

Le sondage indirect

Pierre Lavallée, Statistique Canada – Vendredi 21 novembre de 9h à 12h30

Pour tirer les échantillons nécessaires aux sondages sociaux ou économiques, il est utile de disposer de bases de sondage, c'est-à-dire de listes d'unités destinées à représenter les populations cibles. Malheureusement, il arrive qu'on ne dispose pas directement d'une liste contenant les unités de collecte désirées, mais plutôt d'une liste d'autres unités reliées d'une certaine façon à la liste des unités de collecte. On peut donc parler de deux populations U^A et U^B reliées entre elles où on désire produire une estimation pour U^B . Malheureusement, on dispose d'une base de sondage seulement pour U^A . On peut alors imaginer le tirage d'un échantillon de U^A afin de produire une estimation pour U^B en se servant de la correspondance existante entre les deux populations. C'est ce qu'on peut désigner par *sondage indirect* (Lavallée, 2002, 2007).

L'estimation d'un total (ou d'une moyenne) d'une population cible U^B de grappes en se servant d'un échantillon tiré d'une autre population U^A reliée d'une certaine façon à la première peut constituer un défi de taille, et ce, en particulier si les liens entre les unités des deux populations ne sont pas bijectifs. Le problème vient surtout de la difficulté d'associer une probabilité de sélection, ou un poids d'estimation, aux unités enquêtées dans la population cible. Afin de résoudre ce type de problème d'estimation, on a développé la *méthode généralisée du partage des poids* (MGPP). La MGPP permet d'obtenir un poids d'estimation pour chaque unité enquêtée de la population cible U^B . Ce poids d'estimation correspond en gros à une moyenne des poids de sondage des unités de la population U^A d'où est tiré l'échantillon.

Le cours se veut un survol des différents développements effectués autour du sondage indirect. La théorie entourant la MGPP y sera présentée, mais aussi différentes applications possibles qui accentuent l'attrait de celle-ci, notamment le sondage de populations difficiles à joindre. Dans un premier temps, on donnera une description formelle de la MGPP et on décrira son utilité. Ensuite, on présentera une revue littéraire où on associera la MGPP à différentes méthodes de sondage apparues dans la littérature. C'est ainsi que l'on verra que le sondage indirect est une généralisation de méthodes de sondage telles que le sondage par réseaux (Network Sampling) et le *sondage en grappes adaptatif* (Adaptive Cluster Sampling). On présentera par la suite des résultats théoriques sur la MGPP, notamment l'aspect sans biais de la méthode et la variance des estimations découlant de celle-ci. On examinera ensuite d'autres généralisations possibles de la MGPP. Par exemple, on décrira comment étendre le sondage indirect d'un degré à deux degrés. On verra par la suite l'une des principales applications de la MGPP, soit celle reliée aux enquêtes longitudinales. On décrira aussi comment on peut tenter d'améliorer la précision des estimations découlant de la MGPP en utilisant le calage sur marges. Ensuite, on traitera le cas pratique où de la non-réponse survient lors de la collecte des données. On verra qu'on peut corriger les poids issus de la

Ateliers de formation – 17 et 21 novembre 2014

8ème colloque francophone sur les sondages - Dijon



MGPP en calculant une probabilité de réponse associée aux unités répondantes. Finalement, on terminera par une conclusion qui met l'accent sur de nouvelles applications de la MGPP.

Ateliers de formation – 17 et 21 novembre 2014

8ème colloque francophone sur les sondages - Dijon



Bibliographie

- Lavallée, P. (2002). *Le sondage indirect, ou la méthode généralisée du partage des poids*. Éditions de l'Université de Bruxelles (Belgique) et Éditions Ellipses (France), 215 pages.
- Lavallée, P. (2007). *Indirect sampling*. Springer, New York, 245 pages.