

TRAITEMENT DES DONNÉES INFLUENTES DANS LE CAS D'UN SONDAGE À DEUX PHASES AVEC UNE APPLICATION AU TRAITEMENT DE LA NON-RÉPONSE

Cyril Favre Martinoz ¹ & Jean-François Beaumont ² & David Haziza ³

¹ *LSE, Crest-Ensaï, cyril.favremartinoz@ensai.fr*

² *Division de la recherche et de l'innovation en statistique, Statistique Canada,
jean-francois.beaumont@statcan.gc.ca*

³ *Département de mathématiques et de statistique, Université de Montréal,
David.Haziza@umontreal.ca*

Dans les enquêtes auprès des entreprises, on recueille souvent des variables économiques dont la distribution est fortement asymétrique. Par conséquent, certaines unités (par exemple, celles qui appartiennent à la queue de distribution) peuvent avoir une grande influence sur les estimateurs. Une unité influente est une unité qui, étant donné une variable d'intérêt, un paramètre d'intérêt, un plan de sondage et un estimateur, a un impact significatif sur l'erreur due à l'échantillonnage de l'estimateur considéré. Pour remédier à ce problème, il est nécessaire de construire des estimateurs robustes qui limitent l'impact des unités ayant une grande influence.

Nous allons présenter une généralisation au cas d'un sondage à deux phases des résultats de l'article de Beaumont *et al.* (2011), qui proposent de construire des estimateurs robustes pour des totaux à partir du concept de biais conditionnel d'une unité pour un sondage à une phase. Nous étendons la notion d'influence au cas d'un sondage à deux phases et construisons un estimateur robuste dans le cas d'un tirage à deux phases. L'estimateur robuste construit dépend du paramètre de contrôle de la fonction de Huber ; le choix de ce paramètre est très important et il sera détaillé. Ensuite, nous présentons une application au traitement de la non-réponse. Enfin, nous détaillons des études par simulation afin de comparer la performance des estimateurs robustes et des estimateurs non-robustes en termes de biais relatif et d'efficacité relative.

Bibliographie

Kim, J.K., et Kim, J.J. (2007). Nonresponse weighting adjustment using estimated response probability, *The Canadian Journal of Statistics*, 35 , 501-514.

Moreno-Rebollo, J.L., et Muñoz-Reyez, A.M., et Muñoz-Pichardo, J.M. (1999). Influence diagnostics in survey sampling : conditional bias, *Biometrika*, 86, 923-968.

Moreno-Rebollo, J.L., et Muñoz-Reyez, A.M., et Jimenez-Gamero, M.D., et Muñoz-Pichardo, J.M. (2002). Influence diagnostics in survey sampling : estimating the conditional bias, *Metrika*, 55, 209-214.

Beaumont, J.-F, et Haziza, D., et Ruiz-Gazen, A. (2011). A unified approach to robust estimation in finite population sampling. En révision pour *Biometrika*.