Survey Methodology, December 2006
Vol. 32, No. 2, pp. 157-163
Statistics Canada, Catalogue No. 12-001

Geometric Versus Optimization Approach to Stratification: A Comparison of Efficiency

Marcin Kozak and Med Ram Verma [1]

Abstract

In this paper, the geometric, optimization-based, and Lavallée and Hidiroglou (LH) approaches to stratification are compared. The geometric stratification method is an approximation, whereas the other two approaches, which employ numerical methods to perform stratification, may be seen as optimal stratification methods. The algorithm of the geometric stratification is very simple compared to the two other approaches, but it does not take into account the construction of a take-all stratum, which is usually constructed when a positively skewed population is stratified. In the optimization-based stratification, one may consider any form of optimization function and its constraints. In a comparative numerical study based on five positively skewed artificial populations, the optimization approach was more efficient in each of the cases studied compared to the geometric stratification. In addition, the geometric and optimization approaches are compared with the LH algorithm. In this comparison, the geometric stratification approach was found to be less efficient than the LH algorithm, whereas efficiency of the optimization approach was similar to the efficiency of the LH algorithm. Nevertheless, strata boundaries evaluated via the geometric stratification may be seen as efficient starting points for the optimization approach.

Key Words: Optimum stratification; Geometric Stratification; Numerical Optimization; Lavallée-Hidiroglou algorithm.

1. Introduction

Gunning and Horgan (2004) proposed a stratification algorithm based on a geometric progression. For the sake of simplicity, we will call this technique the "geometric approach to stratification," "geometric stratification," or just "geometric approach." The geometric stratification aims to equalize values of the coefficient of variation of a stratification variable within strata, based on the assumption that the variable is uniformly distributed within each stratum. Gunning and Horgan (2004) showed that their algorithm is much easier to implement and more efficient than the classical cumulative root frequency method (Dalenius and Hodges 1959) as well as the Lavallée and Hidiroglou (LH) algorithm (Lavallée and Hidiroglou 1988). Horgan (2006) compared the geometric stratification with the Dalenius and Hodges' (1959), Ekman's (1959), and Lavallée and Hidiroglou (1988) procedures; again, in their study the geometric stratification occurred to be the most efficient among the procedures compared. Gunning, Horgan and Yancey (2004) applied this method to stratify accounting populations.

Like the cumulative square root frequency method, the geometric approach is an approximate stratification technique, and hence the stratification points it provides may be quite far from optimum stratification points. On the other hand, there exist approaches, especially for univariate stratification, that lead to near-optimum stratification points. These approaches are based on the use of self-implemented algorithms or numerical optimization methods to provide strata boundaries (*e.g.*, Lavallée and Hidiroglou 1988; Lednicki and Wieczorkowski 2003; Kozak 2004). Such methods, however, usually require initial strata boundaries to start an optimization process; approximate stratification methods can be employed to find such initial points. Of course, initial strata boundaries should be of high quality, as their low quality may cause the optimization to provide a local minimum (Rivest 2002).

Many surveys deal with positively skewed study variables. If this is the case, it is important to take into account this attribute when stratifying a population. Many researchers have attempted to create stratification methods that would construct a so-called "take-all" stratum (*e.g.*, Glasser 1962; Hidiroglou 1986), from which all the elements are selected in the sample with probability 1. In stratified sampling, this is the best manner of dealing with positively skewed variables. Such methods are usually more efficient (certainly, only if a population is positively skewed) than stratification methods in which a take-all stratum is not constructed. A take-all stratum is not constructed in the geometric stratification (Gunning and Horgan 2004).

The aim of this paper is to compare the efficiency of the geometric stratification, proposed by Gunning and Horgan (2004), and two optimization approaches to stratification (Lavallée and Hidiroglou 1988; Lednicki and

1. Marcin Kozak, Department of Biometry, Warsaw Agricultural University, Nowoursynowska 159, 02-776 Warsaw, Poland. E-mail: marcin.kozak@omega.sggw.waw.pl; Med Ram Verma, Division of Agricultural Economics & Statistics, ICAR Research Complex for N.E.H. Region, Umroi Road, Umiam (Barapani) Meghalaya, India, Pin 793 103. E-mail: mrverma19@yahoo.co.in.

Wieczorkowski 2003; Kozak 2004), which is based on the use of numerical optimization methods.

2. Stratification Approaches Compared

Suppose we aim to stratify an $N-$element positively skewed population, U, based on an $N-$vector $\mathbf{x} = (x_1, \ldots, x_N)^T$ of values, known at the outset (*i.e.*, prior to the study), of a stratification variable X.

In this paper, we consider two stratification problems. In the first problem, L strata are to be constructed subject to a given sample size n. Suppose we are looking for an $(L+1)-$vector of strata boundaries $\mathbf{k} = (k_0, \ldots, k_L)^T$ $(k_0 < k_1 < \ldots < k_L$, k_0 being the minimum and k_L the maximum value of $X)$ that minimizes the variance of an estimator of the population mean of X under stratified sampling with simple random sampling without replacement within strata (*STSI*) sampling combined with a take-all stratum approach. (Note that we treat the stratification variable as identical to the corresponding survey variable.) The variance of $\bar{x}_{\text{st}}$ is given by

$$V(\bar{x}_{\text{st}}) = \sum_{h=1}^{L-1} \left(\frac{N_h}{N}\right)^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h},$$
$$\bar{x}_{\text{st}} = \sum_{h=1}^{L} \frac{N_h}{N} \bar{x}_h,\ \bar{x}_h = \frac{1}{n_h} \sum_{k=1}^{n_h} x_{kh}\ (h = 1, \ldots, L), \quad (1)$$

where n_h is the sample size from the h^{th} stratum, N_h is the size of the h^{th} stratum, S_h^2 is the population variance of X restricted to the h^{th} stratum, $\bar{x}_{\text{st}}$ is the estimator of the population mean of X under *STSI* sampling, $\bar{x}_h$ is the estimator of the population mean of X in the h^{th} stratum under simple random sampling without replacement (*SI*) sampling, and x_{kh} is the value of X for the k^{th} sample element of the h^{th} stratum and $h = 1, \ldots, L$.

The optimum sample allocation, which is in our problem obtained by minimizing the variance (1) subject to a given sample size n, is given by the following Neyman-optimum formula adjusted to a take-all stratum approach (Lednicki and Wieczorkowski 2003):

$$n_h = (n - N_L) \frac{N_h S_h}{\sum_{h=1}^{L-1} N_h S_h},\ h = 1, \ldots, L-1. \quad (2)$$

The geometric approach to stratification aims to equalize values of the coefficient of variation of X within the L strata. It simply consists of applying the following formula based on a geometric progression (Gunning and Horgan 2004)

$$k_h = ar^h,\ h = 0, \ldots, L, \quad (3)$$

where $a = \min(X)$, $k_L = \max(X)$, and $r = (k_L / k_0)^{1/L}$. The formula (3) is based on the assumption that X is uniformly distributed within each stratum.

The optimization approach applied to this particular stratification problem is based on the numerical optimization of the following problem: Minimize

$$f(\mathbf{k}) = V(\bar{x}_{\text{st}}), \quad (4)$$

where $V(\bar{x}_{\text{st}})$ is the variance (1) under the optimum allocation (2), subject to constraints

$$N_h \geq 2 \text{ and } 2 \leq n_h \leq N_h \text{ for } h = 1, \ldots, L-1, \quad (5)$$

and

$$\sum_{h=1}^{L-1} n_h = n - N_L. \quad (6)$$

Sometimes, when one wants to obtain more or less equal levels of precision of estimation in each stratum, a power allocation may be applied (Bankier 1988; Rivest 2002; Lednicki and Wieczorkowski 2003):

$$n_h = \frac{(n - N_L)(N_h \bar{x}_h)^P}{\sum_{h=1}^{L-1} (N_h \bar{x}_h)^p},\ p \in (0,1];\ h = 1, \ldots, L-1. \quad (7)$$

The optimization approach is more difficult to apply than the geometric stratification approach due in large part to the fact that the algorithm for the geometric approach is significantly more simplistic than for the optimization approach. An optimization method has to be chosen from among various available methods. Lednicki and Wieczorkowski (2003) used the simplex method of Nelder and Mead (1965); however, more efficient methods, which often require self-implemented algorithms (*e.g.*, Kozak 2004), can be applied, too.

Note that the geometric stratification does not take into account the formulae for the variance (1), the sample allocation (2), and the constraints (5). It may happen that one of the constraints (5) is not fulfilled. For these reasons, the geometric stratification is an approximate stratification procedure.

In this study, the algorithm proposed by Kozak (2004) was applied to stratify several populations. It is a random search algorithm adjusted to the problem of stratification. It is a simple algorithm; in each step, a stratum boundary is randomly selected and randomly changed. If the new set of strata boundaries is better than the previous one, the new one replaces the previous one. In the Appendix, the algorithm based on the paper by Kozak (2004) is given in detail.

The second problem considered in the paper is construction of strata that minimize a sample size from a population with respect to a given level of precision of estimation (the precision of estimation being given by the variance of an estimator of the population mean or total). The Lavallée-Hidiroglou (LH) algorithm (Lavallée and Hidiroglou 1988) can be seen as a particular optimization method to solve this particular stratification problem; it does not, however, work in other problems, *e.g.*, in the one considered earlier. For details of the algorithm, see the paper by Lavallée and Hidiroglou (1988). Besides the LH algorithm, the geometric stratification and random search method were applied to construct the strata.

The R language and environment (R Development Core Team 2005) was used to perform all the computation work in the present study.

3. Numerical Comparison of Efficiency of the Approaches in Stratification Under Fixed Sample Size

In this section, we compare two stratification approaches, the geometric stratification (geom) and optimization approach (optim), applied to a problem of searching for the strata boundaries that minimize the variance of the considered estimator with respect to a fixed sample size. In order to perform the comparison, five artificial populations of various sizes (from 2,000 to 10,000) were generated. Their summary statistics are presented in Table 1; the histograms of the stratification variables in the populations are given in Figure 1. In each case, the stratification variable was positively skewed (the skewness ranged between 1.40 for the 1st population to 5.02 for the 5th population). As it is usually the case in real populations, values of the stratification variables were integers. The sample size, n_i, from the i^{th} population was $n_i = f N_i$, where $f = 0.15$ is an assumed sample fraction and N_i is the size of the i^{th} population.

Table 1
Summary Statistics for Studied Artificial Populations

Population	Size	Range	Skewness	Mean	Variance
1	4,000	3–72	1.40	16.11	45.8
2	4,000	243–28,578	2.66	2,823.95	4.8×10^6
3	2,000	6–2,793	3.55	224.12	6.0×10^4
4	10,000	62–74,398	4.20	3,616.41	2.1×10^7
5	2,000	259–186,685	5.02	9,265.36	1.1×10^8

First, each population was stratified using the geometric stratification method into 4, 5, 6, and 7 strata. Then, the optimization approach was applied; as initial parameters in the optimization approach, the strata boundaries determined via the geometric stratification were used.

Like Gunning and Horgan (2004), to compare the efficiency of the two approaches, the relative efficiency was calculated via the formula:

$$\text{eff}_{\text{geom, optim}} = \frac{V_{\text{geom}}(\bar{x}_{\text{st}})}{V_{\text{optim}}(\bar{x}_{\text{st}})}, \tag{8}$$

where $V_{\text{geom}}(\bar{x}_{\text{st}})$ and $V_{\text{optim}}(\bar{x}_{\text{st}})$ are the variances (1) under the geometric and optimization approach, respectively. In addition, we calculated the coefficients of variation of the estimator of the population mean under both approaches:

$$\text{cv}_{\text{geom}} = \frac{\sqrt{V_{\text{geom}}(\bar{x}_{\text{st}})}}{\bar{x}_{\text{st}}}; \; \text{cv}_{\text{optim}} = \frac{\sqrt{V_{\text{optim}}(\bar{x}_{\text{st}})}}{\bar{x}_{\text{st}}}. \tag{9}$$

Table 2 contains the values of the relative efficiencies (8) and the coefficients of variation (9) for each combination studied (population × number of strata).

Table 2
Coefficients of Variation of the Estimator of the Population Mean Under the Geometric Stratification (CV_{geom}) and Optimization Approach (CV_{optim}), and Efficiencies of the Geometric Stratification Relative to the Optimization Approach ($\text{eff}_{\text{geom,optim}}$)

Number of strata L	CV_{geom}	CV_{optim}	$\text{eff}_{\text{geom,optim}}$
	Population 1		
4	0.0086	0.0056	1.53
5	0.0070	0.0042	1.66
6	0.0057	0.0034	1.66
7	0.0051	0.0029	1.75
	Population 2		
4	0.0116	0.0084	1.37
5	0.0095	0.0065	1.47
6	0.0085	0.0051	1.66
7	0.0073	0.0042	1.72
	Population 3		
4	0.0235	0.0133	1.76
5	0.0174	0.0100	1.74
6	0.0146	0.0081	1.80
7	0.0129	0.0067	1.91
	Population 4		
4	0.0104	0.0063	1.64
5	0.0089	0.0047	1.88
6	0.0073	0.0038	1.93
7	0.0064	0.0032	2.00
	Population 5		
4	0.0235	0.0134	1.76
5	0.0185	0.0100	1.86
6	0.0161	0.0080	2.00
7	0.0134	0.0074	1.82

In each case, the optimization approach was more efficient than the geometric stratification. The efficiency was smaller than 1.5 for only two combinations; in the rest of combinations, it ranged between 1.5 and 2. Usually, the more strata constructed the greater the gain in efficiency.

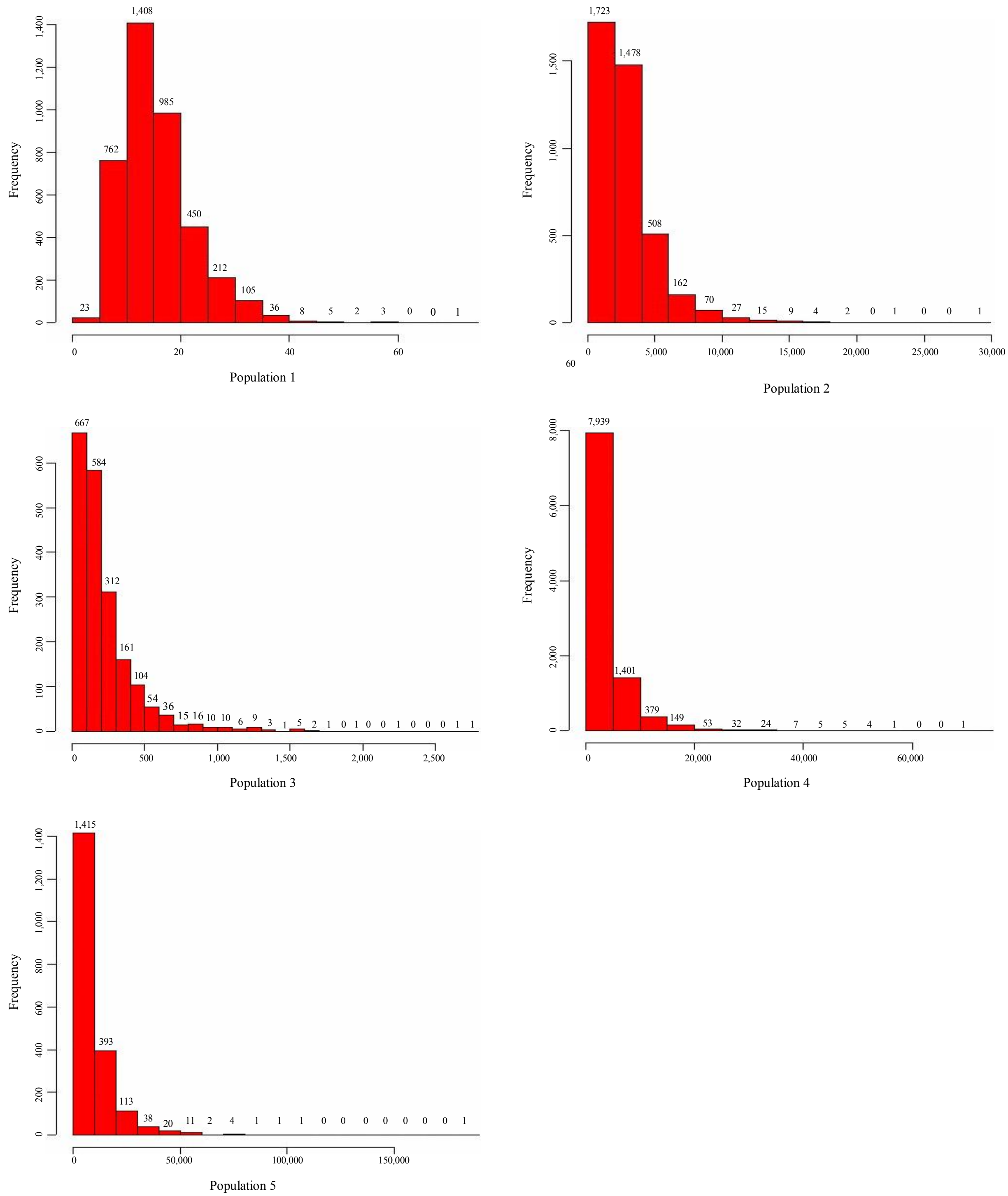

Figure 1. Histograms of stratification variable in studied artificial populations.

4. Numerical Comparison of Efficiency of the Stratification Approaches Under Fixed Level of Precision of Estimation

Gunning and Horgan (2004) and Horgan (2006) compared the geometric stratification with the Lavallée and Hidiroglou (Lavallée and Hidiroglou 1988) algorithm and found that the former was usually more efficient. In this section, we compare the three stratification approaches: the geometric stratification, the LH algorithm, and the optimization approach via a random search method. In this study, the same five populations as in the previous section were used (see Table 1 and Figure 1).

The relative efficiencies of two approaches were evaluated as

$$\text{eff}_{i,j} = \frac{n_i(\text{cv})}{n_j(\text{cv})}, \quad (10)$$

where i and j are the indices of the stratification approaches (i, j = geom, optim, LH), and $n_i(\text{cv})$ and $n_j(\text{cv})$ are the minimum sample sizes required to obtain a desired level of precision (cv) under the i^{th} and j^{th} approaches, respectively.

Using the three approaches, each population was stratified into $L = 4, \ldots, 7$ strata; the required level of precision was 0.01 in each case. Minimum sample sizes required for this level of precision and relative efficiencies (10) are given in Table 3.

Table 3
Minimum Sample Sizes Required to Obtain a Value Equal to 0.01 for the Coefficient of Variation of the Estimator of the Population Mean, Under the Geometric Stratification (n_{geom}), Optimization Approach (n_{optim}), and LH Algorithm (n_{LH}); and Efficiencies of the Geometric Stratification Relative to the Optimization Approach ($\text{eff}_{\text{geom, optim}}$), the Geometric Stratification Relative to the LH Algorithm ($\text{eff}_{\text{geom, LH}}$), and LH Algorithm Relative to the Optimization Approach ($\text{eff}_{\text{LH, optim}}$)

Number of strata L	n_{geom}	n_{optim}	n_{LH}	$\text{eff}_{\text{geom, optim}}$	$\text{eff}_{\text{geom, LH}}$	$\text{eff}_{\text{LH, optim}}$
			Population 1			
4	805	496	496	1.63	1.63	1.00
5	613	344	344	1.78	1.78	1.00
6	460	252	252	1.83	1.83	1.00
7	357	192	192	1.86	1.86	1.00
			Population 2			
4	483	248	259	1.94	1.86	1.04
5	329	154	163	2.14	2.02	1.06
6	224	113	117	1.98	1.92	1.03
7	180	83	83	2.17	2.17	1.00
			Population 3			
4	782	410	411	1.91	1.90	1.00
5	601	303	304	1.98	1.98	1.00
6	495	242	241	2.04	2.05	1.00
7	422	195	195	2.11	2.16	1.00
			Population 4			
4	839	409	409	2.05	2.05	1.00
5	650	301	301	2.15	2.15	1.00
6	552	240	242	2.30	2.28	1.01
7	–[1]	200	200	–	–	1.00
			Population 5			
4	1,768	894	894	1.98	1.98	1.00
5	1,274	628	628	2.03	2.03	1.00
6	949	459	459	2.07	2.07	1.00
7	758	355	355	2.13	2.13	1.00

[1] There were numerical problems with obtaining stratum boundaries (sample sizes from some strata were bigger than the sizes of these strata).

From the results it follows that the optimization approach was more efficient than the geometric stratification; this outcome was obtained for each population and number of strata. The relative efficiency was always greater than 1.6. Moreover, an interesting conclusion follows from the comparison of the efficiency of the geometric and LH stratifications. As already mentioned, Gunning and Horgan (2004) and Horgan (2006) found the geometric stratification more efficient than the LH algorithm. On the contrary, in our study, the LH algorithm was always more efficient than the geometric stratification. This situation occurred also for other generated populations of various sizes and skewness (results not included in this paper). Nevertheless, we do not state that the LH algorithm is always more efficient than the geometric stratification. It may happen that the geometric stratification will be better, as Gunning and Horgan (2004) and Horgan (2006) obtained in their studies.

From the comparison of the LH algorithm and the optimization approach it follows that both approaches provides stratification points leading to similar sample sizes. In some cases, the LH stratification was slightly better and in some other cases slightly worse than the optimization approach. Nevertheless, these differences do not mean that we could indicate either of these two approaches as more efficient. In fact, these two approaches have the same aim (in this particular stratification problem) and they just differ in the algorithm to achieve this aim. In summary, on the basis of our results we conclude that, in general, the LH stratification and optimization approach are more efficient than the geometric stratification.

5. Conclusions

The stratification technique based on a geometric progression proposed by Gunning and Horgan (2004) has a significant advantage; namely, its algorithm is very simple to implement compared to the cumulative square root of frequency method of Dalenius and Hodges (1959) and to other stratification methods. It is, however, an approximate stratification procedure, so the stratification points it provides may lead to poor precision of estimation (or a large sample size required to achieve a required level of precision). Furthermore, it is likely that some of the strata constructed will not fulfill the constraints (5); *e.g.*, some strata may be empty (so they would not comprise any population element) or/and sample sizes from some strata may be smaller than two or greater than their population sizes.

In our study, the optimization approach (via the LH and random search algorithms) was more efficient than the geometric stratification for each population studied and number of strata constructed. Nevertheless, the strata boundaries provided by the geometric stratification can be seen as efficient initial parameters required in the optimization approach; they should not be considered, however, as the optimal or efficient strata boundaries. Furthermore, our results conclusively show that the geometric stratification is less efficient than the stratification presented by Lavallée and Hidiroglou (1988), which is the result opposite to the one obtained by Gunning and Horgan (2004) and Horgan (2006). This problem needs further studies on real skewed populations; investigations on artificial populations univocally show that the LH algorithm and the optimization approach are more efficient than the geometric stratification.

At first look, one could be surprised that the gain in efficiency after applying the LH and optimization approaches compared to the geometric stratification increases after increasing the number of strata. This can be easily explained. The aim of the geometric stratification is to equalize cvs of the stratification variable within the strata. Therefore, this is not the same aim as the aim of stratification, which is to optimize the efficiency of estimation or to minimize a sample size. Furthermore, there is no certainty that under the optimum stratification the distribution of the stratification/survey variable is uniform within the strata. These two sets of strata boundaries (*i.e.*, provided by the geometric and optimization approaches) are not necessarily the same; in fact, they are likely different.

Note that we applied the random search method as the algorithm of the optimization approach to stratification. In fact, Lavallée and Hidiroglou's (1988) algorithm is a representative of optimization approaches, too. When the aim of stratification is to minimize a sample size required to achieve a desired level of precision, the two approaches will likely provide similar results, as they did in our study. Nevertheless, the random search algorithm may be applied to any stratification problem (*i.e.*, any optimization function and its constraints), contrary to the LH algorithm, which is applicable only when a sample size is minimized with respect to a given level of precision. It is to be noted that the random search algorithm, as a global optimization method, provides random results.

Our aim, however, was not to promote any of these two algorithms by showing that they are more efficient than the geometric stratification. In addition, we applied Nelder and Mead's (1965) simplex method to stratify the populations (results not presented in the paper); its results were very similar to those of the LH and random search method algorithms. Each of these methods has some drawbacks. For instance, numerical difficulties may occur while using the LH algorithm (Slanta and Krenzke 1996); the random search method provides random results (Kozak 2004); Nelder and Mead's (1965) method may be inefficient under large number of strata and large populations (Kozak 2004); and, in fact, none of the methods has been proven to provide optimum stratification points. Therefore, there is still a need of constructing a stratification algorithm that would be optimum irrespective of the situation (*e.g.*, of a population size or variable's skewness) and that would provide results that are not random. Our main aim was to prove that the geometric stratification is not optimum, although the stratification points it provides may be useful as initial parameters in other approaches to stratification.

Acknowledgements

The authors are very indebted to the referees and the Associate Editor of *Survey Methodology* for their valuable comments, which helped to improve the first version of this paper.

Appendix

The algorithm given below was proposed by Kozak (2004); we have adapted some of its details to the general stratification problem. In the algorithm, we do not refer to the particular problem of stratification (*i.e.*, we do not define the optimization function and its constraints), since the algorithm works for both problems presented in the paper as well as for other stratification problems. Where required, we refer to "optimization function" (which may be either the variance of an estimator considered or a sample size from a population) and "constraints" (which, depending on the optimization function, may be the constraints (5) and (6), or the constraints (5) combined with the constraint on the level of precision of estimation); certainly, other forms of the optimization function and its constraints may be considered as well.

Let us define a vector **a** as follows. It takes values on the interval $(1, N)$, N being the population size. Provided that a population is sorted by the values of a stratification variable X, two elements a_{h-1} and a_h of the vector **a** define the stratum h in such a way that this stratum consists of the elements with the index I (which gives the order of an element in the population sorted) that $a_{h-1} < I \leq a_h$, $h = 1, \ldots, L, a_0 = 0, a_L = N.$ The algorithm is as follows.

1. Sort the population by the values of the stratification variable.
2. Choose an initial vector **a**, *i.e.*, the vector of initial strata boundaries. You may use random integers that satisfy the constraints, but practice shows that better results may be achieved by using approximate strata boundaries obtained via some approximate stratification methods. Calculate the value of the optimization function. Check the constraints; if they are not fulfilled, the initial points have to be changed.

3. For $r = 0, 1, \ldots, R$ repeat the following step:

   a. Generate point $\mathbf{a}'$ by drawing one stratum boundary a_i and changing it as follows

$$a_i' = a_i + j,$$
$$a_k' = a_k \quad \text{for } k = 1, \ldots, L-1, k \neq i, \tag{11}$$

   where j is the random integer, $j \in \langle -p; -1 \rangle \cup \langle 1; p \rangle$, p being a given integer chosen based on the population size (the larger the population, the larger the p value); usually, it should be between 3 and 5.

   b. Calculate the value of the optimization function.

   c. If the constraints are satisfied and the value of the optimization function under the vector $\mathbf{a}'$ is smaller than the value under the vector $\mathbf{a}$, accept the new vector, *i.e.*, $\mathbf{a}_{r+1} = \mathbf{a}'$ (where $\mathbf{a}_{r+1}$ is the vector of strata boundaries in a next iteration); otherwise do not accept the vector, *i.e.*, $\mathbf{a}_{r+1} = \mathbf{a}$.

4. Finish the algorithm if the stopping rule is fulfilled, *e.g.*, if $r = R$, where R is given number of steps, or if in the last m (for instance, 50) steps the value of the optimization function did not change. Finally, calculate the vector $\mathbf{k}$ (the vector of final strata boundaries) on the basis of the values of the vector $\mathbf{a}$.

References

Bankier, M.D. (1988). Power allocations: Determining sample sizes for subnational areas. *The American Statistician*, 42, 174-177.

Dalenius, T., and Hodges, J.L. (1959). Minimum variance stratification. *Journal of the American Statistical Association*, 54, 88-101.

Ekman, G. (1959). An approximation useful in univariate stratification. *Annals of Mathematical Statistics*, 30, 219-229.

Glasser, G.J. (1962). On the complete coverage of large units in a statistical study. *Review of the International Statistical Institute*, 30, 28-32.

Gunning, P., and Horgan, J.M. (2004). A simple algorithm for stratifying skewed populations. *Survey Methodology*, 30, 159-166.

Gunning, P., Horgan, J.M. and Yancey, W. (2004). Geometric stratification of accounting data. *J. de Contaduria y Administration*, 214, septiembre-diciembre.

Hidiroglou, M.A. (1986). The construction of a self-representing stratum of large units in survey design. *The American Statistician*, 40, 27-31.

Horgan, J.M. (2006). Stratification of skewed populations: A review. *International Statistical Review*, 74(1): 67-76.

Kozak, M. (2004). Optimal stratification using random search method in agricultural surveys. *Statistics in Transition*, 6(5), 797-806.

Lavallée, P., and Hidiroglou, M. (1988). On the stratification of skewed populations. *Survey Methodology*, 14, 33-43.

Lednicki, B., and Wieczorkowski, R. (2003). Optimal stratification and sample allocation between subpopulations and strata. *Statistics in Transition*, 6, 287-306.

Nelder, J.A., and Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7, 308-313.

R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; URL http://www.R-project.org.

Rivest, L.-P. (2002). A generalization of Lavallée and Hidiroglou algorithm for stratification in business surveys. *Survey Methodology*, 28, 191-198 (http://www.mat.ulaval.ca/pages/lpr/).

Slanta, J., and Krenzke, T. (1996). Applying the Lavallée and Hidiroglou method to obtain stratification boundaries for the Census Bureau's annual Capital Expenditure Survey. *Survey Methodology*, 22, 65-75.