

LE PLAN DE SONDAGE ÉQUILIBRÉ ET L'ESTIMATION DU TOTAL D'UNE POPULATION FINIE

Mohammed El Haj Tirari¹

RÉSUMÉ

En présence d'information auxiliaire, un plan de sondage équilibré permet de sélectionner des échantillons vérifiant que l'estimateur de Horvitz-Thompson des totaux d'un ensemble de variables auxiliaires sont égaux aux vrais totaux connus de ces variables. Dans de ce travail, nous montrons comment l'utilisation d'un plan de sondage équilibré peut permettre de garantir la robustesse des estimateurs par régression d'un total, que se soit contre une mauvaise spécification de modèle ou contre les variations d'échantillonnage.

1. INTRODUCTION

Pour estimer le total d'une population finie, selon l'approche statistique considérée et la présence ou pas d'information auxiliaire, plusieurs estimateurs ont été proposés. Les propriétés d'optimalité de ces estimateurs dépendent généralement des critères de choix adoptés, sans que aucun de ces estimateurs ne soit optimal par rapport à deux les critères existants.

En présence d'information auxiliaire, l'estimation par régression est considérée parmi les méthodes les plus efficaces qui permettent de mettre à profit cette information auxiliaire lors de l'estimation des totaux d'une population finie. L'information auxiliaire disponible peut être utilisée également au moment de la sélection de l'échantillon ; c'est le cas par exemple pour les plans de sondage équilibrés. Dans ce travail, en considérant les différentes approches statistiques utilisées en théorie des sondages, nous verrons comment l'utilisation d'un plan équilibré peut être mise à profit de telle sorte à obtenir des propriétés intéressantes pour les estimateurs par régression du total d'une population finie.

2. L'APPROCHE BASÉE SUR LE PLAN DE SONDAGE

Dans une approche basée sur le plan de sondage, en recherchant le coefficient de régression qui minimise la variance asymptotique de l'estimateur par la régression généralisée *GREG*, Montanari (1987) propose un estimateur *GREG* du total qui est optimal selon ce critère. Malheureusement, en plus de dépendre des probabilités d'inclusion d'ordre deux, cet estimateur n'est optimal que lorsque son coefficient de régression est connu, ce qui n'est pas le cas. De plus, le coefficient de régression, de l'estimateur *GREG* proposé par Montanari, est une fonction des estimateurs de la variance et de la covariance, ce qui le rend vulnérable aux variations de l'échantillonnage. Dans ce travail, nous montrons que, si un plan de sondage équilibré est utilisé, on peut obtenir une approximation de l'estimateur optimal de Montanari, qui est indépendante des probabilités d'inclusion d'ordre deux avec un coefficient de régression plus robuste contre les variations d'échantillonnage.

3. L'APPROCHE MODÈLE

Sous l'approche modèle, plusieurs estimateurs par régression du total ont été proposés (voir Royall, 1992 ; Särndal, 1980 ; Särndal *et al.*, 1992 ; Wright, 1983). Quel que soit, l'approche modèle considérée (basée sur le modèle ou uniquement assistée par le modèle), L'inconvénient majeur de ces estimateurs est qu'il

¹ Mohamed El Haj Tirari, Laboratoire de Statistique d'Enquêtes, CREST – Ensai, Campus de Ker-Lann, Rue Blaise Pascal – BP 37203, France, 35172 Bruz Cedex (tirari@ensai.fr).

sont tous vulnérables à une mauvaise spécification de modèle. Dans cette partie, en supposant que le plan de sondage mis en oeuvre est équilibré, nous étudierons les propriétés des estimateurs par régression d'un total proposés sous l'approche modèle. Nous montrons que, sous certaines conditions sur le plan de sondage, le fait de sélectionner des échantillons équilibrés permet de garantir la robustesse de ces estimateurs du total quand le modèle de régression considéré est incorrect.

RÉFÉRENCES

- Berger, Y.G., Tirari, M.E.H., Tillé, Y. (2003), « Toward optimal regression estimation in sample surveys », *Australian and New Zealand Journal of Statistics*, 45, p. 319-329.
- Deville, J.C., Tillé, Y., (2004), « Efficient balanced sampling : The cub method », *Biometrika*, 91, p. 893-912.
- Montanari, G.E. (1987), « Post sampling efficient QR-prediction in large sample survey », *International Statistical Review*, 55, 191-202.
- Royall, R.M. (1992), « Robustness and optimal design under prediction models for finite populations », *Survey Methodology*, 18, 179-185 .
- Särndal, C.E., (1980), « On π -inverse weighting versus best linear unbiased weighting in probability sampling », *Biometrika*, 67, 639-650.
- Särndal, C.E., Swenson, B., Wretman, J.H. (1992), *Model Assisted Survey Sampling*, New York, Springer-Verlag.
- Wright, R.L., (1983), « Finite Population Sampling With Multivariate Auxiliary Information », *Journal of the American Statistical Association*, 78, 879-884.