Comparison of Selection Methods for Lattice Sampling

João Gabriel Malaguti

Abstract: Most of probabilistic sampling theory focuses on one-dimensional populations, however there exist cases in which it is necessary to consider two-dimensional populations. Lattice sampling works by selecting values in both dimensions and obtaining the sampling by doing their cross product. The selections methods of simple random sampling and systematic sampling used are well studied, though the problems regarding precision loss in systematic sampling when periodicities are present have not been studied for these populations. Furthermore, the combination of methods (using simple random sampling in one selection and systematic sampling in another) has not been studied either. This paper uses Monte Carlo simulations techniques to compare the different sampling plans for a numbers of populations, finding indications that for non-periodic populations any lattice sampling plan that uses systematic sampling has higher precision, though for periodic populations the inclusion of systematic sampling causes lower precision in estimation.

Keywords: Cross-classified sampling; Monte Carlo simulation; Design effect.

Most of probabilistic sampling theory focuses on one-dimensional populations, however there exist cases in which it is important to consider two-dimensional populations (BETHLEHEM, 2009). Such populations can be natural (like areas), or constructed by the researchers through combination of two variables of interest, like places of purchases and products for the estimation of price indexes (DALÉN & OHLSSON, 1995) or even days and maternity hospitals, like in the French longitudinal study of infancy ELFE (JUILLARD, CHAUVET & RUIZ-GAZEN, 2015).

Originally created to select units of areas (QUENOUILLE, 1949) and later expanded to sample time and space (VOS, 1964), two-dimensional sampling is a group of sampling plans capable of considering the multidimensionality of the population. Among the methods in this class, two are used when the variables to be samples lack categories or strata: cluster sampling in two stages and lattice sampling (also known as cross-classified sampling). In this paper, the focus is on the latter.

Let a two-dimensional population *D* with *r* rows and *c* columns, such that $N = l \times r$, from which we want to select a sample with *n* elements. Lattice sampling consists in randomly selecting n_r rows and n_c columns. The sample is the cross product of the two sets such that $n = n_r \times n_c$ (JUILLARD, 2016).

The selections methods used are well studied, with results from Bellhouse (1977) proving that the optimum plan involves systematic sampling, though Dunn & Harrison (1993) point out that the problems regarding precision loss in systematic sampling when periodicities exist have not been studied for two-dimensional populations. Furthermore, the combination of methods (using one in the first selection and the other in the second) has not been studied either. This is because much of the literature (OHLSSON, 1996 and SKINNER, 2015, for example) focuses on the creation of variance and standard error estimators for the two-dimensional group and such combinations would introduce more complications.

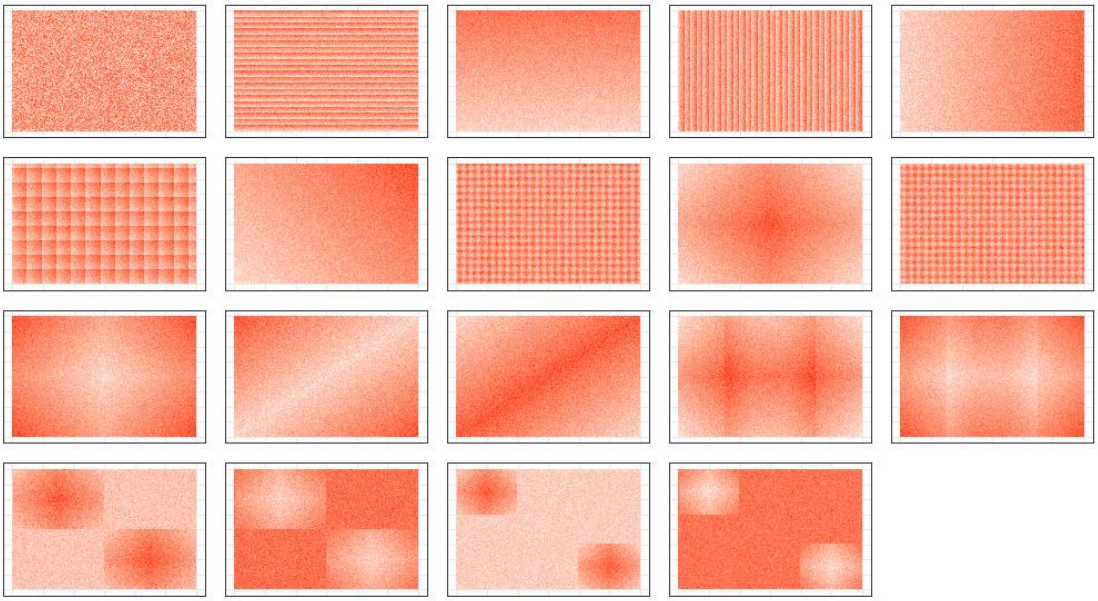
To analyse the behaviour of estimators under different sampling plans, a standard measure is needed. In sampling, there exists a metric with the objective of comparing the efficiency of alternate sampling plans: Kish's design effect (deff) (SKINNER, 1989). A relatively simple way of estimating the design effect is computationally.

Simulation studies, such as Monte Carlo (MC) simulations, are computational statistical methods based on generating different independent samples to obtain approximate answers to stochastic problems (BRANDIMARTE, 2014). The precision of this method is related to the number of samples generated (iterations), such that the higher is the number of samples, the more precise the estimates are. Because sampling plans, like the ones presented before, are by definition probabilistic it follows that we can use MC simulations to reach approximate results.

Monte Carlo simulation also permits the estimation of the standard error (and, therefore, the design effect) without need for a formal equation, placing the plans with combined selection methods on the same level as the plans with unique selection methods, for which approximate equations exist (SKINNER, 2015). For this study, only four combinations were analysed: pure simple random sampling (1), simple random sampling and systematic sampling (2), systematic sampling and simple random sampling (3); and pure systematic sampling (4).

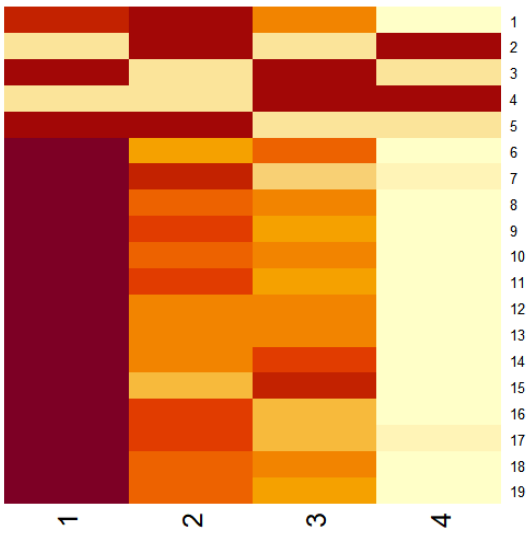
In order to compare the four plans, 19 populations with similar means and variances but different patterns in their distribution were created (Figure 1). From these, another two sets of populations were made, one randomizing its rows and the other, its columns.

Figure 1 – Heatmaps of the 19 populations



The simulation study consisted of taking a total of 100,000 samples for each population and sampling plan and estimating the standard errors through Monte Carlo. These values were then used to calculate the composite design effects using the lattice sampling with simple random sampling selection as the baseline. Plotting these results in a heatmap (Figure 2), controlling by population, we can see that apart from populations 2 and 4 which are highly periodical the lattice sampling with systematic sampling had the lowest values (lighter colours) and therefore the highest precisions, though the plans with combined selection also performed better than the baseline plan for most populations.

Figure 2 – Heatmap of design effect by sampling plan and population



BELLHOUSE, D. R. Some Optimal Designs for Sampling in Two Dimensions. Biometrika, v. 64, n. 3, p. 605–611, 1977.

BETHLEHEM, J. Applied Survey Methods: a statistical perspective. New Jersey: John Wiley & Sons, 2009.

BRANDIMARTE, P. Handbook in Monte Carlo Simulation – Applications in Financial Engineering, Risk Management, and Economics (1a ed.). Hoboken: John Wiley & Sons, 2014.

DALÉN, J.; OHLSSON, E. Variance Estimation in the Swedish Consumer Price Index. Journal of Business & Economic Statistics, v. 13, n. 3, p. 347–356, 1995.

DUNN, R.; HARRISON, A. R. Two-Dimensional Systematic Sampling of Land Use. Journal of Royal Statistical Society., v. 42, n. 4, p. 585–601, 1993.

JUILLARD, H. Two-dimensional sampling in practice. Case Studies in Business, Industry and Government Statistics, v. 6, n. 1, p. 36–49, 2016.

JUILLARD, H.; CHAUVET, G.; RUIZ-GAZEN, A. Estimation under cross-classified sampling with application to a childhood survey. Journal of the American Statistical Association., p. 1–24, 2015.

KISH, Leslie. Survey Sampling, 1ª ed. New York: Wiley-Blackwell, 1965.

OHLSSON, E. Cross-Classified Sampling. Journal of Official Statistics, v. 12, n. 3, p. 241–251, 1996.

QUENOUILLE, M. H. Problems in Plane Sampling. The Annals of Mathematical Statistics, v. 20, n. 3, p. 355–375, 1949.

SKINNER, C. J. Introduction to Part A. In: SKINNER, C. J.; HOLT, D.; SMITH, T. M. F. eds. Analysis of Complex Surveys. 1ª ed. Chichester: John Wiley & Sons, 1989.

SKINNER, C. J. Cross-classified sampling: Some estimation theory. Statistics and Probability Letters, v. 104, p. 163–168, 2015.

VOS, J. W. E. Sampling in Space and Time. Review of International Statistical Institute, v. 32, n. 3, p. 226–241, 1964.