Modeling the Reduction of Sample Size for Spatial Datasets

Ronny Vallejos
Departamento de Matemática
Universidad Técnica Federico Santa María,
Valparaíso, Chile
ronny.vallejos@usm.cl

Abstract

This work focuses on the reduction of sample sizes due to the effect of autocorrelation for spatial/time models. this type of problem is quite common in several disciplines where the goal is to reduce the number of georeferenced observations to be sampled. In the literature the effective sample size (ESS) is defined as the number of independent and identically distributed observations of a spatial process. Recent research extended that definition to more complex models with more general conditions for the theoretical developments. In this talk we motivate the study of ESS with several practical problems illustrated with real-world datasets. Then an overview of the main proposals will be presented, including a general definition of ESS for arbitrary spatial regression processes as a weighted version of the ESS of each column in the design matrix. Examples of different patterned correlation structures are explored in order to establish theoretical properties that hold for the ESS in the original case (constant mean). We study the asymptotic properties of the ML estimates of the ESS for an increasing domain framework. The consistency and asymptotic normality are established under very precise conditions. Monte Carlo simulations are carry out to gain more insight into the asymptotic properties for finite sample sizes. A hypotesis testing is developed to test whether a reduction of the sample size is adequate. In addition, a model-free definition of the ESS is provided in order to estimate the number of observations that are necessary to plot the codispersion map (a new tool to visualize the spatial correlation) on the plane, without excessively increasing the computational time.

Key words: Effective sample size; spatial process; correlation function.