Computational Statistics and Data Analysis 84 (2015) 14–26

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Approximate maximum likelihood estimation of the autologistic model

Marco Bee *, Giuseppe Espa, Diego Giuliani

Department of Economics and Management, University of Trento, Italy

ARTICLE INFO

Article history:
Received 25 December 2013
Received in revised form 13 October 2014
Accepted 16 October 2014
Available online 30 October 2014

Keywords:
Spatial models
Sufficient statistics
Neighborhood relation
Kernel density

ABSTRACT

Approximate Maximum Likelihood Estimation (AMLE) is a simple and general method recently proposed for approximating MLEs without evaluating the likelihood function. The only requirement is the ability to simulate the model to be estimated. Thus, the method is quite appealing for spatial models because it does not require evaluation of the normalizing constant, which is often computationally intractable. An AMLE-based algorithm for parameter estimation of the autologistic model is proposed. The impact of the numerical choice of the input parameters of the algorithm is studied by means of extensive simulation experiments, and the outcomes are compared to existing approaches. AMLE is much more precise, in terms of Mean-Square-Error, with respect to Maximum pseudo-likelihood, and comparable to ML-type methods. Although the computing time is non-negligible, the implementation is straightforward and the convergence conditions are weak in most practically relevant cases.

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 Section 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,

* Correspondence to: via Inama 5, 38122 Trento, Italy. Tel.: +39 0461282296; fax: +39 0461282241.
E-mail addresses: marco.bee@unitn.it (M. Bee), giuseppe.espa@unitn.it (G. Espa), diego.giuliani@unitn.it (D. Giuliani).

http://dx.doi.org/10.1016/j.csda.2014.10.019
0167-9473/

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-i.i.d. sample from the approximated posterior. Although the theory does not require an i.i.d. 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 dependences, from specified conditionals (see, for example, Hughes et al., 2011).