

# Projet Migs 2

## Sondage indirect et méthode du partage des poids

Camelia Goga

premier semestre 2016

Lorsqu'on veut sélectionner un échantillon aléatoire dans une population cible, nous avons besoin d'une liste exhaustive, ou base de sondage, des individus de la population. De plus en plus souvent, il arrive qu'au lieu de disposer d'une base de sondage contenant la population souhaitée  $U^B$ , on a uniquement accès à une base de sondage contenant des unités,  $U^A$ , liées d'une certaine façon à la liste d'unités de  $U^B$ . Dans ces conditions, on sélectionne un échantillon  $s^A$  dans  $U^A$  et en utilisant les liens entre les deux populations, on produit une estimation pour  $U^B$ . On donne à cette approche le nom de sondage indirect (Deville and Lavallée, 2006, Lavallée, 2007).

Supposons par exemple que nous voulons obtenir des estimations pour une population d'enfants, mais que nous ne disposons d'une base de sondage que pour les parents. La population cible est la population des enfants  $U^B$ , mais nous n'avons une liste que pour les parents  $U^A$ . Nous devons sélectionner un échantillon  $s^A$  de parents avant de pouvoir interviewer les enfants.

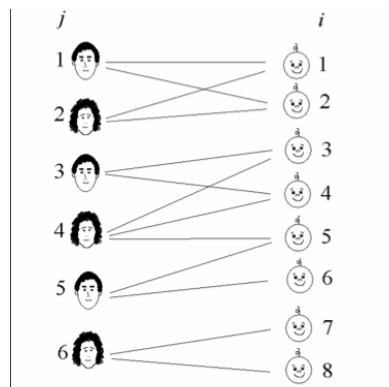


Figure 1. Population  $U^A$  of parents and population  $U^B$  of children with the links between the two.

L'estimation des caractéristiques de la population cible étudiée par sondage indirect peut poser un défi de taille, en particulier si le lien entre les unités des deux populations n'est pas bijectif. Le problème vient surtout de la difficulté à associer une probabilité de sélection, ou un poids d'estimation, aux unités étudiées de la population cible. La méthode généralisée du partage des poids (MGPP) a été mise au point par Deville and Lavallée (2006), Lavallée (2007) afin de résoudre ce genre de problème d'estimation. La MGPP fournit un poids d'estimation pour chaque unité enquêtée de la population cible.

Cette méthode est beaucoup appliquée en pratique. Par exemple, pour estimer le nombre de touristes en Bretagne, un échantillon de boulangeries a été utilisé par Deville and Maumy-Bertrand (2006). L'entreprise *La Poste* envisage également d'utiliser cette méthode pour estimer le nombre quotidien de lettres, colis, ... qui sont distribués en France.

**Travail demandé** : se familiariser avec les plans de sondage de ce type et étudier les propriétés de l'estimateur sans biais pour le total d'une variable  $Y$ .

### Bibliographie

1. Ardilly, P. (2006), *Les techniques de sondages*, éditions Technip.
2. Deville, J.-C. and Lavallée, P. (2006). Indirect sampling : the foundation of the generalized weight share method, *Survey Methodology*, 32, 2, 165-176.
3. Deville, J.-C. and Maumy-Bertrand, M. (2006). Extension of the Indirect Sampling and its Application to Tourism, *Survey Methodology*, 32, 2, 177-185.
4. Lavallée, P. (2007), *Indirect sampling*, Springer.
5. Tillé, Y. (2001), *Théorie des sondages*, Dunod.