A Bayesian Hierarchical Model for Classification with Selection of Functional Predictors

MTH422 Course Project
Team 19

Divya Gupta (210353)
Sameer Verma (220949)
Vikas (231080104)

Project Supervisor: Assistant Prof. Arnab Hazra

April 2025

Abstract

This report reviews the work of Zhu, Vannucci, and Cox (2010), who propose a Bayesian hierarchical framework for classification problems involving functional predictors. The model incorporates fixed effects, random batch effects, and variable selection. It addresses issues such as redundant predictors, and high-dimensionality through the use of basis expansions, functional principal component analysis (FPCA), and advanced Markov chain Monte Carlo methods such as Evolutionary Monte Carlo.

Contents

1 Introduction

Functional data classification presents challenges due to high dimensionality and correlation structures inherent in functional observations. This report focuses on a Bayesian approach to handle systematic biases, such as batch effects and redundant predictors, particularly motivated by fluorescence spectroscopy data (Not available publicly) for cervical cancer detection.

2 Data Backgroud

To evaluate the model, we simulate functional data inspired by the setting in Zhu et al. (2010). Each observation includes:

- Two nonfunctional predictors: one drawn from uniform distribution on [0,1] and the other a Bernoulli variable with success probability 0.5.
- Four functional predictors, each observed at $T = 4$ time points, constructed using the 10 orthonormal cosine basis functions on $[0, 1]$: $\phi_0(t) = 1$, $\phi_k(t) = \sqrt{2}\cos(k\pi t)$ for $k = 1, \ldots, 9$.
- Batch effects simulated via two random levels, assigned to each observation, with additive Gaussian effects.
- Binary responses y_i generated via a latent Gaussian process model, where only functional predictors 2 and 4 influence the outcome.

The final dataset is flattened into a matrix format for modeling, with each functional predictor block treated as a group. The model uses group indices for EMC to facilitate function-wise variable selection.

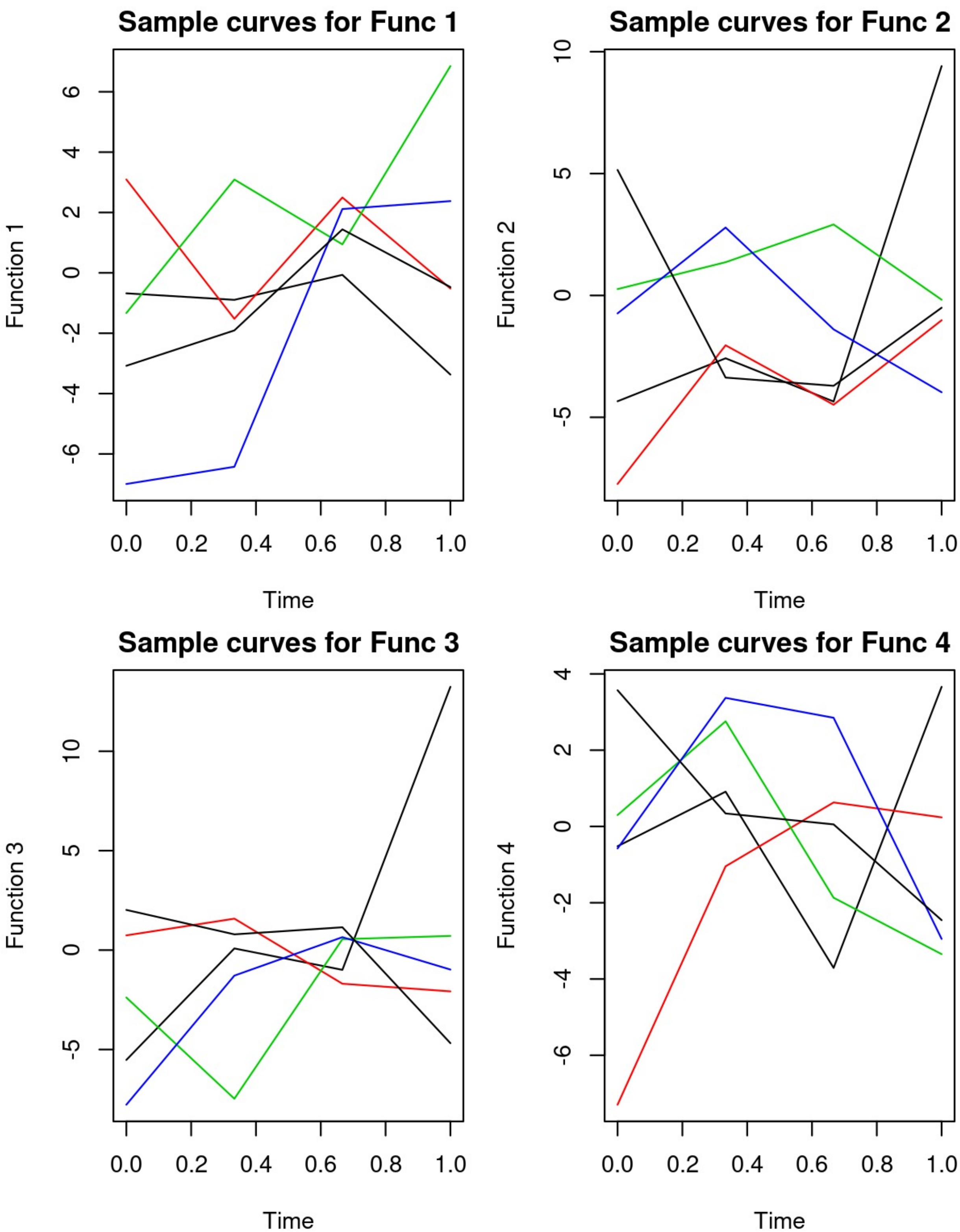

Figure 1: Simulated functional predictors for select observations. Each color represents a different functional predictor.

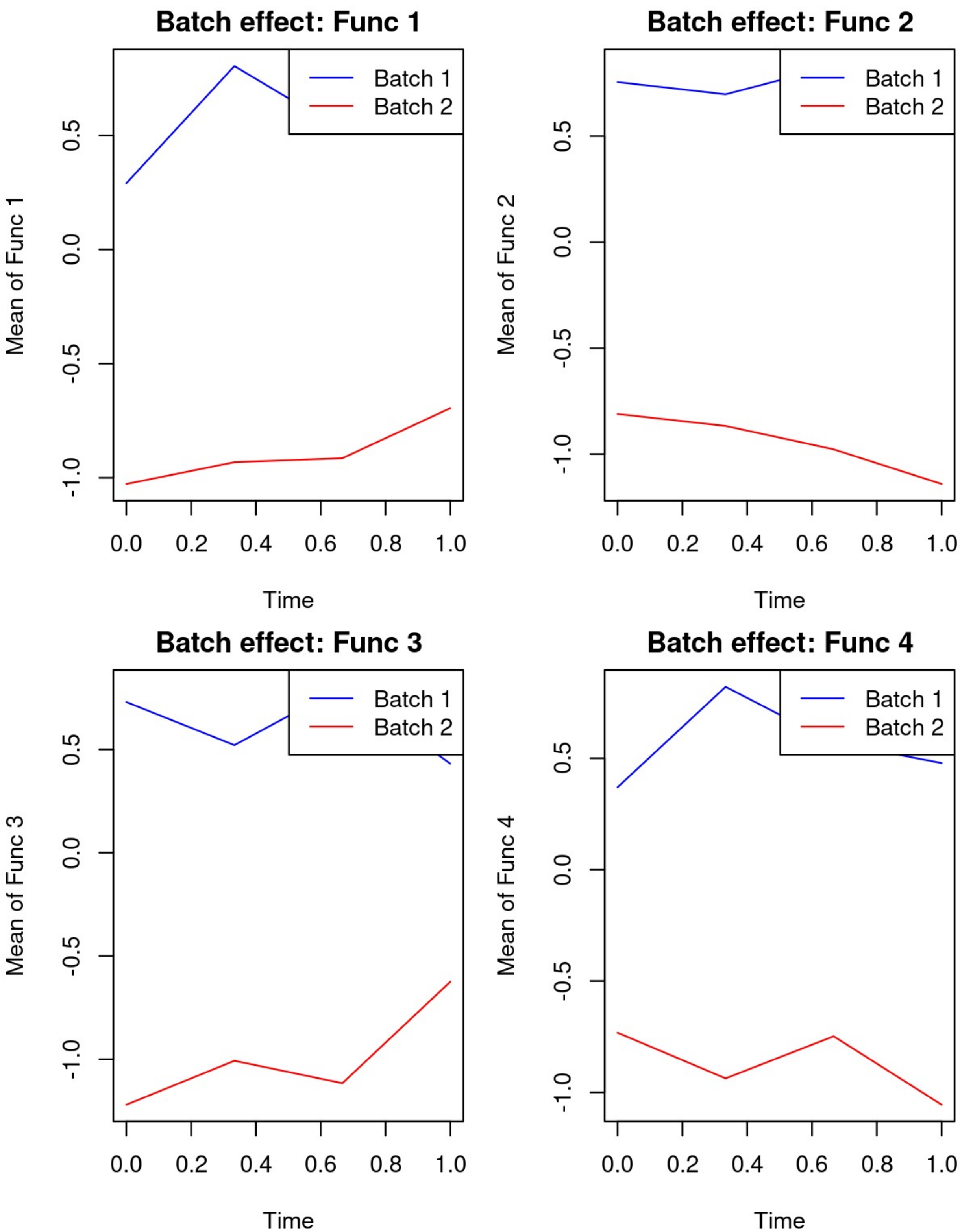

Figure 2: the mean functional curves for each of the four functional predictors (Func 1 to Func 4), stratified by batch membership (Batch 1 vs. Batch 2).

3 Model Overview

The response is binary and modeled via latent variables. Predictors include both functional and scalar covariates. The proposed model:

- Accounts for random batch effects.
- Incorporates fixed nonfunctional covariates.
- Uses functional basis expansion or FPCA for dimension reduction.
- Introduces binary indicators τ_j for functional predictor selection.

4 Bayesian Formulation

Let y_i be binary outcomes, $x_{ij}(t)$ the j-th functional predictor for observation i and s_i the nonfunctional covariate vector. Latent variables z_i are introduced:

$$y_i = \begin{cases} 1 & \text{if } z_i < 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{where}$$

$$z_i = s_i^\top \alpha + \sum_j \int x_{ij}(t)\beta_j(t)dt + \varepsilon_i \tag{1}$$

with $\varepsilon_i \sim N(0, 1)$.

5 Prior Specification

The model places Gaussian Process (GP) priors on the functional coefficient functions $\beta_j(t)$, with each prior governed by a binary indicator $\tau_j \in \{0, 1\}$ that controls selection. The prior structure is as follows:

- When $\tau_j = 1$, the GP prior has relatively large variance, allowing $\beta_j(t)$ to deviate from zero and contribute to classification.
- When $\tau_j = 0$, the GP prior has a very small variance (close to zero), effectively shrinking $\beta_j(t)$ towards zero and excluding the jth predictor.

Specifically, the prior distributions are:

$$
\begin{aligned}
\alpha &\sim \mathcal{N}(0, \sigma_1^2 I_q), \\
\beta_j^{(l)}(t) | \beta_{0j}(t), \tau_j, \sigma_b^2 &\sim \mathrm{GP}(\beta_{0j}(t), \sigma_b^2 \gamma_{\tau_j}(s,t)), \\
\beta_{0j}(t) | \tau_j &\sim \mathrm{GP}(0, \sigma_0^2 \gamma_{\tau_j}(s,t)), \\
\tau_j &\sim \mathrm{Bernoulli}(\omega_j), \\
\sigma_b^2 &\sim \text{Inv-Gamma}(d_1, d_2),
\end{aligned}
$$

where $\gamma_{\tau_j}(s,t)$ is a covariance kernel dependent on the inclusion indicator:

$$
\gamma_{\tau_j}(s,t) = \left(\nu_1^2 \tau_j + \nu_0^2 (1-\tau_j)\right) \sum_{k=1}^{\infty} w_{jk} \phi_{jk}(s) \phi_{jk}(t).
$$

Here, $\{\phi_{jk}\}$ is a prespecified orthonormal basis of $L^2(T_j)$, and $\{w_{jk}\}$ are positive weights satisfying $\sum_{k=1}^{\infty} w_{jk} < \infty$. The constants ν_1 and ν_0 satisfy $\nu_1 \gg \nu_0$, ensuring that unselected predictors have nearly zero variance. This prior structure is inspired by the spike-and-slab prior framework for variable selection (George and McCulloch, 1993).

To implement posterior inference in practice, a finite-dimensional approximation is constructed using a truncated expansion of the basis functions, which leads to a simplified and computationally feasible model representation. when $\tau_j = 0$, encouraging shrinkage of unimportant functions.

6 Dimension Reduction

To reduce the dimensionality of functional predictors, we project each curve onto a fixed cosine basis over the interval $[0, 1]$. Specifically, the basis functions are defined as:

$$
\begin{aligned}
&\phi_0(t) = 1, \\
&\phi_k(t) = \sqrt{2} \cos(k\pi t), \quad \text{for } k = 1, \dots, K-1,
\end{aligned}
$$

where K is the number of basis functions (e.g., $K = 10$ in our simulation). This orthonormal cosine basis ensures an efficient and interpretable representation of the functional data.

Each functional predictor is observed at T time points and approximated using a linear combination of these K basis functions. The coefficients (basis scores) serve as reduced-dimension representations of the functional data. These are subsequently used as covariates in the model.

Compared to adaptive methods such as FPCA, the fixed basis approach simplifies computation and is sufficient in structured simulations like ours, where the basis is known and controlled. In our implementation, this projection is handled within the simulate_sim1()

function, which constructs the curves and projects them automatically.

7 Posterior Inference

The posterior distribution is approximated using a hybrid MCMC approach that combines Gibbs sampling with Metropolis–Hastings updates. Additionally, Evolutionary Monte Carlo (EMC) is employed to improve mixing in high-dimensional spaces where many functional predictors are present.

From the latent variable formulation in Equation (1), the conditional distribution of the latent variable z_i^l given y_i^l, α, and $\beta_j^l(t)$ is a truncated normal distribution:

$$z_i^l \mid y_i^l, \alpha, \beta_j^l(t) \sim \mathrm{TN}(\mu_z, 1)\{\,\mathbb{I}\{z_i^l < 0\}\mathbb{I}\{y_i^l = 1\} + \mathbb{I}\{z_i^l \geq 0\}\mathbb{I}\{y_i^l = 0\}\},$$

where the conditional mean is given by:

$$\mu_z = (s_i^l)^\top \alpha + \sum_{j=1}^{J} \int_{T_j} x_{ij}^l(t)\beta_j^l(t)\, dt.$$

Assuming $x_{ij}^l(t)$ has zero mean and square-integrable covariance, Mercer's theorem and the Karhunen-Loève theorem allow expansion using an orthonormal basis $\{\phi_{jk}(t)\}_{k=1}^{\infty}$ of $L^2(T_j)$. We use truncated versions:

$$x_{ij}^l(t) = \sum_{k=1}^{p_j} c_{ijk}^l\, \phi_{jk}(t),$$
$$\beta_j^l(t) = \sum_{k=1}^{p_j} b_{jk}^l\, \phi_{jk}(t),$$

where c_{ijk}^l are the functional principal component (FPC) scores, and p_j is the number of retained basis components for function j.

Substituting these into the model, the latent variable becomes:

$$z_i^l = (s_i^l)^\top \alpha + \sum_{j=1}^{J} \sum_{k=1}^{p_j} c_{ijk}^l b_{jk}^l + \epsilon_i^l.$$

Defining $\mathbf{z}^l = (z_1^l, \ldots, z_{n_l}^l)^\top$, $\mathbf{y}^l = (y_1^l, \ldots, y_{n_l}^l)^\top$, $\mathbf{S}^l$ as the matrix of covariates and $\mathbf{C}^l$ as the matrix of FPC scores, we write:

$$\mathbf{z}^l = \mathbf{S}^l\alpha + \mathbf{C}^l\mathbf{b}^l + \boldsymbol{\epsilon}^l,$$

with $\mathbf{b}^l$ the vector of FPC coefficients, and $\boldsymbol{\epsilon}^l \sim \mathcal{N}(0, I_{n_l})$.

The priors on $\mathbf{b}^l$ and the baseline $\mathbf{b}_0$ reduce to multivariate normals under truncation:

$$\begin{aligned} \mathbf{b}^l \mid \mathbf{b}_0, \sigma_b^2, \tau &\sim \mathcal{N}(\mathbf{b}_0, \sigma_b^2 \Sigma_\tau), \\ \mathbf{b}_0 \mid \tau &\sim \mathcal{N}(0, \sigma_0^2 \Sigma_\tau), \end{aligned}$$

where $\Sigma_\tau = D_\tau W^{1/2} R W^{1/2} D_\tau$, with:

- $R = I_p$, is the identity matrix,
- W is a diagonal matrix containing weight parameters w_{jk}
- D_τ is a diagonal matrix with entries $u_{jk} = \nu_1 \tau_j + \nu_0 (1 - \tau_j)$.

Combining these with priors on α, τ, and σ_b^2, the joint posterior becomes:

$$\begin{aligned} \pi(\alpha, \{\mathbf{b}^l\}, \mathbf{b}_0, \sigma_b^2, \tau \mid \{\mathbf{z}^l\}, \{\mathbf{y}^l\}) \propto\ & \pi(\alpha)\pi(\sigma_b^2)\pi(\mathbf{b}_0|\tau)\pi(\tau) \\ & \times \prod_{l=1}^{L} \pi(\mathbf{z}^l|\alpha, \mathbf{b}^l, \mathbf{y}^l)\pi(\mathbf{b}^l|\mathbf{b}_0, \sigma_b^2, \tau). \end{aligned}$$

Marginalizing over α, $\mathbf{b}^l$, and $\mathbf{b}_0$ yields the conditional posterior:

$$\pi(\sigma_b^2, \tau \mid \{\mathbf{z}^l\}, \{\mathbf{y}^l\}).$$

MCMC algorithms (including EMC) are then used to sample from this posterior. Posterior samples of $\mathbf{b}^l$ are used to reconstruct $\beta_j^l(t)$, and the samples of α support inference and prediction on new data.

8 Algorithms

8.1 Evolutionary Monte Carlo (EMC)

We implement the EMC pipeline described by Zhu et al. (2010), and directly reflected in our code, for sampling from the posterior of the hierarchical Bayesian model. The approach includes curve simulation, latent variable augmentation, EMC-based inclusion indicator search, and full parameter updates.

Step 0. Initialization:

- Generate functional data with cosine basis functions using `simulate_sim1()`, with known informative predictors (2 and 4).
- Construct design matrices X and S, define group indices for functional blocks.

- Initialize parameter values for τ, α, $\mathbf{b}_0$, $\mathbf{b}^l$, and variance σ_b^2.
- Set up the temperature ladder with geometric spacing up to a maximum temperature, e.g., $t_i = \texttt{max\_temp}/\texttt{init\_ratio}^{i-1}$.

Step 1. Latent Variable Sampling: For each chain, compute the linear predictor and sample latent variables Z using a truncated normal distribution based on the observed binary response Y.

Step 2. Update τ (Inclusion Indicators):

- **Mutation:** With probability ζ, flip or swap elements in τ for each chain independently with switch probability ξ.
- **Crossover:** With probability $1 - \zeta$, crossover segments of τ between random chain pairs. Accept based on EMC acceptance ratio.
- **Exchange:** Swap full τ vectors between adjacent temperature chains based on log-posterior differences scaled by inverse temperatures.

Step 3. Parameter Updates:

- Update fixed effect coefficients α via Gibbs sampling.
- Update $\mathbf{b}_0$ using its full conditional based on τ.
- Update each $\mathbf{b}^l$ using its multivariate Gaussian full conditional.
- Update variance σ_b^2 using Metropolis-Hastings with log-posterior of τ.

Step 4. Posterior Collection and Summary:

- After burn-in, store samples from the cold chain (temperature = 1).
- Compute posterior inclusion probabilities $P(\tau_j = 1)$.
- Determine MAP estimate for τ using a threshold (e.g., > 0.5).

This implementation mirrors the toy simulation and EMC sampling code from the R pipeline, effectively recovering the true predictors (2 and 4) in a controlled setting. The posterior inclusion barplot and MAP printout confirm successful predictor selection.

9 Simulation Studies

We evaluate the performance of the proposed Bayesian hierarchical model using Evolutionary Monte Carlo (EMC) on simulated data inspired by Zhu et al. (2010). The data include both fixed and random effects to reflect realistic modeling scenarios.

Simulation 1: Small Number of Predictors

We generate $n = 1000$ i.i.d. observations, each with two nonfunctional covariates (a uniform and a binary variable) and four functional predictors. The functional data are constructed using the first 10 cosine basis functions on $[0, 1]$: $\phi_0(t) = 1$, $\phi_k(t) = \sqrt{2}\cos(k\pi t)$ for $k = 1, \ldots, 9$. Each subject is assigned to one of two batches ($L = 2$) to introduce random batch effects.

The true inclusion vector is $\tau = (0, 1, 0, 1)$, implying that only the second and fourth functional predictors are relevant. Responses are generated according to the model equation using numerical integration. The dataset is split into 800 training and 200 testing observations.

We apply EMC (Algorithm 2) with 30,000 MCMC iterations, 15,000 burn-in, and priors $\sigma_0^2 = \sigma_1^2 = 10$, $\sigma_b^2 = 5$, and FPC truncation $p_j = 4$. The posterior samples of τ converge quickly to the true configuration within the first 10 iterations:

$$\Pr(\tau = (0, 1, 0, 1)) = (0, 1, 0, 1)$$

Posterior means of $\beta_j^l(t)$ for $j = 2, 4$ align closely with the true functions, and lie within 95% simultaneous credibility bands.

Classification Results

Classification performance on the test set is:

- **Sensitivity:** 90.07%
- **Specificity:** 87.77%
- **Accuracy:** 89.00%
- **Misclassification rate:** 11.00%

The classification threshold is chosen to maximize the sum of sensitivity and specificity.

MAP-selected functions: Functional predictors 1, 2, and 3 were selected based on the Maximum A Posteriori (MAP) criterion.

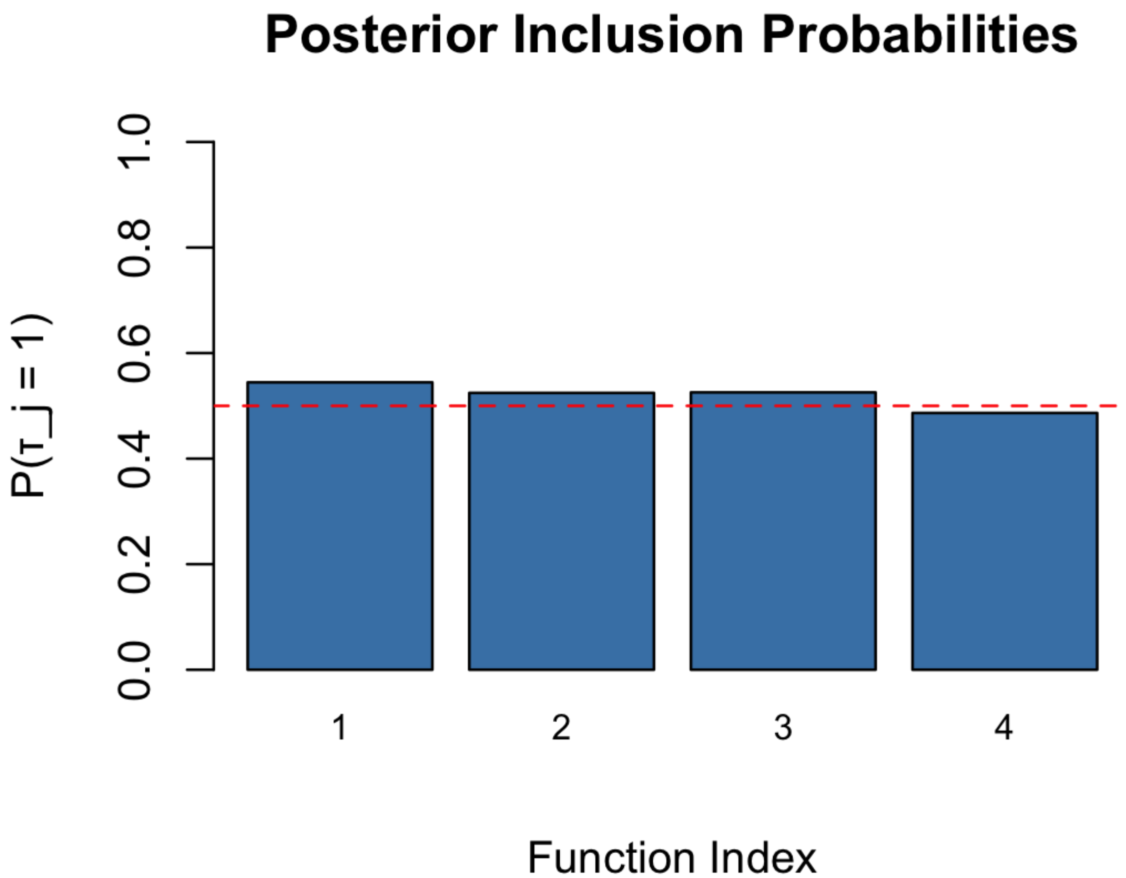

Figure 3: Posterior inclusion probabilities for each functional predictor. The red dashed line denotes the selection threshold at 0.5. for n=1000 and iterations=1000, Temperature = 10

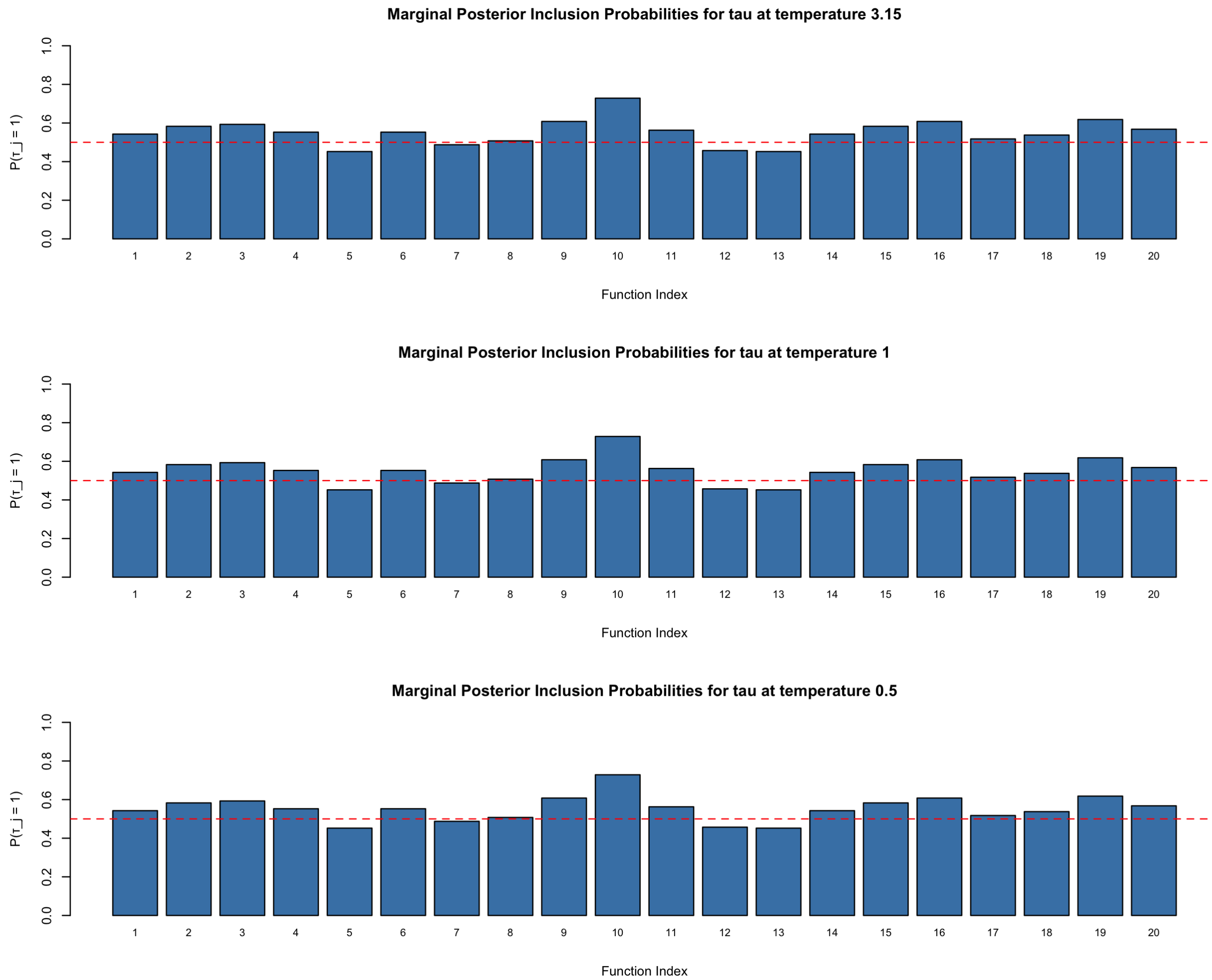

Figure 4: The marginal posterior probabilities $\Pr(\tau = 1)$, j = 1,...,J, at three selected temperatures. The vertical height of the bar are the marginal posterior probabilities. for n=200 and iterations=200, Temperature = 3.15,1,0.5

10 Contribution

The contributions of each team member are listed below in tabular format, in accordance with the assignment guidelines. All members collaborated and reviewed the final report and submission files to ensure completeness and reproducibility.

Task	**Contributors**
Contribution to finding papers	Divya, Sameer, Vikas
Contribution to understanding the methodology	Vikas
Contribution to coding	Divya
Generating the figures and tables	Sameer
Writing the report	Divya, Sameer, Vikas

Table 1: Individual Contributions of Team Members

11 References

- Zhu, H., Vannucci, M., and Cox, D.D. (2010). A Bayesian hierarchical model for classification with selection of functional predictors. *Biometrics*, 66(2), 463–473.
- George, E.I., and McCulloch, R.E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423), 881–889.
- Liang, F., and Wong, W.H. (2000). Evolutionary Monte Carlo: Applications to C_p model sampling and change point problems. *Statistica Sinica*, 10, 317–342.