Spatially Balanced Sampling using the Halton Sequence

Blair Robertson, Jennifer Brown, Trent McDonald and Chris Price

School of Mathematics and Statistics, University of Canterbury and Western EcoSystems Technology

Introduction

A spatial sampling design determines where sample locations are placed in a study area. The main objective is to draw sample locations in such a way that valid scientific inferences can be made to all regions of a study area. A common feature in natural resource sampling is that nearby locations tend to be similar because they interact with one another and are influenced by the same set of factors. This means sample efficiency can be increased by spreading sample locations over the study area. Stevens and Olsen (2004) called this approach spatially balanced sampling.

Spatial Balance

Let $\pi(\boldsymbol{x}) = nf(\boldsymbol{x})$ be an inclusion density function, where $f(\boldsymbol{x}) : [0,1)^2 \to \mathbb{R}_{\geq 0}$ is a bounded probability density function. For a sample, $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \subset [0,1)^2$, the Voronoi polygon for $\boldsymbol{x}_i$ is

$$\omega_i = \{\boldsymbol{x} \in [0,1)^2 : \|\boldsymbol{x} - \boldsymbol{x}_i\| \leq \|\boldsymbol{x} - \boldsymbol{x}_j\| \text{ for all } j = 1, 2, \ldots, n\}.$$

A sample of size n is spatially balanced if

$$v_i = \int_{\omega_i} \pi(\boldsymbol{x}) d\boldsymbol{x} \approx 1,$$

for $i = 1, 2, \ldots, n$.

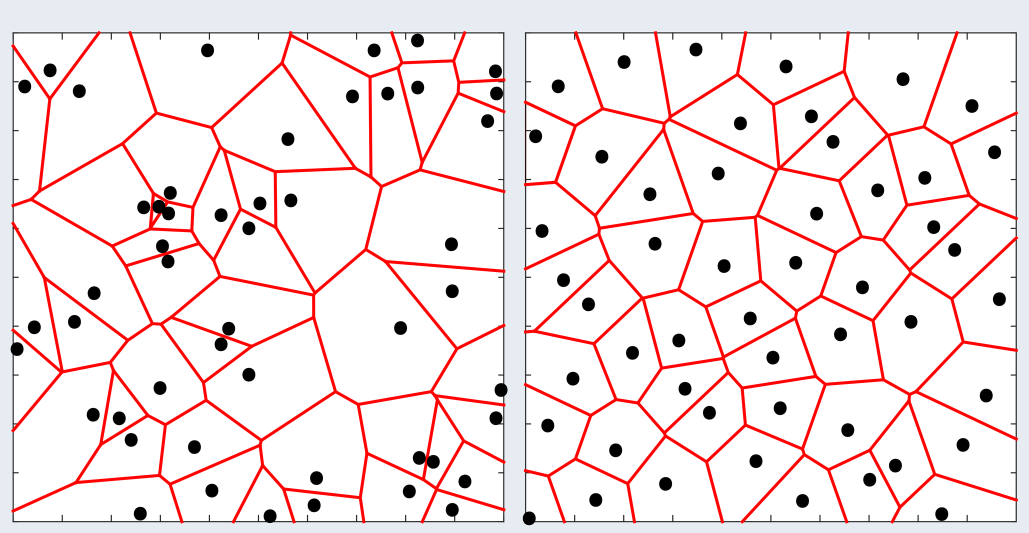

Figure 1:(Left) a simple random sample (SRS) of $n = 50$ points from $[0,1)^2$. (Right) an equal probability BAS sample of $n = 50$ points from $[0,1)^2$. In this case, v_i is proportional to the area of ω_i (shown in red). BAS has far better spatial balance than SRS because the areas of each ω_i are more similar in size.

Halton Sequence

The random-start Halton sequence $\{\boldsymbol{x}_j\}_{j=1}^{\infty}$ in $[0,1)^d$ is defined as follows. The ith coordinate of the jth point in this sequence is

$$x_j^{(i)} = \sum_{p=0}^{\infty} \left\{ \left\lfloor \frac{u_i + j}{b_i^p} \right\rfloor \bmod b_i \right\} \frac{1}{b_i^{p+1}},$$

where u_i is a random non-negative integer, b_i is a positive integer and $\lfloor \cdot \rfloor$ is the floor function. The bases b_i are chosen to be small, co-prime integers to ensure the points are evenly spread over the unit box. The random-start Halton sequence is

$$\{\boldsymbol{x}_j\}_{j=1}^{\infty} = \left\{ \left(x_j^{(1)}, x_j^{(2)}, \ldots, x_j^{(d)} \right) \right\}_{j=1}^{\infty}. \quad (1)$$

Let $B = \Pi_{i=1}^{d} b_i^{J_i}$, where J_i is any non-negative integer. It can be shown [3, 4] that B consecutive points from (1) will have exactly one point in each of the Halton boxes defined by

$$\prod_{i=1}^{d} \left[m_i b_i^{-J_i}, (m_i + 1) b_i^{-J_i} \right), \quad (2)$$

where m_i is an integer satisfying $0 \leq m_i < b_i^{J_i}$, for all $i = 1, \ldots, d$. The sequence is also quasi-periodic (period B) because points of the form $\boldsymbol{x}_{j+\alpha B}$ with $\alpha = 0, 1, \ldots,$ are in the same box.

BAS and HIP

Balanced acceptance sampling (BAS) [2] is a spatially balanced sampling design that draws its sample using the random-start Halton sequence. Consider drawing n sample locations from an areal resource $\Omega \subset [0,1]^2$ with $\lambda(\Omega) > 0$, where λ is the Lebesgue measure. An equal probability BAS sample is simply the first n points from (1) that fall within Ω. If $\boldsymbol{x}_1 \notin \Omega$, discard the sequence and generate another [3]. Unequal inclusion density functions can also be utilized with an acceptance rejection sampling technique [2].

Halton iterative partitioning (HIP) [4] extends BAS to point resources. It partitions the resource into $B \geq n$ boxes that have the same nested structure as (2), but different sizes. These boxes are then uniquely numbered using a random-start Halton sequence of length B. The HIP sample is obtained by randomly drawing one point from each of the boxes numbered $1, 2, \ldots, n$. Unequal probability samples can be drawn by altering the inclusion probability of each box.

Halton Sequence, Halton Boxes and a HIP Partition

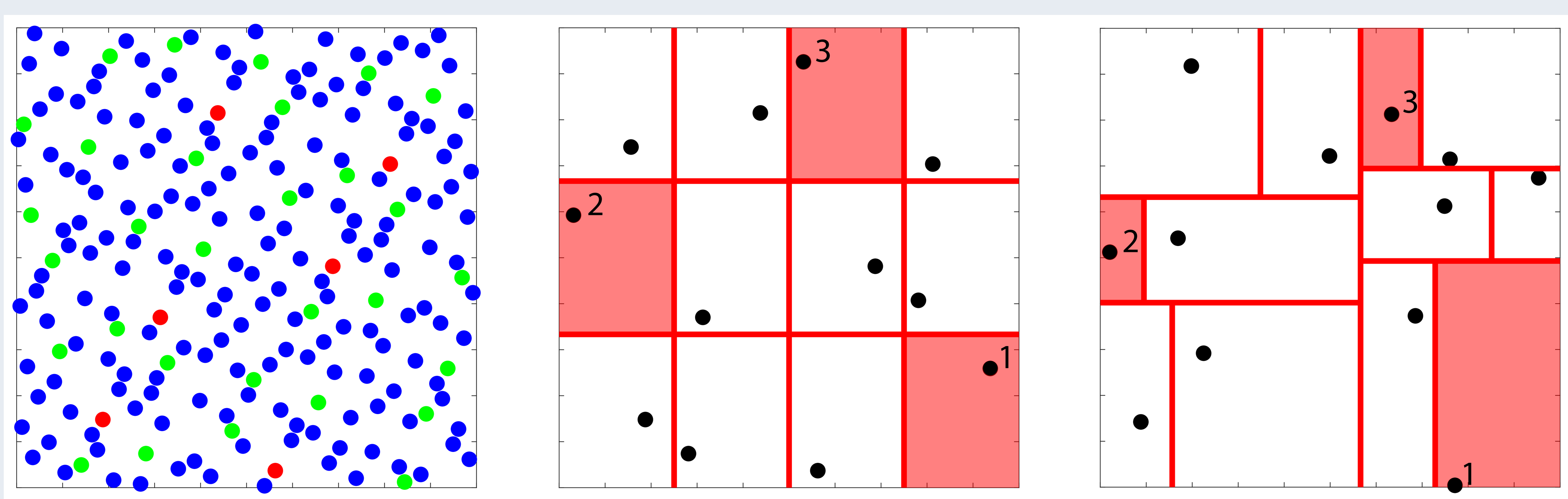

Figure 2:(Left) A BAS sample of $n = 216$ points with $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_6$ in red and $\boldsymbol{x}_7, \ldots, \boldsymbol{x}_{36}$ in green. Hence, contiguous sub-samples are also spatially balanced. (Centre) Halton boxes for $B = 2^2 \times 3 = 12$ and B consecutive points from (1). The first three points in the sequence are numbered and their boxes shaded. (Right) A HIP partition for a $N = 12$ point resource using $B = 12$. Using the box numbering from the centre figure, an equal probability HIP sample of $n = 3$ points from the resource is shown.

Concluding Remarks

Spatially balanced sampling designs are commonly used for sampling natural resources and a variety of designs have been proposed. The main advantages of BAS and HIP include being conceptually simple, computationally efficient and being able to draw spatially balanced over-samples. This makes them particularly useful for sampling natural resources because imperfect sampling frames and accessibility problems result in fewer units being observed than planned. BAS and HIP samples can be drawn using the SDraw package in the R programming language.

References

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

[2] Robertson, B. L., Brown, J. A., McDonald, T., and Jaksons, P. (2013). BAS: Balanced acceptance sampling of natural resources. *Biometrics* 3, 776-784.

[3] Robertson, B. L., McDonald, T., Price, C. J., and Brown, J. A. (2017). A modification of balanced acceptance sampling. *Statistics and Probability Letters* 129, 107-112.

[4] Robertson, B. L., McDonald, T., Price, C. J., and Brown, J. A. (2018). Halton iterative partitioning: spatially balanced sampling via partitioning. *Environmental and Ecological Statistics* 3, 305-323.

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

Contact Information

Email: blair.robertson@canterbury.ac.nz

UC UNIVERSITY OF CANTERBURY Te Whare Wānanga o Waitaha CHRISTCHURCH NEW ZEALAND

WEST Inc.