Kalman Filter for Maximum Likelihood Estimation of GMRFs

Luigi Ippoliti and Luca Romagnoli

Abstract The paper is concerned with maximum likelihood - ML - parameter estimation for Gauss-Markov Random Fields - GMRFs. The ML estimator is straightforward under the unrealistic toroidal boundary conditions but difficult for Dirichlet and other boundary conditions for which, in general, the likelihood is evaluated at the cost of $O(n^3)$ steps for a lattice of n sites. In this paper, by exploiting the recursive structure of the GMRF, we show that using the Kalman filter recursions the ML estimator can be obtained at the cost of $O(n^2)$ steps.

Key words: Gaussian Markov Random Field, Maximum likelihood, Kalman filter

1 Gauss-Markov random fields

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 many agricultural field trials. Different types of models have been proposed for analyzing such data. Here, we consider those that specify the conditional distribution at each site given the values at all other sites - conditional autoregressive CAR models, also known as Gauss-Markov random fields - GMRFs. For convenience, we mainly refer here to Dryden et al. (2002) and Moura and Balram (1992) for known results.

We only consider univariate stationary Gaussian models here - i.e. they have finite variance, and the correlation between the observations at two sites only depends on the relative positions of the two sites. The assumption of normality for analysing

Luigi Ippoliti
Department of Economics, Viale Pindaro 42, 65127 Pescara, Italy, e-mail: ippoliti@unich.it

Luca Romagnoli
Department of Economics, Management and Social Sciences, Via F. De Sanctis, 86100 Campobasso, Italy, e-mail: luca.romagnoli@unimol.it

a data set is convenient and often reasonable (perhaps after transformation). We assume models are defined on an infinite regular rectangular lattice, and applied to data on a complete finite lattice. Then a GMRF is defined by its conditional means.

Suppose that t, u and z are $d-$dimensional vectors, and assume that $\{x(t), t \in Z^d\}$ is a second-order stationary random field on the regular rectangular lattice, with mean zero, autocovariance function $R_x(u) = Cov\{x(t), x(t+u)\}$, and autocorrelation function $r_x(u) = R_x(u)/\sigma_x^2$, where $\sigma_x^2 = R_x(0)$. Provided the sum is finite, the autocovariance generating function - acgf - of x is $\Gamma_x(z) = \sum_{u \in Z^d} R_x(u) z^u, \;\; z \in C^d$, with $\sum |R_x(u)| < \infty$ and $z^u = \prod_{i=1}^d z^{u_i}$. The stationarity conditions ensure that the acgf always exists for $|z| = 1$, where $|z| = \sqrt{(z'z)}$. Let $A(z) = 1 - \sum_{j \in S_p} \alpha_j z^j$ be a finite symmetric Laurent series satisfying $A(z) = A(z^{-1})$, i.e. $\alpha_j = \alpha_{-j}$ for all j, where S_p is a finite subset of Z^d containing neighbours of the origin. Thus $t + S_p$ is the set of neighbours of site 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. Thus, for $d = 2$, 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 $[x(t)|\cdot] = \left[x(t)|x(t-j) : j \in Z^d \backslash \{0\}\right]$ denote $x(t)$ conditional on the values at all other sites. Then, under normality, the conditional autoregression of order p - CAR(p) - is defined by the conditional mean,

$$E[x(t)|\cdot] = \sum_{j \in S_p} \alpha_j x(t-j),$$

where $\alpha_j = \alpha_{-j}$ for all j, and the constant conditional variance $\mathrm{Var}[x(t)|\cdot] = \mathrm{Var}[\eta(t)] = \sigma_\eta^2$, where $\eta(t) = x(t) - E[x(t)|\cdot]$ is the interpolation error process. Then

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

where L is a shift operator on an index, such that $L^j x(t) = x(t-j)$, and $\Gamma_{\eta x}(z) = \sigma_\eta^2$, i.e. the interpolation error $\eta(t)$ is uncorrelated with all x's except $x(t)$. Provided the stationarity condition $A(z) \neq 0$ for $|z| = 1$ is satisfied, the acgf of the CAR is $\Gamma_x(z) = \sigma_\eta^2 / A(z)$, and the acgf of $\eta(t)$ is $\Gamma_\eta(z) = \sigma_\eta^2 A(z)$. Thus the interpolation process is correlated, with $r_\eta(u)$ given by the inverse correlations of the original process $x(t)$ (Ippoliti et al., 2012).

2 The ML estimator and the Recursive structure of a finite GMRF

Suppose $d = 2$ and consider an $(N \times M)$ lattice Ω with $n = N \times M$ sites. Then, let $x(t),\ t \in \Omega$, be the process at sites $(t_i, t_j), i = 1, \ldots, N; j = 1, \ldots, M$. We write the vector, $\mathbf{X} = \left[\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_N^T\right]^T$, where $\mathbf{x}_i = \left[x(t_i, t_1), x(t_i, t_2), \ldots, x(t_i, t_M)\right]^T$ is the i-th

row of Ω. The $(n \times 1)$ vector $\mathbf{X}$ contains the sites in raster scan order - stacking the top row, then the second row, etc. Then, using matrix notation and following equation (1), a non-causal representation of the GMRF is

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

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

$$L(\alpha|\mathbf{X}) = \frac{n}{2}\log(2\pi\sigma_\eta^2) + \frac{1}{2}|\mathbf{A}(\alpha)| + \frac{1}{2\sigma_\eta^2}\mathbf{X}^T\mathbf{A}(\alpha)\mathbf{X}.$$

The ML fit of the model can be found by minimizing the likelihood over the valid parameter space. In practice, optimization can be over the covariance matrix, $\mathbf{R}_x = \sigma_\eta^2 \mathbf{A}^{-1}(\alpha)$, positive definite. There are however several problems associated with this. An important part of the GMRF model specification is the choice of boundary conditions (b.c.) for a stationary process, since elements of $\mathbf{R}_x$ for boundary sites on a finite lattice can be very complicated (Besag and Moran, 1975). When toroidal boundary conditions are assumed, the minimization of the likelihood for parameter estimation can be carried out with only $O(n \log n)$ steps (see for example, Dryden et al., 2002). However, 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. One option is to set values of $x(t)$ outside the lattice $\mathbf{L}$ to 0 (or the mean in general) - sometimes called the Dirichlet conditions. However, in general, apart from specific cases, these methods do not have $\mathbf{R}_x^{-1}$ in a form which can be rapidly used. In fact, although there are computationally fast algorithms for first-order and some second-order GMRFs with the more realistic Dirichlet boundary conditions, computation is expensive and in general takes $O(n^3)$ steps. In this paper we develop a simple method for inference with Dirichlet or other boundary conditions which, exploiting the recursive structure of the GMRF, is computationally fast and requires only $O(n^2)$ steps. For simplicity, we discuss the representation of a first-order GMRF with parameters α_{01} (for horizontal interactions) and α_{10} (for vertical interactions); however extensions to higher orders is straightforward.

Let $\mathbf{H}_M$ be the $(M \times M)$ matrix with entries equal to 1 just above and below the principal diagonal. Also, let $\mathbf{B} = \mathbf{I}_M - \alpha_{01}\mathbf{H}_M$ and $\mathbf{C} = -\alpha_{10}\mathbf{I}_M$, where $\mathbf{I}_M$ is the $(M \times M)$ identity matrix. Then, it follows that $\mathbf{A}(\alpha) = \mathbf{I}_N \otimes \mathbf{B} + \mathbf{H}_N \otimes \mathbf{C}$. The recursive structure of the GMRF consists of two equivalent one-sided representations of $x(t)$ that are obtained by the Cholesky decomposition of the potential matrix $\mathbf{A}(\alpha)$, using a Riccati equation. For example, since $\mathbf{A}(\alpha)$ is positive definite, we define the lower/upper Cholesky decomposition as $\mathbf{A}(\alpha) = \mathbf{U}^T\mathbf{U}$, where $\mathbf{U}$ is upper triangular. Then, from equation (2) we write $\mathbf{U}\mathbf{X} = \varepsilon$, where $\varepsilon = \mathbf{U}^{T^{-1}}\eta$. As a result, we thus have an equivalent one-sided "backward" regressor model from which it follows that $\varepsilon \sim N\big(\mathbf{0}, \sigma_\eta^2\mathbf{I}\big)$ and $E\big[\varepsilon\mathbf{X}^T\big] = \sigma_\eta^2\mathbf{U}^{-1}$.

The importance of this model lies in the fact that the block banded structure of $\mathbf{A}(\alpha)$ is reflected in the structure of $\mathbf{U}$. In fact, since $\mathbf{A}(\alpha)$ is block tridiagonal, $\mathbf{U}$ has only one nonzero off-diagonal block per block row with $(M \times M)$ block entries, $\mathbf{U}_i$ and $\mathbf{W}_i$, on the diagonal and off-diagonal block, respectively. Hence, the one-sided "backward" representation can be expanded as an $(M \times 1)$ vector AR process (see, Moura and Balram, 1992)

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

where $\mathbf{G}_i = \mathbf{U}_i^{-1}$, $\mathbf{F}_i = -\mathbf{U}_i^{-1}\mathbf{W}_i$ and $E\left[\varepsilon_i, \mathbf{x}_j\right] = 0$, for $j < i$. The blocks $\mathbf{U}_i$ and $\mathbf{W}_i$, which represent the spatially varying regressors of the one-sided AR representation, can be obtained as the solution of the following Riccati equations: a) $\mathbf{S}_1 = \mathbf{U}_1^T \mathbf{U}_1 = \mathbf{B}$, b) $\mathbf{W}_i = (\mathbf{U}_i^T)^{-1}\mathbf{C}$ and c) $\mathbf{S}_i = \mathbf{U}_i^T \mathbf{U}_i = \mathbf{B} - \mathbf{C}^T \mathbf{S}_{i-1}^{-1} \mathbf{C}, i = 2, \ldots, N$. Note that equation (3) represents a "backward" state-space row model for a noise-free GMRF and, as in time series analysis, the Kalman filter (Hamilton, 1994) can be used to evaluate recursively the likelihood for parameter estimation. In fact, let $\hat{\mathbf{x}}_{i|i+1} = E\left[\mathbf{x}_i | \mathscr{X}_{i+1}\right]$ the least square forecast of the state vector $\mathbf{x}_i$ conditional on the information observed up to row $i+1$ and $\mathbf{P}_{i|i+1}$ the corresponding mean squared error matrix. Then, the conditional distribution, $\mathbf{x}_i | \mathscr{X}_{i+1}$, is Gaussian with mean $\hat{\mathbf{x}}_{i|i+1}$ and covariance matrix $\mathbf{P}_{i|i+1}$, that is

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

From (4) it is then a simple matter to construct the sample log-likelihood, $\sum_{i=1}^{N} \log f(\mathbf{x}_i | \mathscr{X}_{i+1})$, which can be maximized numerically with respect to the unknown parameters σ_η^2 and α. Note that equation (4) can be evaluated at the cost of $O(n^2)$ steps and can be easily extended to deal with noisy data. Finally, note that a one-sided "forward" representation of the GMRF can also be defined. However, since the two representations are equivalent, only one of them can be considered.

References

1. Besag, J.E, Moran, P.A.P.: On the estimation and testing of spatial interaction in Gaussian lattice processes. Biometrika. **62**, 555–562 (1975)
2. Dryden, I., Ippoliti, L., Romagnoli, L.: Adjusted maximum likelihood and pseudo-likelihood estimation for noisy Gaussian Markov random fields. J. Comput. Graph. Stat. **11**, 370–388 (2002)
3. Hamilton, J.D.: Time series analysis. Princeton University Press, Princeton. (1994)
4. Ippoliti, L., Martin, R.J., Bhansali, R.J.: Rational spectral density models for lattice data. Submitted for publication. Available from the authors. (2012)
5. Moura, J.M.F., Balram, N.: Recursive structure of noncausal Gauss-Markov random fields. IEEE Trans. Inform. Theory. **38**, 334–354 (1992)