

COORDINATION D'ÉCHANTILLONS ET ÉQUILIBRAGE : ASPECTS THÉORIQUES ET MISE EN ŒUVRE DANS LE CADRE DU RENOUVELLEMENT DES ÉCHANTILLONS DES ENQUÊTES D'ÉVALUATION DES ÉLÈVES

Marc Christine ¹ & Thierry Rocher ²

¹ *Insee-DMCSI, 18 boulevard Adolphe Pinard 75014 PARIS, marc.christine@insee.fr*

² *Depp, 61-65 rue Dutot 75732 Paris cedex 15, thierry.rocher@education.gouv.fr*

Ce papier fournit un cadre théorique et des méthodes de résolution d'un problème d'échantillonnage qui intervient dès lors qu'on a construit un 1er échantillon d'enquête à une certaine date et que, postérieurement, on souhaite tirer un 2ème échantillon de caractéristiques fixées (taille, probabilités d'inclusion) et présentant des conditions de recouvrement avec le 1er, mais sans pouvoir agir sur le tirage déjà effectué du 1er échantillon. On souhaite également que le 2ème échantillon ait des qualités de "représentativité" par rapport aux caractéristiques actuelles de l'univers, ce que l'on traduira par des *conditions d'équilibrage*.

Le cadre proposé s'applique à tous les cas où les unités sont tirées à *probabilités inégales* (souvent conditionnées par un facteur de taille) : enquêtes Entreprises, tirage d'unités primaires géographiques dans une enquête Ménages, tirage d'établissements scolaires... Il inclut différents cas de figure usuels : échantillons en deux phases, échantillons disjoints, échantillons avec recouvrement partiel, échantillons où l'on impose des conditions de "représentativité" lorsqu'on travaille soit sur la réunion des deux, soit sur le 2ème seulement.

L'approche théorique repose sur la notion d'**échantillonnages séquentiels conditionnels** et sur la technique **des échantillons équilibrés**. On montrera dans ce papier que l'ensemble des contraintes astreintes au 2ème échantillon sont en général incompatibles, ce qui conduira à **rechercher des solutions approchées**, obtenues en relâchant la contrainte relative aux probabilités d'inclusion finales du 2ème échantillon. Ces solutions approchées posent des problèmes de calcul explicite (en général seulement possible de manière numérique) et de propriétés statistiques des estimateurs en dérivant.

Une application est celle du tirage d'échantillons d'établissements scolaires pour les enquêtes d'évaluation des élèves. En effet, pour faire des comparaisons pertinentes entre deux périodes, il est préférable d'assurer un recouvrement entre les échantillons aux deux dates, mais chacun d'entre eux doit présenter une bonne "représentativité" par rapport à l'univers du moment.

On a donc effectué des simulations de tirage d'échantillons dans la base d'établissements 2009 avec des conditions de recouvrement par rapport à un échantillon tiré dans la base 2000 et des conditions d'équilibrage par rapport à l'univers de 2009. Dans un premier temps, pour chaque tirage d'échantillon dans la base 2000, on calcule les probabilités d'inclusion finale approchées de l'échantillon 2009 qui rendent le problème soluble. Dans un second temps, pour chaque tirage du 1er échantillon, on simule des tirages du 2nd échantillon avec les probabilités conditionnelles adaptées permettant d'obtenir les nouvelles probabilités finales approchées. On calcule des estimateurs de totaux (connus) de variables d'intérêt en utilisant ces nouvelles probabilités d'inclusion et on étudie empiriquement le biais et la précision de ces estimateurs.