Spatially Balanced Sampling

Jennifer Brown[1], Blair Robertson[1], Trent McDonald[2]

[1] University of Canterbury, New Zealand
[2] Western Ecosystems Technology Inc, USA

E-mail for correspondence: `Jennifer.Brown@canterbury.ac.nz`

Abstract: Spatially balanced sampling is an emerging area in statistical sampling. These designs are popular because they are one way to ensure the selected sample has spatial coverage over the entire survey area. This feature of spatial coverage aids in the resultant sample being representative of the population of interest.
One of the first and the most commonly used spatially balanced design is called GRTS (Generalized Random Tessellation Stratified sampling) where sample effort is spread evenly over the target region. The term spread evenly in this context means having coverage of survey effort over the region. The coverage from GRTS has a stochastic component rather than a fixed interval, regularly spaced coverage as in a systematic sampling design.
We have extended the idea of GRTS to a new design called Balances Acceptance Sampling (BAS). The BAS design allows surveys to be balanced in dimensions higher then two (n - dimensional space). Until now, most designs have considered balance in 2-D geographic space. With BAS we can achieve balance in 3-D space, or in higher dimensions. In some applications these dimensions can be features other than the spatial measures of geographic location, and the design allows aspects such as time for repeat surveys to be incorporated into sample balance.

Keywords: Sample; Environmental Sampling; Spatial Sampling.

1 Introduction

Spatial sampling is a broad category referring to designs that incorporate spatial reference in site selection. The purpose of sampling is usually to estimate a population parameter, such as the mean or total of some characteristic of interest. For example, in environmental applications surveys may be to estimate the density or abundance of a plant species of interest. In social science applications the survey may be to estimate household

---

This paper was published as a part of the proceedings of the 32nd International Workshop on Statistical Modelling (IWSM), Johann Bernoulli Institute, Rijksuniversiteit Groningen, Netherlands, 3–7 July 2017. The copyright remains with the author(s). Permission to reproduce or extract any parts of this abstract should be requested from the author(s).

income within a city. Interest in spatial sampling has increased because of the availability of georeferenced data, computation capacity, and the overarching need for surveys to be cost-efficient.
The gold standard of survey designs is simple random sampling (Cochran 1977), where selection of individual units in the sample is at random. In the above examples, individual units could be a quadrat or plot within which plants are counted, or, individual households from which household income is recorded. A spatial sample would appear very similar to a simple random sample except that the selection process would use information on the spatial location of the sample units (Wang et al. 2012).
A feature of many natural and social systems is that characteristics observed in one area are likely to be similar to the ones observed in an adjacent area. For example, the counts of plants in one plot will be similar to the counts in an adjacent plot (Grafström et al. 2012, Stevens and Olsen 2004). This is because of the underlying biotic and abiotic processes that drive the species distribution. In human populations individual neighbourhoods tend to have more similar household incomes than are observed among neighbourhoods, because of socio-economic factors. One way to view this is that there is limited new information provided from a sample unit that is spatially adjacent to another unit that has already been measured. This motivated development of survey designs where the sample is evenly, or near to evenly, spread over the study area. A design that generates samples that are well-spread over the population is called a spatially balanced survey design.
There are many different methods of spatially balanced survey designs (Wang et al. 2012). The design that stimulated the most interest in spatially balanced sampling, and began the use of this phrase, is Generalized Random Tessellation Stratified sampling (GRTS), developed by Stevens and Olsen (2003). In this design an invertible mapping technique is used to transform two-dimensional space into one-dimensional space. Then, a systematic sample is selected along the linear representation. Sampling location geo-references are generated from selecting points at regular intervals in this one-dimensional space (Brewer and Hanif 1983). The one-dimensional space is then mapped back to the two-dimensional original space. By maintaining the spatial properties of the original units, the resultant sample is spatially balanced, with neither no one area being over-represented with high sample intensity nor under-represented with low sample intensity.

2 Other spatially balanced designs

Local Pivotal Method (LPM) is a design for spatially balanced sampling (Grafström et al. 2012). The method is based on a method introduced by Deville and Till (1998). The algorithm involves sequentially updating

each point's inclusion probability. Starting with N points, neighbouring points compete to be included in the sample. The winning point of each competition has its inclusion probability increased and the losing point has its decreased. Eventually n points have inclusion probabilities of one and N-n points have inclusion probabilities of zero. The resultant sample will have points separated spatially because it is unlikely two adjacent points will be included in the final sample of size n. The design can be computationally heavy with large N, and an alternative method, suboptimal LPM, was suggested for these situations (Grafström et al. 2014), where only a subset of neighbours are used for the comparison, rather than all possible points.
Balanced acceptance sampling, BAS, (Robertson et al. 2013) is a more recent design for spatially balanced sampling. The method uses the Halton sequence, a quasi-random number sequence (Halton 1960). In two-dimensional geographic space, Halton points are used to generate the georeferenced locations of the sample units (starting from a randomly chosen position in the sequence). The design uses acceptance/rejection sampling to select sample units. If a generated sample point is beyond the edge of the sample space the sample unit is rejected, otherwise it is accepted. The design is straightforward computationally, and has better spatial balance than the comparable GRTS design (Robertson et al. 2013).
The algorithm can be extended into more than two dimensions (e.g., up to five), and this is an appealing feature for some survey situations. One application is for surveys that involve repeat samples of the same population. Environmental monitoring is an example, where interest is in how population parameters change over time. This involves repeat visits to the survey area, often on an annual basis. Other examples can be in social economic surveys where there is interest in how indicators are changing over time. In these situations, the design can be viewed as having three dimension, the two dimensional georeferenced space and the third dimension that is time. Three dimensional spatial balance can be thought of as a way of ensuring that for any one survey there is spatial balance, and no one area is excessively over- or under-sampled. In addition, over the course of the repeat surveys (e.g., annual surveys for 10 years), there is no one area that is excessively repeatedly over - or under-sampled. Until now many survey designs rely on fixed intervals between site revisits, and here BAS offers a method for randomising the repeat interval.

Acknowledgments: Special Thanks to Peter Jaksons and Naeimeh Abi for their involvemnt in ongoing research, and helpful suggestions.

References

Brewer, K.R.W. and Hanif, M. (1983). *Sampling with Unequal Probabilities. Lecture Notes in Statistics.* New York: Springer-Verlag.

Cochran, W.G. (1977). *Sampling techniques.* 3rd edition. John Wiley Sons.

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

Grafström , A, Saarela, S. and Ene, L. T. (2014).Efficient strategies for forest inventories by spreading the sample in auxiliary space. *Canadian Journal of Forest Research*, **44**, 1156 – 11164.

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

Halton, J.H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematic*, **2**, 84 – 90.

Robertson, B.L., Brown, J.A., McDonald, T. and Jaksons, P. (2013). BAS: Balanced acceptance sampling of natural resources. *Biometrics*, **69**, 776 – 784.

Stevens, D.L. and Olsen, A.R. (2003). Variance estimation for spatially balanced samples of environmental resources. *Environmetrics*, **14**, 593 – 610.

Stevens, D.L. and Olsen, A.R. (2004). Spatially balanced sampling of natural resources. *Journal of the American Statistical Association*, **99**, 262 – 278.

Wang, J.F., Stein, A., Gao, B.B. and Ge Y. (2012). A review of spatial sampling. *Spatial Statistics*, **2,** 1 – 14.