Dynamic Prediction with fPCA

2022-11-3

Dynamic prediction

- With observations up to t_m, predict outcomes (or probabilities of outcome) after that time point
- Prediction updates with new observations
- Mixed model prediction is difficult
  - Not flexible enough for densely measured data
  - Out-of-sample random effects cannot be estimated
- Functional mixed effects model

Functional Concurrent Regression (FCR)

- Goal: to predict future track based on partially observed track
- For a subject i, we observe a function over t

$$Y_i(t) = f_0(t) + b_i(t) + \epsilon_i(t)$$

- Subject-specific random effect

$$b_i(t) = \Sigma_{k=1}^{c} u_{ik} B_k(t)$$

where $\boldsymbol{u_i} \sim N(0, \Gamma)$

- We usually observe Y_i on a series of discrete t_{ij}

$$Y_{ij} = f_0(t_{ij}) + b_i(t_{ij}) + \epsilon_{ij}, \quad j = 1...J_i$$

where $\epsilon_{ij} \sim N(0, \sigma_{\epsilon^2})$.

Connection to fPCA

- When there is no covariate in the model, this is essentially a fPCA problem.
  - $\boldsymbol{B}$ is a matrix of eigenfunctions
  - $\boldsymbol{u}$ is a matrix of PC scores/loadings
- Use fPCA to estimate f_0, Γ and σ_ϵ
- For a new subject with observations up to t_m, estimate its score:

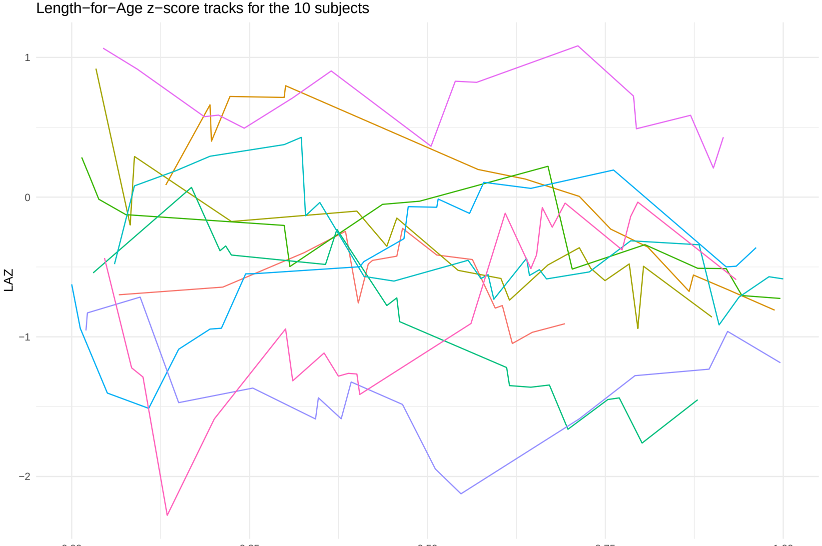
$$\hat{\boldsymbol{u}} = E(\boldsymbol{u}|\boldsymbol{y}) = \hat{\boldsymbol{\Gamma}}\boldsymbol{B}^{\boldsymbol{T}}(\boldsymbol{B}\hat{\boldsymbol{\Gamma}}\boldsymbol{B}^{\boldsymbol{T}} + \hat{\sigma_\epsilon}^2 \boldsymbol{I_m})^{-1}\boldsymbol{y}$$

With the estimated score, we can predict its outcome in following time points

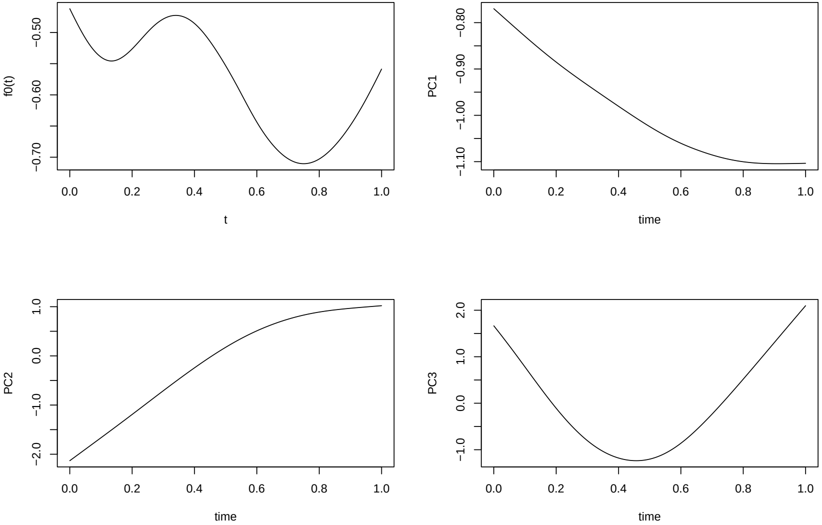
$$\hat{\boldsymbol{Y}} = \hat{\boldsymbol{f}}_0 + \boldsymbol{B}^T\hat{\boldsymbol{u}}$$

Simulated child growth data

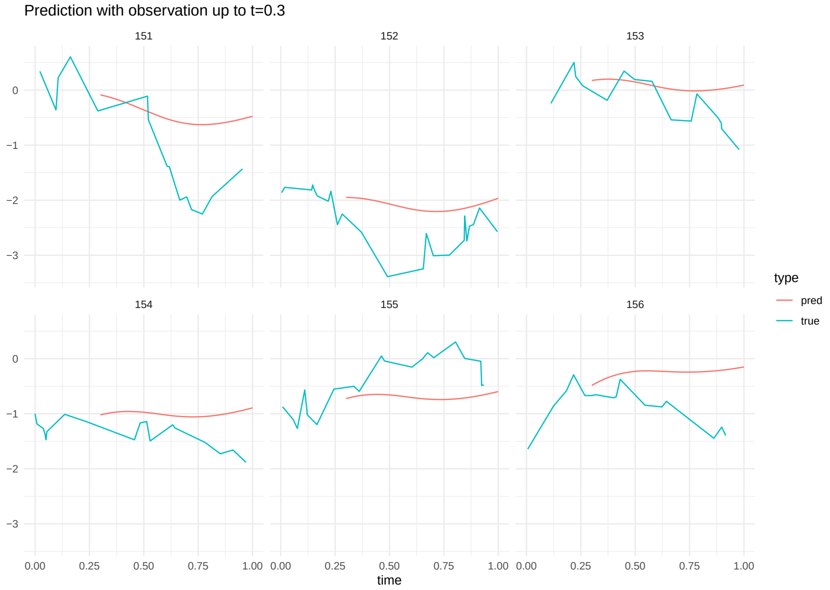
- Predict length-for-age z-score (LAZ), observed with noise
  - Calculated using the age- and sex-specific WHO standard references



FPCA on observed LAZ



Prediction of new partially observed sample



Prediction of new partially observed sample

Prediction of new partially observed sample

Extension to non-Gaussian data

- fPCA on non-Gaussian data is very difficult and computationally intensive
- fPCA on a latent Gaussian variable instead
  - transformation using link function: $g(E(Y)) = \boldsymbol{X}^T\beta$
- Get a smooth latent Gaussian function by pooling a series of GLMM
  - Assume we have regularly observed functions Y_{ij}, $i = 1 \dots N$ and $j = 1 \dots J$
  - Bin the functional domain into small intervals $t \in 1 \dots T$
  - Fit a GLMM in each bin to get a latent variables

$$g(E(Y_{it})) = \beta_{0t} + b_{it}$$

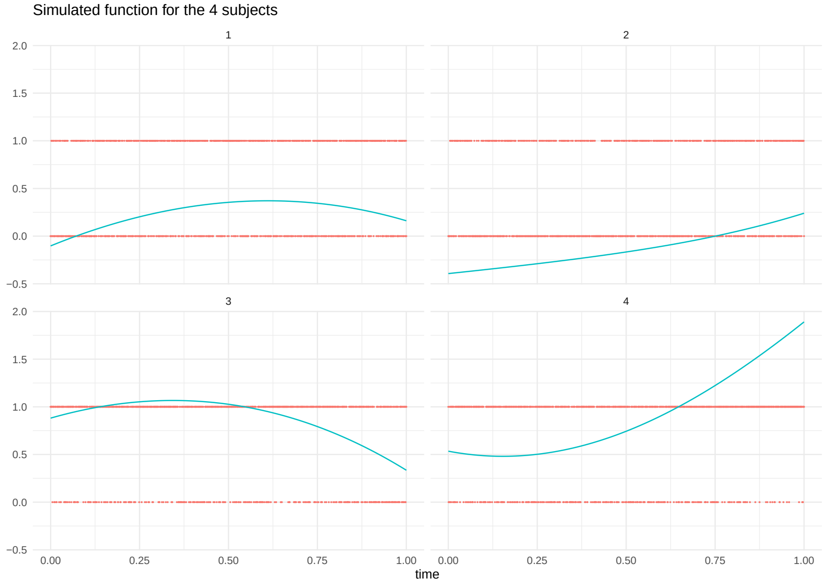
Let $Z_{it} = g(E(Y_{it}))$, we can do Gaussian fPCA on this latent variable

Extension to non-Gaussian data

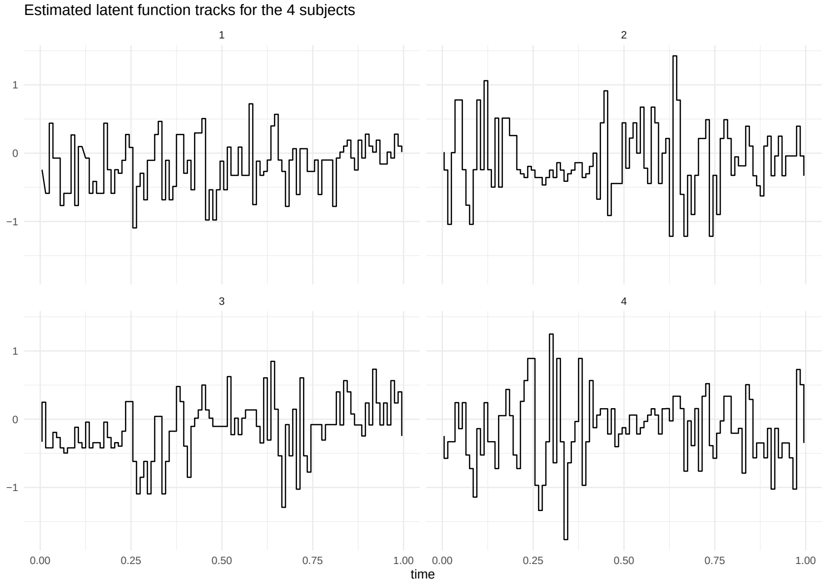
$$Z_i(t) = a_{i1} sin^2(t) + a_{i2} cos^2(t) + a_{i3} t^2 + a_{i4} t$$

$$Y_i(t) \sim Binomial(\frac{exp(Z_i(t))}{1 + exp(Z_i(t))})$$

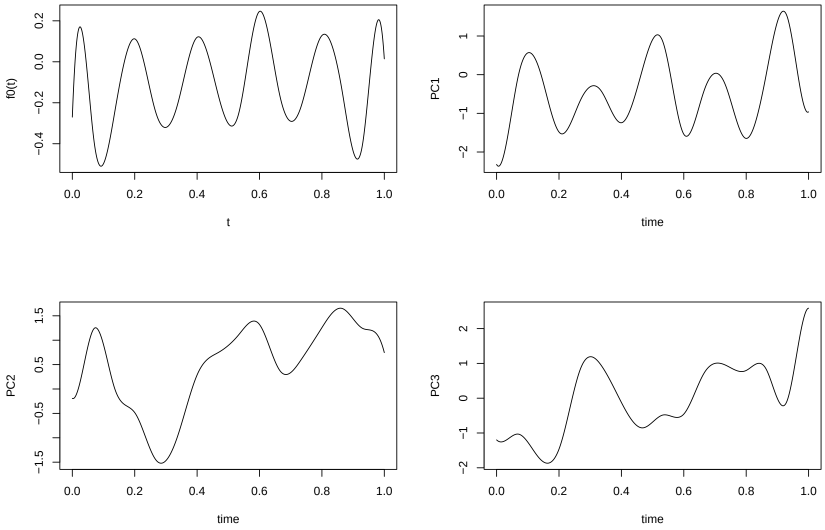
Extension to non-Gaussian data



Extension to non-Gaussian data



Extension to non-Gaussian data



Next steps

- Simulate a non-Gaussian function and implement methods above
  - Improve estiation of latent functions very well
  - Numeric problems
- Transform latent function back to non-Gaussian function
- Inclusion of covariates