Resampling methods for non-stationary spatial data

Magnus Ekström
Department of Forest Resource Management and Geomatics
Swedish University of Agricultural Sciences, Umeå
S-901 83 Umeå, Sweden
Magnus.Ekstrom@resgeom.slu.se

Sara Sjöstedt-de Luna
Department of Mathematical Statistics
Umeå University
S-901 87 Umeå, Sweden
sara@matstat.umu.se

1. Introduction

In this paper a resampling method for *non-stationary* spatial data is proposed, and a consistent estimator of the variance of sample means is provided. It is assumed that the data is observed on a regular lattice in some region $A \subset \mathbf{R}^2$. Remote sensing data from satellites are, for example, on this form. Such collected data may be used for different purposes, with applications to forestry, agriculture, landscape changes, environmental monitoring, etc. Due to the variation on the ground it is reasonable to believe that features measured in different pixels are differentely distributed although nearby observations tend to be similar. Motivated by this, we consider differently distributed locally dependent random variables on a spatial grid, with smoothly varying expected values, or with expected values that can be decomposed additively into directional components. The dependence structure is allowed to differ over the lattice.

2. Estimator of variance

For convenience we assume that the region A is of rectangular shape. Hence we have spatially indexed data $\{X_{\boldsymbol{i}} : \boldsymbol{i} = (i_1, i_2) \in \mathcal{I}_{\boldsymbol{n}}\}$, where $\mathcal{I}_{\boldsymbol{n}} = \{\boldsymbol{i} : i_1 = 1, ..., n_1, \text{ and } i_2 = 1, ..., n_2\}$, $\boldsymbol{n} = (n_1, n_2)$. We want to estimate the variance of the sample mean $\bar{X}$, using a variant of block resampling. Different block resampling estimators of variance have been proposed under moment and mixing conditions, e.g. in Politis & Romano (1993) and Sherman (1996). In these papers the data is assumed to be stationary, although this assumption can be relaxed (Ekström 2001). However, these estimators cannot handle the case of varying expected values.

We assume that, for all $\boldsymbol{n}$, $X_{\boldsymbol{i}} = Y_{\boldsymbol{i}} + r_{i_2} + c_{i_1}$, where $f(i_1/n_1, i_2/n_2) = EY_{\boldsymbol{i}}$ satisfies a Lipschitz condition of order α, and $\sum r_{i_2} = \sum c_{i_1} = 0$. Assume that $X_{\boldsymbol{i}}$ and $X_{\boldsymbol{j}}$ are independent whenever $|i_1 - j_1| > m_1$ or $|i_2 - j_2| > m_2$, for some m_1 and m_2, and that $|X_{\boldsymbol{i}}|$, $\boldsymbol{i} \in \mathcal{I}_{\boldsymbol{n}}$, have uniformly bounded moments of order $2 + \delta$, for some $\delta > 0$. Construct rectangular blocks of size $k_1 \times k_2$, let $K = k_1 k_2$, and denote block averages by $\bar{X}_{\boldsymbol{i}} = \sum_{j_1=i_1}^{i_1+k_1-1} \sum_{j_2=i_2}^{i_2+k_2-1} X_{\boldsymbol{j}}/K$. Compute crosswise differences of the block averages $Z_{\boldsymbol{i}} = \bar{X}_{i_1,i_2} + \bar{X}_{i_1+t_1+k_1,i_2+t_2+k_2} - \bar{X}_{i_1+t_1+k_1,i_2} - \bar{X}_{i_1,i_2+t_2+k_2}$. An estimator of $\gamma_{\boldsymbol{n}} = n_1 n_2 Var(\bar{X})$ is now formed as

$$\hat{\gamma}_{\boldsymbol{n}} = \frac{K}{4N'} \sum_{i_1=1}^{n_1'} \sum_{i_2=1}^{n_2'} (Z_{\boldsymbol{i}} - \bar{Z})^2,$$

where $N' = n_1' n_2'$, $n_l' = n_l - 2k_l - t_l + 1$, $l = 1, 2$, and $\bar{Z} = \sum_{i_1=1}^{n_1'} \sum_{i_2=1}^{n_2'} Z_{\boldsymbol{i}}/N'$. If $t_l > m_l$ and $k_l \to \infty$, as $n_l \to \infty$, $l = 1, 2$, then $\hat{\gamma}_{\boldsymbol{n}} - \gamma_{\boldsymbol{n}} \xrightarrow{P} 0$ under the assumptions given above (plus an additional assumption on at which rate the block size is increasing).

3. Example

General forest parameters can be estimated by combining satellite data with sparsely distributed field data, e.g. estimates of wood volume can be derived from a Landsat TM scene together with Swedish National Forest Inventory (NFI) data, by applying the k Nearest Neighbour (kNN) method. In the kNN method, values of forest parameters are calculated for pixels as weighted averages of the k spectrally most similar plots. The NFI is conducted exclusively as a field survey, and has a low spatial resolution. By using information collected in the NFI, together with satellite data, it is believed that reliable estimates of forest parameters can be obtained for smaller areas than what is possible otherwise. Since an estimate of wood volume per hectare over some area is obtained by averaging kNN estimates given for 25×25m pixels in the Landsat TM scene, estimates of accuracy can be obtained by using methods derived in the current paper. When comparing our estimates of variance with estimates of MSE in Nilsson & Sandström (2001), obtained from a field inventory, it appears as the resampling method yields reasonable results (assuming that the bias is rather low over aggregated areas).

REFERENCES

Ekström, M. (2001). Non-parametric estimation of the variance of sample means based on non-stationary spatial data. Manuscript.

Nilsson, M. & Sandström, P. (2001). Evaluation of a non-parametric method for estimation of wood volume and biomass in boreal forests by combining remote sensing and field data. Manuscript.

Politis, D.N. & Romano, J.P. (1993). Nonparametric resampling for homogeneous strong mixing random fields. *J. Multivariate Anal.*, 47, 301-328.

Sherman, M. (1996). Variance estimation for statistics computed from spatial lattice data. *J. Roy. Statist. Soc. Ser. B*, 58, 509-523.