Accepted Manuscript

Efficient likelihood computations for some multivariate Gaussian Markov random fields

L. Ippoliti, R.J. Martin, L. Romagnoli

PII: S0047-259X(17)30486-4
DOI: https://doi.org/10.1016/j.jmva.2018.07.007
Reference: YJMVA 4387

To appear in: *Journal of Multivariate Analysis*

Received date : 13 August 2017

Please cite this article as: L. Ippoliti, R.J. Martin, L. Romagnoli, Efficient likelihood computations for some multivariate Gaussian Markov random fields, *Journal of Multivariate Analysis* (2018), https://doi.org/10.1016/j.jmva.2018.07.007

Efficient likelihood computations for some multivariate Gaussian Markov random fields

L. Ippoliti[a,*], R.J. Martin[a], L. Romagnoli[b]

[a]*Dipartimento di Economia, Università degli Studi G. d'Annunzio, Viale Pindaro 42, 65127 Pescara, Italy*
[b]*Dipartimento di Economia, Università degli Studi del Molise, Via F. De Sanctis, 86100 Campobasso, Italy*

Abstract

Data collected from spatial locations are often multivariate. Gaussian conditional autoregressive (CAR) models, also known as Gaussian Markov random fields, are frequently used to analyze such continuous data, or as models for the parameters of discrete distributions. Two difficulties in Gaussian maximum likelihood estimation are ensuring that the parameter estimates are allowable values, and computing the likelihood efficiently. It is shown here that, for some commonly-used multivariate CAR models, checking for allowable parameter values can be facilitated, and the likelihood can be computed very quickly.

Keywords: Conditional autoregressive model, Gaussian Markov random fields, Lattice data, Maximum likelihood estimation, Multivariate observations, Regional data.

AMS 2000 subject classifications: 62H11, 62H12, 62H35

1. Introduction

Large amounts of essentially continuous spatial data are associated with the nodes or interiors of a regular rectangular lattice, or with irregularly spaced sites or irregularly shaped regions. For example, pixellated images are associated with the interiors of rectangular lattices, and some spatial sampling is at rectangular grid points; in contrast, epidemiological, ecological and environmental data are usually associated with irregular sites or regions. Frequently, there is more than one variable of interest. Spatial data observed at a few time points can also be treated as multivariate.

It is often reasonable to use, possibly after transformation, a Gaussian distribution to model continuous data. Gaussian models are also frequently used in hierarchical modeling of the parameters of discrete models, such as the Poisson and Logistic. Henceforth only Gaussian models are considered, which are defined by their mean and variance structure. These can be directly specified, but the likelihood can then be difficult to obtain. An alternative specification, which is considered here, and frequently used in applications, is that of conditional autoregressive (CAR) models, also known as Gaussian Markov random fields (GMRFs); see, e.g., Section 6.3.2 of [6] or [20]. These CAR models specify the mean and variance of the values at a site in terms of the values of a set, usually small, of nearby spatial sites.

For lattice data, the number of sites can be very large, and models for the mean-corrected observations are usually simplified, e.g., taken to be (approximately) stationary, or homogeneous. Several sets of dependence neighborhoods can be included, each with an associated parameter.

Many simple models with few parameters have been proposed for multivariate CARs on irregular regions. Examples are [8, 11–13, 18, 22]. In these papers, a simple form for the spatial dependence is used, based on a 0/1 neighborhood adjacency matrix. For irregular regional (or areal) data, the number of sites is often not large, with some models implying that both the strength of the dependence and the conditional variance vary with the number of neighbors.

*Corresponding author
Email addresses: ippoliti@unich.it (L. Ippoliti), luca.romagnoli@unimol.it (L. Romagnoli)