

Méthodes d'imputation aléatoires équilibrées

Guillaume CHAUVET¹, Jean-Claude DEVILLE² et David HAZIZA³

En pratique, il est souvent requis d'estimer des totaux (ou moyennes) ou des quantiles (par exemple, la médiane). Bien qu'une imputation déterministe puisse mener à des estimateurs (approximativement) pour des totaux, elle est fortement déconseillée pour des quantiles car elle distord généralement la distribution de la variable que l'on impute. Les estimateurs de quantiles peuvent être alors considérablement biaisés, surtout si les taux de non-réponse sont élevés. C'est pourquoi on a fréquemment recours à des méthodes d'imputation aléatoires qui tendent à préserver la distribution des variables d'intérêt. Cependant, ces dernières introduisent une variabilité supplémentaire (appelée variance due à l'imputation) due à la sélection aléatoire de résidus. Les estimateurs résultant sont donc souvent inefficaces.

L'objectif visé est de déterminer une classe de méthodes aléatoires pour lesquelles la variance due à l'imputation est considérablement réduite (ou éliminée) tout en préservant la distribution de la variable que l'on impute. Trois approches ont été proposées dans la littérature : (i) l'imputation fractionnelle (voir par exemple, Kalton et Kish, 1981, Fay, 1996 et Kim et Fuller, 2004). (ii) Une procédure d'ajustement des valeurs imputées proposée par Chen, Rao et Sitter (2000) dans le contexte de l'imputation par hot-deck aléatoire. (iii) Une approche de sélection des résidus aléatoires de manière à réduire la variance due à l'imputation (Kalton et Kish, 1981 et Deville, 2005).

Dans cette présentation, nous considérons la troisième approche qui nous amènera à discuter d'une classe de méthodes d'imputation, appelée méthodes d'imputation aléatoires équilibrées. Le but sera de tirer des résidus aléatoirement en satisfaisant certaines contraintes appropriées de manière à réduire ou éliminer la variance due à l'imputation. Ces méthodes étant liées à l'échantillonnage équilibré, nous proposons d'adapter la méthode du Cube proposée par Deville et Tillé (2004) pour arriver à nos fins. L'algorithme proposé sera général en ce sens où il pourra être utilisé pour une grande classe de méthodes d'imputation aléatoires et n'importe quel plan de sondage. Il pourra être utilisé pour un grand nombre de paramètres (totaux, quantiles, etc.). Nous présenterons les résultats d'une étude par simulation afin d'évaluer les propriétés de cette approche.

¹ Ensai/Crest, DR Bretagne, Insee - chauvet@ensai.fr

² Ensai/Crest, Insee - Jean-Claude.Deville@ensai.fr

³ Université de Montréal - david.haziza@umontreal.ca