Summer 2024

Quasi – Monte Carlo Estimation for Functional Generalized Linear Mixed Models

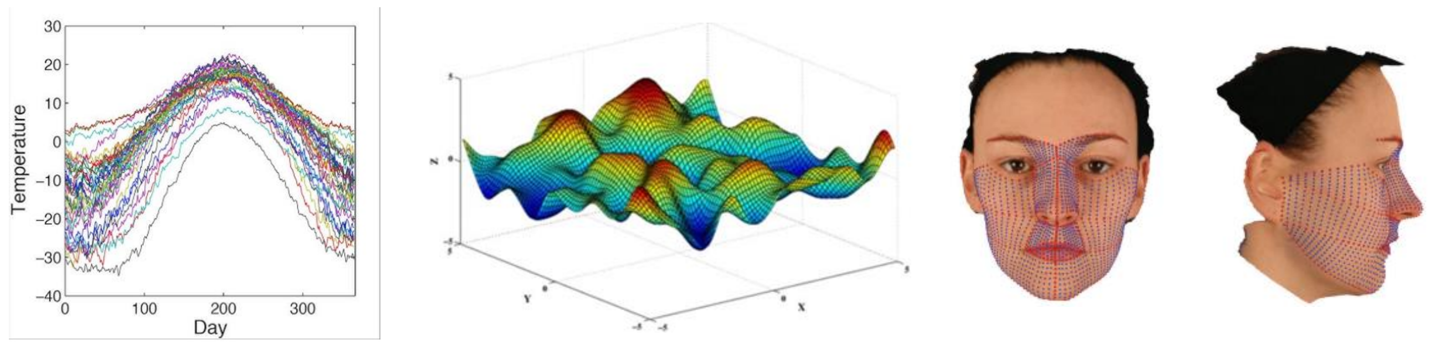
Ruvini Jayamaha
Western Michigan University, ruvinikjayamaha@gmail.com

Quasi – Monte Carlo Estimation for Functional Generalized Linear Mixed Models

Ruvini Jayamaha and Dr. Hyun Bin Kang
Department of Statistics, Western Michigan University, Kalamazoo, MI

1. Proposed Model

- **Functional Data Analysis (FDA)** is a topic of growing interest in the Statistics community. The data in **FDA** are smooth curves or surfaces in time or space which can be conceptualized as functions.



- We propose a **Functional Generalized Linear Mixed Model (FGLMM)** to fit EEG data and estimate the parameters using **Quasi-Monte Carlo Method**.
- This proposed model deals with non-Gaussian scalar response, functional predictor, and random effects. We relax the assumption of link and variance functions.

2. Background and Contribution

- Limited works on scalar response variable in Functional mixed model setting.
- We include the mixed effects and the generalized form of the model by allowing non-Gaussian response to define our proposed model FGLMM.
- Quasi-likelihood allow the non-Gaussian and non-exponential cases.
- We relax the model assumption of known link and variance function.
- Monte Carlo (MC) method allow to approximate the intractable integration point wise.

3. Mathematical framework

- Data representation: $(\{X_{ij}(t), t \in T\},\ Y_{ij})\quad i = 1, \dots, n \quad j = 1, \dots, q$. Assume that these data are from an i.i.d sample.
- Random Intercept model: $g[E(Y)] = \beta_0 + \int_T \beta_1(t) X(t)\, dt + ZU_0$
- Using Karhunen-Loeve expansion with orthonormal basis function $\{\emptyset_k(t)\}_{k=1}^{\infty} \in \mathcal{L}^2(t)$, $t \in T,\ K-$truncated Random Intercept Model is

$$g[E(Y_{ik})] \approx \beta_0 + \sum_{k=1}^{K} b_{1k}\, x_k + ZU_0 \quad i = 1, \cdots, n$$

- **Quasi-Monte Carlo Estimation in FGLMM:**
- Let $P_K = \{c_k : k = 1, \dots, K\} \in C^q$. The QMC approximated log-likelihood is,

$$l(\beta, \theta) = log\left[\frac{1}{K}\sum_{k=1}^{K} exp\left\{\sum_{i=1}^{n} l_i(\beta, \Sigma^{\frac{1}{2}} F^{-1}(c_k))\right\}\right] \qquad (1)$$

Where the conditional log quasi-likelihood is $l_i(\beta, b) \propto \int_{y_i}^{\mu_i} \frac{a_i(y_i - u)}{\phi V(u)}\, du$

- MLE $\hat{\beta}$ of β must be the solution of the score equations:

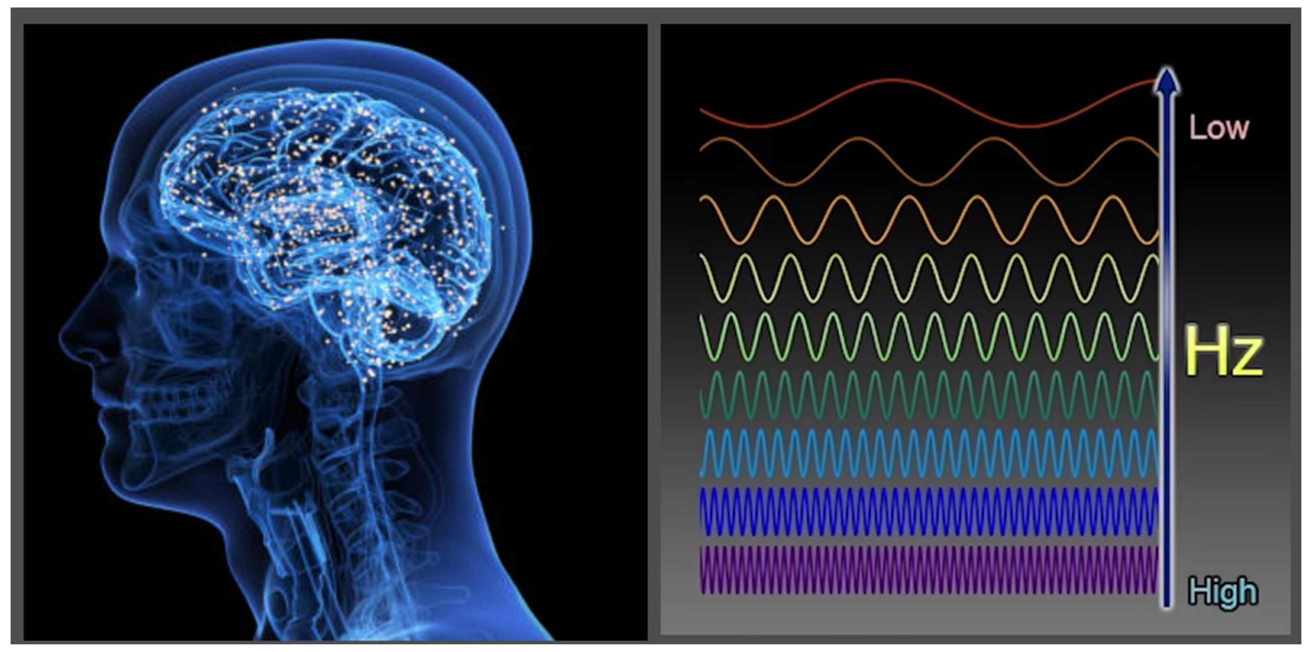
$$\dot{l}_\beta = \frac{\partial}{\partial \beta} l(\beta, \theta) = \sum_{k=1}^{K} w_k \left\{\sum_{i=1}^{n} \frac{a_i(y_i - h(\eta_{ik}))}{\phi v(\mu_{ik}) \dot{g}(\mu_{ik})} x_i\right\} = 0 \qquad (2)$$

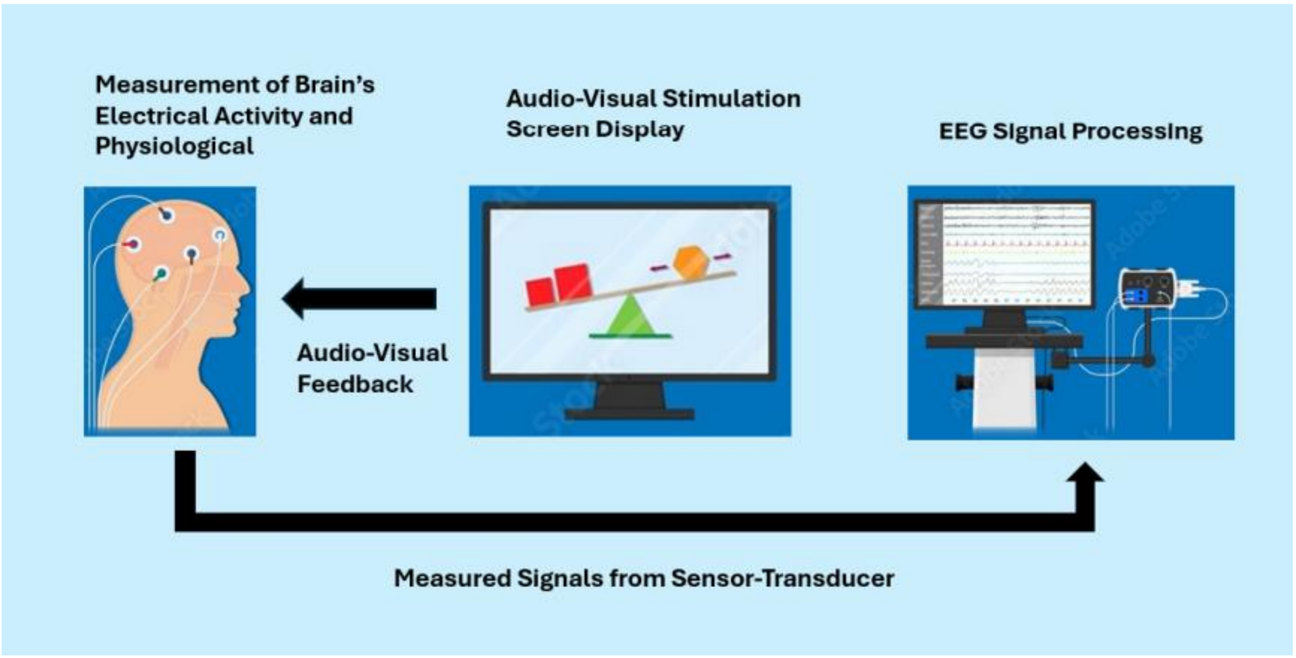
- Score equation for j-th variance component of θ:

$$\dot{l}_{\theta_j} = \frac{\partial}{\partial \theta_j} l(\beta, \theta) = \sum_{k=1}^{K} w_k \left\{\sum_{i=1}^{n} \frac{a_i(y_i - h(\eta_{ik}))}{\phi v(\mu_{ik}) \dot{g}(\mu_{ik})} z_i' \dot{\Sigma}_j^{\frac{1}{2}} b_k\right\} = 0 \qquad (3)$$

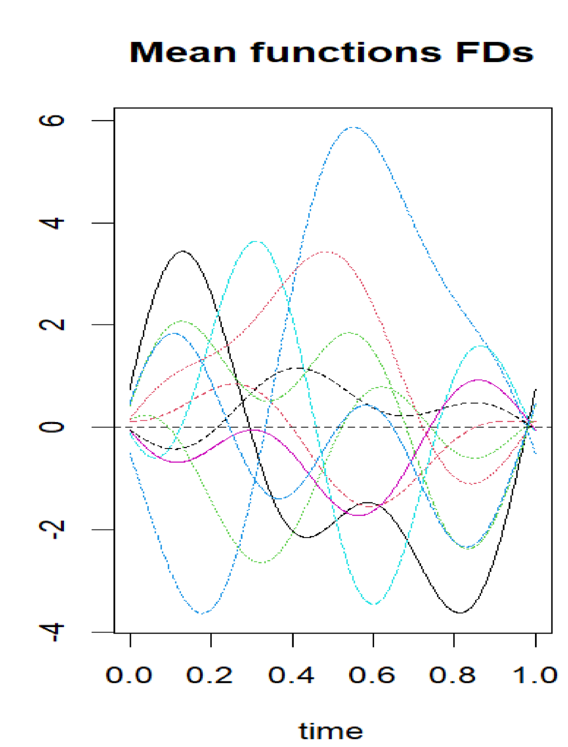
- Use Newton-Raphson algorithm to solve the equations (2) and (3) for given initial (β, θ):

$$\begin{bmatrix}\beta \\ \theta\end{bmatrix}_{new} = \begin{bmatrix}\beta \\ \theta\end{bmatrix}_{old} + \begin{bmatrix} -\ddot{l}_{\beta\beta} & -\ddot{l}_{\beta\theta} \\ -\ddot{l}_{\beta\theta}' & -\ddot{l}_{\theta\theta}\end{bmatrix}_{old}^{-1} \begin{bmatrix}\dot{l}_\beta \\ \dot{l}_\theta\end{bmatrix}_{old} \qquad (4)$$

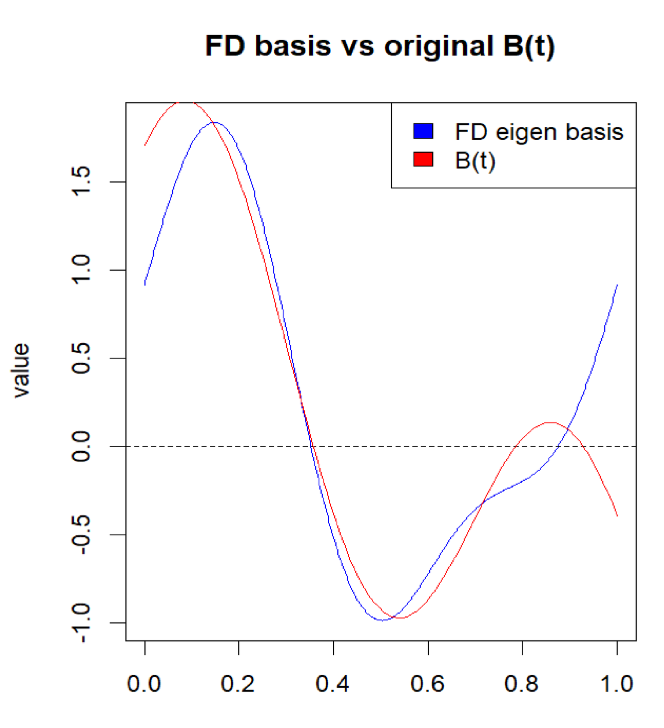




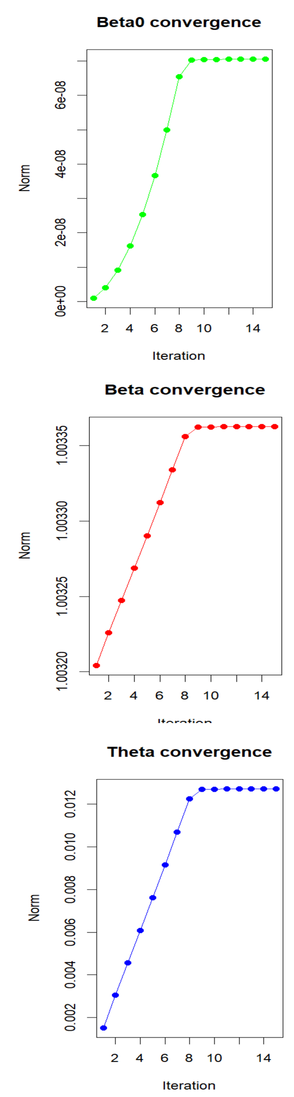
4. Model Performance



Mean Functions of FDs

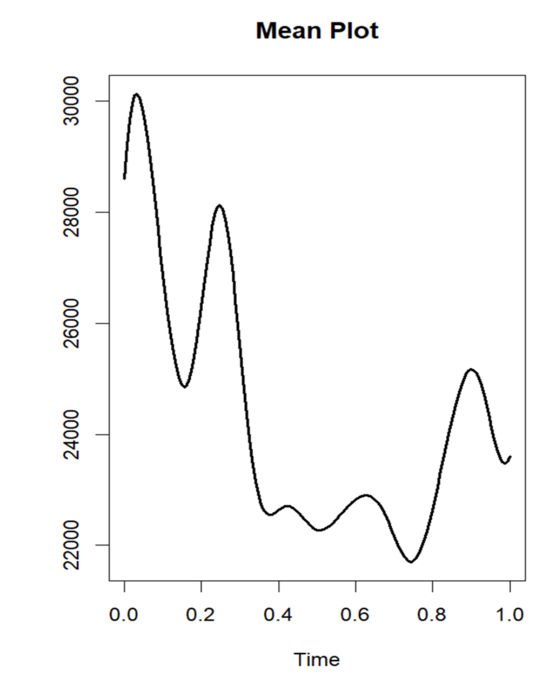


Comparison of FD eigen basis and $\beta(t)$ function.

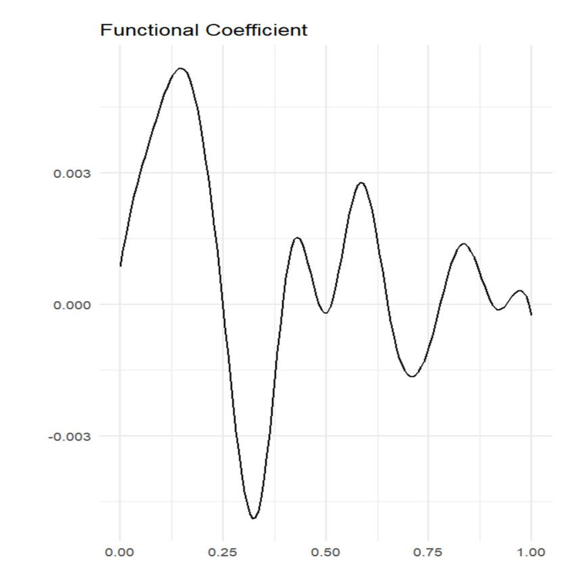


5. Results – Apply model for EEG data set

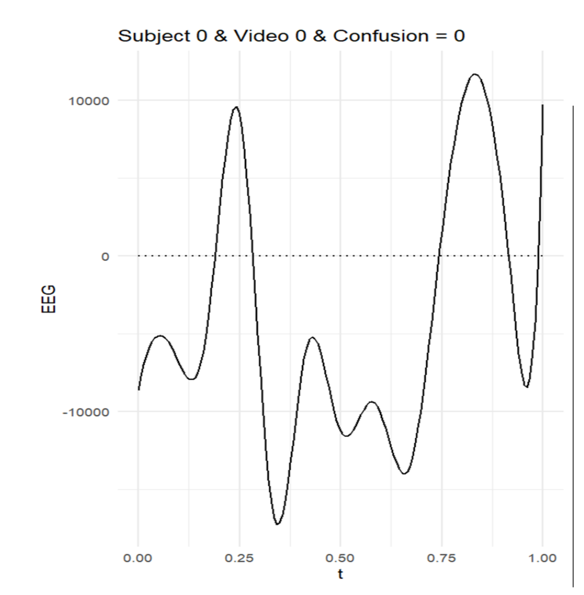
- We are interested with the data set Electroencephalogram (EEG) signals and confusion levels among the college students watching different Massive Open Online Courses (MOOC) videos collected by Wang et al. 2013.



Mean Brain wave Plot

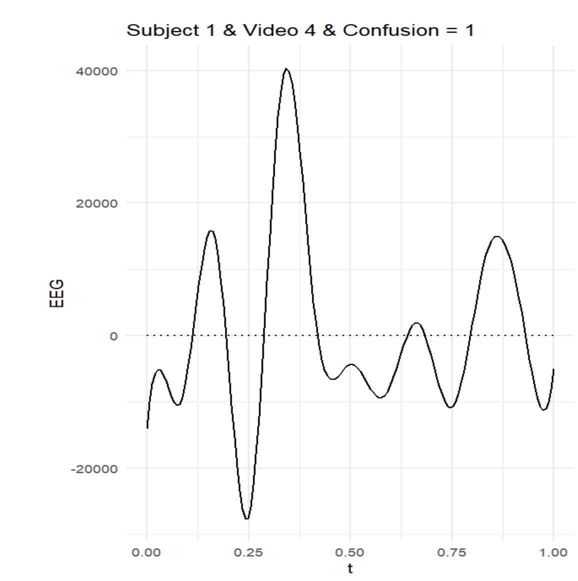


Functional Coefficients



Subject 0, Video 0, Confusion = 0

$\int X(t)\beta(t)dt = -0.78$



Subject 1, Video 4, Confusion = 1

$\int X(t)\beta(t)dt = 0.77$

6. Summary, Conclusion, and Recommendations

- The proposed model proves particularly useful when dealing with data that exhibit temporal or spatial dependencies, as is typical in EEG recordings.
- We can effectively interpret the findings of fitting an FGLMM to EEG data and use functional data plots to gain insights into the process.
- The proposed estimation method, QMC-FGLMM, alleviates the assumptions of usual GLM or GLMM, such as the distribution of the response variable and the link functions, thereby allowing estimation of non-traditional data.
- The model permits the assignment of different coefficients to each group (mixed model), thus facilitating the estimation of distinct behaviors across various groups.

References

- H. Luce. Functional Generalized Linear Mixed Models. PhD thesis, Western Michigan University, 2023.
- H.-G. Müller and U. Stadtmüller. Generalized functional linear models. 2005.
- J. Pan and R. Thompson. Quasi-monte Carlo estimation in generalized linear mixed models. In Proceedings of the 19th international workshop on statistical modelling, pages 239–243. Firenze University Press Firenze, Italy, 2004.
- J. Ramsay, B. Silverman, J. Ramsay, and B. Silverman. Differential operators in functional data analysis. Functional Data Analysis, pages 217–238, 1997