

Un dialogue de sourds dans le théâtre statistique ? Une nouvelle méthode pour étudier les effets de structure avec l'analyse géométrique des données

Nicolas Robette, MCF UVSQ (Printemps), nicolas.robette@uvsq.fr

Olivier Roueff, CR CNRS (Cresppa-Csu), o.roueff@free.fr

Domaine: Analyse des données

Depuis leur diffusion à la fin des années 1960, les techniques de régression sont rapidement devenues dominantes dans les sciences sociales anglo-saxonnes. On peut y distinguer trois usages de la régression : l'un, « mesuré », consiste à contrôler la structure de la population étudiée, i.e. à raisonner « à propriétés sociales égales par ailleurs » (un exemple classique : Vallet, Caille, 1995) ; l'autre, « métrologique » (Desrosières, 2001), consiste à utiliser la régression pour tenter de hiérarchiser l'importance de différents facteurs causaux ; le dernier, « hyper-métrologique », consiste à évaluer le plus précisément possible l'effet « pur » des dispositifs d'action publique ou non gouvernementale, en calquant la démarche sur celle des sciences expérimentales. Cette « sociologie des variables » a été l'objet de nombreuses critiques épistémologiques et méthodologiques, y compris par des figures prééminentes de la sociologie, sans qu'elles atteignent toutefois l'« hégémonie économétrique ». La situation en France est assez différente, du fait de la relativement faible mobilisation des outils statistiques par les sciences sociales, des spécificités de la statistique publique française et de la forte association entre le programme de recherche élaboré par Pierre Bourdieu et l'analyse géométrique des données.

Dans ce contexte, la relation entre analyse géométrique et modèles de régression a tourné au dialogue de sourds. Leur complémentarité est parfois soulignée : à l'analyse des correspondances l'exploration et la description des données, à la régression l'explication, le rejet et la corroboration des hypothèses. Cette approche minore à la fois la capacité explicative de l'analyse géométrique, et l'impossibilité épistémologique d'une explication nomologique « pure » dans les sciences de la société. D'autres auteurs ont proposé une manière d'articuler les deux méthodes dans la perspective d'étudier – et non d'éliminer – les effets de structure : il s'agit de visualiser graphiquement les résultats d'une régression dans le cadre d'une analyse géométrique des données (Rouanet et alii, 2002 ; Lebaron, 2013).

Nous proposons quant à nous une nouvelle démarche en plusieurs étapes, l'Analyse Factorielle Standardisée (AFS), qui introduit l'usage de la régression au sein d'une analyse géométrique des données afin d'observer les effets de structure – la ou les variables souhaitées sont neutralisées en contraignant les axes d'une ACM (ou ACP, AFM...) à être indépendants des variables concernées : leurs modalités projetées en variables supplémentaires apparaissent (ou « disparaissent ») au centre du nuage des individus. Techniquement, l'AFS consiste : 1) à réaliser une ACM « classique » sur des propriétés individuelles ; 2) à régresser les coordonnées individuelles pour chacun des axes, avec comme variable indépendante la propriété à neutraliser (i.e. l'effet de structure que l'on

veut étudier) ; 3) à réaliser une ACP à partir des résidus des régressions et à projeter en supplémentaire les propriétés ayant servi à construire l'ACM. La comparaison de l'ACM « classique » et de l'AFS permet alors d'objectiver d'éventuels effets de structure.