Maximum Likelihood Estimation for Spatial Models

Kanti V. Mardia *

Department of Statistics, University of Leeds, Leeds, LS2 9JT, UK

Overview: The paper gives maximum likelihood (ML) estimation methods for spatial linear models in three forms: Direct Representation (DR), Conditional Autoregressive (CAR) Models and Simultaneous Autoregressive (SAR) Models for the Gaussian Case. We also discuss the computational aspects of the methods. The problem of asymptotic bias is considered for the DR. For the intrinsic random field we obtain the maximum likelihood estimators (MLEs) and indicate a relationship with marginal likelihood. We give the exact MLEs for CAR models on a circle and a torus together with some properties. It is indicated how the same approach applies to the SAR models. For the stationary random field, we discuss the Whittle approximation. We also consider the MLE for intrinsic CAR. We then describe the estimation of a nugget parameter and its asymptotic distribution. We indicate the extension of the method to the multivariate case, block data, missing values in lattice, designs under spatial correlation, and the such. Finally, a general discussion is given.

1. Introduction

Spatial models have become increasingly used for image analysis (see, Mardia, 1989 for references). Previous applications were more in agriculture, forestry, ecology, geography, geology, to name a few disciplines. We mainly study ML estimation for three forms of spatial linear models: Direct Representation (DR), Conditional Auto-regressive (CAR) and Simultaneous Auto-regressive (SAR) when the random field is Gaussian. Finite range covariance schemes lead to a sparse covariance matrix Σ of the random field in DR, whereas CAR and SAR lead to sparse Σ^{-1}. However, the models are interrelated; see Section 2.

Section 3 gives the MLEs and their problems for DR. Section 4 gives the MLEs for the intrinsic model. Section 5 gives CAR models with some comments on the SAR case. Section 6 discusses the problem for a nugget parameter or errors in variable model. Section 7 considers some other uses and extensions such as block data, missing values in a lattice and the use of these methods in experimental design. The last section gives a discussion.

* The author is grateful to Dan Griffith for inviting him to participate in this Symposium. He is also grateful to John Kent, Tim Hainsworth and Alan Watkins for their helpful comments. This work is supported by NSF grant DMS-8803207 to Professor Watson, Princeton University.

2. The model and its three forms

2.1. The spatial linear model

Let $\{X(\mathbf{t})\}$ be a stochastic process where $\mathbf{t}$ represents a point in d-dimensional space where we write $T = R^d$ for the Euclidean space and $T = Z^d$ for points on a regular lattice. Suppose the process is sampled at points $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_n$ to give the sample vector

$$\mathbf{X} = \{X(\mathbf{t}_1), X(\mathbf{t}_2), \ldots, X(\mathbf{t}_n)\}'.$$

Suppose that $\{X(\mathbf{t})\}$ is Gaussian and that

$$\mu(\mathbf{t}) = E\{X(\mathbf{t})\} = \mathbf{f}(\mathbf{t})'\boldsymbol{\beta}, \tag{2.1.1}$$

where $\boldsymbol{\beta}$ is a q-by-1 parameter vector and $\mathbf{f}(\mathbf{t})$ is a vector of known functions, possibly monomials, which may describe a trend. If there is no confusion we will write

$$\mathbf{X} = (X_1, X_2, \ldots, X_n)' \text{ and } \boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)'.$$

Suppose that

$$\{X(\mathbf{t}) - \mu(\mathbf{t})\}$$

is second-order stationary with

$$\mathrm{Cov}\{X(\mathbf{t}),\, X(\mathbf{t}+\mathbf{h})\} = \sigma(\mathbf{h}; \boldsymbol{\theta}), \tag{2.1.2}$$

where $\sigma(\cdot;\, \boldsymbol{\theta})$ is a positive definite function of $\mathbf{h}$, assumed known apart for a p-by-1 vector of parameters $\boldsymbol{\theta}$. [We will discuss some restrictions on $\sigma(\cdot;\, \boldsymbol{\theta})$ for the asymptotic theory in Section 3.] Thus from (2.1.1),

$$E(\mathbf{X}) = \mathbf{F}\boldsymbol{\beta}, \tag{2.1.3}$$

where $\mathbf{F} = \{\mathbf{f}(\mathbf{t}_1), \mathbf{f}(\mathbf{t}_2), \ldots, \mathbf{f}(\mathbf{t}_n)\}'$. Let the covariance matrix of $\mathbf{X}$ be $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ with

$$\sigma_{ij} = (\boldsymbol{\Sigma})_{ij} = \{\sigma(\mathbf{t}_i - \mathbf{t}_j;\, \boldsymbol{\theta})\}.$$

Thus our model is of the form

$$\text{observation} = \text{deterministic trend} + \text{stochastic fluctuation}$$

where trend measures the long-term variation whereas the stochastic fluctuation measures the short-term or the local variation.

2.2. The direct representation

We can specify $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ by modelling $\sigma(\mathbf{h}; \boldsymbol{\theta})$ directly. With this formulation, we will call the spatial linear model the Direct Representation (DR).

Consider the following example. Table 1 gives a topographic data set consisting of 52 points from Davis (1973, pp. 313-314) with $T = R^2$. Figure 1 plots the data, and looking at the values closely, there is some indication of trend. At a smaller scale, we will expect local variation. One way to model the local variation, is to take $\sigma(\mathbf{h}; \boldsymbol{\theta})$ with finite *range*

TABLE 1
GEOGRAPHIC COORDINATES
AND ELEVATIONS OF CONTROL POINTS
FOR EXAMPLE SURVEYING PROBLEM

E–W Coordinate t_1	N–S Coordinate t_2	Elevation $X(t_1, t_2)$	E–W Coordinate t_1	N–S Coordinate t_2	Elevation $X(t_1, t_2)$
0.3	6.1	870	5.2	3.2	805
1.4	6.2	793	6.3	3.4	840
2.4	6.1	755	0.3	2.4	890
3.6	6.2	690	2.0	2.7	820
5.7	6.2	800	3.8	2.3	873
1.6	5.2	800	6.3	2.2	875
2.9	5.1	730	0.6	1.7	873
3.4	5.3	728	1.5	1.8	865
3.4	5.7	710	2.1	1.8	841
4.8	5.6	780	2.1	1.1	862
5.3	5.0	804	3.1	1.1	908
6.2	5.2	855	4.5	1.8	855
0.2	4.3	830	5.5	1.7	850
0.9	4.2	813	5.7	1.0	882
2.3	4.8	762	6.2	1.0	910
2.5	4.5	765	0.4	0.5	940
3.0	4.5	740	1.4	0.6	915
3.5	4.5	765	1.4	0.1	890
4.1	4.6	760	2.1	0.7	880
4.9	4.2	790	2.3	0.3	870
6.3	4.3	820	3.1	0.0	880
0.9	3.2	855	4.1	0.8	960
1.7	3.8	812	5.4	0.4	890
2.4	3.8	773	6.0	0.1	860
3.7	3.5	812	5.7	3.0	830
4.5	3.2	827	3.6	6.0	705

Elevation is measured in feet above sea level. Coordinates are expressed in 50–feet units measured from an arbitrary origin located in the southwest corner, with t_1 being the East–West coordinate and t_2 being the North–South coordinate (from Davis, 1973).

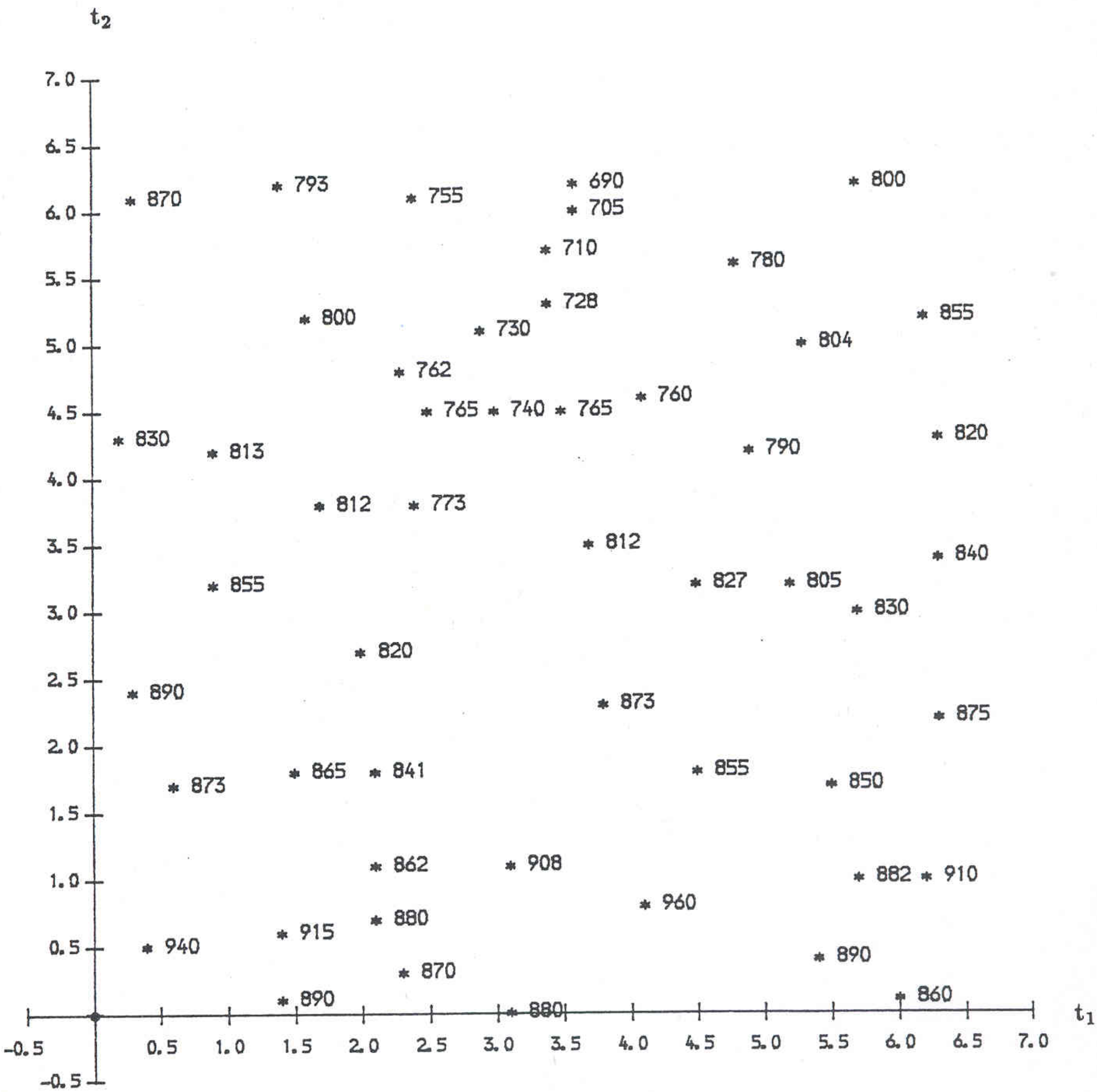
α so that $\sigma(\mathbf{h}; \boldsymbol{\theta}) = 0$ for $|\mathbf{h}| > \alpha$. For example, we could use the power scheme where the covariance function is given by (Mardia and Watkins, 1989)

$$\sigma(\mathbf{h}; \boldsymbol{\theta}) = \sigma^2(1 - \alpha^{-1}|\mathbf{h}|)^4, \quad |\mathbf{h}| < \alpha; \ = 0 \text{ otherwise} \tag{2.2.1}$$

where $\boldsymbol{\theta} = (\sigma^2, \alpha)'$. We note that $\Sigma(\boldsymbol{\theta})$ is sparse if α is less than the maximum distance between the points. Further, the sites can be irregularly distributed.

Figure 1.

Spatial distribution of elevations (Davis, 1973).



2.3. The conditional autoregressive model

Another way to model the spatial linear model is to specify $\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}$ through a conditional autoregression (CAR) model. We will mainly concentrate here on the case when all the n sites are on a regular lattice. On an infinite regular lattice Z^d, the CAR model is defined by (Besag, 1974)

$$E(X_i \mid \text{rest}) = \mu_i + \sum_{j \neq i} \phi_{ij}(\boldsymbol{\theta})(x_j - \mu_j), \tag{2.3.1}$$

$$\text{Var}(X_i|\text{rest}) = \tau^2, \tag{2.3.2}$$

where 'rest' denotes all points $j \in Z^d, j \neq i$, and $\boldsymbol{\Phi}(\boldsymbol{\theta}) = [\phi_{ij}(\boldsymbol{\theta})]$ is a symmetric matrix with $\phi_{ii}(\boldsymbol{\theta}) = 0$, and $\phi_{ij}(\boldsymbol{\theta})$ a function such that the process $\{X(\mathbf{t})\}$ is covariance stationary. In particular, we take

$$\begin{aligned} \phi_{ij}(\boldsymbol{\theta}) &= \theta_{i-j} \text{ for } i-j \in N, \\ &= 0 \text{ otherwise} \end{aligned} \tag{2.3.3}$$

where N denotes a finite symmetric neighbourhood of the origin and $\theta_{i-j} = \theta_{j-i}$. For example, for the first-order neighbourhood in 2-dimensions,

$$N = \{(-1,0),(1,0),(0,1),(0,-1)\}, \tag{2.3.4}$$

and there are two parameters in $\boldsymbol{\Phi}(\boldsymbol{\theta})$, θ_{10} and θ_{01} which we will write as θ_1 and θ_2 respectively. Thus for $\mu_i = 0$; the CAR is defined by

$$E(X_i \mid \text{rest}) = \sum_{j \in N} \theta_{i-j} x_j, \ \text{Var}(X_i \mid \text{rest}) = \tau^2. \tag{2.3.5}$$

Sometimes it will be convenient to write the parameters as

$$\phi_h = \tau^{-2}, \ \mathbf{h} = \mathbf{0}; \ \phi_h = \tau^{-2}\theta_h, \ \mathbf{h} \neq \mathbf{0}. \tag{2.3.6}$$

The simplest example of (2.3.5) is when $\theta_{i-j} = \theta$ so that

$$\boldsymbol{\Phi}(\boldsymbol{\theta}) = \theta \mathbf{W}, \ E(X_i \mid \text{rest}) = \theta \sum_{j \in N} x_{i+j}, \tag{2.3.7}$$

where $(\mathbf{W})_{ij} = 1$ if $i - j \in N; = 0$ otherwise. $\mathbf{W}$ is called the adjacency matrix. We will call this the basic CAR model.

If $\mathbf{X}$ is $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we have

$$E(X_i \mid \text{rest}) = \mu_i + \sum_{j \neq i} (\sigma^{ij}/\sigma^{ii})(x_j - \mu_j), \text{ and } \text{Var}(X_i \mid \text{rest}) = 1/\sigma^{ii}.$$

On identifying these two expressions with (2.3.1) and (2.3.2) respectively, we obtain

$$\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} = \tau^{-2}[\mathbf{I} - \boldsymbol{\Phi}(\boldsymbol{\theta})]. \tag{2.3.8}$$

It should be noted that the infinite matrices $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}$ have to be absolutely convergent and positive definite. $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is certainly p.d. if $|1 - \sum_{h \in N} \theta_h \cos(\boldsymbol{\omega}'\mathbf{h})| < 1$. This holds if

$$\sum_{h \in N} |\theta_h| < 1 \tag{2.3.9}$$

[since $|\cos(\boldsymbol{\omega}'\mathbf{h})| \leq 1$] and therefore in the basic model $|\theta| < 1/\nu$ where ν is the number of neighbours.

In general, it should be noted that for the stationary CAR,

$$\sigma(\mathbf{h}; \boldsymbol{\theta}) = (2\pi)^{-d} \int_{(-\pi,\pi)^d} [\exp(i\boldsymbol{\omega}'\mathbf{h})]/[\sum_{s \in N_0} \phi_s \cos(\boldsymbol{\omega}'\mathbf{s})] d\boldsymbol{\omega}, \tag{2.3.10}$$

where the set N_0 is N with the origin included. Hence the class of stationary process includes the class of CARs. We can use $\sigma(\mathbf{h}; \boldsymbol{\theta})$ in the DR but here $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is complicated. However, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is simple and sparse.

So far we have discussed the CAR on infinite lattice Z^2 but in practice, our sites are on a finite lattice D. Let $C = Z^2 - D$. We first obtain the inverse covariance matrix of $\{X(\mathbf{t})\}, t \in D$.

For the infinite lattice, we can write the inverse covariance matrix (2.3.8) of $\mathbf{X}$ as

$$\tau^2 \boldsymbol{\Sigma}_{\infty}^{-1} = \mathbf{I}_{\infty} - \boldsymbol{\Phi}_{\infty} \equiv \begin{pmatrix} \mathbf{I}_D - \boldsymbol{\Phi}_D & \mathbf{B} \\ \mathbf{B}' & \mathbf{I}_c - \boldsymbol{\Phi}_c \end{pmatrix},$$

where $\mathbf{I}_D$ and $\mathbf{I}_c$ are the identity matrix, $\boldsymbol{\Phi}_D$ matrix for the finite lattice D and so forth. Then Künsch (1983) has shown that the inverse covariance matrix for the finite lattice D is

$$\boldsymbol{\Sigma}_D^{-1} = \mathbf{I}_D - \boldsymbol{\Phi}_D - \boldsymbol{\Gamma}_D,$$

where

$$\boldsymbol{\Gamma}_D = \mathbf{B}_D(\mathbf{I}_c - \boldsymbol{\Phi}_c)^{-1}\mathbf{B}'.$$

provided all matrices converge absolutely. Thus (2.3.8) is not valid for D unless $\boldsymbol{\Gamma}_D = \mathbf{0}$. However, the exact $\boldsymbol{\Sigma}_D^{-1}$ can be obtained through $\boldsymbol{\Sigma}_D$ whose elements are obtained from (2.3.10).

In fact, $\boldsymbol{\Sigma}_D(\boldsymbol{\theta})$ is the covariance function of the marginal distribution of $\{X(\mathbf{t})\}$ on $\mathbf{t} \in D$, and therefore the process is stationary on $\mathbf{t} \in D$ with $\sigma(\mathbf{h}; \boldsymbol{\theta})$ given by (2.3.10). We will call this process an M–CAR. However, $\boldsymbol{\Sigma}_D(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_D(\boldsymbol{\theta})^{-1}$ are both complicated, unlike $\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}$ given by (2.3.8) with $\boldsymbol{\Phi}(\boldsymbol{\theta})$ defined at (2.3.3). We can achieve some simplicity by making boundary adjustments in the following two ways:

(i) T–CAR. Wrap the CAR on a torus.

(ii) C–CAR. For $\mu(\mathbf{t}) = 0$, use the conditional distribution of $\{X(\mathbf{t})\}$ on D given $X(\mathbf{t}) = 0$, $\mathbf{t} \notin D$, *i. e.* use the free boundaries.

Under the C–CAR, the CAR representation (2.3.5) is preserved but we do not have stationarity. Under the T–CAR, we have stationarity as well as the CAR representation but the periodic boundaries are not realistic.

Using the C–CAR can lead to serious bias in estimation for large n. The main reason is that the boundary for $d = 2$ is of order $n^{1/2}$ and the effect of neglecting it can be of order higher than n^{-1}. For some practical examples see Guyon (1982). Martin (1987) has highlighted some confusion in this area. The above result (2.3.8) requires basic knowledge of conditional distributions for the multivariate normal case, especially Theorems 3.2.3 and 3.2.4 in Mardia, Kent and Bibby (1989).

Suppose the sites are irregularly distributed; then the extension of the CAR model is not straightforward, in general (see, Besag, 1975). An interesting particular case is for $\mathbf{\Phi}(\boldsymbol{\theta})$ given by (2.3.7), where $W_{ij} = 1$ if i and j are nearest neighbours, and zero otherwise. Also for the regularized process, equation (2.3.7) can be used, where now

$W_{ij} = 0$ if areas are not contiguous, and

$\propto$ a monotonic function of the length of the common boundary otherwise.

If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\mathbf{W}$, with $\lambda_1 < \cdots < \lambda_n$, then $\mathbf{\Sigma}$ is positive definite if $0 \le \theta < 1/\lambda_n$.

2.4. The simultaneous autoregressive model

For simplicity let us assume the finite lattice case. We have

$$\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\psi}(\boldsymbol{\theta})(\mathbf{X} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}, \ \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}).$$

Thus with $\boldsymbol{\psi} \equiv \boldsymbol{\psi}(\boldsymbol{\theta})$,

$$\mathbf{\Sigma}(\boldsymbol{\theta})^{-1} = \sigma^{-2}(\mathbf{I} - \boldsymbol{\psi}')(\mathbf{I} - \boldsymbol{\psi}), \text{ or } \mathbf{\Sigma}(\boldsymbol{\theta}) = \sigma^2(\mathbf{I} - \boldsymbol{\psi})^{-1}(\mathbf{I} - \boldsymbol{\psi}')^{-1}, \tag{2.4.1}$$

where the defining property requires $|\mathbf{I} - \boldsymbol{\psi}|$ to be non–singular. As in the CAR case, we can take a particular case as $\boldsymbol{\psi} = \theta\mathbf{W}$ so that $\mathbf{\Sigma}(\boldsymbol{\theta})^{-1}$ is sparse. Here $\mathbf{W}$ need not be a symmetric matrix. For almost all values of θ, the class of the SARs is included in the class of the CARs on the infinite lattice. However, note that $\boldsymbol{\psi}$ is not uniquely determined by a given $\mathbf{\Sigma}(\boldsymbol{\theta})$, unlike the CAR. We will not discuss the estimation problems for the SAR in detail.

2.5. Particular cases

We now give these three representations for the Geometric Scheme, where the correlation function is

$$\rho(\mathbf{h}; \lambda, \nu) = \lambda^{|h_1|}\nu^{|h_2|}. \tag{2.5.1}$$

We have its SAR and CAR representations on the infinite lattice as

SAR: $X_{ij} = \lambda X_{i-1,j} + \nu X_{i,j-1} - \lambda\nu X_{i-1,j-1} + \varepsilon_{ij}$,

CAR: $E(X_{ij}|\cdot) = \alpha(x_{i-1,j} + x_{i+1,j}) + \beta(x_{i,j-1} + x_{i,j+1}) - \alpha\beta(x_{i-1,j-1} + x_{i-1,j+1} + x_{i+1,j-1} + x_{i+1,j+1})$,

Var $(X_{ij}|\cdot) = \sigma^2/(1 + \lambda^2)(1 + \nu^2)$,

where $\alpha = \lambda/(1+\lambda^2)$ and $\beta = \nu/(1+\nu^2)$. From $E(X_{ij}|\cdot)$, it follows that the neighbourhood is of the second order.

For the first–order neighbourhood for the CAR in Z^2, we have from (2.3.10)

$$\sigma(h_1, h_2) = \tau^2(2\pi)^{-2} \int_{(-\pi,\pi)^2} \frac{\cos(\omega_1 h_1)\cos(\omega_2 h_2)}{1 - 2\theta_1\cos(\omega_1) - 2\theta_2\cos(\omega_2)}\, d\omega_1 d\omega_2. \tag{2.5.2}$$

Besag (1981) shows that for $\theta_1 + \theta_2 > 0.48$ and $(h_1, h_2) \neq (0,0)$,

$$\sigma(h_1, h_2) \simeq \tau^2\{2\pi(\theta_1\theta_2)^{1/2}\}^{-1}K_0\left((1-2\theta_1-2\theta_2)^{1/2}[\frac{h_1^2}{\theta_1^2}+\frac{h_2^2}{\theta_2^2}]^{1/2}\right),$$

where $K_0(\cdot)$ is the modified Bessel function of the second kind and order zero. In particular, for $\theta_1 = \theta_2 = \theta$, $\sigma(h_1, h_2)$ is closely approximated by a monotonic decreasing function of $|\mathbf{h}| = (h_1^2+h_2^2)^{1/2}$, so that it is almost an isotropic scheme. Further, there is very slow decay of $\rho(\mathbf{h})$ with increasing $|\mathbf{h}|$ whenever $\rho(1,0)$ is moderately high; *e. g.*, if $\rho(1,0) = 0.85$, then $|\mathbf{h}|$ must exceed 2000 before $\rho(\mathbf{h}) < 0.1$. Note that the Geometric Scheme is highly anisotropic and cannot display the type of slow decay for the first-order CAR scheme considered here.

3. ML estimation for DR

3.1. ML equations

In this Section, we follow Mardia and Marshall (1984). Since from Section 2.1 $\mathbf{X}$ is multivariate normal, the log-likelihood function of $\mathbf{X}$ with the parameters $(\boldsymbol{\beta}, \boldsymbol{\theta})$ is

$$\ell = \ell(\mathbf{X}; \boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \frac{1}{2}(\mathbf{X}-\mathbf{F}\boldsymbol{\beta})'[\boldsymbol{\Sigma}(\boldsymbol{\theta})]^{-1}(\mathbf{X}-\mathbf{F}\boldsymbol{\beta}). \tag{3.1.1}$$

On differentiating (3.1.1) with respect to $\boldsymbol{\beta}$, with the help of $\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$ we get

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \mathbf{F}'\boldsymbol{\Sigma}^{-1}\mathbf{X} - (\mathbf{F}'\boldsymbol{\Sigma}^{-1}\mathbf{F})\boldsymbol{\beta}. \tag{3.1.2}$$

Hence, the MLE of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{F})^{-1}\mathbf{F}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}, \tag{3.1.3}$$

where $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$, $\hat{\boldsymbol{\theta}}$ being the MLE of $\boldsymbol{\theta}$. To differentiate (3.1.1) with respect to $\boldsymbol{\theta}$, we first note the following two results:

$$\frac{\partial \log|\boldsymbol{\Sigma}|}{\partial \theta_i} = \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i), \quad \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \theta_i} = -\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1},$$

where $\boldsymbol{\Sigma}_i = \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}$. Hence differentiating with respect to θ_i, with the help of these two results, we have

$$\frac{\partial \ell}{\partial \theta_i} = -\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i) + \frac{1}{2}\mathbf{w}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\mathbf{w}, \quad i = 1, \ldots, p, \tag{3.1.4}$$

where $\mathbf{w} = \mathbf{X} - \mathbf{F}\boldsymbol{\beta}$. Thus the p equations for $\hat{\boldsymbol{\theta}}$ are

$$\hat{\mathbf{w}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}_i\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{w}} = \mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}_i), i = 1, \ldots, p, \tag{3.1.5}$$

with $\hat{\mathbf{w}} = \mathbf{X} - \mathbf{F}\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$. We note that the equations hold even when $\{X(\mathbf{t})\}$ is not covariance stationary.

One does not see an analytical solution to the ML equations (3.1.3) and (3.1.5). However, for a nested scheme such that

$$\sigma(\mathbf{h};\boldsymbol{\theta}) = \sum_{i=1}^{p} \theta_i \sigma_i(|\mathbf{h}|) \tag{3.1.6}$$

some progress can be made, since $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is of the form $\Sigma \mathbf{K}_i \theta_i$, where the matrices $\mathbf{K}_i$ are fixed. These are realistic models in the Analysis of Variance (see, Hocking, 1984), but not so realistic in Spatial Statistics. However, we will just give one example for its mathematical contents, namely the variance component scheme. Suppose the process is stationary, with

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \theta_1 \mathbf{I}_n + \theta_2(\mathbf{I}_r \otimes \mathbf{E}_s), \tag{3.1.7}$$

where $n = rs$, and $\mathbf{E}_s$ is an s-by-s matrix of 1s. Then $\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}}$ and the MLEs of θ_1 and θ_2 are obtained from the solution to

$$r(s-1)\lambda_0^{-1} + (r-1)\lambda_1^{-1} + \lambda_2^{-1} = q_0\lambda_0^{-2} + q_1\lambda_1^{-2},\ s(r-1)\lambda_1^{-1} + n\lambda_2^{-1} = sq_1\lambda_1^{-2},$$

where

$$q_i = \mathbf{X}'\mathbf{H}\mathbf{A}_i\mathbf{H}\mathbf{X}\ (i = 0, 1),\ \lambda_0 = \theta_1,\ \lambda_1 = \theta_1 + s\theta_2,\ \lambda_2 = \theta_1 + n\theta_2, \tag{3.1.8}$$

and

$$\mathbf{H}_n = \mathbf{I}_n - n^{-1}\mathbf{E}_n,\ \mathbf{A}_0 = \mathbf{I}_r \otimes \mathbf{H}_s,\ \text{and}\ \mathbf{A}_1 = s^{-1}\mathbf{H}_r \otimes \mathbf{E}_s.$$

The ML equations can be given in closed form since here $\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}$ can be obtained explicitly.

Also, some progress can be made for a finite lattice with the doubly geometric scheme given by (2.5.1). To obtain numerical solutions in general, we give the standard solving method involving the information matrix, which we now derive.

3.2. The information matrix and asymptotic normality

On differentiating (3.1.4) with respect to θ_j, we find that

$$2\frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j} = -\text{tr}(\mathbf{R}_{ij} - \mathbf{S}_{ij}) - \mathbf{w}'(\mathbf{S}_{ij} + \mathbf{S}_{ji} - \mathbf{R}_{ij})\boldsymbol{\Sigma}^{-1}\mathbf{w}, \tag{3.2.1}$$

where

$$\mathbf{R}_{ij} = \boldsymbol{\Sigma}_{ij}\boldsymbol{\Sigma}^{-1},\ \mathbf{S}_{ij} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j, \boldsymbol{\Sigma}_{ij} = \frac{\partial^2 \boldsymbol{\Sigma}}{\partial\theta_i \partial\theta_j} = \frac{\partial \boldsymbol{\Sigma}_i}{\partial\theta_j}. \tag{3.2.2}$$

Since $E(\mathbf{w}'\mathbf{X}\mathbf{w}) = \text{tr}(\boldsymbol{\Sigma}\mathbf{X})$, after taking the expectation of (3.2.1) we obtain

$$E[-\frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j}] = (1/2)\text{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j) = a_{ij},\ \text{say}. \tag{3.2.3}$$

We also have, on differentiating (3.1.2) with respect to θ_j, $\frac{\partial^2 \ell}{\partial\beta\, \partial\theta_j} = -\mathbf{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\mathbf{w}$. Since $E(\mathbf{w}) = \mathbf{0}$, we find that

$$E[\frac{\partial^2 \ell}{\partial\beta\, \partial\theta_j}] = 0.$$