Composite likelihood methods for space-time data

M. Bevilacqua\(^1\), C. Gaetan\(^2\), E. Porcu\(^3\) and J. Mateu\(^3\)

\(^1\) Dipartimento di Scienze Statistiche, Università di Padova, Italy.
\(^2\) Dipartimento di Statistica, Università Ca’ Foscari di Venezia, Italy.
\(^3\) Departamento de Matemáticas, Universitat Jaume I, Castellon, Spain.

Abstract: We propose two methods based on composite likelihood that can be used to fit space-time data. We perform an intensive simulation study in order to evaluate the performances of the proposed methods, by keeping a constant comparison with well known methods proposed by the Geostatistical literature. The methods we propose are shown to have considerable computational gains.

Keywords: Composite likelihood; Space-time covariance estimation

1 Introduction

Estimation of the covariance structure is fundamental to Geostatistics. The classical estimation methods performances depend on the configuration of the spatial data set. The methods of ML or REML theoretically yield consistent estimators but require a full specification of the probabilistic model and involve inversion of large matrices which request \(O(n^3)\) operations with \(n\) spatial data. Regarding WLS, the method does not require a probabilistic specification for the underlying spatial structure and is computationally easier than likelihood methods. Unfortunately, it depends on the method of moments estimates, that is very sensitive to the choice of the lag distance as well as the tolerance regions.

As far as space-time data is concerned, WLS methods have been applied (Cressie, Huang 1999; Gneiting, 2002) to fit spatio-temporal covariances. Owing to complexity in computation, full likelihood has not been used in the space-time framework, being \(O(n^3t^3)\) the order of computation.

Composite likelihood (Lindsay, 1988) indicates a general class of pseudo-likelihoods based on likelihood of marginal or conditional events. The motivation behind the introduction of composite likelihood methods is that there are a number of situations, such as is the case of spatio-temporal models, where the computation of the full likelihood is very difficult and too time-consuming. We believe that composite likelihood (CL, for short) could be a good alternative in order to overcome the already mentioned computational problems.
for space-time data. We present an intensive simulation study, considering a constant comparison with the other classical estimation methods. In Section 2 we propose two methods, based on CL, to fit space-time data. We show that the methods offer reasonably good estimates and considerable computational gains. Theoretical properties of these CL estimators are not inspected in this work, and it is clearly the next step to be faced in our research.

2 Two possible composite likelihood approaches for space-time data

The first approach follows Curriero and Lele (1999) and extends their methods to the spatio-temporal context. Let $Z(s_i, t), i = 1, \ldots, n, t = 1, \ldots, T$ be a realization of a spatio-temporal process $Z(s, t), s \in R^2$ and $t \in R^1$.

We make the intrinsic stationary hypothesis:

$$E(Z(s, t)) = \mu$$
$$\text{Var}(Z(s_i, t) - Z(s_j, t')) = 2\gamma(s_i - s_j, t - t')$$

Also, we assume that:

$$U(i, j, t, t') = Z(s_i, t) - Z(s_j, t') \sim N(0, 2\gamma(s_i - s_j, t - t'))$$

Let $\gamma(s_i - s_j, t - t'; \theta)$ be a valid spatio-temporal variogram model where $\theta \in \Theta$ is the parameter to estimate. Our goal is to estimate $\theta$ through a CL method. Thus, we consider the sum of the marginal log-density of $U_i, j, t, t'$, namely:

$$CL(\theta) = \sum_{t = 1}^{T} \sum_{t' > t}^{T} \sum_{i = 1}^{n} \sum_{j > i}^{n} l_{i, j, t, t'}(\theta) + \sum_{t = 1}^{T} \sum_{t' > t}^{T} \sum_{i = 1}^{n} \sum_{j > i}^{n} l_{i, i, t, t'}(\theta)$$

where the log-negative likelihood is, up to a constant:

$$l_{i, j, t, t'}(\theta) \propto \log(\gamma(s_i - s_j, t - t'; \theta)) + \frac{1}{2} \frac{U(i, j, t, t')^2}{\gamma(s_i - s_j, t - t'; \theta)}$$

This is a composite likelihood because each of the components $l_{i, j, t, t'}(\theta)$ is a legitimate likelihood. Estimates can be obtained by maximising $CL(\theta)$ with respect to $\theta$. We note that the method does not request any inversion matrix though it involves many sums. The number of operation is $O(n^2 T^2)$. Properties of these estimates are to be investigated.

The second approach is suitable for regular monitoring data with respect to time. Let $Z_t = (Z(t, s_1), \ldots, Z(t, s_n))'$, $t = 1, \ldots, T$ and we assume that
\[
\begin{pmatrix}
Z_t \\
Z_{t-1}
\end{pmatrix}
\sim N\left[
\begin{pmatrix}
\mu_t \\
\mu_{t-1}
\end{pmatrix},
\begin{pmatrix}
V_t & C_{t,t-1} \\
C_{t,t-1} & V_{t-1}
\end{pmatrix}
\right]
\]

where \( E(Z_t) = \mu_t \), \( \text{var}(Z_t) = V_t \) and \( \text{cov}(Z_t, Z_{t-1}) = C_{t,t-1} \). We denote \( l_t \) the logarithm of the density of \((Z_t, Z_{t-1})\) and we define the composite likelihood

\[
CL(\theta) = \sum_{t=1}^{T} l_t
\]

where \( \theta \) is the vector of the unknown parameters.

In this case we have to invert a \( 2n \times 2n \) matrix for \( T \) times, so the complexity is \( O(Tn^3) \). Thus the method is useful for large time and few spatial observation. Clearly, properties of the estimates are to be investigated.

References


