

## AVANT-PROPOS

Pascale KUNTZ<sup>1</sup>, Bruno LECLERC<sup>2</sup>

Ce numéro spécial rassemble cinq textes issus de communications présentées aux Journées francophones sur le thème « La fouille dans les données par la méthode d'analyse implicative ». Ces journées, organisées par Régis Gras et Marc Bailleul, se sont déroulées les 23 et 24 juin 2000 à l'IUFM de Caen. Nous remercions d'abord ces organisateurs d'avoir proposé à *Mathématiques et Sciences humaines* d'être le lieu de publication de cette sélection. Cette offre était particulièrement bienvenue, la revue ayant été, depuis les origines [Flament 1966], et sans doute en raison de certaines spécificités des données issues des sciences sociales, un lieu de publication privilégié pour les travaux concernant la définition, l'extraction et l'interprétation des implications significatives dans un corpus de données □ on trouvera ci-dessous un recensement des articles correspondants, avec, sans prétendre aucunement à l'exhaustivité, quelques autres publications sur ces sujets. Ce numéro permet aussi de renouer avec le thème, longtemps très présent dans nos pages, de la didactique des mathématiques (cf. par exemple, sans trop s'éloigner de notre problématique [Reuchlin 1971 ou Vinrich 1977]). Il s'inscrit aussi dans un courant très dynamique, comme en attestent les nombreux congrès actuels sur la fouille des données – en particulier en France la conférence annuelle «Extraction et gestion des connaissances» ou, par exemple, le «Séminaire Université-Industrie» sur les méthodes booléennes organisé à Paris en mai dernier par Michel Minoux.

Les communications aux Journées portaient sur la méthode d'analyse statistique implicative développée à l'origine par R. Gras [1979] □ [cf. Gras *et al.* 1996] s'inspirant de l'algorithme de la vraisemblance du lien de I.C. Lerman (voir par exemple [Lerman 1981]), et son application aux recherches en Fouille de données. Les articles qui suivent proposent donc pour la plupart de nouveaux développements et algorithmes dans ce cadre. Les bases de cette approche ont déjà trouvé une bonne place dans *Mathématiques et Sciences humaines* comme dans *Mathématiques, Informatique et Sciences humaines* [Lerman, Gras et Rostam 1981a, b, Gras et Larher 1992, Bailleul et Gras 1994].

---

<sup>1</sup> IRIN, École Polytechnique de l'Université de Nantes, la Chantrerie, B.P. 60602, 44306 Nantes cedex □, e-mail □ Pascale.Kuntz@polytechn.univ-nantes.fr.

<sup>2</sup> Centre d'Analyse et de Mathématique Sociales, EHESS, 54 boulevard Raspail 75270 Paris cedex 06, e-mail □ leclerc@ehess.fr.

Au départ, la construction des implications repose naturellement sur des considérations de nature logique et algébrique, comme dans l'analyse booléenne [Hammer et Rudeanu 1966, Flament 1966, 1976, Degenne 1972] ou l'analyse Galoisienne (ou conceptuelle), qui induit la représentation de tout treillis, et donc des treillis de Galois, par un système canonique d'implications [Barbut et Monjardet 1970, Wille 1984, Guigues et Duquenne 1986, Duquenne 1987, Luxenburger 1991, Ganter et Wille 1999, Caspard et Monjardet 2001]. Dans cette direction, de nombreuses questions restent ouvertes, comme celles du lien entre les implications « booléennes » et « latticielles ». Parallèlement, et avant même qu'il n'ait été question de « fouille de données », on a vite mesuré les difficultés algorithmiques posées par cette approche pour des données de taille importante : ceci, et la nécessité de faire le tri entre les implications significatives et des formes plus accidentelles, a conduit à l'introduction d'indices quantitatifs simples [Van Buggenhaut 1987, Guénoche 1986], ou fondés sur des modèles probabilistes élaborés comme celui de l'analyse implicative statistique ou celui, assez différent et également largement présenté dans cette revue, de l'analyse implicative bayésienne [Bernard et Charron 1996a, b, Bernard et Poitrenaud 1999]. On peut ainsi mesurer l'importance, annoncée plus haut, de *Mathématiques et Sciences humaines* dans le développement de chacun de ces courants. Nous sommes heureux de poursuivre dans cette voie avec ce numéro, et nous espérons de nouvelles contributions sur un thème offrant nombre de sujets de discussion et de questions ouvertes.

Parmi les personnes qui nous ont aidé pour la préparation de ce numéro, nous tenons à remercier les rapporteurs qui se sont joints aux membres du comité de rédaction de la revue et aux intervenants de ce numéro, et notamment Jean-Marc Bernard et Stéphane Lallich.

## BIBLIOGRAPHIE

- BAILLEUL, M., GRAS, R. , « L'implication statistique entre variables modales », *Mathématiques, Informatique et Sciences humaines* 128, 1994, p. 41-57.
- BARBUT, M., MONJARDET, B., *Ordre et Classification, Algèbre et Combinatoire*, Paris, Hachette, 1970.
- BERNARD, J.-M., CHARRON, C., « L'Analyse Implicative Bayésienne, une méthode pour l'étude des dépendances orientées. I : données binaires », *Mathématiques, Informatique et Sciences humaines* 134, 1996a, p. 5-38.
- BERNARD, J.-M., CHARRON, C., « L'Analyse Implicative Bayésienne, une méthode pour l'étude des dépendances orientées. II : Modèle logique sur un tableau de contingence », *Mathématiques, Informatique et Sciences humaines* 135, 1996b, p. 1-18.
- BERNARD, J.-M., POITRENAUD, S., « L'analyse implicative bayésienne multivariée d'un questionnaire binaire : quasi-implications et treillis de Galois simplifié », *Mathématiques, Informatique et Sciences humaines* 147, 1999, P. 25-46.

- CASPARD, N., MONJARDET, B., «The lattice of Moore families and closure operators on a finite set», [à paraître 2001 dans *Discrete Applied Mathematics*].
- DEGENNE, A., *Techniques ordinales en analyse des données statistique*, Paris, Hachette, 1972.
- DUQUENNE, V., «Contextual implications between attributes and some representation properties for finite lattices», *Beiträge zur Begriffsanalyse*, B. Ganter et al., eds., Mannheim, Wissenschaftsverlag, 1987, p. 213-240.
- FLAMENT, C., «L'analyse booléenne de questionnaires», *Mathématiques et Sciences humaines* 12, 1966, p. 3-10.
- FLAMENT, C., *L'analyse booléenne de questionnaires*, Paris, Mouton, 1976.
- GANTER, B., WILLE, R., *Formal Concept Analysis - Mathematical Foundations*, Berlin, Springer-Verlag, 1999.
- GRAS, R., Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques, Thèse d'état, Université de Rennes I, 1979.
- GRAS, R., et al., *L'implication statistique*, Grenoble, La Pensée Sauvage, 1996.
- GRAS, R., LARHER, A., «L'implication statistique, une nouvelle méthode pour l'analyse des données», *Mathématiques, Informatique et Sciences humaines* 120, 1992, p. 5-31.
- GUÉNOCHE, A., «Fonctions booléennes sur un tableau en 0/1», *Data Analysis and Informatics*, E. Diday et al. eds., Amsterdam, North-Holland, 1986, p. 443-451.
- GUÉNOCHE, A., MONJARDET, B., «Méthodes ordinales et combinatoires en analyse des données», *Mathématiques et Sciences humaines* 100, 1987, p. 5-47.
- GUIGUES, J.-L., DUQUENNE, V., «Familles minimales d'implications informatives résultant d'un tableau de données binaires», *Mathématiques et Sciences humaines* 95, 1986, p. 5-18.
- HAMMER, P.L., RUDEANU, S., *Boolean Methods in Operations Research and related areas*, Berlin, Springer-Verlag, 1966.
- LERMAN, I.C., *Classification et analyse ordinale des données*, Paris, Dunod, 1981.
- LERMAN, I.C., GRAS, R., ROSTAM, H., «Élaboration et évaluation d'un indice d'implication pour des données binaires I», *Mathématiques et Sciences humaines* 74, 1981a, p. 5-35.
- LERMAN, I.C., GRAS, R., ROSTAM, H., «Élaboration et évaluation d'un indice d'implication pour des données binaires II», *Mathématiques et Sciences humaines* 75, 1981b, p. 5-47.
- LUXENBURGER, M., «Implications partielles dans un contexte», *Mathématiques, Informatique et Sciences humaines* 113, 1991, p. 35-55.

- REUHLIN, M., «Formes vides et contenus informes en recherche pédagogique», *Mathématiques et Sciences humaines* 36, 1971, p. 59-77.
- Van BUGGENHAUT, J., «Questionnaires booléens et schémas d'implications et degrés de cohésion», *Mathématiques et Sciences humaines* 98, 1987, p. 9-20.
- VINRICH, G., «Dépendances didactiques», *Mathématiques et Sciences humaines* 57, 1977, p. 43-58.
- WILLE, R., «Sur la fusion des contextes individuels», *Mathématiques et Sciences humaines* 85, 1984, p. 55-71.