Gaussian Markov Random Fields

Gaussian Markov Random Fields (GMRF) are useful Gaussian processes for data on regular lattices. These models require the specification of **neighborhood structure**. The simplest such structure is a **first order** or **nearest neighbor**.

Gaussian Markov Random Fields

Gaussian Markov Random Fields (GMRF) are useful Gaussian processes for data on regular lattices. These models require the specification of **neighborhood structure**. The simplest such structure is a **first order** or **nearest neighbor**.

Definition: Define

$$z_{\partial_i} = \{z_j : i \sim j\}$$

where $\sim$ denotes neighborhood. A GMRF is defined by its full conditionals

$$\pi(z_i|z_{-i}) = \pi(z_i|z_{\partial_i})$$

where $\pi(z_i|z_{\partial_i})$ is a normal distribution. This implies that all the full conditionals of the process are fully specified by the neighbors of a point.

The former definition of a GMRF is a generalization of the autoregressive processes on the real line.

The former definition of a GMRF is a generalization of the autoregressive processes on the real line.

More formally

Definition: A random vector $x \in \mathbb{R}^n$ is a GMRF with mean μ and precision matrix W iff its density has the form

$$\pi(x) \propto |W|^{1/2} \exp\left\{-\frac{1}{2}(x-\mu)W(x-\mu)\right\}$$

where $W_{ij} \neq 0 \Leftrightarrow i \sim j$.

Gaussian Markov Random Fields

Theorem: Consider a random vector $z \in \mathbb{R}^n$ with density given by

$$\pi(z) = (2\pi)^{-n/2}|W|^{1/2} \exp\left\{-\frac{1}{2}(z-\mu)'W(z-\mu)\right\}$$

Then

$$E(z_i|z_{-i}) = \mu_i - \frac{1}{W_{ii}}\sum_{j\neq i}^{n} W_{ij}(z_j - \mu_j)$$

$$\text{Prec}(z_i|z_{-i}) = W_{ii} \quad \text{and} \quad \text{Corr}(z_i, z_j|z_{ij}) = -\frac{W_{ij}}{\sqrt{W_{ii}W_{jj}}}$$

Gaussian Markov Random Fields

Theorem: Consider a random vector $z \in \mathbb{R}^n$ with density given by

$$\pi(z) = (2\pi)^{-n/2}|W|^{1/2} \exp\left\{-\frac{1}{2}(z-\mu)'W(z-\mu)\right\}$$

Then

$$E(z_i|z_{-i}) = \mu_i - \frac{1}{W_{ii}}\sum_{j\neq i}^{n} W_{ij}(z_j - \mu_j)$$

$$\text{Prec}(z_i|z_{-i}) = W_{ii} \quad \text{and} \quad \text{Corr}(z_i, z_j|z_{ij}) = -\frac{W_{ij}}{\sqrt{W_{ii}W_{jj}}}$$

The proof of the theorem can be found in Rue and Held (2005), Chapter 2.

Thus, for $\mu = 0$,

$$\pi(z_i|z_{-i}) = N\left(-\frac{1}{W_{ii}}\sum_{j\neq i}^{n} W_{ij}z_j, \frac{1}{W_{ii}}\right)$$

which implies that the full conditional of z_i depends only on $z_{\partial i}$.

Thus, for $\mu = 0$,

$$\pi(z_i|z_{-i}) = N\left(-\frac{1}{W_{ii}}\sum_{j\neq i}^{n} W_{ij}z_j, \frac{1}{W_{ii}}\right)$$

which implies that the full conditional of z_i depends only on $z_{\partial i}$.

Notice that the variance $\Sigma = W^{-1}$ (if it exists) provides information about the marginal variances, Σ_{ii} and marginal correlations, $\Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}$. The Precision matrix W provides information about the conditional variances and conditional correlations.

Conditional Autoregressions (CAR)

Suppose we specify a joint distribution by giving the full conditionals as normals with

$$E(z_i|z_{-i}) = -\sum_{j\neq i}^{n} \beta_{ij} z_j \quad \text{and} \quad \text{Prec}(z_i|z_{-i}) = \kappa_i > 0$$

This is known as a **Conditional Autoregression**.

Does this correspond to a multivariate normal distribution?

Conditional Autoregressions (CAR)

Suppose we specify a joint distribution by giving the full conditionals as normals with

$$E(z_i|z_{-i}) = -\sum_{j\neq i}^{n} \beta_{ij} z_j \quad \text{and} \quad \text{Prec}(z_i|z_{-i}) = \kappa_i > 0$$

This is known as a **Conditional Autoregression**.

Does this correspond to a multivariate normal distribution?

The previous theorem suggests that we define a precision matrix W taking

$$W_{ii} = \kappa_i \quad \text{and} \quad W_{ij} = \kappa_i \beta_{ij}$$

and, for symmetry of W, κ and β need to satisfy

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

Brook's Lemma

To answer the question regarding the possibility of specifying a joint distribution through the full conditionals, we need the following lemma.

To answer the question regarding the possibility of specifying a joint distribution through the full conditionals, we need the following lemma.

Lemma: Let $\pi(z)$ be the density of $z \in \mathbb{R}^n$ and let Ω denote the support of $\pi(z)$. Let $z, z' \in \Omega$, then

$$\frac{\pi(z)}{\pi(z')} = \prod_{i=1}^{n} \frac{\pi(z_i|z_1, \ldots, z_{i-1}, z'_{i+1}, \ldots, z'_n)}{\pi(z'_i|z_1, \ldots, z_{i-1}, z'_{i+1}, \ldots, z'_n)}$$

As this equality holds for any fixed point z' in the support of $\pi(z)$, then the lemma implies that the joint density of z is specified by the full conditionals. The proof is in Rue and Held (2005).

To see that a CAR model defines a valid GMRF take $z' = 0$, then Brook's lemma implies that

$$\log \pi(z) = \log \pi(0) - \frac{1}{2} \sum_{i=1}^{n} \kappa_i z_i^2 - \sum_{i=2}^{n} \sum_{j=1}^{i-1} \kappa_i \beta_{ij} z_i z_j$$

a multivariate normal density with precision matrix W.

In practical applications the precision matrix W is specified up to a precision constant λ, so that

$$\pi(z|\lambda) \propto \lambda^{k/2}|W|^{1/2} \exp\left\{-\frac{1}{2}\lambda z' W z\right\}$$

where $|W|$ is defined as the product of the positive eigenvalues of W and k is the rank of W.

An example of a simple GMRF based on linear dependence is

$$\pi(z_i|z_{-i}) = N(\overline{z}_{\partial_i}, 1/n_i) \propto \exp\left\{-\frac{1}{2}\sum_{j\in\partial_i}(z_i - z_j)^2\right\}$$

where n_i is the number of neighbors of z_i. This corresponds to

$$W_{ij} = \begin{cases} n_i & i = j \\ -1 & i \sim j \\ 0 & \text{otherwise} \end{cases}$$

Posterior Distribution

Observed data are usually noisy versions of GMRF, so a typical likelihood for observations y_i over an n grid must include observational error, and would be

$$L(y|z) \propto |W_y|^{1/2} \exp\{-1/2(y-z)'W_y(y-z)\}$$

where the precision of the observational error is usually assumed to be diagonal.

Posterior Distribution

Observed data are usually noisy versions of GMRF, so a typical likelihood for observations y_i over an n grid must include observational error, and would be

$$L(y|z) \propto |W_y|^{1/2} \exp\{-1/2(y-z)'W_y(y-z)\}$$

where the precision of the observational error is usually assumed to be diagonal.

If z follows a GMRF with precision matrix λW, then the posterior distribution of z is

$$\pi(z|y) \sim N(\Lambda^{-1}W_y y, \Lambda^{-1})$$

where $\Lambda = W_y + \lambda W_z$.

Posterior Distribution

The posterior distribution is known in closed for, but for large numbers of grid points it can be prohibitive to perform calculations involving the inverse of Λ. The full conditionals of the posterior distribution, assuming $W_y = \lambda_y I$, are easily obtained as

$$\pi(z_i|z_{-1}, y) \sim N\left(\frac{\lambda_y y_i + \lambda_z \sum_{j\in\partial_i} W_{ij} z_j}{\lambda_y + \lambda_z W_{ii}}, \frac{1}{\lambda_y + \lambda_z W_{ii}}\right)$$

The locally linear GMRF corresponds to a matrix W that has a negative unit for every neighbor in the off-diagonal and diagonal terms that corresponds to the number of neighbors. This implies that $W\mathbb{1} = \mathbf{0}$. W is singular and $\pi(z) = \pi(z + c\mathbb{1})$ for any c. Which means that the distribution of z is not properly specified. This produces a so called **intrinsic** GMRF.

Intrinsic Processes

The locally linear GMRF corresponds to a matrix W that has a negative unit for every neighbor in the off-diagonal and diagonal terms that corresponds to the number of neighbors. This implies that $W\mathbb{1} = \mathbf{0}$. W is singular and $\pi(z) = \pi(z + c\mathbb{1})$ for any c. Which means that the distribution of z is not properly specified. This produces a so called **intrinsic** GMRF.

To obtain a properly specified prior for z we can add a small amount to each diagonal element of the precision matrix, so that W is replaced by $W + \delta I$.

Process Convolutions

GMRF are natural processes to us as priors for the latent process in a process convolution. In the standard formulation we have that the process convolutions is given as

$$y = Kx + \varepsilon, \;\; \varepsilon \sim N(0, 1/\lambda_y I), \;\; x \sim N(0, 1/\lambda_x I)$$

we can replace the last assumption by $w \sim GMRF(\lambda_x W)$. This will create dependencies among the latent process. The parameters of the GMRF, λ_x can be used to induce regularity in the process convolution in a way similar to the kernel width.

Using gamma priors for λ_y and λ_x will produce the posterior distribution

$$\pi(x, \lambda_x, \lambda_y | y) \propto \lambda_y^{a_y + n/2 - 1} \exp\{-\lambda_y (b_y + .5(y - Kx)^T (y - Kx))\}$$

$$\times \lambda_x^{a_x + m/2 - 1} \exp\{-\lambda_x (b_x + .5 x^T W x)\}$$