

Analyse des Proximités et programmes de Codage Multidimensionnel

Gérard d'Aubigny

LABSAD,
Université Pierre Mendès-France,
Grenoble, France

1 Introduction

L'objectif poursuivi par ce texte est de fournir des points de repère méthodologiques et logiciels aux utilisateurs désireux d'analyser des tableaux de données de proximités. On dispose alors d'un ou de plusieurs tableaux de données brutes (\mathbf{R}^i , $i = 1, 2, \dots, n$) dont le terme général r_{jk}^i (r comme raw data) fournit une mesure de proximité entre les objets j et k ($j, k = 1, 2, \dots, p$), relevée dans la situation i . Les objets sont des stimuli en psychologie, des biens en économie, des espèces ou des repères en biologie, des atomes en chimie, des sites, des zones, des lignes en géographie, des sujets en sociologie, etc. Les situations correspondent aux unités statistiques d'un protocole expérimental ou d'une enquête, à des temps ou des lieux d'observation, etc. Nous parlons dans la suite de sujets pour les éléments i et d'objets pour les éléments j .

Parmi les méthodes d'analyse des structures de proximité, on s'intéresse à la classe de modèles qui se proposent de représenter les coefficients de proximité entre objets sous la forme de distances entre points d'un espace affine. On obtient ainsi un graphique dont la consultation, plus attrayante que la lecture d'un tableau de chiffres, permet souvent de dégager des constatations qui seraient passées inaperçues sans représentation, des modes de structuration de l'information et donc des stratégies de lecture du (ou des) tableau(x) de données étudié(s).

Historiquement, les concepts d'axe et de dimension ont joué un grand rôle dans la recherche de structures. Ils apparaissent même à beaucoup d'utilisateurs néophytes comme l'unique finalité de l'emploi de méthodes d'analyse multivariée, l'unique type de régularité à rechercher dans la lecture des graphiques. Ceci explique l'adoption du terme de *Codage Multidimensionnel* pour désigner la classe de méthodes dont l'ambition est de construire une représentation des données dans un cadre géométrique caractérisé par la construction de nuages de points dans des espaces affines de dimension aussi faible que possible et dépendant des données, afin de visualiser les positions relatives d'un ensemble fini d'objets. On parle aussi parfois de *Positionnement Multidimensionnel*, cf. Escoufier (1980), ou encore d'*Echelonnement Multidimensionnel*, cf. Tournois et Dickes (1992)), comme traduction des termes anglais de *Multidimensional Scaling* utilisé en psychométrie et d'*Ordination Method* préféré en biométrie.