Projection-based Methods for Hierarchical Spatial Models

Collaborators:

Yawen Guan, University of Nebraska

Jaewoo Park, Yonsei University, South Korea

Ben Seiyon Lee, Penn State University

GEOMED 2019, University of Glasgow, Scotland

August 2019

Murali Haran

Department of Statistics, Penn State University

Talk Summary

- Gaussian and non-Gaussian spatial data are common.
  Theme of this conference!
- Hierarchical spatial models, including spatial generalized linear mixed models (SGLMMs)
  - Popular for lattice or areal data
    Besag, York, Mollie (1991) $\approx$ 3,000 citations
  - and continuous-domain data
    Diggle et al. (1998) $\approx$ 3,000 citations
- Challenges:
  1. Computational
  2. Regression parameter interpretation is tricky
- Projection-based methods to help address these issues
- (Unusual) I will discuss Bayes *and* ML inference

Key References

- Guan and Haran (2018): A computationally efficient projection-based approach for spatial generalized linear mixed models, *Journal of Computational and Graphical Statistics,* 27:4, 701-714.
- Guan and Haran (2019): Fast expectation-maximization algorithms for spatial generalized linear mixed models

Both on arxiv.org (this week!)

Spatial Generalized Linear Mixed Models

(Diggle et al., 1998)

Example model for count data $Z(\mathbf{s}), s \in \mathcal{D} \subset \mathcal{R}^d$.

1. Data model:

$$Z(\mathbf{s}_i) \mid \beta, W(\mathbf{s}_i) \overset{Indep.}{\sim} \texttt{Poisson}(\mu(\mathbf{s}_i)), i = 1, \ldots, n$$

$$\log\left(\mu(\mathbf{s}_i)\right) = X(\mathbf{s}_i)\beta + W(\mathbf{s}_i),$$

2. Process model: latent variable **W**, impose spatial dependence via Gaussian process resulting in $\mathbf{W} \mid \sigma^2, \phi \sim MN\left(\mathbf{0}, \sigma^2 \Sigma_\phi\right)$,
   - $\Sigma_{ij} = COV(W(\mathbf{s}_i), W(\mathbf{s}_j)) = C(||\mathbf{s}_i - \mathbf{s}_j||)$
   - e.g. $C(||\mathbf{s}_i - \mathbf{s}_j||) = \sigma^2 \exp(-\frac{||\mathbf{s}_i - \mathbf{s}_j||}{\phi})$ (Exponential)

SGLMMs with latent GMRFs

(Besag et al., 1991; Rue and Held, 2005)

2* (Another version) latent variable **W**, impose spatial dependence via Gaussian Markov random field resulting in $(W(\mathbf{s}_1), \ldots W(\mathbf{s}_n))^T \mid \Theta \sim N(\mathbf{0}, Q(\Theta)^{-1})$, $Q(\Theta)$ is a precision matrix based on a graph that describes a neighborhood structure: adjacencies specify dependence

Challenges with SGLMMs

- Bayesian inference is based on $\pi(\theta, \mathbf{W} \mid \mathbf{Z}) \propto$

$$\prod_i^n f\left(Z(\mathbf{s}_i) \mid \boldsymbol{\beta}, W(\mathbf{s}_i)\right) |\sigma^2 \Sigma_\phi|^{-\frac{1}{2}} \exp\left(-\frac{\mathbf{W}' \Sigma_\phi^{-1} \mathbf{W}}{2\sigma^2}\right) p(\boldsymbol{\beta}, \sigma^2, \phi)$$

- MLE based on observed-data likelihood $L(\boldsymbol{\beta}, \sigma^2, \phi; \mathbf{Z}) =$

$$\int_{\mathbb{R}^n} \left\{ \prod_{i=1}^n f_{Z(\mathbf{s}_i) \mid W(\mathbf{s}_i)} \left(Z\left(\mathbf{s}_i\right) \mid \boldsymbol{\beta}, W\left(\mathbf{s}_i\right)\right) \right\} f_W\left(\mathbf{W} \mid \sigma^2, \phi\right) d\mathbf{W}.$$

Let $\theta = (\boldsymbol{\beta}, \sigma^2, \phi)$

- $\mathbf{W}$ has same dimensions as data
- Strong cross-correlations among $\mathbf{W}$ (cf. Christensen, 2004) and between $\mathbf{W}$ and $\boldsymbol{\beta}$ (Reich et al., 2006)

Rich Literature on Fast Computing

- Many ideas, most designed for linear spatial models
  - Multiresolution methods, with parallelizations (Katzfuss, 2017; Katzfuss and Hammerling, 2014)
  - Nearest neighbor process (Datta et al., 2016)
  - Random projections (Banerjee, A., Tokdar, Dunson, 2013)
  - Lattice kriging (Nychka et al., 2010)
- A few approaches that work well for SGLMMs:
  - Predictive process (Banerjee, Gelfand, Finley, Sang 2008)
    - Works well, very general. *Finding knots etc. is challenging.*
  - Stochastic PDEs + INLA (Lindgren et al., 2011)
    - Fast! Approximately integrates out **W**, numerical integration
    - Some models missing, e.g. ordinal, multivariate spatial; data model involving numerical model (not in closed form)

Outline of Projection-based Approach

(Banerjee, Tokdar, Dunson, 2011; Guan and Haran, 2018)

1. Fast approximation to the principal components of Σ_ϕ
2. Reduce dimensions: replace **W** with $UD_m^{1/2}\boldsymbol{\delta}$. $\boldsymbol{\delta}$: much fewer dimensions and less correlated components
3. Project $UD_m^{1/2}\boldsymbol{\delta}$ to $C^\perp(X)$
   - Makes random effects orthogonal to fixed effects
4. Reduced model can be fit
   - In a Bayesian framework as part of an MCMC algorithm
   - ML framework using Monte Carlo expectation-maximization (MC-EM) or Laplace approximation EM (LA-EM)

Analogous to Projection for Latent GMRFs

Projection-based approach for latent GP models is analogous to Hughes and Haran (2013) where Moran's projection was used for latent GMRF models:

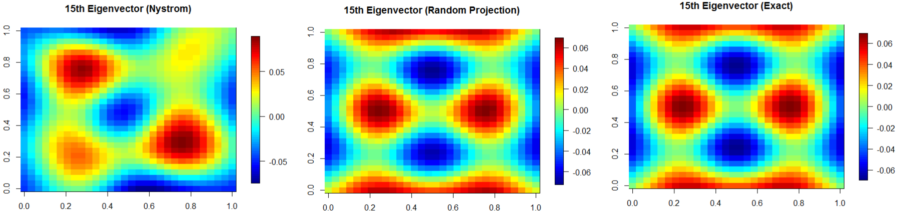
Step 1 Details

For speed we use a fast *approximate* eigendecomposition

Left: deterministic approximation

Center: **random approximation**

Right: exact eigendecomposition



- **Random projections** used in Banerjee, Tokdar, Dunson (2013); also Sarlos (2006), Halko et al. (2009)

Step 2: Reducing Dimensions via Projection

- Approximates the leading m eigencomponents of the covariance matrix Σ_ϕ
- **Replace W with** $UD_m^{1/2}\boldsymbol{\delta}$

Bayesian Inference: MCMC for SGLMMs

- Markov chain is slow mixing (need longer Markov chain) due to strong cross-correlations among **W**
- Let Θ=covariance/precision parameters; $\boldsymbol{\beta}$=regression
- Block updating schemes may help. E.g. blocks:

$$\boxed{\pi(\mathbf{W} \mid \Theta, \boldsymbol{\beta}, \mathbf{Z})} \quad \boxed{\pi(\Theta \mid \boldsymbol{\beta}, \mathbf{W}, \mathbf{Z})} \quad \boxed{\pi(\boldsymbol{\beta} \mid \Theta, \mathbf{W}, \mathbf{Z})}$$

  - Challenging to obtain good proposals for **W**, especially for high-dimensions
  - Computationally expensive per update

See Rue and Held (2005), Christensen et al. (2006), Haran and Tierney (2012)

They do not scale well (when $N > 1000$)

Bayesian Inference: Projected MCMC for SGLMMs

- Block updating scheme with new blocks:

$$\boxed{\pi(\boldsymbol{\delta} \mid \Theta, \boldsymbol{\beta}, \mathbf{Z})} \quad \boxed{\pi(\Theta \mid \boldsymbol{\beta}, \mathbf{W}, \mathbf{Z})} \quad \boxed{\pi(\boldsymbol{\beta} \mid \Theta, \mathbf{W}, \mathbf{Z})}$$

  - $\boldsymbol{\delta}$ is low-dimensional, e.g. 50 for **W** of 3,000 dimensions
  - $\boldsymbol{\delta}$ has low cross correlations and low correlation with X
  - Hence, easy to construct block updates for $\boldsymbol{\delta}$

ML Inference

Question: Why would anyone do this???

ML Inference

Question: Why would anyone do this???

- Providing priors can be a nuisance, e.g. we often do not have good prior information about covariance function parameters; this sidesteps the issue

ML Inference

Question: Why would anyone do this???

- Providing priors can be a nuisance, e.g. we often do not have good prior information about covariance function parameters; this sidesteps the issue
- Some people want to fit their models using ML (not an option for most moderate/large data sets)

ML Inference

Question: Why would anyone do this???

- Providing priors can be a nuisance, e.g. we often do not have good prior information about covariance function parameters; this sidesteps the issue
- Some people want to fit their models using ML (not an option for most moderate/large data sets)
- At least for this narrow class of problems, it was faster than doing Bayes

ML Inference

Question: Why would anyone do this???

- Providing priors can be a nuisance, e.g. we often do not have good prior information about covariance function parameters; this sidesteps the issue
- Some people want to fit their models using ML (not an option for most moderate/large data sets)
- At least for this narrow class of problems, it was faster than doing Bayes
- May provide useful insights on MCEM for other latent variable problems

ML Inference: Projected Monte Carlo EM

- Let $\psi = (\beta, \theta)$, so $f_{Z,\delta}(\mathbf{Z}, \delta; \psi)$ is complete data likelihood

  At t^{th} iteration,

  E-step:

  (a) Simulation step: obtain an MCMC sample $\delta^{(t,1)}, \ldots, \delta^{(t,k_t)}$ from $f_{\delta|Z}(\delta|\mathbf{Z}, \psi^{(t)})$ under current estimate, $\psi^{(t)}$

  (b) Monte Carlo integration step:

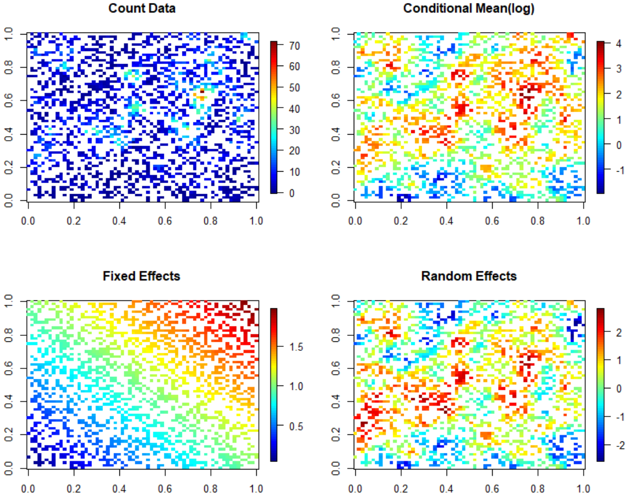
$$Q(\psi, \psi^{(t)}) = E[\ln f_{Z,\delta}(\mathbf{Z}, \delta; \psi)|\mathbf{Z}, \psi^{(t)}]$$
$$\approx \frac{1}{k_t} \sum_{k=1}^{k_t} \ln f_{Z,\delta}(\mathbf{Z}, \delta^{(t,k)}; \psi) = \hat{Q}(\psi, \psi^{(t)}).$$

  M-step: find $\psi^{(t+1)}$ to increase $\hat{Q}$ using Newton-Raphson.

- Stop when increase in Q-function $<$ threshold

Faster: replace Monte Carlo with Laplace approximation

A Simulated Example



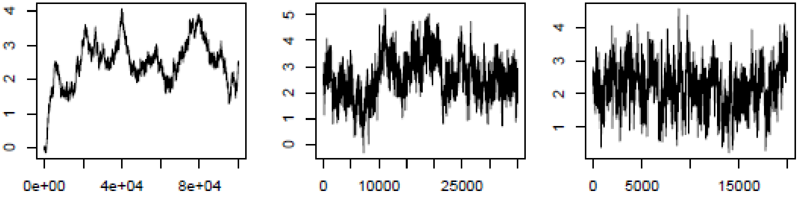
- Matern covariance function with ν= 1.5 and effective range 0.2

- Covariates are the coordinates with β_1=β_2= 1

Results: Rank Comparison

	LA-EM				MCMC-EM			
Rank	β_1=1	β_2=1	$\frac{\tau}{\phi}$=13.7	MSE	β_1=1	β_2=1	$\frac{\tau}{\phi}$=13.7	MSE
40	1.24	1.69	8.34	47.87	1.22	1.71	7.83	65.90
50	1.21	1.65	10.11	66.79	1.20	1.65	9.72	40.10
70	1.14	1.66	10.64	121.21	1.13	1.68	11.71	41.16
90	1.09	1.63	17.01	93.77	1.11	1.62	12.24	24.45
110	1.08	1.61	17.06	75.38	1.08	1.62	15.85	14.07

MCMC-EM Results: Mixing



Trace plots of a random effect for 3 consecutive EM iterations.

- Typically mixing improves
- The first EM iteration uses more sampling effort
- Monte Carlo sample size reduces substantially

The Last Slide

- Described projection-based approach to fit latent Gaussian random field model
- automated approach for
  - Bayesian inference
  - ML inference
- projections apply to
  - latent GMRF models (Moran projections)
  - latent GP models (random projections)
- The approach
  - is fast and automated, good approximations
  - reduces the dimensions of the posterior distribution/integral
  - reparameterization improves mixing of the MCMC algorithm
  - able to adjust for spatial confounding

Laplace Approximation EM

(Eidsvik et al., 2010; Bonat and Ribeiro, 2016)

Obtain MCMC samples $\delta^{(t,1)}, ...\delta^{(t,k_t)}$ from the conditional distribution $f_{\delta|Z}(\delta \mid \mathbf{Z}, \mathbf{s}\psi^{(t)}) \propto f(Z|M_{\phi^{(t)}}\delta, \mathbf{s}\beta^{(t)})f(\delta|\theta^{(t)})$

$\Longrightarrow$ Approximate the conditional distribution with a Gaussian distribution, $f_{\delta|Z}(\delta \mid \mathbf{Z}, \mathbf{s}\psi^{(t)}) \approx f_G(\delta \mid \mathbf{Z}, \mathbf{s}\psi^{(t)})$, and obtain the mode δ^* of $f_G(\delta \mid \mathbf{Z}, \mathbf{s}\psi^{(t)})$

Approximate expectation $E\left[h(\delta) \mid \mathbf{Z}, \mathbf{s}\psi^{(t)}\right]$ with MC averages:

$\Longrightarrow$ Approximate expectation by (1) Taylor expanding $h(\delta)$ around δ^*, then (2) take expectation w.r.t f_G: