

Estimation sur petits domaines par scission des poids : application à une enquête sur la mobilité en France

Toky RANDRIANASOLO¹, Yves TILLÉ², Jimmy ARMOOGUM³

L'approche classique pour réaliser une estimation sur petits domaines consiste à utiliser un modèle. On utilise généralement un modèle linéaire mixte qui prend en compte les domaines dans la partie aléatoire du modèle. Un des problèmes que nous pouvons rencontrer dans ce type d'approche est qu'elle ne fournit pas nécessairement des estimations locales cohérentes avec des estimations globales. On voudrait en effet que la somme des estimations au niveau de ces petits domaines coïncide avec l'estimation au niveau du domaine global.

Nous proposons une nouvelle approche qui consiste à scinder les poids de l'estimateur global. On veut d'emblée respecter la propriété de cohérence. L'idée de l'approche est de créer un poids dépendant à la fois de l'individu et du petit domaine. Chaque individu pouvant ainsi contribuer à tous les domaines. Pour un individu donné, la somme de ses poids relatifs à chaque domaine doit être égale au poids global. Ce poids global peut être l'inverse de la probabilité d'inclusion dans l'échantillon ou un poids résultant d'un calage. De plus, on impose que le système de poids pour un domaine particulier satisfasse des propriétés de calage sur des variables auxiliaires connues pour ce domaine.

On construit ainsi une matrice notée Q dont le nombre de lignes est égal au nombre d'individus dans l'échantillon et dont le nombre de colonnes est égal au nombre de petits domaines. Cette matrice représente la manière dont le poids global de chaque individu est scindé entre les domaines. Cette matrice est stochastique car la somme de toutes ses lignes vaut 1. Il existe deux types de contributions : celle de l'individu à son propre domaine (que nous appelons « auto-contribution ») et celle de l'individu aux autres domaines que le sien (que nous appelons « extra-contribution »). Un individu qui contribue plus aux autres domaines qu'au sien, contribuera moins à son propre domaine et inversement.

Notre approche consiste à construire un certain estimateur composite ayant une partie d'« extra-contribution » construite à l'aide de la matrice de probabilité Q et une partie « domaine » avec une estimation directe (en utilisant directement les poids globaux pour les unités du domaine). Ces deux parties sont pondérées par un certain paramètre obtenu en minimisant la dispersion de la variable d'intérêt au niveau de chaque petit domaine de telle sorte que lorsqu'un petit domaine n'est pas représenté dans l'échantillon, la partie « auto-contribution » ne soit pas considérée et qu'inversement lorsque la taille d'un petit domaine

¹ <toky.randrianasolo@ifsttar.fr>

Département Economie et Sociologie des Transports (DEST) – IFSTTAR

² <yves.tille@unine.ch>

Institut de Statistiques Université de Neuchâtel

³ <jimmy.armoogum@ifsttar.fr>

Département Economie et Sociologie des Transports (DEST) – IFSTTAR

est très grande dans l'échantillon, la partie estimée par d'autres unités que celles du domaine soit négligeable.

On testera cette méthodologie sur les données Suisses du package « sampling » de R, notamment dans l'estimation de la population Suisse par tranche d'âge par canton. On l'appliquera alors dans l'estimation de la mobilité régionale en France à partir de l'Enquête Nationale sur les Transports et les Déplacements de 2007-2008.