Conditional and Intrinsic Autoregressions

2 authors, including:

Håvard Rue
King Abdullah University of Science and Technology
278 PUBLICATIONS   **16,747** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Spatial modeling with R-INLA: A review View project

The R-INLA project View project

13

Conditional and Intrinsic Autoregressions

Leonhard Held and Havard Rue

CONTENTS

13.1 Introduction

The purpose of this chapter is to give an overview of conditional and intrinsic autoregressions. These models date back at least to [3], and have been heavily used since to model discrete spatial variation.

Traditionally, conditional autoregressions have been used to directly model spatial dependence in data that have been observed on a predefined graph or lattice structure. Inference is then typically based on likelihood or pseudo-likelihood techniques [3,17]. More recently, conditional autoregressions are applied in a modular fashion in (typically Bayesian) complex hierarchical models. Inference in this class is nearly always carried out using Markov chain Monte Carlo (MCMC), although some alternatives do exist [8,21].

In this chapter, we will describe the most commonly used conditional and intrinsic autoregressions. The focus will be on spatial models, but we will also discuss the relationship to autoregressive time-series models. Indeed, autoregressive time-series models are a special case of conditional autoregressions and exploring this relationship is helpful in order to develop intuition and understanding for the general class.

This chapter will not describe in detail how to build hierarchical models based on conditional autoregressive prior distributions and how to analyze them using MCMC. For a thorough discussion, see [1,12,20] as well as Chapter 14.

To begin, consider a random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ where each component is univariate. It is convenient to imagine that each component is located at a fixed site $i \in \{1, \ldots, n\}$. These sites may refer to a particular time point or a particular point in two- or higher-dimensional space, or particular areas in a geographical region, for example.

We now wish to specify a joint distribution with density $p(\boldsymbol{x})$ for $\boldsymbol{X}$. A decomposition of the form

$$p(\boldsymbol{x}) = p(x_1) \cdot p(x_2|x_1) \cdot p(x_3|x_1, x_2) \cdot \cdots \cdot p(x_n|x_1, x_2, \ldots, x_{n-1}) \tag{13.1}$$

is, of course, always possible. In a temporal context, this factorization is extremely useful, and—under an additional Markov assumption—further simplifies to

$$p(\boldsymbol{x}) = p(x_1) \cdot p(x_2|x_1) \cdot p(x_3|x_2) \cdot \cdots \cdot p(x_n|x_{n-1}).$$

Indeed, this factorization forms the basis of so-called first-order autoregressive models and can be conveniently generalized to higher orders. However, in a spatial context, where the indices $1, \ldots, n$ are arbitrary and could, in principle, easily be permuted, Equation (13.1) is not really helpful, as it is very difficult to envision most of the terms entering the above product.

It is much more natural to specify the full conditional distribution $p(x_i|\boldsymbol{x}_{-i})$, the conditional distribution of X_i at a particular site i, given the values $X_j = x_j$ at all other sites $j \neq i$. In a spatial context, the Markov assumption refers to the property that the conditional distribution $p(x_i|\boldsymbol{x}_{-i})$ depends only on a few components of $\boldsymbol{x}_{-i}$, called the neighbors of site i. However, it is not obvious at all under which conditions the set of full conditionals $p(x_i|\boldsymbol{x}_{-i})$, $i = 1, \ldots, n$, defines a valid joint distribution. Conditions under which such a joint distribution exists are discussed in [3] using the Brook expansion ([9]), see chapter 12 for details.

By far the most heavily studied model is the Gaussian conditional autoregression, where $p(x_i|\boldsymbol{x}_{-i})$ is univariate normal and $p(\boldsymbol{x})$ is multivariate normal. Gaussian conditional autoregressions with a Markov property are also known as Gaussian Markov random fields ([16,20]). Various Gaussian conditional autoregressions will be discussed in section 13.2. However, there are also nonnormal conditional autoregressions, for example, the so-called autologistic model for binary variables X_i, as discussed in section 13.3. In section 13.4, we turn to intrinsic Gaussian conditional autoregressions, a limiting (improper) form of Gaussian conditional autoregressions of practical relevance in hierarchical models. Finally, Section 13.5 gives a brief sketch of multivariate Gaussian conditional autoregressions.

13.2 Gaussian Conditional Autoregressions

Suppose that, for $i = 1, \ldots, n$, $X_i|\boldsymbol{x}_{-i}$ is normal with conditional mean and variance

$$\mathsf{E}(X_i|\boldsymbol{x}_{-i}) = \mu_i + \sum_{j \neq i} \beta_{ij}(x_j - \mu_j), \tag{13.2}$$

$$\mathrm{Var}(X_i|\boldsymbol{x}_{-i}) = \kappa_i^{-1}. \tag{13.3}$$

Here, μ_i will typically take a regression form, say, $w_i^T \alpha$ for covariates w_i associated with site i. Without loss of generality we assume that $\mu_1 = \cdots = \mu_n = 0$ in the following. Under the additional assumption that

$$\kappa_i \beta_{ij} = \kappa_j \beta_{ji}$$

for all $i \neq j$, these conditional distributions correspond to a multivariate joint Gaussian distribution with mean $\boldsymbol{0}$ and precision matrix $\boldsymbol{Q}$ with elements $Q_{ii} = \kappa_i$ and $Q_{ij} = -\kappa_i \beta_{ij}$, $i \neq j$, provided that $\boldsymbol{Q}$ is symmetric and positive definite.

Such a system of conditional distributions is known as an autonormal system [3]. Usually it is assumed that the precision matrix Q is regular; however, Gaussian conditional autoregressions with singular Q are also of interest and known as *intrinsic autoregressions*, as discussed in Section 13.4.

In many applications the coefficients β_{ij} will be nonzero for only a few so-called "neighbors" of X_i. Let ∂i denote the set of "neighbors" for each site i. We can then write Equation (13.2) (using $\mu_1 = \cdots = \mu_n = 0$) as

$$\mathsf{E}(X_i|\boldsymbol{x}_{-i}) = \sum_{j \in \partial i} \beta_{ij} x_j$$

to emphasize that the conditional mean of X_i only depends on the neighbors ∂i. The random vector $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ will then follow a Gaussian Markov random field, as discussed in chapter 3.1.

13.2.1 Example

Suppose that the X_is follow a zero-mean Gaussian conditional autoregression with

$$\mathsf{E}(X_i|\boldsymbol{x}_{-i}) = \phi \begin{cases} \frac{1}{2}(x_2 + x_n) & \text{for } i = 1 \\ \frac{1}{2}(x_{i-1} + x_{i+1}) & \text{for } 1 < i < n \\ \frac{1}{2}(x_1 + x_{n-1}) & \text{for } i = n \end{cases} \tag{13.4}$$

where $\phi \in [0, 1)$ and $\mathrm{Var}(X_i|\boldsymbol{x}_{-i}) = \kappa^{-1}$, say. At first sight, this looks like a first-order autoregressive time-series model, but by linking the first "time point" x_1 with the last "time point" x_n, the model is defined on a circle. The model is called a circular first-order autoregressive model and is useful for analyzing circular data.

The precision matrix of $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ is

$$Q = \frac{\kappa}{2} \begin{pmatrix} 2 & -\phi & & & & & -\phi \\ -\phi & 2 & -\phi & & & & \\ & -\phi & 2 & -\phi & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -\phi & 2 & -\phi & \\ & & & & -\phi & 2 & -\phi \\ -\phi & & & & & -\phi & 2 \end{pmatrix} \tag{13.5}$$

with all other elements equal to zero. Thus, the precision matrix Q is a *circulant* matrix with base $\boldsymbol{d} = \kappa \cdot (1, -\phi/2, 0, \ldots, 0, -\phi/2)^T$ (the first row of Q) (see [20, sec. 2.6.1] for an introduction to circular matrices). The covariance matrix $\boldsymbol{\Sigma} = Q^{-1}$ of $\boldsymbol{x}$ is again circular. Its base $\boldsymbol{e}$, which equals the autocovariance function of $\boldsymbol{X}$, can be calculated using the discrete Fourier transform DFT($\boldsymbol{d}$) of $\boldsymbol{d}$,

$$\boldsymbol{e} = \frac{1}{n}\mathrm{IDFT}(\mathrm{DFT}(\boldsymbol{d})^{-1}),$$

here IDFT denotes the inverse discrete Fourier transform and the power function is to be understood elementwise. See [20] for a derivation.

The following R-code illustrates, how $\boldsymbol{e}$ is computed for the circulant precision matrix (13.5) with $n = 10$, $\phi = 0.9$, and $\kappa = 1$. Note that the (inverse) discrete Fourier transform is computed with the function `fft()` and that the imaginary parts of the function values are equal to zero.

```
> # function make.d computes the base d
> make.d <- function(n, phi){
+    d <- rep(0.0, n)
+    d[1] <- 1
+    d[2] <- -phi/2
+    d[n] <- -phi/2
+    return(d)
+ }
> # function e computes the base e, i.e. the autocovariance function
> # if corr=T you obtain the autocorrelation function
> e <- function(n, phi, corr=F){
+    d <- make.d(n, phi)
+    e <- Re(fft(1/Re(fft(d)), inverse=TRUE))/n
+    if(corr==F)
+       return(e)
+    else return(e/e[1])
+ }
> n <- 10
> phi <- 0.9
> result <- e(n, phi)
> print(result)

 [1] 2.3375035 1.4861150 0.9649742 0.6582722 0.4978530 0.4480677 0.4978530
 [8] 0.6582722 0.9649742 1.4861150
```

From the autocovariances e we can easily read off the autocorrelations of X. The left panel in Figure 13.1 displays the autocorrelation function for $n = 100$ and $\phi = 0.9, 0.99, 0.999,$ 0.9999. Of course, the autocorrelation function must be symmetric, the correlation between x_1 and x_3 must be the same as the correlation between x_1 and x_{99}, for example. For the two smaller values of ϕ, the autocorrelation is essentially zero for lags around $n/2 = 50$. For the larger values of ϕ very close to unity, there is substantial autocorrelation between any two components of x.
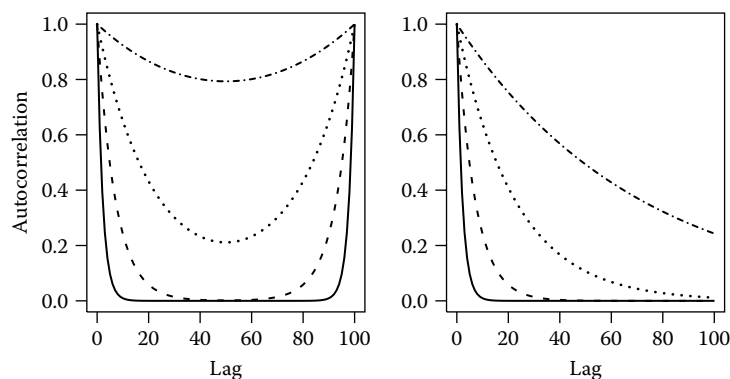
FIGURE 13.1
Autocorrelation function of the circular (left) and ordinary (right) first-order autoregressive model (13.4) and (13.6), respectively, for $n = 100$ and $\phi = 0.9$ (solid line), $\phi = 0.99$ (dashed line), $\phi = 0.999$ (dotted line), and $\phi = 0.9999$ (dot-dashed line). The corresponding coefficients of the ordinary first-order autoregressive model are $\alpha = 0.63, \alpha = 0.87, \alpha = 0.96,$ and $\alpha = 0.99$; compare equation (13.7).

It is interesting to compare the autocorrelations obtained with those from the ordinary first-order autoregressive process defined through the directed definition

$$X_i | x_{i-1} \sim N(\alpha x_{i-1}, \kappa^{-1}), \tag{13.6}$$

where $|\alpha| < 1$ to ensure stationarity. This model has identical neighborhood structure as the circular first-order autoregressive model, except for the missing link between X_1 and X_n. The autocorrelation function is $\rho_k = \alpha^k$ for lag k.

It is easy to show that this directed definition induces the full conditional distribution

$$X_i | \boldsymbol{x}_{-i} \sim \begin{cases} N\left(\alpha x_2, \kappa^{-1}\right) & i = 1 \\ N\left(\dfrac{\alpha}{1+\alpha^2}(x_{i-1}+x_{i+1}), \left(\kappa(1+\alpha^2)\right)^{-1}\right) & i = 2, \ldots, n-1 \\ N\left(\alpha x_{n-1}, \kappa^{-1}\right) & i = n. \end{cases}$$

If we want to compare the circular autoregressive model Equation (13.4) with the ordinary autoregressive model (13.6), we need to equate the autoregressive coefficients of the full conditional distributions. From $\phi/2 = \alpha/(1+\alpha^2)$ it follows that for a given autoregressive coefficient ϕ of the circular autoregressive model, the corresponding coefficient $\alpha = \alpha(\phi)$ of the ordinary first-order autoregressive process is

$$\alpha(\phi) = \frac{1 - \sqrt{1-\phi^2}}{\phi}. \tag{13.7}$$

For example, $\phi = 0.99$ corresponds to $\alpha \approx 0.87$, $\phi = 0.999$ corresponds to $\alpha \approx 0.96$. This illustrates that coefficients from undirected Gaussian conditional autoregressions have a quite different meaning compared to coefficients from directed Gaussian autoregressions.

Figure 13.1 compares the autocorrelation function of the circular autoregressive model with coefficient ϕ with the corresponding autocorrelation function of the ordinary autoregressive model with coefficient $\alpha(\phi)$. A close correspondence of autocorrelations up to lag 50 can be seen for $\phi = 0.9$ and $\phi = 0.99$. The autocorrelations up to lag $n/2$ of the circular model differ from the corresponding ones from the ordinary model not more than $4.5e - 11$ and 0.00072, respectively. For $\phi = 0.999$ and $\phi = 0.9999$, the decay of the autocorrelations with increasing lag is not as pronounced as the geometric decay of the ordinary autoregressive model. This is due to the increasing impact of the link between x_n and x_1 in the circular model.

13.2.2 Gaussian Conditional Autoregressions on Regular Arrays

Suppose now that a conditional autoregressive model is defined on a lattice with $n = n_1 n_2$ nodes and let (i, j) denote the node in the ith row and jth column. In the interior of the lattice, we can now define the nearest four sites of (i, j) as its neighbors, i.e., the nodes

$$(i-1, j), (i+1, j), (i, j-1), (i, j+1).$$

A proper conditional Gaussian model with this neighborhood structure, often called *first-order autoregression*, is based on the conditional mean

$$E(X_{ij} | \boldsymbol{x}_{-ij}) = \alpha(x_{i-1,j} + x_{i+1,j}) + \beta(x_{i,j-1} + x_{i,j+1}) \tag{13.8}$$

with $|\alpha| + |\beta| < 0.5$ and $Var(X_{ij} | \boldsymbol{x}_{-ij}) = \kappa^{-1}$, say. In most practical applications, both α and β will be positive. Assuming that the lattice is wrapped on a *torus*, so that every pixel has four neighbors, this process is stationary. A *torus* is a regular lattice with toroidal boundary
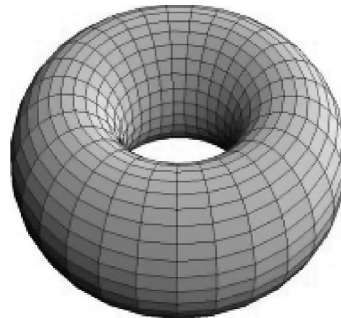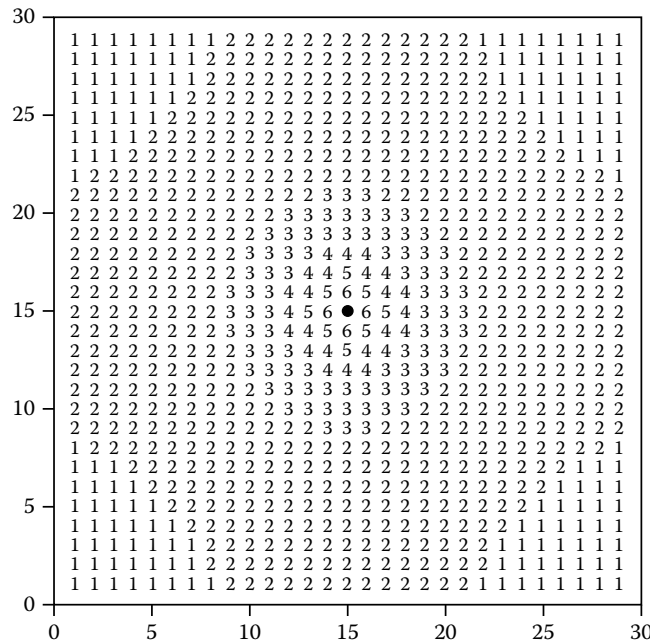
FIGURE 13.2
Illustration of a torus obtained on a two-dimensional lattice with $n_1 = n_2 = 29$ and toroidal boundary conditions.

conditions, which can be obtained in two steps. First, the lattice is wrapped to "sausage." In a second step, the two ends of the sausage are joined such that the sausage becomes a ring. This two-stage process ensures that every pixel of the lattice has four neighbors. For example, pixel $(1, 1)$ will have the four neighbors $(1, 2)$, $(2, 1)$, $(1, n_2)$ and $(n_1, 1)$. For further illustration of toroidal boundary conditions, see Figure 13.2 and the R-code in the following example. Note that an alternative way to study conditional autoregressions is on an *infinite* regular array, in which case the process will be stationary and the *spectral density* is useful. (For details, see [7,17].)

13.2.3 Example

Suppose we set $\alpha = \beta = 0.2496$ in model (13.8), defined on a torus of size $n_1 = n_2 = 29$. The following R-code illustrates the computation of the autocovariance matrix of **X** by simply inverting the precision matrix of **X** using the function solve(). An alternative way would be to exploit the fact that the precision matrix of **X** is *block-circulant*. The two-dimensional Fourier transform can then be used to calculate the base of the autocovariance matrix (see [20, section 2.6.2] for details).

```
> # make.prec computes the precision matrix of a toroidal first-order
> # autoregression on a two-dimensional lattice of size n1 x n2
> # with coefficient coeff
> make.prec <- function(n1, n2, coeff){
+   prec <- diag(n1*n2)
+   for(i in 1:(n1*n2)){
+     j <- ((i-1)%%n1)+1 # column index
+     k <- (n1*(n2-1))   # if i>k we are in the last row
+
+     if(j!=1) (prec[i,i-1] <- -coeff)    # left neighbor
+     else (prec[i,i+(n1-1)] <- -coeff)   # left toroidal neighbor
+
+     if(j!=n1)  (prec[i,i+1] <- -coeff) # right neighbor
+     else (prec[i,i-(n1-1)] <- -coeff)  # right toroidal neighbor
+
+     if(i>n1) (prec[i,i-n1] <- -coeff)  # top neighbor
+     else (prec[i,(j+k)] <- -coeff)     # top toroidal neighbor
+
+     if(i<=k) (prec[i,i+n1] <- -coeff)  # bottom neighbor
```

FIGURE 13.3
Plot of the correlation of a pixel x_{ij} with the pixel $x_{15,15}$ in model (13.8), defined on a torus of size $n_1 = n_2 = 29$ with coefficients $\alpha = \beta = 0.2496$. Shown is 10 times the autocorrelation, truncated to an integer.

```
+      else (prec[i,j] <- -coeff)          # bottom toroidal neighbor
+    }
+    return(prec)
+ }
> prec <- make.prec(n1=29, n2=29, coeff=0.2496)
> # inversion gives the covariance matrix
> cova <- solve(prec)
```

From the autocovariance matrix, we can easily calculate autocorrelations between any pair of sites. Figure 13.3 displays the correlation of pixel x_{ij}, $1 \le i, j \le 29$, with pixel $x_{15,15}$ in the center of the plot. Although the coefficients α and β are close to the border of the parameter space, the correlation between adjacent pixels is only 0.669. The smallest correlation observed, for example, between $x_{1,1}$ and $x_{15,15}$ is 0.186.

13.3 Non-Gaussian Conditional Autoregressions

For binary or count data, direct usage of Gaussian conditional autoregressions is often not possible. Instead, conditional autoregressive models in the form of a logistic or log-linear Poisson model have been proposed. Here, we discuss the autologistic and the auto-Poisson model, which basically adopt the form (13.2) for the conditional mean of $X_i | x_{-i}$ using a link function, as known from generalized linear modeling ([19]). However, consistency requirements imply that for binary data only the logistic link, and for Poisson counts only the log link can be used (see [2] and [3] for details). Only the autologistic model has gained

some popularity in applications, the auto-Poisson has undesirable properties, which make it not suitable for most applications in spatial statistics.

13.3.1 Autologistic Models

Assume $X_i, i = 1, \ldots, n$, are binary random variables with conditional success probability $\pi_i(\boldsymbol{x}_{-i}) = \mathsf{E}(X_i | \boldsymbol{x}_{-i})$. The autologistic model specifies the (logit-transformed) conditional mean

$$\mathrm{logit}\, \pi_i(\boldsymbol{x}_{-i}) = \mu_i + \sum_{j \in \partial i} \beta_{ij} x_j,$$

where $\beta_{ij} = \beta_{ji}$, for consistency reasons. The normalizing constant of the joint distribution, which depends on the β_{ij}s, is very difficult to compute, thus a traditional likelihood approach to estimate the coefficients is typically infeasible. Instead, a pseudo-likelihood approach has been proposed by [4], in which the product of the conditional binomial probabilities is maximized. The model can be generalized to a binomial setting with additional "sample sizes" N_i, say. Also, the model can be extended to include covariates (see [14], for example).

13.3.2 Auto-Poisson Models

Suppose $X_i, i = 1, \ldots, n$, are Poisson random variables with conditional mean $\lambda_i(\boldsymbol{x}_{-i}) = \mathsf{E}(X_i | \boldsymbol{x}_{-i})$. Similar to the autologistic model, the auto-Poisson model specifies the (log-transformed) conditional mean

$$\log \lambda_i(\boldsymbol{x}_{-i}) = \mu_i + \sum_{j \in \partial i} \beta_{ij} x_j.$$

It turns out that a necessary (and sufficient) condition for the existence of a joint distribution with the specified conditional distributions is that $\beta_{ij} \leq 0$ for all $i \neq j$. However, a negative coefficient β_{ij} implies negative interaction between i and j because the conditional mean of X_i decreases with an increase in x_j. This is quite opposite to the intent of most spatial modeling; however, there are applications in purely inhibitory Markov point processes (see [5]).

13.4 Intrinsic Autoregressions

Intrinsic Gaussian autoregressions arise if the precision matrix $\boldsymbol{Q}$ of the Gaussian conditional autoregression (13.2) and (13.3) is only positive semidefinite with rank($\boldsymbol{Q}$) $< n$. For example, if $\beta_{ij} = w_{ij}/w_{i+}$ and $\kappa_i = \kappa w_{i+}$ where $\kappa > 0$ is a precision parameter, $w_{ij} \geq 0$ are predefined weights and $w_{i+} = \sum_{j \neq i} w_{ij}$, $\boldsymbol{Q}$ will be rank deficient. Such weights are quite common in spatial models for areal data. For example, adjacency-based weights are $w_{ij} = 1$ if regions i and j are adjacent (usually denoted by $i \sim j$) and zero otherwise. Other choices are weights based on the inverse distance between area centroids or the length of the common boundary, for example.

For adjacency-based weights, the conditional mean and variance simplify to

$$\mathsf{E}(X_i | \boldsymbol{x}_{-i}) = \sum_{j \in \partial i} x_j / m_i$$

$$\mathrm{Var}(X_i | \boldsymbol{x}_{-i}) = (\kappa \cdot m_i)^{-1},$$

here m_i denotes the number of neighbors of region i, i.e., the cardinality of the set ∂i.

The resulting joint distribution is improper, its density can be written (up to a proportionality constant) as

$$p(x|\kappa) \propto \exp\left(-\frac{\kappa}{2}\sum_{i \sim j}(x_i - x_j)^2\right), \tag{13.9}$$

where the sum goes over all pairs of adjacent regions $i \sim j$. This is a special case of a *pairwise difference prior*, as described in [6]. With $x = (x_1, \ldots, x_n)^T$, the density (13.9) can be written in the form

$$p(x|\kappa) \propto \exp\left(-\frac{\kappa}{2}x^T R x\right), \tag{13.10}$$

where the *structure matrix* R has elements

$$R_{ij} = \begin{cases} m_i & \text{if } i = j, \\ -1 & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}.$$

We immediately see that the precision matrix $Q = \kappa R$ cannot be of full rank because all rows and columns of R sum up to zero.

In the special case where the index $i = 1, \ldots, n$ represents time and each time-point has the two (respectively one) nearest time-points as its neighbors, Equation (13.9) simplifies to

$$p(x|\kappa) \propto \exp\left(-\frac{\kappa}{2}\sum_{i=2}^{n}(x_i - x_{i-1})^2\right).$$

This is a so-called *first-order random walk* model, as it corresponds to the directed formulation

$$X_i|x_{i-1} \sim N(x_{i-1}, \kappa^{-1}),$$

with improper uniform prior on x_1. Obviously this is a limiting case of model (13.6) with $\alpha = 1$. The structure matrix of this model has a particularly simple form,

$$R = \begin{pmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{pmatrix}, \tag{13.11}$$

and forms the basis of some spatial models on regular arrays, as we will see later.

Intrinsic autoregressions are more difficult to study than ordinary (proper) conditional autoregressions. The rank deficiency of the precision matrix does not allow the computation of autocorrelation functions, for example. Similarly, it is not possible to sample from an intrinsic autoregression without imposing additional constraints, so they cannot be models for data. On infinite regular arrays, intrinsic autoregressions can be studied using the *generalized spectral density* (see [7,17] for details).

13.4.1 Normalizing Intrinsic Autoregressions

An interesting question that arises is the appropriate "normalizing constant" of intrinsic Gaussian autoregressions. The constant will depend on unknown parameters in the precision matrix Q and is necessary if those need to be estimated from the data. Of course,

intrinsic Gaussian autoregressions are improper, so there is no constant to normalize the density

$$p(x|\kappa) \propto \exp\left(-\frac{1}{2}x^T Q x\right) \tag{13.12}$$

if Q is not positive definite. The term "normalizing constant" has to be understood in a more general sense as the normalizing constant of an equivalent lower-dimensional *proper* Gaussian distribution.

It is now commonly accepted ([13,15,20]) that for the general model Equation (13.12) with $n \times n$ precision matrix Q of rank $n - k$, the correct "normalizing constant" is

$$(2\pi)^{-(n-k)/2}(|Q|^*)^{1/2},$$

where $|Q|^*$ denotes the *generalized determinant* of Q, the product of the $n - k$ nonzero eigenvalues of Q.

In the special case $Q = \kappa R$ of model (13.10) with known structure matrix R, the "normalizing constant" simplifies to

$$\left(\frac{\kappa}{2\pi}\right)^{\frac{n-k}{2}} \tag{13.13}$$

due to the rank deficiency of R with rank $n - k$. If the neighborhood structure is nonseparable, i.e., every pixel is connected to every other by some chain of neighbors, then $k = 1$.

13.4.2 Example

Suppose data y_i, $i = 1, \ldots, n$, are observed and we assume that

$$y_i|x_i, \sigma^2 \sim N(x_i, \sigma^2) \tag{13.14}$$

are conditionally independent with known variance σ^2. Assume further that, conditional on κ, the unknown mean surface $x = (x_1, \ldots, x_n)^T$ follows a pairwise difference prior (Equation 13.9) with a nonseparable neighborhood structure. The goal is to infer x from y in order to denoise the observed "image" y and to obtain a smoother version. A fully Bayesian analysis would place a hyperprior on κ, usually a conjugate gamma prior $\kappa \sim G(\alpha, \beta)$, i.e.,

$$f(\kappa) \propto \kappa^{\alpha-1}\exp(-\beta\kappa).$$

To implement a two-stage Gibbs sampler (see for example [11]), one would sample from $x|\kappa, y$ and from $\kappa|x, y = \kappa|x$. Note that R is of rank $n - 1$ since the graph is assumed to be nonseparable, so based on (13.9) and (13.13), it follows that

$$\kappa|x \sim G\left(\alpha + \frac{n-1}{2}, \beta + \frac{1}{2}\sum_{i\sim j}(x_i - x_j)^2\right).$$

The other full conditional distribution is

$$x|\kappa, y \sim N(Aa, A),$$

where $A = (\kappa R + \sigma^2 I)^{-1}$ and $a = \sigma^2 y$.

Note that there is no need to include an intercept in (13.14), as the intrinsic autoregression x has an undefined overall level. An equivalent formulation is to include an additional intercept with a flat prior and to use an additional sum-to-zero constraint on x. Note also that omission of the data error, i.e., setting $\sigma^2 = 0$, is not useful, as x_i will then equal y_i and no smoothing will be done.

13.4.3 Intrinsic Autoregressions on Regular Arrays

We now return to conditional autoregressions defined on regular arrays. When fitting model (13.8) to data, the estimated coefficients are often close to singularity (i.e., $\alpha + \beta$ will be close to 0.5) in order to obtain nonnegligible spatial autocorrelations. A limiting case of model (13.8) is obtained if $\alpha + \beta = 0.5$. For example, if $\alpha = \beta = 0.25$, the conditional mean of x_{ij} is

$$\mathsf{E}(x_{ij}|x_{-ij}) = \frac{1}{4}(x_{i-1,j} + x_{i+1,j} + x_{i,j-1} + x_{i,j+1}).$$

This is an intrinsic autoregression and a special case of the pairwise difference prior (Equation 13.9) with conditional variance equal to $1/(4\kappa)$.

However, on regular arrays it is possible to define an anisotropic intrinsic model, which is able to weight horizontal and vertical neighbors differently. The conditional mean in this extended model is still given by Equation (13.8), but the coefficients $\alpha > 0$ and $\beta > 0$ are now allowed to vary subject to $\alpha + \beta = 0.5$. The conditional variance is still equal to $1/(4\kappa)$. This specification defines a valid intrinsic autoregression. In applications, α (or β) can be treated as an unknown parameter, so the degree of anisotropy can be estimated from the data.

To estimate α it is necessary to compute the generalized determinant of the associated precision matrix Q, which can be written as a sum of two *Kronecker products*:

$$Q = \alpha R_{n_1} \otimes I_{n_2} + \beta I_{n_1} \otimes R_{n_2}.$$

Here R_n is the structure matrix (13.11) of an n-dimensional random-walk model and I_n is the $n \times n$-identity matrix. An explicit form for the generalized determinant can be found in [20], page 107.

13.4.4 Higher-Order Intrinsic Autoregressions

All intrinsic autoregressions up to now are of order one, in the sense that the precision matrix Q has a rank deficiency of 1. This is due to an undefined overall level of the distribution of x. An equivalent representation is obtained if x is replaced by $\mu + x$, where x has a density as described above, but under an additional sum-to-zero constraint, and the scalar μ has an improper locally uniform prior. In more complex hierarchical models with more than one intrinsic autoregression, such sum-to-zero constraints are necessary to ensure a proper posterior. Computational routines for sampling from GMRFs under linear constraints are particularly useful in this context for MCMC simulation (see chapter 12 for details).

Intrinsic autoregressions of higher order may also be considered. On regular lattices, such autoregressions can be defined using the closest eight or twelve nearest neighbors, for example. However, appropriate weights have to be chosen with care. It is useful to start with an (improper) joint Gaussian distribution based on squared increments, similar to the squared difference prior (Equation 13.9), and to derive the full conditional from the joint distribution. For example, one might consider the increments

$$\genfrac{}{}{0pt}{}{\circ\,\bullet}{\bullet\,\circ} - \genfrac{}{}{0pt}{}{\bullet\,\circ}{\circ\,\bullet}, \tag{13.15}$$

where the •s enter the difference, but not the ∘s, which only serve to fix the spatial location. Summing over all pixels with well-defined increments, Equation (13.15) thus leads to the joint improper distribution

$$p(x|\kappa) \propto \exp\left(-\frac{\kappa}{2}\sum_{i=1}^{n_1-1}\sum_{j=1}^{n_2-1}(x_{i+1,j+1} - x_{i+1,j} - x_{i,j+1} + x_{i,j})^2\right). \tag{13.16}$$

This is a special case of model (13.10) with structure matrix R defined as the *Kronecker product* of two structure matrices R_1 and R_2 of random-walk type (13.11) with dimension n_1 and n_2, respectively: $R = \kappa \cdot (R_1 \otimes R_2)$. The rank of R is $(n_1 - 1)(n_2 - 1)$, so R has a deficiency in rank of order $n_1 + n_2 - 1$.

The conditional mean of x_{ij} in the interior of the lattice ($2 \leq i \leq n_1 - 1$, $2 \leq j \leq n_2 - 1$) now depends on its eight nearest sites and is

$$\mathsf{E}(x_{ij}|x_{-ij}) = \frac{1}{2}(x_{i-1,j} + x_{i+1,j} + x_{i,j-1} + x_{i,j+1}) \tag{13.17}$$
$$- \frac{1}{4}(x_{i-1,j-1} + x_{i-1,j+1} + x_{i+1,j-1} + x_{i+1,j+1}),$$

while the conditional precision is 4κ. In a more compact notation, the conditional mean is

$$\mathsf{E}(x_{ij}|x_{-ij}) = \frac{1}{2} \begin{smallmatrix} \cdot & \circ & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \circ & \cdot \end{smallmatrix} - \frac{1}{4} \begin{smallmatrix} \circ & \cdot & \circ \\ \cdot & \cdot & \cdot \\ \circ & \cdot & \circ \end{smallmatrix}.$$

Anisotropic versions of this intrinsic autoregression with eight neighbors are discussed in [17].

For illustration, we now describe how to derive the conditional mean (13.17) from (13.16). Clearly, $p(x_{ij}|x_{-ij}, \kappa) \propto p(x|\kappa)$, so in the interior of the lattice four terms in the double sum in Equation (13.16) depend on x_{ij}, hence,

$$p(x_{ij}|x_{-ij}, \kappa) \propto \exp\left(-\frac{\kappa}{2}\left((x_{i+1,j+1} - x_{i+1,j} - x_{i,j+1} + x_{i,j})^2\right.\right.$$
$$+ (x_{i+1,j} - x_{i+1,j-1} - x_{i,j} + x_{i,j-1})^2$$
$$+ (x_{i,j+1} - x_{i,j} - x_{i-1,j+1} + x_{i-1,j})^2$$
$$\left.\left.+ (x_{i,j} - x_{i,j-1} - x_{i-1,j} + x_{i-1,j-1})^2\right)\right),$$

which can be rearranged to

$$p(x_{ij}|x_{-ij}, \kappa) \propto \exp\left(-\frac{\kappa}{2}\left((x_{i,j} - (x_{i+1,j} + x_{i,j+1} - x_{i+1,j+1}))^2\right.\right.$$
$$+ (x_{i,j} - (x_{i+1,j} + x_{i,j-1} - x_{i+1,j-1}))^2$$
$$+ (x_{i,j} - (x_{i-1,j} + x_{i,j+1} - x_{i-1,j+1}))^2$$
$$\left.\left.+ (x_{i,j} - (x_{i,j-1} + x_{i-1,j} - x_{i-1,j-1}))^2\right)\right).$$

A useful identity for combining quadratic forms* eventually gives

$$p(x_{ij}|x_{-ij}, \kappa) \propto \exp\left(-\frac{4\kappa}{2}\left(x_{i,j} - \left(\frac{1}{2}(x_{i-1,j} + x_{i+1,j} + x_{i,j-1} + x_{i,j+1})\right.\right.\right.$$
$$\left.\left.\left.- \frac{1}{4}(x_{i-1,j-1} + x_{i-1,j+1} + x_{i+1,j-1} + x_{i+1,j+1})\right)\right)^2\right), \tag{13.18}$$

from which the conditional mean (13.17) and the conditional 4κ precision can be read off.

---

* $A(x - a)^2 + B(x - b)^2 = C(x - c)^2 + \frac{AB}{C}(a - b)^2$ where $C = A + B$ and $c = (Aa + Bb)/C$.

It is easy to see that the distribution (13.16) is invariant to the addition of arbitrary constants to any rows or columns. This feature makes this distribution unsuitable as a prior for a smoothly varying surface, a defect that can be remedied by expanding the system of neighbors. Indeed, consider now the joint distribution

$$p(x|\kappa) \propto \exp\left(-\frac{\kappa}{2} \sum_{i=2}^{n_1-1} \sum_{j=2}^{n_2-1} (x_{i-1,j} + x_{i+1,j} + x_{i,j-1} + x_{i,j+1} - 4x_{i,j})^2\right), \qquad (13.19)$$

which is based on the increments

$$\vcenter{\hbox{$\cdot\;\bullet\;\cdot$}} - 4 \vcenter{\hbox{$\cdot\;\cdot\;\cdot$}}.$$

The conditional mean

$$\begin{aligned}
\mathsf{E}(x_{ij}|x_{-ij}) = {}& \frac{8}{20}(x_{i-1,j} + x_{i+1,j} + x_{i,j-1} + x_{i,j+1}) \\
& - \frac{1}{10}(x_{i-1,j-1} + x_{i-1,j+1} + x_{i+1,j-1} + x_{i+1,j+1}) \\
& - \frac{1}{20}(x_{i-2,j} + x_{i+2,j} + x_{i,j-2} + x_{i,j+2})
\end{aligned}$$

can be derived for pixels in the interior of the lattice ($3 \le i \le n_1 - 2, 3 \le j \le n_2 - 2$). In our compact notation, the conditional mean is, hence,

$$\mathsf{E}(x_{ij} \mid x_{-ij}) = \frac{1}{20}\left(8\,\vcenter{\hbox{$\cdots$}} - 2\,\vcenter{\hbox{$\cdots$}} - 1\,\vcenter{\hbox{$\cdots$}}\right).$$

The conditional variance is $1/(20\kappa)$, while appropriate modifications for both mean and variance are necessary on the boundary of the lattice (see [20] for a detailed discussion). Anisotropic versions have also been considered ([17]).

This conditional autoregression is based on the 12 nearest neighbors of each pixel. The distribution (13.19) is invariant to the linear transformation

$$x_{ij} \rightarrow x_{ij} + p_{ij},$$

where

$$p_{ij} = \gamma_0 + \gamma_1 i + \gamma_2 j$$

for arbitrary coefficients γ_0, γ_1, and γ_2. This is a useful property, as the prior is often used in applications for smoothing deviations from a two-dimensional linear trend p_{ij}.

This model has some drawbacks, however. First, the four corners—$x_{1,1}, x_{1,n_2}, x_{n_1,1}, x_{n_1,n_2}$—do not appear in Equation (13.19). Secondly, viewed as a difference approximation to a differential operator, model (13.19) induces a so-called anisotropic discretization error, i.e., the approximation error is larger along the diagonals than in the horizontal or vertical direction (for details on this issue, see page 117 in [20]).

A more elaborate model is given by

$$\begin{aligned}
p(x|\kappa) \propto \exp\Bigg(-\frac{\kappa}{2} \sum_{i=2}^{n_1-1} \sum_{j=2}^{n_2-1} \bigg(&\frac{2}{3}(x_{i-1,j} + x_{i+1,j} + x_{i,j-1} + x_{i,j+1}) \\
&+ \frac{1}{6}(x_{i-1,j-1} + x_{i-1,j+1} + x_{i+1,j-1} + x_{i+1,j+1}) - \frac{10}{3}x_{i,j}\bigg)^2\Bigg), \quad (13.20)
\end{aligned}$$

based on the increments

$$\frac{2}{3}\;\substack{\cdot\;\cdot\;\cdot\\\cdot\;\cdot\;\cdot\\\cdot\;\cdot\;\cdot} + \frac{1}{6}\;\substack{\cdot\;\cdot\;\cdot\\\cdot\;\cdot\;\cdot\\\cdot\;\cdot\;\cdot} - \frac{10}{3}\;\substack{\circ\;\circ\;\circ\\\circ\;\cdot\;\circ\\\circ\;\circ\;\circ}.$$

Note that the four corners—$x_{1,1}$, x_{1,n_2}, $x_{n_1,1}$, x_{n_1,n_2}—now enter the joint distribution. The full conditional of x_{ij} depends on 24 neighbors, its conditional expectation is

$$\mathsf{E}(x_{ij}\mid \boldsymbol{x}_{-ij}) = \frac{1}{468}\left(144\;[\text{stencil}] - 18\;[\text{stencil}]\right.$$
$$\left. +8\;[\text{stencil}] - 8\;[\text{stencil}] - 1\;[\text{stencil}]\right)$$

the conditional variance is $1/(13\kappa)$ (see [20] for further details).

13.5 Multivariate Gaussian Conditional Autoregressions

Multivariate Gaussian conditional autoregressions are a straightforward generalization of Equation (13.2) and Equation (13.3). Suppose X_i, $i = 1, \ldots, n$ is a p-dimensional random vector and let the conditional distribution of X_i given $\boldsymbol{x}_{-i}$ be multivariate Gaussian with conditional mean and covariance matrix

$$\mathsf{E}(X_i|\boldsymbol{x}_{-i}) = \boldsymbol{\mu}_i + \sum_{j\neq i} \boldsymbol{B}_{ij}(\boldsymbol{x}_j - \boldsymbol{\mu}_j) \tag{13.21}$$

$$\mathrm{Cov}(X_i|\boldsymbol{x}_{-i}) = \boldsymbol{\Phi}_i^{-1}. \tag{13.22}$$

The matrices $\boldsymbol{B}_{ij}$ and $\boldsymbol{\Phi}_i > 0$ are all of dimension $p \times p$. Without loss of generality, we assume in the following that $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_n = \boldsymbol{0}$. As in the univariate case, the joint distribution of $X = (X_1, \ldots, X_n)$ is multivariate normal with mean $\boldsymbol{0}$ and precision matrix $\boldsymbol{Q} = \boldsymbol{D}(\boldsymbol{I} - \boldsymbol{B})$, provided that $\boldsymbol{Q}$ is regular and symmetric ([18]). Here, $\boldsymbol{D}$ is block-diagonal with entries $\boldsymbol{\Phi}_i$, $i = 1, \ldots, n$, $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{B}$ is $np \times np$ with block-elements $\boldsymbol{B}_{ij}$ for $i \neq j$ and block-diagonal entries equal to zero. More details on this model can be found in [1], sec. 7.4.2.

In practice, we often encounter the situation that we have multivariate observations in each pixel with a fixed neighborhood structure between the pixels. A straightforward generalization of the adjacency-based intrinsic pairwise-difference prior (Equation 13.9) is

$$p(\boldsymbol{x}|\boldsymbol{\Phi}) \propto \exp\left(-\frac{1}{2}\sum_{i\sim j}(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{\Phi}(\boldsymbol{x}_i - \boldsymbol{x}_j)\right) \tag{13.23}$$

with conditional mean and covariance matrix equal to

$$\mathsf{E}(X_i|\boldsymbol{x}_{-i}) = \sum_{j\sim i} \boldsymbol{x}_j/m_i$$
$$\mathrm{Cov}(X_i|\boldsymbol{x}_{-i}) = (m_i \cdot \boldsymbol{\Phi})^{-1}.$$

Multivariate conditional autoregressive models are discussed in more detail in [10] (see also sec. 7.4 in [1]).

Acknowledgments

References

1. S. Banerjee, B.P. Carlin, and A.E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*, vol. 101 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 2004.
2. J. Besag. Nearest-neighbor systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society, Series B*, 34(1):75–83, 1972.
3. J. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B*, 36(2):192–225, 1974.
4. J. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24(3):179–195, 1975.
5. J. Besag. Some methods of statistical analysis for spatial data. *Bullotion of International Statistical Institute*, 47:77–92, 1976.
6. J. Besag, P.J. Green, D. Higdon, and K. Mengersen. Bayesian computation and stochastic systems (with discussion). *Statistical Science*, 10(1):3–66, 1995.
7. J. Besag and C. Kooperberg. On conditional and intrinsic autoregressions. *Biometrika*, 82(4):733–746, 1995.
8. N.E. Breslow and D.G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(1):9–25, 1993.
9. D. Brook. On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems. *Biometrika*, 51(3 and 4):481–483, 1964.
10. A. Gelfand and P. Vounatsou. Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, 4:11–25, 2003.
11. A.E. Gelfand and A.F.M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–509, 1990.
12. D. Higdon. A primer on space-time modeling from a Bayesian perspective. In B. Finkenstädt, L. Held, and V. Isham, Eds., *Statistical Methods for Spatio-Temporal Systems*, p. 217–279. Chapman & Hall, London, 2007.
13. J.S. Hodges, B.P. Carlin, and Q. Fan. On the precision of the conditionally autoregressive prior in spatial models. *Biometrics*, 59:317–322, 2003.
14. F.W. Huffer and H. Wu. Markov chain Monte Carlo for autologistic regression models with application to the distribution of plant species. *Biometrics*, 54(1):509–524, 1998.
15. L. Knorr-Held. Some remarks on Gaussian Markov random field models for disease mapping. In P.J. Green, N.L. Hjort, and S. Richardson, Eds., *Highly Structured Stochastic Systems*, Oxford Statistical Science Series, no 27, pages 260–264. Oxford University Press, Oxford, U.K. 2003.
16. H.R. Künsch. Gaussian Markov random fields. *Journal of the Faculty of Science, The University of Tokyo*, 26(1):53–73, 1979.
17. H.R. Künsch. Intrinsic autoregressions and related models on the two-dimensional lattice. *Biometrika*, 74(3):517–524, 1987.
18. K.V. Mardia. Multidimensional multivariate Gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis*, 24(2):265–284, 1988.

Handbook of Spatial Statistics

19. P. McCullagh and J.A. Nelder. *Generalized Linear Models*. CRC Press, Boca Raton, FL, 2nd ed., 1990.

20. H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*, vol. 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 2005.

21. H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). Forthcoming. *Journal of the Royal Statistical Society, Series B*, 2009.

Update ref. 21.