

AVANT-PROPOS

Du fait de son large domaine d'application, l'Analyse factorielle Multiple (AFM) est maintenant largement utilisée. Le présent livre est un exposé complet de la méthode : il rassemble les aspects théoriques et méthodologiques, des exemples d'application et la mise en œuvre logicielle via un package R (**FactoMineR**).

De même que l'analyse en composantes principales (ACP) ou l'analyse des correspondances multiples (ACM), l'AFM s'applique à des tableaux structurés ainsi : pour un ensemble d'individus (un individu = une ligne), on dispose des valeurs de chacun pour un ensemble de variables (une variable = une colonne). La spécificité de l'AFM tient à la prise en compte, au sein des variables actives, d'une structure en groupes définie par l'utilisateur. De telles données sont désignées par « tableau individus \times variables structurées en groupes ».

Ce format de données est fréquent. En premier lieu parce qu'il correspond bien à la démarche de l'utilisateur lorsqu'il conçoit un recueil de données. Ainsi, le rédacteur d'un questionnaire d'opinion procède par thèmes, qu'il décline chacun selon plusieurs questions (les questions seront les variables et les thèmes les groupes de variables). Cette structure du questionnaire doit bien sûr être présente lors de l'analyse des résultats. En second lieu, parce l'utilisateur souhaite souvent rapprocher des données recueillies sur les mêmes individus statistiques mais dans des contextes (par exemple géographiques et/ou temporels) différents. Ainsi, dans l'industrie agro-alimentaire, on dispose souvent, pour un même ensemble de produits, de profils sensoriels issus de dégustations réalisées dans plusieurs pays et/ou par des (types de) dégustateurs différents. Il est nécessaire d'analyser simultanément ces ensembles de données tout en préservant leur individualité, lors de l'analyse d'abord et lors de l'interprétation ensuite.

L'expérience issue de nombreux contacts avec des utilisateurs variés montre que les tableaux multiples constituent en fait le format standard des données auxquelles on est confronté aujourd'hui lorsque l'on applique des méthodes factorielles. A cette complexité de structure (en groupes de variables), s'ajoute une complexité de nature, les variables pouvant être quantitatives et/ou qualitatives. Il est donc nécessaire de disposer d'une méthodologie d'analyse de tableaux individus \times variables, dans lesquels les variables sont structurées en groupes quantitatifs, qualitatifs ou mixtes. Tel est précisément le champ d'application de l'AFM.

L'AFM est le résultat d'un travail conjoint de Brigitte Escofier et de Jérôme Pagès au début des années 80. Cette méthode est maintenant bien établie si l'on en juge par sa disponibilité logicielle croissante. Citons, sans prétendre à l'exhaustivité, quelques logiciels incluant une procédure d'AFM : SPAD, **FactoMiner** (Package R) **ade4** (Package R), Uniwin (Statgraphics), XLStat.

La disponibilité de la méthode étant acquise, la fréquence du format des données justifiant sa mise en œuvre s'imposant d'elle-même, il reste encore à aider l'utilisateur à appréhender ses données dans leur complexité. Pour cela, une question est centrale : que signifie précisément « prendre en compte la structure en groupes de variables dans une analyse d'ensemble » ? Autrement dit, pourquoi ne

pas mettre en œuvre une analyse factorielle usuelle, par exemple une analyse en composantes principales, et tenir compte de la structure en groupes de variables uniquement dans l'interprétation. En étant un peu réducteur, on pourrait dire que ce livre ne répond qu'à cette seule question.

Les deux premiers chapitres reprennent les méthodes de base de l'analyse factorielle d'un tableau individus \times variables, ACP et ACM.

Le chapitre 3 traite de l'analyse factorielle simultanée de variables quantitatives et qualitatives, sans distinction de groupes. La méthode décrite, dite AFDM (Analyse Factorielle de Données Mixtes), est peu connue; elle est l'occasion d'introduire les éléments techniques permettant de prendre en compte les deux types de variables au sein d'une analyse unique.

Les chapitres suivants, numérotés de 4 à 10, décrivent l'analyse factorielle multiple. Les quatre premiers abordent successivement les points clés de l'AFM dans le cadre de variables quantitatives. Un chapitre est dédié aux données qualitatives et mixtes. Enfin deux chapitres comparent chacun l'AFM à une méthode de référence pour des questions spécifiques : la méthode Statis et l'analyse procustéenne.

Le chapitre 11 présente une extension naturelle de l'AFM : l'AFM hiérarchique (AFMH). Dans cette méthode, les variables ne sont pas structurées par une simple partition, mais par une hiérarchie ou, si l'on préfère, une suite de partitions emboîtées. Un exemple typique de ces données est fourni par les enquêtes dont le questionnaire est structuré en thèmes et sous-thèmes.

Enfin, le chapitre 12 présente, sous la forme de deux fiches, quelques éléments de calcul matriciel et d'espaces métriques utilisés dans ce livre.

Au terme de cet ouvrage il m'est agréable de remercier Sophie Puyo, ingénieure statisticienne, qui a assuré l'essentiel de la mise en forme de ce livre. Première lectrice de ce travail, elle a été aussi à l'origine de bon nombre d'améliorations. Ces remerciements s'adressent aussi à Magalie Houée-Bigot, ingénieure statisticienne, qui a pris le relais de Sophie après l'intervention des relecteurs. Je remercie aussi tout particulièrement Eric Matzner-Løber pour l'accueil qu'il a su réserver à ce livre et les échanges que cela a occasionnés. Il est juste enfin de remercier Annie, mon épouse, qui éclaire ma vie et donc, indirectement, ce livre.

Les données utilisées dans ce livre sont disponibles sur le site du laboratoire de mathématiques appliquées d'Agrocampus Ovest.

Les chapitres 3, 8, 9 et 10 reprennent, en les adaptant au format d'un livre, des travaux initialement publiés dans la revue de statistique appliquée (dont la publication s'est arrêtée en 2006). C'est là une excellente occasion de remercier Pierre Cazes, directeur de cette revue, d'abord pour l'excellent accueil qu'il fit à ces travaux et ensuite pour son encouragement à les reprendre dans un livre.

Les calligraphies sont dues au talent de Richard Delécolle.