Bayesian Multivariate Hierarchical Semiparametric Mixed model with Gaussian Process Priors

Beomjo Park
beomjo@korea.ac.kr

Dept. of Statistics, Korea University

Nov. 11th, 2017

Table of Contents

Motivation

- Multivariate semiparametric regression with Dirichlet process mixture of Ornstein-Uhlenbeck (OU) process error
- Hierarchical representation captures possible similarities across different groups while allowing each group to have distinct smooth curve or surface representation.
- Tensor product cosine basis to relax additivity and linearity assumption on nonparametric regression.

Literature Review

- Lenk (1999), Lenk and Choi (2017) proposed semiparametric regression models (BSAR) with Fourier cosine basis representation. For $x \in [0, 1]$,

$$f(x) = \sum_{j=0}^{\infty} \theta_j \phi_j(x), \qquad \phi_0(x) = 1, \ \phi_j(x) = \sqrt{2}\cos(\pi j x)$$

$$\theta_j \mid \tau, \gamma \sim \mathrm{N}\left(0, \tau^2 \exp[-j\gamma]\right), \qquad j \geq 1, \gamma > 0$$

- Quintana *et al.* (2016) proposed DPM over OU process under univariate linear mixed model.

- Rosen and Thompson (2009) proposed multivariate semiparametric model with OU process error on functional data analysis.

Semiparametric Mixed BSAM

Consider the following semiparametric mixed model :

$$\boldsymbol{Y}_i = \boldsymbol{W}_i\boldsymbol{B} + f(\boldsymbol{X}_i) + \boldsymbol{U}_i\boldsymbol{R}_i + \boldsymbol{Z}_i \qquad (i = 1, \cdots, n)$$

- Response : $\boldsymbol{Y}_i = (Y_{i1}, \cdots, Y_{in_i})^\top \in \mathbb{R}^{n_i \times L}$ for each individual $i \in \{1, \cdots, n\}$ at times $\{t_{i1}, \ldots, t_{in_i}\}$
- Linear effect covariate : $\boldsymbol{W}_i = (\boldsymbol{w}_{i1}, \boldsymbol{w}_{i2}, \ldots, \boldsymbol{w}_{in_i})^\top \in \mathbb{R}^{n_i \times p}$
- Nonparametric covariate : $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{in_i})^\top \in [0,1]^{n_i \times q}$
- Random effect covariate : $\boldsymbol{U}_i \in \mathbb{R}^{n_i \times r}$
- Error process : Multivariate OU process with zero mean

$$d\boldsymbol{Z}_{it} = -A\boldsymbol{Z}_{it} + Bd\boldsymbol{W}_{it}$$

where $\boldsymbol{W}_{it}$ is L-dim Wiener process.

Tensor product basis

Let $\phi_0, \phi_1, \ldots$ be an orthonormal basis for $L^2([0,1])$

$$\mathrm{Span}\left(\{\phi_{j,k}(x_1,x_2) = \phi_j(x)\phi_k(x_2) \ : \ j,k = 0,1,\ldots\}\right) = L^2([0,1]^2)$$

Then,

$$\begin{aligned} f(x_1,x_2) &= \sum_{j,k=0}^{\infty} \theta_{j,k}\phi_j(x_1)\phi_k(x_2) \\ &\approx \sum_{j=0}^{J}\sum_{k=0}^{K} \theta_{j,k}\phi_j(x_1)\phi_k(x_2) = \mathrm{vec}(\boldsymbol{\varphi})^\top \mathrm{vec}(\boldsymbol{\Theta}) \end{aligned}$$

where $[\boldsymbol{\Theta}]_{(j,k)} = \theta_{j,k}$, $\boldsymbol{\varphi} = \boldsymbol{\phi}_J(x_1)\boldsymbol{\phi}_K(x_2)^\top \in \mathbb{R}^{(J+1)\times(K+1)}$

Multivariate Responses

Denote $\boldsymbol{Y} = [\boldsymbol{y}_1 \mid \boldsymbol{y}_2] \in \mathbb{R}^{n\times 2}$, $\boldsymbol{W} = [\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n]^\top \in \mathbb{R}^{n\times p}$, $\boldsymbol{B} = [\beta_1 \mid \beta_2] \in \mathbb{R}^{p\times 2}$, $[\boldsymbol{\Theta}_l]_{(j,k)} = \theta_{l,j,k}$, $\boldsymbol{\Xi} := [\mathrm{vec}(\boldsymbol{\Theta}_1) \mid \mathrm{vec}(\boldsymbol{\Theta}_2)] \in \mathbb{R}^{(J+1)(K+1)\times 2}$ and $\boldsymbol{\Phi} \in \mathbb{R}^{n\times(J+1)(K+1)}$ with each row consists of $\mathrm{vec}(\phi_J(x_{i1})\phi_K(x_{i2})^\top)$, that is, $[\boldsymbol{\Phi}]_{(i,(j,k))} = \phi_j(x_{i1})\phi_k(x_{i2})$.

MvBSAR

$$\boldsymbol{Y}_i = \boldsymbol{W}_i\boldsymbol{B} + \boldsymbol{\Phi}_i\boldsymbol{\Xi} + \boldsymbol{U}_i\boldsymbol{R}_i + \boldsymbol{Z}_i, \qquad \boldsymbol{Z}_i \sim (\boldsymbol{0}, \Sigma_Z)$$

$$\mathrm{vec}(\boldsymbol{\Xi}) \sim \mathcal{N}_{(J+1)(K+1)\times 2}\left(\boldsymbol{0}, \boldsymbol{V}_{0,\Xi} := \left[\begin{array}{c|c} \tau_1^2\Gamma_{12}\otimes\Gamma_{11} & \boldsymbol{0} \\ \hline \boldsymbol{0} & \tau_2^2\Gamma_{22}\otimes\Gamma_{21} \end{array}\right]\right)$$

$$\Leftrightarrow \mathrm{vec}(\boldsymbol{\Theta}_l) \sim \mathrm{N}_{(J+1)(K+1)}(\boldsymbol{0}, \tau_l^2\boldsymbol{\Gamma}_{l2}\otimes\boldsymbol{\Gamma}_{l1})$$

where $\boldsymbol{\Gamma}_{l1} := \mathrm{diag}(\exp(-j\gamma_{l1}))\,|_{j=0}^{J}, \boldsymbol{\Gamma}_{l2} := \mathrm{diag}(\exp(-k\gamma_{l2}))\,|_{k=0}^{K}$

Multivariate OU process

- Under stationarity condition, Σ_Z be a var-cov matrix of $\boldsymbol{Z}_t$ satisfying

$$A\Sigma_Z + \Sigma_Z A^\top = BB^\top$$

- Following Rosen and Thompson (2009) ensuring stationarity, we place prior on A and $C := BB^\top$:

$$\begin{aligned} S_{kk} &= 1, \quad S_{kl} \sim N(0, \sigma_a^2), (k \neq l), \quad \log(\Lambda_{A,kk}) \sim N(0, \sigma_a^2) \\ L_{kk} &= 1, \quad L_{kl} \sim N(0, \sigma_L^2), (k < l), \quad \log(D_{kk}) \sim N(0, \sigma_D^2) \end{aligned}$$

where $A = S\Lambda_A S^{-1}$: eigen decomposition and $C = LDL^\top$: cholesky decomposition.

Dirichlet process mixture of OU process

Denote $\zeta_i := \left[S^i_{kl}(k \neq l),\ \log(\Lambda^i_{A,kk}),\ L^i_{kl}(k < l),\ \log(D^i_{kk}) \right]$ for observation $i = 1, \cdots, n$.

$$\zeta_1, \cdots, \zeta_n \mid G \overset{ind}{\sim} G, \quad G \sim DP(\alpha, G_0)$$

$$G_0 = \prod_{k \neq l} \mathrm{N}(S_{kl}; 0, \sigma_a^2) \times \prod_k \mathrm{N}(\log(\Lambda_{A,kk}); 0, \sigma_a^2) \times \prod_{k<l} \mathrm{N}(L_{kl};\ 0, \sigma_L^2) \times \prod_k \mathrm{N}(\log(D_{kk});\ 0, \sigma_D^2)$$

where α be a dispersion parameter and G_0 be a base measure.

Prior Specification

- For covariance of random effects term Σ_R
  : Hierarchical Half-t prior (Huang *et al.*, 2013)

$$\Sigma_R \sim \mathrm{HIW}_{\mathrm{ht}}(\nu, \boldsymbol{\xi}) \ \Leftrightarrow\ \Sigma_R \mid \Lambda \sim \mathrm{IW}(\nu + 2 - 1,\ 2\nu\Lambda),\ \ \lambda_l \overset{ind}{\sim} \mathrm{Ga}\left(\frac{1}{2}, \frac{1}{\xi_l^2}\right)$$

with $\Lambda := \mathrm{diag}(\boldsymbol{\lambda})$. ($\nu = 2$ yields marginally uniform on correlation.)

- Conjugate prior and T-smoother

$$\begin{aligned}
\boldsymbol{\beta}_l &\sim \mathrm{N}\left(\boldsymbol{m}_{0,\beta}, \boldsymbol{V}_{0,\beta}\right) && (l = 1, 2) \\
\tau_l^2 &\sim \mathrm{IGa}\left(\frac{r_{0,\tau}}{2}, \frac{s_{0,\tau}}{2}\right) && (l = 1, 2) \\
\gamma_{lc} &\sim \mathrm{Exp}(w_0) && (l = 1, 2,\ c = 1, 2)
\end{aligned}$$

Hierarchical Modeling

Consider group assignment $\boldsymbol{g} = \{1, \cdots, g\}$, and corresponding subject assignment $i = 1, \cdots, n_{g_i}$

$$\boldsymbol{Y}_{gi} = \boldsymbol{W}_{gi}\boldsymbol{B}_g + \boldsymbol{\Phi}_{gi}\boldsymbol{\Xi}_g + \boldsymbol{U}_{gi}\boldsymbol{R}_{gi} + \boldsymbol{Z}_{gi}$$

Here we consider hierarchical structure with priors :

$$\begin{aligned}
\mathrm{vec}(\boldsymbol{B}_g) \mid \tilde{\boldsymbol{B}} &\sim \mathrm{N}\left(\tilde{\boldsymbol{B}}, \boldsymbol{V}_{0,\boldsymbol{B}}\right), && \forall g \in \{1, \cdots, g\} \\
\mathrm{vec}(\boldsymbol{\Xi}_g) \mid \tilde{\boldsymbol{\Xi}} &\sim \mathrm{N}\left(\tilde{\boldsymbol{\Xi}}, \boldsymbol{V}_{0,\boldsymbol{\Xi}_g}\right), && \forall g \in \{1, \cdots, g\} \\
\tilde{\boldsymbol{B}} &\sim \mathrm{N}\left(\boldsymbol{m}_{0,\tilde{\boldsymbol{B}}}, \boldsymbol{V}_{0,\tilde{\boldsymbol{B}}}\right), && \tilde{\boldsymbol{\Xi}} \sim \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{V}_{0,\tilde{\boldsymbol{\Xi}}}\right); \\
\tau^2_{lg} &\overset{ind}{\sim} \mathrm{IGa}\left(\frac{r_{0,\tau}}{2}, \frac{s_{0,\tau}}{2}\right), && \gamma_{glc} \overset{ind}{\sim} \mathrm{Exp}(w_0)
\end{aligned}$$

where

$$\boldsymbol{V}_{0,\boldsymbol{\Xi}_g} := \left[\begin{array}{c|c} \tau^2_{g1}\Gamma_{g12} \otimes \Gamma_{g11} & \boldsymbol{0} \\ \hline \boldsymbol{0} & \tau^2_{g2}\Gamma_{g22} \otimes \Gamma_{g21} \end{array}\right], \quad \boldsymbol{V}_{0,\tilde{\boldsymbol{\Xi}}} = \left[\begin{array}{c|c} \boldsymbol{V}_{0,\tilde{\boldsymbol{\Xi}}_1} & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{V}_{0,\tilde{\boldsymbol{\Xi}}_2} \end{array}\right]$$

$$\boldsymbol{V}_{0,\tilde{\boldsymbol{\Xi}}_l} = \mathrm{diag}\left(\frac{w_0}{j + w_0}\right)\Bigg|_{j=0}^{J} \otimes \mathrm{diag}\left(\frac{w_0}{k + w_0}\right)\Bigg|_{k=0}^{K}$$

Posterior Inference

- For τ_{lg}, $\boldsymbol{\lambda}$, Conjugate prior $\Rightarrow$ Inverse Gamma
- For Σ_R, Conjugate prior $\Rightarrow$ Inverse Wishart
- For $\tilde{\boldsymbol{B}}, \boldsymbol{B}_g$, $\tilde{\Xi}, \Xi_g$, $\boldsymbol{R}_{gi}$, Conjugate prior $\Rightarrow$ Normal
- For γ_{glc}, Slice sampling

$$f(\gamma_{glc} \mid \text{Rest}) \propto \exp\left(w_J \gamma_{glc} - \sum_{j=1}^{J} c_{js} \exp(j\gamma_{glc}) \right)$$

- For ζ
  - DP sampling $\Rightarrow$ Algorithm 8 of Neal (2000)
  - Resampling (Bush and MacEachern, 1996) $\Rightarrow$ Random Walk MH with Normal proposal with numerically obtained negative inverse of Hessian as variance covariance matrix.

Empirical Studies : Simulation

Settings

- Bivariate response ($L = 2$) and bivariate smoother with $(5, 5)$ bases
- 100 subjects in total of $2,000$ obs. randomly split into 2 groups
- Time $t_{ij} \sim \mathrm{Unif}[0, n_i]$ for each subjects i
- Linear components $w_1, w_2 \sim \mathrm{Unif}[0, 1]$
- Nonparametric components $x_1, x_2 \sim \mathrm{Unif}[-1, 1]$
- Random intercept ($q = 1$) with $\sigma_R = 2$
- For error distribution, we generated from the two processes and assign randomly among 100 subjects.

$$\boldsymbol{Z}_{it} = -\begin{pmatrix} 2 & -0.6 \\ -0.6 & 2 \end{pmatrix} \boldsymbol{Z}_{it} + \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} \mathrm{d}\boldsymbol{W}_{it}$$

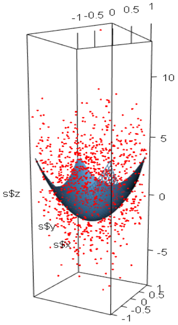
and

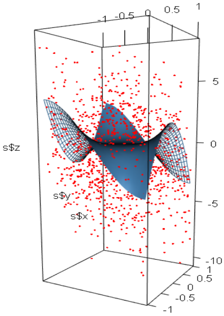
$$\boldsymbol{Z}_{it} = -\begin{pmatrix} 3 & -2 \\ -0.1 & 3 \end{pmatrix} \boldsymbol{Z}_{it} + \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} \mathrm{d}\boldsymbol{W}_{it}$$

Empirical Studies : Simulation

For the first group

$$y_{1it} = [-2w_{1it} - 5w_{2it}] + [3(x_{1it}^2 + x_{2it}^2) - 2\ ] + r_i + z_{1it}$$
$$y_{2it} = [\quad 4w_{1it} + 3w_{2it}] + [2(x_{1it}^3 - 3x_{1it}x_{2it}^2)] + r_i + z_{2it}$$

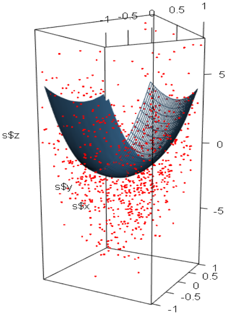


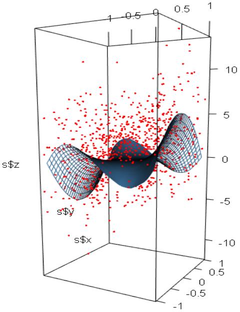


Empirical Studies : Simulation

For the second group,

$$y_{1it} = [-4w_{1it} - 3w_{2it}] + [6x_{1it}^2 + x_{2it}^2 - 2.33] + r_i + z_{1it}$$
$$y_{2it} = [\ \ 2w_{1it} + 5w_{2it}] + [6(x_{1it}^3 - x_{1it}x_{2it}^2)\ \ ] + r_i + z_{2it}$$

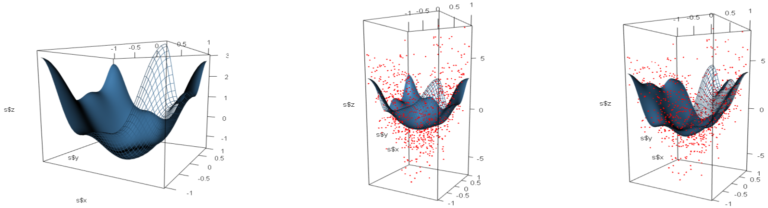




Empirical Studies : Simulation

	β_{11}	β_{12}	β_{13}	β_{14}	β_{21}	β_{22}	β_{23}	β_{24}	σ_R
True	-2	4	-5	3	-4	2	-3	5	2
Fitted	-2.1	3.9	-5.2	3.1	-4.1	1.9	-3.0	5.1	1.66
	(0.12)	(0.13)	(0.11)	(0.11)	(0.12)	(0.12)	(0.13)	(0.11)	(0.20)

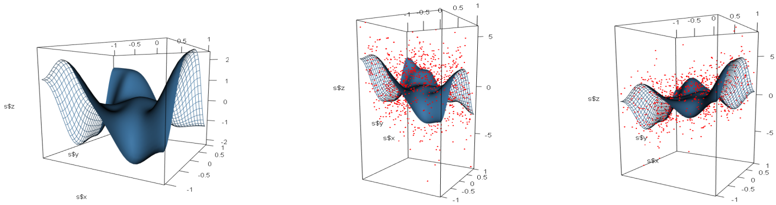
Table: Posterior mean and s.e. of Linear components and s.e. of random effect



(a) Overall (b) Group 1 (c) Group 2

Figure: Fitted nonparametric components on $l = 1$

Empirical Studies : Simulation



(a) Overall (b) Group 1 (c) Group 2

Figure: Fitted nonparametric components on $l = 2$

	G_{11}	G_{12}	G_{21}	G_{22}
RMSE	1.81	1.71	1.60	1.57
RMISE	0.32	0.75	0.86	0.64

Table: Performance Measures

Empirical Studies : Simulation

- DPM identified 97% accuracy on cluster assignment with MAP.
- Posterior mean of parameters are

$$\hat{A}_1 = \begin{pmatrix} 2.26 & -0.58 \\ -0.92 & 2.15 \end{pmatrix}, \hat{C}_1 = \begin{pmatrix} 18.53 & -0.58 \\ -0.58 & 15.90 \end{pmatrix} \Rightarrow \hat{\Sigma}_{Z_1} = \begin{pmatrix} 4.43 & 1.35 \\ 1.35 & 4.28 \end{pmatrix}$$

$$\hat{A}_2 = \begin{pmatrix} 2.94 & -1.64 \\ -0.52 & 3.66 \end{pmatrix}, \hat{C}_2 = \begin{pmatrix} 9.86 & 0.26 \\ 0.26 & 10.39 \end{pmatrix} \Rightarrow \hat{\Sigma}_{Z_2} = \begin{pmatrix} 2.00 & 0.57 \\ 0.57 & 1.50 \end{pmatrix}$$

where true parameters

$$A_1 = \begin{pmatrix} 2 & -0.6 \\ -0.6 & 2 \end{pmatrix}, C_1 = \begin{pmatrix} 16 & 0 \\ 0 & 16 \end{pmatrix} \Rightarrow \Sigma_{Z_1} = \begin{pmatrix} 4.40 & 1.32 \\ 1.32 & 4.40 \end{pmatrix}$$

$$A_2 = \begin{pmatrix} 3 & -2 \\ -0.1 & 3 \end{pmatrix}, C_2 = \begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix} \Rightarrow \Sigma_{Z_2} = \begin{pmatrix} 1.86 & 0.54 \\ 0.54 & 1.52 \end{pmatrix}$$

Discussion

- We present multivariate extension of semiparametric mixed model with tensor product basis and flexible nonparametric mixture on serial correlation.
- Hierarchical representation captures possible similarities across different groups while allowing each group to have distinct surface.
- Further extensions and issues
  - Applicability to Spatio-Temporal Data
  - Time varying coefficients
  - Choice on number of basis
  - Further regularization

References I

Bush, C. A. and MacEachern, S. N. (1996). A semiparametric bayesian model for randomised block designs. *Biometrika*, 83(2):275–285.

Gardiner, C. (2004). *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*. Springer complexity. Springer.

Huang, A., Wand, M. P., et al. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, 8(2):439–452.

Lenk, P. J. (1999). Bayesian inference for semiparametric regression using a fourier representation. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 61(4):863–879.

Lenk, P. J. and Choi, T. (2017). Bayesian analysis of shape-restricted functions using Gaussian process priors. *Statist. Sinica*, 27(1):43–69.

Neal, R. M. (2000). Markov chain sampling methods for dirichlet process mixture models. *Journal of computational and graphical statistics*, 9(2):249–265.

Quintana, F. A., Johnson, W. O., Waetjen, L. E., and B. Gold, E. (2016). Bayesian nonparametric longitudinal data analysis. *Journal of the American Statistical Association*, 111(515):1168–1181.

References II

Rasmussen, C. E. (2006). Gaussian processes for machine learning. MIT Press.

Rosen, O. and Thompson, W. K. (2009). A bayesian regression model for multivariate functional data. *Computational Statistics & Data Analysis*, 53(11):3773–3786.

Smith, M. and Kohn, R. (2002). Parsimonious covariance matrix estimation for longitudinal data. *Journal of the American Statistical Association*, 97(460):1141–1153.

Wasserman, L. (2006). *All of Nonparametric Statistics*. Springer Texts in Statistics. Springer New York.

Appendix : Multivariate OU process [Gardiner, 2004]

- $\boldsymbol{Z}_{it}$ following zero mean Multivariate OU process :

$$d\boldsymbol{Z}_{it} = -A\boldsymbol{Z}_{it} + Bd\boldsymbol{W}_{it}$$

  where $\boldsymbol{W}_{it}$ is L-dim Wiener process.
- Under stationarity condition, Σ_Z be a var-cov matrix of $\boldsymbol{Z}_t$ satisfying

$$A\Sigma_Z + \Sigma_Z A^\top = BB^\top$$

- Letting $\Delta t_{ij} = t_{ij} - t_{i,j-1}$, $(j = 1, \cdots, n_i)$ and $t_{i0} = 0$, the transition density can be obtained :

$$p(\boldsymbol{Z}_{i,t_{ij}} \mid \boldsymbol{Z}_{i,t_{i,j-1}}, \Delta t_{ij}) \propto \mathrm{N}_L\left(\gamma_{t_{ij}} \mid \mathbf{0}, \boldsymbol{\Omega}_{\Delta t_{ij}}\right)$$

  where $\gamma_{t_{ij}} = \boldsymbol{z}_{i,t_{ij}} - \exp(-A\Delta t_{ij})\boldsymbol{z}_{i,t_{i,j-1}}$ and
  $\boldsymbol{\Omega}_{\Delta t_{ij}} = \Sigma_z - \exp(-A\Delta t_{ij})\Sigma_z \exp(-A^\top \Delta t_{ij})$