DEM Discussion Papers

Department of Economics and Management

2013/12

Approximate Maximum Likelihood Estimation of the Autologistic Model

Marco Bee, Diego Giuliani, Giuseppe Espa

ISSN 2282-2801 DEM discussion papers [online]

UNIVERSITÀ DEGLI STUDI
DI TRENTO

DEM Discussion Paper

ISSN 2282-2801 DEM Discussion Papers [online]
Università degli Studi di Trento

Department of Economics and Management, University of Trento, Italy.

Guidelines for authors
Papers may be written in Italian or in English. Faculty members of the Department must submit to one of the editors in pdf format. Management papers should be submitted to R. Gabriele. Economics Papers should be submitted to L. Andreozzi. External members should indicate an internal faculty member that acts as a referee of the paper.

Typesetting rules:

1. papers must contain a first page with title, authors, abstract, keywords and codes. Page numbering starts from the following page;
2. a template is available upon request from the managing editors.

Approximate Maximum Likelihood Estimation of the Autologistic Model

Marco Bee
Department of Economics and Management, University of Trento

Giuseppe Espa
Department of Economics and Management, University of Trento

Diego Giuliani
Department of Economics and Management, University of Trento

1 Introduction

Point estimation of spatial models is well known to be a difficult issue. One general reason is that, from a probabilistic point of view, a spatial model is a random field, typically characterized by a complex dependence structure, of which only a single realization is available for estimation. More specifically, what makes estimation overly complicated is the computational intractability of the normalizing constant of the joint density, even for moderate lattice sizes. This problem is particularly serious for Maximum Likelihood Estimation (MLE) procedures, as the normalization constant depends on the parameters of the model, and thus cannot be ignored in the maximization of the likelihood function.

On the other hand, the conditional distributions at single sites, given values at neighboring locations, usually admit simple representations. Exploiting this idea, Besag (1975) developed the so-called Maximum Pseudo-Likelihood Estimation (MPLE) method, which is still very popular in practical applications and is now usually considered within the context of composite likelihood (Varin et al., 2011).

MPLE is the earliest and simplest approach to estimation of the parameters of the most important model for spatially dependent binary random variables, i.e. the autologistic model. As will be made clear in Sect. 2, MPLE is based on the pseudo-likelihood function, defined by the product of the conditional distributions at all locations given the values at neighboring locations. To obtain the estimators, the pseudo-likelihood is maximized with respect to the parameters as if it were a likelihood, i.e. by means of standard logistic regression estimation techniques. However, the two functions coincide exactly if the observations at different locations are independent, a condition that is only satisfied in trivial cases. As a result, MPLEs are

consistent and asymptotically normal (Geman and Graffigne, 1987; Comets, 1992; Guyon and Künsch, 1992) but not efficient, with a loss of efficiency positively related to the (absolute) value of the spatial dependence parameter.

Given that the difficulties are mostly caused by the normalization constant, research has focused on methods of evaluating this quantity, usually by approximating it. The pioneering work by Ogata and Tanemura (1984) develops various techniques for approximating the likelihood function. Moyeed and Baddeley (1991) use an iterative stochastic approximation approach. Geyer and Thompson (1992) (see also Zheng and Zhu, 2008 and Hughes et al., 2011) introduce a Monte Carlo maximum likelihood approach that maximizes numerically an approximation of the likelihood function. Huffer and Wu (1998) employ Markov Chain Monte Carlo methods. Gu and Zhu (2001) compute MLEs by combining Markov Chain Monte Carlo and stochastic approximation methods. Huang and Ogata (2002) propose a generalization of Maximum Pseudo-Likelihood. Friel and Pettitt (2004) develop a method for exact MLE of the autologistic model that is simulation-free for rectangular lattices with smallest dimension not exceeding 10 and can be extended to larger sizes by means of Monte Carlo techniques. Finally, Wang and Zheng (2013) use expectation-maximization pseudo-likelihood and Monte Carlo expectation-maximization likelihood.

Recently, MLE for intractable likelihoods (see Murray et al., 2006 for a useful classification) has received some attention in the literature, mostly because various simulation-based approaches allow the approximation of the likelihood in these setups (Cox and Kartsonaki, 2012). In this paper we propose to apply the Approximate Maximum Likelihood Estimation (AMLE) method developed by Rubio and Johansen (2013) to the autologistic model. In short, AMLE exploits the potential of Approximate Bayesian Computation (ABC; Pritchard et al., 1999; Beaumont et al., 2002) for likelihood maximization. In a Bayesian setup, ABC produces a sample approximation to the posterior distribution, whose mode is the approximate Maximum A Posteriori (MAP) estimate. Under the parameterization of interest and with a uniform prior, the MAP estimate coincides with the MLE. The most appealing feature of AMLE is inherited from ABC: it allows one to obtain estimators (actually MLEs) without performing a formal maximization of the likelihood function. It is worth noting that Geyer and Thompson (1992)'s Monte Carlo maximum likelihood approach, though based on simulation, is quite different in that, unlike AMLE, it requires the evaluation of an approximation of the likelihood function. Similar considerations hold for the approximated maximum likelihood estimators recently proposed by Feizjavadian and Hashemi (2015).

This work is, to best of our knowledge, the first application of AMLE in the field of spatial statistics. It is also the first time AMLE is based on a non-iid sample from the approximated posterior. Although the theory does not

require an iid sample (Rubio and Johansen, 2013, p. 1636), it is of interest to analyze the performance of the algorithm in this framework. Hence, we study via simulation the properties of AMLE applied to the autologistic model, across various lattice sizes, and give insights into the choice of the basic components as well as the numerical values of the input parameters of the algorithm.

There are two main reasons why AMLE is the ideal candidate for computing MLEs of the autologistic model. First, as no likelihood evaluation is required, it allows one to bypass the problem of computing the normalizing constant. Second, the theory of AMLE is especially elegant and effective when ABC can be based on sufficient statistics. Whereas in most typical applications of ABC sufficient statistics are not available or difficult to compute and finding alternative summary statistics may not be straightforward, for the autologistic model sufficient statistics are readily available. Moreover, provided we can sample data from the model of interest for any value of its parameters, it can be applied to other spatial models with no major modification.

The approach proposed in this paper, besides being approximately efficient (as the estimators are approximate MLEs) is easily implemented and has the distinctive advantage of not suffering from the curse of dimensionality. As a matter of fact, the method works well even for very large dimensions: if the simulation of the autologistic model is based on the Metropolis algorithm, which generates in a sequential manner a single random variable for each location, the only limit is the machine's physical memory. The two last features are non-negligible strengths with respect to existing approaches, often characterized by involved implementation and/or quickly deteriorating performances for large lattice sizes.

The rest of this paper is organized as follows: Section 2 reviews the autologistic model, Section 3 introduces the AMLE methodology and develops its application to the autologistic model, Section 4 presents first the results of extensive Monte Carlo experiments aiming at a comparison of AMLE, MLE and MPLE in terms of Mean Squared Error (MSE) and then a real-data application. Finally, Section 5 concludes.

2 The autologistic model

By spatial model we mean a statistical model for a spatial pattern of data $\boldsymbol{y} = (y_i \in A \subset \mathbb{R}^s, i = 1, \ldots, K$, where $\mathbb{R}^s$ is the s-dimensional Euclidean space) having density

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{e^{-Q(\boldsymbol{y};\boldsymbol{\theta})}}{Z(\boldsymbol{\theta})}, \tag{1}$$

where $Q(\boldsymbol{y};\boldsymbol{\theta})$ is the energy function, which measures the interaction between the y_is. The normalizing constant is $Z(\boldsymbol{\theta}) = \int_{A^K} e^{-Q(\boldsymbol{y};\boldsymbol{\theta})}\mu(d\boldsymbol{y})$, where

μ denotes a suitable base measure - typically an appropriate version of either counting measure in the discrete case or Lebesgue measure in the continuous case. Note that one can construct joint distributions such as (1), as well as more complex joint distributions containing covariates and directional dependencies, from specified conditionals (see, for example, Hughes et al., 2011).

In the following we will be concerned with the autologistic model, which can be seen as a special instance of (1). Let $\tilde{\boldsymbol{y}} = (\tilde{y}_{i,j})$ $(i = 1, \ldots, N;$ $j = 1, \ldots, M,$ $K = MN)$. In this case the joint distribution of $\tilde{\boldsymbol{y}}$ is given by (Strauss, 1992; Casella and Robert, 2004, Example 5.8; Arbia, 2006, Sect. 2.4.2.3)

$$p(\tilde{\boldsymbol{y}}) = \frac{1}{Z(J,H)} \exp\left\{ -J \sum_{(i,j)\in\mathcal{N}} \tilde{y}_{i,j} - H \sum_{i,j} \tilde{y}_{i,j} \right\}, \tag{2}$$

where $J \in \mathbb{R}$ and $H \in \mathbb{R}$ are parameters, $\tilde{y}_{i,j} \in \{-1, 1\}$, the normalizing constant $Z(J,H)$ is given by $\sum_{\tilde{\boldsymbol{y}}\in\{-1,+1\}^K} \exp\{-J \sum_{(i,j)\in\mathcal{N}} \tilde{y}_{i,j} - H \sum_{i,j} \tilde{y}_{i,j}\}$ and $\mathcal{N}$ is a prespecified neighborhood equivalence relation: the sites (i,j) and (k,l) are said neighbors if either $i = k$ and $|j - l| = 1$ or $j = l$ and $|i - k| = 1$. The energy function can be written as $Q(\tilde{\boldsymbol{y}}; \boldsymbol{\theta}) = -\boldsymbol{\theta} V(\tilde{\boldsymbol{y}})$, where $V(\tilde{\boldsymbol{y}}) = \sum_{(i,j),(k,l)\in\mathcal{N}} \tilde{y}_{i,j}\tilde{y}_{k,l}$ is a statistic that counts the excess of like, over unlike, nearest-neighbor of a given configuration $\tilde{\boldsymbol{y}}$. In spatial statistics, the parameter J is the most important one, as it measures the intensity of the interaction, or the degree of spatial dependence, between the $\tilde{y}_{i,j}$s; when $J = 0$, the $\tilde{y}_{i,j}$s are independent. The model is of paramount importance in statistical mechanics, where it is known as the Ising model of ferromagnetism (Cipra, 1987): in this case, J is the interaction strength and H represents the energy involved in the magnetization of the lattice.

Let now $y_{i,j} = (\tilde{y}_{i,j} + 1)/2$. The conditional representation of (2) is

$$\begin{aligned} P(y_{i,j} = 1 | y_{k,l}, (i,j) \neq (k,l)) &= \frac{\exp\{2(H + J \sum_{i,j\in\mathcal{N}} y_{i,j})\}}{1 + \exp\{2(H + J \sum_{i,j\in\mathcal{N}} y_{i,j})\}} = \\ &= \frac{\exp\{\alpha + \beta \sum_{i,j\in\mathcal{N}} y_{i,j}\}}{1 + \exp\{\alpha + \beta \sum_{i,j\in\mathcal{N}} y_{i,j}\}}. \end{aligned} \tag{3}$$

If the $y_{i,j}$s were independent, this would be a logistic regression model. Unfortunately, as they are dependent, MLE must be based on the joint distribution (2), whose normalizing constant $Z(J,H)$ becomes rapidly intractable from the computational point of view as K increases.

3 AMLE of spatial models

3.1 A review of AMLE

Given a sample $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) \in \mathbb{R}^{q \times n}$ from a distribution with density function $f(\boldsymbol{y}; \boldsymbol{\theta})$, let the likelihood function be denoted by $L(\boldsymbol{\theta}; \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ is a vector of parameters. Assume for a moment a Bayesian setup, such that $\pi(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$ and $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ is the posterior, given by

$$\pi(\theta|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} f(\boldsymbol{y}|\boldsymbol{t})\pi(\boldsymbol{t})d\boldsymbol{t}}. \tag{4}$$

Suppose that we can construct the following approximation of the likelihood function:

$$\hat{f}_\epsilon(\boldsymbol{y}|\boldsymbol{\theta}) = \int_{\mathbb{R}^{q \times n}} K_\epsilon(\boldsymbol{y}|\boldsymbol{z}) f(\boldsymbol{z}|\boldsymbol{\theta}) d\boldsymbol{z}, \tag{5}$$

where $K_\epsilon(\boldsymbol{y}|\boldsymbol{z})$ is a normalized Markov kernel and ϵ is a scale parameter. Plugging (5) into (4) we obtain an approximation of the posterior:

$$\hat{\pi}_\epsilon(\theta|\boldsymbol{y}) = \frac{\hat{f}_\epsilon(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} \hat{f}_\epsilon(\boldsymbol{y}|\boldsymbol{t})\pi(\boldsymbol{t})d\boldsymbol{t}}.$$

For a uniform prior on a suitable set $\boldsymbol{D} \subset \mathbb{R}^p$, the maximization of the likelihood function and the maximization of the posterior density are equivalent, provided that the posterior density is written in the parameterization of interest.

Let $\eta : \mathbb{R}^{q \times n} \to \mathbb{R}^l$ be a summary statistic. The typical kernel $K_\epsilon^\rho(\boldsymbol{s}|\boldsymbol{t})$ is defined on the space of these summary statistics as follows:

$$K_\epsilon^\rho(\eta(\boldsymbol{y})|\eta(\boldsymbol{z})) \propto \begin{cases} 1 & \rho(\eta(\boldsymbol{y}), \eta(\boldsymbol{z})) < \epsilon, \\ 0 & \text{otherwise}, \end{cases} \tag{6}$$

where $\rho : \mathbb{R}^l \times \mathbb{R}^l \to \mathbb{R}^+$ is a metric. The simplest version uses $\eta(\boldsymbol{y}) = \boldsymbol{y}$, in which case we obtain the ABC algorithm introduced by Pritchard et al. (1999). For reasons that will be outlined below, when employing this technique in the AMLE setup we shall use the sufficient statistics of the model, if available.

The preceding discussion motivates the following algorithm:

Algorithm 1 *(AMLE)*

1. *Obtain a sample* $\boldsymbol{\theta}_\epsilon^* = (\boldsymbol{\theta}_{\epsilon,1}^*, \ldots, \boldsymbol{\theta}_{\epsilon,m}^*)'$ *from the approximate posterior* $\hat{\pi}_\epsilon(\theta|\boldsymbol{y})$*;* m *is commonly called ABC sample size;*

2. *Use this sample to construct a nonparametric estimator* $\hat{\phi}$ *of the density* $\hat{\pi}_\epsilon(\theta|\boldsymbol{y})$*;*

3. *Compute the maximum of* $\hat{\phi}$*,* $\tilde{\boldsymbol{\theta}}_{m,\epsilon}$*. This is an approximation of the MLE* $\hat{\boldsymbol{\theta}}$*.*

Step 1 can be carried out using any algorithm, as it is not required that the sample be independent. The most common solution is the simple ABC algorithm.

Algorithm 2 *(simple ABC)*

1. *Simulate* $\boldsymbol{\theta}'$ *from the prior distribution* $\pi(\cdot)$*;*

2. *Generate* $\boldsymbol{y} = (y_1, \ldots, y_n)'$ *from* $f(\cdot|\boldsymbol{\theta}')$*;*

3. *Accept* $\boldsymbol{\theta}'$ *with probability* $\propto K_\epsilon(\boldsymbol{y}|\boldsymbol{z})$*, otherwise return to Step 1.*

Rubio and Johansen (2013) study the asymptotic properties of the estimator, which are particularly important because they can give some insight into the choice of the input parameters. The crucial result is that, under a mild condition about $K_\epsilon(\boldsymbol{y}|\boldsymbol{z})$, $\hat{\pi}_\epsilon(\boldsymbol{\theta}|\boldsymbol{y})$ converges pointwise to $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ as $\epsilon \to 0$, for any $\boldsymbol{\theta} \in \boldsymbol{D}$.

A corollary of the preceding result suggests how to choose the summary statistic: if η is a sufficient statistic for $\boldsymbol{\theta}$, the ABC approximation converges pointwise to the posterior distribution. It is therefore clear that, if a sufficient statistic is available for $\boldsymbol{\theta}$, one should use it in the algorithm.

Finally, under the additional condition of equicontinuity of $\hat{\pi}_\epsilon(\cdot|\boldsymbol{y})$ on $\boldsymbol{D}$, and provided $\pi(\cdot|\boldsymbol{y})$ has a unique maximizer, it is possible to show that $\lim_{\epsilon\to 0} \hat{\pi}_\epsilon(\tilde{\boldsymbol{\theta}}|\boldsymbol{y}) = \pi(\tilde{\boldsymbol{\theta}}|\boldsymbol{y})$, where $\tilde{\boldsymbol{\theta}}$ is the unique maximizer of $\pi(\cdot|\boldsymbol{y})$.

Consider a simple random sample $\boldsymbol{\theta}^*_\epsilon = (\boldsymbol{\theta}^*_{\epsilon,1}, \ldots, \boldsymbol{\theta}^*_{\epsilon,m})'$ from the approximate posterior $\hat{\pi}_\epsilon(\cdot|\boldsymbol{y})$ with mode $\tilde{\boldsymbol{\theta}}_\epsilon$ and an estimator $\tilde{\boldsymbol{\theta}}_{m,\epsilon}$ of $\tilde{\boldsymbol{\theta}}_\epsilon$ obtained from $\boldsymbol{\theta}^*_\epsilon$ such that $\tilde{\boldsymbol{\theta}}_{m,\epsilon} \to \tilde{\boldsymbol{\theta}}_\epsilon$ almost surely when $m \to \infty$. From these results it follows that, for any $\gamma > 0$, there exists $\epsilon > 0$ such that $\lim_{m\to\infty} |\hat{\pi}_\epsilon(\tilde{\boldsymbol{\theta}}_{m,\epsilon}|\boldsymbol{y}) - \pi(\tilde{\boldsymbol{\theta}}|\boldsymbol{y})| \leq \gamma$ almost surely.

It is worth noting that $\tilde{\boldsymbol{\theta}}_{m,\epsilon}$ is an approximation of the MLE, with asymptotic variance related to the numerical value of ϵ: in particular, depending on ϵ, the estimator may be more or less efficient than the MLE. Ionides (2005) has indeed shown that the maximization of a smoothed approximation of the likelihood such as (5) may be preferable to the maximization of the likelihood itself.

In step 2, Rubio and Johansen (2013) suggest using Kernel Density Estimation (KDE). The performance of KDE deteriorates as the dimension of

$\boldsymbol{\theta}$ increases, but intractable likelihoods often have few unknown parameters, as is the case of the autologistic likelihood considered in this paper, which has only two unknown parameters. Nevertheless, an increase in the number of parameters (for example, in autologistic models including covariates, directional dependencies or even time-varying dependencies) is likely to cause higher computational costs.

3.2 AMLE of the autologistic model

Let $\boldsymbol{Y} = (y_{1,1}, \ldots, y_{N,M})$ be the $N \times M$ matrix containing the realization of an autologistic model. From now on, we will use the notation $\boldsymbol{y} = \text{vec}(y_{1,1}, \ldots, y_{N,M})$ and $\boldsymbol{\theta} = (\alpha, \beta)'$. The implementation of AMLE requires specific versions of Algorithms 1 and 2 and appropriate choices of the numerical values of the input parameters. We consider Algorithm 2 first.

The basic version of AMLE employs a uniform instrumental prior to enable the sampling procedure, even though any proper prior with mass in some neighborhood of the MLE might be used as well, provided its influence is later compensated for (Rubio and Johansen, 2013, p. 1637). We first need starting values and intervals around them, sufficiently wide to contain the true parameter values. If you can specify a compact set in which the MLE is known to lie, the widths can be approximately computed analytically, otherwise Monte Carlo techniques can be easily implemented.

For the autologistic model, the natural choice of the starting value is the MPLE estimator $\hat{\boldsymbol{\theta}}^{MPLE}$. We then sample candidate values of the parameters from the $U[\hat{\theta}_i^{MPLE} + d_i, \hat{\theta}_i^{MPLE} + u_i]$ distributions, where $d_i \leq 0$ and $u_i \geq 0$ $(i = 1, 2)$ are such that the intervals $[\hat{\theta}_i^{MPLE} + d_i, \hat{\theta}_i^{MPLE} + u_i]$ contain with very high probability the true values of the parameters.

The choice of d_i and u_i can be based on a simulation experiment, consisting in sampling B times the autologistic model with the MPLE estimate of the parameters (see the discussion concerning Step 2 of Algorithm 3 below for a thorough description of the simulation procedure) and computing the 99.9% confidence interval for each parameter. The numerical values of d_i and u_i are then given by the bounds of the confidence intervals. In the present context, we use $B = 10\,000$.

We know from theory that, if η in (6) is a sufficient statistic for $\boldsymbol{\theta}$, the AMLE algorithm has favorable convergence properties. Sufficient statistics for (2) are given by $\sum_i \tilde{y}_i$ and $\sum_{(i,j)\in\mathcal{N}} \tilde{y}_{ij}$. In terms of (3), they can be conveniently rewritten as $S_1 = \sum_{i=1}^{L} y_i$ and $S_2 = \sum_{i=1}^{M-1} \sum_{j=1}^{N} \mathbb{I}_{\{y_{i,j}=y_{i+1,j}\}} + \sum_{i=1}^{M} \sum_{j=1}^{N-1} \mathbb{I}_{\{y_{i,j}=y_{i,j+1}\}}$, where $\mathbb{I}$ is the indicator function. In more general Gibbs random fields, Grelaud et al. (2009) identify that there are finite-dimensional sufficient statistics which are natural candidates for ABC.

We are now in a position to detail Algorithm 2 in the autologistic case:

Algorithm 3

1. *Simulate* $\alpha' \sim U[\hat{\alpha}^{MPLE}+d_1, \hat{\alpha}^{MPLE}+u_1]$, $\beta' \sim U[\hat{\beta}^{MPLE}+d_2, \hat{\beta}^{MPLE}+u_2]$.

2. *Simulate a realization* $\boldsymbol{y}'$ *of the autologistic model with parameters* α' *and* β'*; the simulation is carried out by means of the Metropolis algorithm (Metropolis et al., 1953) using the same setup of Gu and Zhu (2001, p. 346); see also Müller (1991) and Huang and Ogata (2002, p. 6).*

3. *Compute the sufficient statistics* S_1' *and* S_2' *using* $\boldsymbol{y}'$*. If the Euclidean distance* $||\boldsymbol{S}-\boldsymbol{S}'|| = \sqrt{(S_1 - S_1')^2 + (S_2 - S_2')^2}$ *is smaller than* ϵ*, where* $\boldsymbol{S} = (S_1, S_2)$*,* $\boldsymbol{S}' = (S_1', S_2')$ *and* S_1 *and* S_2 *are the sufficient statistics computed with the observed data, accept* α' *and* β'*; otherwise, return to Step 1.*

4. *Estimate non-parametrically the density of the simulated* $\boldsymbol{\theta}$*s using standard kernel density estimation methods;* $\hat{\boldsymbol{\theta}}$ *is given by the mode of the estimated non-parametric density.*

At step 2, the detailed implementation of the Metropolis algorithm is as follows. Take a random initial state, i.e. select independently each $y_{i,j} = \pm 1$ $(i = 1, \ldots, N; j = 1, \ldots, M)$ with equal probability. Suppose that the current value of the process at site (i, j) is $y(i, j)$ and the corresponding energy value is Q, and consider the energy Q^* associated to $y(i, j)^* = -y(i, j)$. The core of the algorithm can be summarized as follows:

1. if $Q^* \leq Q$, replace $y(i, j)$ with $y(i, j)^*$ and Q with Q^*;

2. if $Q^* > Q$, generate $u^* \sim U(0, 1)$ and

   (i) if $U \leq \exp(Q - Q^*)$, replace $y(i, j)$ with $y(i, j)*$ and Q with Q^*;

   (ii) if $U > \exp(Q - Q^*)$, keep $y(i, j)$ and Q.

For a given pair of parameters, a replication of the simulation is performed by selecting each site in lexicographical order and repeating at least $500 \times N \times M$ times steps (1) and (2) above. In order to assess convergence, compute

$$SM_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} m(\boldsymbol{y}^{(t)})/T,$$

where $T \geq 500$ is the number of Monte Carlo steps, $\boldsymbol{y}^{(t)}$ $(t = 1, \ldots, 500)$ is the output from the Metropolis algorithm at the t-th iteration and $m(\boldsymbol{y}) = \sum_{i=1}^{N} \sum_{j=1}^{M} y_{i,j}$ is the so-called magnetic configuration of $\boldsymbol{y}$. Stop when $|SM_T(\boldsymbol{\theta})| < 0.001$.

At step 2, the autologistic model may also be simulated by means of perfect sampling. In this case, Propp and Wilson (1996) show that exact samples can be generated for the Ising model, although the computing time is large if the model is below its critical temperature.

Standard analysis, not reported here, of the output of the Metropolis algorithm at Step 2 of Algorithm 3 shows that the thinned chain is approximately independent. Thus, we use simple ABC even though the output of Step 2 is not truly independent. Hence, more precisely, this way of proceeding is an approximation to simple ABC, as it arises from the use of an MCMC transition instead of iid sampling; as the sample is approximately iid, the impact of this approximation should be negligible. A more rigorous approach would be to use the MCMC-ABC algorithm proposed by Marjoram et al. (2003).

As of step 4, analogously to Rubio and Johansen (2013), we use the kde command of the ks (Duong, 2014) package of R (R Core Team, 2014) to compute the nonparametric estimator $\hat{\phi}$ of the density $\hat{\pi}_\epsilon(\theta|\boldsymbol{y})$. The bandwidth is selected using the hpi (when only one parameter is unknown) or Hpi (when two parameters are unknown) functions, which implement plug-in bandwidth selectors (Wand and Jones, 1994). The kernel density estimation step is not so important if m is large (Rubio and Johansen, 2013, p. 1642); nevertheless, the choice of the plug-in selector is based on the results by Sköld and Roberts (2003) concerning bandwidth selection when data are generated via the Metropolis-Hastings algorithm. Finally, we compute $\hat{\boldsymbol{\theta}} = \arg\max(\hat{\phi})$.

Besides $\boldsymbol{D}$, in order to use the algorithm one has to set the parameters m and ϵ. The choice is problem-dependent and will be examined thoroughly in the next section. Before focusing on the autologistic model, we show the results of a toy simulation experiment that illustrates how the variance of the estimator depends on ϵ.

Example 1. We simulate $B = 100$ times a binomial random variable $X \sim \text{Bin}(n, p)$ with $n = 100$ and $p = 0.5$. At each replication, we estimate p by means of the MLE $\hat{p} = x/n$ and by means of AMLE, using an ABC sample size $m = 1000$, $D = [0, 1]$ and $\epsilon \in \{0.01, 0.05, 0.5\}$. It is well-known from classical MLE asymptotic theory that $\text{var}(\hat{p}) \approx p(1-p)/n = 0.0025$. Table 1 gives the bias, the variance and the MSE of the two estimators, where the variance is given by the sample variance of the B estimates. The rejection rate is defined as $B(1 - m/\sum_{i=1}^{B} T_i)$, where T_i is the number of replications of the ABC algorithm needed to generate m observations from $\hat{\pi}_\epsilon(\theta|\boldsymbol{y})$ at the i-th replication of Algorithm 3. Whereas with $\epsilon = 0.5$ almost all observations are accepted, the computational cost associated to the case $\epsilon = 0.01$ is much larger, because approximately 99% of the simulated parameter values are rejected.

Table 1: Bias, variance and MSE of classical MLE and AMLE of p for $X \sim \text{Bin}(n, p)$ with $n = 100$ and $p = 0.5$. The results shown are averages of 100 Monte Carlo replications. The remaining parameters are $m = 1000$, $D = [0, 1]$.

	MLE	AMLE		
		$\epsilon = 0.01$	$\epsilon = 0.1$	$\epsilon = 0.5$
bias	$2.30 \cdot 10^{-3}$	$3.22 \cdot 10^{-3}$	$4.61 \cdot 10^{-3}$	$4.90 \cdot 10^{-3}$
variance	$2.51 \cdot 10^{-3}$	$2.61 \cdot 10^{-3}$	$3.49 \cdot 10^{-3}$	$6.00 \cdot 10^{-1}$
MSE	$2.52 \cdot 10^{-3}$	$2.62 \cdot 10^{-3}$	$3.51 \cdot 10^{-3}$	$6.00 \cdot 10^{-1}$
rejection rate		0.989	0.799	0.052

Considering that AMLE avoids evaluation of the likelihood function, approaches to the computation of standard errors based on Fisher information are not feasible. On the other hand, standard errors can be approximated in a natural way, albeit with a non-negligible computational burden, by means of parametric bootstrap (Davison and Hinkley, 1997, Sect. 2.2): the procedure consists in estimating $\boldsymbol{\theta}$, simulating B times the model using the estimated parameter vector $\hat{\boldsymbol{\theta}}$, computing B AMLEs $\hat{\boldsymbol{\theta}}^*_1, \ldots, \hat{\boldsymbol{\theta}}^*_B$ and finally obtaining the standard error of the estimator as $\text{se}(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{B} (\hat{\boldsymbol{\theta}}^*_i - \bar{\hat{\boldsymbol{\theta}}}^*)^2/(B-1)$, where $\bar{\hat{\boldsymbol{\theta}}}^* = \sum_{i=1}^{B} \hat{\boldsymbol{\theta}}^*_i/B$.

4 Numerical experiments

We now turn to the simulation experiments concerning the use of AMLE for estimating the autologistic model. As said above, three parameters need to be set as input of the algorithm. In doing this, we have to face a trade-off between computational burden and precision of the estimators. In particular, the precision increases as ϵ gets smaller and m gets larger. $\boldsymbol{D}$ is less important because, provided it includes the true parameter value, it mainly affects the rejection rate (Rubio and Johansen, 2013).

We note first that there is a negative relationship between ϵ and the rejection rate rr, because rr is given by $rr(B) \stackrel{\text{def}}{=} \#\{\boldsymbol{y}' : |\eta(\boldsymbol{y}) - \eta(\boldsymbol{y}')| > \epsilon\}/B$, where B is the number of replications, and

$$\lim_{B \to \infty} rr(B) = P(|\eta(\boldsymbol{y}) - \eta(\boldsymbol{y}')| > \epsilon), \tag{7}$$

where the convergence in (7) is almost sure via the Strong Law of Large Numbers. Application of AMLE to the autologistic model is quite heavy from the computational point of view. Suppose indeed to be interested in implementing Algorithm 3 with $m = 100$ and to use a value of ϵ such that the rejection rate is $rr(B) = 0.99$. Then the total number of autologistic

models that need to be simulated is 10 000. Using the Metropolis algorithm with 500 steps, we have to generate $500 \cdot M \cdot N \cdot 10\,000$ random variables.

The following simulation experiments are based on setups similar to Gu and Zhu (2001), Huang and Ogata (2002) and Friel and Pettitt (2004), so that the results obtained with AMLE are comparable to MLEs produced by their procedures. In detail, the three experiments carried out in this paper are organized as follows:

1. Simulate an $N \times M$ autologistic model with $N = 125$ and $M = 12$. This is the dimension of the real-data analysis carried out in this paper as well as in Gu and Zhu (2001) with the Wiebe's wheat data (see below). The parameters are $\alpha = 0$ and $\beta \in \{-0.4, -0.2, 0, 0.2, 0.4\}$. α is treated as a known parameter, so that we only estimate β.

2. Same as 1, but now the lattice has $M = N = 64$, as in Gu and Zhu (2001)'s and Huang and Ogata (2002)'s Monte Carlo investigations.

3. Analogously to Friel and Pettitt (2004), simulate an $N \times M$ autologistic model with $N = 12$ and $M = 100$. The parameters are $\alpha \in \{-0.3, 0, 0.3\}$ and $\beta \in \{-0.3, -0.1, 0, 0.1, 0.3\}$. Both parameters are considered unknown and therefore estimated from the simulated data.

All the simulations are repeated $B = 15$ times. In order to compare the estimators, we use the Root-Mean-Square-Error (RMSE), given by $\text{RMSE}(\hat{\theta}) = \sqrt{b(\hat{\theta})^2 + \text{var}(\hat{\theta})}$, where $b(\hat{\theta}) = (1/B)\sum_{i=1}^{B} \hat{\theta}_i - \hat{\bar{\theta}}$ is the bias, $\text{var}(\hat{\theta}) = (1/(B-1))\sum_{i=1}^{B}(\hat{\theta}_i - \hat{\bar{\theta}})^2$ is the variance and $\hat{\bar{\theta}} = (1/B)\sum_{i=1}^{B}\hat{\theta}_i$ is the sample mean.

Before carrying out the experiments outlined above, we perform AMLE with $\alpha = 0$ and $\beta = 0.4$ with various values of ϵ, in order to assess the impact of ϵ on the performance of the algorithm. Fig. 1 shows the results for $\epsilon \in \{5, \ldots, 15\}$ and $m \in \{200, 400\}$. For comparison purposes, we also report the RMSE of the MPLE estimator.

The graph suggests that the RMSE is essentially the same for all values of ϵ, as the differences seem to be random and mainly related to sampling variability. This probably means that $\epsilon = 15$ is already small enough to guarantee a good approximation of the approximated posterior to the likelihood (see Rubio and Johansen, 2013, Corollary 1). In the following we use $\epsilon = 9$, as it is the numerical value corresponding to the smallest RMSE, and $m = 200$, as $m = 400$ doubles computing time with no significant precision improvement.

Tables 2 and 3 give numerical values of bias, standard deviation, RMSE and relative performances, measured as $RMSE(\hat{\beta}^{AMLE})/RMSE(\hat{\beta}^{MPLE})$, respectively for $N = 125$, $M = 12$ and for $N = M = 64$. In the latter case, the table also displays the outcomes obtained by Gu and Zhu (GZ) with

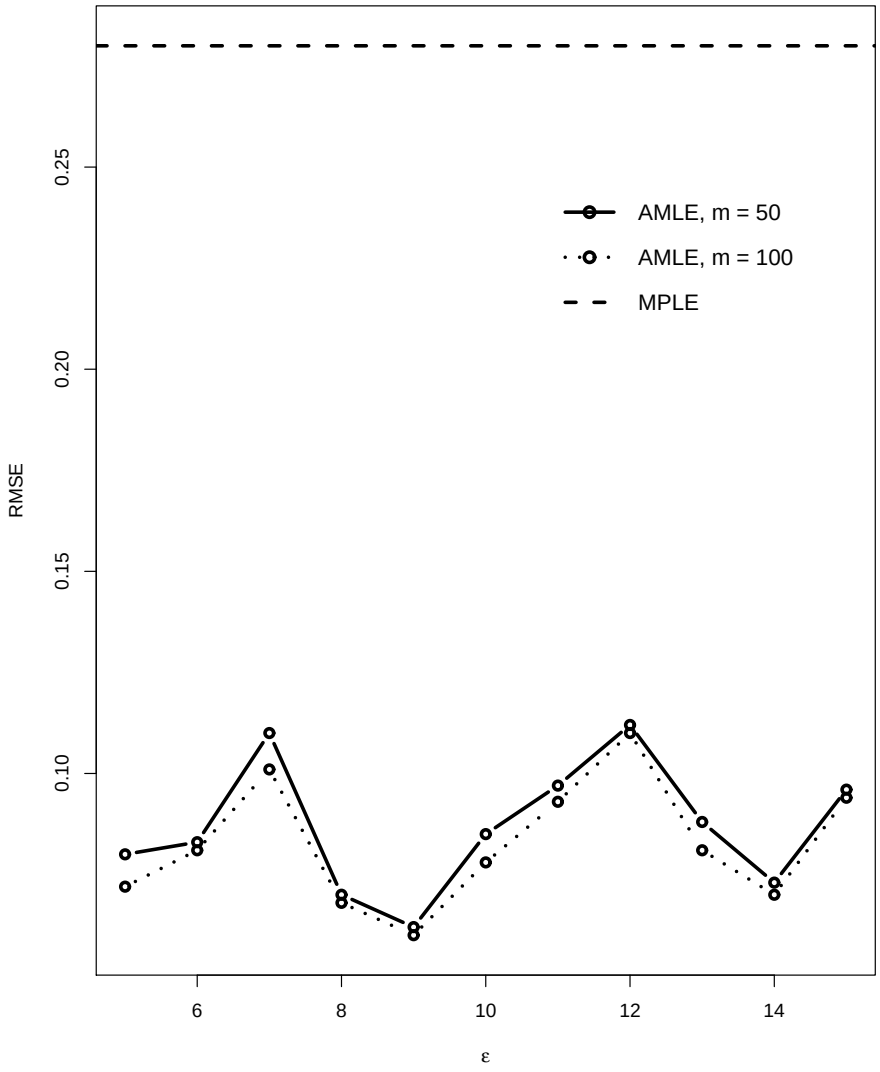

Figure 1: RMSE of three estimators of β (one-parameter case, parameters $\alpha = 0, \beta = 0.4$, sample sizes $m \in \{200, 400\}$).

Table 2: Bias, standard deviation, RMSE and relative performance of $\hat{\beta}^{AMLE}$ vs $\hat{\beta}^{MPLE}$ for various values of β, with $N = 125$, $M = 12$, $\epsilon = 9$, $m = 200$, $B = 15$. Relative performance is defined as RelPerf = RMSE($\hat{\beta}^{AMLE}$)/RMSE($\hat{\beta}^{MPLE}$).

		Bias ($\times 10^2$)	Sd ($\times 10^2$)	RMSE ($\times 10^2$)	RelPerf
$\beta = -0.4$	MPLE	3.39	0.67	3.45	
	AMLE	1.20	0.89	1.49	0.43
$\beta = -0.2$	MPLE	3.06	0.64	3.13	
	AMLE	−1.51	0.95	1.79	0.57
$\beta = 0$	MPLE	−0.37	0.61	0.71	
	AMLE	0.39	0.63	0.74	1.04
$\beta = 0.2$	MPLE	−2.21	0.74	2.34	
	AMLE	−0.16	0.77	0.78	0.34
$\beta = 0.4$	MPLE	−2.78	0.55	2.83	
	AMLE	−0.22	0.62	0.65	0.23

their MCMC stochastic approximation method and by Huang and Ogata (HO) with MLE computed via the DALL optimization subroutine (Ishiguro and Akaike, 1989); see Gu and Zhu (2001, Table 1) and Huang and Ogata (2002, Table 1). Figure 2 shows the relative performances of $\hat{\beta}^{AMLE}$ with respect to $\hat{\beta}^{MPLE}$ both for $N = 125,\ M = 12$ and for $N = M = 64$.

From the tables and the figure, at least three facts emerge clearly. First, the relative performance of AMLE gets larger as β increases in absolute value; this is unsurprising, as MPLE corresponds to MLE when $\beta = 0$, and the loss of efficiency is an increasing function of $|\beta|$. Second, the improvement in relative performance brought by AMLE seems to be different for positive and negative spatial dependence. This outcome is not new in the spatial statistics literature (see, e.g., Schabenberger and Gotway, 2002; Griffith and Arbia, 2010; Arbia et al., 2011; Arbia et al., 2013), where "asymmetric" results corresponding to setups with spatial dependence of the same magnitude but different sign are well known. Finally, the smaller MSE of AMLE is entirely due to a smaller bias, whereas the standard deviation of the two estimators is approximately the same or slightly larger for AMLE, probably because of some extra Monte Carlo sampling variability.

For the 64×64 lattice, AMLE and Gu and Zhu (2001)'s outcomes are very similar to each other, except when $\beta = 0$ and $\beta = -0.2$, and preferable to Huang and Ogata (2002) results. When $\beta = 0$ and $\beta = -0.2$, AMLE seems to perform better than exact MLE. A possible explanation may be that estimators obtained by maximizing smoothed versions of the true likelihood such as (5) in some cases are more stable than MLEs (Ionides, 2005).

Table 3: Bias, standard deviation, RMSE Error and relative performance of $\hat{\beta}^{AMLE}$ vs $\hat{\beta}^{MPLE}$ for various values of β, with $N = M = 64$, $\epsilon = 9$, $m = 200$, $B = 15$. Relative performance is defined as RelPerf = RMSE($\hat{\beta}^{AMLE}$)/RMSE($\hat{\beta}^{MPLE}$). For comparison purposes, the results obtained by Gu and Zhu (GZ) with their MCMC stochastic approximation method and by Huang and Ogata (HO) with MLE are reported as well (see Gu and Zhu, 2001, Table 1 and Huang and Ogata, 2002, Table 1).

		Bias ($\times 10^2$)	Sd ($\times 10^2$)	RMSE ($\times 10^2$)	RelPerf
$\beta = -0.4$	MPLE	3.30	0.62	3.36	
	AMLE	0.27	0.58	0.64	0.19
	GZ	-0.01	0.68	0.68	
	HO	-	-	0.64	
$\beta = -0.2$	MPLE	1.50	0.37	1.54	
	AMLE	0.09	0.49	0.50	0.33
	GZ	0.01	1.00	1.04	
	HO	-	-	1.04	
$\beta = 0$	MPLE	0.11	0.43	0.45	
	AMLE	0.04	0.48	0.48	1.07
	GZ	0.10	1.10	1.15	
	HO	-	-	1.11	
$\beta = 0.2$	MPLE	1.34	0.56	1.45	
	AMLE	0.45	0.75	0.88	0.61
	GZ	0.01	1.00	0.92	
	HO	-	-	1.04	
$\beta = 0.4$	MPLE	2.85	0.52	2.89	
	AMLE	0.43	0.56	0.70	0.24
	GZ	0.10	0.67	0.70	
	HO	-	-	0.67	

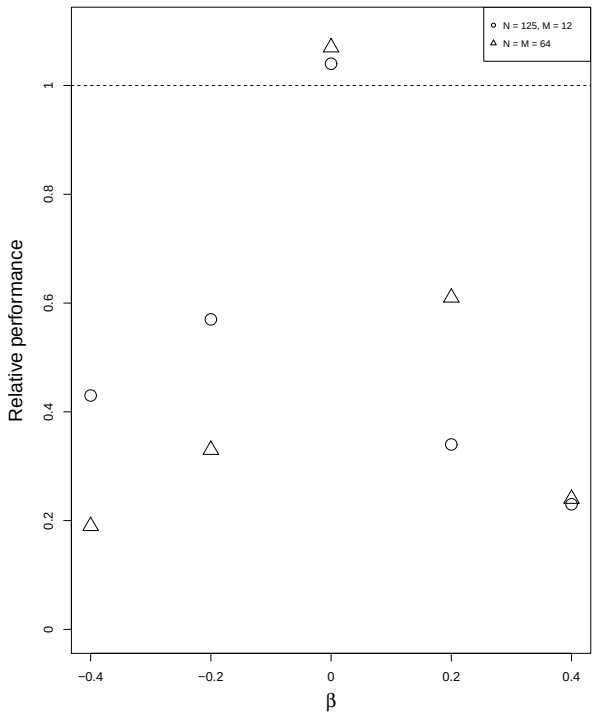

Figure 2: Relative performance of y of $\hat{\beta}^{AMLE}$ with respect to $\hat{\beta}^{MPLE}$, measured as $RMSE(\hat{\beta}^{AMLE})/RMSE(\hat{\beta}^{MPLE})$ in the one-parameter case for $\alpha = 0$ and $\beta \in \{-0.4, -0.2, 0, 0.2, 0.4\}$, with $\epsilon = 9$, $m = 200$, $B = 15$.

Table 4: Bias, standard deviation and RMSE of $\hat{\beta}^{AMLE}$ vs $\hat{\beta}^{MPLE}$ for $\beta = \pm 0.4$, with $N = 125$, $M = 12$, $\epsilon = 9$, $m = 200$, $B = 100$.

		Bias ($\times 10^2$)	Sd ($\times 10^2$)	RMSE ($\times 10^2$)	RelPerf
$\beta = -0.4$	MPLE	2.92	0.64	2.99	
	AMLE	0.73	0.77	1.06	0.35
$\beta = 0.4$	MPLE	-2.22	0.57	2.29	
	AMLE	-0.26	0.45	0.52	0.23

The rejection rate is between 0.991 and 0.993 for the experiments in Table 2 and between 0.992 and 0.994 for those in Table 3. Thus, to obtain a sample of 200 parameter values, the number of simulations of the autologistic model is approximately equal to 25 000 in the first case and to 28 000 in the second case. The amount of time needed ranges between about two and two and a half hours per replication on a 3.16 GHz machine.

The small number of replications on which these results are based may cause some concern. To check the reliability of the outcomes, we repeated two of the experiments with $B = 100$. In particular, in the case $M = 125$, $N = 12$, we reran the simulations with $\beta = +0.4$ and $\beta = -0.4$, obtaining the results displayed in Table 4.

The meaningfulness of the outcomes with $B = 15$ can also be verified by means of the following argument. Consider, for example, the estimates concerning the case $\beta = -0.4$ in tables 2 and 4. As both AMLE and MPLE are asymptotically normal, the 95% approximate confidence intervals for $\hat{\beta}_B^{AMLE}$ and $\hat{\beta}_B^{MPLE}$ are respectively given by $\hat{\bar{\beta}}_B^{AMLE} \pm 2.14\hat{\sigma}^{AMLE}/\sqrt{B}$ and $\hat{\bar{\beta}}_B^{MPLE} \pm 2.14\hat{\sigma}^{MPLE}/\sqrt{B}$, where 2.14 is the 0.975 quantile of the Student t distribution with 14 degrees of freedom, $\hat{\bar{\beta}}_B^{AMLE} = (1/B)\sum_{i=1}^{B} \hat{\beta}_i^{AMLE}$ and $\hat{\bar{\beta}}_B^{MPLE} = (1/B)\sum_{i=1}^{B} \hat{\beta}_i^{MPLE}$. Plugging the estimates displayed in tables 2 and 4 into these formulas we get

$$CI_{\hat{\beta}_{15}^{MPLE}} = -.361 \pm 2.14\frac{.0067}{\sqrt{15}} = [-.3647, -.3573],$$
$$CI_{\hat{\beta}_{15}^{AMLE}} = -.3842 \pm 2.14\frac{.0089}{\sqrt{15}} = [-.3918, -.3842].$$

Similarly, when the true value of β is equal to 0.4, the approximate CIs are given by

$$CI_{\hat{\beta}_{15}^{MPLE}} = .3722 \pm 2.14\frac{.0055}{\sqrt{15}} = [.3692, .3752],$$
$$CI_{\hat{\beta}_{15}^{AMLE}} = .3978 \pm 2.14\frac{.0062}{\sqrt{15}} = [.3944, .4012].$$

Thus, even with a sample size as small as $B = 15$, the difference between the two approaches is significant.

To summarize the pros and cons of AMLE and of Gu and Zhu (2001)'s approach, which has a similar behavior in terms of RMSE, AMLE is easier to implement, whereas Gu and Zhu (2001)'s MCMC stochastic approximation is computationally less heavy.

We now turn to the two-parameter case, using first $\epsilon = 9$ and $m = 200$ as in the one-parameter case. The results are shown in tables 5 and 6. Table 5 also reports the outcomes obtained by Friel and Pettitt (2004) in the same setup. Friel and Pettitt (2004) compute the MSE of an estimator $\hat{\theta}$ as $MSE(\hat{\theta}) = \hat{b}^2(\hat{\theta}) + \sqrt{B}\hat{\sigma}^2_{\hat{\theta}}$, where $B = 15$ and $\hat{b}(\hat{\theta}) = \sum_{i=1}^{B}(\hat{\theta}_i - \hat{\bar{\theta}})/B$ and $\hat{\sigma}^2_{\hat{\theta}} = \sum_{i=1}^{B}(\hat{\theta}_i - \hat{\bar{\theta}})^2/(B-1)$ are respectively the estimated bias and variance. Instead, we compute the MSE as $MSE(\hat{\theta}) = \hat{b}^2(\hat{\theta}) + \hat{\sigma}^2_{\hat{\theta}}$. Hence, in order to make Friel and Pettitt (2004)'s results comparable to ours, we use our formula for computing the MSE, taking as inputs the estimated bias and variance obtained by Friel and Pettitt (2004) in their experiments. For example, when $\alpha = 0$ and $\beta = -0.3$ in Table 5, we compute $MSE(\hat{\beta}^{FP}) = 0.004^2 + 0.00837^2 = 0.000086057$, so that $RMSE(\hat{\beta}^{FP}) = \sqrt{0.000086057} = 0.00927$. Overall, the outcomes in Table 5 show that AMLE performs better than MPLE, but the difference is smaller than in the preceding tables. One reason may be that AMLE needs a smaller ϵ and/or a larger m to reach convergence. However, before considering this issue, it is worth noting that this result is also related to a peculiar feature of the MPLE estimators.

Similar to the one-parameter cases, the improvement is mainly due to a smaller bias component. For some parameter combinations (consider for example $\hat{\alpha}^{MPLE}$ when $\alpha = 0$ and $\beta = -0.3$), MPLE is approximately unbiased, so that the two estimators have a similar MSE. In these instances, concluding that MPLE performs well seems more appropriate than concluding that AMLE performs poorly, as $\hat{\alpha}^{MPLE}$ is approximately unbiased. Continuing the analysis of the $\alpha = 0$, $\beta = -0.3$ case, $\hat{\beta}^{AMLE}$ has a smaller RMSE than $\hat{\beta}^{MPLE}$, because the latter shows a systematic bias. The other cases where the difference is small mostly correspond to $|\beta| = 0.1$ and are therefore not surprising, because, when $|\beta| \to 0$, MPLEs converge to MLEs.

Turning to the convergence issue, the need of a larger m in the two-parameter case may be justified by the fact that kernel density estimation works worse when the dimension of the parameter space is larger. Moreover, according to the simulation results presented by Rubio and Johansen (2013), the numerical value of ϵ is, in general, crucial for the quality of the approximation. Thus, we repeated some of the experiments with $\epsilon = 5$ and $m = 500$. Selected results are displayed in Table 6 and relative performances are shown in Fig. 3.

Now the improvement obtained with AMLE is more significant, but mostly for $\hat{\beta}$. Moreover, the gain in efficiency of AMLE is much larger when

Table 5: Bias, standard deviation, RMSE and relative performance of $\hat{\alpha}^{AMLE}$ vs $\hat{\alpha}^{MPLE}$ and $\hat{\beta}^{AMLE}$ vs $\hat{\beta}^{MPLE}$ for various values of α and β, with $N = 12$, $M = 100$, $\epsilon = 9$, $m = 200$ and $B = 15$. Relative performance is defined as RelPerf = RMSE($\hat{\theta}^{AMLE}$)/RMSE($\hat{\theta}^{MPLE}$), where θ is equal to α or β. For comparison purposes, the results obtained by Friel and Pettitt (FP) with their method are reported as well (see Friel and Pettitt, 2004, tables 2 and 3).

		Bias ($\times 10^2$)	Sd ($\times 10^2$)	RMSE ($\times 10^2$)	RelPerf
$\alpha = 0, \beta = -0.3$	$\hat{\alpha}^{MPLE}$	-0.040	1.860	1.860	1.006
	$\hat{\alpha}^{AMLE}$	-0.206	1.859	1.871	
	$\hat{\alpha}^{FP}$	1.2	0.293	1.235	
	$\hat{\beta}^{MPLE}$	1.362	0.835	1.598	0.689
	$\hat{\beta}^{AMLE}$	-0.466	0.998	1.101	
	$\hat{\beta}^{FP}$	0.9	0.34	0.962	
$\alpha = 0, \beta = -0.1$	$\hat{\alpha}^{MPLE}$	0.181	1.806	1.815	1.000
	$\hat{\alpha}^{AMLE}$	0.123	1.812	1.816	
	$\hat{\alpha}^{FP}$	-2	0.395	2.039	
	$\hat{\beta}^{MPLE}$	-0.085	1.075	1.079	0.992
	$\hat{\beta}^{AMLE}$	0.019	1.070	1.070	
	$\hat{\beta}^{FP}$	-0.7	0.472	0.844	
$\alpha = 0, \beta = 0.1$	$\hat{\alpha}^{MPLE}$	-0.449	1.376	1.448	0.824
	$\hat{\alpha}^{AMLE}$	-0.454	1.103	1.193	
	$\hat{\alpha}^{FP}$	0.1	0.361	0.375	
	$\hat{\beta}^{MPLE}$	0.534	0.850	1.004	0.897
	$\hat{\beta}^{AMLE}$	0.047	0.900	0.901	
	$\hat{\beta}^{FP}$	0.8	0.416	0.902	
$\alpha = 0, \beta = 0.3$	$\hat{\alpha}^{MPLE}$	1.503	1.621	2.211	0.498
	$\hat{\alpha}^{AMLE}$	0.213	1.079	1.100	
	$\hat{\alpha}^{FP}$	1.4	0.337	1.440	
	$\hat{\beta}^{MPLE}$	-3.013	1.432	3.336	0.518
	$\hat{\beta}^{AMLE}$	-1.167	1.274	1.728	
	$\hat{\beta}^{FP}$	0.4	0.216	0.455	
$\alpha = 0.3, \beta = -0.3$	$\hat{\alpha}^{MPLE}$	-1.935	1.819	2.656	0.832
	$\hat{\alpha}^{AMLE}$	-0.779	2.067	2.209	
	$\hat{\alpha}^{FP}$	2.6	0.97	2.775	
	$\hat{\beta}^{MPLE}$	1.753	0.812	1.932	0.814
	$\hat{\beta}^{AMLE}$	-1.331	0.837	1.572	
	$\hat{\beta}^{FP}$	1.9	0.881	2.094	
$\alpha = 0.3, \beta = 0.3$	$\hat{\alpha}^{MPLE}$	2.624	1.389	2.969	0.780
	$\hat{\alpha}^{AMLE}$	1.681	1.594	2.317	
	$\hat{\alpha}^{FP}$	-3.3	0.87	3.412	
	$\hat{\beta}^{MPLE}$	-2.676	0.639	2.751	0.769
	$\hat{\beta}^{AMLE}$	-1.703	1.256	2.116	
	$\hat{\beta}^{FP}$	3.1	0.929	3.236	

Table 6: Bias, standard deviation, RMSE and relative performance of $\hat{\alpha}^{AMLE}$ vs $\hat{\alpha}^{MPLE}$ and $\hat{\beta}^{AMLE}$ vs $\hat{\beta}^{MPLE}$ for various values of α and β, with $N = 12$, $M = 100$, $\epsilon = 5$, $m = 500$ and $B = 15$. Relative performance is defined as RelEff = RMSE($\hat{\theta}^{AMLE}$)/RMSE($\hat{\theta}^{MPLE}$), where θ is equal to α or β.

		Bias ($\times 10^2$)	Sd ($\times 10^2$)	RMSE ($\times 10^2$)	RelEff
$\alpha = -0.3, \beta = -0.3$	$\hat{\alpha}^{MPLE}$	1.106	2.117	2.388	1.013
	$\hat{\alpha}^{AMLE}$	−0.729	2.308	2.420	
	$\hat{\beta}^{MPLE}$	2.991	0.849	3.109	0.305
	$\hat{\beta}^{AMLE}$	−0.036	0.948	0.949	
$\alpha = 0, \beta = -0.3$	$\hat{\alpha}^{MPLE}$	−1.512	1.500	2.130	1.006
	$\hat{\alpha}^{AMLE}$	−1.387	1.788	2.143	
	$\hat{\beta}^{MPLE}$	2.042	0.782	2.186	0.432
	$\hat{\beta}^{AMLE}$	−0.046	0.943	0.945	
$\alpha = 0, \beta = 0.3$	$\hat{\alpha}^{MPLE}$	−1.778	1.270	2.185	0.689
	$\hat{\alpha}^{AMLE}$	−0.975	1.148	1.592	
	$\hat{\beta}^{MPLE}$	−2.507	1.094	2.735	0.413
	$\hat{\beta}^{AMLE}$	0.034	1.130	1.130	
$\alpha = 0.3, \beta = 0.3$	$\hat{\alpha}^{MPLE}$	2.633	1.636	3.100	0.757
	$\hat{\alpha}^{AMLE}$	1.708	1.611	2.348	
	$\hat{\beta}^{MPLE}$	−2.770	0.964	2.933	0.349
	$\hat{\beta}^{AMLE}$	−0.348	0.962	1.023	

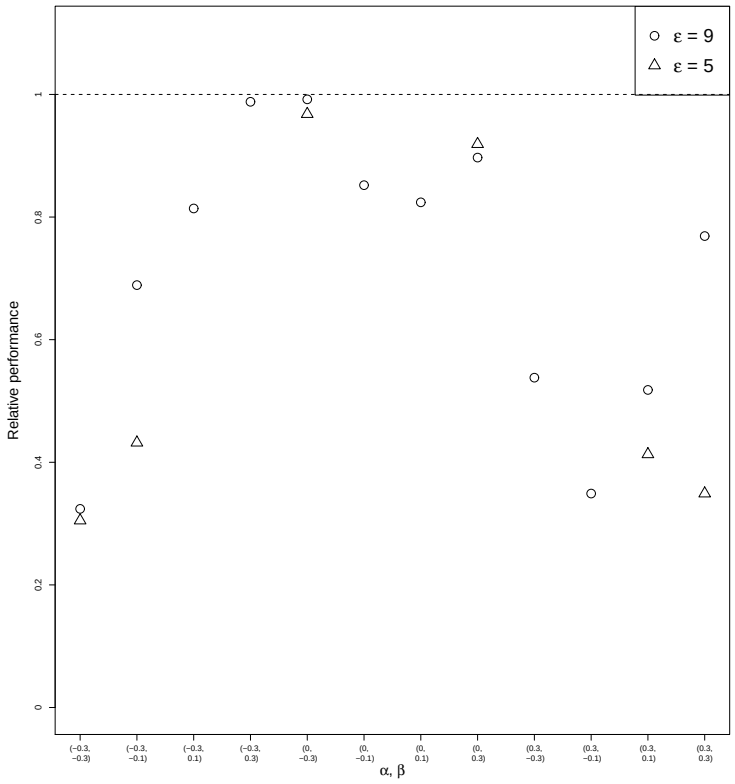

Figure 3: Relative performance of the estimator of β in the two-parameter case for various parameter values, $\epsilon \in \{5, 9\}$, $m = 500$.

$|\beta|$ is larger. The relative performance of the estimators of the intercept in cases with $\beta = 0.3$ and $\beta = -0.3$ is quite different.

Analogously to the single-parameter case, for all the experiments in Table 5 the rejection rate is between 0.991 and 0.993, so that, for simulating $m = 200$ pairs of parameters, the computing times are about the same. As for the setup of Table 6, the rejection rate is between 0.996 and 0.998, so that, in order to get 500 simulated parameter values, one needs to generate between approximately 130 000 and 230 000 autologistic models. The necessary amount of time is between about 12 and 22 hours respectively.

The pattern of the numerical results obtained by Friel and Pettitt (2004) is similar to ours; a notable difference is that their estimates seem to suffer from some systematic bias for large values of $|\beta|$.

4.1 Application: Wiebe's wheat data

We apply the AMLE method to the same dataset analyzed in Gu and Zhu (2001), now available in the `R` package `agridat` (Wright, 2013) using the command `wiebe.wheat.uniformity`. The data are the results of a uniformity trial of 1500 plots of wheat conducted in Idaho in 1927, and refer to a rectangular lattice of dimension 125×12 (for details see Andrews and Herzberg, 1985). Values larger than the mean are set equal to 1, the remaining values are equal to 0, so that the variable of interest is binary. We use AMLE with $m = 500$ and $\epsilon = 9$, obtaining $\hat{\beta}^{AMLE} = 0.383$ and $\hat{\sigma}_{\hat{\beta}^{AMLE}} = 0.0070$. The standard error has been computed via parametric bootstrap (see Sect. 3.2) with $B = 15$ replications. Considering that a 95% approximate confidence interval would be $[0.383 - 1.96 \cdot 0.007, 0.383 + 1.96 \cdot 0.007] = [0.369, 0.397]$, the result is in good agreement with the MLE computed by Gu and Zhu (2001), equal to 0.372 with an estimated standard deviation of 0.012.

5 Conclusion

In this paper we have estimated the parameters of the autologistic model via Approximate Maximum Likelihood. Since the properties of the estimators depend on the parameters of the algorithm, we ran simulation experiments aimed at measuring the precision of the estimators and its relationship to the input parameters.

The choice of the scaling factor ϵ and of the ABC sample size m seems to have a rather limited impact on the performance of the algorithm in the single-parameter case, but is more important in the two-parameter model, probably because kernel density estimation is less effective when the dimension increases.

Overall, the performance of the algorithm is excellent in the one-parameter case, where relative performance of AMLE in terms of RMSE, ranges be-

tween 19% and 43% when $|\beta| = 0.4$ and between 33% and 61% when $|\beta| = 0.2$. In the two-parameter case the big picture is similar, but only under more stringent conditions about ϵ and m, and especially as concerns the estimator of β: the relative performance of the estimator of β is between 0.305 and 0.432 for $|\beta| = 0.3$, and between 0.919 and 0.968 for $|\beta| = 0.1$. It is worth noting explicitly that in the two-parameter example our Monte Carlo evidence suggests that reducing ϵ and/or increasing m gives estimators with smaller RMSE. Therefore these outcomes, which are indeed essentially in line with MLE results obtained in the literature in analogous frameworks, can probably be improved, at the cost of an increased computational burden.

The AMLE method has two major advantages in the autologistic setup. First, it can be used for any dimension of the lattice, as long as the model can be simulated; basically, this means that the only limit is the computer's physical memory. Second, its implementation is easy. The computational cost is rather large, but this is a serious problem mostly in simulation experiments, where the model needs to be estimated several times. In real-data point estimation problems the procedure has to be run only once, so that setting a small ϵ and a large ABC sample size is usually not a major problem.

To conclude, we mention some important issues that require further research. First, as pointed out above, the method can be applied to other classical spatial models for which MLE is difficult. The most immediate examples are binary (or count) Markov random fields and the autonormal model. The former family of models is frequently used in various disciplines including ecology, agriculture, epidemiology, geography, hydrology and image analysis (Kaiser and Caragea, 2009). Application of AMLE to these cases needs to be studied thoroughly, but should critically depend on the possibility of simulating the models and computing summary (hopefully sufficient) statistics.

AMLE can probably also be used in more complicated spatial models where MLE is not available at all, for example when data are not only spatially dependent, but also clustered in groups, and nested or crossed effects are of specific interest (see e.g. the model proposed by Corrado and Fingleton, 2012).

As already noted, Algorithm 3 may be based on the MCMC-ABC approach by Marjoram et al. (2003); assessing whether the use of this algorithm has a major impact on the performance of AMLE deserves further scrutiny.

Building on the theoretical work by Ionides (2005) and on some of the results of our simulation study, it may be important to investigate whether, and, if the answer is positive, under which conditions, AMLE outperforms exact MLE in the autologistic case.

Finally, the determination of the numerical values of the parameters ϵ and m is difficult. Therefore, this issue has to be examined more thoroughly, especially in frameworks where simulating the model is computationally expensive, so that the effort needed to perform simulations with small ϵ and/or

large m is not affordable.

Acknowledgements

The authors would like to thank Fuchun Huang (Victoria University) for useful discussions concerning the simulation results and three anonymous reviewers whose valuable comments considerably improved an earlier version of this paper.

References

Andrews, D. F., Herzberg, A. M., 1985. Data. Springer.

Arbia, G., 2006. Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence. Advances in Spatial Science. Springer.

Arbia, G., Bee, M., Espa, G., 2011. Aggregation of regional economic time series with different spatial correlation structures. Geographical Analysis 43, 78–103.

Arbia, G., Bee, M., Espa, G., 2013. Testing isotropy in spatial econometric models. Spatial Economic Analysis 8, 228–240.

Beaumont, M. A., Zhang, W., Balding, D. J., 2002. Approximate Bayesian computation in population genetics. Genetics 162, 2025–2035.

Besag, J. E., 1975. Statistical analysis of non-lattice data. The Statistician 24, 179–195.

Casella, G., Robert, C., 2004. Monte Carlo Statistical Methods, 2nd Edition. Springer.

Cipra, B., 1987. An introduction to the Ising model. American Mathematical Monthly 94, 937–959.

Comets, F., 1992. On consistency of a class of estimators for exponential families of Markov random fields on the lattice. Annals of Statistics 20, 455–468.

Corrado, L., Fingleton, B., 2012. Where is the economics in spatial econometrics? Journal of Regional Science 52, 210–239.

Cox, D., Kartsonaki, C., 2012. The fitting of complex parametric models. Biometrika 99, 741–747.

Davison, A. C., Hinkley, D. V., 1997. Bootstrap Methods and their Application. Cambridge University Press.

Duong, T., 2014. Kde: Kernel smoothing. R package version 1.9.2.
URL http://CRAN.R-project.org/package=kde

Feizjavadian, S., Hashemi, R., 2015. Analysis of dependent competing risks in the presence of progressive hybrid censoring using Marshall-Olkin bivariate Weibull distribution. Computational Statistics and Data Analysis 82, 19–34.

Friel, N., Pettitt, A. N., 2004. Likelihood estimation and inference for the autologistic model. Journal of Computational and Graphical Statistics 13, 232–246.

Geman, S., Graffigne, G., 1987. Markov random field image models and their applications to computer vision. In: Gleason, A. M. (Ed.), Proceedings of the International Congress of Mathematics. American Mathematical Society.

Geyer, C. J., Thompson, E. A., 1992. Constrained Monte Carlo maximum likelihood for dependent data (with discussion). Journal of the Royal Statistical Society Series B 54, 657–699.

Grelaud, A., Robert, C. P., Marin, J.-M., Rodolphe, F., Taly, J.-F., 2009. ABC likelihood-free methods for model choice in Gibbs random fields. Bayesian analysis 4, 317–336.

Griffith, D. A., Arbia, G., 2010. Detecting negative spatial autocorrelation in georeferenced random variables. International Journal of Geographical Information Science 24, 417–437.

Gu, M.-G., Zhu, H.-T., 2001. Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation. Journal of the Royal Statistical Society B 63, 339–355.

Guyon, X., Künsch, H. R., 1992. Asymptotic Comparison of Estimators in the Ising Model. No. 74 in Stochastic Models, Statistical Methods, and Algorithms in Image Analysis, Lecture Notes in Statistics. Springer, pp. 177–198.

Huang, F., Ogata, Y., 2002. Generalized pseudo-likelihood estimates for Markov random fields on lattice. Annals of the Institute of Statistical Mathematics 54, 1–18.

Huffer, F., Wu, H., 1998. Markov Chain Monte Carlo for autologistic regression models with application to the distribution of plant species. Biometrics 54, 509–524.

Hughes, J., Haran, M., Caragea, P. C., 2011. Autologistic models for binary data on a lattice. Environmetrics 22, 857–871.

Ionides, E. L., 2005. Maximum smoothed likelihood estimation. Statistica Sinica 15, 1003–1014.

Ishiguro, M., Akaike, H., 1989. DALL: Davidon's algorithm for log-likelihood maximization - A FORTRAN subroutine for statistical model builders. Vol. 25 of Computer Science Monographs. The Instittute of Statistical Mathematics, Tokyo.

Kaiser, M. S., Caragea, P. C., 2009. Exploring dependence with data on spatial lattices. Biometrics 65, 857–865.

Marjoram, P., Molitor, J., Plagnol, V., S. Tavarè, S., December 2003. Markov Chain Monte Carlo without likelihoods. In: Proceedings of the National Academy of Sciences (U.S.A.). Vol. 100. pp. 15324–15328.

Metropolis, N., Rosenbluth, A., Rosenbluth, N., Teller, A., Teller, E., 1953. Equations of state calculations by fast computing machines. Journal of Chemical Physics 21, 1087–1092.

Moyeed, R. A., Baddeley, A. J., 1991. Stochastic approximation of the MLE for a spatial point pattern. Scandinavian Journal of Statistics 18, 39–50.

Müller, P., 1991. A generic approach to posterior integration and Gibbs sampling. working paper, Purdue University.

Murray, I., Ghahramani, Z., MacKay, D., 2006. MCMC for doubly-intractable distributions. In: Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence UAI06. pp. 359–366.

Ogata, Y., Tanemura, M., 1984. Likelihood analysis of spatial point patterns. Journal of the Royal Statistical Society. Series B 46, 496–518.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., Feldman, M. T., 1999. Population growth of human Y chromosomes: A study of Y chromosome microsatellites. Molecular Biology and Evolution 16, 1791–1798.

Propp, J., Wilson, D., 1996. Exact sampling with coupled Markov chains and applications to statistical mechanics. Random Structures and Algorithms 9, 223–252.

R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL `http://www.R-project.org/`

Rubio, F. J., Johansen, A. M., 2013. A simple approach to maximum intractable likelihood estimation. Electronic Journal of Statistics 7, 1632–1654.

Schabenberger, O., Gotway, C., 2002. Statistical Methods For Spatial Data Analysis. Chapman and Hall.

Sköld, M., Roberts, G. O., 2003. Density estimation for the Metropolis-Hastings algorithm. Scandinavian Journal of Statistics 30, 699–718.

Strauss, D., 1992. The many faces of logistic regression. The American Statistician 46, 321–327.

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

Wand, M. P., Jones, M. C., 1994. Multivariate plugin bandwidth selection. Computational Statistics 9, 97–116.

Wang, Z., Zheng, Y., 2013. Analysis of binary data via a centered spatial-temporal autologistic regression model. Environmental and Ecological Statistics 20, 37–57.

Wright, K., 2013. Agridat: Agricultural datasets. R package version 1.8. URL `http://CRAN.R-project.org/package=agridat`

Zheng, Y., Zhu, J., 2008. Markov Chain Monte Carlo for a spatial-temporal autologistic regression model. Journal of Computational and Graphical Statistics 17, 123–137.

List of recent papers published un the same series

2013/01 A regular multidistance among fuzzy numbers, Franco Molinari

2013/02 Employer moral hazard and wage rigidity. The case of worker-owned and investor-owned firms, Marina Albanese, Cecilia Navarra and Ermanno Tortia

2013/03 Origins and prospects of the Euro existential crisis, Luigi Bonatti and Andrea Fracasso

2013/04 Uses And Motivations For Credit Derivatives: An Empirical Investigation Into Italian Banks, Eleonora Broccardo, Maria Mazzuca and Elmas Yaldiz

2013/05 At the core of the international financial system, Valentina Feroldi, Edoardo Gaffeo

2013/06 Reassessing the spatial determinants of the growth of Italian SMEs, Roberto Gabriele, Diego Giuliani, Marco Corsino, Giuseppe Espa

2013/07 A spatial and sectoral analysis of firm demography in Italy, Giuseppe Espa, Danila Filipponi, Diego Giuliani, Davide Piacentino

2013/08 Using information markets in grantmaking. An assessment of the issues involved and an application to Italian banking foundations, Edoardo Gaffeo

2013/09 Interbank contagion and resolution procedures: inspecting the mechanism, Edoardo Gaffeo e Massimo Molinari

2013/10 Transatlantic contagion 2010-…, Roberto Tamborini

2013/11 Do middle managers matter?, Elena Feltrinelli, Roberto Gabriele, Sandro Trento