1 Première approche par règles élémentaires

Dans ce modèle, pour respectivement l'implication $a \wedge b \Rightarrow \bar{c}$ et l'implication $a \wedge b \Rightarrow c$, les indices $q_1(a \wedge b,$ c) et $q_2(a \wedge b, \bar{c})$ sont :

$$q_1 = \frac{n_{a\wedge b\wedge c} - \frac{n_{a\wedge b}.n_c}{n}}{\sqrt{\frac{n_{a\wedge b}.n_c}{n}}} \qquad \text{et} \qquad q_2 = \frac{n_{a\wedge b\wedge \bar{c}} - \frac{n_{a\wedge b}.n_{\bar{c}}}{n}}{\sqrt{\frac{n_{a\wedge b}.n_{\bar{c}}}{n}}} \qquad \textbf{(1)}$$

Pour que l'implication $a \wedge b \Rightarrow \bar{c}$ soit de bonne qualité, il est nécessaire que q_1 soit négatif. En effet, le nombre de contre-exemples observé $n_{a\wedge b\wedge c}$ doit être inférieur à celui auquel seul le hasard pourrait conduire, dans l'hypothèse d'indépendance de $a \wedge b$ et de c, soit la moyenne $\frac{n_{a\wedge b}.n_c}{n}$. Des formules (1), on déduit que

$$q_1 = \frac{n_{a\wedge b} - n_{a\wedge b\wedge \bar{c}} - \frac{n_{a\wedge b}.n_c}{n}}{\sqrt{\frac{n_{a\wedge b}.n_c}{n}}} = - \frac{n_{a\wedge b\wedge \bar{c}}.n - n_{a\wedge b}(n - n_c)}{\sqrt{n.n_{a\wedge b}.n_c}}$$

$$= \frac{n_{a\wedge b\wedge \bar{c}} - \frac{n_{a\wedge b}.n_{\hat{c}}}{n}}{\sqrt{\frac{n_{a\wedge b}.n_c}{n}}} = -q_2 . \frac{\sqrt{\frac{n_{a\wedge b}.n_{\bar{c}}}{n}}}{\sqrt{\frac{n_{a\wedge b}.n_c}{n}}}$$

Finalement : $q_1 = -q_2\sqrt{\frac{n_{\bar{c}}}{n_c}}$ ou encore que le rapport $\frac{q_1}{q_2} = -\sqrt{\frac{n_{\bar{c}}}{n_c}}$.

q_1 et q_2 sont bien de signes opposés, ce qui est conforme à l'intuition. Mais de plus, l'amplitude de la positivité de q_2 induit celle de la négativité de q_1.

Au sens de l'intensité d'implication (classique), pour que la règle $a \wedge b \Rightarrow \bar{c}$ soit considérée comme une exception et apparaisse, la différence $\varphi(a \wedge b, \bar{c})$ - $\varphi(a \wedge b, c)$ suivante doit être positive et suffisamment grande :

$$\frac{1}{\sqrt{2\pi}} \int_{q_1}^{+\infty} e^{-\frac{t^2}{2}} dt - \frac{1}{\sqrt{2\pi}} \int_{-q_1\sqrt{\frac{n_c}{n_{\bar{c}}}}}^{+\infty} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{q_1}^{-q_1\sqrt{\frac{n_c}{n_{\bar{c}}}}} e^{-\frac{t^2}{2}} dt \qquad \textbf{(2)}$$

Proposition 10 Il y a donc apparition de règles d'exception, lorsque q_1 est négatif, c'est-à-dire lorsque $\frac{n_{a\wedge b\wedge c}}{n} < \frac{n_{a\wedge b}}{n}.\frac{n_c}{n}$ (sous indépendance) et de qualité d'autant meilleure que l'ensemble C des instances satisfaisant c est supérieur à celui qui satisfont sa négation non(c). La force de l'intensité d'exception sera à la mesure de la valeur de l'intégrale gaussienne sur l'intervalle $[q_1 ; -q_1\sqrt{\frac{n_c}{n_{\bar{c}}}}]$. De même, règle attendue et règle d'exception coïncident lorsque $q_1 = q_2 = 0$.

Ainsi, c'est à l'occasion de l'indépendance de $a \wedge b$ et $\bar{c}$ et donc de $a \wedge b$ et c que disparaît la règle d'exception.

3.1.2 Deuxième approche par R-règles

Dans le cadre de l'approche par R-règles, considérant la règle ($b \Rightarrow c$) comme une variable binaire, c'est-à-dire prenant respectivement la valeur 0 lorsque b=1 alors que c=0 et la valeur 1 dans les autres cas, les contre-exemples à la règle sont en nombre $n_{b \wedge \bar{c}}$. Dans ces conditions, les contre-exemples à la R-règle $a \Rightarrow (b \Rightarrow c)$ apparaissent lorsque a prend la valeur 1 alors que ($b \Rightarrow c$) prend la valeur 0, c'est-à-dire lorsque $b \wedge \bar{c}$ prend la valeur 1. Par suite, le nombre de ces contre-exemples est $n_{a \wedge b \wedge \bar{c}}$ et l'indice d'implication associé à la R-règle, dans une modélisation de Poisson de l'implication statistique, est alors :

$$q_3 = \frac{n_{a \wedge b \wedge \bar{c}} - \frac{n_a . n_{b \wedge \bar{c}}}{n}}{\sqrt{\frac{n_a . n_{b \wedge \bar{c}}}{n}}} = \frac{n . n_{a \wedge b \wedge \bar{c}} - n_a . n_{b \wedge \bar{c}}}{\sqrt{n . n_a . n_{b \wedge \bar{c}}}}$$

On constate que cet indice est différent de celui associé à la règle $a \wedge b \Rightarrow c$ qui est la règle élémentaire attendue de la conjonction $a \Rightarrow c$ et $b \Rightarrow c$. En conséquence, les indicateurs qui nous permettront de prévoir l'existence d'une règle d'exception dans cette approche hiérarchique seront différents de ceux qui nous permettent d'anticiper l'exception dans l'approche par le graphe implicatif.

Rappelons alors que l'indice d'implication de (a et b) $\Rightarrow \bar{c}$ est :

$$q_1 = \frac{n_{a \wedge b \wedge c} - \frac{n_{a \wedge b} . n_c}{n}}{\sqrt{\frac{n_{a \wedge b} . n_c}{n}}} = \frac{n . n_{a \wedge b \wedge c} - n_{a \wedge b} . n_c}{\sqrt{n . n_{a \wedge b} . n_c}}$$

On démontre, par transformation des deux indices, que q_1 est négatif (donc la règle d'exception est valide) alors que q_3 est positif (donc la règle attendue n'apparaît pas) si et seulement si : $\frac{n_{a \wedge b}}{n_a} > \frac{n_{b \wedge \bar{c}}}{n_{\bar{c}}}$, c'est-à-dire si la fréquence conditionnelle de b dans a est supérieure à sa fréquence dans $\bar{c}$.

Proposition 11 : $q_1 < 0$ et $q_3 > 0 \Leftrightarrow \frac{n_{a \wedge b}}{n_a} > \frac{n_{b \wedge \bar{c}}}{n_{\bar{c}}}$

Inversement, si cette inégalité est observée dans l'autre sens, la règle attendue apparaît alors que la règle d'exception n'existe pas.

Dans l'exemple numérique qui illustre la règle d'exception $a \wedge b \Rightarrow \bar{c}$, nous avons :

a) d'une part : $n_{a \wedge b} = 7$, $n_b = 24$, soit $\frac{n_{a \wedge b}}{n_a} = 0{,}29$,

b) d'autre part : $n_{b \wedge \bar{c}} = 8$, $n_{\bar{c}} = 100$ soit $\frac{n_{b \wedge \bar{c}}}{n_{\bar{c}}} = 0{,}08$.

L'inégalité est bien satisfaite.

Ce résultat analytique est différent de celui obtenu par l'approche graphique ce qui confirme bien la différence de signification des deux représentations de l'implication.

3.2 L'intensité d'implication de la conjonction dans le modèle binomial

Posons q'$_1$ et q'$_2$ respectivement les indices respectifs d'implication de $a \wedge b \Rightarrow \bar{c}$ et $a \wedge b \Rightarrow c$, lorsque le modèle de tirage aléatoire des parties *A*, *B* et *C* est binomial. Dans ce cas, par un calcul comparable au modèle précédent, on obtient

$$\frac{q'_1}{q'_2} = - \sqrt{\frac{n_{\bar{c}}\,(n^2 - n_{a\wedge b}.n_{\bar{c}})}{n_c\,(n^2 - n_{a\wedge b}.n_c)}} \qquad \textbf{(3)}$$

Posons

$$k(a,b,c) = \left[\frac{\left(1 - \frac{n_{a\wedge b}.n_{\bar{c}}}{n^2}\right)}{\left(1 - \frac{n_{a\wedge b}.n_c}{n^2}\right)} \right]^{\frac{1}{2}} \qquad \textbf{(4)}$$

Donc,

$$\frac{q'_1}{q'_2} = - \sqrt{\frac{n_{\bar{c}}}{n_c}}.\,k(a,b,c) \qquad \textbf{(5)}$$

et la différence $\varphi(a \wedge b, \bar{c})$ - $\varphi(a \wedge b, c)$ entre les intensités d'implication est

$$-\frac{1}{\sqrt{2\pi}} \int_{q'_1}^{+\infty} e^{-\frac{t^2}{2}} dt - \frac{1}{\sqrt{2\pi}} \int_{-q'_1\sqrt{\frac{n_c}{n_{\bar{c}}}}.\,k\,(a,b,c)}^{+\infty} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{q'_1}^{-q'_1\sqrt{\frac{n_c}{n_{\bar{c}}}}.\,k(a,b,c)} e^{-\frac{t^2}{2}} dt \qquad \textbf{(6)}$$

Proposition 12 : Pour le modèle binomial, la différence entre les intensités d'implication sera non seulement fonction du rapport $\frac{n_c}{n_{\bar{c}}}$ mais aussi de *k*(*a*,*b*,*c*) (**4**). Ce coefficient est d'autant plus grand et renforce ainsi l'effet du rapport $\frac{n_c}{n_{\bar{c}}}$ que $1 - \frac{n_{a\wedge b}.n_{\bar{c}}}{n^2} >> 1 - \frac{n_{a\wedge b}.n_c}{n^2}$, c'est-à-dire $\frac{n_{a\wedge b}}{n}.\frac{n_c}{n} >> \frac{n_{a\wedge b}}{n}.\frac{n_{\bar{c}}}{n}$. Les deux membres de cette inégalité ne sont autres que, de gauche à droite, les probabilités respectives du nombre de contre-exemples aléatoires - dans le modèle binomial où les variables $a \wedge b$ et c seraient indépendantes - des implications $a \wedge b \Rightarrow \bar{c}$ et $a \wedge b \Rightarrow c$. Ainsi, plus on s'attend à une réfutation de $a \wedge b \Rightarrow \bar{c}$, au vu de $n_{a\wedge b}$ et de $\bar{c}$, plus le caractère surprenant, *exceptionnel*, de cette règle est manifeste

par le constat de la modicité des contre-exemples observés à savoir $n_{a \wedge b \wedge c}$. Ceux-ci la valident au détriment de $a \wedge b \Rightarrow c$. L'inégalité montre la « contribution active à l'exception » au rapport $\frac{n_c}{n_{\bar{c}}}$, contribution qu'apportent les instances, de cardinal $n_{a \wedge b}$ dans celles de cardinal $n_{\bar{c}}$.

Cette conséquence, liée au modèle binomial, nous apparaît donc plus riche que celle énoncée dans le modèle de Poisson. En effet, elle nous fournit une relation de contrôle entre les paramètres plus fine du caractère d'exception que dans le modèle de Poisson. Ce phénomène, certes lié au nombre de paramètres de définition du modèle binomial, le gratifie cependant d'un intérêt que le logiciel CHIC permet d'exploiter à travers l'offre de son menu.

Remarque : A titre de comparaison, si nous nous intéressons à un autre indice de mesure de qualité de règle, à savoir la *confiance c*, qui est à la base des principaux autres indices de qualité (Lenca et al. 2004), nous obtenons les propriétés suivantes. Celle-ci s'exprime ainsi :

c($a \Rightarrow c$)= $\frac{n_{a \wedge c}}{n_a}$ (souvent notée : $\frac{\Pr[a \wedge c]}{\Pr[a]}$, autrement dite probabilité conditionnelle de *c* sachant *a*). La relation entre les règles que nous avons examinées est alors :

$$c(a \wedge b \Rightarrow \bar{c}) = \frac{n_{a \wedge b \wedge \bar{c}}}{n_{a \wedge b}} = 1 - \frac{n_{a \wedge b \wedge c}}{n_{a \wedge b}} = 1 - c(a \wedge b \Rightarrow c)$$

La règle d'exception a pour mesure le complément à 1 de la règle attendue. Ainsi, elle est indépendante des valeurs des occurrences.

4 Conclusion

Pour conclure et résumer, lorsque deux variables impliquent une 3ème, que leur conjonction implique plutôt la négation de cette 3ème, nous considérons que cette règle est d'exception, en un sens voisin mais différent de celui de E. Suzuki et Y.Kodratoff (1999). Nous avons étudié et illustré par un exemple numérique et un exemple de génétique, l'expression de ce caractère exceptionnel. Puis nous avons précisé les relations entre les paramètres des variables dans les deux modélisations selon lesquelles est construite l'Analyse Statistique Implicative : un modèle de Poisson et un modèle binomial, l'un et l'autre convergeant vers le même modèle gaussien.

Nous avons évoqué une approche complémentaire pour la détection de ces règles qui se base sur les travaux menés ces dernières années sur les R-règles (Gras et Kuntz, 2005 et chap. 4). La construction associée d'une hiérarchie implicative n'a pas été initialement développée dans ce but. Cependant, elle constitue une piste à explorer tant d'un point vue algorithmique que méthodologique concernant l'interprétation de ce que pourraient être des « R-règles d'exception ». Mais que ce soit pour des règles ou des R-règles d'exception, le signalement de ces semi-paradoxes par rapport au sens commun prouve, s'il en était nécessaire, *le saut conceptuel qu'impose le passage de la logique formelle de l'implication à la logique des quasi-implications,* objet de l'A.S.I..

Chapitre 7 : Extraction[21] de Règles en Incertain par l'Analyse Statistique Implicative[22]

Partant du cadre défini et formalisé par (Zadeh 1979, 2001), par (Dubois et Prade 1987), ce texte vise à étudier les proximités formelle et sémantique des cadres de l'incertain et de l'analyse statistique implicative (A.S.I.) entre variables à valeurs intervalles et variables-intervalles (Gras 2001a). On ne rappellera pas les formalisations classiques des notions premières et de chaque opérateur de la **logique floue**. On s'intéressera plus particulièrement à l'opérateur **«** implication » à l'aide duquel on extrait des règles d'association. Nous considérons celles qui croisent des sujets (ou des objets) et des variables, présentant des modalités nettes ou floues. Rappelons qu'une règle entre deux variables ou entre conjonctions de variables est établie sur la base de la rareté statistique du nombre de ses contre-exemples, dans l'hypothèse de l'indépendance a priori des variables en jeu (Gras 1979), (Lerman et al 1981) et chapitre 1 de la partie 1 de cet ouvrage. La qualité de la règle, avons-nous écrit, sera évidemment d'autant plus grande que ce nombre de contre-exemples sera invraisemblablement petit sous cette hypothèse, eu égard aux occurrences des variables et des instances totales.

Dans le premier sous-chapitre, nous présentons la problématique. Puis, nous construisons, de façon peu classique, une distribution floue à partir de données objectives. Ensuite, nous abordons la recherche de règles d'association dans une situation « floue » en nous appuyant auparavant sur la notion de variables modales. Enfin, nous revenons sur la construction des règles en ramenant les variables floues à des variables-intervalles.

1 Problématique. Un exemple prototypique de situation en incertain

Bien que les applications de la logique floue soient nombreuses en intelligence artificielle (par exemple en matière de diagnostic médical, de reconnaissance des formes ou de recherche de panne), plusieurs questions restent bien souvent latentes : comment obtient-on des distributions des degrés d'appartenance à un intervalle dans le cas de variables numériques ? Sur quelles connaissances sont-elles établies ? Sont-elles données a priori et mises à l'épreuve de la réalité ou bien sont-elles des construits ? S'il s'agit de ce dernier cas, quel processus d'extraction de connaissances à partir de données peut y conduire et quel type de règle peut-on alors extraire dans ce cadre ? Quelle signification peut-on donner à une règle associant deux sous-ensembles ou deux attributs flous ? On rejoint alors une des problématiques du data mining et de la qualité des règles, ce qui justifie notre préoccupation en A.S.I...

[21] Ce texte a été présenté avec une forme voisine et un contenu réduit sous le titre : « Extraction de règles en incertain par la méthode statistique implicative », dans les *Comptes rendus des 12èmes Rencontres de la Société Francophone de Classification, Montréal 30 mai-1er juin 2005, UQA.* Les auteurs étaient Régis Gras, Raphaël Couturier, Fabrice Guillet et Filippo Spagnolo.

[22] Remerciements à Maurice Bernadet pour sa lecture du texte et ses précieux conseils

Voici un exemple prototypique d'une situation où les données sont floues. U est l'univers de référence, du discours dit-on, portant sur 3 **modalités ou attributs flous** relatifs à la taille d'individus : {petit, moyen, grand}et se confondant avec l'ensemble E des observations potentielles ou réalisées des attributs. Si l'on veut accorder un **degré d'appartenance**, de la forme $\mu_T(x)$, aux sous-ensembles associés à ces 3 attributs respectivement : T_1=petit, T_2 =moyen, T_3 = grand, à des individus x de l'ensemble E des sujets, on obtient par exemple selon 3 d'entre eux :

a) $x=i_1$ est un individu dit plutôt petit, alors $\mu_{T1}(i_1)$ =0,8, $\mu_{T2}(i_1)$ =0,5, $\mu_{T3}(i_1)$ = 0,2, c'est-à-dire pour ce sujet, un fort degré d'appartenance à la classe T_1, faible degré à la classe T_3.

b) $x=i_2$ est un individu dit pas très grand, alors $\mu_{T1}(i_2)$ = 0,1, $\mu_{T2}(i_2)$ = 0,6, $\mu_{T3}(i_2)$ = 0,7.

c) $x=i_3$ est un individu dit plutôt grand, alors $\mu_{T1}(i_3)$ = 0, $\mu_{T2}(i_3)$ = 0,7, $\mu_{T3}(i_3)$ = 0,9.

Les données sont ici floues. Aux sous-ensembles de U, on associe une **pseudo-partition** définie par d'autres modalités telles que « pas très grand », « plutôt grand », etc.. « pas très », « plutôt »,… qui sont appelées des **modificateurs linguistiques**. Ils permettent de définir de nouveaux **sous-ensembles flous** à partir des sous-ensembles précédents. On peut dans ce cas considérer qu'à chaque sujet, à chaque observation, correspond un vecteur « net » dans un espace de dimension le nombre des modalités de U.

Traditionnellement, dans la théorie du flou, on représente les distributions des tailles floues d'une façon comparable à la suivante en plaçant les 3 sujets, en fonction de ces distributions supposées ici affines par morceaux :

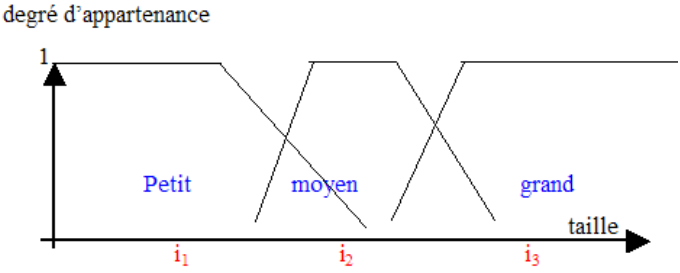


FIG. 20-– *Une représentation d'un exemple*

Mais les fonctions d'appartenance peuvent être différentes de celle choisie ici ; citons : courbe Bêta, courbe de Gauss, courbe triangulaire, etc.

On constate que les sous-ensembles flous ont des intersections généralement non vides.

2 Deux méthodes de construction de distributions floues par extraction de connaissances

Dans le cadre que nous retenons, les distributions des degrés d'appartenance seront le fruit de l'interaction entre connaissances objectives (une vraie valeur de la variable, un

attribut net ou modificateur linguistique *consensuel*) et connaissances subjectives. Dans la littérature, les degrés sont des données. D'où proviennent-elles ?

L'exemple ci-dessus illustre le cas d'un échantillon d'individus donné. On disposera **effectivement** de leur taille s (un nombre) ou des caractères ou attributs nets : « petit », « moyen » et « grand » au vu d'une décision consensuelle du type : les caractères « petit », « moyen » et « grand » seront attribués **objectivement** au regard de leur taille mesurée. Face à ces données, on pourra comparer le point de vue **subjectif** portant sur les mêmes individus qui énoncera qu'un sujet de 179 cm n'est pas petit, mais peut être considéré de taille grande ou moyenne, non contradictoirement.

Différentes méthodes pour définir la distribution des attributs visent à effectuer un processus de « **fuzzification** » (Bernadet, 2004) : définition des classes floues pour chaque attribut, puis mise en correspondance de chaque attribut avec un degré d'appartenance à un sous-ensemble flou, comme nous le voyons dans l'exemple introductif. Dans (Zadeh, 1997), une méthode de discrétisation optimale est donnée. Ici, nous procèderons autrement.

3 Relation entre intervalles nets et attributs flous

Notre objectif, dans ce paragraphe, est de « *fuzzifier* » en quantifiant le degré d'appartenance d'un sujet à un intervalle numérique donné. Pour ce faire, la méthode de type « clustering » que nous proposons consiste, à partir du choix d'un indice de similarité, ici celui de la vraisemblance du lien de I.C. Lerman (1981), d'extraire, tout d'abord, la proximité entre les attributs nets et les attributs flous.

Auparavant, selon le procédé défini dans (Gras 2001) et présenté au chapitre 3, nous choisissons de transformer l'ensemble des valeurs observées sur les sujets en sous-intervalles disjoints de variance inter-classe maximale afin de pouvoir attribuer à chaque sous-intervalle un attribut net de même désignation que celle attribuée aux attributs flous. Cette partition nette est établie par la méthode des nuées dynamiques de (Diday 1972). Enfin, pour chaque classe de similarité entre **attribut net** et **attribut flou**, nous déterminons le degré d'appartenance des sujets à une classe floue à partir de la mesure normalisée de typicalité associée à chaque individu. En effet, cette typicalité, définie dans (Gras et al. 2006) et présentée au chapitre 5, rend compte d'un degré de responsabilité dans la proximité d'attributs, soulignant l'accord entre net et flou. Ainsi, nous disposerons d'une mesure vérifiant les axiomes de Zadeh relatifs au concept de « possibilité ». Mais, son avantage par rapport à la détermination subjective classique est qu'elle est établie à l'épreuve statistique de la réalité et qu'elle varie avec la dilatation de l'ensemble des sujets.

En résumé, les données initiales sont de deux ordres :

- d'une part, des variables **objectives, consensuelles** aux valeurs numériques réparties sur des intervalles auxquels on associe respectivement un **attribut net** ,
- d'autre part, un **attribut flou** attribué **subjectivement** à chaque sujet.

Exemple : Les données portent sur 60 sujets. Leurs tailles T nettes vraies varient de 168 et 198 cm. Appliquant l'algorithme de la variance inter-classe maximale, nous obtenons, par l'algorithme, une partition constituée des intervalles nets: « Tpeti » de 168 à 174, « Tmoy » de 175 à 183, Tgran de 184 à 198. En outre, les attributs flous, issus d'un jugement subjectif, sont notés respectivement TP, TM et TG. Ainsi, par exemple, un sujet de taille vraie 180 cm

sera classé dans la classe nette Tmoy mais aussi simultanément dans les classes floues TM et TG. La hiérarchie des similarités donnée par CHIC entre ces 6 variables, est alors :

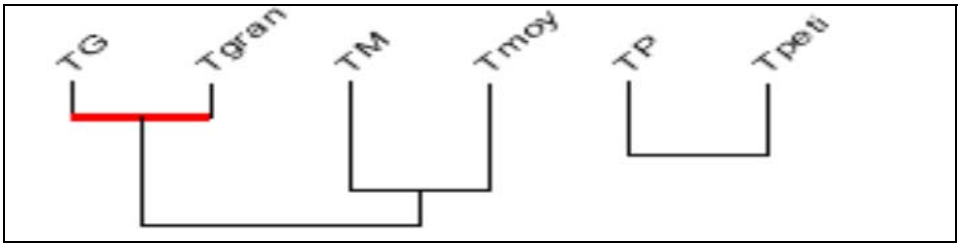


FIG. 21- – *Hiérarchie des similarités entre les intervalles nets et flous*

On note que les attributs nets s'associent aux intervalles flous correspondants, ce que raisonnablement on pouvait attendre d'une subjectivité normale. Un sujet pourra donc posséder TM et TG suivant le point de vue du juge si sa taille n'est pas manifestement grande.

Le logiciel CHIC ((Couturier et Gras 2005) et Partie 2, chap 11 et 12), restitue les mesures de typicalité des sujets selon les 3 classes de similarité. Rappelons qu'un individu est d'autant plus typique d'une classe qu'il a une attitude conforme à la constitution de la classe par la population de sujets.. On observe alors en consultant les calculs de typicalités donnés par CHIC que, par ex., le sujet i06 a une typicalité de 0 sur les tailles petites, 0,056 sur les tailles moyennes et 0,95 sur les tailles grandes. On peut faire de même pour les autres sujets de l'échantillon. Ce sont ces valeurs que nous retenons comme degrés d'appartenance respectifs par rapport aux attributs flous.

Pour itérer le procédé dans le but d'affiner les distributions, il suffit, par ex, de remplacer une des modalités d'attribut ou chacune d'entre elles par 2 modalités. Ainsi, « grand » sera subdivisé en « très grand » et « assez grand », « moyen » en « pas très grand » et « pas très petit », etc. La partition de l'intervalle des tailles se fera sur la base de 6 intervalles.

« Attribuer des degrés d'appartenance à partir des typicalités aux associations observées sur un échantillon » nous paraît atténuer l'arbitraire habituel des affectations de ces degrés.

4 Construction de l'histogramme d'une variable-intervalle à partir des données floues des sujets

Cette fois, on dispose de la distribution des valeurs floues prises par chaque sujet d'une population sur un intervalle. On cherche à en déduire une distribution des degrés d'appartenance sur cet intervalle. L'objectif final est de définir une variable symbolique, dite aussi variable-intervalle, qui soit l'histogramme d'un intervalle sur lequel on pourra déterminer des sous-intervalles optimaux selon le critère de la variance.

Soit f_1, f_2, …,f_n les fonctions d'appartenance respectives des n sujets à un intervalle A. On suppose, par analogie avec les densités, que ces fonctions sont normalisées sur A. Dans la majorité des cas, chaque sujet contribue de la même façon à la densité, sinon une pondération adaptée ramène à un problème analogue. Alors la fonction $f=(f_1+f_2+ \ldots.+f_n)/n$ intègre en un histogramme sur A la distribution des fonctions d'appartenance. Il suffit ensuite de

discrétiser A en une suite de points pondérés selon f ; enfin, d'appliquer sur A l'algorithme de la variance selon la méthode des nuées dynamiques pour obtenir une variable-intervalle a dont on pourra étudier les relations implicatives avec les autres variables du même type.

Par ex., on donne les valeurs floues de notes obtenues sur [0 ; 20] par 3 étudiants i_1, i_2, i_3 : une correction multiple affecte à i_1, i_2 et i_3 respectivement des notes : 5 à 9, 6 à 11 et 8 à 15

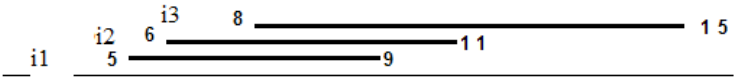


FIG. 22– *Une représentation de l'échelonnement de notes*

Supposant la distribution uniforme des valeurs floues, normalisées sur [0 ; 20], selon chacun des intervalles, on obtient le tableau (TAB. 23) des fonctions d'appartenance : par ex. sur [0,3 ; 0,55], correspondant à l'intervalle de notes [6 ;10], f_2=1/5 sur chacun des 5 intervalles d'amplitude 0.05 et f_2= 0 ailleurs.

individus	Modalités de a				
↓	[0.25; .30]	[0.30 0.40].	[0.40; .45]	[0.45; 0.55]	[0.55; 0.75]
i1	1/4	2/4	1/4	0	0
i2	0	2/5	1/5	2/5	0
i3	0	0	1/7	2/7	4/7

TAB. 23 – *Valeurs prises par les modalités sur les 3 sujets*

par ex. A est discrétisable en 420 (ppcm de 3, 4, 5, 7) points

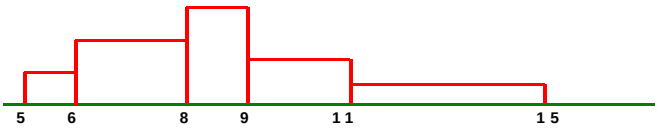


FIG. 23- *Histogramme associé*

5 Règles d'association pour des variables numériques

On suppose dorénavant, à titre d'exemple prototypique, que les distributions des variables floues sont connues selon 2 variables observées sur les mêmes sujets : taille et poids. Mais cet exemple a la vertu de généralité. On veut étudier, maintenant, comme en ASI, les règles de déduction entre l'attribut taille et l'attribut poids, présentant des modalités, l'un **Taille** = {petit, moyen, grand}, l'autre **Poids** = {léger, moyen, lourd}.

On dispose de données sous forme d'un tableau numérique des degrés d'appartenance aux modalités d'attributs flous, valeurs relatives à un échantillon de 20 sujets. Les 3 premiers de ces sujets constituent le tableau (TAB. 24). L'un d'entre eux, i_1, n'est donc pas très grand et pas très lourd, l'autre i_2 assez grand et assez lourd, le dernier i_3 plutôt grand et plutôt lourd.

	taille			**poids**		
	petit T_1	*moyen* T_2	*grand* T_3	*léger* P_1	*moyen* P_2	*lourd* P_3
i_1	8/15	5/15	2/15	7/14	4/14	3/14
i_2	1/14	6/14	7/14	2/15	5/15	8/15
i_3	0	7/16	9/16	1/16	6/16	9/16

TAB. 24 – *Valeurs prises par les modalités sur les 3 sujets*

5.1 Un premier traitement de variables numériques

On propose ici un traitement implicatif, selon l'A.S.I., en considérant les 6 variables tailles-poids comme des variables numériques sur l'ensemble des 20 sujets. On obtient le graphe implicatif en utilisant l'indice de (Lagrange 1998), réactualisé par (Régnier et Gras. 2004) et (Partie 1, chap. 3 et Partie 2, chap. 1). Ainsi, les implications T3 ⇒ P3 et P1 ⇒ T1 sont valides au seuil 0,90 et signifient que les propositions grand ⇒ lourd et léger ⇒ petit, règle qui est sémantiquement contraposée de la première, sont acceptables. Une autre implication à un seuil supérieur à 0,6 apparaît : $P_2 \Rightarrow T_1$.

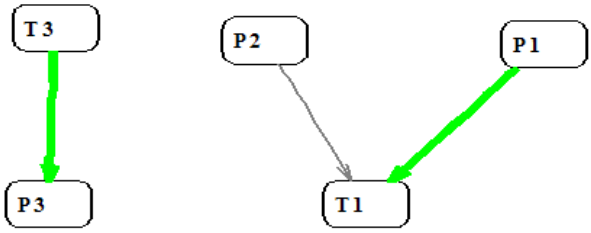


FIG. 24– *Graphe implicatif taille x poids*

Ces résultats ne s'opposent pas, bien entendu, au bon sens. Les autres règles d'association confirment une meilleure adéquation à la sémantique de l'implication qu'avec les approches de Reichenbach et Lukasiewicz (cité dans (Dubois et Prade, 1987)). On ne retrouve pas, par ex. : léger ⇒ grand.

Mais, l'approche proposée ici présente l'inconvénient de considérer que les 6 modalités des variables « taille » et « poids » sont actives dans le traitement et ne restituent pas, ainsi, les nuances de leur structure. Il semble donc intéressant, sémantiquement parlant, de revenir à la considération de modalités de variables de type intervalles où les modalités apparaissent comme sous-intervalles d'une variable-intervalle principale.

5.2 Second traitement par des variables à valeurs intervalles

Ce second traitement (Gras et al. 2001 a) et (chapitre 3) va permettre de prendre en compte de façon plus fine les nuances des observations prises selon des sous-ensembles flous et de répartir leurs valeurs de façon optimale sur un intervalle numérique [0 ;1], selon une partition dont l'utilisateur définit le nombre de classes pour chacun de 20 sujets.

Nous disposons d'un nouveau tableau donnant les distributions des 6 modalités des 2 attributs « taille » et « poids » relativement à chacun des individus et les valeurs binaires prises par 2 variables supplémentaire « Femme », « Homme ». En voici les 2 premières lignes.

	Taille petite pt	**Taille moyenne m**	**Taille grande T**	**Var. supp. Femme**	**Var. supp. Homme**	**Poids léger L**	**Poids moyen o**	**Poids grand P**
$\mathbf{i_1}$	0,7	0,4	0,3	1	0	0,8	0,3	0,1
$\mathbf{i_2}$	0,2	0,5	0,8	0	1	0,1	0,4	0,9

TAB. 25 – *Distributions des attributs flous « taille » et « poids »*

Par ex., le sujet i_1 admet un degré d'appartenance 0,7 à la classe des sujets petits, 0,4 à celle des sujets de taille moyenne et 0,3 à la classe des sujets de grande taille. De plus (variable supplémentaire), ce sujet est une femme et la distribution de ses degrés d'appartenance aux 3 classes de poids, sont respectivement 0,8, 0,3 et 0,1. Le traitement va emprunter cette fois la méthode des variables à valeurs intervalles. Chaque modalité conduira à la construction de sous-intervalles optimaux, c'est-à-dire la détermination de sous-intervalles optimisant, du moins localement sinon globalement, l'inertie inter-classe. Utilisant ensuite CHIC de traitement de ce type de variable, on établit les règles telles que : si un sujet relève de l'intervalle t_i de la modalité t de l'attribut « taille » alors, généralement, il relève de l'intervalle p_j de la modalité p de l'attribut « poids ». Ainsi, si par ex., il a tendance à être plutôt petit, alors il a généralement tendance à être plutôt léger.

Les partitions en 3 sous-intervalles calculées par CHIC sont données dans le tableau ci-dessous.

tailles petites:	tailles moyennes :	grandes tailles :
t1 de 0 à 0.1	m1 de 0.1 à 0.3	T1 de 0 à 0.1
t2 de 0.2 à 0.5	m2 de 0.4 à 0.6	T2 de 0.2 à 0.5
t3 de 0.6 à 1	m3 de 0.8 à 0.8	T3 de 0.8 à 0.9
poids légers :	poids moyens :	poids lourds :
L1 de 0 à 0.2	o1 de 0.2 à 0.3	P1 de 0 à 0.1
L2 de 0.3 à 0.6	o2 de 0.4 à 0.5	P2 de 0.2 à 0.4
L3 de 0.8 à 1	o3 de 0.6 à 0.7	P3 de 0.7 à 0.9

TAB. 26 – *Partitions optimales calculées par CHIC*

Le graphe implicatif au niveau de confiance 0,90 est également donné par CHIC :

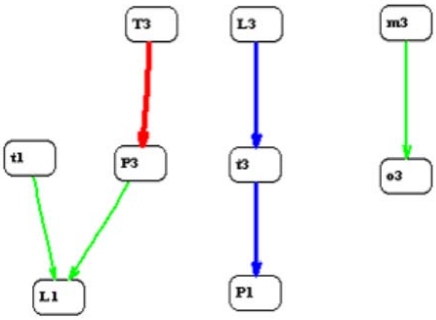


FIG. 25– *Graphe implicatif taille x poids*

On voit par exemple que :

- l'individu de grande taille (T3) admet généralement un poids important (P3) et donc n'est pas considéré comme léger(L1). Ce sont les hommes qui apportent, et de très loin (risque de se tromper = 0.07), la plus importante contribution ;

- l'individu de poids plutôt léger (L3) est généralement de petite taille (t3) ;dans ce cas, ils ne sont que très rarement considérés lourds (P1). Ce sont les femmes qui sont les plus contributives à ce chemin (risque = 0.25) ;

- les deux variables t1 et L1, liées par la règle t1⇒L1, correspondent à des fréquences rares. Si donc, on rencontre un sujet petit alors il est généralement léger. Le sexe Homme contribue à la formation de cette règle.

6 Conclusion

A l'aide de l'A.S.I., nous avons cherché à **objectiver** la notion de degré d'appartenance. Situant le modèle d'implication entre attributs par rapport à des modèles classiques, nous avons mis en évidence par un graphe, les relations implicatives entre des modalités de variables numériques. Nous avons, semble-t-il, amélioré la formalisation de la sémantique en faisant référence à des variables-intervalles. Les règles les plus consistantes ont pu être extraites selon leur qualité. Enfin, la relation entre des variables extrinsèques et ces règles ont permis d'enrichir notre connaissance sur ces règles. Des applications à des situations réelles tenteront de valider cette nouvelle approche de l'incertain. D'ores et déjà, nous savons quel intérêt, par exemple dans le détection de l'origine de pannes les industriels ont accordé au traitement du problème à l'aide de la logique du flou.

Chapitre 8 : Réduction du nombre de variables[23]

1 Introduction

L'Extraction de Connaissances dans les Données (ECD) a pour objectif la découverte de connaissances utiles cachées dans des données volumineuses (Frawley et al., 1992), (Fayyad et al., 1996). Lorsque ces données se ramènent à une table croisant sujets et variables, la notion de volume se traduit par un grand nombre de lignes (sujets), ce qui est statistiquement intéressant, mais aussi un grand nombre de colonnes (variables) ce qui peut l'être moins.

En effet, dès que le nombre de variables devient pléthorique, la plupart des techniques disponibles deviennent impraticables. En particulier, lorsque l'on procède à une analyse implicative par calcul de règles d'association (Agrawal et al., 1993), le nombre de règles découvertes, subit une explosion combinatoire avec le nombre de variables, et devient rapidement inextricable pour un décideur, pour peu que soient demandées des conjonctions de variables. Dans ce contexte, il s'avère nécessaire de procéder à une réduction préliminaire du nombre de variables.

Ainsi, (Ritschard et al., 2000) ont proposé une heuristique efficace permettant de réduire à la fois le nombre de lignes et de colonnes d'une table, à partir d'une mesure d'association servant de critère de quasi-optimalité pour piloter l'heuristique. Cependant, à notre connaissance, dans les différentes autres recherches, le type de situation à l'origine de la nécessité du regroupement de lignes ou de colonnes n'est pas pris en compte dans les critères de réduction, que la problématique et la visée de l'analyste soient la recherche de similarité, de dissimilarité, d'implication, etc., entre variables.

Aussi, dans la mesure où il existe des variables très voisines au sens de l'implication statistique, il pourrait être opportun de substituer à ces variables une seule qui serait leur leader en termes de représentation d'une classe d'équivalence de variables similaires pour la visée implicative.

Nous nous proposons donc, à l'instar de ce qui est fait pour définir la notion de quasi-implication, de définir une notion de quasi-équivalence entre variables, afin de construire des classes d'où nous extrairons un leader. Nous l'illustrerons par un exemple. Ensuite, nous envisagerons la possibilité d'utiliser un algorithme génétique afin d'optimiser le choix du représentant de chaque classe de quasi-équivalence.

2 Définition de la quasi-équivalence

Deux variables binaires a et b sont logiquement équivalentes pour l'A.S.I. lorsque sont simultanément satisfaites, à un seuil donné, les deux quasi-implications : $a \Rightarrow b$ et $b \Rightarrow a$. Nous avons conçu des critères pour évaluer la qualité d'une quasi-implication : l'un est l'étonnement statistique inspiré de la vraisemblance du lien de Lerman (1981), l'autre est la

[23] Ce chapitre s'inspire fortement de l'article intitulé, tout en l'étendant : « Réduction des colonnes d'un tableau de données par quasi-équivalence entre variables », publié dans les actes de EGC 2 (janvier 2002), Cépadués, avec pour co-auteurs Régis Gras, Fabrice Guillet, Robin Gras et Jacques Philippé.

forme entropique de la quasi-inclusion (Gras et al. 2001) qui est présentée dans le chapitre 3 de cette partie de l'ouvrage.

Selon le *premier critère*, on pourrait dire que deux variables a et b sont quasi-équivalentes lorsque l'intensité d'implication φ(a,b) de a⇒b est peu différente de celle de b⇒a. Cependant, sur des ensembles importants (plusieurs milliers), ce critère n'est plus suffisamment discriminant pour valider l'inclusion.

Selon le *deuxième critère*, on se base, sur une mesure entropique du déséquilibre entre, d'une part, les effectifs $n_{a\wedge b}$ (individus qui satisfont a et b) et $n_{a\wedge \bar{b}}$ (individus qui satisfont a et non(b), contre-exemples à l'implication a⇒b) pour signifier la qualité de l'implication a=⇒b, et d'autre part, les effectifs $n_{a\wedge b}$ et $n_{\bar{a}\wedge b}$ pour évaluer la qualité de l'implication réciproque b⇒a.

Ici nous allons utiliser ici une méthode comparable à celle utilisée dans le chapitre 3 pour définir l'indice d'implication entropique.

En posant n_a et n_b, respectivement effectifs de a et de b, le déséquilibre de la règle a⇒b est mesuré par une entropie conditionnelle K(b|a=1), et celui de b⇒a par K(a|b=1) avec :

$$K(b \mid a=1) = -(1-\frac{n_{a\wedge b}}{n_a})\log_2(1-\frac{n_{a\wedge b}}{n_a}) - (\frac{n_{a\wedge b}}{n_a})\log_2(\frac{n_{a\wedge b}}{n_a}) \text{ si } \frac{n_{a\wedge b}}{n_a} > 0{,}5 \quad (\mathbf{1})$$

$$K(b|a=1)=1 \text{ si } \frac{n_{a\wedge b}}{n_a} \le 0{,}5 \quad (\mathbf{2})$$

$$K(a \mid b=1) = -(1-\frac{n_{a\wedge b}}{n_b})\log_2(1-\frac{n_{a\wedge b}}{n_b}) - (\frac{n_{a\wedge b}}{n_b})\log_2(\frac{n_{a\wedge b}}{n_b}) \text{ si } \frac{n_{a\wedge b}}{n_b} > 0{,}5 \quad (\mathbf{3})$$

$$K(a|b=1)=1 \text{ si } \frac{n_{a\wedge b}}{n_b} \le 0{,}5 \quad (\mathbf{4})$$

Ces deux entropies doivent être suffisamment faibles pour que l'on puisse, avec une bonne certitude, parier sur b (resp. sur a) lorsque a (resp. b) est réalisé. Par conséquent leurs compléments respectifs à 1 doivent être simultanément forts.

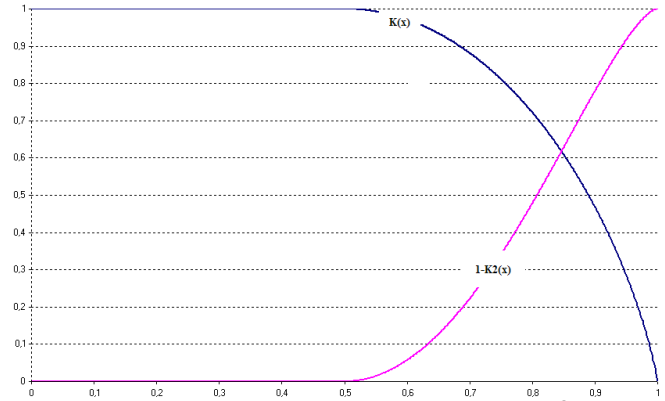


FIG. 26- *représentation graphique des fonctions* K *et* 1-K^2 *sur* [0 ;1]

Définition 36: Un premier **indice entropique d'équivalence** est donné par :

$$e(a,b) = \left(\left[1 - K^2(b \mid a = 1)\right]\left[1 - K^2(a \mid b = 1)\right]\right)^{\frac{1}{4}}$$

Quand cet indice prend des valeurs dans le voisinage de 1, cela traduit une bonne qualité d'une double implication. De plus, afin de mieux prendre en compte $a \wedge b$ (les exemples), nous intégrons ce paramètre à travers un indice de similarité s(a,b) des variables, par exemple au sens de I.C. Lerman (1981). L'indice de quasi-équivalence est alors construit par conjugaison de ces deux notions.

Définition 37: Un second **indice entropique d'équivalence** est donné par la formule

$$\sigma(a,b) = [e(a,b).s(a,b)]^{\frac{1}{2}}$$

Partant de ce point de vue, nous énonçons alors le critère de quasi-équivalence que nous retenons.

Définition 38: On dit que la **paire de variables {a,b} est *quasi-équivalente*** pour la qualité choisie β si σ(a,b)≥ β

Par exemple, une valeur β=0,95 pourra être considérée comme désignant une bonne quasi-équivalence entre a et b.

3 Algorithme de construction des classes de quasi-équivalence

Soit un ensemble V={a,b,c,…} de v variables muni d'une relation valuée R induite par la mesure de quasi-équivalence σ sur l'ensemble des paires de V. On supposera les paires de variables classées selon un ordre décroissant de quasi-équivalence. Si nous avons fixé le seuil de qualité de la quasi-équivalence à β, seules seront conservées les premières des paires {a,b} vérifiant l'inégalité σ(a,b)≥β. En général, seule une partie V', de cardinal v', des variables de V vérifiera cette inégalité. Si cet ensemble V' est vide ou trop réduit, l'utilisateur pourra ramener son exigence à une valeur seuil β plus petite. La relation étant symétrique, on disposera au plus de $\frac{v'(v'-1)}{2}$ paires à étudier. Quant à V-V', il ne contient que des variables non réductibles.

On propose d'utiliser l'algorithme glouton suivant :

1° On constitue une première classe potentielle C_1^0 = {e,f} telle que σ(e,f) représente la plus grande des valeurs de β-équivalence. Si cela est possible, on étend cette classe selon une nouvelle classe C_1 en prélevant dans V' tous les éléments x tels que toute paire de variables au sein de cette classe admette une quasi-équivalence supérieure ou égale à β.

2° On poursuit par :

a) Si o et k formant la paire (o,k) immédiatement inférieure à (e,f) selon l'indice σ, appartiennent à C_1, alors on passe à la paire immédiatement inférieure à σ(o,k) et on procède comme dans le 1°.

b) Si o et k n'appartiennent pas à C_1, on procède comme dans 1° à partir de la paire qu'ils constituent en formant la base d'une nouvelle classe C_2^0.

c) Si o ou k n'appartient pas à C_1, l'une de ces deux variables pourra soit former une classe singleton, soit appartenir une classe future. Sur celle-ci, on pratiquera bien entendu comme ci-dessus.

Après un nombre fini d'itérations, on dispose d'une partition de V en r classes de σ-équivalence : $\{C_1, C_2, \ldots, C_r\}$. La qualité de la réduction pourra être évaluée par un indice brut ou proportionnel à $\beta \frac{p}{k}$. Cependant nous lui préférons le critère défini ci-dessous qui présente l'avantage d'intégrer le choix du représentant.

De plus, on pourrait choisir k variables représentantes des k classes de σ-équivalence sur la base du critère élémentaire suivant : la qualité de liaison de cette variable avec celles de sa classe. Mais, ce critère ne permet pas d'optimiser la réduction puisque le choix du représentant est relativement arbitraire et peut-être signe de trivialité de la variable. Nous allons donc revenir sur cette question.

4 Recherche d'un critère pour déterminer un optimum de la réduction

Nous nous sommes inspirés ici des notions de variance implicative et de cohésion développées dans (Gras et al., 1996 c) et que nous retrouverons dans le chapitre 2 de la Partie 2. En particulier, on montre que si l'on considère chaque variable binaire comme un vecteur de $[0,1]^n$, alors le carré scalaire du vecteur $\vec{ab}$ traduit l'implication a⇒b et sa réciproque b⇒a. Dans le cas de variables binaires, ce carré scalaire ne s'annule que si les deux implications sont strictes (elles prennent les mêmes valeurs, 0 ou 1, selon les mêmes sujets). Toute valeur de ce carré proche de 0 traduit donc une ressemblance de a et b, tout en évitant les biais consécutifs au centrage/réduction requis par exemple pour le calcul du coefficient de corrélation.

Aussi, les k classes d'équivalence étant construites comme ci-dessus, nous pouvons expliciter une certaine inertie que nous désignons par **inertie implicative** I, à partir de l'ensemble $\{a_1, a_2, \ldots a_k\}$ de k variables qui les représentent :

$$I = \frac{1}{\sum_{i=1}^{k} k_i c_i} \left(\sum_{i=1}^{k} k_i c_i \left[\overrightarrow{a_i g}\right]^2 \right)$$

où C_i est une classe, k_i son effectif, c_i sa cohésion implicative, et g le barycentre des représentants. On vérifie que cette inertie croît avec la cohésion et l'effectif de la classe C_i. Cette grandeur I rend compte, dans un contexte implicatif, de la séparabilité des représentants des classes ainsi que de leur consistance.

5 Un exemple illustratif

Une enquête auprès d'étudiants de l'École Polytechnique de l'Université de Nantes a porté sur la recherche de règles d'association entre 41 variables, animaux sauvages ou domestiques, possédant ou non certaines qualités parmi 75 qualités proposées (par ex. « affectueux », « féroce », etc.) auxquelles on associe les 41 animaux. Ainsi «bruyant » sera accordé à « baleine », « canard », etc.. Comme la représentation du graphe exprimant les règles implicatives entre animaux risque de ne pas présenter une qualité de clarté facilitant l'analyse, nous avons utilisé le logiciel REDUCHIC, élaboré par Raphaël Couturier, afin de réduire le nombre de variables selon l'algorithme ci-dessus implémenté dans REDUCHIC. On obtient les informations suivantes :

Animaux		**Classe**
Lapin, Dauphin		*Dauphin*
Poule, Canard		*Poule*
Crocodile, Couleuvre		*Couleuvre*
Loup, Tigre, Requin		*Tigre*
Rat, Mygale		*Rat*
Ours, Lion	représentant de la classe	*Lion*
Renard, Chat		*Renard*
Lynx, Crotale		*Crotale*
Vache, Âne		*Âne*
Vautour, Corbeau		*Corbeau*
Mouche, Cigale		*Mouche*
Mouton, Chien		*Chien*

TAB. 27

Le nouveau fichier des animaux, après cette réduction par équivalence et le choix du meilleur représentant minimisant l'inertie implicative, ne comporte plus que 28 animaux, ce qui est manifestement plus aisé à représenter et à analyser. On observera la vraisemblance des équivalences extraites comme par exemple celle de la classe (Loup, Tigre, Requin)

En opérant de la même façon avec un questionnaire devant faire apparaître des traits de personnalité, nous disposions de 142 variables et 2299 sujets. Une réduction a conduit à un fichier plus compact de 34 variables, plus aisément traitable. Certaines classes d'équivalence contiennent jusqu'à 12 variables dont une seule sera retenue par la suite.

6 Détermination d'un optimum par un algorithme génétique

Notre objectif est de trouver un ensemble de k représentants maximisant l'inertie implicative I (ici de complexité en $(\frac{p}{k})^k$). Pour cela, on dispose de différentes méthodes heuristiques : par exemple, les nuées dynamiques selon la méthode développée par E. Diday et son usage dans le cadre de l'implication statistique dans (Gras et *al.* 2001), la programmation dynamique, un algorithme génétique (Goldberg 1994), etc.. C'est cette dernière méthode que nous utiliserons, car elle a l'avantage d'être efficace pour de grands espaces de recherche. Dans la population considérée par l'algorithme génétique, chaque

individu est codé par chromosome constitué de k gènes dont chaque allèle code une variable représentante d'une classe. Au début du processus chaque représentant est choisi aléatoirement.

Précisons la forme des différents opérateurs génétiques :

a) -**Sélection** : le critère de qualité utilisé pour la sélection est l'inertie implicative I.
b) -**Reproduction** : le cross-over entre deux chromosomes s'effectue par le choix aléatoire d'un rang de coupure.
c) -**Mutation** : La mutation d'un gène se fera en respectant un critère de contiguïté : la probabilité de mutation de a vers b devra être proportionnelle au carré scalaire du vecteur $\vec{ab}$.

7 Conclusion

Une implémentation spécifique a été effectuée par des étudiants de l'École Polytechnique de l'Université de Nantes à l'occasion d'un projet. Nous disposons également d'un algorithme implémenté dans le logiciel REDUCHIC qui a permis de réduire, au gré de l'utilisateur, et aussi sensiblement qu'il le souhaite, le nombre de variables à traiter. Ce logiciel a été suivi de nombreuses expérimentations et applications permettant, d'une part d'en évaluer la pertinence et l'efficacité et, d'autre part, les performances en temps de calcul. Le nouveau fichier réduit est directement intégrable au logiciel C.H.I.C et traitable par lui comme tout autre fichier .csv (Couturier et Gras, 2005). La complexité liée à la prise en considération des conjonctions de variables s'en trouve fortement allégée.

Chapitre 9 : Règles superflues ou redondantes en Analyse Statistique Implicative[24]

1 Introduction

Certaines recherches, aux objectifs différents mais avec le même souci de réduction, ont permis de diminuer sensiblement le nombre de variables en établissant des indices de similarité entre elles. C'est le cas, par exemple, des indices de Jaccard, de Russel et Rao, de Rogers et Tanimoto, de Piatetsky-Shapiro cités dans (Matheus et al, 1996). Couturier et al. (2004) définissent un indice de similarité dans le cadre de l'A.S.I. et comparent, avantageusement dans leur approche, les différents indices visant la réduction de variables. Gras et al. (2002) établissent une relation d'équivalence entre les variables en colonnes visant leur réduction comme nous venons de le voir dans le chapitre 8 précédent. Blanchard et al. (2004) ne traitent pas ce sujet mais, par contre, étudient, via la notion d'entropie, la qualité des règles de la forme a⇒b et de leur contraposée, également dans le cadre de l'A.S.I., en définissant le taux informationnel d'une règle à partir du gain d'information apporté sur b par la réalisation de a. Nous abordons ici le problème de la redondance en suivant une démarche comparable à celle de J. Blanchard. L'objectif de ce chapitre est alors : comment réduire l'ensemble des règles obtenues en ne retenant que celles qui fournissent des informations différentes, donc non surabondantes, voire non superfétatoires ?

En d'autres mots, la problématique de ce chapitre est la suivante :

On suppose que des règles ont été extraites de l'ensemble des données et, qu'en particulier, une classification orientée, dite cohésitive, organise l'ensemble des attributs en règles et méta-règles ou règles généralisées,(Gras et Kuntz., 2004) comme nous l'avons abordé dans le chapitre 4. On souhaite réduire le nombre de ces règles en conservant l'information maximale qu'elles contiennent selon deux types d'indice. Pour cela, nous envisagerons les règles généralisées d'ordre quelconque, c'est-à-dire celles dont la prémisse et la conclusion sont des règles simples, d'ordre 0 entre variables ou bien des règles généralisées, c'est-à-dire des règles de règles, d'ordres supérieurs.

Une première solution, que nous avons présentée déjà dans le chapitre 8, pourrait consister à procéder de la même façon. Il suffirait de définir une relation d'équivalence entre deux règles R et S d'ordre quelconque dès lors que nous aurions à la fois R⇒S et S⇒R à un haut niveau de qualité implicative. Ici, nous privilégions la relation non symétrique d'implication entre règles. Par suite, c'est la qualité de l'information de l'une sur l'autre qui servira plutôt de critère de réduction.

Pour ce faire, nous présentons l'approche de la réduction du nombre de règles et leur redondance au sens de l'entropie de Shannon. Dans la même perspective, nous proposons ensuite l'emploi de l'indice de Gini afin d'évaluer l'information apportée par une règle sur une autre.

[24] Une version voisine de ce texte figure en anglais sous le titre : « Reduction of Redundant Rules in Statistical Implicative Analysis. *Selected Contributions in data Analysis and Classification*, P. Brito, P. Bertrand, G. Cucumel, E. de Carvalho, (eds), Springer, p. 367-376, avec pour auteurs Régis Gras et Pascale Kuntz

2 Entropies de Shannon réduites et conditionnelles

2.1 Entropie d'une règle

Soit 2 règles. R=(a⇒b) et S=(c⇒d) où les variables a, b, c et d peuvent être elles-mêmes des conjonctions d'attributs ou, plus généralement, des règles. Soit p (resp. q) la fréquence de réalisation de R (resp. S) dans l'ensemble E des sujets.

L'entropie, au sens de Shannon, liée à R (resp. S) est :

$$H(R) = -(1-p)\log_2(1-p) - (p)\log_2(p) \text{ et } H(S) = -(1-q)\log_2(1-q) - (q)\log_2(q)$$

On rappelle que H(R) (resp. H(S)) est l'incertitude moyenne de « l'expérience » liée à la réalisation de la règle R (resp. S). En d'autres termes, c'est l'information moyenne attachée à la connaissance du résultat de l'expérience réalisant R (resp. S) ; c'est-à-dire encore, cette information est égale à la quantité d'information contenue dans la réalisation de R (resp . S). Comme ce qui nous intéresse en matière de gain informationnel est le cas où le nombre de contre-exemples à la règle est faible, eu égard aux occurrences en jeu (c'est-à-dire le cas où la fréquence p (resp. q) pour laquelle elle est vraie, est forte), nous limiterons notre étude à p (resp. q) à des valeurs supérieures ou égales à 0,5, ce qui permet une bijection de H sur l'intervalle [0,5; 1].

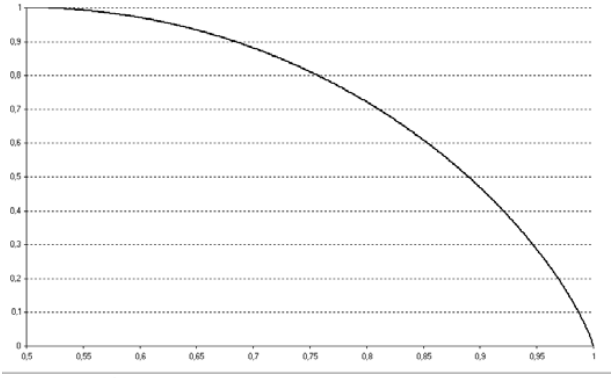


FIG. 27- *représentation graphique de la fonctions* H *sur* [0,5 ;1]

Donc, pour des raisons sémantiques, ne souhaitant pas la symétrie de H par rapport à p ou q égales à 0,5, nous posons la définition suivante.

Définition 39: On appelle ***entropie réduite*** **de la règle** R (resp. S) la quantité $H_r(R)$ (resp. $H_r(S)$) donnée par :

$H_r(R)= H(R)$ si $p \geq 0,5$ et $H_r(R)= 1$ si $p<0,5$

Si p=1 alors $H_r(R)=0$, l'incertitude au sujet de R est nulle car cette règle est certaine.

Si p=0,5 alors $H_r(R)=1$, l'incertitude au sujet de R est maximum. On ne peut faire le pari qu'elle se réalisera.

Les propriétés sont respectivement les mêmes pour S

2.2 Entropie conditionnelle d'une règle

Considérant maintenant le tableau croisant E (ensemble des individus) et les deux règles R et S considérées comme des variables prenant sur chaque individu, la valeur 1 ou 0 suivant que celui-ci vérifie ou non la règle en question.

Notons :

a) p_R (resp. p_S) la fréquence de R (resp. S) ; $p_{\overline{R}}$ (resp. $p_{\overline{S}}$) la fréquence de non(R) ou $\overline{R}$ (resp. non(S) ou $\overline{S}$)

b) p_{RS} (resp. $p_{\overline{RS}}$), la fréquence des règles R et S conjointes (resp. non(R) et S); $p_{R\overline{S}}$ (resp. $p_{\overline{RS}}$) la fréquence des règles R et non(S) (resp. non(R) et non(S))

L'entropie conditionnelle de S sachant R est alors :

$$H(S \mid R) = -p_{RS}\log_2(\frac{p_{RS}}{p_R}) - (p_{R\overline{S}})\log_2(\frac{p_{R\overline{S}}}{p_R}) - (p_{\overline{RS}})\log_2(\frac{p_{\overline{RS}}}{p_{\overline{R}}}) - (p_{\overline{RS}})\log_2(\frac{p_{\overline{RS}}}{p_{\overline{R}}})$$

Proposition 13 si R et S sont indépendantes, alors $H_r(S \mid R) = H_r(S)$

En effet $p_{RS} = p_R p_S$

$$H(S \mid R) = -p_{RS}\log_2(p_S) - (p_{R\overline{S}})\log_2(p_{\overline{S}}) - (p_{\overline{RS}})\log_2(p_S) - (p_{\overline{RS}})\log_2(p_{\overline{S}})$$

$$H(S \mid R) = -p_S\log_2(p_S) - (p_{\overline{S}})\log_2(p_{\overline{S}}) = H(S)$$

Ainsi, la connaissance de la réalisation de l'événement R ne modifie pas l'incertitude sur S.

Autre situation, si les règles R et S sont liées fonctionnellement, comme par exemple :

$$p_{RS} = p_R,\ p_{\overline{R}\,\overline{S}} = p_{\overline{R}},\ p_{R\overline{S}} = p_{\overline{RS}} = 0,$$

alors $H(S \mid R) = 0$ et il n'y a plus d'incertitude sur S.

En effet, nous obtenons en admettant que f(0)=0 est un prolongement de $f(x)=x\log_2(x)$ en x=0 : $H(S \mid R) = -p_R\log_2(\frac{p_R}{p_R}) - (0)\log_2(\frac{0}{p_R}) - (0)\log_2(\frac{0}{p_{\overline{R}}}) - (p_{\overline{R}})\log_2(\frac{p_{\overline{R}}}{p_{\overline{R}}})$

Enfin, la fréquence p_{RS} et les marges étant fixées, les autres fréquences étant déterminées, nous posons alors :

Définition 40: On appelle ***entropie conditionnelle réduite*** **de la règle S sachant R** (resp. la règle R sachant S) la quantité :

$H_r(S \mid R) = H(S \mid R)$ si $p_{RS} \geq 0{,}5$ et $H_r(S \mid R) = 1$ si $p_{RS} < 0{,}5$

$H_r(R \mid S) = H(R \mid S)$ si $p_{RS} \geq 0{,}5$ et $H_r(R \mid S) = 1$ si $p_{RS} < 0{,}5$

Ainsi, la différence $H_r(S) - H_r(S \mid R)$ est la quantité d'information ou l'incertitude contenue dans R au sujet de S, lorsque la fréquence p_{RS} est supérieure à 0,5, puisque c'est

l'accroissement de l'information sur S quand on connaît R. Autrement dit, c'est aussi la diminution de l'incertitude sur S dans l'hypothèse où la réalisation de R est connue.

2.3 Superfluité d'une règle

On pose $h(S) = \frac{\mathrm{H}(S)}{\log_2 N}$ où N est le nombre de valeurs que peut prendre S (ici N = 2 car S(x) = 1 ou 0 suivant que l'individu x satisfait ou non la règle S). Sachant que $H(S) \leq \log_2(N) = 1$ (Roubine, 1970) et que l'égalité n'a lieu que dans le cas uniforme où p=(1-p)=0,5 et, par conséquent, où l'incertitude est maximum, le rapport h(S) est une entropie *réduite relative* de S. Elle est toujours inférieure ou égale à 1. Si cette valeur est très voisine de 0, l'expérience S est quasiment *superflue* puisque l'une des deux probabilités p_S et $1\text{-}p_S$ est beaucoup plus grande que l'autre ; le « pari » sur l'une est inutile car on est presque sûr de l'issue de l'expérience.

On pose encore r(S) = 1 – h(S), pour une simple raison de compatibilité sémantique, à savoir que plus r(S) est grand, plus la superfluité est grande. r(S) est appelé ***coefficient de superfluité*** de S. On définit de la même façon h(R), entropie relative de R, et r(R), puis r(S/R) et r(R/S). Notons que $0 \leq r(S) \leq r(S/R) \leq 1$.

Définition 41: L'expérience S (donc la règle elle-même) est dite ***ε-superflue***, quand on connaît R, lorsque r(S/R), coefficient de superfluité de S sachant R, vérifie $r(S/R) \geq 1\text{-}\varepsilon$

On dit de même que R est ε-superflue, quand on connaît S, lorsque $r(R/S)) \geq 1\text{-}\varepsilon$

A travers cette notion de superfluité, nous possédons un moyen de faire décroître sensiblement le nombre de règles simples et généralisées formées au cours de l'étude hiérarchique des variables. L'utilisateur dispose d'un contrôle de la suppression éventuelle de règles en choisissant une valeur pour ε suffisamment petite. Mais, nous proposerons plus loin, un autre critère plus puissant.

Il est aisé de démontrer la symétrie :

$$H_r(S) - H_r(S \mid R) = \mathrm{I}(S,R) = \mathrm{I}(R,S) = H_r(R) - H_r(R \mid S)$$

c'est-à-dire que l'accroissement de l'information sur S quand on connaît R, autrement dit le ***gain informationnel***, est le même que l'accroissement de l'information sur R quand on connaît S. Cependant, les deux valeurs r(S) et r(R) ne sont pas nécessairement égales. Ainsi, si l'on cherche à éliminer une règle R ou S ou de type $R \Rightarrow S$, il sera bien sûr plus intéressant d'éliminer celle dont le coefficient de superfluité est le plus grand.

2.4 Redondance de règles

Afin de distinguer les gains associés aux expériences conditionnantes, nous considérons maintenant le rapport : $\frac{H_r(S) - H_r(S \mid R)}{H_r(S)}$ qui n'est pas négatif et qui, cette fois, n'est pas symétrique par rapport à R et S. Il représente un gain d'information relativisé par la grandeur de $H_r(S)$. Comme $H_r(S) \geq H_r(S \mid R)$ pour tout R, ce rapport est, par définition, égal à 0 quand $H_r(S) = 0$. Il varie entre 0 et 1.

Si R et S sont strictement liées, alors $H_r(S \mid R) = 0$, car il n'y a aucune incertitude sur S quand R est connue. La réciproque n'est pas vraie et, tout au moins, peut-on suspecter une certaine liaison de S à R.

Par ailleurs, si R et S sont indépendantes, alors $H_r(S) = H_r(S \mid R)$. La réciproque n'est pas vraie ; par contre, on peut suspecter l'indépendance.

Définition 42: Soit deux règles R et S. On appelle ***gain d'information relatif*** de S par R le rapport : $G_r(S \mid R) = \dfrac{H_r(S) - H_r(S \mid R)}{H_r(S)}$ et $G_r(S \mid R) = 1$ si $H_r(S) = 0$

Notons que :

Si $G_r(S/R) = 1$, alors nécessairement $H_r(S \mid R) = 0$. Dans ce cas, la diminution de l'incertitude sur S, soit la quantité d'information contenue dans R au sujet de S est maximum. Il est fort probable que S et R soient liées. S serait ***redondant****e* par rapport à R

Si $G_r(S/R) = 0$, la quantité d'information contenue dans R ou apportée par R au sujet de S est égale à ce qu'elle était sans connaître R. Il est fort probable que R et S soient indépendantes.

Si $H_r(S) = 0$, alors $H_r(S \mid R) = 0$. Un simple calcul montre alors que, de la même façon, $H_r(R \mid S) = 0$ et le prolongement par continuité en 0 de $G_r(S/R)$ se justifie aussi sémantiquement.

Définition 43: Soit deux règles R et S. On dit que S est ***ε-redondante*** *par rapport à R* si $G_r(S/R) \geq 1 - \varepsilon$

Étant donnée une suite de règles produite par une hiérarchie ordonnée, selon l'analyse statistique implicative, cette définition nous permet de réduire sensiblement, au gré de l'utilisateur, par action sur la valeur minimum acceptable 1- ε, le nombre de ces règles en conservant celles qui assurent un maximum d'information. L'algorithme que nous utilisons pour une automatisation de la réduction par rapport à la règle R de plus forte intensité d'implication et de plus forte fréquence dans E. R étant donnée, de proche en proche, on compare les superfluités et les redondances des règles d'intensité décroissante en éliminant celles qui sont superflues ou redondantes à un seuil 1- ε. A l'issue de ces éliminations, on itère le processus avec une règle R' d'intensité inférieure à R et on effectue les mêmes calculs parmi les règles restantes. Et ainsi de suite. Avec le même objectif de réduction du nombre de règles, nous allons maintenant examiner une autre approche dans un cadre conceptuel différent.

3 Information mutuelle au sens de l'indice de Gini

L'indice de Gini permet de mettre en évidence une distribution « inégalitaire » dans une population (comme par exemple les revenus !). Il semble donc adapté, comme l'était, dans notre choix précédent, l'entropie de Shannon, pour signifier et quantifier la dispersion au sein d'une distribution relative à la réalisation ou non de règles dans E. Étalonner d'une deuxième manière le « désordre » et donc la qualité informative d'une telle distribution nous paraît

d'un intérêt certain. Rappelons que l'indice de Gini est un cas particulier, pour α=2 de l'α-entropie de Havrda et Charvàt (1967) définie ainsi pour une variable R dont la distribution des probabilités (ou des fréquences) sur ses k valeurs est (p_1, p_2,…,p_k) ;

$$H(R) = \frac{1}{1-\alpha}\left[\sum_i p_i^{\alpha} - 1\right]$$

L'indice de Gini pour une telle variable binaire est donc :

$$Gini(R) = 1 - \sum_i p_i^2$$

On peut démontrer, par un développement au premier ordre, que quand α tend vers 1, la limite de l'ε-entropie est justement l'entropie de Shannon. D'où la proximité des sémantiques de l'une et l'autre.

Interprétons l'indice de Gini :

a) $Gini(R) = 1 - \sum_i p_i^2$ peut encore s'écrire $Gini(R) = \sum_i (1 - p_i) p_i$ puisque la somme des probabilités est égale à 1. L'indice de Gini peut donc s'interpréter en terme de variance, donc de quantité d'information : celle d'une somme de variables aléatoires indépendantes de Bernoulli de paramètres respectifs p_i ;

b) $Gini(R) = 1 - \sum_i p_i^2$ peut s'interpréter comme l'écart entre les normes de deux vecteurs de dimension k et respectivement de composantes toutes égales à $\frac{1}{\sqrt{k}}$ pour le 1er vecteur, les autres étant égales à p_i pour le 2ème vecteur.

Par exemple, dans le cas qui nous intéresse (satisfaction ou non d'une règle), k est égal à 2 et le minimum de variance ou celui de d'écart est obtenu pour p = 1-p = 0,5. Propriété que nous avions déjà observée avec l'entropie de Shannon. On démontre, par étude de la limite, que lorsque p tend vers 1 (très bonne représentativité de R), la différence entre H(R) et Gini(R) est approximativement 1,6 p(1-p)>0. De plus, quand p>0,5, H(R) reste meilleure que Gini(R). Autrement dit, pour l'intervalle de variation de p qui nous intéresse, l'entropie de Shannon est plus informative que l'indice de Gini. Ou, de façon équivalente, sauf pour l'utilisateur, que l'indice de Gini est plus sévère que l'entropie de Shannon.

Comme nous l'avons fait pour l'entropie de Shannon, on considère l'indice conditionnel de Gini dans la définition suivante (Jaroszewicz et Simovici 2001)[25] :

Définition 44: Deux règles R et S étant données, de modalités respectives avec les fréquences p_i pour R, p_{ij} pour (R, S), q_j pour S, **l'indice conditionnel de Gini** de S sachant R est :

$$Gini(S \mid R) = 1 - \sum_i \sum_j \frac{p_{ij}^2}{p_j}$$

[25] Nous reprenons ici la démarche de Jaroszewicz S. et Simovici D. mais en spécifiant leur approche, comme celle de J. Blanchard portant sur l'indice informationnel entre variables, à des règles généralisées.

On pourra remarquer que la somme a le sens d'une somme de variances conditionnelles généralisées et, donc, contient encore le sens d'une information.

Dans le cas qui nous intéresse ici, les modalités de R et de S sont « vrai » et « faux », suivant que l'individu x satisfait ou non la règle R ou la règle S. La formule donnant l'indice est alors avec les notations adoptées dans ce chapitre :

$$Gini(S \mid R) = 1 - \left(\frac{p_{RS}^2}{p_R} + \frac{p_{R\bar{S}}^2}{p_R} + \frac{p_{\bar{R}S}^2}{p_{\bar{R}}} + \frac{p_{\bar{R}\bar{S}}^2}{p_{\bar{R}}} \right)$$

On peut alors définir, comme avec l'entropie de Shannon, le gain de Gini qui sera aussi un gain informationnel sur S quand on connaît R :

Définition 45: Deux règles R et S étant données, on appelle ***gain de Gini*** *pour S sachant R*, l'accroissement d'information suivant :

$$gainGini(S \mid R) = Gini(S) - Gini(S \mid R)$$

Ce gain apparaît comme une différence de variances et, par conséquent, un indicateur de la qualité de l'information fournie sur S par la connaissance de R. Remarquons que si R et S sont indépendantes, alors $Gini(S) = Gini(S \mid R)$ et $gainGini(S \mid R) = 0$, comme pour le gain de Shannon. La réciproque n'est pas vraie. On peut, comme avec le gain informationnel relatif selon l'entropie de Shannon, relativiser le gain de Gini et nous donner ainsi un deuxième critère pour obtenir une réduction de l'ensemble des règles acceptées sous une qualité implicative ou cohésitive compatible avec l'attente de l'utilisateur. Prochainement, une comparaison de l'efficacité de réduction selon ces deux modes sera conduite par simulation.

4 Conclusion

Dans le cadre de l'ASI et après l'obtention de règles simples ou généralisées, nous avons cherché un moyen qui puisse conduire à la réduction du nombre de telles règles obtenues. Le moyen retenu dans ce texte est centré sur la prise en compte de l'information respective apportée par la réalisation d'une règle si l'on connaît la réalisation d'une autre. Pour cela, nous avons fait appel à deux méthodes qui ne se ramenant pas conceptuellement l'une à l'autre pourraient s'avérer complémentaires : la première en examinant l'entropie conditionnelle d'une règle quand on en connaît une autre ; la deuxième en utilisant l'indice de Gini conditionnel. Comparer ces deux méthodes fera l'objet de prochaines recherches.

Références

Acid S. de Campos, L.M., A. Gonzalez, R. Molina and N. Perez de la Blanca (1991). Learning with Castle, in R. Kruse, P. Siegel (Eds) *Symbolic and quantitative Approaches to incertainty*, Springer-Verlag, 99-106

Ag Amouloud, S. (1992). *L'ordinateur, outil d'aide à l'apprentissage de la démonstration et de traitement de données didactiques*, Thèse de doctorat de l'Université de Rennes 1.

Agrawal, R., T. Imielinsky and A. Swami (1993). Mining association rules between sets of items in large databases, *Proc. of the ACM SIGMOD'93*, 207-216

Amarger, S., D. Dubois and H. Prade (1991). Imprecise quantifiers and conditional probabilities, in R. Kruse, P. Siegel (Eds), *Symbolic and quantitative approaches to uncertainty* Springer-Verlag, 33-37.

Aze, J. et Y. Kodratoff (2001). Évaluation de la résistance au bruit de quelques mesures d'extraction de règles d'association, *Extraction des connaissances et apprentissage*, Hermès, Vol 1, n° 4, 143-154

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

Bailleul, M. et R. Gras (1995). L'implication statistique entre variables modales, *Mathématique, Informatique et Sciences Humaines*, Paris : E.H.E.S.S., n°128, 41-57

Benzecri, J.P. (1973*). L'analyse des données* (vol 1), Paris : Dunod,.

Bernard, J.-M. et S. Poitrenaud (1999). L'analyse implicative bayesienne d'un questionnaire binaire : quasi-implications et treillis de Galois simplifié", *Mathématiques, Informatique et Sciences Humaines*, n° 147, 25-46

Bernadet, M., G. Rose and H. Briand (1996). FIABLE and fuzzy FIABLE : two learning mechanisms based on a probabilistic evaluation of implications, *Conference IPMU'96*, Granada, 911-916

Bernadet, M. (2004). Qualité des règles et des opérateurs en découverte de connaissances floues. *Mesure de qualité pour la fouille de données*, RNTI-E-1, Toulouse *:* Cépaduès Éditions,169-192

Blanchard, J., P. Kuntz, F. Guillet and R. Gras (2003). Implication intensity: from the basic statistical definition to the entropic version , *Statistical Data Mining and Knowledge Discovery*, Chapman & Hall/CRC, Washington, 473-485

Blanchard, J., P. Kuntz, F. Guillet and R. Gras (2004), Mesure de la qualité des règles d'association par l'intensité d'implication entropique, *Mesures de qualité pour la fouille de données*, RNTI-E-1, Toulouse *:* Cépaduès Éditions, 33-44

Blanchard, J., F. Guillet, R. Gras et H. Briand (2004) Mesurer la qualité des règles et de leurs contraposées avec le taux informationnel TIC, *Extraction et Gestion des Connaissances*, Volume 1, RNTI, Cépaduès. 287-298.

Références

Blanchard, J., F. Guillet, H. Briand et R. Gras (2005). Ipee : Indice probabiliste d'écart à l'équilibre pour l'évaluation de la qualité des règles, *Extraction et Gestion des Connaissance : état et perspectives*, RNTI-E-5, Toulouse *:* Cépaduès Éditions, 391-395.

Bodin, A. (1997). Modèles sous-jacents à l'analyse implicative et outils complémentaires. *Prépublication* IRMAR. n°97-32, 1-24

Bodin, A. et R. Gras (1999). Analyse du préquestionnaire enseignants avant EVAPM-Terminales, *Bulletin n°425 de l'Association des Professeurs de Mathématiques de l'Enseignement Public*, Paris 772-786,

Brin, S., R. Motwani and C. Silverstein (1997) Beyond market baskets : generalizing association rules to correlations", *Proc. Of ACM SIGMOD Conf. On Management of Data SIGMOD'97*, 265-276

Brousseau, G. (1986). Fondements et méthodes de la didactique des mathématiques, *Recherche en Didactique des Mathématiques, 4/2*, Grenoble La Pensée Sauvage.

Bruhat, G. (1959). *Optique*, Paris : Masson.

Couturier, R. (2001). Traitement de l'analyse statistique implicative dans CHIC, Actes des Journées sur la *Fouille dans les données par la méthode d'analyse implicative*, IUFM Caen, 33-50.

Couturier, R, R. Gras and F. Guillet (2004). Reducing the number of variables using implicative analysis *In International Federation of Classification Societies, IFCS 2004, Springer Verlag: Classification, Clustering, and Data Mining Applications*, Chicago, 277--285.

Couturier, R. et R. Gras (2005). CHIC : Traitement de données avec l'analyse implicative, *Extraction et Gestion des Connaissances*, Volume I1, RNTI, Toulouse *:* Cépaduès Éditions, 679-684

Cox, D. R. and N. Wermuth (2004) Causality : a Statistical View, *International Statistical Review*, International Statistical Institute, 72, 3, 285-305,.

David, J., F. Guillet, R. Gras and H. Briand (2006). Conceptual hierarchies matching : an approach based on discovery of implication rules between concepts, In Proc. ECAI 2006, 17th European Conference on Artificial Intelligence, IOS Press, Riva del Garda, Italy

Diday, E. (1972). *Nouvelles méthodes et nouveaux concepts en classification automatique et reconnaissance des formes*, Thèse d'État, Université de Paris VI.

Diday, E. et M.O Menessier (1991). Analyse symbolique pour la prévision de séries chronologiques pseudo-périodiques in *Induction symbolique-numérique à partir de données"*, Cépaduès Editions.

Dubois, D. et H. Prade (1987). *Théorie des possibilités. Applications à la représentation des connaissances en informatique*, Paris : Masson.

Durkheim, E.(1897). *Le suicide*, Paris :PUF.

Fayyad, U., G. Piatetsky-Shapiro and P. Smyth (1996). From Data Mining to Knowledge Discovery. In *Advances In Knowledge Discovery and Data Mining, Fayyad U., Piatetsky-Shapiro G., Smyth P, and Uthurusamy R. eds*, AAAI/MIT Press, 1-31,.

Frawley, W., G. Piatetski-Shapiro and C. Matheus (1992). Knowledge discovery in databases : an overview. *AI Magazine*. 14(3), 57-70.

Fleury, L. (1996). *Extraction de connaissances dans une base de données pour la gestion de ressources humaines*, Thèse de doctorat de Université de Nantes.

Gammerman, A. et Z. Luo (1991). Constructing Causal Trees from a medical database, *Technical Report TR 91 002*, Dept of computer Sci. Heriot-Watt, Univ Edimburgh

Ganascia, J.G. (1987). *AGAPE et CHARADE : deux techniques d'apprentissage symbolique appliquées à la construction de bases de connaissances*, Thèse d'Etat, Université de Paris Sud

Ganascia, J. G. (1991). *CHARADE : Apprentissages de bases de connaissances dans "Induction symbolique - numérique à partir de données*, Toulouse *:* Cépaduès Éditions.

Goldberg ,D. (1994). *Algorithmes génétiques*, Addison-Wesley France

Goodman, R.M. et P. Smyth (1989). The induction of probabilistic rule set. The ITRULE algorithm, *Proceedings of sixth international conference on machine learning*, 129-132

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

Gras, R. et A. Larher (1992). L'implication statistique, une nouvelle méthode d'analyse de données, *Mathématique, Informatique et Sciences Humaines*, E.H.E.S.S. Paris, 120, 5-31

Gras, R., H. Briand and P. Peter (1996a). Structuration sets with implication intensity, *Proceedings of the International Conference on Ordinal and Symbolic Data Analysis* - OSDA 95, E. Diday, Y. Chevallier, O. Opitz (Eds.), Paris : Springer, 147-156

Gras, R. et H. Ratsimaba-Rajohn (1996b). Analyse non symétrique de données par l'implication statistique. *RAIRO-Recherche Opérationnelle, 30-3,* AFCET, Paris, 217-232

Gras, R., S. Ag Almouloud, M. Bailleul, A. Larher, M. Polo, H. Ratsimba-Rajohn et A. Totohasina (1996 c). *L'implication Statistique*, Collection Associée à Recherches en Didactique des Mathématiques, Grenoble : La Pensée Sauvage.

Gras, R. (1997a). Nœuds et niveaux significatifs en Analyse Statistique Implicative, Prépublication IRMAR, 97-32, 1-11 (3)

Gras, R., H. Briand, P. Peter and J. Philippé (1997b). Implicative statistical analysis, *Proceedings of International Congress I.F.C.S., 96,* Kobé Tokyo: Springer-Verlag, 412-419

Gras, R., E. Diday, P. Kuntz and R. Couturier (2001a). Variables sur intervalles et variables-intervalles en analyse statistique implicative, *Actes du 8ème Congrès de la Société Francophone de Classification,* Université des Antilles-Guyane, 166-173

Gras, R., P. Kuntz, R. Couturier et F. Guillet (2001b). Une version entropique de l'intensité d'implication pour les corpus volumineux. *Extraction des Connaissances et Apprentissage (ECA*), vol. 1, n° 1-2, 69-80. Hermès Science Publication.

Gras, R., P. Kuntz et H. Briand (2001c). Les fondements de l'analyse statistique implicative et quelques prolongements pour la fouille de données, *Mathématiques et Sciences Humaines*, n° 154-155, 9-29

Gras, R(égis), F. Guillet, R(obin) Gras et J. Philippé (2002). Réduction des colonnes d'un tableau de données par quasi-équivalence entre variables, *Extraction des connaissances et apprentissage*, Paris : Hermès, Volume 1, n°4/2001, 197-202.

Gras, R., P. Kuntz et H. Briand (2003). Hiérarchie orientée de règles généralisées en analyse implicative, *Extraction des Connaissances et apprentissage*, Paris : Hermès, 145-157.

Gras, R., R. Couturier, J. Blanchard, H. Briand, P. Kuntz et P. Peter (2004a). Quelques critères pour une mesure de qualité de règles d'association. Un exemple : l'implication statistique, *Mesures de qualité pour la fouille de données*, RNTI-E-1 Toulouse *:* Cépaduès Éditions, 3-32.

Gras, R., P. Kuntz et J.C. Régnier (2004b). Significativité des niveaux d'une hiérarchie orientée en analyse statistique implicative, *Classification et fouille de données*, M. Chavent et M. Langlais (Eds), *RNTI-C-1*, Toulouse *:* Cépaduès Éditions, 39-50.

Gras, R. et P. Kuntz (2005). Discovering R-rules with a directed hierarchy, *Soft Computing, A Fusion of Foundations, Methodologies and Applications*, Volume 1, Springer Verlag, 46-58..

Gras, R., J. David, J.C. Régnier et F. Guillet (2006). Typicalité et contribution des sujets et des variables supplémentaires en Analyse Statistique Implicative. *Extraction des Connaissances* (EGC'06), Volume2, Toulouse *:* Cépaduès Éditions, 359-370,

Gras, R., J. David, F. Guillet et H. Briand (2007a). Stabilité en A.S.I. de l'intensité d'implication et comparaisons avec d'autres indices de qualité de règles d'association, *Proceedings atelier « Qualité des données et des connaissances »*, EGC 07, Namur

Gras, R., P. Kuntz et E. Suzuki (2007b). Une règle d'exception en Analyse Statistique Implicative, *Extraction des Connaissances* (EGC'07),Volume1, RNTI-E-9 Toulouse *:* Cépaduès Éditions 87-98.

Gras, R. et P. Kuntz P. (2007c). Reduction of Redundant Rules in Statistical Implicative Analysis. P. Brito, P. Bertrand, G. Cucumel, E. de Carvalho, (Eds) *Selected Contributions in data Analysis and Classification*, Springer, 367-376

Havrda, J.H. and F. Charvàt (1967). Quantification Methods of Classification Processes, *Concepts of Structural Entropy*, *Kybernetica*, *3*, 30-37

Hempel, C. G. (1945). Studies in the Logic of Confirmation, *Mind* 54, 1-26

Hipp, J., U. Guntzer and J. Nakhaeizadeh (2000). Mining association rules: Deriving a superior algorithm by analyzing today's approach , *Proc. of 4th Eur. Conf. on Principles of Data Mining and Knowledge Discovery*, Lect. N. in Art. Int. 1910, 160-168.

Horschka, P. et W. Klögsen (1991). A support system for interpreting statistical data. *Knowledge Discovery in Databases*, AAAI Press, 325-345,.

Jaroszewicz, S. and D. Simovici (2001). *A general Measure of Rule Interestingness* Berlin Heidelberg : Springer-Verlag, 253-265

Kendall, M. G. and A. Stuart (1991). *Kendall's advanced theory of statistics*. (Vol. 2) London : Edward Arnold.

Kuntz, P., R. Gras and J. Blanchard (2002). Discovering Extended Rules with Implicative Hierarchies, *Conference on the new frontiers of statistical data mining and knowledge discovery*, Knoxville, Tennessee

Lagrange, J. B. (1998). Analyse implicative d'un ensemble de variables numériques ; application au traitement d'un questionnaire à réponses modales ordonnées, *Revue de Statistique Appliquée*, Paris, 71-93

Lahanier-Reuter, D. (1998). *Étude de conceptions du hasard : approche épistémologique, didactique et expérimentale en milieu universitaire*, Thèse de doctorat de l'Université de Rennes 1.

Lallich S, Lenca P. et Vaillant B. (2005) Variations autour de l'intensité d'implication, *Actes ASI 03*, Université de Palerme.

Larher, A. (1991). *Implication statistique et applications à l'analyse de démarches de preuve mathématique*, Thèse de doctorat de l'Université de Rennes 1.

Lehn, R. (2000). *Un système interactif de visualisation et de fouille de règles pour l'extraction de connaissances dans une base de données*. Thèse de doctorat de l'Université de Nantes.

Lebart, L., M. Piron et A. Morineau (2006). *Statistique exploratoire multidimensionnelle*. (4ème édition), Paris : Science sup, Dunod.

Lent, B., A.N. Swami and J. Widow (1997). Clustering association rules. *Proc. of the 13th Int. Conf. on Data Engineering*, 220-231.

Lenca, P., P. Meyer, P. Vaillant, P. Picouet et S. Lallich (2004). Évaluation et analyse multi-critères de qualité des règles d'association, *Mesures de qualité pour la fouille de données*, RNTI-E-1, Toulouse *:* Cépaduès Éditions, 219-246.

Lerman, I.-C. (1981). *Classification et analyse ordinale des données*. Paris : Dunod,.

Lerman, I.-C., R. Gras. et H. Rostam (1981). Élaboration et évaluation d'un indice d'implication pour des données binaires, I et II, *Mathématiques et Sciences Humaines* ,, n° 74,, 5-35 et n° 75, 5-47

Lerman, I.C. et J. Azé (2004). Indice probabiliste discriminant (de vraisemblance du lien) d'une règle d'association en cas de « très grosses » données, *Mesures de qualité pour la fouille de données*, RNTI-E-1, Toulouse *:* Cépaduès Éditions, 69-94.

Loevinger, J., (1947). A systematic approach to the construction and evaluation of tests of abilities, *Psychological Monographs, 61*, n° 4.

Lotfi, A. et L.A**.** Zadeh (2001). From computing with numbers to computing with words from manipulation of measurements to manipulation of perception, in *Proceedings*

"Human and machine perception" (Thinking, deciding and acting), V. Cantoni, V. Di Gesù, A. Setti e D. Tegolo Eds, Kluwer Academic, New York.

Matheus, C.J., G. Piatetsky-Shapiro and D. Mc Neill (1996). Selecting and reporting what is interesting, *In Advances in Knowledge Discovery and Data Mining*, AAAI press/MIT Press, 495-515

Pearl, J. (1988). *Probabilistic Reasoning in intelligent systems*, San Mateo, CA, Morgan Kaufmann.

Pearl, J. (1995). Causal diagrams for empirical research, *Biometrika*, 82, 4, Great Britain, 669-710.

Polo-Capra, M. (1996). *Le repère cartésien dans les systèmes scolaires français et italien : étude didactique et application de méthodes d'analyse statistiques multidimensionnelles*, Thèse de doctorat de l'Université de Rennes 1.

Ralambondrainy, H. (1991). Apprentissage dans le contexte d'un schéma de base de données in *Induction symbolique-numérique à partir de données*, Toulouse *:* Cépaduès Éditions

Ratsimba-Rajohn, H. (1992). *Contribution à l'étude de la hiérarchie implicative. Application à l'analyse de la gestion didactique des phénomènes d'ostension et de contradiction*, Thèse de doctorat de l'Université de Rennes 1.

Régnier, J.C. et R. Gras (2005). Statistique de rangs et analyse statistique implicative, *Revue de Statistique Appliquée*, LIII, 5-38

Ritschard, G., D. A. Zighed et N. Nicoloyannis (2000). Maximiser l'association par agrégation dans un tableau croisé, R. Gras et Bailleul M. (Eds) *Journées sur la Fouille dans les données par la méthode d'analyse statistique implicative*, , 219-233.

Roubine, E. (1970). Introduction à la théorie de la communication, (Tome III) *Théorie de l'information*, Paris : Masson,

Saporta, G. (2006). *Probabilités, Analyse de Données et statistique*, Paris : Ed. Technip,

Schektman, Y., J. Trejos et M. Troupe (1992). Un générateur de règles floues à partir de bases de données volumineuse, *Actes des 3èmes Journées "Symboliques-Numériques"*, mai 1992, Paris.

Sebag M. et Schoenauer (1991). Un réseau de règles d'apprentissage, *Induction symbolique-numérique à partir de données*, Toulouse *:* Cépaduès Éditions.

Shannon, C.E. et W. Weaver (1949). *The mathematical theory of communication*, Univ. of Illinois Press.

Simon, A., (2000*). Outils classificatoires par objets pour l'extraction de connaissances dans des bases de données*, Thèse de doctorat de l'Université de Nancy.

Spagnolo, F. et R. Gras (2004). A new approach in Zadeh's classification : fuzzy implication through statistic implication, *NAFIPS 2004, 23nd Conference of the North American Fuzzy Information Processing Society, Banff*, AB Canada, 27-30.

Suzuki, E. and Y. Kodratoff (1999). Discovery of surprising exception rules based on intensity of implication. *Principles of data mining and knowledge discovery science.* Springer, 184-195.

Suzuki E. and J. Zytkow (2005). Unified algorithm for undirected discovery of exception rules, *Int. J. of Intelligent Systems*, vol. 20, Wiley, 673-694

Totohasina, A. (1992) *Méthode implicative en analyse de données et application à l'analyse de conceptions d'étudiants sur la notion de probabilité conditionnelle*, Thèse de doctorat de l'Université de Rennes I.

Vergnaud, G., (1994). La théorie des champs conceptuels, *Recherches en Didactique des Mathématiques*, 10/2-3, Grenoble : La Pensée Sauvage, 133-170.

Zadeh, L.A. (1979). A Theory of Approximate Reasoning, J. Hayes, D. Michie, and L.I. Mikulich (Eds)., *Machine Intelligence 9*, New York: Halstead Press, 149-194.

Zadeh, L.A. (1997). Toward a Theory of Fuzzy Information Granulation and its Centrality in Human Reasoning and Fuzzy Logic, *Fuzzy Sets and Systems 90*, 111-127.

Zighed, D. A. et R. Rakotomalala (2000). *Graphes d'induction: apprentissage et data mining*. Paris: Hermes Science Publications.

Summary

This first part, structured in 9 chapters, consists of a general presentation of the Statistical Implicative Analysis (SIA). The authors are both defining the concepts and theorems of this theoretical approach, and also the methodological and epistemological foundations. In particular, the reader will find the definitions of: the relation of quasi-implication, the implication index, the intensity of implication, the index of propensity, the tree of implication, the directed hierarchy of implications... In addition, each concept is illustrated through an example.