Statistics and Probability Letters 129 (2017) 107–112

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A modification of balanced acceptance sampling

B.L. Robertson [a,*], T. McDonald [b], C.J. Price [a], J.A. Brown [a]

[a] School of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand
[b] Western EcoSystems Technology, Inc, Cheyenne, WY, USA

ARTICLE INFO

Article history:
Received 31 March 2017
Received in revised form 3 May 2017
Accepted 7 May 2017
Available online 25 May 2017

Keywords:
Environmental sampling
Halton sequence
Spatial balance

ABSTRACT

This article presents a modification of balanced acceptance sampling (BAS) that causes inclusion probabilities to better approximate targeted inclusion probabilities. A new sample frame constructor for BAS is also introduced from which equi-probable spatially balanced samples are drawn.

1. Introduction

A spatially balanced sampling design selects sample locations that are well spread over the study area, a sample with few 'clumps' or 'voids'. Natural resources are often spatially autocorrelated because nearby locations interact with one another and are influenced by the same factors (Stevens and Olsen, 2004). Hence, spreading the sample over of the study area is known to be efficient, and many variations of spatially balanced designs have been proposed (Stevens and Olsen, 2004; Grafström et al., 2012; Robertson et al., 2013). This article considers balanced acceptance sampling (BAS) (Robertson et al., 2013).

BAS uses a quasi-random number sequence to select spatially balanced samples from either continuous or point resources in multidimensional space. In particular, BAS uses a random-start Halton sequence (Wang and Hickernell, 2000), which maps the natural numbers to vectors $\{\mathbf{x}_k\}_{k=1}^{\infty}$ in $[0,1)^d$ (Robertson et al., 2013). The ith coordinate of each vector in the sequence has an associated base, b_i, and all bases $\{b_1, b_2, \ldots, b_d\}$, are required to be pair-wise co-prime. In this article b_i is the ith prime number. The ith coordinate of the kth point in this sequence is (Price and Price, 2012)

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

where u_i is a random non-negative integer and $\lfloor x \rfloor$ is the floor function — the largest integer that is less than or equal to x. The random-start Halton sequence is

$$\{\mathbf{x}_k\}_{k=1}^{\infty} = \left\{ x_k^{(1)}, x_k^{(2)}, \ldots, x_k^{(d)} \right\}_{k=1}^{\infty}. \tag{1}$$

Choosing the first d prime numbers as bases and setting $u_i = 0$ for all i gives the classical Halton sequence (Halton, 1960).

To observe an equi-probable BAS sample from a continuous resource such as a polygon or geographic region, which we label Ω, a random-start Halton sequence is defined over a minimal bounding box that encloses Ω. If $\mathbf{x}_1 \in \Omega$, the point

* Corresponding author.
E-mail address: blair.robertson@canterbury.ac.nz (B.L. Robertson).

http://dx.doi.org/10.1016/j.spl.2017.05.004

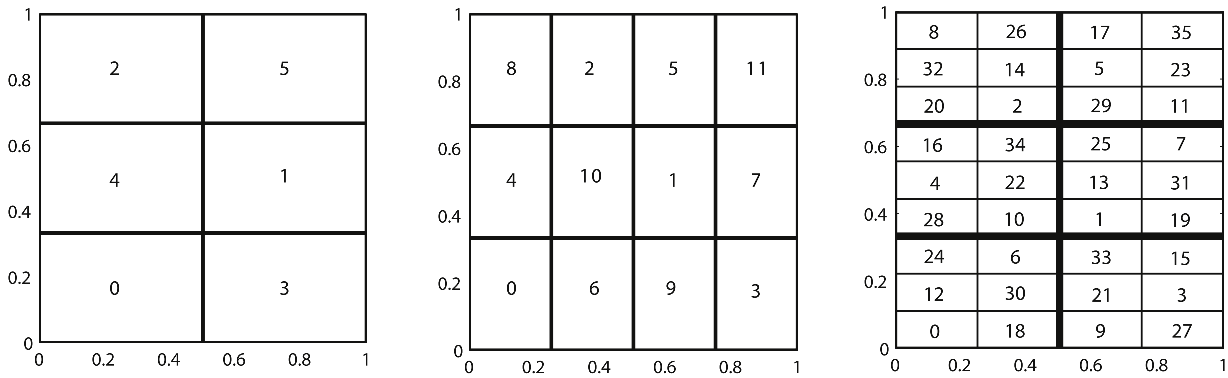

Fig. 1. Halton boxes with $b_1 = 2$ and $b_2 = 3$ for different J_i values. Left: $J = (1, 1)$ and $B = 2 \times 3 = 6$; Center: $J = (2, 1)$ and $B = 2^2 \times 3 = 12$; Right: $J = (2, 2)$ and $B = 2^2 \times 3^2 = 36$. Any B consecutive points from a random-start Halton sequence will have exactly one point in each of the B boxes. Points from the classical Halton sequence ($u_1 = u_2 = 0$) with the same k values (mod B) will be in the Halton box labeled k mod B. For example, $\mathbf{x}_{71}$ will be in the box numbered 5 (left), 11 (center) and 35 (right).

is included in the sample, otherwise the candidate point is rejected. The next point in the sequence is then considered. If $\mathbf{x}_2 \in \Omega$, the point is included, otherwise it is rejected. This process of testing successive points for membership in Ω is repeated until the required number of points has been accepted. Unequal probability samples can be achieved by adding a dimension and using an acceptance/rejection sampling strategy (Robertson et al., 2013).

Sampling point resources using BAS, for example, a collection of coordinate locations on a map or a grid defined over a polygon, is similar. First, however, the N points are replaced with N non-overlapping equally sized boxes with positive Lebesgue measure, where each box contains exactly one point. Then, a random-start Halton sequence is defined over a minimal bounding box containing all N boxes. If $\mathbf{x}_1$ is within a unit's box, that unit is included in the sample. Otherwise, no unit is selected. The next point in the sequence, $\mathbf{x}_2$, is then considered and the method repeats. A BAS sample is realized when the required number of distinct units has been accepted.

BAS has several desirable properties including spatial balance in two or more dimensions, spatially balanced over-samples and admittance of standard design-based estimators (Robertson et al., 2013). However, when sampling point resources, targeted inclusion probabilities are not necessarily achieved because BAS uses an acceptance/rejection sampling technique. The acceptance/rejection technique changes the actual inclusion probabilities and they need to be estimated (or calculated if the population is sufficiently small) for unbiased estimation of population parameters (Robertson et al., 2013). The dangers of not achieving targeted inclusion probabilities include bias and increased variance of the Horvitz–Thompson estimator.

In this article we present a modification of BAS which causes inclusion probabilities to better approximate targeted inclusion probabilities and introduce a new sample frame constructor for point and continuous resources, called Halton frames, which allow exact equi-probable BAS samples to be drawn from these resources. We begin by discussing properties of the Halton sequence that are pertinent to our modification and Halton frames. In Section 4 we present the modification of BAS and show how exact equi-probable BAS samples and cluster samples are drawn from Halton frames. Design-based estimators are given in Section 4.3 and concluding remarks are given in Section 5.

2. Properties of the Halton sequence

It can be shown that for any set of positive integers J_i, any $B = \prod_{i=1}^{d} b_i^{J_i}$ consecutive points from a Halton sequence (1) with co-prime bases b_i, will have exactly one point in each of the boxes determined by

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

where m_i is an integer satisfying $0 \leq m_i < b_i^{J_i}$, for all $i = 1, 2, \ldots, d$ (Price and Price, 2012; Halton, 1960). Indeed, this property ensures the Halton sequence is well-spread over the unit box. We call these boxes Halton boxes and they are illustrated in Fig. 1 for $(J_1, J_2) = (1, 1), (2, 1)$ and $(2, 2)$. The size and shape of these boxes can be altered by choosing different co-prime bases and different J_i values. Increasing J_i reduces the size of the boxes in the ith dimension and varying J_i for different dimensions changes the shape of the boxes.

Another property of the Halton sequence is that it is quasi-periodic. If $\mathbf{x}_k$ is in a specific Halton box, then k mod $b_i^{J_i}$ must take a specific value from the set $\{0, 1, \ldots, b_i^{J_i} - 1\}$ for each $i = 1, 2, \ldots, d$ (Price and Price, 2012; Halton, 1960). To determine the value of k (mod B) for points in each box, a system of congruences is solved using the Chinese Remainder Theorem (CRT). For example, if $b_1 = 2, b_2 = 3, J_i = 2$ (giving $B = 2^2 \times 3^2 = 36$) and k satisfies the following congruences

$$k = 1 \pmod 4$$
$$k = 4 \pmod 9),$$