

# Inférence triplement robuste en présence de données manquantes

*David HAZIZA<sup>1</sup>, Valéry DONGMO JIONGO<sup>2</sup> et Pierre DUCHESNE<sup>3</sup>*

Afin de traiter la non-réponse partielle, plusieurs méthodes d'imputation sont utilisées en pratique (imputation par la régression, imputation par le ratio, imputation par la moyenne, etc). Les estimateurs résultants sont asymptotiquement sans biais si le modèle de non-réponse et/ou le modèle d'imputation est correctement spécifié.

Dans certaines situations, le statisticien d'enquête doit faire face à la présence de valeurs aberrantes. Ce type de problème survient fréquemment dans le contexte des enquêtes auprès des entreprises qui collectent des variables économiques (par exemple, le revenu), dont la distribution est habituellement très asymétrique.

Bien qu'il soit possible de réduire l'impact des valeurs aberrantes à l'étape du plan de sondage (en utilisant par exemple des strates à tout prendre) ou à l'étape de l'imputation (en construisant des classes d'imputation homogènes par rapport à la variable que l'on cherche à imputer), il n'est pas généralement possible d'éliminer le problème des valeurs aberrantes. En présence de ce type de valeurs, les estimateurs imputés ont habituellement une très grande variance.

Une solution naturelle serait de remplacer les valeurs imputées par une version robuste.

Par exemple, supposons que la non-réponse partielle est traitée au moyen d'une imputation par la régression dont le coefficient de régression est estimé au moyen la méthode des moindres carrés pondérés. Il est alors tentant de remplacer ce coefficient de régression par une version robuste (par exemple, un estimateur M).

Cependant, ce type de méthode mènera généralement à des estimateurs biaisés dont l'erreur quadratique moyenne est souvent plus élevée que celle des estimateurs non-robustes.

Nous proposons un estimateur robuste construit au moyen du concept de biais conditionnel d'une unité qui est une mesure de son influence.

Nous montrons qu'une valeur aberrante peut avoir un impact important sur l'erreur due à l'échantillonnage et/ou sur l'erreur de non-réponse. Le biais conditionnel comme mesure d'influence est attrayant car il tient compte du plan de sondage, du modèle de non-réponse ainsi que du modèle d'imputation. Il s'agira alors de réduire l'influence des valeurs qui exhibent un biais conditionnel important. L'estimateur proposé est généralement biaisé mais plus stable que l'estimateur non-robuste.

Cependant, la forme de ce dernier est relativement complexe. Il est donc irréaliste de supposer que tous les utilisateurs de données seront en mesure de le calculer.

Pour contourner ce problème, on aura recours à l'imputation calée qui consiste à déterminer des valeurs imputées finales aussi proche que possible des valeurs imputées initiales et tel que la

---

<sup>1</sup> Université de Montréal

<sup>2</sup> Université de Montréal

<sup>3</sup> Université de Montréal

somme pondérée des valeurs observées et des valeur imputées finales coïncident avec l'estimateur robuste proposé.

Les résultats d'une étude par simulation seront présentés afin d'étudier le comportement de l'estimateur proposé en termes de biais relatif et d'efficacité relative.