Éléments de validité en Analyse Factorielle Multiple

Marine Cadoret, Sébastien Lê et Jérôme Pagès

Résumé. L'AFM propose un certain nombre d'indicateurs d'aide à l'interprétation. A tout indicateur, il est précieux de pouvoir associer une probabilité critique. Pour cela, on peut mettre en oeuvre des épreuves de permutation. L'exposé précise quelques indicateurs clés sur lesquels il est utile de disposer d'éléments de validité et quelques méthodologies à mettre en oeuvre pour leur associer une probabilité critique.

1 Introduction et problématique

L'Analyse Factorielle Multiple (AFM, [1]) s'applique à des données dans lesquelles un même ensemble d'individus est décrit par plusieurs groupes de variables. Dans le cas où les variables sont quantitatives, le coeur de l'AFM est une analyse en composantes principales dans laquelle les variables sont pondérées afin d'équilibrer le rôle des groupes de variables entre eux. Cette ACP particulière fournit outre une représentation des individus et des variables, une représentation des individus décrits par chacun des groupes, ainsi qu'une représentation des groupes de variables. L'AFM permet de mettre en évidence, entre autres, des facteurs communs à tous les groupes, à certains d'entre eux ou spécifiques à l'un des groupes.

Jusqu'ici, il n'existait pas d'éléments de validation aux résultats fournis par l'AFM hormis ceux existant pour le coefficient RV. L'objet de cette communication est de proposer des éléments de validation pour deux indicateurs importants de l'AFM : la valeur propre d'une part, la mesure de liaison $\mathscr{L}_g$ d'autre part. L'importance de ces indicateurs se justifie à travers leur interprétation :

- Si λ_s désigne la valeur propre associée à la composante principale de rang s issue de l'AFM, son interprétation diffère de celle de l'ACP de par la prise

Marine Cadoret
Agrocampus Rennes, Laboratoire de mathématiques appliquées, 65 rue de Saint-Brieuc, CS 84215, 35042 RENNES Cedex e-mail: marine.cadoret@agrocampus-rennes.fr

en compte d'une structure en groupes, qui se traduit en particulier par une pondération des variables visant à équilibrer les groupes entre eux. Dans ce contexte, une valeur propre élevée correspond à une composante principale liée à plusieurs groupes, constituant alors une structure commune aux groupes. On attend donc de λ_s qu'elle apporte une information sur l'existence d'une structure commune entre l'ensemble des groupes de variables.

– Si $\mathscr{L}_g(z_s, K_j)$ désigne la mesure de liaison $\mathscr{L}_g$ entre la composante principale de rang s issue de l'AFM, notée z_s, et le groupe de variables j, noté K_j, la coordonnée du groupe j sur l'axe de rang s dans la représentation des groupes de variables évoquée précédemment s'interprète à la fois comme une contribution et comme la mesure de liaison $\mathscr{L}_g(z_s, K_j)$. Cette mesure apporte une information sur l'existence d'une structure commune au groupe j et au facteur commun z_s.

Ces éléments de validation sont obtenus sous forme de probabilités critiques, calculées composante par composante, sur la base de tests de permutations. Le principe général de ces tests est de générer, par des permutations dans les données, la distribution d'un indicateur sous une hypothèse nulle, et de situer dans cette distribution la valeur observée de l'indicateur. Le nombre de permutations ayant conduit à une valeur supérieure à celle observée nous donne la probabilité critique associée à cet indicateur. Ces tests de permutations font l'objet d'une utilisation courante en statistique [2].

Les questions posées sont :

– concernant la valeur propre de rang s : la dimension s peut-elle être considérée comme représentant une structure commune à au moins deux groupes de variables ?
– concernant la mesure $\mathscr{L}_g(z_s, K_j)$: cette dernière met-elle en évidence une structure commune entre le groupe j et au moins l'un des autres groupes ?

2 Méthodologie

A partir du jeu de données étudié, le principe consiste à perturber les données par des permutations des lignes, de façon indépendante d'un groupe de variables à l'autre.

Cette permutation ne change pas la structure interne d'un groupe en ce sens où les corrélations entre variables de ce groupe restent inchangées, mais a pour conséquence une modification des liaisons entre d'une part les variables du groupe et d'autre part les variables des autres groupes.

Deux choix sont alors envisageables : réaliser des permutations des lignes indépendantes d'un groupe à l'autre pour chacun des groupes ou pour un groupe en particulier. Ces deux choix permettent respectivement d'aborder les deux indicateurs que sont la valeur propre d'une part,et la mesure de liaison $\mathscr{L}_g$ d'autre part.

Concernant les valeurs propres, mesures de liaison globale, les permutations se font pour tous les groupes, indépendamment les uns des autres : les variables d'un même groupe subissent les mêmes permutations, alors que deux variables de deux groupes différents subissent des permutations différentes. Pour chaque ensemble de permutations, on calcule la valeur propre de rang s associée aux données ainsi perturbées. La distribution des valeurs propres obtenue correspond à l'hypothèse nulle suivante : les structures internes de chaque groupe étant données, il n'y a aucune structure commune entre les variables de groupes différents (absence de relation inter-groupe).

Concernant les mesures $\mathscr{L}_g(z_s, K_j)$, mesures de liaison "locale" entre le groupe de variables j et la composante principale de rang s, les permutations ne concernent que les données du groupe j, les autres groupes restant inchangés. Pour un groupe j donné, la distribution des $\mathscr{L}_g(z_s, K_j)$ obtenue correspond à l'hypothèse nulle suivante : le groupe j n'a pas de structure commune avec l'un des autres groupes.

Pour l'un et l'autre des deux indicateurs, on procède comme mentionné précédemment, c'est à dire axe par axe : les permutations sont réalisées d'abord pour le 1^{er} axe dans l'espace des variables, puis pour le 2^{eme} axe dans le sous-espace orthogonal au 1^{er} axe, et ainsi de suite.

3 Application

Cette méthodologie sera appliquée à un jeu de données sensorielles qui a déjà donné lieu à des interprétations, validées empiriquement par l'utilisateur. Nous confronterons ces interprétations aux épreuves de validité décrites ci-dessus.

Littérature

1. B. Escofier, J. Pagès, *Analyses factorielles simples et multiples*, 3^{eme} édition (Dunod, Paris, 1998)
2. F. Kasi-Aoual, S. Hitier, R. Sabatier, J-D. Lebreton : Refined approximations to permutation tests for multivariate inference. Computational Statistics and Data Analysis. **20**, 643–656 (1995)