# Une nouvelle méthode divisive en classification non supervisée pour des données symboliques intervalles

Nathanaël Kasoro*, André Hardy**

*Université de Kinshasa
Département de Mathématique et d'Informatique
B.P. 190, Kinshasa, République Démocratique du Congo
kasoro.mulenda@yahoo.fr
**Université de Namur
Unité de Statistique - Département de Mathématique
8 Rempart de la Vierge - B - 5000 Namur - Belgique
andre.hardy@fundp.ac.be

**Résumé.** Dans cet article nous présentons une nouvelle méthode de classification non supervisée pour des données symboliques intervalles. Il s'agit de l'extension d'une méthode de classification non supervisée classique à des données intervalles. La méthode classique suppose que les points observés sont la réalisation d'un processus de Poisson homogène dans $k$ domaines convexes disjoints de $R^p$. La première partie de la nouvelle méthode est une procédure monothétique divisive. La règle de coupure est basée sur une extension à des données intervalles du critère de classification des Hypervolumes. L'étape d'élagage utilise un test statistique basé sur le processus de Poisson homogène. Le résultat est un arbre de décision. La seconde partie de la méthode consiste en une étape de recollement, qui permet, dans certains cas, d'améliorer la classification obtenue à la fin de la première partie de l'algorithme. La méthode est évaluée sur un ensemble de données réelles.

## 1 Introduction

Le but de la classification non supervisée est de décomposer un groupe d'objets, sur lesquels on mesure un ensemble de variables, en un nombre relativement restreint de sous-groupes d'objets semblables. De nombreuses méthodes de classification ont été publiées dans la littérature scientifique. La plupart d'entre elles utilisent un critère de classification basé sur une mesure de dissimilarité. Pour éviter ce choix (bien souvent arbitraire) d'une dissimilarité nous utilisons un modèle statistique pour la classification basé sur le processus de Poisson homogène (Hardy (1983)). De ce modèle est issue la méthode de classification des Hypervolumes (Hardy (1983)). Pirçon (2004) a développé une nouvelle méthode divisive de classification basée sur le critère de classification des Hypervolumes. Notre objectif est d'étendre cette méthode à des données intervalles. Une variable $Y$ dont le domaine d'observation est $\mathcal{Y}$ est appelée à valeurs d'ensemble si $\forall x_i \in E, \quad Y : E \to \mathcal{B} : x_i \longmapsto Y(x_i)$ où $\mathcal{B} = \mathcal{P}(\mathcal{Y}) = \{U \neq \emptyset \mid U \subseteq \mathcal{Y}\}$.