

Méthodes d'estimation sur bases de sondage multiples dans le cadre de plans de sondage à deux degrés

Guillaume Chauvet, ENSAI (CREST), Campus de Ker Lann, Bruz, chauvet@ensai.fr

Guylène Tandeau de Marsac, INSEE, Direction Régionale de Lille, guylene.tandeau-de-marsac@insee.fr

Lorsqu'on s'intéresse à une population finie, il arrive qu'aucune base de sondage ne la recouvre totalement et qu'il soit nécessaire de tirer des échantillons dans deux bases de sondage (ou plus) pour représenter l'ensemble des individus. Pour mettre en commun ces échantillons, de nombreuses méthodes d'estimation sur bases de sondage multiples ont été proposées (Hartley, 1962; Bankier, 1986; Kalton et Anderson, 1986; Mecatti, 2007; Rao et Wu, 2010); voir également les articles de revue de Lohr(2009, 2011), et les articles référencés, pour un panorama complet. Notons que la méthode de Mecatti (2007) s'inspire des travaux de Lavallée (2002, 2007) sur la Méthode généralisée du partage des poids.

Nous nous intéressons ici au cas de deux échantillons sélectionnés selon un plan à deux degrés, avec un premier degré de tirage commun. Ce cadre correspond aux enquêtes de l'INSEE avec extension : un premier échantillon de logements est sélectionné dans les communes de l'ancien Echantillon-maître (Bourdalle et al., 2000), et un second échantillon est sélectionné et enquêté dans les communes du même Echantillon-maître afin de cibler une sous-population spécifique. On dispose de deux mesures d'enquêtes provenant de deux échantillons indépendants au deuxième degré du plan de sondage. Nous appliquons des méthodes d'estimation sur bases de sondage multiples pour la mise en commun de ces deux échantillons. Nous montrons que les estimateurs étudiés peuvent dans ce contexte être calculés conditionnellement au premier degré de tirage, ce qui simplifie leur calcul notamment pour l'estimateur optimal de Hartley (1962).

Dans le cadre d'une étude par simulations, nous avons étudié les estimateurs de Hartley (1962), de Kalton et Anderson (1986) et de Bankier (1986) pour mettre en commun les échantillons issus de deux vagues d'enquête. Nous avons plus particulièrement étudié le cas où un échantillon représente la population entière (échantillon complètement représentatif), alors que le second n'en représente qu'une partie (échantillon partiellement représentatif). Dans le cadre considéré, l'utilisation de l'échantillon partiellement représentatif ne permet pas de gagner en précision. Ces résultats suggèrent qu'un estimateur simple est parfois préférable, même s'il n'utilise qu'une partie de l'information collectée.

Il s'agit d'un travail en collaboration avec Guylène Tandeau de Marsac, qui a donné lieu à un article à paraître dans *Survey Methodology*.

Références

- Bankier, M.D. (1986). Estimators Based on Several Stratified Samples With Applications to Multiple Frame Surveys. *Journal of the American Statistical Association*, 81, p.1074-1079.
- Bourdalle, G., Christine, M., et Wilms, L. (2000). Échantillons maître et emploi. *Série INSEE Méthodes*, 21, p. 139-173.
- Hansen, M.H., et Hurwitz, W.N. (1943). On the theory of sampling from finite populations. *Annals of Mathematical Statistics*, 14, p. 333-362.
- Hartley, H.O. (1962). Multiple frame surveys. *Proceedings of the Social Statistics Section, American Statistical Association*, p. 203-206.
- Kalton, G., et Anderson, D.W. (1986). Sampling Rare Populations. *Journal of the Royal Statistical Society, A*, 149, p. 65-82.
- 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).
- Lavallée, P. (2007). *Indirect Sampling*. New York: Springer.
- Lohr, S.L. (2007). Recent developments in multiple frame surveys. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 3257-3264.
- Lohr, S.L. (2009). Multiple frame surveys. In *Handbook of Statistics, Sample Surveys: Design, Methods and Applications*, (Eds., D. Pfeffermann and C.R. Rao). Amsterdam: North Holland, Vol. 29A, p. 71-88.
- Lohr, P. (2011). Alternative survey sample designs : Sampling with multiple overlapping frames. *Survey Methodology*, 37, p. 197-213.
- Mecatti, F. (2007). A single frame multiplicity estimator for multiple frame surveys. *Survey Methodology*, 33, p. 151-157.
- Rao, J.N.K., et Wu, C. (2010). Pseudo-empirical likelihood inference for dual frame surveys. *Journal of the American Statistical Association*, 105, p. 1494-1503.