Modeling non-Gaussian spatial data

Jeffrey Doser[1] & Andrew Finley[2]

May 15, 2023

[1]Department of Integrative Biology, Michigan State University.

[2]Department of Forestry, Michigan State University.

Non-Gaussian spatial data

- Often data sets preclude Gaussian modeling: $y(\mathbf{s})$ may not even be continuous
- Examples:
  - Binary: presence or absence of a species at location $\mathbf{s}$.
  - Count: abundance of a species at location $\mathbf{s}$.
  - Categorical: counts of trees by size class at location $\mathbf{s}$.
- Replace Gaussian likelihood by exponential family member (Diggle, Tawn, and Moyeed (1998)).

Hierarchical Bayesian approach

- **First stage**: $y(\mathbf{s}_i)$ are conditionally independent given β and $\mathrm{w}(\mathbf{s}_i)$. Here we use a canonical link function, say $g(E[y(\mathbf{s}_i)]) = \eta(\mathbf{s}_i) = \mathbf{x}(\mathbf{s}_i)^\top \beta + \mathrm{w}(\mathbf{s}_i)$.

Hierarchical Bayesian approach

- **First stage**: $y(\mathbf{s}_i)$ are conditionally independent given $\boldsymbol{\beta}$ and $\mathrm{w}(\mathbf{s}_i)$. Here we use a canonical link function, say $g(E[y(\mathbf{s}_i)]) = \eta(\mathbf{s}_i) = \mathbf{x}(\mathbf{s}_i)^\top \boldsymbol{\beta} + \mathrm{w}(\mathbf{s}_i)$.
- **Second stage**: Model $\mathrm{w}(\mathbf{s}_i)$ as a Gaussian process:

$$\mathbf{w} \sim N(\mathbf{0}, \sigma^2 \mathbf{R}(\phi))$$

Hierarchical Bayesian approach

- **First stage**: $y(\mathbf{s}_i)$ are conditionally independent given β and $\mathrm{w}(\mathbf{s}_i)$. Here we use a canonical link function, say $g(E[y(\mathbf{s}_i)]) = \eta(\mathbf{s}_i) = \mathbf{x}(\mathbf{s}_i)^\top \beta + \mathrm{w}(\mathbf{s}_i)$.
- **Second stage**: Model $\mathrm{w}(\mathbf{s}_i)$ as a Gaussian process:

$$\mathbf{w} \sim N(\mathbf{0}, \sigma^2 \mathbf{R}(\phi))$$

- **Third stage**: Priors and hyperpriors.

MCMC sampling for spatial GLMMs

- Additional GLMM flexibility comes at a computational cost: lose conjugacy of $\beta, \mathbf{w}$
- Requires more Metropolis steps. Particularly costly for $\mathbf{w}$
- Practical consequence: slower, less efficient algorithms
- Prediction and interpolation proceed as with the Gaussian case

Pólya-Gamma data augmentation

- General approach for Bayesian (spatial) logistic regression that yields conjugate updates of β (and $\mathbf{w}$)

Pólya-Gamma data augmentation

- General approach for Bayesian (spatial) logistic regression that yields conjugate updates of β (and $\mathbf{w}$)
- Introduce augmented data $\omega(\mathbf{s}_i)$ for each $i = 1, \ldots, n$, where $\omega(\mathbf{s}_i) \sim \mathrm{PG}(N(\mathbf{s}_i), 0)$, with $N(\mathbf{s}_i)$ the Binomial weights

Pólya-Gamma data augmentation

- General approach for Bayesian (spatial) logistic regression that yields conjugate updates of β (and $\mathbf{w}$)
- Introduce augmented data $\omega(\mathbf{s}_i)$ for each $i = 1, \ldots, n$, where $\omega(\mathbf{s}_i) \sim \text{PG}(N(\mathbf{s}_i), 0)$, with $N(\mathbf{s}_i)$ the Binomial weights
- Define $\kappa(\mathbf{s}_i) = y(\mathbf{s}_i) - N(\mathbf{s}_i)/2$

Pólya-Gamma data augmentation

- General approach for Bayesian (spatial) logistic regression that yields conjugate updates of β (and $\mathbf{w}$)
- Introduce augmented data $\omega(\mathbf{s}_i)$ for each $i = 1, \ldots, n$, where $\omega(\mathbf{s}_i) \sim \text{PG}(N(\mathbf{s}_i), 0)$, with $N(\mathbf{s}_i)$ the Binomial weights
- Define $\kappa(\mathbf{s}_i) = y(\mathbf{s}_i) - N(\mathbf{s}_i)/2$
- Resulting Gibbs sampler is remarkably similar to that of a Gaussian model with response $y(\mathbf{s}_i)^* = \kappa(\mathbf{s}_i)/\omega(\mathbf{s}_i)$ and heteroskedastic variances $\tau^2(\mathbf{s}_i) = 1/\omega(\mathbf{s}_i)$.

Pólya-Gamma data augmentation

- Suppose $y(\mathbf{s}_i) \sim \text{Bernoulli}(\psi(\mathbf{s}_i))$.

$$\begin{aligned}\psi(\mathbf{s}_i)^{y(\mathbf{s}_i)}(1-\psi(\mathbf{s}_i))^{1-y(\mathbf{s}_i)} &= \frac{\exp(\mathbf{x}(\mathbf{s}_i)^\top \boldsymbol{\beta} + \mathrm{w}(\mathbf{s}_i))^{y(\mathbf{s}_i)}}{1+\exp(\mathbf{x}\mathbf{s}_i^\top \boldsymbol{\beta} + \mathrm{w}(\mathbf{s}_i))} \\ &= \exp(\kappa(\mathbf{s}_i)(\mathbf{x}(\mathbf{s}_i)^\top \boldsymbol{\beta} + \mathrm{w}(\mathbf{s}_i)))\times \\ &\int \exp(-\frac{\omega(\mathbf{s}_i)}{2}(\mathbf{x}(\mathbf{s}_i)^\top \boldsymbol{\beta} + \mathrm{w}(\mathbf{s}_i)))^2 \times \\ &\qquad p(\omega(\mathbf{s}_i) \mid 1, 0) d\omega(\mathbf{s}_i),\end{aligned}$$

- $p(\omega(\mathbf{s}_i) \mid 1, 0)$ is the Pólya-Gamma PDF with parameters 1 and 0
- With Gaussian priors on $\boldsymbol{\beta}$ and IG prior on σ^2, full conditionals for $\boldsymbol{\beta}$, σ^2, and $\mathbf{w}$ are available in closed form. ϕ updated with MH.
- **See Polson, Scott, Windle (2013) JASA**

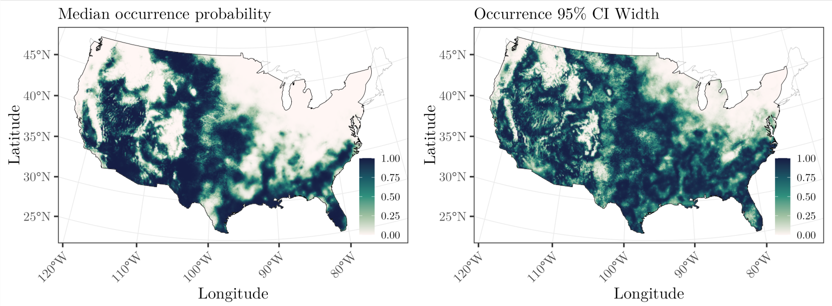
Example: Binary spatial regression

- Objective: predict the distribution of Loggerhead Shrike across the US

$$
\begin{aligned}
y(\mathbf{s}_i) &\sim \text{Bernoulli}(\psi(\mathbf{s}_i)) \\
\text{logit}(\psi(\mathbf{s}_i)) &= \mathbf{x}(\mathbf{s}_i)^\top \boldsymbol{\beta} + w(\mathbf{s}_i) \\
\mathbf{w} &\sim N(\mathbf{0}, \sigma^2 \mathbf{R}(\phi)) \\
\boldsymbol{\beta} &\sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \\
\sigma^2 &\sim IG(a_\sigma, b_\sigma) \\
\phi &\sim \text{Uniform}(l, u) \\
\omega(\mathbf{s}_i) &\sim \text{PG}(1, 0)
\end{aligned}
$$

Example: Binary spatial regression

Posterior predictive inference proceeds as with the Gaussian case



Some practical considerations

- Priors for σ^2 and ϕ may need to be more informative, particularly for binary data.

Some practical considerations

- Priors for σ^2 and ϕ may need to be more informative, particularly for binary data.
- Be careful with non-identity link functions when thinking about priors.

Some practical considerations

- Priors for σ^2 and ϕ may need to be more informative, particularly for binary data.
- Be careful with non-identity link functions when thinking about priors.
- Pólya-Gamma data augmentation works really for binomial data. Computational cost increases as Binomial weights increases.

Some practical considerations

- Priors for σ^2 and ϕ may need to be more informative, particularly for binary data.
- Be careful with non-identity link functions when thinking about priors.
- Pólya-Gamma data augmentation works really for binomial data. Computational cost increases as Binomial weights increases.
- Pólya-Gamma data augmentation also applicable for Negative Binomial count data, but slow for large counts and can be unstable.

Software

- `spBayes`
  - Univariate and multivariate, full GPs or predictive processes
  - Gaussian, Binomial (no Pólya-Gamma data augmentation), Poisson
- `spNNGP`
  - Univariate, NNGPs
  - Gaussian, Binomial
- `spOccupancy`
  - Univariate and multivariate, focus on modeling wildlife distributions, full GPs or NNGPs
  - Bernoulli
- `spAbundance` (`https://github.com/doserjef/spAbundance`)
  - Univariate and multivariate, focus on modeling wildlife/plant abundance, NNGPs
  - Gaussian, Poisson, Negative Binomial