Markov random fields; CAR and SAR models

• Motivation: Markov dependence for time series
  – Decomposing the joint distribution of the AR(1) process
  – Using a neighborhood structure

• Markov random fields (MRFs)
  – Cliques
  – Expanding the energy function

• The conditional autoregression (CAR) or auto-normal model

• The simultaneous autoregressive (SAR) model

• Choosing the form of $B$ and $C$
  – Shrinking the effect of $W$ to gain nonsingular matrices

• CAR or SAR?
Introduction

• Let \( \{ \mathcal{B}_i : i = 1, \ldots, m \} \) denote a partition of \( m \) distinct regions, such that \( \bigcup_{i=1}^{m} \mathcal{B}_i = D \) and \( \mathcal{B}_i \cap \mathcal{B}_j = \emptyset \) for each \( i \neq j \).

• In this section we introduce Markov random fields (MRFs) as a mechanism to construct dependence among the random variables of the areal process:

\[
\{ Z_i = Z(\mathcal{B}_i) : i = 1, \ldots, m \}.
\]

• We will consider conditional and simultaneous representations to define these models.

• We motivate our discussion with a time series model.
Motivation: Markov dependence for time series

• In time series analysis, we often consider the autoregressive process of order one, $AR(1)$, as the simplest model with time series dependence.

• Let $\{Z_t : t \in \mathbb{Z}\}$ denote a white noise process with mean 0 and variance $\sigma^2 > 0$.

• Then $\{X_t : t \in \mathbb{Z}\}$ is an $AR(1)$ process if it can be written as the solution of the following equation:

$$X_t = \phi X_{t-1} + Z_t.$$

• When $|\phi| < 1$, the process $\{X_t\}$ is stationary, and the solution to the equation is unique.
Decomposing the joint distribution of the AR(1)

- Now suppose that we want to write down the joint density/mass function, \( f(\mathbf{x}) \), for the random vector \( \mathbf{X} = (X_1, \ldots, X_n)^T \).

- We can factorize the joint distribution into a product of conditional distributions:

\[
    f(\mathbf{x}) = f(x_1) \times f(x_2|x_1) \times f(x_3|x_2, x_1) \times \cdots \times f(x_n|x_{n-1}, \ldots, x_1).
\]

- But in the AR(1) process the distribution of the current \( X_t \) only depends on the past value \( X_{t-1} \).

  - We say that the process is Markov-in-time of order one.

- Thus the joint distribution can be simplified to

\[
    f(\mathbf{x}) = f(x_1) \times f(x_2|x_1) \times f(x_3|x_2) \times \cdots \times f(x_n|x_{n-1}).
\]
Using a neighborhood structure

- For a time series process \( \{X_t\} \), the \textbf{neighbors} \( N_t \) are defined to be the set of other times for which we can simplify the conditional distribution to

\[
f(x_t|\{x_s: s \neq t\}) = f(x_t|\{x_s: s \in N_t\})
\]

- For the AR(1) process the neighbors are \( N_1 = \emptyset \), with

\[
N_t = \{t - 1\}, \quad \text{for } t \geq 2.
\]

- Thus we have decomposed the joint distribution of the AR(1) process into a product of conditional distributions depending on the neighbors:

\[
f(x) = \prod_{t=1}^{n} f(x_t|\{x_s: s \in N_t\}).
\]

- Now consider the opposite (much harder!) question:

1. Does a product of conditional distributions based on neighborhoods define a valid and proper joint distribution for \( \{X_t\} \)?
2. How does this extend to spatial processes?
Markov random fields (MRFs)

- In general the answer to the first question is not always yes.
  - You cannot write down any old product of conditional distributions.

- In the spatial case, now suppose that we have random variables \( \{Z_i = Z(B_i) : i = 1 \ldots, m\} \) observed on regions \( B_i \) that form a partition of a space \( D \).

- The (spatial) neighbors \( N_i \) are the set of the other areas for which we can simplify the conditional distribution to

  \[
  f(z_i | \{z_j : j \neq i\}) = f(z_i | \{z_j : j \in N_i\}).
  \]

- When a product of conditional distributions based on a spatial neighborhood structure can be used to define a valid and proper joint distribution we call the process a Markov random field (MRF).

Clique

- A **clique** of $D$, $C$, is a set of regions, such that all pairs of regions in $C$ are neighbors of one another or $C$ is a single region.

- Let $cl(D)$ denote the set of all cliques of $D$.

- The Hammersley-Clifford theorem [Hammersley and Clifford, 1971] tells us that $\{Z_i\}$ is a Markov random field on $D$ if and only if

$$f(z) = \frac{1}{K} \exp(E(z)/T),$$

where $K$ is a normalizing constant (also called the **partition function**), $T$ is a constant called the **temperature**, and $E(z)$ is called the **energy function**.

(The names all come from statistical physics.)
Expanding the energy function

- The energy function can be expressed as [e.g. Grimmett, 1973, Besag, 1974, Kaiser and Cressie, 2000]

$$E(x) = \sum_{C \in cl(D)} V_C(x),$$

where the functions $V_C(\cdot)$ are called \textbf{clique potentials}.

- These potentials define the \textbf{local} dependence structure which in turn define the joint \textbf{global} distribution for the MRF.

- By changing the functional form of the potential, we change the distribution and dependence structure of the MRF.

- In his celebrated paper, Besag [1974] defines the class of \textbf{autoregressive models}, based on writing down potential functions corresponding to exponential families (e.g, Normal, Binomial, Poisson, Exponential).

- For the auto-normal, the normalizing constant $K$ is known. In many other cases it is unknown, which complicates the inference.
The conditional autoregression or auto-normal model

• Let

\[ Z_{(-i)} = \{Z_j : j \neq i\}, \]

denote the collection of RVs at all regions except \( B_i \).

• In the conditional autoregressive (CAR) model we write down a model for the distribution of \( Z_i \) conditional on \( Z_{(-i)} \)

\[
Z_i | Z_{(-i)} \sim N \left( \mu_i + \sum_{j=1}^{m} c_{ij} (z_j - \mu_j), \tau_i^2 \right).
\]

where here \( m \) could be infinite or finite.

In the above expression

- The expectation is \( \mu_i = E(Z_i) \), and the conditional variance is \( \tau_i^2 \).

- \( \{c_{ij} : i = 1, \ldots, m, j = 1, \ldots, m\} \) are a set of unknown or known constants with \( c_{ii} = 0 \) for all \( i \).

We can set \( c_{ij} \) equal to zero if \( j \) is not a neighbor of \( i \).
The CAR model, continued

- The conditional density satisfies

\[
f(z_i | z_{(-i)}) \propto \exp \left\{ \frac{\left[ z_i - \mu_i - \sum_{j=1}^{m} c_{ij}(z_j - \mu_j) \right]^2}{2\tau_i^2} \right\},
\]

and the joint distribution can be written as a multivariate normal.

- [Besag, 1974] Suppose that \( c_{ij}/\tau_i^2 = c_{ji}/\tau_j^2 \) (a symmetry condition) for all \( i \) and \( j \).

Let \( \mu = (\mu_1, \ldots, \mu_m) \), \( D = \text{diag}(\tau_1^2, \ldots, \tau_m^2) \), and \( C = (C_{ij}) \).

Then if \( D^{-1}(I - C) \) is positive definite,

\[
Z \sim N_m(\mu, (I - C)^{-1}D).
\]

- **Important remark**: By specifying \( C \) we are actually writing down a model for the inverse of the covariance matrix (also called the **precision matrix**) – compare this to a geostatistical model in which we model the covariance matrix.

  - Implication: CAR modeling can be very computationally efficient.
The simultaneous autoregressive (SAR) model

- Instead of conditioning on the other regions, we write down a model in terms all the regions at once.

- In the **simultaneous autoregressive (SAR) model** let

  \[ Z_i = \mu_i + \sum_{j=1}^{m} b_{ij} [Z_j - \mu_j] + \epsilon_i, \]

  where again \( m \) could be infinite or finite.

In the above expression:

- Again \( \mu_i = E(Z_i) \) for each \( i \).

- \( \{b_{ij} : i = 1, \ldots, m, j = 1, \ldots, m\} \) are a set of known or unknown constants with \( b_{ii} = 0 \) for all \( i \).

- \( \{\epsilon_i : i = 1, \ldots, m\} \) is a set of independent \( N(0, \tau_i^2) \) RVs.

- Let \( \mu = (\mu_1, \ldots, \mu_m) \), \( D = \text{diag}(\tau_1^2, \ldots, \tau_m^2) \), and \( B = (B_{ij}) \).

  If \((I - B)\) is of full rank,

  \( Z \sim N_m(\mu, (I - B)^{-1}D(I - B)^{-T}), \)

  where \( A^{-T} \) is the transpose of the inverse of the matrix \( A \).
Choosing the form of $B$ and $C$

- Most commonly the matrices $C$ and $B$ in the CAR and SAR models respectively are defined using some proximity matrix $W$.

  - But we have to be careful to make sure that we get non-singular covariance matrices.

- Let us take the CAR model as an example.

- Suppose that $W$ is symmetric. Let $c_{ij} = w_{ij}/w_{i\bullet}$ and $\tau_i^2 = \tau^2/w_{i\bullet}$, with $\mu_i = 0$.

Then

$$Z_i | Z_{(-i)} \sim N \left( \frac{\sum_{j \neq i} w_{ij}z_j}{w_{i\bullet}}, \frac{\tau^2}{w_{i\bullet}} \right).$$

We have that the precision matrix is

$$D^{-1}(I - C) = D^{-1} - D^{-1}C = \frac{1}{\tau^2} (D_w - W),$$

where $D_w = \text{diag}(w_{1\bullet}, \ldots, w_{m\bullet})$.

In this case $D_w - W$ is singular and we obtain what is called the intrinsic autoregressive model.
Shrinking the effect of $W$ to gain nonsingular matrices

- We can solve this problem by shrinking the effect of $W$ using a spatial dependence parameter $\rho$.

- Our precision matrix then becomes

$$D_w - \rho W.$$

- This matrix is invertible as long as

$$\rho \in \left[ \frac{1}{\lambda_{\text{min}}}, \frac{1}{\lambda_{\text{max}}} \right],$$

where $\lambda_{\text{min}}$ and $\lambda_{\text{max}}$ denote respectively the minimum and maximum eigenvalue of the matrix $D_w^{-1/2} W D_w^{-1/2}$.

- With $\rho$ in this range

$$Z \sim N_m(0, \tau^2(D_w - \rho W)^{-1}).$$

- There are other parameterizations for the precision.
CAR or SAR?

- Clearly choosing $C$ and $B$ so that the covariance matrices are equal under the two models leads to the same model.
  
  - Not a very sensible comparison in practice as $B$ and $C$ are constructed using the proximity matrix.

- See Wall [2004] for a more in-depth comparison, showing that for a given proximity matrix, that results of both modeling approaches are different, but also can be counter-intuitive.
  (we are used to modeling covariances, not precisions).
References


