Introduction to Non-Normality (and Nonparametrics) in Spatial Statistics

Based on Handbook of Spatial Statistics Ch. 4, 7, 11

John Paige

Statistics Department
University of Washington

April 11, 2017

Topics

- Spatial GLMs and GLMMs
- Gaussian-log-Gaussian (GLG) mixture modeling
- Bayesian Nonparametrics (stick breaking priors)
- Others (not discussed here)
  - Spatial Deformation Method
  - Copulas
  - Transform data directly before modeling (e.g. Box-Cox)

Spatial Generalized Linear Mixed Models (GLMMs): Formulation

- $[Y(\mathbf{s}_i)|\boldsymbol{\eta}, \beta]$, $i = 1, ..., m$ is conditionally independent for any location, $\boldsymbol{s}_i$, with conditional mean $E[Y(\mathbf{s}_i)|\boldsymbol{\eta}, \beta] = \mu(\mathbf{s}_i)$
- $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{H}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$
  - $g(\cdot)$ link function
  - $\mathbf{X}\beta$ linear fixed effects
  - $\mathbf{H}\boldsymbol{\eta}$ latent random effects (spatial process, typically Gaussian)
  - $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 I)$ site-specific random variation

Spatial Generalized Linear Mixed Models (GLMMs): Computation

- Likelihood not directly computable:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{Y}) = \int_{\mathbf{R}^n} \prod_{i=1}^{n} f_i(Y(\mathbf{s}_i)|\boldsymbol{\eta}, \boldsymbol{\theta}) \; d\eta$$

- Possible alternatives:
  - EM (Zhang, 2002)
  - Composite (pairwise) likelihood (e.g. Varin, Host, and Skare 2005)
  - MCMC
    - Langevin updates for $\boldsymbol{\eta}$ suggested (e.g. Diggle and Robeiro, 2007)
    - If ε is included, one can use conjugate MVN Gibbs updates for $\boldsymbol{\eta}$ if σ_ε known or nearly known (e.g. Wikle 2002, Royle and Wikle, 2005)
- geoRglm R package (Diggle and Ribeiro, 2007)

Gaussian-log-Gaussian (GLG) mixture model

- Palacias and Steel (2006) suggest the model:

$$y_i = \mathbf{x}(\mathbf{s}_i)^T \boldsymbol{\beta} + \frac{\eta_i}{\sqrt{\lambda_i}} + \epsilon_i$$

  - $\boldsymbol{\eta} \sim N(\mathbf{0}, \sigma^2 \mathbf{C}_{\boldsymbol{\theta}})$ Spatial process
  - $\log(\boldsymbol{\lambda}) \sim N\left(-\frac{\nu}{2}\mathbf{1}, \nu \mathbf{C}_{\boldsymbol{\theta}}\right)$ mixing variables independent of $\boldsymbol{\eta}$, $\boldsymbol{\epsilon}$
  - $\boldsymbol{\epsilon}$ iid mean zero variance τ^2 normal noise ("nugget")

Gaussian-log-Gaussian (GLG) mixture model

$$y_i = \mathbf{x}(\mathbf{s}_i)^T \boldsymbol{\beta} + \frac{\eta_i}{\sqrt{\lambda_i}} + \epsilon_i$$

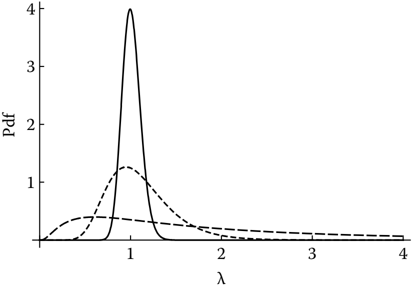

FIGURE 11.1
Marginal probability density function of mixing variables λ_i for various values of ν. Solid line $\nu = 0.01$, short dashes: $\nu = 0.1$, long dashes: $\nu = 1$.

Gaussian-log-Gaussian (GLG) mixture model

$$y_i = \mathbf{x}(\mathbf{s}_i)^T \boldsymbol{\beta} + \frac{\eta_i}{\sqrt{\lambda_i}} + \epsilon_i$$

Nice properties:

- $E[\lambda_i] = 1$
- $\mathrm{Var}\,(\lambda_i) = \exp\{\nu\} - 1$
- $\nu \to 0 \Rightarrow \boldsymbol{\lambda} \to \mathbf{1} \Rightarrow y_i \to$ (Gaussian)
- Large $\nu \Rightarrow$ heavy-tailed, highly skewed distribution
- Easily computable correlation function

Stick Breaking Priors

In order to have nonparametric model in Bayesian setting we need a prior over an infinite-dimensional model space:

Definition

A random probability distribution, F, has a **stick-breaking prior** if:

$$F \stackrel{d}{=} \sum_{i=1}^{N} p_i \delta_{\theta_i}$$

where δ_z is a Dirac measure at z, $p_i = V_i \prod_{j<i}(1 - V_j)$ where $V_1, ..., V_{N-1} \stackrel{iid}{\sim} Beta(a_i, b_i)$, and $\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_N \stackrel{iid}{\sim} H$ for centering (aka base) distribution H. Here, N could be infinite.

Stick Breaking Priors

When $N = \infty$, there are several well-known priors in this class:

- Dirichlet Process Prior (see Ferguson, 1973 and Sethuraman, 1994)
- The Pitman-Yor (or two-parameter Poisson-Dirichlet) process

Generalized Spatial Dirichlet Process

- The main idea is to allow the $\boldsymbol{\theta}_j$'s in our stick-breaking prior to be spatially dependent (Duan, Guindani, and Gelfand 2007 and Gelfand, Guindani, and Petrone, 2007)
- Allow the surface selection to depend on location:

$$F^{(n)} \overset{d}{=} \sum_{i_1=1}^{\infty} \dots \sum_{i_n=1}^{\infty} p_{i_1,\dots,i_n} \delta_{\theta_{i_1}} \dots \delta_{\theta_{i_n}}$$

  for $i_j = i(\mathbf{s}_j)$ are the location indices and θ_{i_j} locations are drawn from centering random field H. Weights $p_{i_1,\dots,i_n}$ are distributed independently from the locations on the infinite unit simplex.
- Constraints on the weights allow for smooth (mean square continuous) random probability measure
- Note: this requires replications in time

Hybrid Dirichlet Mixture Models

- Petrone, Guindani, and Gelfand (2009) take previous model further using species sampling prior, which is more general than stick-breaking prior (the weights, V_i, are allowed to be general). The model is:

$$\mathbf{y}_i|\boldsymbol{\theta}_i \overset{ind}{\sim} N_m(\boldsymbol{\theta}_i, \sigma^2\mathbf{I}_m),$$
$$\boldsymbol{\theta}_i|F_{\mathbf{x}_1,\ldots,\mathbf{x}_m} \overset{iid}{\sim} F_{\mathbf{x}_1,\ldots,\mathbf{x}_m},$$
$$F_{\mathbf{x}_1,\ldots,\mathbf{x}_m} \overset{d}{=} \sum_{j_1=1}^{k} \ldots \sum_{j_m=1}^{k} p(j_1, \ldots, j_m)\delta_{\theta_{j_1,1},\ldots,\theta_{j_m,m}}$$

where:

- $p(j_1, \ldots, j_m)$ is proportion of (hybrid) species $(\theta_{j_1,1}, \ldots, \theta_{j_m,m})$ in the population, and $\boldsymbol{\theta}_j = (\theta_{j,1}, \ldots, \theta_{j,m}) \overset{iid}{\sim} H$
- H typically Gaussian process

Order-Based Dependent Dirichlet Processes (πDDPs)

- Griffon and Steel (2006)
- Unlike previous nonparametric Bayesian methods presented, doesn't require replication of spatial fields
- Define rankings of weights, $\mathbf{V}$, via orderings, $\boldsymbol{\pi}(\mathbf{s})$
- Since $p_i = V_i \prod_{j<i}(1 - V_j)$, we find $E[p_i] \geq E[p_{i+1}]$. Hence similar orderings have similar distributions

Spatial Kernel Stick-Breaking Prior

$$
\begin{aligned}
Y(\mathbf{s}) &= \eta(\mathbf{s}) + x(\mathbf{s})^T \boldsymbol{\beta} + \epsilon(\mathbf{s}) \\
\boldsymbol{\eta}(\mathbf{s}) &\sim F_{\mathbf{s}}(\eta) \overset{d}{=} \sum_{i=1}^{N} p_i(\mathbf{s}) \delta_{\boldsymbol{\theta}_i} \\
p_i(\mathbf{s}) &= V_i(\mathbf{s}) \prod_{j=1}^{i-1} (1 - V_j(\mathbf{s})) \\
V_i(\mathbf{s}) &= w_i(\mathbf{s}) V_i \\
V_i &\sim \text{Beta}(a, b) \\
\boldsymbol{\theta}_i &\overset{iid}{\sim} H \\
w_i(\mathbf{s}) &\in [0, 1]
\end{aligned}
$$

- $w_i(\mathbf{s})$ can be, e.g. kernel function or $\log(w_i(\mathbf{s})/(1 - w_i(\mathbf{s}))) \sim GP$

Spatial Kernel Stick-Breaking Prior

TABLE 11.4

Examples of Kernel Functions and the Induced Functions $\gamma(\mathbf{s}, \mathbf{s}')$, Where $\mathbf{s} = (s_1, s_2)$, $h_1 = |s_1 - s_1'| + |s_2 - s_2'|$, $h_2 = \sqrt{(s_1 - s_1')^2 + (s_2 - s_2')^2}$, $I(\cdot)$ is the Indicator Function, and $x^+ = \max(x, 0)$

Name	$w_i(\mathbf{s})$	Model for κ_{1i} and κ_{2i}	$\gamma(\mathbf{s}, \mathbf{s}')$
Uniform	$\prod_{j=1}^{2} I\left(\lvert s_j - \psi_{ji}\rvert < \frac{\kappa_{ji}}{2}\right)$	$\kappa_{1i}, \kappa_{2i} \equiv \lambda$	$\prod_{j=1}^{2}\left(1 - \frac{\lvert s_j - s_j'\rvert}{\lambda}\right)^{+}$
Uniform	$\prod_{j=1}^{2} I\left(\lvert s_j - \psi_{ji}\rvert \frac{<\kappa_{ji}}{2}\right)$	$\kappa_{1i}, \kappa_{2i} \sim \text{Exp}(\lambda)$	$\exp(-h_1/\lambda)$
Exponential	$\prod_{j=1}^{2} \exp\left(-\frac{\lvert s_j - \psi_{ji}\rvert}{\kappa_{ji}}\right)$	$\kappa_{1i}, \kappa_{2i} \equiv \lambda$	$0.25\left[\prod_{j=1}^{2}\left(1 + \frac{\lvert s_j - s_j'\rvert}{\lambda}\right)\right]\exp\left(-\frac{h_1}{\lambda}\right)$
Squared exp.	$\prod_{j=1}^{2} \exp\left(-\frac{(s_j - \psi_{ji})^2}{\kappa_{ji}^2}\right)$	$\kappa_{1i}, \kappa_{2i} \equiv \lambda^2/2$	$0.5\exp\left(-\frac{h_2^2}{\lambda^2}\right)$
Squared exp.	$\prod_{j=1}^{2} \exp\left(-\frac{(s_j - \psi_{ji})^2}{\kappa_{ji}^2}\right)$	$\kappa_{1i}, \kappa_{2i} \sim \text{InvGa}\left(\frac{3}{2}, \frac{\lambda^2}{2}\right)$	$0.5/\left(1 + (\frac{h_2}{\lambda})^2\right)$