Techniques d'enquête, décembre 1987
Vol. 13, n° 2, pp. 233–236
Statistique Canada

Comparaison de la méthode de Horvitz-Thompson et de la méthode de Hansen-Hurwitz

S.G. PRABHU-AJGAONKAR[1]

RÉSUMÉ

La méthode de Hansen-Hurwitz (1943) est réputée moins efficace que la méthode de Horvitz-Thompson (1952) qui est liée à un certain nombre de méthodes d'échantillonnage avec PIPT (probabilité d'inclusion proportionnelle à la taille). Le présent article démontre de façon simple la supériorité de la seconde méthode et présente donc un intérêt au point de vue pédagogique.

MOTS CLÉS: Méthodes d'échantillonnage; probabilité d'inclusion proportionnelle à la taille; forme quadratique définie positive.

1. INTRODUCTION

Soit U une population finie constituée de N unités identifiables $[U_1, U_2, \ldots, U_N]$. Deux nombres X_i et Y_i se rattachent à U_i, la i-ième unité de la population; les valeurs de X_i sont connues tandis que celles de Y_i sont inconnues mais fixes. De façon générale, X_i est une mesure de la taille de U_i qui est fortement correlé avec Y_i.

Si nous utilisons la méthode de Hansen-Hurwitz (1943) pour estimer le total de la population $T_y = Y_1 + Y_2 + \ldots + Y_N$, nous devons choisir avec remise n unités de population avec probabilité proportionnelle à X_i et utiliser l'estimateur sans biais

$$t_{HH} = \frac{1}{n} \sum_{r=1}^{n} \frac{y_r}{p_r},$$

où $p_r = X_r / T_x$, $T_x = X_1 + X_2 + \ldots + X_N$, et y_r $(r=1, 2, \ldots, n)$ est le résultat du r-ième prélèvement. Étant donné que $\Sigma Z_i = 0$, il est facile de montrer que

$$\mathrm{Var}(t_{HH}) = \sum_{i=1}^{N} \frac{Z_i^2}{np_i}, \tag{1}$$

où $Z_i = Y_i - p_i T_y$, $i=1, 2, \ldots, N$.

Dans le cas d'un échantillonnage sans remise, Horvitz et Thompson (1952) ont proposé l'estimateur sans biais

$$t_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i},$$

[1] S.G. Prabhu-Ajgaonkar, Department of Mathematics and Statistics, Marathwada University, Aurangabad 431004, Inde.

où π_i ($i=1, 2, \ldots, N$) désigne la probabilité d'inclusion de la i-ième unité de population U_i dans l'échantillon. En outre, lorsque π_i est proportionnelle à X_i, nous parlons d'une méthode d'échantillonnage avec PIPT. Pour ce genre de méthode,

$$\operatorname{Var}(t_{HT}) = \sum_{i=1}^{N} \frac{Z_i^2}{np_i} + \sum_{i \neq j=1}^{N} Z_i Z_j \frac{\pi_{ij}}{n^2 p_i p_j} \tag{2}$$

où Z_i est défini en (1) et π_{ij} ($i \neq j=1, 2, \ldots, N$) est la probabilité (conjointe) d'inclusion des i-ième et j-ième unités de population dans l'échantillon. Lorsqu'on définit une méthode d'échantillonnage avec PIPT, π_{ij} peut être exprimée de façon plus simple.

Par (1) et (2),

$$\phi = \operatorname{Var}(t_{HT}) - \operatorname{Var}(t_{HH}) = \sum_{i \neq j=1}^{N} Z_i Z_j \frac{\pi_{ij}}{n^2 p_i p_j}. \tag{3}$$

2. COMPARAISON DES MÉTHODES

Pour estimer T_y, Midzuno (1952), Sen (1952) et Sankaranarayanan (1969) ont proposé des méthodes d'échantillonnage avec PIPT fondées sur l'estimateur d'Horvitz-Thompson t_{HT}. La méthode de Midzuno et Sen est applicable si

$$p_i = \frac{X_i}{T_x} > \frac{n-1}{n(N-1)}, \quad i=1, \ldots, N, \tag{4}$$

tandis que celle de Sankaranarayanan est applicable si

$$\sum_{j \in s} p_j > (n-1)/(N-1) \text{ pour tous } s \in S,$$

qui est une condition moins stricte.

Pour les deux méthodes, les probabilités conjointes d'inclusion sont définies par

$$\pi_{ij} = \frac{n(n-1)}{N-2}\left(p_i + p_j - \frac{1}{N-1}\right).$$

Ainsi, par l'équation (3),

$$\phi = \frac{n(n-1)}{n^2(N-2)}\left[\sum_{i=1}^{N} \frac{Z_i^2}{p_i}\left(2 - \frac{1}{(N-1)p_i}\right) + \frac{1}{(N-1)}\left(\sum_{i=1}^{N} \frac{Z_i}{p_i}\right)^2\right]. \tag{5}$$

L'expression ci-dessus est non-négative si

$$P_i > \frac{1}{2(N-1)}, \quad i=1, 2, \ldots, N,$$

et dans ce cas la méthode de Horvitz-Thompson est supérieure à la méthode de Hansen-Hurwitz. La condition définie en (4) a été établie pour la première fois par Rao (1963) lorsque $n=2$ et que la méthode de Midzuno-Sen est utilisée, mais il est intéressant de constater par l'équation (5) que la condition ne change pas même pour des valeurs de n supérieures à 2.

Chaudhuri (1975) et Mukhopadhyay (1975) ont, chacun de leur côté, obtenu les résultats ci-dessus pour la méthode de Midzuno-Sen.

Brewer (1963), Rao (1965) et Durbin (1967) ont proposé diverses méthodes d'échantillonnage avec PIPT pour $n=2$ avec les mêmes probabilités d'inclusion,

$$\pi_{ij} = \frac{2p_i p_j}{1+k}\left(\frac{1}{1-2p_i} + \frac{1}{1-2p_j}\right) \text{ où } k = \sum_{i=1}^{N} \frac{p_i}{1-2p_i}.$$

Ces méthodes ne sont pas assujetties aux conditions qui s'appliquent aux méthodes précédentes. Par (3),

$$\phi = \frac{1}{1+k}\sum_{i=1}^{N} \frac{Z_i^2}{1-2p_i} \geq 0,$$

de sorte que la méthode de Horvitz-Thompson est de nouveau supérieure à la méthode de Hansen-Hurwitz.

REMERCIEMENTS

L'auteur tient à exprimer sa reconnaissance au rédacteur en chef, M.P. Singh, et à l'arbitre pour leurs nombreux commentaires utiles.

BIBLIOGRAPHIE

BREWER, K.R.W. (1963). A model of systematic sampling with unequal probabilities. *Australian Journal of Statistics*, 5, 5-13.

CHAUDHURI, A. (1975). On some properties of the sampling scheme due to Midzuno. *Bulletin of Calcutta Statistical Association*, 23, 1-19.

DURBIN, J. (1967). Design of multi-stage surveys for the estimation of sampling errors. *Applied Statistics*, 16, 152-164.

HANSEN, M.H., et HURWITZ, W.N. (1943). On the theory of sampling from finite population. *Annals of Mathematical Statistics*, 14, 333-362.

HORVITZ, D.G., et THOMPSON, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.

MIDZUNO, H. (1952). On the sampling system with probability proportionate to sum of sizes. *Annals of Institute of Statistical Mathematics,* 3, 99-107.

MUKHOPADHYAY, P. (1975). PPS sampling schemes to base HTE. *Bulletin of Calcutta Statistical Association*, 23, 21-44.

RAO, J.N.K. (1963). On two systems of unequal probability sampling without replacement. *Annals of Institute of Statistical Mathematics*, 15, 67-72.

RAO, J.N.K. (1965). On two simple schemes of unequal probability sampling without replacement. *Journal of the Indian Statistical Association*, 3, 173-180.

SANKARANARAYANAN, K. (1969). An IPPS sampling scheme using Lahiri's method of selection. *Journal of the Indian Society of Agricultural Statistics*, 21, 58-66.

SEN, A.R. (1952). Further developments of the theory and application of primary sampling units with special reference to the North Carolina agricultural population. Thèse de doctorat, North Carolina State College, Raleigh.