An Algorithm For The Mean Autocorrelation In Regular Arrays

Steen Magnussen, Canadian Forest Service, Pacific Forestry Centre,
Victoria, British Columbia

Background and rationale:

Two-stage cluster sampling is an efficient sampling design for large scale surveys using remotely sensed data. Practical and economic considerations favors large first stage sampling units (clusters). Statistical efficiency, on the other hand, generally increases as cluster size goes down. Deciding which cluster size to use in a survey becomes a balancing act between factors with opposite effects. To make a rational choice the survey planner needs a model predicting how the variance of survey results will vary with the size and spatial configuration of the clusters.

The similarity of observations taken within a cluster is a key determinant of the statistical efficiency of a survey. The higher the similarity the lower the efficiency. Similarity is expressed in terms of an intra-plot correlation. Post-survey estimates of the correlation are obtained from one-way analyses of variance. Prior estimates, however, are needed for planning an optimum cluster size. Information from the autocorrelation of 'close' survey units can be used to develop a model of the correlation between cluster units as a function of the distance separating them.

Once a basic model for the correlation of cluster units has been developed the statistical efficiency of the cluster is found to be inversely proportional to the average correlation between the cluster units. Computing the average correlation between hundreds of cluster units is straightforward but it can be a tedious and time consuming task.

This poster presents a method for a rapid computation of the mean correlation between cluster units when the associated spatial autocorrelation is a power function of the 'Manhanttan Distance' (aka around the block distance). When applied to survey clusters with 120 x 120 units the fast computation was done in less than a second as compared to 328 seconds for the direct approach (results with a 300 Mhz PC with 120Mb RAM).

For a cluster with units arranged in m rows and n columns and an autocorrelation which is a power function of the Manhanttan distance between units the mean autocorrelation is:

$$\overline{\rho}_{m\times n}=\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{i'=1}^{m}\sum_{j'=1}^{n}\rho^{|i-i'|+|j-j'|}-m\cdot n}{m\cdot n\cdot(m\cdot n-1)}$$

To arrive at the fast computing algorithm it is helpful to break this expression into separate easily manipulated sums of within rows, within columns and across column (row) correlation.

Within row correlation is:

$$\Gamma_{row}=2\cdot m\cdot\sum_{j=1}^{n-1}\sum_{j'=j+1}^{n}\rho^{|j'-j|}=-\frac{2\cdot m\cdot\rho\cdot(1-n+n\rho-\rho^{n})}{(-1+\rho)^{2}}$$

A similar expression is available for columns with m and n exchanged (not shown).

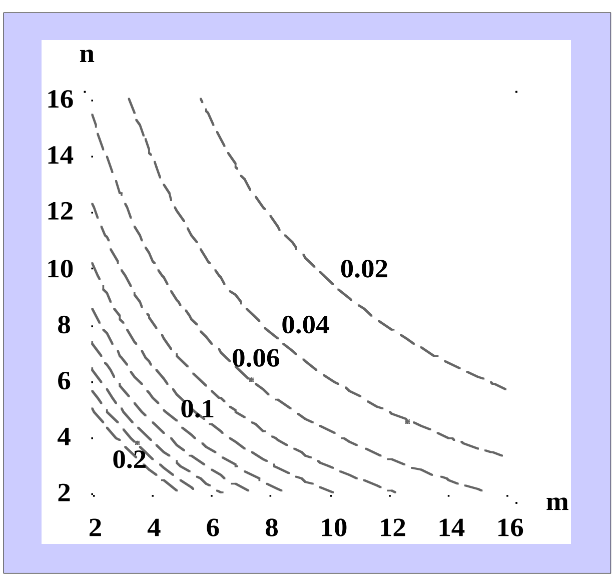
The across row correlation is:

$$\Gamma_{across\ rows}=2\cdot\sum_{k=1}^{m-1}\sum_{j=1}^{n-1}\sum_{j'=j+1}^{n}(m-k)\cdot\rho^{j'-j+k}=2\cdot\frac{\rho^{2}\cdot(1+m(-1+\rho)-\rho^{m})\cdot(1+n\cdot(-1+\rho)-\rho^{n})}{(-1+\rho)^{4}}$$

Correlation across columns would yield an identical term (not shown). Combining the above terms and after some simplification one obtains the final algorithm for the mean autocorrelation:

$$\overline{\rho}_{m\times n}=[-2\rho\cdot((-1+\rho^{m})\cdot(n\cdot(-1+\rho^{2})-2\cdot\rho\cdot(-1+\rho^{n}))+m\cdot(-1+\rho)\cdot(-2\cdot n\cdot(-1+\rho)+(1+\rho)\cdot(-1+\rho^{n})))]/(-1+\rho)^{4}/(m\cdot n\cdot(m\cdot n-1))$$

Further simplifications are possible for square clusters (m=n). See proceedings for further details. A graphic illustration of how the average within plot autocorrelation (contour lines in graph) depends on *m* and *n* is shown below for r = 0.4.



Conclusions:

Simple algorithms for the mean autocorrelation in large plots can be derived for isotropic/symmetric autocorrelation functions. The algorithm(s) provide(s) for faster computing and insight to the relationship between correlation at the unit level and at the aggregate level of a cluster.

Contact:

Steen Magnussen
Pacific Forestry Centre
506 West Burnside Road
Victoria, BC V8Z 1M5
(250) 363-0712
smagnuss@pfc.forestry.ca

Natural Resources Canada | Ressources naturelles Canada
Canadian Forest Service | Service canadien des forêts
Canada