The Local Pivotal Method and its Application on StatVillage Data

Diana Sokurova[1]

[1]University of Tartu, e-mail: diana.sokurova12@gmail.com

Abstract

The purpose of this paper is to give an overview of the local pivotal method and compare it with other well-known sampling methods, applied on the real data. In the theoretical part detailed description of the local pivotal method is given. In the practical part, Monte Carlo simulation is conducted to find out which sampling method gives better estimation for data coming from a hypothetical village StatVillage, which is based on real data.
Keywords: survey sampling, sample survey theory, statistical estimation, pivotal method, local pivotal method

1 Introduction

The aim of a probabilistic survey sampling is to find out the strategy and estimator-function that leads to best estimate of the population's parameter of interest. Very popular sampling methods are simple random sampling, for which all objects in population have the equal inclusion probabilities, stratified sampling, for which population is divided into strata by determined criteria and after that some sampling method is applied in every stratum separately. The stratified sampling method, with simple random sampling applied in each stratum, is called stratified random sampling. The method, where systematic sampling is applied in each stratum, is called systematic stratified sampling. systematic sampling is based on the selection of elements from an ordered sampling frame.

Here the novel sampling method is presented, called the local pivotal method. The local pivotal method is special case of the general pivotal method, what was introduced in (Deville & Tillé, 1998). The local pivotal method was created to achieve spatially balanced sample.(Grafström *et al.*, 2012)

2 Pivotal Method

The pivotal method allows using the unequal inclusion probabilities as the initial ones. Method updates the initial inclusion probabilities interactively so that in each step the inclusion probabilities are recalculated until they becomes equal 0 or 1. The value of 1 means that corresponded unit is in sample, the value of 0 - not in a sample.

To describe the updating rule,some notation have to be introduced. We denote possibly updated inclusion probabilities with π_i', and that the unit i is finished if $\pi_i' = 0$ or $\pi_i' = 1$. Once a unit is finished, it is not used in algorithm again. (Deville & Tillé, 1998)

Algorithm 1. Pivotal Method

1. Choose two units i ja j randomly, $i, j \in U$, where U is the finite population.
2. Update the vector of their inclusion probabilities by the following updating rule.

   (a) If $\pi_i + \pi_j < 1$, then

   $$(\pi_i', \pi_j') = \begin{cases} (0, \pi_i + \pi_j), \text{ with probability } \frac{\pi_j}{\pi_i+\pi_j} \\ (\pi_i + \pi_j, 0), \text{ with probability } \frac{\pi_i}{\pi_i+\pi_j}. \end{cases}$$

   (b) If $\pi_i + \pi_j \geq 1$, then

   $$(\pi_i', \pi_j') = \begin{cases} (1, \pi_i + \pi_j - 1), \text{ with probability } \frac{1-\pi_j}{2-\pi_i-\pi_j} \\ (\pi_i + \pi_j - 1, 1), \text{ with probability } \frac{1-\pi_i}{2-\pi_i-\pi_j}. \end{cases}$$

3. Repeat algorithm until all units are finished(i.e equal to 0 or 1).

Because of the randomly choosing procedure, this method is also called random pivotal method.

3 Local Pivotal Method

The local pivotal methods update the inclusion probabilities according to the updating rule described above, but for two nearby units at each step. There is two different ways to choose the two nearby units i and j. At the first way, it is required that two units are the nearest neighbors to each other (Algorithm 2). At the second way it is sufficed that only one of units is the nearest neighbor to another (Algoritm 3). (Grafström *et al.*, 2012)

Algorithm 2. Local Pivotal Method I

1. Randomly choose one unit i.
2. Choose unit j, a nearest neighbor to i. If two or more units have the same distance to i, then randomly choose between them with equal probability.
3. If j has i as its nearest neighbor, then update the inclusion probabilities according to the updating rule. Otherwise go to 1.
4. If all units are finished, then stop. Otherwise go to 1.

Algorithm 3. Local Pivotal Method II

1. Randomly choose one unit i.
2. Choose unit j, a nearest neighbor to i. If two or more units have the same distance to i, then randomly choose between them with equal probability.
3. Update the inclusion probabilities for the units i and j according to the updating rule.
4. If all units are finished, then stop. Otherwise go to 1.

4 Simulation

To compare Local Pivotal Method according with other sampling methods on the data from StatVillage (Schwarz, 1997) Monte-Carlo simulation was used. For each sampling method, 1000 samples were drawn and 1000 estimates were found. Then Monte-Carlo mean and Monte-Carlo standard error was calculated by following formulas:

$$E_{MC}(\hat{t}) = \frac{1}{1000}\sum_{k=1}^{1000} \hat{t}_k,$$

$$\sqrt{V_{MC}(\hat{t})} = \sqrt{\frac{1}{999}\sum_{k=1}^{1000}(\hat{t}_k, -E_{MC}(\hat{t}))^2},$$

where $\hat{t}_k$ stands for the estimate total t in the simulation step k, $k = 1, \dots, 1000$.

Simulations was done for two study variables: continuous variable household month income (moninch) , and discrete variable household size (hhsize). The block and the house numbers from address of household and number of income recipients were taken as the auxiliary information.

4.1 Results of Simulation

The result of Monte-Carlo simulation for continious variable are listed in the table 1

Table 1: Estimates and standard errors of variable *moninch*

Actual value	4 843 695	
Selection method	$E_{MC}(\hat{t})$	$\sqrt{V_{MC}(\hat{t})}$
Simple random sampling	4 838 280	152 718.57
Stratified random sampling	4 842 568	139 327.05
Systematic stratified sampling	4 845 962	58 380.02
Random pivotal method	4 844 994	156 132.98
Local pivotal method I	4 843 181	50 190.02
Local pivotal method II	4 844 111	49 183.08

Based on the table 1, the most accurate estimates are get from the systematic stratified sampling and both local pivotal methods where the difference in standard errors in local pivotal methods is small.

Below is given results for discrete variable for the various methods of selection.

Table 2: Estimates and standard errors of variable *hhsize*

Actual value	3 000	
Selection method	$E_{MC}(\hat{t})$	$\sqrt{V_{MC}(\hat{t})}$
Simple random sampling	2 999.64	73.22
Stratified random sampling	2 997.93	54.13
Systematic stratified sampling	3 002.18	51.15
Random pivotal method	3 001.38	71.17
Local pivotal method I	3 000.8	60.19
Local pivotal method II	3 001.44	58.42

Based on the table 2 , the most accurate estimate for standard error is derived from a systematic stratified sampling. Both local pivotal methods did not give the best or worst estimate, but the local pivotal method II is a bit more accurate than the local pivotal I.

In conclusion, in the case of a continuous variable, both local pivotal methods provide more accurate estimates, and in this case, it is recommended to use local pivotal method II because it is more accurate and faster in execution. In the case of a discrete variable, each of the local pivotal methods did not provide any better estimates, and in this case, it is advisable to use a systematic stratified selection.

References

Deville, J.-C. & Tillé, Y. (1998). Unequal probability sampling without replacement through a splitting method. *Biometrika* **85**, 89–101.

Grafström, A., Lundström, N. L. P. & Schelin, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics* **68**, 514–520.

Schwarz, C. (1997). Statvillage: An on-line, www-accessible, hypothetical city based on real data for use in an introductory class in survey sampling. *Journal of Statistics Education* **5**.