STAT 8025

Lecture 10: Multivariate Models

Dr. Emily Lei Kang

Division of Statistics & Data Science
Department of Mathematical Sciences
University of Cincinnati

- We have been focused on a *single* process, $Y(s)$. Recall that we introduce spatial dependence structure so that we can borrow strength in space for improved inference
- We often also need to borrow strength across *variables*, by considering multivariate models:
  - Example: PM2.5 and ozone
- We can build and fit a multivariate model even if the responses are misaligned (e.g., at s observing PM2.5 at s but not ozone
  - introducing dependence across process and space
  - improving predictions

Conditional Approach

- We can always write joint models into a sequence of conditional models (recall NNGP, Vecchia)
- We can use this conditional approach to build the joint models for $Y_1(\mathrm{s})$ and $Y_2(\mathrm{s})$, a bivariate process

$$Y_1(\mathrm{s}) = w_1(\mathrm{s})$$
$$Y_2(\mathrm{s}) = \rho Y_1(\mathrm{s}) + w_2(\mathrm{s})$$

  where we assume $w_1(\cdot)$ and $w_2(\cdot)$ are independent GPs with $Cov(w_i(\mathrm{s}), w_i(\mathrm{u})) = C_i(\mathrm{s}, \mathrm{u})$.
- If $\rho = 0$, then $Y_1(\cdot)$ and $Y_2(\cdot)$ are independent.
- A similar framework is used in multi-fidelity computer experiments, called the co-kriging model (Kennedy and O'Hagan, 2000) - computer experiments and spatial statistics are closely related.

- Induced cross covariance is:

$$\begin{aligned} & Cov(Y_1(s), Y_2(s)) \\ = \; & Cov(w_1(s), \rho w_1(s) + w_2(s)) \\ = \; & \rho Var(w_1(s)) \end{aligned}$$

$$\begin{aligned} & Cov(Y_1(s), Y_2(u)) \\ = \; & Cov(w_1(s), \rho w_1(u) + w_2(u)) \\ = \; & \rho Cov(w_1(s), w_1(u)) \\ = \; & \rho C_1(s, u) \end{aligned}$$

Inference

- MLE: Straightforward if we observe both $Y_1(\cdot)$ and $Y_2(\cdot)$ at all observation locations. If not (i.e., misalignment), tricky... EM?
- Bayesian: Convenient, as we can impute missing values in MCMC iterations

Reading assignment: Kennedy, M. C. and O'Hagan, A. (2000), Predicting the output from a complex computercode when fast approximations are available, *Biometrika*, 87, 1-13.

$$Z_t(x) = \rho_{t-1} Z_{t-1}(x) + \delta_t(x), \; t = 2, \ldots, s$$

where

- ρ_{t-1} is kind of regression parameter
- $\delta_t(\cdot)$ is independent of $Z_{t-1}(\cdot), \ldots, Z_1(\cdot)$
- $Z_1(\cdot)$ is a stationary GP independent of $\delta_t(\cdot)$.
- $Z_1(\cdot)$, $\delta_t(\cdot)$, $t = 2, \ldots, s$ are assumed to be independent GP with covariance function

$$C_t(x, x') = \sigma_t^2 \exp\left\{-b_t(x - x')^T (x - x')\right\}$$

Reading assignment: Noel Cressie, Andrew Zammit-Mangion (2016), Multivariate spatial covariance models: a conditional approach, *Biometrika*, 103, 915-935.

- Convolutional type conditional approach:

$$E(Y_2(\mathrm{s})|Y_1(\cdot)) = \int_D b(\mathrm{s},\mathrm{u})Y_1(\mathrm{u})d\mathrm{u}$$

$$Cov(Y_2(\mathrm{s}), Y_2(\mathrm{u})|Y_1(\cdot)) = C_{2|1}(\mathrm{s},\mathrm{u})$$

- Example: $b(\mathrm{s},\mathrm{u}) = b(h)$ with $h = \mathrm{u} - \mathrm{s}$
  - independent: $b(h) = 0$
  - pointwise dependence: $b(h) = AI(h = 0)$
  - diffused dependence: $b(h) = A\left\{1 - (\frac{\|h\|}{r})^2\right\}^2$ for $\|h\| \leq r$
  - asymmetric dependence: : $b(h) = A\left\{1 - (\frac{\|h-\Delta\|}{r})^2\right\}^2$ for $\|h - \Delta\| \leq r$
- Focusing on modeling but not focusing on inference (e.g., not tackling misaligned data)

Separable Model

- The separable model is

$$Cov(Y_i(s), Y_j(u)) = \sigma_{ij} Corr(s, u)$$

- This implies:
  - Variance: $Var(Y_i(s)) = \sigma_{ii}$
  - Cross covariance:

$$Cov(Y_i(s), Y_j(s)) = \sigma_{ij}$$

$$Cov(Y_i(s), Y_i(u)) = \sigma_{ii} Corr(s, u)$$

$$Cov(Y_i(s), Y_j(u)) = \sigma_{ij} Corr(s, u)$$

- Cross covariance between responses is the same across locations; spatial correlation is the same for all responses.

- Consider q-variate process. Suppose we observe it at n locations (no missing, no misalignment). We can show that the covariance matrix of the nq-dimensional observation vector Y is

$$\Sigma_s \otimes \Sigma_y$$

  where Σ_s is $n \times n$ and Σ_y is $q \times q$.
- As with the separable spatio-temporal models, this Kronecker structure can lead to computational simplification.

Linear Model of Coregionalization

- This is like factor analysis for spatial data.
- Consider a q-variate process. We assume that we can represent the q responses through L latent (unobserved) processes:

$$Y_i(s) = \sum_{l=1}^{L} A_{il} f_l(\mathrm{s}) + \epsilon_i(\mathrm{s})$$

  - $\epsilon_i(\cdot)$ iid $N(0, \sigma_i^2)$. We need this to ensure positive definiteness.
- Then for $\mathrm{Y}(\mathrm{s}) = (Y_1(\mathrm{s}), \ldots, Y_q(\mathrm{s}))'$, we have

$$\mathrm{Y}(\mathrm{s}) \sim N(\mathrm{A}f(\mathrm{s}), \mathrm{D})$$

  - A is $q \times L$
  - $f(\mathrm{s}) = (f_1(\mathrm{s}), \ldots, f_L(\mathrm{s}))'$ and we assume $f_i(\cdot)$ are independent GPs
  - $\mathrm{D} = diag(\sigma_1^2, \ldots, \sigma_q^2)$

Identifiability

- A lower trapezoidal matrix with positive diagonal elements:

$$A_{ij} = 0 \text{ if } i < j$$

$$A_{ii} > 0$$

- $f_l(\cdot)$ is GP with variance 1
- Prior, and posterior inference through MCMC

- Suppose $f_l(\cdot)$, $l = 1, \ldots, L$ are GP with mean 0 and variance 1 and common Correlation function $c(\mathrm{s}, \mathrm{u}; \boldsymbol{\theta})$ will result in a separable model. [Try verify this after class!]
- If $f_l(\cdot)$ is with correlation function $c(\mathrm{s}, \mathrm{u}; \boldsymbol{\theta}_l)$, then it is non-separable.
- Nonstationary? allowing A_{ij} to vary across space?

Reading assignment: Zhang, L, Banerjee, S. Spatial factor modeling: A Bayesian matrix-normal approach for misaligned data. Biometrics. 2021; 1-14. https://doi.org/10.1111/biom.13452

- To be presented as one of the final projects

Remarks

- Spatial → multivariate spatial is a typical research route, as spatial → spatio-temporal, and multivariate spatio-temporal further
- How about a large number of responses? Rather an open problem until a few recent investigations:
  - Krock, M., Kleiber, W., Hammerling, D., & Becker, S. (2021). Modeling massive multivariate spatial data with the basis graphical lasso. arXiv preprint arXiv:2101.02404. https://arxiv.org/pdf/2101.02404.pdf
  - Dey et al. (2022) Graphical Gaussian process models for highly multivariate spatial data, Biometrika, 109, 993–1014.

Summary

- Multivariate models

Preview:

- Areal data and models