Kalman Filter for Estimating Bivariate GMRFs on Regular Lattice

Il Filtro di Kalman per la Stima di Modelli Spaziali Condizionali Bivariati su Lattice Regolare

Luigi Ippoliti, Richard Martin and Luca Romagnoli

Abstract Bivariate conditional autoregressive CAR models are widely used classes of multivariate spatial models. In this paper, we consider their maximum likelihood estimation which, under general boundary conditions, can be problematic. As in time series analysis, it will be shown that the Kalman filter provides an alternative and a computationally efficient solution for computing the likelihood.

Abstract *I modelli autoregressivi condizionali costituiscono una classe di modelli ampiamente usati nell'ambito dell'analisi di dati spaziali osservati su reticolo regolare. In questo lavoro, utilizzando una specificazione bivariata, si considera il problema della stima di massima verosimiglianza che, per condizioni di bordo generali, risulta problematica dal punto di vista computazionale. Come nell'analisi delle serie temporali, anche in questo ambito, sfruttando la natura ricorsiva del modello, dimostreremo che la verosimiglianza puó essere calcolata in maniera efficiente mediante il filtro di Kalman.*

Key words: Gaussian Markov Random Field, Maximum likelihood, Kalman filter, Spatial statistics, Image analysis

1 Introduction

Large amounts of essentially-continuous spatial data are associated with the nodes or interiors of a regular rectangular lattice. Examples include pixellated images which occur in many different applications, regularly-sampled spatial data, and

Luigi Ippoliti
University G. d'Annunzio, Chieti-Pescara, Italy, e-mail: ippoliti@unich.it

Richard Martin
University G. d'Annunzio, Chieti-Pescara, Italy, e-mail: r.j.martin26@gmail.com

Luca Romagnoli
University of Molise, Campobasso - Italy e-mail: luca.romagnoli@unimol.it

many agricultural field trials. Different types of models have been proposed for analyzing such data. Here, we consider stationary Gaussian conditional autoregressive (CAR) models, also known as Gauss-Markov random fields - GMRFs. We assume the GMRFs are defined on an infinite regular rectangular lattice, and applied to data on a complete finite lattice.

The paper is concerned with maximum likelihood - ML - parameter estimation of bivariate GMRFs. As discussed in Rue and Held (2005) the Markov property of a GMRF makes it possible to employ numerical methods for sparse matrices to construct fast algorithms operating on the precision matrix. Specifically, if we are prepared to assume toroidal boundary conditions for a lattice, then ML estimation is relatively straightforward. However, for alternative boundary conditions the implementation can be problematic.

Most of the difficulties can be overcome by exploiting the direct connection between the non-zero pattern of the precision matrix and the Markov properties of the GMRF. As discussed by Lavine (1999) and Moura and Balram (1992) by using the recursive structure of the GMRF, and its associated state-space representation, the Kalman filter constitutes an alternative, quick, and an efficient method of parameter estimation. In this paper, we shall show that for some parameterizations this approach is especially useful for estimating bivariate GMRFs.

The format of our paper is as follows. In section 2 we begin by briefly reviewing the univariate GMRF model, its ML parameter estimation and its state-space formulation. Section 3 introduces the bivariate GMRF. In particular, we discuss a specific parametrization for which the joint distribution is obtained as the product of the conditional and marginal distributions assumed to be univariate GMRFs. Here, when the eigenvalues of the precision matrix are not know, it is shown that the use of the Kalman filter offers a computationally efficient solution for computing the likelihood. Section 4 concludes the paper with a discussion of the methodological proposal.

2 Univariate Gauss-Markov random fields

Suppose that $\boldsymbol{t}$, $\boldsymbol{v}$ and $\boldsymbol{z}$ are two-dimensional vectors, and assume that $\left\{x(\boldsymbol{t}), \boldsymbol{t} \in Z^2\right\}$ is a second-order stationary random field on the regular rectangular lattice, with mean zero, autocovariance function $R_x(\boldsymbol{v}) = Cov\{x(\boldsymbol{t}), x(\boldsymbol{t}+\boldsymbol{v})\}$, and autocorrelation function $r_x(\boldsymbol{v}) = R_x(\boldsymbol{v})/\sigma_x^2$, where $\sigma_x^2 = R_x(\boldsymbol{0})$. Provided the sum is finite, the autocovariance generating function - acgf - of x is

$$\Gamma_x(\boldsymbol{z}) = \sum_{\boldsymbol{u} \in Z^2} R_x(\boldsymbol{v}) \boldsymbol{z}^{\boldsymbol{v}}, \;\; \boldsymbol{z} \in C^2$$

where $\sum |R_x(\boldsymbol{v})| < \infty$ and $\boldsymbol{z}^{\boldsymbol{v}} = \boldsymbol{z}_1^{v_1} \boldsymbol{z}_2^{v_2}$. Stationarity ensures that the acgf always exists for $|\boldsymbol{z}| = 1$, where $|\boldsymbol{z}| = \sqrt{(\boldsymbol{z}'\boldsymbol{z})}$. The spectral density function - sdf - of x is defined

here as $f_x(\boldsymbol{\lambda}) = \Gamma_x(e^{i\boldsymbol{\lambda}}) = \sum R_x(\boldsymbol{v})e^{i\boldsymbol{v}'\boldsymbol{\lambda}} = \sum R_x(\boldsymbol{v})\cos(\boldsymbol{v}'\boldsymbol{\lambda})$, where $\boldsymbol{\lambda} \in (-\pi,\pi]^2$; with its integral being $(2\pi)^2\sigma_x^2$. The inverse relationship is

$$R_x(\boldsymbol{v}) = (2\pi)^{-2}\int e^{i\boldsymbol{v}'\boldsymbol{\lambda}} f_x(\boldsymbol{\lambda})d\boldsymbol{\lambda}.$$

Let

$$A(\boldsymbol{z}) = -\sum_{\boldsymbol{j}\in S_p^+} \alpha_{\boldsymbol{j}}\boldsymbol{z}^{\boldsymbol{j}}$$

be a finite symmetric Laurent series with $\alpha_{\boldsymbol{0}} = -1$ and satisfying $A(\boldsymbol{z}) = A(\boldsymbol{z}^{-1})$, i.e. $\alpha_{\boldsymbol{j}} = \alpha_{-\boldsymbol{j}}$ for all $\boldsymbol{j}$, where S_p is a finite subset of $\boldsymbol{Z}^2$ containing neighbours of the origin. Thus $\boldsymbol{t}+S_p$ is the set of neighbours of site $\boldsymbol{t}$. The order of the neighbourhood set is denoted by p, and is defined sequentially by the maximum distance between the origin and a point in S_p. Hence, the first-order ($p=1$) neighbours of a site are those 4 sites which are adjacent to it; and the second-order neighbours ($p=2$) are these plus the 4 diagonally adjacent sites.

Let $[x(\boldsymbol{t})|\cdot] = \left[x(\boldsymbol{t})|x(\boldsymbol{t}-\boldsymbol{j}) : \boldsymbol{j} \in Z^2\backslash\{\boldsymbol{0}\}\right]$ denote $x(\boldsymbol{t})$ conditional on the values at all other sites. Then, under normality, the conditional autoregression CAR(p) or Gauss-Markov random field of order p, is defined by the conditional mean and the constant conditional variance

$$E\left[x(\boldsymbol{t})|\cdot\right] = \sum_{\boldsymbol{j}\in S_p} \alpha_{\boldsymbol{j}}x(\boldsymbol{t}-\boldsymbol{j}), \quad \mathrm{Var}\left[x(\boldsymbol{t})|\cdot\right] = \mathrm{Var}[\eta(\boldsymbol{t})] = \tau_x^2,$$

where $\alpha_{\boldsymbol{j}} = \alpha_{-\boldsymbol{j}}$ for all $\boldsymbol{j}$ and

$$\eta(\boldsymbol{t}) = x(\boldsymbol{t}) - E\left[x(\boldsymbol{t})|\cdot\right]$$

is the interpolation error process. Then

$$A(\boldsymbol{L})x(\boldsymbol{t}) = \eta(\boldsymbol{t}), \quad \boldsymbol{t} \in Z^2, \tag{1}$$

where $\boldsymbol{L}$ is a shift operator on an index, such that $\boldsymbol{L}^{\boldsymbol{j}}x(\boldsymbol{t}) = x(\boldsymbol{t}-\boldsymbol{j})$, and $\Gamma_{\eta x}(\boldsymbol{z}) = \tau_x^2$, i.e. the interpolation error $\eta(\boldsymbol{t})$ is uncorrelated with all x's except $x(\boldsymbol{t})$. We call $A(\boldsymbol{z})$ reflection-symmetric if the $\alpha_{\boldsymbol{j}}$ are equal for all sign changes on the $\boldsymbol{j}_i$, and completely-symmetric if they are also equal for all permutations of $\boldsymbol{j}$. Hence, we refer to a CAR(p) as reflection-symmetric, RS-CAR(p), or completely-symmetric, CS-CAR(p), if its $A(\boldsymbol{z})$ is.

2.1 The likelihood function

Consider an $(N\times M)$ lattice $\mathscr{L}$ with $n = NM$ sites and let $x(\boldsymbol{t})$, $\boldsymbol{t} \in \mathscr{L} \subset Z^2$, be the process observed at sites $(t_i,t_j), i = 1,\ldots,N; j = 1,\ldots,M$. The observed vector is thus written as $\boldsymbol{x} = \left[\boldsymbol{x}_1^T,\boldsymbol{x}_2^T,\ldots,\boldsymbol{x}_N^T\right]^T$, where $\boldsymbol{x}_i = \left[x(t_i,t_1),x(t_i,t_2),\ldots,x(t_i,t_M)\right]^T$.

Following equation (1), by using a lexicographic order for the sites and stacking the observations in the $(n \times 1)$ vector $\boldsymbol{x}$, a non-causal representation of the GMRF can be written as

$$\mathbf{A}(\boldsymbol{\alpha})\boldsymbol{x} = \boldsymbol{\eta}, \tag{2}$$

where $\mathbf{A}(\boldsymbol{\alpha})$ is the $n \times n$ *potential matrix*, with entries equal to 1 along the principal diagonal, $-\alpha_{\boldsymbol{j}}$ if the sites $\boldsymbol{t}$ and $\boldsymbol{t}-\boldsymbol{j}$ are neighbours, and zero otherwise. From equation (2) it readily follows that $\boldsymbol{x} \sim N\big(0, \tau_x^2 \mathbf{A}^{-1}(\boldsymbol{\alpha})\big)$, $\eta \sim N\big(0, \tau_x^2 \mathbf{A}(\boldsymbol{\alpha})\big)$ and the negative log-likelihood of $\boldsymbol{x}$ is

$$L(\boldsymbol{\alpha}, \tau_x^2 | \boldsymbol{x}) = \frac{n}{2}\log(2\pi\tau_x^2) + \frac{1}{2}|\mathbf{A}(\boldsymbol{\alpha})| + \frac{1}{2\tau_x^2}\boldsymbol{x}^T\mathbf{A}(\boldsymbol{\alpha})\boldsymbol{x}.$$

The ML fit of the model can be found by maximizing the likelihood over the valid parameter space to ensure the stationarity conditions. In general, the parameter space where the $\boldsymbol{\alpha}$ can take values is defined by the positive definite condition of the covariance matrix. Hence, in practice, optimization can be either over the covariance matrix, $\mathbf{R}_x = \tau_x^2\mathbf{A}^{-1}(\boldsymbol{\alpha})$, or the precision matrix, $\mathbf{R}_x^{-1}$, positive definite and the constraint that all the eigenvalues are positive is equivalent to positive definiteness.

An important part of model specification and estimation refers to the choice of the boundary conditions. Except for the torus assumption, the result for different boundary conditions is non-stationary - variances are no longer constant, and correlations at a given lag depend on the sites involved. Also, although there are computationally fast algorithms for first-order and reflection-symmetric second-order GMRFs with non-toroidal boundary conditions, computation is expensive. The use of the band Cholesky decomposition of $\mathbf{R}_x^{-1}$ can be used to simplify the problem (see, Rue and Held, 2005; sec 2.3-2.4). However, for very large lattices the algorithm can be computationally demanding.

2.2 The recursive structure of a finite GMRF

By following Moura and Balram (1992), it is shown that the recursive formulation intrinsic to a non-causal GMRF leads to a state-space representation which can greatly facilitate the computation of the likelihood. This representation is obtained using a Riccati equation and is written as

$$\begin{aligned} \mathbf{x}_i &= \boldsymbol{\Gamma}_i\mathbf{x}_{i+1} + \mathbf{G}_i\varepsilon_i, \;\; 1 \le i \le N-1 \\ \mathbf{x}_N &= \mathbf{G}_N\varepsilon_N, \end{aligned} \tag{3}$$

where $\mathbf{G}_i$ and $\boldsymbol{\Gamma}_i$ are parameter matrices obtained through the Cholesky decomposition of $\mathbf{A}(\boldsymbol{\alpha})$, and $E\left[\varepsilon_i\mathbf{x}_j\right] = 0$, for $j < i$.

It can be immediately recognized that equation (3) represents a "backward" state-space row model for a noise-free GMRF. Hence, as in time series analysis, the Kalman filter (Hamilton, 1994; p. 372) can be used to evaluate recursively the likelihood for parameter estimation. In fact, let $\hat{\boldsymbol{x}}_{i|i+1} = E\left[\boldsymbol{x}_i | \mathscr{X}_{i+1}\right]$ the least-squares forecast of the state vector $\boldsymbol{x}_i$ conditional on the information observed up to row $i+1$ and let $\boldsymbol{\Sigma}_{i|i+1}$ denote the corresponding mean squared error matrix. Then, the conditional distribution, $\boldsymbol{x}_i | \mathscr{X}_{i+1}$, is Gaussian with mean $\hat{\boldsymbol{x}}_{i|i+1}$ and covariance matrix $\boldsymbol{\Sigma}_{i|i+1}$, that is

$$f(\boldsymbol{x}_i | \mathscr{X}_{i+1}) \propto |\boldsymbol{\Sigma}_{i|i+1}|^{-1/2} exp\left\{(\boldsymbol{x}_i - \hat{\boldsymbol{x}}_{i|i+1})^T \boldsymbol{\Sigma}_{i|i+1}^{-1} (\boldsymbol{x}_i - \hat{\boldsymbol{x}}_{i|i+1})\right\},\ i = 1, \ldots, N. \quad (4)$$

From (4) the sample log-likelihood is thus obtained as

$$L(\boldsymbol{\alpha}, \tau_x^2 | \boldsymbol{x}) = \sum_{i=1}^{N} \log f(\boldsymbol{x}_i | \mathscr{X}_{i+1}),$$

which can be maximized numerically with respect to the unknown parameters τ_x^2 and $\boldsymbol{\alpha}$.

3 Bivariate GMRFs

In recent years, multivariate spatial models have been proven to be an effective tool for analyzing spatially related multidimensional data arising from a common underlying spatial process. Bivariate GMRF models have been mainly developed in the fields of medicine and public health to study the regional patterns of multiple diseases (see, for example, Kim et al. 2001). Applications with environmental data are instead considered by Sain et al. (2011).

Suppose that $\boldsymbol{u}_v = \left(\boldsymbol{y}^T\ \boldsymbol{x}^T\right)^T$ denotes a second-order stationary zero mean bivariate GMRF of order p represented in variable order (i.e. all n values for variable y and then for variable x). Let

$$\mathbf{D} = \frac{1}{\tau_x^2}\left(\mathbf{I} - \sum_{\boldsymbol{j} \in S_p} \alpha_{x,\boldsymbol{j}} \mathbf{W}^{(\boldsymbol{j})}\right), \quad \mathbf{E} = \frac{1}{\tau_y^2}\left(\mathbf{I} - \sum_{\boldsymbol{j} \in S_p} \alpha_{y,\boldsymbol{j}} \mathbf{W}^{(\boldsymbol{j})}\right)$$

and

$$\mathbf{F} = \beta_y \mathbf{I} - \sum_{\boldsymbol{j} \in S_p} \phi_{y,\boldsymbol{j}} \mathbf{W}^{(\boldsymbol{j})},$$

where $\mathbf{W}^{(\boldsymbol{j})}$ is the incidence matrix for neighbours corresponding to the model parameter matrices, $\mathbf{D} = \mathbf{R}_x^{-1} = \mathbf{Q}_{x.y}$, $\mathbf{E} = \mathbf{R}_{y.x}^{-1} = \mathbf{Q}_{yy}$ and $\mathbf{F} = \mathbf{R}_{yx}\mathbf{R}_x^{-1} = -\mathbf{Q}_{yy}^{-1}\mathbf{Q}_{yx}$. Then, the covariance and the precision matrices of the joint distribution can be written as

$$\mathbf{R}_{u_v} = \begin{bmatrix} \mathbf{E}^{-1} - \mathbf{F}\mathbf{D}^{-1}\mathbf{F}^T & \mathbf{F}\mathbf{D}^{-1} \\ \mathbf{D}^{-1}\mathbf{F}^T & \mathbf{D}^{-1} \end{bmatrix} \quad \textit{and} \quad \mathbf{R}_{u_v}^{-1} = \mathbf{Q}_{u_v} = \begin{bmatrix} \mathbf{E} & -\mathbf{E}\mathbf{F} \\ -\mathbf{F}^T\mathbf{E} & \mathbf{D} + \mathbf{F}^T\mathbf{E}\mathbf{F} \end{bmatrix}.$$

and their determinants can be computed as $|\mathbf{R}_{u_v}| = |\mathbf{D}^{-1}|\ |\mathbf{E}^{-1}|$ and $|\mathbf{Q}_{u_v}| = |\mathbf{Q}_{yy}|\ |\mathbf{Q}_{x.y}|$.

4 Simplifying the computational complexity

In general, under general boundary conditions and for a large p, the computation of the determinant is expensive. However, if we consider the case in which one of the two variables, say x, is assumed to be a predictor of y, the joint distribution can be written through the specification of simpler conditional and marginal forms, that is $f(\boldsymbol{u}) = f(\boldsymbol{y}|\boldsymbol{x})f(\boldsymbol{x})$. In fact, since the conditional distribution of $\boldsymbol{y}|\boldsymbol{x}$ is multivariate normal with mean $\boldsymbol{\mu}_{y|x} = \mathbf{F}\boldsymbol{x}$ and nonsingular covariance matrix $\mathbf{R}_{y.x} = \mathbf{Q}_{yy}^{-1}$, it turns out that the negative log-likelihood can be written as, $L(\boldsymbol{\theta}|\boldsymbol{u}) = L(\boldsymbol{\theta}|\boldsymbol{y}.\boldsymbol{x}) + L(\boldsymbol{\theta}|\boldsymbol{x})$, or equivalently,

$$L(\boldsymbol{\theta}|\boldsymbol{u}) = \log(2\pi) - \frac{1}{2n}\log\big(|\mathbf{Q}_{yy}|\ \ |\mathbf{Q}_{x.y}|\big) + \frac{1}{2n}\left(\boldsymbol{x}^T\mathbf{Q}_{x.y}\boldsymbol{x} + \tilde{\boldsymbol{y}}^T\mathbf{Q}_{yy}\tilde{\boldsymbol{y}}\right) \tag{5}$$

where $\boldsymbol{\theta}$ is the set of all model parameters and $\tilde{\boldsymbol{y}} = \boldsymbol{y} - \mathbf{F}\boldsymbol{x}$. Note that the likelihood has been scaled by $1/n$ for convenience.

Since the conditional and the marginal distributions appear as univariate GMRFs, the log-likelihood in (5) can be computed for general model parameterizations through $L(\boldsymbol{\theta}|\boldsymbol{y}.\boldsymbol{x})$ and $L(\boldsymbol{\theta}|\boldsymbol{x})$ which, in turn, can be evaluated by the Kalman filter as in section 2. The performance of the Kalman filter estimator will be assessed in a complete experimental study in the extended version of this article.

References

1. Hamilton, J.D.: Time series analysis. Princeton University Press, Princeton (1994)
2. Kim, H., Sun, D., Tsutakawa, R.K.: A bivariate Bayes methods for improving the estimates of mortality rates with a twofold conditional autoregressive model. JASA, **96**, 1506–1521 (2001)
3. Lavine, M.: Another look at conditionally Gaussian Markov random fields. In *Bayesian Statistics*, **6**, (Alcoceber, 1998), pp. 371–387. Oxford University Press, New York (1999)
4. Moura, J.M.F., Balram, N.: Recursive structure of noncausal Gauss-Markov random fields. IEEE Trans. Inf. Theory, **38**, 334–354 (1992)
5. Rue H., Held L.: Gaussian Markov random fields. Theory and applications. Chapman & Hall/CRC, Boca Raton (2005)
6. Sain, S. R., Furrer, R., Cressie, N.: A spatial analysis of multivariate output from regional climate models. Ann. Appl. Stat., **5**, 150–175 (2011)