U.P.B. Sci. Bull., Series A, Vol. 82, Iss. 4, 2020 ISSN 1223-7027

RESAMPLING FOR SPACE-TIME RANDOM FIELDS

Dana SYLVAN[1], Constantin TÂRCOLEA[2], Adrian Stere PARIS[3]

Motivated by the "big data" movement, this communication addresses data -intensive bootstrap algorithms for space-time random fields. Over the past decade it has been widely acknowledged that modeling and prediction for space-time data (geostatistical or point patterns) can be a challenging multi-disciplinary task with many open problems. New data science research aimed at finding efficient statistical tools for the exploration, visualization and inference is timely and useful in numerous applied disciplines. In this paper we introduce a space-time moving block-bootstrap method for data observed on a regular grid, also addressing potential extensions to irregularly spaced data. The main ideas shared here may be also used for non-homogeneous Poisson processes where the intensity function varies with time and space. The proposed data -intensive algorithms have applications in many disciplines as to atmospheric and earth sciences, education, engineering, hydrology, public health.

Keywords: quantiles, resampling, Moving Block Bootstrap (MBB).

1. Introduction

The motivation for this paper is linked to the advent of “big data” and the emergent field of data science, supported and constantly improved by increased computing capabilities. In the past decade it has become possible to collect, store and analyze enormous amounts of information. Moreover, the use simulations and resampling techniques such as bootstrap is now quite common and was shown to lead to improved estimations and predictions for highly dimensional stochastic processes with very complex structures. There exists extensive statistical literature on these topics; most relevant for this note is the monograph by Lahiri [1] devoted to resampling for dependent data.

In particular, bootstrap for spatial processes was addressed in recent years in a number of papers focused both on proving asymptotic results, as well as showing applications on real life data. For instance, Mattenfeldt et al. [2] propose a bootstrap variant involving wrapping and tiling to produce a consistent estimator of the spatial intensity together with confidence bands.

[1] Prof., Dept. of Statistics, City Univ. of New York, USA, e-mail: dana.sylvan@gmail.com
[2] Prof., Dept. of Math., UPB, Romania, e-mail: constantin_tarcolea@yahoo.com
[3] Assoc. Prof., Dept. of CCII, UPB, Romania, e-mail: adrian.paris@upb.ro

This paper also presents a very well-designed simulation study where the spatial block bootstrap approach is compared with a parametric scheme involving splitting.

Regarding inhomogeneous point patterns, Guan and Loh [3] use thinned block bootstrap to correctly assess and prove consistency of the variance of the first order intensity function. Resampling for time series was used in Cameletti, Ignaccolo and Sylvan [4] by adapting a seasonal bootstrap scheme in the time domain to improve predictions of threshold exceedance probabilities that were afterwards mapped spatially via a stationary variogram. This was based on the assumption of space-time separability which led to a manageable mathematical framework. However, in many situations this assumption cannot be reasonably justified and thus the need for space-time non-separable approaches. Sylvan, Târcolea and Paris [5] propose a space-time moving block design for modeling quantile surfaces via kernel smoothing, use it in a comparison study of the performance of several space-time quantile estimates, and show that computational time can be significantly reduced with no loss in performance when considering smoothing via non-separable, overlapping space-time moving blocks.

Regarding bootstrap for space-time data, to the best of our knowledge there are no current methods that rely on space-time block bootstrap for random fields and in this respect this paper fills a gap. Resampling ideas have been tried in small number of applied studies. For example, Kim and O'Kelly [6] use bootstrap permutations in a space-time surveillance model, however this empirical approach is appropriate for space-time point patterns and is designed to detect hot spots. For a survey of recent computational methods and tools we refer to Shekhar et al. [7].

The paper is organized as follows. Next section presents the theoretical framework for spatial block bootstrap for processes observed on a regular grid that is extended to spatial-temporal random fields in Section 3. The article ends with a brief discussion including challenges and potential extensions to space-time point patterns.

2. Bootstrap for spatial data

The main idea behind statistical bootstrapping, a technique introduced by Efron [8] for independent, identically distributed observations, is to produce replicates of a statistic by sampling with replacement. When the data are correlated, the sampling region needs to be divided into appropriate blocks in order to preserve the dependence structure, and then the sampling can be carried out in moving blocks. This was a breakthrough in bootstrap for stationary time series and it was introduced in Künsch [9] . The procedure is known as Moving

Block Bootstrap (MBB). In this note we present a method to extend MBB to space-time data.

Initially, consider bootstrapping a stationary spatial process $\{Z(s), s \in Z^d\}$ observed at finitely many locations $\{s_1, . . . , s_n\}$. Here d is the spatial dimension and the observation points are given by the part of the integer grid Z^d that lies inside the sampling region R_n considered to be rectangular, namely $\{s_1, . . . , s_n\} = R_n \cap Z^d$. Let:

$$\mathcal{K}_n = \{k \in Z^d, b_n(k + [0, 1)^d) \cap R_n \neq \emptyset\}, \tag{1}$$

where $b_n \to \infty$ is a sequence of positive integers.

We define a bootstrap version of the spatial attribute Z(.) over R_n by taking its version over each subregion

$$R_n(k) = R_n \cap [b_n(k + [0, 1)d)], k \in \mathcal{K}_n \tag{2}$$

Let $\mathcal{I}_n$ be the index set of all cubes of volume b_n^d in R_n. We take one $R_n(k)$ at a time, resample from a collection of subregions of type k of R_n to define the bootstrap version of $Z(.)$ over $R_n(k)$. Then, $\{i+b_n[0, 1)^d, i \in \mathcal{I}_n\}$ forms a collection of cubic blocks that are overlapping and contained in R_n. For each $i \in \mathcal{I}_n$, the subsample of observations $\{Z(s), s \in Z^d \cap i +b_n[0, 1)^d\}$ is complete, meaning that $Z(.)$ is observed at every point on the integer grid in the subregion $i + b_n[0, 1)^d$. Thus the number of observations in the resampled block $Z_n^*(R_n(k))$ equals b_n^d. The overlapping block bootstrap version $Z_n^*(R_n)$ of $Z_n(R_n)$ is obtained by concatenating the resampled blocks of observations, $\{ Z_n^*(R_n(k)), k \in \mathcal{K}_n\}$.

Based on the above, the bootstrap version of a random variable $T_n = t_n(Z_n , \theta)$ is given by $T_n^* = t_n(Z_n^*(R_n), \tilde{\theta}_n)$, where the same function $t_n(. , .)$ that appears in the definition of T_n is also used to define its bootstrap version.

Here $\tilde{\theta}_n$ is an estimator of θ and is obtained by replicating the relation between the joint distribution of Z_n and θ.

Let

$$T_n = |N_n(h)|^{1/2}(\hat{\theta}_n - \theta), \text{ with } \theta = \mathrm{Cov}(Z(0), Z(h)), \tag{3}$$

is the autocovariance of the spatial process at a fixed lag $h \in Z^d \backslash \{0\}$, and $\hat{\theta}_n$ is the empirical estimator

$$\hat{\theta}_n = \left|N_n(h)\right|^{-1}\sum\nolimits_{s\in N_n(h)} Z(s)Z(s+h) - \left|N_n(h)\right|^{-1}(\sum\nolimits_{s\in N_n(h)} Z(s))^2 , \tag{4}$$

where

$$N_n(h) = \{s \in Z^d : s, s+h \in R_n\} \tag{5}$$

Then T_n is a function of the bivariate spatial process

$$Y(s) = (Z(s), Z(s+h)),' \qquad (6)$$

defined for $s \in R_n \cap (R_n - h)$.
As with MBB for time series, the bootstrap version of the spatial variables can be obtained by considering the vectorized process $\{Y(s), s \in R_n \cap (R_n - h) \cap Z^d\}$.
For a discussion of consistency properties of autocovariance estimators in spatial block bootstrap we refer to Lahiri [1].

3. Space-time Moving Block Bootstrap

We begin by developing a scheme for space-time MBB by extending the spatial block bootstrap previously described to space-time processes observed on a grid.

Let $\{Z(s, t), (s, t) \in Z^{d+1}\}$ be a space-time random field assumed to be non-separable.

We can mimic the previous approach by simply thinking of time as an added dimension, then defining space-time overlapping blocks accordingly. We conjecture that extensive Monte Carlo simulations will inform further adjustments and refinements of the resampling scheme. A significant practical issue in geo-statistics is that the random field of interest may not be observed on a lattice.

However, if the number of spatial locations is sufficient large, we can incorporate a missing data step in the MBB by using spatial interpolation (kriging) and thus impute the attribute at the grid points with no data.

The most important aspect of statistical inference for random fields relies on correct uncertainty assessment in the estimates of the trend (mean function), as well in the estimates of the dependence structure of the variable or attribute of interest.

Under second-order stationarity, the dependence is assumed to be captured by the autocovariance function. Parametric autocovariance models are preferred, then inference is made based on maximum likelihood estimators of these parameters. For non-separable space- time processes, we will employ the flexible Whittle-Matern family of isotropic space-time covariance functions introduced in Gneiting [10]. Typically, the problem of interest is to predict the attribute field at space-time points with no observations and this can be done by using the space-time analogue of kriging. Krigged predictions and the corresponding prediction errors are functions of the autocovariance function. In practice, the auto-covariance matrix is estimated from the same data and this yields additional error.

Analytical expressions of the overall error are very hard, if not impossible to derive. Resampling and/or conditional simulation techniques need to be further employed in order to adjust for all uncertainty sources. We propose a construction of space-time MBB confidence bands based on the resampling scheme previously summarized and thus produce reliable confidence bands for the estimators and predictors under consideration. A large number of recent studies showed that several variants of bootstrap led to improved confidence bands in time series and in spatial models. It seems therefore reasonable to conjecture that the space-time MBB scheme previously described leads to asymptotically efficient estimators of the parameters of the Whittle-Matern space-time autocovariance function.

4. Discussion

Motivated by “big data” and the emergent field of data science, we propose some data-intensive bootstrap algorithms for stationary space-time random fields. Specifically, we show how MBB for spatial data can be extended to space-time random fields for data observed on a regular grid. We also comment on potential extensions to irregularly spaced data and make conjectures on asymptotic properties of parameter estimates of non-separable space-time covariance functions with focus on uncertainty assessment through the resulting space-time MBB confidence bands. Implementation challenges may be significant because of the data-intensive nature of the proposed algorithms, there is a need for future optimized routines. We also believe that the main ideas shared here could be extended to space-time point processes. Specifically, for non-homogeneous Poisson processes where the intensity function varies with time and space, a bootstrap selector of the optimal space-time smoothing parameter may be constructed by adapting the moving block bootstrap scheme previously described. In future work we will start with fixed and time-dependent bandwidths and address edge corrections for the spatial bandwidth matrix. Both overlapping and non-overlapping space-time blocks may be used. In future work we will consider space-time MBB in Monte Carlo simulations and applications to real data. Moreover, as we are learning from the current health crisis, it is extremely important to get new insights to understanding complex space-time point patters.

To conclude, modeling space-time data (geostatistical or point patterns) is a challenging multi-disciplinary task with many open problems. New data science research aimed at finding efficient statistical tools for the exploration, visualization and inference is timely and useful in numerous applied disciplines. This paper proposes data-intensive algorithms with countless applications in many disciplines including but not limited to atmospheric and earth sciences, education, engineering hydrology and public health.

REFERENCES

[1] *S. N. Lahiri*, Resampling methods for dependent data. Springer Verlag, 2003.
[2] *T. Mattfeldt, H. Häbel, and F. Fleischer,* "Block bootstrap methods for the estimation of the intensity of a spatial point process with confidence bounds", Journal of Microscopy, 251, 2013, pp.84-98.
[3] *Y. Guan, J. M. Loh,* "A thinned block bootstrap variance estimation procedure for inhomogeneous spatial point patterns", Journal of the American Statistical Association, 102:480, 2007, pp. 1377–1386.
[4] *M. Cameletti, R. Ignaccolo and D. Sylvan,* "Assessment and visualization of threshold exceedance probabilities in complex space-time settings: A case study of air quality in Northern Italy", Spatial Statistics, vol. 5, 2013, pp. 57–68.
[5] *D. Sylvan, C. Târcolea and A. S. Paris,* "Space-time quantile surfaces of non- stationary random fields: a comparison study", BSG Proceedings 23, Geometry Balkan Press, Bucharest, 2016, pp. 68–74.
[6] *Y. Kim and M. O'Kelly,* "A bootstrap based spacetime surveillance model with an application to crime occurrences", Journal of Geographical Systems, Volume 10, 2008, pp. 141-165.
[7] *S. Shekhar, Z. Jiang, R. Y. Ali, E. Eftelioglu, X. Tang, V. M. V. Gunturi and X. Zhou,* "Spatio-temporal data mining: a computational perspective", International Journal of geo-information, Volume 4, 2015, 2306–2338.
[8] *B. Efron,* "Bootstrap methods: Another look at jackknife". The Annals of Statistics, vol. 7, 1979, pp. 1–26.
[9] *H. R. Künsch,* "The jackknife and the bootstrap for general stationary observations", The Annals of Statistics, vol. 17, 1989, pp. 1217–1261.
[10] *T. Gneiting,* "Nonseparable, stationary covariance functions for space-time data", Journal of the American Statistical Association, vol. 97 (458), 1979, pp. 590– 600.