
Bootstrap avec remise pour l'estimation de variance dans des enquêtes auprès de ménages Principes et exemples

Guillaume CHAUVET ()*, *Pascal BESSONNEAU (**)*, *Gwennaëlle BRILHAULT (**)*, *Cédric GARCIA (***)*

() ENSAI (Irmar), Rennes*

*(**) Ined, Aubervilliers*

*(***) Université Gustave Eiffel, Champs-sur-Marne*

`chauvet@ensai.fr`

Mots-clés. (6 maximum) : Échantillonnage, variance, bootstrap

Domaines. Théorie des sondages aval : estimation de variance

Résumé

L'estimation de la variance dans les enquêtes est un problème complexe. Les poids d'estimation sont le résultat de nombreux traitements statistiques, parmi lesquels le traitement de la non-réponse et le calage, et leur impact sur la variance doit être pris en compte. Le bootstrap est un outil important, conduisant à la création de ce que l'on appelle les poids bootstrap, qui sont publiés avec les données d'enquête. Ces poids peuvent être utilisés pour calculer de façon répétée la version bootstrap du paramètre d'intérêt, conduisant à un estimateur de variance (ou un intervalle de confiance) obtenu par simulations. L'intérêt pour les utilisateurs est qu'aucune information autre que les poids bootstrap n'est alors nécessaire pour l'estimation de la variance. En particulier, une description complète du plan de sondage et du processus d'estimation n'est pas nécessaire, ce qui serait le cas dans le cadre d'une approche analytique où l'estimateur de variance doit être explicitement écrit.

Il existe une littérature abondante sur le bootstrap pour données d'enquête, voir par exemple Shao et Tu [1], Davison et Hinkley [2], Chauvet [3] et Mashreghi et al. [4]. L'une de ces techniques est le "rescaled bootstrap" proposé par Rao et al. [5]. Il implique de calculer des poids bootstrap initiaux, puis de les recalculer afin de reproduire un estimateur de variance sans biais pour l'estimation d'un total (cas linéaire).

Un cas particulier du bootstrap Rao-Wu est obtenu en utilisant une taille de rééchantillonnage égale à la taille de l'échantillon moins un. Cela conduit à ce que l'on appelle le bootstrap des

unités primaires ou bootstrap avec remise (Mac Carthy et Snowden, [6]). Le bootstrap avec remise est assez simple à mettre en oeuvre. Il nécessite de rééchantillonner uniquement les unités d'échantillonnage, et non les unités finales. La prise en compte du traitement de la non-réponse et du calage est assez naturelle. Le bootstrap avec remise conduit à une estimation prudente de la variance, en ce sens que la variance d'échantillonnage au premier degré est généralement surestimée. Il s'agit donc d'une approche prudente dans la production d'intervalles de confiance. Le biais positif de l'estimateur bootstrap de la variance est négligeable si les taux d'échantillonnage au premier degré à l'intérieur des strates sont négligeables. De plus, si l'enquête est répétée au fil du temps, la contribution de la variance d'échantillonnage de premier degré s'estompe avec le temps, alors que la variance due à l'attrition et à la non-réponse totale augmente.

Dans cet exposé, nous expliquerons comment cette méthode de bootstrap peut être appliquée dans le cas d'une enquête de type ménage-individu. Ce travail s'appuie sur un article paru dans *Techniques d'Enquêtes*, et disponible [avec le lien suivant](#).

Abstract

Variance estimation is a challenging problem in surveys. The final weights used at the estimation stage include several statistical treatments, including correction of unit non-response and calibration, and their impact on the variance is to be assessed. Bootstrap is a useful tool, leading to the creation of so-called bootstrap weights released with the survey data set. These weights can be used to compute repeatedly the bootstrap version of the parameter of interest, leading to a simulation-based variance estimator or confidence interval. The interest for practitioners is that no information other than the bootstrap weights is needed for variance estimation. In particular, a comprehensive description of the original sampling design and estimation process is not required, which would be the case under an analytic approach where the variance estimator needs to be worked out.

There is an extensive literature on bootstrap in survey sampling, see for example Shao and Tu [1], Davison and Hinkley [2], Chauvet [3] and Mashreghi et al. [4]. One of these techniques is the so-called rescaled bootstrap proposed by Rao et al. [5]. Initial bootstrap weights are obtained, and these weights are then rescaled so as to reproduce an unbiased variance estimator for the estimation of a total (linear case).

A particular case of the Rao-Wu bootstrap is obtained when the resample size is the sample size minus one. This leads to the so-called bootstrap of Primary Sampling Units (PSUs) or with-replacement bootstrap (Mac Carthy et Snowden, [6]). The with-replacement bootstrap is fairly simple to implement; in particular, it requires to resample the primary sampling units only, and not the final units. Accounting for treatment of non-response and calibration is fairly natural. The with-replacement bootstrap leads to conservative variance estimation, in the sense that the first-stage sampling variance is usually overestimated. This is therefore a prudent approach in producing confidence intervals. The positive bias of the bootstrap variance estimator is expected to be negligible when the first-stage sampling rates inside strata are negligible, which is often the case in phone surveys. Also, if the survey is repeated over time, the contribution of the first-stage sampling variance is likely to fade while the variance due to attrition and unit non-response grows bigger.

In this work, we explain how this bootstrap method may be used for a survey for households and individuals. This work is based on a paper published in *Survey Methodology*, and available [with the following link](#).

Bibliographie

- [1] Shao, J. and Tu, D. S. (1995). The jackknife and bootstrap. Springer Series in Statistics.
- [2] Davison, A. C. and Hinkley, D. V. (1997). Bootstrap methods and their application. Cambridge Series in Statistical and Probabilistic Mathematics.
- [3] Chauvet, G. (2007). Méthodes de Bootstrap en population finie. Thèse de doctorat, Université de Rennes 2.
- [4] Mashreghi, Z., Haziza, D., and Léger, C. (2016). A survey of bootstrap methods in finite population sampling. *Statistical Surveys*, 10, pp. 1-52.
- [5] Rao, J., Wu, C., and Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, 18(2), pp. 209-217.
- [6] McCarthy, P. and Snowden, C. (1985). The bootstrap and finite population sampling. *Vital and Health statistics. Series 2, Data Evaluation and Methods Research*, 95, pp. 1-23.
- [7] Bessonneau, P., Brilhault, G., Chauvet, G., et Garcia, C. (2021). With-replacement bootstrap variance estimation for household surveys : Principles, examples and implementation. A paraître dans *Survey Methodology*.