Combining Euclidean and composite likelihood for binary spatial data estimation

(Article begins on next page)

12 October 2024

Combining Euclidean and composite likelihood for binary spatial data estimation

Moreno Bevilacqua · Federico Crudu · Emilio
Porcu

Abstract In this paper we propose a blockwise Euclidean likelihood method for the estimation of a spatial binary field obtained by thresholding a latent Gaussian random field. The moment conditions used in the Euclidean likelihood estimator derive from the score of the composite likelihood based on marginal pairs. A feature of this approach is that it is possible to obtain computational benefits with respect to the pairwise likelihood depending on the choice of the spatial blocks. A simulation study and an analysis on cancer mortality data compares the two methods in terms of statistical and computational efficiency. We also study the asymptotic properties of the proposed estimator.

Keywords Binary Spatial Fields · Latent Gaussian fields · Euclidean Likelihood · Composite Likelihood

1 Introduction

Estimation of the dependence structure in spatial processes has drawn the attention of many scientists in the last twenty years. Random fields are the mathematical framework for statistical analysis of spatial data and parametric inference based on the maximum likelihood (ML) method is generally considered the best option for the estimation of spatial dependence under this framework. However, fitting models through ML may be troublesome. In the case of Gaussian fields, for instance, ML is impractical when the dataset is large, since its evaluation requires a computational burden of the order $O(n^3)$, where n is the number of spatial locations. This problem is exacerbated in the estimation of the covariance model associated to multivariate random fields (Porcu et al. (2013); Gneiting et al. (2010)). For this reason, in the last years there has been a growing interest in proposing new computationally feasible methods for estimating spatial Gaussian data. In particular, the objective is to find methods that allow for a reasonable compromise between statistical and computational efficiency. See Rue & Tjelmeland (2002) and the survey by Sun et al. (2012) for an exhaustive illustration of the problem.

Working under the assumption of Gaussianity of the underlying spatial field simplifies things considerably, since one only has to focus on the second-order structure of the field. Unfortunately, many applications to ecology and epidemiology show that the assumption of Gaussianity is often violated (see for example Adler (2008)). A widely discussed example is the case of binary spatial data. For instance, P. Lin & Clayton (2005) proposed a quasi-likelihood approach while Albert & McShane (1995) consider a generalized estimating equation approach in the context of generalized linear models. Sherman et al. (2006) in the context of spatial autologistic models proposed the use of composite and Monte Carlo ML estimation. In the geostatistical framework, the case of binary data has been faced through several alternative approaches to the classical kriging predictor, such

Universidad de Valparaiso, Departamento de Estadística

E-mail: moreno.bevilacqua@uv.cl · Pontificia Universidad Católica de Valparaíso, Instituto de Estadística. E-mail: federico.crudu@gmail.com · Universidad Técnica Federico Santa Maria, Departamento de Matemáticas. E-mail: emilio.porcu@usm.cl

as indicator (Journel (1983)) or disjunctive kriging (Matheron (1976)), or again trans-Gaussian kriging (Cressie (1993)).

Heagerty & Lele (1998), instead, following the literature on hierarchical models, derive the joint probability model through the specification of a spatial Gaussian latent process. Specifically, a latent Gaussian field is thresholded in order to obtain a binary field. Thus, the finite dimensional distributions can be deduced through algebraic manipulation. Unfortunately, the joint distribution involves a sum of 2^n-dimensional Gaussian integrals. Once again, the evaluation of the likelihood as n increases becomes computationally demanding.

Under this framework, Heagerty & Lele (1998) proposed a special case of composite likelihood (CL) based on pairs, also known as pairwise likelihood (PL). CL is a class of estimating functions that contains, and thus generalizes, ordinary likelihood methods (see Varin et al. (2011) for a complete review). One of the features of CL is that it is possible to perform inference on the parameters at a reduced computational cost. This makes it a useful estimation method when we deal with large datasets. See also Caragea & Smith (2006), Fuentes (2007), Lindsay (1988), Stein (2008), Stein et al. (2004) and Vecchia (1988).

Lunardon et al. (2012), in the IID case, use empirical likelihood (EL) with moment conditions coming from the score function of the PL in order to overcome the problem of non standard asymptotic distribution of the likelihood ratio statistic of the PL. See also Qin & Lawless (1994) for a general reference on EL estimation for moment condition models.

Our proposal is in the same spirit of Lunardon et al. (2012), but with three main differences. First, we consider an Euclidean likelihood (EU) estimator. Like EL, EU estimators are the result of a complicated optimization problem. However, unlike EL, EU does not need a joint numerical solution of the parameter of interest and the auxiliary parameter. In fact, the estimator of the auxiliary parameter has a closed form and the parameters of the model can be estimated by minimizing a simple quadratic form.

Second, we combine the EU estimator with a blockwise approach. The resulting estimator is the natural extension of the EU estimator in the IID case to the dependent data case. See for instance Kitamura (1997) in the context of time series and Nordman & Caragea (2008) for spatial variogram estimation.

Third, the aim of considering a blockwise Euclidean likelihood is to further reduce the computational cost associated to PL estimation.

The paper is organized as follows. In Section 2, PL and EU estimation methods are briefly reviewed while in Section 3 we describe the spatial binary random field model obtained through a latent Gaussian field. The proposed blockwise EU method with moment condtions coming from the PL approach is discussed in Section 4. In Section 5 we study the asymptotic properties of the proposed estimator while in Section 6 a simulation study compares the PL estimator with the proposed method. Section 7 contains an application to cancer mortality data. Finally, some concluding remarks are given in Section 8.

2 Composite and Euclidean Likelihoods

2.1 Composite Likelihood

CL is a class of estimating functions that contains, and thus generalizes, standard likelihood methods. One of the features of CL is that it is possible to perform inference on the parameters at a reduced computational cost. The most successful case of the CL class is probably the PL function. (see for instance Bevilacqua & Gaetan (2014), Padoan et al. (2010) and Heagerty & Lele (1998)).

Let $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^{\intercal}$ be a realization from a n dimensional random vector with joint distribution $f(\boldsymbol{z}; \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^v$. Let $L_{ij} := L(\theta; A_{ij})$ be the likelihood associated to the probability density function of the pairs $A_{ij} = (Z_i, Z_j)^{\intercal}$. Then the weighted PL function is defined as:

$$PL(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} l_{ij}(\boldsymbol{\theta}) w_{ij}, \tag{1}$$

where $l_{ij}(\cdot) := \log(L_{ij}(\cdot))$ and w_{ij} are positive suitable weights that do not depend on $\boldsymbol{\theta}$. The maximum PL estimate is given by $\widehat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}}\, PL(\boldsymbol{\theta})$.

A distinctive feature of PL is that the associated pairwise score function, $\nabla PL(\boldsymbol{\theta})$, is unbiased, irrespective of the distributional assumptions on the pairs. Here, $\nabla PL(\cdot)$ denotes the gradient vector obtained through partial differentiation with respect to the elements of the vector $\boldsymbol{\theta}$. In the binary case and under some specific conditions on the underlying random field, Heagerty & Lele (1998) show that the PL estimator is consistent and asymptotically Gaussian with covariance matrix equal to the inverse of the Godambe information matrix

$$\boldsymbol{V}(\boldsymbol{\theta}) = \boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{J}(\boldsymbol{\theta})^{-1}\boldsymbol{H}(\boldsymbol{\theta})^{\intercal}, \tag{2}$$

where

$$\boldsymbol{H}(\boldsymbol{\theta}) = -\,\text{E}[\nabla^2 PL(\boldsymbol{\theta})], \qquad \boldsymbol{J}(\boldsymbol{\theta}) = \text{E}[\nabla PL(\boldsymbol{\theta})\nabla PL(\boldsymbol{\theta})^{\intercal}], \tag{3}$$

with $\nabla^2 PL(\theta)$ being the matrix of second derivatives.

2.2 Euclidean likelihood

EU is a statistical method for estimation and inference based on the optimization of a discrepancy function given a set of constraints (see e.g. Owen 2001). This approach has the same spirit as EL (Owen 1988, 1990, 1991, 2001) as it only changes in the specification of the discrepancy function. Similarly to EL, EU allows estimation and likelihood-like inference without the need of specifying a parametric likelihood function. However, while the EL estimator is computed numerically jointly with the estimation of an auxiliary parameter (vector), the EU estimator is computationally more tractable as the auxiliary parameter has a closed form solution and the parameter of interest is calculated via the optimization of a quadratic form.

Let $\boldsymbol{Z}_i = (Z_{1i}, \ldots, Z_{ri})^{\intercal}$, $i = 1, \ldots, n$ be a IID sample as before, and $\boldsymbol{g} : \mathbb{R}^r \times \Theta \to \mathbb{R}^q$ a known vector–valued function for which there exists a fixed $\boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}^v$ such that

$$E[\boldsymbol{g}(\boldsymbol{Z}_i, \boldsymbol{\theta}_0)] = \boldsymbol{0}. \tag{4}$$

The model can be estimated via EU by taking the sample analogue of the expectation and by solving the associated constrained optimization problem. The EU optimization problem is characterized by the following Lagrangian function

$$L = -\frac{1}{2}\sum_{i=1}^{n}(n\pi_i - 1)^2 - n\boldsymbol{\lambda}^{\intercal}\sum_{i=1}^{n}\pi_i\boldsymbol{g}(\boldsymbol{Z}_i, \boldsymbol{\theta}) + \mu\Big(\sum_{i=1}^{n}\pi_i - 1\Big). \tag{5}$$

By taking the first order conditions and standard calculations we can find an expression for the weights π_i as

$$\pi_i(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \frac{1}{n}\Big(1 - \boldsymbol{\lambda}^{\intercal}(\boldsymbol{g}(\boldsymbol{Z}_i, \boldsymbol{\theta}) - \widehat{\boldsymbol{g}}(\boldsymbol{\theta}))\Big), \tag{6}$$

where $\widehat{\boldsymbol{g}}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\boldsymbol{g}(\boldsymbol{Z}_i, \boldsymbol{\theta})/n$, and a close form solution for the auxiliary parameter $\boldsymbol{\lambda}$ as

$$\boldsymbol{\lambda} = \widehat{\boldsymbol{S}}(\boldsymbol{\theta})^{-1}\widehat{\boldsymbol{g}}(\boldsymbol{\theta}) \tag{7}$$

where $\widehat{\boldsymbol{S}}(\boldsymbol{\theta}) = \sum_{i=1}^{n}(\boldsymbol{g}(\boldsymbol{Z}_i, \boldsymbol{\theta}) - \widehat{\boldsymbol{g}}(\boldsymbol{\theta}))(\boldsymbol{g}(Z_i, \boldsymbol{\theta}) - \widehat{\boldsymbol{g}}(\boldsymbol{\theta}))^{\intercal}/n$. By plugging in Equations (6) and (7) into the objective function of the Lagrangian (5), we obtain a quadratic form in $\boldsymbol{\theta}$:

$$\frac{1}{2}\sum_{i=1}^{n}(n\pi_i - 1)^2 = \frac{n}{2}\widehat{\boldsymbol{g}}(\boldsymbol{\theta})^{\intercal}\widehat{\boldsymbol{S}}(\boldsymbol{\theta})^{-1}\widehat{\boldsymbol{g}}(\boldsymbol{\theta})$$

An estimator for $\boldsymbol{\theta}$ is found by minimizing the above quadratic form

$$\widehat{\boldsymbol{\theta}}_{EU} = \arg\min_{\boldsymbol{\theta}}\widehat{\boldsymbol{g}}(\boldsymbol{\theta})^{\intercal}\widehat{\boldsymbol{S}}(\boldsymbol{\theta})^{-1}\widehat{\boldsymbol{g}}(\boldsymbol{\theta}). \tag{8}$$

3 Binary Spatial Data based on latent models

This section is largely expository and readapts to our case the approach proposed in Heagerty & Lele (1998). In the following, we obtain a spatial field by thresholding a Gaussian field $Z(\boldsymbol{s}) + \epsilon(\boldsymbol{s})$, $\boldsymbol{s} \in \mathbb{R}^d$. For a given threshold $c > 0$, we have that

$$Y(\boldsymbol{s}) = \mathbb{I}\left(Z(\boldsymbol{s}) + \epsilon(\boldsymbol{s}) > c\right), \qquad \boldsymbol{s} \in \mathbb{R}^d$$

is a binary spatial field. Here, $\mathbb{I}(A)$ defines the indicator function of the set A, whilst $\epsilon(\cdot)$ is a spatial white noise with zero mean and covariance function $\operatorname{cov}\left(\epsilon(\boldsymbol{s}), \epsilon(\boldsymbol{s}')\right) = \tau^2 \mathbb{I}(\{\boldsymbol{s} = \boldsymbol{s}'\})$. Throughout, we assume that $Z(\cdot)$, being independent of $\epsilon(\cdot)$, is a weakly stationary and Gaussian field, with a given mean $\mu \in \mathbb{R}$ and with a stationary and isotropic covariance function belonging to a certain parametric family. That is,

$$\operatorname{cov}\left(Z(\boldsymbol{s}), Z(\boldsymbol{s}')\right) = \sigma^2 \rho\left(\|\boldsymbol{s} - \boldsymbol{s}'\|; \alpha\right), \tag{9}$$

with $\|\cdot\|$ denoting the Euclidean seminorm, and where σ^2 is the variance of Z, α is a scaling parameter and ρ is a correlation function so that it is identically equal to one at the origin ($\boldsymbol{s} = \boldsymbol{s}'$). The parameter τ^2 allows for accounting the presence of microscale effects and it is commonly known as nugget effect in geostatistics. Since the evaluation of likelihood involves a sum of 2^n dimensional Gaussian integrals, Heagerty & Lele (1998) propose to use PL for estimating the parameters of the latent spatial process Z. First, notice that the marginal and bivariate distributions can be explicitly calculated. It is easy to show that

$$\mathbb{P}\left(Y(\boldsymbol{s}) = 1\right) = \Phi\left(\frac{\mu - c}{\tilde{\sigma}}\right) =: p, \tag{10}$$

where Φ denotes the standard Gaussian distribution and $\tilde{\sigma} := \sqrt{\sigma^2 + \tau^2}$. Also, we have

$$\mathbb{P}\left(Y(\boldsymbol{s}) = a \bigcap Y(\boldsymbol{s}') = b\right) =: p_{ab}, \quad a, b = 0, 1 \tag{11}$$

and

$$p_{11} = \Phi_2\left(\frac{\mu - c}{\tilde{\sigma}}, \frac{\mu - c}{\tilde{\sigma}}, \frac{\sigma^2 \rho\left(\|\boldsymbol{s} - \boldsymbol{s}'\|; \alpha\right)}{\tilde{\sigma}^2}\right), \tag{12}$$

with Φ_2 denoting the bivariate standard Gaussian distribution. In order to solve identifiability problems, Heagerty & Lele (1998) adopt a special parameterization that allows to consider $\sigma^2 \in (0, 1]$ and $\tau^2 = 1 - \sigma^2$, with $c = 0$. Such a result is obtained by standardizing Z and ϵ, and the reader is referred to Heagerty & Lele (1998) for the details. In this case, $\boldsymbol{\theta} = (\mu, \sigma^2, \alpha)^\intercal$ and the PL function is as in (1) with:

$$l_{ij}(\boldsymbol{\theta}) = z_i z_j \log(p_{11}) + (z_j(1 - z_i) + z_i(1 - z_j)) \log(p - p_{11}) + (1 - z_i)(1 - z_j) \log(1 - 2p + p_{11}). \tag{13}$$

4 Blockwise spatial Euclidean likelihood

In recent years a number of papers have discussed the possibility of adapting the EL framework to the case of dependent data. For example, Kitamura (1997) proposes a blockwise EL estimator for estimating equations in the presence of weakly dependent data. L. Lin & Zhang (2001) analyse a similar problem for EU.

Our proposal combines pairwise and Euclidean likelihood estimation methods for the estimation of binary fields in the following way: we suppose the latent spatial process $Z(\boldsymbol{s})$ is observed at n locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ that belong to the sampling region $\mathcal{R}_n \subset \mathbb{R}^d$, $d \geq 1$. Given an increasing sequence $\{\gamma_n\}$ of scaling factors, we generate the sampling space $\mathcal{R}_n$ by inflating a prototype region $\mathcal{R}_0 \subset (-1/2, 1/2]^d$ by the factor γ_n, such that $\mathcal{R}_n = \gamma_n \mathcal{R}_0$. With respect to this scheme the

sampling locations are $\{s_1, \ldots, s_n\} = \mathcal{R}_n \cap \mathbb{Z}^d$. Hence, a set of moment conditions for a known vector-valued function of (subsets) $\widetilde{Z}$ of the observations $Z := (Z(s_1), \ldots, Z(s_n))^\intercal$ are defined as

$$E[g(\widetilde{Z}, \theta_0)] = 0.$$

We now set $\widetilde{Z}_{ij}$ for the generic pair $(Z(s_i), Z(s_j))^\intercal$ and we choose $g(\widetilde{Z}_{ij}, \theta) = \nabla l_{ij}(\theta) w_{ij}$ where $\nabla l_{ij}(\theta) w_{ij}$ is the pairwise weighted score associated to $l_{ij}(\theta) w_{ij}$ and $l_{ij}(\theta)$ is as in (13). Then, we consider the following moment conditions

$$E[g_{ij}(\theta_0)] = 0, \quad \theta_0 \in \Theta. \tag{14}$$

where $g_{ij}(\theta) = g(\widetilde{Z}_{ij}, \theta)$. We implement a blockwise approach similar to Nordman & Caragea (2008), who deal with the case of EL in the presence of spatial data. Data blocking is a nonparametric device that allows us to model the dependence in the data without particular assumptions on the structure of the dependence. This approach translates into a modification of the moment conditions in (14).

Let us now consider a sequence of positive integers $\{b_n\}$ and define a d-dimensional overlapping block as $\mathcal{B}_{b_n}(\kappa) = \kappa + b_n(-1/2, 1/2]^d$, with $\kappa \in \mathbb{Z}^d$. We require that b_n grow at a slower rate than γ_n, i.e. $b_n^2/\gamma_n \to 0$. Subsequently, we are being intentionally sloppy in omitting the subindex n for the following quantitities: this allows to simplify considerably the notation.

Then, a blockwise analogue of Equation (14) is

$$E[m_\kappa(\theta_0)] = 0, \quad \theta_0 \in \Theta. \tag{15}$$

where the moment conditions $m_\kappa(\theta)$ are defined as

$$m_\kappa(\theta) = \frac{1}{b^d} \sum_{\{i,j\} \in \mathcal{A}_{ij\kappa}} g_{ij}(\theta), \tag{16}$$

with $\mathcal{A}_{ij\kappa} := \{(i,j) : (s_i, s_j) \in \mathcal{B}_b(\kappa) \cap \mathbb{Z}^d, i = 1, \ldots, n_\kappa - 1, j = i + 1, \ldots, n_\kappa\}$. Here n_κ is the number of location sites in the block κ. The parameter b is the side length of the spatial block and can be viewed as a bandwidth parameter. Suppose we dispose of N blocks. Then the spatial blockwise Euclidean likelihood (denoted SBEU) estimator is the result of the following constrained optimization problem

$$L = -\frac{1}{2} \sum_\kappa (N\pi_\kappa - 1)^2 - N\lambda^\intercal \sum_\kappa \pi_\kappa m_\kappa(\theta) + \mu \left(\sum_\kappa \pi_\kappa - 1 \right). \tag{17}$$

An estimator for the parameters of interest θ can be obtained in the same way as in the random sample case, this is

$$\widehat{\theta}_{\mathrm{SBEU}} = \arg \min_\theta Q_n(\theta) \tag{18}$$

where

$$Q_n(\theta) = \widehat{m}(\theta)^\intercal \widehat{\Sigma}(\theta)^{-1} \widehat{m}(\theta), \tag{19}$$

and

$$\widehat{\Sigma}(\theta) = b^d \sum_\kappa (m_\kappa(\theta) - \widehat{m}(\theta))(m_\kappa(\theta) - \widehat{m}(\theta))^\intercal / N \tag{20}$$

where $\widehat{m}(\theta) = \sum_\kappa m_\kappa(\theta)/N$. In the following section we give a consistency and asymptotic normality result for the estimator in (18).

5 Asymptotic Results

We require the following assumptions:

C1 The true parameter $\boldsymbol{\theta}_0$ is an interior point of the compact set Θ and $\widehat{\boldsymbol{m}}(\boldsymbol{\theta})$ is twice continuously differentiable in a neighbourhood of $\boldsymbol{\theta}_0 \in \Theta$.

C2 $E[\|\nabla^\ell \boldsymbol{m}_{\boldsymbol{\kappa}}(\boldsymbol{\theta})\|^{2+\eta}] < \infty$, where ∇^ℓ is the ℓ-th derivative of $\boldsymbol{m}_{\boldsymbol{\kappa}}(\cdot)$ with respect to $\boldsymbol{\theta}$ with $\ell = 0, 1, 2$ and $\eta > 0$; $\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla^\ell \widehat{\boldsymbol{m}}(\boldsymbol{\theta}) - \nabla^\ell \boldsymbol{m}(\boldsymbol{\theta})\| \to_p 0$, $\ell = 0, 1$, where $\nabla^\ell \boldsymbol{m}(\boldsymbol{\theta}) = E[\nabla^\ell \boldsymbol{m}_{\boldsymbol{\kappa}}(\boldsymbol{\theta})]$; $\nabla \boldsymbol{m}(\boldsymbol{\theta}_0)$ is full column rank; $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0) \to_p \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$, a positive definite matrix; $b^{-1} + \frac{b^{2d}}{n} \to 0$ as $n \to \infty$.

C3 (CLT): $\sqrt{n}\widehat{\boldsymbol{m}}(\boldsymbol{\theta}_0) \to_d N(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0))$.

Some comments are in order: the mixing conditions are implicitly defined on the type of sets $\mathcal{R}_n$ as defined in Section 2, and we refer the reader to Nordman & Caragea (2008) and Jenish & Prucha (2009) for more details.

Let us define $Q(\boldsymbol{\theta}) = \boldsymbol{m}(\boldsymbol{\theta})^\intercal \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \boldsymbol{m}(\boldsymbol{\theta})$ and state the following theorems.

Theorem 1 *Assume that* **C1** *to* **C3** *hold and that*

$$\boldsymbol{\theta}_0 = \arg \min_{\boldsymbol{\theta} \in \Theta} Q(\boldsymbol{\theta})$$

is unique. Then, $\widehat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$ *and* $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to_d N(\mathbf{0}, \boldsymbol{\Omega}(\boldsymbol{\theta}_0))$.

Theorem 2 *Assume that* **C1** *and* **C2** *hold. Then,* $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}}) \to_p \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$.

Proofs are deferred to the appendix.

6 Simulations

The aim of this section is to compare the PL estimator with the estimator proposed in Section 4 from statistical and computational point of view. In order to do so, we work under two spatial sampling schemes:

A. a regular grid on $[-15, 15]^2$ with unit spacing (i.e. $n = 961$ location sites);
B. an irregular grid, namely $n = 480.5 \times 2^\lambda$ location sites uniformly distributed on $[-15, 15]^2$ with $\lambda = 1$ (i.e. $n = 961$ location sites).

Under these two spatial settings, we perform 1000 simulations of a binary spatial random field $Y(\boldsymbol{s})$ generated by thresholding a latent Gaussian field $Z(\boldsymbol{s})$, with a given mean μ and a correlation function $\rho(\cdot; \alpha)$ of the exponential type, i.e.

$$\text{corr}\,(Z(\boldsymbol{s}), Z(\boldsymbol{s}')) = \rho(\|\boldsymbol{s} - \boldsymbol{s}'\|; \alpha) = \exp\left(-3\frac{\|\boldsymbol{s} - \boldsymbol{s}'\|}{\alpha}\right). \tag{21}$$

The correlation function is parameterized in such a way that α corresponds to the practical range of ρ. The latent Gaussian field $Z(\boldsymbol{s})$ was simulated through Cholesky decomposition. Note that we assume $\sigma^2 = 1$ and $\tau^2 = 0$. The mean μ determines the marginal probability of success (denoted p in Section 3) and the scale parameter α has a direct impact on the joint probabilities of success (denoted p_{11}). We combine two scenarios corresponding to increasing marginal probability of success given by the mean parameter μ and increasing spatial dependence corresponding to different values of the scale parameter α. We consider the following combinations of parameters:

1. let $\mu = -1, 0, 1$, corresponding to marginal probabilities of success of approximately 0.16, 0.50, and 0.84 respectively with $\alpha = 3$;
2. let $\mu = -1, 0, 1$, corresponding to marginal probabilities of success of approximately 0.16, 0.50, and 0.84 respectively with $\alpha = 5$.

For each scenario we estimate the parameters μ and α with PL as in Equation (1) with l_{ij} as defined in Equation (13), using cutoff weights $w_{ij} = 1$ whenever $\|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq d$, and 0 otherwise. When $d = 1$ we consider the weighted version of PL. On the other hand, when $d = +\infty$ we take its unweighted version. The choice of compactly supported weights leads to computational benefits and may improve the statistical efficiency of the PL estimator for the covariance parameters as shown, for example, in Bevilacqua et al. (2012) in the Gaussian case.

We estimate μ and α with SBEU using the same weights. The choice of the block length in the SBEU estimation method can obviously affect the efficiency of the estimator. Following Lee & Lahiri (2002) we choose the spatial blocks by considering both overlapping and non overlapping translates of the set $[C\sqrt{\gamma}, C\sqrt{\gamma}]^2$ where C is a positive constant. This choice comes from Sherman (1996), who shows that the dimension of the optimal block should be proportional to the square root of γ. We choose $C = 0.3651484, 0.7302967, 1.460593$. This is equivalent to choose square blocks of length respectively $b = 2, 4$ and 8. In the literature, the overlapping case has been shown to have a better behavior in terms of statistical efficiency (see Lahiri (1999) for some results about the block bootstrap). Note that when we apply the blockwise scheme a constant ℓ must be specified. The parameter ℓ governs the degree of overlapping of the blocks along the field. A possible choice is $\ell = bp$ where $0 < p \leq 1$. Figure 1 shows how overlapping and non overlapping blocking is performed in practice when $b = 4$ under scenarios A and B and fixing $p = 0.25$ that is $\ell = 1$. Note that when $p = 1$ we are considering the non overlapping case.

Tables 1 to 4 report the results in terms of simulated relative efficiency with respect to the PL. That is, the ratio between the simulated mean square error of the PL and that of the other estimators. Table 1 and 2 show the relative efficiency under scenario A1 and A2 respectively, while Table 3 and 4 show the relative efficiency under scenario B1 and B2.

In general, the PL estimator performs slightly better than the SBEU estimator. Note that for each scenario we considered, the smallest block length (equal to 2) performs better with respect to the other possible choices. When increasing the spatial dependence i.e. increasing α the simulated MSE increases, as expected, for both estimation methods. Moreover, note that the MSE is lower when the marginal probability of success is 0.5 (i.e. the case $\mu = 0$) with respect to the case $\mu = -1, 1$. This is consistent with the results of Heagerty & Lele (1998). Comparing the fully overlapping and the non overlapping case there is, as expected, a general improvement in terms of the simulated MSE when using the fully overlapping setting. Nevertheless, the improvement is not so clear when the block length is equal to 2.

Finally, in order to compare the computational performance of the PL and SBEU estimation methods we consider the scenario B with an increasing $\lambda = 1, \ldots, 6$. That is, we are considering $n = 961, 1922, \ldots, 30752$ locations sites in the unit square. In table 5, for each n, we report the ratio between the time needed to evaluate the PL function and times needed for the other methods specifically weighted PL and SBEU with $b = 2, 4, 8$ in its weighted and unweighted version and using overlapping and non overlapping blocks. The time, in seconds, is obtained considering the elapsed time of the R function system.time(). As expected PL in its weighted version performs better then PL (approximately 46 times much faster when $n = 30752$). Overall SBEU in its non overlapping version is much faster then the overlapping version as expected. Moreover the SBEU in its weighted version is faster than the unweighted version. The best performance is obtained with SBEU in its overlapping and non overlapping version. For instance with $b = 2$, $d = 1$, and considering the non overlapping case the time needed to evaluate the SBEU function is 255 times faster then the PL and 5 times faster then the weighted PL. In general, the computational cost associated to PL is of order $O(n^2)$, where n is the number of location sites. Suppose that the n observations are divided into N blocks with associated blocklength b. The order of computation of each block is $O(b^{2d})$. Therefore, the order of computation of the SBEU estimator is of order $O(b^{2d}N)$. We assume that $b^2/\gamma \to 0$, which implies that $b^{2d} = o(n)$ (see Nordman & Caragea (2008)). Hence, if $N = O(n)$, the order of computation of the SBEU estimator is of order $o(n^2)$. This suggests that the SBEU estimator enjoys some computational advantage over the PL estimator both in the weighted and case. Thus, although we notice a small loss of statistical efficiency in the simulations, this may be balanced by the computational gains described above.

| | | | $d = 1$ | | | | $d = +\infty$ | |
| | | PL | | SBEU | | PL | | SBEU | |
			$b = 2$	$b = 4$	$b = 8$		$b = 2$	$b = 4$	$b = 8$
$\mu = -1$	μ	0.980	0.909	0.830	0.686	1.000	0.912	0.837	0.693
			(0.947)	(0.839)	(0.638)		(0.916)	(0.797)	(0.624)
	α	1.084	1.019	0.909	0.763	1.000	1.157	1.009	0.814
			(1.033)	(0.867)	(0.654)		(1.141)	(0.893)	(0.622)
$\mu = 0$	μ	0.977	0.903	0.825	0.684	1.000	0.906	0.831	0.692
			(0.953)	(0.831)	(0.623)		(0.937)	(0.798)	(0.615)
	α	1.460	1.369	1.251	1.054	1.000	1.369	1.117	0.855
			(1.334)	(1.184)	(0.873)		(1.325)	(1.013)	(0.686)
$\mu = 1$	μ	0.978	0.904	0.821	0.677	1.000	0.910	0.830	0.691
			(0.935)	(0.824)	(0.604)		(0.888)	(0.784)	(0.582)
	α	1.207	1.128	1.034	0.850	1.000	1.216	1.061	0.862
			(1.143)	(0.990)	(0.697)		(1.181)	(0.930)	(0.692)

Table 1: Simulated relative efficiency (with respect to the PL) of SBEU estimator and weighted PL when $\alpha = 3$ under the setting A1. For SBEU estimator relative efficiency is presented for different values of the block length and for its overlapping and non overlapping (in parentheses) version.

| | | | $d = 1$ | | | | $d = +\infty$ | |
| | | PL | | SBEU | | PL | | SBEU | |
			$b = 2$	$b = 4$	$b = 8$		$b = 2$	$b = 4$	$b = 8$
$\mu = -1$	μ	0.982	0.912	0.836	0.697	1.000	0.911	0.839	0.701
			(0.953)	(0.860)	(0.658)		(0.936)	(0.849)	(0.654)
	α	1.299	1.220	1.135	0.969	1.000	1.187	1.034	0.855
			(1.253)	(1.098)	(0.819)		(1.189)	(0.903)	(0.642)
$\mu = 0$	μ	0.977	0.905	0.828	0.685	1.000	0.907	0.834	0.693
			(0.952)	(0.846)	(0.640)		(0.945)	(0.832)	(0.646)
	α	1.516	1.425	1.230	1.059	1.000	1.418	1.148	0.897
			(1.446)	(1.247)	(0.886)		(1.417)	(1.055)	(0.675)
$\mu = 1$	μ	0.978	0.905	0.818	0.669	1.000	0.907	0.821	0.675
			(0.947)	(0.838)	(0.615)		(0.928)	(0.825)	(0.614)
	α	1.379	1.269	1.165	0.991	1.000	1.256	1.081	0.918
			(1.332)	(1.150)	(0.819)		(1.255)	(0.972)	(0.703)

Table 2: Simulated relative efficiency (with respect to the PL) of SBEU estimator and weighted PL when $\alpha = 5$ under the setting A2. For SBEU estimator relative efficiency is presented for different values of the block length and for its overlapping and non overlapping (in parentheses) version.

| | | | $d = 1$ | | | | $d = +\infty$ | |
| | | PL | | SBEU | | PL | | SBEU | |
			$b = 2$	$b = 4$	$b = 8$		$b = 2$	$b = 4$	$b = 8$
$\mu = -1$	μ	0.776	0.470	0.620	0.525	1.000	0.557	0.752	0.639
			(0.354)	(0.671)	(0.478)		(0.428)	(0.829)	(0.610)
	α	0.838	0.463	0.668	0.573	1.000	0.684	1.035	0.852
			(0.241)	(0.666)	(0.489)		(0.290)	(0.778)	(0.591)
$\mu = 0$	μ	0.792	0.494	0.638	0.559	1.000	0.579	0.768	0.678
			(0.395)	(0.676)	(0.499)		(0.458)	(0.831)	(0.620)
	α	1.047	0.562	0.824	0.716	1.000	0.729	1.052	0.853
			(0.348)	(0.802)	(0.605)		(0.396)	(0.891)	(0.648)
$\mu = 1$	μ	0.750	0.453	0.625	0.565	1.000	0.537	0.783	0.705
			(0.368)	(0.651)	(0.492)		(0.425)	(0.834)	(0.650)
	α	0.752	0.432	0.671	0.593	1.000	0.578	0.953	0.870
			(0.266)	(0.619)	(0.462)		(0.300)	(0.728)	(0.604)

Table 3: Simulated relative efficiency (with respect to the PL) of SBEU estimator and weighted PL when $\alpha = 3$ under the setting B1. For SBEU estimator relative efficiency is presented for different values of the block length and for its overlapping and non overlapping (in parentheses) version.
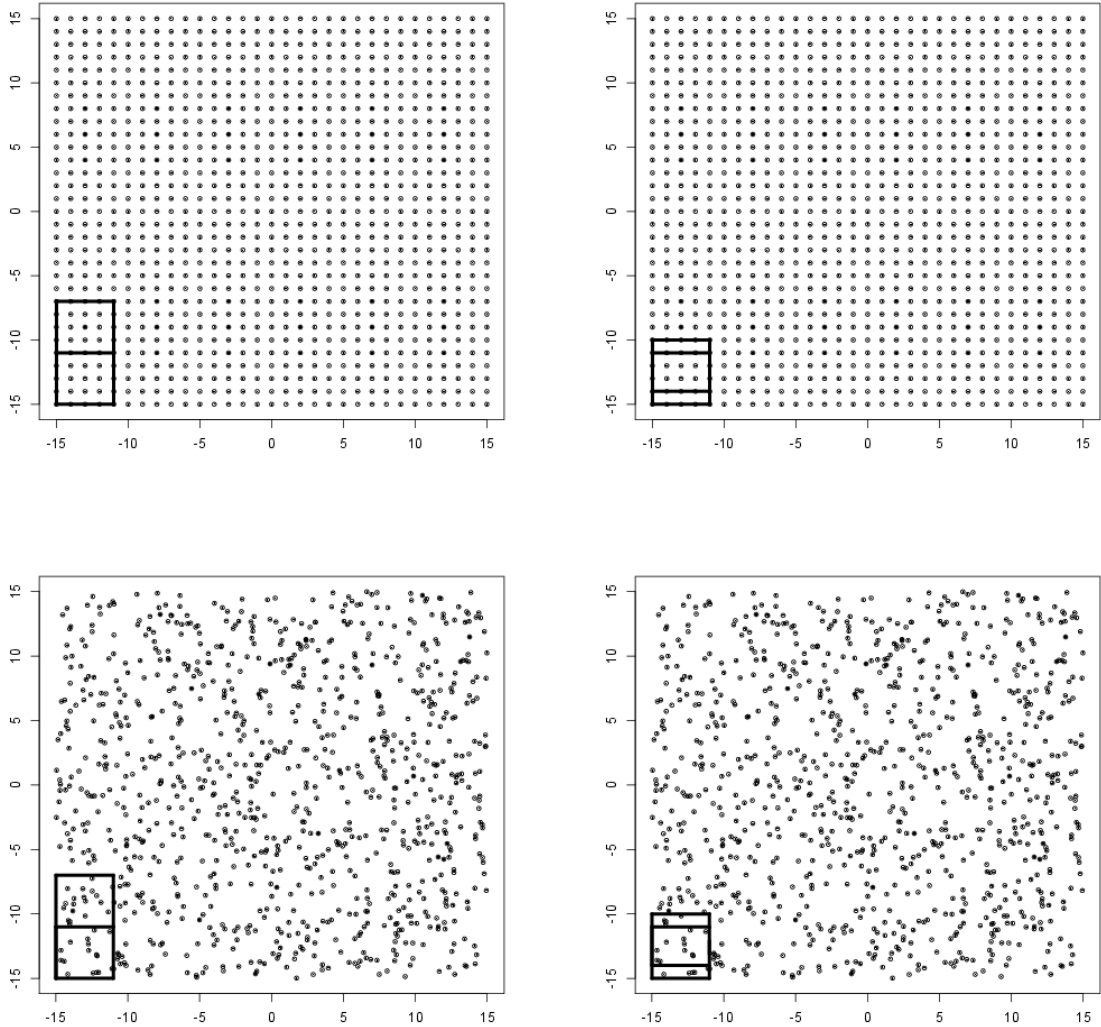
Fig. 1: Non overlapping (left) and overlapping (right) blocks for scenarios A (top) and B (bottom) when $b = 4$.

7 Data Example: Cancer mortality map

We consider the dataset on cancer mortality rates for the United States of Sherman & Carlstein (1994) and re-analyzed by Sherman et al. (2006) and Nordman (2008). The data describes a cancer mortality map shown in Figure 2, constructed using mortality rates from liver and gallbladder cancer in white males during 1950-1959. The sampling region contains 2298 sites on a portion of the integer grid $(0, 66] \times (0, 58] \cap \mathbb{Z}^2$. Here for a given location site s, $Y(s) = 0$ indicates low mortality rate

Sherman et al. (2006) fit an autologistic model employing maximum conditional composite likelihood while Nordman (2008) employs a blockwise EL estimation method. In some sense, our approach combines the previous two approaches since we are considering a specific type of composite likelihood (pairwise) and we are considering blocks in the Euclidean likelihood framework. In order to investigate about the spatial dependence, we consider the lorelogram function (Heagerty & Zeger, 1996). The latter has been introduced in the context of longitudinal data for describing the dependence of binary responses. In this context, the lorelogram plays the same role of the

			$d = 1$				$d = +\infty$		
			PL	SBEU			PL	SBEU	
				$b = 2$	$b = 4$	$b = 8$		$b = 2$	$b = 4$	$b = 8$
$\mu = -1$	μ	0.839	0.575	0.675	0.572	1.000	0.645	0.767	0.650
			(0.472)	(0.738)	(0.535)		(0.541)	(0.850)	(0.633)
	α	0.826	0.452	0.672	0.575	1.000	0.647	0.989	0.858
			(0.225)	(0.641)	(0.467)		(0.315)	(0.761)	(0.636)
$\mu = 0$	μ	0.866	0.618	0.702	0.595	1.000	0.678	0.774	0.658
			(0.551)	(0.772)	(0.574)		(0.605)	(0.856)	(0.644)
	α	1.064	0.602	0.855	0.763	1.000	0.787	1.067	0.878
			(0.363)	(0.811)	(0.630)		(0.414)	(0.895)	(0.633)
$\mu = 1$	μ	0.828	0.567	0.684	0.597	1.000	0.628	0.779	0.681
			(0.506)	(0.742)	(0.560)		(0.546)	(0.847)	(0.664)
	α	0.780	0.434	0.677	0.609	1.000	0.568	0.933	0.904
			(0.238)	(0.656)	(0.516)		0.319)	(0.749)	(0.601)

Table 4: Simulated relative efficiency (with respect to the PL) of SBEU estimator and weighted PL when $\alpha = 5$ under the setting B2. For SBEU estimator relative efficiency is presented for different values of the block length and for its overlapping and non overlapping (in parentheses) version.

		$d = 1$				$d = +\infty$		
	PL	SBEU			PL	SBEU		
		$b = 2$	$b = 4$	$b = 8$		$b = 2$	$b = 4$	$b = 8$
$n = 961$	36.60	10.45	12.33	14.83	1.00	8.01	2.91	1.26
		(43.91)	(121.94)	(121.95)		(84.45)	(40.66)	(13.89)
$n = 1922$	44.94	11.11	12.72	15.58	1.00	7.46	3.19	1.20
		(111.89)	(150.82)	(183.48)		(102.98)	(38.05)	(15.41)
$n = 3844$	45.95	12.21	12.62	14.80	1.00	7.87	3.07	1.16
		(151.28)	(171.23)	(190.47)		(119.33)	(39.79)	(17.24)
$n = 7688$	45.41	14.33	12.94	16.31	1.00	8.84	3.21	1.23
		(206.61)	(176.36)	(196.46)		(132.80)	(40.95)	(17.24)
$n = 15376$	45.62	15.48	14.45	16.04	1.00	9.17	3.23	1.11
		(242.71)	(175.13)	(195.31)		(144.09)	(40.22)	(15.09)
$n = 30752$	46.27	16.23	13.72	15.35	1.00	9.34	3.20	1.12
		(255.10)	(188.67)	(208.33)		(144.50)	(42.35)	(15.14)

Table 5: Ratio between the time needed to evaluate the PL and the other methods when increasing n under scenario B1. For SBEU estimator the ratio is presented for different values of the block length and for its overlapping and non overlapping (in parentheses) version.

variogram in the continuous case (see for instance Kovitz & Christakos (2004)) Specifically, the lorelogram function is defined as the logarithm of the pairwise odds ratio,

$$\pi = \log\left(\frac{p_{11}p_{00}}{p_{10}p_{01}}\right), \tag{22}$$

where p_{ab}, $a, b \in \{0, 1\}$ are as described in (11). For binary data, it provides a better measure of dependence than the covariance function, it is easy to interpret and its upper and lower bounds do not depend on the mean.

We assume the data are a realization from a binary random field generated thresholding a Gaussian random field with mean μ and correlation model as in (21). That is, we specify the marginal probability of success p as in (10) and the joint probability of success p_{11} as in (12). We estimate the parameters of the underlying Gaussian field $(\mu, \alpha)^\intercal$ with weighed PL and weighted SBEU using $b = 2, 4, 8$. The weights have been chosen as $w_{ij} = 1$ whenever $\|s_i - s_j\| \leq 10$, and 0 otherwise.

Table 6 reports the estimates of the underlying random field parameters with the weighted PL method and the non overlapping weighted SBEU method. Note that the estimates are similar and that the estimated marginal probability of success is approximately 0.27 for all the estimation methods. The latter being very close to the empirical probability of success 0.2693647. Figure 2 shows the comparison between the empirical lorelogram and the estimated lorelogram using the
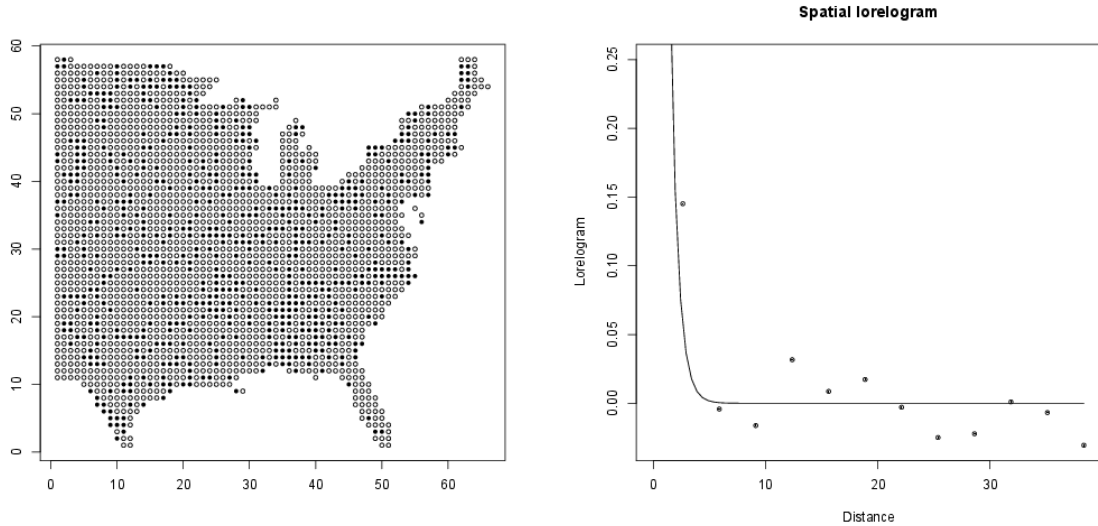
Fig. 2: Cancer mortality map data on the left and empirical lorelogram vs fitted lorelogram using not overlapping SBEU estimates with $b = 4$ on the right.

weighted SBEU with $b = 4$ and highlights the weakness of spatial dependence for cancer mortality data.

	α	μ
WPL	0.672	-0.614
SBEU(2)	0.650	-0.603
SBEU(4)	0.703	-0.596
SBEU(8)	0.695	-0.612

Table 6: PL and SBEU estimates for the Cancer mortality map data .

8 Concluding remarks

In this paper we propose a blockwise EU method for the estimation of a spatial binary field obtained by thresholding a latent Gaussian random field using moment conditions derived from the score of the PL.

The simulation results suggest that the EU-based estimator may be slightly worse in terms of statistical efficiency. However, we observe that the SBEU estimator provides considerable gains in terms of computational efficiency when compared with the PL estimator, especially in its non overlapping version. In this sense the choice of the blocks can be driven by computational considerations.

Moreover, like other estimation methods based on blocks, it can be parallelized in order to obtain further computational benefits by exploiting the approach proposed in Eidsvik et al. (2013). Finally, we outline that the proposed method can be extended to other types of data and to the space time setting using spatio-temporal blocks (Kleiber & Porcu (2014); Martinez-Ruiz et al. (2010); Ruiz-Medina et al. (2008); Yu et al. (2007)). In addition to that, differently from standard likelihood methods, it is well-known that the PL likelihood ratio test does not have a standard limiting distribution (Lunardon et al. (2012); Pace et al. (2011)). On the other hand, it could be shown that EU-based tests are chi square distributed. To our knowledge the relative finite sample properties of those tests are not known. Those are possible topics for future research.

Acknowledgements

Appendix

The appendix collects the proofs of the asymptotic results. First of all, we introduce the following notation: $\nabla_{\boldsymbol{\theta}}$ and $\nabla_{\boldsymbol{\lambda}}$ are the first derivative operators for $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ respectively, while $\nabla_{\boldsymbol{\theta\theta}}$, $\nabla_{\boldsymbol{\lambda\lambda}}$ and $\nabla_{\boldsymbol{\theta\lambda}}$ indicate second and cross derivatives and are defined accordingly. Similarly, for a certain function $\widehat{R}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ defined below, $\widehat{R}_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ is its first derivative with respect to $\boldsymbol{\theta}$. Derivatives with respect to $\boldsymbol{\lambda}$, second derivatives and cross derivatives are defined in a similar manner. In addition to that, we use

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \frac{b^d}{N} \sum_{\boldsymbol{\kappa}} \boldsymbol{m}_{\boldsymbol{\kappa}}(\boldsymbol{\theta}) \boldsymbol{m}_{\boldsymbol{\kappa}}(\boldsymbol{\theta})^{\mathsf{T}}$$

instead of the expression in Equation (20). As noticed in Newey & Smith (2004) this choice does not change the estimator but it simplifies the asymptotic analysis.

Proof (Proof of Theorem 1) We have to show that, for some $\delta > 0$, $P(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| > \delta) \to 0$ as $n \to \infty$. By continuity of $Q(\boldsymbol{\theta})$ and the assumption that $\boldsymbol{\theta}_0$ is the unique minimizer, we have that, for some $\varepsilon > 0$, $\{\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| > \delta\} \implies \{\mid Q(\widehat{\boldsymbol{\theta}}) - Q(\boldsymbol{\theta}_0) \mid > \varepsilon\}$. This is, the latter set contains the former. Hence, $P(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| > \delta) \leq P(\mid Q(\widehat{\boldsymbol{\theta}}) - Q(\boldsymbol{\theta}_0) \mid > \varepsilon)$. By assumptions **C1-C2** and arguments similar to those in Theorem 2 we have the following uniform convergence condition

$$\sup_{\boldsymbol{\theta} \in \Theta} \mid \widehat{Q}_n(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}) \mid \to_p 0. \tag{23}$$

Therefore,

$$\varepsilon < \mid Q(\widehat{\boldsymbol{\theta}}) - Q(\boldsymbol{\theta}_0) \mid = \mid Q(\widehat{\boldsymbol{\theta}}) - \widehat{Q}_n(\boldsymbol{\theta}_0) + \widehat{Q}_n(\boldsymbol{\theta}_0) - Q(\boldsymbol{\theta}_0) \mid$$
$$\leq 2 \sup_{\boldsymbol{\theta} \in \Theta} \mid \widehat{Q}_n(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}) \mid \to_p 0$$

where the latter inequality follows from the triangular inequality and the uniform convergence condition in (23). This implies $P(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| > \delta) \leq P(\mid Q(\widehat{\boldsymbol{\theta}}) - Q(\boldsymbol{\theta}_0) \mid > \varepsilon) \to 0$ as $n \to \infty$. Hence, $\widehat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$. Before showing asymptotic normality we show that the estimate of the Lagrange multiplier $\widehat{\boldsymbol{\lambda}}$ converges to zero in probability. Similarly to (7) we have that

$$\frac{\widehat{\boldsymbol{\lambda}}}{b^d} = \widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}})^{-1} \widehat{\boldsymbol{m}}(\widehat{\boldsymbol{\theta}}). \tag{24}$$

By a mean value argument, assumptions **C1-C2**, results in Theorem 2 and continuous mapping theorem we get

$$\frac{\widehat{\boldsymbol{\lambda}}}{b^d} \to_p \mathbf{0}$$

In analogy with (5) let us define

$$\pi_{\boldsymbol{\kappa}}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \frac{1}{N} \Big(1 - \boldsymbol{\lambda}^{\mathsf{T}}(\boldsymbol{m}_{\boldsymbol{\kappa}}(\boldsymbol{\theta}) - \widehat{\boldsymbol{m}}(\boldsymbol{\theta})) \Big)$$

and

$$\widehat{R}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = -\frac{1}{2N} \sum_{\boldsymbol{\kappa}} (N \pi_{\boldsymbol{\kappa}}(\boldsymbol{\theta}, \boldsymbol{\lambda}) - 1)^2.$$

Then by simple calculations we can write

$$\widehat{R}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = -\boldsymbol{\lambda}^{\mathsf{T}}\widehat{\boldsymbol{m}}(\boldsymbol{\theta}) + \frac{1}{2b_n^d}\boldsymbol{\lambda}^{\mathsf{T}}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})^{\mathsf{T}}\boldsymbol{\lambda}.$$

The first order conditions of $\widehat{R}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}})$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ are

$$0 = \widehat{R}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}}) = -\nabla\widehat{\boldsymbol{m}}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\lambda}} + \frac{\boldsymbol{\lambda}^{\mathsf{T}}}{Nb_n^d}\sum_{\boldsymbol{\kappa}}\boldsymbol{m}_{\boldsymbol{\kappa}}(\widehat{\boldsymbol{\theta}})\nabla\boldsymbol{m}_{\boldsymbol{\kappa}}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\lambda}} \tag{25}$$

$$0 = \widehat{R}_{\boldsymbol{\lambda}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}}) = -\widehat{\boldsymbol{m}}(\widehat{\boldsymbol{\theta}}) + \frac{1}{b^d}\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\lambda}}. \tag{26}$$

Let us now take a mean value expansion of the first order conditions (25) and (26) about the true values $(\boldsymbol{\theta}^{\mathsf{T}}, \boldsymbol{\lambda}^{\mathsf{T}})^{\mathsf{T}} = (\boldsymbol{\theta}_0^{\mathsf{T}}, \boldsymbol{0}^{\mathsf{T}})^{\mathsf{T}}$

$$\boldsymbol{0} = \widehat{R}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}}) = \widehat{R}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \boldsymbol{0}) + \widehat{R}_{\boldsymbol{\theta}\boldsymbol{\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\widehat{\boldsymbol{\lambda}} + \widehat{R}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \tag{27}$$

$$= \widehat{R}_{\boldsymbol{\theta}\boldsymbol{\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\frac{\sqrt{n}}{b^d}\widehat{\boldsymbol{\lambda}} + \frac{1}{b^d}\widehat{R}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

$$\boldsymbol{0} = \widehat{R}_{\boldsymbol{\lambda}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}}) = \widehat{R}_{\boldsymbol{\lambda}}(\boldsymbol{\theta}_0, \boldsymbol{0}) + \widehat{R}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\widehat{\boldsymbol{\lambda}} + \widehat{R}_{\boldsymbol{\lambda}\boldsymbol{\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \tag{28}$$

$$= \sqrt{n}\widehat{R}_{\boldsymbol{\lambda}}(\boldsymbol{\theta}_0, \boldsymbol{0}) + b^d\widehat{R}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\frac{\sqrt{n}}{b^d}\widehat{\boldsymbol{\lambda}} + \widehat{R}_{\boldsymbol{\lambda}\boldsymbol{\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).$$

More compactly,

$$\begin{pmatrix} \boldsymbol{0} \\ \sqrt{n}\widehat{R}_{\boldsymbol{\lambda}}(\boldsymbol{\theta}_0, \boldsymbol{0}) \end{pmatrix} = -\begin{pmatrix} \frac{1}{b^d}\widehat{R}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}}) & \widehat{R}_{\boldsymbol{\theta}\boldsymbol{\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}}) \\ \widehat{R}_{\boldsymbol{\lambda}\boldsymbol{\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}}) & b^d\widehat{R}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}}) \end{pmatrix}\begin{pmatrix} \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ \frac{\sqrt{n}}{b^d}\widehat{\boldsymbol{\lambda}} \end{pmatrix}.$$

By the UWLLN $\frac{1}{b^d}\widehat{R}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}}) \to_p \boldsymbol{0}$, $b^d\widehat{R}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}}) \to_p \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ and $\widehat{R}_{\boldsymbol{\lambda}\boldsymbol{\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}}) \to_p -\nabla_{\boldsymbol{\theta}}\boldsymbol{m}(\boldsymbol{\theta}_0)$. Hence,

$$\begin{pmatrix} \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ \frac{\sqrt{n}}{b^d}\widehat{\boldsymbol{\lambda}} \end{pmatrix} = -\begin{pmatrix} -\boldsymbol{\Omega}(\boldsymbol{\theta}_0) & \boldsymbol{\Omega}(\boldsymbol{\theta}_0)\nabla_{\boldsymbol{\theta}}\boldsymbol{m}(\boldsymbol{\theta}_0)^{\mathsf{T}}\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)^{-1} \\ \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)^{-1}\nabla_{\boldsymbol{\theta}}\boldsymbol{m}(\boldsymbol{\theta}_0)\boldsymbol{\Omega}(\boldsymbol{\theta}_0) & \boldsymbol{\Lambda}(\boldsymbol{\theta}_0) \end{pmatrix}\begin{pmatrix} \boldsymbol{0} \\ \sqrt{n}\widehat{\boldsymbol{m}}(\boldsymbol{\theta}_0) \end{pmatrix} + o_p(1)$$

where $\boldsymbol{\Omega}(\boldsymbol{\theta}_0) = (\nabla_{\boldsymbol{\theta}}\boldsymbol{m}(\boldsymbol{\theta}_0)^{\mathsf{T}}\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)^{-1}\nabla_{\boldsymbol{\theta}}\boldsymbol{m}(\boldsymbol{\theta}_0))^{-1}$ and $\boldsymbol{\Lambda}(\boldsymbol{\theta}_0) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)^{-1} - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)^{-1}\nabla_{\boldsymbol{\theta}}\boldsymbol{m}(\boldsymbol{\theta}_0)\boldsymbol{\Omega}(\boldsymbol{\theta}_0)\nabla_{\boldsymbol{\theta}}\boldsymbol{m}(\boldsymbol{\theta}_0)^{\mathsf{T}}\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)^{-1}$. The result follows from an application of the CLT and the continuous mapping theorem.

Proof (Proof of Theorem 2) Assume $\widehat{\boldsymbol{\theta}}$ consistent and define the following mean value expansion around $\boldsymbol{\theta}_0$

$$\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}}) = \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0) + \nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\Sigma}}(\dot{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).$$

By rearranging, taking the Euclidean norm and the Cauchy-Schwarz inequality we get

$$\|\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)\| \le \frac{2b^d}{N}\Big\|\sum_{\boldsymbol{\kappa}}\boldsymbol{m}_{\boldsymbol{\kappa}}(\dot{\boldsymbol{\theta}})\sum_{\boldsymbol{\kappa}}\nabla\boldsymbol{m}_{\boldsymbol{\kappa}}(\dot{\boldsymbol{\theta}})\Big\|\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|$$

$$\le \frac{2b^d}{N}\Big(\sum_{\boldsymbol{\kappa}}\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{m}_{\boldsymbol{\kappa}}(\dot{\boldsymbol{\theta}})\|^2\Big)^{\frac{1}{2}}\Big(\sum_{\boldsymbol{\kappa}}\sup_{\boldsymbol{\theta}\in\Theta}\|\nabla\boldsymbol{m}_{\boldsymbol{\kappa}}(\dot{\boldsymbol{\theta}})\|^2\Big)^{\frac{1}{2}}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|$$

$$= o_p(1).$$

The last line follows from the consistency of $\widehat{\boldsymbol{\theta}}$ and from the fact that $b^d = o(\sqrt{n})$. Thus,

$$\|\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\| \le \|\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0) + \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\|$$

$$\le \|\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)\| + \|\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\|$$

$$\le o_p(1) + \sup_{\boldsymbol{\theta}\in\Theta}\|\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\|$$

$$\le o_p(1).$$

The result follows from the application of the triangular inequality and UWLLN.

References

Adler, J. (2008). Some new random field tools for spatial analysis. *Stochastic Environmental Research and Risk Assessment*, *22*, 809–822.

Albert, P. & McShane, L. (1995). A generalized estimating equations approach for spatially correlated binary data: Applications to the analysis of neuroimaging data. *Biometrics*, *51*, 627–638.

Bevilacqua, M. & Gaetan, C. (2014). Comparing composite likelihood methods based on pairs for spatial gaussian random fields. *Statistics and Computing, to appear*.

Bevilacqua, M., Gaetan, C., Mateu, J. & Porcu, E. (2012). Estimating space and space-time covariance functions: a weighted composite likelihood approach. *Journal of the American Statistical Association, Theory & Methods*, *107*, 268–280. doi: 10.1080/01621459.2011.646928

Caragea, P. & Smith, R. (2006). *Approximate likelihoods for spatial processes* (Tech. Rep.). Department of Statistics, Iowa State University.

Cressie, N. (1993). *Statistics for spatial data* (revised ed.). New York: Wiley.

Cressie, N. & Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society, Series B*, *70*, 209–226.

Eidsvik, J., Shaby, B., Reich, B., Wheeler, M. & Niemi, J. (2013). Estimation and prediction with block composite likelihoods using parallel computing. *Journal of Computational and Graphical Statistics*, *xx*, xx.

Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association*, *102*, 321–331.

Gneiting, T., Kleiber, W. & Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, *105*, 1167–1177.

Heagerty, P. & Lele, S. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*, *93*, 1099–1111.

Heagerty, P. & Zeger, S. L. (1996). Lorelogram: A regression approach to exploring dependence in longitudinal categorical responses. *Journal of the American Statistical Association, Theory & Methods*, *93*, 150–162.

Jenish, N. & Prucha, I. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics*, *150*, 86–98.

Journel, A. (1983). Nonparametric estimation of spatial distributions. *Journal of the International Association for Mathematical Geology*, 485–468.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Annals of Statistics*, *25*, 2084-2102.

Kleiber, W. & Porcu, E. (2014). Nonstationary matrix covariances: compact support, long range dependence and quasi-arithmetic constructions. *Stochastic Environmental Research and Risk Assessment, to appear*.

Kovitz, J. L. & Christakos, G. (2004). Spatial statistics of clustered data. *Stochastic Environmental Research and Risk Assessment*, *18*, 147–166.

Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods. *Annals of Statistics*, *27*, 386–404.

Lee, Y. & Lahiri, S. (2002). Least squares variogram fitting by spatial subsampling. *Journal of the Royal Statistical Society B*, *64*, 837–854.

Lin, L. & Zhang, R. (2001). Blockwise Euclidean likelihood for weakly dependent processes. *Statistics and Probability Letters*, *53*, 143-152.

Lin, P. & Clayton, K. (2005). Analysis of binary spatial data by quasi-likelihood estimating equations. *Annals of Statistics*, *33*, 542-555.

Lindsay, B. (1988). Composite likelihood methods. *Contemporary Mathematics*, *80*, 221–239.

Lunardon, N., Pauli, F. & Ventura, L. (2012). A note on empirical likelihoods derived from pairwise score functions. *Journal of Statistical Computation and Simulation*, *8*, 1405–1414.

Martinez-Ruiz, F., Mateu, J., Montes, F. & Porcu, E. (2010). Mortality risk assessment through stationary space-time covariance functions. *Stochastic Environmental Research and Risk Assessment*, *24*, 519–526.

Matheron, G. (1976). A simple substitute for the conditional expectation: the disjuntive kriging.

Advanced geostatistics in the Mining Industry, Eds M. Guarascio, M. David and C. Huijbrechts. Dordrecht, Holland, 221–236.

Newey, W. & Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, *72*, 219–255.

Nordman, D. (2008). A blockwise empirical likelihood for spatial data. *Statistica Sinica*, *18*, 1111-1129.

Nordman, D. & Caragea, P. (2008). Point and interval estimation of variogram models using spatial empirical likelihood. *Journal of the American Statistical Association*, *103*, 350-361.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, *75*, 237-249.

Owen, A. B. (1990). Empirical likelihood confidence regions. *Annals of Statistics*, *18*, 90-120.

Owen, A. B. (1991). Empirical likelihood for linear models. *Annals of Statistics*, *19*, 1725-1747.

Owen, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, London.

Pace, L., Salvan, A. & Sartori, N. (2011). Adjusting composite likelihood ratio statistics. *Statistica Sinica*, *21*, 129–148.

Padoan, S. A., Ribatet, M. & Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association, Theory & Methods*, *105*, 263–277.

Porcu, E., Daley, D. J., Buhmann, M. & Bevilacqua, M. (2013). Radial basis functions with compact support for multivariate geostatistics. *Stochastic Environmental Research and Risk Assessment*, *27*, 909–922.

Qin, J. & Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics*, *22*, 300–325.

Rue, H. & Tjelmeland, H. (2002). Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, *29*, 31–49.

Ruiz-Medina, M. D., Angulo, J. M. & Anh, V. V. (2008). Multifractality in space-time statistical models. *Stochastic Environmental Research and Risk Assessment*, *22*, 81–86.

Sherman, M. (1996). Variance estimation for statistics computed from spatial lattice data. *Journal of the Royal Statistical Society B*, *58*, 509–523.

Sherman, M., Apanasovich, T. & Carroll, R. (2006). On estimation of binary autologistic spatial models. *Journal of Statistical Computation and Simulation*, *76*, 167-179.

Sherman, M. & Carlstein, E. (1994). Nonparametric estimation of the moments of a general statistic computed from spatial data. *Journal of the American Statistical Association, Theory & Methods*, *89*, 496–500.

Stein, M. (2008). A modeling approach for large spatial datasets. *Journal of the Korean Statistical Society*, *37*, 3–10.

Stein, M., Chi, Z. & Welty, L. (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society B*, *66*, 275–296.

Sun, Y., Li, B. & Genton, M. G. (2012). Geostastistics for large space-time datasets. *Space-Time Processes and Challenges Related to Environmental Problems, Porcu, E., Montero and Schlather Eds.*.

Varin, C., Reid, N. & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, *21*, 5–42.

Vecchia, A. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society B*, *50*, 297–312.

Yu, H.-L., Kovolos, A., Christakos, G., Chen, J.-C., Warmerdam, S. & Dev, B. (2007). Interactive spatiotemporal modelling of health systems: the SEKS-GUI framework. *Stochastic Environmental Research and Risk Assessment*, *21*, 555–572.