Spatial Sampling Strategies

Luis Ambrosio Flores
Luis Iglesias Martínez
Departamento de Economía y Ciencias Sociales Agrarias, U.P.M. Madrid, Spain

Robert P. Haining
Department of Geography, University of Sheffield, England

ABSTRACT: A practical method to identify a spatial sampling strategy for efficient estimation of the spatial mean of an attribute in a selected area is presented. The (relative) efficiency of a spatial sampling strategy depends on the spatial (variability) correlation structure among the population elements. For undertaking spatial sampling, the following sequence of steps is considered: (i) computing the empirical variogram (or correlogram) using data from a pilot sample, (ii) fitting to these data a theoretical model of the variogram, (iii) computing average correlations between pairs of elements under the fitted model, and (iv) computing the expected efficiencies of the considered sampling strategies under the fitted model. Using data from an agricultural land use map, the method is tested. In this test, the expected efficiency of the sample mean under random sampling, stratified random sampling and systematic random sampling is computed and compared with the observed efficiency. The results are satisfactory.

1. Introduction

We consider the problem of designing a sample to estimate the spatial mean of an attribute in a selected area. The problem is motivated by the need to estimate crop acreages and land use cover to provide official agricultural statistics for an area. We consider the sample mean under three sampling schemes: random sampling (RS), stratified random sampling (STR) and systematic random sampling (SYS). The population that we are considering is arranged in a regular lattice. The elements are grouped in mn blocks with lk elements in each. In the RS scheme, a sample of $\nu\, mn$ elements is selected with equal probability among the $mnlk$ elements of the population. In the STR scheme, each block is considered as a stratum and a sample of size ν is selected in each. In the SYS scheme, (i) ν elements are chosen with equal probability from the lk elements of a block, and (ii) the ν elements in each of the other $(mn\text{-}1)$ blocks are those elements that occupy the same positions within each block as were chosen at (i).

2. The Model-based Relative Efficiency

We will follow a model-based approach to study the efficiencies of these sampling strategies. Given a correlation model, it can be shown [Das 1950] that the expected values of the design-based variance of the sample mean under each one of the three considered sampling schemes are:

RS: $$EV_r(\bar{z}) = \frac{\sigma^2}{mn\vartheta}\left(1-\frac{\vartheta}{lk}\right)\left[1-\Phi(ml,nk)\right]$$

STR: $$EV_{st}(\bar{z}) = \frac{\sigma^2}{mn\vartheta}\left(1-\frac{\vartheta}{lk}\right)\left[1-\Phi(l,k)\right]$$

SYS: $$EV_{sy}(\bar{z}) = \frac{\sigma^2}{mn\vartheta}\left(1-\frac{\vartheta}{lk}\right)\left[1-\frac{mnlk-1}{lk-1}\Phi(ml,nk)+\frac{lk(mn-1)}{lk-1}\Phi_{sy}(m,n)\right] \quad [1]$$

where $\Phi(ml,nk)$, $\Phi(l,k)$ and $\Phi_{sy}(m,n)$ are the average correlations between all pairs of elements of the population, between all pairs of elements of a given stratum and between all pairs of elements of the same systematic sample (cluster), respectively.

Two theoretical models, Spherical and Exponential, have been fitted to the empirical variogram of nonirrigated herbaceous, using a pilot sample (systematic at a rate of 3 percent). The models were fitted in two ways, one using REML covariance estimates and the other assigning to the parameters (nugget, range and sill) the values considered visually/graphically reasonable. The expected relative efficiencies using the REML fitting are shown in Table 1 for $l=k$.

Table 1. Model-based and Observed Relative Efficiencies in the selected area

Relative efficiency		Model/Population	l-values					
			2	5	10	20	50	100
Stratified with respect to Random sampling	Model based EV_r/EV_{st}	Spherical	1.76	1.70	1.61	1.46	1.19	1.05
		Exponential	1.97	1.86	1.71	1.51	1.23	1.08
	Observed V_r/V_{st}	Population	4.01	2.35	1.73	1.42	1.16	1.04
Systematic with respect to Random sampling	Model based EV_r/EV_{sy}	Spherical	1.71	1.67	1.63	1.54	1.37	1.06
		Exponential	1.91	1.82	1.74	1.60	1.32	1.11
	Observed V_r/V_{sy}	Population	3.47	2.06	1.40	1.11	1.44	1.04
Systematic with respect to Stratified	Model based EV_{st}/EV_{sy}	Spherical	0.97	0.98	1.01	1.05	1.15	1.01
		Exponential	0.97	0.98	1.01	1.06	1.07	1.03
	Observed V_{st}/V_{sy}	Population	0.87	0.88	0.81	0.78	1.24	1.00

3. Conclusions

Variogram or correlogram estimates are needed to design good spatial sampling schemes. The variogram can be accurately estimated using a pilot sample. We found that the expected model-based efficiencies are very similar to the observed efficiencies. The results are robust with respect to the chosen theoretical model of variogram and to the fitting procedures.

STR and SYS are more efficient than RS, and their efficiencies increase when the square block side *'l'* decrease, so that the optimum *l*-value is the minimum *l*-value. For STR and SYS, given a population of size *NN* and a sample size $s_0 = \nu nn$, the most efficient strategy corresponds to the minimum *'l'* which corresponds to the maximum of $nn = s_0/\nu$ and this corresponds to the minimum ν $(\nu=1)$. For SYS, the most efficient strategy is to take only one sample ($\nu=1$) of elements a distance apart $l=NN/s_0$ in the row and column directions; in other words, as suggested by several authors, a regular arrangement of the ν starting-points is more efficient than a random selection of the ν starting-points. For the given sample size s_0 and the most efficient strategy, the expected variance of the estimator can be obtained using [1] and the fitted correlation model. For a given *l*-value, the sample size necessary to achieve a desired precision $EV(z)=V_0$ can be estimated using [1] and the fitted correlation function.

References

Das, A.C. (1950), "Two-dimensional Systematic Sampling and the Associated Stratified and Random Sampling," Sankhaya 10, pp 95-108.