

# CONSISTANCE SOUS UN MODÈLE DE RÉPONSE DE LA FONCTION DE RÉPARTITION ESTIMÉE EN PRÉSENCE DE DONNÉES MANQUANTES

Guillaume Chauvet <sup>1</sup> & Hélène Boistard <sup>2</sup> & David Haziza <sup>3</sup>

<sup>1</sup> *Ensaï (Crest)*, [chauvet@ensai.fr](mailto:chauvet@ensai.fr)

<sup>2</sup> *Université Toulouse 1*, [helene@boistard.fr](mailto:helene@boistard.fr)

<sup>3</sup> *Université de Montréal*, [David.Haziza@umontreal.ca](mailto:David.Haziza@umontreal.ca)

On rencontre des problèmes de données manquantes dans les enquêtes quand certaines des unités refusent de répondre, ou quand il est impossible de les contacter. Nous nous intéressons ici au cas où une méthode d'imputation simple est utilisée pour remplacer une valeur manquante par une valeur artificielle. Afin d'étudier les propriétés des estimateurs, nous considérons deux approches : (i) l'approche par le modèle de non-réponse (NM), et (ii) l'approche par le modèle d'imputation (IM) qui nécessite la spécification d'un modèle décrivant la distribution de la variable étudiée. Les procédures doublement robustes offrent, dans une certaine mesure, une protection contre une mauvaise spécification de l'un des deux modèles ; voir par exemple Kang et Schafer (2007), Cao *et al.* (2009) pour la Statistique Classique, Kim et Park (2006) et Haziza et Rao (2006) dans le cadre d'un sondage, et les références dans ces articles.

La littérature mentionnée ci-dessus traite seulement le cas où l'on estime la valeur moyenne de la variable d'intérêt. À notre connaissance, le cas d'un estimateur doublement robuste de la fonction de répartition est peu (ou pas) abordé dans la littérature. Dans ce travail, nous considérons le cas d'un mécanisme d'imputation des valeurs manquantes par la régression aléatoire pondérée (motivé par le modèle IM). Plus précisément, pour chaque non-répondant, un résidu observé pour un répondant est sélectionné au hasard avec une probabilité proportionnelle à un poids d'imputation, et utilisé dans l'imputation de la valeur manquante. Nous étudions les propriétés de double robustesse de la fonction de répartition estimée, en fonction des poids d'imputation utilisés, théoriquement et à l'aide d'une étude par simulations.

## Bibliographie

Cao, W., et Tsiatis, A.A., et Davidian, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data, *Biometrika*, 96, 723-734.

Haziza, D., et Rao, J.N.K. (2006). A nonresponse model approach to inference under imputation for missing survey data, *Survey Methodology*, 32, 53-64.

Kang, D.Y., et Schafer, J.L. (2007). Demystifying double robustness : a comparison of alternative strategies for estimating a population mean from incomplete data (with discussion and rejoinder), *Statistical Science*, 22, 523-539.

Kim, J.K., et Park, H.A. (2006). Imputation using response probability, *Canadian Journal of Statistics*, 34, 171-182.