

Un programme rapide d'échantillonnage équilibré.

Guillaume CHAUVET

Laboratoire de Statistique d'Enquête - CREST - ENSAI
École Nationale de la Statistique et de l'Analyse de l'Information
Rue Blaise Pascal, Campus de Ker Lann, 35170 Bruz, France
chauvet@ensai.fr

La méthode du Cube est une famille d'algorithmes qui permet la sélection d'échantillons équilibrés. Un grand nombre de plans de sondage classiques peuvent être décrits comme des plans équilibrés, comme le plan de taille fixe, à probabilités inégales, stratifié...

Nous proposons ici une nouvelle implémentation de la méthode Cube.

La première partie du travail consiste à accélérer le processus de tirage. Une étape du Cube consiste en effet à décider de sélectionner ou non une unité, ce qui revient essentiellement à choisir un vecteur du noyau d'une matrice ; dans l'ancienne version du Cube, la matrice en question est très grosse, ce qui occasionne des temps de calcul importants. Nous avons montré qu'on pouvait en fait se contenter de travailler sur une matrice de taille beaucoup plus petite, ce qui rend le déroulement plus rapide. Plus exactement, avec l'ancienne version du Cube, on prend à chaque étape une décision pour un individu choisi parmi tous les individus restants et, avec notre version, on prend à chaque étape une décision pour un individu choisi parmi un nombre restreint des individus restants.

Le temps d'exécution ne dépend alors plus que de la taille de la population (et non plus du carré de la taille), ce qui permet de l'utiliser sur des populations de plusieurs centaines de milliers de personnes.

Dans la deuxième partie du travail, les différentes étapes de la méthode ont été décomposées afin de rendre son utilisation plus souple.

Il s'agit de bien isoler les différentes étapes de la méthode du Cube, en particulier la phase dite de vol et la phase d'atterrissage, pour permettre des méthodes particulières d'échantillonnage coordonné dérivées du Cube et nécessitant de pouvoir appliquer séparément ces deux phases (ce qui n'est pas possible avec l'ancienne version). Des programmes de tirage coordonné seront écrits, fondés sur l'article de Favre et Tillé (tirage d'échantillon équilibré dans le complémentaire d'un échantillon équilibré, complémentation d'un échantillon pour l'équilibrer...). D'autre part, des simulations seront effectuées pour vérifier si, après avoir effectué un tri aléatoire sur le fichier puis en appliquant notre méthode, on obtient un tirage à entropie maximale.

Références

Chauvet, G. and Tillé, Y. (2004). A fast algorithm of balanced sampling, *Submitted for publication*.

Deville, J-C. and Tillé, Y. (2004). Efficient balanced sampling : the cube method. *Biometrika, accepted for publication*.

Tillé, Y. and Favre, A.-C. (2004). Co-ordination, combination and extension of optimal balanced samples. *To appear in Biometrika*.