

---

# Performance d'une heuristique d'agrégation optimale bidimensionnelle

**Gilbert Ritschard**

*Département d'économétrie, Université de Genève  
bd du Pont-d'Arve 40, CH-1211 Genève 4  
Gilbert.Ritschard@themes.unige.ch*

---

*RÉSUMÉ. Cette communication propose une discussion et une étude par simulations de l'efficacité de l'heuristique de Ritschard et al. [RIT 01] pour la recherche simultanée des niveaux d'agrégation optimaux de deux variables catégorielles. L'heuristique consiste en une procédure ascendante où l'on procède successivement au regroupement de deux catégories qui maximise l'amélioration du critère choisi. L'heuristique est comparée à l'exploration exhaustive de toutes les possibilités de regroupement. Les simulations réalisées font apparaître que le gain considérable en temps de calcul peut se traduire par un coût relativement important en termes de pourcentage d'optima globaux manqués mais très limité en termes d'écart par rapport à l'optimum global.*

*ABSTRACT. This paper discusses the efficiency of Ritschard et al.'s heuristic [RIT 01] for simultaneously determining the optimal aggregation of two categorical variables. The heuristic is a bottom-up algorithm that proceeds successively to the grouping of two categories that maximizes the increase in the selected criteria. The heuristic is compared to the exhaustive scanning of all possible groupings. The simulations reported show that the dramatic reduction in required computation time has a relatively high cost in terms of the percentage of missed optima. They show, nevertheless, that the missed solutions are very close to the global optimum.*

*MOTS-CLÉS : table de contingence, agrégation, association, discrétisation*

*KEYWORDS: crosstable, aggregation, association, discretization*

---

## 1. Introduction

Quelle relation peut-on établir entre le type d'articles achetés par un client et son âge ? Voilà une question courante à laquelle les moteurs d'extraction de connaissances savent donner des réponses adéquates sous forme de règles simples lorsque le nombre de types d'articles répertoriés et le nombre de classes d'âges ne sont pas excessifs. Cependant, si le nombre d'articles et/ou de classes est grand, il s'agit, pour obtenir des règles simples, de procéder à des regroupements de catégories.

Pour un attribut discret comme l'« article acheté », la démarche usuelle de regroupement conduit à choisir un niveau dans une hiérarchie généralement préétablie de généralisations du concept (voir par exemple [HAN 01] p. 151). Ceci se fait en général avant de soumettre les données au processus d'apprentissage. Pour les variables quantitatives continues comme l'âge, la démarche revient à discrétiser la variable, c'est-à-dire à partitionner les valeurs admissibles en classes de valeurs pertinentes. Cette discrétisation se fait a priori avec des techniques non contextuelles, mais peut aussi, comme par exemple avec Sipina [ZIG 00], être intégrée au processus de recherche des règles en exploitant des techniques contextuelles.

La catégorisation et la discrétisation ont, depuis les travaux de W.D. Fisher [FIS 58], été abondamment étudiés dans le cas unidimensionnel, c'est-à-dire dans le cas où chaque variable est traitée séparément. On trouve par exemple une revue de ces méthodes dans [ZIG 99].

Lorsque, comme dans notre exemple, il s'agit de mettre en évidence le lien entre deux attributs, les regroupements faits sur une variable vont évidemment influencer ceux qui devront être faits sur l'autre. Il est alors naturel de songer à une procédure de regroupement simultané sur les deux attributs et l'on se trouve alors dans un contexte multidimensionnel. La littérature est ici moins riche, bien que le regroupement simultané optimal des lignes et des colonnes d'une matrice remonte aux travaux de W.D. Fisher [FIS 62] [FIS 69] qui cherche à simplifier les modèles économiques de prédiction et de décision à équations multiples. Le problème a également été étudié pour des matrices de données où il s'agit de partitionner simultanément les cas et les variables, voir par exemple [AND 73], [BOC 79] et [GOV 95]. Le cas particulier de tableaux binaires a notamment été traité par Govaert [GOV 84] [GOV 95]. En ce qui concerne plus précisément les tables de contingence, l'agrégation ou le partitionnement simultané des lignes et des colonnes a été examiné de plusieurs points de vue. Benzécri [BEN 73] aborde par exemple la question des partitions des lignes et des colonnes qui maximisent le khi-2 de Pearson. Ce même problème a été traité en particulier par Celeux et al. [CEL 88] au moyen d'une algorithmique de nuées dynamiques appliquée alternativement sur les lignes et les colonnes. On peut également mentionner les travaux de Gilula et Krieger [GIL 83] qui étudient le comportement du khi-2 de Pearson suite à la réduction de tables par agrégation, ainsi que Hirotsu [HIR 83] et Greenacre [GRE 88] qui s'intéressent à mettre en évidence des sous-tables homogènes. Breiman et al. [?] ont quant à eux étudié et implémenté dans leur procédure C&RT la dichotomisation conjointe de deux variables.

A notre connaissance le regroupement simultané de plus de deux variables n'a été abordé que dans le contexte de la discrétisation en data mining. Les contributions les plus significatives semblent ici être celles de Dougherty et al. [DOU 95] qui étudient les regroupements qui optimisent la précision prédictive de classificateurs, et, dans une optique essentiellement non contextuelle, celles de Monti et Cooper [MON 99] et Bay [BAY 01].

Pour le problème qui nous occupe où l'on cherche les regroupements les plus aptes à fournir des règles d'associations simples, c'est naturellement l'optique contextuelle qui est pertinente, l'objectif étant de déterminer à la fois le nombre de groupes et les regroupements qui maximisent un critère donné. Les travaux cités précédemment n'offrent pas de solutions satisfaisantes de ce point de vue, soit parce que, comme les travaux réalisés à la suite de ceux de Benzécri ou encore ceux de Breiman et al., ils supposent le nombre de groupes fixé a priori, soit encore parce qu'ils portent sur des objets, e.g. matrices binaires et tableaux de données, ou des critères, l'homogénéité par exemple, difficilement transposables dans notre contexte.

Nous avons alors dans [RIT 01] proposé une heuristique pour maximiser le degré d'association, heuristique qui peut d'ailleurs très bien s'étendre à l'optimisation d'autres critères tels que par exemple la minimisation de la dispersion intra-classe de variables de contrôle. L'heuristique procède par regroupements successifs. Nous nous proposons dans cette communication d'approfondir, notamment par le biais de résultats de simulations, l'étude du comportement de cette heuristique, c'est-à-dire évaluer d'une part le gain de temps (réduction de complexité) par rapport à une exploration exhaustive qui seule peut assurer l'optimum global et d'autre part le pourcentage d'optima manqués et l'écart entre la solution trouvée et l'optimum global.

La communication est organisée comme suit. La section 2 introduit le cadre formel et les notations. La section 3 introduit l'heuristique et la section 4 en discute la complexité. La section 5 rapporte et commente les résultats des simulations réalisées. Enfin, à la section 6 nous concluons avec quelques indications sur les développements et applications potentielles de l'heuristique.

## 2. Cadre formel et notation

Soit deux variables  $x$  et  $y$  prenant respectivement  $\ell$  et  $c$  états différents. Le croisement des variables donne lieu à une table de contingence  $T_{\ell \times c}$  à  $\ell$  lignes et  $c$  colonnes.

On considère les critères d'association  $\theta_{xy}$  entre  $x$  et  $y$  en tant que fonction des éléments de la table de contingence  $\theta_{xy} = \theta(T_{\ell \times c})$ . Le critère  $\theta_{xy}$  sera ainsi par exemple le  $v$  de Cramer ou le  $\tau_b$  de Kendall calculé sur la table  $T_{\ell \times c}$ .

Soit  $P_x$  une partition des états de la variable ligne  $x$  et  $P_y$  une partition des valeurs de  $y$ . Chaque couple  $(P_x, P_y)$  de partitions donne lieu à une table de contingence  $T$

différente. Le problème général envisagé est alors la recherche du couple de partitions qui maximise la mesure d'association

$$\max_{P_x, P_y} \theta(T(P_x, P_y)) . \quad (1)$$

Pour des variables ordinales, et donc en particulier pour les variables mesurables de type intervalle ou ratio, seules les partitions obtenues par regroupement de catégories adjacentes sont pertinentes. Dans ce cas, on considère le problème restreint

$$\begin{cases} \max_{P_x, P_y} \theta(T(P_x, P_y)) \\ \text{s.c. } P_x \in \mathcal{A}_x \text{ et } P_y \in \mathcal{A}_y . \end{cases} \quad (2)$$

où  $\mathcal{A}_x$  et  $\mathcal{A}_y$  désignent respectivement l'ensemble des partitions par regroupements adjacents des valeurs de  $x$  et de  $y$ . En désignant par  $\mathcal{P}_x$  et  $\mathcal{P}_y$  les ensembles non restreints de partitions, on a, pour  $c, \ell > 2$ ,  $\mathcal{A}_x \subset \mathcal{P}_x$  et  $\mathcal{A}_y \subset \mathcal{P}_y$ . Notons encore que, l'association entre variables ordinales pouvant être négative, le critère  $\theta(T(P_x, P_y))$  à maximiser est dans ce cas la valeur absolue d'une mesure d'association.

### 3. Heuristique

L'heuristique proposée procède itérativement. A chaque étape, on recherche parmi les regroupements de deux catégories ligne ou de deux catégories colonne celui qui maximise le critère d'association  $\theta(T)$  retenu. Formellement, en notant  $(P_x^k, P_y^k)$  la partition obtenue à l'étape  $k$ , il s'agit à chaque étape  $k$  de chercher la solution  $(P_x^k, P_y^k)$  du programme

$$\begin{cases} \max_{P_x, P_y} \theta(T(P_x, P_y)) \\ \text{s.c. } P_x = P_x^{(k-1)} \text{ et } P_y \in \mathcal{P}_y^{(k-1)} \\ \text{ou} \\ P_x \in \mathcal{P}_x^{(k-1)} \text{ et } P_y = P_y^{(k-1)} \end{cases} \quad (3)$$

où  $\mathcal{P}_x^{(k-1)}$  désigne l'ensemble des partitions sur la variable  $x$  obtenues par un regroupement de deux classes de la partition  $P_x^{(k-1)}$ .

Pour des variables ordinales, il convient de remplacer  $\mathcal{P}_x^{(k-1)}$  et  $\mathcal{P}_y^{(k-1)}$  par les ensembles  $\mathcal{A}_x^{(k-1)}$  et  $\mathcal{A}_y^{(k-1)}$  de partitions obtenues par le regroupement de deux éléments adjacents.

En partant de  $T^0 = T_{\ell \times c}$  le tableau initial associé aux catégories les plus fines des variables  $x$  et  $y$ , l'heuristique consiste alors à rechercher successivement les tableaux  $T^k, k = 1, 2, \dots$  définis par les partitions solution de (3). La procédure se poursuit tant que  $\theta(T^k) \geq \theta(T^{(k-1)})$  et est arrêtée dès que cette condition n'est plus vérifiée. En d'autres termes, on procède successivement au regroupement de deux catégories

qui maximise l'accroissement du critère tant que cet accroissement optimal reste non négatif.

Le *regroupement quasi-optimal* est le couple  $(P_x^k, P_y^k)$  solution de (3) à l'étape  $k$  où

$$\theta(T^{(k+1)}) - \theta(T^k) < 0 \quad \text{et} \quad \theta(T^k) - \theta(T^{(k-1)}) \geq 0 .$$

Par convention, on fixe le critère d'association  $\theta(T)$  à zéro pour toute table ayant une seule ligne ou colonne. L'algorithme conduit ainsi à une table  $1 \times 1$  avec une seule valeur si et seulement si toutes les lignes, et donc toutes les colonnes, sont identiquement distribuées.

#### 4. Complexité

Afin de rendre compte du gain que l'heuristique permet en termes de vitesse d'exécution, on compare ici la complexité de l'heuristique à celle de l'exploration exhaustive de tous les regroupements possibles de lignes ou de colonnes, c'est-à-dire de l'ensemble des couples  $(P_x, P_y)$ . Les résultats sont des rappels de ceux donnés dans [RIT 01].

Pour le balayage exhaustif, le nombre de situations à examiner est donné par le produit du nombre de regroupements possibles des lignes par celui des colonnes, soit  $\#\mathcal{P}_x \#\mathcal{P}_y$ .

Pour le cas de variables nominales, le nombre de regroupements correspond au nombre  $B(c) = \#\mathcal{P}$  de partitions de l'ensemble des  $c$  catégories de la variable. Il peut être obtenu récursivement par la formule de Bell [BEL 38]

$$B(c) = \sum_{0 \leq k \leq c-1} \binom{c-1}{k} B(k)$$

avec  $B(0) = 1$ . Le nombre de cas à explorer est donc dans ce cas  $B(c)B(\ell)$ .

Pour des variables ordinales, et donc en particulier dans les problèmes de discrétisation, seuls les regroupements de catégories adjacentes sont pertinentes. Le nombre de cas à explorer s'en trouve dès lors réduit. Le nombre  $G(c) = \#\mathcal{A}$  de groupements différents de  $c$  catégories est

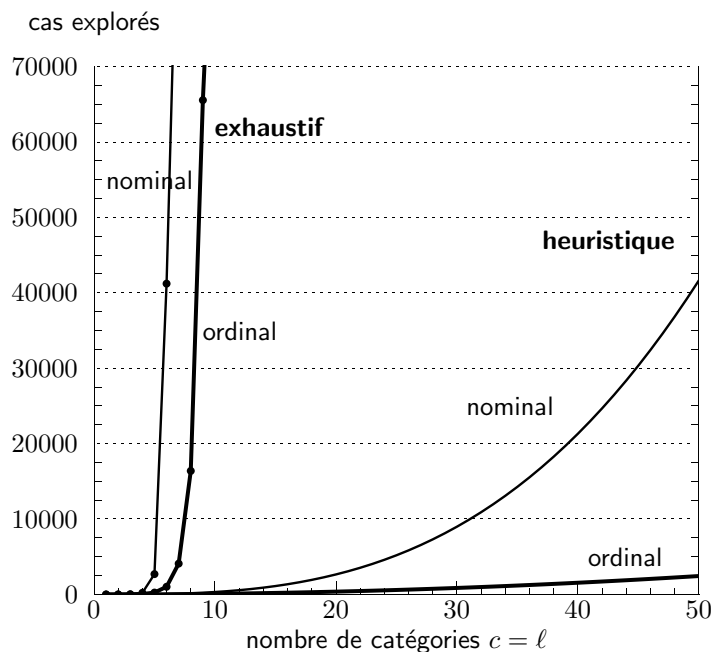
$$G(c) = \sum_{k=0}^{c-1} \binom{c-1}{k} = 2^{(c-1)}$$

et le nombre de cas à explorer dans ce cas ordinal devient  $G(c)G(\ell)$ .

Pour ce qui est maintenant de l'heuristique, la borne supérieure du nombre de cas explorés par l'algorithme est

$$\frac{\ell(\ell^2 - 1) + c(c^2 - 1)}{6} \quad \text{dans le cas nominal}$$

$$\frac{\ell(\ell - 1) + c(c - 1)}{2} \quad \text{dans le cas ordinal .}$$



**Figure 1.** Complexité selon la taille du tableau carré : pour l'heuristique les valeurs indiquées sont des bornes supérieures

Les tableaux 1 et 2 montrent comment le gain en efficacité relative<sup>1</sup> augmente avec le nombre de catégories. Les valeurs données dans ce tableau ainsi que celles reportées dans la figure 1 concernent des tableaux carrés. La complexité de l'heuristique semble, dans la figure 1, croître exponentiellement avec la taille des tables. Il convient cependant de ne pas perdre de vue que le nombre de cas effectivement explorés par l'heuristique est en règle générale bien inférieur à la borne supérieure représentée.

1. Dans les tableaux, la colonne « % réduction » donne le taux de réduction du nombre de cas explorés que permet l'heuristique par rapport au cas exhaustif.

**Tableau 1.** *Nombre de configurations explorées, cas nominal*

$c = \ell$	exhaustive	heuristique	% réduction
2	4	2	50.00%
3	25	8	68.00%
4	225	20	91.11%
5	2704	40	98.52%
6	41209	70	99.83%
7	769129	112	99.99%
8	17139600	168	100.00%
9	447195609	240	100.00%
10	$1.345 \cdot 10^{10}$	330	100.00%
20	$2.675 \cdot 10^{27}$	2660	100.00%
50	$3.449 \cdot 10^{94}$	41650	100.00%
100	$2.264 \cdot 10^{231}$	333300	100.00%

**Tableau 2.** *Nombre de configurations explorées, cas ordinal*

$c = \ell$	exhaustive	heuristique	% réduction
2	4	2	50.00%
3	16	6	62.50%
4	64	12	81.25%
5	256	20	92.19%
6	1024	30	97.07%
7	4096	42	98.97%
8	16384	56	99.66%
9	65536	72	99.89%
10	262144	90	99.97%
20	$2.749 \cdot 10^{11}$	380	100.00%
50	$3.169 \cdot 10^{29}$	2450	100.00%
100	$4.017 \cdot 10^{59}$	9900	100.00%

## 5. Optima globaux manqués

Pour juger de la performance en termes de résultats, nous avons procédé à plusieurs séries de simulations pour voir comment se situe le quasi-optimum donné par l'heuristique par rapport à l'optimum global.

Plusieurs critères d'association ont été examinés. Nous reportons ici les résultats pour le  $t$  de Tschuprow, le  $\tau$  de Goodman-Kruskal, et le  $\tau_b$  de Kendall. Parmi toutes

les mesures que nous avons considérées,<sup>2</sup> le  $t$  de Tschuprow a été retenu car il obtient les plus mauvais scores tant en termes d'optima manqués que d'écart par rapport à l'optimum global. Le  $\tau$  de Goodman-Kruskal a été choisi comme représentant des mesures asymétriques de type PRE (proportion de réduction de l'erreur de prédiction) et le  $\tau_b$  de Kendall comme représentant des mesures ordinales.

La comparaison a été réalisée pour des tableaux carrés de taille 4, 5 et 6. Au delà de 6, la recherche des optima globaux devient irréalisable dans un temps raisonnable.

Pour le  $t$  de Tschuprow et le  $\tau$  de Goodman-Kruskal, nous reportons, respectivement dans les tableaux 3 et 4, des résultats pour le cas nominal et le cas ordinal. Pour le  $\tau_b$  de Kendall au tableau 5, seul le cas ordinal est évidemment considéré.

Pour chaque mesure, taille et type de variable, 200 tableaux ont été générés aléatoirement. Chaque tableau a été obtenu en répartissant une population de 10000 cas au hasard dans les  $c \times \ell$  cellules du tableau selon une distribution uniforme. Notons que cette façon de procéder diffère de celle qui a donné lieu aux résultats présentés dans [RIT 01], où l'on a utilisé des distributions uniformes conditionnelles emboîtées : un pourcentage aléatoire entre 0 et 100% des cas est alloué à la première ligne, puis un pourcentage aléatoire des cas restant à la seconde ligne et ainsi de suite jusqu'à la dernière ligne, le total de chaque ligne étant ensuite réparti de la même façon dans la ligne. La différence est que les tableaux générés ici ont évidemment tendance à être plus proches de la distribution uniforme et donnent par conséquent lieu, en règle générale, à des associations moins importantes entre variables ligne et colonne.

Les tableaux 3 à 5 donnent, pour chaque série de tables générées, le pourcentage d'optima globaux manqués, des indications (maximum, moyenne, écart type, indicateur d'asymétrie) sur la distribution des écarts non nuls d'une part, et pour l'ensemble de tous les écarts nuls et non nuls d'autre part. Les écarts relatifs dont on donne le maximum et la valeur moyenne, sont calculés par rapport à l'optimum global. Les deux dernières lignes indiquent respectivement la moyenne des valeurs initiales du critère pour les tables générées et la moyenne des optima trouvés pour ces tables. Au tableau 5, ces deux valeurs sont des moyennes de valeurs absolues.

A la lecture des tableaux 3 à 5, on constate que le pourcentage d'optima manqués par l'heuristique est assez élevé et tend à augmenter avec la taille du tableau. Ce pourcentage est plus faible pour les mesures de type PRE (le  $\tau$  de Goodman-Kruskal), probablement parce que ces mesures ne peuvent augmenter avec des regroupements sur la variable indépendante (voir [RIT 01]) et que les regroupements se font donc essentiellement sur la variable dépendante. Curieusement, le pourcentage d'optima manqués est cependant, pour les mesures de type PRE, plus élevé dans le cas ordinal que dans le cas nominal.

---

2. Il s'agit des mesures discutées dans [RIT 01], soit  $t$  Tschuprow,  $v$  de Cramer,  $\tau$  de Goodman-Kruskal,  $u$  de Theil,  $\gamma$  de Goodman-Kruskal,  $\tau_b$  de Kendall et  $d$  de Somers, les trois derniers ne s'appliquant évidemment qu'aux variables ordinales. On trouvera une discussion détaillée du comportement de ces mesures dans [OLS 95].



**Tableau 3.** *Simulations :  $t$  de Tschuprow*

Tschuprow	nominal			ordinal		
Taille	4 × 4	5 × 5	6 × 6	4 × 4	5 × 5	6 × 6
Écarts non nuls	39.5%	62.5%	74.5%	23.5%	36%	46.5%
maximum	0.073	0.074	0.077	0.077	0.063	0.108
moyenne	0.025	0.023	0.028	0.019	0.019	0.012
écart type	0.015	0.014	0.016	0.014	0.016	0.015
asymétrie	0.986	0.979	0.598	1.674	0.972	3.394
avec écarts nuls						
moyenne	0.010	0.015	0.021	0.005	0.007	0.006
écart type	0.016	0.016	0.018	0.011	0.013	0.012
asymétrie	1.677	1.062	0.615	3.168	2.211	4.457
Écarts relatifs						
maximum	0.168	0.198	0.221	0.179	0.194	0.307
moyenne	0.079	0.077	0.093	0.063	0.066	0.046
Initial moyen	0.260	0.240	0.226	0.263	0.244	0.228
Global moyen	0.340	0.316	0.303	0.301	0.275	0.250

**Tableau 4.** *Simulations :  $\tau$  de Goodman et Kruskal*

$\tau$ de G&K	nominal			ordinal		
Taille	4 × 4	5 × 5	6 × 6	4 × 4	5 × 5	6 × 6
Écarts non nuls	5%	6.5%	12%	6%	19%	32.5%
maximum	0.013	0.031	0.029	0.076	0.077	0.059
moyenne	0.007	0.010	0.008	0.025	0.016	0.013
écart type	0.004	0.009	0.009	0.021	0.016	0.012
asymétrie	-0.308	1.004	1.181	1.107	2.361	1.908
avec écarts nuls						
moyenne	0.0004	0.0007	0.0010	0.0015	0.003	0.004
écart type	0.0018	0.0033	0.004	0.008	0.009	0.009
asymétrie	5.323	6.471	5.137	6.685	5.040	3.255
Écarts relatifs						
maximum	0.142	0.296	0.318	0.420	0.518	0.401
moyenne	0.062	0.091	0.079	0.216	0.168	0.149
Initial moyen	0.074	0.060	0.048	0.073	0.060	0.050
Global moyen	0.148	0.128	0.113	0.118	0.098	0.084

Ce fort pourcentage d'optima manqués est heureusement tempéré par le faible écart qui sépare le quasi-optimum de l'optimum global. Si l'on considère les seuls écarts non nuls, ils représentent en moyenne moins de la moitié de l'écart entre la valeur initiale du critère et l'optimum global. Dans le cas d'associations plus importantes que celles générées ici avec une distribution uniforme, ce rapport devient largement plus favorable. En regardant les écarts non nuls, on constate par ailleurs que le niveau

**Tableau 5.** *Simulations :  $\tau_b$  de Kendall*

$\tau_b$ de Kendall	ordinal		
Taille	4 × 4	5 × 5	6 × 6
Ecart non nuls	19%	24.5%	32%
maximum	0.596	0.597	0.542
moyenne	0.235	0.182	0.140
écart type	0.195	0.190	0.157
asymétrie	0.076	0.598	0.652
avec écarts nuls			
moyenne	0.045	0.045	0.045
écart type	0.125	0.123	0.111
asymétrie	2.775	2.849	2.445
Ecart relatifs			
maximum	1.954	1.970	1.982
moyenne	0.355	0.259	0.074
Initial moyen (valeur absolue)	0.094	0.078	0.064
Global moyen (valeur absolue)	0.256	0.236	0.215

et la dispersion de ces écarts restent stables quand la taille du tableau augmente. Ces écarts ont évidemment tendance à être plus importants quand la mesure donne des valeurs élevées. Inversement, les écarts relatifs ont tendance à être grands quand la mesure prend des valeurs proches de zéro.

Enfin, relevons que le  $\tau_b$  de Kendall prend ses valeurs sur  $[-1, 1]$ . Les écarts peuvent donc être supérieurs à la valeur de l'optimum global lorsque le quasi-optimum et l'optimum global sont de signes contraires. Ceci explique les écarts relatifs maximaux supérieurs à un.

De façon générale, il ressort de ces simulations que l'heuristique a un coût en termes de fiabilité des résultats qui reste tout à fait raisonnable en regard de l'extraordinaire gain d'efficacité qu'elle offre.

## 6. Conclusion

L'heuristique étudiée a été introduite dans le contexte de la maximisation d'un critère d'association. Elle peut cependant sans autres être transposée à toute situation où le regroupement des catégories peut influencer un critère à optimiser. En particulier, nous envisageons de l'exploiter dans le contexte des indices de loyers ou du marché de l'art où il s'agit de déterminer une stratification optimale des types de logements ou d'art. Le critère à optimiser devient ici la minimisation de la dispersion résiduelle du prix dans les cellules du tableau. Quant aux extensions de la procédure, nous travaillons actuellement à une mise en œuvre qui permettent d'imposer des effectifs minimums dans les cellules. D'autre part, nous travaillons sur une procédure

hiérarchique descendante, où l'on procède par dichotomisations successives à partir d'un tableau totalement agrégé.

Pour terminer, mentionnons encore que la procédure discutée a été implémentée sous forme d'un logiciel pour l'environnement Windows. Ce logiciel, nommé OCA, est disponible à l'adresse <http://mephisto.unige.ch/stats>.

## 7. Bibliographie

- [AND 73] ANDERBERG M., *Cluster Analysis for Application*, Academic Press, New York, 1973.
- [BAY 01] BAY S. D., « Multivariate Discretization for Set Mining », *Knowledge and Information Systems*, vol. 3, n° 4, 2001, p. 491–512.
- [BEL 38] BELL E. T., « The iterated exponential numbers », *Ann. Math.*, vol. 39, 1938, p. 539–557.
- [BEN 73] BENZÉCRI J.-P., *Analyse des données. Tome 2 : Analyse des correspondances*, Dunod, Paris, 1973.
- [BOC 79] BOCK H., « Simultaneous Clustering of Objects and Variables », DIDAY E., LEBART L., PAGES J., TOMASSONE R., Eds., *Data Analysis and Informatics*, Amsterdam, 1979, North Holland, p. 187–203.
- [CEL 88] CELEUX G., DIDAY E., GOVAERT G., LECHEVALLIER Y., RALAMBONDRAINY H., *Classification automatique des données*, Informatique, Dunod, Paris, 1988.
- [DOU 95] DOUGHERTY J., KOHAVI R., SAHAMI M., « Supervised and unsupervised discretization of continuous features », *Proceedings of the Twelfth International Conference on Machine Learning*, 1995, p. 194–202.
- [FIS 58] FISHER W. D., « On Grouping for Maximum of Homogeneity », *Journal of the American Statistical Association*, vol. 53, 1958, p. 789–798.
- [FIS 62] FISHER W. D., « Optimal aggregation in multi-equation prediction models », *Econometrica*, vol. 30, 1962, p. 744–769.
- [FIS 69] FISHER W. D., *Clustering and Aggregation in Economics*, The John Hopkins Press, Baltimore, 1969.
- [GIL 83] GILULA Z., KRIEGER A. M., « The Decomposability and Monotonicity of Pearson's Chi-Square for Collapsed Contingency Tables with Applications », *Journal of the American Statistical Association*, vol. 78, 1983, p. 176–180.
- [GOV 84] GOVAERT G., « Classification simultanée de tableaux binaires », DIDAY E., JAMBU M., LEBART L., PAGES J., TOMASSONE R., Eds., *Data Analysis and Informatics 3*, Amsterdam, 1984, North Holland, p. 223–236.
- [GOV 95] GOVAERT G., « Simultaneous clustering of rows and columns », *Control and Cybernetics*, vol. 24, n° 4, 1995, p. 438–458.
- [GRE 88] GREENACRE M., « Clustering the Rows and Columns of a Contingency Table », *Journal of Classification*, vol. 5, 1988, p. 39–51.
- [HAN 01] HAN J., KAMBER M., *Data Mining : Concept and Techniques*, Morgan Kaufmann, San Francisco, 2001.

- [HIR 83] HIROTSU C., « Defining the Pattern of Association in Two-way Contingency Tables », *Biometrika*, vol. 70, 1983, p. 579–589.
- [MON 99] MONTI S., COOPER G. F., « A latent variable model for multivariate discretization », *The Seventh International Workshop on Artificial Intelligence and Statistics*, 1999.
- [OLS 95] OLSZAK M., RITSCHARD G., « The behaviour of nominal and ordinal partial association measures », *The Statistician*, vol. 44, n° 2, 1995, p. 195–212.
- [RIT 01] RITSCHARD G., ZIGHED D. A., NICOLOYANNIS N., « Maximisation de l'association par regroupement de lignes ou colonnes d'un tableau croisé », *Revue Mathématiques Sciences Humaines*, vol. 39, n° 154/155, 2001, p. 81–97.
- [ZIG 99] ZIGHED D. A., RABASEDA S., RAKOTOMALALA R., FESCHET F., « Discretization Methods in Supervised Learning », *Encyclopedia of Computer Science and Technology*, vol. 40, n° 25, 1999, p. 35–50.
- [ZIG 00] ZIGHED D. A., RAKOTOMALALA R., *Graphes d'induction : apprentissage et data mining*, Hermes Science Publications, Paris, 2000.