Functional Generalized Linear Mixed Models

Harmony Luce
Western Michigan University

FUNCTIONAL GENERALIZED LINEAR MIXED MODELS

Harmony Luce, Ph.D.

Western Michigan University, 2023

With the advancements in data collection technologies, researchers in various fields such as epidemiology, chemometrics, and environmental science face the challenges of obtaining useful information from more detailed, complex, and intricately-structured data. Since the existing methods often are not suitable for such data, new statistical methods are developed to accommodate the complicated data structures.

As a part of such efforts, this dissertation proposes Functional Generalized Linear Mixed Model (FGLMM), which extends classical generalized linear mixed models to include functional covariates. Functional Data Analysis (FDA) is a rapidly developing area of statistics for data which can be naturally viewed as smooth curves or functions. FDA methods exploit the natural smoothness that characterizes the data and achieves greater statistical efficiency compared to multivariate methods.

After introducing the FGLMM, parameter estimation methods using both Frequentist and Bayesian approaches are discussed. The simulation settings of a random intercept model are used to show the performance of the two estimation schemes. Also, the higher prediction performance of FGLMM will be shown compared to the

Functional Generalized Linear Model (FGLM) and the Generalized Linear Mixed Model (GLMM). Also, an application of the FGLMM to EEG brainwave data is discussed.

FUNCTIONAL GENERALIZED LINEAR MIXED MODELS

by

Harmony Luce

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Statistics
Western Michigan University
June 2023

Doctoral Committee:

Hyun Bin Kang, Ph.D., Chair
Joshua Naranjo, Ph.D.
Duy Ngo, Ph.D.
Sangwoo Lee, Ph.D.

ACKNOWLEDGMENTS

I offer my utmost gratitude to my advisor, Dr. Hyun Bin Kang. Her unwavering guidance and patience gradually shaped me to become a better educator and an accomplished PhD student.

I would like to thank the committee members Dr. Joshua Naranjo, Dr. Duy Ngo, and Dr. Sangwoo Lee for providing careful assessment of my work and sharing their valuable insights that helped in improving my dissertation.

I appreciate the unending support from my family and friends. With our constant communication, I am always reminded that I can persevere through challenges and achieve my goals.

Harmony Luce

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Functional Data Analysis (FDA) is a field of statistics that obtains information from functions, curves, surfaces, or anything else that varies over a continuum (Ramsay and Silverman, 2005; Horváth and Kokoszka, 2012; Wang et al., 2016; Kokoszka and Reimherr, 2017). When one of the variables of interest is smooth over a continuum domain, then it can be considered as functional. Examples include temperature curves (Zhang and Chen, 2007), growth curves (Chen and Müller, 2012), or brainwaves (Tian, 2010). With the development of data collection technology, such data over continuum are observed widely in various fields.

As researchers are interested in investigating the relationship between these functional variable(s) and other variables, the functional regression methods have been developed. Functional linear models include scalar-on-function regression (scalar response and functional predictor), function-on-scalar regression (functional response and scalar predictor), and function-on-function regression (functional response and functional predictor). Functional Generalized Linear Models (James, 2002a; Müller and

Stadtmüller, 2005; McLean et al., 2014) are developed to deal with scalar-on-function regressions where the response variable is non-Gaussian. Functional Mixed Models are developed to deal with the cases where repeated measurements are made on the same units or measurements are made on clusters of related statistical units, but most of the functional mixed models concern functional responses (Guo, 2002; Morris and Carroll, 2006b; Antoniadis and Sapatinas, 2007b; Zhu et al., 2011; Chen and Wang, 2011; Ma et al., 2021).

In this dissertation, we take a further step in these developments and propose a Functional Generalized Linear Mixed Models (FGLMM). FGLMM concerns a non-Gaussian scalar response and functional predictor(s) with random effects. FGLMM is an extension of Generalized Linear Mixed Models (GLMM) in functional settings which is often used to analyze data when the observations or the statistical units are grouped or clustered. We take brainwave data Wang et al. (2013) as a motivation and example of our analysis. The data contains the brainwaves of 10 college subjects on 10 different videos and a binary measure of whether a subject was confused or not while watching a given video. Through our model, we will be able to analyze how the brainwave relates to the confusion of the subject.

1.2 Contribution

Functional Generalized Linear Mixed Models (FGLMM) will allow a new avenue of analysis of data over continuum that are observed repeatedly or observed on clusters of related statistical units. We can investigate how the functional predictor affects the non-Gaussian scalar response with the presence of random effects. Also, by treating the data over continuum as functional objects, we can deal with the high dimensionality of data and exploit the innate smooth correlated structure of data. We also provide a general framework of FGLMM so that it can be extended into different specific models. In this manuscript, we introduce two models of FGLMM – random intercept model and random functional slope model.

In parameter estimation, we use both Frequentist approach and Bayesian approach, making our work one of few examples that deals with both pillars in statistics. Through simulation and real data analysis, the results from Frequentist approach and the results from Bayesian approach will be compared. There are increasing interests in using Bayesian methods in FDA framework, and our work will be a good example for such efforts.

Furthermore, our work on FGLMM allows further research on random effects with the functional data. For example, our model can be extended to spatial functional models or spatio-temporal models where the random effects are spatially correlated.

1.3 Organization

Chapter 2 discusses the related literature and background that are required to give support to the building blocks of the proposed model. This includes Functional Data Analysis, Functional Principal Component Analysis, and Linear Models.

Chapter 3 formally introduces the proposed model, Functional Generalized Linear Mixed Model (FGLMM) together with analogous examples from GLMM such as the random intercept model and random functional slope model. Parameter estimation procedures using both the Frequentist and Bayesian approaches are presented here.

Chapter 4 discusses the process of simulating FGLMM data and uses it to implement the estimation procedures mentioned in chapter 3. It seeks to compare the results yielded from the two estimation procedures, Frequentist and Bayesian approaches.

Chapter 5 provides key insights in the application of FGLMM on the EEG dataset.

Chapter 2

Background and Literature Review

2.1 Functional Data Analysis

2.1.1 Functional Data Analysis

Functional Data Analysis (FDA) deals with the analysis and theory of data that are in the form of functions, images and shapes, or more general objects. It arises when one of the variables or units of interest in a data set can be naturally viewed as a smooth curve or function. FDA can be thought of as a statistical analysis of samples of curves.

Kokoszka and Reimherr (2017) cites that implementing FDA comes along with various reasons for its usage: (1) to represent the data in ways that aid further analysis, (2) to display the data so as to highlight various characteristics, (3) to study important sources of pattern and variation among the data, (4) to explain variation in an outcome or dependent variables by using input or independent variable information, and (5) to compare two or more sets of data with respect to certain types of variation, where two sets of data can contain different sets of replicates of the same functions, or different functions for a common set of replicates.

Ramsay (2009) provides an example for which FDA is very much applicable.

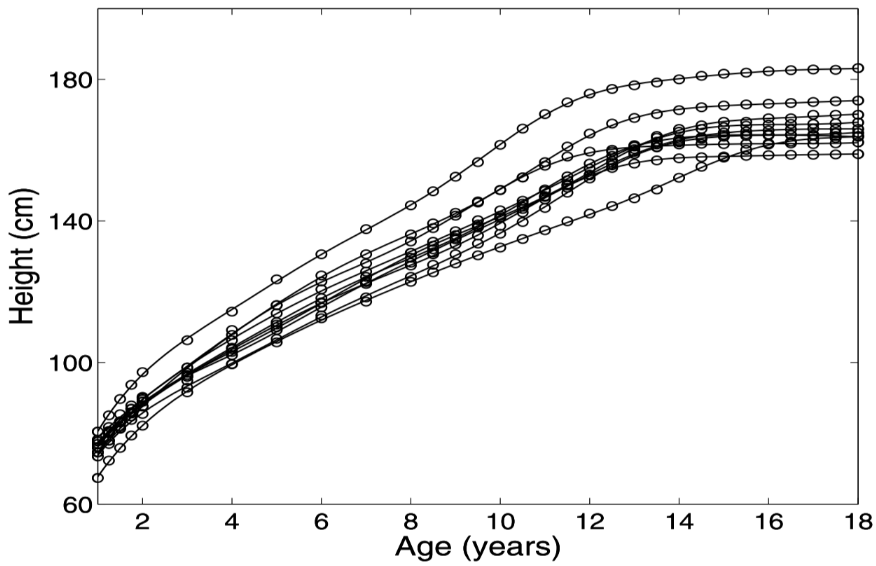


Figure 2.1: Girls' Growth

The above plot illustrates the heights of 10 girls measured at a set of 31 ages in the Berkeley Growth Study. The important observation to make for this particular study is that the height may be modelled using a smooth function and in effect, the data could be instead seen as having 10 functional observations even though it was taken from a finite set of points. Moreover, this can be extended to inspecting the changes in height, as it would be done so with differentiation.

2.1.2 Basis Expansion

With the nature of functions having infinite dimensionality, the best way to address it is by applying basis expansion. The main idea of basis expansion is to approximate the observed functions as linear combinations of some basic shapes. From Ramsay

(1997), it is

$$x(t) \approx \sum_{k=1}^{K} c_k \phi_k(t),$$

where the basis functions are ϕ_k and the coefficients are c_k.

2.1.2.1 Fourier Series

One method for basis expansion is the Fourier Series, which is given by

$$\hat{x}(t) = c_0 + c_1 \sin \omega t + c_2 \cos \omega t + c_3 \sin 2\omega t + c + 4 \cos 2\omega t + \cdots$$

It is defined by the periodic basis $\phi_0(t) = 1$, $\phi_{2r-1}(t) = \sin r\omega t$, and $\phi_{2r}(t) = \cos r\omega t$. Its period of $\frac{2\pi}{\omega}$ is defined by the parameter ω. The Fourier basis is also known to be orthogonal since $\int \phi_i(t)\phi_j(t)dt = 1$ for $i = j$ and $\int \phi_i(t)\phi_j(t)dt = 0$ otherwise.

2.1.2.2 B-Spline

A common option for representing non-periodic functional data or parameters would be by using splines. A spline function is mainly characterized by the order of its polynomial segments and its knots. The simplest description of a spline is that it is a method of approximating segments of observed functions on each subinterval.

2.1.2.3 Wavelet

Morris and Carroll (2006a) describes Wavelets as families of orthonormal basis functions that represents other functions. For any square-integrable function in $(-\infty, \infty)$, an orthogonal wavelet basis can be determined by dilating and translating a mother wavelet ψ as

$$\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$$

for integers j and k. Hence, a function $g(t)$ may be expressed as

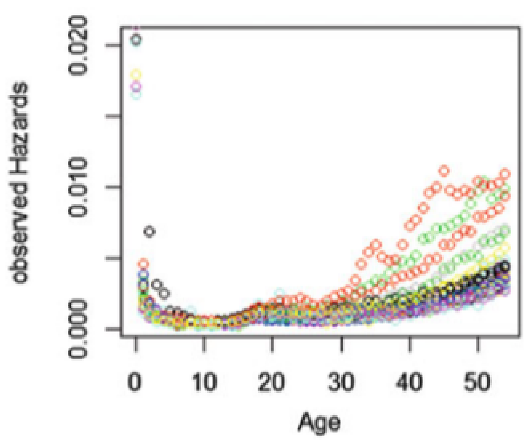
$$g(t) = \sum_{j,k \,\in\, \mathcal{T}} d_{jk}\psi_{jk}(t),$$

where the coefficients are

$$d_{jk} = \int g(t)\psi_{jk}(t)dt.$$

2.1.3 Estimating Smooth Functions

The concepts from the two previous sections combines to allow for estimation of smooth functions. The graph illustrates the annual mortality hazards of boys born in 1960.



Cleophas and Zwinderman (2018) provides the above plot and it illustrates 23 observed hazards from 1960 to 2014. If each hazard function can be represented using a linear combination of coefficients and a choice of basis function (e.g. Fourier, B-Spline) then

$$x_n(t) \approx \sum_{k=1}^{k} c_{nk}B_k(t), \quad n = 1, \ldots, 23, \quad t \in \mathcal{T}.$$

Golub et al. (1979) discussed a method to estimating the coefficients. This is achieved by minimizing

$$\sum_{n=1}^{23}\sum_{j=1}^{54}\left(x_{nj}-\sum_{k=1}^{K}c_{nk}B_k(t_j)\right)^2+\lambda\int L\left[\left(\sum_{k=1}^{K}c_{nk}B_k(t_j)\right)\right]^2 dt.$$

The function $L(\cdot)$ measures the roughness of the $x_n(t)$'s and the smoothing (or penalty) parameter λ must be estimated from the data or may be chosen by implementing generalized cross validation (GCV).

2.2 Functional Principal Component Analysis

From Rencher (2012), the goal of Principal Component Analysis (PCA) is to maximize the variance of a linear combination of variables. It serves as a dimension reduction tool for cases when the number of independent variables is relatively larger than the number of observations; or when the independent variables are strongly correlated. The multivariate PCA method follows these steps:

1. Determine the principal component weight vector $\boldsymbol{\xi}_1 = (\xi_{11}, \ldots, \xi_{p1})^\top$ for which the principal components scores

   $$f_{i1} = \sum_j \xi_{j1} x_{ij} = \boldsymbol{\xi}_1 \mathbf{x}_i$$

   maximize $\sum_i f_{i1}^2$ subject to

   $$\sum_j \xi_{j1}^2 = \|\boldsymbol{\xi}_1\|^2 = 1.$$

2. Compute the weight vector $\boldsymbol{\xi}_2$ with components ξ_{j2} and principal component

scores maximizing $\sum_i f_{i2}^2$, subject to the constraint $\|\boldsymbol{xi}_2\|^2 = 1$ and the additional constraint

$$\sum_j \xi_{j2}\xi_{j1} = \boldsymbol{\xi}_2^\top \boldsymbol{\xi}_1 = 0.$$

3. Repeat the process, as required.

Functional Principal Component Analysis (FPCA) is the analog of PCA in the functional setting. The transition from multivariate PCA to FPCA simply changes the summations into integrals in the aforementioned procedure.

1. Determine the principal component weight function $\xi_1(s) = (\xi_{11}, \ldots, \xi_{p1})^\top$ for which the principal components scores

$$f_{i1} = \int \xi_(s) x_i(s) ds$$

maximize $\sum_i f_{i1}^2$ subject to

$$\int \xi_1(s)^2 ds = \|\boldsymbol{\xi}_1\|^2 = 1.$$

2. Compute the weight function $\xi_2(s)$ with components and principal component scores maximizing $\sum_i f_{i2}^2$, subject to the constraint $\|\boldsymbol{\xi}_2\|^2 = 1$ and the additional constraint

$$\int \xi_2(s)\xi_1(s) ds = 0.$$

3. Repeat the process, as required.

2.3 Linear Model

2.3.1 Linear Model

The Linear Model (LM) aims to model a linear relationship between the y_i's and x_{ij}'s given by a sample of n observations. The Linear Model is commonly written as an equation that relates the response vector $\underset{n\times 1}{\mathbf{Y}}$ with the design matrix $\underset{n\times(p+1)}{\mathbf{X}}$ associated with the explanatory variable/s through the parameter vector $\underset{(p+1)\times 1}{\boldsymbol{\beta}}$. In symbols, it is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where n and $p+1$ are the respective number of observations and number of parameters. The parameter vector is also called the vector of regression coefficients. The additional vector $\underset{n\times 1}{\boldsymbol{\varepsilon}}$ accounts for the errors in the model - these are collectively called as the error terms. With that given structure comes with the following assumptions:

1. The error terms have zero mean, i.e., $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$.
2. The error terms have constant variance (called *homoscedastic*) and are uncorrelated, i.e., $\mathrm{Cov}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top) = \sigma^2\mathbf{I}$ where $\sigma^2 > 0$ and $\mathbf{I}$ is the identity matrix.
3. The design matrix $\mathbf{X}$ is of full rank, i.e. $\mathrm{rank}(X) = p + 1$.
4. The error terms follows a normal distribution, where $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I})$.
5. The error terms and explanatory variables are independent, i.e., $\boldsymbol{\varepsilon} \mid \mathbf{X} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I})$.

A well-known approach to estimating the parameters is by minimizing the Sum of Squared Errors (SSE). It begins by writing the SSE as a function of $\boldsymbol{\beta}$ and applying matrix calculus to get the *least squares estimator*. A brief derivation of the estimator is shown below.

$$\begin{aligned}
\text{SSE} &= \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon} \\
&= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \mathbf{Y}^\top \mathbf{Y} - 2\mathbf{Y}^\top \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \\
\frac{\partial}{\partial \boldsymbol{\beta}} \text{SSE} &= -2\mathbf{X}^\top \mathbf{Y} + 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \\
-2\mathbf{X}^\top \mathbf{Y} + 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} &= \mathbf{0} \\
\hat{\boldsymbol{\beta}} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}.
\end{aligned}$$

The above setup of the Linear Model is specifically called the Multiple Linear Regression Model (MLRM). When $p = 1$, it becomes a Simple Linear Regression Model (SLRM).

2.3.2 Functional Linear Model

The Functional Linear Model (FLM) considers either one of the response or the explanatory variables to be a curve and the other one would be scalar. Another case would be when both variables are curves. This is commonly called as Functional Regression and its three cases are Scalar-On-Function Regression, Function-On-Scalar Regression, and Function-On-Function Regression. Kokoszka and Reimherr (2017) defined and provided outlines for each of the aforementioned FLMs.

2.3.2.1 Scalar-On-Function Regression

The Scalar-On-Function Regression (SOFR) is when the response Y_i is scalar, the regressor $X_i(s)$ is a curve, and the parameter is a regression function $\beta(s)$. Its equation form is taken as

$$Y_i = \int \beta(s)X_i(s)ds + \varepsilon_i, \quad i = 1, 2, \ldots, n.$$

With the inclusion of an intercept, the model will be

$$Y_i = \alpha + \int \beta(s)X_i(s)ds + \varepsilon_i, \quad i = 1, 2, \ldots, n.$$

By applying basis (e.g. fourier basis, bspline basis) expansion on β, the above model would be equivalent to an LM. Assume that

$$\beta(t) = \sum_{k=1}^{K} c_k B_k(t),$$

then

$$\begin{aligned}\int \beta(t)X_i(t)dt &= \sum_{k=1}^{K} c_k \int B_k(t)X_i(t)dt \\ &\approx \sum_{k=1}^{K} x_{ik}c_k.\end{aligned}$$

From there, it will become an LM where the parameter vector $\mathbf{c} = [\alpha, c_1, c_2, \ldots, c_K]^\top$ will be estimated in the same way as that from LM, i.e. $\hat{\mathbf{c}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$.

2.3.2.2 Function-On-Scalar Regression

The Function-On-Scalar Regression (FOSR) takes curves $Y_i(t)$ for responses and scalars x_{ik} for regressors, with the coefficient functions $\beta_k(t)$ as the parameters. It is written as

$$Y_i(t) = \sum_{k=1}^{q} x_{ik}\beta_k(t) + \varepsilon_i(t).$$

The least squares estimator approach may also be applied to estimate $\boldsymbol{\beta}(t)$. Note that in this case, the least squares estimator would be the value of $\boldsymbol{\beta}$ that minimizes

$$\sum_{i=1}^{n} \left|\left| Y_i - \sum_{k=1}^{q} x_{ik}\beta_k \right|\right|^2 = \int \sum_{i=1}^{n} e_i^2(\boldsymbol{\beta}, t)dt$$

. Similarly, $\sum_{i=1}^{n} e_i^2(\boldsymbol{\beta}, t)$ is minimized when

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y}(t),$$

where

$$e_i^2(\boldsymbol{\beta}, t) = \left(Y_i(t) - \sum_{k=1}^{q} x_{ik}\beta_k(t) \right)^2 .$$

2.3.2.3 Function-On-Function Regression

The Function-On-Function Regression (FOFR) has both the responses $Y_i(t)$ and regressors $X_i(s)$ as curves. And $\beta(t, s)$ is the kernel parameter. In equation form, it is

$$Y_i(t) = \int \beta(t, s)X_i(s)ds + \varepsilon_i(t).$$

2.4 Generalized Linear Model

2.4.1 Generalized Linear Model

The Generalized Linear Model (GLM) is an extension to the Linear Model (LM), where the main difference between the two is that GLM is used when the response variable is non-normal. As discussed by Agresti (2015) and McCullagh and Nelder (2019), the GLM can be broken down into three components:

- The *Random Component* that identifies the probability distribution of the response variable $\underset{n\times 1}{\mathbf{Y}}$. The observations $Y_1, Y_2, \ldots, Y_n$ are assumed to be independent.
- The *Linear Predictor* $\mathbf{X}\boldsymbol{\beta}$, where $\underset{n\times p}{\mathbf{X}}$ is the matrix of p explanatory variables from the n observations and $\underset{p\times 1}{\boldsymbol{\beta}}$ is the parameter vector.
- The *Link Function* $g(\cdot)$ such that $g[\mathrm{E}(\mathbf{Y})] = \mathbf{X}\boldsymbol{\beta}$. This is usually denoted by $\boldsymbol{\eta}$ and the expected value is denoted by $\boldsymbol{\mu}$ so that $\boldsymbol{\eta} = g(\boldsymbol{\mu})$.

In the most basic cases of GLM, there is the assumption that $\mathbf{Y}$ comes from the *exponential family*. The probability density (or mass) function for Y_i takes the form

$$f(y_i; \theta_i, \phi) = \exp\left[\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right],$$

where θ_i is the natural parameter and ϕ is the dispersion parameter. Additionally, the link function $g(\cdot)$ may be a monotonic twice differentiable function. Then, $\boldsymbol{\mu} = g^{-1}(\boldsymbol{\eta})$.

2.4.1.1 Exponential Family

Using the likelihood approach for estimating the parameters, the GLM likelihood equations are

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{ij}}{\mathrm{Var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0, \quad j = 1, 2, \ldots, p,$$

where $\eta_i = \sum_{j=1}^{p} \beta_j x_{ij} = g(\mu_i)$. Equivalently, the likelihood equations are

$$\mathbf{X}^\top \mathbf{D} \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0},$$

where $\mathbf{D} = \text{diag}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)$. Agresti (2015) discussed the Newton-Raphson and Fisher Scoring methods for estimation. Here, the generalized least squares estimate for $\boldsymbol{\beta}$ is

$$(\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{z},$$

where $\mathbf{V} = \text{Cov}(\boldsymbol{\varepsilon})$ and the general linear model is $\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Using the derived likelihood equations above, the score equations can be expressed as

$$\mathbf{u} = \mathbf{X}^\top \mathbf{W}\mathbf{D}(\mathbf{y} - \boldsymbol{\mu}).$$

Then the Fisher scoring equations would be

$$(\mathbf{X}^\top \mathbf{W}^{(t)}\mathbf{X})\boldsymbol{\beta}^{(t+1)} = \mathbf{X}^\top \mathbf{W}^{(t)}\mathbf{z}^{(t)},$$

and its solution is

$$\boldsymbol{\beta}^{(t+1)} = (\mathbf{X}^\top \mathbf{W}^{(t)}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{W}^{(t)}\mathbf{z}^{(t)}.$$

2.4.1.2 Quasi-Likelihood

An alternative approach presented in Agresti (2015) and McCullagh and Nelder (2019) makes use of a form similar to that in the likelihood equations for the exponential family assumption. However, in this method, there is no distributional assumption for the response $\mathbf{Y}$. To simplify the notation, the construction of the quasi-likelihood function begins with

$$U = u(\mu; Y) = \frac{Y - \mu}{\sigma^2 V(\mu)},$$

where

$$\begin{aligned} \mathrm{E}(U) &= 0 \\ \mathrm{Var}(U) &= \frac{1}{\sigma^2 V(\mu)} \\ -\mathrm{E}\left(\frac{\partial U}{\partial \mu}\right) &= \frac{1}{\sigma^2 V(\mu)}. \end{aligned}$$

If the integral

$$Q(\mu; y) = \int_{y}^{\mu} \frac{y-t}{\sigma^2 V(t)} dt,$$

then it should behave like a log-likelihood function for μ under certain assumptions. The function $Q(\mu; y)$ is referred to as the quasi-likelihood (or log quasi-likelihood). Since the responses are independent, then the quasi-likelihood for the whole data would be

$$Q(\boldsymbol{\mu}; \mathbf{y}) = \sum_{i} Q_i(\mu_i; y_i).$$

And the quasi-deviance function is

$$D(y; \mu) = -2\sigma^2 Q(\mu; y) = 2 \int_{\mu}^{y} \frac{y-t}{V(t)} dt.$$

The parameter estimation procedure is similar to before as it would still undergo differentiation to obtain the quasi-score function given by

$$\mathbf{U}(\boldsymbol{\beta}) = \frac{1}{\sigma^2} \mathbf{D}^\top \mathbf{V}^{-1} (\mathbf{Y} - \boldsymbol{\mu}).$$

The quasi-score equation

$$\mathbf{U}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$$

can be solved applying the Newton-Raphson with Fisher scoring, i.e. choosing a value for $\hat{\boldsymbol{\beta}}_0$ that is close to $\hat{\boldsymbol{\beta}}$ so that through a number of iterations a convergence will occur and

$$\hat{\boldsymbol{\beta}}_1 = \boldsymbol{\beta} + (\mathbf{D}^\top \mathbf{V}^{-1}\mathbf{D})^{-1}\mathbf{D}^\top \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}).$$

2.4.2 Functional Generalized Linear Model

In the Functional Generalized Linear Model (FGLM), the explanatory variable $\mathbf{X}$ is replaced with a function $\mathbf{X}(t)$. One of the earliest documentations on FGLM is by James (2002b). In their model, the predictor becomes functional so that

$$g(\mu) = \beta_0 + \int \omega_1(t)X(t)dt.$$

With their choice of cubic spline, a parametrization would use

$$X(t) = \mathbf{s}(t)^\top \gamma, \qquad \gamma \sim \mathrm{N}(\boldsymbol{\mu}_\gamma, \Gamma),$$

where $\mathbf{s}(t)$ is the q-dimensional spline basis at time t, γ is the q-dimensional spline coefficients for the predictor. With that parametrization, the model becomes

$$\begin{aligned} g(\mu_i) &= \beta_0 + \int \omega_1(t)\mathbf{s}(t)^\top \gamma_i dt \\ &= \beta_0 + \boldsymbol{\beta}_1^\top \gamma_i. \end{aligned}$$

It is also assumed that

$$x(t) = X(t) + e(t),$$

where $e(t)$ is a zero-mean Gaussian process that represents errors.

Another take on FGLM is by Müller and Stadtmüller (2005), where they consider the functional quasi-likelihood model:

$$Y_i = g\left(\alpha + \int \beta(t)X_i(t)dw(t)\right) + e_i, \qquad i = 1, \dots, n,$$

where

$$\begin{aligned} \mathrm{E}(e|X(t), t \in \mathcal{T}) &= 0, \\ \mathrm{Var}(e|X(t), t \in \mathcal{T}) &= \sigma^2(\mu) = \tilde{\sigma}^2(\eta). \end{aligned}$$

In a similar fashion to how estimation is performed in FLM, basis expansion is used. Assuming an orthonormal basis $\rho_j,\ j = 1, 2, \dots,$ then the predictor $X(t)$ and the parameter function $\beta(t)$ can be expressed as

$$X(t) = \sum_{j=1}^{\infty} \varepsilon_j \rho_j(t), \qquad \beta(t) = \sum_{j=1}^{\infty} \beta_j \rho_j(t),$$

where

$$\begin{aligned} \varepsilon_j &= \int X(t)\rho_j(t)dw(t), \\ \beta_j &= \int \beta(t)\rho_j(t)dw(t). \end{aligned}$$

By orthonormality,

$$\int \beta(t)X(t)dw(t) = \sum_{j=1}^{\infty} \beta_j \varepsilon_j.$$

Due to the infinite dimensionality of the predictors, a series of models would be used to approximate the given model in which the number of predictors is truncated at $p = p_n$.

The truncation is done by letting

$$U_p = \alpha + \sum_{j=1}^{p} \beta_j \varepsilon_j, \qquad V_p = \sum_{j=p+1}^{\infty} \beta_j \varepsilon_j.$$

Additionally, if $\varepsilon_j^{(i)} = \int X_i(t)\rho_j(t)dw(t)$ where the e_i' are the standardized errors, the model becomes

$$Y_i = g\left(\alpha + \sum_{j=1}^{\infty} \beta_j \varepsilon_j^{(i)}\right) + e_i' \tilde{\sigma}\left(\alpha + \sum_{j=1}^{\infty} \beta_j \varepsilon_j^{(i)}\right), \quad i = 1, \ldots, n.$$

Introducing truncated linear predictors η and means μ as such

$$\eta_i = \alpha + \sum_{j=1}^{p} \beta_j \varepsilon_j^{(i)}, \quad \mu_i = g(\eta_i),$$

then the p-truncated model is

$$Y_i^{(p)} = g_p\left(\alpha + \sum_{j=1}^{p} \beta_j \varepsilon_j^{(i)}\right) + e_i' \tilde{\sigma}_p\left(\alpha + \sum_{j=1}^{p} \beta_j \varepsilon_j^{(i)}\right), \quad i = 1, \ldots, n.$$

Since the errors vanish asymptotically, the model that could be used would be

$$Y_i^{(p)} = g\left(\alpha + \sum_{j=1}^{p} \beta_j \varepsilon_j^{(i)}\right) + e_i' \tilde{\sigma}\left(\alpha + \sum_{j=1}^{p} \beta_j \varepsilon_j^{(i)}\right), \quad i = 1, \ldots, n.$$

In the estimation procedure, the score equation

$$U(\beta) = 0$$

must be solved, where

$$U(\beta) = \sum_{i=1}^{n} \frac{(Y_i - \mu_i) g'(\eta_i) \varepsilon^{(i)}}{\sigma^2(\mu_i)}.$$

The following matrices would be key to solving the score equation

$$D = D_{n,p} = \left(\frac{g'(\eta_i)\varepsilon_k^{(i)}}{\sigma(\mu_i)} \right)_{1 \le i \le n,\, 0 \le k \le p},$$

$$V = V_{n,p} = \operatorname{diag}\left(\sigma^2(\mu_1), \ldots, \sigma^2(\mu_n)\right)_{1 \le i \le n}.$$

From which, the score equation would be equivalent to

$$D^\top V^{-1/2}(Y - \mu) = 0,$$

where it can be solved by applying the iterated least squares method, as that from GLM.

2.5 Linear Mixed Model

2.5.1 Linear Mixed Model

The Linear Mixed Model (LMM) incorporates a random effect to the already existing linear predictor, $\mathbf{X}\boldsymbol{\beta}$. The linear predictor becomes $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U}$, where $\underset{n \times q}{\mathbf{Z}}$ is the design matrix associated with the random effects vector, $\underset{q \times 1}{\mathbf{U}}$. The model would now be:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \boldsymbol{\varepsilon}.$$

As outlined by Fahrmeir (2013), with the Gaussian model assumption, the random effects and errors are both normal and independent. Their joint distribution may be expressed as

$$\begin{bmatrix} \mathbf{U} \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim \underset{q+n}{\mathrm{N}} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \right),$$

where $\mathbf{G}$ and $\mathbf{R}$ are block diagonal covariance matrices given by

$$\mathbf{G} = \text{blockdiag}(\mathbf{Q}, \ldots, \mathbf{Q}, \ldots, \mathbf{Q})$$
$$\mathbf{R} = \text{blockdiag}(\sigma^2\boldsymbol{\Sigma}_{n_1}, \ldots, \sigma^2\boldsymbol{\Sigma}_{n_i}, \ldots, \sigma^2\boldsymbol{\Sigma}_{n_m}).$$

With the introduction of the random effects, it must be emphasized that the variances of these random effects are to be estimated alongside the intercept and slopes designated to the fixed effects. Common approaches to estimating the variance and covariance parameters are by maximization of either the profile log-likelihood l_P or the restricted log-likelihood l_R given by

$$l_P(\boldsymbol{v}) = -\frac{1}{2}\left[\log|\mathbf{V}(\boldsymbol{v})| + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{v}))^\top \mathbf{V}(\boldsymbol{v})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{v}))\right],$$
$$l_R(\boldsymbol{v}) = l_P(\boldsymbol{v}) - \frac{1}{2}\log|\mathbf{X}^\top \mathbf{V}(\boldsymbol{v})^{-1}\mathbf{X}|,$$

where $\boldsymbol{v}$ is the vector of unknown parameters. Note that $\mathbf{V} = \mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}^\top$ whenever the parameters in $\mathbf{G}$ and $\mathbf{R}$ are known. Otherwise $\boldsymbol{v}$ will be estimated through the use of either of the above log-likelihood functions. The estimated fixed and random effects are

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \hat{\mathbf{V}}\mathbf{X})^{-1}\mathbf{X}^\top \hat{\mathbf{V}}^{-1}\mathbf{y},$$
$$\hat{\mathbf{U}} = \hat{\mathbf{G}}\mathbf{Z}^\top \hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

2.5.2 Functional Linear Mixed Model

2.5.2.1 Functional Linear Mixed Model

The Functional Linear Mixed Model (FLMM) considers either (or both) the fixed or random effects to be functions.

Morris and Carroll (2006a) introduces the FLMM in the form

$$\mathbf{Y}(t) = X\mathbf{B}(t) + Z\mathbf{U}(t) + \mathbf{E}(t),$$

where $\underset{N\times 1}{\mathbf{Y}(t)}$ is the vector of observed functions, $\underset{p\times 1}{\mathbf{B}(t)}$ is the vector of fixed effect functions, and $\underset{m\times 1}{\mathbf{U}(t)}$ is the vector of random effect functions, while $\underset{N\times p}{X}$ and $\underset{N\times m}{Z}$ are their respective design matrices. And $\underset{N\times 1}{\mathbf{E}(t)}$ is the vector of functions representing the residual processes. Important assumptions for this model includes that $\mathbf{U}(t) \sim MGP(P, Q)$ and $\mathbf{E}(t) \sim MGP(R, S)$ are independent, where both of them follow a multivariate Gaussian process (MGP).

Antoniadis and Sapatinas (2007a) writes the model as

$$Y_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta}(t_{ij}) + \mathbf{Z}_{ij}\boldsymbol{\alpha}^{(i)}(t_{ij}) + \varepsilon_{ij}, \quad i = 1, \ldots, n;\ j = 1, \ldots, m_i,$$

where $\underset{p\times 1}{\boldsymbol{\beta}(t)}$ is a vector of fixed functions $\underset{q\times 1}{\boldsymbol{\alpha}^{(i)}(t)}$ is a vector of stochastically independent random functions, which are modelled as realizations of zero-mean Gaussian processes.

2.5.2.2 Functional Additive Mixed Model

Scheipl et al. (2016) elaborates on the use of Generalized Functional Additive Mixed Model (GFAMM). The model is structured as

$$\begin{aligned} y_{il} &= \mathcal{F}(\mu_{il}, \boldsymbol{\nu}) \\ g(\mu_{il}) &= \sum_{r=1}^{R} f_r(\boldsymbol{\mathcal{X}}_{ri}, t_{il}), \end{aligned}$$

where $i = 1, \ldots, n$, $l = 1, \ldots, T_i$, $f(\mathbf{t})$ is the vector of values upon evaluating the entries in $\mathbf{t}$, and $f(\mathbf{x}, \mathbf{t})$ is the vector of evaluations for all combinations of rows in $\mathbf{x}$ and $\mathbf{t}$.

Similar to how it had been done in every functional model, basis functions would be used to approximate the terms of $f_r(\mathcal{X}_r, t)$. This is done so by using tensor product of marginal bases evaluated on $\boldsymbol{\mathcal{X}_r}$ and $\mathbf{t}$. For this structure,

$$f_r(\boldsymbol{\mathcal{X}_r}, \mathbf{t}) \approx (\boldsymbol{\Phi}_{\boldsymbol{x}r} \odot \boldsymbol{\Phi}_{\boldsymbol{t}r})\boldsymbol{\theta}_r = \boldsymbol{\Phi}_r\boldsymbol{\theta}_r,$$

where $\boldsymbol{\Phi}_{\boldsymbol{x}r}$ is the matrix of evaluations of a suitable marginal basis for the covariates in $\boldsymbol{\mathcal{X}_r}$ and $\boldsymbol{\Phi}_{\boldsymbol{t}r}$ is the matrix that contains the evaluations of a marginal basis in $\mathbf{t}$, provided their respective basis functions K_{xr} and K_{tr}.

2.6 Generalized Linear Mixed Model

2.6.1 Generalized Linear Mixed Model

The Generalized Linear Mixed Model (GLMM) is an extension to GLM, which now includes random effects. As a result, the GLMM combines the ideas of GLM and LMM, where the link function $g(\cdot)$ is such that

$$g(\mathrm{E}(\mathbf{Y})) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U}.$$

The linear predictor will be $\boldsymbol{\eta}$

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U},$$

so that

$$\boldsymbol{\mu} = \mathrm{E}(\mathbf{Y}) = g^{-1}(\boldsymbol{\eta}).$$

To accommodate correlated random effects (Fahrmeir, 2013), the assumption that $\mathbf{U} \sim \mathrm{N}(\mathbf{0}, \mathbf{G})$ is imposed, where $\mathbf{G}$ is a positive definite covariance matrix.

2.6.1.1 Estimation in GLMM

2.6.1.1.1 Known Variance-Covariance Parameters

Under this assumption, Fahrmeir (2013) showed that the estimation procedure would be similar to that outlined for in LMM, where the joint likelihood

$$L(\boldsymbol{\beta}, \mathbf{U}) = p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{U})p(\mathbf{U})$$

would be utilized. It suffices to calculate the posterior mode since

$$p(\boldsymbol{\beta}, \mathbf{U}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{U})p(\mathbf{U}).$$

The penalized log-likelihood would then be

$$l_{\text{pen}}(\boldsymbol{\beta}, \mathbf{U}) = l(\boldsymbol{\beta}, \mathbf{U}) - \frac{1}{2}\mathbf{U}^\top \mathbf{G}^{-1}\mathbf{U}.$$

By the method of iteratively weighted least squares, the estimator would then have a similar form to that of LMM. It is given by

$$\begin{pmatrix} \boldsymbol{\beta}^{(\hat{t}+1)} \\ \hat{\mathbf{U}}^{(t+1)} \end{pmatrix} = (\mathbf{C}^\top \mathbf{W}^{(t)}\mathbf{C} + \mathbf{B})^{-1}\mathbf{C}^\top \mathbf{W}^{(t)}\tilde{\mathbf{y}}^{(t)},$$

where $\mathbf{C} = (\ \mathbf{X}\ \ \mathbf{Z}\ )$, $\mathbf{B} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{pmatrix}$, and

$$\mathbf{W} = \mathbf{D}\boldsymbol{\Sigma}^{-1}\mathbf{D}$$

$$\tilde{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{U}} + \mathbf{D}^{-1}(\mathbf{y} - \boldsymbol{\mu}),$$

where $\mathbf{D} = \text{diag}\left(\ldots, \dfrac{\partial h(\eta_i)}{\partial \eta}, \ldots\right)$ and $\mathbf{\Sigma} = \text{diag}(\ldots, \sigma_i^2, \ldots)$.

2.6.1.1.2 Unknown Variance-Covariance Parameters

In this scenario, the covariance structure for the random effects is instead a function of θ, which is a vector of parameters, so that $\mathbf{G} = \mathbf{G}(\boldsymbol{\theta})$. Breslow and Clayton (1993) showed that the fixed effects $\boldsymbol{\beta}$ and random effects $\mathbf{U}$ can be simultaneously estimated by using the integrated quasi-likelihood function (IQLF)

$$\exp\left[Q(\boldsymbol{\beta}, \boldsymbol{\theta})\right] \propto |\mathbf{G}|^{-1/2} \int \exp\left[-\frac{1}{2\phi}\sum_{i=1}^{n} d_i(y_i; \mu_i) - \frac{1}{2}\mathbf{U}^\top \mathbf{G}^{-1}\mathbf{U}\right] d\mathbf{U},$$

where

$$d_i(y, \mu) = -2\int_y^\mu \frac{y - u}{a_i \text{Var}(u)} du$$

is the deviance measure of fit. If the IQLF is rewritten as

$$\exp\left[Q(\boldsymbol{\beta}, \boldsymbol{\theta})\right] \propto c|\mathbf{G}|^{-1/2} \int \exp\left[-\boldsymbol{\kappa}(\mathbf{U})\right] d\mathbf{U}$$

for some constant c and Laplace's integral approximation is applied, then

$$Q(\boldsymbol{\beta}, \boldsymbol{\theta}) \approx -\frac{1}{2}\log|\mathbf{D}| - \frac{1}{2}\log|\boldsymbol{\kappa}'(\tilde{\mathbf{U}})| - \boldsymbol{\kappa}(\tilde{\mathbf{U}}),$$

where $\tilde{\mathbf{U}}$ is the solution to

$$\boldsymbol{\kappa}'(\mathbf{U}) = -\sum_{i=1}^{n} \frac{(y_i - \mu_i)\mathbf{Z}_i}{\phi a_i \text{Var}(\mu_i)} g'(\mu_i) + \mathbf{G}^{-1}\mathbf{U} = 0$$

which minimizes $\boldsymbol{\kappa}(\mathbf{U})$. Taking the second partial derivative yields

$$\boldsymbol{\kappa}''(\mathbf{U}) \approx \sum_{i=1}^{n} \frac{\mathbf{Z}_i \mathbf{Z}_i^\top}{\phi a_i \mathrm{Var}(\mu_i) \left[g'(\mu_i)\right]^2} + \mathbf{G}^{-1} = \mathbf{Z}^\top \mathbf{W} \mathbf{Z} + \mathbf{G}^{-1},$$

where $\mathbf{W}$ is the diagonal matrix with terms $\mathbf{W}_i = \left(\phi a_i \mathrm{Var}(\mu_i) \left[g'(\mu_i)\right]^2\right)^{-1}$ that are same as that of the iterated weights in GLM. The quasi-likelihood could then be approximated as

$$Q(\boldsymbol{\beta}, \boldsymbol{\theta}) \approx -\frac{1}{2} \log |\mathbf{I} + \mathbf{Z}^\top \mathbf{W} \mathbf{Z} \mathbf{G}| - \frac{1}{2\phi} \sum_{i=1}^{n} d_i(y_i, \mu_i) - \frac{1}{2} \tilde{\mathbf{U}}^\top \mathbf{G}^{-1} \tilde{\mathbf{U}},$$

where the goal is to find $\tilde{\mathbf{U}}$ that maximizes the total of the last two terms. And differentiation with respect to $\boldsymbol{\beta}$ and $\mathbf{U}$ gives the following score equations:

$$\sum_{i=1}^{n} \frac{(y_i - \mu_i)\mathbf{X}_i}{\phi a_i \mathrm{Var}(\mu_i) g'(\mu_i)} = 0$$
$$\sum_{i=1}^{n} \frac{(y_i - \mu_i)\mathbf{Z}_i}{\phi a_i \mathrm{Var}(\mu_i) g'(\mu_i)} = \mathbf{G}^{-1}\mathbf{U}.$$

Using iterated weighted least squares (IWLS), the solution to the above would be equivalent to that for

$$\begin{bmatrix} \mathbf{X}^\top \mathbf{W} \mathbf{X} & \mathbf{X}^\top \mathbf{W} \mathbf{Z} \mathbf{G} \\ \mathbf{Z}^\top \mathbf{W} \mathbf{X} & \mathbf{I} + \mathbf{Z}^\top \mathbf{W} \mathbf{Z} \mathbf{G} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\nu} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \mathbf{W} \mathbf{Y} \\ \mathbf{Z}^\top \mathbf{W} \mathbf{Y} \end{bmatrix},$$

where $\mathbf{U} = \mathbf{G}\boldsymbol{\nu}$ and combining with the normality assumptions in LMM would yield the solutions:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{Y}$$
$$\hat{\mathbf{U}} = \mathbf{G}\mathbf{Z}^\top \mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

2.6.2 Functional Generalized Linear Mixed Model

The Functional Generalized Linear Mixed Model (FGLMM) accommodates a GLMM in the functional setting, where the rows of the design matrices in GLMM are now curves, and the fixed and random effects become functions. The responses are still scalar and the model is generally constructed as

$$g[\mathrm{E}(\mathbf{Y})] = \langle \mathbf{X}, \boldsymbol{\beta} \rangle + \langle \mathbf{Z}, \mathbf{U} \rangle.$$

When the design matrices and effects are scalar, the model is GLMM. If either effects are treated as functions, then the model is FGLMM. More of this model will be discussed in Section 3.1.

Chapter 3

Methodology

3.1 Functional Generalized Linear Mixed Model

Functional Generalized Linear Mixed Model (FGLMM) is an extension of Generalized Linear Mixed Model (GLMM) in functional setting. For each ith subject or sampling unit, we observe $(Y_i, \{t_l, X_{il}\})$ for $i = 1, \dots, n$, $l = 1, \dots, L$ where L number of measurements has been taken over the domain $\mathcal{T}$ where $t_l \in \mathcal{T}$. These $\{t_l, X_{il}\}$ are observed values of functional objects $X_i(t)$. Thus, our data can be viewed as $(Y_i, \{X_i(t), t \in \mathcal{T}\})$ for $i = 1, \dots, n$. We assume that the data form an i.i.d. sample. The dependent variable Y_i is a real-valued random variable that can be either continuous or discrete. The predictor variable $X_i(t), t \in \mathcal{T}$ is a random curve that is square integrable over $\mathcal{T}$. If we concern a logistic regression, then Y_i will be a binary variable, and if we concern a Poisson regression, then Y_i will be a count variable.

We write our FGLMM model as

$$g[\mathrm{E}(\mathbf{Y})] = \langle \mathbf{X}, \boldsymbol{\beta} \rangle + \langle \mathbf{Z}, \mathbf{U} \rangle. \tag{3.1}$$

The link function $g(\cdot)$ is assumed to be monotone and twice continuously differentiable function with bounded derivatives and thus invertible. We use inner product notation $\langle\cdot,\cdot\rangle$ to denote that this inner product can either be in $\mathbb{R}$ or in $L^2(t)$, depending on the corresponding $\mathbf{X}$ or $\mathbf{Z}$ being scalar or functional variable. If $\mathbf{X}$ is functional, the inner product $\langle\mathbf{X},\boldsymbol{\beta}\rangle = \int_{\mathcal{T}} \mathbf{X}(t)^\top \boldsymbol{\beta}(t)dt$, and if $\mathbf{X}$ is scalar, the inner product $\langle\mathbf{X},\boldsymbol{\beta}\rangle = \mathbf{X}^\top\boldsymbol{\beta}$.

3.1.1 Random Intercept Model

As an example of FGLMM, we introduce a random intercept model. For this model, we take $\mathbf{X} = (1, X(t))^\top$, and $\mathbf{Z} = 1$. The resulting form would be

$$g[\mathrm{E}(\mathbf{Y})] = \beta_0 + u_0 + \int_{\mathcal{T}} \beta_1(t)X(t)dt. \tag{3.2}$$

With Karhunen–Loéve exapnsion with basis system $\{\phi_k(t)\}_{k=1}^\infty$, we can represent the functional predictor variable and the functional coefficient as

$$X(t) = \sum_{k=1}^{\infty} x_k\phi_k(t),$$

$$\beta_1(t) = \sum_{k=1}^{\infty} b_{1k}\phi_k(t).$$

Assuming that $\{\phi_k(t)\}_{k=1}^\infty$ are orthonormal basis of functional space $L^2(t), t \in \mathcal{T}$, meaning that $\langle\phi_{k_1},\phi_{k_2}\rangle = \int_{\mathcal{T}} \phi_{k_1}(t)\phi_{k_2}(t)dt = 1$ if $k_1 = k_2$ and 0 if $k_1 \neq k_2$, we can have

$$\int_{\mathcal{T}} \beta_1(t)X(t)dt = \sum_{k=1}^{\infty} b_{1k}x_k.$$

We will approximate the model by a series of models with the expansion truncated at K. We assume that with large enough K, the remainder $\sum_{k=K+1}^{\infty} b_{1k}x_k$ becomes negligible.

$$g[\mathrm{E}(Y_{ik})] \approx \beta_0 + u_0 + \sum_{k=1}^{K} b_{1k} x_k.$$

3.1.2 Random Functional Slope Model

We also introduce a random functional slope model.

$$g[\mathrm{E}(\mathbf{Y})] = \beta_0 + \int_{\mathcal{T}} \beta_1(t) X(t) dt + \int_{\mathcal{T}} u_1(t) X(t) dt. \tag{3.3}$$

In this model, $u_1(t)$ is a random functional slope. We can consider $\mathbf{X} = (1, X(t))^\top$ and $\mathbf{Z} = X(t)$ from Equation 3.1. We assume that $u_1(t)$ follows random stochastic process, for example a Gaussian process. This means that if we take any finite observations of $u_1(t)$, $\{u_1(t_j); t_j \in \mathcal{T}, j = 1, \ldots, J\}$ will follow Gaussian distribution.

Here also, we take an orthonormal basis system $\{\phi_k(t)\}_{k=1}^{\infty}$. Then we can represent the functional objects as

$$X(t) = \sum_{k=1}^{\infty} x_k \phi_k(t),$$

$$\beta_1(t) = \sum_{k=1}^{\infty} b_{1k} \phi_k(t),$$

$$u_1(t) = \sum_{k=1}^{\infty} u_{1k} \phi_k(t).$$

We assume that any finite collection of u_{1k} will follow Normal distribution. With this representation and truncation at K, Equation 3.3 becomes

$$g[\mathrm{E}(Y_{ik})] \approx \beta_0 + \sum_{k=1}^{K} b_{1k} x_k + \sum_{k=1}^{K} u_{1k} x_k$$

where u_{1k} are taken as random effects.

3.2 Parameter Estimation

Both the Frequentist and Bayesian approaches to parameter estimation will be considered in fitting the FGLMM.

3.2.1 Frequentist Approach

The maximum likelihood estimator is considered for the Frequentist approach. We assume that Y follows an exponential family and the link function g is known and invertible. If we do not know the link function, then we should consider a quasi-likelihood model. We assume the random effect also follows Gaussian distribution.

For illustration, we use a random slope model with a binary response variable. We assume that there are q groups each with n observations. For $i = 1, \ldots, n$ and $j = 1, \ldots, q$,

$$
\begin{aligned}
Y_{ij}|u_j &\overset{ind}{\sim} Bernoulli(p_{ij}) \\
\log\left(\frac{p_{ij}}{1-p_{ij}}\right) &= \beta_0 + u_j + \int_{\mathcal{T}} \beta_1(t)X(t)dt \\
u_j &\overset{iid}{\sim} N(0, \sigma^2)
\end{aligned}
$$

With the truncation K where $K \to \infty$, we will use $\int_{\mathcal{T}} \beta_1(t)X_i(t)dt = \sum_{k=1}^{\infty} b_{1k}x_{ik} \approx \sum_{k=1}^{K} b_{1k}x_{ik}$ with appropriate basis $\{\phi_k\}_{k=1}^{K}$. Thus,

$$
\begin{aligned}
Y_{ij}|u_j &\overset{ind}{\sim} Bernoulli(p_{ij}) \\
\log\left(\frac{p_{ij}}{1-p_{ij}}\right) &= \beta_0 + u_j + \int_{\mathcal{T}} \sum_{k=1}^{K} b_{1k}x_{ik}dt \\
u_j &\overset{iid}{\sim} N(0, \sigma^2)
\end{aligned}
$$

And we write the truncated functional parameter as $\beta_{1;K}(t) = \sum_{k=1}^{K} b_{1k}\phi_k(t)$.

Then the likelihood function will look like

$$\begin{aligned}
L(\beta_0, \{b_{1k}\}_{k=1}^{K}, \sigma^2|\mathbf{Y}=\mathbf{y}) &= P(\mathbf{Y}=\mathbf{y}|\beta_0, \{b_{1k}\}_{k=1}^{K}, \sigma^2) \\
&= \int P(\mathbf{Y}=\mathbf{y}|\beta_0, \{b_{1k}\}_{k=1}^{K}, \sigma^2, \mathbf{u}) f(\mathbf{u}|\sigma^2) d\mathbf{u} \\
&= \int P(\mathbf{Y}=\mathbf{y}|\beta_0, \{b_{1k}\}_{k=1}^{K}, \mathbf{u}) f(\mathbf{u}|\sigma^2) d\mathbf{u} \\
&= \int \prod_{i=1}^{n}\prod_{j=1}^{q} P(Y_{ij}=y_{ij}|\beta_0, \{b_{1k}\}_{k=1}^{K}, \mathbf{u}) f(\mathbf{u}|\sigma^2) d\mathbf{u} \\
&= \prod_{j=1}^{q}\int \prod_{i=1}^{n} P(Y_{ij}=y_{ij}|\beta_0, \{b_{1k}\}_{k=1}^{K}, u_j) f(u_j|\sigma^2) du_j \\
&= \prod_{j=1}^{q}\int \prod_{i=1}^{n} P(Y_{ij}=y_{ij}|\beta_0, \{b_{1k}\}_{k=1}^{K}, u_j) f(u_j|\sigma^2) du_j.
\end{aligned}$$

As

$$p_{ij} = \frac{\exp(\beta_0 + u_j + \sum_{k=1}^{K} b_{1k}x_{ik})}{1+\exp(\beta_0 + u_j + \sum_{k=1}^{K} b_{1k}x_{ik})}$$

and

$$P(Y_{ij}=y_{ij}|p_{ij}) = \left(\frac{p_{ij}}{1-p_{ij}}\right)^{y}_{ij}(1-p_{ij}),$$

$$\begin{aligned}
&L(\beta_0, \{b_{1k}\}_{k=1}^{K}, \sigma^2|\mathbf{Y}=\mathbf{y}) \\
&= \prod_{j=1}^{q}\int \prod_{i=1}^{n} \frac{\exp(\beta_0 y_{ij} + u_j y_{ij} + \sum_{k=1}^{K} b_{1k}x_{ik}y_{ij})}{1+\exp(\beta_0 + u_j + \sum_{k=1}^{K} b_{1k}x_{ik})} f(u_j|\sigma^2) du_j \\
&= \prod_{j=1}^{q}\int \frac{\exp(\sum_{i=1}^{n} \beta_0 y_{ij} + u_j y_{ij} + \sum_{k=1}^{K} b_{1k}x_{ik}y_{ij})}{\prod_{i=1}^{n}(1+\exp(\beta_0 + u_j + \sum_{k=1}^{K} b_{1k}x_{ik}))} \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{1}{2}(\frac{u_j}{\sigma})^2\right)}\right) du_i.
\end{aligned}$$

The score of the likelihood function (i.e. $s(\theta) = \frac{\partial \log L(\theta)}{\partial\theta}$) for the above case will contain integration which can sometimes be numerically solved, sometimes not. Therefore, we use the Gauss-Hermite quadrature or the adaptive Gauss-Hermite quadrature to approximate the integration (Jin and Andersson, 2020). It seeks to numerically compute integrals of the form

$$I = \int \exp\left[p\, l(t)\right] dt,$$

where p is scalar and $l(t)$ is a unimodal function that does not depend on p. When t is assumed to be a $K \times 1$ vector, the Adaptive Gauss-Hermite quadrature (AGHQ) approximation with m quadrature points per dimension is given by

$$I^{\text{AGHQ}} = 2^{K/2} \det\left(\hat{L}\right) p^{-K/2} \sum_{j_1,\ldots,j_K=1}^{m} \left\{\prod_{i=1}^{K} w_{j_i} \exp(z_{j_i})\right\} \exp\left\{p\, l\left(2^{1/2}\hat{L} z_{j_1,\ldots,j_K} + \hat{t}\right)\right\},$$

where $z_{j_1}, \ldots, z_{j_K} = \left(z_{j_1}, \ldots, z_{j_K}\right)^{\top}$, $\hat{t}$ is the mode of $l(t)$ and $\hat{L}$ is the Cholesky decomposition of the inverse of $-\partial^2\left(\hat{t}\right) / \left(\partial t\, \partial t^{\top}\right)$. It follows that the AGHQ approximation is evaluated to the log-likelihood.

Using adaptive Gauss-Hermite quadrature will give $\hat{b}_{11}, \ldots, \hat{b}_{1K}$, or in general, $\hat{b}_1, \ldots, \hat{b}_K$. We can create our truncated estimator $\hat{\beta}_K$ from these estimates by $\hat{\beta}_K(t) = \sum_{k=1}^{K} \hat{b}_{1k}\phi_k(t)$. We'll show the consistency of this estimator.

Theorem 1 (Consistency for Frequentist Estimator) *Our Frequentist estimator $\hat{\beta}_K$ converges in probability to the true parameter β as $n \to \infty$ and $K \to \infty$.*

Proof. First, we fix the truncation number K. As in Proposition 3.2.1 of Bianconcini (2014), consistency of maximum likelihood estimators using adaptive Gauss-Hermite quadrature is established, $\hat{b}_{1k} \xrightarrow{\mathcal{P}} b_{1k}$ for every $k = 1, \ldots, K$ as $n \to \infty$. And this means that the truncated $\hat{\beta}_K \xrightarrow{\mathcal{P}} \beta_K$ as $n \to \infty$. Also, we have $\hat{\beta}_K \xrightarrow{\mathcal{P}} \beta$ as $K \to \infty$.

$$\begin{aligned} P\left(\|\hat{\beta}_K - \beta_1\|^2 \geq \epsilon\right) &\leq P\left(\|\hat{\beta}_K - \beta_K\|^2 + \|\beta_{K_n} - \beta_1\|^2 \geq \epsilon\right) \\ &\leq P\left(\|\hat{\beta}_K - \beta_K\|^2 \geq \frac{\epsilon}{2}\right) + P\left(\|\beta_K - \beta\|^2 \geq \frac{\epsilon}{2}\right) \\ &\to 0 + 0 \end{aligned}$$

as $n \to \infty$ and $K \to \infty$. $\square$

3.2.2 Bayesian Approach

The Bayesian approach rewrites the GLMM as

$$\begin{aligned} b_{11}, \ldots, b_{1K} &\overset{\text{iid}}{\sim} \text{MVN}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}\right) \\ p\left(\mathbf{Y} \mid \mathbf{X}_i, b_0, \mathbf{b}_1, u_j\right) &= \prod_{i=1}^{n} p\left(y_{ij} \mid \beta_0, \mathbf{b}_1^\top \mathbf{X}_i, u_j\right), \end{aligned}$$

where the observations from different groups are conditionally independent (Hoff, 2009).

With $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ being only the typical parameters that may have full conditional distributions in GLMMs, the Metropolis-Hastings algorithm is used to approximate the posterior distribution of the parameters. It combines Gibbs steps to update $(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ and Metropolis steps to update all b_{1k}.

In Gibbs sampling, the full conditional distributions of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ depend only on the $b_{11}, \ldots, b_{1K}$, which implies that $p(y_{ij} \mid \mathbf{b}_1^\top \mathbf{X}_i)$ has no effect on the full conditional distributions of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$. By that, their respective full conditional distributions will always be multivariate normal and inverse-Wishart.

The Metropolis step updates b_{1k} by proposing a new value b_{1k}^* based on the current parameter values, then decides to accept or reject it with appropriate probability. A proposed distribution here would be multivariate normal, with a mean being the current $b_{1k}^{(s)}$ value and a proposal variance, $V_{1k}^{(s)}$. The procedure works as follows:

1. Sample b_{1k}^{*} from a MVN $\left(b_{1k}^{(s)}, V_{1k}^{(s)}\right)$.

2. Calculate the acceptance ratio by using the formula:

$$r = \frac{p\left(y_{ij} \mid \mathbf{X}_i, b_{1k}^{*}\right) p\left(b_{1k}^{*} \mid \boldsymbol{\theta}^{(s)}, \boldsymbol{\Sigma}^{(s)}\right)}{p\left(y_{ij} \mid \mathbf{X}_i, b_{1k}^{(s)}\right) p\left(b_{1k}^{(s)} \mid \boldsymbol{\theta}^{(s)}, \boldsymbol{\Sigma}^{(s)}\right)}$$

3. Sample u from $U(0,1)$.

4. If $u < r$, then assign $b_{1k}^{(s+1)}$ to b_{1k}^{*}. Otherwise, assign $b_{1k}^{(s)}$.

Joining the two sampling procedures above, the Metropolis-Hastings algorithm for approximating the posterior $p(b_{11}, \ldots, b_{1K}, \boldsymbol{\theta}, \boldsymbol{\Sigma} \mid \mathbf{X}_1, \ldots, \mathbf{X}_n, \mathbf{Y})$ is given by:

1. Sample $\boldsymbol{\theta}^{(s+1)}$ from its full conditional distribution.

2. Sample $\boldsymbol{\Sigma}^{(s+1)}$ from its full conditional distribution.

3. For every $j \in \{1, \ldots, q\}$, follow the Metropolis steps.

From MCMC, we get $\{b_{1k}^{(m)}\}_{m\in\mathbb{N}}$ where m represents the number of Markov chain. Using them, we can obtain our estimator $\hat{\beta}_K^{(m)}(t) = \sum_{k=1}^{K} \bar{b}_{1k}^{(m)} \phi_k(t)$ where $\bar{b}_{1k}^{(m)} = \frac{1}{m}\sum_{l=1}^{m} b_{1k}^{(l)}$, the mean from the chain. We'll show the consistency of this estimator. We'll note them as $\{b_k^{(m)}\}_{m\in\mathbb{N}}$ and $\hat{\beta}_K^{(m)}$ to address the general case.

Theorem 2 (Consistency for Bayesian Estimator) *Our Bayesian estimator* $\hat{\beta}_K^{(m)}$ *converges in probability to the true parameter* β *as* $n \to \infty$, $K \to \infty$, *and* $m \to \infty$.

Proof. The proof follows Kang et al. (2023). First we fix K. Since $\{b_k^{(m)}\}_{m\in\mathbb{N}}$ is an ergodic and stationary Markov chain, we have their mean $\bar{b}_k^{(m)} \xrightarrow{\mathcal{P}} b_k$ as $m \to \infty$ using the law of large numbers. As

$$P\left(\|\hat{\beta}_K^{(m)} - \beta_K\|^2 \geq \frac{\epsilon}{2}\right) = P\left(\|\bar{b}_K^{(m)} - b_K\|_{\mathbb{R}^K}^2 \geq \frac{\epsilon}{2}\right) \to 0,$$

we have $\hat{\beta}_K^{(m)} \xrightarrow{\mathcal{P}} \beta_K$ as $m \to \infty$ for every $K \in \mathbb{N}$.

Since β_K is the truncated version of β, we have $\beta_K \xrightarrow{\mathcal{P}} \beta$ as $K \to \infty$ and this gives $P\left(\|\beta_K - \beta\|^2 \geq \frac{\epsilon}{2}\right) \to 0$.

$$\begin{aligned}
P\left(\|\hat{\beta}_K^{(m)} - \beta\|^2 \geq \epsilon\right) &\leq P\left(\|\hat{\beta}_K^{(m)} - \beta_K\|^2 + \|\beta_K - \beta\|^2 \geq \epsilon\right) \\
&\leq P\left(\|\hat{\beta}_K^{(m)} - \beta_K\|^2 \geq \frac{\epsilon}{2}\right) + P\left(\|\beta_K - \beta\|^2 \geq \frac{\epsilon}{2}\right) \\
&\to 0 + 0
\end{aligned}$$

as $n \to \infty$, $K \to \infty$, and $m \to \infty$. □

3.3 Computation

Before applying either of the two approaches, we modify the dataset. As we observe $(Y_i, \{t_l, X_{il}\})$ for $i = 1, \ldots, n$, $l = 1, \ldots, L$, we need to turn $\{t_l, X_{il}\}$ into functional objects. Using basis system $\{\phi_k(t)\}_{k=1}^K$, we find

$$x_{ik} = \underset{x_{ik}}{\arg\min} \sum_{i=1}^{n} \left(X_{il} - X_{i;K}(t_l)\right)^2 + \lambda \int \left[LX_{i;K}(t)\right] dt$$

where $X_{i;K}(t) = \sum_{k=1}^K x_{ik}\phi_k(t)$. $X_{i;K}(t)$ is a truncated approximation of true $X_i(t)$. The choice of K and λ can be determined using the generalized cross-validation score

(GCV). If ϕ_k are not orthonormal or if K is too large, we can do FPCA to reduce the dimension and obtain orthonormal eigenbasis. Then we can use the basis coefficients x_{ik} in our analysis as with orthonormal basis, $\int_{\mathbb{J}} \beta_1(t)X_i(t)dt = \sum_{k=1}^{\infty} b_{1k}x_{ik} \approx \sum_{k=1}^{K} b_{1k}x_{ik}$. Here are the steps.

1. *Creation of Functional Data Objects*

   The desired variables in the dataset are converted to functional data objects by smoothing the data using a roughness penalty. This requires appropriate selection of basis expansion and number of basis functions.

2. *Application of FPCA on the Functional Data Objects*

   With emphasis on the importance and helpfulness of dimension reduction, FPCA is performed. Carrying out FPCA allows for the number of independent variables to be lessened and might make the overall parameter estimation less complex.

3. *Recreate the dataset.*

   The recreated dataset includes the functional coefficients computed from the first step, but only the ones that were selected based on FPCA.

4. *Use appropriate estimation scheme.*

   4-1. Frequentist approach.

   The AGHQ approximation is the method used for GLMM estimation in the `R` function `glmer` from the `lme4` package. It is specified through the argument `nAGQ`, which is the number of quadrature points. If that number is 1, the AGHQ approximation becomes the Laplace integral approximation. If it is 0, it uses a faster but less exact form of parameter estimation and it optimizes the random and fixed effects in the penalized iteratively reweighted least squares. Any value larger than 1 computes at a higher accuracy but will take more time. For the random intercept models fitted in Section 4, `nAGQ` was set to 1 for the binary response data, and 0 for the count response data.

4-2. Bayesian approach.

The `jags.model` and `coda.samples` are the primary `R` functions from the `rjags` package that allows for the Bayesian approach to GLMM estimation. It requires the input prior distributions and data to be written in the (Bayesian inference Using Gibbs Sampling) BUGS-language description.

Chapter 4

Simulation Studies

This section discusses the results of simulating and fitting a GLMM over different functional settings.

4.1 Random Intercept Model

4.1.1 Simulation Setting

The outline of the GLMM data simulation is given below. We run the simulation 100 times. We have tried our simulation setting for the combinations of sample size $N = 25, 50, 100, 200$, the number of repetitions $q = 10, 15, 25$, and number of basis generating functional variables $nb = 3, 5, 7$. Our scalar response Y_{ij} will be with $i = 1, \ldots, N$, and $j = 1, \ldots, q$. Our functional variables $X(t)$ are observed at $t_k, k = 1, \ldots, K = 100$ equally spaced points over domain $[0, 1]$.

1. *Determine the Functional Coefficient*, $\beta(t)$.

   We create $\beta(t)$ coefficients of the predictors $X(t)$'s that would be estimated in the usual GLMM setting. We generate $\beta(t)$ as linear combinations of three Fourier basis functions and their coefficients are taken as $\boldsymbol{\beta} = (1, -1, 1)$.

2. *Generate Functional Predictor Observations* (t_k, X_{ik}).

   We first generate the true $X(t)$ using Fourier basis functions with coefficients following multivariate normal. We specify the number of basis generating (nb) functions to be used when generating an $N \times nb$ samples from the multivariate normal distribution. Then we evaluate the functions at each time points $t_k; k = 1, \ldots, K$ with $K = 100$ and we add a Gaussian error to this. Therefore, our observations $X_{ik} = X_i(t_k) + e_{ik}$ where $e_{ik} \sim N(0, 0.1^2)$.

3. *Generate the Response* Y_{ij}.

   We now generate the response Y_{ij} where $i = 1, \ldots, N$, and $j = 1, \ldots, q$. We first calculate the regression lines for random intercept model $g_{ij} = \beta_0 + u_{0,j} + \int_0^1 \beta_1(t) X_i(t) dt$. Here, we generate the random intercept $u_{0,j}$ times following $N(0, 2^2)$, $j = 1, \ldots, q$. We approximate $\int_0^1 \beta_1(t) X_i(t)$ by the trapezoidal rule. We note that here we use the true functions $X_i(t)$, not the measurements X_{ik}. Then we generate $Y_{ij} \sim Binom\left(1, \dfrac{\exp[g_{ij}]}{1 + \exp[g_{ij}]}\right)$ for binary response and $Y_{ij} \sim Pois(\exp(g_{ij}))$ for count response.

After the generation of data, we get $\{Y_{ij}\}$, $\{(t_k, X_{ik})\}$ where $i = 1, \ldots N$, $j = 1, \ldots q$, and $k = 1, \ldots, K$.

In the beginning of estimation, we turn these observations (t_k, X_{ik}) into functional objects $\hat{X}_i(t)$. We used Fourier basis for expansion. The number of basis functions is determined generalized cross validation (GCV), and penalty term is taken as the integrated squared second derivative.

Table 4.1 is a snapshot of binary response with settings $N = 25$, $q = 10$, and $nb = 3$. Since the number of basis given through GCV was also 3, (X1, X2, X3) represent the coefficients for linear combinations of those three Fourier basis for $\hat{X}_i(t)$. Table 4.1 will be used as an input for Frequentist or Bayesian estimation.

Table 4.1: Pre-processed Binary Response Data ($N = 25$, $q = 10$, $nb = 3$) For Model Fitting

	Subject	Item	Y	X1	X2	X3
1	1	1	1	1.0177	2.1211	2.3085
2	1	2	1	1.0177	2.1211	2.3085
3	1	3	1	1.0177	2.1211	2.3085
4	1	4	1	1.0177	2.1211	2.3085
5	1	5	1	1.0177	2.1211	2.3085
6	1	6	1	1.0177	2.1211	2.3085
7	1	7	1	1.0177	2.1211	2.3085
8	1	8	1	1.0177	2.1211	2.3085
9	1	9	1	1.0177	2.1211	2.3085
10	1	10	1	1.0177	2.1211	2.3085
11	2	1	1	0.5896	-2.6545	-2.5715
12	2	2	1	0.5896	-2.6545	-2.5715
13	2	3	0	0.5896	-2.6545	-2.5715
14	2	4	1	0.5896	-2.6545	-2.5715
15	2	5	1	0.5896	-2.6545	-2.5715
16	2	6	1	0.5896	-2.6545	-2.5715
17	2	7	1	0.5896	-2.6545	-2.5715
18	2	8	1	0.5896	-2.6545	-2.5715
19	2	9	1	0.5896	-2.6545	-2.5715
20	2	10	1	0.5896	-2.6545	-2.5715
⋮	⋮	⋮	⋮	⋮	⋮	⋮
241	25	1	0	2.1115	1.0890	-4.0833
242	25	2	0	2.1115	1.0890	-4.0833
243	25	3	0	2.1115	1.0890	-4.0833
244	25	4	1	2.1115	1.0890	-4.0833
245	25	5	0	2.1115	1.0890	-4.0833
246	25	6	1	2.1115	1.0890	-4.0833
247	25	7	0	2.1115	1.0890	-4.0833
248	25	8	1	2.1115	1.0890	-4.0833
249	25	9	0	2.1115	1.0890	-4.0833
250	25	10	0	2.1115	1.0890	-4.0833

4.1.2 Binary Response Results

The relevant metrics that describe and summarize the estimates for the Binary Response Random Intercept Model are shown in Tables 4.2 to 4.8. The Bayesian approach performed better when estimating the fixed intercept. Both approaches did good in estimating the functional coefficient. There seems to be opportunities in the estimation of the random intercept standard deviation for both methods.

Table 4.2: **Binary Response Data** MSE decreases as q increases for $(N = 25, nb = 3)$ for both approaches.

MSE	Random Intercept Standard Deviation							
	Frequentist				Bayesian			
N	nb	$q = 10$	$q = 15$	$q = 25$	nb	$q = 10$	$q = 15$	$q = 25$
25	3	0.3448	0.2263	0.1505	3	0.2682	0.0556	0.0027
	5	0.4388	0.2788	0.1203	5	0.6164	0.0014	0.1116
	7	0.3773	0.2714	0.1176	7	1.4727	0.0799	0.1232
50	3	0.3017	0.1611	0.1109	3	0.0006	0.0877	0.2128
	5	0.2165	0.1843	0.1246	5	0.0048	0.0011	0.0045
	7	0.2684	0.1588	0.1056	7	0.0001	0.0229	0.0352
100	3	0.2581	0.1516	0.0828	3	0.0041	0.0005	0.0003
	5	0.2027	0.1287	0.0860	5	0.0026	0.0034	0.0023
	7	0.2480	0.1636	0.1026	7	0.0045	0.0002	0.0003
200	3	0.2031	0.1689	0.0906	3	0.4257	0.0006	0.0008
	5	0.2733	0.1693	0.0946	5	0.0181	0.0106	0.0060
	7	0.2166	0.1611	0.1079	7	0.0009	0.0018	0.0003

4.1.3 Count Response Results

The relevant metrics that describe and summarize the estimates for the Count Response Random Intercept Model are shown in Tables 4.9 to 4.15. Both approaches appeared to have good estimates for the functional coefficient. The Frequentist approach seemed to have better estimates for the random intercept standard deviation and fixed intercept.

Table 4.3: **Binary Response Data** Majority of the average random intercept standard deviations are higher for Bayesian.

Mean	Random Intercept Standard Deviation							
	Frequentist				Bayesian			
N	nb	$q = 10$	$q = 15$	$q = 25$	nb	$q = 10$	$q = 15$	$q = 25$
25	3	2.4655	1.7311	2.0297	3	2.5178	1.7642	2.0515
	5	2.6773	1.9061	2.2929	5	2.7851	1.9625	2.3341
	7	3.0181	2.1987	2.2927	7	3.2136	2.2827	2.3510
50	3	1.9750	1.8778	2.2801	3	2.0244	2.2962	1.5387
	5	1.2266	1.3140	1.9576	5	2.0692	1.9668	1.9326
	7	1.9360	1.9055	1.7225	7	2.0094	2.1512	1.8124
100	3	1.3226	1.9373	1.9197	3	1.9361	1.9785	2.0160
	5	1.9011	2.0039	1.9962	5	1.9486	1.9416	1.9525
	7	2.1250	2.3310	2.0470	7	1.9326	1.9862	2.0177
200	3	1.2705	1.7301	1.9349	3	1.3476	1.9751	2.0283
	5	1.7843	1.7226	1.8998	5	1.8654	1.8968	1.9222
	7	1.3504	1.8332	1.6478	7	1.9702	1.9573	1.9828

Table 4.4: **Binary Response Data** All of the median random intercept standard deviations for Bayesian are closer to the actual standard deviation of 2, than Frequentist.

Median	Random Intercept Standard Deviation							
	Frequentist				Bayesian			
N	nb	$q = 10$	$q = 15$	$q = 25$	nb	$q = 10$	$q = 15$	$q = 25$
25	3	2.4375	1.7181	2.0353	3	2.4942	1.7574	2.0505
	5	2.6395	1.8963	2.2651	5	2.7208	1.9502	2.3036
	7	2.9080	2.1523	2.2752	7	3.1102	2.2336	2.3261
50	3	1.9493	1.8512	2.2605	3	1.9967	2.2848	1.5243
	5	1.2049	1.2963	1.9657	5	2.0658	1.9602	1.9039
	7	1.9328	1.9035	1.6998	7	1.9771	2.1266	1.7954
100	3	1.3263	1.9346	1.8925	3	1.8912	1.9875	2.0625
	5	1.8923	2.0008	1.9909	5	1.9176	1.9096	1.9545
	7	2.1240	2.2941	2.0520	7	1.8966	1.9394	1.9923
200	3	1.2732	1.7382	1.9318	3	1.3506	1.9356	2.0014
	5	1.7884	1.7211	1.9318	5	1.8473	1.8356	1.9310
	7	1.3632	1.8333	1.6458	7	1.8973	1.9591	1.9288

4.1.4 Model Comparison

The relevant metrics that describe and summarize the estimates for the comparison of the three models, FGLM, GLMM, and GLMM are shown in Tables 4.16 to 4.25. For

Table 4.5: **Binary Response Data** At $N = 100$ & $N = 200$, the Frequentist approach performed better for $N = 25, 50$ while the Bayesian approach did well when $N = 100, 200$.

MSE	Fixed Intercept							
	Frequentist				Bayesian			
N	nb	$q = 10$	$q = 15$	$q = 25$	nb	$q = 10$	$q = 15$	$q = 25$
25	3	0.5778	0.3493	0.2656	3	0.9936	0.0581	0.4523
	5	0.6072	0.3697	0.2146	5	0.2200	0.0001	0.0256
	7	0.7422	0.4326	0.2315	7	0.0080	0.3268	0.7005
50	3	0.3632	0.2450	0.1221	3	0.0135	0.8679	0.0218
	5	0.4076	0.2471	0.1323	5	0.0001	0.0437	0.5595
	7	0.6347	0.3585	0.1994	7	0.4758	0.1227	0.2312
100	3	0.4543	0.2918	0.2035	3	0.0034	0.0004	0.0026
	5	0.3643	0.2518	0.1521	5	0.0024	0.0000	0.0034
	7	0.4445	0.2813	0.1340	7	0.0008	0.0001	0.0000
200	3	0.4000	0.2146	0.1437	3	0.3269	0.0025	0.0021
	5	0.3232	0.2362	0.1655	5	0.0006	0.0004	0.0007
	7	0.4249	0.3295	0.1430	7	0.0006	0.0079	0.0003

Table 4.6: **Binary Response Data** At $q = 10$, most of the average of the fixed intercepts from Bayesian are close to the actual value of 1.

Mean	Fixed Intercept							
	Frequentist				Bayesian			
N	nb	$q = 10$	$q = 15$	$q = 25$	nb	$q = 10$	$q = 15$	$q = 25$
25	3	0.0091	1.2020	1.6434	3	0.0032	1.2411	1.6725
	5	0.4825	0.9669	1.1298	5	0.5309	1.0078	1.1599
	7	0.8366	0.4001	1.7772	7	0.9104	0.4283	1.8370
50	3	0.8620	0.8483	0.3257	3	0.8837	1.9316	0.8525
	5	0.1128	0.4767	1.1597	5	0.9928	0.7909	0.2520
	7	1.0878	1.2227	1.1280	7	1.6898	1.3503	1.4808
100	3	0.9413	1.0938	1.3269	3	1.0584	0.9811	1.0514
	5	2.4728	1.0938	1.3269	5	1.0494	1.0039	1.0584
	7	0.7472	0.8731	0.6069	7	1.0287	0.9891	0.9961
200	3	1.5622	0.9954	0.8153	3	1.5717	0.9500	0.9543
	5	0.4218	0.0844	0.3984	5	1.0249	1.0200	0.9914
	7	2.0119	1.1519	1.1426	7	0.9750	0.9113	0.9833

the count response data, due to the large values that influence MSE, the logarithmic

Table 4.7: **Binary Response Data** For $N = 100$, the medians for the Frequentist approach decrease as q increases when nb is either 5 or 7.

Median	Fixed Intercept							
	Frequentist				Bayesian			
N	nb	$q = 10$	$q = 15$	$q = 25$	nb	$q = 10$	$q = 15$	$q = 25$
25	3	0.0044	1.2028	1.6415	3	-0.0034	1.2476	1.6562
	5	0.5031	0.9872	1.1164	5	0.5379	1.0177	1.1476
	7	0.8251	0.3540	1.7763	7	0.8920	0.3922	1.8450
50	3	0.8703	0.8528	0.3342	3	0.8787	1.9061	0.8485
	5	0.1262	0.4692	1.1551	5	0.9929	0.7887	0.2487
	7	1.0823	1.2325	1.1312	7	1.6585	1.3457	1.4738
100	3	0.9409	1.0996	1.3305	3	1.0377	0.9100	1.0358
	5	2.4803	1.7494	1.6273	5	0.9954	0.9924	1.1113
	7	0.7340	0.8789	0.6037	7	0.9758	1.0592	1.0316
200	3	1.5549	0.9999	0.8156	3	1.5726	0.8653	0.9018
	5	0.4283	0.0861	0.3996	5	1.0402	1.0050	0.9314
	7	2.0200	1.1508	1.1442	7	0.9743	0.9509	1.0353

Table 4.8: **Binary Response Data** Based on the MSE's for the functional coefficient estimates, both approaches seem to work well for $N = 50$, $N = 100$, $N = 200$.

MSE	Functional Coefficient, $\beta_1(t)$							
	Frequentist				Bayesian			
N	nb	$q = 10$	$q = 15$	$q = 25$	nb	$q = 10$	$q = 15$	$q = 25$
25	3	0.1052	0.0761	0.0484	3	0.1403	0.0917	0.0652
	5	0.2265	0.1225	0.0703	5	0.2578	0.1432	0.0892
	7	0.3213	0.1662	0.0987	7	0.5730	0.2250	0.1275
50	3	0.0471	0.0279	0.0153	3	0.0498	0.0377	0.0171
	5	0.0729	0.0373	0.0264	5	0.0766	0.0451	0.0262
	7	0.1032	0.0650	0.0356	7	0.1418	0.0713	0.0371
100	3	0.0224	0.0148	0.0092	3	0.0244	0.0155	0.0097
	5	0.0320	0.0184	0.0122	5	0.0355	0.0195	0.0129
	7	0.0378	0.0265	0.0159	7	0.0416	0.0284	0.0168
200	3	0.0103	0.0065	0.0043	3	0.0118	0.0067	0.0044
	5	0.0158	0.0092	0.0053	5	0.0166	0.0094	0.0053
	7	0.0188	0.0130	0.0081	7	0.0195	0.0133	0.0083

score given by the formula below

$$\text{Logarithmic Score} = -\log\left(P_Y\left(Y = Y_{ij} \mid \lambda = \hat{Y}_{ij}\right)\right)$$

Table 4.9: **Count Response Data** Based on the MSE's, the Frequentist approach performed better in estimating the random intercept standard deviation.

MSE	Random Intercept Standard Deviation							
	Frequentist				Bayesian			
N	nb	$q=10$	$q=15$	$q=25$	nb	$q=10$	$q=15$	$q=25$
25	3	0.2139	0.1318	0.0902	3	0.4298	0.1434	1.0406
	5	0.1962	0.1134	0.0735	5	1.3161	0.6396	0.4405
	7	0.2752	0.1532	0.0759	7	1.0432	0.0168	0.5162
50	3	0.2358	0.1487	0.0859	3	0.8461	0.5609	0.6357
	5	0.1606	0.1274	0.0981	5	0.4975	0.1989	0.1940
	7	0.2292	0.1394	0.0783	7	0.4242	0.2885	0.6813
100	3	0.2021	0.1435	0.0717	3	0.8494	0.7093	1.2520
	5	0.1739	0.1185	0.0813	5	1.3664	0.1155	0.5092
	7	0.2304	0.1487	0.0827	7	0.6921	1.0550	0.3314
200	3	0.1874	0.1461	0.0778	3	1.0205	0.9232	1.2728
	5	0.2450	0.1555	0.0964	5	0.8928	1.0328	1.0019
	7	0.2021	0.1404	0.0944	7	0.8894	0.9734	1.1339

Table 4.10: **Count Response Data** Aside from when $nb = 3$ and $q = 10$ the mean values of the estimated random intercept standard deviation using the Frequentist approach are very close to the actual value of 2, for $N = 100$.

Mean	Random Intercept Standard Deviation							
	Frequentist				Bayesian			
N	nb	$q=10$	$q=15$	$q=25$	nb	$q=10$	$q=15$	$q=25$
25	3	2.3429	1.7065	2.0256	3	2.6555	2.3787	3.0201
	5	1.9975	2.0678	1.7995	5	3.1472	2.7997	2.6637
	7	1.7095	1.6829	2.1980	7	3.0214	2.1297	2.7185
50	3	1.9129	1.8677	2.2670	3	2.9198	2.7489	2.7973
	5	1.2147	1.2837	1.9329	5	2.7053	2.4460	2.4405
	7	1.8604	1.8328	1.6870	7	2.6513	2.5371	2.8254
100	3	1.3254	1.9685	1.9229	3	2.9216	2.8422	3.1189
	5	1.8738	1.9955	1.9726	5	3.1689	2.3399	2.7136
	7	2.0925	2.2751	2.0480	7	2.8319	3.0271	2.5757
200	3	1.2632	1.7167	1.9192	3	3.0102	2.9609	3.1282
	5	1.7774	1.7293	1.8855	5	2.9449	3.0162	3.0010
	7	1.3254	1.8275	1.6386	7	2.9431	2.9866	3.0648

was instead used as a metric to compare the prediction performance of the three models. Whereas for the binary response data, the accuracy was calculated as the proportion of 1's upon assigning 1 for when the predicted $p_{ij} \geq 0.5$ and 0 otherwise.

Table 4.11: **Count Response Data** Except for when $N = 25$, $q = 15$, and $nb = 7$, the median values of the estimated random intercept standard deviation by the Bayesian approach are way above the actual value of 2.

Median	Random Intercept Standard Deviation							
	Frequentist				Bayesian			
N	nb	$q = 10$	$q = 15$	$q = 25$	nb	$q = 10$	$q = 15$	$q = 25$
25	3	2.3407	1.7034	2.0253	3	2.4329	2.0083	2.8103
	5	1.9939	2.0676	1.7976	5	2.8450	2.5524	2.5276
	7	1.7070	1.6757	2.1963	7	2.8914	1.8649	2.5276
50	3	1.9113	1.8689	2.2642	3	2.7188	2.5415	2.6945
	5	1.2159	1.2856	1.9332	5	2.4846	2.3075	2.2092
	7	1.8607	1.8333	1.6876	7	2.4378	2.3022	2.7378
100	3	1.3243	1.9665	1.9236	3	2.8635	2.7167	3.0099
	5	1.8741	1.9955	1.9724	5	2.9752	2.1399	2.5583
	7	2.0919	2.2750	2.0479	7	2.6281	2.8339	2.4258
200	3	1.2628	1.7168	1.9186	3	2.7710	2.7774	2.9421
	5	1.7768	1.7294	1.8854	5	2.6712	2.8737	2.7725
	7	1.3253	1.8275	1.6381	7	2.7358	2.7059	2.8498

Table 4.12: **Count Response Data** The Frequentist approach clearly performed better in estimating the fixed intercept.

MSE	Fixed Intercept							
	Frequentist				Bayesian			
N	nb	$q = 10$	$q = 15$	$q = 25$	nb	$q = 10$	$q = 15$	$q = 25$
25	3	0.5330	0.3255	0.2099	3	3.4073	2.6885	3.0421
	5	0.4233	0.2743	0.1897	5	2.9118	1.5879	3.7211
	7	0.5238	0.3395	0.1571	7	4.2994	2.3088	1.6623
50	3	0.3408	0.2317	0.1147	3	5.3675	7.4139	3.5561
	5	0.3249	0.2308	0.1245	5	0.5673	6.9103	0.6145
	7	0.5091	0.3054	0.1662	7	6.0907	4.0310	4.6020
100	3	0.4380	0.2884	0.1956	3	7.1555	2.2410	5.2155
	5	0.3245	0.2494	0.1324	5	0.8140	3.7209	0.9523
	7	0.3970	0.2671	0.1296	7	0.2207	3.9783	2.4589
200	3	0.3886	0.2273	0.1380	3	4.5950	4.1096	4.9630
	5	0.3167	0.2309	0.1605	5	4.1742	4.8064	4.3998
	7	0.4135	0.3170	0.1351	7	3.2567	3.7199	4.3455

Table 4.13: **Count Response Data** For the Frequentist approach, the combination of $N = 50$ and $nb = 7$ appear to have the mean estimated fixed intercepts be close to actual value of 1.

Mean	Fixed Intercept							
	Frequentist				Bayesian			
N	nb	$q = 10$	$q = 15$	$q = 25$	nb	$q = 10$	$q = 15$	$q = 25$
25	3	0.0106	0.7312	1.4093	3	-0.8459	-0.6397	-0.7442
	5	1.6969	1.5060	0.9732	5	-0.7064	-0.2601	-0.9290
	7	1.4418	0.4727	0.9722	7	-1.0735	-0.5195	-0.2893
50	3	0.8486	0.8417	0.3182	3	-1.3168	-1.7228	-0.8858
	5	0.0966	0.4501	1.1592	5	0.2468	-1.6287	0.2161
	7	1.0742	1.1877	1.1002	7	-1.4679	-1.0077	-1.1452
100	3	0.9548	1.0905	1.3233	3	-1.6750	-0.4970	-1.2838
	5	2.4387	1.7489	1.6140	5	0.0978	-0.9290	0.0241
	7	0.7294	0.8507	0.6119	7	1.4698	-0.9948	-0.5681
200	3	1.5521	0.9892	0.8025	3	-1.1436	-1.0272	-1.2278
	5	0.4237	0.0837	0.3938	5	-1.0431	-1.1923	-1.0976
	7	1.9947	1.1404	1.1413	7	-0.8046	-0.9287	-1.0846

Table 4.14: **Count Response Data** Majority of the median of the estimated fixed intercepts when $q = 25$ are quite near the actual value of 1, under the Frequentist approach.

Median	Fixed Intercept							
	Frequentist				Bayesian			
N	nb	$q = 10$	$q = 15$	$q = 25$	nb	$q = 10$	$q = 15$	$q = 25$
25	3	0.0073	0.7291	1.4108	3	-0.6856	-0.3454	-0.6078
	5	1.6968	1.5048	0.9739	5	-0.4844	-0.0896	-0.9071
	7	1.4450	0.4755	0.9704	7	-1.0331	-0.2886	-0.2585
50	3	0.8483	0.8423	0.3183	3	-1.2292	-1.6143	-0.8807
	5	0.1003	0.4526	1.1612	5	0.2344	-1.6465	0.2693
	7	1.0722	1.1874	1.1097	7	-1.3399	-0.8822	-1.1724
100	3	0.9534	1.0891	1.3246	3	-1.6969	-0.5260	-1.2472
	5	2.4372	1.7501	1.6132	5	0.1298	-0.7982	-0.0635
	7	0.7296	0.8501	0.6112	7	1.6402	-0.9383	-0.5971
200	3	1.5519	0.9875	0.8020	3	-1.1173	-1.0750	-1.1815
	5	0.4233	0.0830	0.3948	5	-0.9360	-1.1444	-1.0655
	7	1.9951	1.1404	1.1423	7	-0.9977	-0.7274	-1.0326

Table 4.15: **Count Response Data** Both approaches performed well in estimating the functional coefficient.

MSE	Functional Coefficient, $\beta_1(t)$							
	Frequentist				Bayesian			
N	nb	$q = 10$	$q = 15$	$q = 25$	nb	$q = 10$	$q = 15$	$q = 25$
25	3	0.0004	0.0002	0.0002	3	0.0018	0.0003	0.0052
	5	0.0006	0.0004	0.0003	5	0.0047	0.0012	0.0247
	7	0.0015	0.0011	0.0005	7	0.0067	0.0278	0.0262
50	3	0.0002	0.0001	0.0001	3	0.0048	0.0002	0.0004
	5	0.0002	0.0002	0.0001	5	0.0002	0.0003	0.0058
	7	0.0003	0.0002	0.0002	7	0.0008	0.0032	0.0062
100	3	0.0001	0.0001	0.0001	3	0.0001	0.0001	0.0001
	5	0.0001	0.0001	0.0001	5	0.0001	0.0001	0.0001
	7	0.0001	0.0001	0.0001	7	0.0005	0.0001	0.0001
200	3	0.0000	0.0000	0.0000	3	0.0000	0.0000	0.0000
	5	0.0001	0.0001	0.0001	5	0.0001	0.0001	0.0001
	7	0.0001	0.0001	0.0001	7	0.0001	0.0001	0.0001

Table 4.16: **Count Response Data** The assumed underlying distribution is poisson, so the logarithmic score takes the negative log of the assigned probability using the poisson pmf to an observed response given that the mean is the estimated response. It means that that the closer the estimate is to the actual value, the higher the probability, and the lower the logarithmic score. In any case, FGLMM performed best.

Log Score	Predicted Y ($\hat{Y}$)				
	Frequentist, 1000 Simulations				
N	nb	Model	$q = 10$	$q = 15$	$q = 25$
25	3	FGLM	2729.5229	2857.2455	2769.7430
		GLMM	2448.4714	2171.1622	2112.2923
		FGLMM	2.0273	2.0378	2.0383
	5	FGLM	1607.5917	1701.5307	2686.0154
		GLMM	1490.2607	1664.1315	1930.9377
		FGLMM	2.0251	2.0339	2.0375
	7	FGLM	2356.8517	2556.1595	2519.4851
		GLMM	2307.2538	2211.2136	2072.2850
		FGLMM	2.0072	2.0189	2.0170
50	3	FGLM	1881.8495	2140.5976	2554.0509
		GLMM	2037.3829	2233.3366	2263.0955
		FGLMM	2.0541	2.0586	2.0640
	5	FGLM	2662.7187	2766.0124	3091.1086
		GLMM	2791.5692	2436.6150	2308.2403
		FGLMM	2.0468	2.0410	2.0401
	7	FGLM	1512.7567	2530.9869	2680.9133
		GLMM	2028.6597	2400.8565	2493.0049
		FGLMM	2.0360	2.0367	2.0439
100	3	FGLM	1928.3181	2103.1492	2485.6178
		GLMM	2707.5405	2701.2378	2822.2093
		FGLMM	2.0886	2.0888	2.0969
	5	FGLM	3919.7225	3726.5886	3527.0234
		GLMM	2715.6362	2807.7284	2867.8125
		FGLMM	2.0728	2.0702	2.0798
	7	FGLM	3843.9564	3637.3017	3388.4254
		GLMM	4642.6298	4043.1137	3476.9340
		FGLMM	2.0591	2.0599	2.0574
200	3	FGLM	2526.6866	2631.5458	2878.5550
		GLMM	3702.5964	3531.1081	3444.3256
		FGLMM	2.1221	2.1237	2.1180
	5	FGLM	3318.2945	3499.2631	3466.5280
		GLMM	3298.2110	3580.4438	3549.6687
		FGLMM	2.0776	2.0892	2.0993
	7	FGLM	3311.0659	3352.1753	3131.6957
		GLMM	3622.5830	3569.8599	3389.6367
		FGLMM	2.1005	2.0908	2.0976

Table 4.17: **Binary Response Data** FGLMM appears to be the best model among the three for the binary response data, since accuracy values calculated for it are at minimum around 0.86 and the best accuracy for either FGLM or GLMM is about 0.79.

Accuracy	Predicted Y ($\hat{Y}$)				
	Frequentist, 1000 Simulations				
N	nb	Model	$q = 10$	$q = 15$	$q = 25$
25	3	FGLM	0.7857	0.7835	0.7810
		GLMM	0.7572	0.7583	0.7582
		FGLMM	0.8757	0.8750	0.8744
	5	FGLM	0.7895	0.7856	0.7828
		GLMM	0.7564	0.7564	0.7574
		FGLMM	0.8780	0.8758	0.8746
	7	FGLM	0.7911	0.7854	0.7831
		GLMM	0.7555	0.7577	0.7562
		FGLMM	0.8799	0.8776	0.8762
50	3	FGLM	0.7880	0.7837	0.7815
		GLMM	0.7485	0.7492	0.7494
		FGLMM	0.8703	0.8695	0.8694
	5	FGLM	0.7894	0.7851	0.7822
		GLMM	0.7486	0.7480	0.7480
		FGLMM	0.8717	0.8701	0.8697
	7	FGLM	0.7901	0.7862	0.7830
		GLMM	0.7498	0.7492	0.7494
		FGLMM	0.8728	0.8718	0.8709
100	3	FGLM	0.7860	0.7823	0.7802
		GLMM	0.7448	0.7457	0.7463
		FGLMM	0.8663	0.8671	0.8665
	5	FGLM	0.7857	0.7821	0.7811
		GLMM	0.7454	0.7454	0.7465
		FGLMM	0.8673	0.8665	0.8668
	7	FGLM	0.7860	0.7827	0.7798
		GLMM	0.7446	0.7451	0.7455
		FGLMM	0.8679	0.8669	0.8667
200	3	FGLM	0.7861	0.7834	0.7807
		GLMM	0.7438	0.7443	0.7436
		FGLMM	0.8651	0.8653	0.8646
	5	FGLM	0.7866	0.7842	0.7813
		GLMM	0.7418	0.7426	0.7432
		FGLMM	0.8648	0.8646	0.8646
	7	FGLM	0.7860	0.7833	0.7803
		GLMM	0.7441	0.7438	0.7447
		FGLMM	0.8654	0.8652	0.8655

Table 4.18: **Count Response Data** Here, FGLMM is still the clear winner among the three models since its MSE values across the board are well lower than that of the other two models and these values are all even smaller than 1. It is also observed that it gets better with the increasng values of q.

Mean (MSE)	Fixed Intercept (Actual: $\beta_0 = 1$) Frequentist, 1000 Simulations				
N	nb	Model	$q = 10$	$q = 15$	$q = 25$
25	3	FGLM	2.4763 (3.1180)	2.6259 (3.3406)	2.7101 (3.3993)
		GLMM	3.1979 (8.0662)	3.2164 (7.9731)	3.2164 (7.9344)
		FGLMM	0.9822 (0.4409)	1.0004 (0.2873)	1.0014 (0.1776)
	5	FGLM	2.4733 (3.0537)	2.5878 (3.1776)	2.7017 (3.4054)
		GLMM	3.2365 (8.1070)	3.2468 (8.0626)	3.2502 (8.0357)
		FGLMM	1.0165 (0.3993)	1.0265 (0.2755)	1.0301 (0.1746)
	7	FGLM	2.4341 (2.9113)	2.5789 (3.1527)	2.6877 (3.3496)
		GLMM	3.2080 (8.2564)	3.2227 (8.1857)	3.2084 (7.9923)
		FGLMM	0.9887 (0.3944)	1.0055 (0.2597)	0.9913 (0.1586)
50	3	FGLM	2.4523 (2.9333)	2.5785 (3.1550)	2.7120 (3.4165)
		GLMM	3.3312 (11.1046)	3.3394 (11.1169)	3.3416 (10.8174)
		FGLMM	0.9822 (0.4051)	0.9907 (0.2664)	0.9924 (0.1397)
	5	FGLM	2.4543 (3.0108)	2.5409 (3.0317)	2.6555 (3.1862)
		GLMM	3.3518 (11.9179)	3.3382 (11.7692)	3.3314 (11.5712)
		FGLMM	1.0076 (0.4181)	0.9940 (0.2745)	0.9872 (0.1561)
	7	FGLM	2.4503 (2.9900)	2.5728 (3.2083)	2.7079 (3.4627)
		GLMM	3.3437 (12.1834)	3.3347 (12.0274)	3.3459 (12.0883)
		FGLMM	1.0021 (0.4176)	0.9939 (0.2760)	1.0060 (0.1713)

Table 4.19: **Count Response Data** Upon observation, the pattern continues for FGLMM, the MSE decreases with increasing q.

Mean (MSE)	Fixed Intercept (Actual: $\beta_0 = 1$)				
	Frequentist, 1000 Simulations				
N	nb	Model	$q = 10$	$q = 15$	$q = 25$
100	3	FGLM	2.4958 (3.1053)	2.6141 (3.2994)	2.7430 (3.5366)
		GLMM	3.6946 (8.8768)	3.6878 (8.7193)	3.7000 (8.6912)
		FGLMM	1.0069 (0.3705)	1.0003 (0.2530)	1.0120 (0.1536)
	5	FGLM	2.4833 (3.0574)	2.6041 (3.2247)	2.7232 (3.4595)
		GLMM	3.6933 (8.9080)	3.6849 (8.7346)	3.7036 (8.7763)
		FGLMM	1.0054 (0.3789)	0.9968 (0.2520)	1.0155 (0.1596)
	7	FGLM	2.5190 (3.2831)	2.6451 (3.4636)	2.7385 (3.5533)
		GLMM	3.6789 (8.9313)	3.6829 (8.7734)	3.67772 (8.6053)
		FGLMM	0.9879 (0.4022)	0.9921 (0.2724)	0.9862 (0.1601)
200	3	FGLM	2.4610 (3.0085)	2.5949 (3.1788)	2.6919 (3.3234)
		GLMM	3.7559 (10.6122)	3.7653 (10.5246)	3.7528 (10.3156)
		FGLMM	0.9975 (0.4038)	1.0071 (0.2654)	0.9946 (0.1547)
	5	FGLM	2.3890 (2.7347)	2.5170 (2.9141)	2.6451 (3.1758)
		GLMM	3.7254 (10.4892)	3.7434 (10.4911)	3.7524 (10.3944)
		FGLMM	0.9621 (0.3804)	0.9805 (0.2566)	0.9895 (0.1639)
	7	FGLM	2.4873 (3.0934)	2.5794 (3.1924)	2.7141 (3.3992)
		GLMM	3.7728 (10.9242)	3.7546 (10.7252)	3.7706 (10.7452)
		FGLMM	1.0114 (0.3970)	0.9934 (0.2711)	1.0092 (0.1565)

Table 4.20: **Count Response Data** The two models that have a random intercept are only GLMM and FGLMM and it appears here that both models performed just as good as each other and have relatively very low MSE values.

Mean (MSE)	Random Intercept SD (Actual: SD(u_0) = 2)				
	Frequentist, 1000 Simulations				
N	nb	Model	$q = 10$	$q = 15$	$q = 25$
25	3	GLMM	1.8643 (0.2195)	1.9208 (0.1422)	1.9460 (0.0766)
		FGLMM	1.8646 (0.2198)	1.9211 (0.1424)	1.9462 (0.0766)
	5	GLMM	1.8355 (0.2202)	1.8861 (0.1456)	1.9225 (0.0896)
		FGLMM	1.8357 (0.2204)	1.8863 (0.1457)	1.9228 (0.0898)
	7	GLMM	1.8384 (0.2222)	1.8977 (0.1432)	1.9390 (0.0887)
		FGLMM	1.8387 (0.2226)	1.8980 (0.1434)	1.9393 (0.0888)
50	3	GLMM	1.8440 (0.2109)	1.9019 (0.1521)	1.9493 (0.0859)
		FGLMM	1.8441 (0.2110)	1.9020 (0.1521)	19495 (0.0861)
	5	GLMM	1.8284 (0.2367)	1.8775 (0.1616)	1.9262 (0.0866)
		FGLMM	1.8286 (0.2368)	1.8776 (0.1617)	1.9264 (0.0867)
	7	GLMM	1.8452 (0.2284)	1.9022 (0.1499)	1.9454 (0.0861)
		FGLMM	1.8452 (0.2284)	1.9023 (0.1500)	1.9455 (0.0861)

Table 4.21: **Count Response Data** Even with increased N, the two models GLMM and FGLMM performed just as good as each other.

Mean (MSE)	Random Intercept SD (Actual: SD(u_0) = 2)				
	Frequentist, 1000 Simulations				
N	nb	Model	$q = 10$	$q = 15$	$q = 25$
100	3	GLMM	1.8516 (0.2155)	1.9135 (0.1453)	1.9591 (0.0793)
		FGLMM	1.8517 (0.2157)	1.9136 (0.1454)	1.9592 (0.0794)
	5	GLMM	1.8582 (0.2092)	1.9121 (0.1330)	1.9457 (0.0840)
		FGLMM	1.8583 (0.2092)	1.9122 (0.1331)	1.9458 (0.0840)
	7	GLMM	1.8619 (0.2395)	1.9176 (0.1526)	1.9580 (0.0826)
		FGLMM	1.8620 (0.2397)	1.9177 (0.1597)	1.9581 (0.0827)
200	3	GLMM	1.8428 (0.2205)	1.8978 (0.1414)	1.9360 (0.0830)
		FGLMM	1.8428 (0.2205)	1.8978 (0.1414)	1.9360 (0.0830)
	5	GLMM	1.8205 (0.2178)	1.8719 (0.1397)	1.9215 (0.0852)
		FGLMM	1.8205 (0.2178)	1.8719 (0.1397)	1.9215 (0.0852)
	7	GLMM	1.8491 (0.2170)	1.9017 (0.1396)	1.9517 (0.0813)
		FGLMM	1.8491 (0.2170)	1.9017 (0.1397)	1.9517 (0.0813)

Table 4.22: **Binary Response Data** On the contrary, FGLMM performs better on the estimated random intercept for this type of data. The MSE values also appear to improve with increasing combinations of N and q.

Mean (MSE)	Random Intercept SD (Actual: SD(u_0) = 2)				
	Frequentist, 1000 Simulations				
N	nb	Model	$q = 10$	$q = 15$	$q = 25$
25	3	GLMM	1.0307 (1.0727)	1.0594 (0.9849)	1.0773 (0.9120)
		FGLMM	1.9282 (0.3602)	1.9598 (0.2469)	1.9694 (0.1285)
	5	GLMM	1.0136 (1.1075)	1.0366 (1.0228)	1.0571 (0.9515)
		FGLMM	1.9363 (0.3908)	1.9430 (0.2571)	1.9526 (0.1449)
	7	GLMM	1.0053 (1.1264)	1.0470 (1.0023)	1.0627 (0.9422)
		FGLMM	1.9633 (0.4056)	1.9842 (0.2388)	1.9791 (0.1445)
50	3	GLMM	1.0239 (1.0487)	1.0583 (0.9568)	1.0889 (0.8761)
		FGLMM	1.8720 (0.2765)	1.9164 (0.1870)	1.9632 (0.1130)
	5	GLMM	1.0202 (1.0661)	1.0491 (0.9755)	1.0769 (0.8974)
		FGLMM	1.8833 (0.3137)	1.9203 (0.2019)	1.9512 (0.1100)
	7	GLMM	1.0266 (1.0472)	1.0573 (0.9538)	1.0849 (0.8829)
		FGLMM	1.9028 (0.2893)	1.9385 (0.1855)	1.9722 (0.1121)

Table 4.23: **Binary Response Data** The pattern continues with increased N, but this time the observed MSEs for $N = 100, 200$ are relatively smaller than that for when $N = 25, 50$.

Mean (MSE)	Random Intercept SD (Actual: SD(u_0) = 2)				
	Frequentist, 1000 Simulations				
N	nb	Model	$q = 10$	$q = 15$	$q = 25$
100	3	GLMM	1.0414 (0.9998)	1.0778 (0.9053)	1.1027 (0.8386)
		FGLMM	1.8597 (0.2440)	1.9268 (0.1588)	1.9643 (0.0897)
	5	GLMM	1.0450 (0.9905)	1.0780 (0.9012)	1.0968 (0.8511)
		FGLMM	1.8751 (0.2371)	1.9231 (0.1474)	1.9543 (0.0958)
	7	GLMM	1.0480 (0.9965)	1.0801 (0.9066)	1.1040 (0.8388)
		FGLMM	1.8898 (0.2721)	1.9336 (0.1757)	1.9709 (0.0955)
200	3	GLMM	1.0423 (0.9940)	1.0742 (0.9081)	1.0948 (0.8498)
		FGLMM	1.8481 (0.2391)	1.9030 (0.1519)	1.9382 (0.0871)
	5	GLMM	1.0279 (1.0157)	1.0573 (0.9353)	1.0868 (0.8638)
		FGLMM	1.8309 (0.2282)	1.8777 (0.1455)	1.9259 (0.0871)
	7	GLMM	1.0445 (0.9859)	1.0748 (0.9048)	1.1053 (0.8311)
		FGLMM	1.8567 (0.2308)	1.9107 (0.1463)	1.9576 (0.0858)

Table 4.24: **Count Response Data** The functional treatment on the coefficient appears to be better than the conventional scalar, with MSE values of both the FGLM and FGLMM being all very close to 0.

MSE	Coefficient (β_1: Scalar, $\beta_1(t)$: Functional)				
	Frequentist, 1000 Simulations				
N	nb	Model	$q = 10$	$q = 15$	$q = 25$
25	3	FGLM	0.0003	0.0003	0.0002
		GLMM	2.5455	2.5444	2.5452
		FGLMM	0.0003	0.0003	0.0002
	5	FGLM	0.0007	0.0005	0.0003
		GLMM	2.5314	2.5312	2.5306
		FGLMM	0.0007	0.0005	0.0003
	7	FGLM	0.0015	0.0008	0.0006
		GLMM	2.5451	2.5454	2.5453
		FGLMM	0.0015	0.0008	0.0006
50	3	FGLM	0.0002	0.0001	0.0001
		GLMM	2.6620	2.6720	2.6661
		FGLMM	0.0002	0.0001	0.0001
	5	FGLM	0.0002	0.0002	0.0001
		GLMM	2.7377	2.7423	2.7430
		FGLMM	0.0002	0.0002	0.0001
	7	FGLM	0.0003	0.0002	0.0002
		GLMM	2.7830	2.7859	2.7909
		FGLMM	0.0003	0.0002	0.0002
100	3	FGLM	0.0001	0.0001	0.0001
		GLMM	2.2224	2.2224	2.2224
		FGLMM	0.0001	0.0001	0.0001
	5	FGLM	0.0001	0.0001	0.0001
		GLMM	2.2215	2.2214	2.2214
		FGLMM	0.0001	0.0001	0.0001
	7	FGLM	0.0001	0.0001	0.0001
		GLMM	2.2189	2.2189	2.2189
		FGLMM	0.0001	0.0001	0.0001
200	3	FGLM	0.0001	0.0001	0.0001
		GLMM	2.3175	2.3178	2.3175
		FGLMM	0.0001	0.0001	0.0001
	5	FGLM	0.0001	0.0001	0.0001
		GLMM	2.3181	2.3182	2.3181
		FGLMM	0.0001	0.0001	0.0001
	7	FGLM	0.0001	0.0001	0.0001
		GLMM	2.3394	2.3400	2.3397
		FGLMM	0.0001	0.0001	0.0001

Table 4.25: **Binary Response Data** The MSE values for GLMM appear to be relatively at the same level as with the count data. Still, the functional coefficients appear to be the better choice over scalar.

MSE	Coefficient (β_1: Scalar, $\beta_1(t)$: Functional)				
	Frequentist, 1000 Simulations				
N	nb	Model	$q = 10$	$q = 15$	$q = 25$
25	3	FGLM	0.3655	0.3680	0.3721
		GLMM	2.2080	2.2063	2.2061
		FGLMM	0.1195	0.0782	0.0448
	5	FGLM	0.3647	0.3633	0.3609
		GLMM	2.2062	2.2063	2.2062
		FGLMM	0.2008	0.1169	0.0675
	7	FGLM	0.3831	0.3827	0.3682
		GLMM	2.2113	2.2079	2.2032
		FGLMM	0.3088	0.1766	0.0990
50	3	FGLM	0.3433	0.3637	0.3723
		GLMM	2.1980	2.1986	2.1938
		FGLMM	0.0521	0.0316	0.0184
	5	FGLM	0.3417	0.3490	0.3631
		GLMM	2.1970	2.1967	2.1958
		FGLMM	0.0733	0.0466	0.0265
	7	FGLM	0.3539	0.3597	0.3679
		GLMM	2.1981	2.1957	2.1958
		FGLMM	0.1006	0.0624	0.0361
100	3	FGLM	0.3468	0.3600	0.3730
		GLMM	2.1921	2.1908	2.1892
		FGLMM	0.0220	0.0155	0.0088
	5	FGLM	0.3482	0.3662	0.3698
		GLMM	2.1912	2.1908	2.1898
		FGLMM	0.0312	0.0195	0.0116
	7	FGLM	0.3483	0.3629	0.3743
		GLMM	2.1928	2.1917	2.1918
		FGLMM	0.0394	0.0255	0.0151
200	3	FGLM	0.3419	0.3546	0.3636
		GLMM	2.1888	2.1870	2.1881
		FGLMM	0.0110	0.0072	0.0040
	5	FGLM	0.3314	0.3440	0.3610
		GLMM	2.1882	2.1877	2.1893
		FGLMM	0.0149	0.0094	0.0057
	7	FGLM	0.3461	0.3553	0.3728
		GLMM	2.1889	2.1875	2.1878
		FGLMM	0.0184	0.0122	0.0074

Chapter 5

Real Data Analysis

5.1 Data Description

Electroencephalogram (EEG) signal data (Wang et al., 2013) were collected from 10 college students as they watched Massive Open Online Courses (MOOC). Videos came in two categories – those assumed to be not confusing and those that are expected to be confusing for college students. Videos assumed to be not confusing included topics from basic algebra or geometry; whereas the ones that are expected to be confusing talked about Quantum Mechanics or Stem Cell Research. Each video had a 2-minute duration and they were taken in the middle of the discussion, to add confusion. Ten videos were prepared for each category. Students wore an apparatus called MindSet, which measured frontal lobe activity. After every session, the student rated his/her confusion level based on a scale of 1 to 7, with one being least confusing and seven being most confusing. These ratings are normalized and tagged as either confusing or not confusing and recorded as self-labelled confusion. This is on top of the study's predefined label of confusion.

The EEG dataset contains 12811 observations taken from 10 students. It also has

Table 5.1: EEG Dataset Variables

SubjectID	Student ID
VideoID	Video ID
Attention	Proprietary measure of mental focus
Meditation	Proprietary measure of calmness
Raw	Raw EEG signal
Delta	1-3 Hz of power spectrum
Theta	4-7 Hz of power spectrum
Alpha 1	Lower 8-11 Hz of power spectrum
Alpha 2	Higher 8-11 Hz of power spectrum
Beta 1	Lower 12-29 Hz of power spectrum
Beta 2	Higher 12-29 Hz of power spectrum
Gamma 1	Lower 30-100 Hz of power spectrum
Gamma 2	Higher 30-100 Hz of power spectrum
Predefined Label	Predefined label of confusion
User-defined Label	User-defined label of confusion

15 variables.

5.2 Analysis Results

Upon inspection of the dataset, it was found that there were varying number of observations per subject, per video. There were 10 subjects and 10 videos. A snapshot of the responses of and observations on subject 0 are summarized in Table 5.2. It shows there that subject 0 was not confused with videos 0, 3, 4, 7, and 9. There were 144 observations on video 0 for subject 0, which means that 144 measurements were recorded from him by the MindSet apparatus (i.e. 144 observations of each: Attention, Meditation, Raw, Delta, Theta, Alpha 1, Alpha 2, Beta 1, Beta 2, Gamma 1, and Gamma 2). The predefined label is assigned 1 for confusing and 0 for not confusing. These values were predetermined, as the study was designed. The user-defined label is the normalized (or categorized) value related to how the subject rated the video on how confusing it is.

Table 5.2: Subject 0, Count of Observations

VideoID	Predefined Label	User-defined Label	Number of Observations
0	0	0	144
1	0	1	140
2	0	1	142
3	0	0	122
4	0	0	116
5	1	1	123
6	1	1	116
7	1	0	112
8	1	1	124
9	1	0	122

The succeeding step in the analysis requires that the observations be made into functional data objects. The desired output here are curves, which are represented through the use of an appropriate basis expansion procedure. The ideal choice here was a B-spline basis expansion because the functions are not perceived to be periodic. The Beta 1 signal was the selected variable because Beta rhythms were associated with focused attention (Kropotov, 2009). Functions like coordination among multiple representations in the cortex; inhibition of movement and motor planning; preservation of the status quo; and signaling of decision making.

After having decided which basis expansion procedure would be used, the data smoothing process presented a minor challenge in the creation of the functional data objects since it was illustrated previously that there are varying number of observations per subject, per video. This meant that the respective argument values to the observations will naturally vary. The situation was remedied by choosing a common domain of $[0, 1]$ for all curves, but with different lengths based on the number of observations per subject, per video. For reference, Figure 5.1 is a plot of the functional data object created for subject 0 at video 0.

Thereafter, Functional Principal Component Analysis (FPCA) was performed

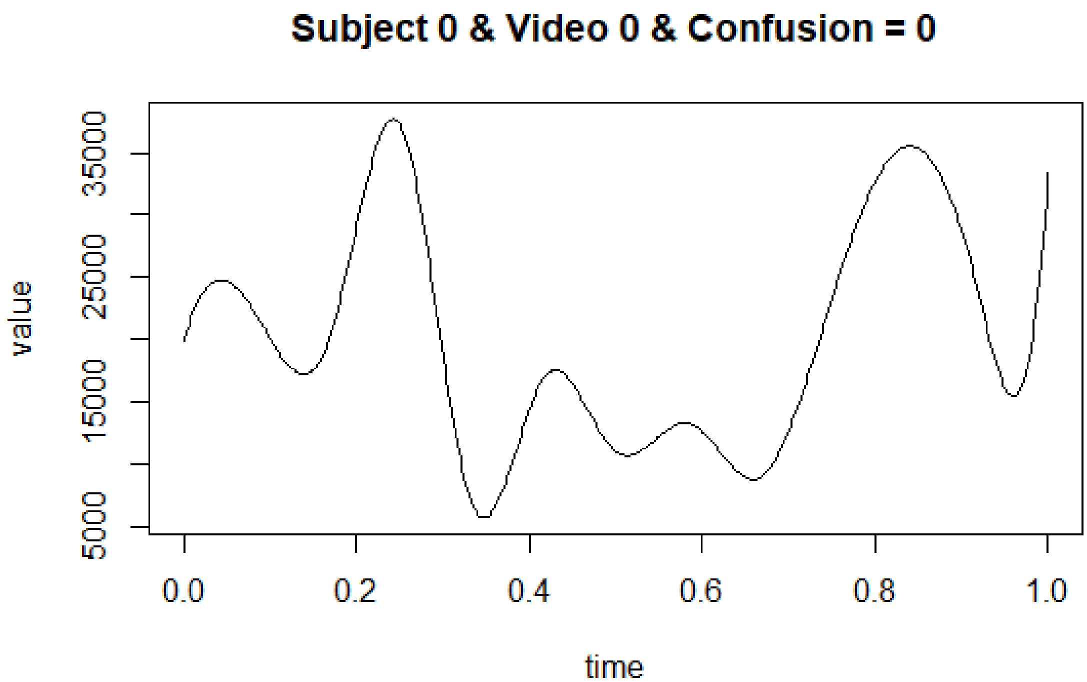


Figure 5.1: Subject 0 was not confused with Video 0. Since the total time of a video is 2 minutes or 120 seconds, the time domain of 0.2 could be interpreted as the 24^{th} second of the video and 0.4 is the 48^{th}. With the value of the function being at a high level in that interval, it means that the Beta 1 signals are also high, indicating a high focused attention from the subject.

Table 5.3: The estimated fixed intercept and random intercept standard deviation values are higher for the Bayesian approach.

Parameter	Frequentist	Bayesian
Fixed Intercept	0.0487	0.0648
Random Intercept SD	0.3790	0.6590

on the created functional data objects. The chosen number of basis functions in the creation of the functional data objects was 15. It follows that there were 15 functions that underwent the FPCA. The resulting total variation for the first four eigencomponents was 80% and it was 90% for the first seven eigencomponents. The first four eigencomponents were selected, with centered coefficients.

A portion of the reconstructed data for the GLMM fitting is shown in Table 5.4. Both the Frequentist and Bayesian approaches were used to fit a random intercept model to the data. It is represented as

$$g[\mathrm{E}(Y_{ij})] = \beta_0 + u_{0j} + b_{11}X_1 + b_{12}X_2 + b_{13}X_3 + b_{14}X_4.$$

The response variable here is whether the subject was confused or not, so the link function $g(\cdot)$ is logit. Both Frequentist and Bayesian approaches to estimation were explored in the GLMM fitting. The important estimates that can be quickly compared are those for the fixed intercept and random intercept standard deviation, which are summarized in Table 5.3

Finally, these estimates were used to create functional coefficients data, which were centered and plotted. With the goal of fitting an FGLMM, the comparison of the estimates for the GLMM coefficients were best illustrated by plotting the functional coefficients data, as seen in Figure 5.3.

We show the plots of selected $X_{ij}(t)\beta(t)$ for the i-th subject and j-th video in Figure 5.4 to 5.6, and give the value of $\int X_{ij}(t)\beta(t)dt$.

Table 5.4: The coefficients calculated for the first four principal components are merged with the Subject ID, Video ID, and Confused indicator (based on the user-defined label).

Subject ID	VideoID	X_1	X_2	X_3	X_4	Confused
0	0	-846.25929	1008.8372	6320.5328	4238.446	0
0	1	-946.48103	217.9511	-1187.2162	-1591.009	1
0	2	-84.30612	-1886.6392	5262.4712	8388.714	1
0	3	2120.19043	-703.5000	-178.2416	4223.297	0
0	4	1838.20018	1787.8160	1112.8336	2482.846	0
0	5	-4041.64312	-6126.0397	-2940.4154	-4274.564	1

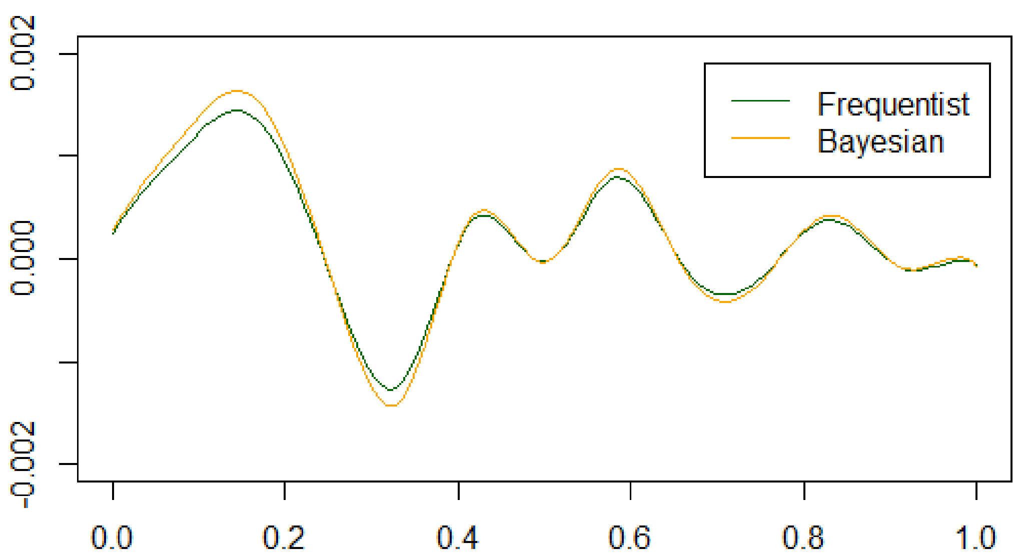


Figure 5.2: The functional coefficients plots for both approaches appear to be similar. The curves being alike indicates that the estimates from either approach are close with each other.

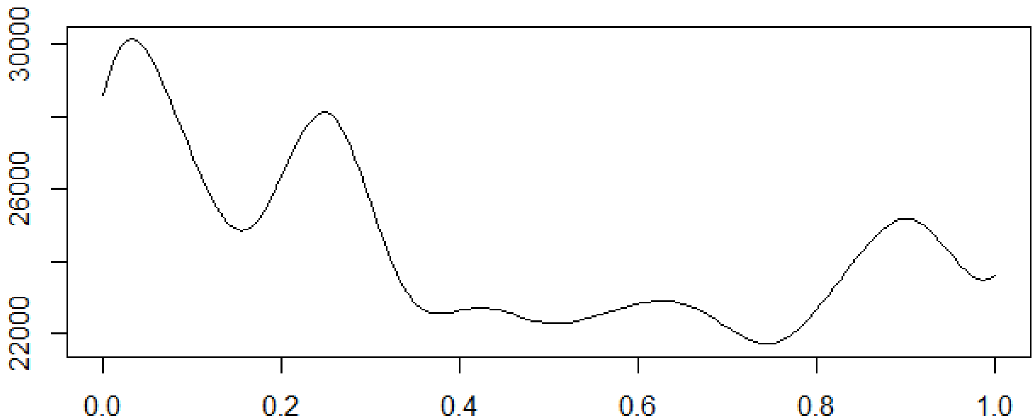


Figure 5.3: The **Mean Plot** suggests that on average, subjects tend to be more focused at the onset of watching the video, which happens within the first 30 seconds.

The accuracy of the predicted probability of getting confused or not per subject across the ten videos they watched has been summarized in Table 5.5. The predicted probability of confusion is assigned 1 when its estimated value is at least 0.5 and 0 otherwise. The calculated accuracy for each subject here is the combined number of assigned 1's and 0's that correctly matched the observed 1's and 0's in the User-defined label divided by 10 videos.

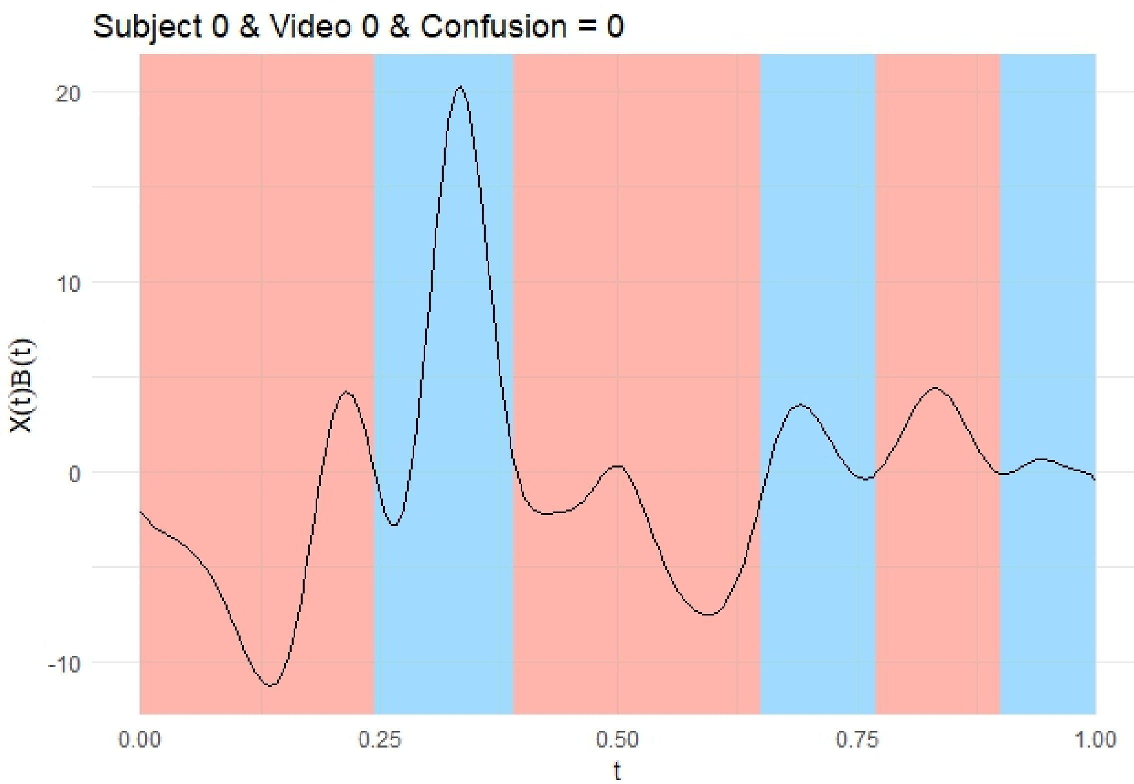


Figure 5.4: Subject 0 was not confused upon watching video 0. The calculated value of $\int X_{00}(t)\beta(t)dt$ is -36.55859. Recall that the inverse logit takes $\frac{\exp(g_{ij})}{1+\exp(g_{ij})}$; as a result, if we put -36.55859 in there, the value would be essentially zero, which means that the prediction is that subject 0 would not be confused when watching video 0.

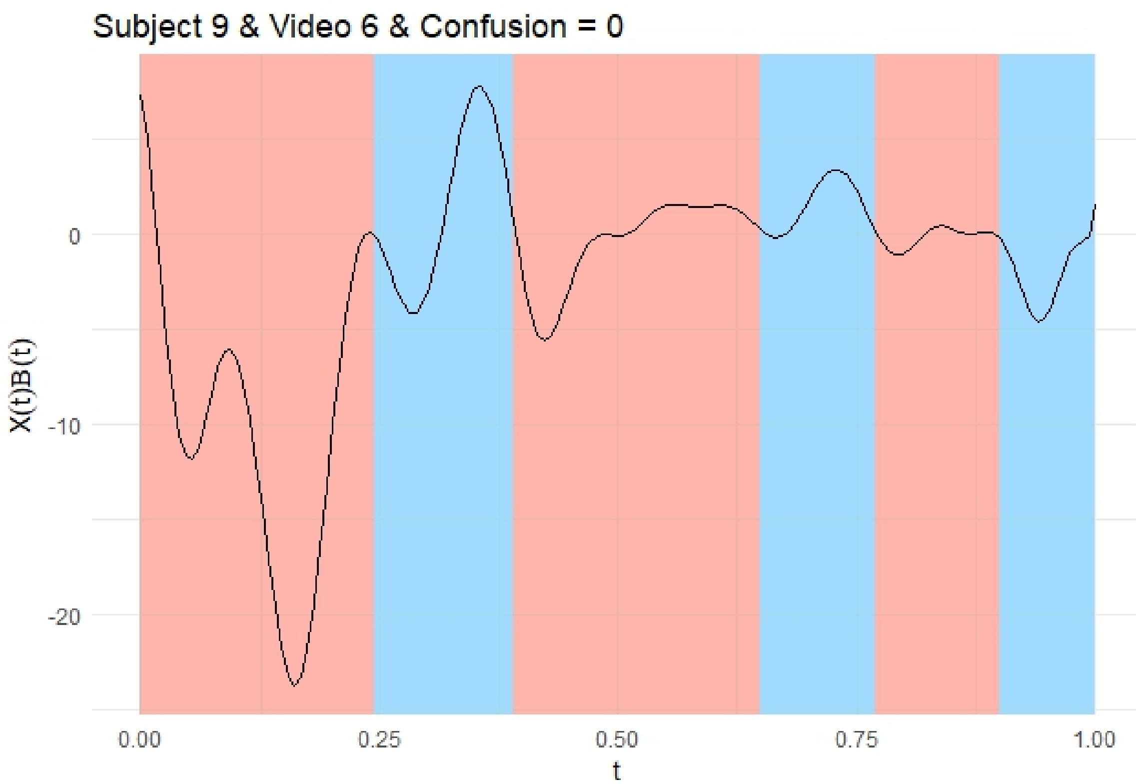


Figure 5.5: Subject 9 was not confused upon watching video 6. The calculated value of $\int X_{96}(t)\beta(t)dt$ is -348.1599. Recall that the inverse logit takes $\frac{\exp(g_{ij})}{1+\exp(g_{ij})}$; as a result, if we put -348.1599 in there, the value would be essentially zero, which means that the prediction is that subject 9 would not be confused when watching video 6.

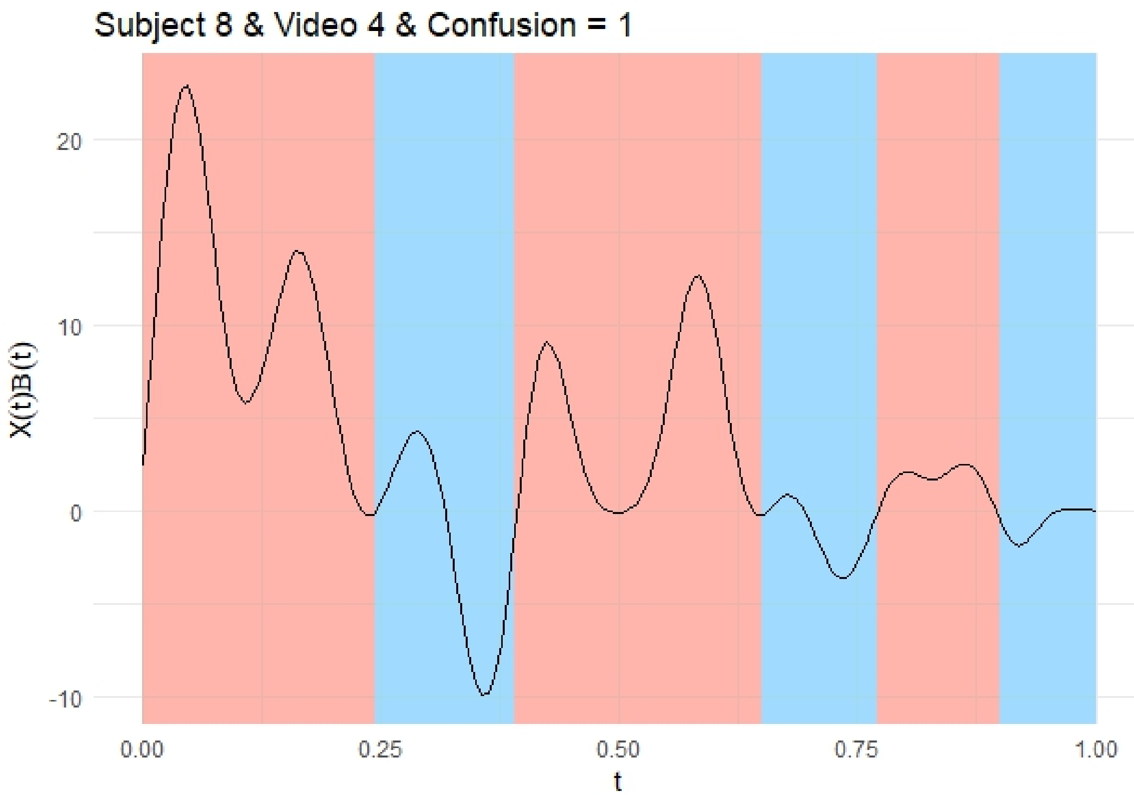


Figure 5.6: Subject 8 was confused upon watching video 4. The calculated value of $\int X_{84}(t)\beta(t)dt$ is 527.4457. Recall that the inverse logit takes $\frac{\exp(g_{ij})}{1+\exp(g_{ij})}$; as a result, if we put 527.4457 in there, the value would be almost equal to 1, which means that the prediction is that subject 8 would be confused when watching video 4.

Table 5.5: The overall accuracy of the predicted probability of confusion is 0.63. It was only for subject 3 that the proportion of the prediction of being confused or not is below half. Our random intercept model did best for subjects 5 and 6.

Subject ID	Accuracy
0	0.60
1	0.70
2	0.60
3	0.40
4	0.60
5	0.80
6	0.80
7	0.50
8	0.60
9	0.70
Overall	0.63

Chapter 6

Conclusion and Future Work

6.1 Conclusion

I. We proposed a framework for Functional Generalized Linear Mixed Models (FGLMM) and gave the Random Intercept Model as an example.

II. The Frequentist approach performed better in parameter estimation for the count response data, while the Bayesian approach performed better in parameter estimation for the binary response data.

III. The predictive performances of FGLMM were better than that of FGLM and GLMM for both the binary and count response data.

IV. We fitted an FGLMM to real data (EEG) to show where it can be applied and used the functional data plots to interpret the findings.

6.2 Future Work

I. We plan to simulate and estimate Random Functional Slope model data.

II. We plan to try Random Functional Slope model to EEG data, where each Video ID has a random functional slope.

Bibliography

A. Agresti. *Foundations of linear and generalized linear models*. John Wiley & Sons, 2015.

A. Antoniadis and T. Sapatinas. Estimation and inference in functional mixed-effects models. *Computational statistics & data analysis*, 51(10):4793–4813, 2007a. ISSN 0167-9473.

A. Antoniadis and T. Sapatinas. Estimation and inference in functional mixed-effects models. *Computational statistics & data analysis*, 51(10):4793–4813, 2007b.

S. Bianconcini. Asymptotic properties of adaptive maximum likelihood estimators in latent variable models. 2014.

N. E. Breslow and D. G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25, 1993. ISSN 0162-1459.

H. Chen and Y. Wang. A penalized spline approach to functional mixed effects model analysis. *Biometrics*, 67(3):861–870, 2011.

K. Chen and H. Müller. Conditional quantile analysis when covariates are functions, with application to growth data. *Journal of the Royal Statistical Society, Series B*, 74:67–89, 2012.

T. J. Cleophas and A. H. Zwinderman. Functional data analysis i: Gene expression levels for predicting drug efficacy outcomes. In *Regression Analysis in Medical Research*, pages 393–406. Springer International Publishing, Cham, 2018. ISBN 331971936X.

L. Fahrmeir. *Regression Models, Methods and Applications / by Ludwig Fahrmeir, Thomas Kneib, Stefan Lang, Brian Marx.* Springer Berlin Heidelberg : Imprint: Springer, 1st ed. 2013.. edition, 2013. ISBN 3-642-34333-3.

G. H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.

W. Guo. Functional mixed effects models. *Biometrics*, 58(1):121–128, 2002.

P. D. Hoff. *A first course in Bayesian statistical methods*, volume 580. Springer, 2009.

L. Horváth and P. Kokoszka. *Inference for Functional Data with Applications*. Springer, New York, 2012.

G. M. James. Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):411–432, 2002a.

G. M. James. Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):411–432, 2002b.

S. Jin and B. Andersson. A note on the accuracy of adaptive gauss–hermite quadrature. *Biometrika*, 107(3):737–744, 2020.

H. B. Kang, Y. J. Jung, and J. Park. Fast bayesian functional regression for non-gaussian spatial data. *Bayesian Analysis*, 1(1):1–32, 2023.

P. Kokoszka and M. Reimherr. *Introduction to functional data analysis*. Chapman and Hall/CRC, 2017.

J. D. Kropotov. Chapter 3 - beta rhythms. In J. D. Kropotov, editor, *Quantitative EEG, Event-Related Potentials and Neurotherapy*, pages 59–76. Academic Press, San Diego, 2009. ISBN 978-0-12-374512-5. doi: https://doi.org/10.1016/B978-0-12-374512-5.00003-7. URL https://www.sciencedirect.com/science/article/pii/B9780123745125000037.

W. Ma, L. Xiao, B. Liu, and M. A. Lindquist. A functional mixed model for scalar on function regression with application to a functional mri study. *Biostatistics*, 22(3):439–454, 2021.

P. McCullagh and J. A. Nelder. *Generalized linear models*. Routledge, 2019.

M. W. McLean, G. Hooker, A.-M. Staicu, F. Scheipl, and D. Ruppert. Functional generalized additive models. *Journal of Computational and Graphical Statistics*, 23 (1):249–269, 2014.

J. S. Morris and R. J. Carroll. Wavelet-based functional mixed models. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 68(2):179–199, 2006a. ISSN 1369-7412.

J. S. Morris and R. J. Carroll. Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2):179–199, 2006b.

H.-G. Müller and U. Stadtmüller. Generalized functional linear models. *The Annals of Statistics*, 33(2):774–805, 2005.

H.-G. Müller and U. Stadtmüller. Generalized functional linear models. *the Annals of Statistics*, 33(2):774–805, 2005.

J. Ramsay. *Functional Data Analysis with R and MATLAB by James Ramsay, Giles Hooker, Spencer Graves.* Use R! 1st ed. 2009.. edition, 2009. ISBN 1-282-29264-1.

J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer, 2005.

J. O. J. O. Ramsay. *Functional data analysis J.O. Ramsay, B.W. Silverman.* Springer series in statistics. Springer, New York, 2nd ed. 2005.. edition, 1997. ISBN 9786610612420.

A. C. Rencher. *Methods of multivariate analysis Alvin C. Rencher, William F. Christensen.* Wiley series in probability and statistics. Wiley, Hoboken, N.J., 3rd ed.. edition, 2012. ISBN 1-118-39167-5.

F. Scheipl, J. Gertheiss, and S. Greven. Generalized functional additive mixed models. *Electronic Journal of Statistics*, 10(1):1455–1492, 2016.

T. S. Tian. Functional data analysis in brain imaging studies. *Frontiers in psychology*, 1:35, 2010.

H. Wang, Y. Li, X. Hu, Y. Yang, Z. Meng, and K.-m. Chang. Using eeg to improve massive open online courses feedback interaction. In *AIED Workshops*, 2013.

J.-L. Wang, J.-M. Chiou, and H.-G. Müller. Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295, 2016.

J. T. Zhang and J. Chen. Statistical inferences for functional data. *The Annals of Statistics*, 35(3):1052–1079, 2007.

H. Zhu, P. J. Brown, and J. S. Morris. Robust, adaptive functional regression in functional mixed model framework. *Journal of the American Statistical Association*, 106(495):1167–1179, 2011.