Non-Gaussian data

ST733 – Spatial Statistics

Non-Gaussian Data

- We need new methods to handle non-Gaussian data
- Spatial binary data: $Y(\mathbf{s}) = 1$ if site $\mathbf{s}$ is diseased and $Y(\mathbf{s}) = 0$ otherwise
- Spatial count data: $Y(\mathbf{s}) \in \{0, 1, 2, ...\}$ is the number days at site $\mathbf{s}$ with high temperature
- Do you know of an n-dimensional Poisson distribution with an $n \times n$ covariance matrix?

Non-Gaussian Data

- Spatial dependence is often induced by adding Gaussian random effects to a generalized linear model (GLM)
- Two equivalent representation of the Gaussian model:
  - Conditional model:
  - Marginal model:
- Only the conditional model is available for non-Gaussian data, but this is sufficient

GLMs with spatial random effects

- Non-spatial GLM:

- Spatial GLM:

- Spatial GLM for binary data:

- Spatial GLM for count data:

GLMs with spatial random effects

- Should we include a nugget term in the spatial random effect model?

- How to measure spatial dependence?

GLMs with spatial random effects

How to fit the model

$$Y(\mathbf{s})|\lambda(\mathbf{s}) \overset{indep}{\sim} \text{Poisson}(\lambda(\mathbf{s})) \text{ where } \log(\lambda(\mathbf{s})) = \mathbf{X}(\mathbf{s})\beta + \mu(\mathbf{s})$$

and μ is a GP?

- MLE:
- Bayes:
- Large n

Copulas

- Adding spatial random effects induces dependence, but except for the normal model they lead to complicated marginal distributions
- A spatial copula maintains a desired (continuous) marginal distribution while allowing for spatial dependence
- For example, say we want the marginal distribution of $Y(\mathbf{s})$ to be Gamma(a, b) for all $\mathbf{s}$
- The copula model is:

Copula

- Say $(Z_1, ..., Z_n) \sim \text{Normal}(0, \Sigma)$ and $Y_i = Q[\Phi(Z_i)]$ where Q is the Gamma(a, b) quantile function

- Derive the joint distribution of $(Y_1, ..., Y_n)$

Spatial probit model for binary data

- A special case where Bayesian computing works nicely is the probit regression model for binary data[1]
- Non-spatial probit regression is

$$\text{Prob}(Y_i = 1) = \Phi(\mathbf{X}_i\beta)$$

  where Φ is the standard normal PDF
- This is equivalent to $Y_i = I(Z_i > 0)$ where

$$Z_i \stackrel{indep}{\sim} \text{Normal}(\mathbf{X}_i\beta, 1)$$

- The variance of Z_i must be one for identification:

[1] Polya-gamma priors are similar for spatial logistic models, though a bit more complex

Spatial probit model for binary data

- Spatial extension:

- MCMC imputes the Z_i each step from a truncated normal distribution

- Conditioned on these latent variables the model is equivalent to Gaussian spatial model