# Dynamic Prediction of Generalized Functional Data

Ying Jin, Andrew Leroux

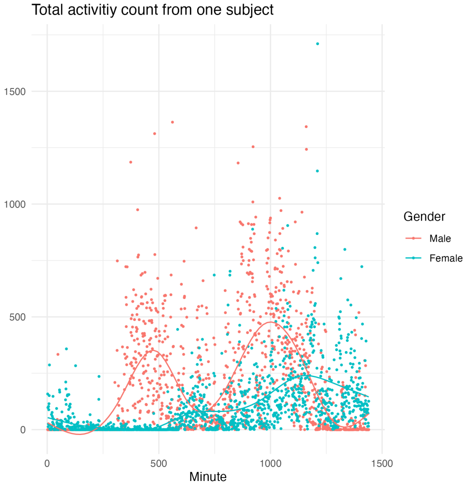
December 8th, 2023

# Introduction

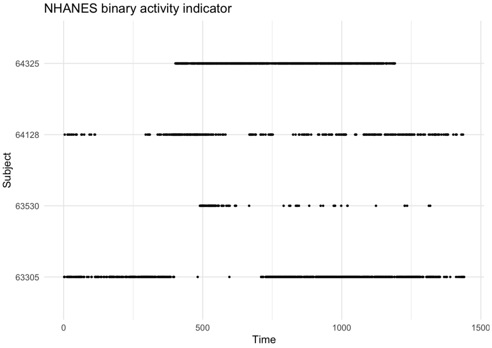
# Functional data

- Functional data
  - Unit of observations is a function for each subject
  - Arising from a smooth underlying function: $E(Y(t)) = \mu(t)$
  - Observed as repeated measures densely collected across the study domain
- Generalized functional data
  - Functional data with discrete value (e.g., binary outcome)
  - Following exponential family distribution characterized by a continuous latent function: $g(E(Y(t))) = \eta(t)$

# Examples



Total activitiy count from one subject
1500
1000
500
0
0
500
1000
1500
Minute
Gender
Male
Female

NHANES binary activity indicator
64325
64128
63530
63305
Subject
0
500
1000
1500
Time

# Dynamic prediction

- Dynamic prediction
  - To predict future outcomes based on historical data from the same subject
  - Prediction updates as extra measures are collected
- Challenges of generalized functional data
  - Density Density of NHEANES binary indicator: minute by minute. 1440 measures /day for each subject
  - Complexity
  - Estimation of out-of-sample random effect
- Goal: to develop a fast, scalable method for dynamic prediction of generalized functional outcomes

# Method

# Assumptions

For each subject i in the population

- A generalized outcome $Y_i(t)$ is generated along a variable t (for example, time), where $t \in (0, T]$
- The outcome, at any specific t, follows an exponential family distribution characterized by a (latent) continuous function $\eta_i(t)$
- The latent function $\eta_i(t)$ consists of a functional fixed effect (population-level) and a random effect (subject-level)

$$g[E(Y_i(t))] = \eta_i(t) = \beta_0(t) + b_i(t)$$

# Observed data

- In practice we would observe the discrete realization of $\{Y_i(t), t\}$ along a dense grid
- Domain of t: $(0, T]$
- Measurement index $j = 1...J$
- Value of t at jth observation: $t_j$
- Value of outcome at jth observation: $Y_i(t_j)$

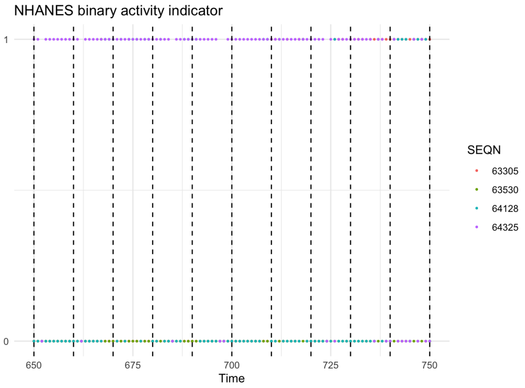
# Generalized Functional Principal Component Analysis (GFPCA)

- Functional PCA of non-Gaussian functional data

$$g(E(Y_i(t))) = f_0(t) + \sum_{k=1}^{K} \xi_{ik}\phi_k(t)$$

  - $f_0$ is the population mean function
  - $\phi_k$ are orthogonal principal component functions (PC)
  - $\xi_{ik}$ are mutually independent scores/loadings. $\xi_{ik} \sim N(0, \lambda_k)$
- Existing methods tend to be slow in implementation
- Fast implementation of FPCA exists for Gaussian outcomes (e.g., FACE)

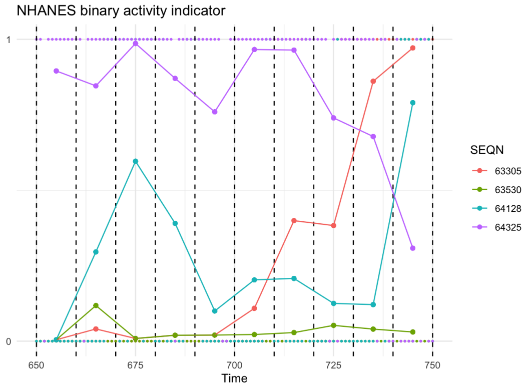
# Fast implementation of GFPCA (fGFPCA)




Bin the observed outcomes in to small, non-overlapping, equal length bins.

- Bin index: $s \in \{1, 2, ...S\}$
- Midpoint index of the sth bin: $m_s$
- Value of t at bin midpoints: $t_{m_s}$
- If bins have equal length w, then the sth bin is $(t_{m_s} - \frac{w}{2}, t_{m_s} + \frac{w}{2}]$

# Fast implementation of GFPCA (fGFPCA)

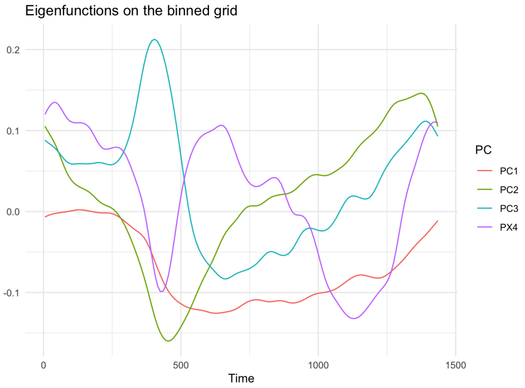



Fit a local, intercept-only generalized linear mixed model at every bin:

$$g[E(Y_i(t_j))] = \eta_i(t_{m_s}) = \beta_0(t_{m_s}) + b_i(t_{m_s})$$
$$t_j \in (t_{m_s} - \frac{w}{2}, t_{m_s} + \frac{w}{2}]$$

Estimate subject-level latent function tracks on the binned grid $\hat{\eta}_i(t_{m_s})$

# Fast implementation of GFPCA (fGFPCA)



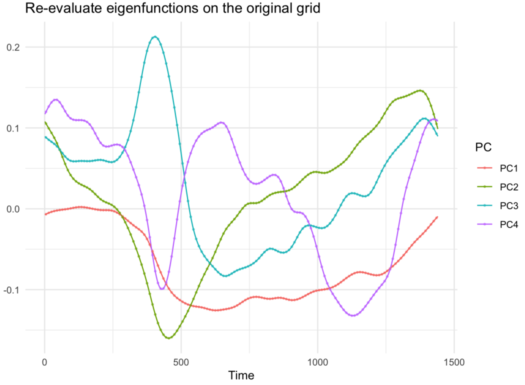

Fit FPCA on the estimated latent functions:

$$\hat{\eta}_i(t_{m_s}) = f_0(t_{m_s}) + \sum_{k=1}^{K} \xi_{ik}\phi_k(t_{m_s}) + \epsilon_i(t_{m_s})$$

to obtain a set of estimates:

- Basis functions $\hat{\mathbf{\Phi}} = \{\hat{\phi}_1(t_{m_s}), ..., \hat{\phi}_K(t_{m_s}))\}$
- Variance of scores $\hat{\lambda}_1 ... \hat{\lambda}_K$

# Fast implementation of GFPCA (fGFPCA)




- $\hat{\phi}_K(t_{m_s})$ are estiamted on the binned grid
  - Use spline basis the to re-evaluated on the original grid $t_j$
- $\hat{\phi}_K(t_{m_s})$ and $\hat{\lambda}_k$ are biased by a multiplicative factor
  - Use a GLMM model to de-biased
- This step also re-estimates the population mean function $\hat{f}_0$

# Out-of-sample dynamic prediction

- Assume we have a new observations $u$ who is partiall observed with $J_u$ measures ($J_u < J$)
- Prediction of unobserved track: $\hat{\eta_u}(t_j) = \hat{f}_0(t_j) + \sum_{k=1}^{K} \hat{\xi}_{uk}\hat{\phi}_k(t_j), J_u < j \leq J$
- Since the outcome follows an exponential family distribution $p(Y_i(t)|\eta_i(t)) = h(Y_i(t))exp\{\eta_i(t)T[Y_i(t)] - A(\eta_i(t))\}$, the Log-likelihood of this new subject:

$$
l_u(\boldsymbol{\xi}_u) = \sum_{j<J_u} log(h(Y_u(t_j))) + \eta_u(t_j)T(Y_u(t_j)) - log(A[\eta_u(t_j)])
$$

$$
\boldsymbol{\xi}_u = [\xi_{u1}, ..., \xi_{uK}]
$$

- Use Bayes method to maximum the likelihood:
  - Prior distribution: $\xi_{uk} \sim N(0, \hat{\lambda}_k)$
  - Posterior distribution: the likelihood of $P(Y_u(t_j)|\xi_u) = l_u(\boldsymbol{\xi}_u)$

# Simulation

# Simulation

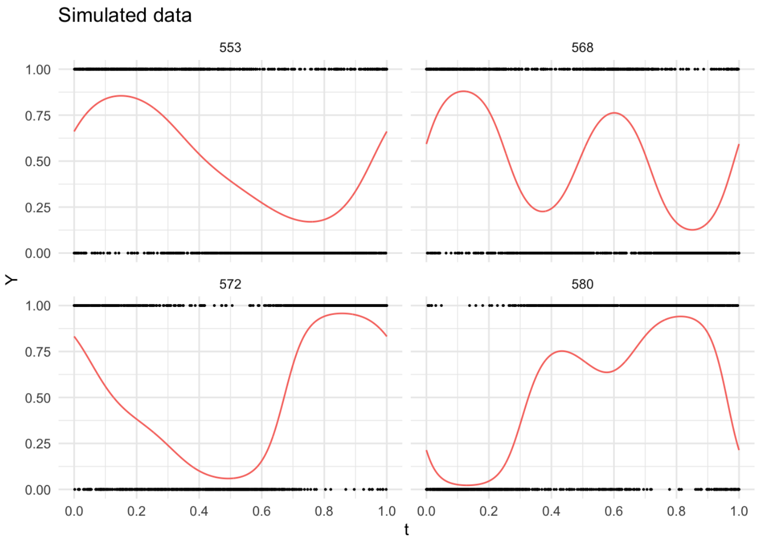
- 500 datasets, each with 500 subjects (N=500)
- Each subject has 1000 observations along $t \in (0, 1]$ (J $= 1000$)
- Binary functional outcomes:

$$Y_i(t) \sim Bernoulli(\frac{exp(\eta_i(t))}{1 + exp(\eta_i(t))})$$

$$\eta_i(t) = f_0(t) + \xi_{i1}\sqrt{2}sin(2\pi t) + \xi_{i2}\sqrt{2}cos(2\pi t) + \xi_{i3}\sqrt{2}sin(4\pi t) + \xi_{i4}\sqrt{2}cos(4\pi t)$$

  - $f_0(t) = 0$
  - $\xi_{ik} \sim N(0, 0.5^{k-1}),\ k \in \{1, 2, 3, 4\}$

# Example

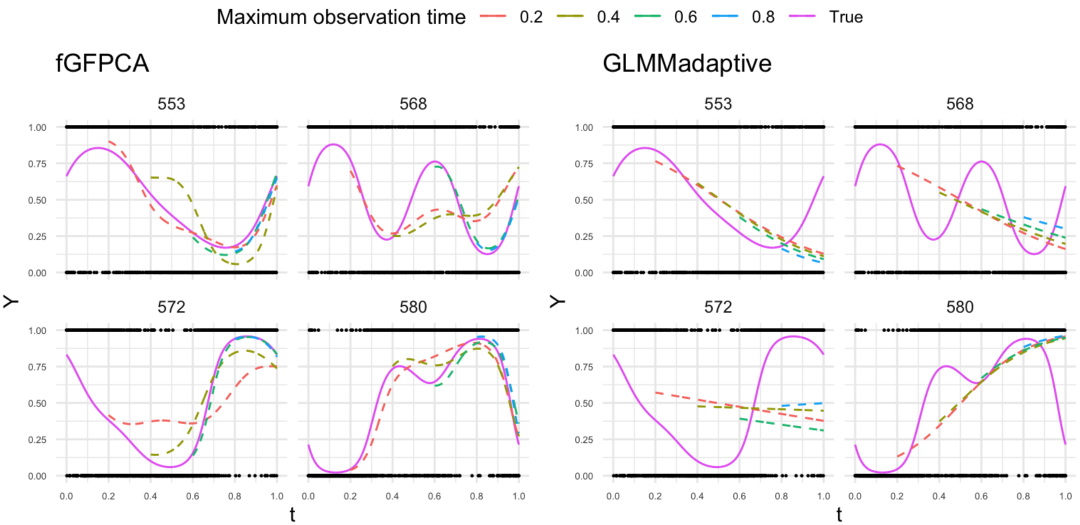


Simulated data
553
568
572
580
1.00
0.75
0.50
0.25
0.00
Y
0.0
0.2
0.4
0.6
0.8
1.0
t

# Evaluation

- Compare predictive performance of two methods
  - fGFPCA
  - GLMM using Adaptive Gaussian Quadrature (GLMMadaptive)
    - Limited in terms of model flexibility: $g(E(Y_i)) = \beta_0 + \beta_1 t + b_{i0} + b_{i1} t$
    - Incorporate spline basis: computationally unfeasible
- Evaluation metrics
  - Integrated Squared Error (ISE)
  - Area-Under-the-Receiver-Operator-Curve (AUC)

# Predictive performance



# Prediction performance

Table: Integrated squared error

| | Maximum observation time | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | fGFPCA | | | | GLMMadaptive | | | |
| Window | 0.2 | 0.4 | 0.6 | 0.8 | 0.2 | 0.4 | 0.6 | 0.8 |
| (0.2, 0.4] | 146.407 | | | | 387.708 | | | |
| (0.4, 0.6] | 183.967 | 74.977 | | | 291.579 | 269.799 | | |
| (0.6, 0.8] | 218.265 | 49.275 | 15.776 | | 315.778 | 282.736 | 278.242 | |
| (0.8, 1.0] | 108.918 | 77.981 | 17.747 | 12.005 | 563.011 | 477.485 | 597.746 | 600.34 |

# Prediction performance

Table: Area under the ROC curve

| | Maximum observation time | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | fGFPCA | | | | GLMMadaptive | | | |
| Window | 0.2 | 0.4 | 0.6 | 0.8 | 0.2 | 0.4 | 0.6 | 0.8 |
| (0.2, 0.4] | 0.748 | | | | 0.591 | | | |
| (0.4, 0.6] | 0.664 | 0.734 | | | 0.524 | 0.596 | | |
| (0.6, 0.8] | 0.715 | 0.790 | 0.803 | | 0.669 | 0.694 | 0.687 | |
| (0.8, 1.0] | 0.740 | 0.755 | 0.781 | 0.784 | 0.514 | 0.556 | 0.526 | 0.564 |

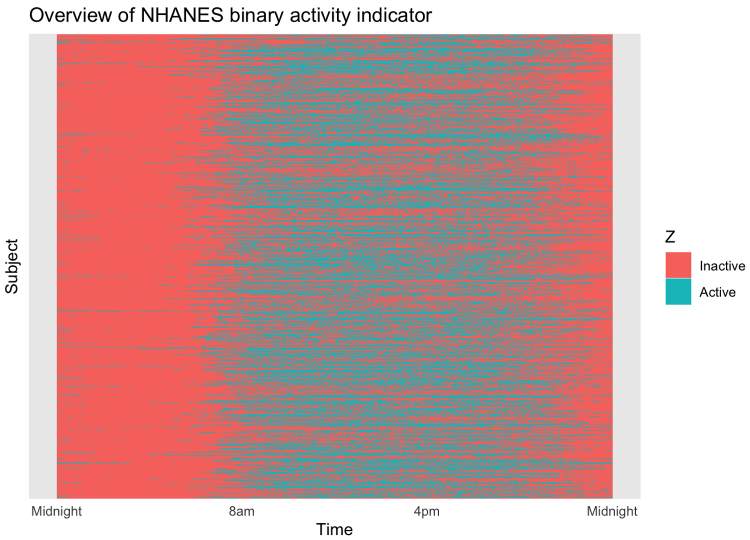
# Computation time

Table: Computation time (minutes)

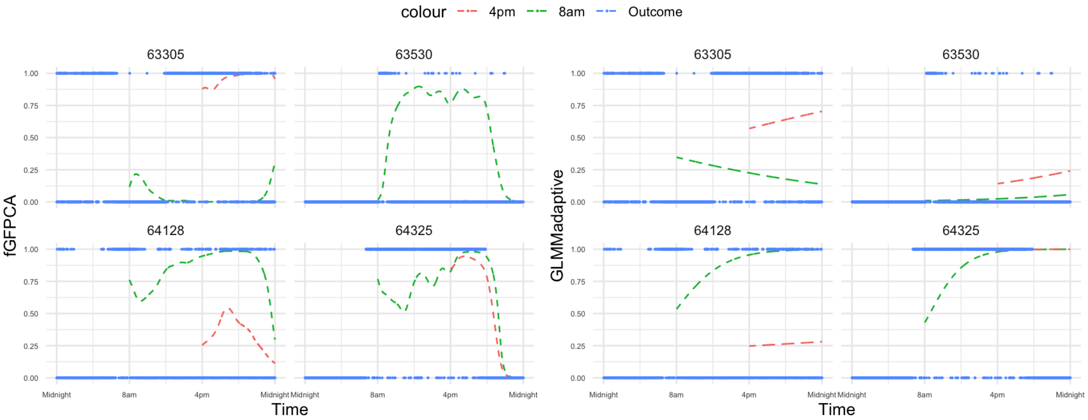
| Method | Fit | Prediction |
|---|---|---|
| fGFPCA | 0.725 | 1.592 |
| GLMMadaptive | 2.287 | 0.017 |

# Data application

# NHANES binary activity indicator



# Prediction performance



# Prediction performance

Table: Area Under the ROC curve

| | Maximum observation time | | | |
|---|---|---|---|---|
| | fGFPCA | | GLMMadaptive | |
| Window | 8am | 4pm | 8am | 4pm |
| 8am-4pm | 0.587 | | 0.628 | |
| 4am-midnight | 0.680 | 0.766 | 0.448 | 0.613 |

# Discussion

# Discussion

- fGFPCA can accommodate more flexible correlation structure between repeated measure
- fGFPCA reduced time spent on model fitting
- However, when ever larger dataset, fGFPCA is still not efficient enough
  - The GLMM model for re-evaluation of estimates in step 4
  - Laplace Approximation for out-of-sample prediction

**Thank you!**

# References


Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker.
Fitting linear mixed-effects models using lme4.
*Journal of Statistical Software*, 67(1):1–48, 2015.

Jeff Goldsmith, Fabian Scheipl, Lei Huang, Julia Wrobel, Chongzhi Di, Jonathan Gellar, Jaroslaw Harezlak, Mathew W. McLean, Bruce Swihart, Luo Xiao, Ciprian Crainiceanu, and Philip T. Reiss.
*refund: Regression with Functional Data*, 2022.
R package version 0.1-28.

Mithat Gonen and Glenn Heller.
Concordance probability and discriminatory power in proportional hazards regression.
*Biometrika*, 92(4):965–970, 2005.

Patrick J. Heagerty and Yingye Zheng.
Survival model predictive accuracy and roc curves.
*Biometrics*, 61(1):92–105, 2005.

# Appendix

# Example: binary data

- Maximize the posterior log-likelihood:

$$l(\boldsymbol{\xi}_u|\boldsymbol{Y}_u,\hat{\boldsymbol{\Theta}}) \propto l(\boldsymbol{Y}_u|\boldsymbol{\xi}_u,\hat{\boldsymbol{\Theta}}) + l(\boldsymbol{\xi}_u|\hat{\boldsymbol{\Theta}})$$

- Log-likelihood of $\boldsymbol{Y}_u$:

$$l(\boldsymbol{Y}_u|\boldsymbol{\xi}_u,\hat{\boldsymbol{\Theta}}) = \sum_{s=1}^{s_{max}} h_s\eta(s) - \sum_{s=1}^{s_{max}} n_s log(1+exp(\eta(s))), \quad \eta(s) = \hat{f}_0(s) + \sum_{k=1}^{K} \xi_{uk}\hat{\phi}_k(s)$$

  - $n_s$ indicates the number of observation in the sth bin
  - $h_s$ indicates the number of events/sucesses in the sth bin
  - $t_m$ is in bin $s_{max}$
- Log-likelihood of $\boldsymbol{\xi}_u$:

$$l(\boldsymbol{\xi}_u|\hat{\boldsymbol{\Theta}}) \propto -\boldsymbol{\xi}_u^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\xi}_u/2, \quad \boldsymbol{\Gamma} = \begin{bmatrix} \hat{\lambda}_1 & \dots \\ \dots & \dots \\ \dots & \hat{\lambda}_K \end{bmatrix}$$

# Example: binary data

$$l(\xi_u|\mathbf{Y}_u, \hat{\mathbf{\Theta}}) \propto \sum_{s=1}^{s_{max}} h_s \eta(s) - \sum_{s=1}^{s_{max}} n_s log(1 + exp(\eta(s)) - \xi_u^T \mathbf{\Gamma}^{-1} \xi_u / 2$$

$$\frac{dl(\xi_u|\mathbf{Y}_u, \hat{\mathbf{\Theta}})}{d\xi_u} = \sum_{s=1}^{s_{max}} h_s \phi(s) - \sum_{s=1}^{s_{max}} n_s \frac{exp(\eta(s))}{1 + exp(\eta(s))} \phi(s) - \xi_u^T \mathbf{\Gamma}^{-1} = 0$$

- The numeric solution of the score equation can be found efficiently