Spatial Spread Sampling Using Weakly Associated Vectors

(2020) Journal of Agricultural, Biological and Environmental Statistics volume 25, pages 431-451

Raphaël Jauslin & Prof. Yves Tillé

2 December 2020

Notation

- Population : $U = \{1, \dots, k, \dots, N\}$
- Sample : $s \subset U$,
  Example : $U = \{1, 2, 3, 4, 5\}$, sample $s = \{2, 3, 5\}$, or $\mathrm{s} = (0, 1, 1, 0, 1)^\top$.
- Inclusion probabilities $\pi_k = \mathbb{P}(k \in S) = \sum_{s \ni k} p(s)$.
- Spatial coordinates $\mathrm{z}_k^\top = (z_{k1}, z_{k2}, \dots, z_{kp}) \in \mathbb{R}^p$
- Total $Y = \sum_{k \in U} y_k$ and Horvitz-Thompson estimator $\widehat{Y} = \sum_{k \in S} \frac{y_k}{\pi_k}$

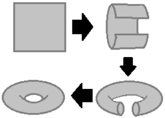
Euclidean and tore distance

Euclidean distance

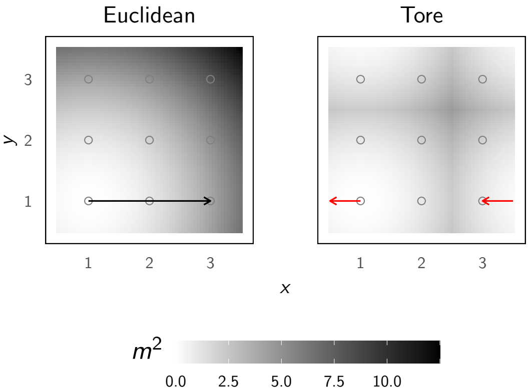
$$
\begin{aligned}
m_E^2 : U \times U &\to \mathbb{R}^+ \\
(k,\ell) &\to (\mathbf{z}_k - \mathbf{z}_\ell)^\top (\mathbf{z}_k - \mathbf{z}_\ell)
\end{aligned}
$$

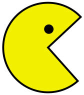
Euclidean distance

$$m_E^2 : U \times U \rightarrow \mathbb{R}^+$$
$$(k, \ell) \rightarrow (z_k - z_\ell)^\top (z_k - z_\ell)$$



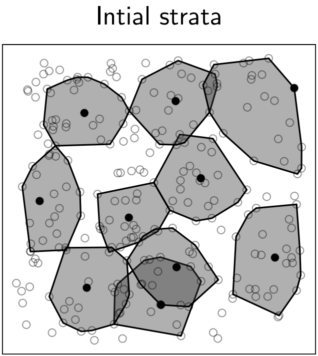
Let $U = \{1, \dots, 9\}$ be on a regular grid of size 3×3.

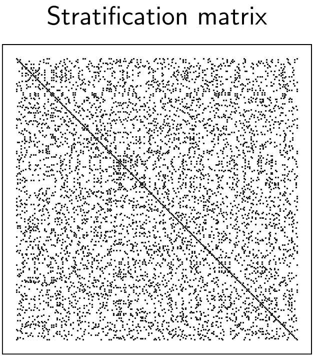




Strata illustration

Figure: The overall population size is equal to $N = 250$ and the inclusion probabilities are identical and equal to $\pi_k = 1/25 = 0.04$. Number of units in each stratum is equal to 25.





Intuition of WAVE sampling

Algorithm is based on the cube method (Deville and Tillé, 2004). The idea is to modify recursively the vector of inclusion probabilities $\boldsymbol{\pi}$.

Intuition of WAVE sampling

Algorithm is based on the cube method (Deville and Tillé, 2004). The idea is to modify recursively the vector of inclusion probabilities $\boldsymbol{\pi}$.

1. Start from $\boldsymbol{\pi}$.

Intuition of WAVE sampling

Algorithm is based on the cube method (Deville and Tillé, 2004). The idea is to modify recursively the vector of inclusion probabilities $\boldsymbol{\pi}$.

1. Start from $\boldsymbol{\pi}$.
2. Compute the stratification matrix with $\boldsymbol{\pi}$ and spatial coordinates $\mathbf{z}_k$, $k \in U$.

Intuition of WAVE sampling

Algorithm is based on the cube method (Deville and Tillé, 2004). The idea is to modify recursively the vector of inclusion probabilities $\boldsymbol{\pi}$.

1. Start from $\boldsymbol{\pi}$.
2. Compute the stratification matrix with $\boldsymbol{\pi}$ and spatial coordinates $\mathbf{z}_k$, $k \in U$.
3. Find the weakly associated vector of the stratification matrix.

Intuition of WAVE sampling

Algorithm is based on the cube method (Deville and Tillé, 2004). The idea is to modify recursively the vector of inclusion probabilities $\boldsymbol{\pi}$.

1. Start from $\boldsymbol{\pi}$.
2. Compute the stratification matrix with $\boldsymbol{\pi}$ and spatial coordinates $\mathbf{z}_k$, $k \in U$.
3. Find the weakly associated vector of the stratification matrix.
4. Modify the vector $\boldsymbol{\pi}$ following the direction of the latter.

Intuition of WAVE sampling

Algorithm is based on the cube method (Deville and Tillé, 2004). The idea is to modify recursively the vector of inclusion probabilities $\boldsymbol{\pi}$.

1. Start from $\boldsymbol{\pi}$.
2. Compute the stratification matrix with $\boldsymbol{\pi}$ and spatial coordinates $\mathbf{z}_k$, $k \in U$.
3. Find the weakly associated vector of the stratification matrix.
4. Modify the vector $\boldsymbol{\pi}$ following the direction of the latter.
5. Update the stratification matrix, and recompute everything until there is not anymore inclusion probabilities such that $0 < \pi_k < 1$.

Voronoï (Stevens Jr. and Olsen, 2004)

Let v_k be the sum of inclusion probabilities of the units belonging to the Voronoï polygon of the unit k. The spatial balance measure is

$$B(S) = \frac{1}{n} \sum_{k \in S} (v_k - 1)^2 \qquad (1)$$

Moran's I index (Tillé et al., 2018)

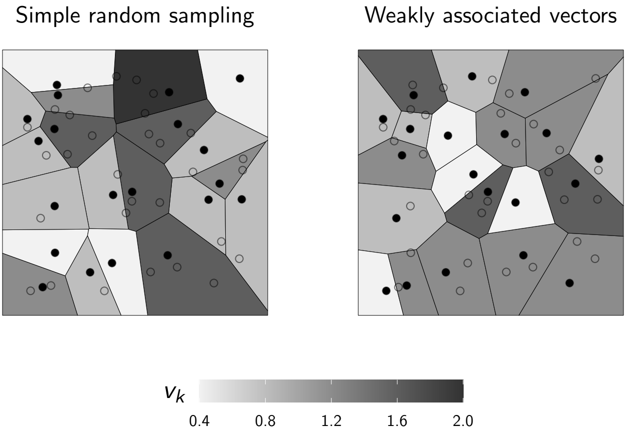
$$I_B = \frac{(\mathrm{s} - \bar{\mathrm{s}}_w)^\top \mathrm{W} (\mathrm{s} - \bar{\mathrm{s}}_w)}{\sqrt{(\mathrm{s} - \bar{\mathrm{s}}_w)^\top \mathrm{D} (\mathrm{s} - \bar{\mathrm{s}}_w) \quad (\mathrm{s} - \bar{\mathrm{s}}_w)^\top \mathrm{B} (\mathrm{s} - \bar{\mathrm{s}}_w)}}.$$

where $\mathrm{W} = (w_{k\ell})$, w_{ij} indicates how close is j to i, $w_{ii} = 0$, D be the diagonal matrix containing $w_{i.}$

$$\mathrm{A} = \mathrm{D}^{-1}\mathrm{W} - \frac{\mathrm{1}\mathrm{1}^\top \mathrm{W}}{\mathrm{1}^\top \mathrm{W}\mathrm{1}}, \quad \mathrm{B} = \mathrm{A}^\top \mathrm{D}\mathrm{A}.$$

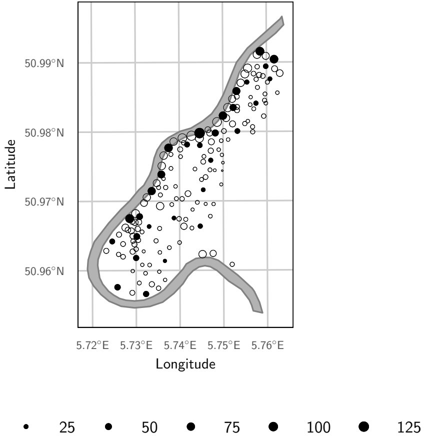
Voronoï example

Figure: The population and sample sizes are respectively equal to $N = 50$ and $n = 20$, the inclusion probabilities are identical and equal to $\pi_k = 0.4$.



Meuse dataset package "sp" (Pebesma and Bivand, 2005)

Figure: The overall population size is equal to 155. The inclusion probabilities are proportional to copper level variable and the sample size is equal to 30.



GitHub (Jauslin and Tillé, 2020)

```
library(sampling)
library(WaveSampling)
data("meuse",package = "sp")
X <- as.matrix(meuse[,1:2]) # spatial coordinates
pik <- inclusionprobabilities(meuse$copper,30)
s <- wave(X,pik)
head(s,8)

##      [,1]
## [1,]    0
## [2,]    1
## [3,]    1
## [4,]    0
## [5,]    0
## [6,]    1
## [7,]    0
## [8,]    0
```

Spatial Balance

Table: Spreading measures results based on 10000 simulations on the Meuse dataset. The population size is equal to 155. We compare different usual methods for spatial sampling such as Generalized Random Tessellation Stratified or Local Pivotal Method (Grafström et al., 2012).

	Sampling design										
	Equal probabilities						Unequal probabilities				
	wave	lpm1	scps	grts	hip	srswor	wave	lpm1	scps	grts	maxent
I_B											
$n = 15$	-0.518	-0.338	-0.351	-0.226	-0.230	-0.030	-0.354	-0.244	-0.247	-0.153	0.009
$n = 30$	-0.664	-0.427	-0.427	-0.266	-0.259	-0.019	-0.427	-0.290	-0.283	-0.154	0.048
$n = 50$	-0.796	-0.519	-0.473	-0.302	-0.248	-0.011	-0.455	-0.305	-0.263	-0.181	0.060
B											
$n = 15$	0.119	0.125	0.118	0.170	0.160	0.379	0.115	0.121	0.120	0.170	0.387
$n = 30$	0.118	0.123	0.126	0.164	0.159	0.359	0.120	0.121	0.120	0.162	0.345
$n = 50$	0.139	0.132	0.143	0.174	0.194	0.329	0.138	0.133	0.141	0.160	0.281

UNIVERSITÉ DE NEUCHÂTEL

Bibliography I

Deville, J.-C. and Tillé, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika*, 91:893–912.

Grafström, A., Lundström, N. L. P., and Schelin, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics*, 68(2):514–520.

Jauslin, R. and Tillé, Y. (2020). Spatial spread sampling using weakly associated vectors. *Journal of Agricultural, Biological and Environmental Statistics*, 25:431–451.

Pebesma, E. J. and Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, 5(2):9–13.

Stevens Jr., D. L. and Olsen, A. R. (2004). Spatially balanced sampling of natural resources. *Journal of the American Statistical Association*, 99(465):262–278.

Tillé, Y., Dickson, M. M., Espa, G., and Giuliani, D. (2018). Measuring the spatial balance of a sample: A new measure based on the Moran's *I* index. *Spatial Statistics*, 23:182–192.