
Estimation de la variance pour un plan produit

Jean RUBIN (*), Guillaume CHAUVET (**)

(*) Insee (ENSAI)

(**) ENSAI (Irmarr), Rennes

jean.rubin@ensae.fr

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

Domaines. Théorie des sondages aval : estimation de variance

Résumé

Cet article propose d'étudier un type de plan de sondage appelée *plan produit* (ou *cross-classified sampling* en anglais) relativement similaire à l'échantillonnage à plusieurs degrés et pouvant apparaître lorsque la notion de hiérarchie entre les populations "primaires" et "secondaires" est moins apparente. Ce plan de sondage est notamment utilisé dans l'Etude Longitudinale Française depuis l'Enfance (ELFE), cf. Juillard et al. [2].

En reprenant les termes d'Ohlsson [1], on considère une population dite de *lignes* et une population dite de *colonnes*. Le plan produit consiste à tirer un échantillon dans ces deux populations de manière indépendante, échantillons que l'on notera S_R et S_C . On observe ensuite les variables d'intérêt correspondant aux lignes et colonnes sélectionnées, ie. $\{y_{i,k}, i \in S_R, k \in S_C\}$. Comme indiqué par Skinner [3], ce type de plan de sondage peut par exemple apparaître lorsque l'on souhaite échantillonner des enseignes et des objets afin de comparer les prix d'objets identiques dans des enseignes différentes. Un autre exemple présenté par Juillard et al. [2] correspond à la sélection d'un échantillon de maternités et d'un échantillon de dates sélectionnés indépendamment, afin de recueillir des informations sur les maternités à des dates identiques.

Cette méthode de tirage crée une interdépendance des lignes et des colonnes dans la sélection des unités, ce qui complique les estimations de la variance par rapport à un tirage à deux degrés classique. Plusieurs approches ont été proposées pour traiter ce problème. Ohlsson [1] et Skinner [3] ont proposé des estimateurs sans biais de variance pour le cas où les échantillons sont sélectionnés selon un sondage aléatoire stratifié dans chaque dimension. Skinner [2] a également proposé une méthode de bootstrap pour le cas où les échantillons dans chaque dimension sont tirés avec remise. Juillard et al. [2] ont proposé des estimateurs de variance en se basant sur une comparaison avec un tirage à deux degrés. Ces derniers estimateurs ont l'avantage d'être toujours positifs.

Dans ce travail, nous considérons le cas d'un plan produit avec un nombre fini de dimensions, possiblement supérieur à 2. En nous basant sur la décomposition de Hoeffding-Sobol, nous montrons que l'estimateur de Horvitz-Thompson sous un plan produit peut s'écrire comme une somme de variables aléatoires non corrélées, ce qui généralise la décomposition présentée par Ohlsson [1]. En utilisant un cadre théorique similaire à celui de Juillard et al. [2], nous montrons sous des conditions générales que la plus grande partie de ces termes peut être négligée dans l'estimation de variance. Nous en déduisons en particulier une méthode simple de bootstrap de type pseudo-population (Gross [4]), applicable dans le cas du tirage simple sans remise dans chaque dimension.

Bibliographie

- [1] Ohlsson, E. (1996). Cross-classified sampling. *Journal of Official Statistics*
- [2] Juillard, H., Chauvet, G., and Ruiz-Gazen, A. (2015). Estimateurs de variance issus d'un plan produit pour l'enquête elfe. XII^{ème} édition des Journées de Méthodologie Statistique, 173.
- [3] Skinner, C. (2015). Cross-classified sampling : some estimation theory. *Statistics & Probability Letters*, 104 :163–168.
- [4] Gross, S. (1980). Median estimation in sample surveys. *Proceedings of the Section on Survey Research Methods*, pages 181–184.