

Peut-on construire des régions selon des critères statistiques ?

Marc CHRISTINE¹ (INSEE, DMCSI) et Michel ISNARD² (Insee, Secrétariat général)

La (re)construction de régions constitue, depuis quelques mois, un objet de débats intenses au sein de la classe politique et de controverses dans lesquelles les préoccupations des différents acteurs semblent assez éloignées de considérations statistiques. On a vu fleurir au fil des jours différentes cartes : projets, souhaits, simulations... avec apparition de 12, 13, 14 .. macro-régions, sans parler de la difficulté sémantique de désignation des entités ainsi construites. En novembre 2014 - date de rédaction du présent résumé -, le Sénat a voté un découpage de la France métropolitaine en 15 régions.

Tous les exercices de partitionnement géographique évoqués ci-dessus ont été réalisés dans le cadre extrêmement simple consistant à partir des régions existantes et à en agréger certaines d'entre elles pour arriver à un total de macro-régions de l'ordre de 11 à 12 (si l'on se limite aux régions continentales). Même si le nombre de combinaisons possibles demeure élevé, la contrainte de *contiguïté* des régions à agréger réduit sensiblement le champ des possibles, sans parler de la restriction imposée par les considérations politiques, historiques ou sociales qui « interdiraient » certains regroupements.

Reprenant une problématique initialisée aux JMS 2000³, puis développée aux JMS 2012⁴, le présent article se propose d'explorer des solutions de partitions du territoire en « régions », en procédant par des agrégations d'unités géographiques élémentaires plus fines que les actuelles régions : départements, arrondissements, zones d'emploi ou cantons.

Il s'agit de la mise en application des techniques, des algorithmes et des outils informatiques développés notamment lors de la dernière présentation citée, pour constituer, au sein d'une population de référence (le territoire métropolitain) des sous-ensembles présentant des conditions d'homogénéité ou d'hétérogénéité maximales vis-à-vis de certaines caractéristiques quantitatives. Mais les classes ainsi constituées doivent aussi respecter la contrainte forte de *connexité*, relativement au positionnement géographique des unités élémentaires appréhendé par le concept de *contiguïté*, qui constituera une relation binaire entre les unités de la population. Enfin, le nombre total de classes à constituer est fixé a priori (par exemple par le pouvoir politique).

On simulera donc différents partitionnements, selon les critères d'optimisation et selon les jeux de variables retenus : le cas standard est celui où l'on cherche à minimiser ou maximiser la variance intra-classes pour une variable quantitative standard : ceci illustre le cas de la distance euclidienne et d'une inertie s'interprétant en termes de dispersion.

D'autres critères seront analysés : avec plusieurs variables d'intérêt, quantitatives ou qualitatives ou avec des statistiques non linéaires.

Au final, il sera intéressant de comparer les cartes obtenues au découpage des régions actuelles ou aux projets de regroupement en cours d'élaboration et de vote au Parlement.

¹ marc.christine@insee.fr

² michel.isnard@insee.fr

³ « Un algorithme de regroupement d'unités statistiques selon certains critères de similitude », Marc CHRISTINE et Michel ISNARD, VII^{èmes} Journées de Méthodologie statistique, 4-5 décembre 2000.

⁴ « Agrégation optimale sous contrainte de contiguïté : aspects théoriques et mise en œuvre avec applications à des cas pratiques », Marc CHRISTINE et Michel ISNARD, XI^{èmes} Journées de Méthodologie statistique, 24-26 janvier 2012.