

# Affectation pondérée sur des données de type intervalle

Chérif Mballo<sup>\*,\*\*</sup> & Edwin Diday<sup>\*\*</sup>

<sup>\*</sup> ESIEA, Pôle E. C. D, Laval France

<sup>\*\*</sup> LISE-CEREMADE, Université Paris Dauphine, 75775 Paris cedex 16 France  
mballo@esiea-ouest.fr; diday@ceremade.dauphine.fr

**Résumé.** On s'intéresse à la construction d'arbres de décision sur des données symboliques de type intervalle en utilisant le critère de découpage binaire de Kolmogorov-Smirnov. Nous proposons une approche permettant d'affecter un individu à la fois aux deux nœuds fils générés par le partitionnement d'un nœud non terminal. Le but de cette méthode est de prendre en compte le positionnement de la donnée à classer par rapport à la donnée seuil de coupure.

## 1 Introduction

Désignons par  $\mathfrak{I}$  l'ensemble des intervalles fermés bornés de  $\mathfrak{R}$  (ensemble des nombres réels) :  $x \in \mathfrak{I}$ , on note  $x = [i(x), s(x)]$ . Indiquons par  $\Omega$  un ensemble d'individus ou objets. Une variable  $X$  est de type intervalle (Bock et al., 2000) si  $X(w) = [\alpha, \beta]$   $\forall w \in \Omega$ , où  $\alpha, \beta \in \mathfrak{R}$  et  $\alpha \leq \beta$ . Différentes méthodes de construction d'arbres de décision sur des données symboliques de type intervalle ont été proposées ((Limam, 2005) ; (Périnel, 1996)) en utilisant le critère de Gini et le likelihood. Nous privilégions ici le critère de découpage binaire de Kolmogorov-Smirnov noté KS dans la suite. Ce critère a été introduit par (Friedman, 1977) pour une partition binaire à expliquer sur des données continues. Il a été étendu aux données qualitatives classiques par (Asseraf, 1998). Nous avons examiné son adaptation aux données symboliques de type intervalle ((Mballo et al., 2004), (Mballo et Diday, 2005 c)) mais dans cette approche, un individu est affecté entièrement à un nœud de l'arbre de décision. Nous proposons dans ce papier une approche permettant d'affecter un individu à la fois aux deux nœuds fils générés par le découpage d'un nœud non terminal. Nous terminerons par un exemple pour illustrer cette approche.

## 2 Présentation du critère de Kolmogorov-Smirnov

Supposons que l'ensemble  $\Omega$  défini précédemment est une population théorique de  $n$  objets destinés à être classés par un arbre de décision. Ces objets sont décrits par  $(p+1)$  variables :  $p$  variables de type intervalle  $X_1, X_2, \dots, X_p$  (variables explicatives) et une