Theory and Computations Behind the spBayes Package

Jim Faulkner

Space-Time Reading Group
University of Washington

Oct 24, 2017

Outline

1. Available models
   - Full-rank univariate Gaussian spatial regression models
   - Low-rank predictive process models
   - Multivariate Gaussian spatial regression models
   - Non-Gaussian models
   - Dynamic spatio-temporal models
2. Bayesian inference and MCMC
3. Computations

Univariate Gaussian spatial regression

- Assume we have observations $y(\mathbf{s})$ measured at set of n spatial locations $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, where $\mathbf{s} \in \mathcal{D} \subseteq \mathbb{R}^2$ and

$$y(\mathbf{s}_i) = \boldsymbol{x}(\mathbf{s}_i)^T \boldsymbol{\beta} + w(\mathbf{s}_i) + \epsilon(\mathbf{s}_i)$$

- $\boldsymbol{x}(\mathbf{s})$ is a set of spatially referenced covariates
- $w(\mathbf{s})$ is a spatial Gaussian process with mean zero and covariance function $C(\mathbf{s}, \mathbf{t}|\boldsymbol{\theta})$. For the set of n observation locations, $\boldsymbol{\alpha} = (w(\mathbf{s}_1), \ldots, w(\mathbf{s}_n))^T$ and $\mathbf{K}(\boldsymbol{\theta})$ is its $n \times n$ covariance matrix.
- $\epsilon(\mathbf{s})$ is an independent white-noise process to capture measurement error and micro-scale variation, where $\epsilon(\mathbf{s}_i) \sim \mathrm{N}(0, \tau^2)$

Low-rank predictive process models

For large data sets, using a low-rank representation of the spatial field will speed computations.

- Consider a set of r knots $\mathcal{S}^* = \{\mathbf{s}_1^*, \ldots, \mathbf{s}_r^*\}$, which may but need not be a set of the entire collection of observed locations in $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, but $r \ll n$.
- Assume that $w(\mathbf{s}) \sim \mathrm{GP}(0, C(\cdot|\boldsymbol{\theta}))$ and let $\boldsymbol{w}^*$ be a realization of $w(\mathbf{s})$ over $\mathcal{S}^*$. That is, $\boldsymbol{w}^* \sim \mathrm{N}(\mathbf{0}, \boldsymbol{C}^*(\boldsymbol{\theta}))$
- Define the *predictive process* as

$$\tilde{w}(\mathbf{s}) = \mathrm{E}\left[w(\mathbf{s}) \mid w(\mathbf{s}_i^*),\, i = 1, 2, \ldots, r\right]$$

- The realizations of $\tilde{w}(\mathbf{s})$ are the kriged predictions conditional on a realization of $w(\mathbf{s})$ over $\mathcal{S}^*$

Multivariate Gaussian spatial regression

- Consider case where m point-referenced outcomes are measured at each location $\mathbf{s}_i$ and regressed on a known set of predictors. Then for $j = 1, 2, \ldots, m$

$$y_j(\mathbf{s}_i) = \boldsymbol{x}_j(\mathbf{s}_i)^T \boldsymbol{\beta}_j + w_j(\mathbf{s}_i) + \epsilon_j(\mathbf{s}_i)$$

- where $\boldsymbol{\epsilon}(\mathbf{s}) = (\epsilon_1(\mathbf{s}), \epsilon_2(\mathbf{s}), \ldots, \epsilon_m(\mathbf{s}))^T \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Psi})$,
- spatial variation follows a $m \times 1$ GP: $\boldsymbol{w}(\mathbf{s}) = (w_1(\mathbf{s}), w_2(\mathbf{s}), \ldots, w_m(\mathbf{s})) \sim \mathrm{GP}(\mathbf{0}, \boldsymbol{C}_w(\mathbf{s}, \mathbf{t}))$
- $\boldsymbol{C}_w(\mathbf{s}, \mathbf{t})$ is a cross-covariance matrix with entries being covariance between $w_j(\mathbf{s})$ and $w_k(\mathbf{t})$.
- Use linear model of coregionalization to specify cross covariance, which assumes $\boldsymbol{C}_w(\mathbf{s}, \mathbf{t}) = \boldsymbol{A}\boldsymbol{M}(\mathbf{s}, \mathbf{t})\boldsymbol{A}$, where $\boldsymbol{A}$ is $m \times m$ lower triangular and $\boldsymbol{M}(\mathbf{s}, \mathbf{t})$ is $m \times m$ diagonal with correlation functions as entries.

Non-Gaussian models

- Spatial GLMs, where dependent (response) variables are non-Gaussian.
- Replace the Gaussian likelihood with the assumption that $\mathrm{E}[y(\mathbf{s})]$ is linear on a transformed scale. That is,

$$\begin{aligned} \eta(\mathbf{s}_i) &\equiv g(\mathrm{E}[y(\mathbf{s}_i)]) \\ &= \boldsymbol{x}(\mathbf{s}_i)^T\boldsymbol{\beta} + w(\mathbf{s}_i) \end{aligned}$$

  where $g(\cdot)$ is a suitable link function.
- spBayes also provides low-rank predictive process models for spatial GLMs.

Dynamic spatio-temporal models

- Assume continuous space and discrete time.
- The model equations are

$$y_t(\mathbf{s}_i) = \boldsymbol{x}_t(\mathbf{s}_i)^T \boldsymbol{\beta}_t + u_t(\mathbf{s}_i) + \epsilon_t(\mathbf{s}_i)$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t$$
$$u_t(\mathbf{s}_i) = u_{t-1}(\mathbf{s}_i) + w_t(\mathbf{s}_i)$$

- where

$$\epsilon_t(\mathbf{s}) \sim \mathrm{N}(0, \tau_t^2)$$
$$\boldsymbol{\eta}_t \sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$$
$$w_t(\mathbf{s}) \sim \mathrm{GP}(\mathbf{0}, C_t(\cdot|\boldsymbol{\theta}_t))$$

- and $\boldsymbol{\beta}_0 \sim \mathrm{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and $u_0(\mathbf{s}) \equiv 0$

Bayesian inference

- Given a likelihood $p(y \mid \boldsymbol{\theta})$ and a prior $p(\boldsymbol{\theta})$ we can use the relationship $p(y, \boldsymbol{\theta}) = p(y \mid \boldsymbol{\theta})p(\boldsymbol{\theta})$ and Bayes Rule to find the posterior distribution $p(\boldsymbol{\theta} \mid y)$.
- By Bayes Rule the posterior distribution is:

$$\begin{aligned} p(\boldsymbol{\theta} \mid y) &= \frac{p(y \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(y)} \\ &= \frac{p(y \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(y \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \end{aligned}$$

- The integral in the denominator is almost always intractable, but we don't need to worry about it. We can use Markov chain Monte Carlo (MCMC) to sample from the posterior even if we do not know the exact form of the posterior distribution.

MCMC: Metropolis-Hastings

- We can use a Markov chain with stationary distribution equal to the posterior distribution to generate samples from the posterior, where the samples are a sequence of draws $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(k)}$
- One way to do this is with the Metropolis-Hastings algorithm:
  1. Start with arbitrary starting value $\theta^{(0)}$ where $p(\theta^{(0)} \mid y) > 0$. Set $k = 1$
  2. Draw proposed value θ^* from proposal density $q(\theta^*|\theta^{(k-1)})$.
  3. Set $\theta^{(k)} = \theta^*$ with probability

     $$\alpha = \min\left\{1, \frac{p(\theta^*|y)}{p(\theta^{(k-1)}|y)} \frac{q(\theta^{(k-1)}|\theta^*)}{q(\theta^*|\theta^{(k-1)})}\right\};$$

     otherwise set $\theta^{(k)} = \theta^{(k-1)}$.
  4. Set $k = k + 1$ and go back to 2.

MCMC: Gibbs Sampling

- If the complete conditional marginal distributions have known forms we can use Gibbs sampling.
- Gibbs Sampler:
  1. Set arbitrary starting values $\left\{\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_m^{(0)}\right\}$ and set $k = 1$.
  2. Draw $\theta_1^{(k)})$ from $p(\theta_1|\theta_2^{(k-1)}, \theta_3^{(k-1)}, \ldots, \theta_m^{(k-1)})$
  3. Draw $\theta_{(2)}^{k}$ from $p(\theta_2|\theta_1^{(k)}, \theta_3^{(k-1)}, \ldots, \theta_m^{(k-1)})$
  4. Draw $\theta_{(3)}^{k}$ from $p(\theta_3|\theta_1^{(k)}, \theta_2^{(k)}, \ldots, \theta_m^{(k-1)})$

     $\vdots$

  m. Draw $\theta_{(m)}^{k}$ from $p(\theta_m|\theta_1^{(k)}, \theta_2^{(k)}, \ldots, \theta_{m-1}^{(k)})$

General model description

Assuming normally-distributed observations, the posterior of many of the models can be written:

$$\mathrm{N}(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}(\boldsymbol{\theta})\boldsymbol{\alpha}, \boldsymbol{D}(\boldsymbol{\theta})) \times \mathrm{N}(\boldsymbol{\beta}|\boldsymbol{\mu}_{\beta}, \boldsymbol{\Sigma}_{\beta}) \times \mathrm{N}(\boldsymbol{\alpha}|\mathbf{0}, \boldsymbol{K}(\boldsymbol{\theta})) \times p(\boldsymbol{\theta})$$

- where $\boldsymbol{K}(\boldsymbol{\theta})$ and $\boldsymbol{D}(\boldsymbol{\theta})$ are families of $r \times r$ and $n \times n$ covariance matrices, respectively, and
- $\boldsymbol{Z}(\boldsymbol{\theta})$ is $n \times r$ with $r \ll n$, and $\boldsymbol{\theta}$ are unknown process parameters
- Assume that $\boldsymbol{\mu}_{\beta}$ and $\boldsymbol{\Sigma}_{\beta}$ are known hyperparameters
- Inference carried out by sampling from posterior of $\{\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\theta}\}$

Sampling process parameters

- First integrate $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ from the model and sample $\boldsymbol{\theta}$ from

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto \mathrm{N}(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\mu}_{\beta}, \boldsymbol{\Sigma}_{y|\theta}) \times p(\boldsymbol{\theta})$$

- where $\boldsymbol{\Sigma}_{y|\theta} = \boldsymbol{X}\boldsymbol{\Sigma}_{\beta}\boldsymbol{X}^T + \boldsymbol{Z}(\boldsymbol{\theta})\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{Z}(\boldsymbol{\theta})^T + \boldsymbol{D}(\boldsymbol{\theta})$
- Use Gibbs sampling and random walk Metropolis
- Avoid calculating inverses by using Cholesky factorizations and solving triangular systems. Use matrix-vector multiplications and sparse matrix calculations where possible.

Sampling slope and random effects

- Once have marginal posterior samples of $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\boldsymbol{y})$, draw $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ using composition sampling.
- Suppose $\left\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots \boldsymbol{\theta}^{(M)}\right\}$ are M samples from $p(\boldsymbol{\theta}|\boldsymbol{y})$
- Drawing $\boldsymbol{\beta}^{(k)} \sim p(\boldsymbol{\beta}|\boldsymbol{\theta}^{(k)}, \boldsymbol{y})$ and $\boldsymbol{\alpha}^{(k)} \sim p(\boldsymbol{\alpha}|\boldsymbol{\theta}^{(k)}, \boldsymbol{y})$ for $k = 1, 2, \ldots, M$ results in M samples from $p(\boldsymbol{\beta}|\boldsymbol{y})$ and $p(\boldsymbol{\alpha}|\boldsymbol{y})$
- Uses Gibbs sampling
- Use efficient matrix calculations and other tricks to simplify computations
- Computations can become prohibitive for large n

Low rank models

- To reduce computational burden of large n, specify $\boldsymbol{Z}(\boldsymbol{\theta})$ with $r \ll n$ in predictive process model.
- First integrate $\boldsymbol{\alpha}$ from the posterior to get

$$p(\boldsymbol{\beta}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \mathrm{N}(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\mu}_{\beta}, \boldsymbol{\Sigma}_{y|\beta,\theta}) \times \mathrm{N}(\boldsymbol{\beta}|\boldsymbol{\mu}_{\beta}, \boldsymbol{\Sigma}_{\beta}) \times p(\boldsymbol{\theta})$$

- Use Gibbs sampling to alternately update $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$
- When have posterior samples for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, can now draw samples of $\boldsymbol{\alpha}$ from its full conditional distribution given both $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ with methods similar to those for full rank model.

References

- Finley, A., S. Banerjee, and B. Carlin. 2007. spBayes: an R package for univariate and multivariate point-referenced spatial models. Journal of Statistical Software 19(4):1-24.
- Finley, A.O., S. Banerjee, and A. E. Gelfand. 2015. spBayes for large univariate and multivariate point-referenced spatio-temporal data models. Journal of Statistical Software 63(13):1–28.
- Banerjee, S., B. P. Carlin, and A. E. Gelfand. 2015. Hierarchical modeling and analysis for spatial data. Second edition. CRC Press, Taylor & Francis Group.